Oxford Textbook of Surgery

Contents

Editors

Preface to the First Edition

Preface to the Second Edition

Contributors

Acknowledgements

Volume 1 | Volume 2 | Volume 3

VOLUME 1

Section 1 Response to injury
Douglas W. Wilmore

Section 2 Trauma and shock
David V. Feliciano, Richard J. Mullins, and Grace S. Rozycki

Section 3 Surgical nutrition
Timothy A. Pritts and Josef E. Fischer

Section 4 Surgical infections and AIDS

4.1 Surgical infections
Dietmar H. Wittmann, and Robert E. Condon

4.2 The surgery associated with HIV infection
C. Wastell and Peter A. Davis

Section 5 Blood

5.1 Haematological problems
Paul L. F. Giangrande and T. J. Littlewood

5.2 Blood transfusion in surgical practice
M. F. Murphy

Section 6 Wound healing
George W. Cherry, Margaret A. Hughes, Mark W. J. Ferguson, and David J. Leaper

Section 7 Molecular and cell biology in surgical practice
D. J. Weatherall

Section 8 Design and interpretation of clinical trials
William C. Wood

Section 9 Evidence-based approach to surgical decisions
Kieran J. O'Flynn, Miles Irving, and David Sackett

Section 10 Anaesthesia and the operating room

10.1 Anaesthesia for surgeons
John W. Sear and Oliver J. Dyar

10.2 Approaches to local and regional anaesthesia
Michael Dobson

10.3 Surgical diathermy
John T. B. Moyle

10.4 Electrical safety in anaesthesia and surgery
John T. B. Moyle

10.5 Interpretation of lung-function tests in the surgical patient
M. Jocelyn Morris

Section 11 Care of the critically ill patient

11.1 Guidelines and scoring systems
Christopher S. Garrard

11.2 Cardiovascular aspects
Christopher S. Garrard and Nigel R. A. Clarke

11.3 Renal aspects
Christopher S. Garrard

11.4 Respiratory aspects
Christopher S. Garrard

11.5 Infection in the intensive care unit
Christopher S. Garrard

11.6 Central nervous system aspects
Christopher S. Garrard

11.7 Gastrointestinal aspects
Christopher S. Garrard

11.8 Grief: breaking bad news: offering organ donation
Patricia M. Franklin

Section 12 Medical problems in the surgical patient

12.1 Respiratory problems
Rafael Barrera and Roger S. Wilson

12.2 Cardiological problems
Pierre Foëx

12.3 Renal problems
Christopher G. Winearls and Peter J. Ratcliffe

12.4 Hepatic problems
Susan N. Cullen, Carol A. Cobb, and R. W. Chapman

12.5 Psychological care
Owen S. Surman and Elizabeth Gray

Section 13 Imaging in surgical practice

13.1 Conventional radiology
Basil J. Shepstone

13.2 Computed tomography
Stephen Golding

13.3 Magnetic resonance imaging
Jeff Fidler and David D. Stark

13.4 Gastrointestinal radiology
Daniel J. Nolan

13.5 Ultrasound imaging
David R. M. Lindsell

13.6 Imaging in children
David R. M. Lindsell

13.7 Nuclear medicine techniques
Basil J. Shepstone

13.8 Nuclear magnetic resonance spectroscopy
Bheeshma Rajagopalan and Peter Styles

13.9 Positron-emission tomography
John M. Hoffman, Raghuveer K. Halkar, and Ernest V. Garcia

13.10 Imaging guidelines

Section 14 Endoscopy

14.1 Upper gastrointestinal endoscopy
Timothy Simon and Aaron S. Fink

14.2 Biliary and pancreatic endoscopy
Julian Britton

14.3 Colonoscopy
Kenneth A. Forde

14.4 Diagnostic laparoscopy
Barry A. Salky

14.5 Percutaneous endoscopic gastrostomy and jejunostomy
Timothy Simon and Aaron S. Fink

14.6 Bronchoscopy
Malcolm K. Benson

14.7 Thoracoscopy
Anthony P. C. Yim and Mohammad Bashar Izzat

Section 15 Endoscopic surgery
Timothy M. Farrell and John G. Hunter

Section 16 Transplantation

16.1 Basic transplantation immunology
Hugh Auchincloss, Jr.

16.2 Organ procurement
Francis L. Delmonico

16.3 Organ and tissue preservation for transplantation
Vernon Marshall, Alan Saunder, David Scott, Brian Howden and Paula Jablonski

16.4 Immunosuppression
Peter J. Morris

16.5 Kidney transplantation
Peter J. Morris

16.6 Liver transplantation
Sandy Feng, Dicken S. C. Ko and A. Benedict Cosimi

16.7 Heart and heart–lung transplantation
George V. Letsou and John C. Baldwin

16.8 Small-bowel transplantation
D. L. Bigam and D. R. Grant

16.9 Vascularized pancreatic transplantation
T. Pearson and C. Larsen

16.10 Pancreatic islet and fetal pancreas transplantation
Derek W. R. Gray

16.11 Xenogeneic transplantation
Hugh Auchincloss, Jr.

16.12 Reconstruction of the central nervous system by neural transplantation
Cesario V. Borlongan, Thomas B. Freeman, and Paul R. Sanberg

Section 17 The arteries

17.1 Pathobiology of atherosclerosis
Jawaharlal W. B. Senaratne and Fiona R. Green

17.2 Pathobiology of vasculitis

Jawaharlal W. B. Senaratne and Gavin P. Spickett

17.3 Vasculitis
Gavin P. Spickett

17.4 The non-invasive vascular diagnostic laboratory
David S. Sumner

17.5 Obliterative atheromatous disease

17.5.1 Aortoiliac occlusive disease
Joseph S. Giglia and David C. Brewster

17.5.2 Femoral and distal arteries
John A. Murie

17.5.3 Mesenteric arteries
Peter H. Lin and Thomas F. Dodson

17.6 Aneurysms

17.6.1 Abdominal aorta
Jack Collin

17.6.2 Femoral artery
Richard P. Cambria and Margaret L. Schwarze

17.6.3 The popliteal artery
Linda Hands

17.6.4 Carotid artery
Antony Fox

17.6.5 Subclavian artery
Antony Fox

17.6.6 Visceral arteries
Jack Collin

17.7 Extracerebral vascular disease

17.7.1 Carotid artery
Peter J. Morris

17.7.2 Vertebrobasilar, subclavian, and innominate arteries
Peter J. Morris

17.7.3 Carotid body tumours
Linda Hands

17.8 Vascular disease of the upper limb

17.8.1 Thoracic outlet obstruction
Peter J. Morris

17.8.2 Raynaud's syndrome
Peter M. Lamont

17.8.3 Sympathectomy in the upper limb
Jack Collin

17.9 Renovascular disease
Linda Hands and Peter J. Morris

17.10 Arterial emboli

17.10.1 Arterial emboli in limbs
Bruce Campbell

17.10.2 Arterial emboli: mesenteric arteries
Leslie W. Ottinger

17.11 Arterial and venous injuries
David V. Feliciano

17.12 Infection and arterial surgery
Christopher P. Conlon and Jack Collin

17.13 Minimally invasive vascular surgery
Victor J. Weiss and Alan B. Lumsden

17.14 Thrombolytic therapy
Farzin Adili and Glenn M. LaMuraglia

17.15 Neointimal hyperplasia, thrombophilia, and graft thrombosis
Changyi Chen and Alan B. Lumsden

17.16 Vascular access and other techniques for dialysis and chemotherapy
Derek W. R. Gray

17.17 Limb amputation
Derek W. R. Gray

Section 18 The veins

18.1 Venous disorders, vascular malformations, and chronic ulceration in the lower limbs
David J. Tibbs and John H. Scurr

18.2 Deep vein thrombosis and pulmonary embolism
Lazar J. Greenfield and Mary C. Proctor

VOLUME 2

Section 19 Abnormalities of the lymphatic system
Kevin G. Burnand and Catharine L. McGuinness

Section 20 Endocrine disease

20.1 Genetic aspects of endocrine disease

Terry C. Lairmore and Samuel A. Wells

20.2 The thyroid gland

20.2.1 Physiology of the thyroid gland
J. A. Franklyn

20.2.2 The thyroid gland
Gregory P. Sadler and Nicholas Dudley

20.3 The parathyroid gland

20.3.1 Parathyroid disorders: molecular genetics and physiology
R. V. Thakker

20.3.2 Surgery of the parathyroid glands
Gregory P. Sadler and Nicholas Dudley

20.4 The adrenal gland

20.4.1 The adrenal gland
Justin A. Roake

20.4.2 Surgery of the adrenal gland
Collin J. Weber and C. Daniel Smith

Section 21 The breast

21.1 Benign conditions of the breast
Michael J. Greenall

21.2 Cancer of the breast
Michael J. Greenall and William C. Wood

21.3 Abnormalities of the male breast
P. Jane Clarke and Linda Hands

Section 22 The oesophagus

22.1 Dysphagia
J. Shapiro

22.2 Benign oesophageal stricture, reflux, and hiatus hernia

22.2.1 Reflux disease and hiatus hernia
Reginald V. N. Lord and Tom R. DeMeester

22.2.2 Benign oesophageal strictures
Reginald V. N. Lord and Tom R. DeMeester

22.3 Diverticula and achalasia

22.3.1 Oesophageal diverticula
John C. Wain

22.3.2 Achalasia
John C. Wain

22.4 Oesophageal perforation, Boerhaave's syndrome, and Mallory–Weiss syndrome
Joseph I. Miller, Jr.

22.5 Benign and malignant tumours
Cameron D. Wright, Henning A. Gaissert, and Douglas Mathisen

Section 23 The stomach and duodenum

23.1 Peptic ulcer—stomach and duodenum
Alan G. Johnson

23.2 Carcinoma of the stomach
Hugh Barr and Michael J. Greenall

23.3 'Gastritis'
Alan G. Johnson

23.4 Gastric volvulus and acute gastric dilatation
Hugh Barr

23.5 Foreign bodies and bezoars
Michael N. Margolies

Section 24 The small intestine

24.1 Small bowel obstruction
Stephanie Rohovsky and Ronald Bleday

24.2 Crohn's disease

24.2.1 Crohn's disease of the small intestine
D. P. Jewell and J. Satsangi

24.2.2 Surgery for Crohn's disease of the small intestine
Timothy A. Cook and Neil Mortensen

24.3 Fistulas
Gordon Carlson and Miles Irving

24.4 Diverticular disease of the small bowel
Michael N. Margolies

24.5 Malabsorption syndromes
D. P. Jewell

24.6 Short bowel syndrome
A. Bianchi

24.7 Radiation enteritis
Christopher G. Willett and Marcelo Mester

24.8 Small-bowel tumours
T. R. Worthington and R. C. N. Williamson

24.9 Foreign bodies

Michael N. Margolies

24.10 Pneumatosis cystoides intestinalis
Andrew Mitchell

Section 25 Surgery for obesity
D. Michael Grace

Section 26 The colon and rectum

26.1 Diverticular disease: diverticulitis, bleeding, and fistula
Bruce D. George

26.2 Inflammatory disease of the colon

26.2.1 Pathology of non-specific chronic idiopathic inflammatory bowel disease
Bryan F. Warren

26.2.2 Colonic inflammation-diagnosis and management
D. P. Jewell

26.2.3 Surgical management of ulcerative colitis
Timothy A. Cook and Neil Mortensen

26.2.4 Crohn's disease of the colon
Neil Mortensen and Timothy A. Cook

26.3 Colorectal tumours
Alfred M. Cohen

26.4 Large-bowel obstruction
John P. Welch

26.5 Volvulus of the colon
Alan R. Berry

26.6 Chronic constipation in adults
Neil Mortensen and Timothy A. Cook

26.7 Angiodysplasia
Bruce D. George and Paul C. Shellito

26.8 Prolapse of the rectum
Timothy A. Cook and Neil Mortensen

Section 27 The appendix

27.1 Acute appendicitis
Charles M. Ferguson

27.2 Primary appendiceal malignancies
Charles A. Staley

Section 28 The anus

28.1 Haemorrhoids or piles
W. Hamish Thomson

28.2 Pruritus ani
C. G. Marks

28.3 Anal fissure
Emin A. Carapeti and R. K. S. Phillips

28.4 Anorectal abscesses and fistula-in-ano
Julio Garcia-Aguilar and David A. Rothenberger

28.5 Faecal incontinence
Norman S. Williams and B. James Mander

28.6 Cancer of the anus
Michael J. Greenall

Section 29 Sacrococcygeal pilonidal sinus
Jan Rakinic and Robert Fry

Section 30 The liver

30.1 Surgical anatomy of the liver and biliary tree
Adrian Savage and Julian Britton

30.2 Hepatic trauma
David V. Feliciano

30.3 Abscesses — pyogenic and amoebic
Robert H. Rubin

30.4 Cancers of the liver
George W. Daneker, Jr. and Mark S. Talamonti

30.5 The Budd-Chiari syndrome
J. Michael Henderson

Section 31 The biliary tract

31.1 Benign diseases of the biliary tract
Julian Britton, Kenneth I. Bickerstaff, and Adrian Savage

31.2 Malignant diseases of the biliary tract
Steven A. Ahrendt and Henry A. Pitt

Section 32 Portal hypertension

32.1 Aetiology, presentation, and investigation
George Hamilton

32.2 Emergency treatment of bleeding oesophageal varices
J. R. Galloway

32.3 Surgical shunt procedures
J. R. Galloway

32.4 Non-shunt procedures in management of variceal bleeding

George Hamilton

32.5 Ascites and its complications
Robert M. Strauss and Jules L. Dienstag

Section 33 The pancreas

33.1 Acute pancreatitis
Oscar Joe Hines and Howard A. Reber

33.2 Chronic pancreatitis
C. W. Imrie and S. J. Govil

33.3 Trauma to the pancreas
Alan R. Berry

33.4 Pancreatic cancer
Charles J. Yeo and John L. Cameron

33.5 Insulinomas and other tumours
Jeffrey A Norton and Hop Le

Section 34 The abdomen

34.1 The acute abdomen
Julian Britton

34.2 The peritoneum and peritonitis
D. McWhinnie

34.3 Trauma: injuries to the abdominal wall and mesentery
Ralph L. Warren and Charles J. McCabe

34.4 The omentum
D. McWhinnie

34.5 Retroperitoneal fibrosis
David Cranston

34.6 Retroperitoneal neoplasms
Herbert C. Hoover, Jr.

Section 35 Hernias of the abdominal wall

35.1 Inguinal and femoral hernias
Clare Cheek and Andrew Kingsnorth

35.2 Umbilical, epigastric, and rarer abdominal wall hernias
C. Daniel Smith

35.3 Wound dehiscence, incisional hernia, and parastomal hernia
Adrian Savage and Peter M. Lamont

Section 36 Gynaecology and obstetrics in general surgery

36.1 Acute vaginal bleeding
I. Z. Mackenzie

36.2 The management of acute pelvic pain
Michael D. G. Gillmer and Bruce C. Dunphy

36.3 Acute pelvic sepsis and tubo-ovarian abscess
Robert G. Forman

36.4 Ovarian accidents
Michael D. G. Gillmer

36.5 Ovarian tumours
Ira R. Horowitz

36.6 Uterine fibroids
William D. Boyd and F. Mark Charnock

36.7 Endometriosis
David H. Barlow

36.8 Cancer

36.8.1 Uterine malignancy
C. O. Granai, Walter H. Gajewski, and Bruno Arena

36.8.2 Cervical malignancy
C. O. Granai, Stephen S. Falkenberry, Hector M. Tarraza, Walter H. Gajewski, Bruno Arena, Margarett Ellison, and Michael A. Steller

36.8.3 Radical hysterectomy
C. O. Granai and Hector M. Tarraza

36.9 Perineal abscesses
William D. Boyd and F. Mark Charnock

36.10 Late spontaneous abortion
I. Z. Mackenzie

36.11 Caesarean section
Gordon M. Stirrat

36.12 Third-stage complications
Michael D. G. Gillmer

36.13 Surgery for infertility
Arlene J. Morales and John A. Rock

36.14 Rupture of the pregnant uterus
J. P. Anthony and F. Mark Charnock

Section 37 Soft tissue tumours
William C. Wood

Section 38 Skin

38.1 Hamartomas, vascular malformations, and congenital defects
Calvin O. McCall and Thomas G. Cropley

38.2 Benign tumours and cysts of the skin and its appendages

Calvin O. McCall and Thomas G. Cropley

38.3 Benign acquired pigmented lesions of the skin
Arthur J. Sober and Richard G. B. Langley

38.4 Premalignant and malignant tumours of the skin
Calvin O. McCall and Thomas G. Cropley

38.5 Cutaneous melanoma

38.5.1 Recognition, staging, and prognosis of cutaneous melanoma
Arthur J. Sober and Richard G. B. Langley

38.5.2 Surgical management of malignant melanoma
Grant W. Carlson

38.6 Cutaneous problems occurring in the surgical patient
David A. Wrone, Richard G. B. Langely, and Arthur J. Sober

38.7 Therapies used in dermatology
Calvin O. McCall and Thomas G. Cropley

Section 39 Urology

39.1 Genitourinary anatomy
Francis J. McGovern and John D. Seigne

39.2 Voiding function and dysfunction
Michael H. Safir, Angelo E. Gousse, and Shlomo Raz

39.3 Trauma to the genitourinary tract
David M. Rudnick and Alex F. Althausen

39.4 Urinary complications of spinal abnormalities and injuries
Marcus Drake and Jeremy Noble

39.5 Urinary stone disease
David M. Rudnick and Stephen P. Dretler

39.6 Male sexual dysfunction and infertility
Lewis Kriteman and Jerry Yuan

39.7 Benign prostate disease
D. Cranston and Damian Hanbury

39.8 Cancer of the prostate
John Naitoh, Jean B. deKernion, and Daniel Shoskes

39.9 Tumours of the urinary tract

39.9.1 Tumours of the kidney
W. Turner and D. Cranston

39.9.2 Urothelial tumours
T. S. O'Brien

39.9.3 Tumours of the testicle
Sam D. Graham, Jr. and D. Cranston

39.9.4 Carcinoma of the penis
D. Cranston

39.10 Torsion of the testicle
D. Cranston

39.11 Vasectomy
D. Cranston

Section 40 Cardiac surgery

40.1 Development of surgery of the heart and great vessels
Stephen Westaby

40.2 Cardiopulmonary bypass and myocardial protection
Stephen Westaby

40.3 Circulatory support for the failing heart
Francisco A. Arabía, Jack G. Copeland, and Richard G. Smith

40.4 Vascular and cardiac biological and non-biological prostheses
Graeme L. Hammond and Michael A. Coady

40.5 Principles of Blood conservation, cardiac anaesthesia, and cerebral protection
Catherine R. Grebenik and Michael E. Sinclair

40.6 Cardiovascular monitoring and postoperative care
Catherine R. Grebenik and Michael E. Sinclair

40.7 Congenital heart disease

40.7.1 Management of congenital heart disease
Nicholas Archer and Ingegerd Östman-Smith

40.7.2 Patent ductus arteriosus and aortic coarctation
Emile A. Bacha and Gus J. Vlahakes

40.7.3 Atrial, ventricular, and atrioventricular septal defects
Ravi Pillai and Kenny Sin

40.7.4 Malalignment anomalies of the interventricular septum
Andrew Parry and David Hamilton and Pankaj Mankad

40.7.5 Abnormalities of the atrioventricular valves
Ravi Pillai and Kenny Sin

40.7.6 Stenosis and atresia of the arterial valves
Marshall L. Jacobs

40.7.7 Transposition of the great arteries
Lester C. Permut and Hillel Laks

40.7.8 Congenital anomalies of the thoracic aorta and main pulmonary arteries
Stephen Westaby

40.7.9 Anomalies of pulmonary and systemic venous return

Emile A. Bacha and Gus J. Vlahakes

40.7.10 Miscellaneous congenital cardiac defects
Arvind K. Agnihotri and Gus J. Vlahakes

40.8 Coronary artery disease

40.8.1 Coronary artery disease: risk factors, diagnosis, and medical treatment
J. C. Forfar and B. Gribbin

40.8.2 Interventional cardiology
Douglas C. Morris

40.8.3 Coronary artery bypass
Robert A. Guyton

40.8.4 Surgery for complications of ischaemic heart disease
Matthew J. Wall Jr. and John C. Baldwin

40.9 Acquired valvular disease

40.9.1 Pathology of cardiac valves
M. J. Davies

40.9.2 Surgery for acquired aortic and pulmonary valve disease
Graeme L. Hammond and Michael A. Coady

40.9.3 Diagnosis and surgical management for tricuspid and mitral valve disease
Kenneth L. Franco and Graeme L. Hammond

40.10 Surgery of the thoracic aorta
Stephen Westaby

40.11 Surgery for cardiac rhythm disturbances
T. Bruce Ferguson, Jr.

40.12 Miscellaneous cardiac diseases
Andrew Parry

VOLUME 3

Section 41 Thoracic surgery

41.1 Surgical anatomy and radiology of the chest
Stephen Westaby and B. Trotman-Dickenson

41.2 The trachea
Douglas Mathisen and Hermes C. Grillo

41.3 Benign and malignant tumours of the lung
John C. Wain

41.4 Primary chest wall neoplasms
Edwin C. McGee, Jr., Richard I. Whyte, and Douglas Mathisen

41.5 Mediastinal tumours
Henning A. Gaissert and Douglas Mathisen

41.6 The diaphragm
Daniel P. Doody and Ashley C. Moncure

41.7 Lung volume reduction surgery and pulmonary transplantation
Joseph I. Miller, Jr.

41.8 Pneumothorax
Ashley C. Moncure

41.9 Chylothorax
Ashley C. Moncure

41.10 Respiratory infections
Malcolm K. Benson

41.11 Principles and practice of thoracic drainage
Stephen Westaby

41.12 Thoracic trauma
Stephen Westaby

Section 42 Paediatric surgery

42.1 General care of the paediatric patient, especially the newborn infant
Andrew R. Wilkinson

42.2 Thoracic surgical problems
Robert C. Shamberger and W. Hardy Hendren

42.3 Evaluation of acute abdominal pain in children
Richard R. Ricketts

42.4 Disorders of the gastrointestinal tract in children
D. Mervyn Griffiths

42.5 Duplications of the gastrointestinal tract and mesenteric cysts
Robert C. Shamberger

42.6 Paedatric hepatobiliary disorders — surgical aspects
Eric L. Lazar and R. Peter Altman

42.7 Anorectal disorders

42.7.1 Hirschsprung's disease
Richard R. Ricketts

42.7.2 Other anorectal problems
Richard R. Ricketts

42.8 The abdominal wall

P. K. H. Tam

42.9 Surgery of the urinary tract in children
W. Hardy Hendren and Joseph G. Borer

42.10 Paediatric solid tumours
Robert C. Shamberger

42.11 Fetal surgery
Craig T. Albanese and Michael R. Harrison

42.12 Paediatric otolaryngology
Ellen M. Friedman and Ronald W. Deskin

42.13 Spina bifida
Gordon Tang and Daniel L. Barrow

Section 43 Spleen and haemopoietic tissues

43.1 Surgery of the spleen
P. Jane Clarke and Peter J. Morris

43.2 Lymph node diagnosis
Armando E. Giuliano and Pond R. Kelemen

43.3 Surgery in patients with leukaemia
P. Jane Clarke

Section 44 Neurosurgery

44.1 General introduction to neurosurgery
C. B. T. Adams

44.2 Head injury
G. M. Teasdale and L. T. Dunn

44.3 Intracranial tumours
Jeffrey J. Olson, Daniel L. Barrow, Mark R. Gilbert, and Nelson Oyesiku

44.4 Surgery for epilepsy
C. B. T. Adams

44.5 Surgery for Parkinson's disease and other movement disorders
Roy A. E. Bakay, Mahlon R. DeLong, and Jerrold L. Vitek

44.6 Trigeminal neuralgia
C. B. T. Adams

44.7 Intracranial infection
C. B. T. Adams

44.8 Hydrocephalus
Daniel L. Barrow

44.9 Intracranial aneurysms and arteriovenous malformations
Thomas A. D. Cadoux-Hudson and Andrew Molyneux

44.10 Spinal surgery
C. B. T. Adams

44.11 Neurosurgery for intractable pain
Peter J. Teddy

Section 45 Head and neck surgery

45.1 Eye emergencies
Bradford J. Shingleton and Mariana D. Mead

45.2 Nose and sinuses
Richard R. Orlandi and David W. Kennedy

45.3 The evaluation of a neck mass in the adult patient
C. M. Norris, Jr. and Marvin P. Fried

45.4 Salivary glands
Joseph Califano and David W. Eisele

45.5 Oral and oropharyngeal cancer
C. M. Norris, Jr., James W. Lucarini and Marshall Posner

45.6 Nasopharyngeal cancer
Walter W. K. King and Arthur K. C. Li

45.7 Cancer of the larynx and hypopharynx
Mark C. Weissler

45.8 Surgery of the ear and temporal bone
Andrew P. Freeland and Martin J. Burton

45.9 Temporomandibular joint dysfunction
Richard P. Juniper

45.10 Oral and maxillofacial surgery
Joseph W. Wilkes, David A. Keith, Carol A. Lorente, and John A. Buehler

45.11 Facial injuries
R. J. Gates and S. A. Wall

45.12 The surgical treatment of snoring and obstructive sleep apnoea
Robert J. O. Davies, Grant Bates, and John R. Stradling

45.13 Pharyngeal pouch surgery
Grant Bates

45.14 Surgical technique of tracheostomy, minitracheostomy, and cricothyroidotomy
Martin. J. Burton

Section 46 Orthopaedics

46.1 Sports medicine and athletic injuries
Eric W. Carson and Brant Bair

46.2 Adult Orthopaedics

46.2.1 Musculoskeletal trauma

Andrew H. Schmidt and Marc F. Swiontkowski

46.2.2 The adult shoulder
A. J. Carr

46.2.3 The hand and wrist
Edward Akelman and Arnold-Peter C. Weiss

46.2.4 The arthritic hip
James R. Roberson

46.2.5 The knee
Richard S. Limbird and Michael Belanger

46.2.6 Degenerative disorders of the spine
Philip R. Lucas

46.2.7 Spinal trauma
Philip R. Lucas

46.2.8 Inflammatory disorders of the spine
Thomas E. Whitesides

46.2.9 Spinal infections
Mitchell F. Reiter and John G. Heller

46.2.10 Spinal cord monitoring
Jerrold Rosenberg

46.3 Paediatric orthopaedics

46.3.1 Paediatric fractures
Patricia M. Solga

46.3.2 Developmental dysplasia of the hip
Michael G. Ehrlich and L. Brett Babat

46.3.3 Legg-Calvé-Perthes' disease
Michael G. Ehrlich and Craig P. Eberson

46.3.4 Slipped femoral capital epiphysis
Michael G. Ehrlich and Dennis C. Crawford

46.3.5 Congenital foot disorders
Michael G. Ehrlich and Justin K. Greisberg

46.3.6 Congenital spinal deformity
John B. Emans

46.4 Management of bone tumours
Dempsey S. Springfield and Henry J. Mankin

Section 47 Some special aspects of surgery in developing countries

47.1 Surgery of bacterial disease

47.1.1 Abscess, cellulitis, and necrotizing bacterial infections
Ian O'Rourke, Phillip Carson, and Gary Lum

47.1.2 Typhoid fever and other salmonella infections
Christopher S. Grant, William J. K. Huizinga, and A. S. Daar

47.1.3 Abdominal tuberculosis
N. Rangabashyam, B. S. Anand, and R. OmPrakash

47.1.4 Tropical ulcer
H. Foster

47.1.5 Otitis media in high-risk populations
Peter S. Morris

47.1.6 Leprosy
John Hargrave

47.1.7 Pyogenic liver abscess
Pai-Ching Sheen and King-Teh Lee

47.2 Surgery of protozoal and parasitic disease

47.2.1 Amoebiasis
Prasit Watanapa ,Arun Pausawasdi, and R. C. N. Williamson

47.2.2 Schistosomiasis (bilharziasis)
A. S. Daar and Euan M. Scrimgeour

47.2.3 Schistosomiasis (bilharziasis) of the genitourinary tract
Imtiaz Husain

47.2.4 Hydatid disease
William K. J. Huizinga, Christopher S. Grant, and A. S. Daar

47.2.5 Ascariasis and other intestinal nematode infections
Oluwatope A. Mabogunje and A. S. Daar

47.2.6 Biliary and intestinal trematodes
Oluwatope A. Mabogunje, Christopher S. Grant, and A. S. Daar

47.2.7 Chyluria
A. P. Pandey

47.3 Splenic surgery in the tropics
H. Foster

47.4 Obstetric vesicovaginal fistulas
Gordon Williams and Ambaye Michael Wolde

47.5 Common surgical eye conditions in the developing world
Gordon J. Johnson and Allen Foster

47.6 Intrahepatic stones
C. M. Lo, S. T. Fan, and J. Wong

47.7 Some special aspects of plastic surgery in developing countries
T. E. E. Goodacre

47.8 Surgery of advanced disease and late presentation
A. S. Daar and Arjuna Aluwihare

47.9 Surgery with limited resources
A. S. Daar, S. Adib Rizvi, and S. Anwar Naqvi

47.10 Special paediatric problems in developing countries
Hugh Grant

Section 48 Cancer

48.1 Cancer genetics
K. L. Nathanson, T. C. Smyrk, M. S. Brose, H. T. Lynch, and B. L. Weber

48.2 Avoidable cancer
Richard Doll

48.3 Chemotherapy of cancer
I. Craig Henderson

48.4 Cancer gene therapy
James S. Economou

48.5 Emergencies in cancer
James Carmichael

48.6 Access for chemotherapy
Charles A. Staley

48.7 Irradiation in cancer

48.7.1 Principles of radiation oncology
David Morris and Joel Tepper

48.7.2 Principles of radiation protection
Edward W. Webster

Section 49 Pain relief

49.1 Care of the dying patient
Robert G. Twycross

49.2 Cancer pain relief
Robert G. Twycross

49.3 Relief of chronic non-malignant pain
Henry McQuay

Section 50 Plastic surgery

50.1 Principles and practice of plastic surgery
Michael J. Yaremchuk, and G. Gregory Gallico III

50.2 Craniofacial reconstruction for congenital, traumatic, and neoplastic conditions
Michael D. Poole

50.3 Acute burns
Robert L. Sheridan and Ronald G. Tompkins

Section 51 How should surgical trainees be selected?
Thomas F. Dodson

Edited By

Sir Peter J. Morris
Nuffield Professor of Surgery,
University of Oxford

and

William C. Wood
Joseph Brown Whitehead Professor and Chairman,
Department of Surgery,
Emory University School of Medicine,
Atlanta

Preface to the First Edition

The *Oxford Textbook of Surgery* aims to present a picture of surgery in its totality. To achieve this it has brought together the experience of surgical practice from two major clinical schools on opposite sides of the Atlantic, namely the Oxford University Clinical School and the Massachusetts General Hospital Clinical School of the Harvard Medical School, for most of the contributors are on the staff of one or the other or have worked in one or other institution in the past. In some instances former visiting Professors have been brought into play but the vast majority of contributors can claim some allegiance to one or other institution. We have found, not surprisingly perhaps, that, with one or two notable exceptions, such as, for example, in the approach to cancer of the prostate, the practice of surgery in the United Kingdom and the United States does not differ remarkably. Each subject has been covered by an expert in the field writing from his or her own wisdom and experience, but remembering past needs and difficulties during training days.

In this book we have approached surgery from a practical point of view and have attempted to cover most aspects of general surgery (an entity less easily defined nowadays), as well as giving an overview of the various specialist branches of surgery ranging from orthopaedics and neurosurgery through cardiothoracic surgery to plastic surgery. To achieve detailed coverage of every aspect of surgery in one textbook is obviously impossible, but our aims were to produce a book which could be used as a source of reference by general surgeons, either trained or in training, both in the Western world and in developing countries, and equally to provide a ready source of reference for surgeons in specialist branches of surgery, such as neurosurgery and orthopaedic surgery. We hope that with this approach general surgeons will not only get a feeling for current practice in their own areas of interest but will also be able to obtain sufficient information about problems presenting in their patients that involve, or might involve, another speciality. Similarly the specialist surgeon can also find sufficient material in the Textbook concerning some problem in general surgery or another specialist branch of surgery to be able to reassure a patient or to use for teaching purposes. We feel strongly that there is still a major need for a comprehensive textbook of surgery in this age of increasing specialization. Our approach should also prove invaluable to the medical undergraduate as well as to the surgeon practising in developing countries, where the general surgeon remains truly a generalist.

We have endeavoured to make the text become alive and attractive by the widespread use of colour for illustrations and tables, a major innovation for a textbook of surgery. This aim has certainly been achieved in our opinion by our publishers, Oxford University Press. Our thanks are due to their staff.

We are grateful to all our contributors who have been most patient during the gestation of the book. We especially wish to thank Mr Steve Westaby and Dr John Baldwin who organized the cardiac surgery section, Mr Chris Adams who planned the neurosurgery section, Dr Steve Dretler for the urology section, and Dr Michael Ehrlich of Brown University (formerly MGH) who, with his department, dealt with the whole of the orthopaedic section.

We hope that you will find this book an attractive and readily approachable treatise on surgery in its entirety, whether you be practising, training, or studying in the Western world or indeed in a developing country with limited resources.

Peter J. Morris
Oxford

Ronald A. Malt
Boston

Preface to the Second Edition

Following the enormous success of the first edition of the *Oxford Textbook of Surgery* it appeared inevitable that a second edition would be called for, and so proved to be the case. A great deal has changed in the past six years, not the least being the selection of a new North American editor, William C. Wood, to replace Ron Malt on his retirement. This has added a new dimension to the contributions which, in the first edition, reflected to a considerable extent thinking from the Massachusetts General Hospital and Oxford. In this second edition the contributors are drawn from a much wider field with obviously a greater input from Emory as well as elsewhere in North America, and also from further afield in the United Kingdom and the rest of the world. Again, it is striking how similar the approaches to surgery are on both sides of the Atlantic.

As before, we have tried to present surgery in its totality. We feel that for the practising surgeon, be he/she a specialist surgeon or a general surgeon practising in a rural area, there is still need for a textbook to which surgeons can refer easily, either for specific information about their own interests, or to ascertain the state of the art in another discipline. Although this book should serve as a source of reference for the surgical trainee, it was our intention to include sufficient detail so that experienced surgeons will find it more useful than a basic textbook of surgery.

There have been major changes in this edition with the inclusion of much new material such as sections on Response to injury, Trauma and shock, and Surgical nutrition, which were not covered comprehensively in the first edition. In addition, there are new and important sections directed at evidence-based approaches to surgical decisions, and molecular biology. The explosion in minimally invasive surgery, which had not got into its full stride at the time of the first edition, is now dealt with extensively in appropriate areas. Virtually all sections have been updated or completely rewritten, and all authors were asked to be sure that they had included data from randomized trials in their areas of expertise, together with complete references in suggested Further Reading at the end of each chapter.

A criticism of textbooks is that they soon become out of date or even may miss important contributions which appear between the preparation of a chapter and its appearance in print a year later. We accept that this is true in part, but much of what we do in surgery does not change rapidly and hence a textbook such as this serves as a valuable source of reference. In addition, all readers should know where they can obtain more information, such as systematic reviews and reports of randomized trials that may have appeared after the book was published or even while it was being published. This is covered in detail in the section entitled ‘Evidence-based approach to surgical decisions’, but we recommend that readers of this, or indeed any textbook, get into the habit of using the Cochrane Library (published by Update Software, Oxford) and in particular The Cochrane Database of Systematic Reviews and The Cochrane Controlled Trials Register.

The first edition was the first surgical textbook to be produced in full colour and once again Oxford University Press have done a magnificent job in producing an even more beautiful tome to read and appreciate. We hope it will be an adornment to the bookshelves of all surgeons, not only as source of knowledge, but also as a work of art.

TABLE cellSpacing=0 cellPadding=0 width="100%" border=0> Peter J. Morris William C. Wood University of Oxford Emory University

Contributors

C. B. T. Adams Senior Consultant Neurosurgeon, Radcliffe Infirmary, Oxford, UK

44.1 General introduction to neurosurgery

44.4 Surgery for epilepsy

44.6 Trigeminal neuralgia

44.7 Intracranial infection

44.10 Spinal surgery

Farzin Adili Division of Vascular Surgery, Massachusetts General Hospital, Boston, USA

17.14 Thrombolytic therapy

Arvind K. Agnihotri Cardiac Surgical Unit, Massachusetts General Hospital, Boston, USA

40.7.10 Miscellaneous congenital cardiac defects

Steven A. Ahrendt Department of Surgery, Medical College of Wisconsin, Milwaukee, USA

31.2 Malignant diseases of the biliary tract

Edward Akelman Vice Chairman of Orthopedics, University of Orthopedics, Providence, Rhode Island, USA

46.2.3 The hand and wrist

Craig T. Albanese Department of Pediatric Surgery, UCSF Fetal Treatment Center, San Francisco, California, USA

42.11 Fetal surgery

Alex F. Althausen Department of Urology, Massachusetts General Hospital, Boston, USA

39.3 Trauma to the genitourinary tract

R. Peter Altman Director, Pediatric Surgery, Babies and Children's Hospital of New York, New York, USA

42.6 Paediatric hepatobiliary disorders - surgical aspects

Arjuna Aluwihare Vidya Jothi Professor of Surgery, University of Peradeniya and Chairman, Board of Study in Surgery at the Postgraduate Institute of Medicine, Sri Lanka

47.8 Surgery of advanced disease and late presentation

B. S. Anand Professor, Digestive Diseases Section, Baylor College of Medicine, Houston, USA

47.1.3 Abdominal tuberculosis

J. P. Anthony Consultant Obstetrician and Gynaecologist, Northampton General Hospital NHS Trust, UK

36.14 Rupture of the pregnant uterus

Francisco A. Arabia Section of Vascular and Thoracic Surgery, University of Arizona Health Sciences Center, Tucson, USA

40.3 Circulatory support for the failing heart

Nicholas Archer Consultant Paediatric Cardiologist, John Radcliffe Hospital, Oxford, UK

40.7.1 Management of congenital heart disease

Bruno Arena Visiting Fellow in Gynecologic Oncology, Women's & Infants' Hospital, Brown University, Providence, Rhode Island, USA

36.8.1 Uterine malignancy

36.8.2 Cervical malignancy

Hugh Auchincloss Jr. Transplantation Unit, Massachusetts General Hospital, Boston, USA

16.1 Basic transplantation immunology

16.11 Xenogeneic transplantation

L. Brett Babat Rhode Island Hospital, Providence, USA

46.3.2 Developmental dysplasia of the hip

Emile A. Bacha Harvard Medical School Cardiothoracic Program, Massachusetts General Hospital, Boston, USA

40.7.2 Patent ductus arteriosus and aortic coarctation

40.7.9 Anomalies of pulmonary and systemic venous return

Brant Bair Department of Orthopedics, Emory University School of Medicine, Atlanta, Georgia, USA

46.1 Sports medicine and athletic injuries

Roy A. E. Bakay Vice Chairman of Neurosurgery, The Emory Clinic, Inc., Atlanta,Georgia, USA

44.5 Surgery for Parkinson's disease and other movement disorders

John C. Baldwin Dean, Dartmouth Medical School, Hanover, New Hampshire, USA

16.7 Heart and heart-lung transplantation

40.8.4 Surgery for complications of ischaemic heart disease

David H. Barlow Nuffield Professor of Obstetrics and Gynaecology, John Radcliffe Hospital, Oxford, UK

36.7 Endometriosis

Hugh Barr Professor and Dean, Cranfield Postgraduate Medical School in Gloucestershire, Gloucester Royal Hospital, UK

23.2 Carcinoma of the stomach

23.4 Gastric volvulus and acute gastric dilatation

Rafael Barrera Memorial Sloan-Kettering Cancer Center, New York, USA

12.1 Respiratory problems

Daniel L. Barrow Chairman of Neurosurgery, The Emory Clinic Inc., Atlanta, Georgia, USA

42.13 Spina bifida

44.3 Intracranial tumours

44.8 Hydrocephalus

Grant Bates Consultant Otolaryngologist and Head and Neck Surgeon, The Radcliffe Infirmary, Oxford, UK

45.12 The surgical treatment of snoring and obstructive sleep apnoea

45.13 Pharyngeal pouch surgery

Michael Belanger University of Orthopedics, Providence, Rhode Island, USA

46.2.5 The knee

Malcolm K. Benson Consultant Physician, Osler Chest Unit, The Churchill Hospital, Oxford, UK

14.6 Bronchoscopy

41.10 Respiratory infections

Alan R. Berry Consultant Surgeon, Northampton General Hospital, UK

26.5 Volvulus of the colon

33.3 Trauma to the pancreas

A. Bianchi Specialist Paediatric and Neonatal surgeon, St Mary's Hospital for Women & Children and Royal Manchester Children's Hospital, Manchester, UK

24.6 Short bowel syndrome

Kenneth I. Bickerstaff Consultant Surgeon, Princess Royal Hospital, Telford, Shropshire, UK

31.1 Benign diseases of the biliary tract

D. L. Bigam The Toronto Hospital, Ontario, Canada

16.8 Small bowel transplantation

Ronald Bleday Chief of Colorectal Surgery, Beth Israel-Deaconess Medical Center, Boston, Massachusetts, USA

24.1 Small bowel obstruction

Joseph G. Borer Children's Hospital, Boston, Massachusetts, USA

42.9 Surgery of the urinary tract in children

Cesario V. Borlongan Senior Staff Fellow, Intramural Research Program, National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland, USA

16.12 Reconstruction of the central nervous system by neural transplantation

William D. Boyd Senior Registrar, Obstetrics and Gynaecology, John Radcliffe Hospital, Oxford, UK

36.6 Uterine fibroids

36.9 Perineal abscesses

David C. Brewster Clinical Professor of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, USA

17.5.1 Aortoiliac occlusive disease

Julian Britton Consultant Surgeon, John Radcliffe Hospital, Oxford,UK

14.2 Biliary and pancreatic endoscopy

30.1 Surgical anatomy of the liver and biliary tree

31.1 Benign diseases of the biliary tract

34.1 The acute abdomen

M. S. Brose Fellow in Hematology and Oncology, Hospital of the University of Pennsylvania, Philadelphia, USA

48.1 Cancer genetics

John A. Buehler Oral and Maxillofacial Surgery, Harvard Vanguard Medical Associates, Brookline, Massachusetts, USA

45.10 Oral and maxillofacial surgery

Kevin G. Burnand Professor and Chairman of Surgery at Guy's, King's, and St Thomas', Guy's, King's, and St Thomas' Medical School of King's College, London, UK

19.0 Abnormalities of the lymphatic system

Martin J. Burton Consultant Otolaryngologist, The Radcliffe Infirmary, Oxford, UK

45.8 Surgery of the ear and temporal bone

45.14 Surgical technique of tracheostomy, minitracheostomy, and cricothyroidotomy

Thomas A. D. Cadoux-Hudson Consultant Neurosurgeon, The Radcliffe Infirmary, Oxford, UK

44.9 Intracranial aneurysms and arteriovenous malformations

Joseph Califano Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

45.4 Salivary glands

Richard P. Cambria Associate Professor of Surgery, Massachusetts General Hospital, Boston, USA

17.6.2 Femoral artery

John L. Cameron Chairman, Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

33.4 Pancreatic cancer

Bruce Campbell Consultant Vascular Surgeon, Royal Devon and Exeter Hospital, Exeter, UK

17.10.1 Arterial emboli in limbs

Emin A. Carapeti Resident Surgical Officer, St Mark's Hospital, Harrow, Middlesex, UK

28.3 Anal fissure

Gordon Carlson Consultant Surgeon and Senior Lecturer, Hope Hospital, Salford, Manchester, UK

24.3 Fistulas

Grant W. Carlson Chief, Department of Surgery, Emory University School of Medicine, Atlanta, Georgia,USA

38.5.2 Surgical management of malignant melanoma

James Carmichael Professor of Clinical Oncology, University of Nottingham, UK

48.5 Emergencies in cancer

A. J. Carr Consultant Orthopaedic Surgeon, Nuffield Orthopaedic Centre, Oxford, UK

46.2.2 The adult shoulder

Eric W. Carson Department of Orthopedics, LSUMC-Medical Center, New Orleans, Louisiana, USA

46.1 Sports medicine and athletic injuries

Philip Carson Division of Surgery, Royal Darwin Hospital, Casuarina, Australia

47.1.1 Abscess, cellulitis, and necrotizing bacterial infections

R. W. Chapman Consultant Gastroenterologist/Honorary Senior Lecturer, John Radcliffe Hospital, Oxford, UK

12.4 Hepatic problems

F. Mark Charnock Consultant Obstetrician and Gynaecologist, John Radcliffe Hospital, Oxford, UK

36.6 Uterine fibroids

36.9 Perineal abscesses

36.14 Rupture of the pregnant uterus

Clare Cheek Specialist Registrar in General and Colorectal Surgery, Birmingham Heartlands Hospital, UK

35.1 Inguinal and femoral hernias

Changyi Chen Assistant Professor and Co-Director of the Vascular Research Laboratory, Emory University School of Medicine, Atlanta, Georgia, USA

17.15 Neointimal hyperplasia, thrombophilia, and graft thrombosis

George W. Cherry Chairman, Wound Healing Programme, Oxford Wound Healing Institute, The Churchill Hospital, Oxford, UK

6 Wound healing

Nigel R. A. Clarke Specialist Registrar in Cardiology, John Radcliffe Hospital, Oxford, UK

11.2 Cardiovascular aspects

P. Jane Clarke Consultant Surgeon, Nuffield Department of Surgery, John Radcliffe Hospital, Oxford, UK

21.3 Abnormalities of the male breast

43.1 Surgery of the spleen

43.3 Surgery in patients with leukaemia

Michael A. Coady Department of Surgery, Yale University School of Medicine, New Haven, USA

40.4 Vascular and cardiac biological and non-biological prostheses

40.9.2 Surgery for acquired aortic and pulmonary valve disease

Carol A. Cobb John Radcliffe Hospital, Oxford, UK

12.4 Hepatic problems

Alfred M. Cohen Chief of the Colorectal Service, Memorial-Sloan Kettering Cancer Center, New York, USA

26.3 Colorectal tumours

Jack Collin Consultant Surgeon and Reader in Surgery, University of Oxford and John Radcliffe Hospital, Oxford, UK

17.6.1 Abdominal aorta

17.6.6 Visceral arteries

17.8.3 Sympathectomy in the upper limb

17.12 Infection and arterial surgery

Robert E. Condon Emeritus Professor and Chairman, The Medical College of Wisconsin, Milwaukee, USA

4.1 Surgical infections

Christopher P. Conlon Consultant in Infectious Diseases, John Radcliffe Hospital, Oxford, UK

17.12 Infection and arterial surgery

Timothy A. Cook Senior Registrar, St Mark's Hospital, Harrow, Middlesex, UK

24.2.2 Surgery for Crohn's disease of the small intestine

26.2.3 Surgical management of ulcerative colitis

26.2.4 Crohn's disease of the colon

26.6 Chronic constipation in adults

26.8 Prolapse of the rectum

Jack G. Copeland Marshall Foundation Artificial Heart Program, University of Arizona Health Sciences Center, Tucson, USA

40.3 Circulatory support for the failing heart

A. Benedict Cosimi Professor of Surgery, Harvard University Medical School; Chief Transplantation Unit, Massachusetts General Hospital, Boston, USA

16.6 Liver transplantation

D. Cranston Consultant Urological Surgeon, The Churchill Hospital and Honorary Senior Lecturer, University of Oxford, Oxford, UK

34.5 Retroperitoneal fibrosis

39.7 Benign prostate disease

39.9.1 Tumours of the kidney

39.9.3 Tumours of the testicle

39.9.4 Carcinoma of the penis

39.10 Torsion of the testicle

39.11 Vasectomy

Dennis C. Crawford Rhode Island Hospital, Providence, USA

46.3.4 Slipped femoral capital epiphysis

Thomas G. Cropley Associate Professor of Medicine, Division of Dermatology, University of Massachusetts Medical Center, Worcester, USA

38.1 Hamartomas, vascular malformations, and congenital defects

38.2 Benign tumours and cysts of the skin and its appendages

38.4 Premalignant and malignant tumours of the skin

38.7 Therapies used in dermatology

Susan N. Cullen Part-time Specialist Registrar, Department of Gastroenterology, John Radcliffe Hospital, Oxford, UK

12.4 Hepatic problems

A. S. Daar Professor of Surgery, College of Medicine, Sultan Qaboos University, Sultanate of Oman and Hunterian Professor, Royal College of Surgeons of England

47.1.2 Typhoid fever and other salmonella infections

47.2.2 Schistosomiasis (bilharziasis)

47.2.4 Hydatid disease

47.2.5 Ascariasis and other intestinal nematode infections

47.2.6 Biliary and intestinal trematodes

47.8 Surgery of advanced disease and late presentation

47.9 Surgery with limited resources

George W. Daneker, Jr.
Georgia Surgical Associates, Atlanta, USA

30.4 Cancers of the liver

M. J. Davies Professor of Cardiovascular Pathology, St George's Hospital Medical School, London, UK

40.9.1 Pathology of cardiac valves

Robert J. O. Davies Consultant Physician and Senior Lecturer, Osler Chest Unit, The Churchill Hospital, Oxford, UK

45.12 The surgical treatment of snoring and obstructive sleep apnoea

Peter A. Davis Senior Registrar, Hammersmith Hospital, London, UK

4.2 The surgery associated with HIV infection

Jean B. deKernion Chairman of Urology, UCLA School of Medicine, Los Angeles, California, USA

39.8 Cancer of the prostate

Francis L. Delmonico Associate Professor of Surgery, Massachusetts General Hospital, Boston, USA

16.2 Organ procurement

Mahlon R. Delong Chairman of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA

44.5 Surgery for Parkinson's disease and other movement disorders

Tom R. DeMeester Professor and Chairman, USC Healthcare Consultation Center, Los Angeles, California, USA

22.2.1 Reflux disease and hiatus hernia

22.2.2 Benign oesophageal strictures

Ronald W. Deskin Associate Professor of Otolaryngology and Pediatrics, University of Texas Medical Branch, Galveston, USA

42.12 Paediatric otolaryngology

Jules L. Dienstag Associate Professor of Medicine, Harvard Medical School and Physician, Massachusetts General Hospital, Boston, USA

32.5 Ascites and its complications

Michael Dobson Consultant Anaesthetist and Senior Clinical Lecturer, John Radcliffe Hospital, Oxford, UK

10.2 Approaches to local and regional anaesthesia

Thomas F. Dodson Associate Professor, Department of Vascular Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

17.5.3 Mesenteric arteries

51 How should surgical trainees be selected?

Richard Doll Emeritus Professor of Medicine, University of Oxford, UK

48.2 Avoidable cancer

Daniel P. Doody Associate Visiting Surgeon, Department of Pediatric Surgery, Massachusetts General Hospital, Boston, USA

41.6 The diaphragm

Marcus Drake Clinical Lecturer in Urology, University of Newcastle, Newcastle upon Tyne, UK

39.4 Urinary complications of spinal abnormalities and injuries

Stephen P. Dretler Director, Kidney Stone Center, Massachusetts General Hospital, Boston, USA

39.5 Urinary stone disease

Nicholas Dudley Honorary Consultant General Surgeon, John Radcliffe Hospital, Oxford, UK

20.2.2 The thyroid gland

20.3.2 Surgery of the parathyroid glands

L. T. Dunn Senior Lecturer and Honorary Consultant, Department of Neurosurgery, University of Glasgow, UK

44.2 Head injury

Bruce C. Dunphy Director and Chief, Division of Gynaecology, Foothills Hospital, Calgary, Alberta, Canada

36.2 The management of acute pelvic pain

Oliver J. Dyar Consultant Anaesthetist, John Radcliffe Hospital, Oxford, UK

10.1 Anaesthesia for surgeons

Craig P. Eberson Rhode Island Hospital, Providence, USA

46.3.3 Legg-Calve-Perthes' disease

James S. Economou Division of Surgical Oncology, UCLA School of Medicine, Los Angeles, California, USA

48.4 Cancer gene therapy

Michael G. Ehrlich Surgeon-in-Chief, Department of Orthopedics, Rhode Island Hospital,Providence, USA

46.3.2 Developmental dysplasia of the hip

46.3.3 Legg-Calve-Perthes' disease

46.3.4 Slipped femoral capital epiphysis

46.3.5 Congenital foot disorders

David W. Eisele Director, Division of Head and Neck Surgery, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

45.4 Salivary glands

Margarett Ellison Third Year Fellow in Gynecologic Oncology, Women & Infants Hospital, Profidence, Rhode Island, USA

36.8.2 Cervical malignancy

John B. Emans Department of Orthopedic Surgery, Children's Hospital, Boston, Massachusetts, USA

46.3.6 Congenital spinal deformity

Stephen S. Falkenberry Gynecologic Oncologist/Breast Surgeon, Women & Infants Hospital, Providence, Rhode Island, USA

36.8.2 Cervical malignancy

S. T. Fan Department of Surgery, University of Hong Kong

47.6 Intrahepatic stones

Timothy M. Farrell Clinical Fellow in Laparoscopic Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

15 Endoscopic surgery

David V. Feliciano Chief of Surgery, Grady Memorial Hospital, Atlanta, Georgia, USA

2 Trauma and shock

17.11 Arterial and venous injuries

30.2 Hepatic trauma

Sandy Feng Assistant in Surgery, Massachusetts General Hospital, Boston, USA

16.6 Liver transplantation

Charles M. Ferguson Wang Ambulatory Center, Massachusetts General Hospital, Boston, USA

27.1 Acute appendicitis

Mark W. J. Ferguson Professor, School of Biological Sciences, University of Manchester, UK

6 Wound healing

T. Bruce Ferguson, Jr.
Associate Professor of Surgery, Department of Cardiothoracic Surgery, LSU School of Medicine, Marrero, Louisiana, USA

40.11 Surgery for cardiac rhythm disturbances

Jeff Fidler Department of Radiology, University of Nebraska Medical Center, Omaha, USA

13.3 Magnetic resonance imaging

Aaron S. Fink Chief, Surgical Service, Department of Surgery, Atlanta Veterans Affairs Medical Center, Decatur, Georgia, USA

14.1 Upper gastrointestinal endoscopy

14.5 Percutaneous endoscopic gastrostomy and jejunostomy

Josef E. Fischer Chairman, Department of Surgery, University of Cincinnati Medical Center, Ohio, USA

3 Surgical nutrition

Pierre Foex Nuffield Professor of Anaesthetics, University of Oxford, UK

12.2 Cardiological problems

Kenneth A. Forde Department of Surgery, College of Physicians and Surgeons, Columbia University, New York, USA

14.3 Colonoscopy

J. C. Forfar Consultant Physician, John Radcliffe Hospital, Oxford, UK

40.8.1 Coronary artery disease: risk factors, diagnosis, and medical treatment

Robert G. Forman Director, Centre for Reproductive Medicine, London, UK

36.3 Acute pelvic sepsis and tubo-ovarian abscess

Allen Foster Senior Lecturer in International Eye Health, London School of Hygiene and Tropical Medicine, UK

47.5 Common surgical eye conditions in the developing world

H. Foster Director, Special Surgical Unit, John Hunter Hospital, Newcastle, Australia

47.1.4 Tropical ulcer

47.3 Splenic surgery in the tropics

Antony Fox Consultant Vascular Surgeon, Royal Shrewsbury Hospitals NHS Trust, UK

17.6.4 Carotid artery

17.6.5 Subclavian artery

Kenneth L. Franco Associate Professor, Yale University School of Medicine, New Haven, Connecticut, USA

40.9.3 Diagnosis and surgical management for tricuspid and mitral valve disease

Patricia M. Franklin Clinical Nurse Specialist and Psychologist in Transplantation, The Churchill Hospital, Oxford, UK

11.8 Grief: breaking bad news: offering organ donation

J. A. Franklyn Professor of Medicine, University of Birmingham, UK

20.2.1 Physiology of the thyroid gland

Andrew P. Freeland Consultant Otolaryngologist, The Radcliffe Infirmary, Oxford, UK

45.8 Surgery of the ear and temporal bone

Thomas B. Freeman Professor, Department of Neurosurgery, Harbourside Medical Tower, Tampa, Florida, USA

16.12 Reconstruction of the central nervous system by neural transplantation

Marvin P. Fried Otolaryngologist-In-Chief, The Joint Center for Otolaryngology, Boston, Massachusetts, USA

45.3 The evaluation of a neck mass in the adult patient

Ellen M. Friedman Chief of Service, Department of Pediatric Otolaryngology, Baylor College of Medicine, Houston, Texas, USA

42.12 Paediatric otolaryngology

Robert Fry Divrector, Division of Colon and Rectal Surgery, Jefferson Faculty Foundation, Philadelphia, Pennsylvania, USA

29 Sacrococcygeal pilonidal sinus

Henning A. Gaissert University Surgical Associates, Brown University School of Medicine,Providence, Rhode Island, USA

22.5 Benign and malignant tumours

41.5 Mediastinal tumours

Walter H. Gajewski Assistant Professor, Brown University School of Medicine, Providence, Rhode Island, USA

36.8.1 Uterine malignancy

36.8.2 Cervical malignancy

G. Gregory Gallico III
Associate Professor of Surgery, Massachusetts General Hospital, Boston, USA

50.1 Principles and practice of plastic surgery

J. R. Galloway Associate Professor of Surgery, Department of Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

32.2 Emergency treatment of bleeding oesophageal varices

32.3 Surgical shunt procedures

Ernest V. Garcia Director, Emory Center for PET, Atlanta, Georgia, USA

13.9 Positron emission tomography

Julio Garcia-Aquilar Colon and Rectal Surgery Associates, St Paul, Minnesota, USA

28.4 Anorectal abscesses and fistula-in-ano

Christopher S. Garrard Director of Intensive Care, John Racliffe Hospital, Oxford, UK

11.1 Guidelines and scoring systems

11.2 Cardiovascular aspects

11.3 Renal aspects

11.4 Respiratory aspects

11.5 Infection in the intensive care unit

11.6 Central nervous system aspects

11.7 Gastrointestinal aspects

R. J. Gates Overseas Fellow, Department of Plastic Surgery, The Radcliffe Infirmary, Oxford, UK

45.11 Facial injuries

Bruce D. George Consultant Colorectal Surgeon, John Radcliffe Hospital, Oxford, UK

26.1 Diverticular disease: diverticulitis, bleeding, and fistula

26.7 Angiodysplasia

Paul L. F. Giangrande Consultant Haematologist, Oxford Haemophilia Centre, The Churchill Hospital, Oxford, UK

5.1 Haematological problems

Joseph S. Giglia University of Cincinnati School of Medicine, Ohio, USA

17.5.1 Aortoiliac occlusive disease

Mark R. Gilbert The Emory Clinic, Inc., Atlanta, Georgia, USA

44.3 Intracranial tumours

Michael D. G. Gillmer Consultant Obstetrician and Gynaecologist, John Radcliffe Hospital, Oxford, UK

36.2 The management of acute pelvic pain

36.4 Ovarian accidents

36.12 Third-stage complications

Armando E. Giuliano Chief of Surgical Oncology, John Wayne Cancer Institute at the Saint John's Health Center, Santa Monica, California, USA

43.2 Lymph node diagnosis

Stephen Golding Lecturer in Radiology, University of Oxford and Honorary Consultant Radiologist, John Radcliffe Hospital, Oxford, UK

13.2 Computed tomography

T. E. E. Goodacre Consultant Plastic and Reconstructive Surgeon and Honorary Senior Lecturer, The Radcliffe Infirmary, Oxford, UK

47.7 Some special aspects of plastic surgery in developing countries

Angelo E. Gousse Attending Surgeon, San Francisco General Hospital, California, USA

39.2 Voiding function and dysfunction

S. J. Govil Consultant Surgeon, Christian Medical College for Lower India, Vellore, India

33.2 Chronic pancreatitis

D. Michael Grace Department of General Surgery, London Health Sciences Center, London, Ontario, Canada

25 Surgery for obesity

Sam D. Graham, Jr.
The Virginia Urology Center, Richmond, USA

39.9.3 Tumours of the testicle

C. O. Granai Director of Gynecologic Oncology, Brown University/Women and Infants Hospital, Providence, Rhode Island, USA

36.8.1 Uterine malignancy

36.8.2 Cervical malignancy

36.8.3 Radical hysterectomy

Christopher S. Grant Head of Department of Surgery, College of Medicine, Sultan Qaboos University, Sultanate of Oman

47.1.2 Typhoid fever and other salmonella infections

47.2.4 Hydatid disease

47.2.6 Biliary and intestinal trematodes

D. R. Grant The Toronto Hospital, Ontario, Canada

16.8 Small bowel transplantation

Hugh Grant Consultant Paediatric Surgeon, John Radcliffe Hospital, Oxford, UK

47.10 Special paediatric problems in developing countries

Derek W. R. Gray Reader in Surgery and Consultant Surgeon, John Radcliffe Hospital, Oxford, UK

16.10 Pancreatic islet and fetal pancreas transplantation

17.16 Vascular access and other techniques for dialysis and chemotherapy

17.17 Limb amputation

Elizabeth Gray Department of Psychiatry, Massachusetts General Hospital, Boston, USA

12.5 Psychological care

Catherine R. Grebenik Consultant Anaesthetist, Oxford heart Centre, John Radcliffe Hospital, Oxford, UK

40.5 Principles of blood conservation, cardiac anaesthesia, and cerebral protection

40.6 Cardiovascular monitoring and postoperative care

Fiona R. Green Nuffield Department of Surgery, John Radcliffe Hospital, Oxford, UK

17.1 Pathobiology of atherosclerosis

Michael J. Greenall Consultant Surgeon, John Radcliffe Hospital, Oxford, UK

21.1 Benign conditions of the breast

21.2 Cancer of the breast

23.2 Carcinoma of the stomach

28.6 Cancer of the anus

Lazar J. Greenfield Department of Surgery, University of Michigan, Ann Arbor, USA

18.2 Deep vein thrombosis and pulmonary embolism

Justin K. Greisberg Rhode Island Hospital, Providence, USA

46.3.5 Congenital foot disorders

B. Gribbin Honorary Consultant, Department of Cardiology, John Radcliffe Hospital, Oxford, UK

40.8.1 Coronary artery disease: risk factors, diagnosis, and medical treatment

D. Mervyn Griffiths Consultant Paediatric Surgeon and Senior Lecturer in Paediatric Surgery, Wessex Regional Centre for Paediatric Surgery, Southampton General Hospital, UK

42.4 Disorders of the gastrointestinal tract in children

Hermes C. Grillo Chief of Thoracic Surgery, Massachusetts General Hospital, Boston, USA

41.2 The trachea

Robert A. Guyton Chief of Cardiothoracic Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

40.8.3 Coronary artery bypass

Raghuveer K. Halkar Division of Nuclear Medicine, Emory University Hospital, Atlanta, Georgia, USA

13.9 Positron emission tomography

David Hamilton Professor of Cardiac Surgery, University of Edinburgh, UK

40.7.4 Malalignment anomalies of the interventricular septum

George Hamilton Consultant and Honorary Senior Lecturer in Vascular Surgery, Royal Free Hospital School of Medicine, London, UK

32.1 Aetiology, presentation, and investigation

32.4 Non-shunt procedures in management of variceal bleeding

Graeme L. Hammond Department of Surgery, Yale University School of Medicine, New Haven, Connecticut, USA

40.4 Vascular and cardiac biological and non-biological prostheses

40.9.2 Surgery for acquired aortic and pulmonary valve disease

40.9.3 Diagnosis and surgical management for tricuspid and mitral valve disease

Damian Hanbury Consultant Urological Surgeon, The Lister Hospital, Stevenage, Hertfordshire, UK

39.7 Benign prostate disease

Linda Hands Clinical Reader and Consultant Surgeon, John Radcliffe Hospital, Oxford, UK

17.6.3 The popliteal artery

17.7.3 Carotid body tumours

17.9 Renovascular disease

21.3 Abnormalities of the male breast

John Hargrave Formerly of the Northern Territory Medical Service, Australia

47.1.6 Leprosy

Michael R. Harrison Director, UCSF Fetal Treatment Center, San Francisco, California, USA

42.11 Fetal surgery

John G. Heller Director of Education, Emory Spine Center, Decatur, Georgia, USA

46.2.9 Spinal infections

I. Craig Henderson Alza Corporation, Palo Alto, California, USA

48.3 Chemotherapy of cancer

J. Michael Henderson Chairman, Department of General Surgery, Cleveland Clinic Foundation, Ohio, USA

30.5 The Budd-Chiari syndrome

W. Hardy Hendren Chief of Surgery Emeritus, Children's Hospital, Boston, Massachusetts, USA

42.2 Thoracic surgical problems

42.9 Surgery of the urinary tract in children

Oscar Joe Hines Assistant Professor, Gastrointestinal Surgery, UCLA School of Medicine, Los Angeles, California, USA

33.1 Acute pancreatitis

John M. Hoffman Department of Neurology, Eory University School of Medicine, Atlanta, Georgia, USA

13.9 Positron emission tomography

Herbert C. Hoover, Jr.
Chairman, Department of Surgery, Lehigh Valley Hospital, Allentown, Pennsylvania, USA

34.6 Retroperitoneal neoplasms

Ira R. Horowitz Director, Division of Gynecologic Oncology, Emory University School of Medicine, Atlanta, Georgia, USA

36.5 Ovarian tumours

Brian Howden Melbourne, Australia

16.3 Organ and tissue preservation for transplantation

Margaret A. Hughes Clinical Scientist, Oxford Wound Healing Institute, The Churchill Hospital, Oxford, UK

6 Wound healing

William J. K. Huizinga Associate Professor in Surgery, Sultan Qaboos University, Sultanate of Oman

47.1.2 Typhoid fever and other salmonella infections

47.2.4 Hydatid disease

John G. Hunter Director, Emory Endosurgical Center, Emory University School of Medicine, Atlanta, Georgia, USA

15 Endoscopic surgery

Imtiaz Husain Bedford Hospital, UK (formerly Associate Professor of Urology, King Saud University, Riyadh, Saudi Arabia)

47.2.3 Schistosomiasis (bilharziasis) of the genitourinary tract

C. W. Imrie Professor, Department of Surgery, Royal Infirmary, Glasgow, UK

33.2 Chronic pancreatitis

Miles Irving Professor, Department of Surgery, University of Manchester Hope Hospital, Salfored, Manchester, UK

9 Evidence-based approach to surgical decisions

24.3 Fistulas

Mohammad Bashar Department of Surgery, Prince of Wales Hospital, The Chinese University of Hong Kong

14.7 Thoracoscopy

Paula Jablonski Melbourne, Australia

16.3 Organ and tissue preservation for transplantation

Marshall L. Jacobs Section of Cardiothoracic Surgery, St Christopher's Hospital for Children, Philadelphia, Pennsylvania, USA

40.7.6 Stenosis and atresia of the arterial valves

D. P. Jewell Professor of Gastroenterology, John Radcliffe Hospital, Oxford, UK

24.2.1 Crohn's disease of the small intestine

24.5 Malabsorption syndromes

26.2.2 Colonic inflammation-diagnosis and management

Alan G. Johnson Professor of Surgery, University of Sheffield, Royal Hallamshire Hospital, Sheffield, UK

23.1 Peptic ulcer - stomach and duodenum

23.3 ‘Gastritis’

Gordon J. Johnson Rothes Professor of Preventive Ophthalmology, Institute of Ophthalmology, University College London, UK

47.5 Common surgical eye conditions in the developing world

Richard P. Juniper Director, Postgraduate Dental Education, Oxford, UK

45.9 Temporomandibular joint dysfunction

David A. Keith Department of Oral and Maxillofacial Surgery, Massachesetts General Hospital, Boston, USA

45.10 Oral and maxillofacial surgery

Pond R. Kelemen John Wayne Cancer Institute at the Saint John’s Health Center, Santa Monica, California, USA

43.2 Lymph node diagnosis

David W. Kennedy Department of Otorhinolaryngology, University of Pennsylvania School of Medicine, Philadelphia, USA

45.2 Nose and sinuses

Walter W. K. King Chief and Consultant in Head and Neck/Plastic and Reconstructive Surgery, Prince of Wales Hospital, Hong Kong

45.6 Nasopharyngeal cancer

Andrew Kingsnorth Professor of Surgery, University of Plymouth, Derriford Hospital, Plymouth, UK

35.1 Inguinal and femoral hernias

Dicken S. C. Ko Assistant in Surgery and Urology, Massachusetts General Hospital, Boston, USA

16.6 Liver transplantation

Lewis Kriteman Chief Resident, Department of Urology, Emory University School of Medicine, Atlanta, Georgia, USA

39.6 Male sexual dysfunction and infertility

Terry C. Lairmore Department of Surgery, Washington University School of Medicine, St Louis, Missouri, USA

20.1 Genetic aspects of endocrine disease

Hillel Laks Chief, Cardiothoracic Surgery, UCLA School of Medicine, Los Angeles, California, USA

40.7.7 Transposition of the great arteries

Peter M. Lamont Consultant Vascular Surgeon and Honorary Senior Clinical Lecturer in Surgery, Bristol Royal Infirmary, UK

17.8.2 Raynaud’s syndrome

35.3 Wound dehiscence, incisional hernia, and parastomal hernia

Glenn M. LaMuraglia Division of Vascular Surgery, Massachusetts General Hospital, Boston, USA

17.14 Thrombolytic therapy

Richard G. B. Langley Instructor in Dermatology, Massachusetts General Hospital, Boston, USA

38.3 Benign acquired pigmented lesions of the skin

38.5.1 Recognition, staging, and prognosis of cutaneous melanoma

38.6 Cutaneous problems occurring in the surgical patient

C. Larsen Transplant Immunology, Emory University School of Medicine, Atlanta, Georgia, USA

16.9 Vascularized pancreatic transplantation

Eric L. Lazar Senior Clinical Fellow, Division of Pediatric Surgery, Babies and Children’s Hospital of New York, USA

42.6 Paediatric hepatobiliary disorders - surgical aspects

Hop Le San Francisco VA Medical Center, California, USA

33.5 Insulinomas and other tumours

David J. Leaper Professor, Professorial Unit of Surgery, North Tees Hospital, Stockton on Tees, Tyneside, UK

6 Wound healing

King-Teh Lee Associate Professor of Surgery, Kaohsiung Medical College Hospital, Taiwan, Republic of China

47.1.7 Pyogenic liver abscess

George V. Letsou Associate Professor, Department of Surgery, Texas Medical Center, Baylor College of Medicine, Houston, USA

16.7 Heart and heart-lung transplantation

Arthur K. C. Li Professor and Vice Chancellor, Chinese University of Hong Kong

45.6 Nasopharyngeal cancer

Richard S. Limbird University of Orthopedics, Providence, Rhode Island, USA

46.2.5 The knee

Peter H. Lin Vascular Surgery Division, Emory Clinic, Atlanta, Georgia, USA

17.5.3 Mesenteric arteries

David R. M. Lindsell Consultant Radiologist, John Radcliffe Hospital, Oxford, UK

13.5 Ultrasound imaging

13.6 Imaging in children

David R. M. Lindsell Consultant Radiologist, John Radcliffe Hospital, Oxford, UK

13.10 Imaging guidelines

T. J. Littlewood Department of Haematology, John Radcliffe Hospital, Oxford, UK

5.1 Haematological problems

C. M. Lo Department of Surgery, University of Hong Kong, Queen Mary Hospital, Hong Kong

47.6 Intrahepatic stones

Reginald V. N. Lord USC Healthcare Consultation Center, Los Angeles, California, USA

22.2.1 Reflux disease and hiatus hernia

22.2.2 Benign oesophageal strictures

Carol A. Lorente Oral and Maxillofacial Surgery, Harvard Vanguard Medical Associates, Brookline, Massachusetts, USA

45.10 Oral and maxillofacial surgery

James W. Lucarini Ear, Nose, and Throat Department, Kaiser Permanente, Atlanta, Georgia, USA

45.5 Oral and oropharyngeal cancer

Philip R. Lucas University of Orthopedics, Providence, Rhode Island, USA

46.2.6 Degenerative disorders of the spine

46.2.7 Spinal trauma

Gary Lum Division of Surgery, Royal Darwin Hospital,Casuarina, Australia

47.1.1 Abscess, cellulitis, and necrotizing bacterial infections

Alan B. Lumsden Chief of Vascular Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

17.13 Minimally invasive vascular surgery

17.15 Neointimal hyperplasia, thrombophilia, and graft thrombosis

H. T. Lynch Professor and Chairman, Department of Preventive Medicine and Public Health; Professor of Medicine, Director of Creighton Cancer Center, and Director of Creighton Hereditary Cancer Institute, Creighton University School of Medicine, Omaha, Nebraska, USA

48.1 Cancer genetics

Oluwatope A. Mabogunje

47.2.5 Ascariasis and other intestinal nematode infections

47.2.6 Biliary and intestinal trematodes

I. Z. Mackenzie Reader in Obstetrics and Gynaecology, John Radcliffe Hospital, Oxford, UK

36.1 Acute vaginal bleeding

36.10 Late spontaneous abortion

B. James Mander Specialist Registrar, South-West Thames Region, London, UK

28.5 Faecal incontinence

Pankaj Mankad Royal Hospital for Sick Children and Royal Infirmary, Edinburgh, UK

40.7.4 Malalignment anomalies of the interventricular septum

Henry J. Mankin Orthopedic Service, Massachusetts General Hospital, Boston, USA

46.4 Management of bone tumours

Michael N. Margolies Professor of Surgery, Massachusetts General Hospital East, Charlestown, USA

23.5 Foreign bodies and bezoars

24.4 Diverticular disease of the small bowel

24.9 Foreign bodies

C. G. Marks Professor of Surgical Oncology, Royal Surrey County Hospital, Guildford, UK

28.2 Pruritus ani

Vernon Marshall Professor, Department of Examinations and Training, Royal Australasian College of Surgeons, Melbourne, Australia

16.3 Organ and tissue preservation for transplantation

Douglas Mathisen Chief, General Thoracic Surgery Unit, Massachusetts General Hospital, Boston, USA

22.5 Benign and malignant tumours

41.2 The trachea

41.4 Primary chest wall neoplasms

41.5 Mediastinal tumours

Charles J. McCabe Associate Chief of Emergency Services, Massachusetts General Hospital, Boston, USA

34.3 Trauma: injuries to abdominal wall and mesentery

Calvin O. McCall Department of Dermatology, Emory University School of Medicine, Atlanta, Georgia, USA

38.1 Hamartomas, vascular malformations, and congenital defects

38.2 Benign tumours and cysts of the skin and its appendages

38.4 Premalignant and malignant tumours of the skin

38.7 Therapies used in dermatology

Edwin C. McGee, Jr.
Massachusetts General Hospital, Boston, USA

41.4 Primary chest wall neoplasms

Francis J. McGovern Clinical Instructor in Surgery, Massachusetts General Hospital, Boston, USA

39.1 Genitourinary anatomy

Catherine L. McGuinness Senior Lecturer/Consultant Surgeon, St Thomas' Hospital, London, UK

19 Abnormalities of the lymphatic system

Henry McQuay Professor of Pain Relief, Pain Relief Unit, The Churchill Hospital, Oxford, UK

49.3 Relief of chronic non-malignant pain

D. McWhinnie Consultant Surgeon, Milton Keynes General Hospital, uckinghamshire, UK

34.2 The peritoneum and peritonitis

34.4 The omentum

Mariana D. Mead Ophthalmic Consultants of Boston, Massachusetts, USA

45.1 Eye emergencies

Marcelo Mester Gastroenterology and Oncologic Surgery, Sao Paulo, Brazil

24.7 Radiation enteritis

Joseph I. Miller, Jr.
Chief of General Thoracic Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

22.4 Oesophageal perforation, Boerhaave's syndrome, and Mallory-Weiss syndrome

41.7 Lung volume reduction surgery and pulmonary transplantation

Andrew Mitchell Consultant Surgeon, Milton Keynes General Hospital, Buckinghamshire, UK

24.10 Pneumatosis cystoides intestinalis

Andrew Molyneux Consultant Neuroradiologist, The Radcliffe Infirmary, Oxford, UK

44.9 Intracranial aneurysms and arteriovenous malformations

Ashley C. Moncure Visiting Surgeon, Massachusetts General Hospital, Boston, USA

41.6 The diaphragm

41.8 Pneumothorax

41.9 Chylothorax

Arlene J. Morales Crawford Long Hospital, Atlanta, Georgia, USA

36.13 Surgery for infertility

David Morris University of North Carolina School of Medicine, Chapel Hill, USA

48.7.1 Principles of radiation oncology

Douglas C. Morris Director of the Emory Heart Center, Emory University School of Medicine, Atlanta, Georgia, USA

40.8.2 Interventional cardiology

M. Jocelyn Morris Osler Chest Unit, The Churchill Hospital, Oxford, UK

10.5 Interpretation of lung function tests in the surgical patient

Peter J. Morris Nuffield Professor of Surgery, John Radcliffe Hospital, Oxford, UK

16.4 Immunosuppression

16.5 Kidney transplantation

17.7.1 Carotid artery

17.7.2 Vertebrobasilar, subclavian, and innominate arteries

17.8.1 Thoracic outlet obstruction

17.9 Renovascular disease

43.1 Surgery of the spleen

Peter S. Morris Menzies School of Health, Casuarina, Northern Territory, Australia

47.1.5 Otitis media in high-risk populations

Neil Mortensen Consultant Colorectal Surgeon and Clinical Reader in Colorectal Surgery, John Radcliffe Hospital, Oxford, UK

24.2.2 Surgery for Crohn's disease of the small intestine

26.2.3 Surgical management of ulcerative colitis

26.2.4 Crohn's disease of the colon

26.6 Chronic constipation in adults

26.8 Prolapse of the rectum

John T. B. Moyle Consultant Anaesthetist and Chartered Engineer, Milton Keynes General Hospital, Buckinghamshire, UK

10.3 Surgical diathermy

10.4 Electrical safety in anaesthesia and surgery

Richard J. Mullins Professor, Oregon Health Sciences University, Portland, USA

2 Trauma and shock

John A. Murie Consultant Vascular Surgeon, The Royal Infirmary of Edinburgh, UK

17.5.2 Femoral and distal arteries

M. F. Murphy Consultant Haematologist and Senior Clinical Lecturer in Blood Transfusion, National Blood Service, Oxford Radcliffe Hospitals, UK

5.2 Blood transfusion in surgical practice

John Naitoh Assistant Clinical Professor of Surgery and Urology, UCSD School of Medicine, La Jolla, California, USA

39.8 Cancer of the prostate

S. Anwar Naqvi

47.9 Surgery with limited resources

K. L. Nathanson Division of Medical Genetics, University of Pennsylvania Health System, Philadelphia, USA

48.1 Cancer genetics

Jeremy Noble Department of Urology, The Churchill Hospital, Oxford, UK

39.4 Urinary complications of spinal abnormalities and injuries

Daniel J. Nolan Consultant Radiologist, John Radcliffe Hospital, Oxford, UK

13.4 Gastrointestinal radiology

C. M. Norris, Jr.
Assistant Professor of Otolaryngology, Harvard Medical School, Boston, Massachusetts, USA

45.3 The evaluation of a neck mass in the adult patient

45.5 Oral and oropharyngeal cancer

Jeffrey A. Norton Chief, Surgical Service, San Francisco VA Medical Center, California, USA

33.5 Insulinomas and other tumours

T. S. O'Brien Consultant Urological Surgeon, Guy's and St Thomas' Hospitals, Guy's Hospital Campus, London, UK

39.9.2 Urothelial tumours

Kieran J. O'Flynn Consultant Urological Surgeon, Hope Hospital, Salford, Manchester, UK

9 Evidence-based approach to surgical decisions

Ian O'Rourke Professor, Division of Surgery, Royal Darwin Hospital, Casuarina, Australia

47.1.1 Abscess, cellulitis, and necrotizing bacterial infections

Jeffrey J. Olson Section of Neurosurgery, The Emory Clinic, Inc., Atlanta, Georgia, USA

44.3 Intracranial tumours

R. OmPrakash Western General Hospital, Edinburgh, UK

47.1.3 Abdominal tuberculosis

Richard R. Orlandi Assistant Professor, Department of Otolaryngology, University of Michigan School of Medicine, Ann Arbor, USA

45.2 Nose and sinuses

Ingegard Östman-Smith Consultant Paediatric Cardiologist, John Radcliffe Hospital, Oxford, UK

40.7.1 Management of congenital heart disease

Leslie W. Ottinger Formerly Visiting Surgeon, Massachusetts General Hospital, Boston, USA

17.10.2 Arterial emboli: mesenteric arteries

Nelson Oyesiku Department of Neurosurgery,The Emory Clinic, Inc., Atlanta, Georgia, USA

44.3 Intracranial tumours

A. P. Pandey Professor of Urology, Christian Medical College Hospital, Vellore, India

47.2.7 Chyluria

Andrew Parry Consultant Paediatric Cardiac Surgeon, Bristol Royal Hospital for Sick Children, UK

40.7.4 Malalignment anomalies of the interventricular septum

40.12 Miscellaneous cardiac diseases

Arun Pausawasdi Professor of Surgery, Faculty of Medicine, Siriraj Hospital, Bangkok, Thailand

47.2.1 Amoebiasis

T. Pearson Transplant Immunology, Emory University School of Medicine, Atlanta, Georgia, USA

16.9 Vascularized pancreatic transplantation

Lester C. Permut Assistant Professor of Surgery and Pediatrics, New York Medical College, Briarcliff Manor, USA

40.7.7 Transposition of the great arteries

R. K. S. Phillips Consultant Colorectal Surgeon, St Mark's Hospital, Harrow, Middlesex, UK

28.3 Anal fissure

Ravi Pillai Consultant Cardiothoracic Surgeon and Honorary Senior Lecturer, University of Oxford, John Radcliffe Hospital, Oxford, UK

40.7.3 Atrial, ventricular, and atrioventricular septal defects

40.7.5 Abnormalities of the atrioventricular valves

Henry A. Pitt Department of Surgery, Medical College of Wisconsin, Milwaukee, USA

31.2 Malignant diseases of the biliary tract

Michael D. Poole Professor, The St George Hospital, Kogarah, Australia

50.2 Craniofacial reconstruction for congenital, traumatic, and neoplastic conditions

Marshall Posner Dana Farber Cancer Institute, Boston, Massachusetts, USA

45.5 Oral and oropharyngeal cancer

Timothy A. Pritts Department of Surgery, University of Cincinnati Medical Center, Ohio, USA

3 Surgical nutrition

Mary C. Proctor Senior Research Associate, Department of Surgery, University of Michigan, Ann Arbor, USA

18.2 Deep vein thrombosis and pulmonary embolism

Bheeshma Rajagopalan Reader in Medicine, MRC MRS Unit, John Radcliffe Hospital, Oxford, UK

13.8 Nuclear magnetic resonance spectroscopy

Jan Rakinic Program Director, Colon and Rectal Surgery, Department of Surgery, Jefferson Faculty Foundation, Philadelphia, Pennsylvania, USA

29 Sacrococcygeal pilonidal sinus

N. Rangabashyam Madras, India

47.1.3 Abdominal tuberculosis

Peter J. Ratcliffe Professor, Renal Unit, The Churchill Hospital, Oxford, UK

12.3 Renal problems

Shlomo Raz Department of Surgery/Urology, UCLA School of Medicine, Los Angeles, California, USA

39.2 Voiding function and dysfunction

Howard A. Reber Chief, Gastrointestinal Unit, UCLA School of Medicine, Los Angeles, California, USA

33.1 Acute pancreatitis

Mitchell F. Reiter Assistant Professor, Department of Orthopedic Surgery, University of Medicine and Dentistry of New Jersey, Newark, USA

46.2.9 Spinal infections

Richard R. Ricketts Chief, Pediatric Surgery, Emory Children's Center, Decatur, Georgia, USA

42.3 Evaluation of acute abdominal pain in children

42.7.1 Hirschsprung's disease

42.7.2 Other anorectal problems

S. Adib Rizvi

47.9 Surgery with limited resources

Justin A. Roake University Department of Surgery, Christchurch School of Medicine, New Zealand

20.4.1 The adrenal gland

James R. Roberson Emory Clinic Sports Center, Decatur, Georgia, USA

46.2.4 The arthritic hip

John A. Rock Chairman of Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, Georgia, USA

36.13 Surgery for infertility

Stephanie Rohovsky Chief Resident in Surgery, Beth Israel-Deaconess Medical Center, Boston, Massachusetts, USA

24.1 Small bowel obstruction

Jerrold Rosenberg Clinical Assistant Professor, Brown University Medical School, Department of Orthopedics and Rehabilitation, Rhode Island Hospital, Providence, USA

46.2.10 Spinal cord monitoring

David A. Rothenberger Colon and Rectal Surgery Associates, St Paul, Minnesota, USA

28.4 Anorectal abscesses and fistula-in-ano

Grace S. Rozycki Director, Trauma/Surgical Critical Care, General Surgery, Grady Memorial Hospital, Atlanta, Georgia, USA

2 Trauma and shock

Robert H. Rubin Chief of Infectious Disease for Transplantation, Massachusetts General Hospital, Boston, USA

30.3 Abscesses - pyogenic and amoebic

David M. Rudnick Fellow in Endourology, University of California at San Francisco, USA

39.3 Trauma to the genitourinary tract

39.5 Urinary stone disease

David Sackett Director, Centre for Evidence-based Medicine, Oxford, UK

9 Evidence-based approach to surgical decisions

Gregory P. Sadler Consultant Endocrine Surgeon, John Radcliffe Hospital, Oxford, UK

20.2.2 The thyroid gland

20.3.2 Surgery of the parathyroid glands

Michael H. Safir Attending Surgeon, San Francisco General Hospital, California, USA

39.2 Voiding function and dysfunction

Barry A. Salky Clinical Professor of Surgery, Division of Laparoscopic Surgery, Mount Sinai Hospital, New York, USA

14.4 Diagnostic laparoscopy

Paul R. Sanberg Professor, Department of Neurosurgery, University of South Florida College of Medicine, Tampa, USA

16.12 Reconstruction of the central nervous system by neural transplantation

J. Satsangi Professor of Gastroenterology, University of Edinburgh, UK

24.2.1 Crohn’s disease of the small intestine

Alan Saunder Melbourne, Australia

16.3 Organ and tissue preservation for transplantation

Adrian Savage Consultant Surgeon, Russells Hall Hospital, Dudley, West Midlands, UK

30.1 Surgical anatomy of the liver and biliary tree

31.1 Benign diseases of the biliary tract

35.3 Wound dehiscence, incisional hernia, and parastomal hernia

Andrew H. Schmidt Department of Orthopedics, Hennepin County Medical Center, Minneapolis, Minnesota, USA

46.2.1 Musculoskeletal trauma

Margaret L. Schwarze Massachusetts General Hospital, Boston, USA

17.6.2 Femoral artery

David Scott Melbourne, Australia

16.3 Organ and tissue preservation for transplantation

Euan M. Scrimgeour College of Medicine, Sultan Qaboos University, Sultanate of Oman

47.2.2 Schistosomiasis (bilharziasis)

John H. Scurr Consultant Surgeon, Middlesex and University College Hospitals, London, UK

18.1 Venous disorders, vascular malformations, and chronic ulceration in the lower limbs

John W. Sear Reader in Anaesthetics, John Radcliffe Hospital, Oxford, UK

10.1 Anaesthesia for surgeons

John D. Seigne Moffitt Cancer Center,Tampa, Florida, USA

39.1 Genitourinary anatomy

Jawaharlal W. B. Senaratne Clinical Research Fellow/Lecturer in Surgery, John Radcliffe Hospital, Oxford, UK

17.1 Pathobiology of atherosclerosis

17.2 Pathobiology of vasculitis

Robert C. Shamberger Senior Associate in Surgery, Children's Hospital, Boston, Massachusetts, USA

42.2 Thoracic surgical problems

42.5 Duplications of the gastrointestinal tract and mesenteric cysts

42.10 Paediatric solid tumours

J. Shapiro Joint Center for Otolaryngology, Boston, Massachusetts, USA

22.1 Dysphagia

Pai-Ching Sheen Professor of Surgery, Kaohsiung Medical College Hospital, Taiwan, Republic of China

47.1.7 Pyogenic liver abscess

Paul C. Shellito Assistant Professor of Surgery, Harvard Medical School; Associate Visiting Surgeon, Massachusetts General Hospital, Boston, USA

26.7 Angiodysplasia

Basil J. Shepstone Reader in Radiology, The Radcliffe Infirmary, Oxford, UK

13.1 Conventional radiology

13.7 Nuclear medicine techniques

Robert L. Sheridan Assistant Chief of Staff, Shriners Burns Institute, Boston, Massachusetts, USA

50.3 Acute burns

Bradford J. Shingleton Ophthalmic Consultants of Boston, Massachusetts, USA

45.1 Eye emergencies

Daniel Shoskes Division of Urology, Harbor-UCLA Medical Center, Torrance, California, USA

39.8 Cancer of the prostate

Timothy Simon General Surgery Resident, Emory University School of Medicine, Atlanta, Georgia, USA

14.1 Upper gastrointestinal endoscopy

14.5 Percutaneous endoscopic gastrostomy and jejunostomy

Kenny Sin National Heart Centre, Singapore

40.7.3 Atrial, ventricular, and atrioventricular septal defects

40.7.5 Abnormalities of the atrioventricular valves

Michael E. Sinclair Consultant Anaesthetist and Senior Clinical Lecturer, John Radcliffe Hospital, Oxford, UK

40.5 Principles of blood conservation, cardiac anaesthesia, and cerebral protection

40.6 Cardiovascular monitoring and postoperative care

C. Daniel Smith Chief, GI and General Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

20.4.2 Surgery of the adrenal gland

35.2 Umbilical, epigastric, and rarer abdominal wall hernias

Richard G. Smith Marshall Foundation Artificial Heart Program, University of Arizona Health Sciences Center, Tucson, USA

40.3 Circulatory support for the failing heart

T. C. Smyrk Associate Professor of Pathology, Mayo Clinic, Rochester, Minnesota, USA

48.1 Cancer genetics

Arthur J. Sober Assistant Chief of Dermatology, Massachusetts General Hospital, Boston, USA

38.3 Benign acquired pigmented lesions of the skin

38.5.1 Recognition, staging, and prognosis of cutaneous melanoma

38.6 Cutaneous problems occurring in the surgical patient

Patricia M. Solga Assistant Professor, University of Orthopedics, Inc., Providence, Rhode Island, USA

46.3.1 Paediatric fractures

Gavin P. Spickett Regional Department of Immunology, Royal Victoria Infirmary, Newcastle upon Tyne, UK

17.2 Pathobiology of vasculitis

17.3 Vasculitis

Dempsey S. Springfield Department of Orthopedics, Mount Sinai Medical Center,New York, USA

46.4 Management of bone tumours

Charles A. Staley Assistant Professor, Department of Surgical Oncology, Emory University School of Medicine, Atlanta, Georgia, USA

27.2 Primary appendiceal malignancies

48.6 Access for chemotherapy

David D. Stark Professor and Chairman, Department of Radiology, University of Nebraska Medical Center, Omaha, USA

13.3 Magnetic resonance imaging

Michael A. Steller Gynecologic Oncologist, Women & Infants Hospital, Providence, Rhode Island, USA

36.8.2 Cervical malignancy

Gordon M. Stirrat Professor of Obstetrics and Gynaecology, University of Bristol, UK

36.11 Caesarian section

John R. Stradling Professor, Osler Chest Unit, The Churchill Hospital,Oxford, UK

45.12 The surgical treatment of snoring and obstructive sleep apnoea

Robert M. Strauss Northwest Georgia Gastroenterology Associates, Marietta, USA

32.5 Ascites and its complications

Peter Styles Acting Director, MRC Magnetic Resonance Spectroscopy Unit, John Radcliffe Hospital, Oxford, UK

13.8 Nuclear magnetic resonance spectroscopy

David S. Sumner Chief, Section of Peripheral Vascular Surgery, Southern Illinois University School of Medicine, Springfield, USA

17.4 The non-invasive vascular diagnostic laboratory

Owen S. Surman Department of Psychiatry, Massachusetts General Hospital, Boston, USA

12.5 Psychological care

Marc F. Swiontkowski Professor and Head, University of Minnesota Department of Orthopedic Surgery, Minneapolis, USA

46.2.1 Musculoskeletal trauma

Mark S. Talamonti Department of Surgery, North Western University Medical School, Chicago, Illinois, USA

30.4 Cancers of the liver

P. K. H. Tam Professor, Department of Surgery, Queen Mary Hospital, Hong Kong

42.8 The abdominal wall

Gordon Tang Resident, Department of Neurosurgery, Emory University School of Medicine, Atlanta, Georgia, USA

42.13 Spina bifida

Hector M. Tarraza Maine Medical Center, Portland, USA

36.8.2 Cervical malignancy

36.8.3 Radical hysterectomy

G. M. Teasdale Professor of Neurosurgery, University of Glasgow, UK

44.2 Head injury

Peter J. Teddy Consultant Neurosurgeon, The Radcliffe Infirmary, Oxford, UK

44.11 Neurosurgery for intractable pain

Joel Tepper Chairman of Radiation Oncology, Uuniversity of North Carolina School of Medicine, Chapel Hill, USA

48.7.1 Principles of radiation oncology

R. V. Thakker Professor, MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital Campus, London, UK

20.3.1 Parathyroid disorders: molecular genetics and physiology

W. Hamish Thomson Consultant General Surgeon, Gloucestershire Royal Hospital, Gloucester, UK

28.1 Haemorrhoids or piles

David J. Tibbs Honorary Consulting Surgeon, John Radcliffe Hospital, Oxford, UK

18.1 Venous disorders, vascular malformations, and chronic ulceration in the lower limbs

Ronald G. Tompkins Chief of Burn and Trauma Services, Massachusetts General Hospital, Boston, USA

50.3 Acute burns

B. Trotman-Dickenson Research Fellow in Radiology, Royal Brompton National Heart and Lung Hospital, London, UK

41.1 Surgical anatomy and radiology of the chest

W. Turner Consultant Urologist, Addenbrooke's Hospital, Cambridge, UK

39.9.1 Tumours of the kidney

Robert G. Twycross MacMillan Clinical Reader in Palliative Medicine, Sir Michael Sobell House, The Churchill Hospital, Oxford, UK

49.1 Care of the dying patient

49.2 Cancer pain relief

Jerrold L. Vitek Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA

44.5 Surgery for Parkinson's disease and other movement disorders

Gus J. Vlahakes Associate Professor of Surgery, Massachusetts General Hospital, Boston, USA

40.7.2 Patent ductus arteriosus and aortic coarctation

40.7.9 Anomalies of pulmonary and systemic venous return

40.7.10 Miscellaneous congenital cardiac defects

John C. Wain Director, Lung Transplant Program, Massachusetts General Hospital, Boston, USA

22.3.1 Oesophageal diverticula

22.3.2 Achalasia

41.3 Benign and malignant tumours of the lung

S. A. Wall Consultant Plastic and Reconstructive Surgeon, Oxford Craniofacial Unit, The Radcliffe Infirmary, Oxford, UK

45.11 Facial injuries

Matthew J. Wall, Jr.
Associate Professor, Department of Surgery, Baylor College of Medicine, Houston, Texas, USA

40.8.4 Surgery for complications of ischaemic heart disease

Bryan F. Warren Consultant Gastrointestinal Pathologist, John Radcliffe Hospital, Oxford, UK

26.2.1 Pathology of non-specific chronic idiopathic inflammatory bowel disease

Ralph L. Warren Director of Trauma and Surgical Critical Care, Massachusetts General Hospital, Boston, USA

34.3 Trauma: injuries to abdominal wall and mesentery

C. Wastell Professor of Surgery Emeritus, Imperial College of Science, Technology and Medicine at Chelsea Westminster Hospital, London, UK

4.2 The surgery associated with HIV infection

Prasit Watanapa Professor, Department of Surgery, Siriraj Hospital, Bangkok, Thailand

47.2.1 Amoebiasis

D. J. Weatherall Regius Professor of Medicine, University of Oxford, UK

7 Molecular and cell biology in surgical practice

B. L. Weber Professor of Medicine and Genetics, University of Pennsylvania School of Medicine, Philadelphia, USA

48.1 Cancer genetics

Collin J. Weber William C. McGarity Professor of Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

20.4.2 Surgery of the adrenal gland

Edward W. Webster Department of Radiological Sciences, Massachusetts General Hospital, Boston, USA

48.7.2 Principles of radiation protection

Arnold-Peter C. Weiss University of Orthopaedics, Providence, Rhode Island, USA

46.2.3 The hand and wrist

Victor J. Weiss Emory University School of Medicine, Atlanta, Georgia, USA

17.13 Minimally invasive vascular surgery

Mark C. Weissler Division of Otolaryngology, University of North Carolina, Chapel Hill, USA

45.7 Cancer of the larynx and hypopharynx

John P. Welch Connecticut Surgical Group, P.C., Hartford, USA

26.4 Large-bowel obstruction

Samuel A. Wells Executive Director, American College of Surgeons, Chicago, Illinois, USA

20.1 Genetic aspects of endocrine disease

Stephen Westaby Consultant Cardiac Surgeon, Oxford Heart Centre, John Radcliffe Hospital, Oxford, UK

40.1 Development of surgery of the heart and great vessels

40.2 Cardiopulmonary bypass and myocardial protection

40.7.8 Congenital anomalies of the thoracic aorta and main pulmonary arteries

40.10 Surgery of the thoracic aorta

41.1 Surgical anatomy and radiology of the chest

41.11 Principles and practice of thoracic drainage

41.12 Thoracic trauma

Thomas E. Whitesides Emory Spine Center, Decatur, Georgia, USA

46.2.8 Inflammatory disorders of the spine

Richard I. Whyte Head, Division of Thoracic Surgery, UCSF Stanford Health Care, California, USA

41.4 Primary chest wall neoplasms

Joseph W. Wilkes Professional Affairs, Harvard Vanguard Medical Associates, Brookline, Massachusetts, USA

45.10 Oral and maxillofacial surgery

Andrew R. Wilkinson Professor of Paediatrics, John Radcliffe Hospital, Oxford, UK

42.1 General care of the paediatric patient, especially the newborn infant

Christopher G. Willett Department of Radiation Oncology, Massachusetts General Hospital, Boston, USA

24.7 Radiation enteritis

Gordon Williams Consultant Urologist, Hammersmith Hospital, London, UK

47.4 Obstetric vesicovaginal fistulae

Norman S. Williams Professor of Surgery, St Bartholomew's and the Royal London School of Medicine and Dentistry, London, UK

28.5 Faecal incontinence

R. C. N. Williamson Professor of Surgery, Imperial College School of Medicine, Hammersmith Hospital, London, UK

24.8 Small bowel tumours

47.2.1 Amoebiasis

Douglas W. Wilmore Frank Sawyer Professor of Surgery, Brigham and Women's Hospital, Boston, Massachusetts, USA

1 Response to injury

Roger S. Wilson Chairman, Department of Anesthesiology and Critical Care, Memorial Sloan-Kettering Cancer Center, New York, USA

12.1 Respiratory problems

Christopher G. Winearls Oxford Kidney Unit, The Churchill Hospital, Oxford, UK

12.3 Renal problems

Dietmar H. Wittmann Associate Professor of Surgery, Medical College of Wisconsin, Milwaukee, USA

4.1 Surgical infections

Ambaye Michael Wolde Obstetrician and Gynaecologist, Addis Abab Fistula Hospital, Ethiopia

47.4 Obstetric vesicovaginal fistulae

J. Wong Professor of Surgery, University of Hong Kong, Queen Mary Hospital, Hong Kong

47.6 Intrahepatic stones

William C. Wood Chairman, Department of Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

8 Design and interpretation of clinical trials

21.2 Cancer of the breast

37 Soft tissue tumours

T. R. Worthington Imperial College School of Medicine, Hammersmith Hospital, London, UK

24.8 Small bowel tumours

Cameron D. Wright Assistant Professor of Surgery, Massachusetts General Hospital, Boston, USA

22.5 Benign and malignant tumours

David A. Wrone Clinical Fellow, Dermatology Department, Massachusetts General Hospital, Boston, USA

38.6 Cutaneous problems occurring in the surgical patient

Michael J. Yaremchuk Chief of Craniofacial Surgery, Massachusetts General Hospital, Boston, USA

50.1 Principles and practice of plastic surgery

Charles J. Yeo Professor of Surgery and Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

33.4 Pancreatic cancer

Anthony P. C. Yim Professor, Department of Surgery, Prince of Wales Hospital, Hong Kong

14.7 Thoracoscopy

Jerry Yuan Urology Associates of North Georgia, Suwanne, USA

39.6 Male sexual dysfunction and infertility

Acknowledgements

We gratefully acknowledge permission to use material from other published sources as detailed below.

Section 1 Response to injury

Figure 2 adapted from: Frank SM *et al.* The catecholamine, cortisol, and hemodynamic response to mild perioperative hypothermia. *Anesthesiology* 1995; **82**: 83–93.

Figure 3 and **Figure 6** reproduced from Wilmore DW. Homeostasis. In: Sabiston DC, Jr, ed. *Textbook of surgery*, 15th edn. Philadelphia: Saunders, 1997: 55–67.

Figure 4 reproduced from Souba WB, Jr, Wilmore DW. Diet and nutrition in the care of the patient with surgery, trauma and sepsis. In: Shilo ME, Olson JA, Shike M, ed. *Modern nutrition in health and disease*. Philadelphia: Lippincott, Williams and Wilkins, 1994: 1222.

Figure 5 reproduced from Garlick PJ, Wernerman J. Protein metabolism in injury. In: Cooper GJ, Dudley HA, Cann DS, Little RA, Maynard RL, ed. *Scientific foundations of trauma*. Oxford: Butterworth Heinemann, 1997: 696.

Figure 7 reproduced from Fleck A, Colley CM, Meyers MA. Liver export proteins and trauma. *British Medical Bulletin* 1985; **41**: 265–73.

Figure 8 reproduced from Plank LD, Connolly AB, Hill GL. Sequential changes in severely septic patients during the first 23 days after onset of peritonitis. *Annals of Surgery*, in press.

Figure 9 reproduced from Wilmore DW, Goodwin CW, Aulick LH, Powanda MC, Mason AD JR, Pruitt BA, Jr. *Annals of Surgery* 1980; **192**: 491–504.

Figure 10 reproduced from Wolfe RR. Substrate metabolism in injury. In: Cooper CJ, Dudley HAF, Cann DS, Little RA, Maynard RL, ed. *Scientific foundations of trauma*. Oxford: Butterworth Heinemann, 1997: 1681.

Figure 11 reproduced from Wilmore DW, Aulick LH, Mason AD Jr, Pruitt BA, Jr. Influences of the burn wound on local and systemic responses to injury. *Annals of Surgery* 1977; **186**: 444–58.

Figure 12 reproduced from Wilmore DW, Aulick LH. Systemic responses to injury and the healing wound. *Journal of Parenteral and Enteral Nutrition* 1980; **4**:149–51.

Figure 16 reproduced from Wilmore DW. Catabolic illness: strategies for enhancing recovery. *New England Journal of Medicine* 1991; **325**: 695–702.

Figure 17 reproduced from West MA, Hackman DJ, Baker J, Rodriquez JL, Bellingham J, Rotstein OD. Mechanism of decreased *in vitro* response during macrophage cytokine release after exposure to carbon dioxide. *Annals of Surgery* 1997; **226**: 179–90.

Table 3 adapted in part from Wannemacher RW, Jr. Protein metabolism (applied biochemistry). In: Ghadimi H, eds. *Total parenteral nutrition*: *premises and promises*. New York: Wiley, 1975: 113.

Table 5 adapted from Wilmore DW, Aulick LH, Mason AD, Jr., Pruitt BA, Jr. Influence of the burn wound on local and systemic responses to injury. *Annals of Surgery* 1977; **186**: 444–58.

Table 7 reproduced from Michie HR, Wilmore DW. Sepsis, signals and surgical sequelae (a hypothesis). *Archives of Surgery* 1990; **125**: 531–6.

Table 8 adapted from Bessey PQ, Watters JM, Aoki TT, Wilmore DW. Combined hormonal infusion simulates the metabolic response to injury. *Annals of Surgery* 1984; **200**: 264–81.

Section 2 Trauma and shock

Figure 4 reproduced from Feliciano DV. *Advances in Trauma* 1987; **2**: 179–206.

Figure 6 reproduced from Feliciano DV, *et al.* Vascular injuries of the chest and abdomen. In Rutherford RB, *et al.*, eds. *Vascular surgery*. 3rd edn. Philadelphia: WB Saunders, 1989: 588–603.

Figure 8 reproduced from Rozycki GS, *et al.* Surgeon-performed ultrasound for the assessment of truncal injuries. Lessons learned from 1540 patients. *Annals of Surgery* 1998; **228**: 557–67.

Figure 10 reproduced from Feliciano DV. Selective management of splenic injuries. In Hurst JM, ed. *Common problems in trauma*. Chicago: Year Book Medical Publishers, 1987: 194–203.

Figure 11 reproduced from Feliciano DV. Abdominal trauma. In Schwartz SI, Ellis H, eds. *Maingot's abdominal operations*, 9th edn. Norwalk: Appleton and Lange, 1989: 457–512.

Figure 12 reproduced from Martin TD, *et al.* Severe duodenal injuries. Treatment with pyloric exclusion and gastrojejunostomy. *Archives of Surgery* 1983; **118**: 631–5.

Figure 13 reproduced from Cushman JG, Feliciano DV. Contemporary management of pancreatic trauma. In: Maull KI, *et al.* eds. *Advances in trauma and critical care*. St Louis: Mosby, 1995; **10**: 309–36.

Table 4 reproduced from Barnes FC. *Cartridges of the world*., 7th edn. Northbrook, IL: DBI Books, Inc., 1993.

Table 5, **Table 6** and **Table 7** reproduced from Feliciano DV. Patterns of injury. In: Feliciano DV, Moore EE, Mattox KL, eds. *Trauma*, 3rd edn. Stamford: Appleton and Lange, 1996: 85–103.

Section 3 Surgical nutrition

Figure 1 reproduced from Munro MN. Parenteral nutrition: metabolic consequences of bypassing the gut and liver. In: Green HL, Holliday MA, Munro MN, eds. *Clinical nutrition update: amino acids*, 1977: 141, by permission of the American Medical Association.

Figure 2 reproduced from West JB, ed. *Best and Taylor's physiological basis of medical practice*, 12th edn. Baltimore: Williams and Wilkins, 1991: 730.

Figure 3 reproduced from Gann DS, Foster AH. Endocrine and metabolic responses to injury. In: Schwartz SI, ed. *Principles of surgery*, 6th edn. New York: McGraw-Hill, 1994: 32.

Figure 4 reproduced from Kinney JM, ed. The application of indirect calorimetry to clinical studies. In: *Assessment of energy metabolism in health and disease*, 1980: 42. Used by permission of Ross Products Division, Abbott Laboratories, Inc., Columbus, Ohio 43216, USA.

Figure 5, **Figure 11**, **Figure 12**, and **Figure 13** used by permission of the Department of Parenteral and Enteral Nutrition, University of Cincinnati Medical Center.

Figure 6 and **Figure 7** redrawn from Fischer JE, Freund HR. Central hyperalimentation. In: Fischer JE, ed. *Surgical nutrition*. Philadelphia: Lippincott-Raven, 1983: 696–7.

Figure 8 and **Figure 9** redrawn from MacLennan SE, Vascular access techniques. In: The University of Cincinnati Residents, ed. *The Mont Reid manual*, 4th edn. St Louis: Mosby-Year Book, Inc., 1997: 678–9.

Figure 10 redrawn from Shires GT, Shires GT, III, Lowry SF, Fluid, electrolyte, and nutritional management of the surgical patient. In: Schwartz SI, ed. *Principles of surgery*, 6th edn. New York: McGraw-Hill, 1994: 91.

Figure 14 reproduced from S Berry. Nutrition. In: The University of Cincinnati Residents, ed. *The Mont Reid manual*, 4th edn. St Louis: Mosby-Year Book, Inc., 1997: 60.

Table and **Table 2** reproduced from Cerra FB. Metabolic and nutritional support. In: Feliciano DV, Moore EE, Mattox EL, eds. *Trauma*, 3rd edn. Stamford, CT: Appleton and Lange, 1996: 1157.

Section 4 Surgical infection and AIDS

4.1 Surgical infections

Figure 2 reproduced from Wittmann DH. *Intra-abdominal infection: pathophysiology and treatment*. New York: Marcel Dekker, 1991.

Figure 3 Data for exogenous bacteria taken from Wittmann DH. Wertigkeit der Antibiotika bei posttraumatischen Knocheninfektionen (evaluation of antibiotics in bone infections). In: Cotta H, Braun A, eds. *Knochen-und Gelenkinfktionen (bone infections)*. Berlin: Springer Verlag.

Data for endogenous bacteria from Howard R. Microbes and their pathogenicity. In: Simmons RL, Howard RJ, eds. *Surgical infectious diseases*. Norwalk CT: Appleton Century Crofts, 1986:11–28.

Table 8 reproduced from Walker A, *et al. Archives of Surgery*, 1988; **123**: 908–11. Copyright, 1988. American Medical Association

4.2 The surgery associated with HIV infection

Figure 1 redrawn after Jollies *et al. British Medical Journal* 1996; **312**.

Section 5 Medical problems in the surgical patient

5.1 Respiratory problems

Figure 1 reproduced from Nunn JF. *Applied respiratory physiology*. London: Butterworth, 1987.

Figure 2, Figure 3 and Figure 4 reproduced from Froese AB, Bryan AC. Effects of anesthesia and paralysis on diaphragmatic mechanics in man. *Anesthesiology*, 1971; **41**: 242.

Figure 5 reproduced from Stevenson GW, *et al.* The effects of anesthetic agents on the human immune response. *Anesthesiology*, 1990; **72**: 542.

Section 10 Anaesthesia and the operating room

10.1 Anaesthesia for surgeons

Figure 1 adapted from Wood JH, Kee DB. Hemorheology of the cerebral circulation in stroke. *Stroke* 1985; **16**: 765–72, with permission of authors and publishers.

10.4 Electrical safety in anaesthesia and surgery

Figure 2 reproduced from IEC 479–1. *Effects of passing current through the human body*. Geneva: International Electrotechnical Commission, 1984, who retain the copyright.

Section 11 Care of the critically-ill patient

11.1 Guidelines and scoring systems

Figure 1 reproduced from Garrard CS, Foëx P, Westaby S, ed. *Principles and practice of critical care*. Oxford: Blackwell Science, 1997: 83.

11.4 Respiratory aspects

Table 2 reproduced from Murray JF, *et al.* An expanded definition of the adult respiratory distress syndrome. *American Reviews of Respiratory Disease*, 1988; **138**: 720–3.

Section 12 Medical problems in the surgical patient

12.3 Renal problems

Figure 4 redrawn from Forni LG, Hilton PJ. Continuous haemofiltration in the treatment of acute renal failure. *New England Journal of Medicine* 1997; **336**: 1303–9.

Section 14 Endoscopy

14.7 Thoracoscopy and video-assisted thoracic surgery

Figure 1 reproduced from The history of thoracoscopic surgery. *Annals of Thoracic Surgery*, 1993; **56**: 610–14, by permission of the Society of Thoracic Surgeons.

Figure 2 and **Figure 5** reproduced from Yim APC, Thoracoscopic major lung resection revisited. *Journal of Endoscopic and Laparoscopic Surgery of Asia*, 1996; **1**: 49, by permission of The Society of Endoscopic and Laparoscopic Surgeons of Asia.

Section 15 Endoscopic surgery

Figure 1 adapted from Hunter JG, ed. *Baillière's Clinical Gastroenterology*, Vol. 7, No. 4, *Laparoscopic Surgery*, 1993: 758.

Figure 5 and **Figure 6** reproduced from Toouli J, Gossot D, Hunter JG, eds. *Endosurgery*. Edinburgh: Churchill Livingstone, 1997: 42.

Section 16 Transplantation

16.1 Basic transplantation immunology

Figure 3 reproduced from Bjorkman PJ, *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, 1987; **329**: 506–12. © 1987 Macmillan Magazines Ltd.

Figure 6 and Figure 10 reproduced from Auchincloss H, Jr, Sykes M, Sachs DH. Transplantation immunology. In: Paul WE, ed. *Fundamental immunology*. Philadelphia: Lippincott-Raven, 1998: 1175–236.

16.3 Organ and tissue preservation for transplantation

Figure 1 reproduced from Carrel A, Lindbergh C. *The culture of organs*. London: Hamish Hamilton, 1938. Every endeavour was made to clear permission, and the publisher would be very happy to hear from the copyright holder whom we were unable to trace.

Figure 2 reproduced from Morris PJ, ed. *Kidney transplantation: principles and practice*. 3rd edn. Philadelphia: WB Saunders Co., 1988.

16.5 Kidney transplantation

Figure 1, Figure 13, and Figure 16 reproduced from Cecka JM. The UNOS Scientific Renal Transplant Registry. In: Cecka JM, Terasaki PI, eds. *Clinical transplants 1998*. UCLI Tissue Typing Laboratory, 1999: 1–16.

Figure 2 reproduced from Morris PJ, Johnson RJ, Fuggle S, Belger MA, Briggs D. Factors influencing outcome of primary cadaveric renal transplantation in the UK. *Lancet* 1999 (in press).

16.6 Liver transplantation

Figure 2 reproduced from Cosimi AB. The donor and donor nephrectomy. In: Morris PJ, ed. *Kidney transplantation: principles and practice*, 3rd edn. Philadelphia: WB Saunders Co., 1988: 110.

Figure 3, Figure 5, Figure 6, and Figure 9 adapted from Jenkins RL. Recipient hepatectomy. In: Warren KW, Jenkins RL, Steele GD, eds. *Atlas of surgery of the liver, pancreas, and biliary tract*. Norwalk, CT: Appleton & Lange, 1991.

Figure 4 reproduced from Tzakis A, *et al.* Venous grafts for liver transplantation in patients with portal vein thrombosis. *Transplantation*, 1989; **48**: 531. Copyright Williams & Wilkins, 1989.

Figure 7 reproduced from Jenkins R, Benotti P, Bothe A, Rossi RL. Liver transplantation. *Surgical Clinics of North America*, 1985; **65**: 103.

Figure 10 reproduced from Emond JC, *et al.* Transplantation of two patients with one liver. *Annals of Surgery*, 1990; **212**: 16.

Figure 12 reproduced from Terpstra OT, *et al.* Auxiliary partial liver transplantation for end-stage chronic liver disease. *New England Journal of Medicine*, 1988; **319**: 1508.

Section 17 The arteries

17.4 The non-invasive vascular diagnostic laboratory

Figure 1 reproduced from Kohler TR, Nance DR, Cramer MM, Vandenburghe N, Strandness DE Jr. Duplex scanning for diagnosis of aortoiliac and femoropopliteal disease: a prospective study. *Circulation* 1987; **76**: 1074–80.

Figure 4 reproduced from Sumner DS, Mattos MA, Hodgson KJ. Surveillance program for vascular reconstructive procedures. In: Yao JST, Pearce WH, eds. *Long-term results in vascular surgery*. Norwalk, Connecticut: Appleton & Lange, 1993: 33–59.

Figure 5 reproduced from Baker JD, Barker WF, Machleder HI. Evaluation of extracranial cerebrovascular disease with ocular pneumoplethysmography. *American Journal of Surgery*, 1978; **136**: 206–8.

Figure 7 reproduced from Sumner DS, Thiele BL. The vascular laboratory. In: Rutherford RB, ed. *Vascular surgery*, 4th edn. Philadelphia: WB Saunders, 1995; 45–64.

Figure 8 reproduced from Yao ST. Haemodynamic studies in peripheral arterial disease. *British Journal of Surgery* 1970; **57**: 761–6.

Figure 11 reproduced from Sumner DS. Rational use of noninvasive tests in designing a therapeutic approach to severe arterial disease of the legs. In: Puel P, Boccalon H, Enhalbert A, eds. *Hemodynamics of the limbs – 2*. Toulouse, France: GEPESC, 1981: 369–76.

Figure 12 reproduced from Rutherford RB, Lowenstein DH, Klein MF. Combining segmental systolic pressures and plethysmography to diagnose arterial occlusive disease of the legs. *American Journal of Surgery*, 1979; **138**: 211–18.

Figure 13 reproduced from Strandness DE Jr, Sumner DS. *Hemodynamics for surgeons*. Grune & Stratton, New York, 1975.

Figure 16, Figure 17, Figure 19, and Figure 28 reproduced from Sumner DS. Clinical evaluation of peripheral vascular disease. In: Ritchie WP, Jr, Steele G Jr, Dean RH, eds. *General surgery*. Philadelphia: JB Lippincott, 1995: 551–70.

Figure 18 reproduced from Ramsey DE, Manke DA, Sumner DS. Toe blood pressure: a valuable adjunct to ankle pressure measurement for assessing peripheral arterial disease. *Journal of Cardiovascular Surgery* 1983; **24**: 43–8.

Figure 20 and Figure 22 reproduced from Sumner DS. Noninvasive assessment of upper extremity and hand ischemia. *Journal of Vascular Surgery* 1986; **3**: 560–4.

Figure 21 reproduced from Sumner DS, Lambeth A, Russell JB. Diagnosis of upper extremity obstructive and vasospastic syndromes by Doppler ultrasound, plethysmography, and temperature profiles. In: Puel P, Boccalon H, Enhalbert A, eds. *Hemodynamics of the limbs – 1*. Toulouse, France: GEPESC, 1981: 365–73.

Figure 23 reproduced from Sumner DS, Mattos MA, Hodgson KJ. Surveillance program for vascular reconstructive procedures. In: Yao JST, Pearce WH, eds, *Long-term results in vascular surgery*. Norwalk, Connecticut: Appleton & Lange, 1993: 33–59.

Figure 25 and Figure 27 reproduced from Sumner DS. Is color coded Doppler a waste of money? In: Bernstein EF, ed. *Vascular diagnosis*, 4th edn. St Louis: Mosby, 1993: 125–32.

17.5.2 Femoral and distal arteries

Figure 10 reproduced from Michael JA. Choice of material for above-knee femoropopliteal bypass graft. *British Journal of Surgery*, 1989; **76**: 7–14, by permission of

the publishers, Butterworth & Co.

17.7.1 Carotid artery

Figure 14 reproduced from The ECST Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgical Trial (ECST). *Lancet* 1998; **351**: 1379–87.

Figure 15 reproduced from Barnett HJM *et al.* Benefit of endarterectomy in patients with symptomatic, moderate or severe stenosis. *New England Journal of Medicine* 1998; **339**: 1415–25.

17.7.3 Carotid body tumours

Figure 1 reproduced from McPherson *et al.* Carotid body tumours and other cervical paragangliomas. *British Journal of Surgery*, 1989; **67**: 34, with permission from Blackwell Scientific Publications, Oxford.

17.9 Renovascular disease

Figure 14 reproduced from Bell, Jamieson, and Ruckley. *Surgical management of vascular disease*. London: WB Saunders Co., 1990: 761

17.10.1 Arterial emboli: limbs

Figure 1 and Figure 6 reproduced from Campbell W. Acute limb ischaemia. *Surgery*, 1990; **81**: 1937–41 by courtesy of The Medicine Group (UK) Ltd.

17.11 Arterial and venous injuries

Figure 3 reproduced from Feliciano DV. Trauma to the aorta and major vessels. *Chest Surgery Clinics of North America*, 1997; **7**: 305–23.

Figure 9 reproduced from Feliciano DV. Vascular injuries. In: Maull KI, Cleveland HC, Strauch GO, Wolferth CC, eds. *Advances in trauma*. Chicago: Year Book Medical Publishers, 1987; **2**: 179–206.

Figure 10 reproduced from Feliciano DV. Injuries to great vessels of the abdomen. In: Wilmore DW, Cheung LY, Harken AH, Holcroft JW, Meakins JL, eds. *Scientific American surgery*. New York: Scientific American, Inc., 1999: Section IV, Chapter 9, 1–9.

Section 20 Endocrine disease

20.1 Genetic aspects of endocrine disease

Figure 2 redrawn from Lairmore TC, Wells SA Jr. Multiple endocrine neoplasia syndromes. In: Sabiston DC Jr. *Sabiston textbook of surgery: the biological basis of modern surgical practice*. Philadelphia: WB Saunders Company, 1997: 640.

Figure 4 redrawn from Lairmore TC, Wells SA Jr. Molecular biology of endocrine tumors. In: DeVita VT Jr, Hellman S, Rosenberg SA. *Cancer: principles and practice of oncology*, 5th edn. Philadelphia: Lippincott-Raven Publishers, 1997: 1621.

Figure 5 reproduced from Pierotti MA, *et al.* Characterization of an inversion on the long arm of chromosome 10 juxtaposing D10S170 and RET and creating the oncogenic sequence RET/PTC. *Proceedings of the National Academy of Sciences of the USA*, 1992; **89**: 1616–20.

Figure 6 redrawn from Arnold A. Genetic basis of endocrine disease 5: molecular genetics of parathyroid gland neoplasia. *Journal of Clinical and Endocrinological Metabolism*, 1993; **77**: 1108–12.

20.4.2 Surgery of the adrenal gland

Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 10, and Figure 13 redrawn from Peplinski G, Norton J. *Current techniques in general surgery—adrenalectomy*, Vol. 3. New York: Lawrence Dellacorte, 1994.

Figure 7, Figure 8, Figure 9, Figure 11, and Figure 12 redrawn from Weber CJ, Amerson DR, Smith D. Laparoscopic adrenalectomy: the new gold standard. *World Journal of Surgery*, 1999; **23**(11).

Section 22 The oesophagus

22.2.2 Benign oesophageal strictures

Figure 2 reproduced from Zaninotto G, *et al.* Esophageal function in patients with reflux induced strictures and its relevance to surgical treatment. *Annals of Thoracic Surgery*, 1989; **47:**362–70.

Figure 6 redrawn from Dawson SL, *et al.* Severe esophageal strictures: indications for balloon catheter dilatation. *Radiology*, 1984; **153**:631–5.

22.3.1 Oesophageal diverticula

Figure 1 redrawn from Negus VE. Pharyngeal diverticula. Observations on their evolution and treatment. *British Journal of Surgery*, 1950; **38**:129–46, by permission of the publishers, Butterworth-Heinemann Ltd.

Figure 3 and Figure 6 redrawn from Trastek VF, Payne WS. Esophageal diverticula. In: Shields TW, ed. *General thoracic surgery*. 3rd edn. Philadelphia: Lea & Febiger, 1989: 989–1001.

22.3.2 Achalasia

Figure 2 reproduced from Katz PO. Achalasia. In: Castell DO, Richter JE, Dalton CB, eds. *Esophageal motility testing*. Norwalk, CT: Appleton & Lange, 1987: 107–17.

Figure 3 redrawn from Ellis FH, *et al.* Esophagomyotomy for esophageal achalasia: experimental, clinical and manometric aspects. *Annals of Surgery*, 1967; **166**: 640–56.

22.5 Benign and malignant tumours

Table 1 reproduced from Roth JA, Ruckdeschel JC, Weisenburger TH, eds. *Thoracic oncology*. Philadelphia: WB Saunders Co., 1989: 296.

Table 2 adapted from Akiyama H, *et al.* Principles of surgical treatment for carcinoma of the esophagus. Analysis of lymph node involvement. *Annals of Surgery*, 1981; **194**: 438–46.

Table 6 adapted from Muller JM, *et al.* Surgical therapy of oesophageal carcinoma. *British Journal of Surgery*, 1990; **77**: 845–57.

Section 23 The stomach and duodenum

23.4 Gastric volvulus and acute gastric dilatation

Figure 1 and Figure 2 redrawn from Buchanan J. Volvulus of the stomach. *British Journal of Surgery* 1930; **18**: 98–108.

Section 24 The small intestine

24.2.2 Surgery for Crohn's disease of the small intestine

Figure 11 reproduced from *The atlas of gastrointestinal endoscopy*, 2nd edn. London: Gower Medical Publishing.

Section 26 The colon and rectum

26.2.2 Colonic inflammation – diagnosis and management

Figure 3 reproduced from *The atlas of gastrointestinal endoscopy*, 2nd edn. London: Gower Medical Publishing.

26.2.3 Surgical management of ulcerative colitis

Figure 4 reproduced from *The atlas of gastrointestinal endoscopy*, 2nd edn. London: Gower Medical Publishing.

26.2.4 Crohn's disease of the colon

Figure 2 reproduced from *The atlas of gastrointestinal endoscopy*, 2nd edn. London: Gower Medical Publishing.

26.3 Colorectal tumours

Figure 19 reproduced from Haggit RC, Glotzbach RE, Soffer EE, *et al.* Prognostic factors in colorectal carcinomas arising in adenomas: implications for lesions removed by endoscopic polypectomy. *Gastroenterology* 1985; **89**: 328.

Section 28 The anus

28.1 Haemorrhoids or piles

Figures in this chapter have been reproduced from the following publications:

Thomson WHF. *British Journal of Surgery*, 1975; **62**: 542–52, by permission of Blackwell Scientific Publications, Oxford.

Thomson WHF. Haemorrhoids. *Surgery*, 1985; **19**: 453–6, by permission of The Medicine Group (Journals) Ltd.

Thomson WHF. Clotting of the sacculated venous plexus of the anus. *Surgery*, 1992; **10**: 2:25–7 by permission of The Medicine Group (Journals) Ltd.

Thomson WHF. *Schweizerische Rundschau fur Medizin (PRAXIS)*, 1982; **82**(71): 107.

Thomson WHF. *Lancet*, 1982; **2**: 467–8. ©The Lancet Ltd.

Thomson WHF. Haemorrhoids and all that. *The Practitioner*, 1982; **April**: 619–28.

Thomson WHF. *Archives Francaises des Maladies de l'Appareil Digestil*, 1976; **65**: 523–7. ©Masson, Paris.

Thomson WHF. A new look at hemorrhoids. *Medical Times*, 1976: November.

28.4 Anorectal abscesses and fistula-in-ano

Figure 1, Figure 2, Figure 4, and Figure 8 reproduced from Goldberg S, Gordon, Nivatvong S, eds. *Essentials of anorectal surgery*. Philadelphia: JB Lippincott Company, 1980.

Figure 3, Figure 5, Figure 6, Figure 7, and Figure 9 reproduced from Schwartz *et al. Principles of surgery*, 5th edn. New York: McGraw-Hill.

Figure 10 reproduced from Goldberg S, Garcia-Aguilar J. The cutting seton. In Phillips RKS, Lunniss PJ, eds. *Anal fistula: surgical evaluation and management*. London: Chapman & Hall, 1996: 99.

Figure 11(b) reproduced from Matos D, Lunniss PJ, Phillips RKS. Total sphincter conservation in high fistula-in-ano. *British Journal of Surgery*, 1993; **80**: 802.

Section 29 Sacrococcygeal pilonidal sinus

Figure 2 adapted from Nivatvongs, S. Pilonidal disease. In Gordon P, Nivatvongs S, eds. *Principles and practice of surgery for the colon, rectum, and anus*. St Louis: Quality Medical, 1992. reproduced from Rakinic J. Sacrococcygeal pilonidal sinus. In Cameron J, ed. *Current surgical therapy*, 6th edn. Philadelphia: Mosby, 1998.

Section 30 The liver

30.1 Surgical anatomy of the liver and biliary tree

Figure 1, Figure 2, and Figure 3 redrawn from Bismuth H. Surgical anatomy and anatomical surgery of the liver. In: Blumgart LH, ed. *Surgery of the liver and biliary tract*. Edinburgh: Churchill Livingstone, 1988.

Figure 4 and Figure 5 redrawn from Smadja C, Blumgart LH. The biliary tract and anatomy of biliary exposure. In: Blumgart LH, ed. *Surgery of the liver and biliary tract*. Edinburgh: Churchill Livingstone, 1988.

30.2 Hepatic trauma

Table 1 reproduced from Moore EE, Cogbill TH, Jurkovich GJ, *et al. Journal of Trauma* 1995; **38**: 323–4.

Section 31 The biliary tract

31.1 Benign diseases of the biliary tract

Figure 1, **Figure 14**, **Figure 15**, **Figure 42**, and **Figure 49** redrawn from Blumgart LH, ed. *Surgery of the liver and biliary tract*. Edinburgh: Churchill Livingstone, 1988.

Figure 2 redrawn from Sherlock S. *Diseases of the liver and biliary system*. 6th edn. Oxford: Blackwell Scientific Publishers, 1981.

31.2 Malignant diseases of the biliary tract

Table 1 and **Table 3** reproduced from *AJCC Manual for staging of cancer*, 4th edn. Philadelphia: Lippincott-Raven Publishers, 1992.

Figure 3 reproduced from Gall FP *et al.* Radical operations for carcinoma of the gallbladder: present status in Germany. *World Journal of Surgery* 1991; **15**: 328–36.

Figure 4 and **Figure 5** reproduced from Tsukada K, *et al.* Outcome of radical surgery for carcinoma of the gallbladder according to the TNM stage. *Surgery*, 1996; **120**: 816–22.

Figure 6 reproduced from Nakeeb A, Pitt HA, *et al.* Cholangiocarcinoma: a spectrum of intrahepatic, perihilar, and distal tumors. *Annals of Surgery*, 1996; **224**: 463–73.

Figure 7 and **Figure 8** redrawn from Cameron JC, *et al. Atlas of surgery*, volume 1. St Louis: Mosby, 1990.

Figure 9 reproduced from Sugiura Y, *et al.* Extensive resection of the bile ducts combined with liver resection for cancer of the main hepatic duct junction; a cooperative study of the Keio Bile Duct Cancer Study Group. *Surgery*, 1994; **114**: 445–7.

Section 32 Portal hypertension

32.5 Ascites and its complications

Table 3 modified from Garcia-Tsao, *Gastroenterology Clinics of North America* 1992; **21**: 257.

Table 4 reproduced from Epstein M. Hepatorenal syndrome. In: Epstein M, ed. *Diseases of the kidney*. 3rd edn. Baltimore: Williams & Wilkins, 1988: 92.

Section 33 The pancreas

33.4 Pancreatic cancer

Figure 1 and **Figure 2** reproduced from Wilentz RE, Hruban RH. Pathology of cancer of the pancreas. *Surgical Oncology Clinics of North America* 1998; **7**: 43–65.

Figure 3, **Figure 4**, **Figure 5**, **Figure 6** and **Figure 7**, **Figure 9**, and **Figure 10** reproduced from Yeo CJ, Cameron JL. Pancreatic cancer. *Current Problems in Surgery*, in press.

Figure 8 reproduced from Howard JM, Idezuki Y, Ihse I, Prinz RA, eds. *Surgical diseases of the pancreas*, 3rd edn. Baltimore: Williams & Wilkins, 1998: 510 (Fig. 52.8).

Figure 11 and **Figure 14** reproduced from Cameron JL. *Atlas of surgery*, Vol. 1. Toronto: BC Decker, 1990: 427 (Image V), 435 (Image H).

Figure 12 reproduced from Yeo CJ, Cameron JL. The pancreas. In: Hardy JD, ed. *Hardy's textbook of surgery*, 2nd edn. Philadelphia: JB Lippincott Co., 1988: 718 (Fig. 28–9).

Figure 13 reproduced from Yeo CJ *et al.* A prospective randomized trial of pancreaticogastrostomy versus pancreaticojejunostomy after pancreaticoduodenectomy. *Annals of Surgery* 1995; **222**: 582 (Fig. 1).

Figure 15 reproduced from Yeo CJ *et al.* Pancreaticoduodenectomy for pancreatic adenocarcinoma: postoperative adjuvant chemoradiation improves survival: a prospective, single institution experience. *Annals of Surgery* 1997; **225**: 627 (Fig. 2).

Table 1 modified from Gold EB, Goldin SB. Epidemiology of and risk factors for pancreatic cancer. *Surgical Oncology Clinics of North America* 1998; **7**: 67–91.

Table 3 and Table 4 reproduced from Hruban RH, Petersen GM, Ha PK, Kern SE. Genetics of pancreatic cancer: from genes to families. *Surgical Oncology Clinics of North America* 1998; **7**: 1–23.

Table 6 adapted from Japan Pancreas Society. *General rules for surgical and pathological studies on cancer of the pancreas*, 3rd edn. Tokyo: Kanehara and Co., 1987.

Table 7 modified from Lillemoe KD, Sauter PK, Pitt HA, Yeo J, Cameron JL. Current status of surgical palliation of periampullary carcinoma. *Surgery, Gynecology, and Obstetrics* 1993; **176**: 1–10.

Table 8 reproduced from Yeo CJ. Management of complications following pancreaticoduodenectomy. *Surgical Clinics of North America* 1995; **75**: 913–24.

Section 34 The abdomen

34.6 Retroperitoneal neoplasms

Figure 2 and **Figure 3** reproduced from Karakousis CP, Velez AF, Gerstenbluth R, Driscoll DL. Resectability and survival in retroperitoneal sarcomas. *Annals of Surgical Oncology* 1996; **3**: 150–8 with permission from Lippincott-Raven Publishers.

Section 35 Hernias of the abdominal wall

35.1 Inguinal and femoral hernias

Figure 1, **Figure 4**, **Figure 5** and **Figure 8** redrawn from Devlin HB and Kingsnorth AN. *Management of abdominal hernias*, 2nd edn. London: Lippincott Williams & Wilkins, 1998.

Figure 2 and **Figure 7** redrawn from Devlin HB. *Management of abdominal hernias*. Oxford: Butterworth-Heinemann.

Section 36 Gynaecology and obstetrics in general surgery

36.5 Ovarian tumours

Table I reproduced from International Federation of Gynaecology and Obstetrics. Annual report on the results of treatment in gynaecological cancer. *International Journal of Gynaecology and Obstetrics* 1989; **28**: 189–90.

36.7 Endometriosis

Figure 4 reproduced from *Fertility and Sterility*, 1985: 351–5. The American Fertility Society.

36.9 Perineal abscesses

Figures in this chapter were redrawn from:

Nichols DH, Randall CL, eds. *Vaginal surgery*, 3rd edn. Baltimore: Williams & Wilkins, 1989: 165.

Monaghan JM, ed. *Bonney's gynaecological surgery*. 9th edn. London: Bailliere Tindall, 1986:120–1.

36.13 Surgery for infertility

Figure 1, **Figure 2**, **Figure 3**, **Figure 4** and **Figure 5** reproduced from Rock JA, Thompson JD. *TeLinde's operative gynecology*, 8th edn. Philadelphia: Lippincott-Raven, 1995.

Table 2 reproduced from Thompson JD, Rock JA. *TeLinde's operative gynecology*, 7th edn. Philadelphia: Lippincott-Raven, 1992.

Section 37 Soft tissue tumours

Table 2 adapted from Suit H, *et al. Journal of Clinical Oncology*, 1988; **6**: 854–62.

Section 38 Skin

38.3 Benign acquired pigmented lesions of the skin

Table 1 and **Table 2** adapted from Sober AJ, Koh HK. Melanoma and other pigmented skin lesions. In: Wilson J, *et al.*, eds. *Harrison's principles of internal medicine*. 12th edn. New York: McGraw-Hill, 1991:1634–8.

Table 3 adapted from Kang S, Sober AJ. Disturbances of melanin pigmentation. In: Moschella SL, *et al.*, eds. *Dermatology*. 3rd edn. Eastbourne: WB Saunders Co., 1992:1442–74.

Table 4 adapted from Lucky AW. Pigmentary abnormalities in genetic disorders. In: Noidlund JJ, ed. *Dermatologic Clinics, Pigmentation Disorders*. Eastbourne: WB Saunders Co., 1988: 201.

38.4 Premalignant and malignant tumours of the skin

Table 1 modified from Lupton GP. Cutaneous metastases. *Journal of the Association of Military Dermatology*, 1990; **16**: 18–25.

38.5.1 Cutaneous melanoma

Table 1 reproduced from Koh HK, *et al.* Sunlight and cutaneous malignant melanoma: evidence for and against causation. *Photochemistry, Photobiology*, 1990; **52**: 765–79.

Table 3, **Table 6**, and **Table 7** reproduced from Ho VC, Sober AJ. Therapy for cutaneous melanoma: an update. *Journal of the American Academy of Dermatology*, 1990; **22**: 159–76.

Table 4 reproduced from Sober AJ, Koh HK. Melanoma and other pigmented skin lesions. In: Wilson JD, *et al.*, eds. *Harrison's principles of internal medicine*. 12th edn. New York: McGraw-Hill, 1991: 1734–8.

Table 5 adapted from Beahrs OH, *et al*, eds. *Manual for staging of cancer*. 3rd edn. Philadelphia: JB Lippincott, 1988: 143.

Table 8 and **Table 9** adapted from Ho VC, Sober AJ. Therapy for cutaneous melanoma: an update. *Journal of the American Academy of Dermatology*, 1990; **22**: 159–76.

Section 39 Urology

39.3 Trauma to the urinary tract

Figures in this chapter have been redrawn from Burke JF, Boyd RJ, McCabe CT, eds. *Trauma management*. Chicago: Mosby-Year Book, 1988: 126–39.

39.5 Urinary stone disease

Figure 8 reproduced from Ueno A, *et al. Urology*, 1977; **10**: 545.

33.7 Male sexual dysfunction and infertility

Figure 1 redrawn from Wespes E. Cavernovenous leakage. In: deVere White R, ed. *Problems in urology*. Philadelphia: JB Lippincott Co., 1987: **1**, No. 3.

39.8 Cancer of the prostate

Figure 2 and **Figure 3** redrawn from Walsh PC. Radical prostatectomy. In: *Campbell's urology*. 5th edn. Philadelphia: WB Saunders Co.

Table 1 reproduced from Chapter 34, *Manual for staging of cancer*. 5th edn, 1998. Used with the permission of the American Joint Committee on Cancer (AJCC®), Chicago, Illinois. The original source for this material is the *AJCC® Cancer staging manual*. 5th edn, 1997 published by Lippincott-Raven Publishers, Philadelphia, Pennsylvania.

Table 2 reproduced from Richardson TD, Oesterling JE. Age-specific reference ranges for serum PSA. *Urologic Clinics of North America* 1997; **24**: 349.

Table 4 reproduced from Shipley WU, *et al.* Treatment related sequelae following external beam radiation for prostate cancer: a review with an update in patients with

stages T1 and T2 tumor. *Journal of Urology* 1994; **152**: 1799–805.

Section 40 Cardiac surgery

40.1 Development of surgery of the heart and great vessels

Figure 1 reproduced from O'Malley CD, Saunders JB de CM. *Leonardo da Vinci on the human body*. New York: Schuman, 1952: 217.

40.2 Cardiopulmonary bypass and myocardial protection

Table 4 reproduced from Hearse DJ, *et al. Protection of the ischemic myocardium: cardioplegia*. New York: Raven Press, 1981.

40.4 Vascular and cardiac biological and non-biological prostheses

Figure 6 reproduced from Hammond GL, Laks H., Geha AS. Development of aortic valve prostheses. *Connecticut Medicine*, 1980; **44**: 348–52.

Figure 7, **Figure 8**, and **Figure 9** reproduced from Hammond GL, Geha AS, Kopf GS. Biological versus mechanical valves. *Journal of Thoracic and Cardiovascular Surgery*, 1987; **93**: 182–98.

40.7.5 Abnormalities of the atrioventricular valves

Figure 1, **Figure 2** and **Figure 3** reproduced from Becker AE, Anderson RH. *Cardiac pathology*. Edinburgh: Churchill Livingstone.

40.7.6 Stenosis and atresia of the arterial valves

Figure 1 and **Figure 2** redrawn from Edmunds LH, Norwood WI, Low DL. *Atlas of cardiothoracic surgery*. Philadelphia: Lea & Febiger, 1990.

40.7.7 Transposition of the great arteries

Figure 1 redrawn from Laks H, *et al.* The influence of coronary anatomy on the arterial switch operation in neonates. *Journal of Thoracic and Cardiovascular Surgery*, 1992; **104**: 708.

Figure 7 and **Figure 8** redrawn from de Leval MR. *Surgery for congenital heart defects*.

Figure 9 and **Figure 10** redrawn from Bune, *et al.*, eds. *Glenn's thoracic and cardiovascular surgery*. 5th edn. Norwalk CT: Appleton & Lange, 1991.

40.7.8 Congenital anomalies of the thoracic aorta and main pulmonary arteries

Table 2 and **Table 4** reproduced from Graham TP, Jr. The Eisenmenger reaction and its management. In: Roberts WC, ed. *Congenital heart disease in adults*. Philadelphia: FA Davis Co., 1979.

Table 5 modified from Oram S, ed. *Clinical heart disease*. London: William Heinemann Medical Books, Ltd., 1971, and Fowler NO, ed. *Cardiac diagnosis and treatment*. Hagerstown: Harper & Row, 1975.

40.8.1 Coronary artery disease: risk factors, diagnosis, and medical treatment

Figure 1 reproduced from Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: risk factor changes and mortality results. *Journal of the American Medical Association*, 1982; **248**:1465–77, and the National Heart, Lung, and Blood Institute. The cholesterol facts. A summary of the evidence relating dietary fats, serum cholesterol, and coronary heart disease. *Circulation*, 1990; **81**:1721–33. Copyright 1990 American Heart Association.

Figure 4 reproduced from Weiner DA, *et al.* Prognostic importance of a clinical profile and exercise test in medically treated patients with coronary disease. *Journal of the American College of Cardiology*, 1984; **3**: 772–9.

40.8.2 Interventional cardiology

Figure 1 reproduced from Heart and Stroke Facts: 1998 Statistical Supplement, 1997. Copyright American Heart Association.

Table 4 reproduced from Herrmann HC, Feldman T. Preface: percutaneous transvenous mitral commissurotomy with the Inoue balloon. *Catheterization and Cardiovascular Diagnosis*, 1994; Supplement 2: 1.

40.8.3 Coronary artery bypass

Figure 3, **Figure 4** and **Figure 5** redrawn from Yusuf S, Zucker D, Peduzzi P, *et al.* Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists' Collaboration. *Lancet* 1994; **344**: 563–70.

Figure 14 redrawn from Kupersmith J, Holmes-Rovner M, Hogan A, Rovner D, Gardiner J. Cost-effectiveness analysis in heart disease, Part III: ischemia, congestive heart failure, and arrhythmias. *Progress in Cardiovascular Disease* 1995; **37**: 307–46.

40.9.2 Surgery for aortic and pulmonary valve disease

Figure 1 and **Figure 2** reproduced from Harlan JH, Starr A, Harwin FM. *Manual of cardiac surgery*. New York: Springer Verlag 1995: 188.

Figure 3 reproduced from Rapaport E. Natural history of aortic and mitral valve disease. *American Journal of Cardiology*, 1975; **35**: 221.

Figure 4, **Figure 5**, **Figure 6**, **Figure 7**, **Figure 8**, **Figure 9**, **Figure 10** and **Figure 11** redrawn from Massimiano PS, Hammond GL. Aortic valve disease and hypertrophic myopathies. In: Bune, *et al.*, eds. *Glenn's thoracic and cardiovascular surgery*. 5th edn. Norwalk CT: Appleton & Lange, 1991; **2**:1702–5.

Figure 12 reproduced from Elkins RC, *et al.* Pulmonary autograft replacement of aortic valve: an evolution of technique. *Journal of Cardiac Surgery* 1992; **7**: 108–15.

Figure 13 and **Figure 14** reproduced from Cohn LH, Adams DH, Couper GS, Bichell DP. Minimally invasive aortic valve replacement. *Seminars in Thoracic and Cardiovascular Surgery* 1997; **9**: 331–6.

Figure 15 redrawn from Misbach GA, Turley K, Ebert PA. Pulmonary valve replacement for regurgitation after repair of tetralogy of Fallot. *Annals of Thoracic Surgery*, 1983; **36**: 686.

Figure 16 redrawn from Laks H, Marin-Garcia J, Willman VL. *In situ* placement of a valved conduit at the pulmonary annulus. *Journal of Thoracic and Cardiovascular Surgery*, 1977; **74**: 642.

Figure 17 redrawn from Karl TR, *et al.* Surgical treatment of absent pulmonary valve syndrome. *Journal of Thoracic and Cardiovascular Surgery*, 1986; **91**: 593.

40.9.3 Diagnosis and surgical management for tricuspid and mitral valve disease

Figure 1 and Figure 5 redrawn from Bharati S, *et al.*, eds. *Cardiac surgery and the conduction system*. New York: John Wiley & Sons, 1983: 45.

Figure 2, Figure 3, Figure 8, and Figure 9 redrawn from Frankl WS, *et al.*, eds. *Valvular heart disease—comprehensive evaluation and management*. Philadelphia: FA Davis Co., 1986: 382.

Figure 4 redrawn from Cohn LH, *et al.*, eds. *Modern techniques in cardiothoracic surgery*. New York: Futura Publishing Co., 1979: 22.

Figure 6 and Figure 7 redrawn from Cosgrove DM, *et al.* Mitral valvuloplasty. *Current Problems in Cardiology*, 1989; **14**: 377.

Figure 10 redrawn from Cooley DA, *et al.*, eds. *Techniques in cardiac surgery*. Philadelphia: WB Saunders Co., 1984: 209.

Figure 11 redrawn from Cosgrave DM. Partial sternotomy for mitral valve operations. *Operative Techniques in Cardiac and Thoracic Surgery*, 1998; **3**: 33.

Figure 13 redrawn from Cohn LH. Parasternal approach for minimally invasive aortic valve surgery. *Operative Techniques in Cardiac and Thoracic Surgery*, 1998; **3**: 56.

Figure 14 redrawn from Pompili. Port access approach for minimally invasive mitral valve surgery. *Operative Techniques in Cardiac and Thoracic Surgery*, 1998; **3**: 63.

40.11 Surgery for cardiac rhythm disturbances

Figure 1 adapted from Cox JL, Gallagher JJ, Cain ME. Experience with 118 consecutive patients undergoing surgery for the Wolff-Parkinson-White syndrome. *Journal of Thoracic and Cardiovascular Surgery*, 1985; **80**: 496.

Figure 2 redrawn from Anderson RH, Becker AE. Cardiac anatomy for the surgeon. In: Danielson GK, ed. *Lewis' practice of surgery*. Hagerstown MD: Harper & Row, 1979: 331.

Figure 3 redrawn from Cox JL, Ferguson TB Jr. Surgery for the Wolff-Parkinson-White syndrome: the endocardial approach. *Seminars in Thoracic and Cardiovascular Surgery*, 1989; **1**: 34–46.

Figure 4 reproduced from Kramer JB, *et al.* Arrhythmia and conduction disturbances: simultaneous computer mapping to facilitate intraoperative localization of accessory pathways in patients with Wolff-Parkinson-White syndrome. *American Journal of Cardiology*, 1985; **56**: 571.

Figure 5, Figure 6 and Figure 7 adapted from Cox JL. The surgical management of cardiac arrhythmias. In: Sabiston DC Jr, Spencer FC, eds. *Gibbons' surgery of the chest*. 5th edn. Philadelphia: WB Saunders Co., 1990.

Figure 8 redrawn from Cox JL, Holman WL, Cain ME. Cryosurgical treatment of atrioventricular node reentry tachycardia. *Circulation*, 1987; **76**: 1329. Copyright; 1987 American Heart Association.

Figures 11 redrawn from Lowe JE, *et al.* Surgical management of chronic ectopic atrial tachycardia. *Seminars in Thoracic and Cardiovascular Surgery*, 1989; **1**: 58–66.

Figure 12 adapted from Williams JM, *et al.* Left atrial isolation: new technique for the treatment of supraventricular arrhythmias. *Journal of Thoracic and Cardiovascular Surgery*, 1980; **80**: 373.

Figure 13 redrawn from Canavan TE, *et al.* Computerized global electrophysiological mapping of the atrium in a patient with multiple supraventricular tachyarrhythmias. *Annals of Thoracic Surgery*, 1988; **45**: 232–5.

Figure 14 reproduced from Ferguson TB Jr. Schuessler RB, Hand DE, Boineau JP, Cox JL. Lessons learned from computerized mapping of the atrium: surgery for atrial fibrillation and flutter. *Journal of Electrocardiology* 1994; **26**: 210–19.

Figure 18 reproduced from Kawaguchi AT, *et al.* Factors affecting rhythm after the maze procedure for atrial fibrillation. *Circulation* 1996; **94** (Suppl. 9): II 139–42.

Figure 20 modified from Cox JL. First decade of ventricular tachycardia surgery. *Circulation*, 1989; **79**: 1160–70. Copyright; 1989 American Heart Association.

Figure 21 and Figure 22 redrawn from Cox JL. Surgical treatment of ischemic and non-ischemic ventricular tachyarrhythmias. In: Cohn LH, ed. *Modern technics of surgery*. Mount Kisco NY: Futura Publishing Co., 1985.

Section 41 Thoracic surgery

41.1 Surgical anatomy and radiology of the chest

Figure 2, Figure 3, Figure 5, Figure 6, Figure 7, and Figure 8 reprinted from *The CIBA Collection of Medical Illustrations*, illustrated by Frank H. Netter, M.D., 1969. All rights reserved. Copyright © 1980 CIBA-Geigy Corporation.

41.2 The trachea

Figure 2 redrawn from Cantrell JR, Guild HG. Congenital stenosis of the trachea. *American Journal of Surgery*, 1964; **108**: 297.

Figure 3 redrawn from Grillo HC, Mathisen DJ. The trachea—tumors, strictures and tracheal collapse. In: Bune *et al.*, eds. *Glenn's thoracic and cardiovascular surgery*. 5th edn. Norwalk CT: Appleton & Lange, 1991: 615–32.

Figure 4 and Figure 6 redrawn from Grillo HC. Surgery of the trachea. *Current problems in surgery*. Chicago: Year Book Medical Publishers, 1970.

Figure 5 redrawn from Grillo HC. Surgical treatment of postintubation tracheal injuries. *Journal of Thoracic and Cardiovascular Surgery*, 1979; **78**: 860.

Figure 7 and Figure 8 redrawn from Grillo HC. Primary reconstruction of airway after resection of subglottic laryngeal and upper tracheal stenosis. *Annals of Thoracic Surgery*, 1982; **33**: 3.

Figure 9 redrawn from Mathisen DJ, *et al.* Management of acquired nonmalignant tracheoesophageal fistula. *Annals of Thoracic Surgery*, 1991; **52**: 759–65.

Table 1 and Table 2 reproduced from Grillo HC, Zannini P, Michelassi F. Complications of tracheal reconstruction. Incidence, treatment, and prevention. *Journal of Thoracic and Cardiovascular Surgery*, 1986; **91**: 322–8.

41.5 Mediastinal tumours

Figure 3 redrawn from Grillo HC, *et al.* Combined approach to 'dumbbell' intrathoracic and intraspinal tumours. *Annals of Thoracic Surgery*, 1983; **36**: 402–7.

41.7 Lung transplantation and volume reduction for pulmonary failure

Table 2 adapted from Naunheim KS, Fergason MK. The current status of lung volume reduction operations for emphysema. *Annals of Thoracic Surgery* 1996; **66**: 601–12.

Table 5 adapted from Trulock EP. Recipient selection. *Chest Surgical Clinics of North America* 1993; **3**: 1–18.

Table 8 and Table 9 adapted from *St Louis International Lung Transplant Registry Report*. Washington University School of Medicine, St. Louis, MO, 1996.

Table 10 and Figure 2 reproduced from Yusen RD, Lefrak SS. Evaluation of patients with emphysema for lung volume reduction surgery. *Seminars in Thoracic and Cardiovascular Surgery*, 1996; **8**: 83–93.

41.9 Chylothorax

Figure 1 redrawn from Meester TR, Lefontaine E. The pleura. In: Sabiston DC, Spencer FC, eds. *Surgery of the chest*. 5th edn. Philadelphia: WB Saunders Co., 1990.

Section 42 Paediatric surgery

42.2 Thoracic surgical problems in infants and children

Figure 2(a-g, j) redrawn from Shamberger RC, Welch KJ. Surgical repair of pectus excavatum. *Journal of Pediatric Surgery*, 1988; **23**: 614–22.

Figure 2(h,i) redrawn from Shamberger RC. Chest wall deformities. In: Shields TW, ed. *General thoracic surgery*, 4th edn. Baltimore, MD: Williams & Wilkins, 1994: 529–57.

Figure 6(a,c) redrawn from Shamberger RC, Welch KJ. Surgical correction of pectus carinatum. *Journal of Pediatric Surgery*, 1987; **22**: 48–53.

Figure 6(b) redrawn from Shamberger RC. Surgical correction of chondromanubrial deformity. *Journal of Pediatric Surgery* 1988; **23**: 319–22.

Figure 33 redrawn from Hendren WH, Hendren WG. Colon interposition for esophagus in children. *Journal of Pediatric Surgery*, 1985; **20**: 829–39.

Figure 35 redrawn from Kim SH, Hendren WH, Donahoe PK. Gastroesophageal reflux and hiatus hernia in children: experience with 70 cases. *Journal of Pediatric Surgery*, 1980; **15**: 443–51.

Figure 38 redrawn from Hendren WH, Kim SH. Pediatric thoracic surgery. In: Scarpelli EM, Auld PAM, Goldman HS, eds. *Pulmonary disease of the fetus, newborn, and child*. Philadelphia: Lea & Febiger, 1978.

42.7.1 Hirschsprung's disease

Figure 4 adapted from Hendren III WH, Kim SH. Abdominal surgical emergencies of the newborn. *Surgical Clinics of North America* 1974; **54**: 499.

Figure 5 adapted from Ricketts RR, Pettitt BJ. Management of Hirschsprung's disease in adolescents. *American Surgeon* 1989; **55**: 223.

42.10 Paediatric solid tumours

Figure 8 reproduced from Green DM. *Diagnosis and management of malignant solid tumors in infants and children*. The Hague: Martinus Nijhoff Publishing, 1985.

Figure 16 reproduced from Altman RP, Randolph JG, Lilly JR. Sacrococcygeal teratoma: American Academy of Pediatrics Surgical Section Survey 1973. *Journal of Pediatric Surgery* 1974; **9**: 389–98.

Section 44 Neurosurgery

44.3 Intracranial tumours

Table 2 reproduced from DeAngelis LM. Management of brain metastases. *Cancer Investigation* 1994; **12**: 157, by courtesy of Marcel Dekker, Inc.

Table 3 reproduced from Fine *et al.* Primary central nervous system lymphoma. *Annals of Internal Medicine* 1993; **119**: 1093–104.

44.8 Hydrocephalus

Figure 1 reproduced from Scott RM, ed. *Concepts in neurosurgery* Vol. 3, *Hydrocephalus*. William and Wilkins, Baltimore, 1990; 3: 13.

Section 45 Head and neck surgery

45.1 Eye emergencies

Figure 4 and Figure 5 redrawn from Shingleton BJ, *et al. Eye trauma*. St Louis: Mosby Year Book Publisher, 1991.

45.2 Nose and sinuses

Figure 1, Figure 2, and Figure 4 modified from Bailey BJ, ed. *Head and neck surgery—otolaryngology*. Philadelphia: JB Lippincott, 1993. Original illustration by Anthony Pazos.

Figure 3 modified from Cummings CW, ed. *Otolaryngology—head and neck surgery*. St Louis: Mosby-Year Book, 1993.

Figure 8 modified from Paparella MM, ed. *Otolaryngology—head and neck surgery*. Philadelphia: WB Saunders, 1991.

45.3 The evaluation of a neck mass in an adult patient

Figure 1, Figure 2, Figure 3, Figure 4, Figure 5 and Figure 6 redrawn from Fried MP. Evaluation of the adult neck mass. *Medical Times*, 1982; January issue.

45.4 Salivary glands

Figure 8 redrawn from Johns ME. *Atlas of head and neck surgery*. St Louis: Mosby 1990: 333–9.

45.5 Oral and oropharyngeal cancer

Figure 1 and Figure 3 reproduced from Medina JE. A rational classification of neck dissections. *Otolaryngologic and Head and Neck Surgery* 1989; **100**: 169–76

Figure 2 reproduced from Medina JE, Byers RM. Supraomohyoid neck dissection: rationale, indications, and surgical technique. *Head and Neck* 1989; **11**: 111–22.

45.6 Nasopharyngeal cancer

Table 1 reproduced from McGuire LM, Lee JCK. The histopathologic diagnosis of nasopharyngeal carcinoma. *Ear, Nose and Throat Journal* 1990; **69**: 229–39.

Table 4 reproduced from Morton RP, *et al.* Transcervico-mandibulo-palatal approach for surgical salvage of recurrent nasopharyngeal cancer. *Head and Neck* 1996; **18**: 352–8.

Figure 3 redrawn from Harnsberger HR. CT and MRI of masses of the deep face. In: Keats TH, ed. *Current problems in diagnostic radiology*, 1987; **XVI**:149.

45.7 Cancer of the larynx and hypopharynx

Figure 1 and Figure 8 reproduced from *Manual for staging of cancer*, 5th edn, 1998: 37–9 and 45–6. The original source for this material is the *AJCC cancer staging manual*, 5th edn. Philadelphia: Lippincott-Raven Publishers, 1997.

45.12 The surgical treatment of snoring and obstructive sleep apnoea

Figure 1, Figure 2, Figure 3, and Figure 6 reproduced from Bleach, Milford, and Van Hassell, eds. *Operative otolaryngology*. Oxford: Blackwell Science Ltd., 1997: 301.

Section 46 Orthopaedics

46.4 Management of bone tumours

Table 2 reproduced from A system for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopaedics*, 1980; **153**: 108.

Section 47 Some special aspects of surgery in developing countries

47.2.2 Schistosomiasis (bilharziasis)

Figure 3 redrawn from *Infectious diseases*. London: Mosby, 1999.

Section 48 Cancer

48.8.1 Principles of radiation oncology

Figure 2 after Puck TT, Marcus PI. *Journal of Experimental Medicine*, 1956; **105**: 653–66.

Figure 4 reproduced from Sinclair WK. *Radiation Research*, 1964; **21**: 584–611.

Section 49 Pain relief

44.2 Cancer pain relief

Figure 1, Figure 2 and Table 3 reproduced from Twycross RG, Lack SA. *Alimentary symptoms in far-advanced cancer*. Edinburgh: Churchill Livingstone, 1986: 304–5.

Figure 3 reproduced from *Cancer pain relief*. Geneva: World Health Organization, 1986 (Annex 1, Figure 1. The analgesic ladder for cancer pain management), p. 51.

Figure 4 reproduced from Twycross RG, Lack SA. *Symptom control in far-advanced cancer: pain relief*. London: Pitman Books, 1983:**110**, 272.

Section 50 Plastic surgery

50.1 Principles and practice of plastic surgery

Figure 53 adapted from Goldman MP, Fitzpatrick RE. *Cutaneous laser surgery*. St Louis: Mosby, 1994:20.

50.2 Craniofacial reconstruction for congenital,traumatic, and neoplastic conditions

Figure 2 redrawn from Poole M, *et al.* Craniofacial surgery. *British Medical Journal*, 1985; **290**: 693, and Davies D, ed. *ABC of plastic and reconstructive surgery*, 1985.

Section 51 How should surgical trainees be selected?

Figure 1 and Figure 2 reproduced from Rutkow IM. *Surgery: an illustrated history*. Mosby-Year Book, Inc. in collaboration with Norman Publishing, 1993.

Table 1 and Table 2 modified from John KD, Modlin IM. A brief historical perspective and a comparison of the current systems of surgical training in Great Britain, Germany and the United States of America. *Surgery, Gynecology and Obstetrics*, 1993; **177**: 628, 629.

1 Response to injury

Douglas W. Wilmore

Introduction
Response components
- **Hypermetabolism**
- **Altered thermoregulation**
- **Protein redistribution and accelerated nitrogen loss**
- **Increased gluconeogenesis**
- **Accelerated lipolysis**
- **Hyperdynamic circulatory state with alterations in regional blood flow**

Response manifestations: host determinants and treatment effects
Wound signals: how do we know when we're injured?
- **Afferent nervous stimuli**
- **Circulating factors**

The role of the central nervous system
The role of hormones and inflammatory factors mediating injury responses
Modifying post-traumatic responses
- **Feeding the patient**
- **Warming the patient; minimizing thermal stress**
- **Using minimally invasive procedures**
- **Providing selective nutrients**
- **Administering anabolic agents**
- **Epidural anesthesia and other anesthetic techniques**

Further reading

Introduction

Injury produces a fairly predictable and reproducible set of responses that alter the clinical course and influence clinical care; these local and systemic changes in circulation, metabolism, and immunological factors apparently have evolved to enhance recovery. John Hunter, the famous eighteenth century surgeon, was the first to comment on this unique situation: ‘There are circumstances attending accidental injury which does not belong to disease. The injury done has in all cases a tendency to produce both a disposition and a means of cure.' Thus, it appears that this innate genetic program serves to enhance recovery of the injured tissue, support the inflammatory response, and minimize associated infection and other complications.

Injury in the context of a surgical service refers to patients who have sustained tissue damage that is the consequence of an operation, is accidental, or due to a disease. The injury usually requires some type of surgical repair. Injuries can often be graded by severity: for example, a clean surgical incision is a minor injury for it results predominantly in injured cells along the wound with minimal inflammation; crush injury or flame burns are major injuries that result in more extensive tissue damage associated with an exaggerated inflammatory response. These examples demonstrate that injury responses are related to the extent of tissue damage and the subsequent inflammatory response; a traumatic injury resulting in open wounds will clearly elicit greater responses than those observed in the uncomplicated patient undergoing an elective operation. In fact, with present-day anesthesia, minimally invasive surgical procedures, and modern techniques for postoperative pain control, injury responses after most elective procedures are minimal. Intervention by the surgeon is rarely necessary and can even be meddlesome. In more seriously ill patients cared for in a burn or trauma unit or in the surgical intensive care unit, injury responses predominate. Care is directed toward supporting the overworked heart and lungs, and attenuating catabolic responses. The reason for studying and understanding injury responses is so the physician can learn to minimize adverse post-traumatic events while maintaining the recuperative components that facilitate wound healing, tissue repair, and convalescent recovery.

Response components are categorized in Table 1. The clinical and laboratory manifestations of injury responses are common variables that are frequently monitored because they reflect the clinical condition of the patient. Study of the metabolic manifestations requires more specialized approaches and this methodology will be discussed later in more detail. The physiological consequences of injury responses serve as the basis for modern intensive care: support adequate cardiac output; optimize ventilation, often using a mechanical ventilator; monitor fluid balance and organ function; and provide adequate nutrition.

Clinical manifestations
Fever
Tachycardia
Tachypnea
Presence of a wound or inflammation
Anorexia
Laboratory changes
Leucocytosis/leucopenia
Hyperglycemia
Elevated C-reactive protein/altered acute-phase reactants
Hepatic/renal dysfunction
Metabolic manifestations
Hypermetabolism
Accelerated gluconeogenesis
Enhanced protein breakdown
Increased fat oxidation
Physiologic consequences
Increased cardiac output
Increased ventilation (e.g. increased gas exchange)
Increased membrane transport
Weight loss/if prolonged will contribute to prolonged convalescence
Wound healing

Table 1 Response components

All of these events occur while the wound is healing, and this endpoint (wound closure) is usually achieved unless the patient has associated diseases that interfere with wound vascularization or inflammation, or they are malnourished, or there is an intervening complication. Thus, the purpose of the injury response is to mobilize stored fuel (adipose tissue) and other tissue components (primarily skeletal muscle-derived amino acids) and transport this substrate to the wound to support the healing process, to immunologically active tissue to support cellular function and replication, and to other vital organs to optimize their function in supporting recuperation (Fig. 1). In short, we break down our own tissue to ensure tissue repair. With mild to moderate injury this catabolic response causes minimal debility; with larger, more extensive injury, such use of body tissue, especially protein, may prolong convalescence and even contribute to increased mortality.

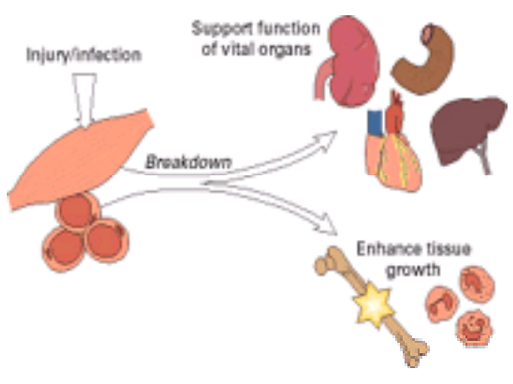


Fig. 1. An overview of the metabolic response to injury. The breakdown of adipose tissue and protein-containing tissue (primarily skeletal muscle) provides fuel and building blocks that support vital organs and enhance proliferation of immunologic tissue and wound repair.

Response components

Hypermetabolism

At rest and in the fasting steady state, energy derived from the oxidation of body fuels is almost completely used for heat production and, when balanced with heat loss, this process maintains a near-constant body temperature. Rates of energy metabolism can be measured and derived from the determination of oxygen consumption and carbon dioxide production, a technique that quantifies respiratory gas exchange and is referred to as indirect calorimetry. Gas exchange is usually measured using a canopy hood placed over the patient's head to capture all expired air. If the patient is being supported by a mechanical ventilator, devices are now available for the continuous monitoring of oxygen utilization and carbon dioxide production. These values are then converted into energy equivalents.

Oxygen consumption changes little after elective operations, but with long-bone fractures, peritonitis or burn injury, the metabolic rate increases (Table 2). The hypermetabolism varies with time after the injury: it is normal or below basal in the shock phase of injury, then rises with resuscitation and peaks at 5 to 10 days after injury, when inflammation is at its height. As the wound heals, the metabolic rate returns to normal.

Mild malnutrition	↓ 0–15%[a]
Postoperative recovery following elective operation	↑ 0–5%
Cancer patients	↓ 10% ↑ 20%[b]
Patients with peritonitis	↑ 5–25%[b]
Patients with severe infection or multiple trauma	↑ 30–55%[b]
Thermal injury (40–100 per cent of total body surface burn)	↑ 40–80%[b]

[a] ↓ decrease, ↑ increase, [b] the rise is proportional to the extent of disease.

Table 2 The effect of various surgical conditions on metabolic rate

The etiologic background of this response is multifaceted. Thyroid hormone concentrations are normal, but catecholamine concentrations are elevated. The metabolic rate will fall toward normal with effective a- and b-blockade. In addition, when normal individuals are infused with catecholamines, their metabolic rate rises. The increased sympathetic output from the central nervous system can be blocked by some general anesthetics, especially high-dose narcotics, and this blockade normalizes the metabolic rate.

When the quantity of oxygen utilized by various organs and regional beds is determined, it has been found that oxygen consumption is elevated in all regions of the body. However, most of the regions (an extremity or the splanchnic bed, for example) of seriously injured patients consume about the same percentage of total body oxygen as in normal individuals; this suggests that the increased energy expenditure that follows injury is a generalized response, involving both visceral and peripheral tissues. The hypermetabolism is associated with heightened cardiorespiratory activity, increased splanchnic metabolic function, inefficient substrate cycling, cellular proliferation, wound repair, and a variety of other synthetic and transport processes.

Altered thermoregulation

Concomitant with the development of hypermetabolism in the trauma patient is the development of fever, a 1 to 2°C elevation in core temperature. This response is thought to be unrelated to systemic infection, but is a result of the inflammatory response and the generation of cytokines. Cytokines [particularly interleukin (**IL**)-1, IL-6, tumor necrosis factor (**TNF**), and interferon] affect the central nervous system and result in an upward resetting of the central reference temperature. With this change, two mechanisms exist in the body to increase core temperature: vasoconstriction decreases heat loss and hypermetabolism increases heat production. Peripheral vasoconstriction is the usual response observed in a normal individual in a warm or thermally neutral environment. Behavioral responses also minimize heat loss, and patients will add blankets or change posture to minimize their surface exposure to a cool environment. Such adjustments are frequently unavailable to critically ill patients, who are often exposed and paralyzed or heavily sedated in air-conditioned intensive care units. Thus, the major physiologic response available to a seriously ill patient for maintaining the new elevated core temperature is to increase their metabolic rate: patients respond by becoming hypermetabolic.

The hypermetabolic response is not fever dependent (it does not occur solely to generate a fever). However, caring for patients with large wounds or extensive burn injuries in a warm environment can reduce their thermoregulatory drive and lessen their hypermetabolism. Small infants and frail elderly patients will also benefit by being cared for in warmer ambient temperatures (27–29°C).

Patients undergoing operations lasting 2 h or more often succumb to intraoperative hypothermia (Fig. 2); this results in stimulation of the sympathetic nervous system and activation of the pituitary–adrenal axis. By placing these patients on a warming device kept on the operating table in the cool operating room, the stress related to environmental cooling can be greatly reduced and outcome improved. One group has reported that warming patients in this manner is associated with a significant decrease in morbid cardiac events, including arrhythmias and myocardial infarctions. Others have noticed that keeping the patient warm in the operating room significantly reduces the incidence of wound infections.

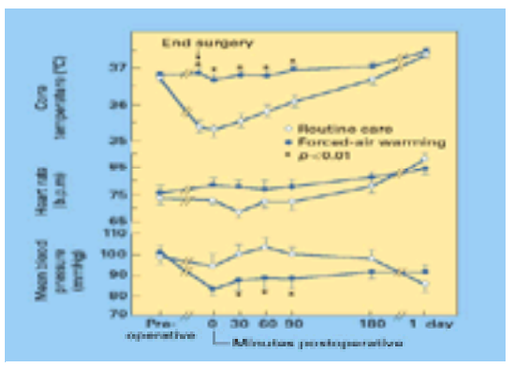


Fig. 2. The effect of routine care and forced air warming on core temperature, heart rate, and blood pressure throughout a major operative procedure.

The advantages of post-traumatic fever are generally unknown but it will enhance vasodilation, improve blood flow to surface wounds, and enhance wound healing. Studies in ectoderms show that survival after bacteremia is positively correlated with body temperature. In contrast, the absence of a febrile response in humans is usually associated with a poor outcome. Because of this, a fever of up to 38.5°C is generally not treated if there are no signs of infection and the patient is asymptomatic, and if the increased core temperature does not overburden a compromised cardiovascular system. Fever above this is usually evaluated because of infection; it is generally treated with antipyretics and antibiotics after diagnosis.

Protein redistribution and accelerated nitrogen loss

Normal body composition is usually described in terms of a four-compartment model: adipose tissue, minerals (primarily bone), water, and protein (Fig. 3). Protein represents the active, functioning tissue of the body, contributing both to structure (for example, skeletal muscle) and biochemical activity (enzymes). Protein occurs in a hydrated form, and for every gram of protein within the body there is an associated 3 g of water.

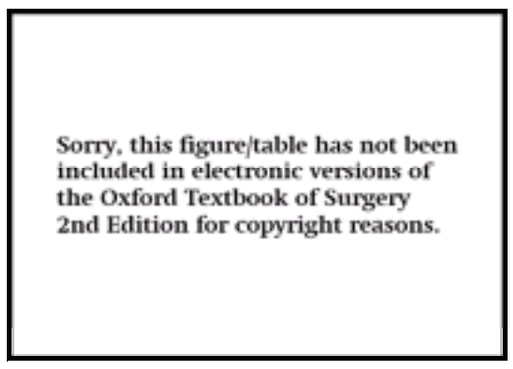


Fig. 3. The four-compartment model of body composition, with water further divided into the inter- and extracellular compartments. The mineral compartment is in large part bone. Lean body mass represents the body mass devoid of fat and minerals. Body cell mass is lean mass without the extracellular fluid.

The protein content of the body is also divided between skeletal muscle and viscera, and this regionalization is important when considering the changes that occur with fasting, sepsis or injury, for there is an active transfer of amino nitrogen from skeletal muscle to visceral tissues. Skeletal muscle mass represents 30 to 50 per cent of total body protein, is greater in men than women, and declines with age. Between the age of 20 and 80 years, the total muscle cross-sectional area declines by about 40 per cent; this, of course, affects strength and endurance with aging, although some of these effects may be reversed with exercise. However, in the context of injury responses, the quantity of a patient's muscle mass may determine their long-term ability to withstand a catabolic disease. This is one of the reasons why elderly patients, with their reduced skeletal muscle mass, are so vulnerable during prolonged periods of muscle breakdown.

Unlike carbohydrate and fat, fuels that are stored in the body (e.g. as glycogen and adipose tissue), protein is not a stored fuel. Protein loss from the body represents loss of a structural or functional component of a living organism. While in the short term, protein loss may be of minor significance, long-term wasting will result in delayed wound healing, immunosuppression, and prolonged convalescence.

After an operation or an injury the patient increases their urinary excretion of nitrogen. The nitrogen is primary in the form of urea, which represents about 85 per cent of the urinary nitrogen loss, although this proportion may vary widely. Creatinine, ammonia, uric acid, and amino acids are also found in the urine in greater quantities than normal. The nitrogen molecule is used as a surrogate marker of protein because the relation between the two substances is relatively fixed; that is:

$$\text{Nitrogen (g)} = \text{protein (g)}/6.25.$$

Net protein catabolism is determined by two standard techniques. Over the short term, balance studies determine body protein loss or gain; this approach measures all the nitrogen consumed (via food, enteral tube feedings, parenteral feedings, blood products, etc.) and all the nitrogen lost in drainage, urine, and stool, which are collected and measured. These losses are then analyzed for total nitrogen and the balance determined:

$$\text{Nitrogen balance} = \text{nitrogen in} - \text{nitrogen out}$$

(a plus sign means body protein gain; a minus sign indicates body protein loss).

Although nitrogen balance is commonly dependent on protein intake, in the postinjury state nitrogen balance is usually negative even when feeding is provided to a patient. In severely ill patients, this loss of body protein often continues for weeks. A comparison of the nitrogen loss that occurs after some of the post-traumatic and catabolic states evaluated in the past is shown in Table 3, and a typical balance study is presented in Fig. 4.

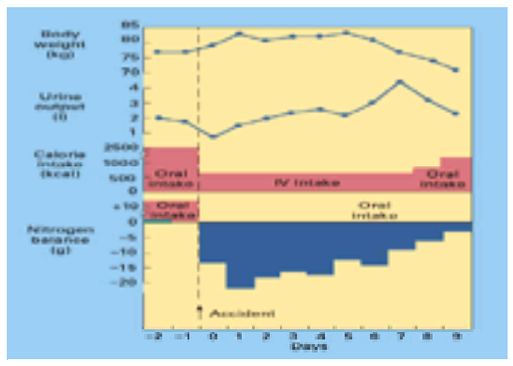


Fig. 4. The injury response in an individual following a moderate accidental injury requiring an abdominal operation. Nitrogen intake is plotted above the zero line of the nitrogen balance chart. Negative nitrogen balance is plotted by the blank bars, and moves toward balance with time and food intake.

Precipitating factor	*Cumulative nitrogen loss (g)*
Injury	
Major burn	170
Multiple injury	150
Peritonitis	136
Simple fracture	115
Major operation	50
Minor operation	24
Infection	
Typhoid fever (untreated)	116
Pneumonia (untreated)	99
Tuberculosis (treated)	57
Q fever (treated)	40
Sandfly fever (untreated)	16

Table 3 Estimates of nitrogen loss following catabolic illness (first 10 days, *ad lib* feedings)

The second method of analyzing body protein loss is to perform serial studies of body composition. Newer techniques utilizing whole-body counters allow quantitation of body nitrogen. Others have measured total body potassium and used this element as a marker of body protein; this is possible because the total potassium is generally related to the amount of protein within the body, and this proportion is relatively fixed. A third approach is to use a scanning device that utilizes dual X-ray absorptiometry. This approach measures several components of the body including ‘lean body mass' a term that describes both body protein and water together. Because water retention occurs in the post-traumatic state, the direct measurement of protein mass is not possible using this technique. However, if an investigator also measures total body water, this quantity can be subtracted from the measured lean body mass and the size of the protein compartment approximated.

Compositional techniques are often used to study patients who have a prolonged and persistent catabolic state. While repeated measurements may be made and a trend established over time, the usual protocol measures the individual at the beginning and end of a particular period of time, usually 10 days to 2 weeks or more. By subtracting the mass of a particular compartment obtained before the experiment from that obtained at the end, the change in body protein (or fat) can be determined.

Compositional measurements on critically ill patients are performed in only a few laboratories throughout the world. The equipment is expensive, is rarely located adjacent to intensive care, and not designed to hold a seriously ill patient, who often requires frequent monitoring and is surrounded by supportive equipment. However,

some serial compositional studies have been performed and one example will be presented later.

Net loss or gain of body protein is a general measure of the catabolic state of the patient. Yet the maintenance of protein within an individual tissue is a balance between protein synthesis and breakdown, with both processes occurring simultaneously. Synthesis and breakdown can become mismatched, resulting in protein loss. This occurs by one of two mechanisms: synthesis falls but breakdown continues in a normal manner, or synthesis remains the same and breakdown is increased (Fig. 5).

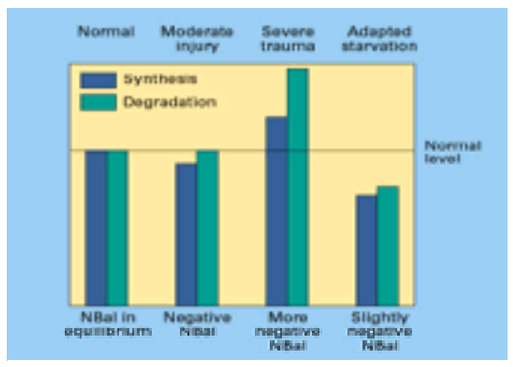


Fig. 5. Whole-body rates of protein synthesis and breakdown are compared in normal individuals, individuals with two degrees of trauma, and patients who have fasted. Note when breakdown outstrips synthesis, negative nitrogen balance occurs, even though rates are elevated as occurs in the trauma patients. Conversely, adaptation to starvation causes depression of both synthesis and breakdown.

Whole-body protein synthesis is reduced after elective surgery but protein breakdown remains the same. In contrast, after severe trauma, patients have enhanced breakdown of protein and whole-body synthesis is normal or only slightly elevated.

As individual tissues lose or gain protein, there is uptake or release of amino acids into various regional vascular beds. By studying the arterial and venous concentrations of amino acid in vessels supplying and draining particular organs or regions, and determining the blood flow to that organ, the rate of uptake or release of amino acid nitrogen can be determined:

$$\text{Net flux (mm/min)} = [\text{arterial concentration (mm/ml)} - \text{venous concentration (mm/ml)}] \times \text{blood flow (ml/min)}$$

(a minus sign denotes release; a plus sign uptake).

A variety of substrate flux studies have been performed in normal and critically ill patients. These investigations have demonstrated that there is a coordinated movement of amino acids between organs. During the acute phase of injury, amino acids are released from the noninjured extremities of injured patients. Since skeletal muscle is the major protein-containing tissue in the extremity, this release represents the accelerated muscle proteolysis that is a characteristic catabolic response to injury. Other studies have shown that amino acids are avidly extracted from the bloodstream of the splanchnic bed; this is, in large part, an extraction of amino acids by the liver for the synthesis of structural, plasma, and acute-phase proteins. In addition, ureagenesis is accelerated and the urea is eventually excreted in the urine, thereby accounting for the increased post-traumatic nitrogen excretion (Fig. 6).

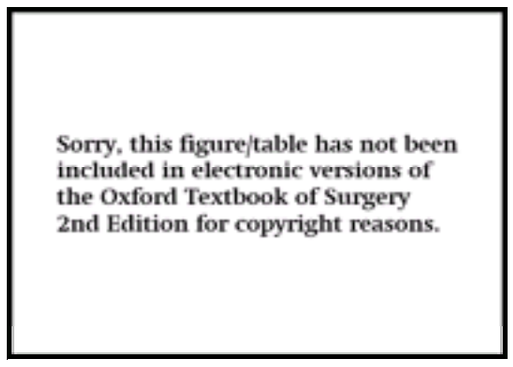


Fig. 6. Major pathways for the metabolism of amino acids following injury: both alanine (ALA) and glutamine (GLN) serve as major gluconeogenic precursors, but GLN also is metabolized by the gut and kidney.

These observations of interorgan flux are consistent with other studies of the rates of tissue protein synthesis following isotropic infusions into stressed animals. Taken together, these studies confirm that protein turnover responds in a manner that redistributes body protein to satisfy the body's needs. Thus, the synthetic rate is decreased in ‘non-essential' tissues (such as limb skeletal muscle or the gut), while the synthetic capacity is maintained or enhanced in tissues in which work or synthetic capacity are increased (respiratory and cardiac muscle, lung, liver, spleen). Most of these observations have been made in animals, and only a few studies of this type have been performed in humans. However, all the data (regional flux studies, compositional studies, and tissue data from animals) fit the general thesis that serious injury stimulates enhanced protein turnover, resulting in a translocation of protein from skeletal muscle to the viscera (primarily liver, spleen and heart) that are vital for survival. Additional nitrogen is processed as ‘waste' and converted to urea for excretion.

The composition of the amino acids released from skeletal muscle is unique. The mixture does not mimic the amino acid pattern found in skeletal muscle protein but rather reflects a rather specialized pattern. Two amino acids predominate, alanine and glutamine, and these two account for approximately 50 to 75 per cent of the amino acid nitrogen transported out of skeletal muscle (Table 4). Alanine is an important glucose precursor and thereby indirectly provides this fuel source, which is essential for several key tissues; glutamine too is a gluconeogenic substrate but also serves as a primary substrate for immune cells and enterocytes, participates in acid–base homeostasis, and serves as a precursor for glutathione (an important intracellular antioxidant). Generation of glutamine results from upregulation of the glutamine synthetase enzyme, which is present in all tissues but is of physiologic importance in the lungs in the short term and skeletal muscle over the long term. Glucocorticoids activate this enzyme and thus the stress response results in the rapid generation of glutamine in these selective tissues.

	Percentage of total
	(mean, range)
Essential amino acids	23 (19–28)
Branched chain amino acids	9 (7–12)
Non-essential amino acids	76 (65–80)
Glutamine	26 (8–34)
Alanine	23 (14–32)

Table 4 The composition of the amino acid nitrogen released from an extremity (representing primarily skeletal muscle) during a critical illness

It has been hypothesized that the tissue requirements for glutamine may outstrip the ability of a given tissue (particularly skeletal muscle) to produce it and hence a relative deficiency state exists. This state is characterized by a fall in glutamine concentrations in both the plasma and tissue compartments. The provision of exogenous glutamine in these stress situations when inflammation persists corrects the plasma and tissue stores, improves nitrogen balance, and often reduces morbidity and mortality. Thus, glutamine is considered a conditionally essential amino acid, the conditions being the presence of inflammation and/or injury that increases the demand for glutamine.

The liver performs many functions following injury but one unique role is to alter dramatically the plasma concentration of certain circulating proteins called ‘acute-phase' reactants. These proteins may be present in the bloodstream under normal conditions but their concentrations change rapidly, owing to rapid increases or decreases in their hepatic synthesis. The signals for these changes include the elaboration of glucocorticoids, IL-6, other cytokines, and possibly nervous signals arising from the brain.

Some acute-phase proteins rise after injury (C-reactive protein, g-antiglobulin, fibrinogen) and others fall (transferrin, albumin); a general time course for these responses is shown in Fig. 7. While it is thought that these substances contribute to overall host defense, their specific role in post-traumatic responses is still debated. However, C-reactive protein serves as a very specific marker of the extent of inflammation and as such can generally be used to quantitate the degree of stress or inflammation that occurs in the patient.

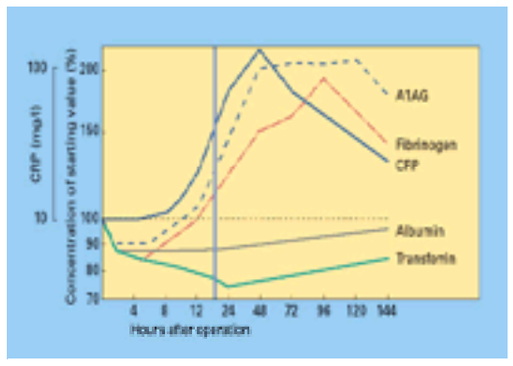


Fig. 7. Alterations in plasma concentrations of acute-phase proteins following an operation; the *y*-axis showing percentage change utilizes a logarithmic scale, except for C-reactive protein, which is in actual concentrations (mg/l).

What are the clinical consequences of the increased net protein catabolism that follows injury and infection? In the short term, a well-nourished individual will sustain minimal debility from the negative nitrogen balance that accompanies an elective operative procedure. In the long term this response may be more deleterious. For example, Hill and associates studied a group of seriously ill patients with peritonitis. Body compositional studies revealed that the patients lost a total of 1.5 kg of body protein over the first 3 weeks of their illness, despite ‘adequate' nutritional support (Fig. 8). Most of this deficit occurred in skeletal muscle, which would eventually be manifest in decreased respiratory muscle function, decreased strength and activity, and prolonged convalescence. Such a response is of major clinical importance in older patients, who have already lost skeletal muscle mass with aging and have little reserve to sustain function after a catabolic event. Research is needed to evaluate a variety of treatments that can be used to attenuate or reverse the protein catabolic response to injury.

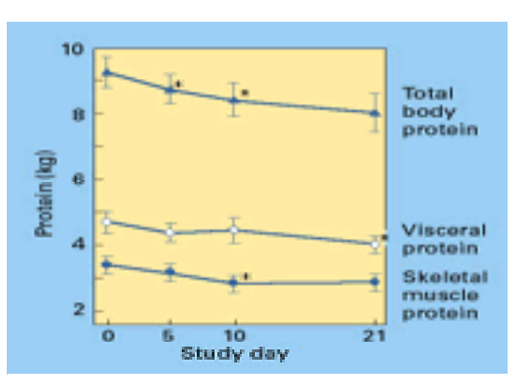


Fig. 8. Changes in total body protein and visceral and skeletal muscle protein over 21 days in patients with peritonitis receiving ‘adequate nutrition'; note the significant fall in body protein that occurs with time[* denotes differences ($p < 0.05$) from initial values].

Increased gluconeogenesis

Following injury, glucose concentrations are usually elevated; with time, these may be near normal during fasting, but increase with the infusion of isotonic, glucose-containing solutions or parenteral feedings. However, if glucose tolerance studies are performed they may be normal or ‘diabetic-like'. This response was initially referred to as the ‘diabetes of injury', but subsequent studies demonstrated that insulin was present but a state of insulin resistance had arisen.

Glucose is produced primarily by the liver, with this 6-C compound arising from stored glycogen and from 3-C precursors. In fasting normal individuals, the 200 g of glucose released each day from the liver is principally utilized by obligate glucose-consuming tissue, such as the brain, red cells, and renal medulla. A little glucose is consumed by resting skeletal muscle and some is stored as adipose tissue.

The classic technique for measuring the rate of glucose production is to determine the arterial–venous glucose difference across the liver (e.g. splanchnic bed) and also hepatic blood flow. This approach is limited in that it measures only ‘net' glucose production by the liver (that is, the total glucose produced minus the glucose utilized by hepatic tissues) and also does not account for the glucose produced by the kidneys (although the renal contribution is usually quite small).

Another technique for measuring glucose production is to infuse stable isotopes of glucose and determine their steady-state concentrations; glucose production is then calculated. This technique measures whole-body glucose production (from all sites) but relies on assumptions about the size of the glucose compartment. However, in general, the two techniques compare favorably, with the catheterization technique yielding results slightly lower than the isotopic. Because the tracer technique does not require catheterization of hepatic vessels, it is now used more commonly then the arterial–venous difference.

Both approaches have been used to describe the response in glucose metabolism following injury. Using arterial-venous catheterization techniques, Wilmore and colleagues examined substrate flux across the splanchnic bed in noninfected burn patients and burn patients with bacteremia; rates of glucose production were elevated approximately 60 to 100 per cent above normal. Thus, if a normal postabsorptive individual produced approximately 200 g glucose/day, a noninfected patient with a 50 per cent surface burn would release approximately 320 g/day. In patients who became infected, glucose production rates rose another 35 to 40 per cent above that observed with injury alone. Catheterization studies of the splanchnic bed also allowed measurement of substrates that could potentially be utilized for gluconeogenesis. The splanchnic uptake of lactate and pyruvate was two to three times greater than that observed in normal individuals. Moreover, splanchnic amino acid extraction was increased: assuming complete conversion of gluconeogenic amino acids to glucose, these precursors accounted for as much as 20 to 35 per cent of the new glucose produced. With bacteremia, the contribution of lactate, pyruvate, and amino acids to new glucose was further increased (Fig. 9).

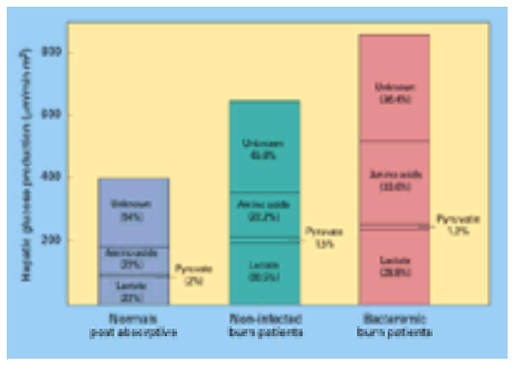


Fig. 9. Hepatic glucose production in normal individuals, burn patients, and burn patients with bacteremia; an estimate of the per cent contribution of gluconeogenic precursors is also shown.

The suppressability of the hepatic glucose produced was evaluated in a subset of injured patients. In normal individuals, the administration of exogenous glucose at a rate of 4 mg/kg per minute almost totally suppresses hepatic glucose production; arterial and hepatic venous concentrations become similar and thus the arterial–venous difference approaches zero. Infusions of glucose at these rates in injured patients only suppressed glucose production by about 50 per cent.

Similar studies were performed in trauma patients by Wolf *et al.* using stable isotopic techniques. They found that glucose production rates were increased about twofold following flame burn and accidental injury (Fig. 10). Infusing exogenous glucose completely suppressed glucose production in normal volunteers, but failed to do so completely in the injured patients. Using similar isotopic techniques, it was found that glucose–lactate–glucose recycling could account for about one-third of the glucose produced.

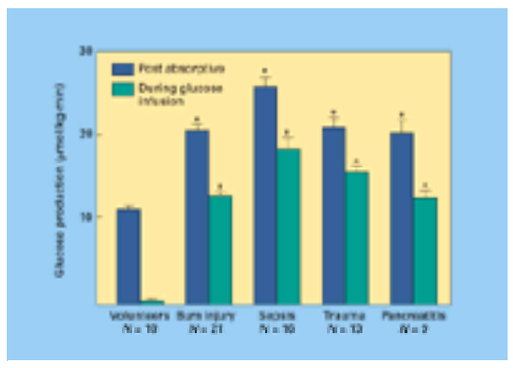


Fig. 10. Glucose production measured in normal individuals and patients with catabolic surgical diseases: all patient groups have increased production rates in the basal state with glucose infusion, glucose production is almost totally suppressed in normal individuals, but is only slightly reduced in the catabolic patients.

Where is the increased glucose being utilized? Using determinations of arterial–venous difference and blood-flow measurements in burn patients, it was shown that brain glucose uptake was near normal but renal uptake slightly increased (25 per cent). When glucose uptake was studied in uninjured or slightly injured limbs of injured patients, minimal uptake was observed. In contrast, glucose extraction increased markedly when similar measurements were performed on injured extremities (Table 5). Approximately 80 per cent of the glucose consumed was converted to lactate, presumably by the inflammatory cells and fibroblasts that comprised the wound. Thus, the processing of glucose to lactate by the wound accounted for the increased metabolism of glucose in the peripheral tissue, and completed the description of the cycling of glucose (e.g. glucose–lactate–glucose) that occurs in injured humans.

	Small leg burn	Large leg burn
Number of participants	8	7
Age (years)	28	27
Weight (kg)	73.2	76.1
Total body surface burn (%)	40.5	42.0
Total leg burn (%)	9.5	50.0
Leg blood flow (ml/100ml per min)	4.220 ± 0.43	8.020 ± 0.51*
Leg oxygen consumption (ml/100ml per min)	0.187 ± 0.037	0.240 ± 0.010
Leg glucose uptake (mg/100ml per min)	0.037 ± 0.010	0.134 ± 0.077*
Leg lactate production (mg/100ml per min)	0.0930 ± 0.095	0.299 ± 0.073*

* $p < 0.05$

Table 5 Comparison of glucose uptake in the legs of patients with small leg injuries vs extensive surface injuries

What accounts for the diminished clearance of glucose in injured patients? While insulin concentrations may be low in the early stages of injury, they rise to normal or supernormal in nondiabetic trauma patients following adequate resuscitation. Using the euglycemic insulin-clamp technique, Black and associates studied normal individuals and trauma patients to examine the kinetics of glucose disposal. At plasma insulin of less than 100 µU/ml, responses in both groups were similar. However, when the insulin doses infused were increased, maximal disposal rates in the patients reached only two-thirds of those observed in the normal controls [9.17 mg/kg per minute vs 14.3(± 0.78) mg/kg per minute]. A similar pattern was observed when exogenous glucose was infused and glucose was cleared by stimulating endogenous insulin release. The investigators concluded that insulin resistance occurred in peripheral tissue, probably in skeletal muscle, and that this was consistent with a postreceptor deficit. These conclusions were subsequently confirmed by studies that examined glucose uptake in the forearm of volunteers receiving various doses of insulin infusion.

The etiology of the resistance to insulin is unknown. However, Bessey and associates performed insulin-clamp studies and examined forearm glucose uptake (primarily representing skeletal muscle) in normal individuals with and without short-term infusions of epinephrine (adrenaline). Glucose uptake was reduced from 0.66(± 0.8) to 0.18(± 0.13) mg/100 ml per forearm tissue/min with epinephrine infusion. In a subsequent study, similar resistance was observed after 3 days of glucocorticoid administration. Thus, it appears that the hormonal environment plays a major part in the insulin resistance observed in peripheral tissues. However, another aspect of post-traumatic metabolism may also contribute to insulin resistance. Several decades ago, Randel proposed that elevated fatty acids inhibited the entry of glucose into skeletal muscle, and conversely the elevation of glucose impaired the entry of fatty acids. That there is competition between substrates for entry is generally true, but this hypothesis remains controversial and its clinical relevance is unknown. However, rapid mobilization of fatty acids from adipose tissue after injury provides more than enough substrate for skeletal muscle, and this may have some regulatory role in enhancing insulin resistance.

What is the clinical relevance of these findings of increased gluconeogenesis and insulin resistance in injured patients?

1. Injured patients are prone to develop hyperglycemia. Their blood glucose should be monitored and every effort should be made to maintain it below 200 mg/dl. This is important, for increased infection rates have been positively related to blood glucose concentrations.
2. Exogenous insulin is often required in trauma patients receiving glucose-based feedings. Because of the insulin resistance that is present, and the increased clearance of insulin from the bloodstream that has been observed, more insulin is required in the post-traumatic state than in most other conditions (excluding patients with diabetes mellitus).
3. Glucose infused at low doses is oxidized; but as the quantity infused increases, much of the exogenous glucose is stored as body fat. This process places

increased energy demands on the injured patient and may contribute to increased carbon dioxide production and dependency on mechanical ventilation. In general, the rate of glucose infusion should not exceed 4 mg/kg per minute, providing no more than 400 g glucose/day.

Accelerated lipolysis

Despite the increased gluconeogenesis and hyperinsulinemia in the postresuscitation phase of injury, in the fasting state the injured patient oxidizes fat as a primary fuel source. This process is best observed by measuring the respiratory quotient, which is the ratio of carbon dioxide produced to oxygen consumed. When the respiratory quotient approaches 1.0, carbohydrate is being oxidized; at 0.7, fat is the primary fuel source. In the normal individual, the respiratory quotient is about 0.83, suggesting a mixed fuel source; in the patient receiving excess glucose (more than 4 mg/kg per min) the respiratory quotient exceeds 1.

When the respiratory quotient is determined in fasted, injured patients it approaches 0.7, indicating fat oxidation. Because these patients are also hypermetabolic, this indicates that not only is fat the primarily oxidized fuel source but also that fatty acids are being oxidized at a greater rate than in normal controls. This general concept was confirmed by isotopic turnover studies that demonstrate an increased release of fatty acids of 50 to 100 per cent above normal in patients following accidental injury or flame burn. This increased release appeared to be under the influence of adrenergic control, for b-adrenergic blockade acutely reduced it to the normal range. In contrast to the turnover rates, fatty acid concentrations are usually normal or slightly elevated in the trauma patient.

When stored fat (triglycerides) undergo hydrolysis, both fatty acids and glycerol are released. Because re-esterification can occur with the fatty acids but not the glycerol, the disparity in the rate of release between these two compounds is a reflection of the rate of lipolysis within the adipocyte. In injured patients, glycerol turnover is elevated two to three times above that of normal controls, and this rate is not suppressed by glucose infusion, again demonstrating insulin resistance. In addition, about 70 per cent of the released fatty acids are not oxidized but are re-esterified; and this process occurs primarily in the liver. Thus, increased cycling of fatty acid triglyceride exists in the injured patient, a process that utilizes excess energy and may also contribute to the known increase in hepatic fat deposition.

Because of the increased availability of fatty acids after injury, this substrate becomes the primary oxidizable fuel in the injured patient. On supplying a mixed fuel source in the form of enteral or parenteral nutrition, the body gradually shifts to oxidize the exogenous fuel, particularly if glucose is administered and insulin is stimulated. As the rate of exogenous energy supplied approaches the metabolic requirement, the respiratory quotient approaches 1.0, indicating that primarily carbohydrate is being oxidized. Because of the increased insulin insensitivity and the heightened sympathetic drive, the fatty acid cycle remains operative even in this fed state. As fatty acids and triglycerides are being synthesized from the excess carbohydrate, they are stored, adding to adiposity and supporting hepatic fat deposition. Liver dysfunction may follow. Understanding this response emphasizes the danger that can occur with overfeeding in the injured. It should also be emphasized that the energy component of the feeding should contain both carbohydrate and fat calories.

Starvation in normal individuals results in the mobilization of fatty acids and the formation of b-hydroxybutyrate and acetoacetate, or ‘ketone bodies'. These water-soluble forms of fat are then utilized by the brain, muscle, and kidney as a fuel source, thus sparing glucose and proteolysis. Following injury, ketosis is greatly limited even during fasting; this occurs despite the abundance of free fatty acids to be utilized as a precursor substrate. The decreased ketosis after injury appears to result from alterations in the hormonal environment, particularly the increase in insulin.

Hyperdynamic circulatory state with alterations in regional blood flow

Injury is associated with loss of blood and edema formation, events that decrease the effective circulating blood volume and thus cause cardiac output to fall. With resuscitation and operative maneuvers that stabilize the circulating blood volume, cardiac output gradually rises. With time (1–5 days), cardiac output is increased above normal, and the central arterial–venous oxygen difference is narrowed. In part, this demand for additional blood flow is in response to the increased oxygen demands. In addition, there are other thermoregulatory demands and the blood flow to specific regional beds is enhanced. Finally, as the area of inflammation or injury becomes vascularized and granulation tissue forms, blood flow to these regions is greatly enhanced.

During this hyperdynamic phase of injury, cardiac output is related to the size or extent of injury; in addition, it tends to be increased more than the rise in oxygen consumption. Much of the additional blood flow is directed to the periphery or the site of injury; this was best demonstrated in an elegant series of studies by Aulick and associates, in which blood flow was measured using whole-leg plethysmography in thermally injured patients. Blood flow was measured in nine normal individuals and 26 thermally injured patients. It was essentially normal in the uninjured legs (3 ml/100 ml leg volume per minute) but increased in a curvilinear manner and approached a plateau of 8.0 ml in patients with burns that exceeded 60 per cent of the leg surface (Fig. 11). Studies of muscle blood flow utilizing xenon wash-out techniques demonstrated that muscle blood flow in the uninjured limb or in the muscle underlying the wound was essentially normal. These studies confirmed the interpretation of the whole-body studies that most of the increased peripheral blood flow following thermal injury is directed to the surface wound. This flow, however, is both a result of dilation of existing vessels and of the perfusion of new vessels secondarily to the formation of a neovasculature within the wound. The flow does not appear to be the result of increased metabolic demands, for the arterial–venous difference across the wound remains narrow because the metabolism of wound tissue is primarily anaerobic (see Table 5).

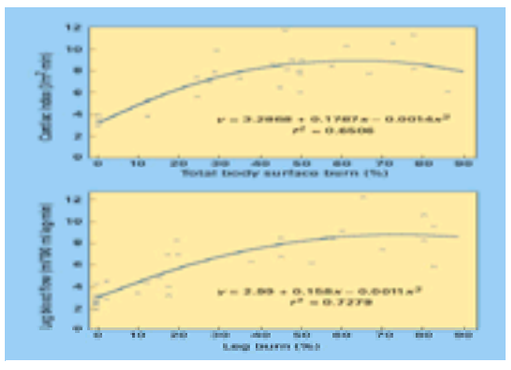


Fig. 11. Blood flow to the whole body (cardiac index) is related to the size of the surface wound, in a manner similar to the relation between leg blood flow and the size of the leg injury.

Studies of the perfusion of other vascular beds show that both renal and splanchnic blood flow are increased and that coronary arterial blood flow is elevated in proportion to cardiac output. A generally partitioning of flow in a normal individual and a burn patient is shown in Fig. 12.

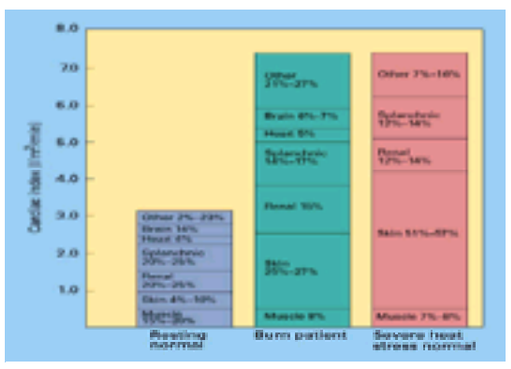


Fig. 12. In a resting normal individual, the cardiac index supplies oxygen requirements for visceral organs and skeletal muscle under thermally neutral conditions. In the burn patient, these demands increase and, with the increased perfusion of the burn wound, the cardiac index rises. This shift of total body circulation to the peripheral tissues from the central core is similar to that observed in a normal individual under conditions of severe heat stress.

The accelerated cardiac output satisfies both the demands of the healing wound and the metabolic and thermoregulatory demands of the host. Maintenance of a full

blood volume and support of the hyperdynamic state is essential for optimal wound healing and a successful outcome.

Response manifestations: host determinants and treatment effects

A variety of factors determine the eventual manifestations of injury responses (Table 6). Age, sex, and nutritional status all have an important and modifying role. In general, responses are dampened in infants and toddlers and are maximal in teenagers and young adults. While injury responses to submaximal stress may be quite similar between young and old individuals, responses generally diminish with age in patients who experience a more severe stress, such as those with major injuries and burns. Maximally stressed patients who are over 70 years of age have much greater limitations in physiological responses than those in their 20s; this is due in large part to changes in body composition with age (such as loss of muscle mass) and to limitations in function (such as a decrease in maximum cardiac output). The decline in muscle strength generally parallels the erosion in muscle bulk, although other factors may also contribute to the functional impairment. Even in good health, an 80-year-old woman is at or near maximal quadriceps strength for rising from a low, armless seat.

Extent of injury/inflammation
Time course
Patient-related effects:
Age, sex, disease, nutrition status
Treatment effects:
Operations, steroids, drugs

Table 6 Factors that modify injury responses

Many of these effects on muscle strength can be reversed by submaximal exercise. The carbohydrate intolerance observed in the elderly person is closely related to the overall level of exercise and general fitness. A weekly exercise program will maintain maximum oxygen consumption and cardiac output when compared to those of age-matched controls who are sedentary. Preoperative training of this sort is possible in patients scheduled for elective surgery, but is generally not possible in the injured patient or the one with postoperative complications.

Sex also accounts for differences in body composition and this effect may also alter injury responses. Because lean body mass and muscle mass are smaller in women than in age-matched men, the resting metabolic rate is lower and nitrogen excretion is less in women than men following an elective operation or accidental injury. However, following severe stress, women may have a survival advantage. In a retrospective compilation of data from four trials involving sepsis, Bone found that the preponderance of morbidity and mortality was in males not females. Some of these effects may be due to the deleterious influence that testosterone may have on the immune system, which will be discussed later.

Finally, under- or malnutrition has marked effects on injury responses. Changes in intermediate metabolism occur, with weight loss and protein wasting. Nitrogen loss is greatly attenuated in malnourished patients following colectomy compared to that in well-nourished individuals. While mild weight loss (less than 10–12 per cent of the usual body weight) may not alter responses significantly or effect eventual outcome, weight loss of more than 15 per cent has been associated with a greater incidence of postoperative complications; preoperative feeding, if appropriate, is indicated.

Underlying disease will also modulate injury responses. The decline in cardiac and respiratory function with age limits the hyperdynamic circulatory response that occurs after injury. Changes in respiratory compliance and respiratory muscle strength with age account for the increased need for assisted mechanical ventilation in older patients. Because insulin plays an important part in effecting anabolic responses, comorbid conditions such as diabetes mellitus greatly accelerate protein catabolism and alter glucose tolerance. The resulting hyperglycemia often limits the quantity of carbohydrate calories that can be provided to the seriously ill diabetic patient. Patients with diseases associated with wasting, such as cancer, human immunodeficiency virus infection, autoimmune disease requiring glucocorticoid administration, and neuromuscular disorders, all have muscle loss that will greatly lessen their protein catabolic responses to disease. Because of impaired strength, mechanical ventilation is often necessary in these patients, and rehabilitation is arduous and prolonged.

Responses are often modified by multiple drug effects, operations, and complications, particularly those of an infectious nature. Drugs such as dopamine and glucocorticoids are frequently used in critically ill patients and greatly affect intermediate metabolism. Van de Berge and associates have been at the forefront in demonstrating that in prolonged critical illness there is suppression of the pulsatile elaboration of the anterior pituitary hormones: thyroid-stimulating hormone, growth hormone, and prolactin. These changes are even more severe with the administration of the inotropic agent dopamine. This group hypothesizes that such changes in anterior pituitary function contribute to the severe wasting and catabolism in serious illnesses.

Operations are commonly required after accidental injury and these may be performed as rapidly as the patient is stabilized. Immediate operation after accidental injury causes a second stressful stimulus and this response has been studied experimentally. For example, the sensitivity of the adrenal cortex to adrenocorticotrophic hormone (**ACTH**) is significantly greater after a second procedure performed one day after the first. Likewise, repeated hemorrhage elicits an augmented catecholamine response.

Early pinning of fractures, drainage of intra-abdominal abscesses, or excision of burn wounds also changes the ‘natural' time course related to the insult and greatly shortens convalescent recovery. Similarly, complications such as infection and multiorgan system failure often add to the initial stress and cause physiologic and metabolic responses to approach their limits.

Wound signals: how do we know when we're injured?

Since the early descriptions of injury responses in the 1930s, interest has been directed toward identifying those factors that mediate the metabolic and physiologic changes. Information on the exact mediators of the response is as yet imprecise but our knowledge has increased over recent years. The response mediators can be divided into two general categories: afferent nervous signals and circulating signals. These signaling systems affect a variety of organ systems, but their major impact appears to be on the central nervous system, which then integrates the information. Interaction between nervous and circulating signals has only recently been recognized.

Afferent nervous stimuli

The most rapid communication between an injured area and the brain is via the afferent nervous system. With injury we feel pain, which sets into motion the complex array of injury responses. Hume and Egdahl injured the distal extremity of an animal before and after denervating its hind limb. When the peripheral nervous system was intact there was a brisk pituitary–adrenal response; after nerve transection the response did not occur. Nociceptive pain signals are transmitted primarily by small myelinated (Ad) and small unmyelinated (C) sensory afferent fibers to the dorsal horn of the spinal cord. Subsequently, these impulses are transferred cephalad to the ventral–posterior nucleus of the thalamus. It has been suggested that fast-conducting fibers may also be involved in injury-response signals. During abdominal operations, transmission via the spinal cord and sympathetic pathways seems to dominate over signals transmitted via the vagus.

Other factors generated at the site of injury may reduce the threshold for nociceptive nerve stimulation. Local production of, for example, substance P, vasoactive peptides, and arachidonic metabolites facilitates afferent neural impulses from the damaged peripheral nerves and contributes to postinjury hyperalgesia.

This information has been applied to clinical practice: spinal and epidural anesthesia is useful in interrupting this afferent pathway; when operations are performed on the lower extremities or in the lower abdomen, spinal blockade will greatly attenuate the stress response (that is, reduce stimulation of the pituitary–adrenal axis) and possibly reduce postoperative convalescence.

Circulating factors

Studies in patients undergoing operations on denervated extremities demonstrate that many responses still remain; this suggests that circulating factors coming from the wound or site of injury initiate these responses, which include changes in the plasma concentrations of the acute-phase proteins, fever, granulocytosis, and stimulation of the coagulation cascade. The responses appear to be initiated and propagated by a variety of factors arising from the wound, including complement products, arachidonic acid metabolites, platelet-activity factor, and cytokines. These substances may arise as a result of the local injury or occur because of the presence of ischemic tissue. The signals may serve in an apocrine/pericrine manner to facilitate local inflammation and repair. As the severity of injury increases, these substances may diffuse into the bloodstream to affect other tissues and organs such as the brain, bone marrow, and liver.

Cytokines have attracted much attention in recent years and these proinflammatory factors (IL-1, TNF, IL-6, IL-8, and interferon-a) may effect some of the responses. These mediators primarily act locally; they affect immunologic cells and stimulate T-cell proliferation and hemopoiesis. However, if they reach the bloodstream, systemic responses are observed. Infusion of IL-1 or TNF elicits fever, alterations in glucose metabolism, negative nitrogen balance, and hepatic synthesis of acute-phase proteins; IL-6 may also elicit some of these events. The actions of these cytokines may be due to their direct effects on a specific tissue or, in the case of TNF, the effects may be secondary through the stimulation of the brain and the elaboration of the counter-regulatory hormones, glucagon, cortisol, and catcholamines.

Thus it is thought the cytokines and other wound factors serve as signals that arise from the area of inflammation directing the body's local and systemic responses to facilitate tissue repair. However, we do not know which specific wound factors are the most important in initiating and propagating the injury responses. In the bloodstream, IL-1 and TNF are often elevated acutely but their concentrations fall rapidly after the initial traumatic event. TNF may appear intermittently in the bloodstream later in the course of a catabolic patient, but this is usually associated with the onset of an infectious complication. In contrast to IL-1 and TNF, IL-6 in the bloodstream is elevated over a much longer period of time, and its concentration is generally related to the extent of injury. However, experiments in ‘knock out' mice who lack the *IL-6* gene demonstrate no major alteration in the stress response in the absence of this cytokine.

Thus, it appears that the redundancies in inflammatory mediators allow the wound to signal the host to mount injury responses. Additional research is needed to identify specific mediators and determine their exact role in initiating these metabolic events.

The role of the central nervous system

The brain is responsible for integrating both the nervous and hormonal signals it receives. The primary site for these responses is the hypothalamus; removing the brain above this level in experimental animals does not effect the pituitary–adrenal response to stress. However, the brainstem is necessary for trauma responses: administering large doses of opiates that greatly dampen hypothalamic function to injured patients significantly reduces oxygen consumption, heart rate, core temperature, and minute ventilation.

After its integration of afferent signals, the hypothalamus employs two major effector arms to regulate physiologic responses, the sympathoadrenal axis and the hypothalamic–pituitary–adrenal axis. The sympathoadrenal axis provides a mechanism for rapid response in the cardiovascular, respiratory, and metabolic systems via the sympathetic nervous system. Signals from the sympathetic area of the hypothalamus are transmitted via the intermediolateral column of the spinal cord to the sympathetic efferent nerves. Preganglionic splanchnic nerves innervating the adrenal medulla mediate the outpouring of epinephrine (adrenaline) and other catecholamine products into the bloodstream. Postganglionic nerves communicate with other organs, blood vessels, and cells via the release of norepinephrine (noradrenaline) from nerve endings. The effects of catecholamines vary, in large part because of the dual system composed of both a- and b-receptor sites, and the fact that circulating epinephrine predominantly exerts b-effects at low concentrations and a-effects at high concentrations.

Plasma concentrations of catecholamines and their urinary excretion products tend to increase after injury and infection. During periods of hemodynamic stability, typically referred to as the flow phase of injury, the extent of catecholamine elaboration tends to correlate with the severity of injury and the metabolic rate of the patient.

In the second effector system, the hypothalamic–pituitary–adrenal axis, regulatory peptides arise from the median eminence of the hypothalamus and are transported through a portal system of blood vessels to the anterior lobe of the pituitary gland. These releasing factors result in the elaboration of ACTH, growth hormone, thyrotropin, luteinizing hormone, and prolactin into the bloodstream. Following critical illness, there is a prompt elaboration of ACTH but, in general, suppression of the release of the other pituitary hormone. This suppression is amplified by the administration of a commonly infused catecholamine, dopamine. It is unknown if deleterious effects occur because of this pharmacological interaction.

The hormones of the anterior pituitary gland, and antidiuretic hormone elaborated from the posterior pituitary, circulate throughout the body and exert effects on various tissues including other endocrine glands. In addition, other factors such as the opiate peptides and somatostatin are elaborated, and these factors and the releasing hormones all exert systemic effects on physiological functions.

One area of central nervous regulation only recently recognized is the part that cytokines may play in influencing brain function during inflammatory states. Astrocytes and microglia have been shown to express several cytokines and cytokine receptors, and to increase cytokine release following a variety of stimuli such as the administration of lipopolysaccharide or exposure to a hypoxic environment. Acute injections of IL-1 or IL-6 into the lateral ventricle of the brain or into specific brain areas cause transitory pyrexia, anorexia and adrenal cortical activation, similar to responses observed after injury. In a chronic infusion model, intracerebroventricular administration of IL-1, but not of equimolar amounts of IL-6, caused anorexia, negative nitrogen balance, and weight loss (Fig. 13). These responses were not attenuated by ‘clamping' the glucocorticoids at low concentrations by adrenalectomy and replacement with corticosterone pellets.

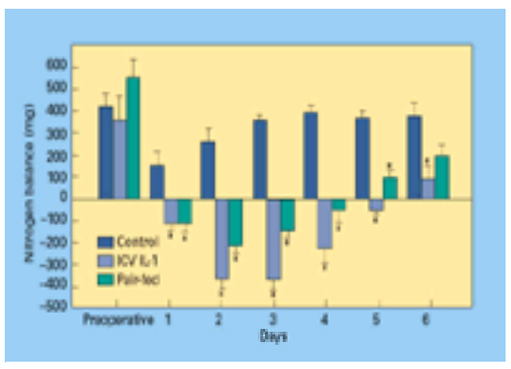


Fig. 13. Nitrogen balance in groups of rats receiving intracerebroventricular infusion: controls received buffered saline; the experimental group received buffered saline but was pair-fed to the IL-1 group (*$p < 0.05$ vs pair fed) (adapted from unpublished data provided by AG Hill, L Jacobson, J Gonzales, J Rounds, JA Majzours, and DW Wilmore).

In other studies, IL-1 was injected subcutaneously into experimental animals and the central nervous effects blocked by intraventricular infusion of an IL-1 receptor antagonist; the central blockade significantly attenuated the protein catabolism that occurs with IL-1, but did not affect the leucocytosis (Fig. 14). In similar studies, lipopolysaccharide was given as the provocative stimulus and increased glucose production was observed; when an IL-1 receptor antagonist was infused into the central nervous system, attenuation of the increased glucose flux was observed. Thus the brain appears to be a major modifier of the stress response through the action of cytokines.

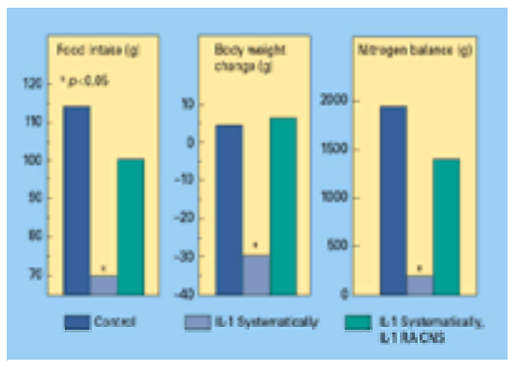


Fig. 14. The effect of peripherally infused IL-1 and its central nervous blockade on chow intake, body weight change, and nitrogen balance in rats. Each animal received two constant infusions, one peripherally and the second into the brain ventricle. The control animals received saline in both infusions. The IL-1 animals received IL-1 systemically and saline centrally. The third group received IL-1 systemically and blocking antibodies to IL-1 intracerebroventricularly. *$p < 0.05$ when compared to the other two groups. Note the attenuation of these stress responses with central IL-1 blockade. (Adapted from unpublished data provided by AG Hill and DW Wilmore.).

Cytokines are large, lipophobic molecules that do not have ready access to the central nervous system, but they may circulate in the bloodstream and be actively transported into specific regions of the brain. Alternatively, cytokines may penetrate specific areas of the blood–brain barrier at one or more of the circumventricular organs. The vascular endothelium may act as a transducer for cytokines that generates signals which stimulate brain pathways, or the cytokines may damage the integrity of the vascular endothelium (or the blood–brain barrier) and stimulate central neural circuits. Finally, cytokines may stimulate peripheral neural pathways, such as the vagus, to send afferent signals to the central nervous system.

There is increasing evidence that vagal pathways are utilized as the communication link between the peritoneal cavity and the central nervous system, especially during episodes of intra-abdominal infection. Many of the central nervous effects induced by intraperitoneal administration of lipopolysaccharide or IL-1, such as fever, increased elaboration of ACTH, and the induction of IL-1 RNA within the brain, can be blocked or attenuated by subdiaphragmatic vagotomy (Fig. 15). Receptors appear to be present within the peritoneal cavity, as increased electrical activity has been observed in the vagus following intraportal injection of IL-1b. In addition, sectioning the subdiaphragmatic vagus blocked the hyperalgesic effects to IL-1b, and IL-1 receptors have been identified on abdominal paraganglia. These findings suggest that cytokines may be somewhat compartmentalized throughout the body and communicate with the brain via the vagus and other afferent nerves.

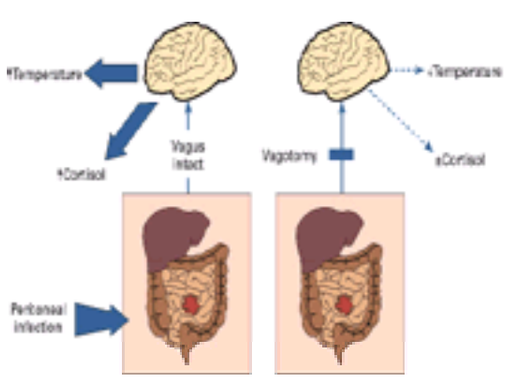


Fig. 15. The role of nervous afferent signaling via the vagus nerve following intraperitoneal inflammation.

In summary, outflow of impulses via the sympathetic nervous system and by way of the pituitary–adrenal axis occurs after injury and other critical illnesses. It appears that cytokines are activated in the brain during these events, although the exact mechanism is unknown. Experimental evidence suggests that infection within the peritoneal cavity may signal the brain via afferent vagal impulses.

The role of hormones and inflammatory factors mediating injury responses

It appears that both nervous and circulating inflammatory factors stimulate the brain to initiate injury responses. Circulating factors may bypass the brain and stimulate responses directly in selective tissues (Table 7). However, a reasonably coordinated set of signals reaches the tissues to initiate catabolic changes and one major component of this signaling process are the counter-regulatory hormones glucagon, catecholamines, and glucocorticoids. As has previously been mentioned, concentrations of these factors are all elevated in the bloodstream following severe stress. These hormones play a major part in the increased gluconeogenesis, insulin resistance, accelerated lipolysis, and proteolysis that occur in other diseases and after injury and infection. This was clearly demonstrated by Bessey and colleagues, who infused these hormones into normal volunteers over a 3-day study period to achieve hormone concentrations similar to those observed in injured patients. On a separate occasion, the volunteers were infused with saline and that 3-day period served as a control study. Food intake was comparable and constant throughout both study periods. During the hormonal infusion, negative nitrogen balance, increased gluconeogenesis, and hypermetabolism were observed, associated with salt and water retention (Table 8). Similar studies, in which the individual hormones were infused, found minimal alterations in metabolism and circulation, suggesting that these three catabolic hormones exerted synergistic and sustained effects when infused together.

Dose of TNF infused (g/m²)	Response	Clinical correlate
1	Hypothermia	Subclinical infection
	Myalgia and headache	
	Anorexia	Influenza
30	Fever	Acute appendicitis
	Tachycardia	
	Elevated acute phase proteins	
	Rigors	
	Elevated stress hormones	Intra-abdominal abscess
> 500	Fluid retention	Major thermal injury
	Lymphopenia	
	Hypotension	
> 630	Decreased consciousness level	Septicemia
	Profound hypotension	Severe acute pancreatitis
	Pulmonary edema	Infected massive burns
	Oliguria	

Low doses may have direct tissue effects, while higher doses produce symptoms predominantly mediated by the central nervous system.

Table 7 Effects of varying doses of tumor necrosis factor (TNF) infused into humans

	Saline	Hormones[a]
Metabolic rate (kcal/m² per h)	32.36 ± 0.05	38.42 ± 1.31
Nitrogen balance (g/day)	-0.20 ± 0.4	-3.30 ± 0.4
Protein catabolic rate (g 70 kg/day)	38.40 ± 1.9	68.50 ± 1.5
Plasma glucose (mg/100 ml)	94.00 ± 2.0	133.00 ± 4.0
Serum insulin (μunit/ml)	8.00 ± 1.0	22.00 ± 3.0
Glucose production rate (mg/kg per min)	2.08 ± 0.06	2.55 ± 0.06
Insulin-mediated glucose disposal (mg/kg per min)	8.62 ± 0.51	3.22 ± 0.51
Insulin-mediated forearm glucose uptake (mg/100 ml per forearm/min)	1.03 ± 0.09	0.48 ± 0.12

[a] All significantly different from control values.

Table 8 Effects of hormonal infusion on metabolism and circulation in humans

Table 8 The effect of infusing cortisol, glucagon, and catecholamine on metabolism

Responses that did not occur during the hormonal infusion were fever, alterations in acute-phase protein concentrations, and leucocytosis. When the inflammatory stimulant etiochoanolone was injected into the volunteers, it caused local inflammation associated with fever, alterations in C-reactive protein, hypoferremia, and leucocytosis. Administering the inflammatory stimulus while infusing the counter-regulatory hormones resulted in a complete manifestation of injury responses, showing that both inflammatory and endocrine mediators are responsible for mediating the responses to critical illness.

In these simulations, nitrogen loss was moderate, and not as brisk as observed after comparable injury; this may have been related to the rise in insulin that occurred with the triple hormone infusion. In a subsequent study, Bessy and Lowe infused the catabolic hormones and simultaneously blocked insulin release by administering somatostatin. A marked negative nitrogen balance occurred, suggesting that the withdrawal of (or resistance to) anabolic hormones is an important component of the protein catabolic response.

Growth hormone is another hormone with anabolic effects that may be attenuated after injury. Its primary anabolic effects occur through the stimulation of a second hormone, insulin-like growth factor-1 (**IGF-1**). Operations and accidental injury result in a pronounced fall in IGF-1. After major surgery, the nadir of IGF-1 occurs at about the same time as the peak muscle breakdown, suggesting a causal relation. In addition, it has been shown that similar patients have decreased expression of the gene for the growth hormone receptor in skeletal muscle. A decrease in receptor numbers could contribute to resistance to growth hormone, which has also been described in critically ill patients.

These data, when taken together, suggest that both the increase in catabolic hormones and the withdrawal of (or resistance to) anabolic hormones account for many of the metabolic alterations observed in the injured patient. These endocrine events may be mediated via the brain and/or effected directly by cytokines and other inflammatory factors. In addition, cytokines have direct effects on tissues and these signals interact within this hormonal environment to further modulate catabolic events (Table 9).

Endocrine effects
Hypermetabolism
Hyperglycemia
Hyperinsulinemia
Insulin resistance
Negative nitrogen balance
Accelerated protein flux
Attenuation of inflammation
Inflammatory effects
Fever
Alterations in acute-phase proteins
Hypoferremia
Stimulation of endocrine responses
(with moderate–high cytokine activity)

Table 9 Mediators of the trauma response

Modifying post-traumatic responses

Moderate to severe catabolism, if unchecked and prolonged, can result in severe debilitation and wasting, and prolonged convalescent recovery. Therefore, a variety of approaches have been utilized to modify these untoward post-traumatic responses.

Feeding the patient

Providing exogenous nutrients, either by the enteral or parenteral routes, has become a standard of care for seriously ill patients over the past decade. Feeding prevents weight loss and attenuates the protein wasting that occurs in injured or infected patients. However, in spite of presumed adequate feedings, loss of body protein still occurs in patients with catabolic diseases (see Fig. 8). While feeding attenuates the loss of body nitrogen (Fig. 16), it has also been associated with the maintenance or gain of body water and total body fat.

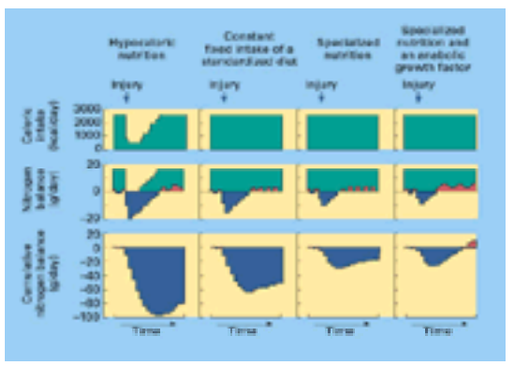


Fig. 16. Post-traumatic negative nitrogen balance can be attenuated by a constant diet, specialized nutrition, or growth factors: stippled bars represent caloric intake in the top panels and nitrogen intake in the middle panels; solid bars represent negative nitrogen balance, and hatched bars a positive nitrogen balance.

Feeding hypercaloric diets to seriously ill patients is not without potential risks, and severe complications have been associated with nutritional support. In patients on enteral feeding, abdominal distension, diarrhea, and aspiration resulting in pneumonia commonly occur. Even then, providing sufficient enteral feeding to meet caloric goals is often difficult or impossible. Patients who receive parenteral nutrition may experience hyperglycemia, catheter sepsis, a variety of electrolyte abnormalities, an increased incidence of infections, and severe liver dysfunction. Thus, the safe use of either of these feeding techniques in a critically ill patient requires experience and patience on the part of the provider.

Several general rules are helpful in achieving safe nutritional support in critically ill patients.

1. Feeding is rarely a priority. Attend to circulatory and cardiorespiratory problems before initiating feedings.
2. Don't overfeed. Complications are usually related to increasing caloric intake. Set limits of 30 to 35 kcal/kg a day and 1.5 g protein/day; in most cases, start slowly to achieve these goals gradually.
3. Use combined enteral–parenteral feeding, thus minimizing the risk of high-calorie feeding by either route, while providing adequate nutrient support with this combined approach.
4. Constantly attempt to shift feeding toward the enteral route. The cost of enteral feeding alone is about one-fourth the cost of parenteral feeding.

Warming the patient; minimizing thermal stress

As previously noted, trauma and burn patients have reset their central reference temperature and utilize physiological and behavioral mechanisms to keep warm. Modern hospitals use environmental engineering techniques to maintain a cool working temperature (usually around 23–5°C) for the staff and visitors, which is undesirably cool for most patients. Moreover, patients who are sedated and mechanically ventilated are unable to utilize many of their physiological mechanisms to

maintain their internal thermal balance. Patients who are particularly vulnerable to cold stress are those who have sustained accidental flame burns or moderate to severe trauma, those who undergo operations lasting more than 2 h, babies, and elderly persons. These patients (and possibly others) should be insulated with bed covers and, if possible, cared for in thermostatically controlled rooms that allow them to be warmed to their comfort level. Unlike the usual hospital environment, patients generally prefer a temperature of 27 to 29°C, depending on their illness.

Special care should also be taken when patients are transported to areas of the hospital that are traditionally cold, such as the operating theater, endoscopy or radiographic rooms. Simple precautions that prevent cooling reduce both morbidity and mortality in inpatients.

Using minimally invasive procedures

With the introduction of laparascopic and thoracoscopic procedures, it has become evident that patients no longer sustain the postoperative discomfort and complications observed when similar procedures were done with open techniques. Several reports have shown no differences in stress hormone concentrations (cortisol, glucagon, catecholamines) between patients undergoing open or laparascopic procedures. However, there does appear to be an attenuation of the inflammatory response following laparascopic procedures, as manifest by reduced amounts of IL-6.

In an effort to determine the mechanism of this effect, West and associates examined the ability of peritoneal macrophages to produce cytokines in room air and while incubated under CO_2, a gas commonly used for intraperitoneal insufflation. When exposed to air and stimulated by lipopolysaccharide, the macrophages produced large quantities of TNF and IL-1. When exposed to CO_2, the stimulated macrophages were essentially unable to produce cytokines (Fig. 17). Comparable studies in an animal model confirmed that these *in vitro* studies were reproducible *in vivo*: CO_2 insufflation greatly attenuated cytokine production by peritoneal macrophages. Thus, the ability to dampen the inflammatory component of the injury response by the use of CO_2 insufflation during laparoscopic surgery may play a major part in modifying postoperative catabolic responses.

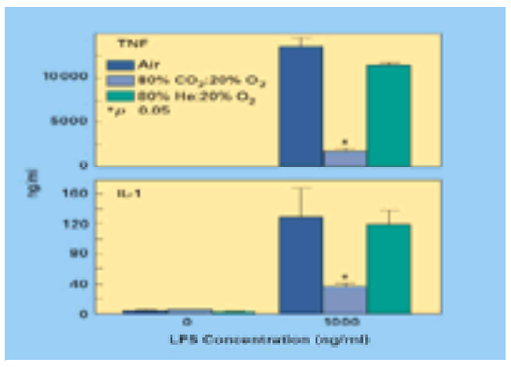


Fig. 17. The production of TNF and IL-1 following stimulation of peritoneal macrophages with lipopolysaccharide while the cells were cultured in various environments; note the attenuation of the cytokine response when the cells were exposed to 80 per cent CO_2/20 per cent O_2.

Further work is needed to evaluate other pharmacologic methods that may attenuate the response of immunologic tissue and facilitate postoperative recovery.

Providing selective nutrients

It is known that a variety of nutrients can modify the inflammatory response. Compounds such as the antioxidant vitamins, the amino acids L-glutamine and L-arginine, w-3 fatty acids, and ribonucleotide have all been evaluated in animal and cell-culture studies, and found to have modifying effects on immunologic responses. Clinical investigators have incorporated these nutrients into feeding formulations administered to patients in the postoperative or post-traumatic state in an attempt to modify catabolic responses and minimize infections.

The most consistent, positive clinical findings have been associated with supplementation of feedings with L-glutamine. This amino acid is normally synthesized by skeletal muscle and released into the bloodstream. It is utilized by inflammatory cells, and the enterocyte and colonocyte, as a primary fuel. Glutamine also participates in acid–base homeostasis and the biosynthesis of the intracellular antioxidant glutathione. Randomized, controlled, double-blind prospective trials have demonstrated that glutamine reduces the infection rate in patients following bone marrow transplantation, in premature infants, in trauma patients, and in athletes after extended exercise. In postoperative patients, glutamine enhanced protein synthesis in skeletal muscle. In a large, randomized study of 84 elderly patients who presented primarily with gastrointestinal sepsis, a significant reduction in 6-month mortality was demonstrated in a glutamine-supplemented group.

An additional approach is to utilize vitamins that have antioxidant potential and theoretically could reduce the oxidant stress associated with a surgical procedure. Rabl *et al.* administered such a preparation to patients undergoing revascularization procedures of the lower extremities and compared the outcome with a placebo group in a controlled trial. Biochemical products of lipid oxidation were greatly attenuated in the vitamin-supplemented group and limb swelling was significantly reduced when compared to that in controls.

It therefore appears that specific and appropriate nutrient supplementation can alter the postinjury course, limit the incidence of infection, and improve outcome in severely catabolic patients.

Administering anabolic agents

A variety of substances specifically target skeletal muscle to enhance net protein synthesis, increase muscle mass, and thus enhance function. Such agents, if administered safely, could be used in the perioperative period to enhance strength or could be given postoperatively to facilitate recovery. Hormones studied to date include testosterone, anabolic steroids, growth hormone, and IGF-1, together with several other drugs and/or hormones that stimulate the endogenous elaboration of growth hormone.

In one clinical study, Byrne *et al.* administered growth hormone (0.14 mg/kg per day) to patients with benign esophageal lesions and compared the changes in body composition to those observed in comparably fed controls. After 3 weeks of receiving a repletion diet (50 kcal/kg and 2 g protein/kg per day), both groups gained body weight (4.65 kg in the growth hormone group and 3.06 kg in the controls). The growth hormone-treated patients gained minimal body fat but had significantly more lean mass (4.3 kg compared to 2.0 kg in the controls) and more protein (1.4 vs 0.1 kg). In addition, the patients receiving standard therapy tended to deposit a greater proportion of body weight as extracellular water and significantly more fat than the growth hormone-treated group. Growth hormone enhanced the efficiency of protein deposition by 66 per cent, thereby accelerating nutritional repletion. Such a response should shorten the period of convalescence required by a malnourished patient who has had a major surgical procedure.

A number of investigators have given growth hormone to postoperative patients. In one such study, Jiang and coworkers administered growth hormone (0.06 mg/kg per day) or placebo to postoperative patients following gastrectomy or colectomy who were receiving hypocaloric parenteral nutrition (20 kcal/kg and 1 g protein/kg per day). The patients receiving growth hormone lost significantly less body weight (1.3 kg as compared to 3.2 kg in controls) and nitrogen (cumulative 8-day balance, -7.1 g in the growth hormone group and -32.6 g in controls). Analysis of body composition revealed that the patients given growth hormone maintained their lean body mass despite the major surgical procedure. The controls lost about 10 per cent in hand-grip strength in the postoperative period, while the patients given growth hormone maintained their grip force throughout that period.

Others have shown that postoperative administration of growth hormone enhances recovery after total hip replacement and major abdominal surgery. Herndon *et al.* administered growth hormone or placebo to burned children in a double-blind, placebo-controlled clinical trial. Protein anabolism was improved and wound healing accelerated. Patients with a 60 per cent total body-surface burn demonstrated a more rapid healing time and could be discharged from the hospital approximately 2 weeks earlier than the controls. This reduction in hospital stay accounted for a significant savings of hospital costs, even after including the cost of the anabolic agent.

A variety of trials are now in progress to evaluate the effect of growth hormone and other anabolic agents on postoperative and postinjury recovery.

Epidural anesthesia and other anesthetic techniques

The use of spinal or epidural anesthesia to block afferent signals to the brain during pelvic and lower-extremity surgery has already been discussed. In addition to dampening the pituitary–adrenal response, nitrogen balance was improved and postoperative morbidity reduced. When both spinal and epidural blockade were combined with laparascopic colectomy, a dramatic reduction in postoperative debility was observed. Kehlet and associates studied eight elderly patients (average age 78 years) using this approach. Bowel function returned within 12 h after the procedure, and the patients were eating and walking within 24 h of their operation. Average time to discharge from hospital after the procedure was 48 h.

In another study on semielective surgical patients, neonates requiring cardiac procedures were randomized to receive high-dose opioid anesthesia or a lighter anesthesia with halothane and morphine. Those receiving the deep opioid anesthesia had an attenuated stress-hormone response, and the incidence of sepsis and metabolic acidosis was greatly reduced when compared with the controls. Mortality was also significantly decreased.

These studies demonstrate that, in selected units, new approaches in anesthesia and pharmacologic management can alter injury responses and enhance and/or accelerate recovery. Such methods have so far only been used in patients undergoing elective and semielective operative procedures, not in those admitted in emergency. Newer approaches also need to be developed and used in patients with accidental injuries and severe infection.

Further reading

Anand KJS, Hickey PR. Halothane-morphine compared with high-dose sufentanil for anesthesia and postoperative analgesia in neonatal cardiac surgery. *New England Journal of Medicine* 1992; **326**: 1–9.

Askanazi J, Elwyn DH, Silverberg PA, Rosenbaum SH, Kinney JM. Respiratory distress secondary to a high carbohydrate load. A case report. *Surgery* 1980; **87**: 596–8. [Describes the major problems associated with glucose loading in sick patients.]

Aulick LH, Wilmore DW. Increased peripheral amino acid release following burn injury. *Surgery* 1979; **85**: 560–5.

Aulick LH, Hander EH, Wilmore DW, Mason AD, Jr, Pruitt BA, Jr. The relative significance of thermal and metabolic demands on burn hypermetabolism. *Journal of Trauma* 1979; **19**: 559–66.

Aulick LH, Wilmore DW, Mason AD, Jr, Pruitt BA, Jr. Influence of the burn wound on peripheral circulation in thermally injured patients. *American Journal of Physiology* 1977; **233**: H520–6.

Aulick LH, Wilmore DW, Mason AD, Jr, Pruitt BA, Jr. Muscle blood flow following thermal injury. *Annals of Surgery* 1978; **188**: 778–82.

Aulick LH, Wilmore DW, Mason AD, Pruitt BA, Jr. Metabolic and thermoregulatory responses to thermal injury. I: Effectors of thermogenesis. *Experientia* 1978 (Suppl. 32; Girardier L, Seycdoux J, ed.): 259–304.

Barbanel G, Ixart G, Szafarczyk A, Malaval F, Assenmacher I. Intrahypothalmic infusion of interleukin I-b increases the release of corticotropin-releasing hormone (CRH) and adrenocorticotropin hormone (ACTH) in free moving rats bearing a push-pull cannula in the median eminence. *Brain Research* 1990; **516**: 31–6.

Bardram L, Funch-Jensen P, Jensen P, Crawford ME, Kehlet H. Recovery after laparascopic colonic surgery with epidural analgesia, and early oral nutrition and mobilization. *Lancet* 1995; **345**: 763–64.

Bessey PQ, Lowe KA. Early hormonal changes affect the catabolic response to trauma. *Annals of Surgery* 1993; **218**: 476–91.

Bessey PQ, Brooks DC, Black PR, Aoki TT, Wilmore DW. Epinephrine acutely mediates skeletal muscle insulin resistance. *Surgery* 1983; **94**: 172–9.

Bessey PQ, Watters JM, Aoki TT, Wilmore DW. Combined hormonal infusions simulate the metabolic response to injury. *Annals of Surgery* 1984; **200**: 264–81. [A classic study in which hormones were infused into normal volunteers and many of the injury responses were observed.]

Birkhahn RH *et al.* Whole-body protein metabolism due to trauma in man as estimated by L-[^{15}N]alanine. *American Journal of Physiology* 1981; **24**(241): E64–71.

Black PR, Brooks DC, Bessey PQ, Wolfe RR, Wilmore DW. Mechanisms of insulin resistance following injury. *Annals of Surgery* 1982; **196**: 420–35.

Bone RC. Toward an epidemiology and natural history of SIRS (systemic inflammatory response syndrome). *Journal of the American Medical Association* 1992; **268**: 3452–5.

Brooks DC, Bessey PQ, Black PR, Aoki TT, Wilmore DW. Post-traumatic insulin resistance in uninjured forearm tissue. *Journal of Surgical Research* 1984; **37**: 100–7.

Buzby GP for the VA TPN Cooperative Study Group. Perioperative total parenteral nutrition in surgical patients. *New England Journal of Medicine* 1991; **325**: 525–32. [A large, multicenter trial to evaluate the effect of perioperative feeding on outcome.]

Byrne TA *et al.* Anabolic therapy with growth hormone accelerates protein gain in surgical patients requiring nutritional rehabilitation. *Annals of Surgery* 1993; **218**: 416–18.

Cahill GF, Jr. Starvation in man. *New England Journal of Medicine* 1970; **282**: 668–75.

Carpentier YA *et al.* Effects of hypercaloric glucose infusion on lipid metabolism in injury and sepsis. *Journal of Trauma* 1979; **19**: 649–54.

Castell LM, Poortmans JR, Nesholme, EA. Does glutamine have a role in reducing infections in athletes? *European Journal of Applied Physiology* 1996; **73**: 488–90.

Christensen T, Stage JG, Galbo H, Christensen NJ, Kehlet H. Fatigue and cardiac and endocrine metabolic response to exercise after abdominal surgery. *Surgery* 1989; **105**: 46–50.

Frank SM *et al.* Perioperative maintenance of normothermia reduces the incidence of morbid cardiac events. *Journal of the American Medical Association* 1997; **227**: 1127–34. [Warming patients in the operating room reduces mortality.]

Goodwin CW, Jr, Aulick LH, Powanda MC, Wilmore DW, Pruitt BA, Jr. Glucose dynamics following severe injury. *European Journal of Surgical Research* 1980; **12**(Suppl. 1): 126–7.

Griffiths RD, Jones C, Palmer TE. Six-month outcome of critically ill patients given glutamine-supplemented parenteral nutrition. *Nutrition* 1997; **13**: 295–302.

Hagberg JM, Allen WK, Seals DR, Hurley BF, Ehsani AA, Holloszy JO. A hemodynamic comparison of young and older endurance athletes during exercise. *Journal of Applied Physiology* 1985; **58**: 2041–6.

Hammarqvist F, Wernerman J, Ali R, von der Decken A, Vinnars E. Addition of glutamine to total parenteral nutrition after elective abdominal surgery spares free glutamine in muscle, counteracts the fall in muscle protein synthesis, and improves nitrogen balance. *Annals of Surgery* 1989; **209**: 455–61.

Hermansson M *et al.* Measurement of human growth hormone receptor messenger ribonucleic acid by a quantitative polymerase chain reaction-based assay: demonstration of reduced expression after elective surgery. *Journal of Clinical Endocrinology and Metabolism* 1997; **82**: 421–28. [Important application of cell biology to a surgical problem.]

Herndon DN, Barrow RE, Kunleel KR, Broemeling L, Rutan RL. Effect of recombinant growth hormone on donor-site healing in severely burned children. *Annals of Surgery* 1990; **212**: 424–9.

Hill AG, Jacobson L, Gonzalez J, Rounds J, Majzoub JA, Wilmore DW. Chronic central nervous system exposure to interleukin-1b causes catabolism in the rat. *American Journal of Physiology* 1996; **271**: R1142–8.

Hill AG, Siegel J, Rounds J, Wilmore DW. Metabolic response to interleukin-1. *Annals of Surgery* 1997; **225**: 246–51.

Hill GL, Monk DN, Plank LD. Measuring body composition in intensive care patients. In: Wilmore DW, Carpentier YA, ed. *Metabolic support of the critically ill patient*, pp. 3–18. Springer, Berlin, 1993. [State-of-the art methodology for body composition analysis.]

Hoffman AF. Defective biliary secretion during total parenteral nutrition: probable mechanisms and possible solutions. *Journal of Pediatric Gastroenterology and Nutrition* 1995; **20**: 376–90.

Houdijk APJ *et al.* Randomized trial of glutamine enriched enteral nutrition on infectious morbidity in patients with multiple trauma. *Lancet* 1998; **352**: 772–6. [An extremely well done feeding trial in trauma patients.]

Hume DM, Egdahl H. The importance of the brain in the endocrine response to injury. *Annals of Surgery* 1959; **150**: 697–712.

Jiang ZM *et al.* Low-dose growth hormone and hypocaloric nutrition attenuate the protein-catabolic response after operation. *Annals of Surgery* 1989; **210**: 513–24.

Johnson IDA. Role of nutrition in perioperative protein sparing. In: Meng HC, Law DL, ed. *Parenteral nutrition*, pp. 26–34. Charles C Thomas, Springfield IL, 1970.

Kapcala LP, He JR, Gao Y, Pieper JO, DeTolla LJ. Subdiaphragmatic vagotomy inhibits intra-abdominal interleukin-1b stimulation of adrenocorticotropin secretion. *Brain Research* 1996; **728**: 247–54.

Kehlet H. Epidural anesthesia in surgical practice. *Current Practice in Surgery* 1990; **2**: 223–6.

Kluger MJ, Ringler DH, Anver MR. Fever and survival. *Science* 1975; **188**: 166–8.

Kurz A, Sessler DI, Lenhardt R. Perioperative normothermia to reduce the incidence of surgical wound infection and shorten hospitalization. *New England Journal of Medicine* 1996; **334**: 1209–15. [Warming patients reduces morbidity.]

Kushner I. The phenomenon of the acute phase response. *Annals of New York Academy of Sciences* 1982; **389**: 39–48.

Lacey JM, Wilmore DW. Is glutamine a conditionally essential amino acid? *Nutrition Reviews* 1990; **48**: 297–309. [The initial hypothesis proposing that glutamine is a conditionally essential amino acid in seriously ill patients.]

Lang CH, Cooney R, Vary TC. Central interleukin-1 partially mediates endotoxin-induced changes in glucose metabolism. *American Journal of Physiology* 1996; **271**: E309–16.

Lexell J, Taylor CC, Sjostrom M. What is the cause of ageing atrophy. Total number, size and proportion of different fiber types studied in whole vastus lateralis muscle from 15 to 83 year-old men. *Journal of Neurological Science* 1988; **84**: 275–94.

Lilly MP, Engeland WC, Gann DS. Adrenal response to repeated hemorrhage: implications for study of trauma. *Journal of Trauma* 1982; **22**: 809–14.

Lukaszewicz G, Abcouwer SF, Labow BI, Souba WW. Glutamine synthetase expression in the lungs of endotoxin-treated and adrenalectomized rats. *American Journal of Physiology* 1997; **273**: L1182–90.

Macfie J, Smith RC, Hill GL. Glucose or fat as a non-protein energy source? A controlled clinical trial in gastroenterological patients receiving intravenous nutrition. *Gastroenterology* 1981; **80**: 103–7.

Max SR *et al.* Dexamethasone regulates glutamine synthetase expression in rat skeletal muscle. *American Journal of Physiology* 1988; **255**: E397–402.

Mealy K, Gallagher H, Barry M, Lennon F, Traynor O, Hyland J. Physiological and metabolic responses to open and laparoscopic cholecystectomy. *British Journal of Surgery* 1992; **79**: 1061–4.

Michie HR *et al.* Tumor necrosis factor and endotoxin induce similar metabolic responses in human beings. *Surgery* 1988; **104**: 280–6.

Muhlbacher F, Kapadia CR, Colpoys MF, Smith RJ, Wilmore DW. Effects of glucocorticoids on glutamine metabolism in skeletal muscle. *American Journal of Physiology* 1984; **247**: E75–83. [The first demonstration that the stress response increases the production of glutamine from skeletal muscle.]

Neu J *et al.* Enteral glutamine supplementation for very low birth weight infants decreases morbidity. *Journal of Pediatrics* 1997; **131**: 691–9.

Niijima A. The afferent discharges from sensors for interleukin 1-b in the hepato-portal system in the rat. *Journal of Physiology* 1992; **446**: 236.

Nordenström J *et al.* Free fatty acid mobilization and oxidation during total parenteral nutrition in trauma and infection. *Annals of Surgery* 1983; **198**: 725–35.

O'Keefe SJ, Sender PM, James WP. Catabolic loss of body nitrogen in response to surgery. *Lancet* 1974; **ii**: 1035–8. [Elective operation decreases whole-body protein synthesis.]

Plank LD, Connolly AB, Hill GL. Sequential changes in the metabolic response in severely septic patients during the first 23 days after the onset of peritonitis. *Annals of Surgery* 1998; **228**: 146–58. [Serial body-composition determinations coupled with cytokine determination in critically ill surgical patients.]

Rabl H, Khoschsorur G, Petek W. Antioxidative vitamin treatment: effect on lipid peroxidation and limb swelling after revascularization operations. *World Journal of Surgery* 1995; **19**: 738–44. [An interesting approach to the use of vitamin therapy.]

Randle PJ, Garland PB, Hales CN, Newsholme EA. The glucose and fatty acid cycle. *Lancet* 1963; **i**: 785–9.

Redmond HP, Watson RW, Houghton T, Condron C, Watson RG, Bouchier-Hayes D. Immune function in patients undergoing open vs. laparoscopic cholecystectomy. *Archives of Surgery* 1994; **129**: 1240–6.

Rogy MA *et al.* Persistently elevated soluble tumor necrosis factor receptor and interleukin-1 receptor antagonist levels in critically ill patients. *Journal of the American College of Surgeons* 1994; **178**: 132–58.

Ross RJM, Chew SL. Acquired growth hormone resistance. *European Journal of Endocrinology* 1995; **132**: 655–60.

Streat SJ, Beddoe AH, Hill GL. Aggressive nutritional support does not prevent protein loss despite fat gain in intensive care patients. *Journal of Trauma* 1987; **27**: 262–6. [Overfeeding makes patients fatter and overhydrated.]

Taylor JW, Hander EW, Skreen R, Wilmore DW. The effect of central nervous system narcosis on the sympathetic response to stress. *Journal of Surgical Research* 1976; **20**: 313–20.

Tsuji H, Shirasaka C, Asoh T, Takeuchi Y. Influences of splanchnic nerve blockade on endocrine-metabolic responses to upper abdominal surgery. *British Journal of Surgery* 1983; **70**: 437–9. [A study that could only be performed by individuals trained both in surgery and anesthesiology.]

Van de Berghe G, de Zegher F. Anterior pituitary function during critical illness and dopamine treatment. *Critical Care Medicine* 1996; **24**: 1580–90. [Problems in the brain with dopamine therapy.]

Van de Berghe G *et al.* Thyrotropin and prolactin release in prolonged critical illness: dynamics of spontaneous secretion and effects of growth hormone-secretagogues. *Clinical Endocrinology* 1997; **47**: 599–612.

Watkins LR *et al.* Blockade of interleukin-1 induced hyperthermia by subdiaphragmatic vagotomy: evidence for vagal mediation of immune-brain communication. *Neuroscience Letters* 1995; **183**: 27–31. [A provocative article which, if true, has interesting applications for surgeons.]

Watters JM, Wilmore DW. Role of catabolic hormones in the hypoketonaemia of injury. *British Journal of Surgery* 1986; **73**: 108–10.

Watters JM, Bessey PQ, Dinarello CA, Wolf SM, Wilmore DW. Both inflammatory and endocrine mediators stimulate host responses to sepsis. *Archives of Surgery* 1986; **121**: 179–90. [A classic study in volunteers to determine the effects of hormones and the effects of inflammation on injury responses.]

West MA, Hackam DJ, Baker J, Rodriquez JL, Bellingham J, Rotstein OD. Mechanism of decreased *in vitro* murine macrophage cytokine release after exposure to carbon dioxide: relevance to laparoscopic surgery. *Annals of Surgery* 1997; **226**: 179–90.

Wilmore DW, Aulick LH, Mason AD, Jr, Pruitt BA, Jr. Influence of the burn wound on local and systemic responses to injury. *Annals of Surgery* 1977; **186**: 444–58. [Blood flow is specifically directed to the area of injury.]

Wilmore DW, Goodwin CW, Aulick LH, Powanda MC, Mason AD, Jr, Pruitt BA, Jr. Effect of injury and infection on visceral metabolism and circulation. *Annals of Surgery* 1980; **192**: 491–504. [Infection increases the catabolic response to injury.]

Wilmore DW, Long JM, Mason AD, Jr, Skreen RW, Pruitt BA, Jr. Catecholamines: mediators of the hypermetabolic response to thermal injury. *Annals of Surgery* 1974; **180**: 653–69.

Wilmore DW, Orcutt TW, Mason AD, Jr, Pruitt BA, Jr. Alterations in hypothalamic function following thermal injury. *Journal of Trauma* 1975; **15**: 697–703.

Windsor JA, Hill GL. Weight loss and physiological impairment: a basic indicator of surgical risk. *Annals of Surgery* 1988; **207**: 290–6.

Wolfe RR, Allsop JR, Burke JF. Glucose metabolism in man: responses to intravenous glucose infusion. *Metabolism* 1979; **28**: 210–20.

Wolfe RR, Durkot MJ, Allsop JR, Burke JF. Glucose metabolism in severely burned patients. *Metabolism* 1979; **28**: 1031–9. [The first study to show that endogenous glucose production could not be suppressed with glucose infusion.]

Young A. Exercise physiology in geriatric practice. *Acta Medica Scandinavica* 1986; (Suppl. 711): 227–32.

Ziegler TR *et al.* Clinical and metabolic efficacy of glutamine-supplemented parenteral nutrition after bone marrow transplantation: a randomized, double-blind, controlled study. *Annals of Internal Medicine* 1992; **116**: 821–8.

2 Trauma and shock

David V. Feliciano, Richard J. Mullins, and Grace S. Rozycki

Trauma
Definition
Epidemiology
Trauma prevention
Mechanisms of injury
Patterns of injury
Trauma scoring systems
Trauma systems
Resuscitation
Injury to the cranium and brain
Injury to the face and eye
Injury to the neck
Injury to the chest
Injury to the abdomen
Shock
Introduction
Pathophysiology of shock
Hypovolemic shock
Systemic inflammatory response syndrome (SIRS) and septic shock
Cardiogenic shock
Neurological shock
Further reading

Trauma

Definition

Derived from the Greek word ‘tpavua', trauma refers to an injury (in + jus, ‘not right') or wound. Either of these is characterized by a structural alteration and/or physiologic imbalance that results when energy is imparted or vital functions compromised during interaction with physical or chemical agents. Trauma encompasses a wide range of mechanisms of injury (Table 1).

Blunt	Motor vehicle crash, Fall, Assault, Crush
Penetrating	Stab, Gunshot, Shotgun, Laceration by foreign body
Poisoning	Solids, Liquids
Drowning	
Burns	Thermal, Electrical, Radiation
Bites and stings	

Table 1 Mechanisms of injury

Epidemiology

There is a lack of interest in trauma prevention and a lack of recognition of the magnitude of trauma as a public health problem in both developed and third world countries. This is surprising because the data documenting the significant adverse effects of trauma on society have been available for many years (Table 2). In ‘Accidental Death and Disability: The Neglected Disease of Modern Society', a monograph authored by the United States National Academy of Sciences/ National Research Council in 1966, it states that ‘Accidents are the leading cause of death among persons between the ages of 1 and 37; and they are the fourth leading cause of death at all ages.'. In the 1996 data for the United States, available from the Centers for Disease Control and Prevention, unintentional injuries (excludes homicides and suicides) were the leading cause of death among persons between the ages of 1 and 44 and, when homicides and suicides are included, the fourth leading cause for all ages. In addition, the years of potential life lost due to premature deaths from unintentional injuries, suicides, and homicides in the United States exceed those due to malignancies, heart disease, or infection with human immunodeficiency virus.

Unintentional 94 948 (major categories below)	
• Motor vehicle crash	43649
• Falls	14986
• Poisoning	9951
• Fire and flames	3741
• Drowning	3488
• All other causes	19133
Intentional 51 874 (major categories below)	
• Suicide	30903
• Homicide	20634
• Legal intervention	337
Undetermined/death at war	3479

* From Injury Facts. 1999 Edition. National Safety Council, Itasca, Illinois.

Table 2 150 298 deaths due to injury in the United States in 1996*

The National Safety Council in the United States has calculated that there were 256 daily deaths from unintentional injuries during 1996. If suicides and homicides are included, nearly 395 individuals died from injury each day in the same year. Of interest, there were 20.7 million unintentional injuries during the same time period. The overall estimate of cost related to unintentional injuries in the United States in 1996 was $444.1 billion, including $74.6 billion in medical expenses.

Trauma prevention

From 1992 to 1996, the number of annual deaths from unintentional injuries in the United States increased slightly to a total of 93 400. There were 43 300 deaths in motor vehicle crashes in 1996, accounting for 46.4 per cent of the total and 16.3 deaths per 100 000 population. While these figures clearly document the continuing magnitude of the injury problem related to motor vehicle crashes, they represent a ‘20th century public health achievement' according to the Centers for Disease Control and Prevention. This is because the number of deaths from motor vehicle crashes represented a 23.1 per cent decrease from the 56 278 reported in 1972. In similar fashion, the number of deaths per 100 000 population represented a nearly 40 per cent decrease from the 26.9 per 100 000 reported in 1972. Factors thought to

contribute to these impressive decreases include the following: (1) improvements in the design of and equipment available on motor vehicles (Table 3); (2) laws mandating the use of restraint devices; (3) the temporary or permanent lowering of the speed limit on highways to 55 miles per h; (4) random stops by law enforcement agencies to assess vehicle safety and driver sobriety; and (5) increased public and judicial awareness due to the efforts of groups such as Mothers Against Drunk Driving (MADD). In truth, all of these are factors (human host, vehicle or agent, environment) that affect the pre-event, event, and post-event phases of an incident leading to injury. The ‘matrix' that combines all these factors was originally described by the late William Haddon, Jr. in 1968. Of interest, the failure to pursue a similar broad-based public health approach to the problem of handgun-associated violence in the United States is documented by the annual 17 000 to 18 000 homicides by firearms.

1954	Safety padding on dashboard
1955	Safety door latch
1964	Lap belts standard
1965	Highway Safety Act
	National Traffic and Motor Vehicle Safety Act
1967	Collapsible steering column
1968	Cross shoulder belts
1969	Side-door beams
1970	Ignition interlock (eliminated later)
1973	Sustain 5 mile/h crash
1984	Mandatory use of seat belt (New York)
1988	Air bag on Chevrolet
1995	Side air bag on Volvo
1995	Day running lights on General Motors

*Source: Detroit Public Library, Detroit, MI.

Table 3 Improved safety of motor vehicles and driving in the United States, 1954–1995*

Mechanisms of injury

Blunt trauma

Victims of motor vehicle crashes, falls, or assaults are injured by impact force and deformation related to deceleration and compression. Determinants of impact force are magnitude (essentially, kinetic energy and area of application), duration of application, and direction of application. The physical deformities that result from impact force are known as strains. Strains are divided into those that are tensile (stretching), shearing (opposing forces across an object), or compressive (crushing). When the elasticity (tendency to regain original condition) or viscosity (resistance to change in shape during motion) of a tissue or organ is exceeded by applied strains, disruption results. In biomechanical terminology, disruption occurs at the elastic limit or break point. Disruption causes injuries to the skin (abrasion, contusion, chop, puncture, incision, laceration), buckling or fracture of bones, and visceral or vascular ruptures.

Other than impact force, factors that determine the magnitude of injury after blunt trauma include gender, impact resistance of body parts, fixation of body parts causing deformation during deceleration, and anatomic protection of body parts. The impact of gender is not clear at this time, but the lighter body skeleton and smaller muscular development of females may be important. In one study the fatality risk in motor vehicle crashes for females aged 15 to 45 years was 25 per cent greater than for males. Bones such as the first and second ribs, sternum, scapula, pelvis, and femur are fractured only when significant impact forces are applied. Therefore, associated injuries should always be suspected when such a fracture occurs. For example, associated injuries to the head, chest, and abdomen occurred in 53, 64, and 33 per cent of patients, respectively, with fracture of the first rib in one review. Fixation of the descending thoracic aorta to the ligamentum arteriosum, the liver to the falciform ligament, and the small bowel to adhesions, the ligament of Trietz, and the retroperitoneal cecum are thought to contribute to injuries occurring in these structures during blunt deceleration trauma. Finally, the close associations of the brain and skull, the right ventricle and the sternal area, and the spleen and ribs 9 to 11 are all related to injuries occurring in these structures.

Penetrating trauma

The kinetic energy of stab wounds is low, and death occurs only if a critical organ such as the heart or a major blood vessel is injured. Injuries from missile wounds are caused by a combination of factors including the following: (1) missile (weight or mass, shape, velocity, kinetic energy at impact); (2) medium (drag or resistance of the medium and coefficient of drag); and (3) pattern of flight of missile upon impact (whether yaw, tumbling, precession, or nutation is present). The most important of these is kinetic energy (KE), which is defined by the following formula:

$$KE = \frac{\text{missile weight (grains)} \times \text{missile velocity}^2 \text{ (ft/s)}}{45\,380}$$

It is obvious that doubling missile velocity results in a quadrupling of the kinetic energy. Therefore, it is helpful to divide muzzle velocities of civilian or military weapons into those that are low (<1100 ft/s), medium (1100–2000 ft/s), or high (>2000 ft/s). Representative cartridges, missiles, muzzle velocities, and muzzle kinetic energies are listed in Table 4.

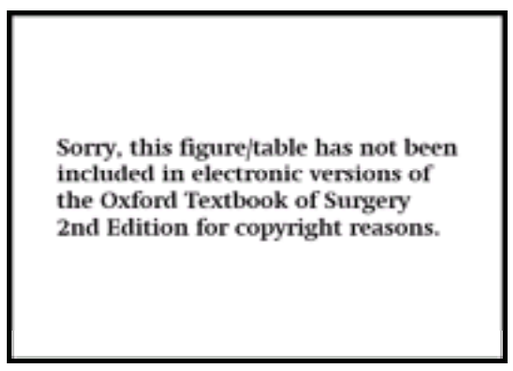


Table 4 Representative cartridges

Direct mechanisms of injury from missile wounds or fragments include cutting or laceration and transfer of heat. Indirect mechanisms include longitudinal low-displacement shock or sonic pressure waves and temporary cavitation from transverse high-displacement shear waves. The low-displacement wave does not appear to cause damage to tissues, while the high-displacement wave associated with medium- or high-velocity missiles significantly increases indirect damage by causing cavitation. This is thought to be due to the transfer of kinetic energy causing alternating collapsing and reforming of the cavity after the missile passes. Tissues that are relatively inelastic such as the brain, liver, and spleen are damaged the most by cavitation. Other factors that increase the magnitude of injury include fragmentation of the missile after striking the victim or the creation of secondary missiles such as fragments of teeth or bone.

Patterns of injury

Blunt trauma

When the entire body is considered, the patterns of injury for victims of motor vehicle crashes are determined by the size of the car, type of accident, position of the victim in the car, body habitus of the victim, and whether or not the victim was wearing a restraint device or had an air bag for protection. Unrestrained drivers who survive high speed frontal crashes have a consistent pattern of injuries which is readily predictable. The four most common areas of injury in such individuals include fracture of the femur (65 per cent), fracture of the pelvis (46 per cent), injury to the thoracic body wall or contents (46 per cent), and fracture of facial bones (37 per

cent).

The pattern of injury for a driver or front seat passenger is modified significantly by the use of restraint devices or the presence of an air bag. In addition, restraint devices have been documented to lower mortality from motor vehicle crashes by up to 50 per cent in several reviews. It should not be surprising, however, that sudden deceleration of the victim into a restraint device or air bag may actually cause certain injuries (Table 5).

Injuries to:	
Compression	Torn or avulsed heart valves
	Ruptured or insufficient valves
	Thrombosis of the coronary or abdominal aorta
Hyperflexion	Closed fracture of thoracic vertebrae
Should be to:	
Subsequently hyperextension and rotation	Internal tear or thrombosis in mesenteric, carotid, subclavian, or vertebral arteries
	Fracture or dislocation of cervical spine
Compression	Internal tear or thrombosis in subclavian artery
	Fracture of ribs
	Pulmonary contusion
	Rupture of upper abdominal viscera
Air bag	
Contact	Corneal abrasions, keratitis
	Fracture of face, neck, chest
Concussive interaction	Ruptured eardrum
Thorax (seat belt)	Fracture of cervical or thoracic spine*
Hyperextension (unrestrained)	Fracture of cervical spine*

Table 5 Injuries related to restraint devices, especially when worn improperly

Another pattern of injury is present in pedestrians struck by motor vehicles. If the vehicle is not braking, fractures of the upper third of the tibia and fibula with or without an associated dislocation of the knee may result. The pedestrian then impacts with the hood or windshield of the vehicle and may sustain injuries to the ribs and enclosed upper abdominal viscera. Finally, the victim slides off the car onto the pavement where further injuries to the head and upper extremities may occur. When two components of the triad described are present, diagnostic efforts should be directed toward determining the presence or absence of the third.

Isolated blunt injuries to the head, neck or trunk or to the extremities also cause consistent patterns of injury (Table 6 and Table 7).

Diagnosed injury	Rule out associated injury
Fracture of parietal bone	Epidural hematoma
Fracture of mandible	Whiplash, fracture of cervical spine
Fracture of first and second ribs or sternum	Cardiac contusion or rupture, rupture of descending thoracic aorta
Fracture of sternum	Compression fracture of thoracic spine
Fracture of scapula	Fracture of ipsilateral ribs, ipsilateral pulmonary contusion
Fracture of ribs 8 to 12 on right	Lacerated liver
Fracture of ribs 8 to 12 on left	Lacerated spleen
Fracture of pelvis	Ruptured bladder, transection of main urethra

Reproduced from Feliciano DV: Patterns of injury. In: Feliciano DV, Moore EE, Mattox KL, eds. Trauma, Textbooks Standard. Appleton and Lange, 1996: 155-183. Copyright The McGraw-Hill Companies, Inc. In no event shall McGraw-Hill have any liability to any party for special, incidental, tort, or consequential damages arising out of or in connection with the McGraw-Hill Material, even if McGraw-Hill has been advised of the possibility of such damages.

Table 6 Patterns of injury in blunt trauma: isolated injuries to head, neck, or trunk

Diagnosed injury	Rule out associated injury
Anterior dislocation of shoulder	Traction on axillary nerve
Fracture of shaft of humerus	Traction on radial nerve
Fracture of supracondylar area of humerus	Intimal tear or thrombosis in brachial artery
Fracture of distal radius	Compression of median nerve
Posterior dislocation of hip	Traction on sciatic nerve
Fracture of supracondylar area of femur	Intimal tear or thrombosis in popliteal artery
Posterior dislocation of knee	Intimal tear or thrombosis in popliteal artery
Fracture of neck of fibula	Traction on peroneal nerve
Open tibial fracture	Intimal tear, thrombosis, or avulsion of popliteal artery or tibial vessel

Reproduced from Feliciano DV: Patterns of injury. In: Feliciano DV, Moore EE, Mattox KL, eds. Trauma. 3rd edn. Stamford: Appleton and Lange; 1996: 155-183. Copyright The McGraw-Hill Companies, Inc. In no event shall McGraw-Hill have any liability to any party for special, incidental, tort, or consequential damages arising out of or in connection with the McGraw-Hill Material, even if McGraw-Hill has been advised of the possibility of such damages.

Table 7 Patterns of injury in blunt trauma: isolated injuries to extremities

Penetrating trauma

Patterns of injury in patients with penetrating wounds are less helpful than with blunt trauma. The fundamental principle when operative management is chosen, particularly for cervical (between the clavicles and angle of the mandible), abdominal, or extremity wounds, is complete exploration of all structures in the possible path of the missile.

Trauma scoring systems

In order to compare results of trauma care in one center over different time periods or between multiple centers, many scoring systems have been developed. These scoring systems have been used for prehospital triage, description of injuries to single organs, description of multiple injuries and correlation with survival, and patient status or care in the intensive care unit and correlation with survival (Table 8).

Prehospital
CRAMS Scale (circulation, respiration, abdomen/thoracic, motor, speech)
PHI or Prehospital Index (blood pressure, heart rate, respiratory rate, consciousness, mechanism)
RTS or Revised Trauma Score (Glasgow Coma Scale, systolic blood pressure, respiratory rate)
Magnitude of injuries – single organ
AIS or Abbreviated Injury Scale – 1990 Revision, UPDATE 98 (severity 1–6)
GCS or Glasgow Coma Scale (3–15, eye opening, verbal response, motor response)
OIS or Organ Injury Scale (severity 1—V or VI)
Magnitude of injuries – multiple organs
AI or Anatomic Index (from ICD-9-CM)
AP or Anatomic Profile (head/spinal cord, thorax/thoracic cord, other serious and nonserious injuries)
ISS or Injury Severity Score (1–75, sum of squares of three highest AIS scores from three different regions)
NISS or New Injury Severity Score (1–75, sum of squares of three highest AIS scores from anywhere in the body)
PATI or Penetrating Abdominal Trauma Index (organ, magnitude of organ injury)
PII or Pediatric Trauma Index (Injury Severity Score, Glasgow Coma Scale, Pediatric Trauma Score)
PTS or Pediatric Trauma Score (size, airway, systolic BP, CNS, open wound, skeletal)
Intensive care unit
APACHE II (Acute Physiology and Chronic Health Evaluation)
TISS-76 Therapeutic Intervention Scoring System

Table 8 Scoring systems used for trauma and surgical critical care

The availability of widely-utilized scoring systems has allowed for better analysis of trauma outcomes. All quality or performance improvement programs in trauma centers also use the scoring systems. In the TRISS method, survival after injury is predicted by using a combination of mechanism, patient age, Revised Trauma Score from the field, and Injury Severity Score. Concerns about TRISS methodology led to the development of ASCOT or A Severity Characterization of Trauma. This system uses mechanism, a scaled interval component for age, the Revised Trauma Score and a modified Anatomic Profile. Finally, ICISS or ICD-9 Injury Severity Score uses ICD-9 discharge diagnoses and survival risk ratios for each diagnosis.

Trauma systems

Since 1976, continuous updates of the monograph now entitled *Resources for optimal care of the injured patient: 1999* have been prepared by the Committee on Trauma, American College of Surgeons. This monograph describes ideal trauma systems, prehospital care, and trauma centers. Particular emphasis is placed upon essential or desirable criteria for level I to IV trauma centers and the proper development of programs in prevention, performance improvement, education and outreach, and research in such centers. Hospitals that care for trauma patients anywhere in the world can use this monograph as a starting point for developing a trauma center.

The *Model trauma care system plan* of the Health Resources and Services Administration, US Department of Health and Human Services, was published in 1992. This document describes the administrative, clinical, and operational components of a regional trauma system. The emphasis is ‘to match a facility's resources with a patient's needs so that optimal and cost-effective care is achieved' (*Optimal Care* monograph, 1999). The inclusive system described recognizes that the majority of injured patients in the United States are cared for in smaller hospitals, many of which are not trauma centers. Also, there is clear recognition that the 15 per cent of patients who are severely injured will receive the most benefit from care in a Level I (regional) or Level II (area) trauma center.

Resuscitation

The Advanced Trauma Life Support Course (ATLS) of the American College of Surgeons Committee on Trauma was developed in 1980 to improve the care of injured patients during initial in-hospital evaluation. After prehospital triage to the appropriate trauma center or local hospital, the injured patient undergoes a simultaneous primary survey and resuscitation. Hemodynamically stable patients then undergo a secondary survey, followed by transfer to another center, if appropriate, or to another area of the same hospital.

Primary survey/resuscitation

The rapid evaluation of the injured patient in the ATLS schema follows the time-honored and physiologically appropriate sequence known as the ‘ABCs'. In this sequence, the ABCs represent the following:

A Airway maintenance with protection of cervical spine
B Breathing and ventilation
C Circulation with control of hemorrhage
D Disability or neurologic status
E Exposure or completely undress the patient, but prevent hypothermia

Asphyxiation causes injury to the brain within minutes, and evaluation of the airway supersedes all other concerns in the multiply-injured patient. Patients who are awake, but cannot respond verbally to a question or respond with a hoarse or stridulous voice, should be suspected of having obstruction of the airway. Any unconscious or comatose patient who presents with gurgling sounds or stridor during attempts at breathing should be suspected of the same. Maneuvers to establish an airway in the injured patient are summarized in Table 9. Surgical cricothyroidotomy is the emergency invasive airway of choice in the ATLS course. ‘Classical' tracheostomy is reserved for patients with crush injuries of the larynx, cricotracheal separation, or perforation/laceration of the trachea with air leaking through the wound in soft tissue. The cervical spine is maintained in the neutral position during any of the airway maneuvers listed.

Single
Return cervical spine to neutral position
Chin lift or jaw thrust
Remove foreign body manually or with suction
Insert oropharyngeal or nasopharyngeal airway
Long tubes
Laryngeal mask airway (for cervical spine injury)
Oral endotracheal tube, over fiberoptic bronchoscope, if necessary
Nasotracheal tube
Esophageal-tracheal Combitube*
Surgical
Needle cricothyroidotomy (wilderness medicine)
Surgical cricothyroidotomy
Tracheostomy

*Sheridan Catheter Co., Argyle, NY.

Table 9 Emergency airway maneuvers in injured patients

With an airway established, failure of one side of the chest to move with spontaneous or assisted respirations is diagnostic of a mechanical problem with ventilation. This may be due to compression of the lung by intrapleural air or blood, inadequate oxygenation and ventilation secondary to a combined flail chest and pulmonary contusion, or bronchial obstruction by a mucus plug. If the patient is agitated, complains of air hunger, and has an oxygen saturation below 90 per cent on fingertip oximetry, a tension pneumothorax should be suspected. It may be necessary in such a patient to perform an emergency ‘blind' needle thoracentesis in the second or third intercostal space in the midclavicular line without a preliminary chest radiograph. Clinical improvement following a release of air through the needle is followed by insertion of a 38-French thoracostomy tube in the midaxillary line at the level of the male nipple. This tube is then connected to an underwater seal drainage system. A patient with an open pneumothorax larger than the glottic opening is best treated by covering the defect with vaseline-impregnated gauze, inserting a thoracostomy tube away from the area of injury, and moving the patient to the operating room for closure. In more stable patients, a chest radiograph will document the presence of a closed pneumothorax or hemothorax that is treated with tube thoracostomy, also. Patients with multiple sequential segmental fractures of the ribs of one side of the chest wall may have a flail chest. In the absence of acute respiratory failure from an associated pulmonary contusion, analgesia for the fractured ribs and pulmonary toilet in the surgical intensive care unit will help to avoid intubation in 60 to 70 per cent of patients. In the absence of intrapleural air or blood or a flail chest with pulmonary contusion, a mucus plug in a bronchus or a bronchial injury is diagnosed during fiberoptic bronchoscopy.

In order to control external hemorrhage, particularly from an extremity, direct finger pressure, a compression bandage, or a proximal blood pressure cuff is applied. The old aphorism that ‘there is no blood vessel outside the human trunk that is larger than the human thumb' still holds, and only shotgun wounds or mangled extremities may require the other techniques listed. Management of the associated circulatory shock is discussed elsewhere in this chapter.

The assessment of disability or neurologic status during the primary survey involves checking the pupillary size and reaction and recording a score according to the Glasgow Coma Scale (GCS) (Table 10). A mild head injury is defined as one with a GCS of 14 to 15, a moderate one with a GCS of 9 to 13, and a severe one with a GCS of 8 or below. Serial recordings of pupillary size and reaction and GCS through the initial assessment to definitive care or transfer are highly accurate techniques in assessing the magnitude of injury to the brain. CT scans of the brain are generally performed in hemodynamically stable patients with a loss of consciousness at the time of trauma.

Assessment area	Score
Eye opening (E)	
Spontaneous	4
To speech	3
To pain	2
None	1
Best motor response (M)	
Obeys commands	6
Localizes pain	5
Normal flexion (withdrawal)	4
Abnormal flexion (decorticate)	3
Extension (decerebrate)	2
None (flaccid)	1
Verbal response (V)	
Oriented	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
None	1

GCS Score = (E+M+V); best possible score = 15; worst possible score = 3.
Adapted from Teasdale G, Jennett B. *Lancet* 1974; ii: 81–4 and Teasdale G[illegible] *Journal of Anaesthesia* 1976; 48: 760–6.

Table 10 Glasgow Coma Scale

Exposure refers to complete disrobing of the patient. This allows for a comprehensive physical examination during the secondary survey or for assessing the number of knife or missile wounds in the operating room prior to an emergency operation. Because of the risk of hypothermia during initial assessment in the emergency center or an emergency operation, the maneuvers listed in Table 11 are used to prevent or reverse this added problem.

[illegible]
Warm trauma resuscitation room
Increase operating room temperature > 85°F
Cover patient's head with a radiant-air warming device
Cover body parts not being examined or out of the operative field with warm blankets or a warming device such as the Bair Hugger® or Life-Air® system
Infuse crystalloid solutions and blood through a warming device such as the [illegible] Fluid Warmer®
Irrigate nasogastric and thoracostomy tubes with warm saline during laparotomy
Irrigate open peritoneal cavity, pleural cavities, and peritoneal cavity during simultaneous thoracotomy or thoracotomy and laparotomy
Turn off ventilatory circuit on anesthesia machine

[a] Augustine Medical, Inc., Eden Prairie, MN.
[b] Medical Products Group, Marshall, MN.
[c] Level 1 Technologies, Inc., Rockland, Mass.

Table 11 Maneuvers to prevent or reverse hypothermia during initial assessment or ‘damage control operations'

Other catheters and monitoring lines are placed during the resuscitation phase. These include a nasogastric or orogastric tube that decompresses gastric dilatation, decreases the risk of aspiration, and documents gastric perforation or rupture. A urethral catheter is inserted, as well, if there are no signs of injury to the male urethra, i.e. blood at the urethral meatus, a large scrotal or penile hematoma, or a ‘high-riding' prostate on rectal examination.

Emergency operation instead of secondary survey

Patients with persistent hypotension despite control of external hemorrhage and infusion of resuscitation fluids during the primary survey/resuscitation phase have internal hemorrhage 97 to 98 per cent of the time (Table 12). Intrapleural hemorrhage exceeding 1200 to 1500 ml in the first 15 min after the insertion of a thoracostomy tube or continuing at a rate of 100 to 200 ml/h mandates a thoracotomy. Persistent hypotension and the presence of fluid on a surgeon-performed abdominal ultrasound or more than 10 to 20 ml blood on a diagnostic peritoneal tap indicates that a laparotomy should be performed. Compression of the iliac wings causing a grinding sound in the comatose patient or eliciting a complaint of pain in the awake patient suggests that a Malgaigne's pelvic fracture is present. Such a patient should undergo application of an external fixator in the emergency center or operating room or transfer to the angiography suite for therapeutic embolization of bleeding pelvic vessels.

Site	Rapid diagnosis by
Intrapleural	Physical examination
	Surgeon-performed ultrasound
	Chest radiograph
	Tube thoracostomy
Intraperitoneal	Surgeon-performed ultrasound
	Diagnostic peritoneal tap
Pelvis	Physical examination
	Pelvic radiograph
Both femurs fractured	Physical examination
	Pelvis and thigh radiographs
External hemorrhage (scalp, mangled extremity, etc.)	Physical examination

Table 12 Sites of hemorrhage causing hypovolemic shock in severely injured patients

Persistent hypotension in the 2 to 3 per cent of patients without trauma-related hemorrhage is caused by one of the following: (1) injury to the spinal cord; (2) cardiac arrhythmia, blunt cardiac contusion or perforation, myocardial infarct; or (3) pre-event use of medications that suppress cardiovascular responses.

Secondary survey

Following completion of the primary survey and resuscitation phases, hemodynamically stable patients undergo a secondary survey. This is a head-to-toe physical examination that initiates a series of radiographs or other diagnostic studies to determine the presence of all injuries.

Transfer

Hemodynamically stable patients in the trauma center should be transferred to CT, angiography, the surgical intensive care unit, or the operating room after the secondary survey is completed. The location will depend on the injuries diagnosed during the primary and secondary surveys.

Hemodynamically stable patients with multiple or complex injuries requiring specialized care should be transferred from rural or other small hospitals to the closest area or regional trauma center.

Injury to the cranium and brain

Cranium

Injury to the scalp results in significant hemorrhage that may cause hypovolemic shock in children and some adults. Rapid control of hemorrhage in the emergency center is attained by applying metal skin clips or a continuous suture. Deeper lacerations in hemodynamically stable patients are closed in layers including the galea aponeurotica, connective tissue, and skin.

The calvaria or cranial vault consists of the frontal and parietal bones and the squamous portion of the occipital and frontal bones. The base of the skull is formed by the occipital and sphenoid bones. Significant injury to the calvaria or base of the skull may result in a fracture with or without significant injury to the underlying brain.

Closed fractures of the calvaria have traditionally been diagnosed on radiographs of the skull. If skull films are obtained and a fracture, intracranial air or shift of the pineal gland to one side is present, a CT scan of the head should be obtained. In the modern era, CT of the head is available in most hospitals and is used in all patients with loss of consciousness at the time of trauma or a GCS of 13 or less at initial assessment. Therefore, the diagnosis of a fracture in the calvaria or base of the skull is most commonly made with this modality.

Patients with closed fractures of the calvaria or base of the skull are admitted to the hospital for serial assessments of GCS, pupillary size and reactivity, and neurologic status. Depressed skull fractures extending beyond the thickness of the normal skull bone or open skull fractures undergo operative management. Underlying hematomas compressing the brain or tears in the dura mater are treated simultaneously.

Brain

In most patients who survive after sustaining blunt multisystem trauma, the magnitude of injury to the brain is the major determinant of functional outcome. In general, 10 per cent of patients with a GCS of 14 to 15 will have some chronic disability, while approximately 10 per cent of patients with a GCS of 9 to 13 will die and more than 50 per cent will have some chronic disability. The mortality rate for patients with a GCS of 8 (coma) or lower is 30 per cent, and nearly all survivors have chronic disability.

It is now clear that hypoxemia and hypotension in the field or in the hospital will have adverse effects on the patient with injury to the brain. These secondary insults will more than double the mortality when a severe injury to the brain is already present. Therefore, early intubation, administration of Ringer's lactate solution, and diagnosis and control of ongoing hemorrhage are mandatory in such patients.

The diagnosis of moderate or severe injury to the brain is first suspected by a GCS of 13 or less in the field or emergency center. Once vital signs have been stabilized, a CT of the head is performed to determine whether cerebral edema, subarachnoid or intraventricular hemorrhage, an extracerebral hematoma, or an intracerebral hematoma or contusion is present. The diagnosis of cerebral edema on a CT is generally followed by the insertion of a monitor to determine intracranial pressure. Confirmation of a significant elevation in intracranial pressure (normal = 10 mmHg) or a fall in cerebral perfusion pressure (mean arterial blood pressure—ICP, normal = 80 mm), is treated with hyperventilation to a $PaCO_2$ of 30 mmHg, intravenous administration of 20 per cent mannitol at 1 g/kg, and use of an intravenous loop diuretic. Drainage of cerebrospinal fluid and the institution of barbiturate coma are used when standard measures are ineffective and there is a risk of uncal herniation. While subarachnoid hemorrhage in association with a parenchymal injury in the brain demands no specific treatment, intraventricular hemorrhage is treated with the insertion of a ventriculostomy catheter and drainage of cerebrospinal fluid. A patient with a lens-shaped epidural hematoma on CT has often presented with the sequence of losing consciousness at the time of trauma, having a ‘lucid interval' in the emergency center, and then rapidly becoming comatose. Most commonly associated with transverse parietal fractions of the skull, an epidural hematoma is drained and the middle meningeal artery ligated through a limited craniotomy. Subdural hematomas are present in 30 per cent of patients with severe injuries to the brain and are thought to be due to tearing of the veins between the cerebral cortex and dura. As the underlying brain is often damaged and the extent of the clot is greater than with an epidural hematoma, a more extensive craniotomy is indicated. An intracerebral hematoma noted on CT may require ultrasound-guided surgical drainage through a limited craniotomy when compressive effects are significant (Fig. 1). Frontal or temporal intracerebral contusions have a ‘salt and pepper' appearance on CT, do not require surgery, and may lead to in-hospital behavioral problems for the patient.



Fig. 1. Right-sided intracerebral hematoma on CT of the brain.

Even in the presence of a normal CT of the head, a cerebral concussion or diffuse axonal injury (DAI) may be present. A cerebral concussion is present when the patient has lost consciousness at the time of injury and suffers retrograde amnesia for the event. While most patients recover fully with the exception of the amnesia, others develop a post-concussion syndrome with dizziness, nausea, and problems with memory. DAI has become a ‘catch-all' diagnosis for those patients with severe sequelae of an injury to the brain (posturing, sweating, hyperpyrexia, hypertension), but no evidence of an ischemic or mass lesion on CT. Prognosis is ominous as predicted by use of the Glasgow Outcome Scale.

After reducing elevations in intracranial pressure and maintaining cerebral perfusion pressure, there are other critical components of neurointensive care. Included among these are avoiding hypoxemia, hypotension, hyperglycemia and hyperpyrexia, treating seizures aggressively, and instituting early gastric or enteral feedings to counteract the severe catabolism that occurs in many of these patients.

Injury to the face and eye

Face

Injuries to the midface or lower face may compromise the patient's upper airway. If an individual with extensive experience in emergency orotracheal or nasotracheal intubation is unsuccessful at a first attempt in a patient with facial injuries, cricothyroidotomy in the emergency center is indicated. Control of hemorrhage from injuries to soft tissue is obtained by a pressure dressing or rapid insertion of a continuous stitch in patients who will require further diagnostic evaluation in the emergency center. Other hemodynamically stable patients with isolated severe lacerations or avulsions of soft tissues in the face are moved to the operating room for repair. This location allows for safe clamping of major vessels, careful inspection of problematic anatomical areas such as the parotid duct, facial nerve, eyelids and lacrimal canaliculi, and meticulous debridement and cosmetic closure (Fig. 2).

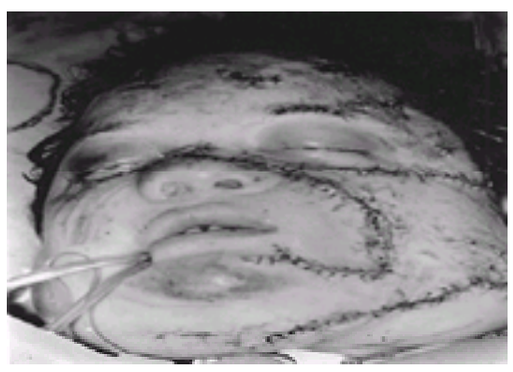


Fig. 2. Extensive facial lacerations secondary to impalement in windshield were repaired in the operating room.

Fractures of facial bones may be detected on gentle palpation of the nose, orbital rims, maxillary bones, zygomatic bones, and mandible. Movement of the nasal bones or midface or the presence of malocclusion is pathognomonic of fractures. A fracture of the mandible can be confirmed on a panographic radiograph available in the emergency center, while other fractures are confirmed on axial and coronal CT scans in the modern center. Most facial fractures are combinations of the patterns (I – alveolar ridge; II – pyramidal; III – craniofacial dissociation) described by the French surgeon LeFort in 1901. Current principles of repair include cosmetic incisions, reduction of fragments, and use of plates and screws for stabilization.

Eye

Most injuries of the eye are readily detected by a careful examination of the orbit, eyelids, pupils, globe, and visual acuity. B-mode ultrasound is useful in determining whether intraocular hemorrhage or a foreign body is present, while axial and coronal CT scans are useful in analyzing orbital blowout-type fractures.

Ocular injuries that may be cared for by general surgeons in hospitals without ophthalmology staff include foreign bodies of the cornea, corneal abrasions, limited lacerations of the eyelids, and preliminary care of chemical burns. Moist cotton swabs are used when attempting to remove foreign bodies of the cornea. When this is unsuccessful, an ophthalmologist with a slit lamp must be consulted. Corneal abrasions are confirmed by applying fluorescein to the injured eye and noting the presence of greenish fluorescence under a Wood's light or special flashlight. Limited abrasions are treated with the administration of cycloplegic and antibiotic solutions, while larger abrasions are then patched until a re-examination. Nonmarginal lid lacerations may be closed by the general surgeon, while marginal lid lacerations, lacerations near the medial canthus, and injuries to the globe (Fig. 3) are referred to an ophthalmologist. Chemical injuries involving either alkali or acid burns are managed by immediate irrigation with a solution of neutral pH after topical anesthesia has been applied.



Fig. 3. Stab wound to the left eye.

Injury to the neck

Blunt trauma

Blunt trauma to the neck may injure the laryngotracheal complex, carotid or vertebral arteries, or the cervical vertebrae. The possible presence of this last injury in all patients with blunt cervical trauma complicates all emergency airway maneuvers.

Fractures of the larynx occur as unrestrained drivers make contact with the upper rim of the steering wheel in frontal crashes. Dysphonia, hoarseness, or stridor may result, and an emergency tracheostomy may be necessary for control of the airway. This is especially true if cricotracheal separation occurs, as an attempt at intubation may cause collapse of the tenuous airway that remains. Patients without airway distress may be evaluated by a cervical CT. Management ranges from voice rest to operative laryngofissure.

Blunt cerebrovascular injuries result from a victim ‘submarining' (hyperextension and rotation of the neck) under the shoulder harness component of a restraint device in a motor vehicle crash. In one study from the Denver Health Medical Center, independent predictors of a blunt injury to the carotid artery included GCS of 6 or less, a diagnosis of diffuse axonal injury, fracture of the petrous bone, or a LeFort II or III fracture. Another study from the same group documented that injuries to the cervical spine (fracture, dislocation, or ligamentous) were present in 71 per cent of patients with a blunt injury to the vertebral artery, but there was no consistent level or pattern of injury. Heparinization is indicated in patients with blunt carotid or vertebral injuries in the absence of injuries to the brain and lowers the incidence of neurological deterioration or stroke.

The incidence of injuries to the cervical spine is surprisingly low (< 2 per cent) in patients with injuries to the brain, though this increases when there is a GCS score below 10. In contrast, 35 per cent of patients with injuries to the cervical spine have an associated injury to the brain. An awake and alert patient without complaints of pain or tenderness in the cervical spine has an incidence of injury of less than 0.3 per cent. In symptomatic or unresponsive patients, in line stabilization with a rigid cervical collar is maintained from the scene of injury until final in-hospital clearance on routine radiographs or a cervical CT. Routine lateral, anteroposterior, and open mouth odontoid cervical radiographs are adequate to identify 95 to 98 per cent of vertebral fractures in this location. CT is used when routine radiographs are inadequate because of patient size or when findings are equivocal. In order to restore alignment in the patient with a fracture-dislocation of the cervical spine, axial traction using weights is applied through Gardner-Wells tongs in the emergency situation. Maximal recommended weights are 30 lb for the upper and 60 lb for the lower cervical spine. Long-term stabilization is maintained by the subsequent application of a halo vest.

An injury to the cervical spinal cord in association with a fracture-dislocation of the spine should always be suspected. The diagnosis is confirmed by a careful evaluation of the motor and sensory systems, testing the bulbocavernosus reflex, and assessing tone of the anal sphincter in awake patients. In the comatose patient with an injury to the cervical spine, a cervical MRI may be necessary to confirm compression, distortion, or transection of the spinal cord when physical findings cannot be properly assessed. Whenever an injury to the spinal cord at any level is suspected or confirmed, the current standard of care mandates that an intravenous bolus of 30 mg/kg of methylprednisolone be administered as soon as possible after injury. After an intravenous bolus within 3 h of injury, a continuous intravenous infusion of methylprednisolone at 5.4 mg/kg/h is maintained for 23 h. Patients receiving the bolus 3 to 8 h after injury are maintained on the infusion for 48 h. The results of the NASCIS (National Acute Spinal Cord Injury Study) 2 and 3 in the United States are still being discussed, but steroids are currently administered for both medical and medicolegal reasons. In-hospital care for the patient with an injury to the cervical spinal cord includes the following: (1) pulmonary toilet to avoid atelectasis and pneumonia; (2) treatment of suctioning-induced bradycardia; (3) institution of bladder and bowel regimens; (4) frequent turning or use of a rotating bed to avoid sacral and heel decubiti; and (5) recognition of poikilothermy.

Penetrating

The cervical region is divided into penetrating wounds below the sternal notch (zone I), between the sternal notch and the angle of the mandible (zone II), and above the angle of the mandible (zone III). Most discussion focuses on injuries in zone II, the most common site of injury. One subset of patients present with overt symptoms or signs, including a rapidly expanding pulsatile hematoma, internal bleeding into the airway, external bleeding, or loss of the airway from compression or direct injury. Management in the emergency center includes emergency intubation or cricothyroidotomy if hemorrhage or compression is affecting the airway or tracheostomy if direct injury to the trachea has occurred. Either internal or external hemorrhage from the carotid artery is controlled by pressure at the base of the neck or over the area of injury as the patient is moved to the operating room.

Mandatory cervical explorations in asymptomatic patients with only stab wounds penetrating the platysma muscle in zone II result in a greater than 50 per cent ‘negative' exploration rate. For this reason, such patients undergo nonoperative management and serial physical examinations. Indications for a delayed operation include a new onset of a pulsatile hematoma, dysphagia, extensive crepitus, or signs of inflammation. Patients with gunshot wounds traversing the anterior triangle in zone II are rarely asymptomatic, and such patients are almost always managed with more extensive diagnostic testing or a cervical exploration.

A third group of modestly symptomatic patients have a penetrating injury in proximity to the carotid-jugular vessels, a stable hematoma, complaints of hoarseness, dysphagia, or odynophagia, or palpable crepitus. When there is missile or pellet proximity to the carotid sheath or a stable hematoma, a carotid arteriogram or color flow Doppler study is indicated (Fig. 4). In patients with hoarseness, dysphagia, odynophagia, a positive ‘sip' test (increased pain with swallowing a mouthful of water), or palpable crepitus, a contrast esophagogram with a water soluble dye or a mixture of such dye and barium is followed by esophagoscopy and/or bronchoscopy in the

operating room.



Fig. 4. Gunshot wound with occlusion of the left internal carotid artery in an asymptomatic patient.

In general, patients with penetrating injuries in zone I or III that are in proximity to the carotid arteries undergo an arch aortogram or color flow Doppler study.

The operative approach in unstable patients with presumed vascular injuries in zone I is a high (above the nipples) bilateral anterolateral thoracotomy. Stable patients with arterial injuries visible on arteriography undergo a median sternotomy with cervical or supraclavicular extensions, as needed (Fig. 5). Unilateral injuries in zone II are approached through an oblique cervical incision on the anterior edge of the sternocleidomastoid muscle, while gunshot wounds with bilateral cervical traverse are usually managed through a high ‘collar' incision with lateral extensions, as needed. An injury to the high carotid artery that is actively hemorrhaging in zone III is approached through an oblique cervical incision for proximal vascular control. Direct visualization of the injury is obtained through a vertical ramus osteotomy of the mandible with separation of the segments. Small carotid pseudoaneurysms or carotid-jugular fistulas in zone III are being treated with intraluminal stents in some centers.

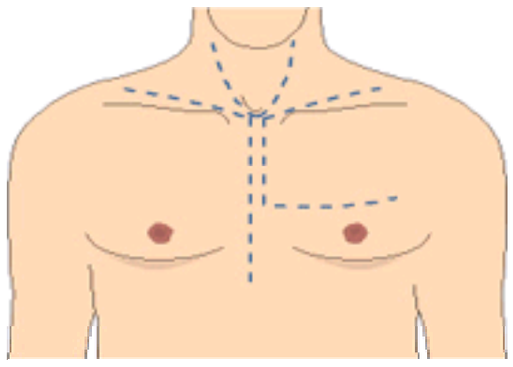


Fig. 5. Cervicothoracic incisions used in patients with lower cervical-upper thoracic injuries.

Penetrating injuries to the cervical trachea in zone II are managed by limited debridement and transverse closure with interrupted 3–0 absorbable sutures with the knots outside. Extensive perforations may be amenable to an end-to-end anastomosis after a laryngeal-lowering procedure has been performed. The trachea has an acceptable rate of healing if the lateral blood supply is preserved, and a tracheostomy is not indicated to ‘protect the repair'. Pressure on the repair is minimized by the presence of a balloon cuff on the endotracheal tube.

Penetrating injuries to the cervical esophagus in zone II are closed transversely with one or two layers of absorbable suture. When an associated injury to the trachea or carotid artery is close to the esophageal repair, a muscle flap using the sternal head of the sternocleidomastoid muscle is mobilized and sutured in place between the repairs. Because of a significant leak rate, a drain is usually inserted close to the esophageal repair. Such a drain should exit anteriorly and never be placed in contact with the carotid artery or sheath. On rare occasions, extensive loss of tissue from the cervical esophagus is best treated by converting the wound to a side ‘blowhole' esophagostomy exiting through the cervical incision. This technique eliminates the high leak rate after an emergent end-to-end anastomosis and may allow for simple delayed closure as edema and inflammation resolve over several weeks.

Injury to the chest

Blunt trauma

As previously noted, compressive forces to the chest wall may cause fractured ibs. Pain control is a primary goal in the elderly patient, the smoker, or in any patient with multiple or bilateral fractured ribs. This is accomplished by intercostal nerve blocks with 0.25 per cent bupivacaine, intrapleural instillation of 0.50 per cent bupivacaine, or the insertion of an epidural catheter for continuous administration of fentanyl citrate. An associated problem in patients with multiple rib fractures is blood loss, which has long been reported to be 125 ml/rib (source unknown).

The first rib is known as the ‘rib of superlatives' as it is the highest, shortest, flattest, most curved, and strongest. A patient who sustains a fracture of the first rib has sustained significant compressive forces to the cervicothoracic area, with a mortality exceeding 35 per cent in older studies. This mortality has been due to associated injuries to the brain, heart, descending thoracic aorta and other great vessels, and upper abdominal viscera (Fig. 6). With many studies documenting that fractures of the first rib are not statistically significant predictors of rupture of the descending thoracic aorta, mandatory thoracic aortography is no longer performed in such patients. Contrast-enhanced thoracic CT is more commonly used as a noninvasive screening tool to evaluate the superior mediastinum and great vessels.

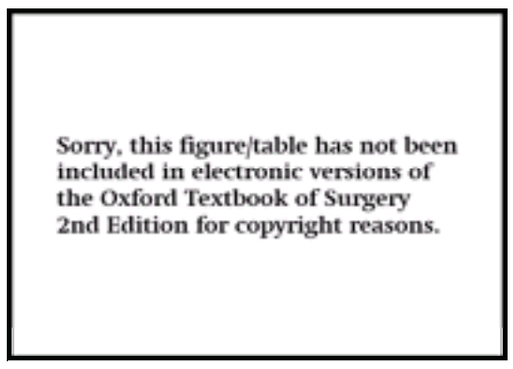


Fig. 6. Widening of the superior mediastinum, especially towards the right side, in a patient who also has fractures of the right first and second and left second ribs (arrows) and a deceleration tear of the innominate artery.

Multiple fractures of the ribs may cause blunt pulmonary injuries such as contusions or hematomas. Contusions are usually self-limited unless vigorous resuscitation with crystalloid solutions has been necessary to treat hypovolemic shock from other injuries. In stable patients, treatment of the chest wall pain, use of an incentive spirometer or suctioning to prevent atelectasis, and use of maintenance fluids only is appropriate during the 3 to 5 days it takes for the contusion to clear. Pulmonary

hematomas displace normal parenchymal tissue, take longer to resolve, and may lead to a post-traumatic pulmonary pseudocyst with an air–fluid level.

Blunt cardiac injury (formerly known as ‘cardiac contusion') of clinical significance is extraordinarily rare. The only initial diagnostic evaluation necessary to determine whether or not this injury is present is an electrocardiogram (ECG). Patients with direct blows to the sternum or even sternal fractures do not have a blunt cardiac injury when there is a normal ECG. In patients with direct blows to the sternum and presumed new-onset ECG changes (arrhythmia or block), unexplained hypotension, or a history of medical or surgical cardiac disease, admission to a telemetry unit is appropriate. Neither creatine kinase–myocardial band nor cardiac troponin T and I contractile protein levels need to be measured at any time, and radioisotope cardiac scans are never indicated. On rare occasion, a transesophageal echocardiogram (TEE) is indicated when hypotension is resistant to treatment and an intracardiac defect is suspected.

Blunt rupture of the heart from compression generally occurs at the junction of the right atrium and superior or inferior vena cava, in the left atrial appendage, or at the right ventricular outflow track. When tamponade causes a cardiac arrest in the prehospital period, there are essentially no survivors. In patients arriving with signs of life after sustaining blunt thoracic trauma in a frontal crash, a surgeon-performed ultrasound evaluation of the pericardium will allow for early detection of tamponade. In the reasonably stable patient with this diagnosis, rapid transfer to the operating room for a median sternotomy and cardiorrhaphy is indicated.

An emergent or urgent thoracotomy is necessary in approximately 7 to 8 per cent of all patients with blunt thoracic injuries. Classical indications include intrapleural hemorrhage from an injury to an intercostal or internal mammary artery or to the hilar area of the lung, blunt rupture of the heart, or rupture of the descending thoracic aorta.

Penetrating trauma

Penetrating wounds of the lung resulting in a pneumothorax and/or hemothorax are treated successfully with a thoracostomy tube 85 to 90 per cent of the time. This is because the systolic blood pressure in the pulmonary artery is only 20 per cent of the systemic systolic blood pressure. Also, coaptation of the visceral and parietal pleural surfaces is extremely effective in sealing air leaks or areas of peripheral hemorrhage from the lung. Continuing hemorrhage of 100 to -200 ml/h through a thoracostomy tube during a period of observation indicates that an urgent thoracotomy is necessary (Fig. 7). Persistence of a significant hemothorax after insertion of a thoracostomy tube is managed with evacuation through video-assisted thoracoscopy in the modern trauma center.



Fig. 7. A patient with a gunshot wound to the lateral left chest (skin marker) eventually needed a thoracotomy to control hemorrhage from the lung.

Penetrating wounds of the heart result in fatal prehospital tamponade or exsanguination in 50 to 75 per cent of victims. In agonal patients arriving in the emergency center with a precordial or transmediastinal wound, an emergency center left anterolateral thoracotomy is performed at the inferior edge of the male nipple (the left breast is retracted upward in females). Should the entrance would be to the right of the anterior midline, a bilateral anterolateral thoracotomy is performed. Upon entrance into the left hemithorax, the left lung is manually elevated and the descending thoracic aorta clamped in the posterior mediastinum. A cardiac injury is present when there is an obvious wound in the pericardium, obvious clot outside the heart in the pericardial sac, or the heart is asystolic. A pericardiotomy is performed with scissors in a longitudinal direction superior to the left phrenic nerve. The heart is swung into the left chest, and a rapid visual and manual inspection is performed to determine the location of the perforation. Small perforations of the atria are isolated over a Satinsky clamp, while small perforations of the ventricles are sealed with a finger or closed with a stapler. Larger or posterior holes in the ventricles are controlled with the insertion of an 18 French bladder catheter with a 5-ml or 30-ml balloon. Inflation of the balloon combined with traction will seal most deficits. Cardiorrhaphy is performed using 3–0 or 4–0 polypropylene sutures. Wounds of the ventricles, particularly the thin-walled right ventricle, are repaired with polypropylene sutures passed through rectangular Teflon pledgets. Prior to completion of any ventricular suture line, the ventricle is elevated to allow for the evacuation of retained air. Removal of the clamp on the descending thoracic aorta is performed gradually as intravenous infusions of crystalloid solutions or blood and bicarbonate are administered.

In stable patients with precordial or transmediastinal wounds, surgeon-performed ultrasound of the pericardium is performed as previously described (Fig. 8). False-negative examinations in reasonably sized patients with pericardial blood are almost never reported, while occasional false-positive examinations from pericardial fluid do occur. When clinical suspicion for a cardiac injury is low despite the presence of a ‘positive' surgeon-performed ultrasound in the emergency center or when a suboptimal ultrasound view is obtained in a large patient, a subxiphoid pericardiotomy performed in the operating room under general anesthesia is indicated. Survival rates for patients undergoing thoracotomy in the emergency center or operating room for penetrating cardiac wounds obviously depend on the mechanism of injury, vital signs upon admission, the speed of diagnosis, and the quality of surgical therapy. In patients with small stab wounds of the heart and reasonable vital signs upon arrival, survival will be in the 90 to 100 per cent range after a median sternotomy in the operating room. In contrast, the survival rate for a profoundly hypotensive or agonal patient with a gunshot wound of the heart will be in the 0 to 15 per cent range after a left anterolateral thoracotomy in the emergency center.



Fig. 8. Surgeon-performed ultrasound is ‘positive' for tamponade in patient with a parasternal stab wound.

Penetrating wounds of the esophagus are uncommon, but will cause death if treatment is delayed. Any wound in proximity to the thoracic esophagus should be evaluated by an immediate CT-esophagogram or routine contrast swallow. Confirmation of a negative study is obtained by a ‘negative' flexible esophagoscopy in the operating room. Wounds of the thoracic esophagus are managed through a right or left posterolateral thoracotomy, depending on the level. A primary transverse closure is buttressed with a rhomboid muscle flap in the upper esophagus, a Thal fundal patch or Nissen fundal wrap in the distal esophagus, or a Grillo pleural flap or intercostal muscle flap at any level.

Injury to the abdomen

Blunt injury

Obvious indications for an emergency celiotomy in a patient who has suffered blunt abdominal trauma include hypotension in combination with a rigid distended abdomen, peritonitis, or evisceration. The problem in evaluating many patients with possible blunt abdominal injuries, however, is that the physical examination is compromised by one of the following: (1) altered sensorium (injury to the brain, ingestion of alcohol or illicit drugs); (2) altered sensation (injury to the spinal cord); (3)

injuries to adjacent structures (ribs, pelvis, thoracolumbar spine); or (4) physical examination is equivocal. In the groups described, additional diagnostic tests are necessary to confirm a diagnosis of intra-abdominal injury.

Hypotensive patients with possible intra-abdominal hemorrhage undergo either a diagnostic peritoneal lavage or surgeon-performed ultrasound. An open or closed diagnostic peritoneal tap/lavage is performed through an infraumbilical midline site (supraumbilical if there is pelvic fracture) after the insertion of a nasogastric tube and bladder catheter. The return of 10 to 20 ml of gross blood or of bile, succus entericus, stool, or food material is a positive ‘tap', and immediate celiotomy is indicated. In the hypotensive patient with a grossly negative ‘tap', the value of a subsequent time-consuming lavage with 1000 ml of normal saline solution is questionable. Diagnostic peritoneal lavage is invasive, has an accuracy of 95 to 98 per cent in detecting intraperitoneal hemorrhage, and has a complication rate of 0.5 to 1 per cent.

The previously described surgeon-performed ultrasound of the pericardium is extended onto the right upper quadrant, left upper quadrant, and suprapubic area to complete the ‘FAST' (focused assessment for the sonographic examination of the trauma patient) in hypotensive (or stable) patients. As an alternative to diagnostic peritoneal lavage, a surgeon-performed ultrasound is a noninvasive and rapidly-performed assessment of the abdomen in the injured patient. An anechoic image (fluid or blood) in either subphrenic area, Morison's pouch, the splenorenal fossa, or the pouch of Douglas/pararectal areas in a hypotensive patient is essentially 100 per cent accurate in confirming the need for an emergency celiotomy.

Hemodynamically-stable patients in whom the physical examination is compromised by any of the previously described factors are first evaluated by a surgeon-performed ultrasound. In the absence of any intraperitoneal fluid on a first and subsequent examination, the need for a follow-up examination of the abdomen with a CT scan is questionable – even in the intoxicated patient. The presence of intraperitoneal fluid on the ultrasound mandates a follow-up ab-dominal CT to localize the source of hemorrhage. Findings on anabdominal CT with contrast that indicate the need for an urgent celiotomy, even in a hemodynamically stable patient, are listed in Table 13.

Diagnosed hemodynamic injury	
Active intraperitoneal or retroperitoneal hemorrhage	
AAST Organ Injury Scale Grade IV or V injury to the liver, spleen, or kidney with ‘blush sign' (active hemorrhage) or large hemoperitoneum	
AAST Organ Injury Scale Grade III, IV, or V to the pancreas (ductal transection or massive disruption involving head of gland)	
Contrast extravasation from duodenum, small bowel, colon, or intraperitoneal bladder	
Renal capsular enhancement only, suggesting blunt thrombosis of renal artery (controversial)	

AAST = American Association for the Surgery of Trauma

Table 13 Findings on contrast CT of the abdomen that indicate the need for an urgent celiotomy after blunt trauma

When an abdominal contrast CT is not available to evaluate the hemodynamically-stable patient, a chest radiograph, flat plate radiograph of the abdomen, or contrast study of the gastrointestinal or genitourinary tract will detect all of the injuries in Table 13 except those to the liver, spleen, or pancreas.

As previously noted under ‘Patterns of injury', it is important to evaluate patients with marks of a restraint device across the lower abdomen after blunt deceleration/compression trauma carefully. In a former era, many of these patients would have died in a frontal collision. By surviving in the modern era of restraints, death is exchanged for deceleration/compression injuries in the abdomen. In one review of 61 children with a ‘linear ecchymosis across the abdomen' (seatbelt sign) after a motor vehicle crash, 14 children (23 per cent) injured a hollow viscus, 13 (21 per cent) had an injury to the lumbar spine, and five (8 per cent) had injury to both.

Penetrating trauma

Approximately 25 to 33 per cent of patients with stab wounds of the anterior abdominal wall (between anterior axillary lines) do not have penetration of the peritoneal cavity. Therefore, stable and cooperative patients without obvious indications for a celiotomy (see below) should first undergo a wound exploration under local anesthesia in the emergency center. In the absence of penetration of the anterior fascia or, if possible to determine, penetration of the peritoneum despite a deep track, the patient is discharged after the stab wound site is irrigated and closed. Penetration of the anterior fascia in large patients or the anterior peritoneum in thin patients mandates further evaluation. The most common option chosen around the world is serial physical examinations for 24 h by a surgeon or senior resident. This noninvasive approach results in a delay to definitive operation in only 5 to 6 per cent of patients with intra-abdominal injuries. A second option is to perform a standard diagnostic peritoneal tap/lavage with positive results being the same as described for blunt abdominal trauma. This invasive technique results in a certain number of false-positive results (bleeding from the site of the stab wound), occasional false-negative results (early lavage after smallstab hole of midgut), and has an overall accuracy of 88 to 94 per cent. In the 45 to 50 per cent of patients who are originally asymptomatic despite having penetration of the anterior peritoneal cavity, 50 per cent of these will eventually come to a celiotomy based on a changing physical examination or on a positive tap or lavage.

Approximately 50 to 55 per cent of patients with anterior stab wounds penetrating the peritoneal cavity have the same obvious indications for an emergency celiotomy as do patients with blunt abdominal trauma. In addition, patients with the following should also undergo celiotomy: (1) new onset hematemesis, proctorrhagia, or hematuria; (2) evidence of a left-sided diaphragmatic defect on finger palpation prior to insertion of a thoracostomy tube; or (3) contrast radiography evidence of an injury to the kidney (significant injury), ureter, or bladder.

The management of stab or gunshot wounds to the flank (between anterior and posterior axillary lines from sixth intercostal space to iliac crest) or to the back (posterior to posterior axillary line from tip of scapula to iliac crest) has changed over the past 20 years. Because of the large bulk of muscles in this area in young males, the routine celiotomies that were performed in the past were often negative. In patients in whom a local wound exploration does not reveal the end of a stab wound track, either serial physical examinations or double (intravenous and oral) or triple (add rectal and colon) contrast CT is performed. Serial examinations result in a false-positive (unnecessary celiotomy)/ false-negative (delayed celiotomy) rate of 5 per cent. Examination of most or all retroperitoneal viscera and vascular structures using double or triple contrast CT has an overall accuracy rate of 96 to 97 per cent.

Older data documented that gunshot wounds traversing the peritoneal cavity resulted in visceral or vascular injuries needing surgical repair in 96 to 98 per cent of patients. In recent years, it has become clear that 15 to 30 per cent of patients with gunshot wounds in proximity to the peritoneal cavity or visceral–vascular retroperitoneum actually have missile tracks that pass through the body wall or anterior–lateral extraperitoneal area, only. In addition, some centers are observing isolated gunshot wounds to the liver or kidney in stable patients in whom an emergency CT mostly rules out associated injuries to the gastrointestinal tract. When the hemodynamically stable patient without peritonitis presents with a possible extraperitoneal gunshot wound, serial physical examinations rather than an emergency celiotomy are appropriate. If available, a surgeon-performed ultrasound documenting intraperitoneal fluid (blood in the ‘asymptomatic patient') would be followed by celiotomy rather than serial examinations.

Diaphragm

Both blunt and penetrating injuries to the hemidiaphragm have been missed in all trauma centers. Early laparoscopy with a 30° scope or thoracoscopy should be performed whenever a chest radiograph or abdominal CT cannot rule out this injury (Fig. 9). Acute repairs are performed through a celiotomy with 1 polypropylene suture. Long posterolateral blunt ruptures are best performed using an interrupted suture technique. Should the entire posterior aspect of the hemidiaphragm be avulsed from the body wall, interrupted polypropylene sutures passed through the free edge and around a posterior rib will allow for secure fixation. A chronic post-traumatic left diaphragmatic hernia is approached through a lateral or posterolateral thoracotomy to permit safe division of adhesions between abdominal viscera and pleural surfaces.



Fig. 9. Acute post-traumatic left diaphragmatic hernia following blunt trauma.

Spleen

In adult patients with blunt trauma and a splenic injury, approximately 60 to 65 per cent are hemodynamically stable and are managed nonoperatively with a 90 to 98 per cent success rate. Failure of nonoperative management is likely when there is active hemorrhage on the contrast CT (‘hot spot') or when there is an American Association for the Surgery of Trauma Organ Injury Scale (AAST OIS) Grade III or IV injury with a ‘large' hemoperitoneum. Operative management is necessary in 35 to 40 per cent of patients with blunt injuries, and 40 to 45 per cent of these spleens undergo operative splenorrhaphy with a 98 per cent success rate (no reoperation for bleeding). Splenorrhaphy can be performed with the application of topical hemostatic agents, continuous suture using chronic or polypropylene material, or use of the argon beam coagulator in 85 per cent of patients (Fig. 10). Another 5 to 10 per cent of patients will have no hemorrhage from the injured spleen, while fewer than 5 per cent will require partial or hemisplenectomy. The latter procedure is only useful in maintaining postoperative ‘splenic immunity' when the remaining fragment is perfused by a main branch of the splenic artery.

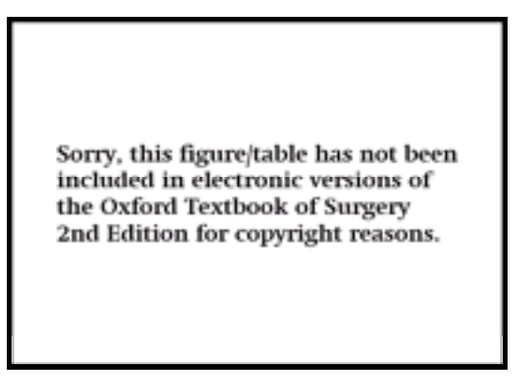


Fig. 10. Suture splenorrhaphy and topical agent in a patient with blunt trauma to the spleen.

Operative repairs in patients with stab wounds or gunshot wounds to the spleen are possible in 75 per cent and 30 to 35 per cent of patients, respectively. Much as in patients with blunt splenic ruptures, splenorrhaphy is appropriate only in hemodynamically stable patients with few other organ injuries in the abdomen.

Stomach

A large blunt rupture of the stomach is a rare entity occurring secondary to compression after a meal, while most gastric perforations follow penetrating wounds and are small. One or two-layer closure with absorbable or nonabsorbable suture material is appropriate, and the ‘rule of two's' (entrance and exit wounds) should always be assumed when a gunshot wound is the mechanism of injury. An occult second perforation at the gastroesophageal junction or on the lesser or greater curvature is easily diagnosed by filling the stomach with a mixture of one ampule of methylene blue dye in 200 to 300 ml saline while a noncrushing intestinal clamp is placed across the pylorus.

Duodenum

Retroperitoneal air outlining the C-loop of the duodenum in the right upper quadrant on a flat plate radiograph of the abdomen is pathognomonic of blunt rupture (Fig. 11). Even if this finding is not present, a patient with a direct blow to the upper abdomen (handlebars in a child/lower rim of the steering wheel in an adult) should have a CT with oral contrast or a standard gastrografin upper gastrointestinal series to evaluate the duodenum. A patient with a blunt intramural hematoma (‘coiled spring sign') is managed nonoperatively.



Fig. 11. Retroperitoneal air in right upper quadrant outlining C-loop of duodenum is pathognomonic of rupture.

At operation for blunt trauma, the presence of retroperitoneal crepitus or bile-staining is diagnostic of rupture. In patients with penetrating wounds in the upper abdomen, a track of a knife or gunshot wound should always be followed to see if there is injury to any portion of the duodenum.

A large blunt rupture in the second or third portion or a penetrating wound in any portion that does not involve the papilla is closed with two layers of sutures in a transverse direction. A viable omental pedicle is sutured over the repair, and, with larger injuries, a closed suction drain is left in the right upper retroperitoneum in proximity to the repair for 7 to 10 days. When there has been a delayed diagnosis of a duodenal injury, when repair results in significant narrowing or appears tenuous, or when there is a combined pancreatoduodenal injury, some type of diversion procedure is appropriate. Pyloric exclusion with gastrojejunostomy is the easiest technical approach and has been the most frequently utilized in major trauma centers. Through a dependent gastrotomy, the pyloric muscle ring (not the prepyloric antral mucosa) is closed with a continuous 1 polypropylene suture. An antecolic gastrojejunostomy completes the procedure (Fig. 12). Upper gastrointestinal radiographic series performed 10 to 21 days after pyloric exclusion document that 95 per cent will open within 2 to 3 weeks despite suturing. Should a leak from the duodenal repair occur in the interim (incidence 2 to 6 per cent), this will be an end rather than side fistula until the exclusion opens. Alternate approaches include truncal vasotomy, antrectomy, and gastrojejunostomy (‘duodenal diverticulization') or gastrostomy, retrograde duodenostomy, and antegrade jejunostomy (‘triple tube'). Deaths directly related to postoperative duodenal leaks occur in 1 to 4 per cent of patients in large clinical series.



Fig. 12. Pyloric exclusion by internal suture and an antecolic gastrojejunostomy are used in patients with a delayed diagnosis of a duodenal injury, a tenuous duodenal repair, or a combined pancreatoduodenal injury.

Pancreas

CT remains the diagnostic method of choice in patients with blunt abdominal trauma, and nonoperative management is chosen for AAST OIS Grade I or II pancreatic injuries. A suggestion of transection on the CT mandates a celiotomy. When the CT is equivocal and there is persistent hyperamylasemia in the first 48 h after injury, an ERCP is indicated.

Complete operative exposure of the pancreas in patients with either blunt or penetrating trauma includes a Kocher maneuver, division of the gastrocolic omentum, and, on occasion, division of the retroperitoneum along the inferior edge of the gland. Parenchymal injuries not involving the duct are filled with a pedicle of viable omentum, and a closed suction drain is left in proximity to this area for 7 to 10 days by many surgeons. Blunt transections near the neck are usually obvious once the overlying hematoma and mesothelial capsule of the gland have been removed. When a penetrating wound is near the duct, the presence of extensive fat necrosis in the lesser sac, clear fluid draining from the perforation, or the appearance of clear fluid drainage after the administration of intravenous secretin (2 mg/kg) is strongly suggestive of a ductal injury. Ductal injuries to the left of the mesenteric vessels are managed with distal pancreatectomy and splenectomy in adults (Fig. 13). In many centers, the stump is closed with 4.8 mm staples, covered with a viable pedicle of omentum, and a closed suction drain left in proximity to the stump for 7 to 10 days. On rare occasions in stable patients, transections over the superior mesenteric vein or to the right are managed with distal Roux-en-Y pancreatojejunostomy and oversewing of the proximal segment at the point of transection. A pancreatoduodenectomy is indicated only when there has been maceration or devascularization of the pancreatoduodenal complex and should only be performed in normothermic patients without a significant base deficit or an intraoperative coagulopathy.

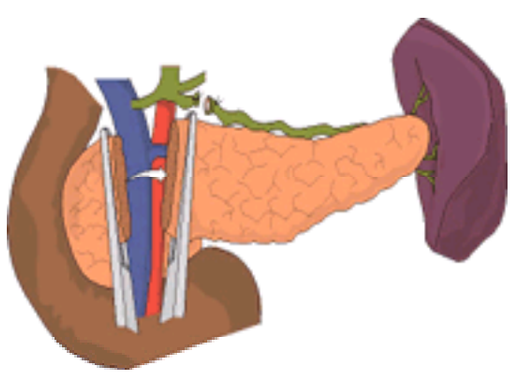


Fig. 13. Pancreatic ductal injuries to the left of the mesenteric vessels are managed with distal pancreatectomy and splenectomy in adults.

Postoperative pancreatic fistulae (some elevation of amylase level in drainage) occurs in 25 to 33 per cent of patients after major resections. These are treated with drainage, subcutaneous injections of somatostatin-analog, and jejunal feedings. The mortality from pancreatic injuries is often related to associated upper abdominal injuries, but ranges from 3 to 5 per cent for stab wounds to 15 to 20 per cent for blunt trauma or gunshot wounds.

Small bowel

Blunt rupture of the small bowel occurs near the ligament of Trietz, near the ileocecal valve, or at points of adhesions. CT with oral contrast has missed small injuries in 7 to 12 per cent of patients in large series, and serial physical examinations are necessary for 24 h after a ‘negative' study. Injuries to the small bowel are found in 30 and 50 per cent of patients undergoing celiotomy for stab or gunshot wounds, respectively.

Adjacent perforations from stab or gunshot wounds are connected and closed in a transverse direction in one- or two-layers. Large mesenteric hematomas, large areas of deserosalized bowel, or jejunal segments with large numbers of perforations (i.e. shotgun wounds) are resected, with a reanastomosis performed with sutures or staples (Fig. 14). The postoperative leak rate after repair of the small bowel is approximately 1 per cent.

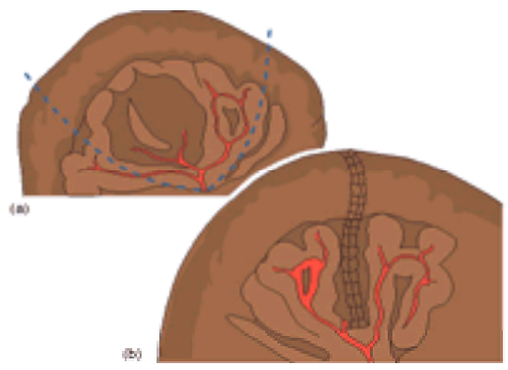


Fig. 14. Large mesenteric hematoma with compromised blood supply to the overlying small bowel is treated with resection and reanastomosis. (Copyright, Baylor College of Medicine, 1983 and used by permission.)

Colon

The repair of colonic perforations using one- or two-layers of sutures or resection of the right colon with immediate ileocolostomy is now performed in 90 to 95 per cent of patients with blunt or penetrating injuries. Resection of the transverse colon or left colon with immediate colocolostomy without an intraoperative bowel prep is performed in some centers, but remains controversial. Resection of the injured transverse or left colon segment with an end colostomy is utilized in the 5 to 10 per cent of patients with large injuries and significant fecal contamination.

The fecal fistula rate for patients undergoing primary repair (excluding colocolostomy) or resection with end colostomy is 1 to 2 per cent. With most incisions being packed open, the wound infection rate is 4 to 6 per cent. Postoperative abdominal abscesses and mortality are both 5 to 6 per cent in patients with primary repairs. In

patients undergoing resection with end colostomy, postoperative abscesses and mortality are both 15 to 17 per cent.

Rectum

Extraperitoneal gunshot wounds in which the perforation is deep in the narrow male pelvis are not repaired. Management includes proximal loop sigmoid colostomy, manual evacuation of stool from the rectum, and insertion of a 1-inch Penrose drain into the deep pelvis through a presacral incision. The excellent blood supply of the rectum leads to healing of the unrepaired rectal perforation in more than 50 per cent of patients within 7 to 10 days. This allows for same-admission colostomy closure in those patients without contraindications to another general anesthetic.

Trauma damage control

Patients with shock from exsanguination related to abdominal injuries and massive transfusion develop pre- or intraoperative hypothermia, persistent metabolic acidosis, and a coagulopathy. This so-called ‘vicious cycle of metabolic failure' is irreversible as long as the patient is in the operating room with the abdomen open during a prolonged procedure. Pre- or intraoperative markers that confirm ‘metabolic failure' and suggest that a ‘damage control' operation should be performed instead of a definitive laparotomy are listed in Table 14. The concept of ‘damage control' operations has now been used for over 15 years and includes three phases (Table 15). Alternate closures of the abdominal incision or coverage of the open abdomen are used in most patients undergoing a ‘damage control' operation (Table 16). Prospective randomized trials to confirm the value of ‘damage control' are unlikely to be performed as the concept is now widely accepted. Also, it appears to allow for the salvage of severely injured patients who died when prolonged definitive first celiotomies were performed.

Factor	Level
1. Initial body temperature	<35°C
2. Initial acid-base status	
• Arterial pH	<7.2
• Base deficit	<−8 mmol/l in patients <55 years of age or <−6 mmol/l in patients >55 years of age
• Arterial lactate	>5 mmol/l
3. Cause of coagulopathy	Prothrombin time and/or partial thromboplastin time >50 per cent of normal

Modified from Rotondo MF, Zonies DH. Damage control: an approach for improved survival in exsanguinating penetrating abdominal injury. J Trauma 1997; 43: 1062–70.

Table 14 Intraoperative indications to perform damage control operations

Sorry, this figure/table has not been included in electronic versions of the Oxford Textbook of Surgery 2nd Edition for copyright reasons.

Table 15 Three phases of ‘damage control' celiotomy

Skin closure only
Towel clips 1/2–3/4 inches apart
Continuous suture
Temporary silos (with or without zipper)
Steri—Drape*
Silastic sheet[b]
Genitourinary irrigation bag
X-ray cassette bag
Absorbable mesh for conversion to open abdomen
Polyglactin mesh
Polyglycolic mesh
Late reconstruction of incisional hernia
Marlex mesh[c]
Polytetrafluoroethylene body wall patch[d]

*3M, St. Paul, MN.
[b]Dow Corning Corp., Midland, MI
[c]Bard, Cranston, RI.
[d]W.L. Gore, Flagstaff, AZ.

Table 16 Alternate techniques of incisional closure at completion of ‘damage control' celiotomy

Shock

Introduction

Shock is best defined as a deficit in oxygen delivery to the mitochondria of cells. It is not a single phenomenon but rather a group of syndromes, each characterized by a particular physiologic dysfunction that must be treated (Table 17). Restoring intravascular volume and red cell mass, pharmacologic support of cardiac function, and adjustments in systemic vascular resistance enable physicians to resuscitate patients in hypovolemic shock from trauma, cardiogenic shock from a myocardial infarct, and septic shock from a bacteremia. Reversal of shock may not be enough, however, to achieve recovery in all patients. Reperfusion after critical periods of ischemia often leads to an inflammatory response. Endogenous inflammatory events are essential for repair of damaged tissues and recovery of normal homeostasis, but can become exaggerated following resuscitation from shock and cause organ failure when directed toward normal tissues. Infection and multiple organ failure are secondary insults which have been theoretically linked to the duration and intensity of the initial shock insult. Therapy may even be required after the initial reversal of shock to ameliorate the inflammatory consequences of such reperfusion.

- The first minute: immediately identify cause and palliate
- The first hour: select therapy which optimally restores perfusion to cells
- The first days: modulate inflammatory mediators amplified during reperfusion
- The first weeks: sustain and nurture during recovery and repair

Table 17 Strategy for successful treatment of the injured patient in shock

Pathophysiology of shock

Cellular bioenergetic failure

Physicians evaluate and treat patients in shock guided by organ function; however, the extent of disruption in cellular metabolism determines the magnitude of risk to the patient. There is a deficit in oxygen delivery to the mitochondria of cells in shock, and, consequently, the rate of aerobic metabolism fails to meet the patient's energy needs. In severe shock, cells approach bioenergetic failure. Cellular dysfunction in turn causes organ failure and, in profound shock, quickly leads to the patient's death. The principal source of cellular energy is oxidative phosphorylation which occurs in mitochondria. Electrons are extracted from molecular oxygen during this process, and the chemical energy released is directed into higher-energy chemical bonds, most commonly the terminal phosphate in adenosine triphosphate (ATP). The energy in ATP is released during hydrolysis of the terminal phosphate bond with production of adenosine diphosphate (ADP). Examples of cellular work dependent upon ATP include biochemical synthesis, muscle contraction and movement, transport of chemicals across cellular membranes, and processes intrinsic to the inflammatory response. There must be a continuous conversion of ADP back to ATP via aerobic metabolism, and this process of remaking biochemical energy must occur at a rate sufficient to sustain metabolism. In shock, the rate-limiting factor for ATP resynthesis is a molecular oxygen deficit.

Without adequate oxygen, glucose can be metabolized in cells by anaerobic means. The rate of ATP synthesis is meager, however, and anaerobic production is useful only as an interim means of sustaining minimal cellular work during a period of mitochondrial hypoxia. In aerobic circumstances, each mole of glucose completely oxidized to carbon dioxide within mitochondria generates 36 to 38 moles of ATP. Anaerobic metabolism of one mole of glucose produces only two moles of ATP. The severity of cellular bioenergetic stress during shock is proportional to the increase in ADP levels. Following successful treatment of shock, patients have oxygen delivery restored, and recovery corresponds to repletion of cellular ATP levels (Table 18).

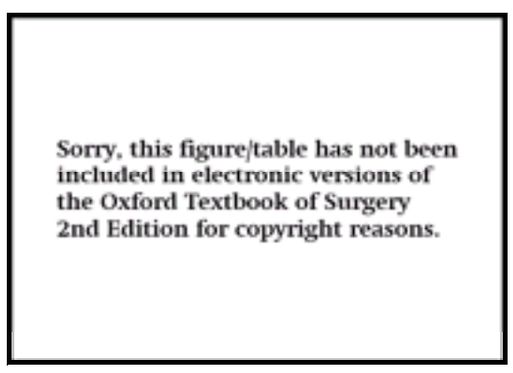


Table 18 ATP-ADP cycle

It is appropriate for the treating physician to monitor organ function and blood biochemical tests which indicate successful restoration of oxidative phosphorylation and sufficient ATP to sustain cellular work.

Acidemia provides an index of severity of shock

The proton (H^+) is central to energy metabolism. During oxidative phosphorylation, ADP is chemically bonded to a proton and an organic phosphate to produce ATP in the mitochondria. During release of the high-energy bond of ATP, a proton is released. When aerobic metabolism is not sustained, protons accumulate, and patients develop acidosis in proportion to the severity of shock. Excess H^+ is efficiently buffered by several molecular species in the cytosol when pH is 7.40. As more protons accumulate in patients with shock and buffer is consumed, there can be an accelerated decline in arterial pH. Interpretation of blood pH is complicated due to the independent influence of $PaCO_2$. In patients with excess H^+, increased alveolar ventilation can cause a compensatory increase in pH. A more clinically useful measure for evaluating excess H^+ in patients with shock is the calculated value of base deficit in a sample of blood. The calculated base deficit is the amount of bicarbonate (as mmol/l) required to correct the patient's blood pH after adjusting the $PaCO_2$ to a normal value of 40 mmHg. In the medical literature, there are differences in the mathematical sign used when reporting base deficit. For this chapter base deficit is described as a positive number; in the circumstance of excess bicarbonate, a '+' sign is used.

A treating physician can classify a patient's risk for death and development of multiple organ failure based upon base deficits. In one study, injured patients whose base deficits were between 6 and 9 had a 23 per cent mortality, while those with base deficits exceeding 10 had a 49 per cent mortality (Table 19). Others have reported on trends in base deficit as a guide to resuscitation. Organ failure and death are more frequent among patients who fail to correct a base deficit within 24 h. These studies may be interpreted to indicate that physicians should pursue additional resuscitative or corrective actions in trauma patients who present in shock and do not respond to resuscitation, as evidenced by a failure to correct base deficit within the first 24 h. A persistent base deficit indicates continued failure to deliver sufficient oxygen to mitochondria for sustaining aerobic metabolism at the required rate. Monitoring base deficit as an index of resuscitation may provide the only indication of incomplete resuscitation from shock with the attendant risk of death from organ failure.

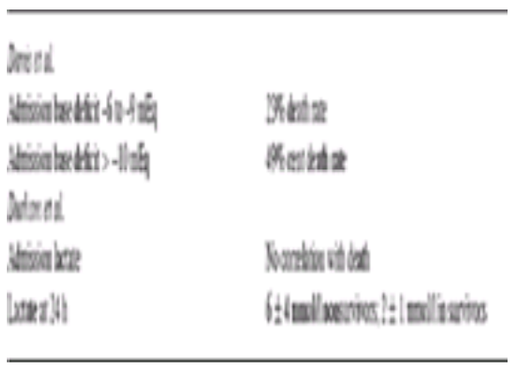


Table 19 Predictive value of acidemia in injured patients in shock

Several studies have reported that serum lactate concentration is equivalent or superior to base deficit for providing the treating physician an indication of the magnitude of shock. Glycolysis is the biochemical sequence by which glucose is converted to pyruvate, and these biochemical processes support cellular bioenergetics by generating two moles of ATP per mole of glucose consumed. Patients in shock accumulate pyruvate because aerobic metabolism is depressed, and pyruvate cannot proceed through the tricarboxylic acid cycle for oxidation. As pyruvate concentration in the cytosol increases, the enzyme lactate dehydrogenase generates lactate from pyruvate at an increasing rate. Anionic lactate buffers the excess H^+ which accumulates in shock, and thus a measurable lactic acidemia develops. The liver and kidney are two organs capable of utilizing blood lactate and converting it back into glucose. An alternate form of lactate metabolism occurs in the liver, where lactate can be oxidized to CO_2 and water. Patients in shock have serum lactate levels above the normal value of 2 mmol/l. Lactate elevations in the range of 4 to 5 mmol/l are common among patients resuscitated from shock associated with multiple injuries, and longer delays in returning lactate levels to normal correspond with a greater risk of death. The value of serial lactate levels as an indication of resuscitation has not been proven superior to serial monitoring of base deficit. Many clinicians

prefer using base deficit correction as an indicator of resuscitation because it is readily available from arterial blood gas analysis.

Septic shock and accelerated glycolysis

While scholars widely accept that lactic acidosis is an index of the severity of hypoperfusion in shock states characterized by low cardiac output, there is debate regarding the mechanism of lactic acidemia in patients with septic shock. Septic shock is commonly associated with a high flow pattern of hemodynamic response. Inflammatory mediators released by the body in response to bacterial toxins cause a reduction in systemic vascular resistance, and hypotensive patients with invasive infection may have cardiac outputs two- to threefold greater than normal. These patients follow a characteristic pattern of high mixed venous oxygen saturation, progressive base deficit, and elevation of lactate. One possible explanation for this pattern holds that vasodilation in peripheral microcirculations produces shunts, and blood flows from artery to vein without contact with the microcirculation to allow oxygen diffusion to mitochondria. A second theory postulates that alterations in glucose metabolism cause a several-fold increase in the rate of glycolysis. This results in excessive synthesis of protons and pyruvate, and pyruvate is shifted to lactate by the enzyme pyruvate dehydrogenase. Studies of septic patients using isotopes and calculated rates of clearance indicate that the lactic acidemia of sepsis relates to accelerated glycolysis within the cytosol and excessive pyruvate production, which is then cleared into the extracellular fluid as lactic acid. This concept is supported by observations in patients whose hemodynamic measurements deteriorated until they died. Future therapy for acidemia of sepsis may depend upon joint efforts to improve substrate delivery, along with manipulation of the cellular bioenergetic pathways.

‘Supranormal' oxygen delivery after shock

When oxygen delivery to mitochondria is insufficient to sustain normal cytosol levels of ATP during shock, some investigators suggest that an oxygen debt develops. This hypothesis also implies that following restoration of normal blood flow, there must be a period of supranormal oxygen delivery; i.e., not only can aerobic metabolism be sustained at a normal rate, but supplemental oxygen is also available to support the recovery of ATP levels depressed during the period of shock. Evidence supporting the benefits of supranormal oxygen delivery in critically ill trauma patients includes the observation that if oxygen delivery is increased, oxygen consumption also increases. In one study, critically ill patients were randomized into either a control group in which cardiac index values were increased to a proscribed level or a group in which patients were treated to maintain a mixed venous oxygen saturation exceeding 70 per cent. The authors reported successful achievement of the proscribed goals in most patients. There was no evidence of improved survival in the entire study cohort, nor among the subgroup noted to have multiple trauma with massive blood loss. Whether critically ill patients actually benefit from supranormal oxygen delivery has yet to be resolved. Others have suggested that using gastric tonometry as an index of splanchnic circulation provides an indication of satisfactory resuscitation and eventual survival superior to more global indices of underperfusion.

Hypovolemic shock

Hypotensive injured patients commonly have a depleted intravascular volume, which is either the sole cause or a major contributing factor to their shock. Hypovolemia is most commonly related to blood loss while fluid and protein shifts from the plasma compartment as inflammatory exudate also contribute to the intravascular volume deficit. Treatment of hypovolemic shock aims to restore normal intravascular volume, but the restoration of intravascular deficits without control of hemorrhage in trauma patients is not only futile but may prove dangerous.

Pathophysiology

The blood volume makes up 7 to 8 per cent of the body weight, and approximately two-thirds is located in the venous side of the circulation. Multiple physiologic mechanisms enable compensation for substantial hemorrhage and temporary maintenance of perfusion pressure to vital organs such as the brain and heart. A 70-kg individual in normal health with 5000 ml of intravascular blood volume can sustain a blood loss of 1000 ml with minimal change in blood pressure. Endocrine and neurologic reflexes (pain, fear, and hypovolemia of injury) mediate a surge in epinephrine release from the adrenal medulla via central nervous system reflexes to compensate for the loss of intravascular volume. In response to a decline in blood pressure, baroreceptors mediate increased tone in the sympathetic nerves. Accelerated release of norepinephrine from postganglionic terminals vasoconstricts veins, arterioles, and larger arteries. Constriction of venous capacitance vessels reduces the size of the venous compartment while sustaining the venous pressure returning blood to the right atrium. Catecholamines also stimulate increased myocardial contractility to sustain cardiac output when end diastolic volume is reduced in the ventricular chambers. Measurements in normal humans following experimental withdrawal of 500 to 1000 ml of blood demonstrate multiple compensatory reflexes which can be very effective up to the point of a critical deficit. When the loss exceeds the critical deficit, precipitous and profound hypotension, including syncope, can result.

Shifts of fluid and protein from the interstitial compartment to the intravascular space also compensate for the hypovolemia of shock. While fluid shifts are slower mechanisms of compensation for hypovolemia, they lead to sustained expansion of intravascular volume, decrease in hematocrit, reduced vasoconstriction, and improved blood flow to tissue. The shift of fluid from the interstitium to the intravascular compartment occurs because lymph flow through the thoracic duct continues while the driving forces for interstitial fluid replacement decline in the capillary. At the end of a period of hypovolemic shock, the interstitial fluid space is contracted. During the recovery phase from shock, experimental evidence indicates there must be a net gain in interstitial fluid to expand the interstitial space. This increase in interstitial fluid is essential in the period following control of the hemorrhage which caused the shock and has been designated as a ‘fluid uptake phase.' Without the fluid uptake during this second phase of resuscitation, plasma volume remains depleted, and organ failure occurs due to underperfusion. The universal sensation of severe thirst reported by injured patients with blood loss indicates how imbibing water to restore blood volume is a strong reflex which, prior to the interventions in intensive care units over the past 50 years, played a major role in restoring depleted intravascular volume. Research has established the normal course of uncompensated shock in models of hemorrhage; however, in most patients treated for serious injury at a trauma center, early institution of fluid therapy alters the pathophysiology. The post-traumatic fluid uptake phase should seldom exceed 48 h and should be followed by a period of spontaneous diuresis to restore normal interstitial and intravascular volumes.

Diagnosis

Hypotension is a key finding when determining whether a patient is in hemorrhagic shock. Tachycardia, commonly associated with shock in injured patients, is neither a sensitive nor specific indicator of shock, and many injured patients without blood loss are tachycardic. This is because increased vagal tone in patients with severe hemorrhage can cause heart rates to remain below 100 beats per min. A decrease in capillary refill, diaphoresis, and a weak, difficult to palpate femoral or carotid artery pulse are useful clinical indicators of shock when evaluating an injured patient. Hypotension in an injured patient may be transient and respond to the infusion of intravenous fluids. Confounding the clinical evaluation of shock is the variation in physiologic response related to age and medical conditions. The elderly are more sensitive to blood loss, in part because many are on medications which alter the normal physiologic reflexes. Evaluation of infants and children can be confusing because they have extraordinary compensatory mechanisms. The pregnant patient suffering blood loss from injury can herself have adequate perfusion while the fetus is subjected to an ischemic insult. The timely diagnosis of shock in injured patients often depends upon repeated evaluations for signs of hypoperfusion.

A useful component of the clinical evaluation of patients in hypovolemic shock is the quick determination whether evidence exists of ‘significant' blood loss. A patient with even a small skin wound can lose a substantial volume of blood at the scene. Puncture of a major arterial or venous vessel can lead to rapid blood loss until the patient's hypotension and coagulation mechanisms stop the bleeding. When the patient arrives in a hospital for treatment, active hemorrhage may have ceased. A report of a large pool of blood at the scene or blood-soaked clothing on the victim is strongly suggestive of prior exsanguinating hemorrhage. While complete division of major vessels in an extremity during a traumatic amputation is usually followed by retraction and occlusion of the vessel, lacerated or perforated major vessels will continue to bleed or precipitously re-bleed when resuscitation restores perfusion. Direct occlusion of the bleeding site is usually effective. A proximally placed blood pressure cuff functioning as a tourniquet may be required when there is substantial tissue destruction. Scalp and facial hemorrhage can cause shock, particularly in infants and children. To control hemorrhage from the scalp, face, or oral mucosa, infiltration with topical anesthetic agents containing epinephrine is appropriate. Vasoconstriction may stop the bleeding, and the wound can be quickly and temporarily closed with skin staples or a continuous suture until evaluation of the patient is complete. After the patient has stabilized, the wound can be reopened and managed in a more detailed fashion at a deliberate pace. Active hemorrhage from the nares or mouth, particularly when related to a basilar skull fracture, is difficult to control and may lead to asphyxiation or exsanguination. Packing of the oral cavity or nasal vestibule may be required, and consideration should be given to early use of an angiographic study with embolization of the bleeding site.

Massive hemorrhage from major vessels and the heart is usually immediately lethal within a few heartbeats and is rarely a clinical problem that can be treated at the hospital. Patients with a hemopneumothorax and bleeding from the lung, intercostal vessels, and through the diaphragm from a intra-abdominal source are initially managed with insertion of a large diameter chest tube. In adults, a 38 French chest tube is inserted in the posterior aspect of the pleural cavity. Initial blood loss of greater than 1200 to 1500 ml or continuing hemorrhage of more than 100 to 200 ml/h from a chest tube is an indication for thoracotomy. Exceptions to that guideline include those patients with transdiaphragmatic wounds whose bleeding sites reside in the abdomen.

In an adult, over a liter of blood can hemorrhage into the thigh from a fracture of the femur. This degree of hemorrhage can occur without injury to the superficial femoral artery or vein, with blood loss occurring from disruption of smaller vessels and the bone marrow. The capacity of the thigh to expand into a distended cylinder is greater if the ends of the bone override one another and the thigh foreshortens. Thus, there is hemostatic value to placing the leg in a traction splint to realign the bone and return the thigh to a normal configuration. Also, immobilization of the fracture reduces the risk of injury to adjacent vessels from shards of bone. Compound fractures can decompress the hematoma in the thigh, but will exacerbate blood loss. Multiple extremity fractures can also be associated with sufficient blood loss to

produce shock. The clinical diagnosis of fracture of the long bone is simple and should prompt the surgeon to provide additional resuscitation fluid in proportion to the number of fractures and size of associated hematomas.

Fractures of the pelvis can cause life-threatening hemorrhage. For example, 10 to 20 per cent of patients with pelvic fractures require transfusion of more than five units of blood in the first 24 h after injury. Fortunately, less than 3 per cent of patients with pelvic fractures have profound hemorrhage, present in shock, and are at risk for immediate exsanguination. Patients at great risk for shock are those with an open wound which decompresses the pelvic hematoma. Of interest, 10 to 40 per cent of patients with pelvic fractures, severe blood loss, and shock suffer hemorrhage from another site, particularly in the abdomen. Thus, laparotomy to control bleeding from the spleen or liver is essential to effective treatment of shock. Finally, an ominous characteristic of a pelvic fracture on an admission radiograph is a measurable gap at the fracture of greater than 0.5 cm.

There is controversy regarding which interventions are effective for controlling hemorrhage from pelvic fractures and reversing hemorrhagic shock. Most patients will stop bleeding spontaneously, and shock is avoided by appropriate blood transfusion to restore deficits. In the medical literature over the past 10 years, advocates have reported that external compression of the pelvis with pneumatic anti-shock trousers can arrest hemorrhage. Orthopaedic surgeons have advocated external frames attached to screws set in the iliac crest as devices which reduce pelvic volume, compress the hematoma, and arrest hemorrhage. A directly effective hemostatic intervention which will stop bleeding in the subset of patients with life-threatening arterial hemorrhage is transcatheter embolization of the bleeding site.

Following blunt trauma, injury to liver, spleen, or mesenteric vessels is the most common source of intra-abdominal hemorrhage that causes shock. After penetrating intra-abdominal trauma, perforation of any large vessel or group of vessels can lead to rapid and lethal blood loss. Rapid identification of a hemoperitoneum in a patient with complex multiple injuries frequently mandates use of diagnostic peritoneal lavage or surgeon-performed ultrasound. Diagnostic peritoneal lavage has high sensitivity for correctable bleeding in a hypotensive injured patient when gross blood spontaneously drains from the catheter as it is inserted into the peritoneal cavity. Abdominal ultrasound can also provide immediate assessment of the presence of fluid in the peritoneal cavity and, in competent hands, is preferable to diagnostic peritoneal lavage as a noninvasive, quickly performed test. More stable patients with a hemoperitoneum can be managed nonoperatively using a follow-up abdominal CT to assess the magnitude of visceral injuries.

Venous access

Therapy for patients in hemorrhagic shock depends upon intravenous access for the infusion of fluid and blood to restore intravascular volume. When readily available, veins in the antecubital fossa which accommodate large bore catheters are preferable access sites. When injury to the arm or profound vasoconstriction precludes use of veins in the arm or hand, alternative sites should be selected (Table 20). Central venous access into the femoral vein has proven very satisfactory for initial resuscitation, even in patients with abdominal injury. Access via the femoral vein carries no risk of pneumothorax, and this vessel is away from the upper torso where management of the airway and breathing are often being carried out during the initial period when venous access is emergently needed. The preferred catheters for femoral or subclavian central venous access are the 8.5 French size ‘introducers' used for insertion of pulmonary artery catheters. To insert these large catheters, access to the vein is first accomplished with a small catheter, a guidewire is passed and the introducer is inserted over the wire into a large vein. Femoral vein catheterization does increase the risk of venous thrombosis, although the risk seems to be very low in the first 24 h following insertion. The use of subclavian venous access is enthusiastically supported by some whose expertise enables them to insert these catheters with a minimum number of complications. Pneumothorax and misplacement of the catheter are serious complications which make subclavian venous access less desirable in unstable patients. A cutdown on the saphenous vein at the ankle is a reliable method of establishing venous cannulation with a large catheter even in patients with a collapsed peripheral circulation. Intraosseous infusion into the marrow of the tibia is useful in toddlers and infants in whom venous cannulation can be very difficult.

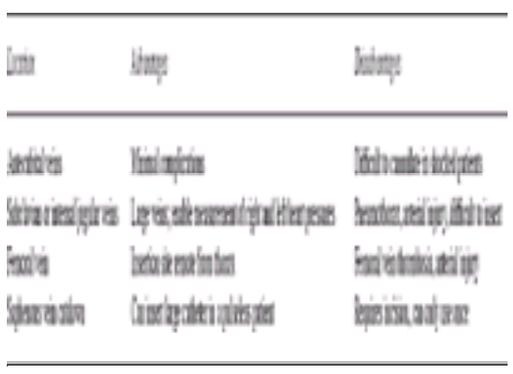


Table 20 Principal sites of venous access in injured patients in shock

Use of isotonic fluid or blood

The preferred fluid for the initial resuscitation of injured patients in shock has been debated over the past 40 years. Advocates of isotonic saline recommend the use of large volumes of balanced electrolyte solutions, while proponents who prefer colloid solutions argue that colloids sustain oncotic pressure and less infusate is required to achieve an equivalent expansion of plasma volume. Randomized control trials in trauma patients treated for hemorrhagic shock have established that balanced electrolyte solutions are as effective or superior to isotonic solutions that contain a colloid. Isotonic saline replacement after hemorrhage lowers the plasma colloid osmotic pressure, and most of the infusion fluid is transferred to the interstitial space. Immediately following resuscitation with isotonic saline for hemorrhagic shock, patients will have edema apparent on physical examination. While edema of the skin, muscle, and bowel is common, these patients rarely develop hydrostatic pulmonary edema characterized by copious endotracheal fluid. A portion of the excess sodium and water required by patients to maintain vital organ perfusion during the phase of fluid uptake following resuscitation from hemorrhagic shock may be intracellular and indicate the severity of depletion of intracellular ATP. It is clear that substantial crystalloid infusions have reduced the prevalence of renal dysfunction and failure among injured patients.

Patients in severe hemorrhagic shock require blood transfusion to avoid profound hemodilution. The hematocrit threshold to guide transfusion decisions in trauma patients resuscitated from hemorrhagic shock has not yet been determined. In one study, increasing the hematocrit of injured patients resuscitated from shock from 28 to 37 per cent had no benefit on oxygen consumption. In another randomized, controlled clinical trial conducted in critically ill patients, there was no advantage to transfusing patients to keep their hemoglobin above 10 g/dl over a policy of accepting a lower hemoglobin between 7 and 9 g/dl. Clearly, there is great value to transfusing blood into patients with shock from a reduction of more than 25 per cent of their intravascular volume. It is less clear whether hematocrit should be considered a trigger which prompts the physician to order a blood transfusion in a patient with normal intravascular volume.

Patients with profound blood loss require red cell transfusion to restore hemoglobin. These patients are preferably transfused with blood that has been typed and cross-matched but, in circumstances of impending lethal exsanguination, the transfusion of type O blood has minimal risk. Girls or women of childbearing age should be given O-negative blood to avoid Rh sensitization. The rapid infusion of blood does have potential complications, especially hypothermia. When body temperature is less than 34°C, the risk of coagulopathy and death increases. In circumstances where stored units of blood are rapidly infused, it is preferable to deliver them through warming devices. The other concern is that restoring intravascular volume to normal while the patient still has active hemorrhage may have adverse effects. It has been suggested that the alternate tactic for management of the hypotensive trauma patient with penetrating torso injuries is withholding transfusion until the operation to control hemorrhage has begun. A more appropriate approach is minimizing time in the emergency center and performing laparotomy within 10 min of arrival when hypotension is present. Hypotension is an independent risk factor for adverse outcome among patients with brain injury, and the role of deferring restoration of perfusion pressure in patients with blunt trauma and shock prior to definitive control of hemorrhage has not been determined. The transfusion of type-specific blood into a hypotensive injured patient is appropriate as a method of restoring intravascular volume, although, at the same time, the trauma surgeon must determine if the patient requires an intervention to terminate persistent hemorrhage.

An acquired coagulopathy is common after transfusion of more than 10 units of blood due to diminished coagulation factors in transfused red cells and dilution secondary to crystalloid infusions. Patients with hemorrhagic shock and multiple bleeding sites develop rapid consumption of coagulation factors by their wounds. The surgeon operating upon the coagulopathic patient may note a diffuse oozing in the wound which will not stop. Reversal of hypothermia and replacement of coagulation factors in patients receiving a massive transfusion is usually required if wound bleeding is to stop. When oozing is imminent or recurring, ‘damage control' procedures are appropriate.

Damage control (Table 14, Table 15, Table 16)

As previously noted a tactical solution to the problems of hypothermia, acidemia and coagulopathy in seriously injured patients undergoing surgery has been termed

‘damage control'. The conceptof damage control is that selected patients will tolerate better a series of abbreviated operative procedures to achieve complete repair of all injuries rather than a single prolonged operation. Thus, in patients with severe abdominal injuries, the first operation should accomplish control of major bleeding sites, insertion of shunts or major vascular repairs required to save life or limb, and closure of bowel perforations. Following the principles of damage control, sites of oozing are packed with gauze to achieve tamponade. The abdominal incision is then closed relatively quickly using towel clips or suture on the skin edges or the insertion of a plastic silo to avoid the adverse consequences of the abdominal compartment syndrome. Patients managed by damage control are transferred to an intensive care unit where treatment emphasizes correction of hypothermia and coagulation deficits and restoration of intravascular volume. Critically ill patients whose major hemorrhage has been stopped will usually improve over 24 h before the surgeon returns the patient to the operating room. By deferring definitive repair of injuries, these precarious patients are given an opportunity to stabilize following hemorrhagic shock, and subsequent repairs can be performed in tissues which are no longer ischemic. Evidence from several clinical series has established that ‘damage control' is a prudent alternative to attempting one definitive operation.

Systemic inflammatory response syndrome (SIRS) and septic shock

Invasive infection and sepsis are principal causes of late death among hospitalized injured patients. Events at the time of injury, however, can increase the risk of subsequent sepsis. Prolonged and profound hemorrhagic shock may contribute to later infection due to suppression of immunologic function. Evidence from experimental studies suggests that blood transfusions are immunosuppressive, as well. Dilution of proteins and cells normally essential in endogenous responses to infection can reduce the capacity to resist infection, particularly in ischemic wounds. Finally, improper management of the patient's injuries or missing an injury during the initial phase of resuscitation and therapy may initiate fatal sepsis.

Patients in septic shock from invasive infection have profound vasodilation, a cardiac index several-fold greater than normal, tachycardia, and hypotension with a wide pulse pressure. This hyperdynamic syndrome is typically seen when patients have had adequate infusion of fluid to expand the intravascular compartment. Septic shock can be attributed to the magnitude of the microbiologic challenge and the intensity of the patient's inflammatory response. Endotoxin, a component of the wall of Gram-negative bacteria, has been demonstrated to initiate hormonal and cellular responses on normal human volunteers similar to those that occur in patients with septic shock. The inflammatory response to invasive infection may simultaneously represent a benefit to the patient as endotoxemia is cleared from the body and invading bacteria are destroyed and a hinderance as the excessive inflammatory response has been hypothesized to account for the vasodilation of septic shock. Modifying the exaggerated inflammatory chain of events has become a major focus of research intended to develop new therapies for injured patients.

Infecting organisms can express antigen molecules on cell membranes or release soluble antigen molecules. Macrophages respond to these antigens in an intermediate step, characterized by amplification, and are part of an orchestrated cellular immune response of lymphocytes, phagocytic cells, microvascular endothelium, and other cells. This inflammatory response is directed by a battery of diffusible mediators including cytokines, immunoglobulins, and products of the cyclo-oxygenase and lipoxygenase pathways of arachidonic acid metabolism. Very high concentrations of diffusible mediators are associated with hemodynamic instability caused by vasodilation.

Clinical observations of patients suffering serious infection have been used to categorize populations based upon the intensity of their inflammatory responses to infection. Four levels of response have been defined by experts. Among patients with the systemic inflammatory response syndrome (SIRS), essential clinical findings include abnormal body temperature, tachycardia, tachypnea, and an aberrant white blood cell count. Sepsis is the circumstance of a patient with findings consistent with SIRS plus a documented infection. This distinction between SIRS and sepsis underscores that patients can have physical findings consistent with an exaggerated inflammatory response to infection without a demonstrable focus of bacterial, viral, or fungal invasion. Severe sepsis occurs when a patient with sepsis develops hypotension and sustains critical organ dysfunction. Septic shock is the most serious clinical problem and exists if a patient has severe sepsis, is resuscitated appropriately with intravenous fluid infusion, and yet remains in shock. These four classifications of the inflammatory response to infection carry increasing risk of complications, including irreversible shock and a higher prevalence of death. The death rate among patients who meet septic shock criteria is 40 to 50 per cent. These four strata of risk correspond to progressively greater compromises in tissue perfusion. Each escalation in shock is hypothesized to correspond to a more intense autodestructive inflammatory response. While infection is a common source of the inflammatory response syndrome, other insults which activate an exaggerated inflammatory response including pulmonary aspiration, burns involving a large body surface area, and pancreatitis can produce SIRS.

Multiple mediators of the inflammatory response have been identified and characterized. A quality common to the exaggerated response to infection is the circumstance of several concurrent cascades of biochemical events, which interact and amplify one another, producing clinical findings observed in SIRS. Circulating cytokines are released from macrophages, endothelial cells, and granulocytes. These increase permeability of microvascular membranes, cause vasodilation of arterioles, adherence of polymorphonuclear cells to endothelium, and infiltration of the same to the site of infection. There, invading bacteria are destroyed with enzymes and toxic oxygen radicals. Pro-inflammatory cytokines include tumor necrosis factor (TNF) and interleukins (including IL-1, IL-2, IL-6, and IL-8). Anti-inflammatory immune mediators also exist. Endogenously produced circulating receptors for inflammatory mediators and cytokine IL-10 are included in the group of naturally protective agents which suppress the intensity of the inflammatory reaction. Other components of the inflammatory response are linked to the cascade of enzymes which cause thrombosis and fibrinolysis. As in most physiologic response systems, a balance forms between those factors and pathways which amplify the response and those which modulate, blunt, or terminate events. The complex interaction of multiple pathways involved in severe sepsis and septic shock makes it difficult to select a specific agent as the target for a precise therapeutic effect.

The hypotension endangering patients with severe sepsis and septic shock is often the consequence of excessive vasodilation. Hemodynamic measurements in these patients commonly reveal low systemic vascular resistance and a high cardiac output as previously noted, and this pattern is known as the hyperdynamic circulation of sepsis. Despite increased cardiac output, evidence shows that myocardial function is impaired in patients with hyperdynamic septic shock. Initial resuscitation of patients with septic shock should include infusion of intravenous fluids. Persistent hypotension has been managed with variable success by the infusion of drugs, principally adrenergic and dopaminergic agents, to increase myocardial contractility and vasoconstriction. Dopamine, at doses to improve cardiac contractility, can effectively increase the mean arterial pressures of septic patients who remain hypotensive following fluid infusion to restore intravascular volume. Adjusted dose infusion of norepinephrine, which has more a- than b-adrenergic activity, is useful in patients with septic shock because it combines vasoconstriction and inotropic support. Carefully conducted clinical trials confirm that norepinephrine can be used to increase the mean arterial pressure of many patients with septic shock while also maintaining perfusion of vital organs. The optimal therapeutic balance which provides lifesaving support to a patient in septic shock must be individualized. When selecting the type and dosage of hemodynamically active drugs, surgeons resuscitating patients should be guided by evidence of benefit available as hemodynamic data measured from pulmonary artery catheters. Specific endpoints that can be measured include oxygen delivery, individual organ function, and resolution of acidemia in general and lactic acidemia in particular.

Nitric oxide (NO) is an endogenous inflammatory agent that experimental evidence indicates is a principal mediator of the vasodilation which constitutes the pathophysiologic problem of septic shock. NO, a small molecule synthesized in endothelial and inflammatory cells, is released from endothelial cells into the microvascular lumen. It diffuses to vascular smooth muscle cells encircling the endothelium, where it induces guanylate cyclase to synthesize cyclic guanosine monophosphate, a mediator of smooth muscle relaxation. The endothelial cell-produced NO is the vasodilatory counterbalance to vasoconstriction from sympathetic tone, catecholamines, and endothelins. Endothelins are a recently discovered class of endothelial polypeptides that cause vasoconstriction. Thus, NO and endothelin are competing paracrine mediators of microvascular tone.

NO is generally produced in small, rapidly cleared amounts and is quickly converted to inactive nitrites and nitrates. Endogenous nitric oxide synthetase (eNOS) is a closely regulated calcium-dependent enzyme that synthesizes NO from L-arginine and oxygen. Another nitric oxide synthetase uses similar substrates, and this enzyme, induced by inflammatory mediators and synthesized in endothelial and smooth muscle cells, is designated iNOS. This enzyme is considered an isoform of eNOS but is capable of a substantially higher rate of NO synthesis. The rapid and large production of NO by iNOS can cause the refractory vasodilation characteristic of septic shock. Other endogenous agents contributing to iNOS activity include bradykinin and acetylcholine. Of interest, blood nitrate concentrations in septic patients are elevated in proportion to the decrease in systemic vascular resistance.

Pharmacologic interventions that reduce the concentration of nitric oxide in patients with septic shock have been attempted to improve the mean arterial pressure. Hypotensive patients with sepsis who were administered iNOS inhibitors do experience an increase in profoundly depressed systemic vascular resistance; however, evidence of a survival benefit for patients treated with blockade of NO synthesis is not available.

SIRS can occur within hours of injury. In patients experiencing hemorrhagic shock, measurement of serial cytokine levels confirms that mediators of the inflammatory response are elevated in the blood. The magnitude of the inflammatory response is proportional to the severity of the hypotensive insult and may reflect the reperfusion of ischemic endothelium. Observations of human volunteers given endotoxin reveal a transient response consistent with SIRS along with an activated cascade of cytokines. Characteristic of the cascade, the initial cytokine TNF-a does not remain elevated, but induces a sequence of other agents and effects responsible for the inflammatory response. One proposed explanation for delayed-onset multiple organ failure syndromes among trauma patients is termed the ‘second hit' hypothesis. The initial insult from injury, including shock, is speculated to change the biochemical status of circulating inflammatory cells, such as polymorphonuclear cells. These cells become primed or modified to respond with greater intensity to a noxious insult. The second hit hypothesis proposes that an additional insult, such as infection, in a trauma patient several days after injury may induce an exaggerated inflammatory response directed against organs remote from the initial site of injury.

Anti-inflammatory agents

Therapy for patients in shock from sepsis must be directed to accomplish multiple purposes. Patients hypotensive from septic shock should receive intravenous fluids

to expand the intravascular volume. In the hypotensive patient with an adequate intravascular volume an inotropic agent such as dopamine should be given as a constant infusion titrated to achieve an acceptable mean arterial pressure. In severe septic shock, a vasopressor that causes vasoconstriction may be necessary. Appropriate surgical and antimicrobial therapy directed against infection is a critical initial step to reduce the burden of pro-inflammatory activation. The status of anti-inflammatory therapy, however, remains primarily experimental.

Several therapies directed toward blocking the exaggerated inflammatory response of patients in septic shock have been proposed and studied in experimental models. Multiple, randomized controlled trials in humans, however, have failed to provide evidence that a given anti-inflammatory agent will reduce the death rate. Corticosteroids administered in pharmacologic doses to patients with septic shock have not improved survival in prospective studies. Focused anti-inflammatory therapies, including antiendotoxin and anticytokine therapy, have also been attempted in patients with septic shock, but with disappointing results. Several recent analyses of the disappointing results from anti-inflammatory therapy have led to the conclusion that single-agent-based blockade therapy will never be effective. Optimal use of anti-inflammatory treatment may require targeted interventions timed to specific points in the cascade of events following an inflammatory insult.

Precipitous infections noted in trauma patients

Septic shock caused by invasive infection can develop rapidly in acutely injured patients. The most common clinical circumstances which account for septic shock within 48 h of injury are incomplete debridement of dead tissue or failure to recognize and appropriately manage a perforation of the esophagus or bowel. Wounds heavily contaminated with pathogenic bacteria combined with tissue necrosis can be sources of overwhelming bacteremia and toxemia. A heavily contaminated compound fracture that is inadequately debrided of dead muscle can be rapidly transformed into a lethal nidus of infection. Efforts to preserve mangled extremities are another source of serious infections in recently injured patients.

Cardiogenic shock

Cardiac failure as a cause of shock in trauma patients can result from pre-existing heart disease or direct trauma to the heart (blunt cardiac injury including myocardial contusion, valvular disruption, or myocardial perforation with tamponade). The most common traumatic cause of acute heart failure is a direct blow to the precordium, hemorrhage into the myocardium, and a sudden cardiac arrhythmia or impairment of contractility in bruised muscle. Anatomical disruption of a cardiac valve can precipitate a catastrophic reduction in ventricular ejection fraction, as well. Cardiac tamponade is defined as compression of the heart during diastole that restricts filling and, in most circumstances, is associated with a penetrating wound or blunt compressive rupture of the heart. Optimal resuscitation of cardiogenic shock in a trauma patient depends upon precise identification of the cause of cardiac dysfunction and therapy directed to correct the physiologic or anatomic problem.

Pre-existing cardiac disease

An injured patient who has sustained blunt trauma and presents in shock with ECG evidence of ischemia should be suspected of having acute myocardial ischemia. Myocardial infarction is confirmed by an increase in cardiac troponin levels measured in serum. Treatment of cardiac dysfunction in injured patients can be complex as they may be receiving drugs that impair the normal response of the cardiovascular system. Blood pressure and pulse rates can be depressed in patients using b-blockers, calcium channel blockers, diuretics, and angiotensin-converting enzyme inhibitors. Patients with heart failure and angina who had optimal drug therapy prior to their injury may have inadequate reserve to compensate for even a small-volume hemorrhage or may develop unstable angina progressing to myocardial infarction. Angina or syncope from an arrhythmia may follow injury, also. Patients suffering the stress of a painful injury may develop angina that progresses to infarction due to an induced catecholamine surge. Several studies have shown that post-trauma deaths in geriatric patients are often the consequence of cardiac failure rather than the original injury. An appropriate protocol in such patients includes emergent transfer to an intensive care unit where directed resuscitation can be performed based upon invasive hemodynamic monitoring. In this protocol, extensive diagnostic evaluation and non-essential interventions are deferred until the patient has been stabilized.

Cardiogenic shock following acute myocardial infarction has a mortality rate of over 70 per cent. Patient survival depends not only upon treating the shock, but also upon restoring perfusion through or around the obstructed coronary vessel. Patients in cardiogenic shock may have excruciating chest pain radiating to both the left and right arms, are hypotensive and diaphoretic, and have distended neck veins. An electrocardiogram will reveal elevated ST-segments and development of a new Q-wave, although patients may have hidden ischemic changes if they have a left bundle branch block.

Successful, rapidly rendered treatment reduces death rates and prevalence of cardiogenic shock among patients suffering acute myocardial infarction. The cornerstone of therapy in acute coronary thrombosis has been immediate treatment with intravenous thrombolytic therapy, though secondary hemorrhage is a potential complication. Percutaneous transluminal coronary angioplasty can also reperfuse ischemic myocardium, but this procedure can only be optimally performed within 2 h by an experienced interventional cardiology team. The use of heparin in the injured patient with a myocardial infarct is often contraindicated, but aspirin with proven benefit as an antiplatelet drug may be given to selected patients. Additional treatments for acute myocardial infarction that improve survival include b-blockade and, for pain control, intravenous narcotics.

Cardiogenic shock is present when there is a systolic blood pressure less than 90 mmHg, a cardiac index less than 2.2 l/min/m^2 and a pulmonary artery wedge pressure exceeding 15 mmHg. Patients in cardiogenic shock who expire often have autopsy findings of greater than 40 per cent myocardial necrosis. Dobutamine or dopamine can be titrated to achieve improved organ perfusion without inducing a further increase in work load, and this averts a ‘vicious cycle' of deteriorating cardiac function. In patients with a marked elevation in systemic vascular resistance, vasodilator therapy can restore blood flow while reducing end diastolic pressure. Occasionally, the patient with acute cardiogenic shock is aided by the added diastolic perfusion generated by an intra-aortic balloon pump until the primary cardiac pathology can be corrected or the reperfused myocardium recovers. Arrhythmias are added risks to patients in cardiogenic shock, although prophylactic antiarrhythmic treatments are not indicated. Cardiogenic shock is a problem best identified early in its course, when therapy can still be effective.

Optimal therapy for injured patients with high-risk coronary artery disease involves prevention of acute myocardial infarction. b-Blockade can reduce myocardial oxygen consumption and the increased ischemia associated with tachycardia. Withdrawal of a b-blocker can precipitate ischemia, and injured patients on such medications may require intravenous dosing to sustain blockade. Effective prophylactic interventions in patients with occlusive coronary artery disease include prompt correction of hypovolemia, avoidance of hypoxia, and pain control including general anesthesia via endotracheal intubation and administration of large doses of narcotics. Evidence shows that nitrate therapy assists patients with angina not only by relieving pain, but also by improving collateral myocardial perfusion.

Pulmonary embolism can be a specific, unusual cause of sudden cardiogenic shock. Acute right ventricular dysfunction secondary to a sudden increase in pulmonary vascular resistance occurs in hospitalized injured patients who have a large load of emboli. Increased risk of thromboembolic disease has been demonstrated in injured patients with fractures in the lower extremities, prolonged bed rest, and hypercoagulopathy. A substantial embolus which occludes much of the pulmonary artery circulation not only causes acute hypoxia, but also can cause a sudden overdistension of the right ventricle. The ECG of a patient in shock following acute pulmonary embolism may show changes consistent with strain on the right ventricle, including a right bundle branch block, a shift of the transition zone to V_6, and Q-waves in III and aVP, but not II. An acute change in ECG findings in a high-risk patient can document the rapid progress to a fatal myocardial infarction. Akinesis of the dilated right ventricle on echocardiography confirms that the problem is acute right heart failure. Therapies for treatment of acute pulmonary embolism include infusion of thrombolytics to dissolve the thrombus or an embolectomy of the pulmonary arteries during cardiopulmonary bypass.

Cardiac tamponade

Cardiac tamponade is a pathologic problem caused by fluid, usually blood, which tensely distends the pericardial cavity. In the circumstance of abrupt filling of the pericardial cavity, the inelastic pericardium becomes restrictive and impairs cardiac output. At the end of diastole, the ventricular chambers compressed by the volume of blood in the pericardium do not fill. End-diastolic myocardial muscle tension is insufficient, and contractile power of the ventricle declines. Tamponade-induced cardiogenic shock can develop over minutes to hours. Cardiac output declines at a rate proportional to the rate of accumulation of fluid or blood in the pericardial cavity. Near-fatal cardiac tamponade arises as mean arterial pressure declines and the ejection fraction of the compressed left ventricle is reduced below a critical level. Death is imminent when the left ventricle fails to generate sufficient intraventricular pressure to open the aortic valve. In patients with cardiac wounds not immediately lethal, there is progressive deterioration as tamponade causes increasing obstruction to cardiac return and venous pooling as cardiac output declines.

Hypotension develops earlier in injured patients with tamponade and low intravascular volume, but will respond to intravenous infusion of 1 to 2 liters of isotonic saline. Patients who present to the hospital in full cardiac arrest after a penetrating presternal or other thoracic wound in proximity to the heart and who have had signs of life during transport to the hospital can be saved by rapid treatment consisting of endotracheal intubation and immediate thoracotomy in the emergency department. If the patient has cardiac tamponade, incision of the pericardium anterior and parallel to the course of the left phrenic nerve allows for evacuation of blood and clot. Inspection reveals either a regularly beating heart with good contractions or an agonal heartbeat. Patients can be restored to life by a period of internal cardiac massage to restore circulation until intravascular volume is expanded and a regular rhythm is restored, usually by administration of intravenous epinephrine. The frequency of successful resuscitation of patients who endure tamponade and asystole from a cardiac wound is low when the mechanism is penetrating trauma and nil in patients with blunt trauma.

Hemorrhage from the heart or a major vessel that causes pericardial tamponade can usually be controlled by direct compression. Large wounds involving the atrium

can be incorporated in a vascular clamp. A large perforation of a ventricle can be temporarily occluded by inserting a balloon catheter, inflating the balloon, and pulling up on the catheter to compress the balloon against the cardiac wound. Sutures used to repair cardiac wounds should be passed through pledgets which reduce the risk of tearing a fragile myocardium, particularly that of the thin-walled right ventricle. A caveat regarding blunt rupture of the heart (usually the atrium) is that a large defect will not cause tamponade to occur due to decompression of the pericardial blood back into the cardiac chambers.

Patients with either stable or minimally depressed blood pressure and precordial or parasternal stab or gunshot wounds can prove diagnostically challenging. Patients with a small to moderate amount of blood in the pericardium may not have distended neck veins and only a modest elevation in central venous pressure of 10 to 15 mmHg. These patients can suffer from tamponade which is not hemodynamically critical, but a further small deterioration in cardiac function can induce a sizeable decline in blood pressure. A surgeon-performed ultrasound of the pericardium using a 3.5 MHz transducer is the diagnostic test of choice and has an accuracy of 97 to 100 per cent in single center and multicenter studies. The subxiphoid pericardial window procedure performed in the operating room enables definitive determination whether there is blood in the pericardium whenever surgeon-performed ultrasound is unavailable or the result of the study is equivocal.

Cardiac contusion (blunt cardiac injury)

Death from a traumatic cardiac contusion is extraordinarily rare after an injured patient is brought to the emergency room. Some patients presumably die quickly at the scene of injury from acute pump failure or a cardiac arrhythmia following a blow to the myocardium. Patients who present to the emergency department hypotensive from a cardiac contusion are only a small proportion of the originally small population of patients who sustain a ‘true' (symptoms or ECG changes present) cardiac contusion. Patients in cardiogenic shock caused by a cardiac contusion have a depressed systolic pressure that cannot be attributed to other injuries and does not improve with the administration of intravenous fluids. Limited motion of the ventricular wall is demonstrable on transthoracic or transesophageal echocardiography in such patients. Patients in shock from cardiac contusion should undergo invasive hemodynamic monitoring and be managed in a surgical intensive care unit with intravenous fluids and vasoactive drugs. If cardiogenic shock persists, an intra-aortic balloon pump may provide a temporary improvement in cardiac output pending recovery of the bruised muscle.

Several tests have been advocated as diagnostic of cardiac contusion in patients who do not exhibit cardiovascular instability, but a normal ECG on admission to the emergency center or within 6 h essentially precludes the diagnosis. Even patients with abnormalities on the admission ECG tolerate general anesthesia for repair of other injuries with no evidence of increased risk of death.

Neurological shock

Clinical diagnosis of injury to the spinal cord

The Advanced Trauma Life Support® manual prepared by the Committee on Trauma of the American College of Surgeons distinguishes between the terms neurogenic shock and spinal shock. Neurogenic shock is the hypotension that occurs in patients whose loss of normal sympathetic tone from an injury to the spinal cord causes loss of arteriolar vasoconstriction, dilation of venous capacitance vessels, and a reduced cardiac sympathetic tone. Spinal shock refers to flaccidity without reflexes in skeletal muscle, the initial neurologic findings in extremities below the patient's injury to the spinal cord.

Pathophysiology

The physiology of hypotension in neurogenic shock patients is complex. As a consequence of loss of sympathetic tone in venous capacitance vessels, blood return to the heart declines as venous reservoirs expand. Patients with injuries to the spinal cord can be particularly sensitive to position changes which favor pooling in veins in the lower extremities. The loss of normal sympathetic tone in arterioles results in decreased systemic vascular resistance. While these patients are often hypotensive, they may have increased cardiac output secondary to vasodilation of the cutaneous circulation to the hands and feet. Furthermore, with poorly regulated cutaneous vasodilation, the patient with neurogenic shock is at greater risk for dissipation of body heat and the development of hypothermia. Patients with injuries to the spinal cord at the T_2 to T_5 level have unopposed vagal parasympathetic tone to the heart, and tachycardia in response to hypotension may not occur due to the depressed sympathetic cardiac tone. In selected patients, bradycardia may require treatment with atropine to sustain cardiac output.

Management of loss of sympathetic tone

The treatment of hypotension in patients with injuries to the spinal cord is the infusion of 2 to 3 liters of isotonic fluid. There is usually an improvement in systolic blood pressure in response to expanding intravascular volume; however, a minority of patients with pre-existing cardiac dysfunction may be at risk for development of heart failure due to loss of cardiac sympathetic tone.

In addition, patients with hypotension and an injury to the spinal cord should always be evaluated for hemorrhagic shock. Absence of sensation in the lower torso can mask pain from peritonitis or a pelvic fracture, and substantial hemorrhage can occur without symptoms. For example, hemorrhage is the most common cause of shock among patients with penetrating injury to the spinal cord. In patients with confirmed neurogenic shock whose systolic blood pressure does not exceed 90 mmHg, mean arterial perfusion pressure can be increased by an intravenous infusion of dopamine or an a-adrenergic agent such as norepinephrine. Invasive hemodynamic monitoring in hypotensive patients with injuries to the spinal cord may be necessary to identify patients who need volume expansion or inotropic support. Also, rapid restoration of good perfusion may benefit injured but not destroyed spinal cord tissue. Vasoconstrictors in hypotensive patients with neurogenic shock should be avoided because increased afterload may exacerbate cardiac dysfunction. In most circumstances, hypotension associated with injuries to the spinal cord resolves within 24 to 48 h. Although blood pressure may return to normal, patients still have impaired cardiovascular physiology. In particular, anesthetic management for emergency surgery in quadriplegic patients is challenging because of impaired compensatory responses to sudden hypovolemia. Patients with permanent paraplegia or quadriplegia can go on to develop autonomic hyper-reflexia with hypertension, bradycardia, and compensatory vasodilation and sweating in the skin above the injury to the cord.

Further reading

Trauma

American Association of Neurological Surgeons. *Guidelines for the management of severe head injury*. New York: Brain Trauma Foundation, 1995.

Barnes FC. *Cartridges of the world*. 7th edn. Northbrook, IL: DBI Books, Inc., 1993.

Biffl WL, *et al.* The unrecognized epidemic of blunt carotid arterial injuries: Early diagnosis improves neurologic outcome. *Annals of Surgery* 1998; **228**: 462–70.

Block EFJ. Diagnostic modalities in acute trauma. *New Horizons* 1999; **7**: 10–25.

Bracken MB, *et al.* Efficacy of methylprednisolone in acute spinal cord injury. *Journal of the American Medical Association* 1984; **251**: 45–52.

Brasel KJ, Ku J, Baker CC, Rutherford EJ. Damage control in the critically ill and injured patient. *New Horizons* 1999; **7**: 73–86.

Burch JM, *et al.* The injured colon. *Annals of Surgery* 1986; **203**: 701–11.

Chesnut RM, *et al.* The role of secondary brain injury in determining outcome from severe head injury. *Journal of Trauma*, 1993; **34**: 216–22.

Committee on Trauma, American College of Surgeons. *Advanced trauma life support for doctors*. 6th edn. Chicago: American College of Surgeons, 1997.

Daffner RH, Deeb ZL, Lupetin AR, Rothfus WE. Patterns of high-speed impact injuries in motor vehicle occupants. *Journal of Trauma* 1988; **8**: 498–501.

Demetriades D, *et al.* The management of penetrating injuries of the back. A prospective study of 230 patients. *Annals of Surgery* 1988; **207**: 72–4.

Demetriades D, *et al.* Selective nonoperative management of gunshot wounds of the anterior abdomen. *Archives of Surgery* 1997; **132**: 178–83.

Demetriades D, Rabinowitz B. Indications for operation in abdominal stab wounds. A prospective study of 651 patients. *Annals of Surgery* 1987; **205**: 129–32.

Fabian TC, Patton JH Jr, Croce MA, Minard G, Kudsk KA, Pritchard FE. Blunt carotid injury. Importance of early diagnosis and anticoagulant therapy. *Annals of Surgery* 1996; **223**: 513–25.

Feliciano DV, *et al.* Management of combined pancreatoduodenal injuries. *Annals of Surgery* 1987; **205**: 673–80.

Feliciano DV, Burch JM, Spjut-Patrinely V, Mattox KL, Jordan GL Jr. Abdominal gunshot wounds. An urban trauma center's experience with 300 consecutive patients. *Annals of Surgery* 1988; **208**:

362–70.

Feliciano DV, *et al.* Splenorrhaphy. The alternative. *Annals of Surgery* 1990; **211**: 569–82.

Feliciano DV. Diagnostic modalities in abdominal trauma. Peritoneal lavage, ultrasonography, computed tomography scanning, and arteriography. *Surgical Clinics of North America* 1991; **71**: 241–56.

Feliciano DV. Patterns of injury. In: Feliciano DV, Moore EE, Mattox KL, eds. *Trauma*. 3rd edn. Stamford: Appleton and Lange, 1996: 85–103.

Feliciano DV, Rozycki GS. Advances in the diagnosis and treatment of thoracic trauma. *Surgical Clinics of North America* 1999; **79**: 1417–29.

Feliciano DV, Moore EE, Mattox KL. Trauma damage control. In: Mattox KL, Feliciano DV, Moore EE, eds. *Trauma*. 4th edn. New York: McGraw-Hill, 2000: 907–31.

Martin TD, Feliciano DV, Mattox KL, Jordan GL Jr. Severe duodenal injuries. Treatment with pyloric exclusion and gastrojejunostomy. *Archives of Surgery* 1983; **118**: 631–5.

Mattox KL, Feliciano DV, Moore EE. *Trauma*. 4th edn. New York: McGraw-Hill, 2000.

Narayan RK, Wilberger JE, Povlishock JT. *Neurotrauma*. New York: McGraw-Hill, 1996.

National Academy of Sciences/National Research Council. *Accidental death and disability: The neglected disease of modern society*. Washington, DC: National Academy of Sciences, 1966 (Public Health Service Publication No 1071-A-13).

National Research Council/Institute of Medicine. *Injury in America. A continuing public health problem*. Washington, DC: National Academy Press, 1985.

National Safety Council. *Injury facts. 1999 Edition*. Itasca, IL: National Safety Council, 1997.

Nesathurai S. Steroids and spinal cord injury: Revisiting the NASCIS 2 and NASCIS 3 trials. *Journal of Trauma* 1998; **45**: 1088–93.

Newman PG, Feliciano DV. Blunt cardiac injury. *New Horizons* 1999; **7**: 26–34.

Pachter HL, Guth AA, Hofstetter SR, Spencer FC. Changing patterns in the management of splenic trauma: The impact of nonoperative management. *Annals of Surgery* 1998; **227**: 708–19.

Payne JE, Berne TV, Kaufman RL, Dubrowskij R. Outcome of treatment of 686 gunshot wounds of the trunk at Los Angeles County-USC Medical Center: Implications for the community. *Journal of Trauma* 1993; **34**: 276–81.

Phillips T, Sclafani SJ, Goldstein A, Scalea T, Panetta T, Shaftan G. Use of the contrast-enhanced CT enema in the management of penetrating trauma to the flank and back. *Journal of Trauma* 1986; **26**: 593–601.

Poole GV. Fracture of the upper ribs and injury to the great vessels. *Surgery, Gynecology and Obstetrics* 1989; **169**: 275–82.

Renz BM, Feliciano DV. Gunshot wounds to the right thoracoabdomen: a prospective study of nonoperative management. *Journal of Trauma*, 1994; **37**: 737–44.

Renz BM, Feliciano DV, Sherman R. Same admission colostomy closure (SACC). A new approach to rectal wounds: A prospective study. *Annals of Surgery* 1993; **218**: 279–93.

Rozycki GS, Ballard RB, Feliciano DV, Schmidt JA, Pennington SD. Surgeon-performed ultrasound for the assessment of truncal injuries: lessons learned from 1540 patients. *Annals of Surgery* 1998; **228**: 557–67.

Rozycki GS, *et al.* Early detection of hemoperitoneum by ultrasound examination of the right upper quadrant: a multicenter study. *Journal of Trauma* 1998; **45**: 878–83.

Rozycki GS, *et al.* The role of ultrasound in patients with possible penetrating cardiac wounds. A prospective multicenter study. *Journal of Trauma* 1999; **46**: 543–52.

Sivit CJ, *et al.* Safety-belt injuries in children with lap-belt ecchymosis: CT findings in 61 patients. *American Journal of Radiology* 1991; **157**: 111–14.

Velmahos GC, Tatevossian R, Demetriades D. The ‘seat belt mark' sign: a call for increased vigilance among physicians treating victims of motor vehicle accidents. *American Journal of Surgery* 1999; **65**: 181–5.

Shock

Abramson D, Scalea TM, Hitchcock R, Trooskin SZ, Henry SM, Greenspan J. Lactate clearance and survival following injury. *Journal of Trauma*, 1993; **35**: 584–8.

ACCM Task Force of the American College of Critical Care Medicine, Society of Critical Care Medicine. Practice parameters for hemodynamic support of sepsis in adult patients in sepsis. *Critical Care Medicine*, 1999; **27**: 639–60.

Adams JE III, Davila-Roman VG, Bessey PQ, Blake DP, Ladenson JH, Jaffe AS. Improved detection of cardiac contusion with cardiac troponin 1. *American Heart Journal*, 1996; **131**: 308–12.

Bickell WH, *et al.* Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. *New England Journal of Medicine*, 1994; **331**: 1105–9.

Bishop MH, *et al.* Prospective, randomized trial of survivor values of cardiac index, oxygen delivery and oxygen consumption as resuscitation endpoints in severe trauma. *Journal of Trauma* 1995; **38**: 780–7.

Califf RM, Bengtson JR. Cardiogenic shock. *New England Journal of Medicine* 1994; **330**: 1724–30.

Canizaro PC, Prager MD, Shires GT. The infusion of Ringer's lactate solution during shock. *American Journal of Surgery* 1971; **122**: 494–501.

Choi PTL, Yip G, Quinonez LG, Cook DJ. Crystalloids vs. colloids in fluid resuscitation: a systematic review. *Critical Care Medicine* 1999; **27**: 200–10.

Cryer HM, Miller FB, Evers BM, Rouben LR, Seligson DL. Pelvic fracture classification: correlation with hemorrhage. *Journal of Trauma* 1988; **28**: 973–80.

Davis JW, Kaups KL, Parks SN. Base deficit is superior to pH in evaluating clearance of acidosis after traumatic shock. *Journal of Trauma* 1998; **44**: 114–18.

Durham RM, Neunaber K, Mazuski JE, Shapiro MJ, Baue AE. The use of oxygen consumption and delivery as endpoints for resuscitation in critically ill patients. *Journal of Trauma* 1996; **41**: 32–9.

Gattinoni L, *et al.* A trial of goal-oriented hemodynamic therapy in critically ill patients. *New England Journal of Medicine* 1995; **333**: 1025–32.

Hebert PC, *et al.* A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *New England Journal of Medicine* 1999; **340**: 409–17.

Heyland DK, Cook DJ, King D, Kernerman P, Brun-Buisson C. Maximizing oxygen delivery in critically ill patients: a methodological appraisal of the evidence. *Critical Care Medicine* 1996; **24**: 517–24.

Illig KA, Swierzewski MJ, Feliciano DV, Morton JH. A rational screening and treatment strategy based on the electrocardiogram alone for suspected cardiac contusion. *American Journal of Surgery* 1991; **162**: 537–44.

Ivatury RR, Kazigo J, Rohman M, Gaudino J, Simon R, Stahl WM. ‘Directed' emergency room thoracotomy: a prognostic prerequisite for survival. *Journal of Trauma* 1991; **31**: 1076–82.

Ivatury RR, Simon RJ, Islam S, Fueg A, Rohman M, Stahl WM. A prospective randomized study of end points of resuscitation after major trauma: global oxygen transport indices versus organ-specific gastric mucosal pH. *Journal of American College of Surgeons* 1996; **183**: 145–54.

Kilbourn R. Nitric oxide synthase inhibitors: a mechanism-based treatment of septic shock. *Critical Care Medicine* 1999; **27**: 857–8.

Kincaid EH, Miller PR, Meredith JW, Rahman N, Chang MC. Elevated arterial base deficit in trauma patients: a marker of impaired oxygen utilization. *Journal of American College of Surgeons* 1998; **187**: 384–92.

Klein SR, Saroyan RM, Baumgartner F, Bongard FS. Management strategy of vascular injuries associated with pelvic fractures. *Journal of Cardiovascular Surgery* 1992; **33**: 349–53.

Lam AM. Acute spinal cord injury: monitoring and anaesthetic implications. *Canadian Journal of Anaesthesia*, 1991; **38**: R60–73.

Levi L, Wolf A, Belzberg H. Hemodynamic parameters in patients with acute cervical cord trauma: description, intervention and prediction of outcome. *Neurosurgery* 1993; **33**: 1007–17.

Lualdi JC, Goldhaber SZ. Right ventricular dysfunction after acute pulmonary embolism: pathophysiologic factors, detection and therapeutic implications. *American Heart Journal* 1995; **130**: 1276–82.

Lucas CE. Resuscitation of the injured patient: the three phases of treatment. *Surgical Clinics of North America* 1977; **57**: 3–15.

Lucas CE, Benishek DJ, Ledgerwood AM. Reduced oncotic pressure after shock. *Archives of Surgery* 1982; **117**: 675–9.

Mackenzie CF, Shin B, Krishnaprasad D, McCormack F, Illingworth W. Assessment of cardiac and respiratory function during surgery on patients with acute quadriplegia. *Journal of Neurosurgery* 1985; **62**: 843–9.

Marik PE, Mohedin M. The contrasting effects of dopamine and norepinephrine on systemic and splanchnic oxygen utilization in hyperdynamic sepsis. *Journal of the American Medical Association* 1994; **272**: 1354–7.

Mayberry JC, Mullins RJ, Crass RA, Trunkey DD. Prevention of abdominal compartment syndrome by absorbable mesh prosthesis closure. *Archives of Surgery* 1997; **132**: 957–62.

McCormick M, Feustel PJ, Newell JC, Stratton HH, Fortune JB. Effect of cardiac index and hematocrit changes on oxygen consumption in resuscitated patients. *Journal of Surgical Research* 1988; **44**: 499–505.

Mikulaschek A, Henry SM, Donovan R, Scalea TM. Serum lactate is not predicted by anion gap or base excess after trauma resuscitation. *Journal of Trauma* 1996; **40**: 218–23.

Mizock BA, Falk JL. Lactic acidosis in critical illness. *Critical Care Medicine* 1992; **20**: 80–93.

Moreno C, Moore EE, Rosenberger A, Cleveland HC. Hemorrhage associated with major pelvic fracture: a multispecialty challenge. *Journal of Trauma* 1986; **26**: 987–94.

Natanson C, Hoffman WD, Suffredini AF, Eichacker PQ, Danner RL. Selected treatment strategies for septic shock based on proposed mechanisms of pathogenesis. *Annals of Internal Medicine* 1994: **120**: 771–83.

Ochoa JB, *et al.* Nitrogen oxide levels in patients after trauma and during sepsis. *Annals of Surgery* 1991; **214**: 621–26.

Parrillo JE. Pathogenetic mechanisms of septic shock. *New England Journal of Medicine* 1993; **328**: 1471–7.

Rangel-Frausto S, Pittet D, Costigan M, Hwang T, Davis CS, Wenzel RP. The natural history of the systemic inflammatory response syndrome, SIRS. *Journal of the American Medical Association* 1995; **273**: 117–23.

Rudis MI, Basha MA, Zarowitz BJ. Is it time to reposition vasopressors and inotropes in sepsis? *Critical Care Medicine* 1996; **24**: 525–37.

Rutherford EJ, Morris JA, Reed GW, Hall KS. Base deficit stratifies mortality and determines therapy. *Journal of Trauma* 1992; **33**: 417–23.

Scalea TM, *et al.* Geriatric blunt multiple trauma: improved survival with early invasive monitoring. *Journal of Trauma* 1990; **30**: 129–36.

Zipnick RI, *et al.* Hemodynamic responses to penetrating spinal cord injuries. *Journal of Trauma* 1993; **35**: 578–82.

3 Surgical nutrition

Timothy A. Pritts Josef E.Fischer

Introduction
Nutritional Requirements
Protein
Calories
Fat
Vitamins and trace elements
Metabolic Changes In Starvation
Metabolic Changes in Surgery, Sepsis, and Trauma
Catabolic phase
Early anabolic phase
Late anabolic phase
Other Causes of Inadequate Nutrition
Poor oral intake
Inadequate absorption
Increased losses
Hospital food
Nutritional Assessment
Research methodologies
Clinical methodologies
Indications for Nutritional Support
Enteral Nutrition
Enteral versus parenteral routes
Methods
Products
Administration
Complications
Parenteral Nutrition
Indications for parenteral nutrition
Role in primary therapy
Role in secondary therapy
Route of delivery
Composition of formulas
Administration
Complications of parenteral nutrition
Nutritional Pharmacology
Further reading

Introduction

While most patients undergoing elective operations easily tolerate a brief period of perioperative starvation, as many as 50 per cent of hospital patients may be nutritionally compromised, depending on the definition. Nutritional status directly impacts on patient care and adequate attention to nutritional issues can help to minimize morbidity and complications. The field of nutritional support has advanced significantly in the late twentieth century with the advent of total parenteral nutrition. The initial enthusiasm for parenteral nutrition has now become justifiably tempered, with primary use of enteral nutrition where possible and parenteral nutrition generally assuming a supplemental and secondary role. More recently, the role of nutritional needs in individual disease states has given rise to the concept of nutritional pharmacology, with nutritional supplementation guided by pathophysiologic principles.

Nutritional requirements

Protein

Protein is perhaps the most important nutrient. Although protein can be degraded and used for gluconeogenesis, this yields only one-fourth of the energy required for protein synthesis and is thus a wasteful process. A primary goal in nutritional support is to provide adequate non-protein sources of fuel so that protein catabolism is minimized.

Between 15 per cent and 20 per cent of body weight is protein. The daily protein turnover is approximately 3 per cent of this, about 275 g in a 70-kg man. A portion of the body's protein is in constant flux in the amino acid pool, providing a supply for new protein synthesis or conversion into other biologically active compounds. Nearly all ingested amino acids are absorbed. These ingested proteins contribute 25 g to the amino acid pool. The other 250 g are provided by endogenous breakdown. If adequate energy supply is present, body protein breakdown is approximately equaled by protein synthesis and the patient is in nitrogen equilibrium.

Protein requirements

The average normal protein requirement of ‘high biologic value protein' is 0.8 g/kg or 56 to 60 g/day. Normal protein intake in affluent Americans is twice this amount. Protein requirements may more than double in stressed, burned, or multiple trauma patients. Assuming an adequate non-protein energy supply, most amino acids can be recycled. In this fashion, only small amounts of essential amino acids are needed in order to maintain nitrogen equilibrium. In adults, 19 to 20 per cent of protein intake should be essential amino acids. This percentage should increase with depletion or injury.

Amino acids

Humans cannot synthesize the essential amino acids. Thus, they must be obtained from the diet. In addition, several other amino acids are conditionally indispensable, as their low synthetic rates may be exceeded by increased requirements, especially in infants. The six truly non-essential amino acids, whose needs are fully met by synthesis, are alanine, aspartic acid, asparagine, glutamic acid, glycine, and serine. The eight essential amino acids are valine, leucine, isoleucine, lysine, methionine, phenylalanine, threonine, and tryptophan. Conditionally indispensable amino acids are cysteine, which is synthesized from methionine, tyrosine, which is synthesized from phenylalanine, and histidine, proline, hydroxyproline, glutamine, and arginine. The liver plays a major part in the catabolism of the essential amino acids, with the exception of the branched chain amino acids valine, leucine, and isoleucine, which are degraded by skeletal muscle.

The semi-essential amino acid glutamine has recently received a great deal of attention. Glutamine is abundant in the circulation and serves as a precursor for other amino acids and proteins. Glutamine also serves as the major energy substrate for the intestinal mucosa, as a nitrogen transporter between organs, and as an important route of ammonia detoxification during acidemia. Enteral glutamine supplementation may promote small bowel adaptation and lead to increased intestinal mucosa villous height and enterocyte protein content, but this effect is not universally seen with parenteral glutamine. In addition, those studies utilizing glutamine dipeptides given parenterally yield improved nitrogen balance which is statistically, but not biologically significant. Despite the probable beneficial effects of glutamine supplementation, glutamine is not presently available in standard parenteral nutrition formulations, partly because it is not stable at room temperature.

Nitrogen losses and balance

One gram of nitrogen equals 6.25 g of protein. Obligate nitrogen losses are 56 to 57 mg/kg per day. The largest proportion of this (37 mg/kg) occurs in the urine. Twelve milligrams per kg of nitrogen is lost in the stool, 5 mg/kg from the integument, and 2 to 3 mg/kg to evaporation. A representative estimation of daily nitrogen flux (in equivalent protein) in a 70-kg man is shown in Fig. 1. Nitrogen balance may be calculated as the difference between nitrogen intake and nitrogen losses, as shown in the equation:

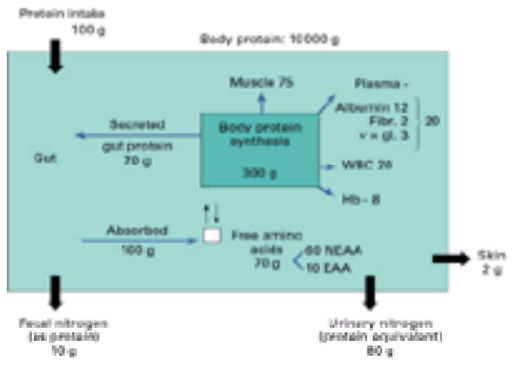


Fig. 1. Schematic diagram of daily protein flux in a 70-kg man.

$$N_{balance} = N_{intake} - N_{loss}$$

Clinical estimation of nitrogen balance requires accurate assessment of nitrogen intake, both enterally and parenterally, and there must be an appropriate number of calories to ensure proper utilization. Accurate determination of nitrogen loss is very difficult to determine in the clinical setting, but nitrogen losses can be estimated by the equation:

$$N_{loss} = 24\ h\ urinary\ urea\ nitrogen + 4\ g/day\ (fecal\ and\ non\text{-}urinary\ loss)$$

Nitrogen loss may be greater than this formula would indicate due to other sources such as wound and fistula drainage. In general, healthy adults maintain nitrogen equilibrium over time. A positive nitrogen balance is usually associated with anabolic states, such as childhood growth or recovery after an elective operation. Negative nitrogen balance is associated with catabolic states and may frequently occur in surgical patients as a result of injury, underlying illness, or infection. Skeletal muscle wasting due to protein breakdown exceeding protein synthesis is a common clinical manifestation of negative nitrogen balance and catabolism. A major goal of nutritional support is to allow protein to be used for anabolic processes by providing adequate non-protein sources of fuel.

Calories

There are three major sources of energy: protein, fat, and carbohydrates. Of normal daily energy expenditure, 85 per cent is from fat and carbohydrates, and 15 per cent from protein.

Protein as an energy source

Of the 15 per cent of normal daily energy expenditure supplied by protein, approximately 50 per cent of this is through direct oxidation of branched chain amino acids to high-energy phosphate. The remainder is via gluconeogenesis. Protein breakdown yields 4 kcal/g, but is an inefficient source of energy as this yields only one-fourth of the energy required for protein synthesis. Protein exists in the body in functional forms only; there is no specific storage form. Overutilization of protein as a fuel source over a prolonged period of time, as occurs in sepsis, can lead to adverse sequelae such as muscle wasting, impaired wound healing, and increased susceptibility to infection.

Carbohydrates as an energy source

Digested sugars are absorbed by the small intestine and transported to the liver, which occupies a central role in the regulation of plasma glucose levels (Fig. 2). Here, the sugars galactose and fructose are rapidly converted into glucose. Glucose yields about 4 kcal/g through glycolysis and the tricarboxylic acid cycle. Glycogen breakdown, due the addition of water and electrolytes, yields only 1 to 2 kcal/g. Most carbohydrates in parenteral nutrition are supplied in the form of dextrose, which provides 3.4 kcal/g.

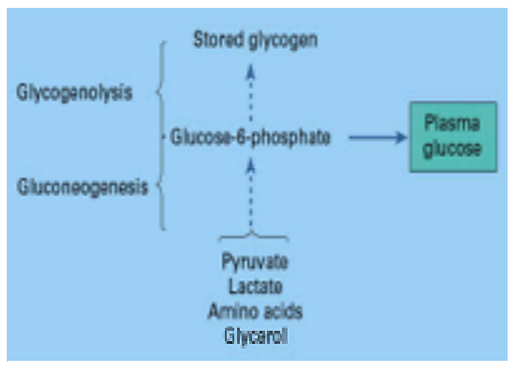


Fig. 2. Hepatic contributions to plasma glucose. The liver serves a central role in glucose metabolism.

The liver and skeletal muscle can store glucose in the form of glycogen. Glycogen stores, however, are exhausted within 24 h of initiation of a fast, and the body then becomes dependent on gluconeogenesis as a source of glucose. Three sources of substrate for gluconeogenesis are glycerol from triglyceride breakdown, amino acids from proteolysis, and lactate from anaerobic glycolysis. Although adaptation may occur later, the brain, red and white blood cells, and renal medulla are initially dependent on glucose as a source of energy during a fast. If there are inadequate alternative sources of glucose, protein is catabolized for gluconeogenesis. At least 400 cal/day in the form of exogenously administered glucose are required to minimize proteolysis.

Fat as an energy source

Fat provides between 25 per cent and 45 per cent of the calories in the American diet. As the body's available glucose and glycogen stores are limited and rapidly exhausted, most body tissues depend on fat for fuel. Fat is an energy dense substance and utilization provides about 9 kcal/g.

Ingested lipids are absorbed by the mucosa of the small intestine by a complicated process involving hydrolysis of triglycerides, absorption, and resynthesis. Once absorbed, short and medium chain fatty acids directly enter the portal venous system. Long chain fatty acids are transported in the lymphatics in the form of triglycerides. Lipids are cleared from the bloodstream through the action of lipoprotein lipase, located on the endothelial surface, and then may enter adipose tissue for storage as triglycerides. In addition to lipids, glucose also contributes to energy storage in adipocytes by providing energy for triglyceride synthesis and by contributing to the formation of glycerol.

Triglycerides consist of three fatty acids bound to glycerol. Hydrolysis of triglycerides is regulated by hormone-sensitive lipase, present only in adipose tissue. The only major hormone that suppresses hormone-sensitive lipase activity is insulin. Epinephrine, norepinephrine, glucagon, growth hormone, and glucocorticoids upregulate hormone-sensitive lipase. The action of epinephrine, norepinephrine, and glucagon is rapid, while the action of growth hormone and gluco-corticoids is more delayed. Together, these hormones ensure increased hormone-sensitive lipase activity under stressful conditions to provide adequate energy sources.

After lipolysis, free fatty acids are released into the circulation. Free fatty acids actually circulate bound non-covalently to albumin. Nearly all tissues, with the notable exception of the brain, can utilize fatty acids as an energy source. They may be oxidized directly in the mitochondria after activation in the cytoplasm by condensation with coenzyme A or provide calories as ketone bodies manufactured in the liver. In the liver, free glycerol can be converted into glucose via gluconeogenesis to provide

energy for tissues unable to utilize other substrates. An overview of the relationship between the liver and other tissues and fatty acid metabolism is shown in Fig. 3.

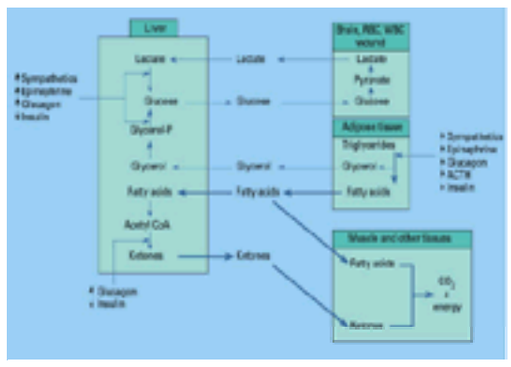


Fig. 3. Interactions between the liver and other tissues during starvation. Free fatty acids are used as an energy substrate except in the brain, wound, and red and white blood cells. Glycerol from triglyceride breakdown is used for gluconeogenesis.

In severe stress or sepsis, fat oxidation and/or utilization may become impaired. Fat oxidation appears to become impaired relatively early, but fat clearance remains normal until later. This may result in fat overload syndrome, characterized clinically by chill, fever, pulmonary insufficiency, back pain, and reticuloendothelial system blocking. To ensure adequate fat clearance, either serum lipemia or the respiratory quotient (see below) should be monitored.

Determination of caloric needs

Caloric needs are related to the metabolic rate, which in turn is demonstrated by the formula:

$$\text{metabolic rate} = 4.83 \times V\text{o}_2$$

where $V\text{o}_2$ is oxygen consumption in liters per unit of time. Furthermore, the respiratory quotient (RQ), a ratio of carbon dioxide produced to oxygen consumed, can be used to estimate utilization of the various caloric sources by the following formula:

$$\text{RQ} = V\text{co}_2 / V\text{o}_2$$

An RQ of 1 is consistent with pure carbohydrate utilization. An RQ of 0.7 is consistent with utilization of fat. An RQ of less than 0.7 indicates ketogenesis. In clinical practice, and RQ of greater than 1 is uncommon, but this may indicate overfeeding. An RQ of between 0.8 and 1, indicating mixed substrate utilization, is desirable.

The basal energy expenditure (BEE) can be estimated from the Harris–Benedict equation:

$$\text{BEE (men)} = 66.47 + [13.75 \times W] + [5 \times H] - [6.76 \times A]$$

$$\text{BEE (women)} = 655.1 + [9.56 \times W] + [1.85 \times H] - [4.68 \times A]$$

where W is weight in kilograms, H is height in centimetres, and A is age in years.

The additional caloric needs due to illness or other metabolic stress can then be calculated by multiplying the BEE by an injury factor:

$$\text{minor operation} = 1.2 \text{ (20 per cent increase)}$$

$$\text{skeletal trauma} = 1.35 \text{ (35 per cent increase)}$$

$$\text{major sepsis} = 1.6 \text{ (60 per cent increase)}$$

$$\text{severe thermal injury} = 2.10 \text{ (110 per cent increase)}$$

This estimate can be further refined to account for activity by multiplication by 1.2 if the patient is confined to bed and 1.3 if the patient is not confined to bed. Therefore, caloric needs equal:

$$\text{BEE} \times \text{injury factor} \times \text{activity factor}$$

This formula may overestimate caloric needs.

Calorie to nitrogen ratio

Calories must be provided in adequate excess to nitrogen so that protein synthesis is permitted. A calorie to nitrogen ratio of between 100 and 150 to 1 (i.e. 100 non-protein calories per gram of nitrogen) is suggested in normal states. This ratio may vary with different disease states. In sepsis a ratio of 100 to 1 is appropriate. In patients with uremia, an increase in the calorie to nitrogen ratio to between 300 and 400 to 1 is suggested.

Fat

In addition to its utility as a caloric source, fat is required for normal immune responses, growth, and development. The essential fatty acids linoleic acid and a-linoleic acid cannot be synthesized and thus must be supplied in the diet in the diet. Fatty acids also serve as precursors for compounds associated with the inflammatory response, including prostaglandins, thromboxanes, leukotrienes, and prostacyclins. Future methods of modifying the inflammatory response may include therapeutic uses of fatty acids.

Vitamins and trace elements

Vitamins and trace elements are required in small amounts to maintain normal body structure, function, and metabolism. Vitamins may function as antioxidants (C and E), in genetic regulation (A, K, and D), serve as coenzyme precursors (the B complex vitamins), or serve other roles that are not yet understood. Trace elements may serve as components of body structures, including bone, or as components of metalloproteins and metalloenzymes. The United States Recommended Daily Allowances (RDA) for vitamins and trace elements are given in Table 1.

Vitamin/trace element	Enteral RDA	Intravenous RDA
Trace element	RDA	RDA
Vitamin A (IU)	4000–5000	3300
Vitamin D (IU)	400	200
Vitamin E (IU)	12–15	10
Vitamin C (mg)	45	100
Folacin (μg)	400	400
Niacin (μg)	17–20	40
Riboflavin (μg)	1.1–1.8	3.6
Thiamine (mg)	1.0–1.5	3.0
Pyridoxine (mg)	1.6–2.0	4.0
Cyanocobalamine (μg)	3	5
Pantothenic acid (mg)	5–10	15
Biotin (μg)	150–300	60
Potassium	70–160 m g	5 mg
Calcium (mg)	1000	
Magnesium (mg)	350	
Iron (mg)	15	1
Zinc	15	
Iodine (μg)	150	
Copper (mg)	2–3	
Manganese (mg)	2.5–5.0	
Chromium (mg)	0.05–0.2	
Selenium (μg)	0.03	0.2
Molybdenum (μg)	0.15–0.5	
Phosphorus	1000	

Table 1 Recommended daily allowances (RDAs) for vitamins and trace elements

Metabolic changes in starvation

As discussed previously, the body's glycogen stores are essentially depleted after 24 h of starvation. In the initial several days of fasting, fat stores and protein supply caloric needs. Fatty acids, mobilized from adipose tissue, provide energy for many tissues and also serve as a substrate for gluconeogenesis in the liver. The products of proteolysis, most of which occurs in skeletal muscle, are also used for gluconeogenesis, which is in turn used to supply glucose for the brain and the other obligate glucose consuming tissues. This results in an obligate nitrogen loss of 10 to 15 g/day.

After several days of starvation, the body begins to adapt. The brain begins to utilize ketones as an energy source rather than glucose, which decreases the need for proteolysis. The obligate nitrogen loss is then reduced to about 4 g/day.

During periods of starvation, protein breakdown can be minimized by exogenous glucose administration. As little as 100 g of glucose per day will decrease proteolysis. This appears to be mediated by insulin, as glucose infusion in diabetic patients will not achieve protein sparing in the absence of insulin. Glucose administration will not decrease proteolysis beyond 50 per cent. The other 50 per cent appears to be amino acid sensitive.

Metabolic changes in surgery, sepsis, and trauma

The body's response to insults such as surgery, sepsis, and trauma differs from the response to simple starvation (Table 2). Critically, the protein sparing effect of glucose is diminished. This leads to increased and prolonged protein catabolism, which may lead to organ failure if allowed to continue untreated. The events in the response to surgery, sepsis, and trauma can be divided into three phases: (1) catabolic, (2) early anabolic, and (3) late anabolic.

	Starvation	Stress hypermetabolism
Resting energy expenditure	↓	↑↑
Respiratory quotient	Low (0.65)	High (0.85)
Mediator activation	–	+++
Regulatory responsiveness	++++	+
Primary fuels	Fat	Mixed
Proteolysis	+	+++
Branched-chain oxidation	+	+++
Hepatic protein synthesis	+	+++
Ureagenesis	+	+++
Urinary nitrogen loss	+	+++
Gluconeogenesis	+	+++
Ketone body production	++++	+

[a]Patients fall in a continuum between extremes of starvation and stress hypermetabolism.

Table 2 Differences in metabolic responses in starvation and stress[a]

Catabolic phase

The catabolic phase immediately follows an insult and may last from 3 to 8 days following uncomplicated elective surgery. Following severe stress, such as multiple trauma or thermal injury, this phase may persist for weeks. The catabolic phase is governed by the classical sympathoadrenal response. In the immediate period after an insult, metabolic demands increase and there is a mobilization of protein to serve as a substrate for gluconeogenesis. Resting energy expenditure may increase dramatically depending on the severity of the insult (Fig. 4). Urinary nitrogen excretion increases to levels greater than those seen in simple starvation and may exceed 20 g/day. Lipolysis and fatty acid oxidation are also increased.

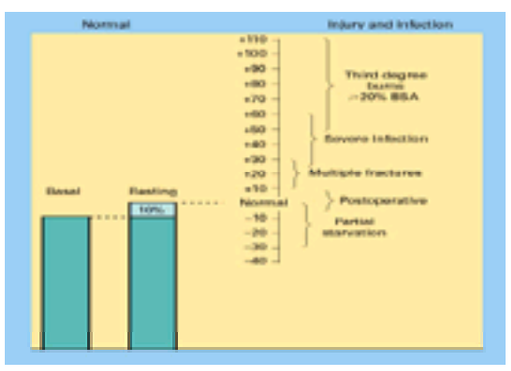


Fig. 4. Differences in resting energy expenditure in patients under normal conditions and in situations of starvation and various levels of metabolic stress.

Although glucose infusions decrease nitrogen loss in patients following an acute insult, this effect is much less pronounced than the protein sparing effect seen in starvation. It appears that the normal response to insulin is lost. The slope remains normal, but the baseline is depressed and the patient becomes insulin resistant. The normal response of skeletal muscle to insulin is decreased by up to 50 per cent.

Early anabolic phase

If the patient survives the catabolic phase, the body enters a period of anabolism. The transition from catabolism to anabolism is marked by a decrease in nitrogen loss, diuresis of retained water, and an improvement in the patient's overall subjective condition. This transition period normally lasts 1 to 2 days.

Following the transition period, nitrogen balance becomes positive, and a net protein synthesis occurs to recover the infrastructure previously utilized for energy needs during catabolism. A corresponding gain in weight and muscle strength occurs. Although it takes place over a more prolonged period of time, the net influx of nitrogen eventually equals the nitrogen losses from previous catabolism. This phase of initial recovery from an insult may last from a few weeks to a few months.

Late anabolic phase

Following the initial recovery from the catabolism seen in response to an insult, the body enters a late anabolic phase, which may last several months. During this period, net influx of nitrogen is gradually replaced by nitrogen balance as the body's protein structures are repleted. If caloric intake is in excess of caloric expenditure, the fat stores previously utilized are also restored. In most patients, body weight will eventually return to baseline.

Other causes of inadequate nutrition

Poor oral intake

Other than the catabolism associated with illness, there are many other possible etiologies of an inadequate nutrition status. There may simply be a lack of food intake, especially in alcoholic patients. In both developed and developing countries, inadequate food supply may lead to malnutrition. More common, however, is poor nutrition secondary to other illness. Anorexia is commonly associated with several chronic illnesses, especially cancer, liver disease, and obstruction. Obstruction may in turn be secondary to other disorders, such as esophageal stricture due to reflux, tumor at any location in the gastrointestinal tract, or strictures associated with inflammatory bowel disease.

Inadequate absorption

Many disease states may lead to inadequate absorption of ingested nutrients. This may include short-bowel syndrome, which may be due to any of a number of etiologies, including massive bowel resection for volvulus or ischemic bowel. Motility disorders, gastric resection, and connective tissue disorders such as scleroderma may also lead to malabsorption.

Increased losses

Other disease states may lead to increased losses from the gastrointestinal tract. This may occur in patients with gastrointestinal fistulas. In these individuals, the loss of nutrients via the fistula must be considered when assessing their nutritional status and nutritional needs. In patients with inflammatory bowel disease, excess loss of protein may occur secondary to mucous, bleeding, and mucosal slough.

Hospital food

A common cause of malnutrition in hospital patients is the hospital diet itself. Hospital food is commonly bland, unappetizing, awkwardly scheduled, and delivered cold. Oral intake is often withheld for a multitude of reasons, both justified and trivial (such as chest or abdominal roentgenogram). Also, diets may be advanced too slowly following procedures.

Nutritional assessment

The goal of nutritional assessment in the surgical patient is to predict the patient at risk for complications due to inadequate nutrition. Ideally, intervention can then be taken to improve nutritional status. Multiple methods are available to determine the nutritional status of patients. Many require little or no specialized equipment, but some are available only in research centers and are not clinically practical. While it is possible to predict accurately nutritional risk for populations of patients, individual risk cannot be accurately predicted.

Research methodologies

Displacement

This is probably the most sensitive method of determining lean body mass. The patient is submerged in water and the composition of various body compartments is estimated. This method assumes that fat is less dense than other body components and that the density of the fat-free portion of the patient is constant. As was previously discussed, the latter assumption is inaccurate in serious illness due to catabolism. Furthermore, the submersion of a critically ill patient to determine displacement may not be practical. This method is not available in most institutions.

Labeled ion exchange

This procedure is also only available in research centers. Using this technique, known volumes of radiolabeled isotopes are injected and used to determine the volumes of various body compartments. Tritiated water is used to determine total body water. Extracellular water volume is determined using radiolabeled sodium and intracellular water volume can be determined using radiolabeled potassium. The volumes of these compartments are then used to estimate lean body mass.

Total body counters

This method utilizes ^{40}K, a naturally occurring radioisotope that occurs in a fixed ratio to the more stable forms of potassium, ^{39}K and ^{41}K. The spontaneous decay of ^{40}K is measured using a total body counter and reflects lean body mass. This method is not clinically useful, as few total body counters exist and the patient is required to remain still for prolonged periods of time.

Clinical methodologies

History and physical examination

The importance of an accurate history and physical examination can never be underestimated. The clinician should pay special attention to a history of chronic illness, unintentional or intentional weight loss, anorexia, and disease processes that may affect oral intake. The patient should also be queried about his or her daily activities and whether recent ability to pursue these has changed. On physical examination, the examiner should pay close attention to signs of poor nutritional status. A cachectic appearance, loose fitting clothes, edema, muscle wasting, loss of body fat, thenar muscle atrophy, glossitis, skin lesions, and pallor may all point to nutritional deficits. The overall impression of an experienced clinician is as accurate as extensive laboratory testing in predicting the patient at risk.

Anthropometrics

Various body measurements can provide supplementary data to corroborate the findings of history and physical exam. These measurements include body weight, weight to height ratio, midarm muscle circumference, and trifold skin thickness. They are useful mainly to supply information concerning fat stores.

Protein measurement

The levels of the circulating proteins albumin, transferrin, prealbumin, and retinol binding protein can be measured in most hospital laboratories. Changes in the levels of these proteins can reflect changes in nutritional status, with respect to their half lives. The half lives of these proteins vary, with albumin the longest at 20 days, transferrin next at 8.5 days, prealbumin next at 1.3 days, and retinol-binding protein the shortest at 0.4 days.

Nitrogen balance

As was previously discussed, a positive nitrogen balance is associated with anabolic states, whereas a negative nitrogen balance is associated with catabolism. Accurate determination of nitrogen balance in the research setting involves the collection of nitrogen lost from the wound, integument, urine, and stool, as well as measurement of expired nitrogen. While this is impractical except in research settings, determination of nitrogen balance in the clinical setting is less cumbersome. As discussed previously, nitrogen losses are determined by measuring 24 h urinary urea nitrogen with the addition of 4 g/day for fecal and insensible losses.

Immunologic function

Delayed cutaneous hypersensitivity has been shown to reflect underlying immune status and correlate with mortality. This can be determined by skin testing with a battery of antigens. While this does not strictly correlate with nutritional status, anergic patients are likely those who are also malnourished.

Radiologic imaging

Computed tomography and magnetic resonance imaging may provide supplementary information concerning the nutritional status of the patient through direct visualization of various body compartments. Both intra-abdominal and subcutaneous fat stores can be estimated by these techniques, but cost is a limiting factor.

Indications for nutritional support

There are no strict criteria indicating which patients need nutritional support. Several factors must be considered. From a practical standpoint, several elements of nutritional status indicate that a patient may be at special risk for nutrition-related complications. Two main indicators are a serum albumin of less than 3 g/dl in a stable patient and weight loss of greater than 10 per cent of body weight or a current body weight at less than 80 to 85 per cent of ideal body weight. Additional supporting findings are anergy to injected skin antigens, transferrin less than 200 mg/dl, and a history of functional impairment.

Several other factors must be considered. The age of the patient is important. Patients less than 60 years old can tolerate 12 to 14 days of fasting with minimal deleterious sequelae. This time interval decreases to 7 to 8 days for patients over 60 years old, and 5 to 6 days for patients over 80 years old. The patient's overall state of health (healthy or otherwise), other coincident medical problems (especially diabetes and liver disease), the magnitude of the anticipated insult, and the likelihood of resuming near normal oral intake in the near future must also be taken into account when deciding which patients may benefit from nutritional support.

Enteral nutrition

Enteral versus parenteral routes

The enteral route is the preferred route for delivering nutritional support. There are several advantages to enteral feeding over parenteral nutrition. Enteral feeding is more physiologic than the parenteral route because it does not bypass the portal system, thus allowing the liver to function in its normal manner. Several studies have shown that enteral feeding confers advantages over parenteral nutrition in critically ill patients in terms of both morbidity and mortality. It appears that the body better utilizes enterally delivered nutrition than nutrition delivered by vein. Enteral nutrition attenuates gut mucosal atrophy and maintains the gut-associated lymphoid tissue. Thus, gut barrier and immune functions are better maintained and infectious complications are decreased. Enteral nutrition theoretically protects against translocation of bacteria from the gastrointestinal tract, but this concept has not yet been demonstrated in human trials. Finally, the enteral route is less expensive than the parenteral route. The enteral route should be used preferentially, and even when full enteral nutritional support is not practical, a portion of the patient's nutrition should still be given enterally if possible. The advantages of the enteral route are probably obtained when as little as 20 per cent of nutritional needs are given enterally.

Methods

Enteral nutrition may be given into either the stomach or small intestine. The small intestine is safer as aspiration may complicate gastric feeding. There are several methods of access available. Choice of enteral access in a given patient depends upon the anticipated duration of treatment and other factors that should be considered on a patient by patient basis.

Short-term supplementation

These are patients in whom the anticipated need for nutritional support is for a relatively short period, often less than 6 weeks. A variety of commercially available feeding tubes can be used to access the stomach or small intestine. For patient comfort, soft small bore (7 to 9 French) tubes should be used. These tubes may be placed nasogastrically if adequate gastric emptying and an intact gag reflex is present. They may also be placed nasoenterically in patients with a higher risk of aspiration. In our institution, nasointestinal tubes are most commonly used and are placed in either the intensive care units at the bedside or in the radiology department using fluoroscopic guidance. They may also be placed endoscopically, with the aid of pH guidance, or in the more traditional ‘blind' fashion if the patient is alert and co-operative. Owing to the small size of these feeding tubes and the often compromised mental status of the patients, great care must be taken to confirm the enteral placement of the tube prior to use (see ‘Complications' below). In addition, these tubes tend to clog easily, especially with medications, and use should be limited as much as possible to nutritional products in order to prevent the need for frequent replacement.

Long-term supplementation

Patients in whom the anticipated need for nutritional support is greater than 6 weeks may benefit from more permanent enteral access. Gastrostomy tubes may be placed either operatively or percutaneously with endoscopic guidance. This route of feeding requires that gastric emptying is present and is contraindicated by evidence of gastroesophageal reflux and absence of a gag reflex. One advantage of a gastrostomy tube is that feedings may be administered either continuously or as intermittent boluses, thus potentially simplifying care, especially in the outpatient setting.

Jejunostomy tubes are commonly placed operatively, either via laparotomy or with laparoscopic assistance. They may either be permanent (end Roux-en-Y type) or temporary (Witzel). In selected cases, a jejunostomy may be placed endoscopically or using the needle catheter technique. A small bore feeding tube may also be passed through an existing gastrostomy tube to create a functional jejunostomy. Jejunostomy feedings are given continuously, rather than as boluses.

Products

Many formulas have been developed for enteral supplementation. These vary in osmolarity, caloric content, protein complexity and density, and fat content. Typical formulas contain 1 to 2 kcal/ml and between 30 and 60 g of protein per liter.

Oral supplements

These are designed for patients who require nutritional supplementation and are able to tolerate oral feedings. These formulas must be palatable, a consideration which increases both cost and osmolarity. Examples of oral supplements commercially available in the United States are Ensure®, Sustacal®, and Carnation Instant Breakfast®.

Tube feedings

Blended formulas

These are pureed diets and are mainly used in patients with gastrostomies. Examples of these formulas available in the United States are Compleat-Modified® and Compleat-Regular®.

Polymeric formulas

These are iso-osmolar and are generally well tolerated by patients. They provide a complete diet with intact protein. Most are lactose free and have a caloric density of 1 kcal/ml. Some also contain fiber. Examples include Jevity®, Ultracal®, Osmolite®, and Isocal®.

High-caloric density formulas

These generally contain 2 kcal/ml and are useful in patients who have increased caloric needs and decreased volume tolerance. They are generally lactose free and contain intact protein. They are also hyperosmolar and can cause diarrhea when given directly into the small bowel. Available formulas in the United States include

Magnacal® and TwoCalHN®.

Monomeric formulas

These are low residue and require essentially no digestion. They are absorbed completely in the small intestine. The protein is in the form of amino acids with or without peptides. Examples include Vivonex TEN® and CriticareHN®.

Disease-specific formulas

These are intended for such conditions as renal and hepatic compromise. Renal formulations of enteric feedings include AminAid®, Nepro®, and Suplena®. They are hyperosmolar, have a caloric content of 2 kcal/ml, and contain essential l-amino acids and reduced nitrogen. Although they may be administered orally, they are relatively unpalatable are often best tolerated as tube feedings. Amin-Aid® is available in flavored forms for supplemental oral administration.

Hepatic formulations available in the United States include HepaticAid II®. HepaticAid II® is low in aromatic and sulphur containing amino acids and enriched with branched chain amino acids. The caloric content is 1 to 2 kcal/ml and the non-protein calorie to nitrogen ratio ranges from 148:1 to 800:1. HepaticAid II® is also exists a pudding for oral consumption.

Immunomodulatory formulas

Immunomodulatory formulas, such as Impact®, are intended for critically ill patients at risk for infections and for use in the immediate postoperative period. They contain increased protein, immunostimulatory amino acids, and lipids, and may decrease infectious complications.

Administration

As discussed above, nutrition should be supplied to the gut if at all possible. Even if total nutritional support cannot be achieved enterally, as much of the patient's needs as possible should be supplied via this route. The exact proportion of nutritional needs that must be supplied enterally to achieve the physiologic benefits of enteral nutrition is not known, but it may be only 20 per cent. Therefore, even in a patient dependent on parenteral nutrition, at least 20 per cent of the protein and caloric needs should be supplied enterally if possible.

Before the initiation of enteral nutritional support, baseline laboratory studies and patient parameters are obtained. Appropriate studies, such as an abdominal roentgenogram may be needed to confirm the position of feeding tubes placed non-operatively. The initial rate of feedings should be 40 ml/h, and if tolerated this can advanced by 20 ml every 12 h until the target goal is reached. Our practice is to start with either iso-osmolar or hypo-osmolar feedings to avoid osmotic diarrhea. The patient should be positioned with the bed at an angle of 30 to 45 degrees. If gastric feedings are being administered, residuals should be checked every 4 h and feedings held for a residual greater than 50 per cent of the total volume given over the previous 4 h.

After feedings are started, care should be taken to prevent the tube from plugging by flushing regularly with water and minimizing use of the feeding tube to deliver medication. Laboratory studies should be repeated at intervals and the patient should be weighed three times weekly in order to assess recovery from nutritional depletion and appropriateness of the current supplementation. Feedings should be delivered with a controlled, measurable feeding pump to regulate accurately the input of nutrients. In our institution, a preprinted form for standing orders in the form of a checklist delineating these practices has been developed (Fig. 5). This aids in the appropriate use of resources and serves as a protocol.

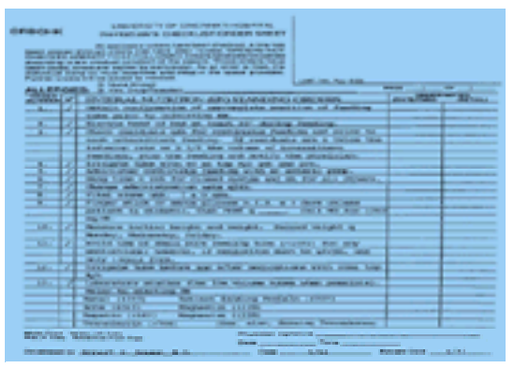


Fig. 5. Enteral Nutrition Standing Orders as employed at University Hospital, Cincinnati.

Gastric feedings

Bolus feedings may be used when nutrients are delivered into the stomach. From a practical standpoint, the stomach tolerates a higher osmotic load earlier better than the small bowel. When a patient is fed through a nasogastric tube or gastrostomy, feeding is begun with formulations that are made slightly hypo-osmolar by dilution with water. The osmolality is increased first, then the volume is increased until the target for nutritional support is reached.

Small bowel feedings

The small bowel lacks the diluting capacity of the stomach and the protective mechanism of the pylorus, so a higher osmotic load may not be well tolerated initially. When the small bowel is used, the volume of the feedings is increased prior to increasing the osmolality. Feedings into the small bowel should be continuous.

Complications

It is generally believed that enteral feedings are safer than parenteral nutrition. This belief may not be correct. A host of complications, both major and minor, can and do occur with enteral nutritional support. Avoidable complications can be minimized with attention to detail and adherence to protocols.

Diarrhea

This is the most common complication of enteral feeding and occurs in about 30 per cent of patients. Diarrhea may be caused by advancing hyperosmolar tube feedings too rapidly, but it may also be the result of infection. Clean technique during formula preparation and time limits on formula life should be followed in order to minimize bacterial overgrowth in the tube feedings as an infectious source. After excluding infectious etiologies, including antibiotic-associated enterocolitis, and metabolic sources, such as pancreatic insufficiency, diarrhea may be treated initially by decreasing the infusion rate. Antidiarrheal agents such as kaolin pectin, diphenoxylate, loperamide, or psyllium may also be used, but only if infectious etiologies have been excluded. If diarrhea persists and is untreated and feedings continue, pneumatosis and/or bowel necrosis and perforation may occur, resulting in mortality.

Catheter complications

Catheter-related difficulties are also common. Great care must be taken in the non-operative placement of small-bore nasoenteric feeding tubes. Many of these soft tubes utilize a stylet to make the tube more rigid in order to aid insertion. This can lead to inadvertent perforation into the pleural cavity if the tube is passed into the tracheobronchial tree during placement. Care must also be taken to confirm proper location of the tube prior to use. Insufflation of air into the tube with auscultation over the abdomen by itself is insufficient to exclude non-enteral placement of a feeding tube. A roentgenogram should always be obtained prior to initiation of feeding if the tube was placed blindly. In our institution, the use of fluoroscopy has minimized complications from nasoenteric feeding tube placement.

As discussed previously, frequent flushing and avoiding use of the feeding tube for medication administration more than absolutely necessary can minimize clogging of feeding tubes and prolong their life span. If a tube does become clogged, it can often be cleared by gently flushing it with warm water or small amounts of carbonated

beverages, utilizing a small syringe to generate additional pressure.

Not infrequently, feeding tubes of all types will be inadvertently dislodged or displaced. Nasoenteric tubes are replaced in the manner with which they were first introduced. If a gastrostomy or jejunostomy has recently been placed, operative replacement may be required. If a well-established tract exists, a new feeding tube may be introduced gently along the existing tract. If any doubt exists regarding the position of the tube, it must be confirmed using a radiographic study with water-soluble contrast.

Aspiration

Pneumonitis and pneumonia resulting from aspiration can be a devastating complication of enteral feeding. This may be a major cause of morbidity and mortality from enteral feedings and it should be remembered that one death from aspiration pneumonia can equal the cumulative mortality occurring in 2 or 3 years in a parenteral nutritional program. The risk of aspiration can be minimized by selection of the appropriate route for feeding as discussed previously, and by attention to proper patient positioning.

Intolerance

The symptoms of feeding intolerance are cramping, diarrhea, abdominal distension, nausea, and emesis. These may be relieved by diluting feedings or decreasing the infusion rate. The sudden appearance of feeding intolerance may indicate the onset of sepsis.

Metabolic complications

Hyperglycemia

This may occur with the onset of feedings or, like sudden feeding intolerance, may herald the onset of sepsis. After initiating feedings, blood glucose levels should be measured twice daily for at least 3 days. Hyperglycemia can be treated with frequent assessment of serum glucose levels and appropriate administration of short-acting insulin or an intravenous insulin drip.

Hyperosmotic non-ketotic coma

This results from untreated hyperglycemia. The obligatory renal solute load leads to dehydration and a hyperosmotic state. Treatment consists of restoration of proper intravascular volume and osmolarity and control of the precipitating hyperglycemia.

Parenteral nutrition

Indications for parenteral nutrition

If enteral nutrition is not possible in the malnourished or at-risk patient, the parenteral route must be utilized. Parenteral nutrition may be used for either primary or supportive therapy. Efficacy for the use of parenteral nutrition has been established in some circumstances, but not in others. Other areas remain under investigation. The indications for total parenteral nutrition are summarized in Table 3 and are discussed in greater detail below.

Efficacy established	*Efficacy established*
Gastrointestinal fistula	Radiation enteritis/chemotherapy toxicity
Short bowel syndrome	Hyperemesis gravidarum
Acute renal failure	Prolonged ileus
Hepatic insufficiency	
Efficacy not established	*Efficacy not yet established*
Inflammatory bowel disease	Preoperative therapy
Anorexia nervosa	Cardiac cachexia
	Pancreatitis
	Respiratory insufficiency
	Under investigation
	Cancer
	Sepsis

Table 3 Indications for parenteral nutrition

Role in primary therapy

Efficacy established

Gastrointestinal-cutaneous fistula

In patients with fistulas, parenteral nutrition allows bowel rest, decreases fistula output, and improves nutritional status. Parenteral nutrition increases the rate of spontaneous fistula closure, thus avoiding operations in some patients. Even if spontaneous closure does not occur, the patient will be in improved nutritional status and overall condition for operative repair. Parenteral nutrition has not been shown to decrease overall mortality in centers experienced in the care of patients with fistulas, but it probably does lead to decreased mortality in patients treated in other institutions. If possible, a portion of the patient's nutritional needs should still be given enterally in order to obtain the benefits of enteral nutrition. Some patients with distal gastrointestinal fistulas can be supported with low residue enteral diets.

Short bowel syndrome

This may be the result of major enterectomy for volvulus or mesenteric thrombosis or from repeated bowel resection for conditions such as Crohn's disease. As a general guide, at least 60 cm of small bowel are required for a patient to survive eventually solely with enteral nutrition (this is reduced to 45 cm if the ileocecal valve is preserved). For many patients, there is simply no alternative to long-term total parenteral nutrition. Indeed, the advent of total parenteral nutrition has allowed many patients who would have previously perished to survive at home. Patients with less severe disease and a greater remnant length of small bowel may require parenteral nutrition on a less permanent basis until adaptation of the remaining bowel occurs.

Acute renal failure

Parenteral nutrition has been shown to decrease mortality in the setting of acute renal failure. In transient acute tubular necrosis, parenteral nutritional support may result in earlier recovery from renal failure. Parenteral formulations of essential amino acids and high concentrations of dextrose can provide maximum calories in a minimum volume, slow increases in blood urea nitrogen, and minimize the proteolysis seen in renal failure. This may also delay the need for dialysis and can be especially advantageous in the patient in whom the ultimate need for dialysis is unclear or who is not hemodynamically stable enough to tolerate dialysis. Once the need for dialysis has been established, many experts suggest adding non-essential amino acids to the parenteral nutrition formulation.

Hepatic insufficiency

Aggressive nutritional support can also decrease mortality in patients with hepatic failure. These patients require protein to ameliorate their malnutrition, but they are also often intolerant of protein due to encephalopathy. This observation led to the development of the false neurotransmitter hypothesis of hepatic encephalopathy, which proposed that decreased hepatic function leads to plasma amino acid imbalances, with increased levels of aromatic amino acids and decreased levels of branched chain amino acids. This imbalance in turn leads to altered neurotransmitters within the central nervous system, resulting in hepatic encephalopathy. Use of solutions enriched with branched chain amino acids and decreased aromatic amino acids has been shown in several studies to be as or more effective than neomycin

or lactulose in the treatment of hepatic encephalopathy.

Efficacy not established

Inflammatory bowel disease

Patients with ulcerative colitis or Crohn's disease may develop exacerbations of their disease with severe diarrhea, bleeding, and inability to tolerate oral nutrition. Although randomized prospective trials have not evaluated the efficacy of parenteral nutrition in this setting, parenteral nutrition is useful in these patients to allow bowel rest and improvement in nutritional status. Crohn's disease confined to the small intestine appears to respond especially well to this approach, with remissions seen in 75 per cent of patients. The course of ulcerative colitis is not affected, but bowel rest and parenteral nutrition may allow optimization of a patient's condition prior to surgical cure with total colectomy and ileoanal pullthrough.

Anorexia nervosa

Patients with this condition may literally starve themselves to death due to an altered body self-image. Although patients with anorexia nervosa may require nutritional support in conjunction with psychotherapy for successful treatment, they are prone to self-sabotage and are difficult to treat. The use of parenteral nutrition for this condition has not been fully evaluated and is presently recommended only for patients with severe malnutrition in whom enteral nutrition is not possible.

Role in secondary therapy

Efficacy established

Radiation enteritis/chemotherapy toxicity

Radiation treatments or chemotherapy may result in a severe enteritis that precludes oral intake. Parenteral nutritional support allows the patient to survive until the gut mucosa recovers. In some cases, patients with chronic radiation enteritis may develop multiple strictures necessitating prolonged home total parenteral nutrition.

Hyperemesis gravidarum

In this setting, parenteral nutrition may be needed to supply nutrients for both the mother and the developing fetus. Maternal nutritional status should be monitored closely with a weight gain of at least 1 kg/month in the second and third trimesters suggested to ensure the nutritional well being of the fetus. Fetal ultrasound studies may be useful to monitor fetal growth.

Prolonged ileus

Patients may develop a prolonged ileus (greater than 5 days) following abdominal operations. If nutritional needs cannot be achieved orally, parenteral nutrition is useful to maintain nitrogen balance and avoid excessive catabolism until the enteral route can again be utilized.

Efficacy not yet established

Preoperative support

In malnourished patients, the benefits of nutritional supplementation must be weighed against the increased rates of infectious complications seen in several studies. The patients that appear to benefit the most are the severely malnourished, in whom a decrease in surgical complications is seen when compared with controls. As discussed earlier, while it is possible to identify the group of patients at risk for complications due to poor nutrition, it is more challenging to identify individual patients at risk. The optimal duration of pre-operative nutritional repletion is unknown, but it is likely to be between 5 and 7 days.

Cardiac cachexia

Cardiac muscle is not immune to the effects of malnutrition. As with any other muscle, poor nutrition leads to decreased performance. Although current evidence is mainly anecdotal, pre-operative nutrition may help to restore cardiac function, but probably requires prolonged nutritional support.

Pancreatitis

Parenteral nutrition does not alter the course or outcome of acute pancreatitis. Most episodes of pancreatitis are self-limited and do not mandate nutritional supplementation. In severe episodes of pancreatitis, nutritional therapy is needed for supportive purposes. If enteral feeding is not possible due to exacerbation of symptoms, parenteral nutrition may be necessary.

Respiratory insufficiency

Critically ill patients may rapidly develop malnutrition, leading to increased mortality. This can lead to decreased respiratory muscle strength and pulmonary defense mechanisms. Nutritional support in these patients should be by the enteral route if possible. If not, parenteral nutrition is indicated. Nutritional components have also been shown to affect directly respiratory function. The branched chain amino acids appear to increase ventilatory drive.

In patients with respiratory failure, overfeeding with carbohydrates may lead to increased carbon dioxide production, and in rare cases this may make ventilator weaning more difficult. This is especially common in nutritionally depleted septic patients who are suddenly given a large glucose load. Monitoring the patient's respiratory quotient with indirect calorimetry, as discussed previously, can help avoid this situation.

Indications under investigation

Cancer

Patients with malignancies often suffer cachexia and weight loss even with a relatively small tumor burden. In addition, chemotherapy, radiation therapy, and operative intervention may contribute to poor nutritional status. Although nutritional support may be helpful in this group of patients, there is justifiable concern that tumor growth may be stimulated. Some evidence has indicated that nutritional supplementation in patients undergoing radiation or chemotherapy may actually lead to decreased recurrence-free interval. Currently, nutritional supplementation is indicated only in patients who are so severely malnourished that they would be less likely to survive an upcoming operation or course of radiation or chemotherapy. In the future, formulations may be available that contain medications that limit tumor growth and uptake of nutrients by tumors.

Sepsis

As with the patient with hepatic disease, altered amino acid levels may be seen in septic patients. The branched chain amino acids are used preferentially by the skeletal muscle for fuel, leading to decreased levels, with relatively increased levels of the aromatic and sulphur containing amino acids. Use of solutions enriched with branched chain amino acids should theoretically decrease proteolysis and improve nitrogen balance, but clinical studies have thus far demonstrated only marginal improvements in these parameters, usually in critically ill patients, and no differences in clinical outcomes.

Route of delivery

Peripheral parenteral nutrition

This route is utilized mainly in institutions without nutritional support teams. It may be useful in situations in which the definite need for parenteral support has not been established or nutritional support is needed for only a very brief period of time. Although it may be difficult to avoid peripheral vein thrombophlebitis, meticulous

attention to catheter insertion and care and use of large veins for access may minimize this. It is very difficult to supply the nutritional needs of a severely ill patient by this route.

Central parenteral nutrition

This is the preferred route in most institutions. Catheters placed into the central venous system terminate in the vena cava. Catheters may be placed either percutaneously or using operative open techniques.

Temporary catheters

These catheters, which may be percutaneously placed at the patient's bedside, are appropriate for most patients requiring parenteral nutrition. The insertion of a catheter for nutritional use is never an emergent procedure. Proper preparation of the patient will decrease complication rates. Prior to catheter insertion, a discussion of the risks, benefits, possible complications, and alternatives should be carried out with the patient or his surrogate and proper written informed consent should be obtained. Care should be taken that the patient is adequately hydrated and lightly sedated. Possibly coagulopathies must be addressed and corrected. Ample assistance should be available. Catheters may be placed utilizing the subclavian or the internal jugular approaches. The venous anatomy of the thoracic inlet is shown in Fig. 6.

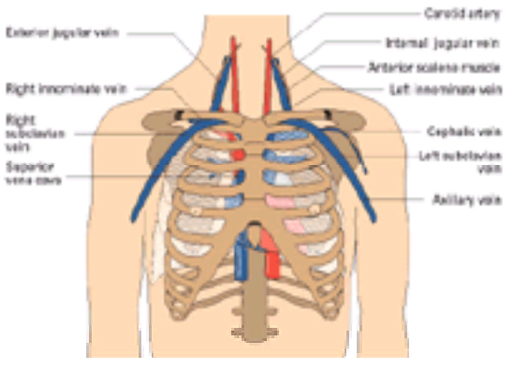


Fig. 6. The venous anatomy of the thoracic inlet. The subclavian veins lie in a transverse fashion that is more pronounced on the left. Insertion of catheters into the left subclavian vein is usually easier for right-handed individuals.

The subclavian approach is preferred for insertion of catheters for administration of parenteral nutrition, as it allows easier care of the catheter site and appears to be more comfortable for many patients. The patient is placed in the supine position with a small towel roll between the scapulae and the bed is placed in 15 degrees in the Trendelenburg position. The patient's arms are placed at his side with his head turned slightly away from the insertion site. The upper anterior chest and neck are carefully prepped and draped. The point of skin insertion is 1 cm caudad and 1 cm medial to the midpoint of the clavicle, with the needle entering the vein at approximately the junction of the medial and middle thirds of the clavicle (Fig. 7). The anticipated track is infiltrated with local anesthesia and the vein is located with a 22-gauge needle, passing the needle at an angle of no more than 10 to 15 degrees from the horizontal toward a point one fingerbreadth superior to the suprasternal notch. After locating the vein with the smaller needle, the 18 gauge Seldinger needle is passed along the same route into the subclavian vein until a ‘pop' is felt and free venous blood return is noted. The patient is then asked to perform a Valsalva maneuvre, the syringe is removed from the needle, and the soft J-tipped guidewire is passed through the needle. The guidewire should thread easily. The needle is then withdrawn, the track is dilated, and the catheter is threaded over the guidewire. The wire is removed and the catheter is secured in place with a single 3-0 or 4-0 suture. The catheter is then capped, flushed, and a sterile dressing is applied.

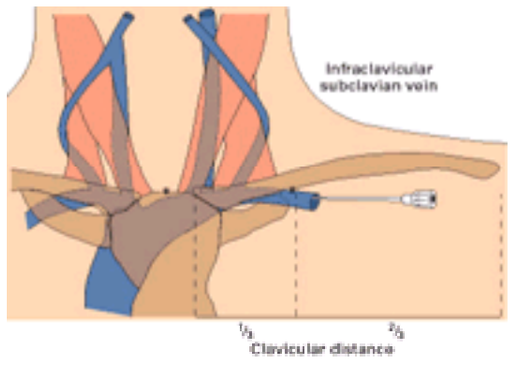


Fig. 7. Cannulation of the left subclavian vein. The vein is entered at the junction of the medial and middle thirds of the clavicle.

Following catheter insertion, a chest roentgenogram to confirm positioning is obtained. The catheter tip should reside in the superior vena cava or at the junction of the vena cava and the subclavian or innominate vein. If the catheter tip lies in the opposite subclavian vein or has traveled into the internal jugular vein in a retrograde fashion, it should be repositioned to avoid an increased risk of thrombosis. Likewise, the presence of the catheter tip in the right atrium is also unacceptable as the catheter may erode through the atrium.

Catheters may be placed into the internal jugular vein using either the middle or posterior approach. Preparation and positioning of the patient is similar to the subclavian approach. In the middle approach, the point of initial insertion is located near the apex of the triangle formed by the two heads of the sternocleidomastoid muscle (Fig. 8). The procedure is carried out as described above, with the needle passed parallel to the clavicular head at a 45-degree angle toward a point between the suprasternal notch and the ipsilateral nipple. With the posterior approach, the internal jugular vein is approached along a path beginning three fingerbreadths above the clavicle along the posterior border of the sternocleidomastoid (Fig. 9). The needle is passed long the posterior belly of the muscle toward the suprasternal notch until the vein is entered.

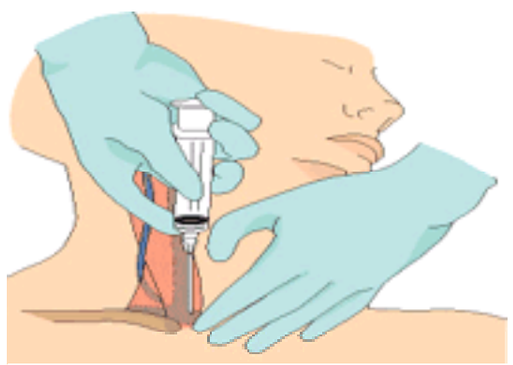


Fig. 8. The middle approach to cannulation of the internal jugular vein. The vein is entered by passing the needle along the medial border of the clavicular head of the sternocleidomastoid muscle toward the ipsilateral nipple.

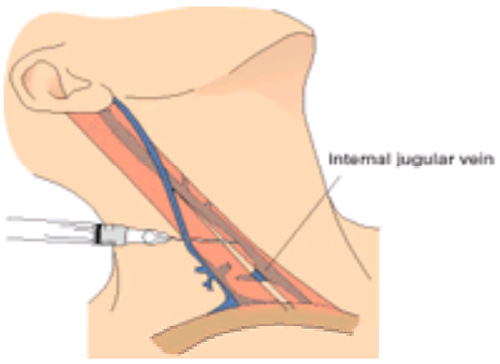


Fig. 9. The posterior approach to cannulation of the internal jugular vein. The needle is passed along the posterior portion of the sternocleidomastoid toward the suprasternal notch.

Indwelling catheters

Patients with a prolonged need for parenteral supplementation frequently benefit from more permanent indwelling catheters. These catheters are placed in the operating room and tunneled to an exit site in the skin (Fig. 10). They may be placed percutaneously, or via cutdown of branches of the axillary vein. The latter approach is especially advantageous in that the risk of pneumothorax is avoided and it provides a superior cosmetic result.

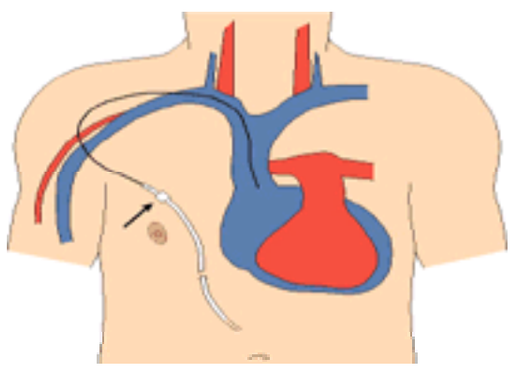


Fig. 10. Long-term parenteral access may be obtained using implantable tunneled catheters. The cuff is usually located closer to the skin entrance. These catheters may be placed either percutaneously or through cutdown of tributaries of the cephalic, axillary, external, or internal jugular veins.

Composition of formulas

The components of various parenteral nutrition formulas available at the University Hospital, Cincinnati are shown in Table 4. The various components are discussed below.

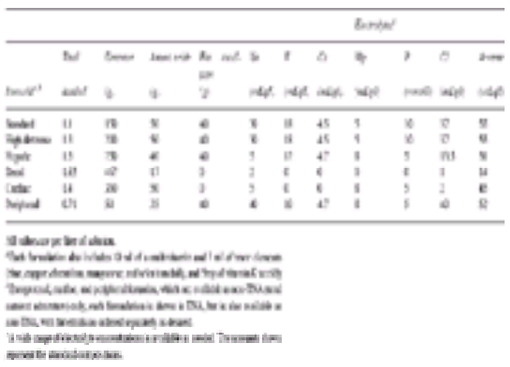


Table 4 Components of parenteral nutrition formulas available at the University Hospital, Cincinnati. The dextrose, fat, amino acid, and electrolyte compositions vary in the specialized formulas. Each of these components can also be varied as needed for each individual patient.

Carbohydrates

In the United States, carbohydrates are supplied exclusively as dextrose. The usual concentration is 15 per cent, but this may be as high as 47 per cent in renal formulations or as low as 5 per cent in peripheral formulations.

Amino acids

Amino acids may be supplied as a balanced composition or a special compositions and formulations in disease-specific formulas. Amino acids usually account for 5 per cent of the parenteral nutrition solution. Modified compositions of amino acids are available in hepatic and renal parenteral nutrition formulations.

Lipids

Lipid emulsions are available as 10 to 20 per cent solutions. Infusion of 100 ml of a 10 per cent solution weekly is adequate to prevent essential fatty acid deficiency. It is considered safe to provide 20 to 60 per cent of total calories as lipids. With the use of total nutrient admixture, lipids are supplied in the same solution as dextrose and amino acids, eliminating the need for separate lipid administration.

Vitamins and trace elements

Administration of vitamins and trace elements can generally prevent deficiencies of these substances. Vitamin K should also be administered weekly in the patient receiving parenteral nutrition.

Insulin

Many patients receiving parenteral nutrition either have or develop abnormal carbohydrate metabolism. Hyperglycemia can be treated by intermittent injection of short-acting insulin preparations. If the patient is consistently and predictably hyperglycemic, insulin can be added to the parenteral nutrition formulation as necessary.

Electrolytes

The range of electrolytes administered in formulations is variable. Standard amounts of electrolytes are shown in Table 4. The amounts of these substances should be modified based on each individual's situation and monitoring of serum levels.

Renal formula

Acute renal failure patients are catabolic, often from both the renal failure itself and associated critical illness. Parenteral nutrition in this setting is indicated in patients with acute renal failure who are not yet being dialysed. The goal of the renal formulation is to address metabolic needs while minimizing accumulation of fluid, urea, and electrolytes. The electrolyte composition offsets abnormalities commonly seen in acute renal failure. The base is 47 per cent dextrose, allowing caloric administration in a limited volume. Once chronic dialysis has begun, some advocate change to a standard formulation.

Hepatic formula

This special formulation may be useful for patients with acute hepatic failure or acute-on-chronic hepatic disease who develop hepatic encephalopathy. Their requirement of 1.1 g of protein equivalent per kg is almost double that of the usual population. Hepatic formula parenteral nutrition is enriched with 35 per cent branched chain amino acids and reduced amounts of aromatic and sulphur-containing amino acids. Patients with mild encephalopathy (grade I) may tolerate restricted amounts of standard amino acid infusions, but will probably not reach nitrogen equilibrium. The modified mixture is indicated in patients with grade II encephalopathy (impending stupor) or higher, and in all patients in whom it is important to reach nitrogen equilibrium.

Administration

Initiation

After the infusion catheter has been placed and its position confirmed with a roentgenogram, infusion of the chosen parenteral nutrition formulation is begun at 40 to 50 ml/h. At the University of Cincinnati Medical Center, the Department of Parenteral and Enteral Nutrition has developed specialized order forms to ensure that orders for parenteral nutrition are accurate and standardized. These forms are available for both standard (Fig. 11) and specialized formulations (Fig. 12). If blood glucose levels are controlled, the rate of infusion can be increased by 20 to 25 ml every 8 to 24 h until caloric needs are met. Renal formulations should begin at 30 ml/h due to the higher glucose content, and are then advanced by 10 ml/h every 24 h.

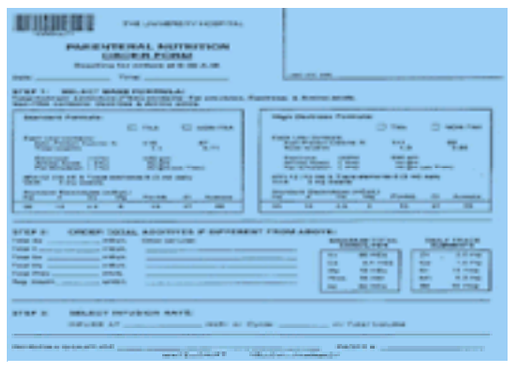


Fig. 11. Order form for standard and high dextrose parenteral nutrition formulations at the University Hospital, Cincinnati.

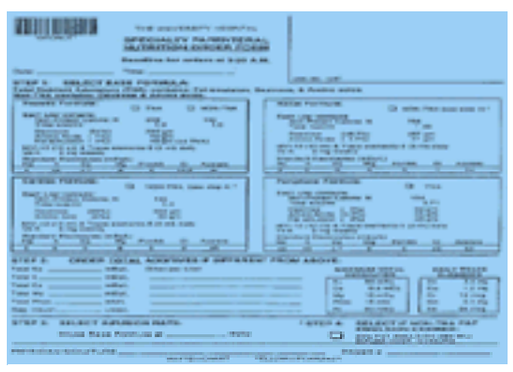


Fig. 12. Order form for speciality parenteral nutrition formulations at the University Hospital, Cincinnati.

Monitoring

Before initiation of parenteral nutrition, weight, vital signs, and baseline laboratory studies are obtained, including transferrin and retinol-binding protein. After initiation, vital signs and intake and output are measured every 8 h. Serum glucose is measured twice daily. Weight is measured three times each week. Selected laboratory studies are repeated weekly and serum electrolytes are obtained at least twice each week. To assist in the proper administration and monitoring of parenteral nutrition, a preprinted order form can be employed (Fig. 13). If the patient is markedly catabolic or the serum electrolytes are not stable, these parameters should be measured more often in order to prevent electrolyte abnormalities.

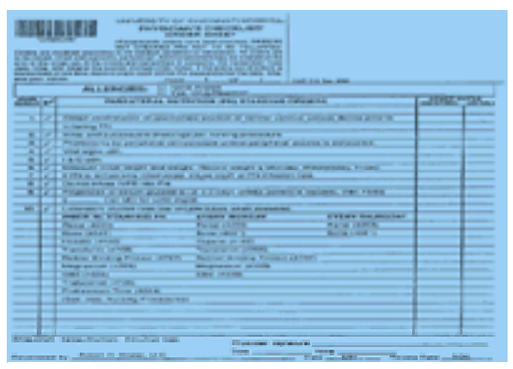


Fig. 13. Standing orders for parenteral nutrition from the University Hospital, Cincinnati.

Complications of parenteral nutrition

Technical complications

Complications related to catheter insertion can be minimized with experience and meticulous placement. Most of these difficulties are related to nearby structures in the

thoracic inlet.

Pneumothorax

This complication occurs in 5 to 6 per cent of central catheter insertions and can be minimized by careful attention to technique and adequate hydration of patients. Its occurrence is more common in elderly and cachectic patients, in whom an internal jugular catheter may be indicated.

Arterial lacerations

These are relatively rare and can be minimized by keeping the angle of the needle no more than 10 to 15 degrees from the horizontal and by use of a small finder needle rather than the larger 18-gauge needle to locate the vein.

Hemothorax

This may result from leakage of blood from the subclavian vein and may require tube thoracostomy.

Hydrothorax

Infusion of fluid through a catheter placed into the pleural cavity or mediastinum may result in hydrothorax.

Thoracic duct injury

This is a relatively uncommon complication, but it may occur with left-sided cannulation.

Air embolism

This may result from improper technique and can be minimized by ensuring that the patient is properly hydrated, placed in Trendelenburg position, and the hub of the needle is covered at all times.

Guidewire embolism

This can result from manipulating the guidewire excessively through the 18-gauge needle, resulting in a portion of the guidewire embolizing distally. Retrieval under fluoroscopy may be necessary. Care must also be taken never to release the guidewire, even momentarily, during placement.

Subclavian vein thrombosis

This late complication occurs in 5 to 10 per cent of patients. Signs and symptoms include pain at the base of the neck and upper arm swelling. If thrombosis is suspected, the catheter must be removed and the diagnosis confirmed with imaging studies. Thrombolysis should be considered. The patient is then anticoagulated for 6 months to ensure patency of the vessel.

Septic thrombosis

This complication is a potentially life-threatening event. If no improvement occurs with anticoagulation and antibiotics, Fogarty catheter embolectomy or excision of the subclavian vein may be required.

Septic complications

These events are probably responsible for most parenteral nutrition related deaths. The incidence of infection can be reduced to less than 2 per cent in a well-run program with careful attention to antisepsis. The incidence of catheter sepsis in our institution decreased from 27 per cent to 0.6 per cent with the establishment of a hyperalimentation team and rigid adherence to protocols. Higher rates of sepsis may be seen with multiple port catheters, especially pulmonary arterial catheters. If a multiple port catheter is required for non-nutrition-related vascular access, one port should be dedicated solely to parenteral nutrition infusion.

Catheter sepsis is directly related to catheter care. Quantitative cultures of pericatheter skin have shown a direct relationship with catheter sepsis and the degree of contamination, with the presence of 10^3 or greater organisms correlating to catheter infection.

Bacteremia from distant sites may also lead to seeding of the fibrin sheath that forms around indwelling catheters. This is more common with gram-positive organisms than with gram-negative organisms. If a patient has established gram-positive bacteremia, the catheter should be replaced over a guidewire, but more likely will need to be removed. This is not necessary with Gram-negative organisms, as they implant less readily, and treatment with antibiotics may suffice.

Fungemia in the patient receiving parenteral nutrition is a very serious complication. *Candida*, the most common fungal pathogen, probably enters the bloodstream through the gastrointestinal tract. *Candida* colonization, defined as positive cultures at two separate sites (i.e. urine and skin), should be treated with antifungal agents. Established *Candida* infection necessitates removal of the catheter and cessation of all intravenous nutritional regimens.

A suggested algorithm for management of catheter sepsis is shown in Fig. 14. If a sudden fever occurs in a previously afebrile patient, the parenteral nutrition solution should be stopped and the solution and tubing should be cultured and changed. Blood cultures should be obtained, and assessment for other sources of fever should ensue. If the catheter site appears infected, the catheter should be removed. If gram-positive bacteremia is present or there are no other sources of infection, the catheter may be replaced over a wire and the catheter tip is cultured. If the catheter tip culture is positive, the central line should be removed. It may be replaced after waiting at least 24 h if the patient is afebrile and on antibiotics.

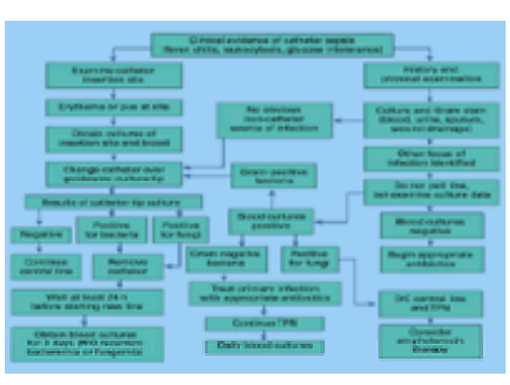


Fig. 14. Suggested algorithm for the diagnosis and treatment of catheter related sepsis.

Metabolic complications

Electrolyte abnormalities

These are relatively common and result from excessive losses followed by inadequate replacement, excessive electrolyte administration, or inadequate secretion. Anabolic patients may require relatively large amounts of potassium, magnesium, and phosphorous. Electrolyte abnormalities can be minimized by frequent monitoring

of serum electrolyte levels, especially in critically ill patients or those beginning parenteral nutrition.

Hyperglycemia

Initiation of parenteral nutrition formulations too rapidly may cause elevated blood glucose levels. This can be minimized by beginning the infusion at 40 ml/h of 25 per cent glucose based solution and increasing the infusion rate by 20 ml/h every 24 h, depending on glucose tolerance. Glycosuria may be seen in the first 2 days even in patients with normal glucose tolerance. Hyperglycemia can be treated with intermediate administration of short-acting insulin based on serum blood glucose levels. Insulin may also be added to the parenteral nutrition formulation. Sudden new onset hyperglycemia should prompt consideration of infectious etiology, as hyperglycemia may be seen 24 h prior to the onset of sepsis.

Hypoglycemia

This may be caused by excessive insulin administration or a sudden slowing of parenteral nutrition infusion. It can be corrected by modifying insulin regimens and/or administration of exogenous glucose in symptomatic patients. If parenteral nutrition must be discontinued suddenly for any reason, hypoglycemia can be prevented by the intravenous administration of 5 per cent dextrose solutions.

Liver dysfunction

Hepatic dysfunction is common in patients on parenteral nutrition. The etiology is not well understood and is probably multifactorial. These derangements include cholestasis, steatosis, elevated liver function tests (with the exception of bilirubin), and hepatomegaly. Most patients manifest a combination of these abnormalities. Elevated bilirubin levels are uncommon except as a result of sepsis, except in children where they may accompany cholestasis or cirrhosis.

Deficiency states

Deficiencies of fatty acids and trace elements may occur in patients on parenteral nutrition. Essential fatty acid deficiency may manifest itself as dry, flaky skin and alopecia. It may be prevented by parenteral administration of 4 to 6 per cent of daily calories as safflower or soybean oil emulsion, or oral administration of 25 to 50 ml/day of margarine or corn, safflower, or sunflower oil.

Zinc deficiency may occur in individuals with excessive diarrhea, fistula losses, or extreme catabolism. It is characterized by impaired wound healing, taste disturbances, darkened skin creases, and a perioral pustular rash. In patients with normal stool losses, 3 to 6 mg of elemental zinc are required per day. This amount increases to between 12 and 20 mg in individuals with excessive diarrhea or short bowel syndrome.

Chromium deficiency may occur in patients on long-term parenteral nutrition with negligible oral intake. Chromium is required for normal glucose utilization. Deficiency results in sudden onset hyperglycemia that is difficult to control. Peripheral neuropathy and encephalopathy may also occur. To treat chromium deficiency, 150 µg of chromium is given daily for 7 days.

Biotin deficiency may occur, but is very rare and almost never occurs in patients taking any portion of their daily requirements by mouth.

Nutritional pharmacology

The term ‘nutritional pharmacology' should probably be credited to J Wesley Alexander of the University of Cincinnati and reflects the use of specific nutrients to modify the host response. Early examples of nutritional pharmacology, which antedate the use of the term, include the use of special formulations of enteral and parenteral nutrition modified for patients in hepatic and renal failure. More recently, investigation has centered on the use of nutrients to augment and modify host defenses in critically ill patients. Other work suggests that nutrient modification may prolong transplant allograft survival. This represents one of the most exciting areas of nutritional investigation today.

Further reading

Alexander JW *et al.* Beneficial effects of aggressive protein feeding in severely burned children. *Annals of Surgery* 1980; **192**: 505. [This study demonstrated that early, aggressive enteral feeding might confer a survival advantage in severely burned children.]

Bower RH *et al.* Early enteral administration of a formula (Impact®) supplemented with arginine, nucleotides, and fish oil in intensive care unit patients: Results of a multicenter, prospective, randomized clinical trial. *Critical Care Medicine* 1995; **23**(3): 436. [This multicenter trial examined the use of an immunomodulatory enteral nutritional formula in critically ill patients, showing reductions in hospital stay and acquired infections.]

Buzby GP (Veterans Affairs Total Parenteral Nutrition Cooperative Study Group). Perioperative total parenteral nutrition in surgical patients. *New England Journal of Medicine* 1991; **325**: 525. [This multicenter trial demonstrated that 7 to 10 days of preoperative parenteral nutrition in severely malnourished patients decreased operative septic complications.]

Dudrick SJ, Wilmore DW, Vars HM, Rhoads JE. Long-term total parenteral nutrition with growth, development, and positive nitrogen balance. *Surgery* 1968; **64**: 134. [This paper was one of the first to describe normal growth achieved with the use of high-dextrose total parenteral nutrition delivered centrally (in puppies).]

Fischer JE. *Nutrition and metabolism in surgical patients.* Little, Brown and Company, Boston, 1996. [This text addresses the basic science and clinical principles of metabolism and both enteral and parenteral nutritional support.]

Fischer JE, Rosen HM, Ebeid AM, James JH, Keane JM, Soeters PB. The effect of normalization of plasma amino acids on hepatic encephalopathy in man. *Surgery* 1976; **80**: 77. [This represents an early use of nutritional pharmacology to modify hepatic encephalopathy by correcting amino acid profiles with a solution rich in branched chain amino acids.]

Kudsk KA *et al.* Enteral vs parenteral feeding: effects on septic morbidity after blunt and penetrating abdominal trauma. *Annals of Surgery* 1992; **215**: 503. [Enterally fed trauma patients were found to have a decreased incidence of infection, pneumonia, and intra-abdominal abscess.]

Moore FA *et al.* Early enteral feeding, compared with parenteral, reduces postoperative septic complications: The results of a meta-analysis. *Annals of Surgery* 1992; **216**: 172. [This meta-analysis comparing enteral versus parenteral nutrition in high-risk surgical patients found a decrease in infectious complications in blunt trauma patients, but not penetrating trauma patients or those undergoing elective operations.]

Wilmore DW, Dudrick SJ. Treatment of acute renal failure with intravenous essential l-amino acids. *Archives of Surgery* 1969; **99**: 669. [This is one the earliest examples of nutritional pharmacology, which is the use of nutrients to affect to course of a disease.]

4.1 Surgical infections

Dietmar H. Wittmann and Robert E. Condon

Definition
Physiology
Local phase of infection
Systemic phase of infection: sepsis
Diagnosis
History and physical examination
Hematology, urinalysis, and radiology
Bacteriology
Therapy of surgical infections
General principles of therapy
Focus control, incision, and drainage
Antibiotic therapy
Risk factors
Long-term treatment in the intensive care unit
Scoring severity of infection
Skin and soft-tissue infections
Cellulitis, erysipelas, and phlegmon
Folliculitis, furunculosis, and carbuncle
Hydradenitis suppurativa
Bite wounds
Synergistic gangrene
Tetanus
Infections of the hand
Breast abscess
Pilonidal disease
Fox den disease
Perianal abscess
Perirectal abscess
Thoracic infection
Lung
Pleural infections
Infections of the heart
Abdominal infection
Peritonitis
Abdominal compartment syndrome
Biliary infections
Liver abscess
Pancreatitis, infected pancreatic necrosis, and pancreatic abscess
Appendicitis
Diverticulitis
Infections of the female genital tract
Infections of the genitourinary tract
Bone and joint infection
Post-traumatic osteitis
Osteomyelitis
Infection of the central nervous system
Subdural empyema
Brain abscess
Postoperative infections
Wound infection: unequivocal definitions
Postoperative peritonitis
Necrotizing fasciitis
Pulmonary infections
Mediastinitis
Infection of the urinary tract
Line infection: intravenous catheters, mediports
Antibiotic prophylaxis
Single-dose prophylaxis
The risk of infection
Specific recommendations
Other ‘clean' operations
Further reading

Definition

Surgical infections are those best treated by operative intervention, or those that follow surgical procedures and occur in the operative wound or at a distant site. They may be classified as primary, occurring spontaneously after tissue injuries, or as secondary, following accidental or planned surgical trauma such as an operation. Further useful definitions are given in Table 1.



Table 1 Definitions

Surgical differ from nonsurgical infections because in surgical infections there is always a source of infection (focus) that may be treated operatively to reduce or eliminate the bacterial inoculum, including necrotic tissue and adjuvants. An infectious focus, by definition, contains pathogenic micro-organisms, which may escape into the circulation or adjacent tissue. Typically, after the focus has been eliminated operatively, antibiotic therapy is either unnecessary or may be short term to treat small inocula of micro-organisms that escaped control by the operative actions at the source.

Physiology

The unique feature of all surgical infections is tissue necrosis. In post-traumatic surgical infection, necrosis is induced by mechanical or other physical trauma; in primary or spontaneous surgical infection a pathophysiologic process induces necrosis. Inflammation is the response to tissue necrosis, leading to the events visible at the surface that were well described by Celsus and refined by Galen as *rubor* (redness), *tumor* (swelling), *calor* (heat), *dolor* (pain), and *functio laesa* (loss of function). These symptoms describe the effects of the host responses, which, when controlled and properly regulated, result in elimination of necrotic material and prepare the way for tissue repair. The same mechanisms of inflammation are employed to eliminate invading micro-organisms.

Inflammation (see Section 1) is characterized by increased blood flow and increased vascular permeability. Phagocytes are recruited, preformed mediators released, and additional mediator compounds generated, concentrating biochemically active compounds and inflammatory cells. The magnitude of the inflammatory response and of its symptoms depends on the burden of tissue injury, and on the number and pathogenicity of the invading micro-organisms. If toxins or other bacterial products continuously destroy tissue, so exceeding the host's capacity to confine the challenge locally, the inflammatory process will continue and break host defense barriers to invade adjacent tissue and escape into circulation. The process may then elicit a systemic inflammatory response, resulting in multisystem malfunction, cell death, and sequential organ failure.

The early sequence of events in the inflammatory process results from the activation of interrelating systems that provide considerable duplication of biologic effects. For example, the peptides of the clotting, kinin, and complement systems may each produce vasodilatation. Duplication maintains the integrity of the response even in the face of dysfunction of one its components; it also provides the basis for additive and even deleterious effects recognized clinically as autoaggressive, multisystem organ failure (Fig. 1).

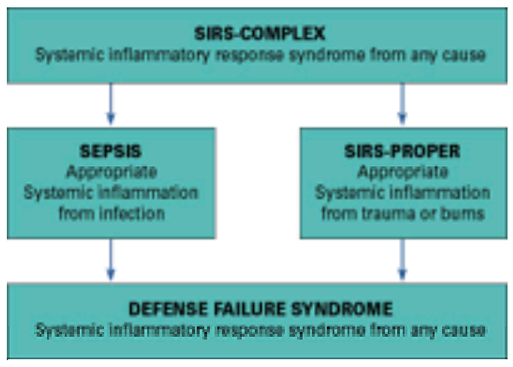


Fig. 1. A systemic inflammatory response may be induced by excessive trauma or infection. If the challenge persists or additional infectious challenge or trauma is added (e.g. badly timed surgical trauma), an appropriate inflammatory response may deteriorate into an inappropriate one, which is best termed ‘defense failure syndrome'.

Local phase of infection

Surgical infections, once initiated, take a relatively uniform course. First, there is local inflammation following the initial tissue injury. Macrophages may not be capable of phagocytosing all the dead cells and detritus, and any remaining necrotic tissue is an excellent nutrient medium for bacterial growth. Bacteria, in turn, release toxins that destroy additional tissue and thus fuel the infectious challenge. They may invade surrounding tissue slowly or rapidly, depending on their production of specific toxins (spreading factors). The host, in turn, answers with further inflammation in an attempt to confine the infection. If successful, dead tissue, foreign bodies, and micro-organisms are destroyed and removed, and a scar is formed.

If the extent of tissue injury and number of bacteria exceed the capacity of the host to terminate an infection locally, an abscess may form (Fig. 2). During the early phases of inflammation there is exudation of plasma, and thus fibrin, through widened spaces between endothelial cells and through open, injured vessels. As the infection progresses, this inflammatory process spreads centrifugally from the initial focus of infection, with macrophages and fibrin deposition attempting to confine the infection faster than bacterial toxins can destroy the tissue. The progress of the infection is eventually stopped through the formation of a pyogenic membrane, inside which dying phagocytes and bacteria release toxins that liquefy the contents of the abscess. The resulting high osmolarity attracts water, increasing the pressure inside the abscess capsule (apparent clinically when incising an abscess), while oxygen and nutrients diffuse poorly through the capsule, promoting anaerobic glycolysis. An abscess is therefore characterized by high pressure, low pH, and low oxygen tension, and is an ideal environment for the multiplication of anaerobic bacteria. It is poorly permeated by antibiotics: this is particularly true of aminoglycosides, which are inactive at the pH commonly found in an abscess. The best treatment of an abscess is drainage: local host defenses are so intensely concentrated around the abscess capsule that additional antibiotic therapy is rarely indicated.

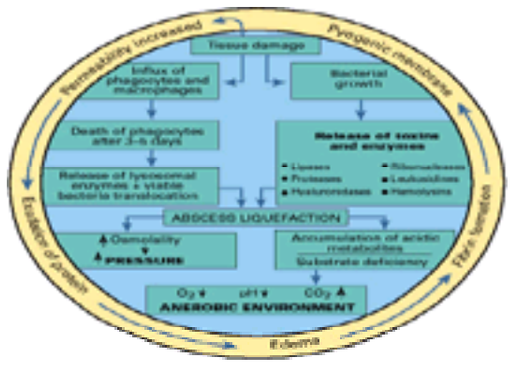


Fig. 2. Pathophysiologic events in an abscess: the pyogenic membrane forms as a result of the events of early inflammation; proteases and other enzymes are released from phagocytes and bacteria, leading to liquefaction of the contents of the abscess cavity; the increased osmolality attracts water, increasing the pressure within the abscess; the environment is basically anaerobic, with accumulation of acidic metabolites and an oxygen deficit—in this environment many antibiotics are ineffective.

Systemic phase of infection: sepsis

If local circumscription of infection is impossible either by removing the bacteria or by abscess formation, micro-organisms eventually invade the bloodstream and may reach distant organs. The presence of bacteria in the bloodstream (bacteremia) occurs transiently in healthy individuals too. In patients, bacteria often will be found on the intravascular portion of catheters. Nontoxin-producing, mostly nonmultiplying bacteria can sometimes be isolated by blood culture, but these cause no or only mild systemic symptoms. Bacteremia may, however, progress to systemic disease, especially in immunocompromised and postoperative patients. If the condition persists and is associated with multiplication of bacteria in the bloodstream, and large numbers of bacteria die as they are attacked by host defense mechanisms, then large quantities of bacterial cell-wall structures (endotoxins) are liberated and a serious state of infection termed sepsis ensues.

Sepsis is characterized not only by invasion and multiplication in the bloodstream of large numbers of bacteria but also by the potential for subsequent sudden overload of the host with endotoxins and cytokines, leading to septic shock. Key to this process is a small molecule – nitric oxide – that when excessively produced will block mitochondrial energy generation by inhibiting Krebs' cycle enzymes and cells are running out of fuel for life. Clinically we then observe sequential organ dysfunction.

Sepsis is the clinical, symptomatic state resulting from the host response to bacteremia. Liberated bacterial exo- and endotoxins are deleterious to many organ functions; equally, cytokine mediators of host defenses are potentially damaging if their downregulation fails. If not treated successfully, the patient may die immediately of septic shock or later following multisystem organ failure: 1 ng endotoxin/kg body wt results in irreversible shock and death within 2 h.

Clinical symptoms of sepsis include the following. Fever is usually is high, spiking, and accompanied by chills. Tachycardia accompanies or precedes the fever and is proportional to it. The total leukocyte count may not be particularly abnormal in sepsis and may even be low, due to the consumption of polymorphs. The differential

count is more reliable: there is always a shift to the left. Petechial lesions may be seen in the skin or conjunctiva of patients suffering from sepsis caused by streptococci, meningococci, or pseudomonads. Anemia secondary to hemolysis may appear rapidly when sepsis is due to staphylococci, pseudomonads, coliforms, or clostridia. During the initial, hyperdynamic phase of septic shock the peripheral vasodilation is explained by a circulatory response that aims to compensate for the inability of cells to use oxygen. The last stage of septic shock is hypodynamic, due to cell death. Shock is common in sepsis caused by Gram-negative organisms, but occurs relatively less often with Gram-positive infections. Metastatic abscesses, especially of the bone, brain, or spleen, are not unusual after a septic episode: any injured tissue is easily infected during sepsis. Diagnosis is aided by a high index of suspicion.

Thermolabile exotoxins are released by living bacteria, particularly the Gram-positive; thermostable endotoxins are released by all bacteria after death. Endotoxins are complex moieties of high molecular weight consisting of phospholipids, polysaccharides, and proteins derived from the outer cell wall, particularly of Gram-negative rods such as *Escherichia coli*. Clinically measurable effects of endotoxin include fever, consumptive coagulopathy, increased vagotonus, hyperglycemia followed by hypoglycemia, leucopenia or leukocytosis, increased plasma lipids, release of hepatic enzymes, thrombocytopenia, and a reduced serum iron.

Low doses of endotoxin primarily affect the reticuloendothelial system. Animal studies have shown a marked reduction in the clearance of particulates such as colloidal carbon during endotoxemia. Mediators such as collagenases, pyrogenic prostaglandins, and coagulation factors are released from macrophages; after 7 days, antibodies against endotoxin are produced. Endotoxins act directly on the hypothalamic temperature-regulation center to cause fever, reinforcing the activity of pyrogenic substances released from dying neutrophils. Erythropoiesis is shifted from the bone marrow to the spleen, resulting in leucopenia followed by leukocytosis after 2 to 6 h. In small doses, endotoxins increase phagocytic activity and bacterial killing. Thrombocytopenia, accompanied by aggregation and lysis of thrombocytes, results in the release of ADP, vasoactive amines, histamine, serotonin, and platelet factor III, which in turn may lead to consumptive coagulopathy. In the extrinsic coagulation system, endotoxins cause release of a tissue factor derived from macrophages, as well as platelet factors and thromboplastins. In the intrinsic system, factor XII (Hageman factor) is activated, leading to disseminated intravascular coagulation.

Endotoxin has a profound effect on metabolism. Initially it induces hyperglycemia, which is followed after several hours by hypoglycemia. Hyperlipidemia results from altered metabolism of free fatty acids, cholesterol, phospholipids, and triglycerides. Protein synthesis by the liver is stimulated; lactate dehydrogenase, transaminases, and phosphokinases are released, increasing their serum concentrations. Release of adrenocorticotrophic hormone, cortisone, and growth hormone is increased; thyrotropin and luteinizing hormone are not affected. Plasma iron and total iron-binding capacity are reduced. A vagotonic effect results in loss of thirst and appetite, stomach emptying is delayed, and diarrhea may occur.

Diagnosis

History and physical examination

The early accurate diagnosis of surgical infections is essential: delayed treatment can result in dissemination, overwhelming sepsis, and multisystem organ failure. The history and physical examination are the surgeon's most important diagnostic tools. The classic signs of tumor, rubor, calor, dolor and functio laesa are indicative of localized surgical infections. Clinical symptoms of systemic sepsis include disturbed sensorium, tachypnea, tachycardia, hypotension, fever, oliguria, and high-output heart failure. In postoperative patients, the sudden appearance of tachypnea and hypotension suggests Gram-negative sepsis. This condition has a potential mortality of 30 to 50 per cent, but early diagnosis and treatment markedly improves the chances of survival.

The entire body must be examined; all dressings should be removed. Inspection and palpation of a suspicious area may reveal the first three of the classical signs of infection. Removal of the dressing around an intravenous cannula may reveal purulent drainage or thrombophlebitis. Rectal examination may show tenderness and induration as signs of a developing pelvic abscess. Auscultation of the chest may reveal the presence of pneumonia before it is evident on a chest radiograph.

The patient should be examined for clues to the source of the infection, such as pain or redness in the surgical wound or at an intravenous infusion site, or purulent sputum, cough, pleuritic pain, rales, or dullness in the chest, diarrhea, dysuria, or flank pain. A foul-smelling odor may lead to the site of an anaerobic infection. Pain in the shoulder and an immobile diaphragm suggest a subphrenic abscess. A pelvic or prostatic mass on rectal examination may indicate an abscess, and headache or nuchal rigidity may indicate an infection of the central nervous system.

Hematology, urinalysis, and radiology

Most bacterial infections produce leukocytosis and, more importantly, a shift to the left in the differential count or a relative lymphopenia. This increase in the proportion of the more immature forms of polymorphonuclear leukocytes may signal infection before an abnormal total leukocyte count is evident. The differential count may also reveal lymphocytosis in viral infections, monocytosis in tuberculosis, eosinophilia in parasitic infections or hypersensitivity reactions (drug allergy), and toxic degranulation of leukocytes in acute bacterial infection. A low white count or a leukemoid response (a total white count of over 25 000 cells/mm^3) may be seen in sepsis in general, and in pneumococcal pneumonia, liver abscess or cholangitis, infected pancreatic necrosis, necrotic bowel, or retroperitoneal phlegmon in particular. Leukopenia is a sign of overwhelming bacterial infection and carries a bad prognosis. Viral infection, typhoid perforation of the bowel, or tuberculosis may also present with leukopenia. Anemia may be associated with infection caused by bacteria, such as *Clostridium perfringens*, group A streptococci, or coagulase-positive staphylococci, that produce hemolytic enzymes.

Routine chest films may reveal generalized or focal atelactasis, or may indicate intra-abdominal infection through signs of gastrointestinal leakage or free air identified under the diaphragm. In the investigation of patients with suspected intra-abdominal infection, flat, upright, and decubitus films may reveal a localized air–fluid level, suggesting an intra-abdominal abscess, or a spreading air-bubble pattern suggestive of infection with a gas-producing organism. Specialized radiologic procedures may be helpful in confirming the diagnosis of intra-abdominal abscess. These studies include ultrasonography and computed tomography (**CT**). Although a gallium scintiscan may be helpful in special circumstances, this examination is subject to appreciable error and is difficult to interpret in a patient who has had a recent operation.

Bacteriology

Observation of exudates and secretions such as wound drainage, urine, and sputum for odor, color, and consistency may be useful in diagnosis. Grape-like odors occur with pseudomonal infections, urea-like odors with *Proteus* infections, and feculent odors with anaerobic organisms such as *Bacteroides*, fusobacteria, clostridia, and peptostreptococci. A Gram stain offers the earliest clue to the cause of an infection, particularly when a specific monobacterial infection is suspected. Since surgical infections are mostly due to multiple infecting organisms that are obligate or facultative anaerobes, the Gram stain usually shows a variety of pathogenic bacteria. Note should be taken of the numbers of polymorphonuclear leukocytes on the slide (few, many, loaded) and whether organisms can be seen inside them. Acid-fast and fungal stains can be used if such infections are likely. Pathogens recovered most frequently from the exogenous and endogenous flora are shown in Fig. 3. Bacteria often responsible for intra-abdominal infections are listed in Table 2.

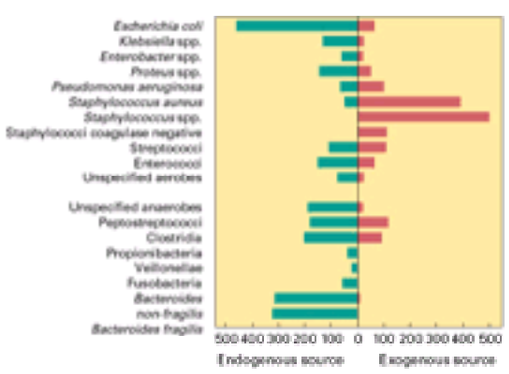


Fig. 3. Bacteria isolated from surgical infections, contrasting those stemming from an exogenous (bone and joint) and an endogenous (intra-abdominal) source.

Aerobic bacteria	%	Obligate anaerobic bacteria	%
Escherichia coli	51	Bacteroides, unspecified	72
Proteus spp.	16	Fusobacteria	7
Klebsiella spp.	14	Veillonella	2
Enterobacter spp.	6		
Pseudomonas spp.	7	Propionibacteria	5
Streptococci	12	Clostridia	23
Staphylococci	5	Peptostreptococci	21
Enterococci	17		
Other	8	Other	21

Data from: Wittmann DH. *Intra-abdominal infections: pathophysiology and treatment*. Marcel Dekker, New York, 1991.

Table 2 Pathogens isolated from 900 patients with intra-abdominal infections

Technique of obtaining the specimen

Purulent material from the deepest aspect of the wound should be aspirated into a syringe and any air evacuated. Pus is the best medium in which to preserve bacteria for transport to the laboratory. The capped syringe is sent for aerobic and anaerobic culture, and for assay of antibiotic sensitivity. Alternatively, a moist swab can be used to obtain bacteria from a site of suspected infection. Ideally, anaerobic specimens should be transported immediately in a CO_2-filled tube and plated within 1 h of sampling; fastidious organisms may otherwise die, resulting in a false-negative culture result. If the specimen is held overnight, it should be placed in an anaerobic sterile vial or tube; under no circumstances should an anaerobic specimen be refrigerated. Generally speaking, *E. coli* (aerobe) and *Bacteroides fragilis* (anaerobe) are the usual causes of wound infection following gastrointestinal or gynecologic operations, while streptococci, staphylococci, and peptostreptococci are the usual causative organisms when intra-abdominal viscera have not been resected or opened.

Blood cultures are helpful in guiding the specific antibiotic therapy of serious infections if the empirically started, initial antibiotics fail. Following careful disinfection of the venepuncture site with an iodophor preparation, blood samples should be obtained for aerobic and anaerobic culture. Blood should not ordinarily be drawn for culture through an existing intravenous needle or catheter. It is important to obtain a number of blood cultures from different sites and at different times. Once the patient chills and a fever spike is observed, most bacteria have already been killed by host defense mechanisms and blood cultures will be negative. It is possible, however, to predict the time of the next bacteremic episode, because fever spikes occur intermittently. Drawing four blood samples at hourly intervals before the next peak will increase the likelihood of a positive culture. If the patient is receiving treatment with antimicrobial drugs, a drug-removing device is helpful in obviating antimicrobial action during culture.

Sensitivity tests need to be interpreted appropriately and with caution since they are not always reproducible and are an oversimplification of the complex foundations upon which antimicrobial chemotherapy is based. Disc diffusion tests are highly sensitive to small technical and environmental changes. Their results may not correlate well with the actual minimal inhibitory or bactericidal concentration of an antibiotic, or with the concentration of antibiotic achieved at the site of infection with the chosen dosage. While important for epidemiologic purposes, routine disc sensitivity tests are generally of little value in guiding an individual patient's antibiotic therapy. The minimal inhibitory concentration (**MIC**), or the minimal bactericidal or fungicidal concentrations, are more useful clinically, because antimicrobial dosing can be adjusted to achieve and sustain antibiotic concentrations at the focus that are three to four times in excess of the concentration required to kill bacteria in the test-tube.

Therapy of surgical infections

All wounds, whether made at the operating table or resulting from trauma, expose normally sterile tissue and provide an environment for bacterial growth. Infections can be minimized if wound management follows these principles.

1. Tissue should be handled gently, and operative trauma kept at a minimum.
2. Further contamination should be minimized by use of aseptic techniques.
3. Devitalized tissue, debris, and traumatic foreign bodies should be removed.
4. Complete hemostasis should be achieved.
5. Blood supply is essential for healing and should not be impaired.
6. Formation of dead space should be avoided during closure.
7. The wound should be closed by layer-to-layer approximation without tension.
8. Operative time should be kept to a minimum to reduce the numbers of bacteria entering the wound.
9. The wound may be irrigated with liberal amounts of sterile saline/Ringer's lactate solution prior to closure.

Resistance of the host to infection is intimately involved with the magnitude of trauma and the early inflammatory response. Three primary factors interact in infection:

1. the extent of tissue injury;
2. the inoculum (quantity) and toxic products (quality) of infecting micro-organisms;
3. the host's defense capacity.

Meticulous atraumatic technique is important: the extent of tissue injury is crucial in the development of subsequent bacterial infection. Macrophages that are preoccupied with the phagocytosis of traumatized dying tissue are not available for killing invading micro-organisms and smaller bacterial numbers may then lead to infection. Debridement is important to reduce the amount of necrotic tissue. If the initial wound is contused and fringed with loose tissue, generous excision by scalpel of the wound edges leaves less traumatized tissue. Any fresh wound almost invariably becomes contaminated as long as it is exposed to the environment. In this context, the operating room is not sterile, and bacteria can be isolated from the wound exudate of even primarily healing wounds. Bacteria in any wound require 6 to 8 h to multiply to a numbers that are sufficiently virulent to invade adjacent tissues if the wound is closed. Up to 8 h after injury, debridement of necrotic tissue and primary wound closure carry little risk of infection. However, infections frequently complicate the healing of wounds if primary closure is completed later than 6 to 8 h after injury (Friedrich's 8-h rule).

General principles of therapy

Non-localized infections (erysipelas, cellulitis, lymphangitis) should be treated with immobilization and elevation as well as antibiotics, preferably penicillin G in high doses (10 million units every 6 h). Local, moist heat relieves pain and increases blood and lymph flow: it is best applied by intermittent moist compresses of short duration. Surgical incision and drainage are usually not indicated.

Focus control, incision, and drainage

In a surgical infection the most important therapeutic modality is the operative elimination of the focus of invading and multiplying bacteria, including factors promoting bacterial growth. Alternatively, incision and drainage of an infectious focus will create a route permitting such material and bacteria to escape while host defense mechanisms establish a wall to halt bacterial invasion into healthy tissue. Antimicrobial therapy is effective only in conjunction with operative reduction of the bacterial inoculum by means of source control. Incision and drainage is thus indicated whenever an infection is localized in a closed space. In most superficial abscesses the finding of fluctuance signals the appropriate time for drainage. When in doubt, needle aspiration may be diagnostic, especially in deeper infections. The incision must be large and placed in the most dependent area of the abscess, with respect for important structures. The incision must be kept open to prevent skin closure before the deepest site is cured. Superficial wound abscesses should be packed lightly with gauze after drainage; deeper abscesses are kept open by sump drains or tubes.

Antibiotic therapy

The goal of antibiotic therapy is to achieve a concentration of antibiotic in the infected tissue that exceeds the MIC by three to four times for at least three-quarters of the time between successive doses (time above the MIC). Drug pharmacokinetics vary considerably with patient age and disease. Antibiotics are usually given for too long and in insufficient doses, particularly at the onset of treatment, if manufacturers' recommendations and disc-sensitivity data only are used to determine the dose and dosing interval. Clinically appropriate doses are recorded in our 1991 review (see Further reading), which should be consulted for further information. Suggestions about initial empiric therapy for surgical infections originating from various sources are recorded in Table 3.

Bacteria or Gram stain	Source	Initial calculated antibiotic therapy
In chains, Gram +ve cocci	Soft tissue	Penicillin G (high dose; 10 million units/day)
In chains, Gram +ve cocci	Soft tissue	Penicillin G (high dose; 10 million units/day) Clindamycin or methicillin
Gram −ve rods	Abdomen	Cefotaxime
	Pulmonary	Cefotaxime
	Biliary tract	Ciprofloxacin or other quinolones
Gram +ve rods	Soft tissue	Penicillin G (high dose; 10 million units/day)
Mixed Gram +ve and −ve	Intra-abdominal abscess	Cefotaxime + metronidazole, or ampicillin–sulbactam
	Multiple injuries, especially extremities	Cefotaxime + clindamycin

Table 3 Initial empiric therapy for surgical infections

Systemic antibiotics are never indicated for the treatment of uncomplicated wound abscess which can be drained through an incision that does not open new tissue planes or expose new tissue to contamination by the contents of the abscess. Since overlying tissues must be opened to drain deep abscesses, antibiotics should be administered during the period immediately before and during incision and drainage. A longer period of therapy with parenteral antibiotics is indicated in immunocompromised patients or when focus control is difficult, such as in infected pancreatic necrosis. If infection persists, the first question to be asked before systemic antibiotics are given is whether the source-control operation or the incision and drainage have been adequate, and whether there may be another, unrecognized surgical infection.

In general, surgeons treat infections for too long because we have problems distinguishing between contamination, infections, and continuing inflammation after all bacteria have died. The issue has been recently addressed by an expert forum, which concluded that the duration of antibiotic therapy should be short, depending on the success of operative focus control. The guidelines provided are listed in Table 4.

Contamination: no postoperative antibiotics
1. Gastroduodenal/peptic perforations operated within 12h
2. Traumatic enteric perforations operated within 12h
3. Peritoneal contamination with bowel-contents during elective or emergency procedures
4. Appendicitomy for early or phlegmonous appendicitis
5. Cholecystectomy for early or phlegmonous cholecystitis

Treatable infection: 2dh of postoperative antibiotics
1. Appendicitomy for gangrenous appendicitis
2. Cholecystectomy for gangrenous cholecystitis
3. Bowel resection for ischemic or strangulated 'dead' bowel without frank perforation

'Mild' Infection: 48h of postoperative antibiotics
1. Intra-abdominal infection from diverse sources with few clinical pus formation
2. 'Late' (more than 12h) traumatic bowel lesions and gastroduodenal perforation with no 'established' intra-abdominal infection

'Moderate' infection: up to 5 days of postoperative antibiotics
1. Diffuse 'established' intra-abdominal infection from all sources

'Severe' infection: more than 5 days of postoperative antibiotics
1. Severe intra-abdominal infection with the source not easily controllable (i.e., infected pancreatic necrosis)
2. Severe intra-abdominal infection treated with planned relaparotomies
3. Postoperative intra-abdominal infection

Adapted from recommendations of consensus paper: Schein M, Wittmann DH, Looser W. *European Journal of Surgery* 1997; 163 (Suppl.).

Table 4 Recommendations for the duration of antibiotic therapy in abdominal infections following source control

Antimicrobial treatment should be specific, that is, directed against the causative pathogens based on either the clinical findings or a specific bacteriologic diagnosis. Most surgical infections, however, are caused by a mix of many obligate and facultative anaerobic bacteria, and culturing all these organisms may become extremely difficult. On the other hand, rapid (within 2 h or so) bacterial identification is not yet available clinically and in most cases antibiotic therapy must be started before bacteriologic results are available. When therapy must be started empirically it should be calculated and targeted at the most likely pathogens and the following points should be considered:

1. the spectrum of pathogens known to be typical;
2. the pathogenicity, synergism, and antagonism exhibited by bacteria in various mixed infections;
3. the concentration of antibiotic that can be achieved at the site of infection;
4. the side-effects of antimicrobials;
5. the negative interaction of antibiotics with host defense mechanisms;
6. the results of well-controlled clinical studies of unselected patients.

Antibiotics that reliably kill bacteria should be given preference, such as penicillin for infections by group A streptococci or clostridia and no penicillins, or extended-spectrum penicillins with and without b-lactamase inhibitors, against *E. coli* and *Klebsiella*.

Risk factors

Several factors increase the risk of a patient acquiring infection (Table 5). These risk factors may be related to the patient's capacity to defend against an infectious threat, the infectious challenge itself as represented by the number and pathogenicity of the bacteria, the extent of any associated injury, and environmental factors such as the hospital bacterial flora. The surgeon should be alert to these factors and tailor therapeutic or preventive strategies to the specific circumstances. For example, antibiotic prophylaxis is usually not indicated before a clean or aseptic operative procedure but may very well be useful in an insulin-dependent diabetic. The dilemma for the surgeon is to decide when bacteriologic contamination of an open ulcer, for example, becomes a clinically significant risk or is associated with tissue-penetrating disease.

Patient related
Malnutrition
Immune compromise
Obesity
Renal insufficiency
Pulmonary dysfunction
Cardiovascular disease
Endocrine and metabolic disorders
Injury related
Location, extent
Environment related
Contamination, superinfection
Hygiene
Long-term stay in intensive care unit

Table 5 Risk factors for infection

Improper preoperative management is an organizational problem and can increase the rate of postoperative infection. Common errors include failure to give a preoperative bath with antiseptic soaps or solutions, and shaving of the operative site the night before operation. Shaving is probably not indicated in most patients; when practised, it should be limited to the immediate preoperative period.

Long-term treatment in the intensive care unit

Changes in the microbial flora of the skin, respiratory tract, and gastrointestinal tract are seen in the most seriously ill patients, regardless of the underlying disease.

Resistant Gram-negative organisms, staphylococci, and fungi usually colonize such patients shortly after admission to the hospital. Factors increasing the incidence of colonization are antibiotic administration, the use of inhalation-therapy equipment, immunosuppressive or irradiation therapy, and depressed neurologic status. The administration of multiple, potent antimicrobials disturbs the patient's own microbial ecologic system, since not only pathogenic but also symbiotic bacteria are eliminated. For example, therapy with imipenem often leads to the elimination of too many bacteria, and, in the susceptible patient, allows superinfection with fungi or other resistant bacteria that are normally of low pathogenicity. These situations should be avoided by discontinuing antimicrobials at the earliest possibility or changing early to a narrow-spectrum regimen directed only at pathogenically important micro-organisms.

Scoring severity of infection

Many scoring systems have been developed to assess the severity of disease. The Surgical Infection Society currently proposes the use of the Acute Physiology Score and a Chronic Health Evaluation (**APACHE**)-II system to compare treatment regimens (Fig. 4), in which there are disease-specific weighting factors to calculate mortality. For patients with intra-abdominal infection, a score of 20 correlates with a mortality risk of 52 per cent (Fig. 5). Scoring, when reported, should always include associated mortality in a tested and representative patient population. It is a helpful tool to assess current and new treatment by comparing it to the score-predicted mortality. To calculate the predicted mortality for intra-abdominal infection, one may use the logistic regression formula in the following equation:

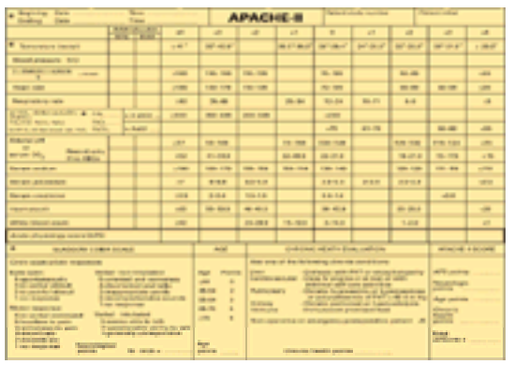


Fig. 4. Form to record data to determine the APACHE-II score.

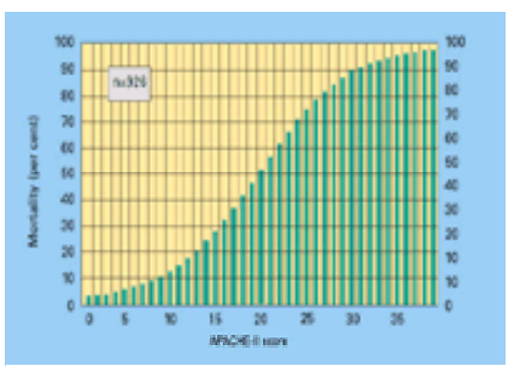


Fig. 5. Correlation of APACHE-II score and mortality risk in intra-abdominal infection (unpublished data from D.H. Wittmann).

$$R = \frac{\text{Exp}\left((\text{APACHE}) - \text{II}^{*}0.146 - 3.517\right) - 3.517)}{1 + \text{Exp}((\text{APACHE-II}^{*}0.146 - 3.517)}$$

Further details are given in Knaus et al. (1985) (see Further reading), where weights for specific diseases are given. All further attempts to improve severity-of-disease scoring have resulted in minimal improvements at the expense of inappropriately increasing the amount of input data required.

Skin and soft-tissue infections

Infections with a single organism usually follow minor trauma, and are limited to the skin and subcutaneous tissues. However, even minuscule lesions may become life threatening: examples are streptococcal erysipelas, cellulitis, phlegmon or lymphangitis; clostridial infections including gas gangrene and tetanus; and some staphylococcal infections.

Cellulitis, erysipelas, and phlegmon

Group A streptococci may cause cellulitis, erysipelas, and spreading phlegmons: once inoculated beneath the skin, the defensive barriers are easily breached by the toxins released by streptococci; in addition the lymphatic system is frequently involved. Clinically, there is edema with reddening of the skin (Fig. 6). Penicillin G is the treatment of choice since it kills all group A streptococci, a remarkable property that has not changed since its introduction more than 45 years ago. Before the discovery of penicillin, invasive group A streptococcal infection had a mortality rate of 90 per cent. Anaerobic streptococci (peptostreptococci) are part of the normal flora of the mouth and gastrointestinal tract. In contrast to other streptococcal wound infections, these organisms produce a thin, brown discharge, often with crepitation in the infected tissue (anaerobic cellulitis). Treatment consists of wide incision and drainage, and the administration of 10 million units (6 g) of penicillin G 6-hourly. Cephalosporins, clindamycin, chloramphenicol, and metronidazole are second-choice agents against anaerobic cocci.



Fig. 6. Secondary wound erysipelas due to hemolytic *Strep. pyogenes*: a superficial wound infection had previously been treated by drainage and packing with resulting granulation of the subcutaneous tissues; nonetheless, and reflecting the invasive capacities of streptococci, this infection began 4 days after opening of the wound; it responded to oral penicillin.

Folliculitis, furunculosis, and carbuncle

These infections are usually due to *Staphylococcus aureus*, although in patients receiving antibiotics, Gram-negative bacteria and *Candida* may be the cause. Folliculitis originates within one hair follicle; furunculosis represents infection of several hair follicles in a circumscribed area; a carbuncle (boil) is a confluent infection

involving multiple contiguous follicles in which the infection is limited to the subcutaneous tissue by thick overlying skin and dense subcutaneous fascia. Carbuncles are usually found on the back of the neck and torso. Warm skin compresses and good local hygiene are usually sufficient therapy for folliculitis and most cases of furunculosis. Carbuncles require incision for drainage and treatment with antistaphylococcal penicillins, erythromycin, or clindamycin.

Hydradenitis suppurativa

This infection of apocrine sweat glands is usually seen in young adults and is due to staphylococci or anaerobes (especially peptostreptococci). Often, only complete excision of the infected tissue down to deep fascia, with subsequent grafting or delayed closure, is curative.

Bite wounds

Human bite wounds are contaminated with a combination of aerobic nonhemolytic streptococci, anaerobic streptococci, *B. melaninogenicus*, spirochetes, and staphylococci. The original wound must be treated by debridement, thorough irrigation, and immobilization. Systemic antibiotics, usually penicillin, must be administered. If the infection becomes established, radical debridement of the infected area is imperative and must be accompanied by antibiotic therapy.

Infections of dog and cat bite wounds are caused by *Pasteurella multocida* in 25 to 50 per cent of cases; otherwise, the spectrum of bacteria is the same as that seen in human bite wounds. The antibiotics of choice are high-dose penicillin G, amoxicillin/clavulanic acid, or oral cefuroxime.

Synergistic gangrene

Chronic progressive bacterial gangrene is caused by the synergistic action of microaerophilic, nonhemolytic streptococci and aerobic, hemolytic staphylococci (Fig. 7). The incubation period is 7 to 14 days. Cellulitis is followed by gangrenous ulceration that is progressive unless treated. Radical excision of the ulcerated lesion and its gangrenous borders is imperative, along with administration of large systemic doses of penicillin. Burrowing ulcers are caused by a combination of microaerophilic streptococci and staphylococci (Meleney ulcer). Such lesions have a characteristic metallic sheen, cause necrosis of large areas of skin, and may produce sinus tracts in the underlying tissue. These should be incised, drained, and treated with high doses of penicillin (10 million units every 6 h).



Fig. 7. Synergistic gangrene caused by a mixed infection by anaerobic streptococci and staphylococci: this infection was caused by self-injection with heroin; there was extensive destruction of the deep tissues extending to the borders of this photograph.

Non-clostridial gangrenous cellulitis caused by *B. melaninogenicus* and anaerobic streptococci is typified by a progressive gangrenous infection of the skin and adjacent areolar and fascial tissues. Prompt incision and drainage, and large doses of penicillin, are necessary. Supportive treatment is imperative, since toxemia with dehydration, fever, and prostration rapidly develops.

Clostridial cellulitis is a serosanguineous, crepitant, septic process of subcutaneous, retroperitoneal, or other areolar tissue, caused principally by *C. perfringens* (also known as *C. welchii)*. It differs from gas gangrene in that the infection does not involve muscle (Fig. 8), but spreads rapidly via fascial planes. Extensive gangrene results from vascular thrombosis. Systemic effects are moderate if the infection is treated promptly with early surgical debridement and penicillin.

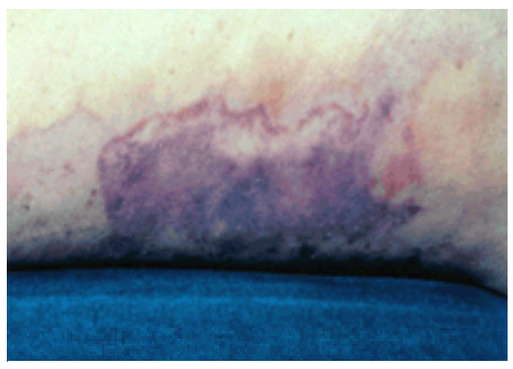


Fig. 8. Cellulitis due to *C. perfringens*: note the ecchymotic discoloration of involved subcutaneous fat and the advancing border of cutaneous infection defined by a line of erythema.

Clostridial myonecrosis (gas gangrene) is an anaerobic infection of muscle characterized by profound toxemia, extensive local edema, massive necrosis of tissue, and a variable degree of gas production (Fig. 9). The causative organisms are the clostridia that abound in soil, dust, and the alimentary tracts of most animals and are usually saprophytic. *Clostridium perfringens*, which is the most common cause, produces a variety of potent toxins, including hyaluronidase, collagenase, four different hemolysins, five necrotizing lecithinases, and six other lethal necrotizing toxins. All clostridia owe their pathogenicity to the elaboration of such soluble exotoxins, which destroy tissue and blood cells. Clostridia enter a wound, multiply in the presence of devitalized muscle, and use iron from myoglobin to produce necrotizing exotoxins. Disruption and fragmentation of normal muscle cells and capillaries result in further necrosis, hemorrhage, and edema. There is no fibrin formation or polymorphonuclear leucocytic reaction. The affected muscles are at first red and friable, but progress to a purplish black, stringy, pulpy mass. The presence of gas is variable. The affected area swells and discharges a brownish, malodorous fluid. The overlying skin initially shows blotchy ecchymoses (marbling), then blackens, and finally sloughs.

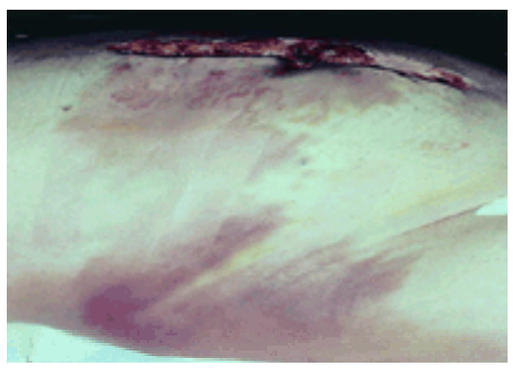


Fig. 9. Myonecrosis due to *C. perfringens*: note the bronzing of skin; this infection spread to this state in only 24 h following an operation for pelvic infection.

The diagnosis of gas gangrene is based on typical clinical findings, as well as on the presence of large, Gram-positive rods in the wound fluid. Delay in diagnosis, even for just a few hours, greatly increases the mortality. Immediate removal of involved muscle groups is necessary: amputation is indicated if the remaining viable muscles are insufficient for useful function. High intravenous doses of penicillin and whole blood are given preoperatively and postoperatively. Multiple treatments with hyperbaric oxygen (oxygen at 3.03 kPa) may reduce the amount of debridement necessary and lower the mortality, but muscle resection should not be delayed in anticipation of hyperbaric therapy. Untreated gas gangrene is always fatal; the fatality rate in treated patients ranges from 25 to 40 per cent.

Tetanus

This is caused by a spore-forming obligate anaerobe, *C. tetani*, found in the feces of humans and animals, and capable of prolonged survival in soil. Two exotoxins are produced: tetanospasmin, a neurotoxin, and tetanolysin, a hemolysin. Dead muscle and clotted blood provides an ideal culture medium for the germination of tetanus spores, and compound fractures with devitalization of muscle are very susceptible to such infection, as are small puncture wounds harboring a clot deep in the tissues. Locally produced tetanolysin contributes to optimal growth conditions through its lecithinase, gelatinase, esterase, and lipase activities. Tetanospasmin, the neurotoxin responsible for the clinical features of the disease, does not act peripherally or locally but is carried to, and acts on, the central nervous system. In order to neutralize blood-borne toxin, antitoxin must be present before tetanospasmin becomes fixed by nerve cells. Antitoxin given when symptoms are apparent only limits further intoxication of nerve cells and cannot reverse developing symptoms.

There is considerable variability in the progression of the disease from its onset. There may be a prodromal period of headache, stiff jaw muscles, restlessness, yawning, risus sardonicus, and wound pain, typically beginning 1 to 2 weeks after trauma but occasionally as early as 1 day or as late as 2 months after injury. The active stage follows in 12 to 24 h, with trismus, facial distortion, opisthotonos, pain, clonic spasms, and seizures. Acute asphyxia is a major hazard: it may result from either spasm of the respiratory muscles or aspiration. The shorter the incubation period, the poorer the prognosis.

Human immune globulin, 3000 units intramuscularly, should be given immediately to neutralize circulating toxins. An additional 1000 units are injected into and immediately proximal to the wound, followed by wide debridement. Five hundred units of intramuscular immune globulin may subsequently be given daily. If symptoms persist for longer than 2 weeks, the large initial doses of immune globulin may be repeated. An airway must be established: tracheostomy will be needed in every patient with more than prodromal symptoms and should be performed before the situation becomes urgent. Respirator support and oxygenation may be needed. Muscle spasms may be controlled with intravenous midazolam or diazepam, or with intramuscular meprobamate or chlorpromazine. If spasms persist, curare or another muscle blocker should be given. Additional sedation is usually not needed if muscle relaxant therapy has been adequate, but is best achieved with intramuscular barbiturates if required. The patient should be placed in a quiet room, with environmental stimulation kept at a minimum to avoid triggering seizures. Intravenous thiopental (Pentothal) may be needed to control seizures. High doses of penicillin G (5–10 million units) will establish a sufficient antibacterial tissue concentration. With appropriate early care, 75 per cent of patients survive with no neurologic impairment.

Principles for tetanus prophylaxis

Tetanus is absolutely preventable by prior active immunization. Effective active immunization (not in association with a fresh wound) is accomplished by the injection of 0.5 ml fluid or adsorbed toxoid, repeated after 2 and 20 months.

Immediate meticulous surgical care of the fresh wound is of prime importance. Removal of devitalized tissue, blood clots, and foreign bodies, obliteration of dead space, and prevention of tissue ischemia in the wound are the objectives of initial treatment. If the wound is grossly contaminated, penicillin should be administered. Wounds seen late or grossly contaminated should be left unsutured after debridement, protected by a sterile dressing for 3 to 5 days, and closed by delayed primary sutures if the tissues appear clean and healthy. Active and passive immunization should be accomplished with 0.5 ml tetanus toxoid and tetanus immune globulin, 250 to 1000 units, as outlined in Table 6. Patients not previously immunized should receive injections of 0.5 ml fluid or adsorbed toxoid, repeated after 2, 6, and 20 months.

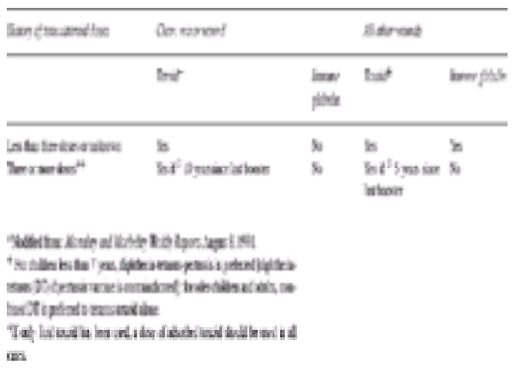


Table 6 Tetanus prophylaxis in wound management*

Infections of the hand

The most common hand infections are paronychia, pulp infection, and subcutaneous abscesses, including felon and bacterial tenosynovitis. The therapeutic goal is to restore full function to the hand. Elevation, immobilization in the position of function, and heat in the form of hot, wet packs changed every 2 to 4 h are helpful. Hot soaks may be used for 20 min every 4 h. Ten million units of penicillin 6-hourly are effective in severe infections. Abscesses and other local collections of pus should be drained promptly; generally the volar aspect is incised. Important principles of managing hand infections are summarized in Table 7.

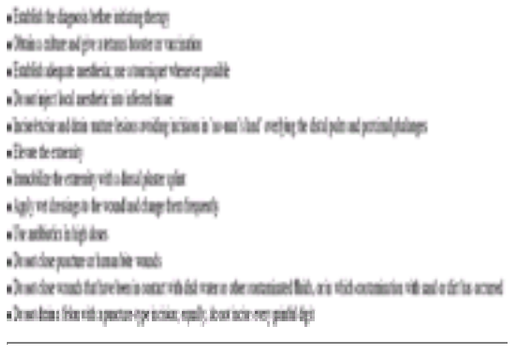


Table 7 Principles of management of hand infections

Breast abscess

Anaerobic and staphylococcal infections of the breast occur in two forms: puerperal (post partum) and spontaneous (nonpuerperal). A puerperal breast abscess is usually caused by staphylococci; nonpuerperal breast abscesses (Fig. 10) may also be due to anaerobes, usually *Peptostreptococcus magnus* (Table 8). Excision of

the involved duct through a periareolar incision is indicated for persisting or recurring, nonpuerperal breast abscess.



Fig. 10. Nonpuerperal breast abscess: the pus was viscid and creamy yellow; the culture grew *B. ureolyticus*, two strains of *Peptostreptococcus* and an ampicillin-sensitive *E. coli*; despite the clinical appearance of the pus, no staphylococci were present; this case illustrates the necessity of documenting infections by culture.

Group	No. of organisms
Aerobes (facultative anaerobes):	
*Staphylococcus**	24
Streptococcus	4
Bacillus	2
Corynebacterium	1
Escherichia	1
Proteus	1
Pseudomonas	2
Anaerobes (obligate anaerobes):	
Actinomyces	3
Clostridium	2
Fusobacterium	1
Mitsuokella	1
*Bacteroides**	8
Lactobacillus	3
Eubacterium	3
*Prigomonbacteriaceae**	85
*Peptostreptococcus**	34
Veillonella	2

*Predominant species.

Table 8 Bacteria recovered in nonpuerperal breast infection

Pilonidal disease

A subcutaneous area located in the intergluteal cleft over the sacrum becomes infected when ingrown hair leads to the formation of a foreign-body sinus and triggers the growth of mixed bacterial species, often with a predominance of anaerobes. Initial lesions are treated by unroofing or wide excision, and allowing secondary healing. This therapy results in a broad scar. The lack of skin structures such as hair follicles in a scar prevents recurrences. No additional antibiotic therapy is required, although wide excision and healing by secondary intention may be required if infection is recurrent or persistent.

Fox den disease

In patients with a disposition to acne conglobata, a sinus-forming and fistulating pyoderma may develop that is different from hidradenitis and pilonidal disease. It consists of epithelialized fistulas and sinuses that spread epifascially over long distances, forming a tract system similar to a fox den. The fistulas contain mostly obligate anaerobic bacteria. Because fistulas and sinuses form epithelium, fox den disease cannot heal spontaneously and requires wide, *en bloc* excision of the entire tract system. It usually starts in the perineum and may extend into the scrotum and anterior abdominal wall. It has also been observed on the upper trunk.

Perianal abscess

Anaerobic infections of the perianal soft tissue originate from anal ducts and glands; they present as localized, subcutaneous foci immediately adjacent to the pigmented epithelium of the anal verge and canal. They do not burrow into deeper tissues, but must be clearly differentiated from perirectal and deeper abscesses by digital rectal examination. If tenderness precludes adequate investigation, a thorough examination can be completed under anesthesia. Before fluctuance occurs, antibiotic therapy for 2 to 3 days together with twice-daily sitz baths will often resolve the infection. If a fluctuant collection is present, perianal incision allows drainage; antibiotics are discontinued but sitz baths should continue until all inflammation has resolved.

Perirectal abscess

Burrowing, intersphincteric, ischiorectal, and supralevator abscesses originate in the rectal crypts and glands at the upper end of the anal canal. Goodsall's rule is incorrect: burrowing abscesses may originate at any point on the anorectal circumference. Typically, the abscess burrows directly and radially from the infected gland or crypt of origin to present on the perianal skin at a slight distance from the anal verge. Burrowing may also occur in the subcutaneous tissues around the anus to form a more or less complete ‘horseshoe abscess.' Alternatively, the infection may burrow between the anal sphincters or through the sphincters into the ischiorectal space, from which it may extend superiorly into the supralevator space, before dissecting subcutaneously to perianal skin. All forms of perirectal abscess are potentially more dangerous than a simple perianal abscess, and the two infections should not be confused. Examination under anesthesia (regional block or general) is essential to make an accurate diagnosis. Drainage through a wide perianal incision is the treatment of choice. *Bacteroides* spp., clostridia, and peptostreptococci predominate as causative organisms; the most common causative aerobe is *Proteus mirabilis*. The drainage wound should be packed open, packs changed frequently, and sitz baths taken twice daily. Additional antibiotic therapy after completion of drainage is not necessary in patients with normal immunologic function.

Thoracic infection

Lung

Lung infection may follow chest trauma that includes pulmonary contusion or rib fractures. Pulmonary infections may be due to a variety of micro-organisms: aspiration following major trauma is a common cause of bronchopneumonia, usually caused by oral anaerobes. Prevention is key and includes total analgesia to allow for optimal breathing, so enhancing pulmonary mechanisms of bacterial clearance. Epidural analgesia is of benefit; early intubation and artificial respiration may be required. Broad-spectrum antibiotics such as imipenem or cefotaxime, in combination with clindamycin, are preferred.

Pleural infections

Pleural empyema may follow hemothorax as well as pneumonia or other pulmonary infections, and is usually due to the same organisms as caused the lung infection. It may also follow elective thoracotomy. Empyema always requires closed-tube drainage. If the lung does not completely expand, open pleural debridement (decortication is a misnomer) is essential. Antimicrobial therapy is directed against the causative bacteria, which may include anaerobes. Pulmonary infections, particularly those that follow aspiration of oral or gastric fluids, may progress to intrapulmonary abscess. Obligate oral anaerobes are often involved, but these may be difficult to isolate in routine culture. Percutaneous needle aspiration is often helpful for diagnosis. Antibiotic therapy should include anaerobic coverage with metronidazole; open drainage may be required.

Infections of the heart

Surgical infections of the heart include endocarditis and pericarditis. Tuberculous pericarditis may require pericardiectomy; endocarditis due to *Strep. viridans*, pneumococci, and other bacteria also may require operative treatment. Enterococcal endocarditis takes a less aggressive course. Subacute bacterial endocarditis is

usually caused by streptococci of the viridans group (70 per cent of cases), *Enterococcus faecalis*, or group D streptococci. All streptococci are sensitive to parenteral penicillin at a dose of 6 million units (3.6 g) every 24 h for 4 weeks. The penicillin sensitivity of *Ent. faecalis* is variable; ampicillin is the best treatment for infections with this organism. Enterococci are generally resistant to cephalosporins and aminoglycosides. Vancomycin-resistant *Ent. faecium* as a result of vancomycin overuse has become a recent problem in many hospitals in the United States of America. While of low pathogenicity, enterococci represent a dangerous pool of resistance genes that may transfer to other bacterial species.

Abdominal infection

Peritonitis

Contamination of the abdominal cavity may follow penetrating abdominal trauma or blunt trauma associated with intestinal perforation or rupture. Such patients usually undergo early operation, and intra-abdominal infection does not occur; unless operated on within 12 h of traumatic perforation, an intra-abdominal infection is likely. Secondary peritonitis may follow a variety of pathologic conditions, including perforated peptic ulcer, pancreatitis, perforation of the gallbladder, bowel ischemia due to strangulation or vascular compromise, appendicitis, perforation of the small or large bowel, genitourinary infection, and perforation of an intra-abdominal abscess. The source of infection must be controlled by closure, excision, or exteriorization (focus control); the abdominal cavity must be cleansed, purging bacteria, toxins, and adjuvants such as bile, mucus, barium, blood, and necrotic tissue. Further influx of bacteria or adjuvants must be prevented. Standard operative therapy consisting of eliminating the infectious source as well cleansing the abdominal cavity from pus, feces, and other toxic products is sufficient in most cases. About 10 per cent of patients present with advanced disease. They often have mortality rates of over 40 per cent predicted by their APACHE-II score. They may have progressive deterioration, and it may not be possible to close the source at a single operation. In some cases the inflammatory edema and ileus may have caused increased intra-abdominal pressure, leading to an abdominal compartment syndrome; closure of the abdomen may then be impossible without exerting undue tension on the fascia. The staged abdominal repair (STAR) consisting of multiple planned relaparotomies and final, fascia-to-fascia, abdominal closure, which is so different from the devastating open abdominal procedures, may be used under these circumstances to improve outcome. The Wittmann Patch® may be useful to bridge the gap temporarily. It allows for easy entry and reclosure of the abdomen and avoidance of fistual and hernia formation.

Abdominal compartment syndrome

Sustained increased intra-abdominal pressure above 10 mmHg (abdominal hypertension) impairs renal, hepatic, cardiovascular, and pulmonary function. Vena caval flow is diminished, reducing cardiac preload while increasing the afterload. Increased pressure extends into the pleural spaces, reducing ulmonary compliance and causing atelectasis leading to pneumonia. Impaired liver function may lead to insufficient production of the proteins required for an adequate inflammatory response. Reduced intestinal perfusion impairs anastomotic healing and leads to an ischemic hollow viscus that in turn will liberate bacteria to close a vicious circle. Abdominal decompression reverses effectively all pressure-induced dysfunctions. The deleterious effects of increased intra-abdominal pressure deserve more clinical attention.

Biliary infections

These are almost always associated with calculous disease. The infection progresses from acute cholecystitis to cholangitis or empyema of the gallbladder and, occasionally, to internal fistula formation; alternatively, the acute infection may settle into repetitive episodes of chronic cholecystitis. Uncomplicated gallstones are associated with bacteribilia (culture-positive bile) in 30 to 50 per cent of cases. The most important and frequently recovered pathogenic bacteria are *E. coli*, klebsiellae, and clostridia. Treatment consists of cholecystectomy and drainage of the bile-duct system, following which infection usually resolves with antibiotic therapy. In immunocompromised patients, broad-spectrum penicillins and third-generation cephalosporins are indicated.

Liver abscess

This may be due to amoebae, salmonellae, or to a mixed bacterial population, and usually follows appendicitis, bacterial or other forms of colitis, or biliary-tract infection. Echinococcal cysts occasionally become secondarily infected. Obligate anaerobes are found in over 50 per cent of liver abscesses. Percutaneous drainage is needed if the abscess has developed a rind or wall within the liver. Specific antibiotic treatment will cure the infection, provided that the underlying condition is eliminated.

Pancreatitis, infected pancreatic necrosis, and pancreatic abscess

Pancreatitis begins as a chemical inflammation, but more than half of the fatalities are due to infection. Infectious complications typically are seen clinically during the second week. Only antibiotics such as imipenem given at the onset of acute pancreatitis effectively prevent noninfected pancreatic necrosis from becoming infected. Infected pancreatic necrosis and pancreatic abscesses develop within the necrotic pancreatic tissue and require drainage; staged abdominal operations have been advocated for severe cases. Antibiotic therapy includes a third-generation cephalosporin combined with metronidazole; imipenem, or quinolone/clindamycin combinations, are good alternatives.

Appendicitis

Appendicitis may present as a simple infection (Fig. 11(a)) or if not treated it may progress to devastating, often lethal, disease (Fig. 11(b)). Appendectomy is required. Antibiotics given prior to surgery are effective mainly for prophylaxis of infection in the incisional wound, and in uncomplicated appendicitis need not be continued after the operation has been completed. Continued administration of antibiotics is indicated when the disease has progressed to gangrene, perforation, abscess, or diffuse peritonitis. Abscesses may be drained percutaneously with the help of ultrasound or CT, or open drainage by laparotomy may be established. Obligate anaerobes are always present. *Escherichia coli* may cause lethal sepsis. In severe cases of diffuse peritonitis, a third-generation cephalosporin must be combined with metronidazole for therapy; ampicillin/sulbactam has also been successful.

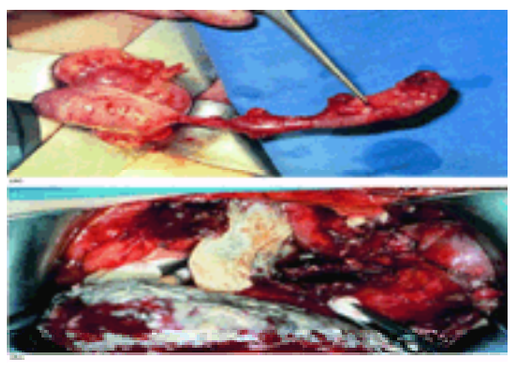


Fig. 11. The range of infection associated with appendicitis is broad: (a) illustrates a case of simple appendicitis treated by early operation; (b) illustrates a case of neglected, perforated appendicitis with formation of a complex lower abdominal abscess associated with necrosis of adjacent tissues.

Diverticulitis

More than half of patients over 50 years of age in the Western world have colonic diverticula, but only a minority develop symptoms. Diverticulitis may occur at any time and is usually treated with bowel rest and antibiotics. The disease is due to anaerobic organisms; 500 mg metronidazole every 12 h is the treatment of choice. If septicemia develops, a third-generation cephalosporin may be included. Other drug combinations active against *B. fragilis* and *E. coli* may be used. Perforation of a diverticulum results in either diffuse peritonitis or a peridiverticular abscess (Fig. 12). The abscess itself may perforate secondarily and cause diffuse peritonitis; laparotomy and resection of the diseased colon segment is the treatment of choice. Primary anastomosis has been successful in selected cases; if this causes problems, staged abdominal repair may be preferred to an end-colostomy, but Hartmann's procedure should be avoided if possible. Later restoration of bowel continuity entails no fewer risks of mortality and morbidity than ‘second look' procedures, but the magnitude of such risks is generally exaggerated in the literature.

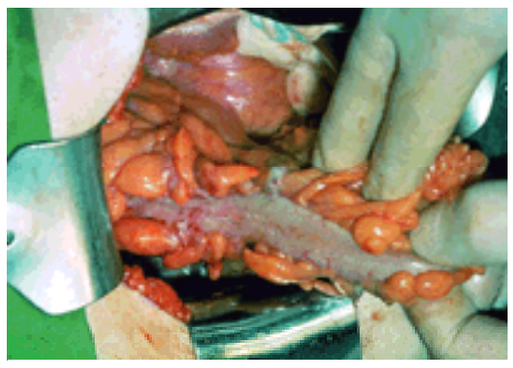


Fig. 12. Diverticulitis: diffuse peritonitis following perforation of a sigmoid diverticulum; the peritoneal edema and fibrin deposits as well as some additional diverticula are clearly visible.

Infections of the female genital tract

Infections of this region are mostly nonsurgical (in the sense that operative drainage or debridement of tissue are not usually necessary to control them), and the mainstay of therapy is the administration of appropriate antibiotics. Sexually transmitted infections are diagnosed by the presence of specific micro-organisms in cultures of cervical exudate. Pelvic inflammatory disease is usually due to gonococci, while endometritis, adnexitis, and myometritis are caused by mixed infections with aerobic and anaerobic intestinal bacteria. These infections are treated with antibiotics; transvaginal drainage is needed in some cases. Occasionally, a persisting tubal abscess may need resection.

Infections of the genitourinary tract

The most common form of urinary-tract infection involves only the urinary bladder, occurs primarily in women due to the relatively short length of the urethra, and is caused by coliform organisms. Such infections usually respond readily to increased fluid intake and administration of an oral penicillin or cephalosporin, although recurrence is common. Infection of the urinary bladder is also common following the placement of a Foley catheter. Although closed urinary drainage systems delay the onset of bladder infection, infection always occurs eventually if the catheter remains in place, the route of infection being through the urethra external to the catheter. Coliform bacteria usually cause such infections. Antibiotics should not ordinarily be used to treat this form of urinary bladder infection unless the catheter can be removed; the continued presence of the catheter assures continuing infection and administration of antibiotics only results in the emergence of resistant organisms. Acidifying bladder washes are partially effective in controlling bladder infection in this situation.

If the ureteropelvic junction is incompetent, backwash of infected bladder urine into the ureter and renal pelvis may result in acute pyelitis, accompanied by septicemia and even septic shock. Chronic pyelonephritis is usually hematogenous in origin; it may result in a nephric or perinephric abscess, which may rupture and cause peritonitis. Drainage is required; antibiotic treatment should be specific and should include a drug active against anaerobes, such as metronidazole.

Bone and joint infection

Post-traumatic osteitis

This is seen primarily in unstable fractures in which moving fragments continuously produce friction and necrotic tissue; unless the fracture is stabilized, the infection cannot be controlled. Staphylococci are the predominant causative micro-organisms but *Pseudomonas aeruginosa* is isolated in 10 per cent of cases. Obligate anaerobes are present in more than 10 per cent of the cases but are usually missed by routine culture techniques; Gram-negative facultative anaerobes may become a problem. Following operative stabilization of the fracture, specific intravenous antimicrobial therapy should be continued for 4 weeks. The bone concentration of the chosen antibiotic should be greater than its MIC for the infecting bacteria.

Osteomyelitis

In adults, antimicrobial therapy should be specific for the infecting organism and should achieve a bone concentration above its MIC. A swab is insufficient as a bacteriologic specimen; open biopsy to obtain a piece of infected tissue for culture and the determination of antibiotic sensitivity is preferred. Hematogenous osteomyelitis in adults (usually immunocompromised individuals) may require operative drainage and sequestrectomy. Hematogenous osteomyelitis in children can be treated successfully if parenteral antibiotics are administered early in the disease, but if a large, subperiosteal abscess or sequestra form, operative management is required. Osteomyelitis of the spine may present as a groin abscess and is usually seen in immunocompromised patients: staphylococci, mycobacteria, and salmonellae are the most common causative micro-organisms. Specific antimicrobial therapy is mandatory; bone biopsy may be required to provide adequate samples for culture and sensitivity studies. The chosen antibiotic needs to reach bone concentrations above their MIC for the pathogen involved.

Infection of the central nervous system

Subdural empyema

Subdural empyema accounts for 10 to 32 per cent of intracranial suppurations. Acute frontal sinusitis and mastoid infections are the most common antecedent cause, but infected subdural collections occur in 2 per cent of patients following meningitis and are most commonly caused by Gram-positive cocci. Early drainage is essential to prevent a further increase of intracranial pressure (neurosurgical emergency).

Brain abscess

Metastatic and primary brain abscess may follow meningitis, or may be due to hematogenous or direct spread from mastoiditis or nasal sinus infection. Anaerobic bacteria and *Staph. aureus* are most commonly found in these conditions. Prompt incision and drainage is required.

Postoperative infections

Wound infection: unequivocal definitions

In 1992, the United States Center for Disease Control (**CDC**) published definitions for wound infections. It renamed ‘wound infection' as surgical-site infection, which may reduce the confusion associated with the previously used term ‘deep wound infection'. The new CDC definitions offer themselves for use in all types of postoperative infections. Investigators increasingly use them to allow for proper assessment of problems as a basis for therapeutic improvements. We highly recommend that the CDC definitions be used by every surgeon to allow for comparison of local and international data.

CDC definitions include superficial incisional, deep incisional, and organ-space surgical-site infections (Fig. 13).

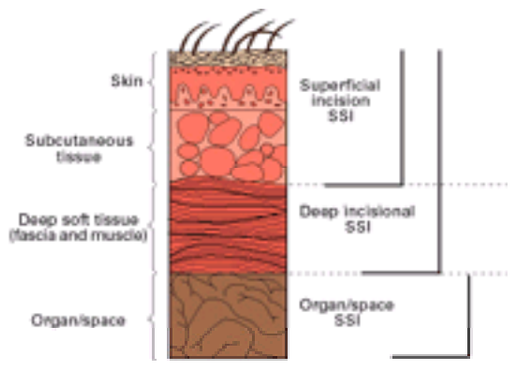


Fig. 13. Surgical-site infections: CDC definitions, drawing from CDC.

Superficial surgical-site infection must meet the following criteria.

1. The infection occurs within 30 days after the operation.
2. The infection involves only skin and subcutaneous tissue.
3. The incision must meet at least one of the following:
   a. purulent drainage from a superficial incision;
   b. organisms isolated from an aseptically obtained culture of fluid or tissue.
4. The incision must meet at least one of the following signs or symptoms of infection:
   a. pain or tenderness;
   b. localized swelling;
   c. redness or heat;
   d. superficial incision is deliberately opened by a surgeon unless the incision is culture negative; or diagnosis of superficial incisional surgical-site infection is made by surgeon or attending physician.

Superficial incisional surgical-site infection

Any purulent discharge from a closed surgical incision (not just from around a suture) with inflammation of the surrounding tissue, with or without a positive culture, should be considered as a wound infection. The rare case of sterile fat necrosis will be included, but this is better than overlooking a true infection. Superficial (suprafascial) wound infection (surgical-site infection) in the early postoperative period is diagnosed on clinical findings, as well as on the results of a Gram stain of needle-aspirated material. The earliest sign of such infections is induration, accompanied by erythema and increasing pain: excessive wound pain is a commonly overlooked sign, particularly in patients with wound infections caused by Gram-negative organisms. Immediate therapy consists of reopening the wound and evacuating the pus; antibiotics are not usually required.

Deep incisional surgical-site infection

Wound infections involving fascia and muscle are more serious, and are usually accompanied by infection in one of the main body cavities, or in bone or a joint; they usually result from a technical error. Drainage, control of any source of continuing infection, and antibiotic therapy are indicated.

Organ-space surgical-site infection

These infections occur within the target organ of the initial operation, as opposed to incisional infection in structures incised to reach the target organ. Examples are postoperative peritonitis from an anastomotic leak or an infected hip joint. These infections are rare. Most organ-space infections bring about severe problems of management, and are associated with high mortality and severe, often lifelong, morbidity.

Postoperative peritonitis

Postoperative peritonitis is an organ-space surgical-site infection. Fifteen to 30 per cent of all intra-abdominal infections occur following an operation. The diagnosis is usually delayed. The most common cause is a technical error compromising the vascular supply to an anastomosis, resulting in necrosis and leakage of intestinal contents into the peritoneal cavity. Iatrogenic perforation of a hollow viscus is another cause. An intra-abdominal hematoma may become secondarily infected, resulting in an abscess. Treatment is operative, as in other forms of secondary peritonitis, although nonoperative drainage with ultrasound or CT guidance is a valid option for an abscess not associated with an anastomosis (Fig. 14). Antimicrobial therapy is more difficult, due to the possible selection of resistant bacteria by preoperative antibiotic therapy. Any antibiotics administered should not only cover the specific bacteria isolated but should also be effective against possible pathogens from the facultative and obligate anaerobic bowel flora; usually a third-generation cephalosporin or a quinolone plus metronidazole are sufficient. Other options are imipenem or one of the extended-spectrum penicillin/b-lactamase-inhibitor combinations. If a resistant *Enterobacter*, or *Serratia*, present problems for treatment, the synergism seen between aminoglycosides and b-lactam antibiotics should be exploited. Enterococci rarely cause a problem and need not be treated specifically when otherwise adequate aerobic and anaerobic antibiotic coverage is provided.

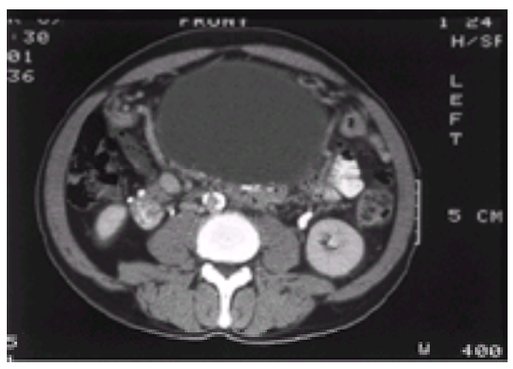


Fig. 14. Large intra-abdominal abscess demonstrated by CT; it may be easily drained percutaneously.

Necrotizing fasciitis

This is a serious mixed infection due to hemolytic streptococci or staphylococci and peptostreptococci, and associated with excessive collagenase production, leading to dissolution of connective tissue. The infection involves the epifascial tissues of an operative wound, laceration, abrasion, or puncture. It may be immediately fulminant or may remain dormant for six or more days before beginning to spread rapidly. Subcutaneous and fascial necrosis accompanies extensive undermining of the skin, resulting in gangrene. Treatment is by excision of the entire area of fascia affected, administration of large doses of penicillin (10 million units every 6 h), and appropriate systemic support.

Pulmonary infections

These are common following thoracic and upper abdominal operations. Pain and the supine posture interfere with adequate diaphragmatic and chest-wall respiration, resulting in atelectasis and subsequent bronchopneumonia or lobar pneumonia. Treatment of mild cases involves physical therapy and adequate analgesia to allow for full expansion of the lungs. If antimicrobial therapy is likely tobe required for suspected pneumonia, transtracheal aspiration should be used to obtain a specimen for culture before therapy is started. If indicated, infected secretions can be removed by bronchoscopy, and a reliable specimen can be obtained for Gram stain, culture,

and antibiotic sensitivity. A variety of micro-organisms, such as streptococci, pneumococci, staphylococci, meningococci, Gram-negative anaerobes, and fungi, may be isolated. The result of the Gram stain will help in choosing the initial antibiotic for calculated therapy. Aspiration pneumonia is usually due to anaerobic oral bacteria. If gastric juice is also aspirated, a severe infection may result (Mendelson syndrome), which has a high mortality. Pleural empyema may develop following thoracic or abdominothoracic operations, or following postoperative pneumonia. The contribution of anaerobic organisms is usually underestimated. Tube thoracostomy drainage and rethoracotomy are treatment options. Initial antimicrobial therapy should be based on the result of a Gram stain and should also include a drug such as metronidazole or clindamycin, active against anaerobes.

Mediastinitis

This carries a high mortality and is most commonly seen following esophageal resection. Treatment consists of adequate operative drainage and the administration of antimicrobials fully active against endotoxin-producing, Gram-negative bacteria and obligate anaerobes; a third-generation cephalosporin, combined with metronidazole, will cover most pathogenic bacteria. Imipenem may be required. Interpretation of bacteriologic results is difficult, since antibiotics will usually have been given before a proper specimen can be obtained during the operative drainage procedure. The gaps in the bacterial coverage of the previously given antibiotics should influence the current choice of antibiotic. Sternal infection is seen following median sternotomy and most commonly is due to staphylococci. If antibiotic treatment is not initially successful, the sternum must be reopened to allow for drainage.

Infection of the urinary tract

A Foley catheter, if required, should be connected to a closed drainage system and should be removed as soon as possible. Specimens of urine should be sent for culture and sensitivity testing every 5 days, and at the time of catheter removal. Suprapubic catheterization is preferable for long-term bladder drainage, as the risk of infection is considerably reduced. A colony count of 10^6/ml bacteria in fresh urine is highly suggestive of active infection. Dysuria is not always present. *Escherichia coli* usually causes hemorrhagic cystitis. If the infecting organism is unknown, a Gram stain will help in selecting an appropriate antibiotic.

Line infection: intravenous catheters, mediports

One-third of intravenous catheters become colonized with bacteria within 2 days of placement. Bacteremia, rarely sepsis, will occur in 1 per cent of patients with an intravenous catheter in place for longer than 48 h, and the risk increases to 5 per cent with increasing duration of catheterization. An intravenous catheter should always be removed and cultured whenever bacteremia is suspected. Mediports and intra-arterial catheters may also be the site of sepsis and should be similarly handled.

Infections of vascular prostheses are most commonly due to staphylococci. Treatment may require removal of the artificial vascular graft. Antimicrobial wound treatment without graft removal succeeds in selected situations.

Antibiotic prophylaxis

Miles and Burke laid the scientific basis for the use of prophylactic antibiotics in surgery in the late 1950s. They were able to show that infections could be prevented only when antimicrobials were given prior to, or at the time of, the infectious challenge. Antibiotics given 3 h following a challenge with infectious bacteria were ineffective in preventing infection. A surgical incision exposes normally sterile tissues to a nonsterile environment; some contamination occurs with any operation. Bacteria may start multiplying before effective host defenses are established, and if initially present in a concentration exceeding 100 000 organisms/g tissue, may exceed the host defense capacity. Host defenses also recognize damaged and dead tissue, which are eliminated by the same humoral and cellular mechanisms as those used to defeat invading bacteria. Thus, there is a need for gentle operative technique to minimize the volume of damaged tissue and for adherence to principles of aseptic surgery to reduce the amount of bacterial contamination.

Following closure of the wound, local intravascular coagulation and the events of early inflammation that initiate wound healing seal its environment: this may explain why the postoperative administration of antibiotics is ineffective in preventing wound infection. Antibiotics administered preoperatively diffuse into the peripheral compartment, in this case the wound fluid; since the wound is saturated with antimicrobials at the time it becomes contaminated, potentially invading bacteria are inhibited from multiplying and many are killed.

The principles of antibiotic prophylaxis are as follows.

1. Use an antibiotic with efficacy against the bacteria likely to contaminate the wound as demonstrated in a controlled clinical trial.
2. Use full doses of the chosen antibiotic.
3. Administer the antibiotics preoperatively at a time such that effective tissue concentrations will have been achieved when intraoperative contamination occurs.
4. If the operation is prolonged beyond 3 or 4 h, give another dose. Otherwise, single-dose prophylaxis is effective in most clinical situations.
5. Employ antibiotic prophylaxis whenever the risk of wound infection is increased.

Single-dose prophylaxis

The first prospective, controlled trial that investigated the proper postoperative duration of antibiotic prophylaxis was performed by Strachan and colleagues in 1977. A single preoperative dose of cefazolin was compared with a regimen of cefazolin given for a period of 5 days after operation. The infection rate following the single dose of antibiotic was 3 per cent; that following multiple postoperative dosing was 5 per cent. This slight numerical difference in favor of the single dose was also found in many subsequent studies that tested the same hypothesis for various indications. The most prudent and conservative interpretation of the results of all of these studies is that, at the very least, single-dose prophylaxis is as effective as multiple dosing, and is preferable because it is less likely to alter the antibiotic-resistance patterns of bacteria in a hospital. It is astonishing in the presence of this evidence (Table 9) that multidose prophylaxis still is used today in many institutions.

Single dose	
Operations performed	510
Infections seen	22
Infection rate	4.3%
Multiple dosing	
Operations performed	493
Wound infections seen	34
Infection rate	6.9%

Table 9 Single-dose versus multidose prophylaxis: results of prospective randomized trials in colon operations (27 studies)

The risk of infection

The risk of developing a wound infection has traditionally been determined by stratifying operations into classes (clean, clean contaminated, contaminated, and dirty) based on the relative degree of intraoperative bacterial contamination. This scheme ignores other important factors, such as the functional state of the host defenses and the amount of tissue trauma engendered by the operation, that also help to determine the risk of developing a wound infection. Regression analysis of outcome in a very large number of surgical patients has indicated that the factors controlling the risk of developing a wound infection are heavy bacterial contamination as reflected in an operation classed as contaminated or dirty, impaired host defenses as reflected in an ASA score of 3 or higher, and long duration of operation. The critical time is procedure-specific (Table 10).

T (hours)	Typical operations
1	Appendectomy, limb amputation, cesarean section
2	Hernia repair, cholecystectomy, exploratory laparotomy, open reduction of fracture, hysterectomy, ventricular shunt, mastectomy
3	Thoracotomy (all), gastric, small bowel, colon operations (all), spinal fusion, joint prosthesis, nephrectomy, vascular surgery (all)
4	Hepatobiliary–pancreatic operations (all), prostatectomy, head and neck surgery, craniotomy
5	Cardiac surgery (all)
7	Organ transplantation

Data from Cruse PJE et al. American Journal of Medicine, 1980; 69(Suppl. 3B): 153–158.
If the duration of operation exceeds T hr, T is the 75th percentile of duration for each specific procedure the risk of infection is increased.

Table 10 Critical values for duration of operation

Any patient exhibiting one or more of these risk factors should be given antibiotic prophylaxis. In addition, prophylaxis should be administered whenever a prosthesis is to be inserted during the operation.

Specific recommendations

In the discussion of selected specific clinical circumstances, we have used the results of published controlled studies that fulfil most of the following criteria: prospective conduct of the study, definition of entrance criteria and drop-out conditions, consideration of risk factors, unbiased randomization, double-blind assessment of clinical and bacteriologic outcome, and appropriate statistical analysis.

Colorectal operations

The risk of wound infection following colorectal operations without prophylaxis is of the order of 40 to 60 per cent. A single dose of parenteral antibiotics or effective bowel preparation with oral antibiotics can reduce the incidence of postoperative infection to less than 10 per cent. The two essential steps in effective prophylaxis for an elective colon operation are thorough mechanical cleansing of the bowel, and oral administration of antibiotics: neomycin and erythromycin base (Table 11). Additional parenteral antibiotics further reduce the rates of postoperative infection. Since the risk associated with administering a single parenteral dose of a b-lactam antibiotic is minimal, we now employ combined oral and systemic prophylaxis (Table 9; Fig. 15). The use of parenteral antibiotics alone is not recommended in elective surgery but is the only regimen available for emergency operations.

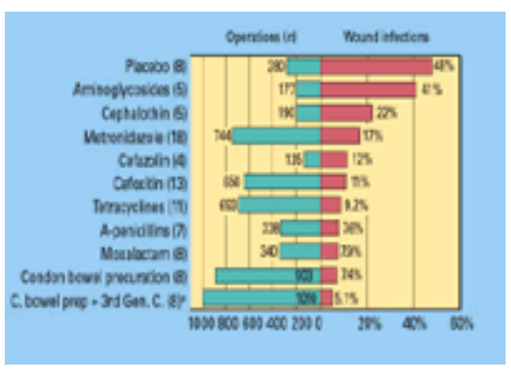


Fig. 15. Surgical site infections following colon operations and antibiotic prophylaxis (numbers in parentheses indicate number of studies). A, extended-spectrum penicillins; *Condon–Nichols bowel preparation (see Table 11) combined with either single dose of 2 g cefotaxime or 2 g ceftizoxime or 2 g cefmenoxime.

Preparation	Description
	[illegible]
	[illegible]
Day 1	Complete evacuation of bowel by normal operation
	[illegible]

Table 11 Preparation for elective colorectal operation: the Condon Nichols prep

Appendectomy

Antibiotic prophylaxis for patients undergoing surgery for acute appendicitis includes an element of treatment for the intra-abdominal infection as well as an element of prophylaxis in relation to infection of the abdominal wound. Antibiotics are of questionable therapeutic value in simple acute appendicitis, but may be therapeutic when perforated appendicitis and peritonitis are present. For all other circumstances, single-dose prophylaxis reduces the risk of wound infection to less than 10 per cent. Following perforation, rates of surgical-site infection are higher and the subcutaneous wound may be left open for healing by secondary intention. Skin closure of such wounds should not be attempted before the seventh postoperative day.

The best results are obtained with antibiotics active against aerobic as well as anaerobic bacteria. In studies in which an aminoglycoside or a cephalosporin were used in combination with either clindamycin or metronidazole, only seven of 261 patients (2.7 per cent) developed a wound infection. The best current regimen seems to be metronidazole combined with a cephalosporin.

Gastroduodenal operations

The incidence of postoperative wound infections following stomach operations correlates directly with the number of bacteria within the stomach, which, in turn, correlates with the acidity of the stomach and the underlying pathology. The ‘acid barrier' that effectively kills swallowed bacteria in normal individuals may be altered by H_2-receptor antagonists or proton-pump inhibitors given during the 12 h before operation. All groups of antibiotics have been studied as prophylaxis for gastroduodenal operations. First-generation cephalosporins are as effective as any other group of b-lactam drugs, and are as effective as aminoglycosides.

Recommended prophylaxis for gastroduodenal operations involves interdicting the use of acid-secretion blocking or -neutralizing agents for 1 or 2 days preoperatively, including when ordered as part of the anesthetic premedication, and the administration of 2 g of cefazolin on induction of anesthesia.

Biliary-tract operations

Although the normal biliary tree rarely harbors any bacteria, in the presence of disease the biliary tract should be viewed as bacterially contaminated. The number of

bacteria is influenced by a variety of factors, primarily the presence of biliary stones or a stricture (Table 12). Of the bacteria isolated from the biliary tree, only few are associated with mortality. Fifty per cent of patients with lethal outcome die from sepsis due to *E. coli* and 25 per cent for clostridial infections. The risk of surgical-site infection following an operation for biliary disease without antibiotic prophylaxis is about 16 per cent. Numerous clinical trials have employed a variety of antibiotic regimens in biliary operations (Table 13).

Bacteria	*No. of isolates*	*% distribution*
Aerobes	1183	95
E. coli	434	35
Klebsiella	134	12
Enterobacter	45	4
Proteus	99	8
Pseudomonas	19	2
Other Gram-negative bacteria	56	5
Streptococci	162	13
Enterococci	150	13
Staphylococci	75	6
Obligate anaerobes	63	5
B. fragilis	9	1
Peptostreptococci	20	2
Clostridia	29	2
Other anaerobes	1	
ALL BACTERIA	1248	100

Data from: Condon RE, Wittmann DH. The use of antibiotics in general surgery. In: Wells, SA, ed. *Current problems in surgery*. Mosby Year Book, Chicago, 1991.

Table 12 Bacteria isolated from biliary tract

Prophylactic agent(s)	*No. of trials*	*No. of patients*	*No. of infections*	*Rate (%) of infections*
Cefuroxime	1	90	13	14
Methicillin	2	359	24	7
Cefotetan/cefoxitin	5	431	24	6
Gentamicin	2	102	11	5
Cefazolin	7	494	26	5
Cefotaxime	2	99	4	4
Cefoxitin	2	92	3	3
Ceftriaxone	5	353	10	3
Piperacillin	1	50	1	2
Cefotaxime	2	87	0	0

Data from: Condon RE, Wittmann DH. The use of antibiotics in general surgery. In: Wells, SA, ed. *Current problems in surgery*. Mosby Year Book, Chicago, 1991.

Table 13 Antibiotic prophylaxis in biliary operations

Single-dose antibiotic prophylaxis is appropriate in surgery of the biliary tract including laparascopic procedures. Although enterococci are frequently recovered from bile, their presence can be ignored in choosing an antibiotic for prophylaxis when cholecystectomy is being performed for calculus disease. The agent of choice remains 2 g of cefazolin given about 30 to 45 min prior to the operation. Second-choice agents may be cefuroxime, cefamandole, sulphamethoxazole–trimethoprim, or cefotaxime. The use of ceftriaxone is not recommended because of its long half-life. Cholangitis (fever, chills, jaundice) involves a more feculent spectrum of bacteria, including the regular presence of anaerobes. In this circumstance, ampicillin–sulbactam, 2 g, is preferred.

Vascular operations

The clinical setting in which most vascular reconstructions are performed increases the risk of infection: electively for arterial insufficiency, implying some degree of tissue ischemia, or as an emergency for trauma. The frequent need for insertion of a prosthesis for bypass or repair further disables host defense mechanisms and increases the risk of infection. Overall, the risk of infection following vascular surgery is of the order of 4 per cent.

The bacteria of concern are *Staph. aureus, E. coli* and other enteric Gram-negative rods, which are the cause of early postoperative infections, and the *Staph. epidermidis* group of organisms, which cause indolent, late graft infections. A few controlled clinical trials have been reported (Table 14), but recent increases in resistance among staphylococci, especially *Staph. epidermidis*, mandate care in the choice of antibiotic for prophylaxis.

Prophylactic agent(s)	*No. of trials*	*No. of patients*	*No. infected*	*Rate (%)*
Cephaloridine	1	27	6	22
Cephradine	1	75	0	0
Cefazolin	1	215	2	1
Cefotaxime	1	134	4	3
Oxacillin	1	168	4	2
Cephalothin	1	232	2	1

Data from: Condon RE, Wittmann DH. The use of antibiotics in general surgery. In: Wells, SA, ed. *Current problems in surgery*. Mosby Year Book, Chicago, 1991.

Table 14 Vascular operations and antibiotics used as prophylactic agents

Recommended prophylaxis is cefazolin administered at the induction of anesthesia, as long as two-thirds or more of *Staph. epidermidis* group isolates in the hospital are sensitive to this antibiotic. If *Staph. epidermidis* resistance is a problem, ampicillin–sulbactam or sulphamethoxazole–trimethoprim are currently effective alternatives. If methicillin-resistant staphylococci are a problem, vancomycin (for the staphylococci) plus cefotaxime or aztreonam (for the coliforms) should be effective.

Other ‘clean' operations

The median infection rate is 3.7 per cent for clean operations when objective, nonbiased assessment methods are used. In six hospitals, infection rates were 4 per cent and higher. It has been shown that in these cases, prophylaxis can reduce infections.

Antibiotic prophylaxis is generally not recommended in clean operations unless a prosthesis is to be inserted during the operation or the patient is immunosuppressed or has impaired host defense mechanisms from another cause. The CDC recommendations identify patients who are at increased risk. One of the major reasons for not recommending antibiotic prophylaxis for most patients in clean operations has been the inability to demonstrate benefit, due to the low rate of infection (less than 3 per cent) thought to characterize such procedures. If infection rates are higher in a specific institution or in the hands of a particular surgeon (Fig. 16), antibiotic prophylaxis seems to be justified. The inability to demonstrate benefit, coupled with the risks of administering antibiotics, which, while low, are not zero for each patient and are clearly negative for the environment, support a judgement that the potential benefits do not outweigh the potential risks. Recent reports, however, document that, for selected operations in the ‘clean' category, the infection risk may be higher than has been assumed, and that antibiotic prophylaxis may be of benefit. Many more data are needed to settle this issue. If it turns out that antibiotic prophylaxis is warranted in some clean operations, it will only confirm the present practice of many surgeons. In many clean procedures (e.g. vascular grafting, total hip-joint replacement) the small reduction of calamitous complications justifies prophylaxis. We recommend prophylaxis in a clean operation if the objectively assessed infection rates surpass 4 per cent. We further recommend support for an objective wound surveillance program with feedback to the surgeon to help in assessing the infection rate objectively as a foundation for prophylaxis.

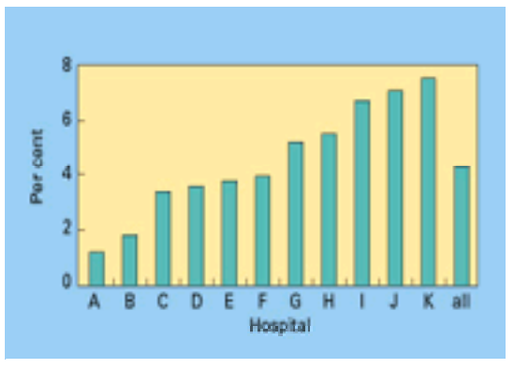


Fig. 16. Infection rates after the same operation (hernia repair) but by different surgeons in different hospitals. *Source*: Wittmann et al. (1995).

Further reading

Bernard HR, Cole WR. The prophylaxis of surgical infection: the effect of prophylactic antimicrobial drugs on the incidence of infection following potentially contaminated operations. *Surgery* 1964; **50**: 161–8. [The first clinical paper dealing with effective prophylaxis.]

Bohnen JMA, Solomkin JS, Dellinger EP, Bjornson HS, Page CP. Guidelines for clinical care: anti-infective agents for intra-abdominal infection: a surgical infection policy statement. *Archives of Surgery* 1992; **127**: 83–8. [Consensus on antibiotic therapy for diffuse peritonitis.]

Burke JF. The effective period of preventive antibiotic action in experimental incisions and dermal lesions. *Surgery* 1961; **50**: 161–8. [Foundation for antimicrobial prophylaxis.]

Condon RE, Wittmann DH. The use of antibiotics in general surgery. In: Samuel A. Wells, ed. *Current problems in surgery*. Mosby Year Book, Chicago, 1991. [Specific guidelines for therapy and prophylaxis of surgical infections.]

Culver DH *et al.* Surgical wound infection rates by wound class, operative procedure, and patient risk index. *American Journal of Medicine* 1991; **91** (Suppl. 3B): 152–7S. [Classification of operations by risk factors.]

Finald M, McGowan JE, Jr. Nosocomial infections in surgical patients. Observations on effects of prophylactic antibiotics. *Archives of Surgery* 1976; **111**: 143–5. [Classical paper based on the analysis of many operations.]

Friedrich PL. Zur bacteriellen Aetiologie und zur Behandlung der diffusen Peritonitis. *Archiv fur Klinische Chirugie* 1902; **68**: 524. [Fundamental paper on the bacterial synergism in surgical infections in general and peritonitis in particular.]

Horan TC, Gaynes RP, Martone WJ, Jarvis WR, Emori TG. CDC definitions of nosocomial surgical site infections, 1992: a modification of CDC definitions of surgical wound infections. *American Journal of Infection Control* 1992; **20**: 271–4. [Most accepted definition of postoperative surgical infection.]

Kanus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine* 1985; **13**: 818. [Most accepted score for risk factor assessment and disease specific mortality prediction.]

Karl RC., Mertz JJ, Veith FC, Dineen P. Prophylactic antimicrobial drugs in surgery. *New England Journal of Medicine* 1966; **275**: 305–8. [Classical paper on prophylaxis.]

Miles AA, Miles EM, Burke J. The value and duration of defense reactions of the skin to primary lodgment of bacteria. *British Journal of Experimental Pathology* 1957; **38**: 79–96. [Fundamental experiments for single preoperative dosing of antimicrobial prophylaxis.]

National Nosocomial Infections Surveillance (NNIS) System. Nosocomial infection rates for interhospital comparison: limitations and possible solutions. *Infection Control and Hospital Epidemiology* 1991; **12**: L609–21. [Addresses limitations of wound surveillance.]

Nyström PO *et al.* Proposed definitions for diagnosis, severity scoring, stratification, and outcome for trials on intraabdominal infection. *World Journal of Surgery* 1990; **14**: 148–58. [Consensus on definitions for intra-abdominal infections.]

Polk HD, Jr, Lopez-Mayor JF. Postoperative wound infection: a postoperative study of determinant factors and prevention. *Surgery* 1969; **66**: 97–103. [Fundamental paper on antibiotic prophylaxis.]

Schein M *et al.* The abdominal compartment syndrome. The physiological and clinical consequences of elevated intra-abdominal pressure. *Journal of the American College of Surgeons* 1995; **180**: 745–53. [Important collection of papers requested from European and American experts addressing the global overuse of antibiotics and focusing on the benefits of short therapeutic courses of 1 to 5 days.]

Seabrook GR. Pathobiology of graft infections. *Seminars in Vascular Surgery* 1990; **3**: 81–8.

Simchen E *et al.* The Israeli study of surgical infection of drains and the risk of wound infection in operations for hernia. *Surgery, Gynecology, Obstetrics* 1990;**170**: 331–7. [Excellent study on factors influencing postoperative infection rates.]

Strachan CJ *et al.* Prophylactic use of cefazolin against sepsis after cholecystectomy. *British Medical Journal* 1977; **1**: 1254–6. [First randomized trial on single dose prophylaxis.]

Wittmann DH. *Intra-abdominal infection: pathophysiology and treatment*, 3rd edn Marcel Dekker, New York, 2000. [Synoptic and detailed overview on operative and nonoperative therapy of intra-abdominal infections.]

Wittmann DH. Peritonitis, intra-abdominal infection, and intra-abdominal abscess. In: Schwartz SI, Shires GT, Spencer FC, ed. *Principles of surgery*, pp. 1449–84. McGraw-Hill, New York, 1993. [Detailed book chapter on diffuse suppurative peritonitis.]

Wittmann DH. Compartment syndrome of the abdominal cavity. In: Irwin RS, Cerra FB, Rippe JM, ed. *Intensive care medicine*, 4th edn. Lippincott-Raven, Philadelphia, 1998, pp.1888–904. [First textbook chapter on etiology, pathophysiology, and treatment of the abdominal compartment syndrome.]

Wittmann, DH. Prevention of infection. In: Hau T, Fahrtmann E, ed. *Textbook of surgical infection*. McGraw-Hill, New York, 1998. [Overview on infection control and wound surveillance.]

Wittmann DH, Wittmann-Tyler. Scope and limitations of anti-microbial therapy of sepsis. *Langenbecks Arch fur Chirurgie* 1998; **383**; 15–25. [Focus on antimicrobial concentrations required in infected tissue to kill bacteria causing surgical infection.]

Wittmann DH *et al.* Calculated empiric antibiotic treatment of acute surgical infections. *Infection* 1991; **19**: S345–50. [Guidelines for empiric antibiotic therapy of surgical infections.]

Wittmann DH *et al.* Antibiotic prophylaxis in abdominal wall hernia surgery: never, always, or selectively? *Problems in General Surgery* 1995; **12**: 47–55. [Rationale for antibiotic prophylaxis in clean (aseptic) operations.]

Wittmann DH *et al.* Pyoderma fistulans sinifica (fox den disease): a distinctive soft-tissue infection. *Clinical Infectious Diseases* 1995; **21**(1): 162–70. [Example of the complex bacteriology of a typical spontaneous surgical infection.]

Wittmann DH *et al.* Management of secondary peritonitis. *Annals of Surgery* 1996; **224**(1): 10–18. [Comprehensive summary of current knowledge on the topic.]

4.2 The surgery associated with HIV infection

C. Wastell and Peter A. Davis

Introduction
Origin of the virus
Clinical horizon
Neoplasms associated with HIV/AIDS
Opportunistic infections in HIV/AIDS
The world situation
Classification of HIV disease
Recognition of the HIV-positive patient
The HIV-positive health care worker
Measures to minimize the risks of nosocomial infection
Protection of health care workers in the operating theatre
Safe practice
Occupationally acquired HIV
Transmission from health care worker to patient
Sharps injuries and the risk of infection
Postexposure prophylaxis
Surgery on HIV-positive patients
Indications for surgery
Conclusion
Further reading

Introduction

Origin of the virus

A plasma sample taken from a male from the Democratic Republic of the Congo in 1959 was shown, by Ho in 1986, to be seropositive for the human immunodeficiency virus-1 (**HIV-1**). This plasma sample was subsequently re-examined and the HIV-1 sequences submitted to phylogenetic analysis which suggested that subtypes A to J share a common ancestor dating from the late 1940s or early 1950s.

Subsequently, the virus was shown to be an RNA retrovirus similar to those responsible for simian disease. It contains the enzyme reverse transcriptase capable of altering the RNA nucleus to a DNA template. The virus has an outer protein envelope, the capsid, containing the GP120 protein capable of attaching to the CD4 receptor on the surface of the CD4 T lymphocyte. Following a process of inclusion, the viral RNA is subjected to the transcriptase altering it to DNA prior to inclusion into the genome of the host T_4 lymphocyte. The virus exists in two forms, HIV-1 and HIV-2. HIV-1 is responsible for the majority of the disease in the world. HIV-2 is found almost exclusively in West Africa.

Clinical horizon

The pandemic was heralded by the description by Gottlieb in 1981 of five homosexual males in California who all had *Pneumocystis carinii* pneumonia—rare except in circumstances associated with cellular immune suppression.

From a clinical standpoint, primary infection may be associated with a flu-like illness and the infected persons' blood can be shown to contain viral RNA. The significance of this timing in the sequence is that presumably there is a point following viral implantation before inclusion in the CD4 T cell when postexposure prophylaxis may be effective in preventing infection. After about 12 weeks the majority of infected persons mount an antibody response and so become seropositive (Fig. 1). There then follows a period of 10 to 12 years during which very little in the way of symptoms occur. However, at the end of this ‘silent period', during which the CD4 T count has been gradually diminishing and viraemia is generally but not always undetectable, the patient becomes significantly immunosuppressed. At this time with a CD4 T count of less than 400/mm^3, diseases associated with symptomatic HIV infection become evident. Following this, in those patients in whom the disease is not modified by treatment (a situation for the majority of those infected), a full picture of AIDS emerges and the patient dies.

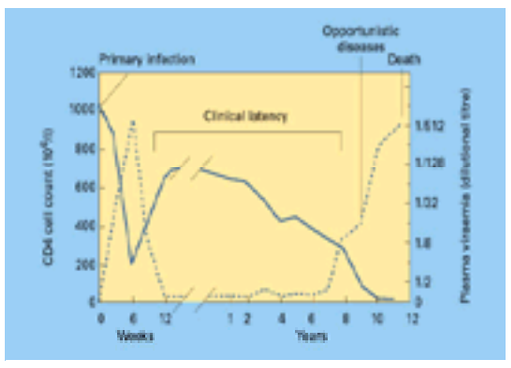


Fig. 1. The natural history of HIV infection. After Jollie, et al BMJ 312; 1244. 1996. — CD4 T lymphocyte count; ------- viral titre.

Neoplasms associated with HIV/AIDS

Reduced cellular immunity is associated with the development of certain neoplasms. Kaposi's sarcoma was present in 14 per cent of the AIDS diagnoses in Table 1. It usually presents as a pigmented, multifocal skin lesion (Fig. 2). It may occur anywhere on the surface of the skin and also affect any tissues of the body. When it involves the lung it has a characteristic fuzzy shadow on radiography (Fig. 3), which is associated with a clinical course that is slower than other acute conditions like bronchopneumonia. It also occurs in the gut from the mouth to the anus and in this situation may be associated with abdominal pain. Diagnosis is by biopsy. It is sensitive to radiotherapy but it may be too diffuse for this to be effective.



Fig. 2. Kaposi's sarcoma (by courtesy of Professor B Gazzard, Chelsea Westminster Hospital).



Fig. 3. Kaposi's sarcoma of the lung (by courtesy of Dr S. Padley, Chelsea Westminster Hospital).

Pneumocystis carinii pneumonia	22
Candidiasis	14
Kaposi's sarcoma	14
Mycobacterium tuberculosis	10
Wasting	5
Encephalopathy	3
Lymphoma	9
Mycobacterium—other	7
Cytomegalovirus	3
Toxoplasmosis	2
Other opportunistic infections	9

Diseases which indicated the presence of AIDS in 373 adult males with 413 diagnoses in England, Wales, and Northern Ireland when the interval between the diagnosis of HIV positivity and AIDS was 3 months or greater, from data to the end of March 1998. *Communicable Disease Report* 1998, 8: 17.

Table 1 Diseases indicating the presence of AIDS (in percentages).

Perhaps not directly due to AIDS, reduced cellular immunity is associated with infection by the human papilloma virus (**HPV**), of which types 16 and 18 and possibly others may be associated with epithelial neoplasia. Affected tissues include the anal mucosa and the cervix.

Another opportunistic infection is the Epstein–Barr virus (infectious mononucleosis), which may be found in B cells in the peripheral blood and in tumour cells in Burkitt's lymphoma and nasopharyngeal carcinoma. It is implicated in the B-cell polyclonal lymphoproliferative disorder seen in patients on immunosuppression after receiving transplants and with lymphoma of the B-cell variety occurring with AIDS. This may affect a wide variety of tissues including lymph nodes, gut, lung, and brain. Treatment is by a combination of chemotherapy and radiotherapy.

Opportunistic infections in HIV/AIDS

Viral

- Herpes simplex types 1 and 2
- Varicella zoster (shingles)
- Cytomegalovirus
- Molluscum contagiosum

Bacterial

- (*Mycobacterium tuberculosis*
- *Mycobacterium avium-intracellulare*
- *Mycobacterium kansasii*
- Folliculitis
- Impetigo
- Dental/gum disease

Fungi

- Candida
- Cryptococcus

Protozoa

- *Pneumocystis carinii*
- Toxoplasma
- Cryptosporidia
- Isosporidia

The world situation

It is estimated that there are 30.6 million people living with HIV/AIDS. Since 1981, 11.7 million have died of the disease leaving 8.2 million orphans. There is some evidence that public health measures, the employment of ‘safe sex' by the use of condoms, and awareness of the nature of the disease together with new treatment systems with drugs have resulted in a reduction in the incidence of new AIDS diagnoses in parts of the developed world, for example Western Europe and the United States. Unfortunately, however, the disease continues in the greater part of the world unchecked and there are an estimated 16 000 new infections worldwide per day (5.8 million new infections in 1997) of which 90 per cent occur in developing countries.

The distribution of HIV/AIDS in various parts of the world together with the principal mode of transmission is given in Table 2. More detailed information on the mode of transmission for the United Kingdom, United States, and Canada is provided in Table 3. All figures are estimates and the distribution is not uniform; for example, although the adult prevalence in Europe is 0.3 per cent this varies between a much lower rate in rural areas and a higher rate in London and other urban centres.

Region	Epidemic started	Adults and children living with HIV/AIDS	Adult prevalence (%)	Major mode of transmission
Sub-Saharan Africa	Late 1970s–early 1980s	20.8 million	7.4	Heterosexual
South and South-East Asia	Late 1980s	6.3 million	0.6	Heterosexual
Caribbean	Late 1970s, Early 1980s	340,000	1.9	Heterosexual, MSM
Western Europe	Late 1970s, Early 1980s	530,000	0.3	MSM, IDU
North America	Late 1970s, Early 1980s	860,000	0.6	Heterosexual, MSM, IDU
Australia, New Zealand	Late 1970s, Early 1980s	12,000	0.1	MSM, IDU

Heterosexual—sexual intercourse between men and women; IDU—intravenous drug abuse; MSM—sexual intercourse between men.

Table 2 Data from the Report on the Global HIV/AIDS Epidemic by the Joint United Nations Programme on HIV/AIDS (UNAID) and World Health Organization (WHO) November 1997.

Country	Adult (%)	Adult & [illegible] (%)				
		Heterosexual	MSM	IDU	Blood	Mother to child
UK	0.1	0.1	7.0	1.0	0.5	0.1
USA	0.6	0.1	5.2	1.0	0.2	0.1
Canada	0.3	0.1	2.1	1.0	0.2	0.1

[illegible] adult and blood products. Heterosexual, sexual intercourse between men and women. IDU intravenous drug use. MSM sexual intercourse between men.

Table 3 Data from the Report on the Global HIV/AIDS epidemic by UNAID and WHO June 1998.

Classification of HIV disease

No really satisfactory classification of HIV/AIDS has been developed because of the difficulties in relating it to a system of staging. It is however important to provide a framework whereby groups of patients can be compared and if possible providing an indication both for prognosis and treatment. One of the current systems by the Centre for Disease Control and Prevention (**CDC**) combines three groups of CD4 T cell counts:

1. more than 500/mm^3 (or CD4 more than 28 per cent);
2. 200 to 499/mm^3 (or CD4 14 to 28 per cent); and
3. less than 200/mm^3 or (CD4 less than14 per cent);

with three categories of clinical groupings. A simplified version of these groups is as follows:

A. Asymptomatic HIV infection

Persistent generalized lymphadenopathy (Fig. 4)



Fig. 4. Persistent generalized lymphadenopathy. The ‘x' merely marks the node to be biopsied.

(Acute (primary) HIV infection

B. Bacillary angiomatosis

Candidiasis—oropharyngeal
Candidiasis—vulvovaginal poorly responsive to therapy
Cervical dysplasia
Constitutional symptoms, such as fever for more than 1 month
Oral hairy leucoplakia
Herpes zoster involving two episodes or more than one dermatome
Idiopathic thrombocytopenia purpura
Listeriosis
Pelvic inflammatory disease
Peripheral neuropathy

C. Includes the clinical conditions listed in the 1993 AIDS surveillance case definition, some of which are included in Table 1.

Recognition of the HIV-positive patient

Patients from areas of high prevalence, such as sub-Saharan Africa (the adult rate in Namibia, Zambia, and Zimbabwe is 19 to 26 per cent, for example) and the Caribbean, carry greater risk of being HIV positive than those from low prevalence areas. Homosexuals and their partners, partners and children of known HIV-positive patients, intravenous drug addicts, haemophiliacs, and those who have received blood or blood products including packed red cells, plasma, factor VIII and IX concentrates, and platelets, before 1986 are all more likely to be HIV positive. It should be noted that neither albumin nor immune globulins carry HIV.

As with all diagnoses the history elicited from the patient is important and difficult to obtain particularly if there are aspects of lifestyle consistent with behaviour likely to result in HIV transmission.

In the ‘silent period' there are no reliable physical signs relating specifically to HIV infection, though since a large number of referrals from the HIV-positive population are for anorectal problems of one sort or another, these together with evidence of intravenous drug use by needle puncture marks should raise suspicions of high-risk behaviour.

Following an acute HIV infection an antibody response is mounted in the majority of patients by the 12th week. However, Gerbeding reports a delayed response in two patients, one only mounting a detectable response by 6 months and the other by 1 year. This serological response can be detected by an enzyme-linked immunosorbent assay. In addition it is possible to measure the viral load either by detection of the p24 core-protein or polymerase chain reaction.

Routine testing of all patients prior to surgery is not recommended by the Centre for Disease Control and Prevention. However, when dealing with high-risk patients, testing is advisable so that appropriate infection control measures can be undertaken. In addition, from the patients' own point of view, treatment can now be expected

to reduce morbidity and prolong life as well as providing prophylaxis against such infections as *Pneumocystis carinii* pneumonia. However, testing is only permissible with the explicit consent of the patient. It is always accompanied by appropriate counselling. If testing is refused, the patient is automatically considered to be at high risk of being HIV positive. An exception to this rule is when a sharps injury has occurred to a health care worker from a patient unable to give consent to testing, for example when the patient is unconscious. Since postexposure prophylaxis depends on the definition of the risk, HIV testing in this circumstance is permissible without patient consent.

The HIV-positive health care worker

If a health care worker believes that he or she may be HIV positive, it is their duty to inform their employing authority and to place themselves under the care of an appropriate physician. Routine testing of health care workers, even if working in high prevalence areas, is not advised. There are cost issues together with the risk of false positive or negative results and always the difficulty of the time between infection and the development of an antibody response, which can result in a false negative test.

Measures to minimize the risks of nosocomial infection

Since AIDS was first recognized as being due to a viral agent, methods have been developed to prevent infection of health care personnel working in an environment in which HIV, hepatitis B virus (**HBV**), and other viruses are prevalent. Since 1985 CDC has developed the strategy of ‘universal precautions', which broadly speaking includes the idea that if all patients coming to an invasive procedure are regarded as being potentially infective and if all precautions are taken to prevent infection then nosocomial infection will not occur. This approach has the virtue of simplicity and clarity but suffers the disadvantage that truly universal precautions include the use by surgeons at least of double gloves, eye protection, and the non-handling directly of sharp instruments, for example needles, which some regard as a disadvantage. Another disadvantage is that truly universal precautions are ‘universally' ignored. Nevertheless, it is axiomatic that an effective barrier be created between patient, surgeon, and indeed all health care workers. It is difficult however for surgeons in particular to accept measures which interfere with an operation when working in an area of low endemicity on a patient in whom infection with HIV or other viruses is unlikely. A more realistic approach, certainly in the operating theatre, is to employ precautions that are applicable to every patient (which are summarized below), and then additional measures with high-risk patients.

Protection of health care workers in the operating theatre

1. With all cases—universal precautions

    Scrub-up ritual
    Use of impervious gown, cap, mask, shoes/boots
    Use of gloves impervious to viruses
    No hand to hand passage of sharps
    Needles only to be resheathed using safe system
    Finger not to be used as a needle guide
    Eye protection against blood or tissue aerosols (such as procedures employing high-speed drills)
    Correct disposal of sharps
2. Additional with high-risk patients—full kit

    Double gloves
    Eye protection
    No hand-held needles, the use of blunt needles where possible
    Footwear protection

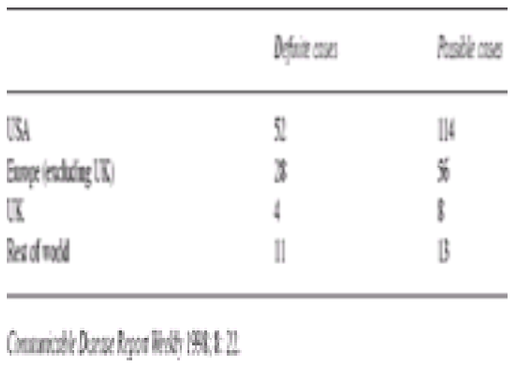
Safe practice

Although most of the above lists are self-explanatory, it should be emphasized that it is the responsibility of the health care worker who undertakes an invasive procedure to ensure disposal of any sharps in an appropriate receptacle. In the operating theatre this rule is modified as a sharp is handed in a transport vehicle to the scrub assistant. Double gloving has been shown to confer some protection and markedly reduces the incidence of skin contamination following perforation of the outer of the two gloves. It is advised that the inner glove be half a size larger than the usual size which is worn on the outside. This, perhaps surprisingly, has been found to be the most comfortable method. A useful variation is to employ Biogel Reveal gloves in which the inner glove is coloured green and fluid entering from a puncture in the outer glove to the interface is readily apparent, in which case both gloves need to be changed. Eye protection is necessary when blood or tissue aerosols are likely, for example when high-speed drills or power tools are used, and in any case when operating on known high-risk patients.

In a situation where high-risk patients are undergoing operation other precautions need to be taken. First, the procedure should be carried out where all staff are trained to understand the nature of the risks. Surgical trainees are more likely to suffer sharps injuries when operating and therefore fully trained surgeons should undertake these procedures. Patients who are infected with HIV are more likely to have a high viral load if they have a low CD4 count and AIDS and it is these who represent the highest risk. It should be noted that all surgeons and other operating department workers should be protected against HBV by vaccine and be aware of the fact that hepatitis C virus (**HCV**) is more infective than HIV and probably represents a greater hazard.

Occupationally acquired HIV

This was first reported in 1984 from Africa and the current situation is summarized in Table 4. Eighty-nine per cent of the definite cases followed percutaneous exposures and most of both definite and possible cases were related to venepuncture. Where data are available, nurses accounted for 39 per cent of cases, doctors or medical students 16 per cent, clinical laboratory workers 13 per cent, and dental workers 3 per cent.



	Definite cases	Possible cases
USA	52	114
Europe (excluding UK)	28	56
UK	4	8
Rest of world	11	13

Communicable Disease Report Weekly 1998; 8: 22.

Table 4 Definite and possible cases of occupationally acquired HIV infection cumulative to December 1997.

Considering the numbers of invasive procedures that have been undertaken worldwide these figures of occupationally acquired infection are low. However, it should be pointed out that reportage is likely to be much more accurate in developed countries where the adult rate is low and the capacity to mount effective infection control measures is high. The rate in the developing world, particularly sub-Saharan Africa, is unknown.

The fact that the majority of cases occur surrounding venepuncture underlines the need for a rational approach to infection control in the operating theatre. Whilst not forgetting the threat posed by other viruses (particularly HCV), expensive, unnecessary, and especially untested measures are not effective.

Transmission from health care worker to patient

Transmission of HIV from an infected health care worker to patient is excessively rare. There is the carefully documented transmission from a dentist working in Florida

to six of his patients. However, this cluster is so remarkable in the face of an otherwise rare event as to suggest extraordinary circumstances in the transmission. There is only one other case that has been reported, and this from a French orthopaedic surgeon to one of his patients.

Sharps injuries and the risk of infection

It should be noted that the transmission of HIV and other blood-borne viruses including hepatitis B, C, D, and E as well as other infective agents is similar. In the health care setting transmission may be by blood, infected body fluids, or concentrated virus either being implanted by percutaneous inoculation or coming into contact with an open wound, non-intact skin, or mucus membranes. Body fluids that may contain HIV include blood and all blood-contaminated fluid as well as amniotic, pericardial, peritoneal, pleural, and synovial fluids, cerebrospinal fluid, semen, and vaginal secretions. The average risk of those who suffer percutaneous exposure to HIV-infected blood has been calculated to be 0.25 to 0.3 per cent, to mucus membrane exposure 0.09 per cent, and although skin exposure constitutes a risk, the precise figure is not calculable. Although seroconversion by injury with a solid sharp such as a knife blade has been reported, by far the most dangerous sharp injury is with a hollow needle. There is some evidence that host defence mechanisms influence the risk of seroconversion.

Postexposure prophylaxis

In 1996 and 1998 CDC issued guidelines with regard to the procedures that should be undertaken when a health care worker comes into contact with any of the hazards mentioned previously. The requirements can be summarized as follows:

(1) the generation of an ‘exposure report';
(2) management of the exposure event itself:
site—wash with soap and water,
assessment of the risk of infection which is summarized in Fig. 5 and Fig. 6,

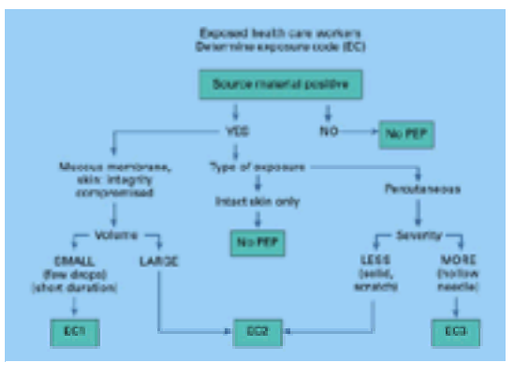


Fig. 5. Algorithm to determine the exposure code. EC, exposure code (1, 2, 3); PEP=postexposure prophylaxis.

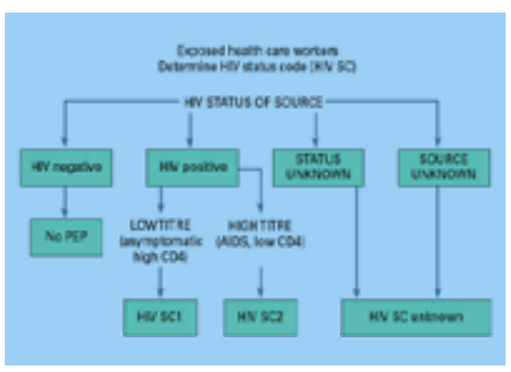


Fig. 6. Algorithm to determine the HIV status code of the source of the exposure. HIV SC=HIV status code (1, 2); PEP=Postexposure prophylaxis.

the need for postexposure prophylaxis is summarized in Fig. 7,

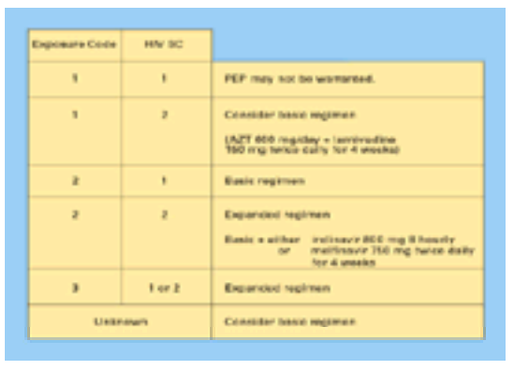


Fig. 7. Guidelines for postexposure prophylaxis. EC=exposure code (1, 2, 3); HIV SC=HIV status code (1, 2); PEP=Postexposure prophylaxis.

follow-up for HBV and HCV;
(3) baseline testing of health care worker for HIV antibody plus retesting at 3 and 6 months;
(4) pregnancy testing if appropriate; and
(5) counselling and education.

If it is decided that prophylactic drugs should be offered then these should be started with as little delay as possible. Both the basic regimen of zidovudine (**AZT**) and lamivudine and the expanded regimen including the addition of either indinavir or melfinavir should be continued for 4 weeks (Fig. 7).

Surgery on HIV-positive patients

In a series of over 4000 patients with HIV/AIDS, Davis found that nearly 25 per cent underwent a surgical procedure of one sort or another and that a significantly greater percentage required an outpatient surgical opinion. By far the most common procedure in these patients was the establishment of venous access to facilitate the administration of foscarnet for cytomegalovirus retinitis. This condition was found to be the commonest but by no means the only ocular complication of AIDS.

Another important surgical requirement was biopsy either of the lymph node to differentiate between persistent generalized lymphadenopathy (Fig. 4), lymphoma, and mycobacterium tuberculosis for example or biopsy to confirm Kaposi's sarcoma. Other minor procedures included inguinal hernia repair and varicose veins.

A considerable surgical load is represented by the anorectal complications of this disease. Anal discomfort often associated with diarrhoea, anal warts (Fig. 8 and Fig. 9), fissuring and perianal ulceration, and a persistent mucus leak are grouped together as the ‘AIDS anus syndrome'. Since outpatient examination was often difficult and painful, examination under anaesthetic often had to be resorted to and was carried out in the above series on 92 patients.

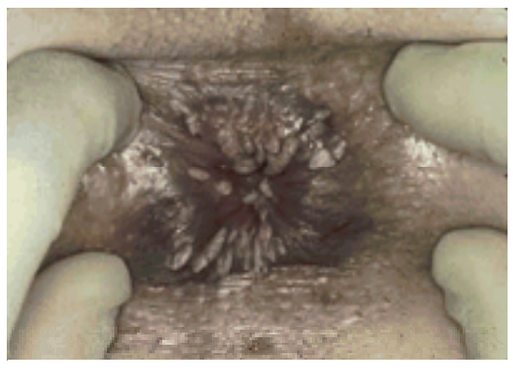


Fig. 8. Anal warts.



Fig. 9. Confluent anal warts.

As might be expected with reduced immunity, perianal abscesses are common. One surprising finding has been that perianal ulceration responds to ulcer excision, if necessary on more than one occasion. Anal warts are a common and distressing problem. Types 16 and 18 of the human papilloma virus and possibly others are associated with the progressive development of epithelial neoplasia. Various degrees of intraepithelial neoplasia occur. They do not seem necessarily to develop into an infiltrating carcinoma of the anus and should be kept under close surveillance. Typical carcinoma both of the anus and rectum together with lymphoma and Kaposi's sarcoma does occurand requires biopsy followed by appropriate therapy. In dealing with the anorectal problems the first priority is to screen for specific infections which include syphilis, gonorrhoea, herpes, and cryptococcus and then to treat the surgical conditions along distinctly conservative lines. The best approach is to avoid surgery if at all possible.

Major surgery is not infrequently required in these patients. Some but not all is HIV/AIDS related. Total colitis due to cytomegalovirus can progress from the type of appearance seen on sigmoidoscopy (Fig. 10) to acute toxic dilatation of the colon (Fig. 11). This latter state precedes perforation, which is invariably fatal, whereas total colectomy is perfectly feasible in spite of CD4 counts well below 20 with patient recovery. Appendicitis is common and may be related to cytomegalovirus involvement of the appendix. Because of the severity of anorectal disease, a terminal left iliac colostomy may improve a patient's quality of life. Biopsy of intra-abdominal masses defined by ultrasound, CT, or MRI may be required, though needle biopsy is often successful and is less invasive (Fig. 12). Other more ordinary procedures such as cholecystectomy may be indicated and should simply be dealt with on their merits.

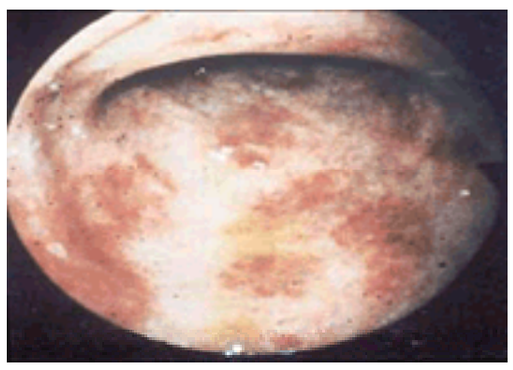


Fig. 10. Flexible sigmoidoscopic view of proctitis due to cytomegalovirus (by courtesy of Professor B. Gazzard, Chelsea Westminster Hospital).



Fig. 11. Plain radiograph of the abdomen showing toxic dilatation of the colon (by courtesy of Professor B. Gazzard, Chelsea Westminster Hospital).

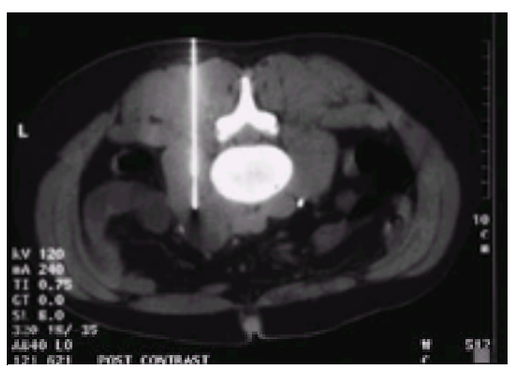


Fig. 12. CT-guided biopsy of large para-aortic node involved with mycobacterium tuberculosis in an HIV-positive male (by courtesy of Dr S. Padley, Chelsea

Westminster Hospital).

Abdominal pain in HIV/AIDS is a common problem and certainly in patients with very low CD4 counts poses a serious diagnostic problem. Laparoscopy is occasionally of use, providing an opportunity for biopsy of any suspicious lesion and the liver. The liver is a frequent source of right upper quadrant pain, particularly if involved with AIDS-related sclerosing cholangitis (Fig. 13). Splenectomy may be required for hypersplenism with a haematological cytopenia associated with HIV and AIDS with satisfactory haematological response.



Fig. 13. Endoscopic retrograde cholangiopancreatography showing AIDS-related sclerosing cholangitis (by courtesy of Professor B. Gazzard, Chelsea Westminster Hospital).

Persistent spontaneous pneumothorax can occur as a complication of *Pneumocystis carinii* pneumonia and may require pleurodesis (Fig. 14). Intrapulmonary neoplasms, for example Kaposi's sarcoma or lymphoma (Fig. 3 and Fig. 15), are seen and require biopsy so that appropriate therapy can be mounted.

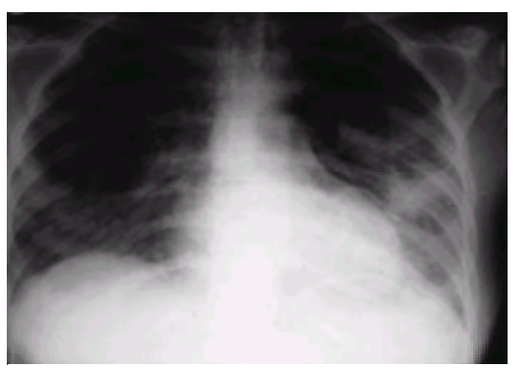


Fig. 14. *Pneumocystis carinii* pneumonia. There is a small pneumothorax at the left apex and a pneumomediastinum indicated by the double outline of the upper left heart (by courtesy of Dr S. Padley, Chelsea Westminster Hospital).

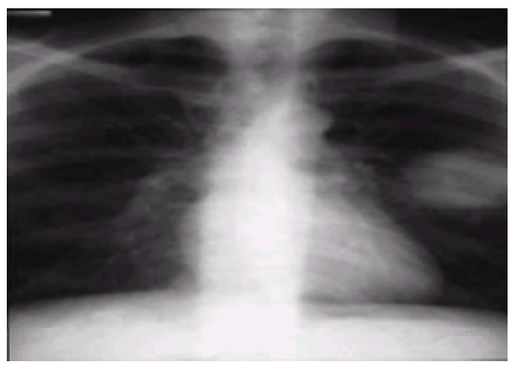


Fig. 15. Lymphoma of the left lung (by courtesy of Dr S. Padley, Chelsea Westminster Hospital).

Indications for surgery

The existence of AIDS is not a contraindication for surgery. Patients submitted to major abdominal operations are able to withstand the trauma and heal their wounds perfectly satisfactorily. Any advantage gained in terms of quality of life should not be denied simply on the basis that the patient has AIDS. Although wound healing has been reported to be deficient, particularly in patients with AIDS and severe anorectal disease and in patients who have both AIDS and abdominal malignancy, apart from reasonable surgical caution wound healing is not an insuperable problem.

Conclusion

Perhaps the details of HIV/AIDS are not directly important so far as surgical care is concerned, they do however serve to underline the complexity and diverse nature of the diseases which are AIDS related that in turn relate to a failure of cellular immunity. HIV/AIDS conditions require the care of specialized physicians skilled in the knowledge of the many drugs that are now available and are changing the scene. The surgeon's role is thus not to treat HIV/AIDS as such but to recognize it and its associated diseases for what they are; often to assist in making specific diagnoses by, for example, biopsy and in treating ‘surgical' conditions in HIV-positive patients appropriately and safely.

Further reading

CDC: Public Health Service Guidelines for the Management of Health-Care Worker Exposures to HIV and Recommendations for Postexposure Prophylaxis, 1998. *Morbidity and Mortality Weekly Reports* 1998; **47** (RR-7).

Communicable Disease Report Weekly 1998; **8**: 17.

Communicable Disease Report Weekly 1998; **8**: 22.

Davis P, Corless DJ, Gazzard B, Wastell C. Wound complications following surgery in HIV seropositive patients. *Digestive Surgery* 1999; in press.

Gerbeding JL, Littell C, Tarkington A, Brown A, Schechter WP. Risk of exposure of surgical personnel to patients' blood during surgery at San Francisco General Hospital. *New England Journal of Medicine* 1990; **322**: 1788–93.

Gottlieb MS *et al.* Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men. *New England Journal of Medicine* 1981; **305**: 1425–31.

Hammond JS *et al.* HIV, trauma and infection control: universal precautions are universally ignored. *Journal of Trauma* 1990; **30**: 555.

Heinemann MH. Ocular manifestations of human immunodeficiency virus and acquired immunodeficiency syndrome. *Problems in General Surgery*, Vol. 11, pp. 267–73. JB Lippincott Company, Philadephia, 1994.

Ho D. *Nature* 1998; **391**: 594–7.

Ho D. *Lancet* 1998; **351**: 351–423.

Morbidity and Mortality Weekly Reports 1989; **38** (S-6).

UNAID/WHO. Report on the global HIV/AIDS epidemic. Joint United Nations Programme on HIV/AIDS (UNAID) and World Health Organization (WHO). November 1997.

UNAID/WHO. Report on the global HIV/AIDS epidemic. Joint United Nations Programme on HIV/AIDS (UNAID) and World Health Organization (WHO). June 1998.

5.1 Haematological problems

Paul L. F. Giangrande and T. J. Littlewood

Introduction
Blood coagulation
Laboratory tests of haemostasis
Acquired disorders of coagulation
Disseminated intravascular coagulation
Liver disease
Chronic renal failure
Vitamin K deficiency
Anticoagulants
Massive blood transfusion
Cardiopulmonary bypass
Congenital disorders of coagulation
Haemophilia
Other congenital disorders of coagulation
Investigation of the patient with a bleeding tendency
Thrombophilia
Management of individuals with thrombophilia
Haemoglobinopathies
Sickle cell disease
Sickle cell trait
Thalassaemia
Myeloproliferative disorders
Lymphoproliferative disorders
Myelodysplasia
Autoimmune disorders
Idiopathic thrombocytopenic purpura (ITP)
Autoimmune haemolytic anaemia (AIHA)
Further reading

Introduction

Bleeding in association with surgery is a common problem encountered by surgeons and a cause of referral to haematologists. An understanding of the mechanism of blood coagulation is important in order to understand the basis of disorders of haemostasis and the common laboratory tests.

Blood coagulation

The fundamental step in blood coagulation is the formation of insoluble fibrin strands. The cleavage of small polypeptide chains from the soluble parent fibrinogen molecule is sufficient to achieve this transformation. However, this is only the last step in a series of enzymatic reactions that take place during coagulation (Fig. 1). The coagulation cascade is initiated in two ways. The extrinsic arm is activated when tissue factor forms a complex with factor VII. The resultant complex activates factor X directly, which has a central role in both pathways. Factor X may also be activated through the intrinsic pathway, when negatively charged subendothelial collagen is exposed and activates factors XII and XI. Factor X forms a complex with calcium and factor V. This complex cleaves prothrombin to produce thrombin, which in turn cleaves fibrinopeptides A and B from soluble fibrinogen to yield insoluble fibrin strands.

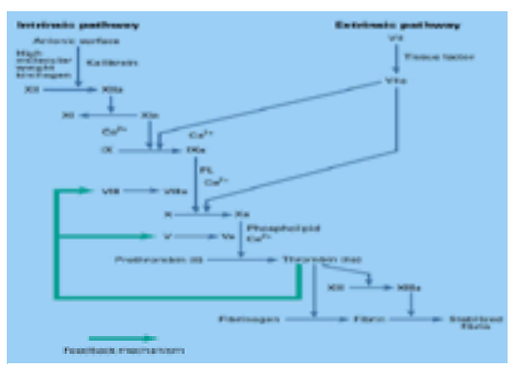


Fig. 1. Blood coagulation pathway.

The basic tests included in a clotting screen are the prothrombin time and the activated partial thromboplastin time (APTT): the former tests the extrinsic pathway, and the latter tests the intrinsic pathway (see below).

There are naturally occurring anticoagulants. The most important of these are antithrombin III, protein C, and protein S. Deficiency of these factors may result in a thrombotic tendency (see below).

Platelets are essential for normal haemostasis. They are particularly important in the formation of the primary response to vascular injury, when activated platelets coalesce to form a platelet plug in the transected vessel.

Laboratory tests of haemostasis

Coagulation tests are carried out on blood anticoagulated with sodium citrate. In contrast to other laboratory samples, it is very important that the correct volume of blood is collected in the appropriate tube. Tests must be carried out on fresh samples as coagulation factors are labile, and if there is undue delay in sending a specimen to the laboratory there will be spurious prolongation of the clotting times. Contamination with heparin (e.g. from indwelling cannulas) will also result in spurious results. It is not possible to state normal ranges which are universally applicable for most clotting tests, as individual laboratories are likely to use slightly different techniques and reagents. Individual laboratories will issue their own normal ranges.

Prothrombin time

Brain extract (rich in tissue factor) and calcium are added to the test plasma and the time taken for a clot to develop is measured. The prothrombin time is a test of the function of the extrinsic pathway and is sensitive to isolated or combined deficiencies of factors II, V, VII, X, and fibrinogen.

Prothrombin time ratio

The results of a prothrombin time are usually expressed as a ratio, comparing the result obtained on the patient test plasma with that obtained on a normal plasma sample.

The normal prothrombin time ratio should be close to 1.0. The International Normalized Ratio (INR) is for practical purposes the same as the prothrombin time ratio.

Activated partial thromboplastin time (APTT)

Phospholipid, calcium, and an activator (such as kaolin) are added to the test plasma, and the time taken for a clot to appear is measured. Kaolin activates factor XII, just like collagen. The kaolin cephalin clotting time thus tests function of the intrinsic pathway and is sensitive to isolated or combined deficiencies of factors XII, XI, X, IX, VIII, V, II, or fibrinogen.

Thrombin time

Thrombin is added to patient plasma and the time taken for a clot to develop is measured. The thrombin time is prolonged when the fibrinogen level is low, or in the presence of inhibitors of thrombin (e.g. heparin).

Platelet count

Normal range is 150 to 400 × 10^9/l. Significant bleeding may occur when the count falls below 80 × 10^9/l. Abnormal bleeding may occasionally occur if platelet function is abnormal, even when the platelet count is normal or even increased (e.g. myelodysplasia, polycythaemia).

Bleeding time

The bleeding time is a simple test of platelet function. A sphygmomanometer cuff is inflated to 40 mmHg around the upper arm. A 5-mm incision is made with a special blade. The incision is wiped with blotting paper every 30 s, and the bleeding time is taken as the time when blood stops oozing. The normal bleeding time is less than 9 min. The bleeding time will be prolonged in association with thrombocytopenia and disorders associated with defective platelet function (e.g. von Willebrand's disease, some cases of myeloproliferative disease).

Acquired disorders of coagulation

Acquired disorders of haemostasis are encountered much more frequently than congenital defects. In order to exclude the possibility of a congenital disorder of coagulation (for which specific therapy may be available) it is important to try to establish from a personal history whether there have been spontaneous haemorrhagic problems (e.g. epistaxis, haemarthrosis, gastrointestinal bleeding) in the past or bleeding after previous surgery (e.g. tonsillectomy, appendicectomy) or dental extractions. The family history should also be elicited. Symptoms of congenital disorders of haemostasis usually appear early in life. However, it should be borne in mind that mild forms of congenital disorders such as haemophilia may only become evident after surgery or major trauma. Acquired disorders of haemostasis may present in a number of ways, ranging from sudden life-threatening bleeding after surgery or childbirth at one end of the spectrum to minor purpura or an increased bruising tendency at the other.

When unexpected bleeding is encountered during surgery, the following possibilities should be considered.

Disseminated intravascular coagulation

Pathogenesis

In most circumstances, initiation of coagulation is a local phenomenon and this is an appropriate reaction to local vascular injury which has resulted in bleeding, for example at the site of a surgical incision. Disseminated intravascular coagulation is a consequence of explosive activation of the coagulation cascade throughout the vascular tree. Paradoxically, this results in a bleeding tendency and symptoms related to vascular occlusion are relatively rare. This is because the initial thrombus formed in response to the triggering of disseminated intravascular coagulation undergoes very rapid lysis. If the initial trigger persists, further cycles of coagulation and instantaneous lysis rapidly result in the depletion of coagulation factors, including fibrinogen, and consumption of platelets.

The principal causes of disseminated intravascular coagulation that are encountered in clinical practice are listed in Table 1.

Severe infections (e.g. septicaemia)
Malignant disease (metastatic carcinoma, leukaemia)
Shock
Severe burns
ABO incompatible transfusion
Obstetric disorders
Eclampsia
Abruptio placentae
Amniotic fluid embolism
Retained dead fetus
Intravascular haemolysis (e.g. infusion of hypotonic saline, systemic absorption of hypotonic fluids after prostatectomy)

Table 1 Principal causes of disseminated intravascular coagulation in clinical practice

Most cases of disseminated intravascular coagulation are triggered by septicaemia. Gram-negative organisms are often implicated. Malignant disease is also an important cause, particularly when there are multiple metastases. Carcinoma of the lung, pancreas, stomach, and prostate are particularly associated with this complication. Promyelocytic leukaemia is also frequently complicated by disseminated intravascular coagulation.

Clinical features

In overt cases, there is widespread bruising with extensive purpura. There may be persistent oozing of blood from surgical wounds and venepuncture sites. Bleeding from mucosal surfaces (e.g. epistaxis) is common. Occasionally, there are signs of vascular occlusion in the distal limbs.

Laboratory features

In fulminant cases of disseminated intravascular coagulation the blood is incoagulable. Both the APTT and prothrombin time are markedly prolonged, but will be corrected by the addition of normal plasma to the patient's plasma. The thrombin time is also prolonged, reflecting depletion of fibrinogen. The fibrinogen level is usually below 1.0 g/l (normal range 2 to 4 g/l). Levels of fibrin degradation tests in the blood will be very high, reflecting hyperfibrinolysis. The platelet count is usually reduced, and may be less than 50 × 10^9/l in severe cases. Examination of the blood film typically reveals the presence of fragmented erythrocytes.

Treatment of disseminated intravascular coagulation

It is important to identify the underlying cause (see Table 1). Treatment of the underlying condition removes the stimulus for further consumption of coagulation factors.

Fresh frozen plasma should be infused, as this is a good source of coagulation factors, including fibrinogen. As a rough guideline, three to four packs will be required initially. Cryoprecipitate is a very good source of fibrinogen and has the advantage of being more concentrated so that volume overload may be avoided. If available, cryoprecipitate should be given as well as fresh frozen plasma. Platelet concentrates should also be transfused: 10 to 12 packs should suffice as initial therapy. There is no convincing evidence that administration of antithrombin concentrates is beneficial. It is important not to overlook the fact that patients often need blood in addition to plasma products. Maintenance of circulating blood volume and an adequate haemoglobin level are important objectives, as tissue hypoxia will only exacerbate disseminated intravascular coagulation.

Contrary to what might be imagined, administration of inhibitors of fibrinolysis (e.g. tranexamic acid) is of no value in the treatment of disseminated intravascular coagulation. The use of such agents may precipitate overt thrombosis. In the past, low doses of heparin were administered in cases of disseminated intravascular coagulation, in an attempt to break the vicious cycle of initial thrombosis and subsequent lysis. Such therapy is no longer widely advocated and it is usually possible to control the activation process by judicious use of blood products and treatment of the underlying cause.

Liver disease

Haemostatic abnormalities in liver disease

The liver is the principal site of synthesis of coagulation factors. Both acute and chronic liver diseases are thus frequently associated with haemostatic abnormalities. Thrombocytopenia of moderate severity (50 to 100 × 10^9/l) is a frequent finding in patients with chronic liver disease. Another factor which contributes to the bleeding tendency of chronic liver disease is increased fibrinolytic activity, associated with a decreased plasma level of naturally occurring a_2-antiplasmin.

The most frequent haemorrhagic problems are oesophageal and gastrointestinal haemorrhage, as well as bleeding from biopsy sites and during and after surgery. Bleeding into soft tissues is only rarely encountered.

The most common laboratory findings are a marked reduction in the plasma levels of all coagulation factors except factor VIII. Both the prothrombin time and APTT are prolonged.

Therapy

Vitamin K should be administered when it is suspected that the haemorrhagic disorder is due, at least in part, to deficiency of the vitamin (see below). Fresh frozen plasma contains all of the coagulation factors and inhibitors present in blood and is suitable for the correction of the multiple abnormalities associated with liver disease. Usually two to three bags of fresh frozen plasma should suffice to correct the haemostatic defect.

Platelet concentrates are usually of little use in patients with liver disease and thrombocytopenia, as the infused platelets are rapidly sequestered in the liver and spleen. Desmopressin (DDAVP) at a dose of 0.3 µg/kg can be used to shorten the bleeding time before invasive procedures such as biopsy or laparoscopy.

Inhibitors of fibrinolysis, such as tranexamic acid, may be useful in the management of upper gastrointestinal bleeding.

Chronic renal failure

Patients with chronic renal failure often have a bleeding tendency, due in part to poor platelet function. This may be corrected by dialysis. Those with nephrotic syndrome, by contrast, may develop thrombotic complications such as renal vein thrombosis and deep venous thrombosis. The most frequent haemorrhagic manifestations observed in uraemic patients, whether on chronic haemodialysis or not, are usually from mucosal surfaces (gastrointestinal bleeding, epistaxis, menorrhagia). Retroperitoneal haemorrhage, bleeding into the pericardial and pleural spaces, and intracranial haemorrhage may develop occasionally. Patients do not usually bleed after surgical procedures, but renal biopsies are sometimes complicated by the formation of an intrarenal haematoma.

Prolongation of the bleeding time associated with a normal or only slightly reduced platelet count and normal coagulation tests are the usual findings in uraemia. There is an inverse relationship between the haematocrit and the bleeding time in renal failure. Haemodialysis or transfusion of platelet concentrates may produce transient shortening of the bleeding time. The infusion of 8 to 10 bags of cryoprecipitate in uraemic patients is usually followed by shortening or even return to normal of the bleeding time. Desmopressin at a dose of 0.3 to 0.4 µg/kg usually restores the bleeding time to normal in uraemic subjects 1 h after intravenous infusion. Transfusion of red cell concentrates in order to correct anaemia and maintain the haematocrit above 0.30 also shortens the bleeding time. Administration of erythropoietin has a similar effect.

Vitamin K deficiency

Vitamin K is necessary for the synthesis of coagulation factors II (prothrombin), VII, IX, and X. Vitamin K is fat soluble and is absorbed effectively only in the presence of bile salts. Some vitamin K is also synthesized by colonic bacteria. Little vitamin K is stored in the body and in certain conditions symptoms of deficiency may become evident within a few weeks.

Deficiency of vitamin K is associated with prolongation of both the prothrombin time and the activated partial thromboplastin time. The thrombin time and plasma fibrinogen concentration are normal, which helps in the exclusion of disseminated intravascular coagulation, and the platelet count is normal. Typical haemorrhagic manifestations are easy bruising and bleeding from sites of injury or from the gums or gastrointestinal tract.

Debilitated patients undergoing surgery are particularly vulnerable, as dietary deficiency may be compounded by the administration of broad-spectrum antibiotics which kill off gut bacteria that synthesize the vitamin. Vitamin K will not be effectively absorbed from the gastrointestinal tract when there is obstruction of the bile duct. Malabsorption of vitamin K may also occur in a number of other conditions, for example coeliac disease and intestinal fistulas. Vitamin K deficiency as a consequence of partial biliary tree obstruction may contribute to the development of impaired haemostasis in chronic hepatic disorders such as cirrhosis. An injection of vitamin K may shorten an abnormally long prothrombin time in such cases.

Anticoagulants

Obviously, the consumption of anticoagulants will result in a bleeding tendency. Where unexpected bleeding occurs during or after surgery, consideration should be given to the possibility that the patient is on anticoagulant therapy. This problem may arise when urgent surgery is carried out without full medical details of the patient being available (e.g. the patient is unconscious).

Where the possibility is suspected, further details may be sought from others involved in the medical care of the patient. Patients taking oral anticoagulants usually carry a medical advisory card. The diagnosis may be confirmed by the finding of a significantly prolonged prothrombin time. The APTT is often only slightly prolonged and the thrombin time is normal. The platelet count will also be normal.

Warfarin and its congeners are competitive inhibitors of vitamin K. Measures which have been recommended to reverse anticoagulant therapy are shown in Table 2.



Table 2 Recommendations on reversal of oral anticoagulant therapy

Bleeding after cardiopulmonary bypass may be due to the presence of heparin (see below). Contamination with heparin of blood samples drawn from venous cannulas for clotting studies is a very common cause of spurious laboratory results. This may lead to further time-consuming tests and delays in surgery before it can be

confirmed that the patient does not have a disorder of haemostasis. Blood for clotting studies should be drawn from a peripheral vein if venous cannulas are flushed with heparin. Inadvertent full heparinization of patients prior to surgery has been reported. In these cases, full-strength heparin was used to flush indwelling venous cannulas, rather than the specific dilute preparations. The use of low-dose subcutaneous heparin does not significantly alter laboratory tests of haemostasis and does not result in a generalized bleeding tendency during surgery.

Administration of streptokinase as thrombolytic therapy (e.g. for myocardial infarction) within 10 days or so after surgery or other invasive procedure such as biopsy may be hazardous as serious bleeding may ensue.

Massive blood transfusion

Blood collected in citrate phosphate dextrose with added adenine (CPDA1) has a shelf life of 35 days at 4°C. However, levels of all the coagulation factors decline during storage. Platelets in stored blood also rapidly lose their viability.

For these reasons, haemorrhagic problems may develop when a patient's blood is replaced by large quantities of stored blood within a short period of time. Microvascular bleeding is a typical manifestation of the impaired haemostasis. Examples include bleeding from mucous membranes, oozing from catheter sites which persists after application of pressure, continuous oozing from surgical wounds, and generalized petechiae. Impaired haemostasis is, of course, only one of several important problems encountered in patients receiving a massive blood transfusion. These include hypocalcaemia due to citrate overload, hyperkalaemia, and hypothermia.

When there is no underlying medical complication, replacement of up to one blood volume (8 to 10 units of blood in an adult) is not likely to be associated with significant haemostatic problems. Laboratory tests of coagulation may help to identify patients who need additional blood components to improve haemostasis when larger volumes of blood are transfused. A platelet count of less than 50×10^9/l, prothrombin time ratio of 1.8 or more, and a plasma fibrin-ogen level of 0.5 g/l or less are strongly associated with microvascular bleeding. Platelet support may be required once the patient has received 15 or more units of blood. Fresh frozen plasma is a source of coagulation factors, including fibrinogen.

Cardiopulmonary bypass

Excessive bleeding in association with cardiopulmonary bypass surgery is often a problem. Several factors contribute.

Thrombocytopenia is often present during cardiopulmonary bypass. There is also impairment of platelet function, associated with prolongation of the bleeding time. Platelet dysfunction is related to contact with the synthetic surface of the oxygenator, and probably the induced hypothermia. Many patients with coronary artery disease may be taking aspirin. Patients taking aspirin before cardiopulmonary bypass surgery are at risk of excessive blood loss during the procedure. It is therefore recommended that aspirin be discontinued at least 5 days before cardiopulmonary bypass. Despite abnormalities in platelet number and function, there is no evidence that routine perioperative transfusion of platelet concentrates is necessary.

Cardiopulmonary bypass is associated with a drop in the plasma levels of most coagulation factors, which is primarily attributable to haemodilution. As with platelet concentrates, it is not necessary routinely to transfuse fresh frozen plasma during cardiopulmonary bypass.

Heparin is routinely administered in order to prevent extracorporeal clotting in the oxygenator. Protamine sulphate is administered at the end of surgery in order to neutralize remaining heparin. In gross excess protamine itself acts as an anticoagulant. Following initial adequate heparin neutralization, the reappearance of active heparin in the bloodstream may occur 2 to 6 h later. This rebound effect is caused by the delayed return of sequestered extravascular heparin which occurs when peripheral perfusion improves. Thrombocytopenia may complicate heparin therapy in about 5 per cent of patients receiving the drug. The onset of thrombocytopenia is typically between 6 and 12 days after exposure to the drug. This is due to the development of an antibody which induces platelet activation. The possibility of heparin-induced thrombocytopenia should always be considered in the differential diagnosis when thrombocytopenia develops after cardiopulmonary bypass. Platelet concentrates should only be transfused if there are bleeding complications as arterial thrombosis has been reported after transfusion of platelets.

The administration of aprotinin, an inhibitor of plasmin, significantly reduces intraoperative and postoperative blood loss associated with cardiopulmonary bypass. The requirement for blood transfusion is reduced and the actual operating time may be shortened.

Congenital disorders of coagulation

Haemophilia

Haemophilia is the most common congenital disorder of coagulation and affects 1 in 10 000 males. Haemophilia A is caused by a deficiency of factor VIII in the circulating blood. In its severe form (less than 2 per cent of the normal factor VIII) it is characterized by recurrent joint bleeds, intramuscular bleeding, and excessive bruising after trauma. Recurrent bleeds may result in crippling arthritis. Haemophilia B (Christmas disease) is a clinically identical disorder caused by deficiency of factor IX. Both conditions show X-linked inheritance, but in one-third of cases of haemophilia there is no preceding family history as the condition results from a new mutation. Female carriers invariably have a high enough factor level to protect them from spontaneous haemorrhagic complications, although treatment with coagulation factor concentrate may be required before surgery.

Inhibitory antibodies

Some 10 per cent of patients with haemophilia A and 1 per cent of those with haemophilia B develop inhibitory IgG antibodies directed against the infused factor. Management of patients with high titre antibodies can be difficult. A satisfactory clinical response may often be obtained by increasing the dose of human factor VIII administered, although the antibody titre may be subsequently boosted. Porcine factor VIII may be effective where the antibody titre against human factor VIII is very high.

The patient's plasma must be screened for the presence of inhibitory antibodies prior to surgery. Elective, non-urgent surgery is not advisable if a patient is known to have inhibitory antibodies.

Liver disease in haemophilia

Chronic liver disease is a problem in many haemophiliacs. This is usually due to persistent infection with hepatitis B or C. All patients who are likely to receive coagulation factor concentrates should be vaccinated against hepatitis B. Documentation of abnormal hepatic function is of practical importance to the surgeon for a number of reasons.

1. In established liver disease, the synthesis of a number of coagulation factors is disturbed. Fresh plasma may be required in addition to factor concentrate to ensure haemostasis.
2. The dosage of certain drugs may need to be modified.
3. There is a potential risk to staff of infection through needlestick injuries.

Liver transplantation has been carried out in a small number of haemophiliacs having serious liver disease, and this has the added bonus of curing haemophilia as both factors VIII and IX are synthesized by hepatocytes. Of course, this radical treatment cannot be routinely applied to haemophiliacs.

HIV infection and AIDS

Approximately one-third of haemophiliacs in the United Kingdom were infected with the human immunodeficiency virus (**HIV**) between the years 1979 and 1985. The extent to which haemophiliacs in various countries were infected with HIV is very variable. Infection with HIV is associated with a progressive decline in the number of CD4+ T lymphocytes in the blood. Cellular immunity is compromised and patients may eventually develop opportunistic infections. Thrombocytopenia is seen in approximately 10 per cent of patients infected with HIV. Clearly, thrombocytopenia is particularly dangerous in haemophiliacs who already have a bleeding tendency. Zidovudine therapy usually raises the platelet count considerably.

There is no evidence that surgery accelerates progression to AIDS by promoting the decline in the number of CD4+ T lymphocytes. Many surgeons are concerned about the possibility of transmission of HIV (see also Chapter 4.2). However, the risk of infection to health care workers through body fluids or inoculation is much less

than that associated with hepatitis B. A prospective study of needlestick injuries among health care staff and involving patients known to be infected with HIV has demonstrated a risk of infection of around 0.3 per cent. The vast majority of such injuries are associated with the resheathing of needles: this should be avoided, and all used needles should be discarded in a suitable container immediately after use. Needlestick injuries among surgeons due to suture needles appear to be less dangerous than injuries associated with hollow needles used for venepuncture. The prophylactic administration of zidovudine after a needlestick injury does not guarantee protection against infection and the drug has a number of toxic effects. The risk of transmission through blood or other bodily fluids coming into contact with intact skin is negligible.

Practical management of a patient with haemophilia scheduled for surgery

Wherever possible, elective surgery should be carried out in a centre experienced in the management of haemophilia. All registered patients should carry a medical card which states the precise diagnosis, factor level, inhibitor status, and blood group. The HIV status of a patient will not be stated on the card. It must be remembered that female carriers of haemophilia A or B may themselves have low levels of factor VIII or IX, respectively, and may require treatment prior to surgery or dental work.

Patients scheduled for elective surgery should be admitted 1 or 2 days beforehand for blood tests (Table 3). Patients with severe haemophilia are likely to need infusions of concentrate for at least 7 to 10 days after any type of surgery. They will often need, therefore, to stay in hospital for a longer period than other patients and this must be planned for accordingly. It is all too common to see patients scheduled for discharge 2 days after ‘minor' surgery. There is no such thing as ‘minor' surgery for haemophiliacs.

1. Check supplies of concentrate are adequate for treatment period
2. Is any other surgical work scheduled?
3. Blood tests for:
   Baseline factor level
   Inhibitory antibody screen
   Hepatitis HBsAg status
   Antihepatitis B antibody (anti-HBs) status
   Antihepatitis C antibody (anti-HCV) status
   Anti-HIV status
   Liver function tests (including prothrombin time)
   Full blood count
   Group and save plasma or cross-match as appropriate

Table 3 Preoperative check list for patients with haemophilia

All patients receiving coagulation factor concentrates (even high-purity ones) should be vaccinated against hepatitis B. If some time has elapsed since completion of the course of vaccination, immune status should be checked and the antibody level (anti-HBs) checked. Surgeons dealing with haemophiliacs should certainly be vaccinated against hepatitis B. No vaccine is as yet available against hepatitis C.

In some countries where concentrate is not readily available, some units have achieved considerable savings by good co-ordination of activities. If a patient is scheduled to have surgery and will be in hospital and having concentrate for some days thereafter, other procedures can be carried out at the same time (e.g. dental work, excision of skin lesions, vasectomy).

On the day of surgery, concentrate should be infused within 1 h before surgery: a dose of 50 IU/kg of factor VIII or 75 IU/kg of factor IX should suffice to boost the plasma level to well within the normal range. Pre- and postinfusion plasma levels of factor should be checked in the laboratory to confirm that a suitable level has been achieved for surgery. After surgery, twice-daily bolus infusions of factor VIII will be required for at least 5 days. Further daily doses may be required for a further 5 days or so, until soft tissue healing is achieved. Factor IX has a longer half-life, and only one infusion a day is necessary. Analgesics must not be given by intramuscular injection and aspirin is contraindicated.

To minimize the risk of bleeding from wounds, it is sensible to change dressings after infusion of a dose of concentrate. Similarly, physiotherapy is best scheduled immediately after infusion. Intramuscular injections (e.g. premedication prior to surgery, postoperative analgesia) must not be given, as a large intramuscular haematoma is likely to develop.

Deep venous thrombosis is not a problem in patients with haemophilia A and prophylaxis (e.g. heparin, low-dose warfarin) is not necessary, even in orthopaedic surgery. Such treatment may also interfere with laboratory monitoring of plasma factor levels after infusion of concentrate. Thrombosis has been reported in patients receiving prothrombin-complex concentrates, and high-purity factor IX concentrates are to be recommended for patients with haemophilia B undergoing surgery.

Other congenital disorders of coagulation

Very occasionally, patients with isolated congenital deficiencies of other coagulation factors may be encountered. Deficiencies of fibrinogen, factor V, X, VII, XI, or XIII may be associated with a serious bleeding tendency. Similar practical guidelines to those set out above apply. Specific plasma-derived concentrates of most of these coagulation factors are available commercially.

Investigation of the patient with a bleeding tendency

It is not necessary to perform tests of haemostasis on all patients scheduled for elective surgery. However, if a history of a possible bleeding tendency emerges during outpatient investigation, patients should certainly be referred for investigation prior to surgery.

Basic tests of coagulation include the prothrombin time, kaolin cephalin clotting time, and the platelet count. Where the platelet count is normal (or even elevated) a bleeding time should be measured to exclude the possibility of congenital or acquired defects of platelet function. It is important to try to distinguish between congenital disorders on the basis of the personal and family history. Table 4 is intended only as a guideline to how the screening test may help in the diagnosis of specific defects.

Disorder	Prothrombin time	APTT	Thrombin time	Platelets	Bleeding time	Notes
Heparin	N	↑↑	↑↑	N	N	Reptilase time normal
Diffuse	↑↑	↑↑	↑↑	↓	↑	Fibrin degradation tests relevant Intravascular coagulation Erythrocyte fragments on blood film
Liver disease	↑	↑	N or ↑	N or ↓	N or ↑	Transaminases abnormal
Functional platelet defect	N	N	N	N	↑↑↑	
Vitamin K deficiency	↑↑	↑	N	N	N	Ecbis carinatus time normal Similar findings with warfarin
Haemophilia	N	↑↑	N	N	N	Confirm with specific factor assay
von Willebrand's disease	N	↑↑	N	N	↑↑↑	Confirm with specific factor assay

N, normal
↑, increased.
↓, decreased.

Table 4 Laboratory investigation of a patient with a bleeding tendency

Where bleeding is encountered during surgery and there is no obvious source of haemorrhage, a prothrombin time, kaolin cephalin clotting time, and platelet count should be requested in the first instance. The administration of two to three units of fresh frozen plasma is a useful practical measure until the cause is identified. Platelet concentrates should be given if the platelet count is below 50×10^9/l.

Thrombophilia

The term thrombophilia is used to describe familial or acquired disorders of the haemostatic mechanism which predispose to thrombosis. Inherited deficiencies of antithrombin III, protein C, and protein S are important causes of a thrombotic tendency. Antithrombin III inactivates factors IX, X, XI, and thrombin. Protein C and its cofactor protein S are vitamin K-dependent proteins which inactivate factors V and VIII. Such patients are particularly vulnerable to the development of deep venous thrombosis and/or pulmonary embolism after surgery if adequate precautions are not taken.

At present, the great majority of thrombotic episodes remain unexplained in that no underlying disorder may be identified. Occasionally, acquired blood disorders associated with a definite thrombotic tendency may be diagnosed after an episode of thrombosis. These include thrombocytosis secondary to myeloproliferative disease, polycythaemia, and the lupus anticoagulant. The development of a lupus anticoagulant is associated with prolongation of the APTT and, in about a quarter of cases, prolongation of the prothrombin time. The former is not corrected by the addition of normal plasma to the patient's own plasma. There may be thrombocytopenia. Despite these findings, the patient is paradoxically at risk of thrombosis and there is no undue risk of haemorrhage (even in association with surgery). Although originally described in association with systemic lupus erythematosus, it is important to appreciate that the great majority of cases arise in people who have no evidence of the disorder and who are perfectly well.

Consideration should be given to investigating patients encountered in clinical practice who have thrombosis at an unusually early age, or in an unusual site (e.g. mesenteric vein thrombosis), or where there is a suggestion of a familial tendency to thrombosis. Laboratory testing for thrombophilia is difficult once a patient is anticoagulated, as heparin reduces the plasma level of antithrombin whilst warfarin boosts the plasma level of antithrombin III but lowers the level of proteins C and S. Investigation is best deferred until the patient has been off anticoagulants for at least 2 months.

Management of individuals with thrombophilia

By no means all patients with documented thrombophilia will experience spontaneous thrombosis during their lifetime. Studies suggest that approximately half of such patients will experience spontaneous thrombosis. Where deficiency of antithrombin, protein C, protein S, or lupus anticoagulant are identified in patients without a history of thrombosis, it is not usual practice to initiate long-term prophylactic warfarin therapy as this treatment itself is not without risks. However, such patients should certainly receive some form of prophylaxis to cover surgery and pregnancy. Standard low-dose subcutaneous heparin therapy is perfectly satisfactory for general surgery. Higher doses of heparin or low-intensity anticoagulation with warfarin are advisable to cover major orthopaedic procedures. Affected women should not take oestrogen-containing oral contraceptives or hormone replacement therapy after the menopause. Plasma-derived concentrates of antithrombin III are available and may be given in addition to subcutaneous heparin to cover surgical procedures. Concentrates of protein C and S are not yet available.

An episode of thrombosis in a person with documented thrombophilia should be treated with heparin and warfarin in the standard fashion. Patients with antithrombin deficiency may require more heparin than usual in order to achieve an adequately prolonged kaolin cephalin clotting time. Consideration should be given to long-term oral anticoagulation following an episode of thrombosis, if there are no contraindications.

Haemoglobinopathies

Sickle cell disease

Sickle haemoglobin is found throughout tropical Africa, scattered in countries bordering the Mediterranean, in the Middle East, and India. With the ease of world travel every surgeon may encounter a patient with sickle cell disease who requires surgery.

Operating on a patient with sickle cell disease significantly increases the risk to the patient of sickle-related complications. These may vary in severity from a painful limb crisis to life threatening neurological or respiratory crises.

Dehydration, anoxia, hypothermia, and hypotension must be avoided. Preoperative blood transfusion has been considered essential in order to minimize the risk of postoperative sickle-related complications although data supporting this practice from randomized trials does not exist. However, a number of recent studies of blood transfusion practice in patients with sickle cell disease have been published. The co-operative study of sickle cell disease analysed retrospectively 1079 operations on 717 patients with sickle cell disease (67.5 per cent had Hb SS, 22.4 percent Hb SC, and 10.1 per cent had Sb thalassaemia). It was found that postoperative complications unrelated to sickle cell disease occurred with the same frequency as in patients without sickle cell disease. The overall mortality was 1.1 per cent.

The six commonest procedures performed were splenectomy and cholecystectomy, dilatation and curettage, caesarian section and hysterectomy, tonsillectomy and adenoidectomy, hip replacement, and myringotomy. The sickle-related complications (painful crisis, acute chest syndrome, and cerebrovascular accident) for these procedures varied considerably; In patients undergoing the gynaecological and obstetric procedures the sickle-related complication rate was 11.1 to 18.6 percent and in the ear, nose, and throat and orthopaedic operations the rate was 0.0 to 14.3 per cent. Cholecystectomy is one of the commonest operations performed in patients with sickle cell disease. An analysis of 364 cholecystectomies in patients with sickle cell disease was recently reported. The total complication rate was 39 per cent with sickle-related events occurring in 19 per cent. A small group of patients in this report had not received preoperative blood transfusion and suffered a higher incidence of sickle-related events (32 per cent).

Another study directly compared an aggressive to a conservative transfusion approach in patients with sickle cell disease undergoing surgery. The target haemoglobin was 10 g/dl in both groups but the aim was to reduce haemoglobin S to less than 30 per cent in the aggressively transfused group. Exchange transfusion was commonly used to obtain this effect. No target reduction in haemoglobin S was set for the conservatively transfused group. The conclusion from this randomized study was that the conservative approach was as effective as the aggressive approach in preventing perioperative complications and the conservative approach resulted in fewer transfusion-related complications.

The effect of the type of anaesthesia on postoperative complications has also been studied and the data suggest that general or local anaesthesia are safer than regional anaesthesia.

It is difficult to draw precise conclusions from the published studies. A reasonable approach is to transfuse adult patients undergoing anything other than minor surgery using a conservative approach as outlined above. Children being prepared for ear, nose, and throat surgery probably do not require transfusing. Other situations need to be considered on their merits. Because it has become standard practise in the Western world to transfuse patients with sickle cell disease preoperatively this will remain the commonest option but a randomized study comparing no transfusion with a conservative transfusion approach is urgently needed.

Finally, patients with sickle cell disease may have minor cardiac and renal impairment but this will be rarely severe enough, at least in younger patients, to occasion any alarm peroperatively.

Sickle cell trait

The risk for patients with sickle cell trait undergoing surgery is not much greater than for normal individuals. However, simple precautions, for example avoiding dehydration or anoxia, should be followed. Preoperative blood transfusion is not needed.

Thalassaemia

Patients with thalassaemia trait can undergo surgery without any special precautions. In contrast, patients with b-thalassaemia major or anaemic patients with thalassaemia intermedia (a variety of thalassaemic disorders with a severity somewhere between thalassaemia major and thalassaemia trait) should be transfused up to a normal haemoglobin preoperatively.

The optimal routine treatment for patients with b-thalassaemia is regular blood transfusion to maintain a normal haemoglobin. Those patients who have not simultaneously undergone a rigorous iron chelating programme may become iron overloaded and the clinical consequences of this, especially the risk of impaired cardiac function, will be of importance to the surgeon and anaesthetist in a patient needing surgery. There have recently been some encouraging reports of improvement in cardiac function in iron-overloaded patients after aggressive high-dose therapy with desferrioxamine. Therapeutic benefit may take many months to achieve but is worthwhile if time allows.

Splenectomy is a common operation in thalassaemic patients. The increased risk of infection after splenectomy is well described. Briefly patients should receive Pneumovax preoperatively and then penicillin by mouth for life postoperatively. A rare thrombotic problem has been identified in some patients after splenectomy.

Persistent thrombocytosis after splenectomy occurs predictably in patients where anaemia persists. The thrombocytosis may be associated with life-threatening thromboembolic complications and this phenomenon has been well documented in patients with HbH disease. HbH disease occurs where there is a failure of three of the four a-globin genes to form a-globin chains with the clinical phenotype, that of thalassaemia intermedia. A number of reports cite persistent thrombocytosis in these patients after splenectomy and the subsequent development of thrombophlebitis and thromboembolism.

In patients with a persistent thrombocytosis (>500 × 109/l) after splenectomy it is wise to use lifelong aspirin postoperatively and treat with warfarin those patients who develop signs of superficial thrombophlebitis or other thrombotic events despite the aspirin.

Myeloproliferative disorders

Among these disorders are included chronic myeloid leukaemia (CML), primary proliferative polycythaemia (formerly known as polycythaemia rubra vera (PRV)), essential thrombocythemia (ET), and myelofibrosis. In each of these disorders the platelet count may be significantly raised (often greater than 1000 × 109/l) and in the perioperative period this may greatly enhance a thrombotic tendency. In addition, patients with polycythaemia will have a high haemoglobin associated with an increase in red cell mass which will further enhance the risk of thrombosis. The situation may be further complicated by an increased haemorrhagic tendency due to abnormal platelet function in each of the above disorders giving rise to the situation where a single patient is simultaneously at increased risk of thrombosis and bleeding.

Wherever possible surgery in such patients should be planned and appropriate treatment given to the patient to render the blood counts normal or near normal preoperatively. Once this is achieved it is good practice to give antithrombotic prophylaxis peroperatively. Preoperative coagulation tests should include, in addition to the platelet count, a bleeding tine, prothrombin time, and activated partial thromboplastin time (see later). Patients with a prolonged bleeding time, with the implication of impaired platelet function, should have platelets available to be transfused in the operative and postoperative period as necessary.

Patients with uncontrolled myeloproliferative disorders who are admitted with acute surgical problems requiring emergency surgery are at significant risk of thrombotic complications. Patients with polycythaemia may be venesected down to a normal haemoglobin relatively quickly (isovolumetric replacement of red cells by colloid is recommended) but a high platelet count cannot be quickly corrected. Peroperative anticoagulation is essential to reduce the risk of thrombosis. Bleeding peroperatively can be managed as above with platelet transfusions and fresh frozen plasma if there is prolongation of the prothrombin or activated partial thromboplasin time.

Perhaps the single most commonly performed operation in patients with myelofibrosis is splenectomy. The indications for this operation include local symptoms from the enlarged spleen, portal hypertension, refractory haemolysis, or thrombocytopenia. A number of reports of splenectomy in the management of myelofibrosis have been published. Mortality rates vary from 4 to 28 per cent and complication rates from 47 to 56 per cent. The major reported causes of morbidity were haemorrhage despite apparently appropriate precautions, followed by infection and thrombosis. A preoperative coagulation screen indicating disseminated intravascular coagulation (DIC) should be a contraindication to splenectomy in this patient group. DIC is probably caused by the effects of splenic infarction and will usually settle spontaneously within a few weeks. Bleeding after splenectomy may also be caused by defective platelet function and a prolonged prothrombin time caused by liver dysfunction, an isolated factor V deficiency, or the presence of circulating anticoagulants. Operating in the presence of DIC or other coagulation disturbance will increase the risk of major bleeding. Infections in one series occurred in 21 of 96 patients after splenectomy with the respiratory system, wound, and subphrenic space being the affected sites. Other postsplenectomy complications in patients with myelofibrosis include accelerated hepatomegaly, an increased incidence of leukaemic transformation, extreme thrombocytosis, and large vein thrombosis. The high mortality and morbidity rates for splenectomy in patients with myeloproliferative disorders compare unfavourably with the relatively trouble free course seen in patients undergoing splenectomy for autoimmune disorders such as idiopathic thrombocytopenic purpura and autoimmune haemolytic anaemia (see later).

Lymphoproliferative disorders

Fewer patients with lymphoproliferative disorders now require splenectomy as part of their management than was the case a number of years ago. Staging laparotomies and splenectomies in patients with Hodgkin's disease are performed rarely now compared with 10 to 20 years ago and splenomegaly with hypersplenism or local symptoms from an enlarged spleen due to tumour infiltration in patients with lymphoma can sometimes be treated successfully with chemotherapy.

Nevertheless there remains a steady trickle of patients, often elderly and with massively enlarged spleens, who require surgery either for local symptoms induced by the splenomegaly, hypersplenism, or both. Thrombocytopenia is the commonest haemostatic abnormality and platelet transfusions may be needed peroperatively. Infection and bleeding are the commonest complications but the postoperative risks are substantially less in patients with lymphoproliferative disorders compared to myeloproliferative disorders.

Myelodysplasia

Myelodysplasia is a clonal disorder of bone marrow resulting in the production of morphologically and functionally abnormal red cells, white cells, and platelets. Infection and bleeding are common complications. The median survival is between 2 and 3 years from diagnosis and approximately one-third of patients will develop acute myeloid leukaemia in the terminal stages of their illness. In the early stages a typical patient may be noted to be mildly anaemic (often with a raised mean cell volume), leucopenic, and thrombocytopenic, or any combination of these abnormalities.

From a surgical point of view the major risks are of haemorrhage and infection. Even if the platelet count is normal platelet function may be abnormal. As with the myeloproliferative disorders platelets for transfusion should be available peroperatively and close attention given to prompt treatment of postoperative infection.

Autoimmune disorders

Idiopathic thrombocytopenic purpura (ITP)

Splenectomy is a standard form of treatment in patients with idiopathic thrombocytopenic purpura who have failed to achieve a sustained response to primary drug therapy. In childhood idiopathic thrombocytopenic purpura up to 80 per cent of patients will remit spontaneously or after treatment with prednisolone and/or intravenous immunoglobulin (IV IgG). In contrast, first-line therapy results in sustained complete remission in only around 25 per cent of adults with this disorder.

Up to 90 per cent of patients with idiopathic thrombocytopenic purpura who have failed to respond to prednisolone, or who relapse after the withdrawal of prednisolone, will obtain a rapid rise in platelet count after treatment with IV IgG; unfortunately the response is usually short-lived. However, maintenance therapy with IV IgG may achieve long term remission in 40 to 50 per cent of patients.

The expense of maintenance IV IgG means that splenectomy is the most common treatment option in patients who have failed to achieve a sustained response with prednisolone. Improvement in the platelet count preoperatively, if considered clinically necessary, can be achieved with prednisolone in those patients who have previously shown a response to this drug, or with IV IgG which improves the platelet count in up to 90 per cent of patients with idiopathic thrombocytopenic purpura.

The risk of adrenal suppression in patients treated with prednisolone for weeks or months preoperatively would necessitate increasing doses of steroid peroperatively. Also patients treated with prednisolone may have an increased risk of perioperative infection. Intravenous IgG does not carry these risks. In preparing a patient for surgery IV IgG is the treatment of choice where the means and facilities allow it to be given. A typical dose of IV IgG is 0.4g/kg body weight/day for 5 days. Platelet transfusion peroperatively is rarely appropriate.

Splenectomy will return the platelet count to normal, or substantially improve it, in around 70 to 80 per cent of patients. Only a small number (approximately 10 per cent) will be left with continuing severe thrombocytopenia and haemorrhagic problems.

Autoimmune haemolytic anaemia (AIHA)

Splenectomy is the treatment of choice for the patients with autoimmune haemolytic anaemia who fail to respond to treatment with prednisolone. Splenectomy is effective in more than 50 per cent of cases. In contrast to patients with idiopathic thrombocytopenic purpura IV IgG has little or no place in the management of patients with autoimmune haemolyic anaemia. There are no specific precautions to be taken preoperatively. Severe anaemia can be corrected by blood transfusion although the half-life of the transfused red cells will be markedly shortened by immune destruction.

Postoperatively a transient increase in platelet count (often to >1000 × 109/l) is common and peroperative antithrombotic prophylaxis should be considered.

Further reading

Bickerstaff KI, Morris PJ. Splenectomy for massive splenomegaly. *British Journal of Surgery* 1987; **74**: 346–49.

Blood Transfusion Task Force, British Committee for Standardization in Haematology. Guidelines for transfusion for massive blood loss. *Clinical and Laboratory Haematology* 1988; **10**: 265–73.

Bloom AL, Forbes D, Thomas DP, Tuddenham EGD, eds. *Thrombosis and haemostasis*, 3rd edn. Churchill Livingstone, Edinburgh, 1993.

British Committee for Standards in Haematology. Guidelines for the use of fresh frozen plasma. *Transfusion Medicine* 1992; **2**: 57–63.

Cohen AR, Mizanin J, Schwartz E. Rapid removal of excessive iron with daily dose intravenous chelation therapy. *Journal of Pediatrics* 1989; **115**: 151–5.

Dacie JV. *The haemolytic anaemias*, Vol. 2. Churchill Livingstone–Longman Group UK Ltd, Edinburgh, 1988.

Danforth DN, Fraker DL. Splenectomy for the massively enlarged spleen. *American Surgeon* 1991; **52**: 108–13.

Duthie RB, Rizza CRC, Giangrande PLF, Dodd C. *Management of musculoskeletal problems in haemophilia*. University Press, Oxford, 1993.

Gill PG, Souter RG, Morris PJ. Splenectomy for hypersplenism in malignant lymphomas. *British Journal of Surgery* 1981; **1**: 29–33.

Haemostasis and Thrombosis Task Force for the British Committee for Standards in Haematology. Guidelines for oral anticoagulation therapy (3rd edn). *British Journal of Haematology* 1998; **101**: 374–87.

Haberkern CM, Neumayr LD, Orringer EP *et al.* Cholecystectomy in sickle cell anaemia patients: perioperative outcome of 364 cases from the national preoperative transfusion study. *Blood* 1997; **89**: 1533–42.

Hirsch J, Dacie JV. Persistent post-splenectomy thrombocytosis and thrombo-embolism. A consequence of continuing anaemia. *British Journal of Haematology* 1966; **12**: 44–53.

Jacobs P, Wood L, Dent DM. Splenectomy in the chronic myeloproliferative syndromes. *South African Medical Journal* 1992; **81**: 499–503.

Koshy M, Weiner SJ, Miller ST *et al.* Surgery and anaesthesia in sickle cell disease. *Blood* 1995; **86**: 3676–84

Mitchell A, Morris PJ. Surgery of the spleen. *Clinics in Haematology* 1983; **12**: 565–90.

Morris PJ, Cooper IA, Madigan JP. Splenectomy for hypersplenism in malignant lymphoma. *Lancet* 1975; **ii**: 250–2.

Poller L, Thomson JM, eds. *Thrombosis and its management*. Churchill Livingstone, Edinburgh, 1993.

Reilly JT. Idiopathic myelofibrosis: pathogenesis, natural history and management. *Blood Reviews* 1997; **11**: 233–42.

Rizza CRC, Lowe GDO, eds. *Haemophilia and its management*. Ballière Tindall, London, 1994.

Tefferi A, Silverstein MN. Current perspective in agnogenic myeloid metaplasia. *Leukemia and Lymphoma* 1996; **22** suppl. 1: 169–71.

Vichinsky EP, Haberkern CM, Neumayr L *et al.* A comparison of conservative and aggressive transfusion regimens in the perioperative management of sickle cell disease. *New England Journal of Medicine*, 1995; **333**: 206–13.

Woodman RC, Harker LA. Bleeding complications associated with cardiopulmonary bypass. *Blood* 1990; **76**: 1680–97.

5.2 Blood transfusion in surgical practice

M. F. Murphy

General principles
Indications for transfusion
Blood groups and red cell antibodies
Procedure for blood transfusion
Pretransfusion compatibility testing
Blood ordering
Elective surgery
Emergencies
Blood, blood components, and blood products
Whole blood
Blood components
Blood products
Strategies for avoiding or reducing the use of blood tranfusion
Autologous transfusion
Complications of blood transfusion
Immediate haemolytic transfusion reactions
Delayed haemolytic transfusion reactions
Non-haemolytic (febrile) transfusion reactions
Urticaria and anaphylaxis
Transmission of infection
Immunosuppression
Further reading

General principles

Blood loss due to surgery or trauma does not necessarily require replacement by blood transfusion. Transfusion is warranted where blood loss is considerable, especially when it is rapid, when further blood loss or surgery is anticipated, or if the patient is suffering from symptomatic anaemia.

Blood transfusion is associated with a number of complications. Each potential recipient requires individual assessment to determine whether the benefits of transfusion outweigh the risks. Information about blood transfusion should be provided to patients as part of the explanation of their overall management. In an emergency, blood transfusion can be given without the patient's agreement unless there is a known pre-existing objection to it.

Indications for transfusion

Although guidelines for red cell transfusion exist, there are few randomized clinical trials in this field. There is evidence of wide variation in the use of blood transfusion for the same surgical procedures, raising concern about patients' safety and the economic use of blood. Studies are needed to investigate outcomes for patients offered different transfusion protocols.

Measurements of the haemoglobin or haematocrit and the volume of blood loss are the only generally accepted and available indices to influence the decision to transfuse and a recent study showed that a transfusion threshold of 70 g/l was as safe and possibly superior to one of 100 g/l in critical care patients. A minimum preoperative haemoglobin of 100 g/l is no longer regarded as essential, as many patients with a lower haemoglobin tolerate surgery and seem to recover just as well. Data from Jehovah's Witnessses, who refuse transfusion, indicate that morbidity and mortality rises when the haemoglobin falls below 70 g/l, but there is no agreement about when transfusions should be given or what optimum concentration of haemoglobin should be achieved.

Replacement of oxygen-carrying capacity and volume is vital when blood loss is greater than 30 per cent of estimated blood volume, or less in elderly patients and those with cardiac or respiratory disease. Blood is now almost exclusively provided as red cell concentrates; there is no evidence that fresh whole blood has ‘magic' haemostatic properties. There has been controversy for many years over the use of colloids or crystalloids for the fluid resuscitation of hypovolaemic patients. A recent systematic review of randomized, controlled trials of resuscitation with colloids compared to crystalloids for the volume replacement of critically ill patients, including those with trauma, burns and undergoing surgery, concluded that the use of colloids was associated with an increase in fatality.

Patients with massive blood loss, defined as those requiring transfusion of a volume of blood greater than their blood volume within 24 h, present difficulties in addition to those of red cell and volume replacement. Stored blood has no functioning platelets, and dilutional thrombocytopenia occurs after transfusion of 1.5 times the patient's blood volume. In order to maintain haemostasis, the platelet count should be maintained above 50×10^9/l. Depletion of coagulation factors is unusual, because stored blood contains adequate amounts of all except for factors V and VIII, which fall during storage. There is no evidence that that prophylactic replacement regimens of fresh frozen plasma, based on formulas of so much fresh frozen plasma per units of blood transfused, prevent the onset of abnormal bleeding or reduce transfusion requirements. Disordered haemostasis is more likely to be due to coexisting disseminated intravascular coagulation. Abnormal prothrombin and partial thromboplastin times (e.g. greater than 1.5 times normal) should in theory indicate the need for the use of fresh frozen plasma to replace coagulation factors, but in practice these tests correlate poorly with bleeding. Continued bleeding together with severely abnormal coagulation tests or laboratory evidence of disseminated intravascular coagulation requires more intensive replacement of coagulation factors with large doses of fresh frozen plasma, and possibly cryoprecipitate.

Blood groups and red cell antibodies

The cells and proteins in the blood express antigens that are controlled by polymorphic genes. The blood groups are determined by antigens on the surface of red cells. More than 400 blood groups have been found.

The ABO system (Table 1) is the most important blood group system because anti-A and anti-B antibodies are capable of producing rapid and severe intravascular haemolysis of incompatible red cells. ABO antibodies are naturally occurring, made in the first few months of life in response to environmental antigens present in food and bacteria. They are usually IgM antibodies capable of fixing complement and causing intravascular haemolysis.

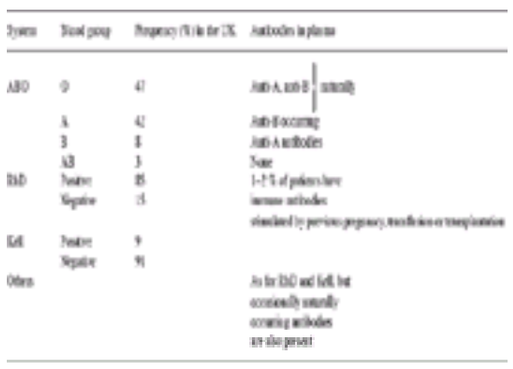


Table 1 Distribution of the most important blood groups

Blood transfusion, pregnancy or transplantation may immunize recipients against other red cell antigens that they lack (alloimmunization). These antibodies, such as Rhesus, are termed immune antibodies. They are usually IgG and do not generally fix complement. They cause removal of antibody-coated cells through their attachment to Fc receptors in the spleen and elsewhere in the reticuloendothelial system (extravascular haemolysis). Incompatibilities involving many blood groups

(including Rhesus, Kell, Duffy, Kidd) may cause haemolytic transfusion reactions and haemolytic disease of the newborn.

Procedure for blood transfusion

The safety of blood transfusion depends on meticulous attention to detail at each stage leading up to and during the transfusion. Avoidance of simple errors involving patient and blood sample identification at the time of collection of the sample for cross-matching and at the time of transfusion would avoid most serious haemolytic transfusion reactions, almost all of which involve the transfusion of ABO-incompatible blood.

Pretransfusion compatibility testing

1. **Blood grouping.** The ABO and RhD groups of the patient are determined.
2. **Antibody screening**. The patient's plasma is tested against red cells from at least two group O donors, expressing a wide range of red cell antigens, using a direct agglutination test (to detect IgM antibodies) and an indirect antiglobulin test (to detect IgG antibodies). About 10 per cent of patients have a positive result, in which case further testing is done with a comprehensive panel of typed red cells to determine if clinically significant antibodies are present (which they are in about 20 per cent of instances of a positive result in the screening test, and in 2 per cent of all patients). These further investigations may take some time and there may be a delay in the provision of compatible blood.
3. Donor blood of the same ABO and RhD group as the patient is selected.
4. **Cross-matching.**
   - *Patients without atypical red cell antibodies*. The full cross-match involves testing the patient's plasma against a sample of the red cells from the donor unit in a direct agglutination test and also using an indirect antiglobulin test. In some hospitals, this has been shortened to an immediate spin cross-match where the patient's serum is briefly incubated with the donor red cells, followed by centrifugation and examination for agglutination. Where blood-bank computer systems can confirm electronically that the patient has the same ABO and RhD groups as in a historical record and the patient has never had *red cell* antibodies, the cross-match may be safely omitted completely (electronic or computer cross-matching).
   - *Patients with atypical red cell antibodies*. Donor blood lacking the relevant red cell antigen(s), as well as having the same ABO and RhD groups as the patient, should be selected. A full cross-match should always be made in case there are additional antibodies to rare blood-group antigens present on the selected donor cells.

Blood ordering

Elective surgery

Sufficient time should be allowed for the laboratory to carry out pretransfusion testing, as provision of blood for the 2 per cent of patients with atypical antibodies may be difficult. Most hospitals have guidelines for blood ordering for elective surgery (maximum surgical blood-ordering schedules), aiming to reduce unnecessary cross-matching and the amount of blood that eventually becomes outdated. Many operations in which blood is only occasionally required for unexpectedly high blood loss can be classified as ‘group and save serum'; this means that, where the antibody screen is negative, blood is not reserved in advance but can be made available quickly if necessary. If a patient has atypical antibodies, compatible blood should always be reserved in advance; finding compatible blood may take several days if unusual or complex mixtures of antibodies are present. Surgical blood-ordering schedules should be regularly reviewed by audit of current blood usage.

Emergencies

There may be insufficient time for full pretransfusion testing. The options include:

1. **Blood required immediately**—use of 2 units of O RhD-negative blood (‘emergency stock'), to allow additional time for the laboratory to group the patient.
2. **Blood required in 10–15 min**—use of blood of the same ABO and RhD groups as the patient.
3. **Blood required in 45 min**—most laboratories will be able to provide fully cross-matched blood within this time.

Blood, blood components, and blood products

Blood collected from donors is processed into:

1. Blood components, such as red cell and platelet concentrates, fresh frozen plasma and cryoprecipitate, which are prepared from a single donation of blood by simple separation methods such as centrifugation, and transfused without further processing.
2. Blood products, such as coagulation factor concentrates and albumin and immunoglobulin solutions, which are prepared by complex processes using the plasma from many donors as the starting material.

It is good transfusion practice to transfuse only the blood component or product required by the patient rather than using whole blood (component therapy). This is the most effective way of using donor blood, which is a scarce resource, and reduces the risk of complications from the transfusion of portions of the blood that are not needed (Table 2).



Table 2 Complications of transfusion

Whole blood

The average volume of blood withdrawn is 470 ml (recently increased to this from 450 ml), taken into 63 ml of anticoagulant. Blood stored at 4°C has a ‘shelf-life' of 5 weeks, during which at least 70 per cent of the transfused red cells should survive normally. Whole blood is rarely used in modern transfusion practice; red cell concentrates plus crystalloids are an acceptable alternative for acute blood loss.

Blood components

Red cell concentrates

Virtually all the plasma is removed and it is replaced by about 100 ml of an optimal additive solution, such as SAG-M, which contains sodium chloride, adenine, glucose, and mannitol for optimal storage of the red cells. The mean volume is about 330 ml and the mean packed cell volume is about 0.57, but the viscosity is low as there are no plasma proteins in the additive solution allowing fast administration if necessary.

Buffy coat-depleted red cell concentrates

These are prepared by removal of the buffy coat, which contains most of the leucocytes and platelets. They are useful in preventing febrile reactions in patients with a

previous history of reactions and in patients likely to receive multiple transfusions, for example those with haematological diseases.

Cytomegalovirus-seronegative red cell and platelet concentrates

These are used to minimize the risk of transmission of cytomegalovirus in susceptible recipients such as cytomegalovirus-seronegative bone marrow and renal transplant recipients, and for fetal and neonatal transfusions.

Leucocyte-depleted red cell concentrates

These are usually prepared by filtration. They are used in patients to prevent HLA alloimmunization in some potential bone-marrow transplant recipients who are at risk of marrow rejection due to HLA alloimmunization, for example those with severe aplastic anaemia. They are also acceptable as a substitute for cytomegalovirus-seronegative blood components, where these are unavailable, as cytomegalovirus is leucocyte-associated.

Washed red cell concentrates

These are preparations of red cells suspended in saline, produced by cell separators to remove all but traces of plasma proteins. They are used in patients who have had severe recurrent urticarial or anaphylactic reactions.

Platelet concentrates

Platelet concentrates are now presented as ‘adult therapeutic doses' in a single pack prepared either by processing pooled buffy coats from four whole blood donations or by plateletpheresis of single donors using cell separators. They may be stored for up to 5 days at 22°C. They are used to treat, or to prevent prophylactically, bleeding in patients with severe thrombocytopenia.

Granulocyte concentrates

These are prepared from single donors by using cell separators. They are used for patients with severe neutropenia with definite evidence of bacterial infection where antibiotic therapy has failed. They are now only rarely used.

Fresh frozen plasma

This is prepared by freezing the plasma from 1 unit of blood at -30°C within 6 h of donation. The volume is approximately 200 ml. Fresh frozen plasma contains all the coagulation factors present in fresh plasma and is mostly used for replacement of those factors in patients with bleeding due to acquired deficiencies such as disseminated intravascular coagulation.

Cryoprecipitate

This is obtained by allowing the frozen plasma from a single donation to thaw at 4–8°C and removing the supernatant. The volume is about 20 ml and it is stored at -30°C. It contains factor VIIIC, factor VIII:von Willebrand factor, and fibrinogen. It is no longer used for the treatment of haemophilia A and von Willebrand disease because of the greater risk of virus transmission than with other available products, but it may be useful in disseminated intravascular coagulation if the fibrinogen is very low.

Blood products

Factor VIII and IX concentrates

These are conventionally freeze-dried preparations of specific coagulation factors prepared from large pools of plasma. They are used to treat patients with haemophilia and von Willebrand disease. Recombinant factor VIII is now recommended as the treatment of choice for patients with haemophilia A because of concerns about the safety of pooled plasma products.

Albumin

There are two preparations:

1. Human albumin solution (4.5 per cent), previously called plasma protein fraction (PPF), contains 45 g/l albumin and 160 mmol/l sodium. It is produced in 50-, 100-, 250-, and 500-ml bottles.
2. Human albumin solution 20 per cent, previously called ‘salt-poor' albumin, contains approximately 200 g/l albumin and 130 mmol/l sodium and is produced in 50- and 100-ml bottles.

Human albumin solutions are generally considered to be inappropriate fluids for acute volume replacement or the treatment of shock because they are no more effective in those conditions than crystalloids. However, albumin solutions are indicated for the treatment of acute, severe hypoalbuminaemia and can be used as the replacement fluid for plasma exchange. The 20 per cent albumin solution is particularly useful for patients with nephrotic syndrome or liver disease who are fluid-overloaded and resistant to diuretics. Albumin solutions should not be used to treat patients with malnutrition or chronic renal or liver disease.

Normal immunoglobulin

This is prepared from normal plasma. It is used in patients with hypogammaglobulinaemia to prevent infections, and in patients with immune thrombocytopenia.

Specific immunoglobulins

These are obtained from donors with high titres of antibodies. Many preparations are available, such as anti-D, antihepatitis B, antivaricella-zoster.

Strategies for avoiding or reducing the use of blood tranfusion

Concerns about the safety of transfusion have led to increased interest in strategies for avoiding or reducing the use of donor blood. Consideration should be given to factors in the patient's past medical history or drug therapy that indicate an increased risk of excessive blood loss, and to attempting to correct them preoperatively, for example by discontinuing antiplatelet and anticoagulant drugs, if possible, several days before surgery. Anaemia, if present, should be investigated and treated appropriately in advance of elective surgery. Intraoperative measures include the use of meticulous surgical and anaesthetic techniques, a cautious use of anticoagulants during surgery, and the use of drugs to enhance haemostasis, for example aprotinin.

A comprehensive strategy for avoiding transfusion of donor blood also includes the employment of strict criteria for the use of blood components and blood products, and consideration of autologous transfusion.

Autologous transfusion

An alternative to the use of donor blood for the replacement of blood lost during surgery is the use of the patient's own blood. Interest in autologous transfusion was mainly stimulated by concern, in the 1980s, about transmission of infection, especially human immunodeficiency virus, by blood transfusion. However, as well as reducing the risk of transfusion-transmitted infection, other hazards of transfusion are avoided, such as alloimmunization and immunosuppression. It should be noted that some risks are not less than with the transfusion of donor blood, such as giving an autologous unit to the wrong patient because of a simple clerical error, and bacterial contamination of the unit.

There are three types of autologous transfusion:

1. **Predeposit**. The patient donates 2–5 units of blood at approximately weekly intervals before elective surgery.
2. **Preoperative haemodilution**. One or two units of blood are removed from the patient immediately before surgery and retransfused to replace operative losses.
3. **Blood salvage**. Blood lost during or after surgery may be collected and retransfused. Several techniques of varying levels of sophistication are available, and an example of one system is shown in Fig. 1. The operative site must be free of bacteria, bowel contents, and tumour cells.

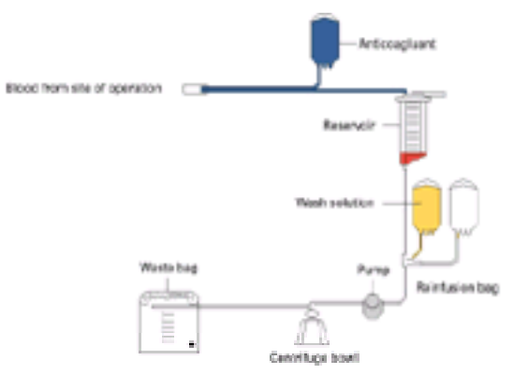


Fig. 1. Diagram of the Haemonetics cell saver blood salvage system.

Comprehensive guidelines exist in the United Kingdom for preoperative autologous donation and for perioperative haemodilution and cell salvage, including the selection of suitable patients and the practical aspects of collection, storage, and transfusion.

There has been little demand for autologous transfusion in the United Kingdom, as blood is generally perceived as being ‘safe'. In addition, there are considerable costs in setting up a predeposit autologous-transfusion service, which are perceived as being of benefit to only a minority of patients. Another problem is the difficulty of guaranteeing that elective surgery will not be cancelled. The limited use of autologous transfusion in the United Kingdom is in contrast to the United States of America, where autologous transfusion accounts for up to 5 per cent of all blood transfused in some regions.

Complications of blood transfusion

In the United States of America, it has been mandatory to report transfusion-associated deaths to the Food and Drug Administration since 1975; such reports have provided useful data that have contributed to efforts to improve the safety of blood transfusion. Similar reporting schemes under the term ‘haemovigilance' have been set up in other countries, including the Serious Hazards of Transfusion (SHOT) scheme that produced its first report in the United Kingdom in 1998.

Immediate haemolytic transfusion reactions

This is the most serious complication of blood transfusion and is usually due to ABO incompatibility. There is complement activation by the antigen-antibody reaction, usually due to IgM antibodies, leading to rigors, lumbar pain, dyspnoea, hypotension, haemoglobinuria, and renal failure. The initial symptoms may occur a few minutes after starting the transfusion. Activation of coagulation may also occur; bleeding due to disseminated intravascular coagulation is a bad prognostic sign. Emergency treatment may be needed to maintain the blood pressure and renal function.

At the first suspicion of any serious transfusion reaction, the transfusion should always be stopped and the donor units returned to the blood transfusion laboratory with a new blood sample from the patient to exclude a haemolytic transfusion reaction.

Delayed haemolytic transfusion reactions

These may occur in patients alloimmunized by previous transfusions or pregnancies. The antibody titre is too low to be detected by pretransfusion compatibility testing, but a secondary immune response occurs after transfusion, resulting in destruction of the transfused cells, usually by IgG antibodies. The patient may develop anaemia and jaundice about a week after the transfusion, although many are clinically silent.

Non-haemolytic (febrile) transfusion reactions

Febrile reactions are a common complication of blood transfusion in patients who have previously been transfused or pregnant. The usual cause is the presence of leucocyte antibodies in the recipient acting against transfused leucocytes, leading to release of pyrogens. Typical signs are flushing and tachycardia, fever (>38°C), chills, and rigors. Paracetamol may be used to reduce the fever. Febrile reactions may be prevented after further transfusions by the use of buffy coat-depleted blood.

Potent leucocyte antibodies in the plasma of donors, who are usually multiparous women, may cause severe pulmonary reactions (called transfusion-related acute lung injury or TRALI) characterized by dyspnoea, fever, cough, and shadowing in the perihilar and lower lung fields on the chest radiograph.

Urticaria and anaphylaxis

Urticarial reactions are often attributed to plasma protein incompatibility but, in most cases, they are unexplained. They are common but rarely severe; stopping or slowing the transfusion, and intravenous chlorpheniramine 10 mg (adult dose), are usually sufficient treatment.

Anaphylactic reactions occasionally occur; severe reactions are seen in patients lacking IgA who produce anti-IgA that reacts with IgA in the transfused blood. The transfusion should be stopped and adrenaline 0.5 mg intramuscular and chlorpheniramine 10 mg intravenous should be given immediately; endotracheal intubation may be required. Patients who have had severe urticarial or anaphylactic reactions should receive either washed red cells, autologous blood, or blood from IgA-deficient donors for those with IgA deficiency.

Transmission of infection

The incidence of post-transfusion hepatitis was estimated to be about 1 per cent in the United Kingdom before testing for antibodies against hepatitis C virus was introduced in 1991. As most cases were non-A, non-B hepatitis due to hepatitis C virus, the incidence of post-transfusion hepatitis has decreased since then to about 1 in 200 000 units transfused. Each donation has been tested for hepatitis B surface antigen for many years, and the incidence of transmission of hepatitis B virus is also about 1 in 200 000 units transfused. Other viruses that may cause post-transfusion hepatitis include cytomegalovirus, Epstein–Barr virus, and hepatitis G virus (HGV), and there are likely to be other as yet unidentified viruses.

In the United Kingdom the incidence of transmission of human immunodeficiency virus by blood transfusion is extremely low, probably less than 1 in 3 million units transfused. Prevention is based on self-exclusion of donors in ‘high-risk' groups and testing each donation for antibodies to human immunodeficiency virus.

Transfusion-transmitted syphilis is now very rare in the United Kingdom. Spirochaetes do not survive for more than 72 h in blood stored at 4°C and each donation is tested using the *Treponema pallidum* haemagglutination assay (TPHA).

Bacterial contamination of blood components is potentially a very serious event, and although rare it is one of the most frequent causes of death associated with transfusion after haemolytic transfusion reactions. Some organisms, such as *Yersinia enterocolitica*, can proliferate in red cell concentrates stored at 4°C, but platelet concentrates stored at 22°C are a more frequent cause of this problem. Several systems have been proposed to reduce the risk of bacterial contamination, including automated culture systems and bacterial antigen-detection systems, but none is currently in routine use.

Following the problems with transmission of human immunodeficiency virus and hepatitis C virus in the 1980s, there is current concern about the risk of transmission by transfusion of the agent responsible for new variant Creutzfeldt–Jakob disease (**CJD**). Recent data suggest that new variant CJD and bovine spongiform encephalopathy are caused by the same abnormal prion protein, a variant of a normal protein present on cell surfaces. So far there are only 39 confirmed cases of new variant CJD in the United Kingdom, but it is not possible to predict the number of cases in the next few years. There has been no proven case of transmission of classical or new variant CJD by blood transfusion.

In relation to measures to reduce the risk of transmission of new variant CJD by blood transfusion, there are no known criteria for exclusion of blood donors, and, as yet, no screening test. There are data to suggest that the abnormal prion protein is associated with lymphocytes; it has been detected in the tonsils of patients suffering from new variant CJD, and recently it has been demonstrated that B lymphocytes are crucial in neuroinvasion. For these reasons, universal leucocyte-depletion of blood is being implemented in the United Kingdom. However, the infective dose of abnormal prion protein is unknown, and the reduction in risk which leucocyte depletion might produce is uncertain.

Given present knowledge, the most logical precaution available against transmission of new variant CJD by blood transfusion is, where practical, not to use blood components or plasma donated in countries where clusters of cases of that disease are present. In 1998, the Department of Health, on the advice of the Committee on Safety of Medicines, announced that United Kingdom donor plasma will not be used for manufacture of medicinal products, and that plasma fractionation centres in the United Kingdom will be allowed to import plasma from countries without clusters of cases of new variant CJD. The Medicines Control Agency will inspect the suppliers of this plasma to determine that their source donors, even if paid, are as safe as donors in the United Kingdom in respect of viral transmission.

While stringent measures are being taken to minimize the risk of transfusion-transmitted infection in the United Kingdom, it may never be possible to guarantee that donor blood is absolutely ‘safe'. The current approach to the safety of blood components and plasma in the United Kingdom is extremely cautious, but no absolute guarantee of safety.

Immunosuppression

Ever since observations were made of the favourable effect of transfusion on the survival of subsequent renal allografts, the basis of transfusion-induced immunomodulation has been the subject of debate. It is assumed that allogeneic leucocytes are required to cause transfusion-induced immunosuppression, but the underlying mechanisms remain uncertain. There has been considerable interest in other clinical effects caused by transfusion-induced immunosuppression, such as postoperative infection and tumour recurrence. However, the findings have been conflicting, as outlined next.

Postoperative infection

In a study of patients with colorectal cancer, the rate of postoperative infection was greater in those receiving whole-blood transfusions than in non-transfused patients, but patients receiving leucocyte-depleted red cell concentrates had the same incidence of infection as the non-transfused. However, there were conflicting results from two randomized studies comparing rates of postoperative infection in patients receiving leucocyte-depleted or buffy coat-depleted, red cell concentrates: one study found no difference, and one found a significantly lower incidence in the patients receiving leucocyte-depleted blood.

A recent analysis of trials in this area concluded that any transfusion effect on infection was small, and that leucocyte-depletion did not confer any convincing benefit over other components, and another agreed that no definite conclusion could be made at the present time. If this effect of blood transfusion on postoperative infection is significant, the clinical and economic impact could be substantial. There may be an opportunity to study this question in a large clinical trial, while universal leucocyte-depletion of blood is being implemented in the United Kingdom.

Cancer recurrence

Retrospective observational studies of cancer patients are conflicting on whether tumour-free survival is influenced by transfusion, with almost equal numbers for and against the hypothesis. There was no difference in tumour-free survival when leucocyte-depleted and buffy coat-depleted, red cell concentrates were compared in patients with colonic cancer. As with the effect of transfusion on postoperative infection, it will be a major advance if the risk of cancer recurrence can be reduced by avoiding transfusion of donor blood or using leucocyte-depleted blood.

Further reading

Barbara J, Flanagan P. Blood transfusion risk: protecting against the unknown. *British Medical Journal* 1998; **316**: 717–18.

Blajchman MA. Allogeneic blood transfusions, immunomodulation, and post-operative bacterial infection: do we have the answers yet? *Transfusion* 1997; **37**: 121–5.

British Committee for Standards in Haematology. Guidelines for implementation of a maximum surgical blood ordering schedule. *Clinical and Laboratory Haematology* 1990; **12**: 193–201.

British Committee for Standards in Haematology. Guidelines for transfusion for massive blood loss. In Roberts, BE, ed. *Standard haematology practice*, pp. 198–206. Blackwell Scientific, Oxford, 1991.

British Committee for Standards in Haematology. Guidelines for autologous transfusion. I. Pre-operative autologous donation. *Transfusion Medicine* 1993; **3**: 307–16.

British Committee for Standards in Haematology. Guidelines for autologous transfusion. II. Peri-operative haemodilution and cell salvage. *British Journal of Anaesthesia*, 1997; **78**: 768–71.

Brown P. Can Creutzfeldt–Jakob disease be transmitted by transfusion? *Current Opinion in Haematology* 1995; **2**, 472–7.

Bruce ME *et al.* Transmissions to mice indicate that ‘new variant' CJD is caused by the BSE agent. *Nature* 1997; **389**: 498–501.

Blumberg N, Heal JM. Blood transfusion immunomodulation: the silent epidemic. *Archives of Pathology and Laboratory Medicine* 1998; **122**: 117–19.

Consensus Conference on Autologous Transfusion, Royal College of Physicians, Edinburgh. Final consensus statement. *Transfusion* 1996; **36**: 667.

Consensus Conference on Red Cell Transfusion, Royal College of Physicians, Edinburgh. Consensus statement on red cell transfusion. *Transfusion Medicine* 1994; **4**: 177–8.

Cooley DA. Conservation of blood during cardiovascular surgery. *American Journal of Surgery* 1995; **170**: 53–9S.

Herbert PC *et al.* A multicenter, randomized, controlled clinical trial of transfusion requirements in critical care. *New England Journal of Medicine* 1999; **340**: 409–17.

Hill AF, Zeidler M, Ironside J, Collinge J. Diagnosis of new variant Creutzfeldt–Jakob disease by tonsil biopsy. *Lancet* 1997; **349**: 99–100.

Hill AF *et al.* The same prion strain causes vCJD and BSE. *Nature*, 1997; **389**: 448–50.

Houbiers JG *et al.* Randomised controlled trial comparing transfusion of leucocyte-depleted or buffy-coat-depleted blood in surgery for colorectal cancer. *Lancet*, 1994; **344**: 573–8.

Jensen LS *et al.* Postoperative infection and natural killer function following blood transfusion in patients undergoing elective colorectal surgery. *British Journal of Surgery* 1992; **79**: 513–16.

Jensen LS, Kissmeyer-Nielsen P, Wolff B, Qvist N. Randomised comparison of leucocyte-depleted versus buffy-coat-poor blood transfusion after colorectal surgery. *Lancet* 1996; **348**: 841–5.

Klein HG, Dodd RY, Ness PM, Fratantoni JA, Nemo GJ. Current status of microbial contamination of blood components: summary of a conference. *Transfusion* 1997; **37**, 95–101.

Klein MA *et al.* A crucial role for B cells in neuroinvasive scrapie. *Nature* 1997; **390**: 687–90.

McClelland DBL. Red cell transfusion for elective surgery: a suitable case for treatment. *Transfusion Medicine* 1994; **4**: 247–9.

Murphy MF, Stevens W, Green ES, Allison P, Smith D. Universal leucocyte-depletion of blood components—con. *Infusion Therapy and Transfusion Medicine* 1999; **25**: 305–11.

Opelz G, Sengar DPS, Mickey MR, Terasaki PI. (1973). The effect of a blood transfusion on subsequent kidney transplantation. *Transplantation Proceedings* 1973; **5**: 253–9.

Popovsky MA, Chaplin HC, Moore SB. Transfusion-related acute lung injury: a neglected, serious complication of hemotherapy. *Transfusion* 1992; **32**: 589–92.

Sanguis Study Group. Use of blood products for elective surgery in 43 European hospitals. *Transfusion Medicine*, 1994; **4**: 251–68.

Sazama K. Reports of 355 transfusion-associated deaths: 1976 through 1985. *Transfusion* 1990; **30**: 583–90.

Schierhout G, Roberts, I. Fluid resuscitation with colloid or crystalloid solutions in critically ill patients: a systematic review of randomised trials. *British Medical Journal* 1998; **316**: 961–4.

Vamvakas EC. Transfusion-associated cancer recurrence and postoperative infection: meta-analysis of randomised, controlled clinical trials. *Transfusion*, 1996; **36**: 175–86.

Vamvakas EC, Carven JH. Allogeneic blood transfusion, hospital charges, and the length of hospitalization: a study of 487 consecutive patients undergoing colorectal cancer resection. *Archives of Pathology and Laboratory Medicine* 1998; **122**: 145–51.

Vamvakas E, Moore SB. Perioperative blood transfusion and colorectal cancer recurrence: a qualitative statistical overview and meta-analysis. *Transfusion*, 1993; **33**: 754–65.

6 Wound healing

George W. Cherry, Margaret A. Hughes, Mark W. J. Ferguson, and David J. Leaper

Introduction
Evidence-based medicine in wound healing
Wound Debridement
Factors Affecting Quality of Healing
Healing Phases
Sequence in inflammation
Platelets
Accumulation of neutrophils
Accumulation of macrophages
Lymphocytes
Epithelial cells
Formation of Granulation Tissue
Fibroblasts
Angiogenesis
Re-epithelialization and basement membrane formation
Contraction
Collagen—Matrix Formation and Remodelling
Extracellular matrix
Synthesis of collagen
Types of collagen
Collagenolysis
Collagen and wound healing
Fetal wound healing
Wound remodelling—pharmacological control
Experimental wound-healing models
Regulation of Wound Healing by Growth Factors
TGF-b1 and scarring
Acute wounds
Chronic wounds
Clinical studies
Anti-growth factor therapy
Healing Deficits
Steroid impaired healing
Irradiation
Cytotoxics
Diabetes
Rheumatoid arthritis
Trace element imbalance
Non-Cutaneous Tissue
Wound Dressings
Tissue Engineering
New and Alternative Treatments
The Future
Further reading

Introduction

Nowhere is the gap between basic research and clinical application more glaring than in the biology of wound healing

Earl A. Peacock Jr (1983)

The consensus definition of a wound, agreed by specialists from different disciplines involved with wound healing, is that it is a disruption of normal tissue structure and function. Wounds can result from injurious processes beginning either internally or externally to the involved organs. These can range from controlled acute disruption of tissue by the surgeon's knife (Fig. 1, Fig. 2) to widespread trauma such as burns. The difference between acute wounds and chronic wounds is that in the former healing occurs through an orderly and timely process leading to restoration and functional integrity, while, on the other hand, chronic wounds fail to follow this course and are often associated with underlying pathology. The process of wound healing has been defined by Davidson as a succession of cellular events which are co-ordinated by the release and recognition of soluble mediators. One or more of these mediators may be rate-limiting for the healing process at one of a number of stages.

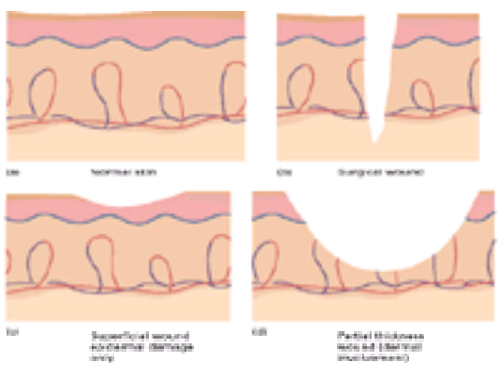


Fig. 1. Types of wounds. Healing takes days or months, depending on size. (a) Normal skin. (b) Incisional wound can be closed by sutures. The edges of the various tissues need to grow together, but a minimal amount of new tissue is required. (c) The wound is restricted to the epidermis, healing can take place by regeneration of tissue and there is no scarring. (d) Open wound that is too large to suture. It must be filled with new tissue, requiring formation of new connective tissue, epidermis, and basement membrane.

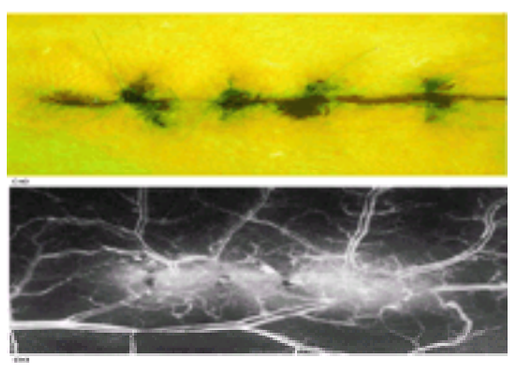


Fig. 2. (a) Experimental freshly made surgical wound in which the sutures have been purposely tied tightly, after which intravenous fluorescein dye was given. The non-stained dark areas are avascular due to the tight sutures. (b) Microangiogram of same wound 7 days later, demonstrating avascular areas where the sutures were

tied tightly, but also illustrating the normal increase in angiogenesis to the wound induced by the surgery. Angiogenesis is a key factor in wound healing and when impaired results in poor healing.

The aim of this chapter is to review the basic science of wound healing, as well as to illustrate how research is influencing the clinical management of wounds at the end of the twentieth century.

Evidence-based medicine in wound healing

As in other medical specialties there is now an emphasis on the need for evidence-based practice in wound healing. The initial quotation from Peacock in 1983, used to open the chapter on wound healing in the first edition of the *Oxford textbook of surgery*, is now beginning to be challenged. The gap between basic wound healing research and its clinical application is at last narrowing. A number of factors are contributing to this change, but probably the most important is the evidence from randomized, prospective clinical trials of new therapies which have led to improved healing. An instance of this is the case of diabetic wounds (Fig. 3), where recombinant human platelet-derived growth factor has recently become the first growth factor to gain approval from the Federal Drug Administration to treat this type of wound. Another example is the use of tissue-engineered skin equivalents for difficult wounds such as venous leg ulcers and diabetic foot ulcers (Fig. 4, Fig. 5).



Fig. 3. Typical diabetic foot wound where recombinant platelet-derived growth factor has recently been shown in controlled clinical trials to enhance healing of this type of wound.

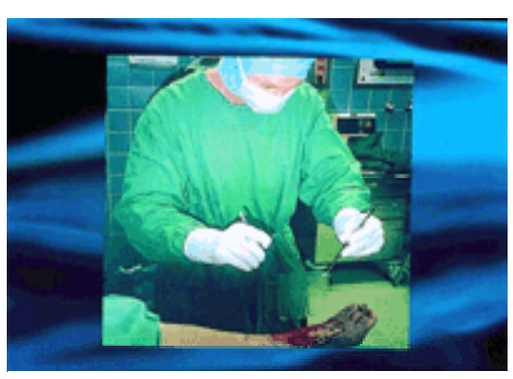


Fig. 4. Human cultured autologous keratinocytes on an esterified hyaluronic acid membrane (Vivoderm) being transferred to patient.



Fig. 5. The healing of a diabetic heel wound with Dermagraft, which consists of cultured neonatal fibroblasts on a polyglycolic acid/polyglactin-9,10 mesh. The Dermagraft was applied weekly for a period of 8 weeks.

The Cochrane Wounds Group has established a database of systematic reviews with complete protocols including studies on pressure sores, venous ulcers, oral zinc treatment, and other wound healing topics. Information on these can be obtained from the website (web page at: http:/www.york.ac.uk\health/depts/hstd/projects.htm wounds).

Wound debridement

Debridement is an important aspect of the treatment of wounds. The modern meaning of debridement, removal of devitalized tissue, is quite different from its early definition which was the opening of a wound as for drainage. In a recent multicenter study by Steed and colleagues examining the effect of platelet-derived growth factor on the healing of diabetic foot ulcers, where surgical debridement was part of the standard therapy, they observed a lower healing rate in centers which performed less frequent debridement. A European Tissue Repair Society working party on wound debridement reviewed this technique as well as the different methods of debridement currently available. These methods included surgical, collagenase, autolytic with interactive dressings, and the revival of larvae (maggots) therapy. The latter was extremely popular during the 1930s and 1940s in North America prior to the advent of antibiotic therapy, particularly in the treatment of osteomyelitis. In the United Kingdom, Europe, and North America this technique has regained popularity in the treatment of difficult chronic wounds (Fig. 6). The antibacterial effect of maggots is now being researched extensively and *in vitro* studies have shown that maggots are effective against methicillin-resistant *Staphylococcus aureus* as well as other species of bacteria.



Fig. 6. (a) Toe wound on a patient with diabetes before and after 3 weeks of maggot therapy: 61-year-old man with diabetes receiving surgical and intravenous antibiotic therapy for several weeks without improvement of his foot ulcer. After 3 weeks of maggot debridement therapy (MDT), his wound was debrided and healing rapidly (reproduced by kind permission of R. A. Sherman, University of California at Irvine, USA). (b) Sloughy venous ulcer (superior) before initiating treatment with larvae (maggot) therapy. (c) Same ulcer 6 days later with marked increase in granulation tissue as well as in the inferior ulcer being treated with the same therapy. The hydrocolloid dressing around the ulcer is used for protection of the surrounding skin.

The difficulty of assessing the extent of the devitalization in wounds in order to determine the extent of debridement needed is an ongoing debate. Methods investigated to determine devitalization more accurately than routine clinical assessment include tissue pH, Po_2, and perfusion (using for example laser Doppler or fluorescein dye techniques). These latter techniques at this stage are mainly research tools.

One of the objectives of treating chronic wounds is to convert these wounds to an acute stage. Sharp surgical debridement is often looked upon as the treatment of choice. However, many patients with chronic wounds never reach a hospital environment and means of providing this type of debridement in the community is difficult, but would be of great benefit. A number of studies examining the analgesic effect of topical EMLA® (5 per cent cream, idocaine and prilocaine) for surgical debridement have been published and show that this agent is effective in controlling pain in patients undergoing debridement, particularly in arterial and venous ulcers.

In healthy individuals, tissue repair after acute injury is thought to occur at a maximum rate and is not really impaired unless adversely affected by localized or systemic conditions (Table 1).

Local	Systemic
Impaired blood supply	Peripheral vascular disease (arterial, venous, and microcirculation)
Bacterial and fungal infection	General malnutrition, or deficiency in:
	proteins, vitamins (especially A and C), trace elements
Complications following radiotherapy	Chemotherapy, or whole body irradiation
Excessive use of topical corticosteroids	Diabetes mellitus, rheumatoid arthritis, metabolic diseases
	jaundice, anaemia
Foreign body reactions	Marfan's syndrome
	Immunosuppressant drugs and corticosteroids
	Anticoagulants
	Ageing (fewer protein synthesis, longer duration of inflammation)
	Obesity
	Mental stress, lifestyle
	Systemic shock leading to impaired oxygenation of tissues

Table 1 Local and systemic factors that adversely affect wound healing

An augmentation in the healing rate of wounds would not only be beneficial to the patients but also to health care budgets. For example the costs to the National Health Service in the United Kingdom of just two types of chronic wounds are: leg ulcers £230 to 400 million per annum (at 1991 prices); prevention and treatment of pressure ulcers between £600 000 and 3 million per annum for a single 600-bed hospital. The total cost of treating chronic wounds in the United Kingdom has been stated to be equal to that of the cost of tobacco-related diseases. With reference to the cost of acute wounds, a number of which fail to heal, for example dehisced or infected wounds, the latter are costly in financial terms. In critical cases such as anastomotic leaks of bowel anastomoses due to poor blood supply or other complications these may lead to prolonged morbidity or mortality.

Factors affecting quality of healing

To a number of patients the most important perception of successful healing is not necessarily acceleration of closure or the degree of tensile strength, but the appearance of the resultant scar.

The way in which surgical wounds are closed can affect both cosmesis and the quality of healing. Staples are used to close surgical wounds and have been used to close skin wounds since Roman times. Staple devices (later chapters) allow low anterior resection and trans-abdominal oesophagojejunostomy, for example. Such devices can shorten operative times, but are expensive. However, simple applicators are very inexpensive (Michel clips after thyroid surgery). With staples there is an associated low wound infection rate. Steri-Strips can be used to close lacerations. They have the advantage over sutures, particularly in children, and are associated with good cosmetic results and low infection rates. The favorable results for Steri-Strips are probably due to the lack of interference to the blood supply of the wound, which can be affected by tying sutures tightly (Fig. 2). A disadvantage of tape closures is that they need a dry skin edge for adherence and have not been popular immediately after major operative surgery. However, to improve cosmesis in the healing of facial wounds or those on other visible sites, sutures are often replaced by Steri-Strips after a few days. Subcuticular skin closure gives good cosmetic results by avoiding the ‘herringbone' scars of interrupted sutures. It is not clear whether an absorbable suture (such as polydioxane) or a non-absorbable suture (such as polypropylene) which needs removal is better.

For the patient with burns, in addition to the cosmesis of the scar, the functional recovery of the injured tissue is of great importance. This goal has been accomplished for a number of years by using pressure garments to manage excess scar tissue formation in these patients. Research workers in Australia, Great Britain, and the United States have shown that covering hypertrophic scars with a Silastic gel sheet results in both clinical and elastometrical improvement of the scars, without exerting pressure. The mode of action of this therapy in preventing hypertrophic scar formation is still not known, but studies by Quinn and colleagues have ruled out any pressure effect, temperature alteration, oxygen tension change, or occlusion as a factor. Some of the more exciting research on scar control is that related to the use of neutralizing antibody to transforming growth factor-b (**TGF-b**) and to the use of calcium antagonists. These are discussed in more detail in later sections.

Another example where basic wound healing research has eventually been utilized clinically is that of the practice of maintaining a moist wound environment under occlusive dressings to enhance healing. An increase in healing compared with wounds left exposed to air was demonstrated in animal studies in 1962 (Fig. 7) and a year later in humans. However, there is a still a query as to whether primary sutured wounds of skin need to be covered or left exposed to the air.

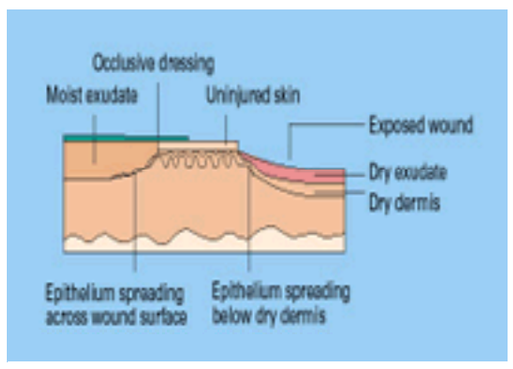


Fig. 7. Diagram of moist wound healing, illustrating the migration of epithelium in a skin wound covered with an occlusive dressing.

Healing phases

In order to describe the complex cascade of events that follows injury, it is convenient to look at this process as a number of overlapping phases: inflammation, formation of granulation tissue with angiogenesis, and scar formation (extracellular matrix remodelling) (Fig. 8, Fig. 9).

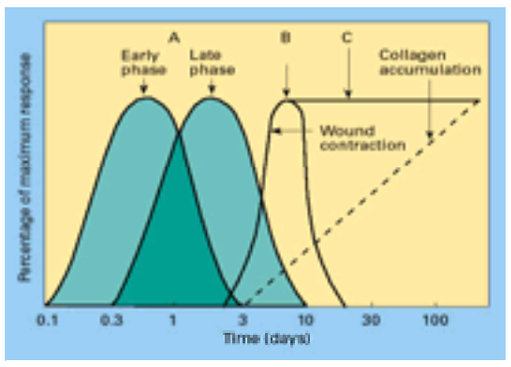


Fig. 8. Phases of wound repair. (A) Early phase of inflammation — accumulation of neutrophils, late phase of inflammation – accumulaton of macrophages; (B) repair phase — formation of granulation tissue; (C) scar formation and maturation.

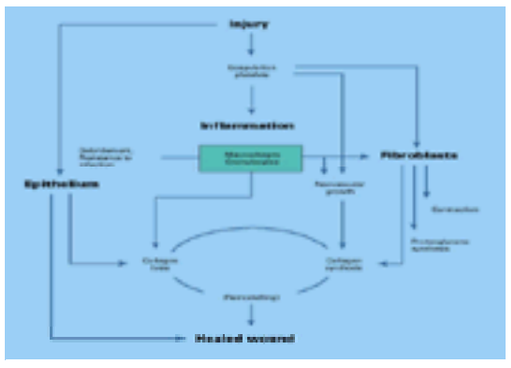


Fig. 9. A general flow diagram of wound healing. Note the central role played by inflammation

Injury to tissue leads to loss of structural integrity, instigating the coagulation cascade to prevent localized haemorrhage. In skin, mucosa, and gut especially, injury is also complicated by the invasion of micro-organisms. These events play an important role in initiating the defense and repair mechanisms by sealing off severed vessels and transferring blood constituents, circulating cells, and bioactive substances to the site of the wound (Fig. 10). This transfer and the ensuing defense processes constitute the early aspect of wound healing, commonly referred to as the inflammatory phase.

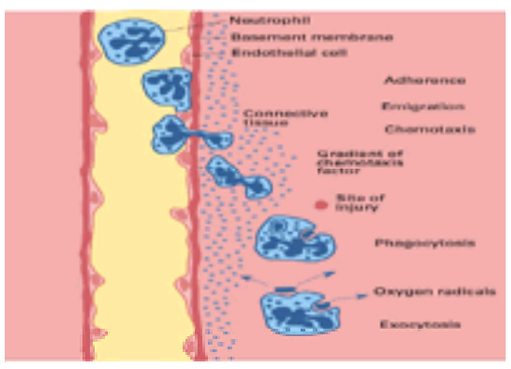


Fig. 10. Scheme depicting the various stages of neutrophil accumulation and participation in the early inflammatory phase of wound healing.

It has been stated by Peacock and others that the inflammatory reaction in soft tissue, which begins literally seconds after injury, is the same whether caused by a surgeon's sterile blade or by invading bacteria after a street injury. Qualitatively, inflammation is the same, but it is likely to be more prolonged in the latter case. More specifically, the mechanism of leucocyte adhesion to the vascular wall after injury followed by diapedesis, a major part of the inflammatory response, is essentially the same in all wounds whether resulting from surgery or trauma.

Sequence in inflammation

The inflammatory phase is triggered by two classes of mediators (soluble signal factors): those controlling vessel permeability and those attracting or trapping cells. The clinical signs of inflammation are caused by changes in blood vessels with dilatation leading to erythema, and endothelial cell separation allowing plasma extravasation, producing localized swelling. There are overlapping stages but, in general, the order of arrival at the wound site from an intravascular space is thought to occur in the following sequence: plasma with soluble components and cellular constituents, first platelets, then neutrophils, followed by monocytes and lymphocytes. The migration of epithelial cells to resurface the injured tissue begins during this phase, mediated by the above events. Alterations in microvascular permeability after injury allow both fluid and plasma components to pass to the tissue. Vasoactive amines and peptides (including histamine from mast cells, serotonin from platelets, and bradykinin from neutrophils) cause the reversible opening of junctions between endothelial cells and allow the passage of neutrophils and monocytes.

Hageman factor (factor XII), a plasma glycoprotein, is activated by adsorption onto fibrillar collagen, leading to the generation of bradykinin and initiation of the complement cascade. The complement system is composed of 20 interacting soluble proteins in the serum and extracellular fluid, which can be activated by IgM and IgG antibodies bound to antigens on the surface of micro-organisms or by bacterial lipopolysaccharides. Large quantities of IgM or IgG antibodies lead to complement fixation by the classical pathway, whereas endotoxin released from bacteria and small quantities of IgG antibody enhance the activation process by the alternate pathway. These proteins are the substances responsible for the acute inflammatory reaction. IgM can lyse Gram-negative bacteria and neutralize viruses. The C5 to C9 factors of complement combine to form a large protein complex that mediates lysis of bacterial cell walls. Complement factors also opsonize invaders (coat their antigen with antibody), making them recognizable to phagocytic cells. The factor C5a is also chemotactic, attracting polymorphonuclear cells, neutrophils, to the site. The complement component C3b binds to specific receptor proteins on phagocytic cells and to microbial cell walls and enhances the ability of the phagocytes to bind, ingest, and destroy micro-organisms.

Platelets

The earliest circulating cell or cell fragment detected in the injury site is the platelet. Platelets contain three types of organelles involved in haemostasis and initiation of the inflammatory phase.

1. a-Granules, which contain adhesive glycoproteins such as fibrinogen, von Willebrand factor, fibronectin, thrombospondin, and also growth factors: platelet-derived growth factor (**PDGF**), transforming growth factors-a and -b (TGF-a and TGF-b), and platelet factor 4.
2. The ‘dense body', the main storage site of serotonin, also contains adenine nucleotide, calcium, and pyrophosphates.
3. Lysosomes, containing neutral and acid hydrolysases, elastase, collagenase, antitrypsin, and a_2-macroglobulin.

The above substances are released when the platelets are activated by various factors. When injury occurs, contact is made between platelets and insoluble components of the subendothelial matrix, particularly collagen, promoting the release of the a-granule contents which then trigger the coagulation process. The activation of platelets is enhanced by some of the complement factors and by bacterial lipopolysaccharides. The latter produce a 50-fold increase in the amount of serotonin released.

Activated platelets become sticky and aggregate to form a plug that temporarily occludes small vessels. Both damaged platelets and tissues release thrombokinase, which converts prothrombin to thrombin, and this in turn ensures the conversion of soluble fibrinogen to insoluble fibrin. The release of serotonin and adenine nucleotides contained in the dense bodies of the platelets induces the aggregation of platelets, which interact with the fibrin network to form a clot that is stronger and more durable than the initial platelet plug. If the clot is allowed to dehydrate, it transforms into a dry eschar covering the wound.

Other substances released by the a-granules, such as PDGF, and by the dense body, such as cyclic adenosine monophosphate, are chemotactic for neutrophils; both TGF-b and PDGF are chemotactic for macrophages, while TGF-a and TGF-b are angiogenic factors. The role of PDGFs in enhancing both experimental and clinical wound healing has been highlighted recently in a number of publications, and is elaborated in more detail below.

Accumulation of neutrophils

Adhesion

Interaction between damaged tissue and serum releases the complement factor C3, and the C3e fragment of this provokes the release of neutrophils from the bone marrow. At the same time, circulating leucocytes near the wound site, particularly neutrophils, cease to flow and adhere to the endothelium. It has been shown *in vitro* that adherence is enhanced by inflammatory mediators, such as C5a (the fifth component of complement), platelet-activating factor, and leucotriene. There is a very fast initial response, with onset of adherence as early as 30 s after injury and with a maximum response at 2 min.

The binding of leucocytes to endothelium results from the interaction of complementary receptors in both cell types. Their expression is enhanced by cytokines and bacterial lipopolysaccharide. Physical factors, such as haemodynamic shear stress, also influence adherence. This first stage of adherence is critical. While there is some evidence that some wounds can heal without the presence of neutrophils, patients with leucocyte adhesion deficiency, lacking an essential glycoprotein, are unable to mobilize neutrophils or monocytes, and exhibit decreased pus formation and impaired wound healing.

Diapedesis

Vasopermeability factors act on actin microfilaments inside the endothelial cells and effect the reversible opening of junctions so that neutrophils are able to pass between the endothelial cells to the extravascular space. It is suggested that the secretion of elastase and other enymes by the neutrophils enables them to degrade elastin and components of the endothelial basement membrane.

Migration

Molecules released by platelets following disruption of the blood vessels, such as kallikrein (an enzyme that leads to the formation of vasodilating peptides) and fibrinopeptides, diffuse to the site of the wound and set up a concentration gradient of chemotactic factors which attract the neutrophils that have traversed the endothelium through the extracellular space to the injury site.

Phagocytosis

At the site the neutrophils form the first line of defense against the invading micro-organisms. The neutrophils phagocytose bacteria, then kill the ingested cells by the production of microbiocidal substances, for example oxygen metabolites such as hydroxyl radicals, hydrogen peroxide, and the superoxide ion. Release of some of these substances to the outside of the cell may also lead to tissue damage and prolong the inflammatory phase. Some bacteria may be killed by non-oxidative mechanisms, but these are not defined *in vivo.* If bacterial contamination is low, the density of neutrophils declines, but if numbers of micro-organisms persist, the bacterial lipopolysaccharides continue to promote the arrival of further neutrophils. The neutrophils are unable to regenerate their enzymes and so themselves decay after phagocytosis.

Accumulation of macrophages

The macrophage is indispensable in the degradation of injured tissue debris and in the reparative phase of wound healing (Fig. 11). If the macrophages are inhibited, wound healing is radically impaired. Normal tissues contain very few macrophages, but, in response to chemotactic factors released after injury, circulating monocytes are attracted to the site of injury several hours after the first neutrophils arrive. Endothelial cells in wounded tissue also play a role in this process, and have been shown to regulate the preferential adhesion of monocytes and lymphocytes to endothelium.

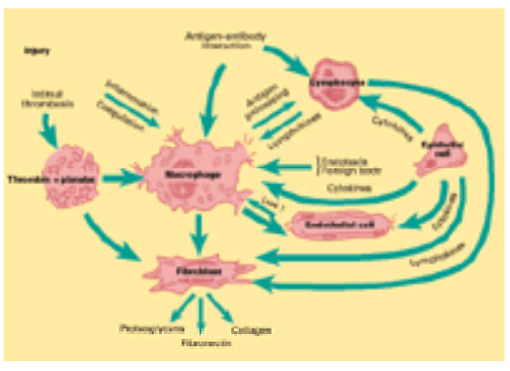


Fig. 11. Cellular biology of wound healing. Note the central role played by the macrophage.

At the injury site, monocytes differentiate into macrophages. One of the signals promoting this differentiation is the binding of fibronectin to surface receptors on monocytes, which induces the activation of the receptors for phagocytosis. Macrophages develop functional complement receptors and undertake similar operations to the neutrophils. However, further interactions with the interferons, and subsequently with bacterial or viral products, induce further differentiation into a fully activated phenotype. Interferons enhance endocytosis and phagocytosis and modulate the surface receptor functions of newly migrated macrophages. Ingestion of bacteria by endocytosis triggers the primary oxygenase that converts molecular oxygen to the superoxide, which then reacts to produce hydrogen peroxide and hydroxyl radicals required for microbiocidal activity. Oxygen is essential. If the partial pressure of oxygen falls below 30 mmHg, macrophages are inactivated; their phagocytosing potential is reduced. The relationship between oxygen pressure and healing has been shown to be linear, explaining the beneficial role of oxygen pressure in repair.

The activated macrophage is the major effector cell for degrading and removing damaged connective tissue components, collagen, elastin, and proteoglycans. Initial degradation takes place extracellularly—up to several millimeters from the macrophage. Collagen and other fragments are then ingested and degraded by the

cathepsin enzymes and other peptides. In contrast to neutrophils, macrophages can continue to synthesize the necessary enzymes, thus persisting for a longer time. They also phagocytose the decaying neutrophils.

Apart from their role in debridement, macrophages secrete chemotactic factors which bring additional inflammatory cells to the wound site. Macrophages also produce prostaglandins, which are strongly vasodilatory and affect the permeability properties of microvessels. The macrophages act after the amines and kinins, and are produced on demand, prolonging the inflammatory phase. Prostaglandins also augment the adenyl cyclase activity in T lymphocytes, which accelerates the mitosis of other cells.

The angiogenesis stimulated in the early phase of wound healing has been shown to be related to the presence of macrophages. Increased levels of lactate production, up to 15-fold, have been found in wounded tissue, and have caused macrophages to produce and release angiogenic substances. The macrophages also produce growth factors, such as PDGF, TGF-b, and fibroblast growth factor (**FGF**), which are necessary for the initiation and propagation of granulation tissue. In this way the macrophages mediate the transition from the initial inflammatory response to the early repair phase of wound healing.

Lymphocytes

B lymphocytes may be absent from the wound site. However, helper T cells are activated following injury, when they recognize any foreign antigen on the surface of antigen-presenting cells, such as Langerhans cells in skin, and certain types of macrophage.

The T lymphocytes migrate into the wound along with the macrophages. Advances in the past 5 years have helped to elucidate the role of the T cell in wound healing. Monoclonal antibody staining has permitted the identification of sets and subsets of lymphocytes, and cell culture and biochemical studies have identified and characterized some of the lymphokines, molecular messengers secreted by lymphocytes, which influence other cells, particularly macrophages and fibroblasts (Fig. 11). Thus, lymphocytes can produce macrophage chemotactic factor, macrophage inhibiting factor that regulates movement, macrophage activating factor, and interleukin 2 (**IL-2**), which enables the T cells to proliferate by an autocrine mechanism. TGF-b, produced by the a-granules of platelets, is chemotactic for both fibroblasts and macrophages, and interferon-g modulates the surface receptor function of newly migrated macrophages and enhances their phagocytic activity, and also activates macrophage oxidative metabolism and antimicrobial activity. T lymphocytes also produce colony stimulating factors (**CSF**). These are glycoproteins that act on neutrophils and macrophages through specific receptors which have recently been identified: granulocyte-CSF, macrophage-CSF, granulocyte/macrophage-CSF, and interleukin-3 (**IL-3**).

The colony stimulating factors are very potent, being effective at very low concentrations (pg/ml). They are involved in the stimulation of proliferation, and of the commitment of the monocyte to differentiation and maturation. They stimulate the function of phagocytosis, and the production by macrophages of substances such as prostaglandins, tumour necrosis factor-a, interferon-g, and further colony stimulating factors. As quantities are very small, it is not known whether all cells are able to produce colony stimulating factors. They are induced *in vivo* by the presence of micro-organisms. Colony stimulating factors are currently in clinical use for the treatment of neutropenia, both congenital and induced by cancer therapy. It has been suggested that there could be a prophylactic role for them in abdominal and genitourinary surgery, where infections are common.

Macrophages and lymphocytes have been shown to be present from day 1 in wounds, although lymphocytes are fewer in number than macrophages. In a study on human wounds by Martin and colleagues, macrophages peaked between 3 and 6 days and lymphocytes between 8 and 14 days. Thus they persist into the early repair phase of wound healing. Both macrophages and lymphocytes disappear from mature wounds by an unknown mechanism, but in abnormal scars both these cell types persist long afterwards. In hypertrophic scars, macrophage and lymphocyte levels have been found to be very high 4 to 5 months after wounding, and lymphocytes were still present at 40 per cent of the high level after 2 years. It has been suggested that control of lymphocytes might be a useful approach to control of scarring. It is of interest that minoxidil, a drug that has been shown *in vitro* to inhibit collagen lattice contraction, has been shown to inhibit DNA synthesis and the production of leucocyte migration inhibition factor by T lymphocytes.

Epithelial cells

Epithelial cells are important in the inflammatory phase as well as in the later repair aspect of wound healing. In partial thickness wounds, epithelial cells migrate from the edges of the wound and from the epithelial linings of hair follicles, sebaceous glands, and sweat glands and begin to proliferate. In full thickness wounds, only the epithelial cells at the edges of the wound are available to migrate, because of the destruction of dermal appendages, and closure takes longer. In sutured surgical wounds epithelial migration begins within the first 24 h of injury and may be completed as early as 72 h in healthy individuals.

Closure of the wound is not the only function of epithelial cells in the inflammatory phase (Fig. 9, Fig. 11). The development of techniques in molecular biology has led to unequivocal identification of many involved cytokines. Keratinocytes have been shown to produce the granulocyte/macrophage colony stimulating factor (GM-CSF), as well as TGF-a, fibroblast growth factor, and vascular endothelial growth factor (VEGF). They also produce interleukins 1, 3, and 6. IL-1 stimulates fibroblast proliferation and enhances the production of type I and III collagen mRNA and of an angiogenic factor. Thus they help to prepare and promote the next phase of wound healing. IL-6 induces the synthesis of proteins, some of which act to terminate the inflammatory phase.

By definition, chronic ulcers have a deficit in epithelialization. This could arise through reduced cell proliferation, or excess cell loss. Early studies of mitotic frequency at the edge of superficial ulcers failed to show any difference between those which healed expeditiously with treatment and those which did not. It is therefore probable that the surface extracellular matrix of such wounds governs the process of wound closure by forming an environment which may be either permissive of, or prohibitive for, epithelial cell adhesion and migration. The nature of these interactions remains relatively unexplored.

Formation of granulation tissue

Various chemotactic, growth, and activating factors produced in the inflammatory phase are concerned in the initiation and development of granulation tissue which lasts from about day 4 to day 21 after wounding. Granulation tissue comprises a loose matrix of fibrin, fibronectin, collagen, and glycosaminoglycans, particularly hyaluronic acid, containing macrophages, fibroblasts, and ingrowing blood vessels. In deep wounds, granulation tissue serves as a scaffold for new tissue ingrowth. In incisional wounds during this phase the wound begins to gain tensile strength, although it is during this early period that wound dehiscence and evisceration most frequently occur.

Fibroblasts

Fibroblasts migrate into the wound site 24 h after injury. During this initial phase of healing the fibroblasts are activated and undergo a burst of proliferative and synthetic activity, initially producing high amounts of fibronectin, and then synthesizing the other protein components of the extracellular matrix, including collagen and elastin, and glycosaminoglycans. The fibroblasts align themselves along the wound axis and form cell to cell links, which contribute to the contraction of the wound.

There has been much discussion about the type and origin of fibroblasts that appear in the wound. These fibroblasts have characteristics in between those of normal resting fibroblasts and smooth muscle cells. This altered phenotype, which has been called the ‘myofibroblast', is more mobile and more contractile than the inactivated fibroblast, and disappears on the completion of wound healing. Early distinctions between fibroblasts and myofibroblasts were based on ultrastructural criteria, but immunochemical analyses have, more recently, led to identification of subspecies of myofibroblasts based on permutations of expression of vimentin, desmin, and a smooth muscle actin. It has now been shown that smooth muscle cells in culture can reversibly modulate from contractile to synthetic cells, that is, the reverse of the myofibroblast development, and this may reflect changes occurring *in vivo.* In addition, it has been demonstrated that smooth muscle genes can be switched on transiently in certain circumstances by other non-muscle cells, including macrophages and some epithelial cells. It is still not known what controls the change in phenotype.

Complex factors influence the behavior of the fibroblasts in the formation of granulation tissue. Migration is promoted by TGF-b (produced by platelets and keratinocytes). Proliferation is promoted by thrombin, by serotonin (produced by platelets), by IL-1 produced by keratinocytes, by fibroblast growth factor from macrophages, and by epidermal growth factor. Synthetic activity is promoted by IL-1 and factor XIII for collagen, and by thrombin, epidermal growth factor, and TGF-b for fibronectin, while lysyloxidase activity is augmented by serotonin. Some remodelling of the extracellular matrix may take place at this stage. Degradative activity, which is also necessary for remodelling, is enabled by the promotion of collagenase synthesis by prostaglandin E_2.

Angiogenesis

Research into factors influencing angiogenesis has been directed at means of inhibiting new vessel growth in regard to tumour metastasis or, in the case of wound repair, means of stimulating angiogenesis to enhance healing.

Hypoxia following injury, if not so severe as to lead to tissue death associated with ischaemia, acts as a major stimulus for angiogenesis, which is required for restoration of blood flow. Along with fibroblast proliferation, neovascularization is a common feature of granulation tissue in the early phase of healing. One stimulus for new vessel growth is fibroblast growth factor, while other angiogenic factors, such as those secreted by macrophages and other cells, also contribute to the neovascularization. The growth of vessels in surgical wounds starts from capillary loops a few days after surgery, and vascularization may be complete in 6 to 7 days. In burns, development is later and may be complete in 12 to 16 days. The secondary wound (reopened and resutured) revascularizes at a significantly faster rate than a control wound which has not been reopened and resutured. This aspect of the importance of angiogenesis in wound healing has been observed in other types of wounds where differences in regional vascularity and healing are directly related.

The endothelial migration seen in granulation tissue is supported by the increased fibronectin in this tissue. Mitotic activity leads to the formation of capillary buds which sprout from blood vessels adjacent to the wound and extend into the wound space. There is a gradual establishment of flow. Endothelial cell proliferation is stimulated by a low wound oxygen tension (Po_2) in the early stages, but growth of vessels is later enhanced by a high wound Po_2, which is also essential for the synthesis of collagen necessary for the complete formation of the vessels. The pattern of vascular growth is probably the same in the healing of skin, muscle, and intestinal wounds. In fractured bones, vessel growth can be stimulated by repeated muscle contraction, which increases bone blood flow, while vascularization is reduced by immobilization.

Modulation of angiogenesis is currently a very active area of research; inhibition of vessel growth in tumours and its promotion in wounds are both appealing therapeutic strategies. Progress has been hampered by difficulty in quantifying the dynamic process of neovascularization without interfering with it. Laser-Doppler flow measurements have been found to correlate with vessel counts in experimental wounds, and may offer a non-invasive and repeatable method. However, conventional laser-Doppler techniques show large variations between readings even at adjacent points. Recently, a scanning laser-Doppler device has been developed, which overcomes this difficulty and which allows the imaging of blood flow in surface wounds. It has been used to study the evolution of blood flow over time in experimental and clinical wounds.

Re-epithelialization and basement membrane formation

Garlick has defined re-epithelialization as ‘the reconstitution of cells into an organised stratified squamous epithelium that permanently covers a wound defect and restores function'. Epithelial cells at the free edge of the wound begin to migrate across the provisional matrix of the granulation tissue. The advancing cells then become stationary and cells above and behind them then migrate over them onto the provisional matrix—a type of cell migration termed the ‘leap-frog' model. Epithelial cells behind the migrating cells shortly begin to proliferate to provide further cells for migration into the defect and for differentiation to form suprabasal epithelial layers. In a partial thickness wound epithelial cells will also migrate from the residual hair follicles. When cells moving into the wound cease to migrate they begin to reconstitute the basement membrane, which consists of a large number of adhesion complexes linked to a network structure. The cells establish the basement membrane by forming hemidesmosomes and depositing the protein components, such as collagen IV and laminin V. Subsequently, they secrete collagen VII, needed for the anchoring fibrils. The integrity of the basement membrane is essential for the stable association of epidermis to dermis and until the basement membrane has been entirely reconstituted the attachment of the new epidermis to the matrix remains tenuous and the new epithelium is easily sheared off by mechanical forces. Proteolytic enzymes produced in chronic wounds by colonizing bacteria or by persistent neutrophils from prolonged inflammation have been postulated to be additional factors which impair re-epithelialization.

Contraction

Wound closure by contraction, the inward movement of the edges of the injured tissue, is a normal part of the healing process. However, in some wounds, such as full thickness freeze injury, contraction does not occur. Wound contraction begins between 8 and 10 days after injury (Fig. 8). It is controlled both by the fibroblasts and by the extracellular matrix, and is due to the fibroblasts applying tension to the surrounding tissue matrix. *In vivo* it has been demonstrated that with contraction there is constant centrifugal tension. The rate of contraction has been shown to be constant for animals of a particular species or strain, and independent of the shape of wounds. However, there is marked interspecies variation. Contraction makes a much greater contribution to closure of full thickness wounds in rats than in man, which adds to the difficulty of extrapolating from experimental studies to the clinical situation.

Collagen—matrix formation and remodelling

Collagen is the major component by weight of the extracellular matrix of the skin, accounting for about 60 to 80 per cent of the dry weight of the tissue. There are known to be 19 different genetically distinct collagen types, 13 of which occur in human skin (Table 2). Collagen synthesis plays an important role in the early stages of healing and the formation of the granulation matrix. Production of collagen remains a major process in wound repair for several weeks after wound closure, and the major collagens continue to undergo remodelling for 2 years or more until the injured tissue is finally restored.

Collagen type	Chain composition	Tissue distribution
I	[α1(I)]$_2$α2(I)	Ubiquitous in most connective tissues, including skin, bones, tendons, ligaments, etc.
II	[α1(II)]$_3$	Skin, blood vessels, perichondrium in fetal tissues and in early wounds
III	[α1(III)]$_3$ or [α1(III)]$_2$α2(III)	Basement membranes, anchoring plaques
IV	[α1(IV)]$_2$α2(IV)	Ubiquitous
	[α1(V)]$_3$	
VI	α1(VI)α2(VI)α3(VI)	Extracellular microfibrils
VII	[α1(VII)]$_3$	Skin, fetal membranes
VIII	[α1(VIII)]$_3$	Newborn: throughout dermis
		Child and adult: papillary dermis
IX	[α1(IX)α2(IX)α3(IX)]	Reticular dermis in skin (proximal)
		Scalp: fan-shaped at base of hair follicles
X	[α1(X)]$_3$	Basement membrane zones of skin and other tissue
XI	[α1(XI)α2(XI)α3(XI)]	Synthesized by dermal fibroblasts; part of dermal extracellular matrix; not associated with fibrillar structures
XII	[α1(XII)]$_3$	Epidermis: transmembrane collagen of basal keratinocytes, component of hemidesmosomes
XIII	[α1(XIII)]$_3$	Basement membrane of epithelium and endothelium
XIV	[α1(XIV)]$_3$	Component of dermo-epidermal basement membrane zones. Cross-bridge between fibrils and other extracellular matrix molecules

Table 2 Collagen types present in skin (modified from Uitto, 1989)

Extracellular matrix

The extracellular matrix of tissues is composed of various polysaccharides and proteins and their complexes. These are secreted by cells *in situ* and different amounts and types are assembled to form diversely organized structures related to the functions of the particular tissue. The matrix not only serves as a support, but has a role in influencing the behavior of the cells in contact with it, affecting their development, migration, proliferation, shape, and metabolism, all of which are important with regard to wound healing.

The polysaccharides are glycosaminoglycans—long, unbranched chains of disaccharide repeating units. They fold with wide curvature in a random fashion and absorb large amounts of water, filling much of the extracellular space. Proteoglycans are formed by the combination of a number of glycosaminoglycan chains with a protein, and may contain up to 95 per cent (w/w) carbohydrate. Glycoproteins, on the other hand, are composed of short, branched oligosaccharide chains, containing from 1 to 60 per cent carbohydrate. The proteins of the extracellular matrix are principally structural proteins such as collagen and elastin, and adhesion proteins such as fibronectin, laminins, and integrins.

All collagen molecules are composed of three polypeptide a-chains, with a left-handed triple-helix configuration. The chains have about 1000 amino acids and have a distinctive amino acid composition of 33 per cent glycine and 20 per cent of the imino acids, proline and hydroxyproline, with a particular repeating trimeric sequence of glycine X Y, where either X or Y is often proline.

Synthesis of collagen

The synthesis of fibrillar collagens involves a progression in the combination of amino acids to form chains which associate to form molecules, and then association to form fibrils which aggregate into fibres or bundles. Fibroblasts are the major cell type to synthesize the dermal collagens. The first stages of synthesis take place intracellularly, to produce procollagen molecules which undergo activation stages in the extracellular space (Fig. 12). Some types of collagen molecules do not form fibrils, but rather networks or other structures and they may or may not be associated to fibrils. Many form part of basement membrane structures.

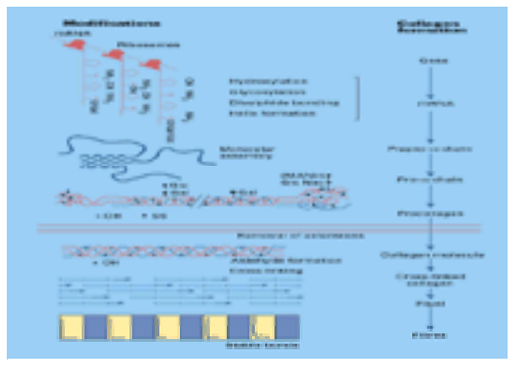


Fig. 12. Collagen biosynthesis: intracellular and extracellular stages; the double line represents the cell membrane; Glc, glucos; Gal, galactose; GlcNac, *N*-acetylglucosamine; SS, disulphide.

Intracellular synthesis

When the genes in the nucleus are activated, mRNAs specific for single polypeptide chains pass into the cytoplasm and are translated on the ribosomes of the endoplasmic reticulum, the three polypeptide chains being synthesized simultaneously (Fig. 12). The three a-chains may be identical (as in type III collagen), or a hybrid molecule consisting of two identical chains and a different third chain (as in type I), or three different chains (as in type VI). The molecule is a triple-chain molecule by the time it is detached from the ribosomes. The small, regularly spaced glycine residues situated in the central area of the chains allow them to pack tightly together to form the triple helix. The molecules then undergo post-translational modifications, principally the hydroxylation of a large number of the proline and lysine residues by the enzymes lysyl hydroxylase and 3- and 4-prolyl hydroxylase. Hydroxylation is a rate-limiting step for collagen secretion, and appears to be tightly regulated by tissue levels of oxygen and lactate. The hydroxy groups of hydroxyproline residues form interchain hydrogen bonds, which contribute to the stabilization of the triple-stranded helix. This procollagen molecule has extension non-helical peptides of 15 to 20 amino acids in non-collagenous sequences at both ends of the chains which are given a globular form by disulfide bonding. The procollagen molecule is then transported to the Golgi apparatus, enclosed in vesicles, and taken via the microtubules to the cell surface.

Extracellular synthesis

The processing of the procollagen molecules to collagen fibres takes place in the extracellular space. The molecules are activated by the cleavage of amino- and carboxy-peptide ends by amino- and carboxyl-propeptidases (Fig. 13). Lack of, or defects in, one of these enzymes results in defective fibres, for example type VIII Ehlers–Danlos disease lacking the aminoprotease. An important step in the assembly of collagen into fibres is the conversion of certain lysine and hydroxylysine residues to aldehydes by the extracellular enzyme lysyl oxidase and the formation of covalent bonds between the short, non-helical end of the collagen molecules, thus cementing the overlaps. The polymerization of many molecules in a staggered arrangement gives rise to the typical periodicity of 60 nm seen in electron microscope sections (Fig. 14) and this arrangement maximizes the tensile strength of the structure. At this stage type III fibrils have diameters of 40 to 60 nm and type I of 100 to 500 nm. The build-up of the peptides released in the transformation of procollagen to collagen inhibits collagen synthesis and thus provides a feedback for switching off the process of synthesis. The failure of this feedback system may be a contributory factor in excessive scarring.

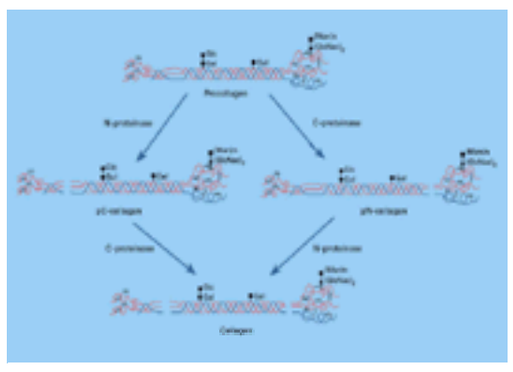


Fig. 13. Cleavage of the procollagen molecule to collagen. There is no obligatory sequence.

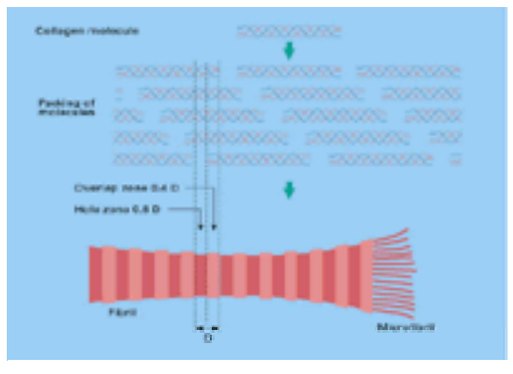


Fig. 14. Staggered packing of collagen molecules to form fibrils.

Cross-links

The aggregates of collagen molecules formed in the extracellular space then undergo cross-linking. The extent and types of cross-links, or ratios of types, vary from tissue to tissue, with age, and in disease. In skin, the collagen produced after injury is initially stabilized by cross-links derived from hydroxyallysine. During normal wound healing there are changes in the nature of the cross-links in the course of time to give a more stable structure. Both the type and the number of cross-links influence the stiffness of the tissue.

Abnormal numbers or types of cross-links can lead to malformation of tissue. The hydroxypyridinium cross-link prevalent in bone also occurs in hypertrophic scars, but is not found in normal skin or in normal mature scar tissue. Increased hydroxylation of lysine residues can occur in hyperglycemic states, and affects the types of cross-links as well as the number. In one study of patients with diabetes, the numbers of five different types of cross-link were shown to increase between three- and sixfold. Clinical assessment of hand contracture was associated with an increased number of cross-links, which was also correlated with the duration of diabetes and with skin changes.

The organization of the collagen in tissues is also influenced by the kinds and amounts of non-collagenous macromolecules that the cells secrete along with the collagen. In addition, the fibroblasts have a mechanical role in the assembly, crawling over the collagen, pulling and compacting it. The final architecture of the collagen network is related to its function. Thus collagen fibres in the papillary dermis are aligned in thin bundles almost perpendicular to the basement membrane and they hold up the dermal papillae. Some of the fibres glide into the loops of the anchoring fibrils attached to the basement membrane. In the reticular dermis the thick, undulating bundles are nearly parallel to the epidermal plane, and are connected by interlacing fibres, allowing the tissue to resist stress in all directions. In the hypodermis interlacing collagen fibres surround the adipocytes.

Types of collagen

Types I and III

Types I and III, the interstitial collagens, are the major types of collagen in skin, the rod-shaped molecules which provide its tensile strength. It is suggested that the fibrils are hybrids of types I and III, with type I present throughout the body of the fibril and type III round the periphery, and that type III plays a role in the regulation of fibril diameter. It has been possible to control fibre size *in vitro* by changing the ratio of type III to type I collagen, a higher proportion of type III producing thinner bundles. The ratios of type I to type III in normal human skin vary with age increasing from 0.8 to 1 in a 15-week fetus and 3.6 to 1 in a 3-month-old child up to 6 to 1 in adult skin. In healing wounds, the percentage of type III collagen in the early stages is higher than in normal skin, with type I collagen increasing as healing progresses, mirroring the changes during fetal development. There is a linear decrease in the total amount and density of collagen with age, and the amount of skin collagen is lower in females than males.

Ratios of the amounts of type I to type III collagen also vary in abnormal healing, being only 2:1 in hypertrophic scars, but 19:1 in keloid. Excessive levels of type III collagen may be associated with increased contraction. Levels are raised in the nodules and contractures of patients with Dupuytren's disease and in patients with diabetes the type III:I ratio is also high, and finger contraction occurs in some cases. Insulin treatment is known to cause expression of type III collagen in the mesangial matrix.

Type IV

Type IV collagen is a major component of the basement membrane in the dermal–epidermal junction to which basal keratinocytes attach preferentially. It is also produced by endothelial cells and forms an essential element of the microvascular wall. The procollagen molecules are not cleaved after secretion, they retain their propeptides and therefore their globular regions which associate ‘head to head', and with further lateral associations form a sheet-like polygonal mesh.

Type V

Type V collagen is ubiquitous in the dermis and interfaces between the cell surface and the surrounding matrix and associates with types I and III to form heterotypic fibrils.

Type VI

Type VI collagen is a molecule of short chain length and is homogeneously distributed throughout the dermis.

Type VII

Type VII collagen is the predominant component of the anchoring fibrils (Fig. 15), localized in the subbasal lamina of the dermal–epidermal junction. Discontinuities within the triple-helical domain allow the molecules to be flexible, aggregating in bundles of various diameters and degrees of curvature. The carboxylate terminals of the fibrils insert into the lamina densa and may extend into the dermis. These fibrils are the strongest mechanism for the adherence of the epidermis to the dermis, stronger than the attachment effected by the hemidesmosomes. They are absent or reduced in the skin of patients with recessive or dominant dystrophic epidermolysis bullosa; this leads to easy blistering.

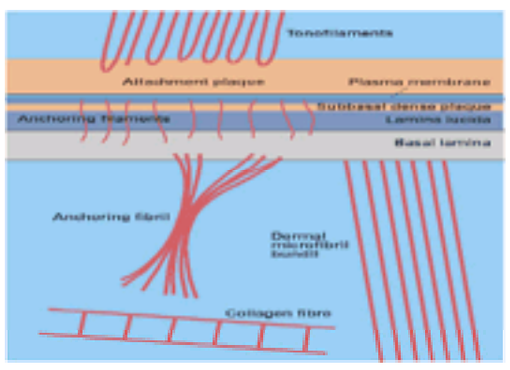


Fig. 15. Epidermal–dermal junction in the region of a hemidesmosome.

Types XII to XIX

Collagen types XII and XIV to XIX all occur in skin: XII, XIV, and XVI in the dermis; XVII (the bullous pemphigoid antigen) in the epidermis; XV, XVIII, and XIX in the basement membrane. The exact function of some of these collagens still needs to be elucidated, but there are indications that these may play an important role in the mechanical strength of the tissues and thus their importance in the repair process.

Collagenolysis

The remodelling of the extracellular matrix, from the early stages of wound healing to the maximal restoration of the injured tissue (which may be many months later) is a dynamic process. Degradation of collagen, as well as its synthesis, is continually occurring. The cells secrete collagenase and other proteases as well as collagens. In the extracellular space, procollagenase is transformed to collagenase which cleaves the helical portion of the collagen molecule, causing the fragments to unwind. The single-stranded polypeptides are then susceptible to degradation by other extracellular proteases and peptidases, or undergo endocytosis and are degraded by intracellular enzymes. Mature collagen may also be absorbed by phagocytosis. The peak levels of collagen phagocytic activity have been shown to correspond with the period when a change of configuration and fibre orientation was occurring.

In health, the synthesis of collagen and its degradation by collagenase and other proteases is finely regulated to maintain the optimal level of collagen. It has been said that ‘an unregulated protease in a biological system is like a loose cannon on a ship'. Imbalance between synthesis and degradation of collagen is associated with disease states, such as scleroderma, as well as alterations in wound healing such as hypertrophic scarring or ulceration. The failure of regulation also causes a problem in the formation of keloids, when the synthesis and breakdown of collagen are both stimulated, but the synthesis to a greater extent, thus leading to imbalance and overgrowth. Increased collagenolysis may occur even with normal levels of collagenase if the collagen itself is abnormal, for example collagen synthesized under hyperglycemic conditions, as in diabetes, is more susceptible to degradation and subsequent tissue breakdown. The remodelling of scar tissue also requires the degradation and synthesis of collagen.

Collagen and wound healing

In the adult, the normal repair of wounds occurs by the formation of granulation tissue and its organization to a scar. Scar is a dynamic, metabolically active tissue. Precise regulation of collagen metabolism during the repair process is exerted by cytokines (see below) and by the interactions of the extracellular matrix with fibroblasts. *In vitro*, fibroblasts in contact with collagen fibrils in a three-dimensional lattice show decreased production of types I and III collagen, but enhanced gene expression and activation of collagenase. Collagen synthesis is maximal between 14 and 21 days, although increased collagen deposition may be provoked to occur earlier by electric stimulation. After 21 days the rate of synthesis and the volume density of collagen in the wound return to the normal level. However, the tensile strength of the tissue continues to increase for a considerable time, up to 60 days or even 1 year. Wound tensile strength is a physical measurement which reflects the degree of intermolecular cross-linking, rather than collagen biosynthesis. Its increase is due to the formation of further cross-links and a change to more stable forms which lead to reorientation of the direction of the bundles. Breaking of the tissue occurs by disruption of the connection between fibres in the bundles and by slipping of molecules one over the other in the fibre. In normal wound healing, as the wound closes and the scar forms, there is a striking decrease in the cellularity between 3 weeks and 2 months after the injury with both fibroblasts and endothelial cells undergoing apoptosis, a type of programmed cell death in which the nucleus fragments and cytoplasmic organelles degenerate. The apoptotic cells are removed by phagocytosis either by macrophages or by neighbouring cells. If the cells within the

granulation tissue are not eliminated this can lead to pathological scarring, such as the development of hypertrophic scars and keloids, characterized by different patterns of high cellularity.

Fetal wound healing

In contrast to the healing by scar formation in the adult, the healing of fetal wounds up to the early third trimester of gestation proceeds without scar formation. Collagen is deposited more quickly in the fetal wound than in the adult, but is rapidly organized and is not excessive. It is thought that this might be due to glycosaminoglycans, which are also deposited. The nature and ratios of the glycosaminoglycans, which affect the cross-linking of collagen fibrils and the migration of fibroblasts, vary in different stages of wound healing. Hyaluronic acid, in particular, the content of which is high in the fetal wound matrix and which is found wherever there is tissue regeneration or repair, has been shown *in vitro* to facilitate the movement of fibroblasts. While hyaluronic acid is present only in the early stages in adult wounds, it is present throughout the process in fetal healing and the wound is closed by mesenchymal ingrowth onto the hyaluronate-enriched matrix. An *in vitro* study of the activity of a hyaluronic acid stimulating factor in sheep detected levels 10-fold higher in the fetal serum than in the serum of normal adults. In animal models there is variation in the open-wound healing response in the fetus of different species—wounds in sheep contracting, while those in the rabbit and monkey heal without contraction, as in the human.

The relevance of reduced scar formation in fetal wound healing and the potential application to clinical healing has been highlighted in recent work in Manchester by one of the authors (Ferguson) and colleagues. These workers have stated that fetal wound healing is characterized by a reduced growth factor profile and have demonstrated that neutralizing TGF-b activity by antibodies results in decreased wound scarring in adult rats. The clinical application of this work is described subsequently in the section on anti-growth factor therapy.

Wound remodelling—pharmacological control

One of the problems with scar tissue is that it tends to remain somewhat weaker and more brittle than previously unwounded tissue. A second problem is that it tends to contract abnormally. A third problem in certain people is that of overhealing, leading to hypertrophic scars or keloids (Fig. 16, Fig. 17). Some clinical differences between keloids and hypertrophic scars are listed in Table 3. While some researchers consider that there is spontaneous regression of hypertrophic scars eventually, others maintain that many such scars persist indefinitely. Biochemical differences between hypertrophic scar, keloid, and normal skin have also been characterized. Keloids have a greatly increased proportion of type I to type III collagen compared with normal skin, whereas in hypertrophic scars the ratio is lower than normal. Glycosaminoglycan contents are also abnormal in hypertrophic scars. One study determined the level of hyaluronic acid in hypertrophic scar to be less than half that of normal skin, while the level of chondroitin-4-sulphate was six times higher, as was also the case in granulation tissue. Either the chondroitin sulphate continues to be formed in the hypertrophic scar or it is not removed.



Fig. 16. Hypertrophic scar on a young child.



Fig. 17. Keloid scar.

Hypertrophic scars	*Keloids*
Develop soon after surgery	May not begin for many months
Usually subside with time (maturation)	
Limited boundary	Overgrow their boundaries
Size commensurate with injury	Minor injury may produce large lesion
Occur with motion (compression)	Independent of motion
Usually occur across flexor surfaces (joints, abdomen, etc.)	Areas of high predilection (ear lobes, presternal skin), rarely across joints
Improve with appropriate surgery	Often worsened by surgery

Table 3 Hypertrophic scars and keloids (from Brody 1981)

One reason for the imperfect characteristics of scar tissue is that the newly formed collagen pattern in a wound is abnormal, and that rapid intermolecular cross-linking while fibres are being formed leads to their being irreversibly fixed. The use of scanning electron microscopy to study collagen morphology led to the hypothesis that if the cross-linking could be temporarily delayed or slowed, a more nearly physiological collagen pattern would have time to develop.

Attempts to delay or inhibit the cross-linking of collagen have met with some success, for example b-aminoproprionitrile and other nitriles inhibit lysyl oxidase, the enzyme that deaminates the lysyl e-amino group to aldehyde, an early step in the formation of intermolecular cross-links. d-Penicillamine reacts with aldehyde groups and so blocks cross-linking of newly formed collagen, but it may also make existing cross-links more labile. Improvement has also been effected using corticosteroids which inhibit fibroblast migration and proliferation and collagen synthesis. Compression has also been of benefit in some cases.

As described in the introduction, Silastic gel is being used in the treatment of abnormal scars and is now licensed for use in the United States. Interferon-g, which suppresses collagen synthesis *in vitro* and in animal models *in vivo*, leads to improvement in keloids in a small group of patients. An *in vitro* study has shown that calcium antagonists reduce the rate of collagen synthesis and increase that of collagenase in a connective tissue equivalent. Lee and colleagues subsequently reported that treatment of over 30 patients with verapamil in a plastic surgery scar clinic in Chicago led to great improvement of hypertrophic scars in a number of patients after four or five injections. This group is now trying combinations of this treatment with conventional modes of therapy. Cryosurgical treatment has also led to

considerable improvement in one group of 45 patients. Increased understanding of the regulation of scarring has led to the suggestion that the inhibition of helper T cells by cyclosporin or compounds with similar properties should be investigated. A further approach to the control of different phases of wound healing, including hypertrophic scar and keloid formation, is that of influencing cell-matrix adhesion by using monoclonal antibodies or synthetic peptides.

Experimental wound-healing models

A wound is a very complex biological system and detailed studies of repair processes and the factors that influence them require both *in vivo* and *in vitro* models. An *in vitro* model is more of a closed system, allowing the isolation of a tissue and investigation of modulation of the system without the complication of other factors. While *in vitro* models are only partial, and cannot be extrapolated directly, they do allow perceptions which may be valid *in vivo.* They can also allow the study of human tissue or cells. Extrapolation from animal studies to the human situation may not be valid. For example, metabolism may be different; metabolites produced from drugs *in vivo* vary in different species; all other mammals except the guinea pig and primates, including humans, can synthesize ascorbic acid.

There are *in vitro* models for many of the various processes involved in wound healing: cell proliferation and differentiation, cell migration, angiogenesis, re-epithelialization, wound contraction/ contracture, and matrix remodelling. The fibroblast-populated collagen lattice (**FPCL**) provides a simple model for connective tissue which can be modified in various ways. An acidic type I collagen solution polymerizes *in vitro* when raised to physiological pH and ionic strength, forming a three-dimensional gel. If fibroblasts are incorporated into the polymerizing solution, they organize and align collagen fibrils and compact them, causing the lattice to contract. The three-dimensional FPCL provides a more '*in vivo* like' ambiance for the fibroblast than does monolayer culture and the metabolism of the fibroblasts more closely resembles that of the cells *in vivo.* Such a lattice has been used as an *in vitro* model for connective tissue contraction and scar contracture in wound healing. The degree and speed of contraction of the lattice are related to the cell density, to the concentration of collagen and of serum, and also to the type of cell involved, and to the presence of drugs or other factors.

Tamoxifen, a drug used extensively in the treatment of breast cancer, has also been used with some success at similar doses for treating fibrotic benign tumours. *In vitro* studies have shown that tamoxifen inhibits the proliferation of fibroblasts from keloid biopsies and decreases their rate of TGF-b production and collagen synthesis in monolayer culture. A recent *in vitro* study using the FPCL model has shown that tamoxifen, at concentrations achieved in tissues by oral clinical doses, inhibits the contraction of collagen lattices in a dose-dependent manner (Fig. 18). The inhibition of contraction is reversible at certain doses after being washed out at different periods of time. The inhibition correlates with the morphology of the cells, which need to be elongated to induce contraction. Another drug which has been shown to control FPCL contraction is minoxidil which also alters the morphology of the cytoskeleton in monolayer culture of fibroblasts (Fig. 19). The results indicate that both of these drugs might be of use in the clinical situation to control remodelling of the extracellular matrix in overscarring tissue (see later section on TGF-b and scarring).

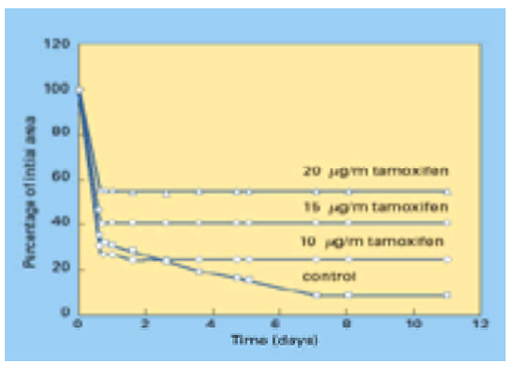


Fig. 18. Dose-dependent inhibitory effect of tamoxifen on collagen lattice contraction by human skin fibroblasts.

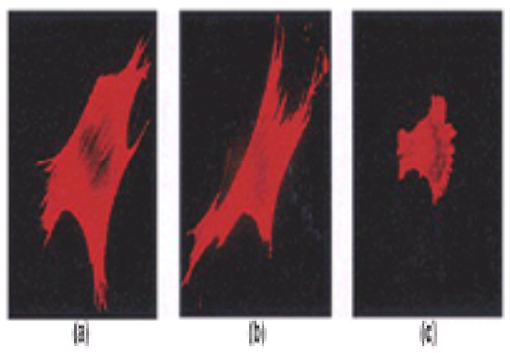


Fig. 19. Morphology of normal human skin fibroblasts in a monolayer, demonstrating actin fibres stained with rhodamine-phalloidin (x 400). (a) Control; (b) minoxidil, 100 µg/ml — little change in morphology, reflecting minor inhibition seen in gel contraction with this dose; (c) minoxidil 800 µg/ml — marked alteration in structure correlating with complete inhibition of gel contraction at this dose, but reversible up to 2 days.

Another study of the collagen lattice model showed the contraction of type III collagen lattices to be faster and to a greater degree than type I hydrated collagen lattices with identical numbers of cells. This could well be correlated with the high type III:I ratios in patients with Dupuytren's contraction and those with diabetes, certain of whom suffer hand contracture. In another *in vitro* study using hydrated collagen lattices, fibroblasts from patients with Dupuytren's disease showed a consistently higher contractility than normal skin fibroblasts.

Keratinocyte migration is a complex physiological process greatly influenced by the nature of the surrounding extracellular matrix as well as by other factors. One of the problems encountered in chronic wounds is that, even after the formation of healthy granulation tissue, keratinocytes fail to migrate from the wound edges to re-epithelialize the lesion. There are no standard *in vitro* models for chronic wounds. However, a scratched monolayer may serve as a simple model of re-epithelialization for initial investigation or to study migration of other cell types. Reproducible straight scratches are made in a confluent monolayer and the migration of cells across the denuded area to 'heal' the wound is monitored by video and can be quantitated (Fig. 20). A similar technique has been useful in examining some of the mechanisms of endothelial cell repair.

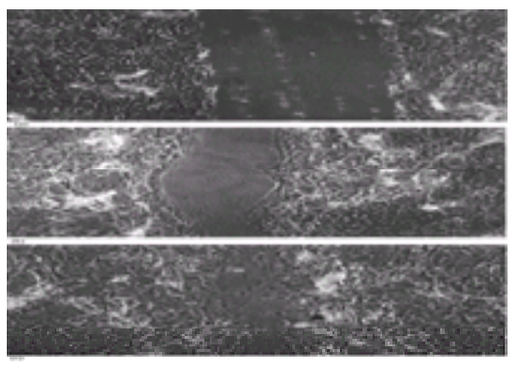


Fig. 20. Scratch technique: effect of Eupolin extract from *Chromolaena odorata* on migration of human keratinocytes (× 100). (a) before therapy. (b) Control at 34 h, 50 per cent closure of 'wound'. (c) In 10 µg/ml Eupolin extract at 34 h, 'wound' is almost closed.

Regulation of wound healing by growth factors

Growth factors are peptide signaling molecules. They play a critical role in the orchestration and integration of all cellular and molecular events during wound healing. They affect all cells involved in the wound healing process: for example inflammatory cells, keratinocytes, endothelial cells, and fibroblasts, and both individually and collectively exert selective effects on such cells, including stimulation or inhibition of cell division, migration, differentiation, and extracellular matrix synthesis/degradation. As such, growth factors play a pivotal regulatory role in the wound-healing cascade and have been the subject of intensive research. Such studies have included investigations of the distribution of various growth factors and their receptors at various times during wound healing: the effects of exogenously adding or deleting such growth factors on wounds in experimental animals and also some limited human studies, predominantly involving growth factor addition. Pharmaceutical interest in the role of growth factors has been bolstered by the fact that these molecules can be made using recombinant DNA technology and consequently they represent potential therapeutic agents for topical or systemic administration to augment the wound healing process. Furthermore, as growth factor receptors and signaling mechanisms have been elucidated, these also have become the target for pharmaceutical intervention, for example in the development of selective receptor or second messenger agonists or antagonists. These extensive preclinical studies have the potential to develop into new human pharmacological therapies for controlling the speed and quality of wound healing in man.

In general, growth factors are small peptides, secreted or stored by various cell types. There are literally hundreds of different growth factors. These fall into several multigene families such as transforming growth factor-bs (TGF-bs), platelet-derived growth factors (PDGFs), fibroblast growth factors (FGFs), insulin-like growth factors (IGFs), epidermal growth factors (EGFs), colony stimulating factors (CSFs), and hepatocyte growth factors (HGFs). Within each family there are often multiple isoforms and family members, for example the TGF-b family consists of well over 30 members including TGF-b1 to 3, bone morphogenetic proteins, and activin, all of which are involved in the wound healing process. Furthermore, an individual isoform of a growth factor may be subject to differential gene splicing or different subunits assemble in different combinations thereby adding further diversity to this already large and complex system. All of the growth factor families mentioned above are involved at some stage in the wound healing process.

The large numbers of growth factors involved in the wound healing process are matched by an equally large number of growth factor receptors present on the surface of target cells. Upon binding the growth factor these receptors signal, usually via an intracellular second messenger system which involves the phosphorylation of tyrosine or serine kinases, sparking a series of intracellular signaling events that end up with a nuclear signal, such as the translocation of a transcription factor to the nucleus. Even though individual growth factors signal through specific cell surface receptors, there is often integration of this signaling via part of the second messenger system. Thus regulation of the transcription factor nuclear factor Kappa-B through the inhibitory Kappa B cascade is a common focus for many proinflammatory cytokines, whilst members of the TGF-b and HGF families converge on the intracellular second messenger SMAD 7.

It should be evident, therefore, that in a wound there are many different growth factors present simultaneously and these may signal in a co-ordinated fashion in parallel at several receptors on the cell surface. *In vitro* culture experiments usually determine the effects of individual growth factors. Often these data do not translate well into the *in vivo* situation where there is a complex mixture of cytokines signaling in parallel. Furthermore, in tissue culture, cells are often in a ‘wounded' environment, including loss of cell contact and high serum levels, and they respond by upregulating their endogenous synthesis of growth factors and growth factor receptors, for example TGF-b 1 and the type 2 TGF-b receptor, which in turn distort the results. A consequence of this is that direct animal experimentation and early human clinical studies are required to elucidate fully the role of growth factors in wound healing and the effects of their experimental addition or deletion.

This complexity in growth factor signaling is enhanced by the observation that many growth factors are secreted in an inactive form as a precursor molecule, such as the latent TGF-b complex, which is then stored either intracellularly or by binding to extracellular matrix molecules such as fibronectin. Activation, for example by protease cleavage, occurs immediately upon wounding. This activation of stored growth factor is as important a regulatory step as transcription, translation, and secretion of newly synthesized growth factor. Several growth factors, such as TGF-b and PDGF, are stored in the granules of platelets and released immediately upon platelet degranulation during clotting. Other growth factors are stored in the granules of keratinocytes of the superficial layers of the skin and their immediate release from such storage is very important in the restitution of skin barrier function (by rapidly affecting the lipid composition in the membranes of adjacent keratinocytes), following minor trauma, such as a scratch, as well as being released during normal wound healing. Several growth factors, such as TGF-b1, bind to extracellular matrix molecules, for instance fibronectin or fibrin, and thus represent a stored reservoir whereby secretion of the factor from a cell at one point in time can affect cells later. The system is made even more complex by the observation that many growth factors are associated with binding proteins which may either render the growth factor inactive or potentiate/augment its action when bound to a cell surface receptor. Members of the insulin-like growth factor family are associated with an equally large family of insulin-like growth factor binding proteins and the effects of the growth factors are often determined more by the biology of the binding protein than by the growth factor itself. Interestingly, exogenous application of insulin-like growth factors, when associated with an appropriate binding protein, is much more effective than exogenous application of the growth factor alone. Indeed, exogenous application of the binding protein alone can be an efficacious therapeutic strategy, as it sequesters and presents to the receptors natural free growth released during the wound healing process.

With this complexity in mind it will be clear that early studies of growth factor activity in wound healing which attempted to ascribe one set of cellular functions, such as keratinocyte migration or extracellular matrix biosynthesis, to a single growth factor were naive. It is often asked ‘why are there so many growth factors involved in the wound healing process?' This question is best thought of in an evolutionary context. Wound healing is a critical evolutionary function: if animals or humans do not heal their wounds properly they rapidly die, for example from bleeding or septicemia. Consequently, evolution has selected for a highly redundant process, whereby many factors have overlapping functions, thus ensuring that minor variations or disturbances in any one part of the system do not have a catastrophic effect. In brief, evolution has selected for speed of wound healing under dirty conditions. Furthermore, in protein evolution it is clear that new molecules arise by gene duplication and subtle alterations to the structure of such molecules. A key feature in this process is that the new molecules generated are compatible with the rest of life, but may assume a specific function in a specific tissue at a certain time under a specific set of circumstances. This contributes to the redundancy of signaling. Thus a proper appreciation of the role of growth factors in wound healing can be gained from a mental paradigm which accords as much importance to the timing of the appearance or disappearance of growth factors during wound healing and their effects in association with other molecules as it does to their individual distribution and effects in isolation.

Figure 21 illustrates the various ways in which growth factors may signal and co-ordinate cellular events during wound healing. These vary from long range endocrine effects, such as systemic levels of TGF-b1 affecting the wound healing process, through to shorter range cell-to-cell signaling or even signaling within a single cell. Furthermore, the second messengers activated within the cell upon growth factor signaling interact with other intracellular signalling mechanisms such as those derived from integrin binding to extracellular matrix molecules. Even at the cell surface there are complementary effects, for instance binding of cells via their integrin receptors to extracellular matrix molecules such as fibronectin can cluster growth factor receptors at these focal contacts, thereby enhancing their interaction with matrix-bound growth factors.

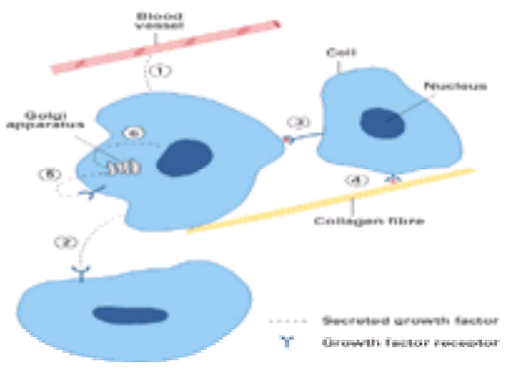


Fig. 21. The principal methods of growth factor signalling: (1) endocrine (release of the factor from the cell into the bloodstream to affect a distant target); (2) paracrine (release of the factor from one cell to affect a neighbouring cell); (3) juxtacrine (the factor is secreted and bound to the surface of one cell and binds to a receptor on an immediately adjacent cell without release of the factor from the cell surface); (4) matricine (a growth factor bound to the extracellular matrix, such as fibronectin, binding to a receptor on the cell surface); (5) autocrine (release of the growth factor from a cell which binds to a cell surface receptor on the same cell); (6) intracrine (synthesis of a growth factor in the cytoplasm of the cell which is then translocated intracellularly, for example to the nucleus, where it directly affects gene transcription without being secreted from the surface of the cell).

Such complexity often gives rise to apparently paradoxical actions of growth factors. Thus the effects that a particular growth factor may signal should be seen in the context of the cells' environment. In essence all growth factor biology must be viewed in context. Thus, for example, small amounts of TGF-b1 are chemotactic to monocytes and macrophages, causing them to be attracted to the wound area early in the wound healing cascade. However, higher amounts of TFG-b1 inhibit the migration of more mature monocytes and macrophages. Thus it can be seen that in two different contexts, at two different times during wound healing, the same

growth factor TGF-b1 is either proinflammatory or anti-inflammatory. Such self-regulating feedback loops are typical of growth factor biology. They normally act physiologically. Thus the enzyme plasmin is a key activator, releasing TGF-b1 from its precursor molecule. The binding of active TGF-b1 to its cell surface receptor rapidly induces the plasminogen-activator inhibitor system, which is turn shuts down TGF-b activation. These feedback loops are particularly important in the interaction between growth factors and extracellular matrix molecules, which may be thought of as part of the same cell-signaling mechanisms.

Thus the response of a particular cell to a particular growth factor can depend critically upon which extracellular matrix molecule it is bound to and upon the shape of the cell. Furthermore, growth factors such as TGF-b1 can induce the synthesis of specific extracellular matrix molecules such as fibronectin which in turn can affect the response of the cell to such growth factors. Extracellular matrix molecules often contain growth factor-like sequences within their sequences, for example tenascin has epidermal growth factor-like sequences and these may act *in situ* with growth factor-like effects. Particular extracellular matrix molecules such as decorin may bind to specific growth factor receptors such as the epidermal growth factor receptor and signal, whilst other growth factors such as bone morphogenetic proteins may act as enzymes for the processing of extracellular matrix molecules such as collagen. It is clear therefore that there are extensive interactive loops in growth factor and matrix biology. Whilst these operate in a physiological context during normal wound healing, aberrations in such growth factor loops often give rise to pathology. Thus it is thought that an inappropriate feedback loop between the TGF-b family, its receptors, and the synthesis of extracellular matrix molecules may underlie pathological fibrotic disorders including hypertrophic scars and keloids.

Most growth factors present during wound healing are rapidly inactivated either by binding and internalization at a cell surface receptor or by being sequestered by binding to an extracellular clearance protein such as a_2-macroglobulin. Growth factor signaling is therefore very short lived. Additionally, proteases present at the wound site may degrade peptide growth factors before they even bind to their receptor. This loss of specific sets of growth factors during wound healing may be as important a regulatory process as the release of the growth factors themselves. Thus, for example, in the resolution of the granulation phase of wound healing where there is widespread cell death (apoptosis), particularly of inflammatory and endothelial cells, death pathways are probably triggered by the loss of specific growth (maintenance) factors from the wound site.

Most growth factors are thought of as signaling in a local context. However, there are data emerging which indicate that systemic levels of growth factors may also be important. Thus, for example, systemic administration of TGF-b1 24 h before wounding causes a major acceleration of the speed of subsequent healing. The mechanism of such systemic effects may include priming of inflammatory cells to respond to the signals released upon wounding. Furthermore, endogenous systemic levels of growth factors may also be important. There are often genetic polymorphisms in growth factor promoters, giving rise to high and low responders.

TGF-b1 and scarring

There are data, for example, which indicate that such polymorphisms for instance in TGF-b1 may be important in the severity of subsequent scarring or in the fibrotic response, such as that following organ transplantation. Furthermore, interesting data indicate that systemic hormonal levels, for instance oestrogens, may regulate specific growth factors such as TGF-b1. Thus postmenopausal women with low levels of oestrogen, also have low levels of TGF-b1 and their wounds heal more slowly, but with a better quality of scarring. Systemic or local oestrogen application results in an elevation of TGF-b1 levels and an acceleration of wound healing, although with a deterioration of scar quality. Thus there are racial, sexual, and individual variations in the levels of specific growth factors and their receptors. Such variations may explain, for example, the predisposition of African and Mongoloid races to hypertrophic scars and keloids. Furthermore, within an individual subject there is also intense variation in growth factor levels and responses with age. In embryonic and fetal wounds, the distribution and activity of growth factors are very different from those in adult wounds, largely as a result of a decreased inflammatory process. Consequently, embryonic wounds heal perfectly with no visible scarring. During childhood, when the immune and inflammatory responses are at their peak, the levels of growth factors within wounds are high and the wounds heal very quickly but with a poor cosmetic result, that is, bad scarring. In old age the profile of growth factors at the wound site again changes, for instance there are decreases in TGF-b1 and elevations of TGF-b3, in part determined by the altered inflammatory response with ageing and in part by hormonal effects. Consequently, in old age wounds tend to heal more slowly but often with a better quality of scarring.

Acute wounds

In an acute incisional dermal wound, growth factors such as PDGF, TGF-b1, TGF-a, and fibroblast growth factors are released immediately upon wounding by the release of stored growth factor from degranulating platelets and injured keratinocytes. These released growth factors have immediate effects, such as attracting inflammatory cells, predominately neutrophils, monocytes, and macrophages to the wound site, which in turn release further growth factors. They also regulate vascular permeability, e.g. vascular endothelial growth factor (VEGF) and re-epithelialization e.g. keratinocyte growth factor. Many of these stored growth factors, such as TGF-b1, also bind to the fibrin clot and thus represent a slow release depot of growth factors during the early phases of wound healing. Manipulation of growth factor levels, such as exogenous addition or deletion, is often best effected during this very early phase of wound healing when there are few autoregulatory and interacting loops of growth factor action. Thus the timing of manipulation is important. The early wounding event also stimulates the synthesis and release of growth factors from surrounding cells, predominantly inflammatory cells of the monocyte and macrophage lineage, but also from neutrophils and fibroblasts. Such growth factors may be induced by hypoxia, for instance the VEGF promoter has a hypoxia response element, by alterations in pH, by the formation of oxygen free radicals, or by changes in intracellular calcium levels following cell permeabilization on wounding. These growth factors then serve to orchestrate the subsequent events during wound healing including initial stimulation of the influx of inflammatory cells and their death at the end of the granulation phase; stimulation of fibroblast proliferation, migration, and extracellular matrix synthesis; and stimulation of keratinocyte division and migration.

Growth factors such as the epidermal growth factor family or keratinocyte growth factor are important in stimulating re-epithelialization, the FGF family and VEGF in stimulating angiogenesis, whilst the TGF-b and PDGF families are important in stimulating fibroblast migration and extracellular matrix deposition. However, as explained earlier, the actions of individual factors must be viewed in the context of other factors or matrix molecules and the state of differentiation of the cell.

Experimental studies have investigated the distribution of growth factors and growth factor receptors both in experimental animal wounds and in human biopsies. These often show evidence for paracrine signalling mechanisms, for instance synthesis of the growth factor by fibroblasts (e.g. keratinocyte growth factor), but of the receptor by epithelial cells (e.g. the keratinocyte growth factor receptor). Experimental animals have also been used extensively to investigate the role of growth factors in wound healing: by exogenous addition of the growth factor; by exogenous deletion using neutralizing antibodies, antisense oligonucleotides, or ribozymes; or by alteration of the effects of the growth factors using transgenic dominant negative growth factor receptors or transgenic animals that over- or underexpress the growth factor. These experimental studies have given considerable insight into the role of various growth factors during wound healing. To date there have been few human studies where exogenous growth factors have been administered to acute, for instance surgical, wounds.

Chronic wounds

There have also been extensive investigations of the distribution of growth factors in chronic wounds such as diabetic ulcers, venous ulcers, and pressure sores. It was once believed that such chronic wounds were deficient in growth factors. That view is no longer tenable. Instead, chronic wounds exhibit a chronic inflammatory phenotype and both the distribution and composition of growth factors and extracellular matrix molecules are different in chronic wounds from those in acute wounds. Thus chronic wounds often have high levels of growth factors typical of a chronic inflammatory/infective response such as tumour necrosis factor-a or interferon-g. They usually exhibit a high turnover pathology, for example rapid synthesis and degradation of extracellular matrix molecules and growth factors, rapid cell proliferation, and death. Thus the pathology of a chronic wound is not a simple growth factor deficiency and exogenous applications of growth factors to stimulate chronic wound healing must be seen in this biological context. It is likely that effective therapies for stimulating healing of human chronic wounds will involve therapeutic manipulation of the profile of growth factors present, such as adding certain growth factors and neutralizing others. Diabetic ulcers, for example, show a persistent absence of TGF-b1 and a high level of TGF-b3; these observations may explain in part the failure of diabetic wounds to heal. In addition, all chronic wounds have an underlying systemic pathology. Thus, in venous ulceration there is often evidence of repeated ischaemia–reperfusion injury and complex extracellular matrix cuffs around blood vessels which bind active growth factors, whilst in diabetic ulcers there are major problems with atherosclerosis and microthrombi in the microcirculation. The activity of growth factors in such chronic wounds may also be different, for instance being rapidly activated by altered wound pH and/or hypoxia, or degraded by wound proteases or activated/inactivated by free oxygen radicals.

Clinical studies

The small number of clinical studies using exogenous recombinant growth factors to stimulate chronic wound healing have been disappointing. In part, this is probably due to naivety about the biology of human chronic wound healing (there are no good experimental animal models) and in part to other factors such as trial design and delivery. Delivery of exogenous recombinant growth factors to human wounds is a particularly critical problem. The chronic wound environment is characterized by high levels of proteases, which rapidly degrade matrix molecules such as fibronectin, elastin, and collagen and will very rapidly degrade exogenously added peptides. The choice of delivery vehicle is therefore crucial. Protection of the growth factor so that it gets to its target, often deep in the wound, is also vital. The development of future growth factor delivery vehicles needs to address such issues and these treatments may be combined, for example, with wound debridement, either surgical or enzymatic. Human clinical trials of any agent to treat chronic wounds, including recombinant growth factors, show a significant placebo effect, as many ulcers will heal when the underlying pathology is treated correctly, for example stable glucose control in patients with diabetes or effective compression bandage therapy in the case of venous ulcers. Most clinical trials of growth factors treat all chronic wounds, for instance venous ulcers, as having the same pathology, whereas in fact there are probably different pathologies and molecular problems in varying subsets of such chronic wounds, yet to be defined. Single growth factor therapy is likely to depend

critically on the timing of application in order to achieve effective therapeutic results. Exogenous manipulations using combinations of growth factors, or exogenous growth factors together with agents which neutralize other growth factors, are likely to be biologically more relevant than the manipulation of single growth factors.

Despite early disappointing results, partly as a result of naivety about the complexity of the system, particular recombinant growth factors have now been approved for commercial sale and clinical usage to stimulate human wound healing, for example recombinant PDGF, so future prospects appear promising.

Anti-growth factor therapy

Anti-growth factor therapy may turn out to be as important in man as exogenous addition of growth factors themselves. For example, it has been shown that addition of neutralizing antibodies to TGF-b1 and -b2 or addition of exogenous TGF-b3 are potent antiscarring therapies. Scarring is a major problem following wound healing, where it often causes adverse functional, growth, and cosmetic effects, particularly in children. Prevention of scarring, even following acute surgical wounds, would be an enormous therapeutic advance. It is also important in preventing hypertrophic scars, keloids, and other abnormal forms of scarring, particularly following burn injury. Scarring of tissues other than the skin is also very important, thus corneal scarring can lead to blindness; glial scarring within the central nervous system can prevent neuronal reconnection, for instance following trauma or neurosurgery; whilst scarring in the form of strictures and adhesions in the abdominal and pelvic viscera can be life-threatening. In experimental systems, neutralizing TGF-b1 and -b2 or elevating TGF-b3 levels prevents such scarring: such neutralization may be by neutralizing antibodies, antisense oligonucleotides, ribozymes, receptor blocking molecules, etc.

In addition to the neutralizing strategies cited earlier, the activation of TGF-b1 can be inhibited by competing for binding of the precursor latency associated peptide molecule to the mannose-6-phosphate receptor (a key event in TGF-b activation) using exogenously applied mannose-6-phosphate. Extensive antiscarring animal studies in tissues such as the skin, eye, brain, and viscera have now led to several of these potential therapies entering clinical trials as antiscarring agents, for example human recombinant phage neutralizing antibodies to TGF-b1 and -b2, mannose-6-phosphate, and recombinant TGF-b3. It is highly likely that, within the next few years, we will be able to manipulate acute wound healing and mimic the more regenerative form of scar-free healing observed in embryos. This would have major implications for contemporary surgery, not only in giving a better cosmetic and functional result, but also in altering the various surgical approaches, particularly in exposed sites such as the face. Furthermore, whilst previous research on growth factors has focused on their use in chronic wounds, it is highly likely that their most effective use will be in acute surgical wounds because, as explained earlier, the timing and context of their application are paramount in eliciting specific effects. Surgery is trauma by appointment and, in contrast to chronic wounds, the timing can be carefully controlled. Thus pharmacological manipulation of acute wound healing by the surgeon, either to accelerate the speed of healing or improve the quality, is likely to become a major part of the surgical armamentarium in the next few years.

Healing deficits

In animals, a wound-healing deficit can be defined artificially and the impairment in healing can be successfully ameliorated by growth factors. However, impaired acute healing in animal models bears an indeterminate relationship to the common human problems of chronic wound healing failure. In essence, it has not yet proved possible to create good analogues of venous, arteriopathic, or decubitus ulcers, if only because laboratory animals do not live long enough to achieve the tissue changes seen in some of these conditions in man.

Steroid impaired healing

The wound-healing deficit induced by the use of corticosteroids is well documented. In steroid-sensitive animals, treatment causes a prolonged monocytopenia, preventing macrophage migration into the wound and thus diminishing this essential element of the wound-healing cascade. The *in vitro* effects of steroid treatment are to depress fibroblast proliferation and inhibit procollagen and matrix protein synthesis. A single local application of TGF-b_1 (10 to 40 pmol/wound) in a collagen suspension in the rat reverses a 50 per cent wound-healing deficit resulting from methylprednisolone treatment. PDGF in the same model fails to reverse the wound-healing deficit, but does increase fibroblast numbers in the wound. However, these fibroblasts lack enhanced expression of procollagen type I. Wound macrophages remain absent from both PDGF- and TGF-b_1-treated wounds. In steroid-treated pigs, local applications of exogenous TGF-b_1 reverse the depression of matrix protein synthesis, procollagen mRNA, and TGF-b_1 mRNA. In methylprednisolone-treated rats, daily injection of 5 mg epidermal growth factor in an albumin carrier into a polyvinyl alcohol sponge restores collagen and matrix protein levels to normal.

Irradiation

Local irradiation impairs wound healing by depleting dermal fibroblasts and decreasing the proliferative potential of endothelium, whereas total body irradiation depresses bone marrow-derived elements, virtually eliminating wound macrophages. A single application of 2 to 10 mg of PDGF in a collagen vehicle partially reverses the surface irradiation wound-healing deficit in the skin of rats 7 days after creation of the wound. PDGF does not reverse the wound-healing deficit seen with total body irradiation. These studies support the hypothesis that PDGF requires the presence of activated wound macrophages for activity *in vivo.*

Cytotoxics

Cytotoxic treatment decreases circulating white cells and impairs the formation of granulation tissue in a wound chamber. Adriamycin treatment reduces the level of TGF-b and PDGF-like activity in aliquots of wound fluid removed from wound chambers. Injection of TGF-b into the wound chamber returns granulation tissue formation to a normal level. Incisional wound healing is impaired by Adriamycin treatment, and the strength of wounds can be returned to normal by a single intraincisional application of 2 mg of TGF-b or 50 mg tumour necrosis factor-a in a collagen vehicle.

Diabetes

Between 5 and 10 per cent of all patients with diabetes have foot ulcers, many of them with full-thickness plantar ulcers which often lead to amputation. It has been stated that diabetes accounts for up to 50 per cent of non-traumatic leg amputations. Patients lose the ability to synthesize normal tissue matrix proteins. In addition, the microcirculation is altered in such a manner that, following injury, patients with type 1 insulin-dependent diabetes fail to illicit a hyperaemic response, resulting in impaired neovascularization, which is essential in the healing process. In implanted Hunt–Schilling wound chambers in diabetic rats, PDGF restores granulation tissue formation and angiogenesis to normal. The influx of connective tissue cells, DNA synthesis, and collagen deposition are increased, effects that are augmented by the addition of insulin to PDGF. In diabetic mice, 0.5 mg bFGF applied locally to an open wound once a day increases granulation tissue thickness, infiltrated cells, capillary number, and tissue strength of full-thickness punch biopsy wounds. Moreover, in diabetic rats, the wound-healing deficit is improved in a differential manner by bFGF and TGF-b: collagen synthesis in polyvinyl alcohol sponges increases 136 per cent at day 9 by a single application of 2 mg TGF-b but tensile strength is unaltered, whereas bFGF suppresses collagen synthesis and increases cellularity. Epidermal growth factor induces a selective increase in the synthesis of type I collagen, whereas insulin returns collagen activity to normal and causes temporary inhibition of proteolytic activity directed primarily at type I collagen.

Rheumatoid arthritis

Patients with arthritis suffer from healing problems. While the prevalence of chronic leg ulcers among the general population in the United Kingdom is about 1 per cent, the prevalence among patients with rheumatoid arthritis is 9 per cent. In addition to a higher prevalence, the ulcers are even more resistant to healing than those in the general population.

Trace element imbalance

General nutrition among patients is often poor on admission to hospital. An investigation conducted in Dundee, Scotland in 1994 found that 40 per cent of patients studied were malnourished on admission. Such malnutrition is likely to involve a significant degree of trace element imbalance. Adequate levels of trace elements are essential for general health as various metals are involved in general cell function. Zinc, magnesium, and calcium are basic requirements in all phases of healing. However, requirements may be modified during the various phases of wound healing, with iron, copper, and manganese being important in the inflammatory stage and in the formation of granulation tissue. This is particularly important in patients who have malignancies, undergoing chemotherapy or impaired nutrition following major surgery. In the later phase of wound healing an adequate level of selenium is also necessary for optimal wound repair. Selenium is a constituent of glutathione peroxidase, an enzyme that is active in the detoxification of reactive oxygen species, such as peroxides, which are involved in inflammation and tissue damage, degrading collagen and hyaluronic acid, for example, and damaging cell membranes. Peretz reviewed a number of studies, several of which demonstrated low serum levels of selenium in patients with rheumatoid arthritis, particularly in Scandinavian countries where the general population had low levels. Since 1984, however, selenium has been added to soil fertilizers there. Patients in France also had particularly low levels, but Peretz did not report on any study in the United Kingdom. Trace element status can also be adversely affected through interaction with different drugs.

A recent survey of 50 patients with large leg ulcers (> 100 cm^2) indicated that 16 per cent of them in whom venous or arterial problems were not detected, did, however,

have a serious deficiency of zinc and ascorbate which may have contributed to the failure of healing. Zinc is probably the trace element that has been most studied in relation to wound healing. Topical zinc, in the form of calamine, was a therapy that was used for wound healing by the ancient Egyptians. In the porcine wound model topical zinc oxide was shown, in arecent study, to increase the level of mRNA for insulin growth factor 1, essential for re-epithelialization, to 163 per cent of controls. It is not only the absolute value of metal levels which matter, but the ratio of concentrations of certain metals may be crucial, for example $Cu/Zn_{(serum)}$ is altered in some skin diseases—the ratio is 63 per cent higher in patients with decubitus ulcer (*n* = 31) compared with controls (*n* = 48).

Oral zinc supplementation to improve wound healing has been shown to be effective in zinc-deficient patients, but the evidence of its value in improving wound healing in zinc-replete individuals is equivocal. However, some of the earlier studies in this field measured only serum zinc levels as opposed to tissue levels, which are more truly indicative of zinc status.

As with other components, a correct balance within limits is important and too much metal can be detrimental, for example extravascular deposition of iron in the legs of patients with venous disease. The level of free iron in hypertrophic scar tissue has been shown to be six times higher than in normal skin. While iron is an essential cofactor for collagen synthesis it can also catalyze the production of a noxious free radical species (•OH) which is involved in tissue destruction.

Non-cutaneous tissue

The maturation of intestinal wounds differs from the maturation of skin wounds, and is mainly reflected in the rate of accumulation of collagen. Fibroblasts are present in low numbers in intestinal submucosa and collagen is generated by smooth muscle cells. TGF-b augments collagen production in smooth muscle cells by 100 per cent and non-collagen proteins by 40 per cent *in vitro.* Epidermal growth factor has no effect on collagen synthesis in smooth muscle, indicating that modulation of wound repair in intestine differs from that in skin. The gain in strength of intestinal wounds is far more rapid than in skin whether assessed by measurement of anastomotic bursting strength or breaking strength of linear enterotomy wounds. Continuous intraperitoneal delivery of 0.5 mg epidermal growth factor/kg.day increases by 20 per cent the tensile strength of linear enterotomy wounds in stomach, ileum, and colon at 5 days. This is accompanied by an increase in cellularity. In the rabbit stomach, topical application of TGF-b (0.1 to 2 mg per wound) to partial thickness (excluding mucosa) longitudinal wounds accelerates wound breaking strength by 4 days. PDGF (10 mg/wound) does not enhance gastric wound strength but increases cellular influx 2.9-fold, whereas TGF-b does not affect cellularity.

Wound dressings

A material which, when applied to the surface of a wound, provides and maintains an environment in which healing can take place at the maximum rate.

Thomas (1986).

Wound dressings have been utilized since the beginning of time, and some of the dressings used today in plugging and concealing wounds, such as lint and cotton wool, would not seem out of place to practitioners of medicine in early civilizations. In the classification of wound-healing products, these dressings have been referred to as passive. New dressings, such as polymeric films, polymeric foams, particulate and fibrous polymers, hydrogels, and hydrocolloids, have been classified as interactive dressings, providing a microenvironment which is conducive to healing. One of the hydrocolloid dressings (DuoDerm/Granuflex), to which a considerable amount of clinical and experimental research has been devoted, provides a wound-healing environment that, in addition to improving healing, also stimulates angiogenesis (Fig. 22, Fig. 23). The third part of this classification is that of products which actively stimulate healing beyond that of the normal biological maximum.

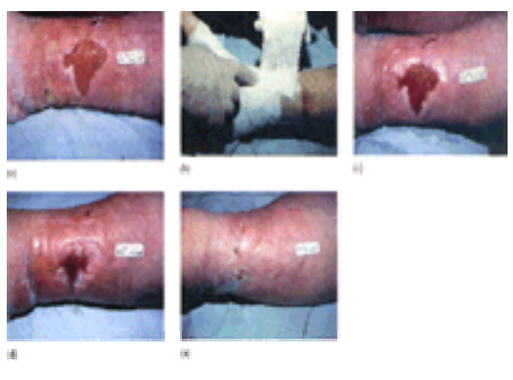


Fig. 22. (a) Venous ulcer before treatment with hydrocolloid dressing under a zinc paste bandage and outer compression bandage. (b) Paste bandage being applied over a hydrocolloid dressing. (c) Four days after treatment with hydrocolloid dressing. Note the increased granulation tissue compared to (a). (d) A marked decrease in the size of the ulcer 4 weeks after using hydrocolloid dressing. (e) Ulcer completely healed.



Fig. 23. Photographs are useful in assessing the healing of chronic ulcers, to measure wound area changes with treatment, but similar assessments can be made by tracing ulcer size using plastic films.

Tissue engineering

Tissue engineering has been defined by a well known worker in this field as ‘the use of biological and/or synthetic materials in conjunction with cells to create biological substitutes to serve as functional tissue replacements'. The variety of organs in which this technique is presently used or is at the experimental stage is extensive (Fig. 24).

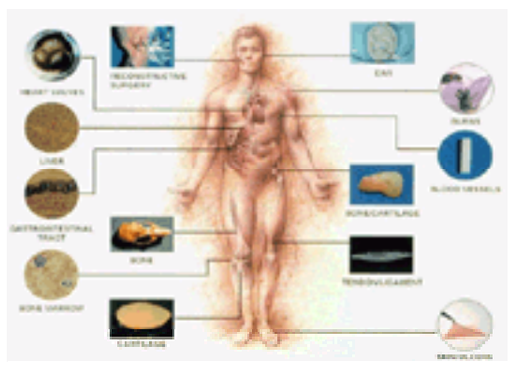


Fig. 24. Illustration of present and future application of tissue engineering in the repair of a variety of tissues and organs in man.

Skin grafts have been used for the coverage of wounds for more than three thousand years. Towards the end of the nineteenth century techniques such as mincing epithelium and spreading it on the wound to stimulate healing were used in Germany. During the 1970s Rheinwald and Green and colleagues developed methods for culturing keratinocytes and preparing epithelial sheets suitable for grafting. This pioneering work led to the successful use of autologous grafts for severely burned patients. The technique has been extended to allogeneic keratinocyte grafts. The latter grafts, although initially thought not to be rejected by the host, have since been shown to be replaced rapidly by host keratinocytes, thus acting as a temporary wound coverage but significantly stimulating the healing process. Disadvantages of this technique include the labour intensity and the need for cell culture facilities. A number of groups have tried to circumvent this by growing the cells directly on a carrier membrane, such as fibrin, polymeric dressing materials, and hyaluronic benzyl ester. The last is commercially called Vivoderm (ConvaTec) and a skin biopsy from a patient is sent to the company for culture of the autologous keratinocytes and growth on the membrane which is then returned to treat the patient (Fig. 4).

At the present time there are a number of living skin equivalents available, comprising dermal and/or epidermal moieties (Table 4). It is recognized that the dermal component is important in skin grafting, and leads to less contracture and improved healing when compared with grafts consisting mainly of epidermis. Clinical studies on treating different types of wounds have been performed using a number of these skin substitutes. Improved healing of diabetic foot ulcers has been demonstrated in a randomized prospective trial using Dermagraft, comprising neonatal fibroblasts grown on a biosorbable suture mesh (Fig. 5, Fig. 25). This product, like some of the other living skin equivalents, has been shown to contain growth factors, matrix proteins, and glycosaminoglycans, components occurring in normal skin which are essential for healing. Another product on which extensive study has been carried out is Apligraf (Novartis) which is a bovine collagen/fibroblast-containing matrix integrated with a sheet of stratified human epithelium. Apligraf has been evaluated in treating wounds caused by surgical removal of skin cancers as well as in randomized prospective trials on the healing of venous leg ulcers (Fig. 26).

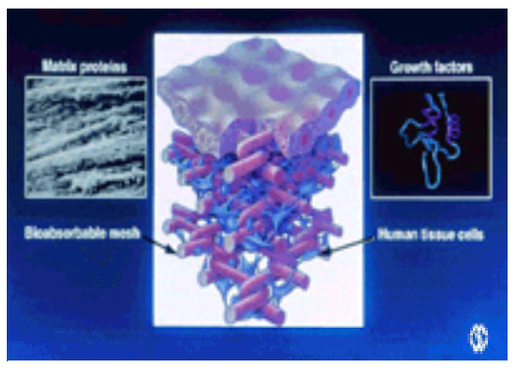


Fig. 25. Platform technology used to bioengineer Dermagraft.

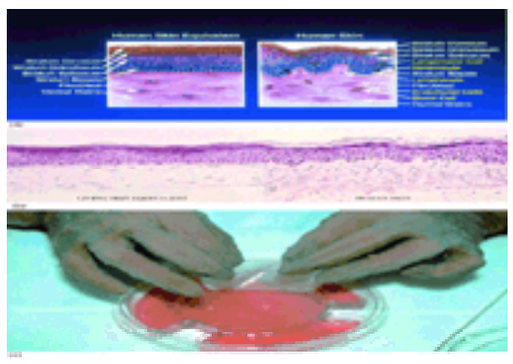


Fig. 26. (a) Schematic comparison between human skin equivalent (Apligraf) and normal human skin (provided by Novartis Pharmaceuticals Corp., East Hanover, NJ, USA). (b) Histological comparison of living skin equivalent (Apligraf) with normal skin. (c) Final product for application to the wound (Apligraf)

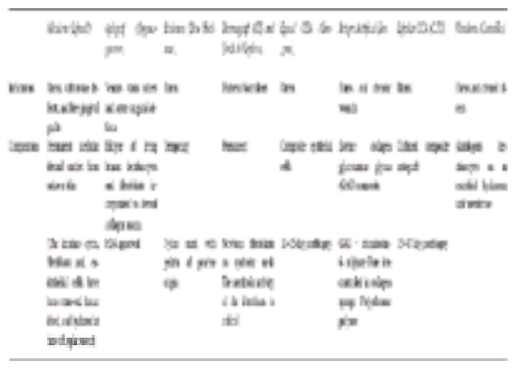


Table 4 Tissue engineered skin

The biomaterial, Integra (Table 4), invented by Burke and Yannas (Harvard-MIT) has been commercially available for several years. This has been used extensively by Dr John Dawson in the Regional Burns Centre in Augusta, Georgia for scar revision and chronic wounds (Fig. 27).

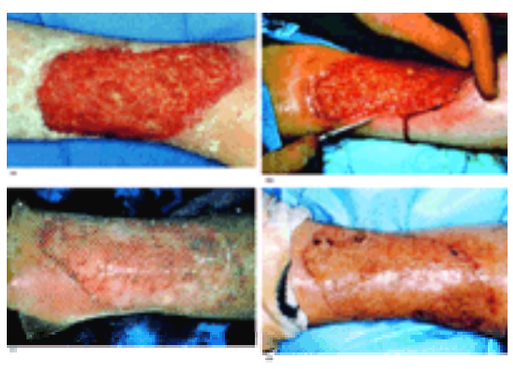


Fig. 27. (a) A 68-year-old diabetic patient with rheumatoid arthritis and venous ulceration, present for a number of years. (b) Excision of devascularized tissue. (c) Integra applied to wound for stimulation of dermal tissue. (d) Result 22 days postoperative following skin graft to new dermis.

New and alternative treatments

One of the recent developments in the delivery of care for wounds is that of wound care centers. This system of delivering care for wounds has gained popularity in the United States where one company has more than 50 such centres. These centres are staffed by multidisciplinary teams. Initially they were set up to offer platelet lysate therapy, but have since concentrated on giving standard wound assessment, treatments, and follow-up. In the United Kingdom there are wound healing treatment centers which have been established since the 1970s in places such as Cardiff, Edinburgh, London, Manchester, Newcastle, and Oxford.

Another new therapy is that referred to as ‘warm-up therapy' using an instrument which produces a warm wound environment by a non-invasive procedure. The theory behind this, based on studies by Hunt and Arbkin, is that localized warming increases blood flow and oxygen to the wound. Work has been published in which this technique has been shown to improve the healing of patients with venous ulcers.

Gene therapy is also being used in experimental wound healing studies, but is still in its infancy.

The majority of the world's population relies on traditional medicine for basic health care, including the treatment of wounds. At Oxford we have had, in our Wound Healing Unit, an exchange programme between researchers from China and Vietnam, and a number of these have been burn surgeons. This has given us the opportunity of looking at the effect of these traditional agents on healing using *in vitro* techniques, as previously described. A number of publications have resulted from this exchange demonstrating that these agents do indeed have a positive effect on the cells involved in the wound healing process (Fig. 20). At the clinical level there is a need for the efficacy of these traditional agents to be demonstrated in proper prospective, randomized, control trials. The latter are beginning to take place and recently such a controlled trial, jointly carried out by Vietnamese and Swedish workers, demonstrated that an extract from the bark of *Choerospondia axiallaris* was successful in the treatment of second degree burns.

The future

The concept of a surgeon working single handed is a thing of the past. Certainly the skills of operative surgery with scrupulous care to join structures such as blood vessels or intestine accurately have to be learnt and audited for best practice. The surgeon who can handle tissues with care and cause the least blood loss for example, should expect the best results, but the sutured skin wound and its successful healing are what the patient perceives as the hallmark of excellence. It is very difficult to measure surgical skill but a working knowledge of materials and tools of the trade are a basic requirement. The days of surgeons telling us that they know how to heal all wounds or who never have a wound failure are long gone. The surgeon who has the prudence to re-explore an abdomen early when there may be an anastomotic failure is the one we admire.

The holistic team approach may have a hackneyed ring to it but the expanding field of wound healing represents an area like no other that depends on a team, as previously discussed concerning wound healing centers. This team should be able to call for expertise from surgeons of all specialties, pharmacists, scientists, nurses and professionals allied to medicine (with regard to rehabilitation and nutrition), and physicians (with regard to endocrinology and care of the elderly). Above all managers need to make resources available and help to put them to optimal use. The number and breadth of wound management and tissue repair societies testifies to this team approach.

Tissue engineering is going to offer the greatest advances in surgical practice in the next few years. Keratinocytes, as previously discussed, can be grown on scaffolds of hyaluronic acid, collagen, or glycosaminoglycans or absorbable polymeric materials, for use on chronic diabetic, pressure, and venous ulcers. Living skin equivalents, stratified squamous epithelia on collagen seeded with fibroblasts, will be evaluated for use on excised burn sites and to cover wide skin excisions for cancer (as well as in chronic wounds). Cells grow well in an extracellular matrix rich in collagen, laminin, and glycosaminoglycans. This tissue engineering will extend to the replacement of bone and cartilage by osteocytes and chondrocytes, as well as other organs.

Technology allowing clinical measurements will also advance, particularly methods of measuring tissue perfusion and oxygenation together with increasingly less invasive imaging. The comprehension of biochemical control of wound healing is more completely and generally understood by the teams involved in wound care. The modulation of control is probably still not at its full potential but the use of growth factors has been disappointing.

The need for multicenter clinical trials is clear, particularly for the evaluation of new technologies. Studies involving dressings have rarely achieved their goals and it is a complex issue to be sure which dressing is appropriate for an individual wound at a specific phase of healing. Evidence-based treatments will be required increasingly. Surveillance and audit, together with quality of life measurements need to be incorporated to bring together the advances of sciences and the pragmatism of clinical practice.

The delivery of wound care in the future, as well as today, will not only depend on efficacy of these new treatments, but just as importantly, on cost-effectiveness. In addition to budgetary restraints, the quality of life issue to patients with wounds needs also to be taken into consideration.

We strongly recommend that in order to appreciate the effect of wounds, particularly severe burns, on the quality of life of individuals, the book *Changing faces* by James Partridge which highlights the complexities of injury from the patient's aspect, should be read.

Further reading

Abatangelo G. Editorial. *Wound Repair and Regeneration* 1996; **4**: 309.

Abatangelo G, Davidson JM, eds. *Cutaneous development, aging and repair*. Liviana Press, Padova, 1989. [A brief comment on some cultured grafts/composites.]

Arbiser JL. Angiogenesis and the skin: a primer. *Journal of the American Academy of Dermatology*, 1996; **34**: 486–97. [A clinical review of angiogenesis in relation to wound healing and other processes.]

Ashcroft GS *et al.* Estrogen accelerates cutaneous wound healing associated with an increase in TGF-b1 levels. *Nature Medicine* 1997; **3**: 1209–15. [A description of oestrogen effects on wound healing in rats and man.]

Ashcroft GS *et al.* Age related differences in the temporal and spatial regulation of matrix metalloproteinases in normal skin and acute cutaneous wounds of healthy humans. *Cell and Tissue Research* 1997; **290**: 581–91. [One of a series of papers from this group documenting age related changes in wound healing.]

Brody GS. Hypertrophic scar contracture. *Plastic and Reconstructive Surgery* 1981; **67**: 673–84. [A classical clinical description including the difference between hypertrophic and keloid scars.]

Burke JF *et al.* Successful use of a physiologically acceptable artificial skin in the treatment of extensive burn injury. *Annals of Surgery*, 1981; **194**:413–28. [A classical study on the use of artificial skin.]

Clark RAF. *The molecular and cellular biology of wound repair*, 2nd edn. Plenum Press, New York, 1996. [An excellent comprehensive text on biological aspects.]

Cowin AJ, Brosnan NP, Holmes JM, Ferguson MWS. Endogenous inflammatory response to dermal wound healing in the fetal and adult mouse. *Developmental Dynamics* 1998; **212**: 385–93. [Compares macrophage numbers and type recruited to fetal and adult wounds.]

Eaglstein WH, Falanga V. Tissue engineering and the development of Apligraf—, a human skin equivalent. *Clinical Therapeutics* 1997; **19**: 894–905. [An overview of tissue engineering related to wound healing.]

Ehrlich HP, Krummel TM. Regulation of wound healing from a connective tissue perspective. *Wound Repair and Regeneration* 1996; **4**: 203–10. [A good summary of dermal contribution to wound healing.]

Ehrlich HP *et al.* Morphological and immunochemical differences between keloid and hypertrophic scar. *American Journal of Pathology* 1994; **145**: 105–13. [A good summary of biological differences between these two types of scars.]

Falanga V *et al.* Rapid healing of venous ulcers and lack of clinical rejection with an allogeneic cultured human skin equivalent. *Archives of Dermatology* 1998; **134**: 293–300. [A prospective randomized multicentre study on the improved efficacy of healing venous ulcers.]

Garlick JA, Taichman LB. Fate of human keratinocytes during reepithelialization in a organotypic culture model. *Laboratory Investigation* 1994; **70**: 916–24. [Emphasizes need for fully differentiated epithelium *in vitro* for wound healing models.]

Garrett B. The proliferation and movement of cells during re-epithelialisation. *Journal of Wound Care* 1997; **6**: 174–7. [A brief description of epidermal aspects of wound healing.]

Hendriks T, Mastboom WJB. Healing of experimental intestinal anastomoses: parameters for repair. *Diseases of the Colon and Rectum* 1990; **33**: 891–901. [Healing problems related to intestine.]

Heimbach D *et al.* Artificial dermis for major burns. A multi-center randomized clinical trial. *Annals of Surgery* 1998; **208**: 313–20.

Herrick S, Ashcroft G, Ireland G, Horan, McCollum, Ferguson MWJ. Up-regulation of elastase in acute wounds of healthy aged humans and chronic venous leg ulcers are associated with matrix degradation. *Laboratory Investigation* 1997; **777**: 281–8. [Demonstrates the harsh proteolytic environment of aged skin and chronic wounds.]

Hinman C, Maibach H, Winer G. The effect of air exposure and occlusion on experimental skin wounds. *Nature* 1963; **200**: 377–9. [A classical study which led to a change of approach in wound treatment.]

Holst RG, Kristofferson A. Lidocaine-prilocaine cream (EMLA Cream®) as a topical anaesthetic for the cleansing of leg ulcers. The effect of length of application time. *European Journal of Dermatology* 1998; **8**: 245–7. [A single blind, randomized trial.]

Jannsen H, Rooman R, Robertson JIS, eds. *Wound healing*. Wrightson Biomedical, Petersfield, 1991. [A general survey on wound healing, bridging the gap between biologists and clinicians.]

Krasner D, ed. *Chronic wound care: a clinical source book for healthcare professionals*, 2nd edn. Health Management Publications, Inc., Wayne, PA, 1997. [Up-to-date clinical management of chronic wounds.]

Ksander GA, Sawamura SJ, Ogawa Y, Sundsmo J, McPherson JM. The effect of platelet releasate on wound healing in animal models. *Journal of the American Academy of Dermatology* 1990; **22**: 781–91. [Experimental growth-factor study.]

Lazarus GS *et al.* Definitions and guidelines for assessment of wounds and evaluation of healing. *Archives of Dermatology* 1994; **130**: 489–93. [Report from a concensus group on definition and assessment of wounds.]

Leaper DJ. Adjuncts to surgery. In: Kirk RM, Mansfield AO, Cochrane J, eds. *Clinical surgery in general. RCS course manual*, 2nd edn. Churchill Livingstone, 1996.

Leaper DJ, Gottrup F. Surgical wounds. In: Leaper DJ, Harding KG, eds. *Wounds: biology and management*. Oxford University Press, 1998. [Describes needles, materials, and methods of incision and closure of surgical wounds.]

Lee RC, Ping J. Calcium antagonists retard extracellular matrix production in connective tissue equivalent. *Journal of Surgical Research* 1990; **49**: 463–6. [Report on a promising new approach to scar treatment.]

Lee RC, Doong H, Jellema AF. The response of burn scars to intralesional verapamil. Report of five cases. *Archives of Surgery* 1994; **129**: 107–11. [Report on a promising new approach to scar treatment.]

Martin CW, Muir IFK. The role of lymphocytes in wound healing. *British Journal of Plastic Surgery* 1990; **43**: 655–62. [Describes aspects of inflammatory cell contribution to wound healing.]

Martin P. Wound healing—aiming for perfect skin regeneration. *Science* 1997; **276**: 75–81. [A recent appraisal of the potential for improving the repair process.]

Myers MB, Cherry GW. Rate of revascularisation in primary and disrupted wounds. *Surgery, Gynecology and Obstetrics* 1971; **132**: 1005–8. [A classical study of the importance of the blood supply in healing.]

NHS Centre for Reviews and Dissemination, University of York, Financial Times Healthcare. *Effective Health Care (Bulletin) 1997; 3(4).* [A bulletin on the effectiveness of health service interventions in relation to evidence-based medicine.]

Ohlsen L, Grafford K, Evers H. EMLA cream as a topical anaesthetic for ulcer debridement and simultaneous split-skin grafting. *European Journal of Plastic Surgery* 1994; **17**: 277–82. [Advantages of local pain control for providing wound coverage through surgery.]

O'Kane S, Ferguson MWJ. Transforming growth factor-bs and wound healing. *International Journal of Biochemistry and Cell Biology* 1997; **29**: 63–78. [A major review of transforming growth factor-b and wound healing.]

Partridge J. *Changing faces: the challenge of facial disfigurement*. Penguin Books, Harmondsworth, 1990. [A patient's perspective of the consequences of severe burn injury.]

Peters RM, Peacock EE, Benfield JR, eds. *The scientific management of surgical patients*, pp. 27–63. Little Brown & Co., Boston, MA, 1983. [A classical chapter on wound healing associated with surgery.]

Phan TT, Hughes MA, Cherry GW. Enhanced proliferation of fibroblasts and endothelial cells treated with an extract of the leaves of *Chromolaena odorata* (Eupolin), a herbal remedy for treating wounds. *Plastic and Reconstructive Surgery* 1998; **101**: 756–65. [Report of *in-vitro* enhancement of wound processes.]

Robson MC. A time to integrate the complete wound team: from bench to bedside and beyond. *Wound Repair and Regeneration* 1996; **4**: 187–8. [Makes a plea for a wound team to manage particularly difficult or recalcitrant wounds.]

Shah M, Foreman DM, Ferguson MWJ. Neutralisation of TGF-b1 and TGF-b2 or exogenous addition of TGF-b3 to cutaneous rat wounds reduces scarring. *Journal of Cell Sciences* 1995; **108**: 985–1002. [Describes key experiments translating findings in embryonic wound healing to therapeutic improvement of scarring in adult wound healing.]

Shah M, *et al.* Role of elevated plasma TGF-b1 levels in would healing. *American Journal of Pathology* 1999; **154**: 1115–24. [Study using transgenic mice to show importance of systemic and local TGF-b levels.]

Stark GB, Horch R, Tanczos E, eds. *Biological matrices and tissue reconstruction.* Springer-Verlag, Berlin, 1998. [a review of advances in bioengineering and wound healing.]

Steed DL and the Diabetic Ulcer Study Group. Clinical evaluation of recombinant human platlet-derived growth factor for the treatment of lower extremity diabetic ulcers. *Journal of Vascular Surgery* 1995; **21**: 71–81. [Randomized, multicentred trial demonstrating the efficacy of PDGF.]

Steed DL, Donohoe D, Wesbster MW, Lindsley L, and the Diabetic Ulcer Study Group. Effect of extensive debridement and treatment on the healing of diabetic foot ulcers. *Journal of the American College of Surgeons* 1996; **183**: 61–4. [Randomized prospective study demonstrating enhancement of healing with frequent debridement.]

Thomas S. The role of foam dressing in wound management. In: Turner TD, Schmidt RJ, Harding KG, eds. *Advances in wound management*. John Wiley, Chichester, 1986. [Summary of wound dressings.]

Uitto J, Olsen DR, Fazio MJ. Extracellular matrix of the skin: 50 years of progress. *Journal of Investigative Dermatology* 1989; **92**: 61s–77s. [Good overview of biology and biochemistry of extracellular matrix in health and disease.]

Vickery C, Harding KG, Leaper DJ. Wound and tissue healing and sutures. In: Taylor I, Karan SJ, eds. *Surgical principles*. Arnold, London, 1996. [Gives a current surgical view of wound healing and adverse factors together with materials and methods.]

Yang L, Qiu CS, Ludlow A, Brummer E., Ferguson MWS. Active transforming growth factor-b in wound repair. Determination using a new assay. *American Journal of Pathology* 1999; **154**: 105–11. [Measurement in tissue sections of TGF-b during wound healing.]

Zouboulis CC, Orfanos CE. Cryosurgical treatment of hypertrophic scars and keloids. *Core Journals in Dermatology* 1991: 11–12. [Describes a different approach to treatment.]

7 Molecular and cell biology in surgical practice

D. J. Weatherall

Introduction
An outline of the structure, function, and regulation of genes
Structure and function
Regulation
Approaches to the isolation of genes, and for studying their function and regulation
Hybridization
Restriction enzymes and Southern blotting
Gene cloning
Gene libraries
Speeding up DNA analysis: the polymerase chain reaction
Finding genes of unknown function
Examining gene expression and function
The Human Genome Project
Clinical applications of recombinant DNA technology
Monogenic disease
Improving cytogenetic analysis
Complex multigenic disorders
Cancer
Genetic manipulation directed at the treatment of disease
Strategies
Vehicles for gene transfer
Receptor-mediated gene transfer
Targeted gene therapy
Interfering with gene function
Cell destruction by genetic manipulation
Artificial chromosomes
Reactivation of fetal genes
Molecular diagnostics
Postscript
Further reading

Introduction

The change in emphasis in the study of human disease from patients or their organs to their cells and molecules has a relatively short history (Table 1). Linus Pauling coined the term ‘molecular disease' in 1949, after his discovery that the structure of sickle-cell haemoglobin differed from that of normal haemoglobin. Until the development of recombinant DNA technology in the mid-1970s, knowledge of events inside the cell nucleus, particularly how genes function, could only be the subject of guesswork based on observations of the synthesis and function of their protein products. However, as soon as it became possible to isolate human genes and study their structure and function, the picture changed dramatically. Over the last 20 years enormous progress has been made towards a better understanding of disease at the cellular and molecular levels. Although, like many recent advances in the basic biological sciences, much of this work has yet to be translated into clinical practice, there seems little doubt that in the none too-distant future this new technology will impinge on many aspects of surgery.

1944	DNA carries genetic information
1953	Structure of DNA
1966	Genetic code established
1972	Recombinant DNA technology
1978	Human gene library
1985	Human gene isolated by positional cloning
1992–7	Linkage map of approximately 95% of human genome

Table 1 The development of modern molecular genetics

This short chapter will review briefly what has been learnt about the structure of genes, and outline the ways in which it is possible to isolate them, analyse their regulation and function, and assess the properties of their products. It will then try to anticipate the clinical applications of this new knowledge for surgical practice in the future.

An outline of the structure, function, and regulation of genes

Structure and function

Essentially, a gene is a particular length of DNA, a double-stranded molecule in which each strand consists of the four bases, guanine (G), cytosine (C), adenine (A) and thymine (T), together with sugar–phosphate backbones. It is the order of these bases, organized as a triplet (three-base) code, that determines the order of amino acids in the protein products of genes. Genes may exist in isolation, as it were, or in families with related functions. For example, the globin genes comprise two separate families, on chromosomes 11 and 16; in each case the order of genes along the chromosome reflects the order in which they are activated at different stages during maturation. There are also so-called superfamilies, that is many different groups of genes found in different regions of the genome, which, although they often subserve disparate functions, have arisen during evolution from limited numbers of forebears and have gradually become more specialized; the immunoglobulin gene family is a prime example.

The human genome consists of somewhere in the region of 3000 Mb of DNA (1 Mb = a million basepairs), spread among 23 pairs of chromosomes, 22 autosomes and two sex chromosomes, X and Y. It is believed that within the human genome there are somewhere in the region of 50 to 100 000 genes, most of which are probably between 1 and 2 kb in size (1 kb = 1000 basepair). It is currently estimated that about 10 per cent of human DNA is made up of coding sequences, while the remainder consists of long strings of repetitive and other sequences, the function of which, if any, is not understood. The distribution of genes and repetitive DNA is not uniform across the chromosomes, an observation that is reflected by their staining properties with Giemsa, a dye that divides chromosomes into a series of light and dark bands. The light bands tend to contain a greater proportion of coding regions, that is functional genes, whereas the dark bands, or Giemsa-positive regions, harbour fewer genes and contain large regions of repetitive DNA.

Most mammalian genes consist of coding regions, or exons, interspersed with variable numbers of non-coding regions called intervening sequences, or introns. In addition they have conserved sequences at both their 5' (left-hand) and 3' (right-hand) ends, which, since they are involved primarily in their regulation, will be discussed later. Many gene clusters made up of genes of similar function also contain pseudogenes, that is genes which resemble structural genes but which contain mutations that render them non-functional. It is thought that they reflect burnt-out evolutionary remnants of once-active genes.

The flow of information between DNA and protein is summarized in Fig. 1. When a gene is transcribed, mRNA (**mRNA**) is synthesized from one of its strands, a process that begins by the formation of a transcription complex consisting of a variety of regulatory proteins together with an enzyme called RNA polymerase. The primary transcript is a large mRNA precursor that is a mirror image of the DNA strand on which it is copied. Because of the rules of base pairing, cytosine always pairs

with thymine and guanine with adenine. The structure of mRNA therefore reflects a faithful copy of the DNA codon on which it is synthesized; the only difference is that RNA uracil (U) replaces thymine (T). Thus it contains both intron and exon sequences. While in the nucleus this molecule undergoes a variety of modifications. First, in a multistep process the introns are excised and the exons spliced together. The now definitive mRNA is further modified by the addition of a complex CAP structure at its 5' end and a string of adenylic acid residues [poly(A)] at its 3' end. The now-processed mRNA then moves into the cytoplasm where it acts as a template for peptide-chain synthesis.

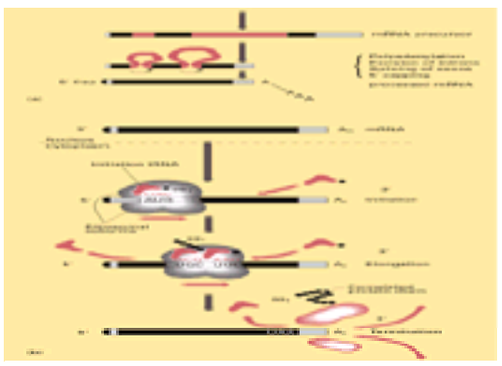


Fig. 1. Gene action: (a) the structure of a gene and the processing of its transcript; (b) the different steps in the translation of mRNA.

Amino acids are transported to the mRNA template on carriers called transfer RNAs (**tRNA**); there are specific tRNAs for each amino acid. Furthermore, because the genetic code is degenerate, that is more than one codon can encode for a particular amino acid, there are several different individual tRNAs for at least some amino acids. The order of amino acids in the peptide chain is determined by the order of codons in the mRNA. tRNAs contain three bases, anticodons, which are complementary to mRNA codons for particular amino acids. They carry their amino acids to the template where they find the appropriate position by codon/anticodon base pairing.

When the first tRNA is in position an initiation complex is formed between several protein initiation factors together with the two subunits that constitute the ribosomes. A second tRNA moves in alongside and the two amino acids that they carry form a peptide bond between them; the peptide chain is now two amino-acid residues long. This process is continued along the mRNA from left to right, and the growing peptide chain is transferred from one incoming tRNA to the next, that is mRNA is translated from 5' to 3'. During this time the tRNAs are held in an appropriate steric configuration with the mRNA by the two ribosomal subunits. There are specific initiation (AUG) and termination (UAA, UAG and UGA) codons. When the ribosomes reach the termination codon, translation ceases, the completed peptide chains are released, and the ribosomal subunits are recycled. In the case of proteins that are built from two different peptide-chain subunits, after synthesis each peptide chain unites with its fellow and folds to assume its characteristic, three-dimensional configuration.

For some proteins the primary gene product may undergo a considerable amount of post-translational modification. Proteins have to be transported to the particular cellular compartments in which they function. This is achieved by an extremely complex cytoplasmic membrane-bound system, involving in particular the endoplasmic reticulum and Golgi apparatus. Proteins that are secreted carry so-called signal sequences that enable them to cross the membrane of the endoplasmic reticulum. Those without these sequences remain in the cytosol. Some of those that enter the endoplasmic reticulum go through the Golgi apparatus to the plasma membrane, or enter special secretory vesicles. At some stage the signal sequences are removed.

Regulation

Since every cell in the body has the genetic information with which to produce an entire individual it is evident that in most cell types the numbers of genes that are expressed must be rather limited. Some, called housekeeping genes, are expressed in most cells. Others are highly specialized and are only expressed in particular cell populations In addition, in some gene families different members are expressed at different stages of development. Although knowledge of how these extremely complex programmes are regulated is still extremely limited, there is some understanding of the general principles involved.

The regulation of gene expression seems to be mediated mainly at the transcriptional level, with some fine tuning during translation and post-translational modification of gene products. DNA not involved in transcription is held tightly packaged in a compact, chemically modified form that is inaccessible to regulatory factors and polymerases, and is heavily methylated. Activation of a particular gene or gene family is reflected by changes in the structure of the surrounding chromatin that can be identified by enhanced sensitivity to nucleases. For example, in the cells of the bone marrow, erythroid lineage-specific nuclease-hypersensitive sites are found at several locations in the cluster of genes that controls the structure of the different b-like chains of haemoglobin. Four are distributed over 20 kb upstream from the family of genes that form the b-globin gene cluster. This marks out a vital regulatory region, called the locus control region, which is able to establish a transcriptionally active domain spanning the entire gene cluster. Similar ‘master' regulatory sequences have been found in proximity to other gene clusters. It is believed that the locus control region becomes opposed to the promoters (see below) of individual genes to increase their rate of transcription. In this sense it behaves like a typical enhancer sequence, that is a region of DNA which upregulates adjacent genes on the same chromosome, despite being situated at some distance from them. In addition to these regulatory elements, each individual gene has highly conserved sequences at its 5' end. The first, the ATA box, is about 30 bases upstream from the initiation codon. The second, the CCAAT box, is about 70 bases upstream; many genes have additional promoter elements, often about 80 to 100 bases further upstream. All these DNA sequences are involved in the initiation of transcription of individual genes. Also, at their 3' non-coding regions most genes have the sequence AATAAA, which is the signal for cleavage and poly(A) addition to mRNA transcripts.

In addition to these basic regulatory sequences, which are on the same chromosome, there are families of regulatory proteins that are encoded at some distance from the structural genes, often on other chromosomes. These so-called transcription factors act in both *cis* and *trans*, that is on both pairs of homologous chromosomes. They fall into several classes. Some are ubiquitous and seem to be involved in the activation of many genes in different tissues. Others are lineage-specific, that is they only act in certain specialized tissues, the bone-marrow progenitor line destined to become red cells, for example. In each of the enhancer regions, round the promoters, and in other regions close to structural genes, there are specific DNA motifs with which these transcription factors bind. Some of them are likely to be developmental stage-specific. Thus gene regulation and activation is currently thought of as a highly complex process in which one or more DNA-binding proteins interact with the promoter regions and associated enhancers, and, together with a number of protein-transcription factors, produce what is called a transcription-initiation complex. In addition to these various positive regulatory elements a number of suppressor, or silencer, regions have been found in close proximity to specific genes or gene families.

But, although the basic anatomy of gene regulation is being worked out, the mechanisms by which these different transcriptional regulators work in a concerted action to turn on specific batteries of genes involved in a similar function, or genes expressed at specific times of development, are not yet understood.

Approaches to the isolation of genes, and for studying their function and regulation

The exploration of the molecular pathology of disease has been made possible by the development of a wide variety of new technologies over the last 20 years. It is beyond the scope of this short account to deal with them in detail; readers who wish to learn more about them are referred to the references at the end of this section. There follows a simple outline of what is possible.

Hybridization

The earliest approaches to the investigation of human genes were indirect, relying on the fact that it is possible to form hybrids between DNA sequences that are almost identical, or between DNA and RNA. Hence, using radioactively labelled DNA or RNA probes, it was feasible to explore the human genome to look for major deletions or changes in the number of copies of genes in different diseases. The first deletion to be demonstrated directly in man, that which underlies a severe form of a-thalassaemia, was demonstrated in this way. However, these approaches were indirect and at the best could only provide semiquantitative data.

Restriction enzymes and Southern blotting

A seminal advance towards the direct exploration of a gene structure was the discovery of bacterial restriction enzymes, a family of enzymes, produced in different types of bacteria, which are able to cut DNA at particular sequences. In this way it was possible to fractionate human DNA into pieces of predictable size and in which the sequence that had been cut was always the same. Another major step forward was the development of a technique called Southern blotting, named after its

inventor Edwin Southern (Fig. 2). Here, mixtures of fractionated DNA are separated by virtue of their size by gel electrophoresis, after which the DNA fractions on the gel are blotted on to nitrocellulose paper, on which they can be immobilized. They are then exposed to radioactively labelled gene probes, lengths of DNA that have base sequences homologous to the gene that is being sought. They will only hybridize with the appropriate sequences; the position of the hybrids can then be identified by autoradiography. This was a huge technical advance because it allowed a variety of human genes to be studied directly and deletions or other major rearrangements to be identified as the cause of several monogenic diseases.

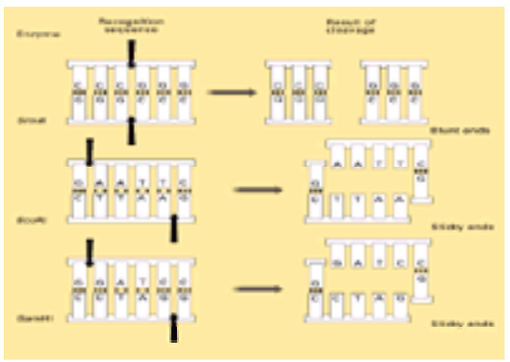


Fig. 2. Restriction enzyme mapping (Southern blotting).

Gene cloning

The next step, and the one that opened the era of recombinant DNA technology, was the ability to introduce fractionated human or other DNA into vectors that are able to replicate in bacterial cells. In this way it is possible to generate, or clone, fractions of DNA containing genes that were being sought, and to produce large quantities for further study. The first cloning vectors were bacterial plasmids, small circles of DNA that had been studied earlier for their properties in transferring antibiotic resistance between different strains of bacteria. It was relatively easy to insert foreign DNA into plasmids, although their value as cloning vectors was limited by the relatively small size of the DNA fragments that they could accommodate. Later, these problems were overcome by the development of more sophisticated vectors, including single-stranded bacteriophage, bacteriophage-l, and cosmids—vectors that were able to incorporate relatively large DNA fragments by combining the small size of a plasmid vector with the limited requirements for packaging DNA into a bacteriophage-l infectious particle. Later it became possible to clone even larger fragments into artificial chromosomes. This advance was based on an increasing understanding of the various elements required for chromosomal stability in yeast, including the centromeric sequences required for chromosome segregation at meiosis, the telomeres (sequences that define the end of the chromosomes), and the origins of replication (**a**utonomously **r**eplicating **s**equences, or ARS). This knowledge, combined with the development of methods that made it possible to introduce DNA into yeast more efficiently, led to the emergence of a new family of cloning vectors called **y**east **a**rtificial **c**hromosomes, or YACs.

Gene libraries

The next step was to construct a set of recombinant clones from a source known to contain a particular gene or other DNA sequence of interest. This type of collection, referred to as a gene library, can be designed to contain all the human genes. These libraries are of two types: cDNA, which are DNA copies of the mRNA populations within a cell, and genomic, which are derived from the total genomic DNA of a particular organism. By developing a wide range of ingenious screening procedures it was possible to isolate individual bacterial colonies containing genes of interest and then to grow them individually so that sufficient DNA could be obtained for structural or functional studies. The development of rapid sequencing techniques soon made it possible to determine the structure of genes isolated in this way.

Speeding up DNA analysis: the polymerase chain reaction

Another technique that revolutionized the analysis of human DNA is the application of the polymerase chain reaction (**PCR**) for the chemical amplification of nucleic acid sequences (Fig. 3). This is achieved through multiple cycles, in each of which a double-stranded DNA template is first denatured by heating. After cooling, specific oligonucleotide primers designed to cover the DNA region of interest are annealed to the template strands. An enzyme called DNA polymerase is then used to copy the two DNA strands, starting at the primers. The result is a two-stranded copy of the particular DNA region. This process is then repeated cycle after cycle; after about 30 cycles 2^{30}, or 10^9, DNA fragments are generated from each initial template. This remarkable technique was facilitated by the discovery of a heat-stable DNA polymerase and the development of thermocyclers, instruments that are capable of rapid cycling between the various temperatures required. The PCR product is usually detected by gel electrophoresis of the total reaction products. This technique has innumerable applications, including the development of methods for rapid DNA sequencing.

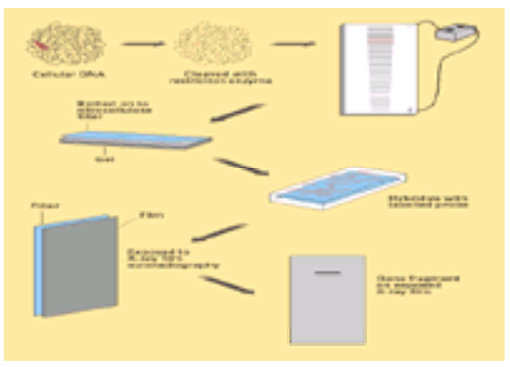


Fig. 3. The polymerase chain reaction. Specific DNA sequences are amplified in the following way. The double-stranded DNA is heated and the separated chains are allowed to bind the primers, which then initiate the sequences of two new chains complementary to the originals. This series of events is repeated 20 to 30 times, with each cycle giving a doubling of the DNA.

Finding genes of unknown function

Early successes in isolating human genes, both normal and associated with monogenic disease, were restricted to those about which something was known of the function of their products. This made it easier to develop probes and to screen gene libraries. However, there were many monogenic diseases, and even more multigenic diseases, in which the function of the offending gene was not known. The way in which this problem has been approached is based on ideas that have been well known to human geneticists for many years; it is the new DNA technology that has made it possible to apply them successfully for the identification of many different genes the function of which has been hitherto unknown.

In the early 1940s the geneticist J. B. S. Haldane suggested that one way to track human genes was to use one that could be identified, eye colour or a blood group for example, and to follow it through a family together with the particular phenotype, a disease for example, associated with the gene that was being sought. If the two stayed together they must be linked on the same chromosome; if the location of the gene that can be identified is known, it should then be possible to find the gene of unknown function, since it must be close by. The problem was that, until DNA technology became available, there were relatively few genetic markers with which to carry out linkage studies of this type. However, once the structure of human DNA was analysed it became clear that there was remarkable individual variability. Over recent years a variety of different families of repetitive sequences have been identified and a linkage map covering most of the human genome has been developed. This, together with the availability of rapid screening techniques, has made the search for human genes by linkage both feasible and, at least to some extent, rapid. Once a linkage has been established it is then possible to move up and down the chromosome and, with luck, to isolate the particular gene that is being sought. From its sequence it is then possible to make an educated guess about its likely function. This approach, originally called reverse genetics and later rechristened positional cloning, has allowed the identification of a wide variety of genes responsible for important monogenic diseases (Table 2), and has enabled a start to be made in defining loci involved in complex multigenic disease, a topic to which we will return later.

Duchenne/Becker muscular dystrophy	1985
Chronic granulomatous disease	1986
Retinoblastoma	1986
Cystic fibrosis	1989
Neurofibromatosis 1	1990
Colon cancer DCC APC*	1990
Fragile X syndrome	1991
Huntington's disease	1993
Polycystic kidney disease 1	1994
Tuberous sclerosis	1994

*DCC, deleted colon cancer; APC, familial adenomatous polyposis.

Table 2 Some susceptibility genes for common disease

Examining gene expression and function

Once it was possible to isolate human genes it was important to learn how they are regulated. As already mentioned, there are key regulatory regions both adjacent to, and at a distance from, genes. In order to study their function, and further to define these regulatory regions, it became necessary to transfer genes into new environments. Initially, key regulatory regions were identified by so-called transient transfection systems, that is by inserting the gene into a particular cell line and studying its expression over a short period. It was in this way, combined with techniques for mapping the primary transcript of different genes, that the promoter regions and other important regulatory regions were defined.

Another major advance was the development of methods for defining the regulatory proteins involved in the control of gene expression. Such protein–DNA interactions may affect the transcription of a particular gene positively or negatively. It became possible to identify their presence by detecting them in crude cellular extracts by gel electrophoresis, a technique known as a ‘gel shift' assay; if a region of DNA is bound to any particular protein the protein retards its mobility so that DNA–protein complexes appear as discrete bands on electrophoresis.

For more sophisticated studies of gene function it was necessary to transfer genes into foreign cells. One of the earliest approaches was to insert genes into retroviruses, since they have all the genetic machinery to enter a cell and insert their genetic material into its genome. A variety of retroviral vectors were constructed for this purpose and a considerable amount learnt about gene function in this way.

Later, methods were developed to change the genetic make-up of animals by the insertion of foreign genes. Several approaches have been used for the introduction of foreign DNA into mice. The first involved the construction of so-called transgenic mice, through the microinjection of recombinant DNA into one of the pronuclei of a fertilized egg; the egg is then reimplanted into a pseudopregnant female mouse. DNA can also be introduced into mice on retroviral vectors. Later it became possible to use mouse embryonic stem cells, totipotent cells capable of differentiating into various cell lineages and giving rise to all the tissues of an adult animal.

Embryonic stem cells can be manipulated in culture and then transferred to a mouse. Recombinant DNA can be introduced into embryonic stem cells in culture by a variety of techniques. The cells are then injected into mouse blastocysts and as the embryo develops the embryonic stem cells carrying the inserted genes contribute to the production of various tissue types. The result is an adult mouse that is chimeric, some cells being normal and other transgenic. By cross-breeding it is possible to produce animals which are homozygous for the transgene and hence express it in all tissues. Using this technique a particular cellular gene can be replaced with a modified one or the function of a gene can be ‘knocked out'. This approach has been of enormous value in studying tissue-specific gene regulation and for the understanding of normal development. It has also contributed to the development of animal models for the study of human disease.

The Human Genome Project

With the further development of sophisticated approaches to DNA sequencing, and its automation by the use of robotics and other methods, there has been much interest in a project directed at producing the complete sequence of human DNA. The first goal, which has been more or less achieved, is to construct a linkage map that covers most of the human chromosomes having useful markers. The next step, the creation of a physical map of the entire human genome, is well under way, and the project is expected to be completed in the early years of the new millennium. The information that this will provide, together with equally remarkable developments including microchip technology for the rapid screening of gene products, is likely to completely revolutionize this field. Current problems that are inhibiting progress, and will have to be solved, include our limited ability to work out protein structures based on information derived from DNA sequences, and the lack of biomathematical approaches to start to build the kind of models that will be required to help us understand how the activities of our different genes are integrated into the complex cellular-control circuitry underlying gene action in higher organisms.

Clinical applications of recombinant DNA technology

The sections that follow will examine briefly some of the clinical potentials of the recombinant DNA field. It is only possible to touch on this vast topic, which will undoubtedly revolutionize many aspects of clinical practice in the new millennium; readers who wish to learn more about specific topics are referred to the reference list at the end of this chapter.

Monogenic disease

It is not surprising that recombinant DNA technology has had its most spectacular successes so far in the elucidation of the molecular pathology of monogenic disease. Indeed, it seems likely that, although there will undoubtedly be some surprises, we have a reasonable idea of the repertoire of the different mutations that underlie human genetic disease.

As shown in Table 3, mutations may interfere with gene function at many levels. Missense mutations, which result in the insertion of a different amino acid in a protein, particularly if they occur in regions that alter function, may result in a wide variety of diseases. For example, amino acid substitutions in various parts of the haemoglobin molecule may give rise to diverse phenotypes ranging from sickle-cell anaemia, through hereditary polycythaemia or cyanosis, to a crippling haemolytic anaemia due to instability of the altered haemoglobin molecule. Similarly, amino acid substitutions in the a_1-trypsin molecule may lead to its deficiency or even convert it to a powerful anticoagulant, while subtle differences of substitutions in the fibroblast growth factor-receptor gene may be responsible for the completely different skeletal deformities of Apert or Crouzon syndromes.

Transcription
Structural changes: deletions, insertions, inversions
Promoter-box mutations
Transcription-factor gene mutations*mRNA processing*
Mutations at GA/AG intron/exon junctions
Mutations of consensus sequences near junctions
Activation of cryptic splice sites in introns or exons
Poly(A) addition-site mutations*Translation*
Initiation codon mutations
Missense or nonsense mutations
Termination codon mutations
Post-translational
Unstable products
Malfunctional products

Table 3 Levels of defective gene function

An even more diverse series of mutations results in the absence, or a reduction in the amount, of a gene product. Nonsense mutations, that is base substitutions in

exons that lead to premature stop codons in mRNA, or mutations that alter the reading frame of the genetic code, lead to premature chain termination and an absence of gene product. There is a wide variety of different mutations that involve the processing of mRNA; some interfere with the splicing mechanism completely so that no gene product is produced while others lead to the generation of alternate splice sites in introns or exons so that both normal and abnormal mRNA are produced. These mutations can give rise to a wide variety of phenotypes of varying severity. Similarly, mutations in the promoter regions of genes can down-regulate them to varying degrees, again producing many different disease phenotypes.

Some monogenic diseases result from much more extensive structural changes in particular genes. Most commonly these involve complete or partial deletions. However, there are more subtle changes. For example, several monogenic diseases have been shown to result from major gene inversions, that is regions of DNA being in the opposite orientation to normal. The first of these to be identified was a large inversion involving several of the human b-like globin genes. There are also rare examples of genes being interrupted by the insertion of DNA sequences along their length. In clusters of genes with similar structure, the globin genes for example, it is not uncommon for slippage and mispairing to occur at meiosis, with the production of fusion genes. Depending on the structure of the N-terminal partner of the fused genes, they may or may not be effectively transcribed.

Other families of mutations involve complex regulatory molecules that govern gene function. Some of them are on the same chromosome as the genes that they control and, if mutated or lost by deletion, completely inactivate single genes or gene families. Mutations of transcription factors may cause the abnormal function of whole families of genes on different chromosomes. Several of these have been defined as the cause of severe mental retardation with or without widespread abnormalities of development. The further understanding of how these proteins function during normal development promises to yield valuable insights into hitherto unexplained mental retardation associated with dysmorphological syndromes.

The ability to identify mutant genes so easily has already been applied widely in clinical practice. It offers a much more accurate way of identifying carriers and for the prenatal detection of severe diseases of this kind. DNA can be isolated by chorion villus sampling from the ninth to tenth weeks of pregnancy or from cells in amniotic fluid. By the use of PCR and specific gene probes it has been possible to increase the speed of analysis of these samples; prenatal diagnosis for many monogenic diseases can now be made within 2 h of receiving a chorion villus biopsy.

Improving cytogenetic analysis

DNA technology has had a major impact on cytogenetics. By the use of techniques involving *in situ* hybridization, together with the development of more sophisticated forms of microscopy, it is possible to identify individual genes on chromosome spreads and hence to describe cytogenetic abnormalities that cannot be seen by conventional light microscopy. As well as providing a valuable approach to the localization of genes, this technology has wide application in the clinic, particularly for identifying small deletions, those involving the tips of chromosomes for example. It has also provided a valuable way of identifying subtle translocations and other cytogenetic abnormalities that underlie developmental disorders.

Complex multigenic disorders

Disorders such as heart disease, hypertension, major psychoses, dementia, osteoporosis, rheumatism, epilepsy, and most forms of cancer are not, except with a few rare exceptions, transmitted through families in a way that suggests that they are caused by a single mutant gene. Rather, these disease have a variable component of genetic susceptibility, involving several different genes, while in many cases the environment seems to be of equal or even more importance. For example, twin studies suggest that vascular disease has a genetic component of about 30 per cent, whereas, in the case of type 2 diabetes, heritability is almost 100 per cent. There is, however, considerable interest in defining the particular genes involved in these multigenic disorders. It seems likely that a better understanding of their function would lead to the definition of the cellular mechanisms that underlie these conditions and hence to more logical forms of treatment. As a by-product, it may also be possible to define subgroups of individuals who are at particularly high risk of exposure to environmental insults and hence to concentrate public-health measures for their prevention in a more logical way.

There are several ways of approaching this problem. Where there are obvious candidate genes, those involved in glucose metabolism in the case of diabetes for example, it is possible to establish linkage studies to determine whether variation at these loci plays a part in the generation of the diseases. Where no candidates are known it is feasible, either by the use of sibling pairs or the study of extended families, to apply the linkage methods mentioned earlier in this chapter to try to define some of the genes involved. By the combined use of these approaches there has been some success in identifying at least part of the genetic component of some of these multigenic disorders.

Studies of the genetics of type 1 diabetes, both in humans and in mice, have given some indication of the complexity of these multigenic systems; in mouse and man there are over a dozen loci involved in susceptibility to this condition. Fortunately, however, at least in the mouse and almost certainly in man, it appears that only two or three are of major importance. Two that have been identified with certainty are the HLA-DR complex and its mouse equivalent and, in humans, a DNA sequence close to the insulin gene, which is probably involved in its regulation. It appears from these studies that these genetic polymorphisms increase an individual's susceptibility to particular environmental triggers. Having this genetic make-up does not necessarily mean that an individual will inevitably suffer from the particular disease, and, similarly, many people who are not ‘genetically susceptible' in this way will nevertheless contract them. Some of the other genes that have been defined in these multigenic systems are summarized in Table 4.

Disease	Genes
Cardiovascular	*LDL(R); ApoB, C, E*
	ACE; LP(A); fibrinogen
Type 1 diabetes	HLA-DR; insulin; *IDDM1, 4, 5*
Type 2 diabetes	Glucokinase; *IVS1(R); MIT*
	Glucagon (R)
Epilepsy	Cystatin B
Obesity	*ob*
Alzheimer's	*ApoE4; APP; PS1; PS2*
Osteoporosis	Vitamin D(R)
Atopy	*FCεRIβ*
Cancer	*BRCA1* and 2; *Rb*; *WT*; *APC*; *LMS3R*; p53
	Many other oncogenes

R, receptor.

Table 4 Some susceptibility genes for common disease

While progress has been remarkably rapid in starting to define some of the major genes in multigenic diseases, a word of caution is necessary. Many false leads have been followed and it is clear that a lot more work will be required before a clear picture of the genetic susceptibility to environmental agents is obtained. A major problem that has been encountered, notably in psychiatric genetics, is the definition of phenotypes. This has led to a number of false assignments of susceptibility or resistance genes to particular chromosomal regions. Work on polymorphisms of the vitamin D receptor, which seemed to be involved in susceptibility to osteoporosis, has shown that inconsistent results may be obtained in different populations and that subtle changes in the environment may make genetic polymorphisms of this kind more or less important.

Cancer

Except for some rare familial cancers it appears that the majority are caused by somatic mutations, that is changes in our genes that we acquire during our lifetime and which may be passed on to the progeny of cell populations. Indeed, it is now believed that most cancers are initiated in this way, a concept that has revolutionized our thinking about the cause of malignant transformation. Exposure to environmental mutagens, or to the generation of endogenous mutagens within the body, may damage the DNA of key regulatory genes that are ubiquitous among all living organizms.

This family of genes, called cellular oncogenes, encompasses growth factors and their receptors, pathways of signal transduction, and a variety of DNA-binding transcription factors. In the case of certain familial cancers, we may be born with mutations involving oncogenes but their deleterious effect is masked by the normal partner on the opposite of the pair of homologous chromosomes. The function of such tumour-suppressor genes may be inactivated by a second mutation, hence leading to a malignant transformation. But it seems likely that most of the common cancers result from the acquisition of several mutations; in the case of colon cancer up to six oncogene mutations are required before the full neoplastic phenotype develops.

There has been considerable progress in linking the action of oncogenes to the regulatory mechanisms that control the way in which the cell cycle is organized. Cell division reflects a complex series of steps during which DNA is synthesized, cells divide, and either continue to do so or move out of cycle into a resting phase. Every

step needs to be carefully regulated and monitored. For example, cells that have acquired a serious mutation or other form of damage to their DNA should not be allowed to continue to divide before the defect is repaired. The cell cycle is regulated by a family of proteins called cyclins, which are activated by another battery of regulators called cyclin-dependent kinases. Those kinases are themselves controlled by other regulatory proteins. The oncogene, p53, sits in the middle of these pathways. It appears to have the ability to slow down the cell cycle when DNA repair is required, identify noxious environmental conditions such as hypoxia, and to decide whether a cell should go into cycle or should die by programmed cell death (apoptosis). Several other oncogenes have a similar role in the regulation of the cellular behaviour.

The abnormal activation of oncogenes can occur in a variety of ways, including the acquisition of point mutations, deletions, duplications, and as part of major chromosomal changes, including translocations. Hence, by identifying the chromosomal location of oncogenes and defining the chromosomal changes that occur regularly in particular cancers, modern cancer biology has brought together the fields of cytogenetics and oncogene function to provide the beginnings of a picture of the mechanisms and complexities of the generation of many common forms of malignancy.

As well as its therapeutic possibilities, to which we will return later, this new information in the cancer field has considerable promise for development in the clinic. The identification of the mutations that underlie the familial cancers has provided invaluable information for genetic counselling and early diagnosis. There is much interest in the development of PCR-based techniques for identifying oncogene mutations in screening programmes using exfoliative cytology. This includes the early identification of cancers of the gastrointestinal tract, lung, and prostate.

Genetic manipulation directed at the treatment of disease

Over the last 20 years, although the initial dreams of gene therapy, that is genetic manipulation directed at the treatment of disease, have been transformed into the reality of more than 175 clinical trials and over 2000 patients treated, there is, as yet, still no conclusive evidence for efficacy. However, sufficient progress has been made to suggest that this type of approach to treatment will slowly find a place in clinical practice.

In discussing genetic manipulation it is important to distinguish between germ-cell and somatic-cell gene therapy. Germ-cell therapy involves insertion of a gene into fertilized eggs for the correction of a genetic disease. Because these genes are dispersed throughout the tissues of the egg, they end up in the germ cells as well as the somatic cells of the fetus, and hence are passed on to future generations. This approach has been banned in most countries. The other form of gene therapy, somatic-cell therapy, involves the insertion of genes or otherwise manipulating the genetic machinery of a cell to treat a disease. In this case the cells are restricted to the population that has been treated and any genetic change remains restricted to these cells and is not passed on through the germ-cell line.

Strategies

The general approaches to somatic-cell gene therapy are outlined in Table 5. Ideally, gene therapy should emulate transplantation surgery in removing the mutant gene and replacing it with a normal one. Since this is not technically possible, another way of achieving the same end, gene correction, entails the specific alteration of a mutant gene sequence using nature's way of exchanging genetic material, that is by site-directed recombination. Although this is the ideal approach to correcting a genetic defect or otherwise altering the function of a gene, and it has been achieved in cultured cells, current methods have such a low level of efficiency that it may be a while before they can be used in practice.

Physical:
Direct injection of DNA
Calcium phosphate transfection
Electroporation
Liposome-mediated DNA transfer
Retroviral vectors
Other viral vectors
Receptor-mediated gene transfer
Targeted gene transfer
Artificial chromosomes
Activation of genes of related function

Table 5 Methods for gene therapy

Because of these technical difficulties much current work is directed towards gene augmentation, that is introducing a gene into cells in a way that will allow it to produce sufficient of its product to compensate for loss of expression of its defective counterpart. At the same time, and particularly in the fields of cancer and infectious diseases, attempts are being made to interfere with the expression of genes in order to either modify the behaviour of cells, or to kill them.

Finally, because certain genetic diseases are due to mutations in genes which have counterparts that are active at earlier stages of development, there is growing interest in the reactivation of these ‘fetal' genes in order to take over the role of their defective adult homologues.

Vehicles for gene transfer

Some of the major approaches that are being pursued for transferring genes or other agents designed to interfere with gene function into cells are summarized in Table 5. They fall into two main classes, viral and non-viral.

Several classes of viral vectors have been subjected to clinical trial. Because they integrate DNA into the genome, much early research focused on retroviral vectors, retroviruses from which many of the viral genes had been removed or altered so that no viral proteins can be made in the cells that they infect. Viral-replication functions are provided by packaging cells which contain ‘helper' viruses that produce all the viral proteins required, but which themselves have been altered so that they are unable to produce infectious viruses. A great deal of experimental work has been based on the mouse Moloney leukaemia virus. Vectors constructed from this virus transduce cells in culture. However, most cells *in vivo* are quiescent and because this virus only invades dividing cells it has limited uses for gene therapy. More recently, a chimeric Moloney–human lentiviral vector has been constructed that can transduce at least some quiescent cells, including neurones.

Because of these problems, adenoviruses have been exploited as vectors for gene therapy, particularly for the treatment of cystic fibrosis and cancer. By eliminating a particular region of this virus it is possible to insert human genes and to render the virus incapable of replication. Hence after gene transfer the virus cannot spread. The problem with this family of vectors is that they produce antigens which elicit an immune response that tends to eliminate the transduced cells and hence the inserted human gene. However, a considerable amount of work is continuing in an attempt to prevent the production of these immunogenic proteins. Another family of viruses that is showing considerable promise for gene transfer are the adeno-associated viruses, which appear to be able to transduce brain, skeletal muscle, and other cells. Their major disadvantage is the relatively small amount of foreign DNA that can be inserted.

A variety of other viral vectors are being explored, based on Epstein–Barr virus, herpesvirus, simianvirus 40, papillomavirus, and others. However, it has to be said that, to date, the ideal retroviral or viral vector for gene therapy has still not been developed.

Of the non-viral vectors listed in Table 5 most work has been directed at cationic, lipid-based delivery systems called liposomes. They have the advantage that they are relatively easy to prepare and do not evoke an immunogenic response. They do not carry the dangers of retroviruses, particularly their oncogenic potential. On the other hand they are a relatively inefficient way of transferring genes into the nucleus and it is not possible to target them at specific tissue.

There are several potential applications of this kind of technology that may find a place in surgical practice in the future. For example, a number of growth factors have been identified that are involved in angiogenesis, vascular endothelial growth factor, fibroblast growth factor, and angioprotein 1, for example. In the cancer field, because of the importance of tumour vascularization, attempts are being made to inhibit angiogenesis. On the other hand, there is much interest in the possibility of stimulating this process as an approach to the management of ischaemic heart disease and peripheral vascular disease. Since the direct administration of these growth factors has serious limitations, efforts are being directed at targeting them to affected tissues by inserting their genes into the kinds of vectors mentioned earlier. Following successful animal experiments, several phase I and II clinical trials are underway to determine whether, by injecting these genes in adenoviral vectors into the coronary circulation, it is possible to improve the vascularization of the ischaemic myocardium. This kind of approach could have wide applications for vascular

surgery in the future.

Receptor-mediated gene transfer

Because most of the experimental work involving the use of gene transfer vehicles outlined in the previous section has been done by transfecting cells *ex vivo*, and with the growing realization that this will often be inefficient and inapplicable in the case of many organs, there has been much interest in the concept of receptor-mediated gene transfer. In theory, this would entail designing a vector that could be injected into the bloodstream and that would home in on an appropriate tissue.

Because the liver is of such central importance in metabolism, and is therefore an obvious target for gene therapy, much work has focused on hepatic-specific receptors, in particular the asialoglycoprotein receptor. Transferrin–polylysine–DNA conjugates, which bind transferrin receptors, have been constructed for this purpose. DNA is packaged in an electrostatic complex using asialo-orosmucoid–polylysine in place of lipid. This conjugate efficiently binds to asialoglycoprotein receptors on the surface of the liver and can be given intravenously. The problem that been encountered with this approach is that, after they enter the hepatic cells, the conjugates fuse with vesicles, forming endosomes in which the DNA is largely degraded. Hence a variety of ingenious approaches are being explored to avoid the endosome-shuttling stage.

A related method involves the use of the haemagglutinating virus of Japan (**HVJ**; a form of Sendai virus). This attaches to the plasma membrane at the cell surface, after which fusion occurs releasing the viral contents directly into the cell cytoplasm. Although transient expression is still a major disadvantage of the HVJ–liposome systems so far developed, a variety of other ways to improve its utility are being explored.

Targeted gene therapy

As mentioned earlier, the use of homologous recombination to make alterations in gene structure offers the most specific approach to gene therapy. Although this has been achieved in human cell lines and non-transformed primary human cells, the efficiency is still very low.

Interfering with gene function

Because of the different objectives involved in controlling cancer or infectious disease compared with monogenic disease, a number of approaches are being explored that aim to alter the function of particular genes involved in neoplasia or in the virulence of infectious agents.

Various techniques directed at the manipulation of the genes of cancer cells in an attempt to restore a normal phenotype are being investigated. Some forms of neoplasia reflect defective function of tumour-suppressor genes, including *TP-53*, the protein product of which is p53, *MTS-1*–p16, *RB-1*, and *APC. TP-53* and *MTS-1*–p16 are defective in a variety of tumours, while *RB-1* mutations have been implicated in retinoblastoma, prostate carcinoma, and osteosarcoma. *APC* mutations predispose towards adenomatous polyposis and a proportion of sporadic colorectal tumours. Attempts are being made to replace these defective tumour-suppresser genes with the normal, ‘wild-type', genes, and, at least in cultured cells, some successes have been achieved.

Another approach to modulating the expression of oncogenes is through the use of so-called antisense agents, which are short, complementary, single-stranded oligonucleotides that interfere with gene function by binding specifically to their corresponding cellular mRNA partners. A related technique involves the specific binding of oligonucleotides to gene targets before transcription, so forming a triple helix. Attempts are also being made to attack the transcription machinery of cells. A related method makes use of ribozymes, agents that are able to cleave pre-mRNA at specific sites. All these approaches have been used to attempt to interfere with oncogene function in malignant cell lines.

Antisense and ribozymal technologies are also being applied to the treatment of viral infections, notably human immunodeficiency virus.

Cell destruction by genetic manipulation

In the fields of cancer and infectious disease it is becoming clear that the most effective treatments will involve destruction of the target cells rather than manipulating their genes with shorter-term objectives. A variety of ingenious ideas are being explored.

One of the most promising approaches to cancer therapy is to utilize transcriptional differences between normal and cancer cells to drive the selective expression of a metabolic suicide gene to confer sensitivity to a prodrug. For example, the herpes simplex virus thymidine kinase converts gangiclovir to its active form. The idea is to target a viral vector carrying this particular gene to express itself in malignant tissue. There are two methods: transduction targeting, which relies on preferential gene delivery to actively dividing cells using a viral vector, and transcriptional targeting, which depends on unique tissue-specific or tumour-specific transcriptional elements to drive the expression of a toxic protein only in cells that contain factors capable of activating the promoter elements. These techniques have been combined and have shown considerable promise for the management of cerebral tumours in experimental animals. A number of human trials have been initiated. One of the unexplained benefits of this type of therapy is the so-called bystander effect, in which non-transduced cells are killed as well as those that have been transduced; a variety of ways are being explored to take advantage of this phenomenon.

The other general approach to killing cancer cells is to attempt to increase their immungenicity or to introduce cytokines and thus stimulate the immune system to reject the tumour without systemic toxicity. A number of experimental systems have demonstrated that the growth of tumours can be inhibited by the insertion of genes encoding different agents of this type.

Artificial chromosomes

The transfer of a large genetic unit containing a particular gene and its control regions in the form of a self-replicating molecule, that is an artificial chromosome, would have many advantages over other forms of gene therapy. It would overcome limitations on the length of DNA that can be inserted, and would, at least in theory, not be associated with the dangers of inserting material into nuclear DNA. Considerable progress has been made in this field, particularly with respect to the creation of appropriate telomeric sequences. The centromere remains a problem, although this is likely to be overcome. The major difficulty will undoubtedly be devising methods to try to transfer such large constructs into cells at the frequency required for effective gene therapy.

Reactivation of fetal genes

Interest in this form of gene therapy has been mainly confined to the field of haemoglobinopathy, although more recently there has been interest in using a similar approach for the management of Duchenne muscular dystrophy.

Most of the important inherited disorders of haemoglobin are due to mutations at the b-globin gene locus. In fetal life this only functions at a low level and the g-globin genes of fetal haemoglobin are active. The switch from fetal to adult haemoglobin production is a reflection of the neonatal decline in g-chain synthesis and the full activation of b-chain production. It has been known for many years that individuals with sickle-cell anaemia or b-thalassaemia who have persistent g-chain synthesis are partly or even completely protected from the effects of their disease. Thus attempts are being made to try to reactivate or increase the activity of g-chain production in adult life.

A variety of agents are capable of stimulating fetal haemoglobin production in adults. They include the demethylating agent 5-azacytidine, hydroxyurea, and several butyrate analogues. In a recent clinical trial it was found that the administration of hydroxyurea reduced the frequency of painful crises in patients with sickle-cell anaemia while at the same time almost doubling their steady-state concentration of fetal haemoglobin. Although this remarkable result cannot be attributed entirely to the elevation of fetal haemoglobin, it shows that this can be achieved. Similar studies in b-thalassaemia have been less successful, although in a few cases, with particular underlying mutations, there have been some spectacular successes by the use of a combination of hydroxyurea and butyrate analogues.

The protein that is defective in Duchenne muscular dystrophy, dystrophin, has a fetal counterpart called utrophin. Recent work suggests that utrophin can ‘make up' for a deficiency of dystrophin in mice with muscular dystrophy.

Molecular diagnostics

While the use of DNA probes for the identification of monogenic diseases and the genetic component of at least some multigenic disorders is already established, this approach is likely to be used much more widely in the future. Genes probes are becoming available for diagnostic purposes right across the field of infectious disease and there is considerable interest in their use as screening tools for the early identification of neoplasms, those involving the gastrointestinal and renal tracts, and the

respiratory system, for example. Currently their place in programmes of cancer prevention still has to be evaluated and compared with simpler and cheaper methods. Early studies indicate that DNA diagnosis for the early identification of colon cancer may not be much more sensitive than the use of faecal occult-blood testing, though a great deal more work is required to determine the ideal conditions required for analysing the contents of the gastrointestinal tract for mutated oncogenes. On the other hand, there is increasing evidence that the use of genetic analysis may be of considerable value in clinical oncology, both for diagnostic and prognostic purposes, and for assessing the progress of various forms of therapy.

Despite early enthusiasm the current position on the place of diagnostic DNA analysis in the clinic remains to be clarified. Its role in the cancer field is likely to be extremely important but it is, as yet, uncertain to what degree it will be of value in exfoliative studies to identify premalignant states.

Postscript

This short account has looked at some of the directions in which molecular and cell biology are moving towards clinical application. This field offers a completely different approach to the investigation and management of many of the intractable conditions that make up the major part of modern surgical practice. here seems little doubt that its new technology will gradually assume a greater importance in the clinic and therefore it is essential that clinicians in every field keep abreast of the remarkably rapid progress in molecular medicine.

Further reading

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. *The molecular biology of the cell*, 3rd edn. Garland Publishing, New York and London, 1994.

Latchman D. *Basic molecular and cell biology*, 3rd edn. BMJ Publishing, London, 1997.

Lemoine NR, Cooper DN. *Gene therapy*. Bios Scientific, London, 1996.

Lewin B. *Genes VI*. Oxford University Press, Oxford, 1997.

Raskó I, Downs CS. *Genes in medicine*. Chapman and Hall, London 1995.

8 Design and interpretation of clinical trials

*William C. Wood**

Comparison populations
Published series
Historical controls
Concurrent, non-randomized controls
Randomized controls
Types of comparisons in clinical trials
Outcomes evaluated
Measurement of results
The importance of moderate treatment effects
Are moderate effects on survival ‘worthwhile'?
Design of clinical trials
Types of trials
Avoiding bias
Avoiding random error
Obtaining large-scale randomized evidence
Simplification of the entry procedures: the ‘uncertainty principle'
Minimizing bias and random error: systematic overviews of randomized trials
Ethical issues
Interpretation of clinical trials
Quality of life analysis
Cost–benefit analysis
Further reading

The history of medicine and surgery in the last century has been one of progressive change driven by the replacement of older methods of diagnosis, evaluation, and treatment with a myriad of newer approaches and technologies. In some instances the changes reflect the opening of entire new fields such as antibiotic therapy or genetic diagnosis. Other changes reflect growing evidence that a former treatment was unnecessary or ineffective. The move from *a priori* reasoning or anecdotal evidence to prospective clinical trials as a basis for evaluation represents one of the most significant contributions of science to daily living. Clinical trials represent a tremendous investment. Not primarily (although not insignificantly) the financial investment, but the brave willingness of patients to participate in clinical trials for the ultimate benefit of their sons and daughters—both literal and figurative. This contribution is beyond calculating as are the efforts of investigators who participate in these largely volunteer activities in order to offer better therapies to their patients in the future. With so much at stake, it is essential that clinical trials be optimally designed and carefully interpreted. A critical review of any current journal demonstrates that this does not inevitably follow. A clinical trial that is improperly performed or wrongly interpreted may be worse than no trial. For ignorance is easily addressed, but false information can obscure an awareness of ignorance, sometimes for decades. Conclusions not actually warranted by the results of the study being reported are consequently the villains most sought by journal editors and reviewers.

Any recommendation to a patient, a colleague, or those third parties to the doctor–patient relationship such as economists, lawyers, insurers, or hospital managers must be supportable by evidence. This evidence may be weak or strong, not only of the efficacy of a course of action, but the costs and benefits as compared with alternative possibilities.

The randomized clinical trial is the most reliable means of assessing the safety and efficacy of alternative treatments, so guiding clinical care decisions. However, progress still is needed to promulgate the philosophy of randomized trials, and to ensure that those trials that are undertaken are designed optimally with the potential to provide the reliable information needed in order to guide clinical practice. This chapter is intended primarily for clinicians who need to apply the results of such trials in their routine practice, and who need to be able to distinguish between research that provides reliable evidence upon which they can base their clinical decision making, and that which is unreliable and potentially dangerous.

Comparison populations

Published series

Many surgical procedures and other therapies are considered standard therapy without ever having been subjected to rigorous evaluation. The term clinical trial is often used loosely to apply to any description of the results of a given operation or treatment together with associated toxicities or morbidities. In order to discuss the results of such series in some context, it is almost irresistible to compare the experience of the series with other series published by others. In surgery the common example is a surgical series previously published. This allows the most common flaw in comparative evaluation: comparing two different groups and attributing the difference in outcome to a single factor such as the ‘new procedure'. Examples abound of comparisons between different populations with different ages, different stages of disease or even different diseases, dif-ferent criteria for exclusion or inclusion, and different criteria for response or toxicity.

Historical controls

A closer comparison is with patients treated in the same institution, by the same physicians. This comparison with historical controls is even more invidious because of the belief of the authors that nothing has changed except the ‘new procedure' or drug. It is impossible to recognize drifts in stage of disease, in better results associated with improved perioperative care, or evolving exclusion criteria that can produce greatly superior results between the present series and the past series, even in the absence of any ‘new procedure' or therapy. Rarely, the results of a new approach are so overwhelming as to be incontrovertible even with historical controls. If virtually all had a bad outcome, and with the introduction of a new therapy, this is reversed, the evidence can speak for itself. Multivariate analysis compares several different actors that are associated with outcome among the comparison groups to see if the treatment is one of those associated with a difference that achieves statistical significance. Even such analyses cannot establish that an observed difference between the treatment and control groups is caused by the treatment method.

Concurrent, non-randomized controls

Some surgical series compare a single group of patients treated over a single period of time in the same institution where one surgeon uses one technique and others another. If the patients are assigned by patient selection, surgeons' choice, alternating dates, or odd or even registration numbers, it is all too easy for bias to allow different populations to be selected.

Randomized controls

The best method of achieving balance between two groups of treatment subjects available for clinical trials is a large number of participants, assigned by randomized selection. The best evidence derives from carefully controlled, randomized clinical trials, with large numbers of participants. The larger the treatment difference, the more confidence may be assigned to the treatment effect. When different randomized trials produce similar results, the evidence is even stronger.

Types of comparisons in clinical trials

The intervention to be tested may be compared with a placebo or sham procedure if there has been no clear demonstration of an effective treatment. If there is a standard procedure or treatment of proven efficacy for the condition, the control group receives standard therapy. This allows both a comparison of efficacy and a comparison of toxicity or morbidity. A comparison may be made with no intervention, termed observation only. This is a common comparison for an unproved surgical procedure, as no sham procedure would be performed in a clinical trial if it would involve risk or discomfort. If the intervention is a medication, the comparison may be with an alternative dose or schedule.

In planning a clinical trial the choice of the comparison group is crucial both from the standpoint of the scientific question to be addressed and from the feasibility of performing the trial. An example of a randomized, clinical trial that has been attempted but never completed is a comparison of prostatectomy versus radiation for the treatment of apparently localized prostate cancer. It is not feasible to perform a large trial if either the patients or the clinicians are unwilling to participate, regardless of the scientific or clinical interest in the outcome of the trial.

Outcomes evaluated

Patients are followed over a planned period of time for outcomes such as:

- survival
- freedom from disease recurrence or progression
- improvement in symptoms
- quality of life
- side-effects of the intervention(s).

It is essential to chose an outcome appropriate to the purpose of the intervention under study. A study of adjuvant systemic therapy for cancer, for example, considers whether it is better to treat all the patients with a given stage of a tumor at the time of primary therapy with systemic therapy such as chemotherapy, or treat only those who recur at the time that recurrence is diagnosed. If the therapy has any efficacy at the time of recurrence, there will almost certainly be a delay in the appearance of disease in the adjuvantly treated group. This delay is achieved at the expense of treating the entire group including both those who would ultimately fail and those who would not. This delay in average time to failure does not mean that it is better to treat all of the patients with adjuvant therapy at the time of surgical therapy rather than waiting and only treating for salvage those who ultimately fail. The important endpoint in an adjuvant trial is survival. If survival is not affected, there is no need to subject those who would not recur to unnecessary therapy.

Because many diseases have a relatively indolent course, the desire for results prior to the final course of the disease has led to an increasing use of surrogate endpoints. An example would be a rise in a serum marker such as prostate specific antigen or carcinoembrionic antigen as an event or surrogate event even before there was other clinical evidence of recurrent tumor. The follow-up intervals of the measured outcomes must also be balanced between the treatment arms or the more frequently monitored group of patients will appear to experience events, other than death, at an earlier time than the less frequently monitored group.

Measurement of results

As described above, masking a study so that results are measured without either the investigator or the experimental subject being aware of which intervention is being used (double blind) allows the most objective measurement. Masking is of particular importance when the outcome may be influenced by awareness of treatment assignment and when there is any subjective aspect of the outcome (e.g. patient comfort or level of activity). It is less important with an objective outcome (e.g. death).

The importance of moderate treatment effects

Of critical importance is that clinicians should be realistic about the scale of improvement that might be achieved from novel interventions. One of the greatest obstacles to the design and conduct of research—including the conduct of clinical trials—has been the overoptimism regarding potential improvements from new treatments, improvements that generally turn out to be short lived. For, if the size of effect that is expected is unrealistically large, then the study is likely to be too small to detect a more moderate, and yet more realistic, effect size. Realistically moderate expectations of what new treatments might achieve tend to foster the design of studies aiming to distinguish between differences in outcome that are just moderate—but still clinically worthwhile—and between differences that are so small as to be unimportant clinically. To achieve this objective, there must be strict control of bias, which requires proper randomization and statistical analysis, and strict control of random error, which generally involves large-scale randomized evidence.

Are moderate effects on survival ‘worthwhile'?

There may be some treatments that produce just a moderate effect, but have an unfavorable side-effect profile or are so costly that the scale of the benefit alongside these disadvantages seems not to be worthwile. However, moderate reductions in mortality that are likely with many currently available, relatively non-toxic interventions may be worthwhile, particularly if they are beneficial to the large proportion of patients with a particular disease and if that disease is common: consider tamoxifen in the treatment of early breast cancer, for example. To some clinicians, improving survival by about 6 per cent might not seem particularly worthwhile. However, worldwide, about 6 million women develop early breast cancer each decade, the larger proportion of whom have hormone-sensitive disease, potentially responsive to tamoxifen. Five years of adjuvant tamoxifen for 6 million women would avoid 1 million relapses, and half a million deaths from breast cancer. Currently, tamoxifen is used by more than 1 million women worldwide and the falls in breast cancer mortality being seen in some countries are probably attributable in part to the use of tamoxifen. So it is crucial that studies be designed to allow the detection of ‘moderate' differences in survival, for the absolute gains that can be achieved are substantial.

Therefore to confirm or refute any moderate treatment effects which it is realistic to expect from new approaches to care, moderate biases and random error need to be minimized, and this can be achieved through large-scale randomized evidence.

Design of clinical trials

Types of trials

A clinical trial is a planned scientific experiment performed in people either healthy or ill to evaluate the efficacy and safety of a procedure or agent for diagnosis, treatment, or prophylaxis. Trials are usually described today in terms of the phase I to IV clinical trial terminology that arose from the testing of investigational new drugs (Table 1). Most clinical trials performed in surgery follow the phase III randomized design (Table 2).

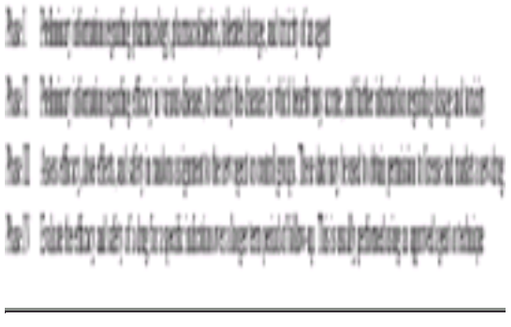


Table 1 Types of new agent (drug) trials

1	Study chairperson, collaborators, addresses, E-mail, telephone, fax numbers
2	Introduction, scientific background
3	Study design, including an algorithm or schema, method of random assignment
4	Eligibility criteria and criteria for exclusion from study
5	Treatment plan for each arm of the study (methods section)
6	Toxicities to be evaluated, modifications of therapy allowed
7	Required monitoring during treatment and in follow-up
8	Criteria for evaluation of outcomes
9	Statistical considerations including size of trial, interim analyses, monitoring committee
10	Informed consent forms
11	Data collection forms
12	References

Table 2 Elements of a clinical trial protocol

Avoiding bias

When seeking to evaluate the effects of a new treatment (or indeed an existing but untested treatment) relative to standard care, the patients receiving the alternative interventions should be as similar as possible in all respects except for the treatments that they are receiving and that are being compared. In this way, any difference in outcome between the patient groups can be attributed to the differing effects of the treatments rather than some other systematic differences between the types of patients in the different treatment groups. This is the fundamental rationale underlying randomization that avoids this kind of bias. Non-randomized studies cannot guarantee in general that the types of patients receiving the study treatment do not differ in some systematic way from those receiving control treatment with which the study treatment is being compared. Nor can the detailed collection of information in non-randomized studies on prognostic factors that could influence outcome deal with this adequately. For, some prognostic factors may be recorded or even measured differently according to the treatment group. Moreover, even if mathematical adjustment to account for known prognostic factors could be made, this would not address the possibility of systematic differences as yet unknown in prognostic factors in the groups being compared.

No prior knowledge of the next treatment allocation

In any properly randomized trial there must be no prior knowledge of the next treatment allocation. The decision to enter the patient should be made irreversibly and in ignorance of the randomly allocated treatment the patient will receive. Otherwise, bias will be introduced—for, if a clinician is aware of the next treatment allocation in a study, (s)he may think this is inappropriate for the particular patient concerned—for example, if the patient is seriously ill, this could influence the type of treatment the clinician feels that the patient might benefit most from. In addition, any prognostic factors that are recorded prior to entry should not be changed after randomization, particularly if any analyses that might be undertaken are going to use prerecorded prognostic variables.

No bias in patient management or outcome assessment

It is important to avoid, as far as possible, any difference in the care that is given over and above the study and control treatments, so that any difference in outcome can be attributed to the study group treatment rather than some other aspect of care being given differentially to the study groups. Similarly, it is important that assessment of the outcome is also made in the same way in the different study groups, and generally this can be achieved—for example, by the use of placebo controls and by blind assessment (this is less important when mortality is the main endpoint, but becomes more relevant when morbidity and quality of life are being assessed). Non-randomized comparisons and especially comparisons using a historical control group can suffer from this type of bias as mentioned above. Double blind refers to assessments of outcome when both the research subject and the investigator are blinded to the treatment that is given.

Intention-to-treat analysis of all randomized patients

Unnecessary biases may be introduced by inappropriate statistical analysis. One of the most frequent and serious errors in this context is the undue emphasis often placed on just one part of the available data—this often happens if the novel treatment seems to be particularly effective in one subgroup of patients. In addition, if certain patients are excluded from the final analysis, this can also introduce bias if those patients excluded differed in some systematic way from those patients included in the analysis. Each of these can distort the evidence and thus lead to misinterpretations of the true effects of the study treatment(s). Therefore, the fundamental statistical analysis should compare all of those originally allocated to one treatment group with all of those allocated to the comparison group—even if, in either group, some of the patients may not have received the treatment to which they were allocated.

Data-dependent emphasis on particular results is usually statistically unreliable

While it is of value to be able to answer the question reliably of whether a treatment is of benefit to a wide range of patients, it would also be useful to be able to refine this generality for application at the level of the individual patient or category of patients. In an ideal situation, in a trial that recruited a heterogeneous group of patients—for example, younger and older women with breast cancer, and women with node-positive and node-negative disease—it would be desirable to know the effects of the intervention being tested in one particular category of patient—for example, the node-negative, younger woman with breast cancer—rather than to the overall study population. However, while seemingly ideal, the direct use of trial results in particular subgroups of patients can be extremely unreliable because the statistical methods used are relatively insensitive. It is much more reliable to focus on the totality of the data. The results in different subgroups are likely, unless there are clear *a priori* reasons for thinking otherwise, to be qualitatively similar even though they might differ quantitatively. If there are *a priori* reasons for anticipating that the effects of treatment might differ in different circumstances, then a limited number of subgroup analyses may be undertaken, but these must be prespecified in the study design for analysis, and not based *a posteriori* on data arising from the study. Even in the largest trials and overviews of trials, subgroup analyses can provide misleading results; for example, earlier overviews of randomized trials of tamoxifen in early breast cancer that attempted to establish whether there was an interaction between the subject's age and the effect of tamoxifen concluded that younger women (less than 50 years) were unlikely to benefit from the drug. Subsequent overviews, when more data were available on younger women, demonstrated clearly that the earlier conclusion had been premature, and that younger women can benefit to the same degree as older women if they are given a modern tamoxifen regimen of about 5 years and if they have hormone receptor-positive disease. Not only are these subgroup analyses statistically unreliable, they have important implications for clinical decision making and patient welfare. On the other hand, subgroup analysis may be legitimate in order to generate hypotheses for future trials. The danger lies in taking such data-derived differences as evidence rather than as merely raising questions. In surgical trials it has been particularly appealing, when no difference has appeared between the treatment groups in the overall study, to seize upon a subgroup that seems to show the anticipated benefit. It must be remembered that if the overall effect was null, those subgroups not showing this putative benefit are suggesting a detriment to the intervention. Play of chance is a more likely explanation for such subgroup differences unless there is a plausible biologic explanation stated in advance.

Avoiding random error

Random errors can be avoided by studying sufficiently large numbers of patients. It is often not appreciated just how large trials need to be in order to minimize the non-systematic random fluctuations that will occur between treatment groups. This can be illustrated by a hypothetical trial in cardiology that is inadequate in size to detect the moderate-sized effect that exists, but which, by modern standards, might be considered to be quite a large study. A 20 per cent reduction in mortality (from, say, 10 to 8 per cent) is supposed to be detected in 2000 patients with myocardial infarction (1000 treated and 1000 controls). One might predict 100 deaths (10 per cent) in the control group and 80 deaths (8 per cent) in the treatment group. Even if this difference were observed, it would not be significant at $P = 0.01$, indicating that even if there is no real difference between the results of the treatments being compared, this result could relatively easily have arisen by chance. The play of chance might also increase the difference to make it significant (e.g. 110 compared with 70 deaths) or dilute and possibly obliterate the difference (e.g. 90 compared with 90 deaths) or even reverse it. So, while 2000 patients may seem a large number, the play of chance can still play an important effect and distort the true effect size that exists. Trials seeking to detect moderate differences between the comparison groups, or to demonstrate equivalence, require large numbers of subjects.

Tables are available to allow a reasonable prediction of the size of the trial required, for instance for an 80 per cent likelihood of detecting a 20 per cent improvement with a certain level of significance (a two-tailed P value < 0.05). These tables require a reasonable estimate of the rate of outcome events in the control population. Much larger numbers are required to state with confidence that treatments are equivalent in efficacy than to demonstrate a benefit. Consequently, most trials address whether or not a benefit achieves a certain level. That level is often a compromise between the benefit that would be sufficient to be of clinical utility, and the feasibility of completing a clinical trial in a few years. The study population must be available for follow-up and all those to be enrolled must be able to provide informed consent to both the treatments involved and to the trial itself.

Obtaining large-scale randomized evidence

In order to obtain large-scale randomized evidence, megatrials, or meta-analyses of all existing trials are needed. Many trials suffer from inadequate size. This sometimes results from overcomplex designs and abundant data collection. Both of these factors can deter trial subjects from entering a trial. Megatrials are effective in obtaining large numbers of patients, but rely on simplicity of design and study procedures and the involvement of many collaborators. This strategy has been effective in many areas of medicine, but the approach should be adopted more widely in clinical trials, including those in surgery. Many trials can be less complicated than they are. The need is for more large, simple, randomized clinical trials. The trials of the recent past and even present, with unduly complicated eligibility criteria and informed consent procedures and documentation together with excessive data collection on prognostic factors, compromise the potential for the trial to produce scientifically valuable results.

Simplification of the entry procedures: the ‘uncertainty principle'

Many trials have unnecessarily complex entry procedures that make the results of the trial less easily generalized, as they are derived from a narrow subset of patients. Such trials are particularly onerous for the clinician. The use of broad eligibility criteria and, in particular, the use of the ‘uncertainty principle' would facilitate the

recruitment of large numbers of patients. This principle states that the fundamental eligibility criterion is that both patient and doctor should be substantially uncertain about the appropriateness of each of the trial treatments for that particular patient. For ethical reasons, patients cannot have their treatment decided at random if they or their doctors feel certain about what treatment should be given. However, if there is uncertainty about what is the most appropriate choice of treatment, then randomization can be the most ethical and scientific way of resolving this uncertainty about how to manage the patient. Uncertainty may exist for different categories of patients, and it follows that a wide range of patients would be eligible for a trial based on this uncertainty. However, such heterogeneity is helpful because the results of the trial will be more medically relevant addressing the varying degrees of uncertainty that exist among doctors about their different patients. Where there is certainty about what is appropriate management, such patients will not be randomized. Where there is uncertainty, the patient will be randomized and evidence will be obtained which can then be used to inform future patient care.

The uncertainty principle simultaneously meets the requirements of ethics, heterogeneity, simplicity (for the clinician randomizing does not have to check complicated and lengthy eligibility criteria, basing his decision on whether there is substantial uncertainty), and maximal trial size. Moreover, with such uncertainty, informed consent can also be facilitated because the degree of informed consent required will probably not differ greatly from that which would exist outside the context of the trial in the routine clinical setting in discussing the management options with the patient.

Minimizing bias and random error: systematic overviews of randomized trials

Archie Cochrane was one of the first to emphasize the importance of reviewing the totality of randomized evidence when drawing conclusions about the effectiveness of alternative health-care interventions. Where several trials have addressed broadly similar research questions (for example, the effects of adjuvant hormonal therapy in early breast cancer in improving long-term survival and reducing the risk of recurrent disease) an overview of their results, with individual patient data, can minimize the random error resultant on small trials and also avoid the bias that might be introduced by focusing on just a subset of the trials available. Megatrials are not always practicable, but where many smaller trials have been undertaken that individually cannot provide a definite answer, pooling the data from these trials in a overview is of value not only in establishing effectiveness of different approaches to care but also in identifying still unresolved issues. Such collaborative efforts also foster partnership among research groups.

Ethical issues

The subject of the clinical trial must understand the design of the trial and whatever risks may be associated with participation. To protect both subjects and investigators (patients and doctors), hospitals and research institutions have institutional review boards or human study committees. These typically involve a panel knowledgeable about clinical trials and include physicians, scientists, nurses, and lay representatives. They review proposed trials and their consent documents, and must provide approval prior to any patient entry. For large trials, particularly those involving more than one institution, data monitoring committees review the ongoing results of the trial at predetermined points in the course of the trial. It is these groups' responsibility to stop the trial if dramatic benefit or harm for participating subjects appears. Biostatisticians are valued members of such committees to ensure that slight differences which may occur by play of chance are not given undue significance and studies stopped prematurely. The design of a large randomized trial should define the boundaries that the results must exceed for early stopping; for example, O'Brian–Fleming models, which demand very large deviance from the expected when numbers are very small at the start of a study to minimize the likelihood of inappropriate cessation by a few good or bad results.

Interpretation of clinical trials

Inferential statistical analysis suggests how likely an observed difference in outcome between groups is a reflection of the treatment effect rather than simply the play of chance. Many medical journals require a *P* value < 0.05 by a commonly accepted two-sided test of significance to allow the authors to conclude that an observed difference is significant. This means that the observed difference reported might be based not on the treatment under study but simply by the chance allocation of patients with a more favorable prognosis to the treatment group less than one time in 20. That level is still quite possible to occur by chance alone.

It is important to publish not only descriptive statistics, such as means and medians, but actual data as well. Publications describing the results of a clinical trial deserve tables of clinical parameters and tables of outcomes by each treatment or control group to allow proper evaluation and interpretation by the critical reader.

When the data from a clinical trial are analyzed, the trial may have been designed with stratification to evaluate an anticipated difference in outcome among the strata, such as men compared with women, premenopausal compared with postmenopausal women, or receptor-positive compared with receptor-negative tumors (Fig. 1 and Fig. 2). If this analysis is in response to a hypothesis stated in the study protocol and is a planned part of the analysis, it is legitimate. Searching through the data for subgroups that may show benefit, or lack of benefit, can be profitable in order to generate hypotheses for future trials. It can be grossly misleading, however, if such data-derived differences are taken as evidence. It has been particularly appealing when no difference has appeared between the treatments in the overall study, to seize upon a subgroup that seems to show the anticipated benefit. It must be remembered that if the overall effect was null, those subgroups not showing this putative benefit are showing a detriment.

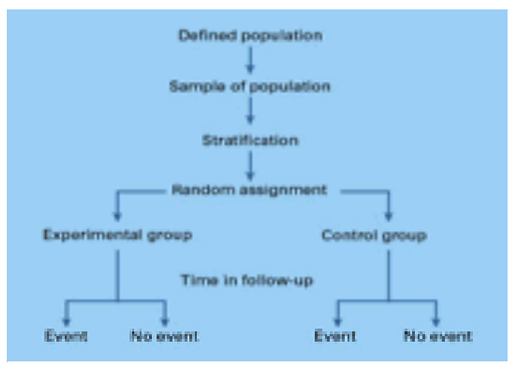


Fig. 1. Clinical trial as population cohort experiment.

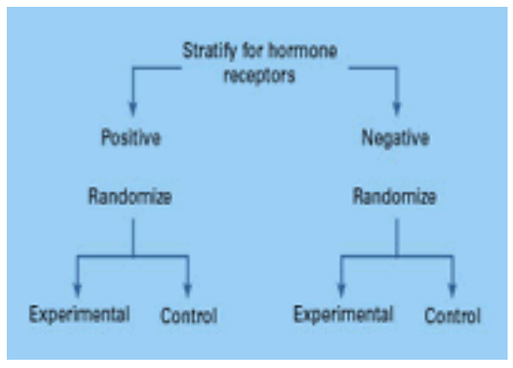


Fig. 2. Patients with breast cancer.

Quality of life analysis

Because both patients and physicians have long appreciated that certain therapies may be marginally superior in terms of disease control, but so morbid in side-effects or late toxicity that simply providing results of survival or freedom from recurrent disease are not sufficient information for clinical decision making, quality of life measures are more frequently part of the outcomes to be assessed. Although instruments for analysis of quality of life are improving, many clinicians and patients remain uncertain as to how to integrate this new information. Instruments may be as simple as analog pain scores or scales of activity. Many are complex questionnaires. It is important that these be validated before great effort is used to integrate them into clinical trials. If they are more objective, they may be easier to

analyze, but they may also be more arbitrary.

Cost–benefit analysis

This term is used increasingly not to describe the risk–benefit of a given therapy, but the financial cost of one method of treatment compared with another. As governments and third-party payers are increasingly involving themselves in health-care decision making, such financial decisions may assume increasing importance. Clinical investigators must take care that such analyses are (i) comprehensive, (ii) based on cost of care and not charges, and (iii) are long-term analyses. A less effective but inexpensive therapy may be expected to appear preferable in the short term. Placing a financial value on ‘wellness' or survival for purposes of comparison is both highly subjective and ethically treacherous.

Special thanks to Christina Davies, BM Bch, University of Oxford, for her thoughtful contributions to this chapter.

Further reading

Baar J, Tannock I. Analyzing the same data in two ways: a demonstration model to illustrate the reporting and misreporting of clinical trials. *Journal of Clinical Oncology* 1989; **7**: 969–78.

Burris HA III. Objective outcome measures of quality of life. *Oncology* 1996; **10**(Suppl.): 131–5.

Cochrane AL. 1931–1971: a critical review, with particular reference to the medical profession. In: *Medicines for the year 2000*, pp. 1–11. Office of Health Economics, London.

Collins R, Peto R, Gray R, Parish S. Large scale randomized evidence: trials and overviews. In*:* Weatherall DJ, Ledingham JGG, Warrell DA, eds, *Oxford Textook of Medicine*, 3rd edn, pp.21–33. Oxford University Press, 1996.

Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomized trials. *Lancet* 1998; **351**: 1451–67.

Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. *Journal of the National Cancer Institute* 1988; **319**: 1681–92.

Makuch R, Simon R. Sample size requirements for evaluating a conservative therapy. *Cancer Treatment Reports* 1978; **62**: 1037–40.

McLeod RS *et al.* Clinical review. Randomized controlled trials in surgery: issues and problems. *Surgery* 1996; **119**: 483–6.

Peto R *et al.* Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part I: Introduction and design. *British Journal of Cancer* 1976; **34**: 585–612.

Peto R *et al.* Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part II: Analysis and examples. *British Journal of Cancer* 1977; **35**: 1–39.

Silverman WA. *Where's the evidence? Debates in modern medicine*. Oxford University Press, New York, 1998.

Yancey JM. Ten rules for reading clinical research reports. Editorial opinion. *American Journal of Surgery* 1990; **159**: 533–9.

Yusuf S, Collins R, Peto R. Why do we need some large simple randomized trials? *Statistics in Medicine* 1984; **3**: 409–20.

9 Evidence-based approach to surgical decisions

Kieran J. O'Flynn, Miles Irving, and David Sackett

Introduction
Evidence-based surgery in practice
Finding the evidence
On line searching
Journals of secondary publication
The Cochrane Collaboration
Safety and Efficacy Register of New Interventional Procedures (SERNIP)
Appraising the evidence
Assessing a systematic review and meta-analysis
Assessing a randomized, controlled trial
Assessing a case series
Assessing a case-control study
Assessing an article on diagnosis
Assessing an article on prognosis
Conclusion
Further reading

Introduction

The concept of an evidence-based approach to surgical practice has a sound historical basis through the founder of scientific surgery John Hunter, and the first advocate of surgical outcomes research Ernest Amory Codman. However whilst embracing the concepts, neither used the term ‘evidence based'. The evidence-based movement became established in Britain with the foundation of the Centre for Evidence Based Medicine at the Nuffield Department of Medicine in the University of Oxford. This coincided with the NHS Research and Development Initiative launched by the British Government. This initiative followed a report from the House of Lords Committee on Science and Technology which revealed that the research supported by the NHS up to 1988 did not address the major issues that were most relevant to the British Health Service. Among the important observations of the committee were that there was a lack of articulation of NHS research needs and that there was a low level of research relevant to the NHS. The committee also stressed the need for a culture change in the NHS so that research findings could be transferred into practice. This report was followed by establishment of the post of NHS National Director of Research and Development supported by an annual allocation of up to 1.5 per cent of the NHS budget (approximately £450 million) to produce an evaluative culture in the NHS and to ensure that the content and delivery of care is based upon the highest quality research.

Two of the earliest manifestations of this venture were the establishment of Directors of Research and Development in every NHS Region in England and subsequently in every hospital. One of the first acts of the national director was the creation of the NHS Cochrane Centre in Oxford, which has metamorphosed to become the International Cochrane Collaboration with 15 Cochrane centres established world-wide.

Evidence-based medicine has been defined by Sackett as follows: ‘the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients'. This definition is appropriate for surgeons for it emphasizes the importance of both diagnostic and clinical expertise which is so important in surgery.

One of the biggest obstacles to the practice of evidence-based medicine is the explosion in medical knowledge which has lead to an exponential growth in the volume of medical literature. It is estimated that there are approximately 3 to 4 million new biomedical publications per year in over 20 000 journals, and Smith has estimated that the sum of medical knowledge doubles every 19 years. Such is the speed with which new developments occur that our textbooks are frequently out of date by the time they are published. Surgical journals are disorganized with an overt reliance on case series and anecdotal information. In spite of the increasing complexity of surgical management, reading patterns among both trainee and consultant surgeons vary widely and many have difficulty gaining access to a quality library and have poor access to information on the ward. Adding to these difficulties is the absence of relevant research on specific clinical problems.

Few surgeons have been trained in the skills of critically appraising research literature. The volume of surgical literature and limited time and resources mean that most clinicians cannot keep up to date with recent advance in their field. It is little wonder that continuing medical education programmes (CME) have flourished although there is little evidence that traditional CME modifies our clinical behaviour or improves patient care. The evidence-based movement, with its emphasis on self-directed clinical problem solving, has become increasingly accepted by the medical profession not least because it is seen to provide a pragmatic approach to the difficulties clinicians face in day to day medical practice.

Evidence-based surgery in practice

An evidence-based approach can be practised in any situation where there is doubt about an aspect of clinical diagnosis, prognosis, or management. The process comprises five distinct steps:

(i) the information required is converted into an answerable question;
(ii) the literature is searched with maximum efficiency for the best evidence with which to answer the posed question—in an ideal world, the best quality evidence would come from properly conducted, randomized trials and well constructed meta-analyses (see Table 1);

Level 1	Meta-analysis Randomized controlled trial
Level 2	RCTs (confidence intervals cross unity) Non-randomized control trials Retrospective cohort studies Case-control studies Convincing non-experimental evidence
Level 3	Observational studies Case series, case reports

Table 1 Levels of evidence

(iii) the evidence obtained from the literature search is critically appraised for its validity (closeness to the truth) and usefulness (clinical applicability);
(iv) the results of the appraisal are implemented into clinical practice;
(v) the clinician evaluates his/her own performance.

Finding the evidence

On line searching

Since its introduction less than a decade ago, the World Wide Web has become the most rapidly adopted communications medium ever. Using a web browser (e.g. Mosaic, Netscape, or Internet Explorer), web documents, including text, graphics, images, and sound, can be accessed at various web sites. The number of web sites has grown exponentially over the past few years. On a web page, various words and phrases are highlighted or underlined indicating that there are hypertext links that may be simply accessed by clicking the mouse button.

As the cost of journal subscriptions increase and library resources dwindle, access to the Internet with its bewildering number of web sites and medical resources,

represents a feasible way for surgeons to keep abreast of new developments in their specialty. Some of the more useful web sites are listed in Table 2.

Resource	*Web site*	*Comment*
American College of Physicians Online	http://www.acponline.org	This allows you to search the ACP Journal club and Evidence Based Medicine
The National Institutes of Health	http://www.nih.gov	Access to this site can provide consensus statements, authoritative guides to current medical issues, information about cancer and its treatment, and details of current research activities
NHS Centre for Reviews and Dissemination	http://www.york.ac.uk/inst/crd/welcome.htm	A database of systematic reviews concerned with the effectiveness and cost-effectiveness of health-care interventions
OMNI	http://omni.ac.uk/	A good place to start a web search if you're looking for evidence based material
OncoLink	http://cancer.med.upenn.edu/	This database provides detailed information on all aspects of cancer. All trials currently being conducted by the National Cancer Institute are cited and discussed
National Coordinating Centre for Health Technology Assessment	http://www.hta.nhsweb.nhs.uk/	The HTA Programme identifies important gaps in the NHS' knowledge about clinical and cost effectiveness, primitives answerable questions, and commissions high quality research
Physicians Home Page	http://php.nlm.nih.gov/index.com	This site provides unlimited access to the National Library of Medicine's (NIH) LINE database, the largest biomedical bibliographic database dating back to 1966 (subscription required)
Cochrane Collaboration Home Page	http://hiru.mcmaster.ca/cochrane/default.htm	This site contains detailed information about the Cochrane centres, the collaborative review groups and their fields of interest, register of RCTs and systematic reviews, and detailed MEDLINE strategies on many EBM issues
SCHARR Guide	http://www.shef.ac.uk/uni/academic/R-D/scharr/train.htm	A comprehensive introduction to evidence based practice on the internet
Centre for Evidence Based Medicine	http://cebm.jr2.ox.ac.uk	The website gives info details on EBM issues, a 'toolbox' assessment of materials useful for practitioners of EBM, and details on centre members

Table 2 Useful surgical resources on the Internet

Journals of secondary publication

The past few years has seen a new type of publication, the journal of secondary publication. Articles in these publications are screened for relevance to clinical practice and have passes critical appraisal filters and are methodological sound. Typically, a structured abstract of the paper occupies one page of the journal along with a commentary from an established clinician in the field, giving the bottom line. Two such journals are the *ACP Journal Club* and *Evidence Based Medicine*, the latter having a greater emphasis on recently-published surgical trials.

The Cochrane Collaboration

This collaboration comprises an ever-expanding international group of clinicians, epidemiologists, statisticians, and consumers who have come together to produce systematic reviews, which are updated each time an important new trial is published. *The Cochrane Library*, first published in 1995, is an electronic library, comprising four databases:

(i) The Cochrane Database of Systematic Reviews (CDSR);
(ii) The York Databases of Abstracts of Reviews of Effectiveness (DARE);
(iii) The Cochrane Control Trials Register (CCTR)—this database comprises a list of ongoing randomized, controlled trials (RCTs) registered with the Cochrane collaboration;
(iv) The Cochrane Review Methodology Database (CRMD).

The Cochrane Library is available in two formats, CD-ROM for windows and 3½ inch disk for windows and is updated every quarter.

Safety and Efficacy Register of New Interventional Procedures (SERNIP)

This scheme, established under the auspices of the Royal Colleges in 1995, aims to protect patients from the inappropriate application of new interventional procedures whose safety and efficacy has not yet been established. The scheme is voluntary and clinicians are encouraged to register new piloted interventional procedures. SERNIP maintains a safety and efficacy registrar of piloted interventional procedures whose safety and efficacy is the subject of ongoing assessments. Clinicians planning to undertake a piloted procedure can contact SERNIP, who can then notify them of the current status of the procedure. Those clinicians who wish to participate in an observational study or RCT can contact the co-ordinator of the assessment of the procedure with a view to joining the relevant study.

Appraising the evidence

Assessing a systematic review and meta-analysis

Any surgeon attempting to keep abreast of current developments in his field must find ways of dealing with the huge amount of published literature. One commonly used solution to this problem is to track down and appraise a review article. Reviews are frequently written by experts who work in a narrow field and by virtue of this are prone to be selective in their appraisal of the current literature and may generate incorrect conclusions and inappropriate or harmful clinical recommendations, thereby potentially delaying the introduction of efficacious treatment. Mulrow studied 50 reviews published in four major journals between 1985 and 1986 and found that there was no statement of methods in 49 papers and that the summary was inappropriate in 47. Clearly these reviews should be read selectively and critically (see Table 3).

Are the results valid?
Did the overview address a focused clinical question?
Were the criteria for article inclusion appropriate?
How likely is it that relevant studies were missed?
Was the validity of the included studies appraised?
Was the assessments of studies reproducible?
Were the results similar from study to study?
What are the results?
What are the overall results of the review?
How precise are the results?
Will the results help me to care for my patients?
Can the results be applied to my patients?
Were all clinically important outcomes considered?
Are the benefits worth the harms and costs?

Table 3 Assessing an overview or meta-analysis

A systematic review is an overview of primary studies that uses explicit and reproducible methods. A meta-analysis is a mathematical synthesis of the results of two or more primary studies that address the same hypothesis in the same way. The rationale for systematic reviews and meta-analyses is based on a number of premises. Firstly, given the huge volume of literature, information on a particular topic must be reduced into easily digestible pieces. The systematic review should separate information that is salient and critical from the insignificant, unsound, anecdotal, or simply biased. For the aspiring researcher a systematic review is an essential tool, as a structured review may prevent unnecessary duplication of research or point out areas for research that have not yet been explored.

In a systematic review the question(s) to be answered and the methods used must be clearly stated. Thorough data collection is of vital importance in the preparation of a systematic review, whether or not a meta-analysis is a part of the review. Complete identification of all relevant studies is particularly important and, where possible, the original patient data should be reassessed. Some of these studies may not have been published for reasons related to the findings. Authors are less likely to submit randomized, controlled trials with negative results for publication and when submitted the trials are less likely to be published. A comprehensive review of randomized, controlled trials relying solely on Medline searches will omit about half of the available studies. The reasons for this are multifactorial. Part of this problem results from inadequate indexing. For example it was only in 1990 that the term *Randomized Control Trial* was introduced as a descriptor term and in 1991 as a publication type. Despite this, the term may not be used consistently by indexers acting for the National Library of Medicine. Further problems arise in that some authors may not have described their methodology clearly enough to allow accurate indexing. Retrieval of high quality studies will depend on the searching skills of the authors. For these reasons authors of systematic reviews will frequently state that hand searching of the relevant literature was performed and efforts made to retrieve relevant articles that were not published in English.

The results of meta-analysis tend to be presented graphically in a fairly standard form (see Fig. 1). Commonly used terminology in meta-analysis and systematic

reviews are discussed briefly in Box 1. Each trial is represented by a horizontal line, the width of which represents the 95 per cent confidence interval of the estimate. The ‘blob' in the middle represents the point estimate of the difference between the groups. The vertical line down the middle is the line of no effect, representing a relative risk of 1 (i.e. no difference between the two groups). If the confidence interval of the result crosses the vertical line, this means that either the sample size was too small to detect a difference between the two groups or there was no significant difference between the treatments. The tiny diamond below the horizontal lines represents the pooled data from all the trials.

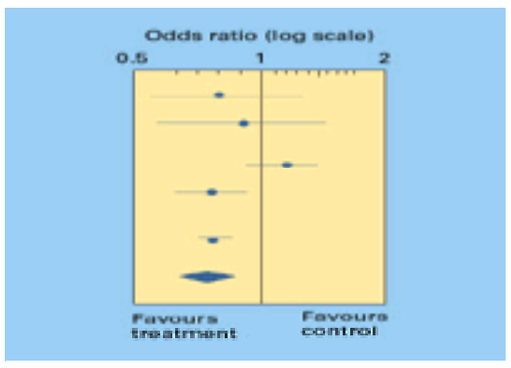


Fig. 1. Odds ratios.

Box 1 Commonly used terminology in meta-analysis

Odds Ratio Suppose that a group of cases and controls are studied to assess the development of an adverse event following a new surgical intervention. An approximate estimate of the relative risk of the adverse event associated with the procedure can be obtained from the odds ratio (OR), given as OR = ad/bc. OR are always used in case-control studies where disease prevalence is not known; the apparent prevalence depends solely on the ratio of sampling cases to control, which is totally artificial.

Study group	Adverse event		Totals
	did occur	did not occur	
New intervention	a	b	a+b
Standard treatment	c	d	c+d
Totals	a+c	b+d	a+b+c+d

Confidence intervals (CI) indicate the imprecision of the sample study estimates as population values. For example in a study with 18% events during the study period and a 99% CI 14.5–21.6, values less than 14.5 and greater than 21.6 are not excluded, but they are less likely to occur than those inside the CI. The middle half of the confidence interval is more likely to contain the population value than the extreme two quarters (i.e. the middle half forms the 67% CI). Regardless of the width of the CI the sample estimate is the best indicator of the population value. The width of the CI is determined by the sample size (larger sample sizes will give more precise results with narrower CI) and the variability of the characteristic being studied (between subjects, within subjects, measurement error etc.)

Homogeneity and heterogeneity Meta-analysis is only properly applicable if the data assessed are homogenous i.e. treatments, patients and endpoints must be similar or comparable. Statistical homogeneity means that the results of each individual trial are mathematically compatible with the results of any of the others. Clinical heterogeneity may arise as trials differ in their patient selection, disease severity, operative management, and duration of follow-up. Other problems may arise in mechanism of randomization and in the way withdrawals and losses to follow-up are handled in the original paper.

Systematic reviews have the advantage that they are generally more reliable and accurate because of the explicit methods used. Thus, large amounts of information can be quickly assessed by the discerning reader. Results of different studies can be formally compared to see if the finding are consistent and, if this is not the case, new hypotheses may be generated about particular subgroups.

On a cautionary note, the finding of some meta-analyses have later been contradicted by large randomized, controlled trials. This lends support to the argument of Eysenck and others that the results of meta-analysis are only applicable if the data summarized are homogeneous, that is treatment, patients, and end points must be similar or at least comparable. Misleading meta-analysis may also be due to the existence of publication bias and other biases that are introduced in the process of finding, selecting, and combining studies. The use of funnel plots, plots of trials' effect estimates against sample size, may be useful to asses the validity of a meta-analysis.

Assessing a randomized, controlled trial

Published opinion suggests that there is an over-reliance on case series in the surgical literature, with insufficient evidence derived from randomized trials. The purpose of randomization is to avoid selection bias and to generate groups that are comparable to each other. For this reason, evidence from a randomized, controlled trial is the soundest we can obtain about causation (whether it concerns aetiology, therapeutics, or the role of a new surgical procedure). A number of criteria must be met for the proper conduct of an RCT; that is no foreknowledge, no bias in patient management, no bias in the outcome assessment of the patient, and an intention-to-treat analysis of the data with no postrandomization exclusions (Table 4).

Are the results valid?
Were the patients properly randomised?
Were all patients who entered the trial properly accounted for at its conclusion?
Were patients and clinicians 'blind' to treatment?
Apart from the experimental intervention, were the groups treated equally?
Were the groups similar at the start of the trial?
What are the results?
How large is the treatment effect?
How precise is the estimate of treatment effect?
Will the results help me to care for my patients?
Can the results be applied to my patients?
Were all clinically relevant outcomes considered?
Are the benefits worth the harms and costs?

Table 4 Assessing a randomized, controlled trial

In a properly controlled trial, the decision to enter a patient is made in ignorance of which of the trial treatments the patient will be allocated. Known prognostic factors should be recorded before the treatment is allocated, especially if it is to be used in an analysis of treatments. There should be no bias in patient management, although this can obviously pose significant problems in surgical trials. In an intention to treat analysis (sometimes called a ‘pragmatic trial'), the randomization not only decides the allocated treatment but also how the patients data will be analysed. If a medical therapy is being compared with surgery, all the surgical programme is included (e.g. delay prior to surgery, death before surgery), and the data is analysed irrespective of whether or not the patient received the prescribed treatment. In essence, by the end of a trial there are four groups of patients (see Fig. 2) and the trial will compare the total number assigned to receive one treatment compared with the other.

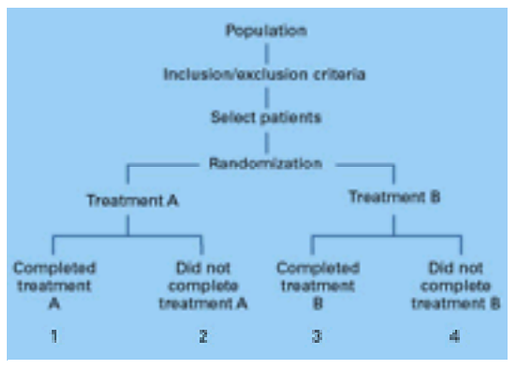


Fig. 2. Scheme for a randomized, controlled trial. An intention to treat analysis will compare groups 1 and 2 with groups 3 and 4. An efficacy analysis will compare groups 1 and 3.

Why is there a relative paucity of high quality RCTs in the surgical literature? Many surgeons argue that RCTs in surgery pose particular problems. Surgery is a craft specialty and surgeons invest a great deal of time and effort in acquiring technical skills and may be reluctant to trial a new technique that will initially be unfamiliar and may be time consuming. In most medical trials, all patients receive standardized treatments (i.e. a fixed dosage of a drug). The treatment is independent of any particular skill of the physician. The operating skill of the surgeon and his experience clearly affect the outcome and care needs to be taken that surgeons participating in RCTs are adequately trained during a prerandomization period, which should ideally last until the surgeon is fully conversant with the new technique. The surgeon should have an acceptable rate of postoperative complications from the procedure. This issue was successfully covered in the North American Symptomatic Carotid Endarterectomy trial where participating surgeons had audited low complication rates from the procedure prior to joining the study. It follows that whenever a new surgical technique which requires training undergoes evaluation, we know of no surgeons who would be willing to randomize their first few patients! These precautions, designed to reduce bias in the surgical trial may, however, lead to a restriction in the number of participating surgeons. Yet the more surgeons who participate in a trial, the more convincing the results and ideally the trial should aim at obtaining results that can be generalized. Proper randomization is pivotal in the conduct and assessment of a RCT as non-randomized trials tend to overestimate the effects of treatment.

Clearly, it is ethical to randomize patients in surgical trials when we are uncertain as to whether a specific procedure might do more harm than good. Conversely, patients cannot have their treatment chosen at random, if either they or their surgeon are already reasonably certain about what treatment they prefer. Such an attitude has lead to the widespread acceptance of radical prostatectomy by patients and doctors as the treatment of choice for localized prostatic cancer in the United States despite a lack of evidence. This has resulted in major problems in recruiting patients into a RCT designed to assess its efficacy in comparison to watchful waiting.

When surgery is being compared to a non-surgical treatment it is impossible to blind the carers or the patients at the time. Unblinded studies overestimate the effects of treatment and ideally the outcome assessment should be performed by a clinician who is unaware of the procedure, and the patient should not divulge these details to the assessor. The decision to endeavour to recruit a patient into a RCT may dramatically alter the surgeon–patient relationship. Surgeons with their extrovert and optimistic approach may have difficulty explaining to their patients the basis for their uncertainty and the need for a randomized trial. Likewise patients, used to the benign paternalism of the surgical fraternity, may not readily accept their surgeons (well founded) clinical uncertainty.

A properly-conducted RCT should have an explicit methodology section, with evidence of credible patient enrolment, clear objectives, and clear outcome measures. It is important that the outcomes being measured are valuable to doctors and/or patients. Large studies with a prestudy calculation of the required sample size are preferable to small studies which may not be of sufficient size to show treatment effect. Studies with large number of dropouts should probably be treated with circumspection, unless there is a good reason to the contrary.

A good RCT will have a clear statement adequately describing the statistical procedures used. It should be remembered that a statistical significance of $p<0.05$ is only 1:20. Any study, where the authors choose a single positive statistic out of many should probably be discarded. Treatment effects are frequently recorded as relative risk reductions (RRR). However, RRR fail to discriminate high absolute treatment effects from small ones. This is because the RRR discards the underlying susceptibility or baseline risk of patients entering randomized trials. Issues relating to treatment effects are discussed in Box 2.

Box 2 Treatment effects

Relative risk reduction (RRR) is the proportional reduction in rates of bad outcomes between the experimental and control participants in a trial (calculated as EER-CER/CER) and accompanied by a 95% confidence interval (CI). RRR fails to discriminate large treatment effects from small ones as it discards the underlying susceptibility (or "baseline risk").

Absolute risk reduction (ARR) is the absolute difference in rates of outcomes between the experimental and control participants in a trial (calculated as EER-CER) and accompanied by a 95% confidence interval (CI). ARR retain the underlying susceptibility of patients and provide more complete information than RRR.

Hypothetical RCT of 6 weeks versus 6 months treatment with anticoagulation to prevent a further DVT over two years

	Adverse events				
	6 weeks control event rate (CER)	6 months experimental event rate (EER)	Relative risk reduction (RRR)	Absolute risk reduction (ARR)	NNT Number needed to treat 1/ARR = NNT
DVT	0.2	0.10	50%	.10	10

Number needed to treat The number of patients who need to be treated to achieve one additional favourable outcome. In the example above, treating patients with anticoagulation for a documented DVT for a period of 6 months (instead of 6 weeks) will reduce by 50% (relative risk reduction) the number of patients who will get a further DVT over 2 years. The absolute risk reduction is 0.1 and the NNT is 10, i.e. 10 patients will need to be treated for 6 months to prevent one further DVT. The relation between number needed to treat and relative risk reduction is shown below.

Outcome		Risk reduction		
Control (%)	Treatment (%)	Relative (%)	Absolute (%)	No needed to treat
50	40	20	10	10
5	4	20	1	100
0.5	0.4	20	0.1	1000

That randomized, controlled trials in surgical practice are possible and valuable is shown by past experience in this area. The famous trial of operative treatment of duodenal ulcer carried out by Goligher and his colleagues clearly demonstrated the advantages and disadvantages of the different surgical procedures. More recently, Majeed *et al.* published their randomized, prospective, single-blind comparison of laparoscopic versus small incision cholecystectomy and showed little difference in terms of hospital stay or postoperative recovery between the two procedures. Sadly, however, many of the trials that are conducted in surgery are not of the highest quality. Pollock has shown that in the area of antibiotic prophylaxis in colorectal surgery there are relatively few RCTs and even fewer of good quality.

With increasing commercial pressures, there is a pressing need to assess the efficacy of new (and frequently more expensive) surgical interventions with established surgical therapies. Black pointed out, in a systematic review of 841 articles on laparoscopic cholecystectomy, that only 9 per cent were reports of randomized, controlled trials or studies using the newer observational methods. The message is clear. New technology should be carefully assessed and validated prior to its introduction into health care.

Assessing a case series

Case series abound in the surgical literature and consequently it is important to be aware of their strengths and weaknesses. Many medical or surgical interventions cannot be evaluated by a randomized, controlled trial. There is an important role for carefully-performed observational studies, which may be used to show clinical uncertainty, to generate a new hypothesis, or to modify and refine a new surgical technique. Case series have the advantages that they are that they are cheap, quick, and easy to perform. A case series will frequently provide important preliminary information, paving the way for a methodologically sound randomized, controlled trial. Some therapeutic interventions have an impact so large that observational data alone are sufficient to show it (e.g. burr holes for extra-dural haematomas). Infrequent adverse outcomes may not be detected during the timescale of a randomized, controlled trial and observational data provides a realistic means of assessing the long-term outcome of a surgical intervention, as evidenced by the recent withdrawal of the 3M hip joint from clinical use. Careful observation and reporting of the evaluation of a new surgical procedure is valid evidence, as it may reduce the learning curve of other surgeons eager to improve their skills.

While often thought provoking, particularly when describing an author's experience with a new surgical technique, case series are sometimes authoritarian and prone to

overinterpretation by the authors. Case series provide the weakest evidence for assessing the efficacy of a treatment as they are subject to uncontrolled biases, particularly in patient selection. The results quoted are infrequently supported by more robust investigation and may not be generalizable. Case series should be read with caution.

Assessing a case-control study

In a case-control study the subjects are selected on the basis of whether they do or do not have the particular disease under study. The groups are then compared with respect to the proportion who have had a history of exposure or characteristic of interest. If those cases (patients) who have had an adverse outcome were more likely to have undergone the treatment, this may constitute some evidence that the treatment might cause or precipitate the adverse outcome.

Some of the advantages of case-control studies are that they are easy to carry out and will become even easier in the future as more and more patient records are computerized. Case-control studies also provide a feasible method for the evaluation of rare diseases and for assessing rare and late adverse effects of drugs. This study design is efficient in both time and costs relative to other analytical approaches.

For a case-control study to provide sound evidence of whether there is a valid statistical link between an exposure and disease, comparability of cases and controls is essential. The major issues to be considered in the design and assessment of a case-control study are the composition of the study group and the information about exposure and disease. Strict diagnostic criteria for the disease are mandatory. The individuals with the condition must be identified. Commonly there are two sources; firstly patients treated at a particular hospital (hospital-based case-control study) and secondly selecting persons in a defined population during a given period of time (population-based case-control study). The advantage of a population based study is that it avoids bias from what ever factor led the patient to be treated in a particular hospital by a particular physician (referral bias, medical expertise, etc.) and it allows a description of a disease in the population and direct assessment of rates of disease in exposed and non-exposed individuals.

For the control group, the crucial requirement is that they are comparable to the target population of the cases and that any exclusions or restrictions made in the identification of cases applies equally to the controls and vice versa.

Case-control studies cannot provide information about the absolute risk of an event but only about the relative risk. However, case-control studies can provide useful information and are useful when the disease outcome is rare or the duration of follow-up is long.

Assessing an article on diagnosis

Information derived by history taking, physical examination, or by diagnostic test is used by all surgeons to make a clinical diagnosis. In the United Kingdom, approximately £40 billion is spent yearly in the NHS of which 4 per cent, or £1.6 billion, is spent on laboratories or laboratory testing. We use laboratories with ever increasing frequency to order tests, sometimes inappropriately, so we should be concerned about their use. Diagnostic data may also be used for several different, but interrelated, purposes. Diagnostic information is necessary to assess the severity of an illness and to predict the subsequent clinical course and prognosis of the condition (e.g. the presence of liver metastases in bowel cancer). Such information may also help estimate the potential for response to therapy (e.g. the presence of oestrogen receptors in breast cancer) and to determine the actual response to therapy.

Appraising an article on diagnosis should aim to assess the validity and applicability of the test (Table 5). The diagnostic test should, in the first instance, have been compared with the gold standard; that is a definitive diagnosis attained by biopsy, surgery, autopsy, or other accepted standard to decide its validity. The test should be applied to patients known to have the target disorder and also applied to a second group of patients known not to have the target disorder; that is all patients should have definitive verification of their disease status. The ‘true' situation is given independently by the reference test, so that the false negatives and false positive rates of the screening test can be obtained (Table 6). This can pose practical difficulties, particularly in cancer screening. It may be impossible to say if a cancer is or is not present in those patients who are negative on screening and accordingly the false negative rate for the test cannot be directly measured. Where this occurs, positive samples may be over-represented in the verified sample leading to inflated estimates of sensitivity (verification bias). Some authors may get around this problem by continuing to follow up the patients for a period of time to identify cancers that may have been present despite a negative result on screening. In this kind of assessment using follow-up, estimates of sensitivity will decrease over time as more cases of cancer are discovered in patients who had negative test results; these are then counted as false negative results. Some authors may make the erroneous assumption that all unverified patients are disease free and in situations where the preponderance of unverified patients have negative test results, this may results in an inflated estimation of the test's specificity (see below).

Are the results valid?
Was there an independent, blind comparison with a reference standard?
Was the test evaluated in an appropriate spectrum of patients (like those in whom it would be used in practice)?
Was the reference standard applied irrespective of the diagnostic test result?
Are the valid results of this diagnostic study important?
Will the results help in patient care?
Is the diagnostic test affordable, accurate, and reproducible in my practice?
Are the results applicable to my patients?
Will the test change my management?
Will my patients be better off because of the test?

Table 5 Assessing an article on a diagnostic test

	Disease present	*Disease absent*
Test positive	True positive = a	False negative = b
Test negative	False negative = c	True negative = d

Sensitivity = a/(a+c). Positive predictive value = a/(a+b)
Specificity = d/(b+d). Negative predictive value = d/(c+d)
Likelihood ratio for a positive test result (LR+) = $\frac{\text{sensitivity}}{1-\text{specificity}} = \frac{a/(a+c)}{1-(d/(b+d))}$
Likelihood ratio for a negative test result (LR-) = $\frac{1-\text{sensitivity}}{\text{specificity}} = \frac{1-(a/(a+c))}{d/(b+d)}$

Table 6 Calculating sensitivity, specificity, positive predictive value, and likelihood ratios

The primary determinants of a diagnostic test benefit are its sensitivity and specificity. Sensitivity may be defined as the proportion of people with the target disorder who have a positive test and specificity as the proportion of people without the target disorder who have a negative test (see Table 6). Sensitivity and specificity are characteristics of the test alone and will depend on the artificial cut-off point that defines a positive and negative test result (Fig. 3). For example in screening tests for prostate cancer, a serum prostate specific antigen (PSA) level of 4 ng/ml is frequently used. Reducing this cut-off level to 2 ng/ml would increase the test's sensitivity but reduces the test's specificity. This increases the false test positives and creates anxiety for the patient as a work-up is performed to decide if they suffer from prostate cancer. Tests with high values of sensitivity may be very useful as negative results effectively rule out the diagnosis. Similarly, when a test has a very high specificity, a positive result effectively rules in the diagnosis.

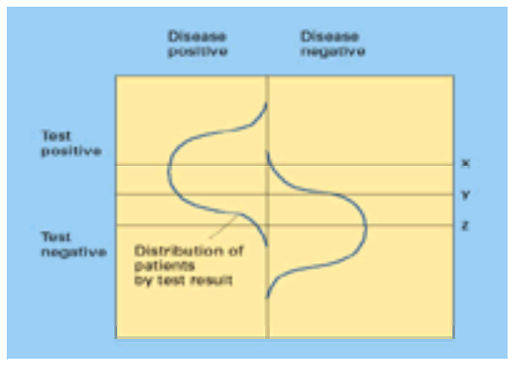


Fig. 3. As the arbitrary cut-off (y) is moved towards x or z the sensitivity and specificity of the test will be altered.

The biology of the condition will also determine the sensitivity of a diagnostic test. For example in colorectal cancer, a faecal occult blood test cannot detect neoplasms that do not bleed. Some colorectal cancers may bleed infrequently or not at all. Hence the sensitivity of a single screening test may be very low, but repeated testing could still produce a large benefit if bleeding occurred intermittently.

The likelihood ratio (LR) is sometimes cited in articles on diagnostic tests. It expresses the odds that a given finding would occur in a patient with, as opposed to without, the target disorder or condition (Table 6). With the LR above 1, the probability of the condition or disease being present goes up; when it is below 1 the probability of the disease being present goes down and when the LR is 1, the probability is unchanged. LR can also be calculated for the negative as well as positive. Likelihood ratios can be used to support or exclude a diagnosis and can be used to generate post-test probability (i.e. the proportion of patients with that particular test result who have the target disorder).

An important consideration in test efficacy studies is the extrapolation of the results. To be worthwhile, the diagnostic test should have been evaluated in a patient sample that included an appropriate spectrum of patients with mild or severe, treated or untreated disease. Diagnostic studies may be conducted by doctors with a special expertise which may not be available in the wider community and this may limit its potential usefulness. The characteristics of a diagnostic test may change over time owing to innovation or improvements in the equipment being used. For example in the Breast Cancer Detection Demonstration Project (BCDDP), conducted in the 1970s, 85 per cent of breast cancers were detected by mammography compared with 56 per cent in the Health Insurance Program in the 1960s. Some diagnostic tests will produce unintepretable results. For example in a screening study using abdominal ultrasound, the presence of bowel gas may hinder an adequate examination. Uninterpretable test results should not be ignored as it may lead to a biased assessment of the test characteristic and give an inadequate assessment of the costs of performing the test, as the test may need to be repeated or an alternative test used. In assessing a diagnostic test, the clinician must decide if the test is applicable to the patient's clinical situation and if the test result will alter patient management.

Assessing an article on prognosis

Prognosis refers to the possible outcome of the disease and the frequency with which it is expected to occur (e.g. death, survival, etc.). Frequently there are characteristics of the patient that we use to predict the patients eventual outcome (e.g. in acute pancreatitis), called prognostic factors. Prognostic factors need not cause the outcome but may be associated with it strongly enough to predict their development. Ideally, the best study design to identify the presence of a an increased risk associated with a prognostic factor is a cohort study. In an ideal cohort study, investigators follow a group of individuals who have not yet had the adverse event and monitor the number of outcome events over a long period of time.

In assessing an article about prognosis, it is important to determine if a well-defined sample of patients at a similar point in the course of their illness were included (Table 7). There are likely to be systematic differences in the results of population-based and speciality-clinic-based studies on the clinical course and prognosis of a disease as referral may be influenced by disease severity. A hospital's reputation may result from its particular expertise in a specialised area of clinical care. ‘Referral bias' results from the referral of patients with a particular condition to a clinician or tertiary centre with acknowledged expertise in the area concerned. This may increase the likelihood of adverse or non-favourable outcomes. Inception cohorts at tertiary care centres yield useful information to other clinicians who work in such settings, but it may not be possible to generalize the results to the wider population. For example in a ‘referral' population of 401 patients with extensive ulcerative colitis one study cites a risk of developing cancer of the colon as 3 per cent after 15 years to 9 per cent after 25 years, although it is unclear if the ‘inception cohort' of patients were followed from very early in their illness. Stewenius *et a1*. reported a population based study with a small increase in cancer risk in patients with severe or extensive ulcerative colitis, but the absolute risk was very low at 1.4 cases per 1000 patient years.

Are the results valid?
Was a defined, representative sample of patients assembled at a common point in the course of the disease?
Was follow-up sufficiently long and complete?
Were objective and unbiased criteria used in assessing the outcome?
If subgroups were identified, were there adjustments for important prognostic factors?
What are the results?
How likely are the results over time?
How precise are the prognostic estimates?
Will the results help me in caring for my patient?
Were the study patients similar to my own?
Are the results sufficiently important to alter management or advise the patient?

Table 7 Assessing an article on prognosis

To allow a true assessment of outcome patients should be at a similar, well-defined point in the course of their disease. The follow-up period of the study should be sufficiently long to detect the outcome of interest (e.g. late recurrence of tumour). Under ideal circumstances the authors will report on all patients entered into the study, but in practice this rarely occurs. Patients may fail to return for follow-up for a wide variety of reasons (death, ill health, relocation, etc.). The larger the number of patients whose fate is unknown the greater the threat to the study's validity. In general, fewer than 5 per cent loss probably leads to little bias and greater than 20 per cent probably threatens the validity of the study. The lower the risk of a prognostic outcome, the greater the potential effect of patients who are lost to follow-up.

The examination for important prognostic outcomes should be carried out by clinicians who were ‘blind' to the other features of these patients. This avoids bias on two counts. Firstly, a clinician who knows the patients has a prognostic factor may carry out a more detailed search for the relevant condition (diagnostic-suspicion bias). Secondly, clinicians (pathologists, radiologists, and surgeons!) may have their judgement influenced by prior knowledge of the case (expectation bias). Ideally in a published report on prognosis, the diagnostic suspicion bias will have been avoided by subjecting all patients to the same diagnostic studies. In surgical studies on prognosis, the occurrence of death is frequently a cited outcome. Judging the cause of death is very prone to error (especially when based on death certification) and assigning a cause to death may be subject to both diagnostic-suspicion and expectation biases.

Articles on prognosis in surgery frequently claim an altered prognosis for different subgroups. These groups' prognostic factors may be age, gender, extent of disease, TNM stage, or comorbidity. It should be remembered that these prognosic factors need not cause the outcome, they may only be associated with its development strongly enough to predict it. When a ‘new' prognostic factor is identified, there is no guarantee that it hasn't resulted from a ‘quirk' in it's distribution between patients with different prognoses. This initial group is called the ‘training set'. In assessing articles on prognosis, there should be a statement (in the methods section of a paper) of a prestudy intention to examine this specific possible prognostic factor (the test set). If this second independent study confirm the prognostic finding in the test set, one can feel more confident about the validity of the evidence.

One final point in studying prognosis: in most disorders, medical interventions will interfere with the natural history of the disease and it should be clear from the publication what interventions the patients underwent.

Conclusion

An evidence-based approach for surgeons has the potential to help busy clinicians with the day to day uncertainty of clinical practice. To the evidence-based approach must be added two other essential elements of clinical practice. The first is the clinical judgement and expertise required for proper diagnosis and the determination of the patients' risks and potential responsiveness to treatment. The second is the understanding and incorporation of the patients' values, beliefs, and preferences about the care they are offered.

Surgeons will continue to face an ever expanding volume of literature, the implementation of new technologies, and increased public scrutiny of their standards of care. The evidence-based approach will require new skills of the surgeon and these skills will need to be incorporated into the postgraduate surgical curriculum if the practice is to be disseminated and integrated into everyday surgical practice. More importantly, a paradigm shift will be required if the resources and skills necessary for an evaluative culture in individuals and organizations are to be developed.

Further reading

Antman EM, Lau J, *et al.* A comparsion of results of meta-analyses of randomised control trials and recommendations of clinical experts. *Journal of the American Medical Association* 1992; **268**: 240–8.

Beahr OH, Shapiro S, Smart C. Report of the working group to review the National Cancer Institue American Cancer Society Breast Detection Demonstration Project. *Journal of the National Cancer Institute* 1979; **62**: 639–99.

Black N. Surgical research. *Lancet* 1996; **347**: 1481–2. letter, comment].

Egger M, Davey Smith G. Misleading meta-analysis: lessons from ‘an effective safe, simple' intervention that wasn't. *British Medical Journal* 1995; **310**: 752–4.

Egger M, Davey Smith G, Scheider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *British Medical Journal* 1991; **315**: 629–34.

Evans M, Poliock AV. The inadequacy of published random control trials of antibacterial prophylaxis in colorectal surgery. *Diseases of the Colon and Rectum* 1987; **30**: 743–6.

Eysenck HJ. Problems with meta-analysis. In: Chalmers I, Altman DG (eds). *Sytematic reviews*. BMJ Publishing Group, London, 1995: pp. 64–74.

Goligher JC *et al.* Clinical comparison of vagotomy and pyloroplasty with other forms of elective surgery of duodenal ulcer. *British Medical Journal* 1968; **2:** 787–9.

Horton R. Surgical research or comic opera: questions but few answers. *Lancet* 1996; **347**: 984–5.

Jaeschke R, Guyatt GH, Sackett DL for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. VI. How to use an article about a diagnostic test: Are the results of the study valid? *Journal of the American Medical Association* 1994; **271**: 389–91.

Laupacis A, Wells G, Richardson WS, Tugwell P for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. V. How to use an article about prognosis. *Journal of the American Medical Association* 1994; **272**: 234–7.

Lennard-Jones JE, Meiville DM, Morson BC, *et al.* Precancer and cancer in extensive ulcerative colitis: finding among 401 patients over 22 years. *Gut* 1990; **31**: 800–6.

Majeed AW, Troy G, Nicholl JP, Smythe A, Reed MW, Stoddard CJ, Peacock J, Johnson AG. Randomised, prospective, single-blind comparison of laparoscopic versus small-incision cholecystectomy. *Lancet* 1996; **347**: 989–94.

Muirhead-Ailwood SK. Lessons of a hip failure. *British Medical Journal* 1998; **316**: 644.

Mulrow C. The medical review article:state of the science. *Annals of Internal Medicine* 1987; **106**: 485–8.

NASCET Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis.. *New England Journal of Medicine* 1991; **325**: 445–53.

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology. A basic science for clinical medicine*, 2nd edn. Little Brown and Company,1991, pp.180–1.

Sackett DL, Richardson WC, Rosenberg W, Haynes RB. *Evidence based medicine. How to practise and teach EBM*. Churchill Livingstone, New York,1997.

Sackett DL, Rosenberg WMC, Gray JAM, Hynes RB. Evidence-based medicine: what it is and what it isn't. *British Medical Journal* 1996; **312**: 71–2.

Smith R. What information do doctors need? *British Medical Journal* 1996; **313**: 1062–8.

Stewenius J *et al.* Incidence of colorectal cancer and all cause mortality in nonselected patients with ulcerative colitis and indeterminate colitis in Mamo, Sweden. *International Journal of Colorectal Disease* 1995; **60**: 330–3.

Strauss SE, Badenoch D, Scott Richardson W, Rosenberg W, Sackett DL. *Practising evidence-based medicine. Learners manual*. Radcliffe Medical Press, Oxford, 1998.

van Wairaven C, Naylor CD. Do we know what inappropriate laboratory utilization is? A systematic review of laboratory clinical audits. *Journal of the American Medical Association* 1998; **280**: 550–80.

Wilt TJ, Brewer MK. The prostate cancer versus observation trial, a randomized trial comparing radical prostatectomy versus expectant management for the treatment of clinically localized prostate cancer. *Journal of Urology* 1994; **152**: 1910–14.

10.1 Anaesthesia for surgeons

John W. Sear and Oliver J. Dyar

Introduction
Preoperative assessment
- Cardiovascular system
- Respiratory system
- Liver disease and anaesthesia
- Chronic renal failure
- Anaemia
- Sickle-cell diseases
- Diabetes mellitus
- Drug therapy and pre-existing conditions
- Inherited conditions
- The preoperative period

General anaesthetic pharmacology and practice
- Induction
- Maintenance

Local anaesthetics
- Systemic toxicity
- Allergic reactions

Specific aspects of anaesthesia
- Outpatient anaesthesia
- Management of the patient with a severe head injury requiring extracranial surgery

Postoperative pain relief
- Physiological effects of pain
- Opioid drugs
- Approaches to pain management

Perioperative fluid balance
- Abnormalities of fluid balance during surgery and the perioperative period

Perioperative complications
- Intraoperative complications
- Postoperative complications of anaesthesia and surgery
- Perioperative deaths

Further reading

Introduction

Contemporary surgical practice has developed in parallel with advances in the preoperative assessment and perioperative management of patients with increasingly complex underlying medical problems. In this chapter, we will consider some of these medi-cal problems as well as briefly discussing current anaesthetic practice and postoperative care.

Preoperative assessment

The assessment of the patient with coexisting medical disease is covered in detail elsewhere and hence only the key points are discussed here.

Cardiovascular system

Ischaemic heart disease

The clinical history will include determining the presence of angina (crucially whether on exercise or at rest) and whether the patient has suffered known myocardial infarction in the past. Evaluation of exercise tolerance may be important in a patient with angina who is undergoing major surgery and who may be subjected to stress factors such as large fluid shifts or aortic cross-clamping. The resting electrocardiogram is often normal in patients with ischaemic heart disease, but may show evidence of previous infarction (Q waves, T-wave inversion, or persistent elevation of the ST segment suggesting a ventricular aneurysm); this is significant, as the poor contractility of damaged myocardial tissue is further depressed by the effects of most volatile anaesthetic agents (see below). Further tests to evaluate left ventricular function include either multiple-gated acquisition or thallium scanning. The predictive value of these in the preoperative assessment of the patient undergoing major vascular surgery is discussed elsewhere.

In patients with clinical or electrocardiographic evidence of previous myocardial infarction, the combined effects of anaesthesia and surgery increase the risk of reinfarction in the early postinfarct period. However, by 6 months the risk falls to that seen in comparable patients not undergoing surgery, and hence elective surgery should be postponed for that period of time.

Care should be taken to maintain optimal drug therapy in the patient with ischaemic heart disease, who may be receiving b-adrenoceptor-blocking drugs, calcium-channel blockers, or nitrates. Since abrupt cessation of therapy may worsen pre-existing angina, all medication should be continued up to the time that preoperative drugs are given. Patients with angina at rest may require additional peri-operative medication, including transdermal or intravenously administered nitrates.

Postoperatively, the effects of abdominal surgery may decrease the gastrointestinal absorption of orally administered drugs, and medication may have to be provided by other routes, such as intravenous infusions, sublingual or intranasal preparations.

Tests must include plasma electrolytes and haemoglobin, the electrocardiogram, and chest radiograph. Other tests may be indicated by the presence of coexisting or predisposing factors, such as diabetes mellitus or hypertension. For elective procedures, assessment by a cardiologist is indicated for the patient whose pattern of angina has changed. In particular, the development of unstable angina or angina at rest may warrant cardiological investigation and treatment before less urgent surgery is undertaken.

Hypertension

A similar strategy is useful in the assessment of patients with hypertension. Adequacy of medication and the presence or absence of the end-organ manifestations (hypertensive heart disease with left ventricular hypertrophy and myocardial ischaemia; hypertensive nephropathy and retinopathy; cerebrovascular disease manifest as strokes or transient ischaemic episodes) should be evaluated. Current drugs used in the treatment of hypertension include thiazide diuretics, b-adrenoceptor and calcium-channel blockers, angiotensin-converting enzyme inhibitors, and other vasodilators. As in the patient with myocardial ischaemia, abrupt cessation of treatment should not occur, as it may lead to severe hypertensive episodes; this is especially important in hypertensive patients treated with the a_2-agonist clonidine.

The surgeon (and anaesthetist) is often faced with an untreated hypertensive patient presenting for elective surgery, and it is therefore important to have a strategy in deciding whether the operation is cancelled or allowed to proceed. Based on the World Health Organization definition of hypertension (systolic pressure above 160 mmHg or diastolic pressure above 95 mmHg), we can recognize four groups of patients: those with diastolic pressure less than 110 mmHg but more than 95 mmHg on at least three occasions following admission; those with diastolic pressure above 110 mmHg but less than 120 mmHg on at least three occasions following admission; those with diastolic pressure over 120 mmHg on at least three occasions following admission; and those in whom diastolic pressures are normal, but systolic pressures are elevated (often 200–250 mmHg).

There is evidence of increased risk associated with anaesthesia and surgery in hypertensive patients (diastolic pressure > 95 mmHg), who suffer significantly more silent myocardial ischaemia than do non-hypertensive patients. The risk appears to be proportionate to the degree of hypertension. It is not clear whether this increased occurrence of silent myocardial ischaemia is associated with increased cardiovascular morbidity or mortality. Recent approaches to the management of mild

hypertension include the prescribing of a single dose of a b-adrenoceptor-blocking drug as part of premedication, or the administration of preoperative clonidine.

In patients with a diastolic blood pressure of 110 to 120 mmHg, but no evidence of end-organ disease, there is a need to balance the urgency of the surgery against the potential risks associated with anaesthesia. These patients should be carefully monitored before, during, and after surgery; their blood pressure should be actively managed. If end-organ disease is present, the operation should be cancelled and appropriate treatment instituted.

Anaesthetists and physicians agree that individuals with a diastolic pressure of over 120 mmHg should have their elective surgery postponed while appropriate measures are taken to control the blood pressure. Although patients with a systolic arterial pressure of above 160 mmHg are defined as hypertensive, they are usually suffering not from hypertension but from arteriosclerosis. Whether any reduction in the perfusion pressure may lead to organ underperfusion, with resulting ischaemia and dysfunction of the heart, kidneys or brain, is debated.

When medical treatment before surgery is thought to be appropriate, the operation should be postponed for 4 to 6 weeks to allow optimization of therapy and resetting of the autoregulatory mechanisms of the body. Uncontrolled hypertension and tachycardia in the postoperative period require active and aggressive treatment: both increased blood pressure and tachycardia may cause an imbalance of coronary supply and demand, with the development of myocardial ischaemia.

Cardiac failure

Elective anaesthesia is contraindicated in the presence of acute heart failure. The combination of congestive cardiac failure with general anaesthesia will result in depression of cardiac muscle and pronounced hypokinesia. The resulting drop in cardiac output may cause profound hypotension, which further impairs coronary perfusion. Regional techniques such as extradural and intrathecal blocks may partly or completely abolish the sympathetic outflow; this reduces blood pressure because systemic vascular tone and venous return (hence cardiac output) are reduced. The beneficial reduction in afterload may be completely offset by the fall in coronary perfusion pressure.

For emergency surgery, whatever anaesthetic approach is used, intraoperative inotropic support with drugs such as dopamine, dobutamine or dopexamine, adrenaline (epinephrine), and noradrenaline (norepinephrine) may be required. A significant increase in systemic vascular resistance caused by inoconstrictor drugs can lead to further reductions in stroke volume, compounding cardiac failure.

Clinical and radiological evidence of left ventricular failure in patients with ischaemic heart disease is linked with a poor clinical prognosis. Cardiac failure should optimally be treated preoperatively, and the patient should be at their best before surgery is undertaken. Physical examination should be directed toward eliciting the signs of florid cardiac failure. Dependent oedema, hepatic congestion (with right ventricular failure), pleural effusions, basal crepitations, and the presence of a third heart sound, often manifest as a gallop rhythm (with left ventricular failure), are the classic findings. In addition to chest radiography and a 12-lead electrocardiogram, plasma creatinine and electrolyte concentrations should be measured before surgery, particularly when diuretic therapy has been used.

The anaesthetist may be asked to provide sedation or analgesia for patients with severe heart disease undergoing endoscopic procedures such as gastroscopy, cystoscopy, and colonoscopy; this may be in an attempt to reduce the patient's risks by avoiding ‘anaesthesia'. It may well be possible to avoid general anaesthesia or major neuraxial block, with their attendant haemodynamic challenges, in this way. However, the goal must be to intervene safely to minimize the sympathetic response to the procedure. Inadequate analgesia and inadequate oxygenation may expose the patient to the very risks it was intended to avoid. Whatever the type and depth of anaesthesia employed, the provision of assistance and facilities for monitoring and administration of general anaesthesia for these highly vulnerable patients must be the same as for all patients undergoing anaesthesia and surgery.

Valvular heart disease

The valves most commonly involved are the aortic and the mitral, though right-sided lesions may occur more frequently in association with intravenous drug abuse. The lesions are either stenosis (with a reduction in valve orifice size, hypertrophy of the upstream chamber, and later dilatation), or incompetence (with volume overload of the upstream chamber leading to a degree of hypertrophy, but more significantly, earlier dilatation). The stenotic lesions tend to have a longer time course, with gradual narrowing of the orifice. The symptoms of low, fixed cardiac output are similar in both aortic and mitral stenosis. Aortic stenosis is also associated with severe impairment of perfusion of the hypertrophied left ventricle, predisposing to syncope and fatal arrhythmias. Valvular incompetence may be acute (endocarditis, aortic dissection, mitral valve cord rupture) or chronic (rheumatic). Acute-onset regurgitation may cause severe and intractable heart failure.

For patients with valvular lesions, the maintenance of normal heart rate, blood pressure, and oxygenation is particularly important. However, if a deviation of these is to be caused, then the stenotic lesions are better with a lower heart rate, promoting forward flow and maintenance of systemic vascular resistance, while for the regurgitant lesions forward flow is greater with a faster heart rate and reduction of downstream resistance.

Patients with valvular heart disease present several other important problems to the surgeon and the anaesthetist. Appropriate antibiotic prophylaxis against bacterial endocarditis should be provided to cover at-risk procedures, which include oral and dental, genitourinary, and bowel. Appropriate regimens are shown in Table 1; they should be given before the induction of anaesthesia. Anticoagulant therapy (aspirin, warfarin, heparin) must be reviewed. For elective surgery, warfarin should be discontinued some days in advance, and heparin introduced as the international normalized ratio normalizes. The heparin can be stopped 4 to 6 h before surgery to reduce the risk of intraoperative haemorrhage. It can be restarted as clinically indicated some hours after surgery when haemostasis is assured.



Table 1 Recommendations for antibiotic prophylaxis for prevention of endocarditis before surgical or interventional procedures

Arrhythmias

A number of arrhythmias may be detected on electrocardiographic monitoring undertaken as part of the preoperative assessment of the elective general surgical patient.

Tachyarrhythmias

These may occur as a result of primary cardiac disease (ischaemic, valvular, or congenital accessory pathways), other systemic disease (thyrotoxicosis), or drug therapy. Drug and metabolic causes should be sought and treated. If atrial fibrillation is present, the ventricular rate should be controlled to about 100 beats/min or less, and there should be a ventriculoradial deficit of less than 20 beats. Atrial flutter should be treated before surgery with digoxin or amiodarone, or by cardioversion. Ventricular tachycardia is life threatening and, if haemodynamically compromising, should be cardioverted as a matter of emergency. For the haemodynamically stable patient, drug therapy may be possible (amiodarone), but urgent cardioversion is more usual.

Bradyarrhythmias

These may occur because of sinus node dysfunction (primary or secondary to drugs) or various degrees of block in the conducting pathways in the heart. Again, drug causes (b-blockade, digoxin) must be excluded or treated. The anaesthetist and surgeon must consider the need for preoperative insertion of a temporary (or permanent) pacing wire in patients with any type of heart block. A permanent artificial cardiac pacemaker should be inserted when the patient has complete (third

degree) heart block, the sick sinus syndrome, or symptomatic bradycardia (less than 35 beats/min).

Temporary pacing should be instituted for patients with any symptomatic heart block, Mobitz type II block, trifascicular block (indicated by axis deviation, right bundle-branch block, and prolongation of the PR interval on the electrocardiogram), or sick sinus syndrome.

Respiratory system

The induction of anaesthesia is associated with changes in pulmonary function that combine to promote arterial hypoxaemia; this occurs whether the patient is allowed to breathe spontaneously or is ventilated using intermittent positive-pressure ventilation.

In the awake patient, ventilation and perfusion both increase from the top to the bottom of the lung. In other words, the most dependent parts of the lung are both best perfused and best ventilated. Several factors combine, during anaesthesia, to disrupt the normal close matching between ventilation and perfusion in the lung; this *V*/*Q* mismatch is a major cause of hypoxia. The lung volume at the end of normal tidal exhalation is the functional residual capacity, made up of the expiratory reserve volume and the residual volume. The closing volume is that lung volume at which the airways in the most dependent areas of the lung start to close. The reduction of functional residual capacity that occurs on induction, with the assumption of the supine posture, and in relation to the age of the patient all combine to make the closing volume approach, and indeed exceed, the functional residual capacity. Because of this, closure of dependent airways occurs in most anaesthetized patients during tidal ventilation; this predisposes to atelectasis, and adversely affects the normally favourable distribution of ventilation and perfusion in the lung. Additional factors that compound these effects are smoking, other pre-existing lung disease, the site of surgery (upper abdominal far more than body-surface surgery), and the use of crystalloid infusions.

These effects primarily contribute to changes in ventilation, with the development of hypoxic areas in the lung. Under normal conditions, hypoxic pulmonary vasoconstriction redistributes blood flow away from hypoxic areas; most volatile but not intravenous anaesthetics inhibit this reflex. Thus perfusion of underventilated, dependent areas of lung and the preferential ventilation of the less well-perfused, non-dependent areas occur during general anaesthesia. Fortunately, increasing the inspired oxygen content (FIO_2) may compensate for *V*/*Q* mismatching.

Awareness of these changes and their causes enables one to predict which groups of patients and procedures are more likely to be associated with intra- and postoperative pulmonary problems. Elderly individuals demonstrate a significant reduction in functional residual capacity and sensitivity to standard doses of anaesthetic drugs. Obese patients have reduced pulmonary compliance, leading to increased work of breathing as well as a reduction in functional residual capacity. Malnourished patients not only have reduced respiratory muscle mass but also may have impaired ventilatory drive in response to hypoxia and hypercapnia. Tobacco smoking causes destruction of pulmonary elastic tissue and chronic hypersecretion of mucus. Other chronic lung diseases, especially if associated with chronic or recurrent infections, are obvious risk factors. Upper abdominal surgery, particularly when using a midline incision, predisposes to postoperative diaphragmatic splinting, resulting in atelectasis. Prolonged surgery is a further risk, particularly if intestinal distension (particularly common when nitrous oxide is employed) results in further impairment of diaphragmatic function.

Clinical features of lung disease

The cardinal features of lung disease should be assessed before anaesthesia and surgery are undertaken. The symptoms are cough, sputum production, wheeze, and dyspnoea. These symptoms should be quantified, if present, in order to assess the severity of the underlying pathology. Thus dyspnoea on climbing three flights of stairs indicates much less severe pulmonary dysfunction than dyspnoea on getting dressed. It is also vital to distinguish symptoms that may be due to disturbance in other systems. Thus dyspnoea, as well as being a cardinal symptom of pulmonary disease, may be due to inadequately treated left ventricular failure. The distinction is highly important, as the therapies, and possibly the risk associated with anaesthesia and surgery, are very different in the two groups.

Physical examination should elicit the signs of severe disease. Cyanosis (unreliably detected and best sought and/or confirmed by pulse oximetry), respiratory rate and character (rapid, shallow ventilation of the ‘pink puffer' the typically pursed lips of the chronic bronchitic imposing his or her own positive end-expiratory pressure), the use of accessory muscles and tracheal tug indicating increased work of breathing and possibly airway obstruction, should all be sought. Elevation of the jugular venous pulse should prompt a search for the parasternal heave of ventricular hypertrophy. Auscultation may reveal signs of acute infection, with bronchial breath sounds over a region of consolidation, the expiratory wheeze of asthma, or the fine, late-inspiratory crepitations of fibrotic lung disease. The heart sounds may reveal the presence of a gallop rhythm, indicative of left ventricular failure.

Certain investigations are specifically triggered by the history of preoperative lung disease: these include spirometry, which may dictate suitability for major lung resection and also enable assessment of any reversible component of the obstructive lung disease, chest radiography (which without a history of pre-existing lung pathology is an insensitive screening tool), and arterial blood-gas measurement.

Asthma

Asthma is a chronic disorder characterized by episodic, reversible bronchoconstriction. There is hyper-reactivity of the bronchial airways in response to a wide variety of inhaled substances. In the acute episode, airway obstruction is caused by smooth muscle constriction, mucosal oedema, and mucous hypersecretion with plugging. The disease is increasing in incidence and mortality, and it is the most common chronic respiratory illness worldwide. Treatment is with inhaled b_2-agonist drugs, inhaled steroids, and inhaled prophylactic therapy (e.g. cromoglycate). Preoperative assessment should elicit the frequency and severity of asthma attacks, factors provoking them, drug history, and whether the patient has ever required ventilatory support. Elective surgery should not be undertaken unless the asthma is well controlled, with peak flow values above 250 to 300 1/min, and minimal or no diurnal ‘dips' in peak flow. In patients with a marked seasonal fluctuation in symptoms, it is better not to undertake elective surgery during the poor season. For the severe asthmatic with intractable airway obstruction, especially those already on inhaled steroid therapy, a short course of systemic steroids (prednisolone 40–80 mg/day, reducing) may be useful. Where feasible, regional anaesthetic techniques should be employed. Where this is not possible, general anaesthesia may be undertaken using a minimal intervention regimen or ‘balanced anaesthesia'. In the minimal intervention, spontaneous ventilation, possibly facilitated by the use of a laryngeal mask airway, is supplemented, where appropriate, by regional anaesthetic techniques. Opioid analgesia and sedation are kept to a minimum. The aims are to avoid stimulating an irritable tracheobronchial tree and to minimize respiratory depression. Alternatively, asthmatics are managed using standard ‘balanced anaesthesia' with a few modifications. The patient receives preoperative physiotherapy and inhaled salbutamol before going to theatre. The induction dose is higher than usual to avoid irritation of the trachea during intubation. Histamine-releasing drugs such as atracurium, and morphine, are avoided.

Chronic infections

Bronchiectasis is characterized by the overproduction of purulent sputum from disorganized and dilated bronchi. Causes in the older adult population include whooping cough, measles, or tuberculosis. There may be some degree of airway obstruction, often poorly responsive to bronchodilator therapy. Patients should be admitted at least 3 days before surgery for postural drainage, physiotherapy, and treatment with antibiotics. Where appropriate, regional anaesthesia is safest; if it is not possible, patients should be intubated to allow for controlled ventilation and clearance of secretions. Endobronchial anaesthesia may prevent soiling of the non-infected lung by the infected one. Postoperative care is important, and the recently introduced minitracheostomy tube has assisted in the clearance of secretions in these patients when effective coughing is limited by pain from the operative site.

General anaesthesia is not contraindicated in younger children with cystic fibrosis, but care should be given to timing, such that surgery is undertaken at the optimal time of the child's well being. Early discharge prevents or reduces the risk of pulmonary colonization by hospital-acquired organisms.

Liver disease and anaesthesia

The presence of liver disease significantly increases the risk of anaesthesia and surgery. Acute liver disease is commonly drug or virus induced. In acute viral hepatitis, anaesthesia and surgery carry unacceptably high mortality (in excess of 95 per cent in some series); this precludes virtually all surgical intervention for at least 30 days. The most common chronic liver disease encountered is cirrhosis, which may be the end result of alcohol abuse or chronic viral hepatitis. The severity of cirrhosis relates directly to perioperative mortality.

The liver receives about 25 per cent of the cardiac output. It has two blood supplies: the hepatic vein, which supplies 65 to 75 per cent of blood flow but only 50 to 60 per cent of the liver's oxygen requirement, and the hepatic artery, which supplies 25 to 30 per cent of the blood flow but 40 to 50 per cent of the oxygen. The four main functions of the liver are metabolism of glucose, amino acids, fatty acids, and cholesterol; production of bile, which is involved in the absorption of drugs and fat from the intestines; detoxification of drugs and waste products; and synthesis of proteins (including albumin, clotting factors, and transport proteins). Alterations in these functions occur to a varying degree with liver disease and may impact on the risk posed by surgery and anaesthesia.

Effects of anaesthesia on liver function

Modern anaesthetic drugs have relatively few direct adverse effects on the liver. There are uncommon instances of immune-mediated liver damage following exposure to volatile agents. The effects of anaesthetic agents on liver function largely stem from alterations in liver blood flow. Halothane causes the greatest disruption, with reduction in cardiac output and marked reduction in portal venous and hepatic arterial blood flow, which predisposes to hepatic ischaemia. Enflurane and isoflurane, by comparison, causes much less disruption and is much less likely to be associated with ischaemia.

The liver is the principal organ of drug metabolism, and changes in liver blood flow, hepatocellular function, and both the plasma protein concentration and the extent of drug binding will all affect drug pharmacokinetics. Patients with cirrhosis show a variable degree of alteration in drug metabolism; drugs undergoing phase 1 metabolism (oxidation, reduction, hydroxylation) have a greater impairment in clearance than drugs metabolized through conjugation reactions.

Preoperative assessment

This should include evaluation of the patient's general condition, the presence of jaundice, ascites, the state of nutrition and hydration, the presence of encephalopathy, and evidence of clotting disorders. Child's grading of the severity of cirrhosis in surgical patients linked these features and hypoalbuminaemia with surgical risk. Even the lowest risk group (A) had a mortality rate of 10 per cent, rising to over 40 per cent in group C. More recently, the presence of coagulopathy (prothrombin time more than 2.5 s above the control) has been associated with a mortality rate of 83 per cent in cirrhotic patients undergoing cholecystectomy. The presence of cardiac failure is associated with mortality in excess of 90 per cent. The chief causes of death in these patients are sepsis, renal failure, coagulopathic bleeding, and hepatic encephalopathy.

Cardiorespiratory status

Patients with cirrhosis generally have reduced systemic vascular resistance, increased cardiac output, and both systemic and pulmonary shunting of blood. If the disease is alcohol induced, there may also be an alcoholic cardiomyopathy. Ascites occurs, due in part to secondary hyperaldosteronism causing marked retention of sodium and water. There is usually a mixed-pattern alkalosis (i.e., both respiratory and metabolic). Pulmonary function may be disrupted by the presence of tense ascites, causing diaphragmatic splinting, basal atelectasis, and an increase in the pulmonary closing volume. Pleural effusions may compound the problem. In addition, the pulmonary shunting referred to above will cause refractory hypoxia.

Renal status

Preoperative renal impairment is associated with adverse outcome, particularly in this group of patients. There is thus a prognostic importance in the preoperative creatinine; a value of more than 130 µmol/l is associated with a worse prognosis. Patients with cirrhosis are liable to develop the hepatorenal syndrome, the cause of which is complex. It is a peculiar functional lesion of the kidney manifesting as oliguria and avid sodium retention, with normal urine microscopy. It carries a mortality in excess of 90 per cent. Measures to prevent its occurrence include maintenance of fluid and sodium balance, avoidance of nephrotoxins, and aggressive treatment of infection.

Haemostatic status

Haemorrhage is a major cause of death in these patients. Predisposing factors include a pre-existing coagulopathy, a dilutional coagulopathy due to rapid fluid replacement, an exaggerated fibrinolytic response, and the complicating metabolic problems of hypothermia, hypocalcaemia, and acidosis following massive transfusion. In addition, there may be thrombocytopenia and defects of platelet function.

Patients with liver disease have reductions in clotting factors II VII, IX, and X, which may respond to the parenteral administration of vitamin K. Factors V and XIII, and fibrinogen, are also often reduced, and this should be treated by the infusion of fresh frozen plasma. Similarly, platelet deficiency can be overcome by platelet infusions. Factor VIII often remains normal in severe liver disease as it is synthesized by the reticuloendothelial system; a low concentration is usually indicative of disseminated intravascular coagulation. Preoperative laboratory tests should include the prothrombin time (measuring the extrinsic pathway and common components), activated partial thromboplastin time (intrinsic pathway and common components), and platelet count. A properly performed bleeding time may indicate the presence of abnormal platelet function.

Neurological status

Hepatic encephalopathy is a deterioration in conscious level of the patient with advanced liver disease; this may culminate in coma, when it is associated with a very high mortality. It may be precipitated by worsening liver function, sepsis or drugs. In particular, central depressants (particularly opioids and benzodiazepines) may precipitate encephalopathy.

Preoperative investigations

A full blood count, prothrombin time, partial thromboplastin time, creatinine and electrolytes, albumin, and transaminases are essential preoperative tests. Electrocardiography and chest radiography are also necessary. Arterial blood gases may be needed, depending on respiratory status.

The history, clinical examination, and laboratory tests can together provide an indication of the degree of risk. Increased mortality is associated with a serum albumin of below 30 g/l, the presence of infection, a white blood-cell count above 10 000/ml, treatment with more than two antibiotics, a serum bilirubin above 50 µmol/l, the presence of ascites, malnutrition, and the need for emergency surgery.

Conduct of anaesthesia

Premedication

Most anaesthetists favour an oral short-acting benzodiazepine or no premedication. Gastric acid suppression is likely to be persistent.

General anaesthesia

The anaesthetic sequence currently used by most units treating significant numbers of patients with liver disease is thiopentone (thiopental), atracurium, and isoflurane for maintenance, with ventilation to normocapnia. Isoflurane maintains hepatic oxygen supply better than other agents, while atracurium depends on breakdown by plasma cholinesterase and Hoffman degradation rather than by hepatic metabolism. Occasional patients show resistance to atracurium due to increased binding by raised concentrations of globulin. Although the enzyme plasma cholinesterase may be deficient in patients with liver failure, there is no contraindication to the administration of suxamethonium when clinically indicated, but the duration of its effect may be prolonged. Analgesia is usually provided by incremental doses of fentanyl or sufentanil. Many anaesthetists do not use nitrous oxide during prolonged surgery because of the development of bowel distension through its diffusion into the gut. Intraoperative fluids (normal saline, blood, 5 per cent dextrose) should aim to maintain normovolaemia, a haemoglobin above 10 g/dl, and normoglycaemia (4–7 mmol/l). Mannitol may be necessary to maintain a urinary output of about 1 ml/kg per hour.

Regional anaesthesia

Regional anaesthesia is advocated by some as a means of avoiding large amounts of central depressant drugs (including narcotics) in the perioperative period. Clearly, clotting abnormalities must be corrected before neuraxial or major regional block is embarked upon. In addition, thoracic epidurals, like general anaesthetics, compromise liver blood flow if the systemic blood pressure drops.

Chronic renal failure

Renal function may be impaired by parenchymal disease, or the result of ageing. Patients with impaired renal function present a number of clinical problems of importance to the anaesthetist and surgeon.

Clinical problems

Acid–base and electrolyte imbalance

Patients with chronic renal failure are unable to maintain pH, electrolyte, and water homeostasis; this may result in hyponatraemia, hyperchloraemia, and hyperkalaemia. Most patients have a moderate, compensated, anion-gap acidosis. Hyperphosphataemia occurs secondarily to the low serum calcium concentrations and increased parathormone secretion in addition to the failure to excrete phosphate. Although the uraemic patient can tolerate mild to moderate degrees of hyperkalaemia, most authorities suggest that a serum potassium of greater than 5.5 mmol/1 should be reduced before anaesthesia and surgery; this is important, as a number of intraoperative factors may further increase the plasma potassium, including hypoventilation (causing respiratory acidosis), administration of suxamethonium, and infusion of stored blood. Methods of reducing high serum potassium concentrations include glucose–insulin infusion, administration of bicarbonate, calcium resonium enema (30–60 g), and haemodialysis or haemofiltration.

Anaemia

Patients with chronic renal failure may present with a normochromic, normocytic anaemia of complex aetiology. Causes include decreased red-cell production as a consequence of reduced erythropoietin synthesis and release, bone-marrow depression by uraemia, decreased red-cell life-span, repeated blood losses during haemodialysis, and aluminium toxicity. To compensate for the low haemoglobin, the patient with renal failure has an increased cardiac output and increased amounts of 2,3-diphosphoglycerate (so shifting the oxygen dissociation curve to the right). Other factors that may exacerbate the anaemia include a deficiency of iron, folate, and vitamins B_6 and B_{12}. These problems are now less prominent with the use of biosynthetic erythropoietin, which can increase haemoglobin concentrations towards the normal.

The combination of anaemia and any associated arteriovenous fistulas will lead to a hyperdynamic circulation, with a fixed, low systemic vascular resistance and impaired circulatory reserve so that these patients are poorly tolerant of any myocardial ischaemia or sepsis.

Coagulation status

Uraemic patients may present with bleeding problems due to platelet dysfunction and thrombocytopenia, as well as decreased platelet factor III, which reduces platelet adhesiveness. Any preoperative abnormalities of coagulation (seen as a prolonged bleeding time, but unaltered prothrombin or partial thromboplastin times) should be treated by platelet transfusion, cryoprecipitate, or infusions of arginine vasopressin (DDAVP) (0.4 mg/kg)

Hypertension and ischaemic heart disease

These conditions affect 60 to 70 per cent of patients with chronic renal failure. Patients requiring treatment for hypertension (other than adequate haemodialysis) often show refractory hypertension, necessitating large doses of combination antihypertensive drugs (b-adrenoceptor blockers, calcium-channel antagonists, vasodilators, and angiotensin-converting enzyme inhibitors). Anaesthetic agents may interact with each of these, resulting in exaggerated hypotensive or bradycardiac responses. In addition, these patients show exaggerated pressor responses to laryngoscopy and tracheal intubation, and to surgical stimulation. Medication must, therefore, be optimized before any elective surgery is undertaken. Patients with chronic renal failure (especially those on haemodialysis) are also liable to suffer from accelerated atherosclerosis (secondary to the associated hyperlipidaemia), uraemic cardiomyopathy, and pericardial effusions and pericarditis; the latter two complications are usually well controlled by regular dialysis.

Central nervous system

Uraemia initially manifests as malaise, fatigue, and reduced mental ability; patients may later progress to myoclonus and fitting, coma, and death. Peripheral neuropathies are common in the lower limbs and may involve the autonomic nervous system, leading to postural hypotension, silent myocardial ischaemia, and delayed gastric emptying. Major surgery, gastrointestinal bleeding, or infection may precipitate an acute encephalopathy.

Gastrointestinal tract

Common gastrointestinal symptoms in patients with uraemia are anorexia, nausea and vomiting, gastrointestinal haemorrhage, diarrhoea, and hiccups. Patients with chronic renal failure also show a delayed gastric-emptying time, as well as increased volume and acidity of the gastric contents. Preoperative use of antacids and histamine H_2-receptor antagonists is indicated; despite this, peptic ulceration occurs in up to 25 per cent of patients on regular dialysis.

Protection of veins, shunts, and fistulas

All functioning shunts and fistulas must be protected during anaesthesia and surgery, with the sphygmomanometer cuff being placed on the contralateral arm. The cuffs of non-invasive blood-pressure monitors also must not be placed on the arm containing fistulas or shunts. The present generation of non-invasive blood-pressure monitors has reduced the indications for intra-arterial monitoring, but damage to future vascular access sites by arterial cannulation is unlikely with the new polyurethane cannulas, especially for short-term intraoperative monitoring.

Venous access should be restricted (if at all possible) to distal sites on the dorsum of the hand, with preservation of all veins in the forearm and antecubital fossa. Central venous cannulation (useful as a guide to maintaining intraoperative normovolaemia) is best achieved via the subclavian or internal jugular routes.

Anaesthesia in patients with chronic renal failure

The introduction of routine preoperative haemofiltration or dialysis has improved greatly the medical status of surgical patients with chronic renal failure. Preoperative laboratory investigations should include estimation of plasma electrolytes and creatinine, haemoglobin and haematocrit, platelets, and a clotting screen. In the diabetic patient with renal failure, preoperative monitoring of the blood glucose is imperative. In all patients, dialysis should be carried out the day before surgery to avoid acute fluid and electrolyte shifts. Preoperative blood transfusion is indicated to treat acute blood loss, or for the patient with a poor cardiorespiratory reserve and a haematocrit of less than 25 per cent undergoing major surgery. Blood should be given during dialysis to avoid fluid overload and hypertension.

All intercurrent therapy should be continued up to the morning of surgery. Because of the likelihood of delayed gastric emptying, all patients should receive antacids and/or H2-receptor-antagonist drugs. Rapid-sequence induction (modified by the use of a non-depolarizing relaxant such as rocuronium in the face of hyperkalaemia) is more commonly used in patients with renal impairment for the same reason. Care should be exercised with the use of centrally acting antiemetic drugs (phenothiazines, buty-rophenones) as they can result in prolonged sedation and extrapyramidal side-effects in the patient with chronic renal failure.

Uraemic patients may show unusual responses due to intercurrent drug therapy, reduced plasma-protein binding of intravenous drugs, and low plasma pseudocholinesterase activity. Electrolyte imbalance may affect the successful reversal of competitive neuromuscular blockade. Premedication should be chosen carefully. The orally ad-ministered, short-acting benzodiazepine temazepam offers suitable anxiolysis and mild sedation. Intramuscular opioid premedication is best avoided because of the propensity to bleeding and the increased sensitivity of the uraemic patient to these drugs.

All patients must be preoxygenated and given an adequate fluid load before the induction of anaesthesia. Induction is best achieved with the combination of hypnotic and opioid using sleep doses of either etomidate or thiopentone (thiopental). In patients undergoing major abdominal or body-cavity surgery who have a history of recent myocardial infarction or poor left ventricular function, moderate to high doses of opioid (fentanyl or sufentanil) followed by a reduced dose of hypnotic may give greater haemodynamic stability. Both regimens will attenuate the haemodynamic responses to induction, laryngoscopy and intubation, and surgical stress. Propofol (di-isopropylphenol) should be used with care in these patients because of the marked hypotension that can occur due to peripheral vasodilatation.

Neuromuscular blockade is best achieved with incremental doses of agents that rely minimally on the kidney for their elimination (atracurium, mivacurium, or vecuronium), and with careful monitoring of the extent of the blockade. Suxamethonium is not contraindicated if the serum potassium is below 5.0 mmol/l. Similarly, doses of analgesic drugs should be titrated against effect. Since active metabolites of pethidine and morphine accumulate in patients with chronic renal failure, the intraoperative use of drugs of the anilino-piperidine group (fentanyl, alfentanil, sufentanil, or remifentanil) is to be preferred. Because free fluoride ions may accumulate in patients with poor excretory function, enflurane and sevoflurane are generally avoided in patients with renal failure; supplementation of nitrous oxide should be with isoflurane or perhaps desflurane.

Neuraxial anaesthesia (spinal or epidural) is not contraindicated if any coagulopathy is first corrected, but its use may be associated with an increased risk of hypotension and infection. When any accompanying sympathetic block wears off in patients receiving epidural or spinal anaesthesia, the sudden increase in systemic vascular resistance can precipitate pulmonary oedema.

Anaemia

The normal range of haemoglobin concentration is 13.5 to 17.5 g/dl in men and 11.5 to 15.5 g/dl in women. Anaemic patients may be asymptomatic, although lassitude and decreased exercise tolerance are common features. The patient may be pale. If anaemia is moderate or severe, physical examination may reveal cardiomegaly and functional heart murmurs. Tissue oxygenation is dependent upon arterial oxygen content, capillary blood flow, and the position of the oxyhaemoglobin dissociation curve.

Oxygen content

This is determined by the partial pressure of oxygen and the percentage haemoglobin saturation according to the following equation: Cao_2 = (Hb × 1.34 × Sao_2) + (Pao_2 × 0.0225) ml/100 ml blood; where Hb = haemoglobin content (g/dl), Sao_2 = oxygen saturation, Pao_2 = partial pressure of oxygen, and 0.0225 = ml oxygen dissolved/100 ml blood/kPa oxygen tension.

In normal arterial blood, haemoglobin is almost fully saturated (96–98 per cent) in the patient breathing room air, and carries 18 to 21 ml oxygen per 100 ml blood. In the patient with anaemia, the decreased oxygen content (due to the decreased amount of haemoglobin) can be partially offset by increasing the inspired oxygen concentration.

Capillary blood flow

This varies with cardiac output, systemic vascular resistance, and blood viscosity: with anaemia, although capillary blood flow increases because of reduced viscosity, this is offset by the reduction in oxygen content of the blood. Hence any factor causing a further fall in oxygen delivery, such as blood loss or myocardial depression, will predispose to tissue hypoxia.

Oxyhaemoglobin dissociation curve

At rest, basal oxygen consumption is about 250 ml/min; this is achieved through extraction by the tissues of about 5 ml oxygen per 100 ml blood. Mixed venous saturation is approximately 75 per cent and the mixed venous Po_2 about 5.3 kPa. The oxyhaemoglobin dissociation curve will be shifted to the left by hypothermia, reduced amounts of 2,3-diphosphoglycerate, or alkalosis; this leads to further reduction of Pvo_2 at 75 per cent saturation and impaired tissue oxygenation. A right shift of the oxyhaemoglobin dissociation curve will allow oxygen to be released to the tissues from haemoglobin, but only at higher $Paco_2$.

Provided that cardiac output, and therefore tissue perfusion, remain unaltered, anaemia causes a reduction in both capillary and tissue Pao_2. In patients with chronic anaemia a number of important physiological compensatory mechanisms come into play. Firstly, the concentration of 2,3-diphosphoglycerate in red cells is increased, shifting the oxyhaemoglobin dissociation curve to the right. In addition, cardiac output is increased, due to reduced blood viscosity; this effectively reduces peripheral vascular resistance. There is also capillary dilatation in response to the reduced oxygen content of the blood.

Packed cell volume versus viscosity

The relation between haemoglobin concentration, blood viscosity, and oxygen flux in the tissues is shown in Fig. 1. The normal physiological response to anaemia is an increase in cardiac output, primarily as a result of the decrease in systemic vascular resistance and an increase in stroke volume. Although arterial oxygen content increases linearly as packed cell volume increases, there is also an accompanying increase in viscosity, which reduces capillary blood flow. As shown in Fig. 1, there is an optimal packed cell volume at which the balance between viscosity and arterial oxygen content results in the maximum volume of oxygen being transported to the tissues in a given time. The greatest relative oxygen-transport capacity occurs at a packed cell volume of about 30 per cent and a haemoglobin concentration of 8 to 9 g/dl. As the haemoglobin falls further, global oxygen delivery can only be maintained if there is an increase in heart rate. In turn, this will increase both myocardial work and myocardial oxygen consumption. In the absence of coronary arterial disease, the normovolaemic patient can compensate down to a haemoglobin of about 8 g/dl.

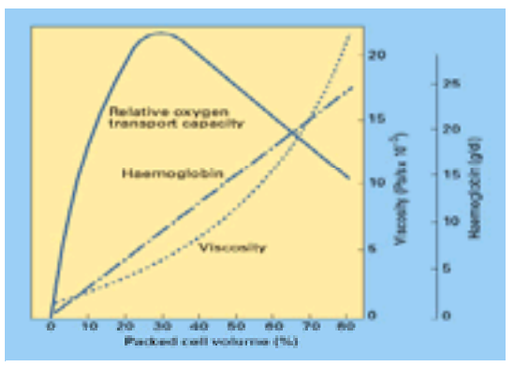


Fig. 1. The relations between haemoglobin concentration, blood viscosity, and oxygen flux in body tissues.

Effects of anaemia on anaesthesia and surgery

There is no substantial evidence to show that normovolaemic anaemia increases the morbidity associated with surgery, although risks are increased in the presence of other factors that may impair tissue oxygenation, such as pre-existing lung disease or hypovolaemia. Similarly, there are insufficient data to suggest that anaemia is associated with poorer wound healing. It has been suggested that healing of colonic anastomoses is optimal if the haemoglobin is maintained in the range 10 to 12 g/dl.

Since inhaled anaesthetic agents have a reduced solubility (15–25 per cent) in anaemic blood, the concentration of anaesthetic in the blood will increase more rapidly than normal. Although this increases the rates of onset of, and recovery from, anaesthesia, there is also the risk of overdosage or toxicity. The depressant effect that these drugs have on the cardiovascular system may be exaggerated in the anaemic patient

Respiratory alkalosis (due to excessive mechanical ventilation) and hypothermia shift the oxyhaemoglobin dissociation curve to the left, reducing tissue oxygen release, while intraoperative hypovolaemia (secondary to blood loss) may decrease cardiac output and cause splanchnic vasoconstriction. Blood lost during surgery must therefore be replaced promptly to restore the circulating blood volume. Pain and shivering in the postoperative period increase systemic vascular resistance and result in tissue hypoxia. Thus, good pain therapy as well as oxygen should be administered in the immediate postoperative period. Attention must be paid to cardiovascular function, avoiding hypovolaemia by appropriate administration of crystalloid, colloid or blood, and by treating arrhythmias or reduced cardiac output promptly.

Although there are potential advantages to maintaining the perioperative haemoglobin at around 9 g/dl, blood transfusion itself is associated with significant risks of perioperative morbidity and mortality; these risks include the following:

1. immunosuppression of a number of effectors in the immune pathways (including increased production of prostaglandin E_2; decreased release of interleukin 2; decreased production of tumour necrosis factor, interferon-g, and granulocyte/macrophage–colony-stimulating factor; a decreased T4/T8 ratio; diminished natural killer-cell activity; suppressed antigen-presentation function by macrophages), but it is not clear whether these effects are due to transfusion itself, or to other factors related to the transfusion;
2. increased incidence of postoperative infection following allogenic transfusion to patients undergoing abdominal surgery;
3. reduced duration of tumour-free survival after cancer surgery;

4. risks of transmission of human immunodeficiency virus, hepatitis B and C; and non-haemolytic and haemolytic transfusion reactions.

Sickle-cell diseases

These are caused by the inheritance of the sickle-cell gene, either alone or in combination with another haemoglobinopathy. The most common genotypes are homozygous sickle-cell disease (SS), sickle-cell trait (SA), sickle-cell haemoglobin C (SC), and sickle-cell thalassaemia (S thal). Under conditions of reduced oxygen tension, sickle haemoglobin undergoes gelation, altering the malleability of the erythrocyte, which assumes an elongated, sickle shape. These misshapen cells cause sludging of the blood in the capillaries, leading to increased viscosity, thromboembolic episodes, and cell haemolysis. To reduce the risk of sickling during surgery, elective operations should be performed only when the patient is not having a crisis and is free from infection. Safe and simple general anaesthesia in these patients requires adequate oxygenation, ventilation to normocapnia, maintenance of the circulating volume, and good postoperative care. The induction of hyponatraemia may reduce the risk of sickling, as may deliberately raising the pH. In patients with severe sickle-cell disease scheduled for major surgery, preoperative transfusion to decrease the absolute percentage of HbS to less than 40 per cent reduces the risk of vaso-occlusion as well as suppressing the bone marrow, thereby resulting in decreased production of sickle cells.

Preoperative assessment should include the prescribing of adequate premedication, as anxiety may precipitate crises due to the associated vasoconstriction. The use of 50 per cent oxygen in nitrous oxide, supplemented by isoflurane, is the current technique of choice. Intraoperative monitoring should include oximetry and capnography. The use of automated, non-invasive blood-pressure cuffs has been questioned because of a possible increase in problems due to stasis and sickling while a sphygmomanometer cuff is inflated. Postoperatively, arterial hypoxaemia is common, and this may be exacerbated by the depressant effects of postoperative analgesia. Oxygen should be given for 24 to 48 h after surgery. Local hypoxaemia and acidosis must be avoided, and tourniquets should not be applied for limb surgery.

Although patients with the sickle-cell trait are less at risk, they can still sickle, although at a lower arterial oxygen tension. Again, adequate oxygenation is essential throughout the perioperative period.

Diabetes mellitus

Between 2 and 2.5 per cent of the population are known to be diabetic and there may be an equal number of undiagnosed diabetic individuals. Diabetes affects about one in six patients over the age of 65 years, and one in four over the age of 85 years.

The perioperative mortality of diabetic patients varies between 4 and 13 per cent; most deaths result from the diseases associated with diabetes, including atherosclerosis, nephropathy, hypertension and ischaemic heart disease, and infection.

Hyperglycaemia may result from a deficiency of insulin, insulin resistance, and accelerated hepatic glucose production due to increased glucagon (non-insulin-dependent diabetes mellitus, type II), or from a lack of insulin (insulin-dependent diabetes mellitus, type I). Ninety per cent of diabetic patients have type II disease: they are often obese and have normal or elevated plasma insulin. These patients are not prone to ketosis, but they may develop non-ketotic hyperosmolar coma.

The diagnosis of diabetes is made on the basis of a fasting venous blood glucose above 140 mg/dl (7.8 mmol/1), and an abnormal glucose tolerance test: blood glucose concentration in excess of 200 mg/dl (11.0 mmol/1) 2 h after ingestion of 75 g of glucose.

Perioperative management

Management of the diabetic patient does not only depend on the careful assessment of the state of glycaemia, but requires attention to the other sites of end-organ disease.

Cardiovascular system

There are three main complications affecting diabetics.

1. Hypertension This is seen in between 30 and 60 per cent of diabetic patients, and may be part of the so-called syndrome X comprising hypertension, hyperlipidaemia, and insulin resistance. Many of these hypertensive patients will have an associated autonomic neuropathy; they are at increased risk of gastroparesis and consequent aspiration, as well as sudden cardiorespiratory arrest. During anaesthesia, the diabetic patient with hypertension may be prone to hypotension, especially if the blood pressure was poorly controlled before the operation. The combination of preoperative hypertension and intraoperative hypotension is associated with an increased incidence of postoperative renal and cardiovascular complications.

2. Coronary arterial disease This is the most common cause of death in elderly diabetic patients. Because of the functional denervation of the cardiac sympathetic afferent fibres, many cases of infarction are ‘silent'.

3. Cardiorespiratory autonomic neuropathy This serious complication of diabetes occurs in over 50 per cent of patients, and is an indicator of poor prognosis both long term and for cardiorespiratory adverse events in the immediate perioperative period. Apart from silent ischaemia, the neuropathy may lead to orthostatic hypotension and cardiovascular instability during anaesthesia. One component of the neuropathy is a deficiency of catecholamine release associated with the induction of anaesthesia and during tracheal intubation; this may result in significant hypotension and bradycardia, especially in the presence of large doses of opioid drugs such as fentanyl. Another aspect of the cardiac neuropathy is prolongation of the QT interval, and analysis of the associated QT dispersion has been suggested as a useful, non-invasive marker for estimating the perioperative risk of cardiac arrhythmias and sudden cardiac death.

There is also an abnormality of the hypoxic drive mechanisms, leading to episodes of hypoxaemia and associated cardiac ischaemia, as well as impaired respiratory protective reflexes against inhalation of gastric contents. Most diabetic patients will also have a mild pulmonary restrictive defect due to reduced elastic recoil and a reduced carbon monoxide pulmonary diffusing capacity. Although this is not a significant feature in the normal course of events, the residual effects of general anaesthesia combined with the effects of major abdominal surgery can result in major postoperative respiratory complications.

Renal system

Diabetic nephropathy results from the combination of diabetic glomerulosclerosis, chronic pyelonephritis, obstructive uropathy, and accelerated renovascular atherosclerosis. During anaesthesia, metabolic decompensation leads to hyperglycaemia, an osmotic diuresis, and marked reduction in the intravascular volume. As a result, renal perfusion falls and this may lead to acute perioperative renal failure, especially if the alternating combination of hypotension and hypertension is present.

Hypoproteinaemia is also seen in these patients (as a sequel to the microalbuminuria); this will lead to increased free-drug concentrations (and hence the propensity to adverse effects) for drugs such as the opioids, benzodiazepines, and local anaesthetics.

Gastrointestinal system

Delayed gastric emptying occurs in 20 to 30 per cent of all diabetic patients, usually in association with an autonomic neuropathy. Prescription of metoclopramide preoperatively is indicated.

Stiff joint syndrome

This is seen in about 30 per cent of young diabetic patients. It arises from the glycosylation of tissue proteins secondary to chronic hyperglycaemia and may affect many joints of the body. Of importance to the surgical patient is the possibility of limited mobility of the temperomandibular and atlanto-occipital joints.

Metabolic response to anaesthesia and surgery

Surgery produces a glycaemic stress response, the magnitude of which is dependent on the site and duration of surgery; the severity of the underlying pathology; and the type of diabetes. The main catabolic effects are due to four hormones, adrenaline (epinephrine), noradrenaline (norepinephrine), glucagon, and cortisol. There is also an associated relative or absolute deficiency of insulin. The resulting hyperglycaemia is due to the combination of increased glycogenesis and decreased

glycolysis, coupled with ketosis caused by lipolysis and ketogenesis, as well as protein catabolism. As a result, the diabetic patient is very unlikely to develop intraoperative hypoglycaemia unless treated with large doses of insulin.

It is now recognized that acute hyperglycaemia during surgery may result in a number of undesirable sequelae: (a) altered host defence mechanisms, (b) extracellular dehydration and electrolyte imbalance, (c) intracellular dehydration, and (d) impaired wound healing. The blood glucose is best maintained in the range 4 to 8 mmol/l, with 6 mmol/l being optimal. There is now much evidence linking chronic hyperglycaemia and the end-organ pathology responsible for the long-term complications of diabetes. There are no findings that perioperative control of glycaemia is of importance in this respect; but hypoglycaemic symptoms can be masked in the sedated and anaesthetized patient.

Treatments

There are three main strategies for the treatment of diabetes mellitus.

1. Patients maintained by dietary treatment alone require no special medication before surgery, although some patients may need insulin therapy at least transiently during the postoperative period.
2. Oral hypoglycaemic agents in common usage are the sulphonylureas and biguanides. Drugs in the former group (which includes tolbutamide, chlorpropamide, glibenclamide, and glipizide) enhance insulin secretion by the pancreas. The pharmacological characteristics of these drugs are shown in Table 2. Because of the longer duration of action of chlorpropamide, the drug should ideally be stopped 24 to 36 h before surgery. The mechanism of action of the biguanides (e.g. metformin) is ill defined, but they appear to decrease gluconeogenesis and increase peripheral, extrasplanchnic glucose utilization by a shift from oxidative to anaerobic metabolism. As such, they have been associated with the development of lactic acidosis, especially in patients with renal impairment. Other oral hypoglycaemic drugs include acarbose, which is an inhibitor of intestinal a-glucuronidase, and acts to delay the digestion and absorption of starch and sucrose.

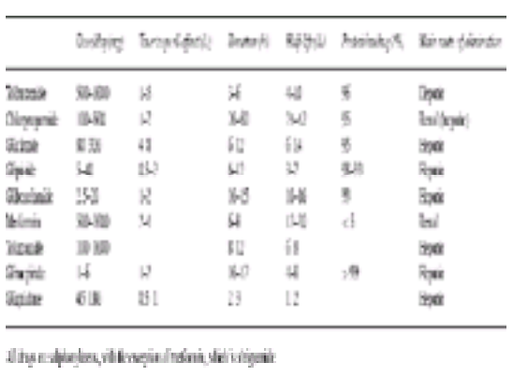


Table 2 Normal daily dosages, duration of effect, elimination half-life, protein binding, and main site of metabolism for commonly used oral hypoglycaemic agents

3. About 25 per cent of all patients with diabetes require insulin, which has a number of different physiological effects. As well as its major effect of increasing peripheral uptake and utilization of glucose, and reducing glycogenolysis, it also decreases gluconeogenesis, increases formation of protein from amino acids, and increases synthesis of glycogen. There is an increasing use of human insulin derived from recombinant DNA sources, or semisynthetically from porcine insulin. Insulin preparations vary in their times to onset, peak and duration of action, as well as in their origin. Short-acting preparations are soluble, while the longer-acting insulins are formulated either as a zinc suspension, or as a suspension with protamine sulphate (isophanes) (Table 3).

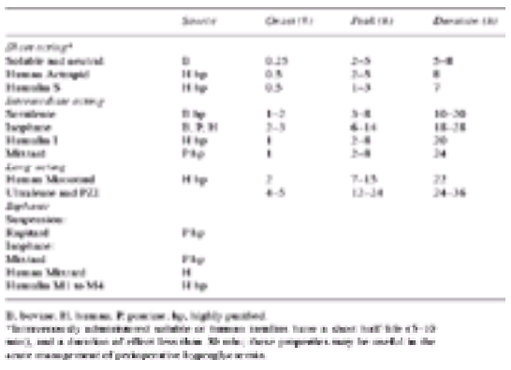


Table 3 Types of insulin preparation used in man, with their times to onset of effect, time of peak effect, and duration of effect

Preoperative management

Laboratory tests should include full electrolytes, urea, and creatinine, blood count with white-cell differential, chest radiograph, and electrocardiogram. Many patients will be receiving other coincidental therapy, which should be optimized before surgery (e.g. calcium-channel blockers, angiotensin-converting enzyme inhibitors, other antihypertensive agents, and digoxin). Plasma and whole-body potassium may be decreased in patients with uncontrolled diabetes, and additional supplementation needs to be given before surgery.

Perioperative management

Elective surgery in the non-insulin-dependent diabetic patient

The acceptable upper limit for the fasting blood glucose in the non-insulin-dependent diabetic is 7.8 mmol/l. In all patients, the oral hypoglycaemic agents should be discontinued on the morning of surgery and recommenced whenever possible the next day. The exceptions are those patients receiving chlorpropamide or metformin, where the longer half-lives of the drugs make it sensible for them to be discontinued on the day before surgery.

Overall there is controversy over the management of this group of patients. They have a normal cortisol response to surgery, which will have the effect of increasing the blood sugar. However, most non-insulin-dependent diabetics have sufficient endogenous insulin secretion to carry them through minor surgery (i.e., surgery not involving penetration of a body cavity or transection of a major limb bone) and so avoid the need for transient insulin therapy. Indeed, some studies have concluded that in the non-insulin diabetic undergoing minor surgery, there is little need for insulin provided that glucose-containing solutions are withheld. For major surgery, exogenous insulin will usually be necessary to maintain glucose homeostasis. A threshold blood glucose of 11 mmol/l has been proposed as a suitable trigger for insulin therapy. Recent studies by Raucoules-Aime and colleagues have shown that for type II diabetic patients, intravenous bolus doses of insulin every 2 h are as effective as infusions in maintaining normoglycaemia.

Elective surgery in the insulin-dependent patient

Various regimens of perioperative management for the patient with insulin-dependent diabetes have been suggested. These include:

1. a *laissez-faire*, minimal intervention approach;
2. a ‘split-normal-dose' regimen where the patient receives 1/4 to 1/2 the usual morning dose of insulin plus 5 per cent dextrose solution at a rate between 100 and 200 ml/h;
3. the intensive glucose–insulin–potassium regimen as described by Alberti and colleagues;
4. (4) the regular intravenous administration of boluses of insulin or an infusion of insulin with the simultaneous infusion of 5 per cent dextrose (with constant infusion of 100 ml/h of 5 per cent glucose, and adjustment of the rate of infusion of the insulin based on frequent blood glucose estimations).

With all these methods of diabetic control, frequent blood glucose determinations are mandatory. If the patient is normally maintained on a long-acting evening dose of

insulin, this should be given together with the meal on the evening before surgery to achieve basal normoglycaemia and then intraoperative intravenous insulin alone used to maintain the peroperative plasma glucose at between 4 and 8 mmol/l.

Fluid replacement should be given as normal saline. A higher blood glucose (and therefore higher insulin requirements) are found in patients receiving either 5 per cent dextrose or Hartmann's solution (Ringer's lactate), the latter being an important gluconeogenic precursor, especially in the starved or catabolic patient.

The same regimen should be used for postoperative diabetic control. Use of sliding scales based on urinary glucose content has no justification today. If a regimen of no insulin until the patient commences oral feeding is adopted, there is a high risk of them developing ketosis.

For patients undergoing surgery whilst receiving local anaesthesia alone (i.e., without sedation), there are modifications to the above regimens. For type I insulin-dependent diabetes, most patients cope well if they receive between 50 to 60 per cent of their normal intermediate or long-acting insulin, with a variable-rate infusion or intravenous boluses of a short-acting insulin if the blood glucose concentration exceeds 300 mg/dl (16.7 mmol/l). For type II patients, the best policy is to omit their oral medication before surgery, and then give it on return to the ward before they receive their normal diet.

Drug therapy and pre-existing conditions

Cardiovascular agents

Many patients presenting for surgery will be receiving long-term therapy for the treatment of hypertension or atherosclerotic heart disease. Continuation of all these agents into the perioperative period has shown a significant improvement for cardiovascular morbidity and mortality by protecting the myocardium against hypotension and ischaemic episodes.

However, there are two areas of current interest. There is debate over the efficacy of calcium channel-blocking drugs in providing perioperative protection against myocardial ischaemia (see Chapter 12.2). The continuation of angiotensin-converting enzyme inhibitors on the morning of surgery has resulted in abnormal responses to the induction of anaesthesia and poor tolerance of hypovolaemia in some patients.

It is the authors' view that both types of drugs be continued up to the time of induction of anaesthesia. Consideration should be given to the possible addition of a single dose of a b-adrenoceptor-blocking drug (such as atenolol) with the premedication in patients with severe coronary arterial disease treated with calcium entry-blocking drugs. In patients on angiotensin-converting enzyme inhibitors, preoperative fluid loading with 500 ml of crystalloid or colloid will attenuate the marked hypotensive vascular response to the induction of anaesthesia seen in some.

Steroid therapy

In the normal individual there is a clearly defined neuroendocrine response to surgery and other stresses. Serum corticotropin and cortisol increase rapidly after surgery, with a return to baseline by 24 to 48 h postoperatively. The magnitude of this response correlates well with the extent of the operation. The incidence of acute total adrenal insufficiency after routine surgery is low (0.01–0.7 per cent), but partial insufficiency may occur, because of:

1. pre-existing or previously undiagnosed disease;
2. acute destruction of the adrenal gland (due to massive retroperitoneal bleeding secondary to anticoagulant therapy or induced thrombocytopenia);
3. drug-related effects, such as increased metabolism of cortisol due to the coadministration of enzyme inducers [e.g. phenytoin, phenobarbitone (phenobarbital), rifampicin]; decreased synthesis due to drugs such as ketoconazole, metyrapone, mitotane, trilostane (or long-term administration of the anaesthetic agent etomidate); and interference with corticotropin action by the antitrypansomiasis drug suramin;
4. blockade of the peripheral glucocorticoid receptor by the abortifacient mifepristone.

Patients at risk of acute adrenal insufficiency can be divided into two categories on the basis of the reason for their receiving steroid medication:

1. for the treatment of chronic autoimmune or inflammatory disease, or chronic lung or gastrointestinal disease;
2. for replacement secondary to local pathology (for example, following hypothalamic irradiation, pituitary excision for adenoma, or adrenalectomy for management of metastatic disease).

Patients treated with corticosteroids for autoimmune or inflammatory disease

Although it is well recognized that pharmacological doses of glucocorticoids are needed as a supplement in place of the normal response to surgery, the original recommendations of an approximate fourfold increase in drug administration are clearly excessive. Such large doses may result in their own adverse side-effects, such as decreased rates of tissue repair, decreased glucose tolerance, and an increased susceptibility to infection. Preoperative stimulation tests will identify those patients unable to mount a response due to exogenous steroid suppression of native adrenocortical synthesis. In these individuals, maintenance therapy must be supplemented with additional doses of hydrocortisone. The normal patient undergoing a major stress will not secrete more than 300 mg cortisol in 24 h, and hence doses of this order will be required in individuals where the adrenal gland is either suppressed or absent.

Current regimens for supplementation in patients undergoing major abdominal or body-cavity surgery are 25 mg cortisol (or its equivalent) given at the time of induction and a further 100 mg cortisol given by continuous infusion over the next 24 h (or hydrocortisone, 100 mg intravenously 8-hourly). For minor surgery, the most appropriate regimen is the normal daily dose given either intravenously at induction or orally immediately after the end of surgery. For complex procedures such as pancreaticoduodenectomy, oesophagectomy and cardiopulmonary bypass, replacement therapy of 100 to 150 mg cortisol per day for 3 days postoperatively is indicated.

The routine use of preoperative stress tests in these patients is not indicated, as neither basal nor adrenocorticotrophic hormone-stimulated serum cortisol concentrations adequately predict the changes seen after surgery. However, there is one group of surgical patients needing special mention, those receiving corticosteroids by topical administration (e.g. by inhalation, intranasally, transdermally, or by enema). Hypothalamo–pituitary–adrenal suppression is extremely rare in these patients, and it is safe to withhold additional supplementation unless they suffer a complicated clinical course postoperatively.

Patients with known hypothalamo–pituitary–adrenal insufficiency

All these patients require glucocorticoid supplementation during surgery; and in those already subjected to adrenalectomy or hypophysectomy, mineralocorticoid supplements may also be required.

In all patients, serum creatinine (or urea) and electrolytes, and blood glucose, should be monitored before surgery, and fluid balance should be carefully checked. Adrenal insufficiency in the postsurgical patient may be manifest as true collapse, or as malaise, lethargy, muscle weakness, and postural hypotension. Treatment should be by fluid infusion of colloid and crystalloid. These patients are often resistant to inotropes and large doses of steroids may be necessary.

Drugs acting on the central nervous system

Therapies for affective disorders

Several new classes of drugs that are of importance to both surgeons and anaesthetists have been introduced in recent years.

Tricyclic antidepressants Severe arrhythmias, including ventricular ectopic beats, may occur in the spontaneously breathing patient receiving volatile inhalational agents, especially halothane. Similarly, the administration of sympathomimetic amines may provide exaggerated responses. The incidence appears greatest in patients undergoing oral surgery. There is agreement that, provided caution is used, continuation of treatment into the perioperative period is a reasonable line of approach.

Monoamine oxidase inhibitors The intraoperative administration of vasopressor agents to patients receiving treatment with monoamine oxidase inhibitors may result in hypertensive crises. These inhibitors may also interact with opioids, resulting either in ‘excitatory signs' (hypertension, hyperpyrexia, convulsions) or marked central nervous depression (especially of the respiratory centre).

Two subtypes of monoamine oxidase are now recognized: type I(A), which preferentially deaminates noradrenaline (norepine-phrine) and serotonin, and type II(B),

which deaminates phenylethylamine. Although the accepted practice was the discontinuation of all monoamine oxidase inhibitors before elective surgery, this is not logical as many agents cause an irreversible antagonism of the type I enzyme, requiring that the enzyme be given time to regenerate before amines can be safely used. It is now accepted that this is not necessary if the anaesthetist avoids certain opioid drugs (viz. pethidine and dextromethorphan), and if only direct-acting sympathomimetic agents (e.g. isoprenaline and phenylephrine) are used. As an alternative, patients can be converted preoperatively to one of the newer, reversible monoamine oxidase inhibitors, for example moclobemide, which has a short half-life of about 2 h.

For patients receiving type II(B) monoamine oxidase inhibitors for the treatment of Parkinson's disease, serious interactions with anaesthetics or analgesics are rare, but again, pethidine (meperidine) is best avoided.

Lithium This treatment for manic depression may potentiate the effects of both depolarizing and non-depolarizing muscle relaxants. Preoperative determination of the plasma lithium concentration is desirable. Meticulous attention to hydration is essential, and non-steroidal anti-inflammatory drugs are best avoided.

Antipsychotics

The major tranquillizers all possess dopamine receptor-blocking properties; while some (e.g. thioridazine) are also arrhythmogenic or possess a-adrenergic receptor-blocking properties leading to intraoperative arrhythmias or hypotension. There is no reason, however, to discontinue their administration before surgery. All of these drugs are central depressants and will therefore result in reduced requirements for most general anaesthetic agents.

Anxiolytic drugs

Apart from a reduced requirement for induction and maintenance agents, anaesthesia in patients on long-term benzodiazepine therapy should be unaltered. Prolonged cessation of benzodiazepines during the perioperative period can result in a withdrawal syndrome.

Antiepileptic drugs

Surgery and anaesthesia may influence the disposition of these agents, resulting either in subtherapeutic drug concentrations, which increase the risk of seizures, or toxicity. Postoperatively, binding of phenytoin is reduced, owing to changes in the plasma concentrations of albumin, a_1-acid glycoprotein and free fatty acids. Drug binding and distribution may also be influenced by changes in acid–base status. Another important side-effect of chronic antiepileptic therapy [especially with phenobarbitone (phenobarbital), phenytoin, and carbamazepine] is hepatic microsomal enzyme induction: patients may require increased doses of both hypnotic and analgesic drugs for adequate anaesthesia and pain relief. There are no published data on any interaction between anaesthetics and the newer antiepileptic agents (such as gabapentin, lamotrigene, and vigabatrin).

Drugs for Parkinson's disease

Many patients have adverse effects from the unwanted peripheral actions of dopamine and other catecholamine metabolites of levodopa; these effects include cardiac arrhythmias. However, the combination of levodopa and carbidopa should not be withheld before surgery as any arrhythmias are easily treated by intravenous b-adrenoceptor-blocking drugs. Withdrawal of levodopa can also be associated with a neuroleptic malignant-like syndrome.

Anticoagulant therapy

General anaesthesia is not contraindicated in patients receiving anticoagulants; but for elective surgery oral drugs such as warfarin should be discontinued 3 to 4 days preoperatively and replaced with an infusion of heparin at a rate sufficient to achieve an activated partial thromboplastin time of twice normal. Before induction of anaesthesia, the international normalized ratio should be below 1.5. For the emergency patient, any residual effect of warfarin should be checked for, and reversed with fresh frozen plasma or clotting-factor concentrates.

Heparin is usually discontinued just before surgery; the best moment is at the time of oral premedication (the biological half-life of the anticoagulant being dose dependent, but ranging between 0.5 and 3 h). The activated partial thromboplastin time should be checked before the induction of anaesthesia. Postoperatively, the heparin infusion should be recommenced once the danger of haemorrhage has passed. Once the patient is tolerating oral intake, warfarin treatment can be started with an oral loading dose of 8 to 10 mg, repeated 24 h later. The international normalized ratio should be checked 48 h later and, if in the therapeutic range (2–2.5), the heparin may be discontinued.

For perioperative prophylaxis against deep venous thrombosis or pulmonary embolism, 5000 units of heparin should be given subcutaneously with the premedication and then continued 12-hourly for 3 days after surgery. The efficacy of this therapy is not usually monitored, but the risk of haemorrhagic episodes is increased in patients concurrently receiving antiplatelet therapy such as low-dose aspirin.

Low molecular-weight heparins (certoparin, dalteparin, enoxa-parin) are as effective and safe as unfractionated heparin in the prevention of venous thromboembolism, but they have a longer duration of action and are therefore given as a once-a-day subcutaneous dose. There is consensus that discontinuation of low molecular-weight heparins is not necessary before regional anaesthetic techniques.

Additional prophylaxis may include elasticated or inflatable stockings and calf stimulators.

Aspirin and other non-steroidal anti-inflammatory drugs

The non-steroidals cause a dose-dependent but reversible inhibition of the enzyme cyclo-oxygenase; aspirin induces an irreversible inactivation of the enzyme that lasts for 7 to 10 days after the cessation of treatment (requiring new platelet synthesis). There is controversy over the management of patients on these drugs and the conduct of epidural or spinal anaesthesia because of the increased tendency towards bleeding and the development of epidural haematoma. Many authorities advise that if the preoperative ‘bleeding time' test is greater than 10 min, then central neuraxial conduction blockade should not be undertaken.

However, there is often an anaesthetic advantage to continuing aspirin and other antiplatelet agents into the perioperative period as they act to reduce the peroperative requirements for anaesthesia and analgesia as well as affording protection against myocardial ischaemia and infarction in those with coronary arterial disease.

Oral contraceptives

The combination of surgery and oral contraception increases the incidence of postoperative deep venous thrombosis and the oestrogen-containing oral contraceptive pill should be discontinued at least 4 to 6 weeks before elective major surgery or surgery requiring prolonged immobilization, and before all surgery to the legs. If emergency surgery is required in a woman currently using oral contraception, prophylaxis against deep venous thrombosis may take the form of low-dose heparin, intraoperative infusion of dextran 70, and elasticated stockings. These recommendations do not apply to minor surgery involving brief immobility, or to women taking the progesterone-only pill.

Hormone-replacement therapy

As this is given in physiological (as opposed to pharmacological) doses, there should be no indication for it to be stopped before surgery. However, there is an increased incidence of deep venous thromboembolism in patients on hormone-replacement therapy, and it may be prudent for the surgeon to review the need for the therapy in particular circumstances.

Smoking

Chronic smoking is associated with the increased likelihood of postoperative chest infections and a reduction in the oxygen-carrying capacity of haemoglobin due to the presence of carboxyhaemoglobin. Hence, many surgeons and anaesthetists feel that smoking should be stopped at least 3 to 4 weeks before surgery.

Alcohol

Although the consumption of alcohol in moderate amounts presents no problems to the surgical patient, excessive quantities may be associated with enzyme induction

or liver dysfunction. Alcoholics require increased doses of anaesthetic, opioid, and other drugs given during the perioperative period to achieve therapeutic effects. Abrupt withdrawal of alcohol over the perioperative period may precipitate symptoms, which may be treated by judicious doses of alcohol or chlormethiazole (clomethiazole). In severe cases, seizures may occur, which require intravenous diazepam and respiratory support before commencing chlormethiazole.

Inherited conditions

There are a number of genetic and familial conditions that influence anaesthetic practice, three of which are of major importance.

Malignant hyperpyrexia

This is an autosomal-dominant disease that mainly affects children and young adults, which arises from an uncontrolled increase in intracellular calcium in skeletal muscle. It is a heterogeneous genetic disorder with linkage to genes found on chromosomes 1, 3, 7, 17, and 19. The disease is characterized by heat production (greater than 1°C rise/h), muscle rigidity, excess lactate and CO_2 production, hypoxia, hyperkalaemia, respiratory and metabolic acidosis, and myoglobinuria. The incidence is about 1:200 000 surgical patients in the United Kingdom.

For the disease to be expressed, a genetically sensitive individual must be exposed to a trigger agent (all inhalational anaesthetic agents, suxamethonium, and possibly nitrous oxide). A greater incidence of the condition is found among patients undergoing the surgical correction of squints and hernias, and with minor orthopaedic procedures. Mortality is high (up to 10 per cent even in developed countries), but can be limited by early detection through careful patient monitoring, a high level of suspicion, and the use of dantrolene. Acute management involves intravenous dantrolene in a dose of 1 mg/kg initially, repeated as necessary up to a cumulative maximum dose of 10 mg/kg; withdrawal of likely precipitating agents; and supportive therapy, including control of hyperkalaemia and acidosis, surface and peritoneal cooling, and treatment of cardiac arrhythmias. Dantrolene sodium may also be used as prophylaxis in susceptible individuals. As the disease is genetically inherited, at-risk individuals should be screened by the *in vitro* examination of a muscle biopsy taken under local anaesthesia. Abnormal contracture following exposure to caffeine is the principal diagnostic test, combined with electron microscopy.

Conditions that mimic malignant hyperpyrexia in presenting with hypertension, acidosis, and rhabdomyolysis include:

1. interaction between a monoamine oxidase inhibitor type I (A) and pethidine (meperidine);
2. cocaine toxicity;
3. heat stroke;
4. serotoninergic syndrome;
5. phaeochromocytoma;
6. neuroleptic malignant syndrome.

Plasma pseudocholinesterase deficiency

Deficiency of this enzyme, which is responsible for the metabolism of suxamethonium, mivacurium, procaine, 2-chloroprocaine, and tetracaine, may be either acquired or inherited. The inherited type may be expressed as both homozygous and heterozygous conditions, and causes variable degrees of prolongation of the duration of suxamethonium, leading to prolonged apnoea and the need for continued intermittent positive-pressure ventilation. In affected patients and, if necessary, in their relatives, the genotype can be delineated by measuring the percentage enzyme inhibition caused by dibucaine (10^{-5} M) and fluoride. A silent gene lacking enzyme activity, and other genetic variants, exist (e.g. C_5).

Porphyria

This group of diseases is caused by inherited defects or acquired dysfunction of the enzymes responsible for haemopoiesis. Hepatic (acute intermittent) porphyria presents the most important problems to the anaesthetist and occurs about 1 in 100 000 live births in the United Kingdom, although higher rates occur in other countries (e.g. Sweden, southern Africa). The condition is due to a deficiency of uroporphyrinogen I synthase, and is diagnosed by an increased urinary concentration of aminolevulinic acid and porphyrobilinogen. Acute attacks may be precipitated by a number of drugs including barbiturates, flunitrazepam, the contraceptive pill, and prochlorperazine.

Preoperative examination of the porphyric patient should include careful assessment of the cardiovascular system for tachycardia and hypertension, and of the central nervous system for neuritis, neuropathies (motor, sensory, and autonomic), and bulbar palsy. Regional anaesthesia is contraindicated as neurological complications may arise. Other, less common forms of the disease that also have anaesthetic implications include hereditary coproporphyria and porphyria cutanea tarda. The same drugs may precipitate crises in these patients.

The preoperative period

Preoperative laboratory testing

Preoperative laboratory testing of patients scheduled for elective surgery allows quantitation of any abnormality detected by history and physical examination, the detection of significant abnormalities not revealed by history and examination, and offers medicolegal protection to medical staff.

There is general agreement that an assessment of the severity of pre-existing disease is important: this allows the patient to be restored to optimal health before surgery, any support likely to be needed in the perioperative period to be predicted, and also assists in risk–benefit decisions. A detailed discussion of preoperative preparation of patients with the various commonly encountered diseases is outside the scope of this section.

Screening of asymptomatic patients in the belief that detection of abnormalities will improve perioperative course and outcome is unproven. Until recently, minimal data were available to support the assertions made about ther value of screening, but there has been increased evaluation of the benefits of preoperative laboratory testing in the last decade. A retrospective study of patients undergoing inguinal herniorrhaphy or stripping of varicose veins found only 63 abnormal results in 1972 tests, and in no instance was patient management influenced by those results. Others have found similar results and it is concluded that history and physical examination determine which tests are appropriate in all but a small minority of patients.

The use of a battery of tests to provide medicolegal protection has also been questioned. One review concluded that there was no benefit to patient management or cost benefit from routine screening of asymptomatic patients. In addition, potential legal problems may be generated when false-positive results occur. Retrospective studies have shown that unexpected positive results in asymptomatic patients are generally ignored: for example, anaemia found on routine screening of children was ignored in 74 per cent of cases.

The tests recommended for asymptomatic patients undergoing routine surgery are shown in Table 4.

	FBC	PTs/PTT	Cr and E	Glucose	LFT	CXR	ECG
Age > 75 years	X		X	X		X	
Smoking	X						
Cardiovascular disease			X			X	X
Respiratory disease	X					X	X
Liver disease	X	X		X	X		
Hepatitis B/C exposure					X		
Renal disease	X		X				(X)
Bleeding disorders	X	X					
Diabetes mellitus			X	X			X
Drugs:							
Diuretics			X				X
Digoxin			X				X
Steroids			X	X			
Anticoagulants	X	X					

FBC, full blood count; PTs/PTT, prothrombin time/activated partial thromboplastin time; Cr and E, creatinine and electrolytes; LFT, liver function tests; CXR, chest radiograph; ECG, electrocardiograph; B/C, hepatitis B or C antigen exposure.

Table 4 Recommended screening tests for patients undergoing surgery

Premedication

Agents used for premedication form part of the anaesthetic technique, providing analgesia or supplementing volatile agent–nitrous oxide or hypnotic infusion–nitrous oxide anaesthetics. The primary indications for preoperative medication are anxiolysis, sedation (especially in child patients), analgesia, amnesia, vagolysis to reduce salivary secretions, and prophylaxis against postoperative nausea and vomiting, and hence aspiration pneumonitis.

Anxiolysis

Most patients experience some anxiety before surgery; this is best allayed by the combination of careful and sensitive discussion of the issues of concern to them and the use of antianxiety drugs, the most common agents being the benzodiazepines (diazepam, lorazepam, or temazepam) and some of the opioids (morphine, pethidine).

Sedation

This may also be achieved by administration of benzodiazepines or opioids. In children, antihistamines such as trimeprazine (alime-mazine) (which can be given orally) are also used.

Amnesia

Many anaesthetists regard amnesia of the events leading up to anaesthesia and surgery as desirable, especially in the very young and in patients having repeated general anaesthetics. Amnesic drugs may also reduce the risk of awareness during anaesthesia, allowing a lighter depth of anaesthesia to be maintained. The most effective amnesic agents are the benzodiazepines (especially lorazepam and midazolam). The anticholinergic agent hyoscine also exhibits amnesic properties.

Analgesia

The routine use of analgesic drugs as part of the premedication, except in patients experiencing severe acute or chronic preoperative pain, has been criticized by some anaesthetists. However, the preoperative administration of analgesic drugs (both opioids and non-steroidal anti-inflammatory drugs) reduces the dose of induction and maintenance agents needed for clinically adequate anaesthesia. In addition, the use of analgesia (either as part of the premedication, or intraoperatively) provides patient comfort in the initial postoperative period. The main disadvantages of all opioid drugs are their association with nausea and vomiting, and the risk of respiratory depression.

Antivagal actions

The response to many noxious stimuli is the development of a vagally medicated bradycardia. Intramuscularly administered anticholinergics are ineffective in preventing these responses. Routine use of antisialogogues has major disadvantages: the patient suffers an unpleasant dry mouth during the preoperative period, and dry mucous membranes are sticky and easily damaged during laryngoscopy and intubation. Present use of these agents tends to be reserved for infants, and those situations where a dry mouth is advantageous (such as for intraoral surgery). Hyoscine is the most potent of the antisialogogues available and has the additional advantage of producing amnesia and sedation. In the elderly patient, however, there is a significant incidence of perioperative confusional states.

Antiemesis

Drugs used for premedication may either reduce the emetic effects of the anaesthetic agents employed or aid gastric emptying. The first group includes the antihistamines and butyrophenones, as well as hyoscine; gastric emptying can be facilitated by administration of metoclopramide, and the gastrokinetic drug cisapride. Additional effective medications include transdermal hyoscine, and the 5-HT_3 (5-hydroxytryptamine) antagonists ondansetron and congeners.

Agents to decrease gastric acidity are normally given only to those patients at risk of regurgitation of gastric contents, such as the pregnant woman, those with a hiatus hernia or strong clinical history of gastric reflux, the obese, diabetics, patients with renal failure, those with clinical evidence of gastric stasis, and in those procedures associated with a high incidence of nausea and vomiting, such as laparoscopy. Antacid therapy aims to reduce acid production and to raise the pH of residual gastric contents to above 2.5. The first is achieved by histamine H_2-receptor or hydrogen-pump antagonists given over several hours preoperatively; the second by alkalis such as sodium citrate given orally 15 to 30 min before induction.

Antihistamine drugs, such as the phenothiazines promazine, chlorpromazine, and promethazine, may also offer protection against the drug-mediated release of endogenous histamine and other autocoids in the atopic individual. Their protection is usually incomplete, and coadministration of sodium cromoglycate and hydrocortisone may be of benefit in the highly sensitized patient.

Methods of premedicant drug administration

Although the most popular routes of administration are the intramuscular (the traditional method) and the oral, alternative routes such as intravenous, rectal, and transdermal have specific indications.

Orally administered premedicants may be less effective than those given parenterally. Gastric emptying and drug absorption will be influenced by pain, anxiety, opioids, and intra-abdominal pathology. However, the oral route is considered by many to be the route of choice for children, and it is the preferred route in patients with bleeding disorders (patients receiving anticoagulants, haemophiliacs, or those with other coagulation defects, and those with liver and renal failure).

The timing of the premedication is of importance, since maximum antisialogogue and antivagal effects need to be achieved before induction of anaesthesia: thus the majority of drugs are given 1 to 2 h preoperatively. For more rapid onset of anxiolysis, the short-acting benzodiazepine temazepam has been advocated, especially for use in day-care anaesthesia.

Starvation and anaesthesia

Although it is commonly stated that the period of ‘nil by mouth' for clear fluids and solids should be 6 to 8 h before the induction of anaesthesia, these guidelines have recently been questioned and updated. The time for gastric emptying of liquids is between 1 and 2 h, while that for solids may exceed 6 h. Furthermore, there is clear evidence that recovery in ambulatory patients may be delayed by any relative hypovolaemia secondary to dehydration. Hence the current guidelines are:

1. no solids from 6 to 8 h before induction;
2. no restriction on clear fluids up to 2 h before the scheduled start of surgery; oral medications may be taken 1 to 2 h before surgery with up to 150 ml water;
3. histamine H_2 antagonists in patients considered at risk of pulmonary aspiration.

Patients at risk of aspiration include those undergoing emergency surgery; those in whom there is likely to be ‘gagging' at intubation or a projected difficult intubation (such as in patients with limited jaw opening, or with anatomical or congenital abnormalities of the jaw and upper airway); those with likely increased gastric contents (i.e., those with gastric outflow obstruction); those with increased intragastric pressure; those with a decreased lower oesophageal sphincter tone (e.g. the parturient, the obese); and those with gastro-oesophageal reflux.

General anaesthetic pharmacology and practice

Safe general anaesthesia depends upon the administration and maintenance of an ‘adequate' anaesthetic dose without the development of unwanted or adverse side-effects. The monitoring, where possible, of the depth of anaesthesia is by clinical signs (such as heart rate, blood pressure, or autonomic responses) or a derivative of brain activity, such as the electroencephalogram or evoked potentials. These should relate the presence of clinically adequate anaesthesia to a defined plasma drug concentration. Safe anaesthesia also requires prompt recovery of normal brain activity through the use of drugs that undergo rapid elimination and have no residual depressive effects on the central nervous or other systems.

Induction

Unconsciousness may be induced by intravenous hypnotics alone or in combination with other agents (including opioids). Hypnotics (with the exception of ketamine) do not possess analgesic properties, nor do they relax skeletal muscle. Although they are rarely used alone for the maintenance of anaesthesia, their use will reduce the requirement for other anaesthetic agents. There are at present a number of agents available for induction of anaesthesia:

thiopentone
etomidate
propofol
opioids such as fentanyl or alfentanil or sufentanil
methohexitone
ketamine
midazolam
thiamylal

Thiopentone (thiopental)

This is still the archetypal hypnotic drug, providing a rapid, effective, and safe onset of anaesthesia. Its non-hypnotic properties include cardiorespiratory depression, and, in large doses, prolonged recovery. Absolute contraindications are rare and include proven barbiturate allergy or a history of any acute porphyria. Thiopentone should be used cautiously, with reduced dosage and careful titration of dose to effect, in patients who have inadequate airway control before induction, in severe cardiovascular collapse or shock, and in those with status asthmaticus, cardiovascular disease (ischaemic heart disease, hypertension, valvular heart disease), hypovolaemia, acute adrenocortical insufficiency, or uraemia.

Methohexitone (methohexital) and thiamylal

These barbiturates have faster clearance rates than thiopentone and both are broken down to inactive metabolites. The side-effects of methohexitone include pain on intravenous injection, a tendency to venous thrombophlebitis, and exaggerated involuntary movements, especially in the unpremedicated patient and those with epilepsy.

Etomidate (a carboxylated imidazole)

This agent has the advantage of not inducing histamine release, and is therefore indicated for use in asthmatic and atopic patients. Its use in induction causes only minimal haemodynamic and respiratory changes. Pain on injection and myoclonic activity during induction have limited its use for routine surgery.

Ketamine

Ketamine is the only true hypnotic agent that also has analgesic effects. In contrast to other agents, it increases sympathetic nervous activity, so affording rapid induction of anaesthesia in patients who require a high level of sympathetic tone to maintain cardiovascular function (for example, in the presence of pericardial tamponade or hypovolaemic hypotension). It is not as rapidly acting as some of the other induction agents already mentioned, and recovery from anaesthesia may be complicated by emergence reactions including hallucinations in up to 30 per cent of adult patients and an even higher proportion of children. These side-effects can be minimized by pretreatment with a small dose of a benzodiazepine. Because of its sympathomimetic properties, ketamine can be used for induction (and maintenance) of anaesthesia in the poor-risk surgical patient (e.g. hypovolaemia). It also has an important place in the provision of repeated anaesthesia (and analgesia) in patients with burns (for debridement of wounds, painful dressing changes, and skin grafting), as well as for short diagnostic and surgical procedures (including cardiac catheterization). Ketamine is also active when given by the intramuscular route, so allowing analgesia to be achieved when there is poor venous access at sites of major injury.

Ketamine should not be used in patients with raised intracranial pressure. Care should be taken over its use in the patient with renal impairment, where its effect may be pronounced and prolonged because of the accumulation of active metabolites.

Propofol

Propofol (di-isopropyl phenol; Diprivan) has the advantage over thio-pentone (thiopental) of being rapidly broken down to inactive metabolites; as such its se is accompanied by rapid recovery, so making it the agent of choice for day-case and short-stay anaesthesia and surgery. The cardiovascular and respiratory depressant effects of propofol are similar to those seen with thiopentone when used for induction.

Use of propofol is often accompanied by pain on injection, especially when it is given into small veins on the dorsum of the hand. This unwanted effect can be limited by pretreatment with a small dose of lignocaine (lidocaine) or an opioid such as fentanyl or alfentanil, or by premixing local anaesthetic and hypnotic agent immediately before injection.

Other side-effects of propofol are coughing and hiccuping during induction, and involuntary movements. There have been occasional episodes of seizure activity (convulsions, myoclonus, or opistho-tonos) after propofol (although the drug has well-defined antiepileptic activity in animals), and therefore it should not be used for induction in known epileptics who currently hold a driving licence. This side-effect usually occurs within a few hours of anaesthesia, but there are reports of it being delayed for up to 6 days.

The low incidence of nausea, vomiting, and headache following propofol anaesthesia is important for the ambulant patient. However, because of its effects on the heart (where it causes depression of myocardial contractility and increased vagal tone), propofol should be used with care in patients with compromised cardiac function (e.g. ischaemic heart disease, arterial hypertension) or hypovolaemia.

Midazolam

This is a water-soluble benzodiazepine used to provide sedation, amnesia, and sleep but not usually anaesthesia. It has a shorter duration of action than diazepam, but may cause unexpected ventilatory and cardiovascular depression, especially in patients with chronic obstructive pulmonary disease, hypotension associated with hypovolaemia, and in elderly individuals.

Maintenance

Anaesthesia can be maintained by volatile agents or hypnotic infusions as supplements to either nitrous oxide or opioids. In some special circumstances it can be maintained by opioids alone with adjuvant drugs such as diazepam or midazolam.

Intravenous infusions of hypnotic agents

There is at present much interest in both Europe and the United States in the use of infusions of hypnotic agents (especially propofol) to supplement nitrous oxide or opioids for the maintenance of anaesthesia. Compared to the older hypnotic agents such as thiopentone (thiopental), propofol has a high rate of clearance from the body and is not biotransformed to pharmacologically active metabolites.

Drugs may be delivered either by empirically altering the rate of infusion according to clinical effect or with stepped infusions aimed at achieving a particular blood concentration. This requires the clinician to have knowledge of the individual drug's disposition kinetics in specific groups of patients, and then to calculate the necessary infusion rate on the basis of blood concentration = infusion rate/ clearance. Clearly this approach does not allow for interindividual variations in drug handling and drug dynamics. A newer approach is the concept of ‘target-controlled infusion', where the microprocessor-controlled infusion pump drives a motorized syringe driver to deliver the amount of drug needed to achieve a given blood concentration.

Since the effect site for anaesthetic agents is within the brain (not in the blood!), recent developments include the targeting of the concentration in the brain, based on *a priori* modelling in comparable groups of patients.

Some of the indications for the use of intravenous (as opposed to volatile-supplemented) techniques include:

1. day-case or ambulatory anaesthesia;

2. military or ‘field' anaesthesia (where nitrous oxide is not available);
3. as a supplement to cardiopulmonary bypass to reduce the incidence of awareness;
4. as a sedation supplement to regional anaesthesia (incorporating major neuraxial blockade or plexus blocks) or as part of the concept of ‘monitored anaesthesia care' that is widely practised in the United States (this can be defined as the administration of subanaesthetic doses of anaesthetic drugs to achieve amnesia and anxiolysis during a surgical procedure performed under local anaesthesia; routine anaesthetic monitoring is employed).

Gaseous and volatile agents

More commonly, volatile or inhalational agents provide the basis for general anaesthesia. The differing physicochemical properties of the various volatile agents will affect their pharmacological effects. Thus, agents with low blood–gas solubilities (e.g. nitrous oxide, desflurane, and sevoflurane) will cause rapid onset and recovery from central effects. The oil–gas solubility provides one index of potency, defined in terms of the minimum alveolar concentration, which is the alveolar concentration preventing a response to a surgical incision in 50 per cent of individuals (Table 5).

Agent	MAC	Oil/gas	Blood/gas
Nitrous oxide	105–110	1.4	0.47
Halothane	0.76	224.0	2.35
Enflurane	1.68	96.0	1.91
Isoflurane	1.15	91.0	1.40
Desflurane	6.00	18.7	0.42
Sevoflurane	2.05	47.0	0.63

MAC = concentration of anaesthetic in oxygen required to inhibit response to the initial surgical stimulus in 50% of individuals, i.e., ED_{50}).

Table 5 Minimum alveolar concentrations (MAC) (indices of potency) and solubility coefficients of inhalational anaesthetics

Nitrous oxide is a weak anaesthetic agent (minimum alveolar concentration 105–110 per cent), which cannot alone suppress the somatic, autonomic, or haemodynamic responses to noxious stimuli during clinical use. However, the addition of nitrous oxide at inspired concentrations of 60 to 70 per cent significantly decreases the dose requirements for other, more potent, volatile agents (as well as those of the hypnotic and opioid drugs). Relative contraindications to the use of nitrous oxide include the need for high inspired oxygen concentrations (for example, during bronchoscopy and upper airway surgery). It is also relatively contraindicated where its high diffusibility into air-filled spaces so increases their volume or pressure as to cause problems: this happens in the presence of a pneumothorax, after pneumoencephalography, after tympanoplasty, in patients with grossly dilated bowel due to intestinal obstruction, and where air embolism may occur. Nitrous oxide should be avoided if possible during the first 4 weeks of pregnancy, because of its effects on organogenesis.

The three established volatile anaesthetics in widespread use are halothane, enflurane, and isoflurane. They are all highly potent volatile agents, but with the disadvantage of having narrow margins of safety between their anaesthetic effects and cardiovascular depressive properties. They therefore have to be administered by calibrated vaporizers.

All three separate components of the anaesthetic triad (anaesthesia, analgesia, and relaxation) may be achieved with these agents; however, the concentrations required to achieve appropriate anaesthetic depth may result in excessive hypotension through the combination of myocardial depression, vascular smooth-muscle relaxation, and depression of the sympathetic nervous system. Thus, the inspired drug concentration tends to be regulated according to an adverse effect of the volatile agents, that is, cardiovascular depression; this is most significant with enflurane, and may be exaggerated in the patients with hypovolaemia or congestive cardiac failure. Cardiovascular depression is less severe with halothane; other side-effects that may favour the choice of halothane as the volatile supplement include low pungency and laryngeal irritability, bronchodilatation, and uterine muscular relaxation.

Isoflurane exhibits all of the favourable properties of the other two volatile agents combined with faster uptake and elimination, and minimal biotransformation, with its associated reduced risk of renal and hepatic toxicity. Although isoflurane produces vasodilatation, this is accompanied by a reflex increase in sympathetic activity that either maintains or increases both heart rate and cardiac output. However, hypotension may occur in patients with decreased sympathetic reserve, such as those receiving b-adrenoceptor-blocking drugs.

In appropriate concentrations, all three agents are able to control the hyperdynamic responses to noxious stimulation during surgery. Dose requirements for volatile agents can be reduced by combination therapy with other centrally acting drugs (hypnotics, sedatives, and opioids), as well as by achieving relaxation through the use of myoneural-blocking drugs.

Other non-hypnotic effects of these three volatile agents include possible alterations in myocardial blood-flow distribution with development of ischaemia (the so-called coronary steal phenomenon) with isoflurane; increased cerebral blood flow and intracranial pressure (seen with both halothane and enflurane in clinical concentrations, but with isoflurane only at concentrations above 1.5 per cent); risk of seizures during enflurane anaesthesia, especially in presence of hypocapnia; hepatotoxicity (with halothane, enflurane, and possibly isoflurane); fluoride-induced nephrotoxicity due to enflurane biotransformation; malignant hyperpyrexia (may be triggered by all three agents); and impaired cellular-mediated immunity. In addition, halothane particularly sensitizes the myocardium to the arrhythmogenic effects of both endogenous and exogenous catecholamines.

There have been two recent introductions, desflurane and sevoflurane, that offer advantages in some clinical circumstances.

Desflurane

This is another halogenated ether, similar in structure to isoflurane. It has important physicochemical properties that confer rapid onset and recovery: these are low blood–gas and other tissue partition coefficients; thus it is easier with desflurane to titrate the dose needed to produce the desired clinical effect. Because of its low oil–gas solubility coefficient (see Table 5), it has a lower potency than the other volatile agents. Furthermore, its boiling point (22.8°C) and high vapour pressure require that it be delivered using a temperature-controlled and pressurized vaporizer.

Its cardiovascular and respiratory effects are similar to those of isoflurane, but, in contrast to the other volatiles so far discussed, the use of desflurane is associated with sympathetic overactivity, especially when the inspired agent concentration is increased in a stepwise manner. This effect is manifest as a tachycardia. Unlike isoflurane, desflurane does not induce a ‘steal phenomenon', that is, vasodilatation of normal coronary arterioles taking blood away from areas of ischaemic myocardial tissue supplied by narrowed vessels.

Other individual characteristics of desflurane include a low ability to sensitize the heart to endogenous or exogenous catecholamines; a dose-related increase in cerebral blood flow (the drug is therefore probably best avoided in the patient with a head injury or increased intracranial pressure); augmentation of the action of non-depolarizing and depolarizing neuromuscular-blocking drugs; and the adverse effects of inhibition of the hypoxic pulmonary vasoconstrictor reflex, and of malignant hyperthermia. Unlike halothane and enflurane, the use of desflurane is accompanied by a high incidence of upper airway irritation (coughing, salivation, and laryngospasm), making it inappropriate for gaseous induction of anaesthesia. Because of its properties in onset and recovery, desflurane can be useful especially in the maintenance of anaesthesia for ambulatory surgery in adult patients. However, its use has been associated with an increased incidence of postoperative nausea and vomiting.

Sevoflurane

This is also a fluorinated ether; but does not have the very low blood and tissue solubilities of desflurane. As a result, its anaesthetic potency (minimum alveolar concentration, ED_{50}) is about 50 per cent less than that of isoflurane, but three times that of desflurane. The cardiorespiratory effects of sevoflurane are similar to those of isoflurane and desflurane, but its use is not accompanied by sympathetic overactivity as with desflurane. However, like isoflurane and desflurane, it does not potentiate cardiac arrhythmias induced by adrenaline (epinephrine). The extent of the dose-related depression of myocardial contractility is similar to that seen with isoflurane, but less than that associated with halothane and enflurane. There is no evidence that the ‘coronary steal' phenomenon occurs with sevoflurane.

As with desflurane, sevoflurane causes an increase in cerebral blood flow; but unlike isoflurane and desflurane, it does not alter the regulation of that flow. It does, however, cause greater respiratory depression than is seen with the other volatile anaesthetic agents.

One important feature of sevoflurane is its metabolism. It is readily broken down by the microsomal cytochromes in the liver hepatocytes to inorganic fluoride ions, which were indicted as the cause of the nephrotoxicity of methoxyflurane (since withdrawn from clinical practice) and enflurane. However, because of the lower tissue solubility of sevoflurane compared to both of those agents, the peak fluoride concentrations after even prolonged sevoflurane anaesthesia do not attain those usually associated with nephrotoxicity (^{3} 50 mmol/l).

Because sevoflurane is less pungent than isoflurane or halothane, it has assumed a position as the agent of choice for the inhalational induction of anaesthesia, where its use is associated with a lower incidence of airway complications (such as breath-holding, coughing, laryngospasm) and excitement, and quicker onset of a surgical plane of anaesthesia.

Intra- and postoperative analgesia

Opioid drugs may be given either as a premedication or by intravenous supplementation to achieve pain relief during surgery. As premedication, they are normally given by the intramuscular route because of the variability of oral drug absorption in the preoperative patient. Small doses of opioids (such as morphine 0.03–0.15 mg/kg or pethidine 0.5–1.0 mg/kg) provide intraoperative analgesic supplementation and reduce the dosage requirements for both the inhaled and other intravenous drugs.

Intraoperative supplementation is usually provided by drugs with a faster time to onset of effect than pethidine or morphine. Thus the anaesthetist uses synthetic opioid drugs (such as fentanyl, alfentanil, sufentanil, or the newer remifentanil), all with a fast onset, higher potency, and fewer side-effects, and a greater margin of safety. However, because all opioid drugs depress respiration, the anaesthetist needs either to provide controlled ventilation or to monitor closely the adequacy of oxygenation and gas exchange (by pulse oximetry and capnography), and of ventilation (by respirometry).

Too large a dose of intraoperative analgesic will lead to postoperative ventilatory depression necessitating respiratory support; the use of opioid antagonists (such as naloxone) may be accompanied by hypertension and tachycardia as a reflection of inadequate analgesia. Furthermore, the duration of action of single doses of naloxone is shorter than that of the opioid agonists, and hence ‘renarcotization' may occur as a late recovery event.

Adverse effects of general anaesthetic agents

Anaesthetic agents affect the central nervous system through a single (or series of) non-specific action(s); that is, they do not appear to be directed at a single receptor site. Hence, adverse (or non-hypnotic) effects occur with all of these agents. The effects are predominantly on the central nervous system (where they may cause excitatory signs or seizure activity; or emesis); on the cardiovascular system, where all agents (with the possible exception of ketamine) have a depressant effect, manifest as a fall in cardiac output, decreased arterial pressure, and a normal or increased heart rate; and on the respiratory system, where all agents cause a dose-related depression of respiration, an impaired ventilatory response to increased carbon dioxide tensions or hypoxia, as well as apnoea at high concentrations. These effects all tend to be predictable from the known pharmacology of the agents, and are dose dependent.

Overdose may be absolute, due to acute excess administration, or to accumulation as a result of reduced elimination or metabolism; or relative, due to patient factors such as hypovolaemia, extremes of age, intercurrent disease (especially cardiac, renal, or hepatic impair-ment), and drug–drug interactions. Exaggerated pharmacological effects that may occur include excessive hypotension or bradycardia; these are seen after the administration of large doses of the intravenous hypnotic agents and opioids, respectively.

Other effects on specific organs or tissues tend to be related to one or more agents rather than generic; they can be broadly grouped into three main areas:

1. effects not predictable from the known pharmacology of the drug and not showing a simple dose–response relation;
2. those associated with long-term drug therapy, and hence dependent on the total cumulative dosage;
3. delayed effects such as carcinogenicity, teratogenicity, and perhaps organ toxicity, which are dose independent.

There are many examples of idiosyncratic reactions that are not predictable from the basic pharmacology of an individual drug: these include porphyria, malignant hyperpyrexia, plasma pseudocholinesterase deficiency, glucose-6-phosphate dehydrogenase deficiency, and the genetic polymorphism of drug metabolism due to different alleles of the metabolizing enzymes (e.g. fast and slow acetylators of hydralazine, slow and fast hydroxylators of desbrisoquine, or propranolol).

Among the intravenous anaesthetic agents there are other examples of adverse effects, such as those due to bacterial contamination of the agent or the use of incompatible drug mixtures. For example, precipitation occurs if thiopentone and *d*-tubocurarine, or vecuronium, are administered in the same syringe or via the same intravenous cannula without flushing between drugs. A further example of drug toxicity is the development of hypersensitivity reactions.

Anaphylactoid and anaphylactic reactions

In general, the incidence of adverse drug responses (as outlined above) is higher than that of the true allergic or immunological reactions. Anaphylaxis is an example of a type I hypersensitivity reaction and as such constitutes a medical emergency requiring immediate resuscitation of the patient. It usually follows a second exposure to an antigen, although a clear history of previous exposure may not be obtained.

Mechanisms of anaphylaxis Most drugs administered during anaesthesia are of low molecular weight (less than 1000 Da). Alone they are not capable of initiating an immune response with the production of specific antibodies. However, many drugs are able to combine *in vivo* with a carrier protein (hapten) to form an antigenic moiety. The extent of basophil and mast-cell degranulation induced depends on the amount of drug injected, the affinity of the drug for the antibodies, and the amount of cell-bound antibodies produced.

Anaphylactic reactions These depend upon classical antigen–antibody interactions that activate the complement cascade via the classical pathway. Complement activation results in mast-cell disruption and the release of histamine and other autocoids, which increase vascular permeability, and so allow the passage of phagocytes from extravascular sites to enter the bloodstream and break down the antigens.

Anaphylactoid reactions These occur as a result of a direct pharmacological effect or some other non-immunological response by which the drug causes release of histamine. Some drugs, particularly those with basic properties (e.g. *d*-tubocurarine, suxamethonium, morphine, thiobarbiturates, and trimetaphan), cause direct competitive displacement of histamine (another basic molecule) from the mast cell. As will be discussed later, predisposing factors exist that may also play a part in the genesis of these anaphylactoid responses.

The main causes of these anaphylactoid hypersensitivity reactions to drugs can be subdivided into activation of the alternative pathway of complement and direct pharmacological effects.

Alternative complement pathway

Unlike the classical pathway, preformed antibodies to a particular antigen are not necessary for alternative pathway activation. Hence prior exposure to the drug or endotoxin is not necessary.

Direct pharmacological effects of drugs

Although adverse drug reactions may involve activation of one or other of the complement pathways, the majority are better described as anaphylactoid in nature, and occur as a result of a dose-related, direct effect of the drug upon mast cells and basophils. Such reactions may be manifest as either local cutaneous signs, or as more severe systemic signs and symptoms of histamine release. With drugs such as thiopentone (thiopental) and some of the neuromuscular-blocking agents, a flush and weal reaction is often seen following their intravenous injection. The systemic features of the reactions are usually mild (hypotension, tachycardia), and without the other stigmata of histamine release. Why some of these cases (for which no immunological basis will be found on subsequent testing) progress to more generalized, systemic symptoms is uncertain.

Predisposition to allergic reactions Factors predisposing to the development of allergic reactions include patient factors, possible pharmacological factors, and potential at-risk patient groups.

Patient factors Adverse reactions are rarer in children than in adults, but there is probably no sex difference in their incidence. The apparently higher incidence in gynaecological patients may be related to pregnancy.

There is a high incidence of adverse reactions on first or subsequent exposure to certain intravenous anaesthetics in the pregnant woman, which may be due to the choice of drugs used for short gynaecological procedures. The immune status of these patients is also altered, and this may play a part.

Hypersensitivity reactions to intravenous anaesthetic drugs are more common in atopic individuals and asthmatics. The increased incidence of hypersensitivity reactions to first exposure in the atopic individual is small compared with the higher incidence of hypersensitivity reactions seen in patients receiving repeat administrations of the same anaesthetic agent. The impression among some anaesthetists that adverse reactions are more common in nervous, anxious patients is not supported by any data. Apart from patients with a history of previous anaphylaxis or a proven sensitivity to one or more drugs, there are no other predictors of at-risk groups.

Pharmacological factors Many of the intravenous anaesthetic agents are poorly water soluble, and hence glycols, macrogols, and non-ionic surfactants are added to increase their solubility; these solvents may trigger complement activation.

Clinical features of allergic reactions There is considerable variation in the severity and magnitude of the clinical features among different reactors. Factors that influence the severity of the symptoms include the amount of drug injected, the reactivity of basophils and mast cells, the responsiveness of the bronchial and vascular smooth muscle, and the activity of the autonomic nervous system. Control of peripheral autonomic nervous activity is also regulated by higher centres. Thus, the increased emotional activity of the asthmatic patient before anaesthesia and surgery may both enhance any existing symptoms and exaggerate any adverse response to intravenous drug administration.

The various symptoms that may occur during adverse reactions are shown in Table 6. The maximum intensity of the symptoms often arises rapidly, usually within 30 min of intravenous drug administration. All of the clinical features can be attributed to mast-cell and basophil degranulation, and this knowledge has been used in some of the methods for preoperative prophylaxis in patients with a history of drug allergy. The usual order of the main clinical symptoms is skin changes, hypotension with tachycardia, and bronchospasm with resulting arterial hypoxia.

Cutaneous flushing or development of an overt rash
Pruritus
Cough
Tachycardia
Bronchospasm
Hypotension
Cardiac arrest

Table 6 Symptoms and signs occurring during acute hypersensitivity reactions to intravenous anaesthetic agents

Skin changes are characterized by the classic triple response described by Lewis. Erythema is caused by dilatation of capillaries over the face, arms, and mantle; weal formation arises from increased vascular permeability causing fluid transudation; flares are presumably due to an axon reflex response. Hypotension and tachycardia result from the transudation of fluid from the vascular to extravascular spaces and from vasodilatation. Histamine also causes vascular smooth-muscle dilatation, and hence venous pooling of blood may occur. The accompanying tachycardia is probably not a baroreflex response, but rather the chronotropic effect of increased circulating catecholamines following their release by histamine from the adrenal medulla. Bronchospasm is the most life threatening of the symptoms, and treatment must primarily be aimed at preventing severe arterial hypoxia.

Following the successful treatment of an allergic reaction, the causative agent must be identified wherever possible. Laboratory diagnosis depends on the results of a number of routine haematological tests, including measurement of the haematocrit. Changes in platelet and white-cell numbers may be small but are significant. Disappearance of basophils (expressed as a percentage of the total leucocyte count) is indicative of a type I response. Serial measurements should be made of the plasma concentrations of total IgE antibodies and complement proteins (C2, C4, the breakdown products C3a, C3b, and C1 esterase inhibitor) during the 72 h after an adverse reaction. Metabolism of both C3 and C4 is indicative of an immune reaction that does not involve IgE antibodies, while C3 conversion alone is an indicator of a specific non-immune mechanism. In contrast to atopic patients, who have increased amounts of IgE, some individuals (10–20 per cent) have very low concentrations of this immunoglobulin. Such people appear to be prone to non-immune-mediated clinical reactions, and may produce positive intradermal tests to drugs that have a high potential for inducing histamine release.

Other *in vivo* and *in vitro* diagnostic tests that may be useful for later assessment to establish the identity of the drug responsible for the reaction include intradermal skin testing, the IgE inhibition test, the leucocyte histamine release test, the basophil degranulation test, the mast-cell tryptase, radioallergosorbent test, and delineation of specific antibodies.

Treatment of allergic reactions to drugs The management of severe histamine-release reactions has been comprehensively reviewed. The aims of treatment must be to identify (where possible), and stop administration of, the causative agent, correct hypoxia and restore the vascular fluid volume (preferably with colloids), and inhibit further release of chemical mediators. When the combination of adrenaline (epinephrine) and adequate volume does not produce improvement, then noradrenaline (norepinephrine) may be life saving. Table 7 lists a management approach to allergic reactions.

Drug administration of antigen
Cease anesthesia
Ensure intravascular fluid volume
Inhibition of further release of chemical mediators
Monitor
Airway
Inhaled oxygen/oxygen
Adrenaline (epinephrine)
Intramuscularly or intravenously (depending on the severity of reaction) or via the endotracheal tube if venous access not available (Over 1/2–1.0 mg (ca. or 50–100 μg i.v. every 1 min, in cardiovascular collapse, 1/2–1.0 mg i.v. or a rate of 0.1 mg/min, titrate dose to blood pressure response
Fluids
Infuse crystalloid or colloids to restore circulating volume
Bronchodilators
If there is bronchospasm, give salbutamol, aminophylline, or nebulized β-blocking anaesthetic, consider use of sevoflurane, diethyl ether, ketamine, or isoflurane for relief of bronchoconstriction
Positive-pressure ventilation
Especially in cases of pulmonary oedema
Correct acidosis
Second-line therapy
Antihistamines, corticosteroids

i.m., intramuscularly; i.v., intravenous.

Table 7 Management of allergic reactions

There is no place for the routine administration of steroids as a first-line treatment. Cortisol does not inhibit the allergen–antibody interaction or the release of vasoactive amines from mast cells and basophils. Thus, steroids are probably only indicated for the treatment of severe bronchospasm. Anticholinergic drugs, such as atropine, may attenuate the release of histamine by decreasing the intracellular concentrations of the second messenger, guanosine monophosphate. However, only limited data are available from which to draw conclusions. Bronchodilators such as salbutamol, and even general anaesthetic agents with bronchodilator properties (diethyl ether, sevoflurane, isoflurane, or ketamine) may be required for refractory bronchospasm. Persistent vasodilatation can be reversed by dopamine, dobutamine,

ephedrine, or adrenaline, given in repeat bolus doses or by infusion. There is no evidence of any value in the acute administration of H_1- and H_2-blocking drugs, but these may have a role in prophylaxis.

Management of these patients after recovery should include counselling and an explanation of the significance of hypersensitivity, reassurance about future anaesthesia, and registration of the patient with Medic Alert or a similar organization. In addition, a clear summary of the reaction should be placed in the patient's notes, and sent to the general practitioner.

Management of the patient who has previously suffered an allergic reaction Although the occurrence of a first-exposure reaction to any drug is unpredictable, steps can be taken to prevent hypersensitivity reactions occurring on subsequent exposures. These precautions include the use of local rather than general anaesthesia where possible, although hypersensitivity to the ester type of local anaesthetics has been reported. When general anaesthesia is essential, adequate preoperative anxiolysis and premedication may reduce stress. Preoperative prophylaxis should be aimed at both reducing histamine release and blocking its systemic effects. Sodium cromoglycate stabilizes mast cells and so prevents degranulation and histamine release, salbutamol prevents bronchospasm, and hydrocortisone antihistamines (H_1-receptor antagonists, e.g. chlorpheniramine or terfenadine, and H_2-receptor antagonists, e.g. cimetidine or ranitidine) may also be administered. The last two groups of drugs inhibit histamine release, but also compete at the receptors to attenuate the decreases in systemic vascular resistance.

Repeat exposure to any drug that the patient has received intravenously in the recent past should be avoided and inhalational agents rather than intravenous agents should be used whenever possible, although repeat exposures to halothane should be avoided. Plasma expanders such as gelatine suspensions and dextrans should not be used if they have been implicated in the previous episode.

Other toxic effects of anaesthetic agents

The toxicity of volatile anaesthetic agents has been known since the introduction of diethyl ether and chloroform. Concerns over modern agents started with the introduction of halothane and led to the National Halothane Survey in the United States by Bunker in 1969, which showed that the incidence of hepatic toxicity following halothane was no greater than that after earlier agents. However, there is undoubtedly a problem of ‘unexplained hepatitis following halothane (**UHFH**)' as well as other toxic effects on various body organs after both volatile and intravenous agents.

Halothane hepatitis The quoted incidence of UHFH varies in different countries and with different age populations. The rate in adults is between 1:25 000 and 1:36 000 exposures, with a lower rate in children (1:82 000–1:200 000 exposures). Severe hepatic necrosis is characterized by fever, jaundice, grossly elevated serum aminotransferases, and massive centrilobular necrosis. The condition is fatal in about 75 per cent of cases. Halothane does not function as a classical liver toxin: there is no clear dose dependence for liver damage, no recognized mechanism of action (but see later), and nor are halothane's effects consistent in different animal species.

Halothane is metabolized by both oxidative and reductive pathways in the smooth endoplasmic reticulum of the hepatocyte, with the production of trifluoroacetic acid and inorganic chloride, bromide, and fluoride. The toxicity leading to hepatic necrosis is thought to be an immunological phenomenon initiated by oxidative metabolism. The unstable intermediate compound trifluoroacetyl bromide binds to liver macromolecules (probably proteins), so forming trifluoroacetylated protein (**TFA**–protein) neoantigens. In susceptible individuals, these neoantigens stimulate the formation of anti-TFA–protein antibodies, and upon subsequent exposure to halothane these mediate the hepatic necrosis. Although the oxidative pathway of halothane metabolism causes TFA–protein formation in most individuals, antibodies are only produced in a few. A crucial variable in determining the antigenicity of halothane seems to be the rate of halothane oxidation and the extent of TFA–neoantigen formation. The main enzyme responsible for the oxidative metabolism of halothane is an isozyme of cytochrome P_{450}, CYP2E1.

Halothane can cause two types of liver damage: mild damage manifest as small, transient increases in serum transaminases, seen in up to 50 per cent of all patients, and massive hepatic necrosis. The risk for massive necrosis is increased with repeated exposure to the agent (especially when re-exposure is within 6 weeks), women and older patients are at greater risk, and other predisposing factors include obesity and enzyme induction. The CYP2E1 isozyme is present in increased quantity in the morbidly obese, and is readily inducible by ethanol and isoniazid. This enzyme has been shown to be responsible for the *in vivo* metabolism of halothane and other volatile agents (methoxyflurane, enflurane, isoflurane, and sevoflurane). Other epidemiological factors for halothane hepatitis (genetic predisposition, age over 40 years, female sex) are probably not related to CYP2E1 metabolism.

The exact incidence of halothane hepatitis is still uncertain, but repeated exposure should be avoided unless there are good clinical indications for the agent's use (for example, in patients with chronic obstructive lung disease, and in the relief of life-threatening bronchoconstriction and hypoxia in patients with severe acute asthma). The recommended [by the Committee on Safety of Medicines] interval of 3 months between halothane anaesthetics would seem illogical if there were a presumed immune background in at least some cases. It is also recommended that halothane be avoided if undefined fever occurred after a previous halothane anaesthetic. Cirrhosis, chronic hepatitis, and liver tumours are not contraindications *per se*.

Other volatile agent-associated hepatic damage Enflurane, isoflurane, and desflurane have all been reported to cause hepatic necrosis, although less often than after halothane. There is also some evidence of cross-reactivity between antibodies in the serum of patients with halothane-associated hepatotoxicity and antigens produced by the metabolism of these other agents. TFA metabolites are also produced by metabolism of isoflurane and desflurane; there is some evidence that enflurane metabolism can also produce TFA–proteins that would be recognized by the antibodies from patients with halothane hepatitis.

The incidence of hepatitis following enflurane is low (fewer than 1:800 000 patients), although there are several case reports of patients who have antibodies cross-reacting to both halothane and enflurane, and isolated cases of presumed isoflurane-induced hepatitis. There have been three case reports of unexplained postoperative liver injury following sevoflurane anaesthesia in the Japanese literature. Their exact cause is unclear, as sevoflurane metabolism is not thought to proceed via reactive metabolites. There is only one unequivocal case of desflurane-associated liver damage.

Renal dysfunction Enflurane is metabolized to form significant amounts of free inorganic fluoride ions, with maximal concentrations normally of the order of 25 to 30 µmol/l. However, increased nephrotoxic concentrations (> 50 µmol/l) may occur during prolonged enflurane anaesthesia, in patients being treated with enzyme-inducing drugs (e.g. ethanol or isoniazid), and in chronic renal failure. There have been occasional reports of patients with pre-existing renal disease developing renal failure after enflurane anaesthesia.

In recent reports, even high concentrations of enflurane or isoflurane were not associated with nephrotoxicity or major impairment of renal tubular function; this would suggest that it might not be the peak serum fluoride concentration that is important, but rather the area under the concentration–time curve. The hepatic breakdown of sevoflurane by CYP2E1 gives hexafluoro-isopropanol and inorganic fluoride ions. However, because of sevoflurane's low tissue solubility and rapid elimination, even after prolonged anaesthesia, peak serum fluoride concentrations are not associated with nephrotoxicity. The absence of any organ toxicity despite a serum fluoride of greater than 50 mmol/l suggests that the classical argument for the ‘peak fluoride concentration' being the causative factor may not correct. An alternative theory is that there is intrarenal breakdown of these anaesthetic agents, and that it is the intrarenal not the plasma concentration that matters. Whichever of these theories is correct, the anaesthetist should probably use isoflurane as the maintenance agent of choice.

Sevoflurane is also broken down to haloalkene by-products (described as compounds A and B) by carbon dioxide absorbers (e.g. soda lime). Although these products are nephrotoxic in rats, there are no data to show this effect in man. However, the problem is best avoided by the anaesthetist using a fresh gas flow of greater than 2 l/min. Carbon monoxide formed by the degradation of the volatile agents by soda lime has also been detected; this can be overcome by regular changing of the carbon dioxide absorbent, ensuring that it is at all times moist.

Local anaesthetics

The provision of pain relief, both intraoperatively and postoperatively, can be achieved by a number of different approaches (see earlier). These include:

1. gaseous anaesthetic agents [nitrous oxide, and the 50:50 mixture of nitrous oxide and oxygen (Entonox)];
2. volatile agents (e.g. isoxane);
3. intravenous analgesic drugs;
4. non-steroidal anti-inflammatory agents; and
5. local anaesthetic agents, which can be used to provide topical analgesia, wound infiltration, nerve blocks, plexus blocks, as well as neuraxial blockade.

All of the local anaesthetic agents produce a transient and completely reversible blockade of nerve function and hence an interruption of sensory perception. They produce analgesia by blocking sodium channels in the axon membrane and inhibiting sodium conductance, whilst having minimal effects on potassium currents. Other pharmacological actions relate to their interaction with calcium ions, perhaps by inhibiting their binding to phosphatidylserine.

The anaesthetics bind to a receptor site at the internal opening of the sodium channel, and some drugs such as benzocaine may actually penetrate the nerve membrane and cause conformational changes that lead to a decrease in the diameter of the sodium channel. Most of the local anaesthetic drugs have a pK_a close to the physiological pH, and will therefore exist in both ionized and un-ionized forms. Drugs can only diffuse through the epineurium and nerve membrane in an un-ionized form, the fractions in the two forms being governed by the Henderson–Hasselbach relation (pH = pK_a + log ionized/un-ionized). Blockade of the sodium channels is dependent upon the presence of the ionized form of the local anaesthetic. Following such blockade there is a decrease in the rate and degree of the depolarization phase of the action potential, with failure to achieve the threshold potential and therefore no development of a propagated action potential.

For further consideration of the pharmacological properties and use of local anaesthetic agents, the reader is referred to Chapter 10.2.

Systemic toxicity

Of great importance to surgeons who use local anaesthesia either alone or in combination with general anaesthesia are the problems of toxicity. Accidental intravascular injection of even small amounts of local anaesthetic can cause profound systemic effects. Toxicity is related to the type of local anaesthetic agent, the total dose administered, the site of injection, the rate of injection, and the use of vasoconstrictors. Toxicity is increased in shock, where relatively more of the cardiac output is directed to the heart and brain. Acidosis also increases toxicity.

The toxic effects that occur with lignocaine (lidocaine) are directly related to the blood concentration, as shown in Table 8.

Dose (μg/ml)	*Reaction*
2–4	Antiarrhythmic
	Circumoral numbness
	Dizziness
	Irrational behaviour
6	Visual disturbances
8	Muscular weakness
10	Unconsciousness
20	Respiratory arrest
> 25	Cardiovascular collapse

Table 8 Toxic effects associated with lignocaine (lidocaine)

In general, the cardiovascular system is more resistant to the effects of local anaesthetic drugs than is the central nervous system, the dose ratio being of the order of 3.5 to 6.7:1. However, this ratio is low for bupivacaine, and this resulted in cases of cardiovascular collapse following release of the tourniquet in patients given bupivacaine for intravenous regional analgesia. It also occurred following the use of 0.75 per cent bupivacaine in obstetric epidural practice. Cardiotoxicity appears to be related to the physicochemical characteristics of the local anaesthetic agents (high potency, high lipid solubility, and high plasma-protein binding), and appears to be increased in the pregnant patient. The predisposition to ventricular arrhythmogenicity is due to the presence of butyl groups within the local anaesthetic side-chains (seen with bupivacaine but not with mepivacaine). Toxicity is influenced by acid–base status. Hypercapnia and acidosis both reduce the threshold for convulsive activity and for cardiac depression. Other systemic effects, apart from the allergic reactions and those listed above, include the development of methaemoglobinaemia (with prilocaine), neurotoxicity (to individual nerves), and the initiation of episodes of malignant hyperpyrexia (see Table 9).

Drug	*Dose*	*IM administration (plain)*	*Dose/systemic side-effects*
Amethocaine	100		Cardiac depression; may cause systole and ventricular fibrillation(Now only used as topical/paracaine)
Bupivacaine	175	250	Cardiotoxicity, drowsiness, convulsions
Chloroprocaine	800	1000	Neurotoxicity; prolongation of effect of neuromuscular
Cocaine	200		Sympathetic stimulation, excitement, restlessness, headache, vomiting, myocardial depression, cardiac arrhythmias
Lignocaine (lidocaine)	300	500	Convulsions, twitching, psychotic reactions
Mepivacaine	350	500	Less cardiotoxic than bupivacaine
Prilocaine	400	600	Methaemoglobinaemia
Ropivacaine	200		Possible convulsions in large doses accidentally injected intravascularly. Few side-effects reported to date (end of 1999).

Amide drugs (bupivacaine, lignocaine, mepivacaine, prilocaine) have all been reported to give epidural malfunction (hyperpyrexia).

Table 9 Maximum doses and signs of toxicity for commonly used local anaesthetic agents

If systemic toxicity does occur, oxygen should be administered, with careful monitoring of cardiovascular effects, and convulsive or central excitatory activity should be treated with diazepam or midazolam. Facilities for intubation and ventilation as well as drugs for support of the circulation should be available at all times. No regional anaesthetic should be administered without secure intravenous access and the facility to resuscitate the patient.

Allergic reactions

Allergic responses to local anaesthetic agents in current use are rare (constituting probably less than 1 per cent of all reactions to drugs). In most patients, other causes of drug reactions are responsible, for example systemic toxicity, simple fainting, or reactions to added adrenaline (epinephrine). The recommended maximum dosage of these agents is shown in Table 9.

The ester group of local anaesthetics is more liable to provoke adverse reactions as these drugs contain both a *p*-aminobenzoic acid group, and methyl- or propyl-paraben as the added preservative. The systemic effects of adrenaline are undoubtedly the cause of some of the reported adverse reactions to the local anaesthetics. Use of 1/80 000 adrenaline in dental anaesthesia is frequently accompanied by tachycardia, palpitation, and chest tightness, although angina is rare.

Hypersensitivity reactions to the amide anaesthetic agents are unusual, but allergy to lignocaine has been reported.

Specific aspects of anaesthesia

There are two further areas of anaesthetic practice where the surgeon may benefit from an awareness of the possible problems encountered by the anaesthetist.

Outpatient anaesthesia

Outpatient anaesthesia has progressed in both the United States and Europe from the performance of simple procedures under local anaesthesia to the total anaesthetic care of patients with complex medical problems. This allows costs per patient per operation and disruption of the patient's personal life to be reduced, and also decreases the risk of exposure to hospital-acquired infections.

Furthermore, a number of key developments have improved outcome from day-case anaesthesia: the introduction of appropriate short-acting drugs; the introduction of the laryngeal mask airway and the cuffed oropharyngeal airway; and the concept of a multimodal approach to the management of perioperative pain.

Preoperative evaluation and patient selection

Patients scheduled for outpatient surgery must be willing and able to comply with the pre- and postoperative instructions. Although initial experiences in most day-surgery units were limited to **ASA** (American Society of Anesthesiologists classification of physical status) group I and II patients, many units are now accepting medically stable ASA group III patients. The imposition of a rigid age range is illogical, as there appears to be no age-related increase in recovery time or the incidence of postoperative anaesthetic complications. However, there are a number of patient groups that it is inappropriate to consider for day-case surgery:

1. patients with unstable medical conditions (ASA III or IV; especially those with disease affecting the cardiorespiratory; hepatic or renal systems, and ‘brittle' diabetes mellitus);
2. known cases of malignant hyperpyrexia or siblings/ relatives of affected patients who have not been previously screened for the condition;
3. patients receiving type A monoamine oxidase inhibitors (e.g. tranylcypromine, phenelzine, isocarboxazid);
4. patients with complex morbid obesity (weight more than 160 per cent for height or body mass index greater than 35);
5. patients liable to acute substance abuse;
6. social difficulties (for example, the single patient with no carer for the first night after surgery);
7. patients with congenital or acquired anatomical abnormalities of the face and upper airway (but these may be acceptable if reviewed by an anaesthetist before day-case anaesthesia and surgery).

Careful selection of patients is of prime importance to the efficient running of a day-surgical unit. Details of patients' past medical and drug histories can be obtained readily by a simple questionnaire completed at the outpatient clinic. In many centres, this questionnaire alone is used as the major screening tool in deciding suitability for day-case anaesthesia and surgery. A full history and physical examination are required before surgery and this can be conducted either by the surgeon in the outpatient clinic or by the anaesthetist. The anaesthetist may see patients either in a special preanaesthetic assessment clinic or by prior arrangement in the day-care unit.

Preoperative assessment should also include the facility for performing simple laboratory tests, depending on the patient's age, state of health, and intercurrent drug history. For young, healthy outpatients undergoing body-surface surgery, there is no evidence of the value of any routine laboratory tests for males, and only a haemoglobin estimation is worthwhile for females. Patients with controlled chronic diseases (hypertension, diabetes mellitus) will require additional laboratory screening (electrolytes, urea or creatinine, blood sugar) as appropriate.

Premedication

The use of premedication in the outpatient setting has been the subject of much debate and interest. There is no evidence that it necessarily prolongs the recovery period, and indeed the use of appropriate short-acting analgesic or sedative drugs may decrease recovery times as a result of their ability to reduce anaesthetic requirements. Suitable sedative anxiolytics include midazolam (1–3 mg intravenous) or zopiclone (2.5–7.5 mg orally).

Because outpatients may have a greater residual gastric volume than inpatients at the time of induction, many authorities have recommended the routine administration of oral antacids, with or without gastrokinetic agents, before outpatient surgery. Unfortunately, colloid antacid suspensions can produce serious pulmonary sequelae if aspirated, and the reliability and efficacy of a single dose (30 ml) of sodium citrate has been questioned. Moreover, the use of oral antacids will *per se* increase the residual gastric volume. Histamine H_2-receptor antagonists and the gastrokinetic agents metoclopramide and cisapride are of greater efficacy.

Since prolonged fasting does not guarantee complete gastric emptying, there is further controversy over the length of the period of fasting before outpatient surgery. There is good evidence that preoperative hunger and thirst contribute significantly to preoperative anxiety, as well as increasing postoperative drowsiness, dizziness, fatigue, and nausea. Furthermore, prolonged fasting may result in the patient arriving in the anaesthetic room with a plasma glucose concentration of less than 4 mmol/l. Ingestion of 150 ml of water as late as 2 h before surgery has been reported to decrease significantly the severity of thirst without increasing the gastric volume in fasted outpatients.

Anaesthesia

The ideal outpatient anaesthetic should comprise a rapid and smooth onset of hypnosis, intraoperative amnesia and analgesia, good surgical conditions, and a short recovery period with minimal or no complications. For induction, propofol is the drug of choice as it has two added advantages over other induction agents by providing a feeling of euphoria on awakening and reducing the incidence of nausea and vomiting.

Although propofol by infusion may also be used for maintenance, the volatile agents are the most commonly used world-wide, with the recently introduced sevoflurane offering faster awakening and fewer side-effects than isoflurane. Use of desflurane (while offering the advantage of fast recovery) may be accompanied by respiratory-tract irritation leading to coughing, postoperative nausea and vomiting, and sympathetic stimulation. Intraoperative analgesia should be provided by short-acting opioids (fentanyl, alfentanil, or remifentanil) supplemented by non-steroidal anti-inflammatory drugs and local anaesthesia. The increased incidence of headache after dural puncture in this population and the duration of action of currently available agents (bupivacaine) make spinal anaesthesia a less suitable option for day-case surgery.

There are no good reasons to exclude endotracheal intubation as part of any outpatient anaesthetic technique, although the introduction of the laryngeal mask airway and the cuffed oropharyngeal airway has reduced the need for intubation in many circumstances and hence the use of neuromuscular-blocking drugs. Together these have had a significant effect on the incidence of postoperative sore throats and the use of drugs to reverse neuromuscular blockade (e.g. atropine, glycopyrronium), which are associated with an increased incidence of postoperative nausea and vomiting. If required, muscle relaxation should be provided by short-acting drugs such as mivacurium and rocuronium.

Of greatest importance to the efficient and safe conduct of outpatient anaesthesia and surgery is the seniority of the operators. The outpatient facility is not the remit of either junior anaesthetists or surgeons in training; cases should be managed by senior personnel experienced in the types of anaesthesia and surgery best suited to the patients.

One of the major causes of surgical readmission after outpatient surgery is inadequate pain relief; hence the provision of adequate analgesia is of paramount importance. Although the injection of long-acting local anaesthetic drugs (bupivacaine, mepivacaine) at the site of incision may help to decrease requirements for postoperative analgesics, pain can be controlled in many patients with conventional oral analgesic drugs such as codeine, paracetamol, and non-steroidal anti-inflammatory agents.

Postoperative nausea and vomiting is a further problem after general anaesthesia and can delay discharge as well as leading to unexpected hospital admissions from outpatient facilities. Factors that increase the incidence of nausea and vomiting include body habitus, type of surgery (laparoscopy, orchidopexy, strabismus surgery, therapeutic termination of pregnancy), assisted ventilation using a face mask (with the resultant passage of air into the stomach), and poor choice of anaesthetic agents. Droperidol (5–7.5 mg/kg intravenous) has been found to be an effective prophylactic antiemetic in both children and adults undergoing treatment as outpatients. Other regimens to prevent nausea and vomiting in ‘high-risk' patient groups include intraoperative metoclopramide or ondansetron.

Discharge criteria

There is a need for all day-case units to establish clear guidelines for the discharge of patients. There are no ‘gold standard' tests available, but clinical assessment should include the following:

1. patient awake, alert, orientated, and responsive;
2. minimal pain;
3. no active bleeding;
4. vital signs stable;
5. minimal nausea and no vomiting;
6. oxygen saturation greater than 94 per cent on room air; or in patients with respiratory disease, returned to baseline.

Management of the patient with a severe head injury requiring extracranial surgery

Often there is need for the provision of anaesthesia and surgery to the patient who has suffered a severe head injury. In such cases, as well as those problems that are

relevant to any neurosurgical patient, there are five other important aspects of management:

1. possible presence of a full stomach;
2. possible associated neck and/or facial injuries leading to potentially difficult intubation;
3. associated extracranial injury resulting in hypovolaemia, anaemia, shock, aspiration, other organ injury (e.g. lung, heart, liver, spleen);
4. the presence of raised intracranial pressure;
5. postoperative intensive-care management.

Pathophysiology of space-occupying lesions

Trauma produces a large increase in brain bulk, due to clot formation, localized oedema from damaged cells, and cellular disruption; gunshot wounds produce the greatest rise in intracranial pressure because of the effects of the velocity of the projectile. However, an episode of bleeding is followed by a marked rise in pressure due to the clot formation and cerebral oedema associated with damage to neurones.

Control of intracranial pressure

Before any extracranial surgery the surgeon should liaise with neurosurgical colleagues to discuss the need for control of the intracranial pressure.

Osmotic diuretics produce a transient reduction in total body water, including that in the cerebrum. Mannitol is the principal agent currently employed, a dose of 1 to 1.5 g/kg being given intravenously over about 20 min. Mannitol may cause a rise in blood pressure if administered rapidly, and in the presence of damage to the blood–brain barrier it may leak out and cause regional fluid retention.

Hypoxia and hypercapnia have a synergistic effect, increasing cerebral vasodilatation and hence intracerebral blood volume. Arterial oxygen tension should be maintained above 100 mmHg and $PaCO_2$ should not exceed 40 mmHg (5.3 kPa). During anaesthesia, $PaCO_2$ should be maintained in the range 30 to 35 mmHg (3.5–4.5 kPa). The elderly patient is less tolerant of prolonged hypocapnia, which has been shown to lead to postoperative memory impairment.

Hypothermia is effective in reducing cerebral metabolic needs and hence provides protection against cerebral hypoxia. However, the possibility of overshoot to below a core temperature of about 32°C in the multiply traumatized patient can lead to an increased risk of ventricular irritability and difficult-to-treat ventricular arrhythmias occur.

Although steroids (e.g. hydrocortisone and dexamethasone) may be useful in reducing the oedema that occurs around some intracranial tumours, they have no role in reducing oedema in cases of acute head injury.

Anaesthesia for patients with severe head injury

Patients with a post-resuscitation Glasgow coma score of less than 8 have, by definition, a severe head injury. Some patients who have a higher score may be very difficult to assess, due to the lack of co-operation accompanying confusion associated with raised intracranial pressure.

A detailed account of the anaesthetic considerations in all neurosurgical conditions is not possible here, but aspects relating to the more frequently encountered problems will be mentioned.

Before the start of anaesthesia and surgery, all trauma patients with a head injury should have a thorough examination to exclude possible sites of concealed blood loss or active bleeding. Assessment should also include the possibility of damage to the cervical spine, pneumothorax, or lung trauma, as these will have a marked bearing on intra- and postoperative management. The combination of significant head and chest injuries is usually an indication for postoperative respiratory support because of the importance of maintaining optimal oxygenation postoperatively and the difficulties of performing adequate chest physiotherapy in people with impaired consciousness. Furthermore, patients suffering profound rises in intracranial pressure may manifest pulmonary oedema; the cause is uncertain but it is believed to be mediated centrally via catecholamine release. Hypertension also frequently accompanies severe raised intracranial pressure.

Laboratory investigations should include serum urea and electrolytes, haemoglobin, and clotting studies. Chest radiography forms part of the usual assessment of the traumatized patient. Electrocardiography may indicate the presence of myocardial contusion.

Management of trauma patients is influenced by the need to protect the airway because of faciomaxillary damage and impaired conscious state, the need for respiratory support because of impaired respiratory drive or thoracic problems, the management of acute blood loss from other sites, and the possibility of coexisting injury to the cervical spine. Anaesthesia is usually induced on the operating table with full monitoring equipment attached to the patient. All trauma patients should be assumed to have a full stomach and thus a rapid-sequence induction technique is used. Damage to the cervical spine should usually be assumed unless it is actually excluded. Maintenance of anaesthesia will vary with the anaesthetist's preference and patient's condition, but the use of agents causing cerebral vasodilatation or an increased cerebral metabolic rate should be avoided.

In addition to routine anaesthetic monitoring, intra-arterial pressure is mandatory.

The aim of good cardiovascular management is to maintain a cerebral perfusion pressure (mean arterial pressure – intracranial pressure) of at least 70 mmHg.

Patients with severe multiple injuries or significant combined head and chest injuries usually require respiratory support in the intensive care unit. No hard and fast rules can be given for postoperative management, but the maintenance of optimal respiratory function is of paramount importance. The PaO_2 should be maintained above 100 mmHg (13.3 kPa) and the $PaCO_2$ should be at the low end of normal range (40 mmHg, 5.3 kPa). Antiepileptic medication is frequently given intraoperatively because of the risk of seizures.

Postoperative pain relief

There is an increased awareness of deficiencies in the traditional management of peroperative pain. In recent years there has been rapid expansion in knowledge of both the mechanisms of pain and alternative approaches to pain control. Major problems in pain management have always been the objective assessment of pain and ‘observer bias' by medical and nursing staff. By defining pain as ‘what the patient says hurts', we can remove this bias. The inadequacy of the traditional intramuscular opioid regimen is generally acknowledged and the incidence of inadequate pain relief by this method varies between 25 and 70 per cent.

A patient's analgesic requirements may be affected by a large number of factors: these include site of surgery and type of incision, anaesthetic factors (including type of anaesthesia and incidence of vomiting), and patient factors (including age and sex, concurrent drug therapy, and psychological factors, such as cultural background, past experience, the understanding of pain, fear, and anxiety, and the ability of the patient to cope). Past bad experiences can affect later treatment of pain. [Drugs should be given for pain relief on a regular basis rather than as requested; in this way, the quality of pain relief is improved. Clearly the patient must also be able to exercise their right of declining any medication.]

Physiological effects of pain

Pain and its inadequate control may affect many body systems, resulting in morbidity. Reduction of alveolar minute ventilation and functional residual capacity caused by diaphragmatic splinting and impaired function of intercostal muscles cause decreased Pao_2 and increased $PaCo_2$. There may also be decreased cough, sputum retention, atelectasis, and infection. Sympathetic stimulation causes tachycardia and hypertension, which can lead to myocardial ischaemia, increased peripheral resistance, and increased myocardial O_2 consumption. Reduced mobility because of pain may increase the incidence of deep venous thrombosis and pulmonary thromboembolism. Increased gastric stasis and reduced intestinal motility increase the incidence of postoperative vomiting. Other problems include urinary retention, restlessness, anxiety, increased postoperative confusion, and impaired sleeping. The endocrine effects include increases in cortisol, catecholamines, aldosterone, and antidiuretic hormone; this results in sodium and water retention, and the anti-insulin effects of these catabolic hormones worsen the control of diabetes.

Opioid drugs

Opioid receptors were first identified in 1973 and the endogenous opioids were isolated in 1975. These receptors are one of the main mechanisms of action of

analgesic drugs. For a recent review of this expanding topic, the reader is referred to papers in a recent (1998) postgraduate issue of the *British Journal of Anaesthesia* (editor: DG Lambert; see Further Reading). One of the main problems in the provision of good analgesia with parenteral or oral opioids is the avoidance of side-effects. These are generally dose dependent, with the major adverse effects being respiratory depression and suppression of the cough reflex. There are also effects on the central nervous system include sedation, euphoria, miosis, nausea and vomiting, and muscle rigidity after high doses of opiates.

Other effects include myocardial depression (dose dependent), bradycardia, reduced myocardial oxygen consumption, vasodilatation due to direct histamine release, delayed gastric emptying, and reduced gastrointestinal motility. Some opiates may cause spasm of the sphincter of Oddi. Retention of urine and itching may also occur. Potency varies between opiates, but the maximal effect is similar.

Approaches to pain management (Table 10)

Opioids
Intramuscular (as required)
Oral opioids
Intravenous infusion
Patient-controlled analgesia
Epidural opiates, bolus or via syringe pump
Local anaesthetic agents
Nerve blocks
Interpleural catheter
Plexus blocks
Indwelling epidural catheter
Non-steroidal anti-inflammatory drugs (NSAIDs)
Combination techniques
Epidural opioids and local anaesthetic agents
NSAIDs and opioids

Table 10 Different therapeutic approaches available for the management of acute postoperative pain

The so-called multimodal approach to pain relief employs combinations of different agents and techniques to control postoperative pain. This method of using analgesic drugs allows the doses of opioids to be reduced, with an associated reduction in the incidence of undesirable effects such as sedation and respiratory depression. Combining opioids with local anaesthetic agents and non-steroidal anti-inflammatory drugs further reduces total opioid dosage and increases the safety margin. The role of anxiety in postoperative pain and the importance of reducing anxiety by explanation and reassurance should be appreciated and incorporated in overall pain management.

Enteral opioids

These can be given as solutions or tablets, or as sustained-release preparations (e.g. morphine sulphate, sustained-release tablets, MST). However, the vomiting and delayed gastric emptying that follows some surgery means that drugs are poorly absorbed and have variable efficacy. In addition, some opioids undergo extensive first-pass or presystemic metabolism in the walls of the intestine and the liver before entering the circulation; this reduces their bioavailability. The effective dose is less than the administered dose.

Intramuscular opioids, as required

With this regimen a fixed, prescribed dose of opioid is given as required; there are many problems with this. Failure may occur because of pharmacokinetic factors, such as variability in absorption, distribution, metabolism, and excretion. Plasma concentrations of morphine are a poor reflection of the drug's concentration in the central nervous system. Pharmacodynamic factors may also result in failure of pain relief: the minimum effective analgesic concentration shows a four- to fivefold variability in surgical patients undergoing the same operation. Administration factors also play a part. The delay between the patient perceiving pain and nursing staff being able to administer an opiate means that there is frequently a significant painful period between doses. The interpatient variability in dose requirement is not appreciated and patients with higher dose requirements are inadequately treated.

Intravenous infusions

One of the disadvantages of intramuscular or intravenous bolus dosing is the constantly changing plasma drug concentration, causing either toxic effects or subtherapeutic, ineffective concentrations over a period of time. Using programmed intravenous infusions to achieve therapeutic concentrations may alleviate these problems. A drug infused at a constant rate takes five half-lives to reach steady state. For morphine, which has a $T_{½}$ of 1.5 to 4 h, 20 h may be needed to reach a stable analgesic level. Thus a loading dose is needed at the start of an infusion and whenever an increase in plasma concentration of the drug is required, yet this is often forgotten when altering infusion rates. Problems arise when a plateau is established that is significantly above or below the ‘therapeutic window'. In addition, drug or metabolites may accumulate, particularly when rates are not reduced with reducing requirements. In general, requirements fall markedly after 24 to 48 h and at night during sleep, allowing for a reduction of infusion rates.

Respiratory depression and hypoxia are a significant complication with all opioid administration techniques but are especially seen with constant-rate intravenous infusions; hence careful respiratory monitoring and assessment of the level of sedation are mandatory. Because the rate of increase of the plasma opioid concentration is slow, the infusion allowed may not provide total analgesia, so necessitating intravenous bolus top-up doses.

Patient-controlled analgesia

This modification of intravenous infusion techniques uses a computerized syringe pump to deliver bolus doses whenever the patient presses a button. A lock-out period ensures that the full effect of the dose is achieved before the patient can deliver more drug. Advantages include a reduction of swings in blood concentration and fewer side-effects, the removal of observer bias or judgement of analgesic requirements, and the tailoring of drug requirements to changing analgesic needs. Some studies have also suggested a reduced total opioid requirement.

Local anaesthetic blocks

The most commonly used local anaesthetic agents are lignocaine (lidocaine), prilocaine, and bupivacaine. The duration of effect varies with the site of the block and the drug used. Most blocks will provide 2 to 4 h of analgesia; occasionally the effect will last for 8 to 12 h. The advantages of this technique are that profound analgesia can be produced without the side-effects of opioids and without respiratory depression. However, the toxic side-effects of local anaesthetics limit the quantity of agent that can be used. There is a risk of damage to adjacent structures or tissues, causing for example pneumothorax, haemorrhage, infection, and, uncommonly, permanent nerve damage. There is also a risk of intravascular injection causing toxicity at surprisingly low doses.

Techniques for the administration of local anaesthetic blocks include wound infiltration, nerve blocks, epidural blocks, paravertebral blockade, and interpleural blockade. The advantage of the last two techniques is that, unlike epidural blockade, there is no effect on limb or bladder innervation and less likelihood of sympathetic blockade.

Enteral non-opioid drugs

These may be classified into the peripherally acting non-steroidal anti-inflammatory drugs and centrally acting drugs such as paracetamol and nefopam. Paracetamol (acetaminophen) has analgesic and antipyretic effects, but no anti-inflammatory activity; high doses can cause hepato- and nephrotoxicity. Non-steroidals all act by inhibition of prostaglandin synthase, but their potency as inhibitors of this enzyme does not parallel their therapeutic efficacy; other factors are involved. Since they are weak acids, they accumulate in the parietal cells of the stomach and inhibit local prostaglandins that have a protective role in the mucosa. This action is possibly responsible for the high incidence of gastric irritation, although newer drugs have a lesser effect on the stomach. It may be that inhibition of cyclo-oxygenase 2 (the inducible form of the enzyme) is responsible for therapeutic effects. Newer drugs showing selectivity for cyclo-oxygenase 2 include meloxicam; they may have a reduced incidence of gastrointestinal and renal problems.

Established drugs that have a role in postoperative pain relief are diclofenac, piroxicam, and ibuprofen. They have several important properties: when used as

adjuvants to conventional therapy; they have been shown to exhibit a ‘morphine-sparing' effect; in conjunction with opioids, they may lessen the incidence of opioid-induced side-effects; and they can be given orally, rectally, or intramuscularly. These drugs are the current first-line treatment for mild to moderate pain, and especially for bone pain.

Spinal and epidural opiates and local anaesthetic agents

The epidural space is bounded by the ligamenta flava and the dura. By placing a catheter in this space, continuous administration of epidural drugs (local anaesthetics and opioids) is possible. More lipid-soluble opiates tend to remain at the dermatome segments where administered, whereas morphine (water soluble) may spread via the cerebrospinal fluid to the brainstem. Delayed respiratory depression, including apnoea, has been documented up to 18 to 24 h after morphine administration, with incidences of between 0.1 and 0.4 per cent after epidural dosing and 4 to 7 per cent after intrathecal administration. This side-effect is less of a problem when more lipid-soluble drugs (for example fentanyl and sufentanil) are used. The advantages of the epidural route over conventional intravenous techniques include superior analgesia, improved lung function, less sedation and respiratory depression, and improved mobility, as well as earlier discharge from hospital. Disadvantages include pruritus (in up to 85 per cent of patients receiving epidural opiates), nausea and vomiting (in up to 50 per cent of patients), and urinary retention with opiates. All of these side-effects can be rapidly reversed with naloxone.

Local anaesthetics such as bupivacaine block afferent (sensory), motor and pain fibres, and can also cause sympathetic blockade, which may lead to hypotension. Reduced concentrations are used to minimize these complications. Combinations of local anaesthetic and narcotic are synergistic at the spinal-cord level and are being used increasingly.

Other complications of the epidural route include epidural haematoma, neurological sequelae at the time of catheter insertion or removal, and, rarely, infection.

Perioperative fluid balance

The healthy adult requires approx. 2500 ml per day of fluid to replace losses (100–200 ml via the gastrointestinal tract; 1000 ml by insensible respiratory and cutaneous losses; and 1000 to 1200 ml as urinary loss). Renal sodium conservation results in a daily average requirement of about 75 mmol; potassium conservation requires a daily intake of about 50 mmol/day.

Intraoperative fluid losses must compensate for any acute reductions in the effective intravascular and interstitial fluid volumes that arise from trauma, haemorrhage, and tissue manipulation. There are also surgical losses from the site of the operation, as well as fluid sequestration (the so-called third-space effect), which is found in association with wound or burn oedema, ascites, and loss into the lumen of the intestines. For example, following upper gastrointestinal surgery, there may be loss of up to 15 per cent of the interstitial fluid volume (i.e., about 1.5–2.0 l). Thus, while minimally traumatic surgery requires a fluid replacement of about 4 ml/kg per hour, it may be as high as 10 to 15 ml/kg per hour during abdominal or aortic surgery. These interstitial fluid losses should be replaced by either crystalloid or colloid, the latter having the advantage of remaining longer in the circulation. However, colloids have been associated with significant side-effects (for example, possible coagulopathy following volumes of dextran or starches in excess of 20 ml/kg; impairment of blood cross-matching for subsequent transfusion; anaphylactic and anaphylactoid reactions).

In addition, the surgical patient will also start off with a preoperative fluid deficit secondary to fasting and bowel preparation. This deficit should be replaced in tandem with any intraoperative requirement on the basis of 50 per cent during the first hour and 25 per cent each during the second and third hours of surgery, most suitably given as Hartmann's solution.

Abnormalities of fluid balance during surgery and the perioperative period

The homeostatic control of reabsorption and excretion of water and salts by the kidney is achieved through a number of different hormones and mechanisms, several of which are influenced by the stress responses to anaesthesia and surgery.

Water excretion is impaired during the perioperative period, as a result of increased secretion of antidiuretic hormone or arginine vasopressin. Sodium excretion is also reduced, while excretion of potassium and nitrogen is increased as a result of the increased concentrations of corticosteroids (glucocorticoids and mineralocorticoids) present in the body during this period. The increased release of antidiuretic hormone, which may persist for up to 72 h postoperatively, is due both to osmotic and non-osmotic factors. Sodium retention occurs as a result of the neuroendocrine response to anaesthesia and surgery. There is an increase in plasma aldosterone release from the adrenal glomerular cells, as well as an alteration in renal haemodynamics with redistribution of renal blood flow away from the tubular regions.

All anaesthetic agents cause a decrease in urinary volume and an associated increase in urinary osmolality. The decrease in urinary volume is brought about by a decrease in renal blood flow and a secondary decrease in glomerular filtration rate. This reduction in urinary volume influences the quantity of perioperative fluid required by the patient.

In contrast to those factors above that promote salt and water retention, atrial natriuretic peptide, a peptide that induces a diuresis, natriuresis, and moderate kaliuresis, has now been isolated from cardiac atrial tissue. It increases the glomerular filtration rate without altering total renal blood flow. It also causes vasoconstriction of the efferent glomerular vessels, inhibits angiotensin II, antagonizes the effects of noradrenaline (norepinephrine), and blocks renin release and hence aldosterone secretion. Data are conflicting on the effects of anaesthesia and surgery on concentrations of atrial natriuretic peptide in man. Other hormones that promote intrarenal vasodilatation and salt excretion include the prostaglandins (D_2, E_2, and I_2) and the kinins (bradykinin and kallidin), which enhance the effects of the prostaglandins and modulate the renin–angiotensin system.

Other perioperative hormonal changes influencing fluid balance include increases in growth hormone and prolactin in response to the stress and trauma of surgery. Both of these hormones may lead to salt and water retention.

Perioperative complications

Intraoperative complications

Since 1980, numerous studies have been published on critical incidents during anaesthesia. The following complications are amongst the most significant.

Patient care

The anaesthetized patient has lost his or her normal protective reflexes and is therefore vulnerable to a variety of traumatic incidents including slips or falls during transfer, pressure sores, corneal ulcers, diathermy burns, electrocution, and peripheral nerve palsies. The anaesthetized patient is vulnerable to nerve injury because of the loss of protective reflexes. Especially vulnerable are the ulnar nerve at the elbow, the lateral popliteal nerve during lithotomy, the brachial plexus (lower nerves during abduction, and upper plexus in the Trendelenburg position), and the supraorbital nerve.

Aspiration of gastric contents

This is a major cause of anaesthetic morbidity. As outlined above, the traditional period of fasting before surgery has become increasingly questioned in recent years. It may be harmful to deprive children of fluids for this amount of time, and more recent recommendations involve feeding with clear fluid to within 2 h of surgery. If emergency surgery is required, then waiting 4 h does not guarantee that gastric emptying will occur: after trauma, gastric emptying effectively ceases and hence delaying surgery does not decrease the risks of aspiration. Most anaesthetists would now elect to induce anaesthesia as early as practicable and then use Sellick's manoeuvre (cricoid pressure) combined with a rapid-sequence induction to reduce the incidence of aspiration. If aspiration occurs, prompt intubation and airway suctioning, and the maintenance of adequate oxygenation, are the cornerstones of therapy. Cricoid pressure should be applied to reduce the chance of further aspiration if intubation is delayed. The urgency of the intervention and the extent of inhalation should be considered when deciding whether to proceed or to defer surgery. If saturation remains satisfactory postoperatively in the recovery room, no further treatment is necessary. Inadequate or falling saturation is an indication for monitoring in a high-dependency area. The role of steroids is debated and the use of prophylactic antibiotics is no longer encouraged.

Failed endotracheal intubation

This is a leading cause of anaesthetic-associated death, particularly in obstetric practice. The surgeon should be prepared to take a role in the management of cases where the trachea proves impossible to intubate and the patient cannot be oxygenated. Emergency tracheostomy is a hazardous procedure, and associated with

significant morbidity and mortality in this setting.

Transurethral resection of the prostate (TURP) syndrome

Absorption of irrigation fluid following transurethral resection of the prostate may lead to neurological and cardiovascular changes. The incidence of this complication has been estimated to be as high as 3.9 per cent of patients undergoing urological or gynaecological endoscopic surgery. Water has been replaced by glycine as an irrigant because of problems due to water toxicity. Glycine, however, may also be absorbed into the circulation, causing dilutional hyponatraemia, hyperammonaemia, and fluid overload, which cause neurological and circulatory changes. Neurological symptoms include apprehension, disorientation, nausea and vomiting, visual disturbances, and coma or seizures. Symptoms occur between 15 min after surgery has commenced up to several hours after it has ended. Cardiovascular changes include bradycardia, raised central venous pressure, hypertension, angina, and electrocardiographic changes.

Several factors affect the onset of the TURP syndrome, including the hydrostatic pressure of the irrigant, the experience of the surgeon (which affects the number of venous sinuses opened), duration of surgery, peripheral venous pressure, and the type of fluid used for irrigation. The clinical effects of hyponatraemia relate both to the speed of onset and the extent of the fall in serum sodium. A gradual fall is better tolerated neurologically than is an abrupt one. As glycine is metabolized to other amino acids, ammonia is produced, and this contributes to the development of neurological sequelae, in combination with hyponatraemia (Table 11). In addition, an elevated serum glycine (above 4000 mmol/l) has been associated with visual disturbances.

Na^+ (mmol/l)	Neurological state
134–143	Normal range
< 120	Restless, confused, mild disorientation
< 115	Nausea, drowsy
< 100	Seizures, coma

Table 11 Effect of alterations of body chemistry during transurethral resection of prostate on level of consciousness; relation between serum sodium concentration and neurological state

Air embolism

This may occur when veins at a level above the heart are held distended by bone rather than collapsing due to atmospheric pressure. Conditions predisposing to this include neurosurgical procedures performed in the sitting position and prone laminectomies, and during reaming of the femur in total hip replacement. It may occur in any operation involving veins above the level of the right atrium. Rarely, gases may be injected into the circulation directly, for example during laparoscopy (carbon dioxide embolism). The features of air embolism depend on several factors, including the volume of gas involved, speed of gas entrainment, and the presence of a patent foramen ovale through which gas may reach the left side of the heart, causing symptoms in the coronary and cerebral circulation.

Delayed return of consciousness following general anaesthesia

Although the most common cause of delayed return of consciousness is drug overdosage, both the surgeon and the anaesthetist should be aware of other possible causes (Table 12).

1. Drugs
2. Hypercapnia and/or hypoxia
3. Metabolic, e.g. diabetes mellitus (uraemic or non-uraemic coma, hyperglycaemia), hypothyroidism
4. Hypothermia
5. Raised intracranial pressure, e.g. in patients with undetected space-occupying lesions who are subjected to significant fluid shifts such as those accompanying major blood loss (e.g. during gastrointestinal haemorrhage or vascular surgery, during transurethral resection of prostate, or associated with surgery for major trauma)
6. Psychiatric

Table 12 Causes of delayed return of responsiveness following surgery

Postoperative complications of anaesthesia and surgery

If complications occur in the recovery room or the early postoperative period, the importance of consultation with the anaesthetist who gave the anaesthetic cannot be overemphasized. The anaesthetist may be able to suggest other causes for the problem, and may wish to see the patient to discuss these problems further.

Respiratory

The rate of postoperative respiratory complications is very variable (ranging from 3 to 76 per cent in different series). About 4 per cent of patients will require airway support in the recovery area to avoid obstruction. Respiratory depression is most commonly due to opioids used for pain relief. However, other causes may include oversedation, recurarization, or the development of pulmonary oedema.

When respiratory depression is severe, immediate respiratory support is necessary, using a self-inflating bag-valve mask with supplementary oxygen. Factors that increase the need for mechanical ventilation include a significant history of tobacco smoking, low preoperative arterial blood saturation, and large intraoperative blood loss and accompanying fluid replacement. Table 13 lists some of the common postoperative respiratory problems and their causes.

Complication	Cause
Hypoventilation	Residual anaesthetic effects
	Pain
Hyperventilation	Intracranial pathology
	Pain
Bronchospasm	Pre-existing disease
	Secretions
	Airway instrumentation
Pulmonary oedema	Fluid overload
	Cardiac failure
	Acute respiratory distress syndrome
Pulmonary embolism	Deep venous thrombosis
Aspiration	Vomiting
	High residual gastric volume

Table 13 Common postoperative respiratory complications

The primary cause of postoperative hypoxaemia is pulmonary shunting of blood secondary to the decrease in the functional residual capacity, abdominal distension leading to impairment of diaphragmatic function, and use of the prone position during surgery. Oxygen saturations of less than 90 per cent are commonly seen, and require treatment by supplemental oxygen therapy as well as optimal analgesia and intensive physiotherapy. Occasionally, bronchoscopy may be required to remove sputum. Bronchospasm is usually only seen in patients with a past history of asthma or chronic obstructive pulmonary disease, but may rarely present in the recovery area as a complication of an allergic or anaphylactic reaction.

Pulmonary embolism is rarer, but should be suspected as the possible cause of any unexpected cardiopulmonary collapse. Precipitating factors include obesity, the oral contraceptive pill, old age, prolonged immobility, recent hip surgery, and intra-abdominal malignancy.

Cardiovascular

Instability of the cardiovascular system is not uncommon in the postoperative period. Cardiac failure occurs when the heart is unable to cope with the additional stress of fluid shifts and drug-induced depression of myocardial contractility. Clinical manifestations range from dyspnoea, which may mimic asthma in mild cases, to frank pulmonary oedema with frothy sputum. Management involves optimization of oxygenation, posture and diuretics, and pre- and afterload reduction with glyceryl trinitrate; in severe cases, intermittent positive-pressure ventilation may be required. The electrocardiogram should be reviewed, as ischaemia or arrhythmias will worsen cardiac function.

Postoperative hypertension may be due to pain, or to the preoperative withdrawal of antihypertensive medication. Optimal pain relief should be ensured before further antihypertensive medication is given. Initially, drugs should be given intravenously to reduce delays and to ensure that reliable blood concentrations are achieved.

Hypotension is most commonly due to haemorrhage coupled with inadequate fluid replacement. Drain tubes should be checked for correct function and concealed blood loss should be excluded. Following spinal or epidural anaesthesia, the functional sympathectomy may cause hypotension.

In the absence of demonstrable fluid problems, ischaemia, arrhythmia, and drug-induced myocardial depression should be excluded. Uncommon causes of postoperative hypotension include relative cortisol deficiency in steroid-dependent patients and subclinical hypothyroidism.

Atrial fibrillation is the most common arrhythmia arising postoperatively. Patients previously maintained on digoxin may suffer arrhythmias on cessation of therapy or due to poor absorption in the presence of abdominal pathology. Following electrocardiographic confirmation of the arrhythmia, specific therapy should be commenced. Rapid atrial fibrillation with haemodynamic instability may require intravenous verapamil or, in severe cases, d.c. countershock.

Pre-existing cardiac disease, pain, poorly controlled hypotension, intraoperative events, and suboptimal oxygenation, especially in combination with hypertension or tachycardia, may lead to ischaemic events in the perioperative period.

Other significant arrhythmias include tachycardias (both sinus and nodal) and bradycardias. Often the latter are of a junctional type and arise through the interaction of volatile anaesthetic agents and neostigmine. These rhythms may resolve spontaneously, but should otherwise be treated: tachyarrhythmias using small doses of b-adrenoceptor-blocking drugs or d.c. shock, and bradyarrhythmias with atropine or isoprenaline.

Temperature control

Abnormalities of temperature in the postoperative recovery period are critical, as hypothermia prolongs the effect of anaesthetic agents as well as increasing the incidence of shivering, with its effects of increased myocardial oxygen consumption and patient discomfort. Hypothermia also adversely affects coagulation.

Fluid and electrolyte imbalance

Complications of fluid and electrolyte imbalance are often seen in the elderly or debilitated patient, in the hypertensive patient treated with diuretics, and in diabetic and neurosurgical patients. Hyponatraemia, hypocalcaemia, and hypermagnesaemia have all been implicated in delayed return of consciousness.

Nervous system

Confusion is common in the perioperative period, especially in the elderly patient. Diagnosis is frequently difficult and management often suboptimal. Diagnosis is frequently made by exclusion of possible causes and in many cases no obvious cause for the acute brain syndrome is ever discovered (Table 14). Relatively inexperienced house staff (interns) often have to manage patients with acute postoperative confusional states. Hypoxia must be excluded, either by oximetry or blood-gas estimation. Review of the anaesthetic chart or recovery-room notes will often reveal a likely cause, but in the majority none is ever ascertained. Management involves reassurance of the patient and staff, combined with measures to prevent damage to suture lines, intravenous equipment, and wound drains. Sedation should be used cautiously if at all. Adequate analgesia must be ensured.

Hypoxia
Pain, particularly in patients who are still drowsy
Alcohol withdrawal
Drugs:
Benzodiazepines (especially in elderly patients)
Ketamine: emergence delirium can occur, especially in bright, noisy surroundings
Antiparkinsonians (dopaminergic)
Scopolamine premedication in 'elderly' patients
Drugs of addiction. Especially in emergency cases, illicit use may have preceded the accident that necessitated the operation
Pre-existing psychological or psychiatric conditions unfrequently overlooked
All reversible organic causes must be excluded or treated before attempting to sedate the patient

Table 14 Causes of postoperative confusion in the general surgical patient

Miscellaneous complications

Vomiting

This is one of the most common and distressing postoperative complications. The incidence of vomiting ranges from 10 to 50 per cent, depending on the type of surgery. Many factors contribute to the incidence of vomiting, including the use of opiates and volatile anaesthetic agents, the type of surgery (gynaecological surgery has a very high incidence), gastrointestinal distension (due to ileus), early ambulation, and female sex. Recent advances in reducing the high incidence of what is an unpleasant side-effect include the preoperative administration of drugs to improve both gastric emptying and prevent nausea (e.g. metoclopramide); the prevention of inadequate perioperative hydration; and the recognition of high-risk surgical procedures, such as laparoscopy, strabismus repairs, orchidopexy, and termination of pregnancy.

Urinary retention

This is a frequent complaint, especially in male patients confined to bed after surgery. Inability to pass urine may be related to fluid deficiency, pain, or difficulties

managing bottles and bedpans, especially in noisy or crowded wards.

Dental damage

This is not uncommon during airway instrumentation. The anaesthetist should be notified immediately, to allow an early dental consultation postoperatively. Crowned, capped, and carious teeth are especially vulnerable to damage during anaesthesia and surgery. Damage may be caused at intubation, by oral airways, or during suctioning.

Rashes

Rashes may be caused by reaction to anaesthetic agents, antibiotics, adhesive dressings, or skin preparation solution. Management is generally conservative, but well-demarcated lesions related to areas of adhesive or skin preparation require follow-up to prevent recurrence in future operations.

Sore throat

The incidence of sore throat following endotracheal intubation varies between 2 and 70 per cent of cases. Predisposing factors are the use of red-rubber endotracheal tubes, cigarette smoking, difficult or traumatic intubation, prolonged intubation, and prior laryngeal pathology. Conflicting results have been found with high-volume, low-pressure cuff designs used for short-term intubation. The management of postintubation sore throat is conservative; reassurance is usually all that is required. Sore throat also occurs with the laryngeal mask airway and other airways.

Muscle pains

The development of muscle pains is common in fit, ambulant, muscular young individuals given suxamethonium to facilitate endotracheal intubation. The pain may be quite severe and resembles that caused by unaccustomed exercise. Management involves notification of the anaesthetist concerned, reassurance of the patient, and simple analgesics.

Perioperative deaths

Ever since the first cases of death associated with anaesthesia, both surgeons and anaesthetists have either formally or informally audited their complications. At present, in the United Kingdom, it is the responsibility of the anaesthetist to report to the coroner (or his/her equivalent in Scotland and Northern Ireland) cases where either death has occurred within 24 h of surgery, or where it has occurred later but the patient has failed to recover complete consciousness after the operation. Similar mechanisms exist for reporting perioperative mortality in the United States and many other countries in the Western world. However, until recently, there were few data available to allow determination of the incidence of perioperative mortality in different patient groups.

In 1982, Lunn and Mushin published the results of their study on deaths occurring within 6 days of anaesthesia. This was a voluntary study encompassing five regions in the United Kingdom. Their conclusions were that 0.6 per cent of patients die within 6 days of surgery and that in only 1 in 10 000 cases was anaesthesia solely responsible for death.

In 1987, the Confidential Enquiry into Perioperative Deaths (CEPOD) was published. This study spanned the 30 days after surgery and involved three regions in the United Kingdom; it excluded certain patient groups (e.g. cardiac surgical patients, children and neonates, obstetrics patients). The last group is audited in the United Kingdom by the Triennial Confidential Enquiry into Maternal Mortality. Only three deaths were judged solely due to anaesthesia, giving a rate of 1 in 185 000. The main conclusions of the enquiry were the need for the availability of proper recovery facilities; the need for provision and utilization of essential monitoring; and the need for adequate supervision of trainees. Of particular importance is the consultation between surgeon and anaesthetist regarding emergency cases and those patients suffering from intercurrent illness. Too frequently, there is inadequate notification to allow optimization in management of those patients suffering from multisystem disease.

Subsequent surveys have stressed the high mortality associated with emergency or urgent surgery, as well as the need for adequate thromboembolic prophylaxis in high-risk surgical patients. Another common theme of these reports has been the need to develop specific protocols for the care of the critically ill surgical patient. This incorporates the need for a team approach involving surgeons, anaesthetists, and specialist physicians.

It is inappropriate to adopt these guidelines for emergency and urgent surgery if hospitals do not have appropriate postoperative care facilities.

Further reading

Aitkenhead AR, Jones RM, ed. *Clinical anaesthesia*. Churchill Livingstone, Edinburgh, 1995.

Baron JF *et al.* Combined epidural and general anesthesia versus general anesthesia for abdominal aortic surgery. *Anesthesiology* 1991; **75**: 611–18. [In a study of 173 patients receiving either intraoperative general anaesthesia or general anaesthesia supplemented with epidural analgesia, there was no difference in the postoperative incidence of cardiovascular and respiratory complications.]

Bode RH *et al.* Cardiac outcome after peripheral vascular surgery: comparison of general and regional anesthesia. *Anesthesiology* 1996; **84**: 3–13. [Choice of three anaesthetic techniques was not associated with differences in cardiac or overall mortality in patients undergoing peripheral vascular surgery.]

Cheng DC *et al.* Early tracheal extubation after coronary artery bypass graft surgery reduces costs and improves resource use. A prospective randomized controlled trial. *Anesthesiology* 1996; **85**: 1300–10. [In a randomized study of 100 patients, use of an early extubation anaesthetic technique was associated with reduced stay in and costs of intensive care by 25 per cent; it did not appear to alter the incidence of, or cost associated with, complications.]

Christopherson R *et al.* Perioperative morbidity in patients randomized to epidural or general anaesthesia for lower extremity vascular surgery. Perioperative Ischemia Randomized Anesthesia Trial Study Group. *Anesthesiology* 1993; **79**: 422–34. [Carefully conducted epidural and general anaesthesia were associated with similar rates of cardiac and other major complications in these surgical patients, but epidural was associated with a lower incidence of reoperation for inadequate tissue perfusion and may therefore be advantageous for this group.]

Frank SM *et al.* Perioperative maintenance of normothermia reduces the incidence of morbid cardiac events. *Journal of the American Medical Association* 1997; **277**: 1127–34. [In 300 patients with either documented coronary artery disease (CAD) or at high risk of CAD, and undergoing abdominal, thoracic or vascular surgical procedures, perioperative maintenance of normothermia was associated with a reduced incidence of morbid cardiac events and ventricular tachycardia.]

Kurz A, Sessler DI, Lenhardt R. Perioperative normothermia to reduce the incidence of surgical wound infection and shorten hospitalization. *New England Journal of Medicine* 1996; **334**: 1209–14. [In a double-blind, randomized study of 200 patients undergoing colorectal surgery, maintaining normothermia intraoperatively was associated with a significant decrease in surgical wound infections (6 vs 19 per cent) and a reduced duration of hospital stay.]

Lambert DG, ed. Recent advances in opioid pharmacology. *British Journal of Anaesthesia* 1998; **81**: 1–84. [A postgraduate educational issue.]

Millar JM, Rudkin GE, Hitchcock M. *Practical anaesthesia and analgesia for day surgery*. Bios Scientific, Oxford, 1997.

Modig J, Borg T, Karlstrom G, Maripuu E, Sahlstedt B. Thromboembolism after total hip replacement: role of epidural and general anesthesia. *Anesthesia and Analgesia* 1983; **62**: 174–80. [Epidural anaesthesia reduced the postoperative incidence of deep venous thrombosis and pulmonary embolism, as well as reducing blood loss.]

Morgan M, Hall GM, ed. *Short practice of anaesthesia*. Chapman and Hall Medical, London, 1998.

Sear JW. Practical treatment recommendations for the safe use of anaesthetics. *Drugs* 1992; **43**: 54–68.

Tangkanakul C, Counsell CE, Warlow CP. Local versus general anaesthesia in carotid endarterectomy: a systematic review of the evidence. European Journal of Vascular and Endovascular Surgery 1997; 13: 491–9. [In a meta-analysis of 17 non-randomized and 3 randomized trials comprising about 6100 patients, local anaesthesia was associated with significant reductions (about 50 per cent) in stroke, myocardial infarction and pulmonary complications, as well as reducing hospital stay.]

Tramer M, Moore A, McQuay H. Meta-analytic comparison of prophylactic antiemetic efficacy for postoperative nausea and vomiting: propofol anaesthesia vs omitting nitrous oxide vs total i.v. anaesthesia with propofol. *British Journal of Anaesthesia* 1997: **78**: 256–9. [In a meta-analysis of over 10 000 patients, propofol anaesthesia alone or omitting N_20 had similar effects on both early and late postoperative vomiting; propofol also reduced the incidence of nausea; there are insufficient data with total intravenous anaesthesia using propofol to indicate whether it has a low emetogenic potential.]

Wheatley RG, Shepherd D, Jackson IJ, Madej TH, Hunter D. Hypoxaemia and pain relief after upper abdominal surgery. *British Journal of Anaesthesia* 1992; **69**: 558–61. [In a comparison of 40 patients recovering from upper abdominal surgery, a significantly greater proportion (47 per cent) in the patient-controlled analgesia group rated their analgesia as excellent compared with the control

intramuscular group (10 per cent); there were no differences between the groups in the incidence of postoperative hypoxaemia.]

White PF, ed. *Outpatient anesthesia.* Churchill Livingstone, New York, 1990.

Yeager MP, Glass DD, Neff RK, Brink-Johnsen T. Epidural anesthesia and analgesia in high risk surgical patients. *Anesthesiology* 1987; **66**: 729–36. [In a randomized, double-blind study of 53 patients, use of epidural anaesthesia and postoperative analgesia was associated with a significant decrease in the overall postoperative complication rate, and in the incidence of cardiovascular failure and major infectious complications.]

10.2 Approaches to local and regional anaesthesia

Michael Dobson

Definition
Advantages
Local anaesthetic drugs
The right drug and formulation—the right concentration
Risks
Discussing the options—patients, surgeons, and anaesthetists
Sedation during regional techniques
Combined regional and general anaesthesia
Explaining the procedure to the patient
Performing the block
Testing the block
Monitoring the patient and anaesthetic during surgery
Problem-solving
Unwilling patients
'Failed block'
Complications
The postoperative period
Conclusion
Further reading

Definition

Regional anaesthesia (also called local or conduction anaesthesia) consists of the application of drugs to parts of the nervous system to interrupt the passage of nerve impulses. Because drugs interfere in a differential way with transmission, many anaesthetists prefer to use the term regional analgesia. Almost all nerve pathways in the body can be interrupted at some point, so the scope for regional techniques is considerable.

Advantages

By definition, a regional technique affects only part of the body, and is therefore inherently less disruptive. The patient can remain conscious, and by doing so provides assurance of a safe airway and avoids complications caused by anaesthetic interference with breathing function. In contrast, most techniques for general anaesthesia interfere with the control of at least three major systems: consciousness, cardiovascular and respiratory function. For many patients it is the loss of control of these functions that causes them the greatest anxiety in relation to surgery.

A regional technique avoids some of the potential complications of general anaesthesia, including the life-threatening ones of hypoxia from airway obstruction or oesophageal intubation. Induction of general anaesthesia is also associated with abrupt changes in cardiovascular function: hypotension from vasodilatation, sometimes followed by hypertension and cardiac arrhythmias associated with tracheal intubation. During prolonged surgery the patient under general anaesthesia may be exposed to gas mixtures that predispose to postoperative lung collapse; there is a fall in functional residual capacity resulting in shunting and hypoxia, while the persistence of respiratory depression and depressed reflexes in the postoperative period can lead to further deterioration in gas exchange. In contrast, the patient who remains awake is alert in the postoperative period, requires less nursing care in the recovery ward, and can return to the ward or be discharged from hospital earlier.

Regional techniques provide the highest possible quality of postoperative analgesia without the side-effects of respiratory depression and nausea associated with potent opiate analgesics. Pain is easier to control when it can be prevented. The patient waking from general anaesthesia may be aware of pain as the first sensation of returning consciousness. For this reason, regional techniques are often used in addition to general anaesthesia, specifically to provide good analgesia in the early postoperative phase, and increasingly these techniques continue into postoperative care on the ward. Good pain relief can result in early ambulation with a reduction of the complications of lying immobile in bed, and an earlier discharge from hospital. For patients having major abdominal or thoracic surgery, the use of continuous regional anaesthesia for 24 h or more postoperatively allows early extubation and breathing without pain, and full co-operation with postoperative physiotherapy.

When effective regional anaesthesia is established before surgical stimulation, the humoral and metabolic responses to stress normally associated with surgery are attenuated or even abolished. If regional anaesthesia is continued into the postoperative period, this can reduce the protein catabolism and negative nitrogen balance normally associated with major surgery.

In some areas of surgery, for example hip replacement and prostatectomy, the use of epidural anaesthesia has consistently been shown to reduce peri- and postoperative blood loss by around 30 per cent, lessening the need for blood transfusion. There is also an increase in peripheral blood flow in the legs, and an associated reduction in the incidence of postoperative deep-vein thrombosis.

Many patients prefer to remain awake during their surgery, and often perceive the operation as well as the anaesthetic as less threatening if it can be done 'under local'. It is certainly significant that a majority of anaesthetists in a recent survey said they would prefer to have a regional technique if they were themselves the patient!

The components of a regional technique are frequently simpler than the complex apparatus needed to provide general anaesthesia This gives regional techniques an inherent safety advantage: complicated things go wrong more often than simple ones. It is not, however, axiomatic that regional anaesthesia is safe, nor that it is always safer than a well-conducted general anaesthetic. All anaesthetics involve some kind of 'physiological trespass' and require an adequate level of skill and vigilance in the practitioner.

There are, of course, some disadvantages of regional methods. Any technique has a failure rate, although this decreases with the experience and skill of the anaesthetist. In the event of failure, the patient is likely to need general anaesthesia, and must have been prepared for this in advance, both physically and psychologically. This means that the routine for preoperative assessment, preparation, and fasting should in principle not vary, no matter which type of anaesthetic is planned. Such preparation will also safeguard the patient in the event of unforeseen complications related to the regional technique, such as a high spinal or convulsion due to drug toxicity.

Although it is true that there is a failure rate, it is also the case that the success of regional anaesthesia can (and indeed should) be tested before surgery begins. In contrast, one of the most feared complications of general anaesthesia, that of awareness during surgery, can only be discovered afterwards, when it is too late to remedy.

In some situations the establishment of effective regional anaesthesia takes more time than would be needed for a general anaesthetic, and in extreme urgency (e.g. severe fetal distress requiring emergency caesarean section) this can be a significant disadvantage. However, in most instances the extra time spent on a regional technique is recouped in the more rapid recovery and the greater comfort of the patient during the recovery period.

Particular contraindications that may preclude a regional technique include bleeding disorders, infection, and unwillingness on the part of the patient, surgeon or anaesthetist! Patient unwillingness should not be overestimated however: in Oxford, currently, well over 90 per cent of women undergoing caesarean section choose a regional technique. One important reason to avoid a regional technique is when the anaesthetist cannot effectively communicate with the patient, whether because of a language or cultural barrier, or impairment of the patient's cognitive faculties by organic or psychological illness.

Local anaesthetic drugs

Conventional local anaesthetic drugs have membrane-stabilizing properties: they act at sodium channels in cell membranes, preventing the entry of sodium ions that triggers depolarization of the membrane and conduction of a nerve impulse. The pharmacological factors that determine the activity of a drug are its pK_a (speed of onset), lipid solubility (potency), and protein binding (duration of action) (Table 1)

Table 1 Properties of commonly used local anaesthetic drugs

Two chemical classes of local anaesthetic exist: both share the structure of an aromatic ring, an intermediate chain, and an amine tail (Fig. 1). In the aminoester group there is a central ester link; in the aminoamides an amide link. In general the aminoamides are more stable and resistant to autoclaving, and this group contains most of the drugs in common use today.

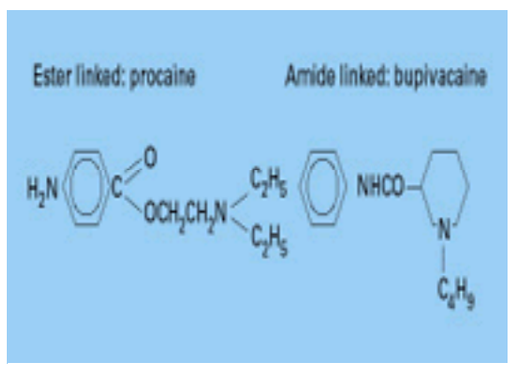


Fig. 1. Structure of local anaesthetic drugs.

The thicker the diameter of a nerve fibre, the greater is the concentration of drug required to block conduction (see Table 2). It is therefore common to observe a differential effect during the development of blockade. This differential is magnified because sensory fibres have a higher firing rate than motor fibres, and this makes them more susceptible to interference from local anaesthetic drugs. If a high concentration of drug is used the block develops, with sequential loss of pain and temperature, muscle-spindle, touch, pressure, proprioceptive, and motor functions. It is not necessary to block all motor fibres in order to produce relaxation, as once the muscle-spindle innervation is affected the reflex arc controlling muscle tone is interrupted, and there is reduced muscle tone even though the muscles are not paralysed.

Table 2 Classification of nerve fibres

The degree of block also depends on the size of the nerve bundle, and the presence and thickness of a fibrous sheath surrounding the nerve. Injected drug will take much longer to act if it has to pass such a barrier or diffuse a long way through the nerve to act on the most central fibres. In most peripheral nerves the fibres are arranged so that those that supply the most distal strictures are centrally placed. It is therefore common for a nerve block to develop first in proximal parts of a limb and to spread distally as the drug diffuses into the central part of the nerve trunk. (Fig. 2)

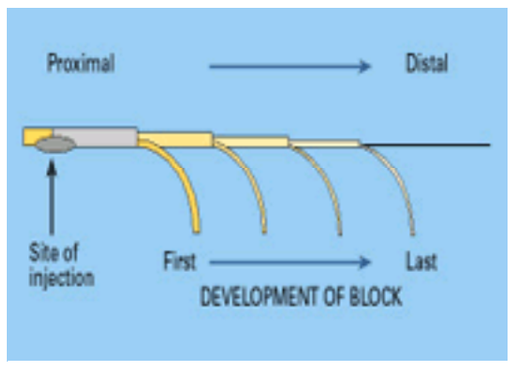


Fig. 2. Development of block in a peripheral nerve.

Local anaesthetic drugs can reach their target by one of a number of routes:

1. infiltration—diffusion to sensory receptors and the finest nerve branches (a similar retrograde diffusion occurs during intravenous regional analgesia);
2. injection close to (preferably not into) a nerve trunk;
3. injection into a plane or compartment through which a nerve runs (e.g. axillary sheath, epidural and subarachnoid spaces).

The right drug and formulation—the right concentration

Choosing the right drug demands knowledge of pharmacology, and of what is actually required in the patient. How important is speed of onset? Lignocaine generally works more rapidly and wears off faster. Long-acting drugs occasionally have a very prolonged offset, which may prevent a day-case patient from being discharged.

The role of added vasoconstrictor must be understood: by delaying the rate of removal of local anaesthetic from the tissues it delays the rise in blood concentrations (and thereby effectively reduces toxicity) and also tends to prolong the duration of the block (this is more noticeable with lignocaine than bupivacaine). Prilocaine is the least toxic agent currently available, and is very suitable for high-dose techniques such as intravenous regional anaesthesia (the resultant methaemoglobinaemia will confuse a pulse oximeter, but adverse effects on the patient are rare). Most agents produce central nervous toxicity at lower concentrations than cardiotoxicity, but with bupivacaine, central nervous and cardiovascular toxicity can occur at similar blood concentrations, and ventricular arrhythmias resistant to treatment can be a feature of overdose. Always calculate the safe dose in mg/kg of drug for your patient before starting to inject.

It is probably true that most of the time the concentration of drug used is higher than necessary, to allow for any inaccuracy in placing the injection. When simple infiltration is used, accuracy of injection is irrelevant, and 0.5 per cent or even 0.25 per cent lignocaine is perfectly satisfactory: the risk of toxicity with 0.25 per cent is negligible. Nerve-blocking techniques (see below) require a higher concentration. Lignocaine 1 per cent or bupivacaine 0.25 per cent are satisfactory, but will probably produce significant motor block, for example quadriceps paralysis in the femoral nerve block that might be used for knee arthroscopy, and may delay the discharge of a day-case patient. More concentrated solutions are available (lignocaine 2 per cent, bupivacaine up to 0.75 per cent) for specialized use.

Multidose bottles of drug are still available in some places, but are not recommended. They invariably contain bactericide, which may cause tissue irritation or even neurotoxicity, in addition to a greater risk for infection. Whenever a glass ampoule is opened, there is a risk of small pieces of glass entering the ampoule; when the injection is to be in a critical site, such as intrathecally, a filter needle should always be used to draw the drug into the syringe.

Risks

There is a widely held view that local and regional techniques are inherently safe, while general anaesthesia is hazardous. This is not the case. Although regional techniques may completely avoid some of the risks of general anaesthesia (e.g. inhalation of gastric acid), local anaesthetic drugs are potentially toxic, especially if accidentally injected into the wrong tissue or compartment. Safe regional anaesthesia therefore requires that all necessary means of resuscitation, including drugs and equipment for tracheal intubation and controlled ventilation, should be available whenever such anaesthesia is planned. Some of the risks of general and regional anaesthesia are shown in Table 3. Toxic reactions and overdoses do occur, often in the context of an inexperienced doctor doing a ‘minor' surgical procedure with the wrong concentration of a drug, unaware of the maximum safe dose. The addition of adrenaline as a vasoconstrictor generally increases the margin of safety by reducing the rate of systemic absorption (provided the drug is not accidentally given into a vessel). If local anaesthetic toxicity occurs it may take the form of central nervous disturbance (restlessness, metallic taste, paraesthesia, convulsions, and respiratory failure), cardiovascular depression or dysrhythmias, or (rarely) true allergy.

General anaesthesia	*Regional anaesthesia*
Airway obstruction	Respiratory depression
Aspiration of gastric contents	Toxicity of drug:
	Depression of the CNS and convulsions
Cardiac depression and dysrhythmias	Cardiac depression and dysrhythmias
Hypotension	Hypotension from neuraxial blocks
Trauma to mouth, pharynx, larynx, and teeth	Accidental intravascular injection
Allergy or hypersensitivity	Allergic reactions
Postoperative hypoxia	Methaemoglobinaemia
Toxic damage to liver or kidneys	Spread of sepsis
Increased intracranial pressure	

Table 3 Complications of general vs regional anaesthesia

Major blocks carry the risk of misplaced injection, for example subarachnoid injection into a dural cuff in the neck, or accidental injection of an epidural dose into a vessel or the subarachnoid space. Major neuraxial blocks are always associated with autonomic blockade, and may precipitate cardiovascular collapse in a patient who is hypovolaemic and depending on vasoconstriction to maintain blood pressure.

In addition to the inherent risks of the drugs, there are also risks associated with the process of injection. Supraclavicular brachial-plexus blocks (and to a lesser extent intercostal blocks) may be associated with the development of a pneumothorax. Accidental puncture of a vessel can result in an expanding haematoma; accidental puncture of the dura with a large, epidural-size needle may cause a severe headache.

Probably the most serious risk is that of human error in injecting the wrong substance by mistake.

In spite of these possible complications, regional techniques can avoid some of the major risks facing ‘unfit' patients. There is minimal interference with respiratory function in those with chronic lung disease. Although the blood pressure may fall, peripheral perfusion may increase as a result of the reduced vascular resistance, and at the same time the work of the heart may fall, reducing the demand on the myocardium.

Discussing the options—patients, surgeons, and anaesthetists

The assumption is often made that regional techniques are linked with minor surgery, and major procedures are associated with general anaesthesia. This assumption is false: regional anaesthesia is used successfully and widely for major procedures such as hip surgery, complex eye surgery, and caesarean section. Although there may be an expectation of a particular type of anaesthesia at the time an operation is initially planned, a definite decision can only be made after the anaesthetist has assessed the patient, in consultation with the surgeon.

There are few absolute contraindications to regional anaesthesia. Refusal by the patient is certainly one of them, but it is important to give the patient a clear understanding of what is involved when explaining the choice of techniques, to state clearly any advantages and disadvantages of a particular mode of anaesthesia, and to make a firm recommendation where there are good medical grounds for preferring one form of anaesthesia.

Often patients need reassurance about complications of which there may be no (or negligible) risk. For example, many people have been told (often by their dentists) that they are ‘allergic' to the adrenaline often added to local anaesthetic solutions. It is not possible to be allergic to adrenaline: such patients have often accidentally received an intravenous bolus of adrenaline-containing solution, and been alarmed by the central nervous effects and tachycardia. They can be normally be assured that they are not allergic, but if genuine allergy to local anaesthetic drug is suspected it can easily be confirmed by skin testing.

The surgeon's preference for a particular anaesthetic technique must be taken into account. The use of regional techniques may often require a modification of surgical technique; gentle handling of tissues and the avoidance of traction on visceral structures, for example. Good communication must be maintained between surgeon, anaesthetist, and patient throughout the procedure. Many surgeons appreciate that the superior conditions of regional anaesthesia during the recovery period (for example the relative absence of coughing, straining, or vomiting) can reduce the incidence of surgical complications.

Many patients dislike injections, and some believe that if they opt for general anaesthesia they will avoid them; the reverse may be true since general anaesthesia, in addition to the necessary pre-anaesthetic intravenous cannula, often results in a need for more intramuscular analgesia postoperatively.

Sedation during regional techniques

In spite of many advantages, surgery can be an ordeal for the fully conscious patient, particularly if they have to remain motionless and the operation is prolonged. Lying perfectly still for 30 min can be very difficult, even on a comfortable bed—and no operating table is a comfortable bed! When the stresses of an operating theatre, and worries about pain and the results of surgery, are added together, it is not surprising that patients may be distressed. It is simple to alleviate such distress without recourse to general anaesthesia. Psychological preparation is the first essential step: explaining to the patient in advance what will happen, asking about their worries and dealing with them. Premedication will also help, providing anxiolysis and also, if necessary, analgesia to cover the discomfort of lying still, stiff joints or backache. Once the block has been inserted, further sedation can be given if desired.

Those who administer sedation must realize that its purpose is to make the patient comfortable and free from anxiety. Patients who are given so much ‘sedation' that

they are unresponsive to voice are actually receiving general anaesthesia by default, often in hazardous conditions in which they may not be adequately monitored.

It is generally helpful to withhold sedation until the quality of the regional block has been assessed, since the patient's full co-operation is needed for this. If the block is found to be inadequate it can be reinforced, or general anaesthesia offered. Sedation will not convert an ineffective into an effective block. If patients experience pain after the start of surgery, it may be appropriate to give an analgesic, but sedation in this situation simply converts a patient in pain into a confused patient in pain, and is unacceptable.

Combined regional and general anaesthesia

There are often good grounds for combining regional and full general anaesthesia. Regional anaesthesia often provides superior postoperative analgesia, and where possible it is more pleasant for the patient if regional blocks can be performed after the onset of unconsciousness (although many major blocks are better done with the patient awake). The use of combined regional and general anaesthesia often provides better operating conditions for the surgeon in terms of good relaxation and reduced bleeding. Even a simple combined technique, such as infiltration of local anaesthetic along a wound line at the end of surgery, can bring significant improvement in recovery and postoperative comfort.

Explaining the procedure to the patient

Getting the patient's confidence and co-operation is a prerequisite. This may be more difficult if there is any barrier to communication, for example with a young child or where there is a language difference, but it is often possible to overcome these difficulties with patience. The patient is understandably anxious; it is doubly important that they know what to expect. Some regional techniques do not produce total neural blockade: for example, the patient having an inguinal hernia repaired under extradural block will probably still be able to move (and even feel) his toes, and may take this as evidence that the block has ‘not worked'. Similarly, a differential block may sometimes result in the retention of touch sensation in areas where pain impulses have been completely blocked; the patient needs to be aware of this. Since all techniques have a failure rate, the patient also needs to be certain that if there is any pain, the anaesthetist will be there and will take immediate action, including, if necessary, general anaesthesia, to abolish it. Most patients will appreciate this specific assurance.

The patient who comes to the anaesthetic room or operating theatre is entering unknown territory. It is only humane to tell them in advance before doing anything to them: not just injecting the block, but also connecting monitoring equipment, measuring the blood pressure, palpating landmarks, inserting intravenous cannulae, and disinfecting the skin. If a large area of skin is to be ‘prepped' (e.g. the back for a spinal), the antiseptic solution should be warmed by placing the bottle in a bowl of warm water.

Performing the block

Most patients will not want to watch needles being inserted, so it is important to tell them just before doing this. Often the most painful part of a regional technique is the initial intradermal injection of local anaesthetic. It is of no importance if the patient moves during this part, and it is therefore wise to say words to the effect ‘I am going to inject some local now; try to keep still, but if you do move it doesn't matter'. Once the initial injection is done the patient can be reassured that the worst part is over, and this often helps them relax. Do not say ‘This is the worst bit' in advance of doing it! Most patients dislike needles, but few are truly needle phobic.

The advent of topical local anaesthetic creams (Emla and Ametop) means that in children, and when necessary in adults, given suitable preparation time, it is possible to give completely painless subcutaneous injections. However, if there is any doubt, it is vital for your credibility and rapport with the patient, especially if a child, to have a truthful approach and tell them that they may feel some brief pain with the injection. Suturing wounds in the casualty department is often difficult, because infiltrating along the wound edge is almost always painful, and because the patient is often already in pain or upset. In this situation a suitable oral premedication can be very useful. Make sure that you wait for several minutes for the drug to work: subcutaneous infiltration does not produce instant analgesia, and painful infiltration followed by painful suturing is as unnecessary as it is unpleasant for both patient and surgeon.

For any regional technique other than minor infiltration it is necessary to have basic clinical monitoring, and to insert a cannula in case intravenous access is urgently needed. Because the insertion of the block demands concentration on the area concerned, it is very helpful to have an assistant to help with the positioning and comfort of the patient. For major blocks such as spinals, extradurals, and brachial plexus, the presence of an assistant is mandatory.

Drugs should in general be drawn up and labelled by the person who will inject them. Injection of the wrong substance is particularly likely to happen if a careless and messy technique is used, with open ampoules and syringes scattered over the work area; it is a good rule to draw up the local anaesthetic drug immediately before giving it (having, of course, checked the ampoule).

Local infiltration blocks the sensory nerve endings and fibres within the tissues, and detailed understanding of local anatomy is probably irrelevant. Where the objective of regional analgesia is to block specific nerve trunk(s), it is preferable to inject local anaesthetic next to the nerve and not directly into neural substance. Pushing a cutting needle into a nerve is undesirable (though it probably happens quite frequently) as it will invariably cut some fibres, although these will normally regenerate.

Placing the local anaesthetic beside the nerve means that a variable time must elapse for the drug to penetrate through the sheath and into the substance of the nerve. This time may be shortened by using a more concentrated solution: up to 2 per cent lignocaine may be needed, for example, to penetrate the sciatic nerve in the thigh. The injection of a large volume of liquid within the fibrous sheath of a nerve will produce pressure effects and further potential long-term damage. For this reason, many anaesthetists avoid blocking any nerve contained in a tough sheath (e.g. the ulnar nerve at the olecranon) for fear of producing permanent damage.

There is considerable variation in the onset times of regional blocks, so it is important to allow enough time for the block to work, and not hurry the patient into the operating theatre as soon as the needle has been withdrawn. Local anaesthetic toxicity is most likely to occur just after the insertion of a block, a further reason for observing the patient quietly for a few minutes before proceeding with surgery.

Testing the block

Before surgery begins it is essential for patient, anaesthetist, and surgeon to be convinced of the effectiveness of the anaesthetic. The anaesthetist will have been monitoring the development of the various stages of the block. Often the first sign is vasodilation of the area concerned (accompanied in the case of neuraxial blocks by a fall in blood pressure), followed by sequential loss of cold sensation, pain, light touch, pressure, and possibly motor function. Patients often spontaneously register the development of anaesthesia, reporting warmth, heaviness, and numbness. It is not necessary to repeatedly apply painful stimulation to check the block; loss of light touch or cold sensation (using ice or ethyl chloride spray) is usually reliable. Explain to the patient that the purpose of testing is to check how quickly and extensively the block is taking effect, and not ‘to find out if the local has worked'. If the block is found to be only partially effective, the patient will then usually have no trouble with a supplementary dose of local anaesthetic. Many anaesthetists prefer to administer an unequivocally painful stimulus (e.g. needle prick) at the anticipated site of incision before surgical preparation starts. This is reassuring for all, most especially for the patient. Surgeons must be dissuaded from making a tentative cut and asking the patient ‘Did you feel that?', as this is bound to cause anxiety to everyone and make the patient more likely to interpret any sensation as a failure of the anaesthetic. Equally, if a patient does report pain, surgery must stop immediately for corrective action to be taken.

Monitoring the patient and anaesthetic during surgery

The most fundamental requirement of monitoring during anaesthesia, whether general or regional, is the presence of an alert anaesthetist at all times with attention focused on the patient. The most effective way of monitoring is by talking to the patient; this also gives them the constant reassurance that all is well, and help is at hand if needed. The most sensitive measures of local anaesthetic toxicity are subjective feelings of light-headedness, numbness of the tongue, and agitation. Similarly, a patient whose blood pressure has fallen rapidly as a result of a spinal or extradural injection will often report symptoms of faintness, dizziness, or nausea well before the fall is detected by blood-pressure monitors.

The level of routine observations and measurements made must relate to the patient's clinical condition as well as to the scope of anaesthesia and surgery: a patient with a ring block of a finger will need rather less attention than one having a hip replacement. Where the regional technique is unlikely to have any significant effect on cardiorespiratory function, including airway integrity, the patient can usually be safely left to take care of their own physiological functions. The patient who has received heavy premedication, sedation, or a major regional block requires monitoring equivalent to that which would be used for general anaesthesia: oxygenation, pulse, blood-pressure and ventilation monitoring as a minimum. For unsedated healthy patients, where pulse oximetry indicates normal oxygenation on air, it is unnecessary to place an oxygen mask, since this is often interpreted by them as evidence that all is not well. Sedated patients do not normally object to an oxygen mask, and benefit from its precautionary use.

In addition to watching the patient's physiological status, the anaesthetist is responsible for monitoring the level and degree of block, since this may change during the

course of surgery. Spinal solutions may continue to ascend for 45 min after injection, especially if the patient's position is altered; if surgery is prolonged, effective blocks may begin to recede, and require supplementation either with more local anaesthetic, or by systemic analgesic or anaesthetic drugs. Monitoring of both the patient and the block continues in the postoperative period to ensure that when the block does wear off, adequate analgesia is available. Increasing use of postoperative extradural infusions (possibly including opiates and other drugs) means that postoperative monitoring of the block may need to continue for many days into the postoperative period.

Problem-solving

Unwilling patients

It is of course unethical (as well as illegal) to force a regional technique on an unwilling patient. If a patient is reluctant, or refuses a regional technique, it is reasonable to try and find out why in a sympathetic way. Many patients have unrealistic fears about what happens during an operation; they may also feel a loss of autonomy or dignity if they are aware of things being ‘done to them' while they are awake. If there are good medical reasons for preferring regional anaesthesia, these must be explained and the patient given sufficient time to consider them, possibly with the advice of relatives. In such situations, patients are often willing to take medical advice, especially if they are promised an anxiolytic premedication and the right to insist on general anaesthesia if they feel the condition is unbearable. Most such patients state afterwards that the regional technique was acceptable, and would not mind having it again.

‘Failed block'

All medical interventions have a failure rate. Patients receiving regional anaesthesia are entitled to know the likelihood of success, and to have explained to them what steps will be taken if the block is not fully effective. Assuming that the anaesthetist has monitored the block (as described above), ‘failed block' will come as a disappointment but not a surprise when it occurs.

If a block is unsatisfactory it is necessary to know whether it is a complete or partial failure. If a complete failure, it is likely that the drug was not injected in the correct place (e.g. if a ‘spinal' injection produced no detectable block after 30 min, it is certain that local anaesthetic was not in fact injected into the subarachnoid space). There may be an option to repeat the initial block if the patient is willing, and the total dose of drug would not exceed the toxicity threshold.

More often a block is produced that is not sufficiently dense to allow surgery at the proposed site. For example, sensation may remain distally in the fingers after an otherwise effective brachial-plexus block. Here, it may be possible to supplement the ‘missing' parts of the block, for example by a suitable combination of nerve blocks at the wrist. It is sometimes the case that only initial skin incision is painful, and local infiltration by the surgeon can cover this. Once the skin has been incised the operation proceeds without further problems, although in such a case it is wise to ask the patient if they are willing to proceed, or would prefer to have general anaesthesia.

If a block begins to recede towards the end of surgery, it is often possible to avoid general anaesthesia by giving the patient systemic analgesics such as fentanyl or morphine by intravenous injection. This is a useful adjunct in caesarean section (after the cord has been clamped), as some surgical manoeuvres (such as swabbing the paracolic gutters or closing peritoneum) can be associated with discomfort even in the presence of quite a high spinal or epidural block.

Complications

Awareness of the possible complications and their incidence is a prerequisite in making an appropriate choice of technique as well as in ensuring that the patient is properly informed and advised.

By far the most common complication of regional anaesthesia is failure of the block, although this complication diminishes with increasing experience on the part of the anaesthetist. Blocks can only fail if the wrong drug is injected, an inadequate dose is used, or injection is made at an incorrect site.

Injecting the wrong drug can be catastrophic for both patient and doctor: most drugs that are not local anaesthetics are likely to be neurotoxic either as a direct effect, or because of the pH or other properties of the formulation injected. As outlined above, it is safe practice to draw up the local anaesthetic immediately before making the injection, after double-checking the label. In any environment where other drugs may have been drawn up, use a tray exclusively for regional needles and syringes.

Injection at the wrong site will normally cause the block to fail. The notable exception is an attempted epidural (large-volume) injection, which, if accidentally injected into the subarachnoid space, will result in a rapidly advancing, total spinal block with severe hypotension and respiratory paralysis. Another ‘wrong' place for the injection to go is into a blood vessel: the toxic doses of local anaesthetics are not calculated on the assumption of intravenous boluses, and hence early toxicity from the local anaesthetic, or of adrenaline mixed with it, is likely. For these reasons, large-volume injections, especially extradurals, are usually preceded by aspiration and a test dose, which it is hoped will reveal subarachnoid or intravascular injection. When an epidural catheter is in use, doses can be given incrementally to reduce the risk of toxicity. Unfortunately, neither a negative aspiration test nor a test dose of either local anaesthetic or adrenaline will detect subarachnoid or intravascular placement in all cases, so it is still necessary to be prepared to treat sudden collapse.

Local anaesthetic toxicity, whether as a result of an overdose oran accidental intravascular dose, is most commonly manifest inthe central nervous system; the patient may notice numbness of the tongue, oral paraesthesias, light-headedness, sensory disturbances, and muscular twitching, progressing to unconsciousness, convulsions, and coma. Cardiovascular toxicity is normally manifest at higher blood concentrations than those which produce central nervous effects, but the latter may be masked if the patient is receiving sedation or general anaesthesia. Compared with other local anaesthetic drugs, bupivacaine is more likely to produce cardiac toxicity and in particular ventricular dysrhythmias, although this risk appears to be less if the laevo isomer rather than the (normally used) racemic mixture is injected. A newer long-acting drug, ropivicaine, also supplied as a pure (laevo) optical isomer, appears to have less cardiotoxicity than bupivacaine. In general, the toxicity of different local anaesthetics is related to their potency; procaine, the least potent, is among the least toxic. Of the amide group, prilocaine is the least toxic, but is not widely used because methaemoglobinaemia results after large doses, although this is rarely harmful to the patient.

The injection of a drug into tissues always carries the theoretical possibility of mechanical damage, either as a result of tissues being cut by the needle, or by other changes secondary to this, for example haematoma or leakage of cerebrospinal fluid resulting in headache. The use of traditional cutting-bevelled needles (such as those used for intramuscular or intravenous injection) is demonstrably associated with a greater degree of nerve damage than the use of short-bevelled, regional-block needles, although some have found that the lesions caused by short-bevelled needles may take longer to regenerate. Leakage of cerebrospinal fluid after spinal anaesthesia, and the resulting ncidence of headache, are much reduced by the use of fine needles (smaller that 25-G) and in particular by the use of pencil-point needles that separate rather than cut the dural fibres, allowing the puncture site to seal almost completely when the needle is withdrawn (see Fig. 3).

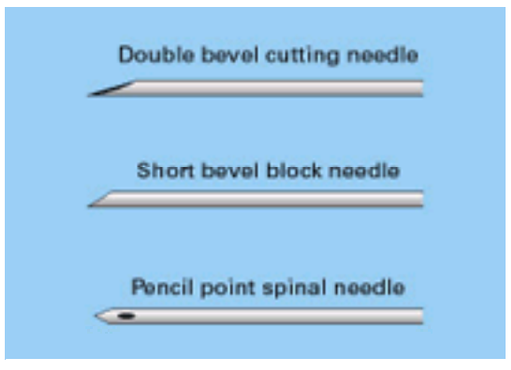


Fig. 3. Needle points and bevels.

At injection sites where a haematoma could form and exert pressure, there is an increased risk of these complications in the presence of clotting abnormalities. Extradural analgesia is therefore generally regarded as inadvisable in the presence of abnormalities of coagulation.

Unless a scrupulously aseptic technique is used, local infection may follow regional techniques. The presence of overt infection at the site of injection is a clear

contraindication to regional anaesthesia.

The postoperative period

Because regional analgesia often persists into the postoperative period it can be used as part of an overall strategy, allowing rapid recovery and early mobilization after almost all types of surgery. The day-surgery patient benefits from a clear-headed recovery, while those having major abdominal or thoracic surgery with extradural analgesia can be weaned more rapidly, and with less pain and distress, from respiratory support; they can co-operate more fully with nursing and physiotherapy staff, and have less chance of developing serious complications.

Increasing use is being made of postoperative infusions for regional analgesia in ward as well as intensive-care/high-dependency settings. Discussion of the application of regional techniques for postoperative anaesthesia, using local anaesthetic drugs, possibly together with other classes of drug such as opiates, is beyond the scope of this short chapter, but in future it will probably play an increasing part in improving the standard of care for patients in the perioperative period. Many hospitals have introduced the concept of an acute-pain service that utilizes regional anaesthesia as a part of a range of techniques to maximize patient comfort and minimize postoperative complications.

Conclusion

Regional anaesthesia has many advantages for the patient, surgeon, and anaesthetist. It is not a ‘minor' form of anaesthesia and demands a high level of knowledge, skill, and the ability to communicate, together with preparedness and ability to treat the rare but serious complications that may arise. The ability to provide rapid recovery and good postoperative pain control means that its use has increased in recent years, with the need for early discharge of inpatients and increased use of day surgery. This trend is likely to continue.

Further reading

Broadman LM, Mesrobian RB, McGill WA. Regional vs general anesthesia: a survey of anesthesiologists' personal preference. *Anaesthesia and Analgesia* 1987; **66**: S20.

Dauphin A, Raymer KE, Stanton EB, Fuller HD. Comparison of general anaesthesia with and without lumbar epidural for total hip arthroplasty: effects of epidural block on hip arthroplasty. *Journal of Clinical Anesthesia* 1997; **9:**, 200–3.

Engquist A, Brandt MR, Fernandes A, Kehlet H. The blocking effects of epidural analgesia on the adrenocortical and hyperglycaemic responses to surgery. *Acta Anaesthesiologica Scandinavica* 1977; **21**: 330–5.

Frank E, Sood OP, Torjman M, Mulholland SG, Gomella LG. Postoperative epidural analgesia following radical retropubic prostatectomy: outcome assessment. *Journal of Surgical Oncology* 1998; **67**: 117–20.

Hall GM. The anaesthetic modification of the endocrine and metabolic response to surgery. *Annals of the Royal College of Surgeons*, 1985; **67**: 25–9.

Jorgensen LN, Rasmussen LS, Nielsen PT, Leffers A, Albrecht-Beste E. Antithrombotic efficacy of continuous extradural anaesthesia after knee replacement. *British Journal of Anaesthesia* 1991; **66**: 8–12.

Tetzlaff JE, Spevak C, Yoon H, Brems J. Patient acceptance of interscalene block for shoulder injury. *Regional Anesthesia* 1993; **18**(1): 30–3.

Voulgaropolos D, Palmer CM. Local anaesthetic pharmacology. In: Prys-Roberts G, Brown B, eds. *International Practice of Anaesthesia*, pp.2/138/1–16. Oxford: Butterworth-Heinemann, 1996.

10.3 Surgical diathermy

John T. B. Moyle

Generators
Waveform
Endoscopic diathermy
Dangers
Further reading

The word diathermy means ‘through heating'. Surgical diathermy is a technique that allows bleeding from small blood vessels to be arrested without the need for mechanical clips or sutures, and enables tissues to be cut without the use of a knife, with sealing of small blood vessels at the same time. It is done by the application of heat to small areas of tissue in a highly controlled way. In microsurgery the heat may be applied with a needle that has been preheated in a flame (for example in ophthalmic surgery), or by applying a piece of resistance wire through which an electric current is passed (as in cautery of the nasal septum). However, the heat is usually generated by passing a radiofrequency electric current through the tissues themselves.

Animal tissue is a conductor of electricity but has a considerable resistivity. Resistivity is the intrinsic property of a material to resist the passage of electricity: the higher the resistivity, the poorer a conductor and vice versa. When an electric current is passed through a poor conductor, energy is lost in the form of heat and the temperature of the material increases. The amount of heat produced depends upon the current (amperes) and the resistance of the tissue (ohms). The heating power developed may be expressed as:

$W = I^2R$

(where W = power (watts, W), R = resistance (ohms, W), and I = current (amperes, A). One watt of power is converted into 1 J/s heat. Surgical diathermy uses 30 to 400 W of power, depending upon the degree of heating required.

Under normal circumstances, passing the required amount of current through the body, either as direct current or at conventional mains frequency (50 Hz in the United Kingdom and Europe, 60 Hz in North America) would cause severe muscle spasm, intense pain, and fatal cardiac arrhythmias. However, as frequency is increased, more current may be passed through the body with fewer of these effects, allowing the use of higher currents and, therefore, higher temperatures (Fig. 1). For this reason, surgical diathermy uses radiofrequency currents in the range 0.5 to 1.5 MHz.

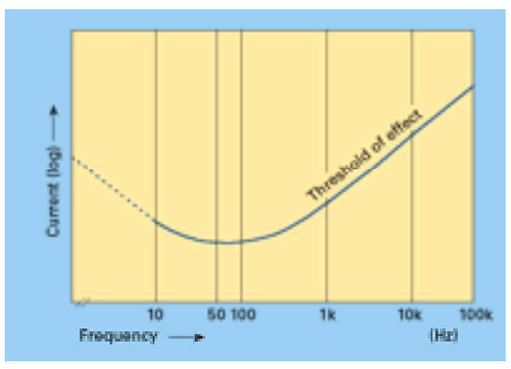


Fig. 1. Variation of the threshold of pathophysiological effects of electric current with frequency.

The application of radiofrequency current across living tissue would cause heating of the whole volume. It is therefore necessary to limit the amount of tissue heated to avoid damaging the surrounding mass. This goal is achieved by ensuring a high current density in the volume of tissue that requires diathermy and a very low current density in all other tissues. The same heating effect will occur along all of the current pathway but the temperature will only rise significantly where the current density is high. Where there is a low current density the rise in temperature will be totally dissipated by the large volume of surrounding tissues. An understanding of the concept of current density is vital to the safe use of radiofrequency surgical diathermy.

There are two ways of delivering a high current density to a localized area: monopolar and bipolar diathermy. (The term ‘monopolar' is inaccurate as it implies a single connection to the patient; however, this would not complete an electrical circuit) With the monopolar technique the current is passed through a large volume of tissue (Fig. 2) from an ‘indifferent' or ‘plate' electrode of comparatively large surface area that is in good electrical contact with a large area of the body. The current then passes through an active electrode of very small contact surface, which is under the control of the surgeon. A very low current density is therefore passed through most of the body, but at the point of contact between the active electrode and the tissues the current density is very high, and therefore has a large heating effect.

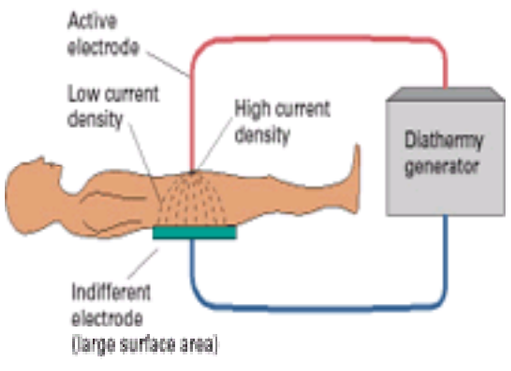


Fig. 2. Two-dimensional representation of monopolar diathermy.

Bipolar surgical diathermy involves the current being passed between two point electrodes, in the form of insulated forceps limbs placed immediately adjacent to the tissue to be heated (Fig. 3). A very high current density, and hence a high heating effect, is produced over a very small volume of tissue, with virtually no heat generated elsewhere in the body. Bipolar diathermy can only be used with relatively low currents and is only suitable for the coagulation of small blood vessels. Its greatest application used to be in microsurgery, especially of the hand, and in neurosurgery, but new equipment is being manufactured for bipolar diathermy during laparoscopic surgery.

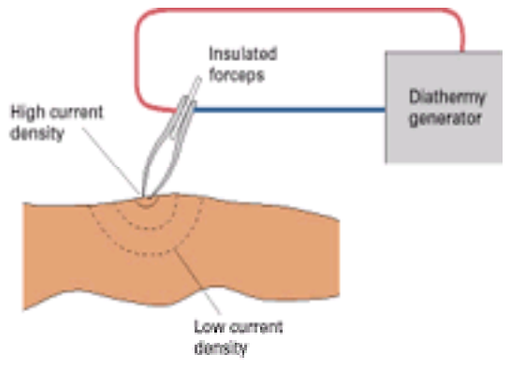


Fig. 3. Two-dimensional representation of bipolar diathermy.

Generators

Generators for radiofrequency surgical diathermy are basically continuous-wave radio transmitters without antennae. The earliest types consisted of a spark gap and a tuned circuit supplied with electricity. Production of an arc across the gap caused the circuit to oscillate at its resonant frequency for as long as current was supplied to it. With the advent of vacuum valves (radio tubes) the quality of the diathermy current improved and it became more controllable. Such valves are expensive and large, requiring power supplies for their filaments as well as high tension for the output current, and are less reliable than the semiconductors that have superseded them.

Modern diathermy generators employ high-power, high-frequency transistors that provide an output equivalent to that available from valve generators. Although the main part of the generator is a high-powered radiofrequency oscillator, semiconductor electronics and microprocessors are used to control the oscillator, monitor its performance, and to monitor safety, especially the connection of the indifferent electrode.

Waveform

The waveforms of the radiofrequency currents used in surgical diathermy have been chosen empirically. Tissue to be cut has to be heated very rapidly, but only in close proximity to the tip of the electrode. Cutting is achieved by striking an arc between the tip and the tissue, thus charring it: a continuous sinusoidal waveform of high power is the most effective. Coagulation requires less heat applied over a slightly greater volume of tissue: this is best achieved by repetitive short bursts of a few cycles of the current. There are some applications, such as transurethral prostatectomy, where a combination of these two waveforms may be blended to achieve more widespread coagulation during cutting.

Many manufacturers produce machines that generate complex waveforms for special situations but their advantages are hard to prove.

Endoscopic diathermy

Radiofrequency diathermy is most commonly used to coagulate small blood vessels, using appropriate forceps, and to cut tissues, using a fine-pointed electrode. Both of these procedures may be done endoscopically using appropriate, insulated electrodes. The most frequent applications for such techniques are the treatment or removal of bladder tumours and hypertrophic prostate tissue. This requires distension of the bladder with a non-conductive fluid, such as a glycine solution, and the use of a blended current. Comparatively high power is required for diathermy in the bladder as the heat energy generated is rapidly dissipated by the large volume of distending liquid.

Appropriate electrodes may also be used via a fibreoptic gastroduodenoscope or a laparoscope.

There has been a dramatic increase in minimal-access or laparoscopic surgery in recent years and at last the manufacturers of equipment are developing diathermy equipment for these techniques. The United States Food and Drug Administration has stipulated that, for monopolar diathermy in laparoscopic surgery, the power of the generator must not exceed 100 W and the peak-to-peak voltage must not exceed 1200 V. Unfortunately, as the diameter of the access decreases in the interests of vanity, so there is an increased energy loss in the abdominal wall. This is because a ‘capacitor' is formed between the tissues of the abdominal wall, the insulation of the cannula, and the electrode; capacitors are insulators to direct current but pass increasing amounts of current as the frequency increases. The more narrow the insulating layer between the tissues and the electrode (as the access is made smaller) the better the capacitive effect. Ideally, bipolar diathermy during laparoscopic surgery reduces the losses as lower power is used.

Dangers

Skin damage is probably the most common surgical problem associated with radiofrequency diathermy. It is caused either by coagulation of blood vessels close to the skin, or by inadvertent contact between a skin edge and a conducting instrument while treating deeper tissues.

Knowledge of the principle of current density obviates the disaster of inadvertent coagulation of pedicles. The classical cause of this problem is the use of monopolar diathermy on the testicle whilst it is supported in the surgeon's hand. The current density is then very high in the spermatic cord, which has a small cross-sectional area and is the sole current pathway between the rest of the body and the testis. Whenever monopolar diathermy is used in this situation, the testicle should remain in contact with the main bulk of the body with, if possible a saline-soaked swab to improve conduction. This disaster can be totally avoided by the use of bipolar diathermy.

Increased current density is also the cause of skin burns through poor contact between the indifferent ‘plate' electrode and the skin. This, and also faults in the indifferent electrode lead, may cause skin burns at other sites as the current attempts to find other pathways to complete the circuit. The generator should have self-diagnosis systems for lead faults, the indifferent electrode must be carefully applied, and the patient's skin must be protected from contact with any other conductive material that may form an alternative current pathway.

The positioning of the dispersive electrode is also important to safety. The site should:

1. have low impedance to the passage of current (fat is a poor conductor/has a high impedance);
2. have a good blood supply to aid dissipation of heat;
3. be reasonably close to the site of surgery, but not be immediately adjacent as a high current density may occur at the edge of the plate nearest the site.

It is also prudent to place the plate away from large metal implants, as these may cause areas of high current density around them.

With the laparoscopic use of diathermy there are some added risks.

1. The electrode insulation may be delicate and must be inspected carefully
2. There is a risk of capacitive currents heating the skin at the entry access.
3. The surgeon places the tip of the electrode using two-dimensional rather than three-dimensional vision.
4. Only the tip of the electrode is visible to the surgeon; there is a risk of inadvertent contact the tissues or instruments outside his field of view. This risk is even greater if the probe insulation has been damaged

There is an enormous variety of internal and external electronic cardiac pacemakers and one must assume that they may all be inhibited or even damaged by radiofrequency diathermy, unless the manufacturer's information states otherwise. If diathermy is vital, then it is necessary to ensure that the pacemaker and its lead are not included in the maximum current pathway: bipolar diathermy is the safer method. There is also an increased risk of arrhythmias and even intracardiac burns due to leakage of current through leads attached to external pacemakers.

Explosion and fire are still risks during diathermy, despite the almost total disappearance of flammable anaesthetic agents. Where flammable anaesthetic agents are

still in use, diathermy should ideally be avoided entirely, although it may be used if it is kept away from the zone of risk.

Factors associated with a high risk of fire include the use of inflammable skin-preparation solutions, especially when there is a likelihood that some of the solution has flowed into skin folds or the perineal area and not been allowed to evaporate. Several such fires have required extensive plastic reconstruction of the groin and perineal area. The atmosphere around a patient during general anaesthesia, especially under the surgical drapes, has an increased concentration of oxygen. Such oxygen rich environments demand that no fuel be provided if a spark is possible. The most obvious example is that all sponges or packing in head and neck surgery be well moistened. Because nitrous oxide supports combustion as well as, if not better than, oxygen, special care must be taken when diathermy is used in the mouth, where the concentration of oxygen/nitrous oxide is always elevated. Methane in obstructed bowel may detonate if struck by a diathermy spark, especially if mannitol has been used as a bowel evacuant.

Further reading

Wattiez A, Khandala S, Bruhat M-A. *Electrosurgery in operative endoscopy*. Blackwell Science, Oxford, 1995.

10.4 Electrical safety in anaesthesia and surgery

John T. B. Moyle

Harmful effects of current electricity

Microshock

Protection against electrical injuries and electrocution
Prevention of gross electrocution in the operating theatre
Prevention of microshock
Further reading

The subject of electrical safety in the operating theatre and elsewhere in health-care facilities is of the utmost importance for the following reasons:

1. electronic equipment is now ubiquitous in health care;
2. in no other sphere of life is the deliberate ohmic connection of electronic equipment to living tissue entertained;
3. the sick patient is more susceptible to the unwanted effects of electrical energy than is the healthy person.

Harmful effects of current electricity

There are a number of ways by which an electric current passing through living tissue may cause damage. With an understanding of these mechanisms, electrical safety enters the realm of common sense.

Risk of electrical damage may occur whenever the body, or part of it, becomes part of an electrical circuit. The amount and variety of morbidity, and the likelihood of fatality, depend upon the magnitude of the current, the time for which it passes through the body, and, to a certain extent, the frequency of the applied current.

From Ohm's law, $I = E/Z$, where I is the current in amperes, E = the potential difference in volts across Z, which is the resistance (or impedance with alternating current). The current passing through the body therefore depends not only upon the magnitude of the applied voltage but also upon the electrical resistance of the body; the lower the resistance, the higher the current. If the skin was perfectly dry, the skin resistance alone would be between 100 000 and 300 000 W. However, water and, especially, perspiration reduce this resistance dramatically, such that the average resistance may be assumed to be between 50 and 10 000 W when assessing electrical risk. Table 1 shows the approximate resistivity of body tissues at 50 Hz.

Material	*Resistivity (Ω/cm)*
Blood	150
Urine	30
Skeletal muscle	300
Cardiac muscle	750
Lung	1275
Fat	3000
Liver	800
Nerve tissue	500
Bone	500–6000

Table 1 Approximate resistivity of body tissues at 50 Hz (derived from Geddes and Baker, 1967)

The adverse effects of excessive currents passing through the living body may be due to one or more of the following.

1. Electrical energy being converted into heat, which will cause damage proportional to the product of time and current; this may even progress to charring (a process made use of in radiofrequency surgical diathermy).
2. Hypoxaemic damage due to respiratory muscle spasm, or temporary cardiac arrhythmia; permanent cardiac damage may also ensue.
3. Chemical burns at contact points due to electrolysis—this type of injury only occurs when there is a direct-current component to the electric current.

Death from electrocution may arise from asphyxia caused by the respiratory muscle spasm, respiratory arrest due to dysfunction of the central nervous system, or because of cardiac asystole or arrhythmia.

Damage to any particular part of the body is dependent upon current density or current per unit cross-sectional area of the current pathway. This concept is illustrated in Fig. 1.

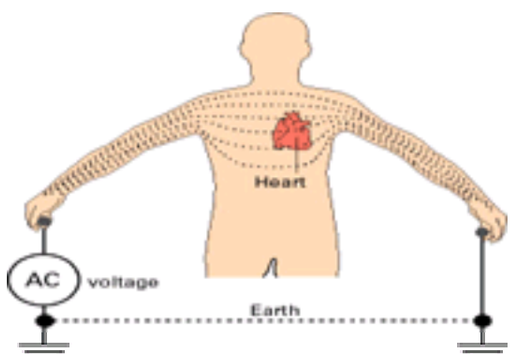


Fig. 1. Current density.

Table 2 shows the effects of increasing current at 50 or 60 Hz through the human trunk. These values vary greatly under different conditions, the most obvious of which is the route that the current takes through the body: the most dangerous is that which passes through the axis of the heart. Other factors causing variation include sex, body weight, and the state of health of the area of contact, and the frequency (see Chapter 10.3), the waveform, and the duration of the electric shock.

Threshold of perception	1 mA
Pain	5 mA
'Let-go' current, severe pain, and muscle spasm	15 mA
Respiratory muscle spasm	50 mA
Ventricular fibrillation	80–100 mA
Sustained myocardial contraction	> 1000 mA

Table 2 The effects of electric current at 50 or 60 Hz through the human trunk

Figure 2 shows the relation between time/current and the probability of the pathological effect.

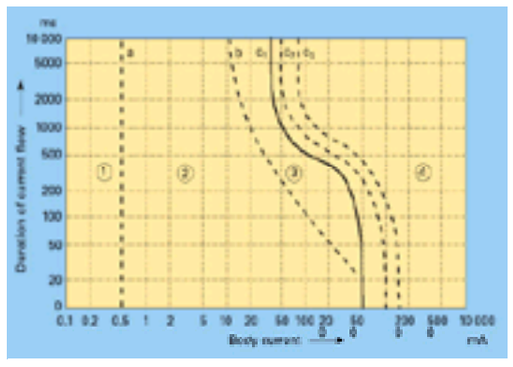


Fig. 2. Time/current zones (see Table 3) of the effects of a.c. currents (15–100 Hz) on humans. The point 500 mA/100 ms corresponds to a fibrillation probability of the order of 0.14 per cent.

Zone	Physiological effects
1	Usually no reaction
2	Usually no harmful physiological effects
3	Usually no organic damage to be expected
	Likelihood of muscular contraction and difficulty in breathing, reversible disturbances of formation and conduction of impulses in the heart, including atrial fibrillation and transient cardiac arrest, without ventricular fibrillation, increasing with current magnitude and time
4	In addition to the effects of zone 3, probability of ventricular fibrillation increasing up to about 5% (curve c2), up to about 50% (curve c3), and above 50% beyond curve c3, increasing with magnitude and time, pathophysiological effects such as cardiac arrest, interruption of breathing, and severe burns may occur

Based on [1].

Table 3 Time/current zones of effects of a.c. currents on humans

Microshock

The concept of microshock is important in anaesthesia, intensive care, and thoracic surgery. If electrical contact is made internally, especially on or close to the heart, very low currents, as small as 10 μA, may initiate arrhythmias. The resistance of skin contact is eliminated and the current density at the interface between the contact and the heart is very high (see Fig. 3). Microshock may occur during thoracic surgery, but the risk is much more common when conductive saline is used as the fluid in a cardiac, pulmonary arterial, or central venous-pressure catheters, as leakage current may pass through a pressure transducer.

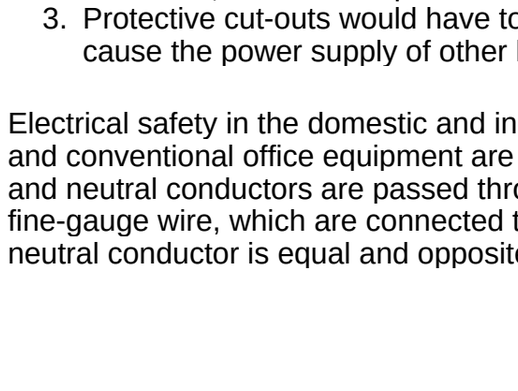


Fig. 3. The concept of microshock.

Protection against electrical injuries and electrocution

The philosophy of electrical safety in medicine has a different bias from that in the domestic or industrial environments. The reasons for this are as follows.

1. Only in the medical environment is direct electrical connection with the body necessary (surgical diathermy, electrocardiogram, electromyogram, electroencephalogram, etc.).
2. Any protective cut-out system must be sensitive enough to protect against currents passing through the body above preset, but very low, values due to fault conditions. Such sensitive devices are very likely to trip-out in non-fault conditions; this would also be dangerous to life in the case of ventilators, dialysis machines, and extracorporeal circulation pumps.
3. Protective cut-outs would have to be built into each piece of electronic apparatus and designed in such a way that a fault condition occurring in one piece did not cause the power supply of other life-supporting equipment to be tripped-out.

Electrical safety in the domestic and industrial environment and for the doctor's surgery (office), where single items of diagnostic apparatus that are not life-supporting and conventional office equipment are used, is nowadays provided by a device called an earth-leakage circuit breaker, the principle of which is shown in Fig. 4. The live and neutral conductors are passed through a ferrite ring, thus forming the single-turn primary of a transformer. The secondary winding consists of many turns of fine-gauge wire, which are connected to a solenoid. Under normal conditions, with no fault in the apparatus supplied through the circuit breaker, the current in the neutral conductor is equal and opposite in polarity to that in the live conductor, and therefore there is no change in the magnetic field and no current is induced to

energize the solenoid.

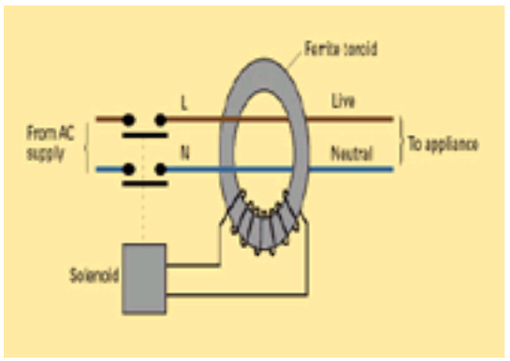


Fig. 4. Principle of the earth-leakage circuit breaker.

If a fault occurs such that electrical leakage current passes out of the system, say, through a human body, then there will be an imbalance of current between the live and neutral conductors and a voltage will be induced to energize the solenoid. Energization of the solenoid mechanically disconnects the live and neutral conductors from the apparatus. The earth-leakage circuit breaker has to be reset manually after the fault condition has been rectified. Earth-leakage circuit breakers are normally arranged to supply more than one socket outlet, which means that a single fault may disconnect the supply from a number of pieces of equipment.

Prevention of gross electrocution in the operating theatre

The earth-leakage circuit breaker is unsuitable for use in the operating theatre and intensive-care unit for the following reasons.

1. Permissible levels of electric current that may be allowed to pass through the sick, and so electrically susceptible, patient are much lower than those permissible in the fit and healthy person.
2. If an earth-leakage circuit breaker is designed to be extra sensitive, it is always more prone to erroneous cut-outs.
3. An electrical current may be deliberately allowed to pass through the patient and may not all return to the equipment in use, but may find an alternative pathway. This may be at extremely low levels in monitoring equipment, but at increasing levels through (muscle or nerve) stimulators, to appreciable values with surgical diathermy.

Because of these problems, the idea in medical electronics is to use ever more rigorous design safety criteria, and patient–circuit isolation rather than earth-leakage circuit breakers.

All equipment that may come into contact with a patient should have been designed and manufactured to comply with the relevant national and international standards. International agreement about safe design is published by the International Electrotechnical Commission (**IEC**) in the Standard IEC 601 (= BS 5724 and BS EN 60601). This standard is published as IEC 601 part 1, which contains all the general requirements for all equipment. An increasing number of parts 2 are being published, which contain particular requirements for each type of equipment.

The main protection against electrocution in the operating theatre and elsewhere in the health-care facility is insulation, which minimizes the leakage current that may pass through the human body. Insulation, which also refers to the protective housing of equipment, must be designed so that possible leakage currents do not exceed levels set out in Standard IEC 601 even if the earth connection is inadequate. Earth conductors at mains-supply outlets should be tested regularly, and the external cables of equipment should also be inspected carefully and regularly. These tests should be made by qualified engineering staff. However, electrical safety is also the responsibility of the users of the equipment, who must:

1. avoid portable distribution boards whenever possible;
2. use ceiling-mounted pendant supplies whenever possible, as they are less likely to be damaged than those on the floor and are unlikely to become wet— keep water and electricity apart.
3. avoid the use of long mains-supply cables, and avoid damage to cables by knotting, equipment wheels, etc.;
4. notify engineering staff of any visible damage to equipment or cables;
5. make sure that regular maintenance records are kept and are available for inspection by the user.

One area where a potential risk of electric shock or burns is not common sense, except to an engineer, is when high-frequency currents are in use, for example with radiofrequency surgical diathermy. In normal use, insulation between two conductors or a single conductor and the body may be entirely adequate for direct current or 50 Hz. However, this insulation may become the dielectric of a capacitor formed between the body and a ‘conductor' at high frequency. (The resistance or impedance of a capacitor decreases as frequency increases.) Thus, at high frequencies, high currents may pass along unexpected routes, causing electrocution or burns. Burns may even occur between the body and metalwork that is not intentionally a conductor but is earthed. Burns due to this mechanism have occurred via the metalwork of operating tables and through the transducers of pulse oximeters (modern pulse-oximeter transducers are isolated so that an earth pathway at high frequency cannot occur).

Thermal heat from the tips of diathermy pencils, ultrasonic scalpels etc. can produce burns apart from electrical shocks. Fibreoptic cables must not rest on patients for this reason, also.

Prevention of microshock

Conventional safety measures may not protect the ‘electrically susceptible' patient sufficiently. IEC 601 classifies the extra requirements in the design of equipment where there is risk of microshock.

Category BF. Equipment having an applied part with intentional connection to the patient (e.g. electrocardiogram, electroencephalogram, electromyogram).

Category CF. Equipment with points specifically designed for application where a conductive connection directly to the heart is established.

In general, the maximum patient leakage current with BF equipment is 100 μA under normal conditions and 500 μA with a single-fault condition. With CF equipment, the maximum currents are 10 and 50 μA, respectively. If CF equipment is connected to a patient, all other devices connected to the patient at that time should be CF rated, otherwise leakage currents may find other routes to the heart.

Particular care must be taken when computing equipment is connected in any way to monitoring apparatus, as most computing equipment is not protected to the same high specification as medical equipment and the safety standard of such monitoring equipment is immediately lowered, by definition, to that of the computing equipment. Some form of electrical isolation should be interposed between the medical device and the computer.

The subject of electrical safety in the health-care environment is complex. Safety is only possible through good maintenance, vigilance, and common sense on the part of the user, and by only using equipment that conforms to IEC 601. However well these guidelines are followed, one must remember that nothing is absolutely foolproof.

Further reading

Geddes LA., Baker LE. The specific resistance of biological material: a compendium of data for the biological engineer and physiologist. *Medical and Biological Engineering* 1967; **5**: 271–93.

IEC 479–1. *Effects of passing current through the human body*. International Electrotechnical Commission, Geneva, 1984.

IEC 601–1. *Safety of medical electrical equipment. Part 1: General requirements*. International Electrotechnical Commission, Geneva, 1988.

Martin TL. *Malice in Blunderland*, p. 5. McGraw Hill, New York, 1973.

10.5 Interpretation of lung-function tests in the surgical patient

M. Jocelyn Morris

Investigations of lung function
Lung mechanics
Gas transfer
Control of breathing
Effects of pulmonary resection
Further reading

Lung-function tests in the surgical patient should allow an appreciation of the type and severity of lung disease, and the risks associated with any surgery-related worsening of the pulmonary condition. In the special case of pulmonary resection, interpretation of the preoperative lung-function tests should lead to a knowledge of whether the patient will survive the loss of lung tissue and if so what the quality of life will be.

In patients scheduled to undergo surgery, lung-function tests are performed to detect those whose lung function is so poor that the risks of general anaesthesia are unacceptable, those whose lung function is moderately impaired such that any further deterioration due to postoperative pulmonary complications will put them immediately at risk, and those with pulmonary disease who may benefit from energetic implementation of respiratory therapy, such as effective bronchodilatation, physiotherapy, and prophylactic steroid therapy, before surgery.

With modern anaesthetic techniques there are virtually no patients who cannot be adequately ventilated and oxygenated during anaesthesia. The problems arise in the postoperative period, when hypoventilation is common and other pulmonary complications such as sputum retention, chest infection, and pulmonary emboli may supervene, or cardiac output may be reduced. The particular vulnerability of patients with chest diseases to these complications arises from their mechanical inability to increase ventilation, from the marked decrease in oxygen content of the blood caused when small decreases of arterial Pao_2 occur in patients who are already hypoxaemic, and in some diseases from reduced sensitivity to rising $Paco_2$, which normally causes ventilation to increase.

The history is important in pinpointing those patients who need to undergo lung-function tests before surgery: the most important warning symptom is shortness of breath. Past history of lung disease may be relevant, including recurrent chest infections, which may indicate those likely to suffer from sputum retention and postoperative chest infection. Other indications for preoperative lung-function testing are smoking, obesity, deformity of the thoracic cage, neuromuscular disease, and proposed thoracic or upper abdominal surgery.

Investigations of lung function

Investigations of lung function can be divided into tests that assess:

1. lung mechanics,
2. gas transfer,
3. control of breathing.

Lung mechanics

The best test for assessing lung mechanics is expiratory spirometry, which, when correctly performed, gives very reproducible results. The equipment for this test is inexpensive and portable. The main disadvantage is that it depends on the co-operation and maximum voluntary effort of the patient, and these are assumed in the interpretation of the results. The elements required for the successful performance of the manoeuvre are maximum inspiration and complete and maximally forced expiration. Several attempts at this manoeuvre may be required, with repeated instructions, encouragement, and demonstrations of the sort of breath needed, before success is achieved. A normal expiratory spirogram is shown in Fig. 1(a). Predicted normal values for forced expired volume in the first second (**FEV1**) and for forced vital capacity (**FVC**) (the total volume expired) depend on age, sex, height, and race, while the ratio FEV_1:FVC is related to age. The measured values should be related to those predicted for a particular patient (Table 1). There is a wide range of normal values, the SD for FEV_1 and FVC being 0.5 l. An initial test may miss a diagnosis of mild disease since the participant may have to lose up to 2 l from their expired volume before falling out of the range into which 95–per cent of normal individuals fit. As there is a large reserve of respiratory function, failure to detect mild disease is not a serious problem in surgical patients.

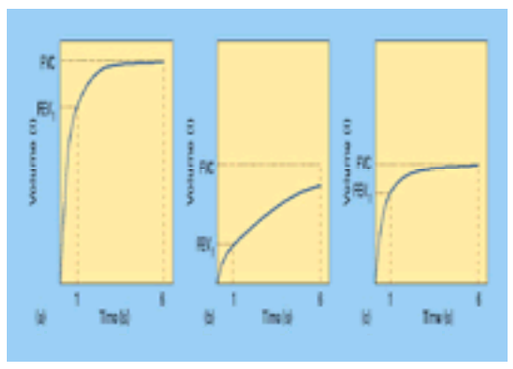


Fig. 1. Typical spirograms demonstrating (a) normal, (b) obstructive, (c) restrictive patterns of lung function.

	Obstructive FEV (percentage predicted)	Restrictive FVC (percentage predicted)
Mild	50–80	50–80
Moderate	35–50	35–50
Severe	<35	<35

FEV$_1$, forced expired volume in the first second; FVC, forced vital capacity.

Table 1 Rough guide to the severity of disturbance in lung mechanics

The results of this test allow lung function to be classed as normal, obstructive disease, or restrictive disease. If the volumes are not normal, distinction between obstructive and restrictive disease can be made from the FEV_1:FVC ratio: if this ratio is reduced there is an obstructive problem, (Fig. 1(b)). Predicted FEV_1:FVC ratio is 83–per cent at age 20 to 25 years, and 67 per cent at age 70 years.

Diseases associated with airflow obstruction in the intrathoracic airways include smoking-induced chronic airflow obstruction (emphysema, chronic bronchitis), asthma, bronchiectasis, and, in infants, bronchiolitis. Central intrathoracic airways are narrowed in carcinoma of the trachea, tracheomalacia, extrinsic compression, carcinoma

of the larynx, vocal-cord paralysis, epiglottitis, croup, large tonsils, and when there is an inhaled foreign body. In obstructive lung disease, maximal expiratory flow rates are reduced, and these are reflected in a reduced FEV_1:FVC ratio. It is useful to obtain a measurement of slow vital capacity (**SVC**), that is maximum inspiration to total lung capacity followed by complete gentle expiration. In patients with obstructive disease this SVC may be considerably greater than the fast (maximally forced) expired volume (FVC). If this is so the ratio FEV_1:SVC will be lower than FEV_1:FVC, emphasizing that the problem is predominantly one of airflow obstruction.

In restrictive lung disease the volumes are reduced but the FEV_1:FVC is ratio is normal, indicating that the airways do not offer abnormal resistance to flow (Fig. 1(c)). The element of the test that is poorly performed is the maximal inspiration: patients with restrictive disease are limited as to the size of the breath they can take in. Causes of inspiratory restriction include weakness of inspiratory muscles, a rigid chest wall (ankylosing spondylitis, scoliosis), pleural disease, stiff lungs (alveolitis, fibrosis, congestion), and exclusion of part of the lung from the manoeuvre(collapse, consolidation, previous resection).

These tests of lung mechanics indirectly evaluate dyspnoea and work of breathing, and give a measure of respiratory reserve, limitations of which may become important postoperatively when additional lung pathology may be present. Assessment of lung mechanics is often not possible before emergency surgery, in seriously ill patients, or in the early postoperative period, because of the maximal breaths required. More sophisticated lung-function tests should be performed in the lung-function laboratory in patients whose results on spirometry indicate severe disease and who are well enough to attend. Hyperinflation, the hallmark of airflow obstruction, can be assessed by the measurement of the static lung volumes, total lung capacity, functional residual capacity, and residual volume, by helium dilution or whole-body plethysmography (Fig. 2). Direct measurements of airway resistance can be made: both the degree of hyperinflation and the increase in airway resistance parallel the decrease in maximum expiratory flow rates in airflow obstruction, and so can be roughly predicted from the simpler maximum-effort tests. Hyperinflation, i.e., an increase in residual volume, has been found to be the best predictor of severe postoperative respiratory complications, together with the current preoperative hypersecretion of mucus and, to a lesser extent, the reduction in FEV_1 and carbon monoxide transfer factor per cent predicted. Peak flow (maximum expiratory flow) will be reduced in both obstructive and restrictive disease and this does not, therefore, distinguish between these two diagnoses.

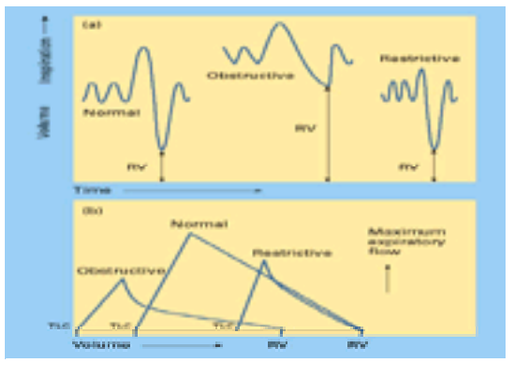


Fig. 2. Schematic representation of static lung volumes with normal, obstructive, and restrictive patterns as shown by (a) full inspiratory and expiratory spirograms and (b) curves for maximally forced expiratory-flow volume.

Gas transfer

Measurements of gas transfer are, on the whole, tests of ventilation perfusion (**V/Q**) matching. V/Q mismatching tends to be severe in patients with significant mechanical problems and major reductions of spirometric lung volumes. Of the tests available to assess gas transfer from alveolus to pulmonary capillary and vice versa, the best is measurement of arterial blood gases. Three items of information are obtained: arterial partial pressure of oxygen (Pao_2), arterial partial pressure of carbon dioxide ($Paco_2$), and the pH. These can be analysed both separately and together to build up a picture of lung function.

If, in a patient breathing air at sea level, the Pao_2 is below the value predicted from the patient's age, hypoxaemia is present:

Predicted Pao_2 (kPa) = 0.133(104 - 0.24 age), SD 1.05

Predicted Pao_2 (mmHg) = 104 - 0.24 age, SD 7.9

For example, predicted Pao_2 is 13.2 kPa (99 mmHg) at age 20 and 11 years, and 11.6 kPa (87 mmHg) at 70 years. A diagnosis of respiratory failure is made when the Pao_2 is 8 kPa (60 mmHg) or below. Respiratory failure is divided into classes. In type 1, $Paco_2$ is normal or reduced (normal range 4.6–6.0 kPa, 35–45 mmHg). This is typically seen in shock, pulmonary embolism, and pneumonia in previously normal lungs, and in asthma, emphysema (‘pink puffer'), and pulmonary fibrosis. In type 2, $Paco_2$ is increased, the most common cause being chronic airflow obstruction. Hypercapnia is occasionally due to muscle weakness or disease of the chest wall, and rarely to failure of ventilatory drive.

Calculation of the alveolar (A) to arterial (*a*) difference for Pao_2, $P(A - a)o_2$, enables some quantification of the disturbance of alveolar–arterial gas transfer. In the ideal lung, $P(A – a)o_2$ will be zero, but in practice in normal lungs breathing room air it should be less than 1 kPa (7.5 mmHg) in young individuals and less than 2.7 kPa (20 mmHg) at age 60–years. In abnormal lungs with mismatching of ventilation and perfusion, $P(A – a)o_2$ will be greater than the above. Assumed alveolar oxygen partial pressure (PAo_2) is calculated as follows:

$PAo_2 = PIo_2 - Paco_2/R$ (R = respiratory exchange ratio, assumed to be 0.8)

PIo_2 = inspired oxygen concentration/100 × (barometric pressure - water vapour pressure)

Therefore when breathing air

$P(A – a)o_2 = (20 - Paco_2/0.8) - Pao_2$ (kPa)

$P(A – a)o_2 = (150 - Paco_2/0.8) - Pao_2$ (mmHg)

When added oxygen is being inspired, even in normal lungs, $P(A – a)o_2$ widens, so comparisons between sequential measurements become difficult if different concentrations of oxygen are being given. Usually hypoxaemia is associated with an increased $P(A – a)o_2$, but if there is global underventilation, for example due to oversedation, $Paco_2$ and $PAco_2$ may rise sufficiently such that PAo_2, and therefore Pao_2 falls. In this condition the calculated $P(A – a)o_2$ is normal, despite the presence of hypoxaemia, indicating that the problem is one of underventilation of normal lungs. More usually, in the type-2 respiratory failure of chronic airflow obstruction, one finds hypoxaemia, hypercapnia, and an increased $P(A – a)o_2$, indicative of the V/Q mismatching seen in abnormal lungs.

A reduced $Paco_2$ indicates hyperventilation; an increased $Paco_2$ indicates hypoventilation. When V/Q mismatching is present there is a disparity between the ventilation required to achieve normal Pao_2 and $Paco_2$. Because of the linear CO_2 dissociation curve, overventilation of one part of the lung (areas with high V/Q) can compensate for underventilation of another part (low V/Q regions). For oxygen, which has a sigmoid dissociation curve, this does not occur (Fig. 3). Breathing hard and increasing the Pao_2 in blood coming from the more normal areas of lung do not increase significantly the oxygen content (saturation) of this blood and therefore of the mixed pulmonary venous blood. This difference between the behaviour of carbon dioxide and oxygen results in type-1 respiratory failure.

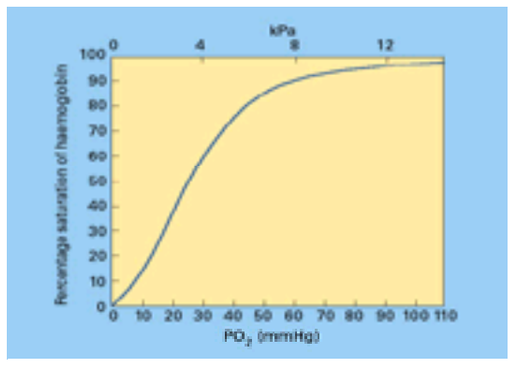


Fig. 3. Haemoglobin oxygen-dissociation curve.

The pH indicates whether an acidaemia or alkalaemia is present, while the *P*aco$_2$ indicates the respiratory contribution to the disturbance of acid–base state. The relation between pH and *P*aco$_2$ indicates whether compensation has occurred (Table 2). For example, if the *P*aco$_2$ is low the patient is hyperventilating and there is a respiratory alkalosis. The pH will indicate whether this is acute and uncompensated (pH > 7.4) or more chronic and compensated (pH near 7.4). It takes several hours for compensation of plasma bicarbonate by renal adjustment to begin, and days to complete. A useful working rule is that compensation does not bring the pH quite back to normal. A slightly acid pH with an elevated *P*aco$_2$ suggests a primary respiratory acidosis with metabolic compensation. A slightly alkaline pH with an elevated *P*aco$_2$ suggests a primary metabolic alkalosis with respiratory compensation. Respiratory or metabolic compensation for alkalosis is less complete and less predictable than that for acidosis. Respiratory compensation for a primary metabolic disorder occurs rather quickly, so the presence of compensation is not useful in determining how long the acid–base disorder has existed. Acute change in blood gases and pH superimposed on chronic changes are more difficult to interpret unless the results of serial measurements of blood gases are available.

pH	$PaCO_2$	
	kPa	*mmHg*
7.6	2.5	20
7.5	4.0	30
7.4	5.5	40
7.3	8.0	60
7.2	10.5	80

Table 2 Approximate predictions of pH in uncompensated respiratory acidosis and alkalosis

Other tests of gas transfer include measurement of oxygen saturation by ear or finger pulse oximetry. This is a valuable non-invasive test for screening or for obtaining serial or continuous measurements. From the shape of the oxygen dissociation curve (Fig. 3) it can be seen, however, that this test will not detect early deterioration in gas exchange well. The *P*ao$_2$ may have fallen considerably before the oxygen saturation falls measurably, and saturation may then fall precipitously with only a further small decrease in *P*ao$_2$. Transcutaneous O_2 and CO_2 electrodes are not yet sufficiently robust in performance for routine use outside the intensive care unit. In the patient able to attend the lung-function laboratory before surgery, measurement of transfer of carbon monoxide is a useful non-invasive screening test, particularly in the assessment of patients with pulmonary fibrosis, in whom CO transfer may be reduced when the lung volumes are still normal.

Acute changes in both lung mechanics and gas transfer are less well tolerated by the patient than are chronic, slowly developing changes.

Control of breathing

Minute ventilation is normally controlled so that *P*ao$_2$ and *P*aco$_2$ remain constant, despite varying metabolic demand for uptake of oxygen and excretion of carbon dioxide. Ventilation is driven by the rhythmic output from the brainstem respiratory centre, the output of which is modulated by a falling *P*ao$_2$ and increasing *P*aco$_2$, by the acidity of the cerebrospinal fluid, and by reflexes arising from airways, alveoli, and chest wall. Formal assessment of the control of breathing is not usually made before surgery, though if the preoperative *P*aco$_2$ is high it indicates either that control mechanisms have failed to increase ventilatory drive or that the mechanical problems are so great that ventilation cannot be increased despite increased drive. Opiates may depress respiratory drive postoperatively, but the benefits of early mobilization, coughing, and chest physiotherapy that are allowed by adequate pain relief are thought to outweigh the disadvantage of this depressive effect. In patients with chronic airflow obstruction, the hypoxic drive to ventilation is important and administration of oxygen may depress ventilation; oxygen must initially be given at low concentrations and its effect on *P*aco$_2$ monitored. It can be seen from Fig. 3 that in any patient with *P*ao$_2$ near the ‘knee' of this curve, a small decrease in *P*ao$_2$ postoperatively due to hypoventilation may cause a sharp drop in oxygen saturation and oxygen-carrying capacity.

Any patient with a hint of breathing difficulties should undergo expiratory spirometry and pulse oximetry before major surgery to assess risk. In addition, baseline measurements made before surgery are helpful for comparison if pulmonary complications do occur postoperatively. A careful history of dyspnoea, persistent production of mucus, and recurrent chest infections will usually indicate those in whom preoperative lung-function assessment is mandatory.

Effects of pulmonary resection

In the special case of pulmonary resection, the surgical intervention necessarily makes the lung-function worse both in the short and the long term. Whether an already respiratorily disabled patient will tolerate this must be assessed. The loss of volume and exercise capacity is less than would be expected from a simple calculation based on the number of segments of lung removed.

Lobectomy leads to an early loss of pulmonary function with later recovery and little permanent defect, about 10 per cent loss in pulmonary volume (vital capacity) and no decrease in exercise capacity. Pneumonectomy leads to an early permanent function loss of about 33 per cent in pulmonary volumes and 20 per cent in exercise capacity. These are average figures. The worst case of loss of 50 per cent of lung volume after pneumonectomy is more likely to occur with a peripheral tumour than with a central one. In patients undergoing lung resection for cancer there is a high incidence of chronic obstructive lung disease, related to smoking, which may be unevenly distributed throughout both lungs. If the tumour is central, there may be collapse/consolidation distal to it. In this case, some of the preoperative impairment will be related to the part of the lung to be removed and will not reflect the function of the lung that remains after surgery. The modern approach to preoperative assessment is to do lung-function tests at rest. If these indicate more than mild disease, split-function studies and exercise tests are undertaken. In split-function testing, postoperative predicted (**ppo**) FEV_1, and transfer factor for carbon monoxide and oxygen uptake at maximum exercise are estimated from a quantitative pulmonary-perfusion radioinuclide scan according to the following formula:

ppoFEV_1 = **preop**(preoperative)FEV_1 × preop perfusion that goes to lung which will remain after surgery ÷ total preop perfusion.

A common recommendation is that surgical resection should not be performed if predicted ppoFEV_1 is less than 0.8 l, but there is no consensus on the minimum ppoFEV_1 for safe resection. Because of the increasing number of women with lung cancer, on theoretical grounds it will be better to express predicted postoperative lung-function measures as per cent predicted normal for each particular patient as this will give a better indication of the severity of disease. For example:

0.8 l = 41% predicted for a woman age 65–years, height 5ft 3 inches; (moderate airflow obstruction)

= 27% predicted for a man age 60 years, height 5ft 9 inches;(severe airflow obstruction).

Because lung volumes are reduced more than exercise capacity, exercise testing is helpful when predicted postoperative lung-function values are marginal.

It is generally agreed that if ppoFEV1 is greater than 40 per cent predicted, pneumonectomy can be undertaken with low morbidity and low mortality risk. Below this, other factors such as age, comorbidity, exercise capacity, and motivation need to be taken into account to an even greater extent in making recommendations about resection.

Further reading

Barisone G, Rovida S, Gazzaniga GM, Fontana L. Upper abdominal surgery: does a lung function test exist to predict early severe postoperative respiratory complications? *European Respiratory Journal* 1997; **10**: 1301–8.

Bolliger CT, Perruchoud AP. Functional evaluation of the lung resection candidate. *European Respiratory Journal* 1998; **11**: 198–212. [Good review of assessment of factors predicting pulmonary and cardiac complications in lung-resection candidates.]

Busch E, Verazin G, Anthowiak JG, Driscoll D, Takita H. Pulmonary complications in patients undergoing thoracotomy for lung carcinoma. *Chest* 1994; **105**: 760–6.

Corris PA, Ellis DA, Hawkins T, Gibson GJ. Use of radionuclide scanning in the preoperative estimation of pulmonary function after pneumonectomy. *Thorax* 1987; **42**: 285–91.

Gibson GJ. *Clinical tests of lung function*, 2nd edn. Chapman and Hall Medical, London, 1996. [Excellent, well-referenced text describing pulmonary physiology and disturbances in disease.]

Ginsberg RJ, Hill LD, Eagan RT, Thomas P, Mountain CF, Deslauriers J. Modern thirty day operative mortality for surgical resections in lung cancer. *Journal of Thoracic and Cardiovascular Surgery* 1983; **86**: 654–8.

Larsen KR, Svendsen UG, Milman N, Brenoe J, Petersen BN. Exercise testing in the preoperative evaluation of patients with bronchogenic carcinoma. *European Respiratory Journal* 1997; 10: 1559–65.

Tisi A. Preoperative evaluation of pulmonary function. Validity, indications and benefits. *American Review of Respiratory Disease* 1979; **119**: 293–310. [Classic, state-of-the-art paper on this topic.]

Zeiher BG, Gross TJ, Kern JA, Lanza LA, Petersen MW. Predicting postoperative pulmonary function in patients undergoing lung resection. *Chest* 1995; **108**: 69–72.

11.1 Guidelines and scoring systems

Christopher S. Garrard

Introduction
Guidelines for admission
Reversibility of illness
Comorbidity
Advance statements or living wills
Mechanical ventilation
Dependency
High-dependency care
Outcome prediction in the intensive care unit
APACHE
SAPS, SAPS II
MPM
Therapeutic intervention scoring system
Organ-system failure scoring systems
Choice of severity scoring systems
Further reading

Introduction

Intensive care is needed for patients who need a higher level of observation and treatment than can be provided in the general ward or high-dependency unit. It is usually reserved for patients with organ failure arising as a complication of an acute illness or injury, or is a predictable outcome of a planned treatment strategy. For an intensive care unit to achieve these aims and function efficiently there should be multidisciplinary approaches to care, integrating medical and nursing staff. In addition there should be:

- a clinical director
- experienced, senior medical staff supporting resident, trainee medical staff, providing 24-h dedicated cover
- a nurse:patient ratio of 1:1.
- comprehensive physiological monitoring and resuscitation equipment
- the ability to support a wide range of organ-system failures including, ventilatory, circulatory, renal, gastrointestinal, haematological, and neurological failure
- a large number and mix of patients to maintain skills and expertise of medical and nursing staff
- technical, clerical, and management support
- active educational, clinical training, clinical audit, and research programmes.

Guidelines for admission

Referrals to the intensive care unit come from emergency departments, general wards, operating rooms, and other intensive care units. These referrals can be either elective (e.g. following planned surgery) or emergency and unplanned. Usually, referral will come from a member of the medical staff, although experienced nursing staff may be as skilled at identifying patients who need intensive care. Thus, a flexible attitude that allows junior medical and nursing staff to request an intensive-care consultation is to be encouraged. This does not alter the physician's role but should reduce the delays in instituting intensive care. Ideally there should be written admission and discharge guidelines that must, by necessity, leave scope for interpretation and the exercise of clinical judgement. The aim of admission guidelines is to exclude patients who are well enough to be adequately cared for on general wards, and those who are so ill that their chances of survival are remote. The middle ground includes patients who need high levels of medical and nursing supervision, and of intervention, and who have largely reversible illness.

Reversibility of illness

Intensive care should be provided for patients who are likely be a benefit from the level of observation and therapy that can be provided in an intensive care unit (Fig. 1). The appropriateness of admission to such a unit largely depends upon the reversibility of the patient's underlying condition. This is not always easy to assess and often the only appropriate course is to admit the patient to the intensive care unit and to evaluate their response to treatment. It should be remembered that, although patients can be supported for prolonged periods, intensive care will be futile unless there is a realistic chance of the their ultimate recovery and return to an acceptable quality of life. If it is apparent that further care is futile, a discussion between medical and nursing staff, the patient (if possible), or the patient's family should assess the appropriateness of continued intervention.

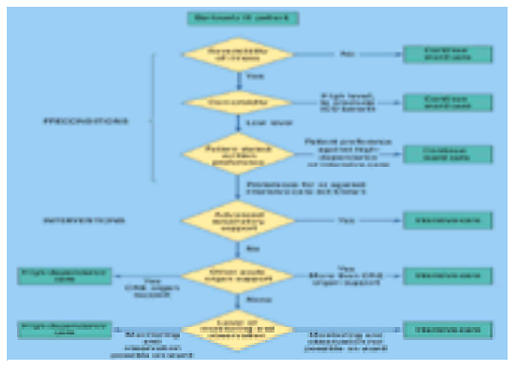


Fig. 1. Guidelines on admission to intensive care outlining the decision process for the appropriateness of intensive and high-dependency care. Three preconditions should be addressed before assessing the need for organ support. Consideration of the patient's preference for or against intensive care (advance directive) can if necessary be delayed until after the need for intensive care has been established.

Comorbidity

It is important to take into account any comorbidity, the level of physiological reserve, and the patient's biological age, which, individually or together, may make recovery highly improbable (Fig. 1). Elderly patients have a higher incidence of cardiovascular and respiratory disease; non-specific immune deficiency can be demonstrated in them, due to poor nutrition, hormonal changes, or other mechanisms that reduce resistance to hospital-acquired infection. The effects of a long, debilitating, critical illness in the elderly person often mean they may have little or no chance of returning to independent life. Ideally, biological rather than chronological age should be considered and, consequently, relatively few intensive care units have an upper age limit for admission. The nature and severity of specific premorbid medical conditions also have a large effect on management and outcome in the intensive care unit. For example, the patient with chronic obstructive pulmonary disease, severe exertional dyspnoea, and grossly impaired lung function is less likely to tolerate long-term mechanical ventilation without barotrauma and infectious complications. Weaning may be protracted or even fail, with the possibility of chronic long-term ventilation support. The presence of advanced, solid, or haematological malignancy may lead the patient, the patient's family, and the doctors to consider intensive care to be inappropriate.

Advance statements or living wills

A patient's stated or written statement against intensive care must always be considered by the referring and unit staff. Where there is uncertainty about the patient's wishes, active and appropriate intensive care should be provided. Later, if their wishes can be established, an appropriate care plan that complies with them can be

implemented.

Mechanical ventilation

The decision whether or not advanced respiratory support is needed rests with the intensive-care doctor and should be based on the clinical condition of the patient. The need for mechanical ventilation is probably the least contentious indication for admission to intensive care. Except under exceptional circumstances or for short intervals, patients should not be ventilated other than in an intensive care unit.

Dependency

The level of dependency of the patient is a crucial factor in determining the appropriateness of intensive care. Patients with life-threatening conditions and who require high levels of nursing dependency should be admitted to intensive care. Examples of life-threatening conditions include acute asthma, cardiogenic shock, and severe diabetic ketoacidosis. Dependency is commonly measured using the Therapeutic intervention scoring system (TISS), which grades the severity of illness by quantifying the therapeutic interventions applied to the patient.

High-dependency care

The appropriateness of admission to high-dependency care will depend partly on what other levels of care, including intensive, are available (Fig. 1). High-dependency care can facilitate appropriate earlier discharge from intensive care but can also be used to provide a level of care higher than that available on a general ward (i.e., a step-up facility). One recent study in the United States showed that establishing a high-dependency unit reduced ward mortality rates by 25 per cent and cardiac arrests by 39 per cent. In hospitals without a high-dependency unit, the intensive care unit may come under pressure to admit patients who could be cared for appropriately in a high-dependency unit. Where there is a high-dependency unit, the policies of the two units should be complementary.

Patients requiring support of a single organ system other than ventilatory support should generally receive high-dependency care. Thus, patients in need of circulatory, renal, or hepatic support can often be managed successfully in a high-dependency unit. When support for two or more organ systems (other than respiratory) is required, intensive care should be provided (Fig. 1).

Patients requiring renal support alone can usually be treated in a renal unit. A key distinguishing feature between intensive, high-dependency, and ward care is the level of nursing required. Usually patient dependency and nursing intensity are directly related.

Outcome prediction in the intensive care unit

The need to predict outcome for critically ill patients has stemmed from society's need to rationalize the allocation of limited medical resources. Intensive care can then be provided for the greatest number of patients who have a reasonable chance of meaningful recovery. Identifying patients who will benefit most from intensive care (see above) and the quality of care provided have become important issues for both intensive-care clinicians and health-care economists. In general, the assessment of potential outcome from critical illness depends upon the severity and extent of the current presenting illness.

Several scoring systems that describe the relation between disease severity and outcome in a group of patients in the intensive care unit have been developed (Table 1). Scoring systems are already commonly used in many areas of medicine such as oncology and surgical specialities. However, scoring illness severity for the wide range of medical and surgical illness in the critically sick patient poses a greater problem. A number of risk-adjustment methods have been applied to intensive care, the best known being **APACHE** (Acute physiology and chronic health evaluation), **SAPS** (Simplified acute physiology score), and **MPM** (Mortality prediction models). These can be used to estimate the expected mortality for a group of patients admitted to intensive care. The ratio of the observed to the expected mortality is termed the standardized mortality ratio. The ratio can be used to compare intensive care units, although there are many confounding factors. The use of a risk-adjustment method to determine outcome in an individual patient requires extremely accurate prediction. To date, such accuracy has yet to be achieved.

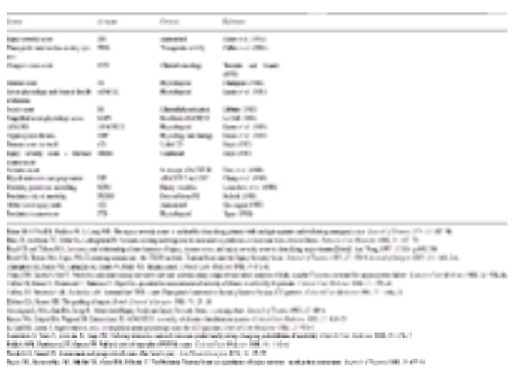


Table 1 Examples of severity and disability classification systems

APACHE

The APACHE severity of disease classification was devised to stratify prognosis in groups of critically ill patients, and to determine the success of treatment. The original APACHE score was based on 34 physiological values (acute physiology score, **APS**), and a subjective assessment of the severity of chronic, intercurrent disease. APACHE II was later developed as a simplified, clinically useful system, using 12 physiological variables (APS) combined with an evaluation of chronic health. It is recommended that the worst physiological values during the first 24 h following admission to the intensive care unit be used.

The 12 variables comprising the APS are:

1. temperature (°C);
2. mean arterial pressure (mmHg);
3. heart rate (beats/min);
4. respiratory rate (breaths/min);
5. alveolar–arterial oxygen gradient (AaDO_2) if fractional inspired oxygen (FIO_2) is 0.5 or greater, or use PaO_2 if FIO_2 is less than 0.5;
6. arterial pH;
7. serum sodium (mmol/l);
8. serum potassium (mmol/l);
9. serum creatinine (mg/l00 ml);
10. haematocrit (%);
11. leucocyte count (cells/mm^3);
12. Glasgow coma score.

Depending on the degree of derangement, a weighted score (0–4 points) is assigned to each variable (except the Glasgow coma score) and summed to produce the APS. The score assigned to the Glasgow coma score is 15 points minus the estimated Glasgow coma score.

An element of clinical judgement has to be applied when estimating the Glasgow coma score. The patient considered to be neurologically unimpaired should be given a Glasgow coma score of 15 even if fully sedated and unresponsive. This provides a much more accurate estimate of the true neurological status. The subjective nature of the Glasgow coma score unfortunately makes it much more liable to bias and therefore a source of significant error in APACHE scoring.

Points are also assigned for greater age, emergency postoperative or non-operative admission, and if there is a prior history of organ-system insufficiency or failure. Double weighting is given to abnormal serum creatinine in the presence of acute renal failure, due to the increased mortality associated with acute renal failure in the critically ill. The maximum possible APACHE II score is 71, although scores rarely ever approach this except as a premorbid condition. Increasing scores correlate with

higher hospital mortality, at each 5-point increment, across a wide range of disease conditions. However, in an individual patient the APACHE II score reflects only the severity of physiological derangement for that diagnostic category at a single point in time.

APACHE II is potentially sensitive to treatment effects preceding admission to the intensive care unit (lead-time bias). Thus, patients who are resuscitated and treated (for hours or days) in one hospital before being transferred to the intensive care unit in another will present to the new hospital with an unrepresentatively low APACHE II in relation to their risk of death. Similarly, a 4- to 6-h delay in admitting a patient to an intensive care unit will be used to improve or normalize their physiological status and falsely lower the admission APACHE score. A trend analysis may therefore be more appropriate, utilizing sequential APACHE scores at fixed intervals (e.g. daily), and noting the rate of change relative to the last score. The newer APACHE III system is allocated a coefficient to adjust for the location of the patient before admission to the intensive care unit. The effects of lead-time bias in North American intensive care units have been considered to be small, although preliminary experience in hospitals in the United Kingdom suggests a much greater effect. The greatest sources of error in APACHE II mortality prediction are associated with the Glasgow coma score and the diagnostic categorization. Inappropriate classification of a trauma patient who has suffered a perforated small bowel as primarily a ‘gastrointestinal perforation' could almost double the predicted mortality. The effect of incorrect classification is especially great if the APACHE II score is low.

Different types of disease are associated with their own particular risk of death or survival. For example, patients with conditions such as diabetic ketoacidosis are expected to survive despite extreme physiological derangement on admission. Similarly, the postoperative outcome in patients with coronary arterial bypass is generally better than the APACHE II prediction. In contrast, equivalent APACHE II scores in patients with septic shock are associated with a much lower survival. By adding the APACHE II score to a logical regression equation incorporating diagnostic categories a prediction of hospital death can be obtained. The coefficients for each diagnostic group were originally derived from North American intensive care units and subsequent data-sets collected in the United Kingdom have yielded different values.

APACHE III

APACHE III is an enhancement of APACHE II that has produced an improved model for the prediction of hospital death. APACHE III was developed on as data-set of 8720 patients in 40 hospitals in the United States and validated on a similar number. Additional indices include the serum albumin, bilirubin, blood glucose, and urine out-put. Some chronic diseases, such as respiratory, have been removed and greater recognition of immune compromise made. The model also includes a mortality prediction for the first 7 days based on daily changes of physiology. Significantly, the 79 principal-diagnosis coefficients are not in the public domain and the use of APACHE III requires payment of a licensing fee. This has lead to some resistance to APACHE III being adopted as a world-wide standard.

Riyadh intensive care programme

This provides an outcome prediction based on a trend analysis of APACHE II in association with organ failure assessment. By following APACHE II over a period of time, improved prediction accuracy is claimed. However, prediction for individual patients falls short of infallibility.

Sickness scoring

The sickness score is a variant of the APACHE II score developed in an attempt to enhance its prognostic accuracy. The modifications to the APACHE score are as follows:

1. units are converted to the SI system;
2. haemoglobin concentration is used rather than haematocrit;
3. oxygenation is assessed using the FIO_2:PaO_2 ratio;
4. the ‘chronic disease' category is redefined to include conditions associated with loss of independent self-care;
5. clinical judgement is used to estimate the Glasgow coma score;
6. haemodynamic instability is assessed to reflect overall abnormalities rather than transient, perhaps drug-induced changes;
7. daily scores are charted, and the trend determined to assess response to treatment.

Using these guidelines, periodic neurological and cardiovascular abnormalities due to sedative drugs could be cancelled out.

SAPS, SAPS II

The original 34 APACHE variables were reduced by Legall to 14 easily definable variables to form the SAPS, similar to the current APACHE II score. Minor modifications to account for the risks of mechanical ventilation were made. The conclusions were similar to results obtained from the APACHE II and Sickness scoring studies.

MPM

Lemeshow devised MPMs based on multivariate statistical analysis of 19 124 adult, general intensive care patients in the United States, Canada, and Europe. The ability of these models to predict risk of death is similar to the scoring systems cited above. The predictive models can be based on admission variables (MPM0) or values during the first 24 h ((MPM24). The system has the advantage of not requiring an admission diagnosis. A series of Yes/No questions is answered, and these are weighted according to their individual contribution to mortality. The predictive ability of sequential MPMs is approximately 74 to 80 per cent. Unlike APACHE II and III, MPM is therapy independent.

Therapeutic intervention scoring system

The Therapeutic intervention scoring system (**TISS**) assigns a score of 1 to 4 points for each of the procedures performed on patients in the intensive care unit to provide an indicator of the severity of illness and nursing dependency. It was suggested that a competent intensive-care nurse could handle 40 to 50 TISS points per day. The num-ber of therapeutic interventions on a given patient would be dependent on the type of care given, and the need for procedures such as invasive monitoring and mechanical ventilation. A high TISS, against a background of continued active treatment, would suggest that discharge from the intensive care unit should not be considered. Whilst TISS has been shown to be valuable from an administrative standpoint, its inability to predict death in an individual patient (like other scoring systems) militates against its use as a outcome indicator. However, TISS has been calibrated against the daily cost of delivering intensive care such that each TISS point can be apportioned a fixed cost. This value can then be used as a crude but simple estimate of subsequent care. Recent developments of TISS have reduced the number of variables to 28.

Organ-system failure scoring systems

Several models have been proposed to quantify the number and the severity of organ systems that fail. Most adopt simple, single measures of organ-system function such as the serum creatinine for renal function and the serum bilirubin for liver function. Popular models include the Marshall score, SOFA score (Sepsis-associated organ failure assessment), and Logistic organ failure score. Most use similar indices, with some minor differences in weighting coefficients. These scoring systems can provide prediction of group mortality but their chief utility may be in quantifying the severity of damage to each organ system.

Choice of severity scoring systems

None of these scoring systems was intended to predict outcome in individual patients. Instead they are used almost exclusively as quality-assessment tools for clinical care or in research as a severity measure. Although APACHE III achieves the best group predictions, it does require the largest data-set and diagnostic-group coefficients are not in the public domain. With the advent of simple computer database tools it is possible to collect all the required variables and derive several severity scores. APACHE II and MPM0 provide the simplest scoring system for general audit purposes in the intensive care unit.

Further reading

Baker SP, O'Neill B, Haddon W, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma* 1974; **14**: 187–96.

Beck DH, Taylor BL, Millar B, Smith GB. Prediction of outcome from intensive care: a prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. *Critical Care Medicine* 1997; **25**: 9–15.

Bion JF, Aitchison TC, Edlin SA, Ledingham IMcA. Sickness scoring and response to treatment as predictors of outcome from critical illness. *Intensive Care Medicine* 1988; **14**: 167–71.

Boyd O, Grounds RM. Physiological scoring systems and audit. *Lancet* 1993; **341**: 1573–4.

Chang RWS, Jacobs S, Lee B, Pace N. Predicting deaths among intensive care patients. *Critical Care Medicine* 1988; **16**: 34–42.

Cullen DJ, Civetta JM, Briggs BA, Ferrara LC. Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Critical Care Medicine* 1974; **2**: 57–60.

Keene AR, Cullen DJ. Therapeutic intervention scoring system: update 1983. *Critical Care Medicine* 1983; **11**: 1–3.

Kilpatrick A, Ridley S, Plenderleith L. A changing role for intensive therapy: is there a case for high dependency care? *Anaesthesia* 1994; **49**: 666–70.

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Critical Care Medicine* 1985; **13**: 818–29.

Le Gall J-R, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European North American multicentre study. *Journal of the American Medical Association* 1993; **270**: 2957–63.

Lemeshow S, Teres D, Avrunin JS, Gage RW. Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Critical Care Medicine* 1988; **16**: 470–7.

Lemeshow S, Teres D, Avrunin JS, Pastides H. A comparison of methods to predict mortality of intensive care unit patients. *Critical Care Medicine* 1987; **15**: 715–22.

Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality probability model (MPM II) based on an International cohort of intensive care patients. *Journal of the American Medical Association* 1993; **270**: 2478–86.

Rogers J, Fuller HD. Use of daily Apache Physiology and Chronic Health Evaluation (APACHE) II scores to predict individual patient survival rate. *Critical Care Medicine* 1994; **22**: 1402–5.

Rowan KM *et al.* Intensive Care Society's APACHE II study in Britain and Ireland. II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *British Medical Journal* 1993; **307**: 977–81.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. *Lancet* 1974, **ii**: 81–3.

11.2 Cardiovascular aspects

Christopher S. Garrard and Nigel R. A. Clarke

Cardiopulmonary resuscitation
Conduct of resuscitation
Circulatory failure
Hypovolaemic
Cardiogenic
Distributive
Physiological measurement in circulatory failure
Flow-directed pulmonary arterial catheter
Lactic acidosis
Types
Alkali therapy of metabolic acidosis
Sodium bicarbonate
Tris-hydroxymethylaminomethane
Sodium carbonate
Sodium dichloroacetate
Further reading

Cardiopulmonary resuscitation

The critical care unit provides an ideal environment for successful cardiopulmonary resuscitation following cardiac arrest. The rate of survival is higher in that unit than in other areas of the hospital because the event is usually witnessed, the rhythm is usually ventricular fibrillation, and all the necessary personnel and equipment are immediately available. The outlook is the most poor when cardiac arrest complicates renal failure, respiratory failure, or other metabolic disorders, including sepsis. Prodromal events such as hypoxaemia, arrhythmias, and electrolyte disturbances should be detected before circulatory failure occurs, and prompt intervention may prevent cardiac arrest.

Whenever possible the cause of arrest should be ascertained. It may be rapidly apparent that the serum potassium had been rising during the proceeding hours or that there had been a prodromal arrhythmia. In the postoperative cardiac surgical patient, electromechanical dissociation due to haemopericardium must always be considered. Such information and the clinical background may influence significantly the approach to resuscitation. Because the patient in the critical care unit is invariably undergoing electrocardiographic monitoring, ventricular fibrillation is usually recognized and treated immediately. Even in the highly advantageous environment of the unit, however, a disciplined approach to basic cardiopulmonary resuscitation still needs to be followed.

Conduct of resuscitation

A resuscitation team leader should supervise the resuscitation. Ideally, this individual should have little active role in procedures but should direct the treatment priorities, co-ordinate the participants, and assess the effectiveness of resuscitation. One team member should maintain cardiac massage while another provides ventilatory support. Another member should establish medication access, preferably by a central venous route, if not previously established, while another records interventions and medications administered.

The ABC (**A**irway, **B**reathing, **C**irculation) order of assessment and resuscitation should be followed. In many cases the patient will already be intubated; if not, the airway must be secured (an oral airway will suffice) and ventilation provided initially by face-mask and Ambu bag (self-inflating ventilation bag). Two hands are required to produce an adequate seal with an anaesthetic face-mask, and another individual should provide ventilation. If the patient is already on a mechanical ventilator, ‘bagging' the patient with 100–per cent O_2 will permit better synchronization with cardiac massage: slow inflations of the lungs sufficient to cause visible expansion of the chest wall are required. A major arterial pulse, such as the carotid, should be checked, since a malfunctioning arterial line may not accurately reflect cardiac output.

Controversy still surrounds the use of the precordial (chest) thump. When a defibrillator is immediately to hand and when the patient is already monitored, it probably has little part to play in the resuscitation procedure.

Circulation support

The ratio of cardiac compressions to each ventilated breath should be 15:2, with a compression rate of about 100/min. Whatever cardiac compression rates are achieved, they must not interfere with effective ventilation. Recent clinical trials have not confirmed the superiority of simultaneous compression–ventilation cardiopulmonary resuscitation, as was suggested by animal studies. It is important that no more than 10 s should be permitted for any manoeuvre that demands temporary cessation of chest compression.

Assessment of the efficacy of cardiopulmonary resuscitation is difficult. An intra-arterial pressure trace is an ideal indicator of cardiac output but is not always available. An end-tidal CO_2 greater than 2 kPa (15 mmHg) has recently been shown to indicate efficient resuscitation and favourable outcome, while one less than 1.5 kPa (10 mmHg) predicts a poor outcome. Administration of bicarbonate during resuscitation will tend to raise end-tidal CO_2, misleading the clinician into believing resuscitation is more effective than it really is.

All medications should be given by a central venous cannula or by the transtracheal route, which is particularly suitable for the administration of adrenaline (epinephrine) (double dose).

Definitive measures

The measures that need to be applied depend upon the underlying cause of cardiac arrest. The electrocardiogram must be interpreted in the clinical context to distinguish between VF, polymorphic VT, and pulseless monomorphic VT. The most common arrhythmia is ventricular fibrillation, which should be treated with immediate defibrillation. If defibrillation at 200 J produces no response, it should be repeated. If there is still no response, defibrillation at 360 J should be performed. If this is also unsuccessful, defibrillation at 360 J should be repeated after intravenous administration of 1 mg adrenaline (epinephrine) and then lignocaine (lidocaine) (100 mg intravenously).

If this is still unsuccessful, alternative paddle positions (antero-posterior) or another antiarrhythmic drug should be tried. In refractory cases, attempts should be made at electrical pacing by either external or internal electrodes.

Cardiopulmonary resuscitation should be attempted for up to 2 min after each drug, and should not be interrupted for more than 10 s, except for defibrillation. Differentiating ‘coarse' from ‘fine' fibrillation has been thought to predict response to defibrillation. However, the standard approach outlined above should still be adhered to. Recent work in animal models shows that the frequency characteristics of the pattern of ventricular fibrillation are determined by the duration of fibrillation. Such a tool may, in the future, serve to predict outcome. Once an effective cardiac output is achieved, a continuous infusion of adrenaline (epinephrine) in doses of 1 to 10 mg/h can be given to support the myocardium and circulation. In the young patient with an otherwise good prognosis, supramaximal doses may be administered. Other inotropes and pressors can be used with equal effect. Arterial blood gases and electrolytes should be monitored, and a chest radiograph obtained.

Adrenaline (epinephrine) has largely replaced lignocaine (lidocaine) as the first pharmacological agent to be given to the patient in cardiac arrest. Adrenaline (epinephrine), and possibly noradrenaline (norepinephrine), may increase cerebral perfusion during basic life support, although they probably do not enhance the efficacy of defibrillation. Lignocaine (lidocaine) is effective in the treatment of ventricular tachycardia and in prophylaxis for ventricular fibrillation. However, evidence to support its use in cardiopulmonary resuscitation is lacking and it may actually render the heart more refractory to electrical defibrillation.

Giving sodium bicarbonate is no longer recommended, except where efficient ventilation can be maintained during prolonged resuscitation. After prolonged resuscitation it may be reasonable to consider giving 50 mmol of sodium bicarbonate (50 ml of 8.4 per cent) if there is persistent metabolic acidosis. Unfortunately, the buffering effect of bicarbonate results in the generation of carbon dioxide, which readily diffuses into the intracellular compartment and may therefore worsen intracellular acidosis. The best way of controlling the acidosis observed during cardiopulmonary resuscitation is to establish effective ventilation and ensure an

adequate circulation. Other alkalinizing agents, such as the combination of sodium carbonate and bicarbonate (Carbicarb®), although not widely available, may offer an alternative method of correcting severe acidosis.

There is no limit to the number of defibrillations that can be attempted, assuming that the diagnosis of the cardiac rhythm is correct. Changing both paddle position and the defibrillator itself should be considered, together with administration of other antiarrhythmic drugs such as bretylium.

Electromechanical dissociation

Electromechanical dissociation is a non-VF/pulseless VT rhythm in which there are QRS complexes without apparent ventricular contractions, as evidenced by a palpable pulse or arterial pressure waves. Caution should always be exercised when using an arterial line to obtain an index of cardiac contractility, since the device may give false information. If in doubt, the carotid or femoral pulse should be checked. Electromechanical dissociation is managed in the same way as any cardiac arrest except that intravenous adrenaline (epinephrine) (1 mg) should be given immediately.

It is essential to exclude and correct hypovolaemia, which may be disguised by the administration of a pressor or may form part of a sepsis syndrome. Pneumothorax/haemothorax should be considered in all victims of trauma, while cardiac tamponade may complicate postoperative cardiac cases. Pulmonary embolism must always be considered in the postoperative patient.

Treatable electromechanical dissociation must be managed aggressively. In many cases, cardiac arrest is unheralded but in retrospect it may be apparent that physiological changes were present, indicative of events such as cardiac tamponade or pneumothorax. Acute and severe blood loss will eventually result in electromechanical dissociation: volume resuscitation requires the insertion of several large-bore cannulas for the transfusion of blood and blood products. In patients likely to survive, emergency thoracotomy, internal cardiac massage, and cross-clamping of the descending aorta may all be required before an effective cardiac output can be achieved.

Any patient maintained on intermittent positive-pressure ventilation who manifests increasing airway pressures may be developing a pneumothorax. There may be insufficient time to obtain a chest radiograph, and tube thoracostomy (possibly bilateral) may have to be performed immediately. Tamponade may develop after cardiac surgery, and the clinician must be alert to the implications of a falling cardiac output and oliguria in the face of a rising central venous pressure. The decision to reopen a sternotomy wound is never made lightly, but prompt intervention may be the only way of saving the patient who is developing tamponade. Cardiac massage is here probably ineffectual at best and at worst may disrupt coronary venous grafts.

If hyperkalaemia or hypocalcaemia are suspected, or if calcium antagonists have recently been administered, calcium chloride (10 ml of 10 per cent) should be given.

Asystole

Asystole is another non-VF/pulseless VT rhythm. It has grave prognostic implications, since treatment is much less effective than for ventricular fibrillation. It is critical that errors in identifying asystole are excluded. Faulty equipment or very fine ventricular fibrillation may lead to the incorrect assumption that asystole is present. If any doubt exists the patient should be treated as for ventricular fibrillation.

Adrenaline (epinephrine) 1 mg, followed by atropine 2 mg, should be given through a central venous line. Isoprenaline may be given if these first two agents fail to re-establish ventricular fibrillation or another rhythm. As with refractory ventricular fibrillation, persistent asystole may respond to either external or internal electrical pacing. There are few, if any, indications for the intracardiac administration of medication: the risk of myocardial damage is high and the benefits questionable.

Cerebral resuscitation

A final indicator of the success or failure of cardiopulmonary resuscitation is the subsequent cerebral function. Many of the factors determining the development of neurological sequelae are self-evident, such as the duration of circulatory standstill. Other factors, such as the rapidity with which circulatory and metabolic homeostasis can be achieved, also have a significant bearing on neurological outcome. Aspects of cerebral resuscitation are considered further elsewhere.

Circulatory failure

Circulatory failure not due to cardiac arrest can be divided broadly into hypovolaemic, cardiogenic, and distributive types. Because of the hypotension and tachycardia associated with these, they are often referred to as shock syndromes. The causes of these three types of circulatory failure are summarized in Table 1.



Table 1 Classification of the circulatory failure syndromes

Hypovolaemic

The clinical response to intravascular volume depletion varies considerably, depending upon the rapidity of depletion and the peripheral vasoconstrictor responses. The classic clinical picture includes orthostatic hypotension, tachycardia, pallor, tachypnoea, cold vasoconstricted peripheries, oliguria, and mental obtundation. Minor volume depletion may be unmasked by the inadvertent administration of a vasodilator such as glyceryl trinitrate or the commencement of positive-pressure ventilation. Prolonged periods of hypovolaemic shock result in permanent tissue injury developing either during the period of hypoperfusion or during subsequent reperfusion.

Management and monitoring

Clinical evaluation of the degree of hypovolaemia precedes any attempts to establish invasive monitoring. The blood pressure, degree of orthostasis, tachycardia, sweating, and peripheral vasoconstriction, together with signs of end-organ hypoperfusion such as neurological obtundation and cardiac arrhythmias, will impress on the clinician the urgency of therapy. A simple clinical classification of the degree of hypovolaemia is shown in Table 2 and provides a guide to the blood-volume deficit. Transfusion must be begunn via large-bore (14 gauge), peripheral intravenous cannulas before time-consuming attempts to establish central venous pressure and arterial lines are made.

	Class 1	Class 2	Class 3	Class 4
Heart rate (beats/min)	< 100	> 100	> 120	> 140
Blood pressure	Normal	Normal	Decreased	Decreased
Pulse pressure	Normal or increased	Decreased	Decreased	Decreased
Capillary refill	Normal	Delayed	Delayed	Very delayed
Blood loss (litres)	Up to 0.75	0.75–1.5	1.5–2.0	2.0 plus
Blood loss (%)	Up to 15	15–30	30–40	> 40

Table 2 Classification of severity of hypovolaemia

Venous access is of prime importance. Small, peripheral venous and triple-lumen central lines are generally inadequate for rapid, large-volume transfusions. Several large peripheral lines combined with a large-bore (8 or 8.5 Fr) central line will usually suffice. If a peripheral cut-down is required, the insertion of the cut end of a sterile giving set (without the connector) directly into a vein ensures rapid transfusion.

The patient should be placed in the ‘feet up–head down posture', and given oxygen supplementation. *In extremis*, with profound hypotension and a cardiac arrest developing, adrenaline (epinephrine) in 1-mg boluses should be considered. In trauma patients, whose injuries may be associated with uncontrollable blood loss, no rate or volume of transfusion will save their life unless emergency surgery is undertaken.

A line for monitoring arterial and central venous pressure greatly facilitates assessment of resuscitation. Correction of blood pressure, resolution of tachycardia, and an increase in central venous pressure (approaching 10–12 mmHg, relative to the right atrium) indicate satisfactory volume expansion; warming of the skin, absence of orthostatic hypotension, urine production of 0.5 to 1 ml/kg per hour, and resolution of metabolic acidosis indicate re-establishment of organ perfusion.

Transfusion fluids

When undertaking fluid resuscitation the clinician must first address the issue of what fluid and how much? Although there is controversy over the selection of crystalloid or colloid, some rationalization can be applied case by case. The supporters of crystalloid fluids point out that hypovolaemia affects both the intravascular and interstitial spaces and that crystalloids are distributed readily to both spaces in a ratio of 1:3. The proponents of colloidal solutions emphasize the urgency of expanding the intravascular space to defend the circulation. Some agents, such as the hydroxyethyl starches, produce a volume expansion greater than the transfused volume, due to their osmotic properties. Obviously, patients with slowly developing hypovolaemia secondary to long-standing gastroenterological losses, and who may be haemoconcentrated, will benefit from balanced electrolyte replenishment. The massively bleeding trauma patient could be initially and briefly supported with any fluid. Very soon, however, haemodilution effects and coagulation defects will mandate blood and blood-product replacement. Clearly, it is the volume and speed of replacement that determines outcome more than the initial selection of the type of fluid.

When massive transfusion is required, blood should always be warmed and continual replacement of clotting factors considered. The difficulty of achieving a balance between under- and over-transfusion should not be underestimated, even with all appropriate monitoring facilities. Remember that prolonged hypoperfusion carries the risk of organ tissue damage, while fluid overload, although not well tolerated in elderly persons, can be reversed with diuretics or haemofiltration.

Cardiogenic

The initial clinical evaluation of cardiogenic failure should determine the nature of the failure, the underlying cause, and the severity of the problem. A full clinical history and physical examination should be performed to detect ischaemic heart disease, hypertension, alcoholism, viral syndrome, and valvular or congenital heart disease. A 12-lead electrocardiogram should be obtained to identify acute myocardial infarction, left ventricular hypertrophy, heart block, or arrhythmias. Chest radiography is required for the assessment of heart size and pulmonary vascular markings, and pleural effusion.

If the patient is suffering from acute pulmonary oedema or cardiogenic shock, treatment should begin immediately. If there is less urgency an echocardiogram provides information about the size, thickness, and performance of the heart. The Doppler mode provides unique, non-invasive assessment of valvular function and allows for detection of intracardiac shunts.

The treatment of cardiogenic failure is based upon the manipulation of myocardial contractility, cardiac preload and afterload. This is achieved by giving inotropes, fluids, diuretics, and vasodilators, alone or in combination. The success or failure of therapy should be measured against clear therapeutic endpoints, such as a doubling of cardiac output, or a 25 per cent fall in pulmonary arterial wedge pressure. The electrocardiogram, systemic arterial pressure, cardiac output, and pulmonary arterial wedge pressure should be carefully monitored throughout the period of therapy. An arterial cannula is essential for monitoring patients receiving vasodilator or inotropic therapy, particularly when the blood pressure is labile. Flow-directed pulmonary arterial catheters are most useful when there is uncertainty about left ventricular filling pressure or cardiac output.

Myocardial preload

Preload is best represented by the ventricular end-diastolic volume, although clinically it is more convenient to use ventricular end-diastolic pressure. The end-diastolic pressure is, in turn, approximated by the atrial pressures—right atrial pressure for the right ventricle and pulmonary arterial wedge pressure for the left ventricle. Increases in preload are associated with increases in both the extent and velocity of muscle fibre shortening, which combine to produce an increase in stroke volume. Heart failure is characterized by a limited increase in left ventricular stroke work for a given rise in left ventricular end-diastolic pressure (Fig. 1).

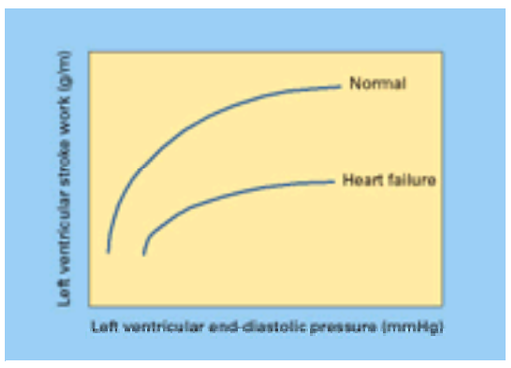


Fig. 1. Ventricular performance curves for normal individuals and patients with congestive heart failure. The failing heart is less responsive to increases in preload than the normal heart.

The correlation between end-diastolic pressure stroke and volume is non-linear, and in the failing heart, contractility is less responsive to increases in end-diastolic pressure. Nevertheless, alterations in preload are important determinants of cardiac performance in both normal and failing hearts. The response of heart muscle to changes in preload produces a functional reserve capable of improving cardiac output. A simple method of confirming the relation between the ventricular end-diastolic volume and the ventricular end-diastolic pressure involves the use of a rapid fluid challenge. Following such a challenge, a significant and persistent rise in end-diastolic pressure without an increase in cardiac output suggests that further fluid loading may be ill advised. Conversely, improved systemic arterial pressure and urine output without much change in end-diastolic pressure suggests that further fluid loading may be beneficial. A summary of factors that alter preload is shown in Table 3.

Factors that increase preload
Venoconstriction
Transfusion
'Head down' position
Factors that decrease preload
Venodilation
Hypovolaemia
Raised intrathoracic pressure:
Positive pressure ventilation
Tension pneumothorax
Cardiac tamponade
Loss of atrial 'pump'

Table 3 Factors that modulate preload

Central venous pressure

Any attempts to modulate preload require continuous and accurate measurement of central venous pressure and pulmonary arterial wedge pressure. Central venous pressure should be transduced and displayed on the bedside monitor: it is difficult to estimate this with confidence by water manometry. Pressure should be measured at the peak of the ‘a' wave (if in sinus rhythm) and at end-expiration. Electronic ‘mean' venous pressures may overestimate central venous pressure in mechanically ventilated patients and underestimate it in those breathing spontaneously. Since the venous pressure is so critical it should be measured by the clinical attendant at the time and not taken from the previously documented clinical record.

Myocardial afterload

Afterload reflects the stress distributed within the ventricular wall during ejection and is determined by impedance factors opposing ejection and by wall tension. Wall tension in turn is determined by the La Place relation between ventricular cavity radius and pressure.

Thus, afterload is not constant during ventricular ejection, but decreases as ventricular volume diminishes. In the normal heart, increases in afterload, such as an elevation in blood pressure or systemic vascular resistance, lead to a compensatory increase in preload to maintain the stroke volume.

The forces resisting left ventricular ejection can be referred to as impedance: these include the resistance of the small arteries and arterioles, the compliance and inertia of the large arteries, the viscosity of blood, and the inertia of the blood itself. Impedance (systemic vascular resistance and pulmonary vascular resistance) is the peripheral component of afterload. Vasoconstriction (or polycythaemia) therefore results in increased impedance and increased afterload. Recognizing the factors that contribute to afterload helps identify avenues of intervention that can support the heart during periods of circulatory failure (Table 4).

Factors that increase afterload
Increased aortic impedance
Arteriolar constriction
Dilatation of ventricular cavity
Factors that decrease afterload
Decreased aortic impedance
Arteriolar dilatation
Reduction in ventricular cavity radius

Table 4 Factors that modulate afterload

Although few therapeutic situations require an increase in afterload, one such arises during isolated right ventricular failure, as might occur after myocardial infarction. Left ventricular preload is low because of impaired right ventricular output. Any factor reducing left ventricular afterload, such as volume depletion, vasodilators, or epidural anaesthetics, may preferentially reduce coronary perfusion of the right heart by reducing systolic pressure. Right ventricular function then becomes further impaired. If right ventricular dilatation then occurs, the displacement of the interventricular septum further reduces left ventricular ejection. Left ventricular function will not improve until left ventricular afterload is increased by the administration of a systemic vasoconstrictor to improve right ventricular coronary blood flow. High right ventricular filling pressures must be maintained to optimize pulmonary flow and venous return to the left heart.

Therapeutic benefit is more commonly derived by reducing afterload in cardiac failure. This is usually achieved by giving vasodilators to reduce impedance. Vasodilators may have different effects on arterial resistance, arterial compliance, and left ventricular volume. The greatest benefit will be seen in hypertensive patients, but significant effect can still be obtained in normotensive patients with heart failure in whom the systolic arterial pressure exceeds 90 mmHg. Several agents are available, including sodium nitroprusside, glyceryl trinitrate, hydralazine, and the angiotensin-converting enzyme inhibitors. In addition to the use of pharmacological agents, aortic impedance can be minimized by avoiding polycythaemia and maintaining a haemoglobin concentration of between 9 and 11 g/dl. The intra-aortic counterpulsation balloon pump is also effective in reducing aortic impedance and has the dvantage of maintaining coronary filling pressure.

Therapeutic manipulation of preload and afterload

Sodium nitroprusside

Nitroprusside is a vasodilator that acts directly on both arteriolar and venous smooth muscle. At least part of its action may be related to the inhibition of platelet aggregation and thromboxane A_2 synthesis, together with it synergism with prostacyclin. It exerts its effects directly on vascular smooth muscle rather than through a receptor system, and, by reducing systemic vascular resistance, it reduces myocardial oxygen consumption and improves myocardial function. Nitroprusside is a direct dilator of vessels within the substance of the myocardium, resulting in increased coronary blood flow and markedly decreasing coronary vascular resistance. Venodilation causes some reduction in preload by increasing the capacity of the venous bed. This in turn may be useful in decreasing pulmonary congestion. A theoretical concern over the use of nitroprusside in patients with myocardial ischaemia is that it might cause preferential shunting of flow into non-ischaemic areas (the ‘coronary steal' phenomenon).

Nitroprusside dilates the cerebral circulation as well as other vascular beds. This is potentially dangerous in patients with pre-existing intracranial hypertension, and may indicate a need for monitoring intracranial pressure. It does not impair normal cerebral autoregulation.

Pulmonary circulation Nitroprusside dilates the pulmonary circulation. By blocking hypoxic vasoconstriction it may reduce arterial oxygenation in 10 to 20 per cent of patients, owing to increased pulmonary venous admixture.

Metabolism Nitroprusside is an iron co-ordination complex that is metabolized in the blood to cyanide and then to thiocyanate (via rhodanese), and excreted in the urine. The thiocyanate metabolite itself has a mild hypotensive effect that is dose dependent and does not appear to form a plateau. Hypotension is generally reversed within 5 to 10 min of stopping infusion of the drug.

Indications Nitroprusside is indicated in the treatment of congestive heart failure, in acute myocardial infarction, aortic insufficiency, acute mitral insufficiency due to papillary muscle dysfunction, or ischaemic ventricular septal defect. It should be avoided in patients with mitral or aortic stenosis. Nitroprusside is particularly

appropriate for hypertensive patients with acute myocardial infarction and persistent chest pain or left ventricular failure, and for normotensive patients with severe pump failure. It should be avoided in hypotensive patients, although it can be used successfully in conjunction with inotropic agents or intra-aortic balloon counterpulsation. Monitoring by pulmonary arterial catheter is generally required. Nitroprusside is particularly suitable for the treatment of hypertensive crises. In view of this propensity to induce excessive hypotension, an arterial line is generally considered a prerequisite in nitroprusside therapy. Almost all hypertensive patients will respond to nitroprusside, although resistance has been described in patients with severe hypertension and renal failure. An additive hypotensive effect is seen with other drugs and it is generally wise to institute other antihypertensive therapy as soon as the blood pressure is controlled.

Dose Generally, only modest doses of nitroprusside are needed: an initial infusion of 10 to 15 µg/min can be increased by 10 µg/min every 5 to 15 min. Most patients with heart failure show a positive response to 70 to 140 µg/min (1–2 µg/kg per min). In patients with pulmonary oedema accompanying congestive heart failure, nitroprusside is also started at a dose of 10 to 15 µg/min, but the dose is more rapidly increased (20-µg/min increments every 3–5 min) in order rapidly to reduce filling pressure and relieve symptoms. During infusion, the pulmonary arterial wedge pressure should not be allowed to fall below 15 mmHg and the mean arterial pressure should be maintained at 70 mmHg (diastolic blood pressure > 50 mmHg). Initial doses of 0.5 to 1 µg/kg per minute are generally effective in the treatment of hypertensive crises.

Response to therapy A positive response to nitroprusside infusion consists of either a drop in the pulmonary arterial wedge pressure or an increase in cardiac output (or both). A decrease of 20 to 50 per cent in the wedge pressure and an increase in cardiac output of 20 to 40 per cent are considered positive responses. If blood pressure falls without an improvement in cardiac output or a decrease in wedge pressure, nitroprusside should be discontinued or an inotrope added.

The dose required to produce a given hypotensive effect in patients with hypertensive crisis is variable, but the maximum recommended is about 8 µg/kg per minute. Since the drug deteriorates in the light, the administration set should be opaque or covered in aluminium foil. Once diluted, nitroprusside remains stable for about 24 h under such conditions. Longer-term medications can be substituted once blood pressure has been controlled with nitroprusside.

Toxicity Because nitroprusside is metabolized to thiocyanate, excessive amounts given over long periods to patients with severely compromised renal function can result in the accumulation of thiocyanate. Side-effects include nausea, vomiting, hiccups, mental confusion, and psychotic behaviour. The clinical manifestations of thiocyanate toxicity are lactic acidosis, confusion, hyper-reflexia, convulsions, tinnitus, and blurred vision. Infusion rates below 3 µg/kg per minute for less than 72 h are not usually associated with toxicity. Monitoring blood thiocyanate may be necessary in patients requiring infusions for longer than 2 to 3 days: concentrations below 10 mg/dl are considered satisfactory.

Glyceryl trinitrate (GTN, TNG, nitroglycerine)

Nitrates cause direct relaxation of vascular smooth muscle. Their predominant effect is on the capacitance vessels (veins), although arterioles are also dilated at higher doses. They may act on smooth muscle by releasing prostacyclin from vascular endothelial cells. Both glyceryl trinitrate and nitroprusside exert a similar effect on preload, with nitroprusside causing a greater decrease in afterload.

Glyceryl trinitrate has an antianginal effect. Several studies suggest that its major effect is due to systemic venodilation, which reduces preload and ventricular size, and in turn decreases left ventricular end-diastolic pressure, intramyocardial wall tension, and myocardial oxygen consumption. The drug also decreases ischaemia by promoting epicardial and collateral coronary blood flow. Within the usual dose range, glyceryl trinitrate has minimal effects on arteriolar tone: blood pressure and cardiac output are rarely affected unless preload is markedly reduced or unless significant myocardial hypoxia is relieved. In the higher dose range, it acts in a similar fashion to nitroprusside, causing a fall in systemic vascular resistance, increasing cardiac output, and decreasing blood pressure.

Metabolism Glyceryl trinitrate is widely distributed in the body and is rapidly metabolized to dinitrates and mononitrates, with a half-life of approx. 4 min.

Indications Because glyceryl trinitrate has little direct effect on blood pressure but has a profound effect on pulmonary arterial wedge pressure, it may be preferable to nitroprusside in patients without hypertension who have heart failure and pulmonary oedema. The intravenous dosage is highly variable but can be started at 10 mg/min and adjusted every 5 to 10 min until the desired haemodynamic response (drop in wedge pressure) is achieved. The dose that is usually effective in patients with heart failure is 30 to 100 mg/min. If no benefit is derived at a dose of 400 to 500 mg/min, other pharmacological options should be considered. High doses of glyceryl trinitrate over several days are well tolerated. Continuous infusion is titrated to produce relief of chest pain or reversal of ischaemic electrocardiographic changes. It is essential that left ventricular filling pressure be adequate, otherwise significant hypotension may ensue. Occasionally, a precipitous fall in blood pressure in response to glyceryl trinitrate may be the first indication that a particular patient is intravascularly depleted.

Toxicity Side-effects include headache, sinus tachycardia, and hypotension. Rare complications include methaemoglobinaemia, paradoxical hypertension with bradycardia, and exacerbation of hypoxaemia.

Angiotensin-converting enzyme inhibitors

Several studies, for example CONSENSUS I, SOLVD, SAVE, have shown that angiotensin-converting enzyme (**ACE**) inhibitors reduced mortality in heart failure. Many other therapies, such as digoxin, amlodipine, and amiodarone, have also been examined, but have not had as good effect on mortality. Despite ACE inhibitors being a very effective therapy for heart failure, they are probably underused for fear of side-effects such as hypotension and renal dysfunction. Patients who are receiving inotropes and afterload reduction in intensive care can be slowly introduced to ACE inhibitors while their blood pressure, urine output, and renal chemistry are closely monitored. Typically, an agent such as captopril is started in low dose, 6.25 mg three times a day, increasing the dose incrementally until a therapeutic effect is obtained. Recently, angiotensin II receptor antagonists have demonstrated similar effects on the circulation and may in future prove to be effective in the treatment of acute and longer-term heart failure.

Toxicity Profound hypotension and renal failure may occur, especially in volume-depleted patients. ACE inhibitors are best avoided in patients with renovascular disease. Other side-effects include dry cough, pancreatitis, and blood dyscrasias.

Diuretics

Short-acting loop diuretics are considered an essential component of the treatment of acute heart failure. Frusemide (furosemide) in doses of 10 to 100 mg should be given intravenously to patients with increased preload. Extremely large doses of frusemide (250–4000 mg/day) have been recommended in patients with reduced renal function but are unlikely to produce a meaningful diuresis: the addition of 2.5 to 5 mg of oral metolazone to frusemide is preferred. Intravenous bumetanide, a potent and short-acting loop diuretic, in doses of 0.5 to 2 mg, may be effective in patients resistant to frusemide. Stable patients with symptomatic heart failure and poor left ventricular function have recently been shown to benefit from a 30 per cent reduction in mortality when taking spironolactone 25 mg daily.

Although the optimal response to diuretic therapy is a reduction in left ventricular filling pressure and increased urine output, the response of the left ventricular pump function to diuretics is variable. Loop diuretics are of great value in patients with decompensated heart failure and should be used as needed to control pulmonary congestion. In the absence of a significant and sustained diuresis in patients with renal failure, short-term reliance on venodilators may buy time while alternative means such as haemofiltration are established.

Myocardial contractility

The terms contractility and myocardial performance should not be confused: they are not synonymous. Myocardial performance is the sum total of preload, afterload, contractility, and heart rate, and is usually measured as cardiac output, while contractility is only one component of myocardial performance. Contractility is the final determinant of cardiac output and can be defined as the force of ejection, independent of afterload or preload. Contractility is not fixed but variable: it improves with adrenergic stimulation and increased coronary blood flow, and worsens due to the action of pharmacological and physiological depressants (barbiturates, general anaesthesia), metabolic abnormalities (hypothyroidism, sepsis), and loss of ventricular substance (myocardial infarction). Some of the factors that influence contractility are listed in Table 5.

Factors that increase contractility
Sympathetic stimulation
Endogenous catecholamines
Exogenous catecholamines and inotropes
Increased myocardial oxygen delivery
Factors that decrease contractility
Myocardial hypoxia
Myocardial ischaemia
Metabolic depressants
Pharmacological depressants

Table 5 Factors that influence myocardial contractility

The direct measurement of contractility is technically difficult, since most techniques, such as ejection fraction, may be influenced by the relative state of preload or afterload at the time the measurement is made. However, the direct measurement is conceptually important because it can be increased by a number of inotropic agents commonly used to treat circulatory failure. The most reliable and commonly used inotropic agents include dopamine, dobutamine, adrenaline (epinephrine), noradrenaline (norepinephrine), isoprenaline, and glucagon.

Therapeutic manipulation of contractility

Dopamine Dopamine is the immediate precursor of noradrenaline (norepinephrine) in the catecholamine synthetic pathway. It has both a- and b-adrenergic effects but differs from other catecholamines by producing vasodilation in renal, mesenteric, coronary, and intracerebral arterial vascular beds. This dopaminergic effect is not blocked by b-blockers. Since dopamine acts on a variety of receptors, and each receptor has a different dose–response relation in different patients, a wide range of haemodynamic effects may be elicited under different conditions. Dopamine exerts b_1-adrenergic activity, mainly by releasing noradrenaline (norepinephrine) from myocardial storage sites. It also increases contractility and heart rate by its direct action on b-adrenergic receptors, effects that are blocked by b-blockers. a-Adrenergic effects causing vasoconstriction occur at high doses. In addition to its direct effect on a-adrenergic receptors, dopamine may also cause contraction of vascular smooth muscle by acting on a serotonin- or tryptamine-sensitive receptor.

Cardiovascular effects Low doses (0.5–2 mg/kg per min) may cause a minimal increase in cardiac output and contractility. Its major effect is an increase in renal blood flow and urine output due to stimulation of dopaminergic receptors. The renovascular effect is probably the most commonly exploited property of dopamine, but it may be lost at infusion rates greater than 5 mg/kg per minute, when there is an increase in cardiac contractility. At even higher doses (> 10 mg/kg per min) the a-adrenergic effects increase. Pulmonary arterial and pulmonary arterial wedge pressure may increase at high doses of dopamine. In addition, dopamine may increase the intrapulmonary shunt fraction, resulting in a fall in oxygenation.

Dopamine selectivity increases renal and mesenteric blood flow by its action on dopaminergic receptors, but some of the renal effects of dopamine may be due to the inhibition of aldosterone secretion resulting in a natriuresis and increased urine output.

Dose Intravenous administration of dopamine results in near steady-state concentrations in 5 min. The half-life of intravenously administered dopamine is approx. 1 min: it is metabolized by both catechol-*O*-methyl transferase and monoamine oxidase enzyme systems, and is ineffective when given orally. Dopamine is inactivated by bicarbonate and other alkaline solutions. The initial dose depends on specific aims of therapy: for the promotion of renal or splanchnic blood flow, doses of 1.5 to 2.5 mg/kg per minute are generally adequate, while a vasopressor effect is only achieved at doses greater than 5 mg/kg per minute. Most patients show a pressor response to dopamine infusion at doses below 20 mg/kg per minute, although some may require infusion rates in excess of 50 mg/kg per minute. At these dose rates there are significant vasoconstrictor effects.

Toxicity Arrhythmias are generally associated with high doses of dopamine or the presence of myocardial ischaemia, metabolic acidosis, or hypoxaemia. Dopamine may increase myocardial conduction and precipitate a rapid ventricular response in patients with atrial fibrillation.

Peripheral gangrene may be seen in patients with profound shock treated for a prolonged period with large doses of dopamine. Hypotension may occur secondarily to vasodilation in certain patients receiving low doses of dopamine: a specific vasopressor should be given until an adequate blood pressure is obtained. Other side-effects include nausea, vomiting, angina pectoris, and, occasionally, hyperglycaemia. Tissue necrosis due to local extravasation is a serious complication and for this reason dopamine should routinely be given via a central vein. If peripheral extravasation does occur, the area should be infiltrated locally with 10 ml of normal saline containing 5 to 10 mg of phentolamine.

Interactions with other drugs Dopamine is metabolized by the monoamine oxidase system and its effects are therefore greatly potentiated in patients receiving monoamine oxidase inhibitors. These effects may persist several weeks after cessation of such inhibitors. Phenothiazines, haloperidol, and tricyclic antidepressants have mild a-adrenergic blocking actions that may reduce the peripheral vasoconstricting effects of dopamine. Propanolol and other b-blocking agents blunt the cardiac stimulation produced by dopamine.

Dobutamine Dobutamine is a synthetic catecholamine that was the product of a systematic attempt to design a pure b_1-adrenergic agent. It is structurally related to isoprenaline and acts directly on b_1-adrenergic receptors in the myocardium. Unlike dopamine, dobutamine does not enhance the release of noradrenaline (norepinephrine) from nerve endings, nor does it act on dopamine receptors. Its action is therefore not potentiated by monoamine oxidase inhibitors. The predominant action on b_1-receptors increases myocardial contractility, and it has less effect on heart rate than does dopamine. Dobutamine has relatively weak b_2- and a-receptor activity, with peripheral vasodilation predominating. It has a short half-life (< 2.5 min) in patients with heart failure: this becomes important if it is necessary to reverse an adverse effect such as ventricular tachycardia. There are no biologically active metabolites of dobutamine.

Cardiovascular effects At low to moderate dose, dobutamine increases myocardial contractility, causes peripheral vasodilation, and augments renal blood flow. Some increase in heart rate can be expected. In addition, a decrease in pulmonary arterial wedge pressure often accompanies the improved cardiac output.

Pulmonary effects Dobutamine increases cardiac output and by doing so increases intrapulmonary venous admixture. Relief of pulmonary venous congestion brought about by a reduction in pulmonary arterial wedge pressure improves pulmonary compliance and reduces the work of breathing.

Metabolism Dobutamine is metabolized by catechol-*O*-methyl transferase and glucuronide transformation in the liver; most of the drug being eliminated unchanged by the kidneys and biliary tract. Its elimination follows first-order kinetics.

Dose Positive inotropic effects occur with doses as low as 0.5 mg/kg per minute, although the usual initial dose is 2 to 2.5 mg/kg per minute. Therapeutic effects of the drug usually plateau at 15 to 20 mg/kg per minute. Dobutamine is useful in the treatment of cardiogenic and septic shock if the associated hypotension is not severe. If hypotension is a problem, noradrenaline (norepinephrine) or high doses of dopamine may need to be added.

Toxicity Cardiac arrhythmias are the most frequent toxic side-effect, but these are less common than with dopamine or isoprenaline. Hypotension may occur, especially if preload is inadequate. Tolerance may be observed after prolonged continuous infusion, an effect that is related to downregulation of b-receptors. Dobutamine is contraindicated in patients with obstructive cardiomyopathy, since it may increase cardiac outflow obstruction. The ventricular response in atrial fibrillation may be increased.

Dopexamine Dopexamine is a synthetic catecholamine with potent b_2- and dopaminergic agonist activity. It has one-third of the potency of dopamine on dopaminergic receptors and, since it is a potent inhibitor of the uptake-l process, it potentiates the actions of endogenous catecholamines. Dopexamine has no a-receptor agonist activity and may even be an antagonist at the a-adrenoceptors. At low to moderate dose, dopexamine increases myocardial contractility, causes peripheral vasodilatation, and augments renal and hepatosplanchnic blood flow. An increase in heart rate can be expected, particularly in the presence of volume depletion. Pulmonary arterial wedge pressure may fall as ventricular function is improved with afterload reduction.

The main features of dopexamine's actions are peripheral, renal, and splanchnic vasodilatation. The reduced peripheral resistance contributes to moderate increases in cardiac output and heart rate. The renal effects of dopexamine include vasodilatation, increased renal blood flow, increased creatinine clearance, and natriuresis. The

DA_1 receptors located on the renal tubules may be responsible for both diuresis and natriuresis. In experimental models of renal failure caused by hypovolaemic shock, dopexamine has been shown to facilitate recovery of renal function.

The inotropic activity of dopexamine results from stimulation of the b_2-receptors, indirect stimulation of the b_1-receptors via inhibition of neuronal uptake of noradrenaline (norepinephrine), and activation of the baroreflex mechanism. In chronic heart failure, b_1- but not b_2-adrenoceptors are downregulated. This reduces the efficacy of b_1-adrenoceptor agonists, but does not prevent the inotropic effect of dopexamine.

Dopexamine is particularly effective in patients with low-output circulatory failure associated with an elevated vascular resistance. Within a dose range of 0.5–2.0g/kg per minute, it produces a reduction in vascular resistance associated with an increase in cardiac output. In addition, it improves renal and splanchnic blood flow. This may improve oxygen delivery to the splanchnic region and thus contribute to a reduction of the risk of multiple organ failure. Indeed, a recent randomized study of the use of dopexamine in the perioperative phase of 107 patients undergoing major surgery resulted in a 75 per cent fall in mortality. This was attributed to improved perfusion and oxygen delivery to organs such as the gastrointestinal tract and the kidneys. Further studies are currently under way to confirm these findings.

It has been stated that dopexamine does not increase the myocardial oxygen cost of increasing cardiac output. However, there are trends towards increased oxygen consumption and reduced lactate extraction. In comparison with dobutamine, dopexamine causes lesser increases in myocardial oxygen consumption.

After cardiac surgery, dopexamine increases heart rate and cardiac output, while it decreases pulmonary and systemic vascular resistance, as well as pulmonary capillary wedge pressure and right atrial pressure. This is associated with a small reduction in mean arterial pressure. The increase in heart rate is transient and diminishes with continued treatment, although the increase in cardiac output is maintained.

Dose Positive inotropic effects occur with doses as low as 0.5 µg/kg per minute; optimal therapeutic effects are seen at l to 2 µg/kg per minute. Hypotension may occur if higher doses are used, especially if fluid loading is inadequate.

Side-effects Tachycardia, ventricular ectopy, and cardiac arrhythmias are the most common side-effects. Hypotension may occur in volume-depleted patients. Dopexamine is contraindicated in obstructive cardiomyopathy, aortic stenosis, thrombocytopenia, phaeochromocytoma, and therapy with monoamine oxidase inhibitors.

Noradrenaline (norepinephrine) Noradrenaline is the neurotransmitter of postganglionic sympathetic nerves and is released from the adrenal medulla. In the heart, it is synthesized and stored in granules in myocardial adrenergic nerve endings. It is a b_1-agonist and has minimal effects on b_2-receptors. a-Adrenergic receptor effects result in marked peripheral vasoconstriction and increased left ventricular afterload.

Cardiovascular effects Cardiovascular effects are not apparent with doses below 2.5 mg/min, but above that, noradrenaline increases systolic pressure proportionately more than diastolic pressure. This results in a significant increase in systemic vascular resistance and mean arterial pressure. Heart rate may be slowed by a baroreceptor reflex and cardiac output remains unchanged or slightly reduced. It is a potent venoconstrictor, and increases venous return by decreasing vascular capacitance. It increases afterload, preload and contractility, and can greatly increase myocardial work and oxygen demand. Coronary blood flow may be increased through an increase in the filling pressure gradient between the mean arterial pressure and the left ventricular end-diastolic filling pressure. Right ventricular perfusion may be greatly enhanced as the majority of right-sided coronary flow occurs during systole.

Pulmonary effects Noradrenaline may be a respiratory stimulant through its action on the carotid bodies. It has little effect on bronchial smooth muscle, but it will increase pulmonary vascular resistance: this is potentially disadvantageous in patients with underlying pulmonary hypertension.

Other vascular effects Physiologically, noradrenaline is a potent vasoconstrictor of the renal arterial bed. This effect can be ameliorated to some degree by concurrent administration of low-dose dopamine. Noradrenaline produces vasoconstriction in the liver and splanchnic beds, resulting in decreased flow. However, in patients with distributive (septic) shock, it may increase renal blood flow and enhance urine production by increasing perfusion.

Noradrenaline does not constrict vessels supplying the brain, although a_2-adrenergic receptors are present on them. Its powerful pressor effects may maintain cerebral perfusion during periods of circulatory collapse; indeed, it may be as effective as adrenaline (epinephrine) in supporting the circulation during cardiac arrest.

Metabolism Noradrenaline is enzymatically degraded in the liver and kidneys. There is also reuptake at b-adrenergic and non-neuronal receptor sites.

Dose The pressor effects of noradrenaline may be beneficial in distributive (septic) shock, where there is systemic vasodilation and peripheral hypoperfusion as evidenced by low systemic vascular resistance and lactic acidosis. It will support myocardial and cerebral perfusion effectively while cardiac output and oxygen delivery are carefully monitored. Although there are theoretical limits to the dose that can be safely administered, it can be increased progressively to achieve the desired effect. It is essential to exclude volume depletion before resorting to pressor therapy. Following overdose of a calcium-channel blocker, intravenous calcium and noradrenaline may quickly restore vascular tone.

Toxicity Systemic vasoconstriction resulting in organ ischaemia, especially of the dermal, renal, and mesenteric vascular beds, may produce irreversible injury. These effects must be weighed against the potential benefits. Palpitations, angina, headaches, hyperglycaemia, and hypocalcaemia may follow administration of noradrenaline (norepinephrine).

Adrenaline (epinephrine) Adrenaline is a naturally occurring catecholamine with a-, b_1-, and b_2-adrenergic activity. For circulatory support it may be infused in doses of between 0.01 and 0.2 mg/kg per minute. Although the oxygen supply:oxygen demand ratio may be adversely affected, adrenaline may improve peripheral perfusion. It is particularly useful in the treatment of distributive circulatory failure, such as the sepsis syndrome and anaphylactic reactions. Although the same effects can be achieved by combinations of agents with predominantly a-adrenergic effects [noradrenaline (norepi-nephrine)] and b-adrenergic effects (dobutamine), adrenaline still has a place in the current management of circulatory failure.

Isoprenaline Isoprenaline has both b_1- and b_2-adrenergic activity, and produces peripheral vasodilation and an increase in myocardial contractility and heart rate. The tachycardia and reduced coronary perfusion pressure result in a much reduced oxygen supply:oxygen demand ratio. It is therefore not the drug of choice for use in patients with myocardial failure, but it may have application in severe bradycardias not associated with myocardial ischaemia.

Glucagon Glucagon is a pancreatic polypeptide that has inotropic and chronotropic effects that are not dependent on b-receptor responsiveness. By directly stimulating adenylcyclase it increases intracellular AMP concentration. Although not a first-line drug for use in circulatory failure, it may be effective when other inotropes have failed.

Phosphodiesterase inhibitors Milrinone and enoximone are powerful phosphodiesterase inhibitors that may be beneficial in patients with severe heart failure who are unresponsive to dobutamine or dopamine. These agents have combined inotropic and vasodilator properties, and are effective even in the presence of b-receptor downregulation. Although they may improve symptoms they have no impact on mortality.

The selection of inotropes and vasodilators in cardiac failure

Dobutamine and dopamine have very different haemodynamic profiles and should not be considered interchangeable. Because dopamine causes substantial peripheral vasoconstriction and can increase pulmonary capillary wedge pressure, it should be used cautiously in patients with acute heart failure who have increased peripheral vascular resistance and elevated wedge pressures. Dobutamine may be preferable to dopamine for the treatment of acute congestive heart failure, while nitroprusside is preferable when the systolic pressure is above 90 mmHg. Combinations of all three drugs are commonly used. A logical regimen in a patient with severe congestive heart failure would be nitroprusside and dobutamine with infusion of low (renovascular) doses of dopamine. Recently introduced inotropic agents such as the b-adrenergic agonists amrinone and dopexamine and the phosphodiesterase inhibitors milrinone and enoximone combine inotropic and vasodilator properties (afterload reduction) and theoretically reduce the risk of increasing myocardial work.

The effects of inotropes, vasodilators, and diuretics on myocardial performance are well represented by the plot of left ventricular stroke-work index and pulmonary arterial wedge pressure. In the failing heart there is greatly reduced myocardial performance for a given preload. As shown in Fig. 2, the depressed curve of heart failure can be shifted toward normal by inotropic drugs or vasodilator drugs.

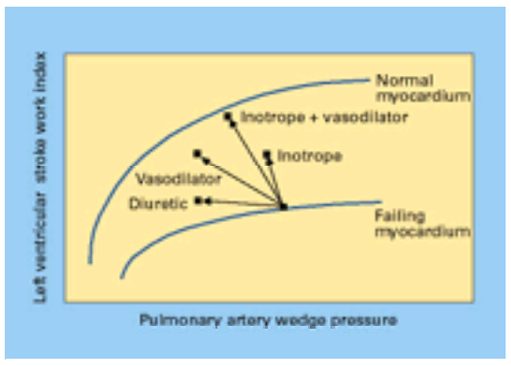


Fig. 2. The effects of inotrope, vasodilator, and diuretic upon the relation of left ventricular performance to pulmonary arterial wedge pressure.

These effects may be complementary when the drugs are infused together. Note that diuretics usually reduce filling pressure without augmenting output. An inotrope such as dobutamine may not only increase contractility but may also reduce pulmonary arterial wedge pressure, while dopamine may increase pulmonary arterial wedge pressure, especially at high doses.

Ultimately, oral inotropes and vasoactive agents will be needed to replace acute systemic therapy. Several effective oral agents, including diuretics, cardiac glycosides (digoxin), nitrates, hydralazine hydrochloride, and ACE inhibitors, are available for the treatment of stable or compensated congestive failure. Of these the combination of diuretics and ACE inhibitors has had the most impact on morbidity. Patients should be supervised while a low dose of an oral agent (such as captopril 6.25 mg or enalapril maleate 2.5 mg) is administered and the intravenous vasodilator drug is gradually discontinued. Cautious dose increases are needed in order to avoid hypotension, especially if the patient is volume-depleted. If the first dose of an ACE inhibitor is tolerated, doses should be increased until a maintenance dose is achieved. Since ACE inhibition tends to increase serum potassium, potassium supplementation should be very cautious and serum concentrations need to be monitored. ACE inhibitors may also cause acute renal failure, particularly in patients with chronic azotaemia and renal arterial stenosis, making it important to monitor serum creatinine and blood urea nitrogen.

Recently b-adrenoceptor antagonists have shown a clear reduction in mortality in patients stabilized with compensated heart failure, in particular bisoprolol in the CIBIS-2 trial and metoprolol in the MERIT-HF trial.

Disturbances in heart rate

In the healthy heart cardiac output can be maintained at low rates by increased stroke volume and increased at high rates by enhanced contractility, over an approximate range of 40–180 beats/minute in sinus rhythm. The impact of an arrhythmia will depend upon the ventricular rate, presence or absence of atrial-ventricular dissociation, presence or absence of underlying heart disease, and the current general condition of the patient. Valvular obstruction, myocardial disease and ischaemic heart disease can adversely affect resistance to flow, contractility, and diastolic compliance. All of these will reduce the threshold at which a given tachycardia compromises cardiac output. For example, some prosthetic heart valves such as the ball and cage or Starr-Edwards valves and those with a relatively small annulus are inherently inefficient and result in significant increases in flow resistance, filling pressures, and myocardial work with heart rates above 130 beats/minute. Inadequate central filling pressures and low peripheral vascular resistance should be considered before hypotension is blamed on the occurrence of a new junctional rhythm or atrial fibrillation.

Arrhythmias

Arrhythmias developing in patients in the intensive care unit are commonly the result of metabolic disturbances, hypovolaemia, increased afterload, hypoxaemia, and the use of inotropes. A concerted effort must be made to identify and correct these disturbances before resorting to treatment with antiarrhythmic agents. Indeed, resistance to these agents can be expected until the underlying abnormalities are corrected. Thereafter, the urgency with which arrhythmia needs to be corrected is determined by the severity of the haemodynamic disruption. Obviously, the onset of shock may necessitate emergency electrical cardioversion, while arrhythmias in the absence of haemodynamic disturbance can be managed with appropriate antiarrhythmic drugs or elective electrical cardioversion. As a general rule, ventricular arrhythmias produce greater haemodynamic disturbance than supraventricular rhythms but there are many exceptions. A summary of treatment regimens for both supraventricular and ventricular arrhythmias is shown in Fig. 3.

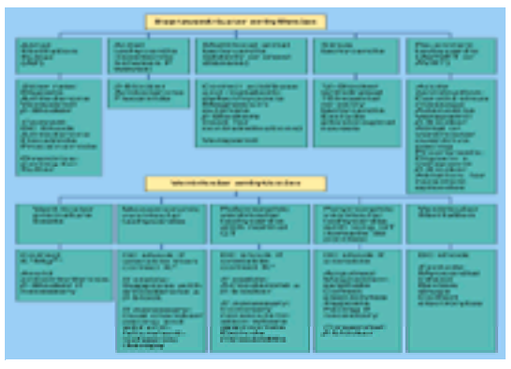


Fig. 3. Algorithms for the treatment of supraventricular and ventricular arrhythmias.

Bradycardias

Symptomatic sinus and junctional or nodal bradycardia will respond to atropine and atrial pacing. Complete block of the atrioventricular node may be poorly tolerated in the presence of impaired left ventricular function because atrial–ventricular synchrony is lost and the ventricular escapte rate is usually less than 45 beats per minute. Ventricular contraction may be impaired if the escape rhythm originates low in the ventricles and does not use the normal His–Purkinje network. The ventricular rate will usually respond to isoprenaline in established complete heart block. Transient atrioventricular block can develop as a result of ischaemia in the atrioventricular node secondary to hypotension. In this, case increasing mean arterial pressure by optimizing central filling pressures and using pressor agents such as adrenaline (epinephrine) may re-establish normal atrioventricular conduction. When cardiac pacing is required a ventricular wire will usually suffice, but cardiac output can be enhanced by temporary dual-chamber pacing which restores atrial–ventricular synchrony.

Tachycardias

Sinus tachycardia is usually an appropriate physiological response but may be due to sinoatrial re-entry or recent withdrawal of b-blockers. Common explanations in the postoperative patient are inadequately controlled pain, infection, bleeding, or hypovolaemia. In the critically ill patient a sinus tachycardia greater than 140 beats/min signifies severe cardiorespiratory distress (e.g. pulmonary oedema, severe pneumonia, tension pneumothorax, pericardial tamponade, massive pulmonary embolism), massive bleeding, or severe infection. The corollary of this is a patient who is relatively well and has a tachycardia of 140 to 150 that is presumed to be due to sinus tachycardia. In this situation the diagnosis is probably incorrect. The more likely diagnosis is atrial flutter with 2:1 atrioventricular node block, atrial tachycardia, or re-entry tachycardia.

The impact of a tachyarrhythmia depends upon the underlying pathological state, the presence or absence of underlying cardiac disease, the maintenance or loss of atrial–ventricular synchrony, and the heart rate. With significant coronary arterial disease, diastolic coronary flow may become inadequate, leading to ischaemia, left ventricular dysfunction, a rising left ventricular end-diastolic pressure, and a drop in mean arterial pressure. Pulmonary oedema or angina commonly develop (despite good systolic function) in elderly patients with a poorly compliant ventricle and rapid atrial fibrillation. Synchronous atrial systole is also important in obstructive valvular

disease, hypertrophic obstructive cardiomyopathy, pulmonary arterial hypertension, and when left ventricular function is impaired.

An incessant tachyarrhythmia can itself induce a cardiomyopathy. Common causes include atrial tachycardia, atrial fibrillation, and atrioventricular re-entry tachycardia.

The urgency of assessment and treatment depends more on the clinical state rather than the origin or rate of the tachycardia. The electrical origin must be defined because it has prognostic and therapeutic implications. Normal initiation and conduction through the heart requires a 1:1 P to QRS relation, a fixed PR interval, and a QRS duration of less than 100 msec (2.5 small squares).

Three discriminatory questions naturally follow.

1. Is the QRS narrow or broad?
2. What is the relation (if present), between P and QRS?
3. Are successive R–R intervals regular or not?

Narrow-complex tachycardias usually originate at or above the atrioventricular node, because a narrow QRS requires conduction to travel normally down the His–Purkinje network. If it is clearly irregular the differential lies between atrial fibrillation, atrial flutter with variable block of the atrioventricular node, multiple atrial extrasystoles, or multifocal atrial tachycardia. The 12-lead electrocardiogram and long sections of rhythm strip should be scrutinized for atrial activity, which is best seen in leads II and VI. If a narrow-complex tachycardia is regular, then the differential includes atrial tachycardia, sinoatrial re-entry tachycardia, atrial flutter, or a re-entry or junctional tachycardia (previously known as ‘supraventricular tachycardia' or ‘SVT') involving the atrioventricular node with or without an atrioventricular accessory pathway. Differentiation of these is beyond the scope of this chapter but relies on the history, position, and polarity of atrial activity, the response to adenosine, comparison with the electrocardiogram in sinus rhythm and, if necessary, electrophysiological studies.

Broad-complex tachycardias should be considered ventricular in origin until proved otherwise, especially if there is pre-existing heart disease. Ventricular tachycardia is normally regular but can vary up to 40 msec (1 small square) between successive R–R intervals if multiple capture and fusion beats are present, a result of a sinus beat completely or incompletely depolarizing the ventricle ahead of the ventricular tachycardia. The main differential of a broad-complex tachycardia includes ventricular tachycardia or a supraventricular tachycardia (in the broad sense), with pre-existing bundle-branch block or rate-related bundle-branch block and pre-excited atrial fibrillation (conducted down an accessory pathway into ventricular muscle). Differentiation depends on the analysis of QRS complexes, looking for evidence of atrial–ventricular dissociation, assessing the response to adenosine, comparison with the electrocardiogram in sinus rhythm and, if necessary, electrophysiological studies.

Patients with atrial tachycardia and monomorphic ventricular tachycardia usually have structural heart disease. Re-entry or junctional tachycardias depend on dual atrioventricular node conduction pathways or accessory pathways. Polymorphic ventricular tachycardia can occur with a normal QT interval or with a long QT interval, the latter association being termed ‘*torsades de pointes*'. The former can occur rarely in the absence of structural heart disease or any provoking factors. Polymorphic ventricular tachycardia can also occur in the period after myocardial infarction where continuing ischaemia is present, and is best treated by revascularization and amiodarone. The long QT variety of polymorphic ventricular tachycardia can be congenital, when it tends to be catecholamine-dependent and responsive to b-blockers, or acquired. The acquired form of *torsades de pointes* is multifactorial and occurs in the presence of hypokalaemia, hypomagnesaemia, bradycardic states, and proarrhythmic drugs, e.g. quinidine, sotalol, non-sedating antihistamines, tricyclic antidepressants, phenothiazines, and cisapride. Macrolide antibiotics and antifungal agents delay the metabolism of the above-mentioned drugs. Treatment entails intravenous magnesium sulphate, correction of hypokalaemia, withdrawal of proarrhythmic drugs, and pacing.

Arrhythmias in the postoperative setting are often precipitated by electrolyte disturbances, hypoxia, hypotension, ischaemia, or inotropic agents. Atrial arrhythmias are commonplace after cardiothoracic surgery, when they may represent a recurrence of a pre-existing problem. The most frequent arrhythmias in the surgical intensive-care patient are atrial fibrillation and atrial flutter. Factors increasing the likelihood of atrial fibrillation include age over 70 years, hypertension, mitral valve disease, cardiac failure, ischaemic heart disease, and chronic obstructive airways disease. At its onset the ventricular rate is often 140 to 180/min and then gradually slows. Up to 60 per cent will self-terminate within 12 h if there is no structural heart disease and the precipitating cause has resolved. The management of a new episode of atrial fibrillation requires consideration of its haemodynamic significance, the risk of thrombus embolization, and the likelihood that it will become persistent. Urgent d.c. cardioversion is required if hypotension, pulmonary oedema or refractory coronary ischaemia are present. If haemodynamic compromise is borderline, then pharmacological cardioversion can be pursued with intravenous amiodarone because of the desirability of preventing recurrences in the acute illness and the tendency of other antiarrhythmic drugs to be proarrhythmic. There must be a clear policy on the length of treatment outside the acute illness. There have been no large-scale, randomized, controlled trials assessing the efficacy of amiodarone acutely to cardiovert atrial fibrillation. Three small studies showed a 70 to 75 per cent cardioversion rate within 3 h. Amiodarone is effective in preventing recurrence once sinus rhythm has been achieved and it will control the ventricular rate in the event of a recurrence. Flecainide has an 80 per cent success rate of cardioverting new-onset atrial fibrillation within 1 h, but it should be avoided in the presence of structural heart disease or hypotension because of its proarrhythmic effects. Intravenous verapamil or b-blockers will slow acutely the ventricular rate and remain useful in the presence of good left ventricular function and coronary ischaemia pending the onset of action of amiodarone. In the acutely ill patient who is pyrexial, hypoxic or thyrotoxic for example, digoxin is unlikely to achieve acute rate control because of the enhanced sympathetic tone, as it works mainly by enhancing vagal tone on the atrioventricular node. Digoxin retains an important role in cardiac failure because of its positive inotropic action. The rate control improves as the cardiac failure is treated, reducing sympathetic drive and atrial stretch.

Echocardiography will demonstrate factors associated with early relapse of atrial fibrillation including impaired left ventricular function, left atrial size greater than 5 cm, mitral valve disease and annular calcification, and hypertensive heart disease. Transthoracic echocardiography may show thrombus but this can only be reliably excluded by transoesophageal echocardiography. Cardioversion within the first 48 h of new-onset atrial fibrillation carries a low thromboembolic risk. If the atrial fibrillation persists longer than this, the usefulness of an elective cardioversion should be considered. This will depend on the expected relapse rate, the haemodynamic significance, and the expected morbidity of long-term anticoagulation. The approximate incidence of stroke in chronic atrial fibrillation is 5 per cent per annum. To minimize the risk of systemic embolization at the time of cardioversion, anticoagulation is required for 4 weeks before (allowing already-formed thrombus to become adherent or lyse) and 4 weeks after cardioversion, at which time atrial electromechanical dissociation will have resolved. If cardioversion is not successful, long-term warfarin will reduce the annual stroke risk by 65 per cent. Digoxin remains useful for rate control in permanent atrial fibrillation. In paroxysmal atrial fibrillation the benefits of antiarrhythmic drugs must clearly outweigh the risks when class 1 drugs (e.g. flecainide, propafenone, disopyramide) and class 3 drugs (e.g. amiodarone and sotalol) are used, because up to a third of patients will withdraw from treatment owing to side-effects. Interestingly both perioperative b-blockade and postoperative magnesium sulphate reduce the incidence of atrial fibrillation, which occurs postoperatively in up to 30 per cent of cardiac surgical patients.

Distributive

The treatment of distributive circulatory failure provides one of the most challenging aspects of critical care. Although there are several causes, including anaphylaxis and neurogenic shock, it is the sepsis syndrome that accounts for most cases. The pathophysiology and principles of management of the sepsis syndrome provide a model that can be applied to most forms of distributive circulatory failure irrespective of cause.

Sepsis syndrome

The sepsis syndrome, and associated multiple organ-system failure, is the most common cause of death in patients on critical care. It represents the host response to endotoxaemia caused by a wide range of micro-organisms. Sepsis syndrome is characterized by low peripheral vascular resistance, which, in the presence of normal cardiac function, is coupled with increased cardiac output (high-output failure) and low filling pressures. However, cardiac function, along with the function of other organ systems, may be greatly impaired. Those patients who do not generate high cardiac outputs in the face of a septic insult generally have a poor prognosis.

Definition

The definition of the sepsis syndrome is based upon the demonstration of signs of infection, shock, and evidence of organ system dysfunction as described in Table 6.

Evidence of host and shock
1. Fever > 38°C or hypothermia to 36.6°C core temperature
2. Tachycardia [1] 90 beats/min
3. Tachypnoea [1] 20 breaths/min or requiring mechanical ventilation
4. Hypotension, systolic blood pressure to 90 mmHg or [2] 40 mmHg fall in systolic pressure, with adequate fluid filling
5. Evidence for localized focus of infection
Together with other clinical evidence of sepsis or end-organ failure
1. Metabolic acidosis, pH or [3] 7.3 accounted for by $Paco_2$ changes, or a base deficit [4] 5
2. Arterial hypoxaemia with Pa_{O_2} to Fi_{O_2}
3. Pao: Fio ratio [5] 33 (Pao to fio or 250 (Pao to mmHg)
4. Increased plasma lactate
5. Oliguria to 0.5 ml/kg per hour for 1 h
6. Coagulation defect with: prothrombin time 1.5 times control or PTT 1.2 times control
7. Thrombocytopenia to 100000 or more than 50% fall within 24 h
8. Acute deterioration in mental status
9. Any other sign of organ-system failure (see Table 7)

After: Bone *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; 101(6): 1644-55.

Table 6 Clinical features of the sepsis syndrome

Pathogenesis and pathophysiology

Sepsis syndrome is the pathophysiological responses to systemic infection or endotoxaemia. The clinical syndrome can be produced by a wide range of organisms including Gram-positive and Gram-negative bacteria, protozoa, viruses, and fungi. In a significant proportion of cases an infective organism is never isolated. Although endotoxaemia is not consistently present in early sepsis, it is usually present in late, severe sepsis syndrome with multiple organ-system failure and is associated with a poor prognosis. Systemic endotoxaemia may also result from translocation of endotoxin and bacteria from the intestinal lumen into the circulation, and may explain the development of sepsis syndrome in patients with trauma and burns (Fig. 4).

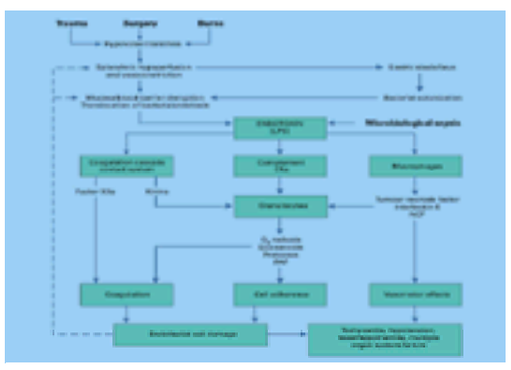


Fig. 4. Possible interactions leading to the sepsis syndrome.

Spill-over of endotoxin into the circulation is known to occur in patients with inflammatory bowel disease such as Crohn's disease and in those with obstructive jaundice. Bile salts may have a beneficial effect by binding endotoxin, and therefore their absence may predispose towards endotoxaemia. The presence of circulating micro-organisms and endotoxins appears to trigger an immunological and inflammatory cascade, with complement activation and the release of a number of host-derived mediators, including tumour necrosis factor, interleukins, and myocardial depressant factor. These mediators are probably responsible for the systemic vasodilation, hypotension, and multiple organ-system failure that are often seen.

The presence of multiple organ-system failure is usually only too apparent to the clinician, although formal diagnostic criteria can be defined and are summarized in Table 7. Not surprisingly, the greater the number of organ systems involved the worse the prognosis.

A. *Circulatory failure*
1. Bradycardia (rate to 50 beats/min)
2. Hypotension (mean blood pressure to 50 mmHg
3. Ventricular tachycardia or fibrillation
4. Metabolic acidosis (pH to 7.2)
B. *Respiratory failure*
1. Respiratory rate to 5 or [2] 40 breaths/min
2. Hypercapnoea ($Paco_2$ [2] 6.7 kPa or 50 mmHg)
3. Hypoxaemia ($(A-a)Po_2$ gradient [2] 50 kPa or 350 mmHg)
C. *Acute renal failure*
1. Urine output to 400 ml/24 h or to 150 ml/8 h
2. Rising serum creatinine [2] 150 mmol/l
D. *Haematological failure*
1. Leucopenia (white blood cell count to 1000 cells/mm^3)
2. Thrombocytopenia (platelets to 20000/mm^3)
3. Anaemia (haematocrit to 20%)
4. Evidence of disseminated intravascular coagulation
E. *Hepatic failure*
1. Prothrombin time increased > twice control)
2. Rising hepatic enzymes
F. *Gastrointestinal failure*
1. Ileus
2. Gastroparesis
3. Haemorrhage
G. *Neurological failure*
1. Depressed consciousness (Glasgow coma score to 6)
2. Seizure disorder

Table 7 Criteria for the presence of organ-system failure

Colonization of the upper gastrointestinal tract Colonization of the upper gastrointestinal tract, particularly with Gram-negative organisms, may be a significant aetiological factor in sepsis. It may also play a part in the perpetuation of the sepsis syndrome. Colonization of the stomach and upper small bowel in the critically ill patient provides a major source of endogenous infecting organisms. Colonization of the upper gastrointestinal tract is promoted by parenteral administration of antibiotics, which are excreted in saliva, bile, and intestinal mucus, and which suppress the endogenous intestinal anaerobic flora. Other factors that promote colonization are listed in Table 8. Colonization of the gastrointestinal tract provides a reservoir of pathogens that not only increase the enteric endotoxin pool but also increase the risk of nosocomial lung infection developing after the aspiration of gastrointestinal contents.

1. Parenteral antibiotics
2. Suppression of gastric acid secretion
3. Gastroparesis and ileus
4. Level of invasive instrumentation
5. Absence of enteral feeding
6. Extremes of age

Table 8 Factors associated with upper gastrointestinal tract colonization

Selective decontamination of the digestive tract Selective decontamination of the digestive tract with orally administered, non-absorbable antimicrobial agents has been recommended as a method of reducing the colonization of the upper gastrointestinal tract in the critically ill patient. This reduces nosocomial lung infections, although

there has not been a consistent reduction in mortality. The procedure appears to be well tolerated and resistant strains of colonizing micro-organisms do not commonly develop. Selective decontamination may, therefore, be recommended in patients with severe multiple injuries and in those in whom persistent endotoxaemia is suspected. Several ‘recipes' have been proposed, including the combination of oral polymyxin, amphoteracin, and gentamicin with a systemic b-lactam such as ceftazidime. This combination is administered as a paste for the mouth and as a mixture to pass through the digestive tract. Polymyxin is particularly effective in reducing amounts of faecal endotoxin.

Oxygen transport considerations in the sepsis syndrome Sepsis syndrome is characterized by defective oxygen utilization in the face of a high cardiac output. The reduced oxygen consumption may be associated with raised blood lactate, suggesting anaerobic metabolism consistent with an intracellular defect in oxygen utilization or microvascular shunting that creates areas of tissue hypoxaemia. It is probable that both elements exist and contribute to the impaired oxygen uptake by the tissues. When tissue hypoxia becomes more severe, as evidenced by a fall in venous oxygen saturation ($S\mathbf{v}O_2$) below 60 per cent, hyperlactaemia can be expected. Tissue hypoxia, however, as evidenced by reduced Svo_2, may be present even in the absence of elevated blood lactate. A normal or high Svo_2 does not exclude tissue hypoxia but merely reflects the degree of reduced tissue oxygen utilization and high-output state.

In health, oxygen consumption is largely independent of delivery until oxygen delivery falls below about half the normal rate, that is, below 7 ml/kg. In the sepsis syndrome, not only is oxygen consumption reduced, but there appears to be delivery-dependent oxygen consumption at all levels of delivery. This claim has been challenged on the grounds that the formulae for calculating oxygen consumption and delivery, using the Fick principle, both contain the common elements of cardiac output and the oxygen content of arterial blood, and that they are therefore inevitably correlated. Indeed, when oxygen uptake is derived from analysis of exhaled gases, no such delivery dependence of oxygen consumption is found. Irrespective of how this controversy is resolved, the underlying principles of using combinations of inotropes and vasopressors to optimize cardiac output and oxygen delivery still pertain, although it remains to be demonstrated conclusively that such methods improve survival. ‘Goal-orientated therapy', in which attempts are made to overdrive the circulation (Table 9), were originally applied to high-risk postoperative patients and victims of trauma but have failed to yield benefit in sepsis. Pathophysiological hypotheses for sepsis syndrome are moving away from possible microvascular phenomena towards identifying metabolic or bioenergetic defects within the cells themselves.



Table 9 Management overview of organ-system support

Management

Management is aimed at organ-system support, the suppression or amelioration of the toxic effects of the septic process, and, most important, the identification and eradication of the septic focus. Some forms of organ-system support, such as mechanical ventilation and haemofiltration, are readily achievable, while the support of the brain, circulation, and gastrointestinal tract has to be approached indirectly by trying to attain adequate oxygen delivery.

Cardiovascular dysfunction Cardiovascular dysfunction in sepsis has rightly attracted a great deal of attention in the past two decades. Much of the interest has focused on the hyperdynamic, high cardiac-output state associated with microvascular shunting. Two peripheral vascular phenomena predominate, both resulting from inflammatory cytokine production. Loss of intravascular volume through leaking capillary beds and arterial vasodilatation induced by nitric oxide produced by endothelial cells. Management is initially directed at fluid resuscitation, usually with colloidal solutions, followed by the judicious addition of pressor agents such as noradrenaline (norepi-nephrine). In many patients there is initially a gratifying clinical response to fluid resuscitation and pressor agents. Thereafter, the patient will either improve and recover or show an increasing requirement for pressor agents, usually associated with clinical signs of poor peripheral perfusion, falling cardiac output, and increasing base deficit. At this point, inotropic agents can be introduced to support the failing myocardium. Sadly, the development of overt myocardial failure usually signals the patient's ultimate death.

Intrinsic myocardial dysfunction in sepsis has been recognized for some years. Superficially, the heart appears to function well in the early phases of sepsis. However, the true state of myocardial dysfunction is masked by the marked reduction in afterload. Only by examining cardiac function in terms of the relation of stroke volume to ventricular filling (pulmonary capillary wedge pressure) can evidence of myocardial dysfunction be found (Fig. 2). Some patients respond to myocardial failure by diastolic dilatation, exploiting the Frank–Starling relation. Those patients who are unable to dilate their ventricle appear to have a worse prognosis.

The chief mechanism for myocardial dysfunction appears to involve circulating and local mediators produced by many cell lines such as bacteria, polymorphonuclear leucocytes, macrophages, and endothelial cells. A myocardial depressant factor in septic shock has been suspected since the 1970s. Today, we believe that not one, but several, separate and interactive factors such as endotoxin, tumour necrosis factor, interleukins 1 and 6, platelet-activating factor, kinins, and nitric oxide probably combine to impair myocardial activity.

In addition to organ-system support, specific interventions aim at the eradication of the septic focus or at inhibiting the inflammatory cascade. Appropriate antibiotics must be selected, but these will eventually prove ineffective unless surgical drainage of large foci of infection is undertaken.

Identification and eradication of the septic focus An aggressive approach to clinical, radiographic, sonographic, and microbiological surveillance is required if the focus of infection is to be identified and eradicated. The investigations should be directed along the lines suggested by a thorough understanding of the clinical problems. Recognizing the limitations of non-invasive investigations should encourage early diagnostic laparotomy. The clinical features of intra-abdominal sepsis may be difficult to elucidate in the sedated patient receiving potent analgesics.

Immunotherapy Many attempts have been made to interrupt the inflammatory cascade of sepsis. Cortiocosteroids appear to be ineffective, as do specific inhibitors of leucotriene and prostenoid production.

Specific monoclonal and polyclonal antibodies against endotoxin, interleukin 6, interferon-g, and tumour necrosis factor-a, soluble receptors for tumour necrosis factor, interleukin1-receptor antagonists, specific antagonists of platelet-activating factor have all failed to realize significant clinical benefit. Recently, attempts to block the elaboration of nitric oxide by the false precursor mono-methyl arginine (LNMMA) demonstrated elevation in blood pressure without improving survival. Why such carefully considered strategies have failed is uncertain, but it apparent that the animal models upon which such strategies were based may not have been comparable to human sepsis. Furthermore, the interactions of the cytokine cascades are much more complex than was at first imagined. Although the numbers of patients recruited were considered adequate, the heterogeneous nature of the population of patients with sepsis probably necessitated even larger studies. Finally, these large, multicentre studies gave little or no consideration to individual host response, assuming all patients broadly followed the models dictated by animal studies.

Genetic characterization of host response

There is wide recognition among clinicians that septic patients can have quite different responses to very similar microbiological insults. The host response to sepsis may be mild, while some patients suffer a severe systemic response with septic shock and multiple organ failure culminating in death. The final pathway explaining the diversity of clinical response has been attributed to cytokines and related molecules that act as essential mediators within the inflammatory cascade. The underlying genetic coding for an increasing number of cytokines has been described. The coding for each gene and therefore an individual cytokine may be varied by adjacent promoter and suppressor regions. Such regulating regions vary from patient to patient and are referred to as gene polymorphisms. Recent evidence has demonstrated genetic polymorphisms for a number of key cytokines in infection susceptibility, immune response, and outcome in patients with surgical sepsis and meningococcal sepsis. So far, certain polymorphisms for tumour necrosis factor and interleukins 1, 6, and 10 have been identified as possibly relevant, and the number is likely to increase. Future work examining host response in terms of the host genetics may offer a method of tailoring therapy to meet the needs of an individual patient.

Physiological measurement in circulatory failure

Flow-directed pulmonary arterial catheter

The flow-directed pulmonary arterial catheter is a multilumen tube used for the catheterization of the right-sided circulation. It incorporates a balloon, of about 1.5-ml capacity, at the tip: this facilitates the passage of the catheter into the pulmonary artery as well as allowing the determination of left-sided filling pressures. A thermistor at the tip enables the determination of cardiac output by the thermodilution technique. Newer catheters include the capacity for sequential atrioventricular pacing as well as continuous monitoring of cardiac output and mixed venous oxygen saturation. Continuous cardiac-output devices use a low-temperature heating coil to warm the blood in a random fashion while the distal thermistor probe detects the flow-dependent changes in temperature. Such catheters avoid the inconvenience of intermittent injection of cool saline or dextrose. Some pulmonary arterial catheters incorporate a rapid-response thermistor, facilitating the estimation of the right ventricular ejection fraction and right ventricular end-diastolic volume. At the proximal, hub end of a flow-directed pulmonary arterial catheter there are several ports and connections: a distal and proximal port, a balloon inflation port, a thermistor connection, and, in some, a fibreoptic connection.

There has been much controversy about the risk compared with the benefit of the pulmonary arterial catheter, fuelled partially by evidence from the United States of America that has suggested an excess mortality associated with their use. It is therefore incumbent on each clinician to determine the true need for its insertion. Indications for pulmonary arterial catheterization are constantly evolving; current indications are summarized in Table 10. To minimize risks associated with the use of these catheters, intensive care units should formulate guidelines and ensure a high quality of training in their insertion. To maximize benefit, the full range of data should be acquired and the best interpretation of the information made by careful training and education of the clinicians.

1. Refractory hypotension
2. Cardiogenic circulatory failure
3. Cardiovascular monitoring and therapeutic guide in the sepsis syndrome
4. Acute right ventricular myocardial infarct
5. Assessment of cardiac valvular and septal defects
6. Assessment of therapy in severe chronic congestive heart failure
7. Electrophysiology studies and cardiac pacing
8. Preoperative staging of high-risk patients

Table 10 Indications for pulmonary artery catheter insertion

The utility of data derived from pulmonary arterial catheterization rests upon the fact that there are limits to the reliability and accuracy of clinical methods of determining pulmonary arterial pressure, pulmonary vascular resistance, left atrial pressure, peripheral vascular resistance, and cardiac output. Several studies have revealed that even when radiographic and clinical criteria are critically evaluated the presence or absence of left ventricular failure cannot always be reliably predicted. Another test of the utility of pulmonary arterial catheterization is whether or not meaningful alterations in therapy result from the information derived from the flotation catheter. It has been estimated that in between one- and two-thirds of patients investigated by this means, changes in treatment are made as a result of the information obtained.

Given the large amount of potentially valuable information obtainable from the catheter, it would be tempting to use it in most critically ill patients. However, the incidence of serious complications is high enough to mitigate against its routine use. Some of the complications related to the pulmonary arterial catheter are listed in Table 11.

1. Overwedging with pulmonary infarction
2. Arrhythmias
3. Air embolus
4. Endocarditis
5. Sepsis
6. Thrombocytopenia
7. Catheter knotting
8. Pneumothorax
9. Tricuspid valve avulsion
10. Pulmonary arterial branch rupture and haemorrhage

Table 11 Some complications associated with the use of pulmonary arterial catheters

The overall incidence of complications is difficult to assess. However, many of the serious complications can be avoided by adhering to standard procedures. Checking the chest radiograph after inserting the catheter should detect pneumothorax and over-distal placement of the catheter tip (overwedging). Furthermore, by ensuring that the pulmonary arterial pressure trace returns after each deflation of the balloon, lung infarction should be prevented. By adopting strict aseptic insertion techniques, using a plastic sheath around the external portion of the catheter, and by removing the catheter within 72 h, the risk of sepsis and endocarditis is reduced.

Insertion

Correct and skilful insertion technique cannot be learnt from a textbook but requires careful instruction by an experienced clinician. However, a few recommendations and comments can be made at this juncture. Full aseptic technique requires the wearing of surgical gown, cap, mask, and gloves, and preparation of as much of the equipment as possible before touching the patient.

This includes attaching three-way taps, flushing all channels with heparinized saline, checking the integrity of the balloon, and completing *in vitro* calibration of the pulmonary arterial catheter oximeter (if used). The use of the internal jugular route, particularly the right internal jugular, facilitates placement. Prior insertion of a pulmonary arterial catheter sheath (8–8.5 Fr) makes later replacement with either another pulmonary arterial catheter or a triple-lumen catheter very much more convenient. As the catheter is introduced its natural curvature should be maintained such that it will be directed through the tricuspid valve and up into the right ventricular outflow tract (with the balloon inflated). The catheter should be advanced slowly, about 2 cm every 2 s, watching its distal portpressure trace. As each heart chamber is entered (right atrial and right ventricle), the pressure should be noted. Once in the pulmonary artery the catheter should be advanced until it wedges, the pul-monary arterial wedge pressure (or pulmonary arterial occlusion pressure) being monitored. When the balloon is deflated the pulmonary arterial trace should reappear. Subsequent inflation of the balloon should initially continue to give a pulmonary arterial trace: as the catheter uncoils and advances, it will again wedge; this should be the ideal catheter position, which a portable chest radiograph will confirm position. Right atrial and pulmonary arterial wedge pressures should be recorded at the top of the ‘a' wave at end-expiration.

Measurement of pressures

Right atrial waveform, pulmonary arterial systolic, diastolic, and mean pressures, and pulmonary arterial wedge waveform should be measured. In the spontaneously breathing patient, the pulmonary arterial wedge pressure reflects left atrial pressure. The relation of pulmonary arterial wedge pressure to left ventricular end-diastolic pressure is more complex, however. Pulmonary arterial wedge pressure provides an accurate indication of left ventricular end-diastolic pressure, provided that the latter is low and there is no mitral valve disease. At high left ventricular end-diastolic pressures (30–35 mmHg), the pulmonary arterial wedge pressure is usually 5 to 10 mmHg lower. The pulmonary arterial diastolic pressure is generally 1 to 2 mmHg greater than the pulmonary arterial wedge and can be used as a crude index of left

atrial pressure, again assuming that there is no mitral valve or pulmonary disease. The pulmonary arterial wedge pressure is not usually higher than the pulmonary arterial diastolic pressure and should never be greater than the mean pulmonary arterial pressure. When this occurs (assuming no mitral regurgit-ation), the discrepancy is usually caused by overwedging of the catheter or by improper measurement of the pulmonary diastolic pressure.

Measurement of cardiac output

This generally requires the averaging of three reasonably close, sequential measurements. Ideally, ice-cooled saline (at 9°C) should be injected at the same point during the respiratory cycle, at end-expiration. This may be difficult to achieve and in practice it is probably satisfactory to adopt random injection timing and derive an average value for cardiac output. Cardiac index is obtained by dividing the cardiac output by the body surface area: the normal range is 2.8 to 3.6 l/min per m^2.

Mixed venous saturation

Blood sampling from the pulmonary artery allows the determination of $S\text{vo}_2$, venous oxygen tension ($P\text{vo}_2$), and venous oxygen content ($C\text{vo}_2$). This is obtained through the distal port of the catheter when it is in position in the pulmonary outflow tract. It is important that the specimen be aspirated slowly (3 ml/min) to avoid obtaining an arterialized specimen. The laboratory determination of $S\text{vo}_2$ is most accurately done with an oximeter and should not be derived from the $P\text{vo}_2$. Pulmonary arterial catheters incorporating direct, continuous oximetry provide a unique facility to monitor $S\text{vo}_2$.

Derived haemodynamic values

Haemodynamic values can be derived using various formulas and a hand-held calculator or microcomputer, or can be calculated directly by newer generations of cardiac-output computers or physiological monitors. The variables needed to calculate derived values are listed in Table 12. These derived haemodynamic values can be classified in two major categories, pump performance and oxygen transport and utilization, and are shown in Table 13.

1. Blood pressure (BP)—systolic, diastolic and mean. Most modern physiological monitoring systems provide accurate mean values. Pulsus alternans, however, can be underestimated as: MAP = diastolic BP + (⅓ systolic BP − diastolic BP)
2. Body surface area (BSA) from appropriate nomogram
3. Heart rate
4. Haemoglobin concentration
5. Arterial blood gases and pH, and barometric pressure (Pb)

Table 12 Additional variables required to calculate derived haemodynamic indices

1. *Data reflecting pump performance*
(a) Cardiac index (CI = CO/BSA)
(b) Stroke volume (SV = CO/HR)
(c) Stroke index (SI = CI/HR)
(d) Systemic vascular resistance (SVR)
(e) Pulmonary vascular resistance (PVR)
(f) Left ventricular stroke work index (LVSWI)
(g) Right ventricular stroke work index (RVSWI)
2. *Data reflecting oxygen transport and utilization*
(a) Oxygen content—arterial (Ca_{O_2}), venous (Cv_{O_2}) and pulmonary capillary (Cc_{O_2})
(b) Arteriovenous oxygen content difference ($a\text{-}vD_{O_2}$)
(c) Oxygen consumption ($\dot{V}_{O_2}$) or oxygen uptake
(d) Oxygen delivery ($\dot{D}_{O_2}$) or oxygen transport or oxygen availability
(e) Oxygen extraction ratio or oxygen utilization ratio ($\dot{V}_{O_2}/\dot{D}_{O_2}$)
(f) Alveolar–arterial oxygen difference (AaD_{O_2})
(g) Shunt fraction ($\dot{Q}s/\dot{Q}t$) or venous admixture ($\dot{Q}vs/\dot{Q}t$)

BSA, body surface area; CO, cardiac output; HR, heart rate.

Table 13 Categories of haemodynamic values derived from pulmonary arterial catheter data

Data reflecting pump performance

Cardiac index (cardiac output divided by body surface area) and stroke index (cardiac index divided by heart rate) relate cardiac output to the patient's body surface area. Systemic vascular resistance and pulmonary vascular resistance represent the peripheral component of afterload (that is, impedance). Their derived calculation is based on a rearrangement of Poiseuille's law (the hydraulic resistance equation):

$$R = \frac{P_i - P_o}{Q}$$

where R = resistance, P_i = pressure at inflow, P_o = pressure at outflow, Q = blood flow.

From this:

$$\text{SVR} = \frac{(\text{MAP} - \text{CVP})80}{\text{CO}}\ \text{dyne/s.cm}^5$$

and

$$\text{PVR} = \frac{(\text{PA} - \text{PAWP})80}{\text{CO}}\ \text{dyne/s.cm}^5$$

where MAP = mean (systemic) arterial pressure, CVP = central venous pressure, SVR = systemic vascular resistance, CO = cardiac output, PA = mean pulmonary arterial pressure, PAWP = pulmonary arterial wedge pressure, and PVR = pulmonary vascular resistance.

Systemic vascular resistance is increased in low-flow states, such as cardiogenic and hypovolaemic shock, secondarily to endogenous or exogenous vasoconstrictors, and in systemic hypertension. It is decreased in states associated with high cardiac output, including trauma, sepsis, burns, liver disease, and anaemia, and in Addison's disease.

Pulmonary vascular resistance tends to be elevated in heart failure, pulmonary embolus, chronic obstructive pulmonary disease, adult respiratory distress syndrome and mitral valve disease, and decreased in vasodilated states and in hypovolaemia.

Left and right ventricular stroke-work indices are derived values reflecting cardiac contractility. They are measures of the external work of the ventricle during each

contraction and are represented as:

$$\mathrm{LVSWI} = \mathrm{SI} \times \mathrm{MAP} \times 13.6$$

and

$$\mathrm{RVSWI} = \mathrm{SI} \times \mathrm{PA} \times 13.6$$

where MAP = mean (systemic) arterial pressure, PA = mean pulmonary arterial pressure, and SI = stroke index (CI/HR).

The left ventricular stroke-work index is elevated in some types of hypertension, aortic stenosis, and stress states (trauma, burns, and sepsis), and decreased in hypovolaemic, cardiogenic, and late septic shock. The right ventricular stroke-work index is usually elevated in patients with pulmonary hypertension and valvular heart disease and decreased in hypovolaemic and cardiogenic shock. Identical stroke-work indices can be obtained by doubling the stroke index and halving the pressure (volume work) or by halving the stroke index and doubling the pressure (pressure work). However, myocardial oxygen consumption is considerably greater when the heart is performing pressure work than when it is performing volume work. The extent of myocardial oxygen demand can be approximated using the rate–pressure product (heart rate × systolic blood pressure). Values above 12 000 are indicative of significantly increased myocardial work and increased myocardial oxygen demands.

Data reflecting oxygen transport and utilization

Oxygen delivery Oxygen delivery is the amount of oxygen leaving the heart to be delivered to the tissues in ml/min (normal range 640–1200 ml/min) or, expressed as an index based on body surface area, the oxygen delivery index (normal range 500–720 ml/min per m^2). Oxygen delivery indicates the integrity of the interactions between the cardiac pump, the oxygenating function of the lungs, and the carrying capacity of red blood cells. It does not provide an index of what the tissues do with the oxygen once it is delivered (oxygen uptake or extraction). Oxygen delivery is calculated from the product of cardiac output and the oxygen content of the blood (Cao_2 ml/dl of blood).

$$\text{Oxygen delivery} = \text{cardiac output} \times Cao_2 \times 10 \text{ ml/min}$$

$$\text{Oxygen delivery index} = \text{cardiac index} \times Cao_2 \times 10 \text{ ml/min.m}^2$$

Oxygen consumption index Oxygen consumption is the amount of oxygen (in ml) consumed by the body's tissues per minute. Normal values range from 190 to 250 ml/min and the oxygen consumption index (corrected for body surface area) ranges from 100 to 160 ml/min per m^2. These variables are calculated as:

$$\text{Oxygen consumption} = \mathrm{CO} \times (Cao_2 - Cvo_2) \times 10 \text{ ml/min.m}^2$$

$$\text{Oxygen consumption index} = \mathrm{CI} \times (Cao_2 - Cvo_2) \times 10 \text{ ml/min.m}^2$$

In normal individuals, resting oxygen consumption is broadly unchanged over a wide range of values of oxygen, the oxygen extraction ratio varying to maintain a stable consumption. The normal physiological response to a fall in oxygen delivery is to increase oxygen extraction. Oxygen uptake, derived from analysis of inspired and expired respiratory gases, during steady-state conditions will produce values equivalent to oxygen consumption.

Arteriovenous oxygen-content difference The arteriovenous oxygen-content difference is calculated as the difference between arterial and mixed venous oxygen content (Cao_2 - Cvo_2). It is increased in conditions of increased oxygen extraction (decreased cardiac output, anaemia, and increased oxygen consumption) and may be decreased in high cardiac-output states, in states associated with significant atrioventricular shunting, and with conditions in which impaired oxygen utilization occurs (sepsis). The oxygen extraction ratio is calculated as the arteriovenous oxygen content difference divided by the arterial oxygen content (Cao_2 - Cvo_2) and reflects the fraction of delivered oxygen that is consumed: the normal range is 22 to 30 per cent. It varies in a similar way to the arteriovenous oxygen-content difference.

The utility of oxygen transport measurement

A characteristic feature of cardiogenic shock is arterial hypoxaemia, which, in association with the reduced cardiac output, causes profound falls in oxygen delivery and a subsequent rise in blood lactate. The administration of oxygen may improve haemodynamic measures and produce a fall in lactate concentrations. The reduced oxygen delivery and hyperlactataemia are associated with an increase in oxygen extraction ratio and a marked reduction in venous oxygen saturation. The profound fall in venous oxygen saturation is of particular importance if the overall mixed venous saturation is less than 50 per cent, indicating that some tissues are significantly hypoxaemic. This results in anaerobic metabolism and increased lactic acid production. The treatment of cardiogenic circulatory failure should therefore be aimed at increasing cardiac output and at increasing venous oxygen saturation. The outcome of cardiogenic shock following acute myocardial infarction is considerably better in those patients whose cardiac index is greater than 2.2 l/min per m^2. Attempts should therefore be made to achieve cardiac indices greater than this using a combination of preload optimization, inotropes, and vasodilators, and then to maximize oxygen on-loading by supplemental oxygen and, if necessary, intubation and mechanical ventilation.

Alveolar–arterial oxygen gradient

The alveolar–arterial oxygen gradient ($AaDO_2$) is a relatively sensitive but non-specific index of cardiopulmonary function. Gradients are normal when hypoxaemia is secondary to high altitude or alveolar hypoventilation (< 2.6–3.3 kPa or 20–25 mmHg). In the presence of a normal cardiac output, the gradient provides a rough index of venous admixture and has the added advantage that a mixed venous blood sample is not required for its determination. However, it does require the FIO_2 to be accurately known. Determination of the gradient may offer predictive information—very high gradients are associated with severe cardiopulmonary disease—as well as offering a means to monitor therapy. It is calculated as:

$$\mathrm{A-aDO_2} = PAo_2 - Pao_2$$

$$PAo_2 = (PB - 47)(\mathrm{Fio_2}) - (1.25 \times \mathrm{PaCo_2})$$

where PAo_2 = partial pressure of oxygen in alveolar air.

Shunt fraction and venous admixture

The shunt fraction (Qs/Qt) measures the fraction of total blood flow that is not oxygenated during its passage through the lungs, and is calculated when the FIO_2 = 100 per cent. When it is calculated at an FIO_2 of below 100 per cent, it is more correctly referred to as the venous admixture (Qva/Qt). Pulmonary conditions that cause shunting, such as pneumonia, atelectasis, and secretions, have primary effects on the shunt fraction. In patients who have diseased lungs, changes in shunt may simply reflect changes in pulmonary blood flow rather than changes in the intrinsic disease process. The effect of blood flow on shunt depends to some degree whether the lungs are normal, diffusely diseased, or have lobar or regional abnormalities. Venous admixture or shunt varies directly with cardiac output in patients with normal or diffusely abnormal lung (the higher the cardiac output, the greater the shunt) and inversely with cardiac output in patients with unilateral lung abnormalities. The resultant effect of cardiac output and shunt on the Pao_2 is complex and depends on whether the alteration in pulmonary shunt predominates over the change in mixed venous blood oxygen content.

The shunt equation is a mathematical derivation, partly based on the Fick equation. The derivation of the shunt equation involves making certain physiological assumptions that are in part theoretical. Since oxygen consumption represents the product of blood flow cardiac output times the atrioventricular oxygen content difference (Cao_2 - Cvo_2), the Fick equation, can be expressed as:

$$Vo_2 = Qt \times (Cao_2 - Cvo_2)$$

The shunt equation is expressed as the ratio of oxygen content differences between pulmonary capillary blood (Cco_2), an assumed quantity based on the knowledge of the partial pressure of oxygen in the alveolus PAo_2 and the oxygen content of arterial and mixed venous blood, respectively. It is represented as:

$$\frac{Q_s}{Q_s} = \frac{(Cco_2 - Cao_2)}{(Cco_2 - Cvo_2)}$$

In order to calculate the shunt fraction one must carry out the following steps.

1. Obtain mixed venous blood through the distal line of the pulmonary arterial catheter together with a direct pulmonary venous saturation if a fibreoptic catheter has been used.
2. Measure arterial blood gases.
3. Calculate Cao_2 = (haemoglobin × 1.34 × Sao_2) + (Pao_2 × 0.022).
4. Calculate Cvo_2 = (haemoglobin × 1.34 × Svo_2) + (Pvo_2 × 0.022).
5. Calculate PAo_2 = (PB - 47) FIO_2 - ($PaCo_2$ × 1.25).
6. Calculate Cco_2 using PAo_2 from step 5, i.e., Cco_2 = (haemoglobin × 1.34 × Sco_2) + (PAo_2 × 0.022) (assume Sco_2 = 100 per cent, a reasonable assumption providing the patient is on supplemental oxygen).

(NB: The oxygen solubility factor is quoted as 0.022 ml/kPa, the equivalent factor for mmHg is 0.003.)

The normal shunt fraction is less than 10 per cent. Shunts below 10 per cent are rarely seen in patients receiving positive-pressure ventilation. Calculated true shunts greater than 30 per cent are considered incompatible with prolonged spontaneous ventilation. Below 20 per cent the weaning process should be considered if the cardiac status is good.

Lactic acidosis

Lactic acidosis is a pathological state characterized by persistent elevation of the serum lactate concentration together with significant acidaemia. Lactic acidosis should always be suspected in a patient with metabolic acidosis and an increased anion gap that cannot be explained by uraemia or ketonaemia. The onset of clinical manifestations may resemble diabetic ketoacidosis, with sudden malaise, weakness, anorexia, nausea, and vomiting. An early sign of lactic acidosis is hyperpnoea or abdominal pain. In contrast to diabetic ketoacidosis, there is no polyuria, polydipsia, or acetone odour, and the fluid deficit is usually less marked.

Lactic acidosis may occur during anaerobic conditions when oxidative metabolism via the Krebs' cycle or gluconeogenesis is prevented. It may be associated with impaired end-organ function, such as liver or kidney disease, which greatly decreases the metabolic capacity for lactate. Lactate is metabolized either via oxidative metabolism (Krebs' cycle) to CO_2 and H_2O, or by gluconeogenesis to glucose. Both of these pathways depend on intact aerobic metabolism and need intact mitochondrial function, a favourable redox state, and adequate ATP. Gluconeogenesis from lactate occurs in the liver and kidneys, and provides a continuous supply of glucose that would be otherwise wasted to tissues.

Increased lactate can exist without acidosis as is seen when lactate production exceeds metabolic capacity, even with intact oxidation metabolism and good end-organ function. When this occurs (for example in hyperventilation and stress), pyruvate is converted to lactate. However, the elevation in lactate is generally modest (5 mmol/l) and acidosis, as noted above, is usually absent.

Types

Type A lactic acidosis is associated with poor tissue perfusion and hypoxaemia, and is the most common type. The initial hydrolysis of a relatively large amount of ATP would initially release more H^+ than lactate.

Type B lactic acidosis not associated with a decreased oxygen supply and lactic acidosis is due to increased glycolysis or decreased gluconeogenesis. Hydrogen ion and lactate production tend to be equimolar. Type B lactic acidosis has been subdivided into three subtypes: type B_1 is associated with disorders such as diabetes mellitus, renal and hepatic disease, infection, and leukaemia; type B_2 is due to drugs, chemicals, and toxins such as phenformin, ethanol, and methanol; type B_3 covers congenital forms of lactic acidosis such as Von Gierke disease.

Normal lactate concentrations are 0.5 to 1.6 mmol/l for arterial blood and 0.5 to 2 mmol/l for venous blood. It is essential that a tourniquet should not be not applied during venesection. Most patients with lactic acidosis have an anion gap that averages 22 to 27 mmol/l. The increase in the anion gap is usually greater than the decrease in HCO_3^-, in contrast to diabetic ketoacidosis, in which the increase in the anion gap is identical to the decrease in HCO_3^-.

Clinical examples

Tissue hypoxia

Lactic acidosis is usually associated with tissue hypoxia, being common when cardiac output is greatly impaired but rare in patients with uncomplicated anaemia or hypoxaemia. As cardiac output decreases to 30 per cent of normal the major compensatory mechanism of increased oxygen extraction is insufficient to prevent tissue hypoxaemia. In the presence of arterial hypoxaemia and even severe anaemia, a tripling of cardiac output and a tripling of oxygen extraction results in a ninefold increase in oxygen delivery and ensures adequate oxygen delivery. Chronic lactic acidosis may develop in patients with severe congestive heart failure, and tissue hypoxia may exist in the presence of high cardiac output if blood distribution is disturbed, as occurs in the sepsis syndrome. The resolution of such a lactic acidosis is then considered a marker of efficacy of therapy, provided liver function is unimpaired.

Seizures

Lactic acidosis may accompany *grand mal* seizures and is generally self-limiting. Specific treatment of the acidosis is unnecessary, although resolution of lactic acidosis may be slower in patients with pre-existing liver disease.

Diabetes mellitus

Lactic acidosis used to be observed in patients with diabetes mellitus following phenformin administration; it is now rare since this side-effect was recognized. Lactic acidosis occasionally develops in patients with diabetic ketoacidosis and extreme volume depletion. It may also be evident in liver disease: basal lactate metabolism is generally normal in cirrhotics but the reserve is much decreased. Rising lactate in acute fulminant hepatitis is associated with a poor prognosis.

Hypoglycaemia

Lactic acidosis in association with hypoglycaemia is generally confined to infants and children, but has been noted occasionally in adults with either hepatic or renal disease.

Malignancy

Lactic acid concentrations are occasionally increased in patients with acute leukaemia, lymphomas, and intra-abdominal neoplasms. This has been attributed to the overproduction of lactate by the tumour and appears to be related to the tumour burden.

Asthmatics

Twenty per cent of asthmatics admitted to hospital may exhibit excess lactate production from the respiratory muscles, with some contribution from lactate

underutilization due to hypoperfusion of skeletal muscle and liver. The presence of lactic acidosis correlates with a peak expiratory flow rate of 60 l/min or less, and might suggest that respiratory fatigue is imminent and that the patient may require ventilatory assistance.

Drugs

Several drugs or toxins including phenformin, salicylates, ethylene glycol, and methanol may be associated with lactic acidosis.

Treatment

Type A

Therapy is directed at correcting or alleviating the underlying cause: this might include the administration of blood, fluids, and drugs to correct circulatory failure by optimizing myocardial performance. The underlying principle should be to maintain adequate levels of oxygen delivery. Mortality is high (> 75 per cent) in patients with persistent lactic acidosis despite appropriate measures.

Type B

No specific form of current therapy has been shown to reduce the mortality rate associated with type B lactic acidosis. Administration of sodium bicarbonate has been recommended to maintain the arterial pH above 7.2: below this there is myocardial depression and reduced cardiac output, while at a pH below 7.0, utilization of lactate by the liver is impaired. The dose of bicarbonate required can be estimated by multiplying the desired increase in HCO_3^-, in mmol, by 50 per cent of the patient's body weight in kilograms. However, bicarbonate stimulates glycolysis whilst depressing other oxidative reactions and therefore enhances lactate production. This increase in lactate production may contribute to the mortality associated with lactic acidosis. In diabetic ketoacidosis, bicarbonate delays the fall in blood lactate and total ketone bodies, even though it improves the pH. The relative risks and benefits of correcting a metabolic acidosis with bicarbonate are more thoroughly considered below.

Haemodialysis and peritoneal dialysis do not correct the cause of lactic acidosis but do restore the buffer pool. Dialysis and haemofiltration may supplement bicarbonate therapy by making space for additional fluid volume to be administered. Haemodialysis and haemofiltration fluids generally use lactate or acetate as the buffer. The buffering effect depends upon conversion to bicarbonate, which may be impaired in patients with severe lactic acidosis or liver disease. In such circumstances a bicarbonate buffer may be preferable, with the above limitations being acknowledged.

Alkali therapy of metabolic acidosis

Severe metabolic acidosis due to diabetic ketoacidosis, sepsis syndrome, or renal failure is a common clinical problem in the critical care unit. Arguments have been voiced both for and against the use of specific alkalinizing agents to correct the acidosis rapidly in an attempt to avoid the complications of the acidotic state. Severe acidosis has several serious and life-threatening effects, including circulatory failure due to reduce myocardial contractility, increased systemic and pulmonary vascular resistance, and enhanced arrhythmogenesis. In addition, the sensitivity of end-organ receptors to inotropes and pressors may be diminished, reducing the efficacy of these agents. Unfortunately, controlled studies of the effects of correction of metabolic acidosis with bicarbonate have failed to demonstrate improved haemodynamics. Additionally, and depending upon the alkalinizing agent selected, certain risks are associated with alkali therapy (Table 14).

1. Sudden electrolyte shifts to and from the intracellular space (particularly calcium and potassium)
2. Increased osmolarity with concentrated alkali
3. Shift in oxyhaemoglobin dissociation curve reducing peripheral tissue oxygen off-loading
4. Increased intracellular acidosis due to diffusion of CO_2 into cells (generated from bicarbonate)
5. Risk of cardiac arrest with pH > 7.55

Table 14 Unwanted effects of alkali therapy

The convenience of arterial blood-gas measurement, reflecting extracellular pH, has unfortunately diverted attention away from the need to correct intracellular acidosis. In addition, when tissue perfusion is poor, as occurs during cardiopulmonary resuscitation, arterial alkalaemia can coexist with venous and intracellular acidosis.

Sodium bicarbonate

Several bicarbonate solutions are available, some of which include salts of weak acids such as citrate (Shohl's solution) or lactate (lactated Ringer's solution) and require the metabolism of these precursors to bicarbonate (Table 15).

Preparations	*Bicarbonate content (mmol/l)*
5% bicarbonate solution	595
8.4% bicarbonate solution	1000
Shohl's solution	100
Lactated Ringer's solution	28

Table 15 Bicarbonate preparations

Hypertonic solutions are theoretically of advantage when hyponatraemia is present or in patients with circulatory overload. Isotonic solutions may be used when there is a risk of producing a hyperosmolar state or when volume expansion is desired. It is not unusual for patients with severe acidosis to require more than 200 mmol of bicarbonate per hour: such resistance to bicarbonate therapy usually denotes continuing acid production, as might occur in a lactic acidosis secondary to sepsis. When bicarbonate neutralizes hydrogen ion, CO_2 and H_2O are produced according to the formula:

$$H^+ + HCO_3^- \leftarrow H_2CO_3 \leftarrow H_2O + CO_2$$

In the presence of an adequate circulation and alveolar ventilation the excess CO_2 is quickly eliminated by the lungs and pH restored to normal. In patients with circulatory collapse or ventilatory failure, CO_2 accumulates in the blood and, although the buffering capacity by bicarbonate is enhanced the effect upon pH is less than expected. Cell membranes are readily permeable to CO_2 (unlike bicarbonate and hydrogen ions), and intracellular pH may paradoxically fall following the administration of bicarbonate. This paradoxical response is of particular importance with regard to cerebral and cardiac function, resulting in raised intracranial pressure and refractory

heart failure, respectively.

Hyperosmolality and hypernatraemia may be serious side-effects of sodium bicarbonate therapy, with each milliequivalent of bicarbonate providing twice the osmolar load. Severe and prolonged hypernatraemia can be associated with serious neurological sequelae including death.

Salt solutions of weak acids, such as acetate and lactate, can be used to improve base deficit indirectly. These act as bicarbonate precursors and are effective because of the subsequent generation of bicarbonate. The onset of effect is clearly slower than the equivalent doses of bicarbonate. Solutions that depend on hepatic metabolism to bicarbonate for their efficiency (e.g. lactated Ringer's, Shohl's solution) should be used with caution in patients with severe liver disease. Acetate is also contraindicated in diabetic ketoacidosis, when acetyl-coenzyme A is already present in excess.

Tris-hydroxymethylaminomethane

Tris-hydroxymethylaminomethane (**THAM**) is an aminoalcohol that has an osmolality similar to that of plasma. A 0.3 M solution is a more powerful buffer than bicarbonate and produces a lower osmolar and sodium load than bicarbonate in equivalent doses. THAM should be infused slowly at rates not exceeding 2 mmol/min, to avoid an excessively rapid fall in arterial Pco_2. After the infusion of 40 to 50 mmol, arterial blood gases can be repeated to assess benefit and determine whether further administration is required.

The buffering activity of THAM differs from bicarbonate in several important respects. By buffering carbonic acid, bicarbonate is generated and CO_2 is removed from plasma. By reducing extracellular and intracellular CO_2, intracellular pH rises, unlike the situation following bicarbonate administration, which may exaggerate intracellular acidosis. The non-ionized fraction of THAM also diffuses directly into the intracellular space and acts as a buffer within the cell. THAM appears to have a positive inotropic and antiarrhythmic effect by correcting myocardial acidosis. Hypoventilation may occur following its administration, due to the reduced CO_2 and resetting of chemoreceptor responsiveness, in which case mechanical ventilation may be required. Other adverse effects of THAM are shown in Table 16.

1. Central venous administration required due to alkaline pH
2. Accumulates in renal failure
3. Increased glycogen oxidation may cause hypoglycaemia
4. May induce hypoventilation

Table 16 Disadvantages of tris-hydroxymethylaminomethane

Experimental evidence suggests that intraneuronal acidosis following head injury can be harmful and that THAM can correct intracerebral acidosis, reduce oedema, and improve the energy state of brain tissue. Preliminary results of a clinical trial of this therapy in severely head-injured and comatose patients showed slightly improved survival.

Sodium carbonate

Sodium carbonate acts as a hydrogen ion acceptor and a bicarbonate precursor. Carbonate preferentially buffers hydrogen ions, producing bicarbonate as follows:

$$CO_3^- + CO_2 + H_2O \rightarrow 2HCO_3^- \text{ (uptake of } CO_2\text{)}$$

$$H^+ + HCO_3^- \rightarrow H_2CO_3 \text{ (uptake of } H^+\text{)}$$

and

$$CO_3^- + H^+ \rightarrow HCO_3^- \text{ (uptake of } H^+\text{)}$$

As with THAM, sodium carbonate produces a fall in extracellular and intracellular CO_2, and therefore reduces intracellular pH. However, it exerts no direct intracellular buffering, due to its highly ionized state.

A combination of sodium carbonate and sodium bicarbonate (Carbicarb®) has the attraction of bicarbonate (rapid activity without the need for metabolism to an active form) and the ability of carbonate to act as a CO_2 acceptor. Generally, a lower dose of Carbicarb is required compared to bicarbonate. In hypoxic lactic acidosis induced in animal models, Carbicarb proved superior to bicarbonate in correcting pH, reducing lactate, and protecting against hypotension.

Sodium dichloroacetate

Dichloroacetate has been shown to improve acidosis and support the circulation in some cases where bicarbonate therapy has failed. Dichloroacetate stimulates phosphodehydrogenase activity, increasing pyruvate oxidation, which in turn generates bicarbonate and reduces blood lactate. By this mechanism it is also able to reduce brain lactate more rapidly than is the case without therapy. Although dichloroacetate decreases the morbidity and mortality of experimentally induced lactic acidosis in dogs, it has yet to be shown to improve survival rates significantly in patients. There appear to be no major short-term complications from dichloroacetate therapy but chronic use carries a risk of drowsiness, paralysis, and polyneuropathy.

Further reading

Cardiac arrest and circulatory failure

Boyd O, Grounds ED, Bennett ED. A randomised clinical trial of the effects of deliberate perioperative increase of oxygen delivery on mortality in high risk surgical patients. *Journal of the American Medical Association* 1994; **270**: 2699–707. [Reduction in mortality and morbidity in high risk surgical patients by increasing oxygen delivery.]

Braunwald E. On the difference between the heart's output and its contractile state. *Circulation* 1971; **43**: 171–4.

Callahan M, Barton C. Prediction of outcome of cardiopulmonary resuscitation from end-tidal carbon dioxide concentration. *Critical Care Medicine* 1990; **18**: 358–62.

Chatterjee K. Vasodilator therapy for heart failure. In: Cohn JN, ed. *Drug treatment of heart failure*, pp. 151–78. York, New York, 1983.

Cohn, JN. Overview of the treatment of heart failure. *American Journal of Cardiology* 1997; **80**(11A): 2–6L.

Faniel R, Shoenfeld P. Efficacy of intravenous amiodarone in converting rapid atrial fibrillation and flutter to sinus rhythm in intensive care patients. *European Heart Journal* 1983; **4**: 180–5. [Acute slowing of AV node conduction and cardioversion within 3 hours.]

Fanning WJ *et al.* Prophylaxis of atrial fibrillation with magnesium sulphate after coronary artery bypass grafting. *Annals of Thoracic Surgery* 1991; **52** 529–33. [Significant reduction in incidence of atrial fibrillation. Main side-effect is hypotension.]

Gaasch WH, Apstein CS, Levine HJ. Diastolic properties of the left ventricle. In: Levine HJ, Gaasch WH, ed. *The ventricle: basic and clinical aspects*, pp. 143–70. Martinus Nijhoff, Boston, 1985.

Gheorghiade M. Digoxin therapy in chronic heart failure. *Cardiovascular Drug Therapy* 1997; **1**: 279–83.

Gheorghiade M *et al.* Pharmacotherapy for systolic dysfunction: a review of randomized clinical trials. *American Journal of Cardiology* 1997; **80**: 14–27H.

Goldberg LI. Cardiovascular and renal actions of dopamine: potential clinical applications. *Pharmacology Reviews* 1972; **24**: 1–29.

Kowey PR *et al.* Meta-analysis of the effectiveness of prophylactic drug therapy in preventing supraventricular arrhythmia early after coronary artery bypass grafting. *American Journal of Cardiology* 1992; **69**: 963–5. [b-Blockade reduces incidence of atrial fibrillation to 10 per cent postoperatively.]

Krischer JK, Fine EG, Weifeldt ML, Guerci AD. Comparison of prehospital conventional and simultaneous compression–ventilation cardiopulmonary resuscitation. *Critical Care Medicine* 1989; **17**: 1263–9.

MERIT-HF Study Group. Metoprolol in heart failure reducing heart failure progression and sudden death. *Lancet* 1999; **353**: 2001–7.

Noc M, Stajer D, Horvat M. Intravenous amiodarone versus verapamil for acute conversion of paroxysmal atrial fibrillation to sinus rhythm. *American Journal of Cardiology* 1990; **65**: 679–680. [Majority of patients with paroxysmal atrial fibrillation converted to sinus rhythm within 3 hours.]

Pitt B. ACE inhibitors in heart failure: prospects and limitations. *Cardiovascular Drug Therapy* 1997; **1**: 285–90.

Pitt B *et al.*The effect of spironolactone on morbidity and mortality in patients with severe heart failure. *New England Journal of Medicine* 1999; **341**: 709–17.

Resuscitation Council (UK). *The 1997 resuscitation guidelines for use in the United Kingdom*. Resuscitation Council, London, 1997.

Roberts SA *et al.* Effectiveness and costs of digoxin for atrial fibrillation and flutter. *American Journal of Cardiology* 1993; **72**: 567–73. [Digoxin takes several hours to control ventricular rate.]

Robertson CE. Advanced life support guidelines. European Resuscitation Council. *British Journal of Anaesthesia* 1997; **79:** 172–7

Royal College of Physicians. Resuscitation from cardiopulmonary arrest. Training and organisation. *Journal of the Royal College of Physicians* 1987; **21**: 1–8.

Strasberg B *et al.* Efficacy of intravenous amiodarone in the management of paroxysmal or new atrial fibrillation with a fast ventricular response. *International Journal of Cardiology* 1985; **7**: 47–55. [Sixty per cent of responders converted to sinus rhythm within 30 minutes.]

Struthers AD. Rationalizing the heart failure trials: from theory to practice. *European Heart Journal* 1997; **18** (Suppl.): E5–8.

Suttorp MJ, Kingma JH, Lie AHL, Mast EG. Intravenous flecainide versus verapamil for acute cardioversion of paroxysmal atrial fibrillation or flutter to sinus rhythm. *American Journal of Cardiology* 1989; **63**: 693–6. [Eighty per cent cardioversion to sinus rhythm within 1 hour with flecainide.]

The CIBIS-II Investigators and Committee. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *Lancet* 1999; **353**: 9–13.

Wilson JR, Reichek N, Dunkman WB, Goldberg S. Effect of diuresis on the performance of the failing left ventricle in man. *American Journal of Medicine* 1981; **70**: 234–9.

Oxygen uptake, sepsis syndrome, and multiple organ-system failure

Alverdy JC, Aoys E, Moss GS. Total parenteral nutrition promotes bacterial translocation from the gut. *Surgery* 1988; **104**: 185–90.

Bone RC *et al.* Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine. *Chest* 1992; **101**(6): 1644–55.

Gutierrez G, Dubin A. Cellular metabolism in sepsis. In: *Update in intensive care and emergency medicine*, Vol. 12, pp. 227–241. Springer, Berlin, 1990.

Ledingham IMcA, Alcock SR, Eastaway AT, McDonald JC, McKay IC, Ramsay G. Triple regime of selective decontamination of the digestive tract, systemic cefotaxime and microbiological surveillance for prevention of acquired infection in intensive care. *Lancet* 1988; **i**: 785–90.

Lefer AM, Rovetto MJ. Influence of a myocardial depressant factor on physiologic properties of cardiac muscle. *Proceedings of the Society for Experimental Biology and Medicine* 1970; **134**: 269–73.

McArdle AH, Palmerson C, Brown RA, Brown HC, Williams HB. Early enteral feeding of patients with major burns: prevention of catabolism. *Annals of Plastic Surgery* 1984; **13**: 396–401.

Matushak GM, Rinaldo JE. Organ interactions in the adult respiratory distress syndrome during sepsis. *Chest* 1988; **94**: 400–6.

Parker MM *et al.* Profound but reversible myocardial depression in patients with septic shock. *Annals of Internal Medicine* 1984; **100**: 483–90.

Parrillo JE *et al.* Septic shock in humans. Advances in the understanding of pathogenesis, cardiovascular dysfunction, and therapy. *Annals of Internal Medicine* 1990. **113**: 227–42

Ramsay G. Endotoxaemia in multiple organ failure: a secondary role for SDD. In: van Saene HKF, Stoutenbeek CP, Lawin P, Ledingham IMcA, ed. *Infection control by selective decontamination*, pp. 135–42. Springer, Berlin, 1989.

Ravin HA, Rowley D, Jenkins C, Fine J. On the absorption of bacterial endotoxin from the gastrointestinal tract of the normal and shocked animal. *Journal of Experimental Medicine*, 1960; **112**: 783–90.

Rush BF *et al.* Endotoxaemia and bacteraemia during haemorrhagic shock. *Annals of Surgery* 1988; **207**: 549–54.

Shibutani K, Komatsu T, Kubal K. Critical level of oxygen delivery in anaesthetized man. *Critical Care Medicine* 1983; **11**: 640–3.

Shoemaker WC, Appel PL, Kram HB, Waxman K, Tai-Shion L. Prospective trial of supranormal values of survivors as therapeutic goals in high-risk surgical patients. *Chest* 1988; **94**: 1176–86.

Stoutenbeek CP, van Saene HKF, Miranda DR, Zandstra DF. The effect of selective decontamination of the digestive tract on colonisation and infection rate in multiple trauma patients. *Intensive Care Medicine* 1984; **10**: 185–92.

Stuber F *et al.* A genomic polymorphism within the tumor necrosis factor locus influences plasma tumor necrosis factor-alpha concentrations and outcome of patients with severe sepsis. *Critical Care Medicine* 1996. **24**: 381–4.

Wellman W, Fink TC, Benner F, Schmidt FW. Endotoxaemia in active Crohn's disease. Treatment with whole gut irrigation and 5-aminosalicylic acid. *Gut* 1986; **27**: 814–20.

Wilmore DW *et al.* The gut: a central organ after surgical stress. *Surgery* 1988; **104**: 917–23.

Zeni F, Freeman B, Natanson C. Anti-inflammatory therapies to treat sepsis and septic shock: a reassessment [editorial; comment]. *Critical Care Medicine* 1997; **25**(7): 1095–100.

Zieglar TR *et al.* Increased intestinal permeability associated with infection in burn patients. *Archives of Surgery* 1988; **123**: 1313–19.

Pulmonary arterial catheter

Baele PL, McMichan JC, Marsh HM, Sill JC, Southorn PA. Continuous monitoring of mixed venous oxygen saturation in critically ill patients. *Anesthesia and Analgesia* 1982; **61**: 513–17.

Boyd KD, Thomas SJ, Gold J, Boyd AD. A prospective study of complications of pulmonary artery catheterizations in 500 consecutive patients. *Chest* 1983; **84**: 245–9.

Connors AF Jr *et al.*The effectiveness of right heart catheterization in the initial care of critically ill patients. SUPPORT Investigators. *Journal of the American Medical Association* 1996; **276**: 889–97

Forrester JS, Diamond GA, Swan HJC. Correlative classification of clinical and hemodynamic function after acute myocardial function. *American Journal of Cardiology* 1977; **39**: 137–45.

Thijs LG, Schneider AJ, Groeneveldt AB. The haemodynamics of septic shock. *Intensive Care Medicine* (United States) 1990; **16** suppl.: S182–6.

Iberti TJ, Fischer EP, Leibowitz AB, Panacek EA, Silverstein JH, Albertson TE. A multicenter study of physicians' knowledge of the pulmonary artery catheter. Pulmonary Artery Catheter Study Group. *Journal of the American Medical Association* 1990; **264**: 2928–32

Pinilla JC, Ross DF, Martin T, Crumb H. Study of the incidence of intravascular catheter infections and associated septicaemia in critically ill patients. *Critical Care Medicine* 1983; **11**: 21–5.

Robin ED. The cult of the Swan–Ganz catheter: overuse and abuse of pulmonary flow catheters. *Annals of Internal Medicine* 1985; **103**: 445–9. [Early critique of the use of pulmonary catheters.]

Shah KB, Rao TK, Langhlin S, El-Etr A. A review of pulmonary artery catheterization in 6,245 patients. *Anesthesiology* 1984; **61**: 271–5.

Stevenson LW, Perloff JK. The limited reliability of physical signs for estimating hemodynamics in chronic heart failure. *Journal of the American Medical Association* 1989; **261**: 884–8.

Swan HJC, Ganz W, Forrester J, Marcus H, Diamond G, Chonette D. Catheterization of the heart in man with use of a flow-directed balloon tipped catheter. *New England Journal of Medicine*, 1970; **283**: 447–51. [First description of flow directed pulmonary catheter for clinical use.]

Vincent J-L, Thirion M, Brimioulle S, Lejeune P, Kahn RJ. Thermodilution measurement of right ventricular ejection fraction with a modified pulmonary artery catheter. *Intensive Care Medicine* 1986; **12**: 33–8.

Lactic acidosis and alkali therapy

Bersin RM, Arieff AI. Improved hemodynamic function during hypoxia with Carbicarb, a new agent for the management of acidosis. *Circulation* 1988; **77**: 227–33.

Effron MB, Guarnieri T, Frederkisen JW, Greene HL, Weisfeldt ML. Effects of TRIS (hydroxymethyl) aminomethane on ischemic myocardium. *American Journal of Physiology* 1978; **235**: H167–74.

Jaffe AS. New and old paradoxes. Acidosis and cardiopulmonary resuscitation. *Circulation* 1989; **80**: 1079–83.

Kruse JA, Haupt, MT, Puri VK, Carlson, RW. Lactate levels as predictors of the relationship between oxygen delivery and consumption in ARDS. *Chest* 1990; **98**: 959–62.

Shapiro JI. Functional and metabolic responses of isolated hearts to acidosis: effects of sodium bicarbonate and Carbicarb. *American Journal of Physiology* 1990; **258**: H1835–9.

Stacpoole PW. Lactic acidosis: the case against bicarbonate therapy. *Annals of Internal Medicine* 1986; **105**: 276–9.

Stacpoole PW, Lorenz, AC, Thomas RG, Harman EM. Dichloroacetate in the treatment of lactic acidosis. *Annals of Internal Medicine* 1988; **108**: 58–63.

von Gazmuri RJPM, Weil MH, Rackow, EC. Cardiac effects of carbon dioxide-consuming and carbon dioxide-generating buffers during cardiopulmonary resuscitation. *Journal of the American College of Cardiology* 1990; **15**: 482–90.

Weil MH, Ruiz CE, Michaels S, Rackow EC. Acid base determinants of survival after cardiopulmonary resuscitation. *Critical Care Medicine* 1985; **13**: 888–92.

11.3 Renal aspects

Christopher S. Garrard

Oliguria
Calculation of fractional excretion of sodium (FE_{Na})
Application of fractional excretion of sodium (FE_{Na})
Management of prerenal azotaemia
Acute renal failure
Management
Anion and osmolal gaps
Anion gap
Osmolal gap
Further reading

Oliguria

P>An appreciation of the significance of oliguria, and its early recognition is essential if steps are to be taken to preserve renal function in the critically ill patient. The obligate solute load that has to be excreted each day determines the minimum effective urine volume. In healthy individuals, about 800 mmol can be eliminated in as little as 650 ml of urine. Even in those taking a high carbohydrate, low sodium, and low nitrogen diet, any degree of renal impairment greatly increases the minimum obligate urine volume. As a rough guide, a urine requirement of 1 ml/kg.h provides adequate solute clearance. At 0.5 ml/kg.h (840 ml for a 70-kg patient) azotaemia will only be avoided if renal function is normal. Urinary tract catheterization is often necessary, but is invasive. Once the potential complications associated with urinary catheters have been accepted, it is all the more important to monitor urine volumes constantly and to take appropriate action if necessary. Table 1 outlines a simple approach to the evaluation and treatment of oliguria. Complete anuria should always, of course, prompt consideration of a blocked urinary catheter.

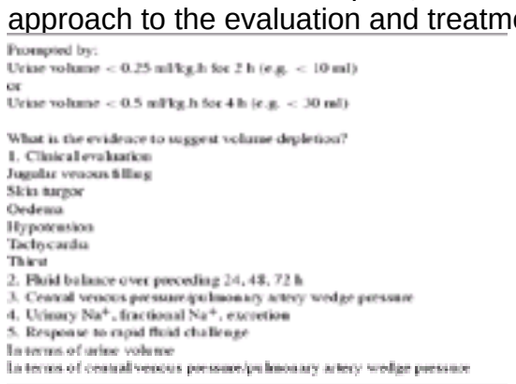


Table 1 Evaluation of oliguria in the critically ill

Relying on only a limited number of the criteria listed in Table 1 may lead to an erroneous conclusion regarding the volume status of the patient. Several other pitfalls may mislead the unwary. In the patient with sepsis, oedema (extravascular fluid) may coexist with marked intravascular hypovolaemia, indicated by a low central venous pressure and positive response to a fluid challenge. Conversely, the central venous pressure and pulmonary artery wedge pressure can easily be overestimated in patients on positive pressure ventilation if measurements are not made at end expiration.

There can be no substitute for personally measuring the venous pressure and, in doing so, confirming the calibration and zeroing of the equipment. Overreliance upon fluid balance records, even when accurately maintained, can mislead since the compliance of the major capacitance vessels can change.

Several urinary measurements may help to distinguish potentially reversible prerenal azotaemia from established intrinsic renal failure; these include sodium concentration (U_{Na}), osmolality (U_{osm}), urine to plasma urea nitrogen ratio, urine to plasma creatinine ratio (U_{cr}/P_{cr}), fractional excretion of sodium (FE_{Na}), and renal failure index. Table 2 shows the differences between these indices for prerenal and renal azotaemia.

Index	*Prerenal azotaemia*	*Acute renal failure*
U_{osm}	High U_{osm}	Isosthenuria
U_{Na}	< 20 mmol/l	> 40 mmol/l
FE_{Na}	< 1 %	> 1 %
U_{cr}/P_{cr}	> 20:1	< 15:1
Urine sediment	Normal	Cellular casts

See text for abbreviations

Table 2 Comparison of prerenal azotaemia and acute renal failure

Calculation of fractional excretion of sodium (FE_{Na})

Filtered and excreted sodium can be calculated as follows:

$$FE_{Na} = \frac{\text{urine Na}}{\text{filtered Na}} \times 100(\%)$$

If urine sodium = urine volume × U_{Na} and filtered sodium = GFR × P_{Na}, then

$$FE_{Na} = \frac{U_{Na} V}{GFR \times P_{Na}} \times 100(\%)$$

$$FE_{Na} = \frac{U_{Na} V}{(Ucr V / Pcr) \times P_{Na}} \times 100(\%)$$

$$FE_{Na} = \left(\frac{U_{Na}}{P_{Na}}\right) \times \left(\frac{P_{cr}}{U_{cr}}\right) \times 100(\%)$$

where GFR is the glomerular filtration rate (ml), P_{Na} is the plasma sodium concentration (mmol/1), U_{Na} is the urine sodium concentration (mmol/1), P_{cr} is the plasma creatinine concentration (mmol/1), U_{cr} is the urine creatinine concentration (mmol/1), and V is the urine flow rate (ml/min).

Since creatinine is essentially not reabsorbed or secreted, the urinary concentration of creatinine is a function of water reabsorption. Sodium, however, is actively reabsorbed by the tubules, such that the final U_{Na} depends upon sodium reabsorption as well as the amount of water reabsorbed by the tubules. The renal failure index (RFI) offers no advantages over the FE_{Na} but may be calculated as follows.

$$RFI = U_{Na} \times \frac{P_{Cr}}{U_{Cr}}$$

Application of fractional excretion of sodium (FE_{Na})

In the oliguric patient, a FE_{Na} below 1 per cent is consistent with prerenal azotaemia. A FE_{Na} of less than 0.4 per cent is even more specific. Conversely, a FE_{Na} of greater than 1 per cent may be consistent with renal damage, although a FE_{Na} greater than 3 per cent is much more indicative of intrinsic renal disease. Values between 1 and 3 per cent are therefore less conclusive. A low FE_{Na} may, under certain circumstances, be an unreliable guide since patients with intrinsic renal failure occasionally have a FE_{Na} of 1 per cent. Conversely, there are hypovolaemic conditions in which the FE_{Na} can be paradoxically high (Table 3).

Causes of FE_{Na} less than 1 per cent in patients with intrinsic renal failure
(i) Non-oliguric acute renal failure
(ii) Early acute glomerulonephritis (without tubular damage)
(iii) Renal failure associated with severe renal vasoconstriction
Hepatorenal syndrome
Sepsis syndrome
Non-steroidal anti-inflammatory drugs
Renal vasculitides
Radiocontrast induced acute renal failure
Myoglobinuria and haemoglobinuria
(iv) Renal hyperperfusion
ACE inhibitors
Renal artery stenosis
Normovolaemic patient with burns
Renal transplant rejection
Cyclosporin nephrotoxicity
Cause of a FE_{Na} greater than 1 per cent in the hypovolaemic patient
(i) Chronic renal disease
(ii) Diuretic use within 12 to 24 h of testing FE_{Na}
Thiazides
Frusemide
Osmotic diuretics
Glycosuria
(iii) Metabolic alkalosis with alkaline urine (pH > 7.0)
(iv) Addison's disease

Table 3 Limits of reliability of FE_{Na} values

The increase in the urine sodium produced by diuretics persists much longer following ingestion of thiazides than after frusemide, ethacrynic acid, or the osmotic diuretics. By 24 h after the last dose of diuretic, FE_{Na} can again be adopted as an indicator of prerenal azotaemia.

There are rare situations where the sodium cation is an unreliable indicator of the nature of renal dysfunction. This is the case when there is sodium wasting such as occurs in the patient with a metabolic alkalosis due to the loss of upper gastrointestinal tract secretions. In such a situation, a urinary chloride level of less than 20 mmol/1 (or FE_{Cl} less than 1 per cent) is indicative of volume depletion.

Management of prerenal azotaemia

Once the clinician is reasonably confident that the observed oliguria is due to volume depletion the obvious response is to administer fluid, the nature of which is dictated by the degree of urgency, electrolyte status, and the haemoglobin concentration of the patient. Colloidal solutions expand the intravascular space rapidly, while electrolyte solutions are distributed across intra- and extracellular compartments. Blood is preferred when the patient is anaemic. If fluid is infused rapidly (500 ml in 30 min), the response to volume expansion is similar regardless of the type of fluid.

A central venous pressure line or pulmonary artery catheter will usually be required to optimize preload and ensure the preservation of renal function. Accurate calibration and zeroing of the water column or transducer is essential if meaningful data is to be obtained. Target central venous or pulmonary artery wedge pressures vary from patient to patient and whether or not the patient is mechanically ventilated. As a rough guide, central venous pressure values of less than 10 mmHg might indicate a degree of underfilling, while the 10 to 20 mmHg range is more consistent with a replete state. The haemodynamic and renal response to a fluid bolus will often resolve any doubts as to the intravascular volume status of the patient. Fluid resuscitation should be performed quickly and under the personal supervision of the clinician. It is insufficient simply to increase the infusion rates of existing fluids and re-evaluate the patient 6 to 12 h later.

Several strategies involving pharmacological agents have been proposed to prevent or reverse progression from prerenal failure to established vasomotor nephropathy. The administration of dopamine in so-called ‘renovascular' doses (about 2.5 µg/kg.min) is popularly believed to reverse the renal vasoconstriction that is a feature of the volume depleted state. Other vasoactive and inotropic agents such as dobutamine and dopexamine have also been proposed as renoprotective agents. Alternatively, infusion (2 mg/h) or boluses of frusemide will induce diuresis in the patient resuscitated with fluid. Despite all these agents being able to induce and maintain a diuresis, there is little evidence that they can increase glomerular filtration rate or prevent the onset of vasomotor nephropathy.

Vasoconstrictor agents have often been considered likely to contribute to renal vasoconstriction and have therefore been contraindicated in prerenal azotaemic states. Contrary to this opinion, recent experience in sepsis has suggested that pressor agents such as noradrenaline, by increasing systemic blood pressure, may encourage a diuresis more effectively than inotropes such as dobutamine that have peripheral vasodilator properties.

Acute renal failure

Less than 10 per cent of patients with uncomplicated acute renal failure die, but in those with multiple organ system failure the mortality rate from acute renal failure rises to above 70 per cent, and may approach 100 per cent when four organ systems fail. Renal failure therefore represents a serious complication in the critically ill patient.

A prevailing challenge for the clinician is to be able to recognize when prerenal failure has progressed to vasomotor nephropathy. Usually this rests on the belief that adequate volume resuscitation has been undertaken, that good perfusion pressures have been maintained (mean arterial blood pressure greater than 70 mmHg), and that obstruction to urine flow has been excluded. Once renal failure is established, additional renal insult must be avoided and consideration paid to the reduction in fluid and drug elimination. Acute renal failure may be grouped into four broad categories: ischaemic injury, nephrotoxicity, glomerular disorders, and vascular disorders (Table 4).

Vasomotor nephropathy	
Sepsis syndrome	
Severe polytrauma	
Massive haemorrhage	
Transfusion reaction	
Myoglobinuria	
Pancreatitis	
Dehydration	
Gastrointestinal losses	
Burns	
Nephrotoxic agents associated with renal failure	
Antibiotics (e.g. aminoglycosides, penicillin)	
Radiographic contrast media	
Heavy metals	
Anaesthetic agents	
Organic solvents	
Insecticides	
Disease of glomeruli and small vessels	
Acute glomerulonephritis	
Systemic lupus erythematosus	
Polyarteritis nodosa	
Henoch–Schönlein purpura	
Wegener's granulomatosis	
Goodpasture's syndrome	
Sickle cell disease	
Malignant hypertension	
Haemolytic uraemic syndrome	
Drug-induced vasculitis	
Abruptio placentae	
Disease of large renal vessels	
Renal arterial thrombosis, emboli, or stenosis	
Bilateral renal vein thrombosis	

Table 4 Some causes of acute renal failure in intensive care

Although more than 80 per cent of renal failure in the patient in the critical care unit will be due to vasomotor nephropathy, the clinician must remain alert to the possibility of other aetiologies. It may be necessary to rule out an obstructive uropathy or an active glomerulitis. Ultrasound visualization of the kidney and collecting system is the least invasive and most convenient method of excluding obstruction. Retrograde urography may still be necessary to identify accurately the level of obstruction and facilitate the placement of stents. Occasionally, renal biopsy may be indicated when the aetiology is obscure and the urine sediment is consistent with a glomerulopathy.

Management

The most immediate clinical problem for the patient who has recently developed renal failure is the maintenance of an appropriate intravascular volume in the face of a large obligate fluid input owing to drug administration and total parenteral nutrition. By concentrating the nutrients, reducing drug diluent volumes, and using syringe pumps instead of volumetric pumps, fluid overload can be minimized. Hyperkalaemia can be temporarily controlled by the use of glucose and insulin infusion (1 unit soluble insulin to each 2 g of glucose) or ion exchange resins. Although high levels of urea and related metabolites are well tolerated in the ambulant or stable patient with renal failure, an early and aggressive approach to renal support with haemofiltration or haemodialysis is preferable in the patient with multiple organ system failure. Thus, the threshold for embarking upon haemofiltration or haemodialysis should be much lower in the patient who is critically ill than in the patient with uncomplicated renal failure. Instead of waiting for volume overload, bleeding, uraemic pericarditis, hyperkalaemia, or acidosis, renal replacement therapy should be started early to maintain the metabolic state as near normal as possible. Such use of renal replacement therapy is associated with an improved outcome, and the overall approach to patient management is greatly simplified once haemofiltration is established. Fluid balance is readily achieved and, provided adequate caloric intake is available, nitrogen intake does not need to be restricted. Full-profile amino acid solutions are probably as effective as the special solutions of essential and branch-chain amino acids. Energy requirements should be met using a balance of carbohydrate and lipid to deliver 45 to 50 kcal/kg.day. Where renal function is impaired for longer than 2 to 3 weeks, some consideration has to be given to supplementation with vitamins and essential elements such as zinc.

Recovery from a reversible renal pathology such as vasomotor nephropathy is typically heralded by a polyuric phase followed by a plateau in serum creatinine lasting a few days, before creatinine levels consistently fall.

Renal replacement therapy

Renal replacement therapy can take the form of peritoneal dialysis, intermittent haemodialysis, continuous ultrafiltration, continuous haemofiltration, or continuous haemodiafiltration. Continuous arteriovenous or venovenous haemofiltration provides relatively simple and effective renal replacement therapy that is particularly well suited to the patient receiving critical care. Although haemodialysis removes certain solutes, such as potassium, more efficiently, haemofiltration causes less haemodynamic disturbance and facilitates the regulation of fluid balance.

Peritoneal dialysis is suitable for patients with acute renal failure who have not had abdominal surgery and are not excessively catabolic. However, peritoneal dialysis has been largely superseded by haemofiltration, although it may still be preferred in patients suspected of carrying hepatitis B virus or HIV, or in very young children.

Indications for renal replacement therapy

The need for renal replacement therapy must be evaluated in each individual patient. Although there are threshold values for hyperkalaemia (> 6.0 mmol/l), blood urea (> 40 mmol/l), or severe acidosis (pH < 7.2) which would justify renal replacement therapy, the overall clinical status, including the need to make room for drugs and nutrition, all need to be considered. The aim of renal replacement therapy is to remove uraemic toxins and maintain electrolyte, acid/base, and fluid balance.

In the patient with uncomplicated renal failure, maintaining the plasma urea below 30 mmol/l would generally be considered acceptable. In the critically ill patient, lower target values approaching 20 mmol/l are preferred.

Continuous haemofiltration

Continuous haemofiltration is achieved by passing heparinized blood at flow rates of between 100 and 200 ml/min through a highly permeable haemofilter. The modern high-flux, biocompatible haemofilters, made from polyamide, polyacrylonitrile, or polysulphone, have negligible effects on platelets, neutrophils, or complement.

Plasma water passes through the haemofilter membrane and is drained into a collecting system (Fig. 1). Blood cells and proteins are not filtered through the membrane and are returned to the patient. The haemofiltrate contains all of the water-soluble components of plasma and thus enough urea and creatinine are removed to control the patient's biochemistry. At haemofilter blood flows of 200 ml/min, about 1 litre of filtrate will be produced each hour. Since the clearance of creatinine and urea by the membrane is close to 100 per cent, a filtration rate of 1 l/h (24 l/day) results in a clearance of about 17 ml/min, which is adequate for all but the most catabolic patient. Since the plasma concentration of potassium is very low, only about 100 mmol can be removed each day. Haemofiltration is therefore not the most efficient treatment for hyperkalaemia although any correction of metabolic acidosis lessens the risks associated with hyperkalaemia.

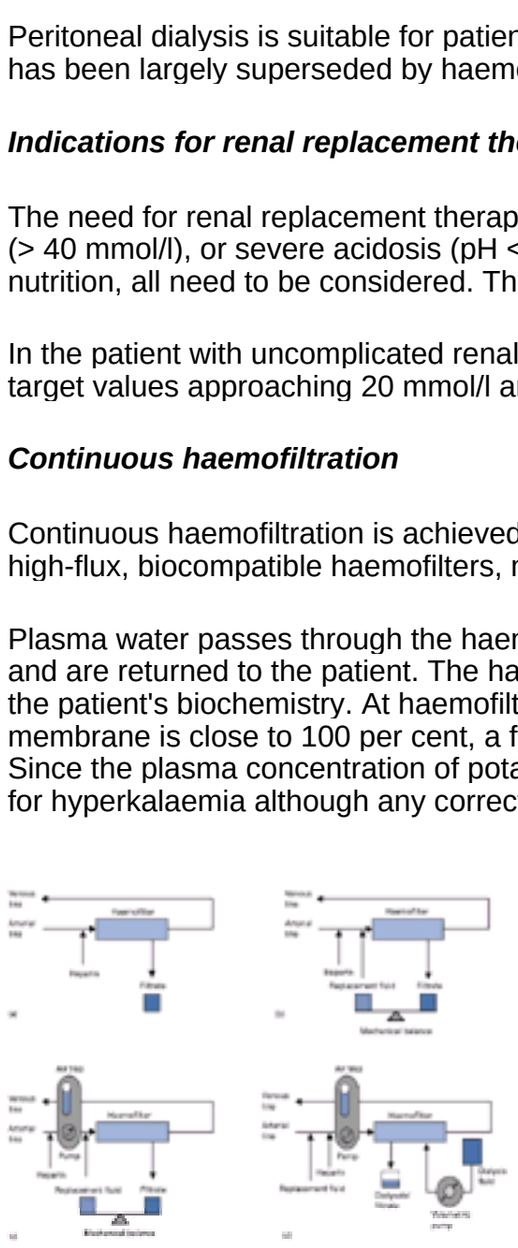


Fig. 1. (a) Schematic of an extracorporeal circuit for spontaneous, continuous ultrafiltration. (b) Schematic of an extracorporeal circuit for continuous arteriovenous haemofiltration. (c) Schematic of an extracorporeal circuit for continuous venovenous haemofiltration. (d) Schematic of an extracorporeal circuit for continuous venovenous haemodiafiltration. Venovascular access is conveniently achieved using a double lumen venous catheter placed within a major large vein. One lumen becomes the ‘arterial line', while th other becomes the venous return line.

A number of uncontrolled studies have demonstrated the removal, by haemofiltration, of circulating cytokines in patients with sepsis. Although clinical benefit has been claimed by this technique (and plasma exchange), it fails to acknowledge that cytokines have both pro- and anti-inflammatory effects and that many are fixed to cell

surfaces. There is no randomized, controlled evidence that haemofiltration improves outcome in sepsis syndrome.

Haemofiltration modes

Several techniques are available to the clinician depending upon the availability of local expertise and the needs of the patient. An early, but less commonly used, mode is spontaneous continuous ultrafiltration, which requires arterial and venous access such as that provided by a Scribner shunt (Fig. 1(a)). Filtered volume is not replaced and relief from fluid overload can be achieved quickly.

The volume of filtrate removed by the haemofilter can be replaced by a suitable haemofiltration replacement fluid which contains a lactate buffer. The replacement fluid may be infused through the return (venous) blood line or proximal to the haemofilter (predilution). The latter technique improves blood flow through the filter and increases haemofiltration efficiency. The reinfusion of a suitable haemofiltration fluid converts spontaneous continuous ultrafiltration to continuous arteriovenous haemofiltration (Fig. 1(b)). The filtration rate is regulated by a gate clamp on the filtrate outflow tubing. Tightening of this clamp reduces the filtration rate as does raising the level of the collection bag.

A feature of these two techniques is that as the blood pressure falls in response to fluid withdrawal, the rate of blood flow and therefore filtration is reduced. The absence of a blood pump further adds to the safety and simplicity of this technique.

An alternative to continuous arteriovenous haemofiltration is continuous venovenous haemofiltration (Fig. 1(c)), which requires the insertion, by the Seldinger technique, of a double lumen venous catheter in the jugular, subclavian, or femoral vein. Infection of these catheters is a potential hazard and they should not be used for any other purpose. All access sites, whether arteriovenous or venovenous, should always be kept uncovered and in direct view of the attending nurse so that disconnections are immediately apparent. Pumped continuous venovenous haemofiltration can maintain clearances approaching 17 ml/min and requires close monitoring and safety alarms. The relative efficiency of this technique and ease of vascular access make it suitable for routine intensive care renal replacement.

Continuous arteriovenous or venovenous haemodiafiltration (Fig. 1(d)) offers yet another alternative. Haemofiltration replacement fluid is pumped by a volumetric infusion pump, in a countercurrent direction, to the filtrate side of the haemofilter membrane. Dialysate fluid flow rates of only 1 to 2 l/h are required to obtain clearance rates that are over twice that of haemofiltration alone.

Haemofiltration protocol

The entire extracorporeal circuit must be thoroughly flushed through with 1 to 2 litres of heparinized saline (5 units of heparin/ml) to exclude all air from the system. This is a simple procedure for continuous arteriovenous haemofiltration but rather more involved with continuous venovenous haemofiltration or haemodiafiltration, which require blood pumps. Suitably trained nurses or specialists in intensive care can undertake all aspects of haemofiltration if specialized renal unit personnel are unavailable. Some recommend administration of 5000 units of heparin to the patient, but this is generally not necessary. Maintenance of fluid balance is achieved by either adding replacement fluid intermittently each hour or by using a mechanical or microprocessor-controlled balance which will ensure a preset fluid balance (such as 0, 500, or 1000 ml). With the mechanical balance system, fluid balance goals can be changed at the end of each 4-litre cycle if necessary. Microprocessor-controlled balances allow continuous adjustment of fluid balance. These automated systems exchange 20 to 30 litres of fluid over an 8- to 24-h period depending upon the blood flow rate through the filter. Judging the volume of fluid to be removed over the length of a haemofiltration period depends upon the preceding fluid balance and perceived fluid balance goals. Overenthusiastic volume removal in a patient requiring intravascular expansion (e.g. in sepsis) will often precipitate hypotension and the need for volume resuscitation. Gradual slowing of filtration rate, as evidenced by a lengthening of the cycle time, usually indicates impending failure of the filter owing to blood clots.

Anticoagulation

Continuous infusion of 500 to 1000 units of heparin per hour into the haemofilter input blood line will generally prevent clotting within the extracorporeal circuit. The dose of heparin represents a compromise between the need to prevent the haemofilter from clotting and the avoidance of systemic effects. In patients at great risk from bleeding the use of low-molecular-weight heparin or prostacyclin should be considered. In the patient with uraemia, the extracorporeal circuit can often be maintained without any anticoagulant for 6 to 12 h. The viability of haemofilter circuits is difficult to predict even with heparin. Some remain functioning for 5 or 6 days while other repeatedly fail after 24 h.

Thrombocytopenia is not uncommon in critically ill patients. In some, this can be attributed to heparin therapy used during haemofiltration. Alternative forms of anticoagulation then need to be considered. Low-molecular-weight heparin is less likely to induce thrombocytopenia, while prostacyclin may protect platelets.

Effects on acidosis and temperature

Most haemofiltration replacement fluids contain lactate as a buffer. Conversion to bicarbonate is impaired in patients with renal and hepatic disease and correction of the metabolic acidosis may therefore be slower than expected. If necessary, non-buffered replacement fluid can be used and bicarbonate infused separately to correct the acidosis. In contrast, a marked metabolic alkalosis may be precipitated by lactate buffer, particularly if the patient is small and the filtration rates high.

Hypothermia (temperature < 35°C) can develop with high flow rates of replacement fluid unless the fluid is warmed with a blood warmer. Even with warmed fluid, the fever associated with sepsis may be abated, suggesting that circulating cytokines are actively removed by the filtration process.

Drug therapy and nutrition

Venovenous haemofiltration can provide the equivalent of a creatinine clearance of almost 20 ml, a value that is often quoted in drug package inserts as a guide for drug dosage. The dose of commonly used drugs that are not protein bound, including antibiotics, must therefore be carefully regulated. Wherever possible drug levels should be monitored. Once haemofiltration is established, dietary nitrogen restriction is no longer necessary and standard parenteral and enteral nutrition preparations can be used.

Anion and osmolal gaps

Anion gap

The anion gap can be a useful guide in resolving biochemical disturbances in the critically ill patient. The anion gap is calculated by the sum of the oncentrations of the principal anions, chloride and bicarbonate, subtracted from the sodium concentration, that is $[Na^+] - ([Cl^-]+[HCO_3^-])$. The normal value is in the range 7 to 18 mmol/l.

Low values are seen in a variety of conditions, including hypoalbuminaemia, bromide or iodide toxicity, and myeloma. However, the changes are sufficiently small as to render the recognition of low anion gaps to be of little clinical use. In contrast, a raised anion gap, due to excess of anions other than chloride and bicarbonate, may be of clinical value. These anions include lactate, citrate, and acetate, present usually as the sodium salts, but may also be inorganic or organic acids. When this is the case the increased anion gap is associated with a metabolic acidosis (Table 5). Not all metabolic acidosis is associated with a raised anion gap, however. When acidosis is caused by the primary loss of bicarbonate or the gain of hydrochloric acid, there are no foreign anions present and the sodium ions are balanced by appropriate though abnormal amounts of chloride or bicarbonate (Table 6).

1.	Renal failure	
2.	Lactic acidosis	
3.	Ketoacidosis	
		β-Hydroxybutyric acid
		Acetoacetic acid
4.	Intoxication	
		Salicylate
		Methanol
		Ethylene glycol
		Paraldehyde
		Sulphur
5.	Non-ketotic hyperglycaemia	
6.	Rhabdomyolysis	

Table 5 High anion gap metabolic acidosis

1. Loss of bicarbonate
(i) Gastrointestinal
Diarrhoea
Pancreatic fistula
Ileostomy
Ureter—bowel diversion
(ii) Renal tract
Renal tubular acidosis
Acetazolamide therapy
Post-hyperventilation
2. Dilutional
Rapid saline infusion
3. Administration of hydrochloride donors
Ammonium chloride
Arginine hydrochloride
Lysine hydrochloride
Hydrochloric acid

Table 6 Normal anion gap metabolic acidosis

In high anion gap acidosis, the fall in bicarbonate approximates the rise in anion gap. In mixed acid–base disorders the change in anion gap may be more or less than that expected from the change in bicarbonate: an oversimplistic approach to the application of anion gap theory may therefore mislead the clinician. Any consensus of opinion reached by the use of anion gap analysis must pass the test of being consistent with the overall clinical picture.

Osmolal gap

The osmolal gap is the difference between the measured and the calculated osmolality.

$$\text{Calculated osmolality} = ([Na^+] \times 2.0) + [\text{urea}] + [\text{glucose}] + [\text{potassium}]$$

The factor of 2.0 applied to the sodium concentration (in mmol/l) allows for the osmolal contribution of the anions that balance sodium cations. Estimates of osmolal gap produce values of 0 ± 6 mosmol/l. The clinical utility of the osmolal gap rests upon identifying patients with toxins, drugs, or certain plasma constituents present in excessive amounts in their serum. Ethanol, methanol, and ethylene glycol are common toxins associated with a raised osmolal gap. The osmolal gap will be raised in patients receiving mannitol for the treatment of raised intracranial pressure and the size of the gap parallels the plasma mannitol concentration.

As with the anion gap, the osmolal gap may help elucidate the underlying cause of certain electrolyte disturbances. However, neither are a substitute for a careful and systematic clinical evaluation of the patient.

Further reading

Bellomo R, Mehta R. Acute renal replacement in the intensive care unit: now and tomorrow. *New Horizons* 1995; **3**: 760–7. [Reviews the controversies surrounding indications and timing for acute renal replacement therapy.]

Bellomo R, Ronco C. Indications and criteria for initiating renal replacement therapy in the intensive care unit. *Kidney International Supplement* 1998; **66**(9): S106–9. [Indications for renal support may be different in acute renal failure.]

Bellomo R, Cole L, Reeves J, Silvester W. Who should manage CRRT in the ICU? The intensivist's viewpoint. *American Journal of Kidney Diseases* 1997; **30**(5 Suppl. 4): S109–11. ['Intensivists' have taken over management of chronic renal replacement therapy.]

Cameron JS. Acute renal failure in the intensive care unit today. *Intensive Care Medicine* 1986; **12**: 64–70. [Review of causes, complications, and outcome from intensive care unit associated renal failure.]

Emmett M, Narins RG. Clinical use of the anion gap. *Medicine* 1977; **56**: 38–54.

Gennari FJ. Serum osmolality uses and limitations. *New England Journal of Medicine* 1984; **310**: 102–5. [Differences between calculated and measured osmolalities provides a useful screen for the presence of foreign low molecular weight solutes in the blood.]

Golper TA, Pulliman J, Bennett WM. Removal of therapeutic drugs by continuous arteriovenous haemofiltration. *Archives of Internal Medicine* 1985; **145**: 1651–2. [Ultrafiltrate drug concentrations are very variable and are affected by many factors.]

Henderson LW, Colton CK, Ford CA. Kinetics of haemodiafiltration: II. Clinical characterisation of a new blood cleansing modality. *Journal of Laboratory and Clinical Medicne* 1975; **85**: 372–91. [Larger molecules are removed by haemodiafiltration, more like the human glomerulus.]

Kierdorf HP. The nutritional management of acute renal failure in the intensive care unit. *New Horizons* 1995; **3**: 699–707. [No indications to reduce nutrition due to fluid restriction issues.]

Klahr S, Miller SB. Acute oliguria. *New England Journal of Medicne* 1998; **338**: 671–5.

Kramer P, Wigger W, Rieger J, Mathaei D, Scheler F. Arteriovenous haemofiltration: a new and simple method for treatment of overhydrated patients resistant to diuretics. *Klinische Wochenschrift* 1977; **55**: 1121–2. [Early description of arteriovenous haemofiltration as a clinical tool.]

Palazzo M. Disturbances of acid–base and electrolyte balance. In: Garrard CS, Foex P, Westaby S, eds, *Principles and practice of critical care*, pp. 225–76. Blackwell Science, Oxford, 1997. [A wide ranging review of acid-base and electrolyte regulation and disturbances.]

Ronco C, Barbacini S, Digito A, Zoccali G. Achievements and new directions in continuous renal replacement therapies. *New Horizons* 1995; **3**: 708–16. [Review of methodology and techniques available for renal replacement therapy.]

Steiner RW. Interpreting the fractional excretion of sodium. *American Journal of Medicine* 1984; **77**: 699–702. [Pitfalls and strengths of using urinary sodium concentration in patients with renal impairment.]

Stevens PE, Riley B, Davies SP, Gower PE, Brown EA, Kox W. Continuous arteriovenous haemodialysis in critically ill patients. *Lancet* 1988; **ii**: 150–2.

Subach RA, Marx MA. Drug dosing in acute renal failure: the role of renal replacement therapy in altering drug pharmacokinetics. *Advances in Renal Replacement Therapy* 1998; **5**: 141–7. [A guide to drug therapy in patients with renal failure.]

Thadhani R, Pascual M, Bonventre JV. Acute renal failure. *New England Journal of Medicine* 1996; **334**: 1448–60.

Zarich S, Fang LST, Diamond JR. Fractional excretion of sodium: exceptions to its diagnostic value. *Archives of Internal Medicine* 1985; **145**: 108–12. [The use of urine sodium concentration in diagnosing the cause of oliguria.]

11.4 Respiratory aspects

Christopher S. Garrard

Acute lung injury and acute respiratory distress syndrome
Pathogenesis
Pathology
Pathophysiology
Management
Outcome
Mechanical ventilation
Indications for intubation and mechanical ventilation
Selection of airway access
Tracheostomy
Minitracheostomy
Cricothyroidotomy
Features and applications of a mechanical ventilator
Clinical monitoring of mechanical ventilation
Specific strategies in ventilator management
Complications of mechanical ventilation
Further reading

Acute lung injury and acute respiratory distress syndrome

Acute lung injury has a wide range of causes and a broad spectrum of severity. In its mildest form the only evidence for lung dysfunction may be an increase in the alveolar–arterial oxygen gradient (A–a) Do_2, without clinical signs of congestion, radiographic changes or a reduction in compliance. The more severe forms of acute lung injury are generally referred to as the acute respiratory distress syndrome. The syndrome was originally derived from adult respiratory disease until it became clear that children (but not premature infants) could develop the syndrome. The most common conditions associated with the development of this syndrome are listed in Table 1.

1. *Circulating factors*
Sepsis syndrome
Shock syndromes of any cause
Severe polytrauma
Pancreatitis
Blood transfusion
Post-cardiopulmonary bypass
Neurogenic pulmonary oedema
Drugs:
Heroin; Paraquat; aspirin; protamine; heparin
2. *Pneumonia*
Bacterial
Viral
Legionella
Protozoa
3. *Lung contusion*
Trauma
4. *Aspiration injury*
Gastric juice
Smoke
Chemicals
Drowning
5. *Embolization*
Amniotic fluid
Fat embolization

Table 1 Conditions associated with acute respiratory distress syndrome

The severity of the disease process depends upon the nature, severity, and duration of the insult to the lungs. The syndrome when associated with aspiration, fat embolism, pancreatitis, or sepsis may be more severe than that associated with a brief episode of hypovolaemic shock or a mismatched blood transfusion. In view of the heterogeneity of disease a simple scoring system, such as that shown in Table 2, should be used to quantify the severity of acute lung injury.

	Value
1. *Chest radiography score*	
No alveolar consolidation	0
Alveolar consolidation confined to one quadrant	1
Alveolar consolidation confined to two quadrants	2
Alveolar consolidation confined to three quadrants	3
Alveolar consolidation in all four quadrants	4
2. *Hypoxaemia score* (Pao_2 in mmHg/unity for Fi_{O_2}, divide by 7.5)	
Pao_2/Fi_{O_2} ≥ 300	0
Pao_2/Fi_{O_2} 225–299	1
Pao_2/Fi_{O_2} 175–224	2
Pao_2/Fi_{O_2} 100–174	3
Pao_2/Fi_{O_2} < 100	4
3. *PEEP score* (when ventilated)	
PEEP 5 cmH$_2$O	0
PEEP 6–8 cmH$_2$O	1
PEEP 9–11 cmH$_2$O	2
PEEP 12–14 cmH$_2$O	3
PEEP ≥ 15 cmH$_2$O	4
4. *Respiratory system compliance score* (when available)	
Compliance ≥ 80 ml/cmH$_2$O	0
Compliance 60–79 ml/cmH$_2$O	1
Compliance 40–59 ml/cmH$_2$O	2
Compliance 20–39 ml/cmH$_2$O	3
Compliance ≤ 19 ml/cmH$_2$O	4
The final value is obtained by dividing the aggregate sum by the number of components that were used	
No lung injury (score = 0)	
Mild to moderate lung injury (score = 0.1–2.5)	
Severe lung injury (score > 2.5)	

PEEP, positive end-expiratory pressure.

Table 2 Acute respiratory distress syndrome—lung injury score

The incidence of acute respiratory distress syndrome is uncertain, but it may affect 5 to 10 per cent of patients at risk. Clearly, the criteria by which the condition is defined will be reflected in the incidence, morbidity, and mortality reported in any particular series. The observation that mortality has changed little over the past two decades may indeed reflect a trend towards a stricter definition. It is evident that patients do not usually die of hypoxaemia but from the complex disturbances that result from multiple organ-system failure. The management of patients with this syndrome is therefore aimed at many facets of organ-system support.

The recognition of moderate to severe acute respiratory distress syndrome is not difficult. The patient has acute respiratory distress, requires an increasing inspired oxygen concentration (Fio_2) (greater than 0.5) to maintain a Pao_2 of more than 7.0 kPa (50 mmHg), and usually has extensive lung infiltrates on the chest radiograph. With mechanical ventilation, the lungs are stiff and require high inflation pressures. Calculated effective, static, and dynamic compliance (C_{dyn} and C_{stat}) are low, usually being less than 20 ml/cmH_2O (30 ml/mmHg). Cardiogenic pulmonary oedema should be excluded by measurement of the pulmonary arterial wedge pressure (< 18 mmHg). The Murray lung injury score will usually be more than 2.5 (Table 2) and calculated venous admixture reveals a shunt fraction of between 30 and 50 per cent.

There is no specific therapy at present for acute lung injury, except to suppress or remove the injurious agent, avoid aggravating the condition, and support the patient until lung function recovers. Acute respiratory distress syndrome may require prolonged periods of mechanical ventilation and can be associated with a mortality of more than 50 per cent. Above all, the clinician must anticipate its development, with the aid of an awareness of the precipitating causes (Table 1). A proactive approach to the management of these underlying disease processes may help prevent or ameliorate progression to respiratory failure.

Pathogenesis

The pathogenesis of acute respiratory distress syndrome is complex, being the product of several processes such as complement activation, neutrophil accumulation, platelet aggravation, and mediator release. In patients with trauma and sepsis, endotoxaemia is re-sponsible for activating complement and other components of the immune–inflammatory cascade. Complement activation causes sequestration of leucocytes in the pulmonary capillaries, which, in turn, injures endothelial cells through

the release of toxic superoxide radicals, arachidonic acid metabolites, cytokines, and proteases.

Unfortunately, monitoring complement activation appears to have little value in predicting the development or the clinical course of the syndrome. It is also clear that it can still develop in the most severely neutropenic patients. Studies with bronchoalveolar lavage have detected neutrophils, proteolytic enzymes, chemotactic factors, antiproteases, and cytokines in lung washings. Activation of the clotting system may further aggravate the pathophysiological disturbances by causing intravascular coagulation and platelet aggregation within the pulmonary vascular bed. Furthermore, the platelet and fibrin aggregates release vasoactive substances, such as serotonin, histamine, and prostaglandins, which, in combination with fibrinogen degradation products, damage the endothelium and pulmonary microvasculature.

Complicating the picture further is the possibility that supportive measures such as oxygen and positive-pressure ventilation may themselves promote or prolong lung injury.

Pathology

Adult respiratory distress syndrome is characterized by diffuse damage to alveolar epithelium and endothelium. In the early stages, the alveolar interstitium becomes infiltrated with inflammatory cells and the alveolar spaces filled with proteinaceous and haemorrhagic fluid. Hyaline membrane formation and capillary microembolism are common features. Leakage of plasma through the damaged pulmonary endothelium results in interstitial and alveolar oedema, and may alter the properties of surfactant. Alveolar type I cells are replaced by cuboidal, microvillous, type II cells, disturbing the architecture of the alveolar wall and exacerbating the effects of injury to the pulmonary capillaries. There may be progression to pulmonary fibrosis, with obliteration of the pulmonary alveolar and microvasculature architecture. Despite the extensive pathological changes that may be found on biopsy, good physiological recovery is possible.

Pathophysiology

The pathological changes result in reduced functional residual capacity, increased venous admixture (shunt), reduced lung compliance, and refractory hypoxaemia. The part played by abnormal surfactant in the pathophysiology of adult respiratory distress syndrome is unclear, but its severity seems to be related to the proportion of the abnormal surfactant. Bronchoalveolar lavage fluid obtained from patients with the syndrome contains aggregated and inactive surfactant. Unlike the experience with respiratory distress in the neonate, surfactant replacement in the adult condition has produced inconsistent results.

Management

The aim of respiratory support is to achieve adequate oxygen on-loading at the lungs without impeding cardiac output. To help maintain oxygen delivery, the haemoglobin concentration should be maintained above 11 g/100 ml of blood. Oxygen, intermittent positive-pressure ventilation, and positive end-expiratory pressure form the basis of respiratory support in acute respiratory distress syndrome. Although constant positive airway pressure can be delivered with a tight-fitting anaesthetic face-mask in the spontaneously breathing patient, endotracheal intubation is generally required. The endotracheal tube should be of the high-volume/low-pressure, cuffed variety and the cuff inflated to a pressure not exceeding 30 cmH_2O.

The ventilator selected to support the patient should be able to deliver the basic modes of ventilation: control mechanical ventilation, assist control, synchronized intermittent mandatory ventilation, and pressure support. The optimal mode is not known but selection should be based upon whichever achieves ventilation goals for the lowest peak and mean airway pressure.

Oxygen

Most patients will require a Fio_2 of more than 0.6 to achieve an acceptable Pao_2 (one that satisfies their individual requirements for oxygen delivery). For example, a patient on 100 per cent oxygen and 20 cmH_2O positive end-expiratory pressure with a haemoglobin concentration of 12 g/100 ml and a Pao_2 of 7.0 kPa (53 mmHg) might appear dangerously hypoxaemic. Yet this low Pao_2 may satisfy oxygen requirements provided the cardiac output is adequate. Oxygen extraction by the tissues drives the mixed venous oxygen content to lower than normal. The clinician must therefore be alert to evidence oftissue hypoxaemia, such as metabolic acidosis, mental obtundation, or cardiac arrhythmias. If such evidence is not present it may be reasonable to assume that oxygen delivery is adequate. If doubt exists about the adequacy of oxygen delivery, a flow-directed pulmonary arterial catheter can be inserted to measure cardiac output and mixed venous saturation, and to calculate oxygen delivery.

Positive end-expiratory pressure (PEEP)

This recruits lung units, partly to restore the reduced functional residual capacity to normal. Positive end-expiratory pressure of 5 to 20 cmH_2O (0.5–2.0 kPa) is added in increments of 5 cmH_2O until an acceptable Pao_2 is reached. Current practice encourages the use of the lowest pressure that will meet oxygenation goals. ‘Best' or ‘optimal' positive end-expiratory pressure is that which maximizes oxygen delivery with the lowest Fio_2, lowest venous admixture, and highest effective compliance. The term ‘positive end-expiratory pressure' is appropriate when the patient's ventilation is supported by intermittent mechanical ventilation. If ventilation is predominantly spontaneous, the term continuous positive airway pressure is usually adopted. Concern over pressure injury (barotrauma) has led to a more conservative approach to the use of positive end-expiratory pressure in recent years.

Ventilator strategies

The ventilation settings of the mechanical ventilator may be as crucial as the level of positive end-expiratory pressure in causing barotrauma. Conventional, high tidal-volume settings of 10 to 12 ml/kg may be inappropriate in the patient with acute respiratory distress syndrome and non-compliant lungs. High peak airway pressures place additional stress upon the alveolar wall, potentially leading to cystic changes in lung morphology. The degree of stress is more related to the changes in lung volume, such that the term ‘volutrauma' (instead of barotrauma) has been commonly applied to the harmful effects of mechanical ventilation. Increasing the ventilation rate and reducing the tidal volume will reduce peak airway pressures at the cost of a higher VD/VT and higher mean airway pressure. A deliberate policy of ‘permissive hypercapnia', allowing the Pco_2 to rise to between 7 and 9 kPa, can be achieved by reducing both the tidal volumes and ventilator rate. This results in lower mean and peak airway pressures. The attendant respiratory acidosis is usually well tolerated and ultimately compensated for.

Prone positioning

A feature of the patient with acute respiratory distress syndrome is the extensive, gravity-dependent lung collapse, which is best visualized by computed tomography. To a large degree this is a consequence of nursing the patient for prolonged periods in the supine position. Sitting the patient as upright as possible minimizes the volume of dependent lung. In recognition that dependent regions of the lung exhibit severe atelectasis and loss of volume, it is logical to alternate the patient between the prone and supine positions every 8 to 12 h. Up to half of patients with severe manifestations of the syndrome show improved oxygenation with such manoeuvres. The major drawback is that vascular access sites and airway access might be dislodged during the turning of the patient.

Non-conventional forms of ventilation such as high-frequency ventilation have potential advantages in supporting patients with acute respiratory distress syndrome. In randomized studies, peak airway pressures are lower but mean airway pressure and mortality are unchanged. Ventilation with an inverted inspiration:expiration ratio is claimed to reduce the positive end-expiratory pressure for a given Pao_2. However, the mean airway pressures are inevitably raised by such techniques.

Nitric oxide therapy

The addition of low concentrations of nitric oxide (2–10 parts/10^6 NO) to the inspired gas mixture will improve oxygenation in about 50 per cent of cases of acute respiratory distress syndrome. Care must be exercised with delivery systems to ensure correct doses, avoid excessive production of toxic metabolites (nitrogen dioxide and methaemoglobin). Although oxygenation can be improved with NO therapy, there is no evidence from randomized studies that survival is better.

Extracorporeal membrane oxygenation

Extracorporeal membrane oxygenation for respiratory failure in adults was abandoned over two decades ago following controlled, randomized investigation. With the availability of improved oxygenators there has been renewed interest in combining extracorporeal, partial removal of CO_2 and low-frequency conventional ventilation. Whether this approach offers significant advantages over techniques for permissive hypercapnia remains unproved, although there are enthusiasts who promote the use of extracorporeal membrane oxygenation in adults. Both techniques are aimed at reducing the deleterious effect of positive pressure upon the alveolar epithelium and may therefore hasten recovery from acute respiratory distress syndrome. In contrast, extracorporeal membrane oxygenation in premature infants with persistent

respiratory distress significantly improves survival.

Intravenous gas-exchange devices (IVOX®) have been developed, which may offer the safest alternative to conventional ventilation and gas-exchange techniques. These devices will require extensive evaluation before they replace or supplement conventional ventilation.

Surfactant

Following the dramatic benefit from the use of surfactant in neonatal respiratory distress, attempts to achieve similar effects in adults have been explored. Despite encouraging anecdotal reports, surfactant has not produced consistent improvement in acute respiratory distress syndrome, particularly when resulting from sepsis syndrome. This failure may stem in part from uncertainty over the type of surfactant (synthetic or animal-derived) and the dose.

Liquid ventilation

Alveolar instability and collapse in acute respiratory distress syndrome stems, in part, from the high surface tensions at the alveolar air/liquid interface. Von Neerguard demonstrated in the 1920s that a liquid–liquid interface in the lung would remove surface-tension effects. Application of this effect can be found in liquid lung ventilation that uses a liquid with a high carrying capacity for both oxygen and carbon dioxide instead of air/oxygen mixtures. Perfluorocarbons (carbohydrate molecules with the hydrogen elements replaced by fluorine) are inert liquids, which, when instilled into the lung, are capable of sustaining gas exchange. Most experience has been gained with infants initially sustained by extracorporeal membrane oxygenation. Much greater experience with this technique is required before more widespread use can be justified

Circulatory support

Fluid administration should be regulated to ensure adequate cardiac output without aggravating the pulmonary oedema. Measurement of pulmonary capillary wedge pressure using a flow-directed pulmonary arterial catheter may provide a more reliable assessment of the filling status of the patient than measurement of central venous pressure. The information provided by measures of cardiac output and derived indices of vascular resistance rationalizes the use of inotropes, pressors, and afterload reduction. Pulmonary hypertension is a not uncommon feature of acute respiratory distress syndrome, and although attempts to reduce the pulmonary arterial pressures have not reduced mortality, significant increases in oxygen delivery can be obtained with agents such as prostacyclin. Despite an increase in venous admixture with prostacyclin, the increases in cardiac output and mixed venous saturation produce enhanced oxygen delivery and improved Pao_2.

There is no strong evidence to favour colloidal over crystalloid replacement fluids. The volume of fluid is more critical: the type of fluid should be selected to ensure electrolyte equilibrium and defend the plasma colloid osmotic pressure. Diuretics should be used to reduce intravascular volume as needed. If the patient becomes oliguric, early haemofiltration will be needed to maintain fluid balance. Nutritional support in the form of enteral feeding or total parenteral nutrition can be started early in the course of acute respiratory distress syndrome.

Specific treatments to inhibit mediator cascades are largely experimental. Multicentre randomized studies have failed to show significant benefit from corticosteroids, non-steroidal anti-inflam-matory agents, or vasodilator prostaglandins. Some reports of high-dose steroids (40–60 mg prednisone/day) in ‘chronic', unresolved acute respiratory distress syndrome suggest some benefit in terms of oxygenation and lung compliance. However, the use of steroids in the acute phase afforded no benefit in multicentre, controlled studies.

Outcome

Provided that multiple organ-system support can be maintained, a positive attitude towards final recovery is justified. Even after periods of mechanical ventilation, a high Fio_2 and positive end-expiratory pressure for up to 3 to 4 months, a good functional outcome is possible. Biopsy-proven pulmonary fibrosis does not inevitably mean that there is fixed, irreversible pathology. A mortality rate of over 50 to 60 per cent is generally quoted for the last decade, although occasional reports have shown a mortality of about 20 per cent. In adults, the most common cause of acute respiratory distress syndrome is sepsis syndrome. In these patients, fatality is determined more by the effects of sepsis and its source than measures of the degree of lung injury (e.g. severity of hypoxia). Thus, it is sometimes said that patients die ‘with' rather than ‘of' acute respiratory distress syndrome. Pulmonary function testing 1 year after recovery may show a reduction in vital capacity with a mild obstructive defect. In many patients the only abnormality may be a reduction in carbon monoxide transfer.

Mechanical ventilation

The provision of efficient and safe mechanical ventilation is a skill that must be mastered by all physicians practising critical care. The basic principles still pertain despite the introduction of complex and sophisticated mechanical ventilators and the overabundance of studies claiming superiority of certain techniques over others. The application of common sense and sound physiological doctrine will serve better than devotion to an attractive technical innovation.

Indications for intubation and mechanical ventilation

Although mechanical ventilation is not to be undertaken lightly, since it is associated with much morbidity and some mortality, failure to intervene promptly can have catastrophic consequences for the patient. The indications for mechanical ventilation fall into two broad categories: inadequate alveolar ventilation with increasing Pco_2, and inadequate gas exchange with increasing alveolar–arterial oxygen gradient and arterial hypoxaemia. Guidelines for mechanical ventilation in acute respiratory failure are shown in Table 3. The physician must always exercise clinical judgement in the interpretation of these guidelines and anticipate problems before they arise. For example, one of the simplest criteria for mechanical ventilation is a respiratory rate of 35 breaths/min or more. If a patient with a respiratory rate of 30 breaths/min is clearly becoming fatigued, an early elective intubation is preferred to an emergency procedure an hour or so later. Similarly, a progressive fall in vital capacity in a patient with myasthenia gravis receiving full medication may indicate a need for ventilatory support, although the critical value of less than 15 ml/kg is not broached.



Table 3 Guidelines for introduction of mechanical ventilation

Selection of airway access

Endotracheal intubation will be the preferred technique in most cases. Orotracheal intubation is particularly suited to emergency intubation; nasotracheal intubation requires a little extra time. A coagulation defect or thrombocytopenia makes nasotracheal intubation inadvisable, owing to the risk of serious haemorrhage. Whatever technique is selected, intubation should be done in a safe and expeditious manner by the most experienced clinician available: this willusually be an anaesthetist or trained specialist in intensive care. Neuro-muscular relaxant drugs should only be used to facilitate intubation by experienced personnel. Complications of endotracheal intubation are due to occlusion or displacement of the tube, and airway trauma. The appropriate size of endotracheal tube for most adult males is 8 to 9 mm internal diameter; and for women, 7 to 8 mm. For children, a rough calculation using the child's age in years divided by 4, plus 4.0, will provide the tube internal diameter in millimetres. These smaller tubes are generally not cuffed.

It is essential that the endotracheal tube be securely anchored and the cuff inflation pressure restricted to less than 30 cmH_2O. The latter can be achieved by always

using high-volume, low-pressure cuffed tubes and allowing a small cuff leak during each ventilator cycle (minimal leak technique). Alternatively, cuff inflation pressures can be measured periodically with an aneroid manometer and adjusted accordingly. Contrary to popular belief, higher cuff pressures do not improve airway protection against aspiration but only serve to damage the tracheal mucosa and risk later stenosis.

Tracheostomy

Tracheostomy should replace endotracheal intubation for specific indications and not merely after the elapse of a predefined time interval. With modern endotracheal tubes and techniques, endotracheal intubation can be tolerated without permanent harm to the airway for months if necessary. It has been shown that most mucosal damage is caused in the first week of intubation, with little additional change thereafter.

However, much can be gained by the judicious selection of patients for tracheostomy either as the preferred primary route for airway access or as a replacement for endotracheal intubation. The common indications for replacement include the need for chronic or permanent ventilation, to help weaning off after previously failed attempts at extubation, to facilitate oral nutrition, or the presence of upper airway complications of endotracheal intubation. Indications for primary tracheostomy are considered elsewhere.

The same principles of management for cuff pressure apply to tracheostomy as to endotracheal tubes. Tracheostomy is associated with fewer but more serious complications than endotracheal intubation. These include tube displacement, pneumothorax, severe haemorrhage, and wound infection.

Minitracheostomy

Minitracheostomy tubes are 3.5- to 4.0-mm diameter, cuffless tubes inserted percutaneously through the cricothyroid membrane, usually under local anaesthesia. A Seldinger technique for the introduction of the minitracheostomy tube offers an alternative to the direct trochar method. Minitracheostomy allows suctioning of lung secretions without the need for formal endotracheal intubation or tracheostomy.

Cricothyroidotomy

A cricothyroidotomy may be preferred in life-threatening, upper airway obstructions where endotracheal intubation is not feasible and there is insufficient time to perform tracheostomy. Under local anaesthesia, a full-sized tracheostomy tube can be inserted (6–8 mm internal diameter) to facilitate mechanical ventilation. In emergency conditions, temporary oxygenation (30 min) can be aided by the intermittent insufflation of 100 per cent oxygen at 345 kPa [50 lbf/in^2 (p.s.i) (15 l/min)] via a 14-gauge cannula inserted through the cricothyroid membrane. Interruption of oxygen flow is regulated by a simple Y-piece adapter.

Features and applications of a mechanical ventilator

Most adult patients are supported on volume/time-cycled, pressure-limited ventilators (volume ventilator or flow generator). The volume/time-cycled ventilators deliver preset tidal volumes regardless of changes in lung compliance or impedance. The price paid for this desirable characteristic is that the inflation pressures must rise to overcome the mechanical load. To protect the patient against inadvertently high pressures, a pressure limit must be set. When this limit is reached the ventilator terminates inspiration regardless of the volume delivered and triggers an alarm.

Neonates and infants may be satisfactorily ventilated with time-cycled, pressure-limited devices (pressure ventilator or pressure generator). The pressure-limited paediatric ventilator offers simplicity and reliable ventilation, although the delivered tidal volume is difficult to measure. In the premature neonate these are not serious limitations and pressure-limited ventilation is the preferred technique.

Specifically designed, compact, lightweight ventilators, driven by cylinder oxygen and using fluid logic circuits, are available for transporting ventilator-dependent patients. They are pressure generators and can be used for both adults and children. By entraining air, a choice of either 60 per cent or 100 per cent oxygen is available.

Drive mechanisms

All ventilators possess a drive mechanism that propels the air/oxygen gas mixture into the patient. Some deliver the gas to the patient directly (single circuit) or indirectly through a dual circuit. There are several types of drive mechanisms, which determine the flexibility, the available ventilation modes, and the ability to deliver the prescribed tidal volume in the face of abnormal lung mechanics.

Schematics of a simple positive-pressure ventilator and ventilator circuit during inspiratory and expiratory phases are shown in Fig. 1 and Fig. 2.

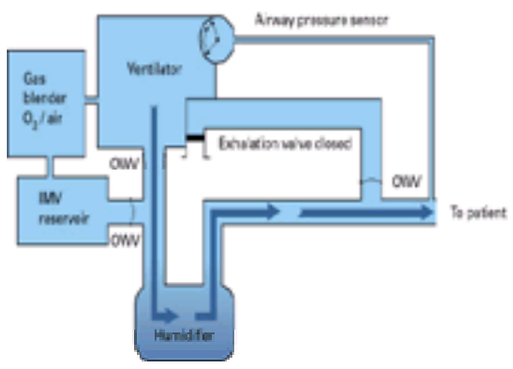


Fig. 1. Schematic of ventilator circuit during positive-pressure inflation of the lungs. The direction of gas flow is indicated by the broad arrows. One-way valves (OWV) assure unidirectional flow and the expiratory valve ensures lung inflation.

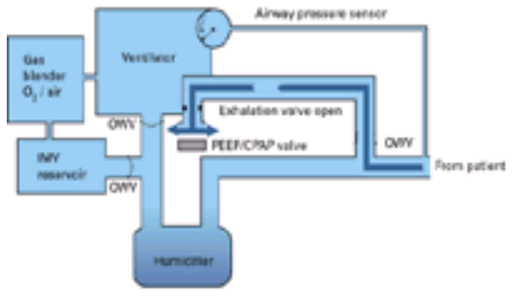


Fig. 2. Schematic of ventilator circuit during exhalation. Expiratory valve is opened permitting passive exhalation. The direction of gas flow is indicated by the broad arrows. Positive end-expiratory pressure/continuous positive airway pressure (PEEP/CPAP) valve partially occludes expiratory flow to produce the desired PEEP/CPAP level.

Control and monitoring mechanisms

The control mechanism that cycles the ventilator from inspiration to expiration may be electromechanical or electronic (using microprocessors). Modern mechanical

ventilators tend to fall into the latter category and offer a degree of sophistication that has greatly improved the safety and efficiency of mechanical ventilation.

Depending upon the indications for ventilation, the clinician must select the mode of ventilation, choose the ventilation variables, and adjust the ventilator alarms. The most commonly available ventilator modes include control mechanical ventilation, assist control (triggered ventilation), intermittent mandatory ventilation (synchronized or not), and pressure support, although some others are used.

Control mechanical ventilation

This provides time- and volume-cycled, pressure-limited breaths at preset rates, but does not allow the patient to breathe spontaneously. This mode is suitable for the paralysed or heavily sedated patient.

Assist-control or triggered ventilation

This synchronizes the ventilator to the patient's own respiratory rhythm, delivering a volume-preset, pressure-tidal volume. A trigger sensitivity must be selected (usually -0.5 to -2.0 cmH_2O) by which the patient can initiate volume preset breaths above the set rates. Patients have a tendency to hyperventilate on assist control. As a safety requirement, a high respiratory-rate alarm is needed and a ‘back up' ventilation rate must be set in the event of apnoea. Assist control is better tolerated than control mechanical ventilation and the patient requires less sedation.

Intermittent mandatory ventilation

This was originally devised for weaning but is now widely adopted as a maintenance mode (see Fig. 3). It provides the opportunity for the patient to breathe spontaneously and supplement the positive-pressure minute ventilation. A number of mandatory breaths (e.g. 6/min) are delivered by the ventilator. In between these mandatory breaths the patient is free to insert additional breaths according to physiological need. In the standard intermittent mandatory ventilation mode, there is a theoretical risk of stacking a ventilator breath on top of a spontaneous breath. However, this does not appear to be a significant problem and an appropriately set pressure limit should prevent this causing inadvertent overinflation of the lungs. More modern ventilators use the triggering or assist facility to synchronize the machine breaths with the patient's own spontaneous breathing pattern (synchronized intermittent mandatory ventilation). This technique is intended as partial ventilation support. With the patient taking spontaneous breaths, it is better tolerated than control mechanical ventilation, results in lower mean airway pressures, has less effect on the cardiovascular system, and allows the patient to regulate their own Pco_2 to at least some degree. Weaning is achieved by gradually reducing the number of breaths in synchronized intermittent mandatory ventilation and allowing the patient to contribute more of the minute ventilation.

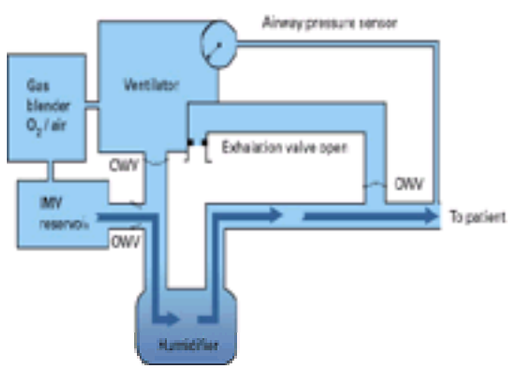


Fig. 3. Schematic of ventilator circuit during spontaneous inspiration in intermittent mandatory ventilation (IMV) mode. The IMV reservoir can consist of a distensible anaesthetic bag that is constantly filled from the air/oxygen blender and results in a continuous gas flow around the ventilator circuit (continuous-flow IMV). Alternatively, access to a reservoir of gas mixture at ambient pressure can be triggered by the patients own inspiratory effort (demand-valve IMV). The direction of gas flow is indicated by the broad arrows. Positive pressure-cycled breaths are triggered by the patient's own ventilatory effort but only up to the set mandatory rate (synchronized IMV).

Pressure support

This uses a triggering facility to deliver not a volume-preset breath as in assist control, but a pressure-limited breath (as with paediatric pressure ventilation). The inspiratory flow rate is usually high to minimize phase lag and the work of breathing. Pressure support may be used alone or in conjunction with synchronized intermittent mandatory ventilation, when it assists spontaneous breaths. Pressure support provides an efficient maintenance and weaning mode that is well tolerated by the patient.

Other modes of ventilation

Pressure-release, high-frequency and inverse inspiration:expiration ratio ventilation have their proponents. High-frequency ventilation in its several forms has been recommended for use following reconstructive laryngeal, tracheal, or bronchial surgery, or for patients with bronchopleural or bronchocutaneous fistulas. Even these applications may not offer much advantage, if any, over conventional modes. Mandatory minute ventilation is an innovative mode whereby the combined spontaneous and mechanical ventilation must reach a minimum preset level. As the patient's spontaneous ventilation increases, the mechanically assisted breaths become fewer. Individual ventilators vary in their ability to achieve successful mandatory minute ventilation. Bilevel positive airway pressure (BiPAP) is an efficient method of providing pressure-assisted ventilation. It is claimed that it is a more efficient mode of ventilation than pressure or flow-triggered pressure support. However, its efficiency probably lies in the efficient mechanics of the device providing the mode of ventilation rather than the mode itself. Evidence to indicate significant superiority of any of these modes over conventional methods of ventilation is not convincing.

Expiratory retard Restriction of the expiratory gas flow through a flow-dependent resistance prolongs expiration and delays the airway pressure drop to atmospheric. Expiratory retard was claimed to prevent premature airway closure in patients with chronic obstructive pulmonary disease and improve lung emptying. It is doubtful whether it contributes any additional benefit beyond that provided by positive end-expiratory pressure or continuous positive airways pressure, and it may increase the work of breathing.

Inspiratory pause Inspiratory pause or hold prolongs the inspiratory phase and delays the onset of expiration. As such it is thought to increase oxygenation by improved ventilation distribution and reduced *V*/*Q* mismatch. Like positive end-expiratory pressure and continuous positive airways pressure, it increases mean airway pressure, and may have cardiovascular effects. It is doubtful whether inspiratory pause contributes significantly to patient management.

Sighs Before the advent of high tidal-volume ventilation and positive end-expiratory pressures and continuous positive airways pressure, sighs were added to the ventilation protocol to prevent progressive atelectasis. Each sigh was delivered two to six times per hour and was equivalent to about twice the conventional tidal volume. The risks of barotrauma probably outweigh the theoretical benefits.

Ventilator variables

Once a ventilation mode has been selected (at least temporarily), ventilatory variables must be set before attaching the patient to the ventilator. The variables include:

1. tidal volume;
2. ventilation rate;
3. inspiratory/expiratory (I:E) ratio;
4. flow waveform;
5. Fio_2 (0.21 to 1.0);
6. pressure limit;

7. positive end-expiratory pressure/continuous positive airway pressure (0–20 cmH_2O).

Tidal volume

The delivered, inspiratory tidal volume may be set at 10 to 12 ml/kg body wt. This should be reduced if the patient has restrictive lung disease or has undergone lobectomy or pneumonectomy. Using respiratory rates of more than 10 breaths/min with such tidal volumes will provide full ventilatory support. If the patient is breathing spontaneously, an intermittent mandatory ventilation mode will be preferred at rates between 4 to 8 breaths/min. If assist-control or pressure support is chosen the respiratory rate will be the patient's spontaneous rate.

I:E ratio, inspiratory flow rate

The rate of inspiratory to expiratory time (**I:E** ratio) will generally range from 1:2 to 1:4. This provides sufficient time for full passive exhalation. In patients with obstructive lung disease, failing to allow adequate time for exhalation results in hyperinflation (auto- or intrinsic positive end-expiration pressure). The higher the set respiratory rate the shorter expiration becomes and the I:E ratio falls. This may lead to the paradoxical situation in the patient with chronic obstructive pulmonary disease where the Pco_2 rises as the ventilator rate is increased.

The I:E ratio can be adjusted in several ways depending upon the make of ventilator. In some, a ratio can be selected directly, while in others the inspiratory flow rate determines the duration of inspiration. An acceptable range for inspiratory flow rates is between 30 and 60 l/min (0.5–1.0 l/s).

Inspiratory waveforms

Many volume- and time-cycled (flow generator) ventilators allow the choice of several waveforms. Although there is little evidence to favour one over the other, a square waveform delivers the tidal volume in the least time and with higher peak pressures. A decelerating flow pattern results in lower peak pressures, longer inspiratory intervals, and lower I:E ratios.

Inspired oxygen concentration

This should be constantly adjusted to provide adequate arterial oxygenation without hyperoxia. Too high a Fio_2 is frequently the cause of failure to wean patients with chronic obstructive pulmonary disease from a mechanical ventilator.

Pressure limit

Setting a pressure limit about 10 cmH_2O above the peak pressure reached during each ventilator cycle protects patients against inadvertently high pressures experienced during coughing or straining. Hitting the pressure limit terminates inspiration and sounds an alarm.

Positive end-expiratory pressure/continuous positive airways pressure (PEEP/CPAP)

Maintaining airway pressure above barometric pressure in a spontaneously breathing patient is called constant positive airway pressure. The same pressure applied to a patient on intermittent positive pressure ventilation is called positive end-expiratory pressure. This technique is used to correct lung volume (functional residual capacity) in conditions characterized by reduced lung volume such as acute respiratory distress syndrome or cardiogenic pulmonary oedema. It may also be of benefit in patients with flail chest segments since it acts to splint the chest wall.

Positive end-expiratory pressure/continuous positive airways pressure is achieved by the inclusion of a resistance at the expiratory end of the breathing circuit. Ideally this resistance should be as close to a threshold resistor as possible, such as an underwater column. In practice most of the valves produce some flow-dependent retardation of expiration, which increases the work of breathing in a spontaneously breathing patient.

Bypassing the oropharynx by an endotracheal tube is known to cause a fall in functional residual capacity in both adults and children. The application of low continuous positive airways pressure (3–5 cmH_2O) reverses this effect and has therefore been suggested as part of routine management of the intubated patient (‘physiological continuous positive airways pressure'). The usual indication for positive end-expiratory pressure is the presence of refractory hypoxaemia due to acute lung injury. Starting at 5 cmH_2O the pressures are increased progressively until satisfactory oxygenation is achieved for a Fio_2 ideally less than 0.6. It is rarely necessary to exceed 20 cmH_2O. The effects of such treatment can be assessed in several ways by calculating the venous admixture or shunt fraction, oxygen delivery or effective static compliance of the lungs. Clearly, following the Pao_2 alone takes no account of the effects of positive pressure upon cardiac output. Continuous measurement of mixed venous Sao_2 with a suitable flow-directed pulmonary artery catheter is a particularly good method of evaluating the response to a change in airway pressure.

Intrinsic positive end-expiratory pressure (PEEPi)

Intrinsic positive end-expiratory pressure is the raising (usually unintentionally) of alveolar pressure above atmospheric pressure consequent upon the interactions of lung airways' time constants and the pattern of mechanical ventilation. It is most likely to occur in patients with airways obstruction (chronic obstructive pulmonary disease and asthma) in whom the expiratory interval is too short to allow full expiration back to functional residual capacity. The effect is to produce air trapping, elevation in end-expiratory and mean airway pressure, and adverse effects on the pulmonary and systemic circulation. Attempts to reduce intrinsic positive end-expiratory pressure by adding extrinsic (external) positive end-expiratory pressure to splint the airways open appear to be largely ineffective.

Ventilator monitors and alarms

The ventilation monitors and alarms that must be set and maintained include:

1. exhaled tidal volume, exhaled minute ventilation;
2. airway pressure/disconnect;
3. peak and mean airway pressures;
4. Fio_2;
5. inhaled gas temperature.

Exhaled volume and minute ventilation

The ability to measure exhaled tidal volume and minute ventilation is a great asset in evaluating the efficiency of ventilation. By setting low-limit alarms the safety of mechanical ventilation is greatly enhanced and supplements the airway pressure/disconnect alarm (Fig. 1(d)).

Airway pressures

Careful monitoring of peak and mean airway pressures is essential in the patient supported with a flow-generating (volume/time-cycled) ventilator. A step increase in peak pressure may indicate a change in lung compliance, airway resistance, displacement of the endotracheal tube into a main bronchus, or even a pneumothorax. The pressure alarm should therefore be set 5 to 10 cmH_2O above the normal peak pressure (Fig. 4).

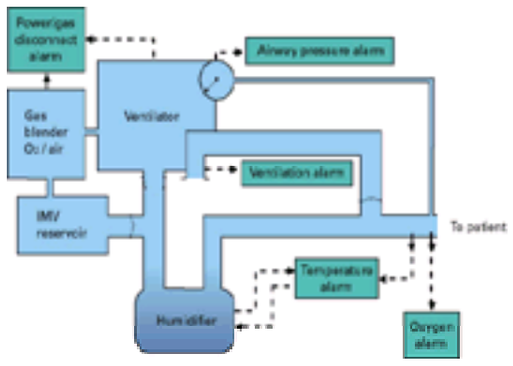


Fig. 4. Schematic of ventilator circuit showing minimum alarm/monitoring facilities on a mechanical ventilator. The pressure alarm monitors inflation pressures (peak and mean) and also detects patient disconnection, which results in the circuit pressure falling to and remaining at zero. Ventilation alarms monitor the expiratory tidal volume and minute ventilation, providing confirmation that the ventilation objectives are being met. The ventilation alarms will also detect disconnection or apnoea. The temperature of inhaled gases is monitored within the humidifier and at the patient's endotracheal tube. Excessively high temperature causes an alarm and automatically downregulates the humidifier heater.

Inspired oxygen concentration

An oxygen sensor just proximal to the endotracheal tube ensures that any sudden fall in Fio_2 is detected immediately and before the patient can desaturate.

Gas temperature

Strict control of the temperature of the inhaled gases is essential to avoid thermal injury and to ensure effective humidification of inspired gases. Modern humidifiers are usually of the ‘wick' variety that efficiently humidify high gas flows. Heat/moisture exchangers placed between the ventilator circuit Y-piece and the endotracheal tube provide an alternative method of humidification, particularly for short-term ventilation. Inspired gas should leave the humidifier at about 40°C and reach the endotracheal tube at close to 37°C (Fig. 4). A low-voltage heating element can be incorporated into the inspiratory limb of the circuit to reduce gas cooling and water condensation.

Clinical monitoring of mechanical ventilation

An essential aspect of monitoring is regular clinical examination of the patient, and inspection of the ventilator and ventilator circuit. Expansion of the chest should be symmetrical with each ventilator-cycled breath (control mechanical ventilation, intermittent mandatory ventilation), assisted breath (assist control or pressure support), or unassisted spontaneous breath (intermittent mandatory ventilation). Auscultation should confirm air entry and detect any added sounds. The patient should be sat up or rolled side to side to allow inspection of the whole of the chest. The endotracheal tube should be secure and as comfortable as possible for the patient. The endotracheal cuff pressure should be checked (< 30 mmHg) or a small leak should be audible with a stethoscope on the side of the neck. The ventilator circuit should feel warm but be free of significant amounts of condensed water. The humidifier temperature and water level should be checked.

The pulse oximeter has contributed significantly to the monitoring and safety of patients on mechanical ventilation. It not only provides a continuous measurement of oxygenation but also reduces the need for arterial blood-gas sampling.

Much can be appreciated from watching the ventilator pressure gauge with each cycle. In addition to evaluating peak inspiratory pressure the clinician will be able to judge whether the patient is ‘fighting' the ventilator. Comparing inspiratory and expiratory tidal volumes may indicate a leak in the circuit, either at circuit connections or at the endotracheal tube cuff. When peak pressures are high, the internal compliance of the ventilator and circuit (about 2–2.5 ml/cmH_2O) may account for much of the volume loss. A rough assessment of the compliance of the lung can be made by following the peak inflation pressures. However, it is preferable to calculate effective static and dynamic compliance in the following way.

Effective dynamic compliance

$$C_{dyn} = \frac{V_T exh - ([P_{peak} - PEEP] \times 2.0)}{P_{peak} - PEEP}$$

Effective static compliance

$$C_{stat} = \frac{V_T exh - ([P_{pause} - PEEP] \times 2.0)}{P_{pause} - PEEP}$$

where C_{dyn} = effective dynamic compliance (ml/cmH_2O), C_{stat} = effective static compliance (ml/cmH_2O), VT_{exh} = exhaled tidal volume (ml), P_{peak} = peak airway pressure (cmH_2O), P_{pause} = inspiratorypause pressure (cmH_2O), PEEP = positive end-expiratory pressure (cmH_2O), and the figure 2.0 represents the internal compliance (ml/cm H_2O) of the ventilator and circuit. The product P_{peak} × 2.0 indicates the volume of gas contained within the ventilator circuit that does not enter the patients lungs but is measured as exhaled *V*T.

The effective dynamic compliance is always less than the effective static compliance by about 10 ml/cmH_2O. Greater differences indicate airways obstruction is contributing to the peak airway pressure. Normal values for effective dynamic compliance range between 50 and 80 ml/cmH_2O, depending upon the size of the patient. Dynamic compliance will fall with both bronchoconstriction or lung restriction as in acute respiratory distress syndrome. Static compliance is relatively unaffected by bronchoconstriction. In severe acute respiratory distress syndrome, static compliance may fall to as low as 10 to 20 ml/cm H_2O from the normal range (60–90 ml/cmH_2O).

Specific strategies in ventilator management

Restrictive lung disease

Patients with restrictive lung diseases such as sarcoidosis or fibrosing alveolitis should be ventilated with small tidal volumes of between 5 and 8 ml/kg at rates of 15 to 20 breaths/min. Oxygen need not be restricted in the manner recommended for patients with chronic obstructive pulmonary disease.

Chronic obstructive pulmonary disease

Although most patients with acute or chronic respiratory failure can be managed successfully without mechanical ventilation, a small proportion fail to respond to conservative measures and require ventilatory assistance. In many cases, the need for mechanical ventilation is the direct result of injudicious oxygen therapy. Low-rate, synchronized intermittent mandatory ventilation (6–8 breaths/min) or low levels of pressure support are ideal for such patients with acute on chronic respiratory failure. The $Paco_2$ should be reduced very slowly towards, but not to, normal. The Fio_2 rarely needs to be higher than 0.35. High ventilator rates (> 14/min) are associated with high values of *V*D/*V*T (> 0.5). Paradoxically, as the ventilator rates are increased in an attempt to increase minute ventilation, the $Paco_2$ may rise. To avoid intrinsic or auto-positive end-expiratory pressure, the I:E ratio should be maintained at 1:2 or higher. Weaning can begin as soon as the precipitating cause of respiratory failure has been corrected. Weaning will be unsuccessful if there is any underlying metabolic alkalosis or if the patient receives sedative or analgesic agents. The $Paco_2$ can be allowed to rise slowly to above normal, provided sufficient time is given for the blood pH to correct and the Fio_2 is kept below 0.35. Carbon

dioxide production can be minimized by providing balanced nutrition, with calories being provided by both lipid and carbohydrate.

Asthma

Probably fewer than 1 per cent of acute severe asthma attacks require mechanical ventilation. However, it is apparent that some patients suffer cardiac arrest and die each year because intubation and mechanical ventilation was not done in time. Hypercarbia alone is generally insufficient indication for ventilation, but a combination of a rising $Paco_2$, fatigue, failure of conservative measures, or arrhythmias does call for elective intubation and mechanical ventilation.

Since asthma has a prevalence of between 4 and 8 per cent in the United Kingdom and the United States of America, it is not uncommon for a postoperative patient on a mechanical ventilator to develop bronchospasm. Asthmatic patients are difficult to ventilate and usually require high inflation pressures. Hypoxaemia may persist despite the addition of high concentrations of oxygen and is the result of mucous plugging of the airways. A philosophy of ‘permissive hypercapnia' or ‘controlled hypoventilation' should be adopted, with the $Paco_2$ remaining elevated (7–8 kPa, 50–70 mmHg). Lower tidal volumes and respiratory rates are therefore possible. Lower inspiratory flow rates result in lower peak pressures and reduced risk of barotrauma. Deaths in ventilated asthmatic patients are usually the result of barotrauma, hypotension in volume-depleted patients, arrhythmias, or lung infection.

Maximal bronchodilator therapy, including corticosteroids, is continued throughout the period of mechanical ventilation, supplemented if necessary with inhalational anaesthetics such as isoflurane or the intravenous anaesthetic ketamine. The inhalational anaesthetics act upon bronchial smooth muscles in several ways but predominantly via calcium channels. In low concentrations (e.g. 1 per cent halothane or isoflurane), rapid bronchodilation occurs such that inflation pressures fall in association with a fall in PCO_2 and a rise in pH. Hypotension will occur if the patient has not been adequately fluid-resuscitated. Rehydration and adequate humidification of inspired gases will usually mobilize secretions and mucous plugs; if not, bronchoalveolar lavage may be indicated.

The use of extracorporeal membrane oxygenation and CO_2 removal has been reported in acute asthma. These must be considered exceptional cases and such techniques cannot be generally recommended. Indeed, a perceived need for extracorporeal membrane oxygenation may grow from improper ventilator management.

Bronchopleural/bronchocutaneous fistulas

Although these fistulas can occur after trauma or lung infection, many arise during the postoperative period following lobectomy or pneumonectomy. It is generally appreciated that early weaning and extubation is preferred in these patients. Occasionally, postoperative complications necessitate a longer period of ventilation when there is significant risk of dehiscence of the bronchial stump. To reduce the risk of this, low tidal-volume, high respiratory-rate ventilation should be adopted to minimize inflation pressures. High-frequency ventilation would appear to be ideally suited to the prevention of bronchopleural fistula, although evidence to prove superiority over conventional ventilation is lacking.

The development of bronchopleural fistula is heralded by clinical deterioration, reduced movement of the chest wall on the affected side, tracheal deviation, and a sudden increase in inflation pressures. Emergency tube thoracostomy must be performed, converting the bronchopleural fistula into a bronchocutaneous fistula. Compensation for the loss of tidal volume through the fistula is easily made by adjusting the ventilator but if the leak is large, endobronchial intubation may be necessary. Bronchopleural and bronchocutaneous fistulas are unlikely to close until the patient is weaned from the ventilator.

Weaning

More than 80 per cent of patients who are ventilated postoperatively can be weaned simply by clinically evaluating their spontaneous ventilation on a ‘T-piece' or similar circuit. The remainder require a progressive reduction in ventilatory support until the ventilation variables can be measured. These variables include the negative inspiratory force and vital capacity. A negative inspiratory force greater than -25 cmH_2O or a vital capacity greater than 10 ml/kg usually indicates sufficient ventilatory reserve for spontaneous ventilation. These variables cannot be applied to patients with severe chronic obstructive pulmonary disease, in whom blood gases have to be followed with each reduction in ventilation support. Failure to wean a patient successfully should prompt the questions in Table 4.

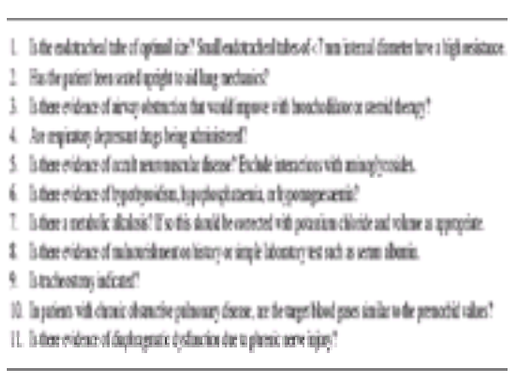

1. Is the endotracheal tube of optimal size? Small endotracheal tubes of <7 mm internal diameter have a high resistance.
2. Has the patient been sat upright to aid lung mechanics?
3. Is there evidence of airway obstruction that would improve with bronchodilator or steroid therapy?
4. Are respiratory depressant drugs being administered?
5. Is there evidence of occult neuromuscular disease? Exclude interactions with aminoglycosides.
6. Is there evidence of hypothyroidism, hypophosphataemia, or hypomagnesaemia?
7. Is there a metabolic alkalosis? If so this should be corrected with potassium chloride and volume as appropriate.
8. Is there evidence of malnourishment on history or simple laboratory test such as serum albumin.
9. Is tracheostomy indicated?
10. In patients with chronic obstructive pulmonary disease, are the target blood gases similar to the premorbid values?
11. Is there evidence of diaphragmatic dysfunction due to phrenic nerve injury?

Table 4 Questions to ask when weaning is difficult

Intermittent mandatory ventilation or pressure-support modes are very suitable for a gradual weaning process. Assist-control can also be used to wean patients by progressively reducing the tidal volume.

Complications of mechanical ventilation

Several complications of mechanical ventilation can be attributed to the local effects of the endotracheal tube upon the airway. These include airway obstruction due to tube displacement and pressure necrosis leading injury to the vocal cords and subglottic stenosis. The risk of nosocomial pneumonia is increased in the intubated patient.

Other complications are the direct consequence of positive-pressure ventilation. Haemodynamic effects such as reduced cardiac output, reduced renal perfusion, and salt and water retention are primarily the result of mechanical, neuroreflex, and humoral factors.

Interstitial lung damage may occur due to positive-pressure ventilation and this has prompted renewed interest in extracorporeal systems for the management of patients with acute respiratory distress syndrome. However, the greatest concern is for the risk of pneumothorax, pneumomediastinum, pneumopericardium, or subcutaneous emphysema. Pneumothorax is the most feared complication because it is associated with rapid deterioration unless dealt with quickly. Tube thoracostomy is mandatory since progression to a tension pneumothorax is very likely. Prophylactic thoracostomy tubes are not recommended, even in the presence of pneumomediastinum. Sudden clinical deterioration associated with a rise in inflation pressures and absence of breath sounds should raise the question of pneumothorax. Emergency decompression with a 14-gauge cannula may produce temporary relief and may have a diagnostic role, but tube thoracostomy should be performed without delay and without radiographic confirmation if necessary. Blunt dissection through the parietal pleura with forceps and digital exploration of the pleural spacebefore inserting the thoracostomy tube is essential if lung damage is to be avoided. Thoracostomy tubes with rigid metal stylets must not be used under any circumstances.

Further reading

Ashbaugh DG, Bigelow DB, Petty TL, Levine BE. Acute respiratory distress syndrome in adults. *Lancet* 1967; **ii**: 319–23. [First description of the syndrome of ARDS.]

Barnes PK. Principles of lung ventilators and humidification. In: Scurr, C. Feldman S, ed. *Scientific foundations of anaesthesia*, pp. 533–43. Heinemann, London, 1982. [Practical guide to mechanical ventilation.]

Bernard GR *et al.* High dose corticosteroids in patients with the adult respiratory distress syndrome. *New England Journal of Medicine* 1987; **317**: 1545–70.

Cameron PD, Oh TE. Newer modes of mechanical ventilatory support. *Anesthesia and Intensive Care* 1986; **14**: 258–66. [Comprehensive review of different modes of ventilation and their application.]

Cuthbertson BH *et al.* UK guidelines for the use of inhaled nitric oxide therapy in adult ICUs. American–European Consensus Conference on ALI/ARDS. *Intensive Care Medicine* 1997; **23**(12): 1212–18.

Danek SJ, Lynch JP, Weg JG, Dantzker DR. The dependence of oxygen uptake on oxygen delivery in the adult respiratory distress syndrome. *American Review of Respiratory Disease* 1980; **122**: 387–95.

Dellinger RP *et al.* Effects of inhaled nitric oxide in patients with acute respiratory distress syndrome: results of a randomized phase II trial. Inhaled Nitric Oxide in ARDS Study Group. *Critical Care Medicine* 1998; **26**(1): 15–23.

Downs JB, Olsen GN. Pulmonary function following adult respiratory distress syndrome. *Chest* 1974; **65**: 92–3.

Downs JB, Block AJ, Vennum KB. Intermittent mandatory ventilation in the treatment of patients with chronic obstructive pulmonary disease. *Anesthesia and Analgesia* 1974; **53**: 437–43. [Use of intermittent mandatory ventilation as a weaning technique for patients with COPD.]

Downs JB, Kelin EF, Desautels D, Modell JH, Kirby RR. IMV: A new approach to weaning patients from mechanical ventilators. *Chest* 1973; **64**: 331–5. [Introduction of intermittent mandatory ventilation as a weaning mode.]

Doyle RL, Szaflarski N, Modin GW, Wiener-Kronish JP, Matthay MA. Identification of patients with acute lung injury. Predictors of mortality. *American Journal of Respiratory and Critical Care Medicine* 1995; **152**(6; Pt 1): 1818–24.

Duchateau J *et al.* Complement activation in patients risk of developing the adult respiratory distress syndrome. *American Review of Respiratory Disease* 1984; **130**: 1058–64. [Demonstration of complement activation and white cell sequestration in the lung in ARDS.]

Elliott CG, Morris AH, Cengiz M. Pulmonary function and exercise gas exchange in survivors of adult respiratory distress syndrome. *American Review of Respiratory Disease* 1981; **123**: 455–92.

Finfer SR, Garrard CS. Ventilatory support in asthma. *British Journal of Hospital Medicine* 1993; **49**(5): 357–60. [Review of mechanical ventilation in respiratory failure in acute asthma.]

Fridrich P, Krafft P, Hochleuthner H, Mauritz W. The effects of long-term prone positioning in patients with trauma-induced adult respiratory distress syndrome. *Anesthesia and Analgesia* 1996; **83**(6): 1206–11.

Gattinoni L *et al.* Treatment of acute respiratory failure with low frequency positive-pressure ventilation and extracorporeal removal of CO_2. *Lancet* 1980; **ii**: 292–4. [Method of reducing potential lung damage from mechanical ventilation.]

Hartog A, Gommers D, Lachmann B. Role of surfactant in the pathophysiology of the acute respiratory distress syndrome (ARDS). *Monaldi Archives of Chest Diseases* 1995; **50**(5): 372–7.

Haynes JB, Hyers TM, Giclas PC, Franks JJ, Petty TL. Elevated fibrinogen degradation products in adult respiratory distress syndrome. *American Review of Respiratory Disease* 1980; **122**: 841–7.

Hickling KG. Extracorporeal CO_2 removal in severe adult respiratory distress syndrome. *Anesthesia and Intensive Care*, 1986; **14**: 45–53.

Hickling KG, Henderson, SJ, Jackson R. Low mortality associated with low volume pressure limited ventilation with permissive hypercapnia in severe adult respiratory distress syndrome. *Intensive Care Medicine* 1990; **16**: 372–7. [Article popularizing permissive hypocapnia in ARDS patients.]

Hirschl, RB, Tooley R, Parent A, Johnson K, Bartlett RH. Evaluation of gas exchange, pulmonary compliance, and lung injury during total and partial liquid ventilation in the acute respiratory distress syndrome. *Critical Care Medicine* 1996; **24**(6): 1001–8.

Hudson LD, Milberg JA, Anardi D, Maunder RJ. Clinical risks for development of the acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine* 1995; **151**(2; Pt 1): 293–301.

Hurst JM, Branson RD, Davis KJ, Barrette RR, Adams, KS. Comparison of conventional mechanical ventilation and high-frequency ventilation. A prospective, randomized trial in patients with respiratory failure. *Annals of Surgery* 1990; **211**: 486–91.

Kariman K, Burns SR. Regulation of tissue oxygen extraction is disturbed in adult respiratory distress syndrome. *American Review of Respiratory Disease* 1985; **132**: 109–14.

Kirby RR *et al.* High level PEEP in acute respiratory insufficiency. *Chest* 1975; **67**: 156–63. [Use of PEEP levels in excess of 20 cmH_20 to improve gas exchange.]

Kumar A *et al.* Continuous positive-pressure ventilation in acute respiratory failure. *New England Journal of Medicine* 1970; **283**: 1430–6. [Use of CPAP in adults with ARDS.]

Leach, CL *et al.* Partial liquid ventilation with perflubron in premature infants with severe respiratory distress syndrome. The LiquiVent Study Group. *New England Journal of Medicine* 1996; **335**(11): 761–7. [Demonstration of ability to support gas exchange in infants with partial liquid ventilation.]

Meduri GV, Headley AS, Golden E *et al.* Effect of prolonged methylprednisolone therapy in unresolving acute respiratory distress syndrome: a randomised trial. *Journal of the American Medical Association* 1999; **280**: 159–65

Montgomery AB, Stager MA, Carrico J, Hudson LD. Causes of mortality in patients with the adult respiratory distress syndrome. *American Review of Respiratory Disease* 1985; **132**: 485–9. [Sepsis is the leading cause of death in ARDS.]

Morris AH *et al.* Randomized clinical trial of pressure-controlled inverse ratio ventilation and extracorporeal CO_2 removal for adult respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine* 1994; **149**(2; Pt 1): 295–305. [No significant improvement in survival with extracorporeal CO_2 removal.]

Murray JF, Matthay MA, Luce JM, Flick MR. An expanded definition of the adult respiratory distress syndrome. *American Review of Respiratory Disease* 1988; **138**: 720–3.

National Heart Lung and Blood Institute, Division of Lung Diseases. *Extracorporeal support for respiratory insufficiency*, pp. 243–5. National Institutes for Health, Bethesda, MD, 1979.

Ognibene FP *et al.* Adult respiratory distress syndrome in patients with severe neutropenia. *New England Journal of Medicine* 1986; **315**: 547–51.

Pepe PE *et al.* Clinical predictors of adult respiratory distress syndrome. *American Journal of Surgery* 1982; **144**: 124–30.

Pepe PE, Hudson LD, Carrico CJ. Early application of positive end expiratory pressure in patients at risk for the adult respiratory distress syndrome. *New England Journal of Medicine* 1984; **311**: 281–6.

Petty TL, Ashbaugh DG. The adult respiratory distress syndrome: clinical features, factors influencing prognosis and principles of management. *Chest* 1971; **130**: 66–71.

Petty TL, Silvers GW, Paul GW, Stanford RE. Abnormalities in lung elastic properties and surfactant function in adult respiratory distress syndrome. *Chest* 1979; **75**: 571–4.

Pontoppidan H, Geffin B, Lowenstein E. Acute respiratory failure in the adult. *New England Journal of Medicine* 1972; **287**: 690–9. [Criteria for defining need for mechanical ventilation in acute respiratory failure.]

Qvist J, Pontoppidan H, Wilson RS, Lowenstein E, Laver MB. Haemodynamic responses to mechanical ventilation with PEEP: the effect of hypervolaemia. *Anesthesiology* 1975; **42**: 45–55.

Rinaldo JE. Mediation of ARDS by leukocytes. Clinical evidence and implications for therapy. *Chest* 1986; **89**: 590–3. [Review of the role of leucocytes in the pathogenesis of ARDS.]

Rinaldo JE, Rogers RM. Adult respiratory distress syndrome. *New England Journal of Medicine* 1986: **315**: 578–80. [Editorial on ARDS.]

Rocker GM, Wiseman MS, Pearson D, Shale DJ. Diagnostic criteria for adult respiratory distress syndrome: time for reappraisal. *Lancet* 1989; **i**: 120–3. [Elastase-antitrypsin complex plasma concentration higher in ARDS patients than in controls.]

Rotman HH, Lavelle TF, Dimcheff DG. Vendenbelt RJ, Weg JG. Long term physiological consequences of the adult respiratory distress syndrome. *Chest* 1977; **72**: 190–2.

Smith RA, Desautels DA, Kirby RR. Mechanical ventilators. In: Kirby RR, Smith RA, Desautels DA, ed. *Mechanical ventilation*, pp. 327–474. Churchill Livingstone, New York, 1985.

Stocker R, Neff T, Stein S, Ecknauer E, Trentz O, Russi E. Prone positioning and low-volume pressure-limited ventilation improve survival in patients with severe ARDS. *Chest* 1997; **111**(4): 1008–17. [Low mortality in ARDS attributed to ventilation techniques and prone positioning.]

Tate RM, Repine JE. Neutrophils and the adult respiratory distress syndrome. *American Review of Respiratory Disease* 1983; **128**: 522–9.

Weiland JE *et al.* Lung neutrophils in the adult respiratory distress syndrome. Clinical and pathophysiological significance. *American Review of Respiratory Disease* 1986; **133**: 218–25.

Weinberg P *et al.* Biologically active products of complement and acute lung injury in patients with the sepsis syndrome. *American Review of Respiratory Disease* 1984; **130**: 791–6.

Wood LH, Prewitt RM. Cardiovascular management in acute hypoxemic respiratory failure. *American Journal of Cardiology* 1981; **47**: 963–72.

Zapol WM *et al.* Extracorporeal membrane oxygenation in severe acute respiratory failure. A randomized prospective study. *Journal of the American Medical Association* 1979; **242**(20): 2193–6. [No benefit demonstrated from the use of extracorporeal membrane oxygenation.]

11.5 Infection in the intensive care unit

Christopher S. Garrard

Nosocomial infection in patients in the intensive care unit
Incidence
Predisposing factors
Infection control
Identification of organisms
Nosocomial pneumonia
Measures to reduce the risk of nosocomial pneumonia
Management
Line-related infection
Urosepsis
Cellulitis
Necrotizing fasciitis
Infection in the immunocompromised host
Failure of phagocytosis
Complement deficiency
Granulocytopenia
Cellular immunodeficiency
Defective humoral immunity
Splenectomy
Sites of infection and organisms
Prevention of infection
Microbiological surveillance
Antibiotic treatment
Fungal infection
Pneumocystis carinii
Toxoplasma gondii
Viral infections
Antibiotic therapy in intensive care
b-Lactams
Penicillins
Cephalosporins
Carbapenems
Monobactams
Aminoglycosides
Quinolones
Anti-anaerobic antibiotics
Antibiotic resistance
Further reading

Nosocomial infection in patients in the intensive care unit

Incidence

Nosocomial infection is one that develops at least 48 h after admission to hospital. It affects about 15 to 40 per cent of patients in the critical care unit, and contributes to approximately 75 per cent of late deaths in such units. The terms colonization, infection, and sepsis are often used interchangeably, but should be used only within the limits of the following definitions. Colonization is defined microbiologically as the presence of a potentially pathogenic organism on two or more consecutive occasions. In contrast, infection is the presence of many pathogenic organisms, with leucocytes and the clinical signs of inflammation. Sepsis is a syndrome comprising the clinical signs of infection (fever and leucocytosis) and organ system failure. An infecting organism may or may not be identified. Wider aspects of the sepsis syndrome are considered in more detail elsewhere.

Predisposing factors

The incidence of nosocomial infection is dependent upon the length of stay in the critical care unit, the number of invasive procedures or degree of intervention, the nature of the underlying illness, and mode of presentation. These factors are to some extent interdependent: the longer a patient remains in the critical care unit the more interventions and procedures are likely to be performed.

The impact of the underlying illness and the type of patient can be judged by examining the different rates of infection in medical and surgical intensive care units. Burns and general surgical critical care units have infection rates of about 30 per cent, compared with less than 15 per cent in general medical units. Coronary care units have rates of infection of 5 per cent, which is similar to that of the general wards.

The performance of invasive procedures has a great effect upon the incidence of nosocomial infection. Endotracheal intubation and mechanical ventilation carries a 20 to 60 per cent incidence of pneumonia, while up to 30 per cent of patients with a urinary catheter develop urinary tract infections. Up to 15 per cent of infections are related to the insertion of intravascular catheters, such as central venous pressure, Swan–Ganz, haemofiltration catheters, and arterial lines. The incidence of wound-related infection depends upon the type of surgery performed but accounts for about 15 per cent of critical care infections.

The longer a patient remains in the critical care unit the higher the incidence of infection: after 24 h about 10 per cent of patients become infected, while more than 90 per cent can be infected after 2 weeks.

Infection control

The keys to infection control include the adoption of preventative measures, the early recognition of infection, and the application of disciplined antibiotic policies. The Centers for Disease Control Guidelines (1985) address the important issues of hand washing, disinfection, and isolation procedures, together with strict control of antibiotic therapy and prophylaxis. These measures are aimed at the elimination of environmental organisms and prevention of cross-infection. Their impact has not been as great as might have been expected, consistent with the recognition that the patient's own endogenous bacterial flora serves as the source of most nosocomial infection.

Identification of organisms

Rigorous surveillance techniques must be applied to identify the organisms responsible for the infection. Specimens must be carefully preserved and transported rapidly to the laboratory: loss of a microbiological specimen through carelessness may adversely affect morbidity or even mortality. Table 1 lists the bacteria most commonly identified in nosocomial infections. Although there is considerable variation from unit to unit the common pathogens generally include *Escherichia coli, Staphylococcus aureus, Pseudomonas aeruginosa*, and enterococci.

Organism	%
Coliforms	20
Staphylococcus aureus	10
Pseudomonas aeruginosa	10
Enterococcus spp.	10
Klebsiella spp., Enterobacter spp., Streptococcus epidermidis, Proteus spp., Candida spp., and others	50

Table 1 Bacterial flora commonly associated with nosocomial infection

The usual sites sampled in the microbiological survey of a potentially infected or septic patient are shown in Table 2. Clearly, surveillance should be directed primarily at the site suggested by the clinical picture. However, much time may be saved by sampling all potential sites of infection simultaneously.

Wounds and discharge sites
Lung and airways
Tracheal aspirate
Blind bronchial lavage
Bronchoscopic lavage
Urine
Blood
Two sites
Fresh venepuncture
Skin preparation with alcohol
Invasive line tips
Cerebrospinal fluid
Others, needle aspirates

Table 2 Antimicrobial surveillance in the patient with sepsis

Nosocomial pneumonia

Nosocomial pneumonia occurs in 0.5 to 1 per cent of patients who stay in hospital. A high proportion of these cases is seen in critical care units where the incidence of nosocomial pneumonia may reach 5 to 25 per cent. Such pneumonias may be responsible for up to a quarter of critical care deaths. Several factors predispose towards nosocomial pneumonia in patients in critical care units: some of these are summarized in Table 3.

Advanced age
Obesity
Cigarette smoking
Low Glasgow coma score
Diabetes mellitus
Malignant disease
Oropharyngeal colonization with Gram-negative organisms
Gastric colonization increased in ileus and reduced gastric acidity
Chronic heart, lung, or renal disease.
Endotracheal intubation

Table 3 Factors predisposing towards nosocomial pneumonia in patients in critical care units

Most reports of nosocomial pneumonia in the patient receiving critical care have used purely clinical criteria to establish the diagnosis. Conventional criteria include fever, leucocytosis, purulent sputum, and the appearance of new or progressive infiltrates on chest radiographs. While these criteria are probably satisfactory for patients in the general hospital they are not specific enough for use in the critically ill. The onset of fever and leucocytosis may be the result of a variety of infectious or even non-infectious pathologies. Since colonization of endotracheal tubes with oropharyngeal bacteria is inevitable within a few hours of intubation, simple endotracheal aspirates may not reliably reflect lung colonization.

A more specific approach to the diagnosis of pneumonia in patients receiving critical care is the adoption of microbiological surveillance based on quantitative cultures of material obtained by a protected specimen brush and bronchoalveolar lavage. A clinically significant infection is suggested by the presence of more than 10^3 colony-forming units (**cfu**) in 1 ml of lavage fluid, or in the 1 ml of saline in which the protected specimen brush has been agitated. Prospective studies of nosocomial pneumonia diagnosed by protected specimen brushing show the incidence increases from less than 10 per cent at 10 days to more than 25 per cent at 30 days. Patients in the critical care unit with pneumonia have a mortality rate about twice that of patients without pneumonia. Bronchoalveolar lavage offers the potential advantage that Gram staining of a centrifuge deposit of the fluid obtained may provide an early indication of pneumonia: if more than 25 per cent of white cells in the centrifuge deposit contain intracellular organisms it is highly likely that the patient has pneumonia. Recent studies have shown that blind bronchial lavage with an undirected catheter is almost as sensitive and specific as fibreoptically guided bronchoalveolar lavage.

Measures to reduce the risk of nosocomial pneumonia

Bacteria enter the lungs by several portals: from contamination of the inspired air by infected equipment, by aspiration of oropharyngeal secretions, or as blood-borne infection. Significant improvement has been achieved in the avoidance of infection from contaminated ventilator circuits and humidification systems. Changing ventilator circuits every day is associated with a higher incidence of nosocomial lung infection than when the circuits are changed every 2 to 3 days; leaving the ventilator circuits unchanged for the entire duration of the patient's admission does not appear to add to the risk of pneumonia. If a humidifier is used in the ventilator circuit, a closed system, which maintains the level of sterile water within the humidifier, is preferable to intermittent topping up. Sheathed suction catheter systems allow a single catheter to be used for 24 h without repeated disconnection of the ventilator circuit from the endotracheal tube. Suctioning can be continued without interruption of either intermittent positive pressure ventilation or continuous positive airways pressure. Although lung infection rates are no lower than those associated with conventional suction catheters, there are benefits from reducing the risk of contamination of nursing staff and cross-infection. Humidification can also be achieved using heat/moisture exchangers that also serve as microbiological filters and may reduce the rate of airway contamination.

Stress ulcer prophylaxis

Antacids and H_2-blockers increase gastric pH and may encourage bacterial colonization of the gastric secretions. Aspiration of gastric fluid is likely to be associated with a higher risk of nosocomial lung infection. The possible adverse effects of H_2-blockers and their questionable efficacy make routine prescription of these agents difficult to justify. Effective stress ulcer prophylaxis can be achieved by instilling sucralfate, a cytoprotective agent that does not increase gastric pH, into the stomach. It has been suggested that introducing small volumes of enteral feed (30 ml every 2 h), even in the presence of an ileus, achieves the same results. Patients receiving

enteral nutrition do not need stress ulcer prophylaxis.

Selective digestive tract decontamination

In view of the potential role of oropharyngeal and gastric secretions in the development of nosocomial pneumonia it is logical to reduce the rate of colonization of these secretions. Combinations of non-absorbable topical antibiotics such as polymyxin, tobramycin, and amphotericin with a short course of a systemic broad-spectrum antibiotic significantly reduce lung infection rates. Surprisingly, selective digestive tract decontamination has little impact upon mortality, except possibly in patients with trauma. Development of microbial resistance does not as yet appear to be a complication of this procedure although anecdotal experience suggests it might encourage the spread of methicillin-resistant staphylococci. The translocation of micro-organisms and endotoxin from the bowel lumen directly into the circulation has been linked with the development and persistence of the sepsis syndrome; this has been cited as a reason for inhibiting or eliminating intestinal colonization. Although routine selective digestive tract decontamination cannot yet be recommended, its use in selected patients might still be considered.

Other measures

Nasojejunal feeding tubes have advantages over the more widely used nasogastric tubes for enteral nutrition since they facilitate such feeding even in the presence of gastric paresis and are associated with a lower risk of aspiration of gastric contents.

The heavily sedated or obtunded patient should be moved regularly from side to side to discourage the development of dependent lung collapse. At other times the patient should be sat upright to reduce the risk of aspiration.

Management

All ventilated patients receiving critical care who are considered to have nosocomial pneumonia on clinical grounds should undergo fibreoptic bronchoscopy for bronchoalveolar lavage or protected specimen brushing. As an alternative, lavage can be performed through a non-directed catheter passed as far into a lobar bronchus as possible. If possible, cytopathological examination of the centrifuge sediment from bronchoalveolar lavage fluid should be performed. If more than 25 per cent of white cells contain organisms, empirical therapy should be commenced, based on the results of Gram staining. The range of bacteria isolated by protected specimen brushing in two studies of nosocomial pneumonia is shown in Table 4.

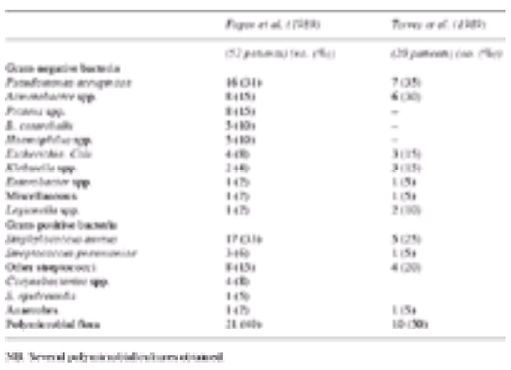


Table 4 Lung bacteria identified by protected specimen brush in patients with nosocomial pneumonia

Any empirical antibiotic regimen for patients with nosocomial pneumonia must provide broad-spectrum cover and minimize the risk of allowing multiresistant organisms to develop. A combination of an aminoglycoside and an antipseudomonal b-lactam largely satisfies these requirements, and has the additional benefit of antimicrobial synergy. These attributes are particularly desirable for treating *P. aeruginosa* infections, which carry a high mortality rate. If a staphylococcal infection is indicated on Gram stain, then antistaphylococcal antibiotics should be included in the regimen. Therapy should be modified or discontinued if less than 10^3 cfu/ml are isolated from quantitative cultures obtained by protected specimen brushing. The duration of antibiotic treatment necessary to eradicate the infecting organism has not been determined: depending upon the clinical response, antibiotics can usually be stopped after 5 to 7 days. Anaerobic or staphylococcal lung infections may require up to 4 weeks therapy.

Line-related infection

Up to 15 per cent of infections are line related and strict rules regarding the insertion and management of arterial and venous catheters need to be rigorously enforced. Recent randomized, controlled studies would suggest that no vascular lines, other than tunnelled feeding catheters and those required for haemodialysis or haemofiltration, should remain in place for more than 8 days. Factors associated with an increased incidence of vascular catheter infections include the infusion of hypertonic solutions, frequent disconnection of infusion lines, and the length of time the catheter remains *in situ.* Peripheral intravenous lines usually need to be changed more frequently because of local non-bacterial inflammation

Urosepsis

The presence of a transurethral bladder catheter in the critically ill patient is the major predisposing factor for urosepsis. Insertion of a urinary catheter must therefore be a clinical necessity and not a convenience. Stringent aseptic techniques must be observed during catheter insertion and during any subsequent disconnection of ‘closed' urinary catheter bag systems. Since many patients receiving intensive care are sedated, the presence of a urinary tract infection may not become apparent unless regular microbiological surveillance is undertaken. Bacteraemia may occur in up to 5 per cent of patients with bacteruria. The commonest infecting organisms are Gram-negative bacteria including *E. coli*, enterococci spp., and *P. aeruginosa*. Bacteruria needs to be treated with systemic antibiotics if there is any suspicion of pyelonephritis, bacteraemia, or sepsis. Ideally the urinary catheter should be removed, although this may not be possible in the sedated or unconscious patient. In the patient with renal failure and severe oliguria or anuria the urinary catheter can be removed awaiting recovery of renal function, which is usuallyheralded by spontaneous micturition.

Cellulitis

Superficial skin infection extending down to the upper subcutaneous tissue can occur with a range of both Gram-positive and Gram-negative organisms, but is most commonly due to *Strep. pyogenes*. Such infections may develop during a stay in intensive care as a complication of invasive intravascular lines, traumatic wounds, or surgical incisions. Unless a specific organism can be identified, empirical therapy with a broad-spectrum antibiotic will be required. Progression or resolution can be assessed by marking the edge of the cellulitic area with an indelible marker and making daily comparisons.

Necrotizing fasciitis

Necrotizing fasciitis is an extremely serious, fulminant infection of deeper skin layers of the abdominal wall, perineum, or extremities. A wide range of causative organisms have been reported including group A streptococci, *Staph. aureus*, and enteric bacteria. It is more likely to occur in patients with diabetes, cirrhosis, or in those who are immunocompromised. Early clinical appearances are those of cellulitis. Blistering may develop and crepitus may suggest a gas-forming organism. The infected area may extend despite appropriate antibiotic therapy and may require extensive debridement to arrest progression of the disease. Sepsis syndrome is a common accompaniment and is associated with a high mortality.

Infection in the immunocompromised host

The most common causes of immunocompromise are malignancy, chemotherapy, corticosteroid treatment, immunosuppressive treatment after transplantation, and acquired immune deficiency syndrome (**AIDS**). Immunocompromise can take the form of failure of phagocytosis, defective cell-mediated immunity, defective antibody-mediated immunity, and splenectomy. Opportunistic infections can usually be managed outside the critical care unit unless respiratory or renal failure intervenes. Some degree of deficiency of both cellular and antibody-mediated immune response is observed in patients with trauma or burns, and this may increase

their susceptibility to infection.

Failure of phagocytosis

A reduction in the number or function of neutrophils, monocytes, or macrophages increases susceptibility to infection. Pyogenic abscesses or osteomyelitis caused by *Staph. aureus*, Enterobacteriaceae, *Serratia marcesens*, and fungi are common. Failure of macrophage function may result from their impaired migration to the infection site (as in diabetes mellitus) or from a failure of opsonization, which requires the interaction of antibody and complement on the surface of the micro-organism. Deficiency of systemic antibodies or complement, whether congenital or acquired, will therefore result in impaired phagocytosis.

Complement deficiency

Congenital deficiencies of complement components 6, 7, or 8 are associated with recurrent systemic infections by *Neisseria meningitidis* and *N. gonorrhoea*. Deficiency of the early components C2 or C3, although rare, may predispose to recurrent viral and bacterial infections. Acquired complement deficiency, which may occur in immune complex diseases such as systemic lupus erythematosus and some subtypes of glomerulonephritis, will encourage the progression of a localized infection to systemic sepsis. In an acute exacerbation of systemic lupus erythematosus, low complement levels are often used as a marker of disease activity. Thus, low complement levels might lead to the conclusion that conditions such as lung infiltrates or encephalopathy are due to the primary disease process and not to infection. Clearly this may not be appropriate.

Granulocytopenia

Granulocytopenia is one of the most common causes of immunocompromise in patients receiving critical care. Patients with fewer than 1000 granulocytes/mm^3, including immature precursors, are considered to be granulocytopenic and are at increased risk of infection. Severely granulocytopenic patients with white cell counts of less than 500/mm^3 are at particular risk. Leucocyte transfusions are not effective in correcting this immune defect, but there may be a place for the use of recombinant granulocyte/monocyte colony-stimulating factor.

Cellular immunodeficiency

Congenital forms of cell-mediated immunodeficiency such as Di George syndrome or Wiskott–Aldrich syndrome are rare, and most cases are due to an acquired defect in the number or function of T lymphocytes. Impaired cell-mediated immunity can be demonstrated by an absence of delayed hypersensitivity to candida, mumps, or tuberculin (PPD) antigens. The T-lymphocyte count (CD3+ and CD4+) may also be reduced.

The causes of acquired cell-mediated immunodeficiency include lymphoproliferative diseases such as Hodgkin's lymphoma, steroid and immunosuppressive therapy, sarcoidosis, and AIDS. Some impairment of cell-mediated immunity is also found in patients undergoing chronic haemodialysis.

Cell-mediated immunity stems from the interaction of effector T cells with antigen-presenting monocytes and macrophages. The presentation of surface antigen (from the pathogen) by presenting cells, to the effector lymphocytes, is an essential prerequisite to the cell-mediated immune mechanism. As even pathogens that reside intracellularly are accessible to the presenting phagocytes, and therefore to the effector lymphocytes, cell-mediated immunity is particularly effective against a wide range of intracellular organisms, some of which are listed in Table 5. Susceptibility to infection with non-intracellular organisms, such as fungi, is also found in cell-mediated immunodeficiency.

Bacteria	
	Salmonella spp.
	Listeria monocytogenes
	Nocardia spp.
	Legionella pneumophila
Mycobacteria	
	M. tuberculosis
	M. avium-intracellulare
Fungi	
	Candida spp.
	Aspergillus spp.
	Mucor spp.
	Cryptococcus spp.
Protozoa	
	Pneumocystis carinii
	Toxoplasma gondii
	Cryptosporidia
Viruses	
	Cytomegalovirus
	Herpes zoster
	Herpes simplex
	Epstein–Barr
Parasites	
	Strongyloides stercoralis

Table 5 Infections associated with cell-mediated immunodeficiency

Defective humoral immunity

Defective humoral or antibody-mediated immunity results from deficient B-lymphocyte maturation or differentiation. The subsequent failure of plasma cells to produce immunoglobulin results in antibody deficiency. Defective antibody-mediated immunity can be demonstrated in the laboratory by assay of IgM, IgG, and IgA and their subclasses. Blood B-cell count may be low and the *in vivo* antibody response to pneumococcus polysaccharide or tetanus toxoid may be impaired. Antibody deficiency occurs in chronic lymphocytic leukaemia, multiple myeloma, and in congenital or acquired hypogammaglobulinaemia. Occasionally a thymoma may underlie an acquired hypogammaglobulinaemia. The usual pathogens are *Haemophilus influenzae*, pneumococcus, *P. aeruginosa*, and *N. meningitidis*.

Treatment of defective antibody-mediated immunity relies upon appropriate use of antibiotics supplemented by immunoglobulin infusion in doses of 200 to 500 mg/kg. Immunoglobulin administration may need to be repeated each month.

Splenectomy

Splenectomy may result from trauma, staging of Hodgkin's disease, or as part of the treatment of hereditary spherocytosis or idiopathic thrombocytopenic purpura. Patients with sickle cell disease may be functionally asplenic.

Pathogens affecting these patients are the encapsulated bacteria such as pneumococcus, *H. influenzae*, and meningococcus. Splenectomized patients, especially children, may develop a fulminant septicaemia that is rapidly fatal unless immediately treated (postsplenectomy syndrome). Prophylactic pneumococcal and *H. influenzae* vaccination is indicated before surgery is undertaken in these patients.

Sites of infection and organisms

The granulocytopenic patient usually experiences infection of the lungs, skin, alimentary tract, oropharynx, and oesophagus. The urinary tract and liver are less commonly involved. Common pathogens include *P. aeruginosa, Klebsiella pneumoniae, E. coli, Staph. aureus, Candida albicans*, and *Aspergillus* spp. Patients with cell-mediated immunodeficiency experience infections in similar sites but with a slightly different spectrum (Table 5). Nocardial infection may be especially common in recipients of cardiac transplants: the lung is the most common site of infection, and bronchoalveolar lavage or transbronchial biopsy is usually required for diagnosis. Treatment is with sulphisoxazole (6 to 12 g/day) or trimethoprim/sulphamethoxazole.

Prevention of infection

Since many infections are caused by the patient's own endogenous bacteria, attempts have been made to reduce infection using isolation procedures and special environmental units: results have been disappointing. Selective decontamination of the oropharynx and upper gastrointestinal tract using non-absorbable antibiotics has failed to increase long-term survival. In contrast, antifungal prophylaxis using ketoconazole has been effective in reducing the incidence of oesophageal candidiasis. As with the immunocompetent patient, the risk of infection is related to the level and number of interventions: invasive procedures should therefore be limited to essential diagnostic investigations. Stool softeners should be administered to prevent bacterial translocation and anorectal abrasion.

Microbiological surveillance

Careful physical examination of the patient is an essential prerequisite to locating and identifying an infection. The oropharynx, ears, and sinuses should be checked daily. Fundoscopic eye examination may reveal fungal infection. Culture samples should be obtained from the sites listed in Table 2, with the addition of oropharyngeal and stool specimens.

The appearance of new infiltrates on chest radiographs indicates an opportunistic pneumonia until proved otherwise. Although sputum should be stained (Gram and special stains) and cultured, fibreoptic bronchoscopy for bronchoalveolar lavage, brushing, and transbronchial biopsy may be required. Occasionally, open lung biopsy is needed to establish a diagnosis.

Antibiotic treatment

The use of antimicrobial prophylaxis suffers from the major disadvantage that bacterial resistance almost always develops. Confirmed or even suspected infection requires immediate and effective treatment: the source or type of infection may not be identified despite careful physical examination and microbiological surveillance and empirical antibiotic therapy may have to be started based on clinical probability. The selected antibiotics need to be effective against Gram-negative and Gram-positive bacteria such as *E. coli, Pseudomonas* spp., *Staph. aureus*, and *Staph. epidermidis*. A combination of a third-generation cephalosporin or penicillin derivative such as piperacillin with an aminoglycoside (gentamicin, netilmicin) and a penicillinase-resistant penicillin such as nafcillin or oxacillin would eliminate most of the likely organisms.

Aminoglycoside resistance may occur, particularly to gentamicin, and may explain a lack of response to treatment. Antibiotic resistance follows regional trends, so that some parts of the world experience antibiotic resistance that is not found elsewhere. The length of antibiotic treatment is an inexact science. In principle, antibiotics should generally be continued until cultures become negative and signs of infection resolve. Prolonged courses of antibiotics should be avoided except in exceptional circumstances such as deep-seated indolent inflammatory masses or bone infections.

Fungal infection

Persistent fever in the face of apparently adequate antibiotic therapy should always raise the question of fungal infection. Systemic fungal infections are difficult to diagnose and carry a high mortality rate. Simultaneous colonization of the skin and other sites such as the oropharynx, oesophagus, and urinary tract makes systemic infection more likely but not inevitable. Blood cultures for fungi can be negative in up to 50 per cent of patients with systemic fungaemia and, unfortunately, serological tests cannot differentiate clearly between colonization and systemic infection. The addition of amphotericin B to the treatment regimen should be seriously considered in cases of suspected candidiasis, despite the risks of nephrotoxicity. A less toxic antifungal agent such as fluconazole might be considered as an alternative, particularly for the treatment of candida infections. Other fungi such as *Aspergillus* and *Mucor* spp. should be considered, especially in patients presenting with a lung infection. Bronchoscopy, bronchoalveolar lavage, and transbronchial biopsy are usually needed to establish the diagnosis of invasive aspergillosis or mucormycosis: treatment is with amphotericin B. Cryptococcal infections usually present with meningitis, pneumonitis, or disseminated infection. Examination of cerebrospinal fluid (Gram stain or India ink preparation with latex agglutination tests for cryptococcal antigen) may confirm the diagnosis. Any skin lesions should be scraped and Gram stained. Transbronchial biopsy is required in the case of suspected cryptococcal pneumonitis. Cryptococcus infection should be treated with amphotericin B and flucytosine.

Pneumocystis carinii

This protozoan is a common cause of pneumonia in the immunosuppressed patient. Patients present with breathlessness, tachypnoea, cough, and hypoxia, and physical examination reveals scattered crackles on auscultation, although chest findings can be minimal. The chest radiographic appearances are non-specific and may take the form of diffuse bilateral interstitial or lobar infiltrates. The infection is generally fatal if not treated. The treatment of choice for *P. carinii* is trimethoprim (20 mg/kg) and sulphamethoxazole (100 mg/kg), given four times a day for up to 14 days. Failure of clinical response within 3 to 6 days or the development of side-effects such as thrombocytopenia may require the therapy to be changed to pentamidine isothionate, 4 mg/kg per day. Side-effects such as neutropenia, uraemia, and cardiotoxicity may be avoided by administering pentamidine as an aerosol. Corticosteroids can be added to reduce the interstitial inflammatory response and have improved survival in patients with AIDS who have developed *Pneumocystis* pneumonia.

Toxoplasma gondii

Toxoplasma gondii disseminates widely through multiple organ systems, with death usually resulting from encephalitis or myocarditis. Serological tests for *Toxoplasma* are not reliable, the diagnosis being best established by identifying the organism in cerebrospinal fluid. Treatment is with a combination of sulphadiazine and pyrimethamine.

Viral infections

Of the viral infections that must be considered, herpes simplex, herpes zoster, and cytomegalovirus are commonly encountered in transplant recipients. If infection with these agents is confirmed, immunosuppressive therapy generally needs to be discontinued and specific antiviral agents administered. Gancyclovir is effective against cytomegalovirus, and acyclovir is effective against herpes simplex and herpes zoster. Both agents are associated with bone marrow suppression and nephrotoxicity.

Antibiotic therapy in intensive care

The major issues that must be addressed when prescribing antibiotics are listed below.

1. Does the nature and severity of infection justify the use of antibiotics?
2. Are antibiotics being used as prophylaxis, specific therapy for an identified organism, or as empirical therapy?
3. What are the appropriate dosage regimens and is combination therapy required?
4. What are the criteria for discontinuing treatment? What is considered to represent a ‘course' of treatment?

Of these, the first and last questions are the most important and the most difficult to answer: whether to start treatment with antibiotics, and when to stop. The selection of an antibiotic for use in critically ill patients should be based on a sound appreciation of the clinical presentation and the microbiological laboratory findings. The patient with fever, leucocytosis, and a site of infection does not necessarily require treatment with antibiotics: inappropriate or unnecessary antibiotic therapy being associated with increased mortality. Conversely, it is not unusual to begin empirical therapy in a patient who is critically ill before an organism or a site of infection can be identified. If possible, empirical therapy is selected according to the nature or suspected site of infection. For example, an aspiration pneumonia will require a different antimicrobial agent or combination of agents than would be selected for intra-abdominal sepsis. After commencing antibiotics, the clinical response to therapy and results of bacterial stains and cultures will determine whether treatment should continue, cease, or be modified. The ideal length of antibiotic treatment has not been determined, but 5 days would generally be adequate for most infections. Beyond this time the possibility of development of resistant microbial strains increases.

b-Lactams

The b-lactams, which include the penicillins, cephalosporins, carbapenems, and monobactamsTable 5, have found wide application in critical care. These antibiotics bind to specific penicillin-binding receptor proteins on the cytoplasmic surface of the bacterial cell wall, the inner membrane, releasing autolysins that disrupt the cell wall structure as the organism replicates.

b-Lactam resistance has evolved by alteration of the penicillin-binding protein receptors, by alteration of the pores that allow egress of the antibiotic through the cell wall, or by production of b-lactamase, a bacterial enzyme which causes hydrolytic destruction of the b-lactam ring with subsequent loss of antibacterial activity. b-Lactamase may be produced by the organism either continuously or after exposure to a b-lactam agent. The development of organisms resistant to multiple antimicrobial agents is a significant problem in the intensive care unit.

Penicillins

Pneumococcal or meningococcal meningitis, streptococcal endocarditis, and anaerobic lung abscesses (but not those caused by *Bacillus fragilis*) may be treated with penicillin G. To maintain minimal inhibitory serum concentrations, 1- or 2-hourly dosing may be necessary in meningitis. The penicillins often cause rashes, anaphylaxis, haemolytic anaemia, potassium-losing nephropathy, and pseudomembranous colitis. In patients with renal failure, high serum concentrations of penicillins

may induce seizures.

Broad-spectrum penicillins

Ampicillin and amoxicillin have an extended range of activity to include Gram-negative bacteria such as *E. coli, H. influenzae, Proteus mirabilis, Salmonella* spp., and *Shigella* spp. Combination of these penicillins with an aminoglycoside for serious enterococcal infections reduces the emergence of resistance.

Penicillinase-resistant penicillins

Resistance to nafcillin and oxacillin seriously restricts the use of these agents in the treatment of staphylococcal infections in the critical care unit. These antistaphylococcal penicillins are indicated primarily when organisms can be reasonably predicted to be sensitive: when resistance is a possibility, an alternative agent such as vancomycin should be prescribed. The addition of an aminoglycoside enhances the efficacy of the penicillins, particularly in serious infections such as endocarditis or osteomyelitis.

a-Carboxy penicillins

Carbenicillin and ticarcillin, are antipseudomonal penicillins effective not only against *P. aeruginosa*, but also against most Enterobacteriaceae, penicillin-sensitive Gram-positive organisms, and several anaerobes. These are generally unsuitable for empirical therapy because of their unpredictable activity against *Klebsiella pneumoniae* and enterococci. Like methicillin, the antipseudomonal penicillins may induce a potassium-losing nephropathy if used in high doses.

Ureidopenicillins

Piperacillin, mezlocillin, and azlocillin are extended-spectrum penicillins. Azlocillin is effective against *P. aeruginosa*, but less predictable in its activity against other Gram-negative bacilli. Mezlocillin and piperacillin are effective against all the organisms sensitive to carbenicillin and ticarcillin, but also very active against *Klebsiella*, enterococci, *Serratia*, and *Citrobacter* spp.

The half-life of ureidopenicillins is about 1.3 h and the recommended dose for critically ill patients is 3 to 4 g every 4 h. Resistance to the ureidopenicillins frequently emerges when they are used as empirical monotherapy and ideally they should be combined with an aminoglycoside such as gentamicin or netilmicin.

Cephalosporins

Due to their broad spectrum of activity, the cephalosporins are among the most frequently prescribed parenteral antibiotics for the critically ill. Because of their relative safety, these agents have been used extensively for surgical prophylaxis as well as treatment regimens.

Cephalosporins have been classified in terms of generations to denote similarities in chemical structure and chronology of release. However, it is probably better to consider the cephalosporins in terms of their potency and spectrum of activity. For example, first-generation cephalosporins are less active against aerobic Gram-negative bacilli than second-generation cephalosporins, while the third-generation agents are effective against a wide spectrum of organisms including Gram-negative bacilli. In addition, first-generation cephalosporins are generally more effective against Gram-positive cocci than are third-generation agents. Thus, first-generation cephalosporins are usually prescribed for staphylococcal or non-enterococcal streptococcal infections in patients with penicillin allergy without anaphylaxis.

All of the cephalosporins are relatively ineffective against enterococci or methicillin-resistant staphylococcal infections. Cross-sensitivity with penicillin allergy may be observed, taking the form of rash, anaphylaxis, neutropenia, drug fever, and pseudomembranous colitis. Occasionally there may be elevation of hepatic enzyme levels and haematological disturbances such as thrombocytopenia, leucopenia, or anaemia.

Cefotaxime, ceftriaxone, and ceftizoxime show similar spectra of activity in ill patients. However, they are relatively ineffective against *P. aeruginosa*, methicillin-resistant staphylococci, and enterococci. Due to its ability to penetrate the blood–brain barrier, cefotaxime has been widely used in the treatment of Gram-negative bacillary meningitis. Ceftazidime has the benefit of being effective against *P. aeruginosa* infections.

Carbapenems

Imipenem is a broad-spectrum carbapenem combined with the dehydropeptidase inhibitor, cilastatin, which inhibits its enzymatic breakdown. Imipenem's activity against aerobic Gram-negative bacilli, including *P. aeruginosa*, is comparable with that of aminoglycosides. Its activity against *Staph. aureus* and streptococci is similar to that of nafcillin. Meropenem has a similar spectrum of activity but does not require an enzyme inhibitor to ensure efficacy.

Resistant strains of *P. aeruginosa* have developed following imipenem therapy, and *Pseudomonas cepacia, Pseudomonas maltophilia*, and *Enterococcus faecium* are also resistant to imipenem.

The broad spectrum of imipenem and meropenem makes these agents highly suitable for empirical therapy for the critically ill, especially when used in combination with an aminoglycoside and metronidazole. Seizures have been reported after the administration of large doses of imipenem or in patients with renal failure.

Monobactams

Aztreonam, is a monobactam antimicrobial, with a somewhat narrower spectrum of activity than other b-lactams. It has an antimicrobial spectrum similar to the aminoglycosides: although effective against *P. aeruginosa*, combination with an aminoglycoside for pseudomonal infections may inhibit the development of resistant strains.

b-Lactamase inhibitors

Several derivatives of the natural penicillins have broad spectra of activity. By combining these penicillins with b-lactamase inhibitors the activity spectrum can be broadened further to create a major class of antibiotics. The addition of the b-lactamase inhibitor clavulanic acid greatly enhances the activity of ticarcillin or amoxicillin, and broadens its spectrum of activity against anaerobes, *Staph. aureus*, and aerobic Gram-negative organisms.

Aminoglycosides

Aminoglycosides, alone or combined with other antibiotics, remain an essential part of therapy for life-threatening Gram-negative infections. Variations in drug distribution mandate close monitoring of serum concentrations to achieve the therapeutic effect without toxicity.

The parenteral aminoglycosides most commonly used in the treatment of serious systemic infections include gentamicin, netilmicin, amikacin, and tobramycin. The *in vitro* antimicrobial activities of these aminoglycosides are similar, in terms of their activity, against Gram-negative organisms such as *P. aeruginosa*.

The combination of an aminoglycoside with a b-lactam or with a monobactam antibiotic has been shown to enhance the *in vitro* susceptibility of Gram-negative bacilli, especially *P. aeruginosa*. Enterococci are generally resistant to penicillin, ampicillin, or vancomycin alone, but these drugs do exhibit synergy with the aminoglycosides. Gentamicin exhibits more synergy than streptomycin, tobramycin, or amikacin.

The plasma half-life and distribution are similar for most aminoglycosides. The pharmacokinetics change in the critically ill patient with sepsis, heart failure, renal failure, or hepatic disease, and similar changes occur following surgery and during pregnancy. These factors tend to increase the volume of distribution so that loading and maintenance doses of the aminoglycosides have to be increased. Aminoglycosides are distributed throughout the extracellular fluid compartment, but because penetration of the blood–brain barrier is minimal, levels within the cerebrospinal fluid and brain tissue are poor. The aminoglycosides are excreted largely unchanged by glomerular filtration.

Gentamicin and netilmicin require a loading dose of between 5 mg/kg and 7 mg/kg (the higher dose for *Pseudomonas* infections) followed by single daily doses according to renal function and aminoglycoside blood levels. Plasma levels should be measured 18 to 20 h after each daily dose allowing time to receive results of the assay and determine the next dose. Once-a-day therapy is believed to reduce the risk of nephrotoxicity although this is not borne out by the available evidence.

However, the convenience of daily dosing and the need for fewer blood level estimations is preferrable to multiple dosing regimens. Although monitoring of aminoglycoside levels is laborious and expensive, combination therapy represents the best way of achieving an antimicrobial effect without encouraging the development of resistant strains.

Side-effects associated with aminoglycoside use include nephrotoxicity, ototoxicity, and accentuation of neuromuscular blockade. Predicting nephrotoxicity is difficult but the presence of factors such as volume depletion and previous aminoglycoside treatment probably increase the risk. Factors associated with nephrotoxicity are blood levels greater than 2 µg/ml, the cumulative dose, and treatment courses longer than 2 to 3 weeks. Some degree of renal impairment may be observed in up to 25 per cent of patients receiving aminoglycoside therapy. However, the true nephrotoxic risk is difficult to assess since aminoglycosides tend to be used in patients already at risk of renal failure from their underlying illnesses.

Quinolones

The fluoroquinolones represent a unique group of antibiotics. Agents such as ciprofloxacin have a much broader spectrum of activity than the earlier quinolones, with significant activity against a wide range of Gram-negative bacilli including *P. aeruginosa*. Approximately 50 per cent of methicillin-resistant strains of *Staph. aureus* are sensitive to ciprofloxacin, although it has limited activity against *B. fragilis* and aerobic streptococci.

Tissue penetration is excellent with both oral and systemic ciprofloxacin, and fluid and tissue concentrations may exceed serum levels, except in bronchial secretions, cerebrospinal fluid, and saliva. A potential drawback of the fluoroquinolones may be the ease with which resistant strains develop. Like the b-lactams, combination therapy with aminoglycosides may be necessary in the critically ill patient.

Anti-anaerobic antibiotics

A wide range of antibiotics, including clindamycin, metronidazole, chloramphenicol, and the extended-spectrum penicillins cefoxitin, imipenem, and piperacillin, are effective in treating anaerobic infections. The third-generation cephalosporins such as cefotetan, moxalactam, and ceftizoxime and more effective against *Bacteroides* spp. than the earlier generations of cephalosporins.

In view of the relative safety of metronidazole the inclusion of this agent in combination therapy is usually justified, especially if lower intestinal pathogens are a potential threat.

Clindamycin is the agent that is most often associated with the development of pseudomembranous colitis although several other antibiotics, including the cephalosporins, may also be implicated. Metronidazole therapy has been associated with neutropenia, polyneuropathy, and convulsions.

Other agents

Several antibiotics that have specific applications can be identified. These include erythromycin, tetracycline, vancomycin, trimethoprim/sulphamethoxazole, and the antituberculous agents. In addition, specific antiviral and antifungal drugs are being used in critically ill patients with increasing frequency.

Erythromycin is the drug of choice for the treatment of *Legionella* infection, which should be considered in patients with pneumonia, renal impairment, gastrointestinal symptoms, or neurological abnormalities. Erythromycin therapy should be started intravenously at 1 g, 6-hourly. Erythromycin offers significant activity against other community-acquired infections, such as pneumococcal and mycoplasmal pneumonia.

Tetracycline remains the drug of choice for patients with Rocky Mountain spotted fever, brucellosis, or non-cholera *Vibrio* infections. The maximum adult daily dose of this agent is 2 g.

Vancomycin is usually effective against serious methicillin-resistant staphylococcal infections. It is also suitable for patients who have a history of anaphylactic reaction to penicillin. Vancomycin may be infused over at least 1 h at a dose of 1 g, repeated 8- to 12-hourly in patients with normal renal function. In critically ill patients with suspected staphylococcal infection, vancomycin can be started before specific identification and sensitivities are available. Flucloxacillin/methicillin can be introduced later when sensitivities are known. As with the aminoglycosides, peak and trough serum concentrations must be closely monitored for dose adjustment after 48 h: dose requirements are much reduced in the presence of renal failure.

Trimethoprim/sulphamethoxazole is most often used in the treatment of *P. carinii* pneumonia. Pentamidine given either parenterally or by nebulization is an alternative therapy, particularly in patients with AIDS. Trimethoprim/sulphamethoxazole has been recommended for use in the management of severe nocardial infections, which characteristically respond to treatment with sulphonamides.

Antibiotic resistance

Resistance to available antimicrobial agents is increasing and includes, in particular, resistance to the glycopeptides (vancomycin and teicoplanin) amongst enterococci, resistance to methicillin and fluoroquinolones amongst staphylococci, and resistance to penicillin and some other b-lactams amongst viridans and pneumococcal strains of streptococci. Until recently *Strep. pneumoniae* was almost universally sensitive to penicillin. However, up to 10 per cent of strains, in certain geographical areas, may now exhibit resistance such that erythromycin or high-dose b-lactam antibiotics may be required for treatment until sensitivities can be defined. With continued widespread use of antibiotics in the critically ill, it must be expected that more resistant strains will appear and provide greater challenges for therapy.

Strains of vancomycin-resistant enterococci (**VRE**) have emerged and spread widely throughout intensive care units during the last few years. Multiresistant strains of *Enterococcus faecium* are especially troublesome because they are often resistant to available antimicrobials. VRE infections occur in patients with severe underlying disease who have undergone invasive procedures and received prolonged courses of broad-spectrum antimicrobial therapy. Because therapeutic options are limited, prevention of spread from patient to patient is the key to control.

Methicillin-resistant *Staphylococcus aureus* (**MRSA**) is also an increasingly common nosocomial pathogen with approximately two-thirds of nosocomial cases and outbreaks occurring in intensive care units. Risk factors for MRSA postoperative wound infections and bacteraemia include severe underlying disease, prior antibiotic therapy, intravascular catheterization, and intensive care unit location. Outbreaks of MRSA can usually be controlled by screening admissions, surveillance, and minimal barrier interventions, while molecular and phage typing of MRSA can be used to support or refute the presence of cross-transmission. Frequently, transmission is from patient to patient with a member of the intensive care staff acting as a vector. Only by the application of meticulous barrier procedures can widespread dissemination within an intensive care unit be prevented. Vancomycin is effective for the treatment of established infection, but colonization does not usually require treatment. MRSA can be the pathogen in ventilator-associated pneumonia, when it is more likely to cause bacteraemia and septic shock than methicillin-sensitive infection.

Further reading

Anon. The choice of antimicrobial drugs. *Medical Letter on Drugs and Therapeutics* 1988; **30**: 33–40. [Good general review of antibiotics and their usage.]

Asensio A, Guerrero A, Quereda C, Lizan M, Martinez Ferrer M. Colonization and infection with methicillin-resistant *Staphylococcus aureus*: associated factors and eradication. *Infection Control and Hospital Epidemiology* 1996; **17**: 20–8. [Case controlled studies from which recommendations for control of MRSA are derived.]

Barza M. Imipenem. First of a new class of b-lactam antibiotics. *Annals of Internal Medicine* 1985; **103**: 552–60. [Review of the nature and role of imipenem.]

Bodnar UR, Noskin GA, Suriano T, Cooper I, Reisberg BE, Peterson LR. Use of in-house studies of molecular epidemiology and full species identification for controlling spread of vancomycin-resistant *Enterococcus faecalis* isolates. *Journal of Clinical Microbiology* 1996; **34**: 2129–32. [Out of 16 patients with VRE, 12 were found to be from a single clone. Implications for the movements of patients.]

Brown AE. Neutropenia, fever and infection. *American Journal of Medicine* 1984; **76**: 421–5. [Review of magnitude of problem assessment and treatment.]

Centers for Disease Control. CDC definitions for nosocomial infections. *American Review of Respiratory Diseases* 1988; **139**: 1058–9. [Standards for defining nosocomial infections.]

Chandrasekar PH, Kruse JA, Mathews MF. Nosocomial infection among patients in different types of intensive care units at a city hospital. *Critical Care Medicine* 1986; **14**: 508–10. [Highest rate of nosocomial infection in surgical ICUs; lowest in coronary care units.]

Chastre JY *et al.* Diagnosis of nosocomial bacterial pneumonia in intubated patients undergoing ventilation: comparison of the usefulness of bronchoalveolar lavage and the protected specimen brush. *American Journal of Medicine* 1988; **85**: 499–506. [Comparison of relative sensitivities and specificities of the two techniques.]

Chastre J *et al.* Quantification of BAL cells containing intracellular bacteria rapidly identifies ventilated patients with nosocomial pneumonia. *Chest* 1989; **95**: 190S–2S. [A rapid method for identifying lung infection from BAL fluid.]

Cormican MG, Jones RN. Emerging resistance to antimicrobial agents in Gram-positive bacteria. Enterococci, staphylococci and nonpneumococcal streptococci. *Drugs* 1996; **1**: 6–12. [A review of common organisms exhibiting antimicrobial resistance.]

Craven DE *et al.* Nosocomial infection and fatality in medical and surgical intensive care unit patients. *Archives of Internal Medicine* 1988; **148**: 1161–8. [Comparison of 526 MICU and 799 SICU patients. More infection in surgical cases but more deaths in medical cases.]

Daschner F. Nosocomial infections in the intensive care unit. *Intensive Care Medicine* 1985; **11**: 284–7. [Types and susceptibilities of ICU acquired nosocomial infection.]

Donowitz GR, Mandell GL. b-Lactam antibiotics. *New England Journal of Medicine* 1988; **318**: 419–26, 490–9. [Uses and limitation of b-lactam antibiotics.]

Dreyfuss D *et al.* Prospective study of nosocomial pneumonia and of patient and circuit colonization during mechanical ventilation with circuit changes every 48 h versus no change. *American Review of Respiratory Diseases* 1991; **143**: 738–43. [Not changing ventilator circuits had no effect on the rate of lung colonization by either Gram-positive or Gram-negative organisms.]

Duma RJ. Aztreonam, the first monobactam. *Annals of Internal Medicine* 1987; **106**: 766–7. [Review of spectrum and applications of aztreonam.]

Fagon JY *et al.* Detection of nosocomial lung infection in ventilated patients. *American Review of Respiratory Diseases* 1988; **138**: 1210–16. [Use of protected specimen brush and quantitative culture in diagnosis of nosocomial lung infection.]

Gerding DN *et al. Clostridium difficile*-associated diarrhea and colitis in adults: a prospective case controlled epidemiologic study. *Archives of Internal Medicine* 1986; **146**: 95–100. [Case controlled study of 149 patients with *Clostridium difficile* diarrhoea. Use of toxin detection and review of causes.]

Gold HS, Moellering RC, Jr. Antimicrobial-drug resistance. *New England Journal of Medicine* 1996; **335**: 1445–53. [Extensive review of patterns and mechanism of antimicrobial drug resistance.]

Hagman HM, Strausbaugh LJ. Vancomycin-resistant enterococci. The ‘superbug' scourge that's coming your way. *Postgraduate Medicine* 1996; **99**(5): 60–5, 69–71. [Review of epidemiology, causes, and management of VRE.]

Hartstein AI, Denny MA, Morthland VH, LeMonte AM, Pfaller MA. Control of methicillin-resistant *Staphylococcus aureus* in a hospital and an intensive care unit. *Infection Control and Hospital Epidemiology* 1995; **16**: 405–11. [Prospective surveillance study using molecular techniques to characterize patterns of infection spread.]

Johanson WG, Pierce AK, Sanford JP, Thomas GD. Nosocomial respiratory infections with Gram-negative bacilli. The significance of colonization of the respiratory tract. *Annals of Internal Medicine* 1972; **77**: 701–6. [Natural history of lung colonization and established lung infection.]

Johanson WG, Seidenfeld JJ, Gomez P, De Los Santos R, Coalson JJ. Bacteriologic diagnosis of nosocomial pneumonia following prolonged mechanical ventilation. *American Review of Respiratory Diseases* 1988; **137**: 259–64. [BAL and protected specimen brush cultures in chronically ventilated baboon model.]

Joshi JH, Schimpff SC. Infections in the compromised host. In: Mandell GL, Douglas RG, Jr, Bennett JE, eds, *Principles and practice of infectious diseases*, pp. 1644–9. John Wiley & Sons, New York, 1985. [A guide to the scope of opportunistic infection and its treatment.]

Leu HS, Kaiser DL, Mori M, Woolson RF, Wenzel RP. Hospital acquired pneumonia. Attributable mortality and morbidity. *American Journal of Epidemiology* 1989; **129**: 1258–67. [A study in 890 cases of hospital acquired pneumonia suggesting a 33 per cent contribution to crude mortality.]

Loudon I. Necrotising fasciitis, hospital gangrene and phagedena. *Lancet* 1994; **344**: 1416–19. [Causes, natural history, and treatment of necrotizing fasciitis.]

McA Ledingham I *et al.* Triple regimen of selective decontamination of the digestive tract, systemic cefotaxime, and microbiological surveillance for prevention of acquired infection in intensive care. *Lancet* 1988; **i**: 785–90. [Controlled study of 324 general ICU patients demonstrating reduction in incidence of lung infection and mortality in certain categories of patients.]

Montecalvo MA, Shay DK, Patel P *et al.* Bloodstream infections with vancomycin-resistant enterococci. *Archives of Internal Medicine* 1996; **156**: 1458–62. [Retrospective review of risk factors for and outcome from VRE bacteraemia.]

Neu HC. Infection problems for the 1990's—do we have an answer? *Scandinavian Journal of Infectious Diseases Supplement* 1993; **91**: 7–13. [Strategies for avoiding MRSA, VRE, and other resistant micro-organism infections.]

Neu HC. The emergence of bacterial resistance and its influence on empiric therapy. *Review of Infectious Diseases* 1983; **5**: 59–520. [Review of the effects of antibiotics on micro-organism selection and the development of resistance patterns.]

Rello J *et al.* Ventilator-associated pneumonia by *Staphylococcus aureus*. Comparison of methicillin-resistant and methicillin-sensitive episodes. *American Journal of Respiratory and Critical Care Medicine* 1994; **150**: 1545–9. [Incidence of bacteraemia, septic shock, and mortality is greater in patients with MRSA infection.]

Schimpff SC. Acquired immunodeficiency syndrome. In: Moossa AR, Robson MC, Schimpff SC, eds, *Comprehensive textbook of oncology*, pp. 605–14. Williams & Wilkins, Baltimore, 1986. [Review of opportunistic infections in patients with AIDS.]

Smith CR *et al.* Cefotaxime compared with nafcillin plus tobramycin for serious bacterial infections: a randomized, double blind trial. *Annals of Internal Medicine* 1985; **101**: 469–77. [Cefotaxime was as effective and less toxic than nafcillin plus tobramycin.]

Stoutenbeek CP *et al.* The effect of selective decontamination of the digestive tract on colonization and the infection rate in multiple trauma patients. *Intensive Care Medicine* 1984; **10**: 185–92. [SDD was able to reduce colonization and lung infection in ICU trauma patients.]

Torres A *et al.* Diagnostic value of quantitative cultures of bronchoalveolar lavage and telescoping plugged catheters in mechanically ventilated patients with bacterial pneumonia. *American Review of Respiratory Diseases* 1989; **140**: 306–10. [BAL and TPC diagnosed pneumonia in ventilated patients with similar accuracy.]

Torres A *et al.* Incidence, risk, and prognosis factors of nosocomial pneumonia in mechanically ventilated patients. *American Review of Respiratory Diseases* 1990; **142**: 523–8. [Twenty-four per cent of ventilated patients developed VAP. Mortality was greater in patients with (33 per cent) versus without (19 per cent) pneumonia.]

11.6 Central nervous system aspects

Christopher S. Garrard

The treatment of brain injury
Mechanisms of brain injury
Brain-orientated intensive care
Regulation of blood pressure
Regulation of temperature
Regulation of intracranial and cerebral perfusion pressure
Jugular venous saturation
Mechanical ventilation
Hyperosmolar therapy
Sedation therapy
Other agents
The patient in coma
Diagnosis
Management
Prognosis
Brain death
Brain-death criteria
Further reading

The treatment of brain injury

Preventative treatment for brain injury (cerebral protection) is only possible in certain circumstances such as in patients undergoing cardiopulmonary bypass, where hypothermia is particularly effective. Therapy commenced during a cerebral insult may be referred to as cerebral preservation; cerebral resuscitation describes any therapy begun after brain injury. Cerebral preservation occurs in the cold-water drowning victim, who makes a good neurological recovery despite a prolonged period beneath the water.

Mechanisms of brain injury

Several processes interact at cellular level to determine the viability of brain tissue following injury. Each of these processes can be altered or ameliorated by brain-orientated intensive care. The rapid re-establishment of cerebral blood flow is of critical importance: brain oxygen stores are depleted in about 10 s, and brain glucose and ATP stores are depleted 5 min after circulatory arrest. Some cerebral neurones are able to tolerate normothermic ischaemia for somewhat longer and the most extreme form of cellular damage, autolysis of brain tissue, begins after 1 to 2 h of no blood flow.

After circulatory arrest of more than 5 min, reperfusion and reoxygenation may cause irreversible brain damage, in the same way that other organ systems may be damaged. However, the widely held belief that a normothermic circulatory arrest lasting longer than 5 min is incompatible with recovery of normal brain function is not always correct. Several studies in animal models have shown good cerebral recovery after periods of circulatory arrest of more than 15 min and there have been reports of humans recovering after arrest times of up to 15 min.

The degree of brain injury, regardless of the nature of the insult, depends first upon the severity and duration of the injury and second, upon the speed and efficiency of resuscitation. To this should be added a third factor, the early application of brain-orientated intensive care.

Brain-orientated intensive care

Brain-orientated intensive care is broadly based and comprises several methods (Table 1). All organ systems must be actively supported and function restored as quickly as possible. The period immediately after cardiopulmonary resuscitation for circulatory arrest may be characterized by persistent metabolic acidosis and impaired cardiac contractility: the myocardium may need to be supported with inotropes and gas exchange maintained by mechanical ventilation until any acidosis resolves. Bicarbonate infusion is best avoided, since alkalinizing agents may worsen cerebral and myocardial intracellular acidosis.

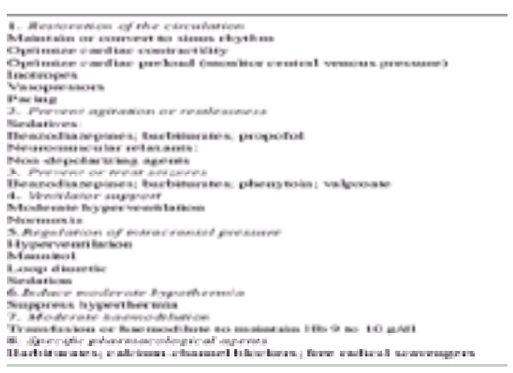

1. *Restoration of the circulation*
Maintain or convert to sinus rhythm
Optimize cardiac contractility
Optimize cardiac preload (monitor central venous pressure)
Inotropes
Vasopressors
Pacing
2. *Prevent agitation or restlessness*
Sedatives:
Benzodiazepines; barbiturates; propofol
Neuromuscular relaxants:
Non-depolarizing agents
3. *Prevent or treat seizures*
Benzodiazepines; barbiturates; phenytoin; valproate
4. *Ventilator support*
Moderate hyperventilation
Normoxia
5. *Regulation of intracranial pressure*
Hyperventilation
Mannitol
Loop diuretic
Sedation
6. *Induce moderate hypothermia*
Suppress hyperthermia
7. *Moderate haemodilution*
Transfusion or haemodilute to maintain Hb 9 to 10 g/dl
8. *Specific pharmacological agents*
Barbiturates; calcium-channel blockers; free radical scavengers

Table 1 Some methods of cerebral intensive care

Regulation of blood pressure

The systemic blood pressure should be supported with vasopressors if autoregulatory mechanisms are impaired. Cardiac arrhythmias must be rigorously controlled. Biochemical abnormalities such as hyperglycaemia should be corrected. If sodium bicarbonate has been used injudiciously during resuscitation, serum osmolality may be significantly increased: the use of mannitol as an osmotic diuretic could further exaggerate the hyperosmolar state and serial measurements of serum osmolality may need to be made. Hyperglycaemia should be controlled with insulin if necessary, since hyperglycaemia has been shown to worsen the outcome of both global and focal cerebral ischaemia.

The management of blood pressure in patients with brain injury and raised intracranial pressure is controversial. Hypotension carries the risk of global hypoperfusion; hypertension, although maintaining perfusion pressure, encourages the development of cerebral oedema. Common sense dictates a strategy aimed at avoiding extremes of blood pressure, with target systolic pressures between 100 and 160 mmHg. To this end, direct intra-arterial monitoring of blood pressure is preferred to non-invasive methods. Theoretically, vasodilators such as hydralazine, nitroprusside, and nitroglycerin (glyceryl trinitrate) should be avoided because they could steal blood flow away from the brain. In practice, however, agents such as nitroprusside and nitroglycerin are very effective in reducing blood pressure.

Regulation of temperature

Hyperthermia unrelated to infection may be observed in the early phases following head injury or brain ischaemia. Active efforts to lower the temperature should be made with antipyretics [paracetamol (acetaminophen)] and cooling. Small intravenous doses of chlorpromazine (2.5 mg for adults) may be effective when other measures fail. There have been several anecdotal reports that hypothermia at the time of anoxic brain injury associated with near drowning may confer benefit. Induction of moderate hypothermia (< 35°C) early in the resuscitation period and for the following 24 h is probably one of the few therapeutic methods that has been shown to improve survival and neurological recovery after head injury.

Regulation of intracranial and cerebral perfusion pressure

Cerebral oedema and the reduction in cerebral perfusion brought about by raised intracranial pressure is a particular problem in trauma and encephalitis. Almost 80 per cent of patients with severe head injury have some elevation in intracranial pressure; 50 per cent of deaths are directly attributable to raised intracranial pressure. It has also been estimated that up to one-third of deaths due to head injury may occur without any such increase. Cerebral oedema following circulatory arrest is less commonly associated with increased intracranial pressure. However, when intracranial pressure is raised after an anoxic or ischaemic insult, there is less evidence that measuring or attempting to control it confirms much benefit. The normal intracranial pressure is less than 15 mmHg at rest, with wide swings according to body activity and position, and is determined by the relation of the intracranial and intraspinal volume to the restricted, rigid skull vault and the more compliant, spinal subarachnoid space. The volume compartments within the skull are the brain, the cerebrospinal fluid, and the intravascular blood. Cerebrospinal fluid and blood are displaced as brain tissue swells. The correlation between intracranial pressure and increases in volume is not linear, but is roughly exponential. At high pressures (> 25 mmHg), a small increase in volume causes a steep rise in pressure, but with little damage until cerebral blood flow becomes restricted. An intracranial pressure sustained above 50 mmHg results in greatly reduced cerebral perfusion and is usually fatal.

Considerable controversy exists about the effects on outcome of monitoring intracranial pressure. Evidence to show that monitoring significantly improves survival after head trauma is less than convincing, but neurological deterioration often appears to follow episodes of raised intracranial pressure and patients with persistently raised intracranial pressure generally have a poor outcome. If attempts to lower intracranial pressure are to be made at all, measurement would seem a prerequisite. The clinician can then titrate therapy to achieve an acceptable cerebral perfusion pressure (mean arterial pressure minus mean intracranial pressure) of greater than the critical 40 mmHg. Perfusion pressures in the range 60 to 80 mmHg are ideal, but are not always attainable.

Clinical signs are poor indicators of intracranial pressure: temporal lobe lesions may be associated with IIIrd nerve signs at a lower pressure than lesions elsewhere and diffuse, bilateral cerebral lesions cause pupillary enlargement only when pressure is considerably greater than 25 mmHg. VIth nerve palsy is commonly described but is not a reliable indicator of raised intracranial pressure.

Techniques for the measurement of intracranial pressure use either a catheter or miniature strain gauge placed in the subarachnoid space or within a lateral cerebral ventricle. Miniature strain gauges (optical or quartz) can also be inserted, by appropriately trained surgeons, anaesthetists or intensivists, across the dura into brain tissue, providing reliable measurements of intracranial pressure for 6 to 8 days. An advantage of these strain-gauge systems is that the pressure reference point is the tip of the catheter and they provide an accurate measure of intracranial pressure independent of the position of the patient or degree of elevation of the head. Intraventricular catheters probably carry a higher risk of infection but also allow drainage of cerebrospinal fluid if needed. With such systems the external transducer must be placed at a fixed reference point such as the external auditory meatus.

A recent innovation, infrared transmission cerebral spectroscopy, may provide a convenient and non-invasive method of evaluating the state of oxygenation of the brain by measuring the haemoglobin saturation of blood within brain tissue of infants. Whether this technique will provide clinically useful information remains to be seen.

Jugular venous saturation

The oxygen saturation of blood within the jugular bulbs reflects the oxygen saturation of brain tissue. It can be measured by placing a 14-gauge catheter retrogradely up one jugular vein till it is just within the jugular bulb. Small samples of blood can be drawn intermittently and oxygen saturation estimated with an oximeter. Alternatively, and more conveniently, a fibreoptic catheter within the jugular bulb can measure venous saturation continuously, although daily calibrations must still be made. Normally, jugular venous saturation ranges between 65 and 75 per cent. Cerebral ischaemia is heralded by saturations below 50 per cent, while persistently high saturations above 95 per cent are consistent with a failure of oxygen extraction and may indicate brain death. Used in conjunction with monitoring of intracranial pressure, the measurement of jugular venous saturation may provide an important adjunct to the monitoring of cerebral perfusion. A serious limitation is that poor signal quality of infrared reflectance may invalidate much of the data obtained.

Mechanical ventilation

Mechanical hyperventilation to reduce the PCO_2 to below 3.5 kPa (< 30 mmHg) will rapidly lower intracranial pressure by causing cerebral vasoconstriction. These effects are seen within minutes, but are dissipated within 2 to 4 h. Further reduction in intracranial pressure requires a further increase in the level of hyperventilation and reducing the level of ventilation may result in an increase in pressure unless the underling pathology has been reversed. An excessive reduction in PCO_2 or elevation in arterial pH may reduce cerebral perfusion and oxygen offloading, and may therefore be counterproductive: this potentially harmful effect can be readily demonstrated by a falling jugular bulb saturation. Hyperoxia should probably be avoided. Hyperventilation necessitates higher mean airway pressures and the ventilator should be regulated to produce the highest alveolar ventilation for the lowest inflation pressures, which can often be attained by using an assisted mode such as volume-assist or pressure-assist in a non-paralysed patient. If the patient is intolerant of thisapproach and becomes agitated, full sedation and, if necessary, neuro-muscular relaxation will allow them to be maintained on synchronized intermittent mandatory ventilation or a control mode of ventilation. Although hyperventilation is an established treatment for raised intracranial pressure, there are no clinical trials proving its efficacy, and one recent study has suggested poorer long-term outcome (3–6 months) in patients hyperventilated following head injury. In experimental models of brain injury, hyperventilation neither reduces brain lactate nor improves the energy state (phosphocreatine:inorganic phosphate ratio) of the brain.

Hyperosmolar therapy

Hyperosmolar therapy with mannitol is widely recommended for production of a sustained reduction in intracranial pressure. Mannitol can be given as a single bolus dose of 0.5 to 1 g/kg body wt, or repeated as an infusion of 0.25 to 0.5 g/kg; this induces an osmotic gradient that reduces intracranial pressure for 3 to 4 h in most patients. Side-effects, due to systemic dehydration and intravascular volume depletion, include the precipitation of a non-ketotic hyperosmolar state in diabetics. Frusemide (furosemide) complements the action of mannitol, but may be less effective when used alone. After 24 h of hyperosmolar therapy, serum osmolality should be monitored, as there appears to little benefit from exceeding osmolalities in excess of 320 mosmol/kg.

Sedation therapy

A wide range of sedative agents can be used to lower intracranial pressure. Continuous infusion and intermittent boluses of benzodiazepines, barbiturates, and newer agents such as propofol are all effective. These agents will also reduce blood pressure if not controlled carefully. Barbiturate administration following acute cerebral injury may have several specific, beneficial effects. Control of agitation and restlessness by sedative agents minimizes changes in intracranial pressure, and seizures should be prevented. Barbiturates may also possess specific cerebral protective properties, although this is contentious and offers an avenue for further investigation. Clinical studies in humans have produced conflicting results, but some cerebral protection appears to have accrued from the use of barbiturates in patients with Reye's syndrome. There are sound reasons why barbiturates should protect the brain after injury (Table 2).

1. Decrease cerebral oedema
2. Reduce the cerebral metabolic rate
3. Reduce intracranial pressure
4. Decrease free fatty acid formation
5. Scavenge free radicals
6. Silence electroencephalographic activity
7. Reduce brain infarct size in focal ischaemia
8. Increase tolerance to carotid occlusion

Table 2 Potential cerebral protective effects of barbiturates

Of the barbiturates available, thiopentone (thiopental) is an excellent sedative agent and anticonvulsant. With prolonged use it is distributed widely throughout the body and its effects are then slow to clear. Until properly conducted clinical studies can convincingly show a specific cerebral protective effect of barbiturates, justification for

their use will be limited to their sedative and anticonvulsant effects.

Other agents

The calcium-channel blockers, particularly nimodipine, improve survival after subarachnoid haemorrhage by reversing cerebral vasospasm. They probably act by reducing the intracellular release and accumulation of free ionized calcium and inhibition of mitochondrial activity that occurs during ischaemia and reperfusion. Studies in animals and in man have examined the potential protective effect of calcium-channel blockers in a much wider range of cerebral injury, with some indication of a beneficial response after cardiac arrest. Despite a degree of cerebral selectivity, the use of calcium-channel blockers may be limited by adverse effects such as systemic hypotension.

Free radicals are short-lived, highly reactive compounds that are released during the reperfusion of ischaemic neuronal issue and initiate sustained lipid peroxidation. Free radical scavengers such as superoxide dismutase, desferrioxamine, thiopentone (thiopental), vitamin E, vitamin C, glutathione, chlorpromazine, mannitol, and some dextrans have been used in an attempt to reduce the effects of reperfusion injury of several organs including the brain. Preliminary studies have shown no protective effect of desferrioxamine and a calcium-channel blocker in a dog ischaemic brain model, although desferrioxamine alone appeared to reduce cerebral damage after shorter periods of cerebral ischaemia. The use of free radical scavengers cannot yet be recommended, however rational their application may seem.

Corticosteroids have no beneficial effect in cerebral oedema associated with head injury, and their effects after cardiac arrest have not been adequately explored. Their use in brain-orientated intensive care, other than to reduce high intracranial pressure associated with brain metastases, cannot be recommended.

The patient in coma

Coma is a state of depressed level of consciousness and should be distinguished from the abnormal content of consciousness that results in confusion, delirium, or psychosis. The drowsy patient (level of consciousness) may also be confused (content of consciousness), although the converse may not be true. The level of consciousness should always be quantified in terms of some standard such as the Glasgow coma score (Table 3). By accurately recording coma scores, small changes in conscious level can be recognized and if necessary, acted upon. Most patients admitted to intensive care in coma will have a score of less than 6.

	Points
Eye opening (maximum 4 points):	
Spontaneous	4
To speech	3
To pain	2
None	1
Verbal response (maximum 5 points):	
Orientated	5
Confused conversation	4
Inappropriate words	3
Incomprehensible sounds	2
None	1
Best motor response (maximum 6 points):	
Obeys commands	6
Localizes to pain	5
Withdraws to pain	4
Abnormal flexion	3
Extensor response	2
None	1

Highest score = 15, lowest = 3.

Table 3 Glasgow coma score

The level of consciousness emanates from the reticular-activating system and the cerebral hemispheres. Transient loss of consciousness (concussion) after head trauma has been attributed to rotation of the hemispheres about the midbrain/diencephalic junction. Ultrastructural changes occur in the brain neurones, even with very transient loss of consciousness after trauma. Prolonged loss of consciousness (coma), whether or not related to trauma, is usually associated with lesions in either both hemispheres, one hemisphere with compression of the upper brainstem, or in the brainstem reticular-activating system. When coma is due to injury to the reticular-activating system, other signs of brainstem dysfunction such as pupillary, caloric, and oculomotor signs re usually apparent.

The predominant causes of coma vary from centre to centre: drug intoxication and head injury may be more common in the inner city while metabolic disease such as diabetic coma, cerebrovascular events, and mass lesions may predominate in other hospitals. A careful clinical history and physical examination are therefore paramount in determining the underlying cause of coma. A list of common causes of coma is shown in Table 4.

Primary cerebral nervous causes
1. Vascular
Haemorrhage; infarction; vasculitis
2. Trauma
Direct neuronal damage
Secondary haemorrhage
Secondary ischaemia/hypoxia
Cerebral oedema
3. Infection
Encephalitis; meningitis; abscess
4. Postictal
5. Malignancy
Primary mass; metastases; oedema
Secondary cerebral nervous causes
1. Drugs
Alcohol; sedatives; narcotics
2. Metabolic
Glucose metabolism:, hyper-/hypoglycaemia
Osmolar disturbances
Endocrine disorders:, pituitary, adrenal, thyroid
Electrolyte disturbances
Hypercalcaemia
Hepatic encephalopathy
Uraemic encephalopathy
Extreme hypo-/hyperthermia
3. Circulatory failure
Cerebral hypoperfusion:
sepsis syndrome
heart failure
hypovolaemic shock

Table 4 Some common causes of coma

Diagnosis

Diagnostic dilemmas often arise when coma is due to drugs or metabolic causes, when physical signs may be consistent with both cerebellar and brainstem lesions, depending upon the severity of the metabolic disturbance. However, pupil size and reactivity are usually well preserved despite severe obtundation. Patients with severe myasthenia gravis, Guillain–Barré syndrome, or basilar arterial thrombosis may appear to be in coma, although they may be receptive. The so-called locked-in syndrome due to cranial nerve and limb paralysis may be extremely difficult to distinguish from coma: the clinical significance of this becomes only too apparent when considering patients as potential organ donors.

Computed tomography of the head may define anatomical abnormalities and confirm a clinical diagnosis. Although the tomographic scan is generally a safe procedure and is available to most critically ill patients, it should not be used indiscriminately, bearing in mind the risks associated with transporting the patient to the radiography department. Conversely, changes in coma score may mean that repeated scans are necessary to determine the requirement for surgical intervention.

The pupillary reflex must be assessed with care and attention: a cursory evaluation with a poor light source is clearly inadequate. Contraction of only 1 mm is sufficient to indicate pupillary responsiveness. Unequal pupils may be found after instillation of mydriatics or direct trauma to the eye.

The presence of full and conjugate eye movements indicates an intact pons and midbrain. Normally, turning the head from side to side will cause the eyes initially to move conjugately in the opposite direction followed quickly by eye movement in the direction of head movement. The presence of ‘doll's eyes', which remain fixed in relation to the head, indicates brainstem damage and should be confirmed by caloric testing. Abducted (outward-looking) eyes are common in stuporous or drowsy patients and should not be interpreted as a bad prognostic sign. As coma deepens the eyes may become fixed in the primary position.

Limb movement affords the most reliable method of detecting asymmetric neurological function. Movements of a limb away from the body (i.e., pushing away) indicate an intact corticospinal pathway and this type of purposeful limb should always be distinguished from flexion, extension, or pronation. The triple flexion response of the hip, knee, and ankle joints in response to tactile stimuli is not necessarily purposeful, since it is integrated at the level of the spinal cord.

Management

Protection of the airway takes precedence over all other aspects of management. Failure to protect the airway and control the cervical spine in a comatose patient with head injury must be considered to be serious omissions. Positioning of the patient in the semiprone or Fowler's position may be adequate in the drowsy patient with an intact gag reflex but as consciousness fades, endotracheal intubation may be required. The technique of intubation should combine speed, safety, and avoidance of factors that may increase intracranial pressure.

Prognosis

Predicting the outcome of coma, regardless of the underlying cause, is a difficult task. Apart from unequivocal brain death there are no clinical signs that confidently predict outcome. Young patients may have many early clinical signs consistent with poor outlook and yet may make a full recovery. Assessment of prognosis based on neurological findings should be interpreted broadly, and in conjunction with the patient's age and previous medical condition. The clinician's approach to management should also incorporate a consideration of the patient's premorbid wishes, if these are known.

Prognostication in cases of head injury is usually more accurate than in non-traumatic coma, for which there are many causes. While more than 90 per cent of patients in whom pupillary or oculomotor reflexes are absent in the first 6 h after injury will die, 4 per cent may still make a significant recovery: this proportion is large enough to make the clinician hesitate to abandon support in the first 24 h. Some favourable and unfavourable signs of recovery in the comatose, head-injured patient are summarized in Table 5. Such signs can be cautiously applied to patients with coma due to other causes, but with less confidence.

Favourable signs of recovery in first 72 h
Early wakening
Verbal response
Intact caloric response
Fending off external stimuli
Normal muscle tone
Unfavourable signs in the first 24 h
Absent spontaneous eye opening
Absent spontaneous eye movement
Absent pupillary response
Absent corneal reflex
Absent caloric response
Absent deep tendon reflexes
Absent muscle tone

Table 5 Predictive signs of outcome in head injury patients

Somatosensory-evoked potentials may offer useful information, particularly in children. Absence of cortical responses with preserved lower-level brain potentials appears to be a reliable indicator of poor outlook in anoxic coma and following head injury.

In practice, an expectant policy allowing time for the development of convincing signs, full evaluation, and neurological consultation to take place seems to be the most acceptable. In those patients in whom the outlook is grave, family and relatives can come to terms with reality. A period of 48 to 72 h may be needed initially. The primary physician should indicate clearly the management policy to nursing staff and house staff, particularly after full and frank discussion with the nearest relatives. In some cases, extubation may be possible and more conservative management plans adopted.

Brain death

The accurate and reliable determination of brain death has become an essential part of clinical practice in the critical care unit. Brain-dead patients need to be identified so that either unnecessary life support can be discontinued or so that transplantation organ donation can be considered and organized. It is unnecessary to establish the diagnosis of brain death, unless organ transplantation or the cessation of life-support measures are deemed appropriate.

The criteria used acknowledge the fact that independent, self-sustaining life is not possible in patients with brainstem death. It is also inevitable that within a short period of time (usually hours, but at most 2 or 3 days), circulatory collapse and cardiac arrest will occur, which will be unresponsive to any medical measures.

Before undertaking brain-death testing in the comatose patient, several preconditions must be satisfied. First, a definitive cause for coma must be established. Second the patient requires mechanical ventilation and third, enough time must have elapsed from the onset of coma to determine that the brain injury is irreversible. This time interval will vary according to the cause of the brain injury. For example, following head trauma or a major intracranial haemorrhage, 6 to 12 h may be sufficient to be sure that recovery is not possible. In the case of hypoxic brain injury secondary to cardiac or respiratory arrest, 24 to 48 h may be needed. If any doubt exists about possible effects of drugs, periods extending up to 7 days may need to be considered. The elimination of drugs and their duration of activity is extremely variable. Some agents, such as nortriptyline, diazepam, methadone, and phenobarbitone (phenobarbital) have prolonged plasma half-lives of between 10 and 150 h, and may therefore have very extended activity.

Brain-death criteria

The tests of brainstem function include many of the cranial nerve reflexes and the activity of the respiratory centre. Pupillary reaction to light depends upon the integrity of the optic nerve, the midbrain, and the oculomotor nerve (III). The corneal reflexes are dependent on the afferent trigeminal (V) and the facial nerve (VIII).

Dysfunction of both cerebral hemispheres disinhibits the brainstem reflex mechanisms for conjugate eye movements. Normally, turning the head from side to side elicits easy or ‘loose' conjugate eye movements in the opposite direction. Brainstem dysfunction cause one or both eyes to fail to move fully and conjugately. Abducens nerve (VI) palsy indicates pontine dysfunction or diffuse increase in intracranial pressure. Oculomotor nerve (III) palsy with absence of full adduction of one eye also indicates midbrain pathology. The pattern of spontaneous respiration may become periodic when both hemispheres are affected, and irregular or absent when the inferior pons and medulla are involved.

Established hospital guidelines for the diagnosis of brain death should be rigorously adhered to. The completion of a standard form to be made part of the patient's record is generally desired. Table 6 lists the most widely accepted brain-death criteria.

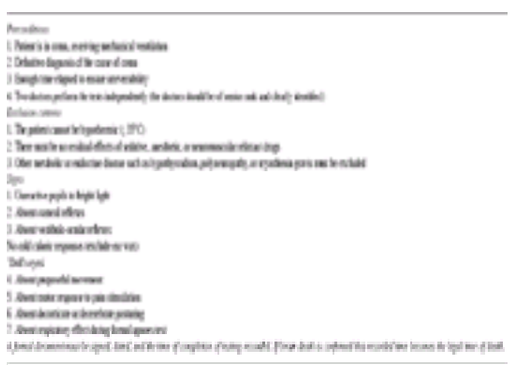


Table 6 Criteria for the diagnosis of brain death

Specific criteria for the diagnosis of brain death may vary in minor details depending on local or national recommendations. These variations may include the requirement for two sets of observations, about 24 h apart, and an electroencephalogram showing electrocerebral silence. An advantage of performing two tests lies not in improved accuracy but in providing time for the co-ordination of transplant teams. Some protocols require the radiographic demonstration of absent cerebral blood flow.

The conduct of the apnoea test may also vary from centre to centre but relies on achieving an indisputable CO_2 stimulus to respiration in the absence of respiratory depressants such as sedatives, narcotics, neuromuscular relaxants, or hypoxia. The patient should be monitored with continuous ECG and pulse oximetry if available. Mechanical ventilation should be adjusted to stabilize the PCO_2 at a normal high level. The patient should then be preoxygenated with 100 per cent O_2 for about 10 min. The ventilator can then be disconnected at the endotracheal tube and a catheter passed down the endotracheal tube to facilitate insufflation of oxygen at 1 to 2 min. The patient is carefully observed to detect any respiratory movement of the chest or abdomen. Arterial oxygen saturation can be accurately monitored with a pulse oximeter. After 5 to 10 min, blood gases can again be sampled. The patient is then temporarily supported with the ventilator until the blood gases confirm a high CO_2 stimulus (6.5 kPa, 50 mmHg) and the absence of hypoxia. In practice PCO_2 between 8 and 10 kPa can be achieved by this protocol. If no respiratory effort has been detected and other signs of brainstem dysfunction have been demonstrated, the patient may be considered brain dead.

It is the responsibility of the doctor performing the tests of brain death to explain the implications of such tests to relatives before the tests are completed. It is essential that the relatives are aware that survival is not possible in the presence of positive brain-death criteria.

Further reading

Abramson NS. *et al.* Randomized clinical study of thiopental loading in comatose survivors of cardiac arrest. *New England Journal of Medicine* 1986; **314**: 397–403.

Ames A III, Nesbett FB. Pathophysiology of ischemic cell death. I. Time of onset of irreversible damage; importance of the different components of the ischemic insult. *Stroke* 1983; **14**: 219–23.

Andrews PJ, Dearden NM, Miller JD. Jugular bulb cannulation: description of a cannulation technique and validation of a new continuous monitor. *British Journal of Anaesthesia* 1991; **67**(5): 553–8.

Bircher N, Safari P. Cerebral preservation during cardiopulmonary resuscitation. *Critical Care Medicine* 1985; **13**: 185–90.

Bruce DA, Gennarelli T, Langfitt T. Resuscitation from coma due to head injury. *Critical Care Medicine* 1978; **6**: 254–8.

Cooper P *et al.* Dexamethasone and severe head injury. A prospective, double blind study. *Journal of Neurosurgery* 1979; **51**: 307–10.

Dearden NM, Midgley S. Technical considerations in continuous jugular venous oxygen saturation measurement. *Acta Neurochirurgica Supplementum (Wien)* 1993; **59**: 91–7.

Farber JL, Chien KR, Mittnacht S. The pathogenesis of irreversible cell injury in ischemia. *American Journal of Physiology* 1981; **102**: 271–81.

Fleischer JE *et al.* Failure of deferoxamine, an iron chelator, to improve neurologic outcome following complete cerebral ischemia in dogs. *Stoke* 1987; **18**: 124–7.

Gentleman D, Dearden M, Midgley S, Maclean D. Guidelines for resuscitation and transfer of patients with serious head injury. *British Medical Journal* 1993; **307**: 547–52.

Gopinath SP *et al.* Jugular venous desaturation and outcome after head injury. *Journal of Neurology, Neurosurgery, and Psychiatry* 1994; **57**: 717–23.

Gudeman S, Miller J, Becker D. Failure of high-dose steroid therapy to influence intracranial pressure in patients with severe head injury. *Journal of Neurosurgery* 1979; **51**: 301–6.

Jennet B *et al.* Predicting outcome in individual patients after severe head injury. *Neurosurgery* 1979; **4**: 283–8.

Jones PA *et al.* Measuring the burden of secondary insults in head-injured patients during intensive care. *Journal of Neurosurgical Anesthesiology* 1994; **6:** 4–14.

Kirsch JR, Helfaer MA, Koehler RC, Traystman RJ. Brain ischaemia and reperfusion injury. In: Bihari D, Holiday JW, ed. *Update in Intensive Care and Emergency Medicine*, pp. 66–84. Springer, Berlin, 1989.

Lim KH *et al.* Prevention of reperfusion injury of the ischemic spinal cord: use of recombinant superoxide dismutase. *Annals of Thoracic Surgery* 1986; **42**: 282–6.

Lind B, Snyder J, Safer P. Total brain ischemia in dogs: cerebral physiological and metabolic changes after 15 min of circulatory arrest. *Resuscitation* 1975; **4**: 97–101.

McCormick PW *et al.* Noninvasive cerebral optical cerebral spectroscopy for monitoring cerebral oxygen delivery and haemodynamics. *Critical Care Medicine* 1991; **19**: 89–97.

Marion DW *et al.* Treatment of traumatic brain injury with moderate hypothermia. *New England Journal of Medicine* 1997; **336**(8): 540–6.

Marshall LF *et al.* Pentobarbital therapy for intracranial hypertension in metabolic coma. Reye's syndrome. *Critical Care Medicine* 1978; **6**: 1–6.

Michael R *et al.* Mechanisms by which epinephrine augments cerebral and myocardial perfusion during cardiopulmonary resuscitation in dogs. *Circulation* 1984; **69**: 822–35.

Murr R, Stummer W, Schurer L, Polasek J. Cerebral lactate production in relation to intracranial pressure, cranial computed tomography findings, and outcome in patients with severe head injury. *Acta Neurochirurgica (Wien)* 1996; **138**: 928–36.

Pallis C, Harley DH. *ABC of brain stem death*. British Medical Journal, London, 1996.

Plum F, Posner J. *The Diagnosis of stupor and coma*. Davis, Philadelphia, 1980.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. *Defining death. Medical, legal and ethical issues in the determination of death*. US Government Printing Office, Washington DC, 1981.

Ritter AM, Gopinath SP, Contant C, Narayan RK, Robertson CS. Evaluation of a regional oxygen saturation catheter for monitoring $SjvO_2$ in head injured patients. *Journal of Clinical Monitoring* 1996; **12**(4): 285–91.

Robertson CS, Cormio M. Cerebral metabolic management. *New Horizons* 1995; **3**(3): 410–22.

Robertson CS, Gopinath SP, Goodman JC, Contant CF, Valadka AB, Narayan RK. $SjvO_2$ monitoring in head-injured patients. *Journal of Neurotrauma* 1995; **12**: 891–6.

Rockoff M, Marshall L, Shapiro H. High-dose barbiturate in therapy humans: a clinical study of 60 patients. *Annals of Neurology* 1979; **6**: 194–8.

Safar P, Bircher N. Cardiopulmonary cerebral resuscitation. In: *An introduction to resuscitation medicine*, 3rd edn. World Federation of Societies of Anaesthesiologists/Baillière Tindall, London, 1986.

Safar P, Grenvik A, Abramson NS, Burcher N, ed. Reversibility of clinical death: symposium on resuscitation research. *Critical Care Medicine* 1988, **16**: 919–1086.

Schanne FAX *et al.* Calcium dependence of toxic cell death: a final common pathway. *Science* 1979; **206**: 700–3.

Shapiro HM. Postcardiac arrest therapy. Calcium entry blockade and brain resuscitation. *Anesthesiology* 1985; **62**: 384–7.

Shiu GK, Nemoto EM. Barbiturate attenuation of brain free fatty acid liberation during global ischemia. *Neurochemistry* 1981; **37**: 1448–56.

Smith DS, Rehncrona S, Siesjo BK. Barbiturates as protective agents in brain ischemia and as free radical scavengers *in vitro. Acta Physiologica Scandinavica* 1980; **42**: 129–34.

Steen PA *et al.* Cerebral blood flow and neurologic outcome when nimodipine is given after complete cerebral ischaemia in the dog. *Journal of Cerebral Blood Flow Metabolism* 1984; **4**: 82–7.

Todd MM, Chadwick HS, Shapiro HM, Dunlop BS, Marshall LF, Dueck R. The neurologic effects of thiopental therapy following experimental cardiac arrest in cats. *Anesthesiology* 1982; **57**: 76–86.

Wei EP *et al.* Effects of oxygen Radical on cerebral arterioles. *American Journal of Physiology* 1985; **248**: H157–62.

Xu RX *et al.* Specific inhibition of apoptosis after cold-induced brain injury by moderate postinjury hypothermia. *Neurosurgery* 1998; **43**(1): 107–14.

11.7 Gastrointestinal aspects

Christopher S. Garrard

Hepatic dysfunction
Total parenteral nutrition
Hepatic hypoperfusion
Drug-induced liver disease
Viral hepatitis
Acute hepatic failure
Aetiology
Clinical features
Treatment
Hepatorenal syndrome
Clinical presentation
Treatment
Ascites
Acute pancreatitis
Clinical presentation
Treatment
Multiple organ-system involvement in pancreatitis
Stress ulcer prophylaxis
Drug therapy
Further reading

Hepatic dysfunction

Varying degrees of hepatic dysfunction, with a variety of causes, are seen in the surgical patient admitted to the critical care unit. There is usually evidence of cholestasis, hepatocellular injury, or a combination of both. Cholestasis is associated with increases in serum alkaline phosphatase and conjugated bilirubin, while hepatocellular damage results in increases in enzymes and the prothrombin time. Extrahepatic cholestasis should be excluded by ultrasonography of the gallbladder and biliary tree.

The most common causes of hepatic dysfunction in the critical care unit are (1)total parenteral nutrition, (2)hepatic hypoperfusion, (3)drug-induced liver disease, and (4)hepatitis.

Total parenteral nutrition

Elevation of the liver enzymes (aspartate aminotransferase), alkaline phosphatase, and serum bilirubin is often seen 2 to 6 days after the start of intravenous nutrition. Ultrasonography may be required to exclude extrahepatic obstruction. The underlying cause is unclear, but histological examination shows elements of fatty infiltration as well as periportal inflammation and intrahepatic cholestasis. Secondary bacterial colonization of the biliary tract may also play a part. Biochemical abnormalities can be quite marked, but the clinical course is generally benign. Transition to enteral feeding is associated with resolution of the biochemical and pathological abnormalities.

Hepatic hypoperfusion

Hypoperfusion of the liver due to sepsis syndrome, cardiogenic shock, haemorrhagic shock, burns, or trauma can result in mild to severe hepatic dysfunction, depending upon the severity and prolongation of the ischaemic insult. Elements of intrahepatic cholestasis may be present in patients with sepsis syndrome, although the major liver injury is probably a consequence of hypoperfusion associated with endotoxaemia.

Drug-induced liver disease

A wide range of drugs may be associated with intrahepatic cholestasis, hepatitis, or even massive necrosis. These include erythromycin, chlorpromazine, tolbutamide, and the anabolic steroids (cholestatic), isoniazid, methyldopa, nitrofurantoin (hepatitis-like), carbon tetrachloride, and paracetamol (acetaminophen) (massive necrosis).

Viral hepatitis

Viral hepatitis should be considered in any patient developing jaundice and the biochemical features of hepatitis. Hepatitis A and B, and less commonly C and E, are responsible for between 50 to 70 per cent of cases of acute liver failure. Serological screening and blood precautions should be undertaken in any patients in the intensive care unit presenting with or developing acute liver failure.

Acute hepatic failure

Hepatic failure that develops acutely over a period of several days or weeks without pre-existing liver disease is referred to as fulminant. It is due to sudden, massive necrosis of hepatocytes, and is followed rapidly by the onset of encephalopathy. In the critical care unit, the likely cause is either viral hepatitis or an overdose of paracetamol (acetaminophen).

Aetiology

Viral hepatitis accounts for up to 70 per cent of cases of acute hepatic failure. The most common causative agent is hepatitis B virus, reported to be responsible for 25 to 75 per cent of cases of viral hepatitis. Hepatitis A is somewhat less common, while hepatitis E accounts for about 20 per cent of cases. Hepatitis C is often found in association with hepatitis B and is now recognized as major cause of chronic progressive liver disease. The association of hepatitis D (delta agent) with hepatitis B is a serious finding and is usually followed by a rapid deterioration in liver function.

Drug reactions and overdose with paracetamol (acetaminophen) account for about 20 per cent of cases of acute liver failure. Other drugs, such as non-steroidal anti-inflammatory agents and antidepressants, are associated with lesser degrees of liver dysfunction but can cause acute liver failure. Halothane anaesthesia is often proposed as a cause of drug-induced hepatic failure, but increased awareness of this and the avoidance of repeated halothane anaesthesia have greatly reduced this problem. Drug overdose explains about 10 per cent of cases of acute hepatic failure, the most common agents being paracetamol (acetaminophen) and ferrous sulphate.

Fulminant hepatic failure is a common feature of *Amanita phalloides* poisoning: ingestion of 50 g of the mushroom is frequently fatal. Other toxins, such as carbon tetrachloride, methylbromide, chloroform and xylene, may occasionally produce acute hepatic failure. Rare causes of fulminant hepatic failure include profound hypovolaemic shock, Budd–Chiari syndrome, fatty liver of pregnancy, Reye's syndrome, Wilson's disease, and hyperthermia.

Clinical features

Jaundice and fetor hepaticus may often be apparent, while other characteristic signs of liver dysfunction, such as spider naevi, ascites and liver palms, are generally absent. The serum bilirubin and transaminases are usually high, although some patients progress to coma before bilirubin becomes significantly elevated.

Hepatic encephalopathy

Hepatic encephalopathy is usually attributed to neurotoxic materials such as ammonia and mercaptans, derived from the metabolism of nitrogenous compounds in the bowel. These compounds enter the systemic circulation having bypassed the liver through anatomical or functional shunts. Entry of aromatic amino acids into the central nervous system may also account for neurological disturbances. The severity of hepatic encephalopathy can be classified into the four stages shown in Table 1.

Factors that may precipitate or be associated with hepatic encephalopathy include the administration of sedative or hepatotoxic drugs, bleeding into the gastrointestinal tract, increased dietary protein, and diuretic-induced hypokalaemic alkalosis.

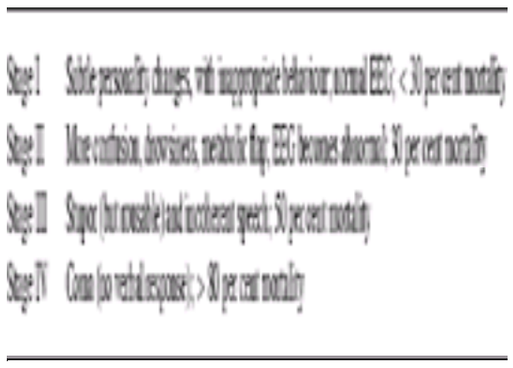


Table 1 Classification and mortality of hepatic encephalopathy

Haemorrhagic manifestations

These result from decreased synthesis of clotting factors II, V, VII, IX, and X, disseminated intravascular coagulation, and splenic sequestration of platelets. Active bleeding from varices, gastritis, and peptic ulcers further depletes clotting factors, and these should be managed aggressively by replenishing clotting factors with fresh frozen plasma and undertaking surgical or endoscopic intervention if necessary.

Renal failure

Renal failure may complicate hepatic failure for several reasons. Gastrointestinal haemorrhage or overuse of diuretics may result in prerenal azotaemia. Severe or prolonged hypotension, sometimes in association with use of nephrotoxic drugs, may result in vasomotor nephritis (acute tubular necrosis). A serious complication of hepatic failure is the hepatorenal syndrome, a poorly understood condition that results in functional renal failure without clear histopathological lesions in the kidneys. If renal replacement therapy is required, continuous haemofiltration is preferred to short periods of haemodialysis, since it is less likely to increase intracranial pressure in the encephalopathic patient.

Pneumonia

Nosocomial infection occurs in up to 30 per cent of patients and is a common cause of death. There may be a qualitative defect in immune defences due to impaired reticuloendothelial cell clearance, and abnormal leucocyte migration and complement-dependent opsonization.

Other features

The adult respiratory distress syndrome is common in patients with endstage liver failure and is almost uniformly fatal. Impaired clearance of vasoactive substances by the hepatic reticuloendothelial system may contribute to the development of acute lung injury.

Spontaneous bacterial peritonitis should always be suspected in the patient with ascites: this may occur in the absence of peritonism, leucocytosis, or fever. A diagnostic abdominal paracentesis is essential.

Hypoglycaemia may develop because of reduce hepatic glycogen reserves and decreased gluconeogenesis.

Treatment

The treatment of hepatic encephalopathy includes reducing the gastro-intestinal protein load by restricting dietary protein intake, preventing gastrointestinal bleeding, and encouraging intestinal emptying with agents such as lactulose. In stage I and II encephalopathy, protein intake should be restricted to fewer than 40 g/day, and this should be further reduced to fewer than 20 g/day in patients with stage III and IV encephalopathy. Caloric intake, preferably as dextrose, should be maintained in excess of 2000 cal (8.4 kJ)/day. Patients with liver damage may be lipid intolerant, and the presence of lipaemic serum requires a reduction of lipid intake. Increased amounts of aromatic amino acids (tryptophan, tyrosine, and phenylalanine) may result in decreased synthesis of normal neurotransmitters and, therefore, enhanced synthesis of false neurotransmitters. These may contribute to encephalopathy, although this process might be reversed by the administration of branched-chain amino acids. Neomycin, 0.5 to 1 g orally every 6 h, suppresses ammonia production by colonic bacteria. The addition of lactulose, initially in doses of 50 ml orally every 2 h, acts as an osmotic laxative and decreases ammonia absorption from the intestinal lumen by increasing gut intraluminal acidity.

Most patients with stage IV coma have intracranial pressures raised to more than 30 mmHg. Intracranial pressure can be conveniently monitored using a microtransducer at the tip of an electrode introduced through a small (3 mm) drill hole in the skull across the dura into the brain substance. Coagulation defects should be corrected before devices for monitoring intracranial pressure are inserted. Control of raised intracranial pressure follows the general recommendations for reducing intracranial pressure in the head injured. The patient should be nursed with the head raised 15° to 20°. Any factor that could obstruct cerebral venous drainage must be avoided. Cerebral perfusion pressure (mean arterial pressure minus intracranial pressure) should be defended by ensuring adequate cardiac preload and the judicious use of pressor agents if necessary. Short-term lowering of intracranial pressure can be achieved by hyperventilation to lower the PCO_2. However, the cerebral vasoconstriction induced by such manoeuvres can potentially reduce cerebral blood flow. Intravenous mannitol (0.5 g/kg) can be given, in repeated doses, providing serum osmolality remains less than 320 mosmol/kg. Thiopentone (thiopental) given as an infusion will reduce intracranial pressure with relatively little effect on blood pressure. Corticosteroids have no effect on raised intracranial pressure in hepatic encephalopathy.

Fluid and electrolyte balance requires continuous and meticulous attention. Patients with hepatic failure are generally intolerant of sodium loads and hypokalaemia is poorly tolerated. Volume depletion must be avoided to minimize the risk of vasomotor nephropathy and monitoring the central venous pressure is almost obligatory. Target central venous pressures of 8 to 12 mmHg will generally ensure adequate preload. Lower pressures, especially in patients receiving assisted ventilation, carry the risk of volume depletion. If oliguria develops, the differential diagnosis includes prerenal azotaemia and the hepatorenal syndrome. Both conditions are associated with a low fractional excretion of sodium (FE_{Na}) so that exclusion of volume depletion is critical to clinical management.

Haemodialysis or haemofiltration should be performed in patients with renal failure if recovery from liver failure is expected or liver transplantation envisaged. It may be difficult to distinguish between hepatorenal syndrome and vasomotor nephropathy. If doubt exists over the precise cause of renal failure, an active approach to management with haemofiltration is indicated.

Use of all types of sedation should be minimized as far as possible. If the use of sedatives cannot be avoided, continuous infusion of a short-acting benzodiazepine such as midazolam is acceptable, provided the infusion is discontinued intermittently to reassess the need for sedation. Coexistent renal failure further reduces the clearance of a wide range of medications and lactate.

Coagulopathy is assessed by daily monitoring of the prothombin time. Although there may be little response to the administration of vitamin K, this should be given empirically in doses of 10 mg intravenously daily for 3 days. Vitamin K stores may be severely depleted in the presence of liver necrosis, and supplementation ensures that vitamin K is available if there is any hepatocellular regeneration.

Fresh frozen plasma should be given to patients with severe bleeding or as prophylaxis before an invasive procedure. A target prothrombin time of less than 1.5 times control is required for most surgical procedures. Clotting-factor concentrates should be avoided since they may precipitate disseminated intravascular coagulation; this is otherwise uncommon unless there is associated sepsis. The most common site of bleeding is the gastrointestinal tract. Bleeding from sites other than surgical wounds is somewhat infrequent.

Hypoglycaemia should be expected in all patients with fulminant hepatic failure. In the encephalopathic patient the symptoms of hypoglycaemia may not be apparent

and it will only be detected by regular monitoring of the blood sugar. Dextrose should be given intravenously as intermittent boluses, or as a continuous infusion if necessary. Administration of glucagon is inappropriate in patients with depleted glycogen stores.

Death secondary to acute hepatic failure usually results from cerebral oedema, sepsis, or haemorrhage. Mortality is related to the stage of encephalopathy (Table 1). Those patients who survive fulminant hepatic failure have a high probability of complete recovery as the liver regenerates over the following 3 to 4 months.

Hepatorenal syndrome

The hepatorenal syndrome refers to renal failure of unknown aetiology that occurs in a fully hydrated patient with severe, often progressive, liver disease. Urine biochemistry is characteristic and renal histopathology unremarkable. The pathogenesis of the hepatorenal syndrome appears to involve intense intrarenal vasoconstriction and alteration in renal cortical blood flow, possibly due to an imbalance of prostaglandins and thromboxane. Early clinical experience suggested that only orthotopic liver transplantation was able to reverse the hepatorenal syndrome, although more recent evidence has reported up to 50 per cent survival in hepatorenal failure following paracetamol (acetaminophen) overdose. Renal failure is reversed after liver transplantation, suggesting that it is due to circulating or systemic factors. Excessive use of diuretics, sepsis, abdominal paracentesis, and gastrointestinal haemorrhage are all associated with an increased risk of the hepatorenal syndrome and suggest a functional aetiology.

Clinical presentation

Typically, patients who develop the hepatorenal syndrome have signs of advanced liver disease, such as icterus, ascites, palmar erythema, spider angiomas, and hepatic encephalopathy. Laboratory findings confirm hepatic failure, with elevated serum bilirubin, liver enzymes, and ammonia. The impaired synthetic function of the liver usually results in hypoalbuminaemia and prolongation of the prothrombin time. Hyponatraemia, hypokalaemia, and alkalosis are common accompaniments. Hyponatraemia results from proximal tubular reabsorption of sodium and water with non-osmotic release of vasopressin. The first sign of hepatorenal syndrome is usually oliguria. The major differential diagnoses include prerenal azotaemia and vasomotor nephropathy (acute tubular necrosis). The patient with hepatorenal syndrome appears well hydrated and volume depletion can usually be excluded. The FE_{Na} is less than 1 per cent and the urinary sodium less than 10 to 15 mmol/1, in contrast to vasomotor nephropathy, which characteristically has a high urinary sodium content—over 30 mmol/1. If prerenal azotaemia is suspected, a cautious fluid challenge consisting of about 250 ml of crystalloid or colloid infused over 30 min can be tried. If doubt remains about the volume status of the patient, a flow-directed pulmonary arterial catheter should be inserted for accurate assessment of left heart filling pressures. The urine sediment should be examined carefully. The presence of cellular casts suggests vasomotor nephropathy rather than hepatorenal syndrome. Obstructive uropathy should be excluded by renal ultrasonography or, if necessary, retrograde pyelography.

Treatment

When treating any patient with advanced liver disease, it is essential to avoid factors that precipitate the hepatorenal syndrome. Adequate nutrition must be maintained and electrolyte abnormalities corrected. Diuretics should be used judiciously and ascites drained only for specific indications. Any intercurrent infection must be treated aggressively. The use of certain drugs, such as the non-steroidal anti-inflammatory agents, angiotensin-converting enzyme inhibitors and tetracyclines, should be avoided, and aminoglycosides should not be used unless this is unavoidable.

There is currently no specific treatment for the hepatorenal syndrome; therapy is directed at supporting the underlying liver disease. Haemodialysis and haemofiltration may control the metabolic disturbances of renal failure but the outcome is determined by the degree of liver impairment. A summary of these supportive measures is shown in Table 2.

Fluid and electrolytes
Include volume depletion/lower central venous pressure or pulmonary arterial catheter
Optimize fluid balance
Water restriction
Sodium restriction
Administer dDAVP
Correct hypokalaemia
Haemorrhage/coagulation
Blood
Fresh frozen plasma
Vitamin K
Encephalopathy
Restrict protein intake
Ensure adequate carbohydrate
Gastrointestinal tract emptying
Branched-chain amino acids
Infection
Antibiotics
Reduce endotoxaemia
Support circulation
Ascites
Diuretics
Paracentesis
Colloid infusion
Peritoneal/venous shunting
Portosystemic shunting
Nutrition and feature

Table 2 Supportive measures in the hepatorenal syndrome

Pharmacological agents that have been used in an attempt to reverse the hepatorenal syndrome include dopamine, isoprenaline, phenoxybenzamine, phentolamine, aminophylline, mannitol, metaraminol, adrenaline (epinephrine), papaverine, and prostaglandin A_1. All have been without effect in double-blind studies. Recovery of renal function has occurred following peritoneojugular (LaVeen) and portacaval shunting. Orthotopic liver transplantation appears to offer the best hope for reversal of the hepatorenal syndrome but is only available for a relatively small group of suitable patients.

Ascites

Abdominal ascites is a common feature of the patient with the hepato-renal syndrome. The common indications for reduction of ascites include abdominal pain, intravascular coagulation and compression of renal veins, gastro-oesophageal reflux (with erosion of varices), and restriction of ventilation by diaphragm splinting. Simple measures such as bedrest and salt restriction are essential preliminary measures. Spironolactone or amiloride, combined if necessary with a loop diuretic, can then be added. In a proportion of patients, direct drainage of ascitic fluid is required. The aim is to drain sufficient fluid to alleviate symptoms without inducing the hepatorenal syndrome. Infusion of colloid should accompany the paracentesis to ensure that the patient does not become volume depleted. Paracentesis is occasionally followed by a brisk diuresis, particularly if the renal veins have been compressed by the ascitic fluid.

Acute pancreatitis

Mild acute pancreatitis has a mortality of less than 3 per cent and rarely requires critical-care management. Ten per cent of patients with pancreatitis have severe disease, however, accompanied by failure of multiple organ systems. Although early reports of severe acute pancreatitis were associated with a mortality of up to 75 per cent, more recent mortality figures have fallen towards 20 per cent. Most cases of pancreatitis are the result of either alcohol abuse or cholelithiasis. Other causes of pancreatitis are listed in Table 3. There is no proven specific therapy for acute pancreatitis save avoiding precipitating factors and applying general supportive measures.

Frequent (80–90 per cent)
Alcoholism
Cholelithiasis
Idiopathic (10 per cent)
Occasional (10–15 per cent)
Postoperative pancreatitis
Trauma
Penetrating peptic ulcer
Drugs (thiazides, sulphonamides, azathioprine)
Virus (mumps, infectious mononucleosis)
Hypercalcaemia
Hyperlipidaemia
Ampullary obstruction
Rare (less than 3 per cent)
Connective tissue disease
Oestrogens
Hypothermia
Malnutrition
Afferent loop syndrome
Translumbar aortography

Table 3 Causes of acute pancreatitis

Clinical presentation

Acute pancreatitis should be considered in any patient presenting with abdominal pain, nausea, and vomiting. Other common clinical signs and symptoms of acute pancreatitis are shown in Table 4. Abdominal pain is invariably present and radiates to the back in 50 per cent of patients. Acute haemorrhagic pancreatitis often presents dramatically with shock, confusion, coma, oliguria, and abdominal ecchymoses. Severity and outcome from pancreatitis can be estimated by simple clinical scoring systems such as those proposed by Ranson and Imrie.

Signs and symptoms	*% cases*
Abdominal pain	90
Nausea or vomiting	70
Fever	70
Abdominal distension/ascites	60
Shock	20
Dyspnoea	20
Icterus	20
Confusion/coma	10
Cullen's or Gray-Turner's sign	< 5
Subcutaneous fat necrosis	< 2

Table 4 Clinical presentation of acute pancreatitis

Laboratory confirmation usually rests on the detection of elevated serum or urinary amylase. Elevated serum amylase also may be found in patients with intestinal obstruction, bowel perforation or infarction, pregnancy, renal failure, major burns, and carcinomas of the lung, oesophagus, or ovary. Isoenzyme analysis may help differentiate pancreatic and salivary sources of amylase.

Treatment

The treatment of severe acute pancreatitis can follow either of two philosophies (see Chapter 33.1). The first involves the surgical removal of necrotic material to reduce the risk of organ system failure and secondary infection. The second adopts a conservative approach, delaying surgery until indicated for complications such as haemorrhage or infection.

Regardless of the management philosophy, the critical-care treatment during the acute, toxaemic phase is directed at the management of primary pancreatic failure (diabetes mellitus), hypovolaemia secondary to fluid sequestration, and multiple organ-system failure. Analgesia can be achieved using agents such as pethidine or a synthetic opiate (alfentanil). Morphine probably should be avoided in view of a propensity to cause spasm of the sphincter of Oddi.

Late complications result from local, intra-abdominal, and systemic infection and often progress to the sepsis syndrome. This late phase has a significant effect on the final mortality from acute pancreatitis and requires all the resources of the critical care unit.

Hyperglycaemia and ketoacidosis require volume replacement, continuous insulin infusion according to blood sugar (2–8 units/h is generally adequate), provision of intravenous dextrose, and correction of electrolyte imbalance. Metabolic acidosis can also result from lactic acidosis and, occasionally, from renal failure.

Hypocalcaemia occurs in about 25 per cent of patients, usually within the first week of the illness. It is associated with hypoalbuminaemia, calcification of areas of intra-abdominal fat necrosis, and parathyroid dysfunction.

Treatment is supportive, there being no specific, effective therapeutic agent available. Drugs such as somatostatin, glucagon, cimetidine, aprotinin, and anticholinergics do not appear to be effective. Peritoneal lavage is favoured by some, as is open drainage of the pancreatic bed following pancreatectomy. Much more extensive and controlled evaluation of these forms of treatment is required.

Multiple organ-system involvement in pancreatitis

Any of the major organ systems can be involved in patients with acute pancreatitis. Organ system failure in pancreatitis is one of the most common indications for admission to the critical care unit. Since organ system failure is responsible for most of the deaths associated with pancreatitis, prevention, or at least early recognition, is paramount.

Cardiovascular system

Hypotension may result from the release of bradykinin and the sequestration of abdominal fluid. The intravascular fluid deficit can be large, requiring several litres for replacement, and monitoring of the central venous pressure is generally required. The nature of the replacement fluid is probably less critical than the volume. Fluid resuscitation should continue until a central venous pressure of about 10 mmHg can be achieved and satisfactory urine output of 0.5 to 1.0 ml/kg per min is maintained. Since much of the intravascular volume deficit is due to fluid extravasation into the interstitial space, there may be considerable oedema despite underlying volume depletion. The electrocardiogram may show prolongation of the ST interval and non-specific T-wave changes. Tachyarrhythmias can develop and are probably related to the metabolic abnormalities of pancreatitis or coexistent alcoholic cardiomyopathy.

Respiratory system

More than half of all patients with pancreatitis will have chest radiographic changes such as patchy infiltrates, collapse, or pleural effusion. Pleural effusions of an exudative type occur in up to 15 per cent of patients, the left hemithorax being most commonly involved. Arterial hypoxaemia is evident in up to 75 per cent of patients: this usually responds to supplemental oxygen but about 20 per cent of patients develop acute respiratory distress syndrome requiring endotracheal intubation and positive-pressure ventilation.

Host defence

Leucocytosis, fever, and complement activation are features of acute pancreatitis. As with all patients under intensive care, nosocomial infection is increasingly likely after the second week of admission. However, use of antibiotics should be reserved for patients with clear signs of infection of sepsis. Prophylactic antibiotics are not recommended.

Gastrointestinal system

Gastric erosions and ileus are common accompaniments of acute pancreatitis. Prophylaxis against stress-associated gastric erosions and total parenteral nutrition should be provided for the mechanically ventilated patient. Ulcer prophylaxis can be provided by cytoprotective agents (sulcralfate), antacids, or H_2-blockers. The last increase gastric pH, and may increase the risk of nosocomial pneumonia.

Blood and coagulation

Anaemia may result from pancreatic or gastric bleeding. Patients with alcoholic cirrhosis and oesophageal varices are at further risk of gastrointestinal haemorrhage. Prolongation of prothrombin time may occur due to coexistent liver failure or disseminated intravascular coagulation: replacement of clotting factors with fresh frozen plasma is indicated, particularly if there is active bleeding. Platelet transfusion also may be required if there is bleeding in the presence of a platelet count of less than 50 000/mm^3.

Renal system

Prerenal azotaemia and vasomotor nephropathy will occur unless rigorous fluid resuscitation is undertaken. Any fall in hourly urine volumes to below 0.5 ml/min per kg should prompt an immediate evaluation of the volume status of the patient using all measures available. In established renal failure, haemofiltration provides a convenient means of controlling the fluid balance, electrolyte balance, and azotaemia. Only rarely is haemodialysis is required.

Central nervous system

Agitation, confusion, and even seizures may occur, possibly due to breakdown of cerebral lipid fractions. General principles of sedation, analgesia and seizure suppression in the intensive care unit are usually adequate.

Stress ulcer prophylaxis

Critically ill surgical patients are at risk from gastrointestinal stress ulceration. Ulcers are most often sited in the fundus of the stomach or the first part of the duodenum. They are small, multiple, superficial, well demarcated, and usually without surrounding oedema. Although all critically ill patients are at risk of bleeding from stress ulceration, the incidence is probably lower than was first estimated. Disease categories particularly associated with stress ulceration include sepsis, burns, spinal and head injuries, coagulopathies, and respiratory, renal and hepatic failure. The regular use of antacids, cytoprotective agents, H_2-receptor antagonists or proton-pump inhibitors is generally considered to have reduced the incidence of bleeding but has not been shown to influence survival. With aggressive, proactive treatment of shock, sepsis and other critical illness, together with greater reliance on enteral feeding, stress ulcer prophylaxis may no longer be a routine.

Targeted prophylaxis with a cytoprotective agent such as sucralfate or H_2-receptor antagonists (ranitidine) should be given to patients considered to be at the highest risk. The pathogenesis of stress ulceration is multifactorial and suggested mechanisms are listed in Table 5. Some early endoscopic studies of asymptomatic, critically ill patients suggest that stress ulceration occurs in 60 to 100 per cent of cases and is present within hours of admission to the intensive care unit. The reported incidence of bleeding from stress ulceration varies according to the patient type and the definition of bleeding. Macroscopic bleeding occurs in about 15 per cent of patients not given prophylaxis compared to 5 per cent who have. However, massive haemorrhage, or the need for surgical intervention, are much less common. In addition, there is general agreement that the incidence of stress ulcer bleeding has fallen significantly over the last two decades. One recent study reported an incidence of less than 1 per cent.

1. Reduced acid production (bacterial colonization)
2. Gastric mucosal ischaemia due to shock or sepsis
3. Failure of protective mucous secretion
4. Formation of free radicals
5. Cytokine activation
6. Reperfusion injury
7. Mucosal damage from refluxing bile salts and pancreatic juice.

Table 5 Suggested mechanisms for gastric stress ulceration

Drug therapy

H_2-receptor antagonists (ranitidine 50 mg thrice daily intravenously or 150 mg twice daily by mouth) reduce stress ulcer bleeding when compared to placebo, and are as effective as antacids. However, there are many potential side-effects and drug interactions of the H_2-receptor antagonists. If gastric bleeding occurs despite ranitidine, a proton-pump inhibitor such as omeprazol (40 mg orally) can be substituted. Of concern is the risk of bacterial colonization of gastric secretions associated with suppression of gastric acid. Aspiration of colonized pharyngeal secretions may then lead to nosocomial pneumonia. The significance of this risk has yet to be fully resolved.

Sucralfate (aluminium sucrose octasulphate) acts a cytoprotective agent by adhering to the damaged gastric mucosa, preventing further damage by acid and pepsin. It also increases the production of mucus and bicarbonate by the gastric mucosa, and may even have a direct antibacterial effect. Sucralfate is as effective as antacids or H_2-receptor antagonists at preventing bleeding from stress ulceration. Sucralfate suspension is given via a nasogastric tube, 1 g 4-hourly. Some studies have suggested that nosocomial pneumonia is less common in patients receiving sucralfate compared to those receiving antacids and H_2-receptor antagonists. On the other hand, recent microbiological studies have disputed the significance of gastric colonization in the genesis of nosocomial pneumonia.

Stress ulcer bleeding is less common in patients who are receiving enteral nutrition; this could reflect a lower severity of illness in view of the maintenance of gastrointestinal absorptive function or be a direct protective effect of nutrients in the stomach and gastrointestinal tract. Enteral feeding does not necessarily render the gastric secretions more acid unless ‘rest periods' from feeding of 2 to 4 h are introduced into the feeding regimen. If enteral feeding has been successfully established, no additional prophylaxis is needed in the absence of gastric bleeding.

Further reading

Boyer TD. Major sequelae of cirrhosis. In: Wyngaarden JB, Smith JH, Jr, ed. *Cecil textbook of medicine*, 18th edn, pp. 847–52. Saunders, Philadelphia, 1988.

Driks MR *et al.* Nosocomial pneumonia in intubated patients given sucralfate as compared with antacids or histamine type 2 blockers. *New England Journal of Medicine* 1987; **317**: 1376–82.

Epstein M, Perez G, Oster JR. Management of renal complications of liver disease. *Journal of Intensive Care Medicine*, 1988; **3**: 71–86.

Ettien JT, Webster PD. The management of acute pancreatitis. *Advances in Internal Medicine* 1980; **25**: 169–98.

Fraser CL, Arieff AI. Hepatic encephalopathy. *New England Journal of Medicine* 1985; **313**: 685–873.

Gordon DG, Tedesco FJ. Recognising systemic manifestations of acute pancreatitis. *Journal of Critical Illness* 1987; **2**: 77–83.

Hawker F. Liver dysfunction in critical illness. *Anaesthesia and Intensive Care* 1991; **19**: 165–81.

Imrie CW, Benjamin JS, Ferguson JC *et al.* A single-centre double blind trial of trasylol therapy in primary acute pancreatitis. *British Journal of Surgery* 1978; **65**: 337–41.

Imrie CW. The management of severe acute pancreatitis. *Clinics in Critical Care Medicine* 1987; **11**: 93–107.

Jensen DM. Portal–systemic encephalopathy and hepatic coma. *Medical Clinics of North America* 1988; **70**: 1081–92.

Marshall JB. Acute pancreatitis. A review with an emphasis on new developments. *Archives of Internal Medicine* 1993; **151**: 1185–98.

Mayer DA *et al.* Controlled trial of peritoneal lavage for the treatment of severe acute pancreatitis. *New England Journal of Medicine* 1985; **312**: 399–404.

O'Grady JG *et al.* Coagulopathy of fulminant hepatic failure. *Seminars in Liver Disease* 1986; **6**: 159–63.

Payne JA. Causes and complications of fulminant hepatic failure. *Journal of Intensive Care Medicine* 1986; **1**: 216–23.

Ranson JH, Rifkind KM, Roses DF *et al.* Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology, Obstetrics* 1974; **139**: 69–81.

Ranson JHC. Etiologic and prognostic factors in human acute pancreatitis: a review. *American Journal of Gastroenterology* 1982; **77**: 633–8.

Rattner DW, Warshaw AL. Surgical intervention in acute pancreatitis. *Critical Care Medicine* 1988; **16**: 89–95.

Runyon BA. Care of the patient with ascites. *New England Journal of Medicine* 1994; **330**: 337–42.

Scharschmidt BF. Acute and chronic hepatic failure and hepatic transplantation. In: Wyngaarden JB, Smith JH, Jr, ed. *Cecil textbook of medicine*, 18th edn. Saunders, Philadelphia, 1988.

Schuster DP. Stress ulcer prophylaxis: In whom? With what? *Critical Care Medicine* 1993: **21**: 4–6.

Sweny P. The hepatorenal syndrome. In: Rainford D, Sweny P, ed. *Acute renal failure*, pp. 83–112. Farrand, London, 1990.

Tryba M. Sucralfate versus antacids or H_2 antagonists for stress ulcer prophylaxis: a meta-analysis on efficacy and pneumonia rate. *Critical Care Medicine* 1991; **19**: 942–9.

Williams R, Gimson A. Management of acute liver failure. *Clinics in Gastroenterology* 1985; **14**: 93–104.

11.8 Grief: breaking bad news: offering organ donation

Patricia M. Franklin

Introduction
The grief process
Common behaviour patterns in the early phase of the grief process
Anger, anxiety, depression, isolation
Healing: new behaviours which will enable the bereaved to continue with their lives
High risk groups: intense bereavement reactions
The needs of the relatives during the crisis time
To have questions answered honestly: to know the prognosis
Support, comfort, and cultural and religious needs
To be near the patient
Communicating with the family
Informer and supporter
Verbal and non-verbal cues
Informing of death
Emotional reactions
Anger
Hysteria
Withdrawal: isolation
Continuing care
Sudden or traumatic death
Brainstem death
Informing of brainstem death
Option of organ donation
Multiple organ donation
When to offer the option of donation
Who should approach the family?
How to approach the family
Staff support
Viewing the body after death
Further care
Further reading

Introduction

Death and bereavement are an integral part of human life and the care of those who grieve is an important part of clinical practice. Regardless of culture, bereavement is a stressful experience which can lead to physical and psychological morbidity. Lazare estimates that 10 to 15 per cent of the patients attending the Mental Health Clinic in Massachusetts General Hospital have an unresolved grief reaction underlying their medical condition.

The way in which relatives are informed of a death and the way in which they are supported in the early phases of grief can have a lasting influence on later adjustment. Care that is appropriate and effective may reduce the incidence of ‘morbidity', particularly amongst ‘high risk groups'.

The purpose of this section is to outline the grief process, to discuss techniques of ‘breaking bad news', and to suggest strategies for supporting the bereaved which clinicians may use to promote a healthier grief reaction, thus reducing some of the damaging consequences of bereavement.

Medical staff receive little training in this area of their work and caring for and supporting families is stressful to all concerned. However, with a knowledge of the grief process and appropriate communication skills, staff can feel more at ease with the situation and offer empathetic and understanding care. Furthermore, such care can be viewed as a major and rewarding challenge for all professionals.

The grief process

Grief is generally described as a psychological process by which people fill the gap in their lives after a large part of their world has been lost. Engel described this process as ‘grief work': the work of mourning by which we can become emancipated from bondage to the deceased, readjust to the environment in which the deceased is missing, and begin to form new relationships.

Lindemann first described ‘The stages of bereavement' in 1944. Other classical texts have supported and expanded this early theory. Most of these writers outlined three stages within grieving: an immediate stage with shock, disbelief, and denial; an intermediary stage with a growing awareness accompanied by anger, anxiety, and depression; and a final stage of resolution, acceptance, and healing.

More recently, theorists have argued that the concept of bereavement in stages is too structured and that such ‘classical texts may not entirely reflect how it is to suffer loss' (McClaren). Each individual will respond to bereavement in a unique way and the concept of stages may negate the individual pattern of coping. The grief process is neither universal or predictable with no two families responding in the same way, and with individual family members reacting with different emotional responses. Therefore, generalizations and comparisons may at best be unhelpful and at worst damaging, particularly if clinicians try to fit individuals into a fixed model of grief.

Thus, grief is now viewed as a ‘process' containing common behaviour patterns and reactions. The intensity of the reactions may be affected by other factors, such as the nature of the relationship between the patient and the bereaved, the age of the deceased, the type of death (expected or sudden), and the bereaved's responses to previous experiences and relationships. Research and clarification regarding the various individual and familial behaviour patterns have been recorded and therefore it is possible to recognize patterns and plan and implement appropriate support and care.

Common behaviour patterns in the early phase of the grief process

Common behaviour patterns in the early phase of the grief process include numbness, panic, shock, denial, inability to concentrate and make decisions, inability to absorb information and use it effectively, demanding and irrational behaviour, aggressive and abusive behaviour, withdrawal, and passivity.

An understanding of these early patterns of behaviour is particularly important to clinicians because such behaviour may occur soon after the death and, thus, at the time the bereaved are meeting with health professionals within the hospital environment.

The phase of stunned numbness is described by a bereaved relative in Speck's book as a ‘cotton wool time when there seems to be an individual blanket between you and the world'. Others speak of being ‘frozen in disbelief' and like a ‘zombie'. There is a safety in this numbness in that it denies the more frightening reactions of helplessness, utter despair, and intense fear.

Denial can be interpreted as a psychological defence mechanism which prevents too much emotional pain at any one moment. Numbness, denial, shock, and disbelief are obviously increased in cases of sudden and traumatic death where there has been no preparation for the terrible news and no possibility of anticipatory grieving.

Numbness, shock, and disbelief may last for hours, days, or weeks and may damage and impede the exchange of information and all forms of communication. Denial may play a role throughout the grief process, emerging and subsiding at different times. Extended denial lengthens the grief process and may result in the bereaved feeling the reality of the death at the time that others appear to have ‘forgotten'.

Anger, anxiety, depression, isolation

The gradual awareness of the reality of the situation is often accompanied by anger and anxiety. Such anger may be directed towards God, the deceased, members of

the caring professions, or it may be internalized and used inwardly against the bereaved themselves. Internalized anger is often linked with feelings of guilt and is most apparent following sudden and traumatic death or the death of a child.

Yearning and searching for the deceased may also occur and is often accompanied by feelings of emptiness and intense isolation. The loneliness may become extreme with thoughts of not being understood by family and friends. Such intense responses may engender a fear in the bereaved that they are ‘going insane' and may result in them becoming totally absorbed with their own feelings to the exclusion of partners and family. Thus, increasing feelings of alienation. Sadness, depression, and exhaustion may gradually develop and may continue for many months.

Healing: new behaviours which will enable the bereaved to continue with their lives

Gradual readjustment and reintegration may occur as the intensity of the emotional pain lessens and the bereaved may start to look forward and find some new purpose in living or new ways of behaving which will enable them to continue with their lives.

Phrases such as ‘letting go of the deceased' and ‘moving on' have been used in the past, but it is widely recognized now that many relatives may wish to find ways of sustaining the bond with the bereaved and integrating this bond into future life.

High risk groups: intense bereavement reactions

Several writers have outlined factors which may indicate a high risk of an intense bereavement reaction requiring additional or specific support. These factors may include:

1. unexpected loss: patient in the younger age group with no previous history of illness;
2. suicide;
3. sudden loss: no preparation for the death;
4. bereaved perceives the family as unsupportive—lack of social network;
5. the relationship between the deceased and the relative is perceived as ambivalent—often manifested as dependency, anger, or guilt; and
6. the death of a child; it is widely recognized that parental grief is more severe, complex, protracted, and traumatic thangrief following any other bereavement.

Research has shown that professional counselling can significantly reduce morbidity in cases of an intense bereavement reaction. The effect of the counselling is to reduce the risk in high risk individuals to that of low risk individuals without counselling.

The needs of the relatives during the crisis time

The needs of the relatives during the ‘crisis' time when the patient is critically ill have been investigated. Molter reported that the five most important needs were:

- to feel there is hope
- to feel that the hospital staff cared about the patient
- to know the prognosis
- to have questions answered honestly
- to be near the patient

Hampe and Breu reported similar needs:

- to be with the dying person
- to be helpful to the dying person
- to be assured that the dying person is comfortable
- to be kept informed of the dying person's condition
- to know of the impending death
- to experience and express emotions
- to comfort and support family members
- to be accepted, supported, and comforted by health care professionals
- to be relieved of anxiety

It may be difficult to meet all these needs; individual needs should be met as and when they arise. Staff need to be flexible. Also, communicating with the family at regular intervals and giving them honest information will help them through the distressing phase when they alternate between hope for recovery and fear of death. Clinicians should focus on the needs of the family and view themselves as a ‘companion', accompanying them through all aspects of the situation.

To have questions answered honestly: to know the prognosis

The family need to know the truth of the situation. Truth in itself is not damaging but its presentation must be carefully planned. Blandy notes that one should ‘Temper the wind to the shorn lamb. Dilute frankness with gentleness and wherever possible offer hope'. Relatives need the facts about the clinical condition and a realistic prognosis with its implications for them as a family. Truth may not be the information that they are hoping for, but it allows them to take control and to select options and make decisions. Staff should strive at this time to develop a rapport with relatives so that trust is established allowing them to inform, support, and offer choices.

It is important that clinicians listen to the family and hear the concerns that the situation has raised for them. Families should be included in discussions concerning care and, if the family wish, children should also be encouraged to be present and involved. Children and relatives that are excluded may imagine a situation far worse than reality.

Support, comfort, and cultural and religious needs

Relatives in a crisis situation require continual support. A relative who is alone should be comforted by an empathetic carer until another family member, friend, or acceptable person can come and support them. While waiting, relatives should be kept as comfortable as possible, preferably in a suitably furnished private room near to telephone and toilet facilities. They should be offered refreshments and allowed to smoke.

To be near the patient

The family should be allowed to sit with the patient as soon as possible and should be encouraged to help with appropriate aspects of care.

Staff must be aware of cultural differences and religious beliefs; interpreters and religious advisers should be contacted to add comfort and assist in communication.

Communicating with the family

Informer and supporter

When communicating with the family it is helpful to use two people: an informer and a supporter. The clinician is often the informer; the supporter is often a nurse, a religious adviser, or another member of the caring team. The roles of the informer and the supporter should be kept separate. The family may blame or reject the informer; should this happen, the supporter can offer physical comfort, repeat information, and offer further support. It is important that the informer does not take such rejection personally. The family is not rejecting them but the message that they have given.

Before communicating with the family it is important for the informer to prepare himself/herself, both physically and mentally. Evidence of trauma such as bloodstains must be removed, as should other barriers which will impede communication (for example surgical masks).

The informer should also familiarize himself/herself with the family situation, noting the names of the principal relatives and their relationship with the patient. If there is

a large group of relatives it is helpful to speak directly to the immediate next of kin, using first names as appropriate.

Meetings with the relatives should be planned so there is time for discussion. They should take place in a private relatives' room where the family can express their thoughts and feelings freely.

Verbal and non-verbal cues

Upon meeting the family the informer and supporter should introduce themselves, shake hands with the relatives, and sit near to them. It is important to maintain a calm, unhurried approach and to offer them the time to ask questions; never hover in the doorway as if ready to make a hasty exit.

Non-verbal cues indicating the gravity of the situation should be used so that the relatives receive some preparation for the information. Facial expressions should be serious, as should the tone of voice. Speak to them in non-technical language and give information slowly, gradually sowing the seeds of the seriousness of the situation. Look them in the eye and speak softly, with spaces in between words and sentences. Never try to overprotect people from unpleasant reality: if there is a possibility of death, it is essential to inform them and help them to prepare. Staff must also be sensitive to relatives' needs and use physical comfort as appropriate, such as holding a hand or placing a comforting arm on their shoulder.

Relatives may try to minimize the seriousness of the situation by misinterpreting information or by only hearing certain parts of the message. Shock and disbelief can block communication and impede understanding; the informer should invite questions in order to find out what has been understood and then clarify and repeat again. Relatives may also confront different staff members with the same questions about the patient's status hoping for a more positive message. It is necessary to maintain good communication between team members so that the same information is given by all.

The family must be encouraged to express their thoughts and feelings. Staff should not tell them how to feel (for example ‘do not upset yourself'); they have something very real to be upset about. It helps at this time to encourage them to talk about themselves and their families; insight gained into their world and their feelings can result in greater empathy and understanding from the carers.

The supporter should arrange further meetings to give the family progress reports while also attempting to resolve any practical problems that arise for them.

Informing of death

In most cases relatives will wish to be at the bedside at the time of death and staff should strive to fulfil this wish, offering them privacy. If it is not possible for the family to be at the bedside, a member of staff who has been in continual contact with the relatives during the ‘crisis' time should ideally be the person to inform them of death. Again, the information should be given in a private area by an informer with a supporter present.

Buckman suggests that the death of a patient may cause clinicians to experience ill-founded feelings of failure, anger, and guilt at not being able to save the life. It is essential that such feelings are recognized, discussed with colleagues, and resolved before the meeting with the family. If these emotions persist they may make the informer defensive and will hinder empathetic communication.

All members of staff will approach this task with trepidation at the thought of giving the message, and with feelings of helplessness at the thought of trying to ameliorate the relatives' suffering. There are no ‘correct' words to use at this time, but it is important to give maximum preparation to the family with a warning of bad news before the actual verbal message: ‘I am afraid that I have bad news for you'—pause—if possible get the relative to say ‘Do you mean that he/she is dead?'. If this response is not forthcoming proceed with ‘We did all that we could to save your wife/husband (use the first name if possible)—pause—'but I am afraid that he/she has died'. The words ‘has died' or ‘is dead' should be used rather than other ambiguous phrases such as ‘passed on' or ‘left us' as these can be misconstrued.

Once the verbal message has been given the carers should anticipate and be prepared for a variety of different emotional reactions. Men and women often have different ways of expressing grief. Men tend to find relief in rage and anger early on, and retire to brood alone; women often need to talk endlessly about the deceased. When everyone within a family circle is devastated they are likely to find it particularly difficult to help one another.

Emotional reactions

Anger

Anger is a frequent reaction to intense feeling and an expression of grief. In order to express such anger the relative may shout, rush about the room, kick and punch the air, the wall, or the furniture. It is important not to do anything to increase this anger. Do not attempt to restrain the relative and do not become defensive and enter into an argument. The best response is to remain calm and to wait for the anger to subside. Staff should show no criticism of this response but should offer support and care.

Hysteria

Regardless of how distressed the relative may be it is important to remember that the outburst will cease after a short time. It is best to remain quiet and calm, and to sit and wait for the hysteria to abate. It is important not to appear judgmental, shocked, or disapproving, but to accept that this is simply another expression of grief. Physical contact and comfort should be offered as the hysteria subsides.

Withdrawal: isolation

Isolation and withdrawal produce perhaps the strongest feelings of helplessness in the carers. It is impossible to communicate adequately or to know how the bereaved relative is thinking and feeling. Bereaved fathers find it particularly difficult to discuss or share their grief, but it is possible to offer a silent yet caring ‘presence' in this situation. Eventually, it may become acceptable to ask gentle questions to establish a rapport and thus elicit a response. It is more helpful to the bereaved to be drawn out and to express reactions rather than to continue suppressing their feelings. The earlier the expression of grief, the healthier the outcome.

Continuing care

Once the initial reaction has subsided it is important to answer the questions that the family may have and to offer them support in the tasks that lie ahead. Never try to console them with platitudes or tell them ‘that you know how they feel'. Grief and its pain is unique to each individual and it is impossible to feel as another does in such a situation. The bereaved will never again have the opportunity to work through this most difficult time and therefore the staff should give them the space and freedom to do so. Hodge states that: ‘The grief work must be done. There is no healthy escape from this—People have a natural protective tendency to avoid the unpleasantness of the grief work, but it is necessary and the more actively it is done, the shorter will be the period of grief'. Simple expressions, such as ‘I am very sorry', bring the most comfort at this time and if spoken with warmth and understanding they impart more than eloquent words or false statements. The family will also gain comfort from the knowledge that the death was peaceful or pain free, and that the deceased was not alone.

Sudden or traumatic death

Sudden or traumatic death robs the family of preparatory grieving and the shock, numbness, and disbelief will be more intense in such situations. During the initial period the bereaved often feel disorientated, powerless, and vulnerable. Therefore, breaking bad news in such circumstances requires empathy, clear communication, and support that will help the relatives emerge from the acute state of shock.

Difficulties in communication may occur because both the clinician and the relative may be influenced by their own fears, thoughts, and feelings. The bereaved may misinterpret the message, may pretend not to hear, or may simply not understand due to their confusion and distress. The clinician may be anxious and not able to put thoughts and feelings into words, thus, speaking too quickly and using language that is too technical. Effective and empathetic communication requires clear non-verbal clues (i.e. serious intonation of voice, serious facial expressions, and caring body posture) combined with simple information using terms that the bereaved will understand.

Following sudden loss it is likely that the family will have many questions that will need to be answered with honesty, as this will help them to make some sense of meaning from the death. The use of ‘open-ended questions', such as ‘how can we help you' and ‘what other information would you like' help to develop rapport and trust, can ease the conversation, and will encourage the relative to seek the answers that they need. Also, acknowledging the relatives feelings and emotions, ‘you

must be very shocked' will help the family to discuss their feelings and thus influence the grief process in a positive way. Again, the aim must be to support, inform, and offer choices because helping the bereaved to make decisions themselves will also help them to regain their coping skills. Active decision-making stimulates a healthy grief process.

Many relatives benefit from a further meeting with the clinician at a later stage so that unanswered questions may be raised and discussed when the numbness and shock have passed. As mentioned earlier, psychological morbidity can be reduced with early counselling, particularly for relatives who have no supportive social networks or who are unable to support each other.

One of the most difficult deaths to understand and accept is the situation where the patient has suffered a major brain insult and is subsequently found to be brainstem dead.

Brainstem death

In the case of brainstem death it is especially important to consider the content and the timing of the information to be given to the family. In this situation the relatives have to understand and accept a new concept of death. Traditional acceptable images of death involve a lifeless body that is cold and asystolic. However, brainstem death presents an image of life in a setting of high technology and hope where the victim is warm and has a heart beat and is breathing, albeit on a machine. Thus, the situation and setting suggest life and hope to the family, in sharp contrast to the message of death that is given to them by the clinician.

Informing of brainstem death

The same preparations and procedures for information giving should apply as mentioned earlier using a dual approach. It is essential that the informer and supporter both understand and accept the brainstem death concept themselves and that they use language that the family can understand. Any hesitation or fudging of the explanation will confuse the relatives and may introduce hope that recovery is possible. The message to be given to them must stress that irreparable damage to the brain has occurred and that there is no hope of recovery; furthermore that death of the brainstem is evident and death of the brainstem is death of the person.

The family must be allowed time to assimilate and accept this information. The central facts may need to be repeated at several meetings before the relatives can understand the diagnosis and its implications.

Option of organ donation

To feel there is hope

As stated earlier, Blandy and others stress that ‘wherever possible one should offer hope to the family'. However, if death has occurred, all hope of recovery for their loved one is lost, but the bereaved can be offered an option of hope and life for others through organ and tissue donation. Tissue donation (i.e. corneas, heart valves, and skin) can be offered in most cases of asystolic death. Furthermore, kidney donation can also follow asystolic death in certain circumstances. Therefore, clinicians should consider the possibility of donation in every case of death and should seek specific advice from the local transplant co-ordinator service.

Multiple organ donation

Brainstem death can offer the family the option of multiple organ donation. For organ donation criteria see Table 1.

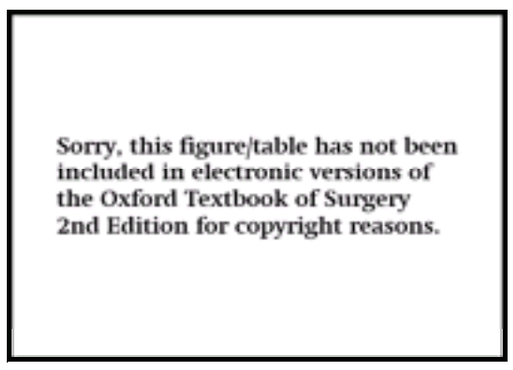


Table 1 Donor criteria

Recent reports suggest that many clinicians are reluctant to introduce the option of donation because they fear that such a suggestion may increase the grief of the bereaved. However, research studies have shown that families gain enormous comfort from the knowledge that their tragedy has resulted in life for others. A survey in New Zealand found that approximately 72 per cent of those questioned had gained some comfort from knowing that others had benefited from their loss. Similar findings were reported in a United Kingdom survey, with 94 per cent of families who had donated feeling that they had made the ‘right decision'. A Dutch study supported the previous surveys and also noted that some families who had refused donation regretted their decision at a later stage. Such research conclusions are further supported by the positive feedback from donor families that is reported by the transplant co-ordinator teams.

Therefore, organ donation can give something positive in an otherwise negative situation. Offering the choice to donate, if performed with empathy, will not increase the distress of the bereaved. They should not be denied this choice or this chance of comfort. A letter from a donor mother reads: ‘it is certainly a source of comfort to me and indeed to all our family to know that our son has been able to touch and enrich the lives of others'.

When to offer the option of donation

In 1991 Garrison and colleagues reported that the timing of the approach may be the critical factor in the potential families' ability to give permission for organ donation. Also, this study from the United States suggests that several factors influence the consent process. First, the longer the patient is in hospital, the more time the family has to appreciate the fact that the patient is critically ill and will not survive. Therefore, it appears to follow that a family that has had more time to absorb and accept the prognosis is better able to move beyond the denial phase and become more receptive to options. Second, the timing of the approach for organ donation has significant consequences. This study reported that if the request for donation was made following notification of death, as opposed to before or simultaneously with the notification of death, the family were more likely to grant consent for donation and this trend appeared to hold true regardless of whoever made the request.

A more recent study in the United Kingdom, which identified reasons for relatives' refusal of donation over a 2-year period, also suggested that consent was more likely when the request was made following completion of the second set of brainstem death criteria when death had been confirmed. The study identified four other factors which may increase the likelihood of consent:

- the presence of a parent at the time of request
- the patient being aged 15 to 24 years or over 65 years
- a plan for cremation of the body
- situations where more than 1 h elapses between the two sets of brainstem death tests

Therefore, it would seem from recent research that it may be good practice to separate the notification of death from the request for organ donation.

Who should approach the family?

There is no one person who is ideal to approach the family because of the enormous variety of individuals and situations. It is most appropriate for the person who has formed a close and trusting relationship with the family to introduce the option of donation. It is essential that this person has a positive commitment to donation

themselves and introduces donation in a positive way.

A United Kingdom study reported that clinicians working in the ‘crisis' areas felt that a lack of training and a lack of experience in offering donation inhibited them in making the request. Conversely, a Canadian study showed that each experience of making the donation request built confidence, ‘practice makes perfect'. Every clinician who was experienced in talking to families about organ donation felt positively about the experience and believed that requesting donation was easier than seeking permission for an autopsy.

It is helpful to remember that the family is being asked to relate the wishes of their relative and whether objections to donation had been expressed, thus, freeing the family from accepting responsibility for the decision.

Many families may have discussed the idea of organ donation previously, perhaps at a time of national publicity; this knowledge of their loved ones' wishes will help them with their response.

It is widely reported that the bereaved strive to fulfil the wishes of their relative at the time of death and the presence of an organ donor card, registration on a donor registry, or a ‘living will' may help the family towards a positive response. Indeed, the bereaved may enquire about the possibility of donation before a formal approach is made.

How to approach the family

Staff are often reluctant to raise the question of donation because they fear that they may increase the family's distress by saying ‘the wrong thing'. However, there are no ‘right' words: each situation is unique and families will have their own individual responses. Requests for organ donation can never be preplanned, although anxiety can be reduced for the person making the request if suitable phrases are considered prior to meeting with the family.

Family: How could this happen? What a terrible waste of a young life.
Response: This is a terrible time for you but it need not be a complete waste; John's death could bring hope to others.
Family: He was a lovely man, he didn't deserve to die.
Response: He sounds like a lovely man, do you think his generosity would extend to helping others through his death?

Families will respond to the option of donation in a variety of ways: whatever the response, the carer should show empathy and understanding. Some families require time to consider their response and should be offered privacy. Many relatives will have additional questions concerning the process of donation and its implications.

Again, it will be helpful to use open-ended questions beginning with ‘how', ‘where', or ‘what' (i.e. what further information would you like) at this time. Such questions offer the bereaved the opportunity to make choices and to gain the information that is important to them.

At this time it may be helpful for them to meet with a member of the transplant team, usually the transplant co-ordinator, who can answer specific questions and start to develop a rapport with them. The family will require reassurance that their loved one will be treated with dignity and respect throughout the donor surgery; that the body will not be mutilated or grossly disfigured; that the surgical wound will be sutured; that they can view the body after surgery; and that the funeral will not be delayed. The transplant co-ordinator will work closely with other health care professionals to answer such questions and to facilitate the wishes of the family. It is often comforting for the family to know that the transplant co-ordinator will be present throughout the donor surgery and will perform the final care in accordance with their wishes.

It is recognized that there will always be families, regardless of the manner in which the request is offered, who will refuse the option of organ donation and health professionals must accept this decision. However, if the family appear undecided or if the immediate response is an angry ‘no', it is acceptable, after a short period of reflection, to explore gently the reasons for such a response. It is frequently found that the family may have specific concerns or unfounded ideas and fears that can be allayed by further information, thus, removing the barriers to permission.

Research suggests that the most commonly quoted reasons for refusal include: the deceased had stated that he/she did not wish to donate; a fear of gross mutilation; a difference of opinion between family members; problems understanding brainstem death; and ‘religious reasons'. However, it is important to note that all the major religions support the act of donation.

If the family agree to organ donation, many relatives will wish to spend time alone with their loved one so that they may say goodbye before the scheduled time of theatre. The opportunity to touch or kiss is especially appreciated: they should be offered privacy and never hurried.

Information following the donation will be provided to the family, unless they express otherwise. This feedback will contain general anonymous information about the recipients and offer further contact and support. Some transplant co-ordinating teams offer post-donation home visits so that ongoing support is activated and any subsequent anxieties or concerns can be addressed, and also in some areas donor family support groups are available.

Most centres will facilitate the exchange of letters between recipients and donor families believing that the bereaved gain comfort from the personal gratitude and obvious well being of the recipient and that recipients need to express their thanks in order to adapt psychologically and assimilate the new organ into their body and their new life. A few centres will help to arrange meetings between the donor family and the recipient; however, such meetings remain controversial.

Staff support

As stated in the introduction, death and bereavement are an integral part of human life and the care of those who grieve is an important part of clinical practice; however, dealing with the dying and their families is stressful for staff and if this stress is unresolved, the individual staff member may become depressed and ‘burnt out'

A supportive environment can reduce this stress; such an environment requires that staff care about each other, listening to each others' problems and offering support across all levels. Health care professionals will have individual coping strategies, but they should also have the opportunity to discuss issues of death and dying together, either formally or informally as requested. Clinicians who do not have this opportunity to replenish their own emotional reserves may find that they do not have anything left to give to future patients and their families.

Viewing the body after death

‘Seeing is believing', and all the families should be offered the opportunity to view the patient after death. If they are reluctant they should be gently encouraged, as it is an important step in accepting the reality of the situation. The body should be carefully prepared and the bereaved given privacy and permission to touch, hold, and kiss as desired. The loss of a young child is particularly distressing and parents may appreciate a lock of hair or a photograph or handprints.

Further care

Before the family return home it is important that they are aware of follow-up arrangements. In most cases this will involve an appointment with the bereavement officer, who will offer help and information concerning the tasks that lie ahead. In some cases it may also be appropriate to arrange a further meeting with medical staff so that additional questions may be answered.

Advice concerning expected grief reactions may also be helpful: relatives can be overwhelmed by the enormity and intensity of their distress. It is important that local support is available and the clinician should alert the family doctor or other support person to the needs of the bereaved. Some relatives may request medication, but in most cases the request should be gently denied as sedation dulls reality and response and inhibits the commencement of grief.

The majority of families recover from the death through the normal phases of grief. If a family member does experience specific problems, further help should be offered. Details of local bereavement organizations which can offer practical advice and experienced counselling should be made available.

Further reading

Blandy J. *Lecture notes in urology*. Blackwell Scientific Publications, Oxford, 1989.

Bowlby J. *Attachment and loss*, Vol. 1. *Attachment*. Penguin, Hardmondsworth, 1971.

Buckley PE. The delicate question of the donor family. *Transplantation Proceedings* 1989; **21**: 1411–12.

Buckman R. Breaking bad news; why is it still so difficult? *British Medical Journal* 1984; **238**: 1597–9.

Engel G. *Psychological development in health and disease*. WB Saunders, Philadelphia, 1962.

Garrison RN. There is an answer to the shortage of organ donors. *Surgery, Gynaecology, Obstetrics* 1991; **173**: 391–6.

Hampe SO. Needs of the grieving spouse in a hospital setting. *Nursing Research* 1975; **24**: 113–20.

Hodge JR. They that mourne. *Journal of Religion and Health* 1972; **11**: 229–40.

Kubler-Ross E. *On death and dying.* Tavistock Publications, London, 1987.

Lazare A. Unresolved grief. In: Lazare A, ed., *Outpatient psychiatry, diagnosis and treatment*, pp. 498–512. Williams & Wilkins, Baltimore, 1979.

Le Poidevin S. The management of bereaved relatives and approaching the next of kin about organ donation. Unpublished thesis (deceased), 1986.

Lewis CS. *A grief observed*. Faber & Faber, London, 1961.

Lindemann E. Symptomatology and management of acute grief. *American Journal of Psychiatry* 1944; **101**: 141–9.

McLaren J. Death of a child. In: Jacobs M, ed., *The care guide. A handbook for the caring professions and other agencies*. Cassell, London, 1995.

Molter NC. Needs of relatives of critically ill patients. A descriptive study. *Heart and Lung* 1979; **8**: 332–9.

Morten JB, Leonard DRA. Cadaver nephrectomy: an operation on the donor's family. *British Medical Journal* 1979; **1**: 239–41.

Parkes CM. *Bereavement: Studies of grief in adult life*. International Universities Press, New York, 1972.

Parkes CM. Determinants of outcome following bereavement. *Omega* 1975; **6**: 303–23.

Parkes CM. Bereavement counselling: Does it work? *British Medical Journal* 1980; **281**: 3–6.

Pincus L. *Death and the family*. Pantheon, New York, 1974.

Raphael B. The management of pathological grief. *Australia and New Zealand Journal of Psychiatry* 1975; **9**: 173–80.

Raphael B. Preventive intervention with the recently bereaved. *Archives in General Psychiatry* 1977; **34**: 1450–54.

Speck P. *Loss and grief in medicine*. Baillière Tindall, London, 1978.

United Kingdom Transplant Co-ordinators Association and British Association of Critical Care Nurses *Report of two year study into the reasons for relatives' refusal of organ donation*. Department of Health G77/030 3085, 1995.

Wakeford RE, Stepney R. Obstacles to organ donation. *British Journal of Surgery* 1989; **76**: 435–9.

Walter T. A new model of grief: bereavement and biography. *Mortality* 1996; **1**: 7–25

Worden WJ. *Grief counselling and grief therapy*. Routledge, London, 1988.

Wright R. *Caring in crisis*. Churchill Livingstone, London, 1986.

Yates DW *et al.* Care of the suddenly bereaved. *British Medical Journal* 1990; **301**: 29–31.

12.1 Respiratory problems

Rafael Barrera and Roger S. Wilson

Introduction
Effect of surgery and anesthesia on respiratory function
Ventilation
Lung volumes
Gas exchange
Host defenses
Clinical evaluation of the patient with pulmonary disease
Cigarette smoking
Obesity
Asthma
Chronic obstructive pulmonary disease
Optimization of pulmonary function
Chest physical therapy
Pharmacologic therapy
Non-invasive respiratory monitoring
Pulse oximetry
Capnometry
Anesthetic technique
Postoperative respiratory failure
Guidelines for management
Modes of mechanical ventilation
Discontinuation of mechanical ventilation
Further reading

Introduction

Pulmonary complications are a major source of morbidity in surgical patients, second only to cardiovascular events as a cause of peri- and postoperative death. Their overall incidence following all types of surgery is approximately 5 per cent. Several risk factors have been associated with a higher rate of pulmonary complications, including age, male sex, emergency surgery, preoperative comorbidity, the length of the surgical procedure, and the American Society of Anesthesiology (**ASA**) classification of physical status.

While these conditions contribute to an overall increase of operative complications, two additional factors predispose specifically to the development of pulmonary complications: pre-existing respiratory diseases and surgery involving the thorax or upper abdomen. Clinically important atelectasis, bronchospasm or exacerbation of chronic obstructive pulmonary disease, and infectious complications, may develop in 20 to 70 per cent of patients with pre-existing respiratory disease. Pulmonary complications, in the most severe cases, can result in acute respiratory failure and the need for prolonged mechanical ventilation.

Various possible risk factors associated with pulmonary complications after elective general surgery have been evaluated. Three have been identified: postoperative nasogastric intubation, preoperative sputum production, and increased duration of anesthesia over 2 h. Age, preoperative oxygenation, and spirometry did not aid in predicting patients at risk.

Patients undergoing elective abdominal surgery are more likely to develop pulmonary complications if there is evidence of lung and cardiac disease. However, spirometry alone did not predict the complications. Other factors that increase the risk for pulmonary complications in this group of patients are an abnormal chest radiograph and their ASA classification. The ASA classification is a valuable tool in predicting complications; it is a subjective, global rating of comorbid disease burden and assesses the clinical status of the patient.

Various risk factors are involved in pulmonary complications in the thoracic surgical patient. Lung cancer is usually a disease of elderly people; it appears that age alone is not a contraindication to thoracotomy, but it could be an additive risk factor in the development of pulmonary complications. Emergency surgery and comorbidities are well known risk factors for developing complications.

Death from pulmonary complications occurs in 7 per cent of surgical patients with moderate to severe chronic lung disease. Such complications occur in 33 per cent of patients with mild to moderate chronic lung disease who undergo upper abdominal surgery, but rates as low as 6 per cent and as high as 76 per cent have been reported. Pulmonary complications after a thoracotomy have decreased over the years: from 1950 to 1960 they were reported as approximately 40 per cent; in the 1990s they are reported as less than 15 per cent.

Special consideration must be given to patients who are at high risk for operative pulmonary complications and to those who are scheduled to undergo lung surgery. Evaluation must consider the effects of anesthesia and surgery on respiratory function in patients with minimal pulmonary reserve. The clinical status of those at risk of developing pulmonary complications should be optimized preoperatively in an attempt to minimize morbidity.

Effect of surgery and anesthesia on respiratory function

Pulmonary function is altered in several ways during and after surgery; important aspects include respiratory function, gas exchange, and defense mechanisms. Such changes occur to some degree in every surgical patient, but have a greater impact in those with pre-existing respiratory diseases who have a poor reserve.

Ventilation

Under normal circumstances, spontaneous minute ventilation increases in response to an elevated arterial CO_2 tension and to a reduction of the arterial O_2 tension. Most drugs used during anesthesia impair this normal ventilatory response to hypercapnia and hypoxemia. Opioids can produce profound respiratory depression: even small doses of morphine can blunt the ventilatory response to hypercapnia and hypoxia, and all the currently available opioid agonists seem to exert a similar effect at equipotent analgesic doses. The inhalational anesthetics halothane, enflurane, and isoflurane also depress respiratory drive; ventilatory response to hypoxemia is virtually abolished at inspired concentrations of halothane as low as 0.1 to 0.2 per cent, which are likely present during early recovery from general anesthesia (Fig. 1). The potent inhalational agents decrease spontaneous minute ventilation, producing a decrease in tidal volume (*V*/*T*) and resultant rapid, shallow breathing. This pattern increases the ratio of dead space to tidal volume (*V*d/**V**t), leading to CO_2 retention, and may facilitate the development of alveolar collapse and atelectasis.

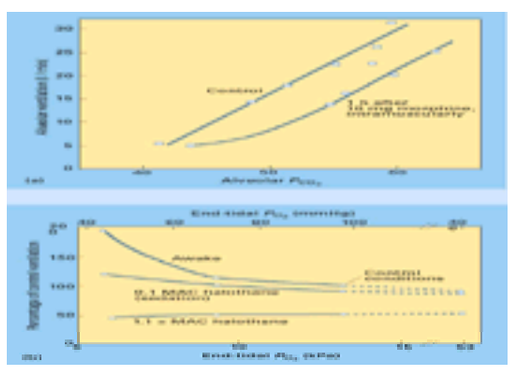


Fig. 1. (a) Effect of opioids on the ventilatory response to hypercapnia: alveolar *P*co$_2$ is plotted against minute ventilation before and 1 h after administration of morphine sulphate, 10 mg intramuscular. (b) Effect of halothane anesthesia on ventilatory response to hypoxia. MAC is the minimal alveolar concentration required for anesthesia.

Lung volumes

Changes in static lung volumes occur during surgery and anesthesia. The first to describe this was Beecher in 1933, who reported a decrease in vital capacity of approximately 45 per cent and in functional residual capacity of approximately 20 per cent after laparotomy; these changes persisted from 1 to 2 weeks postoperatively. Subsequent studies confirmed and extended this original observation. The following characteristics have been established. Functional residual capacity is reduced during general anesthesia by about 20 per cent below the value measured in the awake, supine position. This change occurs early in the course of the surgical procedure, appears not to be influenced by use of muscle relaxants, and is more pronounced in the elderly patient.

Upper abdominal incisions and thoracotomy are associated with the largest decrease in lung volumes postoperatively, followed by lower abdominal surgery. Peripheral procedures do not cause long-term changes in lung volumes. The cause of the reduction in lung volumes was elucidated in 1974 by Froese and Bryan. They described that the diaphragm ascended into the chest by about 2 cm during anesthesia with or without paralysis, and this change correlated with the decrease in functional residual capacity (Fig. 2).

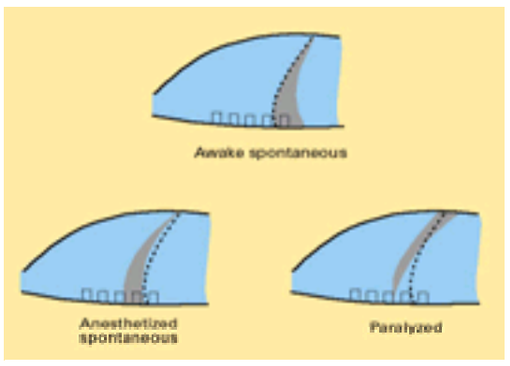


Fig. 2. Diagrammatic comparison of the position of the diaphragm in the awake and anesthetized patient. The broken line is the end-expiratory position of the diaphragm in the awake, supine state; the shaded area indicates the respiratory excursions of the diaphragm.

Breathing at low lung volumes affects gas exchange. A 20 per cent reduction in functional residual capacity added to an equivalent decrease secondary to the assumption of a supine position brings the functional residual capacity quite close to the residual volume. Sighs, which occur 10 times an hour, are decreased postoperatively secondary to the use of sedatives and narcotics. This circumstance may favor the development of atelectasis in the dependent zones of the lung following the induction of general anesthesia. The development of atelectasis increases dead-space ventilation and work of breathing, impairing adequate gas exchange. Atelectasis and small-airway closure impair gas exchange by producing a ventilation/perfusion mismatch and a right-to-left shunt.

Resistance to airflow also increases at lower lung volumes. Airway resistance is inversely proportional to the fourth power of the radius of the airway. At a reduced functional residual capacity, small decreases in volume result in a marked increase in resistance, which may cause hyperinflation and air trapping in patients with asthma or chronic obstructive pulmonary disease.

Gas exchange

Uncomplicated general anesthesia produces abnormalities in gas exchange that may become clinically significant in patients with pre-existing respiratory disease. Adequate gas exchange is dependent upon homogeneous matching of ventilation and perfusion at the alveolar level. Inspired gas delivered to lung regions that have no pulmonary capillary blood flow cannot take part in gas exchange and, conversely, pulmonary blood flow distributed to regions without ventilation cannot be oxygenated. The term ventilation/perfusion ratio (*V/Q*) is commonly used to refer to this relation. Considering the lung as a whole, typical resting values may be 4 l/min for alveolar ventilation and 5 l/min for pulmonary blood flow. Thus, the overall *V/Q* ratio would be 0.8, which happens to be close to the normal ratio between CO_2 production and O_2 consumption (respiratory quotient).

A simplified approach to *V/Q* physiology is shown in Fig. 3. The ventilated but unperfused lung section comprises the dead space, the perfused but unventilated section the shunt, and gas exchange is confined to the working alveoli. In clinical practice, arterial hypoxemia is frequently the result of low *V/Q* ratios rather than of a true shunt and is more correctly referred to as venous admixture. Usually, an increased dead space may be offset by increasing minute ventilation. An abnormal venous admixture up to about 30 per cent may be corrected with higher inspired oxygen concentrations.

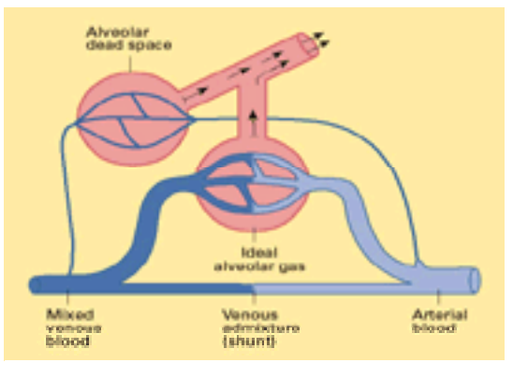


Fig. 3. Diagrammatic representation of the efficiency of gas exchange in the lung, considered as a three-compartment model. Gas exchange occurs only in the ‘ideal' alveoli. The alveolar dead space consists of alveolar units that are ventilated but not perfused. The venous admixture is the result of pulmonary blood flow through alveoli that are not ventilated. In the real situation, infinite gradations of ventilation and perfusion occur.

In the awake individual with normal lungs, the amount of inefficient ventilation causing venous admixture is minimal, resulting in an alveolar/arterial Po_2 gradient of about 10 mmHg. During uncomplicated general anesthesia that gradient may increase to 30 to 50 mmHg. The increase is secondary to the reduction in lung volumes and its effects on the development of atelectasis with alveolar and/or airway closure and resultant *V/Q* mismatch. Furthermore, inhalational anesthetics may impair gas exchange by inhibiting the physiologic mechanism of hypoxic pulmonary vasoconstriction that tends to divert pulmonary blood flow away from areas of low ventilation to regions of increased ventilation, preserving the *V/Q* relation and arterial oxygenation.

Elimination of CO_2 is also affected during general anesthesia, as a result of changes in the ratio *V*d/*V*t. Changes in *V*d/*V*t during general anesthesia are shown in Fig. 4. The anatomic dead space $(V\text{d}/V\text{t})_a$ (oral cavity, pharynx, and conducting airways) is about 30 per cent of the tidal volume in spontaneously breathing individuals. During general anesthesia, the $V\text{d}/V\text{t}_a$ (anatomic dead space) also includes that segment of the anesthetic circuit in which the gas flow is bidirectional, the endotracheal tube or face-mask, and tubing distal to the Y-piece connector. When the trachea is intubated, the $V\text{d}/V\text{t}_a$ is slightly decreased because the upper airway is bypassed. When anesthesia is administered by face-mask, on the other hand, the volume of the mask adds further dead space, and the $V\text{d}/V\text{t}_a$ increases to approximately 40 per cent.

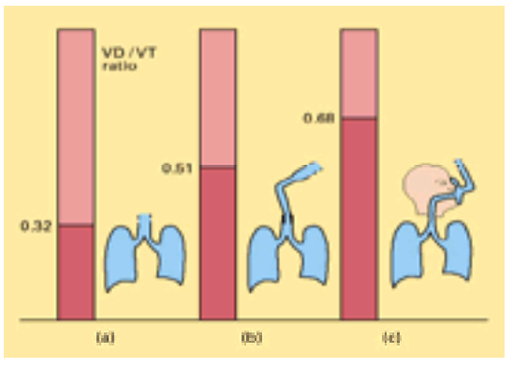


Fig. 4. Changes in dead space as a fraction of *V*/*T* during general anesthesia: (a) from the carina downwards; (b) including the endotracheal tube and connectors; and (c) including the face-mask and connectors.

The term physiologic dead space, $(Vd/Vt)_{phys}$, defines areas of the lung parenchyma beyond the conducting airways where gas exchange does not occur. In the awake individual with normal lungs, no significant $(Vd/Vt)_{phys}$ is detectable. During general anesthesia, the $(Vd/Vt)_{phys}$ increases to approximately 30 per cent of the tidal volume. Thus, a seemingly adequate minute ventilation of 5 to 6 l may result in an alveolar ventilation as low as 2 to 3 l/min, which could cause CO_2 retention.

With the aid of the multiple inert-gas washout technique, the effects of anesthesia on gas exchange can be summarized as follows.

1. The alveolar/arterial Po_2 gradient is increased during anesthesia, and this change is markedly affected by age.
2. The decrease in Po_2 is secondary to an increased distribution of flow to areas of decreased ventilation, most commonly the dependent areas.
3. The increase in *V*d/*V*t seems to be secondary to increased distribution of ventilation to areas of lesser perfusion.
4. The major differences are between the awake and anesthetized state; paralysis and controlled ventilation do not greatly alter overall gas exchange.

Host defenses

Inspired air is normally warmed and humidified in the nasopharynx; this facilitates the clearance of airway secretions by optimizing ciliary function. Pulmonary defense mechanisms are altered during the peri- and postoperative period. Bacterial adherence to upper airway epithelium can be altered after instrumentation. Normal function of the immune system in humans is altered in the immediate postoperative period. Unfortunately, clinical investigation of this change is confronted with the difficulty of separating the effects of the many intraoperative factors that may impact on immunity. It appears that many of the immune changes seen in surgical patients are primarily the result of the surgical trauma and of endocrine responses, rather than of the anesthetic exposure itself. *In vitro* studies suggest that anesthetic agents may have a direct effect on granulocyte and monocyte function and on the release of immunologic mediators. A synopsis of the possible effects of anesthetic agents on different components of the immune response is shown in Fig. 5.

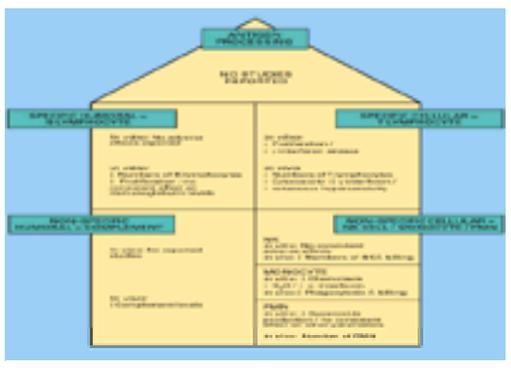


Fig. 5. Effects of anesthetic agents (primarily halothane) on the function of immune components. NKC, natural killer cell; PMNC, polymorphonuclear cells.

Anesthetic gases are essentially dry as they leave the standard anesthesia machine; dryness tends to damage the respiratory epithelium. Endotracheal intubation exacerbates this problem by bypassing the upper airway. The cough mechanism and mucociliary clearance are depressed during anesthesia. Clearance of secretions from the airway is decreased when coughing is blunted. The combination of these factors predisposes to inflammation of the respiratory mucosa and to retained secretions, which may favor alveolar collapse, bronchospasm, and infection. Sedatives, analgesics, and anesthetics may predispose postoperative patients to aspiration of gastric or oral contents, increasing the chances of infection. Prompt diagnosis is crucial if one is to prevent secondary bacterial infections and complications such as pneumonitis, lung abscess or bronchiectasis, which can be catastrophic in patients with pulmonary dysfunction.

Clinical evaluation of the patient with pulmonary disease

Identification of patients at increased risk for operative respiratory complications is important since optimization of their clinical status can decrease morbidity. During clinical evaluation, one should give emphasis to the current respiratory status of the patient, to coexisting cardiovascular disease, to a prior history of treatment for pulmonary disease, to occupational exposures, and to the use of medications. Various ways are available that can help the physician identify high-risk patients, such as pulmonary function tests and arterial blood gases, ventilation perfusion scans, and exercise testing. Pulmonary function tests have been advocated for many years as a preoperative tool to predict outcome; their ability to predict postoperative pulmonary complications has been variable and is evolving. Spirometry is usually recommended before any pulmonary resection. Several guidelines have been proposed to decrease risk of complications. A patient should tolerate a pneumonectomy if the forced expiratory volume in 1 s ($\mathbf{FEV_1}$) is more than 2 l, if the maximal voluntary ventilation is more than 50 per cent of that predicted, and the ratio of residual volume to total lung capacity is less than 50 per cent. In some patients with abnormal spirometric values the diffusion capacity can help the physician predict pulmonary complications. The predicted postoperative diffusion capacity is inversely related to the incidence of pulmonary complications and an important predictor of morbidity.

Spirometry has not been clearly established as a definite test for determining who is a candidate for surgery, but most agree that the detection of severe abnormalities by spirometry should prompt further evaluation. Arterial blood gas analysis is another tool that is used to assess the patient's ability to tolerate a procedure. Hypercapnia, a Pco_2 of more than 45 mmHg, is predictive of an increased risk of respiratory complications and at times is considered a contraindication to a thoracic surgical intervention. Exercise testing is a useful complement to conventional cardiopulmonary evaluation in selecting patients for pulmonary resection who have poor results in pulmonary function tests.

Among the risk factors possibly associated with an increased rate of operative pulmonary complications, the following have been reported consistently.

Cigarette smoking

This is associated with an increased perioperative mortality because of its effects on both the cardiovascular and respiratory systems. A history of tobacco use is a sensitive predictor of lung disease and of postoperative respiratory complications: compared with non-smokers, cigarette smokers have a two- to threefold greater incidence of perioperative pulmonary complications. Smoking also increases the concentration of carboxyhemoglobin in the blood (3–15 per cent), thereby reducing the amount of hemoglobin available for oxygen transport. Cessation of smoking 12 to 18 h preoperatively significantly lowers carboxyhemoglobin and may decrease heart rate, arterial blood pressure, and the concentrations of catecholamines. Ciliary and small-airway function improves slowly over weeks after smoking is discontinued. The incidence of postoperative pulmonary complications remains high in patients that have recently stopped smoking; those who had stopped smoking for over 6 months had the same rates of pulmonary complications as those who had never smoked. Any impact on the incidence of operative pulmonary complications, however, is not detectable unless smoking is discontinued for more than 8 weeks before surgery. This conclusion is in agreement with the observed improvement in respiratory

symptoms and lung volumes that occur over a period of months following the cessation of smoking.

Obesity

Obesity is a serious disorder resulting in significant impairment of health. The body mass index, which is the ratio of weight (kg) to height (m) squared, is the most convenient method of quantifying the degree of obesity. The National Center for Health Statistics has defined overweight as a body mass index over 27.8 in men and 27.3 in women. As obesity is such a pervasive disorder and is an important risk factor for many diseases, it is not surprising that many obese patients have pulmonary complications after surgery. The effect of obesity on lung function is complex: it depends on the degree of obesity, age, and type of body fat distribution. Obesity reduces total lung volume and functional residual capacity, thus predisposing to hypoxemia and atelectasis. The expiratory reserve volume is consistently decreased and the FEV_1:forced vital capacity ratio is increased. The work of breathing in the obese patient is increased due to abnormal chest elasticity, increased resistance of the chest wall and airways, and abnormal diaphragmatic position. Upper airway resistance and increased daily elimination of carbon dioxide are other factors in the obese patient contributing to postoperative pulmonary complications.

Few data are available about the impact of obesity in surgical patients, especially in the critically ill. The risk of aspiration pneumonia as well as pulmonary complications is greatly increased in the postoperative period in the obese patient. Some studies that have employed a precise definition of obesity suggest that postoperative hypoxemia and atelectasis are common in obese patients, but that the risk of severe respiratory morbidity is uncertain. Obese patients that sustain blunt trauma to the abdomen have an eightfold higher mortality rate than non-obese patients. Lower pre- and postoperative arterial oxygen saturation than in non-obese controls was found in patients undergoing jejunoileal bypass for morbid obesity. Obesity was the most important factor for pulmonary embolism.

Pain control with minimal respiratory depression, early ambulation, and close respiratory monitoring are recommended in the obese patient postoperatively.

Asthma

A major concern when managing patients with bronchial asthma is the potential for exacerbation of bronchospasm secondary to stimulation of the airways. Severe bronchospasm may occur during endotracheal intubation, often as a result of light anesthesia, bronchial secretions, and surgical manipulation of the viscera or the airway. Early studies suggested a greater risk of perioperative respiratory complications in patients with bronchial asthma than in non-asthmatics. However, even steroid-dependent asthmatic patients were shown to tolerate major surgery if proper perioperative care was given. Anesthetic considerations in asthmatic patients include the use of volatile agents, which induce a direct bronchodilator effect independent of adrenergic stimulation. Ketamine exhibits bronchodilator properties. Opioids effectively blunt airway reflexes; synthetic opioids, such as fentanyl, do not induce histamine release and may decrease airway irritability during intubation and extubation of the trachea. When possible, regional anesthesia or general anesthesia via a face-mask may be a valuable alternative to endotracheal intubation.

Chronic obstructive pulmonary disease

Patients with chronic obstructive pulmonary disease, as opposed to those with restrictive disease in whom expiratory flow rate and the cough mechanism are preserved, have been identified as a group at increased risk for operative respiratory complications. Many important measures of pulmonary function are associated with an increased incidence of postoperative pulmonary complications. Although it is clear that patients with chronic obstructive pulmonary disease are at higher risk than the normal population, no single test or combination of tests has emerged as a predictor of morbidity. It is well established that a low predicted postoperative FEV_1 appears to correlate with a high risk of developing postoperative pulmonary complications. Abnormal pulmonary function tests seem to be of value as indicators of surgical risk in patients with chronic obstructive pulmonary disease.

Preoperative considerations for pulmonary resection

Virtually all patients with bronchogenic carcinoma of the lung have chronic obstructive pulmonary disease. While surgical therapy offers the best prospect of long-term survival, removal of lung tissue may reduce postoperative function, predisposing to long-term mechanical ventilatory support and possible death. When evaluating patients for lung surgery one should be able to predict postoperative function. The most frequently used method to predict lung function following lung resection is a quantitative ventilation perfusion scan, which can predict postoperative FEV_1(**ppoFEV**$_1$). The selective function of the segment to be resected can be evaluated, and the $ppoFEV_1$ calculated. Historically, patients with a $ppoFEV_1$ of 800 ml or less for thoracic surgery were generally considered to have a poor prognosis, even though this figure was not derived from clinical studies but rather from physiologic principles and observations. Physicians are more inclined to look at various tests to evaluate the postoperative risk of any patient. Postresection disability is related not only to limitations in ventilation but also to alterations in pulmonary blood flow and pressure; thus, criteria that stratify patients according to their ability to exercise have been studied. Measurements of maximum oxygen consumption, pulmonary arterial pressure, and calculated pulmonary vascular resistance during bicycle ergometry have been advocated as possible criteria that could identify operative candidates who would otherwise not be considered for thoracotomy on the basis of pulmonary function tests. There is a vast amount of evidence that no single pulmonary test can accurately predict pulmonary complications in surgical patients. Various investigators are using composite indices to predict complications in surgical thoracic patients. Pierce was the first to demonstrate that the algebraic product of $ppoFEV_1$ per cent × $ppoDL_{CO}$ per cent (less than 1850) had prognostic value in predicting pulmonary complications and mortality. Another index, the predictive respiratory complication quotient, which is the $ppoFEV_1$ per cent × ($ppoDL_{CO}$ per cent)^2 /Pao_2 - Pao_2 gradient (less than 220) can predict pulmonary complications and mortality. The predictive value of exercise testing needs confirmation in larger studies. A careful preoperative assessment of patients with lung cancer is essential for identifying those at high risk secondary to pulmonary dysfunction.

Optimization of pulmonary function

The efficacy of preoperative prophylactic measures in decreasing pulmonary complications has been clearly substantiated. Patients with abnormal pulmonary function tests who underwent a preoperative regimen of bronchodilators and physiotherapy had a postoperative rate of pulmonary complications of 21 per cent, as compared to a 60 per cent incidence in patients with similar pulmonary function tests who did not receive preventative treatment.

Preoperative evaluation would be of limited value unless prophylactic measures were instituted in those patients identified as at increased risk. High-risk categories of patients include those with chronic obstructive pulmonary disease as indicated by abnormal pulmonary function tests (particularly when scheduled for thoracic or upper abdominal surgery), smokers, and patients with acute respiratory infections or active bronchospasm. General measures that may be beneficial include the cessation of smoking, hydration, humidification of inspired gases, antibiotic treatment of infections, and treatment of associated cardiovascular and metabolic disease. More specifically, respiratory function may be improved over a short period of time before surgery by means of a program of chest physical therapy and pharmacologic treatment.

Chest physical therapy

This includes breathing and coughing exercises, postural drainage, ambulation, and the use of mechanical aids to lung expansion. Physiotherapeutic treatment should be planned and carried out by specialized personnel in order to ensure maximal compliance. Preoperative education is likely to be more effective than attempting to teach principles of physiotherapy in the presence of postoperative pain and medications.

It appears that respiratory rehabilitation, physical therapy, and mechanical devices may exert a beneficial effect on the postoperative course of patients who have had chest and abdominal surgery. Controversy remains over the selection of patients for this type of therapy, but high-risk patients with little pulmonary reserve definitely benefit from preoperative pulmonary rehabilitation.

Several kinds of mechanical aid to physiotherapy are available. Intermittent positive-pressure breathing, not now generally used, is not always well tolerated and may transiently decrease pulmonary compliance. It is frequently associated with unpleasant side-effects such as bloating. Forced expiratory exercise by means of blow bottles and similar devices is no longer recommended because of the potential for hyperinflation and air trapping. Currently, maneuvers designed to increase the functional residual capacity, such as incentive spirometry and continuous positive airway pressure by face-mask, are used more frequently. Preoperative intermittent positive-pressure breathing, incentive spirometry, and deep-breathing exercises seems to be equally effective in reducing the incidence of pulmonary complications following upper abdominal surgery when compared with no respiratory treatment.

Pharmacologic therapy

Table 1 summarizes the characteristics of the most common agents used preoperatively in patients with pulmonary disease. Fluidification of bronchial secretions and the resolution of bronchospasm are the main goals. Bronchodilator therapy should be pursued aggressively, even among those patients with chronic obstructive pulmonary disease who do not seem to respond to a single administration during pulmonary function testing. There will be some who benefit from repeated combinations of various pharmacologic treatments over the course of a few days or even a week. Endpoints of preoperative pharmacologic treatment should be the

resolution of acute symptoms such as bronchospasm and dyspnea, and an improvement in the patient's level of activity or exercise tolerance, as well as in their ability to expectorate and perform physiotherapeutic maneuvers. Repeating a set of pulmonary function tests may offer an objective documentation of the effects of treatment.

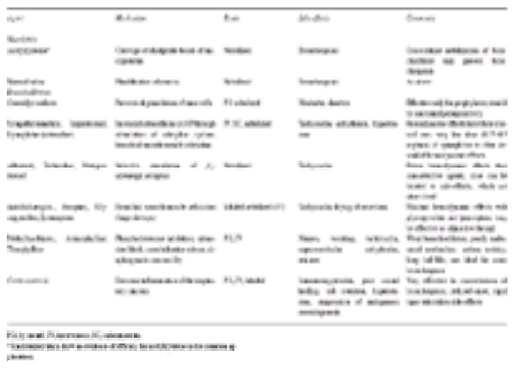


Table 1 Pharmacologic agents in the preoperative treatment of patients at risk for pulmonary complications

Non-invasive respiratory monitoring

Non-invasive monitoring of respiratory function is routinely available in the operating theatre and in the intensive care unit. The extent of its use and the selection of techniques are determined by individual preference, patient status, and proposed operative procedure.

Pulse oximetry

Pulse oximetry provides continuous measurement of arterial oxygen saturation by spectrophotometry. Oxygenated and reduced hemoglobin have different spectra of light absorption: the relative concentration of oxygenated hemoglobin (expressed as percentage saturation) in arterial blood is derived by a microprocessor from the ratio of absorption of two different wavelengths specific for the two forms of hemoglobin. The sensing unit of the pulse oximeter can be applied to areas of the body that a light beam can be shone through and recorded by an emitting and a receiving diode. Adequate arterial blood flow is necessary in order to provide an adequate signal: motion, vasoconstriction, hypothermia, and very low blood pressure interfere with signal detection. Preoperative evaluation by pulse oximetry and exercise desaturation is not recommended routinely, owing to its lack of predictive value. The intraoperative use of continuous oximetry identifies episodes of arterial oxygen desaturation; these can be treated effectively. Postoperatively, pulse oximetry is widely employed during immediate recovery from general anesthesia and in the intensive care unit. Continuous monitoring of arterial oxygen saturation in these settings may reveal episodes of hypoxemia that could otherwise go undetected. During weaning from mechanical ventilation, the recording of arterial oxygen saturation allows for changes in ventilator settings with less frequent arterial blood sampling.

Capnometry

This provides measurement of end-tidal CO_2 (Petco$_2$) and display of exhaled CO_2 waveforms (capnography). The most commonly used capnometers continuously withdraw a small (150 ml/min) sample of gas distal to the Y-piece connector of the breathing circuit. Modifications to face-masks, nasal airways, and nasal cannulas have been designed to facilitate capnography in the awake patient, enabling the capnograph to serve as a monitor for apnea. The Petco$_2$ reflects the Pco$_2$ in the alveolar gas, which in normal circumstances is slightly lower (2–4 mmHg) than the arterial Pco$_2$. A higher arterial Petco$_2$ gradient reflects an increase in Vd/Vt. A sudden fall in the end-tidal Pco$_2$ may be due to an acute decrease in cardiac output, to pulmonary embolism, or to air embolism. Useful applications of capnometry in the operating room include detecting accidental esophageal intubation by the absence of a CO_2 waveform, of inadequate ventilation, of disconnection of a component of the breathing system, and of CO_2 rebreathing from an exhausted CO_2 absorber or a malfunctioning valve. Capnometry may allow early detection of air embolism during sitting craniotomy, spinal fusion or hip surgery by a sudden decrease of Petco$_2$ secondary to a decreased cardiac output or to sampling of gas lower in CO_2, which diffuses into the alveoli from the pulmonary capillaries. Capnography may be useful in the intensive care unit in mechanically ventilated patients, where the adequacy of ventilation in response to physiologic or mechanical changes may be assessed immediately.

Anesthetic technique

There is no evidence that the anesthetic technique affects the outcome of patients with chronic respiratory disease. Advantages of a regional anesthesia (spinal, epidural, extremity blocks) over general anesthesia are the avoidance of the reduction in respiratory volume and consequent alterations in gas exchange, as well as the potential benefit of not intubating the trachea. On the other hand, a regional anesthetic may weaken the function of the respiratory muscles and blunt the proprioceptive reflexes from the diaphragm, causing hypoventilation; oversedation of an awake patient may also lead to hypoventilation, particularly in elderly individuals. Few studies have looked at outcome in patients with respiratory disease with the use of regional as compared to general anesthesia and their results are contradictory: although two large studies report better outcomes following spinal or epidural anesthesia, this finding was not reproduced elsewhere when similar groups of patients were studied. Not surprisingly, a universally applicable answer is not available. When suitable, extremity blocks (brachial, lumbosacral, ankle) are intuitively a safe alternative, since they do not interfere with respiratory mechanics. Spinal or epidural anesthesia may be very effective in a cooperative patient for procedures on the lower extremities or lower abdomen, while they may turn out an unwise choice in elderly patients with dyspnea at rest scheduled for long and traumatizing surgery. Most instances are not so clear cut, and the decision must ultimately be made by an experienced anesthesiologist. The assumption often made, that a patient with severe chronic obstructive pulmonary disease needs spinal anesthesia because he or she may never be extubated is inappropriate.

Another approach now commonly employed for thoracic and upper abdominal surgery utilizes epidural analgesia in combination with light general endotracheal anesthesia. A mixture of a narcotic analgesic and local anesthetic (for example, fentanyl 10 µg/ml in bupivacaine 0.075–0.125 per cent at 3–6 ml/h) is infused into the epidural space via a catheter inserted into the thoracic or lumbar spine, while general anesthesia is maintained with nitrous oxide in oxygen, if necessary supplemented with a low concentration of a volatile agent. The epidural infusion can be continued beyond the operation to provide analgesia. The adjunct of epidural analgesia greatly reduces the amount of general anesthetic, facilitates extubation, and provides excellent pain relief postoperatively without significant ventilatory depression. The possible impact of this combined technique on the development of postoperative pulmonary complications has been assessed in several studies. While a few suggest a benefit when compared to general anesthesia alone, others found no difference; none, however, found higher rates of complications in the epidural groups. These studies are difficult to compare because of their differing definitions of pulmonary complications, and the lack of uniformity of anesthetic techniques used. However, the satisfaction of patients was uniformly superior when epidural analgesia was continued postoperatively, compared with intermittent intramuscular injections of opioids.

Postoperative respiratory failure

Many of the mechanisms that affect respiratory function during anesthesia and surgery persist for a variable time postoperatively. In most instances, pulmonary complications may be viewed as a continuation of physiologic derangements initiated on the operating table. Characteristically, postoperative patients breathe rapidly, with shallow V/T; vital capacity is reduced for several days postoperatively. Prolonged immobilization, tight bandages, splinting from incisional pain, and diaphragmatic dysfunction all contribute to the persistence of low lung volumes.

The most important changes occur after thoracic and upper abdominal surgery. However, any type of surgery, when associated with immobilization, malnutrition, advanced age, and a decreased respiratory reserve, may create grounds for pulmonary complications. Patients with long-standing respiratory failure are particularly vulnerable in the perioperative period: pulmonary or systemic complications may easily result in acute respiratory failure requiring mechanical ventilation.

Causes of an inability to sustain spontaneous ventilation at the end of surgery include excessive administration of volatile anesthetics, narcotic analgesics, or muscle relaxants; prolonged and traumatizing procedures with large fluid requirements; and intraoperative complications such as lung collapse, pneumothorax, and pulmonary edema. In these circumstances, treatment of the intercurrent complication is generally sufficient to allow rapid recovery of respiratory function. Inability to sustain adequate spontaneous ventilation later in the postoperative course is most frequent in patients with severe chronic obstructive pulmonary disease. Frequent causes of acute respiratory failure at this time are pneumonia, tracheobronchitis, aspiration pneumonitis, and cardiac failure. Pneumonia impairs gas exchange and decreases lung compliance; excessive bronchial secretions increase airway resistance and worsen bronchospasm. Patients with limited respiratory reserve tolerate poorly the consequent increased work of breathing and will require prolonged mechanical ventilation.

Guidelines for management

Hypoxemia

Hypoxemia may be treated initially with enriched inspired oxygen concentrations provided by a non-rebreathing face-mask, high-flow oxygen-delivery systems, or by continuous positive airway pressure delivered by mask. Adjustable positive-pressure ventilation triggered by the patient's inspiratory effort can also be delivered mechanically without the use of an endotracheal tube. These systems may be effective in the short-term management of acute respiratory failure secondary to exacerbations of chronic pulmonary disease or to congestive heart failure, but no data are available in relation to surgical patients. The potential for gastric distension from positive airway pressure without endotracheal intubation suggests the need for cautious use of these techniques in patients recovering from gastrointestinal surgery.

Endotracheal intubation

Ultimately, the surgical patient with acute respiratory failure may require endotracheal intubation and mechanical ventilation to avoid exhaustion, acidosis, and hypoxemia. Intubation should be done by an experienced physician. Oral intubation under direct laryngoscopy is generally the easier route, and the one of choice in unstable patients when rapid re-establishment of the airway is essential. Nasal intubation is better tolerated during prolonged mechanical ventilation; the endotracheal tube is more effectively secured and accidental extubation less likely. Potential drawbacks include the possibility of mucosal damage and bleeding, which can make subsequent laryngoscopy very difficult, the limited internal diameter and length of the endotracheal tube, and the possibility of subsequent sinusitis. Nasotracheal intubation should be avoided in patients with coagulopathies, and in those with major fractures of the facial bones or of the base of the skull. Table 2 summarizes issues concerning the oral and nasal approaches to intubation.

	Nasal intubation	*Oral intubation*
Advantages	May be carried out without laryngoscopy when visualization or neck motion are limited Easily secured in place Improved patient comfort Easier oral access and care	More easily carried out by inexperienced personnel Large-bore and shorter-length tubes can be utilized
Disadvantages	Choanal size limits diameter of the tube Occasional difficulty in suctioning Relative contraindications: Risk of epistaxis in patients with coagulopathy Obstruction of paranasal sinuses; long-term use may be associated with bacteremia secondary to sinusitis	Easily dislodged May be occluded if precautions are not taken to prevent the patient from biting on the tube May be poorly tolerated by some patients

Table 2 Nasal versus oral intubation

Nutrition

Patients with acute or chronic respiratory failure are often chronically ill and malnourished, and the added stress of surgery and infection further depletes their metabolic reserves. When tolerated, enteral nutrition may be used instead of total parenteral nutrition. This route seems to be more physiologic, preserves the function of the intestinal mucosal barrier, and avoids the possible complications of parenteral nutrition, including those from venous access.

Physical therapy

As outlined earlier, a program of physical therapy, including postural drainage, deep-breathing exercises, and early mobilization, should be undertaken by specialized personnel. Getting the patient out of bed is a most effective maneuver for lung expansion that may generate a 10 to 20 per cent increase in functional residual capacity. Ambulation with the help of a nurse and a respiratory therapist is possible in cooperative intubated patients.

Analgesia and sedation

Adequate analgesia is essential postoperatively since pain can result in tachycardia and hypertension, and limit respiratory and general activity. Intermittent intramuscular injections do not produce reliable pain control. In patients in the intensive care unit, where respiratory function can be closely monitored, opioids should be given intravenously.

In recent years, neuroaxial administration of opioids has become common practice in many surgical intensive care units. Preservative-free morphine injected intrathecally (1 mg or less) or epidurally (2–4 mg) can give excellent analgesia for 24 to 36 h. The epidural route is more frequently used, since it carries less risk of postdural puncture headache and infection, and thus is safer when prolonged drug delivery is desirable. Continuous epidural infusion through an indwelling catheter allows the use of short-acting, highly lipid-soluble opioids such as fentanyl, thereby limiting the potential for ventilatory depression. Since opioids and local anesthetics block different nociceptors in the spinal cord, mixtures of the two can be used to improve analgesia and minimize side-effects.

With patient-controlled analgesia, standard drugs such as morphine or meperidine are given from bedside devices operated directly by the patient. These devices typically employ infusion pumps with safety limits on infusion rates. Patient-controlled analgesia provides more stable blood concentrations of opioids when compared to intermittent administration as required, and consequently better pain control with smaller amounts of the drug overall.

Regional nerve blocks may also be considered: multiple intercostal nerves blockaded with a local anesthetic [for example, 0.5 per cent bupivacaine with epinephrine (adrenaline) 1:200 000] provide 4 to 8 h of analgesia following thoracotomy and subcostal and flank incisions. Non-steroidal anti-inflammatory agents may be an effective alternative or adjunct to opioids.

Guidelines for postoperative analgesia for the mechanically ventilated patient are summarized in Table 3.

	Agent and dose	Comments
Continuous IV infusion of opioids	Morphine sulphate: 5 mg/h, up to 20–30 mg/h. Fentanyl, 50–100 mg/h	Little or no hemodynamic effects (mild bradycardia); respiratory depression; tachyphylaxis; mild sedative effect; no amnesic effect
Patient-controlled analgesia	E.g. morphine sulphate, 1–2 mg bolus at lockout intervals of 10 min	Little or no hemodynamic effects (mild bradycardia); respiratory depression; tachyphylaxis; mild sedative effect; no amnesic effect
Continuous epidural infusion	Preservative-free morphine: 0.25–1 mg/h. Fentanyl: 3–10 μg/ml in 0.1% bupivacaine, at 5–10 ml/h	Special training for nursing personnel necessary; delayed resp. to 24 h; respiratory depression with morphine
Intercostal nerve blockade	0.5% bupivacaine with epinephrine (adrenaline) 1:200 000, 4–5 ml per level	Repeated every 4–8 h; blocks required; potential for pneumothorax, intravascular injection, infection
Nonsteroidal anti-inflammatory	Indomethacin: 25–50 mg PO or PR	Often effective for pain from thoracotomy and chest tubes
	Ketorolac: 15–60 mg IM	Analgesic effect possibly as potent as intramuscular morphine; potential for coagulopathy

IV, intravenous; PO, by mouth; PR, by way of the rectum; IM, intramuscular.

Table 3 Guidelines for postoperative analgesia in adult intubated patients

Sedation and adequate sleep is also an important aspect of the care of these patients; the practice of withdrawing or minimizing sedation in intubated patients to facilitate weaning is, in many cases, inappropriate. Benzodiazepines and neuroleptic agents are frequently used for sedation. In patients requiring large doses of sedatives because of excessive agitation, a continuous infusion may be appropriate. Midazolam (a short-acting benzodiazepine) or propofol (an ultrashort-acting potent

hypnotic) administered in continuous infusion can provide satisfactory sedation without untoward hemodynamic and respiratory effects; recovery after discontinuation of the infusion is almost immediate in the majority of patients.

Respiratory measures

The extent of acute lung disease is a major determinant of the patient's ability to sustain spontaneous ventilation: extubation will not be possible while major parenchymal damage from pneumonia or trauma persists, causing hypoxemia, increased dead space, and impaired lung compliance. Appropriate antibiotic therapy must be guided by serial cultures. Fluid balance should be evaluated daily, and diuretic therapy administered when clinical evidence of fluid retention develops. Bronchospasm demands aggressive treatment, as outlined earlier: factors capable of precipitating bronchospasm include pulmonary edema, direct stimulation of the airways, pain, agitation, and (although probably not common) b-blockade. Chest physical therapy should be accompanied by adequate analgesia. Excessive amounts of bronchial secretions may hinder extubation; the treatment of tracheobronchitis with antibiotics, inhalation of mucolytics, and chest physical therapy enhances the patient's ability to clear secretions.

Breathing through a long, narrow-diameter tube increases the patient's work and may limit weaning. A major increase in resistance occurs in adults if endotracheal tubes below size 7.0 (internal diameter in millimeters) are used in conjunction with an increase in minute ventilation. Above this size, resistance does not increase appreciably, and the discomfort and risk of changing the tube is not justified in most patients.

The best timing of elective tracheostomy is not well established. Airway damage by the endotracheal tube has become less common since the introduction of high-volume, low-pressure cuffs. It is generally accepted that patients can be safely ventilated through an oro- or nasotracheal tube for 7 to 14 days. Elective tracheostomy has a low incidence of complications and is often unexpectedly welcomed by the patient, who appreciates the lower resistance to breathing and the better comfort as compared with an oro- or nasotracheal tube. Modifications of the tracheostomy tube, for example fenestration, allow the patient to talk when not mechanically ventilated.

Modes of mechanical ventilation

Modern ventilators for use with patients in intensive care units are capable of delivering different modes of ventilation, allowing greater versatility in choice for the individual patient. Mechanical ventilators have been transformed from machines delivering oxygen to a useful tool for helping the patient with the work of breathing and ventilation. Regardless of the mode selected, the first decisions to be made when starting a patient on ventilatory support are about *V*/*T*, respiratory rate, and inspired oxygen concentration. Tidal volumes 50 to 80 per cent larger than the spontaneous are often necessary to compensate for the circuit compressible volume and the increased physiologic dead space, $V\mathrm{d}/V\mathrm{t}_{\mathrm{phys}}$. Since ventilator circuits include humidifiers and long, large-bore corrugated tubing, a considerable percentage of the tidal volume is compressed in the system and never reaches the patient. Most systems have compressible volume loss factors of 3 to 5 ml/cmH_2O airway pressure. Thus, with a peak airway pressure of 50 cmH_2O, 150 to 250 ml of volume is lost. $(V\mathrm{d}/\mathrm{Vt})_{\mathrm{phys}}$ may increase because of dilatation of the airways and decreased cardiac output during positive-pressure ventilation, and because of the lung disease itself.

Slow respiratory rates and large tidal volumes are generally employed in patients with chronic obstructive pulmonary disease, in whom a short expiratory time impairs CO_2 elimination and favors air trapping. Rapid rates may be used safely in patients with restrictive disease.

The duration of inspiration may be controlled in three different ways, depending upon the design of the ventilator. In volume-cycled modes, such as continuous mandatory ventilation, intermittent mandatory ventilation, and assist/controlled ventilation, inspiration stops when the preset volume is delivered. In the absence of a leak, this method guarantees the tidal volume, independent of changes in compliance and resistance. In pressure-cycled modes, such as pressure-limited ventilation, inspiration stops when a preset mouth pressure is reached. In this mode, pressure but not volume has to be set to determine *V*/*T*, which, consequently, is not guaranteed and is affected by changes in respiratory compliance and resistance. In time-cycled ventilators, inspiration stops after a preset time, and tidal volume depends on the inspiratory flow rate; these ventilators are still used in the operating room, but are not practical for prolonged ventilation in the intensive care unit.

Inspired oxygen concentration ($F\mathrm{io}_2$) is generally regulated to the lowest possibility that will maintain adequate arterial oxygen tension (60 to 80 mmHg) in order to avoid oxygen toxicity. The safe $F\mathrm{io}_2$ in humans is not known: oxygen toxicity is directly related to $F\mathrm{io}_2$ and to the length of exposure. Inspired concentrations of O_2 below 50 per cent should be safe even for prolonged periods. Arterial oxygenation may be improved, and $F\mathrm{io}_2$ decreased, by applying positive end-expiratory pressure.

Positive end-expiratory pressure improves oxygenation in patients with acute respiratory failure by increasing functional residual capacity, by redistributing edema fluid to the interstitial compartment, and possibly by decreasing cardiac output and shunting. Although there is a progressive increase in arterial $P\mathrm{o}_2$ with an increase in functional residual capacity, increasing positive end-expiratory pressure has the potential for hyperinflation, increased intrathoracic pressure, decreased preload and increased afterload to the right ventricle, shifting of the intraventricular septum with decreased filling of the left ventricle, and barotrauma. Since the benefits and complications of positive end-expiratory pressure must be balanced, there is no agreement on what is the optimum.

We will illustrate the salient features of a few among the most common modalities of mechanical ventilation, emphasizing that none has been proved to be universally superior and that the choice is dictated by the characteristics of the individual patient, the time course of the respiratory failure, and the preference of experienced physicians.

Controlled mandatory ventilation

This implies that the machine is fully sustaining the task of breathing. Controlled mandatory ventilation is initiated by the respiratory effort of the patient using a pressure or flow trigger, or by the ventilator using a time trigger. To accomplish this mode of ventilation, at times heavy sedation (occasionally with paralysis) has to be instituted. It is reserved for patients with absent respiratory drive or with severe respiratory failure, and is not ideal during the withdrawal of mechanical support. All breaths in controlled mandatory ventilation are delivered at the preset flow. The major advantage is that the *V*/*T* remains constant; the major disadvantages are either overventilation or the risk of barotrauma.

Pressure-controlled ventilation

This mode is pressure-limited and time-cycled; the respiratory rate, inspiratory pressure, and inspiratory time are all preset. When the inspiratory time is reached, the patient can exhale passively. The inspiratory flow rate is variable and changes with patient demands, allowing for a lower tidal volumes in those with poor compliance.

Intermittent mandatory ventilation

This allows spontaneous breathing complemented by a variable input from the ventilator. The amount of ventilation provided can be progressively reduced as the patient's clinical condition improves. Intermittent mandatory ventilation is a simple and effective mode of weaning and is the most widely used means of ventilation support. A potential problem is the inability of some patients to couple their spontaneous activity with the mechanical breaths. In the synchronized modification of intermittent mandatory ventilation, the patient's breath inhibits the mechanical cycle, which is delivered only after a preset pause. Synchronized intermittent mandatory ventilation was developed to prevent auto-positive end-expiratory pressure.

Assist/controlled ventilation

This allows the patient to initiate every breath spontaneously; the machine responds to the negative pressure exerted by their inspiratory effort by delivering a preset tidal volume. By progressively decreasing the size of the assisted breath, the patient is allowed more work of breathing until the input from the ventilator becomes negligible. Although very appealing in principle, assist/controlled ventilation has several drawbacks. The less than perfect design of many demand valves may cause either excessive work to trigger the mechanical breath or, on the other hand, the delivery of breaths in response to stimuli such as coughing or moving, which may cause periods of distinct discomfort. Also, since the size of the assisted breath is the only variable that can be set, this does not always match the patient's pattern of breathing: when their peak inspiratory flow is higher than that supported by the ventilator, the mechanical breath may reach them while exhalation has started, causing discomfort and increased work of breathing.

Inspiratory pressure-support ventilation

This is a relatively new mode of assist ventilation. A preset amount of pressure is delivered in response to the patient's inspiratory effort; the tidal breath is delivered at an inspiratory flow that is always higher than the patient's, thus avoiding dysynchrony. The only preset variable is the inspiratory pressure; the patient controls the entire

breathing cycle, receiving smooth assistance from the ventilator. As long as the inspiratory effort is maintained, the preselected airway pressure stays constant, with a variable flow rate of gas from the ventilator. Tidal volume is determined by the preset pressure level, the inspiratory effort of the patient, airway resistance, and lung compliance. The inspiratory flow rate from the ventilator depends on the patient's demand. Higher pressures can be used to provide total ventilatory support. The patient's ventilatory drive does not cease and the ventilatory muscles continue to contract during the assisted respiratory cycle.

Advantages of this method are a physiologic pattern of breathing, decreased work of breathing, a patient-set respiratory rate, and the potential for less barotrauma.

Continuous positive airway pressure

This allows patients to breath spontaneously with no enforced breaths;the amount of positive pressure is set and maintained constantly while they breathe spontaneously.

Flow-triggering ventilation

Flow-triggering ventilation combines the advantage of both continuous-flow and demand-flow strategies. Flow triggering is defined by two variables, base flow and flow sensitivity. Base flow is the rate of continuous gas flow that circulates through the ventilator system. Flow sensitivity is the rate of inspiratory flow generated by the patient that triggers a breath delivery from the ventilator. The base flow satisfies the patient's demand for initial inspiratory flow, greatly reducing the delay time inherently present in any pressure-triggered system. Before the patient's inspiratory effort begins, the ventilator is already delivering a constant flow of gas into the circuit, which can satisfy the initial phase of the respiratory cycle. Flow triggering reduces the time interval between the start of a spontaneous breath and the delivery of gas through the ventilator circuit. It does not affect the work of the respiratory muscles after inspiration has begun; rather, it decreases the effort necessary to initiate gas flow.

Inverse-ratio ventilation

This is slowly becoming an alternative to conventional modes of ventilation. The concern that lung injury is induced by ventilators led to its development to improve gas exchange by prolonging the inspiratory time. Inverse ratio ventilation alters the intrinsic positive end-expiratory pressure and increases mean airway pressure, redistributing alveolar fluid and improving oxygenation.

Noninvasive ventilation

Noninvasive ventilation is the delivery of assisted mechanical ventilation without the invasive airway device. The development of the nasal continuous positive airway pressure for the treatment of sleep apnea led investigators to try this mode of ventilation in various types of patient. It has the advantage of avoiding intubation, decreasingpeak airway pressure, and preserving host defenses. It can only be used in cooperative patients under strict supervision. It can be used in patients with congestive heart failure, cystic fibrosis, acute asthma and pneumonia, and in those with respiratory insufficiency after extubation. Selection is aimed at those with acute respiratory distress who might need intubation but excluding those with mild exacerbations. Noninvasive ventilation should not be started in patients who have severe respiratory failure or are medically unstable. Thosewho have reasonably reversible causes usually benefit and improve dramatically. Pressure-support and flow-triggering ventilation can be added to noninvasive ventilation.

Discontinuation of mechanical ventilation

Once the patient who has been subjected to prolonged mechanical ventilation finally gets close to being weaned from the machine, some objective criteria to predict the success of extubation are needed. Bedside measurements of the mechanics of the respiratory system are easily obtainable; suggested measures include normal arterial oxygenation, Fio_2 less than 0.5, and a respiratory rate of fewer than 20 breaths/min, a forced vital capacity approaching 1 l/min, ventilation of less than 10 l/min, and a maximum inspiratory pressure greater than -25 cmH_2O. There is no index or combination of variables that can be an accurate predictor for weaning and extubation. The most accurate ratio to predict weaning success is respiratory rate divided by *V*t (less than 105).

Common sense should warn against predetermined criteria: if too stringent, few patients will fail, but many others will be subjected to unnecessarily prolonged mechanical support.

Only the daily observation of their progress, coupled with continuous learning about the underlying pathophysiology of respiratory failure, can improve the success rate in the treatment of these patients.

Further reading

Aitkenhead AR *et al.* Comparison of propofol and midazolam for sedation in critically ill patients. *Lancet* 1989; **ii** : 704–9.

Appelberg M, Gordon L, Fatti LP. Preoperative pulmonary evaluation of surgical patients using the vitalograph. *British Journal of Surgery* 1974; **61**: 57–9.

Bains M. Surgical treatment of lung cancer. *Chest* 1991; **100**: 826–37.

Beecher HK. Effect of laparotomy on lung volume: demonstration of a new type of pulmonary collapse. *Journal of Clinical Investigation* 1933; **12**: 651.

Bersten A *et al.* Treatment of severe cardiogenic pulmonary edema with continuous positive airway pressure delivered by face-mask. *New England Journal of Medicine* 1991; **325**: 1825–30.

Bhavani-Shankar K, Moseley H, Kumar AY, Delph Y. Capnometry and anaesthesia. *Canadian Journal of Anaesthesia* 1992; **39**(2): 617–32.

Brochard L, Isabey D, Piquet J. Reversal of acute exacerbation of chronic obstructive disease by inspiratory assistance with face mask. *New England Journal of Medicine* 1990; **323**: 1523–30.

Celli BR. Perioperative respiratory care of the patient undergoing upper abdominal surgery. *Clinics in Chest Medicine* 1993; **14**: 253–61.

Celli BR, Rodriguez KS, Snider GL. A controlled trial of intermittent positive pressure breathing, incentive spirometry and deep breathing exercises in preventing pulmonary complications after abdominal surgery. *American Review of Respiratory Disease* 1984; **130**: 12–15.

Consensus statement. *Respiratory Care* 1997; **42**: 365.

Epstein S, Faling J, Daly B, Celli B. Predicting complications after pulmonary resection. *Chest* 1993; **104**: 694–700.

Epstein SK, Faling J, Daly B, Celli B. Inability to perform bicycle ergometry predicts increased morbidity after lung resection. *Chest* 1995; **107**: 311–16.

Froese AB, Bryan AC. Effects of anesthesia and paralysis on diaphragmatic mechanics in man. *Anesthesiology* 1974; **41**: 242–55.

Galanos AN *et al.* Relationship of body mass index to subsequent mortality among seriously ill hospitalized patients. *Critical Care Medicine* 1997; **25**: 1962–8.

Gerson MC, Hurst JM, Hertzberg VS, Bayghman R, Rouan GW. Prediction of cardiac and pulmonary complications related to elective abdominal and noncardiac thoracic surgery in geriatric patients. *American Journal of Medicine* 1990; **88**: 101–7.

Goldhaber SZ *et al.* A prospective study of risk factors for pulmonary embolism in women. *Journal of the American Medical Association* 1997; **277**: 642–5.

Gracey DR, Divertie MB, Didier EP. Preoperative pulmonary preparation of patients with chronic obstructive pulmonary disease. *Chest* 1979; **76**: 123–9.

Hall JC, Tarala RA, Hall JL, Mander J. A multivariate analysis of the risk of pulmonary complications after laparotomy. *Chest* 1991; **99**: 923–7.

Keagy BA *et al.* Elective pulmonary lobectomy: factors associated with morbidity and operative mortality. *Annals of Thoracic Surgery* 1985; **40**: 349–52.

Kearney DJ, Lee TH, Reilly JJ, DeCamp, MM, Sugarbaker DJ. Assessment of operative risk in patients undergoing lung resection. *Chest* 1994; **105**: 753–59.

Latimer RG *et al.* Ventilatory patterns and pulmonary complications after upper abdominal surgery determined by preoperative and postoperative computerized spirometry and blood gas analysis. *American Journal of Surgery* 1971; **122**: 622–32.

Lawrence V, Dhanda R, Hilsenbeck S, Page C. Risk of pulmonary complications after elective abdominal surgery. *Chest* 1996; **110**: 744–50.

Loiacono J, Cunneen C. Mechanical ventilation of the postoperative patient. In: Faber P, Wilson RS, ed. *Chest Surgery Clinics of North America*, Vol. 7, pp. 801–4. Saunders, Philadelphia, 1997.

Marik P, Varon J. The obese patient in the ICU. *Chest* 1998; **113**: 492–8.

Markos J, Mullan BP, Hillman DR, Musk AW, Centico VF, Lovegrove FT. Preoperative assessment as a predictor of mortality and morbidity after lung resection. *American Review of Respiratory Disease* 1989; **139**: 902–10.

Meduri GU. Diagnosis of ventilator associated pneumonia. *Infectious Disease Clinics of North America* 1993; **7**: 295–329.

Melendez J, Barrera R. Predictive respiratory complication quotient predicts pulmonary complications in thoracic surgical patients. *Annals of Thoracic Surgery* 1998; **66**: 220–4.

Mitchell C *et al.* Multivariate analysis of factors associated with postoperative pulmonary complications following general elective surgery. *Archives of Surgery* 1998; **133**: 194–8.

Montravers P, Dureuil B, Molliex S, Desmonts JM. Effects of intravenous midazolam on the work of breathing. *Anesthesia and Analgesia* 1994; **79**: 558–62.

Morice R, Peters E, Ryan B, Putman J, Ali M, Roth K. Exercise testing in the evaluation of patients at high risk for complications from lung resection. *Chest* 1992; **101**: 356–61.

Oh SH, Patterson R. Surgery in corticosteroid-dependent asthmatics. *Journal of Allergy and Clinical Immunology* 1974; **53**: 345–51.

Pedersen T, Eliasen K, Henriksen H. A prospective study of risk factors and cardiopulmonary complications associated with anesthesia and surgery: risk indicators of pulmonary morbidity. *Acta Anaesthesologica Scandinavica* 1990; **34**: 144–55.

Peters RM, Turnier E. Physical therapy. Indications for and effects in surgical patients. *American Review of Respiratory Disease* 1980; **122**: 147–54.

Pierce RJ, Copland JM, Sharpe K, Barter C. Preoperative risk evaluation for lung cancer resection: Predicted postoperative product as a predictor of surgical mortality. *American Journal of Respiration and Critical Care* 1994; **150**: 947–55.

Pontoppidan H. Mechanical aids to lung expansion in non-intubated patients. *American Review of Respiratory Disease* 1980; **122**: 109–19.

Reilly JJ, Mentzer SJ, Sugarbaker DJ. Preoperative assessment of patients undergoing pulmonary resection. *Chest* 1993; **103**: 342–5S.

Rock P. The patient with lung disease. *Problems in Anesthesia* 1997; **9**: 165.

Rose DK, Cohen MM, Wigglesworth DF, DeBoer DP, Math M. Critical respiratory events in the postanesthesia care unit: patient, surgical and anesthetic factors. *Anesthesiology* 1994; **81**: 410–18.

Roukema JA, Carol EJ, Prins JG. The prevention of pulmonary complications after upper abdominal surgery in patients with non-compromised pulmonary status. *Archives of Surgery* 1988; **123**: 30–4.

Slutsky A. ACCP consensus conference: mechanical ventilation. *Chest* 1993; **104**: 1833–59.

Smith-Choban P, Weireter LJ, Maynes C. Obesity and increased mortality in blunt trauma. *Journal of Trauma* 1991; **31**: 1253–57.

Stein M, Cassara EL. Preoperative pulmonary evaluation and therapy for surgical patients. *Journal of the American Medical Association* 1970; **211**: 787–90.

Tarhan S *et al.* Risk of anesthesia and surgery in patients with chronic bronchitis and chronic obstructive pulmonary disease. *Surgery* 1973; **74**: 720–6.

Tisi GM. Preoperative evaluation of pulmonary function. *American Review of Respiratory Disease* 1979; **119**: 293–346.

Tremper KK, Baker SJ. Pulse Oximetry. *Anesthesiology* 1989; **70**: 98–108 Williams MH, ed. Disturbances of respiratory control (symposium). *Clinics in Chest Medicine* 1980; **1**: 1–59.

Vaughan RW, Engelhart RC, Wise L. Postoperative hypoxemia in obese patients. *Annals of Surgery* 1974; **180**: 877–82.

Warner MA, Divertie MB, Tinker JH. Preoperative cessation of smoking and pulmonary complications in coronary artery bypass patients. *Anesthesiology* 1984; **60**: 380–3.

Warner MA, Offord KP, Warner ME, Lennon, RL, Conover, MA, Jansson-Schumacher U. Role of preoperative cessation of smoking and other factors in postoperative pulmonary complications: blinded prospective study of coronary artery bypass patients. *Mayo Clinic Proceedings* 1989; **64**: 609–16.

Yang K, Tobin M. A prospective study of indices predicting the outcome of trials of weaning from mechanical ventilation. *New England Journal of Medicine* 1991; **324**: 1445–50.

Zeiher BG, Gross TJ, Kern JA, Lanza, LA, Pererson MW. Predicting postoperative pulmonary function in patients undergoing lung resection. *Chest* 1995; **108**: 68–72.

12.2 Cardiological problems

Pierre Foëx

Cardiovascular diseases
Coronary artery disease
Arterial hypertension
Heart failure
Valvular disease
Cardiomyopathies
Cerebrovascular disease
Peripheral vascular disease
Dysrhythmias and heart blocks
Global assessment of risk
Further investigations
Chest radiograph
Electrocardiography
Nuclear imaging
Cardiac catheterization
Preoperative pulmonary arterial catheterization
Cardiovascular medication
Adrenergic b-receptor antagonists
Calcium-channel antagonists
Angiotensin-converting enzyme inhibitors
Angiotension-II antagonists
Nitrates
Cardiac glycosides
Conclusion
Further reading

Anaesthesia and surgery carry increased risks in patients suffering from cardiovascular disease. This problem is accentuated in elderly individuals. Ischaemic heart disease, chronic infection of the lower respiratory tract, and cardiac failure are the disorders most commonly associated with postoperative deaths. Over the last decade, better preoperative assessment and the introduction of more sophisticated monitoring have increased the safety of surgery in patients with cardiovascular disease, as it has become possible for the anaesthetist to detect changes in the circulation before life-threatening complications occur. Nevertheless, myocardial infarction, progressive myocardial ischaemia, dysrhythmias, congestive cardiac failure, and cerebrovascular accidents continue to occur relatively frequently, reflecting the trend to undertake invasive surgical procedures even in the severely ill patient. Indeed, almost one-third of all surgical patients have either coronary heart disease or associated risk factors. Moreover, despite advances in anaesthetic technique and postoperative pain relief, the perioperative period imposes prolonged stress. Relatively silent cardiac diseases are often initially diagnosed when a patient is admitted for surgery. Medical history, clinical examination, interpretation of the electrocardiogram, chest radiographs, and renal function tests form the basis of the preoperative evaluation. The need for this assessment to be supplemented by echocardiography, radionuclide cardiac imaging, cardiac catheterization, angiography, or ambulatory electrocardiographic monitoring is, however, increasing. Objective data are required to enable decisions to be made about the overall surgical and anaesthetic strategy, including the extent of cardiovascular monitoring and the need for postoperative management in a high-dependency or intensive care unit, as the most conscientious clinical assessment often fails to detect serious physiological abnormalities that should be remedied before anaesthesia and surgery. This requires familiarity with cardiovascular drugs that may modify haemodynamic responses to stresses.

Cardiovascular diseases

Coronary artery disease

Ischaemic heart disease is the most common cause of death in developed countries, accounting for 30 per cent of all deaths in men, and 25 per cent of deaths in women. Its incidence increases markedly with age and its presence is a major threat to surgical patients.

Recent infarction

The risk of perioperative myocardial infarction is increased 30- to 300-fold in patients who have suffered a previous infarction; when surgery takes place within 3 months of the infarction the risk of reinfarction may be as high as 25 per cent. However, the delay between myocardial infarction and surgery is only one factor that needs to be taken into consideration. The type of surgical procedure is also important: reinfarction is more likely to occur after abdominal, thoracic, and prolonged surgery than after body-surface surgery. The quality of the patient's ventricular function is paramount, as an ejection fraction of less than 40 per cent increases the risk of perioperative cardiac complications and death dramatically. Indeed, the presence of left ventricular failure at any time after myocardial infarction has been shown to be the most significant factor associated with death after emergency surgery for appendicitis or hip fracture. Persistence of angina after infarction is also an ominous sign, as it indicates the presence of a compromised myocardium.

The inevitability of high rates of morbidity and mortality associated with anaesthesia and surgery after a recent myocardial infarction has recently been questioned. Extended monitoring of the electrocardiogram, arterial, central venous, and pulmonary capillary wedge pressure, combined with aggressive management of cardiovascular abnormalities during the perioperative period, may reduce the risk of reinfarction to about 5 per cent. To keep variations of heart rate, arterial pressure, and pulmonary wedge pressure within very narrow limits, however, patients must be admitted to an intensive care unit for a relatively long period. It remains safer, in most circumstances, to delay elective surgery for at least 3 months after myocardial infarction.

Angina

Angina is a well recognized risk factor for perioperative complications and death, but there are subsets of patients in whom the risks are particularly high. Unstable angina is associated with risks similar to those of recent myocardial infarction. Unless surgery can be considered life-saving, patients should be made stable before any other procedure is undertaken. Disabling angina increases the risk of perioperative ischaemia, myocardial infarction, and cardiac death, while mild angina is a threat mostly for patients with preoperative abnormalities in the ST–T segment.

Silent ischaemia

Over the last decade it has become obvious that many patients suffer from silent myocardial ischaemia, which can be detected by exercise testing or ambulatory electrocardiographic monitoring.

There are three types of silent myocardial ischaemia. Type 1 is observed in patients without any symptoms and occurs in about 2.5 per cent of the male population aged between 39 and 59 years, about 40 per cent of whom will develop symptomatic coronary arterial disease. Type 2 occurs in about 16 per cent of patients who have suffered myocardial infarction, and type 3 occurs in patients suffering from angina. In these patients, 75 per cent of all episodes of ischaemia are silent, and only 25 per cent are painful.

The absence of pain may be a reflection of less severe ischaemia, a smaller affected area, or a different distribution of impairment of coronary blood flow. However, it is more likely that patients with silent ischaemia have a higher pain threshold, related to higher concentrations of plasma endorphins.

Silent ischaemia has been reported in 20 to 90 per cent of patients presenting for coronary-artery or vascular surgery, and is known to increase the risk of postoperative complications. Silent ischaemia is a predictor of death and myocardial infarction for 2 years and all adverse cardiac events for 5 years. Moreover, preoperative silent myocardial ischaemia is a strong independent predictor of cardiac events for 1 to 2 years after peripheral vascular surgery.

Coronary-artery bypass grafts

The risk of perioperative infarction after successful revascularization of coronary arteries is low, and both morbidity and mortality are close to those of patients without heart disease. Thus, coronary revascularization is often advocated before major surgery is undertaken in patients with coronary heart disease. The low incidence of cardiovascular complications in non-cardiac surgery after a coronary bypass graft may be somewhat misleading, however, since the morbidity and mortality associated with the coronary revascularization itself should be taken into account; this makes the place of coronary revascularization preceding major surgery controversial. When it is indicated in its own right, for example in patients with unstable angina, correctable disabling angina, or triple-vessel disease, coronary revascularization should be done first. When the indications are less clear cut, it is probably advisable only before surgery that is likely to be associated with major haemodynamic instability, such as major vascular surgery. In some patients, coronary bypass surgery improves cardiac function, probably because some areas of the left ventricle are still viable, if not functional. The concept of ‘hibernating' myocardium explains this situation: myocardium with very poor blood supply may cease to contract, yet receives enough blood to maintain cell viability. Reperfusion restores mechanical activity and cardiac function. In patients with pre-existing ventricular dysfunction this does not always occur after coronary graft procedures and left ventricular function may remain severely compromised.

Fleisher and colleagues have developed a decision-analytical approach to the question of coronary arterial revascularization before surgery for abdominal aortic aneurysms. Decision analysis involves modelling each step in a decision process, assigning probability values for each step. Decision analysis is firmly grounded in probabilities obtained from the literature and takes into consideration, in its final stage, the local mortality rate for both repair of abdominal aortic aneurysms and coronary revascularization. Two-way sensitivity analysis delineates the need to perform detailed tests to establish the presence of coronary lesions and their suitability for coronary revascularization, rather than proceed with the planned vascular procedure. For example, in a centre where the mortality from surgery for abdominal aortic aneurysm in patients with coronary arterial disease is 8 per cent and the mortality of coronary revascularization is 2 per cent, decision analysis favours prior coronary revascularization. Conversely, if the mortality from abdominal aortic aneurysm in patients with coronary arterial disease is 5 per cent and the mortality from coronary revascularization is also 5 per cent, the lesser risk is to go ahead with a repair for the abdominal aortic aneurysm.

Causes of ischaemia

Autoregulation of the coronary circulation is impaired in the face of coronary arterial disease and the balance between oxygen demand and supply is easily compromised, leading to regional myocardial ischaemia. Ischaemia may be caused by haemodynamic aberrations, acute reductions in oxygen supply, or acute reductions in arterial blood saturation (hypoxaemia). The major haemodynamic causes of ischaemia are: (i) tachycardia, during which oxygen demand is increased while oxygen supply is impaired because of the shorter duration of diastole; (ii) hypotension, during which coronary perfusion pressure is reduced (as blood flow through narrowed coronary vessels is directly proportional to the perfusion pressure) so that the reduction in perfusion may be greater than the reduction in oxygen demands; (iii) hypertension, which increases oxygen demand, in the face of an insufficient coronary flow reserve; and (iv) left ventricular overfilling, which increases wall tension and oxygen demand, while compromising coronary blood flow because of the augmented tissue pressure. These cardiovascular abnormalities are responsible for many episodes of perioperative myocardial ischaemia and infarction.

Acute reductions in oxygen supply may result from the effects of sympathetic or parasympathetic activity in the presence of damage to coronary endothelium. Release of noradrenaline causes exaggerated vasoconstriction and release of acetylcholine causes muscarinic-dependent vasoconstriction rather than nitric oxide-mediated vasodilatation. Similarly, endothelial damage alters the balance of vasodilators (nitric oxide, prostacyclin) and vasoconstrictors (endothelins, thromboxane A_2, platelet-activating factor), such that vasoconstriction may occur where a dynamic stenosis is present. Finally, many recent studies have shown that nocturnal hypoxaemia occurs frequently after surgery and may be associated with silent myocardial ischaemia.

Abnormalities of left ventricular-wall motion are frequently detected in patients suffering from coronary arterial disease. Non-invasive diagnostic methods relying on radionuclides (radionuclide cineangiography, perfusion scans), and ultrasound (echocardio-graphy) are of proven value and should be used systematically in the preoperative assessment of patients with significant coronary arterial disease. Where substantial abnormalities of wall function are present, or if left ventricular aneurysms are discovered, the negative inotropy of most anaesthetic agents is likely to cause severe depression of global cardiac function.

It is becoming increasingly obvious that the clinical assessment of left ventricular function is difficult, especially in patients whose activity is limited by peripheral vascular disease or other disabilities. Radionuclide studies show that severe dysfunction is frequently present in patients who appear essentially asymptomatic: about one-third of patients presenting for major vascular surgery have ejection fractions below 50 per cent.

Arterial hypertension

It is convenient to consider that a systolic pressure above 160 mmHg, and a diastolic pressure above 90 mmHg define arterial hypertension. Hypertension is present in between 10 and 20 per cent of the adult population and carries the risk of serious cardiovascular complications such as stroke, ischaemic heart disease, peripheral vascular disease, and renal dysfunction.

Clinical examination may reveal an enlarged heart and signs of left ventricular failure. The chest radiograph and electrocardiogram may confirm left ventricular hypertrophy. Signs of previous myocardial infarction, myocardial ischaemia, and disorders of intraventricular conduction are often noted. The cerebral complications of hypertension include transient ischaemic attacks and major cerebrovascular accidents. Renal complications may be shown by proteinuria, and elevated serum creatinine and urea. Inspection of the retina may show diminished vessel calibre, nipping, and flame-shaped haemorrhages. Approximately 90 per cent of hypertensive patients suffer from essential hypertension; less than 10 per cent have secondary hypertension. Some causes of secondary hypertension are listed in Table 1, and it is a possibility that must be kept in mind.

1. *Renal diseases*
(a)Unilateral renal disorders:
Renal artery stenosis
Obstructive nephropathy
Pyelonephritis
(b) Bilateral renal disorders:
Renal artery stenosis
Glomerulonephritis
Interstitial nephritis
Pyelonephritis
Polycystic disease
Collagen vascular disease
Obstructive uropathy
2. *Adrenal disorders*
Primary aldosteronism
Cushing's syndrome
Phaeochromocytoma. 3. *Drug associated*
Oral contraceptives
Corticosteroids
Sympathomimetics
4. *Other causes*
Coarctation of the aorta
Pre-eclampsia
Raised intracranial pressure
Tetanus

Table 1 Causes of secondary hypertension

Long-term treatment of arterial hypertension improves the prognosis of hypertensive patients. This is also true of those with isolated systolic hypertension and of elderly hypertensive patients. In surgical patients, hypertension is associated with a high incidence of perioperative hypertensive crises, dysrhythmias and myocardial ischaemia, and increases the risk of reinfarction.

While maintenance of the antihypertensive drug regimen is essential in order to ensure cardiovascular stability and prevent rebound hypertension, the need to initiate antihypertensive therapy before anaesthesia and surgery in untreated patients is more controversial. Yet this is a frequent problem because 45 per cent of hypertensive patients are not controlled. As postponement of surgery may cause considerable inconvenience to patients and disrupt operative programmes, it is essential to identify those patients in whom the risks of anaesthesia and surgery will be significantly improved by treatment of hypertension. This may be done in considering the severity of the hypertensive heart disease (Table 2).

Category	Systolic (mmHg)	Diastolic (mmHg)
Mild hypertension	140–159	90–99
Moderate hypertension	160–179	100–109
Severe hypertension	≥ 180	≥ 110

Table 2 Classification of severity of hypertension

The presence of mild hypertension increases the risks of anaesthesia and surgery, but there is no evidence that treatment reduces this risk. Moderate hypertension poses a more substantial threat, especially in the presence of target-organ disease, such as significant coronary, cerebrovascular, or renal disease; treatment of hypertension before elective surgery is then recommended. In severe arterial hypertension optimal treatment before elective surgery is essential, as the incidence of dysrhythmias and/or myocardial ischaemia is far higher in untreated than in treated patients. A major cause for concern is the possibility of very large increases in arterial pressure accompanied by tachycardia and increases in the pulmonary wedge pressure at the time of endotracheal intubation, extubation, and during the recovery period. The increase in pulmonary wedge pressure may represent both failure of the left ventricle to eject and reduced ventricular compliance. Hypertensive crises may result in strokes, myocardial infarction, pulmonary oedema, or life-threatening dysrhythmias. Treatment of hypertension blunts these responses, especially when it includes b-adrenoceptor blockers.

If treatment of hypertension is indicated, it should be continued for several weeks before surgery since ‘cosmetic' reductions in blood pressure will be obtained very quickly, but improvement of the underlying vascular hyper-reactivity and cerebrovascular autoregulation will occur gradually over weeks and not days.

Regional anaesthesia is often considered a safe alternative to general anaesthesia in untreated hypertensive patients. This is not true: lumbar epidural anaesthesia has been shown to cause greater reductions in systolic and diastolic arterial pressure in untreated than in treated hypertensive patients.

Hypertension and left ventricular hypertrophy

As left ventricular hypertrophy reduces ventricular compliance, excessive volume loading may cause pulmonary oedema, while reduced filling may result in a dramatic reduction in cardiac output. Moreover, in patients with left ventricular hypertrophy, marked discrepancies may exist between right and left ventricular-filling pressures. Monitoring the pulmonary capillary-wedge pressure may therefore be essential for safe perioperative management after major surgery.

Hypertension in the elderly

Isolated systolic hypertension is due predominantly to the loss of elasticity of the aorta and its major branches. The high pressure increases myocardial oxygen consumption and may cause myocardial ischaemia. However, the treatment of isolated systolic hypertension may be associated with subjective complaints, and on occasion, objective deterioration of cardiac, cerebral, or renal function; in addition titration of blood pressure is often difficult. Treatment of purely systolic hypertension in the elderly is probably not justified before surgery, but adequate cardiovascular monitoring is essential to enable excessive hypo- or hypertension to be detected and treated immediately.

When both systolic and diastolic pressures are elevated, the risk of cardiovascular complications are increased in elderly as well as in younger patients. Antihypertensive treatment reduces the risk of complications but it is important to achieve the reduction gradually in order for cerebral autoregulation to return to normal limits.

Evidence for the risks associated with hypertension in surgical patients

Recently, a case-control study has shown that hypertension was present in 49 per cent of patients who had died within 30 days of surgery, and only 20 per cent of matched survivors. This indicates that the presence of hypertensive disease is a risk factor. Another study has shown that patients with untreated or poorly treated hypertension were substantially more at risk of silent myocardial ischaemia (a known risk factor in surgical patients) than normotensive or well-controlled hypertensive patients. These data suggest that optimizing the control of blood pressure before surgery may reduce the risk of adverse outcome even in patients with mild or moderate hypertension.

Heart failure

Heart failure has become the most frequent cause of admission to hospital in persons 65 years of age or older. Despite effective therapies for heart failure, such as angiotensin-converting enzyme inhibitors, and cardiac transplantation, the prognosis remains poor and deaths due to heart failure have increased.

Incipient right or left ventricular failure, indicated by the presence of a third heart sound or distended jugular veins, increases the relative risk of complications by 2 to 10 times. Overt ventricular failure must be treated before any surgical procedure is undertaken in order to reduce the risk of postoperative pulmonary oedema, and invasive monitoring may be necessary to maintain optimal ventricular filling throughout the perioperative period.

Low ejection fraction is associated with increased risks of anaesthesia and surgery. The author's personal views regarding major high-risk surgical procedures are as follows: when the ejection fraction is greater than 50 per cent, cardiac function is likely to be adequate and the risks are acceptable in the absence of other risk factors, such as reversible myocardial ischaemia and impaired pulmonary or renal functions. When the ejection fraction is between 40 and 50 per cent, the risks are increased, and the presence of additional risk factors should prompt reconsideration of the risk/benefit of surgery, and to consider performing the most limited operation possible. Postoperative care in a high-dependency unit is necessary. When the ejection fraction is between 30 and 40 per cent, whether or not additional risk factors are present, management must be carefully considered. Surgery may have to be postponed to allow optimal medical treatment. Alternatively, less extensive procedures may have to be selected. Surgery may be delayed until the balance of risk is increasingly in favour of surgery rather than abstention, or it may be cancelled altogether. Again, intensive-care management after major surgery is essential. Ejection fractions of less than 30 per cent are usually observed in patients who have intermittent left ventricular failure. Recent episodes of left ventricular failure may require admission of the patient to an intensive care unit before surgery so that full monitoring may be implemented and the circulation optimized with intravenous drugs ahead of major procedures if, and only if, surgery is deemed essential.

Valvular disease

The incidence of rheumatic fever has fallen in the Western world, but many elderly patients still suffer from rheumatic heart disease; others suffer from degenerative valvular disease and the prevalence of valvular disease in people over the age of 65 may be as high as 4 per cent.

Aortic stenosis is now the most common valvular lesion; it leads to the development of a pressure gradient between the left ventricle and the aorta. Left ventricular hypertrophy develops as a result of increased stroke work; hypertrophy is accompanied by an increase in diastolic stiffness and an increase in end-diastolic pressure. Raised ventricular pressure and increased muscle mass augment myocardial oxygen consumption, but myocardial blood flow is impaired due to the reduction in aortic diastolic pressure. During the perioperative period, close attention must be paid to changes in heart rate and arterial pressure. Increases in heart rate are poorly tolerated because they reduce the duration of diastole and thus decrease coronary blood flow. Similarly, interventions that reduce systemic vascular resistance decrease both the coronary perfusion pressure and the coronary blood flow, and may precipitate myocardial ischaemia and sudden death. (Table 3). In patients with aortic stenosis, spinal and epidural anaesthesia, because they decrease diastolic arterial pressure, may cause acute, irreversible, myocardial ischaemia.

Type of lesion	Event	Mechanism	Effect
Mitral stenosis	Bradycardia	Fixed stroke volume	Reduced cardiac output
	Tachycardia	Reduced filling time	Reduced cardiac output
	Atrial fibrillation	Loss of atrial contribution to ventricular filling	Reduced cardiac output
Mitral regurgitation	Peripheral vasoconstriction	Increased regurgitant flow	Decreased forward flow
Aortic stenosis	Bradycardia	Fixed stroke volume	Reduced cardiac output
	Tachycardia	Reduced diastolic time	Reduced coronary flow
	Peripheral vasodilatation	Reduced coronary perfusion pressure	Reduced coronary flow
Aortic regurgitation	Bradycardia	Increased regurgitant flow	Reduced cardiac output
	Peripheral vasoconstriction	Increased regurgitant flow	Reduced cardiac output

Table 3 Adverse haemodynamic interventions in valvular diseases

Aortic regurgitation causes an increase in left ventricular stroke volume with a corresponding increase in left ventricular cavity size. In moderately severe aortic regurgitation, the stroke volume may be double the normal value. End-diastolic pressure in the aorta is low. Forward flow is facilitated by a relatively low vascular resistance, while regurgitant flow is increased by peripheral vasoconstriction, or by the prolonged diastole associated with bradycardia.

Mitral stenosis impedes left ventricular filling. Flow through the narrowed mitral valve depends on the left atrial pressure, which determines early diastolic flow, and the contribution of atrial systole, which determines late diastolic flow. As duration of diastole is an important determinant of ventricular filling, tachycardia is poorly tolerated. Atrial fibrillation further compromises ventricular filling, especially when the ventricular rate is fast. In patients with mitral stenosis, the main benefit of digitalis therapy is to maintain the heart rate below 80 beats/min. In many patients, mean left-atrial pressure is greater than 15 mmHg and volume overload will rapidly lead to pulmonary oedema.

Mitral regurgitation results in dilatation of both left atrium and left ventricle. The regurgitant flow causes high pressure in the atrium during systole, but there is no obstruction to diastolic forward flow except in combined stenosis and regurgitation. Any increase in systemic vascular resistance will limit left ventricular forward ejection and exaggerate retrograde flow into the atrium.

Cardiomyopathies

Cardiomyopathies are myocardial disorders that are not secondary to coronary disease or hypertension. Three main types include hypertrophic, dilated, and restrictive cardiomyopathies. The diagnosis is based on echocardiography. Hypertrophic cardiomyopathy in defined populations has a prevalence of 1 in 500 to 1 in 5000 in the United States. Dilated cardiomyopathy accounts for 2 to 4 per cent of cases of heart failure. Hypertrophic obstructive cardiomyopathy is characterized by massive ventricular hypertrophy associated with failure of the ventricles to relax adequately. Systolic function is maintained and rapid ejection often causes pressure gradients within the cavity of the ventricle. The evolution of hypertrophic cardiomyopathy is often slow over several decades. Life expectancy is not necessarily reduced. Interventions that increase the inotropic state of the myocardium worsen left ventricular function, while this is improved by drugs that reduce contractility, hence the beneficial effects of b-adrenoceptor blockers and calcium antagonists.

Dilated cardiomyopathy implies marked dilatation and impaired contractile performance. The ejection fraction is usually extremely low. Previous viral myocarditis and excessive alcohol consumption are the main predisposing factors. Diuretics and angiotensin-converting enzyme inhibitors are the drugs of choice, digoxin is helpful in advanced cases. Sudden death may result from arrhythmias.

Restrictive cardiomyopathy is characterized by diastolic dysfunction in the absence of systolic dilatation. Amyloid infiltration causes this restrictive syndrome, which has a very poor prognosis.

Cerebrovascular disease

Cerebrovascular disease is responsible for death in 9 per cent of men and 15 per cent of women in the United Kingdom. The symptoms and signs of previous cerebrovascular accidents are easily detected when they have resulted in a significant functional deficit. However, transient disturbances of cerebral function should not be overlooked as they may indicate the presence of atherosclerotic lesions of the cerebral vasculature. Particular attention should be paid to reduced carotid arterial pulsation and to carotid arterial bruits, as carotid arterial stenosis may be responsible for transient ischaemic attacks. Carotid Duplex scan, angiography, and carotid endarterectomy must be considered in patients with symptomatic carotid disease, as hypotensive episodes during any type of surgery may precipitate a stroke in patients with significant stenosis of the carotids. The prognosis of postoperative stroke is poor, with a mortality of about 50 per cent. If carotid disease is asymptomatic, prophylactic carotid endarterectomy is probably not indicated.

Carotid arterial disease may be considered a marker for coronary disease, as the mortality associated with carotid arterial surgery is predominantly due to myocardial infarction.

A high proportion of patients suffer from both cerebrovascular and hypertensive disease. Their management is difficult because their blood pressure can be very labile during anaesthesia and postoperatively. The incidence of neurological deficit is greater in those who develop hypertension after surgery and neurological deficits are much more likely to occur in untreated or poorly controlled hypertensives. Hypertensive crises can cause intracerebral haemorrhages. Gradual control of blood pressure is necessary in order for cerebral autoregulation to return to near normal.

Peripheral vascular disease

Between one-third and two-thirds of patients undergoing peripheral vascular surgery suffer from coronary heart disease, and the morbidity and mortality associated with anaesthesia and surgery are higher in such patients than in those undergoing non-vascular surgery. Clinical evaluation is often inadequate because of physical inactivity, and objective assessment of cardiac function and myocardial perfusion is invaluable, especially in patients presenting for surgery of the abdominal aorta and other intra-abdominal vessels. About one-third of these have a low cardiac output and an ejection fraction of less than 50 per cent.

The high incidence of complications following surgery of the abdominal aorta is a reflection of major haemodynamic disturbances. Cross-clamping of the aorta causes a sudden, large, increase in left ventricular afterload. In patients with compromised left-ventricular function, the pulmonary wedge pressure increases, reflecting acute left-ventricular dilatation associated with acute myocardial ischaemia. Myocardial ischaemia is particularly common when cross-clamping is applied above the renal arteries or the coeliac axis. Administration of vasodilators may be required to prevent or treat myocardial ischaemia during aortic cross-clamping. Removal of the clamps has three effects: firstly, vascular resistance is abruptly reduced; secondly, the sudden return of acidic blood to the heart causes cardiac depression; thirdly, the circulating volume is reduced because vasodilatation has developed in the lower limbs. Adequate volume loading before the gradual removal of cross-clamps is essential.

It is often assumed that surgery of limb vessels is relatively well tolerated because it does not cause major haemodynamic instability. While peripheral vascular surgery is better tolerated than aortic surgery, cardiovascular complications and fatality are still more frequent than in non-vascular surgery, reflecting the association of peripheral vascular disease with hypertensive and coronary disease. Indeed, in elderly patients, vascular surgery for limb salvage carries a mortality rate of up to 16 per cent.

The long-term morbidity and mortality in patients with peripheral vascular disease is the result of coronary events. It is, therefore, legitimate to consider the need for preoperative assessment of the coronary circulation of such patients and their cardiological follow-up if they have not been investigated, as severe coronary arterial lesions may be relatively silent. In this perspective, peripheral vascular disease may be ‘a screening test' for coronary disease and delineates a group of patients warranting further investigations.

Dysrhythmias and heart blocks

Any rhythm other than sinus rhythm is associated with a significant increase in the risk of cardiac complications because preoperative dysrhythmias reflect the underlying heart disease.

Dysrhythmias

Atrial dysrhythmias may precede the development of atrial tachycardia and atrial fibrillation. The loss of correctly timed atrial contraction reduces the ventricular filling, particularly when the ventricular rate is fast. In patients with atrial fibrillation or atrial flutter, the sinus rhythm should be restored if possible, or the ventricular rate controlled, before surgery. Fluid load may be poorly tolerated and, for major surgery, patients with atrial fibrillation may benefit from perioperative monitoring of the pulmonary capillary wedge pressure. Over the past decade, the number of hospital admissions for atrial fibrillation has more than doubled. With this the risk of embolic stroke and heart failure has increased; both conditions are associated with early deaths.

Ventricular ectopic activity is often associated with organic heart disease or digitalis toxicity. Premature ventricular beats frequently occur in patients with ventricular-wall dysfunction and poor ejection fraction. Cardiac function may worsen: many anaesthetic agents depress cardiac performance, and alter the electrophysiology of the myocardium either by a direct effect on the conduction tissue or by enhancing the arrhythmogenic activity of adrenaline.

Heart blocks

Heart blocks (Fig. 1) usually have an organic cause and may be accentuated by vagal stimulation, digitalis, b-adrenoceptor antagonists, and calcium antagonists. Anaesthesia may precipitate, by several mechanisms, the development of complete heart block in patients with a wide range of atrioventricular or intraventricular conduction disorders. First, some anaesthetic agents, particularly the halogenated inhalational anaesthetics, decrease atrioventricular conduction. Secondly, anaesthesia often causes dysrhythmias, and premature beats are known to facilitate the development of postectopic heart block. Thirdly, alterations in the serum potassium may occur: acute hypokalaemia may be caused by respiratory alkalosis (hypocapnic intermittent positive-pressure ventilation), while acute hyperkalaemia may follow the rapid infusion of large quantities of stored blood. These changes exacerbate dysrhythmias and compromise conduction, and the insertion of a temporary pacemaker is often necessary before elective or emergency surgery, even though permanent pacing may not be necessary (Table 4).

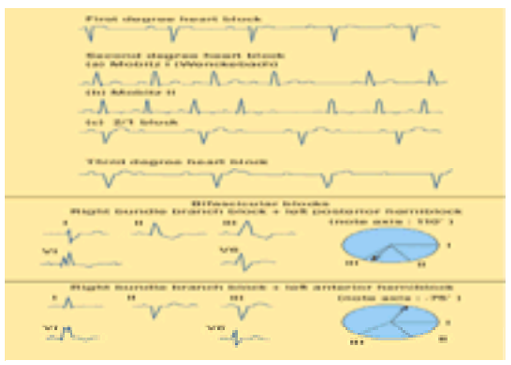


Fig. 1. Typical examples of heart blocks.

Symptomatic first-degree heart block
Symptomatic second-degree Mobitz I heart block
Second-degree Mobitz II heart block
Third-degree heart block
Symptomatic bifascicular block
Left bundle-branch block and first-degree heart block
'Sick sinus' syndrome
Slow rates unresponsive to drugs

Table 4 Indications for temporary pacemakers

The insertion of a temporary pacemaker is not usually recommended in asymptomatic patients with uncomplicated, first-degree heart block. Such patients do, however, occasionally develop profound bradycardia accompanied by hypotension while under general anaesthesia. Temporary pacing should be considered when this type of block is unresponsive to atropine, and must also be considered when first-degree heart block is accompanied by bundle-branch block. In Mobitz type I (Wenckebach) block, temporary pacing is necessary only when the patient is symptomatic. However, temporary (or permanent) pacing is necessary in Mobitz type II, second-degree heart block and in third-degree heart block.

The question of whether temporary pacing is required in patients with a bifascicular block is unsettled. Bifascicular blocks can be identified relatively easily: right bundle-branch block with left axial deviation of greater than -75° is associated with left anterior hemiblock; right bundle-branch block with right axial deviation greater than 110° is associated with left posterior hemiblock. Whenever patients with bifasicular blocks have experienced symptoms that can be attributed to transient complete heart block, temporary pacing is necessary. In totally asymptomatic patients, the risk of anaesthesia facilitating the development of complete heart block is probably very small. Complete left bundle-branch block accompanied by first-degree heart block may be an indication for temporary pacing.

The ‘sick sinus' syndrome represents a group of disturbances of impulse formation in the sinoatrial node that can easily lead to severe arrhythmias, including cardiac arrest during anaesthesia and surgery. The diagnosis should be suspected if there are periodic episodes of slow sinus rate alternating with tachycardia. The establishment of temporary pacing before anaesthesia and surgery is recommended since bradycardia may result from the treatment of tachycardia, and this is usually refractory to drug therapy.

Regional anaesthesia, especially epidural anaesthesia, is not safe in patients with conduction disorders: absorption of lignocaine or bipuvacaine injected into the extradural space may cause sinus bradycardia and induce heart blocks, especially in those with bundle-branch blocks. The same criteria for the insertion of a temporary pacemaker should be applied before regional anaesthesia and before general anaesthesia. Local anaesthesia may also cause a worsening of conduction disorders where large quanitities of local anaesthetics are used.

Patients with permanent pacemakers

Most pacemakers are implanted prophylactically for atrioventricular blocks or ‘sick sinus' syndrome, and they have recently been implanted for the control of recurrent tachydysrhythmias. Over the past 30 years the complexity of pacemakers has increased and they are now classified by the chamber paced, activity sensed, mode of response, and type of programmability. Before anaesthesia and surgery it is essential to establish the type of pacemaker, the risk of programmes being lost during diathermy, and the means of inducing a fixed rate, which may be necessary during the procedure, particularly when electrocautery is used.

The risk of interference by electrocautery is minimized by placing the indifferent diathermy plate well away from the pacemaker, and limiting the duration of electrocautery to 1-s bursts at intervals of at least 10 s.

Global assessment of risk

Several investigators have tried to determine which clinical preoperative factors contribute to the development of cardiac complications in patients undergoing major non-cardiac surgery. The index of cardiac risk in non-cardiac surgery developed by Goldman and his colleagues was based on the relation between postoperative complications and preoperative assessment data in just over 1000 patients. Using discriminant analysis to identify the factors that contributed significantly to the cardiac risks of anaesthesia and surgery, they produced an index that was able to predict the prognosis in over 80 per cent of cases (Table 5). When applied to a

selected population of patients undergoing surgery for abdominal aortic aneurysm, this index tended to underestimate the risk of complications. This is not surprising in view of the marked cardiovascular disturbances that occur during major vascular surgery.

Risk factor	*Risk points*		
S3 gallop, increased venous jugular pressure	11		
Pervious myocardial infarction (less than 6 months)	10		
Premature ventricular ectopics (more than 5 per min)	7		
Arterial dysrhythmias, rhythm other than sinus	7		
Age greater than 70 years	5		
Emergency operation	4		
Severe aortic stenosis	3		
Poor general condition	3		
Intraperitoneal or intrathoracic operation	3		
Class	*Complications %*		
	Cerebrovascular-cardiac surgery (%)	Abdominal-aortic surgery (%)	
Class I	0–5 points	1	7
Class II	6–12 points	7	11
Class III	13–25 points	14	38
Class IV	> 26 points	78	No patients operated on

Sources: Goldman *et al.* (1977); Jeffrey *et al.* (1983).
Depending upon the number of points, patients can be allocated to classes (I to IV) and the risk of complications can be estimated.

Table 5 Multifactorial index of cardiac risk (after Goldman et al. 1977; Goldman 1987)

Another index, developed by Detsky and colleagues, takes coronary arterial disease into account more fully, considering not only recent myocardial infarction (within 6 months) but also older infarcts and severe angina (Table 6). Patients with a low score (0–5) have below-average risk, while higher scores predict an above-average risk of cardiovascular complications.

Coronary heart disease	
Myocardial infarction within 6 months	10
Myocardial infarction more than 6 months	5
Angina class 4 (Canadian Cardiovascular Society)	20
Angina class 3 (Canadian Cardiovascular Society)	20
Unstable angina within 3 months	10
Cardiogenic pulmonary oedema	
Within 1 week	10
Ever	5
Valvular disease	
Suspected critical aortic stenosis	20
Dysrhythmias	
Any rhythm other than sinus rhythm	5
More than 5 premature ventricular beats/min at any time prior to surgery	5
Poor general status	
Age over 70	5
Emergency operation	10

Table 6 Cardiac risk index (after Detsky et al. 1986)

Global indices of risk are useful but have their limitations, as the severity of heart disease is often underestimated by pure clinical assessment. More detailed preoperative assessment, including functional evaluation of ventricular function, is becoming an integral part of the risk assessment in patients undergoing major surgery.

For patients undergoing vascular surgery, L'Italien and colleagues have developed a risk index based on clinical findings. Patients' individual scores are calculated, based on age, angina, history of myocardial infarction, diabetes, history of heart failure, and previous coronary arterial bypass surgery. The predictive value of this risk index is 74 per cent and is improved to 81 per cent when the results of myocardial scintigraphy are included (Table 7).

Clinical risk factors	*Weights*
Age (> 70 years)	0.5
Diabetes	0.8
Angina	0.8
Congestive heart failure	0.6
Myocardial infarction	0.5
Prior CABG (< 5 years)	-2.2

CABG, coronary arterial bypass graft.
Patients' scores are the sum of the weights of the clinical factors. Scores below 1.0 indicate a risk lower than the standard institutional complication rate; scores between 1 and 2 indicate a risk up to 2.5 to 3 times the standard institutional complication rate; scores above 2.5 indicate a risk 3 to 5 times above the standard institutional complication rate.

Table 7 Risk index in vascular surgical patients (after L'Italien *et al.*, 1996)

Further investigations

Chest radiograph

Although the cost-effectiveness of the routine chest radiograph has been questioned, it provides important information in patients with heart disease. Diagnostic alterations of the cardiac shadow, such as prominent left atrium or marked left-ventricular hypertrophy, may be observed in valvular heart disease. Cardiomegaly (heart shadow greater than 50 per cent of the diameter of the thorax) in patients with coronary arterial disease is usually associated with a low ejection fraction, and substantial enlargement of the left ventricle in hypertensive heart disease is associated with poor diastolic ventricular function. Pulmonary congestion and the presence of pleural effusion may also indicate impaired cardiac function. Therefore the routine chest radiograph may provide useful functional information. In addition, hitherto unrecognized lung tumours or aneurysms of the thoracic aorta may be revealed, and their presence, or that of other intrathoracic lesions, may alter the overall management of patients presenting for non-thoracic surgery.

Electrocardiography

Although the resting electrocardiogram is normal in 25 to 50 per cent of patients with coronary heart disease, electrocardiography remains a very important preoperative test in these patients. Characteristic patterns associated with ischaemia, injury, and infarction are easy to recognize (Fig. 2). When Q waves extend to a large area on the precordial leads, left ventricular function is likely to be substantially reduced. Abnormalities of the ST segment or T wave suggestive of myocardial ischaemia are associated with a three-fold increase in perioperative ischaemia. Patients with coronary arterial disease who are admitted for elective surgery following previous investigations should undergo a repeat electrocardiogram: totally silent myocardial infarction often occurs within the days preceding admission. Overlooking a very recent infarction is a most serious hazard.

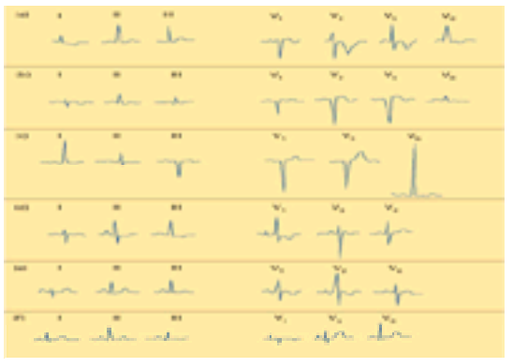


Fig. 2. Typical electrocardiographic patterns of (a) myocardial ischaemia, (b) myocardial infarction, (c) left ventricular hypertrophy, (d) right ventricular hypertrophy, (e) left atrial hypertrophy, and (f) right atrial hypertrophy.

Characteristic patterns of ventricular or atrial hypertrophy are seen in valvular heart disease, and in severe arterial hypertension (Fig. 2).

Holter monitoring

Continuous ambulatory electrocardiographic recording is of considerable value in the diagnosis of dysrhythmias. Short episodes of tachyarrhythmias, including ventricular tachycardia, may indicate that the dysrhythmia is potentially life-threatening and may warrant drug therapy before anaesthesia and surgery. Similarly, in the presence of conduction disorders, continuous electrocardiographic monitoring may reveal episodes of complete heart block demonstrating the severity of what may otherwise be considered a relatively benign conduction disorder.

Over the past decade the use of continuous ambulatory monitoring has revealed that many episodes of myocardial ischaemia are completely silent. Silent ischaemia accounts for 75 per cent of ischaemic episodes in patients with stable angina and 90 per cent in those presenting for coronary arterial bypass grafting. As silent ischaemia is an adverse factor in both stable and unstable angina, 24-h electrocardiographic monitoring is becoming an increasingly important preoperative test in patients with coronary heart disease presenting for surgery. Postoperatively silent myocardial ischaemia is a major predictor of adverse cardiovascular outcomes. As ischaemia often precedes such events by several hours, the use of postoperative electrocardiographic monitoring may become more widespread, especially in patients who have undergone vascular surgery. Continuous electrocardiographic monitoring with ST-segment trending may allow the treatment of ischaemia to be started before serious complications occur. However, prevention by the use of b-blockers may be preferable.

Exercise electrocardiography

Exercise electrocardiography has a high specificity in the prediction of coronary disease and gives an indication of the coronary reserve. Exercise-induced ischaemia usually occurs in the territory supplied by coronary arteries that are moderately or severely obstructed, or that develop vasospasm during exercise. There is a good correlation between poor exercise capacity, positive ischaemic response, and perioperative cardiac complications in patients undergoing vascular surgery. In those with intermittent claudication, the exercise test relies on the upper limbs.

Echocardiography

M-mode echocardiography represents a one-dimensional view of the heart plotted against time, while two-dimensional echocardiography enables evaluation of an entire sector of single-dimensional beams. The latter makes it possible to obtain estimates of muscle mass, ejection fraction, velocity of fibre shortening, and end-diastolic and end-systolic volumes. Segmental abnormalities of wall motion due to myocardial ischaemia can be described qualitatively and quantitatively. Images in multiple planes are necessary to obtain accurate estimations of ejection fraction and volumes.

Both the extent of left ventricular hypertrophy and the degree of diastolic dysfunction can be determined in patients with arterial hypertension. After myocardial infarction the size of ventricular aneurysms can be assessed, as can abnormalities of papillary muscles and the presence of mural thrombi. In valvular heart disease the anatomy of the valves can be examined. Other conditions, such as pericardial effusion, the presence of clots after cardiac surgery, and atrial myxoma, can be detected.

Advances in Doppler technology have made it possible to measure flow across the heart valves; forward and regurgitant flows can be estimated, and pressure gradients across stenoses may be calculated. Such measurements correlate well with direct measurements obtained at cardiac catheterization.

The introduction of transoesophageal echocardiography (the transducer being attached to the distal end of a gastroscope) has made it possible to obtain high-quality images during anaesthesia and surgery. In awake patients, high-quality images of dissection of the thoracic aorta, prosthetic valve dysfunction, intracardiac thrombi, cardiac and paracardiac masses, and acute mitral regurgitation can be obtained. This technique is particularly useful when standard echocardiography is inconclusive because of chest trauma.

Exercise echocardiography is used in the diagnosis of coronary arterial disease as exercise can induce regional myocardial ischaemia, which is characterized by reduced, absent, or paradoxical wall motion. The ejection fraction may also decrease during exercise. All these features indicate that exercise has induced reversible ischaemia in an area supplied by narrowed coronary arteries. Such abnormalities indicate the need for coronary angiography, followed where appropriate by coronary angioplasty or coronary bypass surgery.

Dobutamine echocardiography is also used in the detection of coronary arterial disease. In patients who are unable to exercise, an infusion of dobutamine increases myocardial oxygen demand so that areas of compromised coronary perfusion will exhibit the characteristic features of reversible ischaemia. Further investigations and treatment are justified (see above).

Nuclear imaging

Scintillation cameras detect g-rays emitted by radiopharmaceuticals. Multicrystal cameras (dynamic studies) and single-probe detectors (ejection fraction, cardiac output, pulmonary transit time) extend the range of studies.

Myocardial imaging is used extensively to determine the presence and size of myocardial infarction and the presence of areas of poor perfusion. Hot-spot imaging relies on the avidity of infarcted segments for technetium-99m pyrophosphate. Uptake is detectable as early as 12 to 16 h after infarction, while the maximum abnormality is seen at 48 to 72 h. Cold-spot imaging relies on the uptake of thallium (^{201}Tl) being proportional to regional myocardial blood flow: areas of ischaemia, infarction, or decreased perfusion appear as cold spots. Perfusion scans may be entirely normal at rest, even in the presence of significant coronary stenoses; heterogeneity of thallium uptake may become obvious during metabolic stress or on infusion of a coronary vasodilator such as dipyridamole. Repeat imaging at rest 3 to 4 h later may show that some perfusion defects have disappeared. Reversible defects indicate transient myocardial ischaemia without infarction, and their presence is associated with perioperative ischaemia and adverse outcome in patients undergoing vascular surgery.

Gated blood-pool imaging relies on technetium-99m to label serum albumin or red cells. Following equilibrium of the radionuclide in the intravascular space, activity is counted over 300 to 500 cardiac cycles. This allows gating to a physiological marker (usually the R wave of the electrocardiogram) in order to define end-diastole. The cardiac cycle is then divided into 20 to 30 periods, so that dimensions can be estimated throughout contraction and relaxation. Global and regional ejection fractions may be calculated, as well as the extent and synchrony of contraction. Gated blood-pool imaging is regarded as the ‘gold standard' for measuring the ejection fraction. The severity of coronary arterial disease can be gauged by the changes in function observed during exercise or in response to dipyridamole or dobutamine. A reduction of the ejection fraction during exercise indicates that an area of the myocardium that was functioning normally at rest has become ischaemic and has stopped contributing to ventricular ejection. Radionuclide testing often reveals ventricular dysfunction that was essentially silent because the patient could not exercise or had simply become used to their disability.

Both dipyridamole–thallium-201 imaging (myocardial scinti-graphy) and dobutamine echocardiography are predictors of adverse outcome if a reversible defect or new wall-motion abnormalities are detected. In the presence of such abnormalities, the cardiac event rates are substantially reduced by coronary revascularization, and very high when surgery is performed without coronary revascularization. Thus, these non-invasive screening tests make a significant contribution to the overall management

of the patient, and to the evaluation of the risk:benefit ratio.

Cardiac catheterization

Cardiac catheterization allows the extent of heart disease to be delineated anatomically by the use of contrast media and the measurement of intracardiac pressure, cardiac output, and shunt fraction. The exact site of coronary arterial lesions can be identified. Coronary angiography is not without risks and should be done before non-cardiac surgery only when coronary revascularization is being seriously considered as the first step in a staged management. Recent studies have shown that approximately 25 per cent of patients with coronary arterial disease have a ‘steal prone' coronary circulation, that is, the presence of an occluded artery, a stenosed artery, and demonstrable collaterals between the two. In such patients, arterial vasodilators may cause a ‘steal' phenomenon and induce ischaemia.

Preoperative pulmonary arterial catheterization

The insertion of a balloon-tipped pulmonary arterial catheter allows left and right ventricular-filling pressures, cardiac output, and mixed venous oxygen saturation to be measured and oxygen consumption and physiological shunt to be calculated. There is a high incidence of unrecognized cardiorespiratory abnormalities, particularly in elderly people (Table 8), and the identification and accurate estimation of functional defects makes it possible to postpone surgery in order to begin further treatment, to modify the operation itself, or to cancel surgery altogether if the risks are unacceptable. Extensive published reports have evaluated the benefits of preoperative clinical and non-invasive screening for coronary arterial disease, mostly in patients presenting for peripheral vascular surgery.

Abnormality	% patients
Relative hypoxaemia (Pa_{O_2} < 8.0kPa)	44
Elevated pulmonary capillary wedge pressure (> 7.5mmHg)	46
Raised pulmonary vascular resistance (> 250dyn.s.cm^{-5}.m^{-2})	46
Abnormal venous admixture (> 9%)	82
(> 20%)	22
Increased arteriovenous O_2 content difference (> 5.2ml O_2/dl)	23
Poor left-ventricular function (below normal function curve)	22

Table 8 Physiological abnormalities in the elderly (after Del Guercio and Cohn, 1980)

Cardiovascular medication

The preoperative assessment of patients with cardiovascular disease cannot be complete without considering their long-term cardiovascular medication. Adrenergic b-adrenoceptor blockers, calcium antagonists, angiotensin-converting enzyme inhibitors, angiotensin-II antagonists, nitrates, and cardiac glycosides are commonly used and may interact with drugs given during anaesthesia or during recovery from surgery.

Adrenergic b-receptor antagonists

b-Blockers are useful in both acute and long-term therapy of patients with ischaemic heart disease, arterial hypertension, hypertrophic cardiomyopathies, and certain forms of dysrhythmias. They prevent the cardiac effects of b-adrenoceptor stimulation, thus protecting the heart against the adverse effects of tachycardia and enhanced contractility. Heart rate and cardiac output are decreased by b-blockers in proportion to the patient's pre-existing sympathoadrenal activity.

For many years the prevailing opinion was that b-receptor antagonists should be discontinued before elective surgery. However, detailed haemodynamic studies have shown that arterial pressure and cardiac output are well maintained in hypertensive patients receiving these drugs. Direct benefits of b-adrenoceptor blockade include the blunting of hypertensive responses to laryngoscopy and intubation, and a reduced incidence of dysrhythmias and myocardial ischaemia.

Prevention of tachycardia is particularly important in patients with coronary disease. b-blockers protect the ischaemic myocardium, limit the degree of myocardial infarction, and improve the functional recovery of myocardium subjected to short periods of ischaemia. Maintaining such therapy throughout the perioperative period is therefore strongly recommended. However, b-blockade also modifies some of the responses of the circulation and, if hypovolaemia occurs, the adrenergic stimulation that is normally responsible for tachycardia will be substantially blunted. Reliance on tachycardia as an indicator of the severity of hypovolaemia may therefore cause unnecessary delays in fluid replacement in these patients.

Active prevention of perioperative, silent myocardial ischaemia is possible with b-blockers, a_2-adrenoceptor agonists, and the adenosine modulator acadesine. While all reduce the incidence of silent myocardial ischaemia, only acadesine and atenolol have been shown to improve outcome. Indeed, atenolol, given for 7 days starting on the day before surgery, was shown to reduce the incidence of silent myocardial ischaemia and to improve the long-term prognosis of patients with risk factors for coronary arterial disease. At 2 years, survival was 78 per cent in patients who had received a placebo as opposed to 90 per cent in those who had received atenolol. Thus, guidelines published in 1997 on the assessment and management of patients with coronary arterial disease presenting for vascular surgery now include the advice to determine whether or not b-blockers are contraindicated in such patients. It is generally recognized that too many patients are denied the benefits of b-blockade on the strength of poorly documented contraindications.

Calcium-channel antagonists

Calcium-channel antagonists are used extensively in the management of supraventricular dysrhythmias, angina, and arterial hypertension. As calcium ions are crucial in the electrical activity of the heart and the excitation–contraction coupling of cardiac and vascular smooth muscle, calcium antagonists can be expected to cause marked alterations in the cardiovascular system. The selective inhibition of transmembrane and intracellular Ca^{2+} fluxes is responsible for the depression of sinus-node activity and atrioventricular-node conduction, the negative inotropy, and the vasodilatation that attend the administration of calcium antagonists.

Verapamil causes myocardial depression and decreases the activity of the sinus and atrioventricular nodes. It is effective in slowing the atrial rate in sinus tachycardia and is most effective in terminating re-entrant paroxysmal supraventricular tachycardias. It slows the ventricular rate in atrial fibrillation and atrial flutter. Verapamil is also used in the management of coronary heart disease, as it decreases myocardial oxygen consumption and causes coronary vasodilatation. Verapamil facilitates recovery from brief episodes of myocardial ischaemia.

Nifedipine is predominantly a vasodilator and is ineffective in the management of supraventricular arrhythmias mediated by reflex sympathetic overactivity. It has become a very important drug in the management of arterial hypertension, angina, and coronary spasm. However, short-acting preparations of nifedipine appear to increase the risk of sudden death in hypertensive patients. Long-acting (slow-release) preparations do not have this risk.

The effects of diltiazem are intermediate between those of verapamil and nifedipine. It is used mostly in the management of ischaemic heart disease, though it is also effective in the management of supraventricular dysrhythmias.

Nicardipine and nimodipine, like nifedipine, they are nitrendipine derivatives. Nicardipine is used in the treatment of angina and hypertension, while nimodipine is used mostly to prevent or treat cerebral vasospasm.

The reflex-mediated adrenergic responses to calcium antagonists are suppressed by the addition of b-blockers, and the possibility of adverse interactions between b-blockers and calcium antagonists must be recognized, particularly when they are given intravenously.

Halogenated anaesthetics reduce calcium fluxes in cardiac cells and may potentiate the negative inotropy and chronotropy of the calcium antagonists. This in no way indicates that calcium antagonists should be discontinued before elective surgery, but monitoring must be sufficiently detailed for early-warning signs of adverse effects to be detected immediately.

Calcium antagonists offer little or no protection against perioperative myocardial ischaemia. In this respect they appear to be much less effective than b-blockers, probably due to their inability to protect the heart against the effect of sympathetic overactivity. They do not prevent increases in heart rate.

Angiotensin-converting enzyme inhibitors

Angiotensin-converting enzyme inhibitors cause peripheral vasodilatation and, in patients with normal renal circulation, increase renal plasma flow, decrease renal vascular resistance, and have little effect on glomerular filtration. However, in patients with renal arterial stenosis, these drugs may precipitate renal failure, as the reduction of vascular resistance in the efferent arteriole reduces the glomerular filtration pressure.

Angiotensin-converting enzyme inhibitors are becoming increasingly important in the management of congestive heart failure because they increase cardiac output and exercise tolerance; they also prolong survival. They are also effective in controlling blood pressure in hypertensive patients. Unlike other vasodilators, they do not cause sodium retention because the renin–angiotensin–aldosterone system is blocked.

Converting enzyme inhibitors may minimize the pressor responses to intubation and surgical stimulation without causing exaggerated reductions in blood pressure in response to induction or maintenance of anaesthesia.

Angiotension-II antagonists

Several angiotensin-II antagonists are now available. Losartan, valsartan, and irbesartan offer more complete blockade of angiotensin-II activity than do converting enzyme inhibitors, which do not suppress endogenous angiotensin II produced in tissues. In congestive heart failure, losartan decreases the likelihood of fatality more than captopril. In hypertensive heart disease, angiotensin-II antagonists are effective in reducing blood pressure.

Nitrates

Nitrates cause venodilatation and, to a lesser extent, arteriolar vasodilatation: These effects are the result of their conversion to the free radical nitric oxide, which increases intracellular cyclic guanosine monophosphate causing protein-kinase phosphorylation and a fall in intracellular calcium.

These combined effects decrease left ventricular wall tension and myocardial oxygen consumption. Glyceryl trinitrate causes a reflex tachycardia that may counteract this beneficial effect, unless a b_1-adrenoceptor antagonist is also given. Glyceryl trinitrate redistributes coronary blood flow to the subendocardium by increasing collateral flow. It is effective in the treatment of unstable angina, coronary artery spasm (Prinzmetal's variant angina) and in acute myocardial infarction as a means of limiting infarct size.

Cardiac glycosides

The cardiac glycosides inhibit transport of sodium and potassium ions across cell membranes by inhibiting the Na^+, K^+-ATPase. One major effect is the release of sequestered calcium ions from the mitochondria of the failing heart, thus enhancing calcium availability in the sarcoplasmic reticulum and improving contractile force. This advantage is counteracted by a marked increase in systemic vascular resistance. Thus, in acute-care units, the combination of vasodilators and inotropes such as dopamine or dobutamine is used routinely instead of cardiac glycosides.

Digitalis preparations increase central vagal activity, delaying the atrioventricular conduction time, an effect well suited to the treatment of atrial fibrillation, atrial flutter, and occasionally the termination of supraventricular tachycardia in the Wolff–Parkinson–White syndrome.

Cardiac glycosides may accumulate in patients with poor renal function, and toxic blood concentrations may be achieved by conventional doses. This tendency may be enhanced after anaesthesia and surgery because of the alterations in renal function associated with the stress response to surgery. Digitalis toxicity may be precipitated during anaesthesia because respiratory alkalosis exacerbates the effects of hypokalemia, especially in patients taking diuretics. Nevertheless, the use of digoxin to maintain a normal heart rate (55–70 beats/min) in patients with atrial fibrillation has the advantage over all the other antiarrhythmics of having positive rather than negative inotropic effects. This may be important in patients with poor left-ventricular function.

Preoperative digitalis has been advocated to minimize the effect of potent anaesthetic agents on the heart, and to minimize the incidence of perioperative tachyarrhythmias. However, the risk of intraoperative dysrhythmias is increased. Laver and Lowenstein have advocated the following guidelines: acute digitalization is justified in the patient with preoperative heart failure; digitalis is best continued until the evening before operation in patients with chronic atrial fibrillation if the ventricular rate is in excess of 80 beats/min; preoperative digitalization may be considered for reducing the incidence of supraventricular dysrhythmias in patients undergoing abdominal or intrathoracic operations in whom a previous myocardial infarction is associated with abnormal left-ventricular function; preoperative digitalization should be discouraged if it is intended solely to counteract the cardiac depression of anaesthesia.

Conclusion

Rigorous preoperative assessment is essential in order to identify patients in whom the cardiac risks of anaesthesia and surgery are particularly high. Cardiologists and radiologists are frequently called upon to establish the nature and severity of cardiac disorders and to perform objective tests of cardiac function, as unrecognized poor cardiac function is associated with excessive morbidity and mortality. Sound management rests on communication between surgical and anaesthetic teams in order to define the safest strategy for the perioperative period. Collaboration with cardiologists is essential in order to ensure the best therapy of cardiovascular abnormalities, both pre- and postoperatively.

Further reading

Allman KG, Muir A, Howell SJ, Hemming AE, Sear JW, Foëx P. Resistant hypertension and preoperative silent myocardial ischaemia in surgical patients. *British Journal of Anaesthesia* 1994; **73**, 574–8. [The first study in which a clear association has been demonstrated between arterial hypertension and silent myocardial ischaemia in surgical patients.]

Al Wathiqui MH, Farber N, Pelc L, Gross GJ, Brooks HL, Warltier DC. Improvement in functional recovery of stunned myocardium by long-term pretreatment with oral propranolol. *American Heart Journal* 1989; **117**: 791–8. [The first experimental demonstration that prolonged pretreatment with a b-adrenoceptor blocker enhances functional recovery after transient ischaemia by an effect on heart rate.]

Barnes RW. Asymptomatic carotid disease in patients undergoing major cardiovascular operations, can prophylactic endarterectomy be justified? *Annals of Thoracic Surgery* 1986; **42:**, S36–40. [This article reviews the evidence supporting or refuting the value of prophylactic carotid endarterectomy in asymptomatic patients and concludes that the practice is not justified.]

Bourke ME. The patient with a pacemaker or related device. *Canadian Journal of Anaesthesia* 1996; **43**: R24–41. [A review of the problem associated with pacemakers.]

Braunwald E. Shattuck lecture: Cardiovascular medicine at the turn of the millennium, triumphs, concerns and opportunities. *New England Journal of Medicine* 1997; **337**: 1360–9. [This is a most lucid appraisal of the changing patterns of cardiovascular diseases and their implications.]

Brunner HR, Gavras H, Waeber B, Textor SC, Turini GA, Wauters JP. Clinical use of an orally acting converting enzyme inhibitor, captopril. *Hypertension* 1980; **2**: 558–66. [One of the landmark papers on the use of captopril.]

Carliner NH *et al.*Routine preoperative exercise testing in patients undergoing major non-cardiac surgery. *American Journal of Cardiology* 1985; **56**: 51–8. [This paper contributed to the decline in the use of exercise electrocardiographic testing in the preoperative assessment of patients evaluated before non-cardiacl surgery.]

Chung F *et al.* Calcium channel blockade does not offer adequate protection from perioperative myocardial ischemia. *Anesthesiology* 1988; **69**: 343–7. [Unexpectedly this study showed that calcium-channel blockade was less effective than b-adrenoceptor blockade in reducing perioperative myocardial ischaemia.]

Clements FM, de Bruijn NP. Perioperative evaluation of regional wall motion by transoesophageal two-dimensional echocardiography. *Anesthesia and Analgesia* 1987; **66**: 249–61. [One of the early demonstrations of the value of transoesophageal echocardiography during anaesthesia and surgery.]

Cohn PF. Silent myocardial ischemia, classification, prevalence, and prognosis. *American Journal of Medicin*, 1985; **79** (Suppl. 3A): 2–6. [A key paper drawing attention to the importance of silent myocardial ischaemia and the value of its detection in the prevention of sudden cardiac death.]

Coriat P, Harari A, Ducardonet A, Targot JP, Viars P. Risk of advanced heart block during extradural anaesthesia in patients with right bundle branch block and left anterior hemiblock. *British Journal of Anaesthesia* 1981; **53**: 545–7. [One of the few papers to have addressed the question of possible risks of local anaesthesia in patients with intraventricular or atrioventricular conduction disorders.]

Cutler BS, Leppo JA. Dipyridamole thallium 201 scintigraphy to detect coronary artery disease before abdominal aortic surgery. *Journal of Vascular Surgery* 1987; **5**: 91–100. [An early demonstration of the value of myocardial scintigraphy as a screening test before surgery of the abdominal aorta.]

Deanfield JE, Selwyn AP, Chierchia S, Maseri A, Ribiero P. Myocardial ischaemia during daily life in patients with stable angina, its relation to symptoms and heart rate changes. *Lancet* 1983; **ii**: 753–8. [The story of silent myocardial ischaemia starts with this paper, where the authors showed that only 24 per cent of ischaemic episodes resulted in angina.]

Del Guercio LRM, Cohn JD. Monitoring operative risk in the elderly. *Journal of the American Medical Association* 1980; **243**: 1350–5. [While conventional assessment led to believe that patients were not particularly at risk, physiological data derived from pulmonary arterial catheterization revealed major, often correctable, physiological abnormalities.]

Detsky AS, Abrams HB, Forbath N, Scott JG, Hilliard JR. Cardiac assessment for patients undergoing non-cardiac surgery. *Archives of Internal Medicine* 1986; **146**: 2131–4. [This paper describes a multifactorial risk index that can be easily used to assess patients undergoing non-cardiac surgery.]

Dirksen A., Kjoller E. Cardiac predictors of death after non-cardiac surgery evaluated by intention to treat. *British Medical Journal* 1988; **297**: 1011–13. [A history of cardiac failure is the major predictor of postoperative mortality for emergency surgery (hip fracture, appendicitis) when surgery occurs within a year of previous myocardial infarction.]

Fleisher LA, Skolnick ED, Holroyd KJ, Lehmann HP. Coronary artery revasularisation before abdominal aortic aneurysm surgery. *Anesthesia and Analgesia* 1994; **79**: 661–9. [A novel approach to decision-making about coronary revascularization before surgery of the abdominal aorta, adapted to local mortality rates for both procedures.]

Foëx P. Pharmacology of cardiovascularly active drugs. In Garrard C, Foëx P, Westaby S, eds. *Principles and practice of critical care*, pp. 47–60. Blackwell Science, Oxford, 1997. [A review of the properties of the most frequently used cardiovascular drugs.]

Foëx P, Howell SJ. The myocardium. *Canadian Journal of Anaesthesia* 1997; **44**: R67–76. [A review of the modern concepts of the coronary circulation and their implications for anaesthesia.]

Foster ED, Davis KB, Carpenter JA, Abele S, Fray D. Risk of noncardiac operation in patients with defined coronary disease, the coronary artery surgery study (CASS) registry experience. *Annals of Thoracic Surgery* 1986; **41**: 42–50. [The first large-scale demonstration that after coronary artery bypass surgery the cardiac risks of subsequent non-cardiac surgery are substantially reduced.]

Gifford RW, Jr. Myths about hypertension in the elderly. *Medical Clinics of North America* 1987; **71**: 1003–11. [This review highlights the dangerous misconception that, in the elderly, hypertension does not need to be treated.]

Goldman L *et al.* Multifactorial index of cardiac risk in noncardiac surgical procedures. *New England Journal of Medicine* 1977; **297:** 845–50. [The first study devoted specifically to the assessment of the cardiac risk factors in non-cardiac surgery.]

Haagensen R, Steen PA. Perioperative myocardial infarction. *British Journal of Anaesthesia* 1988; **61**: 24–37. [One of the best reviews on the question of perioperative myocardial infarction.]

Hertzer NR *et al.* Coronary artery disease in peripheral vascular patients. A classification of 1000 coronary angiograms and results of surgical management. *Annals of Surgery* 1984; **199**: 223–33. [This large study dramatically demonstrated the prevalence of significant coronary arterial disease in patients presenting for vascular surgery.]

Howell SJ, Sear YM, Yeates M, Goldacre M, Sear JW, Foëx P. Risk factors for cardiovascular death after elective surgery under general anaesthesia. *British Journal of Anaesthesia* 1998; **80**: 14–19. [This case-control study highlights the major risk factors for cardiovascular death after surgery under general anaesthesia.]

Jeffrey CC, Kunsman J, Cullen DJ, Brewster DC. A prospective evaluation of cardiac risk index. *Anesthesiology* 1983; **58**: 462–4. [In this study the substantial increase in risk that attends vascular surgery was demonstrated by comparing outcome in vascular and non-vascular surgery based on the Goldman multifactorial index.]

Knight AA, Hollenberg M, London MJ, Mangano DT. Perioperative myocardial ischemia, importance of the preoperative ischemic pattern. *Anesthesiology* 1988; **68**: 681–8. [The first study to have demonstrated that most episodes of ischaemia are silent in surgical patients undergoing coronary bypass surgery.]

Leppo J, Plaja J, Gionet M, Tumolo J, Pasakos JA, Cutler BS. Non-invasive evaluation of cardiac risk before elective vascular surgery. *Journal of the American College of Cardiology* 1987; **9**: 269–76. [This study showed that redistribution of perfusion with dipyridamole was highly predictive of serious cardiac events in vascular surgical patients.]

L'Italien GJ *et al.* Development and validation of a Bayesian model for perioperative cardiac risk assessment in a cohort of 1,081 vascular surgical candidates. *Journal of the American College of Cardiology* 1996; **27**: 779–86. [The most recent and arguably most comprehensive in the series on multifactorial risk factors for cardiac complications in vascular surgical patients.]

Mangano DT. Preoperative assessment of the patient with cardiac disease. *Current Opinion in Cardiology* 1995; **10**: 530–42. [One of the most recent and best-prepared reviews of the preoperative assessment of the cardiac patient.]

Mangano DT, Layug EL, Wallace A, Tateo I. for the Multicenter Study of Perioperative Ischemia Research Group. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. *New England Journal of Medicine* 1996; **335**: 1713–20. [To decrease short-term and long-term cardiovascular morbidity and mortality after non-cardiac surgery, i.e., the 7-day administration of a b-blocker starting the day before surgery.]

Oakley C. Aetiology, diagnosis, investigation, and management of the cardiomypathies. *British Medical Journal* 1997; **315**: 1520–4. [The most up-to-date summary on cardiomyopathies.]

Palda VA, Detsky AS. Perioperative assessment and management of risk from coronary artery disease. *Annals of Internal Medicine* 1997; **127**: 313–28. [This article summarizes the views of the American Heart Association on the perioperative assessment and management of patients with coronary arterial disease.]

Prys-Roberts C, Meloche R, Foëx P. Studies of anaesthesia in relation to hypertension. I. Cardiovascular responses of treated and untreated patients. *British Journal of Anaesthesia* 1971; **43**: 122–37. [The study that led to maintenance rather than withdrawal of antihypertensive therapy before surgery.]

Raby KE *et al.* Correlation between preoperative ischemia and major cardiac events after peripheral vascular surgery. *New England Journal of Medicine* 1989; **321**: 1296–300. [One of the major studies to have demonstrated an association between silent ischaemia and adverse cardiac outcome after vascular surgery.]

Raby KE *et al.* Long-term prognosis of myocardial ischemia detected by Holter monitoring in peripheral vascular disease. *American Journal of Cardiology* 1990; **66**: 1309–13. [This study showed that silent ischaemia is a predictor of long-term adverse outcome in patients with vascular disease and not only after surgery.]

Rao TLK, Jacobs KH, El-Etr AA. Reinfarction following anesthesia in patients with myocardial infarction. *Anesthesiology* 1983; **59**: 499–505. [Invasive monitoring and aggressive management of the circulation appear to greatly reduce the risk of perioperative reinfarction.]

Shaw LJ, Eagle KA, Gersh BJ, Miller DD. Meta-analysis of intravenous dipyridamole-thallium-201 imaging (1985 to 1994) and dobutamine echocardiography (1991 to 1994) for risk stratification before vascular surgery. *Journal of the American College of Cardiology* 1996; **27**: 787–9. [A meta-analysis reveals that the prognostic values of dipyridamole–thallium and dobutamine echocardiography are very similar.]

Sorensen MB, Engell HC. Preoperative haemodynamic evaluation of patients submitted for major surgery. *Acta Anaesthesia Scandinavica* 1978; **22**: 391–9. [This early paper shows the value of pulmonary arterial catherization in patients with atrial fibrillation or left ventricular hypertrophy presenting for vascular surgery.]

Steen PA, Tinker JH, Tarhan S. Myocardial reinfarction after anesthesia and surgery. *Journal of the American Medical Association* 1978; **239**: 2566–70. [One of the landmark studies of the risk of reinfarction when surgery is carried out shortly after previous myocardial infarction.]

Yeung AC, Barry J, Orav J, Raby KE, Selwyn AP. Effects of asymptomatic ischemia on long-term prognosis in chronic stable coronary artery disease. *Circulation* 1991; **83**: 1598–604. [Silent ischaemia is one of the strongest predictors of death or myocardial infarction.]

12.3 Renal problems

Christopher G. Winearls and Peter J. Ratcliffe

Surgery in patients with chronic renal disease
- Preoperative assessment
- Anaesthesia and analgesia
- Fluid and electrolyte balance
- Anaemia and bleeding
- Drug prescription

Acute renal failure as a complication of surgery
- Definition
- Incidence
- Aetiology and classification
- Pathophysiology (reviewed by Brady and Singer, 1995)

Management
- Diagnosis
- Immediate management and indications for urgent dialysis
- Prophylaxis and attempts at reversal
- General measures
- Renal replacement techniques
- Prognosis

Further reading

Surgeons and nephrologists share the management of patients with chronic renal failure, acute renal failure, and structural disease of the urinary tract. The presence of pre-existing chronic renal disease requires modifications to the management of patients undergoing surgery and this is discussed first. Acute renal failure is a dreaded complication of surgery: although dialysis and haemofiltration allow many patients to survive and usually recover renal function, the mortality rate remains high. Prevention and management of acute renal failure will be discussed in the second part of this section.

Surgery in patients with chronic renal disease

As the safety of surgery increases and the indications for particular procedures widen, surgery is being undertaken in a greater number of older patients and in those with pre-existing illnesses. Surgeons will therefore be operating on patients with pre-existing chronic renal impairment, either for conditions unrelated to their renal failure or with procedures related to the provision, and complications, of renal treatment (e.g. vascular access operations, parathyroid surgery, renal transplantation).

Preoperative assessment

Modification of pre- and postoperative management will depend on the severity of renal failure, which is best assessed by making an estimate of the glomerular filtration rate. This can easily be done by measuring the urinary creatinine clearance before elective surgery is undertaken, but in more urgent circumstances this opportunity may not arise. If so, the serum creatinine is the best guide to renal function, but reference to age, sex, and body mass (which determine muscle mass and creatinine production rate) is essential in relating the serum creatinine to the adequacy of renal function. An estimate of the creatinine clearance is provided by the Cockroft–Gault formula:

$$\text{creatinine clearance} = [140 - \text{age (in years)}] \times 1.23 \text{ (males) or } 1.04 \text{ (females)} / \text{plasma creatinine } (\mu\text{mol/l}).$$

NB: The formula may be inaccurate if the patient is grossly obese, oedematous, cachetic, pregnant, or has ascites.

A serum creatinine at the upper limit of the ‘normal range' (150 µmol/l) would reflect a creatinine clearance of 75 ml/min in a young male weighing 80 kg, but only 18 ml/min in an elderly female weighing 40 kg. Clearly, the implications of a further decline of 5 to 10 ml/min, which might easily arise during apparently uncomplicated surgery, will be quite different in the two patients.

The preoperative examination should include particular attention to the state of hydration: patients with renal disease have a limited capacity to regulate their salt and water excretion and are therefore particularly liable to imbalances in either direction. Clinical signs may, however, be more difficult to interpret. Tachycardia may be obscured by b-blockade and an acute reduction in systemic blood pressure may have simply brought a previously high blood pressure into the normal range. Useful signs of severe intravascular depletion are cool extremities, peripheral cyanosis, and a postural fall in blood pressure: in very sick patients this may be manifest simply by sitting the patient with their legs over the side of the bed. The jugular venous pressure will be low unless there is coincident myocardial or pericardial disease. Severe overhydration will usually be manifest, as in patients with normal renal function, by a raised jugular venous pressure, and signs of peripheral or pulmonary oedema. Pulmonary oedema is particularly serious, since in those with severe renal failure even large doses of diuretics may not be effective. In such patients, and in any maintenance dialysis patient with pulmonary oedema, emergency dialysis or haemofiltration will be required. In young patients with good cardiovascular function, overhydration is less easily detected and may be manifest as resistant hypertension without significant oedema.

Attention should be directed to the detection and assessment of cardiovascular disease, since severe coronary and hypertensive heart disease is common in patients with renal disease. Coronary disease may be present in young adults, particularly those with renal failure from diabetic glomerulosclerosis in whom myocardial infarction is sometimes painless. If major surgery is contemplated in diabetics with renal failure, an assessment by a cardiologist is advisable. Coronary angiography often reveals remediable lesions, treatment of which (by angioplasty or surgery) reduces the risk of subsequent cardiac events. Hypertension is associated with all forms of renal disease, although the prevalence varies and increases with the severity of renal failure. About 80 per cent of patients with endstage renal disease require antihypertensive treatment. The pathogenesis is poorly understood, but in severe renal failure, sodium retention is an important factor. When this is controlled by dialysis, the number of patients requiring antihypertensive treatment can be reduced to about 20 per cent.

Antihypertensive treatment should be maintained during surgery, unless intercurrent illness has produced severe hypotension. This is particularly important with drugs such as clonidine or b-blockers, the abrupt withdrawal of which may cause rebound phenomena. Hypertension is associated with increased anaesthetic risk, and since there is evidence that treatment, particularly with b-blockers, reduces the incidence of arrhythmias and myocardial ischaemia under anaesthesia, elective surgery is usually deferred when elevated blood pressure is found unexpectedly in the preoperative assessment. It is important, however, to be sure that unexpectedly raised blood pressure in patients with closely monitored renal disease is not simply a manifestation of anxiety. In dialysis patients, unexpected hypertension may be due to fluid overload, which will require removal by dialysis. It is again important to be sure of this diagnosis, since removal of excessive fluid by dialysis before surgery will increase the risk of dangerous intraoperative hypotension.

When assessing patients on maintenance dialysis much can be learnt from the dialysis records. For instance, in patients with limited cardiac reserve, fluid losses and vasodilation during haemodialysis may precipitate sudden and severe hypotension, a problem that may also be manifest during anaesthesia. Records of such problems during haemodialysis may forewarn the surgeon and anaesthetist. In addition, dialysis records should indicate pre- and postdialysis weights and blood pressures, which will be of importance in assessing preoperative fluid balance.

Clinical examination and history should usually be supplemented by simple investigations, which include a full blood count, measurement of serum creatinine and electrolytes, albumin, electrocardiogram, and a chest radiograph. A full blood count will be important in assessing the need for pre- or perioperative transfusions (see below). In acutely ill patients, comparison with previous values may indicate serious blood or other fluid loss.

The use of serum creatinine in the estimation of renal function has been described. This is a less satisfactory guide to the adequacy of dialysis, since the characteristics of dialysis membranes differ from those of the glomerular filter. For instance, in patients maintained by continuous ambulatory peritoneal ialysis a serum creatinine above 1200 µmol/l can often be well tolerated, while such concentrations in a patient on haemodialysis would usually indicate inadequate dialysis. Of much greater importance is measurement of the serum potassium: the risk of cardiotoxicity is very high when this rises above 7 mmol/l, and the concentration that is acceptable preoperatively will depend on whether the patient has renal function or is dialysis dependent, and on the type of surgery planned. For instance, minor surgery in patients with moderate chronic renal failure could proceed if the serum potassium is less than 6 mmol/l whereas in the dialysis patient undergoing major surgery a

preoperative serum potassium of less than 4.5 mmol/l should be the aim.

The electrocardiogram may reveal the presence of hypertensive or ischaemic heart disease; the detection of acute ischaemia and arrhythmias is particularly important. The chest radiograph will show the cardiac diameter. A sudden increase in diameter reflects serious overhydration, but may also indicate acute pericardial or myocardial disease. Pulmonary oedema requires prompt treatment.

Analysis of arterial gases will be helpful in selected patients. In the presence of severe anaemia, arterial hypoxaemia is difficult to detect clinically and is also more serious. Patients with acute surgical pathology and severe renal disease will often have an important metabolic acidosis with respiratory compensation. In this state, sudden and often inadvertent changes in respiration, consequent on sedation, anaesthesia, or analgesia, may cause large changes in arterial pH and consequently the plasma potassium.

Anaesthesia and analgesia

Modifications of anaesthetic technique and drug prescription will depend on the severity of renal failure, the presence or absence of disease in other organs, and the type of surgery. In the main these issues are dealt with elsewhere, but certain simple principles should be understood by all involved in the care of the patient.

The insertion of cannulas into forearm veins for intravenous infusions will damage potential sites for vascular access, and this is a very serious consideration in any patient who is, or may become, dependent on dialysis for life support. It is particularly important to avoid damage to the cephalic venous system. Intravenous infusions should be sited in veins outside the forearm or on the ulnar side of the hand.

Premedication

Delayed gastric emptying is common in patients with serious renal disease, particularly when renal disease complicates diabetes and severe autonomic neuropathy is present. Premedication with agents such as metoclopramide may be helpful in reducing the risk of gastric aspiration. Most renal patients are all too aware of the increased risk they face during intercurrent illness and anxiolytic drugs such as benzodiazepines will often be helpful. These may be given orally, as increased bleeding times in patients on haemodialysis are a relative contraindication to intramuscular agents. Opiate analgesia should be used with care (see below).

Monitoring

If substantial fluid loss is expected, particularly in patients with known cardiovascular instability, central venous pressure should be monitored. Blood pressure is best monitored using an automated sphygmomanometer: intra-arterial monitoring is rarely required, but if it is considered essential, the dorsalis pedis artery is the site of first choice in order to avoid damage to vessels that may be required later for creation of vascular access. Where anaemia and cardiorespiratory disease coexist, the use of pulse oximetry to ensure maintenance of optimal blood oxygenation will be helpful.

Induction of general anaesthesia

Anaesthesia may result in sudden hypotension requiring resuscitation with intravenous fluid or even pressor agents. Care must be taken to avoid overhydration, particularly in patients on haemodialysis in whom preservation of renal function is not an issue and excess fluid will require dialysis or haemofiltration for removal. The risk of hypotension is also present during spinal and epidural analgesia, and in these circumstances it may be even more difficult to control. Even in patients without renal disease, renal blood flow is usually reduced and autoregulation is impaired under surgical anaesthesia. A variety of renal vasoconstrictor mechanisms contribute to this, including activation of the sympathetic nervous system and the renin–angiotensin system. An important defence is provided by the vasodilatory action of renal prostaglandins (I_2 and E_2). In patients with significant renal disease, this mechanism assumes greater importance and prostaglandin synthetase inhibitors should not be used in the perioperative period.

Postoperatively

The most serious, and potentially fatal immediate postoperative complication is respiratory depression. With the use of modern muscle relaxants such as atracurium, which are rapidly cleared by mechanisms independent of renal function, the problem of recurarization due to the action of muscle relaxants outlasting their antagonist should not, in theory, occur. Nevertheless, respiratory depression, and even respiratory arrest, is a very important and rather unpredictable risk after major surgery in patients with severe renal failure. Prolonged and excessive action of sedative and analgesic drugs, and impaired gas exchange arising from undiagnosed pulmonary oedema, may contribute.

The action of most opiate analgesics is prolonged in renal failure. This is particularly striking in the case of morphine and probably arises from retention of active metabolites such as morphine 6-glucuronide. Potentially fatal respiratory depression may also unpredictably complicate the use of relatively modest doses of supposedly milder agents such as dihydrocodeine and dextropropoxyphene. Probably the most predictable agent is pethidine, although retention of the pethidine metabolite, norpethidine, has been reported to account for symptoms of neuromuscular excitability. Facilities for the continuous observation of these patients in the first 2 h after surgery are therefore essential.

Fluid and electrolyte balance

Patients with chronic renal failure but otherwise good health generally maintain sodium and water balance until the glomerular filtration rate is reduced very severely (below approximately 10 ml/min). This is achieved by a corresponding large increase in the proportion of the filtered sodium and water that is excreted. Similarly, to produce a given change in excretion of sodium and water at low glomerular filtration rate requires a magnification of the tubular response: for instance, a response that is effected by kidneys with a glomerular filtration rate of 120 ml/min from a change in fractional excretion of sodium from 0.5 to 3 per cent will require a change from 5 to 30 per cent if the filtration rate is only 12 ml/min. It is therefore not surprising that one of the earliest effects of chronic renal disease is a limitation in the power of the kidney to compensate for changes in sodium and water intake.

In healthy individuals with a normal solute intake, water excretion can be varied from approximately 20 ml to 1500 ml/h; in renal disease this range is reduced in both directions. The reduction in capacity for water excretion will, in most cases, be very much greater than the reduction in glomerular filtration rate, since during surgery other factors, such as non-osmotic release of antidiuretic hormone, impair water excretion. Many patients with severe renal disease will be at risk of water intoxication from commonly prescribed postoperative regimens that include 2 to 3 litres of 5 per cent dextrose/day. Since the precise limitations of renal compensation are difficult to predict, the problem can only be avoided by strict attention to fluid balance, corroborated by daily weighing of the patient.

In some surgical settings, such as the relief of urinary obstruction, and after renal transplantation, massive diuresis and natriuresis may be encountered. In these patients, it may be necessary to adjust the rate of intravenous replacement on an hourly basis, and also to adjust the sodium concentration of the replacement fluid regularly in the light of urinary sodium concentration.

In dialysis patients without residual renal function, great care is required to maintain fluid and electrolyte balance. Patients should be dialysed before surgery, but because of increased arrhythmias and haemodynamic instability immediately after dialysis, where possible, an interval of a few hours should be allowed before induction of anaesthesia. Avoidance of hyperkalaemia is of paramount importance, and can sometimes be difficult, even in patients with only mild renal disease. In general, as with sodium and water, although potassium balance is maintained in chronic renal disease, the capacity for excretion is severely limited. This may be particularly serious in diabetic patients, who are liable to develop hyporeninaemic hypoaldosteronism that further reduces potassium excretion. In patients with significant renal disease, potassium supplements and potassium-sparing diuretic agents should only be used when serum potassium is low or declining.

In dialysis patients, serum potassium should be measured before dialysis is terminated and should be reduced below 4.5 mmol/l. In dialysis patients known to have difficulty in maintaining potassium balance, provision for increased preoperative dialysis should be made and a calcium resonium enema should be given before surgery.

Anaemia and bleeding

Patients with severe renal failure are almost invariably anaemic, but severe anaemia in patients presenting for surgery should be less frequent now that recombinant erythropoietin is available. Anaemia is usually well tolerated, except in patients with coexisting ischaemic heart disease, and major surgery, such as for renal transplantation, is feasible, if not advisable, in many patients with haemoglobin as low as approximately 6 g/dl. Nevertheless, a higher preoperative haemoglobin, in the range 8 to 10 g/dl, will provide a greater safety margin in the event of haemorrhage and itself reduces the bleeding time. Blood should be cross-matched for any procedure during which a brisk haemorrhage is possible and, if the surgical indication is not immediate, consideration should be given to preoperative transfusion,

particularly in patients with cardiac disease. Patients with renal failure have a bleeding diathesis, which is in part a consequence of abnormal platelet function. Such patients are liable to ooze from incisions and particularly careful attention to surgical haemostasis is required. Medically, the problem can be limited by ensuring that the patients are well dialysed, transfused to haemoglobin of 8 to 10 g/dl preoperatively, and that heparin used during dialysis has been cleared or reversed. The administration of platelets is not effective, but there are measures that can be applied in patients who are at particular risk of bleeding or who continue to ooze after surgery. Cryoprecipitate produces a temporary improvement in bleeding time. The synthetic vasopressin analogue l-deamino 8-d-arginine vasopressin (0.3 µg/kg, intravenously) is effective for about 48 h. Conjugated oestrogens provide an improvement in bleeding time that lasts up to 14 days.

Drug prescription

Special care must be taken with the use of all drugs, not only because their elimination may be slowed and their bioavailability altered in patients with renal failure, but also because they may directly or indirectly reduce residual renal function. Before prescribing it is prudent to ask whether these considerations apply and, if they do, whether prescription is essential, whether there is an alternative drug, and whether the dose needs to be modified. A list of commonly used drugs that may adversely affect renal function is given in Table 1.

Drug	*Mechanism*
Antimicrobials:	Tubular damage
Aminoglycosides	
Amphotericin B	
Pentamidine	
ACE inhibitors	Reduce GFR
Corticosteroids	Catabolism
Cyclosporin	Reduce GFR
Cytotoxics	Hyperuricaemia
Diuretics	Hypovolaemia
NSAIDs	Reduce GFR
Tetracyclines	Catabolism
Vitamin D analogues	Hypercalcaemia

GFR, glomerular filtration rate; NSAID, non-steroidal anti-inflammatory drug.

Table 1 Drugs that may aggravate renal failure

Because the elimination of many drugs is dependent on renal function, the dose and frequency of administration must be changed in patients with renal failure (Table 2). In a patient receiving maintenance dialysis the mode of dialysis also has an important effect: for example, haemodialysis removes gentamicin from the circulation very rapidly so that it should be administered at the end of dialysis, following which high concentrations will persist in the plasma until the next treatment. In contrast, peritoneal dialysis will remove gentamicin continuously at a low rate.

Class	*Comment*
Antimicrobial	
Aminoglycosides	Reduce dose; ototoxic
Acyclovir	Reduce dose
Cephalosporins	Reduce dose
Co-trimoxazole	Reduce dose
Ethambutol	Reduce dose or avoid; retinal toxicity
Gancyclovir	Reduce dose
Nitrofurantoin	Avoid; neuropathy
Penicillins	Avoid very high doses; fits
Vancomycin	Reduce dose; ototoxic
Anaesthetic drugs	
Pancuronium	Avoid; prolonged paralysis
Gallamine	Avoid; prolonged paralysis
Opiates	Reduce dose; prolonged effect
Gastrointestinal	
Metoclopramide	Be aware of extrapyramidal effects
H_2 antagonists	Reduce dose
Magnesium salts	Monitor magnesium concentration
Cardiac	
Digoxin	Reduce dose and frequency
β blockers	May need to reduce dose
Spironolactone, amiloride	Avoid; hyperkalaemia
Angiotensin converting enzyme inhibitors	Monitor creatinine and K^+
Clofibrate	Avoid; muscle injury
Oral hypoglycaemics	
Chlorpropamide	Avoid; hypoglycaemia
Tolbutamide	Reduce dose
Biguanides	Avoid; lactic acidosis

Table 2 Drugs to be prescribed with special care in renal failure

Acute renal failure as a complication of surgery

Definition

Acute renal failure following surgery may be defined operationally as ‘a reduction in renal function sufficient to cause clinical problems', to distinguish it from the very much more common occurrence of acute renal functional impairment, often recognized simply by a rise in the plasma creatinine. A reduction in excretory capacity may occur against a background of pre-existing chronic renal functional impairment, when it is referred to as ‘acute on chronic renal failure'. In this condition a relatively minor reduction in renal function can cause clinically important problems.

Incidence

The annual incidence of severe acute renal failure defined as a rise in creatinine to more than 177 µmol/l, 2 mg per cent, is approximately 200 per million population per year. In about one-third it is a complication of either surgery (most often cardiac, vascular, and abdominal) or trauma; one-third of all cases will require renal support in the form of dialysis or haemofiltration. The remainder have transient renal impairment, which can be managed without recourse to dialysis, but which requires careful attention to water and electrolyte balance.

The Leeds experience of acute renal failure from 1984 to 1995 is that the proportion of patients with a surgical cause is falling. Although this explained in part by a rise in the number of medical cases, there is an impression that the absolute risk of renal failure after surgery is lower because of better pre- and postoperative care. This will be difficult to prove because surgery is now being offered to patients with greater comorbidity.

Prospective audit of surgical practice has defined circumstances associated with a particularly high risk of acute renal failure. There is, for example, a 20 per cent incidence of acute renal failure following rupture of an abdominal aortic aneurysm compared to only 3 per cent following elective repair. About 10 per cent of patients undergoing hepatobiliary surgery develop renal impairment but only 3 per cent require renal support. Two other high-risk states are acute pancreatitis and extensive burns: about 20 per cent of patients with greater than 15 per cent burns and 4 per cent of those with acute pancreatitis develop acute renal failure, and in both circumstances the mortality rate is high.

Aetiology and classification

The causes of acute renal failure are generally classified according to whether they act at prerenal, renal, or postrenal sites (Table 3).

1. *Hypoperfusion* (prerenal)
Shock:
Hypovolaemic
Cardiogenic
Septic
Renal arterial disease:
Aortic dissection
Embolism
2. *Renal injury* (renal)
Acute tubular necrosis (following 1):
Toxic
Drugs—myoglobin, haemoglobin
Glomerulonephritis—associated sepsis
Interstitial nephritis—antibiotics
3. *Obstruction* (postrenal)
Bilateral ureteric:
Retroperitoneal fibrosis
Lymphoma
Malignancy—bladder, cervix
Unilateral ureteric:
Stones
Papilla
Tumour
Bladder outflow:
Prostatic hyperplasia or carcinoma
Urethral stricture

Table 3 Aetiology of acute renal failure in surgical practice

Prerenal failure arises in an otherwise healthy kidney as a simple consequence of hypoperfusion. This can be a consequence of volume deficit, cardiac malfunction, haemodynamic imbalance (as, for example, in septic shock, hepatic failure and inappropriate use of angiotensin-converting enzyme inhibitors), or anatomical (e.g. bilateral renal arterial stenosis; Fig. 1). In general surgical practice it most commonly arises from hypovolaemia or sepsis, in contrast to cardiac surgery where the cause is more likely to be left ventricular dysfunction or pericardial tamponade.

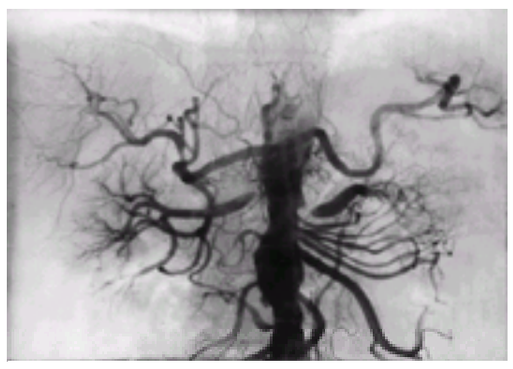


Fig. 1. Aortogram showing bilateral renal arterial stenosis; this patient developed acute renal failure following coronary artery bypass grafting (by courtesy of Dr E.W.L. Fletcher FRCR).

Postrenal causes are those of obstruction to urine flow. For obstruction to cause acute renal failure it must either be bilateral or affect a single functioning kidney. Bilateral ureteric obstruction is most often caused by malignant disease involving both ureters or by retroperitoneal fibrosis. Figure 2 illustrates this problem and its management. Bladder-outflow obstruction is usually a result of prostatic disease or urethral stricture. Renal stones and sloughed papillae cause acute on chronic renal failure when they obstruct the only or better functioning of two kidneys.

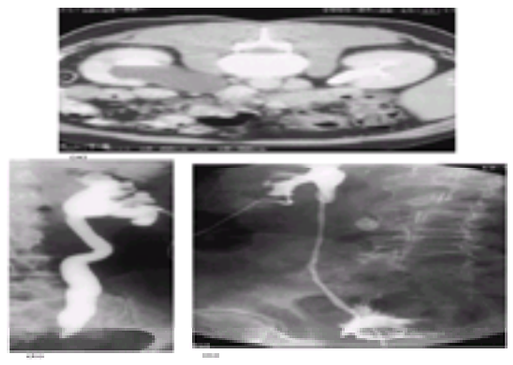


Fig. 2. (a) Computed tomogram demonstrating right-sided hydronephrosis; (b) antegrade pyelogram following percutaneous nephrostomy in a patient with malignant obstruction to the distal ureter, demonstrating hydronephrosis and dilatation of the distal ureter; (c) a double-J stent relieving malignant obstruction after percutaneous nephrostomy and antegrade placement; [(a) and (c) by courtesy of Dr Nigel Cowan FRCR].

Renal causes

Acute renal failure arising from intrinsic damage to the kidney is commonly due to ‘acute tubular necrosis'. Definition is difficult since the pathogenesis in man is poorly understood, but this form of renal disease is generally associated with a severe haemodynamic disturbance or nephrotoxic exposure; acute renal failure is presumed to arise from tubular damage. The term ‘necrosis' is widely used but inaccurate, since frank necrosis of tubular cells is not usually striking when tissue is examined histologically.

In most clinical settings, the risk of acute tubular necrosis is rather unpredictable: a particular reduction of systemic blood pressure, or loss of intravascular volume, cannot be precisely related to the risk of acute renal failure. Following exposure to potential nephrotoxins such as aminoglycoside antibiotics, haem proteins and radiographic contrast media, the risk of acute renal failure is usually low, not clearly dose dependent, and more obviously related to the coincidence of other potentially damaging influences. Many of the agents commonly listed as causing acute tubular necrosis might therefore more accurately be termed risk factors. Risk will also depend on the patient's underlying condition (Table 4): those with conditions such as heart failure, chronic liver disease, obstructive jaundice, chronic hypertension, or diabetes, in which there are pre-existing abnormalities of renal haemodynamics, will be at greater risk from a particular insult. The type of surgery is also important: operations or conditions associated with sepsis or severe tissue injury (for instance after trauma, burns, or major vascular occlusion) are associated with an increased risk of acute tubular necrosis.

Congestive heart failure
Chronic hypertension
Chronic liver disease
Obstructive jaundice
Systemic sepsis
Exposure to aminoglycosides, radiocontrast media
Renovascular disease
Chemotherapy for malignant disease
Multiple myeloma
Diabetes mellitus

Table 4 Conditions predisposing to the development of acute renal failure

Patients with acute renal failure caused by glomerular or interstitial disease usually present to physicians but may be referred to urologists when the striking symptoms are of haematuria or renal pain.

Pathophysiology (reviewed by Brady and Singer, 1995)

Since the kidney has a high resting blood flow, a low arterial–venous oxygen difference, and good autoregulatory capability it is not immediately clear why it should be susceptible to hypoperfusion injury. Regional disparities in oxygenation are well recognized and may leave certain regions at risk despite high overall blood flow, or renal perfusion may suffer an unusually severe reduction in certain forms of haemodynamic shock. The risk of renal injury is strongly dependent on the type of shock, being high in septicaemic shock and low in simple haemorrhagic shock (e.g. following gastrointestinal haemorrhage). The most severe threat to renal perfusion probably arises from multiple interactions such as activation of vasoconstrictor mechanisms, including the sympathetic nervous system and renin angiotensin system, loss of compensatory vasodilation such as by inhibition of prostaglandin production, and vascular injury itself, which may complicate septicaemia and endotoxaemia.

For some risk factors, a direct nephrotoxic action is clear. These include exposure to heavy metals such as *cis*-platinum, organic solvents, polyene antibiotics, amphotericin B, and prolonged administration of aminoglycoside antibiotics. Other precipitating factors such as myoglobin, haemoglobin, and radiographic contrast media have a clear association with acute renal failure, but renal injury is by no means invariable, and the mechanism of damage is not yet adequately explained.

Several mechanisms have been proposed to explain the near complete loss of renal function associated with acute tubular necrosis (Table 5). Tubular injury may prevent function directly, either by luminal blockage or by permitting back leakage of filtrate; or filtration itself may be reduced, either primarily, or secondary to feedback signals arising from impaired tubular transport or raised luminal pressures. Since these mechanisms may all operate at different times or may coincide and interact, it has been impossible to obtain a simple unifying explanation, even for the relatively simple animal models of acute renal failure.

Glomerular (either primary or via feedback from tubular signals)
Reduced perfusion
Reduce ultrafiltration coefficient (probably arising from reduced capillary surface area)
Tubular
Back-leakage of filtrate across damaged epithelium
Obstruction by intraluminal debris or compression of the tubules by swollen cells

Table 5 Mechanisms of excretory failure in acute tubular necrosis

The most immediate consequence of an abrupt loss of renal function is, of course, retention of the waste products of metabolism. The rate of accumulation is dependent not only on the severity of renal failure but on the rate of production. Increased catabolism is observed postoperatively and is particularly severe in patients with sepsis, burns, or trauma. Acute renal failure in this setting is marked by particularly rapid rises in urea, creatinine, urate, phosphate, acidaemia, and most importantly potassium. Massive muscle injury, which is not always traumatic, and may for instance complicate sepsis, ischaemia, prolonged coma, unusually strenuous exercise, or alcoholic intoxication, will lead to the particularly rapid onset of dangerous hyperkalaemia, hypocalcaemia, and hyperphosphataemia.

Management

Diagnosis

Acute renal failure is usually detected by a rise in the plasma creatinine or urea, or as a fall in urine output. Normal urine output alone can be falsely reassuring, since many patients with acute renal failure are not oliguric; it is essential to monitor urea and creatinine in any patient at risk of developing postoperative renal failure.

Essential early diagnostic steps are the exclusion of urinary obstruction, the exclusion or treatment of prerenal failure, and the detection of pre-existing chronic renal disease. In each case, assessment of the clinical history is vital. Urethral obstruction should be excluded immediately by palpation of the bladder.

Exclusion of obstruction

The risk of ureteral obstruction depends on the underlying disease, being increased after pelvic or retroperitoneal surgery for malignancy, or in the presence of a single kidney. It may lead to renal pain or to the classical pattern of alternating polyuria and oliguria. Ultrasonography is the key investigation and will detect almost all cases, with occasional failures when done early after acute obstruction or when the urinary system is encased with malignant tissue. Figure 2 illustrated this problem and its management.

Recognition and correction of prerenal failure

The possibility of prerenal failure should be apparent from clinical examination. Although systemic blood pressure will not always be low in a supine patient, there will generally be evidence of impaired circulation, such as cool extremities and postural hypotension. If external fluid losses are not apparent, the possibility of fluid sequestration, sepsis, or cardiac dysfunction must be considered. The diagnosis of prerenal failure is made by restoration of renal function when the haemodynamic problem is corrected.

The type of replacement fluid (blood, colloid, or crystalloid) should ideally reflect what has been lost. In circumstances where this is not clinically apparent or where blood is not available immediately, circulatory resuscitation should be commenced with isotonic crystalloid or colloid solutions. Controversy still surrounds the arguments for and against crystalloid or colloid solutions; in most circumstances, either is acceptable. Colloid solutions are more expensive and have the potential to cause anaphylactoid reactions, but they are confined, at least partially, to the vascular compartment for a short period of time and can increase plasma oncotic pressure. Although the action on Starling forces may be slight and short-lived, it may be significant in patients with coincident lung or cardiac disease who are on the verge of developing pulmonary oedema. Our practice is to use isotonic crystalloid solution in most circumstances and reserve the use of colloid solutions for resuscitation of shocked patients and those in whom intravascular depletion is combined with interstitial oedema. Colloid solutions are either albumin-containing preparations such as plasma protein fraction or carbohydrate polymer solutions such as Haemaccel. The safety and efficacy of human serum albumin has recently been questioned after a meta-analysis of their use in patients with burns showed a higher mortality. Because of its rapid availability and freedom from the vasodilatory effects sometimes seen with plasma protein fraction, Haemaccel is very suitable until blood becomes available, and up to 1500 ml may be infused. Some concern has been raised by the reporting of acute renal failure apparently occurring as a consequence of dextran infusions. The risk is small and may not be present with all carbohydrate polymers, but it is wise to use plasma protein fraction if large volumes of colloid replacement are to be given.

The aim of correcting the deficit as precisely and rapidly as possible is best achieved by administering fluid rapidly, but in relatively small volumes: about 250 to 500 ml should be given over approximately 30 min, and central venous pressure should be measured accurately after each portion of fluid. As vascular capacitance is filled the central venous pressure will remain low, but it will rise rapidly once repletion is achieved, and 250-ml portions of fluid are appropriate once any shift has been discerned. In some states, such as severe haemorrhage, venoconstriction will be important. This will relax as volume replacement proceeds and may lead to a further fall in central venous pressure, which requires differentiation from continued fluid loss. If renal function does not immediately improve, great care must be taken to avoid overhydration. Once volume replacement is achieved it is common to give an intravenous dose of either mannitol (10–20 g) or frusemide (furosemide) (40–80 mg) to promote a diuresis. This manoeuvre is justified on several grounds. First, it provides reassuring evidence that renal perfusion has been re-established somewhat sooner than might otherwise be the case. Secondly, these agents might actually reduce the risk of acute tubular necrosis (see below). If in a volume-replete patient these manoeuvres do not induce a diuresis of at least 40 ml in the following hour, then ‘acute tubular necrosis' has probably become established.

The management of the volume-replete patient with prerenal oliguria is much more difficult and should ideally be undertaken in the intensive therapy unit, for it will

require trials and titration of inotropic agents guided by continuous blood, central venous, and pulmonary capillary wedge pressures.

Much is made of the importance of urinary electrolyte measurements in the assessment of the oliguric patient. Classically, in prerenal failure, tubular function should be intact and is reflected by a urine specific gravity of above 1.030 and an osmolality greater than 400 mosmol/kg H_2O, a urine:plasma urea ratio above 7, and a fractional excretion of sodium of less than 1 per cent. The distinction is often blurred by diuretic therapy and pre-existing renal disease, and is not a great help in management. The overriding concern is to achieve prompt and precise correction of the haemodynamic problem, something that will often require a period of continuous bedside medical attendance.

Recognition of pre-existing chronic renal failure

The possibility or pre-existing renal failure should have been excluded in the preoperative assessment, but if not may be suspected from a history of renal disease or hypertension, unexplained anaemia, or when reduced kidney size is demonstrated by ultrasonography. If these exclusions are made, renal failure develops postoperatively in a setting of recognized risk, and the urine deposit is not very cellular, a presumptive diagnosis of acute tubular necrosis can be made. In other circumstances, for instance if renal failure is found on admission, diagnostic assessment of the full range of renal diseases will be required. History should involve careful enquiry about drug or toxin exposure. Examination should include a search for cutaneous signs of vasculitis and microscopic examination of fresh urine looking for red cells, red-cell casts, white cells, and bacteria.

Immediate management and indications for urgent dialysis (Table 6)

1. Hyperkalaemia (reviewed by Allen 1993, Further Reading)
(1) Calcium gluconate 10 ml intravenously
(2) Soluble insulin (Actrapid) 10 units with dextrose 50 per cent 50 ml intravenously as an injection
(3) Salbutamol 5 µg/kg intravenously over 15 min (particularly useful in children)—rapid onset of action and lasts approx. 6h
(4) Calcium resonium 30g by mouth or rectally
(5) Sodium bicarbonate 4.2 per cent, 50 ml
(6) Dialysis—preferably haemodialysis
2. *Pulmonary oedema*
(1) Sit the patient up
(2) Oxygen by face mask ± continuous positive airway pressure/ventilator
(3) Frusemide (furosemide) 250mg intravenously
(4) Venodilator intravenously, glyceryl trinitrate 10–200 µg/min
(5) Morphine 5mg intravenously if CO2 / respiratory drive not depressed
(6) Remove fluid by venesection, haemofiltration, or peritoneal dialysis
3. *Metabolic acidosis*
(1) Sodium bicarbonate 4.2 per cent, 100 ml intravenously if sodium load can be tolerated
(2) Haemofiltration if the patient is haemodynamically unstable
(3) Haemodialysis for 2h against bicarbonate dialysate
(4) Monitor calcium as pH is corrected

Table 6 Immediate treatment of complications of acute renal failure

Renal failure is often recognized rather late, presenting the clinician with the need to treat life-threatening complications (hyperkalaemia, pulmonary oedema, and metabolic acidosis) immediately and assess the need for urgent dialysis.

Hyperkalaemia

The risk from hyperkalaemia is of sudden arrhythmic death; early electrocardiographic abnormalities are increased T-wave amplitude, while later, ominous changes are broadening of the QRS complex, and flattening and eventual disappearance of the P wave. Eventually, the electrocardiogram comes to resemble a sine wave and cardiac standstill follows. Treatment is required urgently if the serum potassium is above 6.5 mmol/l or there are electrocardiographic changes (Fig. 3).

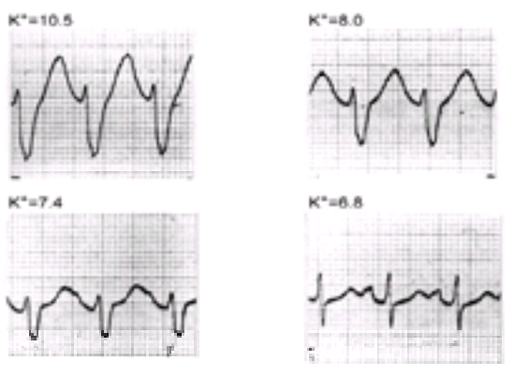


Fig. 3. Electrocardiogram demonstrating the features of severe hyperkalaemia.

The immediate management is intravenous administration of calcium. Calcium gluconate (10 ml of 10 per cent) should be given over 1 to 2 min and repeated until the electrocardiogram improves. Additional measures should be undertaken to reduce the serum potassium. In patients without serious fluid overload, hypertonic sodium bicarbonate (4.2 per cent) may be given intravenously in a dose of 50 mmol (100 ml). Theoretically, entry of potassium into cells should be promoted by correction of acidosis but the sodium load may alone be sufficient. Serum potassium may also be reduced acutely by insulin, which promotes cellular entry by stimulation of Na–K-ATPase. To prevent hypoglycaemia, glucose is given concurrently, a typical regimen being 50 g glucose intravenously with 15 units soluble insulin. Monitoring is required to detect hypoglycaemia in unconscious patients and if this cannot be provided (e.g. during a transfer) it is safer, in non-diabetic patients, to omit insulin and rely on the endogenous insulin response to glucose infusion. In children, salbutamol, 5 µg/kg intravenously, is an effective, alternative short-term (approximately 6 h) treatment of hyperkalaemia. Following control of hyperkalaemia by these measures, provision for urgent dialysis should be made unless renal function has been restored. If hyperkalaemia is resistant or recurs rapidly after dialysis, a site of severe tissue damage (e.g. rhabdomyolysis, gut infarction, limb ischaemia) should be sought.

Fluid overload

This usually manifests as pulmonary oedema (Fig. 4) and is also an immediate indication for ultrafiltration using a dialysis machine, haemofiltration, or peritoneal dialysis, but some holding measures should be put in place immediately.



Fig. 4. Chest radiograph in a patient with acute renal failure showing widespread air-space opacification consistent with pulmonary oedema (by courtesy of Dr Nigel Cowan FRCR).

A marked metabolic acidosis should be treated with sodium bicarbonate only if there is hyperkalaemia or cardiogenic shock. This will have only temporary benefit and is not a substitute for dialysis or measures to improve tissue perfusion.

Dialysis or haemofiltration should be started to prevent the dangerous consequences of acute renal failure rather than as a remedy for them. Absolute indications are: hyperkalaemia (> 6.5 mmol/l), pulmonary oedema, oliguria (< 10 ml/kg per day), metabolic acidosis, a plasma urea of more than 30 mmol/l and rising. Nephrologists and intensivists are now guided by the changes in these measures. If they are moving adversely, treatment is instituted.

Prophylaxis and attempts at reversal

Since many treatments when applied before the insult are effective in ameliorating or preventing acute renal failure in experimental models, it might be expected that prophylactic measures would be important in surgery. However, since many of the measures, such as the use of vasodilators, may exacerbate hypotension, they are difficult or dangerous to apply in clinical practice. Prophylactic administration of dopamine, mannitol, or frusemide (furosemide), coupled with intravenous fluid sufficient to maintain or even expand the plasma volume, will almost certainly reduce the incidence of acute renal failure after high-risk procedures such as aortic surgery and surgery in jaundiced patients. Of these measures, the most important is maintenance of blood volume. The use of prophylactic mannitol is also established practice in aortic surgery, a suitable regimen consisting of infusion of 10 g as a hypertonic 20 per cent solution after the induction of anaesthesia followed by infusion at a rate of 10 g/h during the procedure.

In patients with established acute tubular necrosis, such measures can usually increase the urine output moderately, but there is no evidence that they lessen the severity or duration of renal failure or that they improve survival. However, when renal failure is first detected it is difficult to know whether any reversible element exists. In this case it is usual to attempt to improve renal perfusion and urine flow by the intravenous administration of dopamine (1–2 µg/kg per min) and frusemide (5–10 mg/min) for 1 to 2 h. This practice has been questioned after it was shown, in a double-blind, placebo-controlled trial, that the administration of loop diuretics, although increasing urine flow, made no difference to renal recovery, need for dialysis, or mortality.

A number of approaches to hastening the resolution of acute renal failure are being tested both in the laboratory and the clinic. These include the use of analogues of atrial natriuretic peptide, integrin blockers and mitogens such as insulin-like growth factor I, which would act on various of the pathogenetic mechanisms of oliguric acute renal failure. They have yet to fulfil their promise.

General measures

Apart from the control of fluid and electrolyte balance, a number of measures have become part of the routine management of patients with acute renal failure (Table 7), and are based on common sense and knowledge of the causes of death and morbidity, rather than on the results of rigorous clinical trials.

1.	Avoid and treat infection:
	Remove unnecessary lines and catheters
	Regular surveillance cultures
	Use narrow-spectrum antibiotics
	Regular search for deep sepsis
	Good mouth, wound, vascular-access care
2.	H_2-antagonists to prevent stress ulceration
3.	Nutritional support:
	Enteral if possible
	Parenteral through tunnelled line
	Provide for hypercatabolism
4.	Maintain haemoglobin ¿ 10g/dl
5.	Prescribe prudently:
	Avoid certain drugs
	Modify dose and interval

Table 7 General measures in the management of acute renal failure

The first is the prevention and prompt treatment of infection. To avoid nosocomial infection the number of intravascular catheters should be kept to a minimum and there should be a low threshold for changing these. Bladder catheters should either be removed or connected to a closed drainage system. Samples should be examined frequently for bacteria and *Candida* spp. Fever should be promptly investigated. Antibiotic treatment will often be required before microbiological identification of the pathogen is available. The usual sites of sepsis are surgical wounds, the abdominal cavity, the lungs, central venous catheters, and the urinary tract. Surgeons should have a low threshold for re-exploring the abdomen in patients who have developed renal failure following abdominal surgery or penetrating injury, since in these seriously ill patients the clinical signs and imaging are too insensitive to detect residual sepsis.

The second general measure is the maintenance of adequate nutrition. This is more difficult in patients with renal failure, but with attention to fluid balance it is usually possible to provide both sufficient calories and sufficient nitrogen. The nutritional requirements will depend on the degree of catabolism. Depending on the method of renal replacement being used, adjustments in the potassium, phosphate, and sodium concentrations of parental nutrition solutions will be required. Gastrointestinal haemorrhage was a common cause of death but is seldom fatal today, perhaps because anticoagulation is minimized and the prescription of H_2-receptor antagonists or cytoprotective agents is routine. Most patients with surgical acute renal failure have received broad-spectrum antibiotics and therefore are at risk of developing *Clostridium difficile* toxin-related colitis. Stools should be tested regularly for the toxin, and prompt treatment with oral metronidazole or vancomycin should be instituted.

Finally it should again be emphasized that the metabolism and excretion of many drugs is altered in renal failure. Each prescription should be checked to determine whether dose modifications are required (see Table 2).

Renal replacement techniques

Peritoneal dialysis

Peritoneal dialysis via a rigid catheter inserted percutaneously or by surgical implantation of a soft catheter is still widely used for the treatment of acute renal failure as it is inexpensive and can be instituted without recourse to specialized equipment. Continuous peritoneal dialysis has three possible advantages over haemodialysis: a lower risk of exacerbating bleeding because heparinization is not required, less cardiovascular instability, and a lower risk of disequilibration. In practice these problems can be minimized or overcome with modern haemodialysis techniques. Peritoneal dialysis requires an intact peritoneum, a watertight abdomen, and a safely inserted dialysis catheter. It also needs a significant amount of nursing supervision. Other potential problems include difficulties in maintaining dialysate flow, peritoneal infection, protein losses, limited efficacy in hypercatabolic patients, and, when a rigid catheter is used, immobilization of the patient. The choice of peritoneal dialysis is often made for practical reasons such as lack of haemodialysis equipment, but it is the treatment of choice in infants. It should not be used in hypercatabolic patients or those with any past or recent intra-abdominal pathology, because adhesions make catheter placement hazardous.

Intermittent haemodialysis (Fig. 5)

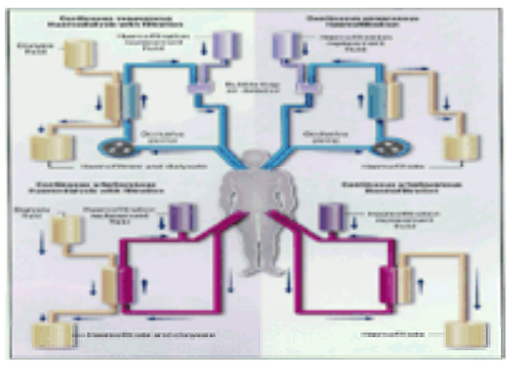


Fig. 5. The mechanisms of haemofiltration and haemodilaysis with filtration: no attempt has been made in this diagram to represent the automatic control of the rates of filtration and infusion of replacement fluid.

Intermittent haemodialysis is the orthodox treatment for acute renal failure and it is the most practical way to treat mobile patients. Access to the vascular system is either via catheters placed in central veins or by the use of an arteriovenous shunt. Treatment should be started before complications make it urgent. The first dialysis treatment is generally short, lasting 2 to 3 h. Depending on whether the patient is severely catabolic, subsequent treatments will need to be undertaken daily or on alternate days, aiming to keep the pretreatment blood urea below 33 mmol/l and to avoid fluid overload. In patients with an unstable cardiovascular system, bicarbonate-buffered dialysate is used in preference to acetate. The dialysis treatment is used as an opportunity to administer blood transfusions, allowing the extra fluid to be removed by ultrafiltration. The choice of dialyser membrane may be of importance. There are prospective studies that show that patients dialysed with ‘bio-incompatible' (usually Cuprophane) membranes have a lower survival rate, a higher risk of sepsis, and a more prolonged period of acute renal failure than those treated with ‘biocompatible' membranes (polymethyl methacrylate or polyacrylo-nitrile). These findings have not been universally accepted and so, because of the cost of the ‘biocompatible' filters, their use is recommended but not considered obligatory.

The risk of bleeding can be minimized if low doses of heparin are used, and if necessary this can be reversed at the end of dialysis by the administration of protamine sulphate. There is some evidence to suggest that the use of low molecular-weight heparin reduces the risk of bleeding, and this is used in some units for patients requiring dialysis in the early postoperative period. Heparin requirements can be further reduced by the use of prostacyclin to inhibit platelet function. The drug is short-lived and is given as an infusion immediately before, and during, the dialysis treatment. In most regimens, prostacyclin does not replace heparin but is used in combination with a reduced dose. The disadvantages are expense and hypotension. If a patient is actively bleeding when dialysis is required, dialysis can even be performed without heparin, provided the lines have been primed with heparin-containing saline and the blood flow is maintained over 200 ml/min.

The particular limitation of intermittent haemodialysis is the need to establish fluid balance for a 24- to 48-h period during a 3- to 4-h treatment. Thus, to control fluid balance, ultrafiltration has to be rapid, thereby risking hypotension, cardiac ischaemia, and arrhythmias. Haemodialysis treatment also needs to be frequent to avoid rapid and large changes in the concentrations of plasma urea and other molecules that may generate transcellular osmotic gradients, leading to ‘disequilibration', a syndrome that probably arises from the rapid entry of water into the brain cells and that leads to coma and fits. Another disadvantage of haemodialysis is the sequestration of white cells that occurs in the lungs as a result of complement activation at the dialysis membrane. Although this sequestration is short-lived, it can aggravate hypoxia.

Continuous treatments

The limitations of intermittent haemodialysis stimulated the development of the continuous treatments, which are particularly applicable to the immobilized patient with multiple system failure being managed in the intensive care or high-dependency unit. The advantages of these forms of treatment are that they can be used in patients with low systolic blood pressures and provide continuous metabolic control and fluid balance with very little perturbation to the cardiovascular system. This means that there are no constraints on the administration of parenteral nutrition, blood products, and drugs. A disadvantage is the need for continuous heparinization, but with low doses and removal by the filter this should not lead to dangerous systemic anticoagulation. The treatments can only be safely used if the patient has continuous bedside nursing supervision, is immobile, and is unlikely to dislodge the vascular access catheters. The use of these catheters is not without its problems. Arterial catheters can cause limb ischaemia in patients with pre-existing peripheral vascular disease, and damage to the vessel wall can lead to thrombosis, aneurysm formation, and leakage of blood into the retroperitoneal space. The advent of pumped venovenous circuits has meant that arterial catheters and arteriovenous shunts are now hardly ever used.

There are a number of forms of continuous renal replacement treatment that rely more or less on filtration or diffusion (see Fig. 4; Table 8). The simplest is continuous arteriovenous or venovenous haemofiltration, with continuous replacement of the filtrate with fluid that has the electrolyte composition of plasma but which is buffered with lactate rather than bicarbonate. Venovenous circuits require a blood pump: since this introduces the risk of potentially fatal air embolism, a venous air-trap detector and automatic safety switch are mandatory. Low-volume haemofiltration (6 l/day) controls fluid balance and substitutes partially for dialysis, but needs to be supplemented by haemodialysis treatments, which can be shorter and less frequent. High-volume haemofiltration (in excess of 15 ml/min) makes haemodialysis unnecessary, but fluid imbalance can occur very rapidly and scrupulous monitoring is needed. This can be achieved by computerized, pumped replacement of haemofiltration fluid or by using a simple gravity-feed device controlled by a balance system that returns fluid in direct proportion to that removed.

Technique	Abbreviation	Indication
Slow continuous ultrafiltration	SCUF	Fluid control
Continuous venovenous or arteriovenous haemofiltration	CVVH or CAVH	Continuous clearance and fluid control
Intermittent venovenous haemofiltration	IVVH	Clearance with intermittent fluid control
Continuous venovenous or arteriovenous haemodialysis	CVVHD or CAVHD	Continuous clearance with fluid control

Table 8 Continuous renal replacement treatment in acute renal failure

Another variation is continuous arteriovenous haemodialysis. The circuit is the same as that for continuous haemofiltration, but a low-resistance dialysis filter is used. The blood flow is kept low (less than 100 ml/min) and the negative pressure across the membrane is adjusted to control ultrafiltration. The additional component is the slow pumping of dialysis fluid, countercurrent through the dialysate compartment of the filter. Full equilibration in the dialyser is necessary to provide optimal clearance of nitrogenous waste products (15–20 ml/min). The effluent from the filter consists of dialysate and ultrafiltrate. Since most of the membranes used have a high ultrafiltration capacity, scrupulous attendance to fluid balance on an hourly basis, or by a continuous replacement system, is again required.

Prognosis

The survival of patients with acute renal failure following surgery depends largely on the underlying disease. The majority of deaths occur in the first week of acute renal failure, and are due to cardiorespiratory or multisystem failure. Sepsis, gastrointestinal haemorrhage, pancreatitis, liver failure, and cerebral damage also contribute to mortality. Overall survival has almost certainly improved over the three decades since dialysis was introduced, although this has not always been easy to demonstrate. Data from the Leeds acute renal failure database reveal a 56 per cent 90-day survival in all patients with surgical acute renal failure but only 39 per cent in the cardiac surgery subset. Mortality is highest in patients with burns (approximately 80 per cent), pancreatitis (approximately 70 per cent), and aortic aneurysm surgery (48 per cent). Treatment for acute renal failure is usually undertaken in the hope that if the patient survives the kidneys will recover. This is not always the case. Persistent renal failure is sometimes seen after severe shock, where the pathology is usually cortical necrosis. Chronicity is also more likely in the elderly patient and in those with pre-existing renal impairment. When acute renal failure occurs in certain settings or is associated with particular complications the prognosis is particularly poor (Table 9). There is, however, no proven method of establishing that the prognosis of a patient with acute renal failure is hopeless, so that decisions to discontinue

renal support in severely ill patients must be made individually.

1.	Superadded adult respiratory distress syndrome
2.	Prolonged requirement for inotropic support
3.	Unresolved sepsis
4.	Jaundice developing during acute renal failure
5.	Underlying premorbid organ dysfunction:
	Cardiac failure
	Chronic liver disease
	Haematological malignancy

Table 9 Patients with a poor prognosis

Further reading

Allon M. Treatment and prevention of hyperkalemia in end-stage renal disease. *Kidney International* 1993; **43**: 1197–209.

Bhandari S, Turney JH. Survivors of acute renal failure who do not recover renal function. *Quarterly Journal of Medicine* 1996; **89**: 415–21.

Brady HR, Singer GG. Acute renal failure. *Lancet* 1995; **346**: 1533–40.

Cochrane Injuries Group Albumin Reviewers. Human albumin administration in critically ill patients: systematic review of randomised controlled trials. *British Medical Journal* 1998; **317**: 235–40.

Forni LG, Hilton PJ. Continuous hemofiltration in the treatment of acute renal failure *New England Journal of Medicine* 1997; **336**: 1303–9.

Himmelfarb J, Tolkoff-Rubin N, Chandran P, Parker RA, Wingard RL, Hakim R. A multi-centre comparison of dialysis membranes in the treatment of acute renal failure requiring dialysis. *Journal of the American Society of Nephrologists* 1998; **9**: 257–66.

Humes HD. Acute renal failure—the promise of new therapies. *New England Journal of Medicine* 1997; **336**: 870–1.

Jacobs C. Membrane biocompatibilty in the treatment of acute renal failure: what is the evidence in 1996? *Nephrology Dialysis Transplantation* 1997; **12**: 38–42.

Klahr S, Miller S. Current concepts: acute oliguria. *New England Journal of Medicine* 1998; **338**: 671–5.

Liano F, Pascual J. Outcomes in acute renal failure. *Seminars in Nephrology* 1998; **18**: 541–50.

Liano F, Pascual J, and the Madrid Acute Renal Failure Study Group. Epidemiology of acute renal failure: a prospective, multicenter, community-based study. *Kidney International* 1996; **50**: 811–18.

Manske CL, Wang Y, Rector T, Wilson RF, White CW. Coronary revascularisation in insulin-dependent diabetic patients with chronic renal failure. *Lancet* 1992; **340**: 998–1002.

Offringa M. Excess mortality after human albumin administration in critically ill patients. Clinical and pathophysiological evidence suggests albumin is harmful. *British Medical Journal* 1998; **317**: 223–4.

Ronco C, Belloma R. Continuous renal replacement therapy: evolution in technology and current nomenclature. *Kidney International* 1998; **53** (Suppl.): S160–4.

Shilliday IR, Quinn KJ, Allison MEM. Loop diuretics in the management of ARF: a prospective, double blind, placebo-controlled, randomised study. *Nephrology Dialysis Transplantation* 1997; **12**: 2592–6.

12.4 Hepatic problems

Susan N. Cullen, Carol A. Cobb, and R. W. Chapman

Introduction
The surgical patient with normal preoperative liver function
Introduction
Overproduction of bilirubin
Gilbert's syndrome
Circulatory failure
Post-transfusion hepatitis
Coincidental viral hepatitis
Drugs
Total parenteral nutrition (TPN)
Fasting
Obesity
Diabetes mellitus
Sepsis
Benign postoperative intrahepatic cholestasis
Extrahepatic obstruction
Acalculous cholecystitis
The surgical patient with pre-existing liver disease
Introduction
Preoperative management
Nutritional status in patients with liver disease
Intraoperative management
Postoperative management
Further reading

Introduction

Postoperative liver dysfunction is a common problem. Although the incidence after elective abdominal surgery is less than 1 per cent, much higher rates occur after major surgery, multiple trauma, and prolonged intervention. The majority of cases are mild, transient, and resolve spontaneously, but occasionally the liver injury may be severe and result in fulminant liver failure and/or chronic liver disease. There are many aetiological factors, and in any one patient the pathogenesis is often multifactorial (Fig. 1).

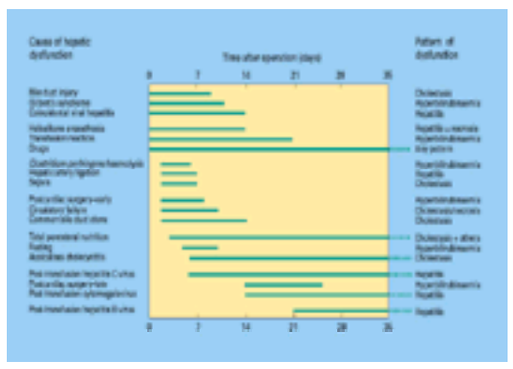


Fig. 1. Onset of hepatic dysfunction in relation to cause and pattern.

In this section the clinical presentation, investigation, diagnosis, treatment, and prevention of hepatic problems in the surgical patient are discussed. Patients with normal preoperative liver function are considered separately from those with pre-existing liver disease. Primary liver diseases, details of hepatobiliary surgery, liver transplantation, and surgical treatment of portal hypertension and ascites are dealt with elsewhere.

The surgical patient with normal preoperative liver function

Introduction

Postoperative hepatic dysfunction in surgical patients with normal liver function can be classified into three groups: those due to (i) overproduction of bilirubin; (ii) hepatocellular dysfunction; and (iii) extrahepatic biliary obstruction (see Table 1).



Table 1 Causes of hepatic dysfunction in the surgical patient with normal preoperative liver function

Overproduction of bilirubin

In a healthy individual the liver conjugates up to 500 µmol of bilirubin per day as a result of the breakdown of red blood cells. The liver is capable of handling several times this quantity without the occurrence of hyperbilirubinaemia and only if haemolysis is severe or occurs in conjunction with hepatocellular insufficiency does jaundice develop. Unconjugated bilirubin comprising 90 per cent of the total is suggestive of haemolysis. When the level of unconjugated bilirubin is excessively high there appears to be a concomitant rise in the conjugated fraction. The cause of significant haemolysis may be haemolytic anaemia, blood transfusion, resorption of hematomas, sepsis, or open-heart surgery.

Haemolytic anaemias

Congenital and acquired haemolytic anaemias can be associated with postoperative jaundice (see Table 2).

Sickle-cell disease
Hereditary spherocytosis
Glucose 6-phosphate dehydrogenase deficiency
Pyruvate kinase deficiency
Non-immune acquired haemolytic anaemias

Table 2 Haemolytic conditions associated with postoperative liver dysfunction

In surgical patients with sickle-cell disease there are increased risks as acute haemolysis and severe pain can be precipitated by infection, dehydration, acidosis, and hypoxia. In the postoperative period the earliest signs of infection must be treated promptly, especially as these patients have splenic hypofunction and are susceptible to bacterial infection. Patients from the African continent, parts of Asia, the Arabian peninsula, and southern Europe should be screened for sickle-cell disease.

Patients with hereditary spherocytosis may also experience a haemolytic crisis following infection, and in the postoperative period this can cause an unconjugated hyperbilirubinaemia. The diagnosis is suggested by the family history or the presence of a raised mean concentration of corpuscular haemoglobin, with more than 1 to 2 per cent of spherocytes on the blood film.

Surgery, infection, acidosis, and many drugs, including antibiotics and analgesics (see Table 3), may precipitate haemolysis in patients with glucose 6 -phosphate dehydrogenase deficiency. At least 10 million people worldwide have this red cell enzyme deficiency, and thus patients from the Mediterranean, South-East Asia, the Middle East, and West Africa should be screened preoperatively. The cresyl blue decoloration test or the methaemoglobin reduction test can be used in screening, and the diagnosis made by enzyme assay.

Definite risk of precipitating haemolysis
Antibiotics
• Nitrofurantoin
• Quinolones (including ciprofloxacin, nalidixic acid, norfloxacin, ofloxacin)
• Sulphonamides (including co-trimoxazole)
Antimalarials
• Primaquine
• Pamaquine
Others
• Dapsone and other sulphones
• Methylene blue
Possible risk of precipitating haemolysis
Aspirin
Menadione
Antimalarials
• Chloroquine
• Quinidine
• Quinine

Table 3 Some drugs that may precipitate haemolysis in patients with glucose 6-phosphate dehydrogenase deficiency

Pyruvate kinase deficiency is another red cell enzymopathy in which infection can precipitate haemolysis. Patients should be aware of their diagnosis but macrocytosis and an abnormal enzyme assay will confirm the diagnosis.

Causes of non-immune acquired haemolytic anaemia include disseminated intravascular coagulation; vasculitis; pneumococcal, meningococcal, and Gram-negative sepsis; *Clostridium perfringens* (was *C. welchii)* infection; burns; drowning; and some drugs. These are covered in other parts of this section.

Blood transfusion

Immediate and delayed haemolytic reactions may occur following blood transfusion. Within 24 h of the transfusion of one unit of stored blood at least 10 per cent undergoes haemolysis. Transfusion of two units of blood should not result in an increase in the serum bilirubin. If transfusion is rapid, massive, or occurs in a patient with impaired liver function, the capacity of the liver to conjugate bilirubin may be exceeded. Jaundice in this situation occurs 10 to 12 h after transfusion. Incompatibility of transfused blood may result in a severe immediate haemolytic reaction, which may occur if there are antibodies to the donated blood in the recipient's plasma. Jaundice appears at 12 h after commencing transfusion, peaks at between 24 and 36 h, and lasts for a total of 4 or 5 days.

Delayed haemolytic transfusion reactions are seen between 3 days and 3 weeks post-transfusion, with the peak reaction being at around 7 to 10 days. They are due to a secondary immune response, and in the majority of cases there has been sensitization to red cell antigens through past transfusion or pregnancy. This response is often to Rhesus and Kidd antigens, and is seen clinically as extravascular haemolysis with fever, jaundice, and anaemia. A serum sample should be screened for antibodies and future transfusions preceded by careful compatibility testing.

Resorption of hematomas

Large hematomas, crush injury, and bleeding from major vessels result in large pools of extravascular blood which, when resorbed, can result in an unconjugated hyperbilirubinaemia. As these patients often have hepatocellular dysfunction due to hypotension, hypoxia, and major surgery, as well as renal impairment, the severity and duration of jaundice may be marked. In a similar way, massive pulmonary infarction can cause hyperbilirubinaemia.

Sepsis

A massive haemolysis can occur in association with *Clostridium perfringens* (was *C. welchii)* infection 24 to 72 h after gastric, biliary tract, or colonic surgery. The typical clinical picture is of a restless hypotensive patient, an acute rise in serum bilirubin, and crepitus around the wound site. Several causes of liver dysfunction probably occur simultaneously in these patients as the conjugated bilirubin level can be greater than the unconjugated. As these cases can be fatal, prompt treatment with massive doses of penicillin and hyperbaric oxygen are imperative. Meningococcal, pneumococcal, and Gram-negative sepsis can cause haemolysis through disseminated intravascular coagulation and secondary microangiopathic haemolysis. As sepsis can also cause intrahepatic cholestasis, a combination of the two factors may cause marked jaundice.

Open-heart surgery

Early and late rises in bilirubin are seen after open-heart surgery. Early onset jaundice may be seen in up to 23 per cent of such patients and the main contributing factors are hypoxia, severity of right-heart failure preoperatively, and number of units of blood transfused. Although it has been suggested that cardiopulmonary bypass and prosthetic valves cause haemolysis, these are probably not significant contributors to the increased bilirubin. Late jaundice due to an autoimmune haemolytic anaemia has been reported where anaemia and jaundice, exacerbated by repeat transfusion, occur a few weeks after surgery. The presence of antiglobulin antibodies confirms the diagnosis; steroids are the treatment of choice.

Gilbert's syndrome

Gilbert's syndrome is a benign, familial, mild, unconjugated hyperbilirubinaemia that affects 2 to 5 per cent of the population. It is more common in men than women and presents in the second or third decade. It is usually diagnosed incidentally at a routine medical examination or when blood tests have been taken for another reason. The serum bilirubin does not exceed 100 µmol/l and is usually less than 50 µmol/l. Serum liver biochemical tests and hepatic histology are normal. There are no abnormal physical signs. Jaundice is usually mild but deepens with fasting and therefore may occur in the postoperative period.

Gilbert's syndrome has an autosomal recessive mode of inheritance. The promoter region of the gene encoding UGT1*1 is lengthened in patients with this condition resulting in reduced UGT1 (bilirubin uridine-diphosphate glucuronosyl transferase) enzyme production. Hepatic clearance of bilirubin is thus reduced. The diagnosis can be confirmed by: a two- to threefold increase in bilirubin induced by a 48-h fast and reversed by the resumption of a normal diet; the fall in bilirubin on taking phenobarbitone, which induces the hepatic conjugating enzyme; and the increase following intravenous injection of nicotinic acid, which raises the fragility of red blood calls. There is reduced bilirubin UDP-glucuronosyl transferase activity in liver biopsies, however Gilbert's syndrome can usually be diagnosed without recourse to biopsy. A diagnosis of Gilbert's syndrome must be considered if an isolated raised level of unconjugated bilirubin is noted after fasting or episodes of vomiting postoperatively.

Circulatory failure

Circulatory failure/surgical shock

Circulatory failure contributes to hepatic dysfunction in many surgical situations, although it is rarely the sole cause of the liver abnormality. Major trauma, burns, sepsis, massive blood loss, and surgery can be precipitants of ‘shock', and these factors often occur together. In particular, gastrointestinal blood loss and septicaemia increase the risk of liver dysfunction when associated with hypotension. Cholestasis is the most common pattern of injury following hypotension, and this is a benign complication with a good prognosis. Prolonged hypotension, which is often associated with increased right atrial pressure, results in an ischaemic hepatitis, for example in open-heart surgery. There is an initial striking elevation of serum aminotransferases up to 200 times the normal level, a marked decrease in prothrombin time, and a typically delayed bilirubin rise. These dramatic changes are seen within hours of surgery, and where no severe liver damage has occurred they revert rapidly to normal with restoration of liver blood flow and oxygenation. However, massive centrilobular hepatic necrosis can occur, and the ischaemic hepatitis can progress to fulminant hepatic failure, which has a high mortality rate. The clinical manifestation of hypoxic liver cell necrosis inevitably postdates the hypoxic event, and other causes, especially a viral hepatitis, must be considered.

Massive haemorrhage in combination with massive transfusion (for example, more than 20 units of blood) puts the liver particularly at risk of damage, should the patient survive. Patients with major trauma are particularly at risk of this form of liver damage as well as that due to direct liver injury. In one study, 2 per cent of patients with major trauma and shock developed significant jaundice.

Patients with major burns form another group in which circulatory failure is an important factor in the aetiology of the associated hepatic dysfunction. Haemolysis often adds to the bilirubin load on the liver.

Hepatic artery ligation

The normal liver usually tolerates hepatic artery ligation without significant sequelae unless the flow of portal-vein blood is inadequate because of vascular stricture and sepsis. Minimal derangements of bilirubin and alkaline phosphatase levels occur, and moderate increases in the aminotransferase levels in the first week may be the only consequence. Hepatic arterial collateral vessels develop very rapidly and this, in combination with the portal circulation, reduces the ischaemic insult. Extensive mobilization of the liver can involve division of the ligamentum and triangular ligament and if this precedes hepatic artery ligation, massive liver necrosis may result. If infarction occurs, the amounts of bilirubin and aminotransferases rise rapidly to high levels.

Post-transfusion hepatitis

The incidence of post-transfusion hepatitis has declined dramatically over the past 50 years by the identification of the viral agents responsible for the majority of cases and the development of antibody assays to screen donated blood for their presence. In countries where volunteer blood donors are used and hepatitis B surface antigen and hepatitis C antibody screening are routine, clinical post-transfusion hepatitis has been virtually eradicated. The viruses known to cause post-transfusion hepatitis are listed in Table 4. Although hepatitis G virus is certainly a transfusion transmissable agent, there is currently no evidence to suggest a causal relationship between HGV infection and hepatitis. HGV screening is therefore not routinely undertaken.

Hepatitis B virus (HBV)
Hepatitis C virus (HCV)
Hepatitis D virus (HDV)
Cytomegalovirus (CMV)
Epstein–Barr virus (EBV)

Table 4 Infective agents that can cause post-transfusion hepatitis

Hepatitis B virus (HBV)

The estimated number of carriers of HBV worldwide is 120 million, and in the United States there are at least 900 000 carriers. Transmission of HBV by blood transfusion has been reduced dramatically by the use of volunteer blood donors and the screening of donor blood, although today parenterally acquired HBV infection can still occasionally occur in transfused patients because genetic mutations in the virus result in false negative screening tests. Hepatitis B screening is currently performed using radioimmunoassay or enzyme-linked immunoassay. Plasma donors in the United States continue to be paid, but the fractionation process used to prepare derivatives together with additional viral inactivation procedures (e.g. heating, detergent treatment) render the plasma products virtually risk-free.

The clinical picture of acute HBV infection is very variable, from asymptomatic to fulminant hepatic failure. A fulminant fatal post-transfusion illness is seen in the elderly. Recovery without sequelae is the rule, but 5 to 10 per cent of infants and children with acute HBV infection develop chronic infection, as do 90 per cent of neonates. The mortality of hepatitis B infection is 1 to 3 per cent, and treatment with a-interferon is available for selected patients with chronic HBV infection. However, postexposure prophylaxis by administration of hepatitis B immunoglobulin and HBV vaccine is also available.

Hepatitis C virus (HCV)

Hepatitis C virus is an RNA virus with homology to the flaviviruses. It was identified in 1989 and found to be the causative agent of more than 90 per cent of non-A, non-B post-transfusion hepatitis. Both plasma and cell products are infective. Serological tests to detect hepatitis C infection were introduced in 1991 and have been progressively improved. The successful screening programme has led to a sharply decreasing incidence of post-transfusion hepatitis with less than 1 in 103 000 blood components transfused resulting in hepatitis C infection. The latest enzyme immunoassay HCV 3.0EIA reduces the seroconversion window to around 10 days leading to further risk reduction.

The incubation period of hepatitis C infection is 5 to 12 weeks, although after transfusions of a particularly large volume of infected blood this may be shortened to between 1 and 4 weeks. Clinically, the disease is usually mild with only a moderate elevation of aminotransferase levels. Rarely, a fulminant hepatitis may occur which has a high mortality. Levels of bilirubin and the aminotransferases may rise for a second, or even third, time in the first few months of the infection. Diagnosis is made

by polymerase chain reaction (**PCR**) in the acute phase as seroconversion may take up to 6 months or more. Tests for antibodies to HCV are widely available but the false positive rate is high. Repeat assays using different varieties of antibody, or the immunoblot technique as opposed to the enzyme-linked immunosorbent assay, improve the specificity. The PCR method will identify viral RNA in the serum and is the most accurate technique. Histologically, there are lymphoid follicles in the portal tracts, showing mild chronic inflammation bordering on chronic active hepatitis. There may be parenchymal inflammation and focal necrosis, as well as some fatty change. Chronic infection is also clinically mild and occurs in 85 per cent of patients or more. At least 41 per cent of patients develop chronic active hepatitis and 20 per cent develop cirrhosis. Chronic disease is more likely to occur in males, patients who acquire the infection at over 40 years of age, and in association with alcohol abuse. HCV also greatly increases the risk of development of hepatocellular carcinoma, four times more so than HBV. The time interval between infection and endstage liver disease may be of the order of 25 years in immunocompetent patients. Patients who are immunocompromised, for example due to HIV disease or immunoglobulin deficiency, have an accelerated course. Treatment of HCV-infected patients with a-interferon in combination with the nucleoside analogue ribovirin has been proved to be effective in approximately 40 to 50 per cent of cases. The response to treatment is correlated with the genotype of the virus.

Prevention was initially through donor education, exclusion of high-risk individuals as donors, and surrogate testing using serum aspartate aminotransferase levels and anti-HCV tests. Blood donors in the United Kingdom are now screened for HCV infection by methods involving the use of antibodies to viral components. Any positive blood is rejected despite the high false positive rate.

Hepatitis D virus (HDV)

Hepatitis D virus only infects individuals already infected with HBV. The incubation period is 30 to 50 days and chronic infection is the most common outcome. Diagnosis is made by finding serum IgM anti-delta. Acceleration towards cirrhosis may occur when patients infected with HBV then acquire HDV infection. Prevention of post-transfusion HDV infection is the same as for HBV.

Cytomegalovirus (CMV)

Cytomegalovirus transmission commonly occurs after transfusion, but infection is usually subclinical and benign. A glandular fever-like illness is typical at 2 weeks to 3 months after exposure. Clinically, fever, splenomegaly, jaundice, raised aminotransferase levels, and atypical lymphocytes may occur. However, massive hepatic necrosis and granulomatous hepatitis have been reported. Patients who are immunocompromised are at most risk and may also develop a fatal pneumonitis or disseminated infection. For these patients white cell depleted blood or CMV antibody negative blood is available. Virus can be cultured from saliva or urine and IgM antibody to CMV can be detected in the serum. In the United Kingdom, 50 to 60 per cent of the population is positive for anti-CMV antibody. Only a small proportion are infective but there is no readily available test for infectivity.

Epstein–Barr virus (EBV)

The EBV causes infectious mononucleosis or glandular fever and can be transmitted parenterally. Clinically there is fever, right upper quadrant pain with or without pharyngitis, and lymphadenopathy. Hyperbilirubinaemia occurs in about one-half of patients, aminotransferases are raised to 20 times the normal level in up to 80 per cent, and one-third have a raised alkaline phosphatase level. The Paul–Bunnell or monospot test is usually positive and the diagnosis is made by finding a raised IgM anti-EBV capsid antibody. The sinusoids and portal tracts are infiltrated with large mononuclear cells. The histology may be similar to that for hepatitis A, B, or C. Fatal acute hepatic necrosis is extremely rare and chronic hepatitis and cirrhosis do not occur. As infection with this virus is common, and infectivity cannot be specifically tested for, screening is notperformed.

Coincidental viral hepatitis

Patients incubating a hepatitic virus may come to surgery and anaesthesia. Postoperative deterioration in liver function frequently occurs and a mortality rate of 31 per cent in 36 such patients has been reported. Thus viral serology should be part of the investigation of a patient with hepatitic liver dysfunction occurring early after surgery.

Drugs

Many drugs used in the peri- and postoperative period have been associated with liver dysfunction. Almost every naturally occurring liver disease that affects humans can be mimicked by the toxic effects of drugs on the liver and this occurs through a wide range of mechanisms. Drugs can affect bilirubin metabolism at any stage causing hyperbilirubinaemia. The drug or its metabolite can be hepatotoxic or can precipitate a hypersensitivity reaction. Hepatocellular dysfunction may be due to cellular necrosis or intrahepatic cholestasis. Factors that increase the risk of drug-induced hepatic injury include pre-existing liver disease, increasing age, female sex, concurrent therapy, and genetic polymorphism.

Early symptoms of drug-induced liver injury are non-specific and include loss of appetite, lassitude, and occasionally right upper quadrant discomfort. There may be few clinical signs however, even in a patient who has biochemical and histological evidence of considerable hepatobiliary damage. Hypersensitivity reactions may be associated with a fever, rash, or eosinophilia. Jaundice in drug-induced liver injury carries a poor prognosis with a fatal outcome of approximately 10 per cent.

The list of potentially hepatotoxic agents is large and ever increasing. Some of the drugs used in the perioperative period are listed in Table 5. The general anaesthetic drugs are discussed separately. A hepatitic serum biochemical pattern must lead to exclusion of a viral aetiology, and the differentiation of intrahepatic and extrahepatic cholestasis is important and should be elucidated with ultrasound scanning. Liver biopsy will only rarely give a diagnosis. Diagnostic challenge with a suspected drug is not recommended as a severe or even fatal reaction can occur.



Table 5 Some hepatotoxic drugs used in surgical patients

General anaesthetic drugs

Halothane is amongst the most important of the idiosyncratic hepatotoxins. It was first introduced in 1956 and within 4 years there had been several reports of postoperative liver necrosis. The National Halothane Study reported the incidence of massive hepatic necrosis to be 1 in 35 000 halothane anaesthetics. Two subsequent studies suggested that the incidence was even higher at 1 in 6000 and 1 in 20 000 uses. After acetaminophen, halothane is the second commonest drug cause of fulminant hepatic failure. Nevertheless, it is still a commonly used general anaesthetic agent with many favorable properties and few adverse effects.

Two types of halothane-induced liver injury appear to exist. Ten to thirty per cent of patients exposed to halothane develop asymptomatic elevations of aminotransferase levels with no clinical features of liver disease. This condition is benign and self-limiting. Its relationship to the rare, severe syndrome of halothane hepatitis is unclear. The latter condition may represent the severe end of a spectrum of liver injury associated with halothane exposure or, more likely, is a separate idiosyncratic reaction.

Multiple exposures are the single most important risk factor for halothane hepatitis. Eighty per cent of patients developing the condition have received halothane more than once, usually in the preceding 28 days. Women are more commonly affected, as is the case with many other types of idiosyncratic hepatic drug reactions. Obesity is also a significant risk factor, possibly due to increasing body stores of halothane or because of higher hepatic activity of P450 2E1, an enzyme which catalyses the metabolism of halothane to reactive metabolites. Concomitant drug therapy with microsomal enzyme inducing agents may also predispose to halothane hepatitis, and

there is evidence of a genetic predisposition to developing the condition.

Fever is usually the initial symptom of halothane hepatitis and this occurs 7 to 14 days after a first exposure to the drug but earlier after multiple exposures. Symptoms of hepatitis occur 2 to 5 days later with anorexia, malaise, nausea, vomiting, and right upper quadrant pain. In most cases, dark urine, pale stools, and jaundice follow, although anicteric cases of halothane hepatitis also occur. Liver biochemistry is typical for acute hepatocellular necrosis with grossly elevated aminotransferase levels (e.g. alanine aminotransferase raised to 10 times the normal level) and elevated serum bilirubin levels, whereas the alkaline phosphatase level is often less than twice normal. Between 10 and 40 per cent of patients develop eosinophilia.

The main histological feature is centrilobular necrosis, varying from a multifocal spotty picture to confluent massive necrosis. Ballooning degeneration of hepatocytes, inflammatory infiltrate, stromal fibrosis, fatty change, and occasionally granulomatous aggregates are also seen. Distinction from viral hepatitis may be difficult.

A number of factors have been postulated in the pathogenesis of halothane hepatotoxicity which include toxic products of metabolism, hypersensitivity, genetic predisposition, regional hepatic hypoxia, and altered calcium homeostasis.

Management, as for all types of drug-induced acute hepatitis, is supportive. Patients with fulminant hepatic failure are best cared for in specialist centers and ideally in proximity to a transplant unit. Patients who have had an adverse hepatic reaction to halothane should be warned about the dangers of future exposure to the drug and advised to wear a MedicAlert bracelet. Up to 90 per cent of cases of halothane hepatitis could be prevented by taking an appropriate history before administering anaesthesia and by adhering to safety guidelines.

Enflurane hepatitis has been described, but on closer examination in many cases the alternative causes of liver injury had not been adequately excluded. True cases are extremely rare. The difference in hepatotoxicity between halothane and the other haloalkane anaesthetics is directly related to their potential to undergo P450-mediated metabolism. Around 30 per cent of halothane is metabolized, whereas the figures for enflurane and isoflurane are 2 per cent and less than 1 per cent, respectively. When enflurane hepatitis does occur, it is similar to halothane hepatitis in clinical presentation and histological features, and the two conditions probably share the same pathogenesis. Case reports of hepatotoxicity associated with isoflurane are extremely rare and so far sevoflurane and desflurane do not appear to have any adverse effects on the liver.

Total parenteral nutrition (TPN)

Since its advent in the 1960s, parenteral nutrition has become safer, more reliable, and progressively more efficient. However, complications still occur and hepatobiliary abnormalities are second only to catheter sepsis in requiring cessation of parenteral feeding. A number of different patterns of liver dysfunction occur.

Short-term TPN

Hepatic steatosis (fatty change) is the earliest and most benign hepatic lesion. It occurs within the first 14 days of TPN administration and is often, but not necessarily, paralleled by a rise in the serum aminotransferase levels. Patients receiving fat-free TPN are much more likely to develop steatosis, and standard TPN regimens now supply a proportion of calories as a lipid emulsion to minimize this problem. The initial change on histology is periportal fat infiltration but this may progress to pan- or centrilobular infiltration. A number of factors influence the accumulation of fat within the liver. Hepatic lipid metabolism is influenced by the balance between insulin and glucagon. High concentration glucose infusions induce high insulin levels and suppress glucagon production. Glycogenesis is therefore favored over lipolysis. Lipid may also accumulate because of increased delivery from peripheral fat stores and from defective production of lipoproteins that transport triglycerides from the liver. Other proposed causes of a fatty liver include excess activity of endotoxins or tumour necrosis factor, glutamine deficiency, a toxic effect of tryptophan metabolites, choline and carnitine deficiency, and increased bacterial translocation from an atrophic gut.

Intrahepatic cholestasis may occur after 2 to 3 weeks of treatment. This manifests as a rise in bilirubin and alkaline phosphatase levels at a time when the serum aminotransferases are returning to normal. The accompanying histological changes are bile duct proliferation, centrilobular cholestasis, and periportal inflammation. Changes in bile acid composition and particularly rises in lithocholic acid are probably responsible for the intrahepatic cholestasis. However, it has also been related to the amount of fat, protein, and total calories provided.

Long-term TPN

Chronic progressive liver disease is rare but well described in patients receiving long-term TPN. One study found 3 of 60 patients on home TPN developed clinically severe liver disease. Of these, one died from hepatic encephalopathy and hepatorenal syndrome after 11 years of TPN and another patient died postoperatively following a cholecystectomy and duct exploration. Patients requiring parenteral nutrition are likely to have multiple other risk factors for hepatic dysfunction such as hepatotoxic drugs, multiple transfusions, and repeated surgery. It can therefore be difficult to isolate TPN as the cause. The histological picture may be similar to that seen in patients on shorter-term TPN, but beyond 6-months therapy a cholestatic picture is the common finding. Cholestasis, hepatocyte necrosis, an alcoholic hepatitis-like picture, steatonecrosis, and early cirrhosis may all occur. The pathogenesis is likely to multifactorial, involving any or all of the mechanisms discussed under short-term nutrition and complicated by the underlying disease process. Patients with short bowel syndrome have the worst prognosis.

Cholelithiasis becomes progressively more common with increasing length of parenteral feeding. Biliary sludging has been found in 100 per cent of patients treated with more than 6 weeks of TPN and 23 per cent of 109 patients developed clinical cholecystitis during TPN treatment. Gallbladder stasis is the most likely cause of gallstone disease in patients on TPN. Gallbladder contractions are reduced by approximately two-thirds during exclusive parenteral nutrition.

Management of TPN-induced hepatobiliary disease

Once TPN has been identified as the most likely cause of deranged liver function, the optimal management is to restart enteral nutrition where possible. Liver function tests will return to normal in most patients within 1 month of cessation of TPN. If however continued parenteral nutrition is unavoidable and liver abnormalities persist and worsen, a number of therapeutic measures can be attempted.

First, a change in the composition of the TPN may be helpful. Lipid emulsions should be administered as approximately one-third of total calories. Patients should not receive more than 3 g/kg per day of the lipid preparation, however, as this may predispose to hepatic fat accumulation.

Second, changing the timing of TPN administration (‘cycling') may improve liver function. TPN solution is given for 8 to 12 h every 24 h rather than as a continuous infusion. This approach reduces the time during which the serum glucose concentrations are high, thereby avoiding persistently high insulin levels which may stimulate hepatic lipogenesis. An improvement in liver function tests will take 2 to 3 weeks to manifest after changing from continuous infusion to cycling. If neither of these two approaches is effective, the total caloric intake will need to be reduced to prevent progressive liver disease.

A number of other therapeutic options are currently under investigation. Metronidazole appears to prevent the development of intrahepatic cholestasis in some adult patients on TPN. The proposed mechanism for this action is the prevention of the intestinal overgrowth of anerobic bacteria allowing the bacterial 7à-dehydroxylation of chenodeoxycholic acid to the potentially hepatotoxic lithocholic acid. Ursodeoxycholic acid has been shown to improve the cholestatic liver function tests of patients with intestinal failure treated with home TPN. Choline supplementation has been shown to reverse hepatic steatosis in a small number of patients on long-term TPN who have low plasma concentrations of free choline.

The most effective approach to the problem of TPN-induced biliary disease has yet to be established. Cholecystokinin, chosen for its prokinetic effect on the gallbladder, has been used in one small study, where it seemed to prevent the formation of biliary sludge in patients treated with TPN. There is no evidence available yet, however, that it has any useful effect on patients with established biliary disease. Ursodeoxycholic acid and chenodeoxycholic acid have been shown to prevent gallstone formation in animal studies of TPN-induced biliary disease.

TPN in infants

Nutritional support of premature infants is now the most common use of TPN in paediatric medicine, and liver dysfunction has been noted in this group of patients throughout the development of this therapy. Children with chronic and/or extensive gastrointestinal disease make up the majority of the remainder. Infants receiving TPN are at risk of five types of hepatobiliary disease: fatty change, intrahepatic cholestasis, biliary sludging, gallstones, and acute acalculous cholecystitis.

Fatty change and hydropic swelling of hepatocytes on liver biopsy may occur early in the course of TPN. These changes are not the precursors of cholestatic disease; they occur infrequently and are reversible on stopping the TPN. Animal and human studies have suggested that the fatty change might be caused by excess

carbohydrate, but deficiency of essential fatty acids may also be a factor.

Intrahepatic cholestasis was first reported in a premature infant in 1971. The incidence of cholestasis increases with the duration of TPN and also with decreasing gestational age and birth weight. If treatment was for more than 60 days, the incidence rose to 60 per cent and to 90 per cent after 80 days. Duration of therapy is often a function of the gestational age, as is weight and therefore these are not independent variables.

The pathogenesis of infant TPN-associated cholestasis is still unclear but predisposing factors include prematurity, lack of enteral feeding, sepsis, and major surgical interventions. It is often difficult to separate several risk factors for hepatic injury in this population of intensively managed patients. Premature hepatic function is probably a primary factor: there is decreased hepatic uptake of bile salts, bile salt synthesis is reduced, ileal uptake of bile salts is inefficient, and the ability to detoxify lithocholic acid is impaired. In addition, most of the contents of TPN have been implicated as hepatotoxins.

Liver biopsy of patients with TPN-associated cholestasis will show bile pigment in the hepatocytes, bile plugs in canaliculi, pseudorosette formation, and a varying degree of triaditis, with or without eosinophils. With longer-term TPN, portal and lobular fibrosis may develop with expansion of portal tracts with portal ductular proliferation, and bile plugs in the interlobular ducts. If TPN is continued, these changes may progress to micronodular cirrhosis.

Cholestasis normally resolves when enteral nutrition is resumed and this is usually accompanied by reversal of most of the histological changes. In some patients, however, enteral feeding does not restore normal bile flow and in these cases protracted cholestasis may lead to irreversible liver damage and the death of the patient from liver failure. Successful operative reversal of intractable TPN-associated cholestasis has been reported, and encouraging results have been demonstrated with the use of intravenous cholecystokinin.

Biliary sludging and gallstones can occur in infants as well as adults on TPN. Endoscopic retrograde cholangiopancreatography (**ERCP**) or surgical disimpaction have both been used successfully in a limited number of cases.

Acalculous cholecystitis occurs most often postoperatively and is associated with hyperalimentation. The gallbladder is free of stones and the technetium-99m disopropyl iminodiacetic acid scan demonstrates non-visualization of the gallbladder. Cholecystectomy may be required.

Monitoring of premature infants on TPN is paramount and the commonly used ‘liver function tests' are not useful. A rise in unconjugated bilirubin occurs late in the course of cholestatic disease in these patients and total bilirubin is commonly raised in premature infants with no liver disease. Alkaline phosphatase is unhelpful as there is a preponderance of the bone isoenzyme in infants, and this isoenzyme is also affected by the child's nutritional status. Aminotransferase levels can also be unreliable as the levels do not correlate with the degree of cholestasis associated with TPN. g-Glutamyl transpeptidase is the most sensitive test, but lacks specificity. Serum bile salts have also been found to be good indicators of cholestasis, but there is a normal developmental delay in bile salt metabolism, giving rise to elevated levels. A combination of g-glutamyl transpeptidase and serum bile salts may be a more specific test. Ultrasound scanning and ERCP will help to exclude extrahepatic cholestasis, and it might also be necessary to perform a liver biopsy. Although the liver histology may be non-specific, a biopsy can identify other causes of cholestasis, such as cytomegaloviral hepatitis or extrahepatic obstruction. It can also indicate the urgency of cessation of the TPN and the prognosis of the liver pathology.

Fasting

Mild hyperbilirubinaemia can be precipitated by fasting and is due mainly to an unconjugated bilirubin rise. The majority of patients showing this effect are probably those with Gilbert's syndrome. Fatty change is also seen, particularly in acute weight loss or starvation. This is related to the increase in serum fatty acids and increased fatty acid turnover precipitated by decreased availability of glucose, a rise in glucagon levels, and increased sympathetic nervous activity. Obese subjects who lose weight rapidly may show a transient elevation of serum liver enzymes.

Obesity

Fatty change in the liver is seen in up to 50 per cent of subjects who are obese, with occasional periportal inflammation and fibrosis. Steatonecrosis and cirrhosis have been reported but this may be due to coexistent diabetes mellitus or alcoholic liver disease. Fifty per cent of patients who are obese can be shown to be glucose intolerant, and this and excess dietary fat and carbohydrate in relation to protein intake may be involved in the aetiology of steatosis. The fatty infiltration is perivenular and diffuse. Liver function tests may be abnormal and reflect more severe histological change. The changes are, in general, benign and non-progressive, and can be reversed by weight loss.

Diabetes mellitus

Patients with diabetes also show fatty change in the liver; the majority of patients being non-dependent on insulin and also overweight. Steatosis is very rare in patients with juvenile-onset insulin-dependent diabetes. Symptoms are rare; an enlarged, slightly tender liver may be found on examination, and liver function tests may be slightly deranged in about 20 per cent of patients with diabetes, but do not correlate with histology. The fatty change is centrilobular and diffuse. Weight loss and good diabetic control will resolve these abnormalities.

Steatonecrosis may also occur and this is seen in the non-insulin-dependent group. It has been suggested that the incidence of cirrhosis among patients with diabetes is twice that of the general population. This suggestion is unproved and may originate in the number of patients with cirrhosis who are glucose intolerant and have wrongly been classified as having primary diabetes.

Emergency biliary surgery in patients with diabetes has a higher than expected mortality. This is due in part to the disruption of glucose control caused by surgery, the increased risk of infection due to leucocyte dysfunction, and poorer wound healing.

Sepsis

Hepatocellular dysfunction occurs early in sepsis despite a hyperdynamic circulation and increased hepatic perfusion. This effect is mediated via Kupffer cell- or macrophage-derived proinflammatory cytokines such as tumour necrosis factor and interleukin 2. Sepsis can produce a deep jaundice, which may be cholestatic and occurs 2 to 4 days after the onset of bacteraemia. Pneumonia, Gram-negative bacteraemia, intra-abdominal abscess, and pyelonephritis can all cause a raised bilirubin. Gram-negative infection in infants frequently causes cholestasis. As in most cases of hepatic dysfunction discussed here, sepsis may be only one element in a multifactorial aetiology.

Biochemically and histologically, the changes are very similar to those observed with circulatory failure, with a moderate rise in conjugated bilirubin, aminotransferases, and alkaline phosphatase levels. However, an increase in the unconjugated bilirubin level also occurs, giving a rise in total bilirubin out of keeping with the increase in liver enzymes. Hepatic histological changes include biliary stasis, fatty change, and periportal inflammation. Extrahepatic biliary obstruction must be excluded. Pneumococcal, meningococcal, and Gram-negative sepsis may cause haemolysis by disseminated intravascular coagulation or a secondary microangiopathic haemolysis, and in these conditions the rise in unconjugated bilirubin will be prominent.

Benign postoperative intrahepatic cholestasis

‘Benign postoperative intrahepatic cholestasis' is unlikely to be a specific entity. It occurs in situations where blood loss is a prominent problem and is probably due to a combination of hypotension and multiple blood transfusions. Caroli in 1950 was the first to describe the occurrence of postoperative cholestatic jaundice. Benign postoperative intrahepatic cholestasis has been included in all lists of causes of postoperative jaundice since about this time. The aetiology of postoperative cholestasis is discussed within this section, and the majority of cases given this label in the past now have a definable cause.

Extrahepatic obstruction

Bile duct injury

Bile duct injury can follow cholecystectomy, common bile duct exploration, or any upper abdominal operation. If unrecognized at operation, jaundice, biliary fistula, or biliary peritonitis will occur in the early postoperative period. ERCP plays an important role in the diagnosis and treatment of postoperative biliary injury.

Common bile duct stones

Retained common bile duct stones after cholecystectomy and/or exploration of the common bile duct are uncommon. In the majority of cases, ERCP will both diagnose and treat this problem by sphincterotomy. Reoperation is required if ERCP fails or is not available. Some practitioners advocate visualization and, if necessary, clearance of the bile duct at ERCP prior to cholecystectomy. Occasionally, blood may collect in the common bile duct and cause obstruction.

Postoperative pancreatitis

Acute postoperative pancreatitis is uncommon and the cause is unknown. Thirty per cent of patients may be jaundiced, and oedema of the head of the pancreas is thought to result in some degree of obstruction and a low-grade hyperbilirubinaemia.

Acalculous cholecystitis

Acute non-calculous cholecystitis can occur after major trauma, burns, surgery that does not involve the upper abdomen, and in patients receiving long-term total parenteral nutrition, especially infants. It accounts for about 1 per cent of all cases of cholecystitis. A Japanese series of acalculous cholecystitis after gastrectomy demonstrated an incidence of 0.64 per cent. The aetiology is unknown, but it has been suggested that biliary stasis is important.

Postoperative cholecystitis occurs most commonly in the fifth to seventh decade, but in patients with trauma or burns this form of cholecystitis is seen most frequently in the second to fourth decade. The sex ratio is also different in these two groups; females predominate in the former, males in the latter. The postoperative form tends to follow a major surgical procedure. This form of cholecystitis can occur up to 1 month after the operation. Right upper quadrant pain and tenderness is usually accompanied by nausea, vomiting, and fever. The observed bilirubin rise is variable but may be up to 85 μmol/l; levels of aminotransferases and alkaline phosphatase are only mildly raised. Ultrasound may show enlargement of the gallbladder and, by definition, no gallstones are seen. ERCP is often necessary to exclude other causes of obstruction, although cholecystectomy should not be delayed in these already seriously ill patients. Histologically, the gallbladder shows vascular dilatation, congestion, and oedema in all layers, without fibrosis. Abscesses of varying size may be seen in the gallbladder wall and the mucosal surface is necrotic and ulcerated. Perforation is frequent.

Acalculous cholecystitis occurs in patients receiving total parenteral nutrition for more than 3 months with an incidence of 4 per cent.

The level of mortality has been given as between 33 and 75 per cent. However, this may pertain only to patients with major trauma and reflects the already much increased mortality in this group.

The surgical patient with pre-existing liver disease

Introduction

Patients with liver disease may require surgery as a consequence of the disease or for unrelated problems. Any surgery involving the liver itself, the biliary system, or the blood vessels associated with the liver has the potential to cause hepatocellular dysfunction, bleeding problems, nutritional difficulties, and infection. Where there is impaired liver function in such patients the problems discussed in this section become doubly important.

Anaesthesia and surgery in a patient with pre-existing liver disease have the potential to cause deterioration in liver function and even acute liver failure. Peri- and postoperative haemorrhage and postoperative infection are the other major complications. Factors that should be considered in these patients are the nature of the liver disease, the preoperative assessment of the liver disease, fluid balance, renal function, drug metabolism and adverse effects, nutrition, maintenance of liver blood flow, and prevention of complications.

Preoperative management

The nature and severity of the liver disease should be confirmed before surgery whenever possible, as there are specific risk factors related to the underlying liver pathology.

In patients with known liver disease, assessment of hepatic function is important preoperatively. Standard liver tests (aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, g-glutamyl transferase) do not give a good guide to hepatic function. Serum bilirubin in patients with cirrhosis can be used as a predictor of liver cell failure. Measurements of serum albumin and liver-dependent clotting factor will assess the synthetic capacity of the liver. The indocyanine green clearance test can also be used to assess function and hepatic blood flow. As haemorrhage and infection are important complications in these patients, platelet count, haemoglobin level, white blood-cell count and differential, blood grouping, and cross-matching of sufficient fresh blood is essential. Serum urea estimation is often unreliable in assessing renal function in liver disease, as hepatic dysfunction causes a low urea. An electrocardiogram, chest radiograph, and screening for any infection are also mandatory. Further specific tests may be required and are mentioned below as appropriate.

Improvement of the preoperative status of patients with liver disease can significantly decrease their operative morbidity and mortality. Specific attention should be given to:

1. correction of coagulopathy to normal by administering vitamin K and fresh frozen plasma;
2. improving the nutritional status;
3. treatment of renal impairment;
4. treatment of infection; and
5. control of ascites.

The nutritional status of patients with liver disease is discussed separately below.

Patients positive for hepatitis B virus (HBV) or C virus (HCV)

The HBV- or HCV-positive patient that comes to surgery is at risk of postoperative deterioration in liver function and is also a risk to all staff involved in his or her care. Hepatitis B surface antigen assay is a good screening test but anti-HBc positivity is more reliable in atypical cases (see above). HCV infection should be determined by PCR. Both HBV and HCV infection are associated with periods of mild, or even subclinical, liver disease, and both can result in chronic liver disease and cirrhosis. Liver function and associated complications of chronic liver disease must therefore be identified prior to surgery.

Acute hepatitis

Surgery should be avoided if possible in the patient with acute hepatitis as acute liver failure may be precipitated, liver function is unpredictable, and the specific cause may not yet be identified (for example, viral serological tests may not become positive until some weeks after the acute illness). One study reported a 61 per cent morbidity and 31 per cent mortality following surgery in patients with undiagnosed acute liver disease. The deaths from hepatic failure occurred in those patients with viral or alcoholic hepatitis.

Obstructive jaundice

Patients with obstructive jaundice are at risk of postoperative renal failure, haemorrhage, and deterioration in liver function. It has been shown that patients with liver disease and hyperbilirubinaemia have abnormal renal structure and function, abnormal circulatory haemostasis, and deterioration in the gastrointestinal barrier to infection, and all these contribute to the postoperative risk of renal failure. Increased levels of unconjugated and conjugated bilirubin and bile salts damage the middle segment of the proximal convoluted tubule, producing changes very similar to those caused by anoxia. Decreased creatinine clearance occurs preoperatively in 30 per cent of patients with obstructive jaundice, and this is seen particularly in patients with coincident sepsis. This is due to decreased renal blood flow, which is more marked in the cortex and results in impaired ability of the kidney to concentrate urine, as well as susceptibility to sodium and water depletion. The cause may be decreased sensitivity to catecholamines. Patients with obstructive jaundice have lowered peripheral vascular resistance, renal salt wasting, some loss of left ventricular function, and pooling of blood in the splanchnic bed. Thus, a small volume of blood loss can result in a marked fall in arterial blood pressure, which will obviously exacerbate any renal failure. Endotoxins have been found in the circulation of 50 to 70 per cent of patients with obstructive jaundice, and these toxins also cause defects in renal structure and function. In particular, they increase renal vascular resistance and cause endothelial swelling, fibrin deposition, and low-grade disseminated intravascular coagulation. Bile salts disrupt endotoxins in the gut lumen, but in obstructive jaundice this is prevented and endotoxins readily reach the

portal circulation. Hyperbilirubinaemia, bile salt retention in the liver, and the endotoxins impair the reticuloendothelial phagocytotic function of the Kupffer cell systemand endotoxins enter the general circulation. Table 6 lists the parameters that help identify the patient at risk of postoperative renal failure.

Age > 60 years
Haematocrit < 30 per cent
White blood-cell count > 10000/mm^3
Albumin < 30g/l
Hyperbilirubinaemia
Raised alkaline phosphatase
Raised serum creatinine
Weight loss
Malignant disease

Table 6 Parameters that identify the patient particularly at risk of renal failure

The principles of preoperative treatment of these patients include the treatment of sepsis, avoidance of nephrotoxic drugs, treatment of renal impairment, and the correction of hypovolaemia, hypoalbuminaemia, hyponatraemia, and anaemia. Preoperative renal failure will require correction of fluid balance, administration of antibiotics, and dialysis if necessary.

Chronic liver disease and cirrhosis

Chronic liver disease can be associated with adequate liver function where there is no hyperbilirubinaemia, hypoalbuminaemia, or coagulopathy. If infection and renal impairment are also excluded, specific preoperative treatment may not be necessary. Perioperatively, particular consideration of fluid balance, ventilation, liver blood flow, and drug metabolism is important. There is still a risk of liver decompensation, haemorrhage, and infection postoperatively. Peri- and postoperative management are discussed below.

Cirrhosis may be associated with well-compensated liver function, assessed as above. However, surgery in patients with cirrhosis has a high mortality, reported to be 80 per cent in patients with advanced cirrhosis undergoing cholecystectomy. Child's classification is a useful guide to the preoperative assessment of risk in these patients, which is made more accurate by the addition of the prothrombin time (see Table 7). Prothrombin time, serum albumin, and the presence of infection (white blood-cell count >10 000/mm^3) are regarded as the most useful indicators of risk. The risk is further increased when these patients undergo gastrointestinal surgery and is highest following hepatobiliary surgery.

	Group A	Group B	Group C
Serum bilirubin (μmol/l)	<40	40–50	>50
Serum albumin (g/l)	>35	28–35	<28
Ascites	None	Mild	Moderate–severe
Encephalopathy	Absent	Grade I–II	Grade III–IV
Prothrombin time (prolonged from control)	Normal	<2.5 s	>2.5 s
Surgical risk	Good	Moderate	Poor
Mortality (%)	1–10	10–30	50–60

Table 7 Child Pugh risk assessment

Cirrhosis is also associated with renal abnormality. The latter is accompanied by proteinuria and hematuria. Renal failure may occur in 50 to 75 per cent of patients with cirrhosis. It may be prerenal, due to acute tubular necrosis, or it may have no obvious cause, when it may be labeled the hepatorenal syndrome. Preoperatively these patients should be assessed daily for weight change, ascites, oedema, and pyrexia. The sodium balance must be estimated; urine volumes, urine and serum osmolarity, serum creatinine, urea, sodium, potassium, and albumin must be checked; and haemoglobin, hematocrit, and liver function must be tested. Deranged coagulation will require daily vitamin K injections, but where hepatocellular function is poor these may not correct the abnormalities and fresh frozen plasma will be needed to cover any invasive tests or bleeding episodes as well as perioperatively.

Infection, hypoalbuminaemia, and ascites should be treated. The last should be treated by nutritional support (see below) along with paracentesis and albumin infusion. Urinary catheterization and a central venous catheter are essential but, again, infection must be avoided.

Nutritional status in patients with liver disease

Perioperative malnutrition increases the morbidity and mortality of any operation. In the patient with liver disease, malnutrition will compound the already significant risk of complications. Obstructive jaundice results in malabsorption of fat and steatorrhoea. If obstruction is prolonged, malnutrition will develop. Inevitably, chronic parenchymal liver disease is associated with protein–calorie malnutrition. Glucose intolerance is seen in 50 to 80 per cent of patients with cirrhosis; deficiencies of vitamins A, C, and E, folic acid, and zinc result, and there is a deficiency of branched-chain amino acids. These, and the nature of the liver disease, result in compromised immune defence, and postoperative infection and wound dehiscence are common problems.

Malnutrition can be identified from a history of weight loss, the patient's height to weight index, body composition, protein turnover, serum albumin, urinary urea, and the immune status, as well as a number of other detailed tests.

Nutritional support can be given as oral supplements, enteral, or parenteral feeding. Oral feeding is obviously the most efficient, cheap, and safe, when it is appropriate. Enteral feeding is well tolerated in obstructive jaundice. Special formula feeds are available for patients with chronic liver disease. They contain a high proportion of branched-chain amino acids which may be deficient and which also help reduce protein catabolism, another factor in the malnutrition in these patients. When such patients require surgery, parenteral feeding is often the only method available. Extensive investigation and preparation often dictates periods of starvation preoperatively and may also be a factor delaying surgery. Parenteral feeding regimens need to be tailored to the patient. If there is no glucose intolerance, glucose solutions can be the major source of calories; but if sugar intolerance exists, fat and protein content must be increased. Lipid emulsions seem to be well tolerated by patients with liver disease. Insulin resistance may be a problem, particularly in patients with infection. In patients with well-preserved liver function, and no past or presenting encephalopathy, standard formulations are acceptable. If there is a risk of encephalopathy, or if it already exists, a mix containing less phenylalanine, tyrosine, tryptophan, and methionine, and more arginine and branched-chain amino acids is recommended. Vitamins A, D, E, K, and C, folic acid, zinc, and copper must be regularly supplemented.

The maximum benefit to be gained from preoperative parenteral feeding comes from 10 to 14 days of therapy. The aim should be to correct the amino acid profile, reach a positive nitrogen balance, and scrupulously avoid infection. Renal failure is often another problem; volume and electrolytes must be carefully monitored, and the formula may need to be adjusted further. If the patient is requiring haemodialysis or haemoperfusion, this will assist the control of fluid balance. In obstructive jaundice, if there is no hepatocellular dysfunction, standard supplemented oral, enteral, and parenteral feeding regimens can be used. Obstructive jaundice should not be left untreated any longer than necessary (see above).

Monitoring is vital and should consist of at least 8-hourly tests of blood sugar; daily weight, urea, creatinine, and electrolyte measurements; biweekly determinations of calcium, phosphate, liver function, and albumin; and weekly measurement of zinc and magnesium levels and culture of blood, urine, and any other drained fluid. The

duration of feeding should be reviewed constantly.

Intraoperative management

The principal intraoperative concern is anaesthetic technique. Basic principles pertain to patients with liver disease, but protein binding, metabolism, and liver blood flow are particularly important considerations. An experienced anaesthetist should be employed where liver disease and hepatobiliary surgery are combined. In patients with well-compensated liver disease, standard drugs may be appropriate. The premedication may be affected by albumin levels, and the response to opiate drugs at this time may be helpful in assessing the dose requirements postoperatively. Narcotic analgesics and the benzodiazepines may have very prolonged action in patients with hepatocellular dysfunction.

During surgery there are several specific problems. Periods of hypotension or hypoxia will compromise liver function and precipitate renal failure more readily than in other patients. There may be derangement of blood coagulation. Rapid or large-volume transfusions of blood products may be required. Hepatobiliary disease and renal impairment will modify most drug pharmacokinetics and dynamics.

Of the haloalkanes, isoflurane is the best agent to use in these patients (see above); however, inhalation agents should be avoided. It should also be noted that prolonged use of nitrous oxide may cause additional liver damage. Liver blood flow, hepatocellular dysfunction, and plasma protein concentration are the factors affecting anaesthetic drug kinetics and dynamics intraoperatively. All anaesthetic techniques cause a reduction in hepatic blood flow, as do hyperventilation, increased sympathetic tone, and surgical manipulation of the liver and abdominal viscera. Good liver perfusion is also important for the clearance of endotoxins and lactate, and metabolic acidosis is another potential problem. A fall in liver blood flow will reduce the clearance of drugs and enzyme activity intraoperatively and the potential for a postoperative deterioration in liver function will be increased.

Control of ventilation must also be precise, as hypoxia, acidosis, and hyperventilation can all affect liver and renal function adversely, both intra- and postoperatively. Central venous catheterization, a urinary bladder catheter, and an arterial line are mandatory. In patients with cirrhosis and renal impairment, sodium-containing fluids should be used judiciously. Fresh blood and fresh frozen plasma must be readily available, and in long operations electrolytes, blood sugar, and coagulation will need to be monitored regularly.

The importance of avoiding infection during all anaesthetic and surgical techniques cannot be stressed too often.

Postoperative management

The immediate postoperative concern is whether to reverse the anaesthetic or whether ventilatory support should be continued in an intensive care unit. Reasons to continue ventilation in these patients include:

1. the persistent effect of anaesthetic and/or neuromuscular blocking agents—in obstructive jaundice, neuromuscular blocking drugs will have prolonged action;
2. massive intraoperative blood loss and transfusion;
3. anticipation of further bleeding;
4. any episode of cardiac arrest during the operation;
5. significant preoperative lung disease;
6. failure to maintain adequate arterial oxygenation; and
7. post-sternotomy, cardiopulmonary, or left-heart bypass.

Regular assessment of liver function in the postoperative period will allow the early identification of deterioration. All the factors discussed in the first part of the section pertain to these patients also.

Central venous catheters should be left in place to monitor fluid balance, especially in relation to risk of bleeding and renal failure. Hematological monitoring must include daily coagulation testing and the prompt correction of abnormalities. Intravenous vitamin K should be given daily.

Constant vigilance against infection is mandatory in postoperative patients with liver disease. All catheters and drainage systems must be watched and handled by experienced staff. Blood, urine, and drained fluids must be monitored regularly for infection. Patients with cirrhosis are particularly prone to chest infection, and chest physiotherapy should be rigorous in these patients. Consideration of peritoneal infection is important, as this can be an occult cause of deteriorating liver function. Biochemical monitoring should be routine as renal function is at risk. When haemorrhage, hypotension, and infection coexist, small volumes of concentrated urine, with a urine to plasma osmolarity ratio greater than 1.05, indicate potential renal failure. This is reversible if recognized and treated promptly. Established renal failure can be oliguric or high output.

Upper abdominal, subcostal, and sternotomy incisions are associated with severe postoperative pain, which is made worse by prolonged procedures. Patients with liver disease often require these operative procedures, and thus pain relief is particularly important. Septicaemia and coagulation abnormalities may limit the administration of analgesia by epidural methods. Doses of other intravenous or inhalational analgesics should be low initially, and they must be monitored carefully. The duration of action of benzodiazepines and narcotic analgesics can be very prolonged in patients with hepatocellular disease.

Further reading

The surgical patient with normal preoperative liver function

Barbara JA, Contreras M. Non-A, non-B hepatitis and the anti-HCV assay. *Vox Sanguinis* 1991; **60**(1): 1–7. [A review of the uses and limitations of the anti-HCV assay.]

Bircher J, Benhamou J-P, McIntyre N, Rizzetto M, Rodes J, eds. *Oxford textbook of clinical hepatology*. 2nd edition. Oxford University Press, 1999. [A useful reference text.]

Farrell GC. *Drug-induced liver disease*. Churchill Livingstone, 1994. [A highly-recommended textbook covering all aspects of this topic.]

Nunes G, Blaisdell FW, Margaretten W. Mechanism of hepatic dysfunction following shock and trauma. *Archives of Surgery* 1970; **100**: 546–56. [Case reports and one of the original reviews of the causes of postoperative jaundice.]

Sherlock S, Dooley J. *Diseases of the liver and biliary system*, 10th edn. Blackwell Science, 1997. [A comprehensive textbook of liver disease.]

Tobler LH, Busch MP. History of posttransfusion hepatitis. *Clinical Chemistry* 1997; **43**(8): 1487–93. [A complete history of the discovery and importance of the transfusion-associated hepatitis viruses since 1940.]

Wang P, Chaudray IH. Mechanism of hepatocellular dysfunction during hyperdynamic sepsis. *American Journal of Physiology* 1996; **270**: R927–38. [A highly readable review of this subject.]

Zimmerman HJ, Fang M, Utili R, Seeff LB, Hoofnagle J. Jaundice due to bacterial infection. *Gastroenterology* 1979; **77**: 362. [Clinical conference and discussion of hepatic involvement in systemic and extrahepatic conditions.]

Total parenteral nutrition

Bowyer BA, Fleming CR, Ludwig J, Petz J, McGill DB. Does long-term parenteral nutrition in adult patients cause chronic liver disease? *Journal of Parenteral and Enteral Nutrition* 1985; **9**: 11–17. [A prospective study of 9 of 60 patients developing abnormal liver function tests on long-term home TPN.]

Capron J-P, Herve M-A, Gineston J-L, Baillon A. Metronidazole in the prevention of cholestasis associated with total parenteral nutrition. *Lancet* 1983; **Feb 26**: 446–9. [A randomized controlled study of the effect on liver function tests of giving metronidazole with TPN in Crohn's disease patients.]

Clarke PJ, Ball MJ, Kettlewell MGW. Liver function tests in patients receiving parenteral nutrition. *Journal of Parenteral and Enteral Nutrition* 1991; **15**(1): 54–9. [A 5-year prospective study of the effect of long-term TPN on liver function tests.]

Fein BI, Holt PR. Hepatobiliary complications of total parenteral nutrition. *Journal of Clinical Gastroenterology* 1994;**18(1)**: 62–6. [An excellent and comprehensive review of this subject.]

Grant JP. *Effects of cyclic vs. continuous TPN in hospitalized patients*. American Gastroenterology Association Postgraduate Course, Nutrition in Gastroenterology, 1987. [Clinically based information on this topic.]

Jeejeebhoy KN, Anderson GH, Nakhooda AF, Greenberg GR, Sanderson I, Marliss EB. Metabolic studies in total parenteral nutrition with lipid in man. Comparison with glucose. *Journal of Clinical*

Investigation 1976; **57**: 125–36. [One of the original studies advocating the use of lipid as the main source of non-protein calories in TPN feeds.]

The surgical patient with pre-existing liver disease

Garrison RN, Cryer HM, Howard DA, Polk HC. Clarification of risk factors for abdominal operations in patients with hepatic cirrhosis. *Annals of Surgery* 1984; **199**: 648–54. [A retrospective review of 100 patients with liver cirrhosis undergoing laparotomy with discussion identifying preoperative risk factors and review of the literature.]

Greenway B, Williams R. Adverse effect of exploratory laparotomy in patients with unsuspected liver disease. *British Journal of Surgery* 1982; **69**: 449–51. [A retrospective review of the outcome for 36 patients referred to the King's College Hospital liver unit after unsuspected liver disease was found at laparotomy.]

Silk DBA. Parenteral nutrition in patients with liver disease. *Journal of Hepatology* 1988; **7**: 269–77. [A review of clinical trials and practical guide to the topic.]

12.5 Psychological care

Owen S. Surman and Elizabeth Gray

Fear of surgery
Origin of preoperative anxiety
  Past trauma
  Identification
  Expectation
Preoperative psychological evaluation
  Personality disorder
  Affective disorder
  Anxiety disorder
  Cognitive impairment
  Alcoholism
Informed consent
  Bonding
  Teaching
  Observing
Innovative surgery
  Patient selection
Reduction of preoperative stress
  Education and support
  Specialized intervention
Preoperative management of psychiatric disorders
  Personality disorder
  Major depressive disorder
  Anxiety disorder
  Panic disorder
  Phobic disorder
  Hex or predilection to death
  Functional psychosis
  Preoperative encephalopathy
  Postoperative encephalopathy
  Alcoholism
  Drug-addicted patients
  Pain intolerance
Awareness during operations
Further reading

Fear of surgery

Two types of fear commonly encountered in the surgical patient are fear of bodily injury and fear of death, typically fear of not awakening from anesthesia (narcosis anxiety). Other common sources of apprehension are fear of pain, fear of cancer being discovered at operation, fear of intraoperative wakefulness, and fears of non-specific factors common to the hospital experience such as separation from job and family. Studies of patients undergoing orthopedic and gynecologic surgery have demonstrated that high levels of anxiety precede hospital admission and persist for several days following operative intervention. One study showed that surgery and anesthesia residents were only able to predict what made their patients fearful about surgery about 50 to 60 per cent of the time.

Origin of preoperative anxiety

Past trauma

The surgical experience recalls early life stress. For some, parental separation at a time of childhood surgery, or unpleasant exposure to a mask for anesthesia induction may trigger abnormal fear of surgery in adulthood. Those with a traumatic past are especially vulnerable.

Identification

Emotional adaptation to surgery may differ, according to expectations derived from the surgical experience of relatives. Patients encountering a similar disease process may observe, share, and compare outcomes in a way that modifies the impact of events: visits from recipients of organ transplants improve the hopes and coping skills of patients awaiting a suitable organ.

Expectation

Surgery is often a source of hope and improved identity, as in the case of cosmetic procedures and transplantation. In other instances, however, surgery may represent a substantial loss. The burden of mastectomy or colostomy has inspired the formation of successful self-help groups. Although there is an implicit gain for the patient whose life is maintained by removal of cancer or whose proximal limb is saved by amputation of a gangrenous distal part, the subjective or symbolic meaning of operative intervention is also of significance to recovery. Emotional outcome is particularly influenced by the patient's knowledge and orientation to perioperative events, particularly the realistic appraisal of what can be expected. Two factors that increase perioperative anxiety are unpredictability and underestimation of pain and risk.

Preoperative psychological evaluation

Patients with psychopathologic states require identification and specialized medical management.

Personality disorder

The different types of personality disorder have in common a basic problem with trust and a pattern of failed or strained relationships. Problems with medical compliance may occur as well as strain in the doctor–patient relationship. The patient may discharge him- or herself. Costly litigation, or even personal injury to the caregiver or a colleague, may follow from the perceived injustice of a malcontent. Some patients who are unlikely to benefit from surgery may seek an operation or a series of operations in a neurotic attempt to gain attention.

In addition to identifying personality pathology, it is important to recognize normal variations in coping style, particularly individual tendencies toward anxiety and in locus of control. Some attribute the outcome of events to external factors beyond control, while others perceive events to be under greater personal influence. Some patients may adopt an avoidance or denial approach to the threat of surgery, while others may exaggerate risks. More stress is encountered among the young, among those with an ‘excess of recent life events', and among those with medical conditions that are relatively demanding. Good outcomes are more likely for those with an active and energetic orientation.

Affective disorder

The depressed patient may be irritable, agitated, or quietly withdrawn. Postoperative mobilization is a challenge and impaired nutrition undermines the process of surgical repair. Treatment of depression requires supportive psychotherapy, psychopharmacologic intervention, and social support. When depression is secondary to surgically correctable physical impairment, successful operative intervention is most often followed by improvement in mood and well being. For example, hysterectomy is often preceded by psychiatric morbidity, which decreases after the operation. Depression is sometimes mistaken for a normal reaction to illness. At times, it may be

inaccurately diagnosed when encephalopathy is the real culprit.

Anxiety disorder

Anxiety may result from misconceptions about surgery, from anniversary reaction to past trauma, or from the impact of new learning or increased physical impairment on established coping skills. Those who deny their disease are especially likely to react anxiously to detailed preoperative information. Isolated phobias such as needle phobias, claustrophobia, or pathologic dread of anesthesia are occasionally encountered, as are generalized anxiety states and multiple phobias with panic attacks. Formal assessment of preoperative anxiety can be performed on rating scales such as a 100-mm visual analog scale which corresponds to the range between extremes of anxiety.

Treatment of anxiety begins with preoperative teaching and formation of a therapeutic alliance. Some patients may derive considerable support from contact with others who have successfully completed a similar operative procedure. Such peer group support has been useful in a variety of surgical settings. Those who are unresponsive to these measures or who have chronic anxiety disorders often benefit from psychotherapy and treatment with anxiolytic agents.

Cognitive impairment

Impaired cognition in patients requiring surgery is most frequently of metabolic origin and is associated with increased risk of postoperative delirium. However, patients who are functionally psychotic, demented, or retarded must be recognized. The Mini-Mental State test is an excellent screening tool for delirium and dementia but has a high rate of false-positive results among those with less than 9 years of education and among people aged 60 or more.

Alcoholism

Patients who are alcoholics are often exquisitely sensitive to rejection and given to pathologic denial. Preoperative recognition of the potential for delirium tremens and of the higher risk of postoperative delirium is important in patients with a history of alcoholism. Postoperative delirium may occur in the form of delirium tremens, as a manifestation of alcohol withdrawal, or delirium may occur *de novo* in an otherwise recovered alcoholic with established sobriety.

Informed consent

Three aspects of the informed consent process can be valuable tools in establishing a good doctor–patient relationship.

Bonding

Along with a statement of risks and benefits, the consent process is a declaration of clinical goals. A factual, caring presentation marks the beginning of a collaborative bond for patient and surgeon.

Teaching

Patients must be informed of discomfort associated with the procedure and about availability of pain relief. The need for intravenous therapy, an indwelling catheter, drains, and an endotracheal tube should be discussed, as well as the customary length of the operation and the anticipated time for recuperation in hospital and following discharge. The surgeon must provide information about necessary postoperative care including medication, diet, activity restriction, and medical visits. The patient should know of complications, including psychological difficulties, commonly associated with the procedure, and the risk of dying. Information should be addressed in a candid but constructive fashion. A visit to areas of the hospital dedicated to postoperative care can be beneficial, as is the opportunity for contact with other patients who have had similar surgery.

Observing

The informed consent process allows the surgeon an opportunity to observe the patient's mental status. Preoccupation with excessive detail may be evidence of anxiety or paranoia. Failed comprehension may signal encephalopathy, internal distraction, or deficient intellect. Dress and deportment are a statement of self-worth as well as personal management and socio-economic status. A despondent, tearful, or lethargic manner signals depression. Attempts by the patient at good-natured evasion of a proper alcohol and drug history may be a clue to pathologic denial. An ingratiating attitude coupled with criticism of former physicians is typical of paranoid individuals. Adjustment problems or evidence of greater psychopathology should be followed by formal psychiatric consultation and by social service intervention when there is a need for additional perioperative support.

Innovative surgery

The pace of medical science has made some former experimental procedures routine. New experiments with the artificial heart and transplantation of multiple organs have aroused the interest of ethicists and health policy planners. The economics of current health care have put a spotlight on quality of life aspects for all such interventions.

Patient selection

The selection process for emerging surgical technology is based on capacity to benefit from the procedure, degree of present need, and time of initial presentation. The need to establish a priority list among patients is a source of stress for physicians; this intensifies as patients die while waiting for treatment. Since patients with a specific medical handicap or psychosocial impairment may have differing levels of limitation, suitability for costly new procedures, such as heart transplantation, is best determined on an individual basis to avoid discrimination. Factors such as older age and prior alcoholism may not be valid reasons for excluding people who otherwise appear to have a good chance of recovery and long life.

Reduction of preoperative stress

Education and support

Emphasis should be on individual concerns. Patients need to know that pain is normal and that early postoperative mobilization is healthy. The surgeon should review past difficulty with specific analgesic drugs, anesthetic agents, or other medications essential to the operation. Patients should be encouraged to request pain medication as needed and they should be reassured about fears of addiction.

Disfigurement is a frequent source of worry but may be couched in understatement such as ‘Will I be able to wear a bikini?'. Other common worries concern sexual function, future childbearing capacity, return to active lifestyles, and loss of privacy. Some may wish to maintain access to aspects of their work. Dietary considerations, visiting arrangements, health care directives, and financial issues should be discussed. A preoperative visit with a medical social worker or dietitian may be helpful. Patients with histories of lengthy medical problems, such as those with juvenile-onset diabetes, often have firm opinions about their medical needs. When they do, preoperative discussion with the nursing staff is beneficial since it helps to accommodate to specific needs. It is equally important to shape expectations and to encourage patients to modify their lifestyles. Timely referral to a smoking cessation program or substance abuse clinic may make a profound difference to postoperative outcome.

Specialized intervention

A preoperative visit by the anesthetist has benefits compared with sedation alone. In a study of patients undergoing abdominal surgery, those who were taught about the normality of postoperative pain and encouraged to request analgesics when needed had greater postoperative comfort and used far fewer narcotics than those receiving no information about pain control. Subsequent studies have employed preoperative interventions such as support, teaching, hypnosis, guided imagery, music, and relaxation training, and have looked at varying measures of postoperative outcome. Although methodological problems exist, such as lack of ‘blind' controls, these studies document a major reduction of time in hospital. For example, in one study the hospital stay of elderly patients undergoing repair of a fractured femur was 12 days shorter in those who received additional care by a psychiatrist compared with a similar group treated a year earlier without psychiatric support. In another study of patients undergoing colorectal surgery, guided imagery significantly reduced postoperative anxiety, pain, and narcotic requirements, and increased patient satisfaction

as compared with controls.

Patients undergoing surgical procedures benefit from information that fosters healthy expectations and from behavioral or cognitive techniques that provide effective coping strategies and an enhanced sense of control. The challenge is to refine these interventions in a manner that allows for differences in individual coping style, variation in operative requirements, and nature of the relationship between the patient and members of the surgical team.

Some researchers have attempted to modify the risk of postoperative delirium. Supportive preoperative visits by a psychiatrist were shown in two studies to reduce the incidence of delirium following cardiac surgery. In a third study, preoperative psychiatric support combined with autohypnosis training failed to produce a statistically significant reduction in postcardiotomy delirium relative to controls with routine care. However, the number of patients who became delirious was insufficient to provide a conclusive result. In a more recent study, 64 patients undergoing cardiac surgery were informed by a nurse investigator of the possibility for unusual postoperative experiences and were better able to cope with changes in cognition or perception.

Preoperative participation of a psychiatrist may be especially helpful in high-risk procedures and when there is a critical demand for patient self-monitoring and compliance. Whenever the surgeon suspects significant psychopathology, or when there is a prior history of postoperative psychiatric difficulty or past medical non-compliance, a psychiatrist should be consulted.

Preoperative management of psychiatric disorders

Personality disorder

When a personality problem is identified, collaboration among primary physician, psychiatrist, and surgeon is necessary. Consistency is essential. The marginally adaptive patient should be enlisted in a specific care plan with a minimum of ambiguity. Good communication among caregivers helps with the demanding dependency of such patients and the emotions they may arouse. Paranoid and obsessive individuals manage best when one member of the team is designated as doctor in charge. The designated physician is ideally one who can relate to the patient and who can expend the time for repeated questions and detailed review of the care plan. At times one can advantageously enlist an interested family member who is a reassuring influence.

Major depressive disorder

If the patient is depressed, psychiatric help should be requested. It is often advisable to postpone the operation and to treat the depressive disorder. However, if a surgically correctable condition is strongly contributory to the mood disorder, there is little benefit in delay. In other instances the surgery may be urgent or the mood disorder intractable. Appropriate antidepressant medication should be given through the first preoperative day, and resumed postoperatively as soon as the patient can safely take sips by mouth. There are isolated reports of adverse interaction between tricyclics and anesthetic agents, but the morbidity of recurrent depression carries a greater risk. Because combined use of halothane and tricyclic antidepressants may increase catecholamine levels, sympathomimetic agents should be used with caution in such situations.

Since antidepressants are metabolized in the liver, careful dosing and measurement of serum levels of the drug are required in patients with abnormal liver function. Imipramine and amitriptyline are available for parenteral administration. Among second-generation antidepressants the selective serotonin receptor inhibitors (fluoxetine, paroxetine, sertraline), citralopam, and bupropion have been popular for their efficacy and infrequent adverse events. Many oral antidepressants can be converted for administration by rectal suppository. In all instances the anesthetist and surgeon should be informed of medication requirements and thought should be given to potential drug interactions. Although there has been some controversy about preoperative use of monoamine oxidase inhibitors, they can be administered safely. Numerous patients have successfully undergone surgery and reported their prior monoamine oxidase inhibitor treatment after the operation.

Monoamine oxidase inhibitors interfere with the breakdown of central nervous system depressants. Potentially fatal adverse reactions are known to occur with meperidine, atropine, or other anticholinergic agents and with barbiturates. However morphine, oxycodone, and codeine can be used postoperatively for analgesia. Because monoamine oxidase inhibitors increase the level of catecholamines in peripheral nerve endings and potentiate the effect of sympathomimetic agents, hypertensive crisis may result from the use of pressors and the diet must be low in tyramine (which is especially high in aged cheese and red wines). Hypertensive crisis may also occur when monoamine oxidase inhibitors are combined with other classes of antidepressants.

Anxiety disorder

Anxiety may be acute in onset and related to fear of surgery, or it may represent a chronic emotional disorder or personality trait. Some patients may not acknowledge distress, or may do so with difficulty because of social custom or personality style. It is best to ask the patient about special concerns. Some patients dread specific types of intervention, such as endotracheal tube placement. Since the average patient knows little about anatomy, much can be learned by eliciting misconceptions. It is helpful to know how patients have coped with prior surgery and with other stressful events. In a recent study of 1420 patients undergoing surgery at the University of Iowa Hospital and Clinics, the best predictor of postoperative psychological distress was found to be preoperative psychological distress.

Anxiety may be increased by an overzealous account of the planned procedure. However, an excessively paternal approach is insufficient to allow for the education and discussion necessary for the patient to make a judgement about the procedure and postoperative outcome.

Minor tranquillizers in the benzodiazepine class are a useful adjunct to psychological support. Shorter acting agents (for example alprazolam, lorazepam, oxazepam) are preferable for use in patients who are elderly or debilitated. Oxazepam and diphenhydramine are most easily metabolized in the liver and are the anxiolytic and sedative hypnotic agents of choice in patients with impaired hepatic function. The prescribing physician should be aware of the half-life and the potency of psychotropic agents and of the patient's past psychopharmacologic history. Patients with debilitating anxiety should be referred for psychiatric assessment.

Panic disorder

Panic disorder, with or without agoraphobia, responds effectively to alprazolam, selective serotonin reuptake inhibitors, imipramine, or monoamine oxidase inhibitors. Ideally, monoamine oxidase inhibitors are best avoided or gradually discontinued 2 weeks prior to the operation. However, they can be continued when there is insufficient time for withdrawal, when anxiety symptoms are severe, or if alternative agents have proved ineffectual. Although selective serotonin reuptake inhibitors have a slower onset of response than does alprazolam, they can be effectively instituted when surgery is elective. Sudden cessation following chronic administration of anxiolytics is often associated with rebound anxiety as well as by withdrawal symptoms. When oral medication cannot be administered, a parenterally administered benzodiazepine can be used to provide sedation. Intravenous lorazepam and oral alprazolam are equipotential for the treatment of generalized anxiety. Lorazepam is also available for sublingual administration.

Phobic disorder

For patients with simple phobias a combination of behavior therapy and anxiolytic agents can be helpful.

Hex or predilection to death

Hackett and Weisman give the example of a farmer whose certainty of postoperative fatality presaged his cardiovascular death 3 days after subtotal gastrectomy. These patients are noteworthy for the absence of anxiety or depression. When such a conviction is evident, surgery is best avoided if possible.

Functional psychosis

Patients who are psychotic often accommodate satisfactorily to the structure of a busy surgical service. Actively suicidal individuals require continuous close supervision by special duty caregivers. Antipsychotic agents should be administered in full dose throughout the pre- and postoperative period. Aliphatic substituted phenothiazines (e.g. chlorpromazine) are more likely to be associated with hypotension than the high potency neuroleptics (e.g. haloperidol). Newer antipsychotic drugs such as risperidone and olanzepine have a reduced incidence of neurologic side-effects. Patients should be managed in a simple direct manner aided, where possible, by supportive family members. The patient should be sheltered from stressful interpersonal relations. Competency should be established or appropriate guardianship arranged in consultation with hospital legal advisers.

Preoperative encephalopathy

Organic central nervous system disorders should be addressed with an appropriate search for underlying metabolic, infectious, and neurologic causes. An unexpected rise in serum ammonia is sometimes evident in patients in whom other liver function tests are relatively mildly impaired. Treatment depends on the cause. Standard techniques should be employed, with reference to clock, calendar, availability of special personal effects, and gentle review of the daily routine. Excessive sedation should be avoided. Supportive nursing techniques and family visits often reduce agitation; intravenous haloperidol can be given at 0.5 to 1.0 mg every hour as required. Caution is necessary when drugs are administered to patients with hepatic dysfunction. Diphenhydramine is a gentle treatment for insomnia. Benzodiazepines may cause confusion in the elderly, but some do benefit from low doses of short-acting agents.

Postoperative encephalopathy

Supportive care should be coupled with a search for specific etiology. Anesthetic agents or intolerance to specific analgesic agents or their metabolites (normeperidine delirium is an example) should be suspected as well as idiosyncratic reaction or toxicity to other medication. Depression of central nervous system function may be evidence of postoperative cardiopulmonary or infectious complication or of endocrinopathy. Postoperative delirium requires energetic treatment when agitation, mood lability, hallucinations, and delusions pose a threat to medical management. When delirium occurs in the intensive care unit, pharmacologic intervention may include intravenous morphine, haloperidol (which can be administered at beginning doses of 0.5 to 1.0 mg intravenously q1 h. prn), or lorazepam. A psychiatric consultant should visit daily, when possible, until the patient's sensorium is clear. Return of normal consciousness may be associated with feelings of shame or guilt among patients who retain memory for delusional material or perceptual aberrations. It is therefore wise to ‘debrief' patients as cognition returns to baseline and to explain events in a supportive fashion. Patients are comforted to know that delirium does not represent a weakness of character or spirit but is rather a product of understandable biologic events and environmental stress.

Although delirium was once a frequent complication of cardiovascular surgery, its incidence following coronary artery bypass grafting and cardiotomy has declined with improvement in surgical technology. The risk of delirium is increased with increasing age and lessened tolerance to decreased perfusion pressure, severity of physical illness, history of myocardial infarction, preoperative organic brain syndrome, duration of extracorporeal circulation, sustained mean arterial pressure under 50 mmHg during bypass, and postoperative hypotension requiring pressure pressors or an intra-aortic balloon pump. Interestingly, gender, type and route of anesthesia, and sleep deprivation were not risk factors. In a study of 23 patients undergoing aortic valve replacement, a previously unreported association was found between low preoperative cholesterol level, which was attributed to a probable catabolic state, and to postoperative psychopathology. A simple bedside test, the Abbreviated Mental Test, has proved both sensitive and specific for postoperative delirium.

Steroid-induced psychosis following organ transplantation has become less frequent. Cyclosporin encephalopathy has been most frequent among liver transplant recipients and is characteristically associated with other signs of toxicity such as tremor, impaired renal function, and elevated blood pressure.

Delirium may occur secondary to alcohol withdrawal in current drinkers or in the recovered alcoholic in the absence of recent alcohol abuse. Another cause of delirium is acute sensory deprivation. Complex partial seizures should also be considered in the differential diagnosis.

Alcoholism

The one reported drink per night may be from a bottomless glass. Preoperative detoxification is best whenever possible; however, prevention of withdrawal is a primary goal. Supplemental nutritional support should be instituted and full doses of chlordiazepoxide should be administered every 4 to 6 h in patients suspected of alcoholic withdrawal. Intravenous alcohol is an alternative. Patients with addictive disorders may provoke an angry response from otherwise caring physicians. Participation of a psychiatric consultant is desirable.

Recovered alcoholics may feel vulnerable and are often apprehensive about a stress-related relapse. Patients who are attending Alcoholics Anonymous may be especially vigilant about perioperative medications and the potential euphoriant effects of narcotic analgesics. When psychotropic agents are advisable, care should be taken to point out the medical indication. Some hospitals sponsor weekly Alcoholics Anonymous meetings. Supportive visits from ‘safe' companions may reduce perioperative stress.

Drug-addicted patients

Treatment is similar to that of the alcoholic patient. Barbiturate requirements for patients addicted to depressants are established with a test dose of phenobarbital. Narcotic addicts who require surgery should not undergo drug withdrawal until after the operation. Pain-relieving medication should be given in addition to that required for daily maintenance. After surgery, drug withdrawal can be approached in a standard method.

Pain intolerance

Inadequate analgesia is the most likely cause of persistent postoperative pain. Tolerance may occur among those receiving frequent daily analgesics over a period of weeks, but addiction is rare. Depression, anxiety, and tolerance to opioids increase postoperative analgesic needs and should be considered whenever pain persists in the absence of postoperative complications. Antidepressants of traditional narcotic adjuvants can be highly beneficial. Psychiatric consultants who are familiar with relaxation techniques and hypnosis can often provide symptom relief.

Awareness during operations

Blacher reported six patients in whom postoperative symptoms of irritability, preoccupation with death, and nightmares were associated with expressed doubts about their sanity. All had experienced wakefulness at some point during their operation, unknown to the anesthesiologist, and all benefitted once a link was established between that event and their postoperative anxiety. Some patients experience postoperative irritability and unease without conscious memory of their intraoperative awareness. Others can quote statements made by the surgical team during the operation.

The overall incidence of awareness during general anesthesia is 0.01 per cent. There is a much greater incidence in cases of major trauma. Awareness is also more likely to occur in obstetric cases and in cardiovascular surgery. The most severe psychiatric sequela is post-traumatic stress disorder, which can prove persistent and debilitating. Denial or avoidance of the issue by the surgical team further aggravates psychological distress. Medicolegal concerns should therefore not deter a timely and candid discussion of events by the surgeon and the anesthesiologist. Surgeons are also cautioned to avoid disparaging or frightening comments during operative procedures.

Further reading

Aberg T *et al.* Adverse effects on the brain in cardiac operations as assessed by biochemical psychometric and radiologic methods. *Journal of Thoracic and Cardiovascular Surgery* 1984; **87**: 99–105.

Abram HS, Gil BF. Prediction of post-operative psychiatric complications. *New England Journal of Medicine* 1961; **265**: 1163.

Anthony JC *et al.* Limits of the ‘Mini-Mental State' as a screening test for dementia and delirium among hospital patients. *Psychological Medicine* 1982; **12**: 397–408.

Ashton C *et al.*. Self-hypnosis reduces anxiety following coronary artery bypass surgery. A prospective randomized trial. *Journal of Cardiovascular Surgery* 1997; **38**: 69–75.

Bailey AR, Jones JG. Patients' memories of events during general anaesthesia. *Anaesthesia* 1997; **52**: 460–76.

Barnason S *et al.* The effects of music interventions on anxiety in the patient after coronary artery bypass grafting. *Heart and Lung* 1995; **March/April**: 124–32.

Blacher RS. On awakening paralyzed during surgery. *Journal of the American Medical Association* 1975;**234**: 67–8.

Clarke DM *et al.* The recognition of depression in patients referred to a consultation-liaison service. *Journal of Psychosomatic Research* 1995; **39**: 327–34.

Chonchubhair AN *et al.* Use of the abbreviated mental test to detect postoperative delirium in elderly people. *British Journal of Anaesthesia* 1995; **75**: 481–2.

Cole MG *et al.* Effectiveness of interventions to prevent delirium in hospitalized patients: a systematic review. *Canadian Medical Association Journal* 1996; **155**: 1263–8.

Craven JL, Bright J, Dear CL. Psychiatric, psychosocial and rehabilitative aspects of lung transplantation. *Clinics in Chest Medicine* 1990; **11**: 247–57.

Cunningham MF *et al.* Introducing a music program in the perioperative area. *AORN Journal* 1997; **66**: 674–82.

Dubin WR, Field AL, Gastfriend DR. Postcardiotomy delirium: a critical review. *Journal of Thoracic and Cardiovascular Surgery* 1979; **77**: 586–99.

Dyer CB *et al.* Postoperative delirium. *Archives of Internal Medicine* 1995; **155**: 461–5.

Egbert LD, Battit GE, Welch GE, Bartlett MK. Reduction of postoperative pain by encouragement and instruction of the patient. *New England Journal of Medicine* 1964; **270**: 825.

Engvall WR. Awareness during anesthesia. *Seminars in Anesthesia* 1988; **7**: 55–61.

Enqvist B, Bjorklund C, Engman M, Jakobsson J. Preoperative hypnosis reduces postoperative vomiting after surgery of the breasts. A prospective, randomized and blinded study. *Acta Anaesthesiologica Scandinavica* 1997; **41**: 1028–32.

Gath D, Cooper P, Day A. Hysterectomy and psychiatric disorder: I. Levels of psychiatric morbidity before and after hysterectomy. *British Journal of Psychiatry* 1982; **140**: 335–50.

Jarvik ME. Drugs used in the treatment of psychiatric disorders. In: Goodman LS, Gilman A, eds, *The pharmacological basis of therapeutics*, 3rd edn, pp. 191–8. Collier-Macmillan, London, 1969.

Johnston M. Anxiety in surgical patients. *Psychological Medicine* 1980; **10**: 145–52.

Johnston M. Pre-operative existential status and post-operative recovery. In: Wise TN, ed. *Advances in psychosomatic medicine*. Guggenheim FG, ed. *Psychological aspects of surgery*, pp. 1–20. Karger, Basle, 1986.

Layne OJ, Yudofsky SC. Postoperative psychosis in cardiotomy patients: the role of organic and psychiatric factors. *New England Journal of Medicine* 1971; **284**: 518.

Lazarus HR, Hagens TH. Prevention of psychosis following open heart surgery. *American Journal of Psychiatry* 1968; **124**: 1190.

Levitan SJ, Kornfeld DK. Clinical and cost benefits of liaison psychiatry. *American Journal of Psychiatry* 1981; **138**: 790–3.

Lundberg SG, Guggenheim FG. Sequelae of limb amputation. In: Wise TN, ed. *Advances in psychosomatic medicine*. Guggenheim FG, ed. *Psychological aspects of surgery*, pp. 199–210. Karger, Basle, 1986.

Marks RM, Sachar EJ. Under-treatment of medical inpatients with narcotic analgesics. *Annals of Internal Medicine* 1973; **78**: 173.

Meninger KA. Polysurgery and polysurgical addiction. *Psychoanalysis Quarterly* 1934; **3**: 173.

Merrikin KJ, Overcast TD. Patient selection for heart transplantation: When is a discriminating choice discrimination? *Journal of Health Politics, Policy and Law* 1985; **10**: 7–32.

Millar K *et al.* Assessment of preoperative anxiety: comparison of measures in patients awaiting surgery for breast cancer. *British Journal of Anaesthesia* 1995; **74**: 180–3.

Mirassou MM. Rectal antidepressant medication in the treatment of depression. *Journal of Clinical Psychiatry* 1998; **59**: 29.

Mumford E, Schlesinger HJ, Glass GV. The effects of psychological intervention on recovery from surgery and heart attacks: an analysis of the literature. *American Journal of Public Health* 1982; **72**: 144–51.

Naber D, Bullinger M. Neuroendocrine and psychological variables relating to postoperative psychosis after open-heart surgery. *Psychoneuroendocrinology* 1985; **10**: 315–24.

O'Hara MW *et al.* Psychological consequences of surgery. *Psychosomatic Medicine* 1989; **51**: 356–70.

Owen JF, Hutelmyer CM. The effect of preoperative intervention on delirium in cardiac surgical patients. *Nursing Research* 1982; **31**: 60–2.

Ray C. The surgical patient: psychological stress and coping resources. In: Biver JR, ed. *Social psychology and behavioral medicine*. John Wiley, New York, 1982.

Shafer A *et al.* Preoperative anxiety and fear: a comparison of assessments by patients and anesthesia and surgery residents. *Anesthesia and Analgesia* 1996; **83**: 1285–91.

Surman OS. Post noxious desensitization: some clinical notes on the combined use of hypnosis and systematic desensitization. *American Journal of Clinical Hypnosis* 1979; **22**: 54–60.

Surman OS. Psychiatric aspects of organ transplantation. *American Journal of Psychiatry* 1989; **146**: 972–81.

Surman OS, Purtilo R. Organ transplants for ugly ducklings: II Allocation of scarce resources to unusual recipients. *Psychosomatics* 1992; **33**: 203–12.

Surman OS, Hackett TP, Silverman EL, Behrendt DM. Efficacy of psychotherapy for patients undergoing cardiac surgery. *Archives of General Psychiatry* 1974; **30**: 830.

Trzepecz P *et al.* Practice and guideline for the treatment of patients with delirium. *American Journal of Psychiatry - Supplement* 1999; **156**(5).

Tusek DL *et al.* A significant advance in the care of patients undergoing elective colorectal surgery. *Diseases of the Colon and Rectum* 1997; **40**: 172–8.

Weisman AD, Hackett TP. Psychosis after eye surgery: establishment of specific doctor–patient relationship in prevention and treatment of black patch delirium. *New England Journal of Medicine* 1958; **258**: 1284.

Weisman AD, Hackett TP. Predilection to death. Death and dying as a psychiatric problem. *Psychosomatic Medicine* 1961; **23**: 232–56.

Wilson SI, Vaughan RW, Stephen CR. Awareness, dreams and hallucinations associated with general anesthesia. *Anesthesia and Analgesia Current Research* 1975; **54**: 606–17.

13.1 Conventional radiology

Basil J. Shepstone

[The chest](#)
[Tumours of the lung and mediastinum](#)
[Infections of the lungs](#)
[Obstructive airways disease](#)
[Sarcoidosis](#)
[The pneumoconioses](#)
[The pleura](#)
[The heart](#)
[Mediastinal opacities](#)
[The postoperative chest](#)
[The genitourinary tract](#)
[Adrenals](#)
[Further reading](#)

Even though it is now over a century since the discovery of X-rays by Wilhelm Roentgen on 8 November 1895 and in spite of the major advances in imaging techniques which are presented in subsequent sections, plain radiographs still command an important place in surgical practice. As the value of plain radiography will be illustrated extensively in specialist chapters, this section will concentrate mainly on the chest radiograph and plain films of the urinary tract.

The chest

The posteroanterior chest radiograph is still the mainstay of the standard medical examination. In hospital practice a routine film on admission will provide a baseline for later comparison, following surgery, when infection, oedema, or collapse might supervene. The lateral projection is no longer carried out routinely, but may be necessary to localize a lesion anatomically (for interventional purposes) or to obtain a view of a structure not seen on the posteroanterior view, for example the right ventricular border, the left atrial border, the oblique fissure, the posterior costophrenic angle, or the spine. It is also useful for the differential diagnosis of mediastinal masses.

The key to successful diagnosis of the chest radiograph is its systematic examination, which must include the soft-tissue shadows (breast, muscles, cutaneous tissues), the diaphragmatic and subdiaphragmatic areas (of great importance to the surgeon), the bony cage (ribs, clavicles, scapulae, spine, proximal humeri), the lung fields, the heart, and mediastinum. Whenever possible, bilateral and symmetrical structures must be carefully compared.

Although it is the duty of the radiographer to produce a technically perfect product, it is still imperative that the film reader checks that this is achieved. For example, positioning must be exact with both left and right costophrenic angles and apices within the film area. The scapulae must not overlap the bony cage. Penetration, as tested by the thoracic vertebrae just being visible through the heart shadow, must be assessed. Underpenetration may produce a useless white film, whereas an overpenetrated film, although looking very black, can be salvaged by the use of a bright light. Rotation with respect to the cassette, which can cause discrepancies in the appearance of the left and right lung fields, must be checked by making sure that the angles made by the clavicle with the vertical are the same on each side. Finally, the degree of inspiration must be checked by counting the number of ribs, preferably on the right side, that appear above the right hemidiaphragm. Ten posterior ribs are the norm. If more than this number can be counted and there is also flatness of the hemidiaphragms, with a decrease in the number and size of the lung markings, chronic obstructive disease may be present.

The routine view is, of course, taken in the posteroanterior position (which refers to the direction of the beam) so that the anteriorly situated heart can be as close to the film as possible (Fig. 1). This enables an accurate estimation of the so-called cardiothoracic ratio to be made (the ratio of the width of the cardiac shadow at its widest point to the diameter of the thorax at its widest point). The normal cardiothoracic ratio is less than 50 per cent.

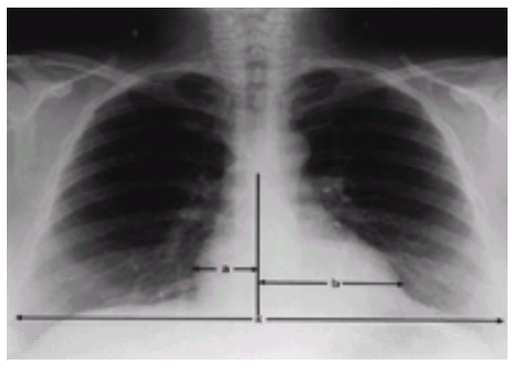


Fig. 1. A normal posteroanterior chest radiograph showing how to calculate the cardiothoracic ratio (equal to $(a+b)/x$), which *should* normally be less than 0.5.

Tumours of the lung and mediastinum

The routine chest radiograph is still the first investigation of choice in the diagnosis of bronchial carcinoma (Fig. 2). Either the primary tumour or one of its secondary effects or complications may be seen, or both. Among the latter is identification of metastases to the mediastinum or elsewhere.

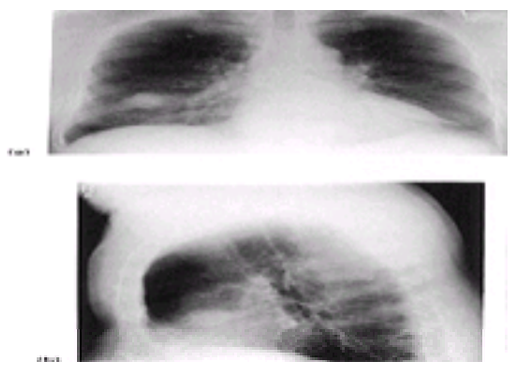


Fig. 2. (a) Primary bronchial carcinoma presenting as a solitary nodule in the right lung (by courtesy of Dr Fergus Gleason). (b) Primary bronchial carcinoma presenting as a solitary nodule in the right lung (by courtesy of Dr Fergus Gleason).

The primary tumour may present as an opaque nodule, but tumours are not the only cause of pulmonary nodules and the differential diagnoses include abscesses or granulomas, infarctions, haematomas, focal collagen disease, retention cysts, sequestrated segments, arteriovenous malformations, and hamartomas. Also, the lesion may be pleural based, when encysted pleural effusions and fibromas are possible. Finally, the lesion may not lie within the pulmonary parenchyma itself, but on the

surface of the body, e.g. a large wart or mole (Fig. 3).



Fig. 3. A skin wart appearing as a solitary nodule in the right upper zone of an otherwise normal chest radiograph.

Malignant tumours may be primary or secondary carcinomas, but may also be lymphomas or plasmacytomas.

A bronchial carcinoma often blocks the bronchus from which it arises and in this case leads to absorption collapse of the affected segments or lobes (Fig. 4). Superimposed infection then also usually occurs. Unless there is another good reason for the collapse in an adult (e.g. bronchiectasis with inspissated pus, asthma with mucoid impaction, or inhaled foreign body) bronchial carcinoma must always be suspected. Cavitating lesions may be due either to tumours or abscesses, although an area of infarction can also break down (Fig. 5).



Fig. 4. Collapse of the right upper lobe due to a bronchial carcinoma, with elevation of the horizontal fissure and the right hilum.



Fig. 5. A cavitating tumour at the left base.

A hilar mass may represent metastases to the mediastinal lymph nodes (Fig. 6) and a pleural effusion may also arise as a result of a malignant tumour (Fig. 7), although infection and infarction can also cause unilateral effusions (exudates). Failure of major organ systems like the heart, kidneys, and liver are also common causes of effusions, but these are usually bilateral (transudates). A hilar carcinoma or metastases to the mediastinal nodes can involve the phrenic nerve, with subsequent elevation and paresis of a hemidiaphragm (Fig. 8).

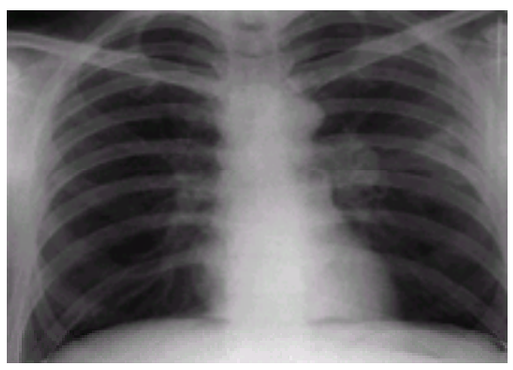


Fig. 6. An enlarged left hilar node due to an associated bronchial carcinoma appearing as an amorphous cavitating lesion.



Fig. 7. A right malignant pleural effusion in a woman who has undergone right mastectomy.

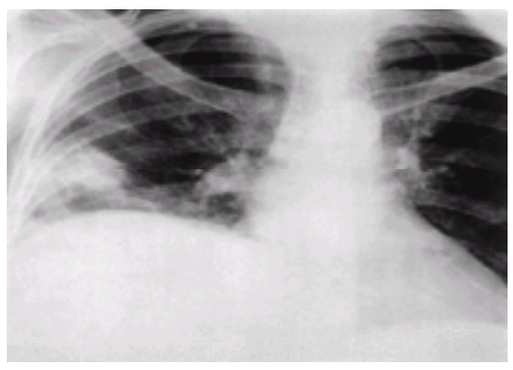


Fig. 8. A raised right hemidiaphragm due to phrenic-nerve paralysis in a patient with carcinoma of the right lung.

In an apical or so-called ‘Pancoast' tumour there is often lysis of the adjacent ribs. Metastases to bone in general may present as lytic or sclerotic lesions, depending on the aggressiveness of the tumour. Alternatively, they may present with pathological fractures.

The lungs are a common site for haematogenous metastases, where they often produce multiple round nodules of different sizes (Fig. 9) (as opposed to granulomatous disease, where the nodules are usually small and of similar size (Fig. 10)). Alternatively, lymphatic spread may show a reticular, almost fibrous-looking appearance, often with septal lines and is referred to as lymphangitis carcinomatosa (Fig. 11).



Fig. 9. Pulmonary metastases in a patient with a primary renal carcinoma.

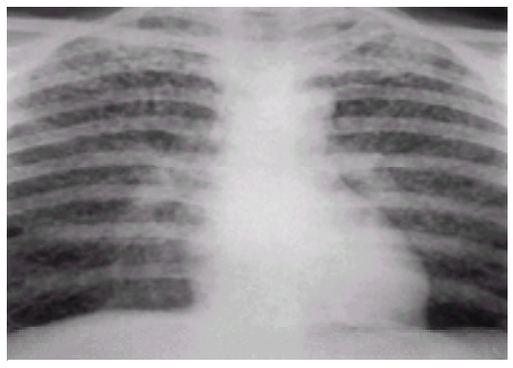


Fig. 10. Miliary tuberculosis.



Fig. 11. Lymphangitis carcinomatosa in a patient with a primary carcinoma of the pancreas. (Note the septal lines, which are not pathognomonic of interstitial pulmonary oedema. They are also seen in the pneumoconioses.)

Lymphoma usually affects the mediastinum. Hodgkin's disease, which is the most common of the lymphomas, normally affects the paratracheal glands, leading to local enlargement and generalized widening of the superior mediastinum (Fig. 12). Leukaemia and sarcoidosis can, however, lead to similar appearances, although sarcoidosis classically enlarges the hilar glands.

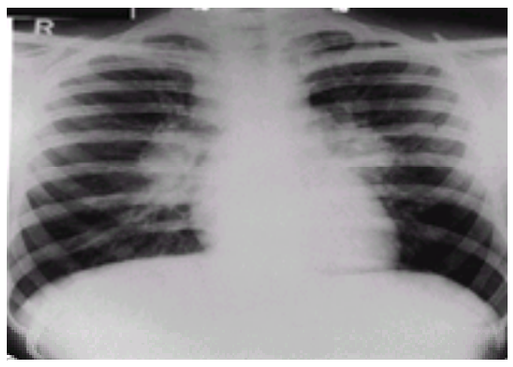


Fig. 12. Bilateral enlargement of the hilar nodes and the right paratracheal nodes in a patient with leukaemia. These appearances are also encountered in lymphoma and sarcoidosis.

Infections of the lungs

These usually present as diffuse, water-dense opacities (Fig. 13). Lobar pneumonia is now not often seen, but when it is, the opacity usually outlines an entire lobe. If resolution is incomplete, fibrosis may appear and bronchiectasis may be seen as a late sequel. The more common bronchopneumonia gives similar densities, but they are usually more patchy and are seen mainly in the lower lobes. Pneumonitis gives a similar picture, with the opacities usually confined to one segment of a lobe.

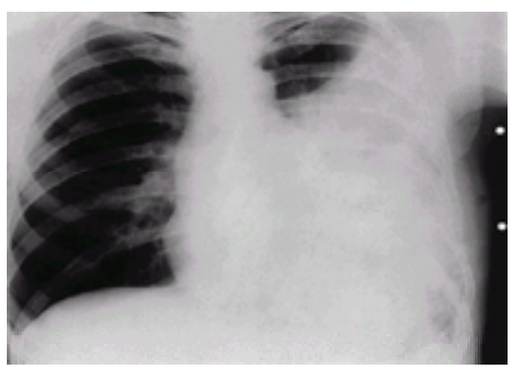


Fig. 13. Widespread left-sided pneumoniac consolidation.

The very common acute or chronic bronchitis may show nothing on chest radiography because bronchi are air-filled structures superimposed on an air-filled background. Radiographic principles dictate that a structure or an area of pathology can be seen only if its radiographic density is different from that of an adjacent tissue.

In bronchiectasis, however, the bronchi do become thick enough to be seen and so may be seen as end-on ring shadows or else small cystic areas, usually at the bases (Fig. 14). Careful computed tomography is necessary before resection is considered.



Fig. 14. Lower-zone bronchiectasis (with collapse of both right middle and lower lobes due to inspissated pus).

Infected areas, especially those distal to an occluded bronchus, may break down to form local abscesses with fluid levels (Fig. 15).

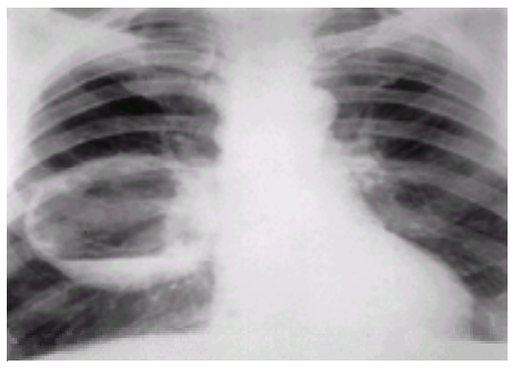


Fig. 15. A large pyogenic cavitating abscess in the right lung.

The characteristic radiographic appearance of tuberculosis in children is an area of water density with enlarged hilar glands in the affected side. It is important to remember that tuberculosis is the only infection which can enlarge hilar lymph nodes enough to make them visible on the routine chest radiograph. Alternatively, the disease can present with a miliary pattern, consisting of multiple small round opacities of similar size throughout the chest (Fig. 10). In adult infections, the disease has a tendency to spread to the posterior segment of the upper lobe and appear as an apical lesion, which can cavitate. A mycetoma may later appear in the cavity (Fig. 16).

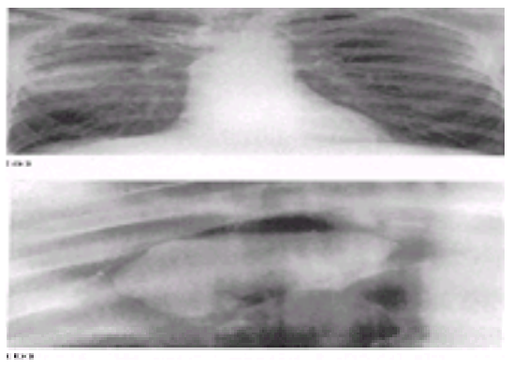


Fig. 16. (a) An aspergilloma in an old tuberculous cavity at the apex of the right lower lobe. (b) An aspergilloma in an old tuberculous cavity at the apex of the right lower lobe.

In patients who are immunosuppressed for any reason, but especially as a result of AIDS, unusual opportunistic infections may occur, for example those from *Pneumocystis carinii* or cytomegalovirus (Fig. 17).

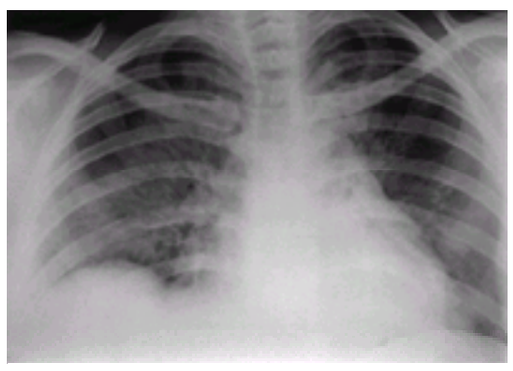


Fig. 17. *Pneumocystis carinii* infection in a patient with AIDS.

Obstructive airways disease

Obstructive airways disease includes the various types of emphysema and asthma and presents radiologically as elongated, translucent lung fields with flattened diaphragms, prominent hilar shadows, and a decrease in the number and size of the lung markings. In emphysema, bullae may occur. These are translucent areas where the alveolar walls have broken down and appear as black areas which are devoid of lung markings and are delineated by fine, curvilinear markings (Fig. 18).

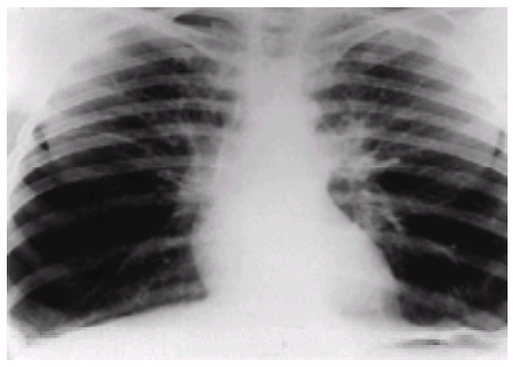


Fig. 18. An elderly man with bullous emphysema.

Sarcoidosis

Like tuberculosis, this disease is characterized by non-caseating granulomatous lesions which can affect the lungs, skin, eyes, and even bone. In the lungs, the disease classically presents with symmetrical, bilateral hilar, and paratracheal gland enlargement (Fig. 12). If the disease progresses, the lung fields may demonstrate a nodular-reticular pattern (Fig. 19), which proceeds to fibrosis and emphysema.

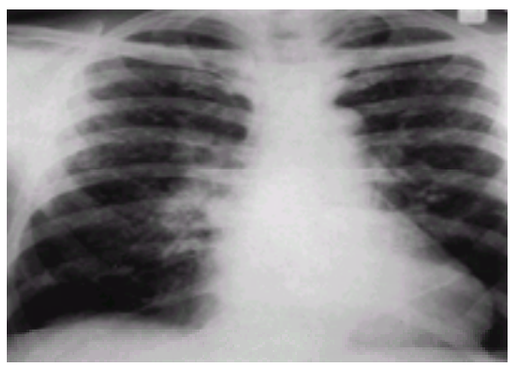


Fig. 19. The nodular-reticular pattern characteristic of sarcoidosis (some hilar lymphadenopathy is still visible).

The pneumoconioses

This is a collective name for a group of conditions usually arising in workers from industrial enterprises such as mining where foreign substances are inhaled. These include silica, iron, tin, talc, asbestos, and beryllium. Silicosis is perhaps the most common of the pneumoconioses. In the early stages there is a widespread small-nodular pattern, which once again may proceed to fibrosis and bullous emphysema. In silicosis the fibrotic areas often occur bilaterally and symmetrically in the upper lung fields and are ‘geometric' in shape—so-called conglomerate masses characteristic of the entity known as progressive massive fibrosis (Fig. 20).



Fig. 20. A retired coal miner with progressive massive fibrosis due to silicosis. The symmetrical rhomboid opacities in the upper zones are called ‘conglomerate masses'.

The pleura

Pleural tumours are rare, but the malignant pleural tumour mesothelioma occurs in workers exposed to asbestos. Such workers may also develop pleural or diaphragmatic calcification and are prone to develop asbestosis *per se*, which gives a characteristically diffuse reticular pattern to the lung fields.

Pleural effusions may be free or encysted (Fig. 7 and Fig. 21, respectively). When free, they manifest as basal opacities which initially fill the costophrenic angles and then rise up the hemithorax towards the axilla, usually with a meniscus. They may be seen following infections, such as pneumonia or tuberculosis, infarctions, or malignancy. They are also seen as complications of cardiac failure, hypoproteinaemic states (like hepatic failure), and renal failure (the ‘uraemic lung'). Empyema, or pus in the pleural space, which was common in tuberculosis, is now rarely seen as a postpneunomic complication.



Fig. 21. An encysted pleural effusion in the right midzone following an infection.

Spontaneous pneumothorax is common in young people and, although sometimes ascribed to the rupture of a bulla, its aetiology is usually unknown (Fig. 22).



Fig. 22. A left-sided pneumothorax in a young man with some shift of the mediastinum to the right.

Pneumothorax may also be due to trauma and is often associated with rib fractures. Tension pneumothorax occurs when air continues to enter the space, but cannot escape because the tear in the pleura acts like a one-way valve. It is therefore essential to observe any shift in the mediastinal structures towards the opposite side and, if so, to equilibrate the pressures immediately.

The heart

The posteroanterior chest radiograph can yield information about the cardiac diameter, enlargement of individual chambers, and the state of the pulmonary vasculature. The cardiothoracic ratio has already been considered.

In the posteroanterior view the right heart border is formed by the right atrium with the superior vena cava entering it from above. On the left side three protuberances may be seen. The top one is the aortic knuckle, the next one the pulmonary outflow tract (opposite the left main pulmonary artery which constitutes the left hilum—normal lymph nodes and bronchi are invisible), and the lower one is the border of the left ventricle. The left atrium lies behind the heart, but occasionally the prominent auricular appendage of an enlarged left atrium will form a further protuberance between the pulmonary outflow tract and the left ventricular border. On the lateral view, the right ventricle forms the anterior border and the left atrium most of the posterior border, where it is a close relation of the invisible oesophagus.

The diagnosis of left-ventricular enlargement is made on the posteroanterior view when its border moves outwards and the apex downwards. Right-ventricular enlargement is diagnosed on the lateral view by noting that the anterior border has risen up higher behind the sternum. On the posteroanterior view, it manifests as a lateral shift of both the left and right heart borders, but with the apex tilted upwards (forming the imaginatively named ‘boot-shaped heart') (Fig. 23). In biventricular enlargement the left and right borders are also displaced laterally, but the apex stays in its normal horizontal plane. The right-ventricular component is again observed on the lateral view, where left-ventricular enlargement shifts the lower posterior border too little for detection.

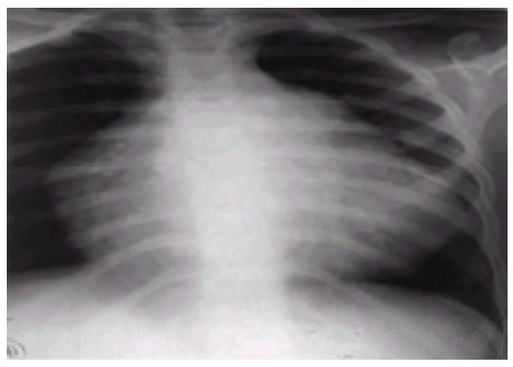


Fig. 23. A child with Fallot's tetralogy showing the misnamed ‘boot-shaped' heart and oligaemic lung fields.

Isolated right atrial enlargement is very rare, but would shift the right-hand border laterally. Left atrial enlargement, arising mainly from mitral-valve disease, forms a double shadow behind the heart as the enlarged chamber penetrates posteriorly and towards the right, surrounded by air-filled lung which provides the contrast. It will also splay the carina, resulting in a left main bronchus pointing laterally rather than downwards towards the left costophrenic angle. The proof of an enlarged left atrium is obtained from looking at a lateral chest after the patient has swallowed some barium. The enlarged left atrium will then indent the contrast-filled oesophagus as it curves round the posterior border of the heart towards the hiatus (Fig. 24). All chamber enlargements can, of course, also be assessed by echocardiography.



Fig. 24. Lateral chest with barium showing indentation of the barium column in a patient with a large left atrium due to mitral stenosis.

Pulmonary appearances in heart disease

The appearance of the lungs provides major clues to the diagnosis of heart disease. Normally the lower-zone vessels are more prominent and larger than the upper-zone vessels and the outer rim of the chest is usually free of lung markings. Because of the disposition of the left and right pulmonary arteries and bronchi, the left hilum is slightly higher than the right and as they emerge from the hila, the main right and left pulmonary arteries have a largely vertical configuration. The pulmonary veins, on the other hand, return to the left atrium, which is about 5 cm below the hila. The veins returning from the upper zones therefore run vertically, whereas those draining the lower zones run horizontally. It is, thus, not possible to differentiate between arteries and veins in the upper half of the lung fields, whereas in the lower half, arteries are vertically disposed and veins horizontally.

In pulmonary venous hypertension (the chief causes of which are mitral-valve disease and left-ventricular failure) both the vertical upper-zone and the horizontal lower-zone veins become larger. The latter are seen coming in right from the periphery, an area where there are normally no markings. As the pressure rises, pulmonary oedema will occur, which first fills the interstitial and then the alveolar spaces. The lower zones become hazy and vessel differentiation is lost. This and true upper-zone venous enlargement contribute to the characteristic ‘upper-lobe diversion', when the upper-lobe vessels appear larger and more prominent than the lower-lobe vessels. Interstitial oedema may appear as septal lines in the costophrenic angles, but these are short-lived as water, accumulating on each side, soon obliterates these distended interlobular lymphatics. A much more certain sign is peribronchial oedema or ‘cuffing' when the normally invisible end-on bronchi become visible as black holes because of the surrounding fluid—a variant of the so-called ‘air bronchogram' (Fig. 25). There may be pleural effusions to a greater or lesser degree, or if no treatment is offered, full-blown perihilar alveolar oedema which presents as widespread cotton-wool-like opacities in the central areas (Fig. 26).

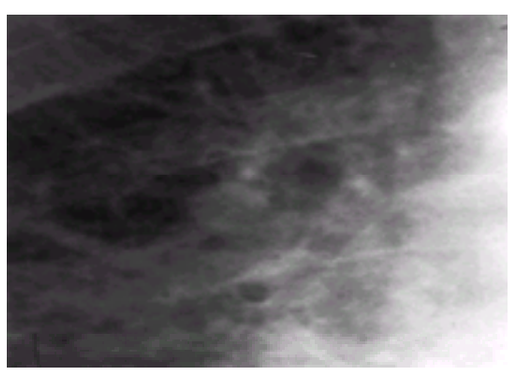


Fig. 25. Detail of a chest radiograph of a patient with pulmonary oedema to demonstrate ‘peribronchial cuffing'—a variant of the ‘air-bronchogram' sign.



Fig. 26. Florid alveolar pulmonary oedema, in this case due to congestive cardiac failure, but which is not pathognomonic of this condition as it can occur wherever there is alveolar damage.

In pure pulmonary arterial hypertension, which arises when there is either increased pulmonary blood-flow (e.g. anaemia, thyrotoxicosis, left-to-right shunts) or peripheral arterial close-down (e.g. primary pulmonary arterial hypertension, pulmonary emboli, chronic obstructive lung disease) there is so-called pulmonary plethora

with an enlarged pulmonary outflow tract, left and right main pulmonary arteries, and the rest of the vertically disposed arterial tree (Fig. 27). Obviously it may also arise from pulmonary venous hypertension and the two varieties often occur together, usually in left-heart failure.



Fig. 27. A patient with pulmonary arterial hypertension due to left-to-right shunt. Note the large pulmonary outflow tract.

Decreased-pulmonary flow (pulmonary oligaemia) is seen in obstruction in the pulmonary outflow tract at or below the pulmonary valves. The classical example is Fallot's tetralogy (Fig. 23).

Pericardial perfusions and cardiac calcifications

These are traditionally classified under the headings inflammatory, non-inflammatory, and malignant, and in fact the causes are very similar to those producing pleural effusions.

The common inflammatory causes are suppurative, viral, tuberculotic, and rheumatic. Non-inflammatory causes include heart failure, uraemia, and myocardial infarction. The effusion could also be due to blood in the pericardial sac from trauma or from a ruptured aortic or left-ventricular aneurysm.

Constrictive pericarditis may be a sequel to viral or tuberculous pericarditis and is sometimes seen after haemorrhage into the pericardium or in one or two of the collagen diseases. Subsequent fibrosis of the pericardium may lead to cardiac tamponade and right-heart failure. About half of the cases lead to calcification (Fig. 28).

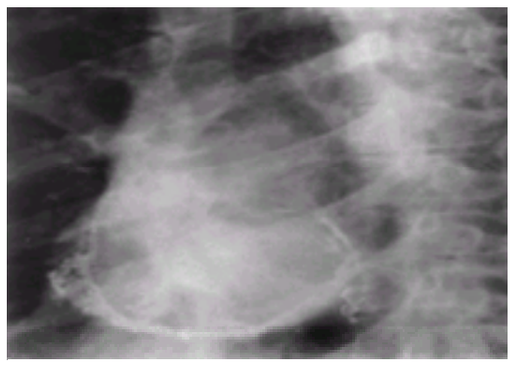


Fig. 28. An old oblique chest film showing a calcified pericardium due to tuberculosis.

Cardiac calcification may also be seen in the mitral and aortic valves, the coronary arteries, and in the left atrium, where it occurs as mural thrombus (Fig. 29).

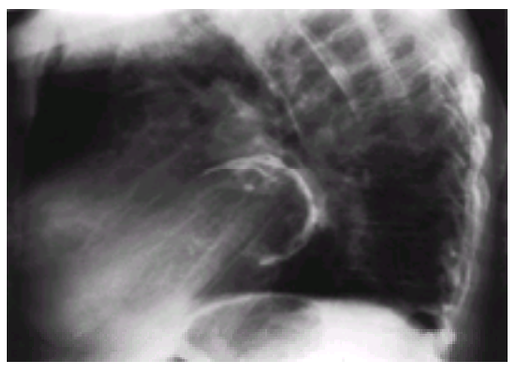


Fig. 29. Calcified left-atrial thrombus.

Mediastinal opacities

A lateral view is mandatory for the differential diagnosis of mediastinal opacities, the usual approach being to separate those occurring in the anterior, middle, and posterior compartments.

Anterior compartment

The normal structures in this compartment are the thymus, retrosternal lymph nodes, the ascending aorta and, occasionally, a retrosternal extension of the thyroid.

Opacities occurring in this compartment may therefore be due to thymomas (Fig. 30), enlarged lymph nodes, aneurysm of the ascending aorta, or retrosternal goitre (Fig. 31). Other possibilities are dermoid tumours, pericardiocoelomic cysts, parathyroid adenomas, and hernia of Morgagni.

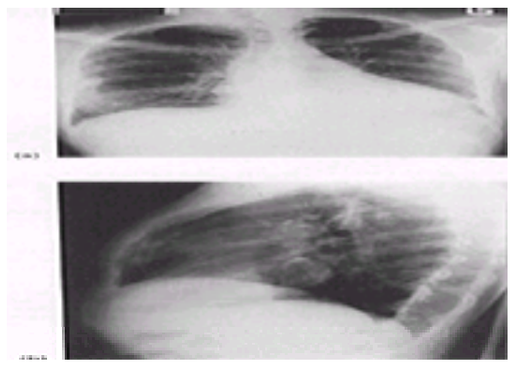


Fig. 30. (a) Posteroanterior chest film of a patient with myasthenia gravis who was found to have a thymoma occupying the anterior translucent space (by courtesy of Dr Fergus Gleeson). (b) Lateral chest film of a patient with myasthenia gravis who was found to have a thymoma occupying the anterior translucent space (by courtesy of Dr Fergus Gleeson).



Fig. 31. Posteroanterior chest radiograph of a patient with a retrosternal goitre which is displacing and compressing (a fundamental sign) the trachea.

Middle compartment

Middle-mediastinal structures include the aortic arch and its branches, the pulmonary artery, the inferior and superior venae cavae, and the heart.

Anomalies may be due to a big left atrium, aortic aneurysm, bronchial cyst, enlarged hilar lymph nodes (due to lymphoma, leukaemia, sarcoidosis, tuberculosis, or metastases).

Posterior compartment

Structures normally occurring in this compartment are the trachea and bronchi, oesophagus, descending aorta, lymph nodes, vagi, spine, and nerves emerging through the intervertebral foramina.

Abnormalities include tumours of neurogenic origin (neuro-fibroma, ganglioneuroma, neuroblastoma, myelocele, and meningo-myelocele), paravertebral abscess, aneurysm of the descending aorta, sequestrated lung segments, reduplication cysts of the oesophagus, loculated pleural effusions, corrective-tissue tumours, Zenker's diverticulum, and oesophageal lesions (achalasia (Fig. 32), hiatus hernia (Fig. 33), leiomyoma, and sub- or epiphrenic diverticula).



Fig. 32. Achalasia of the oesophagus with a fluid level.

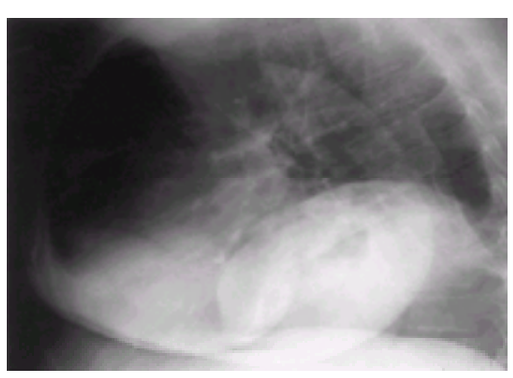


Fig. 33. Lateral chest of an obese woman with a large hiatus hernia occupying most of the posterior mediastinum.

The postoperative chest

After general surgery, poor respiratory effort and retained secretions lead to local collapse, often referred to as ‘plate atelectasis', which usually appears as horizontal lines in the lower zones. Pleural effusions are also common, but both of these entities resolve soon after surgery. General anaesthesia may lead to aspiration pneumonia or even to the full-blown adult respiratory distress syndrome. The radiographic appearances of this syndrome are indistinguishable from pulmonary oedema due to cardiac failure, fluid overload, or any form of alveolar damage. Pulmonary emboli may lead to an oligaemic area of lung with vessel cut-off, but it is usually diagnosed on pulmonary perfusion–ventilation scintigraphy. When the resulting infarction becomes established it manifests as a water-dense wedge which may cavitate.

A subphrenic collection of pus, blood, or fluid can lead to elevation of the hemidiaphragm and an associated ‘sympathetic' pleural effusion. Air under the diaphragm is

common after abdominal surgery, but it may arise from an anaerobic infection or perforation of a viscus. Occasionally, it may arise from the lung or via the vagina in women.

The hallmark of a thoracotomy is a resected 4th, 5th, or 6th rib. Partial regeneration of the rib can take place. After a lobectomy or a sublobectomy resection, the remaining lung should expand to fill the volume previously occupied by the resected segment. This will be accompanied by an appropriate shift of landmarks. However, in spite of drainage, pleural effusions, empyema, and haemothorax can occur and it is not uncommon to see a mixture of expanded lung and organizing fluid densities (Fig. 34). Poor closure of a bronchus can lead to bronchopleural fistula and early surgical emphysema in the soft tissues is a common occurrence. After pneumonectomy there is a combination of hyperinflation of the remaining lung with consequent shift of the mediastinum towards the side of the operation and the accumulation of fluid in the pneumonectomy space. Any chest operation can lead to phrenic nerve damage and elevated hemidiaphragm, but the nerve is often crushed deliberately to reduce the size of the treated hemithorax.



Fig. 34. Chest radiograph of a man who has undergone a left upper lobectomy.

Open-heart surgery is done via a sternal split, radiographic evidence for which is the presence of wire sternal sutures. A widened mediastinum and associated pleural effusions are common and air may enter any of the pleura, pericardium, or peritoneum. Sudden widening of the mediastium may indicate haemorrhage or an aortic dissection. The postpericardotomy syndrome is characterized by pain, fever, and pleurisy and the cardiac outline is enlarged by a pericardial effusion visible with echocardiography.

The genitourinary tract

Plain abdominal radiographs are essential before embarking upon contrast studies of the kidneys. As the perinephric fat is translucent, the kidneys are often visible lateral and parallel to the psoas lines, opposite T12, L1, and L2 and with the left kidney about 1.5 cm higher than the right. As the perirenal fat line persists after nephrectomy, the evidence for this operation is a resected rib and not an absent renal outline.

Even on plain film it may be possible to recognize an enlarged or shrunken kidney. The latter could be due to chronic pyelonephritis, chronic glomerulonephritis, or renal ischaemia, whereas enlargement may be due to hydronephrosis, a tumour, or simply to compensatory hypertrophy. Bilaterally enlarged kidneys maybe due to polycystic disease or to bilateral hydronephrosis. Kidneys are also enlarged in acute situations such as acutepyelonephritis.

Calcification in the renal area is very common and usually due to renal calculi, the majority of which are opaque (Fig. 35). Diffuse calcification of the parenchyma, known as nephrocalcinosis, is uncommon, but may be seen in tuberculosis, hyperparathyroidism, renal-tubular acidosis, medullary sponge kidney, and in milk-alkali syndrome (Fig. 36). Renal cysts and tumours are occasionally identified by calcification in adjacent involved kidney.



Fig. 35. A large staghorn calculus in the right kidney.



Fig. 36. Nephrocalcinosis in a patient with hyperparathyroidism.

Renal stones can travel down the ureters and appear in the bladder or urethra (Fig. 37). Encrustations on papillary bladder tumours may be visible on the plain film (Fig. 38), as can the florid calcification in the walls of the bladder and ureter in schistosomiasis.



Fig. 37. A patient with calculi in a vesical diverticulum and in the urethra. He also has a calculus in a calcified prostate.

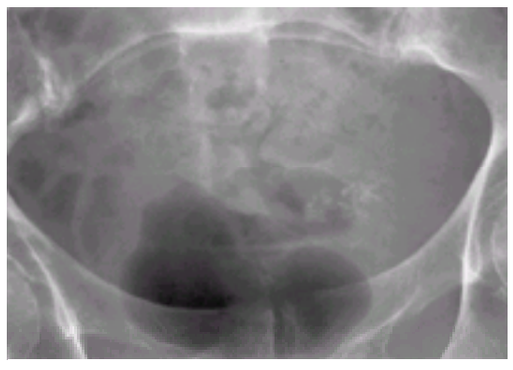


Fig. 38. Plain film of the pelvic outlet showing amorphous calcific deposits on a papillary transitional-cell tumour in the bladder.

Adrenals

The adrenals are nowadays usually investigated by ultrasound, computed tomography, or radionuclide techniques. In areas where tuberculosis is common, this involves the adrenals leading to Addison's disease and calcification in the adrenal areas on the plain film. This could also arise from haemorrhage into the adrenals in infancy and theoretically in certain tumours, notably neuroblastoma.

Further reading

Davidson AJ, ed. *Radiology of the kidney*. WB Saunders Co., Philadelphia, 1985. [A specialist textbook.]

Fraser RG, Pare JAP. *Diagnosis of diseases of the chest*, 2nd edn. WB Saunders Co., Philadelphia, 1977–79. [A specialist textbook.]

Grainger RG, Allison DJ, eds. *Grainger and Allison's diagnostic radiology: a textbook of medical imaging*, Vols 1–3. Churchill Livingstone, New York, 1997. [The current radiologist's ‘bible'.]

Sutton D. *Radiology and imaging for medical students*, 7th edn. Churchill Livingstone, Edinburgh, 1998. [A good reference book for surgeons but less comprehensive than Granger and Allison.]

Sutton D, Young JWR, eds. *A short textbook of clinical imaging*. Springer-Verlag, London, 1990. [A brief introduction to imaging, useful for medical students and registrars in specialties other than radiology.]

13.2 Computed tomography

Stephen Golding

[Introduction](#)
[The technique](#)
- [Preparation](#)
- [Examination technique](#)
- [The image](#)
- [Enhancement](#)
- [Specialized techniques](#)
- [Advantages and disadvantages of CT](#)
- [Relationship to other techniques](#)

[The indications](#)
- [CT-guided interventional techniques](#)

[Costs and benefits of CT](#)
[Further reading](#)

Introduction

In the two decades since computed tomography (**CT**) was introduced by Sir Godfrey Hounsfield it has become established as a powerful diagnostic tool and one that is relevant to many branches of surgery. Used appropriately, CT is capable of making a major impact on management decisions.

The technology of CT is outside the scope of this chapter and the interested reader is referred to Hounsfield or to Pullan. In essence, the scanner rotates an X-ray tube around the patient in an arc and the emergent radiation beam is measured by photoelectric detectors. A computer is used to display the measurements as an image representing a cross-sectional ‘slice' of the patient, based on the density of tissues to X-rays and their ‘attenuation value' (Fig. 1). The image can be thought of as a cross-sectional radiograph, but unlike radiography there is no superimposition of structures and the detector/computer system makes the technique very sensitive. The principal advantages of CT stem from this.

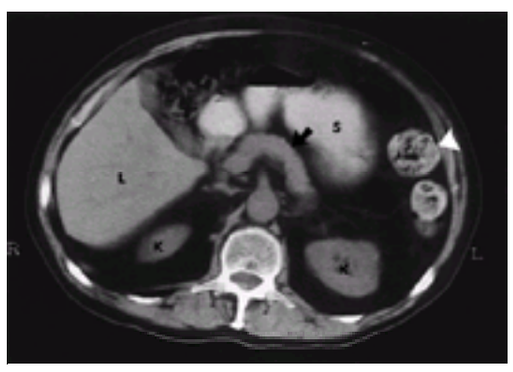


Fig. 1. Normal CT section of the upper abdomen, showing the liver (L), the kidneys (K), and the pancreas (arrow). Note that the organs are identified because the intervening fat has a lower attenuation value than other soft tissues. Oral contrast medium has been given and outlines the stomach (S) and descending colon (arrowhead).

The technique

Preparation

From the patient's point of view examination is simple. In the majority of examinations all the patient has to do is lie on the couch while the machine makes the readings (Fig. 2). Preparation is minimal; some departments prefer to starve their patients for a few hours beforehand if intravenous contrast media are to be given.



Fig. 2. The usual examination position for patients undergoing CT. Greater comfort is provided by using the support behind the knees.

Patients undergoing examination of the abdomen and pelvis are required to drink a dilute contrast medium in order to opacify the bowel and distinguish this from local structures with similar attenuation values (Fig. 1). This contrast medium is also administered rectally for examinations of the pelvic viscera. Examination for gynaecological indications requires a full bladder and the use of a vaginal tampon as an anatomic marker.

Barium retained in the colon after gastrointestinal radiology produces serious artifacts on CT and barium studies should be deferred if CT is required. Large metal prostheses such as hip transplants may cause sufficient interference to prevent successful imaging. Metal clips used for haemostasis or for marking the margins of tumours also produce significant artifacts.

Sedation is only rarely required for patients undergoing CT. Premedication may be useful for those in pain so that they are able to lie still. General anaesthesia is required if involuntary or uncontrollable movement may be a problem, particularly in children. However, many infants can be examined satisfactorily after a feed or under sedation.

Examination technique

Examination is usually carried out with the patient supine, although specialized indications may require specific positions. Exposures typically last a few seconds, and suspended respiration is required when the chest or abdomen are being examined; a diagnostic examination may not be possible in patients who have difficulty holding their breath, as respiratory movement may cause artifact. Other areas can be studied during quiet respiration. In exposures longer than 5 s an antiperistaltic agent (e.g. hyoscine butylbromide, Buscopan, or glucagon) may be given to reduce movement artifact due to peristalsis.

Most CT machines produce a digital radiograph of the examination area (Fig. 3) at the start of the examination. This allows the examination to be planned accurately and also provides a record of the sections.

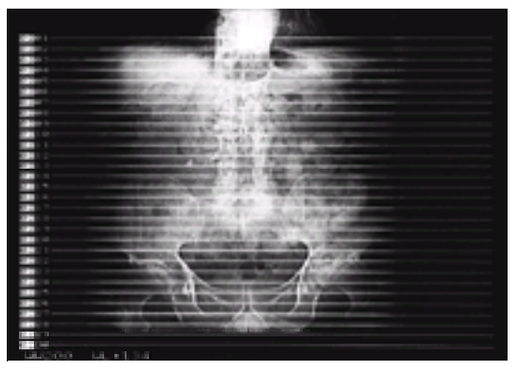


Fig. 3. Digital radiograph obtained at the start of the examination, showing the area to be examined. This examination records 30 sections, taken from the dome of the diaphragm to the pelvic floor.

The examination produces sequential sections, the section thickness and interslice interval being determined by the size of the organ under investigation. Routine examinations of the chest, abdomen, or pelvis usually use sections of 8 to 10 mm in thickness, whereas a specific examination of the adrenal glands may require 5-mm sections, and 1-mm sections may be needed to display the auditory ossicles. Axial sections are standard, although if the patient can be positioned appropriately sections can be obtained in the coronal plane; this is required rarely.

Helical or spiral CT is a recent technique in which the table moves during continuous exposure. Discrete sections are then computed from the data block. Helical CT has the advantage that the examination area can be covered in a short time, often a single breath-hold. It is therefore less susceptible than conventional CT to movement artifact and anatomic misregistration due to physiological move-ment. It also provides superior data for computerized reformatting (see below).

The image

The CT section is a cross-sectional radiograph in which the tissues are displayed on a gray scale according to their attenuation value. Dense structures such as cortical bone appear light, whereas low density areas like air appear dark, as is the case with conventional radiographs. The attenuation of tissues is measured on a wider scale than can be shown effectively on one image and the display console allows the image to be manipulated so that different areas on the attenuation scale can be displayed (Fig. 4).

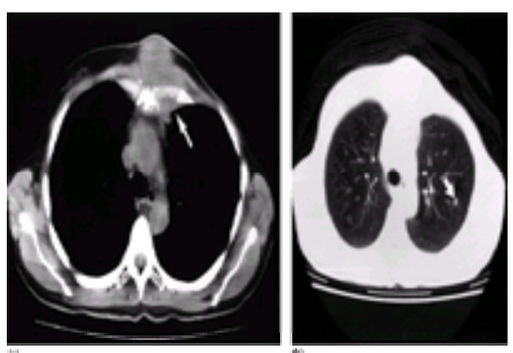


Fig. 4. (a) CT section at the level of the arch of the aorta, displayed at window levels and widths appropriate to soft tissue. The section shows a large soft tissue mass arising in the anterior chest wall, with erosion of the sternum and extension into the mediastinum (arrow). This was due to recurrent carcinoma of the breast. Such images are useful to plan treatment fields for radiotherapy. (b) The same section displayed on lower window levels, showing the structure of the lungs. A small pulmonary metastasis is seen in the left lung (arrow), and was not detected on chest radiographs.

The image data can be manipulated in other ways. Measurements from a small area can be reprocessed to give a high resolution image; this is useful for demonstrating small structures like auditory ossicles or fine detail, as in trabeculae in bone (Fig. 5(a)). Information from contiguous thin sections may be reformatted in different planes or in three-dimensional perspective views (Fig. 5(b)), providing a more anatomical display. Such images are helpful in communicating the orientation of lesions.

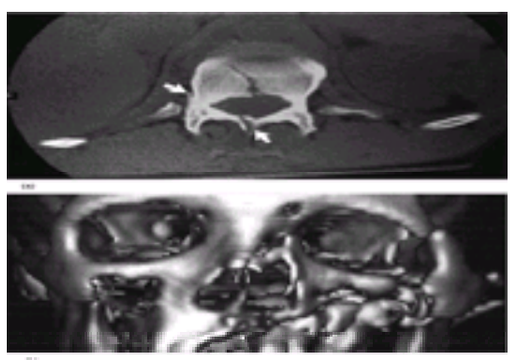


Fig. 5. (a) High resolution CT image obtained in a patient with rotational injury at the thoracolumbar junction. There is a sagittal fracture of the vertebral body. Note that high resolution display also reveals fractures of the neural arch and base of the right pedicle (arrows). (b) Three-dimensional reconstruction produced by surface rendering of CT data in a patient with a comminuted left orbitomaxillary fracture (reproduced by courtesy of Mr S.R. Watt-Smith).

Enhancement

Enhancement refers to the commonly used technique of scanning following the intravenous administration of iodine-containing contrast medium. This increases the attenuation of areas which are perfused, distinguishing them from avascular structures such as abscesses, cysts, and devascularized tissue. It is also used to aid diagnosis by increasing the contrast between normal and abnormal tissues, hence the term ‘enhancement'. This technique is routine in the examination of the liver, kidney, and brain, and in the investigation of abdominal sepsis or trauma (Fig. 6).

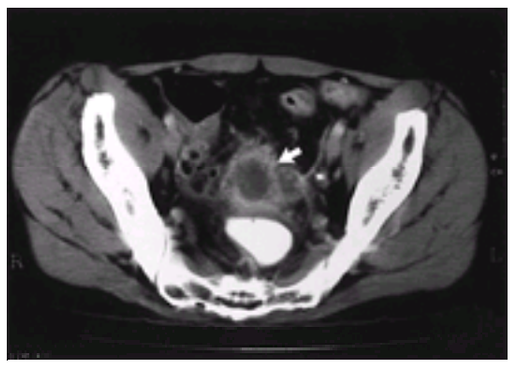


Fig. 6. CT section of the pelvis in a patient with a diverticular abscess. Intravenous contrast medium has been given and outlines granulation tissue around the abscess cavity (arrow). This facilitates identification of inflammatory fluid collections.

Specialized techniques

As circulating blood increases in attenuation after enhancement, examination of blood vessels can be made by taking sections soon after injection, usually during the first or second circulation of the injection bolus. This may be used to distinguish tortuous vessels from other pathology and also to demonstrate vascular disease such as aortic aneurysm or dissection (Fig. 7). Examination during injection through an arterial catheter (CT arteriography) is also recommended in specific indications, although the standard technique suffices for most vascular studies. The ability of helical CT to cover an area quick Jaily makes it particularly valuable for vascular examinations and also for displaying a whole organ (for example the liver) at one stage of enhancement.



Fig. 7. High resolution CT section of the abdomen in a patient with a large aortic aneurysm. The aortic wall is partly calcified. Contrast medium outlines the patent lumen (L) within extensive mural thrombus. Note that the aneurysm has eroded the vertebral body posteriorly and that on the left side there has been partial rupture of the wall.

Many CT machines allow rapidly repeated exposures to be made at one level following an intravascular bolus of contrast medium, with computation of the alteration in density against time. This technique, ‘dynamic CT', produces an accurate measurement of tissue perfusion but is little used in practice, although it is a useful research tool.

CT has also been applied to other contrast procedures in radiology in order to obtain more information, for example after intra-articular injection (CT arthrography). Cross-sectional definition is used to display details which are difficult to assess by the conventional technique.

The precise tissue map of CT images has been used successfully to direct treatment beams in radiotherapy and it is now common for radiotherapy to be planned on the basis of CT images, using special computer hardware.

Advantages and disadvantages of CT

The principal advantage of CT is that it provides a clear, accurate display of tissues without superimposition of structures. Disease processes may be detected at an earlier stage than is possible with other techniques, and lesions may be detected in areas which are difficult to assess with conventional imaging. The technique is not limited to specific organs: since all of the tissues in a body section are displayed it can be used to search for disease sites.

The clinical advantages of CT are well illustrated by its role in neurosurgery, one of its first areas of application. The technique offered for the first time the ability to image cerebral anatomy and pathology directly. More accurate diagnosis was achieved and invasive and indirect tests such as arteriography and encephalography were largely replaced.

Although CT is effective in disease detection and localization, characterization of lesions is more difficult because many have similar attenuation characteristics. For example, it may not be possible to distinguish fibrotic masses from benign or malignant neoplasms on the basis of their CT appearance alone. Biopsy is therefore usually required for definitive diagnosis.

The main disadvantages of CT are the high capital cost of the equipment and the fact that it employs ionizing radiation. The absorbed radiation dose from CT varies according to examination technique, but it is generally similar to that encountered with other major radiological procedures such as angiography or barium enema. CT is therefore used with care around radiosensitive structures such as the eye, or in children and young people, and only for overriding indications in pregnant women.

Relationship to other techniques

In the abdomen, pelvis, and musculoskeletal soft tissues, ultrasound offers an alternative to CT as a sectional imaging technique. The relative strengths of the two are too complicated to discuss in detail here, but in general, if a good quality image is obtained by ultrasound the two techniques are usually comparable in application. However, ultrasound unlike CT is limited by the presence of bowel gas and bone, and if these prevent good images being obtained, CT is more reliable. Ultrasound is also attenuated by fat, making the technique more suitable for slim patients, whereas a moderate amount of body fat improves image quality in CT. Ultrasound is also of limited use in the chest.

The technique which most resembles CT is magnetic resonance imaging (**MRI**). Like CT, MRI is a cross-sectional technique, but unlike CT, examination in any plane is possible. MRI has the advantage of not employing radiation, and it discriminates between soft tissues to a degree unequalled by any other technique. However, MRI is expensive and of limited availability. Movement artifact is a problem in studying the abdominal cavity because scan times are long, and problems in labeling bowel adequately make the techique less applicable than CT in the peritoneal cavity. Little signal is obtained from lung and MRI does not compare with CT in this area.

The indications

It is the golden rule of investigational medicine that no patient is examined unless the results influence clinical management. This is particularly true of CT: examinations should always be tailored to the clinical problem. However, when used appropriately, CT has proved to be a powerful factor in clinical decisions in a wide range of applications. All surgical specialties are major users of CT services. Table 1 lists the clinical indications for which CT is currently recommended; these are divided according to clinical subspecialty although there is some overlap.

Sorry, this figure/table has not been included in electronic versions of the Oxford Textbook of Surgery 2nd Edition for copyright reasons.

Table 1 Applications of computed tomography in surgery

In recent years the indications for CT have had to be reassessed in the light of growing experience in MRI and its availability. MRI has become an established investigational tool in the neurosciences and in orthopaedics, and to a lesser extent in gynaecology (Table 1). For technical reasons MRI is unlikely to be applicable to the diagnosis of conditions affecting the lungs or cortical bone. Haemorrhage can produce confusing appearances on MRI and CT remains the technique of choice in patients who have suffered trauma.

The main advantages of CT stem not only from its accuracy in detecting disease but also from the fact that a convincing normal examination virtually excludes the presence of lesions of any size. When an abdominal mass is suspected, for example, CT is the most accurate technique for demonstrating disease and indicating its organ of origin (Fig. 8), but is also more reliable than ultrasound in excluding disease when none is present. The technique can therefore be used to select patients who require surgical intervention. Similar observations apply to patients with suspected intra-abdominal abscesses (Fig. 6), in whom lesions close to the bowel are difficult to detect by ultrasound; CT more reliably confirms or excludes a focal collection. CT is also a reliable technique for excluding significant damage to intra-abdominal organs in patients who have suffered abdominal trauma (Fig. 9). In all three instances the exclusion of disease has a major effect on the clinical management of the patient.

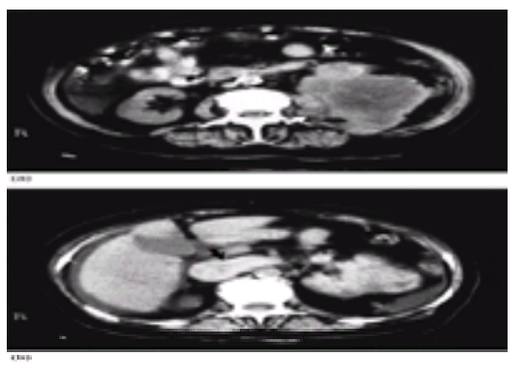


Fig. 8. (a) CT section of the abdomen in a patient who presented with an abdominal mass due to carcinoma of the left kidney. Examination was obtained for tumour staging. The left kidney is replaced by an inhomogeneous mass which has obliterated the fat plane posterior to the kidney and infiltrated the left psoas muscle (arrow), indicating that the tumour is unresectable. (b) A section slightly higher shows tumour expanding the left renal vein and entering the inferior vena cava (arrow). Assessment of the renal vein is an essential component of tumour staging by CT.

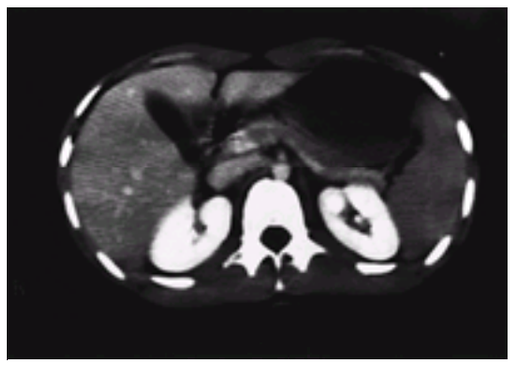


Fig. 9. CT section of the upper abdomen in a patient involved in a road traffic accident. There was extensive laceration to the right side of the trunk. The demonstration by CT of normal anatomy in the body wall and normally enhancing hepatic parenchyma deep to this excludes significant intra-abdominal damage despite extensive right abdominal tenderness and guarding.

The ability of CT to delineate masses accurately has produced a major advance in the management of malignant disease. In addition to the diagnostic role outlined above, the technique is recommended for assessing the local extent of the majority of solid tumours and distinguishes reliably between patients with resectable disease and those in whom attempted resection is pointless (Fig. 8). In addition, CT is used to detect malignant lymph node enlargement in the chest, abdomen, and pelvis (but cannot, unlike lymphography, demonstrate small tumour deposits in nodes of normal size) and has become the technique of choice for demonstrating metastases to the brain, lungs (Fig. 4), liver, and adrenal glands. Staging protocols based on CT have now been defined for the majority of malignant tumors, with the aim of excluding disease spread and therefore identifying those patients suitable for radical treatment.

The combined cross-sectional display of bone and soft tissue has made CT an important technique in orthopaedics, allowing the assessment of stability of serious fractures, demonstrating the disposition of fracture fragments prior to surgical fixation (Fig. 5), and planning corrective surgery in joint disease.

CT-guided interventional techniques

A wide range of percutaneous therapeutic procedures are now performed under CT control. The principal advantage of the technique is that it permits the operator to site an instrument with confidence and safety, even in relatively inaccessible areas of the body. The most common technique is CT-guided biopsy (Fig. 10): aspiration for diagnostic cytology or cutting needle biopsy for histological diagnosis can be undertaken in virtually any area of the body. CT-guided drainage can also be used in the treatment of most deep-seated abscesses and other pathological fluid collections. Guided neurolysis, tumour lysis by alcohol injection, and laser therapy are also possible.

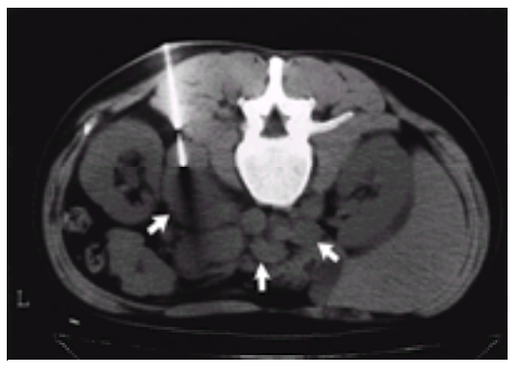


Fig. 10. CT-guided biopsy. The patient presented with abdominal pain and fever. CT showed extensive retroperitoneal lymphadenopathy (arrows). CT guidance was used to site a Trucut needle in the lymph node mass, providing a histological diagnosis of non-Hodgkin's lymphoma.

Costs and benefits of CT

Although the capital and running costs of CT are high, the technique is undoubtedly cost-effective. It can be used to achieve an early diagnosis in patients who would otherwise need to undergo a large number of alternative investigations, and it can be performed on an outpatient basis, reducing costs for inpatient investigation. Moreover, the diagnostic and therapeutic applications of CT frequently replace exploratory laparotomy or other major surgical procedures. Maximization of the cost benefits is heavily dependent on good patient selection, and calls for close liaison between the surgeon and the radiologist.

Further reading

Fishman EK *et al.* Three-dimensional imaging. *Radiology* 1991; **181**: 321–37. [Computing technology for the general reader, with current clinical applications.]

Genant HK. Symposium on computed tomography. *Orthopedic Clinics of North America* 1985; **16**: 357–89. [Useful series of reviews of orthopedic CT.]

Golding SJ. Computed tomography and tumour staging. *Current Imaging* 1990; **2**: 2–8. [An overview of the rationale of imaging in oncologic management].

Golding SJ, Husband JE. CT-guided interventional techniques. *Interventional Radiology* 1990; **5**: 101–23. [Three-part review of the interventional uses of CT.]

Heiken JP, Brink JA, Vannier MW. Spiral (helical) CT. *Radiology* 1993; **189**: 647–56. [Overview of the new technology and clinical uses.]

Hounsfield GN. Computerised transverse axial scanning. *British Journal of Radiology* 1973; **46**: 1016–22. [The ‘creator's' paper, announcing CT.]

Moore AT, Dixon AK, Wheeler T. Cost–benefit evaluation of computed tomography. *Health Trends* 1987; **19**: 8–12. [A seminal study of the factors which make imaging cost-effective.]

Pullan BR. The scientific basis of computerised tomography. In: Lodge T, Steiner R, eds, *Recent advances in radiology*, Vol. 6, pp. 1–15. Churchill Livingstone, Edinburgh, 1979. [A comprehensive account of basic CT technology.]

Williams MP, Scott IHK, Dixon AK. Computer tomography in 101 patients with a palpable abdominal mass. *Clinical Radiology* 1984; **35**: 293–6. [Excellent study relating investigation to clinical need.]

Wing VW, Federle MP, Morris JA, Jeffrey RB, Bluth R. The clinical impact of CT for blunt abdominal trauma. *American Journal of Roentgenology* 1985; **145**: 1191–4. [Another good study relating investigation priorities to clinical need.]

Wittenberg J *et al.* Clinical efficacy of computed body tomography. *American Journal of Roentegenology* 1980; **134**: 1111–20. [One of the first and most important studies of the effect of CT on clinical decision-making.]

13.3 Magnetic resonance imaging

Jeff Fidler and David D. Stark

Introduction
Physical basis of nuclear magnetic resonance
Clinical applications
Brain
Neck
Thorax
Cardiovascular
Abdomen
Biliary tract
Pancreas
Kidney
Adrenal gland
Pelvis
Orthopedics
Further reading

Introduction

Since its clinical introduction in 1981, the use of magnetic resonance imaging (**MRI**) has rapidly expanded. MRI is well suited for the evaluation of many conditions throughout the body. It has inherent advantages over other imaging techniques, such as the lack of ionizing radiation, the ability to image in multiple planes, and contrast resolution.

Recent advances in technology have led to increasing applications, with improved accuracy in diagnosis. Specialized surface and endoluminal coils are now available to examine specific body locations. Stronger and more rapid gradients have led to faster scans, higher spatial resolution, and real-time functional imaging.

In this chapter, we will summarize the clinical applications and findings in various disease processes.

Physical basis of nuclear magnetic resonance

A proton (nucleus of hydrogen) can be compared with a tiny elementary magnet. In a natural environment, the magnetic moments *m* of individual protons in the body are pointed in random directions. It follows that their sum, or macroscopic magnetization, *M*, is zero. If placed in a constant magnetic field B_0, many of the tiny elementary magnets spontaneously undergo a transient orientation in the direction of the field. Therefore, a small magnetization vector **Mz** appears parallel to B_0. The nuclear magnetic resonance phenomenon can be observed when protons absorb radiowaves at a specific frequency (the Larmor frequency). On a macroscopic scale, the magnetization vector **M** will be rotated away from its **Mz** equilibrium position parallel to B_0. The angle between **M** and B_0 will become greater as the applied radiowave has a longer duration or greater power. In MRI, a transmitted radiowave that shifts vector **M** perpendicular to B_0 (transverse or *XY* plane) is called a 90° pulse. As vector **M** is then parallel to the *XY* plane (**M** + **Mxy**), the value of the longitudinal magnetization **Mz** is zero. Resonance occurs when the magnetization rotating in the transverse plane emits radiowaves that can be detected by the imaging system.

When the radiowave excitation pulse ceases, total magnetization *M* slowly returns to its equilibrium state. This phenomenon is called ‘relaxation'. Once the 90° pulse has been terminated, **Mxy** decays and **Mz** grows. This return of longitudinal magnetization (**Mz**) is characterized by the relaxation time T_1. T_1 values vary among tissues according to the mobility of the hydrogen (proton)-containing molecules. Water molecules (such as cerebrospinal fluid) have a very long T_1, while lipid molecules (such as fat tissues) have a very short T_1.

The 90° pulse that shifts equilibrium magnetization from **M** to **Mxy** also places the nuclear magnetic moment ‘in phase', which means that they are moving in exact alignment. However, transverse magnetization is progressively dephased (protons lose alignment) soon after the pulse ceases. This phenomenon, known as transverse relaxation, is characterized by the relaxation time T_2.

Nuclear magnetic resonance in biologic tissues is measured by detecting transverse magnetization **Mxy** before it relaxes (dephases). An electric current is induced in an antenna by the variation of the magnetic field caused by in-phase rotation of protons in the transverse plane (**Mxy**). This electrical signal (the MR signal) represents the magnetic characteristics of each tissue. The amplitude and timing of various signals is mathematically converted to represent the number of protons present in a tissue (proton density) or the relaxation time (T_1 and T_2).

Clinical applications

Brain

General characteristics

Numerous studies in this field have been documented since MRI was first introduced to the clinical environment. MRI continues to be a unique tool, providing high-quality images of the brain with few motion artefacts.

Specific lesions

Cerebral infarction

Cerebral infarction refers to a sudden and focal neurologic deficit with an associated ischemic abnormality in a localized region of the brain due to thrombosis. The hypoxia that occurs with ischemia initially leads to an increase in the intracellular water content; tissue necrosis results if the ischemia is persistent. Prompt diagnosis is imperative in order to identify those patients who may benefit from thrombolytic therapy. Recently, two new pulse sequences have been developed that may allow early identification of stroke and help in the differentiation of salvageable from infarcted tissue. These pulse sequences are perfusion MRI, which produces a measure of blood flow and blood volume within a given area in the brain, and diffusion MRI, which produces a measure of the microscopic motion of water and tissue. Areas of ischemia show altered binding of water to proteins and other cellular structures, which results in a decrease in diffusion. In ischemic patients, comparison of the relative size of perfusion and diffusion abnormalities enables discrimination between normal, ischemic, and infarcted tissue.

Cerebral hemorrhage

Intracranial hemorrhage can appear as various pathophysiologic states, including intraparenchymal hemorrhage, intratumoral hemorrhage, hemorrhagic infarction, subarachnoid hemorrhage, epidural or subdural hemorrhage, and vascular malformation.

Brain neoplasms

Precisely defining the extent of the tumor and grading the malignancy are the most important goals of any imaging study because of the implications for patient management and evaluation of treatment. So many classifications of brain tumors have been provided that a knowledge of general neuropathologic features is essentially for analyzing MR images.

Multiple sclerosis

The clinical course of multiple sclerosis is usually described as a relentless stepwise progression of neurologic dysfunction, manifested by exacerbations and remissions. The plaques are pathologically considered to be gliosis, with edematous change in the acute phase.

MRI findings

Cerebral infarction

Ischemic infarct and/or edema is often visualized within 6 to 12 h of onset as an area of hyperintensity (white signal area) on T_2-weighted images. A well-circumscribed hyperintensity on T_2-weighted images and hypointensity on T_1-weighted images suggest chronic ischemic infarction.

Cerebral hemorrhage

Pathophysiology and consecutive changes in the erythrocytes and hemoglobin in hematomas make MRI findings complex. Within the first few hours after an intraparenchymal hemorrhage, the signal intensity of the hematoma is similar to that of normal tissue. MRI cannot depict acute subarachnoid hemorrhage because the abundant oxygen contained in cerebrospinal fluid prevents the deoxygenation of hemoglobin. In the subacute stage (up to a week), hematomasbegin to show central hypointensity due to the presence of deoxy-hemoglobin and hyperintensity of the periphery corresponding to edema on T_2-weighted images. Over a period of 1 week, hematomas appear as areas of hyperintensity, due to methemoglobin, ringed with hypointensity due to hemosiderin.

Brain neoplasms

Because the tumor may contain cysts, calcification, hemorrhage, and necrosis, a single signal pattern is not found. Peritumoral edema increases relaxation times and makes if difficult to detect the tumor margin. Tumors often extend beyond the rims apparent on computed tomographic (**CT**) scans, and similar difficulties may reasonably be expected when MRI is employed.

Multiple sclerosis

The characteristic MRI findings consist of multiple, usually small, lesions with prolonged relaxation times, most commonly located in the periventricular region. The T_2-weighted spin-echo technique is currently the best method for detecting multiple sclerosis; this is positive in between 80 and 100 per cent of patients with definite multiple sclerosis by established clinical criteria.

Advantages of MRI over other methods

Cerebral infarction

The advantages of MRI over CT are apparent when delineating small infarcts in the brainstem and cerebellum, and lacunar infarcts in the basal ganglia. MRI can detect hyperacute ischemic changes in the brain by 6 h after onset, while CT is usually useful in detecting them after 24 h. Utilizing perfusion and diffusion imaging, the diagnosis can be made even earlier and provide information on tissue salvageability.

Cerebral hemorrhage

MRI is less useful for detecting acute (within 1 day) cerebral hemorrhage. However, it provides a dynamic ‘window' through which a variety of hemorrhagic conditions may be observed.

Brain neoplasms

Multiplanar capabilities and a reduction in artefacts from the skull can demonstrate the relation between brain tumors and the skull better than CT, particularly in patients with infratentorial tumors. However, MRI itself occasionally fails to delineate tumor margins, due to peritumoral edema. A solution would be provided by the use of intravenous contrast agent, gadolinium–DTPA.

Multiple sclerosis

The role of traditional imaging studies in multiple sclerosis has been indirect, to rule out a space-occupying lesion. MRI can be used to determine regression or progression of lesions. Tiny plaques in the brainstem, cerebellum and spinal cord, usually missed with CT, are often detected on T_2-weighted images.

Neck

Parathyroid gland

General characteristics

Normal parathyroid glands cannot be routinely imaged by CT, ultrasonography, or MRI because of their small size and poor contrast against the thyroid. However, they can be visualized by these imaging modalities once neoplastic or hyperplastic enlargement occurs.

MRI findings

Parathyroid adenomas and hyperplastic glands usually show high signal intensity on T_2-weighted images (Fig. 1). Approximately 75 per cent of adenomas can be detected by MRI, but these may be indistinguishable from posterior thyroid adenomas.

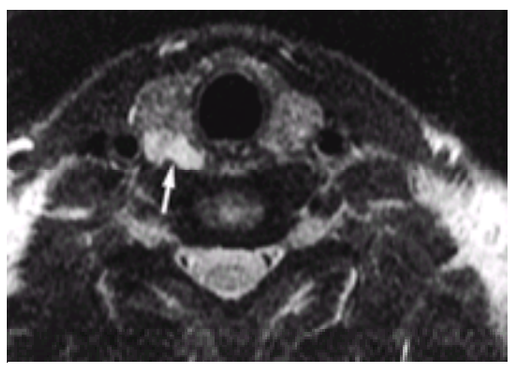


Fig. 1. Parathyroid adenoma: axial image through the region of the thyroid gland shows a high signal-intensity parathyroid adenoma (arrow) located just posterior to the right lobe of the thyroid.

Advantages of MRI over other methods

MRI is complementary to nuclear medical studies in the identification of adenomas. It is particularly valuable in patients with recurrent or persistent hyperparathyroidism following surgical exploration.

Thorax

General characteristics

Enlarged mediastinal lymph nodes are the most common mass lesions, accounting for 25 per cent of all mediastinal diseases. Thymomas, teratomas, neurogenic tumors, and bronchogenic and pericardial cysts are also common. In the evaluation and staging of malignancies, MRI is mainly used as a problem-solving technique, such as evaluating for chest-wall, mediastinal, or diaphragmatic extension of tumor.

MRI findings

Mediastinal lymph nodes are easily visible on T_1-weighted images, standing out clearly against hyperintense mediastinal fat. Mediastinal tumors such as thymoma, germ-cell tumors, and neurogenic tumors usually have signal intensities similar to those of lymph nodes. Cystic lesions containing serous fluid can be distinguished from complicated cysts and solid lesions by their long relaxation times. With tumor extension, the normal high signal-intensity fat planes will be obliterated. On T_1-weighted images, tumor will have intermediate signal intensity and high signal on T_2 weighting (Fig. 2).

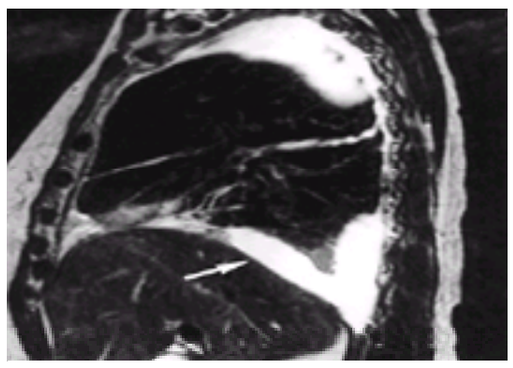


Fig. 2. Mesothelioma: sagittal T_2-weighted image shows a rim of high signal-intensity pleural thickening (arrow) which is mesothelioma; there is no evidence of local extension through the diaphragm into the liver.

Advantages of MRI over other methods

MRI may be more efficient than CT for the detection of hilar adenopathy; the condition can be distinguished from pulmonary vessels and bronchi without the need for intravenous contrast because of the superb contrast afforded by the lack of a signal from flowing blood. However, accurate discrimination between cancer and inflammation is still impossible by MRI: size criteria analogous to those used in CT (larger than 1 cm) are used to identify abnormal nodes. MRI cannot provide tissue-specific diagnosis of mediastinal tumors on the basis of signal intensity, but can indicate spatial relations between tumors and adjacent structures by its multiplanar capabilities. Because of the superior tissue contrast, loss of tissue planes and abnormal signal in tissues may be more apparent than with other techniques.

Cardiovascular

General characteristics

Because of its ability to demonstrate blood flow and its multiplanar capability, MRI is playing an increasing part in the evaluation of cardiovascular disease. Indications include the evaluation of cardiac masses and pericardial disease, the anatomic evaluation of congenital or postoperative abnormalities, functional assessment, and the examination of aortic abnormalities. MRI is also useful in the evaluation of carotid and peripheral vascular disease.

MRI findings

On T_1-weighted spin-echo images, vessels with flowing blood will appear black or demonstrate ‘flow void'. However, in areas of slow flow or clot, the vessel will demonstrate increased signal. Cine images acquired in relation to the QRS waveform demonstrate high-signal blood, which can be examined over a temporal period; this is useful in functional evaluation and for differentiating slow or turbulent flow from clot. Three-dimensional angiographic images can be produced using time-of-flight or phase-contrast techniques.

Advantages of MRI over other methods

MRI is cheaper, less invasive, and has less associated risk than conventional angiography, and can replace angiography in several situations. Images can be obtained without the injection of contrast and can also allow identification of valvular disease. MRI allows improved visualization of certain locations that are difficult to see with echocardiography, such as the pericardium and right ventricle.

Abdomen

Liver

General characteristics

In general, MRI is considered to have greater sensitivity than CT in lesion detection. However, because of the cost and limitations in screening the entire abdomen, MRI is primarily used as a problem-solving technique. Some of the specific indications include lesion characterization and identification, confirmation of fatty infiltration, and in the diagnosis and follow up of hemochromatosis. New hepatoselective contrast agents are becoming available, which may further improve the sensitivity of MRI for the detection of lesions.

MRI findings

Hemangioma On T_2-weighted images, hemangiomas are very well-defined, high signal-intensity lesions. Following dynamic injection of extracellular gadolinium contrast agents, enhancement is similar to that on CT, with peripheral enhancement and centripetal filling. The specificity of MRI in detecting hemangiomas is greater than 90 per cent.

Metastases The MRI appearance of metastases is nonspecific, but is usually low signal on T_1 weighting and higher signal on T_2 weighting, though not as bright as cysts or hemangiomas (Fig. 3). Certain features such as a target-like appearance or rings may suggest the diagnosis. Hemorrhagic lesions may have high signal intensity on T_1 weighting.

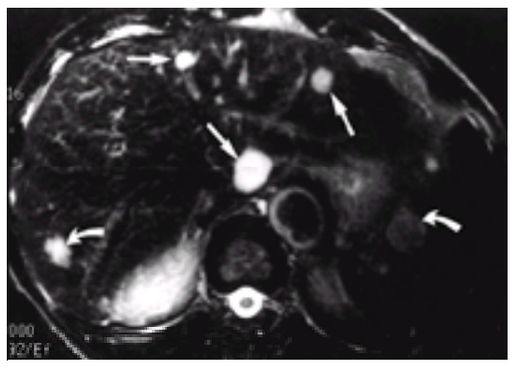


Fig. 3. Hemangiomas and metastases: axial T_2-weighted image throughout the liver and spleen shows well-defined, high signal-intensity hemangiomas (arrow) and lower signal-intensity, ill-defined heterogeneous lesions (curved arrow) representing metastases.

Miscellaneous lesions Adenomas may have areas of hemorrhage or fat that may be high signal on T_1-weighted images and are usually high intensity on T_2 weighting. Focal nodular hyperplasia may show intense enhancement following the injection of dynamic contrast. The presence of a central stellate scan may suggest the diagnosis.

Hepatocellular carcinoma Hepatomas demonstrate signal intensities similar to those of metastases. A small percentage may have high signal on T_1-weighted images because of steatosis. MRI may be useful in characterizing the nodular disorders associated with cirrhosis by comparing the combination of signal intensities on T_1- and T_2-weighted images.

Hepatic steatosis Focal fatty infiltration may have high signal intensity on T_1 weighting, but will often be isointense to normal liver. The most sensitive technique for detecting fatty infiltration is chemical-shift imaging. Fat and water rotate at different rates in a magnetic field. By imaging when fat and water are out of phase, areas of fat will have relative lower signal than images obtained while they are in phase.

Advantages of MRI over other methods

The sensitivity and specificity of MRI depend on many factors, including resolution, motion-correction techniques, and the pulse sequences utilized. Advantages of MRI include the ability to vary pulse sequences to characterize certain types of tissue and to image the liver without iodinated contrast agent.

Hemangiomas MRI is superior to other modalities in characterizing hemangiomas. CT requires iodinated contrast and serial imaging while scans of nuclide-tagged red blood cells have inferior resolution and are unable to identify other lesions.

Metastases High-quality MRI examinations have higher sensitivity in lesion detection than dynamic incremental CT techniques, but newer CT techniques, such as spiral CT, have demonstrated improved sensitivities. Advantages of MRI over CT include improved sensitivity when underlying parenchymal or hemodynamic changes lead to altered enhancement of the liver. These conditions include fatty infiltration, cirrhosis, and patients not able to receive iodinated contrast. MRI is useful in preoperative staging as additional lesions can be detected, often altering therapy.

Biliary tract

General characteristics

New fast T_2-weighted pulse sequences now allow exquisite images to be obtained of the biliary duct (MR cholangiopancreatography). Images can be obtained in any plane and three-dimensional images produced, similar to direct cholangiography.

MRI findings

The bile ducts appear as high signal intensity on these heavily T_2-weighted images. Stones or debris will have lower signal intensity than the surrounding bile.

Advantages of MRI over other methods

MRI allows cholangiographic images to be produced noninvasively without the injection of contrast, and is cheaper and has fewer associated risks than endoscopic retrograde cholangiopancreatography or percutaneous transhepatic cholangiography. However, its exact role is yet to be determined.

Pancreas

General characteristics

MR imaging of the pancreas is indicated to evaluate for pancreatic masses when a CT is equivocal or negative, or when a contrast CT cannot be performed. MR angiography can evaluate the patency of surrounding vessels and MR pancreatography can evaluate the pancreatic duct. MRI also may have improved sensitivity in detecting islet-cell tumors.

MRI findings

One of the most sensitive sequences in identifying small pancreatic tumors are fat-suppressed, T_1-weighted images. Tumors will appear relatively lower signal than the surrounding normal parenchyma. Nonfat-suppressed images can demonstrate extension into surrounding fat and T_2-weighted images the presence of metastases. Islet-cell tumors usually are high signal on T_2-weighted images and may demonstrate hyperenhancement.

Advantages of MRI over other methods

Currently, MRI offers no significant advantage over spiral CT in the local staging of pancreatic malignancies. It can allow improved identification of islet-cell tumors.

MR pancreatography has similar advantages to MR cholangiography. It can diagnose pancreas divisum and dilatation of the main duct in chronic pancreatitis. It can also allow visualization of the duct proximal to strictures when it cannot be visualized by endoscopic retrograde cholangiopancreatography.

Kidney

General characteristics

MRI is indicated for the evaluation of renal masses when other modalities, such as CT or ultrasound, are not diagnostic. A major application of MRI is in the evaluation of vascular invasion. MR urography can allow functional information to be obtained, and can evaluate the ureter and collecting system.

MRI findings

Masses

Renal-cell carcinoma This is usually a solid, enhancing mass. Serial images obtained during dynamic injection of contrast can demonstrate enhancement of these tumors. On gradient echo, extension of tumor into the vessel will appear as lower-signal foci in the high-signal blood (Fig. 4).

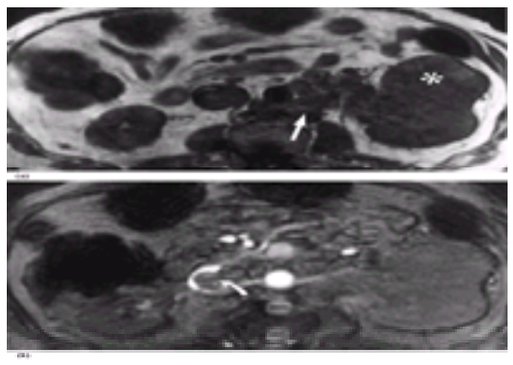


Fig. 4. Renal-cell carcinoma with vascular extension: axial T_1-weighted image (A) and axial gradient-echo images (B) demonstrate a large, exophytic, renal-cell carcinoma (*), retroperitoneal lymphadenopathy (arrow), and vascular extension into the inferior vena cava (curved arrow).

Angiomyolipoma MRI can demonstrate fatty content within a mass (high signal on T_1, suppresses with fat suppression), diagnostic of angiomyolipoma.

Advantages of MRI over other methods

MRI has no significant advantage over contrast-enhanced CT in lesion detection. However, it offers improved evaluation of vascular extension. In patients that cannot receive iodinated, contrasted CT, MRI offers an alternative technique to identify and characterize masses. IT also allows examination in various planes, which can help identify the organ of origin of a suspected renal mass (adrenal, liver, etc.) and extension into adjacent structures. Currently, MR urography has no advantages over other techniques, except in those cases where contrast cannot be administered or when questions cannot be answered by CT or ultrasound.

Adrenal gland

General characteristics

The main indication for MRI of the adrenal gland is to differentiate adrenal adenomas from other masses. It is also useful in helping to identify the organ of origin of the mass and for evaluating any vascular extension of tumor.

MRI findings

Because of their lipid content, adenomas will demonstrate relative signal loss on chemical-shift, opposed-phase images (technique described above in ‘Hepatic steatosis'). Because there is not a large amount of lipid, these tumors are not high signal on T_1-weighted images, unlike a myelolipoma, which usually contains abundantfat. The MR appearance of other tumors is nonspecific. Pheochromo-cytomas are usually large lesions that are very hyperintense on T_2-weighted images and demonstrate enhancement.

Advantages of MRI other over methods

Non-contrasted CT has high specificity in characterizing adenomas. In equivocal lesions, chemical-shift techniques may add further information.

Pelvis

Uterus

General characteristics

MRI provides exquisite anatomic images of uterine zonal anatomy. Indications include the staging of endometrial or cervical carcinoma, and the diagnosis of leiomyomas, adenomyosis, and congenital fusion anomalies (Fig. 5).

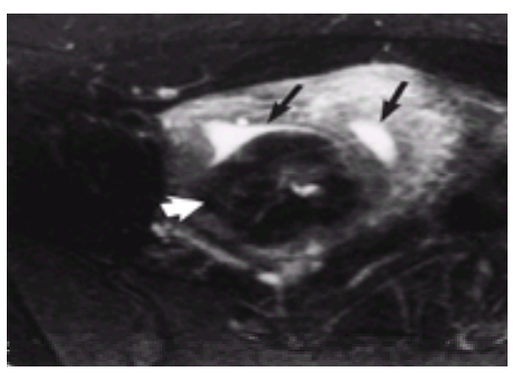


Fig. 5. Leiomyoma and uterine fusion anomaly: axial T_2-weighted image through the uterus shows a characteristic low signal-intensity leiomyoma (short arrow) and two uterine horns (long arrow) in this patient with a septate uterus and associated leiomyoma displacing the right uterine horn.

MRI findings

T_2-weighted images provide superior contrast of the zonal anatomy of the uterus and cervix. Delineation of these structures allows the depth and extent of invasion to be determined. Tumors appear as high-signal areas. On T_2-weighted images, leiomyomas have a characteristic low signal intensity. The MR appearance of adenomyosis consists of a diffuse or focal thickening of the junctional zone, an ill-defined junctional zone, or scattered, tiny focal areas of high signal in the myometrium. Visualization of the uterine zonal anatomy allows congenital fusion anomalies to be characterized by identifying the uterine horns, intervening tissue, and fundal contour.

Advantages of MRI over other methods

The advantages of MRI over CT include its multiplanar capability and superior contrast resolution, allowing better local staging of malignancies.

Prostate

General characteristics

Endorectal coils now allow exquisite high-resolution imaging of the prostate. Indications for MRI include the local staging of prostate cancer.

MRI findings

T_2-weighted images demonstrate the zonal anatomy of the prostate. On axial T_2-weighted images, the peripheral zone is demonstrated as a U-shaped, peripheral, high signal-intensity structure with the central gland less intense. The majority of cancers arise in the peripheral zone and appear as lower-signal foci, but this is nonspecific (Fig. 6).

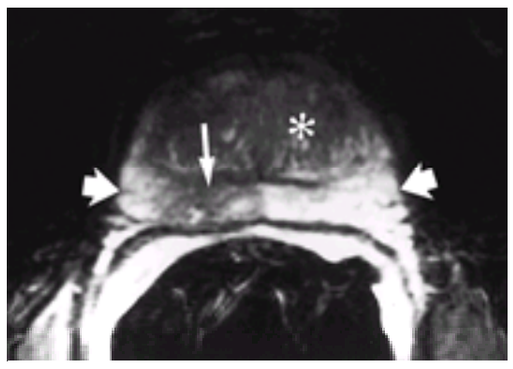


Fig. 6. Prostate carcinoma: axial T_2-weighted image through the prostate shows the high signal-intensity peripheral zone (short arrows) and intermediate signal-intensity central gland (*). Note the low signal-intensity carcinoma (long arrow) involving the right side of the peripheral zone.

Advantages of MRI over other methods

Despite its excellent tissue contrast compared to other techniques, the less than optimal accuracy rates for detecting local capsular penetration have not led to the acceptance of MRI in routine staging algorithms.

Orthopedics

Spine

General characteristics

Disease of the spine can be divided into three categories: intramedullary (intraspinal cord) lesions, intradural extramedullary lesions, and extradural lesions, including diseases of the intervertebral disc.

MRI findings

Intramedullary lesions Syringomyelia and hydromyelia appear as low-intensity images with T_1 weighting and as high signal-intensity lesions spreading longitudinally within the spinal cord with T_2 weighting. In patients with Arnold–Chiari malformation, MRI that includes the posterior fossa of the skull can show the cerebellar tonsils and vermis projecting into the upper cervical spinal canal. Intramedullary plaques of multiple sclerosis may appear as high intensity on T_2-weighted images, but are not visible on T_1-weighted images.

Intramedullary spinal tumors (ependymomas and astrocytomas) have no reliable characteristic MR feature, but appear as lesions with prolonged T_1 and T_2. Spinal arteriovenous malformation may be sometimes demonstrated as a high signal-intensity lesion with a spotty signal void, consistent with abnormal large vessels detectable on T_2-weighted images.

Intradural extramedullary lesions Paravertebral extensions of dumb-bell shaped, intradural, extramedullary tumors (neurofibroma, neurinoma, etc.) are usually well demonstrated on T_2-weighted images but poorly differentiated from contiguous muscle on T_1-weighting. Spinal meningiomas show similar signal intensities to neural tissues on both T_1- and T_2-weighted imaging.

Extradural lesions Degenerative discs usually produce lower signal intensities than normal on T_2-weighted images. Protruded or extruded fragmented discs are easily visible on T_1-weighted and proton density (PD)-weighted images. Involvement of the spine by metastatic tumors is demonstrated as low intensity on T_1-weighted images (Fig. 7).



Fig. 7. Herniated disc: sagittal T_2-weighted image shows the posterior herniated disc (arrow).

Advantages of MRI over other methods

For the investigation of almost all conditions of the spine, MRI is an alternative to contrast myelography and is complementary to radiographs and CT scans. Sagittal scans of the spine and oblique scans along disc spaces are particularly helpful.

Intramedullary lesions Syringomyelia and hydromyelia are demonstrated in their superior and inferior extensions in most cases (about 90 per cent). CT myelography is the most effective way of confirming syringomyelia when the results of MRI are equivocal. MRI is the only method capable of demonstrating the intramedullary plaques of multiple sclerosis. A survey examination of the brain in patients with suspected spinal lesions of multiple sclerosis may demonstrate lesions compatible with

this diagnosis.

Intradural extramedullary lesions Intradural extramedullary tumors can be demonstrated on MRI as well as on CT, but secondary erosive changes of the subarticular canal are less precisely visualized than with CT. Calcification is frequently difficult or impossible to detect on MRI. The intravenous contrast agent gadolinium-DTPA may be valuable, particularly in detecting and estimating the extent of tumors.

Extradural lesions MRI is widely applicable in the investigation of degenerative disc diseases (spondylosis, disc herniation, and canal stenosis). There is an 83 per cent agreement between the results of preoperative MRI and surgical findings in patients with degenerative disc diseases, a rate equivalent to that of CT myelography. T_2-weighted imaging in the sagittal plane is useful in the survey of degenerative discs, which show lower signal intensities than normal. Sagittal T_1-weighted imaging allows simultaneous visualization of multiple metastatic tumors affecting the vertebrae; this is impossible with CT.

MRI findings complement those from plain radiographs in the assessment of vertebral disorders, such as atlantoaxial subluxation, spondylosis, spondylolisthesis, and tethered cord syndrome (myelo-meningocele with dysgenesis of the spine and lipoma).

Knee joint

General characteristics

Injuries of the knee joint, such as meniscal tears, ligament injuries and fractures, arthritis, and neoplasms, are routinely encountered. MRI is the study of choice for internal derangement.

MRI findings

Excellent visualization of menisci and ligaments can be achieved by coronal and sagittal MRI. For example, meniscal tears show disruption of the low signal-intensity meniscus with high-intensity intra-articular fat or fluid on T_2-weighted images (Fig. 8).

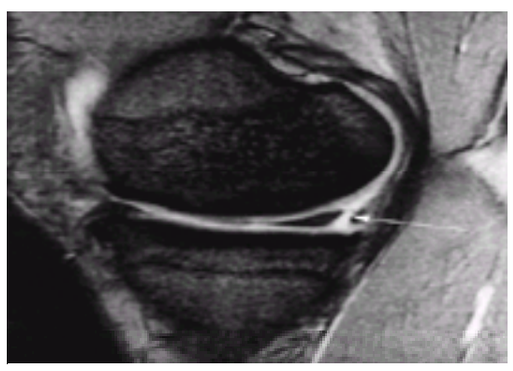


Fig. 8. Meniscal tear: sagittal T_2-weighted image through the medial meniscus shows a large high signal-intensity vertical tear (long arrow) extending through the posterior horn.

Advantages of MRI over other methods

MRI may be helpful in evaluating patients with minor knee injuries and may reduce the need for arthrography and arthroscopy. Meniscal tears that are difficult to detect on arthroscopy can be demonstrated, along with surrounding fluid or blood. Images along the longitudinal direction of the femur, tibia, and fibula are useful to show the extent of marrow involvement in patients with bone tumors.

Glenohumeral joint

General characteristics

The most common clinical indications for MRI are pain or restricted total range of motion in the shoulder, patients referred with suspected rotator-cuff tears or impingement, defects in the glenolabrum, infections, and neoplasms.

MRI findings

Sagittal and oblique coronal, T_2-weighted imaging is well suited to the evaluation of rotator-cuff disease. Tendinitis and fluid collections show as high-intensity lesions, often associated with the cuff tear. Disruptions of the labrum, the main cause of recurrent anterior dislocation, are seen as linear, high-intensity lesions on T_2-weighted imaging (Fig. 9).

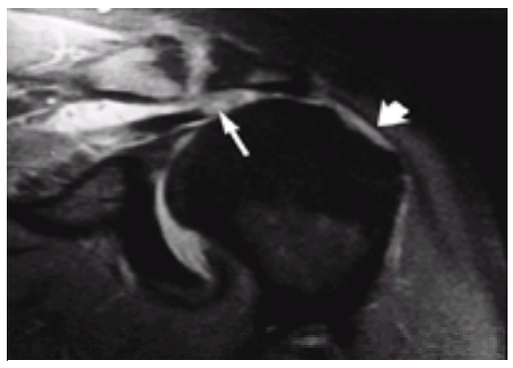


Fig. 9. Rotator cuff tear: oblique coronal views through the left glenohumeral joint show a complete tear with retraction; note the bulbous end of the retracted supraspinatus muscle (long arrow) and high-signal fluid at the expected insertion on the humeral head (short arrow).

Advantages of MRI over other methods

Routine radiography, radionuclide studies, CT, arthrography, and sonography historically have been important techniques for evaluating the glenohumeral joint, but MRI is now the preferred technique for internal ligamentous and muscular injuries.

Hip joint

General characteristics

Among the numerous causes of hip pain, MRI, because of its high sensitivity and specificity, has progressed most rapidly in evaluating patients with suspected

avascular necrosis of the femoral head.

MRI findings

Early in the course of avascular necrosis, an inhomogeneous loss of signal intensity can be seen on T_1-weighted images. A low-intensity line of demarcation may also be evident at the margin of the necrotic zone during the early phase, when radiographs are typically normal.

Advantages of MRI over other methods

MRI permits early detection of avascular necrosis of the hip.

Further reading

Bahn MM, Oser AB, Cross DT III. CT and MRI of stroke. *Journal of Magnetic Resonance Imaging* 1996; **6**: 833.

Barish MA *et al.* MR cholangiopancreatography: efficacy of three-dimensional turbo spin-echo technique. *American Journal of Roentgenology* 1995; **165**: 295–300.

Berquist TH. Shoulder and arm. In: Berquist TH, ed. *MRI of the musculoskeletal system*, 2nd edn, pp. 313–56. Raven, New York, 1990.

Bilbey JH *et al.* NM imaging of the adrenal masses: value of chemical-shift imaging for distinguishing adenomas from other tumors. *American Journal of Roentgenology* 1995; **164**: 637–42.

Bradley WG Jr, Crooks LE, Newton TH. Physical principles of NMR. In: Newton TH, Potts DG, ed. *Modern neuroradiology*, Vol. 2, pp. 15–61. Clavadel, San Francisco, 1983.

Brown JJ, Naylor MJ, Yagan N. Imaging of hepatic cirrhosis. *Radiology* 1997; **202**: 1.

Choi BI, Takayasu K, Han MC. Small hepatocellular carcinomas and associated nodular lesions of the liver: pathology, pathogenesis and imaging findings. *American Journal of Roentgenology* 1993; **160**: 1177–87.

Ebisu T *et al.* Hemorrhagic and non-hemorrhagic stroke: diagnosis with diffusion-weighted and T_2-weighted echo-planar NM imaging. *Radiology* 1997; **203**: 823.

Erickson SJ. High-resolution imaging of the musculoskeletal system. *Radiology* 1997; **205**: 593.

Feinberg DA. Functional magnetic resonance imaging: application of degenerative brain disease and hydrocephalus. *Neuroimaging Clinics of North America* 1995; **5**: 125.

Finn JP *et al.* Imaging of the portal venous system in patients with cirrhosis: NM angiography vs duplex Doppler sonography. *American Journal of Roentgenology* 1993; **161**: 898–994.

Fretz CJ *et al.* Detection of hepatic metastases: comparison of contrast enhanced CT, unenhanced MR imaging, and iron oxide enhanced MR imaging. *American Journal of Roentgenology* 1990; **155**: 763–70.

Gandon Y *et al.* Hemochromatosis: diagnosis and quantification of liver iron with gradient-echo MR imaging. *Radiology* 1994; **193**: 533–8.

Hagspiel DK *et al.* Detection of liver metastases: comparison of supramagnetic iron oxide-enhanced and unenhanced NM imaging at 12.5 T with dynamic CT, intraoperative US, and percutaneous US. *Radiology* 1995; **196**: 471–8.

Higgins CB, Sakuma H. Heart disease: functional evaluation with NM imaging. *Radiology* 1996; **199**: 307.

Irizarry JM, Recht NM. MR Imaging of the knee ligaments and the postoperative knee. *Radiology Clinics of North America* 1997; **35**: 45.

Kreft BP *et al.* Diagnosis of fatty liver with MR imaging. *Journal of Magnetic Resonance Imaging* 1992; **2**: 263–471.

Langlotz C, Schnall NL, Pollack H. Staging of prostatic cancer: accuracy of MR imaging. *Radiology* 1995; **194**: 645.

Loevner LA. Imaging of the parathyroid glands. *Seminars in US CT MR* 1996; **17**: 563.

Low RN *et al.* Abdominal MR imaging: comparison of T_2-weighted fast and conventional spin echo and contrast-enhanced multiplanar spoiled gradient-recalled imaging. *Radiology* 1993; **186**: 803–11.

McFarland EG *et al.* Hepatic hemangiomas and malignant tumors: improved differentiation with heavily T2-weighted conventional spin-echo MR imaging. *Radiology* 1994; **193**: 43–7.

McFarland EG *et al.* Pre-operative staging of cancer of the pancreas: value of MR angiography versus conventional angiography in detecting portal venous invasion. *American Journal of Roentgenology* 1996; **166**: 37–43.

Maldjian C, Schnall NM. Magnetic resonance imaging of the body. *Seminars in Roentgenology* 1996; **31**: 257.

Marks MP, de Crespigny A, Lentz D, Enzmann DR, Albers GW, Moseley ME. Acute and chronic stroke: navigated spin-echo diffusion-weighted MR imaging. *Radiology* 1996; **199**: 403.

Miyazaki T *et al.* MR cholangiopancreatography using HASTE (half-Fourier acquisition single-shot turbo spin-echo) sequences. *American Journal of Roentgenology* 1996; **166**: 1297–303.

Modic MT *et al.* Imaging of degenerative disc disease. *Radiology* 1988; **168**: 177–86.

Murakami T *et al.* Regenerating nodules in hepatic cirrhosis: MR findings with pathologic correlation. *American Journal of Roentgenology* 1990; **155**: 1227–31.

Nelson RC, Chezmar JL, Sugarbaker PH, Bernardino NE. Hepatic tumors: comparison of CT during arterial portography, delayed CT, and MRI imaging for preoperative evaluation. *Radiology* 1989; **172**: 27–34.

Oi H, Murakami T, Kim T, Matsushita M, Kishimoto H, Nakamura H. Dynamic imaging and early phase helical CT for detecting small intrahepatic metastases of hepatocellular carcinoma. *American Journal of Roentgenology* 1997; **166**: 369–74.

Outwater EK, ed. *Topics*, Vol. 8(5), *Magnetic resonance imaging*. Lippincott-Raven, New York, 1996.

Prince MR *et al.* Breath-hold gadolinium-enhanced MR angiography of the abdominal aorta and its major branches. *Radiology* 1995; **197**: 785.

Ros PR, ed. MR imaging of the liver. *Magnetic Resonance Imaging Clinics of North America*, May 1997.

Rothpearl A *et al.* MR urography: technique and application. *Radiology* 1995; **194**: 125–30.

Rubin DA. MR imaging of the knee menisci. *Radiology* 1997; **35**:21.

Rummeny EJ *et al.* Comparison between high-field strength MR imaging and CT for screening of hepatic metastases: a receiver operating characteristic analysis. *Radiology* 1992; **192**: 879–86.

Schiebler ML *et al.* Prostatic carcinoma and benign prostatic hyperplasia: correlation of high-resolution MR and histopathologic findings. *Radiology* 1989; **172**: 131–7.

Schmiedl UP *et al.* MR angiography of the peripheral vasculature. *Seminars in US CT MR* 1996; **17**: 404.

Semelka RC *et al.* Pancreatic disease: prospective comparison of CT, ERCP, and 1.5 T MR imaging with dynamic gadolinium enhancement and fat suppression. *Radiology* 1991; **181**: 785–91.

Semelka RC *et al.* Renal lesions: controlled comparison between CT and 1.5-T MR imaging with non-enhanced and gadolinium-enhanced FAT suppressed spin-echo and breath-hold FLASH techniques. *Radiology* 1992; **182**: 425–30.

Semelka RC *et al.* Islet cell tumors: comparison of dynamic contrast-enhanced CT and MR imaging with dynamic gadolinium enhancement and fast suppression. *Radiology* 1993; **186**: 799–802.

Semelka RC *et al.* Comparison of breathhold T_1-weighted MR sequences for imaging of the liver. *Journal of Magnetic Resonance Imaging* 1994; **4**: 759–65.

Semelka RC *et al.* Hepatic hemangiomas: a multi-institutional study of appearance on T_2-weighted and serial gadolinium-enhanced gradient-echo MR images. *Radiology* 1994; **192**: 401–6.

Semelka RC *et al.* Solitary hepatic metastases: comparison of dynamic contrast-enhanced CT and MR imaging with fat-suppressed T_2-weighted, breath-hold T_1-weighted FLASH and dynamic

gadolinium-enhanced FLASH sequences. *Journal of Magnetic Resonance Imaging* 1994; **4**: 319–23.

Siegelman ES, Schnall MD. Contrast-enhanced MR imaging of the bladder and prostate. *MRI Clinics of North America* 1996; **4**: 153.

Siewart B *et al.* Fast MR imaging of the liver: quantitative comparison of techniques. *Radiology* 1994; **193**: 37–42.

Sipponen JT, Sepponen RE, Sivula A. Nuclear magnetic resonance (NMR) imaging of intracerebral haemorrhage in the acute and resolving phases. *Journal of Computer Assisted Tomography* 1983; **7**: 959.

Soto JA *et al.* Magnetic resonance cholangiography: comparison with endoscopic retrograde cholangiopancreatography. *Gastroenterology* 1997; **110**: 589–7.

Soyer P *et al.* Hepatic metastases from colorectal cancer: detection and false-positive findings with helical CT during arterial portography. *Radiology* 1994; **193**: 71–4.

Stark DD. Hepatic iron overload: paramagnetic pathology. *Radiology* 1991; **179**: 333.

Stark DD *et al.* Magnetic resonance imaging and spectroscopy of hepatic iron overload. *Radiology* 1985; **154**: 137–42.

Stark DD *et al.* Detection of hepatic metastases by magnetic resonance: analysis of pulse sequence performance. *Radiology* 1986; **159**: 365–70.

Stark DD, Moss AA, Goldbert HI, Davis PL, Federle MP. Magnetic resonance and CT of the normal and diseased pancreas: a comparative study. *Radiology* 1984; **150**: 152–62.

Stark DD, Weissleder R, Elizondo G, Hahn PF, Todd LE, Ferrucci JT. Superparamagnetic iron oxide: clinical application as a contrast agent for MR imaging of the liver. *Radiology* 1988; **168**: 297–301.

Stark DD, Wittenberg J, Butch RJ, Ferrucci JT Jr. MR detection of hepatic metastases: a randomized controlled comparison with CT. *Radiology* 1987; **165**: 339–406.

Stark DD, Wittenberg J, Middleton M. Ferrucci JT. Liver metastases: detection by phase-contrast MR imaging. *Radiology* 1990; **158**: 287–90.

Uri DS. MR imaging of shoulder impingement and rotator cuff disease. *Radiology Clinics of North America* 1997; **35**: 77.

Zagoria RJ, Bechtold RE, Dyer RB. Staging of renal adenocarcinoma: role of various imaging procedures. *American Journal of Roentgenology* 1995; **164**: 363–70.

13.4 Gastrointestinal radiology

Daniel J. Nolan

Role of plain abdominal radiographs in diagnosis of the acute abdomen
Examination technique
Pneumoperitoneum
Small-intestinal obstruction
Large-intestinal obstruction
Acute colitis
Mesenteric infarction
Paralytic ileus
Contrast studies of the gastrointestinal tract
Oesophagus
Stomach
Duodenum
Small intestine
Colon
Computed tomography
Angiography
Further reading

Role of plain abdominal radiographs in diagnosis of the acute abdomen

Plain radiographs are the initial imaging procedure performed in most patients who present with suspected acute disorders of the gastrointestinal tract. Radiographs of the abdomen and chest can provide essential diagnostic information.

Examination technique

A supine view of the abdomen and an upright view of the chest are the basic views considered essential in most patients. The upright chest radiograph is an important part of the examination: this is the most reliable view for showing free intraperitoneal air, and pain from pleural or lung disorders may present initially with abdominal pain. Decubitus and upright views of the abdomen are also occasionally helpful. It may be necessary to proceed to contrast studies, ultrasonography, or computed tomography (**CT**) if plain radiographs are unhelpful or inconclusive.

Pneumoperitoneum

Spontaneous pneumoperitoneum normally indicates that a duodenal or gastric ulcer has perforated, or that the colon has perforated due to diverticulitis, acute colitis, carcinoma, or trauma. Perforation of the small intestine is uncommon. Free intraperitoneal air is demonstrated in 60 to 90 per cent of plain radiographs performed in patients with pneumoperitoneum, depending on how carefully the examination is done. The upright posteroanterior chest and left lateral decubitus (right side up) abdominal radiographs are the best views for demonstrating the presence of pneumoperitoneum.

With good radiographic technique as little as 1 ml of intraperitoneal air can be detected. Free intraperitoneal air is demonstrated as a sickle-shaped collection of air between the liver and the diaphragm on the chest view (Fig. 1), and between the liver and the abdominal wall on the left lateral decubitus view. Larger collections of air may outline the liver. When a relatively large amount of free intraperitoneal air is present, characteristic signs may be seen on the supine radiograph. These include gas outlining the outer wall of the intestine (Rigler's sign), a triangular collection of gas between intestinal loops, and gas outlining the gallbladder, the lower border of the liver, lesser sac, and the falciform ligament (Fig. 2). The characteristic ‘football sign' is seen most frequently in infants, when a large amount of gas outlines the lateral limits of the peritoneal cavity. A number of other signs that may be identified on the supine abdominal radiograph include the urachus, lateral umbilical ligament, right upper-quadrant, hepatic-edge, and hyperlucent liver signs, and visualization of the diaphragmatic muscle slips.



Fig. 1. Pneumoperitoneum, chest radiograph: a collection of air is noted between the liver and right hemidiaphragm.

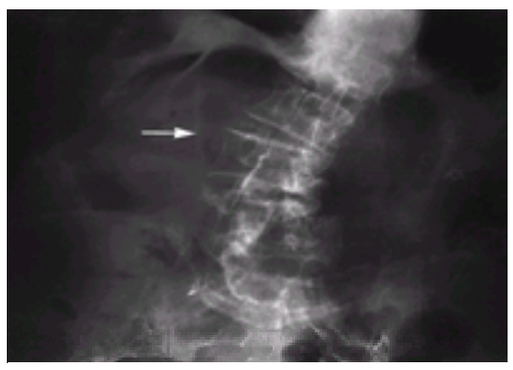


Fig. 2. Pneumoperitoneum: free intraperitoneal gas is seen outlining the falciform ligament (arrow) on a supine view of the abdomen.

Small-intestinal obstruction

Causes of small-intestinal obstruction include adhesions, bands, hernias, strictures, inflammatory lesions such as appendix abscess, diverticulitis, Crohn's disease, and neoplasms.

Obstruction can be diagnosed on plain abdominal radiographs in 60 to 70 per cent of patients, and the supine abdominal view is the most reliable for making the diagnosis. Typical features are gas-distended loops of jejunum and ileum arranged in transverse loops across the central portion of the abdomen (Fig. 3). Little or no gas is seen in the colon in most patients with obstruction of the small intestine, but a moderate or normal amount of colonic gas may be present if the lumen of the small intestine is not completely occluded. If the obstructed loops are fluid-filled they are more difficult to identify, but an upright view in such patients shows the classical ‘string of beads' sign due to multiple small collections of gas above the fluid (Fig. 4). This is diagnostic of small-intestinal obstruction, even in the absence of

gas-distended loops of intestine.



Fig. 3. Small-intestinal obstruction: dilated, gas-filled loops of small intestine are seen in the centre of the abdomen; very little gas is present in the colon. Note the nasogastric tube in the stomach.

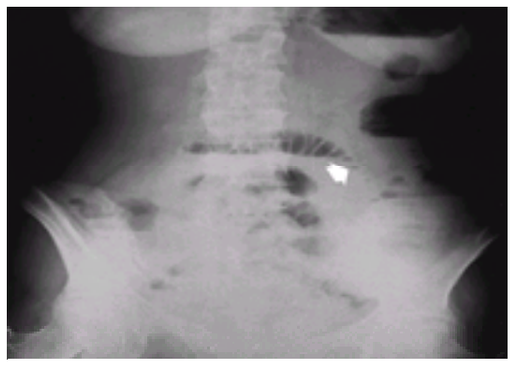


Fig. 4. Small-intestinal obstruction: upright view of the abdomen shows the typical ‘string of beads' sign (arrow).

Plain abdominal radiographs may have a normal appearance in patients with small-intestinal obstruction, due to vomiting in cases of high obstruction or because of the intermittent nature of the obstruction.

Closed-loop obstruction occurs when a single U- or C-shaped loop of intestine is obstructed by a band or twist compressing the adjacently positioned proximal and distal ends of the segment. Gas and/or fluid may be seen in a round or oval loop that remains constant in position on different views.

The characteristic appearances of gallstone ileus, caused by impaction of a gallstone in the small intestine, include evidence of intestinal obstruction, visualization of the obstruction calculus and, in about one-third of cases, air in the biliary tree.

Large-intestinal obstruction

The plain radiographic appearance of obstruction of the large intestine will depend on whether or not the ileocaecal valve is competent. When it is there is usually considerable dilatation of the colon as far as the obstruction, including marked caecal dilatation, usually with no dilatation of the small intestine. The ileocaecal valve is incompetent in most patients, and dilatation of the colon and small intestine is seen, with the caecum only showing slight dilatation. Fluid-filled distension of the proximal colon is seen when the obstructing lesion is proximal to the splenic flexure.

The site of transition between gas- or fluid-filled colon and collapsed, empty distal colon normally identifies the site of obstruction. If there is any doubt about the diagnosis, an instant single-contrast barium or water-soluble contrast enema, done with the contrast medium passing as far as the dilated colonic segments, is likely to confirm the presence or absence of obstruction. When obstruction is confirmed the cause is frequently identified.

Caecal volvulus should be suspected when a haustrated and disproportionately enlarged air-filled viscus is seen anywhere in the abdomen; the caecum is usually absent from the right iliac fossa and distended small intestine is seen to the right of the dilated caecum. Sigmoid volvulus can frequently be diagnosed on plain abdominal radiographs: the characteristic appearance is that of a grossly enlarged, gas-filled sigmoid colon arising from the pelvis and deviating to the left or right flank. The apex of the loop is positioned high in the abdomen, and may lie under and elevate the diaphragm. Three dense, curved lines, representing the walls of the enlarged loop, converge towards the stenosis over the left part of the sacrum (Fig. 5).



Fig. 5. Sigmoid volvulus: the sigmoid colon is grossly distended with gas and the walls of the enlarged loop converge over the left part of the sacrum; the ascending and transverse colon are loaded with faeces and are normal in size. Note Paget's disease involving the left hemipelvis.

Acute colitis

The supine view of the abdomen frequently yields important diagnostic information in patients with acute ulcerative colitis. When air is present in the colon the mucosal edge and haustral pattern give an indication of the severity of the inflammation. In the segments where there is faecal residue, active mucosal disease is unlikely. Patients with toxic megacolon, a potentially lethal complication of ulcerative colitis, show dilatation of the transverse colon (exceeding 5.5 cm in width). Other signs of toxic megacolon include loss of the normal haustral pattern, an irregular contour to the colonic wall, and numerous, broad-based mucosal islands (inflammatory polyps) projecting into the lumen of the dilated segment. Perforation is a serious complication of toxic megacolon.

Mesenteric infarction

Acute mesenteric ischaemia and infarction may occur when emboli, arising in the heart following atrial fibrillation, myocardial infarction, a left atrial myxoma, or from deep venous thrombosis via a patient foramen ovale, lodge in the superior mesenteric artery. Other causes of mesenteric infarction include cardiogenic shock, and penetrating or blunt trauma to the abdomen. Plain abdominal radiographs show distended loops of small intestine shortly after the onset of symptoms; the number and

size of the distended loops increase later. Specific radiological signs may develop, including thickening and oedema of the valvulae conniventes, thickening of the intestinal wall, air in the intestinal wall (Fig. 6), and air in the intrahepatic portal veins.



Fig. 6. Intestinal infarction: the small intestine is dilated and there is extensive gas in the wall of the intestine.

Paralytic ileus

Paralytic (adynamic) ileus is one of the more common forms of intestinal obstruction and usually occurs throughout the gastrointestinal tract, although occasionally involving only one segment. It is a risk that is present for 3 to 4 days after an abdominal operation. Other causes of paralytic ileus include intestinal ischaemia, sepsis, intraperitoneal inflammation such as acute appendicitis, cholecystitis, pancreatitis, retroperitoneal haematoma, fracture of the spine, ureteric colic, thoracic lesions such as basal pneumonia, rib fractures, or myocardial infarction.

Dilated loops of small and large intestine are frequently seen on plain abdominal radiographs. It may be impossible to distinguish adynamic ileus from obstruction, and a contrast study may be required to establish the correct diagnosis. When there is localized inflammation, such as in appendicitis, cholecystitis, or pancreatitis, the ileus may develop in one or two adjacent loops of small intestine called ‘sentinel loops'.

Contrast studies of the gastrointestinal tract

Barium-enhanced examination of the upper gastrointestinal tract is used to evaluate the oesophagus, stomach, and duodenum. The double-contrast barium examination is quick Jai and easy to perform, and takes about 10 to 15 min.

High-density barium is used to coat the mucosal surfaces and a gas-producing agent is used to distend the stomach and duodenum. An intravenous injection of hyoscine butylbromide (Buscopan) or glucagon is given to produce smooth-muscle relaxation. Double-contrast views of the oesophagus are obtained when the barium is swallowed quick Jaily, so that the swallowed air distends the oesophagus enabling mucosal views to be obtained.

Oesophagus

Dysphagia (difficulty in swallowing) is a distressing symptom and can be oropharyngeal or oesophageal. Oesophageal dysphagia is due to a mechanical obstruction caused by narrowing of the lumen or a motility disorder. The barium swallow is the ideal initial procedure for examining patients who present with dysphagia. It is easy and quick to perform, the only preparation required is fasting for a short period beforehand, the patient can resume normal activities immediately afterwards, and both structural and functional oesophageal lesions can be identified.

Zenker's diverticula (Fig. 7) are often related to gastro-oesophageal reflux and there is an association between them and benign oesophageal strictures. For this reason, a barium swallow should always be performed in patients who present with dysphagia, before proceeding to endoscopy, as the diverticulum may be perforated during the procedure with serious consequences.

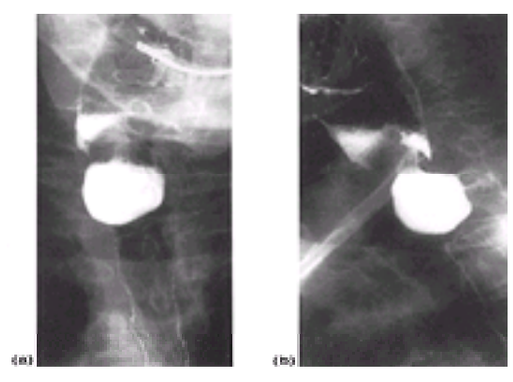


Fig. 7. Zenker's diverticulum: (a, b) anteroposterior and lateral views of the upper oesophagus show a moderate-sized diverticulum.

In patients with mechanical obstruction the site, extent, and nature of the obstructing lesion is shown. Cervical oesophageal webs, a fairly frequent finding not normally diagnosed at endoscopy, are shown as thin, shelf-like filling defects protruding into the lumen of the oesophagus from the anterior wall. Webs are ring-like constrictions covered with squamous epithelium. Some webs are circumferential and are seen with a jet of barium passing through the centre (Fig. 8). There is now no evidence of an association between oesophageal webs and anaemia or sideropenia, the so-called Brown Kelly–Patterson or Plummer–Vinson syndrome.

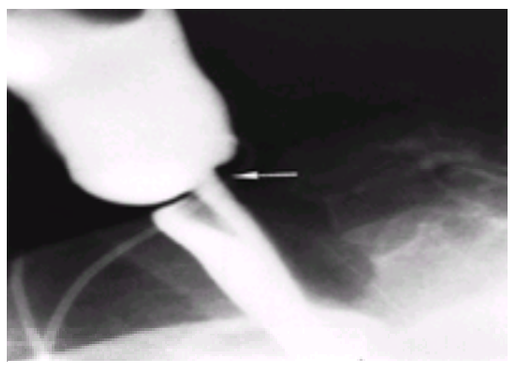


Fig. 8. Oesophageal web: lateral view of the cervical oesophagus shows a typical web with circumferential, shelf-like narrowing; note the jet of barium (arrow) passing through the web. Reproduced from Phillips and Nolan (1995) with permission.

Other causes of mechanical obstruction include carcinoma of the oesophagus, carcinoma of the fundus of the stomach invading the lower oesophagus, and benign

strictures and extrinsic neoplasms compressing the oesophagus. A food bolus may impact in the oesophagus during severe oesophageal spasm.

Carcinoma of the oesophagus appears on barium studies as an irregular stricture with mucosal destruction and shouldering of the margins, as an infiltrating constricting lesion, or as an irregular polypoid mass (Fig. 9).

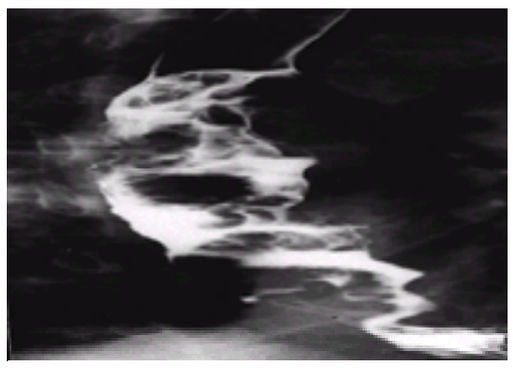


Fig. 9. Carcinoma of the oesophagus: an extensive, irregular, polypoid mass is seen involving the lower half of the oesophagus.

When carcinoma of the fundus invades the lower oesophagus the primary neoplasm may be obvious as a large gastric mass. In other cases a carcinoma is seen at the oesophagogastric junction, with little indication of whether the neoplasm originates in the lower oesophagus or the fundus of the stomach.

The oesophagus may be compressed by enlarged, neoplastic, mediastinal lymph glands or primary carcinoma of the bronchus. Carcinoma of the bronchus occasionally invades the oesophagus, resulting in an oesophagobronchial fistula that can be identified on barium swallow.

Most benign strictures result from gastro-oesophageal reflux and are usually located in the lower oesophagus just above the oesophagogastric junction. There may be an associated hiatus hernia. Benign strictures usually appear as smooth segments of narrowing, although there may be some irregularity of the mucosa without mucosal destruction (Fig. 10). Occasionally it is impossible to distinguish a benign oesophageal stricture from primary carcinoma.



Fig. 10. Benign oesophageal stricture: a short segment of smooth narrowing is noted in the lower oesophagus.

Prolonged nasogastric intubation may result in the development of oesophageal strictures, and the accidental ingestion of corrosive acids or alkalis can result in severe damage to the oesophagus with subsequent stricture formation. Certain medications in tabletform, including tetracycline, quinidine, and a potassium chlor-ide, may lodge in the oesophagus at the level of the aortic arch, causing oesophagitis that occasionally progresses to stricture formation.

Hiatus hernias, classified as sliding and rolling, are seen as herniations of stomach through the diaphragmatic hiatus into the thorax. Sliding hiatus hernias are by far the more common type and are present when both the oesophagogastric junction and stomach herniate into the thorax. While only a small amount of stomach may herniate in some patients, in others the whole stomach is affected. Hiatus hernias are reducible when they move in and out of the thorax and non-reducible when part of the stomach remains fixed in the thorax. The rolling type of hiatus hernia, also known as para-oesophageal hernia, is seen when the oesophagogastric junction remains in its normal position below the diaphragm, the stomach herniating into the thorax beside the normally positioned lower oesophagus.

Oesophageal varices represent dilated venous collaterals, and usually result from portal venous hypertension. They are seen on barium examination as serpiginous or oval filling defects, mostly in the lower oesophagus and extending upwards to involve its middle third (Fig. 11). Obstruction to the superior vena cava may result in upper oesophageal varices.

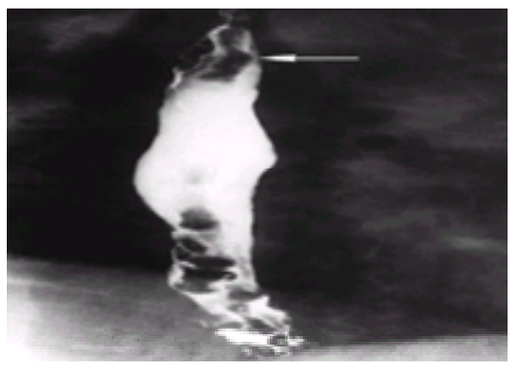


Fig. 11. Oesophageal varices: multiple oval filling defects are seen in the lower oesophagus; a serpiginous filling defect can also be identified in the mid-oesophagus (arrow).

Stomach

The most frequently encountered disorders of the stomach are ulceration and carcinoma. Benign gastric ulcers are seen as small or large, round or oval collections of barium with a surrounding zone of radiolucency due to oedema. Folds frequently radiate from the edge of the ulcer crater. The most frequent sites of gastric ulceration are the lesser curve and the posterior wall of the stomach (Fig. 12). So-called sump ulcers may develop on the greater-curve aspect of the gastric antrum and lower body of the stomach in patients, particularly the elderly individual, taking analgesic medication (Fig. 13). Such ulcers develop because of the combined effect of gravity and the corrosive action of the drugs. Occasionally, these sump ulcers penetrate through the gastric wall and result in the formation of a gastrocolic fistula.



Fig. 12. Gastric ulcer: a small ulcer crater is seen on the posterior wall of the body of the stomach with folds radiating from the edge of the crater.



Fig. 13. Gastric ulcer: a large ‘sump ulcer' (arrow) is seen arising from the greater curve aspect of the lower body of the stomach. Following medical treatment, a repeat examination 5 weeks later showed that the ulcer had completely healed.

Erosions are mostly present in the antrum and appear as small collections of barium with surrounding oedema, often located on gastric mucosal folds.

Carcinoma of the stomach is seen as an ulcerating, polypoid, or infiltrating lesion. Ulcerating carcinomas, sometimes called malignant ulcers, show thickening or distortion of the folds at the edge of the crater and fusion or amputation of the folds by an area of induration at the ulcer edge. Malignant ulcers are often shallow, with a nodular or uneven pattern in the base of the crater, and have ill-defined or irregular outlines. If malignancy is suspected an adequate number of endoscopic biopsies should be obtained at the earliest opportunity.

Polypoid carcinomas appear as irregular, polypoid filling defects in the stomach (Fig. 14). Infiltrating carcinomas characteristically produce marked narrowing of the lumen, and when they involve the whole stomach, show the characteristic ‘linitis plastica' appearance (Fig. 15); mucosal biopsies obtained at endoscopy in patients with linitis plastica often fail to show evidence of malignancy.

Fig. 14. Carcinoma of the stomach: an irregular, polypoid mass is seen in the mid-body of the stomach. Most polypoid carcinomas are larger than this when first detected.

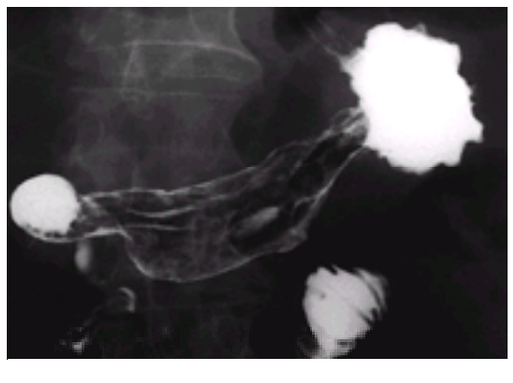


Fig. 15. Infiltrating carcinoma: an extensive infiltrating carcinoma has resulted in contraction of the stomach with irregularity of mucosal folds, the typical ‘linitis plastica' appearance.

The clinical and radiological features of primary gastric lymphoma, which accounts for 2.5 per cent of malignant gastric neoplasms, frequently resemble those of other gastric lesions, particularly carcinoma. Since the prognosis of primary gastric lymphoma is much better than that of carcinoma, accurate diagnosis is important. Characteristic radiological features include a mass that may be partially effaced by the barium, gross hypertrophy of the mucosal folds that become more effaced as the stomach is distended (Fig. 16), and one or more large gastric ulcers seen in association with mucosal hypertrophy. Narrowing and rigidity of the gastric antrum may be seen, sometimes extending across the pylorus into the duodenum. Transpyloric spread of lymphomas of the gastric antrum occurs in 40 per cent of cases.

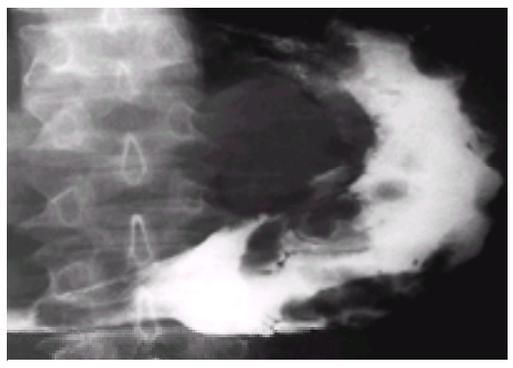


Fig. 16. Gastric lymphoma: there is marked hypertrophy of the mucosal folds in the fundus and body of the stomach.

Duodenum

Benign peptic ulceration is common and is the most frequently encountered disorder of the duodenum. The barium examination is an accurate technique for detecting and demonstrating duodenal ulcers, provided a good double-contrast examination technique is used. Ulcer craters in the duodenum appear as single or multiple, sharply defined, constant collections of barium, sometimes with a surrounding zone of oedema or with folds radiating from the crater. Most ulcers have a diameter of less than 1 cm (Fig. 17). Some patients have considerable deformity of the duodenal cap due to previous ulceration, making it difficult to detect an active ulcer crater. The degree of deformity varies considerably, and, when marked, can result in duodenal stenosis.

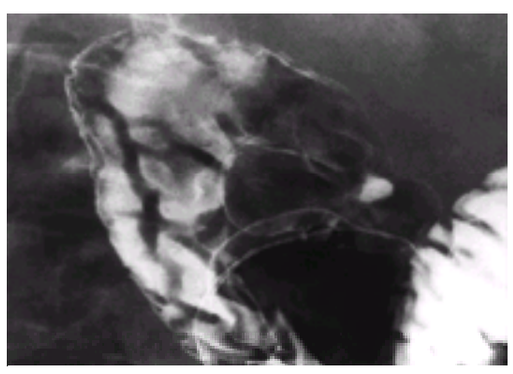


Fig. 17. Duodenal ulceration: a moderate-sized collection of barium is seen outlining a duodenal ulcer crater; there is slight deformity of the duodenal cap and a number of folds radiate from the oedematous edge of the ulcer.

A number of malignant neoplasms may involve the duodenum. Primary neoplasms of the duodenum are uncommon, and can be classified into carcinoma of the papilla of Vater, and true carcinoma of the duodenum. Carcinoma of the papilla of Vater appears as an enlarged papilla with irregular borders, sometimes with ulceration. Non-papillary carcinomas of the duodenum are adenocarcinomas and are seen as ulcerative, polypoid or annular lesions, similar to the appearances of carcinomas in other parts of the gastrointestinal tract.

The duodenum may be invaded by malignant neoplasms from adjacent organs or may be the site of metastatic deposits. Carcinoma of the head of the pancreas frequently involves the duodenal loop, causing widening, a double contour, irregularity of the inner border, or stricture formation. The reversed ‘3' sign of Frostberg is a characteristic but infrequent finding. Carcinoma of the body or tail of the pancreas may invade the distal duodenum. The duodenum may also be invaded by malignant neoplasms in adjacent organs such as the colon, right kidney, and gallbladder. The duodenum may be the site of metastatic deposits from malignancies elsewhere, particularly malignant melanoma. Gastric carcinoma and lymphoma can extend across the pylorus to invade the duodenum. In a recent series, transpyloric spread occurred in 40 per cent of lymphomas and 25 per cent of adenocarcinomas of the gastric antrum.

The duodenum is affected by Crohn's disease in about 4 per cent of patients who have the disease elsewhere in the small intestine or colon. The appearances are similar to those seen in the more distal small intestine. Crohn's disease may cause tubular narrowing of the gastric antrum and proximal duodenum in continuity, resulting in the ‘pseudo post-Billroth I' appearance.

Intramural duodenal haematoma can result from blunt abdominal trauma, anticoagulant therapy, or blood dyscrasias. On barium studies an intramural haematoma is seen as a concentric obstructive lesion in the second or third part of the duodenum, sometimes giving a ‘coiled spring' appearance.

Duodenal diverticula are seen fairly frequently on barium examination and have little clinical significance in the majority of patients.

Small intestine

The barium examination is the investigation of choice in patients with known or suspected disorders of the small intestine who do not present acutely. The barium follow-through and enteroclysis (small-bowel enema) are the principal techniques. The terminal ileum is also shown fairly frequently when barium suspension refluxes through the ileocaecal valve during barium enema investigations. Enteroclysis is now fairly widely used, although the follow-through continues to be used as the method of choice in many centres. Enteroclysis is performed by infusing the barium suspension directly into the intestine through a special tube positioned so that its tip is in the distal duodenum or proximal jejunum. Either a dilute, single-contrast, barium suspension method or a double-contrast technique using a moderate density barium suspension followed by an aqueous suspension of methylcellulose are used. During enteroclysis there is excellent distension and visualization of the jejunum and ileum.

Although the comparative diagnostic yield of both enteroclysis and the follow-through is difficult to assess, the available evidence indicates that enteroclysis is superior to the follow-through for detecting and demonstrating morphological abnormalities in the intestine. We reviewed 1465 patients examined during a 10-year period by enteroclysis and found that the sensitivity was 93.1 per cent and the specificity 96.9 per cent for detecting the lesion responsible for the patient's presenting symptoms.

Disorders that cause morphological changes in the small intestine and that are particularly well shown by enteroclysis include Crohn's disease, neoplasms, chronic radiation enteritis, ischaemia, Meckel's diverticulum, jejunal diverticulosis, tuberculosis, and non-steroidal anti-inflammatory drug enteritis.

The many and diverse signs of Crohn's disease include ulceration shown as discrete ulcers, fissure ulcers, longitudinal ulcers, sinuses and fistulas (Fig. 18) as well as thickening of the valvulae conniventes, cobblestoning, skip lesions, asymmetrical involvement, a featureless outline, and enlargement of the ileocaecal valve.

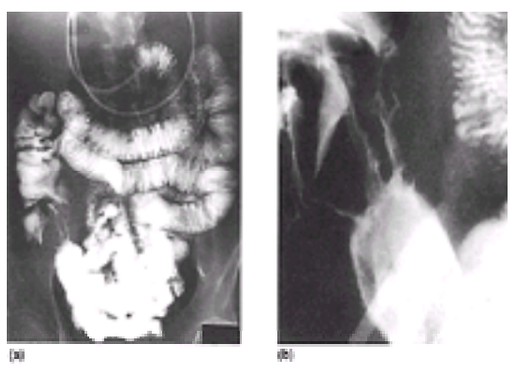


Fig. 18. Ileal Crohn's disease: (a, b) the distal ileum is grossly abnormal with the lumen of the terminal ileum replaced by a number of ileocaecal fistulas; there is slight dilatation of the intestine proximal to the fistulas. The patient present with abdominal pain, weight loss, and diarrhoea; on physical examination a mass was palpated in the right iliac fossa.

Primary neoplasms are uncommon in the small intestine. The available evidence indicates that enteroclysis is superior to the barium follow-through for detecting and demonstrating neoplasms. Primary carcinoma is nearly always located in the jejunum, particularly the proximal jejunum, with radiological appearances similar to those of carcinoma of the colon. Characteristically, carcinoid tumour is shown as an intraluminal or intramural filling defect in the distal ileum. Lymphomas, multiple in 40 per cent of patients, are mostly seen in the ileum, often as an ulcerating or cavitating mass. Leiomyosarcomas and metastases or other malignant neoplasms are frequently shown as a cavitating mass. Benign leiomyomas are mostly seen as round, intraluminal filling defects.

Chronic radiation enteritis develops in a small number of patients following radiotherapy to the abdomen and pelvis. The radiological changes include thickening of the valvulae conniventes, single or multiple stenoses, adhesions, mural thickening, mucosal tacking, sinuses, and fistulas.

Arterial and venous ischaemia result in thickening of the valvulae onniventes, which can be rather marked. Ischaemic strictures, including those resulting from trauma, may present with intestinal obstruction. Meckel's diverticulum is seen as a solitary pouch arising from the antemesenteric border of the ileum. Tuberculosis is uncommon and when seen is shown as ulceration and/or stricture formation, mostly involving the terminal ileum or ileocaecal junction. Non-steroidal anti-inflammatory drugs can cause ‘diaphragm-like' strictures in the small intestine.

Barium examination in patients with small-intestinal obstruction shows the site of obstruction as an abrupt transition between the distended or dilated small intestine proximal to the obstruction site and the collapsed distal loops. Causes of obstruction that may be identified include neoplasms, Crohn's strictures, adhesions, or internal hernias.

Colon

The barium enema remains a widely used technique for the detection of carcinomas and adenomas in the colon, the diagnosis and evaluation of diverticular disease and its complications, and for assessing the extent and severity of inflammatory bowel disease. The double-contrast barium technique is the method of choice in most centres, although a single-contrast examination is performed in patients with suspected colonic obstruction. Water-soluble contrast studies are done mostly to examine the anastomosis in patients who have undergone recent resection of part of the colon. Digital examination of the rectum and sigmoidoscopy are the initial diagnostic procedures in the investigation of colonic disorders, and these should always be done before a barium enema is requested. If a rectal biopsy is performed, an interval of at least 7 days should be allowed before a barium enema to avoid the risk of perforation.

A clean colon, a satisfactory barium suspension, the use of a smooth-muscle relaxant, and good examination technique are necessary to obtain consistently good results. The colon is cleansed by using a combination of cathartics, low-residue diet, and increased fluid intake on the day before the examination. A cleansing enema may be required on the morning of the examination. The examination is made by infusing barium into the colon and replacing much of the barium with air to give a double-contrast effect before taking the radiographs. Hyoscine butylbromide or glucagon is given intravenously to produce smooth-muscle relaxation.

Most carcinomas of the colon have reached a fairly advanced stage by the time the patient presents with clinical symptoms. Carcinoma is shown as a constricting lesion with mucosal destruction, a narrow, irregular lumen and shouldered margins, with a sharp transition between the neoplasm and adjacent normal colon. In some cases the carcinoma is seen as an irregular, intraluminal, polypoid filling defect (Fig. 19); other carcinomas appear as asymmetrical infiltrating lesions with mucosal destruction.



Fig. 19. Carcinoma of the colon: a large, irregular, polypoid filling defect is shown in the distal sigmoid colon on barium enema examination.

Adenomas are shown as either sessile (Fig. 20) or pedunculated, small filling defects, usually less than 1 cm in diameter. Villous adenomas may be larger and have a frond-like appearance.

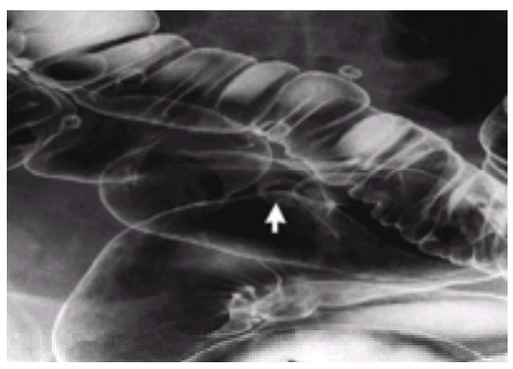


Fig. 20. Adenomatous polyp: barium enema shows a small, sharply defined, irregular filling defect in the mid-sigmoid colon (arrow); a number of diverticula can also be identified.

Flexible sigmoidoscopy, and colonoscopy, are now widely available for examining the colon. The relative roles of these and barium studies have not been established. A small number of centres now perform flexible sigmoidoscopy before barium enemas. The majority of polyps and carcinomas develop in the sigmoid colon and investigation by combined flexible sigmoidoscopy and barium enema improves the detection rate, particularly in patients with diverticular disease, in whom small carcinomas and polyps may be obscured by the diverticula.

Ulcerative colitis, Crohn's colitis, and ischaemic colitis account for the great majority of patients with inflammatory bowel disease who are assessed with barium studies. The diagnosis of ulcerative colitis should be firmly established from a rectal biopsy taken at sigmoidoscopy. The double-contrast barium enema is an excellent technique for demonstrating the extent and severity of inflammation and the presence or absence of an associated carcinoma (Fig. 21). Mucosal ulceration may be mild, moderate, or severe and will extend in a proximal direction from the rectum in continuity to involve part or all of the colon, with associated loss of the normal haustral pattern. Patients with chronic ulcerative colitis show narrowing of the lumen and shortening of the colon.

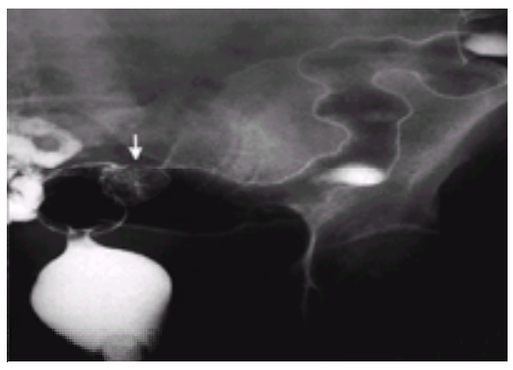


Fig. 21. Ulcerative colitis and carcinoma: the sigmoid colon is shortened and shows minimal mucosal ulceration, consistent with mild active inflammation in chronic ulcerative colitis; a polypoid carcinoma is also seen in the distal sigmoid colon (arrow).

The typical features of Crohn's colitis are inflammation, often in the form of aphthous ulcers, strictures, asymmetrical lesions, skip lesions, and predominant involvement of the right side of the colon. The distal sigmoid colon and rectum are nearly always spared, although perianal sinuses and fistulas (Fig. 22) are a recognized feature. Ischaemic colitis characteristically shows oedematous changes in the splenic flexure that either return to normal in about 4 to 6 weeks, or result in stricture formation.



Fig. 22. Crohn's colitis: there is rather marked shortening of the proximal colon and two tight strictures can be identified in the transverse colon (arrows); note also the perianal sinus track (open arrow).

Diverticular disease is a frequent finding in middle-aged and older patients. Diverticula appear as single or multiple small outpouchings, most frequently in the sigmoid colon, although they may be present in any part of the colon. Acute inflammation (diverticulitis) complicates diverticular disease in a small minority of patients. A paracolic abscess may be shown on barium enema as displacement and narrowing of the intestinal lumen with an altered mucosal pattern. Unlike carcinoma of the colon, the mucosal pattern in the narrowed segment is intact, although it may be distorted, and there is no shouldering of the margins. In some cases it may be impossible to distinguish a paracolic inflammatory mass from carcinoma. A soft-tissue mass, gas lucency, air–fluid level, or barium in an extraluminal cavity may be seen in diverticulitis. The characteristic drape sign is occasionally seen, and is caused by bending of adjacent empty diverticula towards the abscess. Paracolic abscess may result in colonic or sometimes small-intestinal obstruction. Fistulas may be identified as tracks of contrast medium passing from the colon to adjacent viscera. The more common fistulas are colovesical and coloenteric; fistulas to the skin, genital tract, ureter, stomach, hip, perineum, and soft tissues of the thigh are less common. A paracolic fistula is seen as a longitudinal track of barium running parallel to the colon in the paracolic tissues.

Computed tomography

Computed tomography (**CT**) is only occasionally used as the initial investigation when disorders of the hollow organs of the digestive system are suspected. It can, however, provide useful further information about neoplasms and other conditions that involve the gastrointestinal tract.

Initial optimism that CT would be a reliable method for the presurgical staging of oesophageal carcinoma has not been confirmed. It is sensitive in detecting liver metastases and invasion of the tracheobronchial tree, but is unreliable for assessing mediastinal soft-tissue extension and aortic invasion. The newer technique of endoscopic ultrasonography is proving to be more accurate for staging oesophageal carcinoma, since the depth of infiltration can be accurately assessed and lymph-node metastases can be detected.

CT has a limited role in evaluating the stomach; it is most useful for the preoperative staging of gastric carcinoma and for helping to confirm the diagnosis of linitis plastica. The characteristic appearances of gastric varices on CT can be helpful when the diagnosis is difficult at barium studies or endoscopy. The role of CT in visualizing the duodenum is mostly limited to showing changes in adjacent organs, such as the pancreas, that also involve the duodenum.

CT has an increasingly important role in evaluating the small intestine. It provides information about the extraluminal extent of intestinal disorders such as neoplasms and the complications of Crohn's disease, and it is invaluable for diagnosing intestinal ischaemia and infarction. On CT, extensive carcinoid tumour has a characteristic pattern of a mass with a stellate radiating pattern of mesenteric neurovascular bundles. CT is now the preferred imaging technique for evaluating high-grade small-intestinal obstruction. Excellent contrast is provided by the fluid-filled intestinal loops proximal to the site of obstruction, enabling a diagnosis to be confirmed while showing the site and likely cause of obstruction (Fig. 23). CT is particularly useful for diagnosing strangulating obstruction.

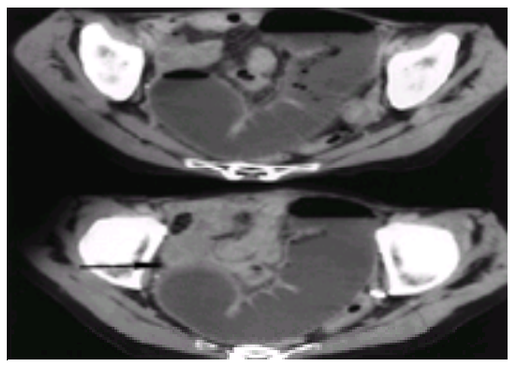


Fig. 23. CT shows small-intestinal obstruction caused by adhesions (arrow) [reproduced from Nolan (1998) with permission].

In the colon, CT is important for staging colorectal carcinoma and for evaluating patients who have undergone surgical resection for carcinoma. It is a sensitive method for detecting local and distant recurrent neoplasm. CT is being used increasingly in the initial evaluation of suspected acute diverticulitis. Sigmoid diverticulitis is seen on CT as localized thickening of the colonic wall in association with inflammatory changes in the pericolic fat or an adjacent abscess. Fistulas, abscesses, intestinal or ureteric obstruction, and peritonitis are complications of diverticulitis that can be identified on CT.

Virtual colonoscopy is a new method of imaging the colonic mucosa that uses interactive virtual-reality computer software to simulate conventional colonoscopy with

images obtained using high-performance, helical CT scanners. This technique is at an early stage of development but in the future could play a leading part in the diagnosis of colorectal cancer.

Angiography

The main indication for gastrointestinal angiography is in the diagnosis and treatment of gastrointestinal bleeding. If endoscopy fails to identify the origin of an acute bleeding episode, selective catheterization of the coeliac axis, superior mesenteric artery, and inferior mesenteric artery is normally performed. Angiography is successful in locating the bleeding site in up to 90 per cent of patients who continue to bleed during the investigation. Embolization may be undertaken during angiography in patients who are unsuitable for surgery. Single vessels such as the left gastric, gastroduodenal, and gastroepiploic arteries can be embolized because of the rich collateral circulation in the upper gastrointestinal tract.

Diagnostic angiography has an important role in detecting the source of chronic bleeding of obscure origin. In many cases obscure bleeding is from the small intestine and superselective arteriography may be required to accurately pinpoint its site. If a lesion demonstrated by angiography in the small intestine is likely to be difficult or impossible to identify at surgery, intraoperative angiography is indicated. The catheter is left in the superselected branch vessel supplying the lesion during the subsequent operation. The anatomical location of the abnormality is then confirmed by an intraoperative angiogram of the individual loops. A modification of the technique involves injecting a small amount of methylene blue through the superselectively placed catheter.

Angiodysplasia, usually located in the caecum and ascending colon, can cause obscure gastrointestinal bleeding. It cannot be identified on barium studies but it may be recognized by an experienced endoscopist. Angiography is an excellent technique for diagnosing this condition.

Meckel's diverticulum, a well-recognized cause of obscure bleeding in adults, can be diagnosed angiographically. Angiography shows a persistent vitellointestinal artery in most patients with Meckel's diverticulum. Superselective catheterization of the distal ileal arteries may be necessary.

Further reading

Balthazar EJ. CT of small bowel obstruction. *American Journal of Roentgenology* 1994; **162**: 255–61. [An excellent review of the role of CT in small-bowel obstruction.]

Baron JH. The Patterson Brown–Kelly syndrome of sideropenic dysphagia does not exist. *Journal of the Royal College of Physicians* (London) 1991; **25**: 361. [This letter summarizes the evidence refuting the existence of the syndrome.]

Cho KC, Baker SR. Extraluminal air: diagnosis and significance. *Radiology Clinics of North America* 1994; **32**: 829–44. [A review of all the radiological signs of free intraperitoneal air.]

Cho KC, Baker SR, Altermann DD, Fusco JM, Cho S. Transpyloric spread of gastric tumours: comparison of adenocarcinoma and lymphoma. *American Journal of Roentgenology* 1996; **167**: 467–9. [This study shows that the spread of gastric neoplasms to the duodenum occurs quite frequently.]

Dixon PM, Roulston ME, Nolan DJ. Small bowel enema: a ten year review. *Clinical Radiology* 1993; **47**: 46–8. [A study on a large number of patients assessing the accuracy of enteroclysis.]

Fenlon HM, Clarke PD, Ferrucci JT. Virtual colonoscopy: imaging features with colonoscopic correlation. *American Journal of Roentgenology* 1998; **170**: 1303–9. [An ideal introduction to virtual colonoscopy.]

Gourtsoyiannis NC, Nolan DJ, eds. *Imaging of small intestinal tumours*. Elsevier, Amsterdam, 1997. [This book describes all aspects of imaging small-intestinal neoplasms and includes a chapter on pathology.]

Greenall MJ, Levine AW, Nolan DJ. Complications of diverticular disease: a review of the barium enema findings. *Gastrointestinal Radiology* 1983; **8**: 353–8. [Probably the last review on the subject.]

Mendelson RM, Nolan DJ. The radiological features of chronic radiation enteritis. *Clinical Radiology* 1985; **36**: 141–8. [A review of the clinical and radiological aspects of radiation enteritis.]

Miller RE, Nelson SW. The roentgenologic demonstration of tiny amounts of free intraperitoneal gas: experimental and clinical studies. *American Journal of Roentgenology* 1971; **112**: 574–85. [An excellent study showing how well tiny amounts of free intraperitoneal air can be shown on plain radiographs.]

Mitchell AWM, Spencer J, Allison DJ, Jackson JE. Meckel's diverticulum: angiographic findings in 16 patients. *American Journal of Roentgenology* 1998; **170**: 1329–33. [Demonstrates the value of meticulous angiographic technique.]

Nolan DJ. Radiological examination of the small intestine. In Gourtsoyiannis NC, Nolan DJ, eds. *Imaging of small intestinal tumours*, pp. 15–27. Elsevier, Amsterdam, 1997. [A detailed description of the enteroclysis technique.]

Nolan DJ. The true yield of the small intestinal barium study. *Endoscopy* 1997; **29**: 447–53. [A comprehensive review of the literature on the accuracy of different barium techniques in the small intestine.]

Nolan DJ. Imaging of the small intestine. *Schweizerische Medizinische Wochenschrift* 1998; **128**: 109–14. [A review of the current methods available for imaging the small intestine.]

Nolan DJ, Traill ZC. Review: the current role of the barium examination of the small intestine. *Clinical Radiology* 1997; **52**: 809–20. [A comprehensive review of the barium study of the small intestine, particularly the current role of enteroclysis with 134 references.]

Phillips AJ, Nolan DJ. Radiology of oesophageal dysphagia. *British Journal of Hospital Medicine* 1995; **53**: 458–66. [A review of the role of the barium swallow in the investigation of dysphagia.]

Rollins ES, Picus D, Hicks ME, Darcy MD, Bower BL, Kleinhoffer MA. Angiography is useful in detecting the source of chronic gastrointestinal bleeding of obscure origin. *American Journal of Roentgenology* 1991; **156**: 385–8. [An excellent study.]

13.5 Ultrasound imaging

David R. M. Lindsell

The abdominal mass
The hepatobiliary system
Gallbladder
Bile ducts
Liver
Pancreas
Pancreatic tumours
Acute pancreatitis
Chronic pancreatitis
Renal tract
Renal tract obstruction
Renal tract tumours
Prostate
Ultrasound in gynaecology
Abnormal uterine bleeding
Pelvic masses
Pelvic pain
Ectopic pregnancy
Infertility
The gastrointestinal tract
Endoscopic and intraluminal ultrasound
Abdominal abscesses
Interventional ultrasound
Other applications of ultrasound in surgery
Thyroid and parathyroid
Male genital tract
Breast
Musculoskeletal system
Ocular ultrasound
Vascular system
Ultrasound contrast agents
Further reading

The applications of ultrasound in surgical practice are many (Table 1). It is not within the scope of this chapter to consider obstetric ultrasound in detail. The ability to obtain images of the fetus and placenta without the use of ionizing radiation has greatly enhanced obstetric practice, and in the United Kingdom most pregnancies will be scanned at some stage during their course. It is possible to assess accurately gestational age, placental position, and the presence of many different fetal anomalies as well as assessing different parameters in the growth-retarded fetus.

Obstetrics and gynaecology
Abdomen
Infant brain
Eye
Thyroid and parathyroid
Salivary glands
Breast
Heart
Pleura
Joints and soft tissues
Scrotum
Vascular system
Trauma
Interventional procedures

Table 1 The applications of ultrasound in surgical practice

The use of non-ionizing radiation makes ultrasound safe; it is also relatively inexpensive and can be performed at the patient's bedside. A number of different organs can be viewed at the same examination and in a short space of time. Ultrasound is therefore widely practised by surgeons, cardiologists, and neonatologists as well as radiologists.

Ultrasound has limitations in that the interpretation of images obtained is dependent on the skill of the operator and image quality is degraded by obesity and by the presence of bowel gas.

The abdominal mass

Patients with a palpable or suspected abdominal mass are investigated in a variety of ways. If there is a strong clinical suspicion that the mass originates in the colon, for example, then a barium enema or colonoscopy may be performed. As it is unusual for the organ of origin of a mass to be known, patients are frequently referred initially for examination by ultrasound, computed tomography (**CT**), or magnetic resonance imaging (**MRI**). Both CT and ultrasound have a positive predictive value of between 95 and 100 per cent in determining the presence or absence of an abdominal mass, and the predicted organ of origin of a mass is correct in 90 per cent of patients examined with either technique. CT is slightly more sensitive in its ability to define the exact nature of a mass and is better than ultrasound in assessing operability of a mass. The ease, lower cost, and lack of exposure to ionizing radiation make ultrasound the initial investigation of choice.

The hepatobiliary system

Ultrasound examination of the right hypochondrium includes a detailed assessment of liver texture, the bile ducts, gallbladder, pancreas, and right kidney.

Gallbladder

Ultrasound has replaced the oral cholecystogram for the detection of gallstones, because of its greater specificity and because it allows examination of other structures at the same time. There is little difference in the sensitivity of the two examinations in the detection of gallbladder pathology, but it is not possible to state the actual sensitivity, since a negative examination by either technique is often the end point and thus false negative diagnoses are only rarely detected.

In acute cholecystitis the gallbladder is distended and thick walled, and stones will usually be seen (Fig. 1). Pericholecystic inflammatory change may also be visible. As it is possible to localize the position of the gallbladder precisely, direct pressure on this area by the transducer will elicit a positive ultrasound Murphy's sign.



Fig. 1. An acutely inflamed thick-walled gallbladder (open arrow) with a large stone (closed arrow) impacted in Hartmann's pouch.

Percutaneous gallbladder drainage using either ultrasound or CT guidance has become common practice in selected groups of patients.

Bile ducts

The ability of ultrasound to detect bile duct dilatation approaches 100 per cent and it is the primary investigation in patients with suspected biliary obstruction (Fig. 2). If biliary dilatation is detected, the level of obstruction needs to be defined: this is possible in 95 per cent of cases, while the cause of the obstruction can be determined in 85 per cent of patients. The sensitivity varies depending on the obstructing lesion (Table 2). Direct visualization of tumours of the ampulla of Vater is rare, as is the precise diagnosis of sclerosing cholangitis.



Fig. 2. A large stone (arrow) within a dilated common bile duct.

	Sensitivity (%)
Pancreatic masses	97
Choledocholithiasis	82
Cholangiocarcinoma	96
Chronic pancreatitis	67

Table 2 Sensitivity of ultrasound in defining the cause of biliary obstruction

In technically inadequate ultrasound examinations, CT or magnetic resonance cholongiopancreatography (**MRCP**) may be necessary.

Once the diagnosis of biliary obstruction has been made it may be appropriate to perform a CT or MRI scan if ultrasound has failed to demonstrate the cause of obstruction or if surgery is contemplated.

Liver

Although ultrasound has a major role in investigating diffuse hepatocellular disease, in surgical practice it is most commonly used to determine the presence of focal liver lesions (Fig. 3). It is able to differentiate between solid and cystic lesions and if there is any doubt about the nature of a mass, guided aspiration/biopsy can be performed to determine whether the lesion is a benign or malignant tumor or whether it is an abscess or cyst. The use of ultrasound at the time of surgery, with a sterile transducer applied directly to the liver, has led to a reappraisal of the sensitivity of preoperative imaging in the detection of liver tumours. Intraoperative ultrasound is currently the most sensitive method of detecting liver tumours: 40 per cent of the lesions detected by this method are not palpable at surgery.



Fig. 3. Focal areas (arrows) of reduced echogenicity within the liver due to metastases.

Preoperative assessment with ultrasound only has a sensitivity of 60 to 70 per cent in the detection of liver tumours. Helical CT and CT portography have greater sensitivity. In some centres MRI is preferred as it has the ability to differentiate between benign capillary haemangiomas and metastases.

Pancreas

Ultrasound fails to produce a good image of part or all of the pancreas in 25 per cent of patients, due to the presence of overlying bowel gas. In the majority of patients in whom the pancreas is clearly seen, detection of abnormalities is sensitive. If the pancreas is poorly visualized or if surgery is contemplated, CT, MRCP, or endoscopic retrograde cholangiopancreatography (**ERCP**) may be needed.

Pancreatic tumours

Focal masses of altered echogenicity are usually due to adenocarcinoma. Occasionally, guided percutaneous biopsy is required to allow differentiation of such masses from focal areas of pancreatitis or rare benign tumours. Endocrine tumours of the pancreas are often small and difficult to detect preoperatively by any technique: intraoperative ultrasound visualizes these with a sensitivity approaching 100 per cent.

Acute pancreatitis

During an acute attack of pancreatitis the coexisting ileus may obscure the pancreas. CT is the preferred investigation, particularly as this allows an assessment of pancreatic necrosis. Ultrasound can be used to evaluate the gallbladder and biliary tree and to monitor pancreatic fluid collections.

Chronic pancreatitis

With the increasing use of helical CT, ERCP, and MRCP, the role of ultrasound in chronic pancreatitis is diminishing. There is, however, poor correlation between the findings on any imaging study and the degree of pancreatic dysfunction other than in patients with the most severe disease.

Renal tract

Ultrasound reliably demonstrates the shape, size, and echogenicity of the kidneys and the nature of mass lesions. It will also demonstrate renal tract dilatation. The normal non-dilated ureter is not visualized. The bladder and prostate are well seen with both transabdominal and transrectal ultrasound. Ultrasound is therefore the primary imaging modality in renal failure, urinary tract infection, and in the assessment of bladder outflow obstruction.

Renal tract obstruction

Ultrasound is a highly sensitive means for the detection of renal tract dilatation. While dilatation usually indicates obstruction, this is not always so, and obstruction in its early phases does not always produce dilatation. Ultrasound provides no functional information and the level of obstruction may be difficult to discern. For these reasons it will often need to be followed by urography, antegrade or retrograde pyelography, or radionuclide renography. Urinary tract calcification, especially in the ureters, may not always be detected by ultrasound and a plain abdominal radiograph should usually accompany any ultrasound examination of the renal tract. Patients with acute symptoms suggestive of ureteric calculi usually undergo urography rather than ultrasound examination.

Renal tract tumours

Ultrasound is more sensitive and more specific than urography in the detection of renal mass lesions other than tumours of the renal pelvicalyceal system, which are also more reliably diagnosed by CT. Ultrasound may be used to determine whether mass lesions detected by urography are cystic or solid (Fig. 4). Tumours of the ureter are poorly detected by ultrasound unless they are causing hydronephrosis. Tumours of the bladder are better detected by transabdominal ultrasound than urography, their sensitivities being 77 and 61 per cent, respectively. The combination of transabdominal and transrectal ultrasound increases the sensitivity to 95 per cent.



Fig. 4. A solid renal cell carcinoma (closed arrow) arising from the lower pole of the kidney (open arrow).

There is still controversy about the imaging protocol for investigating patients with hematuria. A policy that relies on ultrasound, plain abdominal radiography, and cystoscopy alone runs the risk of missing the rare urothelial tumours of the pelvicalyceal system and ureter. For this reason, some urologists and renal physicians still prefer to investigate all these patients with urography. On the other hand, urography is less sensitive than ultrasound in detecting small renal cell carcinomas and bladder tumours, although the latter will be detected by cystoscopy. One possible diagnostic stratagem is to examine patients with ultrasound and abdominal radiography initially, followed by cystoscopy and if these do not reveal the cause of the haematuria, then to proceed to urography or CT at that point.

Prostate

Transabdominal ultrasound permits a rough two-dimensional volumetric assessment of the degree of intravesical enlargement of the prostate, as well as assessment of residual urine volume. It is therefore preferred to urography in patients with bladder outflow obstruction. Transrectal ultrasound gives improved resolution of images of the prostate itself, allowing breaches of the capsule and adjacent lymph node involvement by tumours to be detected. It also permits targeted biopsy of focal abnormalities.

Ultrasound in gynaecology

The uterus and ovaries may be visualized by both transabdominal and transvaginal scanning. Transvaginal ultrasound has the advantage that the transducer is nearer the midline pelvic structures and may therefore be of a higher frequency, which improves the resolution of the image. This method of scanning is preferred in the early detection of pregnancy, in the assessment of abortion and ectopic pregnancy, and in evaluation of the endometrium. The higher frequency limits the depth of penetration of the ultrasound wave and more laterally situated pelvic masses are often better assessed by transabdominal scanning.

Abnormal uterine bleeding

Evaluation of the endometrium has been greatly enhanced by the use of transvaginal scanning. In many series, an endometrial thickness of 4 mm or less has been shown to have a high negative predictive value for the exclusion of endometrial cancer in women with postmenopausal bleeding. If endometrial thickness is 5 mm or more then endometrial sampling is desirable, recognizing that in many cases this will be negative. Certain types of hormone replacement therapy may lead to a degree of thickening of the endometrium.

In premenopausal women with abnormal bleeding, transvaginal ultrasound allows an assessment of the homogeneity and thickness of the endometrium, allowing the diagnoses of endometrial polyps and submucous fibroids to be made more easily. The new technique of hysterosonography, whereby transvaginal scanning is performed at the time that fluid is instilled into the endometrial cavity, gives better differentiation of endometrial polyps from other forms of endometrial pathology.

Ultrasound is of very limited value in assessing carcinoma of the cervix.

Pelvic masses

The role of ultrasound in the assessment of a pelvic mass is to determine the nature and size of the mass as well as its organ of origin. Ultrasound reliably differentiates between solid and cystic lesions in the pelvis. In general terms most ovarian lesions are cystic and most uterine lesions are solid; there is some overlap, however, and the organ of origin is then inferred by visualizing either the normal ovaries or the normal uterus and excluding these as the cause for the mass. A very large mass may obscure normal structures, making the organ of origin more difficult to predict. Features suggestive of malignancy in ovarian lesions are a multiloculated appearance and the presence of solid elements within the mass (Fig. 5). Additional features suggestive of malignancy are the detection of blood flow in solid elements or septi. Much has been written about the use of Doppler indices, such as the resistive index and pulsatility index, to try and differentiate between benign and malignant ovarian lesions. After initial enthusiasm, there is now a general consensus that there is too great an overlap for these indices to be relied upon for differentiation. In addition, the role of ultrasound in screening the general female population for ovarian cancer has still not been adequately evaluated.

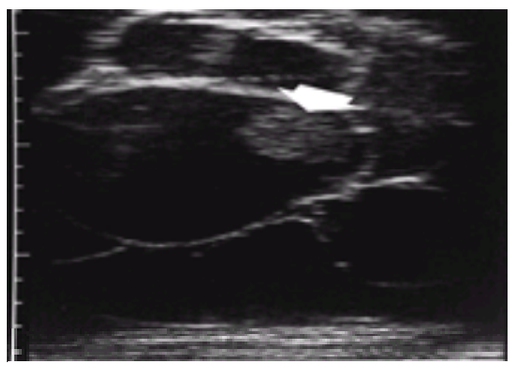


Fig. 5. A multiloculated cystic ovarian tumour containing one small area of solid tissue (arrow).

There does seem to be some value in screening women with a genetically determined higher risk of developing ovarian cancer.

Transvaginal or transabdominal aspiration of adnexal cysts, although somewhat controversial, has been shown to be safe in selected patients.

Pelvic pain

In young women with acute pelvic pain it can be difficult to differentiate between appendicitis, pelvic inflammatory disease, and complications of ovarian cysts. In this clinical setting ultrasound is of value if a positive diagnosis of one of these causes for the pain can be made. In women with chronic pelvic pain, the diagnosis of endometriosis may be made if typical endometriotic cysts are seen. When such cysts are absent, diagnosis is better made by MRI or laparoscopy.

Ectopic pregnancy

The advent of transvaginal ultrasound has increased the likelihood of making a positive diagnosis of ectopic pregnancy. It also permits the earlier diagnosis of an intrauterine pregnancy. Some form of pelvic abnormality will be seen in most patients with an ectopic pregnancy, although the actual ectopic gestation sac and embryo will be seen in a smaller number. As a few patients with ectopic pregnancy will have a virtually normal scan, it is imperative that the scan is always performed in conjunction with an accurate measurement of human chorionic gonadotropin and that there is careful clinical and sonographic correlation.

Infertility

Not only does ultrasound allow an assessment of the morphology of the pelvic organs but it also permits monitoring of cyclical ovarian and endometrial changes. For the purposes of *in vitro* fertilization, transvaginal ultrasound allows accurate needle aspiration of ovarian follicles to harvest ova. More recently, assessment of the patency of the fallopian tubes using transvaginal scanning at the time of instillation of an ultrasound contrast agent into the uterine cavity has become more commonplace.

The gastrointestinal tract

With the exception of the diagnosis of pyloric stenosis in infancy, ultrasound does not have a primary role in the investigation of the gastrointestinal tract. However, tumours of the gastrointestinal tract will often be obvious on ultrasound, although it may be difficult to differentiate these from inflammatory masses such as diverticular masses. Bowel wall thickening and stricture formation is also often apparent in conditions such as Crohn's disease. Ultrasound has a sensitivity of more than 80 per cent and a specificity of 95 per cent in the diagnosis of acute appendicitis without perforation and it has been advocated that it should be used routinely in an attempt to reduce the negative appendectomy rate. The sensitivity is much lower in patients with perforated appendices, although it has been argued that this does not matter as the need for surgery will be more obvious. This fact, and the fact that ultrasound will not detect up to 20 per cent of cases of acute appendicitis, means that it should be used judiciously.

It should probably be confined to patients with equivocal clinical findings and young women, in whom it will exclude a gynaecological cause for the pain. Surgeons should be aware of the high specificity of a positive diagnosis but must equally be aware that a negative result must be treated with caution. CT and MRI are inceasingly also being used in the diagnosis of appendicitis.

Endoscopic and intraluminal ultrasound

Much of the recent development in ultrasound has related to the technological advances that now permit the incorporation of ultrasound transducers into endoscopes and the development of intraluminal transducers. Using transducers attached to upper gastrointestinal tract endoscopes it is now possible to obtain images of the esophagus and heart with a transducer within the oesophagus and to visualize the pancreas, duodenum, and common bile duct with a transducer in the stomach or duodenum. If the endoscope is able to pass through an oesophageal tumour, ultrasound is more sensitive than CT for local staging. It is able to define extension through the layers of the oesophagus, extension to adjacent organs, and spread to local lymph nodes. If the tumour cannot be passed then CT is superior for the detection of mediastinal extension.

Ultrasound examination of the rectum and colon can be performed using either a rectal transducer or a transducer incorporated into a colonoscope. Correlation of the ultrasound results with histological analysis of the resected specimen according to the 1987 TNM classification gives overall accuracy rates of 81 per cent for rectal tumours and 93 per cent for colonic tumours. Some problems have been found with T2 tumours, which may be accompanied by peritumoural inflammation and abscess formation. The possibility of false positive diagnosis of lymph node metastasis is a problem.

Distant metastases within the abdomen can only be detected by abdominal ultrasound, CT, or MRI.

Abdominal abscesses

Ultrasound, CT, MRI, and nuclear medicine techniques may be used to detect intra-abdominal abscesses. No technique is capable of specifically determining whether an identified fluid collection is infected or not, and percutaneous aspiration of the fluid should always be performed. Many comparative studies of the sensitivity of the techniques have been performed: those for ultrasound range from 60 to 100 per cent, for CT from 78 to 100 per cent, and for nuclear medicine techniques from 75 to 100 per cent.

Abscesses in the suphrenic spaces, the right upper quadrant, perirenal areas, and the midline of the pelvis are well seen with ultrasound. CT, unlike ultrasound, has the

advantage of not being affected by open wounds, dressings, and bowel gas and is better able to image the mesentery and retroperitoneum. If ultrasound is inconclusive, then CT will overcome most of the problems that make the ultrasound scan unsatisfactory. Nuclear medicine techniques using either gallium-67 or labelled white cell scanning should be reserved for those patients in whom intra-abdominal sepsis is strongly suspected but in whom a definitive diagnosis cannot be reached by the other two techniques.

Interventional ultrasound

Percutaneous fluid drainage and biopsy procedures can be guided by conventional fluoroscopic screening, CT, or ultrasound. The method chosen will be influenced by the site within the abdomen and the individual preference of the operator. Masses or abscesses within the retroperitoneum, mesentery, or pelvis are often better approached using CT as this provides better visualization of, and therefore avoidance of, the bowel. Superficial lesions, lesions in the liver, and collections in the suphrenic spaces can usually be approached with ease using ultrasound.

The ability to drain abscesses percutaneously has totally changed the management of such patients. Surgery with its greater morbidity, mortality, and cost can often be avoided, or at least deferred until the patient's general condition improves. Success rates of 75 to 85 per cent for percutaneous abscess drainage procedures are common.

Percutaneous biopsy may be performed using either a fine needle to obtain specimens for cytological examination or using larger needles for histological specimens. If accurate cytology is available then fine-needle aspiration biopsy is preferred because of its lower complication rate. Large surveys of the complication rate of fine-needle biopsy in Europe and the United States reveal mortality rates between 0.006 and 0.031 per cent and rates of tumour seeding along the needle track of between 0.003 and 0.009 per cent. Most deaths have been due to haemorrhage following liver biopsy or, less commonly, pancreatitis after pancreatic biopsy.

Many other procedures including nephrostomy, biliary drainage, percutaneous pancreatography, and tumour ablation techniques can be performed using ultrasound guidance.

Other applications of ultrasound in surgery

Thyroid and parathyroid

The ability to differentiate between cystic and solid areas allows assessment of focal or diffuse thyroid enlargement. Most lesions within the thyroid are palpable, and the availability of fine-needle aspiration biopsy for cytology means that there is often no value in imaging thyroid nodules, particularly as ultrasound cannot totally separate benign from malignant lesions.

Ultrasound will usually detect enlargement of the parathyroid glands if they are in a conventional site, but is unable to detect hyperplasia of the glands. In the search for a parathyroid adenoma it should therefore be the first investigation (Fig. 6). Ultrasound can be used to guide aspiration biopsy of impalpable thyroid and parathyroid lesions. Occasionally, ablation of a parathyroid adenoma by ultrasound-guided alcohol injection has been performed in patients unfit for surgery. Any mass in the neck or region of the salivary glands can be evaluated with ultrasound and its nature and organ of origin characterized.



Fig. 6. A parathyroid adenoma (closed arrow) situated posteriorly to the thyroid (open arrow).

Male genital tract

Ultrasound is highly sensitive in the detection of non-palpable testicular tumours and may therefore be used to exclude the presence of a tumor in patients with non-specific scrotal symptomatology. Tumour echo patterns vary and classification of tumour type is not possible (Fig. 7). Some benign testicular conditions such as orchitis may mimic tumours. If a tumour is detected, abdominal ultrasound should be performed to detect lymph node metastases, although actual staging is performed using CT.

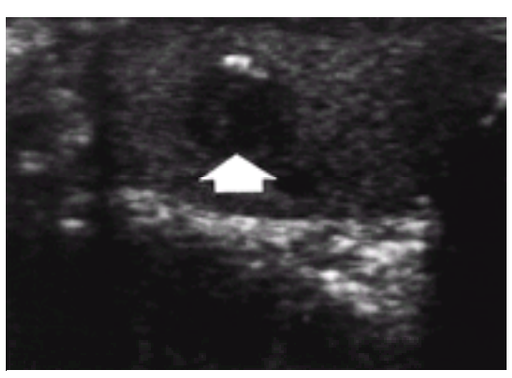


Fig. 7. A small impalpable testicular tumour (arrow) containing a small area of calcification.

Epididymo-orchitis may produce characteristic thickening of the epididymis which may aid in its differentiation from testicular torsion. In addition, the epididymis and/or the testis may show marked hyperaemia on colour flow Doppler in inflammatory conditions. In testicular torsion, the affected testis will usually show no flow or markedly diminished flow when compared with the normal side. However, there is a potential for false negative ultrasound examinations in that if detorsion has occurred the testis may appear relatively normally perfused and if venous rather than arterial occlusion has occurred, then again, an arterial signal may be detectable in the testis. Ultrasound therefore has a high positive predictive value of torsion with a somewhat lower negative predictive value.

Trauma to the testis, such as testicular rupture, and testicular and paratesticular haematoma are well demonstrated.

Ultrasound may also be used in the investigation of penile erectile dysfunction. It is possible to assess peak systolic velocity in the cavernosal arteries as well as diminished flow in diastole as indicators of the adequacy of the veno-occlusive mechanism.

Breast

Although there are those who support the use of ultrasound as a primary imaging technique, it is most often used to clarify the cystic or solid nature of a mass detected on X-ray mammography or in breasts that appear particularly dense on mammography. Ultrasound should be used initially in the assessment of inflammatory breast

disease as abscesses can be localized, and percutaneous drainage can be performed. It may also be used to evaluate breast masses in young patients.

Musculoskeletal system

One of the earliest applications of ultrasound in the evaluation of the musculoskeletal system, and still one of the most commonly performed examinations, is the examination of painful hips, particularly in children. This is an extremely simple and accurate technique for detecting fluid within the joint and, if appropriate, for guiding joint aspiration.

Dynamic examination of the hip in infancy using ultrasound allows detection of congenital dislocation at a time when the femoral head is not ossified and is therefore not visible on the plain radiograph. Other applications of ultrasound to the musculoskeletal system are the detection of non-radio-opaque foreign bodies, the evaluation of rotator cuff and Achilles tendon injuries, and the assessment of soft tissue masses and haematomas.

Ocular ultrasound

High frequency ultrasound is an ideal method for imaging the eye and intraocular structures. This is particularly so when the ocular media is opaque. Ultrasound may be used to evaluate the lens, vitreous detachment, retinal detachment and retrohyloid haemorrhage, choroid detachment, and intraocular tumours.

Vascular system (see also Section 17)

Duplex ultrasound imaging of the extracranial carotid arteries is now widely performed in patients who have experienced a transient ischaemic attack. It may be used to determine which patients go onto angiography and possible surgery. In experienced hands ultrasound has a 92 to 95 per cent accuracy as compared with conventional arteriography in the analysis of extracranial carotid artery disease. If the duplex examination demonstrates a completely normal vessel in a patient with a classic transient ischemic attack or stroke, the cause is unlikely to lie in the extracranial carotid artery. Duplex ultrasound may be used to evaluate the arterial and venous systems of the arms and legs, and in many centres it has replaced venography for the diagnosis of lower limb venous thrombosis. Thrombosis in the popliteal, femoral, and proximal external iliac veins is readily confirmed or excluded, although thrombosis confined to the deep veins of the calf is less easy to demonstrate. Imaging strategies for the detection of deep vein thrombosis are ideally combined with an assessment of pretest clinical probability.

The intra-abdominal vasculature, particularly the hepatic and portal venous systems and the renal vasculature, may be readily evaluated with duplex and colour flow Doppler ultrasound. The altered dynamics of the portal venous system in portal hypertension and obstruction to the hepatic veins are all well demonstrated. Doppler evaluation in possible renovascular hypertension has been disappointing, but the assessment of the renal venous system and the characterization of renal perfusion in renal transplants is more promising.

Not only does duplex ultrasound of the transplanted kidney give an indication of graft failure but it will also detect obstruction of the kidney and the presence of perirenal fluid collections such as lymphoceles and urinomas.

With very few exceptions, ultrasound demonstrates the presence or absence of abdominal aortic aneurysms and can be used for subsequent monitoring of their size. It is less reliable than angiography, CT, and MRI in establishing the relationship to the renal arteries and less reliable than CT and MRI in detecting a leaking aneurysm.

Ultrasound contrast agents

In recent years, ultrasound contrast agents for intravascular use have been developed. These agents are gas microbubble preparations which, because of the high ultrasound reflectivity of the microbubble, enhance the detection of blood flow with both spectral and colour flow Doppler ultrasound. In addition, used in conjunction with harmonic imaging, they may lead to improved tissue resolution with gray scale ultrasound.

As more of these agents come on the market, so the numbers of applications increase. They are widely used in echocardiography, vascular applications such as transcranial Doppler ultrasound, and their role in evaluating the neovascularity of tumours is currently being determined.

Further reading

Berman JM, Beidle TR, Kunberger LE, Letourneau JG. Sonographic evaluation of acute intra scrotal pathology. *American Journal of Roentgenology* 1996; **166**: 857–61. [A pictorial essay.]

Birnbaum BA, Brooke Jeffrey R. CT and sonographic evaluation of acute right lower quadrant abdominal pain. *American Journal of Roentgenology* 1998; **170**: 361–71. [A review of the relative merits of CT and ultrasound.]

Carlson KJ, Skates SJ, Singer DE. Screening for ovarian cancer. *Annals of Internal Medicine* 1994; **121**: 124–32.

Frates MC, Laing FC. Sonographic evaluation of ectopic pregnancy: an update. *American Journal of Roentgenology* 1995; **165**: 251–9.

Karlsson B *et al.* Transvaginal ultrasonography of the endometrium in women with postmenopausal bleeding—a Nordic multicentre study. *American Journal of Obstetrics and Gynecology* 1995; **172**: 1488–94.

Lindsell DRM. Ultrasound imaging of pancreas and biliary tract. *Lancet* 1990; **335**: 390–3.

Liu J-B, Goldberg BB. Endoluminal vascular and nonvascular sonography: past, present, and future. *American Journal of Roentgenology* 1995; **165**: 765–74. [A review.]

McNicholos MMJ, Brophy DP, Power WJ, Griggin JF. Ocular sonography. *American Journal of Roentgenology* 1994; **163**: 921–6. [A pictorial essay.]

Nanda NC, Schlief R, Goldberg BB, eds. *Advances in echo imaging using contrast enhancement*, 2nd edn. Kluwer Academic Publishers, Dordrecht, 1993.

Spencer J, Lindsell D, Mastorakou I. Ultrasonography compared with ntravenous urography in the investigation of adults with haematuria. *British Medical Journal* 1990; **301**: 1074–6.

Stein SM *et al.* Differentiation of benign and malignant adnexal masses: relative value of gray-scale, colour Doppler and spectral Doppler sonography. *American Journal of Roentgenology* 1995; **164**: 381–6.

Wells PS *et al.* Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet* 1997; **350**: 1795–8.

Zanetta G *et al.* Early and short-term complications after US-guided puncture of gynecologic lesions: evaluation after 1000 consecutive cases. *Radiology* 1993; **189**: 161–4.

Zerbey AL, Lee MJ, Brugge WR, Mueller PR. Endoscopic sonography of the upper gastrointestinal tract and pancreas. *American Journal of Roengenology* 1996; **166**: 45–50. [A pictorial essay.]

13.6 Imaging in children

David R. M. Lindsell

Techniques
Gastrointestinal obstruction in the newborn and young infant
Oesophageal atresia and tracheo-oesophageal fistula
Hypertrophic pyloric stenosis
Duodenal obstruction
Malrotation of the intestine
Small bowel obstruction
Meconium ileus
Intussusception
Large bowel obstruction
Hirschsprung's disease
Meconium plug syndrome
Anorectal anomalies
Necrotizing enterocolitis
Gastro-oesophageal reflux
Appendicitis
Acute testicular pain
Abdominal masses
Hydronephrosis
Renal cystic disease
Wilms' tumour
Neuroblastoma
Imaging in urinary tract infection
Biliary atresia
Meckel's diverticulum
Further reading

Techniques

The development of techniques such as ultrasound, computed tomography (**CT**), radionuclide imaging, and magnetic resonance imaging (**MRI**), has greatly enhanced our diagnostic capabilities in children. Many conditions still require conventional radiographic examinations and contrast studies to make a diagnosis. The chosen imaging modality will be the one that provides most diagnostic information in the shortest time while minimizing the discomfort and radiation dose to the child. In this latter respect, ultrasound and MRI have the advantage that they do not use ionizing radiation.

Contrast studies of the gastrointestinal tract are often performed with barium, but when aspiration or intestinal perforation are likely to be present the newer low osmolality, water-soluble contrast media are safer. The use of high osmolality contrast media are rarely indicated, except possibly in the treatment of meconium ileus. Many congenital abnormalities, such as diaphragmatic hernia, duodenal atresia, omphalocele, and congenital hydronephrosis can be diagnosed antenatally by ultrasound, and the obstetrician and paediatric surgeon forewarned. If appropriate, the mother can then be transferred prior to delivery to a paediatric surgical unit. It is beyond the scope of this section to consider antenatal diagnosis and paediatric surgical conditions occurring outside the abdomen.

Gastrointestinal obstruction in the newborn and young infant

Oesophageal atresia and tracheo-oesophageal fistula

The diagnosis of oesophageal atresia is usually suspected clinically and can be confirmed on a chest radiograph with a soft nasogastric tube passed into the oesophagus coiling back on itself. Further confirmation can be obtained by injecting a small quantity of air down this catheter, which will distend the proximal oesophageal pouch. Contrast medium should not be injected as this may lead to aspiration.

A tracheo-oesophageal fistula is present in 85 per cent of infants with oesophageal atresia, and an abdominal radiograph will show gas within the intestine: the abdomen is gasless in those without a fistula. Tracheo-oesophageal fistula can occur in isolation: contrast studies of the oesophagus may be necessary to demonstrate the fistula, which often runs upwards from the oesophagus to the trachea in the lower part of the neck, the so-called H-type fistula.

Imaging studies are also necessary to demonstrate coexisting abnormalities, which are present in about 50 per cent of patients. These include vertebral and renal abnormalities, duodenal and anal atresia, and congenital heart disease.

Hypertrophic pyloric stenosis

Since ultrasound uses non-ionizing radiation, it has now replaced barium studies in the diagnosis of pyloric stenosis in the small percentage of babies in whom the clinical diagnosis is in doubt. The interpretation of ultrasound is operator dependent, and if such expertise is not available a barium study should be performed.

Ultrasound clearly demonstrates the hypertrophied muscle, and the length and thickness of the pyloric muscle can be measured to confirm the diagnosis (Fig. 1(a)).

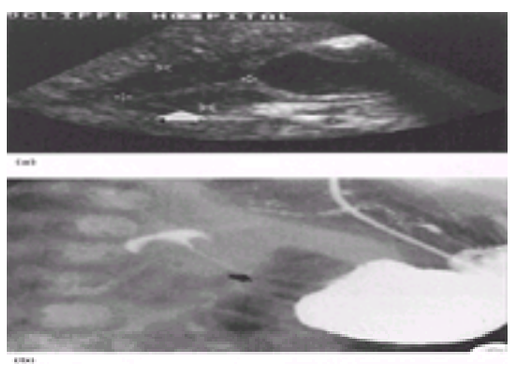


Fig. 1. (a) Ultrasound demonstrating the hypertrophied muscle (white arrow) of hypertrophic pyloric stenosis. (b) Barium meal demonstration of the elongated narrowed pyloric canal (black arrow) of hypertrophic pyloric stenosis.

Barium studies demonstrate a narrow elongated pyloric canal with indentations on the gastric antrum and duodenal bulb from the thickened muscle (Fig. 1(b)).

Duodenal obstruction

Duodenal obstruction may arise from a number of causes (Table 1); the radiographic appearances depend on whether or not the obstruction is complete. If it is complete, as in duodenal atresia, the classic ‘double bubble' appearance of air and fluid in the distended stomach and proximal duodenum is seen (Fig. 2). If the baby has recently vomited or if a nasogastric tube has been passed, the abdomen may appear almost gasless. In this situation the obstruction can be seen on plain radiographs after 15 to 20 ml of air are injected through the nasogastric tube. Plain radiographs may be virtually normal in patients with partial obstruction, and contrast studies may be necessary to demonstrate abnormalities such as a duodenal web or bowel malrotation.

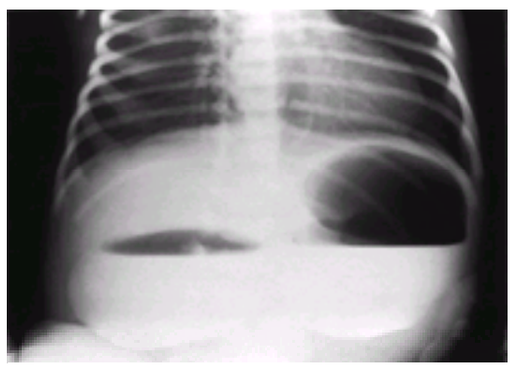


Fig. 2. Double bubble' appearance of two fluid levels in the dilated stomach and proximal duodenum of duodenal obstruction.

Duodenal atresia or stenosis
Duodenal obstruction associated with malrotation
Annular pancreas
Duodenal web
Preduodenal portal vein
Duodenal duplication cyst
Extrinsic compression by adjacent masses

Table 1 Causes of duodenal obstruction

Malrotation of the intestine

During the embryonic period the duodenojejunal loop rotates 270° around the superior mesenteric artery axis in an anticlockwise direction. The cecocolic loop, which initially lies inferiorly to the superior mesenteric artery, also rotates 270° in an anticlockwise direction. Finally, the caecum and ascending colon become fixed to the posterior peritoneum. If this process is interrupted at any point, malrotation or non-rotation results.

Intestinal obstruction, with the possibility of vascular compromise, is due to either an associated volvulus or extrinsic compression from peritoneal Ladd's bands. Abdominal radiographs may either suggest duodenal or small bowel obstruction. The diagnosis of malrotation is made by a contrast study of the upper gastrointestinal tract. This demonstrates the duodenojejunal flexure and ligament of Treitz to be in an abnormally low position or in the right side of the abdomen rather than in their normal position, which is in the left side of the abdomen on a level with the duodenal bulb. Although the position of the caecum is also abnormal, its position is variable in the neonate, and contrast studies of the upper rather than the lower gastrointestinal tract are therefore more reliable.

If a volvulus is present then the affected segment of bowel will appear twisted, whereas Ladd's bands give an extrinsic impression on the bowel outline. Rarely, malrotation presents less acutely, with cyclic abdominal pain and vomiting or malabsorption.

Small bowel obstruction

The main causes of small bowel obstruction are listed in Table 2.

Small bowel atresia or stenosis
Meconium ileus
Hernia (internal or external)
Volvulus (usually associated with malrotation)
Duplication cyst
Intussusception
Intestinal aganglionosis
Inspissated milk curd syndrome

Table 2 Causes of small bowel obstruction

Abdominal radiographs show dilated loops of small bowel with air–fluid levels on erect and decubitus views. It can be difficult to differentiate between small and large bowel obstruction in young infants, because the normal differentiating features of valvulae conniventes of the small bowel and haustra of the colon are not apparent in this age group. The anatomical position of the dilated loops of bowel may be helpful, but it may be necessary to perform a contrast study of the colon simply to define the level of obstruction. It is also important to differentiate between an ileus and obstruction. Infants often develop ileus secondary to septicaemia and metabolic disturbances.

Occasionally a ‘bubbly' appearance is present in the right side of the abdomen; this is due to meconium mixed with air and may be seen in meconium ileus, ileal atresia, or Hirschsprung's disease. If perforation occurs during the intrauterine period, the spilt meconium produces a sterile peritonitis, with the formation of peritoneal calcification or, less often, rim calcification in a ‘meconium pseudocyst'. Ascites may be present.

Meconium ileus

This condition, seen almost exclusively in babies with cystic fibrosis, causes small bowel obstruction due to thick tenacious meconium in the distal ileum. This prevents the normal bowel contents from entering the colon, which, as a result, is often a ‘microcolon'. Ileal atresia is usually associated with a colon of normal calibre, as it is frequently caused by a vascular accident late on in fetal life.

The abdominal radiograph shows dilated small bowel loops, but is often remarkable for the lack of fluid levels due to the very sticky meconium within the bowel. Fifty per cent of babies with meconium ileus have other abnormalities, such as intestinal atresia or volvulus, and this should be suspected if many fluid levels are present on the radiograph.

The diagnosis of meconium ileus is confirmed by a water-soluble contrast study of the large bowel with reflux of contrast into the distal ileum. This will demonstrate ‘balls' of inspissated meconium in the terminal ileum (Fig. 3).

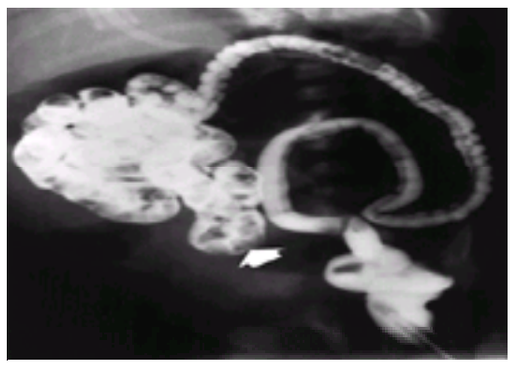


Fig. 3. Large bowel contrast study demonstrating the ‘microcolon' of meconium ileus and reflux of contrast around meconium ‘balls' (arrow) in the distal ileum.

Non-operative treatment of meconium ileus is often attempted. Hyperosmolar water-soluble contrast material is refluxed retrogradely amongst the impacted meconium and may relieve the obstruction both through its lubricating effect and because its osmotic effect pulls fluid into the bowel lumen. Reported success rates for this procedure vary but a rate of 50 per cent can be expected.

The procedure is, however, only appropriate in uncomplicated cases, which account for less than half the total number. The baby has to be very closely monitored as the hyperosmolar effect of the contrast media may lead to considerable fluid and electrolyte imbalance.

In older children with cystic fibrosis, complications include malabsorption and distal intestinal obstruction (meconium ileus equivalent), small bowel volvulus, intussusception, rectal prolapse, cirrhosis, diabetes mellitus, and acquired megacolon. A new entity has been described more recently in which colonic strictures cause large bowel obstruction. This process usually begins in the caecum and ascending colon, but may eventually progress to involve the whole colon. The cause for these strictures has not been definitely established. The relationship of treatment with large doses of pancreatic enzyme supplements to the development of colonic disease is strongly suspected but unproved.

Intussusception

In infants over the age of 3 months the most common causes of small bowel obstruction are intussusception and obstructed hernias. The appearances of intussusception on plain radiographs vary from a virtually normal radiograph to a picture of marked small bowel obstruction. A soft tissue mass of the intussusception itself will be visible in about one-half of the patients, or there will be an absence of the normal colonic gas pattern in the right side of the abdomen (Fig. 4). The diagnosis can be confirmed by either ultrasound or positive or negative contrast examination of the colon.



Fig. 4. Plain radiography showing the soft tissue mass (arrow) of an intussusception.

Once the diagnosis has been confirmed a decision has to be taken whether to attempt non-operative reduction with either barium or gas or whether to proceed directly to surgery. This requires close co-operation between the surgeon and the radiologist. The surgeon must carefully assess the child: signs of peritonitis and bowel ischaemia are contraindications to non-operative reduction, as is radiographic evidence of perforation.

Non-operative treatment has for many years relied on hydrostatic pressure from a barium or water-soluble contrast enema to reduce the intussusception. Recently, gas reduction has replaced barium as the method of choice because of higher success rates and because it is quicker and cleaner. (Fig. 5). Reported series suggest a successful reduction rate using barium of between 50 and 80 per cent, whereas most studies using gas report rather better figures of 75 to 95 per cent. Factors which suggest that the procedure is less likely to be successful are a history of longer than 24 h, the presence of small bowel obstruction, and an intussusception that has reached the distal colon. The incidence of perforation of the bowel during non-operative reduction is between 0.4 and 1.2 per cent. If gas is being used then perforation may lead to respiratory embarrassment due to gross pneumoperitoneum. The spillage of barium and fecal material into the peritoneum was thought to be associated with a very high morbidity: recent studies suggest that an immediate laparotomy reduces the long-term sequelae to a minimum. The rate of recurrence of the intussusception following non-operative reduction is similar to that following surgery (4 to 10 per cent).

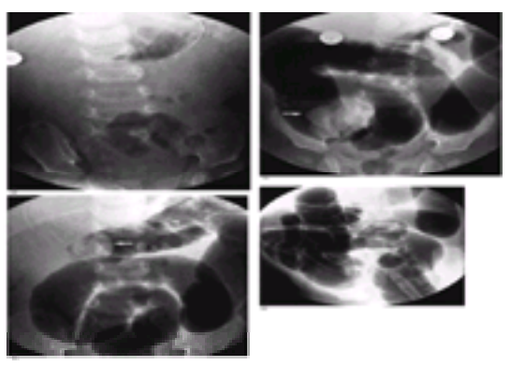


Fig. 5. Plain radiograph of infant with an intussusception shows a relatively gasless abdomen (a); rectal insufflation of air outlines the intussusception in mid-transverse colon (arrow) (b); this is reduced to the caecum (c); complete reduction is indicated by reflux of air into small bowel loops (d).

Large bowel obstruction

The causes of large bowel obstruction are listed in Table 3. As previously discussed the differentiation between large and small bowel obstruction may be difficult using plain radiographs alone. It may be necessary to perform a contrast study of the colon to define the level of obstruction as well as to differentiate obstruction from ileus.

Hirschsprung's disease
Anorectal atresia or stenosis
Meconium plug syndrome
Colonic atresia or stenosis
Extrinsic compression by cysts or tumours

Table 3 Causes of large bowel obstruction

Hirschsprung's disease

The majority of infants with Hirschsprung's disease present with large bowel obstruction in the first week of life. The remainder usually present in the first 3 years with constipation. The definitive diagnosis is made by rectal biopsy. Plain abdominal radiographs suggest a distal large bowel obstruction or, in the rare instance of total aganglionosis of the colon, small bowel obstruction. Contrast studies using low osmolality, water-soluble contrast will define a transition zone in 70 per cent of affected infants (Fig. 6). This transition zone demarcates the proximal normally innervated and dilated bowel from the distal abnormally innervated bowel, which is of normal or reduced caliber.



Fig. 6. Large bowel contrast study in Hirschsprung's disease demonstrating the transition zone (arrow) between distended but normally innervated proximal colon and the collapsed abnormally innervated distal colon and rectum.

Meconium plug syndrome

This condition occurs most commonly in premature infants and infants of diabetic mothers. The colon is normally innervated but plain radiographs suggest large bowel obstruction. Contrast studies demonstrate plugs of meconium obstructing the distal colon. The condition is probably due to functional immaturity of the colon and is self-limiting. The act of performing a contrast study will often dislodge the meconium.

Anorectal anomalies

These include ectopic anus, imperforate anus, rectal atresia, and anal or rectal stenosis. In these conditions, imaging defines the relationship of the atretic segment to the levator ani pelvic sling and the integrity of sphincters and their nerve supply. Both CT and MRI can be used for this purpose. Fistulas can be demonstrated by contrast studies of the bladder, urethra, and vagina, and associated malformations such as renal ectopia, hydronephrosis, spinal and cardiac anomalies, and oesophageal or duodenal atresia can also be visualized.

Prior to complex imaging of the pelvic musculature and renal and gastrointestinal tracts, plain radiographs will show a degree of large bowel obstruction. Spinal abnormalities will also be apparent and if a fistula is present, gas may be visible within the bladder (Fig. 7). A ‘cross-table' radiograph, with the baby prone and with its pelvis raised, is sometimes used to assess the distance between the most distal loop of bowel containing gas and the perineum. This may be misleading as the bowel proximal to the atresia may be plugged by meconium and not visible on the radiograph. This film should be delayed until the baby is about a day old to allow gas to reach the distal bowel. Ultrasound or MRI are now preferred to assess this gap.

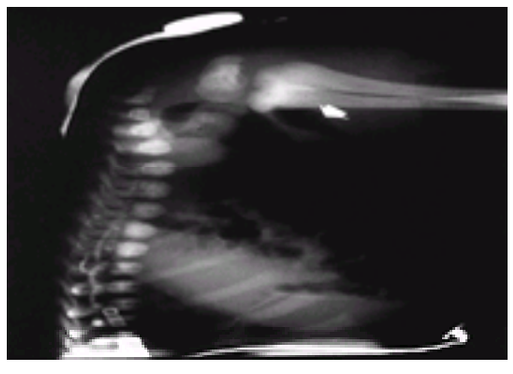


Fig. 7. Lateral invertogram' in a baby with imperforate anus demonstrating air within the bladder (arrow) due to a fistula from the bowel.

Necrotizing enterocolitis

The diagnosis of necrotizing enterocolitis is both a clinical and radiologic one, as radiographic abnormalities may not always be apparent. Radiographic abnormalities include intestinal ileus and distension, and bowel wall thickening. Intramural gas leads to lucencies in the bowel wall which may be linear, circular, or ‘bubbly' (Fig. 8). Air may pass from the bowel wall into the mesenteric venous system and become apparent in the portal vein. If perforation occurs, free air will be visible in the peritoneal cavity and abscesses may form. Free air is best seen on decubitus radiographs but the first signs, including visualization of both sides of the bowel wall, the falciform ligament, and umbilical artery remnants, may be detected on a supine film. Long-term sequelae are stricture formation and, rarely, enterocolonic fistulas. These can be demonstrated by contrast studies of the bowel.

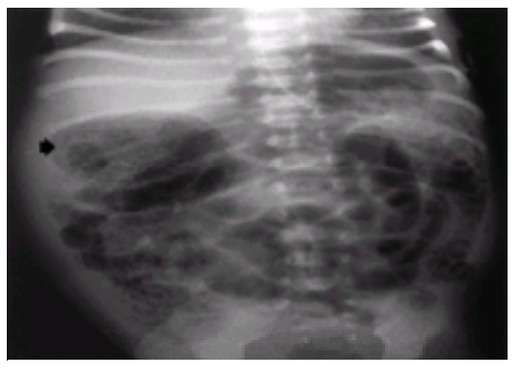


Fig. 8. Intramural air (arrow) in a severe case of necrotizing enterocolitis.

Gastro-oesophageal reflux

Babies and young infants with persistent vomiting or posseting, or failure to thrive in association with anemia, possibly due to severe reflux oesophagitis, need further investigation if conservative methods of treatment fail. These include 24-h pH monitoring in the distal oesophagus, upper gastrointestinal contrast studies, and radionuclide investigation for the detection of reflux as well as quantifying the rate of gastric emptying. As pH monitoring is more physiological in that the monitoring takes place over a 24-h period, this has to a great extent replaced the other techniques.

Appendicitis

The diagnosis of appendicitis is primarily a clinical one. Ultrasound using the graded compression technique has, however, been demonstrated in large series to have high specificity and accuracy and slightly lower sensitivity. Because of the operator dependence of ultrasound, caution needs to be exercised in extrapolating such results to all ultrasound departments.

Acute testicular pain

In adult males, the differentiation between acute epididymo-orchitis and testicular torsion has been improved by the use of colour flow Doppler ultrasound. However, even in young adults, false negative examinations occur due to either partial torsion or detorsion. In the young child these difficulties may be accentuated due to low blood flow rates within the testes. The advent of power Doppler has allowed the detection of lower flow rates than conventional colour flow Doppler ultrasound, but the younger the child the more caution there needs to be in interpreting the ultrasound findings.

Abdominal masses

In the immediate neonatal period most abdominal masses are renal in origin, due either to cystic disease or to hydronephrosis. Other masses occurring at birth are omental, mesenteric or duplication cysts, choledochal cysts, ovarian cysts, and adrenal haemorrhage. Tumours such as Wilms' tumour and neuroblastoma are more common around the age of 2 or 3 years; liver tumours, which are much less common, can occur at any age.

The role of imaging is primarily to define the organ of origin of the mass, to evaluate its nature and, if the mass is a tumour, whether there is evidence of spread. Ultrasound is the initial investigation of choice. This is often supplemented by CT or MRI where appropriate.

Hydronephrosis

Hydronephrosis may be due to obstruction, vesicoureteric reflux, or the prune belly syndrome. Obstruction most commonly occurs at the pelviureteric junction. Ultrasound determines whether hydronephrosis is present, whether it is unilateral or bilateral, whether the ureters are dilated, and whether the bladder and posterior urethra are normal. It gives no information about function. The intravenous urogram provides some functional information, but this is best assessed by radionuclide diuretic renography using technetium-99m either labelled to diethylene-triamine-penta-acetate (DTPA) or mercaptoacetyltriglycine (MAG 3) renography.

Renal cystic disease

There are many forms of renal cystic disease in infancy. The most common cause of an abdominal mass is multicystic dysplastic kidney, which must be differentiated from gross hydronephrosis as renal cysts and dilated calyces can be confused. Multicystic dysplastic kidney is usually unilateral, and a number of non-communicating cysts of varying size are seen. An anomaly of the contralateral kidney, such as pelviureteric obstruction, is occasionally seen. A technetium-99m labeled dimercaptosuccinic acid (**DMSA**) scintigram will demonstrate a non-functioning kidney. A micturating cystourethrogram should also be performed to assess whether vesicoureteric reflux is occurring on the contralateral side. Infantile polycystic kidney disease is almost invariably bilateral. The ultrasound appearances of large highly echogenic kidneys are characteristic, although the actual cysts are usually small and therefore not seen.

Wilms' tumour

Wilms' tumour is the most common urinary tract tumour of childhood and, together with neuroblastoma, it is the most common solid organ tumour outside the central nervous system. Initial imaging with ultrasound or urography demonstrates an intrarenal solid mass which may expand within the kidney or protrude from the surface (Fig. 9). Areas of cystic necrosis may be present. Calcification is seen in about 5 per cent of patients on plain radiographs and in 10 per cent of those examined with CT. Chest radiography, ultrasound, and CT are used for tumour staging. Close attention is paid to the presence of pulmonary or hepatic metastases, lymph node involvement, and involvement of the contralateral kidney, which occurs in 5 per cent of cases.



Fig. 9. Classic urogram of a Wilms' tumour showing a large intrarenal mass lesion.

Neuroblastoma

Neuroblastoma arises in the adrenal glands or anywhere along the sympathetic chain. The imaging protocol depends on the site of origin: about 70 per cent of these tumours occur within the abdomen and half of these originate in the adrenal gland. Within the abdomen, ultrasound demonstrates a solid extrarenal mass which may displace the kidney.

Calcification is seen in 50 per cent of plain radiographs and in up to 80 per cent of CT studies. MRI is also used to assess local spread and is the method of choice for evaluating intraspinal extension. (Fig. 10). Scintigraphy is required to detect skeletal metastases. Iodine-131 or iodine-123-*m*-iodobenzylguanidine (MIBG) may be used to detect the primary tumour and any metastases, as well as monitoring response to therapy.

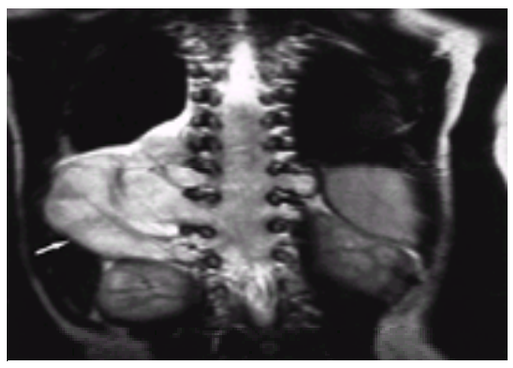


Fig. 10. Coronal, T_2-weighted, fast spin, echo MRI image showing a large right adrenal neuroblastoma (arrow) with extension into the spinal canal.

Imaging in urinary tract infection

Imaging studies in children with urinary tract infection allow the early detection of conditions such as hydronephrosis and vesicoureteric reflux, which may lead to irreversible renal damage. If damage has occurred, the severity can be assessed. Numerous imaging strategies have been proposed. As previously stated, hydronephrosis can be evaluated by ultrasound, intravenous urography, and radionuclide renography. Vesicoureteric reflux is assessed by a micturating cystourethrogram (Fig. 11). This may also demonstrate urethral abnormalities such as posterior urethral valves, ectopic ureters, and uretero-celes. Less commonly, reflux can be demonstrated by direct or indirect radionuclide cystography.

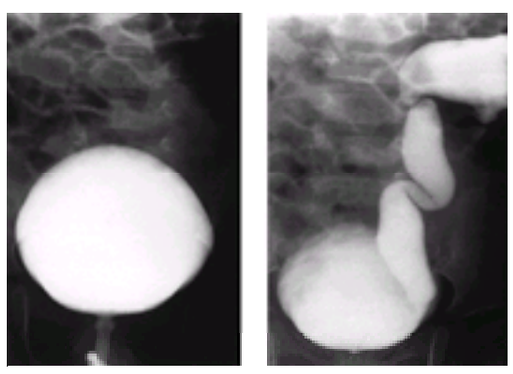


Fig. 11. Micturating cystourethrogram demonstrating vesicoureteric reflux.

Renal scarring is best demonstrated by technetium-99m labeled DMSA scintigraphy (Fig. 12). This also permits quantification of split renal function. Ultrasound is unreliable in the detection of mild or even moderate degrees of renal scarring.



Fig. 12. Technetium-99m DMSA scintigraphy (posterior view) demonstrating renal scarring. Split function: left = 64 per cent, right = 36 per cent.

In the young infant (under 1 year of age) an ultrasound scan to detect hydronephrosis and a micturating cystourethrogram to detect reflux are performed; DMSA scintigraphy is also often carried out. Between the ages of 1 and 5 years there is debate as to the most appropriate investigations. One approach is to perform an ultrasound scan and a DMSA scintigram initially, followed by a cystogram if either show scarring or a dilated urinary tract, suggestive of reflux. After the age of 5 years an ultrasound scan and a plain abdominal radiograph, to detect renal calcification and spinal anomalies, are probably all that is required. In all age groups, if an abnormality is detected, then the appropriate imaging protocol will be followed depending on whether obstruction, reflux, or scarring are suspected.

Biliary atresia

The diagnosis of biliary atresia must be made as soon after birth as possible so that surgery can be performed in the first 2 months of life. Ultrasound is performed initially to exclude other causes of biliary obstruction such as choledochal cyst, although the two may coexist. The gallbladder is not usually visualized in patients with biliary atresia, but examination of the liver is otherwise normal. Following ultrasound examination, biliary excretion is assessed using either [^{123}I]-bromosulphthalein or [^{99}Tc^{m}]-iminodi-acetic acid compounds. Reported sensitivity in differentiating atresia from hepatitis varies between 85 and 100 per cent. The final diagnosis and differentiation may depend on liver biopsy and operative cholangiography. The new technique of magnetic resonance cholangiopancreatography may be of value in assessing the biliary tree in the newborn child.

Meckel's diverticulum

This may occasionally be detected on plain radiographs as a fluid- or air-filled mass containing debris, or a calcified enterolith may be seen. Contrast studies of the small intestine may fill the diverticulum, but the standard method of detection relies on technetium scintigraphy. Provided that the diverticulum contains ectopic gastric mucosa then the scintigram should be positive.

Further reading

Aaronson IA, Cremin BJ. Cysts of the kidney. In: *Clinical paediatric uroradiology*, pp. 76–99. Churchill Livingstone, Edinburgh, 1984.

Aaronson IA, Cremin BJ. Tumours of the urinary tract. In: *Clinical paediatric uroradiology*, pp. 307–34. Churchill Livingstone, Edinburgh, 1984.

Bader TR, Kammerhuber F, Herneth AM. Testicular blood flow in boys as assessed at color Doppler and power Doppler sonography. *Radiology* 1997; **202**: 559–64.

Bell RS, Graham CB, Stevenson JK. Roentgenologic and clinical manifestations of neonatal necrotising enterocolitis. *American Journal of Roentgenology* 1971; **112**: 123–34.

Berdon WE, Slovis TL, Campbell JB, Baker DH, Haller JO. Neonatal small left colon syndrome: its relationship to aganglionosis and meconium plug syndrome. *Radiology* 1977; **125**: 457–62.

Bisset GS, Kirks DR. Intussusception in infants and children: diagnosis and therapy. *Radiology* 1988; **168**: 141–5.

Boechat MI *et al.* Primary liver tumours in children: comparison of CT and MR imaging. *Radiology* 1988; **169**: 727–32.

Bower RJ, Sieber WK, Kiesewetter WB. Alimentary tract duplications in children. *Annals of Surgery* 1978; **188**: 669–74.

David R *et al.* The many faces of neuroblastoma. *Radiographics* 1989; **9**: 859–82. [Review of imaging findings in neuroblastoma.]

Gans SL. Classification of ano-rectal anomalies: a critical analysis. *Journal of Pediatric Surgery* 1970; **5**: 511–13.

Gu L *et al.* Intussusception reduction in children by rectal insufflation of air. *American Journal of Roentgenology* 1988; **150**: 1345–8.

Hahn HB *et al.* Sonography of acute appendicitis in children: 7 years experience. *Pediatric Radiology* 1998; **28**: 147–51.

Hibi S, Todo S, Imashuku S, Miyazaki T. Iodine-131 meta-iodobenzyl guanidine scintigraphy in patients with neuroblastoma. *Pediatric Radiology* 1987; **17**: 308–13.

Hope JW, Borns PF, Berg PK. Roentgenologic manifestations of Hirschsprung's disease in infancy. *American Journal of Roentgenology* 1965; **95**: 217–29.

Houston CS, Wittenborg MH. Roentgen evaluation of anomalies of rotation and fixation of the bowel in children. *Radiology* 1965; **84**: 1–17.

Kluth D. Atlas of oesophageal atresia. *Journal of Pediatric Surgery* 1976; **11**: 901–19. [Anatomic classification of types of esophageal atresia.]

Kohda E, Fujioka M, Ikawa H, Yokoyama J. Congenital anorectal anomaly: CT evaluation. *Radiology* 1985; **157**: 349–52.

Leonidas JC, Berdon WE, Baker DH, Santulli TV. Meconium ileus and its complications, a reappraisal of plain film roentgen diagnostic criteria. *American Journal of Roentgenology* 1970; **108**: 598–609.

Luker GD, Siegel MJ. Color Doppler sonography of the scrotum in children. *American Journal of Roentgenology* 1994; **163**: 649–55. [A pictorial essay.]

Martin GI, Kutner FR, Moser L. Diagnosis of Meckel's diverticulum by radio isotope scanning. *Pediatrics* 1976; **57**: 11–12.

Nadel HR. Hepatobiliary scintigraphy in children. *Seminars in Nuclear Medicine* 1996; **26**: 25–42.

Nussbaum AR, Sanders RS, Hartman DS, Dudgeon DL, Parmley TH. Neonatal ovarian cysts: sonographic–pathologic correlation. *Radiology* 1988; **168**: 817–21.

O'Donovan AN, Habra G, Somers S, Malone DE, Rees A, Winthrop AL. Diagnosis of Hirschsprung's disease. *American Journal of Roentgenology* 1966; **167**: 517–20. [Findings in Hirschsprung's disease on contrast enema.]

Rescorla FJ, Grosfeld JL, West KJ, Vane DW. Changing patterns of treatment and survival in neonates with meconium ileus. *Archives of Surgery* 1989; **124**: 837–40.

Ros PR, Olmstead WW, Moser RP, Dachman AH, Hjermstad BH, Sobin LH. Mesenteric and omental cysts: histologic classification with imaging correlation. *Radiology* 1987; **164**: 327–32.

Royal College of Physicians, working group of the research unit. Guidelines for the management of acute urinary tract infection in childhood. *Journal of the Royal College of Physicians* 1991; **25**: 36–42.

Sato Y *et al.* Congenital anorectal anomalies: MR imaging. *Radiology* 1988; **168**: 157–62.

Snyder WH, Chaffin L. Malrotation of the intestine. In: Mustard WT, Ravitch MM, Snyder WH, Welch KJ, Benson CD, eds, *Pediatric surgery*, pp. 808–17. Year Book Publishers, Chicago, 1969.

Westra SJ, de Groot CJ, Smits NJ, Staalman CR. Hypertrophic pyloric stenosis: use of the pyloric volume measurement in early US diagnosis. *Radiology* 1989; **172**: 615–19.

13.7 Nuclear medicine techniques

Basil J. Shepstone

Basic concepts
Radiopharmaceuticals
Mapping the distribution of radiopharmaceuticals
Imaging disease processes
The skeleton
The urogenital tract
The lungs
Liver and biliary system
The spleen, bone marrow, and lymphatic system
Gastrointestinal tract
The thyroid and parathyroid
Adrenals
Cardiovascular system
The central nervous system
Dacryocystoscintigraphy
Tumour imaging
Infection/inflammation-seeking tracers
Nuclear pathology
Safety of nuclear medical studies
Further reading

Basic concepts

Nuclear medicine uses the properties of radioactive and stable nuclides to diagnose morphological and/or physiological disorders, and to provide therapy using unsealed radioactive sources.

One great advantage that nuclear imaging techniques have over other imaging methods is that they represent function rather than morphology. Another is that, although ionizing radiation is used, adverse reactions are rare. A third is that many of the techniques demonstrate exquisite sensitivity in detecting disease, but this is accompanied by its great disadvantage in that radionuclide studies are seldom specific, that is, with certain notable exceptions, pathognomonic appearances are rare.

It is well known that radioactive substances can emit a- or b-particles, or else g-rays. a-Particles are emitted only by the naturally occurring radionuclides such as radium, which are no longer used in medicine. b-Particles, because of their short path length before they are absorbed, are useful for therapy but useless for diagnosis as they cannot emerge from the body. This leaves only g-rays, which, like X-rays, can pass out of the body and be detected externally.

The currently used unit of radioactivity is the becquerel (Bq), which is the Systéme Internationale (SI) unit and equal to one disintegration per second (dis/s). As this is a very small amount, the commonly used multiple is the megabequerel (MBq), where 1 MBq = 10^6 dis/s. The older unit, the curie (Ci) = 3.7×10^{10} dis/s, so 1 millicurie (mCi) = 37 MBq.

Radiopharmaceuticals

The basic radiopharmaceutical is a g-emitting radionuclide, which is attached to a compound that is specially tailored to target a particular organ system or physiological process. The quantities used are generally small and so neither disturb the process under investigation nor provoke hypersensitivity reactions.

The most popular radionuclide used currently is technetium-99m (where 99 is the mass number and ‘m' means ‘metastable'—it can also be written $^{99}Tc^m$). Like all metastable radionuclides, technetium-99m has the triple advantages of being a pure g-emitter, having a short half-life, and being readily available from a longer-lived generator containing its parent radionuclide. In fact it is now used so widely that g-ray detectors are specifically designed to function optimally at the g-ray energy of technetium-99m.

Technetium-99m originates from the generator as the pertechnetate $^{99}Tc^m0_{4'}$, but it can, for example, be attached to human serum albumin macroaggregates, which will blockade the pulmonary capillaries and can be used to detect pulmonary emboli. It can be linked to phosphates or phosphonates, which undergo chemiabsorption on to bone crystal, in order to image bone. Reticuloendothelial uptake in liver, spleen, and bone marrow is effected by linking it to sulphur or tin colloid, which is phagocytosed. In the form of the pertechnetate it can be taken up by active ion transport into gastric mucosa, thyroid, and salivary glands. Hepatocyte and biliary-tract imaging uses the active cellular transport and excretion of technetium-99m-labelled, substituted carbamoyliminodiacetic acid (**HIDA**). It can also be chelated with, say, dimercaptosuccinic acid (**DMSA**), which will be taken up by the renal cortical cells to produce renal images. Chelated with diphenyltriaminepentaacetic acid (**DTPA**) it can measure glomerular filtration rate. The most exciting, recent compound linked to technetium-99m is hydroxymethylpropylamine oxime (**HMPAO**, Ceretec, Exametazime; Amersham), which crosses the blood–brain barrier and can be used, for example, to investigate neuropsychiatric disorders such as Alzheimer's dementia and schizophrenia.

Other radionuclides are, of course, also used in nuclear medicine. Examples are thallium-201 chloride for myocardial imaging and gallium-67 citrate for the detection of tumours and infective/inflammatory foci. Indium-111 has been used to label white cells to detect infection, monoclonal antibodies to detect targets as diverse as tumours or areas of endometriosis, and somatostatin-receptor ligands to detect lesions containing those receptors. Iodine-123 has replaced one of the original radionuclides used in nuclear medicine, iodine-131, in the diagnosis of thyroid disorders.

All of the above radionuclides are called single-photon emitters because they emit single g-ray photons. However, there is also a class of so-called positron-emitting radionuclides that emit two photons, a positive electron or positron and a conventional negative electron, in coincidence at an angle of 180° to each other. These charged subatomic species cannot exist as such for any length of time and are annihilated by their opposite number, yielding two high-energy g-rays. Among the positron-emitting radionuclides are important biological atoms such as oxygen-15, nitrogen-13, carbon-11, and fluorine-18, which can be incorporated into a variety of tracers. Fluorine-18, for example, has been incorporated into fluorodeoxyglucose (**FDG**), which has revolutionized neurophysiology and neuropathology.

Unfortunately, these potentially most useful radiopharmaceuticals are cyclotron-produced and also have extremely short half-lives of the order of a few minutes. Therefore the in-house equipment and processes needed to produce these include a cyclotron, on-line radiochemical processing, purification and sterilization, and a positron camera—an exorbitantly expensive facility.

It is the ability of radiopharmaceuticals to act as indicators of physiological processes that separates nuclear medicine from other imaging processes such as computed tomography (**CT**) and ultrasonography, which are largely conveyors of morphological information. However, exceptions to this statement do exist in general radiology. The intravenous urogram, for example is a functional as well as a morphological study.

Mapping the distribution of radiopharmaceuticals

The distribution within the body of single photon-emitting radionuclides (i.e., g-emitters) is mapped by means of a scintillation or ‘gamma' camera. (Fig. 1). A modern camera can either accumulate flat (‘planar') images or rotate around the body to reconstruct, via the computer, tomographic images in the transverse, sagittal, or coronal planes.



Fig. 1. Triple-headed camera in use for brain scintigraphy (by courtesy of James King-Holmes).

The basic component of such cameras is a thallium-activated sodium iodide crystal, which can convert g-rays into light photons. The photons are then converted into electrons and subsequently aggregated into amplified, filtered pulses. The pulses, suitably distributed spatially, can then produce maps of the distribution of radioactivity on either X-ray or Polaroid film, or on to light-sensitive paper. Otherwise, pulses may be stored in the memory of a computer for later manipulation.

In order to screen out unwanted g-rays, both from the patient and from other sources of radiation in the imaging suite, or from background radiation, collimators are used. These usually consist of a lead slab bearing many small holes and act like a sieve, keeping out g-rays from any direction bar those from the organ of interest. Further sorting of unwanted energies is done electronically.

The data can therefore be in either digital or analogue form, but the two are interconvertible. So-called regions of interest can, for example, be drawn around areas of physiological study (e.g. the renal cortex, the left ventricle, the oesophagus) and count-rate changes within these regions can be derived as activity–time curves. In this way, for example, renograms can be produced to demonstrate the passage of a radiopharmaceutical such as technetium-99m DTPA through the kidney.

Positron detectors detect the pair of g-rays resulting from the positron–electron annihilation; in order to do this they are placed on either side of the source, usually in a ring formation, and only register a count if both detectors detect photons in coincidence. Image-control techniques are then used to reconstruct the cross-sectional image. Single photon-emission tomography (**SPET**) uses the rotating gamma camera to detect photons from multiple angles (180° or 360°) around the patient and an associated computer to reconstruct multiple slices. Technetium-99m and iodine-123 are the common radionuclides used for SPET studies; the principal organs of interest are the brain and the myocardium.

More recently, a variant on the PET theme has been made possible by combining the increased availability of FDG and double-headed, conventional gamma cameras provided with high-energy collimation and coincidence-counting electronics. This technique is now known as FDG SPET.

The field is advancing by leaps and bounds, with new radiopharmaceuticals being announced almost monthly. In this chapter, however, largely established techniques will be discussed, most of which could be applied in any district general hospital, with only brief reference to PET and immunoscintigraphic applications.

Imaging disease processes

The skeleton

Blood flow and turnover are the major factors that determine the uptake of radiopharmaceuticals into bone. In the resting state, about one-third of the maximum potential blood flow through bone is excluded by sympathetic tone and so processes that alter this tone also affect blood flow. Chemiabsorption on to recently deposited or exposed bone crystal is the principal method of radiopharmaceutical uptake.

Normal bone will therefore be active (‘warm' rather than ‘hot'), with the larger bones such as the pelvis and the joints predominating. Pathology shows up as photon-abundant or ‘hot' areas, owing to increased blood flow or bone turnover, or loss of sympathetic tone. However, a purely lytic area such as an osteoclastoma or an infarct can be photon-deficient (‘cold'), and the extraosseous uptake of bone-seeking radionuclides is well described, for example in paretic states, neuropathies, areas of necrosis, and infarction.

Technetium-99m-complexed, organic and inorganic phosphate compounds are the major bone-seeking tracers, to which hydroxyl groups are often added to improve crystal binding. The currently favoured compounds are the diphosphonates. The skeleton takes up about 60–per cent of the injected dose, the rest being excreted through the kidneys and at the same time providing renal scintigrams, which are crude assessments of renal function. Skeletal scintigraphy is usually done in three phases: detecting flow, equilibrium uptake, and the so-called late crystal phases.

As with most radionuclide studies, skeletal scintigraphy is exquisitely sensitive (10^{14} times as sensitive as plain skeletal radiography), but non-specific. Its main use in modern nuclear medicine is the detection of metastases, and requests for skeletal scintigraphy nowadays constitute nearly half of the routine workload of most general nuclear medicine departments (Fig. 2).

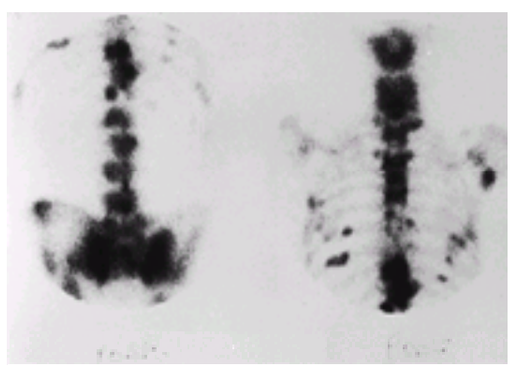


Fig. 2. Part of a skeletal scintigram obtained using technetium-99m diphosphonate. The very black areas are foci of increased uptake due to increased flow to, and osteoblast stimulation around, skeletal metastases from a primary breast cancer.

In the radiological detection of neoplasia, it has been shown that 20 to 30 per cent of bone metastases are seen on skeletal scintigraphy, but not on radiographs. However, in about 5 per cent or less of cases, the reverse is true and it is possible that processes infiltrating the bone marrow (e.g. myeloma and leukaemia) will be missed on skeletal scintigraphy.

A variant of the appearance of metastatic disease in the skeleton is the so-called superscan, when the whole skeleton shows relatively uniform high uptake with little in the way of discrete foci. It might even be reported as normal until it is noted that no renal images are present, an indication of the fact that nearly all the dose has been taken up by rapidly metabolizing bone and very little is left to excrete through the kidneys (Fig. 3). This pattern may, however, also be seen in metabolic bone disease.

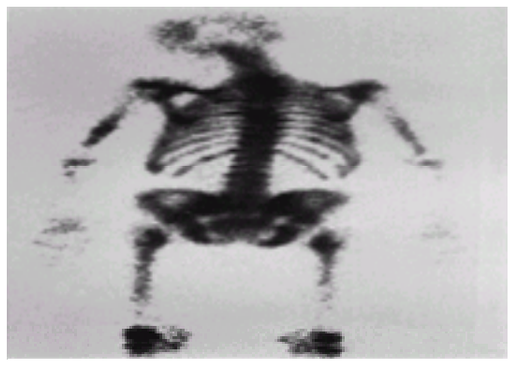


Fig. 3. A skeletal scintigram using technetium-99m diphosphonate showing generalized, increased bone uptake of tracer, but with no sign of renal uptake. This constitutes the ‘superscan' as a result of diffuse metastatic spread from a bronchial carcinoma.

Benign bone tumours show varying degrees of uptake. That in an osteoid osteoma may be almost pathognomonic, with a central, very hot nidus surrounded by an oval of slightly lesser intensity (Fig. 4). Most other bone neoplasia shows varying degrees of uptake and is usually investigated by plain radiography, CT, or magnetic resonance imaging (**MRI**). However, multifocality can be easily demonstrated on scintigraphy.

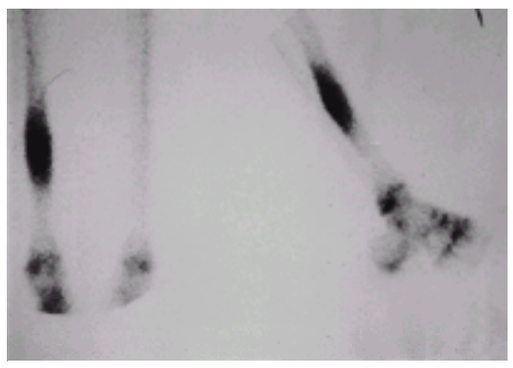


Fig. 4. A skeletal scintigram using technetium-99m hydroxymethylenediphosphonate showing the characteristic high uptake in an osteoid osteoma, with its oval configuration and very ‘hot' nidus.

As far as trauma is concerned, 80 per cent of all fractures show increased uptake by 24 h and nearly all are visible by 3 days. Injury often removes the sympathetic drive, resulting in generally increased, but diffuse, uptake—the reflex sympathetic dystrophy syndrome. Failure to see increased flow at a fracture site a few weeks after injury suggests a lack of bone union. On skeletal radiography, rib fractures usually present as focal areas of increased uptake arranged in a vertical line, as opposed to metastases, which are randomly located. In many parts of the skeleton (for example wrist, ankle, face, base of skull), fractures not seen on conventional radiology may easily be detected on scintigraphy; this includes stress fractures of the tibia.

Avascular necrosis of the femoral head or the scaphoid may be detected as areas of photon deficiency during both the flow and the crystal phases. In hip fractures, account must be taken of the fact that a photon-deficient head can be due to tamponade of the artery by surrounding oedema. To distinguish between necrosis and oedema a scintigram using technetium-99m sulphur or tin colloid, which is phagocytosed by the marrow reticuloendothelial cells, will demonstrate whether marrow has survived or not. It will survive in generalized oedema, but not in avascular necrosis.

Skeletal scintigraphy can detect osteomyelitis both on the flow study, which demonstrates increased blood flow to the affected area, and on the delayed study, owing to bone turnover. Radiographic changes may not occur before 1 to 2 weeks. The technique is also useful in distinguishing between osteomyelitis and cellulitis, as cellulitis produces only increased flow while osteomyelitis shows both increased flow and increased bone turnover. Bone injured in any way can produce misleading images, as reactive bone shows a high affinity for radionuclides for up to 9 months after a fracture. The problem is at its most frustrating in prosthetic hip replacement, which can become loose or infected, or both. It may be solved by using an infection-seeking agent such as gallium-67 citrate or indium-111-labelled white cells, but the issue may be extremely difficult to resolve.

Radionuclides such as technetium-99m hydroxymethylenediphosphonate (**HMDP**) can also be used to investigate metabolic bone disease such as Paget's, renal osteodystrophy, primary hypoparathyroidism, osteoporosis, and a number of other rarer conditions. These can easily be detected, quantified, and followed up using serial scintigraphy or bone absorptiometry.

The skeletal scintigraphic agents, notably technetium-99m pyrophosphate, are often taken up in soft tissue. The best known example is uptake in areas of myocardial or cerebral infarction or ischaemia, while it is also quite common in necrotic tumours and a host of other conditions varying from scar tissue to the lungs of patients in chronic renal failure.

Skeletal scintigraphy can also be used for the detection and follow-up of joint disease, e.g. rheumatoid arthritis, where there is a strong correlation with clinical symptoms.

The urogenital tract

Kidney

Reasonable renal images may be obtained using technetium-99m DMSA, should a patient be allergic to iodine-containing contrast media and so unable to undergo intravenous urography. All renal masses show up as photon-deficient areas. A renal column of Bertin, which may appear as a mass on urography, shows photon-abundance on scintigraphy. The principal use of the technique, however, is for the detection of renal scarring, especially in children, which is notoriously difficult or impossible with intravenous contrast. By measuring the uptake of labelled DMSA by each kidney, an assessment of relative renal function can also be made; renal scintigraphy has also been found to be sensitive in the detection of renal trauma.

A second useful tracer technique for investigating the kidneys is renography. This is usually performed with technetium-99m DTPA or technetium-99m mercaptoacetyltriglycine (**MAG3**). The normal renogram (Fig. 5) is a computer-generated activity–time curve depicting the passage of radionuclide through the kidney and is obtained by defining a region of interest over the kidneys. Its analysis is complex, utilizing so-called deconvolutional analysis and deriving transit times, but it seems to be just as useful if it is interpreted simply. The rising part of the curve (phase I or ‘vascular phase') is a function of radionuclide reaching the kidney; the peak (phase II or ‘handling phase') is related to the efficiency of parenchymal function; the descending part of the curve (phase III or ‘excretory phase') shows the rate of wash-out of tracer from the renal area. Much of the activity ‘seen' by the gamma camera will be background and ideally this should be subtracted from the renal curves.

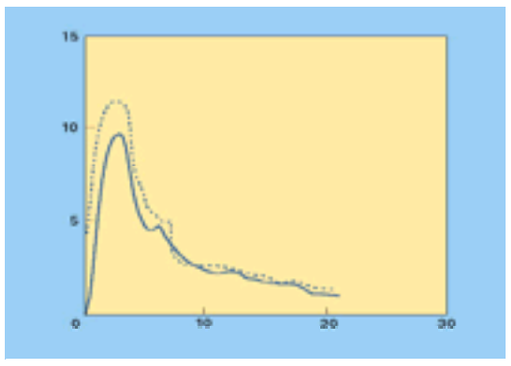


Fig. 5. Normal gamma-camera renogram of left (solid curve) and right (dotted curve) kidneys using technetium-99m MAG_3. Each curve shows a rising vascular phase, sharp handling peak of good amplitude, and a rapidly descending excretory phase (*y*-axis, percentage of injected dose; *x*-axis, time in minutes after injection).

In parenchymal failure there is a normal vascular phase, but the handling peak is of poor count rate, delayed and broadened. One of the most difficult diagnoses to make without invasive contrast angiography is that of renovascular hypertension, but renography following the administration of angiotensin-converting enzyme inhibitors is now reported as increasing the accuracy of the test (Fig. 6).

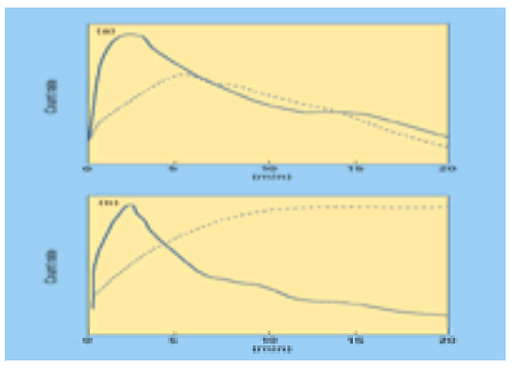


Fig. 6. (a) Technetium-99m MAG_3 renogram in a hypertensive man with right-sided renal arterial stenosis before angioplasty. Abnormal right side (dotted line) and normal left (continuous line). (b) During the administration of an angiotensin-converting enzyme inhibitor (captopril) the left renogram is unchanged, but the right has altered dramatically.

In obstructions one might expect a renogram similar to that shown in Fig. 7(a), where the vascular and handling phases are normal but the excretory phase fails to descend normally and may even continue to rise. However, a similar curve is obtained when there is hydronephrosis without current obstruction. In this instance the tracer simply pours into the dilated pelvicalyceal system. The curve will eventually descend, but not within the half-hour or so normally assigned to a renogram study. There is, however, a very simple way of distinguishing between this ‘floppy-bag' scenario and true obstruction, and that is to perform a diuretic-provocation renogram. There will be prompt clearance of tracer from the pelvis in the cases of non-obstructive hydronephrosis, as shown in Fig. 7(b).

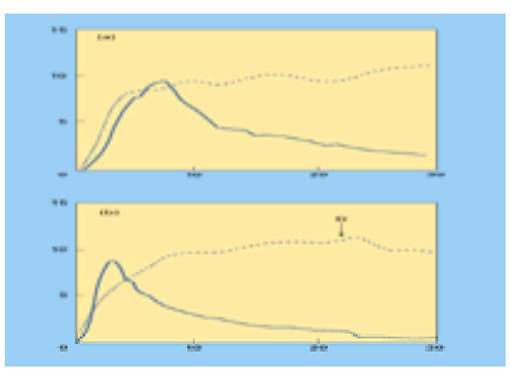


Fig. 7. (a) Renogram showing obstructive pattern on right side. (b) Gamma-camera renogram with diuretic provocation showing normal left renogram, but the right renogram showing little response to diuretic (D). Graphical conventions as in Fig. 5 (the normal renograms).

Radionuclide methods for detecting vesicoureteric reflux can be direct or indirect; they are excellent for follow-up studies as the radiation dose to the patient is much lower than for voiding cystourethrography. The direct test is similar to its radiographic counterpart in that technetium-99m is instilled into the bladder via a catheter and images recorded as the patient voids (Fig. 8(a)). The indirect study can be done as part of a conventional renogram, with the patient being asked to void near the end of the excretory phase. If reflux is present, the curve will show a sharp rise (Fig. 8(b)).

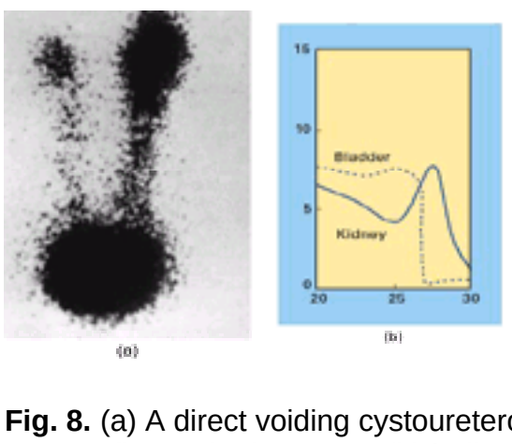


Fig. 8. (a) A direct voiding cystoureterogram using technetium-99m as pertechnetate in saline instilled into the bladder via a catheter while the patient voids; there is gross reflux up both ureters. (b) Indirect voiding cystoscintigraphy. Towards the end of the descending phase of a renogram done with technetium-99m DTPA, this child was asked to void. The activity curve over the bladder drops precipitously, but at the same time the renogram rises to reflect increased activity over the kidney due to reflux.

Serial renograms can be very useful in monitoring renal grafts after transplantation. Apart from being able to detect early rejection 1 to 2 days before biochemical abnormalities become obvious, the method is useful for detecting acute tubular necrosis (in which the renogram steadily rises), injury to renal vessels, ureteral obstruction, and extravasation. Technetium-99m sulphur colloid is also useful in identifying graft rejection, as macrophages accumulating in the failing transplant take

up this agent to the same extent as the liver and spleen.

Kidneys with active pyelonephritis take up gallium-67 citrate for much longer (about 72 h) than normal kidneys (24 h).

Any of the above studies may be coupled with measurements of total glomerular filtration rate and effective renal plasma flow using chromium-51 elthylenediaminetriacetic acid clearance and iodine-123 orthohippurate or technetium-99m MAG_3, respectively.

Testicular imaging and penile blood flow

Radionuclide methods can also be used to investigate acute testicular pain, which can be due to torsion or acute epididimytis. In torsion a cut-off in the flow of the testicular artery is accompanied by decreased vascularity of the testis and increased activity of the surrounding scrotal structures. Epididimytis shows only a low-intensity, semicircular focus or even a small, focal area.

In male impotence involving erectile dysfunction it is possible to measure penile blood flow and venous leakage quantitatively using technetium-99m labelled red cells.

Radionuclide hysterosalpingoscintigraphy

By asking the patient to insert a small syringe containing a low dose of technetium-99m pertechnetate into her vagina like a tampon and imaging her pelvis an hour or so after she has pressed the plunger and removed it, tubal patency can be tested. Tracer is seen in the area of the ovaries if the fallopian tubes are patent.

The lungs

Pulmonary ventilation–perfusion scintigraphy

Pulmonary perfusion imaging for the detection of pulmonary embolism is based on the principle of capillary blockade when labelled particles of human serum albumin, about 20 to 50 µm in diameter, are injected intravenously. These are mixed in the blood after passage through the right side of the heart and are then trapped in the pulmonary capillary bed, with a distribution reflecting pulmonary arterial blood flow. The most popular radiopharmaceutical for the purpose is technetium-99m macroaggregated human serum albumin. The albumin particles are broken down by alveolar macrophages after about 3 to 12 h.

Where there is decreased pulmonary blood flow, for example distal to the site of the embolus, there is therefore a photon-deficient region (Fig. 9). However, many other pathological processes also decrease arterial perfusion, but as they invariably also affect regional ventilation as well, perfusion studies cannot be correctly interpreted without a concomitant ventilation scintigram.

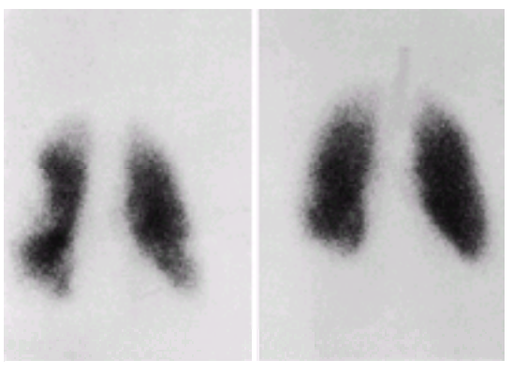


Fig. 9. Perfusion–ventilation scintigrams using technetium-99m macroaggregated albumin and krypton-81m gas, respectively. The labelled macroaggregates lodge in the pulmonary capillary bed. The pair of scintigrams shows an unmatched perfusion defect typical of a pulmonary embolus in the right mid-zone.

Ventilation scintigraphy can be documented using inert gases such as xenon-133 or krypton-81m, or an aerosol such as technetium-99m DTPA delivered through a nebulizer. A more recent agent is technetium-99m pseudogas, which is an ultrafine, mainly monodispersed, aerosol. It is generated when a spray of technetium-99m pertechnetate solution in ethanol is burnt. Krypton-81m gas is the safest and most popular agent, and is simply inhaled in air by the patient standing in front of or lying under the gamma camera. It is derived from a rubidium-81 generator.

The sensitivity of pulmonary perfusion imaging approaches 100 per cent in detecting emboli in a given patient, but when compared with pulmonary angiography, both tests miss some emboli and specificity is low. The following is a simplistic guide to interpretation using perfusion and ventilation scintigraphy and the contemporaneous chest radiograph.

1. *Normal* If pulmonary perfusion is normal, then it can be assumed that both ventilation and chest radiograph willbe normal.
2. *Unmatched perfusion defect/s* Abnormal perfusion with a photon-deficient defect, normal ventilation and a chest radiograph that is either normal or shows oligaemia and vessel cut-off in the affected area; this is diagnostic of early, reversible pulmonary embolus.
3. *Matched perfusion defect(s)* Where a perfusion defect(s) correspond exactly with the ventilation defect(s) the patient will either have known obstructive airways disease (e.g. emphysema or asthma), with the hypoxia causing vessel close-down and so a perfusion defect, or else have opacities on the chest radiograph corresponding to the matched defect(s). The diagnosis is then that of the opacities, for example neoplasia, collapse, infection, oedema, effusion, fibrosis, haemorrhage, and, of course, established infarction. It is therefore possible that a matched defect means a patient has had an embolus, albeit a little late for early anticoagulation.

The above schema is, in fact, very difficult to operate in practice, owing to continuing controversy as to what constitutes a significant perfusion defect. The result is that opinions are usually give as probabilities. In the case of normal perfusion, there need be no further evaluation for pulmonary embolism. Where there is clinical suspicion of emboli, but with a low probability of embolism on scintigraphy, no further evaluation is needed and also no treatment unless there is radiographic evidence of deep venous thromboses. If the probability of emboli is high, anticoagulation is imperative, but angiography is unnecessary. In the intermediate probability range, management depends on whether the clinical suspicion of pulmonary embolism is high or low. If it is high, either venography or angiography are necessary; anticoagulation is necessary if either is positive. However, if the clinical suspicion is low, further angiography or venography is unnecessary and the appearances are usually due to a different disease.

Spiral CT, which can image the actual embolus in its artery, may well replace perfusion–ventilation scintigraphy in the future, but it brings its own problems of high radiation dose, being time-consuming, and of being regarded by many workers as ‘non-physiological'.

Liver and biliary system

The liver–spleen scintigram, using technetium-99m sulphur or tin colloid, was once one of the most frequently used procedures in nuclear medicine. Its decline has not been due to a lack of accuracy, but to the ability of ultrasonography and cross-sectional imaging techniques such as CT and MRI to characterize lesions and their surroundings to a greatly enhanced extent.

However, one existing technique worth describing is that of hepatic arterial perfusion scintigraphy, which is used to map hepatic metastases and involves the delivery of technetium-99m macroaggregated albumin (as described above for use in pulmonary perfusion scintigraphy) through a catheter inserted into the common hepatic artery, just distal to the origin of the gastroduodenal artery. The technique is derived from that used to deliver chemotherapeutic agents. In this way the hepatic arterial perfusion scintigram ‘locks in' the pattern of chemotherapeutic delivery for later imaging. This ensures delivery to the entire burden of liver tumour and identifies inadvertent delivery to the gut or to the systemic circulation via arteriovenous shunting.

Technetium-99m-labelled red blood cells have become valuable for the characterization of known liver masses. The method relies on the fact that most cavernous haemangiomas have decreased perfusion but increased blood-pool activity. This state of affairs is mimicked only be the rare angioscarcoma. The combined used of

this radiopharmaceutical and SPET imaging can detect haemangiomas as small as 1 cm. Advances in SPET technology such as high-resolution and dynamic SPET have improved this figure to less than 1 cm.

The standard liver–spleen scintigram, which relies on the uptake of technetium-99m-labelled colloid by the reticuloendothelial cells of the spleen and their counterpart, the Kupffer cells, in the liver, still serves as a quick and easy procedure for the establishment of hepatosplenomegaly and is highly sensitive in assessing diffuse liver disease. The patterns in alcoholic liver disease (small liver, left-lobe hypertrophy, shift of colloid uptake to an enlarged spleen) and in isolated, hepatic venous obstruction (Budd–Chiari syndrome with its ‘hot' caudate lobe) are particularly well recognized.

In general, the combined use of technetium-99m sulphur colloid imaging, gallium-67 citrate imaging, hepatobiliary imaging (see below), and radiolabelled red-cell imaging allow the histological characterization of mass lesions not yet realized by CT and MRI. These combination methods are very useful in the characterization of focal nodular hyperplasia, hepatic adenoma, focal fatty replacement, cavernous haemangioma, hepatocellular carcinoma, and macroregenerating nodules.

The introduction of technetium-99m HIDA compounds has permitted a non-invasive method of investigating hepatobiliary disorders. These so-called bifunctional agents are extracted from the blood by the hepatocytes and concentrated in bile. In this way one obtains an estimate of hepatocyte function and an image of the biliary excretion pathway. A great advantage of the method is that it can be used when bilirubin is as high as 30 mg per cent.

The technique should be the first choice in the diagnosis of acute cholecystitis where the cystic duct is obstructed. A normal study demonstrates (in temporal sequence over about an hour) the liver, common bile duct, gallbladder, duodenum, and jejunum. If the gallbladder is not seen within 30 min of the intravenous injection of technetium-99m HIDA, acute cholecystitis is present (always provided the patient has a gallbladder!) (Fig. 10). Concomitant ultrasonography is useful, whereby thickening of the gallbladder wall, a pericholecystic collection, or a stone in the cystic duct can be found. A similar technique can be used in jaundiced neonates to distinguish between biliary atresia and neonatal hepatitis. In general, however, jaundice is always investigated by ultrasound in the first instance.

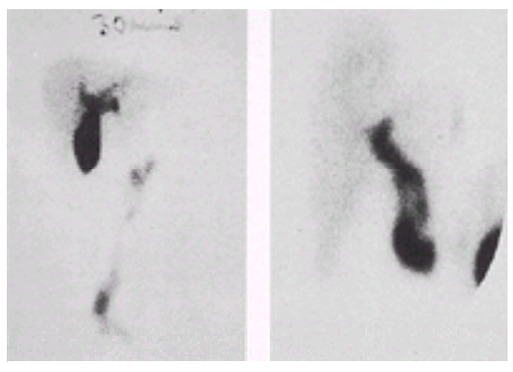


Fig. 10. Technetium-99m HIDA cholecystoscintigraphy in a normal person 30 min postinjection (left) and in a patient with acute cholecystitis (right) 60 min postinjection. The gallbladder is clearly visible early in the study in the normal situation but not even at 1 h in cholecystitis.

The role of the technique in chronic cholecystitis is limited, but in the postoperative patient it is useful in assessing biliary leaks, functional patency of the biliary system, and for detecting cystic-duct remnants. It is also the only reliable way of demonstrating enterogastric reflux quantitatively under physiological conditions.

The spleen, bone marrow, and lymphatic system

As stated above, technetium-99m-labelled sulphur, tin, or antimony trisulphide colloid, which is trapped by the reticuloendothelial cells, can be used to image the liver, spleen, bone marrow, and lymphatic system.

The spleen may be imaged without concomitant hepatic activity by injecting autologous red cells denatured by keeping them in a water bath at 40° for about 1 h and labelled with technetium-99m. The technique is usually reserved for searching for splenuniculi.

The distribution of bone marrow (or at least its reticuloendothelial elements, which usually coincide with the haemopoietic elements) varies with age as the functioning marrow withdraws from the extremities into regions said to be ‘covered by an Edwardian bathing costume'. Any deviation from this pattern, say in myelofibrosis (marked decrease in marrow elements) or in the chronic haemolytic anaemias (extension of the marrow space), can easily be detected. The method is also useful for monitoring bone-marrow transplants and for deciding on optimal sites for marrow aspiration.

Injection of technetium-99m antimony trisulphide between the webs of the toes will enable the extent and integrity of the lower-limb lymphatics to be evaluated. It may therefore be used to assess patients with swollen limbs not thought to be due to venous disease. Similar techniques can be used to evaluate lymphoedema in the arms and to plot the position of the internal mammary chain before irradiation of the chest wall for appropriate, medially situated breast cancers.

The detection of the so-called sentinel node is ‘hot news' in nuclear medicine, but at present is more popular with nuclear medical physicians than with either breast radiologists or surgeons. The technique is based on the concept that tumours will initially drain to one or more, preferred (‘sentinel') nodes, which subsequently drain to a larger group of second-echelon nodes. The implication of identifying the sentinel node is that its detection might obviate the need for axillary dissection. The success of the technique was first established in the case of malignant melanoma and has now spread to the investigation of axillary involvement in breast cancer.

One of the major problems with lymph-node scintigraphy is that the anatomy of nodal distribution and the physiology of colloidal uptake are both capricious. Uptake depends on phagocytic ingestion of a radiolabelled colloid and so it is impossible to know whether non-visualization of an expected node indicates its absence or whether metastatic spread has obliterated its phagocytic population. Conversely, a visible node may not be an metastatically involved node. Be this as it may, many workers have now claimed that, in at least 85 per cent of cases of breast cancer, tracer injected into the breast close to the lesion goes from the tumour to the axillary nodes, and that in all cases the sentinel node is imaged. In a recent controlled series, Veronesi and his colleagues in Italy have accurately predicted axillary lymph-node status in 97.5 per cent of women in whom they found a sentinel node. Similar studies have been undertaken using vital blue dyes, but the false-negative rate is reportedly as high as 20 per cent, whereas with scintigraphy it is as low as 2 to 3 per cent. Several g-probes are now on the market for the detection of these nodes intraoperatively.

In particular, as far as the detection of the sentinel node is concerned, the fact that even the best series fall short of 100 per cent identification means that no surgeon will rely entirely on lymphoscintigraphy. The conclusion at present is that, while the development of distant metastases and patient survival seem to correlate well with positive lymphoscintigraphy, the method cannot be relied upon entirely in clinical practice.

Much research interest has therefore been focused on lymphoscintigraphy using tumour-seeking agents, for example, radiolabelled monoclonal antibodies raised against tumour antigens. Most clinical studies have been on colon cancers and melanoma, but in 1984 Thompson *et al.* used radiolabelled monoclonal antibodies raised against human milk-fat globulin (HMFG-1 and -2): for the detection of axillary metastases the sensitivity (80–86 per cent) and specificity (88–92 per cent) are much better than with routine colloid lymphoscintigraphy. So far the usefulness of these agents has been difficult to assess, owing to a combination of cost, complexity, attainability, and many experimental and case variables.

Many workers, but especially those in Italy, feel that the evaluation of axillary nodal involvement might be the primary indication for FDG PET, which appears to have a negative predictive value of nearly 100 per cent; this translates into sparing about one-third of patients from axillary lymph-node dissection at the cost of missing involved nodes in one patient. The advantage of the technique is that hardly any false-negative nodes have been discovered.

Elective lymph-node dissection in patients with primary melanoma without evidence of clinically involved nodes remains a problem because about 20 per cent of such patients have micrometastases that could later lead to distant metastases and uncontrollable disease. However, looking at this from the opposite point of view, nodal dissection is unnecessary in the majority 80 per cent. As is the case with the breast, imaging of the appropriate sentinel node could potentially solve this dilemma. Using several intracutaneous injections of technetium-99m nanocolloid around the tumour site, the specific drainage pathway of the tumour can be mapped out (Fig. 11) and the sentinel node imaged. Again, the node can be determined during surgery using a portable g-ray probe.

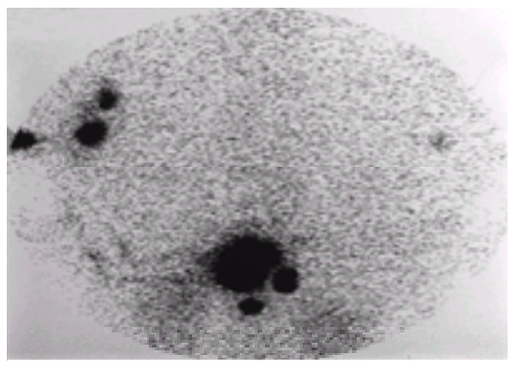


Fig. 11. Posterior view of axillary nodes imaged 3 h after intracutaneous injections of technetium-99m nanocolloid around a melanoma on the right lower back. The first-level node on the left was the sentinel node; this particular lymphatic pathway was unexpected (the focus on the far left is a radioactive marker).

Sentinel nodes can be identified in more than 95 per cent of cases, but discrepancies have been reported between the number of nodes suggested by scintigraphy and the number actually found at surgery. Once again, a reason for missing the sentinel node by scintigraphy could be its complete occupation by tumour and the consequent diversion of lymphatic drainage to other nodes. Nodes may also lie close to the injection sites or may be superimposed. In melanoma particularly, lymphatic plexuses seem common and produce false-positive foci. The head and neck are particularly difficult areas. As with breast cancer, much work still needs to be done and the technique is currently the subject of multicentre trials. Scintigraphy of sentinel nodes in penile and vulval cancers is also being investigated.

Gastrointestinal tract

The liver and hepatobiliary system have already been discussed.

The salivary glands can be imaged after intravenous injection of technetium-99m as the pertechnetate, when the normal parotid and submaxillary glands become visible; the sublingual glands are too small to be seen. The activity can be discharged with a sialogogue. There is no uptake from or discharge into the four glands in true xerostomia or Sjögren's syndrome, whereas these are normal in cases of psychological origin. The method may also be helpful in the diagnosis of Warthin's tumour (papillary cystadenoma lymphomatosum; ‘adenolymphoma'), in which there is no uptake or excretion of tracer.

Structural disease of the oesophagus is correctly investigated by endoscopy and a barium meal, but oesophageal function may be investigated with radionuclides. The principal indications are either gastro-oesophageal pain, dysphagia, or one of the neuromuscular disorders. The two main radionuclide tests for the investigation of these conditions are the measurement of oesophageal transit time and the physiological oesophageal-reflux test. Reflux in babies can be determined by having their milk laced with technetium-99m sulphur colloid or DTPA and taking late images over the chest to look for aspiration of tracer.

Abnormalities of gastric emptying are common after surgery for peptic ulceration; they include dumping syndrome, diarrhoea, and gastric stasis. Early dumping is due to rapid gastric emptying associated with a fall in plasma volume. Late dumping is due to reactive hypoglycaemia resulting from rapid emptying.

Gastric emptying studies using liquid and solid meals labelled with technetium-99m, indium-111 or indium-113m DTPA can assess such problems. They are not indicated in all postoperative patients, but dumping or stasis must be confirmed before any further surgery is contemplated.

In acute gastrointestinal haemorrhage, the angiogram is the conventional method for localizing the site of bleeding, but radionuclide angiography has superior sensitivity for detecting this bleeding and can be a valuable guide to selective abdominal arteriography. The two tracers commonly used for this purpose are technetium-99m-labelled sulphur colloid or red cells. Labelled red cells are better than colloid for detecting intermittent bleeding or bleeding near the liver and spleen. They can also be used for serial imaging over 24 h because they obviously continue to circulate. One disadvantage of the technique is that target:non-target ratios are low, owing to the high circulating background of labelled cells (Fig. 12). Another is that the blood leaking into bowel may be displaced both distally and proximally, and this may result in false identification of the bleeding site. Finally, free pertechnetate can concentrate in gastric or colonic secretions. Radionuclide angiography by this method is reported to detect bleeding rates of 0.1 ml/min, which compares favourably with a rate of only 0.5 ml/min for contrast angiography.

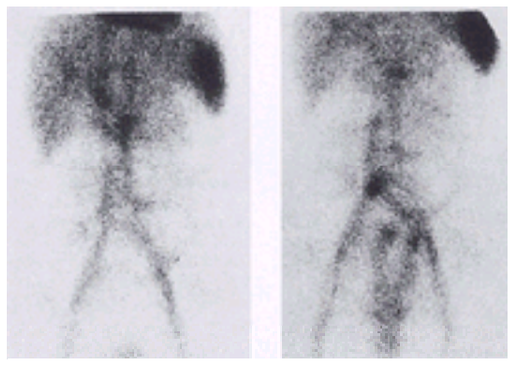


Fig. 12. A search for lower gastrointestinal bleeding using technetium-99m-labelled red cells. There is gradual accumulation of tracer in the distal sigmoid colon and rectum, but the background of circulating labelled red cells can make discrimination difficult. (Left) early; (right) late.

The rationale for the use of radiocolloid to detect bleeding is that intravascular colloid is rapidly cleared by the hepatic and splenic phagocytes, leaving visible any colloid extravasating from a bleeding site. This method is useful for active bleeding in the lower gastrointestinal tract and in monitoring the efficacy of therapy. The use of colloid yields a higher target:non-target ratio than that of labelled red cells, as the background is rapidly cleared of activity. A disadvantage is that bleeding sites near the liver or spleen are difficult to detect.

Rectal bleeding in children may be due to bleeding from ectopic gastric mucosa in a Meckel's diverticulum. Such mucosa, as is the case with the stomach, will take up pertechnetate with an accuracy of between 85 and 95 per cent.

The thyroid and parathyroid

Thyroid

It is well known that the thyroid traps inorganic plasma iodide, which is later organified to form the thyroid hormones triiodothyronine (T_3) and thyroxine (T_4).
Iodine-uptake studies using iodine-131 as sodium iodide were among the very first nuclear medical procedures to be carried out in the clinical setting, shortly after this fission product of uranium-235 became available after the Second World War. Today, the radionuclide of choice is the much safer iodine-123, with iodine-131 used mainly for therapy of toxic nodules as a result of its high output of b-particles.

A number of functional thyroid studies can be performed using iodine-123: uptake tests to distinguish between hypo-, eu- and hyperthyroidism; thyroid-stimulating hormone or thyrotropin-releasing hormone stimulation and repression tests to distinguish between disorders of the thyroid–pituitary–hypothalamic axis and the gland itself; and the perchlorate wash-out test to evaluate whether organification of iodide is normal.

Functional morphology can be assessed by iodine-123 or technetium-99m as pertechnetate (which is trapped, but obviously not organified, by the thyroid). Ectopic thyroid tissue, which may occur in the neck anywhere from the back of the tongue to behind the sternum, can be mapped. Thyroglossal cysts are usually detected clinically, but scintigraphy is employed to ensure that such clinically detected masses do not contain thyroid tissue. Even the normal thyroid image may show

anatomical deviations, with one lobe larger than the other, or, rarely, just a solitary, dominant lobe.

Pyramidal lobes are also seen from time to time, but their visualization is common in the glands of patients with Graves' disease. One should also never underestimate the amount of regrowth of thyroid tissue that can take place after partial, or even so-called complete, thyroidectomy.

Abnormal thyroid images can demonstrate focal or diffuse uptake of tracer, which can, in turn, be photon-abundant (‘hot') or photon-deficient (‘cold'). The division of thyroid nodules into hot or cold is important, as 15 to 20 per cent of cold nodules may be malignant whereas hot nodules are nearly always benign and usually toxic. Unfortunately, there is no way of distinguishing further between the various causes of cold nodules (benign adenomas, carcinomas, metastases, cysts, haematomas, focal thyroiditis, abscesses, or combinations of these) using radionuclide methods, but ultrasound can, of course, distinguish between cystic and solid lesions, and the advent of fine-needle aspiration has in any case enabled a preoperative diagnosis of thyroid nodules to be made. The rarer medullary carcinoma of the thyroid, which produces hypercalcitonaemia, may be detected with technetium-99m in its pentavalent state coupled to DMSA.

Toxic goitres include the hyperplastic, hyperfunctioning gland of Grave's disease. In contrast to this diffuse variety of goitre, there is the uninodular or multinodular toxic goitre, in which the overactive areas suppress normal tissue—the so-called hot nodule. Acute suppurative thyroiditis is rare; subacute thyroiditis is a painful but self-limiting problem that is probably of viral origin. It produces generalized, decreased uptake. Hashimoto's thyroiditis is, by contrast, chronic and painless, and probably due to lymphocytic infiltration causing an organification defect. Early images in this condition show an enlarged gland with normal or enhanced trapping, but later images show poor, uneven uptake.

In cases of hypothyroidism one may not even have sufficient counts to produce an image, but such a ‘non-image' can provide diagnostic information just as useful as in any other, more dramatic imaging situation.

Euthyroid goitres simply show enlargement with normal uptake.

Parathyroids

It is estimated that a surgeon seeking the source of excess parathormone will find it by exploration most of the time but is less than 75 per cent successful on re-exploration, and in that instance nearly all the imaging modalities have to be employed.

The thallium-201/technetium-99m pertechnetate subtraction technique is arguably the most useful imaging test. Thallium-201 is a potassium analogue and, aside from its well-known property of being taken up in the myocardium, is also localized in parathyroid adenomas and a variety of other neoplasms. Normal parathyroid glands are not visible. The problem is that thallium-201 chloride is also taken up in a normal thyroid and so one has to subtract a thyroid image taken after an injection of pertechnetate electronically from one taken previously after an injection of thallium-201 in the same position. The technique is more complex than this simple description would lead one to believe, but it is claimed that about 80 per cent of parathyroid adenomas can be localized by this method. It is unsuccessful in secondary hyperparathyroidism.

As will be noted in the section on myocardial imaging, the radiolabelled isonitrile, technetium-99m 2-hexakis-2-methoxyiso-butylisonitrile (‘sestamibi'), has largely replaced thallium-201 chloride. This is so with parathyroid adenoma imaging as well, the subtraction duo in this case being technetium-99m sestamibi/iodine-123 as sodium iodide. Also, a late 24-h image is taken because sestamibi will persist in an adenoma but not in normal thyroid or parathyroid. Admittedly, the signal from technetium-99m is very faint by this time, but it will just suffice, especially if SPET is used (Fig. 13).

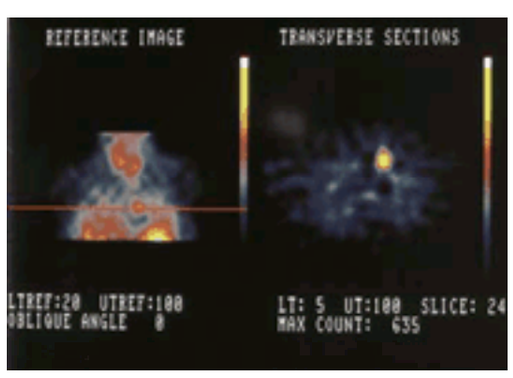


Fig. 13. Planar image (left) and a SPET cross-section (right) using technetium-99m sestamibi, showing a mediastinal parathyroid adenoma.

Adrenals

Medulla

Intra- and extra-adrenal phaeochromocytoma may be imaged using the radiolabelled noradrenaline (norepinephrine) and guanethidine analogue, iodine-131- or -123-labelled metaiodobenzylguanidine (**MIBG**), which is taken up at sites of catecholamine storage granules. This radiopharmaceutical has also proved useful in localizing other tumours of neural-crest origin such as paragangliomas, neuroblastomas, carcinoid, and medullary carcinomas. It has also been used therapeutically in these conditions. The use of radiolabelled somatostatin-receptor ligands in the detection of these tumours will be described later.

Cortex

When used with appropriate suppressive manoeuvres such as the administration of dexamethasone, uptake of the cholesterol analogue, iodine-131- or selenium-75-labelled norcholesterol, can supply evidence of adrenocortical hyperfunction. Adrenal cortical imaging is particularly appropriate to the localization of resectable neoplasia or in the identification of a suspected adrenal remnant following surgery. Symmetrically increased uptake denotes adrenocorticotrophic hormone (**ACTH**)-dependent hyperplasia, while asymmetrical uptake indicates ACTH-independent hypercorticalism. A unilateral image usually represents a neoplasm, commonly an adenoma. Adrenal carcinoma may destroy normal adrenal tissue, with a resultant absence of an image.

The method is particularly useful in distinguishing between ACTH-induced adrenal hyperplasia and primary adrenal tumours that produce hypercorticalism. Results are only slightly less reliable in detecting Conn's tumours producing hyperaldosteronism. The technique can also be used to distinguish between adrenal and ovarian causes of hirsutism.

Cardiovascular system

Modern nuclear medical techniques provide valuable information about cardiac structure, pathology, perfusion, and function. Four principal radionuclide techniques are available: (i) radionuclide ventriculoscintigraphy and the evaluation of ventricular function; (ii) radionuclide or ‘first-pass' angioscintigraphy; (iii) myocardial perfusion scintigraphy; and (iv) infarct-avid scintigraphy.

Radionuclide ventriculoscintigraphy

The cardiac blood pool may be imaged with technetium-99m-labelled red cells. Images of the ventricles (usually the left) in systole and diastole are obtained by gating the gamma camera to the patient's electrocardiographic signal (notably the R wave) and summating the counts obtained over many cardiac cycles.

The most important functional measure of the left (or right) ventricle that can be calculated is the ejection fraction, which is identified as:

$$\text{Ejection fraction} = (\text{end-diastolic counts} - \text{end-systolic counts}) / \text{end-diastolic counts.}$$

In spite of the slight errors inherent in the method, the left ventricular ejection fraction is a sensitive indicator of dysfunction. The value is normally greater than 50 per cent. Patients who have had recent infarction show decreases in their ejection fraction that correlate well with the volume of infarcted tissue and so the measure can serve as a guide to prognosis. It is similarly an excellent indicator of therapeutic effectiveness and can be assessed at rest or during exercise.

The method can also be used to generate contours of the ventricular blood pool in end-diastole and end-systole that can be manipulated to assess wall-motion abnormalities such as hypo-, a-, and dyskinesia. This can also be accomplished by producing phase and amplitude maps from Fourier analysis of the activity–time curves of the summated counts over the several cardiac cycles. Right ventricular scintigraphy is also possible, but less accurate, as there is no one view of the heart (compared with the left anterior oblique view of the left ventricle) on which the right ventricle is completed circumscribed.

Estimates of regurgitant fractions in patients with aortic and mitral valve disease and ventricular volumes can also be derived from radionuclide ventriculography, and the method can be useful in the timing of valve-replacement surgery.

Radionuclide angiography

This is also known as ‘first-pass' angioscintigraphy: it records an injection of technetium-99m as pertechnetate or labelled autologous red cells as it passes through the superior vena cava, the right atrium and ventricle, the lungs, the left atrium and ventricle, and, finally, the aorta. There must be no chamber dilatation, delay in transit, or recirculation (Fig. 14). The method can be used to detect any gross anatomical aberration in patients with congenital heart disease, e.g. right-to-left shunts. By computer analysis of activity–time curves over the lungs, left-to-right shunts can actually be quantified down to those as small as 10 per cent. In this way, if shunts have closed or whether corrective surgery has been successful can be detected.

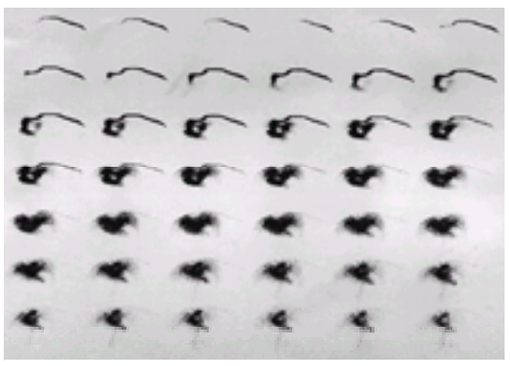


Fig. 14. A first-pass cardiac study using technetium-99m-labelled red cells. The tracer enters the heart by way of the superior vena cava and the right atrium and ventricle, thence via the pulmonary outflow tract to the lungs. The left ventricle and the aorta are then seen before mixing takes place.

First-pass scintigraphy can also be used for screening large, obstructed vascular lesions in the superior vena cava or the aorta, and is occasionally helpful in peripheral vascular obstruction.

Myocardial perfusion scintigraphy

Myocardial cells treat thallium-201 chloride in the same way as they do potassium and so regional myocardial uptake of thallium-201 reflects intact myocardial perfusion to viable myocardium. Regions of ischaemia, acute infarction, or scarring appear as focal decreased uptake. To distinguish between these entities it is essential to derive images after stress and then later at rest; an ischaemic area will ‘fill in' whereas a scar will not.

Myocardial perfusion imaging can be used to differentiate ischaemic from idiopathic cardiomyopathy, evaluate collateral coronary vessels or angioplasty, and assess coronary bypass grafts.

Because of its poor energy levels, thallium-201 is being replaced by the physically more favourable technetium-99m-labelled isonitriles such as sestamibi. The rationale of the study remains the same.

Infarct-avid scintigraphy

This method, which can be used to detect both myocardial and cerebral infarcts between 12 h and 10 days after the event, depends on the property of ischaemic, infarcted, or necrotic tissue to take up certain bone-seeking tracers such as technetium-99m pyrophosphate. A more recent advance embracing the same principle is the use of indium-111-labelled Fab fragments of antimyosin, which will bind to areas of myosin released in acute infarction.

Thrombus detection

While contrast venography provides the ‘gold standard' for detecting deep venous thrombosis in the lower limb, it is not without its contraindications. Iodine-125 fibrinogen could detect clots in the calf only on direct counting and could not be imaged. Technetium-99m fibrinogen could produce a scintigram but has been withdrawn from the market, probably because of the risk of contamination by human immunodeficiency virus. More recently, indium-111-activated platelets have been used for the detection of emboli, thrombi, and atherosclerotic lesions with varying degrees of success. Other, more recent possibilities include radiolabelled a-fibrin peptide, antagonists of the glycoprotein IIB/IIIA receptor, and a technetium-99m-labelled buckminsterfullerene particle (‘Technegas') that binds to a fibrin peptide.

The central nervous system

MRI is now unquestionably the method of choice in delineating cerebral lesions. Conventional cerebral scintigraphy using tracers such as technetium-99m as pertechnetate or glucoheptonate, which only pass through the blood–brain barrier when a lesion is present, is rarely undertaken nowadays but may still be useful in detecting early cerebral infections and in the differential diagnosis of cerebral infarction.

Cerebral radionuclide angiography may, however, still be a very useful, rapid, and non-invasive way of investigating the cerebral circulation. Any vascular lesion, such as an arteriovenous malformation, meningioma or glioma, will be seen as a focal area of increased activity, while photon-deficient areas are seen in cerebrovascular accidents, subdural haematomas, and in massive carotid occlusions. A characteristic sequence in patients who have recently had strokes is the so-called flip-flop phenomenon in which there is diminished perfusion on the side of the brain involved during the arterial phase, but which changes to increased activity during the later venous phase (Fig. 15). Jugular reflux, which occurs in the superior mediastinal syndrome, also produces a characteristic diagnostic picture.

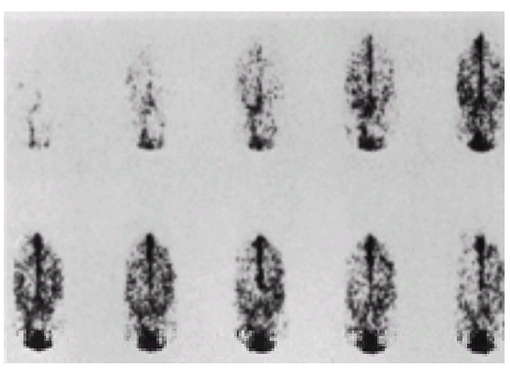


Fig. 15. Radionuclide angiogram using technetium-99m as pertechnetate. Delayed appearance of tracer in left hemisphere (reader's right) due to cerebrovascular

disease in the left carotid and/or left middle cerebral artery. During the later venous phase, activity in the right hemisphere decreases to reveal relatively increased flow in the left hemisphere—the so-called flip-flop phenomenon.

Emission–CT imaging of the brain

This very important branch of nuclear medicine has as yet unfortunately little part to play in neurosurgery, but it is of undoubted importance in the investigation of neuropsychiatric disorders such as the dementias, schizophrenias, the epilepsies, and even the depressive illnesses—none of which has hitherto been amenable to radiological diagnosis. Of all of these conditions, the epilepsies are arguably of the greatest interest to the neurosurgeon because it is possible to resect the sources of seizure disorders if they can be identified anatomically (Fig. 16). Epileptogenic foci show areas of decreased perfusion and metabolism interictally, but become areas of hyperperfusion during a seizure.

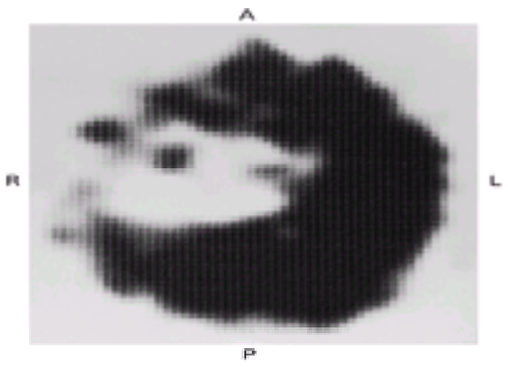


Fig. 16. A section through the brain of an 8-year-old girl with partial-seizure disorder after technetium-99m HMPAO injection, showing a photon-deficient area in the right frontal zone.

The principal agents used for these studies are fluorine-18 deoxyglucose and oxygen-15 in PET studies and technetium-99m HMPAO in SPET studies. HMPAO can also be used as an aid in the diagnosis of cerebral death.

Any isotope can be used in the evaluation of ventriculovenous shunt patency, and indium-111 DTPA cistinography can assist in the diagnosis of normal-pressure hydrocephalus.

Dacryocystoscintigraphy

In the investigations for epiphora, the patency of the nasolacrimal ducts may be investigated under physiological conditions by instilling a drop of normal saline containing 2 to 4 MBq of technetium-99m as pertechnetate into each conjunctival sac and recording the progress of the tracer through the canals into the nose (Fig. 17).

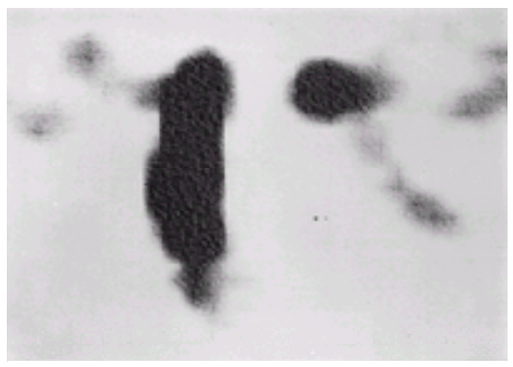


Fig. 17. Dacryocystoscintigraphy showing normal rapid flow of tracer from the conjunctiva through the lacrimal canal on the right, but hold-up on the left.

Tumour imaging

Many of the methods used for detecting benign and primary and secondary malignant tumours have already been discussed, but it is also necessary to describe radionuclide methods specifically designed to image tumours. These include gallium-67 chloride citrate scintigraphy, the use of labelled monoclonal antibodies against tumour-associated antigens, and the new class of indium-111-labelled, somatostatin-receptor ligands. Because breast cancer now affects one in every 11 women, there is also a brief note on the use of radionuclides in detecting primary breast tumours.

Gallium-67 citrate was initially a bone-seeking agent and it was being used for this purpose in a patient with Hodgkin's disease when its propensity to seek out lymphomatous nodes was realized. It is, however, also taken up by a number of normal tissues such as lymphoid tissue (notably Waldeyer's ring), lacrimal glands, liver, spleen, breasts, genitalia, and by faeces. In fact there is little space left to observe any abnormal gallium-67 uptake!

Nevertheless, it is useful in detecting the extent of intrathoracic lymphoma, with a sensitivity of 75 to 80 per cent, and it can detect primary bronchial carcinoma, with a sensitivity of 80 to 85 per cent. However, it was soon found to be an infection/inflammation-seeking agent as well and so it lacks specificity. It also does not concentrate as well in adenocarcinomas as in other types. If primary lung cancer concentrates gallium-67 citrate, detection of its spread to mediastinal nodes has a sensitivity approaching 100 per cent, compared to mediastinoscopy. However, specificity is much lower at 53 to 60 per cent.

With the advent of polyclonal antibodies against tumour-associated antigens, followed closely by the now ubiquitous monoclonal antibodies, hopes of finding tumour-specific radiopharmaceuticals were rekindled and concomitantly the chance of also using them in high activities for therapy. Antibodies can be labelled with specific radionuclides (iodine-123, indium-111, and even technetium-99m as pertechnetate) and provide at least the potential for the specific delivery of radionuclides to a particular cancer for detection and treatment.

Polyclonal antibodies are mixtures of antibodies having affinity for the tumour as well as, unfortunately, cross-reactive affinities towards other tissues. One of the first and best of these to be tried was carcinoembryonic antigen. Monoclonal antibodies, produced by hybridoma technology are, on the other hand, directed against relatively specific antigenic determinants in the neoplastic cell and so are less eager to react with normal cells. However, this lack of cross-reactivity is not absolute and much effort is being expended in producing antibody fragments, e.g. the $F(ab)'_2$ fragments, which have little or no cross-reactivity with normal tissue.

Preliminary results in cancers with well-characterized antibodies to specific tumour-associated antigens have been encouraging, for example in malignant melanoma (Fig. 18), ovarian cancer, and cutaneous T-cell lymphoma. In these instances, success is probably due to the fact that the antigen–antibody reaction may occur on the cell surface.

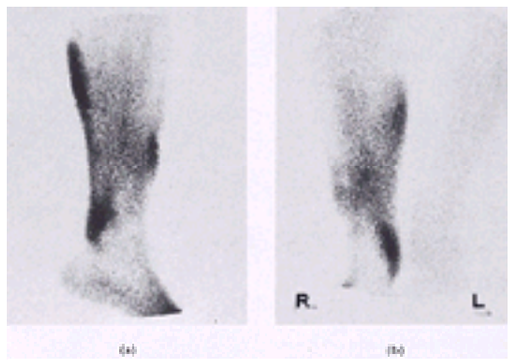


Fig. 18. Scintigrams from a patient with suspected metastatic melanoma on whom immunoscintigraphy was performed using monoclonal antibodies to the high molecular-weight, melanoma-associated antigen labelled with technetium-99m. There were no palpable lesions in the areas of increased uptake. (a) Lateral view of right leg; (b) anterior view of both legs.

Therapeutic studies show some promise, but at the present time both diagnostic and therapeutic applications have not lived up to their initial theoretical promise.

Somatostatin-receptor imaging has recently become well established for the *in vivo* localization of various tumours that contain these receptors, and their metastases. Several tumours, classically specified as either neuroendocrine or non-neuroendocrine, contain high numbers of somatostatin receptors and can be detected using the somatostatin analogue known as octreotide. Also, certain granulomas and autoimmune processes can also be imaged by this method, because they are able to accumulate somatostatin receptor-positive, activated mononuclear leucocytes.

Somatostatin-14, the naturally occurring form, is unsuitable for labelling or therapy due to its very short biological half-life. Curiously, efforts to synthesize an analogue with a longer half-life resulted in a shorter-chain molecule known as octreotide. This analogue proved to be effective in blocking the excess production of hormones by neuroendocrine tumours such as carcinoid, VIPoma (VIP, vasoactive intestinal peptide), gastrinoma, and insulinoma. The latest radiolabelled compound is [indium-111-DTPA-Phe'] octreotide. Structual sequences of these analogues are listed next (their bioactive sites are printed in italics).

Somatostatin-14:

Ala-Gly-Cys-Lys-Asn-Phe-*Phe-Trp-Lys-Thr*-Phe-Thr-Ser-Cys

Octreotide:

D-Phe-Cys-*Phe-D-Trp-Lys-Thr*-Cys-Thr(o1)

[Indium-111-DTPA-D-Phe'] octreotide:

indium-111-DTPA-D-Phe-Cys-*Phe-D-Trp-Lys-Thr*-Cys-Thr(o1)

Somatostatin receptors have been identified on many cells of neuro-endocrine origin, including the somatotroph cells of the anterior pituitary, the thyroid C cells, and the pancreatic islet cells. Also, cells not known as classically neuroendocrine, such as lymphocytes, may possess these receptors. Besides being found in normal tissue, these receptors have been demonstrated in most neuroendocrine tumours, many of which are derived from cells belonging to the amine-precursor uptake and decarboxylation (**APUD**) system, and which contain secretory granules. Somatostatin receptors have also been identified in central nervous, breast, lung, and lymphoid tumours. Imaging of those tumours with radiolabelled octreotide is therefore in competition with the much cheaper noradrenaline analogue, iodine-123 MIBG.

After intravenous injection there is normally faint visualization of the pituitary and the thyroid, with clear images of the liver, spleen, kidneys and bladder. Renal excretion constitutes the major excretory pathway, and in this regard the presence of renal images covering a large area of the abdomen is a disadvantage compared with iodine-123 MIBG. Together with the fact that many APUDomas are found near the liver and spleen, the presence of renal images means that all intra-abdominal, radiolabelled octreotide studies should be subjected to SPET. A small amount of the tracer, excreted via the biliary system, gets into the large bowel by 24 h—unfortunately the time at which pathology is best seen. Laxatives and frequent bladder voiding therefore help to clear the field and improve target:non-target ratios. A normal radiolabelled octreotide scintigram is shown in Fig. 19(a).

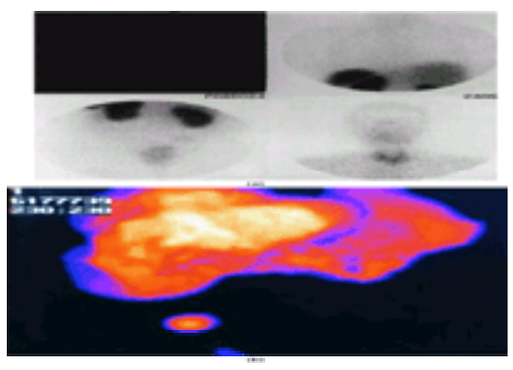


Fig. 19. (a) Planar indium-111 octreotide scintigram showing the normal distribution of the radiopharmaceutical in the kidneys, bladder, thyroid, pituitary, liver, and spleen. The presence of renal images covering a large area of the abdomen is a disadvantage compared with iodine-123 MIBG. (b) Three-dimensional images constructed from SPET cross-sections following indium-111 octreotide, showing a carcinoid tumour in the appendix.

The technique appears to have been most successful in imaging the APUDomas, where sensitivities between 60 per cent (insulinoma) and 100 per cent (gastrinoma, glucagonoma, and paraganglionoma) have been quoted. One-hundred per cent sensitivity has also been claimed for small-cell lung cancer. The three common entities of neuroblastoma, phaeochromocytoma, and carcinoid tumour (Fig. 19(b)) register sensitivities of between 85 and 95 per cent.

Apart from the detection of the tumour site as a prelude to surgical treatment, the technique can also be used to predict the effect of treatment with cold octreotide. Much information has still to be collected, especially as radiolabelled octreotide is still a very expensive radiopharmaceutical.

In the diagnosis of primary breast cancer the indications for radionuclide studies are almost identical to those of MRI and are as follows:

1. in the investigation of dense breasts, where mammography might be inconclusive or to distinguish between solitary and multifocal disease;
2. in the investigation of recurrence as opposed to scarring;
3. in finding lesions in residual breast tissue surrounding implants;
4. in the evaluation of clinically or mammographically suspicious lesions in which core biopsy is contraindicated (e.g. patients on anticoagulant therapy) or very difficult to perform (e.g. near the chest wall).

A number of tracers have been specifically evaluated for the detection of breast cancer, with varying degrees of success; the latest favourites are technetium-99m sestamibi and technetium-99m tetrafosmin. These are lipophilic tracers that accumulate in the cell and mitochondria as a result of negative transmembrane potential. Cells with high energy, however, such as neoplastic cells, are rich in mitochondria and so avidly take up the agents. Sensitivities and specificities of over 90 per cent have been claimed, but few series have yet tried to manage their patients using scintigraphy alone. However, it has a place as a complementary examination in the situations des-cribed above.

As far as breast screening is concerned, it would clearly be physically impossible to perform scintimammography on all women with dense breasts, especially as nearly

one-third of all mammogramsfall into this category and are becoming even more common with the majority of postmenopausal women on hormone replacement therapy.

PET using FDG will map areas of glucose metabolism: this is impaired in tumours, leaving radiolabelled glucose within a lesion. When this method is available, it is considered of choice for the detection of primary breast cancers and for the assessment of lymph-node involvement. Some investigators have reported an overall sensitivity of 100 per cent for both examinations, all in patients with ‘inconclusive mammography'.

The use of labelled hormones, especially 16-a-fluoroestradiol labelled with fluorine-18, has established a clear correlation (88 per cent) between oestrogen-receptor status and histology. Knowledge of such receptors, determined by any method, is important in the evaluation of response to tamoxifen. Studies with radiolabelled progesterone receptors have been less successful.

A definite advantage of FDG PET over other techniques is its ability to detect distant metastases throughout the whole body; it is also useful both in monitoring chemotherapeutic response and evaluating prognosis.

Exciting new developments are taking place in tumour-receptor imaging: for example, radioligands that bond oestrogen and, in breast carcinoma, have been used to image tumours and their secondaries and to monitor, for example, the efficacy of tamoxifen, which can be seen to make oestrogen-positive tumours (about 65 per cent of the total) disappear!

Infection/inflammation-seeking tracers

Intra-abdominal infection and inflammation are major surgical problems, to the resolution of which all imaging modalities (but particularly ultrasonography) can contribute. As far as radionuclide imaging is concerned, the best-known agents for seeking out inflammatory or infective foci have been gallium-67 citrate- and technetium-99m-labelled colloids, but these have been superseded by white cells (either pure polymorphonuclear leucocytes or mixed) labelled *in vitro* with indium-111 chelates or technetium-99m HMPAO. More recently, radiolabelled microcolloids, monoclonal antibodies, immunoglobulins, and antibiotics have been introduced.

Gallium-67 citrate has been used for some time and has a particular reputation for seeking chronic infection. It is simple to use and requires no special preparation, but, as stated above, it targets so many normal structures that there are few areas where the target:non-target ratio is high enough to engender analytical confidence. It is also difficult to image technically as it has three so-called photopeaks that make gamma-camera tuning difficult. This makes its sensitivity for assessing intra-abdominal sepsis rather low (about 60 per cent). Its specificity is also low, for it is a tumour seeker as described, and sequential studies over 3 days have to be carried out to enhance accuracy.

On the other hand, it is now becoming a comparatively simple matter to label leucocytes with either indium-111 oxime, indium-111 tropolone, or technetium-99m HMPAO (the same substance as is used for cerebral perfusion and tumour imaging). HMPAO, because of its availability and the optimal imaging characteristics of technetium-99m, has recently emerged as the labelling agent of choice. Preparation is time consuming and technique has to be meticulous in terms of sterility and the safe handling of blood products, but is nevertheless fairly simple. In brief, the white cells have to be separated from red cells and from platelets, resuspended in the plasma, mixed with the radiolabel, and reinjected into the patients. It would appear that pure granulocytes are not significantly superior for the detection of acute soft-tissue infections so, as their preparation requires an extra stage, it seems just as effective to label mixed white cells.

There is, as expected, normal leucocyte uptake in the haemopoietic areas of the bone marrow, liver and spleen, but no renal or bowel activity as in the case of gallium-67 citrate. It therefore provides a larger clear background in which to identify sepsis. Focal areas close to the liver and spleen may be imaged by using technetium-99m sulphur colloid to outline the liver and spleen, and then subtracting these areas electronically from the overall image.

When the inflammatory response due to intra-abdominal sepsis is marked, early imaging at 1 h after the reinjection of the labelled leucocytes can be obtained. However, with less acute responses, uptake is slower and may be seen at 3 h (Fig. 20). Sometimes even a study at 24 h may be necessary. Delayed imaging can distinguish between abscess and exudate in Crohn's disease and ulcerative colitis. The labelled exudate moves distally, whereas the activity over an abscess obviously remains static. Leucocyte abundance in the bowel is always pathological.

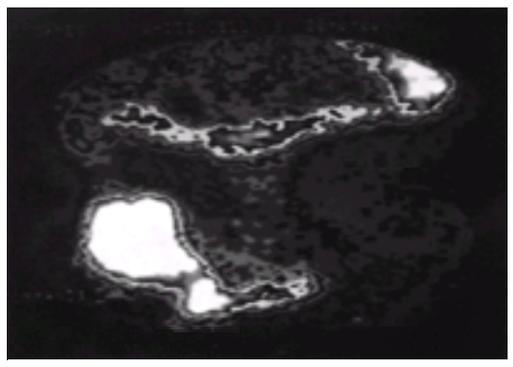


Fig. 20. An early labelled leucocyte study using indium-111 tropolone on a patient with Crohn's disease of the large bowel with a large abscess in the right iliac fossa. This was still visible 24 h later, whereas the rest of the uptake, presumably in exudate, had been passed.

The sensitivity of radiolabelled, mixed white cells is 100 per cent in detecting acute soft-tissue infections but drops to 85 per cent in mixed infections. The specificity is about 92 per cent, with false-positive results due to uptake in inflammatory conditions such as acute pancreatitis and inflammatory bowel disease, necrotic tumours, dissolving haematomas, and in patients on non-steroidal anti-inflammatory agents. Swallowed saliva may also produce misleading appearances.

When intra-abdominal infection is suspected and localizing signs are present, or if scintigraphy is negative or inconclusive, then ultrasound or CT must be used. If not, leucocyte scintigraphy should be the investigation of choice, with a negative study virtually excluding infection. However, a positive study is seldom sufficient to allow surgical drainage without further help from ultrasound or CT.

Nuclear pathology

The non-imaging branch of nuclear medicine is often referred to as nuclear pathology. Although not strictly the province of the surgeon, some of the tests under this heading could be useful in clinical or research settings.

Paramount among these tests is the ability to estimate accurately almost infinitesimal amounts of certain hormones, enzymes and drugs in the blood using the techniques of competitive binding assay, especially those using so-called radioimmunoassay techniques involving radioactive tracers, particularly iodine-125.

There are tests available employing radionuclides to assess a number of physiological variables. These include measurements of body composition (especially blood and plasma volumes), circulation and blood flow, tumour turnover and metabolism.

Safety of nuclear medical studies

The universally adopted policy in radiation protection is to keep radiation dose as low as reasonably achievable (the ALARA principle) and so it is important that the clinician has a reasonable indication that the potential gain for the patient in using a radioactive test will exceed potential risks. Non-essential, repetitive examinations and the use of radionuclides with large b-irradiation components must be discouraged. All investigations with ionizing radiation are to be avoided in pregnant women unless the clinical benefit far outweighs the risk. In addition, it is essential to use sensitive detection equipment, good handling techniques, and highly trained personnel. Using technetium-99m as the basis of most nuclear medicine studies, absorbed doses are of the same order as those incurred in conventional radiography and often far less.

Further reading

Alazraki NP, Taylor AT Jr. *New fundamentals of nuclear medicine: patient management for the 21st century*. The Society of Nuclear Medicine Inc., New York, 1998. [Short introductory text.]

Maisey MN, Britton KE, Gilday DL, ed. *Clinical nuclear medicine*, 3rd edn. Chapman and Hall Medical, London, 1998. [Intermediate text.]

Mistry R. *Manual of nuclear medicine procedures*. Chapman and Hall, London, 1988. [A detailed, how-to-do-it recipe book.]

Morton D, Wen D, Cochran A. Management of early-stage melanoma by intraoperative lymphatic mapping and selective lymphoadenectomy: an alternative routine elective lymphoadenectomy or ‘watch and wait'. *Surgical and Oncological Clinics of North America* 1992; **127**: 392–9.

Murray IPC, Ell PJ, ed. *Nuclear medicine in clinical diagnosis and treatment*, 2nd edn. Churchill Livingstone, London, 1998. [Long, detailed text.]

Thompson C, Stacker S, Salehi N. Immunoscintigraphy for the detection of lymph node metastases from breast cancer. *Lancet* 1984; **ii**, 1245–7.

Veronesi U *et al.* Sentinel node biopsy can avoid axillary dissection in breast cancer patients with clinically negative lymph nodes. *Lancet* 1997; **349**: 1864–7.

13.8 Nuclear magnetic resonance spectroscopy

Bheeshma Rajagopalan and Peter Styles

The basis of magnetic resonance spectroscopy
The information available
Phosphorus (^{31}P) MRS
Proton (^{1}H) MRS
Other nuclei
Obtaining MRS data
The role of MRS in clinical practice
Examples of diseases in which MRS can contribute to surgical management
Transplantation and rejection
Cancer
Epilepsy
Secondary brain damage
Further reading

The basis of magnetic resonance spectroscopy

Magnetic resonance spectroscopy (**MRS**) is an investigative modality that exploits the same basic phenomenon as magnetic resonance imaging (**MRI**). In each case, radiofrequency signals are generated from atomic nuclei that possess the property called ‘spin'. When placed in a strong and uniform magnetic field, these nuclei can be excited by a short pulse of radiofrequency energy and subsequently emit a weak signal that can be detected and recorded in a computer. In MRI, the presence of strong gradients in the magnetic field facilitates the production of images that are derived from the hydrogen nuclei (protons) in body water. In spectroscopy, individual chemicals can be identified due to the fact that the electrons which surround the nuclei modify the external magnetic field and thus provide individual chemicals with a characteristic signal frequency. The data are presented as ‘spectra' in which the strength of the MR signal is plotted against its frequency. Only some nuclei give useful signals, the most notable for human investigation being hydrogen (^{1}H) and phosphorus (^{31}P). It is important to appreciate that spectroscopy signals are very weak, owing to the low concentrations of metabolites in the body compared with the water which is used to give MRI signals. As a rule of thumb, MRS only detects small and mobile metabolites present in concentrations in excess of around 1 mM. This low sensitivity also seriously restricts the spatial selectivity that can be realized using MRS. In MRI a single pixel might measure about 1 × 1 × 3 mm, but MRS spectra will be obtained from volumes ranging typically between a few and a few tens of millilitres for proton and phosphorus, respectively.

The information available

Phosphorus (^{31}P) MRS

Phosphorus MRS detects important metabolites involved in energy metabolism such as adenosine triphosphate (**ATP**), inorganic phosphate (Pi), phosphocreatine (PCr), and some sugar phosphates involved in glycolysis. Of particular note is that the position of the peak for inorganic phosphate is dependent on the ionic state of the phosphate group and may be used as an accurate measure of intracellular pH. The phosphorus spectrum can also contain peaks arising from compounds involved in membrane metabolism, and these occur in the phosphomonoester (PME) and phosphodiester (PDE) regions of the spectrum.

The importance of ^{31}P MRS is that it reports on the very basis of cellular viability, its energy status. In certain tissues one can assess the response of tissue to stress by performing exercise (skeletal muscle), administering an inotropic agent (heart), or providing a metabolic load (liver). Examples from muscle studies in normal individuals and patients with vascular disease are shown in Fig. 1.

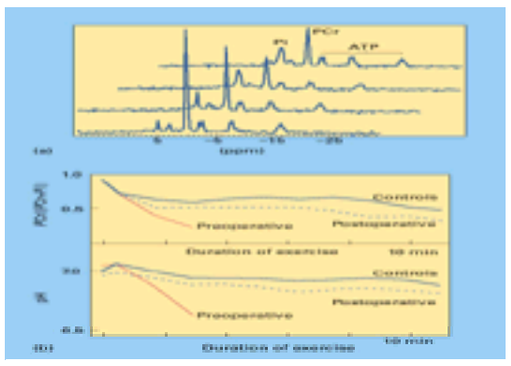


Fig. 1. ^{31}P MRS of skeletal muscle. (a) ^{31}P muscle spectra obtained at rest (lower spectrum) and at increasing time throughout an exercise regimen. Intracellular pH is obtained from the position of the inorganic phosphate (Pi) peak. Note the decrease in phosphocreatine (PCr), increase in Pi, and broadening of the Pi peak that reflects developing pH heterogeneity. (b) Time courses for changes in PCr and pH in patients with claudication before and after vascular surgery. Note that surgery increases exercise duration, decreases PCr utilization, and reduces lactate production as reflected by a smaller pH change. (Data kindly provided by Dr D.J. Taylor and Miss L.J. Hands).

Proton (^{1}H) MRS

Proton MRS is most useful for studying the brain. The spectrum contains signals from *N*-acetyl aspartate (NAA) (function uncertain, but believed to be a marker of neuronal viability), choline-containing compounds (Cho) (which may be involved in membrane metabolism or in the synthesis of acetylcholine), creatine (Cr) (involved in energy metabolism) and myo-inositol (mI) (function uncertain). In situations where anaerobic metabolism is important (e.g. ischaemia, or in the presence of macrophage invasion), it is also possible to observe lactate. Depending on the pulse sequence and echo times chosen for data acquisition, other small molecules can be detected such as g-aminobutyric acid (GABA) and glutamine + glutamate (Glx). Some of these features are illustrated in Fig. 2.

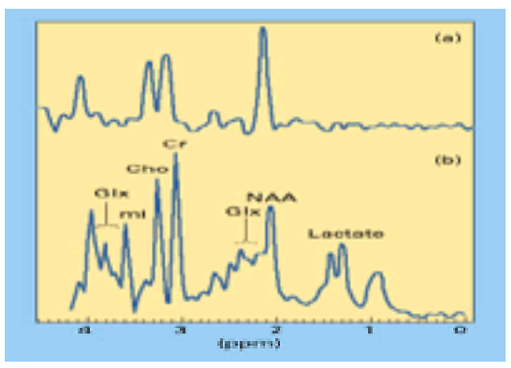


Fig. 2. ^{1}H spectra of human brain: (a) a spectrum of normal brain collected at an echo time of 135 ms; (b) a spectrum collected at a short echo time. This spectrum is from a patient who had suffered from a serious hypoxic episode that resulted in high lactate concentration. Note the extra peaks recorded at the short echo time and the undulating baseline caused by the presence of large molecules that are not detected at longer echo times. (Spectrum B kindly provided by Drs B.D. Ross and R. Kreis, The Huntington Medical Research Institute, Pasadena, CA).

Other nuclei

The ^{13}C nucleus gives magnetic resonance signals, but unfortunately this isotope is only about 1 per cent naturally abundant and so the signals are very weak. In the absence of any isotopic enrichment, spectra are only observed from concentrated molecules such as fat and, under favourable conditions, glycogen in muscle and liver. The use of isotopically enriched substrates such as ^{13}C-labelled glucose has given unique insight into metabolic pathways and energy regulation. Unfortunately, such compounds are prohibitively expensive for routine use.

Fluorine (^{19}F) is a sensitive nucleus, but does not occur naturally in solution in the body. However, the metabolism of fluorinated drugs such as 5-fluorouracil may be studied non-invasively.

Obtaining MRS data

Strategies for obtaining MRS data *in vivo* are largely dictated by the very weak signal strength, which in turn determines the minimum volume of tissue that can be investigated in a reasonable examination time. It is quite possible to obtain full, two- or even three-dimensional images of individual compounds, but the signal-to-noise ratio decreases rapidly with increasing spatial resolution. An alternative approach is to select a region of interest (with reference to a conventional MRI scan) and then use a ‘single voxel' technique in which a spectrum is obtained from this region alone. In cases where the tissue of interest is relatively superficial and homogeneous (e.g. skeletal muscle) it may even be possible to collect the MRS signals from a local surface-coil detector without resorting to more complex localization schemes. More often, however, some additional localization is needed and the choice between single-voxel detection and imaging-type methods (termed chemical-shift imaging) will be dependent on the particular investigation.

The role of MRS in clinical practice

Spectroscopy has traditionally been considered as a research tool rather than a method for routine investigations. However, as an increasing number of clinical MRI instruments are installed with MRS capability, this view is slowly being revised. Of particular interest is spectroscopy of the brain, owing to the fact that medical and surgical intervention is usually undertaken in the absence of biochemical information which, elsewhere, would be available from biopsy. Indeed, our ability to measure dynamic physiological and biochemical processes non-invasively provides unique information in a range of progressive diseases and functional disorders such as epilepsy. It is to be expected that this information will play an increasing part in the management of such patients and assist with surgical planning.

Examples of diseases in which MRS can contribute to surgical management

Peripheral vascular disease and muscle energetics

Energy metabolism of the gastrocnemius–soleus complex has been studied by ^{31}P MRS in patients with intermittent claudication during and after exercise. During exercise (plantar flexion), muscles initially obtain their ATP requirements from the energy stored in phosphocreatine, so that the first change to occur in the spectra is a fall in phosphocreatine and a rise in inorganic phosphate. Further supplies of ATP are obtained from either glycolysis (metabolism of glucose or glycogen to lactate) or oxidative phosphorylation of pyruvate or fatty acids via the Krebs cycle, their relative contributions being determined by the availability of oxygen. As patients with claudication have reduced blood flow, breakdown of phosphocreatine and glycolysis become more important as the sources of ATP. These patients thus show a greater fall in phosphocreatine and pH (measured from the inorganic phosphate peak) than do controls. The synthesis of ATP during recovery, however, is exclusively oxidative. Thus, owing to the reduced supply of oxygen, metabolic recovery in these patients is slower than in normal individuals. Though the metabolic differences in patients with claudication do correlate with the severity of the vascular disease, the correlations are not sufficiently good to use this technique as a clinical tool for assessing patients. However, as repeat studies in a single patient are reproducible, the technique has proved useful in the assessment of therapeutic interventions.

Transplantation and rejection

Attracted by its non-invasive nature, investigators have used ^{31}P MRS to assess the severity of rejection in transplanted hearts, kidneys, and livers. In the heart, rejection is accompanied by a fall in the ratio of phosphocreatine to ATP in heart muscle and the change correlates with the severity of the myocardial damage. These changes, however, are not specific to rejection. The ratio is reduced in cardiac muscle in patients with heart failure due to a variety of valvular and myopathic disorders.

The liver and kidney do not have phosphocreatine in their cells. However, *in vivo* spectra show elevation of the phosphomonoester peak in patients with histological evidence of rejection. *In vitro* studies on biopsies have revealed that the increased phosphomonoester peak is due to increases in both phosphoethanolamine and phosphocholine, both intermediates in membrane metabolism, implying increased turnover of hepatocyte membranes, presumably due to the rejection process. These changes are again not specific for rejection and occur in other conditions such as hepatitis.

Applying ^{31}P MRS to the transplanted kidney, rejection has been shown to increase the inorganic phosphate:ATP ratio and reduce intracellular pH in contrast to acute tubular necrosis, which was associated with a normal pH.

^{31}P MRS has also been used to assess the viability of kidneys and livers before transplantation. The signal from ATP decreases with time, and kidney function immediately after transplantation is better in kidneys with higher amounts of ATP. Similar findings have been reported in the liver. As spectroscopy takes between 30 to 60 min to perform, and transplanted organs with significantly reduced amounts of ATP may recover with time, this technique has not been applied in clinical practice.

^{31}P MRS has, however, proved very useful in the assessment of preservation fluids and cytoprotective agents. Isolated hearts, kidneys, and livers from animals have been perfused with a variety of preservation solutions, and the spectra collected serially over time. As MRS is non-destructive, the response of a single organ to a variety of interventions can be studied. Organ viability can be followed using the amounts of ATP and changes in intracellular pH.

Cancer

Both ^{31}P and ^{1}H MRS have been applied to a variety of tumours to aid diagnosis and follow the response to treatment. In ^{31}P spectra a fairly consistent observation is that the concentrations of phosphomonoesters are elevated in tumours. This has been shown in brain tumours, both primary and secondary tumours in the liver, as well as in solitary tumours elsewhere in the body. This increase has been shown to be due to an increase in phosphocholine, phosphoethanolamine or both. The mechanisms responsible for this increase are not understood, but as both compounds are components of membranes, it is probable that the changes reflect increased rates of membrane turnover. The amount of phosphomonoester decreases following successful therapy in several types of liver tumours including hepatic lymphoma, and relapse is often heralded by an increase in the phosphomonoester peak.

Though ^{31}P spectra have not proved to be useful in determining the type or grade of tumours, studies using ^{1}H MRS have been more successful. These studies have been confined to the brain, and have been largely empirical. ^{1}H spectra obtained from tumours *in vivo* have been analysed using a variety of techniques such as artificial neural networks and pattern recognition, and it has been possible to show that the relative amounts of metabolites such as creatine, choline, *N*-acetyl aspartate, and lactate vary according to tumour cell type and tumour grade. Similar conclusions have been reported from ^{1}H MRS of extracts of biopsies from tumours. The reasons for the phenotypic changes are unknown. Currently no correlative studies between *in vivo* and *in vitro* results in tumours have been published and these techniques remain experimental.

Epilepsy

Surgery for temporal-lobe epilepsy is now well established. A variety of imaging techniques, such as positron-emission tomography, MRI and MRS, have been used to characterize the underlying abnormality, lateralize the main seizure focus, and plan any surgical intervention. ^{1}H MRS has proved to be extremely useful in the assessment of these patients because epileptic activity is associated with metabolic abnormalities. Patients with MRI abnormalities show reduced amounts of *N*-acetyl aspartate in the temporal lobe. Patients with temporal-lobe epilepsy who have normal findings by MRI also have reduced ratios on the side of the abnormal electroencephalogram. This suggests that ^{1}H MRS may be a more sensitive indicator than MRI of temporal abnormalities in these patients, and many centres now use ^{1}H MRS as a part of their standard investigation protocol while planning surgery.

Secondary brain damage

Several brain disorders are characterized by damage to the parenchyma secondary to other primary lesions. Examples include vasospasm secondary to subarachnoid haemorrhage and brain compression due to space-occupying tumours and hydrocephalus. It has been shown by a combination of physiological MRI (perfusion and cerebral blood-volume imaging) and spectroscopy that ischaemia plays a significant part in the pathophysiology of these conditions. A critical question is whether the tissue is fully functional, marginally perfused, or irreversibly damaged. Both ^{31}P and ^{1}H spectroscopy can help with the assessment of brain viability in these cases. In particular, the presence of lactate in the proton spectrum is an indication that the brain is relying on anaerobic metabolism, presumably due to marginal perfusion. There is evidence to suggest that this information can be a useful aid to surgical planning and might indicate an increasing urgency to operate on otherwise slowly developing conditions.

Further reading

Connelly A, Van-Paesschen W, Porter DA, Johnson CL, Duncan JS, Gadian DG. Proton magnetic resonance spectroscopy in MRI-negative temporal lobe epilepsy. *Neurology* 1998; **51**, 61–6. [An example of the application of MRS in epilepsy.]

Gadian DG. *NMR and its application to living systems*, 2nd edn. Oxford University Press, 1995. [A wide-ranging textbook covering the principles and practice of both MRI and MRS.]

Kemp GJ *et al.* Calf muscle mitochondrial and glycogenolytic ATP synthesis in patients with claudication due to peripheral vascular disease analysed using ^{31}P magnetic resonance spectroscopy. *Clinical Science* 1995; **89**, 581–90. [Includes a detailed description of the interpretation of MRS data from muscle.]

Radda GK, Rajagopalan B, Taylor DJ. Biochemistry *in vivo*: an appraisal of clinical magnetic resonance spectroscopy. *Magnetic Resonance Quarterly* 1989; **5**, 122–51. [A review article that includes an explanation of the biochemistry that underlies ^{31}P MRS.]

Robinson SP, Barton SJ, McSheehy PM, Griffiths JR. Nuclear magnetic resonance spectroscopy of cancer. *British Journal of Radiology* 1997; **70**, S60–9. [A good overview of MRS in cancer.]

Various articles in *NMR in Biomedicine*, Vol. 4/5, 1998. [This volume contains several articles dealing with the characterization of tumours using MRS.]

13.9 Positron-emission tomography

John M. Hoffman, Raghuveer K. Halkar, and Ernest V. Garcia

Introduction
Physical principles
Radiopharmaceuticals
Instrumentation
Cardiac imaging
- **Myocardial perfusion**
- **Myocardial glucose metabolism**
- **Myocardial viability**

Epilepsy
Oncologic imaging with FDG-PET
- **Patient preparation for tumor imaging by PET**
- **Cerebral neoplasm**
- **Lung cancer**
- **Colorectal and gastrointestinal malignancy**
- **Breast cancer**
- **Head-and-neck tumors**
- **Lymphoma**
- **Endocrine tumors**
- **Musculoskeletal tumors**
- **Genitourinary tumors**
- **Melanoma**

Further reading

Introduction

Positron-emission tomography (**PET**) is a non-invasive, high-resolution, three-dimensional, functional method of imaging that provides diagnostic information on biochemical and physiologic processes in patients. This is done by imaging and quantifying the distribution of positron-emitting radionuclides in the human body. These radionuclides are used to label substrates, ligands, drugs, antibodies, neurotransmitters, and other compounds that are tracers of specific processes. Since disease states often manifest as biochemical and physiologic changes before structural changes are noted, PET plays an important part in evaluating and managing patients. PET is often used clinically to image tumors, the brain, and the heart to complement other types of anatomic imaging, such as computed tomography (**CT**), magnetic resonance imaging (**MRI**) or angiography. PET is used when the structural information is not enough to detect or monitor a functional disorder.

Recently, two major breakthroughs have reduced the cost of positron imaging so that it is now affordable by most institutions. The wide choice of new positron-imaging devices coupled with the advent of radiopharmaceutical distribution networks (that circumvent the need for an on-site cyclotron) promise to open the field of positron imaging to practically everyone who realizes its diagnostic power.

Physical principles

Systems that create images from the radiation emitted from positron-emitting tracers use the principle of electronic collimation. They take advantage of the fact that positrons (or positive electron antiparticles) emitted from radionuclides usually interact with orbital electrons through a process known as pair annihilation. In annihilation, the mass of the electron and positron pair is converted into energy in the form of two photons that are equal in energy to the rest mass of an electron (511 keV each) but travel in exactly opposite directions.

The general principle of how these photons are used to create a tomographic image is to have detector pairs opposite each other to detect each of the two photons in coincidence. Thus, if the two detectors record an event within a predetermined time ‘window' (coincidence), then the pair annihilation must have taken place somewhere in a straight line between the two detectors. Images are then created as the intersection of these lines, with planes cutting across the object of interest. Since the electronics of the system has provided the direction of the emission it is said to use electronic collimation, in contrast to the mechanical lead collimators with holes used in conventional nuclear medical imaging. In acquisition and reconstruction by two-dimensional PET, only the lines that lie in or next to the planes perpendicular to the axis of the scanner are used.

Radiopharmaceuticals

The positron-emitting radionuclides used clinically are ^{82}Rb (rubidium), ^{15}O, ^{13}N, ^{11}C, and ^{18}F (fluorine). These have short half-lives of between 78 s (^{82}Rb) to 110 min (^{18}F), requiring that either a generator or a particle accelerator be on site for their production. From the list above, only ^{82}Rb is generator-produced; the rest are produced by automated ‘baby cyclotrons' whose advent significantly reduced the capital and operational costs of a PET center. Once these radionuclides have been produced they are rapidly synthesized to a specific molecular compound that satisfies the requirement of a particular biomedical application. By far the most clinically used positron-emitting radiopharmaceutical to date has been [^{18}F]fluorodeoxyglucose (**FDG**), a tracer of glucose metabolism. Robotics systems have been developed and are used widely for the automated synthesis of FDG. Because of FDG's numerous applications in the brain, heart, and oncology, and because of the 110-min half-life of ^{18}F, new commercial companies have set up regional cyclotron distribution networks to provide FDG to sites that do not have cyclotrons. This breakthrough has meant that practically any site that has a way of imaging FDG can order and use it clinically.

Instrumentation

There is a range of systems for positron imaging. Their cost and imaging performance are directly related to the total number and type of detectors surrounding the patient. Conventional, state-of-the-art positron scanners that cost over two million dollars use rings of detectors to surround the patient 360° through an axial field of view of 16 to 20 cm. These detectors usually use bismuth germanate (BGO) crystals, which have a much higher stopping power than the sodium iodide (NaI) crystals used in conventional, single-proton emission, computed tomographic (**SPECT**) systems. Variations from this geometry have been developed to reduce the overall cost of the system while trying to maintain imaging performance. One variation uses a rotating gantry consisting of two arcs of detector banks. Cost is reduced because there is only about half the number of detectors in conventional systems. Since fewer counts per unit time are being recorded, the system uses three-dimensional acquisition and reconstruction to increase the counting statistics of its images. This three-dimensional approach, which is also available in the newer conventional PET systems, measures coincidence events with connecting lines off of the transaxial plane perpendicular to the scanner's axis.

Another approach to reducing cost has been to surround the patient with multiple, NaI single-crystal detectors. Cost is reduced because NaI crystals are cheaper than BGO and because fewer photomultipliers are used to convert the scintillation to electrical pulses. The disadvantage is that fewer counts are absorbed by NaI and the electronics is somewhat slower in counting the events that do take place. To help circumvent these problems the crystals have been made thicker, and three-dimensional acquisition and reconstruction is used. An exciting spin-off of this approach is its incorporation into state-of-the-art, digital, dual-detector SPECT systems. These systems can be used with lead collimators for conventional nuclear medical procedures, or the collimators can be removed and the crystals placed in coincidence to use electronic collimation for positron imaging. Thus, for sites that have already purchased these SPECT systems, a $150 000 to $300 000 upgrade might allow them to perform positron imaging. This type of technology is usually referred as ‘positron or coincidence imaging' to contrast with the ‘PET' approach.

The spatial resolution of the imaging systems described above is between 4 to 6 mm. The main difference between systems is the number of useful events per unit time recorded and thus the total time of acquisition needed to image a patient. Some imaging centers use conventional SPECT cameras with high-energy, multiple-hole lead collimators to create positron images. These devices suffer from both poor spatial resolution (approximately 15–20 mm) and poor counting statistics. Although this approach has been found useful in FDG imaging for myocardial viability, it is of questionable use in tumor or brain imaging.

Cardiac imaging

The major emphasis in cardiology has been in the development of imaging techniques for evaluating myocardial perfusion and metabolism, their relation in the

assessment of coronary arterial disease, and, most importantly, myocardial viability.

Myocardial perfusion

Myocardial perfusion has been evaluated clinically using ^{82}Rb and [^{13}N]ammonia tracers. Rubidium, like thallium-201, acts as a potassium analog. An important advantage of rubidium is that it is generator-produced, thus obviating the need for an on-site cyclotron.[^{13}N]ammonia has properties that mimic those of radioactive micro-spheres. The tracer is rapidly cleared from the blood and is distributed in proportion to regional blood flow in the myocardium, where it becomes trapped. Both tracers generate images in which the regional myocardial intensity is proportional to blood flow. One important advantage of PET myocardial-perfusion imaging over SPECT is that the images are easily corrected for the attenuation by the body of the radiation emitted from the heart. This attenuation creates artefacts in SPECT images. Experts learn to interpret these artefacts, but they still lead to a significant reduction in specificity for detecting the absence of coronary arterial disease. The superior spatial and contrast resolution of PET over SPECT, as well as PET's ability to correct simply for attenuation, result in an increased accuracy for detection of coronary arterial disease. A recent meta-analysis of these methods on data from 13 published studies showed that, although the mean sensitivity of both techniques for detecting coronary arterial disease was rather similar (90 ± 6 per cent for PET and 90 ± 8 per cent for SPECT), the mean specificity of PET was nevertheless statistically significantly superior (91 ± 8 per cent vs 71 ± 20 per cent for SPECT).

Myocardial glucose metabolism

The myocardium utilizes a number of different energy-producing substrates. In the normal heart, especially in the fasting state, fatty acid metabolism provides approximately 70 per cent of cardiac energy requirements while carbohydrates account for most of the other 30 per cent. Following carbohydrate loading, plasma glucose increases and plasma fatty acids decrease. This metabolic change causes the heart to switch over to glucose as a primary energy source. Similarly, in the presence of focal hypoxia or ischemia, myocardial metabolism shifts over to glucose because of increased energy-producing efficiency.

Myocardial viability

FDG-PET has become the ‘gold standard' for assessing myocardial viability in patients with severely impaired function. FDG accumulates in the heart in proportion to glucose utilization by the myocardial cells. In PET images (Fig. 1) of patients who have been glucose-loaded, FDG uptake in normal myocardial segments appears to have a homogeneous distribution. In ischemic segments, FDG uptake appears normal or increased. In necrotic segments, after transmural myocardial infarctions, FDG uptake is absent. In segments with an admixture of necrotic and normal or ischemic myocardium, such as after a nontransmural myocardial infarction, there is a reduction of FDG uptake.

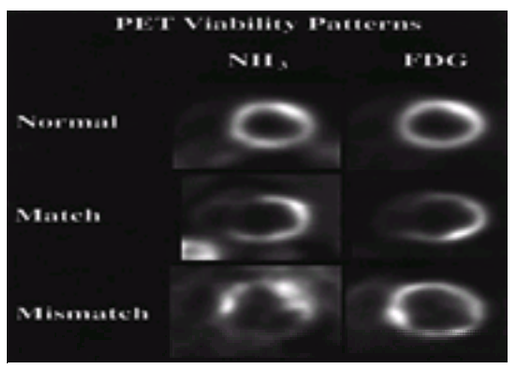


Fig. 1. PET imaging patterns of myocardial blood flow and metabolismfor assessing myocardial viability. The blood flow was assessed using[^{13}N]ammonia as a perfusion tracer and the glucose metabolism using [^{18}F]fluorodeoxyglucose (FDG). Representative short-axis tomograms are shown for three glucose-loaded patients. Upper panel: The left ventricular (LV) myocardium from a normal patient shows a homogeneous distribution of the perfusion (NH_3) and metabolism (FDG) tracers. Middle panel: The LV myocardium from a patient with a transmural myocardial infarction of the septum shows a marked matched reduction of septal uptake by both tracers. Lower panel: The LV myocardium from a patient with a hibernating inferior wall shows a marked reduction of perfusion (NH_3) with a mismatched normal metabolic uptake of FDG.

Clinical assessment of myocardial viability with PET combines the use of a resting myocardial perfusion scan with an FDG scan of myocardial glucose metabolism. In patients with myocardial segments exhibiting resting ischemia or hibernation, a high-risk ‘mismatch' pattern is expected. This pattern is shown as a reduction in uptake of the perfusion tracer and a significantly increased uptake of FDG relative to that of the perfusion tracer (Fig. 1). In patients with myocardial segments that have either transmural or subendocardial infarctions with no residual ischemia, a lower-risk ‘match' pattern is expected, with a relative reduction of uptake for both the perfusion and FDG tracers (Fig. 1).

These perfusion–metabolism patterns have been used in patients with coronary arterial disease and left ventricular dysfunction to predict the recovery of that function after revascularization and the prognosis of adverse cardiac events. A meta-analysis of five reports comprising 99 patients with coronary arterial disease showed that where extensive mismatch regions were detected the patients' left ventricular function improved significantly from a mean ejection fraction of 36 per cent before revascularization to 48 per cent after, whereas in patients who showed small or no mismatch regions, left ventricular function did not improve significantly after revascularization. Another meta-analysis of four reports comprising 388 patients showed that a low rate of adverse cardiac events (range 10–16 per cent) was detected in patients exhibiting the mismatch pattern of hibernating myocardium who were treated surgically, but a high rate (range 33 to 62 per cent) when they received medical treatment only. Patients that did not exhibit the mismatch pattern had relatively low rates of adverse cardiac events (range 6–18 per cent) whether they received medical or surgical treatment. These studies strongly suggest that perfusion–metabolism patterns are useful in management by helping to decide which patients should receive surgical or medical treatment and by predicting the improvement of left ventricular function for those patients who are revascularized.

Initial studies have also shown that perfusion–metabolism patterns can be used in patients with severe cardiomyopathy to help differentiate non-ischemic from ischemic disease. Patients with non-ischemic cardiomyopathy exhibit relatively normal, homogeneous perfusion and metabolism patterns whereas those with ischemic cardiomyopathy exhibit the mismatch and match patterns described above.

Epilepsy

FDG-PET has been established as an extremely useful technique to assist with the lateralization and localization of the temporal-lobe abnormality seen in partial-complex epilepsy. Typically, temporal-lobe hypometabolism (Fig. 2) is found unilaterally, which assists in determining the site of any future operation. This particular hypometabolic abnormality has been consistently correlated with the subsequent electrophysiologic localization of seizure onset and often with morphologic changes on MRI. FDG-PET has also been important in making it possible to avoid the use of depth electrodes for localization and lateralization. The degree of temporal-lobe hypometabolism is also predictive of eventual surgical outcome and subsequent seizure control.

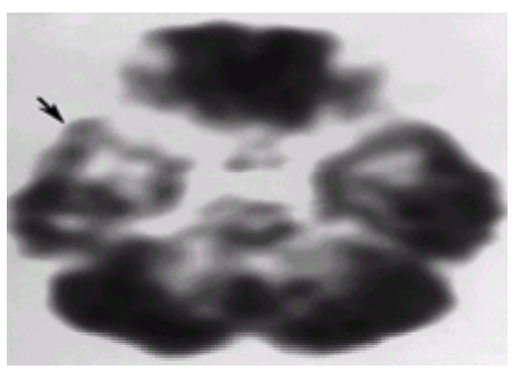


Fig. 2. Patient with long-standing partial-complex seizures; note reduced FDG uptake in the right temporal lobe (arrow) compared to the left.

Oncologic imaging with FDG-PET

[^{18}F]fluoro-2-deoxyglucose (FDG) is an analog of glucose, which is ideally suited for oncologic imaging. Almost all tumor tissues have greater intermediary metabolism than normal tissues. Malignant tissue has altered enzyme functions, with increases in certain glucose-transport proteins, and thus exhibits increased uptake. FDG is a unique tracer in that it is metabolically trapped, and therefore accumulates, in tissue. This provides a functional characterization, independent of the morphologic criteria used in conventional CT and MRI. Almost all tumor types have increased glycolysis, thus having an increased uptake of FDG. This fact is important in oncology because increased uptake can occur in tissue that appears morphologically normal. The current standard of practice for imaging the oncologic patient typically involves CT or MRI, but these may miss signs of early malignant transformation and lymph-node involvement because size and shape are not initially altered in many cases. Much of FDG-PET imaging is used to distinguish malignant from non-malignant tissue, and also to characterize treatment effects, which often cause morphologic changes that are difficult to describe with conventional CT or MRI. FDG-PET is able to differentiate clearly between tumor recurrence and necrosis and between recurrent or residual tumor and scar tissue. This is extremely important in almost all applications of oncologic imaging, since postsurgical and therapeutic effects are typically confounding problems.

Patient preparation for tumor imaging by PET

Patients should be fasted for a minimum of 4 h, which increases uptake into malignant tissue. At least 72 h should also have elapsed since biopsy or surgery because of the possibility of nonspecific FDG uptake secondary to inflammatory reaction. The imaging is typically done 60 min after FDG injection. In cases of colorectal and ovarian cancer, where evaluation of the pelvis is vital, an indwelling Foley catheter in the bladder with irrigation helps to decrease the concentration of FDG in the urine.

Cerebral neoplasm

Like all malignant tumors, brain tumors exhibit increased FDG uptake in the more malignant tissue. To date, the use of FDG-PET has been studied more in patients with brain tumors than in any other oncologic application. The technique is able to differentiate low- from high-grade malignancies (Fig. 3), which provides important prognostic information. FDG-PET is also important in monitoring post-treatment effects; as described above, MRI or conventional CT are not able to differentiate between radiation necrosis and recurrent tumor (Fig. 4). FDG-PET has also been useful in assisting the neurosurgeon to characterize heterogeneous lesions for biopsy localization.

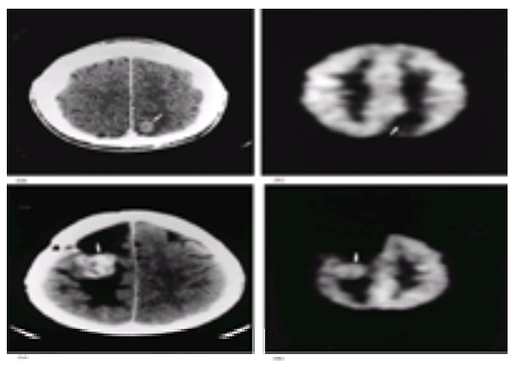


Fig. 3. (a) Anatomic study showing an enhancing abnormality in the left occipital area (arrow). (b) FDG-PET image showing a hypometabolic lesion (arrow) corresponding to the anatomic abnormality; this is a low-grade astrocytoma. (c) Anatomic study showing an enhancing lesion in the right frontal area (arrow). (d) FDG-PET image shows a hypermetabolic lesion (arrow) consistent with a high-grade recurrence of glioblastoma multiforme in the same area.

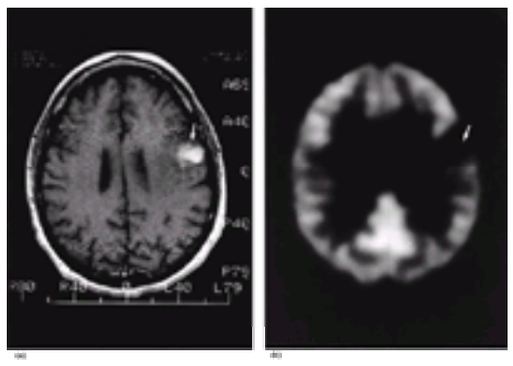


Fig. 4. (a) Anatomic study showing enhancement in the left frontal area (arrow)—recurrence or necrosis?(b) FDG-PET image at the same level showing a hypometabolic area (arrow) consistent with radiation necrosis.

Lung cancer

Solitary pulmonary nodule

Despite advances in morphologic imaging such as CT, the proportion of solitary pulmonary nodules resected that then prove to be malignant is 20 to 40 per cent; PET is far superior in differentiating benign from malignant solitary pulmonary nodules (Fig. 5), with a reported sensitivity of 93 to 100 per cent and a specificity of 78 to 88 per cent. The specificity may vary in relation to the local prevalence of granulomatous diseases such as tuberculosis, histoplasmosis, aspergillosis, coccidioidomycosis, and blastomycosis. Qualitative and semi-quantitative techniques such as the standardized uptake value have been shown to distinguish benign from malignant lesions with great accuracy.

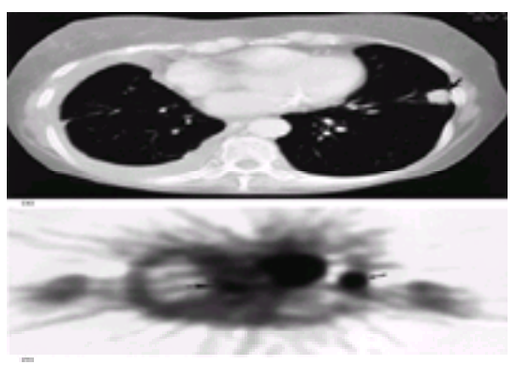


Fig. 5. (a) CT scan of a female patient showing a spiculated mass in the left lung (arrow). The pleural effusion did not show malignant cells and mediastinal lymph nodes were indeterminate. (b) FDG-PET showed focal area of increased FDG uptake in the left lung mass and mediastinal uptake on the right side (small arrow), suggesting lymph-node involvement (large arrow).

Staging

Assessment of hilar and mediastinal lymphadenopathy and distant metastases is vital in the preoperative staging of lung cancer. Several studies have shown the superiority of FDG-PET over CT in assessing lymph-node involvement. Mediastinoscopy is an accepted standard of care in assessing mediastinal involvement, but it is invasive. It is recommended where PET studies are equivocal or in geographic areas where false-positive findings due to granulomatous disease are frequent. Bury *et al.* have shown that, in 109 patients, FDG-PET changed the nodal staging in 33 per cent and the staging for distant metastases in 14 per cent. In an extensive analysis of cost-effectiveness, Gambhir *et al.* have shown that CT + PET saved $1154 per patient, without loss of life expectancy, in a conservative approach in which imaging findings were confirmed by mediastinoscopy. When mediastinoscopy was not used, PET saved $2267 per patient but missed 1.7 per cent of potentially operable patients.

Colorectal and gastrointestinal malignancy

Colorectal cancer is the third most common cancer in the United States, and is resectable in 70 per cent of the patients at the time of presentation. Recent studies have shown that PET has a similar positive predictive value and a better negative predictive value than CT in preoperative evaluation for colorectal cancer.

In recurrent colorectal cancer the size and number of hepatic metastases and the presence of extrahepatic metastasis will decide the prognosis (Fig. 6). As described above, CT and MRI use morphologic criteria, are unable to differentiate between postsurgical changes and local recurrence, and hence have poor accuracy. Many reports have shown that FDG-PET imaging is as good as CT in detecting liver metastasis and is certainly superior to CT for extrahepatic focal metastasis. Detection of unexpected extrahepatic metastasis changes the management and avoids unwarranted operations. The preliminary result of a survey of 14 PET centers and 267 colorectal cancers has shown a significant cost-saving potential of PET imaging.

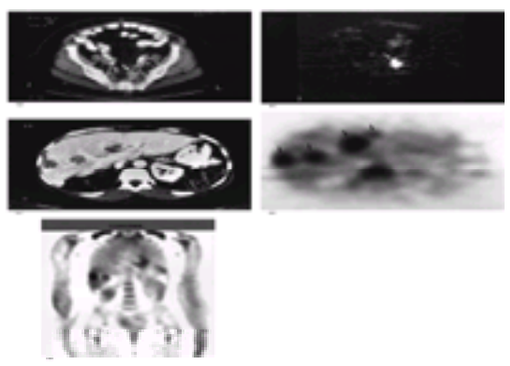


Fig. 6. (a) Anatomic study of the pelvis showing lesion compared with scar (arrow) in patient with colorectal carcinoma; (b) FDG-PET image of the same location showing a hypermetabolic lesion (arrow) consistent with a recurrence. (c) Patient with colorectal carcinoma with numerous lesions on CT scan of liver; (d) corresponding FDG-PET image shows many hypermetabolic lesions (arrow) consistent with liver metastasis. (e) Coronal whole-body FDG-PET image of the above patient showing the capacity of newer scanners to image the entire body.

Breast cancer

Breast cancer remains one of the most serious and common malignancies in women. Despite a national effort to increase awareness and screening, it continues to have an unacceptably high mortality rate. The treatment of breast cancer may involve surgery, radiation, and chemotherapy. At present, a critical issue in the evaluation of women with breast cancer concerns determining the presence of axillary lymph-node involvement. Numerous techniques have been developed in an attempt to assist with this particularly difficult prognostic problem. FDG-PET has been shown to be extremely useful, with high sensitivity, specificity, and predictive value in the evaluation of the axillary lymph nodes, which is essential for eventual surgical treatment. Dissection of these nodes can be expensive and prone to complications, and, in the majority of cases, is negative for tumor involvement. Therefore, many women are undergoing axillary dissection without having nodal metastases and receive no therapeutic benefit from the procedure. FDG-PET is extremely diagnostic in predicting lymph-node involvement in women with breast cancer. In a series by Adler et al. (1997), using FDG-PET, the sensitivity and negative predictive value were both 95 per cent for imaging of the axillary lymph-node system, and it was concluded that patients with a negative PET scan had such a low risk for axillary involvement that dissection was not warranted. A positive FDG-PET study, however, was able to assist in confirming the presence and determining the number of positive lymph nodes eventually found at operation. Since the majority of lymph-node dissections do not reveal malignant tissue, the cost savings from this particular procedure alone would be extremely important.

Head-and-neck tumors

Like other malignant tumors, squamous carcinomas of the head and neck have increased FDG uptake, which appears to correlate with their aggressiveness. FDG imaging is useful both before and after therapy. High uptake in untreated head-and-neck cancer is associated with advanced disease and usually with a poor survival. FDG-PET is also extremely important in monitoring the effects of combined surgery, radiation, and chemotherapy. The treatment of head-and-neck cancer has become a multidisciplinary endeavor including the surgeon, radiation therapist, and medical oncologist. Therefore, a critical assessment of the primary tumor, as well as of lymph-node involvement and therapeutic effectiveness, is extremely important. Conventional imaging with CT or MRI is often difficult, particularly after therapy. Scarring, inflammatory changes, and a lack of lymph nodes that reach the size criteria for positivity often confound the treatment of such patients. FDG-PET provides the clinical information needed to improve both presurgical evaluation and the characterization of lymph-node involvement, and to monitor post-therapeutic effects.

Lymphoma

Malignant lymphoma has traditionally been characterized and staged by conventional imaging techniques such as CT and MRI. The extent of nodal involvement above and below the diaphragm is critical to the eventual treatment. A typical evaluation includes planar chest radiography, lymphoscintigraphy, and mediastinoscopy. Nuclear medical studies using gallium-67 scintigraphy and CT have also been essential in evaluating the patient with lymphoma. Unfortunately, conventional morphologic imaging with CT or MRI is based on criteria for lymph-node size and has been found to miss a significant number of lymph nodes involved with tumor. FDG-PET has been extensively evaluated in patients with lymphoma and appears to have excellent sensitivity and specificity for determining and staging lymph-node involvement.

Endocrine tumors

Thyroid cancer is the most common cancer of endocrine origin and differentiated thyroid cancers have very good prognosis. Thyroglobulin determination and ^{131}I whole-body imaging are the well-established methods of follow-up after the primary tumor has been surgically removed and remnant thyroid tissue ablated with radioiodine. FDG-PET is useful in poorly differentiated thyroid cancer with rising thyroglobulin and negative ^{131}I whole-body images.

For other neuroendocrine tumors, such as carcinoid, gastrinoma, insulinoma and pheochromocytoma, the diagnosis is based on biochemical features; conventional imaging such as CT and ultrasound have poor accuracy for localizing primaries and metastases. Large studies comparing FDG-PET and [^{111}In]octreotide or [^{131}I]MIBG are lacking.

Musculoskeletal tumors

The extent of FDG uptake correlates well with the grade of musculoskeletal tumors. Eosinophilic granuloma and multiple myeloma may be an exception to this. PET is more accurate than CT or MRI in preoperative staging. Monitoring the treatment response in sarcomas has shown that prominent FDG uptake occurs within benign tissues adjacent to treated tumor.

Genitourinary tumors

Amongst gynecological tumors, ovarian cancer is the leading cause of death. The disease is usually advanced by the time of diagnosis. Initial management and staging

is usually by surgery. Recurrence can be monitored by tumor markers such as CA125, but these are not always positive in certain tumors. Conventional imaging such as CT and other techniques such as laparoscopy and coloposcopy still lack accuracy. FDG-PET has shown better specificity and sensitivity than CT.

Prostate cancer is one of the most common malignancies in men. Unfortunately, the results with FDG-PET in prostate cancer have been less favorable. The closeness of the tumor and lymph-node metastases to the bladder, which accumulates urine containing FDG in spite of indwelling catheters and irrigation, is one of the problems. Urinary bladder tumors also suffer from the same problem. However, lymph-node and distant metastases can be detected with greater accuracy.

Melanoma

FDG-PET using whole-body imaging is ideally suited to the patient with melanoma. It allows extensive screening in a minimum of time with high sensitivity and specificity. Numerous series have shown PET to have extremely high sensitivity, specificity, and accuracy. When compared to CT, FDG-PET shows superior sensitivity and specificity for lesion detection, staging, and general screening of the patient with malignant melanoma. Many studies conclude that a single, whole-body, FDG-PET study could replace the standard CT staging tests at significant cost saving in the patient with melanoma.

Further reading

Abdel-Nabi H *et al.* Staging of primary colorectal carcinomas with fluorine-18 fluorodeoxyglucose whole-body PET: correlation with histopathologic and CT findings. *Radiology* 1998; **206**: 755–60. [Results of whole-body FDG-PET imaging in 48 patients with biopsy-proven lesions, with correlation of both PET and CT information; FDG-PET had a high sensitivity and specificity for detecting colorectal carcinoma and was superior to CT in staging.]

Adler LP, Faulhaber PF, Schnur KC, Al-Kasi NL, Shenk RR. Axillary lymph node metastases: screening with [F-18]2-deoxy-2-fluoro-d-glucose (FDG) PET. *Radiology* 1997; **203**: 323–7. [Evaluation of FDG-PET as a screening test for axillary lymph-node involvement in breast cancer. FDG-PET had a high predictive value and can eliminate unneeded node dissections and subsequent morbidity as well as provide significant savings.]

Bergmann SR, Burton S, ed. *Positron emission tomography of the heart*. New York: Futura, 1992.

Brunken RC, Schelbert HR. Evaluation of myocardial substrate metabolism in ischemic heart disease. In: Skorton DJ, Schelbert HR, Wolf GL, Brundage BH, ed. *Marcus cardiac imaging*, pp. 1113–71. Saunders, Philadelphia, 1996. [Comprehensive review of the use of the PET techniques for assessing myocardial viability, including a meta-analysis of the prediction of functional left-ventricular recovery after surgery.]

Brury T *et al.* Whole body 18-FDG positron emission tomography in the staging of non-small cell lung cancer. *European Respiratory Journal* 1997; **10**(11): 2529–34. [109 patients underwent whole-body PET and conventional imaging methods for staging non-small-cell lung cancer. In 33 per cent of patients (22), PET correctly changed nodal staging and in 14 per cent (15 patients) it changed distant metastatic staging.]

Coleman RE, Hoffman JM, Hanson MW, Sostman HD, Schold SC. Clinical application of PET for the evaluation of brain tumors. *Journal of Nuclear Medicine* 1991; **32**: 616–22. [Description of the clinical utility of FDG-PET imaging in brain tumors. Review of the literature with specific examples and discussion of FDG-PET as a prognostic tool.]

Cronin V, Gelantowicz P, Nabi HA. Development of oncology protocol using fluorine 18 FDG: one center's experience. *Journal of Nuclear Medicine Technology* 1997; **25**: 66–9. [Makes recommendations about tomographic quality control, patient preparation, acquisition parameters, bed positioning, and processing of the information qualitatively and quantitatively in an evaluation of 560 patients with various types of cancer.]

Delbeke B, Vitola JB, Sandler MP. Staging recurrent colorectal carcinoma with PET. *Journal of Nuclear Medicine* 1997; **38**: 1196–201. [Prospective assessment of PET, CT, and CT portography in the evaluation of recurrent colorectal cancer before reintervention: in a significant number of patients (around 30 per cent) PET certainly added new information and changed the management, demonstrating the superior advantage that PET has over conventional imaging in the assessment of recurrent colorectal cancers.]

Dewan NA *et al.* PET-FDG imaging and transthoracic needle lung aspiration biopsy in evaluation of pulmonary lesions. A comparative risk benefit analysis. *Chest* 1995; **108**: 441–6. [A study evaluating 35 lung lesions using PET- and CT-based transthoracic fine-needle aspiration biopsy. The overall predictive accuracy was 94 per cent for PET and 86 per cent for needle aspiration. However, aspiration biopsy resulted in pneumothorax in 16 patients needing chest-tube aspiration in nine (26 per cent). This study certainly shows the advantage of doing non-invasive procedures such as PET in a diagnostic work-up.]

Gambhir SS *et al.* Decision tree sensitivity analysis for cost effectiveness of FDG PET in the staging and management of non-small lung cell carcinoma. *Journal of Nuclear Medicine* 1996; 1428–36. [In this cost–effect analysis, two different decision trees were used. In the conservative strategy, CT vs CT plus PET was used in conjunction with mediastinoscopy and biopsy for making a decision on thoracotomy. A less conservative strategy excluded mediastinoscopic biopsy and compared CT with CT plus PET leading to thoracotomy. The outcome is described above.]

Garcia R *et al.* Comparison of FDG PET and technetium-99m MIBI SPECT in evaluation of musculoskeletal carcinomas. *Journal of Nuclear Medicine* 1996; **37**: 1476–9. [48 patients with clinically suspected recurrent residual musculoskeletal carcinomas were examined with FDG-PET and MIBI-SPECT studies within 2 weeks of each other; a positive FDG but negative MIBI scan might suggest multidrug resistance.]Geltman EM. Assessment of myocardial perfusion and metabolism in the cardiomyopathies. In: Skorton DJ, Schelbert HR, Wolf GL, Brundage BH, ed. *Marcus cardiac imaging*, pp. 1171–85. Saunders, Philadelphia, 1996. [Excellent review of the technique for assessing ischemic vs non-ischemic cardiomyopathy.]

Gupta NC *et al.* Solitary pulmonary nodules: detection of malignancy with PET with 18F-deoxyglucose. *Radiology* 1992; **184**: 441–4. [One of the earlier US experiences in using 18-FDG PET studies to assess solitary pulmonary nodules, showing that 18-FDG PET images can distinguish benign from malignant nodules with great accuracy.]

Henry TR, Chugani HT, Abou-Khalil, Theodore WH, Swartz BE. Positron emission tomography. In: Engle J Jr, ed. *Surgical treatment of the epilepsies*, 2nd edn. Raven, New York, 1993. [A comprehensive chapter reviewing the indications and utility of FDG-PET imaging in the evaluation of the patient with partial-complex epilepsy.]

Hubner KF *et al.* Assessment of primary and metastatic ovarian cancer by positron emission tomography using 18-FDG. *Gynecologic Oncology* 1993; **51**: 197–204. [A study of 51 patients who had 18-FDG PET images prior to laparotomy for suspected ovarian cancer. The positive predictive value for PET was 86 per cent and negative predictive value was 76 per cent.]

Larson SM, Cohen AM, Cascade MB. Clinical applications and economical implications of PET in the assessment of colorectal cancer recurrence: a retrospective study. [Abstract]. Proceedings 1994 ICP Meeting, Institutes of Clinical PET, Fairfax, VA.

Leisure GP, Vesselle HJ, Faulhaber PF, O'Donnell JK, Adler LP, Miraldi F. Technical improvements in fluorine 18 FDG PET imaging of the abdomen and pelvis: *Journal of Nuclear Medicine Technology* 1997; **25**: 115–19. [130 patients who had abdominal and pelvic 18 FDG-PET study are analyzed for the quality of images; hydration, administration of furosemide along with a placement of a Foley catheter in the bladder eliminated most of the imaging artefacts originating from the kidneys, ureters, and bladder.]

Lowe VJ *et al.* Pulmonary abnormalities and PET data analysis: a retrospective study. *Radiology* 1997; **202**: 435–9. [Assessment of various methods of assessing FDG-PET images, including visual, qualitative, and semiquantitative determinations in pulmonary abnormalities.]

Minn H *et al.* Prediction of survival with fluorine-18-fluoro-deoxyglucose and PET in head and neck cancer. *Journal of Nuclear Medicine* 1997; **38**: 1907–11. [Thirty-seven untreated patients were evaluated with FDG-PET; a high uptake of FDG in untreated lesions was associated with advanced disease and poor survival.]

Moog F *et al.* Lymphoma: role of whole-body 2-deoxy-2-[F-18]fluoro-d-glucose (FDG) PET in nodal staging. *Radiology* 1997; **203**: 795–800. [Compares FDG-PET and CT in nodal staging of lymphoma.]

Phelps ME, Mazziotta JC, Schelbert HR, ed. *Positron emission tomography and autoradiography*. Raven, New York, 1986.

Rigo P. *et al.* Oncological application of positron emission tomography with fluorine 18 fluorodeoxyglucose. *European Journal of Nuclear Medicine* 1996; **23**: 1641–74. [A comprehensive review of biological characteristics of 18-FDG, methodological considerations and also a detailed survey of various clinical indications for FDG in oncology giving almost all major reference articles published on 18-FDG-PET in tumor detection until late 1996.]

Rinne D, Baum RP, Hor G, Kaufmann, R. Primary staging and follow-up of high risk melanoma patients with whole-body 18F-fluorodeoxyglucose positron emission tomography. Results of a prospective study of 100 patients. *Cancer* 1998; **82**: 1664–71. [Study of 100 consecutive patients with initially diagnosed melanoma, greater than 1.5 mm in thickness.]

Schelbert HR, Demer LL. Evaluation of myocardial blood flow in cardiac disease. In: Skorton DJ, Schelbert HR, Wolf GL, Brundage BH, ed. *Marcus cardiac imaging*, pp. 1093–112. Saunders, Philadelphia, 1996. [Comprehensive review of the PET techniques for diagnosing coronary arterial disease, including a meta-analysis.]

Valk PE *et al.* Staging lung cancer by PET imaging. *Annals of Thoracic Surgery* 1995; **60**: 1573–82. [A study of 99 patients where FDG-PET altered the staging of lung cancer when compared to CT results. Suspected distant metastases were noted, and it was concluded that FDG-PET was superior to CT in staging of lung cancer.]

Wahl RL. Emerging applications of PET in oncology: melanoma, lymphoma and prostate cancer. *Proceedings* 6th Annual International PET Conference. Institute for Clinical PET, Fairfax, VA, 1994. [A review of use of FDG-PET in prostate cancer.]

13.10 Imaging guidelines

Royal College of Radiologists, London*

Introduction
Minimizing radiation dosage
Abbreviations

Introduction

With the development of numerous imaging techniques such as ultrasonography, computed tomography, magnetic resonance imaging, and positron-emission tomography in addition to conventional radiography, contrast radiography, and vascular radiology it is important that clinicians have guidance on their appropriateness and merits. A useful investigation is one in which the result, positive or negative, will alter patient management or add confidence to the clinician's diagnosis. Each unnecessary investigation wastes resources, may add to the irradiation borne by the patient, increases waiting times, and lowers standards. The chief causes of wasteful use of radiological investigations are:

1. imaging when results are unlikely to affect patient management because the anticipated ‘positive' finding is usually irrelevant (e.g. degenerative spinal disease) or because a positive finding is so unlikely;
2. investigating too often, that is, before the disease could have progressed or resolved or before the results influence treatment;
3. imaging when it has already been done, for example at another hospital, in outpatient or accident and emergency clinics;
4. failing to provide adequate clinical information and explain the purpose of the examination, so that the wrong technique may be used;
5. doing the wrong investigation—imaging techniques change all the time.
6. over-investigating.

A guideline is ‘not a rigid constraint on clinical practice, but a concept of good practice against which the needs of the individual patient can be considered'. So, while there have to be good reasons for ignoring guidelines, they should not be regarded as absolute rules. No set of recommendations will command universal support and any problems should be discussed with your own radiologists. The guidelines discussed here reflect the practice in the United Kingdom. They are taken from the fourth edition of *Making the best use of a department of clinical radiology—guidelines for doctors*, published by the Royal College of Radiologists and reprinted with their permission. Copies of the Royal College of Radiologists' guidelines can be obtained from the Royal College of Radiologists, 38 Portland Place, London, W1N 3DG, UK.

The Royal College of Radiologists has taken note of the ‘Appropriateness criteria' issued by the American College of Radiologists. That College lists all possible investigations and awards an appropriateness score between 1 and 9 to each.

Minimizing radiation dosage

A small fraction of the genetic mutations and malignant disease occurring in the population can be attributed to natural background radiation. Diagnostic medical exposures, being the major source of man-made radiation exposure of the population, add about one-sixth to the population dose from background radiation.

Typical effective doses for some common diagnostic radiological procedures vary very significantly. They range over a factor of about 1000 from the equivalent of a day or two of natural background radiation (0.02 mSv for a chest radiograph) to 4.5 years (10 mSv for computed tomography of the abdomen). High-dose examinations such as whole-body computed tomography and barium studies make the major contribution to the collective population dose. It is thus particularly important that requests for these examinations be thoroughly justified and that techniques be adopted that minimize the dose while retaining essential diagnostic information.

Using the guidelines

In general these guidelines only deal with areas of difficulty or controversy. Straightforward indications for examinations are not discussed. A system-based style has been adopted, which provides ready access for most imaging strategies. Trauma, cancer, and paediatric imaging, which cross specialist boundaries, have their own sections. On most pages there are four columns: The first sets out the clinical situation for requesting the examination, the next lists the possible imaging techniques, the third gives the recommendation on whether or not the investigation is appropriate, and the fourth provides explanatory comments.

The recommendations used are:

1. **Indicated**: This shows the investigation(s) most likely to contribute to clinical diagnosis and management. This may differ from the investigation requested by the clinician: for example, ultrasound rather than venography for deep vein thrombosis.
2. **Not indicated initially**: This includes situations where experience shows that the clinical problem usually resolves with time. It is therefore suggested that the study be deferred for 3 to 6 weeks and only performed if symptoms continue. Shoulder pain is a typical example.
3. **Not indicated routinely**: This emphasizes that while no recommendation is absolute, the request will only be carried out if a clinician gives cogent arguments for it. An example of such a justification would be plain radiography in a patient with backache in whom there were clinical findings to suggest something more than degenerative disease (for example, a possible osteoporotic vertebral fracture).
4. **Not indicated**: Examinations in this group are those where the supposed rationale for the investigation is untenable (for example, intravenous urography for hypertension).
5. **Specialized investigation**: These are complex or expensive investigations that usually will be performed only for doctors who have the relevant clinical expertise to evaluate the clinical findings and act on the imaging results. They often justify individual discussion with a senior radiologist.

The strength of evidence for the various statements and recommendations is indicated by:

[**A**]randomized controlled trials (**RCT**s), meta-analyses, systematic reviews
or
[**B**]robust experimental or observational studies,
or
[**C**]other evidence where the advice relies on expert opinion and has the endorsement of respected authorities.

In some clinical situations (for example, the role of ultrasound in normal pregnancy) there are conflicting data within a large body of excellent scientific reports. Thus no firm recommendations are given and the evidence is classified as [C]. It should also be noted that there are very few randomized trials comparing different radiological diagnostic procedures—they are difficult to perform and ethical approval may be denied.

The guidelines are divided into sections as follows:

Section A: Head (including ENT problems)
Section B: Neck [for spine, see sections C (spine) and K (trauma)]
Section C: Spine—general (for trauma, see section K)
Section D: Musculoskeletal system
Section E: Cardiovascular system
Section F: Thoracic system
Section G: Gastrointestinal system (Part 1)
Section G: Gastrointestinal system (Part 2)
Section H: Urological, adrenal, and genitourinary systems
Section I: Obstetrics and gynaecology
Section J: Breast disease
Section K: Trauma (Part 1)
Section K: Trauma (Part 2)

Section K: Trauma (Part 3)
Section L: Cancer (Part 1)
Section L: Cancer (Part 2)
Section M: Paediatrics (Part 1) [for head injury in children, see Trauma (section K)].
Section M: Paediatrics (Part 2)

Abbreviations

MRI = magnetic resonance imaging; CT = computed tomography; NM = nuclear medicine; XR = radiograph; SXR = skull radiograph; US = ultrasound; Ba = barium; AXR = abdominal radiograph; ERCP = endoscopic retrograde cholangiopancreatogram; MRCP = magnetic resonance cholangiopancreatogram; IVU = intravenous urogram; DSA = digital subtraction angiogram.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
A. Head (Including Otolaryngology Problems)			
Congenital disorders ***(For Children see Section M)***	MRI	Indicated [C]	Definitive exam for all malformations & avoids x-irradiation. 3D CT may be needed for bone anomalies. Sedation usually required for young children. Consider US in neonates.
A1			
Cerebrovascular accident (CVA); stroke	CT	Indicated [C]	CT adequately assesses most cases and shows haemorrhage.
	MRI	Specialised [B]investigation	MRI more sensitive in early infarction & for posterior fossa lesions. Some centres use NM (SPECT).
A2	US carotids	Not [C] routinely	Exceptions for: a) those with full recovery in whom carotid surgery is contemplated. b) an evolving CVA where dissection or embolus suspected.
Transient ischaemic attack (TIA) ***(See also B5)***	US carotids	Indicated [B]	If doubt about diagnosis or surgery contemplated. Much depends on local policy & available expertise. US (with colour Doppler) provides functional data about bifurcation disease. Angiography, MRA and CTA are more expensive alternatives.
A3			
Demyelinating and other white matter disease	MRI	Specialised [A] investigation	MRI much more sensitive than CT for demyelinating disease. But MRI may still be negative in up to 25% of those with established multiple sclerosis.
A4	MRI	Specialised [B] investigation	MRI superior to CT in delineating extent and location of other white matter disease.
?Space occupying lesion (SOL)	CT or MRI	Indicated [B]	MRI more sensitive for early tumours, in resolving exact position (useful for surgery) & for posterior fossa lesions. MRI may miss calcification. CT more widely available; and often sufficient in supratentorial lesions and subdural haematomas. MRI superior in the posterior fossa and for vascular lesions.
A5			
Headache: acute, severe	CT	Indicated [B]	CT provides adequate data in most cases of subarachnoid and other intracranial haemorrhage and associated hydrocephalus. N.B. A negative CT does not exclude SAH and where suspected lumbar puncture should follow, assuming no contraindications (e.g. obstructive hydrocephalus). Lumbar puncture may also be needed to exclude meningitis.
A6	MRI or NM	Specialised [C] investigation	MRI better than CT for inflammatory causes. SPECT may be the most sensitive investigation for encephalitis.
Headache: chronic ***(For Children see Section M)***	XR Skull, Sinus, C Spine,	Not [B] routinely	Radiography of little use in the absence of focal signs/symptoms. See A13 below.
A7	CT	Not [B] routinely	Some exceptions for specialists or if evidence of raised intracranial pressure, posterior fossa or other signs.
Pituitary & juxta-sellar problems	MRI	Specialised [B] investigation	Demonstration of microadenomas may not be helpful for management. CT if MRI not available. Urgent referral when vision deteriorating.
A8	SXR	Not [C]	Patients who require investigation need MRI or CT.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Posterior fossa signs	MRI	Indicated [A]	MRI much better than CT. CT images often degraded by beam hardening artefacts.
A9			
Hydrocephalus ?shunt function ***(For Children see Section M)***	CT	Indicated [B]	CT adequate for most cases; MRI sometimes necessary and may be more appropriate in children. US first choice for infants. NM used in some centres (injection into shunt).
A10	XR	Indicated [C]	XR should include whole valve system.
Middle or inner ear symptoms (including vertigo)	CT	Specialised [B] investigation	Evaluation of these symptoms requires ENT, neurological or neurosurgical expertise.
A11			
Sensorineural deafness ***For Children see Section M)***	MRI	Specialised [B] investigation	MRI much better than CT, especially for acoustic neuromas. For deafness in children see M4.
A12			
Sinus disease ***(For Children see Section M)***	Sinus XR	Not [B] routinely	Thickened mucosa is a non-specific finding and may occur in asymptomatic patients.
A13	CT	Specialised [B] investigation	CT is more rewarding and provides unique information about ostial anatomy. Low dose technique desirable. Indicated when maximal medical treatment has failed, when complications arise or if malignancy suspected.

Dementia & memory disorders, first onset psychosis	SXR	Not [B]	Consider investigation if clinical course unusual or in younger patient.
A14	CT or MRI or NM	Specialised [B] investigation	CT and SPECT a good combination for Alzheimer's disease. MRI better for structural changes and assessment of 'normal pressure hydrocephalus'. PET and SPECT readily provide functional data. Cerebral blood flow studies may differentiate Alzheimer's from other forms of dementia.
Orbital lesions	CT or MRI	Specialised [B] investigation	CT provides better anatomical detail, particularly of bony structures (e.g. nasolacrimal duct). MRI avoids radiation dose to lens (but contraindicated when ferromagnetic FB suspected). Consider US for intra-ocular lesions.
A15			
Orbits ?metallic FB (before MRI)	XR orbits	Indicated [B]	Especially for those who have worked with metallic materials, power tools, etc. See Trauma Section K for acute injury.
A16			
Visual disturbances	SXR	Not [C]	Plain XRs rarely contributory. Specialists may require CT or MRI.
A17			
Epilepsy (adult) ***(For Children see Section M)***	SXR	Not [B]	Evaluation requires specialist expertise. Late onset seizures should normally be investigated but imaging may be unnecessary if clearly alcohol related.
A18	CT, MRI or NM	Specialised [B] investigation	Partial/focal seizures may require detailed evaluation if surgery is being considered. Ictal SPECT maximises likelihood of localising focus.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
B. NECK *(For The Spine See Sections C [The Spine] & K [Trauma])*			
Soft Tissues Thyroid nodules	US	Indicated [C]	Demonstrates morphology; allows guided aspiration for cytology or biopsy for histology. Some clinicians will proceed to aspiration with no imaging. NM may be appropriate if biopsy contraindicated.
B1			
Thyrotoxicosis	NM	Specialised [C] investigation	Can differentiate between Graves' disease, toxic nodular goitre and subacute thyroiditis. Provides functional information about nodules.
B2			
?Ectopic thyroid tissue (e.g. lingual thyroid)	NM	Indicated [C]	NM excellent for small ectopic rests of thyroid tissue. In generalised thyroid enlargement or multinodular goitre US readily shows retrosternal extension; real time studies show effect of neck extension, etc. CT/MRI needed to demonstrate full retrosternal extent & tracheal compromise.
B3			
Hyperparathyroidism	Imaging	Specialised [C] investigation	Seek advice. Diagnosis made on clinical/biochemical grounds. Imaging can assist in pre-operative localisation but may not be needed by experienced surgeons. Much depends on local policy and available technology and expertise. US, NM, CT and MRI all accurate in the unoperated neck. MRI probably evolving as the best investigation for ectopic and residual tumours.
B4			
Asymptomatic carotid bruit	US carotids	Not [B] routinely	Significant internal carotid artery lesions are rarely found.
B5			
Swallowed or inhaled foreign body (FB)			See Trauma K 30.
B6			
Mass of unknown origin	US	Indicated [C]	US first line investigation which can also direct biopsy. MRI (or CT) usually only if recommended by radiological or specialist clinical opinion.
B7			
Salivary obstruction	US or Sialogram	Indicated [C]	For intermittent, food related swelling. MR Sialography may be preferred in some centres.
B8	XR	Not [C] routinely	Except in ?calculus in floor of mouth, where XR may be all that is required.
Salivary mass	US	Indicated [B]	US extremely sensitive and, dependent on local expertise, should be first line investigation. MRI excellent for extensive or recurrent disease. CT now of limited use. No indication for CT Sialography.
B9			
Dry mouth ?connective tissue disease	US or Sialogram or NM	Specialised [C] investigation	Not commonly required. Sialogram may be diagnostic but NM provides better functional assessment. MR sialography also used here.
B10			
Temporo-mandibular joint dysfunction	XR	Specialised [B] investigation	Radiographs will demonstrate bony abnormalities, but are normal in great majority, as problems are usually related to articular disk dysfunction.
B11	MRI or arthrography	Specialised [B] investigation	Following failure of conservative treatment when internal derangement suspected. Arthrography offers a true dynamic demonstration.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
C. The Spine ***General (For Trauma See Section K)***			
Congenital disorders, ***(For Children see Section M)***	XR	Specialised [C] investigation	e.g. Full length standing radiograph for scoliosis. See Section M for back pain (M10).

C1	MRI	Specialised [B] investigation	MRI defines all spinal malformations and excludes associated thecal abnormality. CT may be required to delineate bony detail, but remember large radiation burden.
Myelopathy: tumours, inflammation, infection, infarction, etc.	MRI	Indicated [B]	MRI clear first choice for all spinal cord lesions & to evaluate cord compression. CT may be needed if better bony detail is required. Myelography only if MRI is unavailable or impossible. NM still widely used to screen for metastases & for identifying focal skeletal lesions (such as osteoid osteoma).
C2			
Cervical Spine Possible atlanto-axial subluxation	XR	Indicated [C]	A single lateral cervical spine XR with the patient in supervised comfortable flexion should reveal any significant subluxation in patients with rheumatoid arthritis, Down's Syndrome, etc. MRI needed to show effect on cord when XR positive or neurological signs present.
C3			
Neck pain, Brachalgia, ?degenerative change	XR	Not [B] routinely	Degenerative changes begin in early middle-age and are often unrelated to symptoms which are usually due to disk/ligamentous changes undetectable on plain XR. MRI increasingly being used, especially when brachalgia is present.
C4	MRI	Specialised [B] investigation	Consider MRI and specialist referral when pain affecting lifestyle or when there are neurological signs. Myelography (with CT) may occasionally be required to provide further delineation or when MRI is unavailable or impossible.
Thoracic Spine Pain without trauma: ?degenerative disease	XR	Not [B] routinely	Degenerative changes are invariable from middle-age onwards. Examination rarely useful in the absence of neurological signs or pointers to metastases or infection. Consider more urgent referral in elderly patients with sudden pain to show osteoporotic collapse or other forms of bone destruction. Consider NM for possible metastatic lesions.
C5	MRI	Specialised [B] investigation	MRI may be indicated if local pain persists, difficult to manage or if there are long tract signs.
Lumbar Spine Chronic back pain with no pointers to infection or neoplasm	XR	Not [C] routinely	Degenerative changes are common and non-specific. Main value in younger patients (e.g. less than 20, spondylolisthesis, ankylosing spondylitis, etc.) or in older patients e.g.>55.
C6	MRI or CT or NM	Specialised [C] investigation	In exceptional cases where management is difficult. Negative findings may be helpful.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Back pain with possible serious features such as: • Onset < 20, > 55 yrs • Sphincter or gait disturbance • Saddle anaesthesia • Severe or progressive motor loss • Widespread neurological deficit • Previous carcinoma • Systematically unwell • HIV • Weight loss • Intravenous drug abuse • Steroids • Structural deformity • Non-mechanical pain	Imaging	Indicated [C]	Together with urgent specialist referral; MRI is usually the best investigation. Imaging should not delay specialist referral. NM is also widely used for possible bone destruction, and in cases of chronic pain or where infection is suspected. ('NORMAL' PLAIN XR MAY BE FALSELY REASSURING). (For Children see Section M)
C7			
Acute back pain: ?disk herniation; sciatica with no adverse features (see above).	XR	Not [C] routinely	Acute back pain is usually due to conditions which cannot be diagnosed on plain XR (osteoporotic collapse an exception). 'Normal' plain XRs may be falsely reassuring. Demonstration of disk herniation requires MRI or CT and should be considered immediately after failed conservative management.
C8	MRI or CT	Not [B] initially	MRI generally preferred (wider field of view, conus, post-operative changes etc.) and avoids x-irradiation. Either MRI or CT is needed before intervention (e.g. epidural injection). MRI better than CT for post-operative problems.
D. Musculoskeletal System			
Osteomyelitis	XR + NM	Indicated [B]	NM sensitive but non-specific. XR can be normal for first 2-3 weeks. A labelled white cell study may distinguish infection from other lesions.
D1	MRI or CT or US	Specialised [C] investigations	MRI or CT used to establish extent of bony and soft tissue disease and may identify sequestra. Both can demonstrate appropriate site for percutaneous biopsy. US may be helpful, especially in children, if metalware causes artefacts on MRI/CT or if NM non-specific due to recent surgery.
?Primary bone tumour	XR	Indicated [B]	XR may characterise the lesion.
D2	MRI or CT	Specialised [B] investigations	MRI useful for further characterisation and necessary for surgical staging; should be performed before any biopsy. CT can show bony detail better at some sites (e.g. spine) and for some small lesions and is needed if MRI unavailable. MRI more useful for assessment of extent. CT chest if CXR negative to assess pulmonary metastases for many primary malignant lesions. (see L41).
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Known primary tumour ?Skeletal metastases	NM Skeletal survey	Indicated [B] Not [C]	NM readily assesses the whole skeleton and is much more sensitive than plain XR, though less specific. Localised XRs may be needed to exclude other causes of increased activity, e.g. degenerative disease. In prostatic cancer biochemical markers (PSA) can be used to follow up progress of skeletal involvement.

D3	MRI	Specialised [C] investigation	MRI more sensitive and specific than NM, especially for marrow based lesions. However field of view limited.
Soft tissue mass ?tumour Possible recurrence	MRI	Indicated [B]	MRI better than CT for exclusion, detection and staging of soft tissue tumours (superior contrast resolution, multiplanar capability, delineation of neurovascular bundle and muscle/compartment involvement). CT has greater sensitivity for calcification. Increasing interest in US for some anatomical sites.
D4	MRI	Indicated [B]	MR accepted as investigation of choice although US has its proponents and can be used for biopsy.
Bone pain	XR	Indicated [B]	Local view of symptomatic areas only.
D5	NM	Indicated [C]	When symptoms persist and plain XRs negative. MRI used in some centres.
?Myeloma	Skeletal survey	Indicated [C]	For staging and identifying lesions which may benefit from radiotherapy. Survey can be very limited for follow-up.
	NM	Not [B] routinely	Scintigraphy underestimates disease extent.
D6	MRI	Specialised [B] investigation	MRI very sensitive, even limited to spine, pelvis and proximal femora. Particularly useful in non-secretory myeloma or in the presence of diffuse osteopenia. Can be used for tumour mass assessment and follow-up.
Metabolic bone disease	Skeletal survey	Not indicated routinely [C]	Biochemical tests usually suffice. If needed, this should be limited (e.g. hands, CXR, pelvis and lateral lumbar spine). Bone densitometry may be needed. (see D9).
D7			
?Osteomalacia	XR	Indicated [B]	Localised XR to establish cause of local pain or equivocal lesion on NM.
D8	NM	Specialised [C] investigation	NM can show increased ‘activity' and some local complications. Bone densitometry may be needed. (see D9).
Pain ?osteoporotic collapse	XR lateral thoracic and lumbar spine	Indicated [B]	Lateral views will demonstrate compression fractures. NM or MRI more useful in distinguishing between recent and old fractures and can help exclude pathological fractures. Bone densitometry (Dual energy XR apsorptiometry [DEXA] or Quantitative CT) provides objective measurements of bone mineral content; can also be used for metabolic bone disease (see D7, D8).
D9			
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Arthropathy, presentation	XR affected joint	Indicated [C]	May be helpful to determine cause although erosions are a relatively late feature.
	XR hands/feet	Indicated [C]	In patients with suspected rheumatoid arthritis, XR feet may show erosions even when symptomatic hand(s) appear normal.
	XR multiple joint(s)	Not [C] routinely	
D10	US or NM or MRI	Specialised [C] investigations	All accurately show acute synovitis. NM can show distribution.
Arthropathy, follow-up	XR	Not [C] routinely	XRs needed by specialists to assist management decisions.
D11			
Painful shoulder	XR	Not [C] initially	Degenerative changes in the acromio-clavicular joints and rotator cuff are common. Earlier XR if soft tissue calcification is expected.
D12			
Painful prosthesis	XR + NM	Indicated [B]	A normal NM study excludes most late complications. A labelled white cell or gallium study can help distinguish loosening from infection.
D13	US/Fluoroscopy	Specialised [C] investigation	Usually coupled with aspiration/biopsy/arthrography. Such intervention which provides a definitive result is increasingly being used.
Shoulder impingement	MRI	Specialised [B] investigation	Although impingement is a clinical diagnosis, imaging is indicated when surgery is being considered and precise delineation of anatomy is required. But degenerative changes also common in the asymptomatic population.
D14	US	Specialised [B] investigation	Subacromial and acromioclavicular joint impingement are dynamic processes which can be assessed by US.
Shoulder instability	CT arthrography MR arthrography	Specialised [B] investigation	Glenoid labrum and synovial cavity are well delineated by both techniques. Some gradient echo MR techniques can show labrum well without arthrography.
D15			
Rotator cuff tear	Arthrography or US or MRI	Specialised [B] investigation	Much depends on local expertise and surgical plans. All three techniques demonstrate rotator cuff tears.
D16			
?SI joint lesion	XR SI joints	Indicated [B]	May help in investigation of sero-negative arthropathy. SI joints usually adequately demonstrated on AP lumbar spine.
D17	MRI or NM or CT	Specialised [C] investigation	MRI or NM or CT when plain XRs equivocal; MRI carries no radiation dose.
Hip pain: full movement ***(For Children see Section M)***	XR Pelvis	Not [C] routinely	XR only if symptoms and signs persist or complex history (e.g. chance of avascular necrosis, see D20) **N.B. This recommendation does not apply to children.**
D18			
Hip pain: limited movement ***(For Children see Section M)***	XR Pelvis	Not [C] initially	Symptoms often transient. XR if hip replacement might be considered. **N.B. This recommendation does not apply to children.**
D19			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Hip pain: ?avascular necrosis	XR Pelvis	Indicated [B]	Abnormal in established disease.
D20	MRI	Specialised [B] investigation	MRI useful when XR normal, especially in high risk patients. NM and CT can also provide information here.
Knee pain: without locking or restriction in movement	XR	Not [C] routinely	Symptoms frequently arise from soft tissues and these will not be demonstrated on XR. OA changes common. XRs needed when considering surgery.
D21			
Knee pain: with locking, restricted movement or effusion (?loose body)	XR	Indicated [C]	To identify radio-opaque loose bodies.
D22			
Knee pain: arthroscopy being considered	MRI	Specialised [B] investigation	MRI can assist the management decision as to whether or not to proceed with arthroscopy. Even in those patients with definite clinical abnormalities, warranting intervention, surgeons find pre-operative MRI helpful in identifying unsuspected lesions.
D23			
Hallux valgus	XR	Specialised [C] investigation	For assessment before surgery.
D24			
?Plantar fasciitis ?calcaneal spur	XR	Not [B] routinely	Plantar spurs are common incidental findings. The cause of the pain is seldom detectable on XR. US, NM and MRI are more sensitive in showing inflammatory change but the majority of patients can be managed without imaging.
D25			
E. Cardiovascular System			
Central chest pain: ?myocardial infarction	CXR	Indicated [B]	CXR must not delay admission to a specialised unit. CXR can assess heart size, pulmonary oedema, etc. and can exclude other causes. Department film preferable. Subsequent imaging involves specialised investigations (NM, coronary angiography, etc.) and depend on local policy. NM offers myocardial perfusion and ventriculography data. Increasing interest in MRI.
E1			
Chest pain: ?aortic dissection: acute	CXR	Indicated [B]	Mainly to exclude other causes; rarely diagnostic.
E2	CT or US or MRI	Indicated [B]	Seek advice from local radiologists. Much variation. Modern CT systems provide very accurate results. Often coupled with transthoracic US or, better, transoesophageal US. MRI probably the most accurate and increasingly used, despite logistic problems and constraints with some life support systems. Angiography rarely necessary unless above examinations are equivocal.
Aortic dissection: chronic	MRI	Specialised [B] investigation	MRI best investigation to assess change in longitudinal extent. Transoesophageal US and CT recommended.
E3			
CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
?Pulmonary embolus	NM	Indicated [B]	Interpreted along with contemporary CXR. Equivocal findings (e.g. intermediate probability) may necessitate further clarification. Some centres use US to show thrombus in leg veins for further proof.
E4	CT	Specialised [B] investigation	Spiral CT used increasingly as the initial test, especially in patients with co-existing cardiorespiratory disease, and ahead of pulmonary angiography.
?Pericarditis ?pericardial effusion	CXR	Indicated [B]	May be normal; effusion volume/effect not determined.
E5	US	Indicated [B]	Extremely accurate: may be needed urgently for ?tamponade; can show best access for drainage. CT sometimes needed for calcification, loculation, etc.
Suspected valvular cardiac disease	CXR and cardiac US	Indicated [B]	Used for initial assessment and when there is a change in the clinical picture.
E6			
Clinical deterioration following myocardial infarction	Cardiac US	Indicated [B]	US may show remediable complications (VSD, papillary rupture, aneurysm, etc.).
E7			
Follow-up of patients with heart disease or hypertension	CXR	Not [B] routinely	Only if signs or symptoms have changed, when comparison with the CXR obtained at presentation may be helpful.
E8			
?Abdominal aortic aneurysm	US aorta	Indicated [A]	Useful in diagnosis, determination of maximal diameter and follow-up. CT preferable for suspected leak but should not delay urgent surgery.
E9	CT or MRI	Indicated [A]	CT (especially spiral) and MRI for relationship to renal vessels and iliacs. Increasing demand for detailed anatomical information because of increasing consideration for percutaneous stenting.
?Deep vein thrombosis	US lower limb veins	Indicated [A]	More sensitive with colour flow Doppler. Most clinically significant thrombi are detected. Increasing experience with US for calf vein thrombi. May show other lesions.
E10	Venography	Not [C] routinely	Extensive variation according to US expertise and local therapeutic strategy.
Ischaemic leg	Angiography	Specialised [A] investigation	Local policy needs to be determined in agreement with vascular surgeons, especially with regard to therapeutic interventions. US used in some centres as first investigation. Spiral CT and MRI are being developed.
E11			
F. Thoracic System			

Non-specific chest pain **F1**	CXR	Not [C] initially	Conditions such as Tietze's disease show no abnormality on CXR. Main purpose is reassurance.
Chest trauma **F2**	CXR	Not [C] routinely	Showing a rib fracture does not alter management (see Trauma Section K).
Pre-employment or screening medicals **F3**	CXR	Not [B]	Not justified except in a few high-risk categories (e.g. at risk immigrants with no recent CXR). Some have to be done for occupational (e.g. divers) or emigration purposes (UK category 2).
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Pre-operative **F4**	CXR	Not [B] routinely	Exceptions before cardio-pulmonary surgery, likely admission to ITU, suspected malignancy or possible TB. Anaesthetists may also request CRXs for dyspnoeic patients, those with known cardiac disease and the very elderly. Many patients with cardio-respiratory disease have recent CXR available; a repeat CXR is then not usually needed.
Upper respiratory tract infection **F5**	CXR	Not [C]	
Chronic obstructive airways disease or asthma; follow-up **F6**	CXR	Not [B]	Only if signs or symptoms have changed.
Pneumonia adults: follow-up ***(For Children see Section M)*** **F7**	CXR	Indicated [A]	To confirm clearing, etc. Pointless to re-examine at less than 10-day intervals as clearing can be slow (especially in the elderly).
?Pleural effusion	CXR	Indicated [B]	Small effusion can be missed, especially on a frontal CXR.
F8	US	Indicated [B]	To prove fluid consistency; to guide aspiration. CT occasionally needed for better localisation, assessment of solid components, etc.
Haemoptysis	CXR	Indicated [B]	PA plus lateral view.
F9	CT	Specialised [B] investigation	Many centres use CT and then proceed to bronchoscopy; increasing use of CT first (see Cancer L7).
ITU/HDU patient **F10**	CXR	Indicated [B]	A CXR is most helpful when there has been a change in symptoms or insertion or removal of a device. The value of the routine daily CXR is being increasingly questioned.
?Occult lung disease **F11**	CT (HRCT)	Specialised [B] investigation	HRCT can show abnormalities not evident on CXR. Some interest in NM for sarcoid (gallium), alveolitis (permeability) and HIV (gallium).
G. Gastrointestinal System			
Gastrointestinal Tract Difficulty in swallowing **G1**	Ba swallow	Indicated [B]	Ba studies are still recommended before possible endoscopy; they will accurately localise lesions and show the degree of obstruction caused by a stricture and its length. Webs and pouches are well demonstrated. Subtle strictures may be demonstrated by a marshmallow (or other bolus) study. Detailed fluoroscopy or NM needed for motility disorders. Video swallows for suspected pharyngeal dysfunction in conjunction with speech therapists.
Chest pain ?hiatus hernia or reflux **G2**	Ba swallow/ meal	Not [C] routinely	Although Ba swallow useful to demonstrate hernia, reflux and their complications, not all such patients need investigation. Reflux is common and not necessarily the cause of pain. NM may be oversensitive; pH monitoring is generally regarded as the `gold standard' for acid reflux but gives no anatomical information. Metaplasia and oesophagitis are best detected by endoscopy which also allows biopsy. Increasing use of Ba studies before antireflux surgery.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
?Oesophageal pertoration	CXR	Indicated [B]	CXR may be sufficient, unless localisation for surgical repair is planned.
G3	Swallow	Specialised [B] investigation	Swallow should be performed with water-soluble nonionic contrast agents.
Acute GI bleeding: haematemesis	AXR	Not [B]	Of no value.
	Ba studies	Not [A]	Endoscopy provides diagnosis of upper GI lesions, allows injection of varices, etc. Ba studies preclude angiography.
	NM (red cell study)	Specialised [B] investigation	After endoscopy. NM can detect bleeding rates as low as 0.1 ml/min; more sensitive than angiography. Red cell study is most useful in intermittent bleeding.
G4	Angiography	Specialised [B] investigation	When considering surgery or intervention (e.g. embolisation) for uncontrollable bleeding.
Dyspepsia in the younger patient (e.g. under 45 yrs) **G5**	Imaging (Ba meal/endoscopy)	Not [C] routinely	Most patients under 45 yrs can be treated without complex investigations and will undergo a trial of therapy (anti-ulcer or reflux). Either Ba meal or endoscopy for those who fail to respond. Other alarm features pointing to early investigation include unintentional weight loss, anaemia, anorexia, GI bleeding, pain requiring hospitalisation, non-steroid anti-inflammatory drugs, vommiting, no improvement following treatment in those positive for *Helicobacter pylori*.
Dyspepsia in the older patient (e.g. over 45 yrs) **G6**	Imaging (Ba meal/endoscopy)	Indicated [C]	Endoscopy is often the first line investigation. However, Ba meal remains a reasonable alternative. The alternative investigation should be considered whenever symptoms continue after negative result. The main concern is the detection of early cancer, especially submucosal tumours.

Ulcer follow-up **G7**	Ba studies	Not [B] routinely	Scarring precludes accurate assessment. Endoscopy preferred to confirm complete healing and to obtain biopsies (e.g. *Helicobacter pylori*, etc.) where necessary. Some centres use NM studies (Carbon-14 breath test) to assess effect of treatment of Helicobacter pylori.
Previous upper GI surgery (recent) **G8**	Water soluble contrast medium study	Indicated [B]	To assess anastomosis and transit through to small bowel.
Previous upper GI surgery (old) **G9**	Ba studies	Not [B]	Gastric remnant best assessed by endoscopy (gastritis, ulceration, recurrent tumour, etc.). Cross-sectional imaging (US, CT, etc.) may be needed to assess extramural disease. Endoscopic US can demonstrate submucosal recurrence.
Intestinal blood loss, chronic or recurrent	Ba small bowel study	Not [C] initially	Only after upper and lower tract imaging (Ba studies or endoscopy).
G10	NM (red cell or Meckel's study) and/or angiography	Specialised [B] investigation	When all other investigations are negative.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Acute abdominal pain ?perforation ?obstruction **G11**	CXR (erect) and AXR	Indicated [B]	Decubitus AXR to show free air if CXR supine. Supine AXR usually sufficient to establish diagnosis and point to an anatomical level of obstruction. Consider erect AXR if supine AXR normal and strong clinical suspicion of obstruction. Consider CT for small sealed perforations and for establishing site and cause of obstruction.
Small bowel obstruction: **G12**	Contrast studies or CT	Specialised [B] investigation	Studies with non-ionic agents can determine both the site and completeness of obstruction. Some centres use CT in this situation which can determine level and likely cause.
Small bowel obstruction: chronic or recurrent **G13**	Small bowel Ba study	Indicated [B]	Small bowel enema is the examination of choice.
Small bowel disease suspected (e.g. Crohn's disease)	Ba small bowel study	Indicated [C]	Ba follow through tends to give a lower radiation dose than small bowel enema. Some centres use US to assess bowel wall.
G14	NM (white cell study)	Specialised [B] investigation	Labelled white cell scintigraphy reveals activity and extent of disease. Complementary to Ba studies. CT and MRI reserved for complications.
?Large bowel tumour or ?inflammatory bowel disease: pain, bleeding, change in bowel habit, etc. **G15**	Ba enema	Indicated [B]	N.B. Double contrast Ba is only useful if the bowel is properly prepared. Furthermore all patients should undergo rectal examination to assess suitability for Ba enema and to exclude a low rectal tumour. Good practice requires a sigmoidoscopy before Ba enema. Defer Ba enema for 7 days after full thickness biopsy via a rigid sigmoidoscope. Biopsies taken during flexible sigmoidoscopy are usually superficial and the risk of subsequent perforation is low (ideally delay 48 hrs). Some centres use colonoscopy initially, reserving Ba enema for difficult or incomplete examinations. Some centres use CT for the frail elderly patient. Although the irritable bowel syndrome is the commonest cause of a change in bowel habit, Ba enema or colonoscopy is needed to exclude other causes.
Large bowel obstruction: acute **G16**	Enema	Specialised [B] investigation	Single contrast (ideally water soluble contrast medium) study can show narrowed area and exclude ‘pseudo-obstruction'. Some centres use CT which can point to the likely cause.
Inflammatory bowel disease of colon: acute exacerbation	AXR	Indicated [B]	Often sufficient for evaluation.
	NM (white cell study)	Indicated [B]	Labelled white cell study best exam–will reveal activity and extent of disease.
G17	Ba enema	Not [B] routinely	Ba enema is dangerous when toxic megacolon present; unprepared enema in selected cases after discussion with radiologists.
Inflammatory bowel disease of colon: long-term follow-up **G18**	Ba enema	Not [B] routinely	Colonoscopy follow-up preferred to identify developing carcinoma in those at high risk, although Ba enema is still often used, particularly after complex intestinal surgery. Likewise Ba enema preferred for evaluating fistulae etc.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
General Abdominal Problems Acute abdo pain; (warranting hospital admission and surgical consideration) **G19**	AXR plus erect CXR	Indicated [B]	Local policy will determine strategy. Supine AXR (for gas pattern, etc.) is usually sufficient. Erect AXR not indicated routinely. Increasing use of CT as a `catch-all' investigation here.
Palpable mass	AXR	Not [C]	
	US	Indicated [B]	US usually solves the problem and is very reliable in thin patients, right upper quadrant and pelvis.
G20	CT	Indicated [A]	CT is an alternative and useful to exclude a lesion; particularly good in obese patients.
Malabsorbtion	Ba study of small bowel	Not [B] routinely	Imaging is not required for the diagnosis of coeliac disease but may be indicated for jejunal diverticulosis or when biopsy is normal/equivocal. CT may be better if lymphoma suspected.
G21	NM	Specialised [B] investigation	Numerous NM investigations available which should establish presence of malabsorption. Some of these are non-radiological (e.g. breath test).
?Appendicitis **G22**	Imaging	Not [C] routinely	Appendicitis is usually a clinical diagnosis. Imaging (e.g. US with graded compression) can help in equivocal cases or in differentiation from gynaecological lesions. So too can NM (white cell study) and focussed appendix CT (FACT). US recommended in children and young women.

?Constipation ***(For Children see Section M)***	AXR	Not [C] routinely	Many normal adults show extensive fascal material; although this may be related to prolonged transit time it is impossible to assess significance on AXR alone. But AXR can help certain specialists (e.g. geriatricians) in refractory cases.
G23			
?Abdominal sepsis; pyrexia of unknown origin (PUO)	US or CT or NM	Indicated [C]	Seek radiological advice; much depends on local availability and expertise. US often used first (speed, cost) and may be definitive, particularly when there are localising sign; especially good for subphrenic/subhepatic spaces and pelvis. CT probably best test overall: infection and tumour usually identified and excluded. Also allows biopsy of nodes or tumour and drainage of collections (especially recent post-operative). NM particularly good when there are no localising features: labelled WBC good for chronic post-operative sepsis; gallium will accumulate at sites of tumour (e.g. lymphoma) and infection.
G24			
Liver, Gallbladder & Pancreas ?Hepatic metastases	US CT or MRI	Indicated [B] Specialised [B] investigation	The majority of metastases will be demonstrated by US which also allows biopsy. US should be the initial investigation but metastases may show the same reflectivity as the hepatic parenchyma and thus be missed. Hence: CT/MRI used for further exclusion, where US equivocal or surprisingly normal and where full staging is needed or hepatic resection is planned (see also Cancer L13). Recent interest in dual phase spiral CT. MRI being increasingly used here. NM no longer used for this clinical problem.
G25			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
?Hepatic haemangioma (e.g. on US)	MRI or CT	Indicated [B]	Both MRI and CT reliably show further characteristic features of haemangioma and many other solitary hepatic lesions.
G26	NM (red cell study)	Specialised [B] investigation	Not often needed nowadays.
Jaundice	US	Indicated [B]	Sensitive for bile duct dilatation. But dilatation may be subtle in early obstruction and sclerosing cholangitis. Shows gallstones and most forms of hepatic disease. US also shows the level and cause of any obstruction to common bile duct. Discuss subsequent investigations (CT, ERCP, MRCP, etc.) with radiologist.
G27			
?Biliary disease, (e.g. gallstones)	AXR	Not [C] routinely	Plain XRs only show about 10% of gallstones.
	US	Indicated [B]	US allows evaluation of other organs too. Cholecystography is now rarely needed (e.g. poor views at US). CT/endoscopy may be needed for further delineation. Increasing interest in MRCP.
G28	NM	Specialised [B] investigation	Biliary scintigraphy shows cystic duct obstruction in acute cholecystitis. Also useful in chronic cholecystitis.
Pancreatitis: acute	AXR	Not [C] routinely	Unless diagnosis in doubt; then AXR needed to exclude other causes of acute abdo pain (see G19). Some patients presenting with acute pancreatitis have underlying chronic pancreatitis which may cause calcification evident on AXR.
	US	Indicated [B]	To show gallstones and to diagnose and follow pseudocyst development, especially good in thin patients.
G29	CT	Not [B] routinely	Reserved for clinically severe cases (to assess extent of necrosis), in patients who do not improve on treatment or if there is uncertainty as to the diagnosis. Some centres use MRI, especially if repeated follow-up likely.
Pancreatitis: chronic	AXR	Indicated [B]	To show calcification
	US or CT	Indicated [B]	US may be definitive in thin patients; CT will show calcification to good effect.
G30	ERCP or MRCP	Specialised [C] investigation	ERCP shows duct morphology, but considerable risk of acute pancreatitis. Hence current interest in MRCP.
Post-operative biliary leak	NM	Indicated [C]	US will usually have shown the anatomy of the collections, etc. NM study (HIDA) will show activity at site of leak. ERCP will show the anatomy of the leak and may allow intervention (e.g. stent).
G31			
?Pancreatic tumour	US CT	Indicated [B] Indicated [B]	Especially in thin patients and for lesions in the head and body. Increasing use of endoscopic and laparoscopic US. CT (or MRI) good in the more obese patient and where US equivocal or where precise staging needed. ERCP/MRCP may also be indicated.
G32			
?Insulinoma	Imaging	Specialised [B] investigation	When biochemical tests are convincing. MRI emerging as the best examination although arterial phase spiral CT promising. Most centres seek two positive investigations before surgery (out of CT/NM/MRI/angiography). Endoscopic and intra-operative US also useful.
G33			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
H. Urological, Adrenal and Genitourinary Systems			
Haematuria macro- or microscopic	US + AXR IVU	Indicated [B] Indicated [B]	There is a wide variation in local policy. Imaging strategies should be agreed with the local nephrologists and urologists. In many centres US + AXR are the initial studies, but if negative, IVU is still indicated in patients with continuing macroscopic haematuria or in the over 40s with microscopic haematuria. Conversely, patients in whom IVU and cystoscopy are normal who continue to bleed should undergo US, as IVU can fail to show a renal tumour and US will occasionally demonstrate a bladder lesion not seen at cystoscopy.
H1			
Hypertension (without evidence of renal disease)	IVU	Not [A]	IVU is insensitive for renal artery stenosis. See H3.
H2			
Hypertension: in the young adult or in patients unresponsive to medication	US kidneys	Indicated [B]	To assess relative renal size and parenchymal pattern. Doppler US is not sensitive enough for use as a screening tool.
	NM renogram	Specialised [C] investigation	Captopril renography is used in some centres to screen for functional renal artery stenosis.
H3	Angiography (DSA, CTA or MRA)	Specialised [C] investigation	To show stenosis if surgery or angioplasty is considered as a possible treatment.

Renal failure	US + AXR	Indicated [B]	For renal size, structure, obstruction, etc. N.B. a normal US does not exclude obstruction.
H4	NM	Indicated [B]	When appropriate, renography can assess renal perfusion and function.
Renal colic, loin pain	IVU	Indicated [B]	As an emergency examination whilst the pain is present, as radiological signs disappear rapidly after passage of a stone. Delayed films (up to 24 hrs) may be needed to show the site of obstruction. A plain film on its own is of little value. Some centres are now using spiral CT for initial diagnosis.
H5	US + AXR	Not [B] routinely	But of use in those with contraindications to contrast medium or irradiation. Patients need to be well hydrated in this case.
Renal calculi (in the absence of acute colic)	US + AXR	Indicated [C]	AXR alone may be appropriate follow-up for previously demonstrated calculi after an uncomplicated acute attack. An IVU may be required before treatment to show anatomy.
H6			
?Renal mass	US	Indicated [B]	US is good at distinguishing between cystic and solid masses.
H7	AXR + IVU	Not [C]	CT or MRI preferable for staging and assessing complex lesions shown at US.
Prostatism	US IVU	Indicated [B] Not [B]	US can assess upper tract and bladder volumes before and after voiding, preferably with flow rates. It can also show bladder calculi. Transrectal US is not routinely indicated.
H8			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Prostatic malignancy	US	Specialised [B] investigation	Transrectal US with guided biopsies after clinical examination. Some interest in MRI and PET here.
H9			
Urinary retention	US IVU	Indicated [C] Not indicated [C]	US to evaluate the upper tracts (after catheterisation and relief of bladder distension), particularly if urea levels remain raised.
H10			
?Scrotal mass or pain	US	Indicated [B]	Allows differentiation of testicular from extra-testicular lesions.
H11			
?Testicular torsion	US	Not [C] routinely	Torsion is a clinical diagnosis and imaging investigations must not delay the priority that must be given to surgical exploration. Doppler US can be used, when clinical findings are equivocal in the post-pubertal testis.
H12	NM	Specialised [C] investigation	NM techniques can assist with this diagnosis but prompt results essential.
Urinary tract infection in adults ***(For Children see Section M)***	US + AXR or IVU	Not [C] routinely	The majority do not need investigation unless there are recurrent infections, renal colic or failure to respond to antibiotics. Slightly lower threshold to investigate male patients. **N.B. This does not apply to children.**
H13			
Adrenal medullary tumours	CT or MRI	Specialised [B] investigation	Whilst US may identify lesions of this type, CT and MRI provide the best anatomical delineation. Imaging is rarely indicated in the absence of biochemical evidence of such tumours.
H14	NM	Specialised [B] investigation	MIBG locates functioning tumours and is particularly useful for ectopic sites and metastases.
Adrenal cortical lesions, Cushing's and Conn's disease and syndrome	CT, NM or MRI	Specialised [B] investigation	Local advice on the most appropriate examination should be sought. Both CT and MRI can differentiate between the different lesions. NM can distinguish between functioning and non-functioning adenomas. So too can various MRI techniques.
H15			

I. Obstetrics & Gynaecology

N.B. Transvaginal (TV) US equipment should be available in all Departments performing pelvic US

Screening in pregnancy	US	Indicated [C]	Screening US has not been shown to alter perinatal mortality, except where selective termination of pregnancy is applied in the presence of gross fetal abnormality. It does provide useful information about dating and multiple pregnancies. US is also of proven value in assessing placenta praevia and intra-uterine growth. In the specialist care of high risk pregnancies, Doppler US of the umbilical artery assists management. US is now commonly offered as a routine part of antenatal care in the UK, despite the scientific basis for its use remaining controversial. There is wide geographical variation in the number of antenatal US examinations performed. The optimal timing for a single US examination is 18–20 weeks of gestation.
I1			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Suspected pregnancy	US	Not [C]	Pregnancy testing most appropriate. US valuable where molar pregnancy suspected.
I2			
Suspected ectopic pregnancy	US	Indicated [B]	After positive pregnancy test. TV US preferred. Colour flow Doppler increases sensitivity.
I3			
Possible non-viable pregnancy	US	Indicated [C]	Repeat US after a week may be needed (especially when gestational sac < 20 mm or crown rump length < 6 mm). Pregnancy test required. Where doubt exists about the viability of a pregnancy, delay in evacuation of the uterus is essential.
I4			
Suspected pelvic mass	US	Indicated [C]	Combination of trans-abdominal and TV US often required. US should confirm a lesion's presence and determine likely organ of origin. See Cancer Section L. MRI is the best second line investigation, although CT still widely used.
I5			
Pelvic pain, including suspected pelvic inflammatory disease and suspected endometriosis	US	Indicated [C]	Especially when clinical examination difficult or impossible.
I6	MRI	Specialised [B] investigation	Can be useful to localised the larger foci of endometriosis.

Lost IUCD	US	Indicated [C]	
I7	AXR	Not [C]	Unless IUCD is not seen in uterus on US.
Recurrent miscarriages	US	Indicated [C]	Will show the major congenital and acquired problems.
I8	MRI	Specialised [C] investigation	Supplements US for uterine anatomy. Some centres still continue to use hysterosalpingography.
Infertility	US	Indicated [C]	For follicle-tracking during treatment. For assessment of tubal patency. Some centres still continue to use hysterosalpingography.
I9			
Suspected cephalopelvic disproportion	XR Pelvimetry	Not [B]	The need for pelvimetry is increasingly being questioned. Local policy should be determined in agreement with obstetricians. Furthermore MRI or CT should be used wherever possible.
I10	MRI or CT	Specialised [C] investigation	MRI is best as it avoids x-irradiation. CT generally offers a lower dose than standard XR pelvimetry.
J. Breast Disease			
A Screening < 40 yrs	Mammography	Not [A]	Cancer is uncommon under 35 and the sensitivity of mammography in detecting malignancy can be reduced in younger dense breasts.
J1			
Screening 40–49 yrs	Mammography	Not [A]	Recent evidence indicates that whilst cancers can be diagnosed at screening, benefit to the population of this age group is limited. The outcome of the UKCC 40–49 yrs trial is awaited.
J2			
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Screening 50–64 yrs	Mammography	Indicated [A]	Decreased mortality is proven with regular screening in this age group. The UK NHS Breast Screening Programme (NBSP) operates by invitation every 3 years. **N.B. a single view mammogram gives an average breast dose equivalent of about 1.2 mSv. The lifetime risk of induction of a lifetime cancer from one such examination in this age group is about 1:100,000. For women aged 39–49 this risk is approximately doubled.** Controversy about the correct indications for mammography in different age groups is primarily based upon considerations of clinical benefit, not risk. Although mammography is the best method for detecting early breast cancer, it is not 100% sensitive and a negative study cannot exclude breast cancer.
J3			
Screening 65 yrs +	Mammography	Indicated [A]	Definitely indicated. However self referral to the NBSP is required.
J4			
Family history of breast cancer	Mammography	Specialised [C] investigation	At present there is no evidence of benefit but there is some evidence of harm. Screening should only be contemplated after genetic risk assessments and appropriate counselling as to the risks and unproven benefits. Consensus at the moment is that screening should only be contemplated when the lifetime risk of breast cancer is greater than three or four times average. Units should collect and audit their work. This topic is being rigorously discussed at the present time.
J5			
Women < 50 yrs having or being considered for HRT	Mammography	Not [A]	A meta-analysis has shown women < 50 yrs who have received HRT for > 11 yrs are not at increased risk of breast cancer compared to a peer group. Women on HRT 50 yrs and over can be appropriately monitored within the NBSP.
J6			
Augmentation mammoplasty (50 yrs and over)	Mammography	Indicated [C]	As part of the NBSP – best performed at a static unit as there may be a need for extra views or US.
J7			
Symptomatic Patients Clinical suspicion of carcinoma	Mammography	Indicated [B]	Referral to a breast clinic should precede any radiological investigation.
J8	US	Specialised [B] investigation	Mammography ±US should be used in the context of triple assessment – i.e. clinical examination, imaging & cytology/biopsy.
Generalised lumpiness, generalised breast pain, or tenderness, or longstanding nipple retraction	Mammography or US	Not [C]	In the absence of other signs suggestive of malignancy, imaging is unlikely to influence management. Focal, rather than generalised pain may warrant investigation.
J9			
Cyclical mastalgia	Mammography	Not [B]	In the absence of other clinical signs suggestive of malignancy and localised pain, investigation is unlikely to influence management.
J10			
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Augmentation mammoplasty	US MRI	Indicated [B] Specialised [B] investigation	The assessment of integrity of breast implants or coincident masses requires specialist skills and facilities.
J11			
Paget's disease of the nipple	Mammography	Indicated [C]	The prevalence of coexistent breast cancer varies in published studies, but its association is clear and justifies specialist referral.
J12			
Breast inflammation	US	Indicated [B]	US can distinguish between an abscess requiring drainage and diffuse inflammation, and can guide aspiration when appropriate. Mammography may be of value where malignancy is possible.
J13			
K. Trauma			

Head: General

Head Injury: Protocols for management of head injuries are constantly under review and will vary according to local availability of CT, distances involved in transportation to neurosurgical centres, etc. The recommendations given here may need to be adapted following consultation with the neurosurgical centre for your area in the light of local circumstances and policies.

The key management and clinical questions in head injury are:

Clinical: Is there evidence of brain injury?
Is there evidence of intracranial haemorrhage or raised intracranial pressure?
Is there clinical evidence of a skull fracture and, if so, is it depressed?
Are other systems/areas involved?

Management: Does the patient need admission to hospital for observation?
Is CT required?
Is a neurosurgical opinion required?

These questions underline key policies concerning management of patients. Decisions about imaging requirements cannot be separated from non-imaging issues such as admission.

The usual indications for admission include: confusion or depressed consciousness; fracture on SXR; neurological symptoms or signs; seizures; CSF or blood from nose or ear; coagulation disorders; lack of adult supervision at home; patient difficult to assess (?non-accidental injury (NAI), drugs, alcohol, etc.). If a decision is made to admit for observation, imaging becomes less urgent, and the patient will be better examined when sober and more cooperative. CT is increasingly being used as the first investigation in those where there is a medium risk of intracranial injury, in which case SXR is usually unnecessary. Difficulties with image interpretation or the management of the patient may be resolved by referrals via image transfer systems to designated neuroscience centres.

Intracranial abnormalities suggesting need for urgent neurosurgical management include:

- High or mixed attenuation intracranial lesion
- Shift of mid line structures (e.g. third ventricle)
- Obliteration of third ventricle
- Relative dilatation of a lateral ventricle(s)
- Obliteration of basal cisterns
- Intracranial air
- Sub arachnoid or intraventricular haemorrhage.

Children

Head injuries are relatively common in children; in the majority of cases, there is no serious injury: imaging and hospitalisation are unnecessary. If there is a history of loss of consciousness, neurological signs or symptoms (excluding a single vomit) or an inadequate or inconsistent history, imaging is required. CT is the simplest way of excluding significant brain injury. If non-accidental injury is suspected, a skull SXR as part of a skeletal survey is required. In addition, MRI of the brain may be required later to further document timing of the injury.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Head: Low Risk of Intracranial Injury • Fully orientated • No amnesia • No loss of consciousness • No neurological defects • No serious scalp laceration • No haematoma **K1**	SXR CT	Not [C] Not [C]	These patients are usually sent home with head injury instructions to the care of a responsible adult. They may be admitted to hospital if no such adult is available.
Head: Medium Risk of Intracranial Injury • Loss of consciousness or amnesia • Violent mechanisms of injury • Scalp bruise, swelling or laceration down to bone or > 5 cm • Neurological symptoms or signs (including headache, vomiting twice or more, return visit) • Inadequate history or examination (epilepsy/alcohol/child/etc.) • Child below 5 yrs: suspected NAI, ?tense fontanelle, fall more than 60 cm or on to hard surface **K2**	CT or SXR	Indicated [B]	CT is increasingly being used as the first and ONLY investigation in this group of patients, to confidently exclude cranial injury. If no fracture is seen on SXR, patients will usually be sent home with head injury instructions to the care of a responsible adult. If no responsible adult available or if a fracture is present, the patient will usually be admitted. See Section M (M13) for non accidental injury in children. MRI of brain is the preferred investigation for intracranial injuries in NAI.

Head: High Risk of Intracranial Injury • Suspected FB or penetrating injury to skull • Disorientated or depressed consciousness • Focal neurological symptoms or signs • Seizure • Skull fracture or sutural diastasis shown on SXR • CSF from nose or CSF/blood from ear • Unstable systemic state precluding transfer to neurological unit • Diagnosis uncertain **K3**	CT	Indicated [B]	These patients will usually have been admitted for observation. If there is any delay in getting CT on an urgent basis, seek neurosurgical opinion. **N.B. CT should be available within 4 hrs of admission in all patients with a skull fracture.** SXR is not required before CT.
Head: Very High Risk of Intracranial Injury • Deteriorating consciousness or neurological signs (e.g. pupil changes) • Confusion or coma persistent despite resuscitation • Tense fontanelle or sutural diastasis • Open or penetrating injury • Depressed or compound fracture • Fracture of skull base **K4**	CT	Indicated [B]	URGENT NEUROSURGICAL AND ANAESTHETIC REFERRAL INDICATED, which should not be delayed by imaging. **N.B. CT should be performed as an emergency (see K3 above).**

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Face and Orbits Nasal trauma **K5**	SXR XR Facial bones XR Nasal bones	Not [B]	Unless required by a specialist. Poor correlation between radiological findings and presence of external deformity. Management of the bruised nose will depend on local policy: usually follow-up at an ENT or maxillo-facial clinic will determine the need for XR.
Orbital trauma: blunt injury **K6**	XR Facial bones	Indicated [B]	Especially in those where `blow-out' injury possible. MRI or low dose CT may eventually be required by specialists, especially when XRs or clinical signs equivocal.
Orbital trauma: penetrating injury **K7**	XR Orbits US or CT	Indicated [C] Specialised [B] investigation	When: 1. Radio-opaque intra-ocular FB is a possibility (see A16). 2. Investigation requested by ophthalmologist. 3. Suspicion of damage to orbital walls. US or low dose CT may be required; MRI contraindicated with metallic FB (see A16).
Middle third facial injury **K8**	XR Facial bones Low dose CT	Indicated [B] Specialised [B] investigation	But patient cooperation essential Advisable to delay XR in uncooperative patients. In children, XR often unhelpful. Discuss with maxillofacial surgeon who may require low dose CT at an early stage.
Mandibular trauma **K9**	XR Mandible or Orthopantomogram (OPG)	Indicated [C]	For non-traumatic TMJ problems see B11.
Cervical Spine Conscious patient with head and/or face injury only **K10**	XR C spine	Not [B]	In those who meet all of the following criteria: 1. Fully conscious. 2. Not intoxicated. 3. No abnormal neurological findings. 4. No neck pain or tenderness.
Unconscious head injury (see K3/4) **K11**	XR C spine	Indicated [B]	Must be of good quality to allow accurate evaluation. But radiography may be very difficult in the severely traumatised patient and must avoid manipulation (see also K12).
Neck injury: with pain **K12**	XR C spine CT or MRI	Indicated [B] Specialised [B] investigation	Cervical spine XRs can be very difficult to evaluate. Radiography also difficult and: 1. Must show C7/T1. 2. Should show odontoid peg (not always possible at time of initial study). 3. May need special views, CT or MRI especially when XR equivocal or complex lesions. Discuss with Department of Clinical Radiology.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Neck injury: with neurological deficit	XR	Indicated [B]	For orthopaedic assessment.
K13	MRI	Indicated [B]	Some constraints with life support systems. MRI best and safest method of demonstrating intrinsic cord damage, cord compression, ligamentous injuries and vertebral fractures at multiple levels. CT myelography may be considered if MRI not available.

Neck injury: with pain but XR initially normal; suspected ligamentous injury **K14**	XR C spine; flexion and extension	Specialised [B] investigation	Views taken in flexion and extension (consider fluoroscopy) as achieved by the patient with no assistance and under medical supervision. MRI may be helpful here.
Thoracic and Lumbar Spine Trauma: no pain, no neurological deficit **K15**	XR	Not [B]	Physical examination is reliable in this region. When the patient is awake, alert and asymptomatic, the probability of injury is low.
Trauma: with pain, no neurological deficit or patient not able to be evaluated **K16**	XR painful area	Indicated [B]	A low threshold to XR when there is pain/tenderness, a significant fall, a high impact RTA, other spinal fracture present or it is not possible to clinically evaluate the patient. Increasing use of CT and MRI here.
Trauma: with neurological deficit ± pain	XR	Indicated [B]	
K17	MRI	Indicated [B]	Where technically possible. CT often used as patient undergoing CT for other reasons. But MRI best method of demonstrating intrinsic cord damage, cord compression and vertebral fractures at multiple levels.
Pelvis and Sacrum Fall with inability to bear weight **K18**	XR pelvis plus lateral XR hip	Indicated [C]	Physical examination may be unreliable. Check for femoral neck fractures, which may not show on initial XR, even with good lateral views. In selected cases NM or MRI or CT can be useful when XR normal or equivocal.
Urethral bleeding and pelvic injury **K19**	Retrograde urethrogram	Indicated [C]	To show urethral integrity, leak, rupture. Consider cystogram if urethra normal and suspicion of bladder leak.
Trauma to coccyx or coccydynia **K20**	XR coccyx	Not [C] routinely	Normal appearances often misleading and findings do not alter management.
Upper Limb Shoulder injury **K21**	XR shoulder	Indicated [B]	Some dislocations present subtle findings. As a minimum, orthogonal views are required. US, MRI and CT arthrography all have a role in soft tissue injury.
Elbow injury **K22**	XR elbow	Indicated [B]	To show an effusion. Routine follow-up XRs not indicated in `effusion, no obvious fracture' (see also Section M). Increasing use of CT and MRI here.
Wrist injury **K23**	XR wrist NM or MRI	Indicated [B] Specialised [B] investigation	Scaphoid fractures can be invisible at presentation. Most centres repeat XR at 10-14 days if there are strong clinical signs and initial XR negative. Some departments use NM or MRI to exclude fracture earlier than this. Increasing use of MRI as the only examination.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Lower Limb Knee injury (fall/blunt trauma) **K24**	XR knee	Not [B] routinely	Especially where physical signs of injury are minimal. Inability to weight bear or pronounced bony tenderness, particularly at patella and head of fibula, merit radiography.
Ankle injury **K25**	XR ankle	Not [B] routinely	Features which justify XR include: the elderly patient, malleolar tenderness, marked soft tissue swelling and inability to bear weight.
Foot injury **K26**	XR foot	Not [B] routinely	Unless there is true bony tenderness. Even then the demonstration of a fracture rarely influences management. Only rarely are XRs of foot and ankle indicated together; both will not be done without good reason. Clinical abnormalities are usually confined to foot or ankle.
?Stress fracture	XR	Indicated [B]	Although often unrewarding.
K27	NM or MRI	Indicated [B]	Provides a means of early detection as well as visual account of the biomechanical properties of the bone. Some centres use US here.
The Foreign Body (FB) Soft tissue injury: ?FB (metal, glass, painted wood) **K28**	XR	Indicated [B]	All glass is radio-opaque; some paint is radio-opaque. Radiography and interpretation may be difficult; remove blood-stained dressings first. Consider US, especially in areas where radiography difficult.
Soft tissue injury: ?FB (plastic, wood)	XR	Not [B]	Plastic is not radio-opaque: wood is rarely radio-opaque.
K29	US	Indicated [B]	Soft tissue US may show non-opaque FB.
Swallowed FB suspected in pharyngeal or upper oesophageal region ***(For Children see Section M)*** **K30**	XR Soft tissues of neck	Indicated [C]	After direct examination of oropharynx (where most FBs lodge), and if FB likely to be opaque. Differentiation from calcified cartilage can be difficult. Most fish bones invisible on XR. Maintain a low threshold for laryngoscopy or endoscopy, especially if pain persists after 24 hrs (see K33). N.B. For possible inhaled FB in Children see Section M (M23).
Swallowed FB: smooth and small (e.g. coin) **K31**	CXR AXR	Indicated [B] Not [B] routinely	The minority of swallowed FBs will be radio-opaque. In children a single, slightly over-exposed, frontal CXR to include neck should suffice. In adults, a lateral CXR may be needed in addition if frontal CXR negative. Majority of FBs that impact, do so at crico-pharyngeus. If the FB has not passed (say within 6 days), AXR may be useful for localisation.
Sharp or potentially poisonous swallowed FB: (e.g. ?battery)	AXR	Indicated [B]	Most swallowed foreign bodies that pass the oesophagus eventually pass through the remainder of the gastrointestinal tract without complication. But location of batteries is important as leakage can be dangerous.
K32	CXR	Not [B] routinely	Unless AXR negative.
Swallowed FB: Large object (e.g. dentures) **K33**	CXR	Indicated [B]	Dentures vary in radio-opacity; most plastic dentures are radiolucent. AXR may be needed if CXR negative, as may barium swallow or endoscopy. Lat CXR may be helpful.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT

Chest Chest trauma: minor **K34**	CXR	Not [B] routinely	The demonstration of a rib fracture does not alter management.
Chest trauma: moderate **K35**	CXR	Indicated [B]	Frontal CXR for pneumothorax, fluid or lung contusion. A normal CXR does not exclude aortic injury and arteriography/CT/MRI should be considered.
Stab injury **K36**	CXR	Indicated [C]	PA and/or other views to show pneumothorax, lung damage or fluid. US useful for pleural and pericardial fluid.
?Sternal fracture **K37**	XR lateral sternum	Indicated [C]	In addition to CXR. Think of thoracic spinal and aortic injuries too.
Abdomen (Including Kidney)			
Blunt or stab injury **K38**	Supine AXR + erect CXR	Indicated [B]	US valuable for detecting haematoma and possible injury to some organs, e.g. spleen, liver. CT may be needed (see K40,K41 and K42).
?Renal trauma **K39**	Imaging	Indicated [B]	Discuss with radiologist. In agreement with local policy and availability. US often sufficient for minor local injury. Many centres use a limited IVU, merely to ensure normality of contralateral kidney. Some patients with major injury (see below) undergo CT, making IVU unnecessary. Consider renal artery damage, especially in deceleration injuries; arteriography may be needed.
Major Trauma Major trauma–general screen in the unconscious or confused patient **K40**	C spine XR CXR Pelvis XR CT head	Indicated [B]	Stabilise patient's condition as a priority. Perform only the minimum XRs necessary at initial assessment. C spine XR can wait so long as spine and cord suitably protected. Pelvic fractures often associated with major blood loss. See Head Injury K1,K2,K3 and K4
Major trauma–abdomen/pelvis	CXR, Pelvis XR	Indicated [B]	Pneumothorax must be excluded. Pelvic fractures which increase pelvic volume often associated with major blood loss.
K41	CT abdo	Indicated [B]	Sensitive and specific, but time-consuming and may delay Theatre. CT should precede peritoneal lavage. Increasing interest in the use of US in emergency room to show free fluid plus solid organ-injury.
Major trauma-chest	CXR	Indicated [B]	Allows immediate management (e.g. pneumothorax).
K42	CT Chest	Indicated [B]	Especially useful to exclude mediastinal haemorrhage. Low threshold for proceeding to arteriography.

L. Cancer

Many of the clinical problems related to the diagnosis of cancer have already partly been covered within the individual system sections. Brief notes are provided here about the use of imaging in the diagnosis, staging and follow-up of some of the common primary malignancies. Paediatric malignancies are not included as their management is always at specialist level. For breast cancer see Section J. A CXR is necessary at presentation for most malignant lesions to identify possible pulmonary metastases. CXR is also part of many follow-up protocols (e.g. testicular lesions). Follow-up investigations to monitor progress (e.g. post-chemotherapy) are often required; some are driven by trial protocols rather than clinical need and thus should be appropriately funded.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Parotid Diagnosis	US	Indicated [B]	To establish presence of a mass, particularly in superficial lesions.
L1	MRI or CT	Indicated [B]	Useful in the deep portion of the gland and before complex surgery.
Staging **L2**	MRI or CT	Indicated [B]	Especially when complex surgery contemplated; to see relations and involvement of deep lobe.
Larynx Diagnosis **L3**	Imaging	Not [B]	This is a clinical diagnosis.
Staging **L4**	CT or MRI	Indicated [B]	MRI has the advantage of direct coronal imaging. MRI will eventually supersede.
Thyroid Diagnosis **L5**	US and NM	Indicated [B]	See Neck Section B1. US guided core biopsy is increasingly being used.
Staging	CT or MRI	Indicated [B]	To assess local extent (e.g. retrosternal extension and nodes).
L6	NM	Indicated [B]	After thyroidectomy.
Lung Diagnosis	CXR PA and Lat	Indicated [B]	But can be normal, particularly with central tumours.
L7	CT	Specialised [B] investigation	Many centres proceed directly to bronchoscopy which allows biopsy. CT is superior in identifying lesions responsible for haemoptysis.
Staging	CT chest, upper abdomen MRI	Indicated [B] Specialised [B] investigation	Despite limitations in specificity of nodal involvement, etc. Some centres perform NM for possible skeletal metastases. Assists in estimating local invasion of chest wall, particularly for apical and peripheral lesions and mediastinal invasion. Helps distinguish adrenal adenoma from metastasis.
L8	PET	Specialised [B] investigation	A single expensive investigation to identify small metastatic foci may save a lot of other investigations and inappropriate surgery.
Oesophagus Diagnosis **L9**	Barium swallow	Indicated [B]	Before endoscopy in dysphagia.
Staging	CT	Indicated [B]	Despite limitations in sensitivity and specificity of nodal involvement. Simpler than MRI for lung, liver and intra-abdominal nodes.
L10	Transoesophageal US	Indicated [A]	Increasing use of transoesophageal US for local staging where available.
Liver: Primary Lesion Diagnosis	US	Indicated [B]	The majority of lesions will be identified.
L11	MRI or CT	Indicated [B]	If biochemical markers elevated and US negative or liver very cirrhotic. Enhanced MRI and arterial phase CT most accurate in delineating tumour extent.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Staging	MRI or CT	Indicated [B]	MRI probably the optimal investigation in assessing involved segments and lobes. CT arterial portography and intra-operative US useful where available.
L12			
Liver: Secondary Lesion Diagnosis	US	Indicated [B]	US will show the majority of metastases and guides biopsy.
L13	CT or MRI	Indicated [B]	When US negative and clinical suspicion high. MRI better for characterising lesions. CT arterial portography is sensitive but not specific, but many now use triple phase spiral CT techniques following intravenous enhancement. CT and MRI often part of other staging and follow-up protocols.
Pancreas Diagnosis	US or CT or MRI	Indicated [B]	Much depends on body habitus. US usually successful in thin patients; CT better in the more obese. MRI for clarification of problems. Biopsy using US or CT. ERCP or MRCP may also be needed. Endoscopic US, where available, most sensitive.
L14			
Staging	CT or MRI abdomen	Indicated [B]	Especially if radical surgery contemplated. Wide local variation: some centres use angiography, others spiral CT; laparoscopic US also used.
L15			
Colon and Rectum Diagnosis	Ba enema or colonoscopy	Indicated [B]	Much depends on local policy, expertise and availability. See Section G.
L16			Increasing interest in CT of the colon.
Staging	US	Indicated [B]	For liver metastases. Endoluminal US useful for local rectal spread.
L17	CT or MRI abdomen, pelvis	Indicated [B]	Local pre-operative staging to assess rectal lesions before pre-operative radiotherapy. Many centres now treat liver secondaries very aggressively, which may necessitate MRI and/or detailed CT. MRI and CT often complementary; both can assess other abdominal spread. Increasing interest in PET here.
?Recurrence	US liver	Indicated [B]	For liver metastases. Some debate about the value of routine US follow-up in asymptomatic patients.
L18	CT or MRI abdomen, pelvis	Indicated [B]	For liver metastases and local recurrence. Increasing interest in PET here.
Kidney Diagnosis	US	Indicated [B]	See Renal Mass H7.
L19			
Staging	CT or MRI abdomen	Indicated [B]	For local extent, venous, nodal and ureteric involvement, opposite kidney etc.
L20	CT Chest	Not [B] routinely	The presence of lung metastases does not usually, influence management. Increasing interest in PET.
?Recurrence	CT abdomen	Indicated [B]	For symptoms suggesting relapse around nephrectomy bed. Routine follow-up not recommended.
L21			

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Bladder Diagnosis	Imaging	Not [B]	Cystoscopy is the optimal (although not infallible, e.g. diverticulum) investigation.
L22			
Staging	IVU	Indicated [B]	To assess kidneys and ureters for further urothelial tumours
L23	CT or MRI abdomen and pelvis	Indicated [B]	When radical therapy contemplated. MRI is probably more sensitive. CT widely used for radiotherapy planning.
Prostate Diagnosis	Transrectal US	Indicated [B]	Some variation according to local availability and expertise. Transrectal US is widely used together with guided biopsies.
L24			Some interest in MRI and PET here.
Staging	MRI/CT pelvis,	Specialised [B]	Some variation in range of investigative and therapeutic policies. MRI with appropriate coils is sensitive for assessment before possible radical prostatectomy. Staging continued into the abdomen when pelvic disease found.
L25	NM	Indicated [B]	To assess skeletal metastases, when PSA is significantly elevated.
Testis Diagnosis	US	Indicated [B]	Especially when clinical findings equivocal or normal.
L26			
Staging	CT chest, abdomen, pelvis	Indicated [B]	Management now depends heavily on accurate radiological staging.
L27			
Follow-up	CT abdomen	Indicated [B]	Some centres still routinely examine the chest as well, especially for patients without biochemical evidence of disease. Some debate as to whether whole pelvis is needed at follow-up unless there are identified risk factors.
L28			
Ovary Diagnosis	US	Indicated [B]	The majority of lesions are diagnosed by US (including TV with Doppler), laparoscopy or laparotomy. Some are identified by CT/MRI investigations for abdominal symptoms. MRI useful for elucidating problems.
L29			
Staging	CT/MRI abdomen, pelvis	Specialised [B] investigation	Many specialists require CT or MRI in addition to staging by laparotomy. CT is still more widely available.
L30			
Follow-up	CT abdomen, pelvis	Specialised [B] investigation	Usually to assess response to adjuvant therapy. Also use, along with markers, to detect relapse.
L31			

Uterus: Cervix Diagnosis **L32**	Imaging	Not [B] routinely	Usually a clinical diagnosis. MRI may assist in complex cases.
Staging **L33**	MRI or CT abdomen and pelvis	Indicated [B]	MRI provides better demonstration of tumour and local extent. Also better for pelvic nodes. Para-aortic nodes and ureters must also be examined. Para-aortic nodes and ureters must also be examined. Some centres now use transrectal US for local invasion.
?Relapse **L34**	MRI or CT abdomen and pelvis	Specialised [B] investigation	MRI provides better information in the pelvis. Biopsy (e.g. of nodal mass) easier with CT.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Uterus: Body Diagnosis **L35**	US or MRI	Indicated [B]	MRI can give valuable information about benign and malignant lesions.
Staging **L36**	MRI or CT investigation	Specialised [B]	Both CT and MRI can show extra-uterine disease. But MRI can also demonstrate intra-uterine anatomy.
Lymphoma Diagnosis	CT	Indicated [B]	CT good at evaluating nodal sites throughout the body. Also allows biopsy although excision of whole node preferable where possible.
L37	NM	Specialised [B] investigation	NM (gallium) can show foci of occult disease (e.g. mediastinum). PET used in some centres.
Staging **L38**	CT chest, abdomen, pelvis	Indicated [B]	Depending on site of disease, head and neck may also need to be examined. Increasing interest in PET here.
Follow-up	CT or MRI	Indicated [B]	Increasing role for MRI in long term follow-up and residual masses.
L39	NM	Specialised [B] investigation	Consider NM for gallium positive disease. Some centres use PET.
Musculoskeletal Tumours			
Diagnosis **L40**	XR + MRI	Indicated [B]	Imaging and histology complementary. Best before biopsy: See Musculoskeletal Section D.
Staging **L41**	MRI local disease + CT chest	Specialised [C]	See Musculoskeletal Section D. CT for lung metastases.
Metastases from Unknown Primary Tumour			
Diagnosis of primary lesion **L42**	Imaging	Not [C] routinely	Rarely beneficial. Some exceptions for specialists, younger patients or favourable histology.
M. Paediatrics ***Minimise x-irradiation in children, especially those with long term problems*** ***(For head injury in children see Trauma Section K)***			
CNS Congenital disorders **M1**	MRI	Indicated [C]	Definite exam for all malformations and avoids x-irradiation. Sedation usually required for young children. Consider US in neonates. 3D CT may be needed for bone anomalies.
Large head circumference ?hydrocephalus **M2**	US	Indicated [B]	US indicated where anterior fontanelle is open. MRI indicated for older children. (CT may be appropriate if MRI not available.)
Epilepsy	SXR	Not [B]	Poor yield.
M3	MRI or NM	Specialised [B] investigation	MRI usually more appropriate than CT, SPECT also used in some centres.
Deafness in children **M4**	MRI, CT	Specialised [C] investigation	Both MRI and CT may be necessary in children with congenital and post-infective deafness.
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Hydrocephalus ?shunt malfunction (see A10)	XR	Indicated [B]	XR should include whole valve system.
M5	US or MRI	Indicated [B]	US if practical, MRI in older children (or CT if MRI unavailable). NM used in some centres.
Developmental delay ?cerebral palsy **M6**	MRI	Specialised [B] investigation	See also M15 for skeletal investigation of growth failure.
Headaches	SXR	Not [B] routinely	If persistent or associated with clinical signs refer for specialised investigations.
M7	MRI	Specialised [B] investigation	In children MRI is preferable if available because of absence of x-irradiation.
?Sinusitis **M8**	Sinus XR	Not [B] routinely	Not indicated before 5 yrs as the sinuses are poorly developed; mucosal thickening can be a normal finding in children. A single undertilted OM view may be more appropriate than the standard OM view depending on the child's age.
Neck and Spine Torticollis without trauma **M9**	XR	Not [B]	Deformity is usually due to spasm with no significant bone changes. If persistent, further imaging may be indicated following consultation.
Back or neck pain	XR	Indicated [B]	Back pain is uncommon in children without a cause. Follow-up is needed if infection is suspected.
M10	NM	Specialised [B] investigation	When pain continues and XRs are normal. Useful in painful scoliosis.
	MRI	Specialised [B] investigation	See also The Spine Section C. MRI defines spinal malformations and excludes associated thecal abnormality.

?Spina bifida occulta **M11**	XR	Not [B]	A common variation and not in itself significant (even in enuresis). However, neurological signs would require investigation.
Hairy patch, sacral dimple	XR	Not [B]	
M12	US	Indicated [B]	US may be useful in the neonatal period to screen for underlying tethered cord, etc.
	MRI	Specialised [B] investigation	MRI particularly if neurological signs are present.
Musculoskeletal ?Non accidental injury ?child abuse ***(For head injury see Section K)***	XR of affected parts	Indicated [B]	Local policies will apply; close clinical/radiological liaison essential. Skeletal survey for those under 2 yrs after clinical consultation. May occasionally be required in the older child. Consider MRI of brain, even in the absence of cranial apparent injury.
M13	NM	Specialised [B] investigation	Sensitive for occult spine/rib fracture in the younger child where history unavailable.
Limb injury: opposite side for comparison **M14**	XR	Not [B] routinely	Seek radiological advice.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Short stature, growth failure	XR for bone age	Indicated [B] appropriate intervals	2-18 yrs: left (or non-dominant) hand/wrist only. Premature infants and neonates: knee (specialised investigation).
M15			May need to be supplemented with a skeletal survey and MRI for hypothalamus and pituitary fossa (specialised investigations).
Irritable hip **M16**	US	Indicated [B]	US will delineate effusions which can be aspirated for diagnostic and therapeutic purposes. XRs can be delayed, but should be considered when the symptoms are persistent. Consider NM or MRI when Perthes' disease is suspected and plain XRs are normal.
Limp	XR pelvis	Indicated [C]	Gonad protection is used routinely unless shields will obscure area of clinical suspicion. If slipped epiphyses is likely, lateral XRs of both hips are needed.
M17	US or NM or MRI	Specialised [B] investigation	According to local policy, expertise and availability.
Focal bone pain	XR and US	Indicated [B]	XR may be normal initially. US can be helpful particularly in osteomyelitis.
M18	NM or MRI	Specialised [B] investigation	Increasing use of MRI here.
Clicking hip ?dislocation **M19**	US	Indicated [B]	XR may be used to supplement US examination. XR indicated in the older infant.
?Osgood-Schlatter's disease **M20**	XR	Not [C]	Although bony radiological changes are visible in Osgood-Schlatter's disease these overlap with normal appearances. Associated soft tissue swelling should be assessed clinically rather than radiographically.
Cardiothoracic Acute chest infection **M21**	CXR	Not [B] routinely	Initial and follow-up films are indicated in the presence of persisting clinical signs or symptoms or in the severely ill child. Consider the need for CXR in PUO. Children may have pneumonia without clinical signs.
Recurrent productive cough **M22**	CXR	Not [C] routinely	Children with recurrent chest infection tend to have normal CXRs (apart from bronchial wall thickening). Routine follow-up CXR not indicated unless collapse present on initial CXR. Suspected cystic fibrosis requires specialist referral.
Inhaled FB (suspected) ***(see Section K)*** **M23**	CXR	Indicated [B]	History of inhalation often not clear. Bronchoscopy is indicated, even in the presence of a normal CXR. Expiratory films usually sufficient to confirm the presence of air trapping. Chest screening in young children not routinely indicated in view of the high radiation dose.
Wheeze **M24**	CXR	Not [B] routinely	Children with asthma usually have normal CXR apart from bronchial wall thickening. Sudden unexplained wheeze CXR indicated, may be due to inhaled FB (above).
Acute stridor **M25**	XR	Not [B]	Epiglottitis is a clinical diagnosis, but consider FB (above).
Heart murmur **M26**	CXR	Not [C] routinely	Specialist referral needed; cardiac US may be indicated.

CLINICAL PROBLEM	INVESTIGATION	RECOMMENDATION {GRADE}	COMMENT
Gastrointestinal Intussusception	AXR	Indicated [C]	Local policies require close paediatric, radiological and surgical liaison. Where expertise is available, both US and contrast enema (air or barium) can confirm diagnosis and guide reduction.
M27	Further imaging	Specialised [B] investigation	
Swallowed FBs ***(see Section K)***	AXR	Not [C] routinely	Except for sharp or potentially poisonous FBs, e.g. batteries. See Section K.
M28	CXR (including neck)	Indicated [C]	If there is doubt whether the FB has passed, an AXR after 6 days may be indicated.

Minor trauma to abdomen **M29**	AXR	Not [C]	US may be used as initial investigation but CT is more specific, particularly in visceral trauma. XRs may show bone injury in severe trauma. The principles for the investigation of major trauma in children similar to those in adults (see Major Trauma, K40,K41 and K42).
Projectile vomiting **M30**	US	Indicated [A]	US can confirm the presence of hypertrophic pyloric stenosis, especially where clinical findings are equivocal.
Recurrent vomiting **M31**	Upper GI contrast study	Not [C] routinely	This symptom covers a wide range from obstruction in the neonatal period to reflux, posseters and children with migraine. US may be helpful to confirm malrotation. However, upper GI contrast studies may be indicated to exclude malrotation even with normal abdominal XR. Contrast studies in neonates should be undertaken as a specialised investigation.
Persistent neonatal jaundice	US	Specialised [B] investigations	Early and prompt investigation is essential. The absence of dilation in the intrahepatic bile duct does not exclude an obstructive cholangiopathy.
M32	NM		
Rectal bleeding **M33**	NM	Specialised [B] investigation	If Meckel's diverticulum is a possibility do NM first. Small bowel contrast studies may also be necessary. NM also useful in investigation of inflammatory bowel disease. Endoscopy is preferable to Ba enema for assessment of polyps and inflammatory bowel disease. US can be used to diagnose duplication cysts.
Constipation	AXR	Not [C] routinely	Many normal children show extensive faecal material; impossible to assess significance of radiological signs. But AXR can help specialists in refractory cases.
M34	Contrast enema	Not [B] routinely	If Hirschsprung's disease is suspected, specialist referral plus biopsy is preferred to radiological studies.
Palpable abdominal/pelvic mass **M35**	US	Indicated [B]	If malignancy is suspected, further imaging should be performed in a specialised centre.
Uroradiology Enuresis **M36**	Imaging	Not [B] routinely	US and urodynamic studies may be needed in cases of persistent enuresis.
Continuous wetting	US	Indicated [B]	Both examinations may be needed to evaluate duplex system with ectopic ureter.
M37	IVU	Indicated	
CLINICAL PROBLEM	**INVESTIGATION**	**RECOMMENDATION {GRADE}**	**COMMENT**
Impalpable testis	US	Indicated [B]	To locate inguinal testis.
M38	MRI	Specialised [B] investigation	MRI may be helpful to locate an intra-abdominal testis, but increasingly laparoscopy is the investigation of choice.
Antenatal diagnosis of urinary tract dilatation **M39**	US	Indicated [B]	Local protocols should be established. Mild dilatation can normally be monitored by US. Low threshold for specialist referral.
Proven urinary tract infection	Imaging US/NM/ cystography	Specialised [C] investigations	There is wide variation in local policy. Much depends on local technology and expertise. Most patients should remain on prophylactic antibiotics pending the results of investigations. The age of the patient also influences decisions. There is much current emphasis on minimising radiation dose; hence AXR is not indicated routinely (calculi rare).
M40			Expert US is the key investigation in all imaging strategies at this age. Thereafter NM provides data about renal structure (DMSA) and has virtually replaced the IVU here. NM will establish function, exclude obstruction and can also be used for cystography (direct or indirect). Formal direct XR cystography is still needed in the young (e.g. < 2 yrs) male patient where delineation of the anatomy (e.g. urethral valves) is critical.

*We are grateful to Dr D.R.M. Lindsell for arranging for the reproduction of these guidelines and the Royal College of Radiologists for their permission to reproduce them.

14.1 Upper gastrointestinal endoscopy

Timothy Simon and Aaron S. Fink

History
Instrumentation
Monitoring
Indications
Patient preparation
Diagnostic and therapeutic techniques
Diagnostic technique
Tissue sampling techniques
Hemostasis—non-variceal
Hemostasis—variceal
Stricture dilatation
Endoscopic ultrasound
Indications
Further reading

History

Early endoscopes were quite uncomfortable. Using reflected candle light for illumination, they afforded only limited examination views. In 1868, Kussmaul achieved the first practical use of a rigid endoscope while performing gastroscopy on a sword swallower. In 1886, Mikulicz constructed an open-tube rigid esophagoscope which used a miniature light bulb for illumination. Chevalier Jackson modified this into the standard bronchoesophagoscope in the early 1900s.

In the early 1930s, Rudolph Schindler developed a semiflexible gastroscope. Although this instrument allowed better esophageal and gastric imaging, there were many limitations: the need for hyperextension of the neck due to the rigid proximal portion; poor visualization of the fundus, posterior wall, and antrum of the stomach; and the lack of ports for aspiration or biopsy. It was not until 1957 that Hirschowitz introduced the first flexible fiberoptic gastroscope. This instrument revolutionized the practice of endoscopy, allowing therapeutic interventions such as air insufflation, snare biopsy, and hemostasis of bleeding lesions.

Application of endoscopy continues to broaden; future projections suggest that as many as 1 per cent of the population will undergo upper gastrointestinal endoscopy annually. Thus, diagnostic and therapeutic endoscopy will continue to play a significant role in the management of gastrointestinal disorders. This chapter will review the status of gastrointestinal endoscopy in current practice.

Instrumentation

Flexible endoscopic systems transmit light down the endoscope shaft, illuminating the surface to be examined. The image is reflected back and transmitted to the endoscopist either fiberoptically or electronically. In fiberoptic systems, a fixed lens at the end of the instrument shaft focuses the reflected image on an internal fiberoptic bundle. Fiberoptic bundles are 2 to 3 mm wide and are composed of 20 000 to 40 000 individual glass fibers, each approximately 10 μm in diameter. Each fiber is coated with glass of low optical density, which prevents light from escaping. The scope head includes an eyepiece with an adjustable lens; dials which control lens tip deflection; buttons controlling air, water, and suction; and the therapeutic channel inlet.

Most modern systems utilize electronic image transmission. These video endoscopes are very similar to the fiberscopes, offering similar depth of focus, field of view, and tip deflection. The image is reflected on a charge-coupled device (**CCD**) chip mounted on the shaft end. The chip's image contains 30 000 to 150 000 pixels; resolution improves with increasing pixel numbers. The electronic image is transmitted through wires to a video processor in the instrument head.

Newer color video endoscopes use sequential CCD chips, which sequentially illuminate all pixels with light of the three primary colors, alternating each 20 to 30 times/min. Sequential chips can be mounted on smaller endoscopes and offer greater resolution. However, they require larger, more expensive, light source/processor units.

Monitoring

Though endoscopy is generally safe with negligible mortality rates (0.006 to 0.012 per cent of esophagogastroduodenostomies), it may be associated with serious complications. Cardiopulmonary events are most commonly seen. Transient changes in blood pressure and heart rate occur in anywhere between 14 and 100 per cent of cases. Aspiration, oversedation, hypotension, various dysrhythmias, and airway obstruction also occur. Most of these complications are associated with intravenous conscious sedation, particularly in the elderly.

Pulse oximetry, heart rate, respiratory rate, and blood pressure readings should be routinely monitored for every patient. These measures allow detection of arterial oxygen desaturation and altered pulse rate, providing early indication of potential difficulties. ECG monitoring should be considered for high-risk patients and for those that have a history of cardiac dysfunction or significant dysrhythmias.

Although oxygen is primarily recommended in those with significant cardiopulmonary disease and those undergoing prolonged procedures, supplemental oxygen has been shown to decrease the frequency of desaturations during endoscopic procedures.

Indications

When compared with double-contrast barium meal, endoscopy is more sensitive and specific in identifying all types of lesions. Indeed, in a comparative study, when endoscopy was performed first, only 5 per cent of patients required subsequent barium swallow. Nearly 30 per cent of patients undergoing initial barium meal subsequently required endoscopy for definitive diagnosis. Table 1 outlines current indications recommended by the American Society for Gastrointestinal Endoscopy (**ASGE**).



Table 1 Indications for upper gastrointestinal endoscopy.

Upper endoscopy is a very useful tool in the diagnostic evaluation and therapeutic management of upper gastrointestinal symptoms. Indications have been well established (Table 1). Dysphagia/odynophagia, upper abdominal pain, and unexplained vomiting may all be related to anatomic esophageal lesions, which can be readily visualized endoscopically. While lesions such as Zencker's diverticulas may require barium swallow for diagnosis, esophageal webs and tumors are better evaluated endoscopically. Since most esophageal tumors are cancerous, esophageal obstructive symptoms mandate esophagoscopy for potential inspection, biopsy,

and occasionally treatment of the underlying pathology.

Upper gastrointestinal bleeding is one of the major indications for upper endoscopy. Whether bleeding stems from esophageal varices, Mallory–Weiss tears, or peptic ulcer, endoscopy is preferred given its ability to both diagnose and treat the bleeding source in one setting.

Surveillance and follow-up of known pathology are also important roles for upper endoscopy. Potential indications include Barrett's esophagus where detection of dysplasia is critical to appropriate management. Barrett's metaplasia predisposes patients to esophageal adenocarcinoma in about 9 to 15 per cent of cases. The degree of risk increases with segments greater than 8 cm and with presence of high-grade dysplasia within the Barrett's mucosa.

Plummer–Vinson syndrome, lye ingestion, and achalasia also deserve special mention. Plummer–Vinson syndrome is a rare condition characterized by cervical esophageal webs seen primarily in women with associated iron-deficiency anemia. Since cervical esophageal squamous cell cancers develop in 10 to 12 per cent of these patients, surveillance endoscopy can be beneficial. Squamous cell cancers also develop in 2 to 7 per cent of patients with achalasia. Patients with chronic dilatation and stasis warrant aggressive follow-up.

Patients who survive lye ingestion face a 1000-fold increased risk of cancer development compared with the general population. The time from the caustic burn until the onset of cancer averages about 40 years, but may be as short as 9 years. If the damaged esophagus was left *in situ*, endoscopic screening is appropriate beginning 20 years after injury or upon development of dysphagia.

Due to their malignant potential, adenomatous gastric polyps and duodenal polyps merit endoscopic surveillance. Such polyps are frequently seen in patients with colon polyposis syndromes. Note that most gastric polyps in these syndromes represent fundic gland hyperplasia, while associated duodenal polyps are usually adenomatous. These polyps should be followed as described for colonic polyps. Cancer risk in these adenomatous polyps, as with their colonic counterparts, increases with polyp size.

Gastric stump cancer has been reported in 5 to 6 per cent of patients who have previously undergone gastric resection for benign disease. The incidence is related to the development of dysplasia. Because of the 2- to 4-fold increased risk of cancer development, selected surveillance of symptomatic patients has been recommended beginning 15 years after the original operation.

Endoscopy can be used to document healing of gastric and duodenal ulcers after appropriate medical treatment. Endoscopy discriminates patients that have intractable symptoms due to a refractory ulcer from those without endoscopic evidence of an ulcer. The latter finding suggests the need to pursue other etiologies for the presumed ‘ulcer' symptoms.

Patient preparation

Elective upper endoscopy requires at least a 4- to 6-h fast before the procedure. Because of the increased risk of aspiration, patients with known gastric outlet obstruction or delayed gastric emptying due to other causes (e.g. diabetes mellitus) should have clear liquids the preceding day and fast overnight before endoscopy. Gastric aspiration may also be helpful prior to endoscopy in the latter patients. Patients' eyeglasses and dentures should be removed and intravenous access established. Blood bank support must be available for patients with gastrointestinal bleeding.

Endoscopic procedures rarely require antibiotics. Patients at high risk for bacteremic complications such as those with artificial heart valves or prior endocarditis should be given prophylactic antibiotics. Those with recently placed orthopedic or vascular prostheses or systemic–pulmonary shunts should also receive appropriate antibiotics. There is evidence that prophylactic intravenous antibiotics decrease the rate of skin infection associated with percutaneous endoscopic gastrostomy (**PEG**) placement. Guidelines for appropriate use of antibiotics during endoscopy are outlined by the ASGE.

Except for PEG placement, the patient is studied in the left lateral decubitis position. Topical pharyngeal anesthesia is frequently applied to the posterior pharynx to suppress the gag reflex.

Most patients are given conscious sedation for gastroscopy. Meperidine, midazolam, and diazepam are the three agents most commonly used in the United States. Due to the small (3.5 per cent) frequency of thrombophlebitis, midazolam has overtaken diazepam as the benzodiazepine of choice. Midazolam and lidocaine spray synergistically enhance patient tolerance of endoscopy and decrease the incidence of cardiorespiratory disturbance. Elderly patients usually require lower sedative doses, which helps reduce cardiopulmonary risks.

Diagnostic and therapeutic techniques

Diagnostic technique

A 120-cm forward viewing endoscope is favored for routine diagnostic upper endoscopy. It is essential to have a trained assistant at the patient's head throughout the examination. This individual can support and manipulate the patient's head to facilitate introduction of the endoscope. The assistant should also monitor the patient's vital signs and suction the oropharynx as needed to protect the patient's airway. The assistant also can help with various therapeutic maneuvers including biopsies and polypectomies.

After patient preparation and confirmation that the equipment is functioning well, the tip of the endoscope is lubricated and advanced into the patient's esophagus by one of two techniques. The preferred method of intubation is under direct vision. The endoscope is advanced through the center of a previously placed mouthpiece, over the tongue, and past the epiglottis to the cricopharyngeal sphincter. The patient is then instructed to swallow which relaxes the cricopharyngeus muscle, allowing entry into the esophagus.

Some patients may prove difficult to intubate. Blind introduction of the endoscope while depressing the tongue with the second and third fingers of the left hand allows the endoscope to be placed at the cricopharyngeus muscle. The mouthpiece is then positioned, continuing the procedure as described below. This latter technique is less frequently used because of the risk of injury to patient, endoscopist, and endoscope.

The examination continues under direct vision, advancing the instrument towards an open lumen to the desired level. As the scope is advanced, the mucosa is carefully inspected noting anatomic landmarks and their distance from the incisors. Important landmarks include the ‘Z-line', the lower esophageal sphincter, the cardia, the incisura, the pylorus, and superior duodenal angle.

The ‘Z-line' defines the gastro-esophageal junction and is usually 36 to 40 cm from the incisors. Here the white squamous esophageal mucosa meets the red columnar gastric mucosa below. The diaphragmatic ‘pinch zone' usually is seen within 2 cm of the Z-line, identifying the diaphragmatic hiatus. The pinch zone can be appreciated when the patient is instructed to sniff, causing the diaphragm to contract and impinge on the lumen. Separation of the Z-line from the pinch zone by more than 2 cm usually suggests hiatus hernia. The stomach is then entered by deflecting the tip of the instrument upward and to the left revealing the cardia.

After insufflation of air, gastric contents are aspirated and the gastric walls are surveyed. The endoscope is then advanced parallel to the longitudinal folds on the stomach's greater curvature. Corkscrewing the tip to the right, around the vertebral column, allows the antrum to be visualized. The pylorus is now visualized end-on. Even when closed, the pyloric sphincter may open spontaneously upon insufflation of air or with application of gentle pressure.

Passage through the pylorus can be seen as well as felt and is facilitated by using the single-handed technique. With this method the endoscopist uses one hand to manipulate the controls on the instrument head and the other hand to advance the instrument shaft through the pylorus. This allows the endoscopist to benefit from tactile input as well as visual cues. Upon entering the duodenal bulb, its characteristic granular, pale mucosa is recognized. The second portion of the duodenum is then entered by advancing the scope to the superior duodenal angle while simultaneously deflecting the tip and rotating the shaft to the right. At this point, gentle withdrawal of the endoscope paradoxically advances it further down the duodenum as the tip corkscrews around the superior duodenal angle. Once in the descending duodenum, the scope is advanced to its full length. Because of possible bowing in the stomach, the third and fourth portion of the duodenum may not be reached even though the full length of the scope has been inserted.

Optimal mucosal survey is performed upon scope withdrawal. The circumferential folds of the duodenum are noted along with their disappearance at the superior angle. The bile duct papilla can often be seen on the medial wall of the upper third of the descending duodenum. The antrum is carefully inspected, noting passage of peristaltic waves; these should be examined for symmetry of progression. In the gastric corpus, sufficient air is insufflated to flatten the rugal folds, allowing proper mucosal inspection.

Visualization of the cardia, proximal fundus, and the lesser curvature is difficult to achieve with a forward-viewing endoscope. Retroflexion of the endoscope while in the antrum allows visualization of these structures. This maneuver turns the instrument backward on itself and is performed by simultaneously fully flexing the tip upward and to the left while advancing the shaft. The shaft can then be rotated 180° in either direction to attain full visualization of the cardia and fundus.

Tissue sampling techniques

Tissue samples are most commonly obtained by directed biopsy (Fig. 1). Biopsies are taken using cupped forceps that are passed through the endoscope's therapeutic channel. Specimens are best taken from an *en face* position. Lesions that must be approached tangentially (e.g. esophageal) are often better obtained with spiked biopsy forceps. In either case, the forceps are applied with open jaws. Once in proper position, the jaws are carefully closed and withdrawn. For ulcerated lesions, the base and at least four quadrants of the rim are biopsied. Standard biopsy forceps rarely penetrate the muscularis mucosa. When deeper specimens are desired, ‘jumbo' forceps or those with a diathermy snare loop can be used.

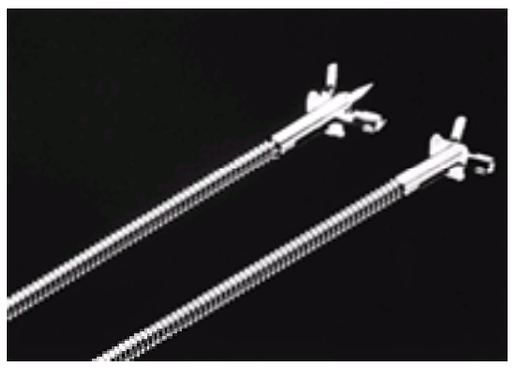


Fig. 1. Samples of two biopsy forceps (by courtesy of Wilson-Cook Winston-Salem, NC).

Brush cytology can also be used to procure tissue samples (Fig. 2). A sleeved brush is passed through the therapeutic channel toward the lesion, advanced out of the sleeve, and scraped over the lesion several times. The brush is then withdrawn back into the sleeve and then the assembly is withdrawn. The specimen is rubbed across glass slides which are rapidly fixed for cytologic processing. Diagnostic yield is increased if disposable brushes are also used. With these disposable brushes, the entire brush head can be transected and placed in fixative. This fluid can then be centrifuged and the cell block analyzed.

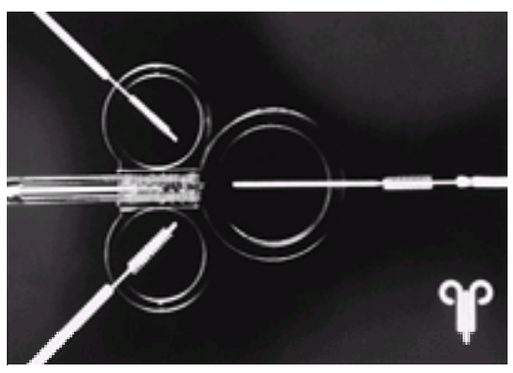


Fig. 2. Various brush cytology instruments (by courtesy of Wilson-Cook Winston-Salem, NC).

Various staining techniques have been described to visualize better mucosal abnormalities. Most involve spraying the mucosa with dyes (e.g. methylene blue, indigo-carmen, Congo red) to identify better mucosal abnormalities.

Hemostasis—non-variceal

Despite new developments in endoscopic therapy, the mortality from upper gastrointestinal hemorrhage has changed little in the past 30 years. The primary predictor of morbidity remains the magnitude of blood loss before the initial assessment. Endoscopy is critical in the initial evaluation and treatment of bleeding and should be performed soon after the patient has stabilized. Hemodynamic instability associated with hemetemesis or hematochezia clearly mandate endoscopic evaluation after stabilizing the patient. Upper gastrointestinal endoscopy accurately identifies the bleeding site and source in about 90 per cent of cases and may also reduce health care cost by accurately triaging patients' care on initial presentation. At the same time, successful treatment can ordinarily be achieved in most settings using a variety of techniques.

A thorough endoscopic examination usually reveals the bleeding source(s), rate of bleeding, and stigmas of recent hemorrhage (see below). Though there is a 70 to 85 per cent rate of spontaneous cessation of hemorrhage, certain endoscopic findings are associated with a higher probability of continued hemorrhage or rebleeding from ulcers. Pulsatile arterial bleeding has an 85 per cent risk of continued hemorrhage. A visible vessel has a rebleeding risk of about 50 per cent (Fig. 3). Lesions with an adherent clot or pigmented protuberance have a 20 to 40 per cent risk of rebleeding. A flat pigmented spot has a 5 to 10 per cent chance of continued bleed-ing. Lesions with a clean base are at low risk (5 per cent) of rebleeding (Table 2).

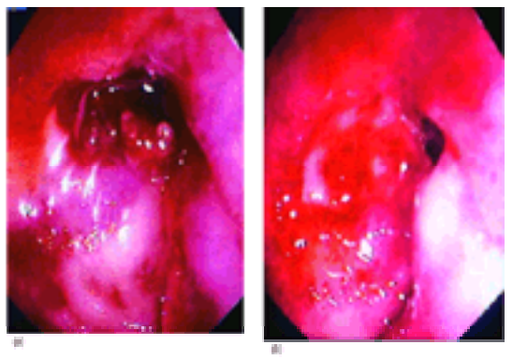


Fig. 3. (a) Endoscopic image of duodenal ulcer with visible bleeding vessel. (b) Duodenal ulcer and vessel after treatment with a heater probe. (By courtesy of Dr Steven Goldschmid.)

Stigmas of recent bleed	*Risk of rebleed (%)*
Active spurting/oozing	85
Visible vessel	50
Adherent clot	40
Pigmented blood spot	5–10
Clean base	5

Table 2 Stigmas of acute bleeding with associated risk of rebleeding.

Endoscopic hemostasis reduces the need for surgery, blood transfusion, and length of hospital stay. Lesions that suggest increased bleeding risk should be treated endoscopically, using either thermal or injection therapy; endoscopic therapy significantly decreases the risk of rebleeding.

Thermal therapies control bleeding by using heat to induce vessel shrinkage via tissue coagulation and collagen contraction. The two main thermal energy sources are light and electricity. With monopolar electrocoagulation, a probe is inserted through the endoscope and placed against the lesion's base. Electric current is transferred from the generator to the tissues and then exits via the patient's ground plate. Current is applied for 2 to 3 s, generating heat due to resistance. With this modality the temperature may reach several thousands of degrees, possibly causing full-thickness tissue damage, one of the major limiting factors of monopolar coagulation. Another drawback is that the probe tip often adheres to the tissue on contact, necessitating probe removal for cleaning. Success is usually achieved, however, in 80 to 96 per cent of cases.

The bipolar electrocoagulation probe, sometimes referred to as the bipolar circumferential active probe (**BICAP**), has at least two active electrodes in the probe tip around which current density is concentrated. Because of this ability to concentrate current locally, coagulation is achieved at lower temperatures (100°C). No grounding pad is needed since the current flows between each pair of electrodes. Therapy is applied in 1- to 2-s 50-watt bursts at lower power settings.

BICAP successfully stops bleeding in 92 to 94 per cent of cases. One advantage to both monopolar and bipolar coagulation is that they allow physical tamponade of the vessel as well as electrical heat application. Animal studies have shown that tissue coaptation before coagulation increases effectiveness. Though rebleeding after treatment can be significant, these methods reduce the overall bleeding recurrence from 50 to about 10 per cent.

The heater probe is another modality of coagulation therapy. With this device, a Teflon-coated cylinder containing an electric coil generates 5 to 30 joules of energy per pulse. As with BICAP and monopolar therapy, the vessel is tamponaded during therapy. Heat is then delivered in 6- to 10-s intervals sequentially at 20 to 30 J each application. With the Teflon coating there is less adherence of the tip to the tissues. Initial hemostasis is achieved in up to 100 per cent with ultimate hemostasis accomplished in 90 to 93 per cent of cases. The heater probe may be more effective than other strategies when used on actively spurting vessels.

Lasers generate heat as light is absorbed by tissues. The argon laser generates light at wavelengths of 500 nm. The argon laser's coagulation depth is limited to 1 mm because hemoglobin in red blood cells absorbs light in this wavelength. This light absorption decreases the amount of laser energy transmitted to the vessel. The neodymium: yttrium-aluminum-garnet (Nd:YAG) laser produces light in the infrared wavelength (1060 nm). It has a greater risk of full-thickness injury because this wavelength of light is not as readily absorbed by the red blood cells. Though equal in efficacy to other thermal therapies, high cost and limited portability are major disadvantages of lasers.

A variety of agents have been utilized as injectable hemostatic agents including 1 per cent polidocanol, 98 to 99 per cent dehydrated ethanol, 3.6 per cent hypertonic saline, 1 to 3 per cent sodium tetradecyl sulfate, and 1:10 000 epinephrine. There is little difference in the effectiveness of these agents. Active bleeding can be controlled in 95 per cent of patients. Shock and ulcer size greater than 2 cm appear to predict failure of injection therapy.

Both thermal and injection methods decrease the incidence of rebleeding, the need for emergency surgery, and even mortality. The benefit of endoscopic therapy is realized mainly in patients with high-risk stigmas (i.e. active bleeding or visible vessels). To date, results of trials comparing various thermal and injection therapies have failed to show significant differences in their efficacy. Recent reports, however, suggest that use of two agents with different mechanisms of action (e.g. coagulation and injection) may be more effective than a single agent alone.

Hemostasis—variceal

Hemorrhage from esophageal varices occurs in 30 per cent of patients with cirrhosis and is associated with a 30 per cent mortality. Endoscopic intervention plays a vital role in management of this serious clinical problem. Initially, volume replacement and appropriate repletion of coagulation factors are essential. The stomach is lavaged to clear blood clots. In patients with mental obtundation, endotracheal intubation will protect the airway. In the presence of massive bleeding, infusion of intravenous vasopressin or balloon tamponade may be necessary. Since complications and transfusion requirements may be increased when eosphageal tamponade precedes endoscopic intervention, this should be avoided when possible.

Upper endoscopy is performed to establish the diagnosis (Fig. 4). Upon confirmation of variceal bleeding, endoscopic treatment is undertaken. Injection sclerotherapy is usually effective and is the initial treatment of choice for control of acute esophageal variceal bleeding in most patients. Sclerotherapy significantly reduces rebleeding and provides a survival advantage when compared with no sclerotherapy. Nevertheless, rebleeding occurs rather frequently (30 to 50 per cent) following sclerotherapy; the rate is significantly greater than that associated with shunt surgery. Complications such as fever, chest pain, pleural effusion, and esophageal stricture are infrequently seen.

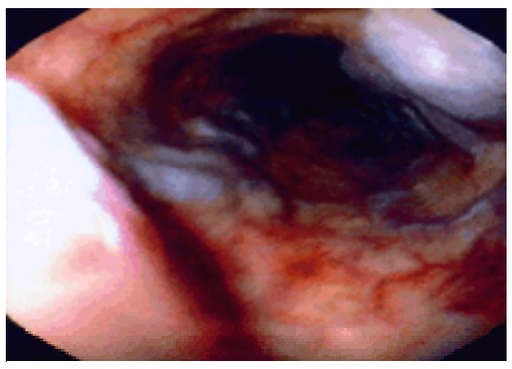


Fig. 4. Endoscopic view of prominent esophageal varices (by courtesy of Dr Steven Goldschmid).

At this time, no sclerosant is universally accepted. In the United States, the most commonly used agents are 0.7 to 3.0 per cent sodium tetradecyl sulfate and 5 per cent sodium morrhuate. A recent, prospective, randomized study reported equal efficacy between sodium tetradecyl sulfate and polidocanol, but favored polidocanol because of fewer complications. Ethanolamine is the most commonly used agent in Europe.

Sclerosants are injected either intravariceally or paravariceally. Intravariceal injection is preferred, since acute variceal bleeding is more effectively controlled (91 compared with 19 per cent, respectively). Injection is begun just above the gastroesophageal junction. The needle is placed into the varix and a test dose of 0.5 ml is injected to ensure intravariceal placement. Once placement is confirmed, an additional 1.5 ml is injected (Fig. 5). This technique is repeated circumferentially around the esophagus until all varices at this level have been treated. Injections are continued 2- to 3-cm proximally, again injecting circumferentially until only small caliber vessels are encountered. The total volume of sclerosant used should not exceed 20 ml per session. If active variceal bleeding is encountered, injections are begun just proximal to the bleeding site and continued until bleeding subsides. The session is usually repeated 5 days later and then continued at 1- to 3-week intervals until

varices are ablated.



Fig. 5. Variceal injection needle used during sclerotherapy.

Because of a notable incidence of complications and rebleeding, alternatives have been sought. Most recently, attention has focused on endoscopic variceal ligation. With this technique, varices are ligated with elastic bands similar to those used for hemorrhoidal banding (Fig. 6). Several reports of the technique suggest superior results in terms of decreased rebleeding, decreased esophageal stricture, and faster variceal obliteration.

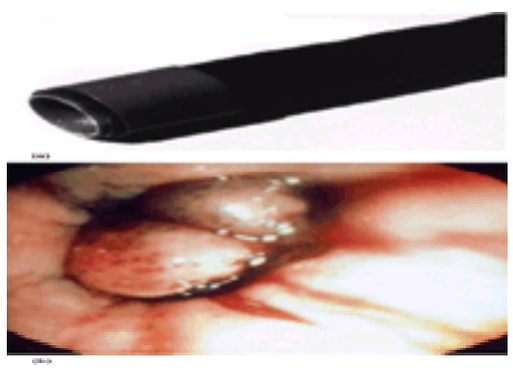


Fig. 6. (a) Variceal band loaded on band applicator. (b) Esophageal varix after band ligation. (By courtesy of Dr Steven Goldschmid.)

Prophylactic endoscopic intervention has been considered as a means of preventing initial bleeding episodes and improving overall survival in patients with cirrhosis. Results cast doubt on any benefit of prophylactic sclerotherapy. The Veterans Affairs Cooperative study was terminated prematurely owing to increased mortality in the prophylactic sclerotherapy group. It is conceivable, however, that prophylactic banding may prove more promising. Several reports demonstrate a reduced incidence of initial bleeding with prophylactic variceal banding and at least one report demonstrated a significant survival advantage. Confirmation of these results awaits future studies.

Propranolol may offer benefit for acute variceal bleeding, slowing acute bleeding as well as preventing hemorrhage. Propranolol may also reduce rebleeding when added to long-term sclerotherapy regimens. However, other studies reported conflicting results indicating the need for more prospective randomized data. Sclerotherapy is superior to propranolol alone in controlling variceal hemorrhage though neither intervention confers survival advantage.

Octreotide is another promising medical therapy which may prevent variceal rebleeding. Octreotide appears to be of greatest utility when combined with either sclerotherapy or endoscopic band ligation.

Stricture dilatation

Patients with symptomatic esophageal narrowing may require dilatation. Initially, a barium swallow should be obtained to visualize the obstruction and its length. Endoscopy with multiple biopsies and brushings then identifies the nature and severity of the stricture.

Esophageal strictures are most commonly a benign stenosis secondary to gastroesophageal reflux. Dilation resolves or improves dysphagia in 80 to 90 per cent of patients with benign strictures. Once moderate resistance is met during dilatation, no more than three dilators of successive size should subsequently be used. Dysphagia is ordinarily relieved when a luminal diameter between 13 and 15 mm (41 to 47 French) is achieved. Several sessions may be required to achieve this diameter.

Currently available dilators include the push type (mercury filled or guidewire driven), which apply both axial and radial forces, and the balloon type, which applies only radial forces. Mercury-filled dilators are most commonly used when dilating symmetric strictures that narrow the esophageal lumen to 12 mm or greater.

Guidewire-directed dilators are particularly useful for long, tortuous, or extremely narrow (less than 10 mm) strictures as well as those within hiatus hernias. These rigid devices are made of polyvinyl chloride and house a hollow core that allows passage over endoscopically or fluoroscopically placed guidewires. With guidewire techniques, routine fluoroscopy is seldom required except when the diameter of the stricture precludes endoscope passage.

Balloon dilators are useful in dilating short-segment or asymmetric strictures as well as stenotic stomas and narrowing due to achalasia. These dilators can be passed over a guidewire or through the endoscope's therapeutic channel. Balloon dilatation for achalasia provides long-term success in 75 per cent of cases, although several sessions may be required.

Malignant esophageal strictures may also benefit from endoscopic intervention. Following endoscopic assessment of tumor location and length, patients with dysphagia can be palliated with dilatation and, occasionally, insertion of an esophageal endoprosthesis. The optimal stent for palliating malignant strictures is not clearly established. Recent data favor expandable metallic stents in lieu of plastic prostheses, since the former appear to have lower morbidity. While laser ablation has also been used in this setting, its use is limited by the frequency of treatment, high expense, and significant incidence of recurrent obstruction. When esophageal tumors erode into the tracheobronchial tree creating a fistula, covered metallic stents offer valuable palliation.

Endoscopic ultrasound

Endoscopic ultrasound has become a valuable tool in the endoscopist's diagnostic armamentarium. A small high-frequency probe (7.5 to 12 MHz) is incorporated into the distal end of a conventional endoscope allowing resolution of structures as small as 2 to 3 mm in size. Due to the high resolution, however, the depth of penetration is limited to 5 to 6 cm beyond the luminal wall; this is usually quite adequate for endoscopic settings. Patients should be prepared as for standard endoscopy. Once the transducer is passed into the area of interest, all intraluminal air is removed; an acoustic interface is achieved between the transducer and the luminal wall by inflation of a water-filled balloon or instillation of water into the lumen. Using a perpendicular scanning plane, five layers of bowel wall are visualized with alternating echogenic and hypoechoic layers (Fig. 7).

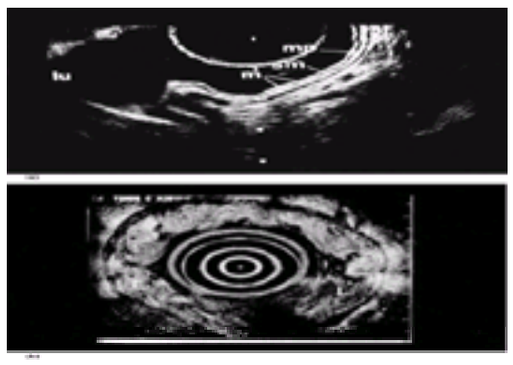


Fig. 7. (a) Layers of the bowel lumen as seen during endoscopic ultrasound. M, mucosal layer; SM, submucosal layer; and MP, muscularis propria layer. (b) An esophageal tumor is seen invading all wall layers.

Indications

There are three primary areas in which endoscopic ultrasound is currently most useful: evaluation of submucosal abnormalities, pancreaticobiliary disease, and staging of gastrointestinal tumors.

Submucosal abnormalities

Endoscopic ultrasound is very useful in its ability to distinguish between intrinsic and extrinsic compression. Since the layers of the bowel wall are clearly discernable using endoscopic ultrasound, visualization of five normal wall layers overlying an abnormality indicates extrinsic compression.

For intrinsic lesions, tumor size, contour, and wall layer of origin can be determined as well as whether the tumor is cystic or solid in nature. Tumors that are greater than 3 cm in size, irregularly contoured, or appear to invade surrounding structures are more likely to be malignant.

Gastric folds and varices are also noted using endoscopic ultrasound. Identification of gastric varices avoids potentially dangerous deep biopsies.

Pancreaticobiliary disease

Though endoscopic ultrasound is unable to differentiate reliably between focal pancreatitis and pancreatic cancer, it is highly sensitive in detection of small (even less than 1 cm) pancreatic tumors. Identification of portal vein invasion by pancreatic cancer can be accurately performed with endoscopic ultrasound; although determination of arterial involvement is less reliable due to less optimal celiac axis visualization. Endocrine tumors can be localized with 80 per cent sensitivity and 90 per cent specificity. Preoperative localization of these tumors helps with intraoperative identification and resection.

Detection of common bile duct stones by endoscopic ultrasound has comparable accuracy as with endoscopic retrograde cholangiopancreatography (**ERCP**). Although the complication rate is reported to be much lower with endoscopic ultrasound than ERCP, the latter allows concurrent extraction of most common bile duct stones.

Gastrointestinal tumor staging

The primary benefit of endoscopic ultrasound in gastrointestinal tumors is its accuracy in regional staging. It is 80 to 90 per cent accurate in estimating the depth of tumor invasion (T stage).

Endoscopic ultrasound is also highly accurate in predicting resectability of esophageal and gastric cancer (91 and 95 per cent, respec-tively). Likewise, with esophageal and gastric cancers endoscopic ultrasound is fairly accurate in staging nodal involvement. In comparative studies, endoscopic ultrasound proved superior to CT scanning in accurate staging of gastric and esophageal cancers. Endoscopic ultrasound is not definitive in assessing distant metastasis (M stage).

Further reading

Anonymous. Esophageal dilatation: guidelines for clinical practice. *Gastrointestinal Endoscopy* 1991; **37**: 122–4. [ASGE guidelines regarding the clinical application of esophageal dilatation.]

Anonymous. Consensus statement on therapeutic endoscopy and bleeding ulcers. *Gastrointestinal Endoscopy* 1990; **36**(5): S62–S65. [ASGE statement on the utility and efficacy of endoscopy in the management of bleeding ulcers.]

ASGE guidelines for clinical application: The role of endoscopy in the management of non-variceal upper gastrointestinal bleeding. *Gastrointestinal Endoscopy* 1992; **38**: 760–2. [Guidelines regarding the clinical application of endoscopy in the managment of non-variceal upper gastrointestinal bleeding.]

ASGE guidelines for training and practice: Sedation and monitoring of patients undergoing gastrointestinal endoscopic procedures. *Gastrointestinal Endoscopy* 1995; **42**: 626–9.

ASGE guidelines for training and practice: Antibiotic prophylaxis for gastrointestinal endoscopy. *Gastrointestinal Endoscopy* 1995; **42**: 630–5. [ASGE guidelines regarding sedation and monitoring practices for patients undergoing endoscopic procedures.]

Brady PG. Esophageal foreign bodies. *Gastroenterology Clinics of North America* 1991; **20**: 691–701.

Cello JP, Grendell JH, Cras RA, Weber TE, Trunkey DD. Endoscopic sclerotherapy versus portacaval shunt in patients with severe cirrhosis and acute variceal hemorrhage. Long term follow-up. *New England Journal of Medicine* 1987; **316**: 11–15.

Ellis FH Jr., Gibb SP, Balogh K, Schwaber JR. Esophageal achalasia and adenocarcinoma in Barrett's esophagus: a report of two cases and a review of the literature. *Diseases of the Esophagus* 1997; **10**: 55–60.

Hirschowitz BI. Development and application of endoscopy. *Gastroenterology* 1993; **104**: 337–42. [Describes the development of endoscopy and the early endoscopes.]

Jenkins SA *et al.* A multicentre randomised trial comparing octreotide and injection sclerotherapy in the management and outcome of acute variceal haemorrhage. *Gut* 1997; **41**: 526–33.

Lee JG *et al.* Endoscopy based triage significantly reduces hospitalization rates and costs of treating upper GI bleeding (UGIB): A randomized controlled trial. *Gastrointestinal Endoscopy* 1998; **47**(4): AB 54.

Lightdale C. Indications, contraindications, and complications of endoscopic ultrasonography. *Gastrointestinal Endoscopy* 1996; **43**(2): S15–S19.

Lin HJ, Wang K, Perng CL, Lee CH, Lee SD. Heater probe thermocoagulation and multipolar electrocoagulation for arrest of peptic ulcer bleeding. A prospective randomized comparative trial. *Journal of Clinical Gastroenterology* 1995; **21**: 99–102.

Llach J *et al.* A prospective randomized trial of heater probe thermocoagulation versus injection therapy in peptic ulcer hemorrhage. *Gastrointestinal Endoscopy* 1996; **43**: 117–120.

McLean A, Fairclough P. Endoscopic ultrasound—current applications. *Clinical Radiology* 1996; **51**: 83–98.

Morrissey JF, Reichelderfer M. Gastrointestinal endoscopy. *New England Journal of Medicine* 1991; **325**: 1142–9.

Offerhaus GJA *et al.* The risk of upper gastrointestinal cancer in familial adenomatous polyposis. *Gastroenterology* 1992; **102**: 1980–2.

Ramirez FC, Dennert B, Zierer ST, Sanowski RA. Esophageal self-expandable metallic stents—indications, practice, techniques, and complications: results of a national survey. *Gastrointestinal Endoscopy* 1996; **45**: 360–4.

Rollhauser C, Fleischer DE. Non-variceal upper gastrointestinal bleeding: An update. *Endoscopy* 1997; **29**: 91–105.

Sarin SK, Guptan RK, Jain AK, Sundarum KR. A randomized controlled trial of endoscopic variceal band ligation for primary prophylaxis of variceal bleeding. *European Journal of Gastroenterology*

and Hepatology 1996; **8**: 337–42.

Silverstein F, Gilbert DA, Tedesco FJ, Buenger NK, Persing J. The national ASGE survey on upper gastrointestinal bleeding, II: clinical prognostic factors. *Gastrointestinal Endoscopy* 1981; **27**: 80–93.

Soll AH. Medical treatment of peptic ulcer disease: Practice guidelines. *Journal of the American Medical Association* 1996; **275**: 622–9.

Steigmann GV *et al.* Endoscopic sclerotherapy as compared with endoscopic ligation for bleeding esophageal varices. *New England Journal of Medicine* 1992; **326**: 1527–32. [Prospective study comparing the use of endoscopic band ligation and endoscopic sclerotherapy in the management of acute variceal hemorrhage.]

Terblanche J, Stiegmann GV, Krige JE, Bornman PC. Long-term management of variceal bleeding: the place of varix injection and ligation. *World Journal of Surgery* 1994; **18**: 185–92. [Describes the role of injection sclerotherapy and variceal band ligation in the long-term management of esophageal variceal bleeding.]

The Veterans Affairs Cooperative Variceal Sclerotherapy Group. Prophylactic sclerotherapy for esophageal varices in men with alcoholic liver disease: a randomized single-blind, multicenter clinical trial. *New England Journal of Medicine* 1991; **324**: 1779–84. [Prospective randomized study of the efficacy of prophylactic sclerotherapy in treating bleeding esophageal varices in alcoholic portal hypertensive patients.]

Webb WA. Management of foreign bodies of the upper gastrointestinal tract. *Gastroenterology* 1988; **94**: 204–16. [Review of the management of foreign bodies of the upper gastrointestinal tract.]

14.2 Biliary and pancreatic endoscopy

Julian Britton

Introduction
Instruments
Indications
Preparation
Technique
Diagnosis
- **Endoscopy**
- **Radiology**

Treatment
- **Endoscopic sphincterotomy**
- **Biliary intubation**

Results
Complications
Further reading

Introduction

Endoscopic retrograde cholangiopancreatography (**ERCP**) is a combined endoscopic and radiologic procedure which plays an essential role in the diagnosis and the management of diseases of the biliary tract and pancreas. The diagnostic technique was first described in 1968 and this was quickly followed by the development of therapeutic endoscopic procedures for the relief of obstructive jaundice and the extraction of stones from the bile duct. Although ERCP requires expensive equipment and technical skill, it is now widely available and it should be regarded as complementary to percutaneous transhepatic techniques.

Instruments

Modern videoduodenoscopes are sophisticated side-viewing endoscopes with a miniature television camera incorporated in the tip of the endoscope. They are designed specifically for use within the duodenum. Older optical duodenoscopes are still available but the quality of the image is poor, particularly when a video camera is attached. The television image enables the endoscopy assistants to see what they are doing and it makes teaching the technique much easier. Duodenoscopes with instrument channels up to 4.2 mm in diameter are available. All of them provide adequate insufflation of air and suction whilst a cannula lies within the channel. They are fully insulated and completely immersible so that the whole instrument can be sterilized.

The basic diagnostic cannula is a 200-cm long plastic tube with an outside diameter of 1.6 mm and a single terminal opening with a rounded tip marked at centimetre intervals for the last 5 cm. This cannula will accept a 0.035-inch diameter guide wire (Fig. 1). Most endoscopic instruments are subsequently introduced over an exchange wire following the initial cannulation.



Fig. 1. A selection of endoscopic instruments. From the left they are: a taper tip cannula, a standard diagnostic cannula, a balloon catheter, a cannula with a 450 cm 0.035 inch diameter Jagwire^{TM} hydrophilic tip guide wire (note the metal collar on the cannula which enables easy visualization on the X-ray screen), and a Dormia basket.

The most important therapeutic instrument is the sphincterotomy knife (Fig. 2). This consists of a fine wire running down inside a plastic cannula and attached at the tip. For the final 3 cm the wire runs outside the plastic tubing. When the wire is shortened within the plastic tubing at the proximal end, the wire at the distal end lying within the ampulla bow strings across the tubing (Fig. 3). Other instruments are baskets and balloons of a conventional design and suitable length and strength (Fig. 1).



Fig. 2. A sphincterotomy knife in the relaxed position emerging from the instrument channel of a videoduodenoscope. The knife has been inserted over a 450 cm 0.035 inch diameter Zebra^{TM} exchange wire.

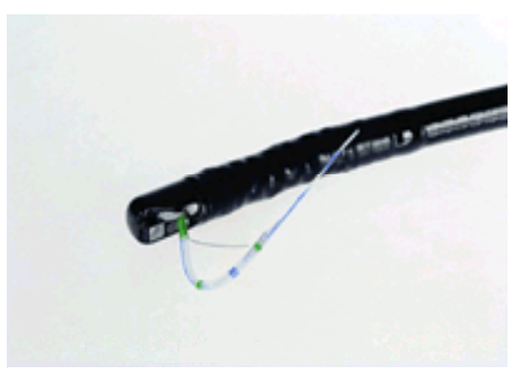


Fig. 3. A sphincterotomy knife in the tightened position.

A good-quality radiological screening unit with an image intensifier that will also take still radiographs is essential. The two television screens, one for the radiological image and the other for the endoscopic picture, should be placed behind the patient's head so that both screens are comfortably visible to the endoscopist and to the radiologist.

Indications

The common indications for ERCP are extrahepatic obstructive jaundice and cholangitis. In most circumstances the endoscopist will attempt to relieve the obstruction or to remove the stones at the same examination. A diagnostic examination is appropriate in any patient in whom biliary or pancreatic disease is suspected provided the same information cannot be obtained in a simpler or safer way. A preliminary ultrasound examination is always essential. In expert hands this is likely to make a diagnosis. Even if this is not possible the findings may influence management during the endoscopy.

An ERCP cannot be performed if the patient has oesophageal, gastric, or proximal duodenal obstruction. There are no other absolute contraindications. Patients with hepatitis or other infectious conditions should be examined at the end of a list and the endoscopes should be cleaned and sterilized immediately afterwards.

Preparation

Most patients of any age and in any condition can tolerate an ERCP. Those with respiratory impairment require careful sedation, but patients who are severely ill from cholangitis for example are often better treated, after resuscitation, by ERCP rather than conventional surgery. The prothrombin time and the platelet count must be within the normal range. Patients with jaundice receive vitamin K routinely and fresh frozen plasma is given immediately before the endoscopy if the prothrombin time remains abnormal. Patients are starved for at least 6 h beforehand and all of them receive intravenous fluids even though renal failure following an ERCP in a patient with jaundice is rare. When a therapeutic procedure is planned the patient must receive prophylactic antibiotics so as to minimize the risk of subsequent cholangitis. We presently use ciprofloxacin at a dosage of 750 mg given orally 2 h in advance or cefuroxime at a dosage of 1.5 g given intravenously 1 h in advance of the examination. Premedication is generally unnecessary.

Children are best examined under general anaesthesia but adults are sedated with intravenous pethidine (50 mg) and midazolam (10 mg). One endoscopy assistant equipped with a dedicated sucker is solely responsible for looking after the patient throughout the procedure. Supplementary oxygen is given to all the patients because respiratory depression can occur. Monitoring of the pulse rate and oxygen saturation are essential as the procedure is conducted in semidarkness. At the end of the examination we routinely reverse the action of the sedative drugs with naloxone (0.4 mg intravenously) and flumazenil (200 µg intravenously).

Patients can eat and drink as soon as they have fully recovered from the sedation. A diagnostic ERCP can be performed as a day case but patients who have a therapeutic procedure should stay in hospital for the night after the examination. Complications, if they occur, usually develop within 12 h of the procedure.

Technique

ERCP is technically the most difficult of all the common flexible endoscopic procedures. Patience, persistence, and perseverance are absolutely essential. Furthermore, a different endoscopist on a different occasion may succeed after an initial failure. An endoscopy assistant who is familiar with and manipulates the instruments is also essential.

The patient lies on the radiography table on their left side with their left arm behind them. The right arm lies up by the head with the intravenous cannula readily accessible to the endoscopist. The right knee and hip are fully flexed to begin with whilst the left leg lies straight. The pelvis then lies more or less vertical and leaves the abdomen free. From this position it is easy to lie the patient onto their face when the endoscope is in the duodenum simply by straightening the right leg. The weight of the patient on the abdomen then maintains the position of the endoscope across the stomach and within the duodenum. By good fortune the X-ray beam then usually lies in exactly the correct plane for radiology of both the bile duct and the pancreatic duct. A diathermy plate is placed around the left thigh when required.

Inserting the duodenoscope requires practice. The endoscopist starts by facing the patient and passing the endoscope into the stomach. The easiest way to enter the duodenum is to push the endoscope around the greater curve of the stomach and obtain a face-on view of the pylorus. The tip of the instrument is then elevated and the endoscope advanced. The pylorus disappears from view like the setting sun and the endoscope enters the duodenum. The tip of the endoscope is then rotated and locked to the right, the endoscopist twists his body and thus the whole instrument to the right and the excess length is withdrawn whilst maintaining a view of the duodenal lumen by manipulating the up/down control. During this manoeuvre the ampulla initially recedes from view but then the tip of the endoscope advances and comes to lie immediately opposite the ampulla. In this ‘short' position the endoscope lies in almost a straight line from the mouth to the duodenum with the ampulla about 65 cm from the incisor teeth (Fig. 4). This is the easiest position in which to cannulate the ampulla. Occasionally the ‘long' position is better, particularly for cannulation of the bile duct. Here the original position of the endoscope around the greater curve of the stomach is maintained and virtually the full length of the endoscope may lie within the patient (Fig. 5).

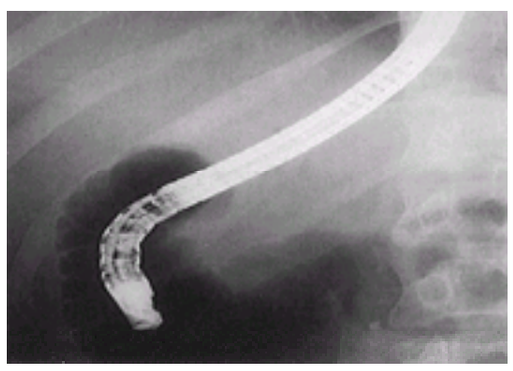


Fig. 4. Duodenoscope in the ‘short' position with the tip immediately opposite the ampulla and in the best position for cannulation.

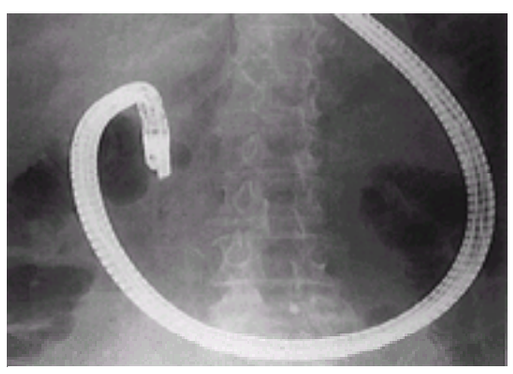


Fig. 5. Duodenoscope in the ‘long' position. It is sometimes easier to cannulate the bile duct from here although the instrument controls are always harder to work.

The key to successful cannulation is to position the endoscope correctly in relation to the ampulla. Until the endoscopist has a clear face-on view of the ampulla (Fig. 6) attempts at cannulation are less likely to be successful. Duodenal peristalsis is a nuisance. Persistent insufflation of air may overcome the contractions but the bowel can be paralysed with intravenous hyoscine-n-butylbromide (40 mg). Intravenous glucagon (1 mg) is an alternative.

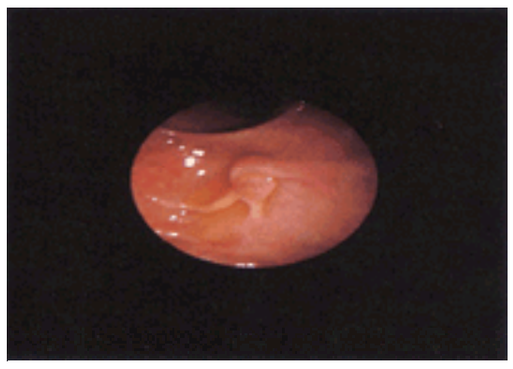


Fig. 6. A normal ampulla. This is a perfect position for cannulation of either duct.

There are a variety of manoeuvres that are sometimes helpful in achieving the correct position. Repeatedly pushing the endoscope further down the duodenum and then withdrawing as described before may help. Hooking the tip of the endoscope around the corner into the third part of the duodenum at the same time as shortening the endoscope is sometimes useful and maintaining the patient on their left side occasionally helps. Patients who have had a previous Billroth II or Polya gastrectomy have to be intubated through the gastroenterostomy and backwards along the afferent loop. This can be difficult if the afferent loop is exceptionally long.

Orientation during endoscopy is always difficult. It is best not to try and orientate oneself in relation to the patient. It is helpful to imagine sitting in the second part of the duodenum facing the ampulla on the medial wall with your legs lying along the third part of the duodenum towards the duodenojejunal flexure. From this position the pancreatic duct runs almost horizontally away from the eye towards the spleen whilst the bile duct passes vertically upwards behind the medial wall of the duodenum towards the liver, which is above the endoscopist's head.

If the ampulla cannot be easily seen the endoscopist should search the medial wall of the duodenum systematically starting distally and working proximally. The endoscope is gently twisted from side to side and any suspicious mucosal folds should be elevated with a cannula. Diverticula require special attention and it may be necessary to place the tip of the endoscope actually within a diverticulum to find the ampulla. The opening itself is usually obvious except with ampullary tumours, when gentle probing with a cannula may help. The accessory ampulla can usually be seen 1 to 2 cm proximal and slightly to the right or anterior side of the main ampulla.

Once the main ampulla has been found the tip of the cannula should be placed just within the orifice and a small bolus of contrast injected whilst screening the upper abdomen. If reflux of contrast into the duodenal lumen occurs, injection should stop immediately and the position of the cannula very gently adjusted before a further injection of contrast. When a duct fills, suitable radiographic pictures are taken. The endoscopist should then withdraw the cannula a little, reposition the endoscope and the cannula, bearing in mind the comments in the paragraph above, and so fill the other duct (Fig. 7, Fig. 8). Sometimes it is immediately possible to advance the cannula some way into a duct. If this happens it is wise to withdraw the cannula a little immediately prior to injecting contrast in order to be sure that the tip of the cannula is not jammed into a side branch of the main pancreatic duct. If cannulation continues to prove difficult, relaxation of the sphincter with glyceryl trinitrate applied either as a spray to the buccal mucosa (800 µg) or as a cream to the skin (16 mg) is worthwhile.

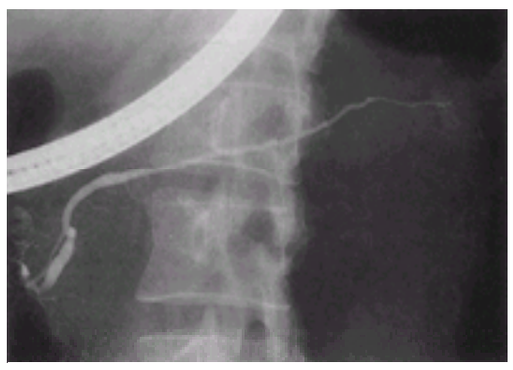


Fig. 7. A normal pancreatogram.



Fig. 8. A normal cholangiogram.

Diagnosis

Endoscopy

Although the examination is primarily radiological, the endoscopic findings should not be ignored. The oesophagus is never properly seen with a side-viewing endoscope and the stomach is rarely fully examined. However, the first and second parts of the duodenum are seen. Benign ulceration, malignant infiltration, and overt cancer may all be visible. Distortion of the duodenal anatomy and stenosis of the lumen may also be appreciated and may on occasion make it difficult to position the endoscope. Observation is particularly important with a carcinoma of the ampulla because histology and cytology may not initially make a definitive diagnosis (Fig. 9). All biopsies should be taken just before removing the endoscope when any bleeding will not interfere with cannulating the ampulla.

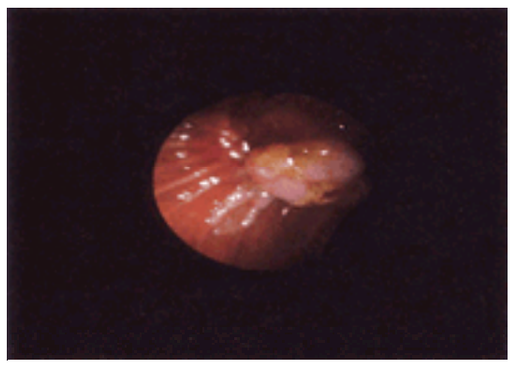


Fig. 9. A typical carcinoma of the ampulla.

The ampulla itself has a variable appearance. The most common appearance is shown in Fig. 6 where the ampulla appears as a nipple. Small mucosal fronds often pout out from the opening and there may be mild inflammation. Sometimes the ampulla is very flat and it is then often hard to find. A patulous, markedly inflamed, or oedematous appearance suggests either the recent passage of a stone or obstruction to lymphatic drainage by a tumour. Very rarely there may be two separate orifices corresponding to the openings to the pancreatic and bile ducts.

Radiology

If chronic pancreatitis is suspected, a plain abdominal radiograph must be taken first as calcification is rapidly obscured by contrast. During the procedure the assistance of a radiologist is invaluable so that the endoscopist can concentrate on the endoscopic appearances whilst the radiologist can position the patient and the X-ray beam to provide the best radiologic image. Any of the standard contrast media can be used but the concentration is important. If the contrast is too dense, stones in the bile duct may be obscured and if it is too dilute, there is a risk of overfilling the pancreatic duct. We use 30 per cent W/V meglumine iothalamate. Some endoscopists prefer to use a non-ionic contrast medium such as iopamide.

Attention to detail during the examination is essential and this requires careful co-ordination between the endoscopy assistant, the endoscopist, and the radiologist. The pancreatic duct must be filled right to the tail of the gland with some filling of the side branches but without filling the acini and producing a parenchymogram which will almost always lead to pancreatitis.

During cholangiography, early screening may be helpful but it may also be deceptive. If a stone is present in the bile duct it may rise up in the duct as the heavy contrast is injected, but the lucent whirlpool which is created when a fine jet of contrast is injected into a dilated duct may mimic a stone. Certainly, no therapeutic procedure should be performed until the diagnosis of choledocholithiasis has been definitely established (Fig. 10).

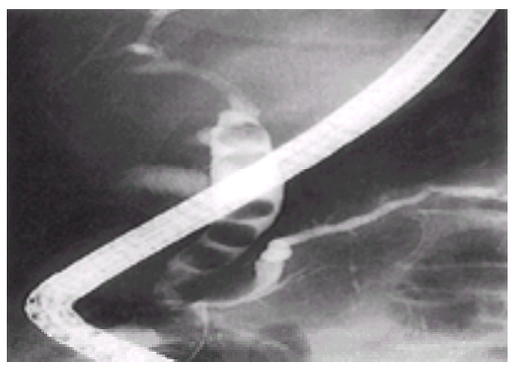


Fig. 10. A cholangiogram showing multiple stones within the bile ducts.

Ideally, contrast should outline both sides of a stricture because the diagnosis may rest almost entirely on the radiologic appearances. If this does not occur in the pancreatic duct, the proximal side branches must be well filled or even slightly overfilled before an apparent narrowing or obstruction of the main duct can be accepted as significant. In the bile duct it may be necessary to jam the cannula deliberately into the bottom end of a stricture and then inject contrast under a little pressure in order to obtain satisfactory pictures (Fig. 11).



Fig. 11. A cholangiogram showing a typical hilar stricture probably due to cholangiocarcinoma.

Sometimes it is impossible to outline any duct. If there is a definite abnormality on ultrasound, it is usually justified to make a small precut in the tip of the ampulla either with a needle knife (Fig. 12) or a sphincterotomy knife. Immediate cannulation of one or both ducts may then be possible, but it is often easier a few days later when there is some oedema of the ampulla. The use of this manoeuvre lessens with increasing experience.

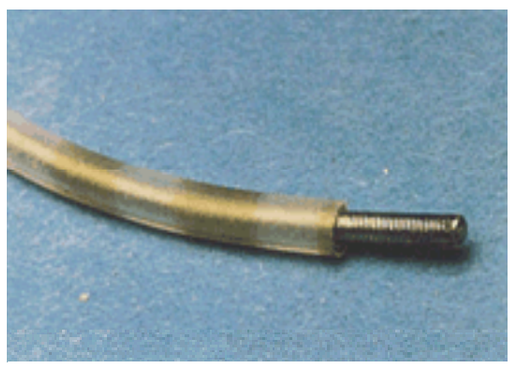


Fig. 12. A needle knife. A bare wire emerges from the end of the plastic cannula. Used for making a precut at the tip of the ampulla.

Treatment

Endoscopic sphincterotomy

The basic therapeutic manoeuvre is to divide the ampullary sphincter and gain access to the bile duct. The relaxed sphincterotomy knife (Fig. 2) is inserted over the exchange wire into the bile duct and the position confirmed on radiography. The knife is slowly withdrawn until about half the wire is visible outside the ampulla. The knife is gently tightened by the endoscopy assistant and the endoscopist divides the sphincter by using short bursts of cutting diathermy current. It is not necessary to remove the exchange wire whilst this is done. Three manoeuvres will extend the cut upwards through the sphincter: elevating the bridge, elevating the tip of the endoscope, and tightening the wire. Only one of these manoeuvres is used at a time so that the sphincter is divided slowly and in a controlled way (Fig. 13). If they are combined there is a real risk that the wire will cut too far too fast. The larger the cut the greater the risk of haemorrhage and perforation. It is rarely necessary to enlarge a sphincterotomy beyond the transverse mucosal fold which lies immediately above the ampulla. It is normally sufficient if the fully tightened knife will come through the opening easily. A sphincterotomy needs only to be big enough to allow removal of the largest stone.

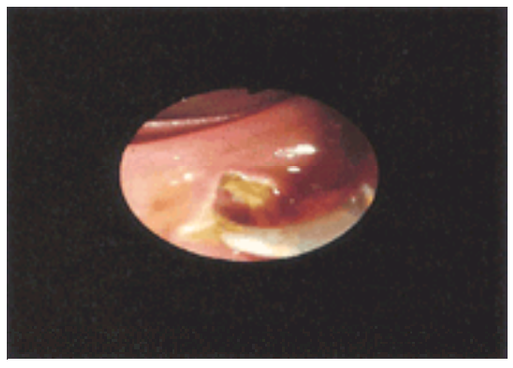


Fig. 13. The ampulla being divided with a sphincterotomy knife.

Stones are removed from the duct with a balloon or a basket (Fig. 14). Very large stones and stones which are wider than the diameter of the proximal bile duct are not easy to remove. Such stones can sometimes be reduced in size by dissolving agents instilled down a nasobiliary drain (Fig. 15), or by crushing baskets, laser light, ultrasound, and extracorporeal shock waves. None of these methods is particularly successful. If the duct cannot be completely cleared it is usually wise to leave a stent in place to allow adequate drainage whilst the further management is considered. In elderly patients with large stones a permanent stent may be the best definitive treatment.

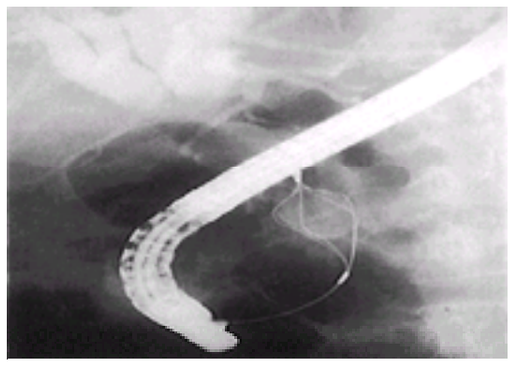


Fig. 14. A large stone being removed from the bile duct using a Dormia basket following a sphincterotomy.

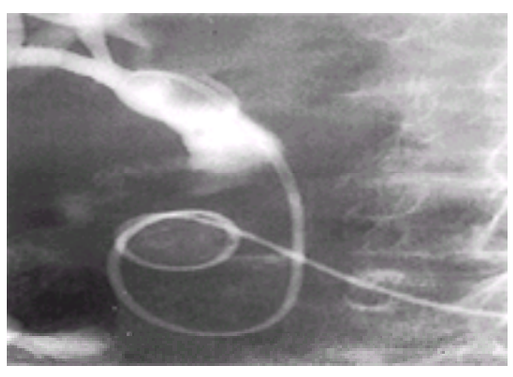


Fig. 15. A nasobiliary drain *in situ.* Mono-octonoin was infused down this tube and reduced the stone in size sufficiently for it to be removed endoscopically.

Biliary intubation

The other fundamental procedure is intubation of a duct with a flexible guide wire carried within a cannula. This is the basis for inserting any form of drainage into a duct and is most often used to place a stent across an obstruction in the bile duct caused by a carcinoma of the pancreas (Fig. 16). The tip of the cannula is placed below the obstruction. A suitable guide wire, preferably with a hydrophilic tip, is manipulated across the stricture by careful co-ordination between the endoscopy assistant and the endoscopist and is then followed by the cannula. The stent is railroaded over the cannula and guide wire across the stricture and carefully positioned.

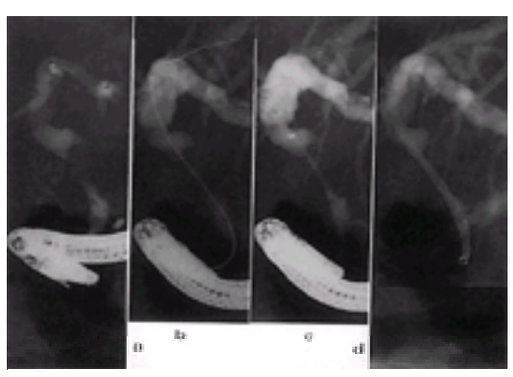


Fig. 16. Intubation of a biliary stricture. (a) Cholangiogram showing the obstruction. (b) Cannula below the stricture and the guide wire across the stricture. (c) Cannula and guide wire both across the stricture. (d) Stent railroaded into place.

Various designs of stent are available. The commonest material in use is polytetrafluoroethylene plastic tubing. Different diameters and lengths of stent are available. The larger the diameter the longer they last. Our standard stent is 10 fr in outside diameter (3.3 mm) and 15 cm long. They last about 4 months. Recurrence of jaundice or cholangitis are signs that the stent needs changing. This is easily achieved by removing the old stent with a basket and inserting a new one using the original technique.

Self-expanding metal stents are also available (Fig. 17). They are placed in an identical manner and either expand as a result of their innate properties or are dilated in place with a balloon. They last longer than plastic stents but cannot be removed. If they occlude, another stent has to be placed inside the old one. They are also expensive. They are useful for the long-term treatment of benign biliary strictures.



Fig. 17. A self-expanding metal stent placed across a malignant biliary stricture. This stent was placed percutaneously.

Results

The ampulla can be found by an experienced endoscopist in 98 to 99 per cent of patients. They will expect to cannulate the desired duct in 90 to 95 per cent of patients although the inexperienced find it slightly easier to cannulate the pancreatic duct. Most series report a 90 per cent success rate in performing a sphincterotomy and complete clearance of stones from the bile duct in 9 out of 10 patients. Half the failures are because of large stones. Endoscopic intubation of the bile duct for malignant obstruction is rather less successful with up to a quarter of attempts failing. Strictures high up and very low down in the bile duct are particularly difficult. Occasionally a combined percutaneous and endoscopic approach is required. The interventional radiologist passes a guide wire through the liver into the bile duct and across the stricture. The endoscopist retrieves the wire in the duodenum and brings it back up through the endoscope. It is then possible to railroad a stent across the obstruction.

Complications

The main complications are cholangitis, pancreatitis, and haemorrhage. The morbidity and mortality for a diagnostic ERCP are about 1 per cent and 0.1 per cent, respectively, whereas following an endoscopic sphincterotomy the figures are 10 per cent and 0.5 per cent. There is a particular risk of haemorrhage after a sphincterotomy and a small proportion of these patients will require a laparotomy. Less frequent problems are retroperitoneal perforation of the duodenal wall, which is a particular risk in the presence of a diverticulum, and impaction of a basket because a stone is trapped but is too large to remove.

Haemorrhage from a sphincterotomy is usually apparent immediately. It can often be stopped with coagulation diathermy applied through a pair of hot biopsy forceps either at the initial examination or later when further haemorrhage becomes apparent. Pancreatitis becomes clinically obvious between 6 and 12 h after the examination. Cholangitis can develop at any time. For all these reasons we record the pulse and blood pressure half hourly for 8 h after a therapeutic ERCP.

Further reading

Classen M, Demling L. Endoskopische Sphinkterotomie der Papilla Vateri und Steinextraktion aus dem Ductus choledochus. *Deutsche Medizinische Wochenschrift* 1974; **99**: 496–7. [The original description of endoscopic sphincterotomy and extraction of bile duct stones.]

Cotton PB. Progress report ERCP. *Gut* 1977; **18**: 316–41. [A review of the early results of ERCP with a comprehensive reference list.]

Cotton PB, Williams CB. *Practical gastrointestinal endoscopy*, 4th edn. Blackwell Scientific Publications, Oxford, 1996. [A detailed description of all endoscopic techniques including ERCP.]

Freeman ML *et al.* Complications of endoscopic biliary sphincterotomy. *New England Journal of Medicine* 1996; **335**: 909–18. [A prospective study of endoscopic sphincterotomy in 2347 patients, of whom 229 suffered a complication and 55 died.]

McCune WS, Shorb PE, Moscovitz H. Endoscopic cannulation of the ampulla of Vater: a preliminary report. *Annals of Surgery* 1968; **167**: 752–6. [First report of ERCP.]

Shepherd HA, Royle G, Ross APR, Diba A, Arthur M, Colin-Jones D. Endoscopic biliary endoprosthesis in the palliation of malignant obstruction of the distal common bile duct; a randomized trial. *British Journal of Surgery* 1988; **75**: 1166–8. [A randomized trial to demonstrate the benefits and disadvantages of endoscopic biliary stents.]

Soehendra N, Reynders–Frederix V. Palliative bile duct drainage—a new method of introducing a transpapillary drain. *Endoscopy* 1980; **12**: 8–11. [The first report of endoscopic biliary stenting.]

14.3 Colonoscopy

Kenneth A. Forde

Introduction
Contraindications
Indications
Equipment
Preparation and procedure
Complications
Perforation
Bleeding
Colonoscopy in selected circumstances
Colostomates
Strictures
Bleeding
Intraoperative colonoscopy
Polypectomy
Endoscopic management of colonic bleeding
Decompression of non-obstructive ileus and volvulus
Conclusion
Further reading

Introduction

Colonoscopy has now become the optimal diagnostic tool for most diseases of the large bowel and a major therapeutic modality as well. Certainly, small lesions (less than 1 cm), and mucosal as well as some submucosal conditions (colitis, angiodysplasia), are better demonstrated by colonoscopy. While barium enema is only a diagnostic tool, with the colonoscope one can biopsy, remove, ablate, dilate, and even at times achieve hemostasis. Barium enema, being more universally available and less costly, requiring less vigorous preparation and perhaps safer, is still frequently compared with colonoscopy. Since in all colonoscopies there are potential blind spots (at flexures, for example) and occasionally the cecum cannot be intubated, there is still occasional need to complement colonoscopy with contrast radiography. However, the barium enema is a weaker modality in the rectosigmoid colon, where extensive diverticulosis and overlapping folds can result in missed lesions. Furthermore, any imaging modality, be it barium contrast radiography, computed tomography, magnetic resonance imaging or (under current development) ‘virtual colonoscopy', will still require confirmation of the diagnosis or institution of therapy (polypectomy) for the resolution of most abnormalities.

All this explains why increasingly the sequence of sigmoidoscopy (rigid or flexible) followed by barium enema is being replaced by colonoscopy in the primary evaluation of the colonic disease.

Contraindications

Colonoscopy is not without risk and there are circumstances in which it should not be undertaken or in which risks should be more thoroughly balanced.

Absolute contraindications include inadequate preparation, fulminant colitis, when perforation is suspected or peritonitis established, and in the presence of a recently created intestinal anastomosis.

Relative contraindications include acute inflammation (a limited, careful examination by an experience endoscopist may be vital to the diagnosis of ischemia or acute colitis) and patients who may not be treated for complications of the examination (e.g. perforation).

Indications

The indications for colonoscopy are ever expanding and now include, for diagnostic colonoscopy, the evaluation of signs, symptoms or history of colitis, resolution of abnormal findings from barium enema, evaluation of all categories of bleeding (although colonoscopy during acute bleeding requires greater skill and experience), surveillance of high-risk patients for adenomas and cancer (Fig. 1, Fig. 2, and Fig. 3), intraoperative localization of non-palpable lesions (in open or laparoscopic procedures), or for evaluation of enigmatic bleeding.

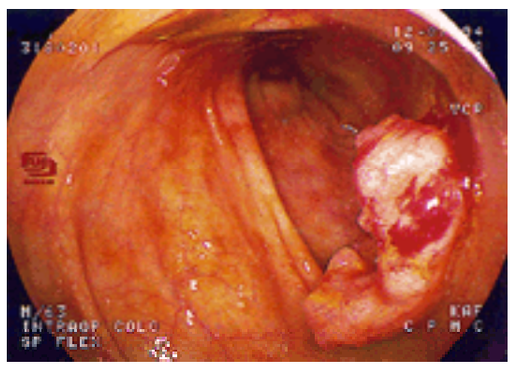


Fig. 1. Ulcerated splenic-flexure carcinoma.

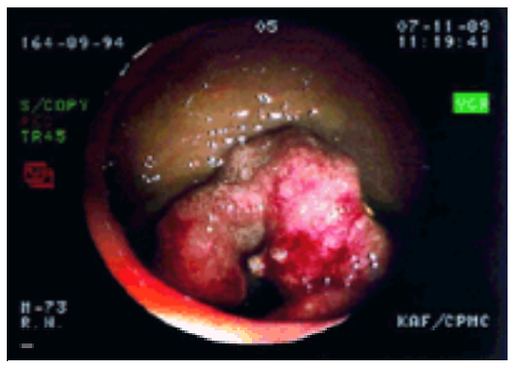


Fig. 2. Large, exophytic carcinoma.

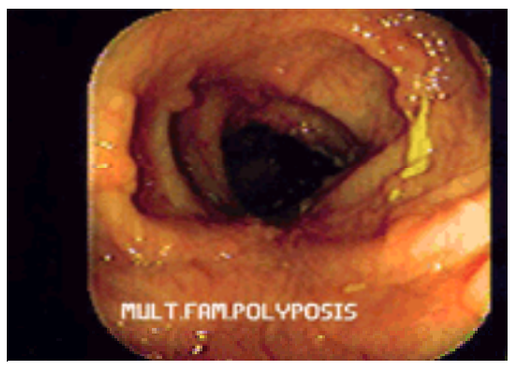


Fig. 3. Multiple familial polyposis: note innumerable tiny polyps, all adenomas.

Indications for therapeutic colonoscopy may be divided into the categories of:

1. hemostasis (vascular malformations, polypectomy site, malignant lesions);
2. resection and ablation (polyps, malignancy);
3. decompression and recanalization (volvulus, pseudo-obstruc-tion, stricture malignancy, foreign body).

Equipment

For the satisfactory performance of colonoscopy, especially involving treatment, it is necessary to have available an array not only of endoscopes but also of important accessories. Colonoscopes are now available as fiberoptic or video instruments. Features of the video colonoscope that make it far superior include interaction with computers to the extent of generating reports, scheduling, maintaining databases as well as providing superior image display. At least 160 cm of working length is desirable and an instrument channel of at least 3.7 mm in diameter. One needs an assortment of biopsy forceps, grasping forceps, cytology brushes, polypectomy snares, retrieval devices (baskets or graspers), injection needles, possibly laser fibers, and at least a monopolar cautery device.

Preparation and procedure

Thorough mechanical bowel preparation is required before colonoscopy and is currently most efficiently obtained with rapid gut lavage by ingestion of a polyethylene glycol electrolyte lavage solution. This is the most physiologic preparation, does not require enemas, and can be achieved within a few hours prior to the procedure. However, alternatives do exist and are employed, including the use of saline cathartics and enemas. Intravenous narcotics and sedatives are commonly administered, but many patients undergo the procedure without any medication. It is customary to monitor the patient during the procedure with pulse oximetry and sphygmomanometry. Electrocardiography is employed when indicated. While some regard fluoroscopy as dispensable, in a surgical endoscopic area where manipulation is sometimes performed (as for management of structures), it is useful to have fluoroscopic capability. It is of course mandatory if one uses the external splinting device that is available for stiffening the sigmoid colon. Fluoroscopy is also helpful in locating lesions that might require surgical intervention. The endoscopist, be they gastroenterologist, surgeon or radiologist, should be qualified to perform the procedure and should have received training in the use of electrocautery as well as fluoroscopic equipment. Staff members in the endoscopy unit should be certified in the performance of cardiopulmonary resuscitation. Modern colonoscopes are constructed so that the instrument controls can be manipulated with the left hand; the right hand is available for insertion of the instrument as well as twisting to the left or right (torquing).

The patient is customarily placed in the left lateral recumbent position to begin the procedure, unless a stoma is being intubated. The perianal area is examined, rectal examination performed, and the instrument is inserted. Once the rectum has been intubated, a small amount of air is inflated to visualize the lumen and the instrument is progressively advanced with the right hand, the tip being deflected with the controls of the left hand. Advancement involves not only forward insertion of the instrument, but also combinations of jiggling the instrument to and fro and any motion that will permit telescoping of the bowel over the endoscope. The goal of insertion is not so much complete visualization but rather to achieve the highest level of intubation desired as efficiently as possible. Careful inspection is made on withdrawal of the endoscope. While some areas of the colon may be clearly identified endoscopically (Fig. 4, Fig. 5, Fig. 6 and Fig. 7), in general the endoscopist tends to be inaccurate in determining the exact location of the instrument, especially in areas of the sigmoid, transverse, and right colon. In fact, reliance on mensuration on the colonoscope is notoriously inaccurate, as is the unfortunately common practice of identifying transillumination in the right lower quadrant as an indication of reaching the cecum. The landmarks on reaching the cecum include visualization of the ileocecal valve and the appendiceal orifice. Precise localization may not be as necessary for a lesion seen anywhere in the right colon, since the surgical procedure will be usually the same. However, it is important to determine where a lesion is in the transverse or sigmoid colon, since different decisions may be made based on such localization. Among methods of location of the site of a lesion are fluoroscopy (if one is skilled in its use and recognizes alterations, for example, in the sigmoid that could have occurred during intubation and appropriate straightening), the injection of vital dyes in the wall of the colon in the vicinity of the lesion, or locating the lesion or site intraoperatively with the colonoscope (Fig. 8). Obstacles to reaching the cecum include poor preparation, strictures, large tumors, and inability to straighten the sigmoid as from previous pelvic operations, radiation or sometimes from extensive diverticular disease (Fig. 9).

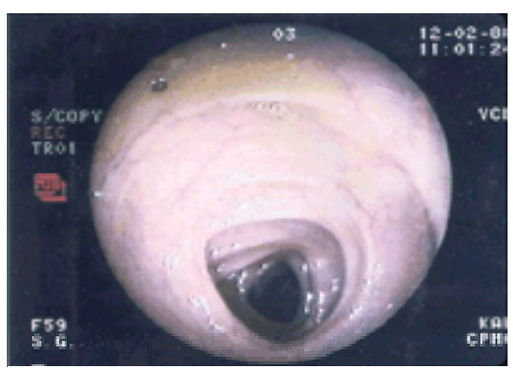


Fig. 4. Endoscopic view of a normal left colon; typical mucosal folds are visible.

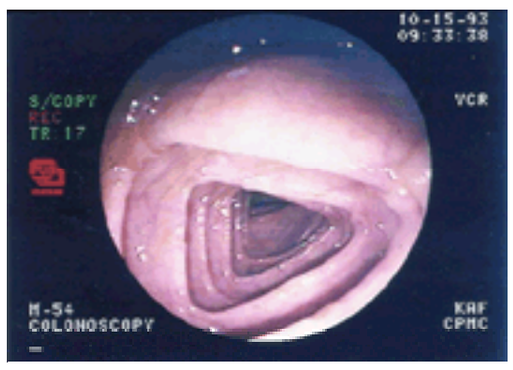


Fig. 5. The transverse colon often displays this triangular appearance.

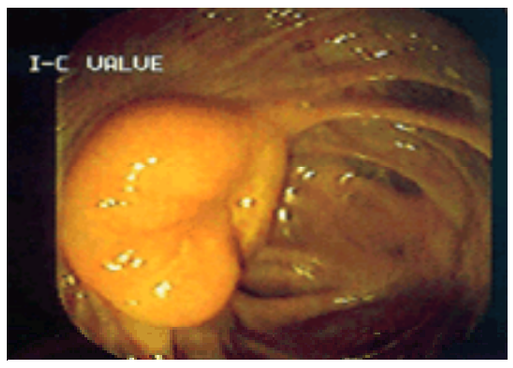


Fig. 6. Cecum, with ileocecal valve, appendiceal orifice, and diverticulum.

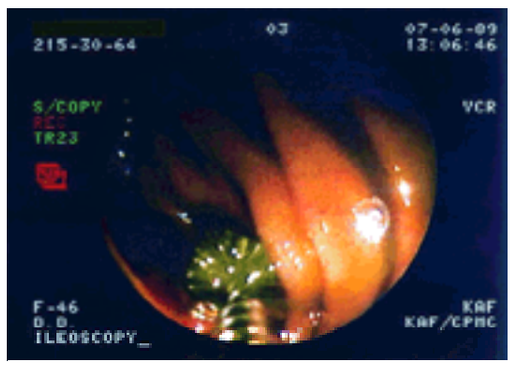


Fig. 7. Terminal ileum, with typical stippled appearance due to villous mucosa and submucosal lymphoid follicles (colonic mucosa, in contrast, is smooth).

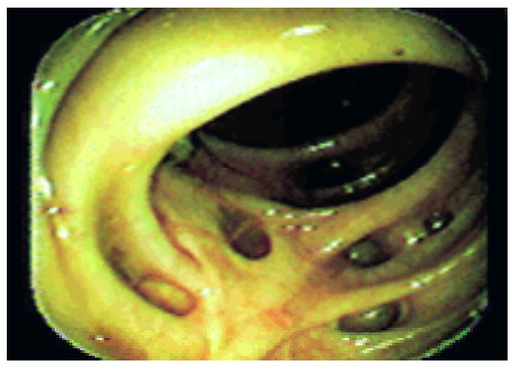


Fig. 8. Diverticulosis.

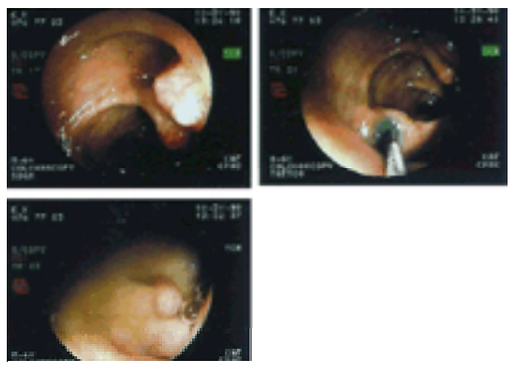


Fig. 9. Suspicious polypoid sigmoid lesion: tattooing the wall opposite the lesion, should resection be later required.

Complications

The most serious complications of colonoscopy are perforation and bleeding.

Perforation

Perforation occurs during diagnostic procedures in some 0.1 to 0.5 per cent of examinations. Causes of perforation include longitudinal tears of the mesentery from excessive looping, especially in the sigmoid, distraction of the bowel wall from a similar maneuver, or impaction of the tip of the endoscope in a diverticulum with subsequent disruption of the bowel wall. Manipulation of irradiated or inflamed bowel, of course, poses a greater risk. This type of perforation at colonoscopy usually requires prompt surgical intervention and repair, often without the need for colostomy as the bowel will have been well prepared mechanically. Perforation following polypectomy, on the other hand, is usually from transmural thermal injury and may be delayed. Management will depend on the patient's status at the time perforation is recognized. Certainly the presence of signs of peritoneal irritation mandate intervention.

Bleeding

While significant bleeding is rare after diagnostic colonoscopy, even when multiple biopsies have been performed, hemoperitoneum from splenic capsular tear has been reported. Postpolypectomy bleeding may be early or late. The overall reported incidence is between 1 and 2.5 per cent. The most immediate hemorrhage may occur from transection of the pedicle of a polyp; regrasping of the pedicle without additional cautery is usually satisfactory to control this, but does require the patience of 10 to 15 min of regrasping. Other techniques for arrest of immediate bleeding include injection of the base of the pedicle or polypectomy site with a 1/10 000 solution of epinephrine (adrenaline). It is rare for the patient to need angiography, intra-arterial infusion, or even surgical intervention. Delayed hemorrhage accounts for most episodes of postpolypectomy bleeding: it occurs most commonly in the first 24 to 48 h, but may also occur up to 14 days following polypectomy, presumably due to separation of the eschar from the polypectomy site. Delayed bleeding is usually controlled with bowel rest. The patient may need transfusion because of rapid loss of a significant amount of the blood volume but, again, rarely is surgical intervention necessary. Numerous other complications have a much lower reported incidence: they range from reaction to medication, instrument failure, bacteremia, incarceration of the endoscope in a hernia, to appendicitis.

Colonoscopy in selected circumstances

Colostomates

In the patient with a colostomy, colonoscopic examination is better tolerated since, in the absence of a sphincter, air is not trapped and usually escapes freely around the instrument. In fact, more insufflation is often necessary to complete the examination. Because the stoma may be somewhat strictured in relation to the tip of the colonoscope, it may be necessary to infiltrate the area with a local anesthetic and either dilate the stoma or release the skin. Subcutaneous herniations of the bowel may frustrate the colonoscopist on beginning the examination.

Strictures

The colonoscopic evaluation and management of strictures requires special expertise and has its own hazards and limitations. In the examination of a patient in whom one encounters a stricture or knows that a stricture exists from prior evaluation or from barium enema, insufflation has to be more cautious to prevent overdistension of the bowel proximally. Overdistension may result in perforation. If one cannot traverse a stricture, cytologic evaluation may be very useful in establishing the diagnosis of malignancy. Biopsy of an area of stricture is usually unproductive unless one obtains tissue from within the strictured area. Some strictures may be treated endoscopically by a variety of methods, depending on the nature of the stricture and the resources of the endoscopist. Techniques vary from through-the-scope balloon dilatation, alongside-the-scope dilatation with esophageal-type dilators, and ablation with electrocautery or laser; even dilatation with a prototype endoscope that has replaceable ‘olives', similar to an Eder–Puestow dilator, has been successful. The management of strictures often involves injection of contrast material or passage of guide wires, and it is therefore necessary to have fluoroscopic capability. Strictures often need management, not necessarily because they are symptomatic but because one needs to gain access to the bowel proximal to them for cancer surveillance or the removal and retrieval of known lesions. While symptomatic strictures, especially above the rectum, often come to surgical intervention, there are some circumstances in which it may be more desirable to manage narrowing of the bowel non-operatively. The patient with recurrent or inoperable cancer is such an example. Using one of the thermal ablation techniques endoscopically, one may be able to avoid a colostomy in the remaining few months of a patient's life. Endoscopic laser ablation of tumor has also been performed preoperatively to avoid a preliminary colostomy in patients with obstructing cancer. This approach, however, is fraught with increased risk: the combination of thermal energy and an obstructed bowel that may contain significant amounts of methane creates is likely to result in explosion.

Bleeding

Colonoscopy has become indispensable for the diagnosis of all categories of suspected lower gastrointestinal bleeding; while, ideally, the patient should be endoscoped when the bowel is not only devoid of fecal material but also completely empty even of blood, there are times when the clinician is faced with the consideration of colonoscopy in the presence of acute, persistent hemorrhage. These are patients who are usually severely ill, sometimes in shock, having often undergone angiography that was negative or for whom this examination has not been available. Since, in the presence of acute bleeding, anatomical landmarks and the whole mucosa often cannot be visualized fully, this type of examination should only be done by an endoscopist with sufficient skill and experience. However, its usefulness in this small group of patients for whom it may be necessary has been amply demonstrated: it may either direct a surgeon to precise limited resection of an affected area or obviate the need for surgical intervention by demonstrating that bleeding is not from the colonic mucosa or from a disorder that can be successfully managed by resection. While it is usually virtually impossible to achieve endoscopic hemostasis in the presence of massive bleeding, lesser degrees of oozing, as from a polypectomy site and even occasionally from a diverticular orifice, may be managed endoscopically by one of the methods mentioned above. Perhaps the greatest advantage of colonoscopy during active colonic bleeding is the assessment of whether bleeding is continuing, whether it is or has been from a localized site or area, or whether from above the level of the colon. Intraoperative colonoscopy may be also performed for acute hemorrhage.

Intraoperative colonoscopy

In addition to acute and unresolved bleeding, sometimes intraoperative colonoscopy is necessary in total endoscopic evaluation of the gastrointestinal tract for enigmatic bleeding that, although not acute, has resulted in the need for multiple transfusions. These are rare circumstances, but require significant preparation and cooperation between the surgeon and endoscopist (whether surgeon or gastroenterologist) and the anesthesiologist. A regular adult colonoscope may be passed orally in the intubated patient and can usually be negotiated to the distal ileum. If necessary, a colonoscope can then be passed per anum and advanced into the distal ileum so that total examination is achieved. However, prior to passage of the instrument in either direction, it is important to place an occlusive clamp on the distal ileum to prevent over insufflation of bowel that is not being examined distally or proximally. The more common reasons for erforming intraoperative colonoscopy are as follows.

1. Identification of a polypectomy site in a patient who has undergone colonoscopic polypectomy with finding of an invasive cancer, where it has been deemed necessary to proceed with colectomy. Since the polyp will have been removed, the surgeon will not be able to see or palpate a lesion and earlier efforts at identification of the site (be it fluoroscopic or injection of vital dye) may not have been successful. The endoscopist as well as the surgeon and the pathologist all have to be able to recognize the varying endoscopic (and gross) appearances of a polypectomy site.
2. Locating a non-visualized or non-palpable tumor at laparoscopic or open operation: now that laparoscopic colon resection is being increasingly performed, the ability of the surgeon to localize even a known tumor is limited since the palpating hand is not available; intraoperative colonoscopy is available to satisfy this need.

Polypectomy

While the rationale for removal of colonic polyps in general is cancer prevention, it is necessary for the endoscopist to recognize that there are a variety of polypoid lesions of the bowel that do not need or require excision. Lesions such as lipomas, inflammatory polyps, areas of lymphoid hyperplasia, small carcinoids, and ureteric stomas in patients who have undergone previous ureterocolonic anastomosis, to name a few. The endoscopist, however, may not be able to identify precisely whether a diminutive polyp is an adenoma or a hyperplastic polyp, or a combination of both, without biopsy at least. Most diminutive polyps (up to approximately 4 mm) may be removed by ‘cold biopsy'. We prefer the biopsy forceps with the central needle that allows the tissue to be impaled and then grasped with the jaws of the forceps, thus preserving the specimen for analysis. While coagulation biopsy forceps are available, overzealous use of them may result in bleeding or perforation and, in general, they are not necessary for removal of these tiny excrescences. Larger polyps, however, require electrocautery and snare removal in one or more fragments. Fortunately, most polyps are pedunculated. As Morson, former pathologist at St Mark's has said, ‘God places the pedicle beneath the polyp not to prevent metastasis but in order to allow the endoscopist to more easily remove it.' Thus even what may turn out to be a malignant polyp may be most adequately sampled by endoscopic polypectomy. The basic technique of removing a pedunculated polyp involves recognition of the stalk, preferably having the polyp in the lower end of the endoscopic field since this is the point of exit of the snare (Fig. 10, Fig. 11 and Fig. 12). A preformed, retractable wire loop in a plastic sheath is passed through the instrument channel of the endoscope, opened, and passed over the head of the polyp, enveloping it at the approximate mid-portion of the stalk. The snare is then progressively tightened until it is snug, but avoiding premature transection of the stalk by avulsion without hemostasis. As electrical current (usually monopolar) is passed through the wire; separation and hemostasis are achieved as the wire is progressively tightened. While some large, sessile polyps may be removed with a snare, often piecemeal, this requires experience and judgment. Transmural burns more easily occur under such circumstances and these may result in frank perforation of the bowel. Remember that large, sessile polyps in the distal rectum may be removed transanally under direct vision; in general, this is superior to attempts at piecemeal removal of such polyps with the endoscope. If a sessile polyp is removed endoscopically and proves to contain invasive cancer, it is generally agreed that the risk of residual cancer at the site or in the regional lymphatics is higher then in a pedunculated adenoma and resection is more commonly advised. However, the management of cancer in a pedunculated polyp that has been removed endoscopically is still a controversial matter. Certainly, the incidence of lymph-node involvement in this small group of patients is low, but the criteria commonly used to decide on which patients should undergo resection are not universally accepted or supported by firm evidence. Most agree, however, that it is a question of balancing the risks of operation against the risks of adverse outcome (lymph-node or more distant metastasis or local recurrence). Factors arguing for surgical resection include cancer at the site of transection of the polyp, if most of the polyp tissue is malignant, if the site is well above the peritoneal reflection, and if the patient is at low risk for colonic resection. Some include cancer that is poorly differentiated and cancer demonstrably in lymphatics.

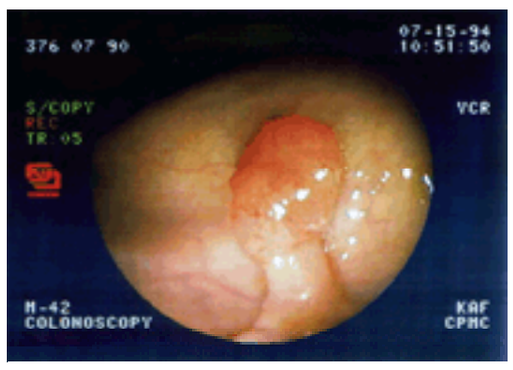


Fig. 10. A pedunculated colon polyp.

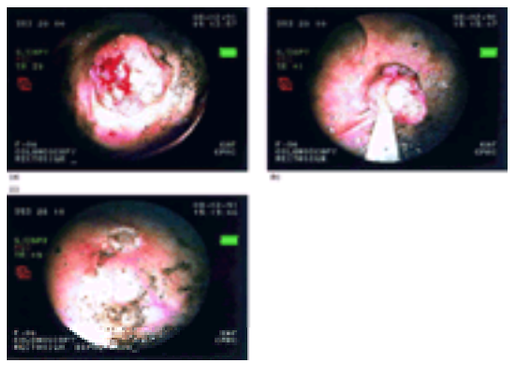


Fig. 11. (a) Polyp on a short stalk; (b) the polyp neck encircled by electroautery snare; (c) polypectomy site after transection.

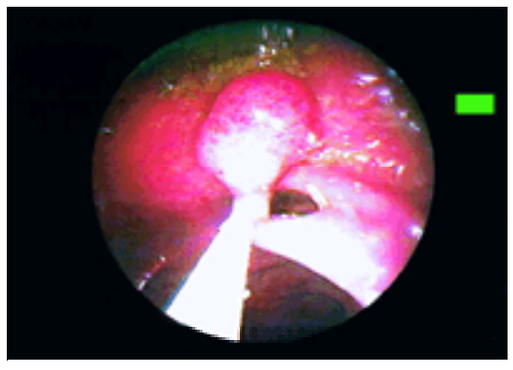


Fig. 12. A snare placed around the polyp.

Endoscopic management of colonic bleeding

In addition to the bleeding colonic polyp, for which polypectomy is definitive management, or the bleeding polypectomy stalk, which may be controlled by regrasping, the colonoscope may be used as a vehicle for other ways of arresting hemorrhage. For example, areas of angiodysplasia (Fig. 13) may be managed by a variety of endoscopic techniques: laser, usually potassium titanyl phosphate (**KTP**), bipolar cautery, or the heater probe. While rebleeding is reported to occur in 10 to 30 per cent of cases, this compares favorably with other treatment modalities. It is, however, not without risk of perforation, occuring in 0 to 7 per cent of series. A bleeding tumor may be palliated with endoscopic laser photocoagulation, usually with the neodymium:yttrium aluminum garnet laser. Hemorrhage from radiation proctitis (Fig. 14, Fig. 15), a more common problem with patients receiving radiation for prostatic or cervical cancer, may be managed endoscopically with laser photocoagulation (usually the KTP). Even diverticular bleeding has at times been managed by electrocoagulation and the production of edema in the vicinity of the bleeding diverticular vessel.

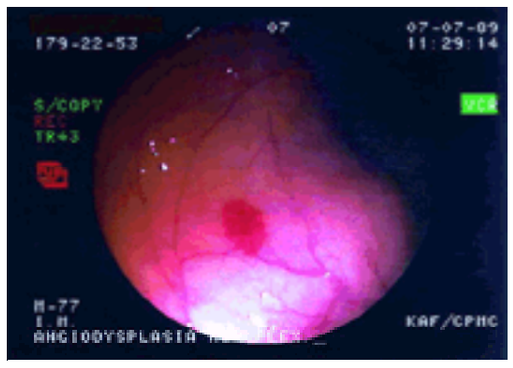


Fig. 13. Angiodysplasia of right colon.

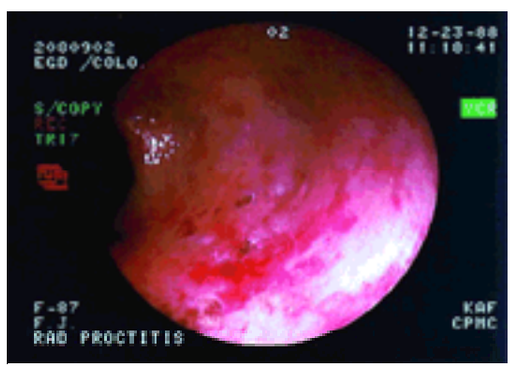


Fig. 14. Bleeding radiation proctitis.

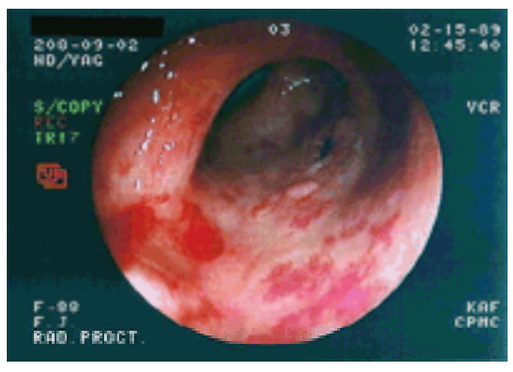


Fig. 15. Radiation proctitis under treatment with laser.

Decompression of non-obstructive ileus and volvulus

When the diagnosis of non-obstructed colonic ileus is made, concern grows about overdistension of the cecum and its possible rupture. It is therefore appropriate to attempt colonoscopic decompression. However, this may be disappointing in that the distension tends to recur. For this reason, it is wise to pass a wide-caliber polyethylene tube through the biopsy channel of the instrument and leave it in place in the proximal colon. While it tends to slip back, if it remains at least in the transverse colon, bowel decompression may last longer. An important caveat in considering colonoscopic decompression is to avoid instillation of barium prior to the examination, since it not only makes visualization difficult but blocks the instrument channels as time passes and the barium congeals.

In patients with volvulus, most commonly sigmoid, emergency rigid sigmoidoscopy has long been in the surgeon's armamentarium for decompression. Colonoscopic decompression is now a useful maneuver. A rectal tube may be inserted alongside the endoscope.

Conclusion

Colonoscopy has thus revolutionized the diagnosis and management of sundry gastrointestinal conditions. Since it has increasing attractiveness as a treatment modality it has become important to the surgeon. Furthermore, even in the era of minimal-access surgery, the endoscope is becoming increasingly important to the surgeon not only as a therapeutic but also as a diagnostic tool, a use that many had thought might be relegated only to the nonsurgeon endoscopist. Further advantages of this technology include the ability to obtain more sophisticated help during procedures, as the assistant can visualize on a television screen the same images seen by the operator. It also provides a superb educational and training tool.

Further reading

Arregui ME, Sackier JM, ed. *Minimal access coloproctology*. Radcliffe Medical, Oxford and New York, 1995. [This overview of the history and development of modern methods of diagnosis and therapy of colonic diseases includes chapters on colonoscopy for the management of benign colorectal conditions and for the management of malignant situations.]

Brunetaud JM, Maunoury V, Ducrotte P, Cochelard D, Cortot A, Paris JC. Palliative treatment of rectosigmoid carcinoma by laser endoscopic photoablation. *Gastroenterology* 1987; **92**: 663–8. [Use of laser energy delivered through the flexible endoscope to ablate tumors.]

Cranley JP, Petras RE, Carey WD, Paradis K, Sivak MV. When is endoscopic polypectomy adequate therapy for colonic polyps containing invasive carcinoma? *Gastroenterology* 1986; **91**: 419–27. [A pathological analysis of the controversial issue concerning the adequacy of endoscopic management alone in the treatment of polyps containg invasive cancer.]

Dent TL, Kukora JS, Buinewicz BR. Endoscopic screening and surveillance for gastrointestinal malignancy. *Surgical Clinics of North America* 1989; **69**: 1205–25. [A review of the role of endoscopic examination and screening for gastrointestinal cancer.]

Forde KA. Colonoscopic management of polypoid lesions. *Surgical Clinics of North America* 1989; **69**: 1287–308. [Endoscopic descriptions with pathologic correlation of polypoid lesions discovered at colonscopy.]

Greene Fl, Ponsky JL, ed. *Endoscopic surgery*. Saunders, Philadelphia, 1994. [The chapter on diagnostic colonoscopy offers a detailed description of the technique of insertion and inspection.]

Guillem JG, Forde KA, Treat MR, Neugut AI, Bodian CA. The impact of colonoscopy on the early detection of colonic neoplasms in patients with rectal bleeding. *Annals of Surgery* 1987; **206**: 606–11. [An aggressive search for the cause of gastrointestinal bleeding, including use of endoscopic techniques results in discovery of significant benign and malignant colorectal disease.]

Mathus-Vliegen EMH, Tytgat GNJ. Laser photocoagulation in the palliation of colorectal malignancies. *Cancer* 1986; **57**: 2212–16. [Series on experience of endoscopically delivered laser therapy and the palliation of advanced colorectal malignancy.]

Nivatongs S, Vermeulen FD, Fang DT. Colonoscopic decompression of acute pseudo-obstruction of the colon. *Annals of Surgery* 1982; **196**: 598–600. [Description and results of the use of the colonoscope and the management of Ogilvie's syndrome.]

Raskin JB, Nord HJ, ed. *Colonoscopy—principles and techniques*. Igaku-Shoin, New York and Tokyo, 1995. [A comprehensive multi-author review of diagnostic and therapeutic colonoscopy with an emphasis on technique.]

Schrock TR. Colonoscopic decompression and treatment of lower gastrointestinal bleeding. *Surgical Clinics of North America* 1989; **69**: 1309–25. [A review of the integration of various diagnostic methods in the evaluation of the patient with various presentations of gastrointestinal bleeding.]

Strodel WE, Brothers T. Colonoscopic decompression of pseudo-obstruction and volvulus. *Surgical Clinics of North America* 1989; **69**: 1327–35. [A review of the experience of colonoscopic decompression of these two conditions.]

Wilcox GM, Anderson PB, Colacchio TA. Early invasive carcinoma in colonic polyps: a review of the literature with emphasis on the assessment of the risk of metastasis. *Cancer* 1986; **57**: 160–71. [A meta-analysis of the risk of metastasis from endoscopically removed polyps containing invasive carcinoma.]

14.4 Diagnostic laparoscopy

Barry A. Salky

Anesthesia
Operative technique
Abdominal pain
Focal liver disease
Masses
Exudative ascites
Second look (prelaparotomy)
Contraindications
Complications
Further reading

The advent of therapeutic laparoscopy has been accompanied by a ‘new' look at diagnostic laparoscopy. The rapid development of new instrumentation and high-resolution television cameras has allowed laparoscopy to become entrenched in surgery. Surgeons once unfamiliar with laparoscopy have rushed into the laparoscopic arena, sometimes without proper training. It is incumbent on the general surgeon to acquire the skills necessary to perform this procedure safely. Diagnostic laparoscopy, once in the realm of the gastroenterologist, now resides firmly in the hands of the surgeon. While the additions of non-invasive diagnostic tests (CT scan, sonography, etc.) have modified the diagnostician's approach to many diseases, laparoscopy continues to play a major role in the evaluation of acute and chronic abdominal pain, focal liver disease, abdominal and retroperitoneal masses, exudative ascites, and prelaparotomy or second-look procedures for evaluation of malignant disease. The diagnostic accuracy of laparoscopy should exceed 95 per cent for all conditions except chronic abdominal pain and second-look procedures. This will decrease the need for exploratory laparotomy with its inherent accompanied morbidity. Table 1 lists the author's experience with diagnostic laparoscopy.

Indication	*Number*	*Accuracy (%)*
Abdominal pain	376	Acute: 99 Chronic: 70
Liver	289	95
Mass	104	96
Ascites	89	97
Second look	57	85

Table 1 The author's experience with the use of diagnostic laparoscopy

Anesthesia

In most cases diagnostic laparoscopy can be accomplished under local anesthesia with intravenous sedation. The author's preference is 0.5 per cent bupivacaine without epinephrine. Appropriate monitoring including continuous blood pressure and pulse oximeter is standard with all laparoscopic procedures. The anesthetic gas of choice when employing local anesthetic is nitrous oxide. Carbon dioxide combines with the fluid in the abdomen to form carbonic acid, which is irritating to the parietal peritoneum. Nitrous oxide is inert and, therefore, it does not cause pain. It does not cause respiratory depression. The absorption from the peritoneal cavity is minimal and can be measured at 0.04 per cent in expired air. Diagnostic laparoscopy under local anesthesia requires some assistance from the patient; therefore, it is not appropriate for the very young or uncooperative patient. It is important not to overdistend the peritoneum with the pneumoperitoneum while under local anesthesia. Profound bradycardia can occur, as the vagal reflexes are still intact in the awake patient. If bradycardia occurs, prompt treatment with atropine is appropriate. Most diagnostic procedures can be performed in the outpatient setting.

Operative technique

The patient is placed supine on the operating room table, and venous compression stockings are placed. An indwelling Foley catheter is not utilized unless the pathology is thought to be located in the pelvis. If performed under local anesthesia, the skin, fascia, muscle, and peritoneum are anesthetized. I prefer the entry to be 2.5 cm above and to the left of the umbilicus (transrectus). It is best to avoid the infraumbilical midline area, especially if the indication is liver pathology. Bleeding from collateral vessels in this area can be difficult to control. Once appropriate anesthesia is obtained, the patient pushes out their abdominal muscles while keeping the spine flat on the operating table. The abdomen is entered with a Veress needle, unless previous surgery dictates an open Hassan entry. A second puncture is always placed in order to facilitate manipulation of the abdominal contents. Extra trocars are placed as needed. If the procedure is performed under local anesthesia, each trocar site must be injected. This is straightforward while the laparoscope is visualizing the parietal peritoneum. The object is to raise a weal of local anesthesia in the parietal peritoneum. The parietal peritoneum in other places should not be touched, as it will be painful to the patient. However, the abdominal contents can be gently manipulated without causing pain. Tilting the operating table in various positions will allow gravity to aid in the displacement of large and small bowel. A thorough evaluation of the abdominal contents is possible under local anethesia.

Abdominal pain

There are multiple reports documenting the accuracy of laparoscopy in both acute and chronic pain syndromes. This becomes even more important in the era of therapeutic laparoscopic surgery. The surgeon not only will be able to make the diagnosis, but also correct the problem in the vast majority of cases. In the author's experience of 376 patients, diagnostic laparoscopy was accurate 99 per cent of the time in acute problems and 70 per cent in chronic pain syndromes. Given a certain level of expertise in advanced laparoscopic surgery, therapeutic laparoscopy will be able to solve the problem in more than 90 per cent of cases. In acute pain syndromes, general anesthesia is utilized, whereas in chronic pain syndromes, local anesthesia is preferred. The patients with chronic abdominal pain often had multiple work-ups including contrast radiography and endoscopies. Previous surgery was present in more than 60 per cent of the patients. Adhesions were not a contraindication to the performance of diagnostic laparoscopy, but lysis of adhesions was frequently necessary. Laparoscopy should be the diagnostic (invasive) procedure of choice in the evaluation of acute abdominal pain requiring surgery.

Focal liver disease

Eighty-five per cent of the liver surface is visible with the laparoscope. It has use in both benign and malignant pathology. In the author's experience of 289 procedures for focal liver disease, a definitive diagnosis was possible in more than 95 per cent of cases. With the addition of laparoscopic ultrasound, the diagnostic rate will increase to 98 per cent. While blind, percutaneous biopsy is appropriate for diffuse disease, laparoscopic-directed biopsy has been shown to have a higher sensitivity and specificity than either CT-guided or blind biopsy in focal disease of the liver. The sensitivity of thin-cut CT scanning is 8 mm, whereas much smaller lesions are visible laparoscopically. The other advantage of laparoscopy is the identification of unsuspected pathology at remote locations, which may impact the decision-making process for therapy in malignant disease. The diagnosis of hemangioma and liver cyst is made visually. Biopsy of these two lesions is contraindicated. Preoperative identification of lesions in the bare area or deep in the parenchyma of the liver are not appropriate candidates for diagnostic laparoscopy. These areas are not visible with laparoscopic techniques.

Masses

Directed biopsy of an intra-abdominal mass is straightforward using laparoscopy. While CT scan and ultrasonographic biopsy have a role in the delineation of abdominal masses, laparoscopy has the additional advantage of controlling hemorrhage, detecting small lesions under 5 mm in diameter, and determining surgical relationships which can impact operative and therapeutic decisions. Retroperitoneal masses can also be approached laparoscopically. In lymphomatous disease, core tissue obtained laparoscopically will negate the need for open laparotomy. If the retroperitoneal mass can be palpated or intra-abdominal viscera are displaced by the mass, in the author's experience, directed biopsy is always possible. With expertise in advanced laparoscopic surgery, the whole retroperitoneum is accessible.

Experience with over 100 patients has yielded the correct diagnosis 99 per cent of the time.

Exudative ascites

Laparoscopy in the presence of ascites requires changes in technique. The initial trocar puncture must be above the umbilicus, or it may be difficult to reach the liver. If tense ascites exists, the fluid must be drained in order to obtain pneumoperitoneum. The instillation of the pneumoperitoneum cannot be into the fluid, or ‘soap bubbles' will form which will preclude visualization of the abdominal cavity. If tuberculosis is suspected, an open entry into the abdominal cavity is recommended. Transudative ascites is almost always caused by a ‘medical' condition, which does not require laparoscopy. Local anesthesia is preferred by the author. In the author's experience, laparoscopy has been diagnostic in 97 per cent of the cases.

Second look (prelaparotomy)

Laparoscopy has a role in defining treatment response in malignant abdominal disease, and it has a role in the selection of some patients regarding operability for cure. Treatment response can be defined accurately with the addition of laparoscopy (as opposed to laparotomy). Biopsy and collection of tumor for chemosensitivity testing will be of increasing importance. The role of diagnostic laparoscopy in patients with carcinoma of the pancreas has been defined by Warshaw. Used properly, approximately 25 per cent of patients will be spared a laparotomy for this disease.

Contraindications

An active infection on the anterior abdominal wall near the planned entry or accessory trocar site is an absolute contraindication to performing diagnostic laparoscopy. Lack of co-operation is a contraindication when utilizing local anesthesia. Uncorrectable coagulopathy is a contraindication to biopsy, but minor degrees of coagulopathy are not a contraindication to diagnostic laparoscopy. Obesity and previous surgery are not contraindications to laparoscopy.

Complications

The risks of injury to abdominal viscera or major vascular structures are low. This is in keeping with the author's personal experience with diagnostic laparoscopy. Avoidance of blind entry into the previously operated abdomen and proper insertion techniques with the Veress needle will go a long way in preventing injury. Bleeding is always a potential problem with any procedure, but laparoscopic-directed biopsy is the safest way to obtain hemostasis. There are a variety of hemostasis techniques available to the laparoscopic surgeon. The fascia of all ports below the level of the umbilicus and 10 mm or larger should be closed to reduce the incidence of postoperative hernia.

Further reading

Boyce HW Jr, Palmer E. *Techniques of clinical gastroenterology*. Charles C. Thomas, Springfield, Illinois, 1975.

Bruehl W. Incidence of complications during laparoscopy and liver puncture under vision: results of a survey. *Deutsche Medische Wochenschrift* 1966; **91**: 2297–9.

Fernandez-del Castillo C, Rattner DN, Warshaw AL. Further experience with laparoscopic and peritoneal cytology in the staging of pancreatic cancer. *British Journal of Surgery* 1995; **82**: 1127–9.

Salky BA. *Laparoscopy for surgeons*. Igaku-Shoin, New York, 1990.

Salky BA, Edye MB. Diagnostic laparoscopy. *Surgical Laparoscopy and Endoscopy* 1993; **2**: 132–4.

Salky BA, Bauer JJ, Gelernt IM, Kreel I. The use of laparoscopy in retroperitoneal pathology. *Gastrointestinal Endoscopy* 1988; **34**: 227–30.

Vilardell F, Seres I, Marti-Vincente A. Complications of peritoneoscopy: a survery of 1455 examinations. *Gastrointestinal Endoscopy* 1968; **14**: 178–80.

14.5 Percutaneous endoscopic gastrostomy and jejunostomy

Timothy Simon and Aaron S. Fink

Technique
Complications
Further reading

Percutaneous endoscopic gastrostomy (**PEG**) is the preferred method for long-term feeding or nutritional supplementation in patients who are unable to swallow. In addition, it is the optimal method for providing chronic gastric decompression. PEG is relatively contraindicated in patients with ascites because of difficulty forming an adequate tract around the tube. Patients with systemic sepsis or intra-abdominal infection should have placement delayed until these problems have resolved.

Technique

Similar patient preparation, monitoring, and sedation principles are followed as for routine esophagogastroduodenostomies (see Chapter 14.1). Antibiotic prophylaxis against Gram-positive cocci decreases the incidence of PEG-site infections. Thus, all patients should receive appropriate intravenous antibiotics before the procedure.

The patient is placed in the supine position, and the left upper quadrant is made sterile and draped. The endoscope is directed into the stomach, which is then distended with air. Complete inspection of the stomach and duodenum is performed to rule out gastric pathology or pyloric obstruction. Using sterile technique, the assistant pushes on the stomach with one finger at the point where maximal transillumination is observed. Indentation of the gastric wall should be observed. Ideally, this point will be two-thirds of the distance from the umbilicus to the midpoint of the left costal margin. A polypectomy snare is passed through the endoscope channel; the wire loop is opened and placed over the bulge created by the assistant's finger. The assistant infiltrates the site with local anesthetic. The fine needle used to anesthetize the skin should be inserted into the abdomen, verifying entry into the stomach. After making a small (1 cm) skin incision, the assistant inserts an intravenous cannula through the incision, which must be visualized entering the stomach. The open snare is tightened around the cannula and the inner stylet is removed. In the ‘pull technique,' a heavy silk suture is placed through the cannula and firmly grasped with the polypectomy snare after releasing the cannula. The cannula is withdrawn, and the endoscope with the tightened snare is removed, bringing the suture out of the patient's mouth. The suture is tied to the well lubricated, tapered dilator affixed to the external end of a gastrostomy tube. The assistant then pulls on the suture until the attached tube exits the abdominal wall. The endoscope is reinserted to view the inner bolster of the tube as the stomach is loosely seated against the abdominal wall. The tube is then secured in place on the abdominal wall with an external bolster. Adequate position of the internal bolster is verified before removing the endoscope.

Two other methods are often used. In the ‘push technique', a guidewire is used instead of a suture. The endoscopist then pushes the gastrostomy tube over the guidewire through the esophagus and into the patient's stomach until it exits the abdominal wall. When the tube emerges from the abdominal wall, it is grasped by the assistant and pulled upward. Tension is maintained at both ends of the guidewire during placement. This technique offers several technical advantages in comparison with the pull method.

In the ‘introducer technique', the stomach is inflated, the insertion site is selected, and the intravenous cannula is introduced into the stomach as described above. A J-tipped guidewire is passed through the cannula into the stomach, and the cannula is then removed. Next, an introducer with a peel-away sheath is passed over this wire, allowing the removal of the wire and introducer. A Foley catheter or other similar gastrostomy tube then is placed through the sheath and its balloon is inflated. The sheath is then removed, after which the catheter is secured to the abdominal wall.

Complications

Although usually minor, complications following PEG placement are not infrequent. Purulent fluid is often observed around the tube insertion site. This is often caused by tension between the inner and outer bolsters, causing cutaneous ischemia and/or necrosis at the PEG site. Care in avoiding excessive tension and proper ostomy skin care normally alleviate this problem. Any pocket of purulence requires drainage. If the infection appears to spread or the site looks particularly dusky, necrotizing fasciitis should be considered. If the latter is diagnosed, the tube must be removed and dead tissue excised. A nasogastric tube should be placed in the stomach for decompression to allow healing of the gastrostomy site. Appropriate antibiotics are given according to the Gram-stain and culture results from the debrided necrotic tissue.

Tube migration causing pyloric obstruction has been reported. This complication should be considered in patients who experience unexplained difficulty with feedings after initially doing well. In this case, the tube is pulled outward until mild resistance is met, and then the outer bolster is repositioned closer to the skin.

Rarely, PEG is complicated by separation from the abdominal wall with resultant peritonitis. Also, gastrocolic fistula may develop due to inadvertent colonic puncture during needle catheter insertion or colonic entrapment between the stomach and the abdominal wall. Diarrhea shortly after feeding may suggest this complication. Diagnosis is confirmed with contrast injection through the tube. Removal of the tube usually results in closure of the tract and fistula, with minimal morbidity.

Theoretically, patients with increased risk of aspiration can be provided with nutrition more safely by extending the PEG procedure to include insertion of a percutaneous endoscopic jejunostomy (**PEJ**). This can be accomplished by passing the jejunal tube through the existing PEG and then endoscopically guiding it into the jejunum with foreign-body forceps. Alternately, a guidewire can be placed through the PEG and endoscopically guided into the jejunum using the foreign-body forceps. The jejunal tube is then inserted into the jejunum over the guidewire. Both techniques are often complicated by unintentional removal of the jejunal tube upon endoscope withdrawal.

A technique of jejunal tube insertion using a nasobiliary catheter (8.5 Fr.) has been described which offers a high success rate for proper jejunal placement. With this technique, however, the narrow luminal diameter of the nasobiliary catheter as well as its tendency to break makes maintaining luminal patency an issue.

With chronic PEG tubes, jejunal tubes can be inserted directly through the stoma site after the latter is dilated to allow passage of the endoscope. The scope is then introduced directly into the stomach and inserted deep into the proximal small bowel. A guidewire is then inserted through the scope, which is left in place as the scope is withdrawn. An assembly consisting of a 12 Fr. feeding jejunal tube preloaded through a 28 Fr. replacement PEG is placed over the guidewire. Peroral endoscopy may be used to verify adequate placement of the jejunal tube (Fig. 1). This should allow increased patency rates compared with the smaller 8.5 Fr. nasobiliary tube. Unfortunately, the approach can only be used in the chronic setting, after the PEG tract has matured (2 to 3 weeks).

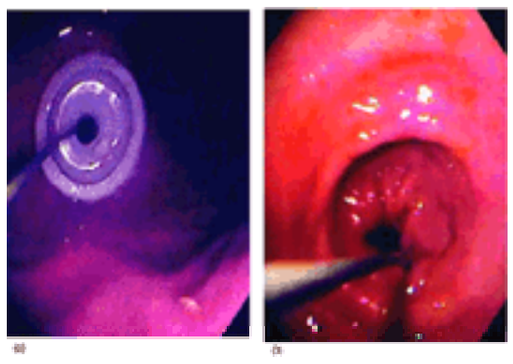


Fig. 1. (a) A 12 Fr. jejunostomy tube seen entering the stomach through the internal bolster of existing PEG tube. (b) PEJ tube passing into the duodenum through the pylorus

Further reading

Anon. Antimicrobial prophylaxis in surgery. *The Medical Letter* 1992; **34**: 5–8. [A comprehensive review outlining guidelines for use of antibiotics during minimally invasive procedures.]

Hogan RB, Demarco DC, Hamilton JK, Walker CO, Polter DE. Percutaneous endoscopic gastrostomy—to push or pull. A prospective randomized trial. *Gastrointestinal Endoscopy* 1986; **32**: 253–8. [A prospective randomized trial which compares the efficacy of the two predominant techniques used when placing PEG feeding tubes.]

MacFadyen BV, Catalano MF, Raijman I, Ghobrial R. Percutaneous endoscopic gastrostomy with jejunal extension: a new technique. *American Journal of Gastroenterology* 1992; **87**: 725–8. [Describes a novel technique for placement of feeding jejunal tubes through existing PEG tubes.]

Ponsky JL, Gauderer MWL. Percutaneous endoscopic gastrostomy: indications, limitations, and results. *World Journal of Surgery* 1989; **13**: 165–70. [The group who first described the PEG present a retrospective review of their experience.]

14.6 Bronchoscopy

Malcolm K. Benson

Bronchoscopy
Fibreoptic bronchoscopy
Rigid bronchoscopy
Therapeutic applications
Endoscopic surgery
Laser photoresection
Cryotherapy
Brachytherapy
Photodynamic therapy
Tracheobronchial stents
Further reading

Bronchoscopy

The debate over the relative merits of rigid versus fibreoptic bronchoscopy has largely been resolved with the recognition that they are complementary, both having advantages and disadvantages. Any experienced bronchoscopist should be familiar with both types of instrument. Rigid bronchoscopy was and still is mainly the province of the thoracic surgeon. Although it can be performed under local anaesthesia, general anaesthesia is felt to be kinder and more appropriate. It retains several advantages over fibreoptic bronchoscopy: the large calibre allows easier suction of retained secretions and better control of bleeding, foreign bodies can be removed more safely, and good control of ventilation and airway patency can be maintained. It can also be used to feel the mobility of the trachea and main bronchi, allowing assessment of operability of tumours.

The flexible fibreoptic bronchoscope, introduced in the late 1960s, can safely be used under local anaesthesia. It can be introduced through nasal or oral passages and can also be passed through an endotracheal tube or via a rigid bronchoscope. Its main advantage, other than the ability to dispense with a general anaesthetic, is the ability to visualize and sample from peripheral bronchi and lung parenchyma (Fig. 1).

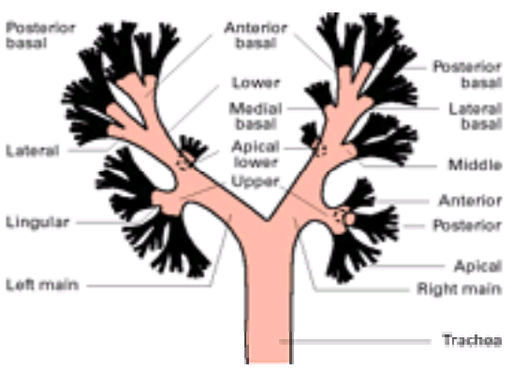


Fig. 1. Diagram of the main branches of the bronchial tree; the increased peripheral vision obtained by use of the fibreoptic bronchoscope is depicted in black.

Fibreoptic bronchoscopy

Indications

The most common indication for bronchoscopy is in the diagnostic assessment of patients with a suspected bronchial carcinoma or other localized obstruction within the bronchial tree. Clinical suspicion may be based on symptoms such as cough and haemoptysis, abnormal radiographic shadowing, or failed resolution of a pneumonia. It is easy to visualize and obtain diagnostic material from lesions in the central airways, including the segmental bronchi. Samples can also be obtained from peripheral lesions by undertaking transbronchial biopsies under fluoroscopic screening or by using bronchoalveolar lavage. Bronchoscopic findings help in making decisions about operability, but even if tumours are felt to be inoperable, a tissue diagnosis may enable more rational use of chemotherapy or radiotherapy.

Fibreoptic bronchoscopy allows access to subsegmental bronchi and is widely used in the investigation of pulmonary infections and interstitial lung disease. Bronchoalveolar lavage and transbronchial biopsy have both proved a useful adjunct in the investigation of patients with immune deficiencies. Both techniques can be used to obtain samples for isolating *Pneumocystis carinii*, cytomegalovirus, and both typical and atypical mycobacteria. In patients with diffuse interstitial lung disease, transbronchial biopsies offer a higher diagnostic yield in conditions that have specific histological features, including sarcoidosis, malignant infiltrates, lymphoma, alveolar proteinosis, and histiocytosis X. Biopsies taken from patients with diffuse fibrotic lung disease are of limited value. Differential cell counts on samples obtained from bronchoalveolar lavage in patients with conditions such as sarcoidosis and fibrosing alveolitis have not been sufficiently specific to be of use diagnostically.

Technique

Whilst there is general agreement about topical anaesthesia for fibreoptic bronchoscopy, the relative risks and benefits from sedation remain the subject of debate. Whatever approach is undertaken, it is essential that the patient be given a careful explanation of the procedure to avoid unnecessary fear. Lignocaine (lidocaine) is the most commonly used topical anaesthetic: there are a number of minor variations in technique used to achieve satisfactory anaesthesia for the nasal and oropharyngeal mucosa, larynx, and tracheobronchial tree. The nasal mucosa is anaesthetized with a 4 per cent spray or lignocaine (lidocaine) gel, the spray also being used for the oropharynx. Further lignocaine (lidocaine) (2 per cent) can be injected through the bronchoscope operating channel directly on to the larynx and tracheal mucosa. An alternative approach is to inject the anaesthetic through the cricohyoid membrane: the initial paroxysm of coughing serves to distribute it on to the larynx and to the tracheobronchial tree.

Sedation is often given to allay patients' anxieties and increase the acceptability of the procedure, although the benefits of routine sedation are not strongly supported by available evidence. In general, sedation should be reserved for those patients most likely to benefit and avoided in those at potential risk from respiratory depression. Risks can also be reduced by ensuring that staff are familiar with the type of sedation used, that there is careful physiological monitoring with oxygen supplementation where necessary, and that full resuscitation equipment is available.

Midazolam is a water-soluble benzodiazepine with a short half-life and rapid onset of action. The dose required is usually estimated on the basis of a patient's weight, although there is considerable variation in response. An initial dose of 2 mg intravenously can be followed by increments of 1 mg/min until adequate sedation is achieved. Flumazenil is a specific benzodiazepine antagonist that can reverse the effects of oversedation. Propafol is an alternative sedative that is used for induction and maintenance of anaesthesia. It may offer some advantages over other sedatives but its use requires familiarity with its administration.

Anticholinergics are now no longer regarded as necessary premedication for fibreoptic bronchoscopy. Their use was largely based on anaesthetic practice for rigid bronchoscopy in order to dry secretions and reduce vasovagal activity, but any apparent benefits are marginal.

Fibreoptic bronchoscopy is generally performed with a patient semirecumbent and the operator standing to one side and facing the patient. An alternative is to stand at the head of a supine patient, an approach that maintains the same spatial orientation of the bronchial tree as in rigid bronchoscopy. Insertion of the bronchoscope via the nose is preferred. In negotiating the nasal passages, the bronchoscope should be passed directly backwards and worked through by gentle manipulation, not force. In 5 to 10 per cent of patients the size of the nasal airway is inadequate and the oral route must be used. The bronchoscope should be protected by passing it through a mouthguard gently clasped between the patient's teeth.

As the bronchoscope is advanced the nares, naso- and oropharynx, and vocal cords should be inspected and the mobility of the cords assessed during phonation. When entering the trachea, coughing can be minimized by adequate local analgesia and avoiding contact with the mucosa as much as possible. Secretions can be aspirated, but if vision is impaired by blood or mucus on the distal lens it can usually be cleared by gently wiping the tip of the instrument on the bronchial mucosa.

Full inspection of the bronchial tree requires a methodical approach and detailed knowledge of bronchial anatomy. Individual segmental bronchi cannot be recognized without appreciating the route taken to that particular airway. Practice in manipulating the instrument on a lung model is strongly recommended: descriptions of normal and abnormal appearances are inadequate substitutes for practical experience. Pathological findings can be recognized by the presence of excess secretions, mucosal abnormalities or distortion of the normal anatomy, usually due to extrinsic pressure. Secretions may vary from excess amounts of normal clear mucus found in patients with chronic bronchitis to thick pus in patients with bronchiectasis or lung abscess (Fig. 2a). Thick, viscid plugs of mucus can be found in patients with asthma. Inflammation may be accompanied by erythema of the bronchial mucosa, which is friable and bleeds easily.

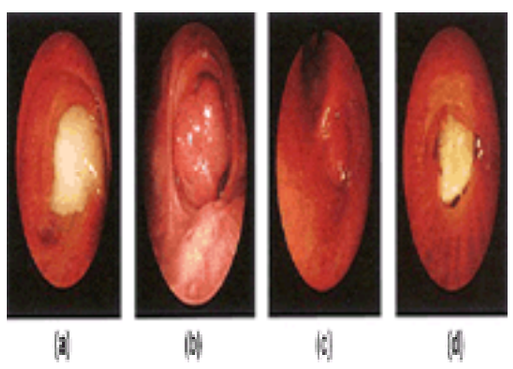


Fig. 2. (a) Postoperative mucous plugging of right main bronchus; mucopurulent secretions were removed at bronchoscopy allowing re-expansion of the lung. (b) Carcinoma of right main bronchus: a large, lobulated, fleshy mass is seen partially occluding the bronchial lumen; there is superficial spontaneous bleeding. (c) ‘Cherry red' bronchial adenoma (carcinoid) protruding from apical bronchus right lower lobe. (d) Foreign body (impacted bone fragment) obstructing right main bronchus. (By courtesy of Dr Peter Stradling.)

The appearance of bronchial neoplasms is variable and they may be difficult to differentiate from an area of local inflammation (Fig. 2(b),Fig. 2(c)). Polypoid tumours are clearly recognizable, and intraluminal obstruction with an ulcerated mass is usually diagnostic. However, small areas of mucosal irregularity and swelling can be due to submucosal tumour. External pressure on the bronchial wall is commonly due to secondary lymph-node enlargement and may result in widening of a carina. The majority of visible tumours prove to be bronchial carcinomas. The rare bronchial adenoma (carcinoid tumour) has a characteristic cherry-red appearance due to its vascularity. Other rare tumours include cylindromas, lipomas, papillomas, and chondromas. The typical appearance of a Kaposi sarcoma, which develops in patients with the acquired immune deficiency syndrome, is of a raised violaceous papule. Endobronchial abnormalities may also be seen in patients with sarcoidosis and amyloidosis.

Sampling techniques

The diagnostic yield from fibreoptic bronchoscopy depends on close co-operation between the bronchoscopist and laboratory-based colleagues. Cytological samples both from central and peripheral lesions can be obtained with a sheathed cytology brush. For peripheral lesions not visible bronchoscopically, diagnostic yield can be increased by sampling under fluoroscopic control. Segmental lavage can also be used to obtain material for cytological examination.

Bronchial biopsies can be taken with a variety of biopsy forceps. The samples obtained are relatively small but when tumour is visible a histological diagnosis can be obtained in more than 90 per cent of patients. In patients with peripheral lung lesions and diffuse parenchymal disease, biopsies can be taken by passing forceps beyond the visible bronchi. Again the procedure is best undertaken with the aid of fluoroscopy. The bronchoscope is wedged into a segmental bronchus and not removed until the biopsy has been taken and haemostasis achieved. Closed biopsy forceps are passed into the lung peripherally to within 1 or 2 cm of the pleural surface. During inspiration the forceps are slightly withdrawn, opened, and then gently advanced during expiration. The forceps are then closed and withdrawn, with the object of obtaining a small amount of the lung parenchyma between two distal airways. Several biopsy samples can be taken in this way if needed.

In order to perform bronchoalveolar lavage, the bronchoscope is again wedged in a segmental bronchus. Sterile, warmed normal saline is injected in approx. 50-ml portions and the fluid is then withdrawn by gentle suction. Up to 50 per cent of the fluid can usually be retrieved. If excess suction is used, trauma to the mucosa will adversely affect the viability of the cells. Despite the fact that culture of the aspirate may be misleading because of contamination with nasopharyngeal commensals, lavage remains the most effective way of isolating a number of organisms including *Pneumocystis carinii*, mycobacteria, cytomegalovirus, and various fungi.

Complications

There are no absolute contraindications to fibreoptic bronchoscopy. There are, however, increased risks in patients with ischaemic heart disease, respiratory failure, asthma, and bleeding diatheses. Bleeding, especially after biopsy, is the most frequent complication. Adenomas and other vascular tumours may bleed profusely and care should be taken in patients in whom cytological brushing induces brisk bleeding. Haemostasis may be achieved by wedging the bronchoscope *in situ*, and inflatable balloon catheters are also available.

Resting arterial oxygen falls by an average of 2.5 kPa during fibreoptic bronchoscopy. Adverse consequences are more likely to occur in patients with pre-existing respiratory failure, and hypoxia may be accompanied by ventricular dysrhythmias. Monitoring of oxygen saturation and supplemental oxygen provide appropriate safeguards. Radiological screening can help to reduce the likelihood of a pneumothorax following transbronchial biopsy. This occurs in approx. 5 per cent of procedures but rarely requires specific treatment.

The risk of cross-infection is small provided that care is taken in cleaning and disinfecting the bronchoscope. Transmission of *Pseudomonas* and mycobacteria between patients has been reported. Cross-infection with human immune deficiency virus is likely to prove less of a problem but adequate safeguards need to be taken by staff performing the bronchoscopy. Gloves, gown, goggles, and mask should be worn by the bronchoscopist to prevent accidental cross-infection.

Rigid bronchoscopy

Although fibreoptic bronchoscopy has an advantage over the rigid bronchoscope as a screening procedure for large numbers of patients, rigid bronchoscopy retains some advantages, which have already been mentioned. It is preferred by most surgeons as a prelude to surgical resection of malignant tumours. The field of vision is more limited than with the fibreoptic ‘scope. Views of upper-lobe bronchi and apical divisions of lower-lobe bronchi require the use of a lateral or oblique viewing telescope.

Technique

Almost all rigid bronchoscopies are done under a general anaesthetic in a fully paralysed patient. Adequate ventilation can be achieved by using the Saunder's oxygen venturi technique. The procedure is performed on a supine patient whose head should be supported on a well-filled pillow and extended so that the chin points vertically. The forefinger and thumb of the left hand protect the teeth and gums from trauma and serve to guide the bronchoscope. The instrument is introduced almost vertically, usually to the right of the mouth. As it follows the contour of the tongue the proximal end is brought downward towards the horizontal. Access to the glottis is obtained by passing posterior to the epiglottis. Once the glottis is seen the bronchoscope is advanced in the midline and the cords approached with the bronchoscope turned through 90°. At this stage the bronchoscope is now directed to the left vocal cord and can be passed gently through the vertical glottic chink. It is essential to avoid the use of force when negotiating the larynx or aligning the bronchoscope within the trachea. The position of the head and bronchoscope needs to be altered in order to intubate the main and lower-lobe bronchi on either side. The left is more difficult and usually requires the head to be turned to the right with the bronchoscope remaining in the right-hand corner of the mouth. The procedure is reversed for examining the right bronchial tree. Correct positioning of lateral or oblique viewing telescopes often proves difficult for the beginner. After methodical inspection of the bronchial tree and the taking of appropriate biopsies, removal of the bronchoscope requires as much care as its insertion.

Therapeutic applications

Although both fibreoptic and rigid bronchoscopy remain essentially diagnostic tools, there are potential therapeutic applications. The removal of foreign bodies is best undertaken with the rigid instrument and there is a variety of forceps and snares available. Care should be taken not to dislodge material into more distal airways. Therapeutic aspiration of inspissated mucus resulting in lobar collapse can also be undertaken with either instrument, although again the rigid bronchoscope has greater advantages. Control of haemorrhage may be achieved by local application of adrenaline (epinephrine) or by the use of balloon catheters to occlude the segment of lung from which the bleeding is occurring.

A variety of endobronchial techniques are used in the management of patients with localized narrowing or obstruction of the trachea or major airways. The endoscopic techniques used depend, at least in part, on whether the stricture is benign or malignant and, if malignant, whether the tumour is intraluminal or causing extrinsic compression. In general, the techniques mentioned are essentially palliative and only appropriate where definitive surgical treatment is technically impossible. The range of techniques available includes closed-loop diathermy, laser photoresection, cryotherapy, brachytherapy, photodynamic therapy, and tracheal stents. Although all have their advocates, most reported studies are from centres that specialize in one or other technique and there is a remarkable paucity of comparative studies.

Endoscopic surgery

Endoscopic resection of the intraluminal component of endotracheal or endobronchial tumours can easily be undertaken by using a rigid bronchoscope to core out tumour. Large fragments can be removed with suitably sized biopsy forceps. There is, however, a significant risk of haemorrhage. Diathermy is probably safer, with less risk of bleeding. Polypoid tumours in which there is a clearly defined pedicle are particularly amenable to this form of treatment. It has been successfully used in the treatment of benign tracheal tumours, although for malignant disease, palliation is the best that can be achieved. It is cheap and readily available and, as such, is a widely used technique.

Laser photoresection

Laser technology has evolved over the past 15 years, and with appropriate patient selection and a skilled operator can achieve valuable palliation. At the centre of the beam, tissue is vaporized at high temperature. The neodymium:yttrium–aluminium–garnet (ND-YAG) laser is widely used and transmitted by flexible optic fibres inserted through either a rigid or fibreoptic instrument. Most operators prefer the rigid instrument. Although mainly used for malignant lesions, it has also been used successfully in the treatment of benign tracheal strictures. It is most likely to be successful for tumours that are within the trachea or major bronchi, have a significant endobronchial component, and are confined to a short segment. It is expensive and requires a somewhat cumbersome delivery system. It is also relatively time-consuming, often requiring repeated application. Potential risks include massive haemorrhage, airway perforation, and systemic air embolism. Endobronchial fire is a rare but life-threatening complication.

Cryotherapy

Cryotherapy is the application of extreme cold to produce local destruction of living tissue. Temperatures of between -20 and -40°C are required to produce cell death. Nitrous oxide is the most commonly used cooling agent. Different sizes and shapes of cryoprobes are available, and they can be used with both the fibreoptic and rigid bronchoscope. The technique is most valuable in palliation of endobronchial tumours but has also been used successfully for benign lesions. Its use in the United Kingdom is limited to relatively few centres.

Brachytherapy

Endobronchial brachytherapy is a technique whereby encapsulated radioactive sources are placed in close proximity to a tumour. It is used as a palliative treatment for malignant tumours causing local narrowing of airways. Since the 1920s, when radium seeds were implanted bronchoscopically into the bronchial lumen, a variety of implanted radionuclides have been used including cobalt-20, radon-222, and iridium-192. Implantation techniques have largely been replaced by afterloading, which involves the temporary placement of a polyethylene catheter adjacent to the tumour, the catheter then being loaded with a radioactive source. Remote afterloading and heavily shielded treatment rooms have improved radiation protection for hospital personnel. A catheter is placed through the tumour at fibreoptic bronchoscopy and the site of treatment identified by passing a graduated guide wire through the catheter. A radioactive source, usually iridium, is then inserted into the catheter under computer control. In general, the treatment is well tolerated and takes about 10 to 15 min. It can be considered as a palliative measure in patients with biopsy-confirmed malignant disease that is inoperable and has not responded to other forms of treatment The most serious complications are massive haemorrhage and fistula formation.

Photodynamic therapy

This involves the use of a photosensitizing agent, which, when exposed to light of appropriate wavelength, forms toxic oxygen radicals that result in cell death. A number of photosensitizers have been tried, of which haematoporphyrin derivatives are most widely used. Photoradiation using an argon pump dilaser is applied bronchoscopically 3 to 5 days after intravenous injection of haematoporphyrin. There are reports of cures in early, superficial carcinomas, although it is largely regarded as a palliative technique for more advanced tumours occluding major airways. Although techniques have been developed over the past 20 years, it has failed to achieve widespread support and its use is limited to relatively few centres.

Tracheobronchial stents

Various stents are used to palliate the symptoms arising from obstruction of a large airway due to malignant disease or benign strictures that are not suitable for surgical reconstruction. The concept is not new and there has been a steady evolution in the design of stents. There are two main types of stent, Silastic tube stents and expanding wire stents. Silastic tubes tend to be more difficult to insert and migration remains a problem. Clearance of secretions is impaired and a side arm is often incorporated through a tracheostomy in order to facilitate removal of secretions. Examples include the Montgomery T-tube and the Westaby Y-stent, which is used for lesions extending into the carina. Self-expanding wire stents were originally designed for biliary and vascular use. They can be inserted through a catheter and positioned using a combination of radiographic screening and bronchoscopic visualization. The most commonly used are the Gianturco and the Wall stent. The Gianturco is a continuous loop of stainless-steel wire with hooks to facilitate anchoring into the airway wall. It is easy to insert but impossible to remove and has been reported as eroding through the bronchial wall with fatal consequences. Use should be limited to palliative treatment of malignant tumours. The Wall stent is an alloy tubular mesh that is available in various sizes. Again, removal is impossible, granulation tissue may form, and tumour may grow through the wire mesh. Erosion of the bronchial wall does not seem to be a major problem.

Further reading

Bert PA, O'Driscoll BR, Notley HM, Barker PV, Stout R. Intra-luminal irradiation for the palliation of lung cancer with high dose rate micro-selectron. *Thorax* 1990; **45**: 765–8. [Good description of technique together with experience in 50 patients with inoperable endobronchial carcinoma.]

Colt HG, Dumon J-F. Airway stents present and future. In: Mathur TN, Beamis JF, ed. *Clinics in chest medicine - interventional pulmonology*. Philadelphia: Saunders, 1995: 465–78.[Review of development of airway stenting techniques together with indications. Limited data on outcomes.]

Hetzel MR, Smith SGT. Endoscopic palliation oftracheobronchial malignancies. *Thorax* 1991; **46**: 325–33. [Useful review of a variety of endoscopic palliation techniques. There is no useful comparative data on outcomes.]

Macha HN, Beca KO, Kemimes HP. Pattern of failure and survival in endobronchial laser resection: a matched pair study. *Chest* 1994; **105:** 1668–72. [This is a prospective audit of outcomes in patients with endobronchial tumours treated with palliative laser resection.]

Maiwand MO, Homasson JP. Cryotherapy for tracheobronchial disorders. In: Mathur TN, Beamis JF, ed. *Clinics in chest medicine - interventional pulmonology*. Philadelphia: Saunders, 1995: 427–14–6. [Overview of new technique of cryotherapy and review of published results together with authors and experience. There are no randomized studies comparing with other palliation techniques.]

Mehta AC, Harris RJ, DeBoer GE. Endoscopic management of benign airway stenosis. In: Mathur TN, Beamis JF, ed. *Clinics in chest medicine - interventional pulmonology.* Philadelphia: Saunders, 1995: 401–14. [A review of endobronchial techniques suitable for use in benign airway stenosis.]

Stradling P. *Diagnostic bronchoscopy*, 6th edn. London: Churchill Livingstone, 1991. [Lavishly illustrated book based on author's personal experience. Outline discussion of techniques with valuable illustrations of normal and pathological bronchoscopic findings.]

14.7 Thoracoscopy and video-assisted thoracic surgery

Anthony P. C. Yim and Mohammad Bashar Izzat

Historical perspective
Basic principles and operative strategies
Indications for VATS
Diagnostic modality
Therapeutic modality
Pleural effusions
Pneumothorax
Mediastinal cysts
Sympathectomy and splanchnicectomy
Pericardial window
Thymectomy
Anatomical lung resections
Other applications of VATS
Future prospects
Further reading

Historical perspective

Thoracoscopy is not new but has been around for almost a century. Hans Christian Jacobaeus, a professor of internal medicine from Stockholm, reported in 1910 using a cystoscope to examine the thoracic cavity under local anaesthesia. He used this technique primarily to lyse adhesions in order to collapse the lungs in the treatment of tuberculosis (Fig. 1). The introduction of this approach was initially received with enthusiasm throughout Europe and America. However, in the 1940s, the advent of streptomycin led to a rapid decline in the use of thoracoscopy for tuberculosis. Thoracoscopy was used only sporadically as a diagnostic tool. In 1976, Lewis and coworkers reported direct diagnostic thoracoscopy on 40 patients using either a rigid bronchoscope or fibreoptic mediastinoscope under general anaesthesia with no mortality and minimal morbidity. The correct diagnosis was achieved in all patients. This stimulated a resurgence of interest in this neglected technique.



Fig. 1. Showing the early use of thoracoscopy by H.C. Jacobaeus with the two cannula technique.

Improved Hopkins lens coupled with the development of solid state systems and microcameras in the 1980s led to the rapid development of videoendoscopic surgery. For the first time, the assistants were able to watch with the operator. The success of laparoscopic cholecystectomy gave impetus to the development of video-assisted thoracic surgery (**VATS**). Selective one-lung ventilation permits easy manoeuvrability of the telescope and instruments. The availability of specially designed endoscopic instruments like the linear staple cutter opened up new vistas for a spectrum of diagnostic and therapeutic thoracoscopic procedures.

Basic principles and operative strategies

Conventional wisdom relates minimal access to limited exposure, but with the advent of videoscopic surgery this is no longer true. The thoracoscope attached to a videocamera unit provides a magnified view of the surgical field with high resolution for details. The chest is the most suitable body cavity for the minimal access approach not only because thoracotomy is a very painful incision but also because once the lung is collapsed (with selective one-lung ventilation) there is plenty of room for instrument manoeuvring. The use of carbon dioxide insufflation, and hence valved ports, is unnecessary. Conventional thoracic instruments can be placed directly through small wounds into the chest. We prefer to use conventional instruments over the dedicated, disposable endoscopic ones whenever possible as they are familiar to the surgeons, easy to use, and cost-effective. For some procedures, a utility thoracotomy is required for retrieval of the specimen. Therefore, VATS should be viewed as a spectrum with the purely endoscopic approach at one end and a video-assisted approach at the other end of a continuum.

Although thoracoscopy can be performed under local anaesthesia, most VATS procedures are performed under general anaesthesia with selective one-lung ventilation. This is achieved using either a double lumen tube or a single lumen tube with a bronchial blocker. For most procedures, the patient is placed in the full lateral decubitus position with flexion of the operating table to open up the intercostal spaces on the operating side (Fig. 2). The patient is prepared and draped as for full thoracotomy.

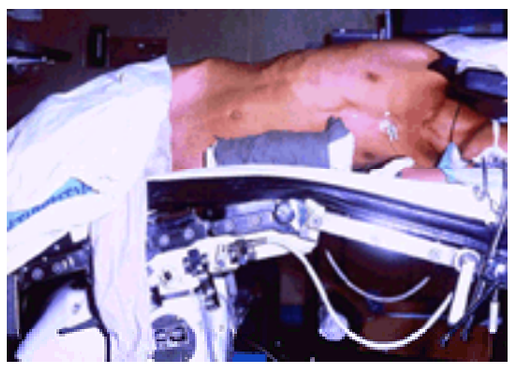


Fig. 2. Flexing the operating table opens up the intercostal spaces for instrumentation.

General exploratory thoracoscopy is usually performed by placing the telescope in the sixth intercostal space, mid-axillary line unless otherwise directed by the location of the pathology. We recommend the ‘finger-clamp' technique to enter the pleural cavity as in the insertion of a chest drain. The inferior placement of the telescope ensures a panoramic view of the chest. For simple procedures like pleural biopsy, an operating telescope (carrying an instrument channel) is recommended. For more complicated procedures, usually two additional instrument ports are required. Their positions depend on the target pathology although some general principles should be followed: (1) instruments should be placed far apart from each other and from the telescope (‘triangulation' strategy) to avoid ‘fencing' during instrument manoeuvring; and (2) avoid paradoxical motion by positioning the telescope and instruments within the same 180° arc, that is, approach the lesion from the same general direction. In situations where paradoxical motion is generated, it can be counteracted by rotating the camera 180°.

Indications for VATS

A survey conducted on North American thoracic surgeons showed VATS has become the preferred or accepted approach over a wide range of thoracic procedures. It is important to emphasize that VATS represents a new approach and not a new operation. Therefore, we adhere to the same basic surgical principles in VATS as we would for open thoracotomy.

Contraindications to VATS are relatively few. In addition to the general contraindications like recent myocardial infarction and severe coagulopathy, specific contraindications include pleural symphysis and inability to tolerate selective one-lung ventilation. The former is relatively uncommon and moderate adhesions can usually be taken down using a combination of sharp and blunt dissection under videoscopic vision. Prior operation on the same side of the chest should not be regarded as a contraindication.

Some of the common procedures performed by VATS are listed in Table 1. As experience is gained, more and more operations are technically approachable by VATS. However, clinical judgment must be exercised so that we are not compromising the long-term benefit for the sake of the minimal access technique. It is important to point out that while there is now a wealth of literature on VATS, there are relatively few publications comparing a VATS procedure with its conventional (thoracotomy) counterpart in a randomized, prospective manner. None the less, because of their low morbidity and good short- and long-term results, many VATS procedures are now well accepted as the approach of choice by the thoracic surgical community. Detailed discussion of each procedure is beyond the scope of this chapter and the readers are referred to our specialized textbook on this subject. We present an overview of the subject in the next few pages.

Diagnostic	
	Pleural biopsies
	Staging of intrathoracic tumours
	Wedge resections of diffuse pulmonary infiltrate
	Mediastinal mass biopsies
Therapeutic	
	Empyema and haemothorax
	Spontaneous pneumothorax
	Sympathectomy/splanchnicectomy
	Pericardial window
	Benign oesophageal disorders
	Resection of benign mediastinal masses
	Anatomical lung resection
	Lung volume reduction surgery
	Thymectomy

Table 1 Some common indications for video-assisted thoracic surgery

Diagnostic modality

The use of VATS as a diagnostic modality is well established and this includes biopsy of the pleura, lung mass, diffuse lung infiltrate, mediastinal mass, pericardium, and vertebral body. For simple procedures, the use of miniaturized instruments (‘needlescopic' instruments of 2 mm or less in diameter) provides an attractive option which could further reduce postoperative discomfort. However, diminished illumination and reduced resolution as well as flexibility of the equipment render it difficult to control fine movements. The use of VATS in staging intrathoracic tumours requires some qualification. VATS is a useful adjunct to mediastinoscopy in the biopsy of several lymph node stations not accessible by the latter approach (like the aortopulmonary, para-oesphageal, and inferior ligament nodes) for patients with primary lung cancer. However, mediastinoscopy should remain the principal diagnostic modality as VATS would not be able to provide information on contalateral mediastinal node involvement (N3 disease or stage IIIB). None the less, VATS plays an important role in excluding patients with pleural metastasis for an unnecessary thoracotomy. It is now our routine to perform VATS exploration in all patients with intrathoracic malignancy. In addition, VATS exploration provides information regarding chest wall invasion and leads to proper planning of the thoracotomy. VATS also plays a role in the secondary staging of lung cancer, that is, in patients with stage IIIA disease who have undergone a course of neoadjuvant chemotherapy in whom repeat mediastinoscopy may be technically difficult and potentially hazardous.

Therapeutic modality

Pleural effusions

For therapeutic procedures, the role of VATS in the drainage of loculated effusions (including empyema and haemothorax) and pleural debridement is well established. It is important to remember that if simple drainage is inadequate, VATS exploration should be recommended early before the empyema progresses from a fibrinopurulent phase to an organized fibrotic phase, resulting in restricted pulmonary function from encasement or fibrothorax. The use of VATS to guide the proper placement of a chest drain should not be underrated.

Pneumothorax

The tremendous success of VATS in the treatment of primary spontaneous pneumothorax has led to earlier referral by physicians and increased acceptance by patients for surgery. Stapled resection of apical bullas followed by mechanical pleurodesis remains the most frequently used technique, although more cost-effective means of eliminating the bullas (like suturing or looping) have been developed. While cases of primary spontaneous pneumothorax are easily approachable by VATS, treatment of secondary spontaneous pneumothorax (with established lung pathology like emphysema or pneumoconiosis) requires more clinical judgment. Patients with difficult adhesions to take down may be more suitable for thoracotomy, while those who are elderly with multiple comorbidities may benefit more from a chemical pleurodesis (we prefer talc slurry) if the lung can be fully re-expanded.

Mediastinal cysts

Benign mediastinal cysts like oesophageal duplication cysts (Fig. 3), thymic cysts, and pericardial cysts (Fig. 4) are surprisingly easy to resect using VATS. Decompression of the cyst with a needle may be necessary to facilitate mobilization. The phrenic nerve should be identified and carefully preserved. For oesophageal duplication cysts, a fibreoptic light source (from a bronchoscope) placed in the oesophagus can help identify this structure thoracoscopically during dissection. As a rule, for benign cysts, the tissue plane is fairly well preserved. In fact, any suspicion of tissue plane invasion by tumour should call for conversion to a full thoracotomy for further dissection.

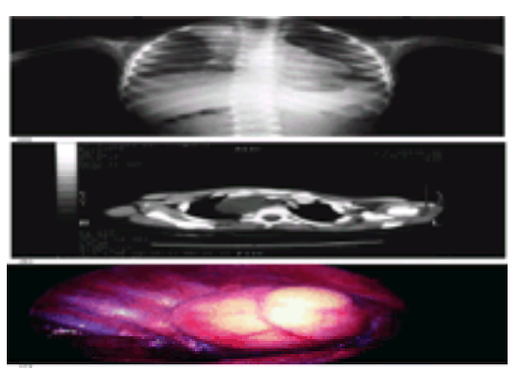


Fig. 3. A 2-year-old female with a right-sided mediastinal mass on chest radiography (a) and CT scan (b). Thoracoscopic view of a cystic structure that was resected thoracoscopically and later confirmed to be an oesophageal duplication cyst (c).

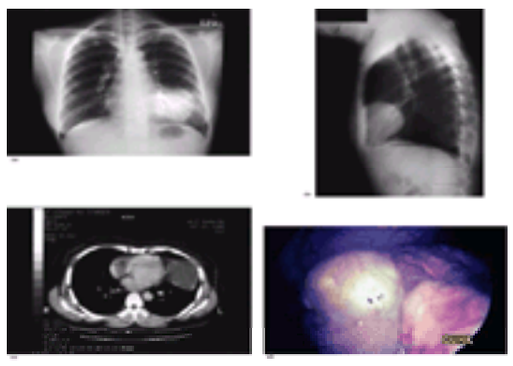


Fig. 4. A middle-aged female with a mass on the left side of the chest on radiography (a,b). CT scan shows a cystic mass next to the pericardium (c). Thoracoscopic view of the pericardial cyst (d).

Sympathectomy and splanchnicectomy

Thoracodorsal sympathectomy is indicated for patients with palmar and axillary hyperhidrosis, and selected patients with reflex sympathetic dystrophy, and vasculopathies like Raynaud's or Buerger's disease. A segment of the sympathetic chain can be excised or ablated (for example with electrocautery), the extent of intervention being dependent on the patient. For example, for palmar hyperhidrosis alone, only the T2 segment needs to be excised, but if axillary hyperhidrosis is also a problem, the level of resection (or ablation) should be extended to include T4. Even with VATS, a variety of surgical techniques are available. We continue to favour full lateral decubitus positioning and general anaesthesia with selective one-lung ventilation using three ports. The first rib can be identified by noting the position of the subclavian artery (which crosses over the first rib to supply the upper extremities). The second intercostal vein usually crosses over the sympathetic chain just underneath the parietal pleura and this has to be cauterized or clipped. We prefer to divide the sympathetic trunk proximally just below the stellate ganglion. Some surgeons still prefer resecting part of the stellate ganglion (because in 10 per cent of patients, sympathetic innervation to the upper extremity occurs through this ganglion). However, because of the risk of Horner's syndrome, this has not been our practice.

The success rate of this procedure for hyperhidrosis is high and ranges from 85 to 95 per cent in large series from Taiwan. Compensatory truncal hyperhidrosis occurs in about 20 per cent of our patients and occasionally can be difficult to treat. This has to be explained fully to the patients before surgery.

Splanchnicectomy is indicated for patients with intractable visceral pain arising from the upper abdomen. The greater splanchnic nerve arises from T5 to T10 and the lesser splanchnic nerve from T10 to T11 of the sympathetic chain. Likewise, the nerves can be excised or ablated. In over half the patients, a unilateral approach from the left is effective and should be tried before resorting to a bilateral approach.

Pericardial window

VATS provides a safe and effective approach to the drainage of pericardial effusion of both benign and malignant aetiology. Large pericardial windows can be created both anterior and posterior to the phrenic nerve for drainage. When pericardial effusion is associated with pleural effusion, the pericardium should be approached from the side with the pleural effusion (or with more effusion in bilateral disease), otherwise a right approach is usually preferred as the right hemithorax is larger than the left, giving a better perspective for VATS evaluation. Caution has to be exercised in the use of monopolar electrocautery to avoid touching the heart and precipitating dysrhythmia. Subxiphoid drainage remains a viable alternative approach which can be performed under local anaesthesia in patients who are very ill and debilitated as well as those with a history of bilateral chest surgery or pneumonia when troublesome adhesions are anticipated. The video-assisted subxiphoid approach using a videomediastinoscope is a recent refinement of the old procedure.

Thymectomy

Thymectomy is an established therapy in conjunction with medical treatment for generalized myasthenia gravis. However, which technique to use continues to be a subject of controversy. Several surgical approaches are currently being used and these include transsternal, transcervical, a combination of the two (for ‘maximal thymectomy'), and recently, VATS. Our own results, as well as the collective data from other centres, showed that there was little difference in symptomatic improvement following thymectomy regardless of which surgical approach was used. The complete remission rate following VATS seemed to be slightly lower than the ‘maximal thymectomy' approach, but we believe this was due to the relatively shorter term follow-up of the patients who received VATS. The natural history of myasthenia gravis following thymectomy is that more patients will go into and stay in complete remission as time progresses. Therefore, long-term follow-up of these patients is important.

Although the thymus can be approached from either side with VATS, we prefer the right-sided approach for several reasons: (1) the superior vena cava forms an easy landmark for dissection; (2) the confluence of the two branchiocephalic veins, which is a difficult area to dissect, can be approached more easily from the right; and (3) it is easier for a right-handed surgeon to dissect the thymus from below from the right. Care has to be taken to dissect out the superior horns completely as they can extend way up into the neck. However, with firm, deliberate downward traction of the mobilized gland (after all the vascular branches and tributaries have been divided), the superior horns can be freed from their fascial attachment (a continuation of the pretracheal fascia). We have found that a dental pledget mounted on a conventional curved clamp is a very useful device for the blunt dissection of the thymus. Complete removal of the thymus, irrespective of the approach, is important to the success of this operation and, therefore, we advocate that the VATS approach should only be performed by surgeons with considerable experience of this technique.

VATS thymectomy definitely has several distinct advantages over the other approaches. It is much less painful than the transsternal approach and recovery is much quicker. The thymus is essentially an anterior mediastinal structure and, therefore, it is more direct to approach it from the chest than from the neck. The transcervical approach has the disadvantage of instrument crowding through a small, single access. Cosmesis following VATS is excellent. Although this on its own should only be regarded as a bonus rather than the main reason for choosing a particular surgical approach, the majority of patients with myasthenia gravis are young females who care a lot about the surgical scars. As there is some evidence that the sooner the patients with generalized myasthenia gravis have surgery, the better is the long-term outcome, it follows that the VATS approach could encourage earlier acceptance of thymectomy by the patients and earlier referral of these patients by their neurologists.

Anatomical lung resections

The application of VATS to anatomical lung resections continues to be a subject of considerable controversy. Even among surgeons practising VATS, only a few use this approach for lobectomy. Anatomical dissection performed through a minithoracotomy in an essentially closed chest raises questions on the safety of the technique; resection for intrathoracic malignancy casts doubt on adequate clearance; the long-term benefits of VATS over a conventional thoracotomy approach are uncertain and the high cost of the consumables and endoscopic equipment questions the cost-effectiveness of this approach in the current era of cost containment.

However, despite all the scepticism, intermediate-term results from several centres performing VATS lobectomy have been very encouraging. The survival figures for lung carcinoma following VATS resection are at least as good if not better than a similar group of patients following resection through a conventional thoracotomy. If these reports can be substantiated by wider experience, it could have an important impact on the surgical management of thoracic malignancy. The reason for the improved survival remains unclear at present, but there is circumstantial evidence that by minimizing chest wall trauma, the body inflammatory response is dampened and immune function better preserved.

The actual technique of VATS lobectomy is by no means a unified approach. Dr Ralph Lewis and his group in New Jersey, America advocate the simultaneous stapling technique for the bronchus and pulmonary vasculature, while other surgeons including us continue the individual ligation technique of the hilar structures. There is little consensus on the size of the minithoracotomy and, more importantly, the use of a rib spreader. We recommend that the term ‘VATS lobectomy' should be reserved for the predominantly endoscopic technique with little or no rib spreading (Fig. 5), while the term ‘minithoracotomy with video-assistance lobectomy' should be used when rib spreading is routine and the surgeons operate by looking through the minithoracotomy. Mediastinal lymph node sampling or dissection by the VATS approach has been described.

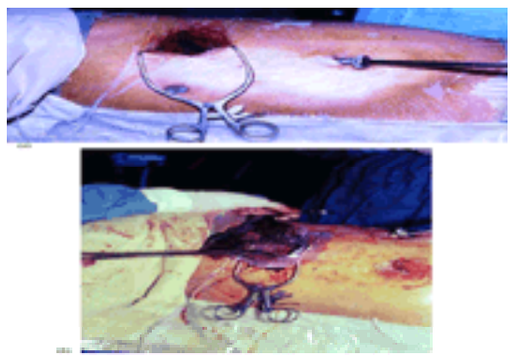


Fig. 5. By maximally opening up the intercostal spaces, a 3.5 cm ‘gap' can be obtained between ribs (a). Only soft tissue retraction is usually necessary for the entire operation, including specimen retrieval (b).

Regardless of the exact technique, it is generally agreed that careful patient selection is essential. Our own patient selection criteria for tumour resection include stage I non-small cell lung cancer (without evidence of endobronchial or chest wall involvement), a tumour size of less than 4 cm, and complete or near complete fissures (a thoracoscopic assessment).

We believe that VATS lobectomy is a feasible and safe procedure in experienced hands and may be of particular benefit to the elderly and patients with multiple comorbidity who are otherwise poor-risk candidates for conventional thoracotomy. The exact role of this procedure, however, awaits its long term results compared with thoracotomy.

Other applications of VATS

Many thoracic procedures are being rediscovered through the thoracoscope, some of these also involve the other surgical subspecialties. Thoracoscopic spinal surgery is receiving increasing attention by the orthopaedic community. For spinal deformity, anterior spinal release as well as instrumentation are now feasible.

VATS is also playing an important role in minimal access cardiac surgery. The left internal mammary artery can be harvested with thoracoscopic assistance through an anterior minithoracotomy. This incision is subsequently used for the anastomosis of the left internal mammary artery to the left anterior decending coronary artery in a procedure now referred to as minimally invasive direct coronary artery bypass grafting. The thoracoscope has also found use in minimal access mitral valve surgery and a totally endoscopic approach to valve replacement and coronary revascularization (port access approach).

Future prospects

Cardiothoracic surgery is undergoing a rapid flux of evolution as the development of videothoracoscopy revolutionalizes its practice. The question now is, does VATS, as we currently practice it, represent an end point that only requires minor refinements or is it an intermediate step to an even less invasive approach? We believe that both views may be correct. VATS represents a spectrum with a purely endoscopic approach at one end and a video-assisted approach (with a utility minithoracotomy) at the other end. For the purely endoscopic procedures, there have been attempts to modify further the surgical access and mode of anaesthesia. The former resulted in the development of 2-mm ‘needlescopic' instruments and the latter in therapeutic thoracoscopy under local anaesthesia. It is entirely possible that in the near future, simple thoracoscopic procedures could be performed under local anaesthesia via an essentially percutaneous route with miniaturized instruments as an outpatient procedure. However, it is important to remember that carefully conducted clinical trials should precede the general acceptance of a new technique or technology, no matter how attractive it may appear initially.

Further reading

Braimbridge MV. The history of thoracoscopic surgery. *Annals of Thoracic Surgery* 1993; **56**: 610–14. [An excellent historical account of thoracoscopy.]

Landreneau RJ *et al.* Video-assisted thoracic surgery: basic technical concepts and intercostal approach strategies. *Annals of Thoracic Surgery* 1992; **54**: 800–7. [An important paper on the basic technique and manoeuvring strategy in video-assisted thoracic surgery.]

Lewis RJ, Kunderman PJ, Sisler GE, Mackenzie JW. Direct diagnostic thoracoscopy. *Annals of Thoracic Surgery* 1976; **21**: 536–9. [An important historical paper in which, for the first time, the thoracic surgeons presented a series of 40 patients who had direct thoracoscopy for diagnosis under general anaesthesia with excellent results.]

Mack MJ, Scruggs GR, Kelly KM, Shennib H, Landreneau RJ. Video-assisted thoracic surgery: has technology found its place? *Annals of Thoracic Surgery* 1997; **64**: 211–15. [A survey on the thoracic surgical community in North America, reflecting the changing attitude of surgeons to evolving technology.]

Yim APC, Izzat MB, Lee TW. Counteracting paradoxical motion in videoendoscopic operations. *Annals of Thoracic Surgery* 1998; **66**: 965. [Description of a simple trick to counteract paradoxical motion in videoscopic operations by turning the video camera.]

Yim AP, Hazelrigg SR, Izzat MB, Landreneau RJ, Mack MJ, Naunheim KS, eds. *Minimal access cardiothoracic surgery*. WB Saunders, Philadelphia, 1999. [The most comprehensive textbook to date on video-assisted thoracic surgery and minimally invasive cardiac surgery, with over 80 chapters by leading experts covering a wide range of topics.]

15 Endoscopic surgery

Timothy M. Farrell and John G. Hunter

Introduction
History of minimally invasive surgery: a history of imaging
 Advances in optics
 Advances in radiology
Rationale for endoscopic surgery
Indications for endoscopic surgery
Training in endoscopic surgery
 Organization of the endoscopic surgical unit
Physiology of endoscopic surgery
 Metabolic effects of pneumoperitoneum
 Mechanical effects of pneumoperitoneum
 The physiology of thoracoscopy
 The physiology of extracavitary endoscopic surgery
Anesthetic considerations
Access for endoscopic surgery
 Intracavitary procedures
 Subcutaneous procedures
 Extraperitoneal procedures
Instrumentation for endoscopic surgery
 Imaging systems
 Energy sources for endoscopic surgery
 Balloons and stents
 Hand instruments
Room set-up for endoscopic surgery
Laparoscopic surgical procedures
 Laparoscopic cholecystectomy
 Laparoscopic appendectomy
 Laparoscopic repair of inguinal hernia
 Laparoscopic fundoplication
 Laparoscope-assisted colectomy
 Laparoscopic splenectomy
 Laparoscopic adrenalectomy
 Laparoscopic exploration of the common bile duct
Special considerations in minimally invasive surgery
 Pediatric endoscopic surgery
 Endoscopic surgery in pregnancy
 Endoscopic surgery and cancer
 Endoscopic surgery in elderly and infirm individuals
Economics of endoscopic surgery
Outcomes of minimally invasive surgery
Legal and ethical issues
Future directions
Further reading

Introduction

Endoscopic surgery is not a surgical subspecialty, but rather refers to a philosophy that encompasses all traditional surgical disciplines. Also dubbed *minimally invasive surgery*, the term endoscopic surgery describes the performance of major operations through small incisions, often using miniaturized, fiberoptic imaging systems. In many cases, endoscopic methods have replaced traditional, open operations because they produce less pain and a more rapid recovery.

The endoscopic surgical revolution has transformed the social structure of medicine. Conventional practice boundaries have been obscured and *non*traditional working alliances have resulted. Internists, emboldened by the accomplishments of interventional radiologists, now routinely make the small incisions necessary for minimally invasive procedures. However, the surgeon's ability to choose the best mode of access for an individual patient, and to deal at once with complications, provide the rationale for keeping minimally invasive therapy in the surgical arena.

History of minimally invasive surgery: a history of imaging

Advances in optics

Although the explosion in the use of endoscopic surgery is relatively recent, the history of the technique dates back nearly 100 years. In 1901, Kelling performed primitive laparoscopy by placing a cystoscope, illuminated by hot elements at its tip, into an inflated abdomen. In the late 1950s, Hopkins, a physicist in Cambridge, UK, described a method for transmitting light through a solid quartz rod without delivering heat and with little loss of light. At approximately the same time, landmark discoveries in the field of fiberoptics permitted rapid development of the first flexible endoscopes. In the early 1970s, flexible endoscopy found widespread application as a diagnostic tool. Later in that decade, rigid and flexible endoscopes were modified to allow interventions for various pathologic conditions. The development of compact, high-resolution, charge-couple devices, which may be mounted on the internal end of flexible endoscopes or the external end of a Hopkins telescope, has catalyzed the explosion of video surgery over the last 10 years. Coupled with high-power light sources, fiberoptic cables, and improved video monitors, charge-couple devices permit the high-resolution video-endoscopy that has dramatically changed our understanding of surgical anatomy and radically altered surgical practice.

Advances in radiology

In addition to advances in invasive optical imaging during the 1970s, substantial technologic improvements in radiologic imaging inspired the development of other innovations. Fluoroscopy has permitted myriad minimally invasive procedures, including the percutaneous administration of balloons and stents for occlusive vascular disease or biliary obstruction. An important example of fluoroscopic, minimally invasive wizardry is the transvenous intrahepatic portosystemic shunt, which has become a standard for portal–systemic decompression.

Few endoscopic procedures more complex than biopsy or the drainage of fluid collections employ ultrasonographic imaging only, owing to its nonintuitive image. Nonetheless, endoscopic and laparoscopic ultrasonograms are becoming invaluable tools for the diagnosis of conditions hidden beyond the view of the endoscope. Tumor staging, determining surgical resectability, and the detection of bile-duct stones are important roles for ultrasound.

Advances in axial imaging techniques, such as spiral computed tomography (**CT**), allow today's interventional radiologists to place precisely catheters for drainage or biopsy, often obviating the need for open operation. The importance of percutaneous procedures for the drainage of intra-abdominal fluid collections, or for the biopsy of lung, liver, pancreas or kidney masses, is too often taken for granted by physicians who did not personally experience the redefinition of medical treatments inspired by these technologic developments. New uses of therapeutic spiral CT scanning include ‘real time' imaging of percutaneous procedures such as discectomy and sympathectomy.

Magnetic resonance imaging (**MRI**) has become an extremely valuable diagnostic tool, especially in neurology and orthopedics. In the next decade, diagnostic endoscopy, especially endoscopic retrograde cholangiopancreatography (**ERCP**), may be replaced with MRI and computer-assisted image reconstruction. However, the long duration of data acquisition and image production, the bulkiness of the scanner, and the obvious incompatibility with metal instruments and monitoring devices, have limited its clinical implementation. The advantages of MRI are its high-resolution images and lack of radiation. ‘Open magnets' allow access to the patient between two large MRI coils, but the narrow separation of the magnets is still too restrictive for most interventions. Neurosurgeons in several centers have begun accumulating experience using MRI to perform frameless stereotactic surgery.

The history of endoscopic surgery is the history of image refinement. The design of specialized surgical instruments and methods for delivering them were of secondary importance, and have followed rapidly behind advances in imaging. In general, today's minimally invasive surgeons perform procedures using optical imaging systems,

rigid telescopes, flexible endoscopes, and operating microscopes. Procedures employing fluoroscopy and ultrasound may now be performed by a surgeon trained in use of the technology and the interpretation of the images. The next generation of surgeons may need to become facile with MR-guided endoscopic surgery as well.

Rationale for endoscopic surgery

To understand better the revolution in endoscopic surgery, one must re-examine the historical emergence of its parent, *open cavitary surgery*. Contemporary principles of open surgery matured after the discovery in the mid-nineteenth century of general anesthesia and in the late nineteenth century of the concept of asepsis. These developments enabled major procedures to be performed without prohibitive complications and fatality. The development of antibiotics, positive-pressure ventilation, and the cardiopulmonary bypass further improved surgical outcomes. Advances in clinical pharmacology and bioengineering necessarily preceded the evolution of organ transplantation and joint replacement. Subsequent advances in parenteral nutrition, advanced antibiotics, and invasive and noninvasive physiologic monitoring have further refined surgical standards. Before the revolution in endoscopic surgery, most surgical research was focused on improving interventions for the more devastating conditions. These interventions frequently resulted in ‘heroic' operations that challenged the limits of human physiology and depended upon advanced supportive measures. The benefits of elective surgery for chronic conditions were often overshadowed by the pain, disability, and disfigurement associated with the incision.

Minimally invasive surgery changed all that. The emergence of endoscopic surgery has been akin to a *populist* revolution, with patients and referring physicians eagerly pursuing the option of a surgical solution to a chronic problem, even where medical therapy is effective. Market forces have promoted the rapid dissemination of endoscopic surgery, as patients have sought out surgeons capable of providing the state-of-the-art service.

Indications for endoscopic surgery

The possibility of an endoscopic approach is indicated when a surgeon is able to achieve similar or better outcomes than with the traditional approach. Conversely, minimally invasive procedures are not indicated when outcomes are inferior to those of traditional surgery. Before the era of laparoscopic cholecystectomy, extracorporeal shockwave lithotripsy and percutaneous biliary lithotripsy failed to displace cholecystectomy because they were neither effective nor definitive for the management of biliary stones. However, laparoscopic cholecystectomy was rapidly acclaimed because it is as efficacious and has fewer complications than open cholecystectomy.

Extremely debilitated or elderly patients are often exceptions to the ‘equivalency rule.' While the most definitive management of choledocholithiasis might be cholecystectomy and exploration of the bile duct (or endoscopic sphincterotomy), even the laparoscopic approach may pose too great a physiologic challenge. In selected cases, the surgeon may choose to provide a ‘less ideal' intervention to minimize iatrogenic harm. For example, acute cholecystitis in a debilitated patient may be managed by tube cholecystostomy, and cholangitis resolved by placing an endoscopic or percutaneous biliary stent. Although drainage is less durable than laparoscopic cholecystectomy and bile-duct exploration, the reversal of life-threatening sepsis will permit more definitive therapy (open or endoscopic) at a safer time.

Training in endoscopic surgery

Traditionally, surgeons master new operations by ‘apprenticing' themselves to a colleague. The arrangements are formal (fellowships) or informal (apprenticeships). The apprentice surgeon scrubs with his or her teacher until both feel sufficient competence has been achieved. New techniques are thus disseminated from an expert surgeon or home institution after apprentice surgeons have completed their training. This system works best for techniques that extend rather than replace traditional operations, such as the ileoanal pouch for ulcerative colitis and transplantation for the failing liver.

In contradistinction, endoscopic surgery has *replaced* many open procedures, including the so-called bread-and-butter operations performed by community surgeons. Thus, it was economically imperative that community surgeons acquire expeditiously skills and judgement in endoscopic surgery. However, the traditional model could not satisfy this need, since there were insufficient experts from whom young surgeons could solicit additional training. Therefore, short courses served to introduce the techniques, but effectively left trainee surgeons to learn the subtle aspects ‘on the job.' While brief training periods may be effective for learning small variations on a conventional procedure, these courses frequently did not allow trainees to learn completely laparoscopic cholecystectomy, for example, as evidenced by the unacceptably long operative times, and high rates of conversions and complications in the early experience. Now, experts in endoscopic surgery are available in most academic hospitals. As advanced endoscopic procedures become the standard of care, it is advisable that established surgeons take prolonged apprenticeships with an expert before venturing out alone. Ultimately, training for minimally invasive surgery will take place in ‘virtual reality' surgical simulators, but the technology and computing speed necessary to accomplish this goal have not yet been realized. Until then, surgical residency programs bear the obligation of teaching basic laparoscopic techniques (laparoscopic cholecystectomy, diagnostic laparoscopy). Although some advanced procedures (laparoscopic fundoplication, laparoscopic adrenalectomy) are also well established, most residency training programs remain unprepared to provide that level of instruction. In this light, there still exists a need for fellowship training in advanced endoscopic surgery to develop those who will then train the next generation.

A final model of training exists, however futuristic. Many believe that a true ‘hybrid' professional could be trained to deliver minimally and maximally invasive therapy without respect for traditional practice boundaries. For example, such an individual might be trained with capabilities in traditional surgery, flexible endoscopy, rigid endoscopy, and radiologic techniques. This individual would have an orientation toward a particular organ system, but would not be tied to a particular technology or access technique. This new breed of surgeon would be best equipped to choose the appropriate approach for his or her patient, and also manage any complications that might ensue. Patients would be spared the frustration of having to visit the gastroenterologist, radiologist, and surgeon to have a problem managed optimally.

In spite of these rational arguments, the large economic and sociologic barriers that separate specialties will not fall quickly. For the time being it appears that most training in endoscopic surgery will be within the current constructs of traditional postgraduate residency and fellowship education. Therefore, the best approximation available in the current system is the ‘team approach.' Multidisciplinary conferences and clinics are becoming commonplace, and interactions between different specialists provide ‘*heterozygous vigor*' to the care of patients and broad-based experience for trainees in endoscopic surgery.

Organization of the endoscopic surgical unit

One model for the delivery of minimally invasive therapy might be termed ‘technology centered.' Often the complex technologies and techniques required for the treatment of various disease processes are similar, and are most efficiently and cost-effectively delivered by specialized providers, as in many interventional radiology and critical-care units. The flaw in this model is that the technologic expert may not be the master or mistress of the disease for which the technique is indicated. Should lymph-node dissection be performed by the best laparoscopist or by the physician most knowledgeable about the pattern of lymphatic dissemination of cancer? Should ultrasound be in the domain of the radiologist or of the surgeon most familiar with the anatomy and pathology of the disease process?

Endoscopic surgery requires a well-organized team directed by an attending surgeon with comprehensive knowledge of its procedures, although he or she may have a specific orientation (hepatobiliary, foregut, colorectal, vascular, thoracic). This individual serves as the medical director, to coordinate the purchase of appropriate equipment and to direct institutional training requirements. An operating-room nurse-coordinator is a essential partner to the medical director. Occasionally, it may be appropriate to allow each nursing specialty to coordinate endoscopic surgery for their section, but the large overlap between laparoscopy, pelviscopy, thoracoscopy, and extracavitary endoscopy impels the medical director (and nurse-coordinator) to oversee all technologic acquisitions. It is critical that the purchase of flexible and video-endoscopes, and reusable and disposable instrumentation, be coordinated centrally to avoid equipment incompatibility.

Other valuable members of the team include a material manager, a video engineer, a hospital safety officer, a research assistant, and an administrative assistant. Generally, areas of endoscopic surgery not controlled by surgeons (flexible endoscopy, interventional radiology) remain outside the domain of the director of endoscopic surgery. In many places, informal alliances between gastroenterologists, radiologists, and surgeons caring for patients with a given disease process have evolved into well-organized multidisciplinary clinics, with shared administrative and technologic resources. This extension of the ‘team' model is optimally designed to maximize the efficiency of patient care and research efforts.

‘Turf wars' over procedural domains are minimized when individuals from different backgrounds work closely together. The optimal care of the patient with an adrenal lesion requires the cooperation of the best laparoscopist and the best surgical endocrinologist. The multidisciplinary ‘team' model is at present the best approximation of the ‘hybrid' professional and offers the greatest benefit for patients.

Physiology of endoscopic surgery

The least invasive endoscopic surgical procedures evoke no significant cardiovascular, endocrinologic, or immunologic alterations. Procedures such as stereotactic breast biopsy or flexible gastrointestinal endoscopy often require only minimal sedation and/or analgesia. Procedures requiring intraperitoneal or intrathoracic manipulations are more invasive and have more profound physiologic impact. In addition, exposure of tissues within the peritoneal cavity requires that the abdominal wall be lifted away from the abdominal organs. The most commonly employed method, positive-pressure pneumoperitoneum, provides further physiologic

derangement.

Metabolic effects of pneumoperitoneum

In early attempts at laparoscopy, air insufflation was achieved by using a sphygmomanometer bulb. The slow reabsorption of the nitrogen fraction of room air by the blood and the peritoneum was believed to cause unnecessary postoperative discomfort. Subsequently, both carbon dioxide and nitrous oxide were tested for abdominal insufflation. Nitrous oxide is physiologically inert, rapidly absorbed, and provided better analgesia for laparoscopy performed under local anesthesia. Carbon dioxide is also rapidly absorbed, and had the added advantage of suppressing combustion. Thus, it became the preferred gas for positive-pressure pneumoperitoneum.

Carbon dioxide, which is rapidly absorbed across the peritoneal membrane into the circulation, creates a respiratory acidosis by the generation of carbonic acid. Physiologic buffers, mostly in bone, can neutralize up to 120 l of carbonic acid, minimizing the hypercarbia during short laparoscopic procedures. Once these buffers are saturated, respiratory acidosis will develop rapidly unless minute ventilation is increased. Patients with normal respiratory function tolerate increases in the ventilatory rate or vital capacity on the ventilator. However, respiratory rates above 20/min may impair gas exchange, and large increases in vital capacitymay induce barotrauma and cause respiratory motion that limits visualization in the upper abdomen. Sometimes the surgeon must diminish or release the pneumoperitoneum to allow the hypercarbia to resolve. Although mild respiratory acidosis is insignificant, if more severe it may cause cardiac arrhythmia, tachycardia, or increased systemic vascular resistance with increased myocardial oxygen demand (Fig. 1).

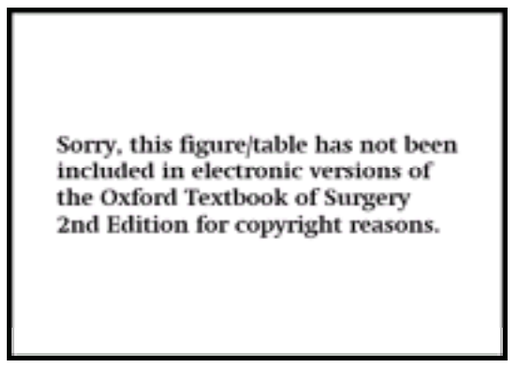


Fig. 1. Carbon dioxide gas insufflated into the peritoneal cavity has both local and systemic effects that may cause complex hemodynamic and metabolic alterations.

It was predicted that the stress response to laparoscopic surgery would be less than for the same operation done via laparotomy. However, several researchers found that during laparoscopic operations the serum cortisol increased more abruptly, and normalized more rapidly, than during identical open procedures. Analysis of cytokine release has revealed a similar trend towards more rapid normalization after laparoscopy than laparotomy. Measures of immune function, such as the delayed-type hypersensitivity response, are more depressed by laparotomy than laparoscopy.

Mechanical effects of pneumoperitoneum

Other variables at work during laparoscopic surgery may affect cardiovascular physiology, especially in the hypovolemic individual (Fig. 1). Increased intra-abdominal pressure transmitted to the inferior vena cava and thorax, the use of a steep, reverse Trendelenburg position, and the loss of muscle tone in the lower extremities may impede venous return and cardiac output. However, if intraperitoneal pressures are maintained below 20 mmHg, cardiac output is rarely affected. Rapid stretching of the peritoneum may induce a vagovagal response with resultant bradycardia and/or hypotension. The appropriate response for any hemodynamic alteration is rapid desufflation of the abdomen, with volume replacement and vagolytic agents (atropine) as necessary. Increased intra-abdominal pressure transmitted through the paralyzed diaphragm will also increase peak inspiratory pressure, but the risk for clinically significant barotrauma and pneumothorax is small after uncomplicated laparoscopic surgery.

Another consequence of positive-pressure pneumoperitoneum and reverse Trendelenburg positioning is diminished venous return from the lower extremities, with increased risk for deep venous thrombosis and pulmonary embolism. Therefore, the use of sequential compression stockings or subcutaneous heparin prophylaxis is recommended. Short laparoscopic procedures such as appendectomy, hernia repair or cholecystectomy may have less risk for venous thrombosis and may not warrant routine prophylaxis, but this possibility remains unproved.

The obstructive effect of increased intra-abdominal pressure on the kidney and renal vein, as well as the subsequent increase in plasma renin and antidiuretic hormone, cause decreased urinary output. While the physical effects of the pneumoperitoneum on renal blood flow are immediately reversible, hormonally mediated changes may persist for up to 60 min. Intraoperative oliguria is common during laparoscopy, and does not usually indicate depletion of the intravascular volume. Since evaporative fluid losses are insignificant during laparoscopy, fluid supplementation is rarely necessary. The anesthesiologist must be taught not to expect large urine volumes during the pneumoperitoneum.

Patients with compromised cardiovascular function are at higher risk from a lengthy laparoscopic procedure. The surgeon should consider an alternative approach, or a reduced insufflation pressure. The use of helium, neon, or argon for insufflation is appealing because these gases are metabolically inert, but their poor solubility in blood (unlike carbon dioxide and nitrous oxide) increases the risk of gas embolism via the venous system. Gas embolism should be suspected if hypotension develops suddenly during insufflation, and the diagnosis may be made by listening for a ‘mill-wheel' murmur with an esophageal stethoscope. Placing the patient in the left lateral decubitus position with the head down traps the gas in the apex of the right ventricle. A rapidly placed central venous catheter may be directed into the ventricle to aspirate the gas.

A number of devices that create a working space by lifting the abdominal wall have been developed. These minimize the physiologic derangements typically associated with positive-pressure pneumoperitoneum, but are uniformly bulky and obstructive. Since abdominal lift devices do not increase intra-abdominal working room as effectively as pneumoperitoneum, the exposure is inferior. Lifting the anterior abdominal wall via the umbilical port causes a ‘pinching in' of the lateral flanks, with anteromedial displacement of the bowel into the operative field. Pneumoperitoneum, by virtue of its evenly distributed displacement of the abdominal wall, provides better exposure for laparoscopy.

The physiology of thoracoscopy

The physiology of thoracoscopic surgery is different from that of laparoscopy. Since the rigid structure of the thorax does not require positive pressure to create a working space, the physiologic impact from gas and positive pressure is not an issue. However, a double-lumen endotracheal tube is needed to deflate the lung on one side. If a positive-pressure pneumothorax is desirable—and in some settings positive pressure may help collapse the ipsilateral lung—one must keep the intrathoracic pressure to less than 10 mmHg to avoid developing a tension pneumothorax.

The physiology of extracavitary endoscopic surgery

Many new endoscopic surgical procedures involve extracavitary planes. Laparoscopic repair of inguinal hernia is generally performed in the anterior extraperitoneal space of Retzius. Laparoscopic nephrectomy may be performed with retroperitoneal laparoscopy. Operations on vessels in the lower extremities and endoscopic plastic surgery require dissection in purely nonanatomic planes. Some techniques depend on high-pressure gas insufflation to create and maintain a working space; others involve balloon dissection and low-pressure gas insufflation or lift devices for the same purpose. While these techniques may cause fewer physiologic perturbations than positive-pressure pneumoperitoneum, the absorption of gas from extraperitoneal sites may also result in significant metabolic acidosis.

Anesthetic considerations

Competent administration of anesthesia during laparoscopic surgery requires full awareness of the physiologic impact of the carbon dioxide pneumoperitoneum. The anesthesiologist and surgeon must communicate effectively during these procedures, since many cardiovascular and pulmonary complications are rapidly treatable by releasing the pneumoperitoneum. Furthermore, the absence of a laparotomy incision diminishes the requirement for narcotics and virtually eliminates the need for intravenous volume replacement for evaporative losses. Since endoscopic procedures are often done on outpatients, short-acting, well-tolerated anesthetic agents are preferable. Preventing nausea, pain, and urinary retention by selecting non-narcotic analgesics (e.g. ketorolac) and using antiemetics prophylactically reduces the need

for postoperative admission to hospital.

Access for endoscopic surgery

Intracavitary procedures

The anatomic orifices are the most natural portals of access for endoscopic surgery. The nares, mouth, urethra, and anus provide access to the respiratory, gastrointestinal, and genitourinary systems. Although these orifices allow entrance without incision, their utility is often limited by their distance from the target tissue. For example, to approach most of the small bowel via the mouth or anus is time-consuming and tortuous, making complex intraluminal procedures impossible at present.

Vascular access, either percutaneous or by ‘cut down,' has permitted general surgeons, cardiologists, vascular surgeons, and interventional radiologists to perform minimally invasive procedures ranging from Hickman catheterization to stent grafting for abdominal aneurysm. Devices may be directed into central sites within the cardiovascular system over a fluoroscopically placed guide wire, the so-called Seldinger technique. Similar techniques are employed outside the vascular tree for abscess drainage, percutaneous endoscopic gastrostomy, transhepatic biliary access, and percutaneous nephrostomy.

Most thoracoscopic procedures require general anesthesia and split lung ventilation. Access for thoracoscopic surgery is similar to that for placing a thoracostomy tube. A small incision is made over a rib and carried down through the pleura under direct vision. The ipsilateral lung is deflated and a plastic trocar is directed across the chest wall to admit a telescope. Once the lung is completely collapsed, subsequent ports are placed under video-endoscopic control. Because insufflation of the chest is unnecessary, simple plastic sheaths are sufficient to ease instrument insertion and to minimize trauma around the incisions.

The creation and maintenance of a pneumoperitoneum for laparoscopic surgery requires a one-way valve to prevent desufflation. Two methods exist for establishing abdominal access during laparoscopic procedures. The first, *direct puncture laparoscopy*, involves elevating the well-relaxed abdominal wall with two towel clips, placing a small incision in the umbilicus, and inserting a specialized, spring-loaded (Veress) needle into the abdominal cavity (Fig. 2). During the insertion of the Veress needle, two distinct layers of resistance are felt as the tip passes through the fascia of the abdominal wall and the peritoneum. The umbilicus is generally the preferred point of access as the abdominal wall is quite thin there, even in obese individuals. The peritoneal space is inflated with CO_2 gas from a pressure-limited insufflator to a maximal pressure of 15 mmHg. Laparoscopic surgery may be performed under local anesthesia using lower pressures and a nitrous oxide pneumoperitoneum, but general anesthesia is preferred for adequate muscle relaxation and patient comfort.

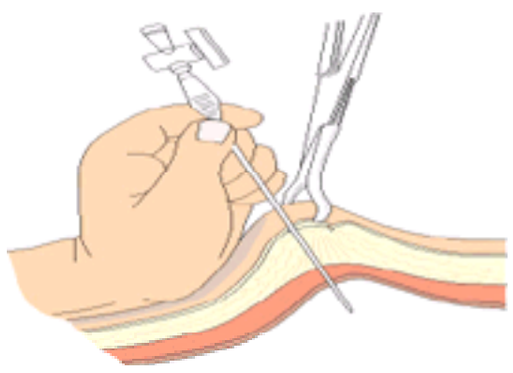


Fig. 2. Insufflation of the abdomen is accomplished with a Veress needle held at its serrated collar by thumb and forefinger.

After peritoneal insufflation, direct access to the abdomen is obtained with a 5- or 10-mm trocar. For safe direct-puncture laparoscopy, the use of a vented stylet (Karl Storz, Tuttlingen, Germany) or trocar with a safety shield (Ethicon EndoSurgery, Cincinnati, OH and others) is advised. During insertion, the trocar must be pointed away from the sacral promontory and great vessels, and directed toward the target area. Although direct-puncture laparoscopy is safe in experienced hands, there appears to be a higher incidence of injury to large vessels than with the alternative approach, particularly in inexperienced hands.

For the occasional laparoscopist, a second method, *direct peritoneal access* (the Hasson technique), is advisable. The surgeon makes a small, transverse incision below the umbilicus and locates the abdominal fascia under direct vision. Kocher clamps are placed on either side of the fascia, and a small incision is made into the peritoneum with curved scissors. A Hasson trocar is passed directly into the abdominal cavity, and secured by sturdy sutures placed on either side of the fascia (Fig. 3). This approach is also preferable for cases in which bowel may be adherent to the abdominal wall. Because of the difficulties in visualizing the abdominal region immediately adjacent to the primary trocar, it is advisable to pass the telescope through a secondary trocar to inspect the entry site before the end of the procedure.

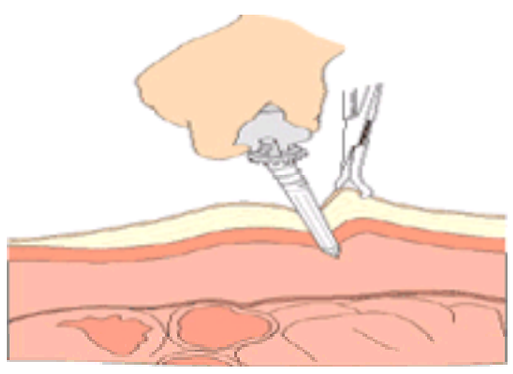


Fig. 3. Direct peritoneal access involves identification and incision of the peritoneum followed by placement of a specialized trocar to maintain a gas seal. The trocar is secured to the fascia by sutures.

After the insertion of the operating laparoscope, secondary punctures are made with 5- and 10-mm trocars under direct vision. Similarly, all trocars are removed under direct vision and the insertion sites are inspected for bleeding at the end of the operation. If bleeding occurs, direct internal pressure with an instrument from another trocar site, or balloon tamponade with a Foley catheter through the bleeding site, will usually stop it. Occasionally, a suture through the full thickness of the abdominal wall is necessary to arrest successfully persistent bleeding from a trocar site.

There is some controversy over which trocar sites require closure by suture at the fascial level. It is generally agreed that 5-mm trocar sites do not require closure while most 10-mm sites do, since failure to provide fascial repair risks visceral herniation. Above the level of the transverse mesocolon, 10-mm incisions are more protected from the risk of hernia formation, and may be closed selectively. Specialized suture delivery systems with ‘crochet' needles have been developed to facilitate wound closure in very obese patients, where it is difficult to visualize the fascia.

Subcutaneous procedures

The subcutaneous plane has been used recently for vascular and plastic surgical procedures. Vascular and cardiac surgeons may now approach the greater saphenous vein and its branches without a long leg incision or multiple small incisions. With endoscopic techniques, the above-knee greater saphenous vein may be harvested through a single incision by using a long retractor that holds a 5-mm laparoscope, allowing coaxial dissection of the vein and coagulation or clipping of side branches.

Minimally invasive techniques have been used for the treatment of venous disease as well. Incompetent perforating veins in the lower leg may be ligated below the level of the fascia (Linton procedure) using a subcutaneous tunneling instrument *via* a small incision above the knee. Plastic surgeons perform several procedures

using endoscopic instruments in the subcutaneous plane, with obvious cosmetic benefits. In some cases, gas insufflation of these soft-tissue planes can assist blunt dissection.

Extraperitoneal procedures

Balloon dissection has been used to create a working space for inguinal herniorrhaphy, adrenalectomy, nephrectomy, lumbar discectomy, or para-aortic lymphadenectomy. Initial access to the extraperitoneal space is as for direct peritoneal access, except that the peritoneum is not entered. A special trocar is introduced into the space and its balloon is inflated (Fig. 4). The balloon trocar is then replaced with a Hasson trocar, and insufflation to 15 mmHg is delivered to provide a working chamber. Generally, extraperitoneal endosurgery provides less working space than laparoscopy but eliminates the possibility of intestinal injury, adhesions, herniation, and ileus. In extraperitoneal endoscopic repair of hernias, the small bowel is never in contact with the prosthetic mesh or a peritoneal membrane that has been closed.

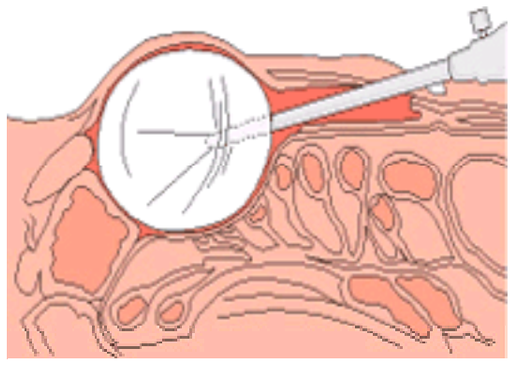


Fig. 4. Balloons are used to create extra anatomic working spaces. In this example, a balloon has been introduced between the posterior rectus sheath and rectus muscle to allow extraperitoneal endoscopic herniorrhaphy.

Instrumentation for endoscopic surgery

The range of complexity in the instrumentation for minimally invasive surgery depends on the degree of visual resolution and the amount of tissue manipulation required. CT-guided drainage of a subphrenic abscess may require no more than a needle, a guide wire, and a plastic drain. Alternatively, laparoscopic exploration of the common bile duct requires dual video imaging and is therefore better performed in an operating room. Discussion of imaging systems, energy sources, and hand instruments follows.

Imaging systems

Those imaging methods conventionally considered ‘radiologic,' including fluoroscopy, CT, ultrasonography, and MRI, are beyond the scope of this chapter. Most endosurgical procedures are guided with a video-endoscopic image. Two established methods of video imaging are commonly used. *Flexible video-endoscopy* depends upon a charged-couple device (**CCD**), placed either on the internal end of a long, flexible endoscope or behind a bundle of thin quartz fibers as a CCD camera mounted on the external end of an endoscope. Most standard gastrointestinal endoscopes now have the CCD chip at the distal end, although small choledochoscopes and nephroscopes are still equipped with fiberoptic bundles. Distally mounted CCD chips have been developed for laparoscopy but are not routinely used.

Imaging for laparoscopy and thoracoscopy employs rigid metal telescopes from 2 to 10 mm in diameter, containing quartz rods of differing optical characteristics (Fig. 5). Since light transmission varies with the square of the cross-sectional area, only limited illumination is possible with small telescopes. In small, white-colored cavities, such as the knee joint, little illumination is necessary. When working in the larger, less reflective confines of the abdominal cavity, the full illumination of a 10-mm laparoscope is generally needed.

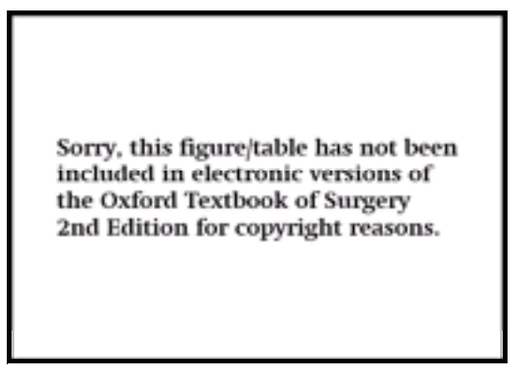


Fig. 5. The Hopkins rod lens telescope includes a series of optical rods that effectively transmits light to the eyepiece; the video camera is placed on the eyepiece to provide the working image.

Rigid telescopes may have a flat or angled distal aspect. A flat end provides a forward view angle (0°), whereas an angled end provides an oblique view (30° or 45°). Rotating an angled telescope allows a wider survey of the operative field through a single trocar site, and lets the operator look ‘around corners.' This advantage is particularly important for visualizing the common bile duct during laparoscopic cholecystectomy or the posterior esophagus during laparoscopic fundoplication. An angled telescope has distinct advantages for most video-endoscopic procedures.

Light is delivered through the endoscope by a fiberoptic light cable. These cables are tremendously inefficient, delivering only 10 per cent of the light provided by the source. Therefore, light sources of 300 W are needed to provide adequate illumination for video-endosurgery.

Video cameras come in two basic designs. The one-chip camera converts gray-scale images to approximate colors by an internal processor. Therefore, perfect color representation is not possible with a one-chip camera. The most accurate color representation is obtained with a three-chip camera, which is technically similar to a color television camera, detects red, green, and blue (**RGB**) input, and provides true color images. RGB imaging provides the highest fidelity for video surgery but is probably not necessary for everyday use. Another advanced option is digital enhancement, which provides a crisper image by emphasizing areas where color or light changes occur between adjacent pixels. Edges are therefore sharpened and image quality is improved. Digital enhancement is available in one- and three-chip cameras. Priorities in an endoscopic video system are illumination first, resolution second, and color third. In the absence of the first two, video surgery is unsafe.

Three-dimensional imaging provides a greater depth of field than with standard endosurgery, and allows more rapid skills' acquisition for endoscopic surgeons learning complex tasks such as suturing. The benefits of three-dimensional systems are less for experienced surgeons. Since these systems require rapid alternation of two slightly offset images, the edges may appear fuzzy and resolution is lost. In our experience, the optical accommodation necessary to rectify these slightly differing images negates any advantage offered by the additional depth of field.

Energy sources for endoscopic surgery

Radiofrequency electrosurgery

The most popular energy source in endoscopic surgery is radiofrequency for electrosurgery. Radiofrequency electrosurgery delivers an alternating current of 500 000 cycles/s (Hz), which does not cause membrane depolarization. Instead, the rapid ‘jiggling' of electrons in the field creates tissue heating that progresses through the phases of coagulation (60°C), vaporization/desiccation (100°C), and carbonization (above 200°C). The radiofrequency can be delivered by monopolar and bipolar

electrodes.

Monopolar electrosurgery requires a remote ground plate on the patient's leg or back to receive the flow of electrons originating at the source. A fine-tip electrode causes a high current density at the site of application and rapid tissue heating. Monopolar electrosurgery is inexpensive and easy to modulate to achieve different effects. A short, high-voltage discharge of current (coagulation current) provides extremely rapid heating. Lower voltage, higher amperage current (cutting current) is more effective for desiccation and vaporization. When the surgeon desires to divide tissue with the least accompanying thermal injury (and least coagulation necrosis), a cutting current is used. With bipolar electrosurgery the electrons flow between two adjacent electrodes, and the tissue between is heated and desiccated. Although there is little opportunity for tissue cutting when a bipolar current is used, the ability to coapt the electrodes across a vessel provides the best method of coagulation for small vessels without thermal injury to adjacent tissues.

Argon-beam coagulation

Another method of delivering radiofrequency electrosurgery is known as argon-beam coagulation. This is a type of monopolar electrosurgery in which a uniform field of electrons is distributed across a tissue surface by a jet of argon gas. The jet distributes electrons more evenly across the surface than does spray electrofulguration alone. This technology has its greatest application for coagulation of diffusely bleeding surfaces, such as the cut edge of liver or spleen. It is less valuable for laparoscopy because the increased intra-abdominal pressures created by the gas jet increase the likelihood of gas embolism.

Lasers

Gas, liquid, and solid-state lasers have been available for medical applications since the mid-1960s. The CO_2 laser (wavelength 10600 nm) is most appropriately used for cutting and superficial ablation of tissues. It is valuable for excisions in remote sites, such as for granulomas of the vocal cords. The CO_2 laser must be delivered with a series of mirrors and therefore is cumbersome to use.

The neodymium (Nd:YAG) laser provides light of 1064 nm wavelength. It is in the near-infrared spectrum and, like the CO_2 laser, is invisible to the naked eye. Nd:YAG laser light is poorly absorbed by most tissue pigments and thus travels deeply into tissue. Deep penetration provides deep heating (Fig. 6), and therefore the Nd:YAG laser is capable of the greatest amount of tissue destruction in a single application, making it ideal for destroying large fungating tumors of the rectosigmoid or tracheobronchial tree. Its disadvantages are the associated risk for perforation of gastrointestinal organs, particularly the cecum.

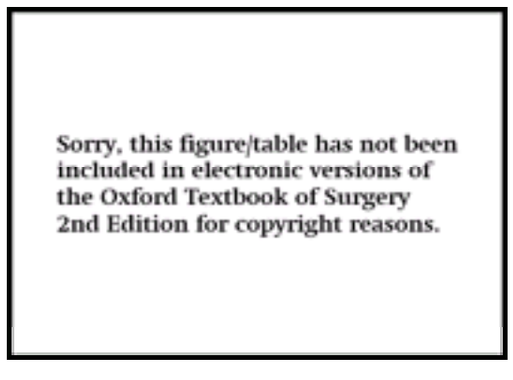


Fig. 6. Plot of light absorption by various tissue compounds (water, melanin, and oxyhemoglobin) as a function of the wavelength; the nadir of the oxyhemoglobin and melanin curves is close to 1064 nm, the wavelength of the Nd:YAG laser.

The frequency-doubled Nd:YAG laser (**KTP** laser) provides 532-nm wavelength light in the green part of the spectrum. Selective absorption by red pigments in tissue, and a depth of penetrance intermediate between that of the CO_2 and Nd:YAG lasers, make this the optimal laser for coagulation (without vaporization) of superficial vascular lesions such as such as hemangiomas and arteriovenous malformations. Argon laser (l; = 632 nm) has properties similar to the KTP laser. Lasers as coagulators have largely been replaced by bipolar electrosurgery, heater probes, and injection. The heater probe is a metal ball that is heated to a temperature (60–100°C) that coagulates bleeding lesions without perforation.

High-energy laser technology provides extremely rapid discharge (less than 10^{-6} s) of large amounts of energy (more than 10^3 V). High-energy lasers, such as the pulsed dye laser, allow the conversion of light energy to mechanical disruptive energy in the form of a shockwave. Such energy may be delivered through a quartz fiber sufficient to fragment kidney stones and gallstones. Shock waves may also be created with miniature electrical sparkplug discharge systems known as electrohydraulic lithotriptors. These waves may be delivered through thin probes for endoscopic application and stone destruction. Lasers have the advantage of pigment selectivity, but electrohydraulic lithotriptors are more popular because they are less expensive and more compact.

Ultrasound

Extracorporeal shockwave lithotripsy creates shock waves that intensify at the focal point of the discharge. When the focal point is directed within the body, these waves are capable of fragmenting kidney stones. Slightly different configurations of this technology may be used to heat tissues internally, and some day may allow *noninvasive* ablation of certain tumors.

Ultrasonic energy has been used to create rapidly oscillating instruments to heat tissue by friction. The harmonic scalpel LCS device (Ultracision, Smithfield, RI) is capable of coagulating and dividing blood vessels with a minimal amount of collateral damage, and is perhaps the most revolutionary new energy technology in laparoscopic surgery. The harmonic scalpel is able to control and divide medium-sized vessels that would otherwise require either bipolar desiccation followed by cutting or clipping and cutting.

Balloons and stents

All branches of endoscopic surgery have need for balloon dilators and stents. Endoluminal balloon dilators may be delivered through an endoscope, or by fluoroscopic guidance. All such dilators have low-compliance balloons, allowing the delivery of high radial pressures as the balloon is inflated, whether for dilation of an atherosclerotic plaque or fibrotic stricture, or disruption of muscular bands (esophageal achalasia). Once a lumen has been dilated, it is often advisable to protect its patency with a stent. Stents are particularly useful for malignant gastrointestinal strictures and certain endovascular procedures (Fig. 7), but are generally not indicated for the long-term management of benign gastrointestinal strictures because tissue ingrowth and inflammatory response may make surgical reconstruction extremely difficult.

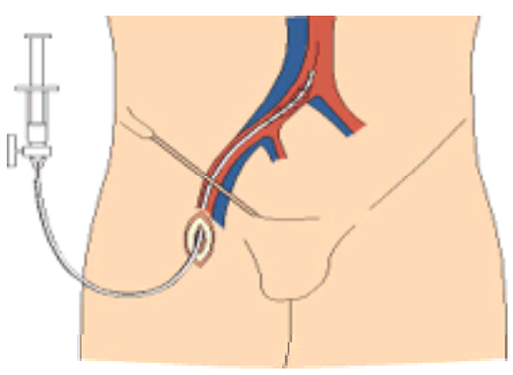


Fig. 7. Common iliac angioplasty is a frequently employed endovascular procedure used to improve inflow to an ischemic lower extremity.

Plastic stents are used to bypass temporarily biliary or urinary obstructions. Metal stents, made of titanium or nitonol, are generally delivered over a balloon and are expanded to the size of the obstructed passage. Metal stents are prone to tissue ingrowth and therefore may reocclude. This is a particular problem in stents for the palliation of malignancies, and may be an obstacle to the long-term use of stents in vascular disease. Attempts at filling the interstices with Silastic or other materials may prevent tumor ingrowth, but may also encourage stent migration. Unique stent designs including hooks and barbs have been proposed in an attempt to minimize stent migration.

Hand instruments

Hand instruments for endoscopic surgery are longer and narrower than conventional surgical instruments. Certain instruments (such as scissors) were easy to reproduce for laparoscopy whereas others (forceps) were virtually impossible. Hence, different configurations of graspers have been developed to replace the various surgical forceps and clamps. Standard endoscopic hand instruments are 5 mm in diameter and 30 cm long, but pediatric and microendoscopic instruments are shorter and narrower. Long shafts (45 cm) are useful in pelvic laparoscopy and obese patients. The only truly unique endoscopic hand instrument developed since the advent of laparoscopic cholecystectomy is the monopolar electrical hook. This device may have suction and irrigation attachments to eliminate smoke and blood from the operative field. The monopolar hook allows tenting of tissue over a bare metal wire with subsequent coagulation and division.

Room set-up for endoscopic surgery

Nearly every endoscopic procedure requires a video monitor. Occasionally, two images are necessary, as for procedures such as ERCP, laparoscopic exploration of the common bile duct, and laparoscopic ultrasonography. Images should be displayed on adjacent video monitors or projected on to a single screen with a 'picture-in-picture' effect to avoid spatial separation. Video monitors should be positioned directly across the patient as the surgeon faces the operative field. The linear association of the endoscope, the field of interest, and the monitor is vital for the surgeon's anatomic and spatial orientation, so that he or she perceives an image between the right and left 'hands' with in-line registration. Therefore, the video monitor should be positioned off the patient's feet for pelviscopy and at the patient's right shoulder for laparoscopic cholecystectomy. It is also valuable to position the insufflation and monitoring equipment across from the surgeon so he or she will recognize variations in insufflating pressure, vital signs, and end-tidal CO_2 tension.

Trocars for the surgeon's left and right hand are ideally placed at least 10 cm apart. This creates an equilateral triangle, with sides of 10 to 12 cm, between the right hand, left hand, and telescope. A diamond is created if the target of the operation (e.g. the gallbladder or gastroesophageal junction) is oriented at the apex of a second equilateral triangle abutting the first (Fig. 8). The surgeon's position at the operating table should allow the elbows to be in at the sides and bent 90°. This orientation provides optimal ergonomics for the operating surgeon, but frequently requires the camera operator (or robotic arm) to reach between the surgeon's hands to guide the telescope. It is usually necessary to alter the position of the operating table, with left or right tilt and Trendelenburg or reverse Trendelenburg depending upon the field of interest.

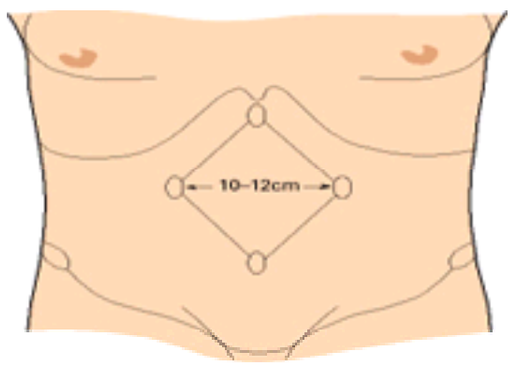


Fig. 8. A diamond is created with the target of the operation, in this case the gastroesophageal junction, at the apex, the laparoscope at the lowest point, and the surgeon's hands 10–12 cm apart so that the elbows are at the sides and arms bent at a 90° angles to provide optimal ergonomics.

Generally, patients are positioned supine for laparoscopic surgery. If the operative field is the gastroesophageal junction or the left lobe of the liver, it is easiest to operate from between the legs, which may be elevated in Allen stirrups or spread apart on leg boards. For pelvic procedures the legs need to be placed in Allen stirrups to achieve access to the perineum. Nephrectomy or adrenalectomy requires a lateral decubitus position with the table flexed. For splenectomy, a 45° tilt of the patient on the table allows easy rotation to provide access to the lesser sac and the lateral peritoneal attachments of the spleen. Thoracoscopy is performed with the patient lateral.

If a patient is to be positioned with legs bent or in the reverse Trendelenburg position, we provide prophylaxis against deep venous thrombosis. Although research findings do not yet compel such prophylaxis in laparoscopy, or favor one regimen over another, it has been shown that sequential compression of the lower extremities during laparoscopy does increase venous return and provide inhibition of thromboplastin activator inhibitor.

Laparoscopic surgical procedures

Many general surgeons spend the majority of their clinical time performing flexible and rigid endoscopic operations. In most *non*surgical, procedure-oriented specialties (e.g. gastroenterology), all therapeutic procedures are endoscopic. In some surgical subspecialties, such as reproductive endocrinology and sports medicine, nearly 100 per cent of interventions are endosurgical. It is beyond the scope of this chapter to discuss the huge range of endoscopic procedures performed by practitioners of all backgrounds. Instead, we will discuss briefly some of the indications and techniques for basic and advanced laparoscopic procedures likely to be performed by general surgeons (Table 1).



Table 1 Laparoscopic surgical procedures

Laparoscopic cholecystectomy

Although flexible endoscopy introduced many general surgeons to video-assisted procedures (e.g. percutaneous endoscopic gastro-stomy), it was laparoscopic cholecystectomy that provided the cornerstone for the recent rapid developments in the field. As outlined earlier, laparoscopic cholecystectomy tested our system of education, reoriented surgical economics, created new political affiliations, and changed the relationship bewteen surgeons and the medical–industrial complex. Laparoscopic cholecystectomy inspired cooperation between competing surgeons to learn the procedure and monitor their results. Surgeons barely out of training were commonly the first to learn the technique, and rapidly assumed a large proportion of the traditional hepatobiliary caseload. Laparoscopic cholecystectomy has

revitalized the previously fragmenting field of general surgery, refocused the clinical and research interests of many established surgeons, and reclaimed many more graduating residents from the clutches of a surgical subspecialty.

One lesson presented by the advent of laparoscopic cholecystectomy concerns the method of training practicing surgeons to perform new, technically sophisticated procedures. Without adequate training and experience, surgeons were ill prepared for the difficulties of laparoscopic surgery, and therefore produced excessive morbidity early in their experience. As future technological developments again reshape surgery, this most important lesson of the laparoscopic revolution must not be forgotten. A system must exist to (1) allow surgeons to acquire the cognitive and psychomotor skills necessary for new procedures in a structured setting, (2) guarantee their possession of these skills upon departure from the trainee setting, and (3) provide proctoring in the surgeon's own hospital environment.

Another valuable lesson provided by the early experience with laparoscopic cholecystectomy was that operative strategies to avoid complications were not intuitive. The common bile duct was injured more frequently than with open cholecystectomy because it was frequently mistaken for the cystic duct. Five steps that minimize the risk of laparoscopic injury to the bile duct include (1) use of an angled (30°) telescope, (2) maximal cephalic retraction of the gallbladder fundus, (3) lateral retraction of the infundibulum, (4) complete dissection of the gallbladder at its junction with the cystic duct, and (5) liberal fluoroscopic cholangiography.

Finally, laparoscopic cholecystectomy has exposed the benefits of endoscopic surgery in general, that it results in little pain, minimal scarring, rapid rehabilitation, and economic advantage. These virtues were understood by laparoscopic gynecologists long before the early 1990s. The procedures listed below represent some of the other endosurgical applications that have entered the general surgeon's practice over the last 10 years.

Laparoscopic appendectomy

Although a ‘one-handed' dissection technique is often adequate for the well-assisted surgeon performing laparoscopic cholecystectomy, laparoscopic appendectomy generally requires proficiency with both hands. It is not necessary to be truly ambidextrous to be a laparoscopic surgeon, but the *non*dominant hand must be trained.

The indications for laparoscopic appendectomy depend on the surgeon's confidence in the diagnosis and the stage of the appendicitis, the patient's characteristics and needs, and the presence of appropriate support personnel. Patients with tenderness in the right iliac fossa will benefit from diagnostic laparoscopy. If uncomplicated appendicitis is found, laparoscopic appendectomy should be performed. Likewise, if a normal appendix and no other pathology are detected, or if the pathology detected is likely to cause recurrent pain (e.g. terminal ileitis), the appendix should be removed. Appendectomy is not performed if another surgical problem (e.g. acute cholecystitis) is apparent. If unsuspected advanced inflammatory conditions of the appendix (gangrene, phlegmon, abscess) are discovered, conversion to laparotomy, or rarely antibiotic therapy and interval appendectomy, are indicated. The perforated appendix may be treated laparoscopically if the inflammatory reaction is minimal and the base of the appendix is normal.

Other consensus indications for a selective approach to laparoscopic appendectomy include obese patients and active patients where rapid rehabilitation is paramount. The advantages of a *non*selective approach to laparoscopic appendectomy have been demonstrated in several prospective, randomized trials. In the majority of these trials, laparoscopic appendectomy resulted in less pain, more rapid recovery, and earlier hospital discharge. Recent meta-analysis of all prospective randomized trials confirmed this finding.

Laparoscopic appendectomy generally requires three ports, one of which is 10 mm in diameter. The surgeon requires one assistant to guide the video-endoscope. Trocars are placed at the umbilicus (10 mm) for the right hand, the suprapubic position (5 mm) for the left hand, and halfway between the umbilicus and pubis (5 or 10 mm) for the telescope (Fig. 9(a)). Working through the umbilical and suprapubic trocars the surgeon mobilizes the appendix, then divides the mesoappendix with clips, bipolar electrosurgery, ligature, or staples. The base of the appendix is controlled with a ligature or an endoscopic linear stapler. When inflammation of the mesoappendix prevents adequate mobilization, it is prudent to create a window in the mesentery to allow stapling across the base of the appendix, and then retrograde dissection of the appendix from the mesoappendix using clips, staples, and bipolar cautery. The appendix should immediately be removed through the 10-mm trocar (if small) or in a specimen bag (if enlarged).

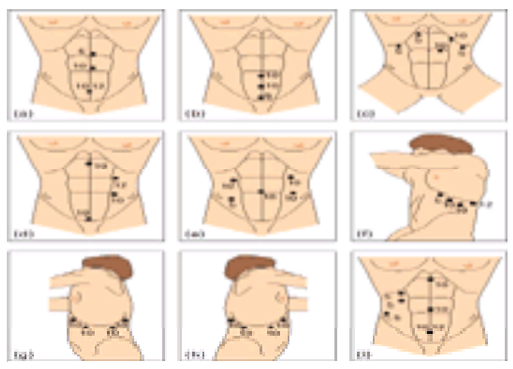


Fig. 9. Trocar positions for advanced laparoscopic procedures:(a) laparoscopic appendectomy; (b) extraperitoneal hernia repair;(c) laparoscopic Nissen fundoplication; (d) right colectomy; (e) laparoscopic low anterior resection; (f) laparoscopic splenectomy (‘hybrid' approach);(g) left adrenalectomy; (h) right adrenalectomy; (i) laparoscopic common bile-duct exploration.

Laparoscopic repair of inguinal hernia

Laparoscopic repair of inguinal hernia is the most controversial of all new endosurgical procedures since it does not shorten the time spent in hospital, its comparative efficacy will take 20 years to evaluate, the technique is difficult to learn, and it costs more than open herniorraphy. The early types of laparoscopic hernial repair (plication of the internal ring; plug of the hernia sac; prosthetic ‘onlay') have been abandoned. Two repairs are commonly performed, the *transabdominal preperitoneal* repair and the *totally extraperitoneal* repair. Although slightly more difficult, many experienced laparoscopic surgeons prefer *totally extraperitoneal repair* except for patients with incarcerated hernias or a history of previous lower-abdominal operations.

The best indication for *totally extraperitoneal repair* is recurrent hernia. Anterior repair of recurrent hernias is associated with a 20 per cent rate of re-recurrence, frequent injury to the ilioinguinal nerve, and possible injury to the testicular blood supply. Before the laparoscopic era, many surgeons advocated open preperitoneal repair for such patients. The *totally extraperitoneal* repair reproduces the open repair, except for the need for a large counter-incision to admit the mesh to the appropriate plane. Preliminary reports suggest that the *totally extraperitoneal* repair will achieve a re-recurrence rate similar to that of the open procedures (less than 5 per cent).

Bilateral hernias are another indication for *totally extraperitoneal repair*. Contralateral dissection adds little time to the procedure and requires no additional incisions. The ‘tension-free' nature of prosthetic hernial repair eliminates the higher recurrence rate associated with simultaneous open repair of bilateral hernias without mesh. The majority of the expense of laparoscopic herniorrhaphy relates to charges for equipment and the time required to gain access. Repair of the contralateral side adds little time or cost of equipment. In fact, given the ease of dissection and repair of the contralateral groin, some surgeons advocate routine bilateral dissection and simultaneous repair if asymptomatic hernias are detected.

The third indication for laparoscopic hernial repair is patient preference. Although many surgeons are skeptical about the value of laparoscopic herniorrhaphy, patients who have undergone conventional repair often request a surgeon with laparoscopic experience. Among patients who have experienced both laparoscopic and open repairs, randomized data support testimonials that laparoscopic herniorraphy causes less discomfort and allows more rapid rehabilitation.

The preferred approach for *totally extraperitoneal* repair requires three trocars in the lower midline (Fig. 9(b)). A small infraumbilic incision is opened through either rectus muscle to the posterior rectus sheath. A large balloon dissector is passed to the pubic symphysis and inflated at this level. The balloon does most of the posterior dissection of Hesselbach's triangle. The balloon is removed and a Hasson cannula and 45° telescope are placed. The spermatic cord is then dissected from any indirect hernial sac. Small sacs may be reduced entirely to remain behind the mesh in the extraperitoneal space. If a large sac is present in the inguinal canal, enough of it must be mobilized to ensure that all intraperitoneal structures are reduced, and a 3–0 *non*absorbable ligature is placed and tied extracorporeally. The sac is then opened widely distal to the knot and left in the inguinal canal. A large piece of mesh, 10 × 15 cm at a minimum, is placed to cover Hesselbach's triangle, the internal ring, and the femoral canal. The mesh is fixed with a fixation device to Cooper's ligament, the pubic tubercle, the posterior rectus muscle, and the transversalis fascia lateral to the epigastric vessels. Palpation of the fixation device through the abdominal wall will minimize the possibility of entrapping the genitofemoral or lateral

femoral cutaneous nerve within a staple.

Laparoscopic fundoplication

Nissen fundoplication was an operation on the way out during the late 1980s, as medical therapy of this condition improved. However, a landmark study from the Veteran's Cooperative Group found that antireflux surgery eliminated gastroesophageal reflux more effectively than medical therapy in patients with severe reflux. These data came to light during the early development of laparoscopic fundoplication. In the last 5 years, outcomes after laparoscopic antireflux surgery have been only marginally different from those of open fundoplication. Like laparoscopic inguinal herniorrhaphy, the ultimate utility of this procedure will require follow-up of 10 to 20 years.

Patient selection and operative technique are the only determinants of the success of laparoscopic Nissen fundoplication. Operation is indicated for severe typical or atypical symptoms of gastroesophageal reflux that (1) require daily proton-pump inhibitors, (2) are associated with stricture and/or Barrett's metaplasia, or (3) are associated with respiratory complications (aspiration, tracheolaryngeal damage). It is important to recognize patients likely to present difficulty and be prepared for those cases. Such patients include those with prior subdiaphragmatic surgery, the obese patient, those with large hiatal hernias (more than 5 cm), and those with significant esophageal shortening (usually in conjunction with Barrett's esophagus and stricture).

Preoperatively, the diagnosis of gastroesophageal reflux is confirmed by detailed anatomic and physiologic evaluation of the foregut. A barium swallow and esophagogastroduodenoscopy are used to demonstrate structural abnormalities associated with the gastroesophageal reflux, such as stricture and hiatal hernia, and to allow mucosal survey for esophagitis, unrecognized malignancy, or gastric pathology. If the diagnosis of gastroesophageal reflux is unclear, 24-h pH monitoring is used. An esophageal motility study characterizes the function of the lower esophageal sphincter and permits the fundoplication to be tailored (partial fundoplication) if a motility disorder is detected. A gastric emptying study is performed only if gastric abnormalities are suspected (diabetes, vomiting, peptic ulcer disease). Abnormalities in gastric emptying may be treated with prokinetic agents or concomitant pyloroplasty at the time of fundoplication.

The operative technique of laparoscopic Nissen fundoplication is well described elsewhere. Several important principles deserve emphasis. Before the first incision, organization of the field of dissection by careful determination of the trocar sites (Fig. 9(c)) and adjustment of the table height are vital to facilitate a successful operation. Intracorporeal rather than extracorporeal knot tying reduces tissue trauma. Postoperative dysphagia is minimized by prescribing a soft diet for 3 weeks. Esophageal dilation may be of value in the first 16 weeks postoperatively for patients with swallowing problems, but is effective only if the fundoplication has not become disrupted.

Laparoscope-assisted colectomy

The appropriate role for laparoscopic colectomy has engendered heated debate, especially its use in colon cancer. At the core of this discussion is the assertion that metastases in the abdominal wall are twice as likely at laparoscopic trocar sites than open colectomy incisions. The current consensus on indications for laparoscopic-assisted colectomy are that it may be appropriate for benign or premalignant disease, for creating a colostomy other than in patients with cancer, and in rectopexy for dysfunction of the pelvic floor. A further creative application of laparoscopy is for the mobilization of the splenic or hepatic flexures during procedures performed primarily in the pelvis, allowing the benefits of low transverse incisions.

Laparoscopic colectomy has unique obstacles. Since the field of dissection often traverses different quadrants of the abdomen, it is frequently necessary to modify the orientation of the instruments during the procedure. Logical trocar placements (based on triangulation) simplify the approach (Fig. 9(e),Fig. 9(e)), but the triangulation techniques ideal for cecal mobilization will not work for the mobilization of hepatic flexures. Using further trocars and changing the telescope ports are two necessary strategies for long-segment colonic resection. Also, a counter-incision is needed to extract the specimen or to create a stoma, which increases pain and ileus, and extends the length of time in hospital for some patients.

Laparoscopic splenectomy

Since elective splenectomy is an infrequent operation, and since not all spleens are amenable to laparoscopic resection, laparoscopic splenectomy has emerged more slowly than laparoscopic fundoplication or colectomy. The most common indication is immunologic disease, particularly idiopathic thrombocytopenic purpura, with a normal or slightly enlarged spleen. Spleens weighing more than 500 g are too difficult to place in a collection bag and their removal requires a generous counter-incision.

Laparoscopic splenectomy was initially described with the patient either supine or in the left lateral decubitus (hanging spleen) position. The ‘leaning spleen' position is a more recently described hybrid, with the patient positioned at 45° on the table. With this approach, the table is tilted so that the patient is supine for the placement of the pneumoperitoneum and trocars (Fig. 9(f)). If necessary, the splenic flexure of the colon is mobilized before placing the last trocar. The splenic artery is identified and ligated in the lesser sac, and platelets are administered as necessary. Then the table is tilted in the other direction to bring the patient into the right lateral decubitus position, allowing the spleen to fall anteriorly. The splenophrenic attachments (peritoneal reflection) are divided to give access to the splenic hilum. Splenic vessels are controlled with a linear cutting stapler or individually ligated. The spleen is most easily manipulated into a strong, nylon retrieval bag if the bag is placed in the left upper quadrant and its mouth is opened to face the laparoscope. The neck of the bag is pulled out at a trocar site and the spleen is morcellated and removed with ring forceps or a mechanized morcellator.

Laparoscopic adrenalectomy

The laparoscopic approach to the adrenal gland has been very successful. Small benign tumors are the most frequent indication. Hyperfunctioning adenomas (including pheochromocytoma) may likewise be laparoscopically resected. If adrenal carcinoma is suspected (lesion diameter more than 5 cm, rapidly enlarging), or if a search for paragangliomas is indicated, a standard open transabdominal approach may be warranted. While a great fear existed about the hemodynamic effects of laparoscopic dissection of pheochromocytomas, the ‘no touch' technique afforded by laparoscopic surgery is ideal for unilateral tumors.

The preferred technique for laparoscopic adrenalectomy places the patient in a lateral decubitus position with table flexion between the ribs and iliac crest. Four trocars are placed beneath the costal margin (Fig. 9(g), h)). For left-sided dissection, the spleen is rolled anteriorly by dividing the splenophrenic ligament, often exposing the yellow–orange adrenal gland. Sometimes, especially in obese patients, endoscopic ultrasound may facilitate localization of the gland. A monopolar electrocautery hook, blunt dissection, and clips are appropriate for dissection around the gland capsule. Dissection with the harmonic scalpel may reduce blood loss in some highly vascular adrenal lesions. Care should be taken not to grasp the gland as the retroperitoneal fat is teased away. The adrenal vein is identified inferomedially, doubly clipped on the stay side, and divided. The specimen is placed immediately into a bag and extracted through one of the trocar sites. The right adrenal dissection is similar except that a liver retractor is needed to hold the right lobe anteriorly. Starting caudally, dissection of the gland away from the vena cava allows identification and safe division of the short right adrenal vein.

Laparoscopic exploration of the common bile duct

At the inception of laparoscopic cholecystectomy, the detection of bile-duct stones during operative cholangiography (routine or selective) usually led to open exploration of the common bile duct. While laparoscopic techniques to explore the common bile duct through the cystic duct are quite effective, laparoscopic exploration of the common bile duct has not been widely implemented. The technical demands of this advanced procedure, and the political and economic value of allowing a colleague to perform ERCP and sphincterotomy, have thwarted the rapid growth of this procedure amongst most surgeons.

Data from several centers show that laparoscopic cholecystectomy, selective cholangiography, and retrieval of bile-duct stones costs less than laparoscopic cholecystectomy and endoscopic sphincterotomy. Endoscopic sphincterotomy and stone retrieval adds $1,000–$2,000 to the cost of therapy, compared to $200–$1000 for laparoscopic exploration of the common bile duct. In an economic environment of capitation payments, incentive always favors the less expensive procedure. Simultaneous intraoperative management of cholelithiasis and choledocholithiasis benefits the patient and is economically prudent.

Laparoscopic exploration of the bile duct begins with fluoroscopic cholangiography. A cholangiogram is performed with a ureteral or modified ERCP catheter to allow the transcystic passage of a guide wire. If stones are detected, a hydrophilic guide wire is passed distally into the duodenum. If the common bile duct and the stones are small (less than 3 mm), the cholangiographic catheter may be advanced to a position just above the stones. Rapid infusion of saline will typically expel them into the duodenum. Intravenous glucagon will help relieve obstruction at the sphincter of Oddi. Stones larger than 3 mm, or those associated with dilatation of the common bile duct, are more easily extracted with a basket. As stone diameters approach 5 to 8 mm, dilation of the cystic duct is appropriate, using an 8-mm balloon passed over a guide wire. The balloon is kept in place for 5 min while the choledochoscope is prepared and an extra 5-mm trocaris placed (Fig. 9(i)). A thin-caliber choledochoscope (8–10 French) is passed over the guide wire and into the distal common bile duct. A 2- to 2.5-mm flat wire basket is passed through the operating channel of the choledochoscope and the stone is entrapped under visual guidance. With practice, this maneuver is easily mastered. For stones lager than 8 mm, choledochotomy, stone extraction, and placement of a T-tube are obligatory (Fig. 10). A number of surgeons use choledochotomy as their primary laparoscopic

approach to stones of the common bile duct. Most bile-duct stones (80–90 per cent) can be extracted successfully by one of these techniques. Laparoscopic exploration of the common bile duct is as effective as endoscopic sphincterotomy, and may reduce the risk of post-procedure pancreatitis and the potential long-term sequelae of sphincter destruction.

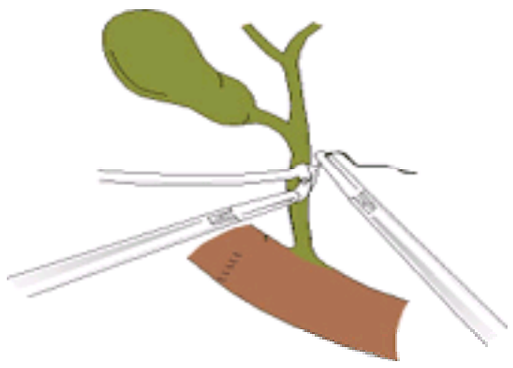


Fig. 10. Laparoscopic choledochotomy and common bile-duct exploration differs from the open procedure in two ways: (1) the gallbladder is left *in situ* to provide a point of traction on the bile duct, and (2) stay sutures are generally not necessary.

Special considerations in minimally invasive surgery

Pediatric endoscopic surgery

Endoscopic surgery in the adolescent is similar to that in the adult, and standard instrumentation and trocar positions usually suffice. However, for the infant or young child a reconsideration of surgical strategies and the use of a specialized instrumentation are required. The instruments are generally shorter (15–20 cm) and narrower (3 mm). Also, since the peritoneal space is smaller in the child than the adult, a 5-mm telescope will provide sufficient illumination for almost any pediatric endosurgical procedure. The development of 5-mm clip appliers and bipolar devices has obviated the need for 10-mm trocars in pediatric laparoscopy. Because the abdominal wall is thin in infants, the pneumoperitoneum pressure may be maintained at 8 mmHg without compromising visualization. Since venous thrombosis is extremely rare in children, prophylaxis for deepvenous thrombosis is probably unnecessary.

Endoscopic surgery in pregnancy

Experimental and clinical data have answered most concerns over the safety of laparoscopic cholecystectomy or appendectomy during pregnancy. Studies on fetal sheep have proved that maternal and fetal pH are linearly related, so fetal acidosis may be avoided by maintaining respiratory alkalosis in the mother. The increased intra-abdominal pressure caused by pneumoperitoneum is much less than the physiologic pressures tolerated by the fetus during mid-pregnancy uterine contractions. A large clinical experience of well over 100 laparoscopic cholecystectomies in pregnancy has produced acceptable results. Nevertheless, all elective operations should be delayed until after delivery. If delay puts the mother at undue risk, operations during the second trimester are the next best option. During all procedures on pregnant patients, the fetus should be carefully shielded from intraoperative X-rays and monitored via transvaginal ultrasound. Initial access to the abdomen in the pregnant woman should be well above the uterine fundus, which reaches the umbilicus at 20 weeks' gestation. Most endoscopic surgeons favor the open (Hasson) approach over direct-puncture laparoscopy to avoid the chance of uterine injury with the Veress needle.

Endoscopic surgery and cancer

P>Endoscopic cancer surgery is the most rigorous test of a surgeon's endosurgical technique, whether for staging, cure, or palliation. Minimally invasive procedures have assisted cancer staging since early in the endoscopic era. Mediastinoscopy preceding thoracotomy is still valuable to assess mediastinal lymph nodes. Laparoscopy is a simple means to assess the liver before attempting curative pancreatic, gastric, or hepatic resections. If cancer is encountered outside the field of a curative resection, chemotherapy, radiation, or less invasive palliative measures are typically indicated. Sometimes, surgical palliation (e.g. gastrojejunostomy) may be performed by an endoscopic approach, if diagnostic laparoscopy indicates no opportunity for curative resection.

The greatest controversy surrounding the role of endoscopic surgery in cancer therapy concerns its potential to serve as a curative approach. Endoscopic colectomy, gastrectomy, pancreatectomy, hepatectomy, esophagectomy, and pneumonectomy are technically possible in malignancy. The unanswered question remains whether endosurgical techniques provide equivalent overall and disease-free survival to those of conventional techniques. It has been proved that the laparoscopic approach to colectomy and gastrectomy permits lymphadenectomy and resection margins equivalent to those of the standard approach. Early reports of trocar-site implantation fueled concerns that endoscopic dissection caused excessive tumor manipulation and dissemination of cancer cells. Prospective randomized trials are under way to determine the appropriate role for endoscopic surgery in cancer. It is appropriate to defer endoscopic cancer resections until these data are available.

Endoscopic surgery in elderly and infirm individuals

Laparoscopic cholecystectomy provides an alternative approach for debilitated patients with symptomatic cholelithiasis. However, its minimally invasive nature does not imply minimal physiologic impact. These individuals require great care in anesthesia, with Swan–Ganz and arterial catheters to monitor the physiologic derangements created by the positive-pressure pneumoperitoneum. In some cases, intraoperative management may be complicated by the laparoscopic access. Much of the advantage of minimally invasive surgery comes to bear after the operation, since a great deal of the morbidity incurred by debilitated patients derives from postoperative pain and impaired postoperative mobilization. Pulmonary complications, urinary-tract sepsis, deep venous thrombosis, pulmonary embolism, congestive heart failure, and myocardial infarction are more likely if patients are unable to walk and reclaim perioperative fluid from the interstitium. By allowing rapid and early mobilization, laparoscopic surgery has lessened the morbidity associated with procedures in elderly and infirm patients.

Patients in septic shock and those with severe cardiopulmonary compromise may benefit from procedures such as percutaneous cholecystostomy and endoscopic papillotomy. Such options may be safer than a laparoscopic operation, especially if there is a high likelihood for conversion to open surgery.

Economics of endoscopic surgery

Endoscopic procedures that are more expensive than the corresponding open procedures must provide substantial evidence of clinical superiority. Hence, there has been a dedicated effort to minimize the expenses of common endoscopic procedures. Most minimally invasive procedures reduce the cost of therapy by shortening the time spent in hospital. Laparoscopic cholecystectomy, fundoplication, splenectomy, and adrenalectomy are examples. Open procedures already performed on outpatients, such as inguinal herniorrhaphy, are unlikely to be cheaper if done endosurgically. Only when the conventional operation requires a 4- to 7-day inpatient stay, such as for colectomy or fundoplication, will the endoscopic approach have a possible cost advantage. Through the responsible use of disposable instruments, and a commitment to the efficient delivery of inpatient services, most endoscopic procedures may be made less expensive than their conventional counterparts.

Outcomes of minimally invasive surgery

Whereas objective outcome measures are available after cancer surgery (time of survival etc.), outcomes after endoscopic surgery for *non*malignant conditions are more frequently qualitative in nature. Alleviation of the symptoms for which surgery was performed (biliary colic or heartburn) is the most direct measure of their utility. Alternatively, the determination of an improvement in *quality of life* may be a more sensitive estimate of the therapeutic benefit of a given procedure. Chronic disease states such as paralysis, arthritis, and blindness have an obvious impact on quality of life. Less obvious, but equally significant, is the impairment of quality of life found in patients with gastroesophageal reflux, inflammatory bowel disease, and other chronic gastrointestinal conditions. Quality-of-life analysis is a spectacular new research tool that has been validated and will guide the acceptance and implementation of future interventions. Comparison of traditional and new therapies will be facilitated. Operative indications for diseases with qualitative symptoms, such as heartburn, will be more easily standardized on a quality-of-life scale.

Legal and ethical issues

The foundation of the patient–surgeon relationship is trust and informed consent. Informed consent obliges a surgeon offering a *non*traditional intervention to disclose

his or her experience and present conventional alternatives. Since experience with endoscopic surgery is shorter than with open surgery, the risk profile and durability of many procedures remain uncertain. Surgeons must provide an honest dialogue about these uncertainties in addition to extolling proven short-term benefits. Finally, all surgeons should keep accurate records and measure long-term outcomes, especially for disease states where other acceptable therapies exist, such as inguinal hernia and gastroesophageal reflux disease.

Procedures perceived by the public as ‘experimental' are at greater risk for litigation in the event of complications. New approaches to traditional operations may present different risks. Injuries to the common bile duct during laparoscopic cholecystectomy continue to occur more frequently than during open cholecystectomy. Surgeons will most effectively protect themselves from legal action by honestly representing their experience, and enlisting the patient's input in selecting the best procedure from a list of options. It must be emphasized to the patient that, even though pain is less and recovery faster, endoscopic surgery does not necessarily reduce the likelihood of intraoperative complications.

Video-endosurgery offers the unique opportunity to videotape every procedure. Initially, many surgeons were using videotape as a marketing tool to promote laparoscopic cholecystectomy. However, when the tapes began to capture injuries to the common bile duct, the surgical community was forced to reconsider this practice. In many regions, videotape is considered part of the medical record and is therefore available for courtroom examination. When available, videotape is more frequently damaging to a surgeon's defense. If the identity of the patient is not affixed to the tape, it may be used for quality control or educational use without personal risk. High-quality videotapes have exceptional educational value, and assist in the rapid, safe dissemination of new endosurgical procedures.

Future directions

Endoscopic surgery represents the beginning of the evolution from open cavitary surgery to totally *non*invasive surgery. Although all the necessary scientific breakthroughs have not yet been made, several existing technologies may be the precursors to totally *non*invasive surgery. Extracorporeal shockwave lithotripsy is a completely *non*invasive intervention, requiring no anesthetic or admission to hospital. However, biliary lithotripsy was ineffective because of biliary colic during the passage of stones, stone recurrence (50 per cent), and failure to fragment all stones. Endoscopic procedures are effective when they achieve the objectives of analogous open cavitary operations. Minimally invasive surgery will withstand the scrutiny of time only if surgeons remain grounded in time-tested principles of surgery during the ascent of this discipline.

Further reading

History

Cuschieri A. ‘A rose by any other name…' Minimal access or minimally invasive surgery. [Editorial] *Surgical Endoscopy* 1992; **6**: 214.

Hopkins HH. Optical principles of the endoscope. In: Berci G, ed. *Endoscopy* pp. 3–26. Appleton-Century-Crofts, New York, 1976.

Katzir A. Optical fibers in medicine. *Scientific American* 1989; **260**(5): 120–5.

Veritas TF. Coelioscopy—a synthesis of Georg Kelling's work with insufflation, endoscopy and lufttamponade. In: Litynski GS, ed. *Highlights in the history of laparoscopy*, pp. 3–21. Barbara Bernert, Frankfurt/Main, 1996.

Wickham J, Fitzpatrick JM. Minimally invasive surgery. [Editorial] *British Journal of Surgery* 1990; **77**: 721.

Indications, training, organization

Buess G. The endoscopic operating room. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 3–8. Churchill Livingstone, New York/London, 1996.

Cuschieri A. Whither minimal access surgery: tribulations and expectations. *American Journal of Surgery* 1993; **165**: 533–5.w

Stiegmann GV. Advanced laparoscopic surgery. *American Journal of Surgery* 1997; **173**: 19–20.

Physiology, immunology, anesthesia

Callery MP, Soper NJ. Physiology of the pneumoperitoneum. In: Hunter JG, ed. *Baillière's clinical gastroenterology: laparoscopic surgery*, pp. 757–77. Baillière Tindall, London/Philadelphia, 1993.

Cullen DJ, Eger EI. Cardiovascular effects of carbon dioxide in man. *Anesthesiology* 1974; **41**: 345.

Cunningham AJ, Turner J, Rosenbaum S, Rafferty T. Transoesophageal echocardiographic assessment of haemodynamic function during laparoscopic cholecystectomy. *British Journal of Anaesthesia* 1993; **70**: 621.

Gossot D. Access modalities for thoracoscopic surgery. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 743–51. Churchill Livingstone, New York/London, 1996.

Ho HS, Wolfe BM. The physiology and immunology of endosurgery. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 163–71. Churchill Livingstone, New York/London, 1996.

Ho HS, Gunther RA, Wolfe BM. Intraperitoneal carbon dioxide insufflation and cardiopulmonary functions. *Archives of Surgery* 1992; **127**: 928.

Hunter JG. Laparoscopic pneumoperitoneum: the abdominal compartment syndrome revisited. [Editorial] *Journal of the American College of Surgeons* 1995; **181**(5): 469–70.

Hunter JG, Staheli J, Oddsdottir M, Trus T. Nitrous oxide pneumoperitoneum revisited: is there a risk of combustion? *Surgical Endoscopy* 1995; **9**: 501.

Schauer PR, Luna J, Ghiatas AA, Glen ME, Warren JM, Sirinek KR. Pulmonary function after laparoscopic cholecystectomy. *Surgery* 1993; **114**: 389.

Smith RS *et al.* Gasless laparoscopy and conventional instruments: the next phase of minimally invasive surgery. *Archives of Surgery* 1993; **128**: 1102.

Wittgen CM, Andrus CH, Fitzgerald SD, Baudendistel LJ, Dahms TE, Kaminski DL. Analysis of the hemodynamic and ventilatory effects of laparoscopic cholecystectomy. *Archives of Surgery* 1991; **126**: 997.

Access

Byron JW, Parkinson G, Miyazawa K. A randomised comparison of Verreìs needle and direct insertion for laparoscopy. *Surgery, Gynecology and Obstetrics* 1993; **177**: 259–62.

Eaves FF. Basics of endoscopic plastic surgery. In: Bostwick, Eaves FF, Nahai, ed. *Endoscopic plastic surgery*, pp. 59–81. Quality Medical Publishing, St. Louis, 1995.

Fletcher DR. Laparoscopic access. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 189–96. Churchill-Livingstone, New York/London, 1996.

Hanney RM, Alle KM, Cregan PC. Major vascular injury and laparoscopy. *Australia and New Zealand Journal of Surgery* 1995; **65**: 533–5.

Himpens J. Laparoscopic preperitoneal approach to the inguinal hernia. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 949–59. Churchill Livingstone, New York/London, 1996.

Kaiser LR. Video-assisted thoracic surgery: current state of the art. *Annals of Surgery* 1994; **220**: 720–34.

Lumsden AB, Eaves FF. Vein harvest. In: Bostwick, Eaves FF, Nahai, ed. *Endoscopic plastic surgery*, pp. 535–47. Quality Medical Publishing, St. Louis, 1995.

Instrumentation/energy sources

Amaral JF, Chrostek C. Comparison of the ultrasonically activated scalpel to electrosurgery and laser for laparoscopic surgery. *Surgical Endoscopy* 1993; **7**: 141–3.

Foutch P, Sivak M. Therapeutic endoscopic balloon dilatation of the extrahepatic biliary ducts. *American Journal of Gastroenterology* 1985; **80**: 575–80.

Hoepffner N, Foerster EC, Hogemann B, Domschke W. Long term experience in wall stent therapy for malignant choledochal stenosis. *Endoscopy* 1994; **26**: 597–602.

Hunter JG. Controversies in laparoscopic cholecystectomy exposure, dissection, and laser versus electrosurgery. *American Journal of Surgery* 1993; **165**: 492–6.

Kozarek RA, Ball TJ, Patterson DJ. Metallic self-expanding stent application in the upper gastrointestinal tract: caveats and concerns. *Gastrointestinal Endoscopy* 1992; **38**: 1.

Lirici MM, Melzer A, Tadeo FS, Buess G, Amgelini L. Tissue approximation in minimally invasive surgery. *Endoscopic Surgery* 1994; **2**: 47–54.

McIntyre RC, Stiegmann GV. Intraoperative, endoscopic and laparoscopic ultrasound. In: Hunter JG, Sackier JM, ed. *Minimally invasive surgery*, pp. 15–21. McGraw-Hill, New York, 1993.

Margulies DR, Shabot MM. Fiberoptic imaging and measurement. In: Hunter JG, Sackier JM, ed. *Minimally invasive surgery*, pp. 7–14. McGraw-Hill, New York, 1993.

Miles WFA, Paterson-Brown S, Garden OJ. Laparoscopic contact hepatic ultrasonography. *British Journal of Surgery*, 1992; **79**: 419–20.

Odell RC. Laparoscopic electrosurgery. In: Hunter JG, Sackier JM, ed. *Minimally invasive surgery*, pp. 33–41. McGraw-Hill, New York, 1993.

Prescher T. Video imaging. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 41–54. Churchill Livingstone, New York/London, 1996.

Trus TL, Hunter JG. Principles of laser physics and tissue interaction. In: Toouli J, Gossot D, Hunter JG. ed. *Endosurgery*, pp. 103–9. Churchill Livingstone, New York/London, 1996.

Voyels CR *et al.* Education and engineering solutions for potential problems with laparoscopic monopolar electrosurgery. *American Journal of Surgery* 1992; **164**: 57–62.

Laparoscopic cholecystectomy

Asbun HJ, Rossi RL, Lowell A, Munson JL. Bile duct injury during laparoscopic cholecystectomy: mechanism of injury, prevention and management. *World Journal of Surgery* 1993; **17**: 547–51.

Deziel DJ, Millikan KW, Economou SG, Doolas A, Ko ST, Airan MC. Complications of laparoscopic cholecystectomy: a national survey of 4,292 hospitals and an analysis of 77,604 cases. *American Journal of Surgery* 1993; **165**: 9–14.

Hunter JG. Laparoscopic antireflux surgery. *Current Surgery* 1994; **51**(3): 182–4. [See review and summary.]

NIH Consensus Conference. Gallstones and laparoscopic cholecystectomy. *Journal of the American Medical Association* 1993; **269**: 1018–24.

Laparoscopic appendectomy

Bonanni F, Reed J, Hartzell G, Trostle D, Boorge R, Gittleman M, Cole A. Laparoscopic versus conventional appendectomy. *Journal of the American College of Surgeons* 1994; **179**: 273–8.

Frazee RC *et al.* A prospective randomized trial comparing open versus laparoscopic appendectomy. *Annals of Surgery* 1994; **214**: 725.

Kum CK, Ngoi SS, Goh PM, Tekant Y, Issac JR. Randomized controlled trial comparing laparoscopic and open appendectomy. *British Journal of Surgery* 1993; **80**: 1599.

Ortega AE, Hunter JG, Peters JH, Swanstrom LL, Schirmer B. Laparoscopic Appendectomy Study Group. A prospective randomized comparison of laparoscopic appendectomy with open appendectomy. *American Journal of Surgery* 1995; **169**: 208.

Semm K. Endoscopic appendectomy. *Endoscopy* 1993; **15**: 59.

Laparoscopic inguinal hernia repair

Fitzgibbons RJ Jr *et al.* Laparoscopic inguinal hernia: results of a multicenter trial. *Annals of Surgery* 1995; **21**: 3–13.

Hoffman HC, Vinton-Traverso AL. Preperitoneal prosthetic herniorrhaphy: one surgeon's successful technique. *Archives of Surgery* 1993; **128**: 965–70.

Phillips EH, Carroll BJ, Fallas MJ. Laparoscopic preperitoneal inguinal hernia repair without peritoneal incision. Technique and early clinical results. *Surgical Endoscopy* 1993; **7**: 159–62.

Stoker DL, Spiegelhalter DJ, Singh R, Wellwood JM. Laparoscopic versus open inguinal hernia repair: randomized prospective trial. *Lancet* 1994; **343**: 1243–5.

Laparoscopic fundoplication

Hinder RA, Filipi CJ, Wetscher G, Neary P, DeMeester TR, Perdikis G. Laparoscopic Nissen fundoplication is an effective treatment for gastroesophageal reflux disease. *Annals of Surgery* 1994; **220**: 472–83.

Hunter JG, Trus TL, Branum GD, Waring JP, Wood WC. A physiologic approach to laparoscopic fundoplication for gastroesophageal reflux disease. *Annals of Surgery* 1996; **223**: 673–87.

Hunter JG, Swanstrom L, Waring JP. Patterns of dysphagia following laparoscopic antireflux surgery. *Annals of Surgery* 1996; **224**(1): 51–7.

Jamieson GG, Baigrie RJ. Indications and evaluation of patients undergoing laparoscopic antireflux surgery. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 301–4. Churchill Livingstone, New York/London, 1996.

Spechler SJ. The Department of Veterans Affairs Gastroesophageal Reflux Disease Study Group. Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. *New England Journal of Medicine* 1992; **326**: 786–92.

Waring JP, Hunter JG, Oddsdottir M, Wo J, Katz E. The preoperative evaluation of patients considered for laparoscopic antireflux surgery. *American Journal of Gastroenterology* 1995; **90**: 35–8.

Laparoscope-assisted colectomy

Böhm B, Milsom JW, Fazio VW. Postoperative intestinal motility following conventional and laparoscopic intestinal surgery. *Archives of Surgery* 1995; **130**: 415–19.

Cirocco W, Schwartzman A, Golub RW. Abdominal wall recurrence after laparoscopic colectomy for colon cancer. *Surgery* 1994; **116**: 842–6.

Fowler DL. Laparoscopic right hemicolectomy. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 665–73. Churchill Livingstone, New York/London, 1996.

Franklin ME Jr. Laparoscopic low anterior abdominal perineal resections. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 685–94. Churchill Livingstone, New York/London, 1996.

Reilly WT, Nelson H, Schroeder G, Wieand HS, Bolton J, O'Connell MJ. Wound recurrence following conventional treatment of colorectal cancer. *Diseases of the Colon and Rectum* 1996; **39**: 200–7.

Whelan RL, Bessler M, Treat MR. Laparoscopic-assisted sigmoid colectomy. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 675–83. Churchill Livingstone, New York/London, 1996.

Laparoscopic splenectomy/adrenalectomy

Böhm B, Milsom JW, Fazio VW. Laparoscopic splenectomy: the initial experience at University of California, San Francisco. *Archives of Surgery* 1995; **130**: 415–19.

Cadiere GB, Verroken R, Himpens J, Bruyns J, Efira M, De Wit S. Operative strategy in laparoscopic splenectomy *Journal of the American College of Surgeons* 1994; **179**: 668.

Delaitre B. Laparoscopic splenectomy: the ‘hanged spleen' technique. *Surgical Endoscopy* 1995; **9**: 528–9.

Gagner M. Laparoscopic adrenalectomy. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*. pp. 623–34. Churchill Livingstone, New York/London, 1996.

Poulin EC, Thibault C, Mamazza J. Laparoscopic splenectomy. *Surgical Endoscopy* 1995; **9**: 172.

Prinz RA. A comparison of laparoscopic and open adrenalectomies. *Archives of Surgery* 1995; **130**: 489–92.

Vaughan ED Jr. Adrenal surgery. In: Marshall FF, ed. *Atlas of urologic surgery*. Saunders, Philadelphia, 1991.

Laparoscopic common bile-duct exploration

Carroll BJ, Fallas MJ, Phillips EH. Laparoscopic transcystic choledochoscopy. *Surgical Endoscopy* 1994; **8**: 310–14.

Hunter JG. Laparoscopic transcystic common bile duct exploration. *American Journal of Surgery* 1992; **163**: 53–8.

Hunter JG. Laparoscopic cholecystectomy and the common bile duct. *Surgical Endoscopy* 1994; **8**: 285–6.

Laycock WS, Hunter JG. Laparoscopic approach to bile duct stones. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 507–19. Churchill Livingstone, New York/London, 1996.

Petelin JB. Laparoscopic approach to common duct pathology. *American Journal of Surgery* 1993; **165**: 687–91.

Stoker ME. Common bile duct exploration in the era of laparoscopic surgery. *Archives of Surgery* 1995; **130**: 265–9.

Special considerations

Georgeson KE. Pediatric laparoscopy. In: Toouli J, Gossot D, Hunter JG, ed. *Endosurgery*, pp. 929–33. Churchill Livingstone, New York/London, 1996.

Holcomb GW III. Diagnostic laparoscopy: equipment, technique, and special concerns in children. In: Holcomb GW III, ed. *Pediatric endoscopic surgery*, pp. 9–19. Stamford, CT: Appleton and Lange, 1993.

Hunter JG, Swanstrom LL, Thornburg K. Carbon dioxide pneumoperitoneum induces fetal acidosis in a pregnant ewe model. *Surgical Endoscopy* 1995; **9**: 272.

Morgenstern L. Ethical considerations in laparoscopic choledocholithotomy. In: Berci G, Cuschieri A, ed. *Bile ducts and bile duct stones*, pp. 161–5. Saunders, Philadelphia, 1997.

Morrell DG, Mullins JR, Harrison PB. Laparoscopic cholecystectomy during pregnancy in symptomatic patients. *Surgery* 1992; **112**(5): 856–9.

Traverso LW, Hargrave K, Kozarek RA. A cost-effective approach to the treatment of common bile duct stones with surgical versus endoscopic techniques. In: Berci G, Cuschieri A, ed. *Bile ducts and bile duct stones*, pp. 154–60. Saunders, Philadelphia, 1997.

16.1 Basic transplantation immunology

Hugh Auchincloss, Jr

Introduction
The antigens that provoke rejection
MHC antigens
Minor histocompatibility antigens
The blood-group antigens
The stimulation of immune responses in transplantation immunology
B cells
T cells
Pathophysiology of the patterns of graft rejection in clinical practice
Hyperacute rejection
Accelerated vascular rejection
Acute rejection
Tolerance
Mechanisms of tolerance induction
Strategies to induce transplantation tolerance
Practical correlates of transplantation immunology
Antigen matching
The cross-match
Sensitization
Blood transfusions
Immunosuppressive treatments based on transplantation immunology
Conclusion
Further reading

Introduction

To a large extent, the clinical practice of transplantation surgery can be performed with a minimal understanding of immunology. This is partly because good operative technique and clinical management are more important than a knowledge of biology in achieving successful organ transplantation, and partly because the suppression of the immune system (which is a unique feature of transplantation surgery) is largely accomplished by the use of non-specific drugs that were not discovered, and are not generally used, on the basis of scientific knowledge. Nonetheless, a basic understanding of transplantation immunology is useful for transplant surgeons for several reasons: (a) some of the preoperative selection of transplant recipients and donors does depend on immunological principles; (b) a few of the treatments used to avoid or reverse rejection episodes are directed at specific elements of the immune system and are therefore applied most effectively if there is some scientific knowledge; and (c) because further advances in the field of transplantation, such as the ability to avoid the long-term use of non-specific immunosuppressive drugs or the ability to turn to animals as organ donors for human transplantation, will require a knowledge of how the immune system normally attempts to destroy non-autologous tissues.

Within the larger field of immunology, ‘transplantation immunology' is, for the most part, simply the application of a more general set of principles to the particular case in which the foreign antigens come from other individuals. However, there are two unique features of transplantation immunology that distinguish it from the study of all other types of immune responses. First, the major histocompatibility complex (**MHC**) antigens (which are among the foreign antigens expressed by organ donors) are a unique set of antigens for the immune system because they are the only foreign proteins that can be recognized by T cells without a requirement that they first be broken down to smaller peptides. Second, organ and tissue transplantation provides the only setting in which the immune system has two different sets of antigen-presenting cells, usually expressing different MHC antigens, available for stimulating an immune response.

This chapter describes the important antigens in transplantation and the elements of the immune response evoked by them that lead to graft rejection. In doing so it attempts to highlight the two unique features of the transplantation immune response and to concentrate otherwise on elucidating those issues in transplantation immunology that are, or are likely to become, most important in patient care.

The antigens that provoke rejection

Activation of the immune system requires that receptors on B cells or T cells recognize foreign antigens. Three types of antigens important in transplantation have been identified: (a) the MHC antigens, (b) the minor histocompatibility antigens, and (c) endothelial glycoproteins including the blood-group antigens.

MHC antigens

As their name implies, the MHC antigens are the most important antigens that cause graft rejection. Their presence was recognized from experiments using inbred strains of mice in which the products of one particular series of genes were found to be especially important in provoking graft rejection. The site within the genome that includes these genes is called the MHC. All species studied have been found to have a MHC and the antigens encoded within it have been well characterized in many of them.

Class I compared to class II MHC antigens

Two types of MHC antigens have been identified, now subdivided as class I and class II antigens. Over the years they have been given different names, as indicated in Table 1. The class I MHC antigens are composed of two chains, one of which is very polymorphic (i.e., varies from individual to individual) while the other, called b_2-microglobulin, is similar for all members of the species. The class II MHC antigens are composed of two polymorphic chains, a and b. Class I antigens exist on almost all cells of the body; class II antigens are less widely distributed, being expressed on dendritic cells, macrophages, B cells, and other cells with ‘antigen presenting' function.

Class I MHC antigens	Class II MHC antigens
Originally called SD antigens	Originally called LD or Ia antigens
Single polymorphic chain	Two polymorphic chains
Mol. wt 45000	Mol. wt 28 000 and 33 000
A, B, and C loci	DP, DQ, and DR loci
Expressed on all cells	Expressed on macrophages, dendritic cells, B cells, and vascular endothelium
Associated with β_2-microglobulin	

Table 1 Characteristics of human MHC antigens

Structure

Despite the differences between the class I and class II MHC antigens, the two types of molecule have a generally similar tertiary configuration. In both, each chain has a short intracytoplasmic tail followed by a transmembrane portion, and in both the two chains, joined together non-covalently, present a relatively large extracellular portion, which can be divided into four regions called domains. The effect is that both types of MHC antigens sit on the cell surface looking schematically like a four-leaf

clover (see Fig. 1).

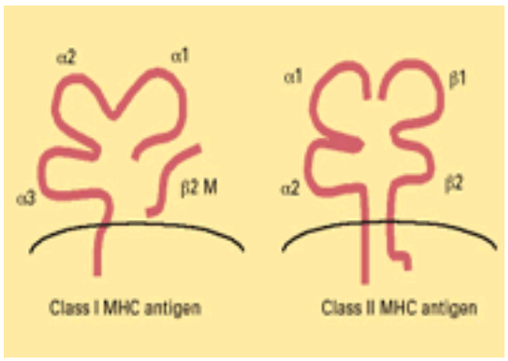


Fig. 1. Schematic diagram of MHC antigens showing their domains.

Nomenclature and genetic organization

At least three different loci within the MHC of humans encode polymorphic class I antigens. Three other loci encode class II antigens. The expression of MHC antigens is codominant (i.e., the genes on both chromosomes of a pair are expressed) and therefore up to 12 different MHC antigens are expressed on the cell surface (six encoded on the chromosome inherited from the father and six on the chromosome from the mother).

In humans the three class I loci are designated A, B, and C and the three class II loci are called DP, DQ, and DR. These six loci are organized on the short arm of chromosome 6 of humans such that the class I loci are together as are the class II loci (Fig. 2).

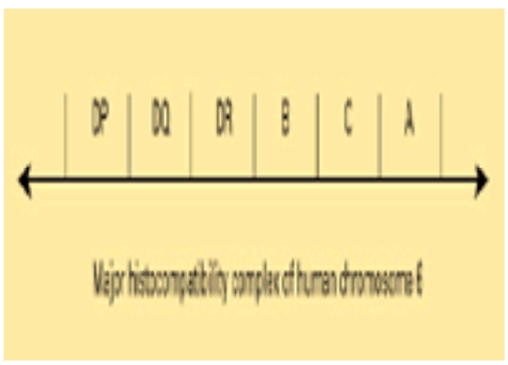


Fig. 2. Organization of the histocompatibility loci of the human MHC.

Polymorphism and the MHC cleft

One of the cardinal features of the MHC antigens is that there is significant variation in the structure of each antigen between individual members of a given species. This variation occurs because there are many different alleles within each species that encode the MHC antigen at any particular MHC locus. For example, there are several dozen different MHC class I antigens that can be encoded within the A locus. This variation in the MHC antigens between individuals is called polymorphism.

Structurally, most of the variation between different MHC antigens occurs in the outer two domains of the molecule. Recent crystallographic data have revealed that these areas together form a ‘cleft' in the exterior of the antigen and this cleft therefore has many slightly different configurations for the different MHC antigens (Fig. 3).



Fig. 3. Crystallographic picture of an MHC molecule.

HLA typing

The many different class I antigens encoded in the A locus have been given numbers such as A2, A3, etc. Similarly, numbers have been assigned to antigens of the other loci. In current clinical practice only the HLA-A, -B, and -DR antigens are usually defined for matching in cadaver transplantation. Since there are two antigens (one from each chromosome) for each of the loci, one person's phenotype might be described as:

HLA-A2, 3; B7, 41; DR1, 4

The process of determining the HLA phenotype is called tissue typing. It is often confused with the cross-match, a procedure that tests for the presence of pre-existing antibodies in a recipient that can react with a donor's MHC antigens. The two techniques may use similar methods but the implications of finding an HLA-matched kidney and one with a negative cross-match are completely different (see below).

Inheritance of MHC antigens

In typical mendelian fashion, every person inherits one copy of the MHC from each parent. Each chromosome copy is called a haplotype. Unless the two parents express some of the same histocompatibility antigens, children will share half of their MHC antigens with each of their parents. This is called a one-haplotype match, and parents and children are often said to be ‘haplo-identical'.

The sharing of MHC antigens between siblings is more variable. Since each parent has two MHC haplotypes and will pass on only one to each child, there are four possible combinations of haplotypes that may be inherited; this is shown diagrammatically in Fig. 4. For a particular child the chance that any sibling will have inherited the same two MHC haplotypes is 1/4, in which case that sibling is genotypic ‘HLA-identical'. The chance that a sibling will have inherited only one of the same MHC

haplotypes is 2/4, in which case that sibling is ‘haplo-identical'. The chance that a sibling will have inherited two different MHC haplotypes is 1/4, in which case that sibling is MHC mismatched. These odds are altered slightly by the possibility of genetic recombination occurring within the MHC, such that it is possible for a sibling to share 1½ MHC haplotypes.



Fig. 4. Inheritance of MHC haplotypes. For a particular offspring seeking an organ transplant, 25 per cent of the siblings will be MHC identical (e.g.a/c+a/c), 50 per cent will be haplo-identical (e.g. a/c+a/d), and 25 per cent will be unmatched (e.g. a/c+b/d).

The function of MHC antigens

The presence of MHC antigens in every species and the strength of the immune response they evoke suggests that they are important components of the immune system. However, despite their discovery in transplantation experiments and their importance in graft rejection, the MHC antigens do not exist in order to cause graft rejection. Despite this point, it was not until the 1970s that immunologists were able to assign a function to the MHC antigens beyond their role in transplantation.

It is now believed that the MHC molecules serve as focusing elements that direct the cellular elements of the immune system to the site where they can do the most good. If one considers a viral pathogen, for example, it would be inefficient to occupy an entire T cell in neutralizing one, or even several, tiny viral particles. Instead, this can be accomplished by the small, soluble, and abundant antibodies produced and secreted by B cells. The bulky machinery of T cells would be better held in reserve to eliminate the much larger cells that have been damaged by viral infection.

The MHC antigens accomplish their purpose by acting as antigen-presenting structures. Invading pathogens are broken down within cells into small peptide fragments and then carried to the cell surface in the polymorphic clefts of MHC molecules. These peptides, sitting in the cleft of MHC antigens, are then exposed to the receptors of passing T cells. The cleft of each MHC antigen is different, and, in each case, it will only have the shape to present some peptide fragments of some foreign pathogens. Therefore, it is to the advantage of an individual to express more than one kind of MHC antigen. If the cleft of one MHC antigen cannot present any peptide from a particular virus, the cleft of another might be able to do so. Furthermore, it is to the advantage of the species to have many different forms of each MHC antigen so that if one individual is unable to present peptides of the pathogen using any of its MHC antigens, another individual, expressing different MHC antigens, may be able to do so. Thus, the antigen-presenting function of MHC antigens appears to explain the multiple loci within the complex and the polymorphism of the antigens encoded there.

While the MHC antigens serve to present the peptides of pathogens on the surface of infected cells, the goal of restricting T cells so that they recognize these peptides only in this setting requires an additional feature. This restriction is imposed on the immune system by selection during maturation only of T cells with receptors that are triggered by peptide fragments presented in association with MHC antigens. Part of the T-cell receptor must recognize the peptide itself and part must recognize surrounding determinants formed by the MHC molecule. Because of this dual recognition by T-cell receptors, cell-mediated responses are said to require ‘associative recognition'. The part that MHC antigens play in limiting the site where peptide peptides can be recognized has led to their description as ‘restriction elements'.

Although MHC antigens do not exist to prevent organ transplantation, they nonetheless can act as the targets of the immune response in organ rejection. These MHC antigens are particularly important, in part because they elicit both T- and B-cell immune responses. Second, because the MHC antigens are so polymorphic, it is very unusual to find two unrelated individuals whose antigens are not different. Third, it is a peculiar feature of transplantation immunity that the strength of cell-mediated immunity to allogeneic MHC antigens is very strong. For example, the precursor frequency of T cells that respond to a foreign MHC antigen is roughly 100-fold higher than the frequency of T cells that respond to the peptides of a pathogenic virus presented with an MHC molecule.

The strength of anti-MHC alloreactivity

The response of T cells to the alloantigens of donor grafts is not surprising, since they would be expected to respond to any foreign antigen. What is not expected, however, is that the allogeneic response should prove to be so powerful. The precursor frequency of T cells that respond to a single determinant of an allogeneic MHC antigen is in the range of two per hundred T cells. In contrast, the precursor frequency of T cells responding to a determinant formed by a foreign virus or other extrinsic antigen, presented in association with a self MHC antigen, is about 1 in 10 000. Thus the allogeneic response is enormously powerful relative to the response to a pathogenic virus and, indeed, the allogeneic T-cell response is the most powerful immune response that has been measured.

An absolute answer is not available to the question of why T-cell reactivity to allogeneic MHC antigens is so strong when the immune system presumably evolved to protect us from viruses and other pathogens and not to protect us from receiving kidney transplants from cadaveric donors. However, a great deal of information has recently been learned about the process of selecting T cells during the development of a mature immune system and this has provided some reasonable explanations for the exceptional strength of alloreactivity.

T-cell development depends on, and occurs in, the thymus. The thymus is also an important site at which T cells that react strongly with self MHC antigens are eliminated before maturation. In addition to these functions, however, the thymus is also critical in positively selecting T cells that do have affinity for modified forms of self MHC molecules. The delicate balance achieved in the thymus between negatively and positively selecting cells, based on their affinity to the same set of MHC molecules, accomplishes the immune system's purpose of eliminating T cells that are autoreactive while developing the capacity to detect foreign peptides within the cleft of self MHC molecules. An important outcome of this process is that intact proteins in the environment cannot be recognized in their native configuration by the T cells of a mature immune system, and that instead they can only be detected after they have been broken down into small peptides and presented in the context of self MHC molecules.

There is one critical exception to this rule that only processed peptides, not intact foreign proteins, can be recognized by T-cell receptors: allogeneic MHC molecules have a configuration so similar to that of modified forms of self MHC molecules that they can, in fact, be recognized by T-cell receptors in their native configuration. This observation is sometimes referred to in transplantation immunology in the shorthand phrase that ‘allo = self + X'. As depicted in Fig. 5, the mature T-cell repertoire that has been selected to be (a) unable to recognize foreign proteins in their native configuration, and (b) to be unable to recognize self MHC molecules presenting peptides of self proteins, and (c) to be able to recognize peptides of foreign proteins presented by self MHC molecules, can, as a result of structural similarity, (d) recognize allogeneic MHC molecules no matter what peptides they might be presenting, without the requirement that the allogeneic MHC molecules themselves be processed into peptides. In transplantation immunology, it is said that allogeneic MHC molecules can be recognized ‘directly', and this is the unique feature of these foreign antigens compared to all other foreign antigens. Direct recognition of allogeneic MHC antigens is one of the two critical aspects of transplantation immune responses that distinguishes them from all other types of immunity.

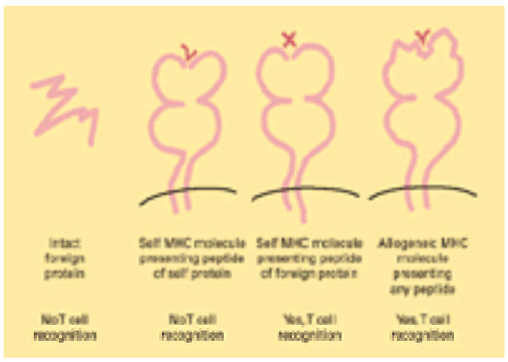


Fig. 5. The T-cell repertoire includes recognition of modified self MHC molecules and of allogeneic MHC molecules.

Although direct recognition of allogeneic MHC antigens is unique, it does not, in itself, explain why the alloreactive response is more powerful than that for modified self MHC antigens. Why should the number of T cells that react with a particular allogeneic MHC determinant be so much larger than the number of T cells that react with a particular ‘self + X' determinant?

Several theories have been proposed to answer this question. One of them suggests that the high precursor frequency of alloreactive T cells does not result from more T cells recognizing allogeneic MHC determinants but rather from the greater number of allogeneic determinants expressed on the surface of an allogeneic antigen-presenting cells. This has sometimes been referred to as the ‘determinant density hypothesis', and its logic focuses on the difference between self antigen-presenting cells and allogeneic antigen-presenting cells, as depicted in Fig. 6(a). An immune response to the measles virus requires that peptides of that virus be presented in association with self MHC antigens on a self antigen-presenting cell. But only a few of the MHC antigens on a given antigen-presenting cell are engaged in the process of presenting measles peptides, while the other MHC antigens are occupied presenting other peptides of autologous proteins, forming determinants that do not stimulate autologous T cells. In contrast, every one of the allogeneic MHC antigens on an allogeneic antigen-presenting cell will be able to stimulate a T-cell response, since these allo-MHC antigens can each be recognized directly. According to this hypothesis, the greater strength of alloreactivity is partly an artefact of our assay systems: there are more copies of a particular allogeneic determinant on allogeneic antigen-presenting cells than of modified self determinants on self antigen-presenting cells. Therefore, alloreactive T cells will be stimulated more easily than those that recognize modified self determinants and their number will appear larger by precursor-frequency analysis.

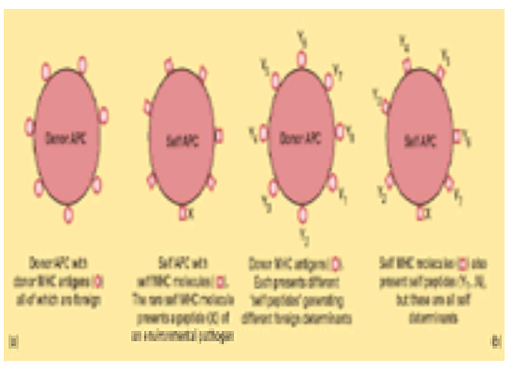


Fig. 6. (a) Determinant density hypothesis. Donor antigen-presenting cells (APC) with donor MHC antigens (○), all of which are foreign (left). Self APC with self MHC molecules (▭). The rate at which self MHC molecules present a peptide (x) of an environmental pathogen is very low (right). (b) Determinant frequency hypothesis. Donor MHC antigens (○), each present different self peptides generating different foreign determinants (left). Self MHC molecules (▭) also present self peptides (Y^1....^{n}). but these are all self determinants.

An alternative explanation, sometimes referred to as the ‘determinant frequency hypothesis', focuses on the notion that almost all MHC molecules are engaged in presenting peptides all of the time, but that most often those peptides are from autologous proteins, forming determinants that are considered self determinants by T cells. For example, as depicted in Fig. 6(b), some MHC antigens may present peptides from serum albumin while others might present peptides from haemoglobin. According to the determinant frequency hypothesis, the MHC antigens of an allogeneic antigen-presenting cell also present these and other peptides of ordinary self proteins in association with the allogeneic MHC molecules. However, the association of an albumin peptide with an allogeneic MHC antigen or of a haemoglobin peptide with that MHC antigen would not form a self determinant. Therefore, a single allogeneic MHC antigen on an allogeneic antigen-presenting cell would generate many different foreign determinants for responding T cells, explaining the higher precursor frequency created by the difference of a single MHC molecule.

These hypotheses are not mutually exclusive, and both the density and the frequency of allogeneic determinants probably contribute to the strength of alloreactivity. Since most studies of alloreactive T cells have suggested that they generally do have a degree of peptide specificity, the determinant frequency hypothesis is probably the stronger of the two. However, debates between these two, and other hypotheses, are much less important than the widely accepted notion that the unusual strength of alloreactivity stems fundamentally from the ability of T cells to recognize foreign MHC molecules directly on foreign antigen-presenting cells, without the requirement that they be broken down into peptides and presented by self MHC molecules on self antigen-presenting cells.

Minor histocompatibility antigens

Minor histocompatibility antigens are simply defined as antigens that cause cell-mediated graft rejection but do not have the structure of MHC antigens or of the closely related, non-classical, MHC-like molecules. As their name implies, individually these antigens stimulate graft rejection weakly, although, under some circumstances, multiple minor antigen disparities may cause graft rejection that is as rapid as when MHC differences exist.

For many years it was assumed that minor histocompatibility antigens would also turn out to be cell-surface glycoproteins, like the MHC antigens, but that something about their structure or their level of expression made them weak stimulators of T-cell responses. It has now become clear that this a completely inaccurate view of the minor antigens. Instead, it is recognized that they are formed from the peptides of cell proteins that have allelic variations between individuals in the population. As depicted in Fig. 7, two different amino acid sequences for cellular protein X (perhaps haemoglobin) could give rise to two distinct peptides that could be presented by the same MHC molecule. Under the circumstances in which the MHC antigens of a donor and recipient were identical, the determinant formed by one peptide with that MHC molecule would be different from the one formed by the presentation of the other. The precursor frequency of T cells that would respond to this ‘self + X' determinant would be expected to be like that of T cells responding to peptides of environmental pathogens, and thus much lower than for allogeneic MHC antigens that can be recognized directly.

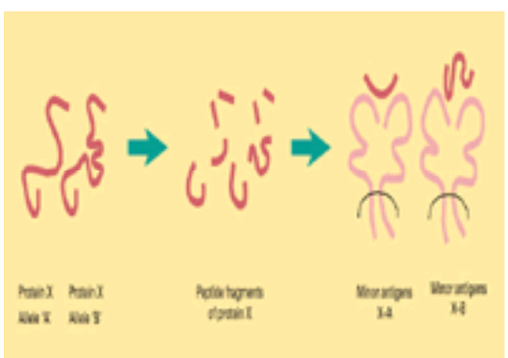


Fig. 7. Minor histocompatibility antigens are peptides of proteins with allelic variation presented in association with MHC molecules.

Although there are probably many hundreds or even thousands of proteins in the body that have minor differences in amino acid sequence between individuals, attempts to measure the number of minor histocompatibility antigens in two MHC-identical individuals have generally suggested that about 30 to 50 loci exist. This relatively low number is probably because a simple difference in an amino acid sequence is not enough to form a minor histocompatibility antigen. The difference must also be one which generates a peptide that is capable of presentation by the individual's MHC molecules and the peptide–MHC complex must be one which is immunogenic to T cells. Furthermore, depending on the type of graft rejection being used to detect the minor antigen dysparities, multiple simultaneous antigenic differences may be required to see an effect. Determining the exact number of minor antigens is not important, however, since it is clear that the number is neither tiny (one or two) nor infinite. Thus, from a practical point of view there are lots of minor histocompatibility antigens, but not so many as to stimulate immune responses of boundless proportions.

The blood-group antigens

The blood-group antigens are carbohydrate determinants expressed on glycoproteins present on the surface of red blood cells and some other cell types. They were identified and characterized because of their importance in blood transfusions, an early example of tissue transplantation.

Three major blood-group antigens are commonly recognized, designated O, A and B, each arising from a single carbohydrate backbone. The genes responsible for the several blood-group antigens encode glycosylation enzymes that modify the underlying substrate differently. Individuals of group O express the unmodified core, those of group A express an additional external sugar, and those of group B add a different additional external sugar. Individuals of group AB have the enzymes for both additional sugar molecules. The inheritance of the major blood-group antigens follows mendelian principles for a single locus with three allelic genes (O, A, B). The relation between genotype and phenotype is shown in Table 2.

Blood group	Genotype	Antibodies
O	OO	Anti-A and Anti-B
A	OA or AA	Anti-B
B	OB or BB	Anti-A
AB	AB	None

Table 2 Blood group antigens

There is no T-cell immune response to the blood-group antigens but there is a B-cell response and antibodies to these structures. The development of blood-group antibodies does not require prior exposure to foreign red blood cells themselves because the carbohydrate determinants of blood groups A and B are also expressed on common bacterial organisms and are thus encountered by everyone, starting early in life. Since individuals do not form antibodies against structures that they express themselves, individuals with group AB do not form antibodies to the other blood-group antigens. However, people with group O develop both anti-A and anti-B antibodies, and A and B individuals develop antibodies to the antigens of each other.

For blood transfusion, there are many other antigens on the surface of red blood cells that must be considered. However, for organ transplantation, only the ABO antigens need to be considered because they are the only ones expressed on vascular endothelium, where they become potential targets for pre-existing antibodies of the recipient that may damage the freshly transplanted organ. Therefore, a general principle in transplantation immunology is that solid organ transplantation should not be performed across blood-group barriers.

In practice, there are three modifications of this general principle that are noteworthy. First, not all organs are equally susceptible to antibody-mediated rejection by blood-group antibodies and the liver, in particular, is sometimes transplanted across blood-group barriers. Second, there are two subgroups of group A, called A_1 and A_2, and individuals of group O or B may not form antibodies that react with the A_2 determinant. Therefore, it is sometimes possible to cross this blood-group barrier in this special case. Finally, removal of blood-group antibodies from potential recipients by plasmapheresis before transplantation has occasionally allowed successful transplantation across blood-group barriers. This approach is not widely applied, however.

The stimulation of immune responses in transplantation immunology

The two components of the immune system most important in the response to transplanted tissues are the B and the T cells. Other components participate as well, including natural killer cells in the case of bone marrow transplantation, and various non-specific cell populations such as macrophages.

B cells

The antibodies produced by B cells that are important in transplantation are directed against either the donor's blood-group antigens (as discussed above) or against the donor's MHC antigens. While the blood-group antibodies develop early in life, the development of anti-MHC antibodies requires exposure to tissues from other individuals, which may occur as a result of pregnancy, blood transfusions, or organ transplantation. For both blood-group and MHC antigens, the initial antibody response is primarily of the IgM isotype, but in the case of anti-MHC antibodies, the B cells tend to shift to the production of various IgG isotypes if they receive help from activated T cells.

The crucial distinction in B-cell responses in transplantation is the timing of the appearance of antidonor antibodies: (a) if the antibodies exist before the time of transplantation they are likely to cause hyperacute rejection; (b) if they appear very rapidly after transplantation they may cause accelerated vascular rejection; and (c) if they appear over the course of weeks to months they may contribute to chronic rejection. Each of these mechanisms of rejection is discussed below.

The technique to determine whether any antidonor MHC antibodies exist before a possible organ transplant is called a ‘cross-match' and is performed in its simplest form by taking serum from the recipient and mixing it with cells from the prospective donor and then adding an exogenous source of complement. If lysis of the donor's cells occurs, then antidonor antibodies have been identified and the cross-match is said to be ‘positive'. Transplants of most types of primarily vascularized organs are not performed across a blood-group barrier or in the presence of a positive cross-match.

T cells

Two primary classes of T cells can participate in graft rejection: CD4+ T cells, which respond either to allogeneic MHC class II antigens directly or to modified forms of self class II molecules, and CD8+ T cells, which can respond to allogeneic MHC class I molecules directly or to modified forms of self class I molecules.

In both cases, the activation of these T cells requires two concurrent signals. The first signal is provided by the interaction of the T-cell receptor with its specific determinant (made up of some combination of an MHC molecule and a peptide). The second has been called ‘the second signal', but it is actually made up of several different elements that include the availability of certain cytokines (such as interleukin 2), and interactions between certain ‘costimulatory' molecules (such as CD28) and their ligands expressed by the stimulating cell population. Some of the components of the second signal are shown diagrammatically in Fig. 8. There are redundancies amongst these components and differing requirements depending on the type and previous history of the T cells involved, so the information in Fig. 8 is schematic and not precise.

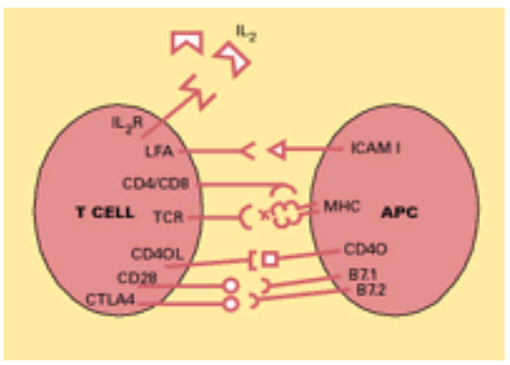


Fig. 8. T-cell activation.

An important consequence of the need for two signals for T-cell stimulation is that not all cell types can actually accomplish this when they come into contact with resting T cells. In the first place, not all cell types even express the class II MHC molecules that are the targets for the antigen receptors of CD4+ T cells. Secondly, not all cells express the necessary costimulatory molecules or generate the appropriate cytokines to provide the second signal. The cells in the immune system that do have both features are called ‘antigen-presenting cells'; they include macrophages and dendritic cells (which in different organs have different specific names, such as Langerhan's cells in the skin or Kuppfer cells in the liver). Other class II-expressing cells, including B cells and endothelial cells, may also develop antigen-presenting functions under certain circumstances.

Two sets of antigen-presenting cells: direct and indirect antigen presentation in transplantation immunology

The need for T cells to interact with antigen-presenting cells to achieve activation is a general requirement for all types of immune response. Another unique feature of transplantation immunology (the second unique feature along with the ability to recognize donor MHC antigens without breaking them into peptides) is that there are two sets of antigen-presenting cells which might potentially be able to stimulate recipient T cells, the set that comes from the donor and the set that comes from the recipient. In most cases the two sets express different MHC molecules. The role of donor-derived antigen-presenting cells in stimulating recipient T cells is made possible because donor MHC antiguns can be recognized directly; therefore, stimulation by donor antigen-presenting cells has been referred to as stimulation by the ‘direct pathway'. Stimulation of T cells by recipient antigen-presenting cells can occur if peptides of the donor's foreign antigens are presented by self MHC molecules on recipient cells. To distinguish this form of stimulation from that occurring through the direct pathway, this presentation of donor antigens by self antigen-presenting cells has been called stimulation through the ‘indirect pathway'.

The terminology developed to identify this critical feature of transplantation immunology is flawed in two important ways. First, the use of the term ‘indirect' to describe the presentation of peptides of donor antigens by self MHC molecules suggests that the process is in some way abnormal or unimportant. Actually, the indirect pathway involves the normal process of an immune response and it is the direct pathway that is unusual. Thus, the appropriate translation of the phrase ‘indirect pathway' should always be ‘normal immunology'. The second problem with the direct/indirect terminology is that confusion has developed over exactly what the terms refer to. Some have used the word ‘indirect' to describe any processing of peptides of donor MHC molecules and their presentation by other MHC molecules. For example, if there was a MHC class I antigen disparity but class II antigen matching between a donor and a recipient, the presentation of the donor's class I peptides by the donor's MHC class II molecules would be considered ‘indirect' recognition by this definition. The problem with this definition is that it would suggest that the determinant formed by the presentation of the donor's class I peptides by the donor's class I molecule would also be a determinant recognized by the indirect pathway, while the determinants formed by all the other peptides presented by the donor's class I molecules would be determinants recognized directly. This is not a useful distinction, and thus a definition based an whether peptides are processed and presented by MHC molecules during alloreactivity is not satisfactory. The best definition of direct as compared to indirect recognition is to distinguish the two pathways by the nature of the antigen-presenting cells that stimulate the T cells. Stimulation by donor antigen-presenting cells defines the direct pathway, while stimulation by recipient antigen-presenting cells defines the indirect pathway.

For many years, transplantation immunology was dominated by consideration of the direct pathway of T-cell stimulation. This was partly because direct stimulation is clearly the stronger immune response when T-cell alloreactivity is measured *in vitro.* In addition, some experimental studies had indicated that depletion of donor antigen-presenting cells allowed prolonged allograft survival without the use of other immunosuppression. Thus, the direct pathway appeared to be the dominant one *in vivo* as well. Finally, the importance of MHC antigens compared to minor histocompatibility antigens in causing graft rejection seemed to demonstrate the importance of the direct pathway, since there are reasonable explanations for there being an unusually strong T-cell response if donor MHC molecules are recognized on donor antigen-presenting cells, but it was not apparent why the peptides of foreign MHC molecules would be any more important than the peptides of any other donor proteins (minor histocompatibility antigens) in stimulating an indirect response.

Despite these powerful arguments, interest in the possibility that indirect responses may play an important part in graft rejection has increased in recent years. This change has come about in part because the availability of genetically altered mice which lack the expression of their own MHC molecules has made it possible to examine the rejection of tissues that cannot stimulate a direct response. The results of these types of experiment have suggested that, in many cases, powerful rejection mechanisms continue to operate. Furthermore, now that large numbers of MHC molecules have been sequenced entirely, it has become possible to perform experiments in which the immune system is manipulated by exposure to peptides of allogeneic MHC molecules without exposure to the intact molecules. These types of experiment have frequently suggested that changes in a recipient's indirect response to donor antigens are sufficient to alter the rejection of subsequently transplanted tissues that should be able to stimulate both direct and indirect responses. Thus, the indirect pathway has appeared to be the dominant one in these cases. Finally, with increasing attention to the problem of chronic rejection, there has been recognition that donor antigen-presenting cells are largely replaced over time in transplanted organs by cells from the recipient's bone marrow. Thus, it has seemed that indirect responses are likely to be paramount during the longer-term rejection of transplanted organs.

The examination of the relative roles of direct and indirect responses has become one of the most important themes in transplantation immunology at the present time. Numerous questions have emerged that have not yet been answered, some of which are listed in Table 3, and this list could easily be extended. These questions are important, partly because the selection of approaches to influence graft rejection by genetic manipulations of donor tissue will depend heavily on their answers and because the choice between various strategies for the induction of transplantation tolerance will be influenced partly by our understanding of which pathway of T-cell stimulation is more important over time.

Which pathway causes more powerful rejection?
Are there differences in the character of the T-cell responses stimulated by the two pathways, perhaps because of differences in the particular antigen-presenting cells involved?
Are there differences in susceptibility to immunosuppression between the two pathways?
Which peptides are most important in stimulating an indirect response?
Can indirect responses by CD4+ T cells contribute to graft rejection?
Under what circumstances can T cells that are stimulated through the indirect pathway interact with and influence T cells that are stimulated by the direct pathway?
To what extent can T cells stimulated by the indirect pathway, which therefore may not recognize determinants expressed by the donor cells, contribute themselves to graft destruction?
How can manipulations of T-cell responses involving only one pathway affect or indirect affect graft rejection, when either pathway alone seems to be capable of achieving graft destruction?
Is the indirect pathway the critical pathway in chronic rejection or do memory responses by T cells stimulated directly play an important role?
Are there circumstances in which either direct or indirect responses may tend to diminish graft rejection?

Table 3 Some unanswered questions about direct compared to indirect pathways of rejection

The development of T-cell effector mechanisms for graft destruction

The activation of T cells by their interaction with antigen-presenting cells is only the first step in the development of cell-mediated effector mechanisms that lead to graft rejection. The process goes on to involve interactions between different populations of T cells, the production of various cytokines, and the development of particular T-cell functions. Figure 9 outlines a simple scheme of the types of T-cell interactions that are commonly thought to be involved in this process, including the possibility

of either direct or indirect sensitization of CD4+ helper T cells. It is a useful scheme to keep in mind. However, it is equally important to understand that the scheme shown in Fig. 9 is unquestionably incomplete and very probably completely wrong.

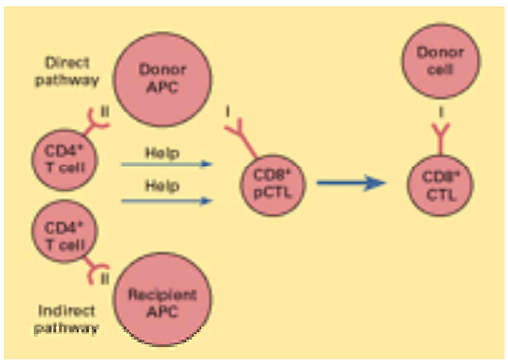


Fig. 9. Simple schematic model of the T-cell interactions leading to graft rejection (II = class II MHC antigen, I = class I MHC antigen).

One example of a problem with the diagram shown in Fig. 9 is its depiction of the final effector mechanism of graft destruction as depending on CD8+ cytotoxic T cells. Actually, transplantation immunologists have been unable to identify how many effector mechanisms there are or the nature of any one of them. Neither the original notion, that T cell-induced graft destruction is due to the development of delayed-type hypersensitivity responses, nor the more recent assumption, that the process depends on the development of cytotoxic T cells, are well supported by the data. Another example is the prediction from Fig. 9 that graft rejection depends on the presence of both CD4+ and CD8+ T-cell populations, the first serving as helper cells for the latter, which are the effector cells for graft destruction. In many experimental systems, however, graft rejection can still occur after either one of the two types of T cells is depleted.

The body of experimental work on the mechanisms of T cell-mediated graft destruction is large and very complicated; key findings include the following. First, several types of experiment have demonstrated that the process of graft destruction is remarkably selective, causing death to foreign tissues while having a minimal impact on recipient tissues in the immediate vicinity. Thus, to a substantial degree the mechanisms of graft destruction are thought to depend on the specificity of the T-cell receptor. Second, numerous studies have indicated that either CD4+ or CD8+ T cells have the capacity to cause graft rejection and that in many cases each population can accomplish this without the participation of the other. On the other hand, it has also been possible to demonstrate conditions in which both populations are required, suggesting that interactions between the two, probably in the form of CD4+ helper cells working with CD8+ effector cells, are sometimes involved in rejection. Third, many clinical and experimental studies suggest a strong correlation between the expression of some of the molecules of T-cell cytotoxicity (such as the granzymes or perforin) and the onset of rejection. Nonetheless, experiments with genetically altered mice that lack the components of T cell-mediated cytotoxicity have shown no effect of these deficiencies and none of the deletions of cytokines or particular cytokine receptors has prevented rejection. On the basis of these findings, it appears that there are probably multiple effector mechanisms for graft rejection, mediated by different populations of T cells, and with interactions between them in many cases.

Pathophysiology of the patterns of graft rejection in clinical practice

Four main patterns of rejection are usually identified in clinical practice: (a) hyperacute rejection, (b) accelerated vascular rejection, (c) acute rejection, and (d) chronic rejection. To a large extent, these clinically recognized patterns appear to correlate with particular underlying immunological mechanisms (Table 4).

Type of rejection	Antigen	Immune response	Time frame
Hyperacute	MHC/blood group	B cells: pre-existing antibodies	Minutes
Accelerated vascular	MHC	B cells: rapidly induced antibodies	Days
Acute	MHC/minor	T cells	Weeks
Chronic	MHC/minor	T and B cells	Years

Table 4 Clinical patterns of graft rejection

Hyperacute rejection

Hyperacute rejection usually occurs within minutes after restoration of the blood supply to a transplanted organ. After an initial period of normal circulation, the transplanted organ becomes swollen and discoloured by interstitial haemorrhage and then undergoes intravascular thrombosis. There is no known intervention that can do more than slightly delay this process once it is begun.

The cause of hyperacute rejection was recognized in the 1960s to be pre-existing antibodies in the recipient's circulation that bind to donor endothelial antigens and then activate the complement system. The membrane attack complex (C5–9) of the complement cascade is particularly important in causing a form of immediate endothelial activation that leads to the separation of one cell from another (hence the haemorrhage and fluid extravasation) and to the release of procoagulant factors (hence the intravascular coagulation).

Although it has devastating consequences, hyperacute rejection is now rarely encountered in clinical transplantation. This is because it can be avoided so reliably by testing the recipient immediately before surgery for the presence of antidonor antibodies. Nonetheless, the likelihood of hyperacute rejection is an important obstacle preventing many potential recipients from ever receiving a transplant because they have developed antibodies that react with so many of the allogeneic MHC antigens. On the other hand, the potential for hyperacute rejection is not an obstacle to all forms of transplantation because the liver seems to be highly resistant to this type of rejection and because cellular transplants do not have a vascular endothelium, which is the target for this type of rejection.

Accelerated vascular rejection

Accelerated vascular rejection is one of several terms for a type of rejection that generally occurs within the first week after surgery and is also quite rare in clinical practice. Other names for the process have included simply ‘vascular' or ‘humoral' rejection. Its three cardinal features are fibrinoid necrosis of the small arteries with evidence of intravascular thrombosis, a relatively scant cellular infiltrate, and the development of a positive cross-match due to the new generation of antidonor antibodies.

It is currently believed that this type of rejection is humorally mediated and depends on the very rapid development of new antidonor antibodies. In contrast to the process in hyperacute rejection, when the vascular endothelium is attacked by a sudden, massive antibody challenge, the onset of this antibody response over several days leads to a different type of endothelial activation, which has been called ‘type II endothelial activation'. This activation depends on the transcription of new genes, especially those regulated through NF-B, and the synthesis of new proteins. Many of the factors produced are those associated with inflammation, including several inflammatory cytokines and several adhesion molecules responsible for attracting non-specific inflammatory cells. Furthermore, type II endothelial activation causes changes in the expression of cell-surface molecules that regulate the clotting cascade. These events account for the histopathological findings of vascular endothelial injury with intravascular coagulation.

This type of rejection does not occur very frequently in clinical practice because its development appears to require unusually rapid induction of antidonor antibodies.

The more typical antidonor B-cell response, which develops over months or years after organ transplantation, does not induce the same type of vascular injury. Indeed, this type of accelerated rejection has been so rare that there was considerable uncertainty about whether it even existed as a separate entity or might simply be an unusually aggressive form of T cell-mediated acute rejection. However, experimental studies of xenogeneic transplantation between closely related species have regularly demonstrated this rejection process and allowed its more rigorous characterization. It now appears that its occurrence in these xenogeneic combinations, and in the unusual allogeneic cases, requires that small amounts of antidonor antibodies already exist before the transplant but at levels too low to be detected by standard assays and too low to trigger hyperacute rejection. The previously sensitized B cells then produce large quantities of them very rapidly when stimulated by the antigens of the organ transplant.

Unlike hyperacute rejection, which can never be reversed once initiated, it is possible to treat accelerated vascular rejection in some cases. The approach typically involves the use of plasmapheresis to remove the antidonor antibodies that already been formed, the addition of reagents especially effective in preventing B-cell responses (such as cyclophosphamide or mycophenolate mofetil), and the augmentation of anti-T-cell immunosuppression as well.

Acute rejection

For most of the history of clinical transplantation the most common type of rejection encountered has been during the first several weeks to months after organ transplantation and its effects have proceeded rapidly over the course of several days. The process is called acute rejection and it is mediated by T cells. Histologically, there is usually a substantial lymphocyte infiltrate when biopsies from organs undergoing acute rejection are obtained. However, the presence of the infiltrate by itself is not sufficient to make a diagnosis of an acute rejection episode, since random biopsies of transplanted organs often demonstrate such infiltrates even when there is no evidence of tissue injury resulting from their presence. One of the signs that active rejection is in progress is evidence of ‘endothelialitis' resulting from an attack by recipient lymphocytes on donor endothelial cells. Thus, all the different mechanisms of rejection seem to target the vascular endothelium of transplanted organs, making the use of the word ‘vascular' to distinguish the various processes unsatisfactory.

In the early years of clinical transplantation a majority of recipients experienced acute rejection episodes despite the use of non-specific immunosuppressive medicines. These episodes were often treatable by increased doses of the same medicines or by additional therapies, but a significant number of patients suffered multiple rejection episodes or proved to be refractory to the various forms of treatment. As a result, loss of transplanted organs to T cell-mediated acute rejection occurred during the first year in 30 to 40 per cent or more of transplant recipients. However, as shown in Fig. 10, there has been a near-constant improvement in the percentage of patients who have achieved 1-year graft survival over the past several decades. This has been accomplished by the discovery of increasingly effective, non-specific immunosuppressive treatments that can both prevent and treat acute rejection episodes. As a result, only a relatively small minority of transplant recipients ever experience acute rejection (somewhere between 10 and 30 per cent in most centres) and only a very small number of patients lose their organ transplants due to immunological causes during the first year.

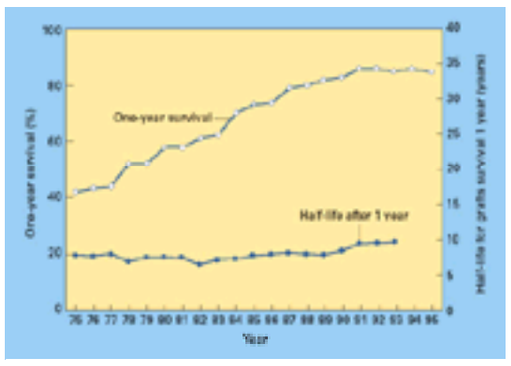


Fig. 10. One-year and long-term cadaver kidney transplant survival over several decades.

One of the most effective ways of treating acute rejection is to use one of several different antibody preparations specific for antigens expressed on T cells. For example, OKT3 is a monoclonal antibody produced in mice against the CD3 antigen, which is present on all T-cell subsets. When OKT3 is given to patients, essentially all CD3+ cells disappear from the circulation and acute rejection episodes are reversed more than 90 per cent of the time. The effectiveness of this specific anti-T-cell reagent is one of the reasons why acute rejection is believed with such confidence to be a T cell-mediated process.

Unfortunately, the reversal of a rejection episode does not guarantee that another will not occur and OKT3 cannot be given indefinitely, partly because of the profound immunosuppression that would result. Thus, some but not all patients have recurrent rejection episodes. From the point of view of transplantation immunology, the notion that rejection can occur as episodic events is curious and poorly understood. Obviously, T-cell sensitization must have occurred at the time of the initial rejection event, but it is not clear what happens to these sensitized T cells or what triggers their reactivation in some cases at certain times, but not in others.

Chronic rejection

Chronic rejection refers to a clinical picture of slow deterioration in organ function over months or years. The deterioration is difficult to control by standard immunosuppression. With the decline in the frequency of acute rejection episodes, this is now probably the most frequent type of rejection encountered in clinical practice. Pathological examination of organs suffering chronic rejection often reveals a relatively sparse lymphocyte infiltrate and may show a characteristic ‘onion-skin' appearance with concentric cellular proliferation and obstruction of flow in the small arteries of the graft, a finding often associated with the presence of antibodies in the recipient specific for donor antigens. These features have suggested that chronic rejection may be caused by antibody-mediated mechanisms. However, experimental studies have suggested that both T- and B-cell responses can contribute to a process of chronic vascular injury and continuing repair that eventually produces the histological features of chronic rejection. Many non-immunological forms of injury probably also contribute to the process.

Tolerance

One of the cardinal features of the immune response is that it does not occur to self antigens under ordinary conditions. This ‘self tolerance' is critical to an individual's survival. Tolerance is also a source of fascination to those interested in transplantation, since its existence suggests that it might be possible to recapitulate nature, instructing the immune system to accept additional antigens (those of the donor graft) as ‘self', thus avoiding the need for non-specific immunosuppression.

The definition of the word tolerance is that the immune system has lost reactivity to one set of antigens while maintaining reactivity to another. Thus tolerance is by definition ‘specific' and the particular set of antigens can be specified, e.g. ‘donor-specific' tolerance. For clarity, the exact nature of the non-reactivity can also be stated, but tolerance in transplantation immunology is generally assumed to imply the long-term survival of a transplanted organ without the use of long-term, non-specific immunosuppression.

Mechanisms of tolerance induction

The effort to achieve tolerance for clinical transplantation has largely been an effort to recapitulate the immune system's own mechanisms for achieving self tolerance. A number of such mechanisms have been identified (Table 5). First, perhaps the most important mechanism for self-tolerance induction is the elimination of lymphocytes with self reactivity at an early stage in their development. For B cells, this occurs during their maturation in the bone marrow, and for T cells it occurs during their development in the thymus when they encounter self antigens on a population of bone marrow-derived, thymic dendritic cells.

Thymic deletional
Peripheral deletional
AICD (activation-induced cell death): FasL-mediated reality apoptotic gene induction
Veto-cell-mediated
Peripheral non-deletional
Suppressor:
T cell receptor-specific
Antigen-specific (? immune deviation)
Non-suppressor (? anergy)

Table 5 Mechanisms of transplantation tolerance

A second mechanism of tolerance involves the deletion of specific cells after they have matured and entered the periphery. In some cases this appears to result from encounters with antigen that cause what has been called ‘activation-induced cell death'. The interaction of a molecule called ‘Fas ligand' with a molecule called ‘Fas' that is expressed on activated T cells appears to be one of the triggers for activation-induced cell death. In other cases, peripheral deletion may be accomplished by the interaction of specific lymphocytes with ‘veto cells'. These cells have the capacity to kill cells that attempt to kill them.

A third category of tolerance-inducting mechanisms involves the disarming of mature lymphocytes that have entered the periphery without actually eliminating them from the system altogether. There appear to be several different ways to accomplish this. In some cases, suppressor cells develop that recognize specific antigens and inhibit destructive immune responses that would otherwise develop in response to these antigens. Exactly how this is accomplished is unknown, but it may sometimes involve the production of various cytokines that inhibit the development of certain types of T-cell responses. In other examples of peripheral, non-deletional tolerance, no suppressor activity can be identified but the T cells that are specific for a particular set of antigens have had their own internal signalling mechanisms altered so that they no longer respond to these antigens in a destructive way.

Strategies to induce transplantation tolerance

The most famous strategy for achieving transplantation tolerance was demonstrated by Medawar and his colleagues in 1953 when they injected donor cells into the uteruses of pregnant mice and showed that the mature offspring could accept skin grafts from the donor strain without additional immunosuppression. This form of ‘actively acquired tolerance', as Medawar described it, almost certainly recapitulated nature's mechanism for deleting ‘self-reactive' T cells in the thymus during the development of the immune system. Since the description of this strategy for transplantation tolerance induction, several dozen additional strategies have been described that work effectively in rodent experiments.

It is difficult to group these many strategies accurately according to the mechanism of tolerance induction they utilize, partly because in many cases the mechanism is not well established and partly because in some cases more than one mechanism may be involved. Furthermore, in this rapidly advancing field, what seems like well-established information at one moment is replaced by new a understanding with remarkable speed. With these caveats in mind, many of the strategies for the induction of transplantation tolerance that have been worked out in rodents can be grouped into five broad categories. First, a number of strategies have attempted to achieve thymic deletional tolerance in adult recipients by first ablating the mature immune system and then allowing a new system to develop in the presence of donor as well as recipient antigens. In its simplest form, this can be accomplished by the production of radiation bone-marrow chimeras, but to avoid the use of lethal, whole-body irradiation, non-myeloablative conditioning regimens have been developed, followed by the introduction of donor bone marrow either intravenously or directly into the thymus. Second, some of the strategies appear to be directed at the augmentation of activation-induced cell death as a means of peripheral deletion. In some cases this has been accomplished by the introduction of donor antigens under certain conditions (including by transplanting an organ in some experiments); in others it has been fostered by the overexpression of genes encoding molecules, such as Fas ligand, that promote activation-induced cell death. The third category of strategies has also attempted to accomplish peripheral deletion for tolerance induction by introducing donor veto cells, especially by adding donor bone marrow in which veto cells are particularly common. The last two categories probably involve peripheral, non-deletional mechanisms. In one case, they have involved the use of a wide variety of molecules that tend to block the development of an adequate second signal and as a group they have been referred to under the heading of ‘costimulatory blockade'. In the other case, they have made use of a variety of anti-T-cell antibodies, including anti-CD3 antibodies (sometimes coupled to a toxin conjugate), non-depleting anti-CD4 antibodies, and anti-interleukin 2-receptor antibodies.

Despite the success of these many strategies in small animals, there is no strategy currently in routine clinical use to achieve transplantation tolerance. This failure to achieve clinical application has many causes: (a) some strategies, such as Medawar's, would not be applicable in adult patients after a mature immune system has developed; (b) some of the treatments that lead to tolerance in small animals are not well tolerated by humans; (c) the degree of success that is scientifically remarkable in experiments on small animals is not necessarily equal to the nearly perfect outcome required in clinical practice; (d) the very success achieved in recent years of clinical organ transplantation in the short term makes it difficult to introduce new strategies that are not well proven; (e) some of the non-specific drugs currently in use to prevent graft rejection may actually prevent the induction of tolerance; (f) some features of the immune system in primates may not be represented well in rodent experiments; and (g) few if any available assays can measure the existence of tolerance in patients other than through simply stopping the immunosuppressive medicines that might turn out to have been essential to the transplanted organ's survival. These represent formidable obstacles to the achievement of transplantation tolerance in clinical practice. Nonetheless, the induction of donor-specific tolerance is viewed as so important by so many in the field that efforts are being made to initiate trials of some of the existing strategies and to develop yet new ones.

Practical correlates of transplantation immunology

This chapter has described some of the concepts in transplantation immunology. While understanding them is often not essential in clinical practice, there are some specific applications of the science that do make a difference.

Antigen matching

One of the frequently debated questions in clinical transplantation is how important it is to achieve matching of MHC antigens between donors and recipients. Although the question is controversial, the related data are reasonably well established. In the first place, organs and tissues can be exchanged between identical twins, who express identical transplantation antigens, without any rejection and without immunosuppression. Second, among living-related kidney donors, there is a demonstrable benefit from matching for all MHC antigens compared to none in terms of the frequency of rejection episodes and the long-term survival of the transplants. Third, the survival of kidneys from cadaveric donors varies according to the degree of MHC antigen matching and this benefit probably increases somewhat as more time elapses following transplantation.

These data support the notion that antigen matching ‘matters'. The controversy arises between those who view the benefits of partial antigen matching as large and those who see them as relatively small in comparison to the disadvantages that occur when seeking better compatibility. Despite their name, the MHC antigens are not the dominant factor in achieving survival of organ transplants in an era of highly effective immunosuppressive drugs. It is unlikely that there is even a 10 per cent difference in the 1-year survival rates of kidney transplants between the best-matched and worst-matched cases in most centres. Is it better to gain a few percentage points in survival for a new patient whose name has recently been added to a long list of waiting recipients on the basis of a more favourable match or to give that kidney to someone who has been waiting a year longer on dialysis? How much improvement in graft survival on the basis of compatibility justifies the cost of transporting kidneys over long distances and the negative effects of longer ischaemic times? These are difficult questions that still lack precise answers and therefore generate substantial disagreement.

In addition to emphasizing the relatively small importance of antigen matching in kidney transplantation, two other critical points are needed to put the issue of antigen matching in context. First, in contrast to organ transplantation, MHC antigen matching is of overwhelming importance in the field of bone marrow transplantation in determining both engraftment and subsequent graft versus host disease. In this case, incomplete matching for even a single MHC antigen dramatically increases the risks of the procedure. Second, in contrast to the heated debate over kidney allocation, MHC antigen matching has essentially no role in the distribution of other organs such as the heart or liver. This is for the simple, practical reason that there is rarely adequate time to obtain the necessary information because these organs must be transplanted with shorter ischaemic times than are acceptable for kidneys.

The cross-match

Antigen matching is controversial and some are uncertain that any effort is necessary to achieve it. The cross-match, on the other hand, is a different issue entirely and

universally recognized to be important in kidney transplantation. A cross-match is performed by mixing serum from a prospective recipient with lymphocytes of a potential donor to determine whether the recipient has antibodies specific for the donor's antigens. A positive cross-match predicts with high accuracy that hyperacute rejection will occur if kidney transplantation is attempted between that donor and recipient.

Although simple in concept, practical difficulties arise in clinical practice. One of these has to do with the sensitivity of the cross-match assay. One way to increase this sensitivity is to add antihuman antibodies from another species after mixing the donor's cells and recipient's serum. This effectively amplifies the binding of antibodies to the donor cells and increases the likelihood of achieving complement lysis. Another way is to perform the assay using fluorescence-activated scanners that can detect the presence of antidonor antibodies in much lower quantities. The problem with these solutions is that they may increase the assay's sensitivity too much, preventing some patients from receiving transplants that they would not actually have rejected. Another problem with the cross-match involves the specificity of the antibodies detected. The standard cross-match is done with T cells from the donor that do not express class II antigens. Some centres also do the assay with donor B cells that do express these antigens. It is clear, however, that not all patients with positive ‘B-cell cross-matches' will suffer hyperacute rejection. Finally, there is uncertainty about what to do about patients who have a negative cross-match with recently obtained serum samples, but a positive cross-match using serum from months or years before. It is known that these patients with an ‘historically positive' cross-match are at higher risk for poor outcomes from their transplant, but the risk is not 100 per cent by any means. These types of issues make the performance and interpretation of the cross-match an important specialty.

Sensitization

Potential recipients of kidney transplants who do not have any antibodies in their serum that react with foreign lymphocytes from any source are said to be ‘unsensitized'. Those who have been exposed to foreign antigens and who have developed antibodies against some or many different foreign MHC antigens are said to be ‘sensitized'. In transplantation this word is almost always used to describe the state of a recipient's B cells, even though in general immunology sensitized lymphocytes can be either B or T cells

Prospective recipients are tested for their level of sensitization by reacting their serum separately with lymphocytes from many different individuals, who are selected because they express a broad range of the human HLA antigens. If the donor's serum causes lysis of 45 per cent (for example) of the panel of different lymphocytes, they are said to have a 45 per cent PRA (panel reactive antibody). Some prospective recipients have antibodies that kill lymphocytes from every individual in the panel. They are said to be ‘highly sensitized' or ‘100 per cent sensitized'. Obviously it is extremely difficult to find donor kidneys for highly sensitized patients that will not generate a positive cross-match. However, it is not really ‘100 per cent' impossible, since a recipient will not have antibodies against the antigens of a donor who is matched for all the HLA antigens.

The process of testing each individual against the panel is useful in predicting the difficulty of finding them a suitable kidney donor. It is also useful in identifying the specificity of the antibodies of the sensitized patient and therefore in determining which HLA antigens are most likely to generate a positive cross-match. This makes it possible to select kidneys, even before a cross-match, that are most likely to be suitable for a highly sensitized recipient, thereby increasing the efficiency of the screening and distribution process.

Many highly sensitized patients end up waiting a long time for a negative cross-match and some never receive transplants at all. As a result there has been great interest in the possibility of removing preformed antibodies from sensitized patients by plasmapheresis. Successful transplants have occasionally been achieved in this manner, but the approach has generally been successful only in the case of ABO-mismatched individuals. On the other hand, some highly sensitized patients show a gradual decline, without any intervention, in both the percentage of ‘panel reactivity' of their serum and the titre of antibody activity against individual cell samples from the panel.

Blood transfusions

Blood transfusions are one of the ways in which individuals may encounter foreign MHC antigens and thus become sensitized. Therefore it seemed reasonable, in the early years of clinical transplantation, to avoid transfusing prospective candidates for kidney transplantation whenever possible. Some patients were asked to struggle with very low haematocrits while waiting for an organ. Despite the sound rationale for this approach, data gathered during the mid-1970s indicated that patients who had received blood transfusions before transplantation and who were then given a cross-match-negative kidney transplant had better survival of their transplants than did those who had never been transfused. Indeed this issue was studied nearly 70 times between 1973 and 1984, almost always with the conclusion that blood transfusions were beneficial. Thus it became standard practice in many transplant centres during the 1980s to insist on potential recipients receiving at least some blood transfusions before transplantation, whether they were anaemic or not.

Several explanations have been considered for the beneficial effect of blood transfusions. Among them, it is possible that the introduction of foreign MHC antigens, intravenously, may sometimes cause downregulation rather than stimulation of the immune response to those antigens, achieving a degree of tolerance. Although this finding encourages hope for tolerance induction, it remains unclear why some patients become sensitized and others may undergo downregulation of their immune response.

Although intentional transfusion of transplant candidates was the norm for many years, beginning in 1986 studies of the benefits derived from this approach suggested that the advantage was becoming very small and in some cases impossible to detect. This apparent change in an outcome that had been so repeatedly verified before was a startling event. The explanation for the shift probably lies in the improved results of transplantation for all recipients, whether transfused or not, derived from new immunosuppressive drugs and other technical advances. Thus it was probably not that transfusions suddenly offered no benefit but rather that the benefit was small, compared to the strength of cyclosporin, OKT3, and other drugs, in preventing or reversing allograft rejection.

Even while the new data were being evaluated, a still more powerful influence caused a decline in the use of intentional blood transfusions before transplantation. The appearance of acquired immune deficiency disease (**AIDS**) and the recognition that it could be transmitted by transfusion caused many patients to balk at transfusions offered in the absence of anaemia. Even when it became possible to assure them that the risk of contracting AIDS was negligible, the new awareness by both patients and physicians of the risks of transfusion diminished enthusiasm for intentional transfusion. In recent years the availability of recombinant erythropoietin for patients on dialysis has also decreased the need for blood transfusion to treat the anaemia associated with chronic renal failure. Thus, the number of patients coming to transplantation without ever receiving a blood transfusion has increased substantially in recent years.

Immunosuppressive treatments based on transplantation immunology

Many of the immunosuppressive medicines used currently in clinical transplantation were discovered by chance (often during drug screening for other purposes), and only subsequently were their mechanisms of action determined. Cyclosporin, for example, was not discovered during a search for immunosuppression and its ability to inhibit the T cells' utilization of interleukin 2 was worked out after its immunosuppressive properties had been identified. However, the use of antibodies directed against T cells, and against particular antigens on the surface of T cells, to achieve immunosuppressive for organ transplantation was clearly the outcome of a progressively sophisticated understanding of transplantation immunology.

The earliest development of antibodies against cells of the immune system was achieved by injecting human lymphocytes (or sometimes selected populations such as thymocytes) into animals such as rabbits or horses. After appropriate purification, the resulting antilymphocyte serum or antithymocyte globulin can be administered to patients to suppress the function of the immune system. These types of preparation continue to be used in clinical practice with good effect.

The next generation of anti-T-cell antibodies became available with the development of the technology to produce large quantities of monoclonal antibodies. The first of these antibodies to be approved for clinical use was OKT3. More recently, a class of monoclonal antibodies directed at the interleukin-2 receptor on T cells has been introduced into clinical practice. The two types of antibodies have quite different physiological functions, but together they illustrate a number of important aspects of clinical therapy with monoclonals.

OKT3 is directed at the CD3 antigen, a complex of proteins expressed on all T cells and part of the signalling apparatus that informs the inside of a cell that the receptor on its surface has been engaged. An important consequence of this function of the CD3 protein is that the first effect of administering OKT3 to a patient is that essentially all of their T cells become activated. The resulting release of cytokines, including interferon-g and tumour necrosis factor, sometimes causes dramatic ‘first-dose effects' of OKT3 therapy. These include high fevers and shaking chills, and the sudden onset of pulmonary oedema. A number of ways have been identified to avoid most of these unpleasant symptoms: (a) patients can be pretreated with high-dose steroids, (b) they can be treated with inhibitors of the various cytokines involved, such as soluble tumour necrosis-factor receptors, and (c) the OKT3 antibody can be modified in its Fc portion to alter it binding to Fc receptors on antigen-presenting cells, because this interaction is critical in achieving the activation of T cells. In practice, the simplicity of steroid administration has made this the standard approach, but the other treatments illustrate some of the further manipulations possible with monoclonal antibody therapy.

Within a few minutes of the administration of OKT3 the number of CD3+ cells that can be detected by fluorescence-activated scanning decreases dramatically. In addition, the number of CD4+ and CD8+ cells also falls, because these are the two main subpopulations of T cells. Because CD3+ cells are no longer in the circulation,

doses of OKT3 given after the first day tend not to cause the same degree of symptoms. The effectiveness of the daily OKT3 injections can continue to be monitored by fluorescence-activated scanning. Typically, this technique shows persistent elimination of CD3+ cells, but it also demonstrates the slow return of CD4+ and CD8+ cells, beginning on about the fifth day. It is believed that this event marks the return of cells of T-cell lineage (as indicated by their expression of the CD4 or CD8 antigens) but of cells that do not express the CD3 antigen, because of the presence of free anti-CD3 antibody in the circulation. This condition has been referred to as ‘modulation'. In the case of OKT3, the return of modulated T cells to the circulation is not of concern because the activation of T cells requires the CD3 protein complex. Thus, these modulated cells cannot be triggered. On the other hand, modulation of T cells for other surface antigens may allow the return of cells that can still cause graft rejection. This may be one of the reasons that OKT3 was such an effective monoclonal antibody in clinical practice.

Frequently, during OKT3 therapy, patients receiving daily doses of the antibody will again begin to develop fevers, on about the eighth or ninth day. Fluorescence-activated scanning usually indicates that CD3+ cells have started to return to the circulation and further testing may reveal that the recipient being treated has developed antibodies against the OKT3 antibody. This is because the anti-CD3 antibody is a mouse protein that is foreign for humans. These antibodies neutralize a portion of the OKT3 antibodies, allowing CD3+ cells to begin to return to the circulation and thereby allowing each subsequent administration of the antibody to cause, once again, first-dose effects. In the short term, this problem can be overcome by increasing the dose of OKT3 and by administering steroids to prevent the activation side-effects. However, repeated or prolonged treatment with OKT3 is limited by this anti-antibody response.

The anti-interleukin-2 receptor antibody was designed with the view that its use might provide more selective immunosuppression. The interleukin-2 receptor is upregulated on T cells only when their activation has been initiated by contact with foreign antigens. Thus, only a very small fraction of T cells express this receptor at any given time, and those that do at the time of organ transplantation are the T cells most likely to be responsive to the donor's antigens. Thus, anti-interleukin-2 receptor antibodies might be able to eliminate just this population of donor-reactive T cells. Fluorescence-activated scanning has indicated that anti-interleukin-2 receptor-positive cells are not detected during treatment with this antibody, but it is not clear whether these cells are actually eliminated or simply instructed to avoid expression of the receptor. It does not appear that donor-reactive cells are eliminated forever by this approach.

Studies on the use of this antibody in larger animals were not originally encouraging when they were done with a murine protein, and the evidence suggested that this might be due to the antimouse antibody response. As a result, anti-interleukin-2 receptor antibodies were not approved for clinical practice until they had been ‘humanized'. The concept of ‘humanization' is to utilize the mouse amino acid sequence that provides the exact specificity of the desired antibody, but to utilize the human amino acid sequence for all of the rest of the antibody protein. This kind of splicing can now be achieved with genetic engineering techniques. The much greater effectiveness of several humanized anti-interleukin-2 receptor antibodies in clinical practice demonstrates the importance of modifying future antibody-dependent strategies to incorporate all of the potential of benefits of genetic engineering to produce ‘designer' antibodies.

Conclusion

At the present time, the performance of clinical organ transplantation is a highly effective, but not an especially scientific undertaking. However, in a few respects described in this chapter, a knowledge of basic principles of transplantation immunology is important in making the best possible decisions for patient care. However, it is much more clear that further studies of transplantation immunology are essential if the field is to expand by utilizing organs from animal donors or if it is to achieve donor-specific tolerance, which would allow both more successful organ transplantation and the treatment of numerous diseases that could be cured by providing new organs or tissues as long as constant immunosuppression was not required.

Further reading

Auchincloss H, Jr, Sykes M, Sachs DH. Transplantation immunology. In: Paul WE, ed. *Fundamental immunology*, pp. 1175–236. Lippincott-Raven, Philadelphia, 1998. [A comprehensive chapter for the graduate student level.]

Bach FH, Auchincloss H, Jr, ed. *Transplantation immunology*. Wiley-Liss, New York, 1995. [A chapter written for surgical residents.]

Bach FH, Widmer MB, Bach ML, Klein J. Serologically defined and lymphocyte-defined components of the major histocompatibility complex in the mouse. *Journal of Experimental Medicine* 1972; **136**: 1430. [One of the early descriptions of class I versus class II MHC antigens.]

Bevan MJ. High determinant density may explain the phenomenon of alloreactivity. *Immunology Today* 1984; **5**: 128–30. [An article providing a theory to explain the exceptional strength of T cell mediated alloreactivity.]

Billingham RE, Brent L, Medawar PB. ‘Actively acquired tolerance' of foreign cells. *Nature* 1953; **172**: 603. [The report of tolerance induction studies that led to a Nobel Prize.]

Bjorkman PJ, Saper MA, Samraovi B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* 1987; **329**: 506–12. [The initial report of the crystallographic description of a MHC molecule.]

Carrel A. Transplantation in mass of the kidneys. *Journal of Experimental Medicine* 1908; **10**: 98. [A very early description of a technical approach used in kidney transplantation.]

Charlton B, Auchincloss H, Jr, Fathman CG. Mechanisms of transplantation tolerance. *Annual Review of Immunology* 1994; **12**: 707–34. [A review describing mechanisms of transplantation tolerance.]

Cobbold SP, Adams E, Marshall SE, Davies JD, Waldmann H. Mechanisms of peripheral tolerance and suppression induced by monoclonal antibodies to CD4 and CD8. *Immunological Reviews* 1996; **149**: 5–34. [An article describing mechanisms of peripheral tolerance induction including suppressors.]

Colvin RB. The renal allograft biopsy. *Kidney International* 1996; **50**: 1069–82. [An overview of the pathology obtained from renal allografts.]

Cosimi AB *et al.* Treatment of acute renal allograft rejection with OKT3 monoclonal antibody. *Transplantation* 1981; **32**: 535–9. [An early description of the use of a monoclonal antibody for treatment of rejection.]

Doherty PC, Zinkernagel RM. H-2 compatibility is required for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *Journal of Experimental Medicine* 1975; **141**: 502. [One of the early articles demonstrating MHC restriction of T cell responses.]

Gould DS, Auchincloss H, Jr. Direct and indirect recognition: the role of MHC antigens in graft rejection. *Immunology Today* 1999; **20**: 77–82. [A discussion of the relative importance of direct and indirect pathways in graft rejection.]

Jeannet M, Pinn V, Flax M, Winn HJ, Russell PS. Humoral antibodies in renal allotransplantation in man. *New England Journal of Medicine* 1970; **282**: 111–17. [An article demonstrating the correlation between humoral antibodies and chronic rejection in kidney transplantation.]

Jenkins MK, Schwartz RH. Antigen presentation by chemically modified splenocytes induces antigen-specific T cell unresponsiveness *in vitro* and *in vivo. Journal of Experimental Medicine* 1987; **165**: 302. [A study demonstrating the development of anergy.]

Kirk AD *et al.* CTLA4-Ig and anti-CD40 ligand prevent renal allograft rejection in primates. *Proceedings of the National Academy of Sciences (U S A)* 1997; **94**: 8789–94. [A description of the combined effect of two co-stimulatory blocking agents.]

Lafferty K, Prowse S, Simeonovic C, Warren HS. In: Paul WE, Fathman CG, Metzgar H, ed. *Immunobiology of tissue transplantation—a return to the passenger leucocyte concept*, pp. 143–73. Annual Reviews, Palo Alto, CA, 1983. [An overview of the concept of passenger lymphocytes.]

Lechler R, Batchelor J. Restoration of immunogeneity to passenger cell depleted kidney allografts by the addition of donor strain dendritic cells. *Journal of Experimental Medicine* 1982; **155**: 31–41. [An early article discussing direct versus indirect recognition.]

Lee RS, Grusby MJ, Laufer TM, Colvin R, Glimcher LH, Auchincloss H, Jr. CD8+ effector cells responding to residual class I antigens, with help from CD4+ cells stimulated indirectly, cause rejection of ‘major histocompatibility complex-deficient' skin grafts. *Transplantation* 1997; **63**: 1123–33. [An article discussing elements of the interaction between direct and indirect pathways in graft rejection.]

Little CC. A possible mendelian explanation for a type of inheritance apparently non-Mendelian in nature. *Science* 1914; **40**: 904–6. [An article describing Little's early insight into the genetic basis of graft rejection.]

Little CC. The genetics of tumor transplantation. In: Snell GD, ed. *Biology of the laboratory mouse* pp. 279–309. Dover, New York, 1941. [The original description of the laws of transplantation.]

Mason DW, Dallman MJ, Arthur RP, Morris PJ. Mechanisms of allograft rejection: the roles of cytotoxic T-cells and delayed-type hypersensitivity. *Immunological Reviews* 1984; **77**: 177. [A review discussing different effector mechanisms.]

Matzinger P, Bevan MJ. Why do so many lymphocytes respond to major histocompatibility complex antigens? *Cellular Immunology* 1977; **29**: 1. [An article discussing a possible theory to explain the strength of alloreactivity.]

Medawar PB. The immunology of transplantation. *The Harvey Lectures (1956–1957)* 1958; **52**: 144. [A famous lecture describing the Nobel Laureate's early experiments.]

Merrill JP, Murray JE, Harrison JH, Guild WR. Successful homotransplantation of human kidney between identical twins. *Journal of the American Medical Association* 1956; **160**: 277–82. [A description of the first successful human kidney transplant.]

Morris PJ, ed. *Tissue transplantation*. Churchill-Livingstone, Edinburgh, 1982. [An important early textbook in transplantation immunology.]

Pescovitz MD *et al.* Effect of class II antigen matching of renal allograft survival in miniature swine. *Journal of Experimental Medicine* 1984; **160**: 1495–508. [An article describing the particular importance of MHC class II antigens.]

Roopenian DC. What are minor histocompatibility loci? A new look at an old question. *Immunology Today* 1992; **13**: 7–10. [An article discussing the nature of minor histocompatibility antigens.]

Rosenberg AS, Mizuochi T, Sharrow SO, Singer A. Phenotype, specificity, and function of T cell subsets and T cell interactions involved in skin allograft rejection. *Journal of Experimental Medicine* 1987; **165**: 1296. [An article describing experiments to define the role of different T cell subsets in skin graft rejection.]

Rosenberg AS, Singer A. Evidence that the effector mechanism of skin allograft rejection is antigen-specific. *Proceedings of the National Academy of Sciences (U S A)* 1988; **85**: 7739. [A study demonstrating the selectivity of allograft rejection.]

Sayegh MH, Perico N, Imberti O, Hancock WW, Carpenter CB, Remuzzi G. Thymic recognition of class II MHC allopeptides induces donor specific unresponsiveness to renal allografts. *Transplantation* 1993; **56**: 461. [A study describing the induction of tolerance using peptides from MHC molecules introduced into the thymus.]

Singer A, Munitz TI, Golding H, Rosenberg AS, Mizuochi T. Recognition requirements for the activation, differentiation, and function of T-helper cells specific for class I MHC alloantigens. *Immunological Reviews* 1987; **98**: 143–70. [An article describing experiments to determine the role of selective T cell subsets in graft rejection.]

Sprent J, Lo D, Gao EK, Ron Y. T cell selection in the thymus. *Immunological Reviews* 1988; **101**: 173–90. [A review describing the processes of T cell selection in the thymus.]

Strom TB *et al.* The Th1/Th2 paradigm and the allograft response. *Current Opinion in Immunology* 1996; **8**: 688–93. [An overview of the question of Th1 and Th2 T cells in graft rejection.]

Sykes M, Sheard MA, Sachs DH. Effects of T cell depletion in radiation bone marrow chimeras II. Requirement for allogeneic T cells in the reconstituting bone marrow inoculum for subsequent resistance to breaking of tolerance. *Journal of Experimental Medicine* 1988; **168**: 661–73. [An article describing experiments demonstrating the deletional lmechanism of tolerance induction.]

Wallny H, Rammensee H. Identification of classical minor histocompatibility antigen as cell-derived peptide. *Nature* 1990; **343**: 275–8. [An article identifying one of the minor histocompatiblity antigens.]

Williams GM, Hume D, Hudson R, Morris P, Kano K, Milgrom F. Hyperacute renal homograft rejection in man. *New England Journal of Medicine* 1968; **279**: 611–18. [An early description of the process of hyperacute rejection.]

16.2 Organ procurement

Francis L. Delmonico

Introduction
The suitable cadaver organ donor
Brain death
Non-heart-beating donation
Age and medical restriction
Age
Malignancy
Poison as a cause of donor death
Infectious contraindications
Laboratory assessment
Social considerations
Procedure from donor referral to organ recovery
Operative events
Sequence of excision
Organ preservation
Further reading

Introduction

The technical success of organ transplantation has evolved in parallel with a surgical expertise that allows multiple organs to be obtained from a single cadaver donor. The procedure for multiple organ procurement widely adopted by transplant surgeons throughout the world now permits more recent innovations that include splitting the liver into two transplantable components. Aside from the surgical advances however, the field of organ procurement has become well established as the career interest of many medical personnel, functioning as medical directors of organ procurement organizations or as co-ordinators of the organ donation process. This donation process now routinely entails a sequence of patient assessment and operative procedures, which include determination of donor suitability, hemodynamic management prior to procurement, orchestration of operative events, and the preservation of the various organs.

The suitable cadaver organ donor

Most organs for transplantation are procured from cadavers following the diagnosis of brain death. This diagnosis has facilitated the procurement of viable organs, free of the warm ischaemic injury seen when their removal follows the arrest of circulation.

Brain death

In October 1967, Dean Robert Ebert of the Harvard Medical School convened a meeting of physicians, (including Joseph Murray, Raymond Adams, and William Sweet), an ethicist, and a legal scholar, chaired by Henry Knowles Beecher (a Professor of Anesthesiology), to examine the characteristics of a permanently non-functioning brain. Several years earlier French clinicians had considered such a condition that was ‘beyond coma', exhibited by patients with non-viable cerebral hemispheres and a non-viable brainstem. However, Beecher's group sought to define such patients as dead, despite the presence of a heart beat. His Ad Hoc Committee did so by the following criteria: (i) unreceptivity and unresponsivity; (ii) no movements or breathing; (iii) no reflexes; and (iv) a flat electroencephalogram.

Establishing a cause of coma (the President's Commission)

The criteria established by the Harvard Committee were later given a corroborative authority, by a report of medical consultants on the diagnosis of death to the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research. Their guidelines supported physicians diagnosing death by the irreversible cessation of either cardiorespiratory function, or brain function, independent of ventilator-derived cardiorespiratory activity. This report was precise in its application of apnea testing and it alerted physicians to be cautious in declaring brain death in children less than 5 years old (see below). A more profound contribution, however, was to emphasize the necessity of determining the cause of the coma. Although this diagnostic element was impliedin the Harvard formulation, in that all of the cases considered bythe Harvard Committee suffered hypoxic and ischemic injury to the brain (personal communication: R. Adams, MD), it was more explicitly stated in the report to the President's Commission. Defining the etiology of the coma provides an accounting for the findings of the clinical examination, it validates an ‘irreversible' condition, and thus, it assures a diagnosis of brain death.

The President's Commission subsequently joined the American Bar Association, the American Medical Association, and the National Conference of Commissioners on Uniform State Laws to propose in every jurisdiction of the United States the adoption of the Uniform Determination of Death Act (**UDDA**). The UDDA has provided a legal foundation that death can be pronounced in the event of either (i) the irreversible cessation of circulatory and respiratory function, or (ii) the irreversible cessation of all functions of the entire brain, including the brainstem. By endorsing a definition of death that diagnostically evaluates the function of the brain, the UDDA emphasized a reality of death, independent of (mechanically supported) circulatory and respiratory function. The UDDA has clearly proclaimed that brain death means the death of a person.

The concept of death

Although the two components of the UDDA are mechanistically distinct, there is a unifying principle that enables both aspects of the definition to provide a single concept of death. The irreversible cessation of circulatory and respiratory function inevitably leads to an irreversible cessation of all functions of the brain. Thus, all death constitutes an irreversible loss of brain function.

Dr William H. Sweet originally espoused this principle in editorial remarks accompanying a comprehensive Medical Progress Report on brain death by Peter Black in the *New England Journal of Medicine* in 1978. Dr Sweet noted: ‘It is clear that a person is not dead unless his brain is dead. The time-honoured criteria of stoppage of the heartbeat and circulation are indicative of death only when they persist long enough for the brain to die'. Thus, the irreversible cessation of circulatory and respiratory function (as stipulated by the UDDA) can also provide a readily identifiable sign that brain function has been permanently lost, rendering the person dead.

The criteria of brain death

If one accepts conceptually that the ultimate measure of life resides in the function of the entire brain, then the criteria, upon which an irreversible loss of brain function is determined, become critically important to everyday medical practice. As noted above, a permanent absence of brain function can be deduced by the irreversible cessation of circulatory and respiratory function. However, in those clinical instances in which the circulatory and respiratory functions are maintained by mechanical support, there are criteria that can also establish the permanent loss of brain function and thus, brain death.

American Academy of Neurology guidelines

The American Academy of Neurology (**AAN**) definition of brain death is ‘an irreversible loss of the clinical function of the brain, including the brain stem'. The AAN has promulgated guidelines for the diagnosis of brain death (AAN Practice Parameters for Determining Brain Death in Adults, E. F. Wijdicks, MD). Consistent with the report to the President's Commission necessitating that the cause of the coma be established, the AAN has stipulated several prerequisites: (i) clinical or neuroimaging evidence of an acute central nervous system catastrophe that is compatible with the clinical diagnosis of brain death; (ii) exclusion of complicating medical conditions that may confound clinical assessment (e.g. no severe electrolyte acid–base or endocrine disturbance); (iii) no drug intoxication or poisoning; and (iv) a core temperature less than 32°C (90°F) (Table 1).

The irreversible cessation of the critical functions of the whole brain (cortical and brainstem)
1. Coma or unresponsiveness from an established cause which clearly accounts for the cessation of whole brain function and irreversibility
2. Absence of brainstem reflexes (pupil, vestibulo-ocular, corneal, pharyngeal, and tracheal)
3. Apnea as demonstrated by formal apnea testing (requires application of specific methodology)
Exclusions
Complicating medical conditions confounding clinical assessment
(no severe electrolyte acid-base or endocrine disturbance)
Drug intoxication or poisoning
A core temperature < 32°C

*Cessation of cerebral blood flow demonstrated by cerebral angiography, isotope scanning, or transcranial Doppler ultrasonography can validate the brain death diagnosis even when these exclusionary conditions are considered.

Table 1 Guidelines for determination of brain death (American Academy of Neurology)

With these clinical issues resolved, there are three cardinal findings in the AAN pronouncement of brain death: (i) coma or unresponsiveness; (ii) absence of brainstem reflexes (pupil, vestibulo-ocular, corneal, pharyngeal, and tracheal); and (iii) apnea. The brainstem (responsible for spontaneous breathing and vasomotor control) has chemoreceptors that monitor the P_{CO2} and pH of the cerebrospinal fluid, which approximate changes in the plasma. An apnea test was stated to be supportive of the brain death diagnosis if the P_{CO2} rises to more than 60 mmHg and no respiratory movement (abdominal or chest excursion) is detected.

Brain imaging with computed tomography in many instances will reveal the etiology of the condition that has led to the ischemic injury of the brain (hemorrhage, massive infarction). In addition, brain imaging (particularly normal studies) can be suggestive of some conditions that masquerade as brain death (such as pseudocoma or ‘locked-in syndrome') and lead to the improper pronouncement of death when it has not occurred. Also, there may be instances in which a transient but complete circulatory arrest to the brain could briefly present a clinical examination that indicates cessation of brain function. Such clinical cases include a patient with intracerebral or subarachnoid hemorrhage, and the patient who is successfully resuscitated after cardiac arrest. A later examination after an interval of time (6 to 24 h) would be necessary to establish the irreversibility of these conditions if other objective tests of cerebral blood flow (as noted below) were not performed. An interval reassessment also enables a sufficient time to rule out a drug overdose by toxic screening of a blood sample.

Other than brain imaging with computed tomography, confirmatory laboratory testing of brain death has included contrast or isotope angiography, isotope scanning, or transcranial Doppler ultrasonography. The objective of these radiographic techniques is to establish an absence of cerebral blood flow. If such a determination is made, the criteria of brain death (coma or unresponsiveness; absence of brainstem reflexes; and apnea) may be observed. However, the cause of the coma must still be established to declare death.

An isoelectric electroencephalogram remains a useful adjunct to the clinical criteria of brain death; however, it is not mandatory in adult patients and it may be omitted if there is no sustained blood circulation to the brain (as demonstrated by radiologic flow studies).

Otherwise, objective evaluation by confirmatory testing of an electroencephalogram is usually recommended in children. The resiliency of a child's brain underscores the necessity of two clinical examinations, apnea tests, and laboratory confirmation studies 24 to 48 h apart depending upon the age of the child. Moreover, it necessitates an evaluation of the whole brain. An absence of brainstem reflexes may be observed in children with meningitis who have intact cerebral function.

Barbiturates are frequently used following acute neurologic injury to reduce intracranial pressure and high blood levels of barbiturates may alter the pattern of the electroencephalogram and interfere with an accurate determination of brain death. However, if radiologic studies confirm an absence of blood flow to the brain, a diagnosis of brain death can be rendered even in children, irrespective of barbiturate blood levels. Cessation of carotid circulation at the base of the skull affects not only the cerebrum, but also the brainstem. Thus, the determination of death in this circumstance adheres to a whole brain concept of brain death.

Brain death by assessment of brainstem function only

Pallis has written that ‘the key functions which define a human being as an independent biological unit (the capacity to be conscious and the capacity to breathe) are subserved by the brainstem'. As a result, Pallis asserts that the diagnosis of brain death can be made by clinical examination of the brainstem reflexes in a comatose individual with apnea. This approach adopted by many physicians in the United Kingdom precludes an evaluation of the cerebrum in establishing brain death (Table 2).

1. Have the preconditions been met?
Is the patient comatose and on a ventilator?
Is there a positive diagnosis of structural brain damage?
Have all possible attempts been made to remedy it?
2. Has adequate attention been given to the necessary 'exclusions' (drug intoxication, primary hypothermia, major metabolic disturbances)?
3. The patient has been on the ventilator for several hours.
Have each of two clinical examinations revealed:
Absent brainstem reflexes?
Persistent apnoea (disconnection test)?
4. If so, the patient can be declared dead, even if the heart is still beating.
5. Once the patient has been declared dead the respirator is ventilating a cadaver.
It is no longer a 'life support system'.
6. The patient is dead when the brainstem is declared dead, not when the cadaver is disconnected from the respirator and the heart stops beating.
7. Neither electroencephalography nor angiography are necessary for a diagnosis of brainstem death.

Table 2 Declaration of brainstem death

On the other hand, the need to assess brainstem activity for the diagnosis of brain death has precluded its application to anencephalic newborn infants, and to some irreversibly comatose individuals: these patient groups are both devoid of cerebral cortex activity, but show spontaneous breathing, which requires intact brainstem function. Brainstem function may also be recognized clinically when pupillary light, corneal, oculocephalic, and oropharyngeal reflexes are present. An active debate has been waged as to whether brainstem function alone, irrespective of cerebral function, should imply the presence of ‘life', and this approach has gained attention following the use of anencephalic donors of hearts and kidneys. At this time however, international society has not accepted a diagnosis of death which disregards brainstem function, that is a capacity to breath spontaneously, and the procurement of organs from anencephalics prior to death has not been sanctioned.

Non-heart-beating donation

As a significant shortage of cadaver organs for transplantation persists, the opportunity for recovering organs from the non-heart-beating donor (**NHBD**, i.e. after a declaration of death by an absence of heart beat and respiratory activity) is being reconsidered more broadly. With the advent of health care proxies, advance directives, and a societal reluctance for extraordinary measures of medical treatment for the terminally ill, an increasing number of NHBDs could become available for organ recovery. Clinical conditions are being recognized in which non-survivable brain injury may be determined, without fulfilling the criteria of brain death.

As with the diagnosis of brain death, the cause of the irreparable brain injury must be determined. With the family's approval (and often at their request), a joint decision with the physician may be made to withdraw life-sustaining support as appropriate care for the dying patient. This practice has become common and it has evolved to a public acceptance, irrespective of organ donation. However, the continuing process of deciding to withdraw terminal care and to consenting subsequently to organ donation after death, may also provide the family with an important consolation at the time of bereavement. Therefore, in certain instances of withdrawal of life support, consideration of organ donation has also become appropriate. D'Alessandro has referred to these clinical situations of carefully orchestrated organ recovery at the time of life support withdrawal as ‘controlled' NHBD, in comparison with ‘uncontrolled' NHBD, which may evolve in the setting of the emergency ward following traumatic death.

Ethical issues of NHBD

The 1997 CBS news magazine *60 Minutes* report on NHBD created an impression that doctors were either taking organs from living people or hastening their deaths in

order to recover organs. The misleading premise of this report was not only false, but it was undermining of the public trust. The controversy of this *60 Minutes* show resulted in the establishment of a Committee on Medical and Ethical Issues in Maintaining the Viability of Organs for Transplantation, by the Institute of Medicine (**IOM**) of the National Academy of Sciences. This IOM Committee was convened in July 1997 by John T. Potts, MD of the Massachusetts General Hospital, who served as Committee Chairman. The IOM Committee heard testimony from leaders of the organ procurement and transplant communities, as well as groups of donor families over a 2-day period. The ethical aspects of several NHBD protocols were examined. The IOM subsequently published an impressive report in December 1997, concluding the NHBD process to be ethically proper and medically effective.

However, there are ethical issues that will require further consideration. These include the duration of asystole that leads to a determination of death, premortem cannulation, the location of the withdrawal of care, and the administration of morphine, regitine, and heparin. Let us examine each of these issues, reflecting upon their importance in accomplishing successful NHBD loss of heartbeat.

Under the law, death must be pronounced at a precise time as the ceasing of a person to exist. However, organs and tissues obviously can remain viable for an extended period after death is declared; otherwise, they could not be transplanted successfully from NHBDs. The central component of life resides in the function of the brain. Thus, the duration of time that the brain can withstand an absence of blood circulation and maintain viability is fundamental to the time of asystole that should elapse before an NHBD is determined to be dead. Although it might be considered arbitrary by some, an asystolic duration of 5 min as proposed by the IOM, could be acceptable to declare death. A total absence of circulation to the brain for 5 min after the events of care withdrawal would be sufficient to conclude that whole brain function was irreversibly lost (see IOM arguments further below). Perhaps every neuron may not have undergone necrosis during that period of time, but the capacity of the brain to sustain its critical functions has passed; and of course, the clinical setting is the withdrawal of care from a patient with a non-recoverable condition.

Thus, the most controversial NHBD protocol issue is this period of asystole that assures that the patient is dead. The 5-min duration of asystole after a declaration of death, as promulgated by the IOM, was done with a background of center protocol variability. For example, Kootstra has suggested that 10 min should elapse before the installation of organ preservation devices, so that there is a clear transition from the care of a patient to the recovery of organs from a cadaver. However, the 10 min period of asystole may prohibit the recovery of organs other than the kidneys. In contrast, the University of Pittsburgh protocol permits the declaration of death 2 min following the cessation of cardiorespiratory activity, providing an opportunity for multiple organ procurement.

The most vocal opposition to NHBD has been forthcoming from those who do not accept the 5-min period of asystole as a sufficient duration to permit a declaration of death. They approach this dispute recounting supportively a unitary concept of death as the irreversible loss of whole brain function. For example, Menikoff has challenged the IOM report, suggesting that ‘the amount of time that passes after the heart stops should be 5 or 10 min longer (than the current IOM recommended time of 5 min), to be certain that the entire brain function has irreversibly ceased functioning'. Otherwise, Menikoff concludes that the dead donor rule would be violated.

The IOM has responded that whole brain death has probably occurred in ‘the cumulative minutes of: (i) after the respirator is stopped until the heart stops; (ii) after the heart stops, 5 more minutes until death is pronounced; and (iii) after the additional time that elapses during the transport from the hospital bed to the operating room, or in the preparation for organ removal even if death is pronounced in the operating room'. Thus, the IOM contends that the evolving clinical scenario as described does not violate the dead donor rule.

Finally, the determination of asystole should be given consistently. The protocol definition of asystole could vary from an absence of a palpable pulse, to electromechanical dissociation, to a flat-line tracing on the cardiac monitor. Irrespective of the method of determining asystole, the physician declaring death should not be a member of the transplant service.

Predicting death after withdrawal of care

By monitoring the ventilatory parameters of tidal volume, vital capacity, and negative inspiratory force, Edwards and Nathan have suggested that a premortem prediction of the time of death can be given (personal communication: John M. Edwards, Delaware Valley Organ Procurement Organization). If the patient cannot generate a negative inspiratory force of greater than 20 cmH_2O, it is likely that the patient will expire within 1 h following the withdrawal of care. This kind of premortem evaluation can be very helpful in determining whether the organ procurement organization should assemble an NHBD team at the donor hospital. However, if 1 h elapses before a declaration of death is made, the organs will have lost sufficient viability to permit successful transplantation. In that circumstance, D'Alessandro has recommended that the organ procurement organization team abandon NHBD efforts (personal communication: A. D'Alessandro, MD).

Violation of the dead donor rule by premortem cannulation

A principle of organ donation is never to violate the dead donor rule by initiating invasive measures intended for organ recovery prior to a determination of death. Thus, unless the patient has been declared dead, the use of premortem cannulation for infusion of preservation fluids in my view is counterproductive to a consistent approach to organ donation. If this procedure is altered by premortem cannulation, the IOM requires special informed consent of the family and the administration of local anesthesia at the site of cannulation. Thus, the importance of this consistency is to not endanger the public trust of the organ donation process. Otherwise, the donation benefit realized is small, as the cannulas can quickly be inserted after death.

Location of withdrawal of care

Should death be declared in the intensive care unit permitting public accountability, or may death be determined in the operating room, enabling the rapid recovery of organs? Withdrawal of life support in the intensive care unit avoids the concern that death might be hastened in the operating room by the transplant team to achieve organ recovery. However, determining death in the operating room affords the most expeditious organ recovery after death. A solution to this aspect of NHBD policy would perhaps be reached were it not for the dilemma described previously: the inability to predict when a patient will stop breathing after the life-support measures are withdrawn. Edwards and Nathan have applied the negative inspiratory force test successfully as noted above. However, in their experience, 10 per cent of potential NHBD have returned from the operating room (still alive) to the intensive care unit because death could not be declared within the 1-h window of organ viability (personal communication: John M. Edwards). Therefore, the uncertainty of this awkward predicament is the principal reason that some advocate the patient remain in the intensive care unit until death is determined, before transporting the patient to the operating room for organ recovery.

There should be no misperception by a family that withdrawal of life support might be intended by the medical staff only as a means for organ recovery. The principal concern of the physician must be the proper care of the dying patient, and there must be a clear separation of the objectives for the withdrawal of life support and the potential of organ donation; otherwise, some physicians have claimed incorrectly that NHBD is ‘tantamount to euthanasia'. Moreover, it must also be clear in the medical and public communities that the recovery of organs may only take place after the patient is declared dead.

Administration of morphine, regitine, and heparin

Once the decision to withdraw support has been made (independent of any decision about donation), narcotic medications such as morphine may be administered to ensure the dying patient experiences no discomfort or awareness of dyspnea. This practice occurs routinely when life support is withdrawn, whether NHBD is feasible or not. The ethical rule of double effect, which has been widely accepted, applies in this clinical situation. Though morphine may hasten death by suppressing respiration, the intent of its administration is to relieve suffering and provide comfort. Since the objective of the morphine treatment is not to expedite death, the goal of assuring patient comfort assumes primacy as a standard of proper care.

The regitine and heparin are given at the time of NHBD death to curtail vessel spasm and to prevent the formation of clot within potentially viable organs. These agents are not given as treatment because a concerted decision has already been made by the primary care team and the family to discontinue any further treatment. They are not intended to hasten death, as death is imminent irrespective of the heparin or regitine.

Outcome of allografts transplanted from NHBD

Cho *et al.* have recently reported a 1-year outcome of 83 per cent for NHBD kidneys compared with an 86 per cent rate of success from kidneys recovered from brain dead donors. The opportunity for successful transplantation was increased by NHBD who had died of traumatic injury. D'Alessandro has reported the successful recovery of 14 pancreases, 13 livers, and a single lung over a 5-year period. The patient and graft survival of the kidney/pancreas transplants was 92.9 per cent (mean follow-up of 26 months). The lung transplant recipient was extubated at 4 days following transplantation. There were two (15.4 per cent) episodes of primary liver allograft non-function.

Uncontrolled NHBD

The policy and practice of physicians in achieving NHBD must withstand full public scrutiny. Therefore, community education about organ donation should include a

discussion of the NHBD procedure. Otherwise, mistrust by society about medical protocols such as NHBD, although technically proper, will endanger the credibility of all organ donation. This assertion is especially relevant to the opportunity of NHBD arising in the emergency ward. Placement of preservation cannulas and/or the initiation of preservation fluids prior to consent from appropriate family may be acceptable, particularly if the person who is dead on arrival or dies in the emergency ward is carrying a donor card. However, the proponents of this approach must be sensitive to a society that is perhaps skeptical that every measure of medical technology was provided to save a life. Nevertheless, with that assurance thoughtfully conveyed to a family, the possibility for NHBD could certainly be expanded.

Age and medical restriction

A detailed social and medical history should be taken from the donor's family by the representative of the organ procurement organization (either physician or co-ordinator) who obtains consent for donation. In addition to being aware of the donor history, procuring surgeons should perform an appropriate examination which might bear upon donor suitability. This includes looking for extremity lacerations that have become secondarily infected, or needle marks indicative of drug abuse, and intraoperative examination of the viscera.

Age

A donor of up to 80 years of age must be considered acceptable for multiple organ donation. We should note, however, that depending upon the donor age, some organs may not be recoverable for transplantation (Table 3). For example, infant donors less than 6 months of age are likely to provide only a heart allograft for transplantation. A donor less than 3 years of age will provide kidneys retained as an *en bloc* preparation for a single recipient. Donors more than 60 years of age are not likely to be suitable for heart or lung recovery, and kidneys that are procured from such older donors are usually assessed by a postrecovery frozen-section biopsy.

Kidney	3–80 years (for single kidney)
Liver	Newborn–80 years
Pancreas	8–55 years
Heart	Newborn–40 years (male)
	Newborn–45 years (female)
Heart/lung	2–35 years
Lung	10–50 years

With no history of pre-existing organ disease.

Table 3 Age restrictions to the multiple organ donor

Malignancy

The risk of transmitting a tumor from an organ donor with a known malignancy is significant (approximately 50 per cent), especially depending upon the type of malignancy. The following cancers have been transplanted to allograft recipients from organ donors: thyroid, bronchial, lung, breast, adenocarcinoma (colorectal and unknown primary), kidney, melanoma, anaplastic, prostatic, and choriocarcinoma. However, cadaver donors with primary brain tumors, a history of non-melanoma skin cancers such as basal or squamous cell, or a history of *in situ* carcinoma of the cervix may be acceptable organ donors. Although none of these tumors has been transmitted to an allograft recipient in my own personal experience, they warrant additional discussion.

Primary brain tumors

Although the spontaneous extracranial spread of a primary brain tumor is rare, presumably because of the blood–brain barrier, not all brain malignancies carry the same minimal risk of metastasis. Moreover, the possibility of metastasis may be influenced by intracranial operative procedures such as a craniotomy, or the insertion of a ventricular shunt. Primary brain tumors have been transmitted from patients with a ventricular shunt; however, the absence of a shunt does not preclude the possibility of the transmission of a brain tumor to a recipient of a solid-organ allograft. If organs are to be recovered from a donor with a primary brain tumor, the thoracic and abdominal cavities must be carefully inspected for metastases. The diagnosis of a primary brain malignancy must also be differentiated from extracranial malignancies such as melanoma, which may have metastasized to the brain. Intracerebral hemorrhage may be the presentation of not only a metastatic melanoma but also a renal cell carcinoma or a choriocarcinoma. Female potential donors with non-traumatic cerebral hemorrhage thought secondary to either a primary brain malignancy or a spontaneous subarachnoid hemorrhage should have a blood level of the human chorionic gonadotropin (**hCG**) determined. If a woman of reproductive age has a measurable blood hCG level, the patient should not be considered as an appropriate donor, unless there is clear evidence (post-mortem) that the patient had either an intrauterine or tubal pregnancy, and the pathology of that gestation does not reveal a choriocarcinoma (personal communication: Ross Berkowitz, MD, Boston, MA).

In situ carcinoma of the cervix

Cervical intraepithelial neoplasia is a dysplastic lesion and a precursor of malignancy that may be suspected by Pap smear, but requires a colposcopic biopsy for definitive staging and management. An absence of invasive cancer permits unrestricted organ donation.

Non-melanoma cancers of the skin

Basal cell and squamous cell carcinomas are very frequent in the general population; however, metastasis to the brain, lung, or liver is rare. If metastases occur, they are usually detected within 6 months of initial treatment. Thus, a careful medical history and visceral examination at the time of organ recovery will establish the safety of donation to prospective allograft recipients when a skin cancer history is given by a donor family or by a medical record.

Donor malignancies may be detected in allograft recipients as early as hours after transplantation (noted in allograft nephrectomy and liver allograft biopsy specimens). Otherwise, metastasis may be evident within 3 months following transplantation (choriocarcinoma metastasized to the lung of a liver allograft recipient) or they may not become evident until more than 3 years following transplantation (the development of melanoma on the chest, arms, and thighs of a renal allograft recipient). The tumors can also grow within the allograft and/or locally adjacent to the allograft: for example, a donor bronchial carcinoma presenting as a large mass adjacent to a renal allograft.

Extrarenal organ recipients are susceptible to the risk of donor-transmitted malignancy, even though the allograft may not be a common site of metastasis: for example, the heart. Presumably theallograft carries the cancer cells and these cells flourish in the immuno-suppressed environment of the recipient. The Penn cancer registry data reveals five heart transplant recipients who developed metastatic malignancy from eight donors with known cancer (personal communication: Israel Penn, MD, Cincinnati, OH). Two of these cases from the Penn registry are informative, as they emphasize that heart transplant recipients are just as vulnerable to the development of donor malignancy as the recipients of other solid organs. First, a 45-year-old cadaver donor with brain hemorrhage underwent a perirecovery liver biopsy which later revealed unsuspected metastatic adenocarcinoma. The heart transplant recipient developed adenocarcinoma metastasis to the spine 6 weeks later. In the second case, four allograft recipients (including a heart recipient who died) developed melanoma from a young cadaver donor who died of brain hemorrhage and unsuspected melanoma. This second case also is revealing in the nature of the melanoma history of the donor, and it underscores the risk of melanoma transmission from potential donors, even though they have not suffered metastatic disease. The donor history retrospectively revealed only ‘benign mole surgery' and there was no evidence of recurrent or metastatic disease at the time of organ donation. In retrospect, a review of the lesion by the Armed Forces Institute of Pathology determined the mole lesion to be a melanoma, but it was not a unanimous decision, as the pathology team was blinded from the donor history and recipient events. We should also note that pathologists may differ regarding even the diagnosis of a thin nevus to be melanoma. Moreover, there have been rare instances in which a thin melanoma (less than 0.5 mm) has metastasized in a non-immunosuppressed patient without other histologic risk factors (personal communication: Martin C. Mihm, MD, Albany, NY). Thus, a donor history of mole surgery by a family member necessitates a careful review of the donor medical record. The diagnosis of melanoma is an absolute contraindication to organ donation, as the risk of transmission of melanoma from either a cadaver or living organ is very high (80 per cent).

Poison as a cause of donor death

Accidental or suicidal poisonings through inhalation, ingestion, or injection of toxic substances result in several thousand fatalities each year. Barbiturate and

benzodiazepine medications are frequently overdosed by suicide victims. Although a majority of these deaths occur following a cardiopulmonary arrest, some patients are resuscitated for a period of time, only to become brain dead. These patients may be considered for organ donation. Table 4 contains a list of organs that have been recovered from donors who died from various poisons.

Cause of death	*Organs successfully transplanted*
Acetaminophen	Heart
Barbiturates	Kidneys
Benzodiazepine	Heart, liver, kidneys
Carbon monoxide	Liver, kidneys
Cyanide	Kidneys
Imipramine	Liver, kidneys
Methanol	Heart, liver, kidneys
Methaqualone	Kidneys
Warfarin (rat poison)	Heart, lungs, liver, pancreas, kidneys

Table 4 A list of poisons as a cause of donor death that did not prevent the successful transplantation of organs (see reference O'Connor and Delmonico).

Each toxic exposure must be evaluated with respect to specific organ injury. Certain drugs such as acetaminophen may render the liver unsuitable for transplantation because of fulminate hepatic necrosis; but other organs may not be injured. The New England Organ Bank (NEOB) has successfully recovered a heart allograft for transplantation from a cadaver donor with acetaminophen toxicity.

Tricyclic antidepressants may cause donor liver injury, and depending upon the serum concentration of the tricyclic, adverse effects not only for donor organs, but also for recipients of organs from a donor with antidepressant medication overdose. A tricyclic-contaminated liver may become the source of a recipient tricyclic blood level that is toxic to the recipient's mental status and cardiac stability. Nevertheless, Fattinger has reported the successful transplantation of kidneys from a patient who died from a trimipramine overdose, and she has suggested that the transplantation of the liver might also be feasible from such a donor if the tricyclic level in the donor serum falls to an acceptable level.

Carbon monoxide expelled by gasoline engines is a common cause of inhalation poisoning. Accidental carbon monoxide poisoning can also occur following exposure to solvents in paint removers and by the release of vapors from faulty home heating systems. Death usually occurs from hypoxia prior to arrival at the emergency ward because carbon monoxide rapidly replaces oxygen from the hemoglobin molecule. However, for those resuscitated patients who subsequently develop anoxic brain death, assessment of renal injury by urine sampling (for myoglobin, protein, and red blood cells) is useful in deciding whether to proceed with kidney procurement, as rhabdomyolysis can affect renal function. Depending upon liver function and morphology by biopsy, liver transplantation should also be considered.

Prompt identification of the toxin and treatment to minimize end-organ damage can be critical life-saving measures. However, these measures can also be consequential to successful organ salvage from patients who still succumb to the poison. For example, cyanide causes cellular anoxia within 30 min of ingestion or within seconds of inhalation. Decreased mitochondrial oxidative metabolism and oxygen utilization lead to lactic acidosis. Unless an antidote regimen of amyl nitrate inhalation and intravenous injection of sodium nitrite and sodium thiosulfate are administered promptly, irreversible central nervous system damage may ensue. However, Barkoukis has reported that use of these cyanide antidotes has resulted in resumption of satisfactory organ function in a patient whose brain injury was beyond resolution following cyanide suicide, enabling the successful transplantation of the heart and kidneys. Liver allografts can also be successfully transplanted from donors with fatal cyanide poisoning, despite initial cyanide toxicity to the liver.

Hemodialysis and ethanol administration have been used to correct the metabolic acidosis and stabilize organ function in a patient who died following methanol ingestion, resulting in the successful transplantation of renal allografts. NEOB transplant centers have successfully transplanted the liver and both kidneys from a donor with fatal methanol toxicity.

Multiple organs have been recovered from an NEOB donor who committed suicide by the ingestion of rat poison. The active ingredient of the rat poison which caused the fatal cerebral hemorrhage was sodium warfarin. The donor's prothrombin time peaked at 200 s, but returned to near normal 48 h later, enabling the successful transplantation of the kidneys, heart, lungs, and liver.

These cases demonstrate that the toxic organ injury from a fatal poisoning may be reversible. Thus, if the adverse effect of the poison to specific organ function is resolved, organ recovery and transplantation is appropriate. The function of organs transplanted from donors who have died from poisoning has not been adversely affected.

Infectious contraindications

Each potential donor must be assessed for infectious organisms. Systemic infection transmitted by a cadaveric donor organ can result not only in a loss of the allograft but also in death of the immunosuppressed recipient.

The history and physical examination of a potential cadaveric donor may reveal a systemic infection present prior to or associated with the patient's death: the latter may have been acquired in hospital. The cause of death may promote an infectious complication: victims of drowning often eventually die of pneumonia, and burn victims may develop cutaneous sepsis.

Infections acquired prior to death

Viral infection

The presence of an active viral infection in the form of hepatitis, perineal herpes, encephalitis or meningitis, pneumonia, varicella zoster, or human immunodeficiency virus (**HIV**) infection is an absolute contraindication to organ donation. A history of a specific viral infection does not preclude donation, but it requires the exercise of clinical judgement after a thorough gathering of historical information. Given the impact of donor-transmitted infections upon the outcome of organ transplantation, detection of infection using serologic screening is an important measure of donor suitability. While donor screening for HIV, HTLV-I, hepatitis B, hepatitis C, and cytomegalovirus is routinely performed in the United States (Table 5), there are no worldwide standards. The practices among organ procurement organizations vary widely, as do the interpretation and the reporting of the results. This review will summarize current knowledge about serologic donor screening for solid organ transplantation, and discuss the utility of screening for organ donor suitability as well as the impact of latent donor infections upon recipient outcome.

HIV antibody
HTLV-I antibody
Hepatitis B virus surface antigen (HBsAg)
Hepatitis B virus surface antibody (anti-HBs)
Hepatitis B virus core antibody (anti-HBc)
Hepatitis C virus antibody (HCV)
Cytomegalovirus antibody (CMV)
Treponemal antigen (syphilis)
Toxoplasma antibody

Table 5 Routine serology screening of the organ donor

HIV infection Human immunodeficiency virus type 1 (HIV) has been transmitted from infected cadaver donors to multiple recipients of organs, resulting in the death of the allograft recipients. United Network for Organ Sharing (**UNOS**) policy states that organs from donors with a positive antibody screening to HIV are not suitable, unless subsequent confirmation testing indicates that the screening test was falsely positive. However, a confirmatory Western blot test cannot usually be performed before the maximum time of cadaver organ preservation elapses. Therefore, for all practical purposes, a cadaver donor serum testing positive for HIV by the preliminary enzyme-linked immunosorbent assay results in the discard of all donor organs for transplantation.

In contrast, there is ample time to perform confirmatory testing upon living donors. All living organ donors should undergo voluntary screening for HIV within 2 to 3 months prior to the planned transplant. During that interval, donors should be instructed to avoid societal encounters that might lead to HIV exposure. For example, HIV has been transmitted from an asymptomatic living donor of a renal allograft (whose serology was HIV negative 8 months prior to the transplant) following homosexual activity between the time of the HIV testing and the transplant.

The American Red Cross is now screening blood and tissue donors for HIV p24 antigen (the core structural protein of the virus) by a Food and Drug Administration licensed method. This new test may be available for organ donor screening in the future. It enables HIV antigen detection within 30 days of exposure, in a window period when a patient may have viremia prior to the development of antibodies. The estimated average time between p24 detection and the development of HIV antibody is approximately 7 days; thus, its temporal value may be limited. However, the importance of p24 antigen detection is underscored when the rare situation arises in which the organ donor is negative for HIV antibody by serology. Simonds has reported the transmission of HIV to seven recipients from organs and tissues from a seronegative donor, who had HIV p24 antigen retrospectively identified in cultured spleen tissue. Presumably p24 antigen would have been detected in a current peripheral blood sample. Five of the seven recipients (including the heart, liver, and kidney recipients) died with HIV infection.

Because of the risk of HIV transmission to blood, tissue, and organ transplant recipients, the Center for Disease Control (**CDC**) has developed guidelines for prevention of HIV to organ and tissue recipients (Table 6). The organ procurement organization staff should obtain a social history on all potential organ donors which addresses these criteria. When the social history of the potential donor reveals any of the CDC criteria, the risk of HIV transmission (despite negative serology screening) must be balanced against the benefit contemplated for a particular allograft recipient. Thus, organs may be recovered without restriction for either recipients with life-threatening illnesses (heart and liver), or for those recipients of non-lifesaving transplants (lung, pancreas, kidney, and bowel). In either circumstance however, the CDC guidelines emphasize a requirement for the transplant center to inform recipients of the potential risk of donor HIV infection, even if that risk is presented only by the behavioral social history.

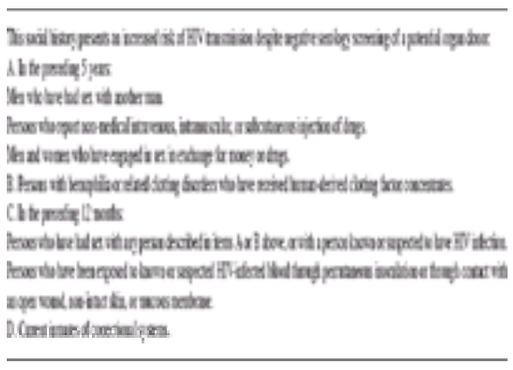

This social history presents an increased risk of HIV transmission despite negative serology screening of a potential organ donor.
A. In the preceding 5 years:
Men who have had sex with another man.
Persons who report non-medical intravenous, intramuscular, or subcutaneous injection of drugs.
Men and women who have engaged in sex in exchange for money or drugs.
B. Persons with hemophilia or related clotting disorders who have received human-derived clotting factor concentrates.
C. In the preceding 12 months:
Persons who have had sex with any person described in items A or B above, or with a person known or suspected to have HIV infection.
Persons who have been exposed to known or suspected HIV-infected blood through percutaneous inoculation or through contact with an open wound, non-intact skin, or mucous membrane.
D. Current inmates of correctional systems.

Table 6 Centers for Disease Control guidelines

Potential donors with a history of incarceration in a correctional institution should be assessed by the usual standards; however, a negative serology for HIV and hepatitis may be sufficient to permit unrestricted donation if 6 consecutive months have elapsed since incarceration. Nevertheless, given the hazard of previous intravenous drug use in this population, some inmates may not be suitable donors because of positive serology screening.

HTLV-I infection Although UNOS policy regarding HTLV-I screening is the same as it is for HIV (potential organ donors are not suitable), not all organ procurement organizations reject a cadaver donor whose serology screening test reveals a positive anti-HTLV-I. The risk of transmission of HTLV-I by solid organ transplantation has not been clearly defined. Human T-lymphotrophic virus type-I/II (**HTLV**) has been transmitted to recipients of contaminated blood transfusions. However, depending upon the emergent nature and life-threatened status of a particular candidate for transplantation (especially heart or liver), a patient may have little alternative but to accept an organ from a donor whose screening for HTLV-I is positive. There is also a possibility that the screening test is falsely positive. The social history may be helpful in discerning whether there is a likelihood of a true positive being subsequently determined by confirmatory testing.

A patient infected with HTLV-I by blood transfusion is at risk for the development of either adult T-cell leukemia or neurologic disorders. Thus, HTLV-I is considered hazardous, and consistent with UNOS policy, many regard a positive donor screening for HTLV-I to be an absolute contraindication to organ donation.

Hepatitis B infection There are several serologic profiles for potential organ donors who have been infected with hepatitis B. Furthermore, the implications of donor infectivity vary according to both the serologic profile and the organ to be donated. The following review is categorized by the possible hepatitis B results reported by an organ procurement organization to the transplant center.

*Hepatitis B surface antigen (**HBsAg**) positive* HBsAg may be identified in the serum of an infected patient within 30 to 60 days of exposure. Hepatitis B virus (**HBV**) transmission has been documented to occur through organ transplantation. Thus, detection of HBsAg in the donor blood sample indicates the donor is at risk for transmitting HBV. However, the organ procurement organization should verify the quality of the serum sample, since a false-positive result can arise from a hemolyzed sample of donor blood.

Liver allografts from HBsAg-positive donors have been transplanted successfully to critically ill HBsAg-negative recipients in life-threatened clinical situations. Renal allografts transplanted from HBsAg-positive donors to HBsAg-negative recipients have also been performed without the development of HBV infection in the recipient; however, the satisfactory clinical course may have been influenced by the presence of previous immunization to HBV. Transplant physicians should recognize that there is a potential for a poor outcome when organs are transplanted from HBsAg-positive donors (deaths from HBV infection) even to HBsAg-positive recipients.

When the HBsAg-positive result is verified by repeat testing of donor blood, organ procurement would probably be considered for heart recipients in a life-threatening emergency status, especially those candidates who are anti-HBs positive (by virtue of immunization or natural immunity).

Isolated anti-HBs (HBsAg and anti-HBc negative) This serologic profile of a donor blood sample could be observed following the vaccination of an individual with hepatitis B vaccine, following the administration of hepatitis B immune globulin (**HBIG**), following the transfusion of a blood product from an immunized donor, or following previous HBV infection.

Antibody to HBsAg, unlike antibody to the HBV core antigen (anti-HBc), does not arise during the acute infection, but rather during convalescence. Thus, up to 6 months may elapse before anti-HBs may be detected following HBV infection. However, anti-HBc is usually detectable soon after exposure to the HBV, and it will usually persist (see below: the now well-accepted importance of detecting anti-HBc).

The donor with isolated anti-HBs is not likely to transmit HBV infection because there is no evidence of active viral replication. However, HBV DNA has been detected by the use of the polymerase chain reaction (**PCR**) in the sera and liver of patients with idiopathic chronic liver disease, whose sera were HBsAg negative, and anti-HBs positive. Moreover, the Mayo Clinic and UCSF (University of California San Francisco) groups in a joint report have observed that a cadaver donor with a serology profile of HBsAg negative, anti-HBs positive, and anti-HBc negative, has transmitted HBV to a liver allograft recipient. They concluded that the sera from such donors may harbor HBV (as detected by PCR-DNA analysis) despite the negative HBsAg. We should also note, however, that this circumstance appears to be rare (less than 1 per cent). Therefore, a donor serology profile of isolated anti-HBs positive does not appear to be a contraindication by itself to the recovery of all organs from such a donor, especially if the medical history of the donor reveals hepatitis B vaccination or the administration of HBIG. However, if no such medical history is determined, there is a rare possibility of HBV transmission from a cadaver donor to a liver allograft recipient with this serologic profile.

Anti-HBs positive, anti-HBc positive, HBsAg negative Anti-HBc (IgM) is the earliest antibody detected following HBV infection (10 to 14 days). Anti-HBc (IgG) can persist for the lifetime of a previously infected patient. Although it has been generally accepted that a serologic profile of HBsAg negative, anti-HBs and anti-HBc

positive reveals recovery and immunity to HBV infection, HBV may still reside in the donor patient's liver. Recent literature reports have emphasized the importance of anti-HBc detection regardless of the anti-HBs result. Although the presence of the anti-HBs may suggest donor immunity to HBV, the detection of anti-HBs does not necessarily reduce the hazard of HBV transmission from donors who are anti-HBc positive, especially among liver allograft recipients. It is now well established that HBV has been transmitted to liver allograft recipients from cadaver donors whose blood sample revealed a serologic profile of anti-HBc and anti-HBs positive (negative HBsAg). The liver continues to harbor HBV (personal communication: Jay Hoofnagle, MD, National Institutes of Health, Bethesda, MD) and following transplantation, HBV can replicate in the newly immunosuppressed environment.

Douglas has recently reviewed the long-term experience with *de novo* HBV infection transmitted from an organ donor, and he has suggested that the clinical course may be relatively mild. It also appears that the risk of HBV transmission from isolated anti-HBc-positive donors is predominantly for the liver allograft recipient. This serologic profile may not carry the same risks for the recipients of other organs. In the UCSF review, only one of 42 kidney and none of seven heart allograft recipients became HBsAg positive following transplantation of organs from isolated anti-HBc-positive donors; whereas, three of six liver recipients developed HBsAg positivity following transplantation from anti-HBc-positive donors. Mount Sinai Medical Center and the New York Regional Transplant Program have also recently analyzed the outcome of 52 HBsAg-negative recipients of allografts from 21 anti-HBc-positive donors. Although four of 16 liver allograft recipients became HBsAg positive, none of the kidney or heart recipient serologies became positive.

Isolated Anti-HBc positive (HBsAg and anti-HBs negative) A profile of isolated anti-HBc positivity may reflect the window of time in which anti-HBs has yet to be serologically detectable in an HBV-infected individual. An insufficient period has elapsed for the anti-HBs to become positive. Anti-HBc may be the only marker indicating current HBV infection, as the HBsAg has decreased in the peripheral blood to levels which are no longer detectable, and the anti-HBs has yet to increase in sufficient titer as to be detectable. The anti-HBc determination is not influenced by exposure to HBV vaccine.

This serology profile is a potential concern for all allograft recipients. HBsAg-negative anti-HBcAb-positive donors can transmit HBV to renal allograft recipients as noted recently by Satterthwaite; thus, anti-HBc-positive donors must be cautiously considered. J. Miller and his associates have proposed testing donor kidney tissue and serum by PCR analysis on HBsAg-negative anti-HBcAb-positive donors, further clarifying the risk of HBV transmission for renal allograft recipients.

Donor liver functions are routinely performed to determine organ suitability, but these may be abnormal because of hemodynamic consequences. Therefore, a pretransplant liver biopsy should be performed on all anti-HBc-positive donors to look for portal piecemeal necrosis consistent with active hepatitis. If the biopsy is positive for hepatitis, a true positive anti-HBc is likely, and the risk of HBV transmission is great.

Some organ procurement organizations do not assay for anti-HBs in the routine serologic assessment of a prospective donor (personal communication: Charles Miller, MD, Mount Sinai Hospital, NY), and rely only on the anti-HBc determination. If the HBsAg is negative, and the anti-HBc is negative, there are no restrictions to organ recovery. If the serologic donor profile reveals HBsAg negative, anti-HBc positive, the anti-HBc positive result is further defined by testing for IgM compared with IgG. IgM-positive donor organs would probably not be recovered because the IgM-positive result indicates recent HBV exposure. However, IgG-positive donor organs may be recovered for life-threatened HBsAg-negative recipients, or recipients who are HBsAg positive. In either of the latter IgG-positive circumstances, Dr Miller treats his recipients with prophylactic HBIG following transplantation.

Caldwell has recently commented upon the implications of detecting an anti-HBc-positive serology in an organ donor. Post-transplantation HBIG was effective in averting allograft-transmitted HBV infection from an anti-HBc-positive donor for a patient who required a transplant because of HBV infection. Lamivudine therapy has also been administered successfully to patients with *de novo* HBV infection following transplantation.

Finally, the presence of anti-HBs in the potential allograft recipient (e.g. following the previous administration of HBV vaccination) may also be an important consideration. The risk of HBV infection by transplantation of a liver allograft from a donor who is anti-HBc positive is reduced for recipients who have pre-existing anti-HBs (personal communication: John Fung, MD, University of Pittsburgh).

Hepatitis C virus The hazard of HCV transmission from a previously infected organ donor is a concern for all allograft recipients. Approximately 5 per cent of all organ donors are anti-HCV positive. The presence of anti-HCV antibody is indicative of an HCV infection as anti-HCV antibody appears in a peripheral blood sample within 2 months of HCV exposure. Most organ procurement organizations have adopted a policy of screening organ donors for anti-HCV antibody. Currently, all organ donors should be tested for anti-HCV antibody using a second-generation enzyme immunoassay which has increased sensitivity and specificity compared with earlier assays. The antibody to HCV does not represent protection from donor organ contamination; however, it is also important to emphasize that detecting anti-HCV in the donor serology panel is not predictive of HCV transmission.

Approximately 50 per cent of anti-HCV-positive patients have detectable hepatitis C viremia by PCR analysis of a peripheral blood specimen. All PCR-positive organ donors will transmit HCV to allograft recipients; however, the risk of HCV transmission from a PCR-negative (anti-HCV antibody positive) donor is unclear. Unfortunately, the use of PCR testing (or branched-chain DNA testing) cannot be accomplished within the time constraint of the preservation period necessary for the release of the donor organs. Thus, exclusion of all anti-HCV-positive organ donors would eliminate the possibility of HCV transmission; however, such an indiscriminate policy would unnecessarily discard some organs that are not infected with HCV.

The consequence of receiving an organ from a donor who is anti-HCV positive is as follows: approximately 50 per cent of the recipients have detectable anti-HCV antibody; 24 per cent have detectable hepatitis C viremia by PCR analysis; and 35 per cent may develop liver disease.

However, notwithstanding the high risk of transmission of HCV to allograft recipients, a positive screening result does not necessarily rule out organ donation. Transplanting an HCV-positive heart allograft when the recipient's life is endangered is an accepted practice even though the recipient is HCV negative. Otherwise, a general restriction of transplanting organs from an HCV-positive donor has been controversial. Since the treatment of HCV infection in an immunosuppressed allograft recipient is usually not successful, many centers regard a positive donor serology for HCV as a contraindication to transplantation. Fishman has suggested a selective strategy of reserving organs from HCV-positive donors for recipients with a previous HCV exposure and detectable anti-HCV antibody. There is evidence in both animal models and human studies that ‘superinfection' with donor strains can occur in HCV-positive recipients, so this approach is not without risk. However, liver transplantation from an anti-HCV-positive donor to an anti-HCV-positive recipient does not appear to cause an increased morbidity or mortality. Furthermore, the kidneys from an anti-HCV-positive donor could also be targeted for those who are sensitized, diabetic, or elderly, and thus may have a limited opportunity for a successful transplant, or a relatively limited lifespan. This approach seems to be prudent, as the long-term risk of chronic hepatitis in renal transplant recipients has now been well documented.

Cytomegalovirus (**CMV**) The seminal work of Ho established the cadaveric organ donor as a most important source of CMV infection in the organ recipient. Thus, all prospective organ donors and recipients should be routinely tested for antibody to CMV. The presence of antibody in either the donor or the recipient indicates the presence of latent viral infection. The specific CMV donor and recipient serologic status has implications for prophylaxis, the highest risk group being CMV-seronegative recipients of CMV-seropositive donor organs (i.e. the so-called primary infection group). Transplantation of an organ from a CMV-positive donor results in subsequent reactivation of the latent virus and replication in the immunosuppressed host. In addition, there is molecular evidence that superinfection from the donor strain can occur in CMV-seropositive recipients. However, recent evidence suggests that donor CMV positive/recipient negative mismatches can be successfully overcome by strategies aimed at CMV prophylaxis.

Historically, several studies in kidney, liver, and lung transplantation demonstrated decreased survival among recipients of CMV-seropositive organs. Centers performing small bowel transplantation have chosen to restrict CMV-seropositive donors from CMV-seronegative recipients because of high morbidity and mortality. The development of CMV disease has been associated with increased morbidity and cost associated with transplantation as well as opportunistic infection. Recipient CMV infection is associated with an increased risk of organ rejection, allograft loss, and death. Nevertheless, data from many of these studies predate more effective prophylactic strategies in the control of CMV infection or disease.

The transplantation of CMV-seropositive donor organs has not been considered an absolute contraindication to transplantation despite the increased risks cited earlier. This is especially pertinent for potential living organ donors. The primary reason for allowing such organs to be transplanted has been the high seroprevalence of the virus in the general population. The prevalence among both organ donors and recipients varies from 40 to 80 per cent depending on age, socio-economic status, and geographic distribution. Fortunately, better control of CMV with refinements in immunosuppression, and the development of effective antiviral agents, has improved the outcome for patients receiving organs from CMV-seropositive donors.

From a practical standpoint, donor testing for CMV should always be performed prior to any blood or plasma product transfusion to an organ donor. This clearly establishes the risk of CMV transmission specifically from the organ donor. Testing of organ donors after blood transfusion presents an additional complexity of assessing the risk of CMV transmission in the blood products given to the donor, even though this risk would be very small (currently less than 1 per cent risk of an organ donor acquiring CMV from blood product transfusion). Furthermore, testing after donor transfusion also introduces the possibility of a false-positive organ donor CMV result, because of anti-CMV antibody passively transmitted in the blood products.

Use of CMV IgM antibody assays in addition to IgG antibody assays has been advocated by some organ procurement organizations. However, it is our experience that

IgM anti-CMV is not present in the absence of IgG antibody. Furthermore, the IgM assay lacks sensitivity; it is not as standardized as the IgG assays that are available and false-positive results are not uncommon. Therefore, from a practical perspective, we would not recommend their use since there is no additional information gained by screening for IgM antibody to CMV.

Other viruses The transmission of neurotropic viruses such as rabies and Creutzfeldt–Jacob disease has been reported from tissue donors. Thus, organ donor deaths associated with these diagnoses exclude donation. Although successful transplantation of renal allografts from donors with Reye's syndrome (encephalopathy and liver failure) has been reported, some transplant centers may be reluctant to expose their recipient to an unknown (presumed viral) etiology of donor death.

Over 95 per cent of potential adult donors are seropositive for Epstein–Barr virus (**EBV**) (IgG antibody); thus, serologic screening of organ donors for EBV has not been routinely performed. However, primary EBV infection, that is, transplantation of an organ from an EBV-seropositive donor to an EBV-seronegative recipient, is associated with an increased risk of post-transplant lymphoproliferative disease. Therefore, recognition of this mismatch in a potential allograft recipient known to be EBV negative may have important prognostic information. Currently there is no effective means of preventing this complication.

Kidneys have been successfully transplanted from donors dying of Reye's syndrome, in which the influenza virus has been implicated. No adverse effects, in particular the development of Reye's syndrome, were evident in the allograft recipients.

Since hepatitis A virus has not generally been considered a virulent pathogen for the allograft recipient, a donor history of hepatitis A virus infection has not been regarded as a contraindication to organ donation. A 3-month interval between active hepatitis A and death is accepted as sufficient to allow for organ donation, particularly if the serum titer of IgM antibodies to hepatitis A virus has returned to normal.

Genital herpes virus infection at the time of death may raise a suspicion that the potential donor is a concomitant HIV carrier, even though the enzyme-linked immunosorbent assay does not reveal the presence of anti-HIV antibody. Aside from the implication of HIV infection, active herpes simplex may be transmitted through renal allografts to the immunocompromised recipient, and organ procurement from these individuals is best avoided.

Acute varicella zoster infection can be lethal in an immunosuppressed allograft recipient and all potential allograft recipients should be screened for antibody to varicella zoster virus prior to transplantation. Allograft recipients who are antibody negative are warned to avoid contact with individuals experiencing acute varicella, and individuals with primary or reactive varicella at the time of death should probably be excluded from donating organs.

Treponemal antigen (syphilis)

The detection of antibody to syphilis by the rapid plasma reagin test (**RPR**) is not a contraindication to organ procurement, but it is a contraindication to tissue procurement. The RPR is reactive in more than 90 per cent of patients with primary syphilis, but it may be negative between 6 and 18 months following primary infection. A Venereal Disease Research Laboratory Test (**VDRL**) can detect non-specific antibodies directed against lipoidal antigens of *Treponema pallidum*. A VDRL or RPR test is routinely obtained on all individuals considered for organ and tissue donation. Although a positive test suggests exposure to *T. pallidum*, the lipoidal antigens used in the assay are found in a number of normal tissues, and a more specific test such as the fluorescent *Treponema* antibody absorption assay may be necessary to exclude a false-positive VDRL result.

Although syphilis can be transmitted by blood transfusion, we are unaware of a recorded infection of a transplant recipient from a syphilitic donor. Moreover, a standard course of penicillin would provide sufficient antibiotic coverage to prevent syphilitic complications in an allograft recipient. This regimen has been reportedly successful in the treatment of two renal allograft recipients intentionally given kidneys from an organ donor whose serology tests for syphilis were positive. Thus, a recipient of any RPR-positive organ donor should be treated for 15 days with penicillin, or doxycycline, tetracycline, or erythromycin for those who are allergic to penicillin.

Parasitic infection

Toxoplasma gondii is an indolent parasite, and infectivity of a potential donor may only become apparent after transplantation. Although toxoplasmosis may now be detected as an associated infection in patients with immunodeficiency syndromes such as AIDS, toxoplasmosis may be transferred from an unsuspected carrier who has died abruptly from trauma. In this latter circumstance detection of donor toxoplasmosis may only be accomplished retrospectively, following the detection of an elevated level of antibodies against *Toxoplasma* in stored donor blood. In general, retention of donor serum for such unforeseen developments may be prudent. Lethal toxoplasmosis has been transmitted to recipients of cardiac allografts from apparently healthy seropositive donors. The possible transmission of the protozoan *T. gondii* from donors with unsuspected central nervous system disease is a concern, especially for heart allograft recipients because of the predilection of this parasite for muscle tissue. Thus, donor seropositivity to *Toxoplasma gondii*, identified in advance of transplantation, may be a contraindication to donation, especially for a seronegative recipient. Fortunately, the use of trimethoprim-sulfamethoxazole for *Pneumocystis carinii* prophylaxis prevents transmission of *T. gondii*.

Response regarding positive serology testing of donors

When serology screening reveals a positive test, confirmatory testing may be subsequently performed by independent laboratories for each pathogen (including *Treponema*). The practice of the New England Organ Bank is to report the confirmed positive findings to the family member who gave consent for donation if there is a hazard that the donor might have transmitted infection to that family member (contact with bodily fluids). This process assures that any family member would be appropriately cared for who is potentially at risk of exposure to HBsAg, hepatitis C virus, HIV, HTLV, or syphilis. Otherwise, such notification would not apply to estranged family or guardians who consent to donation.

Bacteriology and donor sepsis

From an organ bank perspective, sepsis at the time of a patient's death introduces the risk of transmitting infectious disease by donor organs contaminated with bacteria. However, organ procurement organizations now routinely evaluate donor infection on an individual basis, considering the blood culture bacteriology, the presence and duration of central lines, and the nature of the donor infection, before excluding a patient as an organ donor. The risk of bacteremia may be influenced by an extremity cellulitis (especially adjacent to external orthopedic fixatives), or a current intestinal injury or surgery. However, none of these conditions alone prohibits organ donation.

Patients with bacterial meningitis may also be acceptable for organ donation. *Haemophilus influenza, Streptococcus pneumoniae*, and *Neisseria meningitidis* are the most common bacterial pathogens. The administration of at least a broad-spectrum cephalosporin to a potential donor is indicated under these circumstances.

However, other bacterial organisms carry a more significant hazard for blood vessel disruption following transplantation. For example, *Staphylococcus aureus, Bacteroides* spp., *Klebsiella enterobacter, Escherichia coli, Pseudomonas aeruginosa*, and fungal infections such as *Candida albicans, Cryptococcus neoformans*, and *Histoplasma capsulatum* have been transmitted to allograft recipients from infected donors or infected preservation fluids, and implicated in the formation of recipient mycotic aneursyms and disruption of their allograft anastomoses.

In the case of bacteremia, if a sufficient course of antibiotic therapy has been administered which minimizes the risk of bacterial transmission, then organ donation may be considered. Lung recovery may still be accomplished when the donor has a unilateral pneumonia.

In contrast, fungal contamination is very worrisome because the opportunity for adequate donor treatment is limited. Because *Candida* infections are particularly hazardous to the pancreas transplant recipient, many transplant surgeons administer amphotericin B through the nasogastric tube into the donor duodenum during the recovery procedure.

Bacterial infection

Cystitis, as manifested by a positive culture of urine obtained through a Foley catheter, is not a contraindication to donation. If pyuria is observed, ureteral stump cultures may be obtained during the procurement procedure. While bladder contamination following Foley catheterization has not been associated with recipient infection following transplantation, a positive ureteral stump culture raises the possibility of active pyelonephritis, which is a contraindication to transplantation. Good surgical technique, with ligation of the distal ureters following ureteral transection, prevents potentially contaminated urine refluxing from the bladder into the open peritoneal cavity.

A history of pyelonephritis within 3 months of organ donation may increase the risk of transmission of bacteria to the recipient. Clinical judgement regarding the type of organism and verification of treatment should be exercised before recently infected kidneys are accepted for transplantation. *P. aeruginosa* and *Staph. aureus* have

been reported to disrupt the vascular anastomosis of a renal allograft.

A history of chronic respiratory infection or acute pharyngitis does not exclude a patient from donating organs (with the exception of lung donation) unless a virulent organism such as *Staph. aureus* is identified. If bacteremia is not demonstrated by blood culture, donation may be considered to be safe.

Brain death due to mycotic aneurysm necessitates an evaluation of the cause: *C. albicans* and *Staph. aureus* are extremely dangerous to allograft recipients, and blood cultures from potential donors may be negative.

The death of a renal allograft recipient from disseminated tuberculosis transmitted through a cadaver kidney obtained from a donor with unsuspected tuberculosis meningitis has been reported. The etiology of the meningitis only became apparent several weeks after the donor's death, when mycobacteria grew in cultures of his cerebrospinal fluid. The donor's chest radiograph was normal. As well as documenting the transmission of tuberculosis via transplanted organs, this case underscores the danger of accepting organs from individuals with a diagnosis of meningitis of unknown etiology.

Fungal infection

C. albicans frequently colonizes the vagina and perineum of patients who are maintained on broad-spectrum antibiotics for a long period of time. Thereafter, the organism may gain entrance to the bladder through an indwelling Foley conduit. Wound infection and vascular disruption may follow transplantation of organs contaminated with *Candida*, especially to diabetic recipients. Fatal candidal mediastinitis has been reported in recipients of lungs from donors with a heavy growth of *Candida* in cultures from the trachea.

Histoplasma and *Cryptococcus* have been transmitted to renal allograft recipients by organs from donors who died of intracerebral pathology, without evidence of a pulmonary infection. These organisms are difficult to eradicate from the central nervous system unless a protracted course of antifungal therapy (amphotericin B) is administered. Even with careful documentation of proper therapy however, a history of fungal infection (also including cocciodiomycosis and blastomycosis) should exclude an individual from donation, despite the apparent absence of infection at the time of death.

Infections associated with terminal injury

Chest tubes are not a contraindication to organ donation unless culture of the pleural fluid is positive for an organism which poses a high risk to the recipient, such as *P. aeruginosa* or *Staph. aureus*.

Peritoneal drains contaminated with enteric bacteria may assume more significance, especially if they reflect current peritonitis due to injured or devitalized bowel. A retroperitoneal dissection of each kidney can be accomplished through separate flank incisions if necessary. This approach may also be used when gastrostomy tube leakage into the abdominal cavity is suspected in patients who have been nutritionally depleted and who display poor wound healing.

Although perforations of the bowel repaired within 1 week of brain death probably exclude a patient from consideration as a donor, prior repair of the intestine may not, provided that a Gram stain of peritoneal fluid reveals no organisms, blood cultures are negative, and appropriate antibiotic coverage has been administered during the interval.

Extremity fractures in a potential donor may pose a hazard for allograft recipients if metal appliances have been placed to stabilize the fracture and/or there is evidence of cellulitis adjacent to the cutaneous exit site of the appliance. Once again, antibiotics should have been administered for at least 48 h and blood cultures must be free of bacteria before organ procurement is undertaken.

Burn victims who sustain tracheobronchial injury from smoke inhalation are at risk of pneumonitis. Burn injuries also compromise the procurement procedure since incisions cannot be made through damaged skin. Nevertheless, if incisions can be fashioned through areas of skin which have not been burned, organs may be procured from burn victims within 48 h of injury. A longer interval between injury and death increases the risk of septic contamination of visceral organs, through either a compromised respiratory tract or through burned skin.

Infections acquired at the time of death

Cellulitis adjacent to antecubital or subclavian vein intravenous lines signals the possibility of line infestation, which can seed visceral organs with bacteria such as *Staph. aureus*. Management prior to procurement surgery entails the removal of the intravenous line, administration of appropriate antibiotics for at least 24 h following line removal, and the determination of blood cultures.

Laboratory assessment

Traumatic injury to any visceral organ may initially be detected through laboratory testing of donor blood samples (Table 7). Unremitting functional impairment because of shock or trauma generally precludes procurement of the specifically injured organ.

ABO blood type
Complete blood count
Platelet count
Prothrombin time
Partial thromboplastin time
Electrolytes (Na, K, Cl, CO_2)
Blood glucose
Blood urea nitrogen, creatinine
Bilirubin (total and direct)
Serum glutamic oxaloacetic transaminase
Serum glutamic pyruvic transaminase
Alkaline phosphatase
Amylase
Electrocardiogram
Echocardiogram
Creatinine phosphokinase myocardial band
Chest radiograph
Bronchoscopic examination
Blood gases: Po_2, Pco_2, pH
VDRL or rapid plasma reagin test
Hepatitis B surface antigen
Serologic testing for cytomegalovirus, hepatitis C virus, and HIV
Microbiology cultures: blood, sputum, skin effluents, and urinoid stump
Urine analysis

Table 7 Laboratory testing

Candidates for heart donation are usually free of cardiac murmurs and display a normal 12-lead ECG. Pathologic Q waves suggestive of infarction are a contraindication to donation; other electrocardiographic abnormalities, such as ST segment changes, do not exclude donation. Echocardiography may be performed to assess myocardial function and potential valvular pathology in patients requiring pressor support, or in whom a murmur is audible. Mitral valve prolapse without mitral regurgitation is not uncommon and alone does not preclude heart procurement.

Candidates for lung or heart and lung donation should usually be free of a smoking history and have clear lung fields on chest radiographs. A Gram stain and culture of the tracheobronchial secretion should be performed. Blood gas assessment or respiratory function is determined by ventilating the donor with 100 per cent oxygen and no more than 5 cm of positive end-expiratory pressure: under these conditions, the arterial oxygen pressure of a blood sample should be greater than 300 mmHg.

Many transplant centers are willing to accept livers from donors whose liver function tests are not within normal range. A 2- or 3-s elevation in the prothrombin time, and/or mild elevations in transaminase levels are not necessarily a contraindication to procurement. A careful history and serologic testing for HBV and HCV are important elements in the decision.

Small liver lacerations may be observed at the time of organ procurement from victims of trauma with normal blood transaminase levels. A small laceration with no bleeding or bile leakage from the site may not represent a contraindication to liver transplantation.

Liver function is especially susceptible to hypoxia: hemodynamically unstable patients requiring pressor support, with marginal arterial oxygen pressure (less than 100 mmHg), may not be satisfactory candidates for liver donation, particularly if the oxygenation cannot be improved because of pulmonary edema. Consequential

impairment of liver function may be discerned by an increase of at least 5 s in the prothrombin time and a transaminase level that is twice normal.

A coagulopathy detected by thrombocytopenia, low fibrinogen level, and a marked elevation of the prothrombin and partial thromboplastin times suggests a disseminated intravascular coagulopathy, which is known to be associated with head injury, but may also be the result of sepsis. Once again, clinical judgement must be exercised in determining any possible cause of sepsis, impairment of organ function, and the potential for transmission of bacterial organisms. A biopsy of the renal allograft prior to transplantation may be necessary to exclude the presence of microthrombi within the renal parenchyma. Pulmonary dysfunction is a frequent cause of irresolvable hypoxia in patients with disseminated intravascular coagulation because of the deposition of microthrombi in the lungs.

The standard determinations of serum creatinine, blood urea nitrogen, urinalysis, and urine output (more than 1 ml/kg.min) are used to assess satisfactory donor renal function. Creatinine levels above 2.0 mg per cent are not necessarily a contraindication to kidney donation if the cause of renal dysfunction is transient hypotension, and if the creatinine level falls with intravenous fluid repletion. Chronic renal insufficiency, as noted by a persistent creatinine level of greater than 2.0 mg per cent, combined with a urinalysis revealing proteinuria and/or sediment casts, contraindicate renal donation.

A history of diabetes mellitus precludes pancreas donation, but does not necessarily prohibit renal donation, especially if the serum creatinine level is less than 2.0 mg per cent. Mild elevations in serum amylase may not be a reflection of pancreatic injury or pancreatitis, since the enzyme may be derived from a salivary gland source following trauma. Neither mild hyperamylasemia nor mild hyperglycemia are contraindications to pancreas donation.

Lymph node procurement

Many organ procurement agencies have adopted a practice of extracting lymph nodes from the inguinal bed as soon as death has been declared and permission for organ donation has been granted. This practice has facilitated the identification of renal allograft recipients prior to kidney procurement, as accurate tissue typing from peripheral blood is usually difficult, and the subsequent preservation period of cold ischemia.

A 6- to 8-cm incision is made below the inguinal ligament, along the course of the femoral vessels, using sterile technique. Three or four lymph nodes are more than sufficient for HLA typing, and these may be readily identified medial to the femoral vein.

Social considerations

In the same year in which the criteria were developed to determine brain death (1968), a Uniform Anatomical Gift Act was independently promulgated in the United States to facilitate the process of organ donation. The concept of the donor card was derived from the Uniform Anatomical Gift Act, as it permitted the decedent to certify an intent to donate, irrespective of the wishes of the surviving family. Nevertheless, a hierarchy of responsible family members was established for circumstances in which family approval would be requested. The order proceeded from spouse, to offspring, to parent, to sibling, and finally to guardian.

Unfortunately, few people have completed donor cards, even though renewal of a driver's licence affords a simple opportunity to do so. Moreover, the general policy of organ procurement agencies has been to obtain permission from the surviving family member, determined by the hierarchy noted above, even in instances where a donor card has been identified.

Family wishes have prevailed despite properly indicated donor intent. It has also become a standard practice to obtain permission for organ procurement from the local medical examiner or coroner when potential donors have died from an unnatural cause. The medical examiner should be contacted in advance of organ procurement. The operative report may serve as evidence of the postmortem examination of the thoracic and abdominal cavities.

Recently the Department of Health and Human Services of the United States promulgated a regulation that all hospital deaths must be brought to the attention of the local organ procurement organization. Perhaps this initiative will foster a routine evaluation of all deaths for organ and tissue recovery.

Procedure from donor referral to organ recovery

Stratification of the organ procurement organization procedure is based upon potential clinical situations. The activities noted within each section are exclusive and they are presented in a sequential order. The importance of this paradigm is to establish a practice that is both consistent and ethical. There are two cardinal rules that should be emphasized: (i) blood sampling to determine organ donor suitability may be done after the family gives permission, in advance of death; however, (ii) no invasive procedure for the purpose of organ donation should be performed unless the patient is declared dead.

I. Referral for donation by the primary care team

The initial co-ordinator activity of the organ procurement organization is to establish the criteria of donor suitability. Consultation by the co-ordinator of the organ procurement organization with the intensive care team may be given in advance of the brain death declaration, regarding discussion of the family option for donation, and a review of donation procedure.

II. Verbal family interest for donation

Non-invasive procedures for the sole objective of donor evaluation are indicated. Family consent for donor evaluation should be documented in the patient chart by the intensive care team or the co-ordinator of the organ procurement organization, independent of consent for donation. Charges for evaluation are the responsibility of the organ procurement organization. These include clinical blood sampling, blood type and cross-match for possible donor transfusions, urinalysis, sputum and urine cultures, chest radiograph, electrocardiogram, and echocardiogram (Table 7). There should be no written orders by the co-ordinator of the organ procurement organization until a declaration of death and written family consent is obtained.

III. Written family consent for donation, in advance of the diagnosis of death

HLA typing by peripheral blood samples should be performed if at all possible, to expedite the identification of potential recipients. However, this may be contraindicated by red blood cell and platelet transfusions, given to the donor at the time of death. HLA typing is then best accomplished by femoral lymph node procurement after the declaration of death. Serology testing for hepatitis, HIV, HTLV, CMV, and syphilis may now be performed, but it should not be done prior to written family consent.

IV. Written family consent for donation, following the declaration of death

Organ procurement organization orders of donor management are now permissible; so too are invasive procedures which include groin node retrieval, bronchoscopy, cardiac catheterization, and placement of a central line. Thus, HLA typing and cross-matching by lymph node sampling of the donor is indicated. However, no invasive procedures should be performed until a pronouncement of death is made. Organ and tissue procurement can now proceed following medical examiner notification.

Obtaining consent

Perhaps the most significant obstacle to increasing the number of cadaver organs available for transplantation has been obtaining consent. Although there are missed opportunities because the family was never asked to consider organ donation, the lack of family consent is the most common reason why potential donors are lost. Uncoupling a family discussion of the declaration of death by the intensive care team from the request for organ donation by a representative of an organ procurement organization appears to be an approach which permits families to consider donation more favorably. Family consent remains an absolute necessity for organ donation in many communities, irrespective of an individual's intention as indicated by a donor card. However, the identification of a donor card upon a deceased person can influence the decision of the family to donate, especially for individuals from ethnic minorities. Moreover, there are now statutes which give explicit permission to recover organs with a signed donor card, even in the absence of family consent. For example, Pennsylvania currently makes hospitals legally responsible for referring a potential donor and mandates trained personnel of organ procurement organizations to accomplish the request of the family (personal communication: Howard Nathan, Delaware Valley Organ Procurement Organization). Nevertheless, this is not akin to the presumed consent approach that is active in some European countries. In such a scenario, there is the presumption of family consent irrespective of a signed donor card. We do not foresee that the general society of the United States would accept organ donation unless stipulated by a patient donor card or by family consent.

Although public surveys indicate that most people have a positive attitude towards organ donation, there are several misconceptions which can arise at the time of the family request and become obstacles to obtaining consent. These include: a confusion about the meaning of brain death; a belief that the transplant is experimental;

that there is an underworld market for organs; that donor families may face additional medical bills; that the donor's body may be disfigured; and that the medical profession orchestrating the donation is not trustworthy. Thus, to achieve an ideal rate of success the consent process requires structure, consistency, and an experienced person in making the request of a bereaved family.

It is now evident that a discussion of the donation process by the intensive care team poses a conflict for the family which may heighten their reluctance to consent. Also, the intensive care team may not be skilled in the request process. On the other hand, uncoupling the determination of death by the intensive care team from the request for donation by the representative of the organ procurement organization presents a clearer line of responsibility to the family. Family questions regarding the misunderstandings elaborated above are better answered by a representative of the organ procurement organization. Families may also wish to enquire about the outcome of the donation surgery. The co-ordinator of the organ procurement organization can inform the donor family of the result of their donation without revealing the identity of the organ recipients. This opportunity is an appealing benefit to the donor family.

Thus, the elements of best practice to achieve consent for donation are: (i) a thorough discussion by the intensive care team of the irreversibility of the brain injury which has led to the diagnosis of death, (ii) an early referral to the organ procurement organization for assessment of donation potential, and (iii) after the family accepts the determination of death, introduction of a highly trained professional from the organ procurement organization to discuss the process of organ donation.

Directed donation

While it is illegal in the Unites States to restrict the donation of an organ or organs to a class of individuals (e.g. by gender, race, or national origin), directed donation to a specific individual by name may be acceptable. Ideally this individual should be known to the family member requesting the directed donation. Since the buying and selling of human organs is also illegal, the intensive care physician should not knowingly participate in any directed donation involving monetary exchange or other inducement to directed donation. Thus, the nature of the relationship between the next of kin or donor and the individual to whom the organ is being directed should be ascertained. Unless a relationship is substantiated, the possibility of directed donation is precluded. The family should also be informed that ABO incompatibility, as well as other medical considerations, may preclude the possibility of directed donation.

Donor management

Intensive care unit management of the brain-dead potential organ donor is a complex and dynamic process directed towards maintenance of end-organ function and viability. Commonly encountered clinical problems include hypotension, polyuria, electrolyte imbalances, cardiac dysfunction, and hypothermia. Increased blood levels of cytokines occur in brain-dead patients. Although interleukin 1b and tumor necrosis factor-a levels appear to be within normal range, interleukin 6 (**IL-6**) levels have been shown to be abnormal. The consequence of the IL-6 elevation is not known. However, careful monitoring, anticipation of instability, and rapid treatment is required to succeed in maximizing the recovery of transplantable organs.

Initial management goals

The primary donor management goals include restoration and maintenance of normothermia, return of blood pressure to normal, optimization of lung function, restoration of intravascular volume, and correction of acid/base and electrolyte imbalances. Conflicting approaches to ideal management from the various transplant specialties may require the placement of a central line to assure filling pressures which avoid fluid overload but permit optimal organ perfusion. Otherwise, without this sophisticated data, the kidney team may prefer aggressive fluid administration and brisk diuresis, the liver and pancreas team may favor adequate but not excessive portal and arterial perfusion pressures, and at the same time the thoracic teams request volume contraction to minimize the development of pulmonary edema. Therefore, preliminary evaluation of organ suitability may be helpful in guiding management when one or more organs are clearly unsuitable. Subsequent management efforts can then be directed towards those organs that are likely to be utilized.

Poikilothermia

Loss of thermoregulatory function follows hypothalamic dysfunction in brain-dead patients. A shivering thermogenesis is absent. Passive heat loss may lead to progressive hypothermia, which is well known to adversely affect cellular metabolism, oxygen release, and cardiac function.

Due to the relatively high body surface area to body mass ratio in neonates and children, core temperatures may fluctuate much more rapidly than in adults. In all cases, patient temperature should be carefully monitored and hypothermia should be treated early. It is easier to prevent hypothermia than to reverse it. Nevertheless, effective treatment includes maintenance of room temperature at 24°C (75°F), keeping the body and head well covered at all times, using warming blankets, and heating intravenous fluids. The use of a warming blanket in the operating room is essential to counter passive heat loss and maintain normothermia. Heating humidified oxygen (to 45° C) via the ventilator may also be of benefit; however, this is not usually necessary. Since oxygen heating may cause thermal injury to the airway, it is not employed unless the routine measures are unsuccessful.

Hypertension

Severe hypertension (systolic blood pressure greater than 200 mmHg) is infrequently encountered in brain-dead patients. However, when hypertension is observed, it is probably related to brainstem herniation and is therefore self-limiting. The reason to treat sustained hypertension is to avoid transient rhythm disturbances. The goal of treatment should be to maintain the diastolic blood pressure below 100 mmHg.

Sodium nitroprusside (Nipride) infusion is the treatment of choice. Nipride should only be used in cases of severe, persistent hypertension, since prolonged administration may result in cyanide toxicity.

Hypotension

Alternatively, hypotension (systolic blood pressure less than 90 mmHg) is more frequently observed in brain-dead patients, especially at the time of the initial referral. Tilney has detailed the cytokine storm that occurs in association with brain death that can account for donor hypotension. Since brief intervals of hypotension can influence organ function following transplantation, a central venous pressure line and an arterial line are essential in order to manage hypotension appropriately. The most common cause is hypovolemia. However, if a hemodynamically unstable donor persists following volume expansion, a pulmonary artery catheter should be introduced, and serial cardiac outputs and systemic vascular resistance measurements should be determined.

Hypovolemia

Excessive intravascular volume loss (hemorrhage, diabetes insipidus) is common in potential organ donors, especially trauma victims. Neurogenic vasodilatation, common in brain-dead individuals, may exacerbate the hemodynamic instability. Intravascular volume deficits should be corrected prior to the use of vasoactive drugs. Furthermore, patients with severe head injuries may intentionally be kept volume contracted. Low- to moderate-dose dopamine hydrochloride (Intropin) infusion (3 to 5 µg/kg.min) may be required to maintain adequate blood pressure during the initial period of volume repletion. Colloid (albumin) and blood products (packed red blood cells) may be necessary for volume expansion in combination with crystalloid solutions. Monitoring of the central venous pressure (ideally 10 to 15 mmHg) will avoid the development of pulmonary edema, especially in potential lung donors. As the blood volume is restored, vasoactive drugs should be discontinued.

Decreased vascular resistance

Despite adequate volume restoration, some donors will require additional vasoactive drug therapy because of neurogenic vasodilatation. A dopamine dosage between 5 and 10 µg/kg.min may be necessary, since the a-receptor of peripheral blood vessels does not respond with vasoconstriction until this dose range of dopamine is reached. However, a dopamine dose that exceeds 10 µg/kg.min requires insertion of a pulmonary artery catheter for determination of cardiac filling pressures, cardiac output, and systemic vascular resistance. The systemic vascular resistance is calculated by thermodilution cardiac output measurements. If the systemic vascular resistance is less than 400 dyn/s.cm, vasoconstrictor therapy is indicated; otherwise, to avoid the risk of decreased perfusion to visceral organs, vasoconstrictors (e.g. Levophed) should be reserved for low determinations of systemic vascular resistance that are refractory to volume repletion and a dopamine dose between 5 and 10 µg/kg.min. This scenario raises the suspicion of donor sepsis.

Depressed cardiac function

Some donors may exhibit depressed cardiac function, not surprisingly following resuscitation from cardiopulmonary arrest, due to pre-existing cardiac disease or secondary to brainstem herniation. Brain death causes a significant loss of right and left ventricular function. These injuries are greater in the right ventricle and may contribute to early right ventricular failure after transplantation. The deterioration of myocardial performance after brain death correlates temporally with desensitization

of the myocardial b-receptor signal transduction system. Thus, these patients may be hypotensive despite adequate volume restoration. Donors with depressed cardiac function may require inotropic and chronotropic support to maintain an adequate cardiac output. If the cardiac index is less than 2.0 l/min.m, inotropic therapy is indicated, preferably with dopamine at a dose of 3 to 5 µg/kg.min. The use of vasoconstrictors such as Levophed may exacerbate cardiac dysfunction.

Thyroid hormone replacement therapy (tri-iodothyronine: maximal dose 0.6 µg/kg) has been used to improve myocardial function, allowing the use of donor hearts that might otherwise have been considered unsuitable for transplantation, although its efficacy in this setting is controversial.

Mariot has carried out an extensive analysis of thyroid and cortisol levels in brain-dead donors. The free T_3 was low in 105 of 130 donors; the thyroid-stimulating hormone was low in 21, but normal in 99; and the cortisol was low in 19, but normal in 101 patients. These results did appear to support the routine use of either tri-iodothyronine or cortisol to maintain hemodynamic stability.

Respiratory insufficiency

Brain-dead patients require frequent pulmonary hygiene to prevent atelectasis and to maintain adequate oxygenation. In addition, brain death following head trauma may also be associated with an unexplained ‘neurogenic' pulmonary edema. These patients develop an abnormal capillary permeability, with movement of intravascular fluid into the pulmonary interstitium and eventually into the alveolar space, impairing gas exchange and progressively reducing the volume of ventilated lung tissue. Positive end-expiratory pressure of 5 cmH_2O (‘physiologic **PEEP**') may be helpful to maintain alveolar expansion in these circumstances.

Ventilator settings for brain-dead patients should include a tidal volume of 10 to 15 ml/kg and a respiratory rate sufficient to maintain arterial P_{co2} in the 40 to 45 mmHg range. The fraction of inspired oxygen should be kept at 40 per cent or less in potential lung donors to prevent pulmonary oxygen toxicity. Levels of PEEP greater than 7.5 cmH_2O may impede venous return and decrease cardiac output.

Hypoxemia

Arterial oxygen saturations should be no less than 95 per cent. Hypoxemia should be treated aggressively by adjusting the ventilator FIO_2 to maintain PAO_2 levels greater than100 mmHg, in order to minimize injury to transplantable organs.

Hypercarbia

Apnea testing to confirm the diagnosis of brain death can result in significant hypercarbia leading to respiratory acidosis. Careful attention should be paid to the patient's ventilatory status and arterial blood gases, especially following clinical examinations for determination of brain death. The ventilator rate should be increased as needed to return $P(A)_{co2}$ levels to the normal 40 to 45 mmHg range.

Polyuria

Polyuria (urine output greater than 500 ml/h) is frequently seen in brain-dead patients. It may be due to physiologic diuresis, osmotic diuresis (mannitol, hyperglycemia), diuresis caused by hypothermia, partial or complete central diabetes insipidus, loop diuretics, or a combination of the above. Excessive polyuria due to osmotic diuresis or diabetes insipidus may lead to hypernatremia, hypokalemia, and hyperosmolality. The serum potassium should not fall below 3.5 mEq/l, to avoid arrhythmias. Urine and serum electrolyte levels and osmolality aid in the determination of the cause of polyuria.

Aggressive volume restoration may result in a physiologic diuresis. No treatment is warranted, but intake and output should be monitored carefully. Prior mannitol administration or excessive glucose infusion may result in an osmotic diuresis. Glycosuria and hyperglycemia should be treated with a sliding scale of intravenous insulin therapy (4 units of regular insulin for every 100 mg per cent of blood sugar above 200 mg per cent) to return blood glucose levels to normal.

Central diabetes insipidus

Once the above-mentioned causes for polyuria have been excluded, the diagnosis of diabetes insipidus can be made by evaluating the volume of urine output, urine specific gravity, urine and serum electrolyte levels, and urine and serum osmolality. When three of the following findings are noted simultaneously, the diagnosis of diabetes insipidus is established:

- urine output greater than 500 ml/h
- serum sodium greater than 155 mEq/l
- urine specific gravity less than 1.005
- serum osmolality greater than 305 mosm/l

Treatment of diabetes insipidus includes an aqueous pitressin intravenous infusion (10 IU/250 ml D5W) with an initial dosage of 1.2 IU/h (rate: 30 ml/h) and subsequent titration to maintain urine output of 150 to 300 ml/h. Aqueous pitressin should be discontinued for urine output less than 150 ml/h to minimize the possibility of decreased visceral perfusion (especially in liver and pancreas donors). In addition to pitressin therapy, the free water deficit should be calculated and 50 per cent of the calculated deficit should be infused as rapidly as possible, preferably with a hypotonic solution such as dextrose 5 per cent water (D5W) or ½ normal saline. Nevertheless, a low-dose AVP infusion can decrease the plasma hyperosmolality, increase blood pressure, decrease inotrope use, and maintain cardiac output.

Disseminated intravascular coagulation

In addition to the constellation of disorders resulting from brain death many potential organ donors have multiple traumatic injuries which pose further clinical challenges. Disseminated intravascular coagulation is often observed following severe head trauma, associated with the release of tissue thromboplastin into the peripheral blood circulation. Disseminated intravascular coagulation presents two problems for successful organ recovery: (i) the potential for massive blood loss leading to hypovolemic shock, and (ii) the possibility of microvascular damage and thrombosis of the end organs. Close attention to the donor coagulation profile in anticipation of disseminated intravascular coagulation, and aggressive treatment with fresh frozen plasma will ameliorate this potentially devastating complication.

Operative events

Following the transfer of the potential donor to the operating room, manually ventilated with 100 per cent oxygen, ventilator settings should be re-established to maintain the partial pressure of arterial oxygen at greater than 100 torr. A warming blanket may be required to prevent the hypothermia (less than 32°C) associated with brain death, and its destabilizing effect upon cardiac function. The patient's arms should be maximally abducted to permit a simultaneous dissection by thoracic and abdominal teams. Anesthetic management of blood gases, central venous pressure, hematocrit, and urine output are essential for successful organ procurement. A hematocrit of less than 30 per cent should be restored by transfusion of warmed blood, especially if the patient is hypoxic or edematous due to previously infused crystalloid solution. Bicarbonate administration may be necessary to correct metabolic acidosis.

A midline incision is made from the suprasternal notch to the pubis. As the sternum is split, bone wax is applied to minimize bleeding. If only intra-abdominal organs are being procured, the sternal incision is omitted and the midline incision is made from the xyphoid to the pubis. Some surgeons use a cruciate abdominal incision just above the umbilicus for extended exposure. Towel clips may be applied from the apex of each leaf of the abdominal incision, as it is folded to the opposing skin of the abdomen.

The organs of interest, whether thoracic or abdominal, must be inspected by the surgeon for unsuspected pathology or injury.

The thoracic teams generally begin the dissection by opening the pericardium and/or pleural cavities. The great vessels, including the vena cavae, the aortic arch, the innominate vessels, and the pulmonary artery are isolated. Both lungs may be mobilized and inspected by the division of the pulmonary ligaments.

The abdominal dissection may vary, depending upon the viscera to be procured. The liver is usually removed first, with the hepatic arterial circulation being established by dissection of the celiac axis and superior mesenteric arteries. Replaced right or left hepatic arteries originating from the superior mesenteric or left gastric arteries must be preserved. When the donor is hemodynamically stable, the celiac axis and superior mesenteric artery can be isolated to their branches. The portal vein is dissected and the inferior mesenteric vein may be cannulated to permit *in situ* portal cooling, usually with Ringer's lactate solution. Alternatively, the splenic vein or

superior mesenteric vein may be used for portal flushing if the pancreas is not required. The common bile duct is divided at the duodenal brim.

Pancreatic dissection may take place next. The spleen is dissected from its bed, and used as a handle to mobilize the tail and body of the pancreas. On the right side, a Kocher maneuver is used to dissect the C-portion of the duodenum and the head of the pancreas, again to the mesenteric vessels.

Each kidney and ureter is then mobilized by dissecting the poles beneath Gerota's fascia, the number and course of the left and right renal arteries being identified.

The donor is then anticoagulated systemically with 10 000 to 20 000 units of heparin, and the distal aorta above the iliac bifurcation is cannulated. Rarely, an iliac artery may provide a lower pole renal artery branch: this anatomic aberration must be carefully assessed before aortic cannulation and, if present, the contralateral iliac artery can be cannulated for prograde aortic cooling. A cannula may also be placed into the inferior vena cava below the renal veins, to provide a controlled exit for visceral perfusate. Some surgeons prefer to transect the inferior vena cava and allow the perfusate to empty into the abdominal cavity.

In the thoracic cavity, cannulas are finally placed into the main pulmonary artery for cardiopulmonary-plegia and prostacyclin infusion, and into the proximal ascending aorta for cardioplegia. The superior and inferior vena cava are ligated and transected as the plegia solutions are given.

Sequence of excision

The sequence of procurement in multiple organ donation is given in Table 8. Thoracic organs are excised first. Either cardiectomy is performed prior to lung removal (possibly as a double lung block) or the heart and lungs are removed *en bloc* and divided on a back table in the operating room. Alternatively the heart and lungs may be removed as a unit for transplantation to a single recipient. As the aortic arch is cross-clamped (usually with a staple device), perfusion of the intra-abdominal organs is begun through portal and aortic cannulas and this is continued during the thoracic excision.

Heart/lung or heart and lungs
Liver
Kidneys
Pancreas
Bones/tissues

Table 8 Sequence of procurement in multiple organ donation

The liver is the first intra-abdominal organ removed. Although the simultaneous procurement of the liver and pancreas from a single cadaver donor was once considered to be technically impossible because of the same blood supply, it is now routinely accomplished. Transection of the portal vein, 2 cm cephalad to its bifurcation into the splenic and superior mesenteric veins, provides sufficient length of portal vein for transplantation with both the liver and pancreas. The arterial supply can be managed by providing a Carrel patch of aorta either to the celiac axis of the liver or to the celiac axis and superior mesenteric artery of the pancreas. Usually however, the celiac axis and aortic patch are retained with the liver, and the iliac bifurcation graft from the donor is anastomosed to the splenic and superior mesenteric arteries of the donor pancreas.

As the liver is removed, aortic perfusion of the kidneys and pancreas with preservation fluid is continued: approximately 2 liters of preservation fluid are usually given through the aortic cannula.

The method of whole-organ pancreas transplantation which is most commonly employed uses a segment of duodenum to drain exocrine secretions into the bladder. Since the donor duodenum must be transected (usually with the staple device), the pancreas may be removed following *en bloc* nephroureteroectomy, to avoid contamination of either kidney with enteric organisms. The pancreas preparation therefore consists of a segment of duodenum, the entire pancreas, and the attached spleen.

Both kidneys are removed *en bloc* with a cylinder of aorta and inferior vena cava, and with attached ureters of at least 10 cm in length. Once removed the kidneys may be divided for separate packaging and cold-storage preservation, or they may be placed separately (or *en bloc)* on to a pulsatile perfusion machine.

Organ preservation

The development of a consistently effective preservation fluid has dramatically changed the timing of renal, hepatic, and pancreatic transplant procedures. The preservation solution developed at the University of Wisconsin has extended the period of preservation to 48 h, permitting wider geographic sharing of renal allografts, which are allocated principally on the basis of HLA matching. The preservation time for liver and pancreas allografts has also been extended. Preservation times for the various organs are given in Table 9.

Organ	*Hours*
Heart	6
Lung	4
Liver	24
Pancreas	12
Kidney	60

Table 9 Preservation times

The objective of organ preservation is to cool the core temperature of the organ parenchyma so that the demand for oxygen and the requirement for energy (in the form of ATP) can be markedly reduced. During this period of ischemia, cellular integrity of the parenchyma is maintained. These goals can be achieved by flushing of the organs with one of several preservative solutions at a temperature of 47°C.

The constituents of the University of Wisconsin solution include hydroxyethyl starch and lactobionate which suppress cell swelling, glutathione and $MgSO_4$ to stabilize cell membranes during the cold ischemic period, and allopurinol to scavenge oxygen free radicals associated with reperfusion injury and to stimulate ATP synthesis after preservation.

The debate as to whether renal preservation is best achieved by pulsatile perfusion or cold storage has largely subsided. The simplicity of cold storage is a major cost-saving asset. In a randomized study conducted by the New England Organ Bank, renal allografts that were cold stored for less than 36 h functioned as well as the paired kidney preserved by pulsatile perfusion.

Further reading

Alexander J, Zola J. Expanding the donor poor: use of marginal donors from solid organ transplantation. *Clincal Transplantation* 1996; **10**: 1–19.

Amado J *et al.* Blood levels of cytokines in brain-dead patients: relationship with circulating hormones and acute-phase reactants. *Metabolism: Clinical and Experimental* 1995; **44**: 812–16.

American Medical Association Council on Ethical and Judicial Affairs. The use of anencephalic neonates as organ donors. *Journal of the American Medical Association* 1995; **273**: 1614–18.

Anaise D. An approach to organ salvage from non-heartbeating cadaver donors under existing legal and ethical requirements for transplantation. *Transplantation* 1990; **49**: 290–4.

Arras JD, Shinnar S. Anencephalic newborns as organ donors: a critique. *Journal of the American Medical Association* 1988; **259**: 2284–7.

Ascher NL, Bolman R, Sutherland DER. Multiple organ donation from a cadaver. In: Simmons RL, ed., *Manual of vascular access, organ donation and transplantation*, pp. 105–43. Springer-Verlag, Berlin, 1984.

Beecher H *et al.* A definition of irreversible coma: report of the Ad Hoc Committee of the Harvard Medical School to examine the definition of brain death. *Journal of the American Medical Association* 1968; **205**: 337–40.

Belzer FO, Glass NR, Sollinger HW, Hoffmann RM, Southard JH. A new perfusate for kidney preservation. *Transplantation* 1982; **33**: 322–3.

Bernat JL. A defense of the whole-brain concept of death. *Hastings Center Report* 1998; **28**: 14–23.

Case report 32–197. *New England Journal of Medicine* 1997; **337**: 1149–56.

Cho Y, Terasaki P, Cecka M, Gjertson D. Transplantation of kidneys from donors whose hearts have stopped beating. *New England Journal of Medicine* 1998; **338**: 221–5.

Colquhoun S *et al.* Transmission of CNS malignancy by organ transplantation. *Transplantation* 1994; **57**: 970–8.

D'Alessandro A. The non-heart-beating donor. *Graft* 1998; **1**: 23–4.

Delmonico FL. The expanded organ donor controversy. *Graft* 1998: **1**(Suppl. 1): 2.

Delmonico FL, Kaufman DA. Non-heart-beating organ donation: a commentary. *Graft* 1998: **1**: 17–18.

Delmonico FL, Reese JC. Organ donor issues for the intensive care physician. *Journal of Intensive Care Medicine* 1998: **13**: 271–81.

Delmonico FL, Snydman D. Organ donor screening for infectious diseases: review of practice and implications for transplantation. *Transplantation* 1998; **65**: 603–10.

Franz H, Drachman J, DeJong W, Beasley C, Gortmaker S. Public attitudes toward organ donation: implications for OPO coordinators. *Journal of Transplant Coordination* 1995; **5**: 50–4.

Gazdar A. Tumors arising after organ transplantation—sorting our their origins. *Journal of the American Medical Association* 1997; **277**: 154–5.

Gottesdiener K. Transplanted infections: donor to host transmission with the allograft. *Annals of Internal Medicine* 1989; **110**: 1001–16.

Guidelines for the determination of brain death in children. *Annals of Neurology* 1987; **21**: 616–17.

Ho M, Suwansirikul S, Dowling J, Youngblood L, Armstrong J. The transplanted kidney as a source of cytomegalovirus infection. *New England Journal of Medicine* 1975; **293**: 1109–12.

Institute of Medicine, Sciences NAo. *Non-heart beating organ transplantation: medical and ethical issues in procurement.* National Academy Press, Washington, DC, 1997.

Jeevanandam V, Todd B, Regillo T, Hellman S, Eldridge C, McClurken J. Reversal of donor myocardial dysfunction by triiodothyronine replacement therapy. *Journal of Heart and Lung Transplantation* 1994; **13**: 681–7.

Johnson L *et al.* The use of bilateral adult renal allografts—a method to optimize function from donor kidneys with suboptimal nephron mass. *Transplantation* 1996; **61**: 1261–77.

Koogler T, Costarino A. The potential benefits of the pediatric nonheartbeating donor. *Pediatrics* 1998; **101**: 1049–52.

Kootstra G. The asystolic, or non-heart beating donor. *Transplantation* 1997; **63**: 917–21.

Multi-society task force on the persistent vegetative state. Medical aspects of the persistent vegetative state. *New England Journal of Medicine* 1994; **330**: 1499–508, 1572–9.

Novitzky D, Cooper DKC, Chaffin JS, Greer AE, DeBault LE, Zuhdi N. Improved cardiac allograft function following triiodothyronine therapy to both donor and recipient. *Transplantation* 1990; **49**: 311–16.

O'Connor K, Delmonico FL. Organ donation and transplantation from poisoned donors. *Transplantation Reviews* 1999; **13**: 1–4.

Pallis C. ABC of brain stem death: from brain death to brain stem death. *British Medical Journal* 1982; **285**: 1487–90.

Pallis C. Brainstem death: the evolution of a concept. In: Morris P, ed., *Kidney transplantation*, 2nd edn, pp. 101–27. Grune & Stratton, 1984.

Pereira B, Milford E, Kirkman R, Levey A. Transmission of Hepatitis C virus by organ transplantation. *New England Journal of Medicine* 1991; **325**: 454–60.

Pratschke J *et al.* Brain death and its influence on donor organ quality and outcome after transplantation. *Transplantation* 1999; **67**: 343–8.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research Defining Death. US Government Printing Office, Washington, DC, July 1981.

Shewmon A. Recovery from brain death: a neurologist's apologia. *Linacre Quarterly* 1997; **64**: 30–96.

Shewmon D. Chronic brain death. *Neurology* 1998; **51**: 1538–45.

Southard JH *et al.* Important components of the UW solution. *Transplantation* 1990; **49**: 251–7.

Starzl TE *et al.* A flexible procedure for multiple cadaveric organ procurement. *Surgery, Gynecology, Obstetrics* 1984; **158**: 223–30.

Sweet W. Brain death. *New England Journal of Medicine* 1981; **299**: 410–12.

Todres I. Brain death. In: Todres I, Fugate J, eds., *Critical care of infants and children*, pp. 403–5. Little Brown & Company, Boston, 1996.

Truog R. Is it time to abandon brain death. *Hastings Center Report* 1997; **27**: 29–37.

Veatch R. The impending collapse of the whole-brain definition of death. *Hastings Center Report* 1993; **23**: 18–24.

Wijdicks EF. Determining brain death in adults. *Neurology* 1995; **45**: 1003–11.

Wijnen R, Booster M, Stubenitsy B, deBoer J, Heinemen E, Kootstra G. Outcome of transplantation of non-heart beating donor kidneys. *Lancet* 1995; **345**: 1067–70.

16.3 Organ and tissue preservation for transplantation

Vernon Marshall, Alan Saunder, David Scott, Brian Howden and Paula Jablonski

Scope of preservation requirements
Objectives of preservation
Historical review
The problem: effects of ischaemic hypoxia
Ischaemic damage
Reperfusion damage
Hypothermic preservation: a partial solution
Static cold storage
University of Wisconsin solution
Continuous cold perfusion
Normothermic reperfusion
Static cold storage with oxygen delivery
Freezing and vitrification
Phases of clinical preservation
The donor
Extracorporeal storage
The recipient operation
Preservation of individual organs and tissues
Vascularized organs
Summary of major issues in cadaver multiorgan donation
Tissue grafts
The future
Further reading

Scope of preservation requirements

The wide scope of transplantation surgery and its procurement and preservation requirements involve grafting to other sites in the same individual (autografts), between different individuals (allografts), or between different species (xenografts). Grafts of organs and composite tissues (limbs) need vascular anastomoses. Viable autograft and allograft transfer of some tissues (for example skin, cornea, pancreatic islets, other endocrine tissues) is possible without revascularization. Viable non-vascularized tissue transfers also include gametes and embryos. Other non-vascularized tissue allografts (for example bone, cartilage) lose viability but act as useful temporary scaffolding for regenerating host cells to colonize. All these procedures require effective methods of procurement and protection of the tissue, organ, or embryo from removal to reimplantation.

Objectives of preservation

Transfer of disease (sepsis, cancer) must be avoided in all the above circumstances. Effective preservation of function is integral to all techniques involving the transfer of viable organs and tissues. Graft damage must be minimized within the donor in preparation for graft procurement, during graft removal, during storage and transport of the graft, during the transplantation operation, and after reimplantation or revascularization. Some tissues (blood, skin) can be stored for several days or weeks. Vascularized autografts are usually reimplanted immediately, or within a few hours.

Vascularized allografts pose another problem. Not only must viability be maintained but rejection of the grafts must also be prevented by improving the match between donor and recipient tissue and by immunotherapy. These two objectives, preserving viability and preventing rejection, are closely linked. Optimal matching of graft and recipient antigens can require a storage period of 24 h or more. In early years, storage for this length of time was only practicable in kidney transplantation. As prolonged storage becomes more available for liver, pancreas, heart and lung, these grafts may also be allocated on the basis of improved tissue matching. Retrospective analysis of cardiac transplants has indicated that matching would improve the efficacy of the programme, but prolongation of the storage period from the current safe period of 6 to 8 h would be required for this to be achieved.

The process of clinical preservation usually begins with identification of the brain-dead, heartbeating cadaver donor. Effective preservation gains time: time to carry out confirmatory tests of brain death; time to organize operating teams for organ removal; time to type and cross-match donor tissues against a pool of waiting recipients; time to exclude associated transplantable diseases in the donor (particularly viral infections or neoplasms); time to select and locate recipients and arrange their admission and preparation for surgery; time to arrange recipient operating teams; time to transport organs over long distances, even between continents.

Good preservation thus restores order to clinical organ transplantation from the cadaver. Good preservation also provides a ‘window' of diagnostic and therapeutic opportunity during the safe period of extracorporeal storage. The time gained should optimally be available for functional assessment of the graft to confirm its viability, and to predict early function after reimplantation. Good early function is obligatory in the case of heart, lung, liver, and small-bowel grafts. Immediate function of kidney and pancreas grafts, although not obligatory, is highly desirable and facilitates recipient management.

Effective immune manipulation of the recipient or of the graft itself during the period of storage is another (as yet largely unrealized) objective. Treatment of the stored organ to deplete it of passenger leucocytes and other antigen-presenting cells, or other treatments to modify the immunogenicity of parenchymal and endothelial cells, could offer prospects for the dramatic improvement of transplantation results.

Historical review

Attempts to preserve human tissues and organs from putrefaction and decay after death began in antiquity. Preservation of corpses (and comestibles) from the otherwise inevitable rapidity of corruption was most effectively done by cooling and packing with ice. Desiccation by exposure to heat and salt were alternatives in arid climates. Embalming and mummifying techniques were developed to a high art by the Egyptian dynasties.

In the 1930s the classical experiments of Carrel and Lindbergh (Fig. 1) applied perfusion techniques to preservation of organs for transplantation. They established ground rules for organ preservation by continuous *ex vivo* perfusion: expert technology, perfect asepsis, and controlled biological conditions.



Fig. 1. Preservation apparatus of Carrel and Lindbergh.

The Spanish Civil War (1936–9) marked the advent of blood banks and clinical application of tissue-storage techniques. Cold storage diminished metabolic demand. Fortunately, blood, the first widely preserved and transplanted biological substance, lent itself well to extended storage. Refrigerated storage was possible for 3 weeks.

The discovery of cryoprotectants by Polge and coworkers in 1949 ushered in a major extension of preservation times for a variety of simple cells and tissues. Blood cells, gametes (even embryos) could be stored after freezing. Weeks or months later, they could be thawed and successfully reimplanted as autografts or allografts. But in humans, freezing was only successful for isolated cells or undifferentiated, multicellular embryos. The complex and heterogeneous structure of organs, by contrast, rendered them highly sensitive to freezing damage. The biophysical problems of freezing large organs were subsequently defined by Pegg and other workers. To date, no consistent preservation success has been obtained with frozen organs such as kidneys, livers, or hearts. Concepts of frozen humans awaiting revival at an appropriate future time on another planet or in an afterlife remain in the realms of science fiction and quackery. Other species can respond to extreme cold by a freeze/rethaw cycle or by other adaptations. Arctic frogs freeze solid each winter, reviving in the spring thaw. Other warm-blooded animal species (ground squirrels, bears) adapt by varying degrees of winter hibernation or deep sleep, aided by built in cold-tolerant enzyme systems and tissue antifreezes.

Hypothermic organ preservation short of freezing has given more promising results for whole organs in humans. Most early work centred on the kidney. Simple cooling and ice storage gave reliable protection for several hours only. Pegg and Calne in the 1960s showed that preservation could be improved by a cold intravascular flush. Initially the contained blood was replaced with fresh blood, plasma, or extracellular fluid-like flushing solutions containing colloid. In 1969, Collins demonstrated the clear superiority of an ‘intracellular' electrolyte flushing solution, high in potassium, magnesium and phosphate, low in sodium and chloride, and without colloid. Kidney storage for 24 h became clinically practical, provided that the organ did not suffer too much damage by periods of warm ischaemia before cooling. Belzer and coworkers in the early 1970s extended Carrel's work on recirculating machine perfusion. Machine storage extended the time of reliable preservation of kidneys from one day to three days. However, simple hypothermic storage (after flushing) and continuous machine perfusion gave equivalent results for 24-h preservation, particularly if warm ischaemia was avoided. It also became clear that flushing solutions did not need to mimic intracellular composition to be effective. Other solutions, even simpler in composition, based on citrate, sucrose, and other solutes, were shown to be as effective as Collins' solution (Table 1). Clinical kidney preservation and transport could rely mainly on simple hypothermic storage after preliminary flushing with these solutions (Fig. 2); compact cold perfusion machines using modifications of plasma were also available for more extended storage times. Belzer, Southard, and coworkers at the University of Wisconsin gave a further notable stimulus to preservation research in 1987, demonstrating that a complex, multicomponent solution [University of Wisconsin (**UW)** solution] gave significantly enhanced preservation of several organs (particularly pancreas and liver). With the addition of a synthetic colloid (hydroxyethyl starch), the UW solution could be made suitable for both simple flushing and recirculatory machine perfusion. Modifications of UW solution have been successful in extending preservation of pancreas, liver, kidney, heart, lung, and small bowel (see below).

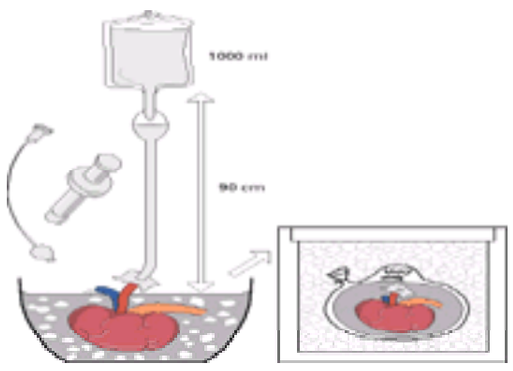


Fig. 2. Organ preservation by static ice storage after cold flushing.

	EuroCollins	*Isotonic citrate*	*PBS*
(mmol/l)			
Sodium	10	80	120
Potassium	115	80	
Magnesium		40	
Citrate		55	
Bicarbonate	10		
Chloride	15		
Phosphate	58		60
Sulphate		40	
Glucose	180		
Sucrose			140
Mannitol		100	
Osmolality (mosmol/kg)	340	300	310
pH (at 0°C)	7.3	7.1	7.2

Table 1 Early organ-preservation solutions: EuroCollins', citrate, phosphate-buffered solution (PBS)

The problem: effects of ischaemic hypoxia

In the cells of normally functioning organs, energy is derived from the oxidation of substrates obtained from the circulating blood (glucose, fatty acids, amino acids, or ketones), or in liver and muscle from breakdown of endogenous glycogen. Energy is stored within cells as phosphate bonds of adenosine triphosphate (**ATP**), creatine phosphate, and other nucleotides. Cellular composition is maintained in homeostasis by numerous enzymatic reactions acting in concert under the directions of hormones or key messenger compounds such as cAMP.

Ischaemic damage

Ischaemia cuts off nutrients and oxygen supply to the organ. The continuing metabolic activity of the organ's parenchymal and other cells causes a cascade of events leading to irreversible cell damage and death. Energy metabolism becomes anaerobic; anaerobic glycolysis causes depletion of high-energy phosphate compounds and only generates two molecules of ATP per glucose molecule, compared with 38 ATPs from aerobic metabolism by oxidative phosphorylation. Degradation of ATP leads to an increased cellular concentration of adenosine, inosine, and hypoxanthine. Depletion of the cell's energy stores inactivates the enzyme systems (Na–K and Ca–Mg ATPases) controlling the active transport pumps of the cell membranes. As fuel reserves disappear, sodium and chloride, freely permeable electrolytes that are normally excluded actively from the cell, now diffuse into the cell down concentration gradients. The osmotic force of non-permeable cellular proteins and anions is no longer balanced by the extrusion of sodium. Water therefore also floods the increasingly swollen cell. Mitochondrial respiration is inhibited. Calcium enters the cytosol and mitochondria. Anaerobic metabolism temporarily uses glucose stores to generate ATP, but lactic acid is also produced, leading to progressive intracellular acidosis; while possibly being initially cytoprotective this ultimately activates lysosomal lytic enzymes, leading to autolysis.

Most organs and cells can tolerate ischaemic hypoxia for 30 to 60 min without permanent damage. Parenchymal cells of most transplantable organs are generally similar in their tolerance to ischaemia. Rapidly metabolizing tissues have lesser tolerance. The heart is particularly vulnerable: it retains ATP more effectively by anaerobic metabolism than does liver or kidney, but the ischaemic heart continues to beat until all energy reserves are depleted. Most organs are irreversibly damaged by 90 to 120 min of ischaemia at body temperatures (warm ischaemia). Vascular damage occurs along with the parenchymal effects. The cells of the vascular endothelial lining bear the brunt of this injury. These effects are slowed, but not reversed, by cooling (Fig. 3).

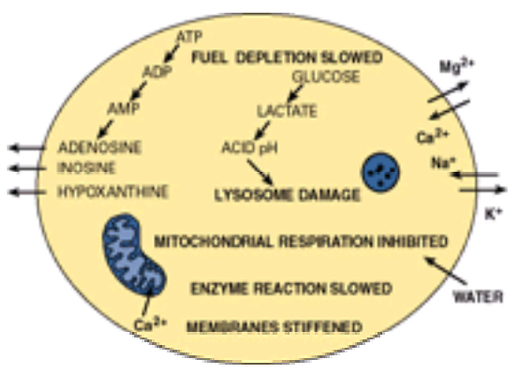


Fig. 3. Cellular effects of cold ischaemia.

Reperfusion damage

Reperfusion injury (Fig. 4) is an added hazard contributing to irreversible damage. Accumulation of metabolic end-products of ATP such as hypoxanthine, under anaerobic conditions, can set the stage for reperfusion injury. When blood flow is restored, oxygen influx leads to the formation of toxic compounds such as hydrogen peroxide, and superoxide and hydroxyl radicals. These active free radicals of oxygen produce further cellular, membrane, and microvascular endothelial injuries. Under normal circumstances the formation of these harmful radicals is short lived, as any formed are cleared by endogenous scavenging mechanisms. Ischaemia depletes these endogenous mediators. Reperfusion injury from blood inflow also activates inflammatory cell messengers leading to platelet aggregation, neutrophil sequestration, and vasoconstriction. After prolonged ischaemia, complete cessation of flow rapidly follows neovascularization (‘no reflow' phenomenon).

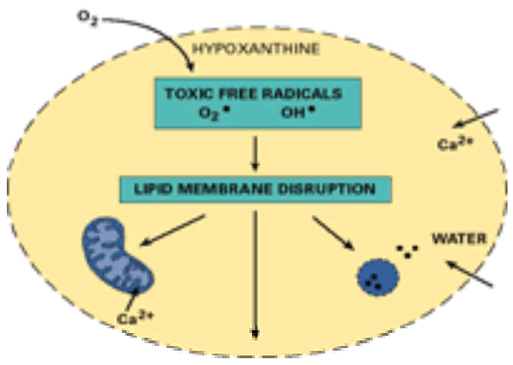


Fig. 4. Reperfusion damage after ischaemia.

Hypothermic preservation: a partial solution

Static cold storage

Simple cooling markedly enhances ischaemic tolerance. All enzymatic activity is temperature dependent: cooling diminishes metabolic activity, curtails oxygen demand, and slows degrading of energy stores. Hypothermia does not stop metabolism, it merely slows the metabolic clock and lessens the gradient at which deterioration occurs. Cooling from 37°C (body temperature) to 0°C (storage temperature) extends the tolerance of most organs to ischaemia from 1 to 2 h to about 12 h.

Unfortunately, cold does not effect a uniform slowing of all biological functions. Hypothermia causes discord in a variety of metabolic processes that occur in concert at 37°C. Transmembrane passive diffusion of ions is not appreciably affected by hypothermia, while active transport mechanisms (for example those governed by Na–K and Mg–Ca ATPases) are inhibited below 10°C. Hypothermia, alone, cannot prevent cell swelling during storage (Fig. 3).

A major requirement for a cold flushing solution is therefore to include an impermeant solute to provide a countering osmotic force against cellular oedema. Large anions such as lactobionate (358 kDa) and gluconate (195 kDa), or non-electrolytes such as the saccharides raffinose (505 kDa) or sucrose (342 kDa), or chelates of citrate and magnesium (approx. 1000 kDa), or protein (70 000 kDa) or amino acids (histidine) can achieve this. Glucose (180 kDa) permeates the cell slowly and can stimulate an undesirable production of lactic acid and hydrogen ions by anaerobic glycolysis. Glucose in flushing solutions can be replaced by impermeant sucrose with benefit; especially for liver and pancreatic grafting, where cell membranes are even more permeable to glucose.

Countering intracellular acidosis with an effective buffer (phosphate, citrate, histidine, bicarbonate, HEPES) is a second major requirement. The importance of these two mechanisms is indicated by the fact that a solution consisting solely of sucrose with an added phosphate buffer (PBS) is remarkably effective in kidney preservation, almost as effective as the very much more complex UW solution.

The electrolyte composition of flushing solutions can vary widely. The freely diffusible anion chloride is preferably replaced by an impermeant anion (lactobionate, gluconate, or chelated citrate). Flushing solutions initially had high potassium (100–130 mmol/l) and low sodium (10–30 mmol/l) concentrations. The high potassium is cardioplegic. These solutions thus cannot be used for early systemic intravenous infusion, and a large systemic bolus from the transplanted organ after revascularization in the recipient could be dangerous. A high potassium concentration is also vasoconstrictive, slowing the flushing and thus the rate of organ cooling. Solutions with the sodium:potassium ratio reversed (Na 130, K 10–30 mmol/l) are as effective and safer, provided suitable impermeants and buffers are present in the solution, and chloride is kept low.

Magnesium has been a useful additive in many solutions; it forms impermeable chelates with lactobionate and citrate. By contrast, calcium has been kept low or excluded. Cell damage is associated with calcium influx (Fig. 5). Calcium (and magnesium) can precipitate in unstable solutions. Calcium is strongly chelated by lactobionate (and citrate), which may account partly for the usefulness of these substances in flushing solutions.

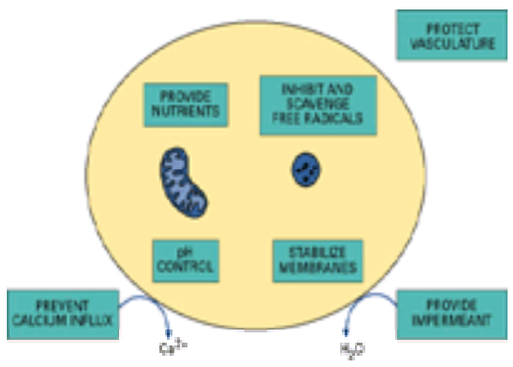


Fig. 5. Requirements of flushing solution for optimal organ preservation

Preservation by simple hypothermic storage is limited by the ultimate exhaustion of nutrients and accumulation of waste products. The preservation of kidneys can be extended to 5 days by simple storage. Preservation of liver and pancreas has proved more difficult, and was not consistently obtained at 24 h until the advent of UW solution.

University of Wisconsin solution

Multiple factors influencing the effectiveness of preservation were brilliantly and serendipitously combined by Southard, Belzer and coworkers at the University of Wisconsin, who produced a solution (UW) containing impermeants (raffinose, lactobionate), buffering capacity (phosphate), free-radical inhibitors and scavengers (glutathione, allopurinol), energy precursors (adenosine), vasoactive agents and hormones (steroids, insulin), and a colloid (hydroxyethyl starch) (Table 2). This complex solution with 13 different constituents was markedly more effective than others (Collins', citrate, phosphate-buffered sucrose (PBS)) for liver and pancreas preservation, and also for kidney and heart grafts. It could be made suitable both for cold-flush storage and continuous perfusion. Not all its components have been found subsequently to be equally important. In particular, hydroxyethyl starch can safely be omitted for cold flushing and simple hypothermic storage. High potassium is

undesirable on several grounds; reversing Na:K ratios does not significantly lower efficacy. Other additives such as the buffer histidine can improve preservation; the polysaccharide raffinose can be replaced by sucrose. Lactobionate is the most effective and most important component of UW solution.

	UW solution	Modified UW solution*
(mmol/l):		
Lactobionate	100	100
Raffinose	30	60
Sodium	30	125
Potassium	125	30
Magnesium	5	5
Chloride	10	10
Phosphate	25	25
Sulphate	5	5
Adenosine	5	5
Allopurinol	1	1
Glutathione	3	3
Insulin (units/l)	100	100
Dexamethasone (mg/l)	8	8
Hydroxyethyl starch (g/l)	50	
Reactive O_2	0.5	0.5
Osmolality (mosmol/kg)	320	320
pH (at 0°C)	7.4	7.4

*Hydroxyethyl starch-free. Shelf life extended.

Table 2 University of Wisconsin (UW) preservation solutions for simple cold storage

Continuous cold perfusion

Continuous cold perfusion by a machine combines the benefits of hypothermia with continuous buffering, continuous wash-out of accumulating toxic metabolites, and continuous provision of oxygen and nutrients. Oxygen is more soluble at lower temperatures, and the lesser energy requirement of hypothermic organs can be provided by acellular perfusates. For all organs studied, viability has been maintained longest by continuous machine perfusion. A heat exchanger is needed to maintain hypothermia at 4 to 9°C, and an atraumatic pump, with oxygenation either by simple surface diffusion or by a membrane oxygenator. Maintaining a constant pH requires an oxygen/carbon dioxide gas mix and an effective buffer. A bubble trap and organ chamber complete the circuit (Fig. 6). Materials coming into contact with the fluid must be sterile, non-toxic plastics or metals. The perfusate must contain colloid to maintain intravascular volume and prevent an ‘exploded' extracellular fluid space. Cryoprecipitated plasma or albumin were initially used, but albumin gradually leaked into the interstitium and led to weight gain, so was replaced by other oncotic colloids (gelatine, dextrans, hydroxyethyl starch, polyethylene glycol). Metabolic fuels improve long-term preservation; glucose has been the most commonly used substrate. Adenosine, adenine and other precursors, or ATP itself, have been used to provide additional sources of energy. Addition of other substrates more specific to individual organs can also improve function after storage (carbohydrates, ketones, pyruvate, essential amino acids, fatty acids). Fuel additives become increasingly important as preservation times are prolonged beyond 24 h.

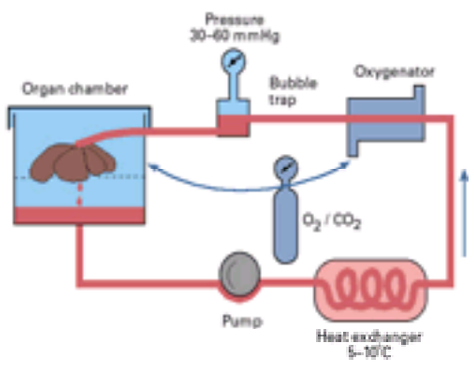


Fig. 6. Principles of continuous hypothermic machine perfusion.

Hypothermic cell swelling due to inhibition of membrane pumps still occurs with cold perfusion below 15°C unless chloride is replaced by impermeants.

Limitations of continuous perfusion develop after about 5 to 7 days. By this stage considerable damage has usually occurred to the vulnerable vascular endothelium, leading to irreversible cellular damage on replantation.

Normothermic reperfusion

Organ perfusion at normal temperature can mimic normal function, but normothermic perfusion as a means of storage has not been possible for prolonged periods without seriously damaging the organ. Kootstra and his colleagues demonstrated that intermittent, brief, normothermic blood perfusion during cold storage improved the preservation of both ice-stored and machine-perfused kidneys. The mechanisms of action are uncertain, but probably relate both to replenishing cellular energy stores and protecting the vasculature.

Static cold storage with oxygen delivery

Continuous persufflation with gaseous oxygen, delivered to the cold-stored organ via its blood vessels or ducts, can partially replenish fuel depletion during storage. The technique is technically simpler than continuous machine perfusion, and has been used to resuscitate kidneys and livers from marginal donors.

Alternatively, oxygen can be delivered during storage without perfusion by suspending the stored organ at the interface between two preservative solutions, one containing a perfluorochemical polyol emulsion providing continuous oxygen transfer. Experience has been mainly confined to pancreas transplantation. Both methods can increase reserves of high-energy phosphate during preservation.

Freezing and vitrification

Freezing is normally lethal to cells. Red cells, lymphocytes, spermatozoa, fertilized ova, embryos, and pancreatic islet cells can all be frozen and thawed satisfactorily using cryoprotectants. Simple tissues such as skin and cornea can also be preserved by freezing. Unfortunately, the freezing of organs, so attractive in concept in approaching true suspended animation and indefinite storage, has not proved feasible to date. The low surface area:volume ratio of organs compared to cells makes effective heat exchange a major problem. Cryoprotectants such as glycerol and dimethyl sulphoxide need high concentrations for adequate cryoprotection and this causes severe organ toxicity and vascular damage. Extracellular ice formation is innocuous when freezing cell suspensions, but can disrupt and severely damage whole organs. The vascular system is especially vulnerable, and the attachment of vascular endothelial basement membrane is disrupted by freezing.

Total vitrification of tissues is an alternative approach, but is also unsuccessful for whole organs. The tissue is first perfused with high concentrations of several cryoprotectants, so that on cooling neither intracellular nor extracellular freezing occurs. The solution vitrifies (solidifies into a glass state) at temperatures of approx. -120°C.

Phases of clinical preservation

The donor

Organ function must be protected before and during organ procurement from the donor. Living donor grafts (e.g. kidney) are protected by maintaining blood and interstitial fluid volumes, by inducing a diuresis with mannitol during the operation, and by avoiding vascular trauma and vasospasm during mobilization and removal of the organ. Warm ischaemia is kept to a minimum.

Management of the heartbeating cadaver donor follows similar protocols. Blood volume is carefully maintained by transfusion, and when necessary by monitoring central venous or pulmonary wedge pressures. Diabetes insipidus is common in brain-dead patients; large volumes of fluid and careful monitoring are required. Over-transfusion and pulmonary oedema must be avoided in patients serving as heart and heart–lung donors. Dopaminergic renal support (2–5 mg/kg per minute) is given. Systemic acidosis must be corrected. Mannitol 25 g, steroids 1 g, chlorpromazine 500 mg, and heparin 10 000 U are given just before *in situ* aortic flush-cooling. Nifedipine should be added if adrenaline (epinephrine) or noradrenaline (norepinephrine) have been given to the donor. Nitric oxide (**NO**) is produced by nitric oxide synthase using L-arginine as substrate, and regulates cellular metabolism and vascular smooth-muscle relaxation. During ischaemia, reduced release of NO contributes to vasoconstriction, platelet aggregation, and increased vascular resistance. Exogenous L-arginine administration during donor organ procurement has diminished vascular damage in liver transplants.

Expeditious and skilful organ removal from a well-prepared, well-hydrated cadaver donor with minimal warm ischaemia is the best guarantee of immediate organ function after grafting; the state of the organ before its removal in such instances becomes the best guide to subsequent early function.

The multiorgan donor

In countries where kidney, heart, lung, liver, and pancreas transplantation is available, up to 80 per cent of all cadaveric donors serve as multiorgan donors. Criteria for the acceptance of donors for heart, lung, liver, and pancreas are more restricted than for kidney transplantation. Transplanted hearts, lungs, and livers must function immediately to support life, whereas kidney recipients can be supported by dialysis while temporary tubular necrosis recovers. An example of criteria required for kidney, liver, pancreas, and heart grafts is set out in Table 3.

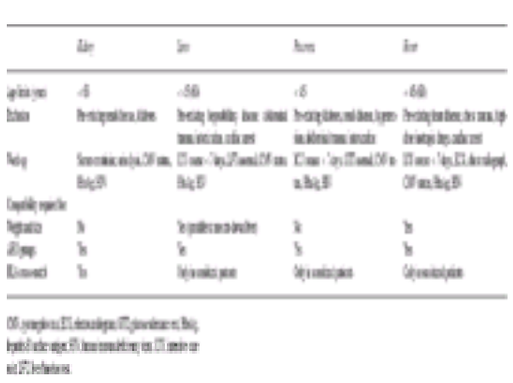


Table 3 Multiorgan donor: organ-specific criteria—kidney, liver, pancreas, heart

General criteria include a donor free from the following.

- Cancer, except for treated skin cancers and malignancies of the central nervous system: systemic embolization of primary brain tumours can occur after ventriculoatrial shunting; such patients should not be used as donors.
- Systemic sepsis (which is particularly likely to occur with a prolonged intensive-care course).
- Hepatitis B surface antigen and hepatitis C virus.
- Human immunodeficiency virus antibodies: any patient with a known or suspected history of homosexual activity or intravenous drug usage is excluded; the patient may be in the latent period before antibodies develop.

A full autopsy should be done following cadaveric donation. Occasionally, unsuspected disease in the donor, such as tuberculosis, indicates the need for a period of drug prophylaxis in the recipient. Occult cancers may also be identified.

Techniques of multiorgan retrieval are outlined elsewhere. The donor organs are skeletonized. All organs require the induction of rapid cooling to restrict initial damage from warm ischaemia. The heart is stopped by infusion of a cold cardioplegic preservation solution, and heart and lungs are removed first. Rapid cooling of other organs is initiated by *in situ* cold flushing via the aorta and portal vein. Liver, pancreas, then kidneys, are removed in sequence during *in situ* flushing. Once the organs have been cooled and removed, the dissection is completed in a bath of ice slush to prepare the vascular pedicles of each organ for transplantation. A final wash-out with 1 to 2 litres of the organ-preserving solution (UW, UW derivatives, histidine–tryptophane–ketarate, EuroCollins, citrate, phosphate buffered sucrose) is done at this stage until the effluent is macroscopically clear and the organ uniformly pallid.

The asystolic (non-heartbeating) cadaver donor

Continuing shortage of suitable heartbeating cadaver donors dying in hospital has led to use of marginal donors presenting with cardiac arrest. Special techniques of intra-abdominal organ cooling by vascular cold perfusion or by intraperitoneal dialysis lavage have been employed to minimize the otherwise inevitably prolonged ischaemic damage. The technique has been used predominantly for kidney grafts, and although the incidence of acute tubular necrosis is higher in such circumstances, long-term function of grafts has been encouraging.

Extracorporeal storage

Simple cold storage

Most organs are subsequently immersed in cold preserving solution and stored at 0 to 4°C in a refrigerated container until reimplantation.

Perfusion storage

Organs are perfused continuously with a recirculating perfusate at 4 to 6°C. This technique is required less commonly because of the increased effectiveness of preserving solutions for simple storage. Perfusion storage is, however, the preferred technique for grafts from marginal donors.

Kidneys are frequently transplanted within 24 h. This period can be extended safely for periods up to 36 to 48 h or longer, if such time is required to find and prepare an optimal recipient and to transport the refrigerated organ over long distances. Liver and pancreas preservation with storage in UW solution is effective for 12 to 24 h. Heart and heart–lung transplantation with current preservation solutions is restricted to a safe storage period of 5 to 8 h.

The recipient operation

Management of the donor is manifestly of great importance to subsequent graft function. Factors in the recipient during operation and after reperfusion may be equally significant. Failure of blood flow to return uniformly to all portions of previously ischaemic tissues (‘no reflow' phenomenon) is a recognized sequel of extended ischaemic damage to organs *in situ*, to autografts and to allografts. This phenomenon has been described after ischaemia of kidney, heart, muscle, brain, and other tissues. Much attention has therefore focused on potential damage suffered by the graft during the implantation operation, and in the early period after revascularization.

Delayed restoration of blood flow after reperfusion can contribute significantly to delayed graft function. Reperfusion of the stored organ, rather than marking the welcome end of its ischaemic insult, may exacerbate the effects of ischaemia and lead to further cellular and vascular damage. The severity of reperfusion injury is dependent on the events of retrieval and storage. Injury can follow prolonged warm ischaemia or extended cold preservation by either static or perfusion storage. The pathophysiology relates to the continuing effects of oligaemic (hypoxic) hypothermic events during organ retrieval and storage under the fresh stimulus of the return of oxygenated blood flow (Fig. 4). Pretreating the recipient, as well as the donor, can obviate some of these effects. Reperfusion injury can be exacerbated by insidious rewarming of the organ during reimplantation. Grafts should be kept cold by moistened cold packs during the recipient operation and vascular anastomoses performed expeditiously. The addition of interposition arterial or venous ‘jump' grafts or other vascular repairs should be done as bench surgery on the preserved chilled organ.

Technical failures of vascular anastomoses, requiring reclamping and revision, are potent contributors to reperfusion injury.

Preservation of individual organs and tissues

Vascularized organs

Kidney

Simple hypothermic storage has employed a variety of flush solutions.

Collins' solution, with high concentrations of potassium, magnesium, phosphate, sulphate and glucose, extended preservation to 48 h, but precipitation of magnesium phosphate was a major deficit. Subsequently, magnesium sulphate was omitted (EuroCollins' solution) with no deleterious consequence. Replacement of glucose by mannitol, sucrose, or raffinose improved function in experimental animals and in humans.

Citrate-based solutions (Marshall/Ross) containing high concentrations of potassium, magnesium, citrate, sulphate and mannitol, were originally devised to overcome the limitations of Collins' solution. Both hypertonic and isotonic solutions are stable and provided successful clinical preservation for 48 h or more. A high magnesium concentration was essential; magnesium–citrate chelates provided the critical semipermeant component.

Bretschneider's histidine–tryptophane–ketarate (HTK) solution used amino acids as major buffers and impermeants.

Phosphate-buffered sucrose (PBS) is an isotonic solution containing only a phosphate buffer and sucrose, an impermeant solute. Clinically and experimentally this solution is highly effective. Preservation at 72 h was better experimentally with phosphate-buffered sucrose than with Collins' or citrate solutions. Similar results were seen with prospective, randomized clinical trials.

UW solution was developed initially for pancreatic preservation. Subsequently, it was found profoundly to extend and improve preservation of the liver; UW became the preferred solution for multiorgan harvest, despite its higher cost. UW solution is highly effective in preserving kidneys. Modifications of the solution, omitting the colloid (hydroxyethyl starch), are equally effective. UW-derived solutions (hydroxyethyl starch-free) can provide successful preservation by simple cold storage of dog kidneys for 5 days. High potassium concentration is not essential, and lactobionate can be replaced by gluconate. Clinically, multicentre trials have indicated that UW solution results in a more rapid reduction of serum creatinine, a higher creatinine clearance rate and less dialysis when compared to EuroCollins (median preservation time 24 h, maximum 48 h). An additional advantage was that sharing kidneys between centres to improve matching improved graft survival without any functional detriment up to 48-h storage.

Prospective, randomized clinical trials comparing simple cold storage and machine preservation using preservation times averaging 24 h have shown no significant differences in early function or long-term survival. With more extended cold storage, kidneys show increasing frequency of acute tubular necrosis as storage times exceed 36 h. Further extension of preservation requires additional resuscitative technology. Perfusion storage can resuscitate kidneys that have been subjected to adverse conditions before being harvested or during storage.

Liver

Liver transplantation has been an established and effective treatment for endstage liver disease since the early 1980s. The pioneering development of liver transplantation by Starzl and Calne involved heroic organizational feats to minimize the ischaemic period. Livers were only able to be stored reliably for 12 h by Collins' or citrate solutions. Preservation of the liver with these solutions was less effective than for the kidney: the hepatocyte is more permeable to glucose and mannitol, leading to reduced osmotic control and increased acidosis due to anaerobic glycolysis (EuroCollins). Pretreatment of the donor with chlorpromazine, and addition of chlorpromazine, diltiazem, or a stable prostacyclin analogue to the flush solution improved preservation.

A major advance in liver preservation occurred with introduction of UW solution. This solution provided effective preservation experimentally in the dog (48 h) and rat (30 h). Safely extended clinical preservation times up to 24 h allowed procurement of liver grafts from distant cities, ample time for histopathological examination of the graft, more perfect recipient hepatectomy, and bench surgery of the graft to tailor adult grafts to fit one or more child recipients.

Experimental studies indicated that essential ingredients of UW were the impermeant anion lactobionate with additional osmolality provided by raffinose. Adenosine, allopurinol, and glutathione seemed also beneficial. Omission of hydroxyethyl starch from UW solution was not detrimental to dog, rat, or human liver grafts. High potassium content was shown to be unnecessary in 48-h preservation of the dog liver. Other buffers such as histidine could replace phosphate.

The hepatocyte is reasonably tolerant of cold ischaemia; the principal damage is to the microvasculature and to the endothelial cells lining the sinusoid. Experimental studies indicate that, although parenchymal cells remain viable with preservation for 48 h, a high proportion of endothelial cells are non-viable and this contributes to poor early function and mortality. UW solution delays but does not wholly prevent cold-induced microcirculatory injury.

More prolonged preservation required perfusion preservation. Continuous perfusion with modified UW solution can successfully store dog livers for 72 h: gluconate replaced lactobionate in this solution and calcium, adenine, glucose, and ribose were included.

Clinical practice requires immediate function for liver grafts. Currently, using simple cold storage, 80 per cent or more grafts function adequately; about 10 per cent show initially poor graft function; and about 5 per cent of grafts show permanent non-function, necessitating life-saving retransplantation. Factors influencing and responsible for poor initial function include fatty change in the donor liver, older donor age, preliminary liver-reduction surgery, and longer cold storage. Although clinical preservation times of 24 to 30 h are possible, it is thus preferable to keep storage times to under 12 h for optimal results. Clinical best practice remains cold storage using UW-derived solutions. When using high-potassium UW solution for preservation, a prevascularization rinse with a further wash-out solution is desirable; solutions designed to protect endothelial function are preferable to the widely used Ringer's lactate rinse. Important considerations, whatever the composition of the solution, are that a brief prevascularization rinse should be performed at room or body temperature; and that an extracellular fluid-like crystalloid composition (Carolina rinse) can be enhanced by the addition of adenosine, allopurinol, desferrioxamine, and glycine. Glucose, fructose, or other nutrients are probably unnecessary in either the cold preserving solutions or the pretransplant rinse.

Assessment of graft quality before implantation by measurement of ATP content, pH, energy charge, or morphology is not necessarily reliable; many of the detrimental changes occur after reperfusion. The volume and quality of bile flow, restoration of clotting, absence of lactate acidosis, blood amino acid clearance, and plasma concentrations of bilirubin and aspartate aminotransferase can distinguish grafts that will recover from those that will not.

Pancreas

Initial methods for pancreas preservation were based on those developed for kidney. The pancreas is more vulnerable to mechanical damage during retrieval, but no greater sensitivity was apparent to either warm ischaemia or cold storage. The problems encountered in pancreatic grafting related to its low circulatory flow rate and to vascular damage on storage and reperfusion. Thrombosis of vascular anastomoses was more likely with poor preservation.

UW solution was, as described above, originally introduced to improve pancreas preservation, which requires preservation of the essential insulin-producing b-cells. Exocrine cell function can provide a marker of viability in the form of amylase excretion. Another important aim of preservation was to minimize the occurrence of ischaemically induced acute pancreatitis, which adds morbidity after transplantation. Preservation of the pancreas using Collins', citrate and other solutions proved difficult, and it was not until the advent of UW solution that reliable preservation for 24 h was achieved experimentally and clinically. Experimental preservation has since been extended to 96 h using UW with additives that inhibit nitric oxide synthase and diminish the production of endothelin and thromboxane. Other variations of UW also gave satisfactory cold storage: variants include removal of hydroxyethyl starch or its replacement with dextran, and combinations of sodium lactobionate with histidine.

The pancreas has also been studied extensively in relation to the benefit of additional oxygen using a two-layer storage technique with a perfluorochemical adjacent to the UW solution. The method enhanced energy reserves and was protective of endothelial function. In small animals, the actual technique of cold storage, whether surrounded by the preserving liquid or wrapped in moistened gauze, also influenced efficacy of preservation. This may relate to the filamentous and non-encapsulated nature of the pancreas, resulting in a greater tendency for harmful oedema and weight gain to occur during storage.

Preservation needs to include the associated duodenal segment containing the duodenal papilla and pancreatic ducts. Fortunately, the duodenum has not proved to be

more sensitive to preservation injury than the pancreas itself.

Clinically, graft survival has not been affected by extending preservation times to 24 h. Grafts stored for over 30 h showed somewhat decreased graft survival. Retrospective analysis of more than 9000 grafts from the International Transplant Registry showed that patients with better HLA-DR matching had significantly better graft survival. Prolonged effective preservation would allow allocation of grafts after matching.

Perfusion storage is inherently more difficult in the pancreas because of its low circulatory flow rate. Early studies using machinery designed for kidney perfusion were unsuccessful.

Heart

Heart preservation clinically has been mostly confined to simple, static cold storage. As 85 per cent of energy is consumed by contraction of the myofibrils, cardioplegic arrest has been an essential feature of heart preservation. The development of heart transplantation from open-heart surgery had a considerable influence on the strategies applied in heart preservation. Cardioplegic solutions developed for cardiac surgery provided safe preservation of myocardial function for only 5 to 6 h.

For transplantation, the heart was excised after the induction of hypothermic cardioplegic arrest by *in situ* flushing with one of the cardioplegic solutions used in open-heart surgery (Table 4). The graft was then stored cold in the flush-out solution. Sinus node dysfunction (usually transient) was common after transplantation and was related to the duration of ischaemia. Microvascular injury was reported in the presence of well-preserved myocytes.

Solution	Main constituents
St Thomas's	Na 120, K 16, Mg 16, Ca 1.2, Cl 160, HCO$_3$ 10; osmolality 320
Stanford	Na 125, K 10, Cl 98, HCO$_3$ 25, mannitol 25, glucose 255; osmolality 400

Table 4 Early cardioplegic solutions: ‘ECF (extracellular fluid)' crystalloid composition

Flushing solutions commonly used for the preservation of other organs are also cardioplegic, but generally contain much higher concentrations of potassium (over 100 mmol/l instead of 20–30 mmol/l) and do not contain any calcium. Initial reluctance to apply such solutions to heart preservation stemmed from earlier studies indicating that potassium-induced contraction band necrosis occurred with potassium concentrations above 40 mmol/l, and that a ‘calcium paradox' effect in cardiac preservation was important (cardiac muscle incubated in calcium-free medium undergoes severe, irreversible damage when reperfused with calcium-containing medium due to a massive influx of calcium). Damage from warm ischaemia enhances calcium influx. The occurrence of calcium paradox at hypothermia is controversial; but removal of calcium from St Thomas's solution was shown to be detrimental, and the calcium paradox has been seen in experimental cardiopulmonary bypass.

Modifications of the UW solution (Table 5) are now successfully used clinically in open heart surgery and for transplantation, following extensive experimental studies. The range of available experimental models included heterotopic transplants to the abdomen or neck in small animals; isolated, perfused, working heart models; metabolic tissue analysis and histology; nuclear magnetic resonance spectroscopy; and allograft function in larger animals. Specimens of human atrial myocardium can be obtained for study during open heart surgery. Successful orthotopic transplants have been reported after 12 to 24 h preservation with UW-derived solutions in non-human primates, and randomized clinical trials have in general confirmed better cardiac preservation with these solutions.

Solution	Main constituents
Cellsor	Lactobionate, mannitol, histidine, glutamate, glutathione Electrolytes 'ECF': Na 100, K 15, Mg 13, Ca 2.5, Cl 40
Columbia University	Glucose, dextran, hydroxyl compounds, glucose, adenosine, cysteine, cAMP, nitroglycerine, verapamil Electrolytes 'ECF': K 120, Mg 5, PO$_4$ 25
Histidine–lactobionate	Histidine, lactobionate, ketoacids, tryptophan Electrolytes 'ECF'
Sodium lactobionate–sucrose	Lactobionate, sucrose, glutathione Electrolytes 'ECF': Na 100, K 5, Mg 5, PO$_4$ 25

ECF, ECF extra-, extracellular fluid

Table 5 University of Wisconsin-derived cardiac preservation solutions

Maintenance of high-energy phosphates (ATP and creatine phosphate) has been another goal, but no linear correlation exists between the concentration of adenine nucleotides after storage and the functional outcome. Below a certain concentration of these high-energy compounds, no functional recovery occurs, but inadequate preservation may occur at normal concentrations. The addition of ATP or its precursors to preserving solutions can improve cardiac function, but this does not correlate with maintenance of tissue ATP concentrations and may have been due to the vasodilatory properties of the additives.

Amino acid enrichment of solutions from added aspartate and arginine enhances ischaemic tolerance. The addition of calcium-channel blockers also improves preservation; these agents may also minimize post-transplant atherosclerosis. The generation of oxygen free radicals has been implicated in ischaemic heart disease and in reperfusion injury. Metabolic inhibitors and free-radical scavengers have been used to pretreat the donor and as additions to the preserving solution. Improvements in left ventricular function, lipid peroxidation, and platelet aggregation have been demonstrated. Prostacyclins dilate coronary vessels, inhibit platelet aggregation, and stabilize lysosomal membranes. A stable prostacyclin analogue added to the cardioplegic solution improved ventricular function after clinical cardiac transplantation.

Continuous microperfusion preservation at low flow rates with oxygenated perfusate improved the preservation of rabbit hearts using a modified UW solution incorporating polyethylene glycol (PEG20). A colloid is essential in machine perfusates, as for other organs.

Heart–lung

Simple cooling has afforded short-term preservation for heart–lung transplants. Lungs were initially flushed with cold EuroCollins' solution via the pulmonary artery immediately after cardioplegic arrest of the heart. Flushing may be a source of injury: care was taken to maintain pressure below that normally found in the pulmonary artery; lung distension was maintained by rhythmic ventilation at 50 to 70 per cent of normal inflation throughout the period of storage. Prostacyclin, superoxide dismutase, and catalase were beneficial when added to EuroCollins' solution.

Clinically, treatment of the donor with prostaglandin E_1 followed by simple hypothermic flushing with EuroCollins' solution provided adequate 6-h preservation. In experimental heart–lung transplants in the dog, treatment of the recipient with superoxide dismutase also improved preservation. UW-derived solutions with a lower potassium concentration improved lung function. High potassium produced pulmonary vasospasm and reduced the efficacy of the flush; low-potassium dextran solution reduced ischaemic/reperfusion damage.

Donor cooling and cardiopulmonary bypass with diluted blood is more complex and has only been effective experimentally for 6 h. The addition of methylprednisolone,

prostaglandin, and isoproterenol improved lung function.

Autoperfused heart–lung preparations at 30°C have afforded the longest lung preservation (12 h), but this procedure is cumbersome and limited in use. Adding metabolic substrates (glucose, ribose) enhanced preservation, as did leucocyte depletion. Lung oedema, activation of complement, and subsequent pulmonary sequestration of leucocytes have been implicated in lung dysfunction.

Lung

Transplantation of the lung alone has been increasingly successful clinically. Lung preservation was only successful for 5 to 6 h using a EuroCollins' flush. After 12 h of storage, pulmonary vascular resistance increased threefold and compliance was reduced; after 24 h the lung was haemorrhagic and oedematous. Addition of methylprednisolone to the donor, the preserving solution, and the recipient extended preservation to 12 h but not to 24 h.

For most organs, preservation techniques initially evolved to maintain the viability of metabolically active cells (hepatocytes, myocytes, renal tubules) rather than the microvasculature, but a well-preserved microcirculation is essential for adequate lung function. Blood perfusion at 25 to 30°C with the addition of metabolites may afford optimum vascular preservation, without an irreversible energy deficit affecting parenchymal cells. Pulmonary surfactant produced by alveolar type II cells is important in microvascular damage. Aprotinin, widely used in open heart surgery to reduce perioperative blood loss, is a membrane stabilizer that may improve endothelial cell function when added to preservation solutions.

Small intestine

Clinical transplantation of the small intestine has had limited but improving long-term success (3-year graft survival is approx. 40 per cent). The hazards of intestinal transplantation preclude it from being as yet an established alternative to home parenteral nutrition. The major problems are rejection and graft-versus-host disease.

Experimental preservation and transplantation of the small intestine was attempted from the late 1950s. Preservation for 24 h, after flushing both the vasculature and the lumen of the graft with a wide variety of solutions, can maintain the physiological and pharmacological properties of the intestinal smooth muscle. More prolonged preservation has not been successful. The concentration of serotonin in luminal effluents increases with prolonged ischaemia and has been used as a measure of ischaemic injury. Treatment with free-radical scavengers or allopurinol has improved preservation. The small intestine has the highest concentration of the enzyme xanthine oxidase; free-radical damage has been implicated after warm and cold ischaemic injury.

Glutamine is the principal fuel of the enterocyte and the addition of 2 per cent glutamine to UW-derived solutions has improved preservation. The intestinal mucosa regenerating after ischaemia has increased glutamine demands for several days. Lipid-peroxidation membrane damage is iron-dependent and inhibited by 21-aminosteroids (lazaroids); these agents, when given to donor and recipient and added to flushing solutions, give added protection.

Leucocyte–endothelial cell interaction and platelet aggregation contribute significantly to postvascularization damage, which is proportional to cold ischaemic times; a preanastomotic rinse (Carolina rinse at 37°C) has attenuated such damage in experimental rat small-bowel isografts. In clinical practice the safe periods of storage tolerance have usually been kept below 12 h.

Summary of major issues in cadaver multiorgan donation

Cold remains the sheet anchor of organ preservation. Preservation solutions are essential complements to cold storage of organ transplants. External cooling alone extends tolerance to ischaemia from 1 to 2 h up to 12 h or more for most organs. Cooling is enhanced and preservation times prolonged to over 24 h by flush-out solutions, which replace the contained blood before storage and transport of the cooled organ. Continuing machine perfusion can also be provided, if required, during the storage and transfer period. Important questions relevant to current clinical preservative solutions can now be summarized.

What are the respective benefits of simple cold storage and continuous perfusion?

Experimentally, storage times are increased for all organs several-fold by a single, preliminary cold flush combined with simple cold storage, from 12 h without a wash-out up to a maximum period of 5 days for kidneys. Continuous cold perfusion can provide oxygen and nutrients, and can extend this time to up to a week of storage. Freezing and vitrification can extend the period of storage indefinitely, but only for selected tissues in humans. Continuous cold perfusion has most benefit and relevance in the preservation of ischaemically damaged organs from asystolic (non-heartbeating) donors or when prolonged (over 36 h) preservation times are required. Continuous perfusion or persufflation are the only techniques currently capable of resuscitating organs during the preservation period. Continuous perfusion can be applied to all organs but is only clinically relevant to kidney preservation. Most organs procured from brain-dead donors are adequately preserved for clinical use by simple cold storage.

What is the relative importance of the different components of established solutions?

Solutions of widely different composition have given approximately similar results experimentally and clinically. The very successful UW solution has 13 individual solutes. But for most organs a very simple buffered sucrose solution (phosphate-buffered sucrose (PBS)) containing a mere two solutes (sucrose and a phosphate buffer) is only slightly inferior. Thus solutes that are impermeable, or only partly permeable, to cell membranes and which minimize cold-induced cell swelling are the most essential components. Such solutes account for more than 60 per cent of the overall effectiveness of preservation. Similarly, all successful preservatives contain a buffer.

What improvements are possible by modifications of established solutions?

The mainstay for all transplantable organs is now UW solution. Not all its components are immutable or necessary, and several modifications are now in clinical use. Components of the original UW solution used clinically in multiorgan donors, as a simple hypothermic flush followed by ice storage, were shown in Table 2, together with their concentrations and functions. Some minor, but significant, variations are necessary for the effective use of UW as a continuous perfusion solution. An oncotic starch colloid (hydroxyethyl starch) is essential in continuous perfusion to maintain vascular volume and to prevent progressive oedema from flooding the interstitium; in contrast, hydroxyethyl starch is not required for simple hypothermic storage. The colloid macromolecule adds nothing to the effectiveness of simple cold preservation, and by increasing the viscosity of the solution hydroxyethyl starch slows and inhibits a one-pass organ flush. The principal component of UW solution is the impermeable lactobionate anion; however, gluconate was found more effective for continuous perfusion and mannitol replaced raffinose. For perfusion preservation more effective buffering was provided by adding HEPES; adenine and ribose replaced adenosine as fuel precursors for ATP, glucose and oxygen were added to promote oxidative phosphorylation. Glutathione and allopurinol were omitted. Important changes in electrolyte composition were to lower potassium concentrations, reversing the sodium:potassium ratio, and to add small amounts of calcium for longer-term storage. Thus, in overviewing electrolyte content, solutions of both ‘intracellular' (high potassium, low sodium) and ‘extracellular' (high sodium, lower potassium) composition can be equally effective.

Reverting to ‘extracellular' colloid-free composition for the flush-out solution has additional advantages by improving the rapidity and ease of multiorgan flush-out, as the high potassium in the standard UW solution is markedly vasoconstrictive and potentially toxic to the endothelium. These modifications of UW solution (sodium–lactobionate–sucrose solution, Celsior, histidine–lactobionate) gave equivalent preservation and have been applied particularly to heart, lung, iver, and kidney transplantation (Table 5).

What are the requirements of parenchymal cells as compared to non-parenchymal endothelial cells?

Parenchymal cells of most organs are reasonably tolerant to ischaemic/anoxic damage. These tolerances are broadly similar across different organs (most can tolerate ischaemia at normothermia for 1 to 2 h before permanent damage occurs). The non-parenchymal cells [Kupffer cells (macrophages) in the liver, and vascular endothelial cells in all organs] are less tolerant and are the predominant cause of ultimate preservation failure.

What additives can enhance endothelial integrity during preservation and on reperfusion?

Preservation-induced endothelial damage is effected by release of cytotoxic oxygen free radicals such as superoxide and hydroxyl ions, by activation of adhesion molecules promoting leucocyte and platelet adhesion and aggregation, and by activation of proinflammatory cytokines (tumour necrosis factor, interleukins, and growth factors) and other locally produced messengers causing vasoconstriction and endothelial damage. Various additives combating these effects have been shown to give significant enhancement of preservation across a range of organs. These additives have comprised the following.

- Free-radical scavengers (superoxide dismutase, mannitol, reduced glutathione).

- Calcium-channel blockers (diltiazem, verapamil, nifedipine, nisoldipine). These agents have important effects in immune modulation as well as in preservation. In preservation they can inhibit calcium-mediated ischaemic damage by preventing calcium from entering the cytosol. Immunological effects include inhibition of calcium-mediated lymphocyte proliferation, and diminished cyclosporin and tacrolimus nephrotoxicity by blocking the vasoconstrictive effects of endothelin and thromboxane A_2 on the afferent arteriole.
- Other vasodilators (b-blockers, angiotensin-converting enzyme inhibitors, pentoxifylline).
- Prostacyclin (prostaglandin GI_2): prostaglandins are unsaturated fatty acids produced by metabolism of arachidonic acid, along with other eicosanoids (thromboxane and leucotrienes). These agents form rapidly, act locally, and are destroyed enzymatically; they are classified as autacoids or prostanoids, to distinguish them from systemic hormones. Prostaglandins participate in and modulate inflammatory responses. Many are vasoconstrictive. Thromboxane is synthesized in platelets; it stimulates platelet aggregation and is a potent vasoconstrictor. Prostacyclin is vasodilatory and inhibits platelet aggregation.
- Corticosteroids and non-steroidal anti-inflammatory drugs inhibit the formation of arachidonic acid, prostaglandin, and thromboxane.
- Benzodiazepines (chlorpromazine, trifluoperazine) and aprotinin act as membrane stabilizers, along with corticosteroids.
- Agents such as lazaroids (aminosteroids) inhibit lipid-membrane peroxidation.
- Other cell-messenger and adhesion molecules (cAMP and cGMP, interleukins, NO and endothelin inhibitors) act to inhibit the complement cascade, and neutrophil chemotaxis and sequestration.

What benefits can derive from a post-storage flush?

This has been studied mainly for the liver. Benefits have been shown from a flush-out with Carolina rinse solution before revascularization. The composition of Carolina rinse has a number of similarities to UW solution and its ‘extracellular' modifications, which may help protect the more delicate endothelial cells from reperfusion injury. Endothelial cells may also be more sensitive to cold-induced damage than are parenchymal cells. Storage at 4°C rather than O°C enhanced endothelial integrity. The optimal temperature for a pretransplant reflush has been at room or body temperatures to facilitate the transition to revascularization.

What are the preservation tolerances for individual organs to simple cold storage after a preliminary flush?

- The optimal flush solution for all organs remains UW solution (Table 6), incorporating its various modifications and additions (hydroxyethyl starch-free, lower potassium, sodium–lactobionate–sucrose solution, Celsior solutions).

	Experimental	*Clinical (h)*
Kidney	5 days (120 h)	24 (to 60)
Liver	2 days (48 h)	12 (to 36)
Pancreas	4 days (96 h)	12 (to 36)
Small bowel	2 days (48 h)	8 (to 12)
Heart	1 day (24 h)	8 (to 12)
Lung	1 day (24 h)	8 (to 12)

Table 6 Current limits of organ preservation—simple cold storage (University of Wisconsin-derived solution best for all organs)

- Experimentally, simple cold storage can preserve kidneys for 5 days, pancreas for 4 days, liver and small bowel 2 days. Heart and lung storage can be extended to 24 h.
- Clinically safe preservation times are preferably under 48 h for kidneys, under 24 h for liver, pancreas and small bowel, and under 8 h for heart and lung.

Does agonal or preservation injury affect the immune response?

The success rate of organ grafts from brain-dead cadaver donors is consistently inferior to those from living donors, even when live donors are unrelated genetically to the recipient. Irreversible brain injury upregulates proinflammatory cytokine mediators and major histocompatibility complex class I and II antigens, facilitating host inflammatory and immunological responses. Preservation-induced injury promotes inflammatory responses, which further potentiate immune damage by upregulation of histocompatibility antigens. Future improvements in preservation solutions and cadaver donor treatments have the potential to minimize or reverse these effects with consequent improved results of transplantation.

Tissue grafts

Composite tissues, limbs

Reconstructive surgery increasingly employs vascularized autografts of composite tissues (skin, subcutaneous fat, muscle, and bone) to fill large defects. Severed limbs and digits are often replaced as autografts. These complex operative procedures can take many hours. Storage of the grafts has predominantly used simple cooling by refrigeration, and by wrapping the tissues in cold saline-soaked packs during implantation. Supplementary vascular flushing has not been widely used for fear of damage to the small vessels requiring microvascular suture. Experimental studies have not indicated significant improvement with vascular flushing. Simple cold storage adequately covers the periods (usually less than 12 h) required in clinical practice. The tolerance of different tissues to ischaemia varies: nerve and muscle are more sensitive than skin and bone. Preservation of the morphology and viability of skin, muscle, and bone was successful experimentally after several days of storage at 4°C by simple hypothermia in a moist environment.

In whole-limb replacement, flushing with an organ-preservation solution can aid cooling and potentially improve preservation. Before the circulation is restored to replanted limbs or other tissues of large bulk, any solution with high potassium should be flushed out with balanced electrolyte solution or plasma; otherwise fatal hyperkalaemic cardiac arrest can occur after release of clamps.

Cornea

Early in the twentieth century it was realized that the cornea was not an inert transparent membrane but a biologically active tissue needing careful handling if it was to survive the transplant operation. Pioneering efforts in the 1930s established the suitability of cadaver corneas for grafting and the feasibility of direct suturing of the graft to host tissues. The relative immunological privilege of the cornea was appreciated at an early stage.

The cornea's function as a transparent reflecting and focusing complex requires the metabolic activity of a monolayer endothelium of mesodermally derived cells on the inner surface bathed by the aqueous humour. Endothelial DNA synthesis has been used to test preservation efficacy experimentally.

For procurement from the cadaver, eyes are enucleated (removed as whole eyes). This simple procedure can be done without elaborate preparation and need not be disfiguring. Originally, corneas collected for transplantation were used immediately after the eye had been enucleated. Short-term storage by refrigeration extended the time. Eyes collected in this way and stored in a moist environment at 4°C under sterile conditions could be used for penetrating (whole-thickness) grafts up to 48 h after the death of the donor.

Intermediate storage, for up to 4 days, was achieved by the use of refrigerated tissue-culture media. MK medium (developed by McCarey and Kaufman) contained dextran 40 and HEPES-buffered tissue-culture fluid and provided reliable preservation for 3 to 4 days. Storage in MK medium has achieved wide acceptance, but still does not give totally sufficient time for the ever more complex screening required for the eye bank: examination of endothelial sheet integrity and disease screening (Creutzfeldt–Jakob syndrome, human immunodeficiency virus antibodies, septicaemia). Tissue typing may prove to be a valuable adjunct to corneal transplantation. The corneal Langerhans cells are HLA-DR positive. Retrospective and prospective surveys have shown that results correlate with antigen match and ABO blood-group compatibility. Further refinement of the tissue-culture medium replaced dextran 40 with chondroitin sulphate, which provided up to 14 days hypothermic preservation with only 3 per cent loss of the endothelial cells. Best preservation was attained with 2.5 per cent chondroitin sulphate free of lower molecular-weight moieties. Long-term storage can be achieved by cryopreservation of the isolated cornea. This technique was developed for prolonged eye banking, but has not gained universal acceptance.

Short-term, hypothermic, whole-eye storage remains the most simple and convenient technique, particularly when there is no shortage of eyes. Surveys of eye banks

in the United States indicate that this is still a common method of short-term banking of whole eyes. Tissue-culture (MK) storage has gained acceptance by the majority of laboratories, and outcomes have not been significantly affected by preservation methods. A minority of departments use cryopreservation, and then only occasionally.

Pancreatic islets

Pancreatic islet preservation and transplantation has been extensively studied experimentally. Intraportal transplantation of islet extracts is now achieving clinical success. Cellular preparations, after collagenase digestion and purification, have been successfully preserved by tissue culture for 24 h at 37°C, by simple ice cooling for 24 h; or for more prolonged periods by freezing with the addition of cryoprotectives.

Skin

Harvested split-skin grafts used as autografts or allografts can be preserved for approx. 2 weeks by simple wrapping and rolling in packs moistened with saline or tissue fluid, which are then stored in the refrigerator. Xenografts of pig skin can be similarly prepared and have been widely used for temporary cover. More prolonged preservation can be provided by freezing, or by cultured epidermal cells that are subsequently reimplanted as a monolayer.

Bone and cartilage

Autografts of fresh cancellous bone taken from iliac crest or other areas remain the most effective source of bone grafts; autografts provide viable osteoblasts and stimulate osteogenesis. Bone is a complex and active metabolic tissue, with osteocytes bearing histocompatibility antigens nourished within a hydroxyapatite matrix. Stored allografts of bone can be cryopreserved for 12 months or more. The biomechanical properties of bone related to its matrix do not seem affected by cryostorage, but preserved allografts are considerably less active in stimulating osteogenesis than fresh autografts.

Alternatively, allografts can be stored in bone banks at room temperature after preparation by defatting in chloroform and methanol, freeze-drying, and sterilization by ethylene oxide or gamma irradiation. Cadaveric allografts or allograft/prosthetic composites have been used to replace long bones excised for bone tumours. Host osteoblasts can subsequently cause new bone to grow along Haversian canals and line the surface with appositional new bone. Sterilized, partially demineralized bone matrix is an alternative source of osteogenic bone extract.

Transplantation of large segments of cortical bone autografts also can provide a rigid bony matrix that is gradually replaced by creeping substitution of new bone. Transfer of viable cortical bone requires a vascularized graft restoring the bone's blood supply at its new site. Vascularized autografts of fibula and iliac crest can replace bony defects resulting from congenital anomalies, disease, or injury.

Cartilage derives most of its nutrition by diffusion from synovial fluids; re-establishment of a blood supply is less critical than for bone. Chondrocytes also possess histocompatibility antigens and evoke an allograft rejection response. Intact cartilage survives better than isolated chondrocytes, showing the importance of the cartilage matrix. Experimentally, cartilage can be stored for 28 days in tissue-culture medium at 4°C. The morphology of the cells and the concentration of glycosaminoglycans and of collagen showed no significant change, and the ability of cells to incorporate [^{35}S]sulphate into glycosaminoglycans and to synthesize collagen was not diminished.

Cryopreservation at subzero temperatures with glycerol or dimethyl sulphoxide showed a less favourable outcome, with loss of chondrocytes and conversion of hyaline cartilage to fibrocartilage. Isolated chondrocytes can be well preserved by freezing, suggesting that poor penetration of cryopreservative occurs in cartilage tissue.

Cartilage or osteochondral allografts have been used clinically to manage degenerative arthritis as onlays or plugs of cadaveric articular cartilage fragments. The methods remain experimental.

Bone marrow

Bone-marrow transplantation is now the preferred treatment for aplastic anaemias and many leukaemias. Marrow allografts usually require minimal preservation, other than that provided by simple hypothermia as for blood collection. Invariably, living donors have been used. Extension of bone-marrow transplantation in the future using cadavers could involve cryoprotective techniques similar to those used for freezing of blood.

Blood vessels

Autogenous veins replace and bypass arteries in aortocoronary and peripheral vascular surgery. Occasionally, autogenous arteries (such as internal iliac) are used for short bypass procedures such as aortorenal grafts. In these situations the graft is a ‘vital' one with the cellular tissues retaining their viability. The graft is usually prepared in heparinized saline at room temperature. Should the expected ischaemic period be in excess of 1 h it is advisable to cool the vessel in iced saline. Blood vessels from cadaver donors are a very important source of extended arterial and venous conduits for renal, liver, and pancreatic allografts. These vessels are adequately preserved by simple hypothermic storage in standard flushing solutions. Low concentrations of deuterium oxide in UW solution have been shown to improve storage up to 72 h.

Vessels can be removed from cadaver donors, treated with glutaraldehyde, and stored indefinitely; they then become essentially collagenous tubes without apparent antigenic activity. This allografted non-living tissue does not cause a rejection reaction but will slowly degenerate and thrombosis or aneurysm formation can develop. It is an inferior option to fresh autogenous vessels for arterial bypass. Glutaraldehyde-treated, porcine heart-valve xenografts similarly show progressive tissue degeneration and have been precoated with fibroblast growth-factor protein to enhance angiogenesis.

Cold ischaemia increases the output of endothelin-1, a vasoactive peptide produced by vascular endothelium and a potent endogenous constrictor of vascular smooth muscle. Topical application of glyceryl trinitrate and verapamil have been used to combat these effects. Other methods aimed at protecting endothelial integrity include the topical application of vascular endothelial growth factor or the use of antisense oligonucleotides of synthetic DNA to block the expression cell-cycle regulatory gene. Experimentally, such agents have been shown to inhibit intimal proliferation after transplantation, to preserve normal endothelial cell phenotype and function, and to reduce the expression of superoxide free radicals. The heterogeneity of vascular tissues may influence preservation efficacy, with differing optimal temperatures of storage. Endothelial-related functions were better preserved by storage at 4°C, while smooth-muscle function was better at 0°C. These findings highlight potential avenues for improved preservation of vascularized organ grafts.

The future

Preservation for transplantation entered a new and exciting phase with the development of multicomponent, UW–lactobionate preserving solutions. The basic requirements of preservation remain rapid cooling aided by impermeants and buffers. Many other metabolic and pharmacological influences have enhanced organ protection. Incorporation of such agents into preservation solutions, as well as their judicious use in the preparation of the donor and in subsequent treatment of the recipient, have extended significantly the preservation times of all organs and tissues. Although heart and heart–lung transplants still require rapid surgery with minimally delayed storage times, prospects for extending safe preservation times for all vascularized organ grafts to 24 h or longer are now within sight. Research continues to look for more effective additives to preserving solutions, and to improvements in perfusion techniques, to extend safe preservation times even longer.

Preparation of the graft or of the cadaver donor by a reduction in antigen-presenting cells and treatment of the recipient before grafting by selected agents both offer potential avenues for more successful transplantation. Successful pretreatment of the cadaver donor to downregulate proinflammatory and immunological messenger molecules is within sight. Advances in extracorporeal protection of parenchymal cells and of the vascular endothelium during subsequent storage, transfer, and reimplantation will contribute significantly to such future developments.

Further reading

Carrel A, Lindbergh GCA. *The culture of organs*. Hamish Hamilton Medical, London, 1938. [Classic early account of organ-preservation techniques by Nobel laureate and aviator technologist.]

Chapman JR, Deierhoi M, Wight C, ed. *Organ and tissue donation for transplantation*. Arnold, London, 1997. [Broad coverage and update of the whole field of organ and tissue donation stressing multidisciplinary approach.]

Collins GM, Dubernard JM, Land W, Persijn GG, ed. *Procurement, preservation and allocation of vascularised organs*. Dordrecht: Kluwer, 1997. [Detailed overview of preservation techniques for vascularized organs.]

Kootstra G. The asystolic, or non-heart-beating, donor. *Transplantation* 1997; **63**: 917–21. [Overview of techniques for procuring and resuscitating organs from donors presenting with cardiac arrest.]

Van Der Werf WJ, D'Alessandro AM, Hoffman RM, Knechtle SJ. Procurement, preservation, and transport of cadaver kidneys. In: Rao VK, ed. *Renal transplantation*, Vol. 78(1), *Surgical Clinics of North America*. Saunders, Philadelphia, 1998. [Detailed overview of kidney preservation at University of Wisconsin, describing current techniques and results.]

16.4 Immunosuppression

Peter J. Morris

[Introduction](#)
[Drugs](#)
- [Azathioprine](#)
- [Steroids](#)
- [Cyclosporin](#)
- [Tacrolimus (FK506)](#)
- [Sirolimus (Rapamycin)](#)
- [Mechanisms of action of cyclosporin, tacrolimus, and sirolimus](#)
- [Mycophenolate mofetil](#)
- [Cyclophosphamide](#)
- [Mizoribine](#)
- [15-Deoxyspergualin](#)
- [Brequinar sodium](#)
- [Leflunomide](#)
- [FTY720](#)

[Biological agents](#)
- [Antilymphocyte or antithymocyte globulins](#)
- [Monoclonal antibodies](#)

[Total lymphoid irradiation](#)
[Complications of immunosuppressive therapy](#)
[Further reading](#)

Introduction

The advances in organ transplantation, so evident in the sections covering transplantation of different organs, can be attributed to a great extent to the advances in immunosuppression over the last 40 years. In the 1950s total body irradiation was the only form of immunosuppression available and patients either died of marrow aplasia and overwhelming infection if given sufficient irradiation to prevent rejection of a renal transplant, or rejected the graft if given lower doses of irradiation. Nevertheless, modest success was achieved in some patients at that time. The introduction of azathioprine in the early 1960s allowed a major advance in renal transplantation and quickly led to a rapid expansion of renal transplant units throughout the world, with prolonged graft survival now being achieved in many patients both in the medium and long term. Steroids were added to azathioprine first to treat rejection and then in combination with azathioprine to prevent rejection. Heterologous antilymphocyte globulin, usually made in horses at the time, was introduced to treat steroid-resistant rejection in the 1970s, but basically the standard immunosuppressive therapy in all units was azathioprine and steroids for nearly 20 years until cyclosporin became generally available in the early 1980s.

Very quickly cyclosporin-based immunosuppressive protocols became standard therapy and remain so to this day. The introduction of cyclosporin led not only to a marked improvement in renal allograft survival (10 to 15 per cent), but also to a striking improvement in liver and cardiac allograft survival. This led to a dramatic increase in the numbers of liver and cardiac transplants throughout the Western world.

During the past 10 years a number of other immunosuppressive drugs have become available or are undergoing trials, such as tacrolimus, mycophenolate mofetil, and sirolimus. In addition the explosion in the production of monoclonal antibodies recognizing different cell surface molecules on cells involved in the immune response promises to add a new dimension to therapy with the hope of producing increasing specificity of immunosuppression. OKT3, a pan-T-lymphocyte monoclonal antibody, has been widely used for many years now to treat steroid-resistant rejection and also as part of induction therapy in many centres. This agent is not a particularly specific antibody in that it recognizes all T cells, but antilymphocyte agents against other lymphocyte targets such as the interleukin 2 (**IL-2**) receptor, CD4, CD40 ligand, and B7.1,2 should allow much greater specificity in immunosuppression to be achieved in clinical practice. Already antibodies against the IL-2 receptor are in clinical practice while the others are undergoing clinical trials.

One of the major problems in clinical transplantation in the coming years will be the evaluation of the large number of potentially valuable new therapies (both drugs and biological reagents) that will become available. For it must be remembered that the results of organ transplantation are now relatively good so that only large multicentre trials will be able to establish the true value or otherwise of new therapies. Furthermore, more sophisticated methods of analysis of outcome than mere graft survival will have to be developed, and although acute rejection has been accepted as a surrogate marker of subsequent graft survival in all trials in recent years, considerable doubt has now been cast on the validity of this assumption.

Drugs

Azathioprine

Azathioprine is a purine analogue, and is essentially an antiproliferative agent, inhibiting both DNA and RNA synthesis by preventing the synthesis of adenylic and guanylic acid from inosinic acid. For 20 years, in association with steroids, azathioprine provided the backbone of immunosuppressive therapy. In the cyclosporin era, azathioprine is still widely used in lower doses with cyclosporin on the assumption that it allows lower doses of cyclosporin to be used, hence decreasing the side-effects of both drugs. The major side-effect of azathioprine is leucopenia, and indeed regular white cell counts are the only method of monitoring the dosage. If azathiaprine is used with steroids alone then the usual starting dose is 3.0 mg/kg daily reducing to maintenance levels of 2.0 mg/kg; lower doses than this will result in increased loss from graft rejection. However, when used as part of a cyclosporin-based protocol it tends to be used in doses of either 100 mg daily or 1.5 mg/kg daily.

Steroids

Prednisolone or prednisone were used routinely with azathioprine in the precyclosporin era, and also are used in most cyclosporin-based protocols. Their mechanism of action in suppressing the alloreaction is unclear, but although immunosuppressive to some extent possibly their major action is an anti-inflammatory one. Certainly their concomitant use with azathioprine was essential to produce appropriate immunosuppression, but that is not clearly so in the cyclosporin era. When first introduced soon after azathioprine became available, steroids were used in high doses and indeed most of the complications of transplantation in the 1960s and 1970s, such as cushingoid changes, avascular necrosis of joints, peptic ulceration, infection, and osteoporosis, could be attributed to the use of high-dose teroids. The demonstration in controlled trials that low-dose steroids were as effective in preventing rejection as high-dose steroids quickly led to the general introduction of low-dose steroid protocols; for instance, 20 mg daily during the first 2 to 3 months after transplantation reducing to maintenance levels of 10 mg daily. These low-dose protocols dramatically reduced the morbidity and mortality of renal transplantation in the azathioprine era.

High-dose steroids are also used to treat rejection and a widely used protocol in this context is either 0.5 or 1.0 g of methylprednisolone given as an intravenous bolus daily for 3 days. This protocol successfully reverses the majority of acute rejection episodes of kidney, heart, and liver, and antilymphocyte globulin or OKT3 is reserved for steroid-resistant rejection.

Cyclosporin

Cyclosporin is a potent immunosuppressive agent, extracted from two strains of Fungi Imperfecti. It has a molecular weight of 1200 and comprises 11 amino acids (Fig. 1). It is a calcineurin inhibitor and its major mechanism of action is to prevent the product of cytokines such as IL-2, which will be discussed in more detail later. When first introduced into clinical practice in renal transplantation and bone marrow transplantation it resulted in superior suppression of rejection and graft-versus-host disease, respectively, which was reflected by better graft and patient survival. However side-effects soon became evident, the major ones being nephrotoxicity and hypertension (Table 1). Nephrotoxicity is a major problem of cyclosporin use in all forms of organ transplantation and bone marrow transplantation and although dose related to a great extent, there is probably no effective dose of cyclosporin that is not nephrotoxic. The nephrotoxicity is due mainly to vasoconstriction of the afferent arterioles of the glomeruli, leading to a decrease in glomerular filtration rate, and these changes appear to be mediated by inhibition of vasodilator renal prostaglandin metabolites and increased production of thromboxane.



Fig. 1. The chemical structure of cyclosporin, tacrolimus, and sirolimus.

Renal	Nephrotoxicity, haemolytic uraemic syndrome
Hepatic	Hepatotoxicity
Neoplastic	Lymphomas, fibrosarcoma of breast, squamous cell carcinoma
Dermatological	Thickening, rashes, hypertrichosis
Gastrointestinal	Anorexia, nausea, failure to gain weight
Metabolic	Hyperkalaemia, hyperuricaemia, hypomagnesaemia, hyperglycaemia
Neurological	Tremor, convulsions, burning sensation in limbs, malaise, depression
Cardiovascular	Fluid retention, hypertension, hypercholesterolaemia, Raynaud's phenomenon, intravascular coagulation
Dental	Gingival hypertrophy
Haematological	Haemolytic anaemia

Table 1 Side-effects of cyclosporin therapy

Because of the nephrotoxicity and other dose-related side-effects of the drug, a number of cyclosporin-based protocols have evolved in an attempt to reduce the incidence of these side-effects (Table 2).

Cyclosporin ± steroids
Cyclosporin conversion to azathioprine after 3 to 12 months
Cyclosporin + azathioprine
Cyclosporin + azathioprine + steroids (triple therapy)
ATG/OKT3 + azathioprine + steroids @ cyclosporin + azathioprine + steroids
ATG/OKT3 + cyclosporin + azathioprine + steroids

Table 2 Cyclosporin-based protocols

Cyclosporin alone (monotherapy)

Although effective, it is possible that higher doses of cyclosporin are required than if used with other agents, and this in turn may lead to a greater incidence of nephrotoxicity.

Cyclosporin and steroids

This is a commonly used protocol, but is not proved to be better than cyclosporin alone, although it has been suggested that the incidence of nephrotoxicity is lower when steroids are used.

Cyclosporin conversion therapy

Cyclosporin is used alone or with steroids and then at some given time after transplantation, such as 3, 6, or 12 months, cyclosporin is replaced with azathioprine and steroids, preferably with some overlap. Undoubtedly renal function improves with conversion but the major drawback of these protocols is the risk of rejection occurring within the 1 or 2 months after conversion. Although these rejection episodes usually respond to steroid therapy, this type of protocol does require close supervision of patients after conversion. Nevertheless, the financial aspects of this protocol as well as the improved renal function are attractive. Thus, this Oxford protocol, as it is often known, is widely used in developing countries.

Low-dose cyclosporin, azathioprine, and steroids (triple therapy)

This is the most widely used cyclosporin-based therapy and produces very acceptable patient and graft survival (Fig. 2). It is not a more potent immunosuppressive protocol than other cyclosporin protocols, but it is relatively free of side-effects and easy to use. A typical protocol (as used in Oxford) would comprise cyclosporin at 8 mg/kg per day in divided doses reducing according to whole-blood trough levels, azathioprine at 100 mg/day, and prednisolone at 20 mg/day reducing to a maintenance level of 10 mg/day, with the further aim of weaning patients off steroids altogether if renal function is stable at 1 year. Indeed our own experience of steroid withdrawal in patients on triple therapy has shown benefits in levels of both hypertension and hypercholesterolaemia.

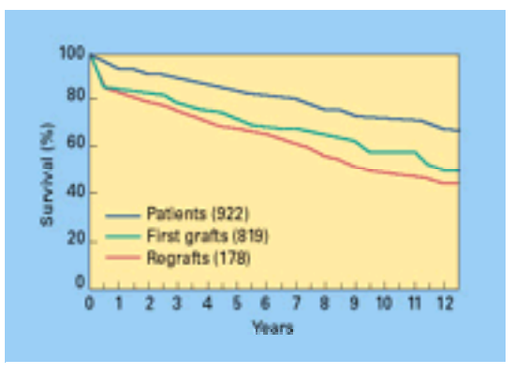


Fig. 2. Overall survival of patients, cadaveric primary grafts, and regrafts treated with triple therapy (cyclosporin, azathioprine, and steroids) in the Oxford Transplant

Centre for the past 15 years. Results in recent years with triple therapy are superior to the earlier years in this experience.

Cyclosporin and azathioprine (double therapy)

This is not widely used and as the addition of steroids is required quite often in the early months it has no advantages over triple therapy.

Antithymocyte globulin (ATG) or OKT3, azathioprine, and steroids with introduction of cyclosporin as renal function is established (sequential therapy)

This type of protocol is widely used in North America in all forms of organ transplantation, but its use remains controversial both because improved graft survival has not been clearly established with the use of this type of protocol, and if subsequently steroid-resistant rejection requires repeated treatment with either ATG or OKT3 then a substantial risk of developing a fatal acute lymphoproliferative disorder has been introduced. Nevertheless, if there is no renal function after transplantation of a cadaveric kidney as a result of acute tubular necrosis, it is an attractive approach to the induction of immunosuppression, for there is reasonably good evidence that the use of cyclosporin in a patient with a non-functioning kidney delays the onset of function and also leads to long-term damage to the kidney.

ATG or OKT3, cyclosporin, azathioprine, and steroids (quadruple therapy)

This regimen is sometimes used in cardiac or liver transplantation. It probably represents an unnecessarily potent immunosuppressive protocol for most patients and is not recommended, except in the highly sensitized patient.

Monitoring of cyclosporin levels

Trough whole-blood levels are measured at either 11 or 24 h depending on whether the patient is taking cyclosporin twice daily or daily. The original formulation of cyclosporin (Sandimmun) in corn oil could be given once or twice a day. If used once per day then trough levels are maintained at 200 to 400 µg/ml in the first 5 weeks and then at 100 to 200 µg/ml once stable graft function is established. Sandimmun has now been largely replaced by a new microemulsion formulation (Neoral) which has a superior bioavailability to Sandimmun but must be given twice daily. Target trough levels are 200 to 300 µg/ml in the first 2 months after transplantation and 100 to 200 µg/ml thereafter. High trough levels of cyclosporin are likely to be associated with toxicity and low levels with an increased incidence of rejection, but this is not inevitable and the trough levels must be used in association with other clinical, biochemical, and histological parameters. Indeed the most accurate measurement of a therapeutic and non-toxic level of cyclosporin is achieved with the measurement of the area under the curve, and this can be carried out approximately in the case of Neoral by measurements at 2 and 6 h after dosing. This is particularly useful if there are apparent problems with absorption as judged by erratic trough levels. One must be aware of possible drug interactions as cyclosporin is metabolized by the cytochrome P450 enzyme system in the liver, and hence any drugs that interfere with that system may lead to an increase (e.g. ketoconazole) or a decrease (e.g. phenytoin) in cyclosporin blood levels.

Tacrolimus (FK506)

Another potent immunosuppressive macrolide antibiotic, tacrolimus or FK506, is derived from *Streptomyces tsukubaensis* and is also a calcineurin inhibitor. It has an entirely different structure to cyclosporin (Fig. 1), but its mechanism of action is similar in that it prevents the production of cytokines in response to antigen recognition (see later). On a weight for weight basis it is 10- to 100-fold more potent than cyclosporin. Its first clinical use in Pittsburgh was for the salvage of liver transplants undergoing rejection despite the use of high-dose steroids and OKT3. In some 70 per cent of instances, conversion from cyclosporin to FK506 resulted in improved liver function tests and survival of the liver allografts, an impressive debut. Its first use as a primary agent with steroids was also explored at Pittsburgh in liver, cardiac, and renal transplantation with excellent results in liver (Fig. 3) and cardiac transplantation. This was followed by two, large, prospective, randomized, non-blinded, multicentre trials in Europe and North America comparing cyclosporin (Sandimmun) and tacrolimus in liver transplantation. Both trials showed a reduction in the incidence of acute rejection in the tacrolimus arms, but graft survival was essentially the same at 1 year. However, more patients were converted to tacrolimus than vice versa.

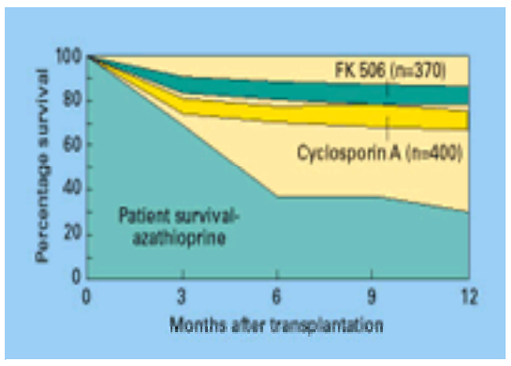


Fig. 3. Survival of patients and grafts after liver transplantation during the azathioprine era, and contrasting with patient and graft survival in Pittsburgh with cyclosporin or FK506 (by courtesy of T. Starzl).

Similarly in two, randomized, prospective, controlled trials in renal transplantation, in which tacrolimus was compared with cyclosporin (Sandimmun), there was a significant reduction of the incidence of acute rejection after transplantation, but no improvement in graft survival was observed at 1 year. Tacrolimus has also proved of value both in liver and renal transplantation for rescue of grafts that are undergoing rejection on cyclosporin therapy.

In addition it has allowed a number of successful small bowel transplants in adults and children to be performed in Pittsburgh without a concomitant liver transplant, which does represent a significant advance in this area. It may also be more effective than cyclosporin in pancreatic transplantation and here too it is tending to become the primary immunosuppressive agent.

Blood trough levels are used for monitoring dosages, the drug being given twice daily with target trough levels being between 10 and 15 ng/ml in the early weeks after transplantation and 5 to 10 ng/ml thereafter. Not unexpectedly there is a similar profile of side-effects of tacrolimus to cyclosporin such as nephrotoxicity and nausea, but there is less hypertension and hyperlipidaemia. On the other hand, there is a greater incidence of hyperglycaemia, often requiring insulin therapy (Table 3). Thus, overall tacrolimus has proved to be a very useful addition to our immunosuppressive armamentarium.

Side-effect	*CSA*	*Aza*	*Steroids*	*FK506*	*MMF*	*RAP*
Nephrotoxicity	++	-	-	++	-	-
Hypertension	+	-	+	-	-	-
Hyperlipidaemia	+	-	+	+	-	+
Leukopenia	-	+	-	-	+	-
Thrombocytopenia	-	+	-	-	-	+
Neurotoxicity	+	-	-	+	-	-
Diabetes	+	-	-	+	-	-

Aza, azathioprine; CSA, cyclosporin; FK506, tacrolimus; MMF, mycophenolate mofetil; RAP, Rapamycin.

Table 3 Side-effects of immunosuppressive drugs

Sirolimus (Rapamycin)

Sirolimus (Rapamycin) is another macrolide antibiotic, produced by *Streptomyces hygroscopicus* isolated from soil samples from Easter Island. It has a very similar molecular structure to FK506 (Fig. 1), but in contrast it has an entirely different mechanism of action in that it inhibits the proliferation of the activated T cell, that is the progression of the T cell from G1 to S phase (see Fig. 5). It is a potent immunosuppressive agent in experimental models and is at present undergoing multicentre phase 3 clinical trials in renal transplantation. In preliminary studies of its use with cyclosporin, with which it appears to be synergistic, there was a highly significant reduction in the incidence of acute rejection to around 10per cent in the first 6 months after transplantation, and this was subsequently confirmed in a large multicentre study. In a small study of cyclosporin compared with sirolimus the incidence of rejection and graft survival was the same in both arms. It has a quite different spectra of side-effects in comparison with the calcineurin inhibitors in that it is not nephrotoxic for example, but it does produce thrombocytopenia and hyperlipidaemia (Table 3).

Another potentially very interesting property is its antiproliferative activity, which may prevent myointimal hypertrophy, the hallmark of chronic allograft failure. It remains to be seen whether this effect in experimental models is found in human allografts, for if so, it would immediately become a very valuable drug. It may also have a role in allowing lower doses of cyclosporin to be used.

Mechanisms of action of cyclosporin, tacrolimus, and sirolimus

There has been a marked increase in our knowledge of how these agents work, the implications of which are enormous in our understanding both of signal transduction after antigen recognition leading to T-cell activation and proliferation of the activated T cell. All three agents bind to a ubiquitous class of proteins within the cytoplasm known as immunophilins. Cyclosporin binds to cyclophilin while FK506 (tacrolimus) and Rapamycin (sirolimus) bind to a different immunophilin, FK-binding protein (**FKBP**). These proteins are isomerases and it was first thought that the immunosuppressive action was due to inhibition of isomerase activity, but as it was realized that very small amounts of cyclosporin or FK506 were immunosuppressive, although only binding to a portion of the total amount of isomerase available, other mechanisms were sought. Then another protein was found to be involved, a calcium- and calmodulin-dependent phosphatase, calcineurin, to which the complexes of cyclosporin plus cyclophilin and FK506 plus FKBP bind, but not the Rapamycin plus FKBP complex; hence presumably the different effect of Rapamycin. The drug-immunophilin complex bound to calcineurin prevents the dephosphorylation of the transcription factors such as NFAT (nucleus factor in activated T cells), and their translocation to the nucleus and to the enhancer region upstream of the gene for IL-2, preventing its transcription and hence the production of IL-2 (Fig. 4). The complex of Rapamycin plus FKBP binds to a kinase, FK-Rapamycin associated protein (FRAP), that plays a critical role in activation of the ribosomal protein S-6 as well as inhibiting phosphorylation of other factors involved in maturation of the cell cycle. The final effect is the prevention of proliferation of the activated T cell, by inhibiting the lymphokine-receptor signal, and the inhibition of the T-cell cycle between the G1 and S phase.

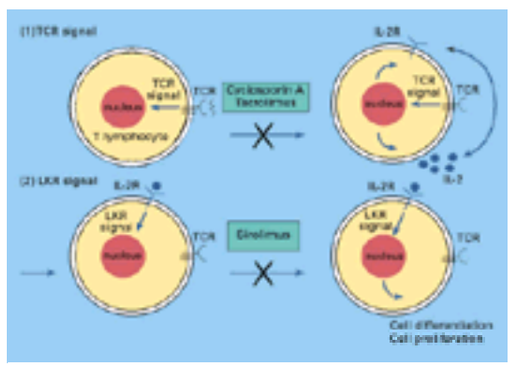


Fig. 4. Mechanism of action of cyclosporin, tacrolimus, and sirolimus. Cyclosporin (CSA) and tacrolimus block the transduction of the signal from the T-cell receptor (TCR) after it has recognized antigen which leads to the production of lymphokines such as IL-2, while sirolimus blocks the lymphokine-receptor signal (LKR), for instance IL-2 plus IL-2R, which leads to cell proliferation.

Mycophenolate mofetil

Mycophenolate mofetil is an ester of mycophenolic acid produced by several species of *Penicillium*. It is converted in the liver to its active moiety, mycophenolic acid. It has a quite specific activity in the purine pathway in that it non-competively inhibits inosine monophosphate dehydrogenase activity and hence the formation of the enzyme guanosine monophosphate. Its action is exerted at the S phase of the cell cycle (Fig. 5), but as the purine salvage pathway is rather less active in lymphocytes than the *de novo* pathway, it is relatively specific for lymphocytes. *In vitro* it has proved to be a powerful immunosuppressive agent, especially when used together with cyclosporin; in addition it inhibits B-cell activity and hence antibody formation. Early experimental use showed that it was a potent immunosuppressive agent and it subsequently underwent clinical trials.

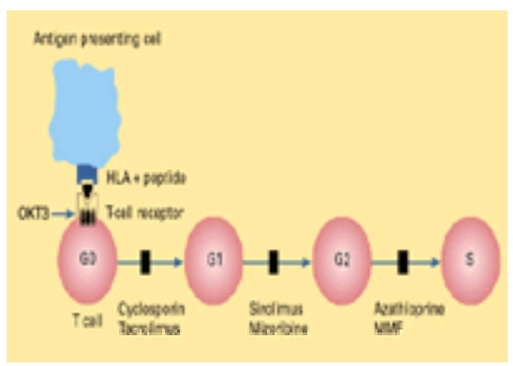


Fig. 5. Sites of action of immunosuppressive agents in the T-cell cycle.

Three, major, multicentre, prospective, randomized, blinded trials have been completed in renal transplantation and more recently one in cardiac transplantation. In the renal transplant trials carried out in North America, Australia, and Europe, mycophenolate mofetil was tested in two doses, either 2 or 3 g/day, in association with cyclosporin and steroids, the control being azathioprine. In all three trials the results were similar in that the incidence of acute rejection in the first 6 months after transplantation was reduced by 30 to 50per cent. However, graft survival was not improved at 1 year. The major side-effects were nausea, diarrhoea, anorexia, and leucopenia. Mycophenolate mofetil has also been shown to be of value in the treatment of refractory rejection even in instances where there has been no apparent response to an antilymphocyte agent. Thus in renal transplantation, mycophenolate mofetil has achieved a role in immunosuppressive therapy and is used in place of azathioprine but with cyclosporin. Its use in the management of chronic renal allograft failure is also being explored, for apart from a possible direct antiproliferative effect on myointimal hypertrophy, it may allow lower doses of cyclosporin to be used thus diminishing cyclosporin nephrotoxicity.

In cardiac transplantation a large multicentre trial has now been completed in which azathioprine was replaced with mycophenolate mofetil, again in combination with cyclosporin and steroids. The incidence of acute rejection was significantly reduced and there was also a superior graft survival at 1 year in the mycophenolate mofetil arms. Mycophenolate mofetil is also widely used with cyclosporin in pancreas transplantation.

Cyclophosphamide

This antiproliferative agent is now seldom used but has been used as a replacement for azathioprine in the presence of hepatotoxicity considered to be due to this drug. It has also been used in India instead of azathioprine because it is cheaper, with success.

Mizoribine

Mizoribine, an imidazole nucleoside antibiotic, which inhibits RNA and DNA synthesis via the purine biosynthesis pathway, has been used successfully instead of azathioprine in Japan for a number of years. It has a similar action to mycophenolate mofetil.

15-Deoxyspergualin

This is an antitumour antibiotic synthesized from spergualin, a fermentation product of *Bacillus laterosporus*. It suppresses macrophage function and inhibits antibody production. In experimental models it is modestly immunosuppressive, while its clinical use has been very limited and is associated with gastrointestinal toxicity and marrow suppression. Uncontrolled trials in Japan have suggested that it is effective in treating rejection, especially in conjunction with the use of methylprednisolone. It might also have a place in the transplantation of sensitized patients or in the desensitization of highly sensitized patients together with plasmapheresis.

Brequinar sodium

This anticancer agent prevents cell proliferation by inhibiting *de novo* pyrimidine synthesis. In experimental models of transplantation it appears to be quite a powerful immunosuppressive agent, not only in allograft models but also in xenograft models used either alone or with other immunosuppressive drugs. It is particularly effective in blocking antibody responses, but phase 1 and 2 studies in patients with cancer and recipients of cadaveric kidneys did show that its use was associated with myelosuppression, especially thrombocytopenia. If it is to be used clinically, it will have to be used in small subtherapeutic doses.

Leflunomide

Leflunomide is an isoxazole derivative and is immunosuppressive in experimental models of autoimmunity and transplantation. It is associated with numerous side-effects in experimental models, such as anaemia, gastrointestinal ulceration, and interstitial nephritis, which preclude clinical use in its present form.

FTY720

This compound was isolated from *Isaria sinclairii* and inhibits lymphocyte proliferation, perhaps by increasing homing to lymph nodes or by causing apoptosis of T lymphocytes. Both in rodent cardiac allograft and canine kidney allograft models it enhanced the effect of cyclosporin. No clinical studies are yet available as the search for less toxic analogues continues.

Biological agents

Antilymphocyte or antithymocyte globulins

Heterologous antisera produced in another species, usually the horse or rabbit, have been used for many years primarily to treat steroid-resistant rejection, but more commonly in recent years as induction therapy to prevent rejection. In general, the major action of antilymphocyte antisera is to reduce the T-lymphocyte count, but being heterologous antisera produced by immunization of horses or sheep with human splenocytes or thymocytes, they inevitably have some general antileucocyte action as well as some antiplatelet activity and may produce varying degrees of leucopenia or thrombocytopenia. However, a good antilymphocyte globulin, and several are available, will successfully reverse a steroid-resistant rejection in 70 per cent of instances, with relatively few side-effects.

Monoclonal antibodies

Monoclonal antibodies are produced by immunizing either rats or mice with the particular antigen against which antibodies are required, for example T lymphocytes. Spleen cells from the rodent are then fused with a myeloma cell line to produce a hybridoma. The hybridomas are screened to find one producing the desired antibody, which can then be cloned, thus providing a continuing source of a highly specific antibody (Fig. 6).

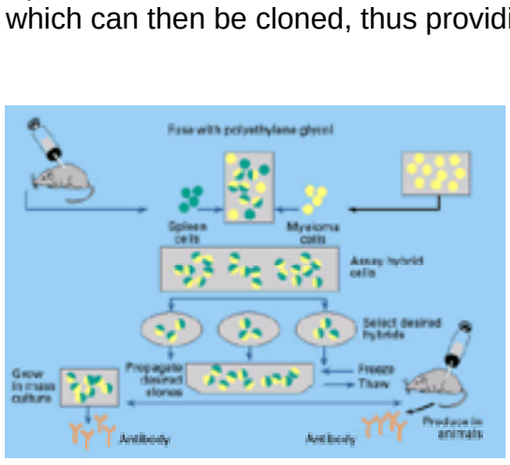


Fig. 6. A summary of the technology of monoclonal antibody production.

Monoclonal antibodies can be produced against a variety of cell surface markers involved in the cellular interactions of the immune response (Fig. 7) and hence allow the possibility of more specific approaches to immunosuppression. Unfortunately, as the antibody is a rat or mouse immunoglobulin, most patients develop anti-idiotypic or anti-isotypic antibodies against mouse or rat globulin which makes them ineffective. However, genetic engineering has allowed chimeric or humanized antibodies to be made, in which the antigen-binding part of the mouse or rat antibody (variable part of Fab) is grafted on to the human constant parts of the heavy and light chains of an immunoglobulin molecule. These genetically engineered antibodies significantly reduce the likelihood of immunization, although anti-idiotypic antibodies may still occur.

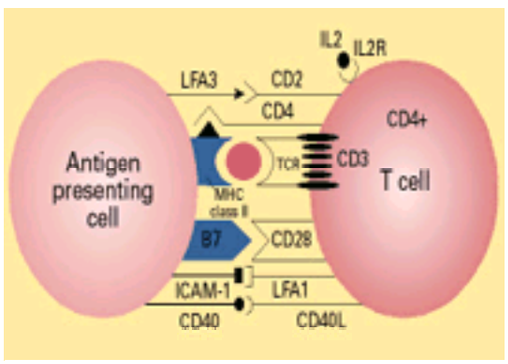


Fig. 7. Examples of some of the potential sites of action of monoclonal antibodies directed against targets on the cell surface of either the antigen-presenting cell or the CD4+ T cell.

Monoclonal antibodies directed against cell surface molecules involved in the cellular interactions of the immune response to an allograft may interfere with the response in a number of ways, such as deletion of a particular type of lymphocyte or blockade of part of the afferent or efferent aspect of the response, which in some instances will allow the development of a regulatory or suppressor cell population and specific unresponsiveness to an allograft.

OKT3

This represents a second generation of antithymocyte globulins, being a mouse monoclonal antibody directed against the CD3 molecule, which is intimately associated with the T-cell receptor and hence it is a pan-T-cell antibody. It modulates the CD3 molecule with its T-cell receptor and is extremely effective in reversing a steroid-resistant rejection in some 70 to 80per cent of instances (but not necessarily more effective than a good antithymocyte globulin).

Two major disadvantages, one specific for OKT3 and the other of all monoclonal antibody therapies, are first the OKT3 syndrome, associated usually with the first dose, and second the development of antibodies to either the idiotype or the isotype of the mouse globulin. The OKT3 syndrome is characterized by a high fever, dyspnoea, nausea, diarrhoea, and even anaphylactic shock, and is mediated in part by tumour necrosis factor. It can be diminished in intensity by prior administration of 1 g of methylprednisolone. Development of antibodies, either isotypic or idiotypic, to OKT3 abrogates the effect of the antibody, and occurs in most patients, preventing its further use.

OKT3 was the first monoclonal antibody to be widely used and was the forerunner of monoclonal antibody therapy in transplantation.

Anti-CD4

Antibodies directed against the T-helper cell (CD4+) population, the pivotal cells in graft rejection, are potent in experimental models of transplantation where they may produce tolerance to an organ allograft. Phase 1 clinical trials of depleting anti-CD4 antibodies have been carried out with encouraging results, but the prolonged depletion of CD4 T lymphocytes which followed a short course, often lasting for more than a year, was felt to be potentially hazardous and the studies were not continued. However, non-depleting antibodies against CD4 may prove to be equally effective. In experimental models, pretreatment of a recipient with donor antigen under the umbrella of anti-CD4 has produced a robust form of tolerance against a subsequent cardiac allograft and phase 1 studies of such a protocol in human recipients of a kidney transplant have been encouraging.

Anti-interleukin-2 receptor (anti-IL-2R)

The IL-2R is expressed on activated T cells, B cells, and macrophages. The IL-2R comprises three subunits: a (55 kDa), b (75 kDa), and g (64 kDa). Antibodies against the appropriate epitope of the a-chain of the IL-2R block the IL-2-driven proliferation of activated T cells, and are extremely immunosuppressive in experimental rodent models. Initial clinical trials of murine or rat anti-IL-2R antibodies used as prophylaxis during the first 10 days after transplantation have in general shown a very modest beneficial effect at best. However, the development of humanized antibodies has circumvented the previous immunization against the rodent antibodies allowing a longer half-life, and two such antibodies have now been shown in randomized trials to reduce the incidence of acute rejection in the first 6 months after transplantation. These antibodies are now available in clinical practice.

Anti-ICAM-1

This is one of a number of antibodies against adhesion molecules which are involved in the interaction between lymphocytes as well as between T lymphocytes and endothelium. Although they have proved to be immunosuppressive in experimental models, clinical trials have not shown them to influence the incidence of rejection nor on the whole have any effect on the outcome of delayed graft function in kidney transplantation, a rather disappointing finding.

Anti-CD40 ligand (anti-CD40L)

Blockade of the CD40/CD40L interaction appears to interfere with an important part of the accessory signalling pathway, and although used initially with blockade of the B7.1,7.2/CD28 accessory pathway by a fusion protein, CTLA4-Ig, to produce profound specific unresponsiveness both in rodent and primate models, it now appears that blockade of the CD40/CD40L pathway is all that is necessary. Clinical trials have commenced in renal and pancreatic islet transplantation with a humanized anti-CD40L antibody.

Anti-CD52 (CAMPATH-1)

CD52 is expressed on virtually all leucocytes except neutrophils and the rat anti-CD52 monoclonal antibody produces a profound depletion of lymphocytes in man. It was first used as a rat monoclonal antibody in kidney transplantation to treat rejection, but was associated with severe infections and the studies were not continued. It is widely used in bone marrow transplantation in Europe to purge bone marrow of T lymphocytes before implantation and has been used successfully in several patients with severe vasculitis to produce a remission. A humanized form is now available and has been used in cadaveric kidney transplantation in an uncontrolled phase 1 trial with low doses of cyclosporin to prevent rejection with encouraging results. More extensive studies are planned.

Total lymphoid irradiation

This is undoubtedly immunosuppressive in experimental models in the rodent and in primates where tolerance to organ allografts can be achieved. In clinical trials a course of total lymphoid irradiation before renal transplantation has proved to be effective in preventing rejection of renal allografts with the use of minimal immunosuppression after transplantation. Indeed there are several well documented cases of long-term survival in patients receiving no immunosuppressive drugs at all! However, the logistics of using total lymphoid irradiation (up to 21 daily treatments) combined with dialysis before transplantation and the advent of more potent immunosuppression with cyclosporin have led the centres pursuing this therapy to abandon it for the time being.

Complications of immunosuppressive therapy

Apart from the specific side-effects of the drugs or antibodies used in transplantation there are three major side-effects of current non-specific immunosuppression, all of which are described in more detail in the following chapters. These are infection (especially viral infections), cancer, and cardiovascular disease. The increased incidence of cancer is particularly evident in those cancers with a putative viral aetiology - non-Hodgkin's lymphoma, squamous cell carcinoma of the skin, Kaposi's sarcoma, and cervical cancer in women - but all cancers show some increase in incidence. Aggressive atheromatous disease occurs in many transplant recipients, and although associated with an increased incidence of hypertension and hyperlipidaemia, which are side-effects of many of the immunosuppressive drugs that we use, it may also be related to the non-specific immunosuppression induced by therapy, a possible mechanism being unclear at this time.

Further reading

Calne RY *et al.* Cyclosporin A in patients receiving renal allografts from cadaver donors. *Lancet* 1978; **ii**: 1323–7. [The first report of cyclosporin in renal transplantation.]

Eugui EM, Almquist SJ, Muller CD, Allison AC. Lymphocyte-selective anti-proliferative and immunosuppressive effects of mycophenolic acid *in vitro*: role of deoxyguanosine nucleotide depletion. *Scandinavian Journal of Immunology* 1991; **33**: 161–73. [An early report of the immunosuppressive effect of mycophenolate mofetil.]

European Mycophenolate Mofetil Cooperative Study Group. Placebo-controlled study of mycophenolate mofetil combined with cyclosporine and corticosteroids for prevention of acute rejection. *Lancet* 1995; **345**: 1321–5. [One of the three classic trials of mycophenolate mofetil.]

Ginns LC, Cosimi AB, Morris PJ, eds. *Immunosuppression in transplantation*. Blackwell Science, Boston, 1999. [A recent book reviewing current immunosuppression in organ transplantation.]

Groth CG, Backman L, Morales JM, *et al.* Sirolimus (rapamycin)-based therapy in human renal transplantation: similar efficacy and different toxicity compared with cyclosporine. Sirolimus European Renal Transplant Study Group. *Transplantation* 1999; **67**: 1036–42. [A randomized trial comparing sirolimus with cyclosporine and showing similar efficacy but with different safety profiles.]

Jones RM, Murie JA, Allen RD, Ting A, Morris PJ. Triple therapy in cadaver renal transplantation. *British Journal of Surgery* 1988; **75**: 4–8. [An early report of combination therapy in renal transplantation.]

Kahan B. Cyclosporin. *New England Journal of Medicine* 1989; **321**: 1725–38. [A comprehensive review of the pharmacokinetic and clinical experience at that time.]

Kahan BD *et al.* Immunosuppressive effects and safety of a sirolimus/cyclosporine combination regimen for renal transplantation. *Transplantation* 1998; **66**: 1040–6. [The first demonstration in the clinic of the potency of cyclosporin and sirolimus.]

Kobashigawa J *et al.* A randomized active-controlled trial of mycophenolate mofetil in heart transplant recipients. *Transplantation* 1998; **66**: 507–15. [Mycophenolate mofetil did not reduce the incidence of rejection but improved graft survival at 1 year.]

Kohler H, Milstein C. Continuous culture of fused cells secreting antibody of predefined specificity. *Nature* 1975; **256**: 495–7. [The first description of the development of monoclonal antibody

technology.]

Mayer AD *et al.* Multicenter randomized trial comparing tacrolimus (FK506) and cyclosporine in the prevention of renal allograft rejection: a report of the European Tacrolimus Multicenter Renal Study Group. *Transplantation* 1997; **64**: 436–43.

Morris PJ. Cyclosporin A. *Transplantation* 1981; **32**: 349–54. [A comprehensive review of the early experimental and clinical experience with cyclosporin.]

Morris PJ, ed. *Kidney transplantation: principles and practice*, 5th edn. WB Saunders, Philadelphia, 2000. [A comprehensive text on all aspects of kidney transplantation.]

Morris PJ, Chan L, French ME, Ting A. Low dose oral prednisolone in renal transplantation. *Lancet* 1982; **i**: 525–7. [The first prospective trial of low-dose oral prednisolone in renal transplantation.]

Ortho Multicenter Transplant Group. A randomized clinical trial of OKT3 monoclonal antibody for acute rejection of cadaveric renal transplants. *New England Journal of Medicine* 1985; **313**: 337–42. [The first trial of a monoclonal antibody.]

Pirsch JD *et al.* A comparison of tacrolimus (FK506) and cyclosporine for immunosuppression after cadaveric renal transplantation. FK506 Kidney Transplant Study Group. *Transplantation* 1997; **63**: 977–83.

Sollinger HW for the US Renal Transplant Mycophenolate Mofetil Study Group. Mycophenolate mofetil for the prevention of acute rejection in primary cadaveric renal allograft recipients. *Transplantation* 1995; **60**: 225–32. [One of the three classic trials of mycophenolate mofetil.]

The Tricontinental Mycophenolate Mofetil Renal Transplant Study Group. A blinded, randomized clinical trial of mycophenolate mofetil for the prevention of acute rejection in cadaveric renal transplantation. *Transplantation* 1996; **61**: 1029–37. [One of the three classic trials of mycophenolate mofetil.]

Waldmann H. Manipulation of T-cell responses with monoclonal antibodies. *Annual Review of Immunology* 1989; **7**: 407–44. [A review of the potential of monoclonal antibodies for therapy.]

Winter G, Milstein C. Man-made antibodies. *Nature* 1991; **349**: 293–9. [A major article describing the approach to genetically engineered antibodies.]

Wood KJ, Pearson TC, Darby C, Morris PJ. CD4: a potential target molecule for immunosuppressive therapy and tolerance induction. *Transplantation Reviews* 1991; **5**: 150–64. [Pretreatment with antigen under an umbrella of anti-CD4 produces tolerance to an allograft.]

Woodle SE *et al.* A multicenter trial of FK506 (tacrolimus) therapy in refractory acute renal allograft rejection. *Transplantation* 1996; **62**: 594–9. [A randomized trial showing that tacrolimus can reverse steroid resistant rejection.]

16.5 Kidney transplantation

Peter J. Morris

[Introduction](#)
[Indications for renal transplantation](#)
[Organ donation](#)
- [Living related donors](#)
- [Living unrelated donors](#)
- [Cadaver donors](#)
- [Removal of kidneys](#)

[Organ preservation](#)
[HLA matching and the crossmatch](#)
[Operation](#)
[Postoperative course](#)
- [Immediate function](#)
- [Delayed or non-function of the transplanted kidney](#)
- [Deterioration of function in a functioning graft](#)
- [Rejection](#)

[Complications of renal transplantation](#)
- [Technical complications](#)
- [Immunological complications](#)
- [Cardiovascular complications](#)

[Results of transplantation](#)
- [Living related grafts](#)
- [Cadaveric grafts](#)
- [Factors influencing graft outcome](#)

[Rehabilitation](#)
[Future developments](#)
[Further reading](#)

Introduction

Transplantation of the kidney has become the treatment of choice for endstage renal failure in most age groups with perhaps the exception of the very young and the very old. That this has happened over the past 40 years represents one of the most significant developments in medicine in this century. Furthermore, advances in immunosuppression, histocompatibility, and organ preservation in the field of kidney transplantation have led to the successful transplantation of other organs, in particular the liver and heart.

The modern era of kidney transplantation began with the pioneer work of David Hume in Boston around 1950 with the transplantation of cadaver kidneys into non-immunosuppressed recipients. The kidneys were implanted in the thigh anastomosing the femoral vessels to the renal vessels with drainage of the ureter on to the skin. Some of these kidneys functioned for several weeks before being rejected. In 1954 Joseph Murray transplanted a kidney between identical twins where no immunosuppression was required, on this occasion placing the kidney retroperitoneally in the iliac fossa, essentially the technique as used today. This operation was a milestone in the field of transplantation as it confirmed that a successfully transplanted kidney in the absence of rejection was capable of essentially normal renal function.

Work continued in Boston, Paris, and the United Kingdom using total body irradiation of the recipient in an effort to prevent rejection of the transplanted kidney with some surprisingly good results in the short term, some kidneys functioning for several months. However, the discovery of the immunosuppressive properties of 6-mercaptopurine by Schwarz and Damashek in Boston in 1959 quickly led to the demonstration by Calne in England and Zukowski and Hume in Richmond, Virginia, that this agent could prevent rejection of kidneys in dogs. Elion and Hitchings, who had produced 6-mercaptopurine at Burroughs Wellcome, then developed azathioprine, of which 6-mercaptopurine is a metabolite. In further studies by Calne, working with Murray in Boston, azathioprine appeared to be perhaps less toxic than 6-mercaptopurine and very quickly was introduced as the standard immunosuppressive drug in clinical renal transplantation. Azathioprine, together with steroids, remained the conventional immunosuppressive therapy for 20 years until cyclosporin emerged on the clinical scene.

Indications for renal transplantation

Today it can be truly said that there are no absolute contraindications to renal transplantation, and all patients with endstage renal failure are potential candidates for renal transplantation (Table 1). Of the metabolic disorders causing endstage renal failure, type I diabetes is by far the most common, ranging from 10 to 25 per cent of all patients coming to transplantation. Renal failure due to oxalosis has been considered an absolute contraindication to transplantation, but successful transplantation has been reported in this condition by methods directed at prevention of the deposition of oxalate in the kidney by maintenance of a copious urine output and the administration of pyridoxine, phosphate, and magnesium solutions. Today the ideal management would require a combined liver and kidney transplant to correct both the metabolic defect and the endstage renal failure.



Table 1 Indications for transplantation

In addition the improved safety and decreased side-effects of modern immunosuppression allows the transplantation both of young children and the elderly patient with very acceptable results at least in the short and medium term. In the case of infants born with renal failure due to congenital abnormalities of the urogenital tract, it is now considered preferable to maintain them on peritoneal dialysis until they are 2 or 3 years of age before giving them a renal transplant.

Recurrence of the original disease has not proved to be as big a problem as originally envisaged in the early years of renal transplantation. Although recurrence is relatively common in certain of the nephritides (e.g. mesangiocapillary type II (dense deposit disease), IgA glomerulonephritis, and focal glomerulosclerosis), loss of the kidney due to the recurrence in these conditions is relatively uncommon, at least within a 5-year period (Table 2).

	Approximate frequency of recurrence (%)	*Loss of graft*
Primary kidney disease		
Focal segmental glomerulosclerosis	30	Common
IgA nephropathy	50	Uncommon
Henoch–Schönlein purpura	80	Uncommon
Membranous nephropathy	10	Uncommon
Mesangiocapillary type I	25	Common
Mesangiocapillary type II	95	Uncommon
Systemic disease with kidney involvement		
Haemolytic uraemic syndrome	50	Common
Diabetes type I	100	Uncommon
Oxalosis	90	Common
Amyloidosis	33	Uncommon

Table 2 Diseases with a high recurrence rate in renal allografts and the likelihood of loss of the graft

Organ donation

Living related donors

From the point of view of graft outcome, living related transplantation still provides better long-term results than cadaver transplantation. In addition the ever-increasing shortage of cadaver donors means that living related transplantation remains an important option for the patient with endstage renal failure. The selection of a donor from within a family is based first on the motivation of the potential donor(s), which should be truly altruistic without coercion from the patient or other members of the family. Second, it requires blood group compatibility and the optimal HLA match in the presence of a negative crossmatch between recipient and donor (see later). The ideal combination is that between HLA identical siblings and next that between siblings disparate for one HLA haplotype or parent to child (see Chapter 16.1). Once it has been established that a family member is a suitable donor on the above psychological and immunological grounds, the donor must undergo an extensive medical evaluation, including a full investigation of renal function, and finally an angiogram to define the renal vasculature on each side.

The risks to the donor are not inconsiderable. First there are the risks associated with a major operation including a small risk of death (0.03 per cent), and second the potential long-term risk to a healthy donor of having one kidney. However, follow-up studies of donors for up to 20 years have revealed no excess risk of hypertension or renal failure over and above that of a normal age-matched population, and indeed there is evidence that living related donors have an increased life expectancy!

Living unrelated donors

There has been considerable controversy in this area and there is obviously opportunity for commercial exploitation, as seen in many developing countries. However, a case can be made for the use of living unrelated donors where there is clearly a strong emotional relationship between donor and recipient, as for example between spouses. Surprisingly, although it would not be expected that the use of a living unrelated donor would provide any greater chance of a successful outcome than the use of a cadaver donor, there is now substantial evidence showing that the outcome of living unrelated transplantation is better than cadaver transplantation. This is assumed to be due to the lack of any damage to the donor kidney from the living unrelated donor in contrast to the cadaver kidney which undergoes damage around the time of death and again as a result of ischaemia reperfusion injury.

Cadaver donors

Most kidney donors are cadaver donors in whom brainstem death has been diagnosed while respiratory and cardiac function is being maintained on a ventilator. The usual cause of brain death is irreversible brain damage due to head injury or intracerebral haemorrhage. It is important that renal function is normal and that urine output is maintained before and after the final diagnosis of brain death has been made. Most of the parameters of brain death reflect brainstem death, for in the absence of brainstem function, respiratory and cardiac function cannot be maintained. The presence of widespread bacterial infection, cancer, or a positive test for hepatitis B, hepatitis C, or HIV would exclude a potential cadaver donor from further consideration. A history of hypertension or diabetes might be a relative contraindication to use of the kidneys, but where doubt exists the kidneys can be removed and an immediate biopsy performed before a final decision is made. In general, very young donors (less than 1 year) or very old donors (over the age of 70) are not considered suitable.

Removal of kidneys

In living kidney donors the selected kidney (usually the left as it provides a longer renal vein) is removed either through a flank incision with the patient in the lateral position or transperitoneally through a midline or transverse abdominal incision. More recently, laparoscopic-assisted nephrectomy has been developed and is being promoted with enthusiasm, in part because of patient wishes, but its true role is still to be evaluated.

Kidneys are removed from the cadaver donor *en bloc* following *in situ* perfusion, with subsequent dissection and perfusion of the kidneys after removal. Today the removal of kidneys is usually part of a multiorgan retrieval procedure, and then the kidneys are the last organs to be removed.

Organ preservation

Although kidneys can be preserved by continuous perfusion with colloid solutions such as oxygenated human albumin for periods up to 72 h, machine preservation has been largely replaced by simple flushing of the kidney with one of a number of preservation solutions and storage in ice slush. This provides satisfactory preservation for up to 48 h, which is generally long enough for selection of the appropriate recipients on medical and matching grounds and transport of the kidneys within a national or even international network of organ exchange. The preservation solutions most widely used are Euro Collin's and hypertonic citrate, both solutions having in common a potassium ion content similar to that within the cell. More recently the University of Wisconsin (**UW**) solution, the preservation fluid of choice in liver transplantation, has become more widely used in kidney transplantation as well. The important components of the UW solution are probably lactobionate and adenosine. Over the last few years the recognition of ischaemia reperfusion injury in transplanted cadaver kidneys has led to efforts to shorten the cold ischaemia storage time after retrieval.

HLA matching and the crossmatch

Matching for HLA, the major histocompatibility complex (**MHC**) in man, presents major logistic problems in cadaver transplantation because of the genetic complexity of the HLA system. For this reason national or international organ exchange networks have developed, such as Eurotransplant, UK Transplant Support Service Authority (**UKTSSA**), and in the United States the United Organ Sharing scheme (**UNOS**), to try and ensure that as many recipients receive a beneficially matched kidney for the HLA A, B, and DR antigens as possible. That this is worth trying to achieve is beyond question as illustrated by recent UNOS and UKTSSA data (Fig. 1 and Fig. 2). However, many recipients with uncommon HLA phenotypes will be excluded on this basis from ever receiving a transplant. The approach to HLA matching can be simplified by matching just for HLA DR, which still provides an improved graft outcome that is not much less than that obtained by matching for the whole of the MHC complex, especially in the patient who is being regrafted. Although the benefits of matching in terms of cadaveric graft outcome are well established, failure to find an acceptable match for a recipient should not preclude that recipient from receiving a mismatched cadaver kidney with the more potent immunosuppression available today. However, this does mean accepting the higher risks of sensitization if the graft fails. Serological typing for HLA is now being replaced by DNA typing, which can be very precise, no slower to perform, and can be done easily on blood samples obtained before death.

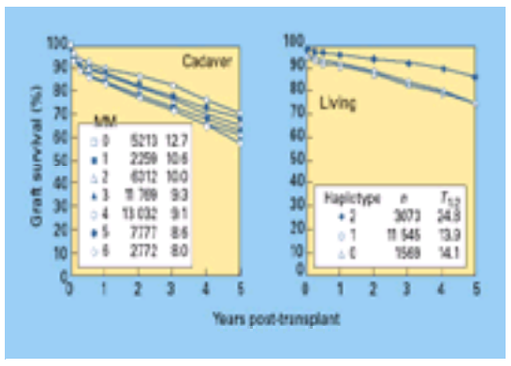


Fig. 1. The effect of HLA matching on outcome of cadaver and living related transplants between 1991 and 1997 in the UNOS Scientific Renal Transplant Registry. MM, mismatches for HLA A, B, and DR. Haplotype, the number of HLA haplotypes shared between donor and recipient. In both the number of transplants analysed and the predicted half-lives are given. (Reproduced with permission from Cecka JM 1999.)

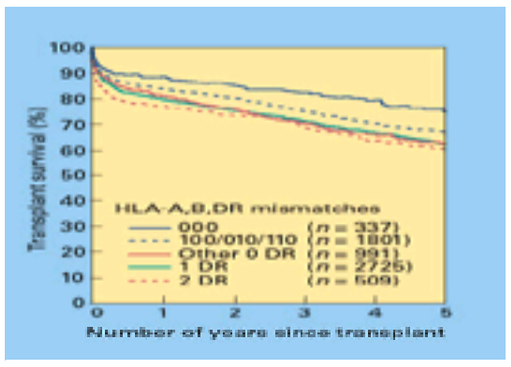


Fig. 2. The effect of HLA matching on outcome of first cadaveric transplants between 1986 and 1993 in the UKTSSA Renal Transplant Registry. (Reproduced with permission from Morris PJ et al. 1999).

Perhaps the most important aspect of the matching procedure in cadaver transplantation is the crossmatch in which donor lymphocytes are added to serum from the recipient in the presence of rabbit serum as a source of complement. If the donor lymphocytes are lysed, this is known as a positive crossmatch. For many years, following the recognition of hyperacute rejection of a transplanted kidney in the presence of a positive crossmatch between donor and recipient, a positive crossmatch was considered to be an absolute contraindication to renal transplantation.

In more recent years the complexity of antibodies found in the sera of sensitized recipients has been recognized, and indeed not all antibodies are directed against HLA antigens. Thus an apparently sensitized recipient may have antibodies, such as autoantibodies, which react with donor lymphocytes giving a positive crossmatch but which do not damage a subsequent kidney from that donor. As most patients who are sensitized by previous pregnancies, blood transfusions, or failed transplants have a mixture of antibodies including antibodies against HLA, it is important to define the class and specificity of these antibodies before transplantation as this helps in the interpretation of a positive crossmatch. Table 3 summarizes the types of antibodies in a sensitized recipient which may give rise to a positive crossmatch, and whether a successful transplant is possible in the presence of that positive crossmatch. This increased understanding of sensitization is the major reason why patients with regrafts do virtually as well as patients receiving a primary graft (Fig. 3).

Specificity of antibody	*Class*	*Current positive*	*Historical positive*	*Transplant contraindicated*
HLA class I	IgG	+	+	Yes
	IgM	+	+	Yes
	IgG	–	+	Yes
	IgM	–	+	No
HLA class II	IgG	+	+	Yes
	IgM	+	+	Yes
	IgG	–	+	?
	IgM	–	+	No
Autoantibody	IgM	+	+	No
	IgM	–	+	No
Non-HLA	IgG	+	+	No
	IgM	+	+	No

A sensitized recipient may have antibodies which are directed against HLA class I or II antigens, autoantigens, or non-HLA antigens, or a mixture of these antibodies. A positive crossmatch between recipient serum and donor lymphocytes may be due to any of the above and may be obtained with serum taken at the time of transplantation (current positive) or at some earlier time, for instance 6 months or a year before transplantation (historical positive). The specificity and class of antibody as well as the timing of the antibody appearance in relation to the transplant are important in determining whether a renal transplant is contraindicated in the presence of a positive crossmatch.

Table 3 The types of antibodies in a sensitized recipient

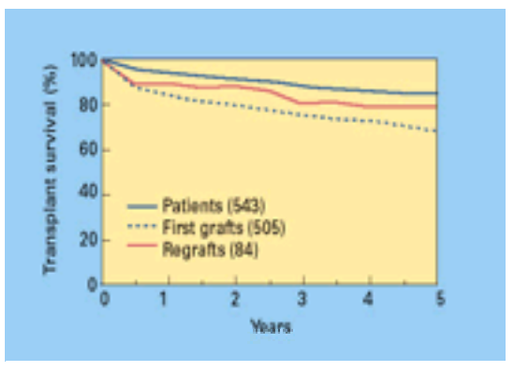


Fig. 3. Actuarial patient and cadaveric graft survival in Oxford from 1991 to 1999. All patients were immunosuppressed with triple therapy (cyclosporin, azathioprine, and prednisolone).

Operation

A renal transplant has been a very standard operation for many years now. The iliac vessels are exposed retroperitoneally through an oblique incision in one or other iliac fossa, the oblique muscles being divided in the line of the incision and the peritoneum reflected upwards and medially (Fig. 4). The renal vein is anastomosed end-to-side to the external iliac vein and the renal artery anastomosed either end-to-end to the divided internal iliac artery or, as is more usual with a cadaver kidney where the renal artery will have a cuff of aorta, end-to-side to the external iliac artery (Fig. 5).

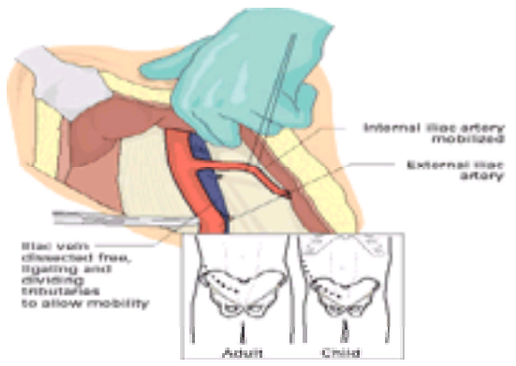


Fig. 4. Incisions for adult and child and the iliac arteries and vein exposed after reflection of the peritoneum upwards and medially.

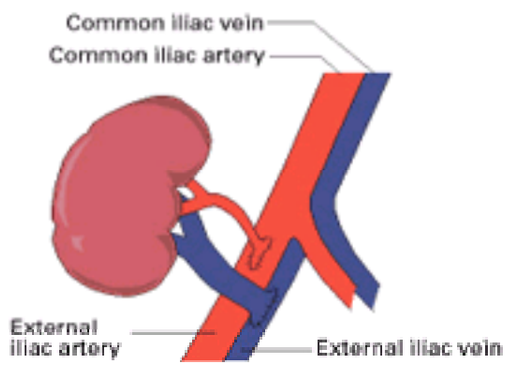


Fig. 5. The vascular anastomoses of a renal transplant showing an end-to-side anastomosis of the renal artery with a cuff of aorta to the external iliac artery and an end-to-side anastomosis of the renal vein to the external iliac vein.

The ureter is implanted in the bladder (ureteroneocystostomy) either through an anterior cystotomy with a submucosal tunnel to prevent reflux or as an extravesical procedure where the ureter is anastomosed to the mucosa of the dome of the bladder with or without a stent and then muscle is drawn over the ureter, again with a view to preventing reflux. Whether the latter technique does prevent reflux is questionable. Another technique used successfully by some units (in particular at Massachusetts General Hospital) is to create a pyeloureterostomy in which the pelvis of the kidney is anastomosed to the recipient ureter.

Although not essential, usually a right kidney is implanted in the left iliac fossa and vice versa as this facilitates the vascular anastomoses. The wound is closed without drainage where possible, and an indwelling catheter left in the bladder for 1 to 5 days depending on the technique used.

Postoperative course

Immediate function

A kidney from a living related donor would always be expected to function immediately whereas only 60 to 80 per cent of cadaver kidneys will do so. Immediate function is usually associated with an osmotic diuresis and urine output over 24 h is anywhere from 5 to 25 litres. This does require meticulous attention to fluid replacement and is made easier if there is a central venous pressure line in place over the first 48 h after surgery. After 48 to 72 h the kidney will begin to concentrate and the intravenous fluid replacement can be slowed and replaced by oral fluids.

Delayed or non-function of the transplanted kidney

In this situation, which is usually due to acute tubular necrosis, other causes of delayed or non-function must be considered (Table 4). Duplex ultrasound examination of the kidney will demonstrate whether the kidney is perfused and also show evidence of obstruction or a urine leak. Finally, hyperacute or accelerated rejection can be excluded by renal biopsy.

Acute tubular necrosis
Renal artery thrombosis
Renal vein thrombosis
Hyperacute rejection
Ureteric obstruction
Urine leak

Table 4 Possible causes of delayed function or cessation of function in the first few days after renal transplantation

Deterioration of function in a functioning graft

Where there is a functioning kidney with a falling or stable serum creatinine which then begins to rise, rejection or cyclosporin nephrotoxicity is the most likely cause. Cyclosporin levels will exclude the latter while the former is confirmed by biopsy. An ultrasound examination will also exclude the presence of ureteric obstruction or leakage. If a renal vein or arterial thrombosis is suspected, both the ultrasound examination and a renogram will clarify the diagnosis.

Rejection

Four types of rejection may be seen after transplantation.

Hyperacute rejection

This is an immediate antibody-mediated rejection occurring within the first 24 h of transplantation in the presence of a positive crossmatch due to HLA class I antibody or an ABO incompatible kidney. The classic hyperacute rejection (which is evident within 60 min of revascularization of the kidney) is not seen today, unless there has been an error in the interpretation of the crossmatch or rarely an incompatible ABO kidney is transplanted in error, but less florid examples are not uncommon and present as non-function of a kidney in a sensitized patient, more usually in a patient receiving a regraft.

Accelerated rejection

This is also rejection in a recipient specifically sensitized against the donor and is mediated by cells. It occurs between 2 and 4 days after transplantation, presenting with a deterioration of renal function often accompanied by a fever. A biopsy will show a florid cellular infiltration within the graft, often associated with arteriolar thrombosis and interstitial haemorrhage in severe cases.

Acute rejection

This is the outcome of the immune response to an allograft in a non-sensitized patient. Most commonly the first rejection episode will be apparent between 7 and 21 days after transplantation, but may present at any time after transplantation. Some 50 per cent of recipients of a cadaver allograft will have at least one acute rejection episode during the first 3 months after transplantation with current cyclosporin-based immunosuppressive protocols. In a classic example of an acute rejection episode the patient would present with a modest fever, the graft would be swollen and tender, and there would be a concomitant fall in urine output and a rise in serum creatinine. This picture is relatively uncommon today with current cyclosporin-based immunosuppression, where there may be a mild fever with minimal or no graft swelling and tenderness, the main feature being a deterioration in renal function as evidenced by a rise in serum creatinine. A biopsy will confirm the diagnosis. The differential diagnosis will be first, cyclosporin nephrotoxicity and second, ureteric obstruction. Blood levels of cyclosporin and an ultrasound examination of the kidney, respectively, can be used to exclude these possibilities.

Chronic rejection

This is an insidious process represented by a steady deterioration in function at any time following the first few months of transplantation. A biopsy will confirm the diagnosis, which is characterized by myointimal hypertrophy of arterioles and arteries within the kidney, glomerulopathy, and interstitial fibrosis. Other diagnoses to be considered are ureteric obstruction, renal artery stenosis, and recurrence of the primary disease.

Immunopathology of rejection

Hyperacute rejection

If a biopsy is taken from the kidney within 1 h of revascularization (i.e. before wound closure), this may reveal a marked infiltration of the kidney with polymorphonuclear leucocytes (Fig. 6). Subsequent biopsies at 24 h and later will reveal widespread glomerular capillary and arteriolar thrombosis with oedema, fibrinoid necrosis of arteriolar walls, and interstitial haemorrhage. Immunofluorescence will demonstrate deposition of IgM, IgG, and complement within the transplanted kidney.

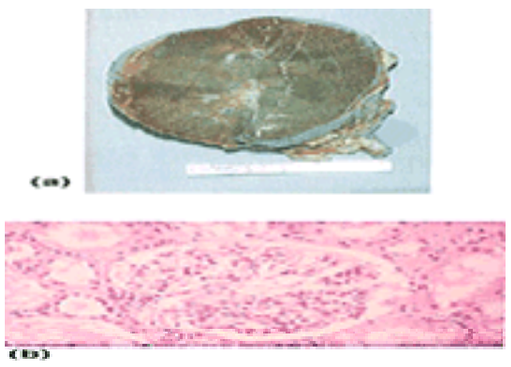


Fig. 6. (a) A hyperacutely rejected kidney on removal at 72+ h. (b) A biopsy 20 min after revascularization showing an intense polymorphonuclear infiltrate within the kidney.

Accelerated and acute rejection

The hallmark of these rejection reactions is the mononuclear cellular infiltrate, the density of which indicates the severity of the rejection, as does infiltration of the tubules themselves (Fig. 7). Associated with the cellular infiltrate are oedema and, in severe cases, interstitial haemorrhage and fibrinoid necrosis of arterioles. The cellular infiltrate contains approximately 65 per cent macrophages and 30 per cent T lymphocytes with CD8+ T lymphocytes being rather more prominent than CD4+ T lymphocytes. A small percentage of the infiltrate comprises natural killer cells and eosinophils. A smaller percentage of the T lymphocytes will exhibit activation antigens such as the IL-2 receptor. In addition there will be an induction of expression of MHC class II antigens on cells such as proximal tubular cells, which do not express these antigens normally.

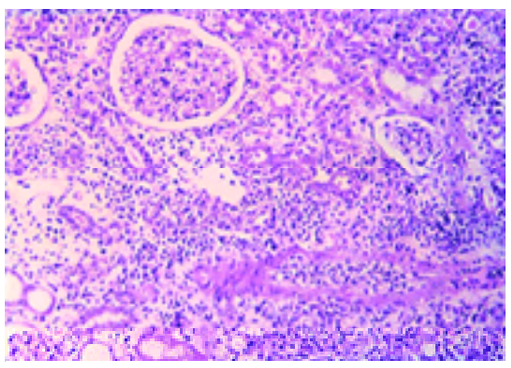


Fig. 7. Severe acute rejection with a dense mononuclear cellular infiltrate 10 days after transplantation.

Chronic rejection

The major features of this chronic process are interstitial fibrosis within the parenchyma of the kidney and intimal hyperplasia of arteries within the parenchyma (Fig. 8).

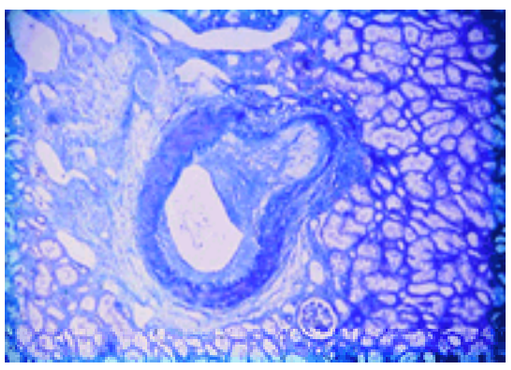


Fig. 8. Changes of chronic rejection in a medium-sized renal artery.

Complications of renal transplantation

The complications of transplantation can be considered as technical or related to the immunosuppressive therapy itself, much of which is secondary to immunosuppression induced in the recipient by the drugs or directly related to the drugs.

Technical complications

Vascular complications

Renal artery thrombosis is very uncommon and usually presents within the first 2 weeks of transplantation as an abrupt cessation of renal function with no other clinical features. Renal vein thrombosis occurs much more commonly than hitherto, and may be related to the thrombogenic potential of cyclosporin. It usually occurs within the first few days of transplantation and is associated with an abrupt cessation of renal function associated with a painful, swollen, tender graft. Very occasionally, immediate re-exploration may enable thrombectomy and salvage of the graft. The prophylactic use of low-dose aspirin (75 mg/day) for the first month after transplantation has virtually eliminated this complication in Oxford.

Renal artery stenosis may present at any time from several months to several years after transplantation, and although the stenosis may occur at the anastomosis, it usually occurs just distal to the anastomosis. The patient presents with poorly controlled hypertension and deteriorating renal function (see later).

Urological complications

A urine leak, either from the bladder or the ureteric anastomosis, occurs within the first 6 weeks of transplantation and is most commonly due to ischaemia of the lower end of the ureter, sometimes due to a poorly removed kidney, with skeletonization of the ureter or interruption of the ureteric blood supply within the hilum of the kidney. A bladder leak will heal with catheter drainage of the bladder, whereas ureteric necrosis is dealt with by excising the necrotic ureter and either reimplanting the ureter in the bladder over a ureteric stent, or anastomosing the ureter to the patient's own ureter either end-to-side or end-to-end again over a stent.

Obstruction of the ureter presents at any time from weeks to years after transplantation and although the site of obstruction is usually in the lower one-third of the ureter, no doubt as a result of ischaemia, a significant number of pelviureteric obstructions are seen. Diagnosis is confirmed by ultrasound examination of the graft and an antegrade pyelogram. Immediate relief of the obstruction can be obtained by the insertion of a percutaneous ureteric stent or a percutaneous nephrostomy if this is not possible. Sometimes a stenosis can be dilated with a balloon, but otherwise corrective surgery is required, involving either reimplantation of the normal ureter if the stenosis is at the entry of the ureter into the bladder or a pelviureteric anastomosis using the patient's own ureter for a more proximal stenosis.

The incidence of urological complications after transplantation is now quite low in experienced units, being no higher than 5 per cent, which represents a marked improvement over that seen 15 to 20 years ago. This improvement is due to better techniques of kidney retrieval and implantation of the ureter, but in particular to the general use of low-dose steroids after transplantation in more recent years.

Lymphocele

This is a collection of lymph around the graft, which arises from divided host lymphatics around the iliac vessels. This complication has also become relatively uncommon since it has been recognized that lymphatics must be ligated with silk rather than occluded with diathermy. A lymphocele is retroperitoneal and as such produces symptoms from pressure on surrounding structures. Thus, it may present with deteriorating renal function due to pressure on the transplanted ureter, a swollen leg on the side of the graft due to pressure on the iliac vein, and tenesmus or strangury due to pressure on the rectum and bladder, respectively. The diagnosis is confirmed by ultrasound examination (Fig. 9). It is treated in the first instance by percutaneous drainage, but if it recurs more than twice after percutaneous drainage it needs to be drained by fenestration into the peritoneal cavity. Sometimes this can be done satisfactorily by a laparoscopic approach rather than an open transperitoneal operation.

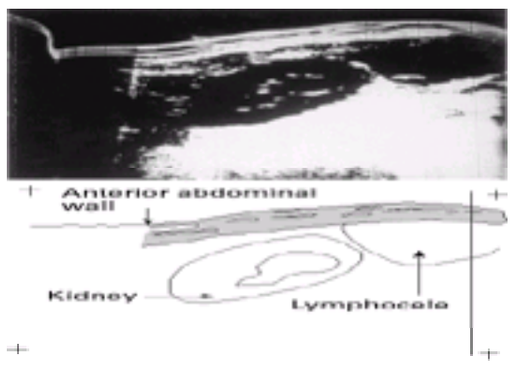


Fig. 9. A lymphocele adjacent to the renal transplant shown by ultrasound.

Immunological complications

Infection

Bacterial infections

Not only bacterial infections, but viral and protozoal infections may occur in transplant patients due in part or wholly to the immunosuppressive therapy given to prevent and treat rejection. The time at which an infection appears in relation to transplantation is important in arriving at a diagnosis, for viral (with the exception of herpes simplex) and protozoal infections are rare in the first month after transplantation. Pneumonia as a postoperative complication and urinary tract infections are common in the early weeks after transplantation. Asymptomatic urinary tract infections are relatively common due mainly to the indwelling catheter left in the bladder for several days after surgery. Wound infections are uncommon with an incidence of 1 to 2 per cent. A preventive antibiotic should be given with the induction of anaesthesia to cover contamination at the time of operation and in particular to cover the possibility of the transplanted kidney being contaminated. An appropriate broad-spectrum antibiotic is used, such as cefuroxime.

Of the mycobacterial infections, tuberculosis is by far the most common and generally presents as a chest infection with fever, cough, and an infiltrate on radiography. However, tuberculosis may uncommonly present as an abscess or joint infection. Any patients who are from areas where tuberculosis is common, such as the Indian subcontinent, or who have a history of tuberculous infection in the past should receive prophylactic therapy for 1 year after transplantation (e.g. treatment with isoniazid).

Viral infections

The most common and potentially the most serious of the viral infections is cytomegalovirus (**CMV**) infection. This may be a primary infection in a CMV-seronegative recipient who receives a kidney from a seropositive donor or reactivation of the virus in a seropositive recipient. The likelihood of a CMV infection occurring is directly related to the potency of the immunosuppression used. The primary infection is more severe, but clinically in both instances the presenting feature will be a high fever (greater than 39°C) associated with a leucopenia. This high swinging fever may last 7 to 10 days and characteristically the patient feels unwell only during the time of the fever. Diagnosis of a primary infection has been confirmed by seroconversion in a seronegative patient or a rise in titre in a seropositive patient, but this does not occur till several days after the onset of fever. Today a diagnosis can be made quite early and before clinical symptoms appear by polymerase chain reaction and/or CMV antigenaemia tests on blood, and hence patients at risk can be monitored every few days and treatment commenced when evidence of infection is provided. More overt infection such as pneumonitis or hepatitis heralds a more serious problem and is potentially fatal in a primary CMV infection. For this reason it is desirable, but often not achievable, to transplant a seronegative recipient with a seronegative kidney, for primary infection does not occur in this situation provided that the

recipient only receives CMV seronegative blood if transfusion is required (Table 5). Vaccination of seronegative recipients before transplantation has been explored and the results of a large, international, randomized, blinded, multicentre trial showed that the rate of infection remained the same, but there were less severe infections in the vaccinated recipients. Prophylactic administration of hyperimmune globulin has also been advocated, but with little evidence to support its use in this way. Ganciclovir, an antiviral agent which has proved valuable in the treatment of an overt CMV infection, is now widely used prophylactically in patients at high risk of developing CMV infections, such as seronegative recipients given a seropositive kidney, especially if given ATG or OKT3 to prevent or treat rejection, when it is almost inevitable that a CMV infection will occur in these patients.

	Recipient CMV seropositive		*Recipient CMV seronegative*	
	Donor CMV+ve	*Donor CMV-ve*	*Donor CMV+ve*	*Donor CMV-ve*
Group total	84	98	60	64
No. infected	52 (62%)	52 (53%)	38 (63%)	0
	Reactivation or reinfection		Primary infection	

Table 5 Effect of donor CMV status on CMV infection in 306 renal allograft recipients at Oxford

Herpes simplex infections around the mouth or less commonly the genitalia are frequent but resolve quickly with prompt treatment with systemic acyclovir. Acyclovir has been used prophylactically in patients with a history of herpes infections in the early weeks after transplantation when immunosuppression is at its highest level, but is probably only justified in the case of a history of genital herpes.

Varicella zoster infections present rarely as chickenpox in recipients never previously exposed, in which case the infection may be fulminating, but present commonly as shingles (Fig. 10). Acyclovir is the treatment of choice.



Fig. 10. Herpes zoster of the left side of the face in a man 2 years after renal transplantation.

Hepatitis B in the donor is a contraindication to transplantation, but it is not uncommon for recipients of a renal transplant to be hepatitis B positive. The outcome in those patients who are positive and immunosuppressed is uncertain in that liver failure may develop in some, but not all, patients after several years.

Hepatitis C, the cause of non-A, non-B hepatitis, in the donor is also a contraindication to transplantation, but is not uncommon amongst transplant recipients (approximately 5 to 10 per cent in some centres). Evidence is now accumulating that this represents a longer-term risk of liver failure in the recipient who is immunosuppressed following transplantation.

Protozoal infections

Pneumocystis carinii is the most important infection in this group and occurs at any time after the first month of transplantation, the patient usually presenting with a dry non-productive cough and a low-grade fever. A chest radiograph may show patchy consolidation in one area of the lung at the early stages of the infection. A bronchoscopy and bronchial lavage should be carried out promptly if no alternative diagnosis is available, for cytological examination of the washings will inevitably confirm the diagnosis of pneumocystis. High-dose septrin is the appropriate treatment. Pneumocystis is rare if prophylactic septrin (0.5 g/day) is used in the immunosuppressed patient and most units now use such a protocol at least for the first 6 months after transplantation.

Fungal infections

Candida infections of the mouth are not uncommon and prophylactic amphotericin lozenges are used during the first few months after transplantation. Aspergillosis is perhaps the most common of the other fungal infections, but is still relatively uncommon, and its usual presentation is as a chest infection. Again the diagnosis is established by bronchoscopy and bronchial lavage.

Cancer

Although there is an increased incidence of most types of cancer in the immunosuppressed renal transplant recipient, this increased incidence is particularly dramatic in the case of tumours with a possible viral aetiology such as non-Hodgkin's lymphoma (100-fold increased incidence), squamous cell cancer of the skin (200-fold), cervical cancer of the uterus (50-fold), and Kaposi's sarcoma (1000-fold). In countries with a high exposure to sunlight such as Australia some 50 per cent of patients will have developed at least one skin cancer within 10 years of transplantation, with a reversal of the normal basal cell to squamous cell cancer ratio of 2:1. Many of the squamous cell cancers are rapidly growing with an increased propensity to metastasize (Fig. 11). They also occur in unusual sites, such as perianal.



Fig. 11. A squamous cell carcinoma of the hand in a renal transplant patient. The lesion reached this size in 6 weeks.

The prevalence of lymphomas is greater with current immunosuppressive protocols, this undoubtedly being related to the potency of the immunosuppression. Two types of lymphoma are seen. The first is an acute lymphoproliferative disorder, usually occurring within the first year of transplantation, which is usually associated with an increase in titre of Epstein–Barr virus. Its course may be rapidly fatal, but also it may remit with reduction or cessation of immunosuppressive therapy together with a course of high-dose acyclovir. Histologically the tumour is usually a polyclonal B-cell lymphoma. This acute lymphoproliferative disorder is more commonly seen in patients who have received ATG and/or OKT3 as part of their immunosuppressive protocol. The second type of lymphoma occurs at later times after transplantation and presents more typically with enlarged nodes. It may be a monoclonal B-cell lymphoma histologically and its response to conventional therapy is not dissimilar to that in a patient who has not received a transplant. Both types of lymphoma not uncommonly involve the base of the brain, which is an unusual site of lymphomas occurring in non-immunosuppressed patients.

Cardiovascular complications

As renal transplant recipients survive longer and become an increasingly more elderly population, ischaemic heart disease and cerebrovascular disease have become the major cause of death after transplantation. The major risk factors in the transplant population are hypertension and hyperlipidaemia. Hypertension requiring treatment after renal transplantation is common with current immunosuppressive protocols, and some 75 per cent of patients are on antihypertensive treatment at 1 year after transplantation. This incidence of hypertension has increased with the introduction of cyclosporin. Good control of hypertension is an essential part of the management of the transplant recipient, and often a combination of a diuretic, b-blocker, calcium channel inhibitor, and angiotensin-converting enzyme (**ACE**) inhibitors is required to achieve control.

Poorly controlled hypertension, especially if associated with deteriorating renal function, should raise the question of a renal artery stenosis in the transplanted kidney (Fig. 12). An angiogram will be necessary to establish the presence of a renal artery stenosis, but even if present this does not mean that the stenosis is a functional one. The cautious introduction of an ACE inhibitor can be used as a diagnostic test in this situation, for if renal function deteriorates and then returns to its previous level on cessation of the drug, this would strongly suggest that a radiological stenosis is a functional one. Renal vein renins from the transplanted kidneys and the host kidneys may be measured, but usually are not helpful.



Fig. 12. A stenosis of the renal artery distal to an end-to-side anastomosis of the artery with a cuff of aorta to the external iliac artery.

Once the diagnosis of a functional renal artery stenosis is made then correction of the stenosis is indicated. In the first instance an attempt to dilate the stenosis by balloon angioplasty should be made. If this is unsuccessful then reconstructive surgery should follow, although the dissection of the renal vessels is usually difficult and requires considerable expertise in both vascular and transplantation surgery. Many types of reconstruction are possible but the most common are reimplantation of the renal artery distal to the stenosis into the iliac artery or insertion of a saphenous vein graft between the common iliac artery and the renal artery distal to the stenosis.

Ischaemic heart disease is common in the elderly and diabetic renal transplant population, but is managed in the same way as in the non-transplant patient. There is no contraindication to major cardiac surgery in these patients if this represents the most appropriate management.

A significant increase in cholesterol levels is seen in many patients after renal transplantation, the rise being essentially in the very low-density lipoprotein fraction with a concomitant decrease in the level of high-density lipoprotein, a major pattern of risk for ischaemic heart disease. The hypercholesterolaemia has been attributed to cyclosporin but may be more closely related to the use of steroids. In any case these patients should be treated energetically with a low cholesterol diet and where high levels of cholesterol remain then one of the statins should be introduced. Similarly, if high triglyceride levels persist after dieting, more active treatment should be considered, as there is now evidence for a strong association between high triglyceride levels and chronic renal allograft failure.

Results of transplantation

Living related grafts

Apart from the rare instance in which transplantation between identical twins is possible, transplantation between HLA identical siblings provides not only the best short- and long-term outcome for graft survival (Fig. 1), but also as less immunosuppression is required there are fewer related problems. Living related donors and recipients sharing one HLA haplotype (parent to child or sibling transplant) also have a better graft survival than cadaveric grafts. It now appears that living related and unrelated donors and recipients mismatched for both HLA haplotypes also have a better graft survival than cadaveric grafts, presumably due to the lack of ischaemic injury to the kidney in the living donor (Fig. 1 and Fig. 13). Because of the better graft survival and also the shortage of cadaver kidneys, living related transplantation remains a justified procedure provided the donor has the appropriate motivation without any financial or emotional coercion. Similar arguments can be advanced for transplantation between emotionally related donors and recipients, such as spouses.

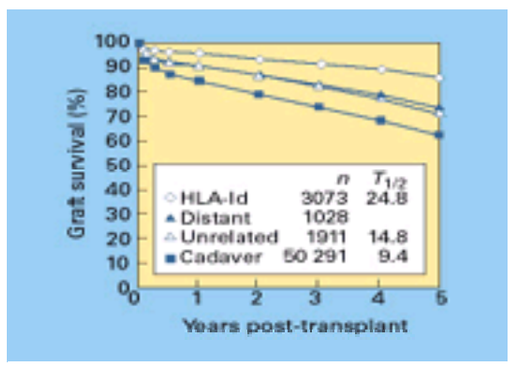


Fig. 13. The outcome of transplants between HLA identical siblings, distant relatives, living unrelated donors and recipients, and cadaver donors and recipients between 1991 and 1997 in the UNOS Scientific Renal Transplant Registry.

Cadaveric grafts

The bulk of renal transplants performed in the Western world are cadaveric (unlike in developing countries where the majority of transplants performed are still living related transplants). Graft survival has steadily improved over the last 20 years due to a variety of reasons so that around 80 to 90 per cent of grafts will be functioning at 1 year and around 65 to 70 per cent at 5 years (Fig. 3 and Fig. 13). Patient survival has not altered appreciably in recent years, being around 90 to 95 per cent at 1 year and 80 per cent at 5 years, and is unlikely to improve now that more elderly patients are being transplanted.

Factors influencing graft outcome

Immunosuppression

Undoubtedly immunosuppression is the major factor in improving graft survival in recent years, with the introduction of cyclosporin in the early 1980s resulting in a 10 to 15 per cent improvement in graft survival at 1 year. The more recent introduction of new immunosuppressive therapies, such as tacrolimus, mycophenolate mofetil, and anti-IL-2R antibodies, has not led to improved 1-year graft survival, although reducing the incidence of acute rejection. Also of note is that half-lives (i.e. the time by which 50 per cent of grafts functioning at 1 year after transplantation have failed) remain remarkably similar in the cyclosporin era compared with the azathiaprine era! In other words, cyclosporin or the newer agents decrease the loss of kidneys from acute irreversible rejection but have no influence on chronic allograft failure.

HLA matching

This still has a definite impact on cadaveric graft outcome in the cyclosporin era, while reliably predicting the outcome of living related transplantation (Fig. 1 and Fig. 2).

Centre effect

The centre in which the transplant is performed remains a major determinant of graft outcome and is still an unexplained phenomenon. It is not obviously associated with the experience of the unit, patient selection, or immunosuppressive protocols but no doubt reflects the influence of all of these and other unidentified factors.

Blood transfusions

That prior blood transfusions in the non-transfused recipient improved graft survival in the precyclosporin era is unquestioned. But in the cyclosporin era it is uncertain whether the transfusion effect still exists, and many units are abandoning their deliberate transfusion policies in non-transfused recipients because of this and the small, but real, risk of transmitting HIV or non-A, non-B hepatitis (hepatitis C) by a blood transfusion. However, an important prospective trial carried out by the European Collaborative Study group suggested that a transfusion effect was still demonstable in the cyclosporin era.

Age

The age of the recipient will influence graft outcome in that the elderly patient is more likely to have or to develop significant ischaemic heart disease and as a result the elderly recipient (over the age of 60) has a lower life expectancy than the younger recipient (Fig. 14). However, the elderly recipient is less likely to lose a graft from irreversible rejection and hence the poorer graft survival in this age group is due entirely to poorer patient survival (Fig. 15). The age of the donor also has an impact on cadaveric graft survival in that kidneys from young (less than 10 years of age) or very old (over 65 years of age) donors show a poorer graft survival.

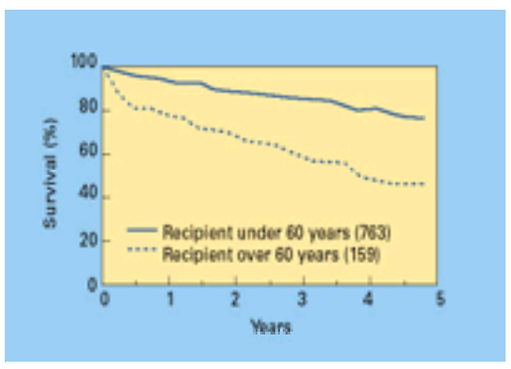


Fig. 14. Actuarial patient survival after renal transplantation in patients in Oxford under or over the age of 60 at the time of transplantation from 1985. All patients were immunosuppressed with triple therapy (cyclosporin, azathioprine, and prednisolone).

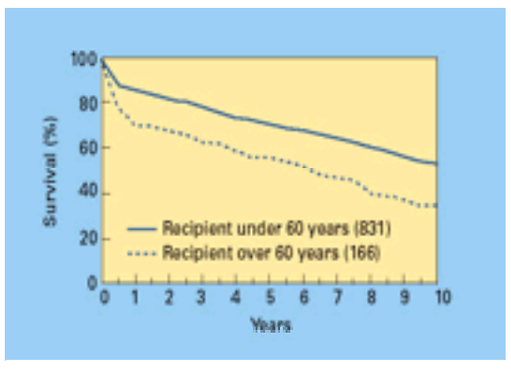


Fig. 15. Actuarial graft survival after renal transplantation in patients in Oxford under or over the age of 60 at the time of transplantation from 1988. All patients were immunosuppressed with triple therapy (cyclosporin, azathioprine, and prednisolone).

Race

The influence of the race of the donor or of the recipient remains uncertain. That a black patient in the United States has a poorer graft survival is undisputed, but distinguishing the possible genetic effect from the socio-economic effect of a less privileged population has proved difficult. Nevertheless evidence has been provided that, after allowing for the socio-economic influence, the black recipient does display a poorer graft survival than his white counterpart. Interestingly, Asian recipients appear to have superior graft survival compared with white and black recipients (Fig. 16).

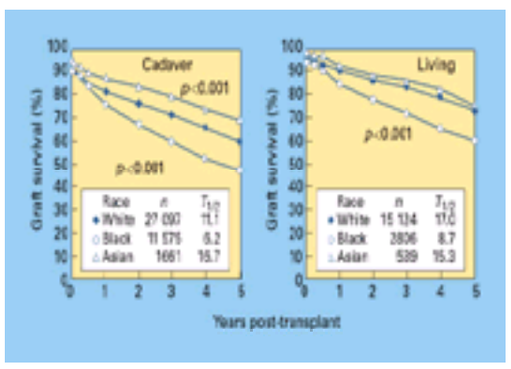


Fig. 16. The outcome of cadaveric and living transplants in Asian, White, and Black recipients in the UNOS Scientific Renal Transplant Registry.

Non-compliance

Cessation of immunosuppressive drug therapy by the patient, either because of unacceptable side-effects or for financial reasons, is now one of the more frequent causes of graft failure and is undoubtedly much more common than realized.

Recurrence of the original disease

Although certain of the glomerulonephritides do recur (see earlier), loss of the graft as a result is relatively uncommon (Table 2).

Other factors

Other factors such as the sex of the donor and recipient, blood group of donor and recipient, and parity of the recipient have little if any effect on graft outcome.

Rehabilitation

The patient with a successful transplant is in general restored to a normal existence with a quality of life as perceived by the patient being little different to that of the normal population. The majority of recipients of a successful graft will return to work, study, or managing the home within 6 months of transplantation.

Pregnancy after transplantation is possible whether the father or the mother is the transplant recipient. Although there is a higher level of spontaneous abortion in the transplant recipient, the incidence of developmental abnormalities in the live births that go to term is not higher. Renal function may deteriorate during pregnancy where pre-eclampsia is more common, and rejection may occur after delivery. Provided that renal function is relatively normal there is no contraindication to pregnancy but careful monitoring of the fetus and renal function in the mother during and immediately after the pregnancy is essential.

Future developments

As graft survival continues to improve with the introduction of more potent immunosuppression more attention will be paid to the side-effects of the immunosuppressive therapy and also to the prevention of cardiovascular disease. In addition, better and more specific immunosuppression may allow chronic rejection to be prevented, for this still remains a significant influence on long-term graft outcome.

Further reading

Bumgardner GL, Matas AJ. Transplantation and pregnancy. *Transplantation Reviews* 1992; **6**: 139–62. [An extensive review of pregnancy after transplantation.]

Bunce M *et al.* Phototyping: comprehensive DNA typing for HLA-A, -B, -C, DRB1, DRB3, DRB4, DRB5 and DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR–SSP). *Tissue Antigens* 1995; **46**: 355–67. [The new technology of tissue typing.]

Cancer report. In: Disney APS, Russ GR, Walker R, Sheil AGR, eds, *Australia and New Zealand Dialysis and Transplantation Registry Report 1997*, pp. 138–46. [The latest report of this outstanding Transplantation Cancer Registry.]

Cecka JM. The UNOS Scientific Renal Transplant Registry. In: Cecka JM, Terasaki PI, eds, *Clinical transplants 1998*, pp. 1–16. UCLI Tissue Typing Laboratory, 1999. [The latest UNOS report.]

Fishman JA, Rubin RH. Infection in organ-transplant recipients. *New England Journal of Medicine* 1998; **338**: 1741–51. [An incisive review of infection after transplantation.]

Ginns LC, Cosimi AB, Morris PJ, eds. *Transplantation*. Blackwell Science, Boston, 1999. [A textbook of organ transplantation.]

Kootstra G. The asystolic, or non-heartbeating, donor. *Transplantation* 1997; **63**: 917–21. [A key article showing that satisfactory kidneys can be obtained from asystolic donors.]

Land W, Messmer K. The impact of ischemia/reperfusion injury on specific and non-specific early and late chronic events after organ transplantation. *Transplantation Reviews* 1996; **10**: 108–27, 236–53. [A comprehensive review of ischaemia/reperfusion injury after transplantation.]

Mathew TH. Recurrent disease after renal transplantation. *Transplantation Reviews* 1991; **5**: 31–45. [A definitive review.]

Morris PJ, ed. *Kidney transplantation: principles and practice*, 4th edn. WB Saunders, Philadelphia, 1993.

Morris PJ, ed. *Kidney transplantation: principles and practice*, 5th edn. WB Saunders, Philadelphia, 2000. [A comprehensive textbook covering all aspects of kidney transplantation.]

Morris PJ, Johnson RJ, Fuggle S, Belger MA, Briggs D. Factors influencing outcome of primary cadaveric renal transplantation in the UK. *Lancet* 1999; **354**: 1147–52. [HLA remains an important factor in determining cadaveric graft outcome.]

Murray JE. Human organ transplantation: background and consequences. *Science* 1992; **256**: 1411–16. [Reflections by a Nobel Laureate on the early days of transplantation.]

Opelz G, Henderson R. Incidence of non-Hodgkin lymphoma in kidney and heart transplant recipients. *Lancet* 1993; **342**: 1514. [An important article documenting the association between non-Hodgkin's lymphoma and immunosuppression.]

Opelz G *et al.* Prospective evaluation of blood transfusions in cadaver kidney recipients. *Transplantation* 1997; **63**: 1155–60. [The transfusion effect still appears to be apparent in the cyclosporin era.]

Penn I. Cancers following cyclosporin therapy. *Transplantation* 1987; **43**: 32–5. [No apparent increase in cancer after transplantation with the introduction of cyclosporin.]

Pereira BJG, Wright TL, Schmid CH, Levey AS. A controlled study of hepatitis C transmission by organ transplantation. *Lancet* 1995; **345**: 484–7. [An important paper concerning the role of hepatitis C in transplantation.]

Racusen LC, Solez K, Burdick JF. *Kidney transplant rejection: diagnosis and treatment*, 3rd edn. Marcel Dekker, New York, 1998. [A textbook dealing with all aspects of rejection and its management.]

Ratner LE *et al.* Laparoscopic assisted live donor nephrectomy—a comparison with the open approach. *Transplantation* 1997; **63**: 229–33. [The first major paper describing this new but as yet unproved technique.]

Solez K, Racusen LC, Billingham ME, eds. *Solid organ transplant rejection. Mechanisms, pathology and diagnosis*. Marcel Dekker, New York, 1996. [A comprehensive textbook about rejection and its histopathological features.]

Terasaki PI, Cecka JM, Gjerson DW, Takemoto S. High survival rates of kidney transplants from spousal and living unrelated donors. *New England Journal of Medicine* 1999; **333**: 333–36. [A seminal paper describing the better outcome of living unrelated transplants in comparison with cadaver transplants.]

Ting A, Morris PJ. Matching for B-cell antigens of the HLA-DR (D-related) series in cadaver renal transplantation. *Lancet* 1978; **i**: 575. [The first paper describing the important effect of matching for HLA-DR alone in cadaveric graft outcome.]

16.6 Liver transplantation

Sandy Feng, Dicken S. C. Ko, and A. Benedict Cosimi

Introduction
History of liver transplantation
Indications for liver transplantation
General indications
Special considerations
Timing of liver transplantation
Donor selection and the procurement operation
Donor selection
Allograft removal and preservation
The recipient operation
Anesthetic considerations
Recipient hepatectomy
Venovenous bypass and ‘piggy back' orthotopic liver transplantation
Orthotopic allograft revascularization
Reduced-size, split liver, and living-donor liver transplantation
Auxiliary liver transplantation
Post-transplantation management and complications
Non-immunologic considerations
Hepatobiliary system
Pulmonary system
Cardiovascular system
Renal and metabolic systems
Hematologic system
Nutrition
Neurologic and psychiatric complications
General surgical care
Immunosuppression and rejection
Results and rehabilitation
Conclusions and the future of liver transplantation
Further reading

Introduction

The liver has a remarkable capacity to regenerate following even extensive necrosis. In some instances, however, progressive cirrhosis develops following injury induced by a variety of conditions, ranging from congenital biliary atresia to viral infections to alcohol abuse. In such patients, fatal complications of portal hypertension, progressive hepatic insufficiency, and infection can be reliably anticipated. Treatment with diet, medications, immunosuppression, or even a few palliative operations is typically of short-term benefit at best. The only life-saving option for such patients, and also for some with acute fulminant hepatic failure, is replacement of the diseased liver with a healthy organ. Nevertheless, as recently as 1980, clinical liver transplantation still remained an investigational procedure which was being performed regularly in only two centers worldwide. Over the next 10 years, the practice of hepatology was completely transformed as hepatic replacement became accepted as not only a hypothetical possibility, but as the preferred therapeutic option that could be practically offered to a significant proportion of the thousands of patients who die annually from irreversible liver failure. As a result, decisions regarding the care of patients with liver disease should now be made with the perspective that future liver replacement may be required. Operative procedures in the portal area, such as portacaval shunts or complex biliary drainage maneuvers, should generally be avoided since they complicate subsequent liver transplantation. Similarly, holding the transplant in abeyance while awaiting the absolute terminal stages of hepatic failure significantly compromises the recipient's chances of survival and is no longer justified for suitable candidates with relentlessly progressive liver dysfunction.

This dramatic change in the recommended approach to the treatment of hepatic insufficiency is based upon the improvement in long-term survival and rehabilitation which has been achieved following transplantation. In marked contrast to the limited rehabilitation provided by other forms of intervention, over 80 per cent of patients who survive transplantation return to full-time employment, schooling, or homemaking. As described below, better definition of the most appropriate indications and timing for liver transplantation, improved methods of donor organ preservation, and refinements of surgical techniques and perioperative management are among the factors which have contributed to reducing morbidity and mortality. The availability of more selective, and therefore less toxic, immunosuppressive protocols, however, has been the most important development in the recent rapid growth of liver transplantation.

The nearly simultaneous introduction into clinical trials of cyclosporine and monoclonal antibodies at the beginning of the 1980s ushered in a new era of solid organ transplantation. Markedly improved results were immediately evident in patients following renal transplantation and even more dramatically following liver transplantation, where a doubling of 1-year survival was seen (Table 1). Stimulated by this encouraging change in clinical outcome, a Consensus Development Conference on Liver Transplantation was convened in the United States at the National Institutes of Health in 1983. Review of the recent results led the participants to conclude that liver transplantation was no longer an experimental procedure but a practical therapeutic approach. This determination encouraged a much wider application of the procedure and, over the next 8 years, more than 150 centers were established worldwide to perform liver transplants for a continually broadening list of indications.

		Survival	
Era	*Etiology of liver failure*	*1 year (%)*	*5 year (%)*
1963–1981	Benign	25–35	5–10
	Malignant	20–30	0–5
1982–1990	Benign	70–75	50–60
	Malignant	55–70	20–25
1991–1994	Benign	75–85	70–75
	Malignant	65–75	45–60
1995–1998	Benign	85–90	n/a
	Malignant	65–90	n/a

Table 1 Recipient survival following liver transplantation

History of liver transplantation

The first reference to hepatic replacement in the scientific literature was by Welch in 1955. Initial experimental efforts were directed towards the transplantation of an extra liver into an ectopic site in the abdomen, typically with systemic venous inflow to the portal system. There are several theoretical advantages of such an auxiliary liver. The inherent risks related to the typically difficult host hepatectomy in the presence of severe portal hypertension and the instability that may develop during the anhepatic phase are avoided. In addition, any residual function of the retained native liver potentially provides a temporary supportive role during postoperative periods of compromised allograft function.

The early canine auxiliary allografts were rapidly rejected, making observation beyond a few days impossible. The more prolonged survival with the introduction of azathioprine immunosuppression revealed a rapid diminution of hepatic allograft mass beginning within 2 weeks after heterotopic implantation. The model thus led to the definition of the importance of hepatotrophic substances including insulin in the splanchnic venous blood. Revascularization techniques that provided portal flow directly from the recipient's alimentary venous return into the transplanted liver proved essential for maintaining the long-term anatomic and functional integrity of the allograft. Perfection of these techniques led to the first clinical trial of auxiliary liver transplantation in November 1964. That attempt, and approximately 10 others over the next 4 years, all met with a fatal outcome, usually as a result of sepsis and hepatic failure. The first successful auxiliary liver transplant was performed in April 1969

at Memorial Hospital in New York. The procedure was performed in a 72-year-old patient with non-resectable cholangiocarcinoma. The allograft functioned normally for 9 months until the patient's death, secondary to infection in her own obstructed liver. The first heterotopic liver allograft to provide unquestionably prolonged survival was performed in 1972 by the same group for treatment of biliary atresia. That patient survived for more than 17 years. Nevertheless, overall clinical results following auxiliary liver transplantation have been poor: only 2 of 50 reported patients who had undergone the procedure by 1986 survived more than 1 year.

The surgical techniques for orthotopic liver transplantation were first studied in canine experiments in 1956. Over the next 7 years, separate teams led by Moore in Boston and by Starzl, first at Chicago and then at Denver, independently identified the exacting technical requirements of the operation. Based upon these extensive studies, the first clinical attempt at liver replacement was undertaken at the University of Colorado in March 1963. Despite the experience of the surgical team, the procedure could not be completed due to massive hemorrhage. Over the 4 years, isolated efforts to accomplish this formidable procedure at several institutions worldwide yielded no long-term successes. Hepatic failure leading to sepsis resulted from ischemic damage and rejection in most of these recipients. It became clear that the successful clinical application of the technical skills already available would have to await the development of more effective immunosuppressive modalities. The addition of antilymphocyte serum to clinical protocols in 1966 helped to overcome this obstacle. With the use of triple-drug therapy including azathioprine, steroids, and antilymphocyte serum, improved results were first reported in renal allograft recipients. Soon afterwards, the first human hepatic allograft recipient to achieve prolonged survival was given a new liver in Denver in July, 1967. This young patient, who underwent the procedure for treatment of an extensive hepatoma, lived with normal allograft function for more than a year before dying with recurrent cancer. Despite such occasional successes, however, frequent postoperative complications continued to plague the procedure and resulted in an early death rate of about 70 per cent. These results remained in stark contrast to the more satisfactory outcome which was being achieved following kidney transplantation. Most centers concluded that the technical complexity of the liver transplant operation would continue to limit its clinical relevance until more highly selective immunosuppressive agents, which would not render the host so vulnerable to infection, became available. Over the next 15 years, therefore, there was a virtual moratorium on the clinical application of liver transplantation except for continued efforts by the two teams in Colorado and Cambridge, England. As noted previously, more specific immunosuppressive agents, including cyclosporine and monoclonal antibodies, were eventually developed and after two decades of disappointments and tragic failures, these finally provided the effective immunosuppression that had been needed to demonstrate the clinical feasibility of hepatic replacement therapy. Further improvements in the management of rejection, such as the introduction of tacrolimus and, more recently, mycophenolate mofetil, promise to continue to reduce the morbidity faced by liver allograft recipients and to extend the indications for the procedure.

Indications for liver transplantation

General indications

Since there is no effective alternative treatment available, liver transplantation should be considered for virtually any patient with advanced hepatic failure and clinical manifestations of cirrhosis. Typical candidates can be divided into five broad groups, as outlined in Table 2. In infants and younger children, biliary atresia has been the most common indication. Initial management of these patients usually includes porticoenterostomy (Kasai procedure) during the neonatal period. Successful biliary diversion and stabilization can be accomplished in 50 to 75 per cent of patients, at least until some growth has occurred. If the Kasai procedure fails, most centers recommend immediate consideration for transplantation since multiple attempts to achieve adequate biliary drainage greatly increase the complications of subsequent hepatic replacement.

Chronic advanced cirrhosis
Primarily parenchymal disease:
Postnecrotic cirrhosis
Alcoholic cirrhosis
Cystic fibrosis
Autoimmune liver disease
Cryptogenic cirrhosis
Primarily cholestatic disease:
Biliary atresia
Primary biliary cirrhosis
Sclerosing cholangitis
Primarily vascular disease:
Budd–Chiari syndrome
Veno-occlusive disease
Fulminant hepatic failure
Viral hepatitis
Drug-induced:
Acetaminophen, halothanes, sulfonamides, phenytoin
Metabolic liver disease:
Wilson's disease, Reye's syndrome
Inborn errors of metabolism:
Glycogen storage disease
α_1-Antitrypsin deficiency
Wilson's disease
OTC deficiency
Primary hepatic malignancies
Hepatocellular carcinoma & cirrhosis
Hepatoblastoma
Hemangioendothelioma

Table 2 Indications for liver transplantation

In later childhood, most candidates for transplantation have inborn errors of metabolism, postnecrotic cirrhosis, and hepatic neoplasms that are otherwise unresectable. In children with benign conditions, a fall-off from the established growth curve is often the major indication to proceed with transplantation. To avoid permanent growth retardation, the optimal time for transplantation for children is often earlier in the evolution of deteriorating hepatic function than for adults.

A wide variety of liver-based congenital errors of metabolism can be successfully treated with liver replacement. Patients with an inborn metabolic defect such as hemophilia who require liver transplantation for other reasons (such as hepatitis) will, therefore, enjoy the additional benefit of complete correction of the underlying defect. Hepatic replacement in adults for conditions leading to endstage cirrhosis, including primary biliary or cryptogenic cirrhosis, sclerosing cholangitis, and autoimmune hepatitis, has been highly successful and is generally accepted as the appropriate therapeutic approach for endstage liver failure.

Selection of the candidates most likely to benefit from liver transplantation relies primarily upon exclusion of coexisting conditions which increase the risk of postoperative death to an unacceptable degree. Absolute contraindications (Table 3) include malignancy or uncontrolled extrahepatic infection, secondary malignancies, or life-limiting cardiac or pulmonary disease. Active drug or alcohol addiction is also considered a contraindication. Experience in patients with active addictions has emphasized that non-compliance with the post-transplant medical regimen almost inevitably leads to unmanageable complications.

Absolute
Uncontrolled infection outside the hepatobiliary system
Malignancy outside the hepatobiliary system
Malignancy metastatic to the liver (e.g. colon, breast)
Uncorrectable, life-limiting congenital anomalies
Uncorrectable, life-limiting cardiopulmonary disease
Active drug or alcohol abuse
Relative
Age over 70 years
Human immunodeficiency virus (HIV) infection
Advanced cardiopulmonary disease
Severe morbid obesity (body mass index > 35kg/m^2)

Table 3 Contraindications to liver transplantation

As surgeons have gained increasing experience with liver replacement, a number of factors which were previously felt to preclude successful transplantation are now no longer considered contraindications. These include, for example, age over 65 years (the oldest reported recipient of a liver was 76 years old at the time of transplantation) and portal vein thrombosis. Extensive clotting of the portal or even mesenteric veins, which previously made revascularization of the allograft impossible, can now be successfully managed through the use of vein grafts (see below). In patients with acute renal failure associated with hepatic decompensation (hepatorenal syndrome), renal function can be anticipated to improve rapidly following successful liver replacement. Patients with chronic renal dysfunction can be managed with simultaneous liver and kidney transplantation with only moderately increased risks. Currently, relative contraindications include age over 70, human immunodeficiency virus (**HIV**) infection, advanced cardiopulmonary disease, and severe morbid obesity (Table 3).

Special considerations

Hepatitis B

Liver transplantation for cirrhosis resulting from chronic hepatitis B virus (**HBV**) infection has been controversial since multiple studies published in the late 1980s and early 1990s reported high rates of graft reinfection resulting in significantly decreased graft and patient survival. Recurrence is nearly universal for patients with evidence of active viral replication (HBeAg and/or HBV DNA positive) at the time of transplantation and occurs in approximately 58 (+/– 7) per cent of patients who have neither index of viral replication. Notably, recurrence is considerably less frequent in patients with d-virus superinfection (32 +/– 5 per cent) and least in those with fulminant hepatitis B infection (17 +/– 7 per cent). In the past, graft reinfection most often resulted in a gradual progression of hepatic dysfunction towards cirrhosis over several years, but occasionally, rapid deterioration of liver function has been observed. Results of retransplantation have been dismal characterized by aggressive recurrence with extremely high patient mortality.

More recent developments in molecular diagnostics and antiviral therapeutics have guided the rational design and investigation of strategies to prevent recurrence. Multiple studies have shown that the strategy of passive immunization using polyclonal hepatitis B immunoglobulin (**HBIg**) is effective in preventing recurrent infection. Reported efficacy rates range from 50 to 90 per cent with variability most likely explained by differences in the dose of HBIg, the frequency and route of administration, and the duration of treatment. It is generally accepted that HBIg treatment needs to be prolonged (more than 6 months) and at doses to maintain serum antibody titers at a minimum of 100 IU/l. There is now evidence suggesting even larger doses may be more efficacious early after liver transplantation. Several reasons have been extended for HBIg failure. First, patients with markers of active viral replication have higher HBIg failure rates. It is postulated that these patients may not have received enough HBIg to neutralize their viral load. Second, immunosuppression may perturb the balance of antigen and antibody concentration and thereby favor escape from HBIg therapy by altering the dynamics of viral replication. It is well established that corticosteroids given in high doses in the immediate peritransplant period or to treat rejection increase HBV replication via a steroid-responsive element in the HBV genome. Finally, a correlation has been identified between the duration of HBIg therapy and the appearance of HBV bearing mutations in the surface antigen gene. It is postulated that HBIg can exert immune pressure leading to the induction and/or selection of mutant virus capable of escaping neutralization by HBIg.

In light of the significant rate of HBIg failure, more recent studies have focused on defining the role of antiviral agents in the prevention of hepatitis B recurrence after transplantation. Currently, the most widely used antiviral agent against hepatitis B is lamivudine, a cytosine nucleoside analog which is a potent inhibitor of viral replication. Although the current studies are non-randomized with small patient numbers, lamivudine alone has been used with success in both preventing hepatitis B recurrence and/or treating recurrence resulting from HBIg failure. Lamivudine failure has been intimately associated with mutations at a particular locus of the hepatitis B DNA polymerase gene. The mechanisms of actions of HBIg and lamivudine suggest that they may work synergistically to prevent hepatitis B reinfection. Inhibition of viral replication would decrease not only the likelihood of saturating the antigen-binding capability of HBIg but also the immune pressure to select for mutants. The humoral immunity provided by HBIg may limit viral spread and, indirectly, viral replication, thereby also decreasing the likelihood of the emergence of lamivudine-resistant virus. However, the theoretical advantage of combination therapy may be countered by the fact that the reading frames of the polymerase and envelop genes overlap. Lamivudine-induced polymerase mutations may alter the antigenicity of HBsAg while HBIg-induced envelope mutations may decrease the susceptibility of polymerase to lamivudine. Nevertheless, data from two non-randomized studies using combination HBIg and lamivudine with a total of 36 patients surviving past the immediate postoperative period indicate no recurrence after a median follow-up of approximately 1 year. While these impressive results need to be confirmed by prospective, randomized studies, there is a strong trend in the transplant community to use combination treatment for the prevention of hepatitis B recurrence after liver transplantation, particularly in the high-risk subset of patients with markers of active viral replication at the time of transplantation.

Hepatitis C

Hepatitis C is now the leading indication for liver transplantation in the United States, accounting for greater than one-third of all transplants performed. As with hepatitis B, the post-transplant course is complicated by recurrent disease. However, unlike hepatitis B, survival rates for patients transplanted for hepatitis C have been comparable with those transplanted for other indications. Thus, the initial impression was that recurrent disease, although not uncommon, was mild and indolent. More recently, it has become clear that viremia is nearly universal after liver transplantation. Clinical recurrence in terms of graft hepatitis develops in the majority of patients within the first year or two after transplantation and rapidly progressive graft dysfunction occurs in 5 to 10 per cent of patients. Although retransplantation and/or death, related directly to recurrent hepatitis C infection, occurs in only approximately 10 per cent of transplant recipients, patients with hepatitis C have been reported to have a significantly higher incidence of lethal infections. It is widely anticipated that long-term (longer than 10-year) survival rates for recipients with hepatitis C will be significantly less than for other indications leading to liver transplantation.

Much attention has been directed towards identifying parameters predictive of incidence and/or severity of hepatitis C virus (**HCV**) recurrence. Unfortunately, several technical issues have contributed to the conflicting data that have emerged from many studies worldwide. Use of non-standardized methods for determining virus genotype and viral titers render it difficult to compare the results of different studies. In addition, the pattern of HCV infection defined by the predominant genotype or average viremia level varies greatly among patients from North America, Europe, and Asia. Finally, studies do not use uniform endpoints to define recurrent disease. Although there are clearly three main criteria (biochemical, virologic, and histologic), most studies present data on only a subset of the three parameters. Nevertheless, consistent threads among studies suggest three primary prognostic indicators: pre- and/or post-transplant HCV titers, HCV genotype, and number of rejection episodes in conjunction with the degree of immunosuppression. In general, high pre- and/or post-transplant viral titers, genotype 1b, and increased episodes of rejection with concomitant intensified immuno-suppression all appear to have a negative impact on the incidence and/or severity of recurrent disease after liver transplantation.

Strategies for treating recurrent HCV infection have paralleled those for the treatment of chronic HCV infection. Several, small, non-randomized studies have shown that interferon-a (3 million units subcutaneously three times weekly) has modest efficacy at best in the treatment of recurrent disease. Biochemical responses (return of transaminases to normal levels) occurred in 9 to 28 per cent of patients; virologic (conversion to negative viral titer) and histologic responses were rare. Furthermore, one study suggested that interferon-a treatment may induce chronic rejection. Recently, the combination of interferon-a and ribavirin, a guanosine nucleoside analog, has been shown to be superior to interferon-a alone for the treatment of chronic HCV. This strategy is beginning to be applied to post-transplant recurrent disease. A pilot study involving 21 patients who received 6 months of combination therapy followed by an additional 6 months of ribavirin monotherapy has shown promising results. After 6 months, a 100 per cent biochemical response, 48 per cent virologic response, and 100 per cent histologic response were observed. At the end of a year, overall there was an 81 per cent biochemical response (94 per cent if only patients completing therapy are included), a 24 per cent virologic response, and an 81 per cent histologic response (again 94 per cent if only patients completing therapy are included). Less than 5 per cent of the patients were unable to complete the course of treatment secondary to anemia, which is a recognized side-effect of ribavirin. There were no episodes of rejection during the study period. Clearly, these results await validation by larger, prospective, randomized studies that are currently in progress.

Attention has also been directed towards efforts to prevent hepatitis C recurrence after liver transplantation. Since high viral titers are associated with decreased efficacy of interferon-a in both pre- and post-transplant patients with HCV and viral titers fall dramatically at the time of transplantation, two randomized controlled studies have addressed the use of interferon-a initiated in the peritransplant period to prevent recurrent disease. The larger study involving 71 patients reported that prophylactic interferon-a for 1 year significantly reduced the risk of recurrence by a factor of 0.4. The smaller study involving 24 patients reported that prophylactic interferon-a for 6 months delayed the occurrence but did not decrease the incidence or severity of recurrent disease. Finally, there are scattered efforts examining the role of reducing or altering immunosuppressive strategies in order to influence the development of recurrent disease. Although the mechanism is not as well elucidated as for hepatitis B, corticosteroids appear to enhance HCV replication and, when administered in high dosages, may be responsible for the rapid rise in viral titers observed shortly after transplantation. In contrast, mycophenolate mofetil, a more recently approved immunosuppressive agent, has been reported to have some antiviral effect and thus could prove to be more appropriate for these patients than azathioprine. Some of the strategies delineated above may be complementary and therefore have additive effects. Prevention and treatment of recurrent hepatitis C certainly represents one of the major challenges currently facing the transplant community.

Alcoholic (Laennec's) cirrhosis

Despite the high prevalence of cirrhosis secondary to alcohol abuse, many of these patients are unsuitable for transplantation. Often, they have multisystem disease and are prone to precipitous clinical deterioration. In addition, non-compliance with essential medical therapy after transplantation has been found to limit long-term success. Alcohol recidivism has been noted to be as high as 15 to 20 per cent, depending on the published series. As a result, hepatic transplantation for this indication requires extensive pyschiatric evaluation and addiction counseling to determine a patient's candidacy for the procedure. Nevertheless, the results of transplantation, combined with a multidisciplinary treatment program for substance abuse in individuals after discontinuing alcohol intake, are comparable with those for other accepted conditions. Thus, liver transplantation for patients with Laennec's cirrhosis is acceptable for carefully selected individuals.

Hepatocellular cancer

Hepatocellular cancer, as the fourth most common cancer in the world, causes substantial morbidity and mortality. Survival periods for patients with symptomatic disease are dismally short. Unfortunately, survival rates even for asymptomatic disease identified by screening or surveillance protocols are also poor. The majority of hepatocellular cancer both worldwide and in the United States is associated with non-malignant liver disease such as cirrhosis and/or viral hepatitis (both hepatitis B and C), which negatively impact on the prognosis of hepatocellular cancer in several ways. First, abnormal hepatic function and diminished hepatic reserve often contraindicate or compromise adequate surgical resection. Second, most chronic liver diseases including cirrhosis result in a ‘field defect' whereby the remaining liver is

at high risk for synchronous or metachronous tumors. Finally, continued deterioration of liver function after successful treatment of hepatocellular cancer carries additional morbidity and mortality. Therefore, liver transplantation, which would remove the malignancy as well as the entire vulnerable field and restore normal hepatic function, has been considered a therapeutic option for patients with non-metastatic hepatocellular cancer. In particular, those with chronic liver disease who often undergo routine surveillance for hepatocellular cancer may have the additional benefit of early diagnosis.

The initial experience with liver transplantation for the treatment of hepatocellular cancer in relatively unselected patients and without adjuvant chemotherapy was discouraging. The worldwide experience with 365 patients reported in 1991 revealed an overall 5-year survival rate of 18 per cent with the 2-year disease-free survival being only 9 per cent. Since then, multiple studies comparing surgical resection and liver transplantation have concluded that patients with small tumors derived the greatest benefit from transplantation in that they had significantly greater disease-free and/or overall survival. These studies emphasize that patients with either a single tumor of less than 5 cm or less than three tumors with each lesion less than 3 cm have survival results comparable with those of other transplant recipients without hepatocellular cancer and are, therefore, appropriate candidates for liver transplantation.

Although the subset of patients with hepatocellular cancer most likely to derive maximal benefit from liver transplantation has apparently been identified, many controversial questions still surround the use of liver replacement therapy to manage hepatocellular cancer. Some within the transplant community question whether the restrictive criterion of small tumors should actually be expanded. Others argue that defining acceptable long-term survival statistics for this subset of patients requires the balanced consideration of the severe organ shortage and the high mortality of non-transplant therapeutic alternatives. The studies mentioned were carried out when the waiting period for liver transplantation was relatively short—approximately 2 months in several of the studies. Moreover, the survival analysis was based upon treated patients rather than on an intention to treat basis. It is likely that current extended waiting times which exceed 1 to 2 years at many United States transplant centers would adversely impact these analyses. Thus, more recent discussions have raised the question of whether special consideration should be extended to patients with hepatocellular cancer over other patients on the transplant waiting list who might be less adversely affected by the prolonged waiting period.

Finally, the role of neoadjuvant therapy for hepatocellular cancer remains unclear. With regard to pretransplant strategies for local tumor control, techniques ranging from surgical resection to transarterial chemoembolization to percutaneous ethanol injection or radiofrequency ablation have been used either alone or in combination depending on the number, size, and location of the tumors, degree of hepatic dysfunction, and institutional preference. Although broadly used, none of these pretransplant therapies have been proved beneficial in preventing recurrence or prolonging survival. There is clearly a need for multicenter, prospective trials to define the optimal strategy for maintaining local tumor control while awaiting liver transplantation.

Retransplantation

As the number of patients undergoing hepatic replacement has increased, the need for hepatic retransplantation has become more common. The reported frequency of retransplantation has varied between 5 and 20 per cent among different centers. Overall survival rates following retransplantation are approximately 20 per cent lower than those following the primary procedure. Repeat liver transplantation is performed primarily in recipients whose initial graft is failing due to technical complications, primary graft non-function, or irreversible chronic rejection. Patients with primary graft non-function, secondary to unrecognized donor organ damage suffered during the agonal prerecovery period or to preservation injury, decompensate rapidly in the post-transplant period. As a result, they may be extremely unstable by the time a second donor is identified. The prognosis following retransplantation for this indication, therefore, is least favorable. Retransplantation for uncontrolled acute rejection is uncommon: more typically, one observes relentlessly progressive chronic rejection manifested by worsening cholestasis and a histopathologic picture of ‘vanishing bile ducts'. Retransplantation has generally been accepted as the only treatment available for these patients, although in the current era of tacrolimus immunosuppression, chronic liver rejection is becoming increasingly rare.

Timing of liver transplantation

The critical decision for all of these patients is the timing of liver transplantation. Until recently, the procedure was often considered so drastic that it was recommended only as a last resort when all other palliative measures had failed. Clearly, allowing these patients to deteriorate to the point of repeated gastrointestinal bleeding, malnutrition, sepsis, and coma compromises the chances of successful transplantation and is no longer acceptable. The role of the hepatologist in choosing the optimal time to refer the patient for transplantation is perhaps the most important determinant of survival following the procedure. Unfortunately, the natural history of these diseases proceeds at various rates in individual patients; the decision to proceed to transplantation in many cases cannot be simply based on some predetermined clinical or biochemical profile. A combination of criteria including the nature and stage of the disease as well as the patient's age, quality of life, and history of previous complications must all be considered. Appropriate candidates with primarily parenchymal conditions often present with hypoalbuminemia, coagulopathy, variceal bleeding, ascites, or hepatic encephalopathy, while serum bilirubin levels are not markedly abnormal. The management of variceal hemorrhage refractory to pharmocologic and endoscopic measures in potential transplant candidates depends upon the severity of liver disease. In general, surgical portal decompression should be cautiously considered for potential transplant recipients. However, a splenorenal or mesocaval shunt may be appropriate for patients with Child's class A cirrhosis. For patients with class B or C cirrhosis, a transjugular intrahepatic portosystemic shunt may be the only viable alternative for refractory hemorrhage secondary to portal hypertension. However, since long-term survival rates after portosystemic shunting are inferior to those achieved by transplantation, patients with advanced liver disease should generally be considered for liver replacement.

Severe hyperbilirubinemia with intractable itching and the disabling complications of hepatic osteodystrophy may be the major indications for liver replacement for patients with cholestatic conditions. In contrast to the patients with primarily parenchymal disease, transplantation may be clearly appropriate for these patients when synthetic function is relatively preserved and only limited evidence of portal hypertension is present.

Patients with fulminant hepatic failure secondary to viral-, toxin-, or drug-induced massive necrosis of a previously healthy liver represent a particularly difficult group to evaluate for transplantation. Although some of these patients recover, a mortality rate of more than 80 per cent can be anticipated for those who progress to stage III/IV encephalopathy. Hepatic replacement in these patients, in order to be successful, should be performed before stage IV encephalopathy and the attendant sequelae of advanced cerebral edema. Transplantation in patients at this grave stage is associated with a mortality rate of more than 50 per cent. Ominous earlier signs of persistent coagulopathy, hypoglycemia, renal failure, and progressive encephalopathy despite maximal medical and nutritional support are indications to refer the patient for urgent liver replacement. Transplantation prior to the development of grave indications could result in the replacement of an occasional liver with reversible damage, which would appear to be justified by a survival rate of 60 to 75 per cent compared with the 20 per cent survival rate of medical management alone.

Donor selection and the procurement operation

Donor selection

Since the liver appears to be shielded from the detrimental effects of many systemic diseases which may compromise the quality of other organs, there are few absolute or relative contraindications to liver donation (Table 4). Ideally, one prefers a donor less than 70 years of age, though older donors have been used with success. There should be no previous history of neoplasia, except for most brain tumors and non-melanoma skin cancer. Donor serologies for HIV, HBV, and HCV may present absolute or relative contraindications to organ utilization. In addition, the gross appearance of the liver at the time of procurement may necessitate a biopsy which can reveal abnormalities such as inflammation and/or fibrosis or macrosteatosis (Fig. 1), which increases the risk of primary graft non-function. The ABO blood group should, ideally, be identical or compatible with the recipient although ABO-incompatible grafts can be considered for younger children (less than 2 years of age) and for critically ill individuals. Reasonable matching for body size between donor and recipient is preferred to ensure adequate hepatic mass and to limit the technical difficulty of the reimplantation procedure. An organ from a donor whose weight is between 50 and 150 per cent of the recipient's is optimal. Since size constraints are particularly exacting for small individuals, particularly young children, and because of the rarity of pediatric donors, one must often consider using adult donors for transplantation into children. This usually necessitates reduction of the graft to a single lobe or less. This approach has been further extended to successful division of a single cadaver donor liver into two viable grafts for implantation into two recipients (discussed later).

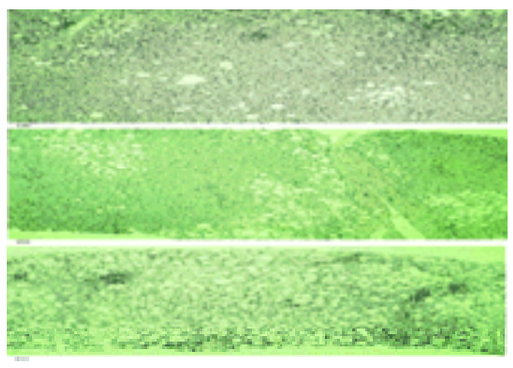


Fig. 1. Histology of macrosteatosis.

Absolute
Human immunodeficiency virus (HIV) infection
Malignancy (excluding most brain and non-melanoma skin cancers)
Hepatitis B surface antigen positive
Relative
Age over 75 years
Macrosteatosis (> 30 per cent)
Hepatitis B core antibody positive (and surface antibody negative)
Hepatitis C antibody positive
Abnormal liver function tests
Biopsy abnormalities (e.g. inflammation, fibrosis, necrosis)
High-risk social behavior

Table 4 Contraindications to liver donation

Histocompatibility matching is not performed for liver transplantation. Typically, the urgency of the recipient operation precludes donor selection based upon tissue matching. Furthermore, retrospective analysis has not shown any significant correlation between matching and outcome. Successful liver transplantation, in fact, has not infrequently been performed even in recipients with measurable serum levels of lymphocytotoxic antibodies reactive with donor histocompatibility antigens. Such a 'positive cross-match' would be anticipated to result in hyperacute rejection of kidney allografts; however, this phenomenon has seldom been documented following liver transplantation. Moreover, only a modestly compromised long-term survival rate has been observed in hepatic recipients of such donor organs.

Allograft removal and preservation

The success of the multiple organ recovery involved during donor hepatectomy requires careful co-ordination among the teams to assure that there is no compromise in organ viability. In addition, it is critical to have anesthesia support to monitor and maintain the cardiovascular integrity of the donor during the meticulous dissection, which may take up to 3 h.

Although the details may differ depending upon the combination of organs to be removed, certain common principles prevail. These include wide exposure, dissection of each organ vasculature while the heart is still beating, placement of cannulas for *in situ* cooling and preservation, and removal of the organs while perfusion continues, usually in the order of heart and/or lungs, liver, kidneys, and pancreas. The organs are exposed through a midline incision extending from the suprasternal notch to the pubis (Fig. 2(a)). Rapid inspection excludes unsuspected sepsis, neoplasia, or other significant pathology, and confirms the gross suitability of the organs to be procured. If the heart is to be used, it is mobilized first so that it can be removed quickly should uncorrectable vascular instability occur during the dissection of other organs. Most abdominal teams employ a rapid perfusion strategy for the infradiaphragmatic organs by isolating the infrarenal aorta and vena cava at their respective bifurcations. The control of these great vessels allows for rapid cannulation and perfusion of all abdominal organs if the donor becomes hemodynamically unstable. If rapid removal is necessary to salvage the organs, the liver, pancreas, and/or kidneys can be excised *en bloc* and separated *ex vivo* on the back table.

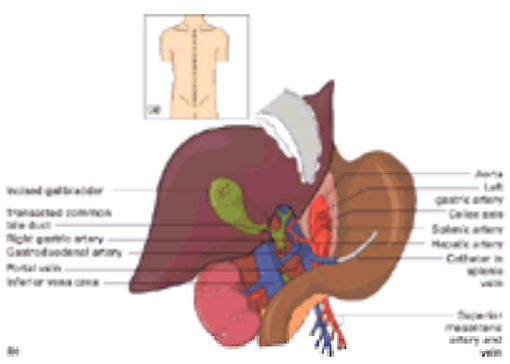


Fig. 2. Diagrammatic depiction of donor hepatectomy. (a) Exposure for multiple organ retrieval is through a long midline incision. (b) The portal area structures have been isolated and a perfusion catheter inserted into the portal vein via the transected splenic vein. For clarity, the pancreas, which usually would be retracted inferiorly, has been omitted.

If *in situ* dissection is to be performed, mobilization of the liver and pancreas begins in the porta hepatis. The common bile duct is isolated and then transected. The gallbladder is incised and flushed to prevent autolysis of the biliary epithelium. The proper hepatic, gastroduodenal, and splenic arteries are identified, thus enabling dissection of the celiac axis to the aorta. If the pancreas will not be used, the neck is divided, exposing the confluence of the splenic and superior mesenteric veins for the placement of a portal perfusion cannula (Fig. 2(b)). Alternatively, a cannula can be introduced into the inferior mesenteric vein and guided up to the portal vein. In these cases, the splenic and superior mesenteric arteries may be ligated (unless a replaced right hepatic artery is present) to increase hepatic perfusion. Mobilization of the liver can be completed by isolating the vena cava above the renal veins. The entire length of the celiac axis, including a Carrel patch of donor aorta, should be taken when the liver is removed.

If the pancreas is to be transplanted, the short gastric vessels must be divided and the spleen mobilized (further discussed in Chapter 16.9). Using the spleen as a handle, the tail and body of the pancreas are carefully dissected out of the retroperitoneum. The splenic and superior mesenteric arteries must be identified and preserved to ensure integrity of blood supply to the pancreas. The infrapancreatic confluence of mesenteric veins can be ligated and divided with a vascular stapling device. Because of the indigenous bacterial and fungal flora of the upper intestine, it is usual to instill antimicrobial solutions such as a combination of betadine and amphotericin B through the nasogastric tube to decontaminate the luminal contents. After irrigation, the second part of the duodenum is divided with a stapling device, completing the pancreatic dissection.

The kidneys are the last organs to be mobilized. Mobilization should be kept to a minimum to decrease the donor operative time; the procuring team simply needs to ensure that the kidneys do not have any masses. An incision into Gerota's fascia will suffice. The ureters, however, must be identified so that they are not injured during the procurement. The donor is then heparinized after which a perfusion cannula is placed in the aorta and a drainage cannula into the infrarenal vena cava. Liver precooling with 1 to 2 liters of crystalloid solution at room temperature via the portal cannula is employed at some centers to decrease the severe temperature shock to the liver with the cold (4°C) University of Wisconsin (UW) solution. In co-ordination with the cardiothoracic team, if present, the supraceliac aorta is then cross-clamped and *in situ* aortic and portal vein flushing is begun using UW solution (Table 5) simultaneous to cardioplegia infusion into the ascending aorta. After adequate flushing, cardiectomy is performed. The liver is removed next. The kidneys and ureters are then removed *en bloc* with the aorta and inferior vena cava and placed in cooled perfusion solution. Donor iliac artery and vein are also removed for possible use as vessel grafts during hepatic reimplantation. In donors from whom whole pancreaticoduodenal procurement is planned, we advise removing this organ block last in order to avoid possible contamination from the transected duodenum.

Constituent	Concentration	Function
K^+ lactobionate	100mmol	Impermeant
NaH_2PO_4	25mmol	Buffer
Adenosine	5mmol	ATP stimulator
$MgSO_4$	5mmol	Membrane stabilizer
Glutathione	3mmol	Membrane stabilizer, antioxidant
Raffinose	30mmol	Impermeant
Insulin	100U	Metabolic enhancer
Trimethoprim/sulfamethoxazole	8mg/40mg	Antibacterial
Modified hydroxyethyl starch	50g	Impermeant
Final concentration:		
Na^+	30mmol/l	
K^+	120mmol/l	
pH	7.4	
Osmolarity	320–330mmol/l	

Table 5 Composition of UW preservation solution

After completion of the donor hepatectomy, the allograft is further flushed through the portal vein and hepatic artery and then immersed in a sterile plastic bag containing UW preservation solution for cold storage until transplantation. The anion (lactobionate) and the trisaccharide (raffinose) contained in the UW solution appear to be particularly effective impermeants for suppression of hypothermia-induced cellular swelling. The introduction of this preservation solution into clinical practice has extended the safe period of cold storage to as long as 24 h. This relatively prolonged preservation time increases the capability for more distant procurement and also relaxes the logistic constraints of the transplantation procedure. The recipient operation can now be scheduled on a semi-emergency basis, thus allowing more flexible use of operating rooms and personnel. Perhaps most importantly, the enhanced margin of safety has made possible the development of new techniques, such as reduced-size and split liver transplantation.

The recipient operation

Anesthetic considerations

One of the most important determinants of perioperative mortality in these patients is the metabolic and resuscitative management by the anesthesia team. The complexity of the procedure requires the establishment of large-bore intravenous access lines and the provision of extensive intraoperative hemodynamic monitoring. In addition to the usual anesthesia equipment, the operating room should have the following devices available: a blood salvage device (Auto-transfuser®), separate pump systems for blood warming and rapid infusion (Level I Rapid Transfuser®), a venovenous bypass apparatus (Medtronic®), a patient warming blanket (BareHugger®), and an argon beam coagulator. Access for infusion and monitoring is typically accomplished via an arterial line and large-bore peripheral and central intravenous lines in the upper extremities, neck, and/or chest. Venous access sites in the legs are avoided since infusion would be unreliable during the period of vena cava occlusion. The use of pulmonary artery catheters is dictated by individual patient indications.

Patients with hepatic failure are predisposed to developing hypoxia from atelectasis due to ascites and from intrapulmonary arteriovenous shunts. Careful preoxygenation and rapid sequence induction anesthesia are commonly used. An appropriately sized, low-pressure cuff endotracheal tube is placed in the event of prolonged postoperative intubation. Adequate padding of the heels, sacrum, elbows, and head is provided to prevent pressure ulceration during the operation. Hypothermia must be anticipated from the prolonged duration of the procedure, the massive volume resuscitation, the extracorporeal blood circulation if venovenous bypass is employed, and the implantation of the cold allograft. Minimizing the time of the surgery, warming blood and other intravenous infusions, flushing the donor liver with room-temperature crystalloid solution or recipient blood prior to reperfusion, covering the head and exposed extremities, and warming with the BareHugger® blanket are essential to limit recipient hypothermia.

The selection of anesthetic agents is governed by the high cardiac output, low peripheral vascular resistance state typically associated with endstage liver disease. Isoflurane inhalational anesthesia will reduce systemic resistance and may be used as tolerated. The preferred maintenance agent for hemodynamic stability is a narcotic anesthetic employing fentanyl and/or morphine. Because some patients may not tolerate administration of routine anesthesia, they are often given a benzodiazepine as well to promote an amnestic response. Lorazepam is frequently chosen since this agent requires only glucuronidation for excretion and thus does not require metabolism by the allograft during the early postimplantation period when hepatic dysfunction may occur.

Some of the intraoperative consequences of the transplant procedure are detailed in Table 6. The patients often develop metabolic acidosis which should not be overcorrected with bicarbonate since they will also receive a large citrate load with administered blood products. Citrate is eventually metabolized to bicarbonate and thus may lead to significant postoperative metabolic alkalosis. Although the need for packed red blood cells and fresh frozen plasma replacement are determined independently, patients typically require approximately equal volumes of each. Platelets are administered based upon pre- and intraoperative platelet counts which are corrected as needed for satisfactory hemostasis.

Prolonged duration of surgery
Major blood loss
Problems related to the anhepatic phase
Lower body volume pooling
Hypotension (without venovenous bypass)
Oliguria
Reperfusion problems
Hyperkalemia
Hypercarbia
Acidosis
Hypoxemia
Coagulopathy
Hypothermia

Table 6 Intraoperative consequences of orthotopic liver transplantation

A critical period for the anesthesia team occurs at the time of initial liver reperfusion. Despite prior flushing of the allograft, the sudden bolus of cold, hyperkalemic, and acidotic blood from the liver, splanchnic circulation, and lower extremities may cause severe pulmonary artery vasoconstriction with resultant systemic hypotension. It is therefore essential to stabilize the patient by correction of acid–base abnormalities and hypovolemia prior to reperfusion of the allograft. Additional administration of calcium chloride just prior to vena caval unclamping is typically employed to antagonize hyperkalemia during reperfusion. As allograft function improves following rewarming and arterial reconstruction, a significant decrease in serum potassium often occurs as a result of both uptake into the revascularized hepatocytes and an increased availability of ionized calcium from hepatic citrate metabolism. Potassium supplementation may then be required to correct the observed hypokalemia. A significant bleeding diathesis may also develop at this stage secondary to the accelerated fibrinolysis which occurs in association with increased blood levels of fibrin degradation products. This consequence of hepatic ischemia and reperfusion may add to the blood loss observed after allograft implantation. The importance of aggressive replacement of coagulation factors during the early reperfusion period, sometimes including antifibrinolytic agents, cannot be overemphasized.

Recipient hepatectomy

Removal of the diseased liver can be a tedious process because of the coagulopathy and portal hypertension inherent in patients with endstage liver disease which may be further exacerbated by highly vascular adhesions from previous surgical procedures. Abdominal exploration in adults is carried out through a bilateral subcostal incision with a midline extension, usually including excision of the xiphoid to allow better exposure of the suprahepatic vena cava. In children, this midline extension is seldom required. Meticulous hemostasis is essential and can be obtained using a combination of electrocautery, suture ligation, and the argon beam coagulator.

After entering the peritoneal cavity, a brief visual and manual exploration is performed. The falciform and left triangular ligaments are divided, sometimes requiring stepwise silk ligatures for hemostasis. The gastrohepatic ligament (lesser omentum) is similarly divided. The right triangular ligament is divided and the bare area of the liver is separated from the diaphragm. In patients with extensive adhesions, it may be necessary to devascularize the liver prior to extensive mobilization. Ideally, the liver is left vascularized until the final stages of hepatectomy to retain baseline levels of clotting factors and glucose metabolism. Rarely, institution of venovenous bypass may be required at this point to help stabilize the patient. Most teams, however, do not find the option of early bypass and rapid hepatectomy attractive as an

intraoperative management option.

Once the hepatic mobilization is completed, the portal structures are isolated. To maximize length for the subsequent biliary anastomosis, the recipient common bile duct is divided above the entrance of the cystic duct. If the patient has had a previous biliary enteric anastomosis, the anastomosis is taken down, protecting the portal area and the Roux-en-Y jejunal loop that will be subsequently used for drainage of the donor liver. The left and right hepatic arteries or the proper hepatic artery are ligated. Often the more proximal common hepatic artery is further exposed for anastomosis leaving the gastroduodenal artery intact (Fig. 3). Finally, the portal vein is skeletonized by ligating and dividing tributaries including the coronary vein and/or pancreatic veins. If the portal vein is found to be thrombosed, portal vein thrombectomy is attempted. If unsuccessful, the more proximal portal vein, the splenic, or the superior mesenteric vein must be explored for a suitable anastomotic site and an iliac vein graft may be necessary (Fig. 4).

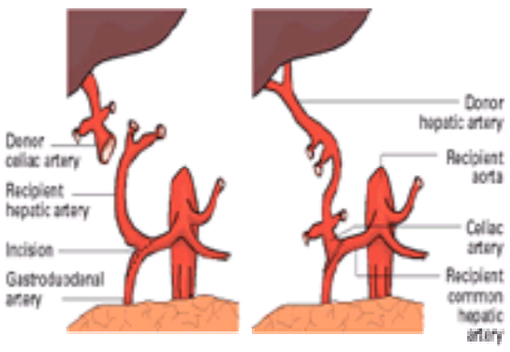


Fig. 3. Isolation of recipient celiac axis for revascularization of liver allograft. (a) The recipient proper hepatic artery has been ligated distal to its bifurcation. (b) The subsequent arterial anastomosis is between donor celiac axis (CA) or Carrel patch of donor aorta to the side of the recipient common hepatic artery at or proximal to the origin of the gastroduodenal artery.

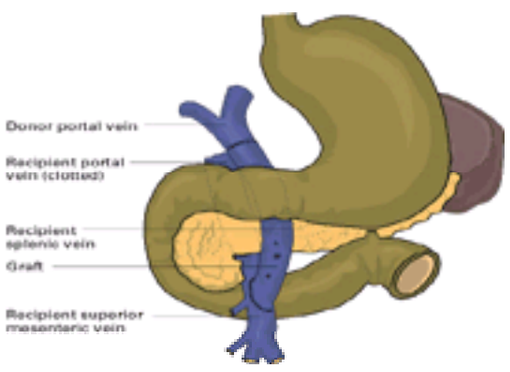


Fig. 4. Use of donor iliac vein to bypass thrombosed recipient portal cannula. In patients with extensive thrombosis, the vein graft may have to be placed anterior to the pancreas to extend to a patent superior mesenteric vein. More commonly a suitable anastomotic site is found behind the pancreas at the confluence of the splenic and superior mesenteric veins.

Once the portal dissection has been completed, the infrahepatic vena cava can be exposed. Because of extensive retroperitoneal collateral vessels, the tissues must often be divided between ligatures, as is done during portosystemic shunting. If the native vena cava is to be resected (see discussion below on alternative ‘piggy-back technique'), the infrahepatic vena cava is carefully encircled just above the right renal vein. The right adrenal vein is identified, ligated, and divided. Then, proceeding proximally, the retroperitoneal attachments of the intrahepatic vena cava are divided up to the suprahepatic vena cava which is encircled close to its exit from the liver. By working at this level, the recipient phrenic veins are usually not encountered—being left attached to the cuff above the site where the vena cava will be divided. After this dissection, the mobilized liver is held in place only by its venous attachments.

Final resection of the recipient liver is then completed by clamping and transecting the portal vein high in the hilum. Vascular clamps are placed on the infrahepatic and suprahepatic cava. A long suprahepatic caval cuff is developed by incising the liver parenchyma and transecting the hepatic veins within the substance of the liver (Fig. 5). The infrahepatic cava is severed close to its junction with the caudate lobe. The liver is removed from the field and any major retroperitoneal bleeding sites are suture ligated. This maneuver typically results in greatly diminished bleeding from the hepatectomy site. Final hemostasis can then be achieved by oversewing or cauterizing the exposed retroperitoneal tissues in the hepatic fossa (Fig. 6).

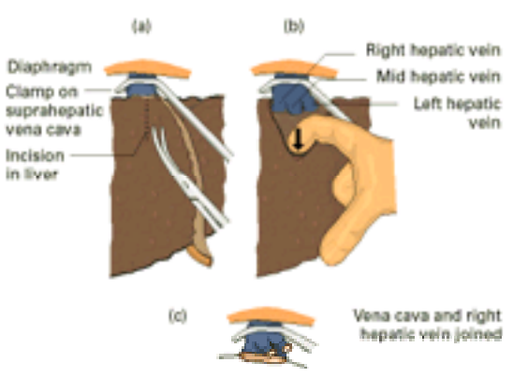


Fig. 5. Transection of recipient suprahepatic vena cava. A long cuff is developed by dividing the hepatic veins within the parenchyma of the liver. (a, b) The veins are then opened into the caval lumen to provide the anastomotic site (c).

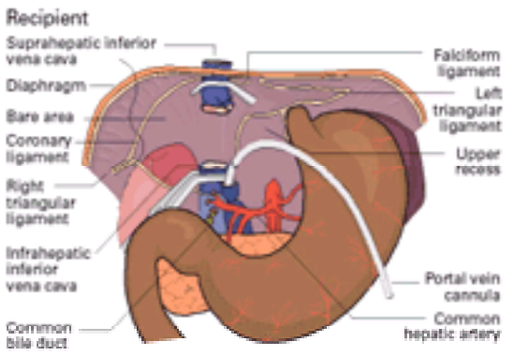


Fig. 6. Control of bleeding in retrohepatic tissues. The ligaments and retroperitoneal adhesions which had been divided to complete the recipient hepatectomy are fully exposed and can be oversewn or cauterized for complete hemostasis.

Venovenous bypass and ‘piggy back' orthotopic liver transplantation

Venovenous bypass was developed early in experimental canine models since canines do not tolerate caval clamping because of hemodynamic instability. Humans, however, tolerate caval clamping and the anhepatic phase much better than the canine. The early use of bypass in humans resulted in numerous embolic complications as a result of non-heparin bonded tubing and the lack of systemic anticoagulation during the procedure. The procedure was thus abandoned and Starzl's group performed over 200 liver transplants without the use of venovenous bypass. However, a consecutive series of complications resulting from intraoperative instability during the anhepatic phase of the surgery prompted Shaw to revisit the use of venovenous bypass in 1982.

The technique of venovenous bypass has since been modified and greatly simplified. Currently, an efferent limb of the circuit composed of heparin-bonded tubing is inserted percutaneously or under direct vision into the recipient's femoral vein for decompression of the lower systemic venous bed. The afferent limb of the circuit is placed into either the axillary, subclavian, or internal jugular vein for blood return to the heart. An in-line centrifugal pump provides flow rates of 1 to 3 1iter/min (Fig. 7). The third cannula representing a second efferent limb of the venovenous bypass may be introduced into the transected portal vein to decompress the splanchnic venous system. This system allows return of 40 to 50 per cent of the cardiac output to the heart from the lower body and/or viscera. In the majority of cases, the use of the femoral afferent limb is all that is necessary to maintain hemodynamic stability. Continuous monitoring of central and systemic pressures are required during this critical period. Hypotension may be caused by hypovolemia, inadequate bypass return, air or thromboemboli from the pump, myocardial depression secondary to hypothermia, or hypocalcemia resulting from citrate intoxication during the rapid transfusion of blood products. Although there is a theoretical possibility of hypoglycemia during the anhepatic stage, the blood sugar typically remains normal to elevated, presumably from the numerous glucose-containing carrier infusions.

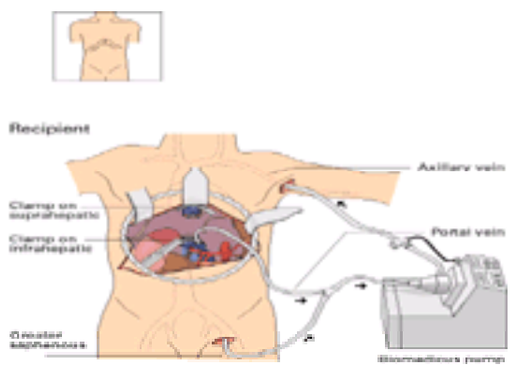


Fig. 7. Anhepatic stage of the transplant procedure with venovenous bypass. Lower body systemic venous blood and splanchnic blood are drained via the femoral and portal cannulas, respectively. Blood is returned to the heart by the in-line centrifugal pump via the axillary vein cannula.

Although venous bypass during the anhepatic phase of the procedure is rarely an essential component of the operation, severe splanchnic venous congestion can occur in its absence (Fig. 8), and some centers routinely use bypass for adult recipients. The advantages and disadvantages of venovenous bypass are summarized in Table 7. The hemodynamic stability provided by high-volume blood return to the heart from otherwise obstructed venous beds allows revascularization of the donor liver in a less urgent fashion. This provides a potential advantage for training of surgeons in this phase of the liver transplant operation.

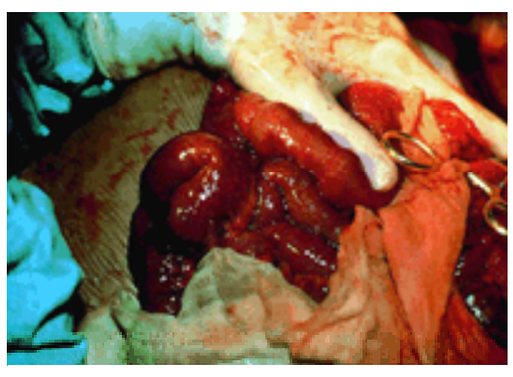


Fig. 8. Hemorrhage and edema of bowel following portal vein occlusion without venovenous bypass. Although such severe splanchnic congestion is not inevitable, routine use of venovenous bypass during the anhepatic phase of the operation essentially eliminates this complication.

Pros
Hemodyamic stability
Improved preservation of renal function
Venous decompression
Decreased blood loss
Cons
Increased operative time
Increased warm ischemia time
Local complications related to venovenous bypass
Increased cost

Table 7 Venovenous bypass in liver transplantation

The use of the piggy-back technique was developed from the experience of liver transplantation in children. As the lack of appropriate-sized organs necessitated transplantation of reduced-sized liver allografts from adult cadaveric donors or living related liver segments into children, innovative approaches to liver revascularization had to be developed (see below). One of the problems encountered, the lack of an adequate donor vena cava, led to a modification of the recipient hepatectomy in which the native vena cava was preserved in its entirety (Fig. 9). This enabled anatomosis of the donor suprahepatic cava or hepatic vein to the recipient's hepatic vein(s) orifices and/or the vena cava. The donor infrahepatic vena cava, if present, was closed to achieve hepatic venous outflow in an end-to-side manner. When applied to the adult population, the advantage of this maneuver is that the hepatectomy can be performed without interruption of the venous return to the right heart. Since hemodynamic integrity is maintained *in vivo*, the piggy-back technique nullifies the need for venovenous bypass and its associated problems, namely increased operative time, local complications related to bypass, increased expense of the additional equipment, and the specially trained personnel to operate the equipment. However, a more meticulous dissection of the liver is necessary for the success of this maneuver. Moreover, prior right upper quadrant surgery or an abnormally large, diseased liver may preclude the use of piggy-back liver transplantation.

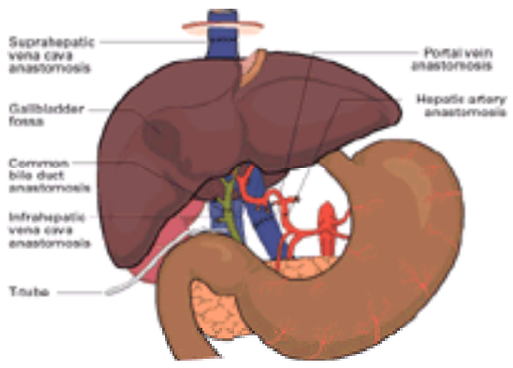


Fig. 9. Completion of the five anastomoses of orthotopic liver transplantation. The end-to-end suprahepatic vena cava, infrahepatic vena cava, and portal vein anastomoses are performed first to allow early reperfusion of the allograft. Arterial reconstruction may be to recipient proper hepatic artery as shown here or, more typically, to the common hepatic artery, as in Fig. 3. The biliary reconstruction is decompressed and stented with a T-tube.

Orthotopic allograft revascularization

The suprahepatic vena caval cuff is tailored by opening the recipient hepatic veins into the caval lumen. Sutures are inserted at the corners of donor and recipient veins and the allograft is gently lowered into the hepatic fossa. The posterior half of the anastomosis is accomplished using everting mattress sutures placed from within the lumen, while the anterior half may be completed in simple over and over fashion. The infrahepatic cava is reconstructed by the usual end-to-end venous anastomotic technique. Prior to completion of the anterior wall of this anastomosis, the allograft is flushed via the donor portal vein with cold or room-temperature Ringer's lactate solution. An alternative approach is to flush the allograft with portal venous blood after completion of that anastomosis prior to unclamping the suprahepatic cava. This important maneuver removes air and any remaining, hyperkalemic preservation fluid as well as acidic radicals that may have accumulated during the ischemic interval.

Preparation for the portal vein anastomosis requires adequate spatulation of the donor and recipient veins to ensure a large unrestricted anastomosis. As noted earlier, this anastomosis may have to be placed at the confluence of recipient splenic and superior mesenteric veins or, in some instances, on to a vein graft. Many surgeons recommend that the suture for the venous anastomoses be tied leaving 1 cm or more of ‘growth factor' between the knot and the vessel wall. This enables expansion of any ‘purse-stringing' effect and limits the likelihood of anastomotic stricture. At this juncture, most surgeons favor release of the venous clamps to restore portal flow to and vena caval flow through the liver. If the venovenous bypass is utilized, it can be discontinued and the remaining catheters are removed to prevent intraluminal clot formation. Hemostasis of the retroperitoneum and the venous anastomoses is ensured at this point.

The hepatic artery anastomosis may be constructed by suturing the Carrel patch of the donor celiac axis to the recipient hepatic artery at the level of the gastroduodenal take-off. Various arterial anomalies frequently require modification in the method of reconstruction, even including insertion of donor iliac artery as a vascular graft to provide inflow directly from the recipient's infrarenal aorta or supraceliac aorta in the case of an inadequate hepatic artery.

Next, prior to biliary reconstruction which significantly impairs the mobility of the liver, another inspection to ensure hemostasis is performed. After cholecystectomy, an end-to-end choledochocholedochostomy using interrupted absorbable sutures is the standard approach for recipients whose native biliary tree is not diseased. At the surgeon's discretion, biliary stenting can be achieved with either a T-tube brought out through the recipient duct (Fig. 10) or a pediatric feeding tube introduced via the donor cystic duct stump. Reconstruction by Roux-en-Y choledochojejunostomy is indicated for patients whose primary hepatic disease involved the biliary tree (such as biliary atresia and sclerosing cholangitis). This anastomosis can also be stented, using a small feeding tube brought out through the side of the Roux limb. Biliary stenting has the advantage of allowing postoperative radiographic studies of the anastomosis if any question of obstruction or leak develops.

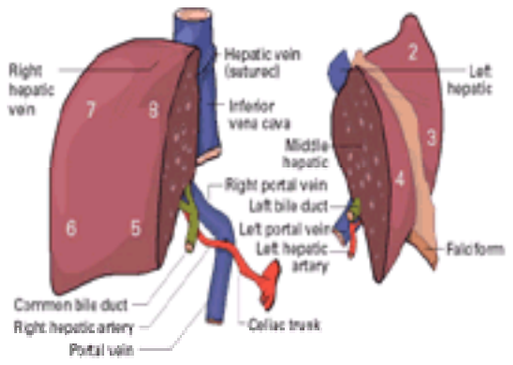


Fig. 10. Split liver preparation for transplantation into two recipients. The portal vein, celiac trunk, common bile duct, and inferior vein cava remain attached to the right lobe. The left lobe will be revascularized and drained by lobar pedicles including the left portal vein, left hepatic artery, left bile duct, and middle and left hepatic veins. Division of the allograft into true left and right lobes leaves the falciform and round ligaments (RL) attached to the left lobe. The numbers indicate hepatic segments.

After all anastomoses are completed, final hemostasis to control surgical bleeding is ensured. Closed suction drains can be placed above and below the liver to prevent accumulation of intra-abdominal fluid collections and the abominal wound is closed in layers.

Reduced-size, split liver, and living-donor liver transplantation

The limited availability of suitably sized organ donors for infants and small children with endstage liver disease awaiting liver transplantation has spurred the development of techniques to adapt larger organs for smaller recipients. The first technique, reduced-size liver transplantation whereby a cadaveric liver is pared down to appropriate size, was first reported in 1984. However, this approach did not gain popularity until the introduction of UW solution in 1988, which relaxed previously stringent time constraints, thereby facilitating the requisite *ex vivo* anatomic dissection of the organ after procurement and prior to implantation. Depending on the donor–recipient mismatch and the liver itself, the left lateral segment (Fig. 11), left lobe, or right lobe can be prepared for transplantation and the remainder discarded. Experience from numerous centers has demonstrated that survival after reduced-size liver transplantation is comparable with that of whole, size-matched grafts.

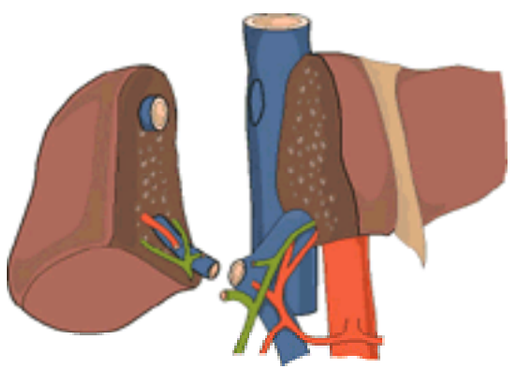


Fig. 11. Procurement of segments 5 to 8 for adult living-donor liver transplant..

Although reduced-size liver transplantation has successfully decreased the mortality of children on the list, concerns surrounding withdrawal of these livers from the

adult donor pool motivated the additional innovation of split liver transplantation, first reported in 1988. Instead of discarding the unused portion, the cadaveric organ would be partitioned into two viable grafts for transplantation into two separate recipients, most frequently an adult and a child (Fig. 11). Initial results of split liver transplantation were disappointing in comparison with those obtained with whole and reduced-size grafts. However, with a more sophisticated approach to donor and recipient selection along with greater technical expertise, recent experience from European and American centers with a significant experience in split liver transplantation has shown considerably improved results.

The technique of split liver transplantation has been further extended to *in situ* division of the liver in the heart-beating donor based on the success of living related liver transplantation (discussion follows) as a modality for pediatric transplantation (Fig. 10). The *in situ* strategy offers multiple advantages over reduced-size liver transplantation and split liver transplantation: (i) decreasing cold ischemia time and avoiding the potential for partial rewarming of the graft during *ex vivo* dissection, thereby decreasing the incidence of poor initial graft function/primary graft non-function; (ii) assurance of hemostasis at the parenchymal division surface; (iii) more precise dissection in the hilum resulting in less biliary devascularization and decreased incidence of post-transplant biliary complications; (iv) enhanced preservation of segment 4 with the ability to monitor its viability during the division procedure; and (v) ease of sharing grafts between distant institutions thereby maintaining equitable access to grafts for adults on the waiting list. Disadvantages of the *in situ* strategy include a longer and more technically challenging procurement operation. However, several studies have shown that with an experienced liver surgeon, the additional time required is modest (1.5 to 2.5 h) and no special equipment is necessary. Results of transplantation for both the left lateral segment graft (segments 2 and 3) and the right trisegmentectomy graft (segments 1 and 4 to 8) have been excellent as reported by institutions with a significant experience.

The use of live liver donors, developed in the 1990s, was similarly motivated by the severe organ shortage and the significant death rate of pediatric patients awaiting liver transplantation. The vast majority of the living related liver transplantation experience has involved procurement of the left lateral segment (segments 2 and 3) from an adult for transplantation into a pediatric patient. Since the results of living related liver transplantation have equaled or exceeded those of whole organ and reduced-size liver transplantation, this technique has become a standard strategy for pediatric transplantation. The current depth of the pediatric armamentarium has virtually eliminated mortality on the pediatric waiting list.

In contrast to the pediatric situation, the mortality of adult patients awaiting liver transplantation continues to increase yearly. Recently, several centers have reported their experience with the use of right lobe (segments 5 to 8) (Fig. 11), extended right lobe (segments 4 to 8), and left lobe (segments 2 to 4) grafts from living donors for adult recipients. Although most of these reports have emanated from Asian countries that have no access to cadaveric donors, this technique has also been adopted by a limited number of North American and European centers in an attempt to expand the donor pool. Widespread application of this strategy is significantly limited by both donor and recipient concerns. The live donor is subject to a small but significant risk of surgical morbidity and mortality. Furthermore, the selection of an appropriate donor for a given recipient requires careful consideration towards ensuring the adequacy of liver mass in conjunction with the patient's clinical condition.

Inherent to all of the above procedures in which a liver is divided prior to transplantation are multiple technical considerations particularly surrounding the division and reconstruction of the vascular and biliary structures which deserve special mention. In general, avoiding the use of grafts for vascular reconstruction has resulted in significantly decreased thrombosis rates. For a right lobe or right trisegmentectomy graft derived from a cadaveric liver, vascular reconstruction is straightforward since the vena cava, the main portal vein, and the celiac axis usually come with the graft. In the case of a living donor, hepatectomy must preserve the recipient's vena cava and the allograft right hepatic vein is typically anastomosed to the recipient's right hepatic vein orifice. The donor's main portal or right portal vein is sewn to the recipient's portal vein. Biliary reconstruction usually involves only a single donor duct, either the common bile duct or right hepatic duct, which can be anastomosed to either the recipient duct or a Roux-en-Y limb of jejunum. All right lobe grafts sit in an orthotopic position in the subphrenic space so there is little concern about torsion of hepatic venous outflow. Vascular and biliary reconstruction of left lateral segment grafts, particularly those derived from an *in situ* split liver or a living donor can be more technically challenging. The donor's left hepatic vein is anastomosed to either the recipient's left hepatic vein or to a common orifice created by incising the septum between the recipient's middle and left hepatic veins. The donor's left portal vein is most often anastomosed to the recipient's main portal trunk, the confluence of the right and left portal veins, or the confluence of the splenic and superior mesenteric veins. Only in the case of an unusually short graft portal vein and/or an unusable recipient portal vein is a vein graft used. The arterial reconstruction for a left lateral segment graft has historically been problematic since the small caliber of donor and recipient vessels led to increased rates of hepatic artery thrombosis. However, the technique of microsurgical reconstruction, typically under 10 times magnification of an operative microscope, using an interrupted suture technique and fine (8–0 to 9–0) polypropylene suture material has dramatically reduced the rate of thrombosis. Adjunctive pharmacotherapy to maintain arterial patency with low-molecular-weight dextran and/or heparin followed by aspirin is commonly employed. Biliary reconstruction often requires the anastomosis of one or more ducts to a Roux-en-Y jejunual limb. Finally, after implantation, the left lateral segment graft must be examined to ensure that it does not have a tendency to fall into the subphrenic space which may then compromise hepatic venous outflow.

Auxiliary liver transplantation

Rarely, there have been instances of hepatic allograft recipients so unstable that they would be predicted not to tolerate hepatectomy and orthotopic replacement. Transplantation of an auxiliary liver into a heterotopic position while leaving the diseased organ *in situ* has been attempted as an alternative for such high-risk individuals. However, long-term success following clinical attempts at heterotopic liver transplantation has been poor, primarily because of size constraints leading to technical complications and infection. Modifications of the procedure, with reduction of graft size by partial hepatectomy, may provide a satisfactory solution. For this approach, the donor liver is prepared *ex vivo* by removing the left lobe and reducing the vena cava to a short segment draining the right and middle hepatic veins. In the recipient, dissection is primarily limited to the portal area, as would be required for a portacaval shunt procedure. The infrarenal aorta is also freed over a length of 3 to 4 cm. The reduced-size liver is oriented in the right subhepatic region with the donor vena cava overlying the recipient cava and the graft hilum facing the aorta (Fig. 12). Vascular anastomoses include donor vena cava to side of recipient suprarenal vena cava, donor portal vein to side of recipient portal vein, and Carrel patch of donor aorta to side of recipient aorta. Biliary drainage is into a Roux-en-Y jejunal loop. Experience with this approach remains extremely limited.

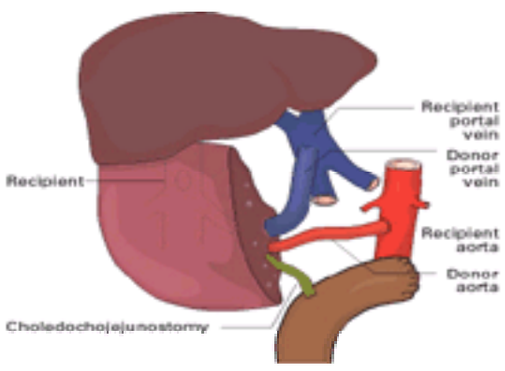


Fig. 12. Auxiliary partial liver transplantation. The donor right lobe is oriented in the subhepatic region with the hilum facing recipient aorta. Vascular anastomoses are: donor vena cava to side of recipient suprarenal vena cava, donor portal vein to side of recipient portal vein, and Carrel patch of donor aorta to side of recipient aorta. Choledochojejunostomy using a Roux-en-Y loop provides bilary drainage.

Depending on the portal vein reconstructive technique used in auxiliary liver transplantation, functional competition has been reported. Of the patients with preservation of the portal blood flow to the native liver, studies have shown that inadequate graft portal flow can occur immediately after the operation. As a consequence of decreased portal blood flow, deterioration of graft function, and allograft atrophy have been observed. Pre-emptive interruption of portal flow to the native liver can optimize flow to the allograft. In this setting, the native liver, supplied only by arterial inflow, may suffer only minimal damage while the portal steal phenomenon is alleviated.

Transplantation of an auxiliary liver into a heterotopic position while leaving the diseased organ *in situ* could be a suitable alternative for patients with hepatic enzyme deficiencies such as ornithine transcarbamylase deficiency or Crigler–Najjar type I disease. Reports have shown that a partial liver segment with normal enzyme activity corrects the metabolic defect and, in the case of allograft failure, the native liver with the ability to recover function is still present. Thus the occasional use of auxiliary liver transplantation may prove to be an alternative and safe approach in these patients with hepatic enzyme deficiencies.

Post-transplantation management and complications

Non-immunologic considerations

The early postoperative management of liver allograft recipients requires much more intensive monitoring and support than is required following kidney or pancreas transplantation. Because of the precarious pretransplant condition, the complicated operation and its attendant massive fluid shifts, and the coagulopathy and

hemodynamic instability that can result if onset of normal allograft function is delayed, these recipients are initially cared for in an intensive care unit. Satisfactory postoperative cardiac output, maintenance of adequate renal function, and correction of abnormal clotting parameters may require aggressive colloid replacement and support with vasoactive drug infusions. Other important aspects of the recipient's care can be considered as they relate to individual organ systems.

Hepatobiliary system

Satisfactory liver allograft function is manifested by the capacity to correct acidosis and hypocalcemia and the restoration of clotting and neurologic function. Production of deeply pigmented bile during the biliary anastomosis or the appearance of such in a biliary catheter, if employed, also suggests adequate hepatic function. These recipients will typically maintain good renal function and will recover rapidly from anesthesia so that successful extubation is possible, usually within 12 to 48 h. If, however, there is minimal or watery, depigmented, so-called ‘white' bile, the early postoperative care will be more complicated. These recipients can be reliably predicted to require assisted ventilation, intensive infusion of fresh frozen plasma to correct severe coagulopathy, and other supportive measures before adequate stabilization may be achieved. As noted above, retransplantation may be the only recourse for some patients with early graft dysfunction. Timely identification of those recipients with irreversible ischemic injury is critical if retransplantation is to be successfully accomplished before fatal cerebral edema develops. As in the pretransplant candidate with fulminant hepatic failure, uncorrectable coagulopathy (prothrombin time over 25 s), hypoglycemia, acidosis, and new onset of renal failure are ominous. These findings, which usually prompt the urgent search for a second allograft, are typically encountered in less than 2 to 3 per cent of liver recipients.

Technical complications most commonly include intra-abdominal bleeding, vascular thrombosis, biliary duct problems, or infected fluid collections. The incidence of postoperative bleeding is directly related to the severity of intraoperative bleeding and the quality of immediate allograft function. Despite meticulous surgical technique and hemostasis in the operating room, diffuse oozing may persist until correction of hypothermia and inadequate serum clotting factor levels. If major blood loss persists despite reversal of coagulopathy, emergency re-exploration is indicated to evacuate hematoma and correct any surgical causes of bleeding. In patients who bleed initially but stop after aggressive medical management, laparotomy may still be required in order to evacuate hematoma which may become secondarily infected in the setting of immunosuppression.

Hepatic artery thrombosis, particularly in small children, should be suspected in all patients who present with acute onset of transaminitis, unexplained fever, biliary leak, or positive blood culture with biliary organisms. Doppler ultrasonography followed by angiography (conventional and/or magnetic resonance) can confirm the diagnosis. If the clinical presentation is one of fulminant liver failure, biliary leak, or relapsing bacteremia, emergency retransplantation is the only viable option. A small proportion of patients who develop mild hepatic dysfunction without abscess formation or recurrent sepsis may be conservatively managed while arterial collaterals to the allograft develop. However, if hepatic artery thrombosis is diagnosed early in the postoperative course, emergency re-exploration and thrombectomy of the clotted vessels is the preferred method of treatment. Early intervention has resulted in successful revascularization of up to 70 per cent of hepatic artery thromboses that occur within 1 week of the initial transplant.

The presenting symptoms of portal vein thrombosis are usually less devastating. Rarely, in the immediate postoperative period, this complication can produce severe biochemical abnormalities and even hepatic necrosis requiring retransplantation. More commonly, portal vein thrombosis is suggested by recurrent variceal bleeding, intractable ascites, or an unexplained elevation in prothrombin time. Successful management of this complication in patients without severe hepatic dysfunction is usually achieved by operative thrombectomy and correction of any technical abnormality encountered. In the late post-transplant period, a splenorenal shunt may be effective.

Bile duct complications, which historically accounted for the majority of postoperative morbidity, have greatly decreased in frequency since reconstruction techniques have become standardized. Placement of a feeding tube or T-tube stent allows ready investigation of any suspected biliary tract complications. However, many centers have abandoned their routine use because biliary peritonitis frequently developed upon removing the stent. Anastomotic leak, usually manifested by appearance of bile in the abdominal drain or an unexplained rise in serum bilirubin, may occur during the first few weeks after transplantation. Minimal, well-drained leaks, not associated with hepatic artery thrombosis, typically close spontaneously. More extensive leaks require endoscopic retrograde cholangiopancreatography for intraluminal stent placement to promote enteric drainage of bile. Failure of endoscopic techniques usually necessitates operative intervention for primary repair, reanastomosis, or even conversion to Roux-en-Y choledochojejunostomy dependent on the intraoperative findings at the time of re-exploration. Even in the current setting of limited donor organs, retransplantation must be considered for leaks secondary to arterial thrombosis. Biliary duct stenosis and/or obstruction usually occur later in the postoperative course, regardless of whether the anastomosis was stented or not. However, those that were stented may have a delayed manifestation of this complication. Short strictures can be dilated via endoscopic or percutaneous approaches. Longer strictures necessitate surgical bypass of the obstruction, usually by conversion to Roux-en-Y choledochojejunostomy. Earlier observations of diffuse obstruction by ‘sludge' within the biliary tree are now seldom encountered. Control of this complication is presumably due to careful biliary flushing during the donor procedure, better cold preservation conditions, and improved methods of reconstruction. An interesting and unusual cause of obstruction has been the development of a tension mucocele in a donor cystic duct remnant which shared a common wall with the bile duct, a phenomenon analogous to the Mirizzi syndrome.

Postoperative non-hepatotropic viral hepatitis, usually caused by cytomegalovirus, can also cause graft dysfunction. Following serologic or hepatic biopsy confirmation of the diagnosis, intravenous ganciclovir therapy is commonly instituted. Other viral infections, including recurrent hepatitis B or C, have been discussed earlier. Occasionally, differentiating these causes of allograft dysfunction from rejection may be difficult until clinical information is analyzed in conjunction with biopsy-derived histopathologic findings.

The prostaglandins have *in vivo* activities of vasodilatation (peripheral, renal, and splanchnic), lysosomal membrane stabilization, and inhibition of platelet aggregation. A number of prostaglandin subtypes have been studied in liver preservation and the treatment of primary graft non-function. Initial reports of prostaglandin use in fulminant hepatic failure and primary graft non-function demonstrated improved patient survival. However, a subsequent prospective, randomized, placebo-controlled trial using prostaglandin E_1 in 160 liver transplant recipients demonstrated similar patient and graft survival as the incidences of acute cellular rejection and primary graft non-function were not significantly different in the two groups. Nevertheless, in patients with surviving grafts, prostaglandin E_1 administration resulted in a 23 per cent shorter length of hospital stay and a 40 per cent shorter length of time in the intensive care unit postoperatively. The purported role of prostaglandin E_1 in vasodilatation may account for the reduced needs for renal supportive therapy as a trend toward improved survival rates was observed in patients with mild renal impairment (preoperative serum creatinine of 1.5 mg/dl or greater). These results suggest that the use of prostaglandin E_1 in liver transplant recipients may reduce the overall morbidity and the cost of transplant-related hospital stays.

Pulmonary system

Aggressive efforts to prevent pulmonary complications are essential in all liver transplant recipients. Intensive chest physiotherapy, postural drainage, and endotracheal suctioning are components of routine postoperative care. If not extubated in the operating room, weaning from assisted ventilation is begun as soon as the recipient is hemodynamically stable and has recovered from anesthesia. Successful extubation can be accomplished usually within 12 to 48 h. Several unusual factors unique to liver allograft recipients may delay weaning. These include right diaphragmatic paralysis, which may result from the intraoperative placement of the suprahepatic vascular clamp, and the metabolic alkalosis which these patients often develop. Because the latter may contribute to compensatory hypoventilation, metabolic correction with potassium chloride, or very rarely hydrochloric acid, may be necessary to allow successful early extubation. Right-sided pleural effusions are common in the early postoperative period. Thoracentesis is rarely required to provide satisfactory lung re-expansion but should not be withheld when necessary, since even limited episodes of pulmonary dysfunction are poorly tolerated in these nutritionally depleted, debilitated, and immunosuppressed patients.

Significant hypoxemia resulting from the presence of arteriovenous malformations in patients with chronic liver disease may persist into the early postoperative period. As the systemic vascular resistance is corrected after liver transplantation, patients with hepatopulmonary syndrome who also have increased pulmonary hypertension may experience right ventricular compromise leading to acute right heart failure which can be extremely difficult to manage. The best preventive measure is to maintain satisfactory oxygenation of the patient and adequate pulmonary toilet so that physiologic intrapulmonary redistribution and shunting does not occur to cause transient pulmonary hypertension. Although the prognosis for such patients is somewhat less favorable, such shunting, even advanced to the stage of digital clubbing, has been observed to reverse following successful liver transplantation.

Cardiovascular system

Initial hemodynamic stability is maintained by volume administration and the addition of inotropic, chronotropic, or vasoactive agents as guided by clinical findings and continuous invasive monitoring of the usual cardiopulmonary variables. Acute onset of arrhythmias is unusual and is typically associated with hypoxemia or severe electrolyte abnormalities. Ischemic cardiac disease in the immediate postoperative period is atypical since the exhaustive preoperative evaluation for liver transplantation includes cardiac risk assessment.

Renal and metabolic systems

Acute renal failure may develop immediately postoperatively due to pre-existing renal insufficiency resulting from acute tubular necrosis or hepatorenal syndrome, intraoperative vena caval occlusion, transient intravascular volume depletion, hepatic allograft dysfunction, or administration of nephrotoxic drugs. All of these factors

are exacerbated by the administration of cyclosporine or tacrolimus. Extensive experience has established that in the presence of satisfactory hepatic allograft function, pretransplant renal dysfunction typically reverses in the first few days after transplantation. For patients with persistent oliguria and renal dysfunction, conversion to OKT3 as a calcineurin-sparing strategy may shorten the requisite time for renal recovery and thereby avoid the need for dialysis. If dialysis is necessary, the technique of continuous venovenous hemofiltration/dialysis is preferable to standard intermittent hemodialysis, particularly in the setting of allograft dysfunction. A small percentage of patients are however left with persistently compromised renal function which gradually progresses under chronic calcineurin inhibitor therapy, even to the need for renal transplantation.

Hematologic system

Leucopenia and thrombocytopenia secondary to hypersplenism typically persist in the early postoperative period. The leucopenia is usually of no clinical consequence; prolonged thrombocytopenia, however, may require platelet transfusions and reduction in the dosage of antimetabolite immunosuppression (azathioprine or mycophenolate mofetil). Recipients of an ABO compatible but mismatched graft can manifest hemolysis 1 to 2 weeks after transplantation resulting from the production of antirecipient isohemagglutinins by the graft. Although hemolysis is usually transient, transfusions using donor ABO-grouped red blood cells may be necessary. If hemolysis is severe and/or persistent, high-dose corticosteroids, plasmapheresis, and even splenectomy represent the heirarchy of therapeutic options.

Another unusual hematologic complication has been the development of aplastic anemia in some patients who undergo liver transplantation for fulminant hepatic failure. The onset of aplastic anemia has usually been within 1 to 6 months following transplantation and subsequent fatal infection is not infrequent. Some patients have had recovery of marrow function, either spontaneously or following antilymphocyte therapy, and others have undergone bone marrow transplantation. The most effective therapeutic approach has not been defined.

Nutrition

Postoperative ileus usually resolves by the third or fourth day, allowing normal enteral alimentation. For obtunded and/or intubated patients, nasoenteric tube feedings are preferred to avoid the risks of aspiration. Hyperalimentation, with its risks for infection and hepatic toxicity, is reserved for the few patients whose gastrointestinal tract cannot be used.

Neurologic and psychiatric complications

Seizures and other neurologic or psychiatric disorders, such as persistent obtundation, expressive aphasia, confusion, and transient psychoses, are common after liver transplantation. Some of the identified etiologic factors have included air embolism, cerebral edema, intracranial bleeding, electrolyte disturbances, cyclosporine or tacrolimus toxicity, and sleep deprivation in an intensive care unit setting. Air embolism is historically included but has seldom been seen since the importance of flushing the liver prior to revascularization was recognized. Calcineurin inhibitors such as cyclosporine and tacrolimus have been thought to lower the seizure threshhold that can be further exacerbated by hypomagnesemia and hypocholesterolemia. Benzodiazepines are used for acute control whereas Tegretol is used for chronic treatment and prophylaxis of seizures. Long-term use of phenobarbital or dilantin is generally avoided because these medications induce the cytochrome P450 microsomal enzyme systems leading to a rapid decrease in blood concentrations of calcineurin-inhibiting immunosuppressive agents.

General surgical care

One or two perihepatic drain(s) can be placed which is (are) removed usually within 48 h after transplantation. The portal drain area is left in place until cholangiography, if a biliary stent has been utilized, or HIDA scan, if a primary unstented anastomosis is employed, confirms the integrity of the biliary tree. If a biliary stent is present, it is usually tied off and left in place for 3 to 6 months to allow adequate time for a tract to develop in a patient on chronic steroid therapy. Earlier removal not infrequently leads to bile spillage and localized peritonitis. In addition to the management of drains and tubes, meticulous wound care is also important to prevent infection during this period of typically intensive immunosuppression.

Immunosuppression and rejection

Development of new immunosuppressive agents over the past three decades has been guided by an ever-growing understanding of the mechanisms of allorecognition and allograft injury. In the 1960s, immunosuppressives were limited to drugs with non-specific actions. Chemotherapeutic agents such as cyclophosphamide and mercaptopurine, which had an effect on nucleic acid metabolism, were utilized to prevent cell proliferation. With increasing knowledge of the human immune system over the ensuing decades, newer immunosuppressive drugs were developed whose actions are more selective. As both Starzl and Murray demonstrated in 1963, the combination of a non-specific immunosuppressive such as a steroid and a nucleic acid synthesis antagonist such as azathioprine can result in prolonged survival of human kidney allografts. This strategy of inhibiting DNA replication was refined recently with the advent of mycophenolate mofetil that inhibits *de novo* but not salvage pathway purine nucleotide synthesis. The advantage of mycophenolate mofetil is enhanced specificity since lymphocytes, the cells responsible for recognition and effector mechanisms of rejection, are unable to utilize the salvage pathway of purine synthesis. Cyclosporine and tacrolimus, although chemically unrelated, are potent calcineurin inhibitors which downregulate interleukin 2 (**IL-2**) gene transcription thereby inhibiting T-cell differentiation and function. In addition, the development of polyclonal (antithymocyte globulin) and monoclonal antibodies (anti-CD3 monoclonal antibody) starting in the early 1970s led to their use in induction immunosuppression protocols and for the treatment of rejection. Newer monoclonal antibodies being evaluated at present, such as IL-2 receptor inhibitors, may find their role in immuno-suppression therapy in transplant recipients.

Currently, immunosuppressive regimens commonly use a combination of steroids, cyclosporine or tacrolimus, and azathioprine or mycophenolate mofetil. Our current protocol for adult liver transplantation is a triple-drug regimen that includes methylprednisolone or prednisone which is started intraoperatively and tapered over 6 days from 200 mg/day to an initial maintenance dose of 20 mg/day. Tacrolimus can be started on the day of transplantation if renal function is adequate. A loading dose of azathioprine (5 mg/kg) is given preoperatively unless there is severe leucopenia or thrombocytopenia, and then continued at 1 to 2 mg/kg per day as determined by the patient's hematologic profile. The monoclonal antibody OKT3 or polyclonal preparations such as antithymocyte globulin can be used as part of a calcineurin inhibitor-sparing strategy for patients with severe or prolonged renal dysfunction. Substitution of mycophenolate mofetil for azathioprine may provide compensatory immunosuppression for patients who cannot tolerate usual levels of tacrolimus because of persistent renal dysfunction. Because of the increased risk of post-transplant lymphoproliferative disease in children treated with tacrolimus, we substitute cyclosporine for this agent in our triple-drug regimen used in patients under 2 years of age.

Despite immunosuppression, 20 to 30 per cent of liver allograft recipients suffer one or more rejection episodes in the early post-transplant period. Typical biochemical and clinical criteria suggesting rejection are listed in Table 8. Differentiating rejection from other disorders, such as allograft ischemia and preservation injury, technical complications, or infection can be difficult. Radiographic examination including hepatic ultrasound with Doppler examination, cholangiography, and/or biopsy may be necessary for definitive diagnosis. Histologic changes of acute rejection include mixed cellular infiltrate invading the portal tracts with bile duct damage and endothelialitis (Fig. 13). With more chronic rejection, manifested clinically by cholestatic jaundice, the hepatic artery intima is infiltrated by foamy macrophages which produce considerable narrowing of the vessel lumen. There may be loss of interlobular bile ducts leading to the term ductopenic rejection and excess fibrosis appears in portal tracts. Unfortunately, many of the biopsy findings can also result from other conditions including cholangitis and viral hepatitis. Thus, the ultimate diagnosis must take into consideration the entire constellation of clinical, biochemical, and histopathologic findings as treatment of multiple suspected rejection episodes exposes the patient to unacceptable risks of infection.

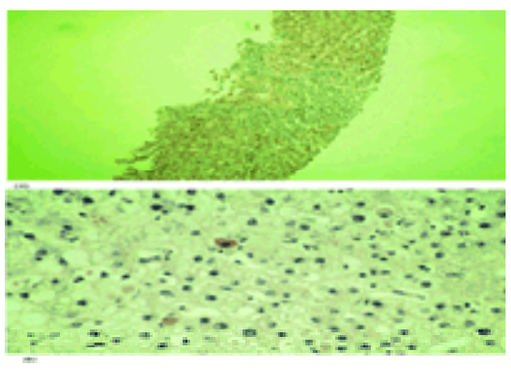


Fig. 13. Histology of recurrent hepatitis C in liver allograft.

Clinical
Fever
Anorexia
Biochemical
Elevated serum bilirubin
Elevated transaminases
Elevated alkaline phosphatase
Radiographic
Cholangiogram
No leak
No obstruction
Doppler ultrasound
No collections
No ductal dilatation
Patent vessles
Pathologic
Portal inflammation
Biliary duct injury
Endotheliabitis

Table 8 Indicators of hepatic allograft rejection

Rejection episodes are usually treated with a short course of pulse corticosteroids (Solumedrol, 500 mg/day for 2 days) with or without a subsequent oral steroid taper to baseline doses. If satisfactory reversal does not ensue, OKT3 monoclonal antibody or another antilymphocyte preparation can be initiated. For patients on cyclosporine, conversion to tacrolimus may reverse rejection. Retransplantation, although rarely required, may be the only recourse for some patients with relentless chronic hepatic allograft rejection.

The intensive immunosuppression required in these debilitated patients, who not infrequently also suffer technical complications, provides an ideal setting for invasive infections. The overall incidence of infection following liver transplantation has been reported to range from 45 to 80 per cent, which greatly exceeds that observed in recipients of other solid organ allografts. These complications are the major source of morbidity and the most common cause of death in liver transplant recipients.

The major infections likely to be encountered are summarized in Table 9. Infections occurring in the early postoperative period may arise from pre-existing recipient conditions, from agents transmitted with the allograft, or from the typical bacterial complications of surgery. Patients with uncontrolled extrahepatic infection should not undergo transplantation. If cultures from the organ donor become positive, specific antimicrobial therapy should be instituted. Postoperative bacterial complications have most commonly been wound infection, abdominal abscess, pneumonia, or bacteremia/diffuse sepsis. The vast majority of these are associated with technical complications such as bleeding, arterial thrombosis, or biliary anastomotic problems. During the later (after 1 month) post-transplant period, viral and other opportunistic infections become more prevalent. Herpes group viruses are the most common: cytomegalovirus infection has been observed in more than 70 per cent of recipients, herpes simplex in approximately 50 per cent, and herpes zoster in 5 per cent. Fortunately, the introduction of ganciclovir has greatly reduced the morbidity and mortality due to these agents. Infection with the Epstein–Barr virus can produce conditions ranging from an infectious mononucleosis syndrome to a life-threatening post-transplant lymphoproliferative disease similar to, or progressing to, B-cell lymphoma. This syndrome may regress if immunosuppression is substantially reduced to steroids alone. Antiviral therapy is frequently administered. If there is no response to these measures, conventional chemotherapy may be necessary.

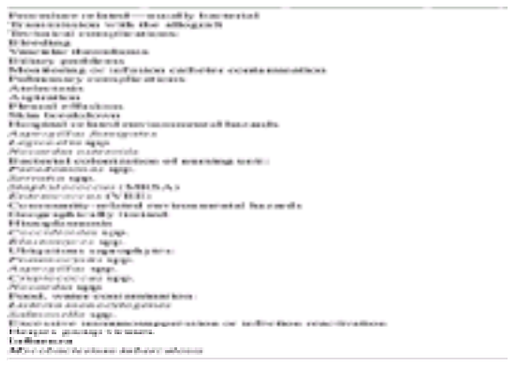


Table 9 Etiology of potential infections observed following liver transplantation

Opportunistic infections are usually manifest as a primary infection of the lungs, most commonly due to *Cryptococcus, Aspergillus*, or *Nocardia* spp. During the later time period, pneumonitis caused by the protozoan *Pneumocystis carinii* may also appear. Since these infections may progress rapidly or spread to other sites, aggressive efforts to identify the etiologic agent and treat the infection are essential. Investigations include examination of induced sputum, bronchoscopic lavage, brushings, or biopsy, or even open lung biopsy, if necessary, to identify the causative agent.

Results and rehabilitation

Overall success rates after liver transplantation in the current era are approximately 85 to 90 per cent at 1 year and 75 to 80 per cent at 3 years. Thus, the expected long-term survival for these previously untreatable patients now exceeds that provided by modern care for many other commonly encountered conditions, including myocardial infarction and most solid tumors. The majority of deaths occur in the first 3 months after liver transplantation, with early graft dysfunction or infection being the most common causes of failure (Table 10). Because we now have prolonged survival data well beyond 10 years, the longer-term prognosis for liver transplant recipients has also been defined. Death in patients who survive at least 5 years after transplantation occurs only rarely and suggests that the prognosis for the approximately three-quarters of recipients currently expected to survive 5 years is indeed excellent. Some of the earlier patients have now survived for over 30 years after transplantation.

	Early (< 3 months)	*Late* (4–36 months)
	(%)	(%)
Infection	6	7
Hepatic failure (primary non-function)	2	–
Technical complications	2	1
Rejection	–	1
Hemorrhage	1	–
Recurrent disease	–	3
Total	11	12

Table 10 Causes of death after liver transplantation

Nearly complete rehabilitation can be anticipated in the majority of surviving patients with demonstrable improvements across all health and psychosocial measures. Adult recipients are often able to return to a productive place in society. Over 80 per cent resume their previous occupations or lifestyle and usually report that their performance is even more effective because of improved stamina and mental clarity. In one study, only 18 per cent of recipients considered themselves not well enough to work. Of the remainder, 57 per cent were employed and the others reported they had taken early retirement (8 per cent), returned to school (3 per cent), were unable to find work (26 per cent), chose not to work (3 per cent), or could not work for family reasons (3 per cent). Not surprisingly, older recipients and those who were continuously out of work before transplantation were the least likely to return to gainful employment after liver transplantation. Children after successful liver transplantation usually have a return of physical health and resumption of normal growth and development. In addition, emotional well being is restored so that these young recipients typically return almost immediately to their normally expected school levels and activities.

Rehabilitation in some patients may be significantly delayed by the complications of osteoporosis which are commonly associated with chronic liver disease. Patients at high risk include those with evidence of hypogonadism, overt calcium malabsorption, previous steroid therapy, and cholestatic liver disease (particularly primary biliary cirrhosis). Despite successful transplantation, severe preoperative osteopenia may progress to compression fractures, back pain, and prolonged disability. Fortunately, this condition generally stabilizes within 3 to 6 months following transplantation. This again emphasizes the importance of the timely consideration of liver replacement before such secondary complications become too advanced. Other barriers to complete rehabilitation are related to immunosuppression, ranging from nephrotoxicity and hypertension to recurrent or *de novo* malignancies to chronic rejection requiring retransplantation.

Conclusions and the future of liver transplantation

Liver transplantation is currently the standard for therapy for patients with endstage cirrhosis who have clinical complications related to their liver disease. The operation is technically difficult and many patients are in poor clinical condition when they are referred for transplantation. Nevertheless, the procedure provides a chance for long-term survival and excellent rehabilitation for the majority of a group of patients whose lifespan would be measured in termsof months with other currently available therapy. The identification of more selective and efficacious immunosuppression together with better means of preserving the donor allograft have been two of the most important factors that have made transplantation the preferred therapy for selected patients with endstage liver disease.

Now that acceptable survival results have established the procedure as an acceptable therapeutic modality, other factors limiting its more widespread application must be addressed. The first of these is the question of cost effectiveness. The average cost of a liver transplant has been estimated to be between $US150 000 and 300 000. In comparison, the costs of withholding liver transplantation are not insignificant. Chronically ill patients with advanced liver disease incur the costs of income loss, medications, hospital stays, palliative surgical procedures, and predictably, terminal care needs. While these expenditures have progressively increased as palliative technology has improved, interestingly, the cost of liver transplantation has actually decreased as improvements in management are defined. For instance, one hospital reduced inpatient, organ procurement, and physician charges for the transplant procedure from $US154 000 in 1991 to 1992, to 103 000 in 1993 to 1995. These savings were achieved mainly by reducing length of stay without sacrificing outcome and quality of care. It is also likely that continued improvements in immunosuppression management can result in long-term savings. Although tacrolimus and cyclosporine-based immunosuppressive regimens provide similar survival rates, tacrolimus may save an average of approximately $US20 000 per transplant in the first year becauseof reduced rejection episodes. In another example, some centers have recommended routine administration of ketoconazole because its effect on cyclosporine metabolism reduces the dosage needed to maintain target levels by 62 per cent at 1 week and 80 per cent at 1 year. This can provide savings of approximately $US5000 in the first year and approximately $US4000 in each subsequent year (inclusive of the cost of the ketoconazole).

Nevertheless, the costs of newer medical approaches such as organ transplantation continue to be reviewed in terms of whether the funds being allocated, despite providing life-saving treatment, might not be more justifiably spent in other areas. In a dramatic example, one United States legislature at one point even voted to discontinue payment for all extrarenal transplantation. The not unexpected nationwide reaction to the death of a 7-year-old boy who was denied a bone marrow transplant subsequently led to a review and reversal of the decision. Interestingly, such critical questions are generally asked only of new medical therapies. In part, this is the fault of the medical community, which perhaps should be more active in identifying areas of established therapy that deserve re-evaluation. It was rather incongruous, for example, to note that in the same journal issue in which the above state legislation decision was reviewed, a report appeared indicating that some of the newer agents used to treat moderate hypertension cost as much as $US20 000 to 30 000 per year of life saved. While it is natural for new medical approaches to be the most likely targets for cost containment, scrutiny will probably soon extend to routinely accepted practices. As has been established in patients with endstage renal failure, the new approach (transplantation) may provide not only a better quality of life but may also be a more cost-effective long-term option than the alternative (dialysis).

The second major issue limiting liver as well as solid organ transplantation is the inadequate supply of donor organs. Currently, 5 to 15 per cent of adult patients die while waiting for a suitable liver, yet allografts are successfully transplanted from fewer than one-third of the estimated 12 000 to 18 000 people who die yearly in the United States in circumstances that could make organ donation feasible. Clearly, new approaches are needed to increase the number of organs available for transplantation and to improve upon the utilization of those organs procured.

Clinical liver transplantation is a procedure that is now recognized to be a life-saving therapeutic modality for endstage liver disease. However, it remains to be influenced by the realities of socio-economic and ethical issues in the determination of its applicability. Ultimately, new approaches, including the use of non-human donor organs and more specific immunosuppressive regimens (for example, sirolimus or more selective monoclonal antibodies such as an IL-2 receptor antagonist) will undoubtedly make the procedure not only more available but also more effective and economically attractive.

Further reading

Anonymous. National Institutes of Health Consensus Development Conference on Liver Transplantation. Sponsored by the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases and the National Institutes of Health Office of Medical Applications of Research. *Hepatology* 1984; **4**(1 Suppl.): 1S–110S.

Asher NL *et al.* Liver transplantation for fulminant hepatic failure. *Archives of Surgery* 1993; **128**: 677–84.

Bismuth *et al.* Hepatic transplantation in Europe. First Report of the European Liver Transplant Registry. *Lancet* 1987; **ii**: 674–6. [Published erratum appears in *Lancet* 1987 Dec 12; **ii**: 1414].

Bismuth H *et al.* Liver resection versus transplantation for hepatocellular carcinoma in cirrhotic patients. *Annals of Surgery* 1993; **218**: 145–51.

Bizollon T *et al.* Pilot study of the combination of interferon-a and ribavirin as therapy of recurrent hepatitis C after liver transplantation. *Hepatology* 1997; **26**: 500–4. [See comments.]

Brems JJ *et al.* Fulminant hepatic failure: the role of liver transplantation as primary therapy. *American Journal of Surgery* 1987; **154**: 137–41.

Broelsch CE *et al.* Liver transplantation, including the concept of reduced-size liver transplants in children. *Annals of Surgery* 1988; **208**: 410–20.

Broelsch CE *et al.* Application of reduced-size liver transplants as split grafts, auxiliary orthotopic grafts, and living related segmental transplants. *Annals of Surgery* 1990; **212**: 368–77.

Calne RY. *Liver transplantation: the Cambridge/King's College Hospital experience*. Grune & 38; Stratton, Orlando, 1987.

Campbell DA, Lucey MR. Liver transplantation for alcoholic cirrhosis. In: Busutill RW, Klintmalm GB, eds, *Transplantation of the liver*, p. 145. WB Saunders, Philadelphia, 1996.

Carrasco L *et al.* Effects of cold ischemia time on the graft after orthotopic liver transplantation. A bile cytological study. *Transplantation* 1996; **61**: 393–6.

Cattral MS *et al.* A pilot study of ribavirin therapy for recurrent hepatitis C virus infection after liver transplantation. *Transplantation* 1996; **61**: 1483–8.

Chung SW *et al.* Liver transplantation for hepatocellular carcinoma. *American Journal of Surgery* 1994; **167**: 317–21.

Coldwell DM *et al.* Multicenter investigation of the role of transjugular intrahepatic portosystemic shunt in management of portal hypertension. *Radiology* 1995; **196**: 335–40. [See comments.]

Cosimi AB *et al.* A randomized clinical trial comparing OKT3 and steroids for treatment of hepatic allograft rejection. *Transplantation Proceedings* 1987; **19**: 2431–3.

Cosimi AB *et al.* A randomized clinical trial comparing OKT3 and steroids for treatment of hepatic allograft rejection. *Transplantation* 1987; **43**: 91–5.

Delmonico FL *et al.* Procurement of a whole pancreas and liver from the same cadaveric donor. *Surgery* 1989; **105**: 718–23. [See comments.]

Delriviere L, Hannoun L. *In situ* and *ex situ in vivo* procedures for complex major liver resections requiring prolonged hepatic vascular exclusion in normal and diseased livers. *Journal of the American College of Surgeons* 1995; **181**: 272–6.

Demetris AJ *et al.* Pathology of liver transplantation. In: Busutill RW, Klintmalm GB, eds, *Transplantation of the liver*, pp. 681–723. WB Saunders, Philadelphia, 1996.

Dienstag JL. Acute hepatitis. In: Isselbacher KJ *et al.*, eds, *Harrison's principles of internal medicine*, p. 1465. McGraw-Hill, New York, 1994.

Dienstag JL *et al.* A preliminary trial of lamivudine for chronic hepatitis B infection. *New England Journal of Medicine* 1995; **333**: 1657–61. [See comments.]

Evans RW. Cost-effectiveness analysis of transplantation. *Surgical Clinics of North America* 1986; **66**: 603–16.

Goss JA *et al. In situ* splitting of the cadaveric liver for transplantation. *Transplantation* 1997; **64**: 871–7.

Henley KS *et al.* A double-blind, randomized, placebo controlled trial of prostaglandin E_1 in liver transplantation. *Hepatology* 1995; **21**: 366–72. [See comments.]

Imigawa DK, Busutill RW. Technical problems: vascular. In: Busutill RW, Klintmalm GB, eds, *Transplantation of the liver*, pp. 626–32. WB Saunders, Philadelphia, 1996.

Iwatsuki S *et al.* Role of liver transplantation in cancer therapy. *Annals of Surgery* 1985; **202**: 401–7.

Kahan BD. Cyclosporine. *New England Journal of Medicine* 1989; **321**: 1725–38.

Kalayoglu M *et al.* Extended preservation of the liver for clinical transplantation. *Lancet* 1988; **i**: 617–19.

Maddrey WC. *Transplantation of the liver*. Elsevier Science, New York, 1988.

Markin RS *et al.* Frozen section evaluation of donor livers before transplantation. *Transplantation* 1993; **56** 1403–9.

Mazzaferro V *et al.* Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *New England Journal of Medicine* 1996; **334**: 693–9. [See comments.]

McHutchison JG *et al.* Interferon-a-2b alone or in combination with ribavirin as initial treatment for chronic hepatitis C. Hepatitis Interventional Therapy Group. *New England Journal of Medicine* 1998; **339**: 1485–92. [See comments.]

Osorio RW *et al.* Predicting recidivism after orthotopic liver transplantation for alcoholic liver disease. *Hepatology* 1994; **20**: 105–10.

Otte JB *et al.* The concept and technique of the split liver in clinical transplantation. *Surgery* 1990; **107**: 605–12.

Penn I. Hepatic transplantation for primary and metastatic cancers of the liver. *Surgery* 1991; **110**: 726–34; discussion 734–5.

Pichlinayr R *et al.* Technique and preliminary results of extracorporeal liver surgery (bench procedure) and of surgery on the *in situ* perfused liver. *British Journal of Surgery* 1990; **77**: 21–6.

Ploeg RJ *et al.* Risk factors for primary dysfunction after liver transplantation—a multivariate analysis. *Transplantation* 1993; **55**: 807–13.

Ramsey G *et al.* Isohemagglutinins of graft origin after ABO-unmatched liver transplantation. *New England Journal of Medicine* 1984; **311**: 1167–70.

Samuel D *et al.* Passive immunoprophylaxis after liver transplantation in HBsAg-positive patients. *Lancet* 1991; **337**: 813–15.

Samuel D *et al.* Liver transplantation in European patients with the hepatitis B. *New England Journal of Medicine* 1993; **329**: 1842–7.

Shaw BW *et al.* Venous bypass in clinical liver transplantation. *Annals of Surgery* 1984; **200**: 524–34.

Starzl TE. Surgery for metabolic liver disease. In: McDermott MV, ed., *Surgery of the liver*, pp. 127–36. Blackwell Scientific, Oxford, 1989.

Starzl TE. Liver transplantation. *Gastroenterology* 1997; **112**: 288–91.

Starzl TE *et al.* Orthotopic homotransplantation of the human liver. *Annals of Surgery* 1968; **168**: 392–415.

Tarter RE *et al.* Quality of life before and after orthotopic hepatic transplantation. *Archives of Internal Medicine* 1991; **151**: 1521–6.

Terpstra OT *et al.* Auxiliary heterotopic liver transplantation. *Transplantation* 1988; **45**(6): 1003–7.

Tzakis A *et al.* Venous jump grafts for liver transplantation in patients with portal vein thrombosis. *Transplantation* 1989; **48**: 530–1.

Wall WJ. Liver transplantation: current concepts. *Canadian Medical Association Journal* 1988; **139**: 21–8.

Wendon J, Williams R. Transplantation for fulminant hepatic failure. In: Busutill RW, Klintmalm GB, eds, *Transplantation of the liver*, p. 93. WB Saunders, Philadelphia, 1996.

Williams JW *et al.* Biopsy-directed immunosuppression following hepatic transplantation in man. *Transplantation* 1985; **39**: 589–96.

16.7 Heart and heart–lung transplantation

George V. Letsou and John C. Baldwin

Heart transplantation
Introduction
Indications
Immunosuppression
Complications after heart transplantation
Late complications
Overall results
Heart–lung transplantation
History
Current status and techniques
Immunosuppression
Complications
Overall results
Further reading

Heart transplantation

Introduction

Cardiac transplantation is the treatment of choice for a variety of patients with endstage cardiac failure; it is a definitive intervention that offers many patients their only chance for dramatic improvement. The procedure has gained in popularity since the introduction of cyclosporine in the early 1980s, which has markedly improved survival rates. In 1997, more than 4000 heart transplants were performed worldwide.

Efforts at cardiac transplantation are almost a century old, beginning with the pioneering experimental efforts of Alexis Carrel and Charles Guthrie. These efforts included both canine heterotopic heart and heart–lung transplants and contributed to the receipt of the 1912 Nobel Prize in Medicine by Carrel. It was not until the 1950s, however, when Demikhov, in the Soviet Union, developed the technique of excision and reanastomosis of the heart at the atrial level that intrathoracic transplantation became feasible. He was the first to demonstrate the ability of a donor heart to support the recipient entirely by excluding the native heart from the circulation. Further advances followed the development of cardiopulmonary bypass techniques. The first successful long-term experimental ‘orthotopic' cardiac transplants (in which the native heart is excised and replaced by the donor heart) were reported in 1960 by Lower and Shumway, half a century after Carrel's work; five of eight animals survived for 6 to 21 days. They emphasized the need for excision of the donor and recipient hearts at the midatrial level, safe cardiopulmonary bypass, and topical hypothermia for cardiac preservation during excision and implantation. This experimental success set the stage for clinical human transplantation, but the concerns about operative technique, graft preservation during transfer from the donor to the recipient, the safety of cardiopulmonary bypass, immunosuppression, clinical care, post-transplant physiology, and legal issues surrounding organ donation had yet to be addressed.

The 1960s saw a steady assault on these problems. The operative technique with midatrial excision and reimplantation was further described, cardiopulmonary bypass was refined, and graft preservation techniques, particularly hypothermia, were improved. These advances led Shumway to comment in 1964 that ‘Enough has been achieved, in fact, to provoke expression of the concept that only the immunologic barrier lies between this day and a radical new era in the treatment of cardiac disease'. The first cardiac transplant into a human recipient soon followed. Hardy and colleagues in Jackson, Mississippi, performed it. Because of legal issues surrounding the concept of ‘brain death', a chimpanzee heart was transplanted into a human recipient who then expired after several hours. The first implantation of a human heart was probably performed by Lower in Richmond, Virginia, when he implanted the heart of a cadaveric kidney donor orthotopically into a baboon in 1966; this key experiment documented the ability of a transplanted human heart to function in a recipient. Barnard in Cape Town, South Africa, using techniques emphasized by Lower and Shumway, performed the first successful human orthotopic heart transplant in 1967. Barnard's operation was sensational, widely publicized, and produced a surge of enthusiasm for cardiac transplantation. However, results of the many subsequent cardiac transplants were dismal and the procedure's popularity waned quickly. Despite these initial discouraging results, Shumway persisted with the development of cardiac transplantation at Stanford through the 1970s. A dramatic improvement in survival rates resulted from refinement of recipient and donor selection, surgical technique, cardiopulmonary bypass, the diagnosis of rejection, and postoperative care during this period. The clinical introduction of cyclosporine in 1980 spurred an exponential growth in solid organ transplantation. By the mid-1980s, 1-year survival for cardiac transplantation approached 90 per cent.

Improved results and standardization of techniques have produced a dramatic resurgence of interest in cardiac transplantation as a treatment for endstage cardiac failure. The number of heart transplants now exceeds 4000 per year worldwide and is limited only by the supply of donor hearts.

Indications

Recipient selection

Appropriate recipient selection is critical for optimal results. Recipients should have endstage heart disease and a life expectancy of 6 to 12 months (i.e. New York Heart Association class III or IV). All modes of conventional medical and surgical management should have been exhausted including the many newer therapies such as angiotensin-converting enzyme inhibitors, b-blockers, automatic implantable cardiac defibrillators, dynamic cardiomyoplasty, dual cardiac chamber pacing, and possibly, left ventricular volume reduction. Recipients are usually less than 65 years of age and without other irreversible systemic illness or organ dysfunction, not including prerenal azotemia or passive hepatic congestion, which are considered reversible. There should be no evidence of infection. All recipients should be emotionally stable, have a realistic attitude towards their illness, and must be able to comply with a rigid postoperative protocol requiring daily immunosuppressive medications for the rest of their lives. These selection criteria were developed to identify the patient population most likely to survive and benefit, given the limited supply of donor organs.

Relative contraindications include recent pulmonary infarction and age above 65. Recent pulmonary infarction is associated with an increased risk of hemorrhage during the systemic anticoagulation necessary for cardiopulmonary bypass. Although patients older than 65 have not been traditionally considered ideal recipients, cardiac transplantation is being carried out with increasing frequency in patients from 65 to 75 years of age. Insulin-dependent diabetes is no longer considered a relative contraindication.

Other relative contraindications include those diseases predisposing to poor outcomes. Amyloidosis rapidly recurs in allografts and is a contraindication to transplantation. Alcoholic cardiomyopathy in patients who continue to drink carries a 25 per cent yearly mortality. Allograft coronary artery disease, which has emerged as a leading cause of late death after transplantation, does not respond well to repeat transplantation.

Preoperative assessment of pulmonary vascular resistance is essential, since elevated pressures increase perioperative mortality. Pulmonary vascular resistance is calculated during right heart catheterization (mean pulmonary artery pressure minus pulmonary capillary wedge pressure divided by the cardiac output in liter/min) and should be less than 8 Wood units, as donor right ventricles are unable to function properly against higher pulmonary vascular resistance. During right heart catheterization, pulmonary vasodilators such as nitroprusside or prostaglandin E_1 may be administered in attempts to decrease the pulmonary vascular resistance. If resistance falls to less than 8 Wood units with such vasodilators, cardiac transplantation can be performed with acceptable, but increased, morbidity and mortality if similar vasodilators are administered in the postoperative period.

Left ventricular ejection fraction is frequently used to evaluate patients with congestive heart failure. Although a decreased left ventricular ejection fraction correlates well with decreased survival in coronary artery disease, the correlation is less dramatic in non-ischemic cardiomyopathy. Ejection fractions of less than 10 per cent are associated with extremely poor survival, but ejection fractions between 10 and 30 per cent are less accurate in predicting survival; functional capacity is a better predictor of survival in patients with left ventricular ejection fractions of 10 to 30 per cent. The survival of patients with severely reduced left ventricular ejection fraction and minimal symptoms is improved with medical management and angiotensin-converting enzyme inhibitor treatment. When such patients have preserved exercise tolerance, the prognosis is relatively good. In many studies, the mortality difference between patients in the New York Heart Association classes I, II, and III varies little, while patients in class IV have a significantly worse (50 to 75 per cent yearly) mortality. Since both physician and patient interpretations bias the assessment of functional status, a more accurate and objective assessment of functional capacity is peak oxygen consumption (V_{o2}max) during stress testing. Peak oxygen consumption is dependent on both cardiac output and peripheral oxygen consumption and relates well with the functional status. V_{o2}max is used as an objective and

independent assessment of prognosis. When V_{o2}max is reduced to less than 15 ml/kg.min, predicted 1-year mortality is high and cardiac transplantation should be considered.

Other factors which correlate with poor prognosis include active coronary ischemia, right ventricular dysfunction, elevated plasma norepinephrine levels, and malignant ventricular arrhythmias. All should mandate the consideration of cardiac transplantation in otherwise appropriate patients. Although cardiac transplantation can be the only life-saving therapy for certain patients, optimal results are less commonly attained with critically ill patients. Data from the United Network for Organ Sharing reveal that ventilator dependency is a highly significant preoperative predictor of mortality and that dependency on intensive care support is also a highly significant predictor of 1-year mortality.

Donor selection

Optimal results depend on proper donor selection. Appropriate donors are usually less than 45 years of age, have been certified brain dead by an appropriate physician under local laws, have not undergone prolonged cardiopulmonary resuscitation, and have normal electrocardiograms. Screening for communicable diseases such as HIV and hepatitis is routinely performed. No blood-borne infection should be present, although many donors have evidence of pulmonary bacterial colonization. If the status of the donor heart is questionable, further evaluation with central venous pressure monitoring, echocardiogram, or even coronary angiography may be necessary. If the potential donor suffered significant chest trauma, cardiac enzyme levels should be within normal limits.

Potential donors should be within 25 per cent of the recipient's weight, although a donor/recipient weight mismatch of as much as 40 per cent can be accepted. Relatively small donors are of greater concern when the pulmonary vascular resistance is elevated and larger donors should be sought since only more muscular hearts will be able to beat sufficiently strongly against the elevated resistance of the pulmonary bed. However, recommendations concerning donor and recipient weight match must be considered in light of the fact that such weights correlate poorly with actual heart size; small donors may have large muscular hearts, particularly younger donors. Care must also be taken when transplanting hearts from female donors into male recipients, as this has been associated with increased mortality.

Donor and recipient should have compatible ABO blood groups. If a preoperative cytotoxic antibody screen using 15 to 100 random donor lymphocytes reveals more than 10 per cent to be cross-reactive, a preoperative lymphocyte cross-match is desirable. HLA matching is useful for research purposes. Retrospective analyses of grafts matched at the HLA A and B loci show slightly improved long-term survival and fewer infections, but no improvement in rejection rate, likelihood of death from rejection, or length of time to first episode of rejection. Donors and recipients matched for HLA DR antigens have fewer episodes of rejection and also a slight increase in survival. However, these increases in survival are not sufficient to warrant preoperative HLA matching.

Operative technique

Donor and recipient operations have been largely standardized.

The donor operation is performed through a median sternotomy. If multiple organs are being procured, the chest portion of the operation is performed first or simultaneously with the abdominal organ procurement dissection. After the pericardium is opened, the heart is inspected for contusions or hematomas, and is palpated to ensure the absence of valvular or coronary artery abnormalities. The aorta, superior vena cava, and inferior vena cava are encircled and controlled. Once the abdominal viscera are dissected, preparations are made to remove all organs. The heart is removed first. Heparin is administered systemically and an ascending aortic cannula is placed for cardioplegia administration. The superior vena cava is ligated and divided. The inferior vena cava and left inferior pulmonary vein are incised, decompressing both ventricular chambers and preventing their distension. The heart is allowed to contract several times and empty completely. An aortic cross-clamp is then placed high on the ascending aorta, just proximal to the innominate artery, and cold crystalloid cardioplegia is infused via the ascending aortic cannula. Preservation is supplemented with topical cold saline. The heart is excised by dividing the aorta just proximal to the aortic cross-clamp; the pulmonary veins and arteries are divided at the pericardial reflection. The heart is packed in sterile saline at 4°C for transport. The other abdominal organs are subsequently harvested.

The recipient operation is also performed through a median sternotomy. Proper timing of the recipient operation is important, as this will minimize the ischemic time of the donor organ; the recipient should be ready for implantation just as the donor heart enters the operating theater. The use of aprotinin in redo operations can be important in reducing blood loss in these patients, who often have right-sided failure and an element of hepatic insufficiency. Once the median sternotomy has been completed, the aorta, superior vena cava, and inferior vena cava are encircled and controlled. After systemic heparinization, separate inferior and superior vena cava cannulas are placed. These cannulas are placed quite laterally on the right atrium to ensure a good cuff of native right atrium for anastomosis to the donor heart. The aorta is cannulated at the base of the innominate artery. Once the donor heart has arrived safely, cardiopulmonary bypass is initiated. The recipient is cooled systemically to 32°C. The aorta is occluded with a cross-clamp and caval snares are applied. The native heart is excised at the midatrial level; the right atrium is entered anteriorly at the atrioventricular groove and the incision extended to the coronary sinus. The aorta is then incised just above the aortic valve. The pulmonary artery is similarly divided at the supra-annular level. The left atrium is entered below the aortic valve, the atrial septum divided, and the native heart removed by dividing the left atrium's posterior wall just below the mitral valve. The donor heart is brought on to the operative field and inspected. It is important to exclude the presence of a patent foramen ovale. Interconnecting the pulmonary vein orifices opens the left atrium. The aorta and pulmonary artery are separated. The heart is properly oriented and placed in the recipient's chest cavity. The left atrial anastomosis is performed first, with a running polypropylene suture. Many surgeons enhance myocardial protection using a saline solution at 4°C infused through a left atrial catheter placed via the left atrial appendage as well as dripping a similar solution on to the new heart. The right atrial anastomosis is completed using running suture. A running suture is used to anastomose donor and native aorta. The pulmonary artery anastomosis is completed in similar fashion. The heart is de-aired and the cross-clamp removed. The pulmonary artery and right ventricle are decompressed and vented for at least 20 min prior to weaning the patient from cardiopulmonary bypass (allowing controlled reperfusion of the new heart). Because the heart is denervated, a slow heart rate is not unusual; isoproterenol is often used to bring the heart rate to between 100 and 110 beats/min. Resuscitation with pressor support completes the procedure.

Yacoub has described an alternative technique for ‘total orthotopic cardiac transplantation'. This method involves a more extensive recipient cardiectomy leaving cuffs of superior and inferior vena cava and right and left pulmonary veins. These structures are anastomosed to the donor heart along with the standard great vessel anastomosis. Proponents of this technique describe a decreased incidence of pacemaker dependence and atrioventricular valve dysfunction, although long-term functional and hemodynamic advantages are not yet clear.

A less common option for heart transplantation is ‘heterotopic' transplantation. In this technique, the recipient heart is left *in situ* and the donor heart is used to augment cardiac output. The procedure is employed for fixed high pulmonary vascular resistance or significant donor/recipient size mismatch. Approximately 40 heterotopic procedures are performed each year. Complications include compression atelectasis of the lung, thromboembolic events from the recipient ventricles, and malignant arrhythmias originating in the native heart.

The postoperative management of cardiac transplant recipients is similar to that of any patient undergoing open-heart surgery. Extubation can usually be accomplished within 6 to 2 h. Early mobilization and ambulation are especially important for pulmonary well being. For severely debilitated patients who have had prolonged waits for suitable organs, a period of reconditioning may be necessary, but should not hinder early mobilization. Usually, isoproterenol is administered intravenously for 4 to 5 days for its chronotropic effects. After 5 to 7 days, if the heart rate does not return spontaneously to 80 beats/min or greater, a permanent pacemaker is inserted. The typical stay in an intensive care unit for a recipient with no complications ranges from 2 to 5 days. Oral nutrition usually can begin within 48 h of surgery. Specialized protection against infection is not necessary. Good hand washing, common sense, and prevention of contact with family members with infections, especially children, is sufficient. A heart transplant recipient with no complications can be discharged from the hospital as early as day 5 to 7.

Immunosuppression

Many immunosuppressant regimens are currently in use. Most centers use at least three agents. Cyclosporine, a fungal metabolite, was introduced clinically in 1980 and is the mainstay of most regimens. Immediately before the operation, 2 to 8 mg/kg are administered; a lower dose should be used when renal insufficiency is present. The drug is continued indefinitely, maintaining a level of 200 to 300 ng/ml as measured by high-pressure liquid chromatography. Nephrotoxicity is a common problem and renal function must be closely monitored.

Azathioprine has been used in all forms of transplantation as an antirejection treatment since the early 1960s. Currently, 2 to 5 mg/kg are given preoperatively, followed by a maintenance dose of 1 to 2 mg/kg.day postoperatively. The white blood cell count mustbe monitored: should it precipitously decrease or drop below 5000 cells/l, azathioprine should be decreased in dosage or discontinued.

Corticosteroids are the third arm of most cardiac transplant immunosuppressive protocols. Many institutions give 500 mg of intravenous methylprednisolone on completion of the aortic anastomosis, and continue intravenous methylprednisolone for 24 h at a dose of 125 mg every 8 h. After 24 h, oral prednisone is started at a dose of 0.6 mg/kg.day and then tapered slowly over several weeks to a maintenance dose of 10 mg each day. Some institutions have successfully abandoned the use of steroids in maintenance immunosuppression, especially for children, because of their effects on growth and development.

Additional immunosuppression with a fourth agent, such as the monoclonal antibody OKT3, is controversial. The protocol may be beneficial in patients with a high likelihood of rejection (such as those patients with a high PRA (percent reactive antibody)). Such four-drug regimens have decreased the frequency of rejection without increasing the likelihood of infection. However, no effect on survival has been documented.

Other immunosuppressants are being more frequently used. Tacrolimus (FK506) is a macrolide antibiotic with potent immunosuppressive properties. Originally used primarily in hepatic transplantation, it is being used with increasing frequency in thoracic transplantation. The mechanism of action is similar to that of cyclosporine. Initially, tacrolimus was used primarily as a rescue therapy in cases of refractory rejection. Recent trials have documented its efficacy as a first-line immunosuppressive agent comparable with cyclosporine. Mycophenolate mofetil is another recently released immunosuppressant. It blocks T- and B-cell lymphocyte replication through inhibition of the enzyme inosine monophosphate dehydrogenase and is used primarily for cases of intractable rejection.

Other modalities of immunosuppression include photochemotherapy, used for refractory rejection, and plasmapheresis, used for both intractable rejection and prophylactically for recipients at high risk of rejection.

Complications after heart transplantation

Technical errors, graft failure, and right ventricular failure are the most common early complications. Although these problems rarely lead to death, they are unusual problems in cardiac surgery. The early manifestations of these problems can go unrecognized.

Preoperative evaluation of heart transplant recipients includes evaluation of pulmonary vascular resistance and severe, fixed pulmonary hypertension is a contraindication to cardiac transplantation. However, the response of pulmonary vascular resistance to cardiac transplantation and increased cardiac output from the donor graft is difficult to predict and right-sided pressures (central venous pressure and pulmonary artery pressure) are commonly elevated in the early post-transplant period. Right ventricular distention must be avoided when weaning from cardiopulmonary bypass and in the first days after transplantation. Graft ischemia, postoperative tricuspid regurgitation, and coronary air embolus all contribute to right ventricular dysfunction and must be minimized. Management of right ventricular failure consists of vasodilator and inotropic agents. Prostaglandin E_1, a relatively selective pulmonary vasodilator, may be infused directly into the pulmonary vasculature. Inhaled nitrous oxide is a much more selective pulmonary vasodilator, which is becoming more widely available and may be more efficacious. These interventions can be started preoperatively in high-risk patients. Right ventricular assist devices may be required in refractory cases.

Acute left ventricular dysfunction may result from prolonged graft ischemia, poor myocardial protection, unrecognized graft injury, infarction, or rarely, hyperacute rejection. Treatment is inotropic and vasodilator support, intra-aortic balloon counterpulsation, left ventricular assist device support, or ultimately, retransplantation.

Technical complications after orthotopic heart transplantation include hemorrhage, pulmonary artery stenosis, and valvular incompetence. Hemorrhage is best controlled prophylactically through meticulous attention to surgical detail; patients who are immunosuppressed tolerate hematomas poorly and tend to become infected. Pulmonary artery stenosis is due to miscalculation of the length or orientation of the pulmonary artery anastomosis and is manifested by right ventricular failure, which may be confused with pulmonary hypertension. Mild atrioventricular valve regurgitation is common due to distortion of the annulus from the biatrial anastamosis. It is usually mild and self-limited.

Pericardial effusions are common, result from the common situation of a large failing heart being replaced by a stronger but smaller organ, and usually resolve with expectant management. They may be followed by echocardiography or aspirated by pericardiocentesis when symptomatic. Constrictive pericarditis is unusual and rarely requires pericardiectomy.

Late complications

In the first months following transplantation, rejection and infection are the most common complications: rejection accounts for 30 per cent of all deaths in cardiac transplant recipients and infections for 20 per cent. Seventy per cent of cardiac transplant recipients have an episode of rejection within the first 3 months after transplantation. After 1 year, the incidence of rejection decreases and less than 10 per cent of transplant recipients per year will have rejection episodes. The majority of rejection episodes respond to appropriate treatment. Approximately 50 per cent of recipients have an infection in the first year. The manifestations of severe systemic infection are often subtle, and monitoring for infection as well as for rejection is of paramount importance.

Rejection was difficult to diagnose in the early years of cardiac transplantation, when a variety of criteria were used, including loss of QRS voltage, occurrence of supraventricular arrhythmias, and clinical signs of low cardiac output. The technique of endomyocardial biopsy revolutionized cardiac transplantation, allowing early diagnosis of rejection, treatment, and improved survival. In the early postoperative period, endomyocardial biopsies are performed weekly. After 1 to 2 months, the frequency can be decreased. Usually, access to the right ventricular septal biopsy site is via the right internal jugular vein. Under fluoroscopic guidance, biopsy forceps are maneuvered to the septal region and at least four samples are obtained for pathologic analysis. Right ventricular pressures and cardiac output are measured. An international grading system has been developed to grade rejection (Table 1). The hallmark of rejection requiring treatment is myocyte necrosis. An infiltrate of lymphocytes is indicative of lesser rejection; myocyte necrosis with extramyocardial hemorrhage is indicative of severe rejection and damage (Fig. 1).

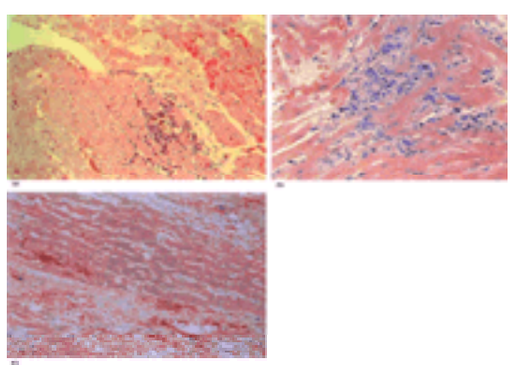


Fig. 1. Photomicrographs of endomyocardial biopsies. (a) Myocardial biopsy showing mild rejection. Note the perivascular infiltrates of lymphocytes. (b) Endomyocardial biopsy showing a moderate rejection with evidence of myocyte necrosis. Note the clusters of lymphocytes throughout the myocardial tissue with areas of myocyte destruction. (c) Endomyocardial biospy specimen showing borderline severe rejection with a diffuse infiltrate of lymphocytes throughout the myocardium and in one area in the upper left-hand corner of the myograft showing diffuse and widespread myocyte destruction.

No rejection	0	No rejection	0
Focal perivascular or interstitial infiltrate without necrosis	1A	Mild rejection	1
Diffuse but sparse infiltrate without necrosis	1B	Mild rejection	2,3
One focus only with aggressive infiltration and/or focal myocyte damage	2	Focal moderate	4,5
Multifocal aggressive infiltrate and/or myocyte damage	3A	Low moderate	6,7
Diffuse inflammatory process with necrosis	3B	Borderline severe	8,9
Diffuse aggressive polymorphous infiltrate, ± edema, ± hemorrhage, ± vasculitis, with necrosis	4	Severe acute	10

Table 1 A comparison of commonly used grading systems for rejection

If myocyte necrosis is detected, antirejection therapy should be begun immediately. Initial treatment at most centers is 1 g of methylprednisolone intravenously for 3

days. Four days after completion of this ‘steroid pulse', endomyocardial biopsy is performed to assess treatment results. If significant rejection persists, monoclonal or polyclonal antibodies are administered. If rejection persists despite all treatments and progresses to hemodynamic instability, retransplantation should be considered. Since the introduction of cyclosporine such refractory rejection has become uncommon.

Diagnosis of infection is difficult because immunosuppression obscures the usual manifestations, and findings are often subtle. Once infection is established, it can progress very rapidly and be extremely aggressive; prompt treatment is necessary. An oral temperature over 38°C must be considered presumptive evidence of infection despite the absence of clinical signs. Such a fever is an indication for thorough culturing and aggressive evaluation.

Graft atherosclerosis is the most important late complication. It is felt to be a manifestation of chronic rejection based on immunologic incompatibility, and manifests primarily as diffuse coronary artery narrowing. The distribution of coronary artery lesions is atypical in that most are distal lesions not amenable to treatment using conventional bypass or angioplasty techniques. Graft atherosclerosis is also unusual in that coronary artery lesions often develop extremely rapidly over several months. Association with cytomegalovirus infection has been documented, but antiviral treatment directed at this agent has not yet been shown to lessen the incidence of graft atherosclerosis. Retransplantation is the only therapeutic option.

Overall results

Cardiac transplantation is extremely effective for the treatment of endstage cardiac failure. The Registry of the International Society for Heart and Lung Transplantation reported that 3471 cardiac transplants were performed in 1997 (Fig. 2). The age distribution of these recipients is shown in Fig. 3. The overall actuarial survival at 1 year is 79 per cent and 5-year actuarial survival is 69 per cent (Fig. 4 and Fig. 5). Survival rates in excess of 90 per cent at 1 year can be anticipated in younger patients with idiopathic or postpartum cardiomyopathies. Cardiac transplantation can no longer be considered experimental and should be offered to all patients who are appropriate candidates.

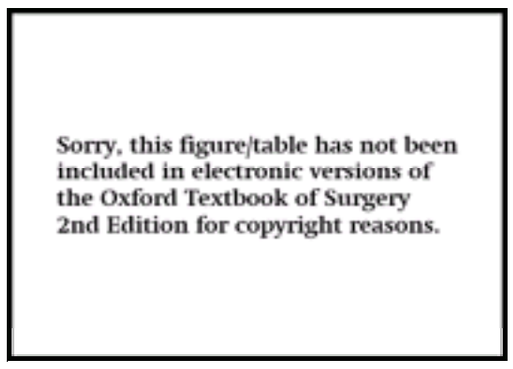


Fig. 2. Heart transplantation volumes and donor age by year (International Society for Heart and Lung Transplantation Registry).

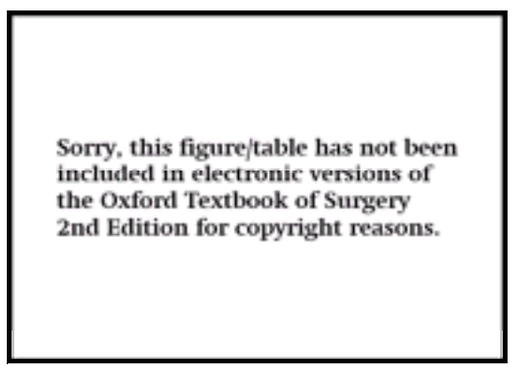


Fig. 3. Age distribution of heart transplant recipients (International Society for Heart and Lung Transplantation Registry).

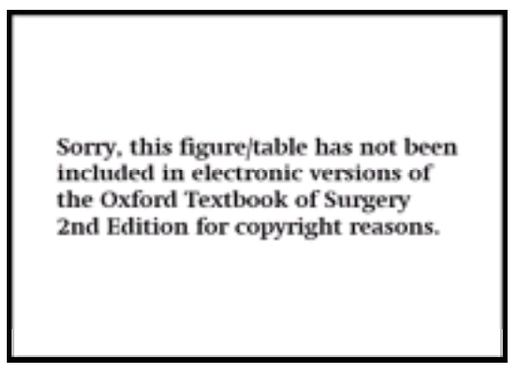


Fig. 4. Total heart transplantation actuarial survival (International Society for Heart and Lung Transplantation Registry).

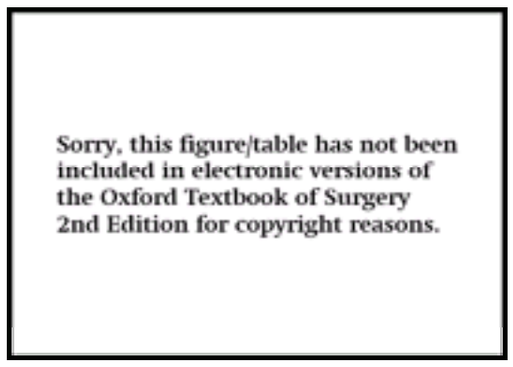


Fig. 5. Adult heart transplantation actuarial survival by era (International Society for Heart and Lung Transplantation Registry).

Heart–lung transplantation

History

The first attempts at cardiopulmonary transplantation date to the initial experiments of Carrel and Guthrie with cardiac transplantation. They performed heterotopic cardiopulmonary transplants in rats as early as 1905. Lower and Shumway performed canine cardiopulmonary transplants in conjunction with orthotopic cardiac

transplants in the 1950s and early 1960s. Canine cardiopulmonary transplants were complicated by the fact that dogs, unlike primates, require intact pulmonary innervation for normal respiratory function: subsequent experiments showed that primates could function normally with denervated lungs. The first human heart–lung transplant was performed by Cooley in 1968 in an infant who died after 14 h. With the development of successful cardiac transplantation at Stanford, further experimental work on cardiopulmonary transplantation as a means to circumvent the problem of elevated pulmonary vascular resistance resulted in the first successful clinical heart–lung transplantation program in 1981. Since that time, cardiopulmonary transplantation has been initiated at many other centers. However, the donor pool for such transplants is small, as both a normal heart and nearly perfect lungs are required. Thus, cardiopulmonary transplantation has not been as widely applied as cardiac transplantation.

Current status and techniques

Recipient selection

As in cardiac transplantation, appropriate recipient selection is critical. Recipients should be severely limited by their pulmonary and/or cardiac disease with a life expectancy of 6 to 12 months. All modes of conventional medical and surgical management should have been exhausted. However, transplantation of patients with endstage disease requiring mechanical ventilation has rarely been successful. Generally, recipients should be less than 45 year of age and without other irreversible systemic illnesses or organ dysfunction, not including prerenal azotemia or passive hepatic congestion, which are considered reversible. Many potential recipients, especially those with cystic fibrosis, have evidence of pulmonary bacterial colonization, but there should be no active systemic infection. Mental stability is very important; all patients should be able to comply with a rigid postoperative immunosuppression protocol.

Previous formal posterolateral thoracotomy is a contraindication to transplantation; this is associated with excessive bleeding after native heart–lung excision, which is difficult to control after implantation of the graft. More limited thoracotomies or previous open lung biopsy are relative contraindications. Recent pulmonary infarction is also associated with an increased risk of hemorrhage and is a relative contraindication. Insulin-dependent diabetes mellitus is no longer a relative contraindication for the same reasons as in cardiac transplantation.

These criteria allow a substantial population to qualify for cardiopulmonary transplantation. There were 151 heart–lung transplants reported to the Registry of the International Society for Heart and Lung Transplantation in 1997. As reported to the Registry, 26 per cent of heart–lung recipients had primary pulmonary hypertension and Eisenmenger's syndrome accounted for 28 per cent (this includes many patients ineligible for orthotopic cardiac transplant because of combined endstage left ventricular failure and elevated pulmonary vascular resistance). Cystic fibrosis accounted for 16 per cent of cardiopulmonary transplants and emphysema for 4 per cent (Fig. 6).

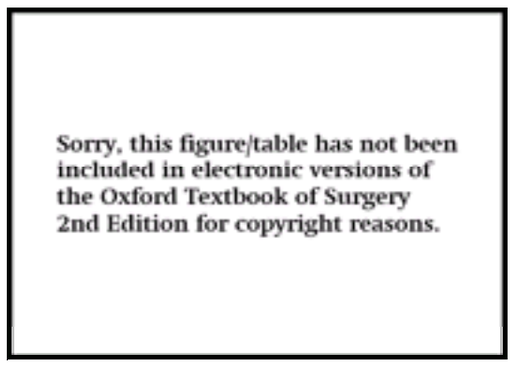


Fig. 6. Heart–lung transplant indications. PPH, primary pulmonary hypertension; AIA, a_1-antitrypsin; CF, cystic fibrosis; ReTX, retransplantation; IPF, idiopathic pulmonary fibrosis.

Donor selection

Criteria for the selection of appropriate donors are similar to those for cardiac transplantation. In addition, heart–lung transplant donors must have a P_{o2} greater than 100 mmHg on 40 per cent inspired oxygen, normal lung compliance with a peak inspiratory pressure less than 30 to 35 mmHg, no major thoracic trauma, and a clear chest radiograph. Any history of smoking warrants careful consideration. There should be no evidence of pulmonary bacterial colonization on sputum Gram stain or culture. A history of lung disease is unacceptable. These criteria are fairly strict due to problems inherent in pulmonary preservation and severely limit the donor pool. However, such criteria are essential if good clinical results are to be obtained.

Operative technique

Donor operation

Since pulmonary preservation is the limiting factor in cardiopulmonary transplantation, the donor operation assumes paramount importance. Throughout the donor procedure, the lungs must be carefully protected. At no time should inspired oxygen be greater than 40 per cent. If it is necessary to increase inspired oxygen above 40 per cent to maintain a P_{o2} greater than 100 mmHg, consideration should be given to aborting the transplant. With the advent of multiorgan procurements, good communication between all procurement teams and the anesthesiologist is essential. As little fluid as possible should be administered intravenously, as the lungs will not be able to clear fluid easily after transplantation. Heart–lung procurement is performed via median sternotomy. The pericardium is opened, the heart inspected and palpated, both anterior pleura are incised, and the lungs are inspected for contusion or hematoma. The aorta and both vena cavas are encircled and controlled. The mainstem trachea is controlled between the superior vena cava and the ascending aorta. After completion of the abdominal dissection, preparations are made to excise the heart–lung block. After systemic heparinization, a cannula for administration of cardioplegia is placed in the ascending aorta. A cannula is placed in the pulmonary artery for the administration of ‘pulmoplegia' (we use a modified Collins' solution with magnesium sulfate and dextrose added) after aortic occlusion. Just prior to aortic occlusion, prostaglandin E_1 is administered intravenously in a slowly increasing dose from 20 to 150 ng/kg.min. The superior vena cava is then ligated and divided. The inferior vena cava and left atrial appendage are incised, decompressing both ventricular chambers. The heart is allowed to contract several times and empty completely. The aorta is occluded with a cross-clamp just below the innominate artery. Cold crystalloid cardioplegia is infused through the ascending aortic root cannula and pulmoplegia is administered via the pulmonary artery cannula while gentle lung ventilation is continued, ensuring adequate pulmoplegia distribution. Preservation is supplemented with topical cold saline. The heart–lung block is excised beginning in the posterior mediastinum just anterior to the esophagus. The dissection is carried up the esophagus from the diaphragm to the mainstem trachea. The aorta is divided just below the cross-clamp. The lungs are inflated to half their total volume and a stapling device placed across the mainstem trachea at least two cartilaginous rings above the carina. The heart–lung block is transported to the recipient at 4°C.

Recipient operation

A median sternotomy is used for the recipient operation. Safe removal of the native organs can be a technical challenge. Great care must be taken to protect phrenic, recurrent laryngeal, and vagus nerves. The aorta is cannulated at the base of the innominate artery; separate inferior and superior vena cava cannulas are placed. After the donor organs have arrived safely, cardiopulmonary bypass is initiated. Atria are divided at the midatrial level, the aorta is divided at the supra-annular level, and the main pulmonary artery is divided. The native heart is removed. Pericardial pedicles preserving the phrenic nerves are created. A 3-cm cuff of pericardium is left anterior to the phrenic nerves, and the posterior incision for the phrenic pedicle is made just anterior to the pericardial reflection of the pulmonary veins. An incision is made in the remaining left atrium to separate the pulmonary veins from the posterior mediastinum. Care must be taken to avoid vagus nerve injury. After division of the left pulmonary ligament, bronchial collaterals to the native lung are divided and the bronchus is transected using a stapling device. The left lung is removed. The right lung is explanted in a similar fashion. At the procedure's conclusion, two pedicles containing the uninjured phrenic nerves remain.

The donor heart–lung block is then prepared for implantation. The tracheal staple line is excised and specimens of donor trachea are obtained for culture. A curvilinear incision is made in the right atrium, avoiding sinoatrial node injury. The recipient trachea is divided two or three rings above the carina in preparation for anastomosis to the donor, which is performed end-to-end using a running polypropylene suture. Right atrial and aortic anastomoses are performed as in isolated cardiac transplantation. The new organs are reperfused for 20 to 30 min with the pulmonary artery decompressed and vented. Weaning from cardiopulmonary bypass is performed with resuscitation of the new heart and lungs, using pressors as necessary. Care must be taken to avoid administration of oxygen in excess of 40 per cent as the newly implanted lungs are exquisitely sensitive to oxygen. Early extubation is desirable. Surgical interruption of pulmonary lymphatic drainage makes rigorous

fluid restriction, early patient mobilization, and meticulous attention to respiratory care essential in the early postoperative period.

Immunosuppression

Immunosuppressant regimens are similar to those used in cardiac transplantation. However, following the initial dose of methylprednisolone, steroids are administered for only 24 h. No further steroids are given for 3 weeks, while tracheal healing progresses. After 3 weeks, oral prednisone at 0.5 mg/kg.day is begun and slowly tapered to 10 mg/day.

As in other solid organ transplants, cyclosporine is the mainstay of treatment. It is administered prior to the operation in a dose of 2 to 8 mg/kg and then continued indefinitely in doses adequate to maintain a level of 200 to 300 ng/ml by high-pressure liquid chromatography. Recipients must be monitored closely for nephrotoxicity.

Azathioprine is administered in a fashion similar to that in isolated cardiac transplantation: 2 to 5 mg/kg is given preoperatively, maintenance dosing is maintained at 1 to 2 mg/kg.day while the white blood cell count is monitored. During the initial postoperative period most centers administer another agent, such as rabbit antilymphocyte globulin, for additional protection in the period when steroids are not administered. Administration of OKT3 is not advocated because of its association with pulmonary edema.

Complications

Technical complications are more common following cardiopulmonary transplantation than after cardiac transplantation. Operative mortality for heart–lung transplantation is still in excess of 10 per cent, as reported to the Registry of the International Society for Heart and Lung Transplantation in 1997.

Thirty per cent of deaths after cardiopulmonary transplantation can be attributed to technical causes, primarily bleeding and phrenic nerve injury. As experience increases, these complications are diminishing. Rejection and infection are common, as in isolated cardiac transplantation. Twenty per cent of deaths are due to rejection and 30 per cent to infection.

Cardiac and pulmonary rejection may occur simultaneously or asynchronously. Cardiac rejection is less common than in patients undergoing isolated cardiac transplantation, although the reasons for this are unclear. Routine endomyocardial biopsy is therefore performed less frequently: usually one or two endomyocardial biopsies are performed in the first months and subsequently only as clinically indicated. Cardiac rejection is graded using the previously described scales and is treated in the same fashion as isolated cardiac transplants. Pulmonary rejection is difficult to diagnose accurately. Although routine transbronchial endoscopic biopsies are performed at some centers, these have not yet demonstrated the same reliability as endomyocardial biopsies. Other techniques to detect rejection include broncheoalveolar lavage and radionuclide perfusion scanning; neither detects rejection reliably. The most reliable indicator of pulmonary rejection currently available is the lung's appearance on chest radiograph, taken in conjunction with clinical suspicion of rejection and pulmonary function testing. If infiltrates develop that are not responsive to antibiotics or antiviral therapies, rejection must be considered and treated presumptively. Despite aggressive treatment of pulmonary rejection, bronchiolitis obliterans, a form of chronic rejection resulting in severely diminished lung function and a pattern of increased interstitial markings on chest radiograph, remains a formidable problem.

As in isolated cardiac transplantation, the diagnosis of infection is made with aggressive evaluation of any fever greater than 38°C orally. Signs of infection are often subtle and appropriate treatment can only be instituted early if attempts at diagnosis are prompt and thorough.

Overall results

Combined heart–lung transplantation is an effective treatment of primary pulmonary hypertension, Eisenmenger's syndrome, and cystic fibrosis. A survival rate is excess of 70 per cent can be expected at 1 year; the 5-year survival rate is approximately 60 per cent.

Further reading

Baldwin JC *et al.* Distant graft procurement for combined heart and lung transplantation using pulmonary artery flush and simple topical hypothermia for graft preservation. *Annals of Thoracic Surgery* 1987; **43**: 670–3. [Prostaglandin E is an effective, relatively selective pulmonary vasodilator permitting distant graft procurement.]

Barnard CN. The operation. A human cardiac transplant: an interim report of a successful operation performed at Groote-Schuur Hospital, Cape Town. *South African Medical Journal* 1967; **41**: 1271–4. [The first successful human heart transplant.]

Billingham ME. The postsurgical heart: the pathology of cardiac transplantation. *American Journal of Cardiovascular Pathology* 1988; **1**: 319–34. [A review of the pathology of cardiac rejection by a leader in the field.]

Bourge RC, Naftel DC, Constanzo-Nordin MR, *et al.* Pertransplantation risk factors for death after heart transplantation: a multicenter study. *Journal of Heart Lung Transplantation* 1993; **12**: 549–62. [Critically ill patients far worse than healthier patients. Donor age is not as important.]

Burke CM *et al.* Twenty-eight cases of human heart–lung transplantation. *Lancet* 1986; **i**: 517–19. [A large series of combined heart-lung transplants.]

Deleuze PH, Benvenuti C, Mazzucotelli JP, *et al.* Orthotopic cardiac transplantation with direct caval anastomosis: is it the optimal procedure? *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 731–7. [A series supporting the use of bicanal anastomoses over the traditional atrial anastomoses.]

Frist WH, Oyer PE, Baldwin JC, Stinson EB, Shumway Ne. HLA compatibility and cardiac transplant recipient survival. *Annals of Thoracic Surgery* 1987; **44**: 242–6. [HLA compatibility improves survival, but not dramatically.]

Glanville Ar, Imoto E, Baldwin JC, Billingham ME, Theodore J, Robin ED. The role of right ventricular endomyocardial biopsy in the long-term management of heart–lung transplant recipients. *Journal of Heart Transplantation* 1987; **6**: 357–61. [A classic paper by the pioneering group in endomyocardial biopsy.]

Grattan MT, Moreno-Cabral CE, Starnes VA, Oyer PE, Stinson EB, Shumway NE. Cytomegalovirus infection is associated with cardiac allograft rejection and atherosclerosis. *Journal of the American Medical Association* 1989; **261**: 3561–6. [Cytomegalovirus infection is a risk factor for cardiac rejection.]

Higenbottam T, Stewart S, Penketh A, Wallwork J. Transbronchial lung biopsy for the diagnosis of rejection in heart–lung transplant patients. *Transplantation* 1988; **46**: 532–9. [Transbronchial biopsy is an effective way to diagnose lung rejection.]

Hunt SA, Schroeder JS. Managing patients after cardiac transplantation. *Hospital Practice* 1989; **24**: 83–100. [An extensive discussion of practical management issues after heart transplantation.]

Kirklin JK, Naftel DC, Kirklin JW, Blakstone EH, White-Williams C, Bourge RC. Pulmonary vascular resistance and the risk of heart transplantation. *Journal of Heart Transplantation* 1988; **7**: 331–6. [One of the surprisingly few papers addressing the issue of increased mortality as pulmonary vascular resistance increases.]

Kreitt JM, Kaye MP. The Registry of the International Society for Heart Lung Transplantation: eighth official report—1991. *Journal of Heart and Lung Transplantation* 1991; **10**: 491–8. [A report of overall survival after heart and heart-lung transplantation.]

Lower RR, Shumway NE. Studies on orthotopic transplantation of the canine heart. *Surgery Forum* 1960; **11**: 18–29. [The classic experimental report of successful heart transplantation.]

Naeije R, Lipski A, Abramowicz M, *et al.* Nature of pulmonary hypertension in congestive heart failure: effect of cardiac transplantation. *American Journal of Respiratory and Critical Care Medicine* 1994; **149**: 881–7. [An evaluation of pulmonary hypertension response to successful cardiac transplantation.]

Opelz G for the Collaborative Transplant Study. The influence of HLA compatibility on graft survival after heart transplantation. *New England Journal of Medicine* 1994; **330**: 816–19. [HLA compatibility improves cardiac graft survival.]

Pfeffer PF, Foerster A, Froysaker T, Simonsen S, Thorsby E. HLA-DR mismatch and histologically evaluated rejection episodes in cardiac transplant can be correlated. *Transplant Proceedings* 1988; **20**: 367–8. [HLA or locus is more important than other HLA loci in determining graft survival.]

Renlund DG, O'Connell JB, Gilbert EM, Watson FS, Bristow MR. Feasibility of discontinuation of corticosteroid maintenance therapy in heart transplantation. *Journal of Heart Transplantation* 1987; **6**: 71–8. [Steroids can be safely discontinued after cardiac transplantation with careful monitoring, particularly in children.]

Shumway SJ, Shumway ME, eds. *Thoracic Transplantation*. Blackwell Science, Oxford, 1995. [A volume devoted to all aspects of transplantation.]

Smith, JA, McCarthy PM, Sarris GE, Stinson EB, Reitz BE, eds. *The Stanford manual of cardiopulmonary transplantation.* Futura Publishing Co., Armonk, NY, 1996. [A practical manual.]

16.8 Small-bowel transplantation

D. L. Bigam and D. R. Grant

Intestinal failure
History of small-bowel transplantation
Indications and recipient selection
Donor selection and the procurement procedure
Recipient surgery
Isolated intestinal transplantation
Small-bowel/liver transplantation
Multivisceral transplantation
Post-transplant management
Rejection
Infection
Nutrition
Complications
Outcome
Graft function
Patient and graft survival
Conclusions and future prospects
Further reading

Intestinal failure

Intestinal failure is the inability to maintain nutrition and/or fluid and electrolyte balance, resulting in malnutrition, weight loss, and dehydration. It can be caused by mechanical or functional deficits (Table 1). The short-bowel syndrome, which results from massive bowel resection, is the most common cause of intestinal failure. Functional disorders include Crohn's disease, radiation enteritis, microvillous inclusion disease, and pseudo-obstruction among others.

	Mechanical (short bowel)	*Functional*
Children	Gastroschisis Intestinal atresia Malrotation/midgut volvulus Necrotizing enterocolitis	Idiopathic pseudo-obstruction Microvillous inclusion disease Intestinal aganglionosis Visceral myopathy Autoimmune enteritis
Adults	Locally invasive neoplasm Mesenteric infarction Trauma	Crohn's disease Radiation enteritis Idiopathic pseudo-obstruction

Table 1 Aetiology of intestinal failure

Initial treatment for intestinal failure is aimed at maintaining hydration and nutrition with total parenteral nutrition until the patient can be sustained on an enteral diet. Irrespective of the length of the remaining small bowel, oral nutrition is always provided to promote mucosal adaptation and improve the gut barrier function. Patients are started on iso-osmolar tube feeds containing glutamine, an essential gut nutrient. Medium-chain triglycerides can be given to provide extra calories. Narcotics, cholestyramine, histamine antagonists, clonidine, and/or somatostatin analogues may be used to decrease the loss of luminal fluid and slow intestinal transit. Some patients benefit from surgical procedures that increase the absorptive capacity of the remaining bowel.

Massive resection of the small bowel after intra-abdominal catastrophe is the most common cause of the short-bowel syndrome. Adults with more than 60 cm of terminal ileum or 30 cm of terminal ileum plus an intact ileocecal valve can usually be weaned off parenteral nutrition, owing to adaptation of the remaining small bowel. The need for lifelong parenteral nutrition in children is more difficult to predict as some are able to adapt and resume normal oral nutrition with as little as 10 to 20 cm of small bowel.

Most patients who require long-term parenteral nutrition at home remain in good health, but this therapy is expensive and it interferes with normal daily activities. Moreover, some cannot tolerate the parenteral-nutrition solutions and they develop severe cholestatic liver disease. Others eventually lose their venous access sites because of line infections and thrombosis.

The complications faced by patients who require long-term parenteral nutrition have stimulated an interest in the development of small-bowel transplantation as a potential cure for intestinal failure. Unfortunately, it has proved more difficult to transplant the small bowel than other solid organs. Problems unique to bowel transplantation include (i) the large number of immunocompetent donor lymphocytes transplanted, (ii) microbial colonization of the graft, and (iii) the lack of a reliable, non-invasive method for diagnosing rejection.

The intestine provides a strong stimulus for rejection because of the constitutive expression of MHC class II antigens on the intestinal epithelium and lymphocytes within the graft. Because of the need for intense immune suppression to prevent rejection, recipients are particularly susceptible to infections. In addition, infectious complications may develop secondary to bacterial translocation from the graft when the gut barrier function is compromised by preservation injury or graft rejection. These problems have led to a slower evolution of small-bowel transplantation compared with other organ transplants.

History of small-bowel transplantation

The first report of experimental small-bowel transplantation was in 1902, when Dr Alexis Carrel successfully transplanted segments of canine small bowel to the great vessels of the neck. In the late 1950s, Lillehei demonstrated that canine small-bowel allografts could provide normal nutrition, despite the lymphatic and neural disruption that accompanies this procedure. The success of renal transplantation in the 1960s stimulated attempts at bowel transplantation in humans, but these procedures were uniformly unsuccessful because of rejection, sepsis and graft failure. In the early 1970s, total parenteral nutrition became available as a treatment for patients with intestinal failure. Interest in intestinal transplantation was revived in the late 1980s after reports of successful, isolated small-bowel transplantation and combined liver/small-bowel transplantation with cyclosporin immunosuppression. These cases were followed by successful, isolated intestinal grafting using tacrolimus (Prograf®). Now, there are more than 30 programs throughout the world performing about 100 intestinal transplants per year.

Indications and recipient selection

The initial step in assessing potential recipients for intestinal transplantation is to confirm that the standard treatments for gut failure, including intensive attempts at enteral feeding, have truly failed. A barium swallow is ordered to visualize the extent of remaining bowel. Doppler studies are performed to assess the abdominal vasculature. Studies of gastric emptying are sometimes helpful in patients with pseudo-obstruction to confirm that there will be adequate intestinal motility into the bowel graft. Patients must also undergo an assessment of their cardiopulmonary and renal status, as well as a psychosocial evaluation.

Small-bowel transplantation has considerable risk, so it is currently reserved for patients with intestinal failure who have failed on conventional therapy. Those on long-term parenteral nutrition who have lost venous access due to repeated catheter infections and/or thrombosis are candidates for isolated intestinal grafting. Those who develop cholestatic liver disease secondary to parenteral nutrition are candidates for combined liver/small-bowel transplant. Some patients who have undergone massive resection of the small bowel because of mesenteric venous thrombosis may benefit from a combined transplant with a new liver that can correct the underlying coagulopathy, such as occurs with protein-C deficiency. With the improved success rate of isolated small-bowel transplantation, there is a move towards earlier transplantation in an effort to avoid parenteral nutrition-induced liver disease and the greater complications associated with a combined transplant. Severe motility

disorders and locally invasive tumours are the most common indications for multivisceral grafting with the liver, stomach, duodenum, pancreas, small bowel and/or colon.

How to decide whether a patient on long-term parenteral nutrition with disturbed liver function requires a combined liver/small-bowel transplant or an isolated small-bowel transplant is currently a matter of debate. The main issue is deciding which patients have irreversible liver dysfunction, thus requiring a combined transplant. Patients with cholestasis or mild fibrosis, but no portal hypertension, are offered isolated intestinal grafting because this procedure can normalize liver function in those who are intolerant of parenteral nutrition.

Contraindications to transplantation include incurable malignancy, significant comorbid conditions such as cardiopulmonary disease or advanced age, and systemic infection. Recurrent infection of the central venous line is one of the indications for small-bowel transplantation, although active infection must be cleared first. The typical patient will have had several bouts of infection efore being considered for a transplant; with the exception of *Candida* infections, which can seed chronic venous thrombi, these are usually controlled easily by removing the indwelling venous catheter.

Donor selection and the procurement procedure

Most cadaveric donors of multiple organs can be small-bowel donors. Many recipients have small abdominal cavities from previous bowel resections. Therefore, it is preferable, where possible, to use a donor who weighs 20 to 30 per cent less than the recipient. Donors are given oral and intravenous antibiotics to decontaminate the gut lumen. Mechanical bowel preparation is not required. An ABO identical match is preferred because of the risk of hemolysis from antibodies produced by B cells within the graft.

The benefits of HLA matching have been demonstrated experimentally, but not clinically. Although some centres give antilymphocyte products to the donor to deplete immune cells from the graft in an attempt to decrease graft-versus-host disease, the benefits are unproven. Although living-related small-bowel transplants have been performed, this procedure entails using even smaller vessels for anastomoses with increased risk of technical complications. None the less, as outcomes improve, the indications for living-related transplantation may expand as they have for the kidney and liver transplantation.

The donor operation for isolated intestinal transplant is usually done as a part of a multiorgan retrieval. Our approach is to remove the liver and small bowel *en bloc*, with careful division of the vasculature on the back table. If an aberrant right hepatic artery originating from the superior mesenteric artery is present, it is retained with the liver and an iliac arterial conduit is used to lengthen the foreshortened, bowel superior mesenteric artery.

The donor operation begins by examining the bowel to ensure its suitability for transplantation. The superior mesenteric vein and portal vein are separated from the pancreas. The proximal jejunum and distal ileum are stapled. Ligatures or clips are placed along the cut edge of the small-bowel mesentery to prevent chylous ascites in the recipient. The graft is perfused via an aortic cannula with University of Wisconsin solution. A Carrel patch of aorta, including the origin of the superior mesenteric artery, is taken to complete the retrieval. Most programs try to keep the cold ischemia time to less than 6 to10 h.

The approach to combined liver/small-bowel retrieval is evolving. Initially, the jejunum and ileum were removed *en bloc* with the liver, connected by the portal vein/superior mesenteric vein and an aortic conduit, including the celiac and superior mesenteric artery as well as a length of thoracic aorta. Recently, the University of Nebraska has recommended including the donor duodenum and head of the pancreas as part of this graft to avoid the need for biliary reconstruction and to facilitate the orientation of the transplanted bowel.

The donor operation for a multivisceral transplant involves an abdominal evisceration anterior to the kidneys, harvesting long lengths of vena cava and the thoracic aorta. The donor spleen is removed, taking care to avoid injury to the pancreas.

Recipient surgery

Isolated intestinal transplantation

There are several methods for restoring the vascular inflow and outflow of the isolated small-bowel graft (Fig. 1). The donor superior mesenteric artery can be anastomosed to the recipient's infrarenal or supraceliac aorta, or to the superior mesenteric artery. The venous end of the graft can be connected to the recipient's superior mesenteric vein, portal vein, or inferior vena cava. The theoretical advantages of portal venous outflow over systemic venous drainage include (i) the delivery of hepatotrophic substances from the gut directly to the liver, (ii) the filtering of organisms and toxins from the allograft by the native liver, and (iii) immune protection from rejection.

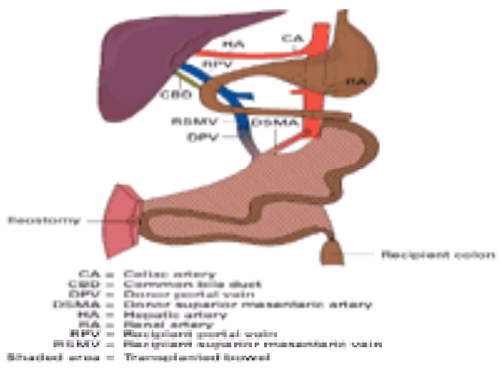


Fig. 1. Isolated small-bowel recipient operation. The donor superior mesenteric artery is anastomosed to the infrarenal aorta while the donor portal vein is anastomosed to the recipient superior mesenteric vein. The proximal and distal ends of the bowel are anastomosed to the native bowel with a diverting loop ileostomy.

The proximal end of the graft is anastomosed to the native bowel. Some centres include a segment of the donor colon in continuity with the small bowel to increase the absorptive capacity of the graft and potentially to decrease the ileostomy output because of the presence of the ileocecal valve. Other centres avoid including colon as doing so may be associated with increased graft loss. The distal end of the intestine can be anastomosed to the native gut, with the creation of a proximal diverting-loop ileostomy, or exteriorized as a stoma. The ileostomy can be closed 3 to 6 months after transplant, once the patient has stabilized with no evidence of rejection or infection.

Small-bowel/liver transplantation

The operation on the recipient for a combined liver/small-bowel transplant begins with the standard suprahepatic and infrahepatic caval anastomoses. The native portal vein, draining the stomach, duodenum, pancreas, and spleen, is anastomosed end-to-side to the donor portal vein. The donor aortic conduit is anastomosed to the infrarenal or supraceliac aorta. When the common bile duct is divided during the retrieval, bile flow is reconstructed by a Roux-en-Y biliary anastomosis with the donor bowel. When the duodenum and head of the pancreas are preserved with the graft, no biliary reconstruction is required (Fig. 2).

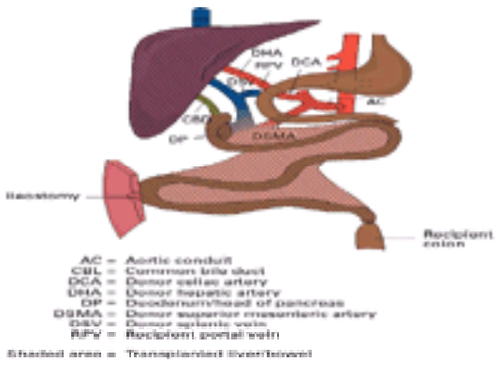


Fig. 2. Combined small-bowel/liver recipient operation. The suprahepatic and infrahepatic caval anastomoses are done as in isolated liver transplantation. The aortic conduit is anastomosed to the infrarenal aorta. The recipient portal vein is anastomosed to the donor splenic vein. Biliary reconstruction is not required because the duodenum and head of the pancreas are included with the graft.

Multivisceral transplantation

The recipient operation begins with total abdominal evisceration anterior to the kidneys, followed by implantation of liver, small bowel, pancreas, and duodenum, with or without the stomach and colon. The caval anastomoses proceed as for liver transplantation. The end of the donor aortic conduit is anastomosed to the side of the recipient's infrarenal aorta. No portal venous anastomosis is required. Proximal and distal gastrointestinal continuity is restored with a diverting-loop ileostomy (Fig. 3).

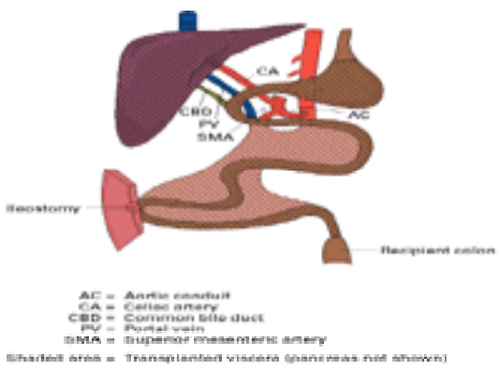


Fig. 3. Multivisceral recipient operation. The suprahepatic, infrahepatic, and aortic conduit anastomoses are done as in combined small-bowel/liver transplantation. Neither portal venous nor biliary reconstruction is required. The pancreas is not shown but is included in the graft.

Post-transplant management

Rejection

Immune suppression

The immunosuppressive regimen for intestinal transplantation continues to evolve. Most recipients are currently induced with tacrolimus and steroids. Many centres add prostaglandins and mycophenolate mofetil (CellCept®) to this regimen.

At our centre, tacrolimus is given orally at a starting dose of 0.3 mg/kg per day, aiming for serum concentrations of 20 to25 ng/ml. The dose is decreased after the first month, or earlier if there are signs of toxicity such as neurotoxicity, nephrotoxicity, or severe glucose intolerance. Methylprednisolone is started at 1 mg/kg per day, and tapered to 0.3 mg/kg per day. We give muromonab-CO3 (OKT3®) intraoperatively and for the first two postoperative days, but many programs do not use OKT3 induction. We also give one unit of donor blood for its potentially immune suppressive effect. Prostaglandin E_1 (Prostin VR®) (0.6–0.8 µg/kg per day) is given intravenously for 2 weeks, followed by oral misoprostol (Cytotec®) (200 µg, three times a day) for immune-suppressive, renal-protective, and vasodilatory effects.

Monitoring for rejection

Early rejection may be asymptomatic or it may present with symptoms ranging from mild fever and non-specific abdominal pain to vomiting, ileus, increased stomal output, acidosis, bacteraemia, and septic shock. The endoscopic appearance of rejection varies from dusky mucosa with oedema, hyperaemia, and decreased peristalsis to ulceration, mucosal sloughing, friability, bleeding, and loss of peristalsis. Microscopically, acute rejection begins with cryptitis and blunting of the villi, then progresses to vasculitis with a mononuclear-cell infiltrate followed by mucosal sloughing, inflammation, and graft necrosis. Transmural rejection may lead to spontaneous perforation of the graft.

Trans-stomal endoscopic mucosal biopsies are taken twice weekly for 1 month or longer to screen for rejection. There is a moderate risk of sampling error as the biopsies may be normal at one site while transmural rejection is occurring at another. At our centre, we screen the entire graft for increased permeability, which is an early sign of graft rejection, once a week by a test with chromium-labelled EDTA.

Chronic rejection presents with refractory diarrhea and weight loss. Endoscopic findings include loss of mucosal folds, pseudomembrane formation, and decreased peristalsis. Biopsies may reveal villous atrophy, subintimal vascular thickening, and fibrosis. Confirmation of chronic rejection may require a full-thickness biopsy.

Rejection episodes are managed by increasing the dose of tacrolimus and giving short-term, high-dose steroids. For refractory or severe rejection, treatment with OKT3 (5 mg/day for 7–10 days) or antithymocyte gammaglobulin (Atgam®, beginning at 500 mg/day with dose titrated to a CD2 count of 100 cells/mm^3 for 7–10 days) is used. Mycophenolate mofetil may be added to reduce the risk of repeated rejection episodes.

There is some recent evidence to suggest that a combined transplant is associated with less rejection than an isolated small-bowel transplant. It should be noted, however, that the bowel graft may undergo active rejection without any evidence of hepatic rejection.

Monitoring for graft-versus-host disease

Graft-versus-host disease is well documented in experimental small-bowel transplantation, but it has not been a common complication clinically. The disease presents with skin, hepatic, or pulmonary changes. Skin involvement varies from a maculopapular rash to exfoliative dermatitis. When the liver is affected, patients present with jaundice and biopsies show a periportal lymphocytic infiltration. Pulmonary involvement usually presents with interstitial pneumonitis. A tissue biopsy is required to confirm the diagnosis. Treatment of graft-versus-host-disease is by increasing immune suppression, usually with short-term, high-dose steroids.

Infection

During the first 3 months after small-bowel transplantation, most patients have at least one bacterial, viral, or fungal infection. To reduce post-transplant infections, patients are given prophylactic perioperative antibiotics, antivirals, and antifungals. At many centres, biopsy confirmation of graft rejection is required before antirejection treatment to avoid giving excessive immune suppression.

Graft rejection results in increased intestinal permeability with the attendant risk of bacterial translocation and systemic sepsis. The sequence of rejection followed by sepsis has been a major cause of death in small-bowel transplant recipients. Therefore, the most important step in preventing bacterial sepsis is vigilant observation for rejection and prompt treatment if found. Other measures to reduce the risk of bacterial infection include early enteral feeds to stimulate mucosal growth and preserve the gut barrier function as well as perioperative oral and intravenous antibiotics.

Bacterial infections of venous lines can be difficult to treat. Most patients who need an intestinal transplant have limited venous access, making it difficult to remove an infected line for lack of another access site. Systemic antibiotic therapy is often relied upon to clear these infections.

Antifungal agents, such as fluconazole (Diflucan®), are given for perioperative prophylaxis. Small-bowel transplant patients who develop multiorgan failure are at high risk for fungal sepsis. Treatment of established infection consists of supportive care and prolonged courses of antifungal therapy. Fungal sepsis is often a lethal

complication.

Viral infections are common after intestinal transplantation with cytomegalovirus and Epstein–Barr virus causing the most morbidity. Donor–recipient mismatches (cytomegalovirus-seropositive donor into cytomegalovirus-seronegative recipient) and the requirement for increased immune suppression to treat acute rejection are risk factors for cytomegalovirus disease. High-risk patients are given prophylactic ganciclovir (Cytovene®) and gammaglobulin during the first post-transplant month. The risk of cytomegalovirus infections can be reduced by implanting grafts from donors who are serologically negative for cytomegalovirus into recipients who are also serologically negative, and by only using cytomegalovirus-seronegative blood products. The clinical manifestations of cytomegalovirus infections include fever, leukopenia, gastrointestinal complications, hepatitis, and interstitial pneumonia. Active disease is treated by intravenous ganciclovir and reducing immune suppression. It can be difficult, and sometimes impossible, to eradicate established cytomegalovirus infection from intestinal grafts, presumably because the gut acts as a massive reservoir for the virus.

Epstein–Barr virus infection can present non-specifically with fever or a mononucleosis-like syndrome. Other presentations include generalized lymphadenopathy, atypical lymphocytosis, or hepatitis. Epstein–Barr virus is a B-cell lymphotropic virus that is maintained within B lymphocytes in a latent state. When patients are immune suppressed, the virus can stimulate B-cell proliferation leading to polyclonal reactive lymphoid hyperplasia that can evolve into a true monoclonal, large-cell lymphoma.

There appears to be a greater risk of developing post-transplant lymphoproliferative disorders after intestinal transplantation than after other organ transplants, probably related to the increased amount of immune suppression required. The Pittsburgh program has reported incidence rates for lymphoproliferative disorders as high as 29 per cent in children or infants with small-bowel grafts. The increased risk in children is probably because a primary infection with Epstein–Barr virus carries a higher risk of developing into a post-transplant lymphoproliferative disorder than does reactivation of an existing infection or reinfection. Ways of decreasing the likelihood of post-transplant lymphoproliferative disorders include minimizing immune suppression, prophylactic treatment with gammaglobulin containing high titres of anti-Epstein–Barr virus antibodies, and the pre-emptive use of ganciclovir when the polymerase chain reaction reveals high titres of the virus genome in the peripheral blood lymphocytes. The initial treatment of established post-transplant lymphoproliferative disorder is a reduction of immune suppression and acyclovir (Zovirax®) or ganciclovir. If there is no response, patients are given systemic chemotherapy.

Nutrition

Parenteral nutrition is continued in the perioperative period, depending on the availability of venous access. Enteral feeds are begun as soon as possible after transplantation via a nasoduodenal feeding tube placed at the time of surgery. Tube feeds provide luminal nutrients that prevent the villous atrophy and increased intestinal permeability which occurs with prolonged parenteral nutrition. Most patients are able to resume full enteral nutrition within 4 to 6 weeks of the transplant.

Complications

The major complications of intestinal transplantation are rejection and infection, but vascular, gastrointestinal, renal, and pulmonary complications may also occur.

Vascular complications include arterial stenosis, thrombosis, pseudoaneurysm formation, and venous thrombosis. Potential risk factors for arterial stenosis and thrombosis include the use of an interposition aortic conduit, the small size of the arteries in children, and pre-existing hypercoagulability. Magnifying loupes or an operating microscope are used for the vascular anastomoses and patients can be anticoagulated postoperatively to improve their patency. Pseudoaneurysm formation has been reported in solid-organ transplantation and may be infectious in origin or secondary to technical errors. This is a difficult complication to manage and, if accompanied by arterial rupture, is associated with significant mortality.

Increased postoperative abdominal pressure due to size mismatches between the donor organ and the recipient's abdominal cavity can lead to an abdominal compartment syndrome with the attendant risk of arterial or venous thrombosis. The compartment syndrome can be avoided by selecting donors that weigh 20 to 30 per cent less than the recipient. Synthetic mesh is used to close the abdomen when the donor small bowel is larger than anticipated.

Small-bowel transplant recipients are also at risk for gastrointestinal complications. Malnutrition, graft preservation damage, and steroid administration are risk factors for impaired healing at the anastomotic sites. Perforation may occur with endoscopic biopsies. Transmural rejection may present with a spontaneous perforation of the graft. Intestinal complications are usually managed by primary repair with or without fecal diversion. Enteral feeding is resumed as soon as possible.

Renal and pulmonary complications can occur as in all types of transplantation. Patient management is complicated by poor venous access, which can make hemodialysis almost impossible.

Outcome

Successful small-bowel transplantation leads to a marked improvement in quality of life. Patients regain the ability to enjoy meals and attain freedom from the 8 to12 h required each day for parenteral nutrition.

An international Intestinal Transplant Registry follows the outcomes of clinical bowel transplantation. Currently, there are 33 intestinal transplant programs, all of whom submit data to the registry. The most recent analysis available updated the results of transplantation to the end of February 1997. To that date, 273 transplants had been performed on 260 patients. Two-thirds of the recipients were children or teenagers. The types of transplants included the small bowel with or without the colon (41 per cent), the intestine and liver (48 per cent), and multivisceral grafts (11 per cent). Maintenance immune suppression was tacrolimus for 212 grafts (78 per cent; 44 of these patients also received mycophenolate mofetil), cyclosporin for 51 grafts (19 per cent), and other or no immune suppression for 10 grafts (3 per cent).

Graft function

The small-bowel graft is capable of functioning such that most patients can be converted to purely enteral feedings. Potential risk factors for poor graft function include preservation damage, neural and lymphatic disruption, rejection, infection, and medication toxicity. Despite these factors, 95 (77 per cent) of the 126 patients surviving intestinal transplant followed in the Transplant Registry have been successfully weaned off parenteral nutrition.

Preservation or ischemic injury to the graft appears to lead to temporary mucosal dysfunction, with resolution expected within the first few days after transplant. The effect of neural disruption does not seem to be clinically significant as intestinal motility is well preserved in the small-bowel graft. Lymphatic disruption and regeneration has been extensively studied in animal models. Lymphatic regeneration appears to be complete by 2 to 4 weeks after the transplant. Early after the transplant the lymph drains into the peritoneal cavity, where it is absorbed. Chylous ascites is a potential risk, although this is usually self-limiting. If chylous ascites persists, patients can be treated with low-fat enteral intake or total parenteral nutrition and a somatostatin analogue.

Patients are given tacrolimus orally immediately after transplantation and adequate serum concentrations are attained quickly. Oral intake is gradually increased until recipients are able to maintain their hydration and nutrition. Rejection, infection, and medication toxicity can lead to increased stool volumes requiring treatment with parenteral fluids and bicarbonate.

Patient and graft survival

Patient and graft survival improved dramatically after the introduction of tacrolimus-based immune suppression. A review of the Intestinal Transplant Registry data up until June 1995 analysed the survival of 180 transplants in 170 patients given cyclosporin or tacrolimus. With cyclosporin, the 3-year graft and patient survival rates were: 11 and 50 per cent for isolated intestinal transplants; 28 and 28 per cent for liver/small-bowel grafts; and 41 per cent for multivisceral grafts. For tacrolimus, the 3-year graft and patient survival were: 29 and 47 per cent for isolated intestinal transplant; 38 and 40 per cent for combined liver/small bowel; and 37 and 43 per cent for multivisceral grafts. Registry data to the end of February 1997 revealed that 131 of the 260 recipients had died. Sepsis and multiorgan system failure accounted for almost 75 per cent of the mortality. Post-transplant lymphomas (10 per cent) and graft thrombosis (10 per cent) were also common causes of death.

The most recent review of the Registry data reveals that transplants after 1991 and at centres that had performed at least 10 transplants had significantly higher rates of graft survival. The 1-year graft survival for all types of intestinal transplants performed between February 1995 and February 1997 was approx. 60 per cent. These results reveal a steady improvement in outcome as experience is gained. Long-term results are still lacking but will be readily available as patients are followed in the Transplant Registry.

Conclusions and future prospects

Intestinal transplantation has finally emerged as a viable clinical therapy. The procedure is still moderately risky and only recommended for patients who have failed conventional treatment. However, as the results improve, small-bowel transplantation will undoubtedly become the treatment of choice for intestinal failure, in the same way that kidney transplantation is currently the treatment of choice for chronic renal failure.

Further reading

Alessiani M, Tzakis A, Todo S, Demetris AJ, Fung JJ, Starzl TE. Assessment of five-year experience with abdominal organ cluster transplantation. *Journal of the American College of Surgeons* 1995; **180:** 1–9. [Description of the techniques and outcomes of cluster transplantation after upper abdominal exenteration.]

Asfar S *et al.* Small bowel transplantation. A life-saving option for selected patients with intestinal failure. *Digestive Diseases and Sciences* 1996; **41:**, 875–83. [Review of outcomes of small-bowel transplantation including contrasting results with cyclosporin versus tacrolimus immune suppression.]

Black RT *et al.* Transplantation of segmental rat intestinal grafts including the ileocecal valve and the ascending colon. *Transplantation* 1994; **57**: 997–1002. [Demonstration of improved function of segmental small-bowel isografts when colon is included.]

Bland P. MHC class II expression by the gut epithelium. *Immunology Today* 1988; **9**: 174–8. [A review of the role of the small-bowel epithelium in the immune response.]

Carrel A. La technique operatoire des anastomoses vasculaire et la transplantation des visceres. *Lyon Medical* 1902; **98**: 859. [The first report of experimental small-bowel transplantation.]

Cohen Z *et al.* Small intestinal transplantation using cyclosporine. Report of a case. *Transplantation* 1986; **42**: 613–21. [ABO-compatible, but not identical, intestinal transplant resulting in clinically significant haemolysis.]

Goulet O *et al.* Small-bowel transplantation in children. *Transplantation Proceedings* 1990; **22**: 2499–500. [Early report of successful human isolated small-bowel transplant.]

Grant D *et al.* Graft-versus-host disease associated with intestinal transplantation in the rat. *Transplantation* 1989; **48**: 545–9. [Demonstration of the presentation and natural history of graft-versus-host disease in a rodent model.]

Grant D *et al.* Treatment with continuous high-dose intravenous cyclosporine following intestinal transplantation. *Transplantation* 1989; **48:** 151–2. [A case report documenting dramatic resolution of parenteral nutrition-induced cholestasis in the immediate post-intestinal transplant period.]

Grant D *et al.* Successful small-bowel/liver transplantation. *Lancet* 1990; **335**: 181–4. [The first successful human combined small-bowel/liver transplant.]

Grant D. Current results of intestinal transplantation. *Lancet*, 1996; **347**: 1801–3. [The first report from the international Intestinal Transplant Registry.]

Grant D. Intestinal transplantation: 1997 report of the international registry. *Transplantation* 1999; **67**: 1061–4).

Howard L, Ament M, Fleming CR, Shike M, Steiger E. Current use and clinical outcome of home parenteral and enteral nutrition therapies in the United States. *Gastroenterology* 1995; **109**: 355–65. [Review of the complications and rising costs of home parenteral nutrition.]

Kocoshis SA. Small bowel transplantation in infants and children. *Gastroenterology Clinics of North America* 1994; **23**: 727–42. [Review of the success and complications of small-bowel transplantation in children.]

Kocoshis SA, Reyes J, Todo S, Starzl TE. Small intestinal transplantation for irreversible intestinal failure in children. *Digestive Diseases and Sciences* 1997; **42**: 1997–2008. [Review of the indications for and outcomes of small-bowel transplantation in children.]

Kurkubasche AG, Row MI, Smith SD. Adaptation in short bowel syndrome: reassessing old limits. *Journal of Pediatric Surgery* 1993; **28**:1069–71. [Further definition of the lower limits of intestinal length required to allow for enteral nutrition.]

Langnas AN *et al.* Preliminary experience with intestinal transplantation in infants and children. *Pediatrics* 1996; **97**: 443–8. [Review of recent experience with small-bowel transplantation at a single centre.]

Lee RG *et al.* Pathology of human intestinal transplantation. *Gastroenterology* 1996; **110**: 1820–34. [Description of the histopathology of human intestinal allografts including the features of rejection.]

Lillehei RC, Goott B, Miller FA. Homografts of the small bowel. *Surgical Forum* 1959; **10**: 197–201. [Demonstration of a successful canine small-bowel allograft model.]

Lillehei RC *et al.* Transplantation of stomach, intestine, and pancreas: experimental and clinical observations. *Surgery* 1967; **62**: 721–41. [Review of this group's extensive work on canine small-bowel transplantation and early clinical experience.]

Nalesnik MA. Posttransplantation lymphoproliferative disorders (PTLD): current perspectives. *Seminars in Thoracic and Cardiovascular Surgery* 1996; **8**: 139–48. [Review of posttransplantation lymphoproliferative disorders including presentation, diagnosis, and treatment.]

Reyes J *et al.* Current status of intestinal transplantation in children. *Journal of Pediatric Surgery* 1998; **33**: 243–54. [The most recent update of the Pittsburgh experience in childhood small-bowel transplantation.]

Thompson JS, Langnas AN, Pinch LW, Kaufman S, Quigley EMM, Vanderhoof A. Surgical approach to short-bowel syndrome. *Annals of Surgery* 1995; **222**: 600–7. [Overview of the surgical approach to short-bowel syndrome.]

Todo S *et al.* Outcome analysis of 71 clinical intestinal transplantations. *Annals of Surgery* 1995; **222**: 270–82. [Review of the largest single institution experience with small-bowel transplantation.]

16.9 Vascularized pancreatic transplantation

T. Pearson and C. Larsen

Introduction
Diabetes
Pancreas transplantation
History
Indications
Surgical technique
Post-transplant management
Results
Conclusions
Further reading

Introduction

Diabetes

General

Diabetes mellitus (DM) occurs either as a result of diminished or absent insulin production or when a decrease in the responsiveness to insulin can not be compensated for by an increase in insulin production. The former case has been classified as insulin dependent diabetes mellitus (IDDM or Type I) and the latter as non-insulin dependent diabetes mellitus (NIDDM or Type II). About 10 per cent of all patients with DM have Type I diabetes. Type I diabetes generally presents during childhood or adolescence and is the result of an autoimmune-mediated destruction of the b-cells in the islets of Langerhans in the pancreas. Ketoacidosis may occur as a result of Type I DM. Exogenous insulin therapy largely prevents these episodes; however, even with close attention to treatment plans it is often difficult to achieve long-term normoglycaemia.

Prevalence

DM is a very common disease with frequent and often severe secondary complications and therefore it has a tremendous impact on the health-care system. In 1997, a total of 15.7 million people had diabetes in the United States which was 5.9 per cent of the total population. Of this total, 5.4 million people had undiagnosed diabetes. Almost 800 000 new cases of diabetes are diagnosed each year in the United States.

Pathogenesis

Type I (IDDM) is characterized by an absolute insulin deficiency. It is the result of an autoimmune-mediated destruction of the beta cells in the pancreas. The immunological response is associated with a lymphocytic infiltration of the pancreas, called insulitis, and clinical diabetes becomes evident when more than 90 per cent of the beta cells have been destroyed. The humoral arm of the immune system, with antibodies directed against islet-cell-associated antigens, also probably plays a role in the pathogenesis of this disease. IDDM clearly has a genetic component as specific HLA antigens are linked to the disease, but an exact pattern of inheritance has not been defined. Viral infections may also play a role in IDDM but whether they precipitate the disease in susceptible individuals is not clear.

Secondary complications

The secondary complications of diabetes are frequent and often severe (Table 1). The death rate of middle-aged people with diabetes in the United States is twice that for the same age group without diabetes, based on death certificate data. It is estimated that diabetes contributed to almost 200 000 deaths in the United States in 1996. Heart disease is the leading cause of diabetes-related deaths. The rate of heart disease in adults with diabetes is about 2 to 4 times higher than that in adults without diabetes. A similar situation exists for stoke rates, and diabetes is the leading cause of new cases of blindness in adults between 20 and 74 years old. Diabetic retinopathy causes between 12 000 and 24 000 new cases of blindness each year. In addition, diabetes is the leading cause of end-stage renal disease in the United States, accounting for about 40 per cent of all new cases. In 1995, almost 100 000 people underwent either dialysis or kidney transplantation. About 60 to 70 per cent of people with diabetes have some form of neuropathy, which in some cases contributes to lower extremity amputations. This factor, combined with the high incidence of peripheral vascular disease, results in almost 70 000 amputations per year in the United States in people with diabetes. These factors combine to make the financial cost of this disease considerable, with estimates for the total cost in the United States in 1997 being almost $100 billion, with $44 billion dollars of this total representing direct medical costs.

Complication	*Cumulative prevalence (%)*
Visual impairment	14
Blindness	16
Renal failure	22
Stroke	10
Amputation	12
Myocardial infarction	21
Median survival after diagnosis of IDDM (year)	36
Median age at death (year)	49

Table 1 The prevalence of complications in patients with IDDM*

Treatment options

The discovery of insulin in 1921 changed diabetes from a disease that frequently resulted in early death secondary to ketoacidosis to one in which the secondary complications of diabetes became the major case of morbidity and mortality. It has been documented by several studies published in the early 1990s that improved peripheral blood glucose control has a beneficial effect on the progression of the secondary complications of diabetes. Unfortunately, administration of exogenous insulin by standard techniques cannot consistently achieve excellent blood glucose control. This has stimulated the development of strategies involving multiple injections of insulin combined with close monitoring of blood glucose, and the use of constant insulin infusions via an insulin pump. However, these techniques are hampered by an increased incidence of hypoglycaemic episodes, technical problems with the pumps, and failure to achieve normal glucose levels as determined by glycosylated haemoglobin values. This provides the opportunity for improved outcome for the diabetic patient with transplantation of the pancreas.

Pancreas transplantation

History

The first clinical pancreas transplant, performed by Kelly and Lillehei at the University of Minnesota in 1966, was a segmental graft placed in the iliac fossa with ligation of the pancreatic duct. This marked the beginning of a period when few pancreas transplants were performed and there was an associated poor graft survival rate. At this time considerable attention was paid to development of the technique and in particular the optimal method for handling the pancreatic duct exocrine secretions. In the mid-1980s, Sollinger and his colleagues at the University of Wisconsin applied, refined, and analysed their results using the technique of draining these secretions into the urinary bladder using a segment of donor duodenum as a conduit. This development, in conjunction with improved immunosuppressive medications, lead to

improved graft survival results and a dramatic increase in the number of transplants performed during the late 1980s and early 1990s (Fig. 1).

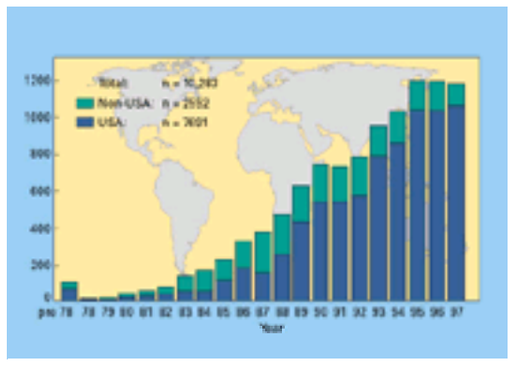


Fig. 1. The number of pancreas transplants performed worldwide that have been reported to the International Pancreas Transplant Registry.

Indications

General

The indications for pancreas transplant remain controversial. However, since the goal of pancreas transplantation is to achieve long-term euglycaemia and thus prevent or slow the progression of the secondary complications of diabetes, it should be performed before the patient has end-stage effects of the disease. In particular, this refers to the vascular disease that so frequently results in morbidity and mortality from myocardial infarctions and strokes. For this reason, a dobutamine stress echocardiogram is performed on all patients being considered for pancreas transplantation and coronary angiography extensively used, as indicated. Absolute contraindications for pancreas transplantation are similar to those for renal transplantation and include malignancy, active infection, advanced cardiovascular disease, and the lack of ability and/or support systems to allow compliance with the post-transplant treatment regimen. Pancreas transplantation can be applied in three different settings, as described below.

Simultaneous renal–pancreas transplantation

Combined renal–pancreas transplant is by far the most common setting (Fig. 2). The patient in this situation is a diabetic who is either on dialysis or immediately predialysis and does not have other severe complications of the disease. This setting has advantages that only one surgical procedure is required, and the immunosuppressive medicines are similar, if not identical, to what the patient would receive for a renal transplant alone. In addition, the kidney and pancreas share donor antigens and therefore rejection in the pancreas may be identified by the diagnosis of rejection in the transplanted kidney. The relatively easy, early diagnosis of rejection in the kidney and its use as an indicator of rejection in the pancreas is an important advantage of simultaneous kidney pancreas transplant from the same donor.

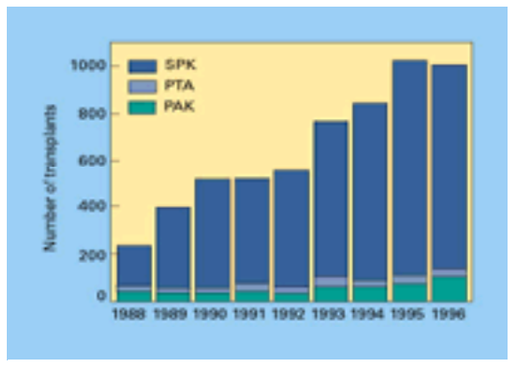


Fig. 2. The number of pancreas transplants performed in the United States based on recipient category: Simultaneous kidney pancreas (SPK), pancreas transplant alone (PTA), and pancreas after kidney transplant (PAK). Data are from the UNOS registry.

Pancreas after renal transplant

Pancreas transplantation may also be considered in the setting of the diabetic patient who has previously received a kidney transplant that retains excellent renal function and who does not have advanced, end-stage complications of diabetes. In this situation the risk of a second surgical procedure must be balanced against the potential benefit derived as a result of slowing the progression of the secondary complications of diabetes. The immunosuppressive regimen after the secondary pancreas transplant would be similar to that after renal transplant alone, therefore the recipient would not be at a substantially increased risk for complications from these medications. However, because the pancreas and kidney would probably not share the same antigens the renal transplant may not function as well as an early indicator of pancreas rejection.

Pre-emptive isolated pancreas transplant

The third setting in which pancreas transplantation may be performed is in the diabetic patient prior to the onset of renal insufficiency. This indication for pancreas transplant has not been widely applied, because in the opinion of many centres the known risks of long-term immunosuppression out weigh the unproven potential beneficial effect of a pancreas transplant on the secondary complications of diabetes. One particular indication may be the diabetic patient who experiences frequent, life-threatening hyper- or hypoglycaemic episode despite good compliance with an appropriately prescribed exogenous insulin regimen.

The pancreas donor

Cadaveric donor

Cadaveric donors are the source of the vast majority of pancreas glands for transplant. As with other potential organ donors, active infection or malignancy constitute a contraindication to pancreas procurement. Specific conditions that make a donor unsuitable for pancreas donation include a history of diabetes or chronic pancreatitis, and traumatic injury of the pancreas. Hyperglycaemia is common in cadaveric organ donors and is not a contraindication for pancreas transplant. Likewise, hyperamylasaemia in the organ donor may result from causes other than injury to the pancreas and is not an absolute contraindication for use of the pancreas. The final decision regarding the suitability of the pancreas for transplantation is made at the time of procurement surgery.

Living donor

The world experience with living donor pancreas transplantation is small and largely confined to one centre, the University of Minnesota. This procedure involves segmental pancreas transplantation in which the tail and body of the pancreas are procured from a living donor by dividing the pancreas and the splenic vessels in a plane overlying the superior mesenteric vessels.

Surgical technique

Donor procedure

Cadaveric donor

The vast majority of pancreas transplants are performed using composite pancreaticoduodenal grafts from cadaveric donors who simultaneously provide kidneys, a liver, and a heart for transplant. The abdominal organs are exposed via a vertical midline incision from the suprasternal notch to the pubis. Our centre prefers the en-bloc technique in which the liver, kidneys, and pancreas are simultaneously removed from the donor and then divided while at 4°C in preservation solution. This requires careful attention to the identification of any aberrant arterial vascular supply to the liver prior to en-bloc removal of the organs, but minimizes the dissection time in the heart-beating donor prior to cross-clamping the aorta. Attention to detail is important in separating the pancreas and liver. The portal vein is divided between the two organs near the level of the coronary vein. The coeliac trunk and patch of donor aorta remains with the liver and the splenic and gastroduodenal arteries are divided (Fig. 3). The gastroduodenal artery is ligated on the pancreas graft and the splenic artery used as the vascular supply to the tail of the pancreas. In most cases, the arterial blood supply is reconstructed with an iliac artery Y-graft to the splenic and superior mesenteric arteries on the pancreas graft (Fig. 3). The duodenum is divided just distal to the pylorus and again in the third portion. This facilitates construction of the composite graft and utilization of the C-loop of the duodenum as a conduit for drainage of the pancreatic exocrine secretions. The pancreas is flushed with University of Wisconsin solution (Viaspan®, Dupont Pharmaceuticals) prior to removal from the donor and then stored in this solution at 4°C prior to transplantation. Many centres attempt to keep the total cold ischaemia time to less than about 16 h, but many successful pancreas transplants have been reported with ischaemia times between 20 and 30 h.

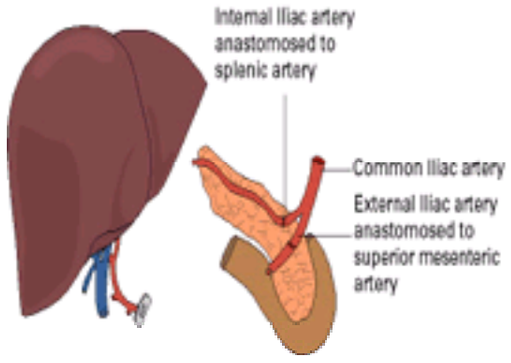


Fig. 3. Vascular reconstruction after simultaneous procurement of the liver and pancreas from a cadaveric donor.

Recipient procedure

Simultaneous renal pancreas transplant

The surgical procedure in the recipient begins with a laparotomy that is performed via a vertical midline incision. The pancreas is generally placed on the patient's right side and the kidney on the left side. The renal vessels are anastomosed to the iliac vessels, generally in an end-to-side fashion with the ureter implanted using an anterolateral extravesicular Liche technique. At the completion of the renal transplant the pancreas is implanted, with a total operative time of approximately 4 h. The iliac artery or brachiocephalic Y-graft from the donor used to reconstruct the superior mesenteric artery and splenic artery is anastomosed end-to-side fashion to the iliac artery of the recipient. Venous outflow from the pancreas is via the portal vein that is anastomosed end-to-side into the iliac vein of the recipient. Alternatively, the portal vein may be anastomosed to a mesenteric vein to provide the venous drainage from the pancreas. The latter approach has the advantage of recreating the portal venous outflow from the pancreas that is found in the native pancreas. While this technique is preferred by several centres, in the absence of a proven benefit most pancreas transplants continue to be performed with systemic venous drainage.

The preferred exocrine drainage of the pancreas is via a side-to-side duodenojejunostomy (Fig. 4(a)). In cases in which enteric drainage is not feasible, duodenocystostomy is the preferred alternative for handling the pancreatic exocrine secretions (Fig. 4(b)). During the evolution of the surgical technique in the late 1980s and early 1990s, bladder drainage evolved as the most popular technique because it was technically simple and associated with a low rate of early complications, and, in particular, a low incidence of intra-abdominal infection. This approach has the additional advantage of providing for the measurement of urinary amylase, a technique that many centres have found a useful means by which to diagnose rejection of the pancreas. The disadvantage of bladder drainage of the exocrine pancreas is the complications that frequently present on longer follow-up. Urological complications including haematuria, cystitis, urethitis and urethral stricture, and urinary tract infections are common. Metabolic acidosis and dehydration secondary to bicarbonate and volume loss in the exocrine secretions are also a significant clinical problem. Therefore enteric drainage of the exocrine pancreas is currently the preferred technique at many transplant centres.

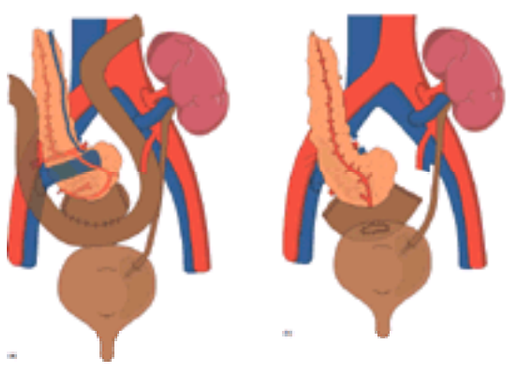


Fig. 4. (a) Simultaneous pancreas and kidney transplantation with drainage of the exocrine secretions of the pancreas into a loop of small intestine. (b) Simultaneous pancreas and kidney transplantation with drainage of the exocrine secretions of the pancreas into the bladder.

Post-transplant management

General aspects

The general guidelines for postoperative management of kidney–pancreas transplant recipients is similar to that established for kidney recipients. The patient is generally monitored in the intensive care unit for the first postoperative night. In contrast to renal transplantation, the transperitoneal approach to pancreas transplantation and the enteric anastomosis of the pancreas results in a postoperative ileus that is managed with nasogastric tube decompression. The ileus generally resolves by the third or fourth day after surgery at which time an oral diet and medications are begun and then slowly advanced.

Glucose control

Euglycaemia in the early post-transplant period is maintained with a continuous intravenous infusion of insulin and frequent monitoring of the blood glucose. The rate of insulin infusion is adjusted to keep the blood glucose between 80 and 120 mg/dl. Patients are frequently able to discontinue exogenous insulin by the third or fourth post-transplant day, but are treated with supplements of subcutaneous insulin via a sliding scale regimen if required as they increase oral intake. Insulin requirements usually decrease to zero as the dose of steroids used to prevent rejection is decreased in the early postoperative period.

Anticoagulation

Thrombosis of the vessels of the pancreatic allograft has accounted for loss of the transplant in the early transplant period in approximately 10 per cent of recipients in

many series. The aetiology of this complication is probably multifactorial and not completely defined but has lead many centres to use anticoagulation as part of the post-transplant regimen. Agents used include heparin and dextran, but their use must be balanced against the risk of a significantly higher incidence of postoperative bleeding. At our centre we currently use subcutaneous heparin (5000 U twice per day) for the first 14 days. We begin aspirin therapy when the recipient is tolerating oral medications.

Immunosuppression

Immunosuppression for pancreas transplantation is similar to that used for renal transplantation. Many centres in the United States use quadruple immunosuppression for pancreas transplant recipients. This consists of induction therapy with an antibody agent such as OKT3 for 7 to 14 days followed by maintenance therapy with the combination of cyclosporine, mycophenolate mofetil (MMF), and prednisone. Many centres now use tacrolimus in place of cyclosporine in this treatment regimen. At our centre we currently do not use induction therapy but begin the patient on intravenous cyclosporine, MMF, and steroids at the time of surgery, which are then converted to oral medications as the patient's postoperative ileus resolves. Acute rejection episodes, especially if they are mild, are treated with intravenous solumedrol, while more severe episodes unresponsive to steroids are treated with OKT3. Patients who have been maintained on cyclosporine are frequently converted to tacrolimus at the time of treatment for a rejection episode.

Complications

Immunological

Rejection is the major cause of pancreas allograft loss, and therefore represents an important diagnostic problem. This problem is made more difficult because we lack a non-invasive, reliable means to diagnosis early rejection in the transplant pancreas. Unfortunately, elevated blood glucose does not occur until late in the process at which time the rejection process is frequently irreversible. Serum amylase and lipase levels may be an indicator of pancreatic injury, but are not specific for rejection. Urinary amylase has been found to be a clinically useful test in this regard by many centres, but this test is only useful in bladder-drained transplants, and not in those patients with enteric drainage. Biopsy of the transplanted pancreas may be performed via a cystoscopic approach in bladder-drained transplants, or percutaneously with CT scan guidance for those with enteric drainage. Defined histological criteria have now been developed for the diagnosis of rejection in the pancreas.

Rejection of renal allografts is usually detected early as a result of careful serial monitoring of serum creatinine values. Therefore monitoring of renal function in recipients of combined renal pancreas transplants from the same donor provides the most practical and reliable means of diagnosing rejection. Initiation of treatment for rejection in the kidney will presumably be adequate to treat early rejection that is assumed to be simultaneously occurring in the pancreas.

Non-immunological

Graft thrombosis Thrombosis of the venous outflow from the pancreas transplant and subsequent congestive necrosis continues to contribute to graft failure in a significant number of patients at many transplant centres. The aetiology of this problem is not completely understood, but is probably multifactorial. As mentioned above, this situation has lead many centres to institute prophylactic anticoagulation therapy in the early postoperative period.

The clinical presentation of graft thrombosis usually includes an acute increase in blood glucose values or insulin requirements associated with abdominal pain that localizes to the area of the transplant. The diagnosis most frequently is made within the first 4 days after transplant and can be confirmed by a doppler ultrasound examination. The diagnosis mandates urgent exploration and graft pancreatectomy to prevent septic complications.

Graft pancreatitis Hyperamylasaemia, presumably as a consequence of injury to the pancreas during procurement, preservation, and transplantation, is common in the early post-transplant period. This is usually self-limited and does not require specific treatment. Pancreatitis may also occur at anytime later in the postoperative period. The patient presents with abdominal pain over the transplant and hyperamylasaemia, and is evaluated with a CT scan of the abdomen to make the diagnosis and to rule-out other diagnoses such as peripancreatic abscess or peritonitis from other causes. Pancreatitis in a patient with bladder drainage of the exocrine pancreas may benefit from placement of a foley catheter, otherwise treatment is supportive with intravenous fluids, pain medications, and careful monitoring.

Anastomotic leak Disruption of the suture line at the anastomosis to drain the exocrine pancreatic secretions may occur after either enteric or bladder drainage. The symptoms of a bladder anastomosis leak may be subtle, and include vague lower abdominal pain, abdominal swelling, and elevated serum creatinine and amylase values. The diagnosis is made by cystogram. Small leaks in favourable clinical situations may be treated by Foley catheter drainage followed by a repeat cystogram to document closure. Large, recurrent or recalcitrant leaks require surgical repair. Anastomotic leak of an enteric anastomosis is an intra-abdominal emergency and requires prompt reoperation. If the clinical situation dictates that the leak can not be safely repaired, then a graft pancreatectomy may be required.

Sepsis Intra-abdominal sepsis generally occurs in the first month after pancreas transplant, possibly as a result of contamination with enteric contents at the time of surgery or as a result of a small anastomotic leak. The resulting peripancreatic abscess is diagnosed by CT scan. Treatment consists of antibiotics based on culture results, and drainage by either CT-scan-directed percutaneous catheters or open surgical approach.

Bleeding Bleeding may occur into the peritoneal cavity from the transplanted pancreas, especially in the first few days after transplant. In most cases this is a self-limited process but on occasion may require re-exploration. Bleeding may also occur from the duodenal segment, which in the case of a bladder drained pancreas results in haematuria. This frequently is the result of ulceration in the duodenal mucosa, and should be evaluated with cystoscopy and biopsy of the ulcer. The histological examination should rule out CMV disease as an aetiologic agent for the ulcer. Treatment of the haematuria may require anti-CMV agents, in addition to supportive care. Severe or recurrent haematuria is an indication for conversion to enteric drainage.

Results

Graft and patient survival rates

Results in the early period of pancreas transplantation were poor, both in terms of a low rate of transplant survival and low numbers of transplants performed. The situation has markedly improved over the last 15 to 20 years. By the end of 1997, more than 10 000 pancreas transplants had been reported to the international pancreas transplant registry (Fig. 1). Pancreas survival rates have dramatically improved in the more recent experience. For example the 1-year survival rates for the patient was 94 per cent, kidney 90 per cent, and pancreas 82 per cent, for the 1620 combined kidney pancreas transplants performed in the United States in 1995 to 96 and reported in the 1998 analysis of the UNOS transplant registry data (Fig. 5). Excellent long-term results are also now being documented as indicated in the large single-centre experience at the University of Wisconsin. Their review of 500 kidney–pancreas transplants performed between 1985 and 1997 documented 1, 5, and 10-year survival rates for the patients of 96.4, 88.6, and 76.3 per cent; for the kidney of 88.6, 80.3, and 66.6 per cent; and for the pancreas of 87.5, 78.1, and 67.2 per cent.

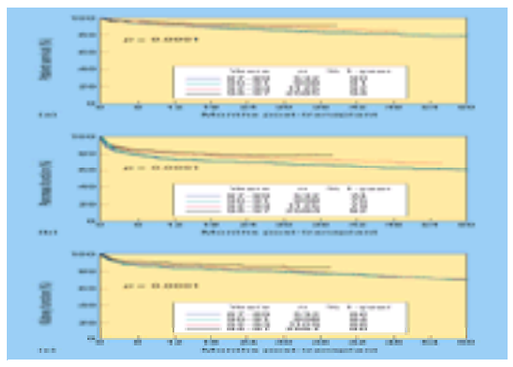


Fig. 5. Patient (a), pancreas (b), and kidney (c) graft survival rates for simultaneous kidney pancreas transplant recipients in the United States according to era. Data are from the UNOS registry.

Further analysis of the UNOS registry data referred to above shows that patient survival rates for kidney–pancreas recipients were similar to those for cadaveric kidney

recipients at 1, 3, and 5 years after transplant. Interestingly, pancreas graft survival rates in combined transplant recipients were significantly better when the donor was less than 50 years old. However, the level of HLA matching appeared to have little impact on either short or long-term kidney–pancreas graft or patient survival rates.

The number of pancreas transplants alone (either prior to or after a kidney transplant) is small in comparison to combined kidney–pancreas transplant procedures (Fig. 6 and Fig. 7). For example the 1995 to 96 cohort in the UNOS transplant registry was only 198 recipients compared to 1620 recipients of combined transplants. The survival results are also less favourable in this group. The 1, 3, and 5-year pancreas graft survival rates were 70, 43, and 32 per cent, respectively.

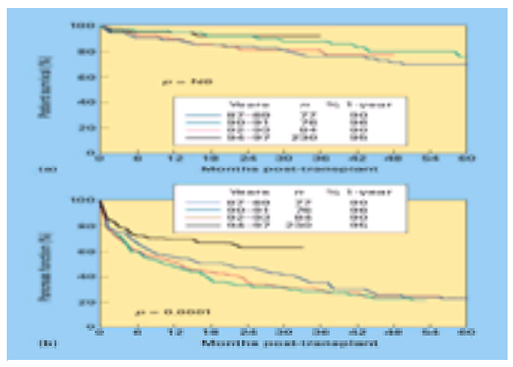


Fig. 6. Patient (a) and pancreas (b) graft survival rates for pancreas after kidney transplant recipients in the United States according to era. Data are from the UNOS registry.

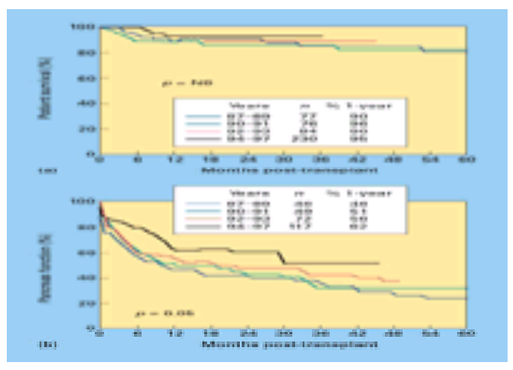


Fig. 7. Patient (a) and pancreas (b) graft survival rates for pancreas transplant alone recipients in the United States according to era. Data are from the UNOS registry.

Metabolic effects

Glucose control

Pancreatic transplantation generally produces normal blood glucose levels within days after transplant. Factors which may contribute to a delay in achieving normal glucose metabolism include ischaemic injury to the islet cells during the transplantation process, infusion of glucose-containing intravenous solutions and the use of relatively high doses of steroids for immunosuppression in the early post-transplant period. Haemoglobin A_{1c} (Hb A_{1c}) values return to normal within 2 months after transplant in the vast majority of patients. Most patients have a normal response to oral and intravenous glucose tolerance tests. Hyperinsulinaemia with elevated levels of C-peptide are found in most recipients with systemic venous drainage of the pancreas, presumably largely due to the lost of the first-pass metabolism of insulin in the liver that is associated with the portal venous drainage of the native pancreas.

Effect on secondary complications

The primary goal of pancreas transplantation is to slow or prevent the progression of secondary complications of diabetes which so frequently result in severe morbidity and ultimately mortality. However, the degree to which pancreas transplantation meets this goal remains to be completely defined. There have been no large-scale, prospective, randomized trails with long-term follow-up to address this important question. Circumstantial evidence for a beneficial effect is drawn from the observation that the pancreas transplant produces excellent glucose control, as attested by normal Hb A_{1c} values, in combination with the results of a large, multicentre trial with long-term follow-up which demonstrated that improved glucose control is associated with a decrease in the rate of progression of the secondary complications of diabetes in non-transplant patients. In addition, there is some direct clinical evidence to support the hypothesis that maintenance of euglycaemia after pancreas transplant may decrease the rate of progression of secondary complications. This is particularly true when the rate of progression of diabetic nephropathy in the transplanted kidney is compared at up to 3 years after transplant of either a kidney alone or a combined kidney–pancreas. There is also some suggestion that diabetic polyneuropathy may be favourably effected by pancreas transplantation. A beneficial effect of pancreas transplantation on diabetic retinopathy has not been demonstrated.

Regardless of the effect of pancreas transplantation on the secondary complication of diabetes, the quality of life is a very important issue for many patients. Subjectively, most recipients are very pleased with their sense of well being after transplant. The absence of insulin injections, the relative freedom of diet and exercise, and the absence of symptomatic hypo- and hyperglycaemic episodes all contribute to an improved quality of life.

Conclusions

Pancreas transplantation has seen progressive improvement in both patient and graft survival rates. Survival rates following combined kidney pancreas transplant are now comparable to those achieved after kidney transplant alone. Improved success has come in association with an increase in the number of transplants performed. There have now been over 10 000 pancreas transplants reported to the International Pancreas Transplant Registry and over 1000 pancreas transplants were performed in the United States alone in 1998.

Pancreas transplant can now be performed safely and successfully achieve euglycaemia in most recipients. This results in an improved quality of life, but the ability of this procedure to slow or prevent the progression of the secondary complications and prolong the lifespan of recipients remains to be defined.

Improved results in pancreas transplantation will undoubtedly lead to an increase in the number of isolated pancreas transplants that are performed, either before renal failure or after a kidney transplant. A challenging problem in this situation is the development of a non-invasive method to detect early rejection in the pancreas.

Research continues to develop islet cell transplantation into a clinical reality. When this occurs and implantation of islet cells can achieve results in terms of glucose control that are similar to those now achieved by whole organ transplant, the cell transplant procedure will replace pancreas transplantation as we now practise it. Until that time, continued improvements in the results of pancreas transplantation will hopefully augment the beneficial effect of this procedure for patients with insulin-dependent diabetes.

Further reading

Barbosa J, Steffes MW, Sutherland DER, Connett JE, Venkateswara Rao K, Mauer SM. Effect of glycemic control on early diabetic renal lesions. A 5-year randomized controlled clinical trial of insulin-dependent diabetic kidney transplant recipients. *Journal of the American Medical Association* 1994; **272**: 600–6.

DCCT Research Group T. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The Diabetes Control

and Complications Trial Research Group. *New England Journal of Medicine* 1993; **329**: 977–86.

Fioretto P, Steffes MW, Sutherland DE, Goetz FC, Mauer M. Reversal of lesions of diabetic nephropathy after pancreas transplantation [see comments]. *New England Journal of Medicine* 1998; **339**: 69–75.

Gaber AO, Shokouh-Amiri MH, Hathaway DK, *et al.* Results of pancreas transplantation with portal venous and enteric drainage. *Annals of Surgery* 1995; **221**: 613–22.

Gruessner AC, Sutherland DE. Pancreas transplants for United States (US) and nonUS cases as reported to the International Pancreas Transplant Registry (IPTR) and to the United Network for Organ Sharing (UNOS). *Clinical Transplants* 1997: 45–59.

Gruessner RW, Sutherland DE, Najarian JS, Dunn DL, Gruessner AC. Solitary pancreas transplantation for nonuremic patients with labile insulin-dependent diabetes mellitus. *Transplantation* 1997; **64**: 1572–7.

Gruessner RW, Sutherland DE, Troppmann C, *et al.* The surgical risk of pancreas transplantation in the cyclosporine era: an overview [see comments]. *Journal of the American College of Surgeons* 1997; **185**: 128–44.

Humar A, Gruessner RW, Sutherland DE. Living related donor pancreas and pancreas-kidney transplantation. *British Medical Bulletin* 1997; **53**: 879–91.

Kelly WD, Lillehei RC, Merkel FK, Idezuki Y, Goetz FC. Allotransplantation of the pancreas and duodenum along with the kidney in diabetic nephropathy. *Surgery* 1967; **61**: 827–37.

Laftavi MR, Gruessner AC, Bland BJ, *et al.* Diagnosis of pancreas rejection: cystoscopic transduodenal versus percutaneous computed tomography scan-guided biopsy. *Transplantation* 1998; **65**: 528–32.

Matas AJ, McHugh L, Payne WD, *et al.* Long-term quality of life after kidney and simultaneous pancreas-kidney transplantation. *Clinical Transplantation* 1998; **12**: 233–42.

Nathan DM. Long-term complications of diabetes mellitus. *New England Journal of Medicine* 1993; **328**: 1676–85.

Navarro X, Sutherland DE, Kennedy WR. Long-term effects of pancreatic transplantation on diabetic neuropathy [see comments]. *Annals of Neurology* 1997; **42**: 727–36.

Nymann T, Shokouh-Amiri MH, Elmer DS, Stratta RJ, Gaber AO. Diagnosis, management, and outcome of late duodenal complications in portal-enteric pancreas transplantation: case reports. *Journal of the American College of Surgeons* 1997; **185**: 560–6.

Odorico JS, Heisey DM, Voss BJ, *et al.* Donor factors affecting outcome after pancreas transplantation. *Transplantation Proceedings* 1998; **30**: 276–7.

Pirsch JD, Odorico JS, D'Alessandro AM, Knechtle SJ, Becker BN, Sollinger HW. Posttransplant infection in enteric versus bladder-drained simultaneous pancreas kidney transplant recipients. *Transplantation* 1998; **66**: 1746–50.

Reichard P, Nilsson BY, Rosenqvist U. The effect of long-term intensified insulin treatment on the development of microvascular complications of diabetes mellitus. *New England Journal of Medicine* 1993; **329**: 304–9.

Robertson RP, Sutherland DE, Kendall DM, Teuscher AU, Gruessner RW, Gruessner A. Metabolic characterization of long-term successful pancreas transplants in type 1 diabetes. *Journal of Investigative Medicine* 1996; **44**: 549–55.

Sindhi R, Stratta RJ, Lowell JA, *et al.* Experience with enteric conversion after pancreatic transplantation with bladder drainage. *Journal of the American College of Surgeons* 1997; **184**: 281–9.

Sollinger HW, Messing EM, Eckhoff DE, *et al.* Urological complications in 210 consecutive simultaneous pancreas-kidney transplants with bladder drainage. *Annals of Surgery* 1993; **218**: 561–8.

Sollinger HW, Odorico JS, Knechtle SJ, D'Alessandro AM, Kalayoglu M, Pirsch JD. Experience with 500 simultaneous pancreas-kidney transplants. *Annals of Surgery* 1998; **228**: 284–96.

Stratta RJ. Immunosuppression in pancreas transplantation: progress, problems and perspective. *Transplant Immunology* 1998; **6**: 69–77.

Stratta RJ. Mortality after vascularized pancreas transplantation. *Surgery* 1998; **124**: 823–30.

Stratta RJ, Taylor RJ, Sindhi R, Sudan D, Jerius JT, Gill IS. Analysis of early readmissions after combined pancreas-kidney transplantation. *American Journal of Kidney Diseases* 1996; **28**: 867–77.

Stratta RJ, Weide LG, Sindhi R, *et al.* Solitary pancreas transplantation. Experience with 62 consecutive cases. *Diabetes Care* 1997; **20**: 362–8.

Troppmann C, Gruessner AC, Benedetti E, *et al.* Vascular graft thrombosis after pancreatic transplantation: univariate and multivariate operative and nonoperative risk factor analysis [see comments]. *Journal of the American College of Surgeons* 1996; **182**: 285–316.

Troppmann C, Gruessner AC, Dunn DL, Sutherland DE, Gruessner RW. Surgical complications requiring early relaparotomy after pancreas transplantation: a multivariate risk factor and economic impact analysis of the cyclosporine era. *Annals of Surgery* 1998; **227**: 255–68.

US Scientific Registry of Transplant Recipients and the Organ Procurement and Transplantation Network. 1998 Annual Report. UNOS/DHHS/HRSA, Rockville, MD, 1999.

West M, Gruessner AC, Metrakos P, Sutherland DE, Gruessner RW. Conversion from bladder to enteric drainage after pancreaticoduodenal transplantations. *Surgery* 1998; **124**: 883–93.

16.10 Pancreatic islet and fetal pancreas transplantation

Derek W. R. Gray

[Rationale](#)
[Isolated pancreatic islet transplantation](#)
- [Current techniques for pancreatic islet isolation](#)
- [The prevention of rejection](#)
- [Sites of implantation of pancreatic islets](#)
- [Prospects for long-term control of glucose metabolism](#)
- [Clinical trials of isolated islet transplantation](#)

[Fetal pancreas transplantation](#)
- [Techniques of fetal pancreas transplantation](#)
- [Prevention of rejection](#)
- [Clinical trials of fetal pancreas transplantation](#)

[Future prospects for isolated islet and fetal pancreas transplantation and xenotransplantation](#)
[Further reading](#)

Although transplantation of isolated pancreatic islets in the insulin-dependent diabetic is not yet a routine clinical procedure, an extensive research literature exists and limited clinical trials are now in progress. Thus, this experimental area of transplantation needs to be considered in the context of vascularized pancreatic transplantation for type I diabetes.

Rationale

Transplantation of the vascularized pancreas has become increasingly respectable in terms of graft and patient survival, such that the results obtained now compare reasonably with those of other organ transplants. But the concept of transplanting an entire organ to cure a non-lethal condition, when what is required is only a small fraction of the transplanted tissue, would appear to be flawed. Although there is evidence that the complications of diabetes are caused by imperfect control of glucose metabolism and that improved glucose homeostasis would probably prevent their onset, there is also considerable evidence that transplantation of the endocrine pancreas at a late stage in the disease has relatively little effect on these complications, although recent evidence has suggested that diabetic nephropathy can be reversed by a successful pancreas transplant, performed at a stage well before dialysis or kidney transplantation is required. However, to make a major impact on the microangiopathic complications of diabetes, transplantation would have to be performed in the early years after presentation in order to influence the development of the complications of diabetes. Vascularized pancreatic transplantation with immunosuppression is unlikely to be an acceptable therapy for the recently-diagnosed, young diabetic, but pancreatic islet transplantation could be acceptable, especially if this could be performed without the need for long-term immunosuppression, as might be possible (see below).

The idea of transplanting insulin-secreting tissue as a free graft for diabetes is older than the discovery of insulin, but little progress was made until relatively recently. In 1965, Moskalewski described the successful isolation of islets from the rodent pancreas using collagenase digestion, after which Lacy, in 1967, perfected the isolation technique to produce sufficient islets for transplantation in rats. The enthusiasm raised by these early experimental results prompted several attempts at human pancreatic islet allotransplantation and autotransplantation. Despite claims to the contrary, it is now generally agreed that these early attempts failed, not least because the technique for islet isolation developed in the rat was not applicable to the human pancreas. These early clinical trials also emphasized the dangers of transplanting unpurified, dispersed pancreatic tissue. Over the last 30 years, techniques for the isolation and transplantation of pancreatic islets have been developed in experimental animals. Significant technical advances in that period were the demonstration of successful transplantation of unpurified islet tissue in pancreatectomized dogs; the use of intraductal collagenase for islet isolation in the dog; and successful transplantation of purified islets in the dog. This work led to the successful development of techniques for islet isolation from the human pancreas. Further modifications continue to improve both the yield and purity of islets obtained from both human and other mammalian pancreases.

An alternative to the use of isolated adult islets is the transplantation of fetal pancreas, arisng out of observations by Coupland in 1960, namely that the endocrine tissue of the fetal pancreas is relatively well developed and survives transplantation while the exocrine tissue is poorly developed and undergoes atrophy after transplantation. However, it was the late Josiah Brown who showed, in 1974, that transplantation of fetal pancreas would cure experimental diabetes in rodents.

Isolated pancreatic islet transplantation

Current techniques for pancreatic islet isolation

Although many techniques for islet isolation have been described in the past, varying from microdissection to dispersion with modified food blenders, almost all groups working with human pancreas, as well as the pancreas of other large mammalian species, now use a variant of intraductal collagenase digestion. The principle behind the technique is the delivery of the collagenase into an exocrine pancreatic duct, which allows selective digestion of the interacinar connective tissue (Fig. 1, Fig. 2, and Fig. 3). Under optimal conditions, for example collagenase delivered at 39°C, the interacinar fibrous tissue is removed within 10 to 30 min. The digestion process can be stopped by cooling at this stage, but the thicker intralobular fibrous tissue remains largely undigested and to release the islets it is necessary to employ the gentlest possible mechanical dispersion. There are now a number of semiautomated machines that use the above principles to disperse human, porcine, and canine pancreas, liberating islets mixed with a large number of exocrine fragments. No entirely successful technique has been developed to separate the exocrine tissue from the islet tissue. The most successful technique relies on the differential density of islet and exocrine tissue, which allows separation of the tissues by centrifugation on a Ficoll or albumin density gradient (Fig. 4, Fig. 5, and Fig. 6). The overall efficiency and reliability of the islet isolation process continues to show slow but steady improvement, with the main recent advance being in the development of purified components of the originally highly impure bacterial collagenase used for the digestion process.

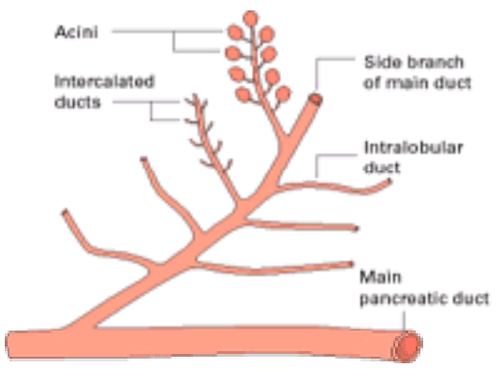


Fig. 1. The simple branch structure of the exocrine duct of the pancreas. Intralobular ducts arise from the length of the main duct and the acini are reached after two further divisions. This simple anatomical arrangement favours the dispersion of retrogradely injected fluids.

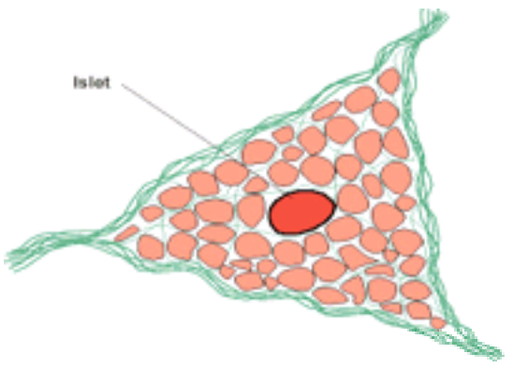


Fig. 2. The microscopic distribution of fibrous tissue (green) within the pancreas. A lobule with acini and a single islet (red) are shown. The fine interacinar islet collagen must be removed by the action of injected collagenase. The removal of the thicker interlobular and perivascular fibrous tissue would require longer digestion.



Fig. 3. Human pancreas distended by retrograde injection of collagenase into the pancreatic duct.



Fig. 4. Equipment designed to automate and speed up human islet isolation (a) stainless steel digestion/filtration chamber (b) continuous gradient automated centrifuge (COBE).

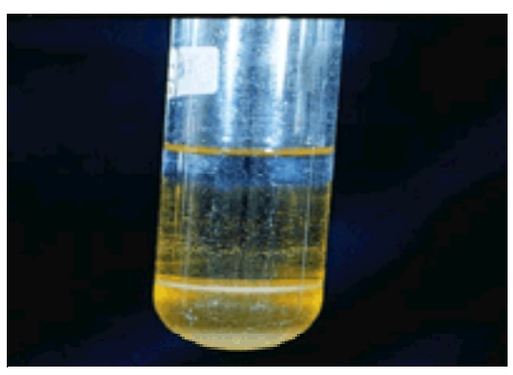


Fig. 5. Ficoll density gradient used to separate human islets from dispersed human pancreatic tissue. The dispersed pancreatic tissue is mixed with the densest layer, then progressively less dense layers of Ficoll are layered over this. After centrifugation the islets rise to a higher level than the bulk of the exocrine tissue.

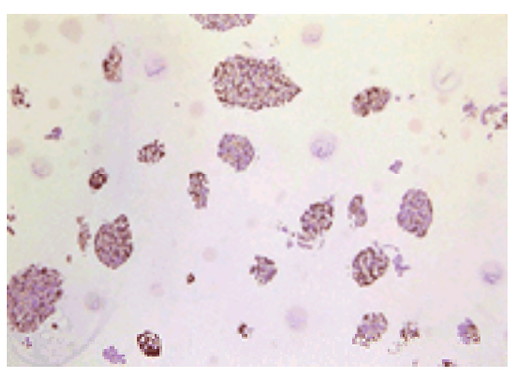


Fig. 6. Histological section of purified human islet tissue. Immunoperoxidase stain for insulin colours the islets brown. Several islets are seen to be intact yet completely separated from exocrine tissue. However, exocrine tissue is still present. Achieving full purification without an excessive loss of yield remains a problem.

The prevention of rejection

Although some endocrine tissues are relatively non-immunogenic, there is no doubt that isolated pancreatic islets transplanted as a free graft induce a rejection reaction. The mechanisms of rejection may be somewhat different to those of a vascularized organ allograft and more akin to those seen in skin allograft rejection. Although prolonged survival of islet allografts has been obtained with a number of immunosuppressive protocols in several animal models, permanent survival of grafts has been difficult to achieve, even in rodent models where tolerance to a vascularized graft has been produced. However, the introduction of cyclosporin, particularly using high-dose parenteral regimens, has resulted in the long-term survival of allografts in rat and dog models. this has also been reported for a related drug, tacrolimus, which is also a calcineurin inhibitor. Cyclosporin or tacrolimus-based protocols, usually as part of a triple or quadruple therapy regimen, are currently being used for immunosuppression in clinical trials of isolated pancreatic islet transplantation.

One of the most exciting aspects of islet transplantation has been the possibility of abrogating the immune response to pancreatic islets by altering their immunogenicity before transplantation. This has been achieved experimentally by tissue culture for several days, either in routine culture conditions or at room temperature, or by treatment with ultraviolet light, high oxygen tension, or antibodies specific for interstitial dendritic cells. Islet allografts treated in these ways have shown delayed or absent rejection in rodent and canine models. These studies have obvious clinical potential, allowing the treatment of diabetics by islet transplantation without the need for long-term immunosuppressive drug therapy. Another possible approach to the prevention of rejection is the microencapsulation of the islets in individual membranes that are impervious to larger molecules, for example antibiotics and lymphocytes, yet are permeable to nutrients and insulin secreted by the islets. Animal models suggested that this technique could allow transplantation without the need for immunosuppressive therapy, and might even allow the transplantation of islet xenografts from animals such as the pig. However, our understanding of graft rejection is advancing rapidly and it is now realized that the processes underlying the immunomodulation and encapsulation approaches are more complex than was once thought.

Sites of implantation of pancreatic islets

Islets have been implanted in most possible sites in experimental rodent models with variable success, but studies in large animal models suggest that the most acceptable sites for clinical application are the liver (via the portal vein), the spleen (either via the splenic veins or by direct puncture), or beneath the kidney capsule (Fig. 7). At present the favoured site for clinical trials is the liver, using the portal vein. This appears to be safe, provided that the preparation is relatively free of exocrine contamination. However, impure preparations could cause portal hypertension with disseminated intravascular coagulation and must be used with great caution in clinical trials.

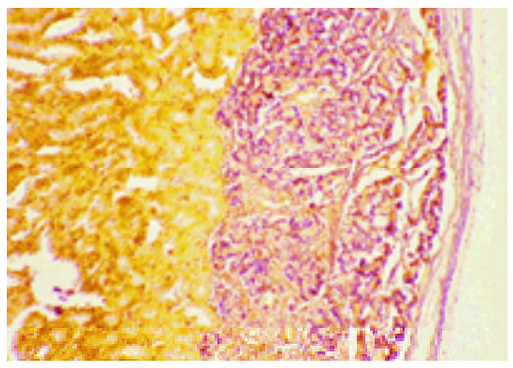


Fig. 7. Rat pancreatic islets transplanted beneath the kidney capsule (syngeneic transplant). The aldehyde-fuchsin stain for insulin is used to demonstrate insulin-containing b-cells, which stain dark mauve.

Prospects for long-term control of glucose metabolism

Transplantation of isolated pancreatic islets can produce virtually normal glucose metabolism in rodent models of experimental and autoimmune diabetes. But in larger animal models the blood sugar, although remaining within the normal range during normal nutrition, becomes abnormal in the presence of stress, probably due to a relatively reduced islet cell mass. Whether this abnormality would be important for prevention of the long-term complications of diabetes, which is the ultimate goal of islet transplantation, remains to be seen.

Animal models of autologous islet transplantation have shown that grafts have tended to fail after 1 to 3 years. This may be due to a process of exhaustion, related to a reduced islet cell mass and the site of implantation. However, islet autotransplantation procedures following total pancreatectomy in patients with severe pain due to pancreatitis have shown function beyond 8 years, reducing the worry that islet grafts cannot function in the long-term when separated from their natural position.

Clinical trials of isolated islet transplantation

As described above, the early trials of clinical islet transplantation were flawed by uncertainty about the viability and actual islet content of the tissue being transplanted. These experiments highlight a problem that has dogged islet transplantation research in general: how to express the actual quantity, the purity, the extent of exocrine contamination, and the viability of the isolated islet tissue in a standard format. Considerable progress has been made in producing more standardized methods of assessment that allow better comparisons between laboratories. Of particular importance has been the development of specific and rapid stains, such as dithizone, for the accurate identification of islet tissue. Over the past 7 years, there have been further clinical trials from centres such as St Louis, Edmonton, Giessen, Milan, and Miami, where the mass of islet tissue, and the purity and viability of the tissue, have been better documented. Most transplants were performed in diabetic patients with renal failure who also underwent kidney transplantation at the same time, but more recent transplants have been performed in diabetic patients with stable, long-term renal allografts. In recent trials the intraportal route of implantation in the liver has been most often used, relying on cyclosporin-based immunosuppression with the addition of antithymocyte preparations to prevent rejection.

The results of these clinical studies have been exciting in one way, since the feasibility of reversing the metabolic abnormalities of type 1 diabetes by islet transplantation has now been proven, with over 13 per cent of patients receiving islet transplants achieving insulin independence, about one-third of patients showing evidence of graft function as defined by detection of C peptide (Fig. 8), and graft function shown to be capable of continuing beyond 3 years (Table 1). However, the overall experience has, ultimately, been disappointing since the proportion of successful islet grafts, defined as insulin independence at 1 year post transplant, remains below 10 per cent (Table 1), which is unacceptable for introduction as a routine clinical procedure. However, research to improve these results continues in several centres.

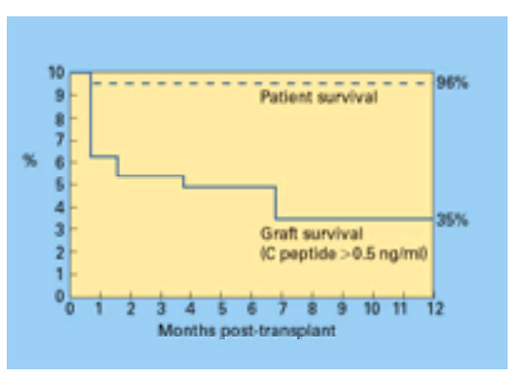


Fig. 8. Plot of patient and graft survival 1990–1997 following cadaveric islet allotransplantation (redrawn with permission of the Islet Transplant Registry, University of Giessen, Germany).

Number of cases	
Total number of cases	252
Institutions registering a transplant in last 3 years	13
Centres performing >30 cases:	Giessen (57 cases)
	Minneapolis (31 cases)
	Pittsburgh (25 cases)
	Milan (23 cases)
Results	
Insulin independent at any point	33/252 (13%)
Insulin independent at 1 year	19/230 (8%)
Longest insulin independence	4 years 2 months
Insulin independent after 1–8 Tx	17/155 (11%)

Data provided by the Islet Transplant Registry, University of Giessen, Germany

Table 1 Adult islet allografts in patients with type-1 diabetes 1990–1998

A useful spin-off from islet transplantation research has been the ability to isolate islets from the pancreas of patients that require total pancreatectomy for non-malignant conditions such as chronic pancreatitis. Infusion of the islets into the portal vein can then prevent the patient becoming diabetic, or at least reduce the severity of diabetes. This procedure, known as islet autotransplantation, has been pioneered in Minneapolis and is now performed in over a dozen centres world-wide.

Functional islet autografts have been documented beyond 8 years. However, the procedure has risks associated with the difficulties of purifying the tissue and the effect of intraportal injection of impure preparations, and should not be undertaken except by those with considerable experience in the field. There is also controversy over the indications for the procedure since there is no doubt that the best results are reported from centres that appear to perform total pancreatectomy at an earlier stage in the natural history of chronic pancreatitis than is normal practice in many centres. Those centres performing the procedure argue that they are relieving their patients of intolerable and otherwise incurable symptoms at an appropriately early stage.

Fetal pancreas transplantation

A major advantage of fetal pancreas transplantation over adult islet transplantation is that minimal processing of the islet tissue is required for transplantation. Although there is a potentially larger supply of human fetal pancreas than human adult pancreas (at least in those countries that permit abortion), several fetal pancreases are required in experimental models of diabetes to correct the diabetic state. Furthermore, the moral and ethical issues of using fetal tissue for transplantation are considerable. Another advantage of fetal pancreatic transplantation is the tremendous potential for growth of the fetal pancreas, unlike adult islets. However, immature fetal islet tissue does inhibit normal insulin secretion in response to glucose. For these reasons a long period of inadequate function is to be expected after transplantation of fetal pancreas into adult recipients, a prediction confirmed in rodent experiments.

Techniques of fetal pancreas transplantation

In comparison to adult islet transplantation, progress in fetal pancreas transplantation is less advanced. Functioning fetal pancreas transplants, both isografts and allografts, have been described and studied in detail in rodent models but, apart from successful fetal pancreatic transplantation in the pig, it has not been possible to repeat these findings in other large animal models. Some poorly documented, successful human cases have been reported from China, but none elsewhere. The fetal pancreas can be transplanted intact in the mouse, or in a few segments in the rat. The tissue is then small enough to survive as a free graft, the most successful site for implantation being under the kidney capsule (Fig. 9). In larger animals, the fetal pancreas is too large to survive as a free graft and must be dispersed, either by sectioning the pancreas into small particles or by digesting it with collagenase. The former provides less usable tissue, whereas it is possible to disperse the tissue fully using collagenase. After a period in tissue culture, which appears to favour islet tissue survival, the dispersed islet tissue forms ‘proislets'. This process has proven particularly successful in the preparation of pig fetal tissue, and has been adapted for neonatal pig pancreas with similar results (but much greater yield).

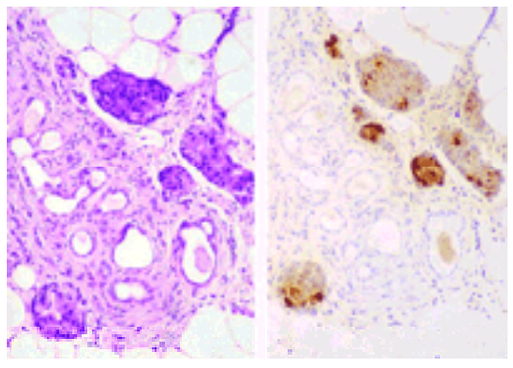


Fig. 9. Fetal rat pancreas transplanted beneath the renal capsule (syngeneic transplant). Adjacent sections are stained with haematoxylin/eosin and the immunoperoxidase stain for insulin. Note the well-preserved islets surrounded by fibrous tissue and ductal remnants. The acinar tissue has atrophied completely.

Prevention of rejection

As was the case for adult islets, initial hopes that fetal pancreatic tissue would be non-immunogenic have proved unfounded, and allografted or xenografted fetal pancreas is certainly rejected vigorously. Rejection can be prevented by various immunosuppressive protocols, although interpretation of these experiments is made difficult by the prolonged time it takes for the grafts to function after transplantation. Tissue culture in the presence of high oxygen tension has been used to reduce the immunogenicity of fetal pancreas grafts, as for adult islets. Long-term survival of allografted fetal pancreas grafts after culture has been described in mice, although this effect is markedly strain-dependent and has not been reproduced in the rat. Histological evidence that these culture techniques can prevent or delay rejection in large animals has also been presented, but there are no functioning models of fetal pancreas transplantation in large animals as a final confirmation.

Clinical trials of fetal pancreas transplantation

There have been a remarkable number of trials of fetal pancreas transplantation in man, stretching back some 15 years. The greatest number have been performed in China and the Eastern bloc countries, and, despite claims of success in some patients, objective evidence of function has not been provided. Clinical trials have been undertaken in Sydney and Denver, as well as sporadic attempts elsewhere. The sites that have been used for transplantation include intraportal injection, muscle pockets, the kidney capsule, and the intra-abdominal omentum. A combination of tissue culture prior to transplantation with a cyclosporin-based immunosuppressive protocol after transplantation has been used to try to prevent rejection. To date, there has been no evidence of continued function either in terms of C peptide production or decreased requirements for insulin therapy.

Future prospects for isolated islet and fetal pancreas transplantation and xenotransplantation

That clinical development of isolated pancreatic islet transplantation (to the point where definite, albeit short-term, function has been produced in a few patients) is certainly an exciting advance. However, the function of these transplanted human islets has been less than satisfactory and a major contributing factor has been a relative lack of tissue for the purpose of both research and transplantation. Recent advances in our understanding of the processes of rejection have raised the possibility of using a non-human tissue source, from an animal such as the pig. Of particular interest has been the potential for altering the donor tissue by genetic engineering to reduce the rejection response after transplantation into man. There are many issues of safety and ethics to be considered, but this approach is under active investigation in a number of centres; indeed actual clinical xenografts of fetal porcine islets have been performed in Stockholm (so far without success).

Further reading

Historical references

Bretzel RG, Brandhorst D, Brandhorst H, Eckhard M, Ernst W, Friemann S, *et al.* Improved survival of intraportal pancreatic islet cell allografts in patients with type-1 diabetes mellitus by refined peritransplant management. *Journal of Molecular Medicine Berlin* 1999; **77**:140–3. [State-of-the-art results, 1999.]

Brown J, Molnar IG, Clark W, Mullen Y. Control of experimental diabetes mellitus in rats by transplantation of fetal pancreases. *Science* 1974; **184**: 1377–9. [First description of function after fetal pancreas transplantation.]

Coupland RE. The survival and growth of pancreatic tissue in the anterior chamber of the eye of the albino rat. *Journal of Endocrinology* 1960; **20**: 69–77. [First description of fetal pancreas surviving transplantation.]

Cretin NLBh, Fournier B, Caulfield A, *et al.* Human islet allotransplantation: world experience and current status. *Digestive Disease* 1998;**15**: 656–62. [Review paper summarizing human islet transplantation to 1998.]

Gray DW, Morris PJ. Transplantation of isolated pancreatic islets. In: Groth CG, ed. *Pancreatic transplantation*. W. B. Saunders, Philadelphia, 1988: 363–90.

Gray DWR, McShane P, Grant A, Morris PJ. A method for isolation of islets of Langerhans from the human pancreas. *Diabetes* 1984; **33**: 1055–61. [First description of large-scale human islet isolation.]

Gray DWR, Morris P. Islet transplantation. In: Alberti, K.G.M.M. *et al.* (eds) *International textbook of diabetes mellitus* (2nd edn). John Wiley and Sons, Chichester, 1997: 965–80. Gray DW, Morris PJ. Transplantation of isolated pancreatic islets. [More comprehensive information regarding islet transplantation.]

Horaguchi A, Merrell RC. Preparation of viable islet cells from dogs by a new method. *Diabetes* 1981; **30**: 455–8. [First description of intraductal collagenase technique.]

Lacy PE, Kostianovsky M. Method for the isolation of intact islets of Langerhans from the rat pancreas. *Diabetes* 1967; **16**: 35–9. [First description of large scale islet isolation from rodent pancreas.]

Lanza RP, Chick WL (eds). *Pancreatic islet transplantation.* RG Landes, Austin, Texas, 1994. [A comprehensive textbook.]

Meyer C, Hering BJ, Grossmann R, Brandhorst H, Brandhorst D, Gerich J, *et al.* Improved glucose counterregulation and autonomic symptoms after intraportal islet transplants alone in patients with long-standing type I diabetes mellitus. *Transplantation* 1998; **66**:233–40. [Islet transplantation brings benefit for a small group of 'brittle diabetes' patients, despite failure to achieve insulin independence.]

Mirkovitch V, Campiche M. Absence of diabetes in dogs after total pancreatectomy and intrasplenic autotransplantation of pancreatic tissue. *Transplantation Proceedings* 1977; **9**: 321–3. [First successful experimental islet transplantation in large mammal.]

Morris PJ, Gray DWR, Sutton R. Pancreatic islet transplantation. *British Medical Bulletin* 1989; **45**: 224–41. [A review of islet transplantation.]

Moskalewski S. Isolation and culture of the islets of Langerhans of the guinea pig. *General and Comparative Endocrinology* 1965; **5**: 342–53. [First description of collagenase used to isolate islets.]

Noel J, Rabinovitch A, Olson L, Kyriakides G, Miller J, Mintz DH. A method for large-scale high-yield isolation of canine pancreatic islets of Langerhans. *Metabolism* 1982; **31**: 184–7. [First descrption of method for preparation of semi-purified islets from large mammal pancreas.]

Sutherland DE. Pancreas and islet transplantation. I. Experimental studies. *Diabetologia* 1981; **20**: 161–85. [A comprehensive summary of the early literature on experimental islet transplantation.]

Sutherland DE. Pancreas and islet transplantation. II. Clinical trials. *Diabetologia* 1981; **20**: 435–50. [A comprehensive summary of the early literature on clinical islet transplantation.]

Warnock GL, Rajotte RV. Critical mass of purified islets that induce normoglycemia after implantation into dogs. *Diabetes* 1988; **37**: 467–70. [A benchmark study quantitating the requirements for reversal of diabetes by islet transplantation in large animals.]

Recent articles of interest

Bretzel RG, Brandhorst D, Brandhorst H, *et al.* Improved survival of intraportal pancreatic islet cell allografts in patients with type-1 diabetes mellitus by refined peritransplant management. *Journal of Molecular Medicine, Berlin* 1999; **7**: 140–3.

Cattral MS, Lakey JRT, Warnock GL, Kneteman NM, Rajotte RV. Effect of cryopreservation on the survival and function of murine islet isografts and allografts. *Cell Transplantation* 1998; **7**: 373–9.

Lakey JRT, Cavanagh TJ, Zieger MAJ, Wright M. Evaluation of a purified enzyme blend for the recovery and function of canine pancreatic islets. *Cell Transplantationt* 1998; **7**: 365–72.

Oberholzer J, Triponez F, Lou J, Morel P. Clinical islet transplantation: a review. *Annals of the New York Academy of Sciences* 1999; **875**:189–99.

Thomas FT, Ricordi C, Contreras JL, Hubbard WJ, Jiang XL, Eckhoff DE, *et al.* Reversal of naturally occuring diabetes in primates by unmodified islet xenografts without chronic immunosuppression. *Transplantation* 1999; **67**: 846–54.

16.11 Xenogeneic transplantation

Hugh Auchincloss, Jr.

Introduction
Clinical history
Fundamental differences between xenogeneic and allogeneic transplantation
Mechanisms of xenograft rejection
Hyperacute rejection
Delayed xenograft or acute vascular rejection
Acute (T cell-mediated) rejection
Other barriers to xenotransplantation
Physiological function in a xenogeneic environment
Infectious risks
Prospects for xenotransplantation
Further reading

Introduction

Xenogeneic transplantation is the transplantation of organs or tissues between donors and recipients of different species. A discussion of xenogeneic transplantation, or ‘xenografting', enters a textbook of surgery mostly as history or as speculation, since clinical xenografting is only rarely taking place as this chapter is written. Still, the notion that animals might donate organs to human beings is fascinating and has received increasing attention (see Fig. 1) as the shortage of human organs has grown more severe. Continuing research advances in the field suggest that clinical xenografting will eventually be performed successfully.

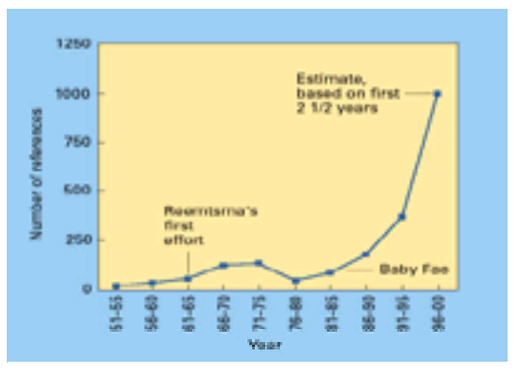


Fig. 1. Fifty years of research on xenotransplantation.

This chapter will briefly review the past history of xenogeneic transplantation and discuss some of the special immunological and non-immunological barriers to xenografting. It will also consider the prospects for successful xenogeneic transplantation in the future.

Clinical history

Xenogeneic transplantation has been attempted approximately 50 times in human beings. None of the organ transplants has survived for a full year, but one patient did maintain good function with a chimpanzee kidney for 9 months until the patient died. In addition, neural cells from a pig were shown at autopsy examination to have survived for 8 months in a patient treated for Parkinson's disease.

Reemtsma was the first to report a xenograft of a chimpanzee kidney to a human in 1963. Starzl reported a clinical series of baboon kidney transplants shortly afterward. Starzl also performed three chimpanzee-to-human and two baboon-to-human liver transplants without success. A xenogeneic heart transplant was first reported in 1964 by Hardy. Several additional heart transplants were attempted, including the case of ‘Baby Fae' in 1985. Most of these clinical efforts involved closely related, non-human primate donors and most were performed during the 1960s. More recently, a number of groups have performed cellular xenotransplants involving neural cells, bone marrow, or islets, and in many cases these clinical efforts have involved pigs as donors. Other variations on xenografting have included temporary perfusion of animal organs, temporary (or ‘bridge') xenotransplants, and the use of cells from animals as components in a variety of ‘bioartificial' organs.

Fundamental differences between xenogeneic and allogeneic transplantation

While it is important to emphasize the differences between the rejection of xenografts and allografts, it is probably more accurate to stress the similarities. In general, xenografts are subject to the same mechanisms of immune rejection as are allografts, but more so. Indeed, the more powerful responses they evoke have provided evidence for types of rejection mechanisms that are only rarely encountered in allogeneic transplantation, and have thus taught us a great deal about basic transplantation immunology.

Two key elements form the basis for the differences between xenograft and allograft rejection: first, xenografts tend to express more new antigens than do allografts; and second, incompatibilities between the molecules of two different species tend to disrupt a series of receptor–ligand interactions that are important in regulating immune responses. This second element is truly at the heart of what is special about xenogeneic transplantation, since molecular incompatibilities between species are critical not only in modifying immune responses, but also in determining whether xenogeneic organs will function physiologically in their new environment. Understanding the two key features that distinguish xenogeneic and allogeneic transplantation is also important for understanding many of the continuing efforts to achieve successful xenografting, since these efforts are often designed to diminish the number of new antigens or to restore the function of regulatory molecular interactions.

Mechanisms of xenograft rejection

The new antigens and the molecular incompatibilities of xenografts lead to at least three types of rejection, each of which has its counterpart in allogeneic transplantation.

Hyperacute rejection

Hyperacute rejection causes destruction of primarily vascularized grafts within minutes to hours after transplantation, and is characterized by interstitial haemorrhage and vascular thrombosis. The process is initiated by an immediate form of endothelial activation (type I) that causes separation of endothelial cells (allowing interstitial haemorrhage) and the loss of anticoagulant factors (leading to thrombosis). Type I endothelial activation is caused by complement fixation, particularly of the terminal membrane-attack complex. In most cases, this complement activation is triggered by the binding of preformed antibodies to endothelial antigens, although in some species combinations the alternative pathway of complement activation may lead to endothelial activation in the absence of antibody binding.

In allogeneic combinations, hyperacute rejection occurs when there are blood-group incompatibilities (in which case the recipient has preformed antibodies to carbohydrate blood-group antigens expressed on the endothelium), or when the recipient has developed antibodies to protein antigens of the major histocompatibility complex (**MHC**) of the donor (as a result of prior exposure to allogeneic tissue through pregnancy, blood transfusions, or previous organ transplantation). However, hyperacute rejection is rarely encountered in allogeneic transplantation because blood-group compatibility can be assured and because potential recipients are tested for anti-MHC antibodies against their donor by a ‘cross-match' assay. In addition, some organs (such as the liver) are highly resistant to the development of hyperacute rejection and in all organs the tendency toward complement activation is countered by the function of a series of endothelial-based, complement regulatory molecules.

As a result, hyperacute rejection is not a universal outcome in the face of alloantibodies, even for susceptible organs such as the kidney.

In xenogeneic combinations, the mechanisms responsible for hyperacute rejection are the same and, thus, as in allogeneic combinations, it typically does not occur in closely related species combinations (called ‘concordant'). However, in more distantly related combinations (called ‘discordant'), the presence of new antigens and of molecular incompatibilities between the species makes the process nearly universal. For example, in the pig-to-human combination, the new antigen is a Gal-a1,3-Gal carbohydrate determinant expressed on pig endothelium as a result of the function of a galactosyltransferase gene, whose function has been lost in higher primates. Essentially, this determinant represents a new blood-group antigen, expressed by all pigs but no humans. As for other blood-group antigens, ‘natural antibodies' develop to the pig foreign determinant as a result of cross-reactions with carbohydrate determinants expressed on common environmental micro-organisms. In addition to this universal blood-group incompatibility between pigs and humans, the process of hyperacute rejection is amplified substantially by the failure of porcine endothelial complement regulatory proteins [such as decay accelerating factor (CD56), CD59, and membrane cofactor protein (CD46)] to function effectively with human complement molecules. As a result, there is reduced counter-regulation of complement activation when pig endothelium is perfused by human blood.

For many years it appeared that hyperacute rejection in discordant species combinations was an overwhelming obstacle to xenotransplantation. However, it has become possible to avoid this mechanism of rejection in pig-to-primate transplants by removing the preformed natural antibodies by perfusing the recipient's blood through pig organs or through columns expressing the a-Gal determinant, before placing the organ transplant. More recently, it has become possible to use genetic engineering techniques to modify donor animals so that they do not express the new antigens or to restore the regulatory functions that are lost in xenogeneic combinations. For example, transgenic expression of human CD56 or CD59 on pig endothelium has been found to prevent hyperacute rejection after pig-to-primate transplantation. In addition, expression of the gene for human H-transferase (which produces a human blood-group determinant from the same substrate used by the pig galactosyltransferase) has been found to diminish the expression of the a-Gal determinant substantially. In the future, it may be possible to eliminate the function of the pig gene entirely by genetic engineering technology. Even today, the combination of an understanding of what causes the vigorous hyperacute rejection in discordant species combinations and the use of modern techniques of genetic engineering have made it possible to overcome hyperacute rejection in clinically relevant models of xenogeneic transplantation. This achievement has been one of the most dramatic developments in the field of xenotransplantation during the past decade.

Delayed xenograft or acute vascular rejection

Under a variety of circumstances in which hyperacute rejection is avoided, the second immunological mechanism of xenograft rejection occurs within a few days after transplantation. This type of rejection is less well understood and has been given a variety of different names, including ‘delayed xenograft rejection', ‘acute vascular rejection', or ‘accelerated vascular rejection' (**DXR/AVR**). It is manifested clinically by fibrinoid necrosis of the donor's vessels and intravascular thrombosis. The fundamental basis for the process appears to be a second type of endothelial activation (type II) that occurs more slowly than in hyperacute rejection and involves the transcription of genes (especially those controlled by NF-kB) and the synthesis of proinflammatory and procoagulant molecules, such as interleukin 1, tissue factor, and various cell-surface selectins. Whereas type I endothelial activation is initiated primarily by complement, type II is initiated primarily by antibody binding (and can thus occur even in the absence of complement).

Type II endothelial activation usually requires a very early antidonor antibody response, and thus rejection by this mechanism rarely occurs following allogeneic transplantation. However, in a few cases where antidonor antibodies already exist (but at levels too low to generate a positive cross-match or hyperacute rejection), accelerated vascular rejection in allografts can occur. This is perhaps most common in cases where previous sensitization is suggested by an ‘historically positive' cross-match, that is a cross-match which is positive using previous serum samples from the recipient, but negative with more recent samples.

On the other hand, an early antibody response following xenogeneic transplantation occurs almost universally. In discordant combinations, this is due to a rapid increase in the preformed natural antibodies (which may have been temporarily depleted to avoid hyperacute rejection), but even in concordant combinations there is a similar increase in the pre-existing antidonor antibodies that were originally too low to be detected. Thus, the routine occurrence of DXR/AVR in xenografts as compared to allografts reflects the presence of additional antigens.

In addition to the presence of additional antigens, the development of DXR/AVR is augmented in some xenogeneic combinations by the loss of regulatory responses resulting from molecular incompatibilities between the species. For example, endothelial expression of tissue-factor pathway inhibitor tends to diminish the tendency toward intravascular thrombosis associated with endothelial activation. However, pig tissue-factor pathway inhibitor fails to interact effectively with human factor Xa to accomplish this purpose. A particularly important molecular interaction which may play a part in DXR/AVR is that between killer inhibitory receptors on natural killer cells and MHC class I molecules. For example, pig cells are highly susceptible to lysis by human natural-killer cells because pig class I molecules fail to provide an inhibitory signal across this species difference. Natural killer cells may contribute to type II endothelial activation and thus to the development of DXR/AVR.

Given the importance of an early antibody response, one approach to the treatment of this type of rejection has been to use drugs that affect B-cell responses, such as cyclophosphamide. This approach has been quite effective in concordant combinations and has been useful even in pig-to-primate combinations, where drug therapy may diminish the effects of endothelial activation even when antidonor antibodies are present. However, this approach requires the use of more vigorous immunosuppression than is typically required for allotransplantation. In addition to pharmacological treatments, efforts similar to those being used for hyperacute rejection are under way to alter donor animals genetically, both to diminish their expression of the additional antigens and to restore the regulatory responses that control DXR/AVR. For example, transgenic expression of human anticoagulant molecules or of human MHC class I molecules or their homologues might prevent intravascular thrombosis or the activation of natural killer cells.

One hopeful aspect of type II endothelial activation is that the tendency toward its development does not last indefinitely. If DXR/AVR does not occur within several weeks, xenografts can survive for prolonged periods despite the presence of antidonor antibodies in the recipient. This phenomenon has been called ‘accommodation' and appears to involve changes in both the endothelium and in the character of the recipient's immune response. Thus, an additional approach to preventing DXR/AVR is to increase the expression of protective ‘antiapoptotic' genes in donor endothelium that may promote the tendency for accommodation to develop.

Acute (T cell-mediated) rejection

As for allografts, xenografts are susceptible to T cell-mediated rejection. Because of the vigorous humoral responses to xenografts described above, most xenografts are destroyed before there is time for cell-mediated responses to develop. However, tissue transplants that are not immediately revascularized (including skin grafts and various cellular transplants) are not susceptible to the mechanisms of rejection described above, and thus can be used to study cell-mediated rejection *in vivo.* Standard *in vitro* assays of cellular immunity have also been used. The most important observation from these types of studies is the basic similarity of the T-cell response to xenografts compared to allografts.

There are differences between the two cases, however, which again appear to reflect both the larger number of foreign antigens and the failure of some molecular interactions between species. The larger number of antigens stimulating T cell-mediated responses are due to the many more disparate proteins between members of different species which therefore generate ‘minor' histocompatibility antigens. This increase alone may cause the exceptional strength of T cell-mediated xenograft rejection, which typically requires higher concentrations of immunosuppressive drugs for its control than are needed for allografts.

On the other hand, the failure of molecular interactions between species has been found to alter the character of cell-mediated rejection in different ways depending on the species combination. For example, when mice receive pig skin transplants, their T cells are unable to respond to stimulation by pig antigen-presenting cells due to molecular incompatibilities between some of the accessory and costimulatory molecules that are involved in T-cell activation. As a result, rejection in this species combination is particularly dependent on the indirect pathway whereby donor antigens are presented by recipient antigen-presenting cells. On the other hand, in the clinically relevant, pig-to-primate combination these same molecular interactions are largely intact, allowing both direct and indirect recognition to occur just as for allografts. A small number of molecular incompatibilities have been identified in the pig-to-human combination, including a diminished capacity of human CD8 molecules to interact with pig class I molecules. This probably makes cell-mediated immunity especially CD4-dependent in this species combination, as it is in others. The most important molecular incompatibility identified so far in the pig-to-human combination is the failure of the killer inhibitory receptors on human natural killer cells to interact with pig class I molecules, allowing unopposed activation of natural killer cells as described above.

The therapeutic implications of our current understanding of cell-mediated xenograft rejection are not clear. It is not likely that genetic engineering can be used to diminish the large number of protein disparities that are inherent in xenogeneic combinations and, other than the interaction between killer inhibitory receptors on natural killer cells and MHC class I, it is not clear which if any defects in regulatory molecular interactions in the pig-to-human combination might be responsible for stronger rejection. On the other hand, there is interest in the possibility of augmenting the capacity of xenografts to downregulate T cell-mediated rejection by ectopic or overexpression of such molecules as soluble CTLA-4 or Fas ligand. There is also particular interest in strategies to induce T-cell tolerance in order to allow xenotransplantation to proceed without a requirement for unusually large doses of immunosuppression.

Other barriers to xenotransplantation

In addition to the immunological responses described above, there are at least two more potential barriers to the clinical application of xenotransplantation.

Physiological function in a xenogeneic environment

Just as molecular incompatibilities between species are critical in determining the nature of the immune responses to xenografts, so too such molecular incompatibilities are likely to be critical in determining which organs can function effectively in a xenogeneic environment. It has not yet been possible to determine fully the importance of this barrier to xenogeneic transplantation because long-term survival of life-sustaining, widely discordant xenografts has not often been achieved. It is known, however, that pig kidneys can maintain the life of a primate for many months, although it appears that pig erythropoietin does not function well in baboons. In addition, it seems likely that pig hearts and pig islet cells can perform their primary physiological functions in non-human primate recipients. On the other hand, several metabolic defects were detected when baboon-to-human liver transplants were performed and it is likely that many more would be encountered after pig-to-human liver transplantation. Thus, defects in the physiological function of xenografts are likely to represent a barrier to xenogeneic transplantation at least in some cases.

Infectious risks

All forms of transplantation carry the risk of transmitting infections from the donor to the recipient and the list of potential infectious agents is quite large when considering animal donors. On the other hand, if animals are bred and maintained in captivity, they can be screened to ensure the absence of essentially all known pathogens. Although the requirement for breeding in captivity has effectively eliminated, for the time being, the possibility of using non-human primates as donors for clinical transplantation, it has made the risk of transmitting a known infection substantially less for xenografts than for allografts.

The one potential infectious risk from animal donors that cannot yet be eliminated is the possibility that endogenous retroviruses might be transmitted to human recipients, where they might be, or become, pathogenic. Because the genetic sequences for these viruses are contained within the germline DNA, their transmission to offspring cannot easily be avoided even by caesarean delivery in captivity. In the case of pigs, several non-pathogenic porcine endogenous retroviruses (referred to as **PERV**s) have been identified, and these viruses are capable of infecting human cells *in vitro.* While there is currently no evidence that PERVs can infect human cells *in vivo* and no indication that they would be pathogenic or transmissible to other humans if they did, the theoretical possibility remains that they could become human pathogens by mutation or recombination. While it would seem likely that if such transmission from pigs to humans were really possible it would have occurred already in nature, the circumstances of xenogeneic transplantation (including the immunosuppressive strategies likely to be used) might generate unique conditions favouring such an event. From the point of view of the individual human recipient, the risks stemming from the remote possibility of PERV infection and mutation to pathogenicity are very small, and thus relatively unimportant. However, from the point of view of humans who are not transplant recipients, but who might be subject to infection by a new pathogen, this theoretical possibility represents a risk, however small, for which there is little corresponding benefit. In the final analysis, determining whether PERV infection represents a real barrier to xenografting can only be accomplished by performing actual clinical trials. For this reason, clinical trials of xenotransplantation have been initiated, but they have been subjected to tight controls by public regulatory agencies to ensure that recipients are screened rigorously for evidence of PERV infection.

Prospects for xenotransplantation

In the interval since the first edition of this textbook, substantial progress has been made in the field of xenotransplantation. Limited clinical trials are now in progress, although these generally involve special cases, such as fetal neural cells into the brain or the use of pig cells in bioartificial organs. At this time, we still appear to be far away from the widespread clinical use of pig organs for human patients. As this chapter indicates, xenografts generally have significant disadvantages over allografts, and thus, even though there is a severe shortage of human donor organs for our population as a whole, individual transplant recipients will still fare better waiting for an allograft in most cases. On the other hand, it is now clear that the competitive disadvantage of xenografts, both immunologically and physiologically, can potentially be eliminated or even reversed by genetic engineering and/or by the induction of tolerance. Thus, there is every reason to believe that further progress in these areas will eventually make widespread clinical xenogeneic transplantation possible.

Further reading

Auchincloss H, Jr. Xenogeneic transplantation. A review. *Transplantation* 1988; **46**: 1–20. [An early comprehensive review of the transplantation literature.]

Auchincloss H, Jr, Sachs DH. Xenogeneic transplantation. *Annual Review of Immunology* 1998; **16**: 433–70. [A review of the recent literature.]

Bach FH. Xenotransplantation: problems and prospects. *Annual Review of Medicine* 1998; **49**: 301–10. [A recent review with a different perspective.]

Bach FH, Auchincloss H, Jr, Robson SC. Xenotransplantation. In: Bach FH, Auchincloss H, Jr, ed. *Transplantation immunology*, pp. 305–38. Wiley-Liss, New York, 1995. [A chapter written for surgical residents.]

Bach FH, Dalmasso AP, Platt JL. Xenotransplantation: a current perspective. *Transplantation Reviews* 1992; **6**:163–74. [A review from an important group working on xenotransplantation.]

Bach FH *et al.* Accommodation of vascularized xenografts: expression of ‘protective genes' by donor endothelial cells in a host Th2 cytokine environment. *Nature Medicine* 1997; **3**: 196–204. [An important article describing changes in xenografts that have undergone accommodation.]

Byrne GW, McCurry KR, Martin MJ, McClellan SM, Platt JL, Logan JS. Transgenic pigs expressing human CD59 and decay-accelerating factor produce an intrinsic barrier to complement-mediated damage. *Transplantation* 1997; **63**: 149–55. [A report of results using some of the transgenic pigs for xenotransplantation.]

Chitilian HV, Laufer TM, Stenger K, Shea S, Auchincloss H, Jr. The strength of cell-mediated xenograft rejection in the mouse is due to the CD4+ indirect response. *Xenotransplantation* 1998; **5**: 93–8. [A study on the strength of the indirect response in xenograft rejection.]

Cooper DKC. Xenoantigens and xenoantibodies. *Xenotransplantation* 1998; **5**: 6–17. [A review of natural antibodies and the antigens they react with.]

Cooper DKC, Kemp E, Reemtsma K, White DJG, ed. *Xenotransplantation, the transplantation of organs between species*, 2nd edn. Springer, Berlin, 1997. [A comprehensive book with contributions from many others.]

Kopp CW *et al.* Effect of porcine endothelial tissue factor pathway inhibitor on human coagulation factors. *Transplantation* 1997; **63**: 749–58. [An important study on factors promoting thrombosis in xenotransplantation.]

Lawson JH, Platt JL. Molecular barriers to xenotransplantation. *Transplantation* 1996; **62**: 303–10. [A recent review of xenotransplantation with a different perspective.]

Moses RD, Pierson RNI, Winn HJ, Auchincloss HJ. Xenogeneic proliferation and lymphokine production are dependent on CD4+ helper T cells and self antigen-presenting cells in the mouse. *Journal of Experimental Medicine* 1990; **172**: 567–75. [An early study demonstrating the role of CD4 cells functioning through the indirect pathway in xenotransplantation.]

Oriol R, Ye Y, Koren E, Cooper DKC. Carbohydrate antigens of pig tissues reacting with human natural antibodies as potential targets for hyperacute vascular rejection in pig-to-man organ xenotransplantation. *Transplantation* 1993; **56**: 1433–42. [An early identification of the alpha gal determinant as critical in xenotransplantation.]

Palmetshofer A, Galili U, Dalmasso AP, Robson SC, Bach FH. a-Galactosyl epitope-mediated activation of porcine aortic endothelial cells. Type II activation. *Transplantation* 1998; **65**: 971–8. [A study of Type II endothelial activation.]

Patience C, Takeuchi Y, Weiss RA. Infection of human cells by an endogenous retrovirus of pigs. *Nature Medicine* 1997; **3**: 282–6. [One of the early studies demonstrating infection of human cells by endogenous retrovirus from pigs *in vitro*.]

Pierson RN, Winn HJ, Russell PS, Auchincloss HJ. Xenogeneic skin graft rejection is especially dependent on CD4+ T cells. *Journal of Experimental Medicine* 1989; **170**: 991–6. [An early study showing the particular importance of CD4 cells in xenograft rejection.]

Platt JL, Logan JS. Use of transgenic animals in xenotransplantation. *Transplantation Reviews* 1996; **10**: 69–77. [A review of the use of transgenic pigs in xenotransplantation.]

Reemtsma K *et al.* Renal heterotransplantation in man. *Annals of Surgery* 1964; **160**: 384. [One of the earliest reports of successful clinical xenotransplantation.]

Schmoeckel M *et al.* Orthotopic heart transplantation in a transgenic pig-to-primate model. *Transplantation* 1998; **65**: 1570–6. [Additional studies using transgenic pigs in primate recipients.]

Sykes M, Lee LA, Sachs DH. Xenograft tolerance. *Immunological Reviews* 1994; **141**: 245–76. [A comprehensive review of tolerance induction in xenogeneic transplantation.]

Zaidi A *et al.* Life-supporting pig-to-primate renal xenotransplantation using genetically modified donors. *Transplantation* 1998; **65**: 1584–9. [A report describing kidney transplantation using transgenic pig donors for non-human primate recipients.]

Zhao Y, Swenson K, Sergio JJ, Arn JS, Sachs DH, Sykes M. Skin graft tolerance across a discordant xenogeneic barrier. *Nature Medicine* 1996; **2**: 1211–16. [A remarkable description of tolerance

induction in a discordant species combination.]

16.12 Reconstruction of the central nervous system by neural transplantation

Cesario V. Borlongan, Thomas B. Freeman and Paul R. Sanberg

Introduction
Parkinson's disease and surgical treatment
The stereotactic techniques of neural transplantation
Neural transplantation, immunosuppression, and alternative graft sources
Neural transplantation and tropic factors
Standardized test for transplant recipients with Parkinson's disease
Double-blind, randomized clinical trials
Summary
Further reading

Introduction

Neurodegenerative disorders have been treated by a variety of surgical procedures, including transplantation in experimental models and in clinical practice. Clinical application of neural transplantation of fetal tissue has been examined as a possible therapy for Parkinson's disease. This experimental treatment has been shown to promote, at best, moderate beneficial effect depending in part on the surgical technique, the viability of the grafted tissue, the host immune response, and the different evaluation parameters used to assess clinical improvement. Clinical centers have used different methods of dissecting the ventral mesencephalon (where the dopamine-rich neurones reside) and transplantion has been performed using either solid, cell suspension, or tissue strand grafts. Disagreement also exists on whether there is a need to immunosuppress the transplant recipient. Finally, comparisons between clinical data have been partly hampered by a lack of consensus on using a standardized test for evaluating clinical improvement. Critical assessment of these issues should optimize techniques on safety and efficacy of neural transplantation for Parkinson's disease. Neural transplantation remains an experimental treatment and its extension to other human disorders should be approached with caution.

Parkinson's disease and surgical treatment

Parkinson's disease is a progressive disorder that is characterized by degeneration of the nigrostriatal dopaminergic pathway. More than 1 million American people are afflicted with Parkinson's disease and about 50 000 new cases are reported yearly. In general, its debilitating course appears unaltered by pharmacologic or surgical procedures, and thus during the progressive deterioration of the patient, treatment can only provide symptomatic palliation. Surgical procedures for Parkinson's disease include neuroablative treatments such as pallidotomy and thalamotomy, and neurorestorative therapies such as deep brain stimulation and neural transplantation. Clinical trials of pallidotomy were performed in the 1950s and were shown consistently to palliate rigidity, but not tremor. On the other hand, thalamotomy, which was initially performed in the 1950s and 1960s, was demonstrated to reduce tremor dramatically. Because tremor is more noticeable, although less disabling than rigidity, pallidotomy was abandoned in favor of thalamotomy. By the late 1960s, the advent of levodopa (L-dopa) had pushed surgical treatment for Parkinson's disease into extinction. Not only does L-dopa abolish tremor and rigidity, it also palliates bradykinesia. However, L-dopa has its own limitations including severe side-effects such as dyskinesia and psychosis, and diminishing efficacy with time. Thus there has been renewed interest in surgical treatments after the early years of experience with L-dopa.

In 1982, enthusiasm in neural transplantation for Parkinson's disease was generated by initiation of clinical trials of adrenal medullary grafts in Sweden. This subsequently gained worldwide attention in the late 1980s following the publication by the Mexican group of a dramatic restoration of function in patients with Parkinson's disease who received transplants. Based on this report, open transplantation procedures via a craniotomy were adopted in the clinic and large amounts of adrenal medulla were directly placed into the striatum. Unfortunately, most neurosurgical groups were not able to replicate the initial report of recovery induced by adrenal medulla transplant and the practice of this procedure has waned considerably. However, cografting techniques or cotreatment with growth factors have shown promise in that better survival of the adrenal medulla tissue grafts appears to occur. In particular, a apanese group (Isao Date and colleagues, Okayama University Medical School, 1996) has demonstrated in patients with Parkinson's disease an enhanced survival of adrenal medulla tissue when grafted together with peripheral nerve tissues. They also showed that grafted adrenal medulla tissues survived better *in vitro* when treated with nerve growth factor.

The use of fetal tissue for treatment of human disorders has a long history; indeed, fetal pancreatic transplants to treat diabetis mellitus were performed in 1928 and fetal thymic transplants to treat lymphopenic immunologic deficiency were initiated in 1969. The pioneering work by Elizabeth Dunn (University of Chicago) in 1904 demonstrates the ability of transplanted fetal tissue to survive in the brain of another animal, but it took another six decades before Lars Olson and Aka Seiger (Karolinska Institute, 1967), using the ‘immunoprivileged' anterior chamber of the eye as the implantation site, provided evidence that the grafted fetal tissue has the capacity to integrate with the host target neurones and that these graft–host connections are functional. The late 1970s and the early 1980s saw a surge in basic science research exploring the utility of fetal tissue transplantation in ameliorating neurologic symptoms in animal models of brain degeneration. The ultimate goal of neural transplantation of dopaminergic neurones is to reconstruct the nigrostriatal dopaminergic circuitry. The rationale for neural transplantation for Parkinson's disease is based on the disease being primarily caused by the loss of a limited number of nigrostriatal dopaminergic neurones which function in a modulatory capacity physiologically, rather than the highly specific, somatotopically organized information that other neural systems possess. Furthermore, parkinsonian symptoms are improved significantly with dopamine replacement therapy, namely L-dopa treatment. The striatum (or caudate– putamen in humans) is comparatively small, thus making it a suitable target for transplantation. Finally, the bulk of the literature on basic research on neural transplantation accumulated over 5 years or so (in early 1980s) has demonstrated that the fetal grafts survive, reinnervate the brain, and ameliorate parkinsonian symptoms in rodent and non-human primates. These experimental results provided the basis for undertaking neural transplantation in patients with Parkinson's disease.

The stereotactic techniques of neural transplantation

The surgical procedure of neural transplantation is in itself routine. However, the collection, dissection, storage, and preparation of the tissue leading to the transplantation surgery, as well as the evaluation of post-transplantation clinical benefit are not well defined at this time. Notwithstanding, direct evidence now exists on transplantation-induced clinical recovery and increased striatal fluorodopa uptake as revealed by positron emission tomography (**PET**) following transplantation of fetal dopaminergic neurones in patients with Parkinson's disease. Parallel clinical outcomes also were reported in transplanted patients with parkinsonism induced by the intake of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP).

In 1989, Swedish neurosurgeons (Ole Lindvall and colleagues, University of Lund) reported successful amelioration of parkinsonian symptoms in patients who received intrastriatal transplants of human fetal ventral mesencephalic dopaminergic neurones. Thereafter, similar surgical operations were conducted in the United States, Mexico, England, Spain, Japan, and France. More than 300 patients with Parkinson's disease around the world have received neural transplantation of embryonic ventral mesencephalic tissue. Of these patients, two have died because of the surgery, while 11 died of causes unrelated to the procedure.

The first two patients in the series of studies by Lindvall and colleagues displayed only minor changes in fluorodopa uptake on PET, but when tissue storage was shortened (see discussion on tissue storage below) and a smaller needle gauge was used, the next two patients exhibited long-term stability of fluorodopa uptake on PET. It is noteworthy that these patients received only unilateral transplants and, therefore, the contralateral non-transplanted side showed a progressive deterioration consistent with the progression of the disease. None the less, even with unilateral transplants, the patients had a pronounced clinical benefit including improvement in the percentage of time in the ‘off' state, rigidity, hypokinesia, and dyskinesia for up to 4 years, before demonstrating a gradual deterioration some 4 to 6 years after transplantation. The long-term graft survival in these patients was indirectly assessed with fluorodopa uptake on PET.

For the first time, a direct correlation of clinical benefits, graft survival, and PET changes became available in 1995 when American neurosurgeons examined the brains from two patients who died after implantation from causes unrelated to the surgical procedure. Robust graft survival was noted that was characterized by variable neuritic outgrowth (2.5 to 7 mm within the striatum and 11 to 12 mm within white matter tracts) extending towards, and forming synaptic connections with, the host tissue. The targeted area revealed confluent reinnervation of up to 78 per cent with the typical patch-matrix pattern of the nigrostriatal system. However, no host sprouting was noted. Clinical improvement and increased fluorodopa uptake on PET correlated with graft survival and host reinnervation. The grafted cells appeared not to be adversely affected by the progression of the disease as observed in the host nigral neurones. Additional pathologic features of the dopaminergic grafts include a delay in maturation rate; presence of metabolic activity; ability to take up, synthesize, and store dopamine; and presence of intact blood–brain barriers. These encouraging results provide a glimmer of hope that Parkinson's disease and other neurodegenerative disorders might be treated effectively by neural transplantation.

Transplantation protocols vary among clinical centers. Curt Freed and colleagues have used the frontal insertion of the cannula to deposit the ventral mesencephalic tissue into the striatum. This technique allows easy access to the anteroposterior striatal targets for implantation of the tissue. On the other hand, Thomas Freeman and colleagues, utilizing the dorsoventral insertion of the cannula, have gained better access to the dorsoventral striatal targets. Another technical difference between the two groups involves the tissue preparation; Freed and colleagues use tissue strand grafts, while Freeman and colleagues have opted for solid grafts. At this time, no direct comparisons on differential effect of each technique have been reported. However, a recent experimental study by Freed and colleagues showed that tissue

strands of ventral mesencephalic tissues survived better in culture and promoted more robust recovery in parkinsonian animals than cell suspensions.

The target area for transplantation is important and, in general, experimental studies have transplanted tissues into the dorsal striatum and obtained optimal functional recovery compared with transplantations into other striatal regions. Clinically, the dorsal striatum may correspond to the postcommisural putamen, which is predominantly depleted of dopamine in Parkinson's disease and receives primary input from the precentral motor fields. Freeman and colleagues have surgically targeted this region in their clinical transplantation procedures. Other clinical centers have targeted both caudate and putamen for transplantation and results from these studies should provide information on whether enhanced clinical benefit is achieved by transplantation into these two regions. Recently, laboratory studies have explored placement of grafts along the nigrostriatal dopaminergic pathway; this bridging graft method has produced robust reconstruction of the lesioned nigrostriatal pathway. However, whether this leads to enhanced functional recovery remains to be determined.

Both techniques of dorsoventral and anteroposterior implantation of grafts have transplanted embryonic tissues of 5.5 to 9 weeks postconception. Tissue donors outside this age range generally result in decreased survival of the grafts. Harvesting of the fetal tissue must coincide with the development of dopaminergic neurones in the subventricular zone and before these neurones have extended their neuritic processes. Of note, xenotransplantation experiments (human tissue into rats) have demonstrated that solid grafts have an optimal donor age of 6.5 to 9 weeks postconception, whereas suspension grafts are most successful with donor ages between 5.5 and 8 weeks. Freeman and colleagues have verified these results in humans.

Differences in dissection methods of the ventral mesencephalon have been shown to result in variable numbers of harvested dopaminergic neurones, as well as differences in viability of the graft following transplantation. Ole Isacson and colleagues have conducted a series of preclinical studies aimed at delineating the specific part of the ventral mesencephalon that will provide the optimal number of dopaminegic neurones. These authors showed that dissection of the medial ventral mesencephalon yields a higher number of dopaminergic neurones than the lateral ventral mesencephalon. Recent studies have shown also that grafts of fetal ventral mesencephalic dopaminergic neurones can survive and integrate better with the host tissue when cotransplanted with embryonic striatal cells. Indeed, a clinical trial was initiated in England using cografts of fetal ventral mesencephalon and striatum with better clinical improvement reported in patients who received the cografts compared with those who received ventral mesencephalon alone.

In the laboratory, Guiddo Nikkhah and colleagues have developed microsurgical implantation tools that can accurately target multiple sites of the striatum requiring minimal ablative surgical procedures. This microsurgical method allows transplantation of a suspension of dissociated ventral mesencephalic cells. There is some debate, however, as to whether the mechanical, as well as, enzymatic dissociation steps involved in the preparation of the cell suspension, compared with the less stringent solid or tissue strand preparation of the tissue, may actually diminish the efficacy of the graft. Furthermore, multiple needle passes may not be appropriate in the clinic because of increased risks. Freed and colleagues have deposited tissues by using 10 to 14 needle passes, while Freeman and colleagues have used a single cortical entry point with grafts injected at 5-mm intervals in a three-dimensional grid array.

Graft efficacy is affected by storage techniques (such as cool storage, ‘hibernation medium', cryopreservation) and time interval between harvesting and transplantation of the tissue. Spencer and colleagues observed only mild improvement (not significantly different from the non-transplanted patients) in patients who received cryopreserved tissues. Furthermore, minimal graft survival was observed in one patient who died at 4 months post-transplantation. Subsequent studies have shown that cryopreservation may be detrimental to the grafted fetal tissue. While cryopreservation does not decrease viability of fetal tissue after thawing and prior to transplantation, it markedly reduces graft survival and neuritic outgrowth following transplantation. On the other hand, tissue storage in ‘hibernation medium' for between 2 and 4 days appears to maintain graft viability. Enhancement of tissue viability during storage has included treatment with lazaroids, antioxidants, free radical scavengers, and tropic factors (see below). If successful tissue storage can be established, then this will allow better pooling and sorting of tissues from multiple donors, and planning for the subsequent surgical procedure.

The number of donors or the dosage of dopaminergic neurones required for effective transplantation remains to be determined. Lindvall and colleagues have transplanted tissues from as many as seven donors, Freeman and colleagues have pooled tissues from three to four donors for grafting in each side, and Freed and colleagues have used a single embryonic donor to transplant into two patients. In both clinical and laboratory studies, survival rate of grafted dopaminergic tissue is about 5 to 8 per cent. While this may represent only a small number of neurones lost in Parkinson's disease, transplanting the full complement of dopamine neurones may not be necessary to achieve clinical benefit because parkinsonian symptoms only become visible after a dopamine depletion of 50 to 80 per cent. Thus, the threshold for dopamine replacement therapy in Parkinson's disease may require only a small number of viable grafted dopaminergic neurones.

Owing to these many variables that may interfere with the functional effects of the graft, coupled with the fact that each clinical center has its own transplantation protocol, it is difficult to interpret inconsistencies in the clinical results of neural transplantation. Preclinical studies in experimental models should be encouraged to reveal which of these factors need to be considered in the clinic.

Neural transplantation, immunosuppression, and alternative graft sources

The central nervous system is an immunoprivileged site, in that grafts may survive in the brain for a prolonged period of time. The immunoprivileged status of the brain is not absolute, and thus immunosuppressive drugs may be required for better graft survival. Cyclosporin A has been the treatment of choice for organ, as well as, neural transplantation. Long-term immunosuppression (for at least 6 months after transplantation) has become an adjunct therapy in most patients receiving transplants for Parkinson's disease. However, Freed and colleagues have shown that immunosuppression may not be needed, and in fact have demonstrated that graft survival does not correlate well with administration of immunosuppression. Freeman and colleagues have administered immunosuppressive drugs to all their patients in the early weeks, but their two patients who revealed increased fluorodopa uptake on PET and robust graft survival received multiple unrelated allografts in the absence of long-term immunosuppression. Preclinical studies support the use of immunosuppressive drugs to enhance survival of the graft (especially those from cross-species donors). Interestingly, the observed increased viability of the transplanted tissue may be related not only to the immunosuppressive effect of the drug, but to its additional neurotropic effect. Recently, cyclosporin A, tacrolimus (FK-506), and their analogs have been demonstrated to promote a tropic effect in parkinsonian animals, and clinical trials are under way. Because of this additional function of certain immunosuppressive drugs, it is possible that these drugs when combined with grafts can promote better survival and even enhance functional recovery in transplant recipients. Indeed, the present authors and others have shown that cyclosporin A enhanced immunoreactivity of tyrosine hydroxylase in normal rats and may be neuroprotective against neurotoxic lesions of the nigrostriatal dopaminergic pathway. Thus, while immunosuppression *per se* may not be needed in neural transplantation, enhancement of graft survival and integration with the host tissue can be facilitated by immunosuppressant-released tropic factors. The limitation of cyclosporin A is that it induces nephrotoxicity and hallucination in some patients. To circumvent these side-effects, other approaches are being developed including eliminating the immunosuppressive property of the drug while retaining its neurotropic feature. The present authors and others have also resorted to finding alternative immunosuppressive agents such as the testis-derived Sertoli cells (which can promote localized immunosuppression), encapsulation of the graft, and masking of the major histocompatibility complex of the graft. Because human fetal tissue may not be the most suitable graft source, these alternative immunosuppressive agents will play a major role in the survival of xenografts (such as porcine cells, genetically engineered cells, and other cell lines). Indeed, Toru Itakura and colleagues have transplanted autologous cervical sympathetic ganglion, and Isacson and colleagues (1996) have transplanted fetal porcine ventral mesencephalic cells into patients with Parkinson's disease. Most recently, attention has focused on using neural stem cells as a source of neural tissue for transplantation therapy. For the first time, cloned human neural stem cells from the ventricular zone of fetal telencephalon have been shown to differentiate into neurones and glia when plated into serum-containing medium after epigenetic or genetic propagation, to develop into regionally specific cell types when implanted into the brain of newborn animals, and to substitute for missing neurones when grafted into a mutant animal with cell deficiency. The demonstration of such multipotentiality in human neural stem cells may be an efficacious alternative to using primary fetal tissue for neural transplantation therapy.

Neural transplantation and tropic factors

In the last 5 years, tropic factors have become useful adjuvant agents in neural transplantation. More recently, glial cell line-derived neurotropic factor (**GDNF**), a member of the transforming growth factor-b superfamily, has been shown to be a specific survival factor for midbrain dopaminergic neurones both *in vitro* and *in vivo*, as well as, in immature and adult animals. Indeed, clinical trials of GDNF in patients with Parkinson's disease began in 1998 at the University of Hawaii. Preclinical experiments have demonstrated that GDNF can protect against MPTP and 6-hydroxydopamine (6-OHDA)-induced damage in the substantia nigra and in the striatum. Moreover, GDNF as a pretreatment for or cotreatment with fetal grafts has been shown to enhance graft survival as well as graft-induced functional effects. A major obstacle in GDNF treatment is its delivery because it does not easily cross the blood–brain barrier. However, new methods, including GDNF conjugation with antibodies to transferrin receptors or incorporation in fibrin glue for slow release, are being explored to circumvent this barrier. Other tropic factors that have been reported to exert tropic and protective effects on dopaminergic neurones, as well as potentiate graft-induced effects, include nerve growth factor and brain-derived neurotropic factor. Stereotactic surgery in combination with neurotropic factor application may prove beneficial for treatment of Parkinson's disease.

Standardized test for transplant recipients with Parkinson's disease

There are many ‘standardized' rating systems for evaluation of transplant recipients with Parkinson's disease and these include the Unified Parkinson's Disease Rating Scale, the Schwab–England Disability Scores, and the Core Assessment Program for Intracerebral Transplantation (**CAPIT**). The CAPIT is supposedly the standardized rating system that was developed to allow comparisons across clinical studies and serves as a registry, but it appears that concession has not been reached among clinical centers to use this test for all neural transplantation protocols. Thus, it is still difficult to analyze data accumulated from different clinical centers and such lack of co-ordination can hinder the evaluation process of the safety and efficacy of neural transplantation. It is hoped that the establishment of scientific

organizations such as the American Society for Neural Transplantation and Repair and the Network of European CNS Transplantation and Restoration will foster better communication among clinical centers. Because of the enthusiasm for the rapid development of this therapy, each clinical center wants to guard its own intellectual as well as technical investment. However, dissemination of basic as well as clinical data is needed to advance this neural transplantation technology. The use of CAPIT by all clinical centers should be viewed as a minimum requirement to proceed with clinical trials of neural transplantation. The fluorodopa uptake on PET analysis can serve as a definitive marker for graft survival, but fluorodopa uptake cannot differentiate between the grafted dopamine neurones and the terminal sprouting induced by the residual host dopamine neurones. While advocating the use of a standardized test, clinicians should be encouraged to conduct additional tests to determine the efficacy of transplantation. For example, the preliminary report on examination of speech and voice in transplant recipients may someday be combined with CAPIT to reveal movement and speech effects of transplantation therapy.

Double-blind, randomized clinical trials

The National Institutes of Health (**NIH**) are currently funding two clinical trials of neural transplantation for Parkinson's disease, one by the University of South Florida and the other by the University of Colorado. The design of these clinical trials has caused a major debate about their suitability for funding. Ethical concerns were raised when the NIH recommended a placebo control group of patients (those that will be exposed to the surgical procedure but will receive no transplants) to be included in the clinical design. The inclusion of this control group has been a routine procedure in evaluating clinical efficacy of new drugs. While it has been reported that drilling holes in the skull is not dangerous, there is still a controversy over the benefits of undertaking such placebo controls. The rationale for using placebo controls is ultimately to delineate transplant efficacy from placebo effect. Indeed, 30 per cent of medically treated patients have displayed improvement following a placebo treatment. The experience with the much publicized, but later aborted (because some people died from the surgical procedure), adrenal gland transplantation for Parkinson's disease also puts pressure on examining the placebo effect. Accordingly, a few patients who undergo placebo treatment can save several thousand patients if neural transplantation is seen not to be efficacious. A limited placebo testing may include drilling holes in the skull plus placebo injection. In addition to the patient being exposed to the placebo treatment, the doctors must also take part in maintaining such double-blind, randomized trials. Notwithstanding, the patients receiving placebo treatment are at some risk (for instance from skin/bone infections, sedative side-effects) and they need to be informed about these adverse affects. Alternatively, experiments on non-human primates at Yale University and the University of Kentucky should provide some definitive answers on the efficacy of neural transplantation, thus avoiding unnecessary harm to human patients.

Summary

A strong research program in basic science is needed to complement the design of clinical studies of neural transplantation. A continued integration of laboratory and clinical research would promote the establishment of a safe and an effective neural transplantation therapy for Parkinson's disease. The efficacy of neural transplantation has recently been explored in Huntington's disease and ischemic stroke. This surgical treatment would benefit from refinement of technical expertise as well as standardization of clinical protocols.

Further reading

Bakay RAE, Sladek JR Jr. Fetal tissue grafting into the central nervous system: yesterday, today, and tomorrow. *Neurosurgery* 1993; **33**: 645–7. [This article provides a historical perspective on neural transplantation from the basic science research to clinical trials.]

Borlongan CV, Stahl CE, Cameron D, Sanberg PR. *Neurological Research* 1996; **18**: 297–304. [This paper reviews alternative non-fetal graft sources, methods in circumventing the host immune response, and genetic engineering for transplantation therapy.]

Date I, Imaoka T, Miyoshi Y, Ono T, Asari S, Ohmoto T. Chromaffin cell survival and host dopaminergic fiber recovery in a patient with Parkinson's disease treated by cografts of adrenal medulla and pretransected peripheral nerve. Case report. *Journal of Neurosurgery* 1996; **84**: 685–9. [This paper presents results of clinical trials on neural transplantation using adrenal medulla: the Japanese experience.]

Deacon T *et al.*. Histological evidence of fetal pig neural cell survival after transplantation into a patient with Parkinson's disease. *Nature Medicine* 1997; **3**: 350–3. [This paper presents the first evidence of survival of xenografted porcine cells in the brain of a patient with Parkinson's disease.]

Dunnett SB, Bjorklund A. Basic neural transplantation techniques. I. Dissociated cell suspension grafts of embryonic ventral mesencephalon in the adult rat brain. *Brain Research Protocols* 1997; **1**: 91–9. [This paper provides the basic dissection and surgical procedure for intracerebral transplantation of dopaminergic neurons.]

Freed CR *et al.* Survival of implanted fetal dopamine cells and neurologic improvement 12 to 46 months after transplantation for Parkinson's disease. *New England Journal of Medicine* 1992; **327**: 1549–55. [This paper presents results of clinical trials on fetal neural transplantation: the American experience I.]

Freeman TB *et al.*. Human fetal tissue transplantation. In: Germano IM, ed., *Neurosurgical treatment of movement disorders*, pp. 177–92. American Association of Neurological Surgeons, Park Ridge, 1998. [This book chapter discusses different variables that may influence the outcomes of a clinical transplant program.]

Freeman TB *et al.* Bilateral fetal nigral transplantation into the postcommissural putamen in Parkinson's disease. *Annals of Neurology* 1995;**38**: 379–88. [This paper presents results of clinical trials on fetal neural transplantation: the American experience III.]

Hoffer BJ, Olson L. Treatment strategies for neurodegenerative diseases based on trophic factors and cell transplantation techniques. *Journal of Neural Transmission* 1997; **49**: 1–10. [This paper examines the efficacy of tropic factors in enhancing neural graft survival and using such combination therapy for treatment of Parkinson's disease and other neurodegenerative disorders.]

Kopyov OV, Jacques DS, Lieberman A, Duma CM, Rogers RL. Outcome following intrastriatal fetal mesencephalic grafts for Parkinson's patients is directly related to the volume of grafted tissue. *Experimental Neurology* 1997; **146**: 536–45. [This paper presents results of the most recent clinical trials on fetal neural transplantation: the American experience IV.]

Kordower JH *et al.* Neuropathological evidence of graft survival and striatal reinnervation after the transplantation of fetal mesencephalic tissue in a patient with Parkinson's disease. *New England Journal of Medicine* 1995; **332**: 1118–24. [This paper presents results of clinical trials on fetal neural transplantation: the American experience II.]

Lindvall O *et al.* Human fetal dopamine neurons grafted into the striatum in two patients with severe Parkinson's disease. A detailed account of methodology and a 6-month follow-up. *Archives of Neurology* 1989; **46**: 615–31. [This paper presents pioneering results of clinical trials on fetal neural transplantation: the Swedish experience.]

Madrazo I, Drucker-Colin R, Diaz V, Martinez-Mata J, Torres C, Becerril JJ. Open microsurgical autograft of adrenal medulla to the right caudate nucleus in two patients with intractable Parkinson's disease. *New England Journal of Medicine* 1987; **316**: 831–4. [This paper presents the first results of clinical trials on neural transplantation using adrenal medulla: the Mexican experience.]

Palfi S *et al.* MRI-stereotactical approach for neural grafting in basal ganglia disorders. *Experimental Neurology* 1998; **150**: 272–81. [This paper presents results of clinical trials on fetal neural transplantation: the French experience.]

17.1 Pathobiology of atherosclerosis

Jawaharlal W. B. Senaratne and Fiona R. Green

Introductory overview
The magnitude of the problem
Risk factors for atherosclerosis
- Age
- Gender
- Race
- Lipid abnormalities
- Smoking
- Hypertension
- Diabetes mellitus
- Homocysteine
- Haemostatic/fibrinolytic abnormalities
- Genetic polymorphisms
- New markers
- Role of infection

Lesions of atherosclerosis
Theories of atherogenesis
- Lipid imbibition theory
- Thrombogenic encrustation theory
- Neoplastic and viral theories
- Lipid theory
- Perinatal theory
- Response to injury

Pathogenesis of atherosclerosis
- Regression of atherosclerosis
- Cellular interactions
- Molecular events
- Pathological phenomena

Animal models of atherosclerosis
Future
- Role of genes in atherogenesis
- Role of gene transfer
- Plaque stabilization

Conclusion
Further reading

At the end of the twentieth century cardiovascular disease remains the main cause of death in Europe, the United States, and much of Asia. Atherosclerosis is the main underlying disease process responsible. The term atherosclerosis derives from the Greek word for porridge (gruel like), *athero*, and *sclerosis* meaning a scarring fibroproliferative response. This aptly describes the condition and, as we understand the process today, it is thought to be an inflammatory fibroproliferative response to an injurious process within the arterial wall leading to occlusion of the lumen. This chapter will describe the pathobiology of atherosclerosis covering the following areas: a brief introductory overview of the process; descriptions of the magnitude of the problems caused by atherosclerosis; risk factors for atherosclerosis; the macroscopic and microscopic lesions of atherosclerosis; the pathogenesis of atherosclerosis including the cellular interactions, molecular events, and pathological phenomena; the uses and limitations of animal models of atherosclerosis; and the future directions and prospects.

Introductory overview

The following is a brief and simplistic overview to a complex disease in order to provide a framework to describe the intricacies. The normal arterial wall consists of three well-defined layers: an outer adventitia consisting of connective tissue and cells providing a plane of separation between the artery wall and the surrounding tissue; a middle media consisting of concentrically arranged layers of smooth muscle; and an inner intima consisting of a single layer of endothelial cells lying on a basement membrane. Atherosclerosis is primarily a disease process affecting the intima, with changes in the media and adventitia, in some instances, late in the disease. The earliest recognizable event in atherosclerosis is adherence of mononuclear cells to the endothelium. The cells migrate into the subendothelial space where they take up lipids, which are thought to play a key role in all stages of atherosclerosis. The cells that accumulate lipids are known as ‘foam cells' and collections of these cells give rise to the first macroscopically recognizable lesion, the fatty streak. As cell death occurs within the lesion, these cells release their lipid leading to the existence of both intra- and extracellular lipid pools. These developing lesions also recruit smooth muscle cells from the media and these lose their contractile properties and assume a new role as synthetic cells producing extracellular matrix. There is continued influx of cells from the lumen as well as the media and, together with lipid accumulation, proliferation *in situ*, and extracellular matrix synthesis, leads to growth of the lesions and formation of distinct plaques—the hallmark of atherosclerostic disease. The growth of the plaques, which is often eccentric, causes narrowing of the lumen and the resulting interference with the blood supply of the tissue concerned is often responsible for the associated symptoms.

A typical plaque is described as consisting of a core of lipid and necrotic material covered by a cap of variable thickness which merges with the uninvolved intima through the shoulder regions. The plaques may be disrupted under the force of blood flow, often through the shoulder regions, breaching the endothelium and exposing the underlying contents of the core, which are highly thrombogenic. Depending on the size of the disruption there is a variable amount of thrombus formed that may lead to occlusion of the artery. This would give rise to symptoms of acute ischaemia in the tissue supplied by the occluded artery. On the other hand, if the breach was minor and the thrombus small, it may only temporarily reduce blood flow and there is some evidence to show that with time such minor thrombi are organized and incorporated into the developing plaque—another mechanism of plaque growth.

The magnitude of the problem

Ischaemic heart disease is still the principal cause of death among adults in the developed world. The principal cause of myocardial infarction is an acute or acute on chronic reduction of blood flow to the myocardium leading to ischaemia and infarction. The main underlying disease process is atherosclerosis, which leads to narrowing of the lumen of coronary arteries. Gradual growth of a plaque leads to progressive reduction in the blood supply, which when it reaches a critical level leads to ischaemia, particularly when increasing demands are made by exercise. This manifests itself as angina—which is the syndrome of pain in the chest with effort—a form of referred pain. The thickness of the plaque cap can vary and a plaque with a thick cap (described as a ‘stable' plaque) can give rise to near occlusion and angina, which could be quite disabling without causing actual infarction. On the other hand, a plaque with a thin cap and relatively large core of lipid is prone to rupture and is described as being ‘unstable'. Rupture of such an unstable plaque would almost certainly give rise to massive thrombosis on top of the plaque, which is already causing arterial narrowing, and would result in catastrophic reduction in blood flow. There is very good evidence that this process underlies myocardial infarction and unstable angina.

Atherosclerosis of the vessels supplying the brain, particularly the extracranial carotid arteries, gives rise to transient ischaemic attacks and cerebrovascular accidents. Critical narrowing due to atherosclerotic plaques, especially at the bifurcation of the common carotid into internal and external, is believed to lead to episodic showers of platelet emboli, which cause temporary occlusion of the vasculature of the brain leading to transient ischaemic attacks. As these emboli are dissolved there is complete restoration of function. If, however, there is more significant occlusion which fails to resolve, some disability would remain and some warning transient ischaemic attacks are followed by progressively increasing disability (stroke in evolution) which eventually stabilizes to leave a degree of permanent loss of function (completed stroke).

In the limb vasculature, particularly in the lower limbs, atherosclerosis leads to claudication (ischaemic pain felt in the muscles, which causes one to stop walking) in the chronic setting—the limb equivalent of angina. If there is occlusion, especially if it is sudden due to an *in situ* thrombosis, then there is limb-threatening ischaemia and gangrene may ensue with eventual limb loss—the equivalent of myocardial infarct. In the visceral vessels, chronic narrowing of the superior mesenteric artery leads to mesenteric angina and acute occlusion causes infarction of the gut with disastrous consequences. Atherosclerosis is one of the causes of renal artery stenosis, which causes hypertension that is difficult to control.

It is easy to appreciate the high level of mortality and morbidity for which atherosclerosis is responsible when one considers the pathology it causes in so many

vascular territories. In terms of actual numbers, atherosclerosis and related disorders account for more than 50 per cent of deaths annually in the developed world. The morbidity and cost of treatment as well as cost in terms of loss of working days would amount to a staggering sum.

Risk factors for atherosclerosis

A risk factor for atherosclerosis has been defined as a habit, trait, or abnormality associated with a sizeable increase in susceptibility to disease associated with severe or extensive atherosclerosis. Risk factors may be associated with disease rather than causal. Identification of causally related risk factors, especially ones that are avoidable or reversible, is important for primary and secondary prevention. A list of hitherto identified risk factors for atherosclerosis is given in Table 1 and a brief description of individual risk factors with special reference to the quantitative importance and possible mechanisms will follow.

Age
Gender
Race
Lipid abnormalities
Smoking
Hypertension
Diabetes mellitus
Obesity
Homocysteine abnormalities
Haemostatic/fibrinolytic abnormalities
Genetic polymorphisms

Table 1 Risk factors for atherosclerosis

Age

Atherosclerosis is a disease of advancing age and of all the factors considered, age has the strongest and most consistent association. Fatty streaks are visible in the aorta by the end of the first decade of life and there is relentless progression of atherosclerotic disease in most arterial territories with advancing age. According to current concepts of the pathogenesis of atherosclerosis, it is an inflammatory fibroproliferative response to injury rather than a mere degenerative process inevitably associated with ageing. It is therefore likely that effects of age represent the cumulative influence of other factors, such as lipid abnormalities, smoking, hypertension, and diabetes.

Gender

Ischaemic heart disease in the middle decades of life and atherosclerotic disease in general are less common among the female sex, but the differences become less prominent in later life, especially after the menopause. This has given rise to much speculation that female hormones are responsible for this apparent protection, but the use of oestrogens in men in secondary prevention trials if anything increased myocardial infarcts as well as non-fatal pulmonary emboli and thrombophlebitis. However, there is strong evidence that postmenopausal hormone replacement therapy in women reduces cardiovascular morbidity. There have been numerous theories put forward to explain these sex differences including differences in lipid metabolism, such as higher high-density lipoprotein (**HDL**) cholesterol and greater lipoprotein lipase activity in females, but to date there is no definite answer.

Race

There are striking geographical differences in the incidence of coronary heart disease, which is a surrogate marker for atherosclerosis. A number of post-mortem studies have provided corroborative evidence showing differences in intimal involvement with atherosclerosis in different populations. The moot point is whether such observed differences are due to racial factors or environmental factors. In one study where three racial groups formed the sample from one location, the black community showed less atherosclerosis than the white or Asian communities. It was still not possible to ascribe these changes purely to racial factors alone as the black community was predominantly of lower social class, implicating socio-economic factors. One interesting scenario that could shed light on the matter is immigrant populations, such as the Japanese resident in the United States. The prevalence of coronary heart disease among this group was much higher than the comparable Japanese living in Japan, which suggests an environmental rather than racial effect. One of the observations from such studies is the fact that within each racial group studied, there is a definite gradient from low to high atherosclerosis. This means that belonging to a particular racial group will not preclude you from having atherosclerotic disease and moreover within each racial group there are genetic as well as environmental factors that could predispose to atherosclerotic disease.

Lipid abnormalities

From the earliest descriptions of atherosclerosis it was apparent that lipid abnormalities would be causally related to the disease. There are several lines of evidence linking atherosclerosis with lipid abnormalities. First, one of the chief constituents of the plaque is of lipid origin, namely cholesterol, in a number of different forms. Second, there is very strong epidemiological evidence to support a link between higher cholesterol levels (especially for communities as a whole) and the higher incidence of atherosclerosis-related disease. Third, in a number of animal models raised levels of cholesterol are associated with the occurrence of atherosclerotic lesions, which resemble the human disease. Fourth, there are a number of genetically determined lipid abnormalities such as familial hypercholesterolaemia (where there is a defect in the low-density lipoprotein (**LDL**) receptor which leads to massive elevation of circulating LDL cholesterol) in which there is premature and extensive atherosclerosis. Finally and crucially, lowering LDL cholesterol leads to a reduction in ischaemic heart disease events. Lipids play a central role in atherogenesis from causing the initial injury to rupture of the plaque and some of the mechanisms of this will be discussed in the pathogenesis section. Atherogenecity of cholesterol, especially the LDL form, is undisputed but the role of other lipid constituents such as triglycerides is still not entirely clear. There are a number of well-documented lipid abnormalities associated with atherosclerosis and they are summarized in Table 2.

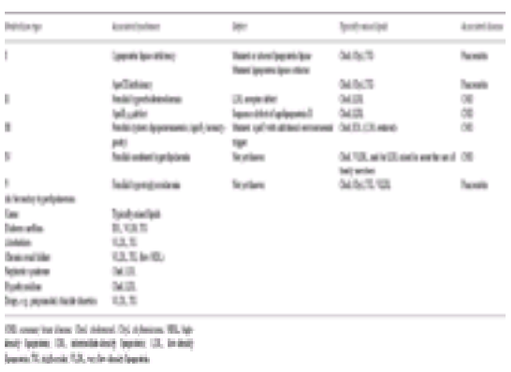


Table 2 Classification of lipid abnormalities

Peroxisomal proliferator-activated receptor-g (**PPAR-**g) is a ligand-activated transcription factor that regulates gene expression in response to various mediators such as fatty acids (e.g. oxidized linoleic acid—9- and 13-HODE), prostaglandins (e.g. 15-deoxy-d12, 14-prostaglandin J2), and a number of pharmacological agents (e.g. the antidiabetic agent troglitazone). Activation of PPAR-g leads to increased expression of plasminogen activator inhibitor type I (**PAI-1**; prothrombotic), increased uptake of oxidized LDL, and a reduction of matrix-metalloproteinase 9 (**MMP 9**) activity which in turn reduces vascular smooth muscle migration—all proatherogenic. Being regulated by fatty acids, PPAR-g is a potential candidate for the link between atherosclerosis and obesity. Central adiposity in particular has been shown to be associated with atherosclerosis.

Smoking

Cigarette smoking has been shown to be associated with increased atherosclerosis in many studies. There have been attempts to quantify the effect of smoking and quoted figures range from a three- to fivefold increased risk of death from ischaemic heart disease for males in their fifth decade who smoked over 15 cigarettes per day. There is some evidence in the literature that the risk is correlated more with the number of cigarettes smoked than the duration of smoking. As there is evidence that the duration of smoking is correlated with the atherosclerosis burden, this indicates that smoking is also related to precipitating events in lesions dependent on the amount of current smoking. As to the mechanisms of this increased atherogenecity, there have been numerous propositions and it is likely that multiple mechanisms operate. The possible mechanisms include: direct injury to endothelium caused by agents in cigarette smoke resulting in blebbing of the luminal surface, microvillus formation, and even detachment of endothelial cells; changes in platelets; increasing fibrinogen and C-reactive protein concentration; and induction of proinflammatory cytokines. In addition to the above, tobacco glycoprotein has been shown to be mitogenic to bovine arterial smooth muscle cells in culture and there are changes in other haemostatic factors such as increased factor VIIIRAG and increased platelet aggregation to adenosine diphosphate.

Hypertension

The increased risk of coronary heart disease among patients with hypertension is quoted to be approximately twofold. Hypertension was shown to be a less important risk factor for ischaemic heart disease in low-risk populations, such as those in the developing countries, where hypertension is associated with haemorrhagic stroke and renal failure (Multiple Risk Factor Intervention Trial). There is also some direct evidence, with more plaques in individuals with hypertension than in those with normal blood pressure, at all ages and in both sexes in the aorta and coronary arteries. Direct endothelial damage under the force of blood pressure is presumed to be the cause, but it is the low-shear areas of the vascular tree with local turbulence of flow and prolonged contact of the blood constituents with endothelium that are involved. Even at branch points the flow divider which experiences the greatest amount of shear is relatively free of atherosclerosis. In a number of primary prevention trials, treatment of hypertension failed to show a significant reduction in ischaemic heart disease.

Diabetes mellitus

As a group, individuals with diabetes are particularly prone to atherosclerosis-related diseases and it is believed that more than three-quarters of deaths in patients with diabetes are due to some form of arterial disease. In one particular study (Helsinki policeman study) for each risk factor and at every level of risk, the mortality from coronary heart disease was threefold higher among those with diabetes than those without. As to the possible mechanisms—diabetes is associated with abnormalities in lipid metabolism which could enhance atherogenesis; in some communities hypertension and smoking, which are independent risk factors for atherogenesis, are more prevalent in those with diabetes; and advanced glycation endproducts (**AGE**), which reflect the abnormal metabolism in diabetes are implicated in injuring the endothelium. These AGEs are known to interact with a cell surface molecule of the immunoglubulin superfamily, named receptor for advanced glycation endproducts (**RAGE**), which leads to enhanced receptor expression and a positive feedback loop whereby prolonged cellular activation of endothelial cells, macro-phages, and smooth muscle cells occurs, setting the stage for chronic cellular activation and tissue damage. The proof that this mechanism operates to a significant degree comes from animal experiments where, in a model of accelerated atherosclerosis associated with diabetes in genetically manipulated mice (streptozocin-induced diabetes in ApoE knockout mice), blockade of cell-surface RAGE by infusion of a soluble, truncated form of the receptor completely suppressed the enhanced formation of vascular lesions. Amelioration of atherosclerosis in these animals by soluble RAGE occurred in the absence of changes in plasma lipids or glycaemia, emphasizing the contribution of a lipid- and glycaemia-independent mechanism to atherogenesis in diabetes. There may also be other, as yet unidentified, mechanisms whereby diabetes leads to increased atherogenesis.

Homocysteine

Patients with elevated plasma homocysteine suffer from accelerated atherosclerosis with myocardial infarctions at a young age. Patients with slightly elevated homocysteine without apparent defects in homocysteine metabolism have an increased risk of atherosclerosis of coronary, cerebral, and peripheral arteries. An elevated homocysteine level is also an independent risk factor for thrombotic disorders. The possible mechanisms include direct toxicity to endothelium, decreasing the availability of nitiric oxide, stimulating collagen production, and being prothrombotic. Dietary supplementation with folic acid can reduce the homocysteine level and trials are underway to establish the role of folic acid in primary and secondary prevention.

Haemostatic/fibrinolytic abnormalities

Elevated plasma levels of several haemostatic factors have been found to be associated with an increased risk of atherosclerotic disease. Elevated plasma fibrinogen is associated prospectively with increased risk of coronary heart disease, carotid arterial disease, and stroke and with peripheral arterial disease. Fibrinogen is an acute-phase reactant as is C-reactive protein, also a risk factor for atherosclerotic disease, and it is likely that elevated levels of these acute-phase reactants reflect pre-existing but asymptomatic, inflammatory atherosclerotic disease. Epidemiological analysis suggests that fibrinogen has an impact on disease over and above the acute-phase effect. This could be explained by the natural functions of fibrinogen—in addition to simply providing the structural framework of the blood clot, it is important in platelet aggregation, is a major determinant of blood viscosity, is involved in monocyte/macrophage binding to cells of the vasculature through the receptor mac-1, and fibrin(ogen) degradation products are chemotactic for various relevant cell types. An elevated plasma level of D-dimer, the plasmin degradation product of cross-linked fibrin, is associated with an increased risk of myocardial infarction and stroke, but the independence of D-dimer as a risk factor is still under debate.

Coagulation factor VII activity has also been implicated in some studies as a risk factor for ischaemic heart disease, but other studies have not validated this. The apparent discrepancy may be accounted for by dramatic differences in the factor VII assays used in these studies, but factor VII level remains a somewhat controversial risk factor. More recently, factor XII levels have also been found to be elevated prospectively, but other studies suggest that this may be because of endothelial damage as a result of pre-existing disease. Factor VIII and von Willebrand factor levels have also been implicated as has soluble thrombomodulin level, but these are not yet widely accepted as risk factors for atherosclerotic disease.

Plasma elevation of plasminogen activator inhibitor type I (PAI-1) is associated with atherosclerotic disease in several cross-sectional studies, and is a predictor of myocardial infarction on a background of pre-existing disease, but its status as an independent primary risk factor still awaits prospective studies for validation. Tissue plasminogen activator (**tPA**) levels have been found to be higher in survivors of myocardial infarction, but it is thought to be a marker of underlying endothelial damage rather than a causal factor. Elevated tPA levels appear to be prospectively associated with thromboembolic stroke. All plasma tPA circulates as a complex with PAI-1 and there is a debate as to whether tPA level is acting as a surrogate marker for PAI-1 measurement.

Other prothrombotic factors, such as factor V and proteins C and S, do not appear to be associated with arterial thrombosis, only with venous thrombosis.

Genetic polymorphisms

Family history is one of the strongest risk factors for atherosclerotic disease and, although twin studies have clearly shown that there is a significant genetic component, this is relatively small compared with a disease like type I diabetes, making it difficult to study. There has been a large amount of research on candidate genes for atherosclerotic disease, particularly myocardial infarction, but so far the findings have been disappointing. A major reason for this is that atherosclerosis is a so-called ‘complex' disease, with both genetic and environmental factors involved in its aetiology. We know that there are likely to be several genes involved; this is not a single-gene disease but is the result of complex interactions among common genetic variants and environmental factors. Each individual genetic variant makes a relatively small contribution to the overall genetic susceptibility to, or severity of, atherosclerotic disease. This makes it difficult to detect the effects of individual candidate genetic polymorphisms without recourse to very large population- or family-based studies.

The number of potential candidate genes currently far exceeds our potential for testing them, and this number is growing with our increasing knowledge of vascular biology. In terms of genetic risk factors actually validated so far, the list is very small. The apolipoprotein E gene, particularly the *epsilon 4* allele, appears to be associated in population-based studies with an increased risk of atherosclerotic disease, probably through its association with elevated LDL cholesterol. The *apoE* gene is further implicated by virtue of the mouse model in which the *apoE* gene has been ablated, where the animals develop atherosclerosis even on chow (i.e. low fat) diet. Much has been made of an association of the angiotensin-converting enzyme insertion/deletion polymorphism with myocardial infarction, but a recent meta-analysis suggests that this may be spurious. The methylene tetrahydrofolate reductase valine-225 allele, which is associated with elevated plasma homocysteine level, has also been implicated, but the data are inconsistent. Another genetic polymorphism that has received considerable attention is the $PI^{A1/A2}$ or Leu/Pro33 polymorphism in the platelet glycoprotein GPIIIa. Although some studies have shown an association of the P33 allele with cardiovascular endpoints, these studies have tended to be rather small, with the larger ones refuting rather than confirming their observations. Although there are data suggesting that the amino acid change may lead to a functional difference in GPIIIa, these data are not entirely consistent and the weight of published evidence is against a simple association of this allele with atherothrombotic disease. Neither the factor V Leiden (Q506) allele nor the prothrombin A20210 allele is associated with atherothrombotic disease, only with venous thrombosis. The literature is awash with small case–control studies on numerous other genetic polymorphisms, but most await confirmation in large, well-controlled studies.

New markers

In an attempt to improve risk stratification, a number of biomarkers are being suggested as new predictors of cardiovascular events. These are linked to pathogenetic mechanisms and include Lp(a) (a link between lipid metabolism and fibrinolysis), markers of inflammation/infection (C-reactive protein), of endothelial dysfunction (the nitric oxide synthesis pathway), and of insulin resistance (plasma insulin levels). There is an active search for markers of other processes important in atherogenesis for which there are currently no markers which can be measured non-invasively, for example apoptosis. Not all these may prove useful, but the hope is that eventually there will be a more accurate risk profile to help in targeting primary and secondary preventative measures.

Role of infection

There has been considerable interest in infection as a potential cause of atherosclerosis and a number of organisms have been implicated including *Chlamydia pneumoniae, Helicobacter pylori*, herpes virus, and cytomegalovirus. The basis of such implication has been either detection of the organism itself in atherosclerotic material (which has not always been consistent) or of antibody responses associated with cardiovascular events. Indirect evidence comes from the use of macrolide antibiotics being associated with improved outcome of ischaemic heart disease, but as these antibiotics also have anti-inflammatory effects, it is not clear what factors are operating. The need for randomized trials has been emphasized in recent publications. One of the objections to infection as a cause of atherosclerosis is the fact that lesions cannot be induced in experimental animals by inoculation of these bacteria or viruses. A recent report of finding *C. pneumoniae* in atherosclerotic lesions of ApoE-deficient mice after intranasal inoculation is suggestive, but still a causal role has not been proved. A suggested mechanism for *C. pneumoniae* causing atherosclerosis is via molecular mimicry involving heat–shock protein 60 (**HSP 60**), where the human protein is mistaken for the bacterial. There is abundant HSP 60 expression in the arterial wall and this causes an immune response, which causes adhesion molecule expression and macrophage activation. This leads to secretion of cytokines, such as tumour necrosis factor-a and matrix-metalloproteinases, which are molecular mediators of atherosclerosis and its complications. However, at present the case for infection as a cause of atherosclerosis remains unproved.

Lesions of atherosclerosis

Macroscopically visible lesions of atherosclerosis are described as fatty streaks, plaques, and complicated lesions, but in reality the process is continuous and it is difficult to say when a fatty streak becomes a plaque. Traditionally, flat lesions are described as fatty streaks and elevated lesions as plaques. Some plaques have a predominance of fibrous tissue while others have large lipid cores. Of the complications, ulceration, rupture and haemorrhage are obvious to the naked eye, but calcification (especially foci of speckled calcification) may not always be recognizable to the naked eye. (Fig. 1 shows some typical examples of atherosclerotic lesions.

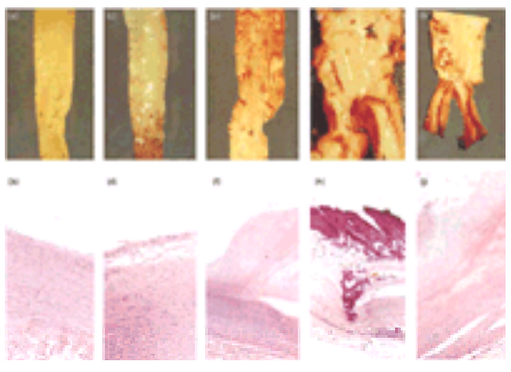


Fig. 1. The range of atherosclerotic disease. (a) and (b), Normal aorta. Note the single layer of endothelial cells in the intima in (b). (c) and (d), Fatty streaks in the aorta. Note the lipids washed away during processing leaving empty spaces in the intima in (d). (e) and (f), Plaques. Note the elevated nature. (g) and (h), Calcified plaques. Calcium deposit burrowing into the medial aspect in (h). (i) and (j), Ulcerated plaque.

Microscopic evaluation of atherosclerotic lesions has been greatly aided by the classification of lesions by the American Heart Association committee on vascular lesions (see further reading). A brief description is given in Table 3.

Type	
I	Initial lesion (adherence of mononuclear cells to intima)
II	Subintimal foam cells
III	Preatheroma (small extracellular pools of lipid)
IV	Atheroma (typical lesion with core and cap)
Va	Fibroatheroma (thick fibrous cap—narrowing lesions)
Vb	Calcified atheroma
Vc	Fibrotic lesion
VI	Complicated lesions with ulceration, rupture, haemorrhage, or thrombosis

Table 3 Histological classification of atherosclerotic lesions

Theories of atherogenesis

Although there has been a large amount of research carried out to unravel the pathogenesis of atherosclerosis, still there is no certainty regarding exact mechanisms. Recapitulating the theories put forward to explain atherogenesis illustrates the progression of the field.

Lipid imbibition theory

Rudolf Virchow (1835) suggested that there is assimilation of circulating lipids into the arterial wall, setting up a chronic inflammatory process involving the intima. It is noteworthy that, even at this early stage, the concept of chronic inflammation was operative.

Thrombogenic encrustation theory

Carl von Rokitanski (1844) suggested that atherosclerosis was due to repeated waves of thrombosis, which occurred from the luminal aspect.

Neoplastic and viral theories

These were put forward by Benditt (1976) and suggested that growth of plaques was due to a monoclonal smooth muscle proliferation and that viral infection was responsible for inducing this cell proliferation. Subsequent experimental data have shown that there are multiple forms of smooth muscle cells and although viruses are still thought to be capable of causing atherogenesis, it is not by inducing uncontrolled cell proliferation but by causing endothelial injury.

Lipid theory

This was put forward with Brown and Goldstein describing the LDL receptor (1973) for which they were awarded the Nobel prize. The transmembrane receptor, encoded by a gene on chromosome 19, is responsible for endocytosis of LDL and when defective leads to accumulation of LDL and cholesterol in the circulation. In the

heterozygous form (approximately 1 in 500 in the developed world) there is moderate to severe hyperlipidaemia (usually about twice the normal) and there are early atherosclerotic complications (usually myocardial infarcts in the fourth and fifth decades). In the homozygous form (familial hypercholesterolaemia, 1 in 1 000 000) there is gross elevation of lipids and extensive and premature atherosclerosis, often causing death in the second or third decade. Heterozygous and homozygous familial hypercholesterolaemia account for only a small proportion of atherosclerotic disease when absolute numbers affected by atherosclerosis are considered, but their discovery amounted to a ‘proof of principle' for the role of lipid metabolism in atherosclerosis.

Perinatal theory

The ‘Barker hypothesis' (1994, fetal origins of adult disease) suggested that individuals who were small at birth due to intrauterine growth retardation are at increased risk of developing cardiovascular disease later in life. This theory is difficult to prove satisfactorily and remains somewhat controversial.

Response to injury

This was proposed originally by Ross and Glomset (1973) and subsequently modified in the light of new information. It states that atherosclerosis is a response to injury, particularly to the endothelium, and is now accepted at least as a working hypothesis by most in the field. Based on this theory, the pathogenesis of atherosclerosis is described in the section that follows.

Pathogenesis of atherosclerosis

Although lesions of atherosclerosis develop as a continuous process, it is helpful to describe the disease in phases of initiation, progression, and complication.

It is not known exactly what initiates atherosclerosis. One of the leading candidates at present is oxidized LDL. LDL is one of the lipo-protein particles and is the major transporter of cholesterol in the bloodstream. There is a large body of evidence that has accumulated over the last decade or so which shows that oxidized LDL is capable of initiating the pathological processes involved in early atherosclerotic lesions. Oxidized LDL is injurious to endothelium and causes expression of adhesion molecules. It is chemotactic to monocyte–macrophages, which also oxidize LDL leading to a self-perpetuating mechanism which is capable of prolonged low-grade injury to the intima. In addition, endothelial cells are capable of oxidizing LDL and, of the number of ways in which this can happen, two mechanisms thought to be important are via superoxide radicals and by the lipoxygenase pathway. Other candidates for initiating lesions are infective agents (*C. pneumoniae* in particular), haemodynamic forces, homocysteine, toxins such as those found in cigarette smoke, and advanced glycation endproducts of diabetes. Endothelial dysfunction leads to increased permeability, permitting entry of lipoprotein particles, and endothelial activation leads to expression of adhesion molecules and recruitment of cells such as monocytes and lymphocytes. Once they are in the subendothelial space, endocytosis of lipids by the cells leads to accumulation of lipids intracellularly and formation of foam cells. Collections of these lipid-loaded cells are visible to the naked eye as fatty streaks, the first macroscopic lesions of atherosclerosis, which are found very early in life in vessels such as the aorta. Death of some of the foam cells releases their lipid contents and leads to the formation of extracellular lipid pools that coalesce to give rise eventually to the necrotic core of the developing plaque.

Smooth muscle cells are recruited from the underlying media and oxidized LDL is chemotactic and mitogenic to them. There are number of growth factors, such as platelet-derived growth factor and fibroblast growth factor, and cytokines such as interleukins and tumour necrosis factor, which are produced by endothelial cells, macrophages, platelets, and smooth muscle cells and regulate smooth muscle proliferation, operating in a complex network fashion. The smooth muscle cells that originally bear a contractile phenotype lose most of their contractile properties and assume a synthetic function. They are responsible for the production of extracellular matrix, an important determinant of the stability of the plaque. Cytokines have also been shown to regulate this synthetic function of smooth muscle cells, for example interferon-g inhibits collagen synthesis by smooth muscle cells *in vitro.* Smooth muscle cells may take up lipid and become foam cells, which look identical to those derived from monocyte–macrophages. By continued lipid and cellular infiltration as well as recruitment and proliferation of smooth muscle cells and matrix synthesis, the plaque continues to grow and can lead to critical narrowing of the artery. By this stage there is a core of lipid which is covered on the luminal aspect by a cap of varying thickness consisting of smooth muscle cells and fibrous tissue. The cap merges into the normal intima through the shoulder regions and these areas contain accumulations of macrophages, with associated matrix-metalloproteinase expression that could lead to plaque instability (Fig. 2).

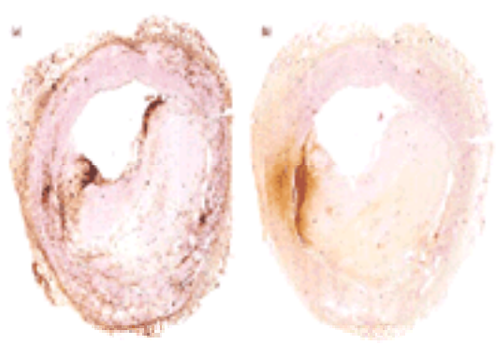


Fig. 2. Macrophages and matrix-metalloproteinases (MMPs) in atherosclerotic plaque rupture. (a), An atherosclerotic artery with an eccentric plaque narrowing the lumen, stained for CD68 (macrophage marker) showing concentrations of macrophages in the plaque shoulders. (b), A serial section of the same artery shown in A stained for MMP-12, showing expression in one of the shoulders, which is a typical finding associated with plaque disruption.

With continuing injury there is endothelial disruption leading to exposure of thrombogenic surfaces to which platelets adhere. These platelets and other constituents of small thrombi forming on the surface of the plaque become organized and incorporated into the growing plaque. The plaques that progress in this fashion sometimes undergo heterotopic calcification. Neovascularization of the plaque occurs with capillaries (often imperfect) forming to supply the deep regions of the growing plaque. This progression phase is followed by the complication phase, which is when most plaques come to light by virtue of the symptoms they cause, which can be quite dramatic and in some cases lead to myocardial infarction or stroke.

In the complication phase there is ulceration (loss of surface endothelium), haemorrhage into the plaque, disruption (rupture of the plaque often through the shoulder region and sometimes through the cap), or any combination of the above. The contributory factors are thought to be haemodynamic stress and weakening of the plaque cap and shoulder regions by the action of matrix-metalloproteinases causing increased degradation of matrix, coupled with reduced production of matrix under the influence of cytokines. Plaques with large necrotic lipid cores and thin caps are particularly prone to rupture and are described as unstable plaques, as opposed to stable plaques with smaller lipid cores and thick caps. With disruption of the plaque, the contents of the core, which are highly thrombogenic, are exposed to the blood. A thrombus forms on top of the disrupted plaque and this may lead to occlusion of the already narrowed lumen leading to acute ischaemia. There is some evidence to suggest that the state of the plaque (whether stable or not) is at least as important, if not more so, than the degree of stenosis in causing significant symptoms, particularly in the coronary circulation. This makes sense as chronic stenosis leads to the development of a collateral circulation and eventual obstruction may not cause a significant change in perfusion, whereas in a lesion that does not cause significant narrowing, with little or no collateral circulation, sudden occlusion will lead to catastrophic ischaemia.

Regression of atherosclerosis

Atherosclerotic lesions are not irreversibly progressive or static; on the contrary they are very dynamic and are reversible to an extent. There is ample evidence from angiographic studies that even quite advanced lesions causing marked stenosis regress substantially under the right conditions. The mechanisms involved include reverse cholesterol transport and remodelling of plaques. Lipid-lowering therapy has been shown to achieve regression of lesions, but this occurs on a longer time scale of months to years, whereas the beneficial effects on clinical event rates are seen much earlier and are thought to be the result of stabilizing lesions.

Cellular interactions

There are a number of cell types that participate in atherogenesis and there are various interactions between them during the process. The roles performed by each of the cell types are described below.

Endothelial cells

Endothelial cells are far more than just a cellular barrier and one of their main functions is to provide a non-thrombogenic, non-adherent surface for the circulating blood cells under normal circumstances. They are also metabolically very active producing many vasoactive substances such as endothelium- derived relaxing factor (now known to be nitric oxide), angiotensin-converting enzyme, endothelin, growth factors such as platelet-derived growth factor, cytokines such as interleukin 1 and 6 (**IL-1** and **-6**), and substances involved in clotting and thrombolysis such as von Willebrand factor, heparan sulphate, prostacyclin, and cyclo-oxygenase. They are capable of oxidizing LDL and, by producing adhesion molecules and cytokines in response to injury, they play an active role in recruiting inflammatory cells into the developing lesion. Endothelial cells express class II major histocompatibility antigens and are involved in antigen presentation. This may be of importance in immunologically mediated injury, especially in view of finding T cells and the more recent observation that some of these cells respond to oxidized LDL. Thus, the endothelium is able to modify and transport lipoproteins, to participate in leucocyte adherence, to form vasoactive substances, to produce growth factors and cytokines, to participate in immune interactions, and to participate in pro- and anticoagulant activity—all of which are part of normal function but when dysregulated become important in atherogenesis.

Macrophages

Macrophages are the principal inflammatory cells of atherosclerotic lesions and it is the adherence of monocytes to the endothelium (once they penetrate and become resident in the arterial wall they are called macrophages) that makes the first detectable change marking the onset of atherogenesis. They are probably the main cell type that is responsible for oxidation of LDL. They act as scavengers and endocytose the lipids in the developing lesion to form foam cells. It would be fair to describe them as the main growth factor and cytokine factories within atherosclerotic lesions, producing an impressive array including platelet-derived growth factor, fibroblast growth factor, transforming growth factors-a and -b, tumour necrosis factor-a (**TNFa**), and IL-1 and -6. Through these mediators the macrophages modulate many smooth muscle cell functions. Macrophages are an important source of extracellular matrix degrading/remodelling enzymes, the family of matrix-metalloproteinases. They are also antigen-presenting cells and, as such, are important in immune-mediated vascular damage.

Smooth muscle cells

Smooth muscle cells form the media of normal arteries and migrate to the intima under numerous chemotactic signals during atherogenesis. They are a major cell type in the enlarging plaque and are responsible for producing extracellular matrix. They are also a source of cytokines and growth factors, which regulate the smooth muscle cells themselves as well as other cells in the plaque environment. Smooth muscle cells are described as being in one of two phenotypic states: cells in the ‘contractile' form are rich in myofilaments and reside in the media in a quiescent state, whilst cells in the ‘synthetic' or ‘proliferative' form, in which the myofilaments are largely lost and endoplasmic reticulum and the Golgi apparatus are well developed, are found in plaques and synthesize extracellular matrix including collagen. This is probably an oversimplification but is helpful in trying to understand the pathogenesis. As one of the major cell types that proliferate during the growth phase of the plaque, as well as being the major producer of extracellular matrix which is important in maintaining stability and forming the plaque cap covering the core, smooth muscle cells play a very important role in atherogenesis.

T cells

There are varying numbers of T cells in atherosclerotic lesions and this has given rise to speculation that atherosclerosis has an immunological basis. There is a large number of the CD4+ helper–inducer subclass of T cells and a lesser number of CD8+ cytolytic T cells. Class II histocompatibility antigen (e.g. HLA DR) expression has been demonstrated in T cells, macrophages, and smooth muscle cells in the plaque, indicating active immunological mechanisms in operation. Although the T cells from plaques have been shown to be polyclonal, it is noteworthy that at least some were reactive against oxidized LDL and HSP 60. These are regarded as possible candidates for initiating immune mechanisms in atherosclerosis.

Platelets

Platelets, which are non-nucleated cell fragments, perform an important function in maintaining vascular integrity, bridging small defects, and preventing spontaneous bleeding under normal circumstances. Platelet adhesion and mural thrombosis have been identified in atherosclerotic lesions where platelets adhere to dysfunctional endothelium, exposed collagen, or activated macrophages. Activated platelets release their granules containing growth factors and cytokines, which modulate a number of functions in the plaque. Platelet activation also leads to prostaglandin synthesis and, in particular, thromboxane A2 causes vasoconstriction and further platelet aggregation, while leukotrienes mediate inflammation. After plaque rupture, it is the platelets that form the initial thrombus and the size of the thrombus will have an effect on outcome. There have been attempts to modulate platelet function, with some success, to prevent thromboembolic complications of atherosclerosis by use of aspirin, dipyridamole, or one of the newer agents acting on platelets (e.g. clopidogrel), after transient ischaemic attacks or arterial bypass grafting. More recently, a better understanding of some of the mechanisms has led to the use of specific antagonists to platelet receptors involved in haemostasis. For example, an antagonist to the platelet glycoprotein IIb/IIIa receptor has been shown to inhibit thrombosis in myocardial infarctions. Platelets are implicated in atherosclerosis throughout the process, playing an important role during the initiation, progression, and complication phases of the disease.

Molecular events

The major advances that have been made over the last two decades in atherosclerosis research have been due to research into the molecular mechanisms involved in atherogenesis. This has led to a better understanding of the mechanisms that operate throughout atherogenesis and in some instances has been translated to more effective therapy. A review of some of the mechanisms thought to be important follows.

Role of adhesion molecules

These molecules are involved in recruiting cells to atherosclerotic lesions. One of the earliest events observed in atherogenesis is mononuclear cells adhering to the intima and this is brought about by expression of adhesion molecules on the endothelial cells which bind ligands expressed on leucocytes (e.g. integrins, sialyl-Lewis x). The molecules involved include selectins (E-, L-, and P-selectins), the immunoglobulin superfamily of adhesion molecules (intracellular adhesion molecules; **ICAM-1** and **-2**), vascular cell adhesion molecule (**VCAM**), platelet–endothelial cell adhesion molecule, and others which are still only partially characterized (vascular cell adhesion protein 1, CD14 on monocytes). Adhesion molecule expression is brought about by activation of the cells involved and this may be by a number of mechanisms including cytokines (TNF-a is known to upregulate ICAM-1 and VCAM-1 on endothelial cells), mechanical factors (low shear stress and turbulent flow are known to induce ICAM-1), and modified LDL (oxidized LDL is known to induce ICAM-1 and VCAM-1). Adhesion molecule expression leads to margination, rolling, attachment, and spreading of leucocytes on the endothelial surface, followed by transmigration through the intima following chemotactic stimuli into the arterial wall.

Role of chemokines and cytokines

Chemokines and cytokines are peptide molecules produced by cells to function as molecular messengers. They elicit wide ranging effects locally as well as systemically and key features include the extremely low concentrations (nano- or picomolar) at which they are effective, their rapid metabolism, and their complex interactions which enable a degree of versatility in the communications they mediate. They are involved in practically all stages of atherogenesis and regulate many functions within the microenvironment of the plaque, with complex interactions. Only some key examples are described here and the reader is referred to further reading for a more comprehensive description.

Monocyte chemoattractant protein 1 (**MCP-1**) is one of the chemokines that has been shown to play a role in recruiting mononuclear cells, where the cells migrate down a concentration gradient of MCP-1. Macrophages themselves produce MCP-1 leading to further recruitment and setting up a potentially self-perpetuating mechanism. Animal experiments where the MCP-1 gene has been knocked out in animals who are prone to atherosclerosis due to induced LDL receptor deficiency (i.e. double knockouts) have shown reduced lesion formation, underlining the importance of this molecule in atherogenesis. Another such experiment in osteopetrotic mice showed the importance of macrophage colony-stimulating factor—a factor important in survival of macrophages.

Cytokines are involved in regulating adhesion molecule expression and also smooth muscle cell functions. For example, IL-1 and IL-6 are known to regulate smooth muscle cell proliferation through a mechanism dependent on platelet-derived growth factor. Interferon-g is known to inhibit collagen synthesis in smooth muscle cells and TNF-a is known to activate matrix-metalloproteinases (**MMPs**), both of which would lead to weakening of the matrix and development of plaque instability. On the other hand, IL-10 is an anti-inflammatory cytokine, deactivating macrophages and stabilizing plaques. Cytokines are inducible molecules and smoking is one of the known mechanisms for induction of cytokine production. Moreover, cytokine genes and their regulatory elements have been shown to have a number of polymorphisms, some of which have been shown to have functional effects, e.g. TNF-a.

Recently, there has been considerable interest in CD40 and CD40 ligand (**CD40L**) interactions. CD40 is an immune mediator of the TNF receptor family which is expressed by macrophages, T cells, endothelial cells, and smooth muscle cells—all the important cell types in atherosclerosis; CD40L, also known as gp39 and CD154, is a TNF-like molecule. There is a body of experimental evidence supporting multiple roles for CD40–CD40L interactions in atherogenesis including increased

expression of adhesion molecules, induction of cytokine release (e.g. IL-1b), and MMP expression (where CD40–CD40L interactions were more effective than TNF-a or IL-1, well known inducers of MMPs). Inhibition of CD40 signalling by means of blocking antibodies to CD40L has been shown to reduce lesion formation in a mouse model (LDL receptor-deficient mice).

Role of matrix metalloproteinases and their inhibitors

Matrix-metalloproteinases are a family of enzymes involved in degradation of matrix proteins important for remodelling of the evolving plaque. These are generally secreted in inactive zymogen form and require activation. There are naturally occurring inhibitors to the MMPs known as tissue inhibitors of MMPs and the balance between activation and inhibition will determine whether there is net degradation or accumulation of matrix. The main producer cells are macrophages and smooth muscle cells and there is tight regulation of expression by cytokines within the plaque. The key MMPs and their main substrates are shown in Table 4.

Class	Substrate	MMP type	Main substrates
Collagenases	Interstitial collagenase	1	Collagen, especially III
	Neutrophil collagenase	8	Collagen, especially III
	Collagenase 3	13	Collagen, especially III
Gelatinases	Gelatinase A	2	Gelatins, collagen, elastin, fibronectin
	Gelatinase B	9	Gelatins, collagen, elastin, fibronectin
Stromelysins	Stromelysin 1	3	Fibronectin, laminin, non-fibrillar collagen, proteoglycans
	Stromelysin 2	10	Fibronectin, laminin, non-fibrillar collagen, proteoglycans
	Stromelysin 3	11	Serpin, fibronectin, proteoglycans
	Matrilysin	7	Laminin, fibronectin, collagen IV
	Metalloelastase	12	Elastin
Membrane-type MMPs	MT1-MMP 1	14	Collagen type IV, gelatin, pro-MMP2
	MT2-MMP 2	15	Collagen type IV, gelatin, pro-MMP2
	MT3-MMP 3	16	Collagen type IV, gelatin, pro-MMP2
	MT4-MMP 4	17	Collagen type IV, gelatin, pro-MMP2
Unclassified		18	

Table 4 Matrix-metalloproteinases (MMP) and their substrates

There is direct and indirect evidence showing the importance of MMPs in atherosclerosis. MMP 1 has been shown to be highly expressed in high-stress regions of the plaque that are prone to rupture. MMPs 1 and 3 have been shown to be expressed more in unstable plaques and, more recently, MMPs 2, 9, and 12 (all capable of lysing elastin) have been shown to be expressed more by symptomatic unstable carotid plaques than by stable ones. Lipid lowering leads to reduction of cardiovascular event rates much more quickly than it leads to regression of angiographically demonstrable lesions; one of the potential mechanisms may be by reducing modified LDL in the plaques, thereby reducing macrophage numbers, which in turn leads to a reduction of MMPs and stabilization of unstable plaques. In theory, inhibition of MMPs should be helpful in stabilizing plaques, but in practice, interference with collagen metabolism in critically narrowed lesions may actually tilt the balance in favour of matrix accumulation and hasten occlusion.

Pathological phenomena

The molecular mechanisms and the cellular interactions described above lead to a number of pathological phenomena observed in atherosclerosis that are helpful in understanding the process of atherogenesis.

Endothelial dysfunction

Endothelial cells are very active cells, not only forming a dynamic barrier but also regulating many functions of the arterial wall. At or near atherosclerotic lesions there is altered function of the endothelium (described as endothelial dysfunction) which leads to deficient vasodilatation and/or an exaggerated vasoconstriction. Mechanisms thought to be involved are an imbalance of mediators such that there is an excess of those causing vasoconstriction and a deficit of those causing vasodilatation. These mediators are known to be produced by endothelial cells themselves. Prostacyclin was one of the earliest identified (by Moncada in 1976) and leads to vasodilatation. In 1980 Furchgott and colleagues described the endothelial-derived relaxing factor (now known to be nitric oxide) in a series of elegant experiments where vasodilatation in response to acetylcholine was shown to be dependent on the presence of endothelium. In 1988 Yanagisawa and colleagues described the potent vasoconstrictor peptide endothelin. Under normal circumstances endothelial-derived relaxing factor predominates, but in diseased arteries there is elevated tone. There have been attempts at quantifying endothelial dysfunction in the research setting by, for example, measuring brachial artery diameter by ultrasound in response to reactive hyperaemia (endothelium dependent) and sublingual nitroglycerin (endothelium independent) with some success. Vasospasm is an important contributory factor in the causation of acute coronary syndromes and endothelial dysfunction is thought to be responsible. Selective endothelial dysfunction has been demonstrated in angiographically normal arteries in patients with hypercholesterolaemia, underlining the importance of this phenomenon as a primary mechanism. Reduction of LDL levels with lipid-lowering therapy results in improved endothelial function with reduction of oxidized LDL, which downregulates endothelial-cell nitric oxide synthase (**ecNOS**), and HMG CoA reductase inhibitors have been shown to stabilize ecNOS mRNA, directly upregulating ecNOS.

Calcification

As plaques advance there are often deposits of calcification and typically they form in the deep intima as plates (this is in contrast to Monkeberg's sclerosis where the calcification is primarily in the media). When present they are visible radiologically, but as calcification by itself has little to do with luminal narrowing, the usefulness of this finding is limited. Previously it was believed to be a passive degenerative process, but there is increasing evidence that it is an active regulated process with many similarities to ossification. As to the mechanisms involved, recent research has indicated roles for molecular mediators of bone calcification such as osteonectin, osteopontin, and osteocalcin. Calcium deposits of atherosclerotic plaques consist of hydroxyapatite and may appear identical to bone. Possible mechanisms are developmental retention of pluripotent cells or osteoblastic immigration. In support of the above hypotheses are findings such as bone morphogentic protein type 2(a potent osteogenic differentiation factor) mRNA by *in situ* hybridization. Recent generation of MGP (matrix g-carboxyglutamic acid (**Gla**) protein) knockout mice which developed extensive and lethal calcification of elastic arteries indicated a role for matrix Gla proteins in vascular calcification. These Gla proteins contain glutamic acid residues, which must be g-carboxylated by a vitamin K-dependent enzyme, and this is a possible mechanism whereby environmental influences such as diet and drugs (e.g. warfarin) may influence vascular calcification. Advanced glycation endproducts found in diabetes accelerate calcification, which would explain at least in part the calcification in association with atherosclerotic disease of the arteries found often among patients with diabetes. Advanced glycation endproducts were shown to have the ability to induce osteoblastic differentiation of pericytes, contributing to the development of vascular calcification in diabetes.

Angiogenesis

Normal intima with a single layer of endothelial cells lying on the vascular basement membrane derives its nutrition from the luminal blood. The media and adventitia of larger vessels have a network of blood vessels (vasa vasorum) that supply these areas. The development of atherosclerotic plaques is associated with neovascularization of the thickened intima and media of vascular walls. The mechanisms responsible for the formation of these intraplaque microvessels are not clearly understood. Postulated mechanisms include the action of smooth muscle-derived angiogenic factors such as vascular endothelial growth factor. There is evidence that hypercholesterolaemia and oxidized LDL impair endothelial cell growth by suppressing basic fibroblast growth factor expression. Furthermore, apolipoprotein A, a factor associated with high atherogenicity, has been shown to suppress angiogenesis. Plaque-associated microvessels are often poorly formed: a recent paper showed that symptomatic carotid disease was associated with increased neovascularization within the atherosclerotic plaque and that these vessels were larger and more irregular and possibly contributed to plaque instability and the onset of thromboembolic sequelae. Neovascularization may thus have a role in the progression of atherosclerotic plaques as well as in the development of intraplaque haemorrhage. Finally, the angiogenesis inhibitors endostatin and TNP-470 have been shown to reduce intimal neovascularization and plaque growth in the apolipoprotein E-deficient mouse model of atherosclerosis, offering promise for the future to be able to control angiogenesis.

Plaque disruption

Coronary thrombosis as a cause of acute myocardial infarction was identified by Herrick as far back as 1912, but only recently was the importance of plaque rupture as a causative factor identified following the work of Davies, Falk, Fuster, and colleagues. Mechanisms involved in plaque disruption have been described in the section on pathogenesis and are summarized in Fig. 3. The potential for achieving plaque stabilization is discussed under future directions.

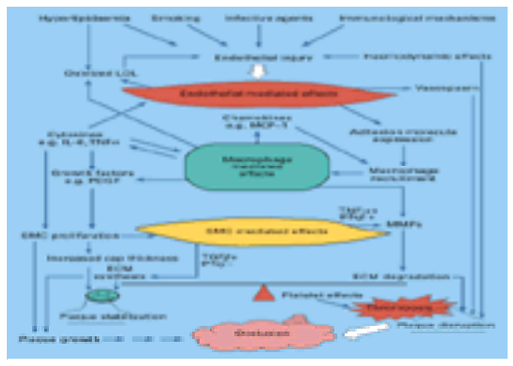


Fig. 3. Schematic diagram of processes in atherogenesis.

Animal models of atherosclerosis

There are a number of animal models of atherosclerosis and these have been used to study the pathogenesis. They also serve to study the effects of new strategies to combat atherosclerosis (e.g. novel pharmacological agents, gene transfer) and their adverse effects. A wide variety of animals has been used including small mammals (mice, rats, rabbits), birds (pigeons), and non-human primates (monkeys). There are advantages and disadvantages in each of these models and the reader is referred to further reading for a detailed discussion on this subject. Briefly, mouse models are more easily manageable and less expensive to work on, but one of the major drawbacks is the ultimate relevance of observations in mice to the human disease. Mice naturally have low levels of LDL and VLDL, carry their cholesterol mainly as HDL, and are relatively resistant to developing atherosclerosis. In order to make them develop lesions of atherosclerosis, gene knockout technology (e.g. ApoE or LDL receptor knockout mice) has been used with or without dietary manipulation (high cholesterol, high fat Western-type diet). ApoE knockout mice develop lesions of atherosclerosis on chow diet after a time and this can be accelerated by fat feeding. These animal models develop lesions that resemble atherosclerotic lesions in humans, but the conditions under which they occur differ, so that direct comparison with human disease should be made with caution. A major problem with the mouse models is the fact that plaque rupture, an event considered the most important in the generation of complications of atherosclerosis in humans, does not occur naturally in any of these animal models.

These animal models have been useful in shedding light on a number of mechanisms operating during atherogenesis. Different animal models are appropriate to study different aspects. For example, cholesterol feeding in rabbits leads to lesions with large cores containing macrophages which resemble human lesions and would be appropriate to study macrophage functions, whereas balloon injury-induced lesions of rat carotid arteries consist largely of vascular smooth muscle cells and very few macrophages and would be suited for studies of intimal hyperplasia and fibrous cap formation. Among the rabbit models, the Watanabe heritable hyperlipidaemic (**WHHL**) rabbit, one of the first animal models of atherosclerosis, models familial hypercholesterolaemia in humans, while the St. Thomas Hospital strain models familial combined hyperlipidaemia. These allow for the assessment of candidate genes for potential use in the treatment of patients with dyslipoproteinaemia. To date, genes for human ApoA-I and lecithin cholesterol acyltransferase have been introduced into the WHHL background and these transgenes have been shown to have significant effects on plasma lipoprotein concentrations. New Zealand White rabbits have low plasma total cholesterol concentrations and lack an analogue of human ApoA-II, providing a system in which to assess the effects of human transgenes on plasma lipoproteins and atherosclerosis susceptibility. These studies provide insights into the mechanisms responsible for the development of atherosclerosis and illustrate the potential uses of animal models in atherosclerosis research.

Future

Role of genes in atherogenesis

The genetic influence on atherosclerotic disease is not as great as for a disease like type I diabetes, so the question arises as to why we should be interested in it. The answer is very clear: this approach will reveal genes and thus highlight biological pathways that are causal in the disease. The actual genetic defects in the pathway may only be important as risk predictors in a relatively small group in any given population, but a novel pathway itself has the potential to provide completely novel points for therapeutic intervention in the population as a whole. Since this is a complex disease, modification of risk at several points may be necessary to reduce overall population risk; lipid-lowering therapy, whilst clearly efficacious, has not yet ‘cured' the disease. There is clearly room for therapeutic improvement, particularly as the incidence of this ‘Western' disease accelerates in the developing world, and genetics may hold the key.

The candidate gene approach, based upon a knowledge of the pathophysiological role of the gene product, has been disappointing. The advent of new mathematical methods for linkage and association analysis and new technologies for genotyping, particularly single nucleotide polymorphisms, means that in the near future, completely novel candidate genes for atherosclerotic disease will be identified on the basis of their genetic relationship to the disease, independently of any knowledge of their biological relevance.

On a more general note, pharmacogenomics and pharmacogenetics, that is, the analysis of polymorphisms determining the efficacy of drugs, whether they influence drug metabolism (pharmacogenomics, e.g. the cytochrome P450s) or whether they influence the response to a drug (pharmacogenetics, e.g. neuronal nicotinic acetylcholine receptor mutations and sensitivity to antiepileptic drugs), will become increasingly important in the prescription of drugs in the not too distant future.

Role of gene transfer

Gene transfer is developing rapidly for therapy of atherosclerotic disease. Potentially therapeutic genes may be delivered to the vasculature or the myocardium though a variety of different techniques. The type of vector used may be naked DNA or oligonucleotides, in the presence or absence of liposomes, or viral vectors, particularly adenovirus for human work and retroviral vectors for animal and *in vitro* work. Naked DNA has the advantage that it does not appear to elicit an immune response, in contrast to adenoviral vectors (although this may depend on the delivery route), but the disadvantage is that gene transfer is less efficient for naked DNA than for adenoviral-mediated transfer. The route of delivery may be important: intramuscular injection, dual and/or coated balloon catheters, coated stents, and adventitial gels have been used, depending on the required site of delivery and ease of access. Adjuvants, such as dimethyl sulphoxide, hyaluronidase, or elastase treatment prior to gene transfer, and physical factors, such as pressure and ultrasound, may also improve transfection/transduction efficiency, but such investigations are still at the research stage. As well as transferring genes that may be beneficial, it is also possible to inhibit genes that may have a detrimental effect. Various options are open to do this; for example, the introduction of antisense oligonucleotides or vectors expressing antisense RNA to inhibit translation of mRNAs encoding the deleterious genes. Another way is to introduce, either directly or through an expression vector, ribozymes that specifically degrade the particular deleterious mRNA; although it acts at the same point—inhibition or prevention of translation—this method is superior to standard antisense inhibition because the stoichiometry is such that one antisense molecule inhibits one mRNA molecule, whereas one ribozyme can degrade multiple mRNA molecules. Antisense constructs or ribozymes can also be delivered in various different ways as for standard gene transfer vectors.

The use of gene transfer techniques to promote angiogenesis and hence restore collateral circulation is advancing rapidly. Phase I clinical trials in humans using direct myocardial injection either of naked DNA or of adenoviral vectors expressing vascular endothelial growth factor have proved successful in reducing symptoms of intractable angina, at least in the short term. Analogous treatment of limb ischaemia is also progressing well. Vein grafts and autologous arterial conduits used for bypass grafting have a tendency to fail and are also candidates for gene transfer technology. The problem with all gene transfer techniques is that the gene expression tends not to persist. This may not matter if there is a specific point at which gene expression is crucial. One example is the disruption of cell division by transfer of antisense inhibitors or ribozymes against cell cycle control genes (e.g. *c-myb*) at the time of balloon angioplasty and/or stent implantation, which has the potential to prevent the intimal hyperplasia that is the cause of the 30 to 50 per cent of restenosis, with the intervention itself providing a convenient and timely means of gene delivery. Other genes that are currently under evaluation for therapeutic transfer into the vasculature include basic fibroblast growth factor, nitric oxide synthase, kallikrein, thrombomodulin, and hirudin. Other systems that may be targeted in the near future include the matrix-metalloproteinases and their inhibitors, the plasminogen activation system, and apoptosis. The safety and efficacy of the phase I trials make gene transfer a promising adjunct, or even alternative in refractory atherosclerotic disease, to surgery, which will be developed rapidly in the next few years.

Plaque stabilization

The importance of plaque disruption in the causation of atherosclerotic disease has been described in previous sections. A recent editorial identified plaque instability as the challenge for atherosclerosis research in the next decade and that plaque instability rather than progression should be the target for new therapies, as this underlies most clinical events. Strategies to achieve plaque stabilization and the mechanisms targeted are summarized in Table 5.

Strategy	Mechanism(s)
Lipid lowering, e.g. HMG-CoA reductase inhibitors—‘statins’	Lowering of LDL and other biologically active lipids. Reduction of lipid content and altering the composition of lipids in the plaque making it firmer. Improvement of endothelial function (upregulating eNOS). Reductions of VCAM expression. Reduction of macrophage numbers in the plaque by reducing monocyte adhesion
Antioxidants, e.g. vitamin E, vitamin C, probucol	Reduction of LDL oxidation
β-Blockers, e.g. atenolol	Controlling blood pressure and lowering heart rate, thereby reducing haemodynamic stress across the plaque
Angiotensin-converting enzyme inhibitors, e.g. lisinopril	Control of blood pressure. Reduction of macrophage accumulation and cholesterol content in the plaques. Increased smooth muscle growth and proliferation. Effects on vessel wall tone. Antithrombotic effects
Matrix metalloproteinase (MMP) inhibitors (experimental)	Inhibition of MMPs, reducing extracellular matrix degradation and increasing cap strength
Fibrinolytic therapy	Dissolution of possible subocclusive coronary thrombi to re-establish full coronary flow. Anti-inflammatory effects
Gene therapy (experimental)	Phase 1 clinical trial strategy involves endothelial growth factors to promote angiogenesis and collateral formation rather than plaque stabilization per se. Future forces endothelial cell repair to improve plaque stability
Others, e.g. interruption of CD40 signaling	Blocking antibodies to CD40 ligand. Reduction of MMP activity

Table 5 Strategies to achieve plaque stabilization

Conclusion

Although atherosclerosis is thought of as a disease of modern living and was described less than 200 years ago, there is evidence that it has existed in humans for much longer. Changes consistent with atherosclerosis have been described in ancient Egyptian mummies from more than 1000 years BC. Although there has been a decline in the mortality caused by atherosclerosis-associated diseases, it remains the number one killer disease of the developed world and its incidence is increasing in the developing world. With a better understanding of its pathogenesis, rational interventions are being implemented. Dietary modifications to reduce cholesterol and to replace saturated fat intake with unsaturated (especially the benefits of oily fish), lipid-lowering drugs, cessation of smoking, better treatment of hypertension and diabetes, and the use of aspirin as an antiplatelet therapy are the cornerstones. The debate on lipid lowering has moved from ‘whether to treat' to ‘who to treat', but lipids are not the whole story. According to a recent review only 50 per cent of the population with cardiovascular disease has abnormal lipids, which may have genetic and/or environmental causes, and so it is not surprising that even if lipids were kept at normal levels in all individuals, atherosclerosis and its complications would still exist. The challenge for the future is to understand all the components in the pathogenesis of atherosclerosis and to devise rational therapies based on this knowledge.

Further reading

Arroyo LH, Lee RT. Mechanisms of plaque rupture: mechanical and biologic interactions. *Cardiovascular Research* 1999; **41**: 369–75. [One of a series of articles in this volume, which is dedicated to the subject (focus on plaque rupture).]

Berliner JA *et al.* Atherosclerosis: basic mechanisms. Oxidation, inflammation, and genetics. *Circulation* 1995; **91**: 2488–96. [A good review article on molecular aspects of atherosclerosis.]

Breslow JL. Mouse models of atherosclerosis. *Science* 1996; **272**: 685–8. [A review article on currently available mouse models of atherosclerosis.]

Brousseau ME, Hoeg JM. Transgenic rabbits as models for atherosclerosis research. *Journal of Lipid Research* 1999; **40**: 365–75. [A review of rabbit models of atherosclerosis.]

Clinton SK, Libby P. Cytokines and growth factors in atherogenesis. *Archives of Pathology and Laboratory Medicine* 1992; **116**: 1292–300. [A major review on the subject outlining the current knowledge and future directions.]

Collins FS. Genetics: an explosion of knowledge is transforming clinical practice. *Geriatrics* 1999; **54**: 41–7. [A commentary on genetics in a clinical setting.]

Davies MJ. Acute coronary thrombosis—the role of plaque disruption and its initiation and prevention. *European Heart Journal* 1995; **16**(Suppl. L): 3–7. [A review of plaque disruption by one of the most respected authorities in the field.]

Falk E, Fernandez Ortiz A. Role of thrombosis in atherosclerosis and its complications. *American Journal of Cardiology* 1995; **75**: 3b–11b. [A good review of the role of thrombosis in atherosclerosis.]

Folkman J. Therapeutic angiogenesis in ischemic limbs. *Circulation* 1998; **97**: 1108–10. [An editorial describing the latest developments in therapeutic gene transfer for atherosclerotic disease.]

Fuster V, Ross R, Topol EJ, eds. *Atherosclerosis and coronary artery disease*, 1st edn. Lippincott-Raven Publishers, Philadelphia, 1996. [A major monograph written by authorities in each of the areas.]

Green FR, Watkins HC. Genetic association studies in coronary disease: the case of GPIIb-IIIa polymorphisms. *European Heart Journal* 1999; **20**: 706–8. [An editorial highlighting the difficulties of population-based genetic studies in atherosclerosis.]

Gupta S. Chronic infection in the aetiology of atherosclerosis—focus on *Chlamydia pneumoniae. Atherosclerosis* 1999; **143**: 1–6. [A review of evidence for infection as a cause of atherosclerosis.]

Libby P, Sukhova G, Lee RT, Galis ZS. Cytokines regulate vascular functions related to stability of the atherosclerotic plaque. *Journal of Cardiovascular Pharmacology* 1995; **25**(Suppl. 2): S9–12. [A short review by outlining molecular events in atherosclerosis.]

Libby P, Geng YJ, Sukhova GK, Simon DI, Lee RT. Molecular determinants of atherosclerotic plaque vulnerability. *Annals of the New York Academy of Sciences* 1997; **811**: 134–42; discussion 142–5. [A major review on the concept of plaque vulnerability.]

Mach F, Schonbeck U, Libby P. CD40 signaling in vascular cells: a key role in atherosclerosis? *Atherosclerosis* 1998; **137** (Suppl.): S89–95. [Review of the role and mechanisms of CD40 interactions in atherosclerosis.]

Ross R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* 1993; **362**: 801–9. [A major review at the time by the proposer of the response to injury hypothesis.]

Ross R. Atherosclerosis—an inflammatory disease. *New England Journal of Medicine* 1999; **340**:115–26. [The latest of the major reviews on atherosclerosis by Russel Ross. (There are previous reviews in the *New England Journal of Medicine* in 1976 and 1986, which together outline the development over the last three decades.)]

Stary HC *et al.* A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation* 1995; **92**: 1355–74. [A consensus document on classification of atherosclerotic lesions.]

Todd JA. From genome to aetiology in a multifactorial disease, type 1 diabetes. *Bioessays* 1999; **21**: 164–74. [A review exemplifying the study of the genetics of a complex disease.]

17.2 Pathobiology of vasculitis

Jawaharlal W. B. Senaratne and Gavin P. Spickett

Introduction
Classification of vasculitis
Pathogenesis of large vessel vasculitis
Pathogenesis of medium-sized vessel vasculitis
Pathogenesis of small vessel vasculitis
Secondary vasculitis
Recent advances in understanding the pathogenesis of vasculitis
Role of adhesion molecules
Role of cytokines
Role of antibodies and immune complexes
Role of T cells
Role of genes
Conclusion
Further reading

Introduction

Vasculitis can be defined as inflammation of the vessel walls accompanied by demonstrable structural change. The inflammation may be acute or chronic and the structural change may be necrosis (often fibrinoid necrosis) or fibrosis. In contrast to atherosclerosis, where the intima of the arteries is the primary area that is affected, in vasculitis the disease starts in the adventitial aspect of the larger vessels (often in the vasa vasorum) or in capillaries, when small vessels are affected. The resulting clinical syndromes, called vasculitides, are a heterogeneous group of disorders with varied aetiopathogenesis, the common feature being inflammation of the vessel wall. The terms vasculitis, angitis, and arteritis are used synonymously.

This chapter will give a brief overview of the scope of vasculitic disorders, provide a system of classification of vasculitides along with brief descriptions of pathologic features, followed by a description of some recent advances in understanding the pathogenesis. The clinical aspects of vasculitis are dealt with in a separate chapter.

Although these syndromes have been known for a considerable period of time, the aetiopathogenesis of many of the different syndromes is not yet fully elucidated. Known aetiological agents include infections, hypersensitivity phenomena, physical and chemical agents. The pathological and clinical features vary widely depending on size and type of vessels involved, their number and distribution, the chronicity of lesions, and the presence of complications. Circulation of the vital organs (brain, heart, kidneys) may be affected and organ failure or death may occur. Acute forms may run a rapid course with all the lesions more or less at the same stage of development. Chronic forms show a more prolonged course with periods of remission and exacerbation and the lesions are often in various stages of development. Complications include thrombosis, vascular occlusion, aneurysm formation, and rupture of vessels with haemorrhage. Vasculitides may present commonly as multisystem disorders and may have symptoms of cutaneous, gastrointestinal, pulmonary, renal, and/or nervous system involvement

Classification of vasculitis

In 1952 Zeek presented the first classification of the vasculitides based on vessel size, which has formed the basis of most subsequent schemes of classification. There were a number of different schemes in use between the 1950s and 1990 and division by Lie into primary (of unknown aetiology) and secondary (due to infection or another disease process, often a connective tissue disease) syndromes formed one of the advances during this time. In 1990 the American College of Rheumatology (**ACR**) proposed a classification with criteria for seven types of systemic vasculitis setting the stage for a new classification system.

In 1993 an international conference was held, known as the Chapel Hill Consensus Conference (**CHCC**), to try and define the vasculitis syndromes. Although a system of classification was put forward, there are limitations to the usefulness of this. One example is that palpable purpura owing to small vessel vasculitis occurs in up to 20 per cent of patients with ‘polyarteritis nodosa' but the above definitions specifically exclude arteriolar involvement in polyarteritis nodosa. Comparison in unselected patients has shown that the ACR criteria and CHCC criteria identify different patients, particularly in polyarteritis nodosa and microscopic polyangiitis. One of the problems associated with the CHCC classification is that it is somewhat restrictive in the clinical setting. This is to be expected, as these classification criteria were not originally developed for diagnosis in individual patients. Another source of problems is that the CHCC classification has not taken into account pathogenic mechanisms, particularly the role of antineutrophil cytoplasmic antibodies (**ANCA**). The definitions are due to be reviewed soon with all these problems in mind.

A recent scheme for classifying vasculitis syndromes based on CHCC criteria is given in Table 1.

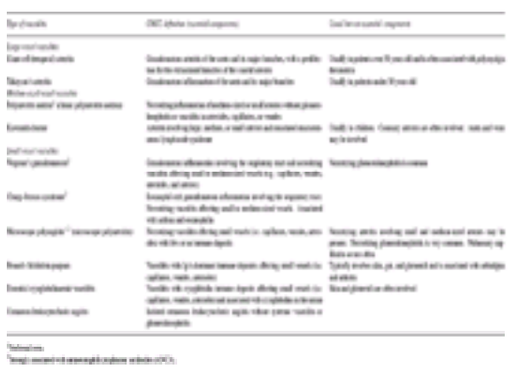


Table 1 Classification of vasculitis

Pathogenesis of large vessel vasculitis

The most important members of this group are giant cell arteritis and Takayasu's arteritis. Both are associated with panarteritis of large and medium vessels, where there is infiltration starting from the adventitia of lymphocytes, macrophages, and later formation of multinucleate giant cells. There is inward progression involving the media, fragmentation of the internal elastic lamina, and intimal proliferation, but fibrinoid necrosis is uncommon (Fig. 1). Healing, with replacement of elastic by fibrotic tissue, may lead to stenosis and occlusion (which may recanalize later) or dilatation and aneurysm formation. Giant cell arteritis typically affects the elderly and there is a female preponderance (3:1). It usually presents with a throbbing temporal headache accompanied by fever and malaise, occasionally leading to blindness and even death. Rarely there is involvement of the thoracic or abdominal aorta and dissecting aneurysms may occur. Takayasu's arteritis typically affects young females (F:M, 10:1), is common in Asia, and usually presents with ischaemia of the upper limbs (called ‘pulseless disease') and non-specific general symptoms. There may be a wide variety of problems due to arterial narrowing (e.g. renovascular hypertension, gut ischaemia, visual disturbance) or dilatation (e.g. aortic regurgitation, aneurysm formation).

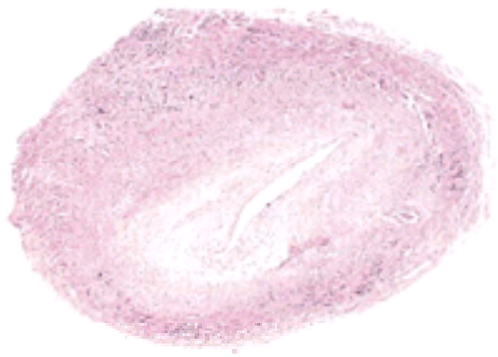


Fig. 1. Giant cell arteritis (large vessel vasculitis). Light micrograph of a transverse section of a temporal artery biopsy showing inflammatory cell infiltration with giant cell granuloma formation and destruction of layers of the wall.

Pathogenesis of medium-sized vessel vasculitis

Polyarteritis nodosa (classic) and Kawasaki disease are examples in this category and polyarteritis nodosa is one of the earliest vasculitic disorders to be described (Kussmaul and Maier 1866). In polyarteritis, small and medium-sized vessels (both arteries and veins) are infiltrated with neutrophils initially, followed by mononuclear cells, and there is fibrinoid necrosis, which heals by fibrosis. The lesions are commonly at branch points and there may be aneurysmal dilatation (leading to nodular enlargement which earned the name polyarteritis nodosa). Immune complex formation is thought to be involved and in some cases hepatitis B virus antigen has been isolated from lesions. Lungs are spared and granulomas do not occur. Multisystem involvement is common and may present with cardiac (angina, ischaemic heart disease), gastrointestinal (haemorrhage, perforation, infarc-tion), renal (renal infarcts and hypertension but glomerulonephritis is characteristic of polyangiitis), or nervous system (neuropathy—mononeuritis multiplex) involvement (Fig. 2).

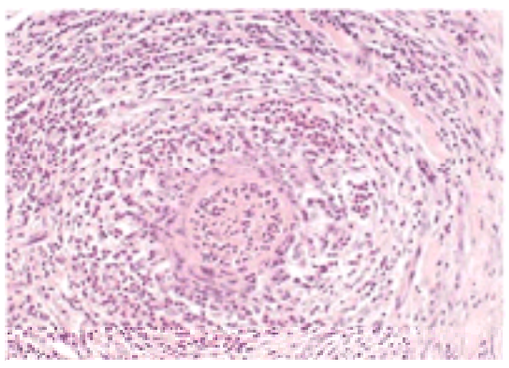


Fig. 2. Polyarteritis nodosa (medium-sized vessel vasculitis). Light micrograph showing the necrotizing vasculitis of polyarteritis nodosa with inflammatory cell infiltration and fibrinoid necrosis destroying the layers of the vessel wall.

Kawasaki disease typically affects children (6 months to 8 years), and is most common in Japan where the disease occurs endemically as well as in epidemics, with seasonal variation, suggesting the involvement of an infectious agent. Recent data suggest involvement of superantigens (e.g. staphylococcal enterotoxins A, B, C, D, and E; streptococcal erythrogeneic toxins A, B, and C) which bind outside the antigen-binding groove on class II major histocompatibility complex molecules and cause T-cell activation and cytokine production. Coronary arteritis is invariable and there is infiltration initially of lymphocytes and neutrophils, through all layers of the arterial wall and even in the periarterial tissue of the larger epicardial arteries in particular, but fibrinoid necrosis is rare. The entire length of the vessels may be involved and damage to the wall may result in diffuse cylindrical or localized saccular aneurysm formation. Thrombosis of these is the main cause of death in acute disease and healing with associated fibrosis and narrowing leads to myocardial ischaemia later. Rarely there are aneurysms of other vessels such as internal mammary, axillary, iliac, and renal arteries.

Pathogenesis of small vessel vasculitis

The syndromes typically representing this group are Wegener's granulomatosis and microscopic polyangiitis. The presence of ANCA with different specificities distinguishes this group, though they are not absolutely specific. ANCA bind to the target antigens, such as protinase-3, on the surface of neutrophils via the F(ab) portion and the unbound Fc portion ligates the Fcg receptors which leads to production of inflammatory mediators such as cytokines, proteolytic enzymes, and reactive oxygen species. It is noteworthy that associations of vasculitides with genetic polymorphisms of FcgRIIa have been described. There is also association with polymorphisms of a_1-antitrypsin, but no HLA associations have been described. It has been suggested that activation of circulating neutrophils by ANCA leads to bystander damage of endothelial cells with development of intravascular thrombosis and inflammation. Indeed, lysed neutrophils and thrombosis in intact capillaries is the earliest identifiable lesion of small vessel vasculitis. Later, mononuclear cell infiltration leads to sustained inflammation, scarring, and at times irreversible organ damage. Vasculitis involves capillaries, arterioles, and venules and at times small and medium-sized arteries. Capillaritis of the lungs leads to pulmonary haemorrhage and glomerular microvascular involvement leads to necrotizing focal segmental glomerulonephritis with crescent formation; both pathologies often coexist. There is some evidence that exposure to silicon may be an environmental trigger in both Wegener's granulomatosis and microscopic polyangiitis. Response of Wegener's granulomatosis to co-trimoxazole suggests a possible infective trigger. In Wegener's there are granulomatous lesions (not seen in microscopic polyangiitis) and necrosis of these may lead to cavitation. Wegener's granulomatosis usually presents with the triad of upper respiratory tract symptoms, such as sinusitis, epistaxis, and otitis media; lung involvement manifesting as cough, haemoptysis, dyspnoea, or pleuritic chest pain; and renal involvement as glomerular nephritis (with no symptoms or with microscopic haematuria) progressing to renal failure (hypertension is uncommon). Microscopic polyangiitis was originally described together with polyarteritis nodosa but there are no significant immune complexes (‘pauci-immune') and the main organs affected are the kidneys and lungs, with skin (purpuric rash), gastrointestinal tract, nervous system, and cardiac involvement being reported (Fig. 3).

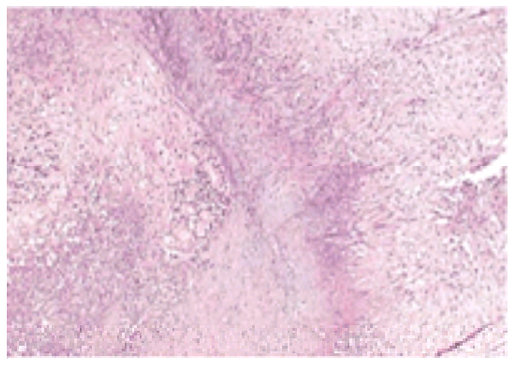


Fig. 3. Wegener's granulomatosis (small vessel vasculitis). Light micrograph showing extensive inflammatory cell infiltration and fibrinoid necrosis in lung tissue typical of Wegener's granulomatosis.

Churg–Strauss syndrome is a multisystem vasculitis with asthma as the most prominent symptom. There is eosinophilia and the lesions resemble polyarteritis nodosa, with which it has been included in some therapeutic trials.

Secondary vasculitis

Some cases of vasculitis are secondary to other diseases such as systemic autoimmune diseases and infections and these are known as secondary vasculitides. In systemic autoimmune diseases, particularly in systemic lupus erythematosus, nuclear antigens and their corresponding antibodies are thought to be pathogenetically involved by forming circulating immune complexes. Such complexes can also form *in situ*, for example by the binding of nuclear histones to the glomerular basement membrane through charge interaction, followed by binding of antibodies. In the case of infections, several mechanisms may operate: immune complex formation may occur in the circulation or *in situ* (e.g. hepatitis C virus causing cryoglo-bulinaemia); microbes may directly invade the vessel wall followed by infectious inflammatory processes (e.g. syphilis); certain viruses such as cytomegalovirus can penetrate endothelial cells and activate them leading to vasculitis; and certain bacterial antigens such as staphylococcal neutral phosphotase can attach to basement membranes and bind IgG non-specifically. Other causes of secondary vasculitis include pharmacological agents, radiation, and malignancy.

Recent advances in understanding the pathogenesis of vasculitis

Role of adhesion molecules

Adhesion molecules are important in recruiting cells to sites of inflammation and there are several classes of adhesion molecules such as integrins (b_1 –VLA-4 and VLA-5, b_2 –LFA-1, Mac-1, p150/95), immunoglobulin superfamily (ICAM-1 and VCAM-1), and selectins (E and L) involved in the pathogenesis of vasculitis. Increasedexpression of Mac-1, LFA-1, and VLA-4 has been reported in vasculitis associated with systemic lupus erythematosus and shown to correlate with disease activity. Increased expression of ICAM-1, VCAM-1, and selectins has been reported in a number of vasculitides but the levels do not always correlate with disease activity. Circulating adhesion molecules, which are often elevated in vasculitides, are not specific as they can also be seen in a number of inflammatory, infectious, and malignant conditions. Thus, adhesion molecules certainly play a role in the pathogenesis of vasculitis but it may not be feasible to use them as markers of specific syndromes, disease activity, or as therapeutic targets.

Role of cytokines

Cytokines are soluble proteins functioning as molecular mediators of inflammation and immunity. Both qualitative and quantitative abnormalities have been described in vasculitis syndromes. Elevated levels of interleukins 1, 2, and 6 (**IL-1**, **-2**, **-6**) and tumour necrosis factor-a in the serum, and these as well as interferon-g and transforming growth factor-b in lesions of vasculitis have been reported. The occurrence of almost identical cytokine mRNA profiles in giant cell arteritis and polymyalgia rheumatica, except that interferon-g was found only in the former, suggests that local production of interferon-g by tissue-infiltrating T cells may be playing a role in granuloma formation. The other possible roles played by cytokines include promotion of leucocyte recruitment via induction of expression of adhesion molecules, activation of endothelial cells and cells mediating immune responses, procoagulant effects such as stimulation of tissue factor expression, and down-regulation of antithrombotic proteins C and S.

Role of antibodies and immune complexes

The discovery of autoantibodies reactive against neutrophil cytoplasmic proteins, first reported by Davis *et al.* in 1982, was a major advance and since then there has been a large number of publications in this area attempting to define the role of these antibodies in the pathogenesis of vasculitis. There are two types of antineutrophil cytoplasmic antibodies: c-ANCA which are directed against a serine proteinase known as proteinase 3 (hence also called PR3-ANCA) that give a characteristic cytoplasmic staining pattern and p-ANCA which are directed against myeloperoxidase (hence also called MPO-ANCA) that give a characteristic perinuclear staining pattern (but this staining pattern can also be seen in the presence of other antibodies against other neutrophil proteins or antinuclear antibodies). These antibodies are found in ‘pauci-immune' (where immune complexes are absent) vasculitis syndromes. There is debate as to whether ANCA are directly involved in the pathogenesis or are merely a disease marker. The case against ANCA being an important mediator includes the finding that in a number of cases of Wegener's granulomatosis there were no detectable ANCA and, when present, the levels did not always correlate with disease activity (although rising titres heralded relapse), in some cases being present for many years after remission, and also the lack of convincing evidence from animal models. There are a number of potential mechanisms for the involvement of ANCA in the pathogenesis of vasculitis. One of them suggests binding of the $F(ab)_2$ region to proteinase 3 or myeloperoxidase and subsequent engagement of the FcgRIIa receptor by the Fc region may lead to activation and neutrophil-mediated vascular injury. There are other autoantibodies such as antiendothelial cell antibodies found in vasculitides but their exact role is not yet clear and some are probably secondary antibodies.

Formation of antigen–antibody complexes and their abnormal deposition leading to inflammation and tissue damage is important in the pathogenesis of several vasculitis syndromes such as polyarteritis nodosa associated with hepatitis B, essential mixed cryoglobulinaemia, Henoch–Schönlein purpura, and isolated cutaneous vasculitis. The typical findings in these include elevated levels of circulating immune complexes, hypocomplementaemia (reflecting the con-sumption), and demonstration of immune complexes in lesions. Complement proteins play an important role in recruiting neutrophils to sites of inflammation as well as causing direct tissue damage from membrane attack complexes.

Role of T cells

Being immunologically mediated diseases, it is not surprising that T cells play very important roles in many vasculitides. In vasculitides where granulomatous lesions are the hallmark (Wegener's granulomatosis, Churg–Strauss syndrome, giant cell arteritis, Takayasu's arteritis) the granulomas consist of T cells (especially CD4+), macrophages, and giant cells. The finding of multiple CD4+ T-cell clones with identical T-cell receptor specificities in anatomically distinct lesions, as well as production of IL-2 and interferon-g by T cells in these lesions, is presented as evidence that giant cell arteritis is mediated by T_{H1} CD4+ T cells responding to an antigen in the vessel wall. There is some evidence that in Kawasaki disease there is superantigen-mediated expansion of polyclonal Vb2+ and Vb8+ T cells leading to immunologic activation and tissue damage secondary to exaggerated cytokine production.

Role of genes

One of the most exciting developments in vasculitis has been the realization that these diseases occur on the background of a genetic predisposition. Familial clustering of at least some types of vasculitides, as well as the association with other types of autoimmune disease, has been known for some time. Most putative HLA associations are not consistent, though this may be due to ethnic variation among the groups studied. More recently, a negative association of ANCA-associated vasculitis with HLA DR13 DR6 was described. In the future, linkage analysis using polymorphic markers in families will hopefully lead to identification of candidate genes.

There are several genetic abnormalities already described which are known to be associated with vasculitis, such as a_1-antitrypsin deficiency. There is evidence of quantitative variation of ANCA antigen (proteinase 3) expression on neutrophils as well as genetic heterogeneity of the ANCA antigens themselves, which may play a role in the formation of ANCA. Genetic variation in the inflammatory mediators (e.g. cytokines, adhesion molecules) may have a role. In this regard, certain polymorphisms of the FcgRIIa receptor for neutrophils have been shown to have high affinity for IgG3 and these were found more frequently in patients with more severe disease. IgG3 ANCA are associated with renal complications and increase prior to a relapse (non-IgG3 ANCA are unrelated to disease activity).

Conclusion

The field is moving towards the idea that many forms of vasculitis are probably genetically based but environmentally triggered. Research is currently directed at trying to unravel the genetic basis and identifying the environmental triggers.

Further reading

Bajema IM, Hagen EC. Evolving concepts about the role of antineutrophil cytoplasm autoantibodies in systemic vasculitides. *Current Opinion in Rheumatology* 1999; **11**: 34–40. [Essential further reading on ANCA.]

Cohen Tervaert JW, Popa ER, Bos NA. The role of superantigens in vasculitis. *Current Opinion in Rheumatology* 1999; **11**: 24–33. [Review of superantigens as an environmental trigger of vasculitis.]

Kallenberg CG, Heeringa P. Pathogenesis of vasculitis. *Lupus* 1998; **7**: 280–4. [Brief review for a quick grasp of the subject.]

Nowack R, Flores Suarez LF, van der Woude FJ. New developments in pathogenesis of systemic vasculitis. *Current Opinion in Rheumatology* 1998; **10**: 3–11. [Most recent paper on the pathogenesis of vasculitis from a series of annual reviews.]

Tervaert JW, Stegeman CA, Kallenberg CG. Silicon exposure and vasculitis. *Current Opinion in Rheumatology* 1998; **10**: 12–17. [Review of relatively recent developments on silicon exposure as an environmental trigger.]

Watts RA, Scott DG. Classification and epidemiology of the vasculitides. *Baillière's Clinical Rheumatology* 1997; **11**: 191–217. [Useful review but due to be updated soon.]

17.3 Vasculitis

Gavin P. Spickett
Professor J. G. G. Ledingham was the coauthor of this chapter in the first edition. Much of the present version is based on that chapter; his contribution to it and his excellent illustrations are gratefully acknowledged.

Introduction
Mechanisms of vascular damage
Investigation of possible vasculitis
Vasculitic syndromes
- **Giant cell arteritis**
- **Takayasu's arteritis**
- **Wegener's granulomatosis**

Polyarteritic syndromes
- **Systemic sclerosis—scleroderma and CREST syndrome**
- **Systemic lupus erythematosus**
- **Behçet's disease**
- **Kawasaki's disease**
- **Buerger's disease (thromboangiitis obliterans)**
- **Lymphomatoid granulomatosis**
- **Henoch–Schönlein purpura**
- **Other causes of vasculitis**

Immunosuppressive treatment in the management of vasculitis
Further reading

Introduction

The term vasculitis embraces a wide variety of conditions in which inflammatory damage to blood vessels is a principal component. The clinical consequences of vasculitis depend on the size and nature of vessels involved, the organs they supply, and the nature of the underlying diagnosis. Most patients with vasculitides will present to physicians in the first instance, but patients with vasculitis affecting major arterial trunks may present to surgeons with, for instance, acute ischaemia of the limbs or abdominal organs. Wegener's granulomatosis is commonly first seen by ear, nose, and throat surgeons since it frequently affects the upper airways and eustachian tubes. Early recognition of a systemic vasculitis is important: most patients respond relatively favourably to medical treatment but have a poor prognosis if left untreated. The diagnosis is often difficult to prove, and in such cases delay in treatment can be particularly hazardous.

The vasculitides can be classified in a number of ways depending on clinical syndromes or histopathological features and size of vessels involved (see Section 7.3). None of these is yet very satisfactory and a retreat to the use of the simple term ‘vasculitis' is increasingly favoured. It is important to recognize the considerable overlap which exists between the primary vasculitides, particularly with regard to the size of vessel affected. Table 1 lists the primary and secondary vasculitides. Primary vasculitides can be roughly subdivided by the size of the predominantly affected vessel and by the presence or absence of granulomas. Such a pathological classification as illustrated in Section 7.3 has little intrinsic merit and does not relate to the underlying triggering event.

Primary	*Secondary*	*Unclassified*
Giant cell arteritis	Systemic lupus erythematosus	Buerger's disease
Takayasu's disease	Systemic sclerosis and variants	Behçet's disease
Wegener's granulomatosis	Rheumatoid arthritis	Kawasaki disease
Polyarteritis	Infection	
(a) Polyarteritis nodosa		
(b) Microscopic polyarteritis		
(c) Churg–Strauss variant		
Relapsing polychondritis	Cryoglobulinaemia	

Table 1 Classification of vasculitis

Mechanisms of vascular damage

The heterogeneous nature of the vasculitides indicates that there is no single explanation for the vascular damage; few primary vasculitic syndromes have a well understood cause. Epidemiological studies and circumstantial evidence suggest that Wegener's granulomatosis, microscopic polyarteritis, and Kawasaki disease may be triggered by infection, but the nature of the agent(s) is unknown. Kawasaki disease strongly resembles toxic shock syndromes associated with streptococcal infection and the effects of a ‘superantigenic' toxin, namely one that promiscuously stimulates all T cells bearing a particular T-cell receptor configuration irrespective of antigenic specificity. Such a superantigen-driven process leads to excess cytokine production which may be damaging to local and distant vascular endothelium either directly or through cellular recruitment. This mechanism may be involved in other types of vasculitis. In the secondary vasculitides the mechanisms of vascular damage are related to the immunological products released as a result of the underlying disorder. The specificity of the disease processes for vessels of particular sizes has not been satisfactorily explained.

The inflammatory response involves non-specific phagocyte cells, macrophages, and neutrophils, which are attracted to areas of vascular damage by specific chemotactic factors (cytokines). These may be released from the vascular endothelium itself, from adherent platelets, or by activation of the complement cascade. Complement is activated locally either by antigen and antibody (classical pathway) or by the damaged endothelium (alternative pathway). The natural amplification of complement activation means that deposition of a small quantity of immune complex rapidly leads to further complement breakdown, release of chemotactic factors, and cellular recruitment. The complement system is linked to the kinin system and the clotting cascade, both of which are activated, leading to increased vascular permeability and thrombus formation. As inflammation progresses, specific T lymphocytes may be recruited: these release lymphokines, further amplifying the cellular response and promoting systemic effects, including the release of acute-phase proteins and fever. Vascular occlusion leads to local tissue infarction, which may be extensive when a major artery is involved. Healing of the vasculitic lesion may lead to local fibrosis, with consequent diminution of blood flow and ischaemic symptoms if the arterial side is involved.

The role of antibody in the vasculitic syndromes is poorly defined. Autoantibodies directed against various components of the cytoplasm of neutrophils (**ANCA**) and against endothelial cells have been described in primary vasculitis: although such antibodies may play a primary pathogenetic role in the disease, rather than being secondary markers of tissue damage, the evidence is far from conclusive Currently, antiendothelial antibodies are viewed as a secondary phenomenon, but some specificities of ANCA have been shown to interact directly with neutrophils to modulate their activation status. Antibody is more obviously involved in the pathogenesis of secondary vasculitis, as in cryoglobulinaemia and the connective tissue disorders. Whereas it has previously been thought that antibody does not enter intact viable cells, it has now been shown that this is untrue for a range of autoantibodies.

Investigation of possible vasculitis

Presentations of vasculitis are protean: features which may lead to a surgical consultation are listed in Table 2. The diagnosis is often difficult to prove, and key investigations required to take a suspected diagnosis further are detailed in Table 3.

Table 2 Presentation of systemic vasculitis to surgical specialties

Table 3 Key investigations in systemic vasculitis

The inflammatory nature of the major vasculitides is manifest by an elevation in plasma levels of the acute-phase response proteins. C-reactive protein, which has a short half-life and rapid response time, is the most useful of these proteins for both diagnosis and monitoring therapy. For reasons that are unclear, certain closely related conditions such as systemic lupus erythematosus (**SLE**) and systemic sclerosis tend to be associated with the production of very little C-reactive protein, while bacterial infection leads to substantial rises in the C-reactive protein levels. Infections masquerading as vasculitis and intercurrent infection in patients with established vasculitic illness may therefore make interpretation of high levels of C-reactive protein difficult. The erythrocyte sedimentation rate, an established marker of inflammation, is useful for diagnosis but of little value in monitoring therapy as it is largely dependent on the serum concentration of fibrinogen, which has a long half-life and a long response time. It is also affected by red cell morphology and the degree of anaemia. The erythrocyte sedimentation rate will nevertheless be raised in most vasculitides.

Serum immunoglobulins tend to be non-specifically elevated; monoclonal immunoglobulins (paraproteins) may be detected in type II cryoglobulinaemia and also in Cogan's syndrome (ocular interstitial keratitis, often associated with various combinations of systemic vasculitis, aortitis, aortic valve disease, and musculoskeletal disease) and, less commonly, in the connective tissue disorders. A paraprotein can be identified rapidly by serum electrophoresis. Some abnormal immunoglobulin molecules (cryoglobulins) precipitate from the serum when it is cooled: if the peripheral skin temperature falls below this critical level *in vivo*, precipitation may then damage cutaneous vessels. To identify a cryoglobulin, a sample must be taken at 37°C and transported at that temperature to the laboratory, where the clot is allowed to retract at the same temperature. The serum is then removed and allowed to cool. Type II cryoglobulins comprise a monoclonal component with rheumatoid factor activity, binding polyclonal immunoglobulins. A high proportion of type II cryoglobulins will have persistent hepatitis C infection: this infection should be sought if a type II cryoglobulin is detected.

Complement levels are valuable indicators which can be used to monitor the activity of systemic lupus erythematosus; in addition, levels are low in some types of cryoglobulinaemia. Both C3 and C4 are acute-phase proteins and their concentrations may be normal even when complement is being consumed; identification of C3 breakdown products may be helpful in demonstrating complement consumption.

Certain syndromes that are accompanied by vasculitis are characterized by the presence of autoantibodies which may be considered to be markers of the disease. This is mainly applicable to the connective tissue disorders (Table 4). In assessing the importance of these autoantibodies when making a diagnosis, it is important to recognize that, as the immune system ages, spontaneous production of raised levels of autoantibodies becomes more common: in middle-aged and elderly patients low levels of these autoantibodies are not necessarily indicative of disease. Autoantibody production may also occur transiently following infection, particularly with Epstein–Barr virus or adenovirus. It is therefore important to check with the reporting laboratory whether the titre of autoantibody is deemed clinically significant. In virtually no cases is the titre of an autoantibody directly related to disease activity, although rising titres of double-stranded DNA antibodies and c-ANCA may herald relapse in systemic lupus erythematosus and Wegener's granulomatosis, respectively.

Table 4 Autoantibodies associated with vasculitis

Rheumatoid factors are autoantibodies directed against the Fc region of human immunoglobulin. High levels are present in seropositive rheumatoid arthritis, and this is associated with the development of nodular and systemic disease. However, they are very poor specific indicators of rheumatoid arthritis since they are produced in response to other inflammatory stimuli and infections, and in the elderly. Their presence does not therefore indicate rheumatoid arthritis unless clinical features are also compatible with this diagnosis.

Antibodies to various neutrophil cytoplasmic antigens have been detected in the plasma of patients with vasculitic syndromes, particularly Wegener's granulomatosis, microscopic polyarteritis, and Churg–Strauss syndrome. In Wegener's granulomatosis and microscopic polyarteritis the antibody which produces a coarse speckled staining of cytoplasm appears to be directed against a lysosomal serum protease (proteinase 3), whereas in necrotizing glomerulonephritis the perinuclear staining antibody recognizes myeloperoxidase, lactoferrin, or elastase in polymorphs. Assays for antibodies to proteinase 3 and myeloperoxidase are now available routinely, to assist in distinguishing between vasculitic syndromes. Atypical staining patterns on human neutrophils may be found in inflammatory bowel disease and sclerosing cholangitis: Crohn's disease certainly has some features of vasculitis histologically. Evidence suggests that titres of cytoplasmic staining antibodies (c-ANCA) may also be used to monitor progress and detect incipient relapse. These IgG antibodies have a half-life of some 3 weeks, so that when following response to treatment it is not useful to repeat assays at intervals of less than 4 to 6 weeks.

The high frequency of renal disease in vasculitic syndromes means that examination of the urine for protein and blood by stick testing, and careful microscopic examination for casts and red cells, is imperative. Plasma creatinine and urea measurements are crude but useful indices of glomerular filtration rate; when progressive

renal disease is suspected, sequential measurements of 24-h urine protein and creatinine clearance should be performed. Isotopic measures of creatinine clearance may be more accurate. Renal biopsy may confirm vasculitis if an involved vessel is sampled, or if the histology is compatible with the renal manifestations of a systemic vasculitic illness. A negative biopsy of renal tissue, however, by no means excludes an arteritic illness.

Arteritis may be confirmed by biopsy of other tissues: those most commonly sampled include skin, muscle, temporal or occipital artery, sural nerve, epididymis, or liver; again it is essential to recognize the patchy distribution of vascular lesions and the need to examine the whole length of any biopsied vessel and to remember that negative findings do not exclude the diagnosis.

Imaging techniques have advanced substantially in the last 10 years. Computed tomography (**CT**) scanning or, better, magnetic resonance imaging (**MRI**) may allow accurate detection of vasculitic lesions in areas such as the lungs and brain, which were previously difficult to investigate. Angiography, including MR angiography, has a major role in delineating the extent of vascular damage, particularly in patients with large vessel diseases uch as Takayasu's disease and in polyarteritis nodosa, where the detection of characteristic aneurysms may be diagnostic. Echocardiography is the diagnostic test of choice in Kawasaki's disease and should be performed sequentially to monitor the size of the coronary artery aneurysms that are a characteristic feature of this illness. Labeled white cell scans may be helpful in localizing sites of inflammation, and often show more extensive disease than is apparent clinically.

Vasculitic syndromes

Giant cell arteritis

This condition, also known as temporal arteritis in view of its predilection for the temporal arteries, was first described in a hospital porter who was unable to tolerate his bowler hat because the pressure it exerted on the inflamed temporal vessels caused severe pain. This disease usually presents with severe headaches, accompanied by localized tenderness of the vessel in its course over the temple and scalp, and sometimes by fever and malaise. It is predominantly a disease of the elderly, particularly women, and is probably the most common form of vasculitis. It is not restricted to the temporal arteries but may involve other major vessels: in the scalp, occipital vessels are occasionally affected. Vasculitis of the facial artery may lead to severe jaw pain which is exacerbated by eating (jaw claudication), while vasculitis of the coronary arteries may cause myocardial ischaemia. Classically, vasculitis may affect the retinal artery, leading to sudden blindness. Vasculitis of arteries of the upper limb may lead to arm claudication, digital ischaemic lesions, and the need for a vascular surgical opinion.

Atypical presentations almost always cause diagnostic delay. There is usually tenderness over the inflamed artery if it can be palpated, C-reactive protein levels and the erythrocyte sedimentation rate are almost always raised, and there is usually a normochromic normocytic anaemia. Biopsy of an affected vessel shows characteristic giant cell granulomas, comprising mainly CD4+ T lymphocytes, but a negative biopsy does not exclude the diagnosis. Angiography may be helpful when upper limb vessels are affected (Fig. 1).

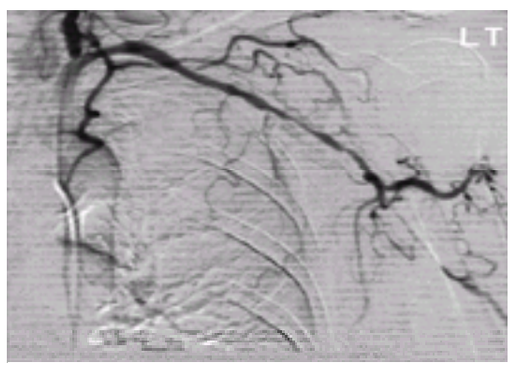


Fig. 1. Angiogram demonstrating the characteristic area of narrowing of the brachial artery in a woman who presented with symptoms of arm claudication and with ischaemic lesions of the fingers. (Kindly provided by Dr E. W. L. Fletcher.)

The response to steroid therapy is rapid, but large doses (60 to 80 mg prednisolone/day) may be required. Immunosuppressive drugs such as cyclophosphamide (2 mg/kg.day) or azathioprine (2 mg/kg.day) have been used on occasions, but there is little evidence to suggest that the combination of prednisolone and these agents conveys substantial advantages over prednisolone alone. The chief value of adding cyclophosphamide or azathioprine is in reducing the total dose of prednisolone needed to control chronic or unresponsive disease. Immediate therapy is mandatory, in view of the risk of visual impairment which occurs in up to 20 per cent of patients with temporal artery disease. The disease tends to regress in activity and may ultimately ‘burn out', and steroids can usually be tailed off after some 1 to 2 years. The disease may relapse, however, and may persist for many years.

The aetiology is obscure: there is an association with polymyalgia rheumatica, and the two conditions may develop sequentially or even coexist in some patients. Polymyalgia rheumatica is a much milder illness, characterized by limb girdle stiffness, and it generally responds to lower doses of corticosteroids.

Takayasu's arteritis

Takayasu first described this disease following observations of retinal changes in a young Japanese woman in 1908. The disease is indeed predominantly one of young women, and although it was first described in Asian women and is still much more common in females (85 per cent of cases), it is by no means racially restricted. It affects particularly the aortic arch and the large vessels of the branches arising from it, but it may extend to or affect only the descending thoracic and abdominal aorta and its primary branches. The pulmonary artery may also be involved.

The illness presents with a generalized inflammatory prodrome in around 70 per cent of patients, with fever, malaise, myalgia, and arthralgia. This is followed after a variable period by symptoms and signs of vascular occlusion including claudication of arms or legs, with which patients commonly present to the vascular surgeon. Chest and back pain, breathlessness, and syncopal attacks may lead to medical consultation. The absence of peripheral pulses and the presence of bruits over affected vessels are characteristic. The renal artery is affected in over one-half of patients: hypertension is therefore common, and renal failure may be a late complication. This disease may also be a cause of inflammatory aortic aneurysms. Visual disturbance and ultimately blindness occurs in chronic cases. Death is usually due to congestive cardiac failure and/or myocardial infarction. The vessel most frequently affected is the subclavian artery (85 per cent). In 12 per cent of patients the abdominal aorta, renal, or mesenteric arteries are involved.

The pathological features are those of a panarteritis, involving all layers of the elastic arteries. Secondary thrombosis and stenotic and aneurysmal lesions are common. Secondary atherosclerotic changes also occur in the damaged vessels.

There is usually evidence of an acute-phase response with elevated levels of C-reactive protein and a rise in the erythrocyte sedimentation rate, but there are no other specific markers. Duplex Doppler ultrasound or MR angiography will demonstrate reduced flow and arteriography will document the extent and nature of the lesions.

Medical treatment with 30 to 50 mg/day of prednisolone will reduce some of the inflammatory response in the early phase, but it is uncertain whether the overall prognosis is affected. A few patients appear to have benefited from cyclophosphamide treatment, but studies of the effects of this drug are confused by the occurrence of spontaneous remission and sudden relapse. Surgical treatment involving appropriate vascular reconstruction is not contraindicated and may be successful in some cases, particularly when local inflammatory change has been damped down or abolished by corticosteroid or other immunosuppressive treatment.

Wegener's granulomatosis

Wegener's granulomatosis can be separated from other forms of arteritis by its characteristic clinical features, by its histopathology, and by its serological marker, the classic antineutrophil cytoplasmic antibody. None of these is, however, specific; even the histopathology is ‘compatible with' rather than ‘diagnostic of' the disease. Firm diagnosis, therefore, rests on the coincidence of typical clinical, histological, and serological features.

The pathology is that of a vasculitis affecting predominantly small arteries and veins, and is marked by a neutrophil and mononuclear cell infiltrate, accompanied by fibrinoid necrosis. Granulomas occur within these vessels and in the surrounding tissues and these have an abundance of multinucleate giant cells. The erythrocyte sedimentation rate and C-reactive protein levels are invariably raised in the acute phase. Serum immunoglobulin levels are also increased, while complement levels are

invariably normal. Antineutrophil cytoplasmic antibodies, usually at high titre, are present in some 90 per cent of patients.

Wegener's granulomatosis can present at any age and either sex can be affected, although the disease occurs predominantly in the middle-aged. Most patients with Wegener's granulomatosis present with malaise, fever, and arthralgia; other features depend on the distribution and activity of the vascular lesions. The upper airways (nose, nasal sinuses, postnasal space, or eustachian tubes) are affected in 90 per cent of patients. These patients usually present to the ear, nose, and throat surgeon with chronic sinusitis, nasal discharge usually with crusting and blood, nasal ulceration, and otitis media. Saddle nose deformity, due to erosion of nasal cartilage, is said to be a typical feature, but it occurs late in the disease.

Parenchymatous lesions in the lung may appear solid on chest radiographs, resembling neoplasms. Central necrosis may lead to fluid levels and the diagnosis of lung abscess or breaking down neoplasm. Infiltrative lesions are commonly misdiagnosed as tuberculosis. These manifestations present with cough, breathlessness, haemoptysis, and chest pain. Reduction of lung diffusing capacity is evidence of generalized interstitial disease, even in the absence of radiographically obvious lesions. Less commonly, submucosal granulomatous lesions sited in the sublaryngeal region may present with the features of extrathoracic obstruction, while localized airways obstruction is caused by lesions lower in the trachea or main bronchi. The clinical presentation of Wegener's granulomatosis affecting the kidney may vary from an abnormal microscopic deposit and modest proteinuria through to rapidly progressive glomerulonephritis, with or without nephrotic features. The disease may be confined to the upper airways, to the lungs, or to the kidney but, in many cases, lesions are found in all three sites. Ocular involvement has been documented with scleritis (Fig. 2), uveitis, and even scleromalacia perforans. When the upper nasal sinuses are affected, granulomas may spread to the retro-orbital space, causing proptosis or disorders of external ocular movement. Cutaneous evidence of vascular damage occurs in around 40 per cent of patients.

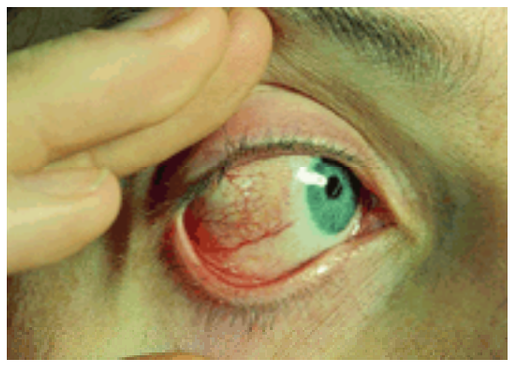


Fig. 2. Episcleritis in a patient with Wegener's granulomatosis.

The treatment of Wegener's granulomatosis is medical and 90 per cent of patients who are untreated will be dead in 2 years and all in 5 years. In patients with life-threatening disease presenting with rapidly progressive glomerulonephritis and/or lung haemorrhage, pulsed methyl prednisolone (1 g daily for 3 days) and pulsed intravenous cyclophosphamide (2 to 10 mg/kg per pulse) can be effective, when followed by continued immunosuppression with prednisolone at 30 to 40 mg/day and cyclophosphamide at 2 to 4 mg/kg.day or azathioprine at 2 to 4 g/kg.day. There is controversy over the relative merits of using longer courses of pulsed intravenous cyclophosphamide compared with oral cyclophosphamide. Cyclophosphamide carries a risk of late bladder cancer and myeloid leukaemia. The risk of the former may be reduced by the concomitant administration of Mesna with intravenous cyclophosphamide. When the presentation appears less immediately dangerous, the same combination of high-dose steroids and cyclophosphamide or azathioprine can be used without the parenteral induction.

The disease appears to be suppressed by such treatment, rather than eradicated, though it may occasionally regress completely and not recur when treatment is withdrawn. More commonly there is a relapse, sometimes after many years. Response to treatment is best monitored by a serial measurement of C-reactive protein and antineutrophil cytoplasmic antibody levels. There is laboratory and clinical evidence that relapse may be precipitated by intercurrent infection. Infusions of pooled immunoglobulin (intravenous Ig) at 0.5 g/kg.day for 4 to 5 days have been shown to be helpful as in Kawasaki disease, although not all patients benefit.

Polyarteritic syndromes

Both microscopic polyarteritis (commonly) and polyarteritis nodosa (less often) may give rise to systemic manifestations, as in Wegener's granulomatosis, of malaise, fever up to 40°C (sometimes the only feature), myalgia, and weight loss. When the disease is confined to these manifestations diagnosis can be particularly difficult, even when it is suspected. Other presenting features, which are mostly medical, depend on the site and severity of the vasculitic process: some are more typical of the nodosa types, while others represent the microscopic categories of the condition. Surgical presentations are not uncommon. The Churg–Strauss variant of polyarteritis comprises evidence of vasculitis, often leading to a severe peripheral neuropathy, accompanied by asthma; blood count demonstrates a significant eosinophilia.

Peripheral vascular disease may occur due to involvement of small vessels, which results in ischaemic lesions or frank gangrene of fingers or toes (Fig. 3). Ischaemic arteritic ulcers may occur particularly in the lower limbs, and there may be associated Raynaud's phenomenon. Embolic lesions or claudication may occur, as in giant cell arteritis, when vasa vasorum vasculitis results in stenosis or aneurysmal lesions of medium-sized vessels.



Fig. 3. The hands of a woman presenting with intractable digital ischaemia, who ultimately died of malignant hypertension as a result of systemic sclerosis involving the kidneys.

Abdominal pain is a well-recognized presentation and probably results from vasculitis in the vessels of the splanchnic circulation. Symptoms are often rather vague and non-specific with ill-defined abdominal pain and occasionally modest diarrhoea or occult blood loss. An area of ischaemic gut or acute appendicitis of vascular origin may perforate, and mesenteric arterial vasculitis may result in a presentation of acute abdomen secondary to infarction of bowel or pancreas.

Renal manifestations present largely to physicians, but infarction of the kidney (in polyarteritis nodosa) may result in loin pain, fever, vomiting, and blood and protein in the urine. This syndrome is often diagnosed as pyelonephritis, despite negative urine cultures, or renal stone disease, despite the absence of firm evidence of a stone. The condition is rare and is best diagnosed by detection of the wedge-shaped infarcts by CT scanning of the kidneys. Clinical vasculitis involving ureter, bladder, epididymis, and testis is rare but may result in hematuria or local pain and inflammation.

Obstructive uropathy due to retroperitoneal fibrosis may be quite commonly associated with a periaortitis (microscopic polyarteritis) surrounding the lower abdominal aorta. Vasculitis is sometimes evident from examination of biopsy material. CT scanning or magnetic resonance imaging in this situation shows soft tissue swelling as well as ureteric obstruction responsive to treatment with corticosteroids; mobilization of obstructed ureters into the peritoneum may be necessary.

The thoracic aorta, particularly the ascending arch, may be affected not only by giant cell arteritis and Takayasu's disease but also in relapsing polychondritis. In this rare condition, vasculitis surrounds tissues in which the glycosaminoglycan content is high. There may, therefore, be inflammation and vasculitic lesions over the cartilage of the ear, nose, trachea, larynx and, more rarely, the sclera and aortic collagen. As well as aortic aneurysm or dissection, perichondritis can result in collapse

of tracheal tissue. Thoracic aortitis is not a common feature of microscopic or nodosa arteritis.

Although coronary arteries are probably rarely affected, this certainly occurs, especially in giant cell arteritis. Vasculitis of the lung, particularly in Wegener's granulomatosis, may be misdiagnosed as a tumour and may present to a surgical team for biopsy or excision. In the Churg–Strauss variant, pulmonary infiltrates, asthma, and eosinophilia predominate and mononeuritis multiplex, which occurs in all forms of vasculitis, is perhaps especially common. Both forms of antineutrophil antibody have been found in this group of patients.

The eye is commonly affected by all vasculitic illnesses, presenting with episcleritis or scleritis more seriously, with vasculitis involving the retinal vessels or the optic nerve, diagnosed by ophthalmoscopy or fluorescein angiography. Anterior uveitis can also occur, but this is much less common. Cranial nerve lesions of the vasa nervorum may produce mononeuritis multiplex of the nerves supplying the external ocular muscles.

Diagnosis of these protean syndromes is often difficult. There are no absolutely diagnostic serological tests but the C-reactive protein level and erythrocyte sedimentation rate are almost always increased and a normochromic normocytic anaemia, together with polymorphonuclear leucocytosis, is common. Complement levels are normal or raised. Selective angiography is often valuable and may demonstrate characteristic small aneurysms, the structures which originally gave rise to the term ‘nodosa'.

Polyarteritis nodosa is considered to be an ‘immune complex disease' based on the link, in some parts of the world, with the presence of hepatitis B virus surface antigen. However, this is not a universal association and most patients in the Western world are not infected with this virus. The treatment is identical to that used in Wegener's granulomatosis, comprising high-dose corticosteroids and immunosuppression with cyclophosphamide or azathioprine. Again, the response to therapy is best monitored by review of clinical manifestations supported by regular review of haemoglobin, white cell count, C-reactive protein level, and erythrocyte sedimentation rate, although the latter is less sensitive.

Systemic sclerosis—scleroderma and CREST syndrome

Although in some patients manifestations of scleroderma are confined to the skin, there is often overlap with the more general organ involvement characteristic of systemic sclerosis. Not all clinicians would classify these disorders as vasculitic in origin, but they are associated with inflammatory cell infiltration and obstruction of small blood vessels with excessive collagen deposition. Again the many manifestations present very largely to physicians. Some particular features often lead to surgical consultations.

Raynaud's phenomenon is particularly common and may be the earliest feature of the disease. It may progress to ischaemic damage of fingers or toes, necrotic ulceration, and ultimately the need for amputation. Other presentations which may be seen by the surgeon include dysphagia due to oesophageal disease; reflux oesophagitis may lead to stricture formation, and a barium swallow study shows the characteristic poor oesophageal motility. Less commonly, sclerosis of the small bowel may lead to disordered mobility, bacterial overgrowth, and malabsorption. Such patients may present with weight loss, nausea, vomiting, and diarrhoea. A small bowel enema shows characteristic local areas of dilatation and pseudodiverticula formation (Fig. 4). In advanced disease there may be complete loss of peristalsis, effectively intestinal obstruction, and some patients progress to a stage at which their survival is dependent on chronic parenteral nutrition. Perforation of the gut has been described and large bowel disease may also present with obstruction or infarction of the colon.



Fig. 4. Small bowel enema in a patient with systemic sclerosis showing dilatation and pseudodiverticular formation.

These syndromes rarely present any difficulty in diagnosis, which is usually obvious clinically. Serologically, they are characterized by a high level of antinuclear antibodies of speckled or nucleolar pattern and a specific autoantibody, Scl-70, which is found particularly in patients with systemic sclerosis. The anticentromere antibody is particularly associated with the **CREST** syndrome (C for Calcinosis circumscripta, R for Raynaud's, E for (o)esophageal disease, S for sclerodactyly, T for telangiectasia), a discrete entity which commonly overlaps with other manifestations of systemic sclerosis. However, the immunology is very distinct as Scl-70 and anticentromere antibodies are never found togther in the same patient. An overlap syndrome of polymyositis and scleroderma may be identified by the presence of the autoantibody PM-Scl. No pharmacological agent is of any proven value in the treatment of systemic sclerosis. Management, therefore, revolves around the amelioration of symptoms and such immediate treatment as the particular clinical manifestations require. Sympathectomy or prostacyclin infusion may be of value, although limited in both time and effect, in the treatment of peripheral ischaemia. Raynaud's symptoms may be controlled by calcium channel blockers, angiotensin-converting enzyme inhibitors, and topical nitrates.

Broad-spectrum antimicrobial (oxytetracycline) therapy is useful in controlling the gastrointestinal manifestations, which depend on overgrowth of bacteria. H2-receptor antagonists or related agents may be helpful in the treatment of reflux oesophagitis and dysphagia, and pacemakers may be needed when disease affects the bundle of His and causes heart block. Total parenteral nutrition can be given if intestinal disease is sufficiently severe.

Systemic lupus erythematosus

Patients with the protean manifestations of this disorder almost always present to the physician, although Raynaud's phenomenon, with or without vasculitis of digital vessels of the upper or lower limbs, may be the only clinical feature in some women, who thus present to a surgical team. The diagnosis is suggested by the pattern of manifestations of the illness. Serological markers include a raised titre of antinuclear antibody, low concentrations of the third and fourth components of complement, increased binding of double-stranded DNA, and/or the presence of anti-Ro, anti-La, and anti-SM antibodies. Elevated levels of antibodies to cardiolipin or antibodies interfering with clotting *in vitro*, the so-called ‘lupus anticoagulant', are not infrequent findings. These antiphospholipid antibodies are mainly associated with otherwise unexplained venous or even arterial thrombosis, and perhaps with a history of recurrent abortion, possibly related to abnormal coagulation in the circulation of the pregnant uterus and resultant placental insufficiency. Major arterial or extensive venous thrombosis in young people should always be followed by investigations for antiphospholipid antibodies.

Behçet's disease

Behçet's disease is a particularly poorly understood illness in which the pathology has a vasculitic component. It is classically characterized by recurrent orogenital ulceration, iridocyclitis with or without retinal vasculitis, and a number of cutaneous lesions. Although it may be more common in Japan and in the countries surrounding the Mediterranean, it is becoming increasingly recognized in the Caucasian population, most commonly in people in the third decade of life. There is a strong immunogenetic background to the disease: it is strongly associated with the HLA B51 antigen in Turkey, Israel, France, and the United Kingdom. HLA DR7 in addition to B51 has been reported to be associated with retinal and neurological manifestations of the disease.

Histopathologically, affected tissues show a lymphomonocytic infiltration around affected epithelia and associated small blood vessels. The latter may be occluded by proliferation of endothelium and associated fibrinoid necrosis.

Mouth ulcers may be difficult to distinguish from those of lesser pathological significance. Ulcers in the genitalia may affect classically the labia, vagina, penis, and scrotum, and there may be an associated epididymo-orchitis. Skin manifestations include classic erythema nodosum as well as pustular lesions which are widely distributed but most common on the face and back. A characteristic feature is the development of pustular lesions at the sites of minor skin damage (‘pathergy').

Among the most important manifestations presenting to surgeons will be those affecting the eye. Uveitis with or without hypopyon, iridocyclitis, retinal vascular lesions

including infarction, optic atrophy, and choroidoretinitis are all dangerous complications and may lead to blindness. Neurological manifestations occur in up to 25 per cent of patients. Any part of the central nervous system may be affected, often in a series of episodes with focal neurological signs; the pontine region is a particular target. If these arise in the hemispheres they may resemble stroke, but manifestations may also occur in the brain stem and cord. On occasion, the findings suggest inflammatory disease of the nervous system, meningitis, or encephalitis with pleocytosis of the cerebrospinal fluid.

Arthralgia and, on occasion, arthritis are commonly reported. Thrombophlebitis of superficial and deep veins may occur and more serious vascular complications include thrombosis, particularly of veins, sometimes such major vessels as the superior and inferior vena cava. Gastrointestinal manifestations are common and include diarrhoea, nausea, anorexia, and abdominal pain. On rare occasions ulceration of both colon and small bowel has been described, as have both malabsorption and pancreatitis. Aneurysmal changes in the aorta and its large branches may also occur. Pulmonary manifestations include shadows on chest radiographs, sometimes associated with haemoptysis and believed to be due to pulmonary vasculitis. Renal manifestations are very rare but both proteinuria and microscopic hematuria have been reported.

Treatment of mouth and genital ulcers with topical ointments or sprays containing both corticosteroids and tetracycline is often very effective, and uveitis may also respond to prednisolone eye drops. The more serious manifestations, particularly those of neurological or retinal disease, are best treated by a combination of corticosteroids and immunosuppressive agents, but the response is uncertain and often disappointing. There are also reports of apparently successful use of colchicine and of aspirin for thrombotic problems. High doses of prednisolone (up to 60 mg a day) appear to be necessary at the onset of serious complications; these are subsequently supplemented by azathioprine in a dose of 2 to 4 mg/kg.day. Cyclophosphamide and low-dose weekly methotrexate have been used, but there is no convincing evidence that they are superior. Cyclosporin A has a particularly valuable role to play in the treatment of retinal and neurological manifestations, and thalidomide, although relatively contraindicated in women of child-bearing age, is beneficial in the treatment of skin and orogenital manifestations. Patients in whom thalidomide is being considered must be counseled, if female, about the risks of pregnancy and all patients should have pretreatment nerve conduction studies, as a peripheral neuropathy is a recognized side-effect.

Kawasaki's disease

This is a systemic illness of children characterized in its onset by lymphadenopathy and mucosal ulceration, with vasculitis sometimes affecting the coronary arteries. The illness presents with fever, conjunctivitis, a skin rash, cervical lymphadenopathy, and oral ulceration. Desquamation of the skin may occur as a late feature. Although originally described in Japan, it occurs worldwide. The cause is likely to be infective, although no pathogen has been identified. The diagnosis is clinical, supported by a non-specific rise in the erythrocyte sedimentation rate and in C-reactive protein levels. Cardiac manifestations of the disease may be detected by electrocardiography (QT prolongation and ST segment changes with or without arrhythmias), and two-dimensional echocardiography or coronary angiography are mandatory to monitor the formation of coronary artery aneurysms and their progress. The pathology of coronary artery aneurysms is characterized by a panarteritis with fibrinoid necrosis and local aneurysm formation. Lymph node biopsy may show necrosis of the node with small vessel vasculitis and proliferation of immunoblasts.

In the acute phase of the disease, symptoms may be settled by high-dose aspirin (30 to 50 mg/kg.day) accompanied by infusions of intravenous immunoglobulin at a dose of 1 g/kg daily for 2 days, but only if given within the first week of the onset of the disease. Corticosteroids are probably of no benefit and may even be harmful.

Buerger's disease (thromboangiitis obliterans)

Whether or not this disorder exists as a specific entity is controversial, especially in the Western world. However, it is a relatively common vascular condition in South-East Asia. The syndrome that suggests its diagnosis most commonly occurs in young men (male to female predominance 9:1) who are smokers. In the lower limb, ischaemic symptoms may produce clinical features ranging from claudication, sometimes felt in the arch of the foot, to ischaemic ulceration of the toes. Less commonly there may be Raynaud's phenomenon or claudication sometimes affecting the hand rather than musculature higher up the limb. Angiography reveals the presence of distal disease, such as obliteration of tibial and/or peroneal arteries in the leg. Histological examination of affected vessels does not reveal atheromatous change, but both arteries and neighboring veins are characteristically infiltrated with neutrophils and small branches may have thrombosed. The only useful therapy is to persuade the patient to abandon smoking.

Lymphomatoid granulomatosis

This condition may be mistaken for Wegener's granulomatosis: the differentiation depends on the report of an experienced histopathologist. In both conditions there is granulomatous vasculitis in affected tissues, which are heavily infiltrated with atypical lymphocytes and plasma cells. Laboratory tests are unhelpful and the erythrocyte sedimentation rate is commonly normal. This disease which, in a significant but uncertain proportion of patients may proceed to malignant lymphoma, may respond well to immunosuppressive treatment with prednisolone and cyclophosphamide or azathioprine.

Henoch–Schönlein purpura

This is a small vessel vasculitis which usually presents in children, although rarer adult forms exist. There are purpuric lesions, especially of extensor surfaces (buttocks, thighs) but the platelet count is normal, arthralgia, hematuria and abdominal pain with gastrointestinal haemorrhage, intussusception, and obstruction. IgA class ANCA may be detected. Poor prognosis is defined by more severe renal involvement and older age. Optimum treatment is undefined: mild cases require no specific treatment, but more severe cases may require immunosuppressive therapy with corticosteroids and cytoxic agents.

Other causes of vasculitis

It is recognized that vasculitis may occur as a secondary event in identifiable infections. These include meningococcal septicaemia, syphilis, tuberculosis, and viral infections, especially hepatitis C. The latter is strongly associated with the development of cryoglobulinaemia as well as other autoimmune phenomena. The cryoglobulins are type II, meaning that there is a monoclonal component with rheumatoid factor activity binding to polyclonal immunoglobulins. Complement C4 is low.

Other causes of ‘pseudovasculitis' include embolization from valve lesions in infectious endocarditis or cardiac myxomas, cholesterol emboli, amyloidosis, and poisonings such as ergotism or excessive methysergide use. Calciphylaxis is a rare syndrome associated with chronic renal failure in which skin necrosis and gangrene occur secondary to a hypsensitivity reaction to calcium deposited in the blood vessels in the presence of parathyroid hormone. It is now recognized that inflammatory bowel disease, ulcerative colitis, and Crohn's disease may be associated with evidence of vascular inflammation, both in the gut and elsewhere.

Immunosuppressive treatment in the management of vasculitis

The mainstay of treatment for most primary inflammatory vasculitides is immunosuppression with high-dose prednisolone and either cyclophosphamide or azathioprine in a dose of 2 to 4 mg/kg.day. High-dose methylprednisolone (500 mg to 1 g daily for 3 days) may be used in conjunction with intravenous cyclophosphamide (1 g/m^2) in patients with life-threatening complications, but the evidence demonstrating the superiority of this approach to high-dose oral therapy is uncertain. Both cyclophosphamide and azathioprine may cause a profound leucopenia and weekly blood counts are mandatory, at least in the initial stages of management. During high-dose and prolonged immunosuppression, administration of prophylactic cotrimoxazole should be considered, since this is known to be useful in preventing *Pneumocystis* pneumonia in transplant recipients. Recent studies have shown that even prolonged high-dose corticosteroids increase the risk of *Pneumocystis* pneumonia. There is some limited evidence that cotrimoxazole may limit relapse in Wegener's granulomatosis, through an unknown mechanism. Cyclophosphamide may also cause alopecia (reversible) and haemorrhagic cystitis, which may be prevented by the concomitant use of mesna. Cyclophosphamide is also toxic to the gonads in both sexes, although the degree of toxicity is variable between individuals. Appropriate warnings must be given to patients, who should be offered the opportunity for sperm banking if a future family is to be considered. Low-dose weekly methotrexate has also been found to be beneficial in a variety of vasculitic syndromes: side-effects include liver and lung toxicity. Long-term use of cytotoxic drugs, particularly chlorambucil and cyclosphosphamide, have been associated with an increased incidence of lymphoid and myeloid malignancy later in life. Cyclosporin A and related immunosuppressive drugs may be helpful under certain circumstances, for example in the control of uveitis and the neuroretinal complications of Behçet's disease. Side-effects include hypertension, renal toxicity, and hirsutism. Newer immunosuppressives such as tacrolimus and mycophenolate are undergoing studies in vasculitis. Plasmapheresis, which has proved effective in myasthenia gravis, antiglomerular basement membrane disease, and in Guillain–Barré syndrome, has also been found useful by some groups, particularly when vasculitis affects the kidney. High-dose intravenous immunoglobulin (1 to 2 /kg) has been shown to be effective in Wegener's granulomatosis, as well as in Kawasaki disease.

Further reading

Ansell BM, Bacon PA, Lie JT, Yazici H, eds. *The vasculitides. Science and practice*. Chapman Hall Medical, 1996.

Barron KS. Kawasaki disease in children. *Current Opinion in Rheumatology* 1998; **10**: 29–37.

Berstein RM. Humoral autoimmunity in systemic rheumatic disease. *Journal of the Royal College of Physicians* 1990; **24**: 18–25.

Chajek T, Fainaru M. Behçet's disease: a report of 41 cases and a review of the literature. *Medicine* 1975; **54**: 179–96.

Chakravarty K. Vasculitis by organ system. *Baillière's Clinical Rheumatology* 1997; **11**: 357–93.

Dahlberg PJ, Lockhart JM, Overholt EL. Diagnostic studies for systemic necrotizing vasculitis. *Archives of Internal Medicine* 1989; **149**: 161–5.

Duna GF, Galperin C, Hoffman GS. Wegener's granulomatosis. *Rheumatic Disease Clinics of North America* 1995; **21**: 949–86.

Hamurydan V, Özdogan H, Yazici H. Other forms of vasculitis and pseudovasculitis. *Baillière's Clinical Rheumatology* 1997; **11**: 335–55.

Hill GL, Moeliono J, Tumewu F, Bratacemadja, Tohandi A. The Buerger syndrome in Java. A description of the clinical syndrome and some aspects of its aetiology. *British Journal of Surgery* 1973; **60**: 606–13.

Hunder GG. Giant cell arteritis and polymyalgia rheumatica. *Medical Clinics of North America* 1997; **81**: 195–219.

Kerr GS. Takayasu's arteritis. *Rheumatic Disease Clinics of North America* 1995; **21**: 1041–58.

Lancet leader: Cogan's syndrome. *Lancet* 1991; **337**: 1011–12.

Lhote F, Guillevin L. Polyarteritis nodosa, microscopic polyangiitis and Churg–Strauss syndrome: clinical aspects and treatment. *Rheumatic Disease Clinics of North America* 1995; **21**: 911–47.

Ohta T, Shinoya S. Fate of the ischaemic limb in Buerger's disease. *British Journal of Surgery* 1988; **75**: 259–62.

Pisani RS, DeRemee RA. Clinical implications of the histopathological diagnosis of pulmonary lymphomatoid granulomatosis. *Mayo Clinic Proceedings* 1990; **65**: 151–63.

Raz I, Okon E, Chajek-Shaul T. Pulmonary manifestations of Behçet's syndrome. *Chest* 1989; **95**: 585–9.

Specks U, Wheatley CL, McDonald TJ, Rohrbach MS, DeRemee RA. Anticytoplasmic autoantibodies in the diagnosis and follow-up of Wegener's granulomatosis. *Mayo Clinic Proceedings* 1989; **64**: 28–36.

Stanford MR, Graham K, Kasp E, Sanders MD, Dummonde DC. A longitudinal study of clinical and immunological findings in 52 patients with relapsing retinal vasculities. *British Journal of Ophthalmology* 1988; **72**: 442–7.

Worrall JG, Snaith ML, Batchelor JR, Isenberg DA. SLE: a rheumatological view. Analysis of the clinical features, serology and immunogenetics of 100 SLE patients during longterm follow-up. *Quarterly Journal of Medicine* 1990; **74**: 319–30.

Zeuner M, Straub RH, Rauh G, Albert ED, Schölmerich AJ, Lang B. Relapsing polychondritis: clinical and immunogenetic analysis of 62 patients. *Journal of Rheumatology* 1997; **24**: 96–101.

17.4 The non-invasive vascular diagnostic laboratory

David S. Sumner

Instruments
Ultrasound
Plethysmography
Laser Doppler
Transcutaneous oxygen sensors
Temperature measurement
Lower extremity arterial disease
Examination methods
Clinical application
Upper extremity arterial disease
Pressure recordings
Blood flow studies
Plethysmographic examination
Documenting vasospasm
Thoracic outlet compression
Acute peripheral arterial obstruction
Peripheral arterial trauma
Cerebrovascular disease
Evaluation methods
Clinical application
Vertebral artery examinations
Transcranial Doppler sonography
Visceral artery disease
Renal arteries
Mesenteric arteries
Aneurysms
Arteriovenous fistulas
Graft surveillance and follow-up studies
Further reading

Over the past three decades, the non-invasive laboratory has assumed an increasingly important role in the work-up of patients with suspected arterial disease. Although clinical evaluation provides valuable clues to the diagnosis, the history is often unreliable, symptoms may be mimicked by other conditions, and the physical examination is subjective and highly dependent on the experience and skill of the examiner. Arteriography—formerly the next step—contributes anatomic information but is of little help in the objective assessment of physiologic impairment. Moreover, arteriography is expensive, time consuming, and carries some risk. By supplying both physiologic and anatomic information, non-invasive testing fills the gap between clinical evaluation and arteriography.

The task of the non-invasive vascular laboratory is to answer the following fundamental questions: is arterial disease present? and, if so, where are the lesions located and how severe and extensive is the obstructive process? Answers to these questions determine the need for further studies (such as arteriography), help decide on the type of and need for therapeutic intervention, and assist in formulating a prognosis. Non-invasive testing also serves as an objective method for following disease progression and for assessing the results of surgical or medical therapy.

This chapter reviews instruments and testing procedures used in non-invasive vascular laboratories and discusses their application to peripheral, cerebrovascular, and visceral arterial disease.

Instruments

A wide variety of non-invasive instruments have been introduced. Some are very simple; others are complicated and incorporate sophisticated technology.

Ultrasound

Of all the diagnostic modalities available in the vascular laboratory, those that employ ultrasound are the most universally applicable. High-frequency sound can be used to study blood flow patterns and to image blood vessels, percutaneously in real-time, without causing discomfort or cellular damage. These attributes make ultrasound ideally suited for non-invasive testing.

Doppler velocity detector

Perhaps the most important single contribution to vascular laboratory instrumentation was the demonstration in 1959 by Satomura in Japan that the Doppler principle can be used to measure blood flow velocity. The concept is simple. High-frequency ultrasound (2 to 10 MHz) generated by a crystal mounted on the end of a probe penetrates the underlying skin and is reflected by acoustic interfaces along its course. Most of these tissue boundaries are stationary, and the frequency of the reflected sound is unchanged. When the beam impinges on a moving target, such as a red blood cell, the frequency of the incident sound wave (f_o) is changed in proportion to the cosine of the angle (q) that the sound beam makes with the velocity vector (v). Frequency increases when the reflector moves toward the sound beam and is decreased when the reflector moves away. In continuous-wave instruments (**CW Dopplers**), the reflected signal is received by a second crystal, also mounted on the end of the probe, and the difference in the frequencies (Df) of the transmitted and reflected signals is calculated electronically.

The following equation expresses these relationships:

$$\Delta f = \frac{2 f_0 \nu \cos \theta}{C}$$

where C is the speed of sound in tissue (about 1.56×10^5 cm/s).

The frequency shift (Df) is in the audible range (20 Hz to 15 kHz) and, therefore, listening to the amplified signal provides a quick and simple method for assessing blood flow. The utility of hand-held ‘pocket' Dopplers depends solely on the audible output. Somewhat more sophisticated instruments provide hard-copy analog tracings of the signal that resemble recordings made with electromagnetic flowmeters. Still more advanced circuitry permits the direction of blood flow to be sensed and displayed on dials or recorded on a strip chart.

Pulsed Dopplers emit time-gated signals that are transmitted and received alternately by the same crystal. By adjusting the time gate, one can focus on blood flow at a certain depth. The sample volume can be restricted to only a millimeter or so, allowing recordings to be made from only the center of the vessel where velocities are usually highest; or the sample volume can be opened up to encompass the entire diameter of the vessel. Pulsed Dopplers are incorporated in duplex scanners and are seldom used alone.

For most clinical and all serious investigative work, analog recordings have been replaced by real-time fast-Fourier analysis of the Doppler frequency spectrum (spectral analysis). Time is displayed on the horizontal axis, and a continuous gray-scale representation of all detected frequency shifts is shown on the vertical axis (Fig. 1). The intensity of the gray scale is proportional to the relative number of red blood cells moving at a specific velocity. The frequency envelope represents the shape of the flow pulse and corresponds to the maximum velocity at the site being interrogated. Forward flow is depicted by frequencies above a zero baseline, and reversed flow by frequency shifts below.

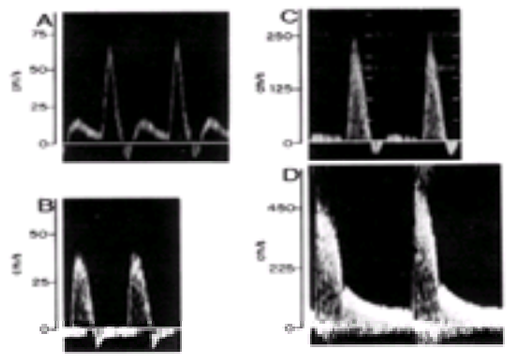


Fig. 1. Spectral analysis of Doppler flow signals. (A) Normal; (B) 1 to 19 per cent diameter stenosis; (C) 20 to 50 per cent diameter stenosis; (D) greater than 50 per cent stenosis.

In any given cross-section of a blood vessel, red cells are moving at a variety of rates; therefore, frequency shifts recorded by the Doppler vary from zero at the arterial wall to a maximum value (usually) at the center of the flow stream. Recordings made with CW Dopplers, which insonate the entire lumen, display a wide range of frequencies simultaneously, completely filling the space enclosed by the maximum frequency envelope. In contrast, pulsed-Doppler recordings made from a small sample volume at the center of a normal artery display a narrow range of frequencies adjacent to the flow envelope. This leaves a large ‘window' of pixel-free space beneath the envelope that is characteristic of an undisturbed flow pattern.

Equation 1 can be rearranged to obtain velocity as a function of the frequency shift:

$$v = \frac{C\Delta f}{2f_0 \cos \Theta}$$

Most Duplex scanners have computer programs that make the calculations necessary to convert frequency shifts into velocities, provided the angle of insonation is known.

B-mode imagers

B-mode imagers measure the time required for a pulse of ultrasound to travel from the probe to an underlying acoustic interface and return to the transducer. Intensity of the echo is shown on a gray scale and time on an axis in line with the probe. As the sound beam is swept over the area of interest, a two-dimensional cross-sectional image appears on the monitor. With these devices, the interface between vessel walls and the flow stream is easily seen. Early instruments produced a static ‘pie-shaped' sector scan when the probe was swept manually though an angle over the area of interest. Now most vascular laboratories use ‘real-time' imagers that either rotate or oscillate the sound beam mechanically to provide a sector scan or use electronic methods to activate sequentially a linear array of crystals mounted on the end of a probe to provide a more easily interpreted rhomboidal scan (Fig. 2).

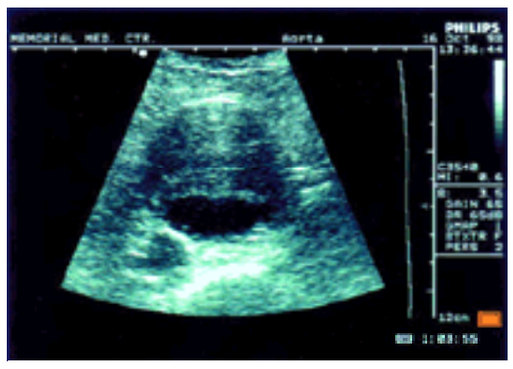


Fig. 2. Real-time B-mode sector scan of 5-cm abdominal aortic aneurysm, showing anterior clot and a residual posterior lumen.

Duplex and color-flow scanners

Next to the hand-held CW Doppler, the most useful instrument in the vascular laboratory is the Duplex scanner. First introduced in the late 1970s, these instruments have undergone many refinements and are now sophisticated electronic machines that provide a wealth of information. Basically, duplex scanners integrate the flow-recording capabilities of the Doppler and the imaging capabilities of the real-time B-mode scanner. These two modalities are synergistic in that the B-mode image permits precise placement of the Doppler sample volume for flow interrogation, and the Doppler signal aids in the identification of the vessel being imaged. The angle that the Doppler sound beam makes with the flow stream is shown graphically on the video monitor and is measured electronically, facilitating the conversion of frequency shifts to velocities. Elsewhere on the screen, the Doppler frequency (velocity) spectrum is displayed (Fig. 3).

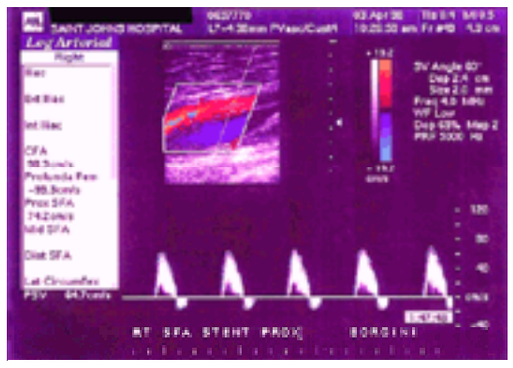


Fig. 3. Color-duplex scan and velocity spectrum of a superficial femoral artery stent with 20 per cent diameter stenosis. Angle of insonation is shown by the diagonal bar and the cursor aligned with the flow stream. Sample volume is indicated by the two parallel bars.

Color-coded flow-mapping was added to duplex scanners in the latter part of the 1980s. Color-flow imaging superimposes a real-time flow map on the B-mode image wherever moving particles are detected. Color (red or blue) identifies the direction of flow relative to the probe, and the color saturation corresponds to the frequency shift. Slow flow gives a deep color, while high velocities approach white (or another color, such as yellow or green, programmed to correspond to a specified velocity). A light blue color, also signifying a high velocity, may be seen when aliasing occurs (Fig. 4). Absence of color (a black image) indicates no flow. Color distinguishes between vascular and avascular structures and between arteries and veins, and it immediately identifies sites where velocities are high or where flow disturbances are present.

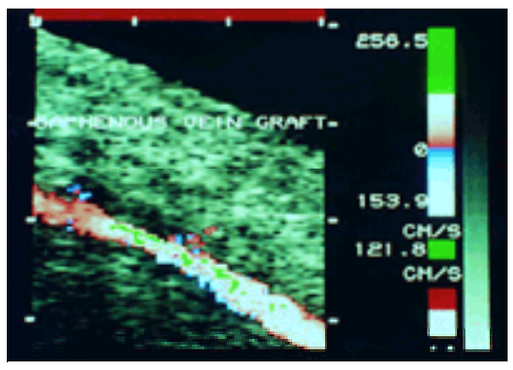


Fig. 4. High-grade stenosis in a saphenous vein graft. Color changes (white, green, and light blue pixels) indicate high-velocity flow.

With the duplex scanner, hard-copy images of specific sites can be made, and critical parts of the study can be recorded on videotape for later review. Volume flow rates can be estimated based on mean blood flow and the diameter of the vessel. This can be accomplished electronically.

Plethysmography

In the early 1960s, plethysmographs were the only instruments available for objective evaluation of the peripheral vascular system. Indeed, many of the fundamental concepts of vascular laboratory testing were based on plethysmographic studies. Although plethysmography continues to be a valuable tool, its role in the modern vascular laboratory is diminishing as ultrasonic instruments assume more of the load.

P>The sole purpose of a plethysmograph is to measure volume change. With the exception of the lung, transient changes in the volume of all organs or body parts are attributable to increases or decreases in their blood content. Plethysmographs are useful, therefore, for measuring changes in venous blood volume and for recording arterial pressure pulsations.

A variety of types are in use. The air plethysmograph (pulse volume recorder) employs an air-filled bladder, which is wrapped snugly around the thigh, calf, or foot. Any change in volume of the enclosed body part causes an increase or decrease in the air pressure within the cuff. The recorded pressure changes are proportional to volume change. This is a rugged and practical instrument, but one that suffers from a low-frequency response.

The mercury strain-gauge plethysmograph consist of a thin Silastic tube filled with mercury (or an indium–gallium alloy). The tube is wrapped around the part being examined. Any change in volume of the part increases or decreases its circumference. This in turn stretches (or relaxes) the mercury-filled tube, changing its resistance. The relative change in resistance is equal to the relative change in volume of the body part. Mercury strain gauges are delicate but very sensitive and very accurate.

The photoplethysmograph is not truly a plethysmograph in that it does not actually measure volume change and cannot be calibrated reliably. Infrared light emitted from a diode is beamed through the skin, and the reflected portion is detected by a photosensor mounted adjacent to the light source. The output is proportional to the concentration of red blood cells in the underlying microvasculature. Amplified signals closely resemble tracings recorded with a mercury strain gauge. Because of its simplicity and sensitivity, the photoplethysmograph has become a popular non-invasive instrument.

Ocular pneumoplethysmographs have a limited role confined to the investigation of cerebrovascular obstruction. These devices employ small plastic cups that are placed over the sclera of each eye and held in place by a vacuum. As the eyeball contracts or expands, pulsations from the ophthalmic artery are transmitted to the cups. Increasing the vacuum to 300 to 500 mmHg distorts the globe and increases intraocular pressure until blood flow to the eye is cut off. As the vacuum is gradually reduced, the point at which pulsations reappear is proportional to the ophthalmic pressure (Fig. 5). An empirically determined relationship is used to convert vacuum levels to ophthalmic pressure (in mmHg).

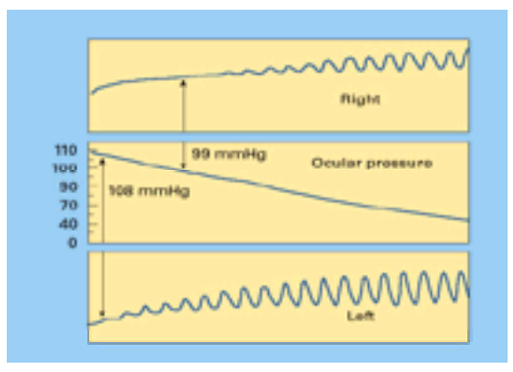


Fig. 5. Ocular pneumoplethysmography in a patient with hemodynamically significant disease of the right internal carotid artery.

Laser Doppler

Laser Doppler flowmeters record shifts in the frequency of monochromatic light reflected from red blood cells in the skin at a maximum depth of 1.5 mm. Although the principle is similar to that on which the ultrasonic Doppler is based, the output depends not only on the velocity of red blood cells but also on the quantity of red blood cells in the sample volume. Moreover, owing to the complex geometry of the cutaneous microvasculature, the light beam intersects underlying capillaries at multiple angles, making quantification impossible. The recorded signal is said to be proportional to ‘red blood cell flux' and is measured in millivolts of deflection or arbitrary units. Despite these drawbacks, the laser Doppler has proved useful in the qualitative assessment of regional cutaneous blood flow.

Transcutaneous oxygen sensors

Polarographic electrodes applied to the skin may be used to measure oxygen tension (transcutaneous PO_2) or carbon dioxide tension (transcutaneous PCO_2). The results are reported in millimeters of mercury and are related to the oxygen supplied by the influx of blood and the rate of utilization in the cutaneous vascular bed. Because oxygen tension diminishes only when blood flow is severely compromised, measurements are useful for evaluating ischemia and the potential for healing.

Temperature measurement

Skin temperature is determined by skin blood flow as well as by a host of environmental factors. In the absence of blood flow, skin temperature approaches that of the room; in hyperemic states, it approaches that of the body core. The relationship between skin temperature and cutaneous blood flow is markedly non-linear, making it useful for qualitative evaluations only. Thermistors applied to the digits are helpful in evaluating vasospastic diseases. Infrared thermography may be useful for mapping regional blood flow patterns.

Lower extremity arterial disease

Physiologic manifestations of arterial obstruction include decreased peripheral arterial pressure, distortion of pressure and flow waveforms, and, at the site of stenoses, flow acceleration and turbulence.

Examination methods

Indirect blood pressure measurement

Ankle pressure

P>No test for evaluating the circulatory status of the leg is more simple to perform, more easily interpreted, or more informative than measurement of the systolic blood pressure at the ankle. This should be the first test performed on any patient with suspected lower extremity arterial disease.

With the patient supine, a pneumatic cuff is wrapped around the ankle, and a hand-held CW Doppler probe is positioned over a pedal artery (Fig. 6). After a flow signal has been obtained, the cuff is inflated until the signal disappears. As the cuff is slowly deflated, the pressure at which the Doppler signal reappears is noted (Fig. 7). Measurements are obtained from both the posterior tibial and the dorsalis pedis artery. The higher of the two pressures is taken as 'the ankle pressure', but both are recorded. In the event that signals cannot be obtained from either of these two arteries, it is often possible to locate the perforating branch of the peroneal artery (lying anterior to the lateral malleolus). Pressures obtained from this artery are probably somewhat less reliable than those obtained from the dorsalis pedis or posterior tibial arteries. When no pedal arterial Doppler signals can be located, a strain gauge or photoplethysmograph applied to a toe may be used to sense the return of flow (Fig. 7).



Fig. 6. Ankle and toe blood pressure measurement by the pneumatic cuff method. Flow in the posterior tibial artery is sensed with a Doppler probe, and toe pulses and volume change are detected with a photoplethysmograph.

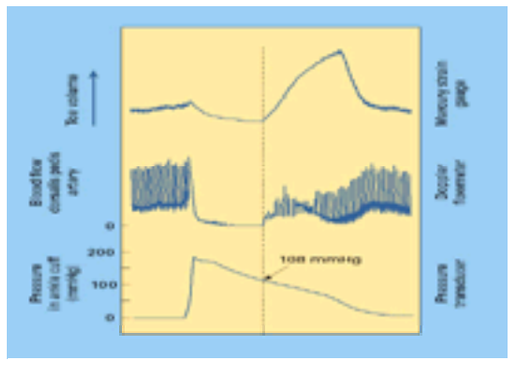


Fig. 7. Systolic blood pressure at the ankle measured simultaneously with a Doppler probe over the dorsalis pedis artery and a mercury strain-gauge plethysmograph on the second toe.

In normal subjects and in patients with lesions that narrow the arterial diameter by less than 50 per cent, the resting ankle pressure usually exceeds the brachial pressure. Because ankle pressure varies with central aortic pressure, results may be normalized by dividing the ankle pressure by the higher of the two brachial systolic pressures to obtain a ratio, called the ankle/brachial index (**ABI**). The normal ABI averages about 1.10 ± 0.10. Values less than 0.92 are indicative of hemodynamically significant arterial obstruction at some point in the arterial tree from the aorta to the ankle. Patients whose only complaint is intermittent claudication have a wide range of ABIs, with average values of 0.60 ± 0.15. ABIs are typically much lower in limbs with rest pain or gangrene (Fig. 8). An absolute ankle pressure of 40 mmHg or less is indicative of severe circulatory compromise, regardless of the ABI.

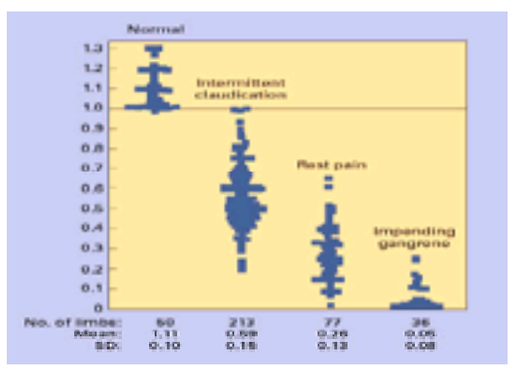


Fig. 8. Relationship of the ankle/brachial pressure index (ABI) to functional impairment produced by arterial obstruction.

The ABI is inversely correlated with arterial resistance and, therefore, lower values are measured when there is multilevel disease and when, in addition to the main artery, entrance or exit points of collaterals are blocked. Occlusions ordinarily reduce the ABI more than stenoses, but there are many exceptions to this rule. In general, ABIs reflect the degree of circulatory impairment at the ankle and are not predictive of the anatomic distribution of disease.

ABI measurements are sufficiently consistent from one examination to the next to make them useful for monitoring the natural history of arterial disease or for assessing the results of interventional therapy. However, variations as large as 0.15 may occur in a hemodynamically stable limb. Although a smaller change in ABI may be significant, the difference between studies should exceed 0.15 before one can be certain.

Ankle pressures may be unobtainable or may be artificially high when distal arteries are rendered incompressible by medial calcification. This problem, which is particularly common in patients with diabetes, is easily recognized. Toe pressures, alone or coupled with Doppler or plethysmographic studies, are helpful in this circumstance.

Toe pressure

Next to the ankle pressure, the toe pressure is the most valuable non-invasive test for assessing lower extremity arterial obstruction. Because digital arteries are seldom calcified, there is little difference between the mean value of toe pressures in patients with or without diabetes with similar symptoms (Fig. 9). For this reason,

toe pressures are particularly informative in patients with diabetes with incompressible ankle arteries. They are also useful in patients in whom pedal or digital arterial disease is the major (or the only) site of obstruction.

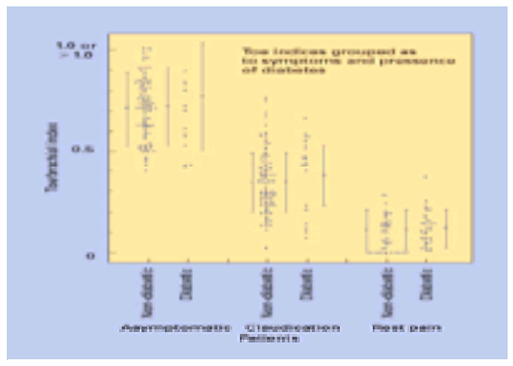


Fig. 9. Toe/brachial pressure indices grouped according to symptoms and presence or absence of diabetes in patients with arterial disease. Vertical bars indicate means and standard deviations for non-diabetic, diabetic, and both groups combined. Data adapted from Ramsey DE, Manke DA, Sumner DS. Toe blood pressure: a valuable adjunct to ankle pressure measurement for assessing peripheral arterial disease. *Journal of Cardiovascular Surgery* 1983; **24**: 43–8.

The technique for measuring toe pressures is analogous to that used at the ankle. A small pneumatic cuff is wrapped around the base of the toe and a photoplethysmograph (or mercury strain gauge) is used as a flow detector (Fig. 6). The reappearance of pulses or, in the absence of pulses, an abrupt rise in the plethysmographic tracing signifies the return of blood flow as the cuff is deflated.

Toe pressures are normally somewhat lower than the corresponding ankle pressure. This is particularly true in older patients and those with diabetes, in whom pedal artery disease is common. A toe/brachial index of 0.70 or greater is probably normal, while an index less than 0.50 is associated with symptomatic arterial disease (Fig. 9). Toe pressures less than 30 mmHg or toe/brachial indices less than 0.20 are consistent with severe ischemia and poor healing potential.

Segmental pressure

Pneumatic cuffs wrapped around the upper thigh, above the knee, and around the upper calf can be used to estimate blood pressure at these segments of the leg. Return of blood flow may be detected with the Doppler probe positioned over any suitable artery below the site of the cuff, but usually the pedal artery location is most convenient. Blood pressures at the same anatomic level are recorded from both legs. At any level a difference between legs of 20 mmHg or more is probably significant. A pressure gradient exceeding 30 mmHg between two adjacent levels of the same leg implies arterial obstructive disease in the intervening segment.

The results of segmental pressure measurements, however, are much less accurate than they are at the ankle. In obese legs, the upper thigh pressure is always spuriously elevated. As a general rule, an upper thigh/brachial pressure index exceeding 1.20 is probably normal, while an index less than 0.80 is probably consistent with inflow disease.

The role of segmental testing (to locate the site or sites of arterial obstruction) has largely been assumed by duplex scanning. In our laboratory, the test is seldom performed.

Plethysmographic studies

The volume of blood at any instant in the peripheral vascular bed is a function of the arterial pressure and, therefore, plethysmographic pulses closely resemble the arterial pressure waveform (Fig. 10). Normal pulses have a rapidly rising upslope, a distinct sharp peak, and a downslope that bows toward the baseline. A ‘dicrotic notch' or wave is frequently present on the downslope. Pulses recorded beyond a hemodynamically significant obstruction have a slower upslope, a rounded peak, and a downslope that bows away from the baseline. The dicrotic notch is lost. As the obstruction becomes more severe, the pulse volume tends to decrease (flatten out) and may disappear entirely in ischemic limbs.

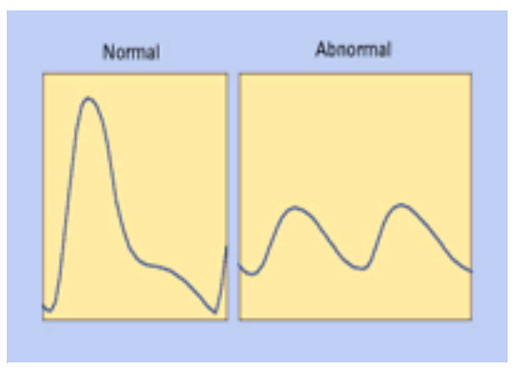


Fig. 10. Plethysmographic pulses from the toes of a normal limb and a limb with occlusive arterial disease.

Digital plethysmograms are obtained by placing a mercury strain gauge around the distal phalanx or by applying a photoplethysmograph to the tip of the toe. An abnormal tracing is indicative of arterial obstruction at the digital, pedal, or more proximal level. The amplitude of the toe pulse, which is roughly proportional to blood flow, provides a qualitative index of local tissue perfusion. Absence of a pulse implies severe disease. Recordings must be made with the patient in a warm room (22°C) to avoid vasoconstriction, which can mimic the changes associated with ischemia. Supplemental warming of the foot may be necessary.

Toe pulses are helpful in conjunction with digital pressures for evaluating disease in arteries distal to the ankle (Fig. 11). Abnormal pulses are very specific for the presence of arterial obstruction but are less sensitive than pressure studies. Normal or near normal waveforms may be present even when the ankle or toe pressure is mildly but distinctly reduced. In patients with pedal or digital artery disease, recording pulses from several or all toes may provide valuable information concerning the distribution of arterial obstruction.

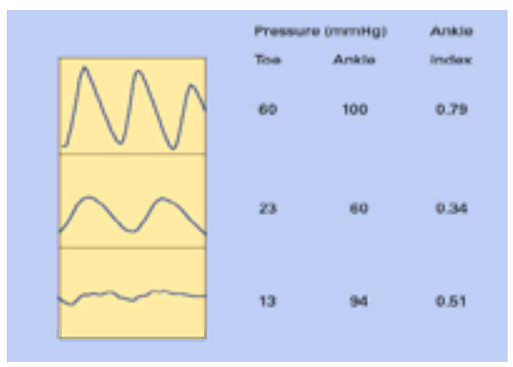


Fig. 11. The contour of the toe pulse correlates closely with the toe pressure but less well with the ankle pressure or ABI (ankle index). *Upper tracing*, good digital

artery perfusion; *middle tracing*, borderline perfusion; *lower tracing*, ischemia.

Segmental plethysmograms may be obtained by positioning air-filled cuffs around the thigh, calf, ankle, or foot. These pulse volume recordings are interpreted in much the same way as toe plethysmograms, but more emphasis is placed on the amplitude of the waves. Based on amplitude and the presence or absence of a dicrotic wave, five categories of obstruction have been defined, ranging from 1 (normal) to 5 (most severe) (Table 1). Pulse volume recordings are primarily useful as a method for determining the approximate location of arterial obstructions (Fig. 12). Now that more precise ultrasonic methods are available, the pulse volume recording is less frequently used.

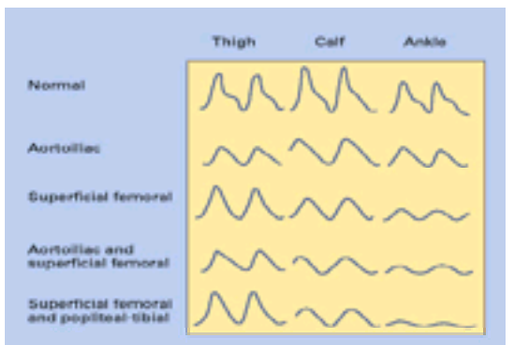


Fig. 12. Tracings of pulse volume recordings from normal limbs and limbs with various combinations of arterial obstruction.

	Chart deflection (mm)	
Category	*Thigh and ankle*	*Calf*
1	> 15°	> 20°
2	> 15[ϕ]	> 20[ϕ]
3	5–15	5–20
4	< 5	< 5
5	Flat	Flat

°With dicrotic notch.
ϕWithout dicrotic notch.

Table 1 Pulse volume recorder categories.

Blood flow studies

Flow may be studied with simple hand-held Doppler devices or with more sophisticated duplex instruments. The information retrieved is helpful not only for recognizing the presence of arterial disease but also for locating lesions and assessing their hemodynamic impact.

Normal (resting) peripheral arterial flow waveforms have a steep upslope culminating in a pointed peak. Flow then rapidly decelerates to the zero level. This is followed by a short period of reversed flow in early diastole and a low-velocity forward flow phase that extends through the remaining pulse cycle (Fig. 13). Audible signals are characterized by an initial high frequency component followed by two lower frequency sounds. These triphasic signals are easily recognized by the experienced observer. After exercise or after restoration of flow following several minutes of ischemia (reactive hyperemia), velocities are greatly increased and the reversed flow component disappears. Audible signals become biphasic or monophasic.

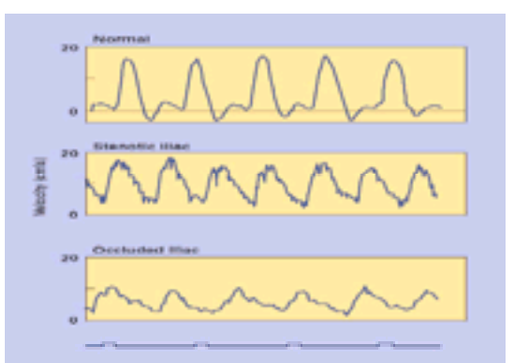


Fig. 13. Analog recordings of Doppler signals obtained from the common femoral artery of a normal subject, a patient with stenosis of the iliac artery, and a patient with iliac arterial occlusion.

Distal to a hemodynamically significant obstruction, waveforms become more rounded. The acceleration phase is less steep, the peak is less clearly defined, the reversed flow component disappears, and velocities remain above the baseline throughout diastole (Fig. 13). Audible signals become monophasic, more noisy, and have a lower peak frequency. At the site of a stenosis, the frequency of the audible signal increases commensurate with the increase in velocity. Absence of a signal implies total arterial obstruction or flow velocities below the threshold level. Just above an occlusion, the signal takes on a ‘thumping' quality, reflecting to and fro motion in the stagnant column of blood.

The hand-held CW Doppler is useful for initial screening and for qualitative assessment of blood flow in the pedal arteries. More accurate and more comprehensive information can be retrieved when pulsed-Doppler signals are subjected to spectral analysis. These recordings are best made in conjunction with duplex scanning, which permits accurate placement of the Doppler sample volume in the center of the flow stream, where velocities are usually maximum. In normal peripheral arteries, the envelope of the velocity spectrum closely resembles that obtained with analog recordings (Fig. 1). Typically, a narrow band of frequencies parallel the envelope, leaving a signal-free window below. If the diameter of the artery is small, if the sample volume is extended to encompass the entire lumen, or if the sample volume is placed near the arterial wall, an extended range of frequencies may be recorded corresponding to a wide range of velocities; and the window is reduced or disappears (spectral broadening). The window is also reduced or absent in the presence of disturbed or turbulent flow.

To estimate the severity of an arterial lesion, the peak systolic velocity in the throat of the stenosis is divided by that in a relatively ‘normal' segment of artery a few centimeters above or below the site of maximum narrowing to obtain a velocity ratio. A velocity ratio exceeding 2.0 is consistent with a diameter reduction of 50 per cent or more. The reversed flow component also disappears, diastolic flow is elevated, and there is extensive broadening of the spectrum with loss of the window (Fig. 1). A velocity ratio greater than 3.0 implies a stenosis exceeding 70 per cent. Less severe narrowing (20 to 49 per cent diameter reduction) is associated with variable spectral broadening, and a velocity ratio in the 1.3 to 2.0 range. These criteria are summarized in Table 2.

Diameter of	*Velocity*		*Diastolic*	*Spectral*
stenosis (%)	*ratio*	*Waveform*	*flow reversal*	*broadening*
Normal	1.0	Triphasic	Present	None
1–19	< 1.3	Triphasic	Present	Minimal
20–49	1.3–2.0	Triphasic	Diminished	Prominent
50–99	> 2.0	Monophasic	Absent	Extensive
Occluded	No flow			

*Modified from Zierler RE. Arterial duplex scanning. In: Rutherford RB, ed., Vascular surgery, 4th edn., pp. 120–30. WB Saunders, Philadelphia, 1995.

Table 2 Duplex scan criteria for classifying lower extremity arterial lesions.*

Conventional and color-flow duplex scanning

B-mode images of the femoral, profunda femoris, popliteal, tibial, and peroneal arteries are readily obtained. The aorta is also easily visualized, but there may be some difficulty securing a good view of the iliac arteries because of their depth and the confounding influence of bowel gas. Gray-scale images of arteries often show atherosclerotic plaques, but estimating their severity requires evaluation of Doppler flow patterns. The color-flow map is a very helpful addition. Color assists in the identification of arteries and veins (Fig. 14). As the artery is scanned, a change of color from red to white (yellow or green) identifies sites of increased velocity, usually caused by a stenosis (Fig. 4). Although the color image may also be narrowed at this point, the degree of stenosis cannot be measured reliably either from the B-mode scan or the color-flow map. This is best accomplished by pulsed-Doppler interrogation of the suspected site of stenosis. Absence of color in a vessel depicted on the gray-scale B-mode image indicates total occlusion. Because clot may be echolucent, the B-mode image alone may fail to detect thrombotic obstruction.

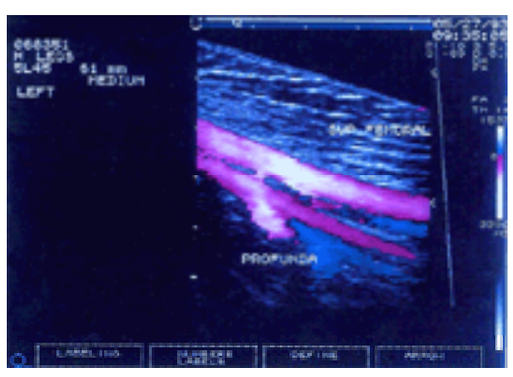


Fig. 14. Color-flow duplex scan of superficial femoral and profunda femoris arteries just below common femoral bifurcation. Arteries are in red, veins in blue.

Laser Doppler studies

Evaluation of laser Doppler flux may be helpful in determining the degree of cutaneous ischemia in specific areas of the foot. Recordings from normally perfused skin are characterized by pulse waves that coincide with the cardiac cycle, vasomotor waves that occur four to six times a minute, and a mean flow velocity (flux) represented by an elevation of the tracing above baseline. Arterial obstruction attenuates pulse waves, decreases mean velocity, and reduces vasomotor waves.

Methods of estimating skin blood pressure have been developed that use compression cuffs and the laser Doppler as a flow sensor, much in the way that plethysmographs are used to record toe pressure.

Oxygen tension

Transcutaneous PO_2 measurements obtained from the foot, calf, or other areas of interest are useful for evaluating the degree of ischemia and potential for healing. Oxygen tension in the skin is a function of arterial PO_2 and, therefore, peripheral measurements should be compared with a well-perfused reference point, such as the infraclavicular skin. In normal limbs, a transcutaneous PO_2 of 60 mmHg is about average. Values less than 20 mmHg are compatible with limb-threatening ischemia. Below 10 mmHg, healing is unlikely to occur.

Exercise testing

Claudication severity can be estimated by having the patient walk on a treadmill at a standardized rate and elevation. Many laboratories use 3.2 km/h (2 mph) and a 12 per cent grade. Most normal subjects, regardless of age, are able to walk for 5 min at this rate without experiencing significant discomfort. Patients with arterial obstructive disease typically develop pain after a few minutes and are usually forced to stop after 2 to 3 min. Prior to walking, patients rest in a supine position for about 20 min. Just before the patient steps on the treadmill, baseline ankle pressures are obtained from both limbs. Immediately after walking, the patient resumes a supine position and ankle pressures are measured as quickly as possible. Thereafter, pressures are measured every minute or two until they return to baseline or until 20 min has elapsed.

In normal limbs, ankle pressures remain relatively stable, perhaps even increasing somewhat with the rise in systemic blood pressure that usually accompanies exercise. In limbs with arterial obstruction, ankle pressures and ABIs decrease (sometimes falling to zero) immediately after exercise (Fig. 15). Recovery to pre-exercise levels is gradual and may require more than 20 min. Claudication time (time to the first definitive symptom), total walking time, the magnitude of the pressure drop, and the time required to return to baseline levels are roughly proportional to the severity of the arterial obstruction.

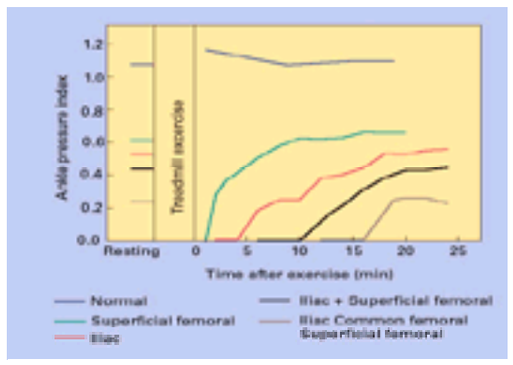


Fig. 15. Ankle/brachial indices of a normal subject and patients with severe arterial obstruction before and after treadmill exercise. Location and extent of obstruction(s) are indicated.

Clinical application

Tests should be selected based on the patient's history and physical examination and should be designed to answer specific questions concerning diagnosis, prognosis, and treatment. Routine use of a battery of tests wastes the resources of the vascular laboratory.

Intermittent claudication

In patients presenting with a typical history of intermittent claudication, finding an abnormal ABI is usually sufficient to establish the diagnosis (Fig. 16). No other tests are required. Treadmill testing is helpful when symptoms are atypical or when the ABI is normal or nearly so. Pain consistent with claudication in the calf or thigh muscles together with a drop in the ABI confirms the diagnosis. Other explanations for exercise-induced leg pain must be sought when the ABI does not fall. Common causes include arthritis and spinal stenosis (pseudoclaudication). Angina, dyspnea, or pain localized to the knee or hip may prove to be the limiting defect despite a drop in the ABI, especially when the magnitude of the pressure decrease seems inconsistent with the severity of the symptoms. In this event, the limiting defect should be addressed before arterial lesions are treated.

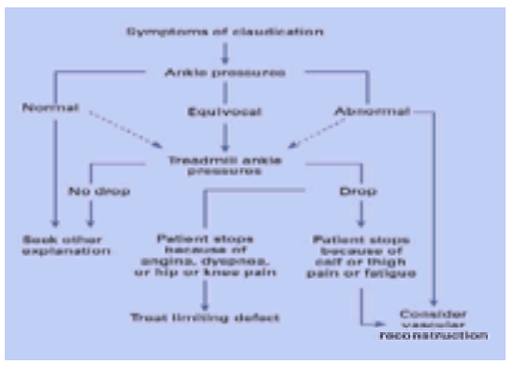


Fig. 16. Diagnostic approach to patients complaining of intermittent claudication.

When arteries at the ankle are incompressible, arterial disease may be ruled out or confirmed by measuring toe pressures, or by plethysmographic studies, Doppler surveys, or duplex scanning.

Ischemia

Measurement of ankle pressure is again the first step in the evaluation of patients with symptoms or signs suggestive of ischemia (rest pain, non-healing ulcers, or gangrene). An abnormal ankle pressure in the ischemic range (30 to 40 mmHg or less) confirms the diagnosis of critically severe obstruction located in arteries proximal to the ankle (Fig. 17). There is no need for exercise testing. Because spontaneous healing or recovery from rest pain is unlikely, further investigations should be directed toward determining the feasibility of interventional therapy.

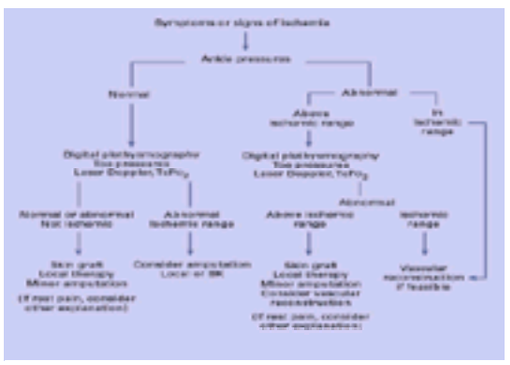


Fig. 17. Diagnostic approach to patients with rest pain, local gangrene, or non-healing ulcers of the foot.

If, on the other hand, the ankle pressure is falsely elevated due to calcification or if it is normal, or abnormal, but above the ischemic range, toe pressures should be obtained (Fig. 17). Digital plethysmography, transcutaneous PO_2 measurements, and laser Doppler studies may also be helpful. Normal results or values well above the ischemic range make it unlikely that pain or ulcers are caused by ischemia. Abnormal findings localized to one or more digits or to specific areas of the foot suggest isolated pedal or digital artery disease. When tissues adjacent to ischemic areas are adequately perfused, local amputation or debridement and skin grafting may be possible. Arterial reconstruction may be beneficial and should be considered in all patients with generalized or localized pedal ischemia when the ankle pressure is significantly reduced (even though it is above the ischemic range). Under these circumstances, a successful reconstruction, by normalizing or improving inflow pressure at the ankle, will often increase blood flow in previously ischemic regions enough to permit healing or to make local amputation possible.

Predicting healing

Healing of ulcers or digital amputations is unlikely when the toe pressure is less than 30 mmHg and practically never occurs when the pressure is below 20 mmHg (Fig. 18). As a rule, the potential for healing is good when toe pressures exceed 30 mmHg. Transcutaneous PO_2 levels less than 10 mmHg or laser Doppler flux less than 40 mV are also poor prognostic signs, but cut-off points are difficult to define and vary from one laboratory to the next. In our laboratory, a transcutaneous PO_2 of 30 to 40 mmHg has generally predicted healing.

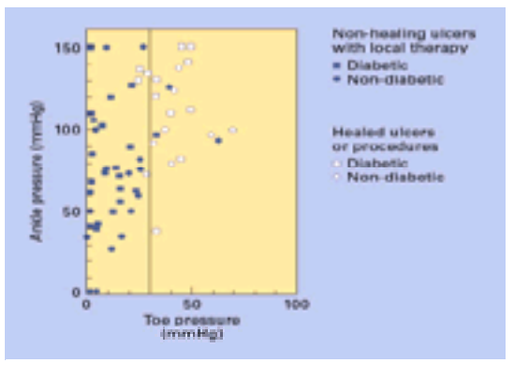


Fig. 18. Ankle and toe pressures in limbs of patients, with and without diabetes, with ulcers that healed with local therapy and ulcers that failed to heal. Note that a toe pressure of 30 mmHg provided good separation between limbs that healed and those that did not, while ankle pressure failed to discriminate between the two outcomes.

Although a normal ankle pressure or one above the ischemic range cannot be used to predict healing of foot or toe lesions, a pressure of 30 mmHg or more at this level provides reasonable assurance that a below-knee amputation will heal.

Locating sites of obstruction

Once the decision has been made to consider interventional therapy, further studies are required to define the anatomic location, extent, and severity of the obstructing lesion or lesions. This information is necessary to determine the feasibility and type of intervention (surgical reconstruction or endovascular therapy) and to plan a therapeutic approach. Although segmental pressure measurements, plethysmographic studies, and Doppler surveys are capable of roughly defining the location and extent of disease, the findings are too imprecise to be of much value in planning intervention. Duplex and color-flow scanning, on the other hand, can identify and locate arterial lesions, sometimes with sufficient accuracy to obviate the need for arteriography. While arteriography continues to be necessary in the majority of cases, duplex findings have proved useful in predicting which lesions may be amenable to endovascular approaches and which lesions are so extensive that endarterectomy or bypass grafting is required.

When tibial arteries or other targets for distal anastomoses are poorly defined arteriographically, their patency and suitability for distal anastomosis may be confirmed with duplex scanning. In other cases, duplex scanning can be used to supplement magnetic resonance angiography.

Evaluating functional results of arterial surgery

After successful reconstructive surgery, there will be an increase in the ABI. If all hemodynamically significant lesions have been removed or bypassed, the index should exceed 1.0. When, however, there are residual sites of obstruction, the ABI will increase, albeit not to normal levels (Fig. 19). Graft patency, therefore, does not necessarily imply an ideal result.

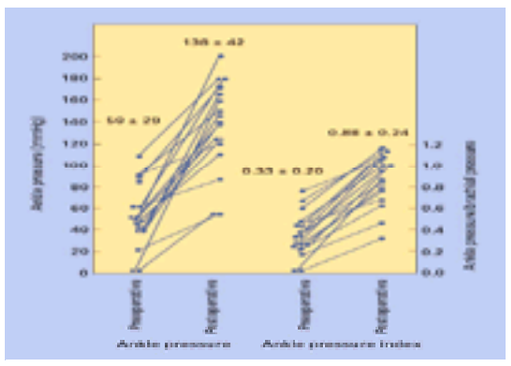


Fig. 19. Ankle pressures and ABIs before and after ‘successful' femoropopliteal bypass. Note that all improved; but normal indices (> 0.90) were achieved in only a fraction of the limbs, owing to residual tibioperoneal disease.

Upper extremity arterial disease

Obstructive arterial disease is less common in the upper extremity than it is in the lower extremity, but the variety of conditions responsible is considerably broader. In addition to ‘fixed' arterial obstruction, patients present to the vascular laboratory with intermittent compression syndromes (at the thoracic outlet) and with episodic peripheral vasospasm (cold sensitivity) of the Raynaud's type. Scleroderma and other connective tissue diseases, giant cell arteritis, vibratory trauma, and percussion injuries are other problems that are encountered not infrequently. Although vascular laboratory tests seldom identify a specific etiology, they can make important contributions to the diagnosis of these diseases and help determine what further tests are necessary.

The main purpose of the vascular laboratory is to differentiate between fixed arterial obstruction and vasospasm and to determine the location and severity of any identified obstruction (Fig. 20). In most respects, the work-up is analogous to that employed in investigating arterial disease of the lower extremity.

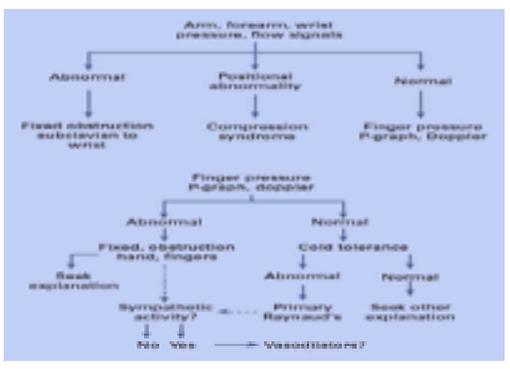


Fig. 20. Approach to the non-invasive diagnosis of upper extremity obstructive and vasospastic disease.

Pressure recordings

Systolic blood pressures are measured in both arms at the brachial, forearm, and wrist levels, using the pneumatic cuff-Doppler technique. At each level, there should be no more than a 20 mmHg difference in the pressures in the two arms. A pressure difference greater than this at the brachial level is diagnostic of an obstructive lesion in the axillary, brachial, or subclavian artery on the side with the lower pressure. A pressure gradient exceeding 20 mmHg between any two segments of the arm implies obstruction in the intervening segment. At any level, a pressure less than 40 mmHg implies severe limb-threatening ischemia.

Finger pressures are very important, especially when symptoms are confined to only one or two fingers or when pressures at all levels of the arm are within normal limits. Finger pressures are measured with pneumatic cuffs placed around the proximal phalanx and a flow sensor (photoplethysmograph) applied to the tip of the finger. A finger/brachial pressure index is calculated by dividing the finger pressure by that in the ipsilateral brachial artery. Indices hover around 1.0 in normal limbs and are almost always above 0.80 (Fig. 21). Similar indices are found in fingers of patients with purely vasospastic disease when their hands are warm. Patients with fixed digital artery obstruction have a wide range of indices extending down to zero. An index less than 0.75 is indicative of obstructive disease.

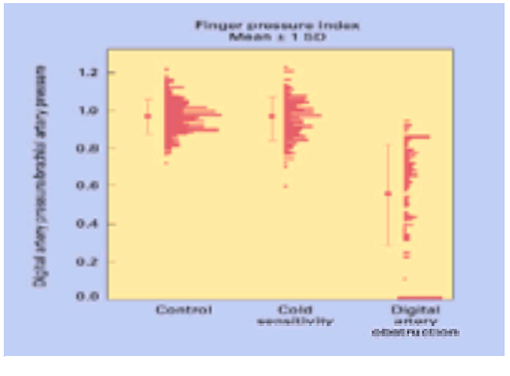


Fig. 21. Finger/brachial indices in normal subjects, patients with cold sensitivity due to vasospastic disease, and patients with digital arterial obstruction. Mean values ± SD are indicated by closed circle and vertical bars.

Blood flow studies

Doppler surveys of the brachial, radial, and ulnar arteries, palmar arch, and digital arteries are useful for detecting and localizing obstructive lesions, which may be widespread or confined to a single finger. Characteristics of normal and abnormal signals are similar to those in the lower extremity. Duplex scanning provides more precise information and can be applied to arteries as small as those in the fingers. The subclavian artery under the clavicle is difficult to examine, but flow signals obtained above and below the clavicle can be used to evaluate the blind area.

Plethysmographic examination

Recordings of digital volume pulsations are easily obtained with a mercury strain gauge or photoplethysmograph applied to the terminal phalanx of the fingers. The tracings closely resemble those obtained from the toes and are interpreted in much the same way (Fig. 22). In addition, a ‘peaked' pulse with an anacrotic notch and a dicrotic notch high on the downslope is often seen in patients with autoimmune diseases or in patients with primary Raynaud's disease during the early stages of cold-induced vasospasm.

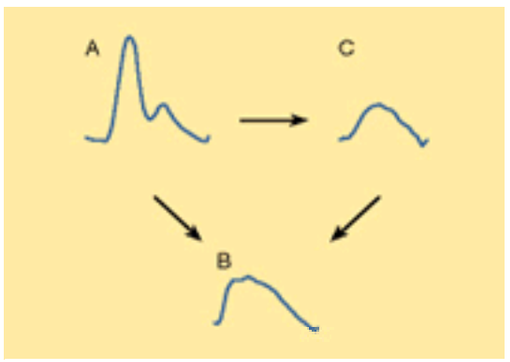


Fig. 22. Plethysmographic pulses from finger tip. A, normal; B, ‘peaked'; C, obstructive. Normal pulses may convert to peaked and peaked to obstructive during cold-induced vasospasm.

Plethysmography is especially useful for identifying obstructive lesions confined to the middle and terminal phalanges, which may not be detected by other methods. In patients with cold-induced vasospasm, the digit pulse is markedly attenuated or disappears entirely during a Raynaud's attack. Patients with ‘primary Raynaud's disease' have normal digit waveforms when their fingers are warm, reflecting the absence of ‘fixed' arterial obstruction. In contrast, patients with Raynaud's phenomenon secondary to arterial obstruction (e.g. scleroderma) have abnormal pulses even when their hands are warm.

Documenting vasospasm

Attempts to provoke a Raynaud's attack in the vascular laboratory may be frustrating. When the history is atypical or when industrial compensation is an issue, objective documentation of cold-induced vasospasm is important. Instruments specially designed for this purpose are commercially available. They incorporate a water-filled finger cuff to cool the digital arteries and a flow-sensor to measure digital arterial pressure. In patients with Raynaud's disease, a trigger point is reached as the finger is cooled to 10 to 20°C that results in a sudden drop in pressure to zero, owing to digital artery spasm. Normal fingers experience only a modest pressure reduction at this temperature. Much the same information can be obtained by measuring finger pressures after immersing the hands in ice water (10°C) for 5 min.

A simpler and more widely available method uses thermistors to measure fingertip temperatures after the hand has been exposed to ice water for 20 s. Normal fingers recover pre-exposure temperatures fairly rapidly, usually within 10 min. The temperature of fingers in patients with vasospasm remains below baseline levels for 20 min or more.

Thoracic outlet compression

A decrease in brachial systolic pressure, Doppler flow signals, or finger pulses during thoracic outlet maneuvers supports the diagnosis of intermittent subclavian artery compression. When compression is complete, pulses and Doppler signals disappear.

Acute peripheral arterial obstruction

Acute arterial obstruction is usually easily recognized based on symptoms (sudden onset of pain) and physical signs (loss of pulses, coldness, paresis). Distinguishing between an embolus and thrombosis of an artery narrowed by a pre-existing atherosclerotic plaque may be difficult, although a history of atrial fibrillation or a recent myocardial infarction would favor the former and a history of claudication would favor the latter. Ankle or arm pressures are seldom necessary to make the diagnosis but may be helpful for documenting the presence of chronic arterial obstruction in the opposite extremity. If present, disease in the opposite limb suggests that the acute occlusion is likely to be thrombotic. In which case, preoperative arteriography is valuable, since extensive arterial reconstruction or preliminary thrombolytic therapy may be indicated rather than simple thrombectomy.

When the occlusion is embolic, a simple Doppler survey often locates the site of obstruction with sufficient accuracy to obviate the need for arteriography. More precise localization can be obtained with the duplex scan.

If Doppler signals are present in the terminal arteries of the involved extremity and if ankle, wrist, finger, or toe pressures exceed 30 mmHg, intervention can be delayed, provided symptoms are not severe. The time gained may be used to optimize the patient's condition or to try thrombolytic therapy. On the other hand, absence of distal arterial signals or a zero or critically low distal pressure implies limb-threatening ischemia, mandating prompt surgical intervention.

Peripheral arterial trauma

Decreased pressure and reduced or absent Doppler flow signals in arteries distal to the site of blunt, penetrating, or iatrogenic trauma suggest arterial injury or vasospasm. However, normal pressures and Doppler signals do not rule out arterial injury in patients with severe hemorrhage, large hematomas, fractures, or penetrating wounds in proximity to a major artery. Since neglecting to treat an arterial injury promptly may have serious consequences, arteriography or surgical exploration is recommended in cases where the index of suspicion is high. When a duplex scanner is available and the clinical situation permits, a good quality scan performed by an experienced operator may suffice to eliminate or confirm arterial injury. Duplex scanning also has the advantage that it can be repeated if sequential or follow-up observations are warranted.

Cerebrovascular disease

Unlike the symptoms and signs of peripheral arterial disease, which are often diagnostic, palpation of carotid pulses is seldom revealing and cervical bruits are unreliable for ruling in or out carotid stenoses or occlusions. Most lesions are asymptomatic. Even total occlusion of the internal carotid artery may occur without symptoms. Localizing symptoms (transient ischemic attacks, amaurosis fugax, and strokes) are usually attributable to emboli originating from atherosclerotic plaques at the carotid bifurcation or, less frequently, from the heart or brachiocephalic vessels. Global ischemia is seldom the cause of symptoms.

Questions to be answered by the vascular laboratory include: Is there disease at the carotid bifurcation and, if so, how severe is the stenosis of the internal carotid artery? Is the internal carotid patent or occluded? What are the surface characteristics and composition of the plaque?

Evaluation methods

Although arteriography is still considered the gold standard for evaluating the cerebral blood supply, it is invasive, subjects the patient to ionizing radiation and contrast allergy, carries a small but not negligible risk of causing a stroke, and is too expensive for routine use—all of which preclude screening of asymptomatic patients and frequent follow-up studies. Moreover, arteriography furnishes little physiologic information and, except for identifying calcification, provides no information regarding plaque composition. Even measurement of stenosis is subject to considerable interobserver variability.

These problems prompted investigators in the late 1960s and throughout the 1970s to seek non-invasive methods to diagnose carotid artery disease. Supraorbital Doppler flow studies, phonoangio-graphy, pulse-delay oculoplethysmography, ocular pneumoplethysmography, static Doppler flow-mapping (‘ultrasonic arteriography'), B-mode scanning, and spectrum analysis of hand-held CW and pulsed-Doppler signals were among the many methods that enjoyed temporary popularity. With the exception of ocular pneumoplethysmography, most are rarely performed today.

Ocular pneumoplethysmography

Criteria for an abnormal study include a difference of 5 mmHg or more between the pressures measured from the two eyes and a ratio of 0.60 or less between the ophthalmic and brachial artery pressures (Fig. 5). With these criteria, accuracies exceeding 90 per cent have been reported for identifying hemodynamically significant stenoses of the internal carotid artery. Because ophthalmic pressure is a function of both the ipsilateral carotid pressure and that of the collateral blood supply, ocular pneumoplethysmography may be used to assess collateral arterial function and the risk of stroke. An ophthalmic artery pressure that drops below 60 mmHg during ipsilateral common carotid pressure suggests a precariously low collateral input.

Some patients are anxious about having suction applied to their eyes and do not tolerate the procedure well. Contraindications include glaucoma, lens implant, and warfarin therapy. Moreover, all physiologic studies, including ocular pneumoplethysmography, are limited in that they are sensitive only to stenoses that exceed 50 per cent diameter reduction. They cannot distinguish between severe stenoses and total occlusion, and furnish no morphologic information. Only 13 per cent of registered vascular technologists in the United States were using ocular pneumoplethysmography in 1994, and the number continues to drop.

Duplex scanning

Duplex scanning has revolutionized carotid diagnosis and has essentially replaced all other non-invasive diagnostic modalities. It is used by 98 per cent of registered vascular technologists and is now widely accepted as the initial diagnostic test in patients with suspected extracranial carotid artery disease. Duplex scanning not only detects atheroma at the carotid bifurcation with a high degree of accuracy (sensitivities and specificities exceed 90 per cent) but also provides reliable information concerning the degree of stenosis, distinguishes between total occlusion and severe stenosis, and identifies minor lesions. The B-mode component provides information regarding plaque morphology and composition.

Technique

Although scanning can be performed accurately with conventional duplex devices, most instruments today incorporate color-flow mapping, which not only facilitates identification of vessels but also alerts the examiner to areas of turbulence and high velocity that suggest the presence of a stenosis (Fig. 23). In practice, the entire cervical carotid artery is scanned longitudinally from the base of the neck to behind the angle of the mandible.

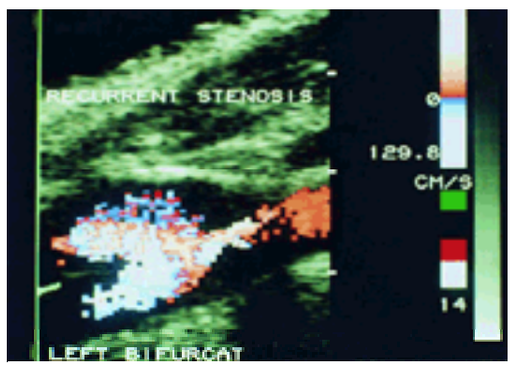


Fig. 23. Color-duplex image of carotid artery bifurcation showing severe recurrent stenosis of distal common carotid and proximal internal and external carotid arteries. Color change from red to white highlights regions of high-velocity flow. The plaque is echolucent (black), suggesting intraplaque hemorrhage.

Spectral analyses of flow patterns in the common, external, and internal carotid arteries are routinely recorded. All regions where a color change from red to white (yellow or green) is noted are also interrogated with the Doppler. The cursor is placed in the center of the flow stream, aligned parallel to the arterial wall (or direction of the flow stream), and the angle of the sound beam is adjusted to approximate 60° (Fig. 24). The external carotid artery is distinguished from the internal by its location, the presence of branches, and by its characteristic ‘high-resistance' flow pattern (which resembles that of peripheral arteries with flow going to zero or reversing in early diastole). In normal internal carotid arteries, ‘low-resistance' flow patterns are observed. Forward flow is present throughout the entire cardiac cycle. The pattern in the common carotid is a cross between those in the external and internal but (normally) more closely resembles that in the internal.

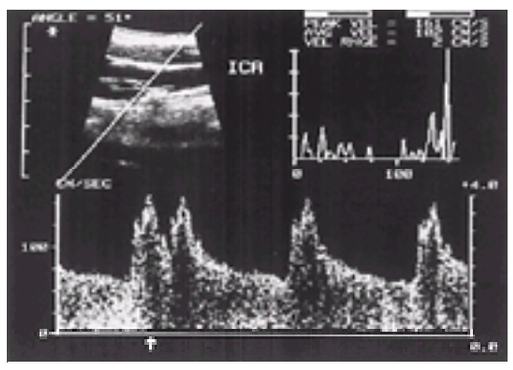


Fig. 24. Spectrum of pulsed-Doppler signal obtained from an internal carotid artery with 50 to 79 per cent diameter stenosis. Peak systolic velocity is 161 cm/s and end-diastolic velocity is 65 cm/s. Plaque is visible on the wall of the internal carotid artery opposite the flow divider. The position of the sample volume and angle of interrogation are indicated. An optimum angle of 60° was not obtainable in this case.

At each site, the spectrum is examined for spectral broadening, and peak systolic and end-diastolic velocities are recorded (Fig. 24). Criteria for predicting the severity of internal carotid artery stenosis based on these parameters vary somewhat from one laboratory to the next, in part related to differences in equipment. Those used in our laboratories are listed in Table 3. Accuracy data are given in Table 4 and Table 5. Discrimination between lower grades of stenosis based on velocity criteria alone is not good. Some help can be obtained from the spectrum or B-mode image. Velocities, however, are much more useful for identifying higher degrees of stenosis and can be used to differentiate between mild stenosis (less than 50 per cent diameter reduction) and moderate stenosis (50 to 79 per cent diameter reduction) and between moderate lesions and severe stenoses of 80 to 99 per cent with acceptable accuracy. Absence of flow in an artery clearly visualized on the B-mode scan is

indicative of total occlusion (Fig. 25).

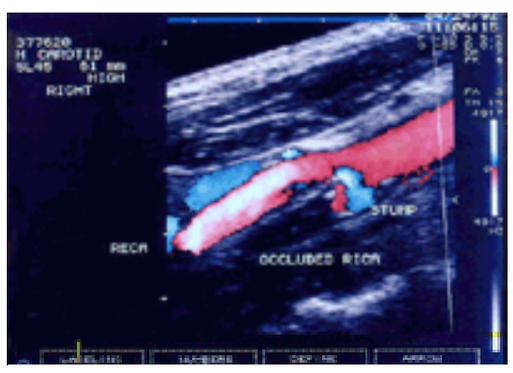


Fig. 25. Color-duplex image of right carotid bifurcation with patent external carotid artery (RECA) and occluded internal carotid artery (RICA). Reversal of flow in the internal carotid ‘stump' is indicated by blue pixels.

Diameter of stenosis (%)	Peak systolic velocity (cm/s)	End-diastolic velocity (cm/s)	Spectral broadening
00–15	< 100	-	-
16–49	> 100, < 130	-	+
50–79	] 130	-	++
80–99	> 130	] 120	+++
100 (Total occlusion)	No flow	No flow	No flow

*Criteria used at Southern Illinois University, Springfield, Illinois, USA, based on ratio of the lumen diameter at the narrowest part of the stenosis compared with the diameter of the lumen of the "normal" internal carotid artery just distal to the stenosis.

Table 3 Duplex velocity criteria for predicting the diameter of the stenosis of internal carotid artery.*

	Arteriography (% stenosis)					
Duplex (%)	0–15	16–49	50–79	80–99	100	Total
0–15	424	11	1	0	0	436
16–49	25	112	6	1	1	145
50–79	4	26	179	9	1	219
80–99	0	0	8	120	8	137
100	0	0	0	0	99	99
Total	453	149	194	130	108	1036

Perfect agreement = 0.91 Kappa = 0.877 ± 0.012
*Data from Peripheral Vascular Laboratory, St. John's Hospital, Springfield, Illinois, USA. Includes all patients with both duplex and arteriography from January 1, 1990 to June 30, 1994.

Table 4 Color-flow duplex scanning compared with arteriography (1036 internal carotid arteries).*

Percentage	Predictive value					
stenosis	Sensitivity %	Specificity %	Positive %	Negative %	Accuracy %	
≥ 0 vs] 15	98	94	95	97	96	
	≥ 49 vs] 50	98	95	93	98	96
≥ 79 vs] 80	95	99	97	98	98	
≥ 99 vs 100	98	100	100	99	99	

*Based on data from Table 4.

Table 5 Accuracy of color-flow duplex scanning for distinguishing between various degrees of stenosis of the internal carotid artery.*

The peak systolic velocity ratio of the internal carotid/common carotid artery has been found by some laboratories to discriminate more reliably between higher degrees of stenosis of the internal carotid artery than do peak or diastolic velocities alone. According to investigators at the University of Oregon, a velocity ratio of 4.0 or higher defines a greater than 70 per cent diameter stenosis with a sensitivity of 91 per cent and a specificity of 87 per cent.

With color off, the B-mode image is examined to evaluate plaque morphology, surface characteristics, and echolucency. Fibrous plaques are echogenic, but soft plaques and intraplaque hemorrhage tend to be more echolucent. Thrombus may be totally echolucent and invisible to B-mode scanning. Calcification produces the brightest echoes and causes acoustic shadows that may interfere with visualization of the lumen and with the recording of velocity data. Using a different transducer angle may get around this problem. ‘Ulcers' are not reliably detected but arteriography is no better.

Clinical application

Indications for carotid scanning include: localizing cerebrovascular symptoms (stroke, transient ischemic attacks, and amaurosis fugax), ill-defined cerebral symptoms compatible with posterior circulation problems (ataxia, dizziness, diplopia, etc.), and asymptomatic cervical bruits. Scans are also indicated in asymptomatic patients who are at a high risk for cerebrovascular disease (patients with coronary artery disease, peripheral vascular disease, hypertension, advanced age, diabetes, smoking, and hypercholesterolemia) and, possibly, in patients scheduled to undergo cardiac or peripheral vascular surgery. If the scan discloses disease exceeding 50 per cent diameter reduction on the side appropriate to well-defined cerebrovascular symptoms, the patient is ordinarily considered to be a candidate for carotid endarterectomy. Similarly, if the scan shows a lesion exceeding 60 to 80 per cent diameter reduction in an asymptomatic patient, carotid endarterectomy may be considered. In our experience it is safe to follow asymptomatic patients with stenoses less than 79 per cent.

Although the usual practice has been to confirm duplex scan findings with arteriography prior to surgery, an increasing number of surgeons are now operating based on the duplex scan alone, provided that the scan is of high quality and the findings coincide with symptoms. The decision to withhold operation can be based safely on the duplex scan in most instances. Patients with moderately severe lesions should be followed periodically to detect advancing disease. Some surgeons feel that it is important to confirm total occlusion to avoid missing a very tight stenosis; but even here a good quality scan may be sufficient to rule out operable disease. Some also advocate confirming the findings of duplex studies with magnetic resonance angiography. If the two studies agree, they are mutually supporting and avoid the need for

arteriography.

Vertebral artery examinations

Vertebral arteries may be examined during the course of a carotid scan. Color is very helpful for locating and identifying these vessels, which lie more deeply in the neck than the carotid bifurcation and are visible only in the intervertebral spaces. Flow patterns resemble those in the carotid arteries. Reversal of flow (or reversal of flow in response to hyperemia of the ipsilateral arm following a period of pneumatic cuff-induced ischemia) is indicative of a subclavian steal. Finding a decreased brachial pressure in the ipsilateral arm is a necessary substantiating finding.

Transcranial Doppler sonography

The direction and velocity of blood flow in the circle of Willis and associated arteries can be studied with low frequency (2 MHz) pulsed-Doppler ultrasound. Recordings from the middle, anterior, and posterior cerebral arteries are made with a hand-held Doppler probe positioned over a transtemporal window, where the bone is relatively thin (Fig. 26). Less consistently, flow in the anterior and posterior communicating arteries can be detected with this approach. The ophthalmic artery and carotid siphon are examined through the orbit, and a suboccipital window (through the foramen magnum) provides access to the vertebral and basilar arteries. The distal extracranial internal carotid artery may be examined through a submandibular window. Vessels are identified by their depth, direction of blood flow, and angle of insonation—a process that requires considerable skill and experience. Increased velocity, damping of the waveform, increased or decreased pulsatility, and reversal of flow are among changes that signify stenosis or collateralization.

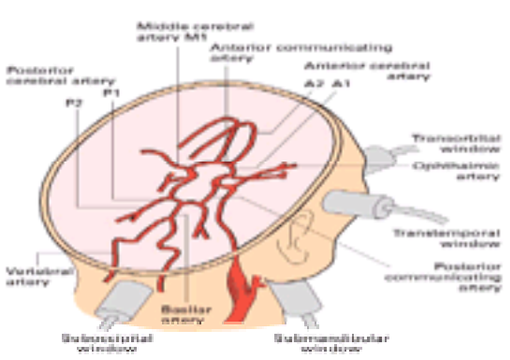


Fig. 26. Arterial components of the circle of Willis and acoustic windows available for transcranial Doppler studies.

Transcranial color-duplex imaging offers significant advantages over conventional systems. Color-coded flow-maps facilitate vessel identification, clarify anatomic relationships, indicate flow direction, and help in the recognition of collateral pathways.

The information derived from these studies can be used to evaluate collateral blood flow in the circle of Willis, to detect intracranial arterial obstructive disease, and to assess the effect of extracranial carotid and vertebrobasilar disease on cerebral perfusion. This information is of some prognostic value and may also be helpful in planning surgical therapy. Transcranial Doppler is especially useful for detecting and monitoring cerebral vasospasm associated with subarachnoid hemorrhage, for detecting arteriovenous malformations, for evaluating hyperperfusion syndromes, and for confirming brain death.

Intraoperatively, continuous flow recordings from the middle cerebral artery may be used to determine the need for a shunt during carotid cross-clamping. Emboli passing through the middle cerebral artery produce characteristic high amplitude ‘chirp-like' signals that are easily recognized and, in the early postoperative period, can alert the surgeon to thrombus or residual defects at the carotid bifurcation.

Transcranial Doppler furnishes more information than ocular pneumoplethysmography and is becoming the preferred method for assessing the intracranial circulation. Although neurosurgeons, neurologists, and clinical scientists have found many applications for transcranial Doppler, its role in the evaluation of patients being considered for carotid endarterectomy is less clearly defined. One drawback is that the studies are more difficult to perform and to interpret than those directed toward the extracranial circulation.

Visceral artery disease

Renal arteries

Although renal artery scanning is one of the more difficult studies performed in the vascular laboratory, as technologic improvements have been made and technologists have become more skilled, these studies are now being performed routinely in many vascular laboratories. Owing to the angle they make with the aorta, their relatively small size, and their depth, renal arteries are difficult to follow longitudinally and only small segments can be visualized at any given position. Accessary renal arteries may not be detected.

Flow in renal arteries (like that in the internal carotid arteries) is positive throughout the cardiac cycle, owing to the low resistance of the renal vascular bed. Absence of flow in a renal artery visualized on B-mode scan indicates total occlusion. Marked spectral broadening and a renal artery/aortic peak systolic velocity ratio greater than 3.5 is indicative of a stenosis exceeding 60 per cent diameter reduction. Using these criteria, investigators report sensitivities exceeding 80 per cent and specificities of 95 per cent or more.

Duplex scanning is being accepted as the initial screening method for investigating the presence of renal artery disease, especially in patients with suspected renal hypertension. It is also useful for follow-up studies and for assessing flow in transplanted kidneys.

Mesenteric arteries

As a rule, the origins of the celiac and superior mesenteric arteries are easily visualized with duplex scanning (Fig. 27). The inferior mesenteric artery is much more difficult to image. The color-flow map helps locate these vessels and a change in color may suggest increased velocities.

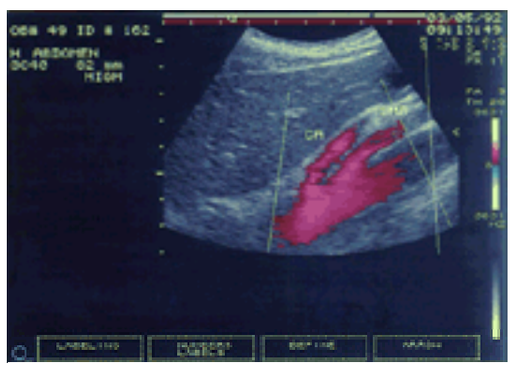


Fig. 27. Color-flow duplex scan of normal upper abdominal aorta showing origins of the celiac axis (CA) and superior mesenteric artery (SMA).

During the fasting state, flow reversal may be observed in the superior mesenteric artery but is not seen in the celiac axis. Severe stenoses are characterized by the

presence of a high-frequency jet with high systolic and diastolic velocities. Peak systolic velocities greater than 200 cm/s in the celiac axis or 275 cm/s in the superior mesenteric artery correlate well with diameter reductions in excess of 70 per cent. Reported accuracies for identifying and ruling out significant disease exceed 80 per cent in the celiac axis and 90 per cent in the superior mesenteric artery.

Aneurysms

B-mode imaging provides a rapid, simple, direct, accurate, safe, and readily available method for detecting and determining the dimensions of abdominal, femoral, popliteal, carotid, and other peripheral aneurysms (Fig. 2). Even aneurysms in small arteries, such as the radial and ulnar, are easily demonstrated. Together with color-flow mapping, B-mode ultrasound readily distinguishes true and false aneurysms from lymph nodes, cysts, hematomas, tumors, and other non-vascular masses. The color map defines the dimensions of the flow stream and, together with the gray-scale image, defines the thrombus load.

Among the advantages that ultrasound has relative to computed tomography (CT) or magnetic resonance imaging (MRI) is that the study is inexpensive and can be repeated as frequently as necessary. Small aneurysms should be followed at 6-month or yearly intervals to detect enlargement that may signify a potential for rupture. In addition, arteries peripheral to the site of an aneurysm should be investigated by methods previously outlined to detect emboli or atherosclerotic plaques that may affect the decision to operate and the design of the surgical procedure.

In many cases, particularly when an operation is not planned, the duplex scan may suffice as the only test. Indeed, it may be sufficient in situations where the anatomy is straightforward, such as it is in many femoral, popliteal, and abdominal aortic aneurysms. When more information is required CT scanning or MRI are unexcelled for evaluating aneurysm morphology.

Color duplex scanning is particularly valuable for the investigation of pulsatile masses in the groin that complicate arterial puncture for cardiac catheterization or arteriography. Simple hematomas are readily differentiated from acute false aneurysms. In the latter, colored pixels ‘swirling' within the mass verify the presence of flow, and a strand of color highlights the communication with the underlying artery (Fig. 28). The majority of these early aneurysms may be treated in the vascular laboratory by applying pressure over the pulsatile mass until the active communication disappears and flow is no longer evident in the aneurysm sac.

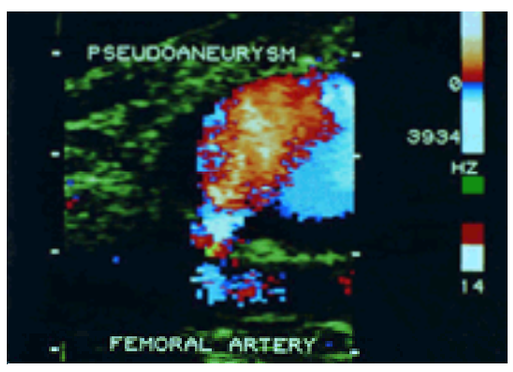


Fig. 28. Color-flow duplex scan of a femoral pseudoaneurysm, showing flow swirling within the aneurysm, clot (black areas), and communication with the femoral artery.

Arteriovenous fistulas

Acquired (traumatic) arteriovenous fistulas are easily recognized by the increased velocity of flow in the artery feeding the fistula and in the vein into which the fistula empties. Flow in the proximal artery is continuous throughout diastole (no reversed flow component) and is turbulent at the fistula site. With color duplex, the communication itself may be visualized. Small iatrogenic fistulas may respond to compression therapy. Others can be followed. Many will thrombose. Larger fistulas require surgical intervention. When the fistula represents failure to ligate a venous tributary during *in situ* bypass grafting, duplex scanning usually locates the communication with sufficient accuracy to obviate the need for arteriography. Most require observation only, but some are locally painful and require surgical correction. They rarely jeopardize graft survival.

Graft surveillance and follow-up studies

Detection of stenoses in infrainguinal bypass grafts and their early correction has been shown to prolong graft survival. With duplex scanning, lesions developing in grafts can be identified before symptoms occur, before pulses disappear, or before the ankle pressure drops. At this stage, a minor surgical procedure will often suffice to correct the problem.

Grafts are scanned longitudinally with cross-sectional images being obtained at selected points. A change in color of the flow map from red to white (green or yellow) identifies the site of a possible stenosis (Fig. 4). Direct Doppler interrogation is the next step. The peak systolic velocity at the site of color change is divided by the peak systolic velocity in a ‘normal' segment of graft a few centimeters above or below the suspected lesion to obtain a ‘velocity ratio'. A velocity ratio of 2.0 or more suggests a greater than 50 per cent diameter stenosis, and a velocity ratio of 3.0 or more suggests a diameter reduction in excess of 70 per cent. All 70 per cent lesions should be repaired, since the risk of thrombosis is high. Lower grades of stenosis may be followed.

Initial studies should be made intraoperatively, and any defects discovered at that time should be corrected. This approach will reduce significantly the number of early thromboses. Postoperatively, scans are obtained before discharge from the hospital, at 3-month intervals for the first year, and then yearly. Most stenoses appear during the first year, but some occur after 2 years. Once a lesion is discovered, arteriography may or may not be necessary prior to surgical correction.

Many vascular laboratories perform follow-up scans of carotid endarterectomies at intervals during the first year and then yearly thereafter. Because the incidence of significant restenosis is so low, the cost-effectiveness of this approach is open to question. Frequently, however, scanning discloses a stenosis of the contralateral carotid artery that initially is not severe enough to merit surgical intervention but that may progress to severe stenosis in the future. Close follow-up may be justified in these cases.

Further reading

Bandyk DF. Essentials of graft surveillance. *Seminars in Vascular Surgery* 1993; **6**: 92–102. [Comprehensive coverage of techniques for graft surveillance.]

Bernstein EF, ed. *Vascular diagnosis*, 4th edn. CV Mosby, St Louis, 1993. [Encyclopedic volume with multiple chapters discussing all aspects of noninvasive diagnosis.]

Bogdahn U, Becker G, Winkler J, Greiner K, Perez J, Meurers B. Transcranial color-coded real time sonography in adults. *Stroke* 1990; **21**: 1680–8. [Discusses advantages of color-flow mapping for transcranial Doppler studies.]

Byrne P, Provan JL, Ameli FM, Jones DP. The use of transcutaneous oxygen tension measurements in the diagnosis of peripheral vascular insufficiency. *Annals of Surgery* 1984; **200**: 159–65. [Provides data concerning oxygen tension values, at rest and after exercise, relative to the severity of circulatory impairment.]

Comerota AJ, Katz ML, Hosking JD, Hashemi HA, Kerr RP, Carter AP. Is transcranial Doppler a worthwhile addition to screening tests for cerebrovascular disease? *Journal of Vascular Surgery* 1995; **21**: 90–7. [Critical evaluation of the limited role of transcranial Doppler in the evaluation of extracranial carotid disease.]

Dawson DL, Zierler RE, Strandness DE Jr, Clowes AW, Kohler TR. The role of duplex scanning and arteriography before carotid endarterectomy: a prospective study. *Journal of Vascular Surgery* 1993; **18**: 673–83. [Presents data supporting the accuracy of duplex scanning in the selection of patients for carotid endarterectomy. Scanning provides sufficient information in 93 per cent of patients.]

El-Barghouty N, Nicolaides AN, Bahal V, Geroulakos G, Androulakis A. The identification of the high risk carotid plaque. *European Journal of Vascular and Endovascular Surgery* 1996; **11**: 470–8. [Discusses the relationship of carotid plaque echogenicity to cerebral infarction.]

Faught WE *et al.* Color-flow duplex scanning of carotid arteries: new velocity criteria based on receiver operator characteristic analysis for threshold stenoses used in the symptomatic and asymptomatic carotid trials. *Journal of Vascular Surgery* 1994; **19**: 818–27. [Optimum peak systolic and diastolic velocity levels for predicting 70 per cent diameter stenosis of the internal carotid artery are presented.]

Fillinger MF *et al.* Carotid duplex criteria for a 60 per cent or greater angiographic stenosis: variation according to equipment. *Journal of Vascular Surgery* 1996; **24**: 856–64. [Velocity thresholds for predicting carotid artery stenosis vary with the duplex scanner being used.]

Gale SS *et al.* Lower extremity arterial evaluation: are segmental arterial blood pressures worthwhile? *Journal of Vascular Surgery* 1998; **27**: 831–9. [Authors conclude that pressure measurements made above the ankle level are of limited utility.]

Harward TRS, Smith S, Seeger JM. Detection of celiac axis and superior mesenteric artery occlusive disease with the use of abdominal duplex scanning. *Journal of Vascular Surgery* 1993; **17**: 738–45. [Velocity criteria that provide 96 per cent accuracy for identifying greater than 50 per cent stenosis of the celiac axis and superior mesenteric arteries are presented.]

Idu MM, Blankenstein JD, de Gier P, Truyen E, Buth J. Impact of a color-flow duplex surveillance program on infrainguinal vein graft patency: a five-year experience. *Journal of Vascular Surgery* 1993; **17**: 42–53. [Overall graft patency rates were better with duplex scanning surveillance than they were with clinical assessment alone.]

Karanfilian RG, Lynch TG, Zirul VT, Padberg FT, Jamil Z, Hobson RW II. The value of laser Doppler velocimetry and transcutaneous oxygen tension determination in predicting healing of ischemic forefoot ulcerations and amputations in diabetic and nondiabetic patients. *Journal of Vascular Surgery* 1986; **4**: 511–16. [Skin oxygen tension levels greater than 10 mmHg and laser Doppler flux greater than 40 mV predicted healing.]

Kempczinski RE, Yao JST, eds. *Practical noninvasive vascular diagnosis*, 2nd edn. Year Book Medical Publications, Chicago, 1987. [This clearly written book describes practical methods for diagnosing vascular diseases.]

Kohler TR *et al.* Can duplex scanning replace arteriography for lower extremity arterial disease? *Annals of Vascular Surgery* 1990; **4**: 280–7. [The authors conclude that therapeutic decisions made with duplex scanning alone are very similar to those made on the basis of arteriography.]

Kohler TR, Nance DR, Cramer MM, Vandenburghe N, Strandness DE Jr. Duplex scanning for diagnosis of aortoiliac and femoropopliteal disease: a prospective study. *Circulation* 1987; **76**: 1074–80. [Duplex scanning was 82 per cent sensitive and 92 per cent specific for detecting stenoses greater than 50 per cent in arteries at knee level or above.]

Langsfeld M, Gray-Weale AC, Lusby RJ. The role of plaque morphology and diameter reduction in the development of new symptoms in asymptomatic carotid arteries. *Journal of Vascular Surgery* 1989; **9**: 548–57. [Patients with heterogenous echolucent plaques were found to be at risk for developing cerebrovascular symptoms.]

Levi CR *et al.* Transcranial Doppler microembolus detection in the identification of patients at high risk of perioperative stroke. *European Journal of Vascular and Endovascular Surgery* 1997; **14**: 170–6. [More than 50 microembolic signals detected per hour in the middle cerebral artery of patients in the early postoperative period predicted ipsilateral focal cerebral ischemia.]

Londrey GL, Spadone DP, Hodgson KJ, Ramsey DE, Barkmeier LD, Sumner DS. Does color-flow imaging improve the accuracy of duplex carotid evaluation? *Journal of Vascular Surgery* 1991; **13**: 659–63. [Authors demonstrate a trend toward improved accuracy when color is added to duplex scanning.]

Lundell AJB, Lindblad B, Bergqvist D, Hansen F. Femoropopliteal-crural graft patency is improved by an intensive surveillance program: a prospective randomized study. *Journal of Vascular Surgery* 1995; **21**: 26–34. [Secondary graft patency at 3 years was 82 per cent in patients undergoing duplex scan surveillance compared to a secondary patency rate of only 56 per cent in those patients with clinical follow-up alone.]

Martin KD, Patterson RB, Fowl RJ, Kempczinski RF. Is the continued use of ocular pneumoplethysmography necessary for the diagnosis of cerebrovascular disease *Journal of Vascular Surgery* 1990; **11**: 235–43. [Ocular pneumoplethysmography was less sensitive than duplex scanning and the combination was less specific than duplex scanning alone. The authors no longer recommend ocular pneumoplethysmography for cerebrovascular evaluations.]

Mattos MA, van Bemmelen PS, Hodgson KJ, Ramsey DE, Barkmeier LD, Sumner DS. Does correction of stenoses identified with color duplex scanning improve infrainguinal graft patency. *Journal of Vascular Surgery* 1993; **17**: 54–66. [Two years after operation, the patency rates of stenotic grafts that had not been revised was only 57 per cent compared to an 88 per cent patency rate of stenotic grafts that had undergone revision.]

Moneta GL *et al.* Accuracy of lower extremity arterial duplex mapping. *Journal of Vascular Surgery* 1992; **15**: 275–84. [Accuracy for detecting 50 per cent stenoses of arteries above the knee and for identifying total occlusion of the tibial arteries with duplex scanning was good.]

Moneta GL, Lee RW, Yeager RA, Taylor LM Jr, Porter JM. Mesenteric duplex scanning: a blinded prospective study. *Journal of Vascular Surgery* 1993; **17**: 79–86. [Peak systolic velocities greater than 275 cm/s in the superior mesenteric arteries and greater than 200 cm/s in the celiac axis accurately predicted stenoses in excess of 70 per cent.]

Moneta GL *et al.* Correlation of North American Symptomatic Carotid Endarterectomy Trial (NASCET) angiographic definition of 70 per cent to 99 per cent internal carotid artery stenosis with duplex imaging. *Journal of Vascular Surgery* 1993; **17**: 152–9. [The authors conclude that an internal carotid artery/common carotid artery peak systolic velocity ratio greater than 4.0 predicts 70 per cent stenoses of the internal carotid artery with a sensitivity of 91 per cent and specificity of 87 per cent.]

Nicholls SC, Kohler TR, Martin RL, Strandness DE Jr. Use of hemodynamic parameters in the diagnosis of mesenteric insufficiency. *Journal of Vascular Surgery* 1986; **3**: 507–10. [Velocity waveforms in the celiac axis and superior mesenteric arteries before and after meals were related to the presence of significant stenoses.]

Orchard TJ, Strandness DE Jr. Assessment of peripheral vascular disease in diabetes. *Circulation* 1993; **88**: 819–28. [This paper reviews various noninvasive methods for detecting and evaluating arterial disease in diabetics and presents a consensus on recommended diagnostic approaches.]

Ouriel K, McDonnell AE, Metz CE, Zarins CK. A critical evaluation of stress testing in the diagnosis of peripheral vascular disease. *Surgery* 1982; **91**: 686–93. [The authors conclude that routine stress testing (treadmill exercise) is not cost effective and should be reserved for evaluating symptomatic patients with normal ankle pressure indices.]

Ramsey DE, Manke DA, Sumner DS. Toe blood pressure—a valuable adjunct to ankle pressure measurement for assessing peripheral arterial disease. *Journal of Cardiovascular Surgery* 1983; **24**: 43–8. [Healing of foot lesions is unlikely when toe pressures are less than 30 mmHg, irrespective of the ankle pressure.]

Rutherford RB, Lowenstein DH, Klein MF. Combining segmental systolic pressures and plethysmography to diagnose arterial occlusive disease of the legs. *American Journal of Surgery* 1979; **138**: 211–18. [This paper presents criteria for diagnosing arterial disease with segmental pressures and plethysmography. The tests were most accurate for identifying normal arteries and isolated arterial occlusions and were less accurate when several levels of obstruction are present.]

Sensier Y, Fishwick G, Owen R, Pemberton M, Bell PA, London DJ. A comparison between colour duplex ultrasonography and arteriography for imaging infrapopliteal arterial lesions. *European Journal of Vascular and Endovascular Surgery* 1998; **15**: 44–50. [In this study, ultrasound interpretations correlated with arteriographic findings with a kappa value of 0.61. Ultrasound was least accurate in the peroneal arteries and in the tibioperoneal trunk.]

Sumner DS. Evaluation of acute and chronic ischemia of the upper extremity. In: Rutherford RB, ed., *Vascular surgery*, 5th edn, pp. 1122–39. WB Saunders, Philadelphia, 2000. [Method for assessing arterial diseases of the upper extremity are covered extensively in this chapter, together with a discussion of the pertinent pathophysiology.]

Sumner DS, Thiele BL. The vascular laboratory. In: Rutherford RB, ed., *Vascular surgery*, 4th edn, pp. 45–64. WB Saunders, Philadelphia, 1995. [Instruments and methods used in the noninvasive assessment of arterial disease are reviewed and the role of the vascular laboratory is discussed.]

Taylor DC *et al.* Duplex ultrasound scanning in the diagnosis of renal artery stenosis: a prospective evaluation. *Journal of Vascular Surgery* 1988; **7**: 363–9. [Sixty per cent stenoses of the renal arteries were detected with a sensitivity of 84 per cent and specific of 97 per cent when the renal artery/aorta peak systolic velocity ratios exceeded 3.5. The authors conclude that duplex scanning can be used to select patients with hypertension or renal dysfunction for arteriography.]

Zierler RE, Sumner DS. Physiologic assessment of peripheral arterial occlusive disease. In: Rutherford RB, ed., *Vascular surgery*, 5th edn, pp. 140–65. WB Saunders, Philadelphia, 2000. [This chapter discusses diagnostic approaches used in the physiologic assessment of peripheral arterial disease.]

17.5.1 Aortoiliac occlusive disease

Joseph S. Giglia and David C. Brewster

Introduction
Clinical presentation
Acute
Chronic
Diagnosis
History and physical examination
Non-invasive vascular studies
Contrast angiography
Magnetic resonance angiography
Computed axial tomographic angiography
Ultrasonography
Treatment
Preintervention evaluation
Initial treatment
Direct aortoiliac reconstruction
Alternatives to direct aortoiliac reconstruction
Endovascular techniques
Summary
Further reading

Introduction

Chronic atherosclerosis of the infrarenal aorta and iliac arteries is a common cause of symptomatic occlusive disease of the lower extremities. Aortoiliac occlusive disease is also frequently associated with infrainguinal occlusive disease. Proper management of patientswith aortoiliac occlusive disease requires a through understanding of the various clinical presentations, the typical patterns of associated infrainguinal disease, the incidence of coexistent cardiopulmonary disease, and the variety of techniques available for therapeutic intervention. Proper patient selection and adherence to well-accepted and standardized indications for intervention will usually result in a favorable clinical result with low risk.

Clinical presentation

The clinical manifestations of aortoiliac occlusive disease can be roughly divided into acute and chronic. Most of this discussion will focus upon chronic disease and its management, but several acute presentations merit consideration.

Acute

Acute aortic occlusion

This can occur by thrombosis of a previously diseased and stenotic aorta or by acute thrombosis of a small aortic aneurysm. Emboliza-tion to the aortic bifurcation from a cardiac source (the so-called saddle embolus) may occasionally occur. Rarely, it can follow direct trauma to a diseased aorta. However, patients more commonly present with a chronic infrarenal aortic occlusion that has only recently become symptomatic. For example, ischemia of the lower extremities can result from progression of disease that has occluded a major collateral, or renal function may deteriorate as the result of ostial lesions or repeated microembolization. Obviously, the limbs of a previously placed aortobifemoral graft can thrombose, leading to an acute presentation of aortoiliac occlusive disease. Graft thrombectomy with correction of distal anastomotic stenoses often corrects the problem in this instance.

Atheroembolization

The aortoiliac segment can be a source of emboli that arise from an unstable atherosclerotic plaque or a small aortic aneurysm. The phenomenon of atheroembolism (blue toe syndrome) is a well-recognized indication for aortoiliac arterial reconstruction. Patients typically present with pain. Often there is a history of invasive procedures involving the aortoiliac segment (e.g. cardiac catheteriza-tion). A clinical history of arterial insufficiency in the lower extremities is often lacking, since the source of the atheroembolism may be a hemodynamically insignificant, ulcerated or unstable plaque. Nevertheless, if the presentation is consistent with atheroembolism an aortogram is indicated. If it reveals irregular, shaggy, or ulcerated atheromatous changes in the aortoiliac system, aortobifemoral bypass and exclusion of the native aortoiliac system may be indicated to avoid further distal embolic events, since repeated episodes of microembolization can result in extensive tissue loss.

Chronic

The typical presentation of aortoiliac occlusive disease is intermittent claudication of the lower extremities. Claudication can affect gluteal, thigh, and calf muscle groups. Often the symptoms are more disabling than those associated with isolated femoropopliteal disease, owing to the greater muscle mass affected in more proximal occlusive disease. Some patients with aortoiliac occlusive disease, particularly those with disease affecting both inflow and outflow vessels, will present with isolated calf claudication. In addition, men may present with Leriche syndrome: originally described by Leriche in 1923 and later in the English literature by Leriche and Moral in 1948, this classic syndrome comprises diminished femoral pulses, claudication in the lower extremities, and impotence.

Although the symptoms experienced by an individual patient are determined by the distribution and severity of the occlusive process, several distinct patterns are recognized (Fig. 1). When disease is confined to the aortic bifurcation and common iliac arteries (type I), limb-threatening ischemia is rare and the symptoms are limited to claudication. Numerous collateral pathways are often found in type I disease, and these may result in only mild to moderate claudication, even with a completely occluded aortoiliac segment. Collateral vessels arising from intercostal and lumbar arteries anastomose with branches of the iliolumbar, gluteal, deep circumflex iliac, and epigastric arteries. Visceral collaterals involving the left colic branch of the superior mesenteric artery may reconstitute flow in the hypogastric artery via the meandering mesenteric artery, the marginal artery of Drummond, and the hemorrhoidal plexus. Complete occlusion of the external iliac arteries may be compensated for by collateral flow through the gluteal branches of the hypogastric artery to the circumflex femoral branch of the profunda femoral artery.

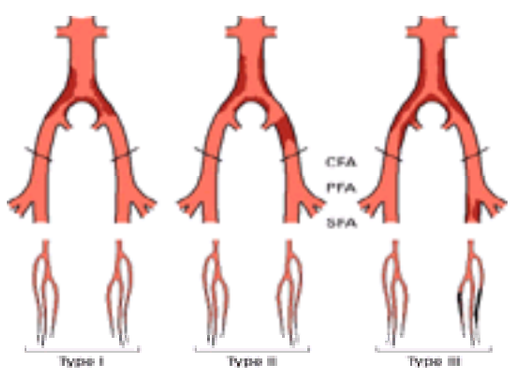


Fig. 1. Distribution of disease: type I, disease localized to the aortic bifurcation; type II, disease also involves external iliac arteries; type III, multilevel disease with infrainguinal arterial disease.

Patients with type I disease who have symptoms severe enough to warrant surgical revascularization are uncommon (5–10 per cent). It is more common in younger

patients, in whom there is a lower incidence of associated coronary disease, hypertension, and diabetes, but often abnormal lipid profiles (e.g. Frederickson type IV hyer-lipidemia). Type I disease is unusual in that nearly 50 per cent of the patients are women.

The majority of patients with symptomatic aortoiliac occlusive disease will have a more widespread condition. Approximately 25 per cent have extension of the disease process into the external iliac arteries (type II); 65 per cent have disease involving both the aortoiliac segments and the infrainguinal arteries (type III). Patients with type III disease present with severe, incapacitating claudication and may also develop limb-threatening ischemia with rest pain and/or ischemic ulcers and gangrene. The indication for revascularization in these patients is often limb salvage rather than claudication alone.

Diagnosis

History and physical examination

The diagnosis of symptomatic aortoiliac occlusive disease is usually established accurately from a complete history and physical examination. Claudication in the proximal muscle groups of the lower extremity, pain after walking a predictable distance, relief of pain with only brief rest, and impotence constitute the classic description. Some patients will present with claudication limited to the calf despite the presence of hemodynamically significant inflow disease; this occurs most often in patients with multilevel (type III) disease, and it is important that they be recognized since inflow must be corrected to improve the symptoms.

The symptoms must be differentiated from those of non-vascular causes of pain in the lower extremities, such as radicular pain cause by nerve-root irritation from spinal stenosis (pseudoclaudication) or herniation of an intervertebral disc. Patients with these conditions will often describe pain induced by standing as well as walking, and this will help distinguish them from claudicants. Patients with pseudoclaudication often need to sit or lie down in order to relieve the pain as opposed to simply stopping walking. A careful history may reveal the sciatic distribution of the pain, suggesting its true cause.

On physical examination there will often be auscultable bruits over the lower abdomen and the femoral regions. Femoral and distal pulses are usually diminished or absent and, in patients with type III disease, dependent rubor, elevation pallor, and trophic skin changes may be seen in the feet and lower legs. A complete physical examination should include a Doppler pulse and pressure examination, which can be performed quickly and inexpensively in the office or clinic with a hand-held device.

Non-invasive vascular studies

These are essential in the diagnosis and management of aortoiliac occlusive disease. They should be performed before and after exercise, as a clinically significant lesion may not register hemodynamically at rest. This phenomenon is the physiologic correlate of Ohm's law: the stenosis of the aortoiliac segment may not be hemodynamically significant at the blood-flow rates seen in the resting patient, but with exercise the rate of flow across the stenotic segment is increased, causing a greater drop in pressure across the stenosis. Pedal pulses that were present at rest may therefore vanish after treadmill exercise, confirming the vascular background to the symptoms. It is, therefore, important to include some form of postexercise measurement of ankle pressure in the evaluation of these patients.

Contrast angiography

Contrast arteriography remains the definitive test in the evaluation of aortoiliac occlusive disease. However, since it is expensive and is associated with the risk of renal dysfunction, vessel injury, and embolization, it should only be undertaken after noninvasive testing suggests the diagnosis and intervention is contemplated. A complete study should include both biplanar aortography and oblique views of the pelvis and groin. It is critical to visualize the origin of the profunda femoris as this vessel often determines long-term graft patency. The lower extremities should be evaluated if that anatomic information is not already known, because of the high incidence of multilevel disease. Pressure measurements proximal and distal to a stenosis help to define its functional significance and can serve as a baseline for subsequent measurements (Fig. 2). A pressure gradient greater than 5 mmHg at rest or a 15 per cent decrease after vasodilation represent a functionally significant lesion.

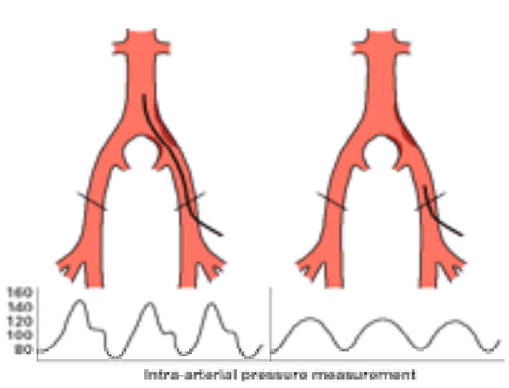


Fig. 2. Intra-arterial pressure measurement: a drop in systolic pressure of more than 5 mmHg across the iliac lesion or a fall in femoral pressure of more than 15 per cent with hyperemic studies identifies a hemodynamically significant iliac stenosis.

Magnetic resonance angiography

Tremendous improvements have been made recently in arterial imaging using magnetic resonance technology. The addition of gadolinium, a non-toxic contrast agent, has further enhanced the resolution of the studies. Magnetic resonance angiographic images are now of sufficient quality in some institutions to be used to plan treatment of aortoiliac occlusive disease in very selected cases. However, slow or turbulent flow leads to signal drop-out, which severely limits the utility of this technique.

Computed axial tomographic angiography

Computed tomographic angiography eliminates the risks of arterial access, but not the possible consequences of a large volume of intravenous contrast. While it is often the sole preoperative study obtained for aneurysm, it has limited utility in planning treatment of aortoiliac occlusive disease. Modern interactive software programs that allow easy, rapid, three-dimensional reconstruction of the axial data may change this in the future.

Ultrasonography

Advances in ultrasound technology and techniques have allowed visualization of the aorta and iliac arteries. Duplex examination of the aortoiliac segment is now routinely performed, but fasting is required to minimize bowel gas artefact. Currently, this modality has a greater role in follow-up than preoperative planning.

Treatment

There is currently a wide range of therapeutic options for aortoiliac occlusive disease, but certain principles should remain paramount. Invasive evaluation should only be undertaken if a therapeutic procedure is contemplated. Also, the overall management of the patient should be directed by a specialist who both understands the natural history of the disease process and is fully capable of treating the potential complications of intervention.

It is generally agreed that limb-threatening ischemia, clinically defined by the presence of rest pain, ischemic ulceration, or frank gangrene, will usually require major amputation; these signs are, therefore, clear-cut indications for arterial reconstruction. In this group of patients there are few contraindications to surgery, since revascularization by some means can usually be accomplished with morbidity and mortality equivalent to or lower than those associated with major amputations. Age is rarely, if ever, a contraindication. If direct aortoiliac reconstruction is deemed too great a risk, alternative techniques for revascularization of the lower extremities, such as extra-anatomic bypass, percutaneous transluminal angioplasty, or a combination of such procedures, may be used.

The need for surgical intervention must be dictated by the circumstances of the individual patient. Incapacitating claudication that prevents them from earning a living or that has negative impact on their desired lifestyle is generally considered an indication, provided that they are not at high risk for surgical complications, do not have a limited life expectancy because of associated medical problems, and have a generally favorable distribution of disease for correction. Patients with stable claudication

will often experience significant improvement in their symptoms following conservative measures such as abstinence from smoking, daily exercise, and weight reduction, if appropriate. Surgical intervention should only be considered if these conservative treatments fail to improve the claudication.

Preintervention evaluation

Preoperative evaluation of the patient with aortoiliac occlusive disease should routinely include assessment of their cardiac, pulmonary, and renal function. Peripheral vascular occlusive disease is a strong marker of significant coronary arterial disease; in fact, at least 40 per cent of patients have significant coronary disease. Myocardial infarction is the cause of more than 50 per cent of the perioperative deaths in patients undergoing peripheral vascular surgery; the detection and management of coronary disease is therefore important. Traditional clinical assessment of cardiac risk may be difficult in the patient with peripheral vascular disease who, due to claudication, has a sedentary lifestyle. Thus, the absence of symptoms of cardiac disease cannot safely be assumed to imply the absence of severe coronary disease. Even coronary angiography may not allow anatomic findings to be related to the perioperative risk of a cardiac event.

Functional cardiac studies

Tests aimed at identifying the subset of patients truly at high risk of suffering perioperative cardiac ischemic events due to coronary arterial disease have been developed. Exercise stress tests are effective but have a limited role in this group of patients with claudication, as already mentioned. Alternatively, echocardiography during dobutamine-induced stress can be used. Nuclear myocardial scintigraphic tests that evaluate the relative distribution of myocardial blood flow before and after a vasodilator (e.g. dipyridamole or adenosine) are sensitive in detecting areas at risk for perioperative ischemia. Collectively these studies help to identify those patients who might require myocardial revascularization before vascular surgery or in whom an alternative to direct aortoiliac reconstruction might be advisable.

Recently there has been a trend to play down the role of preoperative cardiac testing. The cost-effectiveness of these studies now that cardiac morbidity and mortality from aortic surgery have decreased has been questioned. Improved perioperative management, including the liberal use of epidural anesthesia, may have contributed to the observed decrease in morbidity and mortality.

Pulmonary status

The high incidence of cigarette smoking and chronic obstructive airway disease in atherosclerotic persons places them at increased risk for postoperative pulmonary complications. Smoking, even in patients with no detectable lung disease, is associated with increased postoperative atelectasis and pneumonia, abstinence from smoking for 2 weeks before surgery is critical. In patients with clinical evidence of chronic obstructive lung disease, pulmonary function studies are often useful in directing and assessing preoperative preparation, which often includes chest physiotherapy, bronchodilators, and antibiotics. A FEV_1 of less than 800 cc and arterial blood gas tests revealing evidence of CO_2 retention indicate patients at high risk of pulmonary complications. Combined epidural and general anesthesia, with immediate postoperative extubation and postoperative pain management using the epidural catheter, may avoid such complications.

Initial treatment

Initial treatment of stable aortoiliac occlusive disease should include a rigorous exercise program and risk-factor modification including weight reduction, smoking cessation, and treatment of hypercholesterolemia. If these measures fail to lessen the symptoms, then intervention is warranted.

Direct aortoiliac reconstruction

Aortoiliac endarterectomy

Aortoiliac endarterectomy, although frequently done early in the era of aortic reconstruction, is rarely used now. Although a few centers continue to perform sufficient aortoiliac endarterectomies for trainees to feel comfortable with the technique, most vascular surgeons trained within the past several decades will have little or no training and experience with this method. The principal potential benefit of endarterectomy is avoidance of the use of prosthetic grafts, with their possible complications of dilation, infection, anastomotic aneurysm, or other degenerative problems. Fortunately, these are relatively unusual problems with modern techniques and modern vascular grafts.

Endarterectomy is also advocated by some as more likely to improve sexual potency in male patients by more effectively improving hypogastric arterial blood flow. However, this has not been demonstrated in any study, and the more extensive dissection in the region of the aortic bifurcation required in endarterectomy seems likely to result in a higher incidence of neurogenic problems and ejaculatory disturbances.

Aortoiliac endarterectomy may be used for localized disease confined to the distal aorta, aortic bifurcation, and common iliac arteries. In properly selected patients the long-term patency is excellent and equivalent to that of graft procedures. However, more extensive endarterectomy extended into the external iliac arteries or beyond does not have the same durability and patency as bypass grafting. Because the localized aortoiliac occlusive disease most amenable to endarterectomy is encountered in only 5 to 10 per cent of patients requiring aortic reconstruction, the vast majority are not suitable for endarterectomy and will be better treated by grafting.

Currently, endarterectomy should only be considered in young patients with an extended life expectancy or in those with an increased risk of infection. The atherosclerotic plaque should terminate at the common iliac bifurcation, allowing a satisfactory endpoint to the endarterectomy to be achieved without extending the arteriotomy more than 1 to 2 cm in the external iliac artery. The surgeon must verify that a secure endpoint has been achieved, either with or without the use of tacking sutures.

While aortoiliac endarterectomy may give good results in appropriately selected patients, it is important to recognize that many patients with fairly localized disease who were once potential candidates may now be considered for endoluminal treatment with angioplasty, stenting, or endoluminal grafting. All such therapies have further reduced the application of endarterectomy for the treatment of aortoiliac occlusive disease.

Aortobifemoral bypass graft

An aortobifemoral bypass graft continues to be the standard treatment for aortoiliac occlusive disease. It is provenly durable and effective, and should be the ‘gold standard' against which other methods are judged. Refinements in operative technique, improvements in prosthetic graft and suture materials, and in particular, striking advances in preoperative evaluation, intraoperative anesthetic management and postoperative supportive care have all contributed to the steadily improving outcome and generally excellent results attainable in contemporary practice.

Proximal anastomosis

Most vascular surgeons prefer an end-to-end to an end-to-side proximal anastomosis for several reasons (Fig. 3). Theoretically, the end-to-end anastomosis is hemodynamically more sound: there is less turbulence, better flow characteristics, and no chance of competitive flow in the native iliac arteries. The end-to-end technique also carries a lower risk of intraoperative distal embolic events, since flow is not restored through the distal native aorta, where clamping may have loosened atherosclerotic debris or thrombus. It also provides better visualization of the proximal aortic lumen, thus facilitating an aortic thromboendarterectomy if required. Finally, the end-to-end anastomosis allows a short segment of the native aorta to be resected so that the prosthetic graft can lie in the aortic bed, where tissue coverage can more adequately be obtained, thus potentially reducing the risk of a late aortoenteric fistula.

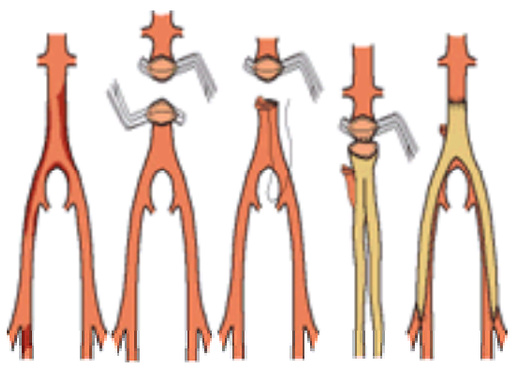


Fig. 3. Aortobifemoral bypass graft: end-to-end proximal aortic anastomosis.

There are, however, specific situations in which the end-to-side configuration is indicated (Fig. 4); primarily these involve patients in whom the distribution of atherosclerotic occlusive lesions would prevent adequate retrograde perfusion of the hypogastric arteries. An end-to-end anastomosis in this setting potentially increases the risk of impotence, ischemic colitis, buttock ischemia, persistent hip claudication, and even paraplegia secondary to ischemia of the spinal cord. Other indications for an end-to-side anastomosis include a patent inferior mesenteric artery with occlusive disease of the superior mesenteric and celiac arteries, an aberrant or accessory renal artery, or a horseshoe kidney. Use of the end-to-side anastomosis in such circumstances allows continued native perfusion of the specific vessels in question while the graft bypasses the aortoiliac occlusive disease. The technical alternative to end-to-side anastomosis in these cases would be reimplantation of, or bypass graft to, the inferior mesenteric, hypogastric, or renal arteries.

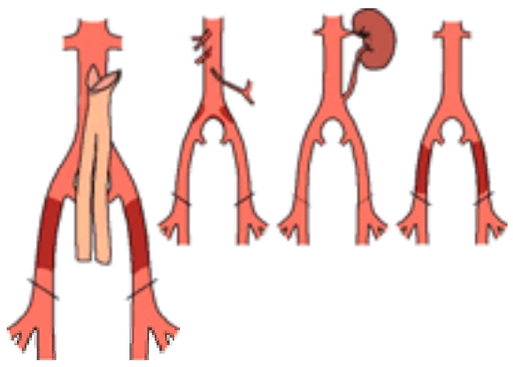


Fig. 4. Indications for end-to-side aortic anastomosis: end-to-side proximal anastomosis is indicated if the distribution of occlusive disease would cause an end-to-end anastomosis to jeopardize pelvic or visceral blood flow.

Distal anastomosis

When an aortic graft is inserted for the treatment of occlusive disease, the distal anastomosis should almost always be to the femoral artery. Even high-quality contrast arteriography with oblique pelvic views tends to underestimate the extent of external iliac disease. Extensive experience has revealed that creating the distal anastomosis in the groin rather than to the external iliac artery avoids the greater rate of graft failure attributable to progression of external iliac disease distal to the aortoiliac anastomosis. The use of prophylactic antibiotics, proper intraoperative skin preparation and draping, and meticulous surgical technique has avoided the increased rates of graft infection that were feared with use of anastomoses at the femoral level.

On completion of the proximal aortic anastomosis, the graft limbs are advanced through the previously prepared retroperitoneal tunnels. In male patients, surgical dissection in the region of the aortic bifurcation and the left common iliac artery is best kept to a minimum to avoid injury to the autonomic nerve plexuses important for erectile and ejaculatory function. On the right side, the retroperitoneal tunnel is created directly anterior to the native iliac artery, placing the graft posterior to the ureter to avoid its entrapment by the limb of the graft. On the left, the tunnel is placed beneath the sigmoid mesocolon in a slightly lateral position to avoid injury to the nerve plexus. Again, care is taken to prevent ureteral entrapment. The distal anastomosis is done end-to-side to the distal common femoral artery.

Profunda disease

It is absolutely critical to graft-limb patency for the distal anastomosis to allow adequate outflow, especially when the graft is placed in a patient with multilevel disease. In such patients with combined-segment disease, chronic occlusion of the superficial femoral artery and orificial stenosis of the profunda femoral artery may often compromise run off distal to the femoral artery. Here, ensuring adequate profunda outflow is the key to maintaining graft-limb patency, and correction of any significant profunda stenosis is of paramount importance. If the profunda is free of orificial stenosis, measures 4 to 5 mm in diameter as determined by gentle insertion of a graduated probe, and has a length of 20 to 25 cm, outflow via it alone is usually adequate to maintain graft patency.

There are several options that can be used to correct a profunda stenosis (Fig. 5). One standard technique involves extension of the femoral arteriotomy down the profunda, thus crossing the orificial stenosis. The end of the graft limb is fashioned with a long, beveled tip allowing the distal end of the anastomosis to extend to the profunda. The anastomosis thus creates a graft patch profundaplasty. Some surgeons prefer to use autogenous tissue (endarterectomized superficial artery or saphenous vein) to perform a separate profundaplasty. In this case, the prosthetic graft is anastomosed to the common femoral artery proximal to the patch profundaplasty, or directly on to the patch itself.

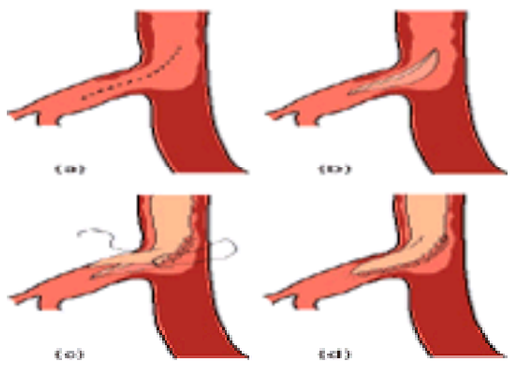


Fig. 5. Femoral profundaplasty: graft patch profundaplasty should be used to ensure adequate outflow via the profunda femoral artery.

Management of multilevel disease

The patient with multilevel (type III) aortoiliac occlusive disease often presents the therapeutic dilemma of whether an outflow procedure (e.g. femoral–popliteal bypass) must be performed simultaneously with an aortobifemoral bypass. Although it is generally accepted that 75 to 80 per cent of patients with multilevel disease will improve with an aortobifemoral bypass, many series have shown that 25 to 33 per cent will fail to have sufficient relief of ischemic symptoms and may require later infrainguinal procedures. If such patients could be identified before surgery, it would often be beneficial to perform simultaneous inflow and outflow revascularization, avoiding the difficulty and potential complications of subsequent reoperation for staged bypass and providing the best chance of relief of ischemic symptoms or salvage of the threatened limb.

While no definitive tests exist to identify patients requiring concomitant distal revascularization, there are some general guidelines. If unequivocally severe inflow disease in the aortoiliac segments is identified by absent or weak femoral pulses confirmed by intra-arterial femoral pressure measurements at the time of angiography, significant improvement may be expected with the inflow procedure alone. However, if clinical evaluation reveals only mild to moderate aortoiliac occlusive disease, a small, diffusely diseased, poorly collateralized profunda femoris, or significant infrapopliteal disease, unfavorable results may be predicted from an isolated inflow procedure.

The single, most important factor in determining the need for a simultaneous distal revascularization appears to be the degree of distal ischemia. Patients with significant tissue loss usually require restoration of pulsatile flow to the foot for healing. If the foot is severely ischemic, as with ischemic necrosis or digital gangrene

likely to require local amputation, it is clear that maximal revascularization is essential if the limb is to be saved.

Intraoperative monitoring of calf or ankle pulse-volume recordings or the ankle brachial index may be used to assess the hemodynamic improvement following completion of the aortobifemoral bypass, although improvement may be delayed in the cold, vasoconstricted lower extremity. Good clinical judgement and surgical experience remain essential in deciding when to proceed with distal reconstruction. Adjunctive intraoperative endovascular techniques such as transluminal balloon angioplasty to address distal occlusive lesions following proximal graft insertion may result in more liberal use of synchronous distal revascularization. Studies are needed to determine if these procedures are safe, reliable, and effective in the long term. At present, we do not commonly use them in these circumstances.

The frequency of combined operation appears to be increasing significantly in contemporary practice. The use of two surgical teams can considerably reduce operative time and thereby minimize the increased risk that was often cited in earlier years as an important reason to avoid simultaneous proximal and distal bypass in favor of a staged approach.

Graft selection

Although subtle differences remain, all modern large-diameter grafts have shown excellent patency and durability. Most believe that Dacron has superior handling characteristics with less hemorrhage at the suture lines. Some studies have suggested increased resistance to infection with polytetrafluoroethylene grafts. We prefer a collagen-impregnated, crimped Dacron graft. Regardless of the material chosen, the selection of graft size is a more important influence on long-term patency. The distal end should closely match the size of the outflow vessel to prevent the formation of a laminated thrombus, which can result from sluggish flow in an oversized graft and lead to graft-limb thrombosis.

Alternatives to direct aortoiliac reconstruction

Nonanatomic grafts have been used for the management of failed or infected grafts and in patients perceived to be at high risk for conventional aortoiliac reconstruction. In these circumstances, the goal is to achieve revascularization by means of grafts that use remote, frequently subcutaneous, pathways and potentially can be performed with lower morbidity and mortality.

Femorofemoral bypass

A femorofemoral bypass graft may be used in unilateral iliac disease. Failure of the graft due to progression of disease in the ‘donor' iliac system is unusual: several studies have suggested that the increase flow through the donor iliac artery that occurs as a result of the femorofemoral bypass graft may actually retard progression of the atherosclerotic disease in the donor iliac system. Previous concerns that the graft might create a ‘steal' phenomenon, significantly reducing perfusion of the donor limb, are unfounded in the absence of any hemodynamically significant stenosis in the proximal donor limb. Severe outflow lesions on the donor side with occlusion or stenosis of the superficial femoral artery may result in ‘steal', but this is not usually significant clinically.

Noninvasive arterial testing is important in assessment of the donor iliac artery. Thigh segmental pressures should be normal, thigh pulse-volume recording traces should have an excellent contour, with brisk upstroke, and femoral arterial Doppler waveforms should be triphasic. Collectively, these noninvasive test results suggest minimal disease in the donor iliofemoral system.

The donor iliac artery must also be evaluated with careful biplanar arteriography to establish clearly the absence of any hemodynamically significant inflow lesions. When superficial and profunda femoral arteries are diseased, noninvasive arterial studies may falsely suggest disease at the iliac level: pull-back intra-arterial pressures obtained at the iliac and femoral levels will disclose the presence of any hemodynamically significant lesions. If arteriography demonstrates an iliac stenosis, hyperemic testing or pressure measurements obtained after the administration of a vasodilator are necessary to be certain that the lesion will not become hemodynamically significant at the increased rates of flow that will occur after placement of the femorofemoral bypass graft.

The status of the outflow in the recipient limb was felt to be critical in determining the long-term patency of such grafts. One study showed that the patency of femorofemoral bypass grafts was reduced from 92 to 52 per cent if the superficial femoral artery in the recipient limb was occluded. Several recent studies have not found this association between the superficial femoral artery and femorofemoral bypass graft patency. A more detailed examination of outflow status, other than just the patency of the superficial femoral artery, may be required to demonstrate an affect on patency.

If the above criteria are applied when selecting patients for femorofemoral bypass graft the results are excellent and morbidity and mortality rates are low, even in high-risk patients. The procedure can be done under epidural, spinal or even local, anesthesia, completely avoiding any myocardial depressant effects of general anesthesia and the need for mechanical ventilation in patients with severe pulmonary disease. The common femoral arteries are exposed through bilateral, vertical incisions in the groin and the graft is placed in a subcutaneous, suprapubic tunnel from the donor to the recipient groin. An 8-mm prosthetic graft is generally chosen. Many currently believe that in this position a reinforced prosthesis is less likely to be occluded by external compression than conventional unsupported grafts. Regardless, the size of the conduit should closely match that of the femoral arteries in order to prevent the deposition of laminated thrombus within the body of the graft. The graft that appears to provide optimal hemodynamics is a gentle ‘C' shape in which the graft rises up though the subcutaneous tunnel and then down again to the opposite femoral artery.

Long-term primary patency rates for femorofemoral bypass grafts of 60 per cent and secondary patency rates of 80 per cent are common. Acceptable long-term patency, technical simplicity, and low morbidity and mortality have led numerous authorities to suggest the use of the femorofemoral bypass graft in low-risk as well as high-risk patients with unilateral iliac disease. In the sexually potent male patient, this approach has the additional advantage of avoiding the postoperative sexual dysfunction that may occur with direct reconstruction of the aorta. In general, however, good-risk patients are still best served by direct aortoiliac reconstruction.

Axillobifemoral bypass

High-risk patients with occlusive disease of the aorta or bilateral iliac arteries who present with limb-threatening ischemia are candidates for axillofemoral bypass as an alternative to either direct aortoiliac reconstruction or primary amputation (Fig. 6). In such patients an axillobifemoral bypass should generally be used in preference to an axillounifemoral graft, since the cumulative 5-year patency of the former is clearly superior; this difference has been attributed to the increased flow rates in the axillary limb of the graft in the bilateral bypass.

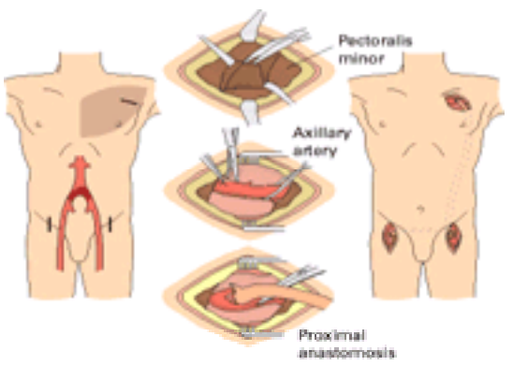


Fig. 6. Extra-anatomic bypass graft: axillobifemoral bypass graft may be the procedure of choice in high-risk patients with aortic or bilateral iliac arterial disease.

The axillary portion of the bypass graft is performed by exposing the first portion of the axillary artery via an infraclavicular skin incision. The pectoralis minor muscle is divided in the tendinous portion near its insertion. This maneuver provides excellent exposure of the artery and also allows the bypass graft to lie at the appropriate angle as it emerges from the subclavicular fossa and enters the subcutaneous tunnel on the chest wall. Bilateral, vertical incisions in the groin are used to expose the femoral arteries. The graft is then passed in a subcutaneous tunnel from the subclavicular dissection to the ipsilateral groin and from that groin to the contralateral femoral artery. This procedure usually requires general anesthesia, at least during the blunt dissection of the subcutaneous tunnels: the subclavicular and femoral dissections may be done under local anesthesia. The prosthetic graft of choice is generally externally reinforced polytetrafluoroethylene, although some surgeons prefer Dacron. Externally supported polytetrafluoroethylene will usually allow better results with graft thrombectomy if occlusion occurs. Prompt thrombectomy produces secondary patency rates of 60 to 70 per cent for occluded grafts. Although inferior to the long-term patency of aortobifemoral bypass or femorofemoral bypass grafts,

axillobifemoral grafts offer a reasonable compromise in the high-risk individual. Paradoxically, the mortality rate of 5 to 10 per cent exceeds that of direct aortic grafting, although it must be remembered that such grafts are undertaken in patients in whom the risks associated with direct operation are almost certainly increased.

Descending aorta–femoral artery bypass

A descending thoracic aorta to femoral artery bypass is an uncommonly performed procedure for reconstruction of the aortoiliac segment. It may be considered where direct reconstruction is not favorable, such as when there has been previous aortic surgery, many abdominal operations, or aortic graft infection.

Endovascular techniques

Recent technologic advances have produced a variety of endovascular techniques that may be used alone or as adjuncts to surgical bypass techniques for treating aortoiliac occlusive disease. These techniques are initially appealing because of their minimal invasiveness and cost savings, but the need for repeated interventions may attenuate these advantages. Life-time treatment costs and device costs need to be considered in the analysis. Few long-term results are available.

Percutaneous transluminal angioplasty

Although some of the decrease in the frequency with which aortobifemoral bypass is performed is attributable to increased use of nonanatomic surgical bypass, it is clear that the major impact has been through the marked increase in the use of iliac percutaneous transluminal angioplasty. In recent years there has been an acceleration of this trend by the rapid growth in the use of stents in conjunction with balloon angioplasty.

The reported long-term success of iliac percutaneous transluminal angioplasty is remarkably variable. A range of 5-year cumulative patency rates from 32 to 92 per cent was reported in a review of a large number of studies comprising more than 6000 procedures. This wide variation and the inconstant reporting standards have clearly contributed to the confusion and controversy over the proper role for this method. Much of these differences relate to nonobjective methods of assessment, differing criteria of ‘success', short follow-up periods, and lack of consistent reporting standards. Many investigators have eliminated initial failures of percutaneous transluminal angioplasty in assessing late results, and ‘patency' may not be the ultimate consideration, because most lesions selected for percutaneous transluminal angioplasty were stenotic and therefore initially patent.

Two carefully performed studies by Johnston *et al.* have done much to assess objectively the results of iliac percutaneous transluminal angioplasty and indicate when this may be considered a reasonable alternative to conventional surgical repair. Their prospective study of 667 iliac percutaneous transluminal angioplasties demonstrated that the chance of long-term success can be predicted accurately by four key variables: indication, site, severity, and run-off. In the most favorable circumstance, that of a localized common iliac stenosis in a patient with claudication alone and good run-off, the 3-year success rate in the Toronto experience was approx. 60 per cent, certainly inferior to that routinely achieved in all surgical series.

A true comparison of iliac percutaneous transluminal angioplasty with aortobifemoral bypass is difficult, because the anatomy (distribution and extent) and severity of disease are usually quite different in patients selected for each method of treatment. Few randomized prospective data exist. In one study by Wilson *et al.* comparing patients with limited disease suitable for treatment by either method, the overall cumulative long-term success rate favored surgery (81 per cent compared to 62 per cent success at 3 years). This statistically significant difference between surgery and percutaneous transluminal angioplasty was due almost entirely to the initial failure rate of the angioplasty (15 per cent) rather than to its late failures. In fact, if such early failures are excluded, the durability of the two methods appears equivalent. In a similar randomized prospective trial by Holm *et al.* in Sweden, immediate and 1-year results of percutaneous transluminal angioplasty and surgery were equivalent, and patients who underwent angioplasty had a significantly shorter hospital stay. It is important to note, however, that only patients who could be treated by both methods were included; these constituted only 5 per cent of the total number of patients treated during the study period.

It appears reasonable to conclude at present that iliac percutaneous transluminal angioplasty does have an established and valuable role in the treatment of selected patients with aortoiliac occlusive disease. It should be considered as potential primary therapy if the chance of long-term success is good. Proper criteria for its use have been fairly well established and consist principally of localized disease, ideally a short stenotic lesion in the common iliac artery. Iliac percutaneous transluminal angioplasty for focal iliac disease is also a valuable adjunct when combined with distal surgical procedures in appropriate patients with multilevel occlusive disease (Fig. 7). Given these indications, iliac percutaneous transluminal angioplasty may be a reasonable alternative to aortobifemoral bypass or related surgical procedures in perhaps 10 to 15 per cent of patients with aortoiliac occlusive disease.

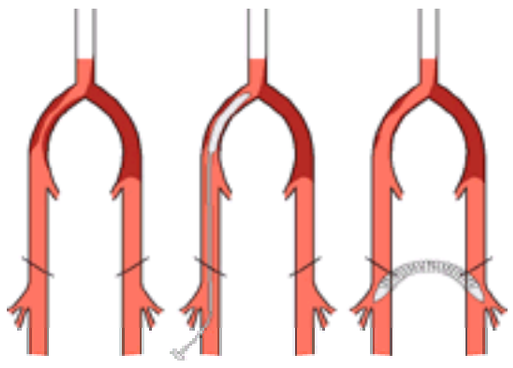


Fig. 7. Combined endovascular and extra-anatomical bypass procedure: transluminal angioplasty of an iliac arterial stenosis provides ‘inflow' for femorofemoral bypass of a contralateral iliac arterial occlusion; this combined technique allows use of a shorter prosthetic conduit than in axillofemoral bypass.

Stent graft

Technologic innovation has allowed the development of new treatments for the management of many forms of vascular disease. Endoluminal stent grafts (Fig. 8) have recently been utilized to treat aortoiliac occlusive disease. The applicability, safety, and durability of these techniques are unknown and await careful long-term analysis.

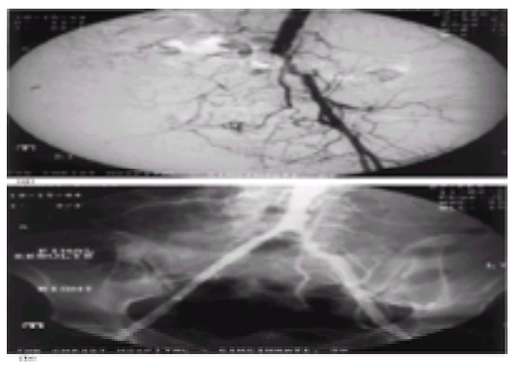


Fig. 8. Example of an iliac arterial occlusion treated with an endovascular graft (by courtesy of Dr J.B. Edwards, University of Cincinnati Medical Center).

Summary

Aortoiliac occlusive disease is a common manifestation of atherosclerosis. It is usually a chronic disease, but several specific acute presentations may occur. The diagnosis is usually suggested by history and physical examination, and confirmed by noninvasive arterial testing. Imaging studies are essential in the planning of

interventions. A wide variety of treatment options exist, including risk-factor modification and exercise, standard open operations, and newer endovascular techniques. Overall management should be based on sound surgical principles and directed by a thoroughly trained specialist who understands the natural history of the disease and is fully capable of providing all treatment options. Generally, if these principles are adhered to, excellent durable results can be expected.

Further reading

Ali M, Norcross ED, Brothers TE. Iliac and femoral artery occlusion by thromboemboli from an abdominal aortic aneurysm in the setting of blunt abdominal trauma. *Journal of Vascular Surgery* 1998; **27**: 545–8.

Bandyk DF *et al. In situ* replacement of vascular prostheses infected by bacterial biofilms. *Journal of Vascular Surgery* 1991; **13**: 575.

Baum RA *et al.* Multicenter trial to evaluate vascular magnetic resonance angiography of the lower extremity. *Journal of the American Medical Association* 1995; **274**: 875–80.

Brener BJ *et al.* Femorofemoral bypass: a 25-year experience. In: Yao JST, ed. *Long-term results in vascular surgery*, pp. 385–93. Appleton and Lange, East Norwalk, 1993.

Brewster DC. Clinical and anatomic considerations for surgery in aortoiliac disease and results of surgical treatment. *Circulation* 1991; **83** (Suppl. I): 42–52.

Brewster DC. Current controversies in the management of aortoiliac occlusive disease. *Journal of Vascular Surgery* 1997; **25**(2): 365–79.

Brewster DC, Darling RC. Optimal methods of aortoiliac reconstruction. *Surgery* 1978; **84**(6): 739–48.

Brewster DC *et al.* Aortofemoral graft for multilevel occlusive disease: predictors of success and need for distal bypass. *Archives of Surgery* 1982; **117**: 1593–600.

Bry JD *et al.* An assessment of the positive predictive value and cost-effectiveness of dipyridamole myocardial scintigraphy in patients undergoing vascular surgery. *Journal of Vascular Surgery* 1994; **19**: 112–24.

Criado E, Keagy BA. Use of the descending thoracic aorta as an inflow source in aortoiliac reconstruction: indications and long-term results. *Annals of Vascular Surgery* 1994; **8**: 38–47.

Criado E *et al.* Femorofemoral bypass grafts: analysis of patency and factors influencing long-term outcome. *Journal of Vascular Surgery* 1993; **18**: 445–50.

Dalman RL *et al.* Simultaneous operative repair of multilevel lower extremity occlusive disease. *Journal of Vascular Surgery* 1991; **13**: 211–21.

DeBakey ME, Lawrie GM, Glaeser DH. Patterns of atherosclerosis and their surgical significance. *Annals of Surgery* 1985; **201**: 115–31.

de Smet A, Ermers E, Kitslaar P. Duplex velocity characteristics of aortoiliac stenoses *Journal of Vascular Surgery* 1996; **23**: 628–36.

Eagle KA *et al.* Dipyridamole–thallium scanning in patients undergoing vascular surgery. *Journal of the American Medical Association* 1987; **257**: 2185.

Holm J, Arfvidsson B, Jivegad L. Chronic lower limb ischemia: a prospective randomized controlled study comparing the 1-year results of vascular surgery and percutaneous transluminal angioplasty (PTA). *European Journal of Vascular Surgery* 1991; **5**: 51–2.

Itoch JR *et al.* Use of magnetic resonance angiography for the preoperative evaluation of patients with infrainguinal arterial occlusive disease. *Journal of Vascular Surgery* 1996; **23**: 792–801.

Johnston KW. Iliac arteries: reanalysis of results of balloon angioplasty. *Radiology* 1993; **186**: 207–12.

Johnston KW, Rae H, Hoss-Johnston SA. Five-year results of a prospective study of percutaneous transluminal angioplasty. *Annals of Surgery* 1987; **206**: 404–13. Kannel WB *et al.* Intermittent claudication: incidence in the Framingham study. *Circulation* 1970; **41**: 857.

Keen RR *et al.* Surgical management of atheroembolization. *Journal of Vascular Surgery* 1995; **21**: 773–81.

Leriche R. Des obliterations arterielles hautes (obliteration de la terminaison de l'aorte) comme cause des insuffisances circulatoires des membres inferieurs. *Bulletins et Memoires de la Societé des Chirurgiens de Paris* 1923 (Decembre).

Leriche R, Moral A. The syndrome of thrombotic obliteration of the aortic bifurcation. *Annals of Surgery* 1948; **127**: 193.

McCarthy WJ *et al.* Descending thoracic aorta-to-femoral artery bypass: ten years' experience with a durable procedure. *Journal of Vascular Surgery* 1993; **17**: 336–48.

Musser DJ, Nicholas GC, Reed JFI. Death and adverse cardiac events post carotid endarterectomy. *Journal of Vascular Surgery* 1994; **19**: 615–22.

Nevelsteen A, Wouters L, Suy R. Aortofemoral Dacron reconstruction for aorto-iliac occlusive disease: a 25-year survey. *European Journal of Vascular Surgery* 1991; **5**(2): 179–86.

Nolan KD *et al.* Femorofemoral bypass for aortobifemoral graft limb occlusion: a 10-year experience. *Journal of Vascular Surgery* 1994; **19**: 851–7.

Ohki T *et al.* Endovascular aortounifemoral grafts and femorofemoral bypass for bilateral limb-threatening ischemia. *Journal of Vascular Surgery* 1996; **24**(6): 984–6; discussion 996–7.

Poulias GE *et al.* Aorto-femoral bypass and determinants of early success and late favorable outcome: experience with 1000 consecutive cases. *Journal of Cardiovascular Surgery* 1992; **33**: 664–78.

Rholl KS, van Breda A. Percutaneous intervention for aortoiliac disease. In: Standness DE, van Breda A, ed. *Vascular diseases: surgical and interventional therapy*, pp. 433–66. Churchill Livingstone, New York, 1994.

Rutherford RB. Aortobifemoral bypass the gold standard: technical considerations. *Seminars in Vascular Surgery* 1994; **7**(1): 11–16.

Rutherford RB, Patt A, Pearce WH. Extra-anatomical bypass: a closer view. *Journal of Vascular Surgery* 1987; **6**: 437–46.

Sanchez LA *et al.* Endovascular grafting for aortoiliac occlusive disease. *Seminars in Vascular Surgery* 1997; **10**(4): 297–309.

Schneider JR *et al.* Femorofemoral versus aortobifemoral bypass: outcome and hemodynamic results. *Journal of Vascular Surgery* 1994; **19**(1): 43–55; discussion 55–7.

Segar DS *et al.* Dobutamine stress echocardiography: correlation with coronary lesion severity as determined by quantitative angiography. *Journal of the American College of Cardiology* 1992; **19**: 1197–202.

Szilagyi DE *et al.* A thirty-year survey of the reconstructive surgical treatment of aortoiliac occlusive disease. *Journal of Vascular Surgery* 1986; **3**(3): 421–36.

Taylor LMJ *et al.* The incidence of perioperative myocardial infarction in general vascular surgery. *Journal of Vascular Surgery* 1991; **15**: 52–61.

Wilson SE, Wolf GL, Cross AP. Percutaneous transluminal angioplasty versus operation for peripheral arteriosclerosis: report of a prospective randomized trial in a selected group of patients. *Journal of Vascular Surgery* 1989; **9**: 1–9.

Wylie EJ. Thromboendarterectomy for arteriosclerotic thrombosis of major arteries. *Surgery* 1952; **32**: 275–90.

17.5.2 Femoral and distal arteries

John A. Murie

Introduction
Presentation and classification
Assessment
Medical treatment
Indications for intervention
Subclinical disease
Intermittent claudication
Rest pain and critical ischaemia
Alternatives to surgery
Operative techniques and graft materials
Reversed autogenous vein bypass
***In situ* bypass**
Femoropopliteal bypass in the absence of suitable long saphenous vein
Other operations
Lumbar sympathectomy
Complications
Early complications
Late complications
Prevention of graft occlusion
Further reading

Mr Michael Callam was the coauthor of this chapter in the first edition. Much of the present version is based on that chapter and his contribution to it is gratefully acknowledged.

Introduction

Although this section deals with infrainguinal occlusive arterial disease, one must appreciate that the pathology affecting the femoral and distal vessels does not exist in isolation; the cardiac, cerebral, and less commonly the mesenteric and renal circulations may be affected. Initial deposits of atheroma in the vessel wall are overlaid with hyaline collagenous material which projects into the arterial lumen. This plaque may ulcerate, leading to superimposed thrombosis which organizes and enlarges, further narrowing the vessel and causing turbulent flow. Turbulence accelerates the process leading to occlusion of the vessel.

The natural history of infrainguinal arterial disease is not a simple steady deterioration towards amputation; it is more often characterized either by stable intermittent claudication or even by symptomatic improvement as collateral channels enlarge. The risk of gangrene or pregangrene within a year of presentation is about 5 per cent, and about 2 per cent per annum thereafter. Of every 100 patients with claudication, approximately 40 will improve, 40 will remain unchanged, and 20 will require active intervention. It is important to appreciate that the general mortality rate of these patients is twice that of age- and sex-matched controls without peripheral arterial disease.

Presentation and classification

The three cardinal features of chronic peripheral lower limb ischaemia are intermittent claudication, rest pain, and gangrene, representing an increasing degree of severity of ischaemia. Intermittent claudication is a cramp caused by inadequate oxygenation of muscle. It is initiated by walking and relieved by rest; generally the calf muscles are most affected. The distance traveled before claudication occurs remains roughly the same unless the underlying condition deteriorates, although the symptom is more pronounced on hurrying or going uphill.

Rest pain occurs when the blood supply is so poor that tissue perfusion is inadequate even at rest. The pain classically affects the toes or forefoot (the most distal part of the limb) although in severe cases it may involve the whole foot or calf. It is usually first noticed in bed, when the patient is horizontal, the beneficial effect of gravity is removed and the foot is warmed, thereby increasing metabolism. Furthermore, external stimuli are reduced at night and so pain may be appreciated more readily. Rest pain is helped by hanging the leg out of bed, standing, or even walking a little. As ischaemia progresses the pain becomes continuous, requiring opiate analgesia for its control. The third clinical feature is gangrene. This is the endstage of ischaemia when the circulation is so poor that necrosis ensues. It usually begins distally in the foot (Fig. 1).



Fig. 1. Ischaemic gangrene affecting the toes of the foot.

In addition to the three cardinal features of ischaemia the concept of ‘critical limb ischaemia' is useful in vascular surgery. This is defined as persistent rest pain requiring analgesia for more than 2 weeks, or ulceration, or gangrene of the foot, plus an ankle systolic pressure below 50 mmHg. In patients with diabetes, owing to the unreliability of ankle pressure recording because of vascular calcification, the pressure criterion is replaced by absence of ankle pulses.

Assessment

Assessment must establish the degree of ischaemia, whether it requires treatment and, if so, the most appropriate treatment. History and examination will usually identify the presence or absence of arterial disease and suggest its severity. Skin temperature, pallor on elevating the limb followed by dependent rubor, and the absence of pulses are particularly important features. The palpation of pulses should give the surgeon a rough idea of the site of arterial occlusion. More exact assessment requires a Doppler ultrasound probe and sphygmomanometer cuff to measure the highest opening systolic pressure of the three ankle arteries (Fig. 2). This is divided by the higher of the brachial systolic pressures to obtain an ankle/brachial pressure index, the normal value of which is greater than 1. Values of 0.6 to 0.9 are typical of claudication, 0.3 to 0.6 of rest pain, and below 0.3 of incipient or actual gangrene. In some individuals with apparently normal values at rest, occult disease may be uncovered if the ankle/brachial index falls after exercise.



Fig. 2. Measurement of ankle/brachial index. The highest systolic pressure of the three ankle arteries is obtained using a hand-held Doppler ultrasound probe. This is divided by the higher systolic pressure of the two brachial arteries.

The need for intervention is easy to assess in those with mild claudication or critical ischaemia. Between these extremes a balance must be struck between the risk to life and limb from any proposed procedure and the compromised lifestyle which may result from conservative management. This balance will be affected by the patient's social circumstances and by the results of further investigations to assess fitness for operation, and by the site and morphology of the occlusion(s). Although simple palpation of pulses usually allows the approximate level of occlusion or stenosis to be recognized, if surgical or radiological intervention is intended, some form of imaging is required to provide more detail: the standard imaging technique is angiography (Fig. 3). Radiographs are exposed after injection of radio-opaque contrast medium into the arterial tree through a fine catheter inserted via the femoral artery in the groin. Current techniques using non-ionic contrast media and narrow gauge catheters are safe, though invasive, and the angiogram remains the investigation of choice. Computed (digital subtraction) angiography has supplanted the basic conventional technique, allowing areas of special interest to be highlighted (Fig. 4). This modern method may employ either an intra-arterial or an intravenous injection. The advantage of the former is that very fine catheters may be used and small quantities of contrast media injected. The advantage of the latter is that the arterial tree does not have to be invaded at all. However, the intravenous technique requires administration of a large volume of contrast agent, which debilitated patients may tolerate poorly. Furthermore, the image quality of intravenous digital subtraction angiography for the leg vessels is poor.

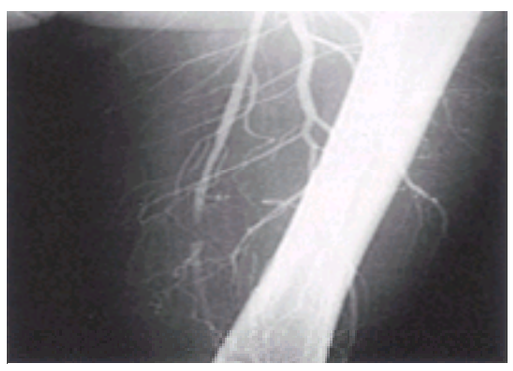


Fig. 3. Conventional angiogram showing classically sited occlusion in the distal superficial femoral artery.

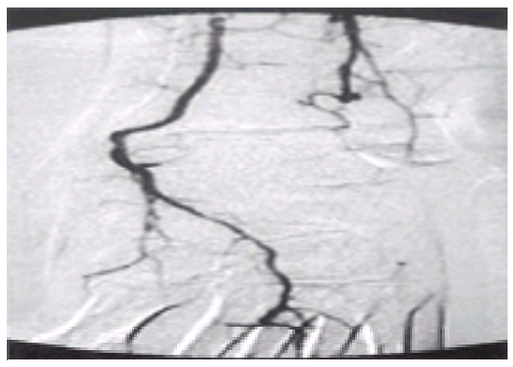


Fig. 4. Absence of complete pedal arch shown by intra-arterial digital subtraction angiography.

It must be stressed that angiography should only be performed if intervention is intended. It allows an assessment of whether intervention is technically feasible and enables the most appropriate form of treatment to be chosen. The appearance of the aorta and iliac vessels is checked to confirm that there is no impairment of inflow to the leg. The sites of stenosis and occlusion in the leg arteries themselves are noted, and patency of the distal arteries (outflow) assessed. The usefulness of angiography in assessing patency of very distal arteries when an upstream occlusion is present has been questioned. The technique of pulse generated run-off may detect patent vessels at the ankle which have not been adequately demonstrated by angiography. A pneumatic cuff around the upper calf is mechanically inflated and deflated rapidly to generate a pseudopulse which is detected by a Doppler ultrasound probe over any patent ankle artery. In patients with critical ischaemia this may allow the possibility of a femorodistal bypass to be recognized in the presence of a negative angiogram. In the absence of a pulse generator system, a reasonable assessment of the availability of ankle vessels for bypass surgery can be made by insonating the vessels with a simple hand-held continuous wave Dopper probe, with the leg dependent. This so-called ‘dependent Doppler' technique is more reliable than angiography for very distal arteries in the presence of a major upstream occlusion. While duplex ultrasound scanning has made significant advances as an arterial imaging tool at many sites, its use in the infrainguinal setting is rather limited. It is probable, however, that the new generation of magnetic resonance imaging equipment will have much to offer in this area. Magnetic resonance angiography is a rapidly expanding field which is likely to render many of today's imaging modalities redundant.

Atheroma has a predilection for certain arterial sites and patterns of disease are recognizable: within the leg these common patterns are femoropopliteal disease (Fig. 3) and distal disease (Fig. 4). Nevertheless, atheroma is never entirely limited to one site, although relieving an obstruction at one site may result in a general relief of symptoms. The most common site of stenosis or occlusion in the femoropopliteal segment is the junction between the middle and distal thirds of the segment, where the superficial femoral artery exits through the adductor hiatus. Typically, an initial stenosis progresses to occlusion, followed by an eventual proximal propagation of thrombus to the origin of the superficial femoral artery; outflow from the common femoral is then solely into the profunda femoris (deep femoral) artery. Although distal propagation of thrombus may occur, the popliteal artery usually stays patent because stasis here is prevented by inflow from the profunda femoris via the geniculate collaterals. There is, therefore, often a patent vessel at this level which can accept the distal end of a bypass graft.

Distal obliterative disease which is not associated with proximal atheroma should raise the suspicion of another pathology, especially diabetes mellitus, arteritis, or previous embolic disease. However, the most common distal disease is found in older patients and is accompanied by extensive proximal atheroma. Distal disease may or may not be amenable to reconstructive surgery, depending on the state of the three (crural) vessels of the lower leg.

Medical treatment

Medical treatment may be indicated when the disease is not of sufficient severity to warrant operation (including angioplasty); when operation is impossible, inappropriate, or unsuccessful; or as an adjunct to operation. Several general measures are applicable to all patients whether or not they have surgery, for instance weight reduction in the obese, correction of anaemia or polycythaemia, treatment of hyperlipidaemia, and control of diabetes. The judicious treatment of heart failure and hypertension may also improve perfusion, but b-blocking drugs should be avoided as they may further compromise a diseased peripheral circulation.

Smoking is the most important correctable risk factor. Stopping smoking may be the only treatment that many patients require; claudication not infrequently improves spontaneously. Stopping smoking may not reduce atheroma that is already present, but continuation of smoking leads to an increased deposition and compromises the development of a collateral circulation. Smoking increases the risk of amputation and the incidence of graft occlusion after surgery.

Exercise is the other arm of effective conservation treatment: it may double the distance that can be walked before pain occurs in up to 80 per cent of patients. It has been suggested that selective exercise of those muscles which are most ischaemic produces the best results and this is best done within a structured exercise programme. Such a programme requires physiotherapist supervision and should employ a thrice weekly regimen in which patients are encouraged to walk to near maximal pain. Even rest pain may occasionally benefit from exercise.

Apart from the above general measures, no other form of conservative treatment is widely accepted as likely to offer significant benefit in claudication or rest pain. A variety of drugs may occasionally offer a modest benefit; the most widely used are oxypentifylline (pentoxifylline) which alters red cell deformability so helping flow in the microcirculation, and naftidrofuryl which has an effect on ischaemic cell metabolism. More recently prostacyclin analogues have been used for their antithrombotic effect.

Indications for intervention

Subclinical disease

In a patient with unilateral symptoms it is not unusual to find angiographic evidence of early disease on the contralateral side. It is not yet known whether intervention for early asymptomatic disease confers benefit over a conservative policy and intervention in this group should occur only in the confines of a clinical trial addressing this question.

Intermittent claudication

Intermittent claudication represents the ‘middle ground'. Any decision to intervene must take into account the possibility that spontaneous improvement may occur, especially if smoking is stopped and exercise adopted. Symptomatic improvement is especially likely within the first 6 months after onset of claudication. Active intervention may be considered in the form of open surgery (generally a femoropopliteal bypass) or percutaneous transluminal angioplasty using a balloon catheter to dilate narrowed or occluded vessel segments. However, while popular with some (especially physicians and radiologists), there is no evidence to suggest that any worthwhile clinical improvement can be obtained using the latter method. Surgery, which may be effective in those with claudication, should only be offered after due consideration of its likely risks and benefits for an individual patient. Operation is likely to be attractive only for a person who experiences claudication after a short walk, has a suitable vein available for use as a bypass conduit, and whose quality of life is greatly affected by his or her symptoms.

Rest pain and critical ischaemia

Rest pain or critical ischaemia despite appropriate medical management requires intervention if at all possible. Revascularization should generally be attempted if angiography, pulse generated run-off, or ‘dependent Doppler' indicates that percutaneous transluminal angioplasty or reconstructive surgery is feasible. Nevertheless, in individual cases the chance of success may be recognizably slim and the decision to proceed requires fine judgement. There is a small group of patients, often unfit, who will benefit from primary amputation.

Alternatives to surgery

Percutaneous transluminal angioplasty is a radiological technique in which a guide wire is introduced percutaneously through the common femoral artery to lie within a stenosis. A catheter with a balloon at its end is introduced over the guide wire and the balloon is inflated within the narrowed segment (Fig. 5). Although introduced by Dotter and Judkins in 1964 for the treatment of atheromatous femoropopliteal disease, percutaneous transluminal angioplasty was made popular primarily by Gruntzig and his colleagues. Today the technique is regularly applied to the femoropopliteal segment and to many sites other than the leg (Fig. 6). It is carried out under local anesthesia and is safe in experienced hands: the main complications of bleeding or thrombosis are unusual and can usually be remedied with surgical assistance. Percutaneous transluminal angioplasty is especially attractive for debilitated patients with severe ischaemia. Generally, occlusions up to 10 cm long may be satisfactorily dealt with by this method. Nevertheless, it is sometimes impossible to push the guide wire and balloon catheter through an occlusion, especially if it is long. Attempts have been made to overcome this situation by burning through the occlusion using a laser (laser angioplasty) to allow proper placement of a guide wire and balloon catheter. Laser angioplasty has not, however, fulfilled its initial promise and has largely fallen into disrepute.

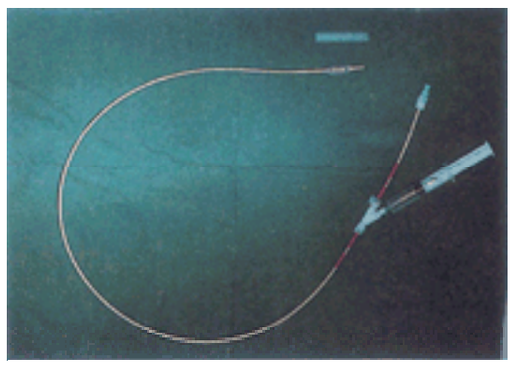


Fig. 5. Percutaneous transluminal angioplasty catheter.

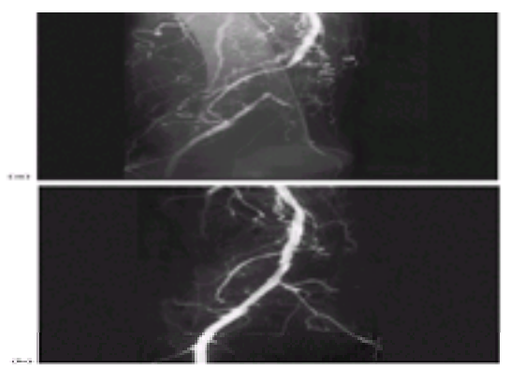


Fig. 6. (a) Angiogram showing superficial femoral artery occlusion. (b) Superficial femoral artery after percutaneous transluminal angioplasty. (By courtesy of Dr Kieran McBride, consultant radiologist, Royal Infirmary of Edinburgh.)

Mechanical atherectomy devices which may be introduced into the artery in a similar fashion to the angioplasty catheter are available. The atherectomy catheter, however, has a cutting mechanism at its tip; a hollow core with side orifice and a sliding blade is a popular pattern, as is a high speed rotating cutter. These devices are used to cut through atheroma, retaining the resulting debris within their core for later removal. Their use is largely confined to a few specialist centres; despite having been used by enthusiasts for several years the results have not encouraged a more widespread acceptance of the technique.

When an acute occlusion occurs in a chronically atheromatous artery (usually the superficial femoral artery), it may cause limb-threatening ischaemia for which emergency surgery used to be the only remedy. Today it is possible, in selected patients, to pass a long catheter percutaneously, via the common femoral artery, into the recent occlusion and to infuse a thrombolytic agent such as urokinase or tissue plasminogen activator directly into the thrombus. If thrombolysis is successful, the

narrowed arterial segment may be improved by percutaneous transluminal angioplasty or selective operation. In recent years the simple infusion technique has been joined by ‘pulse spray' lysis in which the lytic agent is sprayed in pulses into the thrombus. The method requires special equipment and can reduce the time to lysis to a few hours.

Operative techniques and graft materials

The choice of operation for patients with occlusive disease of the lower limb arteries depends on the site of the occlusion(s), the availability of a suitable graft, and the experience of the operator. Although a variety of local bypasses or patch angioplasties may occasionally be desirable, by far the most common procedure is a bypass from the common femoral artery to a distal vessel; this is usually the popliteal artery—either above or below knee level—but it may also be to the tibioperoneal trunk or to any of the three (crural) vessels of the lower leg. The modern technique differs little from that first described by Kunlin in 1948; a shunt is constructed in parallel with the occluded artery using end-to-side anastomoses both proximally and distally (Fig. 7). The rationale is to transport blood around an occluded segment while avoiding operative trauma to collaterals.

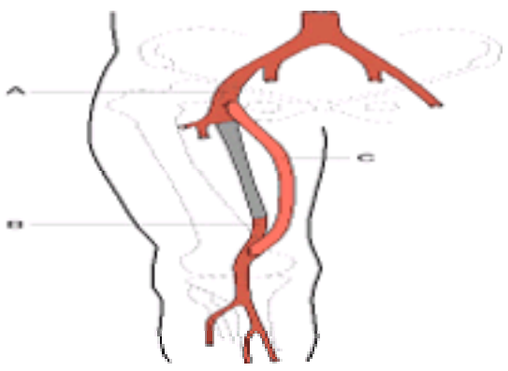


Fig. 7. Femoropopliteal bypass. A, common femoral artery; B, above-knee popliteal artery; C, femoropopliteal bypass graft.

When available, the autogenous long saphenous vein is the best graft material for femorodistal bypass (Fig. 8). However, this vein may be too small in calibre, thrombosed, markedly varicose, or may have been removed surgically in the past. Also, its usable length may be too short for the proposed operation. The short saphenous, the cephalic, and the basilic veins may then be used, individually or in combination with themselves or with a limited length of long saphenous vein. It is common for the vein to be assessed visually at the time of operation, but it is possible to assess the usefulness of the long saphenous vein before surgery, either by duplex ultrasound scanning or by saphenography.



Fig. 8. Long saphenous vein excised, tributaries tied, and dilated by saline injection to check for leaks before being used as a femoropopliteal bypass graft.

If the vein is inadequate it may be necessary to use a graft of synthetic material, the most popular of which is expanded polytetrafluoroethylene (**PTFE**) (Fig. 9). This inert substance has considerable resistance to thrombosis. Early patency rates for femoropopliteal bypass, both above and below knee, using such grafts are similar to those for autogenous vein (Fig. 10). Late patency rates are better when vein is used, especially at the below-knee site. Nevertheless, late graft occlusion is not always associated with limb loss, even when the original operation was undertaken for critical ischaemia, and limb salvage rates for vein and PTFE femoropopliteal bypasses remain broadly similar.

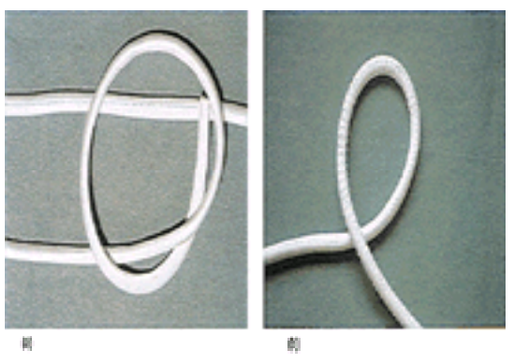


Fig. 9. (a) Polytetrafluoroethylene arterial prostheses. These are 6-mm diameter tubes, one of which has additional stiff external supporting rings (b). (Figure kindly provided by W.L. Gore & Associates.)

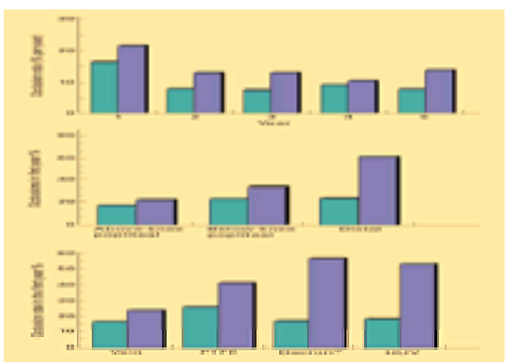


Fig. 10. (a) Occlusion rate for above-knee grafts for each year after surgery: (green) vein grafts, (blue) mixed prosthetic grafts. (b) Occlusion rate in first year against anastomotic site for vein and prosthetic grafts: (green) vein grafts, (blue) prosthetic grafts. (c) Occlusion rate in first year for various graft materials in above- and below-knee situations: PTFE, polytetrafluoroethylene; HUV, human umbilical vein; (green), above knee; (blue), below knee.

An alternative to PTFE is glutaraldehyde-stabilized human umbilical vein supported by an external polyester mesh. This tanned graft is non-antigenic and resists biodegradation reasonably well. Patency rates compare favourably with those of PTFE, but umbilical vein is expensive and aneurysmal degeneration of the graft has been described. Other less popular grafts are available, such as those made from externally supported Dacron velour or ovine collagen.

When a graft of any type is inserted it is good practice to ensure at the end of the operation that the anastomoses, especially the distal one, are technically satisfactory and that flow through the graft is adequate to maintain patency. Several techniques are available. The electromagnetic flow meter has been popular but it is difficult to achieve consistent results with this technique and peroperative Doppler ultrasonography is more satisfactory. To assess the integrity of an anastomosis the choice is between peroperative angiography (Fig. 11), which is cheap and readily available, and angioscopy, which is rather more expensive and sometimes awkward. At the present time the author's choice is to use Doppler ultrasonography to assess the haemodynamics of the situation and angiography to assess the anastomoses but, if the angioscope has been used for graft preparation (see below), it is convenient to use it also for the anastomosis check. It should be appreciated that the non-availability of these techniques should not dissuade the surgeon from operating, especially in patients requiring limb salvage.

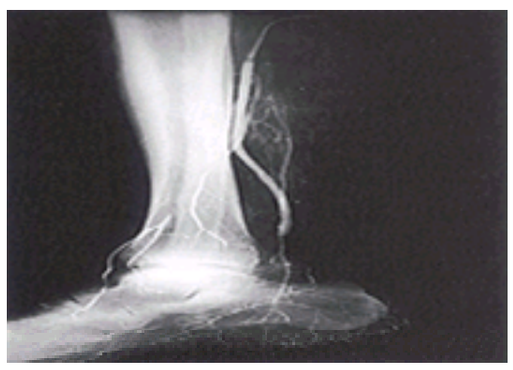


Fig. 11. Peroperative angiogram of femorocrural bypass, showing a good distal anastomosis with run off both distally and proximally into the posterior tibial artery.

Reversed autogenous vein bypass

The most common reconstructive operation for occlusive disease below the inguinal ligament is a bypass from the common femoral to the popliteal artery, performed for occlusion of the superficial femoral artery. A bypass to the popliteal artery above the knee is the preferred option; if the above-knee popliteal artery is very atheromatous or occluded then the below-knee vessel or even the crural vessels may be used.

Femoropopliteal bypass using reversed autogenous long saphenous vein is the archetypal operation. The popliteal artery is exposed via a medial incision and a satisfactory site for the lower anastomosis established. The femoral artery is exposed at the groin. A tunnel beneath sartorius is made between groin and supragenicular politeal fossa, running orthotopically behind the knee if the infrageniculate site is to be reached. An adequate length of long saphenous vein is excised after tying and dividing its tributaries (Fig. 8) and heparin is administered to the patient. It is generally agreed that for use as a reversed femoropopliteal bypass graft, the vein must have a minimum diameter of at least 4 mm. The vein is checked for leaks and if satisfactory is reversed to deactivate the valves, and anastomosed in an end-to-side fashion to the arteries.

Either the proximal or the distal anastomosis may be made first, but the inexperienced surgeon should complete the distal anastomosis before the proximal as this allows the leg to be fully extended with the graft lying in the subsartorial tunnel after one anastomosis in such a way that the exact length of graft which is needed is easily recognized. On completion of both anastomoses the bypass may be checked by assessing the lower anastomosis, either by angiogram or angioscopy via an untied vein tributary which is tied after the check procedure. Adequacy of flow through the graft may be checked using an electromagnetic flow meter or Doppler ultrasonography. In general, there is no need to reverse the heparin at the end of the procedure, nor is there any need for external drains.

In situ bypass

This previously unfashionable operation has enjoyed renewed popularity since the late 1980s, possibly because better instruments for disrupting venous valves have become available. The concept is the same as for the reversed vein operation inasmuch as the sites for proximal and distal anastomoses are the same; the essential difference is that the vein is not excised but left *in situ.*

The vein tributaries must all be recognized and tied off to prevent development of significant arteriovenous fistulas which compromise the graft blood flow distally, cause generalized oedema, and contribute to graft failure. The tributaries may be recognized by a variety of techniques from exposing the whole length of the vein for visual inspection, to methods using Doppler ultrasonography, angiography, or angioscopy. The operator may choose to vary the method of tributary occlusion depending on the length of graft. Visual inspection through a long wound is simple and usually trouble free for above-knee bypass. However, for more distal bypasses, especially those to ankle or foot level, the wound required for direct inspection is very long; such wounds, especially distally, are prone to ischaemic necrosis and infection. For long bypasses tributary recognition is possible using the angioscope; only small incisions directly over tributaries are then needed for their occlusion. Minimally invasive scopes are also available to allow the saphenous vein to be accessed externally in a ‘potential space' within the subcutaneous plane; tributaries may be clipped through such scopes without the need for widespread venous exposure.

When using the vein *in situ* it is best to make the proximal anastomosis first and to declamp and allow the vein to fill with blood down to the first valve (Fig. 12). A valve cutter (valvulotome) is then introduced from the distal end of the vein up to the femoral anastomosis (Fig. 13). Withdrawing the valvulotome disrupts the valves and allows the blood to fill the graft as each set of valves is broken. Eventually the cutter is withdrawn followed by a spurt of pulsatile blood. The graft is then clamped and the lower anastomosis completed. Although this can be done blind, some prefer valvulotomy under direct vision using the angioscope inserted from above; for the direct vision technique the valves are cut before any anastomosis is completed.

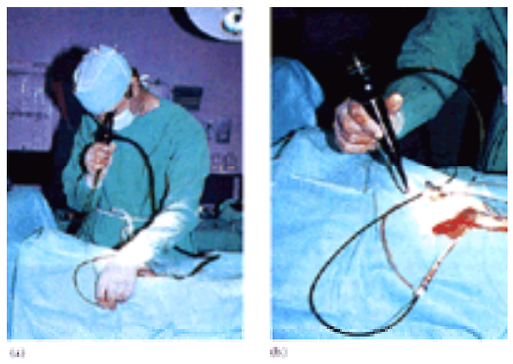


Fig. 12. *In situ* femoropopliteal vein bypass. The long saphenous vein has been anastomosed end to side to the common femoral artery and the clamps released. The vein fills with arterial blood down to the first venous valve (at the tip of the forceps).

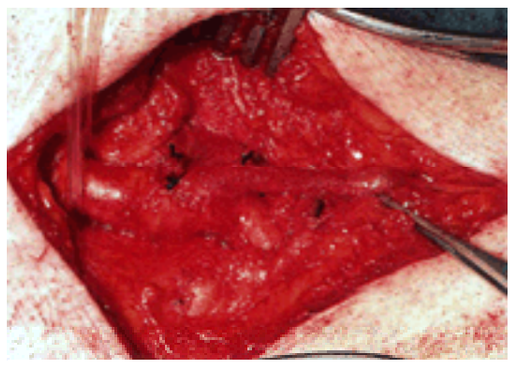


Fig. 13. (a) Hall valve cutter. (b) Close-up of head of valve cutter. The lower cylinder holds the vein open to allow the upper cylinder to disrupt the valves on traction.

The major advantage of the *in situ* technique, in addition to the fact that the vein and its blood supply are left largely undisturbed, is that the vein with the greater calibre in the groin is anastomosed to the large arteries, while the distal vein of smaller calibre is used in the anastomosis to smaller vessels further down the leg. It is therefore haemodynamically attractive and it is possible to achieve good results with a vein diameter of 3 mm or even less. Such grafts may be successful not only when anastomosed to the popliteal and proximal crural vessels, but even when the distal anastomosis is fashioned at ankle level. The technique, however, also has some technical disadvantages. First, arteriovenous fistulas must be recognized and dealt with carefully. Second, the distal anastomosis may need to be made to a very small vessel and experience is required if this is to be successful. Such surgery is best done under magnification (× 2.4 is adequate) and the result must be checked at completion by a peroperative angiogram or by angioscopy (if a thin enough scope is available).

Femoropopliteal bypass in the absence of suitable long saphenous vein

Autogenous ipsilateral long saphenous vein should be used for bypass from the common femoral artery to levels below the knee joint if possible. Some surgeons believe that the superiority of vein is such that if the long saphenous vein is compromised by severe varicosity, thrombosis, or small calibre, a search should be made for another vein source, such as the contralateral long saphenous, the short saphenous, cephalic, or brachial veins. These may be used in combination as a vein–vein composite graft. Many others believe that if the ipsilateral long saphenous vein is unavailable it is reasonable to use a manufactured alternative. The most commonly used are made from PTFE.

PTFE grafts come in a variety of calibres but 6 mm (occasionally 8 mm) is usually chosen for bypasses in the leg (Fig. 9). Some grafts are supported by external rings: these are an attractive adjunct, especially when the knee joint has to be crossed. The neointimal hyperplasia which occurs, especially at the distal anastomosis in PTFE grafts, has long been held to be the principal reason for their poorer patency (compared with vein) and it is likely that this tissue build-up is due to a compliance mismatch between the non-elastic graft and the expansile artery. The interposition of a piece of vein as a collar (Miller collar) between the PTFE and the artery has been suggested to be beneficial, and the results of at least one trial favours its use for grafts inserted below the level of the knee (Fig. 14). Modifications (usually eponymous) of the Miller collar have been suggested by several authors. There is no proof of superiority of one over another and it is likely that all fulfill a function similar to the original construction described by Miller.

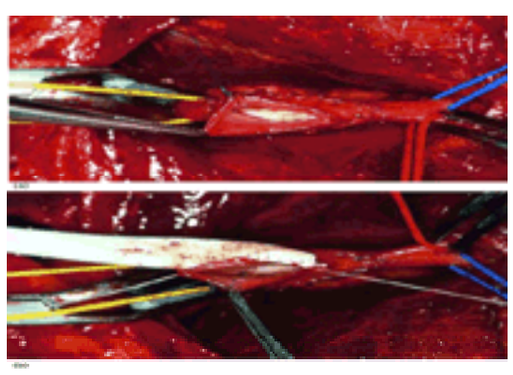


Fig. 14. (a) A collar of vein has been sewn on to an arteriotomy in the infrageniculate popliteal artery. (b) A 6-mm diameter polytetrafluoroethylene vascular graft is now sewn on to the vein collar to complete the lower anastomosis of a femoropopliteal bypass.

Other operations

In the presence of occlusive disease affecting the superficial femoral artery, the profunda femoris artery is the chief collateral channel between the iliac and popliteal arterial systems. In such circumstances, if the profunda is itself compromised, distal ischaemia is increased. Atheroma in the profunda shows a predilection for the area near its origin and reconstruction can be achieved by endarterectomy, by patch angioplasty (profundaplasty) (Fig. 15), or by a combination of both. Such surgery may be carried out in association with inflow reconstruction, such as an aortofemoral bypass in which a graft limb is extended beyond the common femoral artery into the profunda femoris. In general a femorodistal bypass, if feasible, will produce a better result than a profunda reconstruction alone.

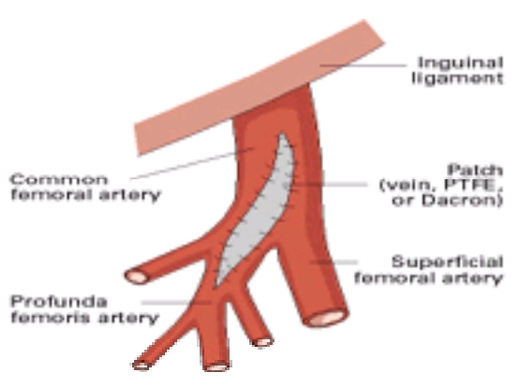


Fig. 15. Patch angioplasty of the origin of the profunda femoris artery (profundaplasty).

Apart from the profunda femoris artery, the only other artery in the leg which is at all frequently managed by endarterectomy is the common femoral. Likewise, patch angioplasty is rarely used at other native artery sites, although it has become fashionable for the relief of stenosis at graft–artery anastomoses and for the treatment of vein graft strictures.

Lumbar sympathectomy

The first lumbar sympathectomy was performed in 1924 and the operation has had variable popularity ever since. The advent of reconstructive surgery has put sympathectomy firmly in second place as a method of treatment of occlusive disease and at present its use can only be recommended as a last resort, for severe ischaemia when arterial bypass or reconstruction is not possible.

If either the collateral or the microcirculation is diseased, sympathectomy is unlikely to be of benefit. Its aim is to excise via a retroperitoneal route the sympathetic chain and ganglia from L4 to L2 (sometimes including L1), thereby increasing blood flow in the limb by decreasing vasomotor tone. The mechanism and effects of sympathectomy have been the subject of discussion for years, and they are complex; it is likely, however, that in suitable candidates the blood flow in the skin capillaries is increased. In contrast, the effect on the blood flow in the skeletal muscle is controversial. Whatever the physiological effect, sympathectomy is extremely unlikely to benefit patients with claudication.

Today, operative sympathectomy is hardly ever warranted. Rather than submit patients to open surgery it is more reasonable to perform a chemical sympathectomy by injecting 5 ml of 1:15 aqueous phenol at two separate sites along the lumbar sympathetic chain (Fig. 16). This requires only local anesthetic and for many years was performed ‘blind', although today the placement of needles should be monitored radiologically. In experienced hands a chemical sympathectomy is as effective as an operative procedure. A recent inovation is to perform an operative lumbar sympathectomy using a minimally invasive endoscopic approach, employing balloon expansion of the retroperitoneal ‘potential space'. It is likely that this represents an improved means to an unimproved end.

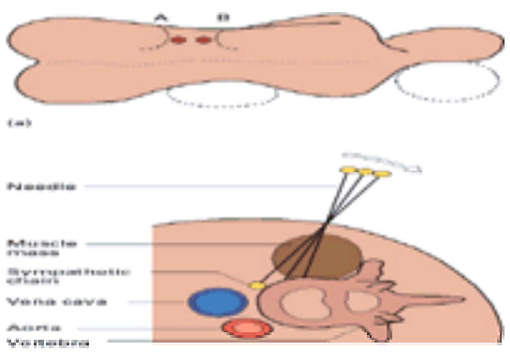


Fig. 16. Chemical sympathectomy. (a) Patient position. A, iliac crest; B, costal margin; ○ represents the two sites for injection. (b) Transverse view. Needles advanced to vertebral body then gradually swung until tangential to vertebral body at which point, after careful checking, injection may be made.

Complications

Early complications

Graft thrombosis

This may be due to inadequate inflow or, more likely, outflow which has not been recognized before surgery. It may also be due to a technically poor anastomosis, especially at the distal end of the graft. A balloon catheter thrombectomy or intraluminal thrombolysis is only likely to be successful in the long term if the underlying cause of the thrombosis is also corrected.

Haemorrhage

This is usually due to technical misadventure and the insertion of further sutures into an anastomosis may occasionally be needed. Very rarely it may herald an early anastomotic infection.

Lymph leak

This typically occurs from groin wounds, although wounds placed more distally in the leg are not immune. Most leaks stop spontaneously. Underlying grafts should be protected from the entry of external bacteria by administration of a broad-spectrum antibiotic for the duration of leakage. Very rarely surgical closure of leaking lymph channels is necessary.

Lymphocele

This too typically occurs at groin level. It is due to leakage within the tissues from lymphatic channels which have been severed at the time of operation. The swelling that presents is non-pulsatile (which distinguishes it from false aneurysm) and may be diagnosed by duplex scan or needle aspiration. If aspiration (of clear or slightly yellow fluid) is chosen, this should be carried out on only one occasion, for diagnositic purposes. Repeated aspiration is likely to introduce infection and prolong the duration of the swelling. The correct management is to await spontaneous resolution.

Oedema

This is not uncommon after bypass surgery in the leg and is often due to cell swelling and an increase in interstitial fluid in reperfused critically ischaemic limbs. Oedema may also be due to lymphatic hold-up, which may be expected to subside rapidly without active intervention. Finally, leg swelling may be due to relative immobility after surgery or to deep vein thrombosis.

Infection

Wound infection is uncommon and occurs most commonly in the groin wound. It is usually superficial. The underlying graft should be protected by administration of an appropriate antibiotic. If the anastomosis or graft becomes infected, the material used for reconstruction will generally need to be removed and a new reconstruction, by an alternative clean route, attempted. This may be difficult to achieve.

Late complications

Thrombosis

This is the main cause of late graft failure and is generally due to progression of atheromatous disease, either proximally or distally. It may also be due to a cellular build-up at the anastomotic site of the graft (neointimal hyperplasia). Strictures may form in vein grafts and compromise flow, finally causing thrombosis. Such strictures are most common within the first year after graft insertion and should be recognized early by regular follow-up so that they may be repaired before thrombosis occurs. This requires either percutaneous transluminal angioplasty or a patch graft.

False or anastomotic aneurysm

When a leak occurs at an anastomosis the fluid may remain within the tissues and form a compartment contiguous with the artery/graft. Blood may flow freely in such a compartment and blood pressure will increase its size with time, forming a false aneurysm. An anastomotic aneurysm is a true aneurysm—it has an intimal lining—and reflects failing strength of the vessel wall at the site of anastomosis. Such aneurysms are more common after infection and may themselves be infected. They are also more common in patients with aneurysmal disease at other sites. Insertion of a graft which is too short, making it pull needlessly on its attached vessels, is a contributory factor in some cases.

Graft degeneration

A good quality autogenous saphenous vein of adequate calibre should rarely degenerate when used as an arterial conduit. Similarly, PTFE has been used in the leg for many years without major problems; this is probably true also for Dacron prostheses. The human umbilical vein graft may be prone to aneurysmal degeneration and

other newer grafts—especially those of animal origin—must be regarded with caution at this time.

Prevention of graft occlusion

Patients who have undergone angioplasty or reconstruction should be subjected to regular follow-up. Occlusion of grafts or angioplasty sites is common and most likely to occur within a year. Although early graft occlusion—within 30 days of insertion—is usually due to technical error and will occur before graft surveillance has started, intimal hyperplasia and fibrotic stricture may be identified in the medium term prior to occlusion. Patency may be prolonged and the life of a graft extended if problems can be identified and corrected by percutaneous transluminal angioplasty or surgery before occlusion; correction after occlusion has occurred is far less likely to be successful.

The form which surveillance should take has been widely discussed. Simple clinical review of symptoms and pulses does not give adequate warning of an impending occlusion. Serial assessment of the ankle/brachial pressure index is easy, non-invasive, and is an effective screening test; a decrease in the index of more than 0.15 suggests that a stenosis is developing. However, even this test will fail to identify graft stenosis before occlusion in some patients. Angiography, especially using digital subtraction techniques, has been suggested as the most effective form of screening, but is both invasive and expensive. Duplex ultrasonography is cheaper and gives an adequate non-invasive assessment of graft function. If the scan shows a suspicious area, particularly a graft segment where the velocity of flow is very high, angiography can be used to confirm the abnormality and allow planning of corrective treatment. The best follow-up at this time entails a duplex scan at 6 and 12 weeks after intervention and then at 3-month intervals to 1 year. The gain from follow-up beyond 1 year is likely to be limited.

Although graft surveillance is important it does not replace the traditional management of a vascular graft. This involves stopping the patient smoking, which significantly improves graft patency, correcting polycythaemia and hyperlipidaemia, and treating diabetes mellitus. Most surgeons do not routinely prescribe long-term antiplatelet or anticoagulant drugs as the evidence that they benefit graft patency is small. The exception perhaps is low-dose aspirin (150 or 300 mg/day) for patients with fabricated grafts below the inguinal ligament.

Further reading

Antiplatelet Trialists' Collaboration. Collaborative overview of randomized trials of antiplatelet therapy—II: Maintenance of vascular graft or arterial patency by antiplatelet therapy. *British Medical Journal* 1994; **308**: 159–68. [Evidence for and against antiplatelet agents after graft insertion.]

Bloor K. Natural history of arteriosclerosis of the lower extremities. *Annals of the Royal College of Surgeons of England* 1961; **28**: 36–8. [Classic study from an era before the massive expansion of vascular intervention.]

Cotton LT, Cross FW. Lumbar sympathectomy for arterial disease. *British Journal of Surgery* 1985; **72**: 678–83. [Extensive review article on sympathectomy.]

Currie IC *et al.* Pulse generated run-off versus dependent Doppler ultrasonography for assessment of calf vessel patency. *British Journal of Surgery* 1994; **81**: 1448–50. [Useful cheap non-invasive tests compared.]

Dormandy J. European consensus on critical limb ischaemia. *Lancet* 1989; **334**: 737–8. [Consensus document about an important concept.]

Dotter CT, Judkins MP. Transluminal treatment of arteriosclerotic obstruction. Description of a new technic and a preliminary report of its application. *Circulation* 1964; **30**: 654–70. [Classic paper from the originators of endoluminal angioplasty.]

Earnshaw JJ, Comerota AJ. Towards international consensus in peripheral arterial thrombolysis. *British Journal of Surgery* 1997; **84**: 1332–3. [International overview of lytic therapy.]

Edmonson RA *et al.* Low molecular weight heparin versus aspirin and dipyridamole after femoropopliteal bypass grafting. *Lancet* 1994; **344**: 914–18. [Evidence for and against anticoagulant/antiplatelet agents after graft insertion.]

Gardener AW, Poehlman ET. Exercise rehabilitation programs for the treatment of claudication pain. *Journal of the American Medical Association* 1995; **274**: 975–80. [Meta-analysis of exercise rehabilitation programs for claudication.]

Gruntzig AR, Hopff H. Percutane Recanalisation chronischer arterieller Verschlusse mit einemneuen Dilatationskatheter. *Deutsche Medizinische Wochenschrift* 1974; **99**: 2502–5. [Classic paper from the popularizers of balloon angioplasty.]

Kannel WB, Shurtleff D. The Framingham study: cigarettes and the development of intermittent claudication. *Geriatrics* 1973; **28**: 61–8. [Large prospective study relating smoking and occlusive arterial disease.]

Kunlin J. Le traitemente de l'arterie obliterante par la greffe veineuse. *Archives des Maladies du Coeur et des Vaisseaux* 1949; **42**: 371. [Classic description of femoropopliteal bypass.]

London NMS, Varty K, Sayers RD, Thompson MM, Bell PRF, Bolia A. Percutaneous transluminal angioplasty for lower limb critical ischemia. *British Journal of Surgery* 1995; **82**: 1232–5. [Consideration of the advantages of the endoluminal technique.]

Lundgren F, Dahllof A-G, Lundholm K, Schersten T, Volkman B. Intermittent claudication—surgical reconstruction or physical training? A prospective randomized trial of treatment efficacy. *Annals of Surgery* 1989; **209**: 346–55. [Major randomized trial of surgery and exercise therapy.]

Moody AP, Edwards PR, Harris PL. *In situ* versus reversed femoropopliteal vein grafts: long-term follow-up of a prospective, randomized trial. *British Journal of Surgery* 1992; **79**: 750–2. [Large comparison of the results of the two commonest vein bypass techniques.]

Perkins JMT, Collin JC, Creasy TS, Fletcher EW, Morris PJ. Exercise training versus angioplasty for stable claudication. Long and medium term results of a prospective randomized trial. *European Journal of Vascular and Endovascular Surgery* 1996; **11**: 409–13. [Major randomized trial of angioplasty and exercise therapy.]

Tyrell MR *et al.* Experimental evidence to support the use of interposition vein collars/patches in distal PTFE anastomosis. *European Journal of Vascular and Endovascular Surgery* 1990; **4**: 95–101. [Laboratory evidence in favor of vein interposition.]

Whyman MR *et al.* Is intermittent claudication improved by percutaneous transluminal angioplasty? A randomized controlled trial. *Journal of Vascular Surgery* 1997; **26**: 551–7. [Major randomized trial of angioplasty and conventional medical therapy.]

17.5.3 Mesenteric arteries

Peter H. Lin and Thomas F. Dodson

Introduction
Clinical presentation
Diagnostic studies
Treatment
- Acute embolic mesenteric ischemia
- Acute thrombotic mesenteric ischemia
- Chronic mesenteric ischemia
- Non-occlusive mesenteric ischemia
- Celiac artery compression syndrome
- Other causes of mesenteric ischemia

Further reading

Introduction

Obliterative atheromatous disease of the mesenteric arteries usually occurs in elderly individuals who are medically debilitated with generalized atherosclerosis. Despite recent progress in perioperative management and better understanding in pathophysiology, mesenteric ischemia is one of the most lethal vascular disorders with mortality rates ranging between 50 and 75 per cent. Delay in diagnosis and treatment are the main contributing factors in its high mortality. It is estimated that mesenteric ischemia accounts for 1 in every 1000 hospital admissions in the United States. The prevalence is rising due in part to the increased awareness of this disease, the advanced age of the population, and the significant comorbidity of these elderly patients. Early recognition and prompt treatment before the onset of irreversible intestinal ischemia are essential to improve the outcome.

There are four major syndromes of visceral ischemia involving the mesenteric arteries, which include: (i) acute embolic mesenteric ischemia, (ii) acute thrombotic mesenteric ischemia, (iii) chronic mesenteric ischemia, and (iv) non-occlusive mesenteric ischemia. Despite the variability of these syndromes, a common anatomic pathology is involved in these processes. The superior mesenteric artery is the most commonly involved vessel in acute mesenteric ischemia. Acute thrombotic mesenteric ischemia frequently occurs in patients with underlying mesenteric atherosclerosis, which usually involves the origin of the mesenteric arteries while sparing the collateral branches. The development of collateral vessels is possible when the occlusive process is gradual rather than a sudden ischemic event. In acute embolic mesenteric ischemia, the emboli typically originate from a cardiac source and frequently occur in patients with atrial fibrillation or following myocardial infarction. Non-occlusive mesenteric ischemia is characterized by a low flow state in otherwise normal mesenteric arteries. In contrast, chronic mesenteric ischemia is a functional consequence of a long-standing atherosclerotic process which involves the main mesenteric vessels: the celiac, superior mesenteric, and inferior mesenteric arteries.

Several less common syndromes of visceral ischemia involving the mesenteric arteries can also cause serious debilitation. Chronic mesenteric ischemic symptoms can occur due to extrinsic compression of the celiac artery by the diaphragm, which is termed ‘the median arcuate ligament syndrome'. Acute visceral ischemia may occur following aortic operation, due to ligation of the inferior mesenteric artery in the absence of adequate collateral vessels. Furthermore, acute visceral ischemia may develop in aortic dissection which involves the mesenteric arteries. Finally, other unusual causes of ischemia include mesenteric arteritis, radiation arteritis, and cholesterol emboli.

Clinical presentation

Abdominal pain out of proportion to physical findings is the classic presentation in patients with acute mesenteric ischemia and occurs frequently following an embolic or thrombotic ischemic event of the superior mesenteric artery. A high index of suspicion is crucial in the diagnosis of acute mesenteric ischemia. Clinical manifestations may include sudden onset of abdominal cramps in patients with underlying cardiac or atherosclerotic diseases. The abdominal pain is often associated with bloody diarrhea, as a result of mucosal sloughing secondary to ischemia. Fever, diarrhea, nausea, vomiting, and abdominal distention are some common but non-specific manifestations. Diffuse abdominal tenderness, rebound, and rigidity are ominous signs and usually herald bowel infarction.

Symptoms of thrombotic mesenteric ischemia may initially be more insidious than those of embolic mesenteric ischemia. Approximately 70 per cent of patients with chronic mesenteric ischemia have a history of abdominal angina. In these patients, the chronicity of mesenteric atherosclerosis is important as it permits collateral vessel formation. The precipitating factor leading to chronic mesenteric occlusion is often due to an unrelated illness which results in dehydration, such as diarrhea or vomiting, which may further confuse the actual diagnosis. If the diagnosis is not recognized promptly, symptoms may worsen and lead to progressive abdominal distension, oliguria, increasing fluid requirements, and severe metabolic acidosis.

Abdominal pain is only present in approximately 70 per cent of patients with non-occlusive mesenteric ischemia. When present, the pain is usually severe but may vary in location, character, and intensity. In the absence of abdominal pain, progressive abdominal distention with acidosis may be an early sign of ischemia and impending bowel infarction. The diagnosis of non-occlusive mesenteric ischemia should be considered in elderly patients with sudden abdominal pain who have any of the following risk factors: congestive heart failure, acute myocardial infarction with cardiogenic shock, hypovolemic or hemorrhagic shock, sepsis, pancreatitis, and administration of digitalis or vasoconstrictor agents such as epinephrine.

Diagnostic studies

Laboratory evaluation is neither sensitive nor specific in the diagnosis of mesenteric ischemia. Complete blood count may reveal hemoconcentration and leukocytosis. Metabolic acidosis develops as a result of anaerobic metabolism. Elevated serum amylase and lactate levels are non-specific findings. Hyperkalemia and azotemia may occur in the late stage of ischemia. Plain abdominal radiographs may provide helpful information to exclude other abdominal pathologies such as bowel perforation, obstruction, or volvulus, which may exhibit symptoms mimicking intestinal ischemia. Radiographic appearance of an adynamic ileus with a gasless abdomen is the most common finding in patients with acute mesenteric ischemia.

The definitive diagnosis of mesenteric thrombosis is made by biplanar mesenteric arteriography, which should be performed promptly in any patients with suspected mesenteric occlusion. It typically shows occlusion or near-occlusion of the celiac and superior mesenteric arteries at or near their origins from the aorta. In most cases, the inferior mesenteric artery has been previously occluded secondary to diffuse infrarenal aortic atherosclerosis. The differentiation of the four types of mesenteric arterial occlusion may be suggested with biplanar mesenteric arteriogram. Mesenteric emboli typically lodge at the orifice of the middle colic artery, which creates a ‘miniscus sign' with an abrupt cut-off of a normal proximal superior mesenteric artery several centimeters from its origin on the aorta. Mesenteric thrombosis, on the other hand, occurs at the most proximal superior mesenteric artery which tapers off at 1 to 2 cm from its origin. In the case of chronic mesenteric occlusion, the appearance of collateral circulation is usually present (Fig. 1). Non-occlusive mesenteric ischemia produces an arteriographic image of segmental mesenteric vasospasm with a main superior mesenteric artery trunk of relatively normal appearance.

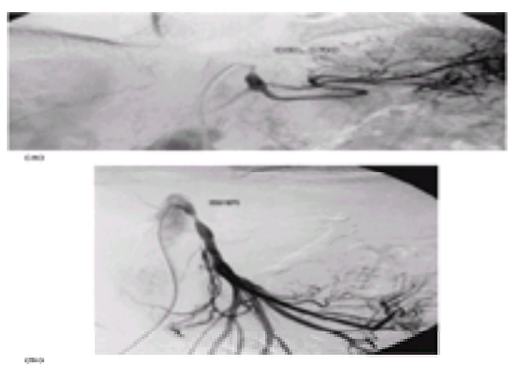


Fig. 1. (a) Selective celiac arteriogram in a 54-year-old patient with intermittent chronic abdominal pain with weight loss. It demonstrated a stenosis in the proximal celiac artery. (b) Selective arteriogram of the superior mesenteric artery in the same patient also demonstrated a tight stenosis in the proximal superior mesenteric artery. The development of collateral circulation is present in the superior mesenteric artery.

Mesenteric arteriography can also play a therapeutic role. Once the diagnosis of non-occlusive mesenteric ischemia is made on the arteriogram, an infusion catheter can be placed at the superior mesenteric artery orifice and vasodilating agents, such as papaverine, can be administered intra-arterially. The papaverine infusion may be continued postoperatively to treat persistent vasospasm, a common occurrence following mesenteric reperfusion. Transcatheter thrombolytic therapy has little role in the management of thrombotic mesenteric occlusion. Although thrombolytic agents may transiently recannulate the occluded vessels, the underlying occlusive lesions require definitive treatment. Furthermore, thrombolytic therapy typically requires a prolonged period of time to restore perfusion and the intestinal viability may be difficult to assess.

Treatment

Initial management of patients with acute mesenteric ischemia includes fluid resuscitation and immediate systemic anticoagulation with heparin sulfate to prevent further thrombus propagation. Significant metabolic acidosis should be corrected with sodium bicarbonate if possible. A central venous catheter, peripheral arterial catheter, and a Foley catheter should be placed for fluid resuscitation and hemodynamic status monitoring. Appropriate antibiotics are given prior to surgical exploration. The operative management of acute mesenteric ischemia is dictated by the cause of the occlusion. It is helpful to obtain a preoperative mesenteric arteriogram to confirm the diagnosis and plan appropriate treatment options. However, the diagnosis of mesenteric ischemia frequently cannot be established prior to surgical exploration and therefore patients in a moribund condition with acute abdominal symptoms should undergo immediate surgical exploration avoiding the delay required to perform an arteriogram.

Acute embolic mesenteric ischemia

The primary goal of surgical treatment in embolic mesenteric ischemia is to restore arterial perfusion with removal of the embolus from the vessel. The abdomen is explored through a midline incision, which often reveals variable intestinal ischemia from the mid-jejunum to the ascending or transverse colon. The transverse colon is lifted superiorly, and the small intestine is reflected toward the right upper quadrant. The superior mesenteric artery is approached at the root of the small bowel mesentery, usually as it emerges from beneath the pancreas to cross over the junction of the third and fourth portions of the duodenum. Alternatively, the superior mesenteric artery can be approached by incising the retroperitoneum lateral to the fourth portion of the duodenum, which is rotated medially to expose the superior mesenteric artery. Once the proximal superior mesenteric artery is identified and controlled with vascular clamps, a transverse arteriotomy is made to extract the embolus using standard balloon embolectomy catheters. In the event where the embolus has lodged more distally, exposure of the distal superior mesenteric artery may be obtained in the root of the small bowel mesentery by isolating individual jejunal and ileal branches to allow a more comprehensive thromboembolectomy. Following the restoration of superior mesenteric artery flow, an assessment of intestinal viability must be made, and non-viable bowel must be resected. Several methods have been described to evaluate the viability of the intestine, which include intraoperative intravenous fluorescein injection and inspection with a Wood's lamp and Doppler assessment of antimesenteric intestinal arterial pulsations. A second-look procedure should be considered in many patients and is performed 24 to 48 h following embolectomy. The goal of a second-look procedure is to reassess the extent of bowel viability, which may not be obvious immediately following the initial embolectomy. If non-viable intestine is evident in the second-look procedure, additional bowel resections should be performed at that time.

Acute thrombotic mesenteric ischemia

The treatment of thrombotic mesenteric ischemia differs from that of mesenteric embolism due to the nature of the superior mesenteric artery. In embolic mesenteric ischemia, the superior mesenteric artery itself is otherwise normal and thromboembolectomy will usually suffice to restore mesenteric circulation. However, thrombotic mesenteric ischemia usually involves a severely atherosclerotic vessel, typically in the proximal superior mesenteric artery. Therefore, these patients require a reconstructive procedure to the distal superior mesenteric artery to bypass the proximal occlusive lesion and restore adequate mesenteric flow. Exploration in patients with thrombotic mesenteric ischemia often reveals a much more extensive intestinal necrosis than in patients with embolic mesenteric ischemia, due to the extensive atherosclerotic process. The saphenous vein is the graft material of choice, and prosthetic materials should be avoided in patients with non-viable bowel because of the risk of bacterial contamination if resection of necrotic intestine is performed. The bypass graft may originate from either the aorta or iliac artery. The supraceliac infradiaphragmatic aorta offers several advantages as the origin of the graft when compared with the infrarenal aorta. The supraceliac aorta is often devoid of atherosclerotic plaque, which minimizes the potential embolic complications associated with clamping of the frequently calcified infrarenal aorta. In addition, revascularization from the supraceliac aorta to the distal superior mesenteric artery permits an antegrade graft placement, which is less prone to kinking when the small intestine is returned to its normal location following the bypass procedure.

Chronic mesenteric ischemia

The therapeutic goal in patients with chronic mesenteric ischemia is to revascularize the mesenteric circulation and prevent the development of bowel infarction. Mesenteric occlusive disease can be treated successfully by either transaortic endarterectomy or mesenteric artery bypass. Transaortic endarterectomy is indicated for ostial lesions of patent celiac and superior mesenteric arteries. This can be accomplished by a left medial visceral rotation to expose the aorta and its mesenteric branches. A lateral aortotomy is performed encompassing both the celiac and superior mesenteric arteries. The visceral arteries must be adequately mobilized so that the termination site of endarterectomy can be visualized. Otherwise, an intimal flap may develop which can lead to early thrombosis or distal embolization.

For occlusive lesions located 1 to 2 cm distal to the mesenteric origin, mesenteric artery bypass should be performed. Multiple mesenteric arteries are typically involved in chronic mesenteric ischemia, and both the celiac and superior mesenteric arteries should be revascularized whenever possible. In general, bypass grafting may be performed either antegrade from the supraceliac aorta or retrograde from either the infrarenal aorta or iliac artery. Both autogenous saphenous vein grafts and prosthetic grafts have been used with satisfactory and equivalent success. An antegrade bypass can also be performed using a small-caliber bifurcated graft from the supraceliac aorta to both the celiac and superior mesenteric arteries, which yields an excellent long-term result.

Non-occlusive mesenteric ischemia

The treatment of non-occlusive mesenteric ischemia is primarily pharmacologic with selective mesenteric arterial catheterization followed by infusion of vasodilatory agents, such as tolazoline or papaverine. Once the diagnosis is made on the mesenteric arteriography, intra-arterial papaverine is given at a dose of 30 to 60 mg/h. This must be coupled with the cessation of other vasoconstricting agents. Concomitant intravenous heparin should be administered to prevent thrombosis in the cannulated vessels. Treatment strategy thereafter is dependent on the patient's clinical response to the vasodilator therapy. If abdominal symptoms improve, mesenteric arteriography should be repeated to document the resolution of vasospasm. The patient's hemodynamic status must be carefully monitored during papaverine infusion as significant hypotension can develop in the event that the infusion catheter migrates into the aorta and leads to systemic circulation of papaverine. Surgical exploration is indicated if the patient develops signs of continued bowel ischemia or infarction as evidenced by rebound tenderness or involuntary guarding. In these circumstances, papaverine infusion should be continued intraoperatively and postoperatively. The operating room should be kept as warm as possible, and warm irrigation fluid and laparotomy pads should be used to prevent further intestinal vasoconstriction during exploration.

Celiac artery compression syndrome

Abdominal pain due to narrowing of the origin of the celiac artery may occur as a result of extrinsic compression or impingement by the median arcuate ligament. This condition is known as celiac artery compression syndrome or median arcuate ligament syndrome. The celiac artery compression syndrome has been implicated in some variants of chronic mesenteric ischemia. Most patients are young females between 20 and 40 years of age. Abdominal symptoms are non-specific but the pain is localized in the upper abdomen and may be precipitated by meals. The treatment goal is to release the ligamentous structure that compresses the proximal celiac artery and correct any persistent stricture by bypass grafting.

Other causes of mesenteric ischemia

While atherosclerotic occlusion of the main mesenteric arteries accounts for the majority of cases of chronic intestinal ischemia, other conditions also exist which can compromise mesenteric circulation and should be considered in the differential diagnosis. They include radiation arteritis, polyarteritis nodosa, lupus erythematosus, Kawasaki's disease, and fibromuscular dysplasia. Patients who smoke heavily or young women who are taking oral contraceptives may be at risk of developing intimal hyperplasia of the visceral arteries which can also lead to mesenteric ischemia.

Further reading

Bech FR, Gewertz BL. Pathophysiology of intestinal ischemia. In: Strandness DE Jr, van Breda A, eds, *Vascular diseases: surgical and interventional therapy*, pp. 745–53. Churchill Livingstone, New York, 1994. [This issue reviews the pathophysiology of visceral ischemia as well as operative and radiological management of intestinal ischemia.]

Beebe HG, MacFarlane S, Raker EJ. Supraceliac aortomesenteric bypass for intestinal ischemia. *Journal of Vascular Surgery* 1987; **5**: 749–54. [The authors described their experience of supraceliac aortomesenteric bypass in the management of visceral ischemia.]

Cherry RD, Jabbari M, Goresky CA, Herba M, Reich D, Blundell PE. Chronic mesenteric vascular insufficiency with gastric ulceration. *Gastroenterology* 1986; **91**: 1548–52. [This text on chronic mesenteric ischemia highlights the clinical significance of the disease including its complication of gastric ulceration.]

Cunningham CG *et al.* Chronic visceral ischemia: three decades of progress. *Annals of Surgery* 1991; **214**: 276–81. [This text reports a single-institutional experience of chronic visceral ischemia, including its diagnosis and treatment.]

Marston A, Clarke JM, Garcia-Garcia J, Miller AL. Intestinal function and intestinal blood supply. *Gut* 1985; **26**: 656–66. [A comprehensive review of the function and vascular anatomy of the intestine.]

Moneta GL, Lee RW, Yeager RA, Taylor LM, Porter JM. Mesenteric duplex scanning: a blinded prospective study. *Journal of Vascular Surgery* 1993; **17**: 79–84. [This report details the use of duplex ultrasound in a prospective blinded study in the detection of mesenteric ischemia.]

17.6.1 Abdominal aorta

Jack Collin

Definition
Epidemiology
Prevalence
Aetiology
Genetic predisposition
Connective tissue degradation
Environmental potentiation
Natural history
Very small aneurysms (less than 4.0 cm diameter)
Small aneurysms (4.0–5.9 cm diameter)
Large aneurysms (greater than 6.0 cm diameter)
Clinical presentation
Symptoms and signs of aortic rupture
Management
Ruptured abdominal aortic aneurysm
Asymptomatic abdominal aortic aneurysm
'Inflammatory' abdominal aortic aneurysm
Investigation of the aneurysm
Investigation of the patient
Operative mortality and morbidity
Operative technique
Further reading

Definition

An aneurysm is by definition an abnormal dilatation of an artery or vein, and the application of this general principle to the abdominal aorta has seldom presented any problems in routine clinical practice. The universal use of abdominal ultrasonography as a basic diagnostic tool, and particularly the introduction of screening programmes for abdominal aortic aneurysms, has recently highlighted the need for a more precise definition to allow appropriate diagnosis of the many marginal aortic dilatations, or small aneurysms, that are now being discovered.

The diameter of the abdominal aorta, like other biological measurements, conforms to a normal population distribution curve. Median aortic diameter increases with age, is greater in men than in women, and is influenced by race, weight, height, and the prevalence of hypertension in the population studied. In men, after 60 years of age, the shape of the curve becomes increasingly skewed to the right as the prevalence of abdominal aortic aneurysm increases. A definition of aneurysm based on deviation from the mean aortic diameter for a population will therefore be of less value than one based on the median diameter.

In some patients with an abdominal aortic aneurysm, dilatation may also involve the suprarenal or thoracic aorta, leading to a diagnosis of thoracoabdominal aortic aneurysm or generalized arterial ectasia, depending on the amount and extent of the dilatation. Any definition that relies solely on comparison of the diameter of a suspected aneurysm with that of adjacent 'normal' aorta will fail in the 5 per cent of patients whose aneurysms are not confined to the infrarenal aorta.

Measurement of aneurysm diameter is commonly made from ultrasonograms, computed tomograms (**CT**) or magnetic resonance images (**MRI**); diameters can be recorded as anteroposterior, transverse, or the maximum diameter any plane. Since aneurysmal aortas elongate as well as dilate, obliquity of the long axis of the aneurysm is common. In consequence, transverse and any-plane diameters measured by MRI and CT are usually greater than anteroposterior diameters and always unreliable. Anteroposterior diameters measured by ultrasonography are usually less than those measured by CT and MRI, and correspond more closely to true diameter obtained by direct measurement at operation.

A modern definition of an aortic aneurysm would take into account the above facts and make due allowance for the inaccuracies of measurement inherent in even the most precise methods of diagnosis. The following definition is proposed: An aortic aneurysm is present when the maximum external anteroposterior diameter of the aorta either (1) is at least 4.0 cm or (2) exceeds the diameter of the adjacent aorta by at least 0.5 cm.

The imprecise but clinically useful term 'aortic ectasia' should be reserved for those cases where the aorta appears abnormally wide but is not aneurysmal by the above definition.

In practice an aneurysm is diagnosed by most clinicians when the aortic diameter is 3.0 cm or more. This pragmatic approach, which takes no account of sex, height, race, age, body morphology, modality or method of measurement, at least has the virtue of simplicity if not reproducibility or validity.

Epidemiology

Aortic aneurysms are very rare in people under the age of 55 years and before that age are virtually confined to patients with Marfan, Ehlers–Danlos, or arteria magna syndromes. The common idiopathic abdominal aortic aneurysm is largely a disease of elderly men. Comparison of fatalities from ruptured abdominal aortic aneurysm by age and sex show that deaths are 13 times more common in men than in women at age 60 to 65 years, but over 80 years of age are only four times more common in men than in women (Table 1). At age 85 almost three times as many women as men are still alive, so among the very elderly the numbers of men and women presenting with ruptured aneurysms are similar. The changing pattern of presentation with age, combined with an increase in the number of elderly people in the populations of most wealthy nations, has led some surgeons to conclude erroneously that abdominal aortic aneurysm has increased in incidence disproportionately in women.

Age (years)	*M:F ratio*
60–64	12.6
65–69	9.1
70–74	7.0
75–79	6.5
80 and over	4.0

Table 1 Male:female ratio of incidence of death from ruptured abdominal aortic aneurysm (England and Wales, 1986)

The annual risk of death from aortic aneurysm increases from $125/10^6$ for men aged 55 to 59 years to $2728/10^6$ at age 85 and over. At age 70 to 74 years, aortic aneurysms are responsible for 2.2 per cent of all deaths in men, and abdominal aortic aneurysms account for 77 per cent of these (Fig. 1). The disease is a particularly common cause of unexpected deaths, and more than 5 per cent of sudden deaths investigated by autopsy in men over 50 years of age are found to be due to ruptured abdominal aortic aneurysm.

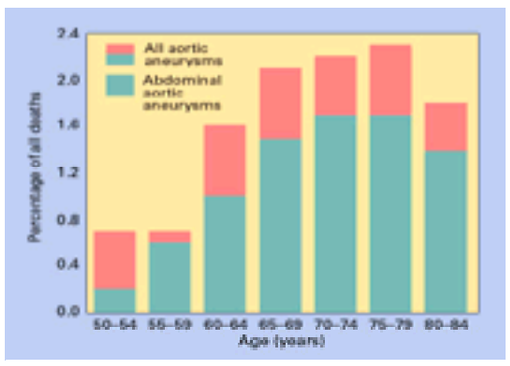


Fig. 1. Death from aortic aneurysm as a percentage of all deaths for men by age.

In the past 30 years there has been a linear increase in the number of recorded deaths from aortic aneurysm in England and Wales. In part this can be explained by the progressive growth in the number of elderly people in the population, but age-specific death rates for aortic aneurysm have also increased over the age of 60 years. Some of the increase might be real, but much of the apparent change is due to enhanced awareness of the disease, improved diagnosis, and altered referral patterns associated with the establishment of specialist vascular units. This issue is difficult to resolve because of the unreliability of national records of cause of death, which ultimately depend on the diagnostic acumen of the reporting doctor, seldom supported by autopsy evidence. In the United States of America a similar, but more rapid, increase in age-adjusted mortality for aortic aneurysm occurred between 1951 and 1968, but the number of recorded deaths then stabilized and, in white ('Caucasian') males, has declined since 1976, possibly as a consequence of the increasing impact of elective surgery for aneurysm.

Reports from the early part of the twentieth century illustrate that there has been a qualitative as well as a quantitative change in abdominal aortic aneurysmal disease. Before the introduction of effective antisyphilitic therapy, the majority of aortic aneurysms were a manifestation of tertiary syphilis, and the mean age of presentation was consequently lower than at present. In developed countries, syphilitic aneurysms are now rare and idiopathic abdominal aortic aneurysms of elderly people represent the vast majority of cases seen.

There has been debate about the influence of racial factors on abdominal aortic aneurysms, with most studies focusing on comparison of black and white populations in the United States and South Africa. There is no doubt that the disease is seen less often in people of African descent than in whites, but in both countries it is difficult to discover whether the differences simply reflect the lower life expectation and mean age of the black population and their poorer access to health care. The high prevalence of hypertension in black (African origin) people gives theoretical grounds for suspecting that, age for age, abdominal aortic aneurysm might well be more common than in whites. At present the evidence for racially determined differences in incidence must be regarded as suspect, and careful epidemiological studies will be required to resolve this issue.

Prevalence

The prevalence of any disease represents the total number of cases, both diagnosed and occult, present in the population at any given time, and should be distinguished from incidence, which is the number of new cases diagnosed over a specified period of time.

The majority of aortic aneurysms are asymptomatic and impalpable; consequently their prevalence in the community can be determined only by systematic screening. The results of many screening studies have been published, but most are fundamentally flawed as measures of prevalence. The incidence of aortic aneurysm increases rapidly with age and is much higher in men than in women. Prevalence studies must therefore differentiate each 5- or 10-year age group and men from women. Studies in patients with hypertension, atherosclerosis, or other diseases cannot produce prevalence data relevant to the whole population. Examination of the data (Table 2) does, however, allow a number of conclusions to be drawn.

	City	Study group	No. screened	Aneurysm definition	Prevalence (%)
Healthy individuals	Oxford	Men 65–74 years	850	> 4.0 cm or 5 mm > segmental aorta	5.7
	Chichester	Men and women 65–80 years	1202	> 3.0 cm	5.8
Hypertensives	London	Men > 50 years	500	> 3.0 cm	7.0
	Birmingham	Men 65–85 years	94	5 mm > segmental aorta	7.4
		Women 65–85 years	71		2.8
Atherosclerosis/ hypertension/ obesity	Portland	Men and women > 50 years, attending cardiology clinic	120	> 4.0 cm	5.0
	Vancouver	Men > 55 years, waist measure > 100 cm, hypertensive or atherosclerotic	65	> 3.0 cm	13.3
	Minneapolis	Men 60–75 years with hypertension and/or coronary heart disease	201	> 1.5 times segmental aortic diameter	9.0
	Nottingham	Men and women attending vascular clinic	104	> 3.5 cm or 5 mm > segmental aorta	7.7

Table 2 Screening studies for abdominal aortic aneurysms

1. In men aged 65 to 74 years the prevalence of abdominal aortic aneurysm of all sizes is around 5.5 per cent, and of diameter 4.0 cm or more is around 2.0 per cent.
2. In patients with hypertension or atherosclerotic occlusive disease of the coronary, carotid, or limb arteries, the prevalence of abdominal aortic aneurysm is 50 per cent higher than in the general population.

Aetiology

The common abdominal aortic aneurysm of elderly men has been labelled as 'atherosclerotic'. This classification has little justification, has paralysed thinking, and needs to be re-examined. It is interesting to note that aneurysmal disease is encumbered by more than itsfair share of unhelpful, or frankly misleading, descriptive terms, among which are atheromatous, mycotic, inflammatory, dissecting, and arteriovenous aneurysms. In the elderly individual the aorta, in common with every other artery, will have obvious features of atherosclerosis but this is not enough evidence to make credible a pathological diagnosis that does not fit with many known facts about the disease.

Tilson has compared patients with abdominal aortic aneurysms and those with occlusive aortoiliac disease. He found that the patients with aneurysm were nine times more likely to be male; were, on average, 11 years older; and were much less likely to have had previous arterial surgery. Patients with occlusive disease were 16 times more likely to require reoperation after aortic surgery. In addition, patients with aneurysm were, on average, more than 5 cm (2 inches) taller than patients with occlusive arterial disease, and had a significantly greater body surface area. In our own experience, aneurysms in patients who have associated occlusive arterial disease are generally smaller and may be less likely to rupture than in patients without severe atherosclerosis. This view is supported by the observation that mean growth rates for small aneurysms are 50 per cent faster when there is no obvious occlusive arterial disease.

Genetic predisposition

Surgeons have been aware for some years of the occasional occurrence of several cases of abdominal aortic aneurysm within families, but proof of the familial pattern of the disease has been difficult to obtain because of the absence of symptoms and the advanced age of onset in most patients. Even carefully elicited family histories will often be unhelpful, since the majority of those with the aneurysmal diathesis will die from other causes, and many who die from rupture of the aneurysm will have the wrong diagnosis recorded unless an autopsy is carried out. The problem is compounded by the absence of a common name for aortic aneurysm, which is consequently unfamiliar to the general public.

Ultrasonographic screening studies have shown a prevalence of abdominal aortic aneurysm of 30 per cent in first-degree male relatives of patients with the disease. Because of the late age of onset, many of those with no evidence of an aneurysm at the time of screening could well develop the disease when they are older, so the

lifetime prevalence in brothers and sons of patients with aneurysm may be as high as 50 per cent. The search for the gene or genes responsible is hindered by the absence of three-generation families with confirmed inheritance of aneurysm for genetic studies. It is likely that, with the present rapid strides in molecular biology, this problem will be solved in the next few years, using techniques such as paired-sibling analysis.

Attention to date has largely been focused on possible variations in collagen genes, particularly type III collagen, but no convincing link with the common idiopathic aortic aneurysm of elderly people has been established. Other target genes are those controlling production of metalloprotease or metalloprotease inhibitors. Given the very strong association of aortic aneurysm with tobacco smoking it is likely that in most patients any genetic predisposition requires an environmental potentiating factor to be present for the phenotype to be manifest. The very late age of onset, the fact that most cases are occult, the need for environmental potentiation, declining rates of tobacco smoking in developed countries, and the possibility of multigene predisposition all conspire to make the search for a genetic cause of aortic aneurysm among the more difficult in molecular biology.

Connective tissue degradation

Research efforts have concentrated on attempts to discover the mechanism of breakdown of collagen and elastin in the arterial wall of enlarging and ruptured abdominal aortic aneurysms. Several studies have shown the presence of proteolytic activity in tissue from aneurysmal aorta, but authentic collagenase has been shown to be present only when the aneurysm has ruptured. It is uncertain whether collagenolysis is the cause or a consequence of rupture. Similar uncertainties surround the detection of elastase in the aortic wall and serum of patients with aneurysm. The discriminant value of such analyses between patients with and without aneurysms has not always been confirmed, although recently a unique metalloprotease elastase has been found only in patients with aneurysm. In one family in which aneurysms occurred in several members at an early age, it has been shown that the disease is linked to a genetically determined defect in type III collagen. It is possible that other genetic variations in type III collagen may account for some, if not all, cases of abdominal aortic aneurysm. Such a finding, although at present speculative, would mirror the situation in osteogenesis imperfecta, where the disorder has been shown to be caused by a large number of different genetic variations in type I collagen.

Recently it has been shown in patients who have had arterial bypass of popliteal aneurysms with autogenous vein that the vein graft has a higher propensity to dilatation, ectasia, and aneurysm formation than comparable vein grafts used to bypass atherosclerotic arterial occlusive disease. Metalloprotease concentrations are also higher in patients with popliteal aneurysms, suggesting a systemic process of vascular collagen breakdown and vessel dilatation.

Environmental potentiation

Abdominal aortic aneurysms have been shown to be associated with:

1. male sex;
2. advancing age;
3. tobacco smoking;
4. hypertension;
5. chronic obstructive airways disease (irrespective of smoking history);
6. occlusive arterial disease affecting coronary, carotid, and limb arteries.

In addition, aortic aneurysm is most common in whites and those who are tall, but these are unlikely to be independent disease determinants and probably reflect ‘racial' differences in population age structure and the relation between height, longevity, and socio-economic status.

The mechanisms by which environmental influences interact with underlying genetic predisposition to produce abdominal aortic aneurysm in an individual patient are at present uncertain, but the first three factors listed above have by far the greatest importance. It is interesting that, in common with other diseases, the marked protective effect of female sex is progressively lost with advancing age, although even in the very elderly person the risk of dying from aortic rupture is three times greater for men than for women.

Natural history

The great majority of abdominal aortic aneurysms are fusiform and are confined to the infrarenal segment. Small, saccular aneurysms are sometimes seen adjacent to atheromatous plaques in patients with predominant occlusive disease (Fig. 2) and rapidly growing, infective, ‘mycotic' saccular aneurysms occasionally occur as a consequence of bacteraemia. Mycotic aneurysms are a local manifestation of systemic disease, require urgent medical and surgical treatment irrespective of size, and have a totally different natural history from that discussed here of the common idiopathic aortic aneurysms of the elderly person.

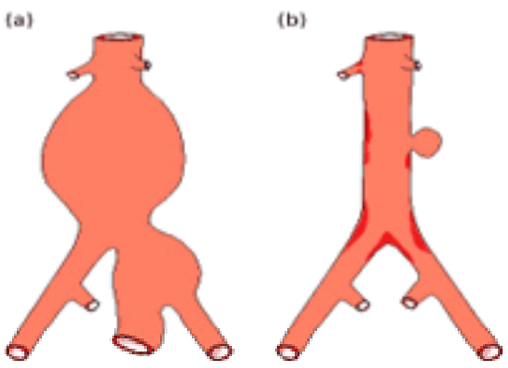


Fig. 2. (a) Large, fusiform, infrarenal abdominal aortic aneurysm continuous with aneurysms of the left common and internal iliac arteries. (b) A saccular aortic aneurysm associated with aortoiliac, atheromatous, occlusive arterial disease.

The median diameter of the infrarenal abdominal aorta increases with age in both men and women. An aortic aneurysm begins as a local accentuation of this normal ageing process. Physical laws predict that the rate of growth will increase with diameter, so once any local accentuation has started it can be expected to increase progressively with time. This explanation accounts for the three types of dilatation common seen:

1. a local aneurysm with normal adjacent arteries;
2. generalized arterial ectasia;
3. local dilatation within an ectatic arterial system.

Serial measurements have confirmed that growth rates increase as abdominal aortic aneurysms enlarge. The development of symptoms, risk of rupture, and clinical management of aortic aneurysms depend largely on their diameter, so it is convenient to discuss the natural history in relation to three somewhat arbitrary size ranges. It is important to remember, however, that the lifecycle of an individual aortic aneurysm is a continuous process from initial development to eventual rupture, the inevitability of which can be prevented only by elective surgery or prior death from some other disease. Looked at in this way, the description of an aortic aneurysm as a cancer of the artery is not quite so fanciful as it might seem.

Very small aneurysms (less than 4.0 cm diameter)

The prevalence of aneurysms of less than 4.0 cm in diameter has become apparent only with the introduction of screening programmes for the disease. Two-thirds of all aneurysms detected by population screening are of this size, the reasons for which are interesting and help in understanding some important features of the disease, as follows.

1. The longest part of the lifecycle of any aneurysm will be when it is small, since incremental growth rates increase as the aneurysm enlarges.
2. Large aneurysms are more likely to be detected and present in routine clinical practice.

3. The larger an aneurysm becomes, the more likely it is to rupture and remove the patient forever beyond the reach of screening programmes.

Very small aortic aneurysms generally enlarge much more slowly than the large aneurysms that present in routine clinical practice, and median growth rates of 0.2 cm per annum are usual. Clinical and autopsy evidence indicates that even these very small aneurysms do sometimes rupture, but there are insufficient data for the risk to be quantified accurately. Several clinical follow-up studies have shown no cases of rupture occurring in such patients while the aneurysms remained very small, but ruptures did occur as the aneurysms grew.

Small aneurysms (4.0–5.9 cm diameter)

Autopsy studies of patients with an abdominal aortic aneurysm show that one-third of aneurysms of less than 6.0 cm in diameter had ruptured and caused death. If we assume that the death rate in the two-thirds of patients who died with, rather than from, their aneurysm is the same as in the general population of the same age, these data suggest a rupture risk of around 2 per cent per annum. A complicating factor in accurate interpretation is that aortic diameter measured at autopsy is likely to have been less than the diameter of the same aneurysms in life. The actual risk of rupture for aneurysms of less than 6.0 cm diameter in life might therefore be substantially less than 2 per cent. Follow-up studies of patients managed conservatively with aneurysms of less than 6.0 cm diameter at the time of diagnosis have demonstrated a mean rupture rate of 6 per cent per annum over 3 years. Rupture rates are initially low but tend to increase progressively with the length of follow-up as the aneurysms continue to expand. For aneurysms of 4.0 to 4.9 cm diameter the mean expansion rate is 0.5 cm per annum, increasing to 0.7 cm per annum for aneurysms of 5.0 to 5.9 cm diameter.

Many of the patients managed conservatively have severe obstructive airways disease that would make operative intervention too hazardous. Obstructive airways disease is associated with more rapid expansion of aortic aneurysms and a higher risk of rupture. Consequently, data from follow-up studies of patients unfit for surgery cannot be used to estimate rupture risks for healthy patients. In addition, diameters measured 2 or 3 years before rupture occurs can give no useful information about rupture risk relative to current diameters.

Recently, data from the United Kingdom Small Aneurysm trial have revealed a rupture risk of not more than 1 per cent per annum for aneurysm of 4.0 to 5.5 cm maximum anteroposterior diameter measured by ultrasonography. Many of the ruptures in fact occurred in aneurysms that had already grown to more than 5.5 cm at the last measurement before rupture.

Large aneurysms (greater than 6.0 cm diameter)

Nowadays, patients with aneurysms of more than 6.0 cm diameter are invariably advised to have elective surgery. What we know of the natural history of large aneurysms comes from studies before operative treatment became possible in 1951, or from contemporary studies in patients too ill to undergo major surgery.

Studies from earlier in the twentieth century of clinically detected, and therefore presumably large and often symptomatic, abdominal aortic aneurysms report that most patients died within 3 years, and two-thirds of all deaths were from rupture of the aneurysm. Since the death rate from all causes in patients of the same age with no aneurysm was around 5 per cent per annum, the risk of rupture in these studies was therefore around 10 per cent per annum. Contemporary studies of patients with severe cardiac, respiratory, or other disease considered to make the risks of elective surgery unacceptably high show that aneurysm rupture accounts for half of all deaths, a rupture risk of 10 to 15 per cent per annum.

Clinical presentation

The majority of abdominal aortic aneurysms are asymptomatic and are often discovered incidentally. The patient may notice a pulsatile epigastric mass for the first time typically while lying relaxed in bed or bath. Large aneurysms in thin patients are readily detected on routine abdominal examination, but most are now discovered by ultrasonography or abdominal radiography performed to investigate unrelated symptoms. Urologists are a frequent source of referrals for many vascular surgeons, since prostatic hypertrophy and aortic aneurysm are both disorders of elderly men and detection of the aneurysm by abdominal palpation is easier during anaesthesia for prostatic resection. It is likely that much of the apparent increased incidence of abdominal aortic aneurysm over the past decade is attributable to general adoption of abdominal ultrasonography as the routine first-line investigation for abdominal symptoms.

In Britain, ruptured abdominal aortic aneurysm still accounts for around a third of all operations for the disease, but in the United States the figure for major vascular centres is currently between 5 and 20 per cent. Community studies have shown that 60 per cent of patients with ruptured abdominal aortic aneurysms do not reach hospital alive, while some of those who do are not operated upon. In Britain, rupture of the abdominal aortic aneurysm is sadly still the way in which more than half of all cases present. In both hemispheres there is a seasonal variation in the incidence of aortic rupture, with more cases occurring in the winter months. The reason for this pattern is unknown, but may be related to the similar observed seasonal variation in mean blood pressure.

Symptoms and signs of aortic rupture

Typically, rupture of an abdominal aortic aneurysm produces the sudden, unheralded onset of severe central abdominal and lumbar back pain. A pulsatile mass may be palpable in the epigastrium. Some patients may have experienced dull back pain of lesser severity for hours or days before, due to acute expansion of the aneurysm immediately before rupture. The lumbar pain may be worse on one side, commonly the left, because of the direction in which the retroperitoneal haematoma spreads. There may be a variable degree of psoas spasm, and sometimes pain in the lower limb, due to compression of lumbar or sciatic nerve roots. Rupture of an internal iliac (hypogastric) arterial aneurysm commonly produces maximal pain in the buttock and, rarely, blood may track with the sciatic nerve through the greater sciatic foramen to produce a gluteal haematoma.

Other early symptoms and signs depend on the volume of acute blood loss. Once the posterior peritoneum is breached, the patient will rapidly bleed to death into the peritoneal cavity, and most immediate deaths are due to intraperitoneal rupture. Survival after rupture depends on an intact posterior peritoneum, tissue tamponade, and early emergency surgery. When the connective-tissue tamponade provided by the retroperitoneum is very effective, or the leak is small, only modest haemorrhage may occur, and these patients can survive long journeys to hospital and several days before exsanguinating haemorrhage occurs. The self-selection of such patients for transfer to distant tertiary referral centres may be partly responsible for the superior results of some units. In most cases, tamponade is less effective and arrests acute haemorrhage only when assisted by hypotension secondary to blood loss. These patients exhibit pallor, sweating, tachycardia, and anuria; transfusion alone, by raising the blood pressure, will result in further haemorrhage. Immediate surgery to clamp the aorta above the site of rupture offers the only chance of survival.

Uncommon presentations

The great majority of abdominal aortic aneurysms will present as described above, but it is a common disease and any vascular surgeon will see several cases in their career, presenting in each of the following ways.

Aortic occlusion

Turbulent blood flow occurs in all aneurysms and slow transit of contrast medium is often seen on angiography. Turbulent flow contributes to the formation of the mural thrombus, which is present in most aneurysms. Sometimes the thrombotic process is more extensive and the aorta may occlude. Occlusion usually does not involve the origins of the renal arteries but is frequently accompanied by acute critical ischaemia of the lower limbs.

Distal embolization

Mural thrombus can become dislodged from within the aneurysm, perhaps as a consequence of direct abdominal trauma, and lodge as emboli in the arteries of the lower limb. One or two per cent of all emboli to the lower limb arise from this source.

Ureteric occlusion

Around 10 per cent of abdominal aortic aneurysms are of the ‘inflammatory' type, with a variable degree of perianeurysmal fibrosis. One or both ureters can become encased in fibrous tissue and occluded, either by being drawn medially towards the aortic aneurysm or, more commonly, where they cross an ‘inflammatory' aneurysm of the common iliac. The patient may present with hydronephrosis or anuria and renal failure.

Aortocaval fistula

This generally occurs in association with aortic rupture into the retroperitoneum, which consequently tends to dominate the clinical picture. In these circumstances the aortocaval fistula is usually only diagnosed peroperatively, when dramatic venous bleeding is seen on opening the aneurysm sac after aortic cross-clamping. Rarely, the aortic aneurysm may rupture only into the inferior vena cava and produce the characteristic clinical picture of venous engorgement and visible arterial pulsation in veins, accompanied by high-output cardiac failure.

Aortoenteric fistula

The majority of aortoenteric fistulas are seen as late complications of aortic surgery, and spontaneous fistulation into the gut from an aorta that has not been operated upon is extremely rare. Primary fistulas usually occur into the duodenum but fistulas in patients with an existing aortic graft tend to be into the first few centimetres of the jejunum and present with haematemesis and melaena. The treatment of this condition is one of the most difficult in vascular surgical practice, since *in situ* graft contamination and infection are inevitable.

Duodenal obstruction

The fourth part of the duodenum and duodenojejunal flexure is intimately adjacent to the abdominal aorta. A large infrarenal aortic aneurysm may therefore be a cause of external compression of the duodenum and high intestinal obstruction. The symptoms are those of duodenal distension with nausea and vomiting, which tends to be intermittent since the obstruction is incomplete.

Management

Symptomatic abdominal aortic aneurysms usually demand urgent or early treatment. The extent of preoperative investigation, assessment, and medical treatment may therefore need to be curtailed and the patient prepared for surgery as well as possible in the time available. The most immediate need for surgery arises in the patient with a ruptured aneurysm, and this is contrasted below with management of the asymptomatic patient. The management of other symptomatic presentations of the disease will fall somewhere between these two extremes, depending on how compelling is the need for surgery.

Ruptured abdominal aortic aneurysm

The key fact to remember is that these patients are in the process of bleeding to death from the moment rupture occurs. More than half will die within the hour from haemorrhage into the peritoneal cavity, and it is unlikely that these patients could ever be saved. The majority of patients arriving at front-line hospitals will be suffering from some degree of circulatory collapse with hypotension. In this condition, blood transfusion without arresting the haemorrhage is as futile as trying to fill a bucket with a hole in the bottom. The diagnosis should be made from the history and clinical examination. Investigations such as abdominal ultrasonography or radiography are unnecessary, time-consuming, and liable to cause fatal delay. The patient should be transferred immediately to the operating theatre, the only permissible investigation being the taking of a blood sample for cross-matching. In the operating theatre all preparations for the operation are carried out before the induction of anaesthesia, which should take place only when the surgeon is poised to make the abdominal incision. Anaesthesia is liable to induce severe hypotension as the vasoconstrictor tone that has been maintaining circulation to vital organs is abolished. At this stage, transfusion is given to the extent necessary to sustain essential functions. Only when the aorta above the rupture has been controlled and securely clamped should full transfusion to restore normal blood pressure be given.

In a number of patients with ruptured abdominal aortic aneurysm the haemorrhage is so well contained by the surrounding connective tissue that there are no obvious clinical signs of blood loss. Such individuals can survive long journeys to tertiary referral centres and may live for several days before the connective tissue finally gives way and fatal haemorrhage occurs. These patients are liable to be misdiagnosed as suffering from other conditions, of which the most common are ureteric colic, pancreatitis, and sciatica. To establish the diagnosis, ultrasonography or CT may be required. Once the diagnosis is certain, operation is required with appropriate urgency since fatal haemorrhage can occur at any time. It is particularly tragic to see a patient who arrived at the hospital in good condition transferred to the operating theatre in a collapsed state after prolonged delay.

Asymptomatic abdominal aortic aneurysm

The only substantial reason for treating the patient with an asymptomatic abdominal aortic aneurysm is to prevent their premature death at some indeterminate future date from its rupture. At present the only treatment known to reduce this risk is elective surgical replacement of the aneurysmal aorta. A decision to recommend treatment must therefore be based on balancing the operative mortality and morbidity against the risk of rupture. The limited information available on the natural history of abdominal aortic aneurysms shows a general relation between aneurysm diameter and rupture risk. For abdominal aortic aneurysms that remain 4.0 to 5.9 cm in diameter the risk of death from rupture is less than 2 per cent per annum and for diameters above 6.0 cm is of the order of 10 per cent per annum. Since the disease is unlikely to produce any distressing symptoms unless the aneurysm ruptures, it seems unreasonable to ask a patient to accept an immediate operative mortality risk greater than the annual expectation of death from the untreated disease. It is essential therefore that every patient should be carefully investigated and the individual risks of surgery assessed so that an informed judgement can be made in each case.

The United Kingdom small-aneurysm treatment trial has now confirmed that for patients whose aneurysms remain less than 5.5 cm in maximum anteroposterior diameter substantially more lives are lost from elective surgery than from aneurysm rupture. So far this is the only randomized, controlled study to have been completed but a similar study is currently being carried out in Veterans Administration Hospitals in the United States. The annual rupture rate for unoperated aneurysms in the United Kingdom study was less than 1 per cent and the mean elective operation mortality in 841 patients was 6.3 per cent. Operative mortality is ‘front loaded', occurring within 30 days of operation, while an annual mortality risk from rupture of the aneurysm is spread over 12 months. In the light of such data a surgeon recommending elective resection for an asymptomatic abdominal aortic aneurysm of less than 5.5 cm in maximum anteroposterior diameter would have some difficulty defending the decision. Surgeons tempted to believe that their operative mortality figures are much less than 6.3 per cent should bear in mind that even if they have had not a single death in their last 50 cases there is only a 95 per cent probability that their overall operative mortality will be less than 6 per cent. A 1 per cent annual rupture risk and a 6 per cent elective operative mortality rate gives a number needed to kill (NNK) for this operation of 20 at 1 year after surgery.

‘Inflammatory' abdominal aortic aneurysm

Inflammatory aneurysms comprise around 10 per cent of all abdominal aortic aneurysms encountered in clinical practice but since they are commonly symptomatic, this probably over-represents the prevalence of the inflammatory variant of aortic aneurysms in the entire population. The pathogenesis of this disorder is still the subject of debate, but the original suggestion that the inflammation is a response to leakage of blood from contained aortic rupture is no longer tenable. The macroscopic appearance at operation is of two types: (1) an angry, hyperaemic, periaortic inflammation, or (2) a chronic, fibrotic, icing-sugar aortic wall. Both types may be seen at different points on the same aneurysm. Histologically, the wall of all aortic aneurysms shows evidence of an inflammatory response and the difference between the macroscopically inflamed and non-inflamed aneurysm is quantitative rather than qualitative. The condition is best regarded as a chronic periaortitis and has much in common with idiopathic retroperitoneal fibrosis. Parums and Mitchinson in Cambridge, England, have advanced the theory that the periaortitis is an immune response to antigens, principally ceroid, leaking from atheromatous plaques into the aortic adventitia. It is unclear whether the liberation of lipoproteins from atherosclerotic plaques is simply a consequence of aortic dilatation or a contributory factor to aneurysm formation.

Inflammatory aneurysms may present with symptoms or signs suggestive of the diagnosis, or they may be discovered incidentally during investigation or at operation for an asymptomatic or ruptured abdominal aortic aneurysm. The belief that inflammatory aortic aneurysms are less likely to rupture is not supported by any evidence and should not weigh heavily in management decisions. Even the thickest aortic walls of inflammatory aneurysms tend to be thin posteriorly where they lie in contact with the vertebral bodies, and rupture at this point is not uncommon. The diagnosis of inflammatory abdominal aortic aneurysm should be suspected in patients presenting with a history of abdominal and back pain, and who have a tender but unruptured aneurysm. An elevated erythrocyte sedimentation rate will be present in half of those with an inflammatory aneurysm, and the diagnosis can be confirmed by demonstrating a thickened aortic wall on CT (Fig. 3) or MRI.

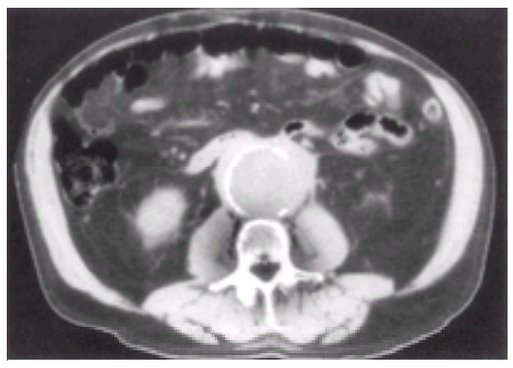


Fig. 3. CT scan of an inflammatory abdominal aortic aneurysm, showing the thick aortic wall.

In some patients, one or both ureters may be obstructed by the periaortitis or, more commonly, where they cross an inflammatory iliac aneurysm. Rarely, such patients may first present with renal failure, and the diagnosis of inflammatory aortic aneurysm be made secondarily. Hydronephrosis due to ureteric obstruction is usually best treated before elective aortic surgery, either by ureteric stenting or nephrostomy.

The presence of a stent in the ureter has the additional advantage of providing a useful guide to identification at operation when the ureters are encased in dense fibrosis. In general, following replacement of the aortic aneurysm, the ureteric obstruction will resolve and operative dissection of the ureters to free them from the periaortitis is seldom necessary.

Operative replacement of an inflammatory abdominal aortic aneurysm is difficult but can be satisfactorily performed in most patients by modification of a standard operative technique, since, fortunately, in the majority of instances the neck of the aneurysm is relatively free of periaortitis. Rarely, an elective operation may be too hazardous to continue when the aorta above the aneurysm is also inflamed. In such patients a case can be made for abandoning the procedure and treating for 3 months with systemic steroids to suppress the periaortitis before a further attempt at aneurysm replacement. ‘He who fights and runs away lives to fight another day.'

Investigation of the aneurysm

The purpose of these investigations is to determine accurately the size and extent of the aneurysm, to note the thickness of the aneurysm wall and the presence of any localized saccular dilatation, and, finally, to assess the importance of coexistent occlusive or aneurysmal arterial disease elsewhere.

Ultrasonography of the abdominal aorta is the first-line investigation; its advantages are that it is cheap, freely available, accurate, reliable, and reproducible. Being non-invasive it can be repeated as often as required, either to confirm the original findings or to monitor growth of the aneurysm. Its main disadvantages are that it is observer-dependent and the permanent images produced can be difficult for anyone, other than the person who performed the scan, to interpret. Visualization of the suprarenal aorta and iliac arteries is often difficult and the study may be impossible in the grossly obese or when large amounts of bowel gas are present. It remains, however, the most useful investigation for measuring the diameter of an infrarenal aortic aneurysm. The maximum anteroposterior diameter measured by ultrasonography is the reference standard for judging the rupture risk of an aneurysm and the need for elective aortic surgery. CT, MRI, and transverse-diameter measurements all tend to overestimate the maximum diameter of an aneurysm and patients should not be recommended for surgical treatment on the basis of such measurements alone.

CT produces excellent permanent records of cross-sectional anatomy that are easy to interpret (Fig. 4). Its main uses are to discover the extent of any suprarenal aortic involvement and the thickness of the arterial wall, in order to detect ‘inflammatory' aortic aneurysms. It can also be used to measure aortic diameter, but inaccuracies can occur if the section is not at right angles to the long axis of a tortuous aneurysmal aorta. An element of tortuosity is present in most aneurysms and transverse or oblique diameters are inherently unreliable. The anteroposterior diameter will usually be less and will more accurately reflect the true maximum diameter. Spiral CT allows a three-dimensional image of the aorta to be constructed so that a diameter at right angles to the long axis of the aneurysm can be measured. This technique overcomes some of the disadvantages of conventional CT but it has yet to be established that diameters measured by spiral CT correspond as closely to true diameters as those obtained by ultrasonography.

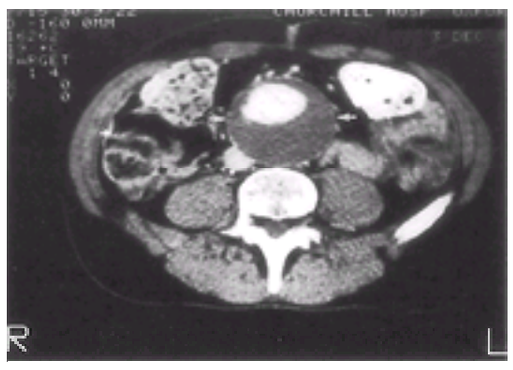


Fig. 4. CT scan of a large abdominal aortic aneurysm, showing mural thrombus.

MRI is now available in most centres, and the quality and definition of the images produced by the newer machines is excellent. Since there is no radiation exposure it could well replace CT in many of its present uses.

Angiography (Fig. 5) is used mainly to discover the relation of the renal arteries to the aneurysm and, by outlining the kidneys, may reveal relevant abnormalities, such as horseshoe or pelvic kidneys. When the aneurysm is large and blood flow turbulent or sluggish, it is sometimes difficult to obtain high-quality angiograms of the limb vessels from aortic injection of contrast, but this information can be of help in planning the extent of any arterial surgery required. The presence of iliac or femoral aneurysms or occlusive arterial disease in iliac, femoral, or more distal arteries of the limbs may be revealed. Routine angiography rarely furnishes information that alters the management of patients with abdominal aortic aneurysm. In general the risks of the procedure now outweigh any benefit obtained and its routine use should be abandoned.



Fig. 5. Angiogram of an abdominal aortic aneurysm, showing the origin of renal arteries.

Investigation of the patient

The purpose of these investigations is to discover how well the patient is likely to tolerate the trauma of a major arterial operation. Of particular importance are cardiac and respiratory function, and the presence of carotid arterial or other coexistent disease.

Attention is paid in the history to symptoms of angina, breathlessness at rest or on exertion, and previous myocardial infarction. All patients should have routine monitoring of their blood pressure and an electrocardiogram. Hypertension should be controlled and, if the history or electrocardiogram suggests possible abnormalities of myocardial function or blood supply, further investigations are essential. Echocardiography is useful for detecting abnormalities of valve or heart-wall function, and the technique of multigated acquisition nuclear imaging allows a ventricular ejection fraction to be calculated at rest and after exercise. In some patients, coronary angiography will be indicated, and any coronary arterial disease discovered may need treatment by angioplasty or bypass grafting before aortic surgery can be safely undertaken. Evidence of recent myocardial infarction is an important reason to recommend delaying elective surgery for all but the largest aneurysms, since the chances of further myocardial infarction and death are substantially increased by operation within 6 months of the infarct. Other relative contraindications to surgery are a low ventricular ejection fraction at rest or one which falls markedly on exercise, indicating inadequate blood supply to the myocardium.

Routine measures of respiratory function, such as peak expiratory flow and spirometry, are simple to perform as an extension of the normal clinical examination and should be a standard part of the assessment of all patients. More complex measurements of gas exchange are rarely necessary but, when required, the services of a respiratory function laboratory may prove helpful. Although poor lung function may be a contraindication to elective surgery, the presence of chronic obstructive airways disease is an important risk factor for aneurysm rupture. With the assistance of a chest physician, most patients can be improved to the point where the risks of elective surgery become acceptable.

The presence of severe, asymptomatic, carotid arterial stenosis presents a more difficult problem to resolve, and debate still goes on about whether carotid endarterectomy should be performed before, during, or after surgery for aortic aneurysm, or indeed, whether it should be performed at all. Each case will need to be resolved on its merits, depending on the relative importance of the aortic aneurysm and carotid stenosis in the individual patient, but, in general, the patient with an asymptomatic stenosis of the internal carotid artery is not greatly at risk of having a stroke during aortic surgery and carotid endarterectomy in such patients is not supported by the evidence available.

Because the majority of patients with an abdominal aortic aneurysm are old, coexistent disease is common. Malignant disease discovered during investigation for the aneurysm presents a particular problem. It is difficult to be dogmatic about the treatment priority, but a rationale for therapy is outlined here. Generally, primary treatment for the cancer is given first, since delay is liable to progressively reduce the chances of cure while, on the other hand, provided rupture has not occurred, a large aneurysm is as easily replaced as a small aneurysm. If primary treatment of the cancer seems to have been successful, then the aortic aneurysm is replaced as soon as the patient has recovered from cancer surgery. If treatment of the cancer is definitely non-curative and life expectation is limited, aneurysm surgery is seldom advised, since in most patients death from rupture of the aneurysm is likely to preferable to death from carcinoma.

Operative mortality and morbidity

In the decade after the first recorded replacement of an abdominal aortic aneurysm in 1951, operative mortality was high, at around 15 per cent. Over the past 30 years, as a result of refinements in surgical and anaesthetic technique and better pre- and postoperative management, elective operative mortality has been consistently reduced to under 5 per cent in some vascular units. It would be a mistake to assume that the low mortality currently reported by some specialist units represents the common experience, and in many hospitals a figure of 10 per cent or more is not unusual. It has been shown that reported operative mortality is higher when data are collected prospectively rather than retrospectively, and higher still when prospective data are collected in the community rather than in the hospitals where the surgery is performed. Prospectively collected community data indicate that the mean 30-day operative mortality of elective surgery for aortic aneurysm is around 8 per cent. Approximately 1 in 20 abdominal aortic aneurysms extends close to or above the origin of the renal arteries. Surgery in these cases may involve clamping the aorta above the renal arteries and sometimes their reimplantation. Operating on suprarenal aneurysms requires special skill and techniques, and inevitably is associated with greater hazard than for the uncomplicated infrarenal aneurysm.

Elective aortic surgery remains a major operation and, even in the uncomplicated case, morbidity is considerable. Most patients can be discharged from hospital within 10 days of operation, but few will be restored to complete well-being in less than 2 months. Currently, there is debate about the long-term outlook for those who have undergone successful replacement for aortic aneurysm. Some follow-up studies have shown a similar life expectation to that of the general population of the same age. This conclusion is disputed by others and seems to be inherently improbable given the known association of abdominal aortic aneurysm with a number of other diseases that impair life expectancy. There is no doubt, however, that the patient who survives surgical replacement of their aneurysm has a greater life expectation than one whose aneurysm is left untreated.

Operative technique

As with any operation, there are many variations in technique used by individual surgeons for routine operations, together with specific variations which may be employed to deal with special situations encountered. The standard techniques used by the author successfully over many years are described below and brief notes on useful or alternative techniques are appended after the main account.

Elective replacement

Preparation for the operation begins days or sometimes weeks before, to ensure that all the clinical information required is obtained and the patient is in the best state of health achievable at the time of operation. The main hazard peroperatively is sudden change in circulatory haemodynamics as a consequence of clamping and unclamping of the aorta, or blood loss. It is therefore essential that adequate monitoring of cardiac function and intravascular volume is in place before surgery begins.

The patient is placed supine on the operating table and the skin is prepared with antiseptic from the nipples to the knees—particular care must be taken to ensure cleaning of the external genitalia. The operation field extends from the xiphisternum to mid-thighs, with the genitalia being securely excluded by towelling and the use of adherent skin drapes. A vertical, midline abdominal incision is made from sternum to symphysis pubis and the peritoneal cavity opened. A complete inspection of all the intra-abdominal organs is made to exclude other pathology. The small intestine is retracted to the right and draped from the operative field, usually being retained within the abdomen, but in an obese patient better exposure is obtained if the intestine is exteriorized within a plastic ‘gut' bag.

Minimal dissection of the retroperitoneum is employed to limit bleeding and it is unnecessary to mobilize adherent duodenum. The neck of the aneurysm is identified by palpation, and the overlying peritoneum and fascia divided in the midline until the aorta is exposed. The inferior mesenteric vein is displaced to the left and seldom needs to be divided. Midline dissection continues until the left renal vein is identified, and fascial division is continued transversely at the lower border of the vein to free it and allow its retraction if required. Blunt dissection on both sides of the aorta in a strictly vertical plane continues until the vertebral body is encountered. Intravenous heparin is administered and the neck of the aorta clamped anteroposteriorly.

When the common iliac arteries are not aneurysmal, their dissection is easily accomplished by division of peritoneum and fascia over their anterior surfaces with blunt finger dissection down each side. The vessels are clamped anteroposteriorly. Common and internal iliac arteries may be aneurysmal and, in these circumstances, the external iliac arteries are clamped and back-bleeding from the internal iliac arteries controlled after the aortic aneurysm is opened.

The aortic aneurysm is inspected and the inferior mesenteric artery is oversewn with a transfixion suture at its origin. The aneurysm is then incised in the midline from its neck to the aortic bifurcation, and the mural thrombus evacuated. At each end of the incision, transverse scissors cuts are made so that half the circumference of the aorta is divided. Bleeding from lumbar vessels is controlled by direct pressure until permanently arrested by oversewing with transfixion sutures. At this stage the operative field should be bloodless and the ends of the aorta can be inspected and cleared of adherent thrombus.

In 80 per cent of cases a tube graft can be used, but where the iliac arteries are aneurysmal a bifurcated ‘trouser' graft will be required. In the latter case it is preferable, and usually satisfactory, to anastomose each limb of the graft either to the termination of the common iliac artery or to the external iliac artery. It is desirable to retain circulation into at least one internal iliac artery to minimize the risk of ischaemia in the gut or spinal cord. The graft is stitched to the neck of the aneurysm, using an inlay technique and a monofilament prolene continuous suture (Fig. 6). Exposure is improved by inserting a self-retaining rake retractor within the aneurysm sac. Three sutures are placed on each side of the midline of the aorta and graft posteriorly. and the graft is ‘parachuted' into place. The suture is continued on each side to the midline anteriorly, particular care being taken to place accurately the corner sutures in the lateral walls of the aorta. The anastomosis is tested for leaks at this stage, since subsequent haemorrhage from the posterior wall is more difficult to deal with.

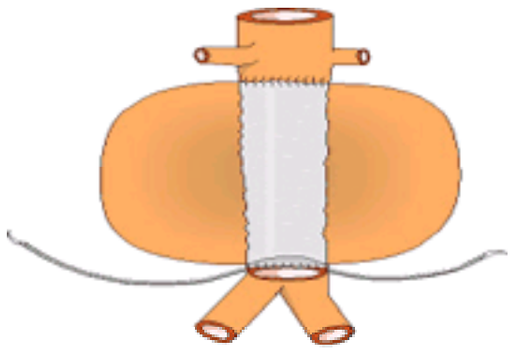


Fig. 6. Diagramatic representation of an aortic tube graft being inserted.

The tube graft is cut to the appropriate length and the distal anastomosis made in exactly the same fashion. A vital step in the procedure is to ensure that no particulate material, atheroma, thrombus, or tissue remains inside the graft or iliac arteries proximal to the arterial clamps. Graft and proximal iliac arteries are therefore irrigated thoroughly with saline before the anastomosis is completed. It is unnecessary to release the distal arterial clamps to achieve this end, and doing so may precipitate arterial thrombosis by exposing blood in the distal vessels to tissue thromboplastin from the operation site.

Five minutes before the anastomoses are completed the anaesthetist is warned that restoration of circulation to the legs is imminent so that circulatory volume can be appropriately and rapidly augmented when required. One distal clamp only is removed and the aortic clamp slowly released. This is a dangerous phase of the operation, and flow through the graft is titrated against the patient's blood pressure and heart filling pressure. Only when pressures are normal, with full restoration of blood flow to one limb, is the second distal clamp slowly released.

The operation is completed by meticulous haemostasis, and reversal of anticoagulation may be necessary to achieve this end.

Technical variations

Transverse incision Many surgeons use a transverse, upper abdominal, dome-shaped incision for abdominal access. The muscles of the abdominal wall are cut in the line of the incision, which commences midway between the umbilicus and sternum and runs parallel to the costal margins. It is claimed that postoperative respiratory complications are reduced by this incision but it is more time-consuming to make and close, bleeding is greater, and access to the iliac arteries is difficult.

Retroperitoneal approach The patient is positioned corkscrewed on the operating table with the pelvis horizontal and the shoulders vertical. An incision is made from the tip of the left twelfth rib to the midline below the umbilicus. Muscles of the abdominal wall are cut in the line of the incision. The extraperitoneal plane is identified and the peritoneum retracted medially to expose the aorta. The left kidney may be retracted with the peritoneum or allowed to remain lying on the psoas muscle. Advocates of this approach point to the facility of the technique in obese patients and claim reduced postoperative morbidity. The disadvantage is that exposure of the right common iliac artery cannot readily be achieved.

The perirenal aneurysm neck Aneurysms extending substantially above the origins of the renal arteries are discussed in the section on thoracoabdominal aortic aneurysms, but aneurysms with a neck at, or immediately above, the origin of the renal arteries can be dealt with by slight modification of the operative approach to the common infrarenal aortic aneurysm. The problem may be encountered unexpectedly at operation in a patient who was thought on preoperative assessment to have a purely infrarenal aneurysm.

The aorta needs to be dissected and clamped above one or both renal arteries. To achieve this exposure the left renal vein must be freed from the aorta and retracted either superiorly or, occasionally, inferiorly. It will be necessary to divide either the left gonadal vein or sometimes the left adrenal vein to permit safe retraction, but division of the renal vein itself is rarely required. The anastomosis between graft and aorta is accomplished expeditiously, with the renal arteries being incorporated within the proximal aorta as a single or double, short tongue. After completion of the proximal anastomosis, the clamp is reapplied to the graft below the renal arteries and blood flow to the kidneys restored before attention is turned to the distal aortic anastomosis.

It is claimed that the retroperitoneal approach allows the perirenal aortic aneurysm to be dealt with as easily as the infrarenal aortic aneurysm and is part of the advocacy for general adoption of this approach.

Types of graft In the early days of aortic surgery, aortic homografts were widely used but were abandoned because their use was inconvenient and they were prone to aneurysmal degeneration. They have been obsolete for 30 years, and have only recently been reintroduced and advocated for use in the presence of established infection of a synthetic graft. Synthetic grafts in common use are of woven or knitted Dacron or polytetrafluoroethylene. Woven Dacron is stiffer and less permeable than the knitted material.

The problem of blood leaking through the graft at operation can be overcome by coating the knitted graft in various ways, but this adds to the cost disadvantage compared with the woven material. Polytetrafluoroethylene grafts are impermeable but suffer from two disadvantages. First, the aortic body of a trouser graft must be cut to the correct length with minimal tolerance if the legs are to lie at an acceptable angle. Secondly, polytetrafluoroethylene is prone to leak at stitch holes for a prolonged time, a problem which is increased when a large needle is used.

If a trouser graft is essential, anastomosis of the limbs to the iliac arteries is preferable, since this avoids additional incisions in the groins and consequently reduces the risk of contamination of the graft and graft infection.

The external iliac arteries invariably and mysteriously remain uninvolved by aneurysmal change, although they may become occluded by atherosclerosis.

Clear indications for aortofemoral grafting are the presence of large femoral arterial aneurysms and stenosis or occlusion of the external iliac arteries.

Replacement of the ruptured abdominal aortic aneurysm

Two-thirds of ruptured aneurysms leak either directly or secondarily into the peritoneal cavity within an hour or so and cause death from exsanguination. These patients usually die at home or on their way to hospital and thus do not come to surgery. Successful surgery is possible only because of temporary tamponade of the rupture by tissue pressure in the retroperitoneum, assisted by hypotension. As described above, these patients are on the brink of death and sustain fatal haemorrhage when the peritoneum ruptures, tissue tamponade fails, or the blood pressure is increased by injudicious transfusion before the rupture is secured. Most lives are saved by immediate transfer from emergency room to the operating theatre, but operative mortality rates improve as the time between rupture and surgery increases. The time delay associated with interhospital transfer selects for treatment those with stable, contained ruptures. Those with poorly contained ruptures die before referral or in transit.

The patient is prepared for surgery while conscious on the operating table. Induction of anaesthesia should occur only when the surgeon is poised ready to make the incision. Blood pressure should be maintained at no more than 100 mmHg systolic until the aorta is controlled. This ideal may require rapid transfusion in order to maintain cerebral and myocardial perfusion on anaesthetic induction when compensatory vasomotor tone is suddenly relaxed. On opening the peritoneal cavity the posterior peritoneum is exposed by displacing the small intestine. The aorta above the aneurysm is identified by palpation and occluded by direct compression with the surgeon's left hand or a ‘snake catcher' instrument (Fig. 7) against the vertebral bodies. Only when the aorta has been securely occluded by compression should the posterior peritoneum over the upper part of the aneurysm be incised. The wall of the aneurysm should be exposed by sharp dissection with scissors through the haematoma and connective tissue, and only when the aortic adventitia is clearly identified is it permissible to use blunt finger dissection in the periadventitial plane to clear the neck of the aneurysm. When the neck of the aneurysm is securely clamped the ‘snake catcher' is removed. The operation then proceeds as for an elective replacement.

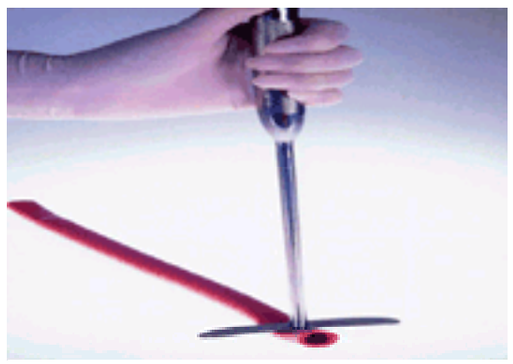


Fig. 7. The ‘snake catcher' instrument for compressing the abdominal aorta against the bodies of the lumbar vertebrae.

Death from ruptured aneurysm is a direct consequence of blood loss. Patients will rarely survive if additional blood loss is caused by extensive, unnecessary, hasty, or careless dissection. Iatrogenic blood loss is most likely to occur from tearing of the left gonadal, left adrenal, or left renal vein by attempts to dissect the aneurysm neck that are too hasty, too wide, and too high. Control of blood flow into the aneurysm must be achieved before any dissection is attempted.

High aneurysm rupture

When rupture of the aneurysm occurs close to its neck, it may be difficult to dissect the neck to clamp it. In this situation the aortic neck can be identified from within the lumen of the aneurysm and a large Foley urethral catheter inserted. The balloon of the catheter is inflated in the neck of the aneurysm and tamponade of the aorta against the vertebral bodies is slowly released to confirm that the balloon is securely impacted. Careful, external, two-handed dissection of the neck of the aneurysm can then be performed and the balloon catheter replaced by an aortic clamp.

Endoluminal surgery

The first endoluminal repair of an abdominal aortic aneurysm was reported in 1991. Since then the technique has been employed on several hundred patients throughout the world using a large number of different home-made, customized, and commercially produced devices. The feasibility of endoluminal surgery in carefully selected patients is unquestionable but technical failures are common and the effectiveness of the procedure in providing useful prophylaxis against rupture of an aortic aneurysm is unproved.

Types of endoluminal graft (Fig. 8)

Aortic tube graft Eighty per cent of open, elective repairs of aortic aneurysms can be accomplished with tube grafting. The avoidance of bifurcated grafts whenever possible has contributed to the reduction in operative mortality and morbidity and chronic graft infection. The first commercial grafts for endoluminal insertion were designed with the hope of retaining these advantages. Unfortunately, endoluminal aortic tube grafts are suitable for only a very small minority of aortic aneurysms since a caudad, normal-diameter aortic cuff of at least 10 mm long above the origin of the common iliac arteries is required for their insertion. In the rare instances when a suitable cuff is present, graft insertion is often technically unsuccessful because of the need to select the exact length of graft required if distal graft fixation without kinking or iliac arterial occlusion is to be achieved.

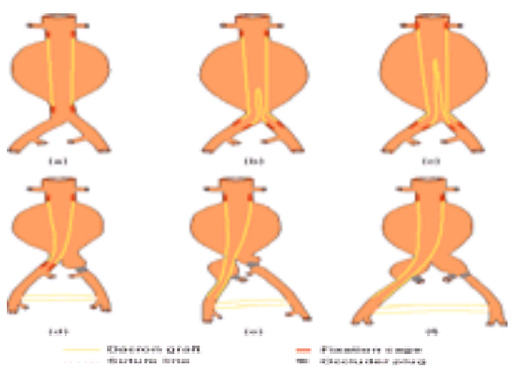


Fig. 8. Diagrammatic representation of the main types of endoluminal abdominal aortic aneurysm exclusion. (a) Aortoaortic tube graft. (b) Aorto-bi-iliac graft. (c). Aorto-bi-iliac graft with cephalad fixation cage crossing the renal arteries. (d) Aorto-uni-iliac graft with femorofemoral cross-over graft. (e) Aorto-unifemoral graft with sutured caudal fixation and femorofemoral cross-over graft. (f) Aorto-unifemoral graft with cephalad fixation cage crossing the renal arteries; femorofemoral cross-over graft and occlusion of both internal iliac arteries.

Aorto-bi-iliac graft Endografts from the aorta to both iliac arteries are either of a one-piece or modular type. With one-piece devices a guidewire attached to the contralateral iliac graft limb is first passed from the ipsilateral groin into the aorta, to be captured and withdrawn by a snare passed from the contralateral femoral artery. With modular devices the ipsilateral graft limb is a full-length trouser leg and the contralateral limb a short trouser leg. The short trouser leg is deployed inside the aortic aneurysm and a long trouser-leg extension is invaginated into it from the contralateral femoral artery.

Deployment of the distal end of the iliac graft limbs may be into the common or external iliac arteries. Usually the common iliac arteries are chosen to allow perfusion of the internal iliac arteries and minimize the risk of retrograde blood flow between the graft and lumen of the aneurysm.

Aorto-unilateral graft with extra-anatomic bypass Aorto-bi-iliac graft insertion is often technically difficult and some patients have only unilateral iliac arteries suitable for endoluminal graft deployment. A customized technique for such patients is to insert a tapered aorto-uni-iliac graft from the most suitable limb. The common iliac artery in the contralateral limb is then occluded and blood flow re-established with a femorofemoral cross-over graft using a conventional, open surgical technique.

An alternative procedure is to first occlude the ipsilateral internal iliac artery and insert a tapered aortofemoral endoluminal graft. The distal limb can then be appropriately tensioned, cut to length, and sewn directly to the femoral artery. The operation is similarly completed by occlusion of the contralateral common iliac artery and insertion of a femorofemoral cross-over graft.

Anatomical constraints on endoluminal surgery

The morphology of the non-aneurysmal aortic cuffs, the aortic and iliac aneurysms, and non-aneurysmal iliac and femoral arteries can all impose constraints on the use of endoluminal grafts. Some of the constraints are important enough to be regarded as absolute by most surgeons, others are relative contraindications that preclude the use of some types of graft or increase the risk of technical failure (Table 3).

All aortic grafts
• An aortic neck below patent renal arteries
• No mural thrombus in aortic neck
• Aortic neck 28mm maximum diameter
• For infrarenal fixation, minimum neck length 15mm
Aortic tube grafts
The distal aortic cuff should be:
• At least 10mm long
• At most 28mm in diameter
• Free of mural thrombus
Aorto iliac grafts
The recipient iliac artery should have a:
• Maximum diameter of 14mm
• Minimum diameter of 8mm

Table 3 Anatomical requirements for endoluminal aortic aneurysm exclusion

All endoluminal grafts for aortic aneurysms require an aortic neck below the origin of patent renal arteries. The neck should be free of mural thrombus, have a maximum diameter of 28 mm and ideally a length of at least 15 mm. The constraint of a maximum diameter was initially imposed by the calibre of deployment system that can usually be introduced through the femoral or iliac arteries. A more fundamental problem is that large-diameter aortic necks are already aneurysmal or ectatic and soon expand away from the graft fixation system, with the development of proximal leaks. Similarly, mural thrombus in the neck is indicative of pathological dilatation, and has the additional disadvantages that initial graft fixation is seldom possible and peripheral embolization of dislodged thrombus is likely. The neck-length restriction can be reduced if an uncovered graft-fixation system is deployed at or above the origin of the renal arteries. The diameter of the wire used for the fixation system is less than the diameter of most renal arterial orifices. The random chance of a narrow orifice coinciding with the soldered junction of two wires is such that renal infarction does not usually occur. The technique of suprarenal graft deployment is for surgeons who are prepared to play Russian roulette with their patients' renal function.

Aortic tube-graft insertion requires a distal aortic cuff of 28 mm maximum diameter and 10 mm minimum length. The presence of such a cuff in aneurysms of clinical significance is unusual and fewer than 5 per cent of endografts currently inserted are aortic tube grafts.

Aortoiliac grafts require that at least one common iliac artery should have a minimum diameter of 8 mm and maximum diameter of 14 mm, and be at least 25 mm long. If the deployment system is introduced from the groin, the minimum-diameter restrictions also apply to the femoral and external iliac arteries. An acute angle at the junction of common iliac artery with the aorta can often be increased by traction on the iliac artery. Persistent angulation of less than 90° can prevent introduction of the deployment system into the aorta.

Mortality and morbidity of endoluminal aortic graft insertion

Thirty-day mortality in published reports from individual series of endoluminal aortic aneurysm exclusion varies from none to 28 per cent with a mean of 5 per cent. It is certain that the true overall operative mortality is currently substantially higher because of reporting bias and the fact that most surgeons and radiologists have performed few procedures, and are on a learning curve for the technique.

National and international registries of endoluminal aneurysm surgery have been established and these have recorded operative mortality of around 10 per cent. Even this figure is more likely than not to under-represent the true mortality, since reporting of cases remains voluntary. It is likely that as case selection becomes more refined and technical expertise improves the operative mortality will be reduced.

As with all major surgery, patients undergoing endoluminal exclusion of aortic aneurysm are at risk of developing cardiac, respiratory, renal, and multiorgan failure. There are also specific problems that are more likely to be associated with endovascular aneurysm surgery. The most common of these appears to be persistent postoperative pyrexia, the cause of which is uncertain. Others are arterial access and instrumentation injuries, thromboembolic events, groin lymph leaks, and endograft or extra-anatomic graft infection.

Early and late technical failure

The reported initial technical success rate for endoluminal aneurysmal exclusion, defined as correct placement of the graft without perigraft blood flow (endoleakage, see next), death or graft occlusion within 30 days of implantation, ranges from 48 to 95 per cent. The most common technical problem is blood flow between the deployed endograft and the aneurysm, now commonly called an endoleak. Endoleaks occur at either the proximal or distal implantation sites or from patent lumbar, inferior mesenteric or iliac arteries. For an endoleak to be demonstrable and persist there must be both an inflow and an outflow for the leaking blood. Some small, peroperative endoleaks seal spontaneously. Others may be closed secondarily by further insertion of an endoluminal stent either at the time of initial surgery or subsequently.

Large peroperative endoleaks, failed graft deployment, or other major technical problems require conversion to open aneurysmal repair. The need to convert to conventional aneurysm surgery varies with the technical expertise of the operator and the difficulty of the procedure attempted. It is generally accompanied by a high operative mortality.

Some early endoleaks are either unrecognized at the time of operation or appear subsequently. It may be possible to seal them by further endoluminal graft insertion. A persistent endoleak means that the operation has failed, aneurysm expansion will continue to occur, and the patient remains at the same risk of aneurysm rupture as there would have been had the operation not been performed.

Late endoleaks occur either as a consequence of graft material failure or continued expansion of the aorta or iliac arteries at the site of graft fixation. Most graft failures to date have been of the metal fixation devices, particularly of fixation hooks, due to metal fatigue. The fabric of some endografts has been chosen as much for its low bulk as its strength and is likely to be less durable than conventional graft fabric. The most fundamental problem, however, is that of continued expansion of the aorta at the neck of the aneurysm. Patients with aortic aneurysms usually have generalized arterial ectasia and continued arterial expansion tends to occur with advancing age. Some surgeons believe that most, if not all, endovascular grafts will become too small as aortic dilatation progresses, with consequential late graft displacement or the development of endoleaks and continued aneurysm expansion.

Does endovascular aneurysm exclusion provide useful prophylaxis against rupture?

Substantially the sole purpose of elective surgery for aortic aneurysm is to provide prophylaxis against unnecessarily premature death from rupture. Net benefit in life years gained is achievable when the death rate from the natural history of the disease is high and when the operative mortality of an effective, durable, prophylactic operation is low. When the operative mortality is the same as the annual mortality from the natural history of aortic aneurysm rupture a net gain in life years will begin to accrue from 2 years after treatment if the operation provides total protection from rupture. Endovascular exclusion of aortic aneurysm is most likely to be technically successful in small aneurysms at little risk of rupture. When used for large aneurysms, early technical failure is common and delayed technical failure from endoleaks is probable. At present, endovascular surgery for aortic aneurysm is a technically challenging and exciting innovation of no established therapeutic value. It remains to be seen whether it will become an effective means of providing prophylaxis against rupture of an aneurysm in some patients.

Further reading

Blankensteijn JB, Lindenburg FP, Van der Graaf Y, Eikelboom BC. Influence of study design on reported mortality and morbidity rates after abdominal aortic aneurysm repair. *British Journal of Surgery* 1998, **85**: 1624–30. [Shows how reported mortality rates are influenced by the methods of data collection.]

Collin J. Aortic aneurysm screening and management. In: Johnson CD, Taylor I, ed. *Recent advances in surgery*. Churchill Livingstone, Edinburgh, 1990. [An overview of justification for AAA screening.]

Collin J, Araujo L, Walton J, Lindsell D. Oxford Screening Programme for abdominal aortic aneurysm in men aged 65–74 years. *Lancet* 1988; **ii**: 613–15. [The first report of screening for AAA.]

Greenhalgh RM, Mannick JA, ed. *The cause and management of aneurysms*. Saunders, Philadelphia, 1990. [Multi-author international conference proceedings.]

Parums DV, Mitchinson MJ. Serum antibodies to oxidised LDL and ceroid in chronic periaortitis. *Journal of Pathology* 1987; **151**: 57. [Deals with the role of ceroid in aortic aneurysm pathogenesis.]

Pierce GE, ed. Abdominal aortic aneurysms. *Surgical Clinics of North America* 1989; **69**: 4. [A mult-author report on AAA.]

Tilson MD. A perspective of research in abdominal aortic aneurysm disease with a unifying hypothesis. In: Bergan JJ, Yao JST, ed. *Aortic surgery*. Saunders, Philadelphia, 1989. [A review of the evidence on pathogenesis of AAA.]

UK Small Aneurysm Trial participants. Mortality results for randomised controlled trial of early elective surgery on ultrasonographic surveillance for small abdominal aortic aneurysms. *Lancet* 1998; **352**: 1649–55. [Confirms that elective AAA repair for aneurysms less than 5.5 cm in diamether is much more dangerous than the natural history of the disease.]

Veith FJ, ed. *Current critical problems in vascular surgery*. Quality Medical Publishing, St. Louis, 1989. [Several chapters relevant to AAA management.]

Woodburn KR, May J, White GH. Endoluminal abdominal aortic aneurysm surgery. *British Journal of Surgery* 1998, **85**: 435–43. [An update of endoluminal surgery from the most active team in the world in Sydney, Australia.]

17.6.2 Femoral artery

Richard P. Cambria and Margaret L. Schwarze

Introduction
Classification
Atherosclerotic femoral aneurysm
False femoral aneurysm
Traumatic pseudoaneurysm
Anastomotic femoral aneurysm
Further reading

Introduction

Aneurysmal disease in the femoral triangle encompasses a spectrum of pathology; it includes both degenerative aneurysm in native vessels and the more commonly encountered false aneurysms. Whereas a true aneurysm is an abnormal dilatation involving all three layers of the vascular wall, a false aneurysm results from a rent in the integrity of either a native vessel or a previous vascular suture line, the ‘wall' of the aneurysm being composed of surrounding scar and inflammatory tissue. The incidence, clinical importance, and management of these diverse lesions vary considerably.

Classification

Femoral artery aneurysms may be true aneurysms due to atherosclerotic degeneration of native vessels or false aneurysms that are either: (1) secondary to penetrating trauma or catheterization of the femoral artery, or (2) anastomotic aneurysms at the site of prior vascular reconstructions (Table 1). Iatrogenic false aneurysm of the native vessel is the most commonly encountered lesion in accordance with the ever-increasing numbers and types of transarterial femoral catheterization procedures. An important complicating factor in the management of the false aneurysm is the presence or absence of infection. Infection can be the principal cause of the false aneurysm but hematogenous seeding can secondarily infect an established femoral aneurysm of any type. Acute necrotizing infection may be a cause of false aneurysm, particularly in intravenous drug abusers who use the femoral vessels as sites of vascular access.

True femoral artery aneurysms—atherosclerotic disease
False femoral artery aneurysms
1. Traumatic aneurysms (catheter injuries, penetrating wounds)
2. Anastomotic aneurysms

Table 1 A classification of femoral artery aneurysm

Atherosclerotic femoral aneurysm

While atherosclerotic femoral aneurysm is the second most common peripheral aneurysm (after popliteal artery aneurysm), it is an uncommon clinical problem. Atherosclerotic femoral aneurysms are generally detected in patients undergoing evaluation for aneurysmal disease elsewhere in the vascular tree. As is the case with the more commonly encountered popliteal aneurysm, atherosclerotic femoral aneurysm has a strong predilection for the male sex, is bilateral in over 70 per cent of the cases, and is seen in association with aortoiliac aneurysmal disease in 85 per cent of the patients. Still, in the largest clinical series reported, only 3 per cent with aortic aneurysm had simultaneous femoral aneurysm.

Diagnostic uncertainty in femoral aneurysm disease generally involves the distinction between diffuse arteriomegaly and genuine aneurysm. Criteria similar to those developed for the definition of a popliteal aneurysm are appropriate. A femoral artery can be considered aneurysmal when its maximum diameter is 1.5 times or more that of the upstream normal external iliac artery. In addition to the physical examination, ultrasound evaluation combined with color Doppler examination provides excellent resolution and a high degree of diagnostic accuracy. The demonstration of laminated thrombus within the aneurysm cavity confirms the diagnosis of a true aneurysm as distinct from an ectatic artery. Ultrasound examination also allows an accurate measurement of size, which may be important in determining whether aneurysm repair is indicated. As is the case in aneurysmal disease in a variety of locations, angiography is not required for diagnosis, but it is essential in planning the technical details of operative repair.

Controversy exists as to the natural history and clinical importance of untreated femoral aneurysm. A 1997 review of the literature on femoral artery aneurysms finds that distal embolization occurs in 0 to 26 per cent of cases, thrombosis in 15 per cent, and rupture in 10 per cent. Local compressive symptoms of both the femoral vein, resulting in venous congestion, and the femoral nerve causing neuropathic complaints are also of consequence. Clearly, symptomatic aneurysms and those larger than 3 cm should be repaired. We favour operations for all lesions larger than 3 cm, particularly if ultrasound examination demonstrates mural thrombus. These criteria are also applied to the patient with simultaneous aortic and femoral aneurysm, although segmental staged resection is often preferred if the intervening external iliac artery is unaffected.

The technical aspects of operative repair relate to the extent of the aneurysm (Fig. 1). Femoral aneurysms have been classified as either type I, terminating proximal to the common femoral artery bifurcation, or type II, where the aneurysm involves the origin of the profunda femoris artery. Type I aneurysms are simply repaired with excision and interposition of a prosthetic conduit, as these arteries will generally be too large to repair with a vein graft. When the aneurysm extends into the femoral artery bifurcation, the details of the repair are somewhat more complex. Since these vessels tend to be large, a short segment of 12 × 6 mm Dacron bifurcation graft for separate anastomoses to the profunda and superficial femoral arteries is used. If concomitant femoral popliteal reconstruction is required to treat distal ischemia, it will generally be necessary to carry out two separate reconstructions, with the femoral popliteal graft arising from the primary femoral artery reconstruction. The temptation simply to excise the anterior wall of the aneurysm and close the defect with the proximal hood of a simultaneous femoropopliteal bypass graft should be avoided, as the remaining femoral artery may become aneurysmal. When femoral aneurysm occurs in association with proximal aortoiliac aneurysms, the principal of segmental resection should be applied. This entails separate end to end graft reconstructions to prevent the development of anastamotic psuedoaneurym and to avoid the sacrifice of pelvic or profunda femoris circulation.

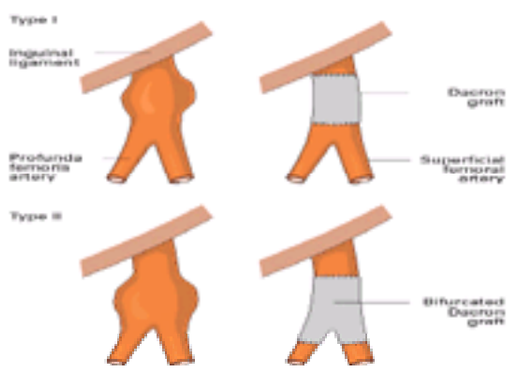


Fig. 1. Classification and repair of femoral artery aneurysms.

False femoral aneurysm

Traumatic pseudoaneurysm

False aneurysm of the native femoral artery, secondary to percutaneous arterial catheterization, is a common problem for the vascular surgeon (Fig. 2). The incidence of this problem increases with the complexity and size of the catheters and sheaths introduced into the femoral artery: reported incidences for simple diagnostic cardiac catheterization are in the 0.5 per cent range, but this figure can increase to greater than 10 per cent after the percutaneous introduction of an intra-aortic counterpulsation balloon pump. Obviously, arterial complications of such procedures are not limited to false aneurysm formation and local thrombotic complications are more common than false aneurysms with the intra-aotic balloon pump. Traumatic femoral arteriovenous fistula formation can also occur, either in association with, or independent of, the false aneurysm. A characteristic systolic/diastolic murmur, audible with the stethoscope, is diagnostic of this condition. Despite the fact that indwelling sheaths may be left in the femoral artery for some period of time, infection in these lesions is distinctly uncommon.

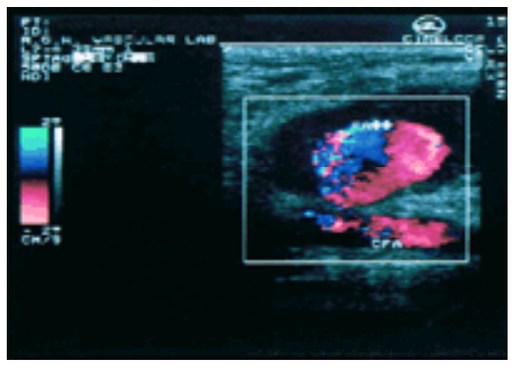


Fig. 2. Color flow duplex examination demonstrating arterial flow in false aneurysm cavity (FA) originating from common femoral artery (CFA).

While the diagnosis of femoral pseudoaneurysm may be straightforward on the basis of a tender pulsatile mass, it may be difficult to distinguish a false aneurysm from an uncomplicated periarterial hematoma in the period immediately following the catheterization procedure. In such circumstances, a color flow Doppler examination is the diagnostic test of choice. Ultrasound can localize the origin of the false aneurysm, characterize its size and shape, and can be used to direct the force of manual compression to treat iatrogenic pseudoaneurysms successfully.

While prompt surgical repair to prevent complications such as progressive enlargement, femoral nerve compression, rupture, and distal embolization has been both effective and routinely applied, recent natural history data and the availability of ultrasound-guided compression therapy have occasioned less frequent application of surgical treatment. It is now appreciated from natural history studies, based on vascular laboratory referrals, that many iatrogenic false aneurysms will thrombose spontaneously within weeks, nearly always with maintenance of femoral arterial continuity. Yet it is clear that some of these lesions enlarge progressively and/or rebleed in the acute phase. Although factors predictive of persistent pseudoaneurysm are still unclear, pseudoaneurysms less than 3 cm in diameter are safe to follow. Patients who are managed expectantly should have close monitoring and rapid enlargement, severe pain, and femoral nerve compression should promote urgent surgical intervention.

As an alternative to expectant management, ultrasound-guided compression therapy is quite successful. Using manual compression directed at the neck of the false aneurysm for periods of 10 min, thrombosis of the pseudoaneurysm is achieved. The procedure requires careful monitoring of distal arterial flow and intravenous narcotics as patient intolerance is often a limiting factor. Recently ultrasound-guided thrombin injection has been shown to be effective in inducing false aneurysm thrombosis.

Since 1991 over 300 cases of successful ultrasound-guided compression therapy have been reported in the literature. Regardless of the size of the aneurysm, the obliteration of iatrogenic pseudoaneurysms with ultrasound-guided manual compression is possible for more than 80 per cent of patients in some studies. The procedure, when performed with proper monitoring, has minimal morbidity. Few cases of spontaneous rupture and distal thromboembolic events have been reported after successful therapy. Manual compression is less successful for patients who require continuous anticoagulation and surgical repair is the best option for these patients. Additionally, those patients who have associated large hematomas, femoral nerve compression, require other surgical procedures such as coronary bypass, and those without adequate follow-up should have surgical repair of their false aneurysm.

Repair under local anesthesia is the preferred technique. Proximal control is obtained at the level of the inguinal ligament; distal control of the superficial and profunda femoris arteries is usually not needed since control generally requires a fair amount of dissection in a patient under local anesthesia. In addition, simply obtaining proximal control and then approaching the arterial defect directly through the aneurysm cavity simplifies the operation. Once the aneurysm cavity is entered, there is generally substantial back bleeding, even though the common femoral artery is clamped proximally. Bleeding can be controlled with a finger or cottenoid dissector, while the contents of the aneurysm cavity are evacuated and the anterior wall of the artery identified. Identification is a key technical maneuver, since failure to expose the arterial surface leads to placement of sutures in the extravascular tissues, and failure of the repair. Following the identification of the defect in the arterial wall, one or two monofilament sutures placed parallel to the long axis of the artery will generally close the defect. Careful preoperative evaluation of the distal circulation is necessary to assess the possible requirement for distal catheter embolectomy, although embolectomy is rarely indicated. It is also critical to assess the adequacy of the distal circulation following repair. Repair is generally assessed by intraoperative pulse volume recordings. If these suggest a problem at the site of the repair, the surgeon must dissect the branches of the femoral artery, and reopen the femoral artery with the longitudinal arteriotomy. This usually occurs in the setting of a highly diseased femoral artery, and local repair with a patch angioplasty is required.

Infected femoral aneurysms present the surgeon with a difficult clinical problem, the goals of treatment being the eradication of intravascular sepsis and preservation of limb viability. Infected aneurysms in the femoral triangle occur in a variety of clinical settings, including infection of iatrogenic false aneurysm, usually secondary to bacterial contamination from indwelling catheters and sheaths, those occurring in intravenous drug abusers injecting themselves about the femoral triangle, and hematogenous inoculation of a true atherosclerotic aneurysm. Since true atherosclerotic aneurysms are uncommon, superinfection is rare. As stated above, despite the frequency of iatrogenic false aneurysm in a large hospital setting, secondary infection of these lesions is also rare.

The worst form of the disease is that found in intravenous drug abusers. Long-standing intravenous drug abuse combined with factors such as subsequent obliteration of available venous conduits, a proclivity for polymicrobial necrotizing infections, and the sociomedical stratum in which this problem is likely to be encountered make this entity a challenging clinical problem; the dual goals of eradication of arterial sepsis and preservation of limb viability may not be feasible because of the nature of the infection. Infected femoral artery aneurysm in a drug abuser is generally associated with extensive arterial sepsis, frequently with the threat of rupture and bleeding through the overlying infected, and sometimes necrotic, skin. These patients frequently present themselves as a true vascular surgical emergency. The diagnosis may not be obvious, and appropriate blood cultures, careful physical examination for evidence of distal emboli, evaluation for concomitant infective endocarditis, and complete angiographic study to look for additional false aneurysms are indicated.

Clearly, the frequency with which the surgeon encounters this problem will be a function of the patient population he serves. Reddy and colleagues, in reporting the largest clinical series of infected femoral aneurysms in drug abusers, have outlined the principles of treatment in this particular clinical setting. These involve complete arterial excision back to healthy arterial wall, and closure of the transected artery with monofilament suture material, in addition to radical debridement of all surrounding necrotic and infected soft tissue. The obvious dilemma in this circumstance involves a decision about the need for, and technical aspects of, arterial reconstruction to maintain limb viability. Above-knee amputation is necessary in one-third of patients in whom arterial excision involves the common femoral artery bifurcation. If the sepsis is limited to the common femoral artery, such that the distal arterial closure can be completed above the femoral bifurcation, the surgeon can then anticipate continued limb viability by virtue of the circumflex femoral and profunda femoris collateral pathways. Thus, the wisdom of arterial reconstruction is tempered by the absolute need for immediate revascularization, and the nature of the patient. Reddy and colleagues advocate a selective approach to revascularization in the intravenous drug abuser and avoid insertion of prosthetic grafts because of the threat of late infection. Others have adopted a more aggressive policy towards limb

preservation and performed routine reconstruction after excision of the infected false aneurysm.

The technical aspects of arterial reconstruction vary with the clinical setting. The bacteriology and extent of sepsis dictate the feasibility of either *in situ* reconstruction or extra-anatomic reconstruction. One would not hesitate to place a vein patch repair or a saphenous vein interposition graft in a contaminated groin, but in the face of necrotizing arterial sepsis a vein patch would be doomed to fail, exposing the patient to the risk of early hemorrhage. The principles of treatment involve debridement of all infected arterial and soft tissues, placement of suture lines, whether arterial closures or anastomoses, in normal arterial tissue, and the interposition of healthy soft tissue coverage. Transposition of the proximal sartorius muscle over femoral arterial repairs in the groin is a simple, convenient, and effective means of providing soft tissue coverage.

Autogenous tissue should be used when placing a vascular reconstruction in a contaminated field. If necrotizing or frankly suppurative arterial sepsis is present, arterial excision and extra-anatomic reconstruction will be necessary. In this circumstance, preoperative arteriography is mandatory to plan the reconstruction. The obturator foramen bypass provides a convenient and effective means to circumvent arterial sepsis in the groin. Inflow for this bypass may be from either native iliac artery, aorta, or the ipsilateral limb of a previously placed aortofemoral bypass graft. Ideally, the extra-anatomic reconstruction is placed first, and groin exploration is accomplished following completion of the ‘clean' procedure. A prosthetic conduit may be acceptable for an obturator foramen bypass, if the distal circulation is adequate to sustain its patency.

Anastomotic femoral aneurysm

Anastomotic aneurysm occurs when either partial or total disruption of a suture line between a native vessel and vascular graft occurs, vascular integrity being maintained by a fibrous tissue capsule which constitutes the wall of the false aneurysm. Because of lack of tensile strength of this fibrous tissue capsule, the natural history of such an aneurysm is to continue to enlarge. In the femoral region, anastomotic aneurysm most commonly occurs at the site of the femoral anastomosis of an aortobifemoral bypass graft (Fig. 3). Numerous clinical series have reported an incidence ranging from 2 to 8 per cent. In most reports, the femoral anastomotic aneurysm appears at a mean of 5 to 6 years after the initial operation, although this figure, as well as the overall reported incidence, varies with duration of follow-up. In the past, the etiology of anastomotic aneurysm was attributed to the use of silk sutures, and poor quality control in the prosthesis itself. Since silk is a biologic material, it is slowly absorbed with time.

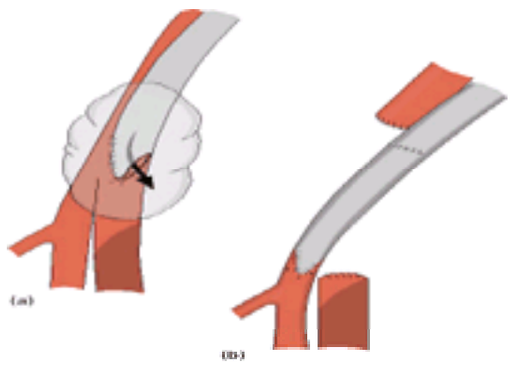


Fig. 3. Anastomotic false aneurysm at femoral anastomosis of prior aortofemoral graft which is the most common location of this lesion.(A) Pathogenesis typically involves partial suture line disruption with false aneurysm contained by surrounding tissues. (B) The most commonly employed method of repair with new interposition Dacron graft sewn end-to-end proximally to old graft and distally to profunda femoris artery in cases with superficial femoral artery occlusion as depicted.

Human beings never ‘heal' a vascular prosthesis; integrity of the suture line is continuously dependent on the integrity of the sutures themselves. With the use of permanent, non-absorbable sutures, and improved prosthetic grafts, the etiology of anastomotic aneurysm has gradually shifted to factors related to the host, rather than the implanted materials. Ongoing degeneration of a diseased arterial wall is the principal cause of development of anastomotic aneurysm in the modern era. However, a number of additional factors and early postoperative complications have been implicated as contributing to false aneurysm formation. These include graft infection, non-infectious postoperative wound complications such as hematoma or seroma communicating with a fresh anastomosis, the need for local endarterectomy at the site of the femoral anastomosis, and inadequate suture bites on a thick, diseased femoral artery. In addition, since the femoral anastomosis is the most common site of an anastomotic aneurysm, physical factors relative to the distraction forces produced by joint motion and compliance mismatch between prosthesis and native artery contribute to false aneurysm formation.

The diagnosis of an anastomotic aneurysm is generally straightforward on the basis of physical examination and other confirmatory tests are not usually needed prior to complete angiography, which is necessary to plan the appropriate surgical therapy. A complete arteriographic study should be performed to search for additional false aneurysms at other suture lines, particularly at the proximal aortic suture line, since a subpopulation of patients afflicted with anastomotic aneurysm are subject to multiple aneurysms and multiple recurrences. As is the case with other false aneurysms, anastomotic aneurysms are likely to enlarge. They can cause symptoms of local compression and arterial ischemia from embolism and *in situ* thrombosis. In addition those located at aortic or iliac suture lines are capable of rupture and/or enteric erosion. Thus, repair of all but the smallest false aneurysms is recommended.

The surgical management of anastomotic aneurysm initially relates to technical steps which are important in prevention of the problem. These have been alluded to above, and involve the use of permanent non-absorbable sutures, large suture bites on diseased femoral arteries, avoidance of excess tension, and careful attention to wound hemostasis and closure at the initial operation. Prophylactic antibiotics are used routinely.

Repair of femoral anastomotic aneurysm may be carried out under either general or regional anesthesia. The old incision is reopened and proximal control of the graft is obtained above the aneurysm. Although dissection in a reoperated groin can be difficult, exploitation of the principles of traction and countertraction, and sharp dissection with a small scalpel blade, are very effective. There is no role for bypassing or avoiding the multiply operated groin simply because it may be difficult: bypassing does not prevent progressive enlargement of false aneurysms, and is likely to result in the sacrifice of the vital profunda femoris artery. Following control of the limb of the old graft and the native common femoral artery, the surgeon decides whether complete dissection of the distal vessels, for distal control, is appropriate. Since distal control may involve extended dissection with the possibility of injury to the profunda femoris artery, we prefer to open the aneurysm after proximal control is obtained and use intraluminal indwelling balloon catheters for distal control. Furthermore, if it subsequently becomes necessary to dissect out additional length of the profunda femoris artery, the balloon catheter serves as a stent of sorts. Initial distal control of the uninvolved profunda femoris artery, and then retrograde dissection to the area of the aneurysm, is neither necessary nor desirable. Following exposure and opening of the false aneurysm, it is important to instill heparinized saline solution into the now clamped limb of the aortofemoral graft, to prevent *in situ* clot development during the period of clamping. At this point, a careful search for contributing factors to false aneurysm formation should be made. If any evidence of infection is present, intraoperative Gram stains should be used to assess the necessity for extra-anatomic reconstruction. If only a portion of the old suture line is disrupted, the surgeon may be tempted to reclose this with additional interrupted sutures. This proclivity is to be avoided, as it increases the chance of a recurrence. Instead, the entire suture line should be taken down, and the native femoral artery trimmed back to healthy artery.

At this point, the surgeon must decide whether a new end-to-side type of reconstruction is feasible, or whether the new reconstruction should be in end-to-end fashion, thereby sacrificing retrograde flow in the native common femoral artery. The guiding principle should be a technically perfect reconstruction to undiseased distal artery, frequently carried out onto the profunda femoris artery in the manner of a profundaplasty (Fig. 3). In such situations, the use of multiple interrupted sutures, at least at the toe and the heel of the reconstruction, is preferred. Reconstruction is generally accomplished with a new short segment of prosthetic graft, which is then simply sutured proximally to the amputated old prosthesis, in end-to-end fashion. Following completion of the reconstruction, adequacy of the distal circulation is ensured before leaving the operating room. Absolute hemostasis is obtained prior to wound closure.

Further reading

Cambria RP, Tilson MD. Obturator foramen bypass grafts in groin sepsis. In: Ernst CB, Stanley JC, eds, *Current therapy in vascular surgery*, pp. 224–30. B.C. Decker, Philadelphia, 1987. [Indications for, and technique of, obturator bypass.]

Cox GS *et al.* Ultrasound-guided compression repair of postcatheterization pseudoaneurysms: results of treatment in one hundred cases. *Journal of Vascular Surgery* 1994; **19**: 683–6. [A large series of ultrasound compression treatment for groin false aneurysm]

Cutler BS, Darling RC. Surgical management of arteriosclerotic femoral aneurysm. *Surgery* 1973; **74**: 764–8. [Emphasizes technical components of repair.]

Feinberg RL *et al.* Color-flow duplex-directed manula occlusion of femoral false aneurysms. *Journal of Vascular Surgery* 1993; **17**: 571–7. [One of the original descriptions of this technique.]

Graham LM *et al.* Clinical significance of arteriosclerotic femoral artery aneurysm. *Archives of Surgery* 1980; **115**: 502–7. [Summarizes the potential complications and indications for treatment.]

Hajarizadeh H *et al.* Ultrasound guided compression of iatrogentic femoral pseudoaneurysm failure, recurrence, and long term results. *Journal of Vascular Surgery* 1995; **22**: 425–33. [Efficacy of ultrasound compression.]

Kent KC *et al.* A prospective study of the clinical outcome of femoral pseudoaneurysms and arteriovenous fistulas induced by arterial puncture. *Journal of Vascular Surgery* 1993; **17**: 125–33. [A valuable study emphasizing the natural history of these lesions.]

Levi N, Schroeder TV. Arteriosclerotic femoral artery aneurysms, a short review. *Journal of Cardiovascular Surgery* 1997; **38**: 335–8. [Most recent comprehensive review.]

Patel KR, Semel L, Clauss RH. Routine revascularization with resection of infected femoral pseudoaneurysms from drug abuse. *Journal of Vascular Surgery* 1988; **8**: 321–8.

Reddy DJ, Smith RF, Elliott JP, Haddad GK, Wanek EA. Infected femoral artery false aneurysm in drug addicts: evolution of selective vascular reconstruction. *Journal of Vascular Surgery* 1986; **3**: 718–24. [These two reports cover the management problems of infected groin aneurysm in drug addicts.]

Schellack J, Salam A, Abouzeid MA, Smith RB, Stewart MT, Perdue GD. Femoral anastomotic aneurysm: a continuing challenge. *Journal of Vascular Surgery* 1987; **6**: 308–17. [A large series of groin anastomotic aneurysm emphasizing modern concepts of pathogenesis and treatment.]

17.6.3 The popliteal artery

Linda Hands

Anatomy
Popliteal aneurysms
Acute thrombosis
Trauma to the popliteal artery
Management
Treatment
Popliteal entrapment syndrome
Diagnosis
Treatment
Functional popliteal entrapment
Outcome of treatment
Cystic adventitial disease of the popliteal artery
Investigation
Treatment
Further reading

Anatomy

The popliteal artery crosses the popliteal fossa closely applied to the back of the femur, the knee joint, and the proximal tibia and tethered proximally by adductor magnus, distally by soleus. This arrangement makes it vulnerable to injury when the knee is dislocated or adjacent bone is fractured. Mechanical stresses imposed by muscle contraction and knee movement could also be responsible for some of the other conditions peculiar to this artery described below.

Popliteal aneurysms (Fig. 1)



Fig. 1. Angiogram of a popliteal aneurysm (by courtesy of Dr E.W. Fletcher).

Popliteal aneurysms rarely endanger life but may threaten limb survival. They are the most common of peripheral aneurysms but occur 15 times less frequently than aortic aneurysms. Nowadays most are ‘atherosclerotic' and 95 per cent occur in men over 60 years. They form part of the spectrum of dilating disease: of those with aortic aneurysms, 10 per cent have popliteal aneurysms; of those with popliteal aneurysms, one-third have aortic aneurysms and 50 per cent have bilateral popliteal aneurysms. Occasionally, mycotic or pseudoaneurysms occur in younger individuals, with pseudoaneurysms sometimes associated with osteochondromas of the distal femur.

The normal popliteal artery measures 0.7 cm in diameter and is often difficult to palpate. A localized expansion over 1.05 cm constitutes a popliteal aneurysm, but may be clinically undetectable until it reaches 2 cm unless ultrasonographic examination is used in patients suspected to harbour the disease.

There are four main complications of popliteal aneurysms: (1) acute thrombosis; (2) distal embolization; (3) rupture; (4) local pressure effects.

Acute thrombosis

Acute occlusion of the aneurysm is probably the most common complication and has a variable outcome. Some patients are unaware of the event and a thrombosed aneurysm is detected during unrelated investigations at a later stage. At the other extreme a significant number of patients require amputation as a consequence of popliteal arterial thrombosis. The latter we know about: the number who undergo silent thrombosis in the community is unknown.

The classical presentation is of a man in his eighth decade, previously untroubled by claudication or other vascular symptoms, who suddenly develops a cold, painful leg. Examination reveals a pale limb with no pulses below the femoral (the thrombosed popliteal aneurysm is usually undetectable), perhaps loss of sensation and movement distally, and in some cases a pulsatile mass behind the other knee. Differentiation between thrombosed aneurysm, embolization from a proximal source such as the heart, and thrombosis on pre-existing atheroma can be difficult, but the treatment is similar whatever the cause.

Propagated fresh thrombus or previous distal embolization (see later) often impedes run-off from the popliteal artery and may jeopardize the success of bypass grafting. Thrombolysis is a useful adjunct to surgery in this situation. Preoperative thrombolysis in the patient with intact sensation and power can open up some or all of the crural vessels. Some patients, however, are made worse by this procedure, probably as a result of loosening and embolization of thrombus from the aneurysm. When there is a large thrombotic load, perioperative thrombolysis of the run-off vessels alone is probably safer. The patient who has a critically ischaemic limb with loss of sensation and movement, or muscle pain, cannot afford the delayed revascularization that preoperative thrombolysis would inevitably incur. Under such circumstances, surgical exploration, with the use, as appropriate, of on-table angiography, an embolectomy catheter, peroperative thrombolysis, and bypass grafting, is indicated.

The popliteal artery can be explored through a medial approach, exposing it above and below the knee, anastomosing a bypass graft in an end-to-side fashion at these two points, and excluding the intervening aneurysmal artery by ligation. The alternative is a posterior approach to expose the popliteal aneurysm directly: the aneurysm is opened and an inlay graft performed. Overall, 5-year graft patency is about 60 per cent, but it is consistently better with vein than with synthetic grafts.

Distal embolization

Acute or chronic distal embolization from a popliteal aneurysm is a common form of presentation. The patient complains of claudication or pain at rest as pedal and crural vessels silt up. Treatment is surgical exclusion of the aneurysm (see above) but bypass grafting may again be frustrated by poor run-off. Thrombolysis is a useful adjunct, but is often less successful because of the long-standing nature of the distal obstruction.

Rupture

This is a rare event accounting for fewer than 5 per cent of presentations of popliteal aneurysm. Blood loss is usually relatively small because the leak is tamponaded by the tight musculofibrous planes in this region. The increase in local pressure may lead to venous compression, occasionally with deep venous thrombosis, and nerve damage. Treatment is emergency surgery as above.

Local pressure symptoms

A large popliteal aneurysm can impede knee flexion or compress the tibial nerve, causing paraesthesias or foot drop. Popliteal aneurysms over 2 cm in diameter compress and displace the popliteal vein but a collateral venous network usually opens up to allow normal venous function. A few of the larger aneurysms eventually cause venous thrombosis. Such aneurysms are usually repaired electively to avoid further complications but the outcome for nerve and venous function is unpredictable.

Asymptomatic popliteal aneurysms

Twenty years ago, those popliteal aneurysms detected were either large or symptomatic. Ultrasonography has brought to our attention a large number of smaller, asymptomatic aneurysms and raised the question of their management.

Some 15 to 20 per cent of the patients who present with thrombosed aneurysms lose their leg. Increased use of thrombolysis may improve this outcome but will do nothing for the 5 per cent or so who have an unsalvageable leg at presentation. Such data have led many surgeons to advocate prophylactic repair of all asymptomatic aneurysms. They argue that a bypass graft excluding the popliteal aneurysm will guard against future catastrophe and should have good run-off at this stage and therefore good long-term patency. Critics of this approach point to the 60 per cent patency of below-knee femoropopliteal bypass grafts and question the inevitability of complications in all popliteal aneurysms.

The problem is that most of our data on initially asymptomatic aneurysms are drawn from patients with large aneurysms, detected during routine clinical examination, or those with generalized aneurysmal disease in whom the detection of one aneurysm often prompts a search for others. Generalized aneurysmal disease and large size are both known to increase the rate of complications from popliteal aneurysms. Even these data, however, show that some patients with popliteal aneurysms can remain asymptomatic for 15 years. Earlier studies, where aneurysms were detected clinically rather than by ultrasound and tended to be larger, found a complication rate of 30 per cent over 3 to 5 years. More recent studies include aneurysms detected by ultrasound, where the mean size is smaller, and report a complication rate of 8 per cent over 3 years.

Such retrospective studies of asymptomatic aneurysms have shown that those over 3 cm in diameter, tortuous aneurysms with significant narrowing of their lumen, and aneurysms with intraluminal thrombus are more likely to threaten limb viability. Vein bypass of an asymptomatic aneurysm has a 90 per cent 5-year patency. An operative policy based on the above criteria represents a reasonable balance between the risks of surgery and of conservative management.

Developments in treatment

Some centres have reported stenting of patent popliteal aneurysms, using a device covered by synthetic graft material. Antegrade percutaneous placement via the femoral artery adds the risk of distal embolization when the device is negotiated through the aneurysm. This has been addressed by retrograde placement via a surgical approach to the distal popliteal artery. Although appealing in its simplicity, the long-term outcome is yet to be determined. We know that synthetic grafts used conventionally to bypass a popliteal aneurysm have disappointing long-term patency. Use of vein-covered stents may provide a more durable alternative at some future stage.

Trauma to the popliteal artery

Damage to the popliteal artery is associated with a high rate of amputation. There are three categories of injury that have been particularly reported in largely retrospective studies: blunt and penetrating injuries, usually in young adult males as a result of civilian or military violence, and injury incurred as a result of total knee replacement in an elderly population.

Military trauma is likely to be a combination of blunt and penetrating injuries, often with a significant delay between injury and repair because evacuation to hospital proves difficult. The nature of civilian violence depends on environment: in cities in the United States penetrating damage due to gunshot wound is common whereas, in other places, blunt injuries from agricultural, industrial, and road traffic accidents predominate.

Injuries to the popliteal artery were treated in the Second World War by arterial ligation and 70 per cent of cases came to amputation. Reconstruction is usually attempted nowadays, resulting in improved rates of limb salvage. Penetrating trauma is associated with a relatively good outcome: amputation rates of 0 to 18 per cent have been reported. Blunt trauma has a poorer prognosis, with amputation rates ranging from 20 to 35 per cent. The likelihood of amputation depends upon the extent of soft-tissue and bony injury, the size and nature of any open wound, the presence of obvious distal ischaemia or accompanying shock, and the age of the patient. Delayed revascularization of 12 h or more is also a poor prognostic factor but found usually only with military injuries. Attempts have been made through the creation of prognostic scoring systems to decide which patients would benefit from primary amputation. Few of these have been evaluated prospectively and, as one might expect from the number of such systems, none is easily and reliably used in the case of individual patients.

Ischaemia associated with total knee replacement is rare (less than 0.2 per cent of cases) but associated with a high rate of amputation (death or amputation in 25 per cent). The cause is usually thrombosis of a diseased popliteal artery, especially if a tourniquet is used. There is often no history to suggest a vascular problem, although preoperative examination may reveal absent pulses. In a few cases, musculofascial compression of the artery occurs when a knee previously held in fixed flexion is extended. Distal embolization of plaque or thrombus from the popliteal artery, arterial transection, and pseudoaneurysm formation occasionally occur.

Management

Skeletal muscle does not tolerate more than 6 h of ischaemia, and, although collateral blood supply may maintain viability beyond that time, it cannot be relied upon when there is extensive soft-tissue injury. It is therefore essential to recognize the possibility of trauma to the popliteal artery with particular injuries, and to diagnose and treat as soon as possible.

Bony injuries around the knee are associated with arterial trauma in 2 per cent cases but in knee dislocation the rate increases to 30 to 50 per cent. ‘Hard evidence' evidence for disruption of the popliteal artery comes from (a) local signs of pulsatile bleeding, an expanding or pulsatile haematoma or the palpation of a thrill from an arteriovenous fistula, or (b) distal ischaemia with reduction or loss of foot pulses. It is important to remember that 10 per cent of patients with significant trauma to the popliteal artery have apparently normal foot pulses.

Is preoperative angiography necessary? Patients with the ‘hard' signs noted above are probably best served by immediate surgery. Preoperative angiography provides little more useful information and delays revascularization. The only exception is with multiple gunshot wounds when the site of arterial injury is unclear. If there is doubt during surgery about the site of arterial injury, perioperative angiography can be performed.

When managing patients with ‘soft' signs—non-expanding haematoma, nerve damage, or proximity of the wound to the popliteal artery but, by definition, full pedal pulses with no evidence of distal ischaemia—immediate intervention is less important, and angiography or Duplex ultrasonography (if feasible) can be pursued. Some argue that angiography is unnecessary in such cases, but the presence of foot pulses does not rule out arterial injury and in one study 75 per cent patients with ‘soft' signs had arterial injury. Narrowing of the popliteal artery on Duplex or angiography should never be ascribed to ‘spasm' even in the presence of compartment syndrome: the vessel should always be explored as there is frequently an intimal tear. Even small intimal tears often lead to later thrombosis or the formation of a false aneurysm.

Treatment

The popliteal artery is usually explored for trauma via a medial incision, with exclusion bypass grafting using vein wherever possible. Some report a posterior approach to the artery but this poses difficulties if a bypass graft to the crural vessels is required. The artery may be completely or partially divided, or have an intimal tear with intact adventitia. It is sometimes possible to perform a direct suture repair of a lacerated or divided artery. Simple endarterectomy for intimal tears rarely provides a satisfactory long-term result. Most cases of arterial trauma require a vein bypass graft, either as an interpositional end-to-end graft or as an end-to-side exclusion graft. Peroperative local or systemic heparinization and even thrombolysis, when necessary, improve the prognosis without a significant increase in risk.

Damage to the popliteal vein from blunt or penetrating trauma occurs in 25 to 50 per cent of cases of trauma to the popliteal artery. The vein should be repaired wherever possible, using saphenous vein from the other leg for both arterial and venous repairs. A third of cases in which the popliteal artery is ligated develop postphlebitic syndrome. In an acutely ischaemic limb with both arterial and venous damage, the insertion of a Javid shunt into the popliteal artery will allow reperfusion

of the limb whilst the vein is repaired.

Four-compartment fasciotomy of the calf is required in approximately a third of cases, particularly those with extensive soft-tissue injury or in whom the vein had to be ligated. Fractures may need to be stabilized or drawn out to length. This should be achieved with external fixation to avoid damage to collateral vessels, preferably after arterial reconstruction. Where necessary, primary nerve repair should be performed after vascular reconstruction.

Popliteal entrapment syndrome

Congenital anatomical abnormalities of muscle and artery in the popliteal fossa together comprise an uncommon cause of compression of the popliteal artery. Compression occurs most commonly because of an abnormal attachment of the medial head of gastrocnemius to the lateral femoral condyle (type 2), causing the popliteal artery to pursue a more lateral course than normal. Sometimes the popliteal artery itself appears at fault, taking an unusual medial course around a normally attached medial head of gastrocnemius. A variety of other causes, including compression by accessory slips of gastrocnemius, popliteus, or plantaris muscle, are also described (Fig. 2).

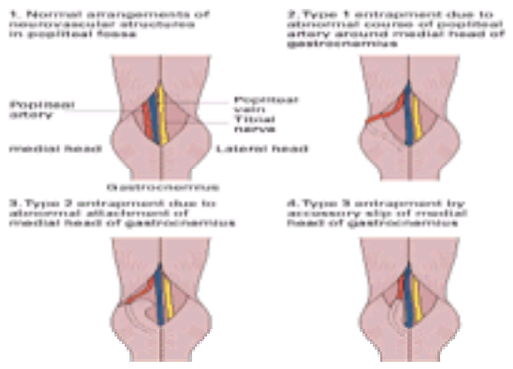


Fig. 2. Types of popliteal artery entrapment.

The prevalence of entrapment of the popliteal artery in post-mortem studies is only 3.5 per cent. Amongst the living it comes to notice when symptoms of restricted blood flow arise. The majority report a cold foot with tingling and cramps following exercise; others report claudication, and relatively few develop rest pain or ulceration. The largest single-centre series amounts to only 30 patients, and entrapment probably accounts for less than 1 per cent patients with claudication. Patients usually present in their late teens or early twenties but some as young as 12 years and as old as 70 years have been described. Males are up to 15 times more likely to be affected than females, and as the entrapment tends to produce symptoms when physical activity is increased, it has been well described in military recruits. The condition occurs bilaterally in somewhere between 20 and 70 per cent of cases.

Post-stenotic aneurysm formation occurs in 15 per cent of patients. Although restricted blood flow at the site of constriction is the usual cause of symptoms, distal embolization of thrombus, from the site of constriction or a post-stenotic aneurysm, leads to impaired run-off in the crural vessels in 30 per cent of patients.

Diagnosis

Claudication-like symptoms in a young adult should raise the suspicion of an entrapment syndrome. Duplex ultrasonography or angiography (Fig. 3) of the popliteal artery during active plantarflexion or passive dorsiflexion of the foot will demonstrate compression of the artery and, in some, an abnormal course of the artery, local thrombus, or post-stenotic aneurysmal dilatation.

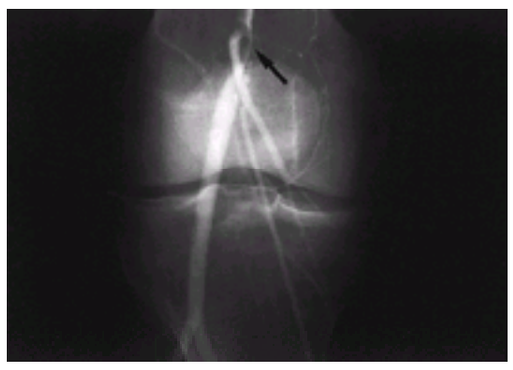


Fig. 3. Angiogram demonstrating a type 1 entrapment of the popliteal artery (by courtesy of Dr E.W. Fletcher).

Treatment

Exploration of the popliteal fossa via a posterior approach allows release of the artery by division of any constricting muscle or fibrous band. In patients without mural thrombus, occlusion, or aneurysmal dilatation (between a third and two-thirds of all patients presenting), this will suffice. Those who have significant arterial damage require bypass grafting. Simple thromboendarterectomy was favoured in the past but the long-term patency was poor compared with saphenous vein bypass. The short saphenous vein is more accessible than the long saphenous vein through a posterior approach but may be too small to use. If a distal bypass graft is required because of run-off disease, a medial approach to the popliteal artery is more appropriate.

Functional popliteal entrapment

Reliance on Duplex ultrasonography alone in making the diagnosis has certain pitfalls. Several studies of healthy young adults, both normally active and athletic, have shown that active plantarflexion produces stenosis of the popliteal artery in all and occlusion in 50 per cent, at the level of the soleal sling (the proximal border of the soleus muscle). Turnipseed reported a series of athletes with calf cramping after exercise in whom surgical exploration revealed no anatomical abnormality but whose symptoms were relieved by division of the soleal sling as it crosses the neurovascular bundle. Compression at this level has since been shown to be almost as common in untrained individuals, but usually without symptoms. Functional popliteal entrapment requires active plantarflexion and is rarely seen with passive dorsiflexion, unlike ‘anatomical' popliteal entrapment. Turnipseed's work would suggest that in those with convincing symptoms, Duplex evidence of compression of the popliteal artery is sufficient to justify surgical exploration.

Outcome of treatment

Those patients requiring only musculofascial release have a good outcome, with 95 per cent patency at 5 years. Unfortunately, those who require bypass grafting fare less well: 65 per cent are patent at 5 years. For young adults this is a significant problem.

Patients who have arterial damage requiring bypass grafting are more likely to have advanced symptoms such as short-distance claudication or rest pain. Patients without arterial damage usually have relatively minor symptoms of cramps and cold feet. It would seem sensible to attempt diagnosis of this condition at an early stage and to take these symptoms seriously in teenage or young adult patients.

Cystic adventitial disease of the popliteal artery

Cystic adventitial disease is a rare disease of uncertain cause. It affects the external iliac, femoral, radial, and ulnar arteries but 85 per cent of cases occur in the popliteal artery. Multiple cysts within the adventitia, perhaps of synovial origin, protrude into the lumen and sometimes impair blood flow.

The disease probably accounts for 1 in 1200 cases of calf claudication. It usually presents in young or middle-aged males who have a history of relapsing and remitting calf claudication over several years but who are completely asymptomatic between episodes. During a symptomatic phase, pedal pulses may be absent, the foot cool, and ankle/brachial pressure indices reduced, but during remission the pulses return and the limb appears normal. Eventually, thrombosis of the popliteal artery may occur, with the development of permanent symptoms and signs. Sometimes there are also symptoms of nerve compression by the cyst.

Investigation

In addition to the peculiar history of ‘disappearing' claudication, ‘Ishikawa's sign' may be positive—flexion of the knee causes the pedal pulses to disappear and the foot to become cold and pale.

During a symptomatic phase, Duplex ultrasonography or angiography (Fig. 4) reveal smooth, rounded, luminal compression—the ‘scimitar sign'. A large cyst may produce an ‘hour-glass' compression of the lumen. In 30 per cent of cases the artery is occluded, with a smooth proximal taper by the time of investigation.



Fig. 4. Angiogram demonstrating cystic adventitial disease of the popliteal artery (by courtesy of Dr E.W. Fletcher).

Computed tomography or magnetic resonance imaging usually confirm the diagnosis by demonstrating cystic changes within the arterial wall.

Treatment

Balloon angioplasty is ineffective because of the compliant nature of the cyst. It may be possible to relieve a stenosis by percutaneous aspiration of the cyst under ultrasonographic control. If this fails, or the artery has thrombosed, the affected segment needs to be excised and replaced by vein graft.

Further reading

Popliteal aneurysm

Dawson I, Sie RB, van Bockel JH. Atherosclerotic popliteal aneurysm. *British Journal of Surgery* 1997; **84**: 293–9. [Systematic review of recent studies since 1980 of patients with popliteal aneurysms.]

Ramesh S, Michaels JA, Galland RB. Popliteal aneurysm: morphology and management. *British Journal of Surgery* 1993; **80**: 1531–3. [Review of experience with popliteal aneurysms presenting to one hospital and including factors predictive of thrombosis.]

Varga ZA *et al.* A multicenter study of popliteal aneurysms. *Journal of Vascular Surgery* 1994; **20**: 171–7. [A multicentre prospective study looking at presentation, management, and outcome of patients with popliteal aneurysms.]

Popliteal artery trauma

Degiannis E, Velmahos GC, Florizoone MGC, Levy RD, Ross J, Saadia R. Penetrating injuries of the popliteal artery: the Baragwaneth experience. *Annals of the Royal College of Surgeons of England* 1994; **76**: 307–10. [Study of popliteal artery trauma due to civilian violence in South Africa.]

Faris IB, Raptis S, Fitridge R. Arterial injury in the lower limb from blunt trauma. *Australia and New Zealand Journal of Surgery* 1997; **67**: 25–30. [Retrospective study of patients presenting with lower limb injury to a single hospital with the purpose of devising a scoring system predictive outcome.]

Holmberg A, Milbrink J, Bergqvist D. Arterial complications after knee arthroplasty: 4 cases and a review of the literature. *Acta Orthopaedica Scandinavica* 1996; **67**: 75–8. [Literature review and four case reports of arterial complications after total knee replacement.]

Popliteal artery entrapment syndrome

Di Marzo L, Cavallero A, Mingoli A, Sapienza P, Tedesco M, Stipa S. Popliteal artery entrapment syndrome: the role of early diagnosis and treatment. *Surgery* 1997; **122**: 26–31. Retrospective study of patients presenting with popliteal entrapment syndrome to one institution, looking at presentation, management, and outcome.]

Erdoes LS, Devine JJ, Bernhard VM, Baker MR, Berman SS, Hunter GC. Popliteal vascular compression in a normal population. *Journal of Vascular Surgery* 1994; **20**: 978–86. [Study of young asymptomatic subjects, either normally active or athletically trained, looking at the effect of active and passive dorsi flexion and plantar flexion of the foot on Duplex detected blood flow in the popliteal artery.]

Turnipseed WD, Pozniak M. Popliteal entrapment as a result of neurovascular compression by the soleus and plantaris muscle. *Journal of Vascular Surgery* 1992; **15**: 285–93. [Review of trained athletes presenting with popliteal artery entrapment to one institution, looking at the usefulness of various investigations and the results of surgery.]

Cystic adventitial disease

Parks RW, D'Sa AA. Critical ischaemia complicating cystic adventitial disease of the popliteal artery. *European Journal of Vascular Surgery* 1994; **8**: 508–13. [Description of two cases presenting with cystic adventitial disease of the popliteal artery and a review of the subject.]

17.6.4 Carotid artery

Antony Fox

Further reading

Aneurysms of the extracranial carotid arteries are uncommon and account for less than 4 per cent of peripheral aneurysms. As a result, few centres have enough experience to provide epidemiological data for carotid aneurysms, the largest single reported series representing 37 of 8500 aneurysms treated at Baylor University over a 21-year period.

Atherosclerosis is responsible for 46 to 70 per cent of all carotid aneurysms. Trauma and previous carotid surgery (endarterectomy) are less commonly responsible, and syphilis has declined in importance. Congenital, mycotic, post-stenotic, postradiotherapy, and arteriovenous aneurysms have been identified. Aneurysms associated with rarer degenerative and dysplastic conditions (Fig. 1), such as Marfan syndrome, Ehlers–Danlos syndrome, and tuberous sclerosis, have also been described.

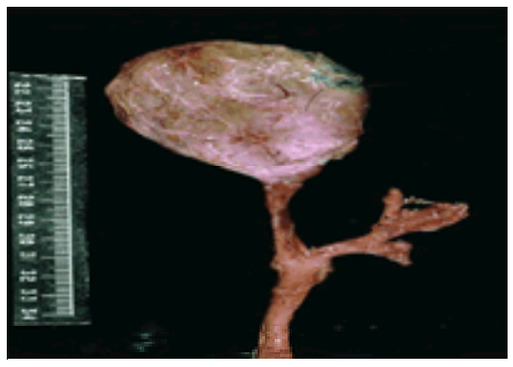


Fig. 1. Large aneurysm of the internal carotid artery found in a 45-year-old man who presented with a dense contralateral hemiparesis.

Most carotid aneurysms present as an asymptomatic, pulsatile swelling in the anterior triangle of the neck just below the angle of the mandible. They may be fusiform or saccular and generally involve the common carotid artery, bulb, or internal carotid arteries. They are bilateral in about 10 per cent of cases. A bruit may be audible. Neurological manifestations are the presenting symptoms in over 50 per cent of patients; these develop secondary to embolization of aneurysmal thrombus (amaurosis fugax and transient ischaemic attacks) or secondary to aneurysm thrombosis (stroke). Carotid aneurysm may also present as a peritonsillar abscess if it progresses medially or compresses the larynx, causing wheezing and dyspnoea. Pain may be referred to the angle of jaw and behind the ear. Hoarseness, dysphagia, weakness of the tongue, and Horner's syndrome may be produced by pressure on the vagus, glossopharyngeal, hypoglossal, and sympathetic nerves. Rupture is rare, but may present in both atherosclerotic and infected false aneurysms.

Elongation with kinking or a twist is the most frequent lesion to masquerade as an aneurysm of the carotid artery (Fig. 2), typically found in the root of the neck on the right in elderly hypertensive women. Other differential diagnoses include a prominent bifurcation of the carotid artery, tumours of the carotid body (chemodectoma) and jugular bulb, lymphadenopathy, branchial-cleft cysts, or other masses overlying the carotid vessels. Careful palpation can discriminate between these entities.



Fig. 2. Angiogram in a 70-year-old woman who presented with a pulsatile swelling in the root of the neck on the right, showing a tortuous origin of the common carotid and subclavian artery.

The investigation of choice is duplex sonography: this is non-invasive, provides useful information about the size and extent of the aneurysm, and allows assessment of the contralateral carotid vessels. Angiography is used to provide accurate morphological details (Fig. 3), with digital-subtraction techniques providing enhanced images. Conventional or spiral CT and MRI can provide excellent images, particularly of the distal extent of highaneurysms.



Fig. 3. Aneurysm that has occurred in a coiled, redundant, internal carotid artery.

Symptomatic aneurysms generally need repair, but the low incidence and diverse causes of asymptomatic aneurysms create problems in defining their need for treatment. Frequent fatality has been demonstrated with an expectant approach, but techniques have now been described that enable discrimination between physiological dilatation and pathological aneurysm formation. Prevention of the neurological sequelae of embolization from the aneurysm wall should also include the use of antiplatelet agents such as aspirin.

The earliest reported surgical treatment was simple ligation of the carotid artery. This operation was first reported by Pare in 1552 for trauma and was carried out by Sir Astley Cooper in 1805 for a carotid aneurysm. His first attempt resulted in hemiplegia and early postoperative death, but he successfully carried out the procedure on a

similar case 3 years later. Despite the high mortality rate from stroke, ligation of the common carotid artery remained the mainstay of treatment for over 150 years.

Resection of the aneurysm with restoration of flow by direct anastomosis or graft is now considered the first choice for treatment. Primary anastomosis may be possible in up to 50 per cent of patients, as elongation of the artery often accompanies aneurysmal dilatation. Where this cannot be achieved, interpositional replacement using autologous external carotid artery, long saphenous vein, or a prosthetic graft is necessary. All patients receive short-term anticoagulation to prevent early distal re-thrombosis. The use of prosthetic materials should be avoided for false aneurysms because infection is a possible aetiological factor. High aneurysms may require the removal of part of the mastoid bone or dislocation of the mandible. Gradual occlusion over a period of days (Crutchfield clamp) can be used. Ligation can then be combined with extra- to intracranial bypass surgery. Interventional techniques are now gaining in popularity, using covered stents, coils, and intraluminal balloon trapping.

Quality-control methods, including stump pressure, electroencephalography, transcranial Doppler and cerebral oximetry, can all be employed during surgery to assess the requirement for a perioperative shunt. Outcome data for operative mortality and disabling stroke using these techniques should be equivalent to those for carotid endarterectomy, that is, less than 1 to 1.5 per cent mortality and less than 5 per cent neurological complications.

Further reading

Bergan JJ, Hoehn JG. Evanescent cervical pseudoaneurysms. *Annals of Surgery* 1965; **162**: 213–7. [Account of presentation and management of false extracranial carotid aneurysms.]

Crandon IW, Teasdale E, Galbraith SL, Hadley DM. Carotid traumatic aneurysm treated by detachable balloon. *British Journal ofNeurosurgery* 1988; **2**: 507–11. [Case of a young man who developed a large infraclinoid carotid aneurysm following blunt trauma presenting with epistaxis.]

de Jong KP, Zondervan PE, van Urk H. Extracranial carotid artery aneurysms. *European Journal of Vascular Surgery* 1989; **3**: 557–62. [Guidelines and suggested criteria for discrimination between physiological and pathological aneurysm formation.]

Hamer JD. Carotid aneurysms. In Horrocks M, ed. *Arterial aneurysms*, pp. 183–93. Butterworth-Heinemann, Oxford, 1995. [A comprehensive guide to arterial aneurysms with subspecialty sections.]

Hunt JL, Snyder WH. Late false aneurysms of the carotid artery: repair with extra-intracranial arterial bypass. *Journal of Trauma* 1979; **19**: 198–200. [Case report presenting alternative approach for the management of false carotid aneurysms.]

Liapis CD, Guugulakis A, Misiakos E, Verikokos C, Dousaitou B, Sechas M. Surgical treatment of extracranial carotid aneurysms. *International Angiology* 1994; **13**: 290–95. [Twenty-three year experience of extracranial carotid bifurcation aneurysms. Presentation and management.]

McCollum CH, Wheeler WG, Noon GP, DeBakey ME. Aneurysms of the extracranial carotid artery. *American Journal of Surgery* 1979; **137**:196–200. [Twenty-one year experience of 37 extracranial carotid aneurysms.]

Mishaly D, Pasik S, Barzilai N, Mashiah A. Repair of internal carotid aneurysm under local anaesthesia. *Journal of Cardiovascular Surgery (Turin)* 1992; **33**: 380–2. [Report of true internal carotid aneurysm repaired with internal jugular vein graft.]

Sharma S, Rajani M, Mishra N, Sampathkumar A, Iyer KS. Extracranial carotid artery aneurysms following accidental injury: ten years experience. *Clinical Radiology* 1991; **43**: 162–5. [Review of 8 post-traumatic extracranial carotid aneurysms over a 10-year period.]

Trippel OH, Haid SP, Kornmesser TW, Bergan JJ. Extracranial carotid aneurysms. In Bergan JJ, Yao JST, eds. *Aneurysms: diagnosis and treatment*, pp. 493–504. Grune and Stratton, New York, 1982. [A review of epidemiology and management of extracranial carotid aneurysms.]

17.6.5 Subclavian artery

Antony Fox

Further reading

Aneurysms of the subclavian artery represent about 1 per cent of all peripheral arterial aneurysms. Their causes, presentation, and treatment can be classified in morphological terms as those of the intrathoracic and those of the extrathoracic portion of the subclavian artery. They are rare in childhood, increasing in incidence with age.

The subclavian artery may be compressed at the thoracic outlet by a cervical rib, an abnormal first rib, or muscular and fibrous bands associated with the scalenus anterior or medius muscles. Post-stenotic dilatation of the subclavian artery may occur. The ‘thoracic outlet syndrome' or ‘scalene syndrome' and previous trauma account for about three-quarters of extracranial aneurysms (Fig. 1 and Fig. 2) and are generally seen in young adults. Atherosclerosis accounts for the majority of the aneurysms in elderly people. Other, rarer causes include congenital (aneurysms of an aberrant right subclavian artery), infective, and degenerative conditions (Behçet's syndrome and Takayasu's arteritis), and there is an increasing number of reports on the adverse sequelae of central venous catheters in relation to subclavian aneurysm.



Fig. 1. Traumatic aneurysm of the left subclavian artery (by courtesy of Professor Sir Peter Morris).



Fig. 2. Angiogram of patient in Fig. 1 showing a traumatic false aneurysm of left subclavian artery (by courtesy of Professor Sir Peter Morris).

An incidental finding of a mediastinal shadow on routine chest radiography is the most common presentation for intrathoracic aneurysms. They are largely asymptomatic, but may present with chest and back pain, Horner's syndrome, venous congestion, and hoarseness secondary to local compression. Distal embolization to the arm is unusual at this site, but is seen in about two-thirds of patients with extrathoracic aneurysms (Fig. 3). The signs must be differentiated from Raynaud's phenomenon, which also presents with episodic symptoms related to the digital vessels. Neurological symptoms of brachial-plexus compression are present in 90 per cent patients with ‘thoracic outlet syndrome' and so the relative importance of the aneurysm may be difficult to determine. Large aneurysms may result in subclavian-steal syndrome. Rarely, the aneurysm ruptures. Intense local pain and swelling may be accompanied by signs of compression of the brachial plexus and ischaemia of the arm. Rupture into the apex of the lung presenting with haemoptysis has been reported.



Fig. 3. Digital necrosis and ischaemia secondary to distal embolization.

Careful physical examination may reveal a pulsatile mass in the supraclavicular fossa, a bruit, or pulse deficits, and there may be evidence of associated compression of the thoracic outlet or upper mediastinum. Coexisting aneurysms should be identified at other sites. Plain radiographs of the thoracic outlet and upper mediastinum may show the extent of the mass, a calcified arterial wall, or cervical ribs. Further morphological data can be acquired by using spiral CT. Magnetic resonance imaging can detect fibrous bands and soft-tissue anomalies causing deviation and compression of local structures. Conventional or digital-subtraction angiography will demonstrate the arterial anatomy, the extent of aneurysmal disease, associated stenosis or thrombus, and the patency of the distal vessels (Fig. 4).

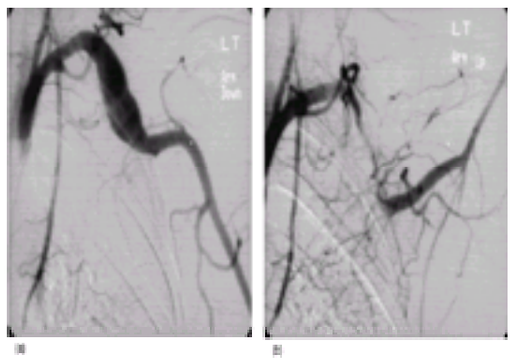


Fig. 4. (a) Post-stenotic dilatation of left subclavian and axillary arteries secondary to thoracic outlet syndrome. (b) Obliteration of subclavian/axillary aneurysm with abduction of the arm.

Intra-arterial thrombolysis may be administered in cases of acute ischaemia of the limb secondary to aneurysmal thrombosis. However, elective surgical treatment is recommended for most subclavian aneurysms. Intrathoracic aneurysms are approached through a lateral thoracotomy on the left or a median sternotomy for the right side. The approach for the extrathoracic subclavian artery varies according to the morphology and cause of the aneurysm. It can be approached through a supraclavicular incision after medial retraction of the phrenic nerve and division of the scalenus anterior. Additional exposure can be achieved by resection of the middle third of the clavicle or by dissection of the axillary artery below the clavicle. In cases of 'thoracic outlet syndrome' it may also be necessary to resect a cervical rib or band, or the first rib if this is implicated. A transaxillary approach may be utilized for this procedure.

The aim of the surgery is to exclude or remove the aneurysm, with restoration of arterial continuity. Direct arterial repair or patch angioplasty may be appropriate in some patients with false aneurysms. In most cases, resection is possible, the arterial supply being maintained by either direct anastomosis or replacement with a venous or prosthetic graft. Rarely, ligation of the artery proximally and distally with excision of the aneurysm can be performed if there is sufficient collateral circulation. This approach can also be used in hazardous cases, the circulation being restored by extra-anatomical axilloaxillary or caroticosubclavian bypass when the collateral supply does not suffice. Percutaneous embolization using steel coils has also been reported.

Further reading

Clagett GP. Upper extremity aneurysms. In Rutherford RB, ed. *Vascular surgery*, 4th edn, pp. 1112–18. Saunders, Philadelphia,1995. [An overview of upper limb arterial aneurysms.]

Dougherty MJ, Calligaro KD, Savarese RP, DeLaurentis DA. Atherosclerotic aneurysm of the intrathoracic subclavian artery. A case report and review of the literature. *Journal of Vascular Surgery* 1995; **21**: 521. [A case of surgically treated asymptomatic atherosclerotic aneurysm of the left subclavian artery, with a review of the literature.]

Fedullo LM, Meranze SG, McLean GK, Burke DR. Embolisation of a subclavian artery aneurysm with steel coils and thrombin. *Cardiovascular and Interventional Radiology* 1987; **10**: 134–7. [Report describing endovascular management for a subclavian artery aneurysm.]

Hobson RW, Israel MR, Lynch TG. Axillosubclavian arterial aneurysms. In Bergan JJ, Yao JST, eds. *Aneurysms: diagnosis and treatment*, pp. 435–47. Grune and Stratton, New York, 1982. [A review of upper limb inflow arterial aneurysms covering epidemiology and management.]

Scher LA *et al.* Staging of arterial complications of cervical rib: guidelines for surgical management. *Surgery* 1984; **95**: 644–9. [A series of 12 patients with 15 arterial lesions is reviewed and a staging system proposed to provide guidelines for managing patients with this condition.]

17.6.6 Visceral arteries

Jack Collin

Introduction
Aetiology
Relative incidence of individual aneurysms
Clinical presentation
Diagnosis
Management
Further reading

Introduction

The visceral arteries comprise the coeliac, superior mesenteric, inferior mesenteric, and renal arteries, together with all their branches. Aneurysms of these vessels are uncommon and in order of frequency rank below aneurysms of the abdominal and thoracic aorta, and those of the iliac, femoral, and popliteal arteries.

In common with other arterial aneurysms the majority are asymptomatic, and since no easy, non-invasive investigation can reliably detect their presence their true prevalence in the community is unknown. Any epidemiological data that exist are suspect since they are obtained from highly selected groups of patients undergoing angiography or computerized tomography, or from autopsy studies in which small visceral aneurysms can easily be overlooked. No single vascular surgical unit has sufficient experience of visceral arterial aneurysms to allow useful analysis of the relative importance of different causes or the indications for, and results of, various management options. Most accounts in the literature comprise single case reports, small series, or analyses of cases collected from literature reviews. Any overview that attempts to give a didactic account of the aetiology, presentation, and management of visceral arterial aneurysms is therefore handicapped by the absence of a sound database and this fact should be borne in mind.

Aetiology

In common with aneurysms of the aorta, the aetiology in the majority of cases is unknown, but the proximate cause is local failure of the connective tissue of the arterial wall to maintain the integrity of the vessel. Most patients are middle aged and, inevitably, atherosclerosis has been implicated, but in contradistinction to abdominal aortic aneurysms, a high proportion of cases occur in young people in their 30s and the majority of patients are women.

The single most interesting fact about this disease is that, in developed countries, ruptured visceral arterial aneurysm is now one of the major causes of maternal mortality. Between 1967 and 1982 there were 14 maternal deaths from this cause (10 splenic, three renal, and one hepatic arterial aneurysm) in England and Wales, with most ruptures occurring in the last weeks of pregnancy. Similar observations in the United States of America have given rise to speculation that systemic and portal hypertension, increased cardiac output, and increased blood flow in some vessels, together with the hormonal and connective-tissue changes in pregnancy, may all be contributory factors.

A number of other causes have been implicated with differing importance in various reports, including congenital malformations, trauma, arteritis, connective-tissue disorders, Behçet's disease and infection. Trauma, either accidental or iatrogenic during percutaneous biopsy, plays a major part in the causation of aneurysms within the liver or kidney but, fortunately, the majority are small and of no clinical significance (Fig. 1). Of the arteritides, Takayasu's disease is particularly important and can give rise to visceral aneurysm, with or without associated occlusive arterial disease.

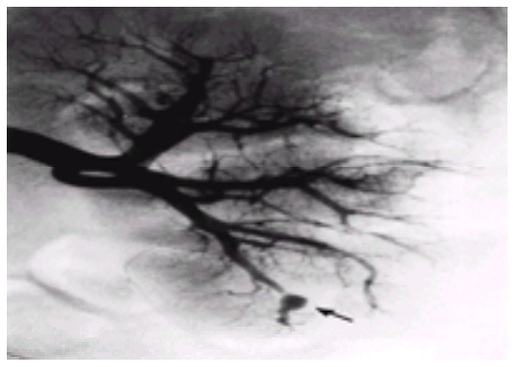


Fig. 1. Small renal arterial aneurysm following a percutaneous renal biopsy.

In the past mycotic, aneurysms were usually a consequence of bacterial endocarditis but more recently intravenous drug abuse and cardiac catheterization have contributed an increasing proportion of cases.

Relative incidence of individual aneurysms

Splenic aneurysms (Fig. 2) account for around two-thirds of all ruptured visceral arterial aneurysms, with most of the remainder being ruptured renal, hepatic (Fig. 3), gastric, superior mesenteric arterial aneurysms (Fig. 4), or on the coeliac artery (Fig. 5). Aneurysms of the inferior mesenteric artery are extremely rare, as are those of the gastroduodenal and pancreaticoduodenal arteries.

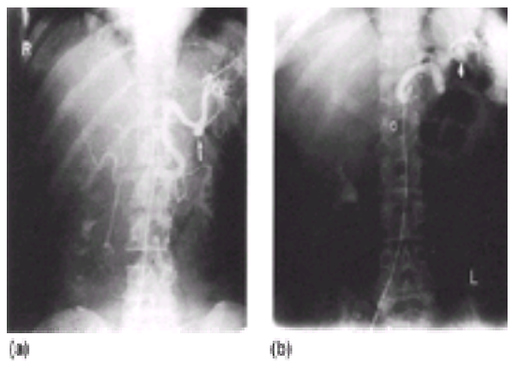


Fig. 2. Splenic arterial aneurysms.



Fig. 3. Hepatic arterial aneurysm.

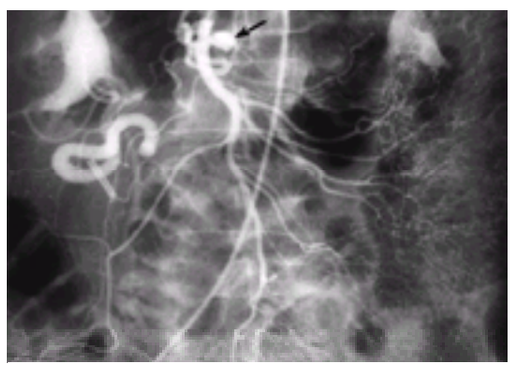


Fig. 4. Superior mesenteric arterial aneurysm.

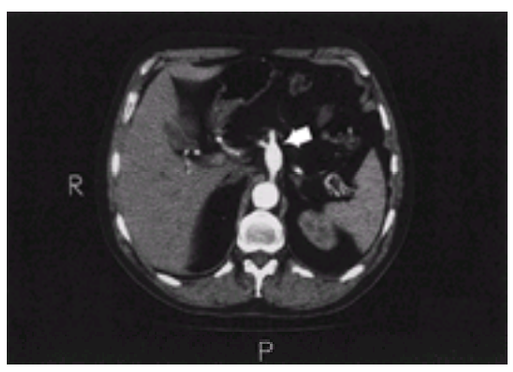


Fig. 5. Computed tomogram showing a coeliac arterial aneurysm.

Clinical presentation

Most visceral arterial aneurysms are asymptomatic or give rise to vague abdominal discomfort, perhaps related to compression or stretching of adjacent nerves by expansion of the aneurysm. Occasionally hepatic, gastroduodenal, or pancreaticoduodenal aneurysms may produce bile-duct compression and jaundice, while splenic aneurysms may be associated with pancreatitis. Mesenteric aneurysms may produce gut ischaemia by releasing emboli from mural thrombus within the aneurysm.

The most dramatic symptoms occur when an aneurysm ruptures. Initially the rupture may be partially contained either by the adjacent connective tissue or, in the case of splenic, gastric, and hepatic arterial aneurysms, within the lesser sac. In such circumstances the main symptoms are abdominal pain and a variable amount of circulatory collapse, depending on the volume of blood lost. Sooner or later the connective tissues that have partially contained the bleeding will give way, with free rupture into the peritoneal cavity and profound circulatory collapse.

Unusual presentations occur when a visceral aneurysm erodes into an adjacent structure, for example the hepatic artery into the common bile duct or a gastric, gastroduodenal, or pancreaticoduodenal artery into the foregut. The resultant haemobilia, haematemesis, and melaena are likely to be attributed preoperatively to other, more common causes of upper gastrointestinal haemorrhage and even at operation the correct diagnosis may not be obvious.

Renal arterial aneurysms occur in the extrarenal (Fig. 6) and intrarenal arteries. They are found in 1 per cent of patients undergoing renal angiography. Extrarenal arterial aneurysms are more likely to present with hypertension secondary to stenosis of the renal artery. For intrarenal arterial aneurysms the most common symptom is haematuria. It is likely that many renal arterial aneurysms discovered by angiography are incidental findings of no clinicopathological significance.

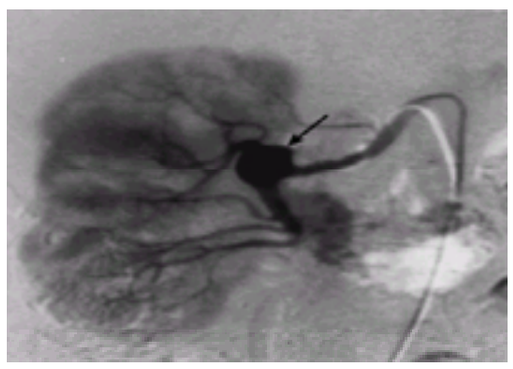


Fig. 6. Renal arterial aneurysm at the bifurcation of the renal artery, which shows evidence of fibromuscular dysplasia.

Diagnosis

The mainstay of precise anatomical diagnosis remains angiography with selective catheterization of the appropriate visceral artery. Angio-nbgrams will also display collateral circulation and allow planning of any operative procedure that may be necessary. Incidental diagnoses of visceral aneurysms will be made from abdominal radiographs if the aneurysm wall is calcified or from abdominal computed tomograms. Computed tomography may also give useful anatomical information or may be diagnostic in cases where a visceral aneurysm is suspected or presents as an abdominal mass.

The diagnosis of a ruptured visceral aneurysm is based on the clinical signs of circulatory collapse from blood loss together with evidence of intra-abdominal bleeding. The patient's condition will usually not permit time to be spent on confirming the diagnosis preoperatively, since emergency laparotomy to control the haemorrhage is

mandatory.

Management

Management decisions are clear-cut and uncontroversial when a visceral arterial aneurysm has ruptured: arresting the haemorrhage, usually by emergency surgery, is essential. There is considerable uncertainty about the best management of an intact aneurysm, particularly when it is asymptomatic. Not enough is known of the natural history of the disease for an informed recommendation to be made. When the aneurysm is very large or causing symptoms, active intervention is usually appropriate unless it is anatomically inaccessible or the patient is in precarious health. The aneurysm can sometimes be occluded by interventional radiological techniques inserting balloons or metal coils. Rarely, it may be possible to insert a covered stent transluminally to exclude the aneurysm and maintain patency of the artery. In most cases the surest, safest, and most cost-effective approach remains operative occlusion, resection, or arterial bypass. An aneurysm of any size discovered in early pregnancy or in a woman who plans to have children is also usually best treated by either transluminal occlusion or operation, because of the risk of rupture during pregnancy.

The greatest uncertainty is caused by a small, asymptomatic, visceral arterial aneurysm in an anatomical location where surgery is potentially difficult or dangerous. Many such aneurysms are best managed conservatively, with regular monitoring of their size and the patient's condition. Others will lend themselves to treatment by interventional radiological techniques, the scope of which continues to expand. Provided the aneurysmal artery can be cannulated, the aneurysm can usually be satisfactorily occluded by the insertion of metal coils or inflatable, detachable balloons.

The majority of renal arterial aneurysms discovered incidentally by renal angiography have a benign prognosis and treatment is usually unnecessary. The risk of rupture of a renal arterial aneurysm under 2 cm in diameter is negligible. Surgery is usually restricted to patients with expanding aneurysms, intractable hypertension, haematuria, renal infarction secondary to microembolism, and to fecund females.

Mycotic aneurysms of visceral arteries have the same sinister prognosis as they do elsewhere, with a tendency to rapid enlargement and rupture. For this reason, elective surgery is best undertaken as early as possible in conjunction with systemic antibiotic therapy to treat the underlying infection. Interventional radiology has no place in mycotic aneurysms since the insertion of coils and balloons into an infected artery guarantees that the infection will never be eradicated.

Further reading

King TA, McDaniel MD, Flinn WR, Yao JST, Bergan JJ. Visceral artery aneurysms. In: Moore WS, ed. *Vascular surgery; a comprehensive review*, pp. 351–65. Grune and Stratton, New York, 1983. [Overview of visceral aneurysms.]

Skudder PA. Visceral artery aneurysms. In: Persson AV, Skudder PA, ed. *Vascular surgery*, pp. 145–73. Marcel Dekker, New York, 1987. [Overview of visceral aneurysms.]

Tareen AH, Hensler M, Jensen F, Gronvall J, Schroeder TV. Solitary intra-renal aneurysm. *European Journal of Vascular and Endovascular Surgery* 1997; **14**; 221–3. [A typical case report and brief literature review of a type of visceral arterial aneurysm.]

17.7.1 Carotid artery

Peter J. Morris

Introduction
Pathophysiology of transient ischaemic attacks in the carotid distribution
Clinical presentation and differential diagnosis
Investigation
Management of transient ischaemic attacks in the carotid distribution or amaurosis fugax
Indications for carotid endarterectomy
Transient ischaemic attack
Retinal infarction
Completed stroke
Evolving stroke
Asymptomatic stenosis of the internal carotid artery
Cerebral hypoperfusion
Carotid endarterectomy
Timing of the operation
Anaesthesia
Monitoring of cerebral ischaemia
Operation
Postoperative monitoring
Complications of carotid endarterectomy
Restenosis
Prevention of stroke
Further reading

Introduction

Atheromatous disease at the origin of the internal carotid artery is a significant cause of strokes. Its recognition as a cause of neurological symptoms is important, for it may be amenable to surgical correction and hence protection from a subsequent stroke by carotid endarterectomy. The usual indication for carotid endarterectomy is a transient ischaemic attack or amaurosis fugax, although it may be performed in patients with an evolving stroke, a completed stroke, or an asymptomatic tight carotid artery stenosis. Transient stroke-like episodes, lasting minutes to hours, have been recognized for more than a century, but it was Chiari in 1905 who first described the association between thrombosis at the carotid bifurcation and distal embolism in the internal carotid artery, this association being further elaborated many years later by Gunning and colleagues in Oxford. Today a transient ischaemic attack is defined as an acute loss of focal cerebral or ocular function lasting less than 24 h and presumed to be due to thromboembolism.

The first report of a reconstruction of the carotid bifurcation in a woman with transient ischaemic attacks was by Eastcott and colleagues in 1954. Over the next 30 years carotid endarterectomy became widely practised for both symptomatic and asymptomatic stenoses of the internal carotid artery. However, there were marked discrepancies in the operative rates in different parts of the world, there being, for example, 20 times more operations per million of population performed in North America than Great Britain in the 1980s. This difference reflected the controversy concerning the value of the operation in stroke prevention, remembering that the operation itself is associated with a significant risk of death and stroke. However, much of this controversy has now been resolved, for the results of two large multicentre trials in Europe and North America showed a marked beneficial effect of operation in the prevention of strokes in patients presenting with transient ischaemic attacks or amaurosis fugax and a tight stenosis of the internal carotid artery.

Pathophysiology of transient ischaemic attacks in the carotid distribution

Atheromatous plaques form at the carotid bifurcation, and especially in the carotid sinus. The lesion in the internal carotid artery is restricted in most instances to the origin of the artery and the carotid sinus, the artery distal to the lesion and the common carotid artery proximal to the lesion being relatively normal. The predilection for this area is undoubtedly due to the turbulence in flow created by the bifurcation and the dilated carotid sinus. The plaque may ulcerate, giving rise to thrombus formation in the bed of the ulcer or, if the plaque eventually produces a tight stenosis of the artery (often precipitated by a haemorrhage into the plaque), thrombus may form at the stenosed area of the plaque (Fig. 1 and Fig. 2). If thrombus is dislodged it will pass downstream, either to the retinal arteries via the ophthalmic artery or to the cerebrum via the middle cerebral artery, giving rise either to amaurosis fugax or a retinal infarct, or to a transient ischaemic attack or a completed stroke. An embolus from the same site in the plaque will tend to lodge distally in the same place, for the bloodstream in the internal carotid artery is laminar rather than randomly distributed. Thus patients tend to suffer repeated attacks of amaurosis fugax or of transient ischaemic attacks but less commonly of both. Haemorrhage into an atheromatous plaque is also frequent and this can lead to sudden narrowing of the lumen as well as to ulceration. Ulceration may be associated with discharge of cholesterol crystals into the lumen, which embolize distally, as well as thrombus formation in the bed of the ulcer. As the stenosis becomes more severe there is an increased risk of thrombosis and occlusion of the internal carotid artery, with a subsequent stroke, although occlusion does not always result in a stroke. Alternatively, flow past the stenosis may become so poor as to produce transient ischaemic attacks on a purely haemodynamic basis. However, the bulk of transient ischaemic attacks in the carotid artery are embolic rather than haemodynamic. Such attacks in the carotid distribution may be due to emboli from proximal lesions located in sites other than the internal carotid artery, such as the heart or the ascending aorta. Three types of emboli can be detected in the retina (Fig. 3): white emboli of cardiac origin, causing segmental blockage and retinal infarction; small white fibrin–platelet plugs, which pass through the retinal circulation within minutes; and small refractile plaques (atheromatous debris).



Fig. 1. Carotid angiogram showing a large atheromatous plaque at the origin of the internal carotid artery with a tight stenosis. The inset shows the endarterectomized specimen with thrombus within the lumen.

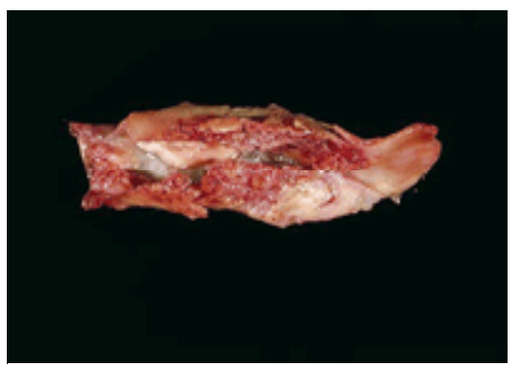


Fig. 2. An internal carotid endarterectomy specimen with thrombus within the lumen at the site of stenosis.



Fig. 3. Fundus of patient who had an episode of amaurosis fugax 24 h before showing a small refractile embolus peripherally (Hollenherst plaque).

Clinical presentation and differential diagnosis

Symptoms of transient ischaemia in the carotid artery territory reflect its distribution to the eye and the anterior two-thirds of the brain. The most common symptoms are weakness, numbness, and clumsiness of the limbs, especially the arm, contralateral to the side of the lesion, or loss of vision on the side of the lesion (amaurosis fugax). The patient characteristically describes the loss of vision as if a blind had been pulled down (occasionally across) the eye and as vision returns, usually in a few minutes, the blind retreats in the opposite direction. Dysphasia associated with a transient ischaemic attack is also common, especially if the left hemisphere is involved in a right-handed person. Amaurosis fugax usually lasts for only minutes while transient neurological defects last from minutes to hours.

Transient neurological deficits may be global rather than focal in nature and it is important to distinguish the two. Global symptoms include faintness, giddiness, vertigo, binocular visual loss, and syncope; these are due usually to a transient fall in blood pressure and hence cerebral hypoperfusion. This seldom results in a focal deficit except where there is a tight carotid stenosis, as discussed earlier. The symptoms of vertebrobasilar ischaemia must also be recognized, but these are usually easily distinguished from transient ischaemic attacks in the carotid artery distribution.

Focal neurological deficits due to causes other than thromboembolism can present in a manner resembling a transient ischaemic attack, but remembering that they exist will usually allow their exclusion (Table 1).

Migraine	Hypoglycaemia
Focal epilepsy	Increased blood viscosity
Intracranial neoplasm	Polycythaemia, thrombocythaemia
Subdural haematoma	Sickle-cell disease
Intracranial aneurysm	Multiple sclerosis
Arteritis	Hysteria

Table 1 Causes of transient focal neurological deficits not due to thromboembolism

Examination of the patient usually confirms that there is no residual neurological deficit. A bruit may be heard over the appropriate carotid bifurcation, but this is a relatively poor sign of an internal carotid artery stenosis, for an external carotid artery lesion may be the cause of a bruit. Furthermore the absence of a bruit does not preclude a diagnosis of internal carotid artery disease as the cause of a transient ischaemic attack, as a bruit may be absent in the presence of a very tight stenosis because the flow is so low. Examination of the heart for evidence of valve disease as a possible source of emboli is also important. In patients who have had an attack of amaurosis fugax not long before being seen, examination of the fundi may be rewarding in that emboli may still be identified even though vision has recovered. If vision remains defective, an embolus that is usually refractile can often be identified as the cause of a retinal infarct.

Investigation

There have been major changes in the investigation of transient ischaemic attacks in the carotid distribution or amaurosis fugax in recent years. The gold standard has always been carotid angiography, most recently digital subtraction arteriography. However, the risk of a stroke being precipitated by angiography is between 1 and 4 per cent when this is performed for extracerebrovascular disease, and this has provided enormous impetus in recent years for the development of non-invasive methods of determining the presence of significant carotid artery disease. The advent of the duplex scan, which combines real-time B-mode scanning with Doppler sound spectral analysis, allows not only the carotid bifurcation to be imaged but also the degree of stenosis to be calculated. The addition of colour has further refined the technique (Fig. 4). This technique is rapidly replacing angiography in the evaluation of the carotid bifurcation, and many vascular units, including our own in Oxford, now perform a carotid endarterectomy on the basis of duplex scanning alone. A CT scan or magnetic resonance image (**MRI**) of the brain is advisable not only to identify any area of infarction but also to exclude other pathology. Transcranial Doppler analysis of blood velocity in the intracerebral arteries can be used to assess collateral cerebral blood flow, and the recent development of probes for the duplex scanner now allows imaging of the intracerebral vessels. In selected patients, especially those with putative cerebral hypoperfusion, transcranial Doppler analysis can provide useful information. Magnetic resonance angiography may be useful if the duplex scan is technically unsatisfactory, particularly to confirm complete occlusion of the internal carotid artery, but it has not provided precise enough images of the carotid bifurcation to become a regular tool in investigation.

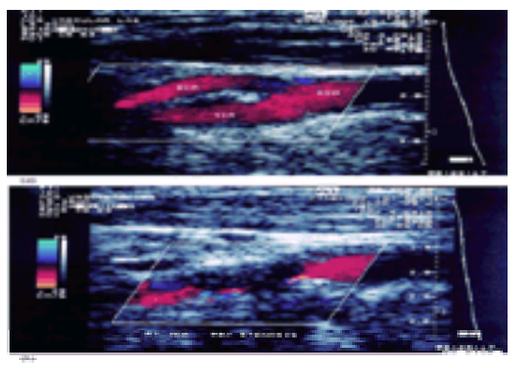


Fig. 4. Duplex scan of carotid bifurcation. (a) Normal bifurcation. (b) Severe stenosis of internal carotid artery (ICA).

If there is any question of a cardiac focus for emboli, an echocardiogram should be performed. Careful cardiac evaluation is always essential, however, for around 20 per cent of patients with transient ischaemic attacks of the carotid distribution die of myocardial infarction within 5 years of undergoing carotid endarterectomy. Other investigations include all those that would be performed in any patient with atheromatous arterial disease, such as a full blood count and blood lipid analysis.

Management of transient ischaemic attacks in the carotid distribution or amaurosis fugax

Once a diagnosis of a transient ischaemic attack in the carotid distribution has been made or even suspected then the patient should be given aspirin, 300 mg/day, and referred for evaluation by either a neurologist or vascular surgeon specializing in this area. Duplex scanning of the carotid arteries will usually enable a diagnosis of a carotid lesion to be established or excluded, and the severity of the stenosis and often the morphology of the plaque to be defined. A CT scan or MRI of the brain is advisable not only to exclude any other intracerebral lesion but also to show any existing areas of infarction.

As a result of the two large multicentre European and North American trials, management of symptomatic carotid artery lesions is now based on sound evidence. If a tight stenosis (greater than 70 or 80 per cent depending on method of measurement) is present, carotid endarterectomy is indicated, provided the patient's general health, in particular their cardiac status, is sound (Table 2). Age of itself is not a contraindication to carotid endarterectomy but the majority of patients will be between 65 and 75 years of age. If the stenosis is minimal (0 to 29 per cent), surgery is not indicated: these patients have a worse outcome in terms of stroke than patients treated medically because the perioperative mortality/stroke rate exceeds the very low risk of a subsequent stroke associated with minimal lesions. These patients should be maintained on aspirin therapy. The appropriate treatment for moderate stenoses (30 to 70 per cent) is dependent on the degree of stenosis and perioperative mortality and stroke rate as the benefits of surgery are very modest, and what benefits can be achieved are dependent on a very low mortality and stroke rate for the surgeon (see later). If transient ischaemic attacks continue to occur during treatment with aspirin, surgery or anticoagulation with warfarin or coumadin should be considered. Whatever the management, any hypertension should be well controlled and hyperlipidaemia treated by diet or, if severe, by cholesterol lowering agents. All patients are encouraged to stop smoking.

Degree of stenosis (according to NASCET)	Management
70–99%	Carotid endarterectomy + aspirin + best medical management
30–69%	Aspirin + best medical management + [carotid endarterectomy (50–69%)]
0–29%	Aspirin + best medical management

Table 2 The management of an internal carotid artery stenosis causing transient ischaemic attacks

Indications for carotid endarterectomy (Table 3)

Transient ischaemic attack

Transient ischaemic attack, including amaurosis fugax
Retinal infarction
Completed stroke
Evolving stroke
Asymptomatic internal carotid artery stenosis
Cerebral hypoperfusion

Table 3 Possible indications for carotid endarterectomy in the presence of a stenosis of the internal carotid artery

A transient ischaemic attack is defined as an acute loss of focal cerebral or ocular function with symptoms lasting less than 24 h and which, after adequate investigation, is presumed to be due to embolic or thrombotic vascular disease. This definition not only includes transient neurological defects but also transient episodes of blindness (amaurosis fugax). The 24-h time period is arbitrary, but has been accepted for many years and indeed the majority of attacks last from minutes to hours. Those due to an atheromatous plaque at the origin of the internal carotid artery (Fig. 1) are by far the most common indication for carotid endarterectomy because of the increased risk of a completed stroke in this situation.

Retinal infarction

If an embolus lodges long enough in the retinal arteries, infarction of part or all of the retina may occur. This is regarded in the same way as amaurosis fugax with respect to an indication for surgery.

Completed stroke

A patient whose neurological deficit lasts longer than 24 h but makes a full recovery, or who is left with a permanent but stable neurological deficit, and who has a stenosis of the internal carotid artery appropriate to the neurological signs is also a candidate for carotid endarterectomy, especially if a good recovery has been made from the stroke. Traditionally, following a completed stroke, surgery is postponed for 6 weeks, but the evidence to support this delayed approach is not really available.

Evolving stroke

This is a controversial indication. Early experience of carotid endarterectomy in this group of patients was accompanied by high perioperative mortality (40 to 60 per cent) and the procedure was abandoned in this situation. However, more recent cautious evaluation of the value of carotid endarterectomy in carefully selected patients with an evolving stroke or a fluctuating neurological deficit, where urgent investigations have confirmed the presence of an appropriate lesion, now suggests that an emergency carotid endarterectomy may have a place in management.

Asymptomatic stenosis of the internal carotid artery

This is another controversial area. It has been an indication for carotid endarterectomy in a significant proportion of patients in the United States, but hardly at all in Great Britain. Retrospective and prospective studies of patients with a known asymptomatic stenosis of the internal carotid artery suggest that the risk of stroke in these patients is relatively low, and indeed less than the risk of myocardial infarction. Furthermore, when strokes do occur they are usually preceded by transient ischaemic attacks. The associated risk of death and stroke from the operation itself may not therefore be outweighed by any benefit of subsequent stroke protection. The recent asymptomatic Carotid Surgery Trial in the United States was stopped because of an apparent benefit of surgery. However, this was probably premature in that the trial was associated with a very low operative mortality and stroke rate in the surgical group in comparison with a medically managed control group with a high stroke rate. A

large European multicentre trial is in progress comparing medical management plus carotid endarterectomy with medical management alone in asymptomatic patients with a tight stenosis of the internal carotid artery and which now has some 3000 patients entered. This trial will be of great importance in helping to define management of patients with a tight asymptomatic carotid artery stenosis. If cardiac surgery is required in a patient with a symptomatic stenosis, this is generally performed first, followed by the carotid surgery at a later date, rather than as a combined procedure, although the latter approach is not uncommon and the protagonists of this approach would say that this can be done without any increased mortality or morbidity.

Cerebral hypoperfusion

This is an extremely complicated clinical problem in that there is no established method of determining whether a tight stenosis of the internal carotid artery, especially if bilateral, is responsible for transient neurological deficits by causing hypoperfusion rather than emboli. Certainly in some instances it must be, and the development of transcranial Doppler analysis and duplex scanning of the circle of Willis may in time allow patients with cerebral hypoperfusion, who would benefit from carotid endarterectomy, to be better defined. It is not likely to be a large number, based on the experience of the intracerebral–extracerebral reconstruction trial of several years ago.

Carotid endarterectomy

Timing of the operation

Once a firm diagnosis is established an elective operation should be planned as soon as possible. The risks of a stroke before surgery after the patient has started taking aspirin are relatively small and the risks of perioperative complications are reduced by taking the time for a careful evaluation of the patient and in particular his or her cardiac status.

Anaesthesia

Most surgeons prefer a general anaesthetic. It is essential that blood pressure control is well established before the patient comes to surgery and it needs to be carefully monitored via an intra-arterial line throughout the procedure. However, in recent years there has been an increasing use of regional anaesthetic blocks (superficial and deep cervical blocks) for carotid endarterectomy as this allows the neurological status to be followed in an awake patient, which is of particular value when the carotid artery is clamped. Thus, if the patient develops a neurological deficit a shunt can then be inserted. Whether this results in a lower risk of perioperative stroke, as claimed, is unproved, but randomized trials are at the planning stage.

Monitoring of cerebral ischaemia

Where the operation is performed under regional anaesthesia it is possible for the patient to report any neurological deficit, provided they have not required so much sedation as to lose their awareness of events. In patients under general anaesthetic, electroencephalography is valuable in units where the appropriate facilities exist (simple three-channel electroencephalography monitoring is of no value). More recently, monitoring of middle cerebral artery velocity on the side of the operation with transcranial Doppler is being widely adopted (Fig. 5). Whether this is an adequate measure of the maintenance of collateral flow is still not established. It may also allow the passage of microemboli, especially during the dissection of the carotid bifurcation, to be detected and thus influence surgical technique.



Fig. 5. Carotid endarterectomy in progress with continuous monitoring of middle cerebral artery velocity on the side of operation with transcranial Doppler.

Operation

The carotid bifurcation is exposed through an oblique incision along the anterior border of the sternomastoid (Fig. 6). Once the carotid bifurcation has been exposed, arterial pressures are measured in the common carotid artery proximal to the lesion and in the internal carotid artery distal to the lesion to establish the presence of a pressure gradient, and then the common carotid artery is clamped and the internal carotid pressure recorded. This is the stump pressure and is a measure of the collateral flow from the other side and the vertebrobasilar system via the circle of Willis. If the pressure is below 50 to 60 mmHg, surgeons who shunt selectively would consider that an indication for a shunt. Blood flow in the internal carotid artery may also be measured, using either an electromagnetic flow meter or more recently an ultrasound probe (OpDop). This provides a baseline for comparison with a similar measurement made after reconstruction.

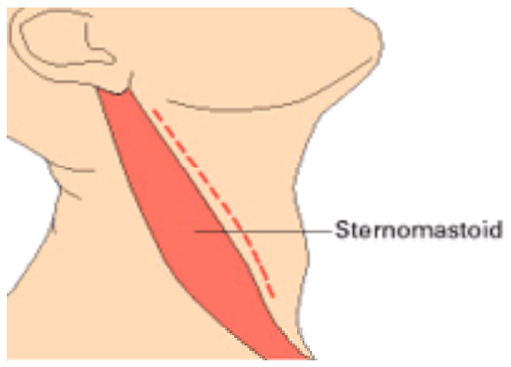


Fig. 6. Skin incision for carotid endarterectomy is made along the line of the anterior border of the sternomastoid.

The patient is then heparinized and clamps are applied to the common, internal, and external carotid arteries. An arteriotomy starting in the common carotid artery below the lesion is carried through the lesion into the internal carotid artery, distal to the disease (Fig. 7). If a shunt is to be used, it is inserted at this time. Some surgeons use shunts routinely in all cases, while others use them selectively. The two most widely used shunts are the Javid shunt and the Pruitt shunt (Fig. 8). The latter is more easily inserted and can be moved more easily out of the way during the endarterectomy (Fig. 9). It is our practice to shunt routinely using a Pruitt shunt if the operation is being performed under a general anaesthetic.

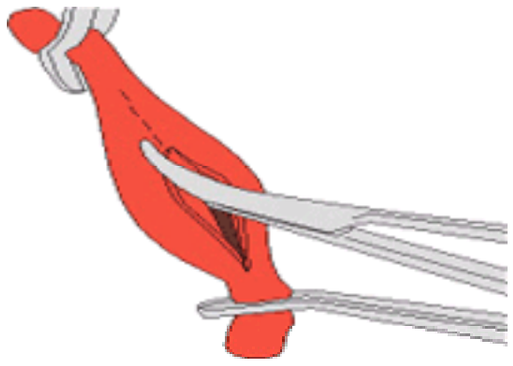


Fig. 7. Arteriotomy commencing in the common artery below the lesion in the internal carotid and carried up the artery through the plaque into the area distal to the lesion.

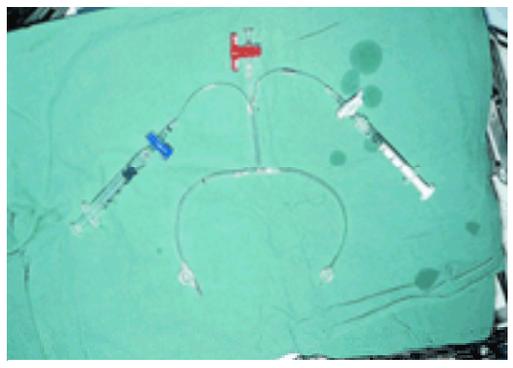


Fig. 8. Pruitt shunt with the balloons inflated.

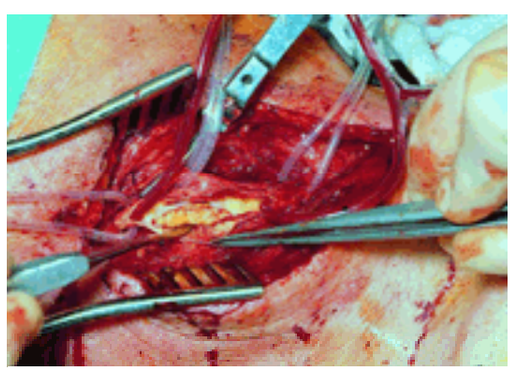


Fig. 9. Pruitt shunt in place while endarterectomy is proceeding.

The endarterectomy requires identification of a plane between the plaque and the media: generally this lies somewhere between the outer one-third and inner two-thirds of the media. If the correct plane is identified, the lesion comes away without difficulty. It is essential to obtain a tightly adherent distal flap in the internal carotid artery and if this is not the case, it should be tacked down with some fine 7/0 Prolene sutures (Fig. 10).

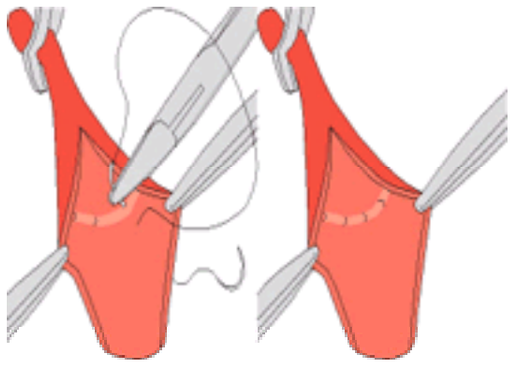


Fig. 10. Insertion of tacking sutures (7/0 Prolene) at distal flap to ensure that it is firmly adherent.

Once the endarterectomy has been completed, the arteriotomy can be closed directly with a continuous Prolene suture (Fig. 11) or with a vein (internal jugular vein or saphenous vein), polytetrafluoroethylene (Gore-Tex or Impra), or Dacron patch (Fig. 12). In recent years many vascular surgeons have routinely closed most arteriotomies with a patch as the recognition of a relatively high incidence of recurrent stenoses, usually at the distal end of the endarterectomy, has become apparent. Occasionally, if the plaque is localized to the origin of the internal carotid artery and the external carotid is free of disease, the bifurcation can be reconstructed by suturing the posterior walls of these two arteries together and then suturing the anterior walls together; in effect this reconstructs a higher bifurcation, but widens the artery at the end of the endarterectomy (Fig. 13).

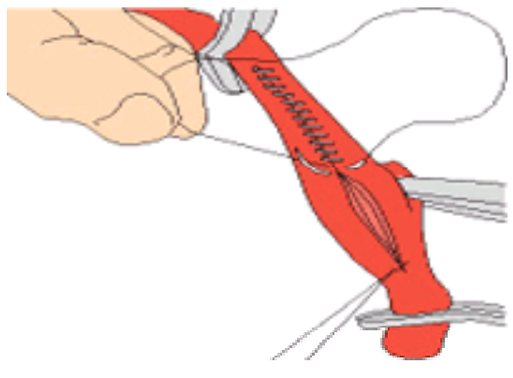


Fig. 11. Closure of arteriotomy with continuous 6/0 Prolene.

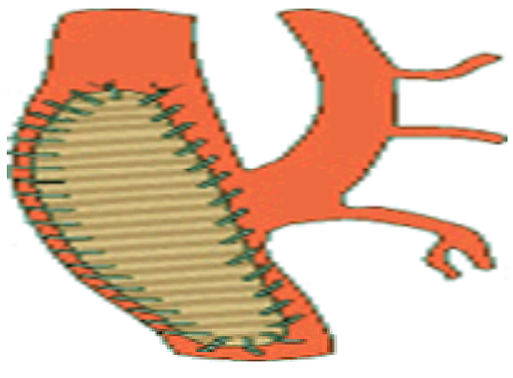


Fig. 12. Closure of arteriotomy with a knitted Dacron patch and 6/0 Prolene.

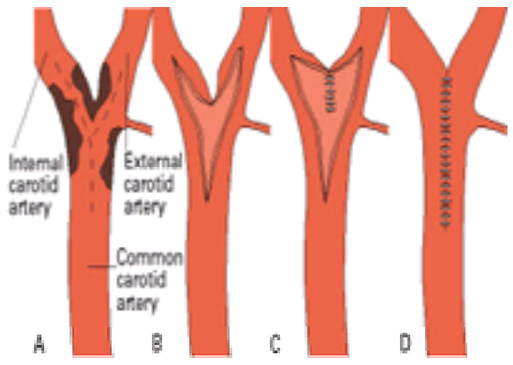


Fig. 13. Closure of arteriotomy using the external carotid artery such that the bifurcation is reconstructed at a higher level providing a wider artery at the site of the endarterectomy.

Postoperative monitoring

Following surgery the patient is kept in the recovery room until awake and hen returned to the ward. As the vast majority of perioperative strokes occur in the first 24 h after operation, neurological observations must be made every 30 min over the first 12 h and then every hour until 24 h. It is also important to maintain a stable blood pressure avoiding any periods of hypo- or hypertension. This should not generally be a problem if the patient has come to operation with well-controlled blood pressure.

If any neurological deficit appears relevant to the side of operation, a rapid return to the operating room is indicated, with exploration of the neck and exposure of the carotid vessels. The endarterectomized site should be re-explored with removal of any thrombus; a careful inspection of the distal flap is mandatory. If thrombosis and occlusion of the internal carotid artery has occurred, restoration of the cerebral flow may increase the likelihood of recovery or limit the extent of the neurological deficit. Although a CT scan of the brain and duplex scan of the operated side would be desirable, it is not practical to do this because of the loss in time in restoring flow in an occluded vessel that would result. Thrombus is found in less than 50 per cent of patients re-explored, and whether removing this limits the evolving deficit is unknown. As any one surgeon's experience in this area is small it will probably never be known, but it would instinctively seem an appropriate course to follow.

Complications of carotid endarterectomy (Table 4)

Death
Stroke
Myocardial infarction
Stroke
Thrombosis
Occlusion
Embolization
Cerebral haemorrhage
Haematoma
Nerve pareses
Hypoglossal
Glossopharyngeal
Facial 'marginal branch'
Recurrent laryngeal
Superior laryngeal
Accessory
Wound infection

Table 4 Complications of carotid endarterectomy

The major complications of this procedure are death, either from stroke or a myocardial infarction, or a stroke. A stroke occurring during the operation or generally within the first 24 h of surgery may be transient or more severe, leading to a permanent disability. The combined perioperative mortality and stroke rate as defined in the European and North American trials (death or neurological deficit resulting in a permanent disability and occurring during or within 30 days of surgery) should be less than 5 per cent. A neurological deficit may be due to thrombosis at the site of endarterectomy, with either occlusion or embolization, or an intracerebral haemorrhage.

As most patients are receiving aspirin, significant haematomas are relatively common, but rarely need draining. Local nerve pareses, especially of the hypoglossal nerve, are also not uncommon but are usually transient, recovering within weeks to months. Retraction is the usual cause of these local nerve pareses, which is often unavoidable, especially in the case of the hypoglossal nerve and a high carotid bifurcation. Retraction of the vagus may result in paresis of the superior or recurrent laryngeal nerve. In general the patient can be reassured that recovery will take place within a month or two.

Restenosis

The availability of duplex scanning after carotid endarterectomy has shown that the incidence of a significant recurrent stenosis (greater than 50 per cent), commonly at the distal end of the endarterectomy, is much greater than hitherto believed, perhaps of the order of 15 to 30 per cent at 5 years. However, the development of symptoms related to the recurrent stenosis is very uncommon. Restenosis may be due to myointimal hypertrophy or atheroma. In the absence of symptoms no treatment other than continued daily aspirin is necessary. The development of transient ischaemic attacks is an indication for reoperation. The incidence of restenosis appears less with patch closures than with a primary closure, but there is no evidence that the lower restenosis rate is associated with a lower stroke rate with the passage of time. However, there is evidence that the use of a patch does reduce the incidence of perioperative strokes compared with primary closure.

Prevention of stroke

As carotid endarterectomy for stenosis of the internal carotid artery is an operation carried out in the expectation that it will prevent a subsequent stroke, it is perhaps surprising that it has taken nearly 30 years to establish that, at least in the case of a tight symptomatic stenosis (greater than 70 to 80 per cent) as determined by angiography, it does indeed do so. The European Carotid Surgery Trial (**ECST**) and the North American Symptomatic Carotid Endarterectomy Trial (**NASCET**) produced very similar results in that both trials showed a significant benefit of surgery in preventing a subsequent stroke. The trials are not strictly comparable in that different methods were used to calculate the angiographic degree of stenosis so that the method used in the ECST underestimates the severity of the stenosis as

measured in the NASCET (for instance a 70 per cent stenosis in the ECST is the equivalent of an 80 per cent stenosis in the NASCET). Both the interim and final analyses of these two classic studies have now been published.

In the ECST, 3024 patients who had had a carotid-territory non-disabling ischaemic stroke, transient ischaemic attacks, amaurosis fugax or retinal infarct together with a stenotic lesion in the relevant internal carotid artery were randomized either to best medical management or best medical management plus carotid endarterectomy. All patients received antiplatelet therapy. For the 778 patients with a severe stenosis (70 to 99 per cent) the total risk of surgical death, surgical stroke, ipsilateral ischaemic stroke, or any other stroke was 12.3 per cent, compared with 21.9 per cent in the non-surgical group ($2p < 0.01$) (Fig. 14). In contrast, in the minimal stenosis group (0 to 29 per cent) there was little risk of ipsilateral ischaemic stroke in the non-surgical group so that any possible benefits of surgery were outweighed by the perioperative risks of surgery. In patients with a moderate stenosis (30 to 70 per cent) there was no benefit of surgery and indeed in the final analysis the benefit of surgery was confined to stenoses of 80 per cent and above, the cut-off point being somewhere between 70 and 79 per cent. The overall risk from surgery of a non-fatal major stroke or death during the first 30 postoperative days was 7 per cent.

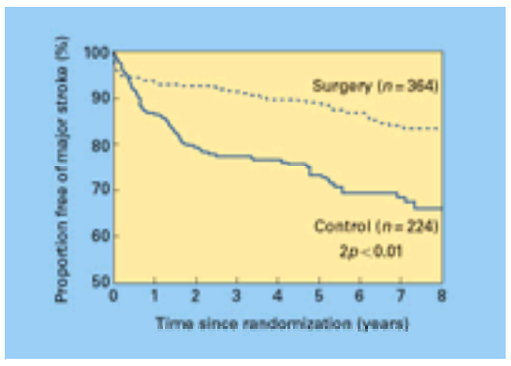


Fig. 14. ECST: Kaplan–Meier survival curves to show survival free of major stroke (with non-stroke deaths occuring more than 30 days after surgery censored) in patients in the surgery and control groups with 80 to 99 per cent stenosis of symptomatic carotid artery.

In the NASCET, 659 patients with a severe stenosis (70 to 99 per cent) who had had a hemispheric or retinal transient ischaemic attack or a non-disabling stroke were randomized to best medical management or best medical management plus carotid endarterectomy. All patients received antiplatelet therapy. The cumulative risk of stroke at 2 years was 26 per cent in the non-surgical group but only 9 per cent in the surgical group ($p < 0.001$) (Fig. 15). Follow-up to 8 years did not show any further benefit. A total of 2267 patients with stenosis of less than 70 per cent were randomly assigned to medical management or surgery. In the moderate stenosis group (30 to 69 per cent) there was a small benefit in patients with 50 to 69 per cent stenosis (the incidence of stroke was reduced from 22.2 to 15.7 per cent). No benefit was demonstrated for stenoses less than 50 per cent. The overall stroke and mortality rate in the first 30 postoperative days was 6.7 per cent.

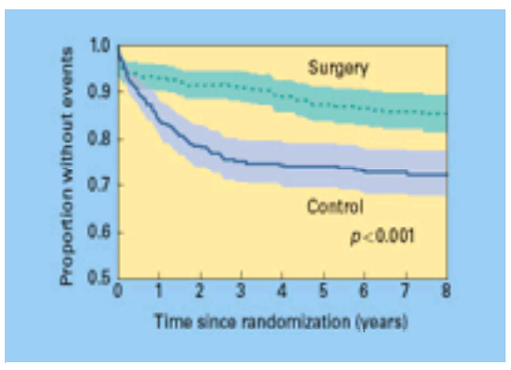


Fig. 15. NASCET: Kaplan–Meier curves for survival free of any ipsilateral stroke, 70 to 99 per cent stenosis with 95 per cent confidence intervals for both curves.

Thus these two trials convincingly demonstrate the beneficial effect of carotid endarterectomy in a patient with a tight symptomatic internal carotid artery stenosis (greater than 80 per cent in the ECST and greater than 70 per cent in the NASCET). Neither study shows a benefit for stenoses of less than 50 per cen, but the NASCET shows a modest benefit of surgery for stenoses between 50 and 69 per cent. In both trials the benefit of surgery was achieved in the first 3 years, the time of greatest risk of an ipsilateral stroke in those patients treated medically; thereafter the risk remained similar in the surgically treated and the medically treated patients. In the NASCET the perioperative risk of stroke or death was increased in the presence of contralateral occlusion, low-dose aspirin (not confirmed in a subsequent randomized trial of high-dose compared with low-dose aspirin which showed a better outcome with low-dose aspirin), no previous cardiac event, diabetes, and diastolic hypertension. In a systematic review of the published literature including an analysis of the ECST data, the ECST group showed that there was a significantly increased operative risk of stroke and death in patients presenting with cerebral symptoms compared with ocular ischaemia, females compared with males, systolic hypertension, peripheral vascular disease, contralateral occlusion of the internal carotid artery, and stenosis of the intracranial ipsilateral carotid artery and the external carotid artery. Although there are differences in the results of these analyses it should be possible in due course to identify those patients at greatest risk from surgery in trying to assess the risk–benefit odds in individual patients.

In patients with a tight asymptomatic stenosis of the internal carotid artery the place of carotid endarterectomy remains uncertain. The Asymptomatic Carotid Atherosclerosis Study (**ACAS**) in the United States was stopped prematurely because of a demonstrable benefit of surgery. A subsequent meta-analysis of carotid endarterectomy for asymptomatic stenoses did conclude that there was a small benefit of surgery in reducing the incidence of an ipsilateral stroke primarily because the risk of stroke in these patients medically treated is relatively low in any case. At this time because the benefit of surgery is so small it seems that medical management is the preferred option for management of most patients with a tight asymptomatic stenosis of the internal carotid artery until a subgroup of patients at greatest risk of a stroke can be identified. It is hoped that the outcome of the large European asymptomatic trial will provide further guidance in time as the target is to enter 4000 patients and entry is nearing completion.

Another controversial area is the management of patients with a symptomatic tight stenosis of the internal carotid artery and symptomatic coronary artery disease. Should the carotid endarterectomy be done first with a greater risk of a myocardial infarction postoperatively, the coronary arteries be revascularized first with a risk of stroke perioperatively, or should both be done together as a combined operation? The balance of available retrospective studies suggests that in the absence of evidence from a controlled trial the combined approach is a reasonable way to deal with this problem.

Finally, the advent of angioplasty and stenting has led to its application to the treatment of symptomatic internal carotid artery stenoses in place of carotid endarterectomy. Early data suggest that the risk of stroke is greater after angioplasty and stenting than after carotid endarterectomy and indeed one randomized trial was stopped because of the high stroke rate in the angioplasty group. Certainly this procedure should only be carried out in the context of a randomized controlled trial until its role in the management of carotid artery disease, if any, can be established.

Further reading

AbuRahma AF *et al.* Prospective randomized trial of carotid endarterectomy with primary closure and patch angioplasty with saphenous vein, jugular vein, and polytetrafluoroethylene: perioperative (30-day) results. *Journal of Vascular Surgery* 1996; **24**: 998–1007. [Evidence that patching reduces the incidence of perioperative strokes.]

Barnett HJM, Haines SJ. Carotid endarterectomy for asymptomatic carotid stenosis. *New England Journal of Medicine* 1993; **328**: 276–8.

Barnett HJM *et al.* Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. *New England Journal of Medicine* 1998; **339**: 1415–25. [The final results of the NASCET trial, now showing a modest benefit of carotid endarterectomy in patients with moderate stenoses (50–69 per cent).]

Benavente O, Moher D. Carotid endarterectomy for asymptomatic carotid stenosis: a meta-analysis. *British Medical Journal* 1998; **317**: 1477–80.

Chassin MR. Appropriate use of carotid endarterectomy. *New England Journal of Medicine* 1998; **339**: 1468–71. [A balanced comment based on the recent results of the NASCET, ECST, and ACAS

trials.]

Chiari H. Uber des Verhalten des Teilungswinkels der Carotis communis bei der endarteritis chronica deformans. *Verhandlungun der Deutschen Gesellschaft für Pathologie* 1905; **9:** 326–30.

Eastcott HHG, Pickering GW, Robb CG. Reconstruction of internal carotid artery in a patient with intermittent attacks of hemiplegia. *Lancet* 1954; **ii:** 994–6. [The first description of a successful operation for transient ischaemic attacks.]

European Carotid Surgery Trialists' Collaborative Group (1991). MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70–99 per cent) or with mild (0–29 per cent) carotid stenosis. *Lancet* 1991; **337:** 1235–43. [A classic trial demonstrating the efficacy of surgery for tight stenoses.]

European Carotid Surgery Trialists' Collaborative Group. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet* 1998; **351**: 1379–87.

Executive Committee for Asymptomatic Carotid Atherosclerosis Study. *Journal of the American Medical Association* 1995; **273**: 1421–8. [This randomised prospective study showed a benefit of carotid endarterectomy in asymptomatic carotid artery stenoses greater than 60 per cent provided the perioperative stroke/mortality rate was less than 3 per cent.]

Gunning AJ, Pickering GW, Robb-Smith AH, Ross Russell R. Mural thrombosis of the internal carotid artery and subsequent embolism. *Quarterly Journal of Medicine* 1964; **33:** 155–95. [One of the first descriptions of embolization of thrombus from the internal carotid artery.]

Hankey GJ, Warlow CP, Sellar RJ. Cerebral angiographic risk in mild cerebrovascular disease. *Stroke* 1990; **21:** 209–22. [Carotid angiograms have a significant risk of stroke.]

Heyman A *et al.* Risk of stroke in asymptomatic persons with cervical arterial bruits. *New England Journal of Medicine* 1980; **302:** 838–42.

Jordan WD Jr, Voellinger DC, Fisher WS, Redden D, McDowell HA. A comparison of carotid angioplasty with stenting versus endarterectomy with regional anesthesia. *Journal of Vascular Surgery* 1998; **28:** 397–403. [Angioplasty and stenting results in a higher perioperative stroke rate than does carotid endarterectomy.]

Lattimer CR, Burnand KG. Recurrent carotid stenosis after carotid endarterectomy. *British Journal of Surgery* 1997; **84**: 1206–19.

Meissmer I, Wiebers DO, Whisnant JP, O'Fallon WM. The natural history of asymptomatic carotid artery occlusive lesions. *Journal of the American Medical Association* 1987; **258:** 2704–7.

Murie JA, Morris PJ. Carotid endarterectomy in Great Britain and Ireland. *British Journal of Surgery* 1986; **76:** 867–70. [An interesting comparison in the rates of surgery for carotid artery disease in the USA and Great Britain and Ireland before the results of the randomised trials were published.]

Naylor AR *et al.* Randomized study of carotid angioplasty and stenting versus carotid endarterectomy: a stopped trial. *Journal of Vascular Surgery* 1998; **28**: 326–34.

North American Symptomatic Carotid Endarterectomy Trial Collaborators. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *New England Journal of Medicine* 1991; **325:** 445–53. [The other classic trial showing the benefit of surgery for a tight stenosis.]

Renton S, Hornick P, Taylor KM, Grace PA. Rational approach to combined carotid and ischaemic heart disease. *British Journal of Surgery* 1997; **84**: 1503–10. [In the absence of a randomized trial, this is a reasoned survey of the approaches to combined disease.]

Rothwell PM, Gibson RJ, Slattery J, Sellar RJ, Warlow CP, for the European Carotid Surgery Trialists' Collaborative Group. Equivalence of measurements of carotid stenosis: a comparison of three methods on 1001 angiograms. *Stroke* 1994; **25**: 2435–39.

Rothwell PM, Slattery J, Warlow CP. Clinical and angiographic predictors of stroke and death from carotid endarterectomy: systematic review. *British Medical Journal* 1997; **315**: 1571–7. [A preliminary attempt to identify patients at greatest risk.]

Tangkanakul C, Counsell CE, Warlow CP. Local versus general anaesthesia in carotid endarterectomy: a systematic review of the evidence. *European Journal of Vascular and Endovascular Surgery* 1997; **13**: 491–9.

Taylor DW *et al.* Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. *Lancet* 1999; **353**: 2179–84. [Patients receiving low-dose aspirin (81 or 325 mg/day) at the time of surgery had a lower perioperative stroke/mortality rate than patients receiving high-dose aspirin (650 or 1300 mg/day).]

Tu JV *et al.* The fall and rise of carotid endarterectomy in the United States and Canada. *New England Journal of Medicine* 1998; **339**: 1441–7.

Warlow C, Morris PJ. *Transient ischaemic attacks*. Marcel Dekker, New York, 1982. [A comprehensive monograph on the subject but particularly describes clinical presentation.]

Winslow CM, Solomon DH, Chassin MR, Kosecoff J, Merrick NJ, Brook RH. The appropriateness of carotid endarterectomy. *New England Journal of Medicine* 1988; **318:** 721–7. [This important paper at the time suggested that around a third of carotid endarterectomies performed in the United States were for inappropriate indications.]

17.7.2 Vertebrobasilar, subclavian, and innominate arteries

Peter J. Morris

Introduction
Pathophysiology
Clinical presentation and differential diagnosis
Investigation
Indications for surgery
Operations
Vertebral artery
Subclavian steal
Innominate steal
Results
Further reading

Introduction

The vertebrobasilar arterial system supplies the brainstem, occipital lobes, and medial aspects of the temporal lobes. Ischaemia causing loss of function in one or more of these segments can produce a complex of symptoms (Table 1) which may be due to causes other than thromboembolism, especially when they occur in isolation. Because of the uncertainty of the diagnosis, surgery has played a far less prominent role in the management of vertebrobasilar transient ischaemic attacks than is the case for those in the carotid distribution. Nevertheless, epidemiological data from the Mayo Clinic in the 1970s suggested that the risk of stroke after a vertebrobasilar transient ischaemic attack was similar to that of patients with a transient ischaemic attack in the carotid distribution. Thus identification of patients who might benefit from surgery is of considerable importance, although in practice this has proved to be extremely difficult.

Vertigo
Ataxia
Hemiparesis
Drop attacks
Sensory loss (unilateral or bilateral)
Diplopia
Binocular visual loss
Dysphagia
Dysarthria
Facial tingling or numbness (unilateral or bilateral)

Table 1 Common symptoms of vertebrobasilar transient ischaemic attacks

Pathophysiology

Obstruction of the origin of a single vertebral artery, usually by atheroma, should not result in distal ischaemia of the vertebrobasilar circulation if the contralateral vertebral artery is normal. In general, stenosis of the origins of both vertebral arteries is necessary to produce vertebrobasilar ischaemia; even then ischaemia may not occur if the internal carotid artery on each side is normal, and provided that there is a normal circle of Willis with intact posterior communicating arteries. However, there is often associated carotid bifurcation disease and intracerebral disease which in association with disease of the origins of the vertebral arteries can result in poor perfusion of the hindbrain. It should also be remembered that an intact so-called normal circle of Willis is found in only about 50 per cent of individuals.

In contrast to transient ischaemic attacks associated with stenoses of the internal carotid artery, those due to vertebrobasilar ischaemia are most often haemodynamic in origin. Whether it is also possible to produce temporary obstruction or kinking of the vertebral arteries in patients with cervical spondylosis by rotation of the neck is uncertain; although obstruction of a vertebral artery by osteophytes has been demonstrated, this should not result in symptoms unless there is associated disease in the contralateral vertebral artery.

A rather uncommon, but better defined, cause of vertebrobasilar ischaemia is the syndrome known as subclavian steal. In this condition there is a significant stenosis or even complete obstruction of the origin of the left subclavian artery, such that when the patient uses the left arm, blood passing up the right vertebral artery passes into the left vertebral artery to feed the subclavian artery distal to the obstruction (Fig. 1).



Fig. 1. Subclavian steal with a stenosis at the origin of the left subclavian artery and stealing of blood from the contralateral vertebral artery down the ipsilateral vertebral artery into the subclavian artery.

In addition, stenosis or occlusion of the origin of the innominate artery may be associated with steal down the right vertebral artery and internal carotid artery (innominate steal), causing symptoms associated with either vertebrobasilar ischaemia or carotid distribution ischaemia, or a combination of both.

Clinical presentation and differential diagnosis

Vertebrobasilar transient ischaemic attacks are difficult to diagnose unless several symptoms occur together during an attack or in separate attacks. Vertigo is the most common symptom of vertebrobasilar ischaemia, but is more often due to other causes, such as disorders of the vestibule (it is important to establish that the patient is having true vertigo). Transient episodes of ataxia are also not uncommon. Other symptoms include diplopia, dysphagia, dysarthria, and drop attacks. The drop attacks are a striking phenomenon when the patient recalls just dropping to the ground without any prewarning symptoms, perhaps, but usually not, losing consciousness transiently, and then recovering immediately. Tingling and numbness of the face and mouth or, indeed, half the body may occur, as also may transient hemiparesis.

Visual loss is the second most frequent symptom after vertigo and is quite variable, ranging from reduced vision in one half field, perhaps accompanied by positive scotomas, to impairment of vision on both sides. Bilateral impairment of vision ranges from total blindness to a generalized mistiness of vision; positive or negative scotomas may occur as spots or moving lights which may be coloured.

The innominate steal or subclavian steal syndromes are classically produced by exercise of the arm on the appropriate side. Apart from bruits in the root of the neck these syndromes will be associated with a distinct pressure gradient between the arms on each side (at least 20 mmHg). However, it must be stressed that the presentation of most patients with innominate or subclavian steal is not classic and the association of symptoms with a radiological finding is often difficult. A differential diagnosis of vertebrobasilar ischaemia includes the same causes as outlined in the differential diagnosis of carotid artery transient ischaemic attacks (Chapter 17.7.1).

Investigation

The investigation of patients with putative vertebrobasilar ischaemia is still based largely on angiography. An aortic arch study with selective viewing of both vertebral and both carotid arteries together with intracerebral views is required. The demonstration of stenoses at the origin of the vertebral artery does not necessarily confirm a diagnosis of vertebrobasilar ischaemia, and there is a lack of adequate functional tests of vertebrobasilar ischaemia. Duplex scanning of the vertebral arteries allows examination of the arteries in different positions of the neck. This, along with transcranial Doppler and duplex scanning of the circle of Willis, is likely to allow more precise definition of the relevance of extracerebral vascular disease, in particular vertebral disease, to symptoms compatible with a diagnosis of vertebrobasilar transient ischaemia.

Angiography remains an essential investigation for vertebrobasilar ischaemia, whatever the proposed cause might be. It demonstrates stenoses or occlusions of the origin of the left subclavian artery and the innominate artery, and also shows retrograde flow down the left vertebral artery in a subclavian steal syndrome (Fig. 2), and down the right carotid and vertebral arteries in an innominate steal syndrome. Reverse flow can also be detected by duplex scanning the appropriate arteries in comparison with the contralateral side. CT scanning or MRI of the brain is essential both to exclude an intracerebral lesion as a cause of the neurological symptoms and to define any infarcts in the posterior hemispheres or cerebellum, MRI being the preferable investigation for lesions in the hindbrain.

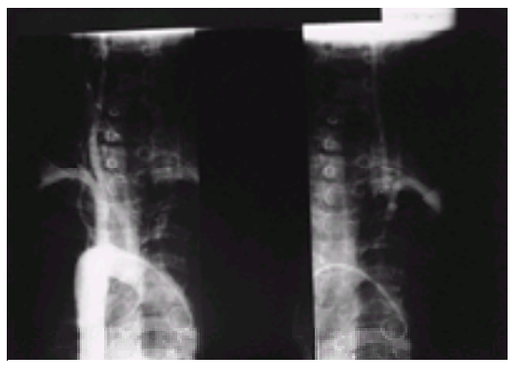


Fig. 2. An angiogram in a patient with a subclavian steal showing obstruction of the origin of the left subclavian artery (a) and later films showing retrograde flow down the left vertebral artery into the subclavian artery distal to the obstruction (b).

Indications for surgery

In the presence of symptoms compatible with vertebrobasilar ischaemia, appropriate lesions demonstrable on angiography, such as bilateral vertebral artery stenoses, left subclavian artery stenosis with reverse flow down the left vertebral artery, or innominate artery stenosis with reverse flow down the common carotid and vertebral arteries on the right side, can be considered an indication for surgery. However, patients in whom a diagnosis of vertebrobasilar ischaemia can be attributed confidently to a surgically correctable lesion are relatively few. The steal syndromes are the most clear-cut diagnoses, but even here the diagnosis often remains speculative and is only confirmed by a satisfactory outcome following surgical correction of the defect.

Operations

Vertebral artery

A stenosis of the origin of the vertebral artery may be approached directly and an endarterectomy performed at the origin or preferably through the subclavian artery (Fig. 3). Alternatively the vertebral artery can be divided distal to the lesion and reimplanted into the common carotid artery (Fig. 4). These procedures are performed through a transverse incision in the root of the neck, after division of the sternomastoid.

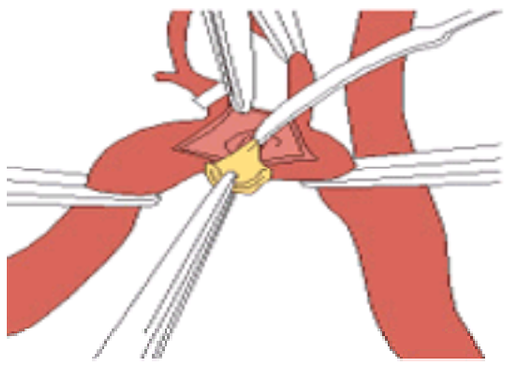


Fig. 3. The subclavian artery has been opened opposite the origin of the vertebral artery and a plug of atheroma extracted from the vertebral orifice.

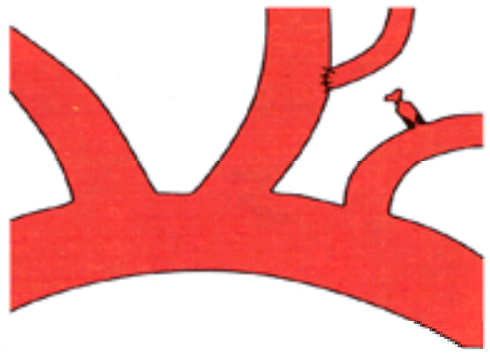


Fig. 4. The vertebral artery has been divided distal to the stenosis at its origin and reimplanted into the common carotid artery.

Stenoses in the more distal vertebral artery can be bypassed in a variety of ways if thought to be the cause of ischaemia. The vertebral artery can be divided and then revascularized with a graft from the common carotid artery, or the external carotid artery can be divided and anastomosed to the distal vertebral artery, or the distal vertebral artery can be anastomosed to the internal carotid artery.

Subclavian steal

Although the original approach to a stenosis of the origin of the subclavian artery was via a left anterolateral thoracotomy with endarterectomy of the artery and closure with a patch, this is rarely performed today. The current operations of choice do not involve opening the chest. A prosthetic or vein graft is inserted end-to-side between the common carotid artery and the subclavian artery distal to the origin of the vertebral artery (Fig. 5). Another approach which avoids clamping the common carotid artery involves running a vein or prosthetic graft from the axillary artery on the contralateral side subcutaneously just below the sternal notch to the axillary artery on the affected side (Fig. 6). Finally the subclavian artery may be divided proximal to the origin of the vertebral artery and anastomosed end-to-side to the common carotid artery (Fig. 7). In recent years most subclavian artery stenoses have been managed by angioplasty. Although initially there was concern that emboli might be dislodged at the time of angioplasty and pass up the vertebral artery, this concern has not been borne out in practice; thus angioplasty would be recommended now as the treatment of choice.

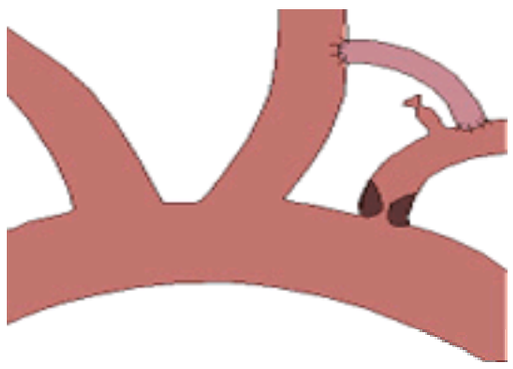


Fig. 5. Left carotid top left subclavian Dacron bypass graft.

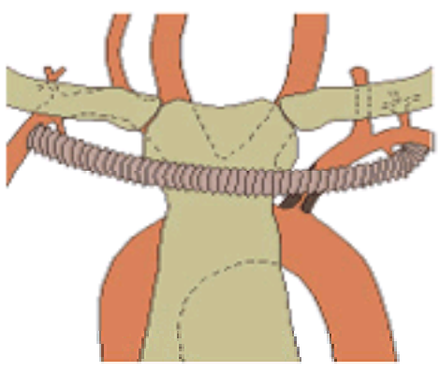


Fig. 6. Axilloaxillary graft, either saphenous vein or prosthetic, passes from between each axillary artery in front of the sternum.

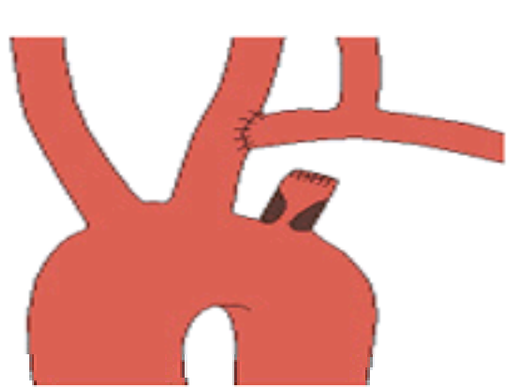


Fig. 7. The subclavian artery is divided in the root of the neck proximal to the origin of the vertebral artery and anastomosed end-to-side to the common carotid artery.

Innominate steal

Although it is possible to perform an endarterectomy directly on the origin of the innominate artery, the simplest approach is to run a Dacron graft off the arch of the aorta and anastomose it end-to-end to the distal innominate artery at its bifurcation (Fig. 8). The innominate artery and its origin from the arch of the aorta is best approached by a median sternotomy. If the patient's condition precludes a median sternotomy then a graft can be inserted between both common carotid arteries in the root of the neck.

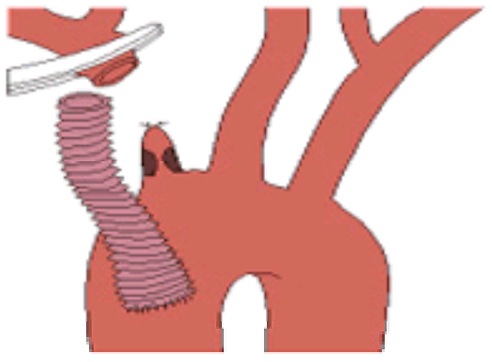


Fig. 8. The innominate artery is divided distal to the stenosis at its origin and a prosthetic graft (Gore-Tex or Dacron) run from the arch of the aorta to the distal innominate artery.

In the presence of complex disease, such as stenoses of the subclavian and innominate arteries, reconstruction can be satisfactorily performed with a bifurcated Dacron graft, the legs of the graft being anastomosed to the distal innominate artery and left subclavian artery distal to the stenosis but proximal to the vertebral artery.

Results

Surgery for vertebrobasilar ischaemia, which now rarely involves opening the chest, can be performed with minimal morbidity. However, it is much more difficult to evaluate the efficacy of surgery in terms of relief of symptoms because of the varied nature of the symptom complex. Nevertheless, with careful selection of patients,

favourable outcomes can be obtained. The outcome of surgery for innominate or subclavian artery steals in terms of symptom relief is much better in general than for vertebral artery surgery. If severe bilateral carotid stenoses are present in a patient with symptoms of vertebrobasilar ischaemia, the most appropriate procedure might well be a carotid endarterectomy.

Further reading

Alpers BJ, Berry RG, Paddison RM. Anatomical studies of the circle of Willis in normal brain. *Archives of Neurology and Psychiatry* 1959; **81**: 409–18. [What is a normal circle of Willis?]

Azakie A, McElhinney DB, Higashima R, Messina LM, Stoney RJ. Innominate artery reconstruction: over 3 decades of experience. *Annals of Surgery* 1998; **228**: 402–10. [A large experience of innominate artery reconstruction to relieve neurological symptoms or right arm symptoms.]

Berguer R. Vertebrobasilar ischaemia: inductions, techniques and results of surgical repair. In: Rutherford RB, ed., *Vascular surgery*, 4th edn, pp. 1574–88. WB Saunders, Philadelphia, 1995.

Berguer R, Morasch MD, Kline RA. A review of 100 consecutive reconstructions of the distal vertebral artery for embolic and hemodynamic disease. *Journal of Vascular Surgery* 1998; **27**: 852–9. [A major experience in reconstruction of the distal vertebral artery.]

Cartlidge NEF, Whisnant JP, Elveback LR. Carotid and vertebro-basilar transient cerebral ischaemic attacks: a community study, Rochester, Minnesota. *Mayo Clinic Proceedings* 1977; **52**: 117–20. [An early epidemiological study of the natural history of vertebro-basilar transient ischaemic attacks.]

17.7.3 Carotid body tumours

Linda Hands

Pathology
Incidence
Presenting symptoms
Diagnosis
Treatment
Conservative
Surgical
Long-term outcome
Further reading

Carotid body tumours arise from the chemoreceptor cells found at the carotid bifurcation. These are part of a widespread system of such cells that may give rise to tumours known as paragangliomas. Other examples of these are tumours of the vagal body or the glomus jugulare, and phaeochromocytomas. Patients can present at any age with a carotid body tumour but do so most commonly in the fifth and sixth decades of life. Although men were originally found to be more often affected, the sex ratio has recently reversed in favour of women.

Pathology

Carotid body tumours consist of epithelioid ‘chief' cells arranged in clusters separated by trabeculae of well-vascularized fibrous tissue. They have well-defined margins but lack a true capsule. Chemoreceptor cells are, in general, capable of secreting adrenaline and noradrenaline, but carotid body tumours are virtually always non-secretory and therefore not associated with symptoms of hypertension. A histological diagnosis of malignancy is based on cellular atypia, mitosis or local invasion, but tends to be unreliable in predicting tumour behaviour. Most behave in a benign fashion: local invasion and lymph-node metastases or haematogenous spread, especially to bone, are reported in less than 20 per cent of cases.

There appears to be a genetic susceptibility to these tumours, and paragangliomas that arise in families are more likely to be bilateral or multiple. Carotid body tumours appear to be inherited on an autosomal-dominant basis. Prolonged residence at high altitude is also associated with an increased incidence of these tumours.

Incidence

Paragangliomas are rare. Carotid body tumours are the most frequent, but even they are uncommon, an incidence of 0.012 per cent of surgical specimens being reported at one hospital. Approximately 10 per cent of these patients have either bilateral or multifocal tumours of the chemoreceptor system. Ten per cent of patients have a genetic basis to their disease and in these the incidence of bilateral/multifocal tumours is approximately 30 per cent.

Presenting symptoms

The most common complaint is of a lump in the neck; this may have been present for a considerable time: one patient was reported to have presented after 47 years! The carotid bifurcation lies close to many important and sensitive structures (Fig. 1) and so expansion of the tumour may lead to paresis of cranial nerves (VII, IX, X, XI, and XII), resulting in such symptoms as dysphagia, choking or hoarseness. Symptoms are reported by approximately 10 per cent of patients.

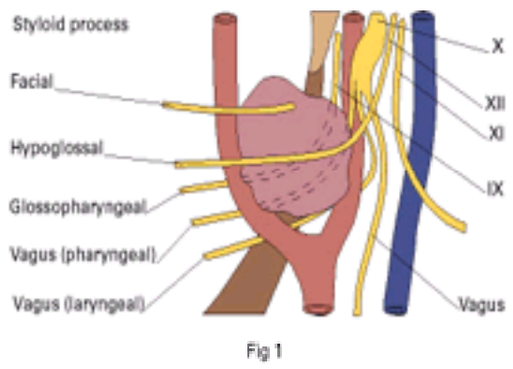


Fig. 1. Distribution of cranial nerves in relation to carotid body tumour. The cranial nerves are at greatest risk during dissection of the upper pole and the back of the tumour.

Diagnosis

It is vitally important to appreciate the possibility of a carotid body tumour in a patient who presents with a neck lump in the region of the carotid bifurcation. A misdiagnosis of lymph-node enlargement followed by an attempt at excision biopsy can result in excessive blood loss and much embarrassment. Fine-needle aspiration in the outpatient department may provide the answer, although this depends on the expertise of the local cytology department. Grey-scale ultrasonograpic imaging will confirm the relation to the carotid bifurcation and should demonstrate splaying of the two branches by the tumour. Duplex scanning is probably more appropriate; it gives more precise identification of the vessels and the extent of their involvement. Angiography is usually performed if duplex or ultrasound suggest a carotid body tumour: it demonstrates splaying of the carotid bifurcation, a tumour blush, and often allows tumour feeding vessels to be identified (Fig. 2). In the future, colour-coded duplex scanning may be able to replace angiography by demonstrating not only the distortion of the carotid bifurcation but also the tumour vascularity. However, confident identification of major feeding vessels for the embolization of large tumours (see below) will still require angiography. Enhanced CT images are also helpful in determining the upper level of a large tumour.

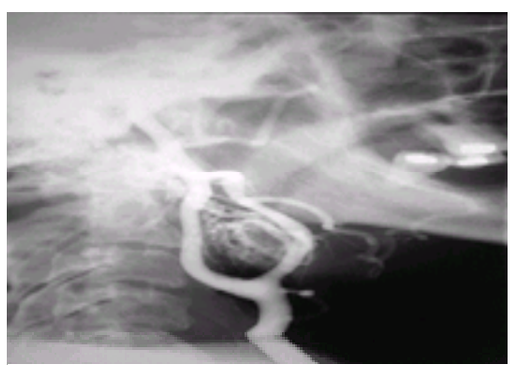


Fig. 2. Carotid angiogram showing the typical splaying of the carotid bifurcation by a carotid body tumour, and the origin of its blood supply from the external carotid artery.

Treatment

Conservative

If tumours are benign, their natural course is of slow enlargement with eventual compression of local structures, resulting in symptoms such as nerve palsies. In the elderly patient without symptoms it may be appropriate to do nothing. However, it is much easier and safer to remove the tumour before extensive local invasion and so in most patients a surgical approach when the tumour is small is advocated. Untreated, 75 per cent of asymptomatic patients eventually develop symptoms and 30 per cent will die from invasion of local structures or metastatic disease.

Surgical

Most patients in whom a diagnosis of carotid body tumour is made should undergo surgical excision.

Preoperative studies

Embolization

These tumours are very vascular and, if large, excision can result in excessive blood loss. In dealing with tumours of greater than 3 cm diameter it may be an advantage to shrink them preoperatively by embolization. It is essential that an experienced radiologist should perform this manoeuvre. There are communications between the main feeding vessels of the tumour from the external carotid and the vertebral artery, so embolization via the external carotid may result in brain embolization. Once the tumour has been successfully devascularized it is important that surgery is done within the next 2 or 3 days or neovascularization will occur and the advantage lost.

Direct and indirect pharyngoscopy/laryngoscopy

This is most important to assess cranial-nerve involvement but is also useful to look for invasion of the pharynx by upward extension of the tumour, a rare occurrence.

CT scan/MRI

Either investigation is used to assess the upward extension of the tumour and to determine any invasion of the skull base. MRI, especially with gadolinium enhancement, may provide some advantages over CT scanning: soft-tissue contrast is probably better, an advantage when assessing the extent of invasion by a large tumour, and greater sensitivity allows detection of tumours down to 5 mm in diameter.

Cross-match

There may be extensive blood loss during the surgery because of the vascular nature of the tumour and it is important that at least 4 units of blood be cross-matched.

Operation

Anaesthesia

General anaesthesia should be used and nasal rather than oral intubation may improve access to a tumour extending up under the mandible towards the base of the skull. Monitoring of central venous pressure is mandatory and an arterial line is useful in view of the possibility of excessive blood loss.

Surgeon

Experience with surgery of the carotid artery is important. It may be necessary to clamp the internal carotid, insert a shunt, and even to replace the artery with a length of saphenous vein. If the tumour is known to extend high in the neck, the assistance of a craniofacial surgeon is often helpful.

Operative procedure

The side of the neck and lower jaw are prepared and draped. The groin should also be prepared in case a length of saphenous vein is required. An incision is made on the anterior border of sternomastoid and dissection extended deeply until the carotid arteries are exposed in their sheath. The common, internal, and external carotids are each exposed and controlled beyond the tumour. Some surgeons advocate dividing the external carotid near its origin as soon as possible to reduce vascularity of the tumour and improve access (Fig. 3). Most would advocate at least clamping the artery at this point, although it is not often readily accessible. Heparin is not required before this manoeuvre.

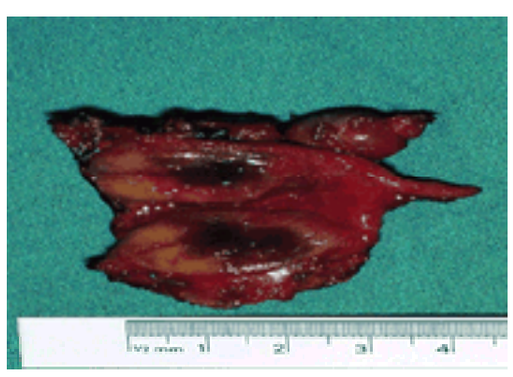


Fig. 3. The same carotid body tumour as in Fig. 2 after resection. In this case the external carotid artery was ligated at its origin, divided, and removed with the tumour.

The Connell technique in which a straight shunt is inserted via the common carotid artery into the internal carotid is another useful approach in that it excludes the external carotid from the circulation, thus effectively devascularizing the tumour. A straight Pruitt shunt is ideal for this purpose.

The tumour is dissected out along the white line that demarcates it from the adventitial tissue of the vessels. This dissection should start in the bifurcation and the tumour should be rolled upwards to get a clear view of the cranial nerves that run across the deep aspect of its upper part. Good visualization is essential, and haemostasis must be meticulous so that cranial nerves can be identified and preserved. However much these nerves are involved, it is usually possible to ‘shell' the tumour off them. The only exception is the tumour diagnosed as a carotid body tumour that proves to be a vagal body tumour. This arises very close to the carotid bifurcation and can be easily mistaken for the former. The vagus nerve itself is involved and therefore has to be sacrificed. If a carotid body tumour extends well up in the neck, fracture of the styloid process will improve access. It may be necessary to dislocate forward the temporomandibular joint to gain access to the upper border of the tumour. This is where preoperative CT scanning or MRI and perioperative assistance from a craniofacial surgeon become invaluable.

The Shamblin classification of carotid body tumours groups them according to the degree of invasion of the arterial wall. Those in group 1 are easy to remove from the vessels, those in group 2 require dissection in a subadventitial plane to remove them from the vessel, and for those in group 3 the tumour encircles and invades the vessel to such an extent that it usually requires complete arterial excision and replacement with a length of saphenous vein. A shunt may be used, but it is often useful to have some means of cerebral monitoring such as EEG or transcranial Doppler to assess distal flow during the procedure. Approximately 50 per cent of carotid body tumours are in Shamblin group 2 and 25 per cent in each of the other two groups.

Once the tumour has been removed and adequate haemostasis achieved the wound should be closed over a vacuum drain, which is left in for about 24 h or until significant discharge has ceased.

Postoperative care

The patient should be closely monitored postoperatively for central and peripheral neurological deficits (especially of cranial nerves IX, X, and XII, cervical sympathetic nerves, and the marginal mandibular branch of VII) and blood pressure well controlled, particularly if the internal carotid artery has been reconstructed. The incidence of cranial-nerve defects associated with surgery has not declined over the years despite improvements in surgical technique (e.g., preoperative embolization to reduce bleeding). Such problems are found particularly after the removal of large tumours and are reported in 10 to 20 per cent of cases, although most recover within a few weeks. Bilateral carotid body tumours should not be removed in the same operation. The incidence of nerve damage following surgery is sufficiently high to make bilateral palsies of the recurrent laryngeal and hypoglossal nerves a significant risk leading to respiratory and swallowing problems postoperatively. For patients who have had surgery to a carotid body tumour resulting in cranial-nerve palsy or carotid occlusion, the risks of operation on a contralateral tumour are considerable and radiotherapy (see below) may be preferable. Cerebrovascular accidents occur in less than 3 per cent of cases in most series.

Although the mortality from removal of carotid body tumours used to be high (in 1950 it was reported as about 30 per cent), it now should be virtually zero because of innovations such as preoperative embolization, clamping of the external carotid artery, the use of shunts, and well-planned strategies for access high in the neck.

Radiotherapy

Radiotherapy has been promoted by some groups as a primary treatment. Few studies have compared this with surgery and none of these has been randomized. Most centres still recommend surgery, with radiotherapy used only for tumours too extensive to excise or as a back-up treatment for those with recurrence. Radiotherapy as a primary treatment appears to slow progression of the disease, but this is difficult to prove in a disease that has a naturally slow progression.

Long-term outcome

Most patients are cured by surgery and any cranial-nerve palsies associated with the procedure are usually temporary, although those caused by prolonged compression from the tumour itself are sometimes permanent. A few patients prove to have a malignant tumour, which may have been diagnosed histologically but usually only comes to light when metastatic disease develops. If the tumour is known to be malignant because of local invasion or metastatic disease, local radiotherapy can be used to prevent local recurrence or to treat metastasis as it occurs. Most tumours are slow growing and survival for many years is possible even with established metastatic disease.

Further reading

Hallett JW, Nora JD, Hollier LH, Cherry KJ, Pairolero PC. Trends in neurovascular complications of surgical management for carotid body and cervical paragangliomas: a fifty year experience with 153 tumours. *Journal of Vascular Surgery* 1988; **7**: 284–91. [Retrospective study of chemodectoma surgery at one institution over three different time periods to assess effect of changes in techniques on mortality and morbidity.]

LaMuraglia EM *et al.* The current surgical management of carotid body paragangliomas. *Journal of Vascular Surgery* 1992; **15**: 1038–44. [Retrospective study of carotid body tumour surgery at one institution looking at the effects of preoperative embolization and perioperative EEG monitoring, in particular on outcome.]

Netterville JL, Reilly KM, Robertson D, Reiber ME, Armstrong WB, Childs P. Carotid body tumors – a review of 30 patients with 46 tumors. *Laryngoscopy* 1995; **105**: 115–26. [Study of presentation, management, and outcome of carotid body tumour surgery at one institution.]

McPherson GA, Halliday AW, Mansfield AO. Carotid body tumours and other cervical paragangliomas: diagnosis and management in 25 patients. *British Journal of Surgery* 1989; **76:** 33–6. [Review of presentation, management, and outcome of a series of cervical paragangliomas managed by one surgeon.]

Vogl T *et al.* Paragangliomas of the jugular bulb and carotid body: MR imaging with short sequences and GA-DTPA enhancement. *American Journal of Roentgenology* 1989; **153**: 583–7. [CT and MR scanning of head and neck paragangliomas.]

Williams MD, Phillips MJ, Nelson WR, Rainer WG. Carotid body tumour. *Archives of Surgery* 1992; **127**: 963–8. [Presentation, management, and outcome of carotid body tumours at one institution over 34 years.]

17.8.1 Thoracic outlet obstruction

Peter J. Morris

Introduction
Anatomy
Clinical presentation
Investigations
Differential diagnosis
Treatment of neurological or arterial thoracic outlet compression
Physiotherapy
Surgery
Results
Treatment of axillary–subclavian vein thrombosis
Further reading

Introduction

Obstruction of the subclavian artery or vein and pressure on the lower trunk of the brachial plexus may occur in the thoracic outlet due to a number of anatomical abnormalities, the best recognized being a cervical rib. The first successful removal of a cervical rib was undertaken by Coote in 1861. The clinical syndrome of thoracic outlet obstruction is a complex of clinical features, which may be either predominantly vascular or predominantly neurological. The syndrome has been given a number of descriptions over the years, each suggesting an aetiology, such as costoclavicular syndrome, scalenus anticus syndrome, and hyperabduction syndrome. However, the varied symptoms and signs with which these patients present are best grouped under the broad term of thoracic outlet obstruction syndrome.

Anatomy

The subclavian vein and artery cross the anterior half of the first rib, separated by the insertion of the scalenus anterior into the scalene tubercle of the first rib (Fig. 1). The brachial plexus lies behind the vessels; the lower trunk of the plexus formed by the roots of C8 and T1 is the most inferior structure of the plexus and lies on the first rib. As these structures cross the first rib they lie behind and below the clavicle and the subclavius muscle arising from the inferior aspect of the clavicle and inserting into the costochondral junction. After crossing the first rib the neurovascular bundle passes into the axilla below the coracoid process of the scapula with its attached pectoralis minor.

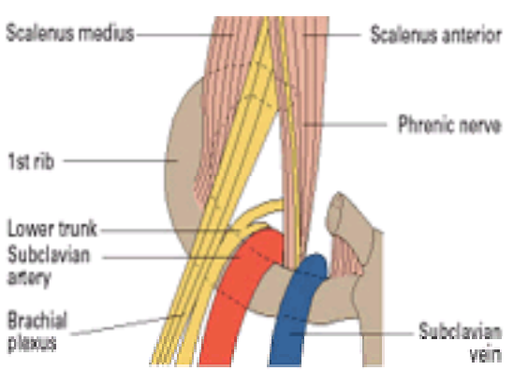


Fig. 1. The subclavian vein crosses the first rib in front of the insertion of the scalenus anterior, which separates it from the subclavian artery and the brachial plexus.

The most common cause of obstruction or pressure at the thoracic outlet is a cervical rib, which passes beneath the brachial plexus and subclavian artery to attach to the scalene tubercle on the first rib. Cervical ribs are relatively common, occurring in around 0.4 per cent of the population. Not infrequently a cervical rib is incomplete but is replaced by a fibrous band, which extends from the tip of the rib again to the scalene tubercle. The lower trunk or the root of T1 of the brachial plexus may be stretched over a cervical rib or a fibrous band, as may the subclavian artery. This may also be squeezed in the angle between the scalenus anterior insertion and the cervical rib in certain positions (Fig. 2). Pressure on the lower trunk of the brachial plexus is sometimes due to a fibrous band lying along the medial border of the scalenus medius as it passes to its insertion in the first rib. The insertion of the scalenus anterior, which is quite fibrous, may extend back along the first rib beneath the subclavian artery and obstruct it when the arm is in certain positions. The space between the clavicle and the first rib may be diminished in those with an old malaligned fracture of the clavicle or a hypertrophied subclavius muscle as occurs in butterfly swimmers. This space also decreases as a result of poor posture: it is not uncommon for women to present with symptoms of thoracic outlet obstruction, either with or without a cervical rib, as they approach middle age and develop a drooping posture around the shoulder girdle.

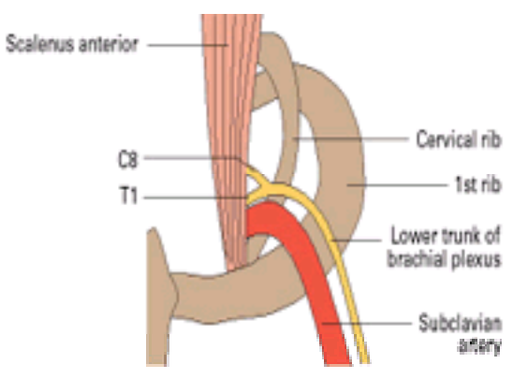


Fig. 2. A complete cervical rib attaching to the first rib at the scalene tubercle, with the subclavian artery elevated over it.

The axillary vein may be obstructed as it crosses the first rib in front of the scalenus anterior to become the subclavian vein if the space between the clavicle and first rib is decreased: this is usually associated with a large subclavius muscle or a fibrous anterior extension of the insertion of the scalenus anterior blending with the fibrous capsule of the costochondral junction.

Clinical presentation

Patients present with any of a variety of symptoms and signs which may be predominantly vascular, predominantly neurological, or a combination of both. The majority of patients are women, usually between the ages of 20 and 40. Those with evidence of arterial obstruction may present with a history of Raynaud's phenomenon, nearly always unilateral, or of inability to use the arm on the involved side above the head for any period of time because of weakness in the arm and hand or pain and numbness in the fingers. On examination the subclavian artery may be prominent in the root of the neck when elevated by a cervical rib, which may indeed itself be palpable, and obliteration of the pulse can be readily achieved by a variety of manoeuvres, such as hyperabduction of the arm and bracing of the shoulders, which will be associated with pallor of the same hand. A bruit may be heard at rest or, more commonly, as the artery is compressed with one of the above manoeuvres. Occasionally, older patients present with features of a subclavian aneurysm (which is really a post-stenotic dilation of the artery beyond an obstruction), giving rise to distal emboli from thrombus within the lumen of the aneurysm. Emboli from a subclavian aneurysm can produce distal signs ranging from splinter haemorrhages to gangrene of the fingertips (Fig. 3).

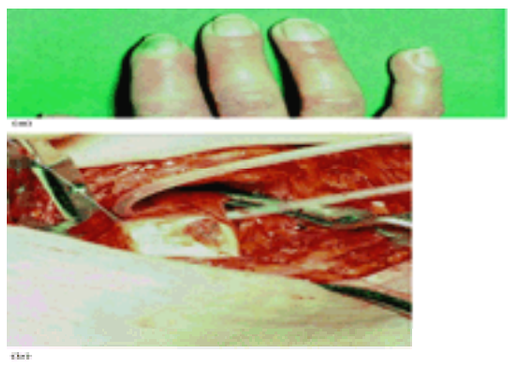


Fig. 3. A 62-year-old woman presented with (a) splinter haemorrhages in fingers (see nail of fifth finger) suggesting a proximal source of emboli. (b) She had a post-stenotic aneurysm beyond a cervical rib compressing the artery, which at operation was found to contain thrombus.

Patients with predominantly neurological symptoms often represent a difficult diagnostic group. Classically, pressure on the lower trunk of the brachial plexus (C8 and T1) results in paraesthesia of the ulnar aspect of the hand and forearm associated with certain postures, such as carrying a heavy basket or attempting to use the hand above the head. By the time of presentation there may be obvious weakness and wasting of the small muscles of the hand. There may be a tender point in the root of the neck over the plexus as it crosses a cervical rib. It is often common for the patient to complain of pain with no appropriate segmental distribution in the shoulder, upper arm, or neck. Many patients present with a combination of both vascular and neurological symptoms, but in which one or other complex of symptoms tends to be dominant.

Venous obstruction at the thoracic outlet is an uncommon presentation, but this diagnosis should always be suspected in young people who present with an axillary vein thrombosis or intermittent swelling of the arm. This is commonly known as subclavian–axillary vein effort thrombosis, first described by Paget and von Schroetter in the late nineteenth century. Patients often present with symptoms in the dominant arm after a period of intense activity. It is also seen in young people who participate in sports that can lead to hypertrophy of the subclavius muscle, such as surfboard paddling or butterfly swimming.

Investigations

A radiograph of the neck and upper chest will demonstrate cervical ribs (Fig. 4), but the absence of a rib does not exclude the diagnosis of thoracic outlet obstruction due to a fibrous remnant of an incomplete rib. A cervical rib, either complete or incomplete, will be visible on the radiograph, as will a prominent transverse process of the seventh cervical vertebra in the presence of a fibrous band. The presence of a cervical rib does not necessarily mean that this is the cause of the presenting symptoms. Cervical ribs are not uncommon (4/1000) and in most people do not give rise to problems.



Fig. 4. A plain radiograph of the neck showing bilateral cervical ribs.

Magnetic resonance imaging (**MRI**) is a useful addition to our diagnostic procedures, as it has the capacity to identify the brachial plexus and angulation of the lower trunk of the plexus over a fibrous band (Fig. 5). Our experience of MRI in the investigation of thoracic outlet obstruction in Oxford suggests that it is a valuable investigation in the neurological cases.

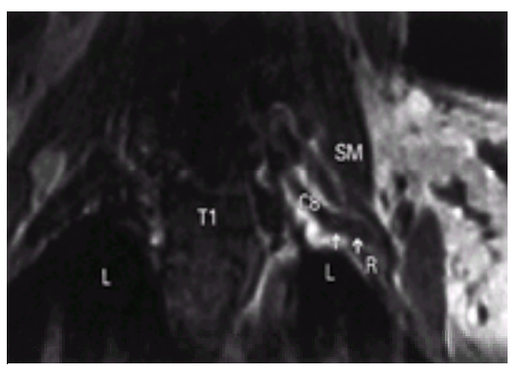


Fig. 5. MRI of thoracic outlet syndrome. A coronal T_1-weighted gradient recalled echo image through the apices of the lungs (L) and the body of the first thoracic vertebra (T1). The root of C8 (C8) is seen passing obliquely on scalenous medius muscle (SM). The C8 root is elevated (arrows) as it crosses the left first rib (R). (By courtesy of Dr N. Moore.)

An arteriogram will demonstrate obstruction of the artery at the outlet with the arm in a variety of positions, such as abduction (Fig. 6), but this can be established clinically and does not need to be confirmed by angiography. The main indication for an arteriogram is to exclude the presence of thrombus within a possible subclavian aneurysm. However, the presence of both an aneurysm and intraluminal thrombus can be established by Duplex scanning, which further reduces the need for angiography.

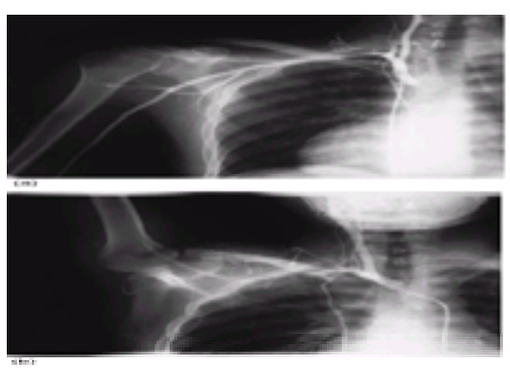


Fig. 6. (a) An arteriogram of the subclavian artery with the arm at the side, and (b) with the arm abducted showing complete obstruction of the artery at the thoracic outlet.

A venogram will always be performed in a young patient presenting with an axillary–subclavian venous thrombosis and in the patient with intermittent swelling of the arm. A characteristic picture may be seen in the presence of thrombosis with collaterals evident in patients with more long-standing symptoms. Sometimes a clinical diagnosis of axillary vein thrombosis will not be confirmed on venography, the clinical features being due to obstruction of the vein only at the thoracic outlet (Fig. 7).



Fig. 7. A venogram in a patient with a diagnosis of subclavian vein thrombosis but in whom the vein was found to be patent at operation but obstructed at the thoracic outlet.

Neurophysiological studies are essential in establishing a neurological component to thoracic outlet obstruction and can provide a high degree of diagnostic accuracy in establishing the site of the problem at the brachial plexus level rather than at the level of the cervical spine or more distally at the elbow or carpal tunnel.

In established ‘classic' thoracic outlet syndrome there is a characteristic pattern of electrophysiological abnormality. The compound muscle action potential recorded from the thenar eminence is low, but median motor conduction velocity in the forearm and distal motor latency are normal. Preserved sensory nerve action potentials from the thumb, and index and third fingers, in the distribution of the C6 and C7 roots of the plexus, with normal conduction velocities across the wrist, serve to differentiate the condition from carpal tunnel syndrome. In contrast, the sensory nerve action potential from the fifth finger, in the C8 dermatome, is relatively reduced in amplitude or is absent. The hypothenar compound muscle action potential may also be reduced in amplitude, but less so than the response from the abductor pollicis brevis. Normal motor and sensory conduction between the mid-arm and wrist exclude a peripheral ulnar nerve lesion.

Electromyography of the intrinsic hand muscles usually shows signs of chronic partial denervation consisting of scanty fibrillation potentials, prolonged polyphasic unstable motor unit action potentials which may have high amplitudes, and reduced density of the interference pattern. These findings are more prominent in the abductor pollicis brevis than in the first dorsal interosseous muscle, and in a few patients similar changes are seen in the forearm flexor compartment.

Other neurophysiological procedures which may be of help in diagnosis include measurement of F response (a late spinal response) latency, reflex latency measurement, recording of sensory nerve action potentials from the medial cutaneous nerve of the forearm, and cortical somatosensory responses following dermatomal stimulation. However, these tests are unlikely to be of value if the more characteristic changes are not observed. Measurement of proximal ulnar motor conduction velocity across the brachial plexus is now generally disregarded as a method of diagnosing the thoracic outlet syndrome. Typical findings in a patient with thoracic outlet compression due to a cervical rib are shown in Table 1.

Motor conduction	
Right median nerve	
Abductor pollicis brevis neurolysis	
Right ulnar nerve	
Abductor digiti minimi	
Compound muscle action potential	11mV (normal)
Conduction velocity	54m/s (normal)
F response latency	27.4ms (normal)
Sensory conduction	
Right thumb	
Amplitude	37μV (normal)
Conduction velocity	50m/s (normal)
Right index finger	
Amplitude	27μV (normal)
Conduction velocity	50m/s (normal)
Right fifth finger	
Amplitude	2μV (low)
Conduction velocity	33m/s (normal)
Mixed nerve action potential	
Right ulnar nerve wrist to mid-arm	
Amplitude	8μV (low)
Conduction velocity	63m/s (normal)
Electromyography	
Right first dorsal interosseous	
Fibrillation potentials, unstable polyphasic motor unit action potentials, reduced interference pattern	
Right flexor digitorum superficialis	
Fibrillation potentials, unstable polyphasic motor unit action potentials, reduced interference pattern	
Right flexor carpi ulnaris	
No abnormal spontaneous activity, normal motor unit action potential outline but high amplitudes (>10mV), normal interference pattern	

Table 1 Results of electrophysiological studies in a patient with thoracic outlet compression due to a cervical rib

Differential diagnosis

The diagnosis of a vascular thoracic outlet obstruction is usually relatively straightforward, but may present problems in patients with bilateral Raynaud's phenomenon. However, symptoms due to thoracic outlet obstruction usually have an earlier onset in one hand. If the presenting feature is unilateral Raynaud's phenomenon, it is important to exclude an occupational cause such as the use of hand-held drills. It should also be remembered that arterial obstruction may occur proximal to the outlet, at the origin of the left subclavian or innominate arteries, and where this possibility exists an arch aortogram will be required.

The two major neurological diagnoses to be excluded are cervical spondylosis and carpal tunnel syndrome. MRI is the investigation of choice in the former, and neurophysiological studies the most useful investigation in the latter, but neither replaces a careful history and examination.

Treatment of neurological or arterial thoracic outlet compression

Physiotherapy

Exercises designed to strengthen the muscles of the neck and shoulder girdle and the adoption of a more upright posture can relieve symptoms in many patients, especially women approaching middle age who often begin to droop across the shoulders, resulting in narrowing of the thoracic outlet. This should be the first approach in most patients unless they have florid symptoms of obstruction, definite neurological signs, or arterial changes, in which case it is better to proceed to surgery.

Surgery

The surgical approach to this condition depends on the cause. Cervical ribs, either complete or fibrous, should be excised. Additional bands in the scalenus medius are divided. Division of the scalenus anterior at its insertion (scalenotomy) has been widely practised in the past as a definitive therapy but is no longer considered adequate treatment in most patients. In any case this muscle will be divided if the neck is being explored by the supraclavicular exposure (see later). However, definite evidence of vascular or neurological thoracic outlet obstruction in the absence of a cervical rib or fibrous band is now treated by excision of the first rib. If there is evidence of arterial damage such as a significant post-stenotic dilatation or an aneurysm of the subclavian artery, that section of the artery is excised and replaced with a short length of vein or a prosthetic graft such as polytetrafluoroethylene. If a graft is to be inserted it may be necessary to expose the axillary artery below the clavicle.

Surgical techniques

There are several techniques by which the above surgical procedures are performed and each has its protagonists. To some extent the technique should be dictated by the cause of the thoracic outlet obstruction.

Supraclavicular approach

With the patient in the prone position and a small sandbag behind the shoulders, a transverse incision is made above and parallel to the clavicle at the base of the anterior triangle. This is carried down through the platysma and the supraclavicular pad of fat is reflected upwards to expose the scalenus anterior muscle. The phrenic nerve crosses the anterior surface of the muscle from lateral to medial and should be carefully separated from the muscle and protected with a sling. Attention should be paid to the possible presence of an accessory phrenic nerve. The scalenus anterior muscle is then divided near its insertion, to expose the subclavian artery, which is immediately behind the muscle.

If a cervical rib, a fibrous remnant, or any other fibrous band is present, it will be obvious on palpation, and can be exposed by further dissection. The major advantage of the supraclavicular approach is that it allows the anatomy of the thoracic outlet to be fully visualized, together with any abnormality. A cervical rib is excised, being careful to take the rib back behind the plexus. This last part of the excision is usually best performed from the lateral aspect of the brachial plexus. A fibrous remnant of a cervical rib is divided and, depending on its length, an incomplete cervical rib is removed to ensure that the lower trunk of the brachial plexus is free of any potential pressure. Any other bands that are apparent, such as along the medial border of scalenus medius are also divided, especially in the absence of a cervical rib or fibrous remnant.

The supraclavicular approach is ideal for removal of a cervical rib or a fibrous remnant, and also allows inspection of the artery. If a segment of the artery has to be replaced, the axillary artery can be exposed below the clavicle for distal control and placing of the distal anastomosis.

The axillary–subclavian vein is not seen through the supraclavicular approach and so it does not allow a possible venous obstruction to be defined; this is best achieved by the infraclavicular approach.

The first rib can be removed from above, but this procedure is less easy as exposure of the rib is often less than satisfactory. Removal of the first rib is best accomplished through the transaxillary approach or the infraclavicular approach.

Transaxillary approach

This is a popular method of resection of the first rib to relieve thoracic outlet compression and was first described by Roos in 1966. Some protagonists of this technique would argue that removal of the first rib and hence detachment of everything that attaches to it, such as the scalenus anterior, cervical rib, and fibrous bands, ensures that the thoracic outlet obstruction is relieved whatever the cause.

With the patient in a full lateral position, the upper arm is abducted to a right angle and forcibly retracted by an assistant. A transverse incision is then made on the medial aspect of the axilla, between the borders of the pectoralis major anteriorly and the latissimus dorsi posteriorly, and deepened to expose the first rib. Care must be taken not to damage the nerve to the serratus anterior. The first rib is then excised after separating it from the attaching structures. The secret of good exposure is vigorous retraction of the arm, which usually requires assistants relieving one another at frequent intervals! This approach has the disadvantage that adequate access to the artery and vein is not possible.

Infraclavicular approach

The positioning of the patient is all important. The patient lies with a sandbag between the shoulder blades and a transverse incision is made beneath the clavicle from the coracoid process laterally towards the midline for some 5 cm. This incision is carried down to the pectoralis major, which is then divided with diathermy to expose the acromioclavicular fascia, incision of which allows the axillary vein and artery to be identified with the plexus behind the artery. The shoulder is then lifted forward, opening up the costoclavicular space like an oyster and allowing the vein and artery to be retracted from the first rib, the anterior half of which is readily visualized. The rib is then excised as far forward as required and then back beyond the plexus. This approach is ideal for patients with venous or arterial obstruction, but has one disadvantage in that as excision of the rib proceeds posteriorly to take it back beyond the plexus, the view of the rib is sometimes poor.

Posterior approach

This is practised infrequently, for although it provides a good view of the neck and posterior part of a cervical rib or a first rib, the artery is not accessible, and it is a much more destructive approach than the others described. It is sometimes useful if a cervical sympathectomy needs to be performed in a patient who has already undergone several neck explorations.

Results

When an operation is performed for arterial obstruction with a clearly established diagnosis, the results are excellent (Table 2). If neurological compression at the thoracic outlet has been diagnosed without doubt, surgical relief of the obstruction will also produce a good outcome. If there is muscle wasting of the small muscles of the hand, recovery will be slow; if this is long-standing it is unlikely to improve, although sensory symptoms and pain are relieved immediately. However, the diagnosis of neurological thoracic outlet compression is often far from clear, and in some instances surgery is performed almost as a diagnostic procedure, when both the patient and neurologist are desperate. Although some vascular surgeons feel that to be certain of relieving symptoms the first rib should always be removed even if a cervical rib is present, the Oxford experience does not support this concept (Table 2). Recurrence of symptoms usually happens in the first 12 months if they occur, but may occur later in a small number of patients due to fibrous scarring resulting in bands which produce pressure on the lower trunk of the plexus once again, or due to regrowth of a resected first rib, which may occur if the rib is removed subperiosteally. For this reason ribs should be removed with the periosteum.

	Presentation	*No.*	*Cured*	*Improved*	*Not improved*
Cervical rib	Vascular	8	6	2	-
	Vascular/neurological	8	8	-	-
	Neurological	5	5	-	-
	Total	21	19	2	-
First rib	Vascular	17	8	8	1
	Vascular/neurological	7	6	-	1
	Neurological	5	3	-	2
	Total	29	17	8	4

Table 2 Symptomatic outcome after removal of a cervical or first rib in 53 patients in Oxford presenting with thoracic outlet compression syndrome of a predominant vascular or neurological nature or a mixture of both

Treatment of axillary–subclavian vein thrombosis

Since the advent of thrombolysis and stenting the management of this condition has become much more controversial. In the patient presenting acutely with so-called effort vein thrombosis the diagnosis of thrombosis will be confirmed by venography. Thrombolysis should be commenced and may result in complete or partial clearance of thrombus. It is then essential to decide whether thoracic outlet obstruction is the underlying cause of the thrombosis. This will commonly be the case in active younger people, and with respect to the vein, this can usually be established by positional venography.

If thoracic outlet compression is present then there is no agreement as to when decompression of the thoracic outlet should be performed, but a reasonable

compromise might be 3 weeks after thrombolysis. At this time the first rib is removed by one of the approaches described earlier, but the author would favour the infraclavicular approach. If thrombolysis has only been partially successful, some surgeons would favour venotomy and thrombectomy at the same time as removal of the first rib. The outcome of surgery for acute axillary–subclavian vein thrombosis due to thoracic outlet obstruction is good.

In the patient who presents with intermittent and long-standing symptoms and who is found to have axillary–subclavian vein thrombosis with well established collaterals and evidence of thoracic outlet compression, it is still worthwhile to remove the first rib to decompress the thoracic outlet as this not only allows better collateral venous drainage in all positions of the arm, but may indeed favour recanalization of the thrombosed vein.

Further reading

Adson AW, Coffey JR. Cervical rib: a method of anterior approach for relief of symptoms by division of the scalenus anticus. *Annals of Surgery* 1927; **85**: 839–57.

Azakie A *et al.* Surgical management of subclavian-vein effort thrombosis as a result of thoracic outlet compression. *Journal of Vascular Surgery* 1998; **28**: 777–86. [An important description of 33 patients with this condition who for the most part were successfully treated by thrombolysis and early decompression of the thoracic outlet.]

Coote H. Exostosis of the left transverse process of the seventh cervical vertebra surrounded by blood vessels and nerves: successful removal. *Lancet* 1861; **i**: 360.

Davies AH, Walton J, Stuart E, Morris PJ. Surgical management of the thoracic outlet syndrome. *British Journal of Surgery* 1991; **78**: 1193–5.

Eastcott HHG. Reconstruction of the subclavian artery for complications of cervical rib and thoracic outlet syndrome. *Lancet* 1962; **ii**: 1243–6.

Kieffer E. Arterial complications of the thoracic outlet syndrome. In: Bergan JJ, Yao JST, eds, *Evaluation and treatment of upper and lower extremity disorders*, pp. 249–75. Grune & 38; Stratton, Orlando, 1984.

Lewis T, Pickering GW. Observations on maladies in which the blood supply to digits ceases intermittently or permanently, and upon bilateral gangrene of digits; observations relevant to so-called Raynaud's disease. *Clinical Science* 1934; **1**: 327–66. [A classic physiological paper related to ischaemia of the upper limbs.]

Murphy JB. Case of cervical rib with symptoms resembling subclavian aneurysm. *Annals of Surgery* 1905; **41**: 399–406.

Murphy T. Brachial neuritis caused by pressure of 1st rib. *Australian Medical Journal* 1910; **15**: 582–5. [An early description of a neurological presentation associated with the first rib.]

Peet RM *et al.* Thoracic outlet syndrome: evaluation of a therapeutic exercise programme. *Mayo Clinic Proceedings* 1956; **31**: 281–7.

Roos DB. Transaxillary approach for 1st rib resection to relieve thoracic outlet syndrome. *Annals of Surgery* 1966; **163**: 354–8. [The original description of axillary resection of the first rib.]

Roos DB. New concepts of thoracic outlet syndrome that explain etiology, symptoms, diagnosis and treatment. *Vascular Surgery* 1979; **13**: 313–21.

Rutherford RB. Primary subclavian–axillary vein thrombosis: the relative roles of thrombolysis, percutaneous angioplasty, stents and surgery. *Seminars in Vascular Surgery* 1998; **11**: 91–5. [A comprehensive review of the various possible treatments of axillary–subclavian venous thrombosis.]

Strange-Vognsen HH *et al.* Resection of the first rib, following deep arm vein thrombolysis in patients with thoracic outlet syndrome. *Journal of Cardiovascular Surgery* 1989; **30**: 430–3.

Thompson JF, Webster JH. First rib resection for vascular complications of the thoracic outlet syndrome. *British Journal of Surgery* 1990; **77**: 555–7.

Tilney P, Griffiths HTG, Edwards EA. Natural history of major venous thrombosis of the upper extremity. *Archives of Surgery* 1976; **101**: 792–6. [An important description of outcome in the absence of treatment.]

Urschel HC Jr, Razzuk MA. Neurovascular compression in the thoracic outlet: changing management over 50 years. *Annals of Surgery* 1998; **228**: 609–17. [A review of an enormous experience of some 15 000 cases evaluated for thoracic outlet syndrome.]

17.8.2 Raynaud's syndrome

Peter M. Lamont

[Introduction](#)
[Incidence](#)
[Pathogenesis](#)
[Associated disorders](#)
- [Connective tissue disease](#)
- [Occlusive arterial disease](#)
- [Trauma](#)
- [Neurovascular lesions](#)
- [Miscellaneous](#)

[Diagnosis](#)
[Differential diagnosis](#)
- [Acrocyanosis](#)
- [Livido reticularis](#)

[Treatment of Raynaud's syndrome](#)
- [General measures](#)
- [Drug therapy](#)
- [Surgical therapy](#)
- [Summary of treatment](#)

[Further reading](#)

Introduction

Maurice Raynaud first described a group of 25 patients with ‘local asphyxia and symmetrical gangrene of the extremities' in 1862. He proposed that the observed changes were caused by vasospasm because most of the patients had palpable pulses at the wrist and patent large arteries were observed in those patients who underwent autopsy. Over the next 70 years it became apparent that this vasospastic condition could occur either in isolation or in association with other diseases, and so a distinction was made between Raynaud's disease (occurring in isolation) and Raynaud's phenomenon (secondary to an underlying disease). More recently this distinction has become less clear cut because some patients may develop collagen disorders several years after the onset of Raynaud's symptoms and patients with apparent primary Raynaud's disease may have low levels of autoantibodies present in their serum. For these reasons both primary and secondary Raynaud's conditions are now classified under the single banner of Raynaud's syndrome.

Raynaud's syndrome therefore describes the changes which result from intermittent vasospasm of the arterioles in the hands or feet; these classically occur following exposure to cold or as a result of emotional stimuli. The vasospasm resolves with warming, leading to the classic sequence of colour changes in the extremity from pallor to cyanosis to redness although, as with many classic descriptions of diseases, the full spectrum of colour changes is seldom seen in one individual patient. The syndrome may be primary, when no other condition can be identified after appropriate investigation, or it may be secondary to a wide variety of underlying disorders.

Incidence

The majority of patients with Raynaud's syndrome are female, both because primary Raynaud's is classically described as a disease of young women and because many of the underlying disorders associated with secondary Raynaud's are more common in women. Of those who actually present with the syndrome, 70 to 90 per cent are female, although many individuals with mild to moderate cold sensitivity may not be disturbed by their symptoms enough to seek medical advice. Population surveys of Scandinavian women aged 18 to 60 years suggest an overall prevalence of Raynaud's syndrome of between 15 and 22 per cent in the general female population—a substantially higher number than is actually treated. Most of the women uncovered by such population surveys have primary Raynaud's syndrome and very few have associated collagen disorders. Secondary Raynaud's syndrome is much more common in patients who present to specialist units with an interest in the disorder. Only 30 per cent of patients presenting to specialist units have primary Raynaud's syndrome; the remainder have secondary Raynaud's syndrome, and one-half of these patients will have overt or suspected connective tissue diseases. This represents the single largest group of patients presenting to specialist units with Raynaud's syndrome. Such data confirm the clinical impression that secondary Raynaud's syndrome results in a greater severity of digital ischaemia than the primary form; patients with secondary Raynaud's syndrome are therefore more likely to present to a clinician for treatment.

Pathogenesis

Distinct colour changes in the hands or feet in response to cold are the hallmark of the syndrome. Cold exposure produces profound pallor and numbness of the digits due to spasm of the digital arteries. The digital microvasculature dilates after a few minutes due to the accumulation of carbon dioxide and the products of hypoxic metabolism. As the vasospasm begins to relax a small amount of oxygenated blood enters these dilated vessels, where it rapidly becomes desaturated and the pallor changes to cyanosis. As the digital vessels relax further, so normal blood flow is re-established and a reactive hyperaemia of the dilated microvasculature ensues as the cyanosis changes to rubor of the digits (Fig. 1).



Fig. 1. The classic appearance of Raynaud's syndrome. The digits exhibit cyanosis changing to rubor as the hand rewarms during a Raynaud's attack.

Cold, emotional stimuli, or even cigarette smoking may induce an attack, and so severe is the vasospasm that its force overcomes the ability of the arterial blood pressure to keep the vessel walls apart, so that the digital arteries are completely closed during the attack (Fig. 2).

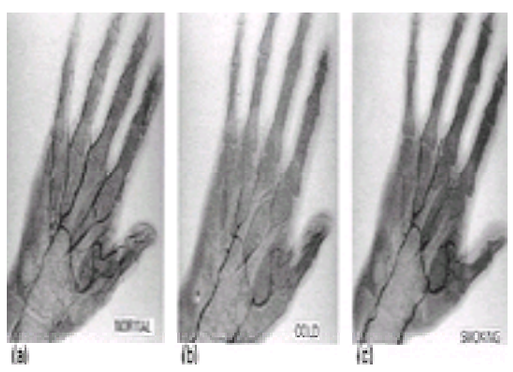


Fig. 2. The induction of digital vasospasm by cold and smoking. (a) The apparently normal angiogram of a 24-year-old female with primary Raynaud's syndrome. (b)

Intense vasospasm prevents filling of the digital vessels after cold exposure in the same patient. (c) Cigarette smoking has the same vasospastic effect.

The definitive picture of the pathogenesis of Raynaud's syndrome has yet to be described. The very number and variety of hypotheses put forward to explain the abnormal sensitivity to cold of the digital vessels attests to the presence of a continuing gap in our knowledge. Maurice Raynaud was the first to put forward his ideas on the pathogenesis of the digital vasospasm when he claimed that it was due to an ‘enormous exaggeration of the excitomotor energy of the gray parts of the spinal cord which control the vasomotor innervation'. This eloquent description of sympathetic overactivity was challenged many years later by Sir Thomas Lewis, who published a series of papers on the subject in the 1930s. Lewis attributed the vasospasm to a local fault in the sensitivity of the digital arteries to cold stimuli and was able to induce attacks experimentally in the presence of an anaesthetized sympathetic supply. Conversely he was able to prevent an attack by keeping the hand warm when the rest of the body was cooled. Thus he concluded that, whatever the state of the sympathetic nervous outflow, provocation of a Raynaud's attack depended upon the local digital temperature. So persuasive were Lewis' experiments that the majority of subsequent work has concentrated on possible mechanisms of a local hypersensitivity to cold, and most workers have accepted that the central thermoregulatory system is normal in Raynaud's syndrome.

The normal vasospastic response to cold may be exaggerated if the digital vessel lumen is already narrowed by structural abnormalities. Microscopic abnormalities, ranging from slight intimal thickening through to frank intimal hyperplasia or even complete luminal obliteration, have been described in patients with this syndrome. The intimal thickening in most cases, however, is no worse than that in age-matched controls without Raynaud's syndrome and a purely structural abnormality seems an unlikely explanation for the excessive response to cold. Much attention has therefore turned to the functional aspects of vasoconstriction, with particular emphasis on sympathetic neurotransmission.

a-Adrenergic receptors sensitive to catecholamines in the digital vessel walls induce vasoconstriction in response to sympathetic stimulation. Sympathetic stimulation also induces the release of histamine, which is a vasodilator, from adjacent mast cells. Histamine H_2-receptors on the sympathetic nerve terminal in turn inhibit a-adrenergic neurotransmission. There is some evidence that this negative feedback control mechanism may be impaired in patients with Raynaud's syndrome. Under normal circumstances the digital vessels dilate rapidly when a cold stimulus sufficient to induce vasoconstriction is withdrawn. In patients with Raynaud's syndrome this rapid postsympathetic vasodilation is delayed or absent and vasodilation occurs only when a warm stimulus is applied. While such an observation may be explained by excessive catecholamine release, by impaired catecholamine inactivation, by histamine depletion, or by mast cell dysfunction, none of these proposals has been clearly shown to underlie a Raynaud's attack. Catecholamine concentrations are certainly higher in the venous blood coming from the hand during a Raynaud's attack, but this phenomenon may be simply the result of a diminished blood flow.

A more recently discovered neurotransmitter, calcitonin-gene-related peptide (**CGRP**) has been found in abundance in perivascular sensory nerve fibers. CGRP is a very potent vasodilator, acting both directly on the vascular endothelium to induce nitric oxide release and on the mast cells to release histamine. As such, CGRP is an important part of the negative feedback reflex vasodilator mechanism that accompanies vasoconstriction. Immunohistochemical studies have revealed a deficiency of CGRP-containing nerve fibers in patients with both primary and secondary Raynaud's syndrome, and the resulting impairment of postsympathetic vasodilatation may be a significant factor in the aetiology of the syndrome.

The issue is complicated further by the observation that sympathetic stimulation normally has a similar effect on blood flow in the hand whether the hand is warm or cold. In patients with Raynaud's syndrome the vasoconstrictor effect of sympathetic stimulation is significantly enhanced when the hand is cold compared with when it is warm. Cold itself may therefore sensitize or enhance local a-adrenergic receptor function in these patients. While these data confirm the conclusions of Lewis that a local phenomenon underlies the onset of a Raynaud's attack, further work is needed to establish the exact mechanism.

Several haemorrheological phenomena have also been noted in the blood of these patients. Both blood and plasma viscosity are increased and there is reduced red cell deformability, abnormal platelet adhesiveness, and reduced activity of the fibrinolytic system. It is difficult to interpret the relevance of these phenomena to the pathogenesis of Raynaud's syndrome.

At normal temperatures there is no significant difference in blood flow through the hands of normal individuals compared with Raynaud's sufferers, despite the increased viscosity in the latter group. As the hand is cooled the blood flow decreases more rapidly in Raynaud's syndrome than in normal controls and may cease altogether at 17 to 22°C, whereas there is still some flow in normal hands at these temperatures. The relative contribution of increased blood viscosity in Raynaud's patients to this cold sensitivity remains a matter for debate and it is not known whether viscosity is crucial to the pathogenesis of the syndrome or whether it is simply a secondary event.

Associated disorders

A wide variety of other conditions may occur in association with Raynaud's syndrome (Table 1) and, in the case of the connective tissue disorders, the Raynaud's attacks may precede other features of the disease by several years. No single common thread has been identified among these disorders which may help to explain the pathogenesis of Raynaud's syndrome and not all of the patients with these disorders go on to develop Raynaud's syndrome. Patients who present with apparent primary Raynaud's syndrome may also exhibit abnormalities on immunological screening of their serum, although the significance of such findings is often uncertain.

1. Connective tissue diseases	
	Scleroderma
	Systemic lupus erythematosus
	Dermatomyositis
	Rheumatoid arthritis
	Vasculitis
2. Occlusive arterial disease	
	Atherosclerosis
	Thromboangiitis obliterans
	Embolism
3. Trauma	
	Vibration injury
	Cold injury
4. Neurovascular lesions	
	Thoracic outlet syndrome
	Carpal tunnel syndrome
5. Miscellaneous	
	Ergotamine intoxication
	Cryoglobulinaemia
	Cold agglutinins

Table 1 Disorders associated with Raynaud's syndrome

Connective tissue disease

Secondary Raynaud's syndrome most commonly occurs in association with the connective tissue diseases. Between 40 and 80 per cent of patients with scleroderma will develop Raynaud's syndrome and the syndrome is often a presenting feature of scleroderma, although it may take several years after the onset of Raynaud's attacks before a definitive diagnosis is made. The incidence of Raynaud's syndrome is reported variously as 35 per cent in systemic lupus erythematosus, 22 per cent in dermatomyositis, and 11 per cent in rheumatoid arthritis. Many of these patients exhibit evidence of structural disease in the digital arteries on arteriography, with irregularity or obstruction of the lumen due to intimal proliferation. Such arteriographic changes are common in the connective tissue diseases and do not correlate with the presence or absence of Raynaud's attacks. Indeed many normal people over the age of 50 years have similar arteriographic findings and it is therefore difficult to implicate organic arterial obstruction on its own as a cause of the syndrome in patients with connective tissue disease.

Occlusive arterial disease

Raynaud's syndrome is seen in patients with arteriosclerotic disease in the limbs, particularly when the arteriosclerotic obstruction is sufficiently severe to reduce the digital blood pressure. When the digital blood pressure is low the normal vasospastic response to cold can more easily close off the vessel lumen against the reduced intravascular blood pressure. Thus although the sympathetic and local responses to cold are entirely normal and appropriate, the effects of cold-induced vasospasm are exaggerated and a Raynaud's attack ensues.

Thromboangiitis obliterans (Buerger's disease) is particularly associated with Raynaud's syndrome because, in addition to obstruction of the medium-sized arteries of the forearm and calf, the digital vessels are commonly involved (Fig. 3). Such segmental involvement in the inflammatory thrombotic process results in a similar exaggeration of the normal vasospastic response to cold described above for arteriosclerosis because of the consequent reduction in digital blood pressure. Raynaud's

attacks are the presenting feature of thromboangiitis obliterans in 12 to 25 per cent of patients and the disease itself may account for up to 12 per cent of the cases presenting to specialist units with Raynaud's syndrome. Thromboangiitis obliterans is particularly associated with smoking and remissions are often induced by abstinence from tobacco. Under these circumstances abnormalities of the digital pulses in the disease have sometimes been noted to disappear when the patient stops smoking, only to recur if the patient starts smoking again.



Fig. 3. Arteriogram of the hand of a male smoker with thromboangiitis obliterans. The proximal digital vessel on the radial aspect of the index finger is severely diseased. There is spasm of vessels in the third and fourth fingers and distal obliteration of vessels in the fifth finger.

Trauma

Raynaud's syndrome is the classic occupational hazard of people who use vibrating tools such as chain saws or pneumatic drills in their job. The attacks do not occur while the tool is being used, but later on when the hand is cooled. There may be a latent interval of several years between the regular use of a vibrating tool and the onset of the syndrome. The prolonged exposure to vibration may damage the endothelium lining the digital arteries and the subsequent subintimal fibrosis leads to widespread palmar and digital artery obstruction (Fig. 4). The suggestion that the vibration may have a direct effect on sympathetic nerve endings has not gained widespread support.

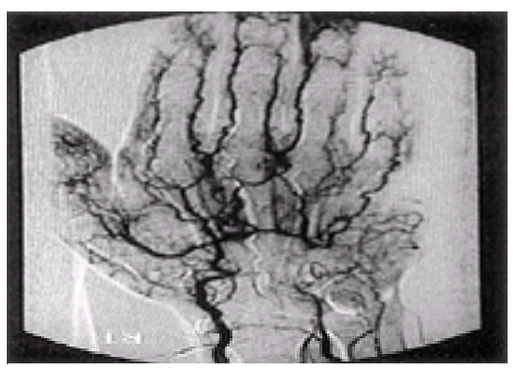


Fig. 4. Arteriogram of the hand of a man with a history of occupational exposure to high speed drills—the ulnar artery in the palm is obliterated and there is irregularity and occlusion of the digital arteries.

Cold injury may also induce Raynaud's syndrome and has also been noted as an occupational hazard, particularly in the frozen food processing industry. Chronic continued exposure to cold is less important than rapid alternate cooling and warming of the hands such as occurs when workers handling frozen foods plunge their hands into warm water every few minutes to keep them warm. Alterations in working practices to avoid this alternate cooling and warming have substantially decreased the incidence of this particular occupational hazard.

Raynaud's syndrome may also be a long-term sequel in digits affected by frostbite.

Neurovascular lesions

Thoracic outlet syndrome may present with Raynaud's syndrome. It is unusual for the neurological and vascular symptoms of thoracic outlet obstruction to coexist in the same patient and it is usually the vascular variety with subclavian artery stenosis, post-stenotic dilatation, and microembolization of the digital vessels which results in Raynaud's syndrome. Raynaud's syndrome may occur in patients with a coincident asymptomatic cervical rib and so it is important to demonstrate a definite vascular abnormality by angiography before attributing the Raynaud's syndrome to thoracic outlet obstruction.

Although carpal tunnel syndrome may be present in up to 15 per cent of patients with Raynaud's syndrome, nerve compression does not appear to contribute to the vasospasm, since division of the flexor retinaculum does not relieve the Raynaud's syndrome. A collagen disorder may underlie both conditions, so that they occur in association with each other without any specific cause and effect relationship.

Miscellaneous

Ergotamine is a potent vasoconstrictor used in the treatment of migraine. Prolonged use, especially in excessive dosages, may result in peripheral vasoconstriction manifesting as Raynaud's syndrome.

The presence of cold agglutinins or cryoglobulins in the blood produces hyperviscosity of the blood upon exposure to cold; blood flow in the digital arteries therefore virtually ceases at low temperatures. Cold agglutinins adsorb on to red blood cells at temperatures between 24 and 34°C and produce a reversible haemagglutination. The digital vessels are then plugged by masses of agglutinated red cells. Subsequent haemolysis of the agglutinated cells may produce haemoglobinuria and haemolytic anaemia.

Cryoglobulins in the blood obstruct the digital vessels by undergoing cold precipitation within them. The cryoglobulins may occur idiopathically but can also be secondary to a malignant reticulosis.

Changes in blood viscosity appear to be the primary mechanism for the obstruction to blood flow in these syndromes rather than any excessive response to sympathetic activity: the episodic digital ischaemia usually disappears if the cold agglutinins disappear.

Diagnosis

The diagnosis of Raynaud's syndrome depends mainly upon the history of colour changes induced by cold exposure or occasionally by emotion. The involved extremity turns pale and numb when cold and makes a slow recovery when warmed, taking 15 to 45 min to pass through the stages of cyanosis and redness back to a normal colour. Pain in the digits is not a usual feature but may occur, particularly in the secondary varieties where there is digital vessel occlusion and a significant amount of ischaemic tissue damage, leading to ulceration and gangrene. This ischaemic damage may become so severe that the patient may require digital amputation despite the presence of a normal radial pulse (Fig. 5).



Fig. 5. Hands of a patient with scleroderma. The right index fingertip has been amputated for ischaemia and a Raynaud's attack is in progress in the left hand.

Such ischaemic tissue damage is extremely rare in the primary form of Raynaud's syndrome where the attacks are solely related to vasospasm and there is no obstructive element. Primary Raynaud's syndrome tends to occur in young women under 30 years and is usually symmetrical, involving either the hands or the feet or both together. No associated disease is present and symptoms continue for 2 or more years without the development of any evident aetiology. Pain is unusual except when bacterial or fungal paronychia complicates the attacks.

Raynaud's attacks may be the presenting feature of a number of associated diseases, evidence for which should also be sought for in the history and examination. Patients with connective tissue disorders such as rheumatoid disease or systemic lupus erythematosus may complain of arthralgia and have stiff, painful, and swollen joints. Patients with scleroderma may complain of dysphagia or diarrhoea and on examination may have a pinched face with a small tight mouth and inelastic skin. Telangiectasia may be present on the lips and hands and there may be missing digits where severe digital ischaemia has necessitated amputation.

Peripheral pulses should be felt for evidence of arteriosclerotic disease, and a history of heavy smoking may suggest the presence of thromboangiitis obliterans in a young male with Raynaud's symptoms. The presence of a subclavian bruit with Raynaud's syndrome in the ipsilateral hand suggests thoracic outlet obstruction, and there may also be differences in blood pressure between the two arms.

A careful history of drug use should be elicited, particularly if the patient takes migraine preparations containing ergotamine. The patient's occupational history is also important as they may have changed their job and only confess to cold exposure or the use of vibrating tools in a previous occupation on direct questioning.

The hands or feet themselves may look completely normal on examination, particularly in the warmth of the consulting room. The history of colour changes is usually sufficient to make the diagnosis, although the extremity can always be immersed in ice cold water for 30 s to induce an attack if confirmation is necessary. Special investigations are mainly directed at the detection of associated disorders: digital pressure and plethysmographic blood flow measurements are more useful as a research tool than for routine clinical use. Arteriography is not necessary in every patient with a clear history, although it can help to distinguish between purely vasospastic Raynaud's syndrome and that with an obstructive element in the digital vessels (Fig. 2 and Fig. 3). Arch arteriography is particularly useful when a more proximal arterial obstructive element is suspected, such as subclavian stenosis with thromboembolic obstruction of the digital vessels, as the proximal obstruction may be amenable to surgical correction.

Comprehensive investigation of the patient with Raynaud's syndrome to rule out associated disorders should include radiography of the thoracic outlet to show the presence of a cervical rib; radiographs of the hands may show the joint disruption characteristic of rheumatoid arthritis or the subcutaneous calcinosis and sclerodactyly of scleroderma. Blood tests include a full blood count and an erythrocyte sedimentation rate, which may be elevated in connective tissue disorders. More specific screening tests include rheumatoid factor, antinuclear factor, anti-DNA antibody, serum protein electrophoresis, cryoglobulins, and cold agglutinin assays.

Differential diagnosis

Acrocyanosis

Acrocyanosis is a separate vasospastic disorder of the hands or feet which occurs mainly in women and may be unilateral. The extremity feels cool, may be slightly oedematous, and has a constant blue discoloration which does not resolve in response to warmth (Fig. 6). Peripheral pulses are normal and ischaemic changes do not occur. The condition is due to cutaneous arteriolar vasospasm which is independent of temperature. The condition may respond to vasodilatory drugs used to treat Raynaud's syndrome and sympathectomy can produce good results if symptoms are severe.



Fig. 6. Unilateral acrocyanosis in the left foot of an elderly woman. The reddish-blue discoloration was constant but the condition was otherwise asymptomatic and the foot pulses were intact.

Livido reticularis

Livido reticularis is also caused by cutaneous arteriolar vasospasm with secondary dilatation of associated capillaries and venules but the distribution is patchy and apparently random. There is patchy reddish-blue mottling of the lower legs and feet or occasionally of the arms and hands which persists irrespective of temperature, although it is often worse in the cold. The majority of cases occur in isolation but the disorder has been described in association with polyarteritis nodosa and systemic lupus erythematosus. No treatment is usually necessary other than reassurance and avoidance of cold.

Treatment of Raynaud's syndrome

General measures

No curative treatment is available for Raynaud's syndrome and the aim is therefore to palliate the symptoms by reducing the frequency and the severity of the attacks. Reassurance that the outlook is generally benign is important, especially in primary Raynaud's syndrome where progression to digital ischaemia and gangrene is extremely rare. The natural history of the syndrome includes long periods of remission, especially over the summer months. The majority of patients can achieve a worthwhile response through the simple avoidance of cold and tobacco smoking. Smoking a cigarette may produce a fall in temperature of 2 to 3°C in the fingertips (Fig. 2) and smoking is absolutely contraindicated in patients with thromboangiitis obliterans, who should be advised that they run a very high risk of amputation if they continue to smoke. Techniques to keep the extremity warm range from advice to wear thick woollen gloves or socks in cold weather to the use of chemically activated handwarmers or electrically heated gloves. Perhaps less practical might be the advice to move to a warmer climate.

While the majority of patients can be managed satisfactorily with these conservative measures a few continue to have frequent or severe attacks or progress to digital

ischaemia. For these patients the choice lies between drug therapy or sympathectomy.

Drug therapy

A variety of vasodilatory drugs with different modes of action are available and many have been advocated in the treatment of Raynaud's syndrome. Unfortunately the syndrome itself may remit spontaneously and there is also a strong placebo effect on the frequency and severity of attacks, so trials of drug therapy have to be very carefully controlled and the results interpreted with caution. A significant response to the drug is seen in many trials but only in 30 to 60 per cent of the patients. In many cases the attacks continue and the drug response is only apparent as a reduction in the frequency or severity of the attacks without complete amelioration of the syndrome. Side-effects are common with these drugs and can be unpleasant. Some therapies are only effective when administered by the intravenous route, which requires hospital admission, and these have to be reserved for the most severely affected patients.

Despite these reservations a trial of the different drug therapies to see if one will induce a response is often worthwhile, because the only other alternative is surgery, the results of which are by no means guaranteed. As the outlook of the syndrome is generally benign, there is plenty of time to give drug therapy a reasonable trial. The overall clinical impression is that marked improvements in symptoms are difficult to detect unless a strict diary of the frequency and severity of attacks is kept to allow comparisons before and after drug therapy to be made.

The best studied oral agents are nifedipine and thymoxamine and effective intravenous agents are iloprost and ketanserin. The use of topical vasodilator creams such as glyceryl trinitrate or hexyl nicotinate (both in concentrations of 1 to 2 per cent) may also be of value in refractory cases.

Nifedipine

Nifedipine is a calcium channel blocking agent that interferes with the inward displacement of calcium ions through the slow channels of active cell membranes. In vascular smooth muscle cells the effect of this interference is a reduction in vascular tone with subsequent vasodilatation. Nifedipine is often used for the treatment of angina and hypertension and is the first choice of many clinicians for the treatment of Raynaud's syndrome. Side-effects are common and include headache, flushing, peripheral oedema, eye pain, and blurred vision. Diltiazem is an alternative calcium-channel blocker.

Thymoxamine

Thymoxamine acts by competitive antagonism of a-adrenoreceptors and blocks vasoconstriction in the skin. There is a low incidence of side-effects with thymoxamine but only 25 to 30 per cent of patients may notice any subjective improvement in symptoms.

Iloprost

The prostanoids prostaglandin E_1 and prostacyclin are potent vasodilators and inhibit platelet aggregation; they are effective in severe Raynaud's syndrome but their rapid metabolism may limit their therapeutic potential. Iloprost is a more stable prostacyclin analog, with a half-life 10 times longer than that of prostacyclin. Iloprost is administered as a 6-h intravenous infusion at a dose of 2 ng/kg.min on three consecutive days. The dose is reduced in 0.5 ng/kg.min increments if side-effects, which may include severe headache, dizziness, vomiting, and diarrhoea, occur. Studies in patients with severe secondary Raynaud's syndrome due to scleroderma report a 30 per cent reduction in the frequency of attacks and a 20 per cent reduction in their severity. Digital ischaemic lesions can heal after iloprost infusion but healing may not necessarily be significantly better than after placebo infusions. The severity of side-effects confines the use of intravenous preparations to hospital inpatients. Although the effects of iloprost on vasodilatation and platelet adhesion disappear almost as soon as the infusion is stopped, the clinical improvement may last for up to 6 weeks and repeated infusions may be given. The results of trials with oral prostanoid preparations are awaited.

Ketanserin

Ketanserin is a serotonin receptor antagonist that antagonizes serotonin-induced vasoconstriction and platelet aggregation and inhibits the amplification by serotonin of vasoconstriction and platelet aggregation by other agents. Intravenous ketanserin improves finger blood flow and relieves ischaemic symptoms in severe secondary Raynaud's syndrome, but the effect is short lived and wears off after the infusion is stopped. Oral ketanserin is also available and although it can improve finger blood flow, many studies have shown no subjective clinical improvement in Raynaud's symptoms over placebo.

Chemical sympathectomy

Reserpine is a rauwolfia alkaloid which depletes norepinephrine from sympathetic nerve terminals. Guanethidine has a similar effect and also prevents norepinephrine release from postganglionic neurons. Reserpine has been used orally to treat Raynaud's syndrome but the high incidence of side-effects associated with long-term use has made it an unpopular choice. Both reserpine and guanethidine may be administered intravenously into an affected extremity distal to a blood pressure cuff which is kept inflated above systolic pressure for 20 to 30 min during and after the injection. This so-called chemical sympathectomy does not produce lasting relief and is not effective in patients with advanced digital vessel obstruction, but may give short-term relief of severe symptoms in patients with predominantly vasospastic disease.

Surgical therapy

The mainstay of surgical treatment in Raynaud's syndrome is sympathectomy, although surgery may also be required when the syndrome is secondary to proximal vascular disease. In the latter case it may be necessary to resect a cervical rib or band to relieve thoracic outlet compression, although first rib resection has also been advocated. Occasionally the subclavian stenosis and poststenotic dilatation are so severe that local excision and vein graft replacement of the artery are required. The transaxillary approach to first rib resection is best avoided in this latter group of patients where the supraclavicular approach gives better proximal control of the artery.

Atherosclerotic disease in the brachial artery is rare but it is possible to bypass severely stenosed or occluded segments with a vein graft if the symptoms warrant it. The majority of patients with severely ischaemic digits due to Raynaud's syndrome have intact distal pulses and tolerate local amputation to control severe ischaemic pain or gangrene well. In many instances the middle or proximal phalanges can be preserved and primary closure of the amputation stump gives good healing. In severe cases digits may require amputation over a period of time as the disease progresses, particularly in patients with scleroderma. The more that can be preserved of each digit therefore, the better the long-term functional result.

Sympathectomy

The results of sympathectomy for Raynaud's syndrome are variable and there is considerable doubt over its long-term efficacy. Subjective clinical improvement is noted in 60 to 70 per cent of patients immediately after sympathectomy, but 10 years later only 30 to 40 per cent of patients are still improved, which is not a statistically significant result compared with non-operated patients.

Good results from sympathectomy may be expected in patients with primary Raynaud's syndrome and in those patients whose secondary Raynaud's syndrome is due to embolic, traumatic, or atherosclerotic obstruction. Good results may also be obtained in patients with thromboangiitis obliterans, but results have been uniformly poor in patients with connective tissue disorders. With the exception of the connective tissue disorders, where sympathectomy is usually best avoided, it is generally agreed that sympathectomy should be reserved for patients with recurrent ulceration of the fingertips or those who are severely incapacitated by vasospastic phenomena in spite of adequate medical management.

Lumbar sympathectomy may be performed either operatively or by percutaneous injection of phenol solution. An operative lumbar sympathectomy is performed through a small transverse muscle-splitting incision just above the level of the umbilicus and lateral to the rectus muscle. The sympathetic chain is approached retroperitoneally. Percutaneous sympathetic blockade is much more difficult in the cervical region and hand symptoms therefore require operative cervical sympathectomy. This operation is technically much easier through the transaxillary route than via the supraclavicular approach. However, the thoracoscopic transaxillary technique is now the procedure of choice (see Chapter 17.8). Both approaches carry the risk of damage to the stellate ganglion resulting in a Horner's syndrome with ptosis, pupillary constriction, and facial flushing and dryness on the affected side. Compensatory sweating may also occur in the trunk or other limbs in a large proportion of patients. Resection of the second, third, and fourth thoracic ganglia appears in practice to produce effective sympathetic denervation of the arm and axilla without risking damage to the stellate ganglion. Although abolished initially after sympathectomy, the vasomotor and sudomotor reflexes may return to the hand a year or more after surgery. Whether this return represents regeneration of nerve fibers or the gradual assumption of a greater role of residual sympathetic pathways is not known, but the phenomenon may account for the variable long-term results of sympathectomy in Raynaud's syndrome.

Summary of treatment

Most patients with Raynaud's syndrome can be managed with advice to stop smoking and to keep the extremity warm. Failure of these simple measures in patients without digital ischaemia is an indication for oral drug therapy with nifedipine, diltiazem, or thymoxamine. The onset of digital ischaemia, especially in the connective tissue disorders, justifies a trial of intravenous iloprost therapy. Continued severe symptoms or ischaemia despite full medical treatment warrants consideration for sympathectomy. Local digital amputation may be the eventual outcome of severe ischaemic damage.

Further reading

Cooke ED, Nicolaides AN. Raynaud's syndrome. *British Medical Journal* 1990; **300**: 553–5. [An overview of the aetiology and treatment.]

Dowd PM *et al.* Raynaud's phenomenon. *Lancet* 1995; **346**: 283–90. [An overview of the etiology and treatment with detailed descriptions of local neurotransmitter mechanisms.]

Lafferty K, Roberts VC, De Trafford JC, Cotton LT. On the nature of Raynaud's phenomenon: the role of histamine. *Lancet* 1983; **ii**: 313–14. [Studies on the aetiology of Raynaud's syndrome.]

Lewis T. Experiments relating to the peripheral mechanisms involved in spasmodic arrest of the circulation in the fingers. A variety of Raynaud's disease. *Heart* 1929; **15**: 7–101. [Classic studies identifying local neurotransmitter faults rather than defects in sympathetic nerve activity as the cause of Raynaud's syndrome.]

Porter JM, Edwards JM. Occlusive and vasospastic diseases involving distal upper extremity arteries — Raynaud's syndrome. In: Rutherford RB, ed., *Vascular surgery*, 4th edn., pp. 961–76. WB Saunders, Philadelphia, 1995.

Raynaud AGM. Nouvelles recherches sur la nature et la traitement de lásphyxie locale des extrémités. *Archives of General Medicine* 1874; **1**: 5. [Raynaud's original description of the syndrome.]

Strandness DE, Sumner DS. Raynaud's disease and Raynaud's phenomenon. In: Strandness DE, Sumner DS, eds, *Haemodynamics for surgeons*, pp. 543–81. Grune & Stratton, New York, 1975.

17.8.3 Sympathectomy in the upper limb

Jack Collin

Hyperhidrosis
Management of essential hyperhidrosis
The technique of sympathectomy
Chapter References

Sympathectomy has been recommended for the treatment of several diverse disorders of the upper limb (Table 1). The theoretical justification for its possible efficacy in most cases is tenuous and objective evidence of benefit is non-existent. The principal use of sympathectomy at present is to provide a permanent cure for patients with disabling palmar hyperhidrosis and those with both palmar and axillary hyperhidrosis. It is now rarely indicated for the treatment of axillary hyperhidrosis alone. Its only other empirical residual role is in the treatment of patients with severe Raynaud's phenomenon with impending or actual digital gangrene, particularly in those with Buerger's disease (thromboangiitis obliterans).

(a) Neurological
Hyperhidrosis
Causalgia
Sympathetic mediated pain syndrome
Brachial neuralgia
Phantom limb pain.
(b) Vascular
Raynaud's syndrome
Thromboangiitis obliterans
Vasculitis
Atherosclerotic occlusive arterial disease
Erythromelagia
Acrocyanosis.

Table 1 Conditions for which sympathectomy of the upper limb has been advocated

Hyperhidrosis

Excessive, unphysiological sweating may be generalized and affect the whole or most of the body but commonly is confined to the hands, axillae, face, or feet. When sweaty feet are the problem patients complain either of the excessive wetness or malodour. Perhaps surprisingly with excessive sweating elsewhere, including axillary hyperhidrosis, it is the volume of sweat not the malodour that causes distress. The explanation seems to be that large volumes of sweat from the thermoregulatory eccrine glands wash away the malodorous secretions of the apocrine glands which have a secondary sexual function and are located predominantly in the axillae and perineum.

The functions of sweating are thermoregulation, to moisturize the skin, and in the hands to increase the friction of palmar skin to facilitate grip. Thermoregulatory sweating is not uniformly distributed over the body surface and under normal circumstances heat loss occurs disproportionately from the hands and face. Heat loss by sweating occurs by virtue of the latent heat of evaporation which is for water 41.1 kJ/mmol (540 kcal/l). It follows that sweating is unphysiological when sweat drips from the body instead of being evaporated since it results in the wasteful loss of fluid and electrolytes without the compensatory benefit of heat loss. The maximum possible sweat loss is 10 litres per day. If all of this volume was evaporated the heat loss achieved would be equivalent to a food intake of 5400 kcal; a daily level of energy expenditure that few individuals are ever likely to experience in a lifetime.

Hyperhidrosis may be idiopathic or secondary to a large number of disorders (Table 2). In almost all patients who present with the sole symptom of excessive sweating the disorder is primary or idiopathic (Table 3). The sex incidence in those who seek medical attention is 2:1 female to male. The onset is typically in childhood or adolescence and sufferers are embarrassed by their wet handshake and wet, salt-stained clothing or shoes. In more traditional schools they have usually been admonished for their smudged and damp work exercise books. Sweat dripping into a computer keyboard and the revulsion of clients and colleagues offered a wet handshake constitute a significant employment disadvantage.

Focal	*Essential (idiopathic, primary)*	
Spinal cord	Injury	
	Syringomyelia	
Gustatory*	Herpes zoster	
	Chorda tympani injury	
	Auriculotemporal (Fey's) syndrome	
Peripheral neuropathy		
Intrathoracic tumour		
Strokes	Cortical (unilateral)	
	Brainstem (symmetrical)	
Generalized	Endocrine	Menopause
	Thyrotoxicosis	
	Phaeochromocytoma	
	Carcinoid syndrome	
	Hypoglycaemia	
Infectious		
Lymphoma		
Primamental angina		
Drugs		

*Physiological gustatory sweating is symmetrical, angby sialogkcall gustatory sweating is usually unilateral

Table 2 Causes of hyperhidrosis

Incidence: female:male, 2:1
Onset in adolescence
High volume sweating, not malodorous
Commonly affects hands, axillae, feet, face
Occasionally generalized sweating
May be constant or intermittent
Provoked by anxiety

Table 3 Features of essential (idiopathic) hyperhidrosis

Management of essential hyperhidrosis

A large number of neurological, endocrine, and other conditions may present with local or generalized hyperhidrosis (Table 2) and these should first be excluded by appropriate investigations. In practice, however, almost all patients presenting with hyperhidrosis to surgeons are suffering from the idiopathic condition. The hands

and/or axillae are most commonly affected and less often the face. Pedal hyperhidrosis sometimes occurs in association with palmar and axillary hyperhidrosis but may be the only, or the major, disability.

For patients with generalized hyperhidrosis, conservative therapies are all that can be offered (Table 4). The simplest and safest of these is to produce waterlogging of the skin and consequently temporary blockage of the sweat ducts by soaking in a bath for 45 min. This ploy can allow the patient a sweat-free special evening out but is impractical for routine daily use. Many patients will already have been prescribed anxiolytic drugs but the value of these is doubtful and there is a risk of inducing dependency. Anticholinergics such as propantheline bromide are sometimes effective but in therapeutic dosage cause an unpleasantly dry mouth.

Generalized hyperhidrosis
Soaking to waterlog the skin
Anticholinergics—propantheline bromide
Anxiolytic drugs (beware dependency)
Localized hyperhidrosis
Antiperspirants—Aluminum chloride hexahydrate
Iontophoresis
Palmar hyperhidrosis
Sympathetic trunkotomy on second rib
Palmar and axillary hyperhidrosis
T2 and T3 sympathetic ganglionectomy
Axillary hyperhidrosis
Intradermal botulinum toxin injection

Table 4 Therapy for essential hyperhidrosis

Topical treatment with antiperspirants such as aluminum chloride hexahydrate can help to control localized hyperhidrosis but skin sensitization is liable to limit its use in most patients. Iontophoresis in which fluid is drawn into the skin by the passage of a small electric current has an application in pedal hyperhidrosis and in those patients with palmar hyperhidrosis who decline operative sympathectomy. Water iontophoresis inhibits sweating only for as long as the skin remains waterlogged. Iontophoresis with an anticholingeric solution can reduce sweating for a few days, but systemic side-effects are similar to those produced by the much simpler, less expensive, and equally effective oral administration of anticholinergics.

For patients who have isolated axillary hyperhidrosis or those in whom sympathectomy for combined palmar and axillary hyperhidrosis has failed to cure the axillary component, the local injection of botulinum toxin is an effective and durable treatment. Botulinum toxin prevents the release of acetylcholine from all nerve endings with which it comes in contact and this is the mechanism of its lethal systemic effect. By intradermal or subcutaneous injection of minuscule doses of toxin the effect can be confined to the innervation of the sweat glands. Axillary sweating is abolished for approximately 9 months by the intradermal injection of 240 mouse units of Dysport® per axilla. The injection is given in 12 separate sites intradermally at 0.1 ml per site to provide complete area coverage. Dysport is not currently licensed for this indication in the United Kingdom and its use for treating hyperhidrosis is on a named patient basis at the risk of the prescribing physician.

Palmar hyperhidrosis can be cured in all cases by sympathetic trunkotomy between the first and second thoracic ganglia. Recurrent hyperhidrosis occurs in up to 5 per cent of patients after a time lapse of 6 to 9 months and is probably caused by the regrowth of divided preganglionic sympathetic nerve fibres. Palmar anhidrosis is the usual consequence of the operation but around a quarter of patients have some residual sweating, an outcome which cannot be deliberately secured but which is preferable to anhidrosis. Residual sweating may be a consequence either of preganglionic outflow to the hand from the T1 spinal cord level or postganglionic outflow from the T2 ganglion passing to the T1 somatic nerve via the nerve of Kuntz. Whatever the explanation for residual palmar sweating, it is surprising that no patient with it in the author's experience has suffered from persistent hyperhidrosis after technically successful sympathetic trunkotomy. Most patients have some reduction in facial sweating after sympathetic trunkotomy below the first thoracic sympathetic ganglion and some have facial anhidrosis. This single sign of Horner's syndrome may be present in the absence of all the eye signs of the syndrome which occur only if the T1 ganglion is damaged. It represents a common anatomical variation in which the facial sweat glands are innervated exclusively by preganglionic fibres leaving the spinal cord caudal to the T1 spinal outflow.

In patients with both axillary and palmar hyperhidrosis the T2 and T3 ganglia must be destroyed to cure excessive axillary sweating. The operation is technically much more difficult and it is probable that the 25 to 50 per cent failure rate reported is largely attributable to failure to secure complete destruction of the ganglia. Following bilateral successful T2 and T3 ganglionectomy compensatory hyperhidrosis is more frequent and troublesome than after bilateral sympathetic trunkotomy. Because of both the high failure rate and the occurrence of compensatory hyperhidrosis sympathectomy is no longer the treatment of first choice for patients with axillary hyperhidrosis unless they also have palmar hyperhidrosis.

The technique of sympathectomy (Fig. 1)

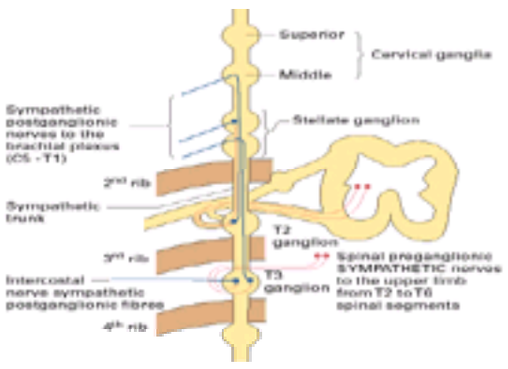


Fig. 1. Schematic representation of the sympathetic nerve supply to the upper limb and axilla. It can be seen that interruption of all the sympathetic nerve supply to the upper limb can be achieved by division of the sympathetic trunk between the first and second thoracic sympathetic ganglia. Sympathectomy of the axilla requires the ablation of the second and third thoracic sympathetic ganglia.

The traditional open surgical operations of transcervical, transthoracic, and paravertebral upper thoracic sympathectomy have now been replaced by thoracoscopic sympathectomy and are virtually obsolete. The sole remaining indication for transcervical upper thoracic sympathectomy is in patients with extensive pleural adhesions preventing collapse of the lung. Such an event is virtually unknown in the young patients requiring sympathectomy for hyperhidrosis and is uncommon in the few patients with impending digital gangrene in whom the operation may confer benefit. The young surgeon unusually called upon to perform a transcervical upper thoracic sympathectomy would be well advised to seek the help of one of the dwindling number of older surgeons who still remember how the operation was done.

Videothoracoscopic sympathectomy has arguably produced the greatest technical improvement of all the ‘key hole' surgical operations introduced in the last few years. It has converted a technically difficult open operation with significant morbidity into a simple, quick procedure leaving a virtually invisible and painless wound. It is the author's practice never to perform the operation bilaterally at the same time. Those who are tempted by the convenience of a single visit to the operating theatre should first consider the consequences liable to ensue if full ventilation of the first collapsed lung is not maintained when the second lung is collapsed. Postoperative pneumothorax is a minor nuisance when it affects one lung in a healthy adult but if bilateral can be fatal. The operation is most conveniently performed with double lumen bronchial intubation so that each lung can be ventilated independently. The patient is placed supine on the operating table with the upper limb abducted to 90 degrees. A Verris needle is inserted into the pleural cavity through the third intercostal space above the fourth rib in the anterior half of the axilla. A measured volume of 1 litre of carbon dioxide is insufflated and simultaneously the anaesthetist withdraws ventilation from the same lung. The Verris needle is withdrawn and replaced by a 10-mm trocar and cannula secure in the knowledge that it will be entering a 1-litre pneumothorax. The trocar is withdrawn and with the cannula open to air, the lung will completely collapse if double lumen intubation has been successfully accomplished. A 10-mm operating telescope composed of a 5-mm telescope and a 5-mm operating channel is introduced. The necks of the second, third, and fourth ribs are visualized by manipulating the upper lobe of the lung downwards with the telescope. The highest rib that can be seen is the second rib. The first rib is covered by a pad of fat and consequently is not visible but can be instrumentally palpated. The lowest rib that can be seen is the fourth. The fourth intercostal space and the fourth thoracic sympathetic ganglion are obscured from the front by the root of the lung and hence inaccessible. Claims by some surgeons that the fourth thoracic sympathetic ganglion is routinely destroyed or ablated should be viewed with incredulity. It cannot be done under direct vision from the front with the patient supine. The sympathetic trunk is identified as a vertically running white cord emerging from the fat pad

covering the first rib and crossing the necks of the second and third ribs. It is essential to identify with certainty the sympathetic trunk before the pleura is incised medial to it on the second rib. The pleural incision is elongated by tearing with forceps. If second and third ganglionectomy is to be performed the sympathetic trunk is exposed caudally as far as the fourth rib by extending the pleural incision. The sympathetic trunk is freed on the second rib and transected with scissors. Division of the sympathetic trunk with electrocautery or the application of electrocautery to the superior end of the divided trunk risks producing the eye signs of Horner's syndrome which is an entirely preventable complication of the procedure. To reduce the chances of reinnervation by regrowth of divided preganglionic nerve fibres, the upper end of the divided trunk is rotated through 180 degrees and buried in the first intercostal space. When the aim is to secure only palmar anhidrosis the lower end of the divided sympathetic trunk is electrocoagulated to inhibit successful nerve regrowth. In patients with axillary hyperhidrosis the lower end of the divided sympathetic trunk is grasped with forceps and by a combination of traction and dissection, the second and third ganglia are mobilized. The sympathetic trunk is then divided or avulsed as it crosses the fourth rib and the two ganglia removed. The operation is completed by reinflating the lung, withdrawing the cannula and telescope, and inserting a single skin suture in the axillary wound.

Chapter References

Bushara KO, Park DM, Jones JC, Schutta HS. Botulinum toxin—a possible new treatment for axillary hyperhidrosis. *Clinical and Experimental Dermatology* 1996; **21**: 276–8. [The first description of the use of Botulinum toxin for producing axillary anhidrosis.]

Drott C, ed. Endoscopic transthoracic sympathicotomy. *European Journal of Surgery* 1998; **164**: Suppl. 580. [The papers and abstracts presented at the second international symposium on thoracoscopic sympathectomy.]

Freeman R, Waldorf HA., Dover JS. Autonomic neurodermatology (part II). Disorders of sweating and flushing. *Seminars in Neurology* 1992; **12**: 394–407. [An excellent review of the causes and physiology of hyperhidrosis.]

17.9 Renovascular disease

Linda Hands and Peter J. Morris

Renovascular hypertension
Pathophysiology
Causes of renal artery stenosis
Renal artery stenosis and renovascular hypertension
Prevalence
Progression of disease
Diagnostic tests for renovascular hypertension
Treatment
Overview of atheromatous renal artery disease
Overview of renovascular disease due to fibromuscular dysplasia
Children with renovascular hypertension
Renal artery trauma
Mechanism of injury
Symptoms and signs
Investigations
Treatment
Renal artery aneurysms
Incidence
Pathology
Consequences of renal artery aneurysms
Treatment
Renal arteriovenous fistulae
Aetiology
Pathology
Symptoms and signs
Investigations
Treatment
Further reading

Renovascular hypertension

Pathophysiology

In a classical experiment in 1934, Goldblatt and his colleagues showed that a clip partially occluding one renal artery of a dog caused hypertension. It has since been demonstrated that reduced renal perfusion stimulates renin release from the affected kidney. This, in turn, increases the conversion of angiotensinogen to angiotensin I. Angiotensin converting enzyme converts angiotensin I to angiotensin II and this sequence stimulates the release of aldosterone from the adrenal cortex (Fig. 1).

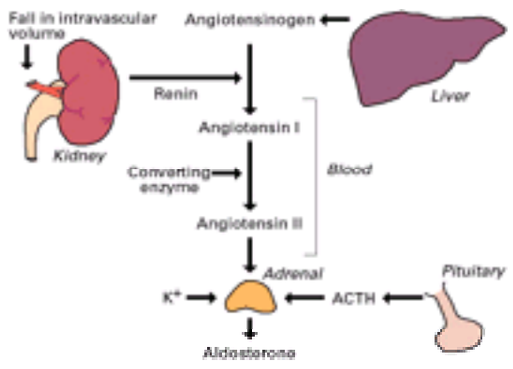


Fig. 1. The renin–angiotensin system.

In fact, it is probably the vasoconstrictor action of angiotensin II which is responsible for the initial stages of hypertension. Salt and water retained under the influence of aldosterone are excreted by the other kidney. However, if there is only one kidney, or if the other kidney has been damaged by prolonged hypertension, the body is no longer able to excrete excess salt and water; hypertension is then due to increased intravascular volume. Thus, in the early stages of this disease, the affected kidney secretes high levels of renin and renin production from the other kidney is suppressed. Later in the disease salt and water retention suppress renin production in both kidneys.

Causes of renal artery stenosis

In humans 70 per cent of renal artery stenoses are caused by atherosclerosis. Patients with this disease are usually elderly, often male, and frequently have other manifestations of atherosclerosis such as ischaemic heart disease and peripheral vascular disease.

The other major cause of renal artery stenosis, probably accounting for 20 per cent of cases, is fibromuscular dysplasia. There are three different forms of the disease, each affecting a particular layer of the arterial wall. The most common is medial dysplasia, characterized by hyperplasia of either the fibrous tissue or smooth muscle of the media (Fig. 2). Up to 70 per cent of cases of medial dysplasia are caused by medial fibroplasia, in which thickened fibromuscular ridges alternate with attenuated media so that the lumen is alternately stenosed and aneurysmal. This produces the so-called ‘string of beads' sign on angiography (Fig. 3). Fibromuscular hyperplasia usually affects the distal two-thirds of the main renal artery and sometimes the primary branches. The disease occurs most often in young women and commonly affects both renal arteries and occasionally the carotid and iliac arteries.

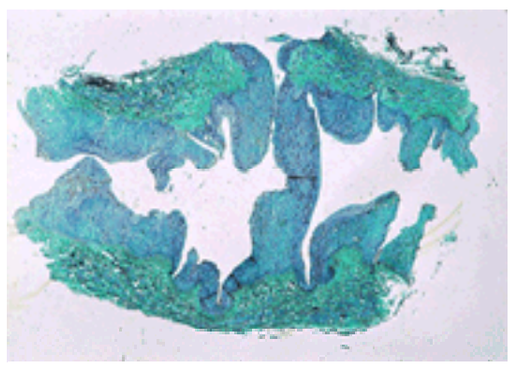


Fig. 2. Histology of the renal artery affected by fibromuscular dysplasia, showing medial proliferation in ridges.



Fig. 3. Angiogram of a renal artery with fibromuscular dysplasia demonstrating the ‘string of beads' sign.

Other, less common causes of renal artery stenosis include trauma, renal artery aneurysm (compressing the adjacent, otherwise normal artery), arteriovenous fistulae (diverting blood from the kidney), neurofibromatosis, Takayasu's disease (an arteritis causing stenosis of arteries arising from the aorta and found mainly in young adults and children of Asian origin), and renal hypoplasia.

Renal artery stenosis and renovascular hypertension

Stenosis of an artery reduces blood flow appreciably only when the luminal diameter is reduced by 70 per cent (Fig. 4). Even demonstration of such a stenosis of the renal artery on duplex scanning or arteriography in a hypertensive patient is not sufficient to warrant a diagnosis of renovascular hypertension. The renal artery disease may be an incidental manifestation of atherosclerosis in a patient with essential hypertension. Many nephrectomies and renal artery reconstructions were performed in patients with hypertension and renal artery stenosis before it was realized that fewer than half benefited from the procedure. The ultimate proof of a renovascular cause of hypertension is its cure by correction of the stenosis, but several tests have been developed that allow patients more likely to respond to such intervention to be selected.

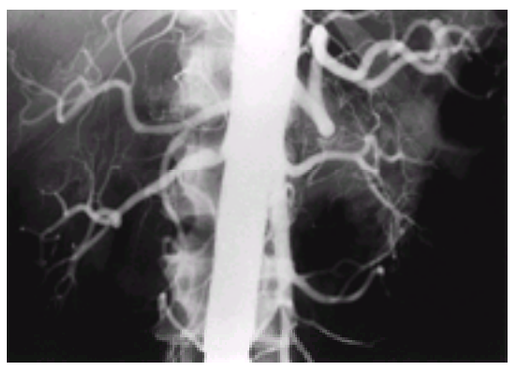


Fig. 4. Angiogram showing a tight stenosis of the right renal artery at the ostium.

Prevalence

In the general hypertensive population the prevalence of renovascular hypertension is only 1 to 5 per cent, but it is one of the few curable forms of hypertension. There are particular groups of individuals who are more likely to have renovascular hypertension and these need to be identified for further investigation (Table 1). In addition, young adults with hypertension, although not necessarily more likely to have a renovascular cause, have a lot to gain by avoiding lifelong treatment with antihypertensive agents if renovascular hypertension is discovered and corrected.

Patients with severe hypertension (diastolic >125 mmHg)
Patients presenting with pulmonary oedema
Patients with generalized atherosclerosis
The elderly patient
The very young patient

Table 1 Hypertensive patients in whom a renovascular cause is more likely to be found

Patients with atherosclerosis often have renal artery stenosis. In patients unselected for hypertension who are undergoing angiography for peripheral vascular disease the prevalence of renal artery stenosis is about 30 per cent; 12 per cent of these have bilateral disease. In hypertensive patients with evidence of peripheral vascular disease the prevalence of renovascular hypertension, as opposed to merely renal artery stenosis, is about 14 per cent.

Patients with severe hypertension are more likely to have renovascular hypertension: the prevalence is 30 per cent in those with a diastolic pressure over 125 mmHg and up to 45 per cent in patients with accelerated hypertension and renal failure.

Age is another important factor. Thirty per cent of patients who develop a diastolic pressure greater than 105 mmHg after the age of 50 years have renovascular hypertension, and it is probably also more common in the elderly with renal failure. Over 70 per cent of older patients with a unilateral atrophic kidney have renal artery stenosis, and nearly 80 per cent of these have significant contralateral stenosis. Many have incipient or overt renal failure as a result, not necessarily accompanied by marked hypertension. At the other end of the scale, renal artery stenosis ranks with coarctation of the aorta as a major cause of hypertension in children.

The presence of a bruit, audible in the upper abdomen or flank, increases the likelihood of renal artery stenosis in any patient.

Progression of disease

This has only been studied in patients receiving medical treatment for their hypertension.

Repeated angiograms taken over 4 to 5 years show increasing stenosis in about 50 per cent of patients with atherosclerotic disease but in only 20 per cent of those with fibromuscular disease. Parallel changes in renal size and serum creatinine can also be detected. When the diameter of the renal artery is reduced by 75 per cent, between 12 and 40 per cent of those with atheromatous stenoses progress to complete occlusion within 1 year. Even at this stage the situation may not be irremedial. Sufficient renal tissue may be maintained by collateral vessels around the capsule and hilum to enable restoration of function if renal artery flow is reinstated (Fig. 5).

One-third of patients with atheromatous disease present with bilateral renal artery stenosis; those with initially unilateral disease have a 40 per cent chance of contralateral renal artery stenosis developing within the next 4 years. Fibromuscular disease is even more likely to affect both renal arteries (60 per cent at presentation) although it is less likely to progress.

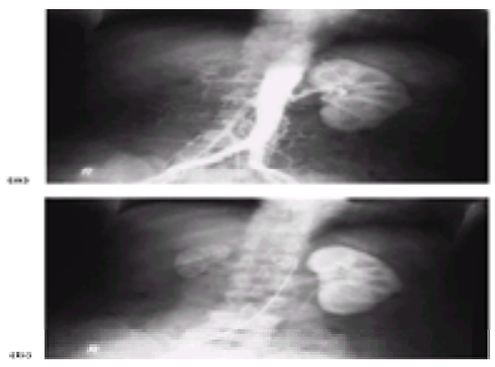


Fig. 5. Early (a) and late phase (b) of an angiogram showing late filling of the right renal vessels and a small right kidney via collaterals because of proximal renal artery occlusion.

While one kidney is functioning normally, unaffected by either stenosis of its artery or hypertensive parenchymal damage, adequate excretion is maintained. However, if that kidney develops functional impairment the patient begins to slip into renal failure. Any kidney affected by renal artery stenosis tends to shrink, with loss of nephrons. This shrinkage can often be reversed by restoring blood supply, but once the kidney measures less than 8 cm in length function is unlikely to return. Thus, intervention can be of benefit to patients with renal failure, provided it occurs before critical renal mass is lost. There is no correlation between renal function and blood pressure, either before or after treatment.

Acute pulmonary oedema occurs in up to 25 per cent of patients who have renal failure due to renal artery stenosis but it is reversed by revascularization of the kidney. Renovascular disease must be excluded in any patients presenting with pulmonary oedema.

Diagnostic tests for renovascular hypertension

Demonstration of renal artery stenosis

Two initial requisites for the diagnosis of renovascular hypertension are hypertension and renal artery stenosis. Confirmation of the latter used to be left to the later stages of investigation because it required arteriography with its attendant risks. Two recent developments have changed that. Duplex ultrasonography is a non-invasive means of detecting renal artery stenosis both by visualization of the vessel and by measurement of the effect of stenosis on blood flow velocity and waveforms. In specialized centres prepared to devote time to each examination this test is both specific and sensitive. However, many examinations are difficult to perform and interpret because of obesity or bowel gas: up to 50 per cent of examinations may have to be abandoned because of technical difficulties. The test also relies heavily on operator expertise. It is therefore unlikely to be useful as a screening test in its present form.

The other major advance is intravenous digital subtraction angiography. This technique requires injection of contrast medium into a vein rather than an artery, which considerably reduces associated morbidity. It relies on computer techniques to demonstrate the arteries when relatively low concentrations of contrast are present. There are, however, several drawbacks to the method. A large amount of contrast has to be injected, which may further compromise renal function in patients with incipient or overt renal failure. Visualization of the vessels depends on cardiac output: output may be impaired in patients with ischaemic heart disease, and even in those with good cardiac output visualization of the renal arteries is often inadequate. Despite these reservations, intravenous digital subtraction angiography is becoming a useful screening tool for renal artery stenosis in patients who have a high probability of renovascular hypertension on clinical grounds.

Renin levels

Once the diagnosis of renal artery stenosis in association with hypertension has been established, further investigations are required. The peripheral blood renin level is often elevated, but this is not always the case. Captopril, an angiotensin converting enzyme inhibitor, increases plasma renin levels in patients with renovascular hypertension and can be used to increase the accuracy of the test.

A more invasive test is measurement of renal vein renin by selective catheterization of each vessel via a femoral vein puncture. If unilateral disease is present the affected kidney should secrete high levels of renin and the contralateral kidney should have suppressed renin production. A ratio between the two kidneys of more than 1.5 has long been used as a highly specific (80–100 per cent) test for renovascular hypertension. Unfortunately, the test lacks sensitivity: false-negative rates of 20 to 50 per cent have been reported. Expressing the higher renin level as an increment relative to the inferior vena cava renin level is said to increase the sensitivity to 74 per cent. If the renal vein renin ratio is measured before and after administration of captopril, sensitivity of the test is increased, but at the expense of specificity.

Renin levels are influenced by fluid balance and by antihypertensive treatment. Ideally these tests should be performed when the patient has been off treatment for 2 to 3 weeks, with controlled sodium intake, and after 12 h of rest but it is often difficult to comply with such restrictions. Impaired renal function may also affect the test. Nevertheless, renal vein renins can provide valuable information if the results are positive and are compatible with unilateral disease, provided that a negative investigation is ignored.

Isotope scanning

Radionuclide renal scanning is a less invasive means of assessing renal function. Labelled hippuran can be used to assess renal blood flow. However, this alone offers no more than other measures of renal artery stenosis. A reduction in renal blood flow following exercise is said to predict a poor response to renal revascularization.

Labelled diethyltetraphenyl aminic acid is used more widely to provide a measure of glomerular filtration rate: in renal artery stenosis, uptake and excretion are delayed. Oral captopril administration produces a further fall in glomerular filtration rate in patients with renovascular hypertension, and makes this a relatively reliable test. Unfortunately, most of these tests are only helpful in unilateral disease.

‘The gold standard'

The ‘gold standard' by which all these tests are assessed is the response of blood pressure to restoration of blood supply. There has been a suggestion that balloon dilatation of the renal artery at the time of intra-arterial angiography should be used as both test and treatment, in view of its low morbidity and mortality reported from certain centres. This approach is not in general use.

Angiography

Intra-arterial aortography is necessary prior to intervention to demonstrate the lesion accurately (Fig. 3, Fig. 4, Fig. 5). Intravenous studies do not provide adequate resolution to plan treatment. A flush aortogram needs to be performed before selective catheterization of the renal arteries, so that any accessory renal arteries, which may be diseased, are detected and the origins of all the renal arteries adequately shown. More recently, there have been significant advances in CT spiral scanning and magnetic resonance angiography which can provide excellent images of renal artery lesions and are likely to replace conventional angiography.

Treatment

Medical

The advent of powerful antihypertensive medications, such as b-blockers, calcium channel blockers, and angiotensin converting enzyme inhibitors, has brought blood pressure under satisfactory control in most patients. In patients with atheromatous disease, the risks of any further intervention are increased because of their

associated disease and age. If there is no evidence of renal failure and the blood pressure can be satisfactorily controlled by medication, there is probably no justification for more aggressive intervention. A reduction in systemic pressure, however, may reduce renal perfusion still further and lead to progressive renal failure. Renal function should be regularly monitored in these patients so that further intervention can be instituted before critical loss of renal substance. Angiotensin converting enzyme inhibitors can cause a reduction in glomerular filtration rate by removing the selective vasoconstrictive action of angiotensin II on the efferent arterioles and so precipitate renal failure. This is particularly likely to happen when there is a critical stenosis in a solitary kidney, for example a transplanted kidney, or bilateral disease.

Patients with fibromuscular dysplasia are often diagnosed before or during middle age. The risks of intervention are lower and the outcome usually better than that in atherosclerotic patients. In addition, the advantage to the patient in terms of years free of antihypertensive treatment, or on reduced medication, are such that an attempt is usually made to correct the defect.

Almost any child or young adult with hypertension due to renal artery stenosis should have the stenosis corrected; the results are good whatever the cause.

Percutaneous transluminal angioplasty

Percutaneous transluminal angioplasty of the renal artery is performed under angiographic screening. A guidewire is fed up to the renal artery from a femoral artery puncture site and then passed across the stenosis (or occasionally across an occlusion). Once in position a catheter bearing a balloon at its tip is slid over the guide wire. Inflation of the balloon disrupts plaque or stretches dysplastic vessel wall to remove the stenosis.

This technique works well in fibromuscular dysplasia: 50 per cent of patients are cured and maintain diastolic pressure below 90 mmHg without antihypertensive medication. Most of the remainder have improved blood pressure control, although they still require medication. Patients with atheromatous disease fare less well. The ‘cure' rate is only about 20 per cent although a further 60 per cent are improved. Atheromatous disease of the renal artery is often an extension of aortic disease and limited to the renal artery ostium where it joins the aorta. In this situation satisfactory dilatation is usually unsuccessful because the balloon tends to bulge back into the aorta and only 25 per cent of patients have any long-term benefit. Percutaneous transluminal angiography also has a place in the treatment of renal failure in this disease. Major centres are reporting a reduced or normal creatinine level in over 85 per cent of patients with renal failure due to either fibromuscular dysplasia or atherosclerosis.

Complications of percutaneous transluminal angiography probably arise in less than 10 per cent of cases, and range from a significant haematoma at the puncture site to renal artery thrombosis or dissection and segmental infarction of the kidney by distal embolization of fractured plaque or thrombus. Although facilities for emergency surgery used to be made available because of the potential risks of the procedure, these risks are small; the chances of successful emergency surgical intervention are even smaller so this is probably no longer necessary. Restenosis occurs in fewer than 10 per cent of technically successful dilatations in patients with fibromuscular dysplasia but in at least 25 per cent of atheromatous lesions (15 per cent of non-ostial lesions). Mortality associated with the procedure ranges between 0 and 3 per cent but in most centres it is less than 1 per cent. Results vary slightly between different centres: those that have a more aggressive approach and use larger balloons tend to have better results without much increase in morbidity.

Renal artery stenting

If balloon angioplasty fails to dilate an ostial stenosis satisfactorily or if such a stenosis recurs, it is possible to place a metal stent to hold the origin of the renal artery open (Fig. 6). This technique is relatively new and follow-up studies are limited in terms of numbers and time scale. Initial technical success has been reported as greater than 90 per cent in some series but restenosis and occlusion rates of 20 to 30 per cent have been encountered over the following 1 to 2 years. Repeated endovascular intervention can sometimes still salvage these stenosed or occluded arteries and one study recently reported a 92 per cent secondary patency rate at 5 years with no major complications.

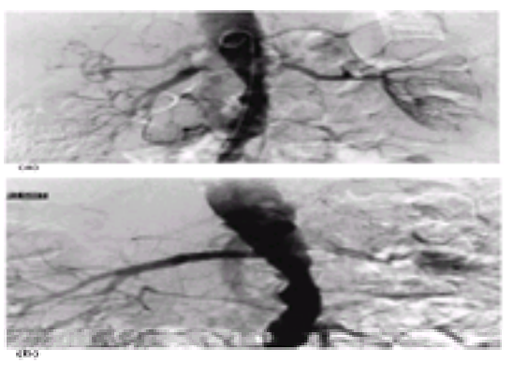


Fig. 6. Bilateral renal artery stenoses in an elderly patient with generalized atheromatous disease before (a) and after (b) angioplasty and insertion of a stent into the right renal artery (by courtesy of Dr Jane Philips-Hughes).

Outcome in terms of blood pressure control is very similar to that of successful angioplasty with about 16 per cent ‘cured' and 62 per cent ‘improved'. There are less data relating to the effect in patients with renal failure but approximately one-third appear to improve and a further third stabilize over the first 9 months.

It is tempting to question whether renal artery stenting will become the preferred treatment for atheromatous renovascular hypertension and renal failure. Preliminary results suggest outcomes that are approaching those of surgical revascularization (see below) without the associated perioperative mortality and morbidity. The answer lies in long-term comparisons of patients randomized to either stenting or surgery and these studies are still awaited.

Surgery

Anaesthesia

Expert anaesthesia for renal artery surgery is critical. Application and removal of an aortic clamp causes large fluid shifts and imposes significant cardiac stress. Adequate monitoring with good intravenous access is essential. Minimal requirements are intra-arterial and central venous lines for measuring pressure, and many would also advocate a Swan–Ganz catheter.

Surgical access

There are several different surgical approaches to revascularization of the kidney. Exposure of the renal arteries may be transperitoneal or retroperitoneal, although the latter method is less suitable for the right renal artery. A combined approach may also be used in which the peritoneal cavity is opened via a transperitoneal incision and then the spleen, pancreas, splenic flexure of the colon, and descending colon are mobilized and swung to the right to expose the aorta, the renal and superior mesenteric arteries, and the coeliac axis (Fig. 7).

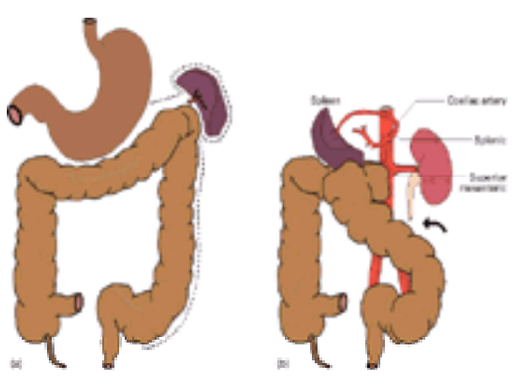


Fig. 7. The left colon, spleen, and pancreas are mobilized and swung to the right to expose the aorta, coeliac axis, and superior mesenteric and renal arteries.

Endarterectomy

Transaortic renal artery endarterectomy is often performed and is perhaps the procedure of choice for bilateral disease. The aorta is clamped and opened at the level of the renal arteries. A local endarterectomy of one or both vessels is performed, and can be extended to the aorta itself at this level if necessary (Fig. 8). This technique is particularly designed for the common situation when the renal artery disease is an extension of aortic plaque, narrowing only the ostium of the renal artery. It cannot be used if the disease is more distal in the vessel.

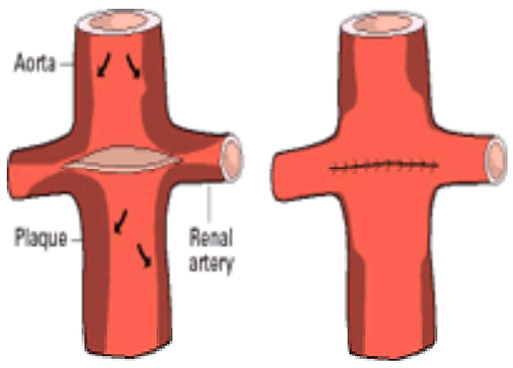


Fig. 8. Transaortic endarterectomy of the renal arteries.

Aortorenal bypass graft

The aorta is clamped, usually below the renal vessels to maintain flow in them as long as possible. A length of reversed saphenous vein is anastomosed to the aorta and then to the renal artery beyond the stenosis, usually end-to-end (Fig. 9). Synthetic graft material, such as polytetrafluoroethylene, may be used instead of vein.

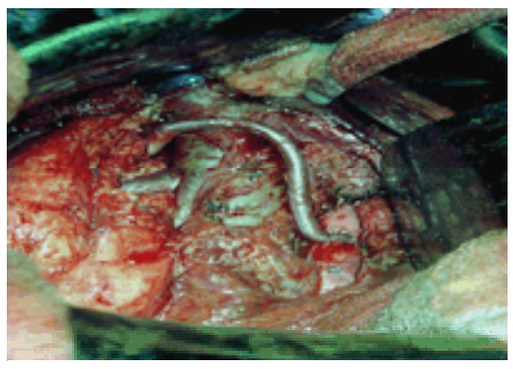


Fig. 9. Aortorenal bypass graft with saphenous vein.

Patients with atheromatous disease sometimes have occlusive aortoiliac disease or an abdominal aortic aneurysm which also requires treatment. In such cases it may be justifiable to replace the lower abdominal aorta with a graft at the same time. Patients with such extensive disease often have bilateral renal artery stenosis. Grafts can be run off the aortic graft to one or both renal arteries (Fig. 10).



Fig. 10. Replacement of diseased aorta with a bifurcated Dacron graft, together with revascularization of the left kidney with a saphenous vein graft from the aortic graft to the distal renal artery.

Ilio- or viscerorenal reconstruction

Occasionally the aorta is so heavily calcified that aortorenal grafting would present serious technical problems. The iliac arteries may be relatively spared, especially on the anterior wall, and an iliorenal graft can be constructed using saphenous vein provided there is a good pulse in the iliac artery below the diseased aorta (Fig. 11). Failing this, or if there is reluctance to clamp the aorta in a high-risk patient, the visceral circulation can be used to donate blood to the kidney. The splenic artery, provided it is relatively free of atheroma, can be transected and anastomosed to the left renal artery leaving the spleen *in situ* (Fig. 12). The hepatic artery can be used to supply the right kidney. A vein graft is usually taken from the side of the hepatic artery just beyond the gastroduodenal artery and run down to the right renal artery (Fig. 13). Alternatively the gastroduodenal artery itself, the communication between the coelic axis and the superior mesenteric circulations, can be divided distally and anastamosed to the right renal artery if it is long enough. Reconstructions based on the visceral arteries should only be performed if significant coeliac axis disease (and in the case of gastroduodenal artery use, superior mesenteric artery stenosis) have been excluded on angiography.



Fig. 11. Iliorenal saphenous vein bypass graft used when the aorta is heavily diseased and unsuitable as a site of graft anastomosis. Note that a stenosis has developed at the anastomosis of the saphenous vein graft to the renal artery approximately 2 years after reconstruction.

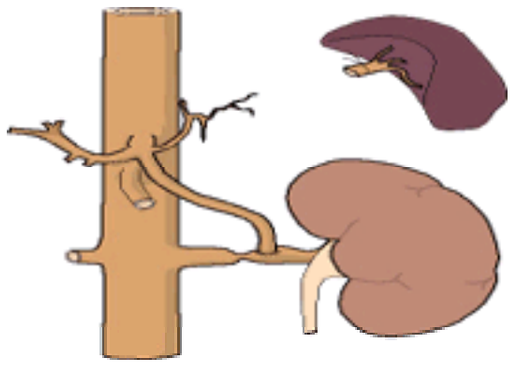


Fig. 12. Revascularization of the left kidney using the splenic artery.

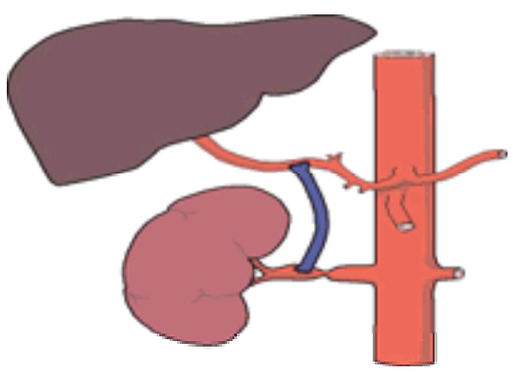


Fig. 13. Revascularization of the right kidney using a vein graft from the hepatic artery to the distal renal artery.

Autotransplantation

In fibromuscular dyspasia particularly, there may be stenoses of the branches of the renal artery, which are not suitable for angioplasty. Correction of these lesions requires microvascular techniques using vein graft; for this to be performed the kidney is usually removed from the body and perfused with cold fluid to protect against ischaemic damage. The reconstruction is performed *ex vivo* before replacing the kidney in the iliac fossa in the same way as a renal transplant. An autotransplant may be used for correction of a renal artery stenosis, in which case the ureter is left intact, the kidney reversed in both its longitudinal and trenasverse axes, and then implanted in the iliac fossa (Fig. 14).

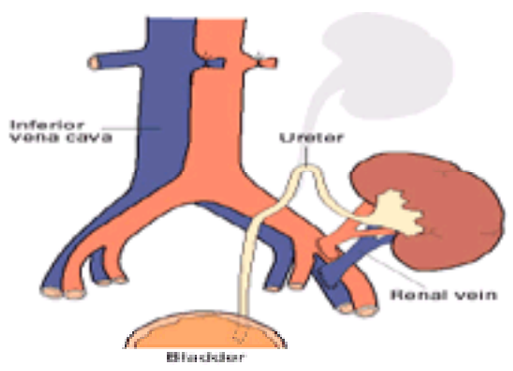


Fig. 14. Autotransplantation of the kidney.

Nephrectomy

Occasionally the kidney has shrunk too far for function to be saved by reperfusion, but it is still a source of renin, causing hypertension. In other cases it proves impossible safely to reperfuse the kidney, either because of the extent of disease or the frailty of the patient, and the other kidney is functioning. In these circumstances nephrectomy may provide a simple and safe remedy for hypertension.

Results

The results of surgery in fibromuscular dysplasia of the renal artery are generally good. About 60 per cent of patients are ‘cured' of their hypertension, 30 per cent have improved blood pressure control, and only 10 per cent fail to respond. Very few patients with fibromuscular dysplasia have evidence of renal failure but if present this is usually also improved, or at least stabilized, by surgical intervention.

Patients with atheromatous disease respond rather less well. Less than 40 per cent are probably ‘cured' of their hypertension, although slightly more are ‘improved', and about 20 per cent fail to respond.

Renal failure occurs in about 15 per cent of patients with renovascular hypertension due to atheromatous disease compared with only about 2 per cent of those with fibromuscular dysplasia. Overall, approximately 50 per cent of patients with renal failure are improved, 40 per cent stabilized, and 10 per cent show further deterioration. These results are in the face of declining renal function prior to surgery and so stabilization of renal function is an improvement attributable to surgery. The benefits of surgery decline with increasing degrees of renal failure. Only 20 per cent of those with a serum creatinine greater than 180 µmol/l are improved, about

60 per cent stabilized, and 20 per cent fail to respond. Of those with a serum creatinine greater than 270 µmol/l, about 30 per cent fail to respond, and the perioperative mortality rate, in one series, increased from 3 per cent to 13 per cent. Occasionally, patients with severe renal failure respond dramatically well to revascularization of the kidney. They tend to have bilateral renal artery occlusions and potentially functional renal tissue which has been maintained via collaterals. It is obviously important to be reasonably sure that such tissue exists before embarking on complicated, relatively high-risk surgery. The angiogram may show filling of the distal renal artery via collaterals (Fig. 4). Occasionally, but not always, a radioisotope scan demonstrates collateral blood flow reaching the kidney. Renal length on ultrasound is a fairly good guide: if this is less than 8 cm, revascularization is unlikely to improve renal failure and may increase renovascular hypertension. Demonstration of viable glomeruli in a renal biopsy, taken either pre- or perioperatively, is also useful.

The mortality rate associated with renal artery surgery is low in patients with fibromuscular dysplasia and hypertension. In patients with atheromatous disease it is 3 to 10 per cent if surgery is confined to only one renal artery but this increases with extended surgery: aortic reconstruction with bilateral renal artery revascularization carries a perioperative mortality rate of up to 15 per cent. As might be expected, the presence of significant atheromatous disease in the coronary and cerebral circulations increases mortality. It is probably also higher in those with renal failure. For these reasons the possibility of successful treatment by angioplasty, with or without stenting, is attractive.

A direct comparison between percutaneous transluminal angioplasty and surgery in the treatment of renovascular hypertension has not been made because the patient groups are dissimilar. The former has been used mainly to treat young patients with fibromuscular dysplasia who are otherwise reasonably fit. It has not been used widely for the patient with widespread atheroma affecting the aorta and extending into the renal arteries, who is also at much greater risk from surgery. The patients with fibromuscular dysplasia who come to surgery are those with complex branch lesions and those in whom percutaneous transluminal angioplasty has failed. However, in recent years there has been a widespread use of angioplasty with or without stenting for atheromatous renovascular hypertension. The majority of atheromatous renal artery stenoses involve the ostium of the renal artery and hence are very resistant to angioplasty or recur quite rapidly. However, the insertion of a stent across the stenosis and the ostium is being used more widely in recent years with encouraging early and medium-term results and may, in due course, replace open surgery in this high risk group of patients. One prospective trial has shown that angioplasty with a stent is technically superior to angioplasty alone.

Overview of atheromatous renal artery disease

This disease encompasses a wide spectrum of patients. At one end is the man in his fifties with significant stenosis of one renal artery causing hypertension, with normal renal function, a slightly irregular aorta, mild asymptomatic peripheral vascular disease, and no evidence of significant heart disease. At the other extreme is the woman in her eighties with severe renal failure, hypertension, rest pain, and a history of several myocardial infarctions. In between are patients with any combination and severity of these problems.

If hypertension that can be controlled satisfactorily with medication is the main problem and the kidneys appear to be functioning adequately, no further intervention is required, provided renal function is monitored regularly. If hypertension is difficult to control or renal function starts to deteriorate and the lesion is a short stenosis or even occlusion, percutaneous transluminal angioplasty is indicated. However, if the lesion is ostial, as it often is in these circumstances, this treatment may fail; if it succeeds the stenosis is likely to recur. In these circumstances many vascular surgeons would proceed directly to surgery although renal artery stenting is an alternative approach that should be considered particularly if the patient is high risk for surgery.

In patients with more extensive atheromatous disease, who are usually older, both renal arteries and the aorta itself are involved. Thus the patient may have renal failure and lower limb ischaemia or an abdominal aortic aneurysm, both of which require treatment. However, these patients also probably have significant myocardial ischaemia and will respond poorly to the stress of an aortic clamp; they also may have cerebrovascular disease, which increases the risk of perioperative cerebrovascular accident. Hypertension may be controllable by medication, rest pain by analgesics, renal failure by dialysis, and the risks of aneurysm rupture can be ignored. However, this approach usually leads to a relatively poor quality of life. It is possible to replace the aorta and revascularize both kidneys, even in the elderly. Obviously the risks associated with such surgery depend on the patient's general health and the skill of the surgeon and anaesthetist. In major centres carrying out such surgery the perioperative mortality has been reported as less than 6 per cent. Renal failure is improved or stabilized in all but 6 per cent, and hypertension improves in over 90 per cent. If the patient's health dictates it and the extent of associated disease allows, the kidneys may be reperfused via the visceral vessels to avoid the dangers of aortic clamping. Although surgical treatment of renovascular disease was previously restricted to the young and middle-aged, it is now being used in the treatment of the elderly, in whom medical treatment or percutaneous transluminal angiography prove unsatisfactory. The attendant risks to heart and brain can be reduced by careful monitoring, and even coronary artery bypass grafting or carotid endarterectomy prior to renal revascularization.

Overview of renovascular disease due to fibromuscular dysplasia

These patients are usually young, and have none of the stigmata of generalized vascular disease. They rarely develop renal failure, but they are at risk from cerebrovascular accidents or heart disease because of their hypertension: this must be detected and treated before the development of irreversible end-organ damage. Younger patients have a less satisfactory response to medical treatment, and surgical intervention is less hazardous. In general, fibromuscular renovascular hypertension responds well to either percutaneous transluminal angioplasty or surgical intervention, but the former is preferred since it is a relatively simple procedure with fewer associated risks. Most dysplastic lesions are amenable to percutaneous transluminal angioplasty, except for those in the smaller branches of the renal artery where the risks of dilatation are greater and the technical success rate diminishes. Surgery is usually only undertaken in patients in whom percutaneous transluminal angioplasty has failed or in whom the lesion is in the distal branches, where reconstruction *ex vivo* (bench surgery) is performed.

Children with renovascular hypertension

The most common cause of renovascular hypertension in children is fibromuscular dysplasia. Other causes include arteriovenous malformations, aneurysms, hypoplasia, Takayasu's disease, and neurofibromatosis. If the lesion is proximal it is occasionally possible to excise the affected artery and reanastomose the vessel to the aorta; however, a more complex procedure is usually required. Aortorenal bypass grafting is an option, but the use of saphenous vein in prepubertal children has led to aneurysmal dilatation or stricture formation in the graft at a later stage. This can be avoided by using autogenous artery, usually one of the internal iliac arteries. An alternative approach is to autotransplant the kidney. This approach is essential when microscopic bench repair of intrarenal abnormalities is required. Hypoplastic renal arteries are often associated with a low coarctation of the aorta, making this vessel unsuitable for bypass grafting. However the iliac arteries are usually spared, and either the kidney can be autotransplanted or an iliorenal graft constructed. Occasionally, especially in those patients with unilateral disease, the kidney is not worth salvaging and the only option is nephrectomy.

The results of surgery for renovascular hypertension in children are excellent, with a cure rate of 90 to 100 per cent.

Renal artery trauma

Mechanism of injury

Deceleration in either a vertical or a horizontal direction imposes stretching and shearing forces on the renal artery because it joins the relatively mobile kidney to the fixed aorta. These forces cause intimal tears, particularly in the middle third of the artery, which may then lead to dissection and thrombosis of the vessel. Occasionally the entire vessel wall is disrupted and a perirenal haematoma forms. There is often remarkably little blood loss (and few concomitant physical signs) because of renal artery spasm. Blunt trauma to the kidney usually causes parenchymal damage, but it may disrupt the renal pedicle.

Complete occlusion or disruption of the main stem of the renal artery leads to renal infarction. Most of these injuries occur in young adult males with previously normal renal arteries. Development of cortical collateral vessels that might maintain viable renal tissue is usually poor. Thus, in most of these patients there is a very narrow window of opportunity, probably within 1 to 2 h of injury and almost certainly within 6 h, in which to revascularize the kidney and recover function. There are reports in the literature of successful revascularization several days after injury, but these are relatively rare. Late revascularization is usually associated with the development of renovascular hypertension and a need for nephrectomy. The patient may develop renovascular hypertension even if no attempt is made at revascularization, since ischaemic renal tissue survives, maintained by a partially occluded renal artery or cortical collaterals. The incidence of hypertension following trauma has been cited as between 10 and 50 per cent, but it is difficult to assess: not all cases of renal trauma are detected, and follow-up is incomplete on those that are detected. It develops between days and years after the injury and in up to 80 per cent of cases appears to be transient.

Symptoms and signs

Renal artery damage is often missed in the initial assessment since there are frequently few signs and often other injuries. It is more likely to be discovered at autopsy, during later urological assessment, or when the patient is evaluated for hypertension. At the initial presentation, patients may complain of unilateral flank pain, and occasionally of abdominal pain. Patients with severe disruption present with shock; those with lesser injuries have at least microscopic haematuria, some have loin tenderness and in a few a bruit can be heard in the upper abdomen or flank.

Investigations

Any trauma patient with even microscopic haematuria must undergo intravenous urography. This will demonstrate any abnormality in renal function and, very importantly, will confirm that the patient has another functioning kidney. Patients in whom the intravenous urogram is normal can be treated conservatively; if the examination shows no function on one side and the patient is haemodynamically unstable, immediate laparotomy is required to deal with the renal artery damage. If the patient is stable but has no function demonstrable in one kidney, angiography or CT scan with contrast to demonstrate the arteries may be helpful as a prelude to reconstructive surgery, provided it does not delay intervention unduly.

A patient who presents with hypertension some time after trauma should undergo angiography to delineate any renal artery defect, in the same way as any other patient with renovascular hypertension.

Treatment

Most patients with renal trauma and normal renal function on intravenous urography can be managed conservatively. Those with severe disruption of the renal pedicle who present with profuse bleeding and absent renal function need emergency laparotomy. In these patients, extensive disruption often precludes reconstruction of the kidney, and nephrectomy is required.

The stable patient with absent renal function can either undergo surgery with a view to reconstruction or be treated conservatively. If more than 6 h have elapsed since the injury, reconstruction is unlikely to be successful and, unless both kidneys are non-functional, is probably not justified. The ischaemic kidney often shrivels and poses no further problem. If the other kidney is normal, renal function is maintained and relatively few patients develop troublesome renovascular hypertension. Laparotomy to remove an ischaemic kidney is unnecessary at this stage.

Reconstruction may be feasible if surgery is performed within 6 h of injury. Thrombectomy of the renal artery and excision of the affected segment with reanastomosis or a vein interposition graft are usually required.

Patients known to have suffered trauma to the renal artery, especially those with non-functioning kidneys on intravenous urography, should undergo regular blood pressure monitoring so that renovascular hypertension can be detected and treated before end-organ damage results.

Renal artery aneurysms

Incidence

Renal artery aneurysms appear to be uncommon. Retrospective studies of postmortem reports put the prevalence at about 0.01 per cent; however many are small and intrarenal, and might not be detected on routine postmortem examination. Angiographic studies detect such aneurysms in up to 1 per cent of individuals (Fig. 15), although these are often selected for study on the basis of symptoms such as pain, haematuria, or hypertension, which might bias the findings. However, one prospective postmortem study found the prevalence to be 9.7 per cent in an apparently unselected series of individuals. Renal artery aneurysms occur with equal frequency in either sex and have been detected at all ages.

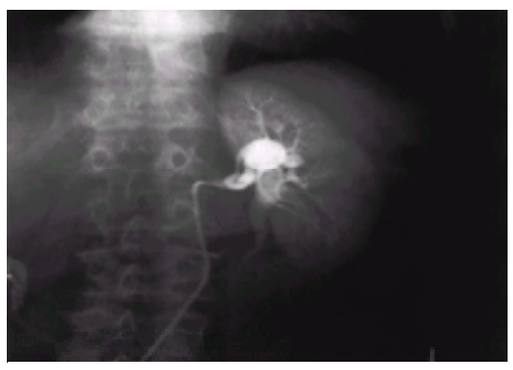


Fig. 15. Renal artery aneurysm (by courtesy of Dr E. W. Fletcher).

Pathology

The majority of these aneurysms are less than 1 cm in diameter and most are saccular. They occur with equal frequency in the main stem artery, in the primary branches, and in the peripheral branches. They are bilateral in up to 10 per cent of patients, and multiple aneurysms occur within one kidney in another 13 per cent. Many of these aneurysms occur at branch points and, as they sometimes occur in children, they may be due to congenital weakness of the arterial wall at these points. They are rarely associated with atherosclerosis, despite the fact that about 20 per cent are calcified.

Fibromuscular dysplasia, and in particular the most common form, medial fibroplasia, is associated with microaneurysm formation in the main stem artery or its primary branches. However, these are usually considered separately from other renal artery aneurysms.

Consequences of renal artery aneurysms

The predominant risk is rupture, but this is very unlikely in small aneurysms (< 1 cm) and uncommon even in larger aneurysms. Rupture is most likely to occur during pregnancy, especially in the third trimester, but even then splenic artery aneurysm rupture is four to five times more common. It presents with abrupt onset of flank pain and hypotension and the mortality is high for both mother (56 per cent) and fetus (82 per cent). Few renal artery aneurysms which rupture in pregnancy are associated with medial hyperplasia, despite its occurrence mainly in young women. Prior to rupture they are usually asymptomatic, although a bruit may be audible.

Aneurysms may thrombose but, because they are saccular, rarely cause arterial occlusion. For the same reason they are unlikely to shed emboli into the distal renal vasculature. Occasionally they are associated with haematuria, presumably because of erosion into the renal pelvis.

Renal artery aneurysms have been associated with hypertension but are rarely the cause of it. Many of the patients in whom they are detected on angiography are investigated because of hypertension and the aneurysm is incidental: surgical treatment of the aneurysm does not usually cure the hypertension unless the aneurysm distorts and compresses the adjacent artery and causes a functional stenosis. Fibromuscular dysplasia may be associated with both hypertension and renal artery aneurysms but it is the stenotic component of the disease that is important in producing hypertension.

Treatment

An aneurysm detected in pregnancy should be repaired before the third trimester, despite the risks to the fetus. Females of childbearing age should also have such an aneurysm repaired. The aneurysm can often be excised and the defect in the side of the artery closed with a vein patch. If the aneurysm is more extensive an aortorenal vein graft may be required. When multiple intrarenal aneurysms are present, *ex vivo* reconstruction becomes necessary. Many aneurysms can be occluded with a Gianturco coil inserted via a femoral artery catheter, but this approach does carry with it the risk of renal artery occlusion or distal embolization.

Unless a renal aneurysm is thought to be the cause of hypertension or troublesome haematuria from the renal artery, they rarely need treating in any other patients. Few patients with aneurysms experience problems, even over long periods of time.

Renal arteriovenous fistulae

Aetiology

Renal arteriovenous fistulae are uncommon. Approximately 30 per cent are congenital; nearly half of the acquired fistulae are secondary to renal biopsy and many follow renal surgery, such as nephrolithotomy, partial nephrectomy, or renal trauma. A few may arise from erosion of a renal artery aneurysm into an adjacent vein and they are sometimes seen in renal tumours. Fistulae occasionally develop in the renal pedicle following nephrectomy, probably due to mass ligature of the artery and vein. Fistula following renal biopsy seems particularly likely to occur if the patient is hypertensive prior to the procedure.

Pathology

Congenital arteriovenous fistulae are often large, complex, cirsoid malformations with multiple communications, and occupy much of the kidney (Fig. 16). Those which develop following biopsy are usually small and intrarenal (Fig. 17). Those associated with trauma may affect the intra- or extrarenal vessels and sometimes even the inferior vena cava.



Fig. 16. Large arteriovenous fistula in hilum of the kidney.



Fig. 17. Small arteriovenous fistula in the kidney following renal biopsy.

Small arteriovenous fistulae usually cause no problems. Larger fistulae are associated with two main problems. The first is due to the increased demands on cardiac output as blood is diverted through the low resistance fistula away from the general circulation. Systolic hypertension with a low diastolic pressure develops, and the patient may suffer high output cardiac failure. The second complication is the steal of blood from normal renal parenchyma to the fistula. Relative ischaemia of renal tissue stimulates renin production and leads to renovascular hypertension, so that diastolic pressure also rises. In a patient with pre-existing renal parenchymal disease, a fistula may also seriously compromise renal function. Rarely, an arteriovenous fistula ruptures to produce an intra- or extrarenal haematoma.

Symptoms and signs

About 60 per cent of patients have microscopic or macroscopic haematuria, but unless they have symptoms associated with heart failure or hypertension there are usually no significant problems in the history. On examination the only sign of the fistula itself is a continuous bruit audible in the upper abdomen or flank.

Investigations

An intravenous urogram may demonstrate reduced opacification of renal substance where blood has been diverted to the fistula, or distortion of the renal pelvis by an intrarenal fistula. In most cases, though, this investigation is of little help. Renal isotope scans may also demonstrate reduced perfusion but offer little more. The definitive investigation is the angiogram: this demonstrates the fistula, its feeding and draining vessels, and the extent of steal from the rest of the kidney.

Measurement of renal vein renin levels may not be helpful in the patient who has also developed renovascular hypertension, because the high flow of arterial blood through the renal vein dilutes renin washed out from the affected kidney.

Treatment

Most small intrarenal arteriovenous fistulae do not require treatment. The majority of those that develop following renal biopsy or trauma close spontaneously within 18 months.

Single intrarenal fistulae which cause problems can often be treated by embolization under radiographic control. Metal coils are usually placed in the segmental feeding vessel to promote thrombosis, often with temporary balloon catheter occlusion of the vessel proximal to this point to prevent the coils being carried through into the vein and systemic circulation.

Congenital arteriovenous fistulae are often too complex to treat by embolization and require surgery. This usually involves partial or complete nephrectomy because of the extent of renal disease.

Fistulae of the pedicle vessels are usually treated surgically. The fistula is excised and standard repair of the artery and vein performed, either by simple suture, vein patch, or interposition graft.

Hypertension associated with renal arteriovenous fistulae is improved or cured in 60 per cent of patients in whom the fistula is closed or the kidney removed. If the fistula is associated with trauma, the response rate increases to 85 per cent.

Further reading

Blum U, Krumme B, Flugel P, *et al.* Treatment of ostial renal-artery stenoses with vascular endoprosthesis after unsuccessful balloon angioplasty. *New England Journal of Medicine* 1997; **336**: 459–65. [Very encouraging results in the medium term with respect to control of blood pressure suggest that this is an effective treatment for renal artery stenosis involving the ostium.]

Dean RH. Surgical management of renovascular disorders. In: Rutherford R.B. ed. *Vascular surgery*, 4th edn. WB Saunders, Philadelphia, 1995, pp. 1371–455. [An excellent series of contributions covering all aspects of renovascular disease.]

Fiala LA, Jackson MR, Gillespie DL, O'Donnell SD, Lukens M, Gorman P. Primary stenting of atherosclerotic renal artery ostial stenosis. *Annals of Vascular Surgery* 1998; **12**: 128–33. [High early

success rate but a significant restenosis rate by 24 months and the authors advise caution in the application of stenting.]

Goldblatt H, Lynch J, Hanzai RF, Summerville WW. Studies on experimental hypertension–I, The production of persistent elevation of systolic pressure by means of renal ischaemia. *Journal of Experimental Medicine* 1934; **59**: 347–79. [The classical experiment underlying the physiology of renovascular hypertension.]

Hansen KJ, Dietch JS, Oskin TC, Ligush J, Craven TE, Dean RH. Renal artery repair: consequence of operative failures. *Annals of Surgery* 1998; **227**: 678–89. [Although the need for secondary reconstructions is low, they are associated with a significant risk of dialysis dependence.]

Maldonado JE, Sheps SG, Bernatz PE, DeWeerd JH, Harrison EG. Renal arterio-venous fistula. *American Journal of Medicine* 1964; **37**: 499–513. [The first review of renal arteriovenous fistulae.]

Novick AC. Surgical correction of renovascular hypertension. *Surgical Clinics of North America* 1988; **68**: 1007–25. [A review of surgical approaches to renovascular hypertension.]

Peterson NE. Review article: traumatic bilateral renal infarction. *Journal of Trauma* 1989; **29**:158–67.

Sos TA. Angioplasty for the treatment of azotaemia and renovascular hypertension in atherosclerotic renal artery disease. *Circulation* 1991; **83** (suppl. I): I1162–6.

Tegtmeyer CJ, Bayne Selby J, Hartwell GD, Ayers C, Tegtmeyer V. Results and complications of angioplasty in fibromuscular disease. *Circulation* 1991; **83** (suppl. I): I155–61.

Tham G, Ekelund L, Herrlin K, Lindstedt EL, Olin T, Bergentz S-E. Renal artery aneurysms. Natural history and progress. *Annals of Surgery* 1983; **197**: 348–52.

van Bockel JH, van Schilfgaarde R, van Brummelen P, Terpstra JL. Renovascular hypertension. *Surgery, Gynecology and Obstetrics* 1989; **169**: 467–78. [A European experience of the management of renovascular hypertension.]

van de Ven PJ, Kaatee R, Beutler JJ, *et al.* Arterial stenting and balloon angioplasty in ostial atherosclerotic renovascular disease: a randomised trial. *Lancet* 1999; **353**: 282–6. [In this prospective, randomized trial, angioplasty with stenting proved superior to angioplasty alone.]

Williamson WK, Abou-Zamzam AM, Moneta GL, *et al.* Prophylactic repair of renal artery stenosis is not justified in patients who require infra renal aortic reconstruction. *Journal of Vascular Surgery* 1998; **28**: 14–20. [The authors suggest that asymptomatic renal artery stenoses in patients requiring abdominal aortic reconstruction may be ignored.]

17.10.1 Arterial emboli in limbs

Bruce Campbell

[Definition and aetiology](#)
[Sites affected by emboli](#)
[Pathophysiology](#)
[Emboli to the limbs](#)
[Diagnosis](#)
[Differential diagnosis](#)
[Treatment](#)
[Other methods of treating emboli](#)
[Long-term anticoagulation after embolism](#)
[Late presentation of arterial emboli](#)
[The mortality and morbidity of peripheral emboli](#)
[Microemboli to the limbs](#)
[Other sites affected by emboli](#)
[The brain and the eye](#)
[Mesenteric emboli](#)
[Further reading](#)

Definition and aetiology

An embolus consists of undissolved material that is carried in the circulation and impacts in a blood vessel, usually blocking it. The most common source of arterial embolism is the left atrium in atrial fibrillation, accounting for two-thirds of all cases. Thrombus forms because of stasis in the enlarged and fibrillating atrium, and fragments detach to enter the arterial circulation (Fig. 1). Thrombi can also form on the damaged endocardium of the left ventricle after myocardial infarction, and arrhythmias cause these to detach and embolize. Other causes of emboli from the heart are shown in Table 1.



Fig. 1. To the left of the picture is embolic material from the left atrium, containing typical pale platelet thrombus. To the right lies a long length of propagated clot, extracted from the superficial femoral artery using the Fogarty embolectomy catheter shown above, with its balloon inflated.



Table 1 Sources of arterial emboli

Emboli can arise from the aorta and its branches (Fig. 2 and Fig. 3).

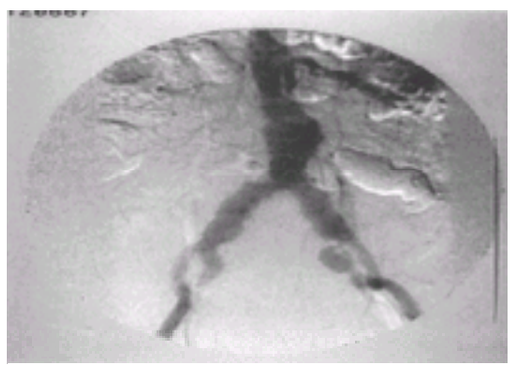


Fig. 2. Digital subtraction arteriogram of an ectatic and diseased aortoiliac system, which was the source of atheromatous emboli to both lower limbs. The patient was treated with an aortobifemoral graft.

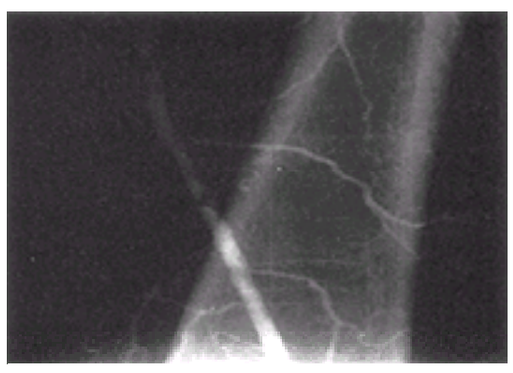


Fig. 3. Arteriogram showing a localized stenosis in the superficial femoral artery. The patient presented with signs of small emboli to the foot. A short segment of reversed saphenous vein was used to replace this segment of the artery. Balloon angioplasty is an alternative method for an isolated stenosis.

Atheromatous plaques may rupture, allowing cholesterol debris to pass distally and block small arteries. Platelet thrombi forming on ulcerated atheromatous stenoses may also embolize. Thrombus forming in aneurysms due to abnormal patterns of blood flow can also form an embolism and this is a particular danger in the case of popliteal aneurysms. Table 1 lists the common sites of origin for arterial ‘atheroemboli', and other rare causes of embolism. The clinical effects of small emboli are described later.

Sites affected by emboli

Emboli usually lodge at the bifurcations of arteries, because the diameter of each major branch is less than that of the main branching vessel.

In surgical practice most arterial emboli affect the limbs, the leg being affected six times more often than the arm (Table 2). Other important sites for embolism include the cerebral, ophthalmic, and mesenteric arteries.

Lower limb	
Aorta and iliac artery	26 %
Femoral artery	45 %
Popliteal artery	15 %
Tibial artery	1 %
Upper limb	
Subclavian artery	1 %
Axillary artery	5 %
Brachial artery	7 %

From a series of 396 emboli (Panetta *et al.* 1986).

Table 2 Sites of occlusion of emboli to the limbs

Pathophysiology

The effects of an embolus blocking a major limb artery depend on the level of the obstruction and on the capacity of collateral arteries to carry blood to the distal tissues. In the absence of good collateral flow there is stasis of blood in the arteries beyond the block and propagated clotting occurs (Fig. 1). Propagated clot also extends proximally to the next major branch. Reflex spasm of distal arteries is another effect of acute arterial occlusion. Clotting and spasm both make ischaemia worse.

Acute ischaemia due to an embolus causes hypoxia of the tissues and a failure to remove waste products; these are particularly damaging to muscle cells, which have a rapid metabolic rate. Muscle death starts to occur after about 6 h. Initially ischaemia causes pain, due to accumulation of metabolites, but as peripheral nerves become increasingly hypoxic, paraesthesiae and eventually complete anaesthesia of the extremity occur.

If the occlusion remains unresolved, venous thrombosis results from stagnation of blood flow: this is a late feature associated with a poor prognosis. ‘Fixed staining' of the skin due to extravasation of blood into the tissues is another late sign (Fig. 4). Continued neglect results in gangrene.



Fig. 4. Fixed staining of the skin: this is a late case of acute ischaemia and the foot could not be salvaged; below-knee amputation was done.

Emboli to the limbs

Diagnosis (Fig. 5)

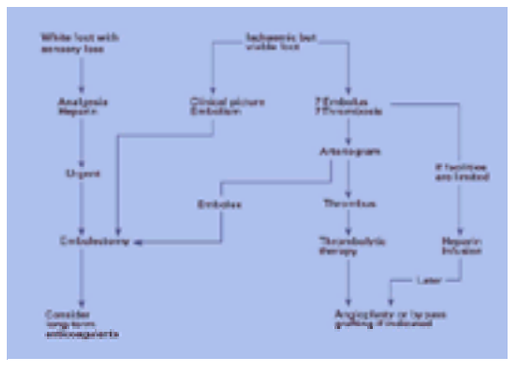


Fig. 5. Scheme for management of the acutely ischaemic limb.

The clinical features are best remembered as the ‘six Ps': pain, pallor, pulselessness, paraesthesiae, paralysis, and perishing cold. The onset of symptoms caused by an embolus is sudden, with pain and pallor occurring first. Colour change is variable and depends on the amount of collateral blood flow. If there are no established collaterals the extremity is white, sometimes with a bluish tinge. If some blood flow is maintained a pink colour remains, but capillary return is slower than normal. The more profound the ischaemia, the sooner paraesthesiae will be followed by anaesthesia. Loss of sensation is a serious sign and an indication for urgent treatment to

restore blood flow. Paralysis is also a sign of advanced ischaemia.

Pulses distal to the occlusion are lost, while immediately proximal to the occlusion the pulse may be enhanced due to the high resistance caused by obstruction.

Acute ischaemia due to an embolus is a clinical diagnosis and special tests should not be necessary, but Doppler ultrasonography confirms absent or poor blood-flow signals in the distal arteries, and the systolic pressure in the extremity is unrecordable or low. When the clinical diagnosis of embolism is obvious, with a white bloodless limb and sensory impairment, together with a likely source (for example atrial fibrillation), no further tests are necessary and the aim should be for embolectomy as soon as possible. When the limb is ischaemic yet viable, and particularly when there is any doubt about the diagnosis of embolism, then an arteriogram should be done (Fig. 6).

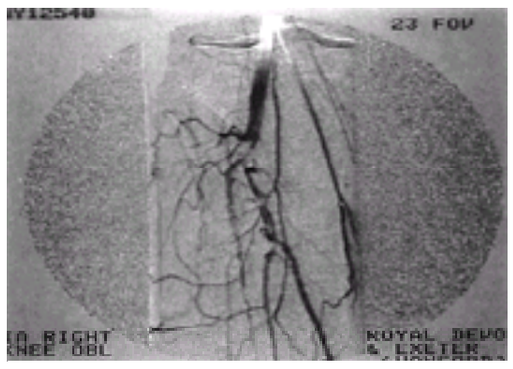


Fig. 6. Digital subtraction arteriogram showing embolic material at the bifurcation of the popliteal artery and in the tibioperoneal trunk.

Differential diagnosis

The main differential diagnosis is acute thrombosis occurring in arteries already narrowed by atherosclerosis (often called thrombosis *in situ* or acute-on-chronic ischaemia). With the increasing prevalence of atherosclerosis as a cause of thrombosis *in situ*, and the decrease in rheumatic heart disease as a source for emboli, thrombosis and embolism are seen with equal frequency as causes of acute limb ischaemia. The onset of ischaemia is often less sudden than in embolism and the degree of ischaemia is less profound, with preservation of a pink colour to the skin and intact sensation, owing to established collateral arteries. There may be evidence of chronic arterial disease with a history of intermittent claudication, and absence of pulses with reduced systolic pressures in the contralateral limb. The absence of an obvious embolic source also supports a diagnosis of thrombosis rather than embolism. A generalized illness with hypotension or dehydration may be evident as a precipitating cause for acute thrombosis.

The distinction between embolism and thrombosis is not always easy, and in any doubtful case an arteriogram should be performed. If arteriography suggests thrombosis, then this can be treated by low-dose infusion of a thrombolytic agent through an arterial catheter. This method of treatment requires radiological facilities and expertise. When the thrombus has been lysed, transluminal angioplasty or bypass grafting may be indicated for repair of the underlying arterial stenoses.

Deep venous thrombosis is sometimes confused with arterial embolism, because it causes pain, reduced ability to move the limb, and colour change. This differential diagnosis should not be difficult: the limb with a venous thrombosis is swollen, sensation is not lost, and Doppler examination will confirm a normal arterial supply. Venous thrombosis occurring as a late sequel of arterial occlusion is accompanied by florid signs of neglected ischaemia.

Treatment

The first priority is relief of pain using strong analgesics such as morphine or pethidine (meperidine) given intramuscularly. An intravenous bolus of heparin should be given to reduce propagated clotting (5000–10 000 units). Continuous infusion of intravenous heparin is then commenced if any delay is anticipated in definitive treatment.

The ischaemic limb should never be actively warmed—this accelerates tissue damage. Measures to cool the limb are sometimes recommended but are generally impracticable. Another traditional recommendation is to nurse the patient with the limb dependent but any improvement from such a manoeuvre is likely to be marginal. Treatment for any acute medical condition, such as cardiac failure or arrhythmias, is commenced.

After these initial steps the aim should be emergency embolectomy as soon as possible to re-establish blood flow to the extremity. Embolectomy is usually done under a local anaesthetic because most patients are elderly and unfit. The area to be anaesthetized should be marked and a dilute local anaesthetic, for example 0.5 per cent lignocaine (lidocaine), chosen so that a large volume can be used. It is helpful to have an anaesthetist in attendance to monitor the patient (electrocardiogram and pulse oximetry) and to administer oxygen or sedation if this is required. An intravenous infusion should be set up before starting the operation.

Transfemoral embolectomy is the best initial approach for all emboli of the lower limbs; in the upper limb, exposure of the brachial artery is required. For femoral embolectomy the incision should be vertical over the femoral artery, and should allow easy access up to the inguinal ligament. The common femoral artery is controlled with a sling as high as possible. Its bifurcation is exposed and slings are passed around the superficial and profunda femoris arteries, and also around any other branches.

In the upper limb it is easiest to approach the brachial artery through a longitudinal incision over the medial aspect of the upper arm. However, an approach through the antecubital fossa is preferable, allowing separate embolectomy of radial and ulnar arteries. This involves an S-shaped incision (medial above the elbow, curving across the antecubital fossa to the midline of the forearm). The bicipital aponeurosis is divided to reach the brachial bifurcation.

If the artery is non-pulsatile it is opened without any clamps in place. For simple embolectomy, transverse incisions in arteries allow closure without the risk of narrowing the lumen. Longitudinal arteriotomy gives greater scope for reconstructive surgery but may require closure with a patch to avoid stenosis in narrow vessels.

A balloon embolectomy catheter of appropriate size is selected (Table 3) and the balloon is tested by inflating it with fluid (from a 1- or 2-ml syringe). The stilette should always be removed before passing the catheter. The catheter is passed, proximally if it is necessary to remove clot and to ensure good inflow, and a proximal clamp is applied. After passing the catheter through an iliac clot into the aorta the contralateral limb must be checked at the end of the operation to ensure that embolic material has not been dislodged distally through the opposite iliac system.

Site	*Size of embolectomy catheter (French)*
Iliac arteries (including aortic bifurcation)	5
Superficial femoral, popliteal, profunda	4
Tibial arteries	2 or 3
Brachial, ulnar, radial	3 or 4

These will vary depending on the calibre of the patient's arteries, but the sizes shown are a guide to those commonly used.

Table 3 Sizes of balloon embolectomy catheters at different sites

The balloon catheter is passed distally and the clot withdrawn, adjusting the balloon inflation pressure depending on the ‘feel' of the catheter as it is withdrawn. Balloon catheters must be used gently, especially in vessels roughened by atheroma, to avoid intimal damage. It should be possible to pass a catheter to the foot and passages are repeated until no further thrombus or clot is withdrawn. A balloon catheter should also be passed down the profunda femoris artery in the thigh. Distal flushing with 50 to 100 ml of heparinized saline (from 500 ml normal saline containing 5000 units heparin) is then performed through a soft catheter (for example a No. 6 umbilical catheter) passed into the distal artery.

The patient should be warned to expect some discomfort during balloon withdrawal and also perhaps during flushing of the extremity with cold heparinized saline. If the first catheter withdrawal causes pain, then 8 to 10 ml of 0.5 per cent lignocaine (lidocaine) can be infused down the distal arteries before further passages of the catheter.

The arteriotomy is closed with a continuous non-absorbable suture (e.g. 0000 or 00000 polypropylene), checking inflow before final closure.

On releasing clamps the extremity should rapidly regain a pink colour, with return of palpable pulses. If the result is unsatisfactory (the catheter fails to pass far enough or the foot is not improved), then an immediate arteriogram should be undertaken on the operating table with a proximal clamp in place to demonstrate the state of the distal vessels. The options thereafter are exposure of the popliteal artery for embolectomy or bypass surgery, or the use of thrombolytic agents. These are best performed by a surgeon with vascular expertise. If the expertise is not available and the foot is viable, then intravenous heparin therapy should be instituted. For the surgeon without vascular expertise it is best to transfer the patient with an ischaemic but viable limb to a vascular unit, rather than embarking on an embolectomy that may turn out to be complex.

If the material removed from the arteries is not typical thrombus it should be sent for histological examination to exclude a tumour embolus.

Swelling of the leg may occur after revascularization, and this can cause increased compartmental pressure leading to muscle necrosis if untreated. If there is any suspicion of raised intracompartmental pressure, a generous fasciotomy should be done. Compartment syndromes are more common after reconstruction for vascular trauma than after embolectomy.

Other methods of treating emboli

Embolectomy is the best treatment for a white, bloodless limb, in which the circulation needs to restored without delay. In these cases there is little to support the claim that embolectomy can ‘do more harm than good', provided that the surgeon is prepared to proceed to on-table arteriography and more distal embolectomy or reconstruction if initial embolectomy is unsuccessful.

Thrombolytic therapy

Low-dose intra-arterial streptokinase, urokinase, or tissue plasminogen activator (known as TPA) can be effective in the treatment of embolism as for arterial thrombosis. This technique should only be used if the extremity is viable (sensation preserved). The infusion should be stopped after 48 h if there is no improvement, but if sequential radiographs show lysis then the infusion may be continued for up to 5 days. This method requires good radiological support.

Heparin therapy

Intravenous heparin alone can be used for the treatment of patients with acute limb ischaemia provided the limb is viable. This is safe management, especially when the facilities or expertise for vascular work are limited, or while awaiting a vascular opinion, or during transfer to a vascular unit. Heparin minimizes propagated clotting and improvement may occur through natural clot lysis and the development of collaterals. The eventual state of the circulation to the limb is likely to be less good after heparin therapy alone than after successful embolectomy.

Percutaneous aspiration thromboembolectomy

This technique employs a specially designed catheter sheath system that can be used alone or in combination with thrombolytic drugs. The catheter is used to aspirate fragments of thrombus from the distal arteries, particularly when lysis has failed to deal with them, or for a rapid result. It is also used to treat iatrogenic emboli resulting from intra-arterial catheterization or balloon angioplasty.

Long-term anticoagulation after embolism

Administration of intravenous heparin in the immediate postoperative period reduces early recurrence of emboli. Thereafter, long-term anticoagulation is traditionally instituted (warfarin by daily oral administration) in the hope of preventing further emboli. Oral anticoagulants certainly should be used in patients in atrial fibrillation, but for those with no obvious embolic source there is conflicting evidence of effect. Nevertheless, anticoagulation seems reasonable for all patients who will take warfarin reliably and from whom regular blood samples can be obtained for clotting studies. If haematological monitoring is difficult or if patients are unlikely to comply properly, then the risk of haemorrhagic complications may outweigh the benefits of long-term anticoagulation.

Late presentation of arterial emboli

Patients with emboli may present many days after the acute event. Embolectomy may be successful even after a delay of a month but is rarely successful thereafter because thrombus adheres to the vessel wall. Thrombolytic therapy may also be successful up to 1 month after impaction of an embolus.

Thrombolysis or intravenous heparin therapy are often used for those patients who present late but whose limbs have remained viable. If the obstruction is not relieved, then arterial reconstruction can be performed at a later date if required.

The mortality and morbidity of peripheral emboli

Patients suffering embolism are often elderly and infirm, and up to 30 per cent die in the postoperative period. Factors associated with a higher mortality are increasing age, recent myocardial infarction, proximal (aortoiliac) occlusions, poor cardiac and pulmonary function, and pre-existing arterial disease.

Amputation is an uncommon sequel to embolism and is associated with delayed presentation. Limb loss following acute ischaemia is more commonly associated with thrombosis in diseased arteries than with emboli.

Microemboli to the limbs

The small emboli that arise from arteriosclerotic arteries or aneurysms present either as isolated ischaemic digits (Fig. 7) or as ischaemic patches on an extremity. The main differential diagnoses are vasculitis and haematological disorders such as thrombocythaemia. ‘Trash foot' is an important complication of grafting for aortic aneurysm and describes the passage of a shower of loose thrombus into the small vessels of the feet (Fig. 8).



Fig. 7. An isolated ischaemic toe resulting from atheromatous embolism to the digital arteries.



Fig. 8. 'Trash foot': a shower of embolic material has been carried to the small arteries of this foot during operation for aortic aneurysm.

Such small distal emboli are impossible to remove surgically. During aortic grafting, prevention is the most important measure: gentle handling of the aneurysm, clamping the iliac arteries first, and flushing of blood both externally and, if possible, internally at the end of the procedure so that no thrombus or clots are allowed to pass into the limb arteries. When a localized arterial stenosis or aneurysm is identified as the cause for microembolism this should be dealt with appropriately, usually by bypass grafting, although balloon angioplasty is sometimes used (see Fig. 2 and Fig. 3).

Other sites affected by emboli

The brain and the eye

The brain is a frequent site of embolism, resulting in stroke. Small emboli passing up the internal carotid artery may also enter the retinal vessels, causing temporary or permanent blindness. If carotid atheroma is the source for recurrent emboli to the eye or brain, then carotid endarterectomy may be required, although many patients are treated with antithrombotic drugs such as aspirin.

Mesenteric emboli (see Chapter 17.10.2)

The superior mesenteric artery is usually affected, with clinical features of severe abdominal pain, sometimes vomiting or diarrhoea, and evidence of an embolic source (Table 1). Other features are a high white blood-cell count and elevated serum amylase. A plain abdominal radiograph will show the absence of normal small-bowel gas shadows centrally.

Urgent laparotomy is required and, if the bowel is still viable, embolectomy of the superior mesenteric artery can be performed. If there is doubt about bowel viability a 'second look' laparotomy should be performed 24 h later. In many cases the bowel is not viable and extensive resection is required. These patients require careful attention to replacement of fluid, electrolytes, and blood. If the whole small bowel has infarcted the patient will die.

Further reading

Campbell WB, Ballard PK. Intra-arterial lignocaine in embolectomy. *British Journal of Surgery* 1996; **83**: 244. [Analgesia during embolectomy.]

Campbell WB, Ridler BMF, Szymanska T. Current management of acute lower limb ischaemia: results of an audit by the Vascular Surgical Society of Great Britain and Ireland. *British Journal of Surgery* 1998; **95**: 1498–1503. [Causation and methods of management.]

Clason AE, Stonebridge PA, Duncan AJ, Nolan B, Jenkins AMcL, Ruckley CV. Morbidity and mortality in acute lower limb ischaemia: a 5-year review. *European Journal of Vascular Surgery* 1989; **3**: 339–43. [Mortality and limb salvage.]

Clavien PA. Diagnosis and management of mesenteric infarction. *British Journal of Surgery* 1990; **77**: 601–3. [Mesenteric emboli.]

Elliott JP, Hageman JH, Szilagyi DE, Ramakrishnan V, Bravo JJ, Smith RF. Arterial embolization: problems of source, multiplicity, recurrence, and delayed treatment. *Surgery* 1980; **88**: 833–45. [Recurrent limb embolism. Anticoagulants.]

Fogarty TJ, Cranley JJ, Krause RJ, Strasser ES, Hafner CD. A method for extraction of arterial emboli and thrombi. *Surgery, Gynecology and Obstetrics* 1963; **116**: 241–4. [Original description of balloon embolectomy catheter.]

Galbraith K, Collin J, Morris PJ, Wood RFM. Recent experience with arterial embolism of the limbs in a vascular unit. *Annals of the Royal College of Surgeons of England* 1985; **67**: 30–3. [Emboli at different sites in the limbs.]

Green, RM, DeWeese JA, Rob CG. Arterial embolectomy before and after the Fogarty catheter. *Surgery* 1975; **77**: 24–33. [Impact of Fogarty catheter embolectomy.]

Hickey NC, Crowson MC, Simms MH. Emergency arterial reconstruction for acute ischaemia. *British Journal of Surgery* 1990; **77**: 680–1. [Arterial reconstruction after embolectomy.]

Kazmier FJ. Shaggy aorta syndrome and disseminated atheromatous embolization. In: Bergan JJ, Yao JST, ed. *Aortic surgery*, pp. 189–94. Saunders, Philadelphia, 1989. [Embolization from the aorta.]

Levin BH, Giordano JM. Delayed arterial embolectomy. *Surgery, Gynecology and Obstetrics* 1982; **155**: 549–51. [Success of delayed embolectomy.]

Luther M, Alback A. Acute leg ischaemia: a case for the junior surgeon? *Annales de Chirugie Gynecologique* 1995; **84**: 373–8. [Need for experienced vascular surgeon.]

Panetta T, Thompson JE, Talkington CM, Garrett WV, Smith BL. Arterial embolectomy: a 34–year experience with 400 cases. *Surgical Clinics of North America* 1986; **66**: 339–53. [Sites of limb emboli.]

Parent FN, Bernhard VM, Pabst TS, McIntyre KE, Hunter GC, Malone JM. Fibrinolytic treatment of residual thrombus after catheter embolectomy for severe lower limb ischemia. *Journal of Vascular Surgery* 1989; **9**: 153–60.[Intraoperative.]

Petersen P, Godtfredsen J, Boysen G, Andersen ED, Andersen B. Placebo-controlled, randomised trial of warfarin and aspirin for prevention of thromboembolic complications in chronic atrial fibrillation. *Lancet* 1989; **i**: 175–9. [Use of warfarin after emboli.]

Scott DJA, Davies AH, Horrocks M. Risk factors in selected patients undergoing femoral embolectomy. *Annals of the Royal College of Surgeons of England* 1989; **71**: 229–32. [Indications for embolectomy, thrombolysis, amputation.]

Silvers LW, Royster TS, Mulcare RJ. Peripheral arterial emboli and factors in their recurrence rate. *Annals of Surgery* 1980; **192**: 232–6. [Recurrent embolism.]

Singh S, Ackroyd R, Lees T, Morris-Jones W, Beard JD. Thromboembolectomy and thrombolytic therapy in acute lower limb ischaemia: a five year experience. *International Angiology* 1996; **15**: 6–8. [Factors affecting outcome.]

Tawes RL *et al.* Arterial thromboembolism. A 20–year perspective. *Archives of Surgery* 1985; **120**: 595–9. [[Poor results of embolectomy for thrombosis. Use of heparin.]

Turnipseed WD *et al.* Percutaneous aspiration thromboembolectomy (PAT): an alternative to surgical balloon techniques for clot retrieval. *Journal of Vascular Surgery* 1986; **3**: 437–41. [Aspiration embolectomy.]

Walker WJ, Giddings AEB. A protocol for the safe treatment of acute lower limb ischaemia with intra-arterial streptokinase and surgery. *British Journal of Surgery* 1988; **75**: 1189–92. [Thrombolytic

treatment.]

Working Party on Thrombolysis in the Management of Limb Ischaemia. Thrombolysis in the management of limb arterial occlusion. Towards a consensus interim report. *Journal of Internal Medicine* 1996; **240**: 343–5. [Role of thrombolysis.]

17.10.2 Arterial emboli: mesenteric arteries

Leslie W. Ottinger

Anatomic considerations
Clinical presentation
Management
Further reading

Mesenteric ischemia and infarction results from interruption of arterial or venous perfusion of all or part of the intestinal tract. The major causes are acute occlusion of the superior mesenteric artery, mesenteric venous thrombosis, and nonocclusive infarction. In general, the prognosis for survival is poor.

Acute occlusion of the superior mesenteric artery may be the result of either thrombosis or an embolus; emboli account for about 20 per cent of all cases of acute ischemia or infarction and about one-half of the cases of acute occlusion. With early diagnosis and proper management, survival with preservation of most or all of the small intestine is readily achieved.

Anatomic considerations

Three vessels constitute the major arterial supply to the intra-abdominal intestine: (1) the celiac axis for the stomach and duodenum, (2) the superior mesenteric artery for the jejunum, ileum, and right and transverse colon, and (3) the inferior mesenteric artery for the left and sigmoid colon. It is unusual for an embolus to either the celiac axis or the inferior mesenteric artery to produce infarction, because of the obtuse angle of the celiac axis to the aorta, the small size of the inferior mesenteric artery, and the adequacy of collateral inflow to both. The superior mesenteric artery, in contrast, originates at an acute angle, is large and, at least when acute occlusion is proximal, seldom has the sufficient collateral flow to preserve viability of the entire area of intestine that it supplies. This artery is one of the most frequent sites of clinically apparent emboli to visceral vessels.

The first branches of the superior mesenteric artery are small, supplying the duodenum, pancreas, and proximal jejunum. The initial branch of any size is the middle colic artery. The superior mesenteric artery tapers rather rapidly distal to it, terminating in the arteries to the distal ileum. The usual sites at which an embolus lodges are at the origin of the middle colic artery, where it may or may not cause occlusion, and in the main trunk of the artery, within 5 cm distal to the origin of the middle colic artery (Fig. 1). Occlusion at the origin of the superior mesenteric artery is almost invariably the result of thrombosis not an embolus.

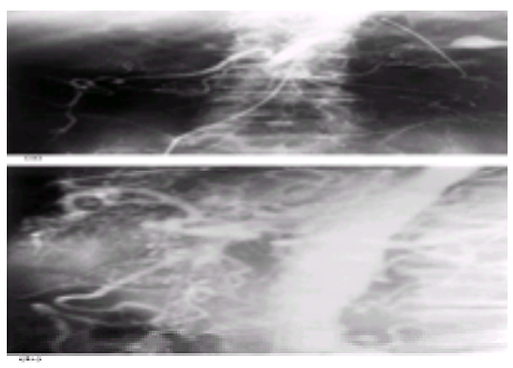


Fig. 1. (a) Selective superior mesenteric artery angiogram demonstrating occlusion of the artery by an embolus distal to the inferior pancreatoduodenal and middle colic arteries; (b) lateral view in the same patient.

As with other peripheral emboli, most mesenteric emboli come from the heart. Common sources of emboli are atrial fibrillation, infarction of the ventricle with mural thrombus, and valvular excrescences. A few emboli come from the aortic wall itself, and some are iatrogenic, following aortic surgery or catheterization.

When occlusion of the superior mesenteric artery occurs gradually, collateral inflow from the other visceral vessels usually protects the intestine from ischemic injury. When occlusion is acute this collateral supply is not sufficient. Depending on the site at which an embolus lodges, the resulting ischemic injury may extend from the proximal jejunum to the left transverse colon or, with a more peripheral site of occlusion, may involve only the ileum and right colon. Anatomic variation in collaterals and the presence of fragments of emboli peripherally mean that the ischemic injury may be patchy and variable in its severity, especially in the early stages. However, the central area of supply, the ileum and cecum, usually sustains the most profound injury. A small embolus that enters a branch of the superior mesenteric artery may lead to a segmental infarct or produce no ischemic injury.

The mucosa is the only layer of the intestinal wall to have little resistance to ischemic injury, becoming damaged rapidly by even relatively mild degrees of ischemia. The first gross evidence of damage is submucosal edema; this proceeds to hemorrhage and sloughing of the mucosa, with a small amount of intraluminal bleeding. If the deeper layers remain viable, or if arterial perfusion is restored, the mucosa regenerates: this may take several weeks, during which there is usually diarrhea and sometimes a degree of malabsorption. In a few patients the presence of deep ulcers leads to major hemorrhage from regenerating areas.

The initial response of the muscularis propria to ischemia is spasm. Depending on the severity and duration of ischemia, atony, infarction, and perforation can follow. Full-thickness, irreversible infarction produces a characteristic, foul-smelling, bloody peritoneal exudate. With restoration of arterial perfusion the viability of the outer layers may be preserved and chronic ulceration may lead to the development of strictures, which may take several weeks to become apparent.

Visceral pain is an invariable symptom of mesenteric ischemia. Only in the obtunded or otherwise mentally impaired patient will it not suggest the diagnosis. In its absence, the diagnosis can be confidently excluded. Other symptoms and signs are less definite and depend on the extent, severity, and location of the ischemic injury.

The course of ischemic injury caused by an embolus is also variable. Short of removal of the embolus, the only factor likely to improve perfusion is remission of arterial spasm and the gradual enlargement of collateral vessels. Conversely, the ischemic injury may be worsened by factors that further decrease perfusion, including hypoperfusion due to systemic hypotension and spasm in visceral arteries. Local factors, including edema, hemorrhage, and distension, can also cause extension of the infarct. Although an embolus in the superior mesenteric artery can lead to immediate infarction and perforation within a few hours, this is not the usual course. Many patients will have symptoms for hours or even days before the development of irreversible infarction.

The collateral flow tends initially to lead to a gradient of severity of ischemic injury, the peripheral areas being most spared. There is a tendency for areas of full-thickness infarction to extend to include the marginal areas with the passage of hours. Eventually the initial patchy involvement will usually become an extended area of full-thickness infarction.

Clinical presentation

Patients with mesenteric emboli almost always present with abdominal pain, which tends to be steady and, being visceral in origin, poorly localized and referred generally to the midabdomen. Although traditionally thought to be sudden in onset and remarkably severe, this is not always true. There may be a vague onset and variable intensity, even with intermittent disappearance. Pain may have been present for several hours or even a few days.

The history may include vomiting, diarrhea, or even tenesmus. Since signs of peritonitis are present only after perforation, there are initially no positive findings on physical examination. The gastric or rectal contents may contain gross or occult blood. At later stages, although the findings are nonspecific, an intra-abdominal catastrophe is obvious, as is the need for laparotomy.

Diagnosis is most elusive during the early stages, when the chance for successful management is high. The decision to proceed with either angiography or laparotomy

must be made on the basis of suspicion, not certainty. The diagnosis must be seriously entertained in any patient with unexplained abdominal pain, especially in those over 60–years old.

Laboratory tests are not so specific as to be useful in a positive way. Many patients, but not all, have a leucocytosis, which may be extreme, counts exceeding 20 000/mm^3. The serum amylase is slightly elevated in 50 per cent of victims. Plain radiographs show no, or only nonspecific, findings in most cases.

Arteriography is especially helpful in the patient with either an embolus or thrombosis of the superior mesenteric artery, but is less so in those with venous or non-occlusive infarction, though it may still yield useful indirect findings. The major application of arteriography is in the patient with a history suggesting an embolus but without systemic or local findings to herald the later stages of infarction. In these patients, arteriography establishes the diagnosis, leading at once to an expeditious laparotomy. When the embolus is small and in a branch, arteriography may also eliminate the need for surgery, although this is quite an unusual site of embolism. As well as providing a certain diagnosis, the arteriogram provides the surgeon with useful information about the site of occlusion and patency of other visceral vessels.

Other radiographic studies, such as computed tomography and indirect flow studies using a Doppler device, show promise in helping to reach a diagnosis. They have not reached a level of development and application where they are generally useful or available.

Management

Even with the availability of angiographic therapeutic techniques, immediate laparotomy is the best approach for the management of the embolus and injured bowel. The ideal operation consists of an embolectomy followed by resection of any segment of nonviable bowel.

The surgeon must first confirm the diagnosis, then make an estimate of the extent and severity of ischemic injury. In the early stages there may be no gross evidence of bowel injury, and there may even be faint peripheral pulses from collateral inflow. The diagnosis of embolism in the superior mesenteric artery should be abandoned only after the presence of a strong pulse in the main trunk is confirmed. In these early cases an embolectomy can allow preservation of all of the intestine. Failure to perform an embolectomy invariably leads to infarction and a poor prognosis.

Many patients have extensive infarction that is clearly irreversible. Experience and judgement are needed to decide whether a resection of most of the small bowel, with or without an embolectomy, should be performed. This decision should reflect the age and general condition of the patient, and the amount of viable jejunum. Parenteral nutrition solutions have made a major contribution to management of patients with very extensive resections, and permanent home hyperalimentation is a measure to be considered. A successful outcome can be achieved in carefully selected patients with extensive infarction.

If the extent of injury is limited, decisions about resection or embolectomy must take into account the site of the embolus. Here, arteriography is helpful (Fig. 2). In a few cases the embolus is in a peripheral vessel and perfusion to adjacent segments of bowel is normal: resection alone is all that is needed. A delayed reanastomosis may be selected for the colon. If the embolus is located centrally, even though the extent of the infarction is limited, resection alone is almost certain to fail. Further infarction can be anticipated and an embolectomy is indicated.

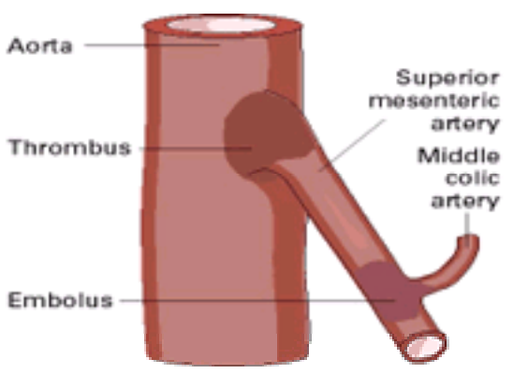


Fig. 2. The usual sites of occlusion with thrombosis and embolization in the superior mesenteric artery.

The technical demands of embolectomy of the superior mesenteric artery are within the skills of the general surgeon. The approach to the vessel is at the base of the mesocolon, lifting the transverse colon in an anterior superior direction. The vein is usually to the right of the artery and most of its left branches pass posterior to it (Fig. 3). The perivascular nerve plexus makes exposure somewhat more tedious than that for an artery in the extremities. This approach will lead to exposure just distal to the middle colic artery. A longitudinal arteriotomy is easier to close, especially in an atherosclerotic vessel, and can also serve as a site for bypass anastomosis if the diagnosis proves to be thrombosis at the origin of the superior mesenteric artery. In the presence of an embolus the pulse will be noted to end at or just beyond the middle colic artery. Embolectomy catheters can be used if the embolus is proximal to the arteriotomy and if there is distal propagation of thrombosis. The mesenteric vessels are fragile and catheters should be used with care. The arteriotomy can usually be closed directly; eversion is neither necessary nor desirable. Rarely, a local endarterectomy or patch graft may be required.

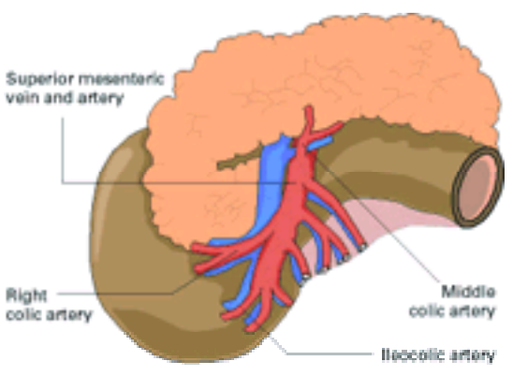


Fig. 3. The superior mesenteric vessels showing the relation of the vein and its branches to the artery.

After mesenteric flow has been restored, a period of observation is needed to determine which ischemic segments of bowel are viable. Although fluorescein injections with observation under an ultraviolet light and determination of pulsatile flow in the bowel wall by a Doppler device make objective assessment possible, these are usually unavailable and observation by an experienced surgeon carries the same degree of accuracy. If no doubt exists, non-viable segments should be resected. Depending on the apparent adequacy of arterial circulation, anastomosis may be performed, or stomas should be established.

When doubt continues after observation, clearly infarcted segments should be extirpated and a second-look operation scheduled. The second operation can be done at a convenient time, a few hours later, when there should be a clear demarcation of infarcted areas. The decision about reoperation should be made during the first operation, as the clinical course over the next few hours cannot serve as an indication that all is well. There is evidence that supports a policy of routine second-look operations after embolectomy.

Postoperative measures to prevent recurrence or extension of thrombus may be appropriate; these include the infusion of antispasm drugs, using angiographic techniques. More important is attention to the management of the cardiovascular system: hypovolemia and peripheral vasospasm must be avoided. Anticoagulants prevent the formation of further emboli but must be used with caution for the first 24 h in the patients with a suture line in the mesenteric artery.

A number of specific postembolectomy complications can be described, all of which are relatively uncommon. Occlusion of the artery by thrombus or a new embolus will cause the original symptoms to recur and this is suggested in the patient who notes recurrence of pain. Early bleeding from the suture line follows anticoagulation;

late bleeding usually means sepsis. Diarrhea, often severe, is common and parenteral nutrition can make an important contribution to survival. Diarrhea subsides as the mucosal injury heals: persistence is indicative of extensive ulceration and stricture formation and a further resection will be needed. During the healing phase, mild bleeding is usual. Massive bleeding from deep ulcers may indicate the need for a resection. Angiography is often helpful in detecting the site of bleeding.

Further reading

Batellier J, Kieny R. Superior mesenteric artery embolism: eighty-two cases. *Annals of Vascular Surgery* 1990; **4**: 112–16.

Bech FR, Gewertz BL. Pathophysiology of intestinal ischemia. In: Strandness DE, Jr, van Breda A, ed. *Vascular diseases: surgical and interventional therapy*, pp. 745–53. Churchill Livingstone, New York, 1994.

Bradbury AW, Brittenden J, McBride K, Ruckley CV. Mesenteric ischaemia: a multidisciplinary approach. *British Journal of Surgery* 1995; **82**(11): 1446–59.

Geelkerken RH, van Bockel JH. Mesenteric vascular disease: a review of diagnostic methods and therapies. *Cardiovascular Surgery* 1995; **3**: 247–60.

Jamieson WG. Acute intestinal ischemia. *Canadian Journal of Surgery* 1988; **31**: 157–8.

Kaleya RN, Boley SJ. Acute mesenteric ischemia. *Critical Care Clinics* 1995; **11**: 479–512.

Kaufman SL, Harrington DP, Sigelman SS. Superior mesenteric artery embolization: an angiographic emergency. *Radiology* 1977; **124**: 625–30.

McKinsey JF, Gewertz BL. Acute mesenteric ischemia. *Surgical Clinics of North America* 1997; **77**: 307–18.

Ottinger LW. The surgical management of acute occlusion of the superior mesenteric artery. *Annals of Surgery* 1978; **188**: 721–31.

Schneider TA, Longo WE, Ure T, Vernava AM, IIIrd. Mesenteric ischemia. Acute arterial syndromes. *Diseases of the Colon and Rectum* 1994; **37**: 1163–74.

Shaw RS, Rutledge RH. Superior mesenteric embolectomy in the treatment of massive mesenteric infarction. *New England Journal of Medicine* 1957; **257**: 595–8.

Sitses-Serri A, Mas X, Rouquita F, Figueres J, Saus F. Mesenteric infarction: an analysis of 83 patients with prognostic studies in 44 cases undergoing a massive small bowel resection. *British Journal of Surgery* 1988; **75**: 544–8.

17.11 Arterial and venous injuries

David V. Feliciano

Mechanism of injury
Location
Types of injuries
Clinical presentation
Diagnosis
Radiologic studies in diagnosis
Nonoperative management
Endovascular stents
General management of vascular injuries
Incisions
Amputation rather than repair
Options for vascular repair
Adjuncts to vascular repair
Carotid artery
Vertebral artery
Thoracic aorta
Innominate, intrathoracic common carotid, and subclavian arteries
Abdominal vascular trauma – midline or zone 1
Abdominal vascular trauma – upper lateral retroperitoneum or zone 2
Abdominal vascular trauma – pelvic retroperitoneum or zone 3
Abdominal vascular trauma – portal and retrohepatic area
Peripheral vascular injuries
Further reading

The extensive research of Alexis Carrel and Charles C. Guthrie at the Johns Hopkins Hospital from 1904–1906 led to the development of vascular operative techniques including repair of the lateral arterial wall, end-to-end anastomosis, and insertion of venous conduits. These were first applied clinically by V. Soubbotitch during the Balkan Wars from 1912–1913, by American, British, and German surgeons in the First World War, and by Ramuald Weglowski during the Polish–Russian war of 1920. The modern experience with vascular repair including suture repair and non-suture insertion of vein grafts or vitallium tubes started during the Second World War and was refined considerably during the Korean and Vietnam Wars. For the past 40 years civilian trauma centers in urban areas, particularly in North and South America, have treated extraordinary numbers of patients with vascular injuries (Table 1).

Neck	694
Thoracic arterial	815
Thoracic venous	153
Thoracic miscellaneous	191
Abdomen	1947
Upper extremity	859
Lower extremity	1102
Total	5760

*Adapted from Mattox KL, Feliciano DV, Burch J, et al. Annals of Surgery 1989 ;209: 698–707.

Table 1 Vascular and cardiac injuries, City of Houston public hospitals, 1958–1987*

Mechanism of injury

The most common cause of vascular injuries in urban trauma centers in the United States is a low velocity missile wound from a handgun. In countries in which firearms are more difficult to obtain, stab wounds account for a significant percentage of penetrating vascular injuries. One other penetrating wound seen less frequently in recent years has been an infected arterial false aneurysm secondary to a missed intravenous injection of illicit drugs.

Vascular injuries from blunt trauma such as fractures, dislocations, contusions, compression, and traction are uncommon. Approximately 15 per cent of victims dying in motor vehicle crashes prior to the introduction of air bags were found to have rupture of the descending thoracic aorta. With injuries in the extremities, the need for surgical repair of an adjacent injured artery has been less than 1 per cent for fractures of the femur and less than 2 per cent for fractures of the tibia.

The increased use of invasive diagnostic and therapeutic techniques by nonsurgeons has caused an increase in in-hospital vascular injuries that also require repair.

Location

The superficial femoral artery and vein in the lower extremity and the brachial artery and vein in the upper extremity are the most commonly injured vessels in both civilian and military series. This is a reflection of the length of these vessels and the fact that direct compression will control hemorrhage and prevent exsanguination prior to arrival in the hospital.

Types of injuries

Defects in the wall, complete transections and arteriovenous fistulae are most commonly caused by penetrating wounds. Contusions or traction injuries causing intimal flaps or disruptions are usually associated with blunt trauma. Spasm may occur after either penetrating or blunt trauma.

Clinical presentation

If the patient with a defect (puncture, laceration, or through-and-through injury) in the vessel wall is thin or the injured vessel has a direct communication with the wound on the skin or with a body cavity, external or internal bleeding and secondary hypovolemic shock occurs. In other patients, wall defects lead to acute hematomas that are tamponaded by soft tissues in the extremities or the lining of body cavities. When an artery is the injured vascular structure, the pulsatile hematoma (early traumatic false aneurysm) may expand rapidly prior to operative repair (Fig. 1).

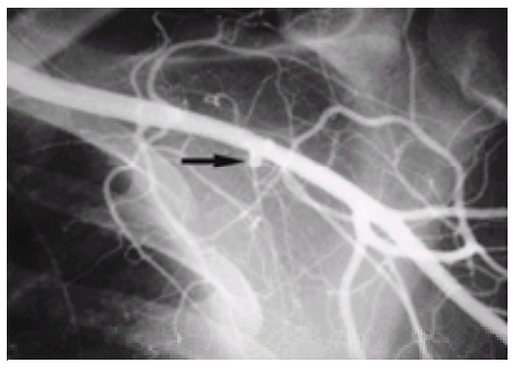


Fig. 1. Preoperative arteriogram demonstrates small pseudoaneurysm of second portion of left axillary artery.

Complete transection with thrombosis of the main artery of an extremity will often lead to the six ‘Ps' – pulselessness, pallor, paresthesias, pain, paralysis, and poikilothermia. If a truncal artery is involved, visceral ischemia or infarction will result. Transections of larger vessels will almost always lead to hemorrhage rather than thrombosis.

An arteriovenous fistula in an extremity will cause a palpable thrill over the area of injury and a continuous audible bruit or murmur through a stethoscope. Sudden compression of a large arteriovenous fistula may cause slowing of the pulse rate (Branham's sign).

Partial intimal flaps will occasionally lead to embolization of aggregated platelets unless anticoagulation is administered. A circumferential intimal flap in a small vessel of an extremity will often thrombose in the first 12 h after injury. If there is circumferential disruption of both intima and media, as in certain deceleration injuries of the descending thoracic aorta or innominate artery, an acute traumatic true aneurysm or ballooning of the remaining adventitial layer may occur.

Diminished or absent distal flow may occur when arterial spasm is present, but the viability of an extremity is usually not compromised unless extensive injuries to arterial collaterals are also present.

Diagnosis

A prehospital history of significant external hemorrhage or hypotension in association with a penetrating truncal wound is pathognomonic of injury to a named vessel or extensive injury to soft tissue or solid viscera. With a history of cervical hyperextension/rotation, hyperflexion, or a direct blow, the presence of a Glasgow Coma Scale score of 6 or less, diffuse axonal injury, fracture of the petrous bone, or a LeFort II or III fracture on in-hospital evaluation, further work-up is needed to rule out a blunt cerebrovascular injury. A history of significant deceleration in a frontal crash or a major side impact mandates some type of evaluation to rule out an injury to the descending thoracic aorta. A significant compression injury to the abdomen or flank in a patient who has or develops guaiac-positive stools or hematuria during in-hospital evaluation suggests that a blunt injury to the superior mesenteric or renal artery is present.

Physical examination is most helpful when hemorrhage, a rapidly expanding hematoma, arterial occlusion, signs of an arteriovenous fistula, or proximity of injury to the major vessels of an extremity are present. Patients with any of these findings except proximity of injury need immediate operation or a radiographic study to localize the site of injury prior to operation.

Physical examination alone has proven to be remarkably accurate in determining the need for operation in patients with penetrating wounds of the extremities. In one study of 310 patients with 366 such wounds, only two patients had missed vascular injuries even though arteriograms were not performed.

Radiologic studies in diagnosis

With the large number of diagnostic studies currently available (Table 2), the choice in a given institution depends on local interest and expertise. In asymptomatic or modestly symptomatic patients with penetrating wounds of the neck in proximity to the carotid artery, color flow Doppler or duplex ultrasound are noninvasive techniques with excellent accuracy. When there is concern about a cerebrovascular injury from blunt trauma, preliminary data suggest that CT angiography may have an acceptable accuracy in diagnosing the small intimal defects that are common. Selective carotid arteriography, however, remains the standard of care for detection of injuries to the carotid or vertebral arteries in most institutions (Fig. 2).

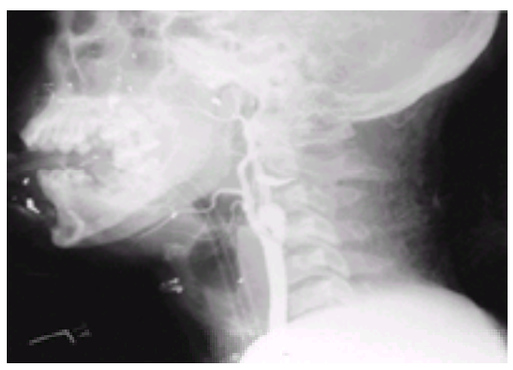


Fig. 2. False aneurysm of left internal carotid artery in a patient who sustained a gunshot wound to the left neck.

Cerebrovascular
Color flow Doppler
Duplex ultrasound
CT angiography
Selective carotid arteriography
Thoracic vascular
Spiral CT
Transesophageal echocardiography
Digital subtraction aortography
Standard aortography
Peripheral vascular
Color flow Doppler
Duplex ultrasound
Emergency center or operating room arteriography by surgeon
Digital subtraction arteriography
Standard arteriography

Table 2 Currently available radiologic techniques for evaluating possible vascular injuries

Contrast-enhanced spiral CT as part of the truncal diagnostic work-up in patients with multisystem blunt trauma has proven to be extremely accurate in detecting deceleration injuries to the descending thoracic aorta. In some centers this technique is used as a screening tool in patients thought to be at ‘low risk' for this injury. Other centers use spiral CT as a definitive diagnostic study and proceed to thoracotomy without confirmation by a thoracic aortogram. Transesophageal echocardiography has proven to be accurate in detecting injuries to the descending thoracic aorta, as well. Limitations to its use include the fact that it is operator-dependent, any contraindication to esophageal intubation, and less than ideal visualization of the ascending aorta behind the right mainstem bronchus, branches of the transverse arch, and the descending thoracic aorta when a pneumothorax is present. In centers that use aortography for diagnosis, digital subtraction

techniques have replaced standard aortography (Fig. 3).

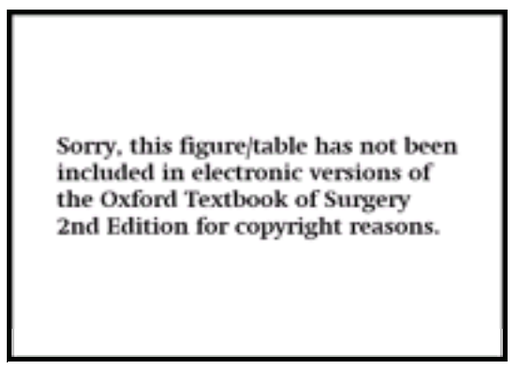


Fig. 3. Aortogram of hemodynamically stable patient with through-and-through gunshot wound to the descending thoracic aorta. The arrow demonstrates extravasation at the more proximal injury.

All the techniques listed in Table 2 are accurate in evaluating patients with possible peripheral vascular injuries (Fig. 4). Surgeon-performed arteriograms (one- or two-shot) in the emergency center or operating room are technically easy and extremely valuable in making rapid diagnoses. Also, they avoid the inevitable delays associated with the return of diagnostic angiography teams to the hospital at night.



Fig. 4. Shotgun wound to right leg with occlusion of the tibioperoneal trunk.

Nonoperative management

With the recognition that arterial injuries caused by interventional diagnostic and therapeutic techniques usually healed without operative repair, many such injuries caused by external trauma are now treated nonoperatively. Injuries such as spasm, small intimal defects (<5 mm), subintimal or intramural hematomas, and even small pseudoaneurysms (<5 mm) with intact distal flow detected on arteriography have been managed nonoperatively for over 10 years. Coumadin or aspirin is indicated in patients with such lesions in the carotid or vertebral arteries in the absence of an associated injury to the brain. The role of these agents during nonoperative management of truncal or peripheral arterial injuries is not clear.

Intimal defects, intimal–medial defects without extravasation (traumatic true aneurysm) and small traumatic false aneurysms in the descending thoracic aorta are now also being managed nonoperatively. Nonoperative management was originally performed in patients in whom a meaningful recovery was doubtful because of associated injuries. Delayed operative or nonoperative management is now routinely used in patients with associated injuries and hypothermia, metabolic acidosis, or a coagulopathy precluding a safe operation. All patients undergoing delayed operative or nonoperative management are maintained on esmolol, labetalol, or metoprolol to maintain mean arterial pressure at 60 to 70 mmHg and decrease shear stress on the area of injury. Failure of a b-blocker to lower mean arterial pressure mandates the addition of further antihypertensive therapy. Follow-up aortography is appropriate to document healing or worsening of the tear in the descending thoracic aorta.

Endovascular stents

There are now preliminary results on the use of endovascular stents placed by interventional surgeons or radiologists as definitive treatment for traumatic true or false aneurysms and arteriovenous fistulae. Most have been inserted in arterial injuries in which advanced techniques of exposure are ordinarily necessary to complete surgical repair. Examples would be intimal or tamponaded injuries to the subclavian or high internal carotid artery. Without long-term follow-up, the use of such stents should be limited to defined protocols at this time. When more readily accessible arterial injuries are present, operative repair remains the standard of care. Of interest, endovascular stents have now been inserted in a small number of patients with blunt rupture of the descending thoracic aorta. Several patients have been followed for over 18 months with satisfactory results.

General management of vascular injuries

In the patient with rapid external bleeding from a peripheral vascular injury, a tight compression dressing over the wound or the application of a proximal blood pressure cuff will usually control hemorrhage. Gloved-finger compression in the entrance and exit sites of a missile or stab wound controls hemorrhage in the operating room until an orthopaedic-type operative tourniquet (for distal wounds) is applied or until the surgical team is ready to operate.

It is mandatory that preparation of the skin and draping should encompass all potential areas of proximal and distal vascular control (Table 3). Also, one or both lower extremities should be draped to the toes to allow for retrieval of the greater saphenous vein from either the groin or the ankle. Draping out the hand or foot of an extremity with an arterial injury in a sterile plastic bag is useful so that color changes can be noted or distal pulses palpated under sterile conditions after completion of the repair. The remainder of the extremity including the area of incision is then covered with an orthopaedic stockinette.

Location of injury	Recommendation
Neck	Mandible to umbilicus; one lower extremity to popliteal area
Thorax	Chin to umbilicus; ipsilateral upper extremity to fingernails; one lower extremity to popliteal area
Abdomen	Chin to both popliteal areas
Upper extremity	Chin to umbilicus from opposite nipple to ipsilateral fingernails; one lower extremity to toenails
Lower extremity	Nipples to bilateral toenails

Table 3 Recommended skin preparation and draping for possible vascular injuries

Incisions

It is appropriate for inexperienced surgeons to use the most extensive incisions to obtain proximal and distal vascular control. For example, an experienced surgeon is usually able to repair an injury to the second or third portion of the subclavian artery through a supraclavicular incision. The inexperienced surgeon might add a median sternotomy or excision of the mid-clavicle on the right or a high left anterolateral thoracotomy and excision of the mid-clavicle on the left (Fig. 5 and Fig. 6).

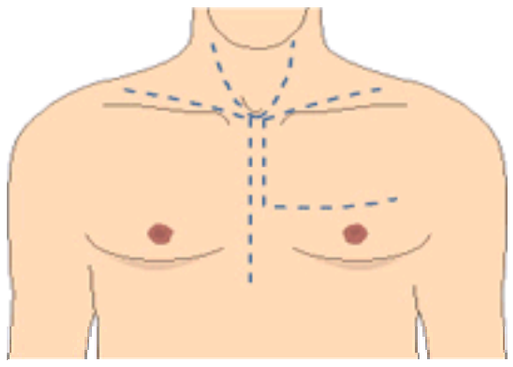


Fig. 5. Cervicothoracic incisions for repair of injuries to the innominate artery or vein, common carotid arteries, or subclavian arteries or veins. (Copyright, Baylor College of Medicine, 1980. Used with permission.)

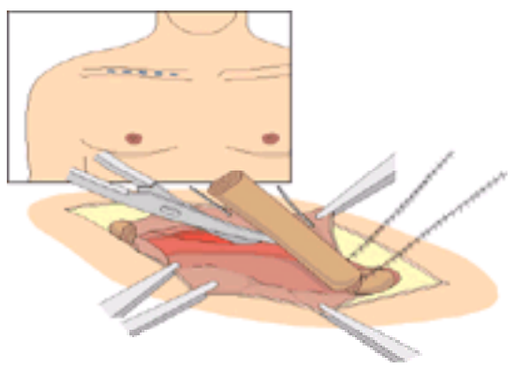


Fig. 6. Excision of the middle two-thirds of the clavicle will improve exposure of injuries to the subclavian vessels. (Copyright, Baylor College of Medicine, 1980 and used with permission.)

Amputation rather than repair

A number of scoring systems are available to aid the surgeon in determining the need for immediate amputation of an extremity with extensive arterial and other injuries. In general, absolute criteria for immediate amputation include the following: (1) patient's life at risk from multiple injuries including those in the mangled extremity; (2) disruption of brachial plexus in the upper extremity or posterior tibial nerve below the knee in the lower extremity; (3) crush injury with cold ischemia (no collateral flow) of duration more than 6 h.

The decision to amputate should be made by attending surgeons in general surgery, orthopaedic surgery, and plastic surgery who have examined the extremity in the operating room. Prior to amputation, operative photographs should be taken to avoid late medicolegal concerns. Also, the senior surgeon should contact the family, if possible, and explain the need for amputation. When there is significant opposition to amputation, a member of the family should be brought to the operating room to view the magnitude of injury.

Options for vascular repair

The choice of one of the options listed in Table 4 depends on the patient's hemodynamic status and magnitude of associated injuries, the extent of injuries to structures around the vascular injury, the level of external contamination or gastrointestinal soiling, and the surgeon's experience.

Lateral arteriorrhaphy or venorrhaphy
Patch angioplasty
Resection with end-to-end anastomosis
Resection with insertion of interposition graft
Bypass graft
Extra-anatomic bypass graft
Temporary intraluminal shunt
Ligation

Table 4 Options for vascular repair

Lateral arteriorrhaphy or venorrhaphy with 5–0 or 6–0 polypropylene suture is used for small lacerations or puncture, pellet, or missile wounds in the smaller vessels of the extremities. Patch angioplasty is infrequently used, but can be performed to avoid graft placement in larger arteries or veins of the proximal extremities or in the trunk.

Resection of the area of injury with an end-to-end anastomosis or insertion of an interposition graft is required in over two-thirds of peripheral arterial injuries. Over-enthusiastic ligation of multiple collateral vessels to complete an end-to-end anastomosis under tension will result in an hourglass appearance at the suture line and possible thrombosis of the repair. An interrupted suture technique is appropriate in children or in any adult vessel with a diameter as small as 4 to 5 mm. When an interposition graft is to be inserted, a reversed autogenous saphenous vein graft from an uninjured lower extremity is the conduit of choice. Polytetrafluoroethylene grafts more than 6 mm in diameter have been used with success in a variety of cervical, truncal, and peripheral locations (Fig. 7). Indications include a saphenous vein that is absent, injured, too small, or of an inappropriate size or when the patient is critically injured and the speed of repair is important. Either polytetrafluoroethylene or Dacron grafts are used for replacement of extensive injuries to the abdominal aorta or iliac artery.

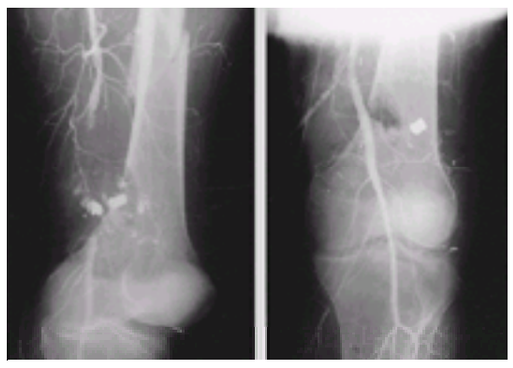


Fig. 7. Gunshot wound of the proximal popliteal artery (left panel) was repaired with insertion of a 6-mm polytetrafluoroethylene graft (right panel). The popliteal vein was repaired, also, and completion arteriography and a below knee four-compartment fasciotomy were performed.

Bypass grafting has primarily been used from the abdominal aorta to the mid-superior mesenteric artery after proximal ligation of this vessel. Extraanatomic bypass grafting is useful when extensive injuries to soft tissues in the antecubital or groin areas are accompanied by injuries to the brachial or femoral arteries (Fig. 8). An extra-anatomic bypass through noninjured tissue around the open wound allows for frequent gauze pack changes in the wound in soft tissue and eliminates the risk of blowout of exposed vascular suture lines.

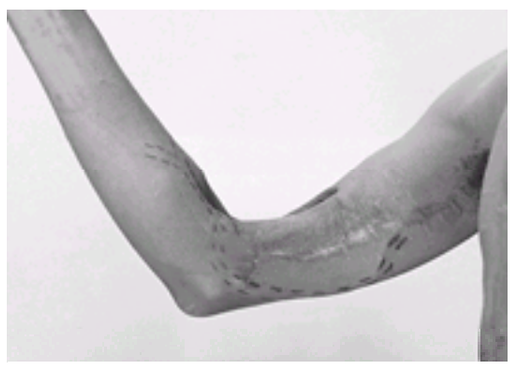


Fig. 8. Blowout of an exposed prosthetic graft–brachial artery anastomosis on the fourth postoperative day prompted reoperation with insertion of an extra-anatomic saphenous vein graft (the path of the graft is marked on the skin). Vigorous wound care and early application of a split-thickness skin graft to the antecubital fossa led to an excellent final result.

The insertion of a temporary intraluminal shunt is appropriate when a rapid ‘damage control' operation in the patient with hypothermia, acidosis, and a coagulopathy is to be performed. Patients with arterial occlusions in association with open fractures (mangled extremities) and those undergoing replants of amputated upper extremities also benefit from the use of temporary shunts.

Ligation is reserved for small distal arteries anywhere in the body, noncritical main arteries below the elbow or knee, and the profunda femoris artery in the thigh. Venous ligation is acceptable in exsanguinating patients unless there is injury to the superior vena cava, the inferior vena cava above the renal veins, the popliteal vein, or the superficial femoral vein in the distal thigh. With ligation of either of the last two veins, an immediate below knee four-compartment fasciotomy is mandatory.

Adjuncts to vascular repair

Completion arteriography using a 20-gauge Teflon-over-metal catheter and 35 ml of meglumine diatrizoate dye is appropriate whenever an end-to-end anastomosis or insertion of an interposition graft has been performed in the lower extremity. As *in situ* distal arterial thrombosis is rare in the upper extremity, return of palpable pulses at the wrist following vascular repair eliminates the need for completion arteriography.

Debridement of contused or devascularized tissues is usually performed after restoration of arterial inflow in the extremities. Irrigation with saline solution containing antibiotics will also lower the incidence of postoperative soft tissue infection around the vascular repair.

Drains are rarely inserted after vascular repair, but may be useful in selected circumstances. Excessive oozing despite reversal of heparinization after a repair of the cervical carotid artery may occasionally warrant the insertion of a closed suction drain away from any vascular suture line for 8 to 12 h. With large blast cavities in the soft tissues of the groin or thigh, muscle should be rotated to separate any vascular repair from the cavity. The cavity is then drained with a closed suction device placed far from the vascular repair.

Massive edema or contamination of soft tissues of an extremity above a vascular repair may preclude immediate closure. A useful technique to prevent contamination and desiccation of exposed vascular repairs is to use a cover of porcine xenograft under antibiotic-soaked gauze. Irrigation of the incision or wound and formal closure by the general or plastic surgeon is then appropriate at 24 to 48 h.

As previously noted, below knee (or elbow) fasciotomy is appropriate whenever a major proximal venous ligation has been performed. Whenever there is a question about the need for fasciotomy, the presence of a measured intracompartmental pressure of more than 30 to 35 mmHg confirms that a fasciotomy is indicated. Only 20 per cent of fasciotomies are performed in the upper extremities, where the extensor and both the superficial and deep volar compartments should be opened. Elevation of the extremity in which a fasciotomy has been performed for 5 to 7 days will allow for closure of the skin incisions in approximately one-third of patients.

Adjuncts for preservation of a ‘borderline' extremity after vascular repair include sympathectomy, arterial infusion with a mixture of heparin, tolazoline, and saline at 30 ml/h, and venous infusion with low molecular weight dextran at 40 ml/h.

‘Damage control' operations for shocked patients with thoracic, abdominal, or peripheral vascular injuries are generally completed using an alternate form of closure of the incision such as towel clip or suture closure of the skin only or the application of an Esmarch bandage or plastic intravenous bag as a temporary silo.

Carotid artery

Current areas of controversy in patients with injuries to the carotid arteries include management with preoperative neurological deficits after penetrating trauma, approach to injuries in zone III (angle of mandible to base of skull) and blunt carotid injuries. In a patient arriving shortly after suffering a gunshot or stab wound to the carotid artery, a neurological deficit may be secondary to hypovolemic shock, cerebral ischemia, or a combination of the two. The only hope for meaningful neurological recovery in patients who do not have coma is to control hemorrhage and restore prograde flow in the carotid artery. Patients with preoperative coma, however, have a poor outcome (persistent coma or death) 75 per cent of the time whether restoration of flow or ligation of the carotid artery is performed.

With minor injuries such as an intimal flap or small pseudoaneurysm on arteriography of the internal carotid artery in zone III, observation, administration of anticoagulants, and, in some centers, insertion of an endovascular stent is performed. With active external hemorrhage or the presence of a pulsatile hematoma, proximal control of the internal carotid artery is obtained in zone II (sternal notch to angle of mandible) of the neck. Exposure of the injury itself is obtained by having the oral surgery service perform a vertical ramus mandibulotomy or temporary subluxation of the temporomandibular joint. Persistent hemorrhage as exposure is being obtained mandates the passage of a Fogarty balloon catheter superiorly in the internal carotid artery.

Intimal lesions or even occlusion of the internal carotid artery from blunt trauma is treated with heparinization in the absence of injury to the brain or multiple associated injuries. With a contraindication to heparinization or the presence of a significant lesion such as a pseudoaneurysm in zone III, the insertion of an endovascular stent is

appropriate.

Vertebral artery

The incidence of injuries to the vertebral artery from blunt injury to the cervical spine is much higher than realized in the past. In recent studies, patients undergoing routine noninvasive testing or vertebral arteriography after sustaining subluxation from a locked facet, destruction of a facet, or fracture of the foramen transversarium have had a 46 to 88 per cent incidence of injury to the vertebral artery. It is now suggested that all patients with fractures in the cervical spine have screening of the vertebral artery, much as in patients with gunshot wounds in proximity to this vessel.

An intimal defect in the second portion of the vertebral artery in its bony canal or the third portion posterior to the atlas is treated nonoperatively with heparinization in the absence of injury to the brain or multiple associated injuries. The presence of a pseudoaneurysm greater than 2 to 3 mm in size mandates coil or balloon occlusion as long as the contralateral vertebral artery is intact.

Active hemorrhage from the first portion of the vertebral artery behind the scalenus anticus muscle in the supraclavicular area is treated with ligation of the vessel.

Thoracic aorta

In patients with significant deceleration or a major lateral impact in a motor vehicle crash, clinical signs suggestive of blunt injury to the descending thoracic aorta include the following: (1) mark of steering wheel on anterior chest; (2) interscapular murmur; (3) upper extremity hypertension and decreased femoral pulses (pseudo-coarctation syndrome); (4) paraparesis or paraplegia; and (5) tenderness over the lower thoracic spine. Even without these clinical signs, the abnormalities on a chest radiograph listed in Table 5 suggest that further diagnostic evaluation is needed.

Widened superior mediastinum ¿ 8 cm on upright film
Opacification of aortopulmonary window
Deviation of the nasogastric tube to the right
Deviation of the trachea to the right
Depression of left mainstem bronchus
Presence of apical pleural cap
Widening of right paratracheal stripe
Widening of paraspinal line
Calcium layering in aortic arch or knob
Massive left hemothorax
Unilateral fracture of first rib and second rib or clavicle
Severely displaced fracture of the first rib
Bilateral first rib fractures
Multiple rib fractures, especially with flail chest
Fracture-dislocation of the sternum
Fracture-dislocation of the lower thoracic spine

Table 5 Signs suggestive of blunt injury to the descending thoracic aorta on a chest radiograph

Contrast-enhanced spiral CT of the thorax has had a 100 per cent sensitivity in detecting blunt rupture of the thoracic aorta in three recent reports. Overall accuracy was 86 and 99.7 per cent in two of these reports. Transesophageal echocardiography had a 93 per cent sensitivity and 98 per cent accuracy in detecting the same lesion in another recent study. Finally, digital subtraction aortography continues to have a sensitivity of 92 to 99 per cent and an accuracy of 97 to 98 per cent.

In hemodynamically stable patients who are to undergo operative repair of a rupture in the descending thoracic aorta because of the presence of a significant true or false aneurysm, operative approaches include clamp/repair, use of an external shunt, use of left atriofemoral bypass via a roller pump, or use of left atriofemoral bypass via a centrifugal pump. In the last and most popular approach, the left atrium or left superior pulmonary vein and the left femoral artery are connected by a centrifugal pump that maintains flow rates at 2 to 3.5 l/min. to the distal abdominal aorta. In a recent multicenter study, only two of 69 patients undergoing aortic repair using the centrifugal pump developed paraplegia. When aortic cross-clamp times are likely to exceed 30 min, heparinization is indicated in the absence of injury to the brain.

Innominate, intrathoracic common carotid, and subclavian arteries

Occasional patients with penetrating injuries to any of these vessels may have exsanguinating hemorrhage into one pleural cavity. With a midline wound, a high bilateral anterolateral thoracotomy above the nipples is performed in the emergency center. A wound to either side of the sternum or thoracic inlet and confirmation of an ipsilateral hemothorax on a surgeon-performed ultrasound is treated with a high unilateral thoracotomy in the exsanguinating patient. Obvious intrapleural hemorrhage from the underside of the right subclavian vessels is managed as an emergency with transpleural apical packing held in place by the surgeon's fist. On the left, clamping of the intrathoracic proximal left subclavian artery is indicated. An unstable patient without intrapleural hemorrhage is moved to the operating room without a preliminary aortogram (Fig. 9).



Fig. 9. Gunshot wound to superior mediastinum injured innominate artery, left innominate vein, left common carotid artery, and posterior aspect of aortic arch. The patient had a contained hematoma and was stable enough to be transported to the operating room.

Hemodynamically stable patients with hematomas in the supraclavicular area on physical examination, the mark of a shoulder restraint, or a hematoma in the superior mediastinum on a chest radiograph undergo an arch aortogram. When a significant true or false aneurysm rather than an intimal defect is present, the choice of incision depends on the vessel injured (Fig. 5 and Fig. 6). Injuries to the innominate, intrathoracic common carotid, or proximal right subclavian artery are best approached through a median sternotomy. The very rare injury to the intrathoracic left subclavian artery is best visualized through a left lateral or posterolateral thoracotomy. Injuries to the second or third portions of the subclavian artery are approached through a supraclavicular incision with division of the clavicle or removal of its middle two-thirds. Intraluminal bypass shunts are not necessary for repairs of the innominate or common carotid arteries in hemodynamically stable patients.

Abdominal vascular trauma – midline or zone 1

Hematomas or areas of hemorrhage in the midline are divided further into supramesocolic or inframesocolic groups. In the supramesocolic area, injuries to the suprarenal abdominal aorta, celiac axis, proximal superior mesenteric artery, proximal renal artery and superior mesenteric vein may be present. Active hemorrhage from this area is approached by retracting the esophagus to the left, splitting the lesser omentum, and clamping the abdominal aorta in the aortic hiatus of the diaphragm. A hematoma in this area in the reasonably stable patient is approached by a left medial mobilization maneuver. This involves dissection with scissors and the fingers to mobilize the left colon, spleen, tail of the pancreas, left kidney, and fundus of the stomach to the midline (Fig. 10). The suprarenal abdominal aorta can then be clamped after removal of the celiac ganglion and the numerous lymphatic channels in this area. If this should prove to be too time-consuming, division of the aortic hiatus of the diaphragm at the 2 o'clock position will allow for visualization and clamping of the descending thoracic aorta in the posterior mediastinum. Distal dissection on the suprarenal aorta will lead to the area of injury in the aorta itself or in one of the visceral vessels. hrough-and-through gunshot wounds of the aorta at

this level are connected and closed in a continuous fashion with polypropylene suture. An extensive injury of the aorta requires segmental resection and insertion of a 12 to 16 mm polytetrafluoroethylene or woven Dacron graft. Injuries to the celiac axis are ligated, while an extensive injury to the proximal superior mesenteric artery that requires ligation is usually followed by an aorto–mid-superior mesenteric artery bypass graft to revascularize the midgut. An alternate approach is to insert a rigid intraluminal shunt into the two ends of the transected or resected superior mesenteric artery to maintain flow until the patient is stabilized in the surgical intensive care unit (‘damage control celiotomy'). Extensive injuries to the superior mesenteric vein under the transverse mesocolon may be treated with ligation. The splanchnic hypervolemia that results will resolve over time, while systemic hypovolemia is treated with the infusion of large volumes of crystalloid solutions.

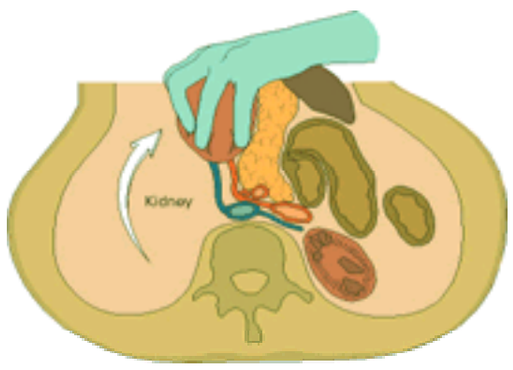


Fig. 10. Left medial mobilization maneuver to expose suprarenal abdominal aorta.

In the inframesocolic area, injuries to the infrarenal abdominal aorta or infrarenal inferior vena cava may be present. Possible vascular injuries are approached in the midline retroperitoneum at the base of the transverse mesocolon. The infrarenal abdominal aorta is clamped just below the left renal vein if a more distal injury is suspected. Aortic repairs at this level are similar to those described for the suprarenal aorta. When there is no injury to the infrarenal abdominal aorta and the midline inframesocolic hematoma elevates the ascending colon as well, injury to the infrarenal inferior vena cava is likely. Ideal exposure of the cava is obtained by a right-sided medial mobilization maneuver in which the cecum and right colon and the C-loop of the duodenum are mobilized to the midline (Fig. 11). Perforations in the inferior vena cava can be grasped with an Allis clamp, elevated, and a Satinsky vascular clamp applied underneath. Transverse repair, if at all possible, is completed with 4–0 or 5–0 polypropylene suture. Exsanguinating hemorrhage from a major infrarenal caval injury is treated with proximal and distal ligation around the perforation. It is worthwhile to perform bilateral below-knee four-compartment fasciotomies in patients who have been in profound shock prior to ligation of the inferior vena cava. Wounds of the cava at the junction with the renal veins usually require proximal and distal control with large vascular clamps and control of the renal veins with silastic vessel loops or angled DeBakey clamps. Wounds at the confluence of the iliac veins may require division of the right common iliac artery and mobilization of the aortic bifurcation to the left for proper visualization. The survival rate for patients with abdominal vascular injuries is listed in Table 6 and Table 7.

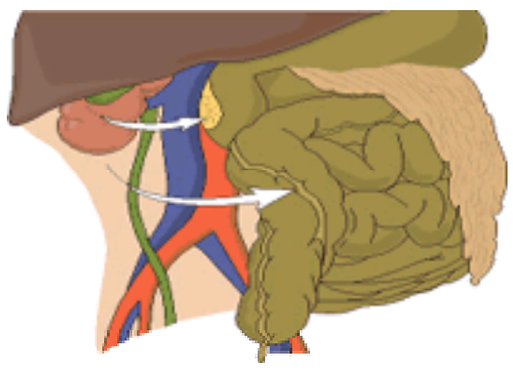


Fig. 11. Right medial mobilization maneuver to expose infrahepatic inferior vena cava. (Copyright, Baylor College of Medicine, 1981, and used with permission.)

Vessel	Survivors/all patients	Survival (%)
Suprarenal aorta	54/155	34.8
Superior mesenteric artery	67/116	57.7
Infrarenal aorta	43/93	46.2
Renal artery	19/22	86.7
Iliac artery	145/236	61.4

Table 6 Survival after arterial injuries of the abdomen*

Vessel	Survivors/all patients	Survival (%)
Superior mesenteric vein	75/104	72.1
Infrahepatic inferior vena cava		
Suprarenal	52/88	59.7
Renal	55/84	65.5
Infrarenal	223/285	78.2
Renal vein	38/43	88.3
Iliac vein	282/404	69.8
Portal vein	67/134	50.0

*Adapted from Feliciano DV, Burch JM, Graham JM: Abdominal vascular injury. In: Mattox KL, Feliciano DV, Moore EE, eds. Trauma. New York: McGraw-Hill, 2000:785–805.

Table 7 Survival after venous injuries of the abdomen*

Abdominal vascular trauma – upper lateral retroperitoneum or zone 2

In patients with blunt trauma and a reasonably normal appearance of a kidney on a preoperative IVP, renal arteriogram, or CT, a perirenal hematoma discovered at a laparotomy for other injuries should not be opened. A perirenal hematoma from a penetrating wound is opened once proximal vascular control of the ipsilateral renal artery and vein at the midline has been obtained. Active hemorrhage from the renal hilum or parenchyma is managed by rapid elevation of the injured kidney out of the retroperitoneum and application of a vascular clamp to the hilar vessels. The renal artery is small and difficult to repair, and significant injuries are usually treated with nephrectomy in the presence of a palpably normal contralateral kidney. The left renal vein can be ligated, if necessary, at the midline to control exsanguination. On the right side, ligation of the renal vein mandates a nephrectomy.

Abdominal vascular trauma – pelvic retroperitoneum or zone 3

Pelvic hematomas from blunt trauma that are unruptured, nonpulsatile, and not rapidly expanding are left intact. A lateral pelvic hematoma from a penetrating wound is opened only after proximal control of the common iliac artery and vein and distal control of the external iliac artery and vein are attained. Injuries to the common or external iliac artery are repaired with the techniques described previously. When extensive fecal contamination from a penetrating abdominal wound accompanies a major injury to the common or external iliac artery, ligation of the uninjured proximal and distal artery is appropriate. This avoids late blowout of an end-to-end anastomosis or anastomosis to a graft should postoperative pelvic cellulitis or an abscess occur. An extra-anatomic crossover femorofemoral bypass graft will maintain viability of the ipsilateral lower extremity. If the patient is unstable and the extra-anatomic bypass will have to be delayed, an ipsilateral below-knee four-compartment fasciotomy is a temporary maneuver to protect the leg.

Abdominal vascular trauma – portal and retrohepatic area

A suspected vascular injury in the hepatoduodenal ligament is treated with a roximal Pringle maneuver and, if possible, the application of a distal vascular clamp at the edge of the liver. Careful dissection of the hepatic artery, portal vein, and common bile duct is necessary before any vascular repair is attempted. Significant injuries to the common, right, or left hepatic artery are treated with ligation as long as the portal vein is uninjured. Transverse venorrhaphy is preferred for small and medium perforations of the portal vein, while ligation is appropriate when there is exsanguination from a major injury. Postoperative management is the same as when ligation of the superior mesenteric vein has been necessary.

Injuries to the retrohepatic vena cava are discussed in Chapter 30.2.

Peripheral vascular injuries

Longitudinal straight incisions are used to expose peripheral vascular injuries, except when the shoulder, antecubital area or popliteal area are involved. Gently curving incisions are used in these locations to avoid postoperative contractures of joints. It may be necessary to approach injured vessels with active hemorrhage directly, while proximal and distal vascular control is obtained when a hematoma is present. Vascular control is maintained using small angio-access clamps or silastic vessel loops. When segmental resection of an injured artery has been necessary, a Fogarty balloon catheter is passed proximally and distally. Prior to reapplication of the angioaccess clamps, 10 to 15 ml of a heparinized saline solution (50 units/ml) is infused into both the proximal and distal ends of the vessel. As previously noted, temporary use of intraluminal shunts, completion arteriography, performance of a below-knee or elbow fasciotomy, and advanced techniques for salvage of a mangled extremity are applied when appropriate.

Further reading

Biffl WL, *et al.* The unrecognized epidemic of blunt carotid arterial injuries. Early diagnosis improves neurologic outcome. *Annals of Surgery* 1998; **228**: 462–70.

Dale WA. The beginnings of vascular surgery. *Surgery* 1974; **76**: 849–66.

DeBakey ME, Simeone FA. Battle injuries of the arteries in World War II. An analysis of 2,471 cases. *Annals of Surgery* 1946; **123**: 534–79.

Demetriades D, *et al.* Penetrating injuries of the neck in patients in stable condition. Physical examination, angiography, or color flow Doppler imaging. *Archives of Surgery* 1995; **130**: 971–5.

Dennis JW, Frykberg ER, Veldenz HC, Huffman S, Menawat SS. Validation of nonoperative management of occult vascular injuries and accuracy of physical examination alone in penetrating extremity trauma: 5- to 10-year follow-up. *Journal of Trauma* 1998; **44**: 243–53.

Duke BJ, Moore EE, Brega KE. Posterior circulation cerebral infarcts associated with repair of thoracic aortic disruption using partial left heart bypass. *Journal of Trauma* 1997; **42**: 1135–9.

Fabian TC, Patton JH Jr, Croce MA, Minard G, Kudsk KA, Pritchard FE. Blunt carotid injury. Importance of early diagnosis and anticoagulant therapy. *Annals of Surgery* 1996; **223**: 513–25.

Fabian TC, *et al.* Prospective study of blunt aortic injury: multicenter trial of the American Association for the Surgery of Trauma. *Journal of Trauma* 1997; **42**: 374–83.

Feliciano DV. Vascular injuries. In: Maull KI, Cleveland HC, Strauch GO, Wolferth CC, eds. *Advances in trauma.* Volume 2. Chicago: Year Book Medical Publishers, 1987: 179–206.

Feliciano DV, *et al.* Management of vascular injuries in the lower extremities. *Journal of Trauma* 1988; **28**: 319–28.

Feliciano DV, ed. A new look at penetrating carotid artery injuries. In: Maull KI, Cleveland HC, Feliciano DV, *et al.*, eds. *Advances in trauma and critical care.* Volume 9. St. Louis: Mosby, 1994: 319–45.

Feliciano DV. Truncal vascular trauma. In: Callow AD, Ernst CB, eds. *Vascular surgery. Theory and practice*. Stamford: Appleton and Lange, 1995: 1059–85.

Feliciano DV. Abdominal vessels. In: Ivatury RR, Cayten CG, eds. *The Textbook of penetrating trauma*. Baltimore: Williams and Wilkins, 1996: 702–15.

Feliciano DV. Trauma to the aorta and major vessels. *Chest Surgery Clinics of North America* 1997; **7**: 305–23.

Feliciano DV. Carotid trauma. When to operate. In: Yao JST, Pearce WH, eds. *Practical vascular surgery*. Stamford: Appleton and Lange, 1999: 391–400.

Feliciano DV. Injuries to the great vessels of the abdomen. In: Wilmore DW, Cheung LY, Harken AH, Holcroft JW, Meakins JL, eds. *Scientific American surgery*. New York: Scientific American, Inc, 1999: Section IV, Chapter 9, 1–12.

Feliciano DV, Burch JM, Graham JM. Abdominal vascular injury. In: Mattox KL, Feliciano DV, Moore EE, eds. *Trauma.* New York: McGraw-Hill, 2000: 783–805.

Frykberg ER, Dennis JW, Bishop K, Laneve L, Alexander RH. The reliability of physical examination in the evaluation of penetrating extremity trauma for vascular injury: Results at one year. *Journal of Trauma* 1991; **31**: 502–11.

Knudson MM, Lewis FR, Atkinson K, Neuhaus A. The role of duplex ultrasound arterial imaging in patients with penetrating extremity trauma. *Archives of Surgery* 1993; **128**: 1033–8.

Larsen PE, Smead WL. Vertical ramus osteotomy for improved exposure of the distal internal carotid artery: A new technique. *Journal of Vascular Surgery* 1992; **15**: 226–31.

Mattox KL, Feliciano DV, Burch J, Beall AC Jr, Jordan GL Jr, De Bakey ME. Five thousand seven hundred sixty cardiovascular injuries in 4459 patients. Epidemiologic evolution 1958 to 1987. *Annals of Surgery* 1989; **209**: 698–707.

Mirvis SE, Shanmuganathan K, Buell J, Rodriguez A. Use of spiral computed tomography for the assessment of blunt trauma patients with potential aortic injury. *Journal of Trauma* 1998; **45**: 922–30.

Schwartz M, Weaver F, Yellin A, Ralls P. The utility of color flow Doppler examination in penetrating extremity arterial trauma. *The American Surgeon*, 1993; **59**: 375–8.

Semba CP, *et al.* Acute rupture of the descending thoracic aorta: repair with use of endovascular stent-grafts. *Journal of Vascular and Interventional Radiology* 1997; **8**: 337–42.

Veith FJ, Marin ML. The present status of endoluminal stented grafts for the treatment of aneurysms, traumatic injuries and arterial occlusions. *Cardiovascular Surgery* 1996; **4**:3–7.

17.12 Infection and arterial surgery

Christopher P. Conlon and Jack Collin

Infections involving native vessels
Mycotic aneurysms
Infected aneurysms in intravenous drug users
Vessels involved in adjacent infection
Prosthetic graft infection
Epidemiology
Pathogenesis
Prevention
Clinical features
Diagnosis
Management
Further reading

The arterial system is prone to infection in a number of ways. Native vessels may become infected either from a distant source, because of a bacteraemia, or because of local sepsis from an adjacent focus. In Western societies with a high prevalence of atheromatous vascular disease, many operations are carried out to replace or bypass diseased arteries. The vascular grafts inserted are at risk of per- and postoperative infection. This chapter outlines the pathogenesis, clinical features, and management of infections involving the arterial tree.

Infections involving native vessels

Mycotic aneurysms

The original description of mycotic aneurysms by Osler related to aneurysms in the aortic arch of a patient with bacterial endocarditis. The term is now used to describe any aneurysm of infective origin resulting either from micro-organisms spreading in the bloodstream from a distant focus, by direct implantation, or by contiguous infection.

Pathogenesis

During an episode of bacteraemia, organisms may adhere to atheromatous plaques. Adherence is a prerequisite for bacterial multiplication and invasion. The resulting inflammation leads to destruction of the arterial intima and subsequent aneurysm formation. Untreated, the infection continues with further arterial damage and eventual rupture of the weakened vessel wall.

Mycotic aneurysms continue to be a complication of infective endocarditis. Infective vegetations on the valves and endocardial surface break off and enter the systemic circulation. These septic emboli often occlude small arteries in the cerebral circulation or in the peripheral arteries of a limb where they can cause local tissue ischaemia. Sometimes the infective emboli stick to the endothelium of a large artery, establishing a local infection and subsequent aneurysm formation. Any artery in the body can be affected. The most common organisms involved in infective endocarditis include *Staphylococcus aureus, Streptococcus viridans*, and other streptococci, and the enterococci. In practice, many bacteria and some fungi can cause endocarditis and may lead to mycotic aneurysms.

More commonly, mycotic aneurysms of the aorta, particularly in the abdomen, arise following bacteraemia from a non-cardiac source. The circulating bacteria adhere to atheroma in the vessel wall. These aneurysms are more common in the elderly who have extensive atheromatous plaques throughout the arterial tree and who are more likely to become bacteraemic, possibly as a consequence of impairment of immunity with ageing. The infecting organisms frequently derive from the gastrointestinal tract. *Salmonella enteritidis* and other salmonella species cause an infective gastroenteritis and are more prone to cause colonic inflammation and to invade the bloodstream in the elderly. These organisms are the most common cause of mycotic abdominal aortic aneurysms. Much less commonly, *Yersinia enterocolitica* bacteraemia occurs following food poisoning and leads to mycotic aneurysms in exactly the same way. Other organisms, such as *E. coli* and *Bacteroides fragilis*, are normally found in the gastrointestinal tract but can enter the bloodstream if the integrity of the intestinal mucosa is impaired, as sometimes occurs in diverticular disease, colonic neoplasms, or inflammatory bowel disease. Rarely, Gram-positive organisms, such as *S. aureus*, lead to aortic aneurysms following an episode of bacteraemia.

Clinical features

Although some mycotic aneurysms present with the features of fever, abdominal pain, and a pulsatile mass, the clinical picture is usually more subtle. Many patients present with a subacute illness including malaise, weight loss, and intermittent fevers and sweats suggestive of chronic sepsis. Others present with fever and severe back pain as the expanding, infected aorta involves the spine posteriorly and causes an infective discitis. On rare occasions, direct contiguous spread from a primary discitis involves the aorta secondarily. A few patients present with severe sepsis and an aneurysm might be obvious clinically or be found during investigations to determine the source of the sepsis. In a few patients the problem is diagnosed only when the aneurysm bursts and pus or severe inflammation are found at operation. There is sometimes a history of recent gastrointestinal infection or bowel disturbance or a recent hospital admission during which intravenous cannulation (a source of staphylococcal sepsis), or other invasive procedures that could have caused a transient bacteraemia, were carried out.

Investigations

These should be aimed at defining the vascular abnormality and at determining the infecting organism and its source. Some patients have a neutrophilia or raised inflammatory markers, such as an elevated C-reactive protein (CRP) or ESR but these investigations are neither specific nor sensitive. Patients with mycotic aneurysms will have been bacteraemic at some stage and a proportion will be bacteraemic at the time of presentation, particularly if febrile. Blood cultures should always be obtained; three sets of cultures taken at different times and from different sites over 24 h maximizes the yield. Computed tomography (CT) or magnetic resonance imaging (MRI) confirms the presence of the aneurysm and allows identification of infective involvement of adjacent structures, such as vertebral bodies (Fig. 1). Occasionally, such imaging reveals a likely source of the bacteraemia, such as a diverticular abscess or a colonic carcinoma. In some cases angiography will be needed to define the vascular anatomy before embarking on surgery (Fig. 2).



Fig. 1. MRI showing abdominal aortic mycotic aneurysm.

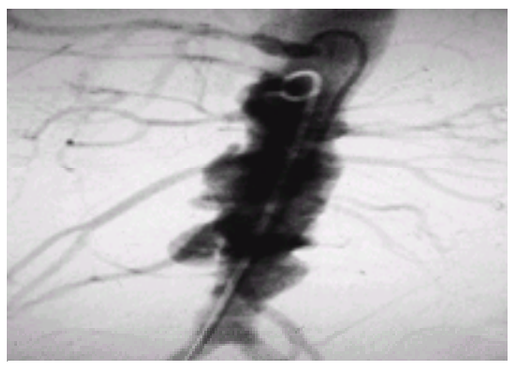


Fig. 2. Angiogram showing mycotic aneurysm in abdominal aorta.

Management

Antibiotics play a crucial role in the management of mycotic aneurysms. If the infecting organism is known preoperatively, appropriate antibiotic treatment should be started early, reducing the risk of sepsis syndrome in the bacteraemic patient and reducing the risk of infection in any new prosthetic graft which is subsequently inserted. In many cases, however, the infecting agent cannot be determined preoperatively. Ideally, in these circumstances, operative specimens should be obtained and sent for culture before antibiotics are administered, maximizing the chances of identifying the infecting organism and the appropriate choice of antimicrobial agent. If the patient is septic, it is necessary to start antibiotics empirically, in which case therapy must be broad spectrum to cover the aerobic and anaerobic organisms most commonly responsible for these infections. The optimal duration of antibacterial treatment has not been established but, as in the treatment of bacterial endocarditis, parenteral therapy for 4 to 6 weeks seems appropriate. With some infections, such as those due to *Salmonella enteritidis*, there are reliably-absorbed, oral antibiotics that can sometimes be used. In our practice, we treat for a minimum of 3 months with antibiotics, with the first 6 weeks being given intravenously. Once the patient is stable, the remainder of the parenteral antibiotic therapy can be administered outside hospital if a suitable home therapy programme is available.

In some cases, antibiotics can sterilize the aneurysm and prevent further damage to the arterial wall, but usually the infection is merely suppressed and arterial destruction is progressive. In small arteries, thrombosis of the aneurysm can occur and, in such cases, antibiotics alone may be curative. In the majority of cases, either interventional radiological techniques or surgery are required to prevent rupture of the aneurysm. The arterial wall may be rapidly destroyed by infection so intervention is usually urgently required and should be undertaken as soon as possible after the diagnosis of mycotic aneurysm has been confirmed.

Mycotic aneurysms in non-essential arteries, where organs have an adequate alternative blood supply (for example the splenic artery), can be treated by proximal and distal occlusion of the artery. This can be achieved by placement of intra-arterial balloons or coils inserted percutaneously. Aneurysms in larger arteries and end arteries supplying vital organs must be treated surgically to remove the risk of aneurysm rupture and to replace or restore the blood supply. For aneurysms in some arteries, a bypass graft can be inserted in an extra-anatomic site, away from the infection, and the aneurysm treated separately by excision or exclusion. This technique is particularly suitable for aneurysms affecting limb vessels where the patency rates for extra-anatomic bypass grafts are comparable to those for *in situ* grafts.

Many surgeons advocate a similar approach, using extra-anatomic bypass grafts, for mycotic aortic aneurysms. This involves the insertion of an axillobifemoral bypass graft with the excision of the aortic aneurysm and proximal and distal aortic occlusion. This operation can conveniently be performed as a two-stage procedure, the extra-anatomic bypass being established first before the abdominal operation to excise the aneurysm (Fig. 3). This approach minimizes the duration of the peroperative limb ischaemia but its main disadvantage is that long-term patency rates for axillobifemoral bypass grafts are substantially less than those for *in situ* aortic grafts.

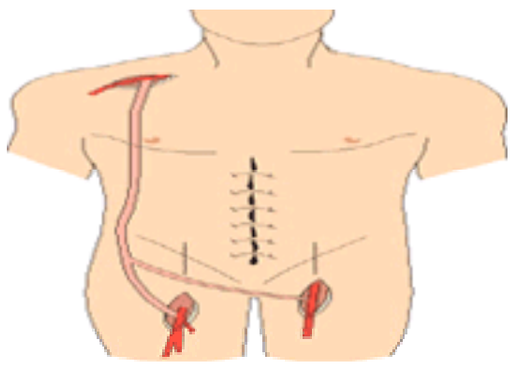


Fig. 3. Diagram showing axillobifemoral bypass graft.

In recent years, there has been an increasing tendency to deal with mycotic aortic aneurysms in the same way that non-infected aneurysms are dealt with; namely by *in situ* aortic tube or bifurcated graft insertion. Provided the patient has been adequately treated with antibiotics perioperatively and extensive debridement of infected and necrotic tissue is performed, the risk of perioperative graft infection appears to be small. In all cases, postoperative antibiotic therapy will be required for several weeks, irrespective of whether an extra-anatomic bypass or an *in situ* graft has been used.

Infected aneurysms in intravenous drug users

Intravenous drug users (IVDU) are at an increased risk of mycotic aneurysms through the use of dirty needles and syringes. Their level of hygiene and state of health is often poor, many are relatively malnourished and generally their resistance to infection is impaired. There is no evidence that human immunodeficiency virus (HIV) infection increases the risk of mycotic aneurysm formation but IVDU with HIV may fare worse than those without HIV should an infected aneurysm occur. Aneurysms can occur at any site but are most common at injection sites where arteries are punctured either inadvertently or deliberately when accessible veins no longer exist.

The same principles of management outlined above apply to aneurysms in IVDU but local injection site aneurysms are often difficult to treat successfully. Extra-anatomic bypass grafts may provide too tempting a target for easy injection, leading to graft infection and further problems. Some surgeons with a large experience of this patient group advocate simple ligation of the injection site aneurysm to prevent exsanguinating haemorrhage, claiming that this procedure rarely compromises limb viability and arterial bypass grafting is unnecessary.

Vessels involved in adjacent infection

Arteries can become secondarily infected following either accidental or iatrogenic trauma. Penetrating injuries with non-sterile objects or those that cross through infected fields can cause infection of adjacent arteries and subsequent aneurysm formation. More commonly, abscesses adjacent to an artery may progress to involve the arterial wall. This occurs, for example, with psoas abscesses caused by pyogenic bacteria and as a complication of tuberculosis or fungal infections.

These patients are less likely to be bacteraemic compared to those with mycotic aneurysms and sometimes do not have obvious symptoms or signs of infection. They often have local pain and swelling or other symptoms and signs if structures, such as nerve roots, are involved in the inflammatory process. The surgical and antibiotic therapeutic strategies are essentially the same as those employed for mycotic aneurysms.

Prosthetic graft infection

Epidemiology

As a consequence of the increasing number of elderly people in developed countries, there has been an increase in the number of arterial operations performed for both occlusive and aneurysmal arterial disease. In addition, it has now been established that useful secondary prevention of stroke can be achieved by carotid

endarterectomy in symptomatic patients with more than 70 per cent stenoses of the internal carotid artery. This procedure often involves the use of prosthetic material for patch closure of the endarterectomy.

In Scotland, approximately 18 grafts/100 000 population are inserted for aortic disease each year. Even in the best hands, infection rates of 1 to 2 per cent can be expected throughout the life of the graft. Prosthetic graft infections are arbitrarily classified as early and late infections. Early graft infections present within 4 months of surgery and late graft infections any time after 4 months. Regardless of the time of presentation, most infections actually occur at the time of surgery. In some cases infection of the native aortic wall is already present and the graft becomes infected by being placed into a non-sterile field. More commonly, there is peroperative contamination of the graft either with organisms from the patient's own skin or, less commonly, by flora from the skin of operating room staff. Rarely, there may be peroperative contamination of the graft by bowel flora, particularly if additional non-vascular operative procedures are undertaken. There is also a slight risk that grafts may be seeded during bacteraemias secondary to the insertion of intravenous cannulae or bladder catheterization in the perioperative period.

Prosthetic grafts are also inserted for the treatment of severe claudication or critical limb ischaemia. Most operations involve an incision and dissection in the groin; a region heavily colonized by bacteria where it is difficult to ensure sterility of the skin. In consequence, peroperative graft infection when the groin is dissected is more common than in aortic surgery confined to the abdomen. Up to 10 per cent of grafts have become infected in some series when surgery involved the groin.

Generally, infection of aortic grafts results in a higher mortality than does infection of peripheral grafts, whereas the latter are more likely to result in limb loss. Either type of graft is more likely to become infected if the surgery is done as an emergency, if the operating time is long, if reoperation is required for any reason, or if there is a superficial wound infection. Obesity, diabetes mellitus, and other conditions with impaired immunity increase the risk of graft infection.

In recent years, intravascular stents have been used with increasing frequency. The majority of these have been placed in coronary arteries following angioplasty for ischaemic heart disease. Transluminal graft insertion is also being used to treat abdominal aortic aneurysms, often in patients who are not fit for general anaesthesia or who are not suitable for an open operation for other reasons. Infection of coronary stents has so far not been a significant problem and the risk appears to be small. On the other hand, placement of a transluminal aortic graft usually involves groin dissection and considerable instrumental manipulation in the groin so there is a greater risk of infection, particularly if the operation is prolonged. Reports of infection of grafts inserted transluminally are increasing but the magnitude of the problem is, as yet, unknown.

Pathogenesis

The majority of prosthetic graft infections occur at the time of surgery and most of those involving aortic grafts are due to coagulase-negative staphylococci or to *S. aureus*, common skin commensals. If infection is related to ischaemic or damaged bowel, Gram-negative bacteria, such as *E. coli*, and anaerobic bacteria are often implicated. Peripheral grafts that are inserted using groin incisions are also likely to be infected by staphylococci but, because the groin is in close proximity to the perineum, faecal organisms can contaminate the wound and these bacteria are usually either anaerobes or Gram-negative aerobic bacteria.

One of the characteristics of bacterial infections of prosthetic surfaces is the formation of a biofilm, or slime. The biofilm is composed of microcolonies of bacteria and bacterial products, mainly glycocalyx, and is adherent to the prosthetic material. The infecting bacteria live in this layer on the surface of the graft and are relatively inaccessible to both antibiotics and the host immune system. These bacteria often exhibit growth characteristics that differ from ‘free-living', or planktonic, bacteria; they become relatively sessile. Such slow growth characteristics means that these bacteria are less sensitive to many antibiotics, most of which act specifically on dividing bacteria. The presence of a biofilm makes it very difficult to eradicate infection.

Different graft materials have varying susceptibility to infection. Dacron grafts are more likely to become infected than grafts made of PTFE (polytetrafluoroethylene). The use of vein grafts instead of prosthetic material greatly reduces the risk of infection.

Survival of bacteria in biofilms is not, in itself, a problem but the organisms spread to involve the suture line between the graft and the native vessel, and the vessel itself. Infection of the suture line may lead to dehiscence and subsequent haemorrhage. Involvement of the native vessel can lead to weakening of the arterial wall and aneurysm formation. In some cases, particularly in smaller arteries, infection causes arterial wall inflammation and promotes thrombus formation with the risk of distal ischaemia.

Prevention

The surest way to prevent infections of vascular graft is by scrupulous adherence to a careful surgical technique. Proper skin preparation, minimizing personnel and movement in the operating room, and careful handling of instruments and tissues all contribute. The avoidance of damage to bowel and the achievement of haemostasis reduce the risk of re-exploration and hence infection. In addition, perioperative antibiotic prophylaxis is mandatory and has been shown to reduce superficial wound infection rates to less than half that achieved without prophylaxis. Very few prosthetic grafts become infected following transient bacteraemias related to subsequent medical or dental procedures after the initial graft insertion and there is no evidence that antibiotic prophylaxis is required to cover such procedures. In contrast, in the early postoperative period, before the graft has completely incorporated, bacteraemia related to intravenous cannulae or urinary catheters might pose a risk. It is good practice to remove such devices as soon as possible after surgery. It is currently fashionable to use prosthetic grafts that have antibiotics bonded to the material, particularly if the graft is being used to replace an infected graft. There is, however, no evidence to confirm that antibiotic-impregnated grafts reduce the risk of subsequent infection.

Clinical features

Unlike the clinical picture seen with mycotic aneurysms, infected prosthetic grafts rarely present with obvious systemic symptoms or signs of sepsis. Early graft infections can present with wound breakdown. This is much more common with peripheral grafts involving groin dissection than it is with abdominal aortic grafts. In severe cases the graft is often visible in the wound or there might be dehiscence of the graft with obvious haemorrhage.

Most prosthetic graft infections present late and this is particularly true of aortic grafts, which can present in a variety of ways. Most dramatically, when the upper end of the graft is infected, the resulting inflammation involves the adjacent bowel (usually the first few centimetres of the jejunum) and an aortoenteric fistula forms. This usually presents as a brisk gastrointestinal haemorrhage. Patients are usually hypotensive and have profuse melaena; haematemesis is uncommon, except in massive exsanguinating haemorrhage. The location of the fistula means that abnormalities can be detected by upper gastrointestinal endoscopy only if a colonoscope is used so that the first part of the jejunum can be visualized. Occasionally, aortic graft infections present with symptoms and signs suggestive of infection. Even when the patient has fever and rigors, blood cultures are rarely positive. More commonly, patients present with an indolent illness and have intermittent fevers and sweats along with general malaise and weight loss. This presentation with ‘pyrexia of unknown origin' is usually not associated with any abnormal signs other than fever.

Aortic graft infections are most likely to present with local complications. Infection in the distal part of an aortobiiliac graft can present with occlusion of one of the limbs of the graft, precipitating critical limb ischaemia. False aneurysms sometimes form at either the proximal or distal anastomoses of an infected graft and these can be detected by postoperative surveillance imaging or, incidentally, when investigations are being performed for unrelated reasons. In some cases, infection is discovered incidentally when vascular surgery is performed to correct a problem not thought to be infective in origin.

Infections of peripheral vascular grafts, such as femoropopliteal bypass grafts, almost always present with local symptoms and signs and are only rarely associated with systemic symptoms of infection. The most common presentation is with wound breakdown in the early postoperative period and the formation of a chronically discharging sinus in the groin (Fig. 4). These infections can also present with false aneurysms at the suture line between graft and native vessel. Occlusion of peripheral arterial grafts is not uncommon and, in some cases, the occlusion is secondary to infection.



Fig. 4. Sinus in groin of a patient with an infected femoropopliteal graft.

Diagnosis

Aortoenteric fistula due to graft infection should be suspected in any patient with melaena and a past history of abdominal aortic surgery. Upper gastrointestinal endoscopy should always be performed to exclude other diagnoses, such as a bleeding peptic ulcer or bleeding oesophageal varices. The site of the fistula is almost always in the first part of the jejunum. Upper gastrointestinal endoscopy, using a colonoscope, is the most useful investigation and is the only way to establish the diagnosis with certainty preoperatively. Unless the endoscope has been passed at least 5 cm beyond the duodenojejunal junction, the fistula will not be seen and the endoscopy should be considered a failed investigation, rather than a negative one. Labelling this failed investigation as a negative test will induce a false sense of security. Abdominal CT is usually non-diagnostic unless there are gas bubbles seen around the aortic anastomosis (Fig. 5). Angiography is unhelpful except in the rare patient who is actively bleeding during the investigation. It is often necessary to proceed to laparotomy on the clinical diagnosis alone.

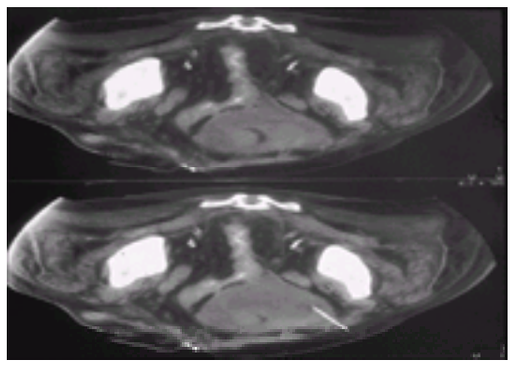


Fig. 5. Abdominal CT showing a perigraft collection of pus in a patient with an infected aortic graft.

In patients presenting with the systemic features of sepsis, blood cultures are sometimes positive. Unlike patients with endocarditis or those with mycotic aneurysms, those with infected prostheses rarely have a sustained bacteraemia. It is thus necessary to take several sets of blood cultures at different times to maximize the chances of identifying an organism. In these cases, and in those presenting more chronically with a pyrexia of unknown origin, imaging of the abdomen with either CT or MRI can be helpful. This might reveal a perigraft fluid collection or a loop of bowel stuck to the graft; sometimes the only clue will be the CT appearance of ‘streaky', or inflamed, fat in the region of the graft. Although there are enthusiasts for using indium-labelled white cell scintigraphy to localize infected grafts, this technique has not been shown to be as useful as CT or MRI and the scintigrams are usually difficult to interpret.

Local complications of infection, such as anastomotic aneurysms and occlusions, can usually be demonstrated by ultrasonagraphy, CT, or angiography. Confirmation that the abnormality detected is due to infection can only be accomplished by macroscopic and histological examination of tissue specimens obtained at surgery along with the culture of the infecting organisms from the graft or arterial specimen.

It is essential to obtain good operative specimens in order to determine which organism is responsible for the prosthetic graft infection. Without this information it is much more difficult to ensure adequate antimicrobial therapy. Specimens should be taken with new instruments (i.e. not those used for the initial surgical dissection) and placed into dry, sterile specimen pots for prompt despatch to the microbiology laboratory. The ideal number and type of specimens to send have not been determined in any properly conducted studies. Clearly, any pus should be collected and sent for culture; a few millilitres of pus are more valuable to the laboratory than a swab. Samples of the affected graft, any intravascular thrombus, and samples of the native vessel at the anastomosis should be sent, all in separate pots.

The presence of bacteria in biofilms presents problems for the accurate diagnosis of infection as these organisms are adherent to the graft and are often not present in sufficient numbers in perigraft tissues to be successfully cultured. Specimens of the infected graft that are sent to the laboratory should be processed in such a way as to maximize the growth of the causative organism. Disruption of the biofilm using high frequency ultrasound and culture of the material in broth rather than on solid media has been shown to increase the diagnostic yield.

Management

The optimal management of prosthetic graft infections is unknown. Most approaches involve some form of surgical procedure in conjunction with antimicrobial therapy. There are no randomized, controlled studies comparing the various approaches.

Aortic grafts

Infection of abdominal aortic grafts usually necessitates surgery. Surgery is mandatory for aortoenteric fistulae and for distal anastomotic rupture and is required for most types of local complication. Rarely, conservative management may be considered if there is perigraft sepsis in the absence of an obvious vascular crisis, particularly in elderly or debilitated patients. For those patients requiring surgery, the decision lies between radical excision combined with extra-anatomic bypass and local *in situ* repair or graft replacement combined with extensive debridement of all infected tissue.

The traditional teaching has been that serious infective complications, such as aortoenteric fistulas, should be treated with radical surgery. The fistula is repaired, the infected graft is totally excised and the aortic stump oversewn. The lower limbs are then revascularized by means of an axillobifemoral prosthetic bypass graft. This approach has the advantage of completely removing the infected prosthetic material and placing the new prosthetic graft in an uncontaminated, new site so that contiguous reinfection is unlikely. There are numerous problems with this approach. Prolonged operating times may lead to acute critical lower limb ischaemia. Constructing the extra-anatomic bypass first, before excision of the graft can eliminate this risk, but this might not be possible in a patient with active bleeding from an aortoenteric fistula. Residual infection in the aortic stump carries the risk of suture line disruption and sudden death from exsanguinating haemorrhage. Longer-term complications include thrombosis or narrowing of the long, axillary bypass graft with consequent critical lower limb or intermittent claudication. A ‘steal' of blood from the upper limb feeding the graft can result in muscle ischaemia and pain when the upper limb is exercised.

The many problems associated with extra-anatomic bypass grafting have led some surgeons to advocate *in situ* graft replacement or repair. Excision of the infected graft and replacement with a new graft is technically easier and quicker than graft excision and extra-anatomic bypass. This has the advantage that the peroperative limb ischaemia time is shorter and graft patency rates are greatly superior to those of axillobifemoral bypass. The major disadvantage is that there is a significant risk that the new prosthesis will become infected from the infected bed of the old graft. Perioperative antibiotic therapy cannot be guaranteed to prevent reinfection. The prosthetic grafts used in such procedures can be impregnated with antibiotics, such as rifampicin to inhibit infection of the new graft. Impregnation of grafts with antibiotics, while theoretically attractive, has not been shown to result in a lower infection rate. It is possible that the use of such antibiotic-impregnated grafts will simply select organisms resistant to the antibiotic used and these resistant bacteria will cause recurrent graft infection. Also, the antibiotic is leeched out of the graft and disappears within 36 to 48 h. In some cases, particularly when the distal limb of an aortofemoral or an aortobiiliac graft is infected, it is possible to excise the infected part of the graft, debride the native vessel at the anastomosis, and replace the excised segment with a new prosthetic graft. More recently, some authors have used femoral veins to construct aortic grafts; others have used aortic homografts.

There are few data in the literature relating to the antibiotic management of prosthetic graft infection. In many papers, although the surgery is described in detail, little attention is paid to either the microbiological flora implicated in the infection or to the antibiotics used. Although many infections are caused by either coagulase-negative staphylococci or by *S. aureus*, some are due to Gram negative organisms and some are caused by anaerobic bacteria. In the absence of good microbiological data, empirical antibiotic therapy needs to be broad spectrum. This is particularly true in cases of aortoenteric fistula when the operative field will inevitably be contaminated with bowel flora.

In Oxford, we have taken the view that prosthetic graft infections should be treated in a manner analogous to the treatment of infected orthopaedic prostheses. In orthopaedic surgery, best results are obtained using a two-stage approach whereby the infected prosthesis is removed, the infection is treated for a period of several

weeks, and then a new prosthesis is implanted. Success can sometimes be achieved using a one-stage procedure if parenteral antibiotic therapy is given for at least 4 weeks after surgery. In most patients with arterial graft infection, the viability of their limb is dependent on the blood flow through the graft. Graft excision without providing an immediate alternative blood supply is therefore, not usually an option. Reinfection of the new graft is liable to occur from the septic focus of the old graft site, even if a virgin, extra-anatomic placement site is chosen to re-establish limb blood flow. The chances of completely eradicating vascular graft infection by surgery are necessarily lower than when treating infected orthopaedic prostheses since debridement of potentially infected local structures is more limited. For these reasons, we administer intravenous antibiotics for 4 to 6 weeks and continue therapy with appropriate oral antibiotics for at least a further 12 months. The initial intravenous therapy is designed to achieve high levels of antibiotic in and around the graft in an attempt to sterilize the field and prevent the establishment of an infected biofilm on the graft. The oral antibiotics are used to suppress the growth of any residual bacteria which might persist in a sessile state in the bed of the old graft. Patients with any clinical or laboratory evidence of on-going inflammation need to remain on antibiotics indefinitely. Many such patients have been maintained in good health without further surgery. Others have required limited local surgery for anastomotic haemorrhage. Few patients have required further major surgery to remove and replace a reinfected graft with an extra-anatomic bypass.

Peripheral graft infections

The principles of managing infections of peripheral (infra-inguinal) grafts are similar to those involved with aortic graft infections. Ideally, the infected graft should be completely excised and replaced in conjunction with appropriate antimicrobial therapy. The use of vein autografts rather than prosthetic material can reduce the risk of reinfection. For non-healing groin wounds or if there is a sinus, surgical debridement and repair of the defect with a rotated tissue flap are required. Conservative surgical approaches combined with local irrigation of the infected graft with antibiotic solutions have been attempted in some centres but there are few data to support this approach. Because the groin is always heavily colonized with bacteria, the risk of reinfection is high and limb graft infections are difficult to eradicate, unless the groin can be avoided. As a result, although the mortality rate from infection of peripheral vascular grafts is lower than that of aortic graft infections, the risk of limb loss is high.

Further reading

Bergamini TM, Bandyk DF, Govostis D, Vetsch R, Towne JB. Identification of *Staphylococcus epidermidis* vascular graft infections: A comparison of culture techniques. *Journal of Vascular Surgery* 1989; **9**: 665–70. [The authors use a dog model of *S. epidermidis* graft infection to examine the most efficient techniques for microbiological diagnosis. Broth culture is favoured over agar methodology and best results are obtained if techniques, such as specimen grinding or ultrasonic oscillation, are used to disrupt bacterial biofilms.]

Braithwaite BD, Davies B, Heather BP, Eamshaw JJ. Early results of a randomized trial of rifampicin-bonded Dacron grafts for extra-anatomic vascular reconstruction. Joint Vascular Research Group. *British Journal of Surgery* 1998; **85**: 1378–81. [A report of a United Kingdom multicentre, randomized study of 279 patients who received either normal prosthetic grafts or rifampicin-soaked grafts for extra-anatomic bypass grafting. There was no significant difference between the two groups but the study had a low overall infection rate and was probably underpowered to detect a difference.]

Brown PM, Jr., Kim VB, Lalikos JF, Deaton DH, Bogey WM, Powell CS. Autologous superficial femoral vein for aortic reconstruction in infected fields. *Annals of Vascular Surgery* 1999; **13**: 32–6. [A novel technique is described in which the saphenous veins are used to re-fashion the aorta, rather than using prosthetic material, when treating infected aortic grafts.]

Costerton JW, Lewandowski Z, DeBeer D, Caldwell D, Korber D, James G. Biofilms, the customized microniche. *Journal of Bacteriology* 1994; **176**: 2137–42. [An excellent review of the nature, behaviour and importance of bacterial biofilms. This will help with the understanding of prosthetic device infection and its management.]

Henke PK, Bergamini TM, Rose SM, Richardson JD. Current options in prosthetic vascular graft infection. *American Surgeon* 1998; **64**: 39–46. [A retrospective series from one centre in the United States assessing all of their graft infections and their management over a 10 year period. The authors favour extra-anatomic bypass grafting over *in situ* replacement whenever possible.]

Maser B, Vedder N, Rodriguez D, Johansen K. Sartorius myoplasty for infected vascular grafts in the groin. Safe, durable, and effective. *Archives of Surgery* 1997; **132**: 522–5 [discussion 5–6]. [An American centre outlines the technique of ipsilateral sartorius myoplasty, along with perigraft debridement, for the management of peripheral graft infections. They describe 15 procedures in 14 patients with 100 per cent wound healing with an average of 36 months follow-up.]

17.13 Minimally invasive vascular surgery

Victor J. Weiss and Alan B. Lumsden

Endovascular instrumentation and technique
Needles
Guidewires
Hemostatic sheath
Catheters
Angioplasty balloons
Stents
Contrast agents
Intravascular imaging
Intravascular ultrasonography
Angioscopy
Endoscopic venous procedures
Saphenous vein harvest
Subfascial endoscopic perforator vein ligation
Endoluminal venous procedures
Venous occlusive lesions
Venous thrombolysis
Extraluminal arterial procedures
Laparoscopic aortic surgery
Endoluminal arterial procedures
Arterial occlusive disease
Endoluminal stent-grafts
Conclusion
Further reading

A minimally invasive approach to the treatment of diseases within the peritoneal cavity has become widely accepted by the surgical community since the initial reports of laparoscopic cholecystectomy in the mid-1980s. The benefits of this technique, including a shorter hospital stay and recovery time, are well recognized. A similar approach to the management of vascular disease was first described nearly 20 years earlier by Dotter and Judkins. The technique of percutaneous access for both the diagnostic and therapeutic management of vascular disease has been the basis for entire subspecialties, including interventional radiology, invasive cardiology, and, most recently, endovascular surgery.

The development of catheter-based and endoscopic technology has laid the groundwork for advances in related instrumentation; this has allowed the endovascular surgeon to operate via an intra-or extraluminal route. From a minimally invasive approach we are now able to treat the full spectrum of vascular pathology, including arteriosclerotic and thrombotic stenoses and occlusions, aneurysmal dilatations, and traumatic lesions. Many of these procedures have only recently been developed, and as such have not been investigated in a manner that would enable accurate comparison with the more traditional methods. Long-term follow-up for these procedures is also frequently lacking.

The objective of this chapter is to provide the reader with an understanding of the current state of endovascular surgery. Endovascular surgical instrumentation and operating room set-up are rather different from those in traditional vascular surgical procedures. As such, a description of the instrumentation necessary for the majority of endovascular procedures is discussed. Diagnostic and therapeutic techniques for both arterial and venous disease are highlighted, with long-term results and comparisons with the traditional surgical approaches provided where possible.

Endovascular instrumentation and technique

Needles

Needles are used for vascular access primarily from a percutaneous location. Their size will be dictated by the diameter of the guidewire used. Most often an 18G needle is used, as it will accept a 0.035 inch guidewire. The most popular access needle is the Seldinger, which can be used for single- and double-wall punctures (Fig. 1).



Fig. 1. Basic endovascular instrumentation: from left to right, Seldinger needle, hemostatic sheath and introducer, angled guide wire, and angioplasty balloon.

The technique for single-wall puncture requires a sharp, beveled needle tip and no central stylet. The anterior wall of the vessel is punctured, with pulsatile back bleeding indicating its intraluminal positioning. This method is most useful for graft punctures, in patients with abnormal clotting profiles, or if thrombolytic therapy is anticipated.

The technique for double-wall puncture requires a sharp inner stylet and a blunt outer cannula. The needle is placed through both the anterior and posterior walls of the vessel, the inner stylet is removed, and the outer cannula is slowly withdrawn until pulsatile back bleeding is noted.

Guidewires

Once the needle is intraluminal, as verified by pulsatile back bleeding, the guidewire may be advanced. This is always passed gently and under fluoroscopic guidance to avoid subintimal dissection or plaque disruption. There is a wide variety of guidewires, differing in their length and diameter, the shape of their tip, their ability to cross tight stenoses, and their overall stiffness. The most commonly used is the Bentson wire, whose straight, flexible tip and variety of sizes make it the most versatile of the guidewires. The ability to cross tight stenoses is facilitated by a guidewire with a hydrophilic coating, angled at the tip to permit steerability, and with a very pliable tip to prevent subintimal dissection. Stiff wires are useful in maintaining a position across a lesion, and as a guide for catheter exchange (Fig. 1).

Hemostatic sheath

The sheath is a device through which endovascular procedures are performed (Fig. 1). It protects the vessel from injury as wires and catheters are introduced. A one-way valve prevents bleeding through the sheath, and a side port allows contrast or heparin flushes during the procedure. Sheaths differ in diameter, with the most commonly used sheaths for percutaneous access having a 5–9 French inner diameter and larger sizes for use when the vessel is exposed surgically. Sheaths also vary in length, but the most commonly used are 10 cm. Longer sheaths are available for interventions in remote locations.

Catheters

There is a wide variety of catheters, differing primarily in the configuration of their tip. The many shapes permit access to vessels of varying dimensions and angulations. Catheters are used in angiography, to protect the passage of balloons and stents, and can be used to direct the guidewire through tight stenoses or tortuous vessels.

Angioplasty balloons

Balloons differ primarily in their length and diameter, as well as in the length of the catheter shaft (Fig. 1). As technology has advanced, balloons with lower profiles (i.e., the size that the balloon assumes upon deflation) have been manufactured. Balloons are used to dilate vascular stenoses (angioplasty), to deploy Palmaz stents (Cordis Inc., Miami, FA, USA), and to assist with additional expansion in self-expanding stents. The lower-profile balloons are less likely to get caught during withdrawal through stents and are easier to pull through sheaths.

Stents

Vascular stents are either self-expanding or balloon-expandable, the most common of which are Wallstent and Palmaz stents, respectively. The self-expanding Wallstent (Schneider, Minneapolis, MN, USA) is a flexible stent deployed by retracting a surrounding protective sheath. Its flexibility allows it to be deployed through a small-diameter introducer, but also accounts for its considerable shortening upon expansion. The Symphony stent (Boston Scientific, Natick, MA, USA) is a self-expanding stent composed of a nickel–titanium alloy known as nitinol. The unique property of this metal is thermal memory, the ability to assume a predetermined configuration upon exposure to body temperature. Currently, this stent is approved by the Food and Drug Administration (United States) for use in the biliary tree, but an approved vascular application is anticipated.

Contrast agents

Before an endovascular treatment is chosen for a particular lesion, the risk of administering intravenous contrast to the individual patient must be assessed carefully. A known allergy to contrast agents or iodine preparations should alert the clinician either to avoid the use of contrast dye or to pretreat the patient with corticosteroids and antihistamines. Of course, all necessary medications should be readily available in case of an unanticipated anaphylactic reaction. Patients receiving contrast while taking the oral hypoglycemic agent metformin occasionally experience profound metabolic acidosis. It is currently recommended to withhold metformin for 48 h prior to contrast administration.

Contrast agents are available in low or high osmolarity, and ionic and nonionic varieties. Low-osmolarity agents are 10 times more expensive than high-osmolarity agents but are less painful on injection, have less associated renal toxicity, and fewer allergic reactions. The incidence of severe adverse reactions to high- and low-osmolarity agents is equivalent at 0.01 per cent. Nonionic contrast agents have a lower associated incidence of adverse reactions, both severe and overall, including a lower incidence of nephrotoxicity in patients with prior renal impairment.

Intravascular imaging

Improvements in endovascular visualization have largely been born out of necessity. Approximately 70 per cent of atherosclerotic plaque is eccentrically positioned within the blood vessel. Because angiography is limited to a uniplanar ‘lumenogram', the images obtained can be rather misleading when attempting to evaluate a stenotic lesion. With increased experience in intravascular stenting, we have discovered that the apposition of the stent to the vessel wall and its location in relation to surrounding branches are poorly evaluated by angiography. Furthermore, angiography exposes the patient to the risks of both ionizing radiation and intravascular contrast. Nevertheless, contrast angiography remains the most common invasive method of vascular investigation for both diagnostic and therapeutic interventions.

Intravascular ultrasonography

Intravascular ultrasonography enables transmural visualization of blood vessels by an ultrasound transducer of 20 to 30 MHz. The luminal dimensions obtained correlate well with both histologic and angiographic findings. Intravascular ultrasound accurately identifies intimal flaps and dissections, and better defines the degree of luminal stenosis (Fig. 2). Perhaps its greatest role is in the assessment of endoluminal stents and in guiding the deployment of stent-grafts. It is free of the many limitations of angiography, including exposure to ionizing radiation or contrast. Furthermore, intravascular ultrasound overcomes the problems inherent in the visualization of a three-dimensional plane in a unidimensional plane by providing an intraluminal, circumferential view.

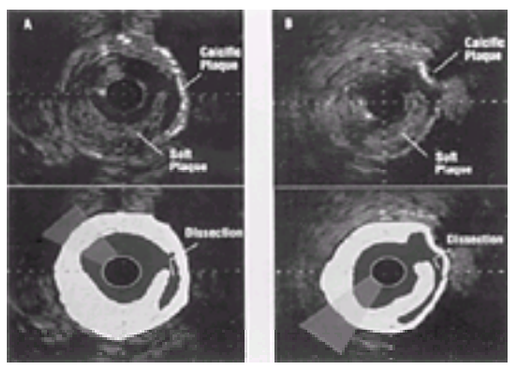


Fig. 2. Intravascular ultrasonograms of (A) calcific plaque and (B) soft plaque with associated dissection.

Catheters for intravascular ultrasonography range in diameter from 2.7 to 9 French, with transducers most commonly in a range from 12.5 to 40 Mhz. Larger vessels such as the aorta, in which penetration of the wall is necessary for adequate visualization, typically require lower-frequency transducers (12.5–20 Mhz). Intravascular ultrasonography provides an accurate evaluation of the anatomy of the vessel wall as well as of plaque morphology, based upon acoustic echogenicity.

From a diagnostic standpoint, intravascular ultrasound is helpful in clarifying equivocal angiographic findings, such as when they do not correlate with the symptom pattern. Intravascular ultrasonographic assessment of stenting has become routine in many centers, due in large part to the inaccuracies associated with angiographic assessment. Intravascular ultrasound used after stenting guided by angiography has demonstrated that the majority of stents are not fully deployed, a factor in contributing to early stent thrombosis (Fig. 3). Endovascular stents and stent-grafts can be placed using intravascular ultrasonography as the only intraoperative imaging, although it is more commonly employed as an adjunct to angiographic assessment. Because of its many advantages over conventional contrast angiography, intravascular ultrasound has secured a niche in vascular imaging.

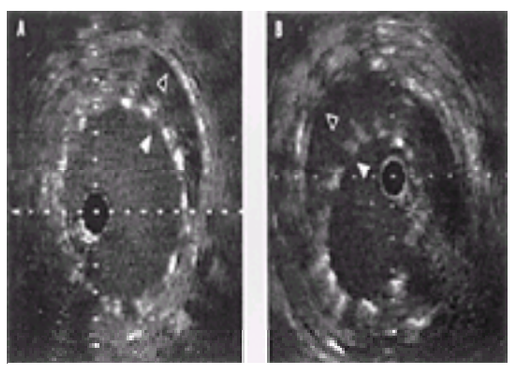


Fig. 3. Intravascular ultrasonogram of an arterial wall (black arrowhead) and an intraluminal stent (white arrowhead) showing imprecise apposition.

Angioscopy

Fiberoptic angioscopy has received renewed popularity in recent years. It can be used to assess the surface anatomy of plaque, it quantifies the degree of ulceration, differentiates between thrombus and embolus, and can be used to verify the results of vascular reconstruction. Angioscopic examination of thrombectomized vessels not only identifies critical stenoses but can also demonstrate residual thrombus missed by angiogram. Several investigators have found that angioscopy may reveal clinically important information not detected by extraluminal inspection, intraluminal probing or angiography in as many as 20 to 30 per cent of cases. With the increasing popularity of *in situ* saphenous vein bypass grafts, angioscopy has become the modality of choice to detect uncut valve leaflets, unligated side branches or significant anastomotic stenoses, which angiography is less able to diagnose. Whether the detection of these technical errors, allowing for their correction, will improve long-term results awaits to be seen.

Angioscopy has permitted a less invasive form of *in situ* saphenous vein bypass; Rosenthal and colleagues have devised two different methods for this. In the first, following angioscopic valvulotomy, an extraluminal, endoscopic-assisted approach is used for the ligation of venous tributaries, limiting the incision to the proximal and distal arterial anastomoses. The second, a variation on the first, is to place platinum coils to occlude the venous side branches from within the lumen of the saphenous vein using an electronically steerable nitinol catheter. Initial reports of both methods have demonstrated their feasibility and fewer wound complications. Their long-term patency, particularly with the likelihood of endothelial injury from excessive venous manipulation, remains to be seen.

Endoscopic venous procedures

Saphenous vein harvest

No surgical procedure is performed through a longer incision than the harvesting of saphenous vein for coronary or extremity bypass. This highly invasive procedure is often done in a leg with pre-existing vascular disease, with predictable consequences. In the only prospective study to date, healing of the harvest site was complicated in 24 per cent of patients undergoing coronary revascularization. While most problems consist of minor delays in wound healing, amputation has been necessary in a few patients suffering more catastrophic complications. Saphenectomy is a procedure that readily lends itself to a less invasive approach.

Endoscopic techniques have been applied to saphenous vein harvest with promising results. Through several small (3–4 cm) incisions, the retraction of overlying soft tissue, the dissection of the greater saphenous vein, and the clipping of venous side branches can be accomplished under direct visualization (Fig. 4). While this should theoretically result in fewer complications related to the length of the incision, experience with this technique is still early. The primary difficulty in endoscopic venous harvesting is working in a coaxial manner, which translates into a much longer time required for endoscopic saphenectomy (average 57.5 min) than for the open method (average 35.5 min). Only when the endoscopic procedure becomes technically easier with shorter operative times will it be more widely accepted.



Fig. 4. Endoscopic saphenous venous harvesting with camera and endoscopic dissector.

Subfascial endoscopic perforator vein ligation

Endoscopic venous exposure has become a popular method for subfascial endoscopic ligation of perforator veins in the treatment of chronic venous insufficiency with ulceration. This is a minimally invasive modification of the Linton procedure, where a long, medial incision in the calf is used to expose and interrupt the veins communicating between the superficial and deep systems. This operation achieved successful results in the healing of venous stasis ulcers, though wound complications and prolonged stays in hospital limited its acceptance.

Via a 10-mm endoscope, the subfascial space of the leg is entered, with a 10-mm port for the camera and a second port for dissection and clipping (Fig. 5). After carbon dioxide insufflation to 30 mmHg, perforating veins are identified and clipped. This endoscopic technique has allowed access to the perforating veins remote from the lipodermatosclerotic skin. Early reports demonstrate a 6 per cent risk of wound infection, with 88 per cent of ulcers healing at 5.4 month follow-up. While this appears to compare favorably with both the conventional Linton procedure and nonoperative therapy, long-term results are not yet available.



Fig. 5. Subfascial endoscopic perforator vein ligation with camera (surgeon's left hand) and scissors (surgeon's right hand).

Endoluminal venous procedures

Venous occlusive lesions

Like their arterial counterparts, venous occlusive lesions in all areas of the body have been treated with a combination of angioplasty, stenting, and thrombolysis. Patients with symptomatic occlusive lesions of the superior vena cava are often successfully treated with these modalities. Although nearly half have some recurrence of symptoms during follow-up, patency is readily restored with the use of thrombolysis and/or further stenting.

In 13 patients with symptomatic occlusive lesions of the superior vena cava (12 with known malignant disease), Crowe *et al.* achieved initial success in 11 patients, while two lesions could not be traversed with a wire. Although nearly half had some recurrence of symptoms during a follow-up period of 14 to 183 days, all had

patency restored with the use of thrombolysis and/or further stenting.

Those most commonly affected by venous occlusive lesions are hemodialysis-dependent patients. This is a large group that may benefit from a minimally invasive approach in the treatment of thrombosed access grafts, and an ideal model for surgeons to gain experience in endovascular techniques (Fig. 6). The endovascular treatment of thrombosed hemodialysis access grafts has largely been under the control of interventional radiologists, with a multitude of techniques for thrombectomy and thrombolysis being employed.

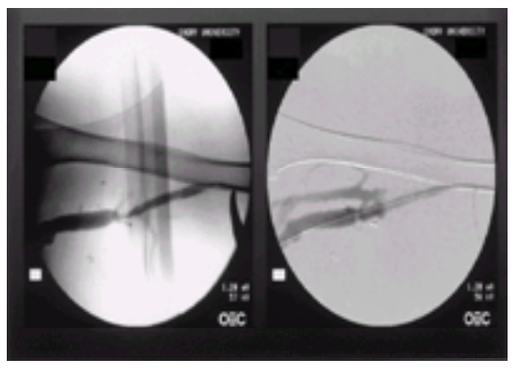


Fig. 6. Arteiovenous hemodialysis access graft venous anastomotic stenosis before (left) and after (right) balloon angioplasty.

Venous thrombolysis

Thrombolysis is a controversial treatment option for deep venous thrombosis. While anticoagulation is effective in preventing pulmonary embolus, many patients go on to suffer the consequences of the postphlebitic syndrome. It is well documented that lytic therapy leads to a more rapid and complete dissolution of clot than with heparin alone. Complete clot dissolution has been observed in approx. 35 per cent of those undergoing lysis as compared to fewer than 5 per cent of those treated with heparin alone. Short-term follow-up (3–6 months) has revealed the preservation of valve function in half of a thrombolytic group compared with approx. 7 per cent of a heparin-treated group. There appeared to be no differences in mortality or in the incidence of pulmonary embolism, though bleeding complications were seen in 17 per cent of the lytic group and 4 per cent of the heparin-treated group. Long-term follow-up of patients for an average of 6.5 years after they had randomly received either streptokinase or heparin for deep venous thrombosis demonstrated that only 20 per cent of the entire group had normal venograms, all those belonging to the streptokinase group. Three-quarters of patients in the lytic group were clinically normal on examination compared with one-third of the group treated with heparin. In another study, patients with proximal deep venous thrombosis treated with either heparin or streptokinase had no major hemodynamic benefit after a 2.5 year follow-up as measured by foot volumetry.

The optimal treatment for deep venous thrombosis is thus not completely clear. While its occlusive effects can quickly and effectively be treated with thrombolytic therapy, bleeding complications are significantly increased. Whether lytic therapy lessens the destructive effects of deep venous thrombosis on valve function and leads to significantly improved clinical outcome is the crucial question that remains to be answered.

Extraluminal arterial procedures

Laparoscopic aortic surgery

Laparoscopic exposure of the aorta for dissection of lymph nodes in both Hodgkin's lymphoma and gynecologic malignancy is a well-recognized method for tissue biopsy. Laparoscopically assisted aortic reconstruction in humans was first reported in 1993 and has since been detailed in sporadic case reports. Early approaches utilized peritoneal insufflation with carbon dioxide and a 10-cm minilaparotomy incision through which vascular control was achieved and aneurysmorrhaphy performed. Despite demonstrating feasibility, peritoneal insufflation resulting in gas embolism remains a concern. In one study, after creating a 1-cm incision in the vena cava of euvolemic dogs under pneumoperitoneum, 18 per cent were found to have gas bubbles in the right heart by transesophageal echocardiography, suggesting a prohibitive risk in vascular procedures performed under pneumoperitoneum.

Because of the concern over gas embolism during laparoscopic vascular bypass, a gasless technique of laparoscopically assisted aortobifemoral bypass employing a mechanical lifter for the abdominal wall has been described. This gasless technique provides less exposure and requires a larger minilaparotomy incision. One transperitoneal laparoscopic aortobifemoral bypass without minilaparotomy incision has been performed in a human. The procedure was not unexpectedly prolonged (total operative time 12 h 10 min, with an aortic cross-clamp time of 3 h 30 min) yet the patient had minimal requirements for analgesia and was discharged on the third postoperative day.

The technical challenge of the transperitoneal approach is primarily due to difficulty in adequate retraction of the small bowel. In order to circumvent this problem, a gasless approach via the retroperitoneum using the Laprolift (Origin MedSystems, Menlo Park, CA, USA) for retraction has been described. Laparoscopic-assisted repair of an abdominal aortic aneurysm with the creation of a 10-cm minilaparotomy incision has been reported as well.

Some of the difficulties with the transperitoneal laparoscopic approach to the aorta have been relieved by the use of a retroperitoneal approach assisted by the endoscopic retractor, including problems with retraction of the small bowel and the potential for gas embolism. However, aortic anastomosis does not easily lend itself to performance under the spatial confines of a laparoscopic procedure. As advances in laparoscopic instrumentation continue, laparoscopic anastomosis might become less technically challenging. Until that time, laparoscopic or laparoscopic-assisted aortic surgery remain in an experimental stage, with only case reports appearing in the literature. With the expansion of trials of stent-grafts for treatment of abdominal aortic aneurysms and the growing use of stents and stent-grafts in the treatment of iliac occlusive disease, enthusiasm has shifted from the laparoscopic-assisted aortic surgery to these other techniques.

Endoluminal arterial procedures

Arterial occlusive disease

Transluminal access to treat occlusive arterial disease is not a new concept; Dotter and Judkins first described this technique over 33 years ago. Since that time there has been an explosion in catheter-based technology, enabling the interventionalist access to treat occlusive disease in nearly any vascular bed. While the technique of balloon angioplasty has been applied to nearly all areas of the body, its success is often site specific and dependent upon several factors. Most importantly, lesion morphology is predictive of immediate and long-term success. ‘Ideal' angioplasty lesions are non-occlusive, short stenotic lesions. Patency improves when these lesions occur in larger vessels with a greater flow velocity.

The best long-term follow-up data are for iliac arterial angioplasty. The common iliac arteries are ideal for angioplasty, owing to their large diameter and high flow rates. Patency rates for iliac stenoses are 87 to 74 per cent at 1- to 4-year follow-up, respectively. However, when the initial treatment failures are included in these figures, patency rates fall to 78, 69, and 32 per cent at 1, 3, and 5 years, respectively. Totally occlusive lesions account for the majority of technical failures.

As vessel diameters and flow rates diminish, so do success rates for angioplasty. This correlation is exemplified by the differences in long-term patency after angioplasty of the common iliac and external iliac arteries. Following angioplasty of the common iliac artery, patency at 1 and 6 years is 81 and 52 per cent, respectively. Patency after external iliac arterial angioplasty at 1 and 4 years is 74 and 48 per cent, respectively.

Iliac arterial angioplasty is associated with 2 to 4 per cent major complications and 4 to 15 per cent minor complications. Many of these minor complications are related to the arterial puncture site (hematoma, pseudoaneurysm, etc.). Mortality from this procedure ranges from 0.3 to 0.8 per cent.

Compared with conventional bypass, common iliac angioplasty has a 10 to 20 per cent lower patency overall. The patency of aortoiliac bypass is 65 to 95 per cent at 5 years and 62 to 78 per cent at 10 years, based on a review of 19 previously published reports. Despite its lower long-term success, it is a useful procedure in patients with focal disease and mild symptoms in whom a major surgical revascularization is not justified. Angioplasty of the iliac vessels can be a useful adjunct to distal surgical bypass as well, increasing the success of distal revascularization and eliminating the risks associated with aortoiliac bypass. Thus, with long-term patency

slightly less than, but comparable to, open surgical bypass and with acceptable morbidity rates, angioplasty has become a well-accepted treatment for iliac occlusive disease.

Angioplasty of the infrainguinal vessels has not had the success that iliac angioplasty has enjoyed. Patency in this region is more dependent upon factors such as the presenting symptoms (claudication as opposed to limb-threatening ischemia) and the status of the distal run-off vessels, in addition to lesion morphology. Initial success in femoropopliteal angioplasty is 80 to 90 per cent, while failure to cross a lesion occurs in 7 per cent of stenoses and 18 per cent of occlusive lesions. While the initial technical success is better for stenoses than occlusions, long-term patency rates for stenoses and short occlusions are equivalent. Patients with optimal angioplasty lesions (stenoses less than 2 cm in length) appear to have slightly inferior long-term results than those obtained with surgical revascularization and with less morbidity and at a lower cost.

In a review of 12 published studies, the initial (30 day) success rate of femoropopliteal angioplasty for claudication was 89 per cent, while for limb salvage it was 77 per cent, but at 6 months, patency had fallen to 81 and 60 per cent, respectively. At 3 years, patencies of 62 and 43 per cent were seen. Overall, angioplasty of femoropopliteal lesions provides 5-year patency of 38 to 68 per cent. More recent studies have found similarly disappointing results with angioplasty of superficial femoral and popliteal arteries (performed primarily for claudication), with a 24-month cumulative patency of 46 per cent. The success of angioplasty is also related to the status of the distal vessels. If at least two tibial arteries are open, the 5-year patency for femoropopliteal angioplasty is from 53 to 77 per cent, but with poor run-off, patencies fall to from 31 to 59 per cent.

In most patients, infrapopliteal disease is extensive and multilevel, which is not well suited to angioplasty. Angioplasty of infrapopliteal short-segment stenoses is usually reserved for those with rest pain or tissue necrosis, provides acceptable results, and may be a viable option in appropriately selected patients. In the largest review of infrapopliteal angioplasty, technical success was obtained in 77 per cent of cases. Two-year success rates (improvement by at least one clinical stage) were 83 per cent for a single stenosis, 67 per cent for multiple stenoses, and 52 per cent for occlusions, with an overall 2-year patency of 62 per cent. The rate of major complications was rather high in this series, (11.3 per cent). Although percutaneous transluminal angioplasty is an important part of an aggressive, multidisciplinary approach to limb salvage, 81 per cent of patients with limb-threatening ischemia require operative treatment at some point, leaving only 19 per cent who can be treated with percutaneous transluminal angioplasty alone.

Proponents of angioplasty for short-segment disease list cost efficacy among its benefits over surgical endarterectomy. A cost analysis has compared balloon angioplasty with endarterectomy for short occlusions of the femoropopliteal segment. Following the primary procedure, 3-year patency was 87 per cent for the group undergoing endarterectomy and 44 per cent for the balloon angioplasty group. When evaluating the cost of maintaining patency by either method over a period of 3 years (the cost:patency ratio), initial failures of balloon angioplasty resulted in a significantly higher overall cost than for endarterectomy. In appropriately selected patients with ideal lesions located in higher-flow vessels, angioplasty is a viable alternative to bypass procedures from a cost standpoint.

Renovascular hypertension as a result of renal arterial stenosis is a disease in which angioplasty has been widely applied. The precise role of angioplasty and stenting in the renal vasculature has yet to be clearly defined however, primarily because of the variation among renal arterial lesions, which vary in location (main renal artery, branch lesions or osteal lesions) and etiology (atherosclerotic or fibromuscular dysplasia). These differences may well affect technical success and long-term patency.

Patients with fibromuscular lesions tend to have better long-term success and fewer procedural complications than those with atherosclerotic renal arterial disease. Tegtmeyer's group reported on 66 patients with 85 renal arterial lesions due to fibromuscular dysplasia. After a mean follow-up of 39 months, 39 per cent of the hypertensive patients were cured and 59 per cent were improved. Angioplasty appears to be the best initial treatment for such lesions if confined to the main renal artery, while bypass is more appropriate if the dysplastic lesions extend to branch vessels.

In the treatment of atherosclerotic renal arterial lesions, the exact location of the occlusive lesion influences the success of angioplasty. Martin and colleagues from Emory University compared the results of angioplasty of 129 osteal and 31 nonosteal lesions. At a mean follow-up of 38 months, 10 per cent of patients with osteal lesions were cured of their hypertension while 46 per cent were improved. In those with nonosteal lesions, 22 per cent were cured while 48 per cent were improved. Primary clinical success rates for these osteal lesions were 78 per cent at 1 year, 63 per cent at 2 years, and 54 per cent at 3 years; repeat angioplasty in 16 patients improved patency to 87, 75, and 66 per cent, respectively.

The indications for intervention can also effect success rates. Weibull reported on 71 renal arterial angioplasties, 57 per cent of patients having hypertension and azotemia, and 43 per cent hypertension alone. The overall technical success rate was 86 per cent. After a mean follow-up of 54 months, 15 per cent of hypertensive patients were cured and 75 per cent improved. For those with both hypertension and azotemia, 38 per cent were improved, 47 per cent un-changed, and 15 per cent worse.

Angioplasty has taken a primary role in the treatment of atherosclerotic and fibromuscular lesions of the main renal artery. The optimal management of osteal atherosclerotic lesions or fibromuscular lesions extending to branch vessels is less clear, with surgical bypass having better long-term success in exchange for greater morbidity.

Lesions of the supra-aortic truck are in certain aspects ideally suited to endovascular intervention. The mainstay of treatment for atherosclerotic occlusive lesions of the innominate and proximal subclavian arteries traditionally has been bypass or endarterectomy via median sternotomy. The potential advantage of decreasing morbidity by avoiding a chest-splitting incision appears great. The morphology of the innominate lesions is well suited to angioplasty, as these often occur as short-segment stenoses rather than occlusions. The prime concern focuses upon the risk of embolization to the cerebral circulation during angioplasty.

The surgical treatment of innominate occlusive disease is associated with a morbidity of 7 to 67 per cent and a mortality of 0 to 5.5 per cent. Over an average follow-up of 43 months, Cherry and colleagues at the Mayo Clinic found 13 per cent restenosis, while the Texas Heart Institute reported early graft occlusion in 7 per cent, failed endarterectomy in 9 per cent, and stroke in 2 per cent. These are the results to which angioplasty must be compared. The first report of angioplasty of the innominate artery appeared in the surgical literature in 1981 (Lowman *et al.*). In an attempt to protect the cerebral circulation from embolic debris, the right common carotid artery was exposed and clamped during angioplasty. Since that time, angioplasty has been performed without carotid occlusion with a very low risk of neurologic sequelae. The difficulty in evaluating innominate angioplasty is with the small number of published cases, precluding accurate comparison with surgical alternatives. Innominate angioplasty appears safe and avoids the potential morbidity of a thoracic incision. The risk of cerebral embolism remains uncertain however, and the precautions necessary to prevent neurologic deficit are undetermined.

Occlusive disease is more common in the subclavian than the innominate arteries, but the morbidity of surgical treatment is similar in both. Beebe reported a 23 per cent complication rate for both intra- and extrathoracic subclavian revascularization. Bypass is a durable procedure, with reported 5-year patencies of 83 to 94 per cent when using synthetic graft material. It has proved difficult to re-establish patency in completely occluded subclavian segments, with reported success rates of 0 to 56 per cent. If the procedure was initially successful, patency rates equivalent to that of surgical bypass have been demonstrated up to 3 years later. Thereafter, patencies for percutaneous transluminal angioplasty fall more rapidly to a 5-year actuarial value of 54 per cent. It has proved to be a safe procedure, with a major complication rate of 0.2 per cent and minor complication rate of 4.1 per cent reported for over 400 subclavian angioplasties. When treating subclavian ‘steal' syndrome, there is a delay of 20 s to 4 min in flow reversing from retrograde to antegrade in the vertebral artery. This acts to protect against embolization via the vertebral artery during subclavian percutaneous transluminal angioplasty. While totally occlusive lesions of the subclavian artery have been resistant in the past, the adjunctive use of thrombolytic therapy and intraluminal stents may improve upon the mediocre results.

The treatment of occlusive disease of the carotid artery by percutaneous angioplasty is probably the most controversial topic in vascular surgery today. The first large series appeared in the literature in 1987 (Theron *et al.*). In this report, 48 patients with atherosclerotic stenoses or postsurgical restenosis were treated. The procedure was successful in 45, with three complications including one case of monocular blindness, and two acute occlusions with one stroke. In 1990, this same author reported on an additional 13 patients treated with carotid percutaneous transluminal angioplasty using a triple-axial catheter system to protect against cerebral embolization. This technique was successful in preventing neurologic complications in this small group.

In 1991, Kachel reported on 43 percutaneous transluminal angioplasties for carotid stenosis of unspecified etiology, including 35 lesions of the internal carotid. Complications included one transient ischemic attack as a result of carotid spasm and one case of platelet deposition treated with emergent endarterectomy. There were no lasting neurologic symptoms or residual stenoses in an undefined percentage of patients followed for 3 to 109 months.

Bergeron and colleagues treated 38 patients with carotid percutaneous transluminal angioplasty, including 24 with stenosis of the internal carotid and seven with lesions of the common carotid. The majority of these lesions were postoperative restenoses (19 cases) and atherosclerotic (10 cases). In the entire series, there was one death due to hyperperfusion syndrome, one reversible stroke from an internal carotid dissection, and one transient ischemic attack due to thrombosis. Other complications included arterial spasm, acute occlusion necessitating redilation with stenting, and a silent cerebral infarct. All the complications occurred in the group treated solely with percutaneous transluminal angioplasty, whereas lesions that were stented were all successful and symptom free.

A comparison between angioplasty with stenting and endarterectomy was reported by Jordan *et al.* in 1997. There was no significant difference between the groups in the frequency of symptomatic lesions being treated. There was no difference in the risk of ‘major stroke' after percutaneous transluminal angioplasty/stent or carotid

endarterectomy (1.9 and 1.8 per cent, respectively), but the risk of ‘minor stroke' was significantly higher in the group treated with percutaneous transluminal angioplasty (6.5 per cent as compared with 0.6 per cent for carotid endarterectomy).

The role of angioplasty and stenting in the carotid artery is controversial. Many in the surgical community argue that carotid percutaneous transluminal angioplasty is an unproved treatment with unknown results and risks. Furthermore, it is being compared with perhaps the most intensively scrutinized operative procedure to date. In those patients at increased risk of operative complications (i.e., a history of ipsilateral neck dissection or radiation), endovascular therapy may well have an acceptable role in the treatment of carotid disease. Further refinement of techniques and instrumentation may permit carotid angioplasty eventually to become an accepted treatment.

Endoluminal stent-grafts

The most exciting of all the new vascular technologies is the transluminally placed, stented graft. This allows insertion, via an access site remote from the vascular lesion, of a new prosthetic vascular conduit, which is brought into position from within the vessel lumen. The prosthetic conduit is then anchored in place with appropriate fixation devices. This is a significantly less invasive approach to the treatment of a variety of vascular lesions, including aneurysms, traumatic pseudoaneurysms, and occlusions. The concept is based upon the work of Juan Parodi, an Argentinean surgeon who performed the firstendoluminal graft repair of an aortic aneurysm in 1990.

Since Parodi's initial paper, numerous reports have attested to the technical feasibility and short-term patency of stent-grafts placed for arterial trauma, for occlusive aortoiliac lesions, isolated aortic, iliac, and combined aortoiliac aneurysmal disease. In summarizing their initial experience with stent-grafts, Marin had a technical success rate of 91 per cent with both aneurysmal and occlusive lesions and 100 per cent with traumatic arterial lesions. Although follow-up was brief, 18-month primary and secondary patencies for aortoiliac occlusive lesions were 77 and 95 per cent, respectively.

Several different types of abdominal aortic endoluminal stent-grafts are currently being evaluated throughout Europe, and in the United States by Food and Drug Administration trials. These stents differ in graft composition, the location of the metallic support (internal or external, and fully supported or supported at the ends only), and overall design (tube or bifurcated). Currently, the most common reasons for precluding the use of these devices are unfavorable aortic anatomy (too short a proximal neck) or iliac anatomy (extreme tortuosity or severe occlusive disease). As stent-graft technology improves, a greater number of patients with aneurysms of the infrarenal abdominal aorta will become suitable candidates.

Conclusion

As biomedical technology continues to grow, we can anticipate that many of the devices used today will be improved upon. This will probably result in a greater ability to open completely occluded lesions, but whether this will translate into acceptable rates of patency will have to be answered. Technologic advance will also provide lower-profile instrumentation, permitting percutaneous access for procedures now requiring arterial cut-down (e.g. stent-grafting for abdominal aortic aneurysm). This raises the possibility that diseases now treated by vascular surgeons may soon be in the treatment realm of anyone adept at catheter-based intervention.

Vascular surgeons are becoming increasingly aware of the part that endovascular procedures will play in the care of the vascular patient. This fact, along with the biomedical industry facilitating product development and physician education, has led to rapid growth in a short time. The treatment of hemodynamically significant atheromatous lesions has suddenly become much easier for the surgeon and more tolerable for the patient. As surgeons become more involved in angiography, a host of lesions easily amenable to angioplasty will be identified. We are all familiar with the natural history of, and appropriate treatment for, these lesions from a surgical standpoint. The appropriate treatment of minimally symptomatic though significantly stenotic lesions becomes more problematic when one realizes how much more difficult they are to treat once they have gone on to occlude. This is one of the more difficult questions raised by this new modality. We will be forced to reappraise our indications for the treatment of specific lesions by a minimally invasive approach associated with fewer and less severe complications than with traditional open procedures. Until then, we must adhere to the current principles of vascular surgery and continue to investigate carefully the potential application of our new technology.

The versatility of endovascular surgery allows for the treatment (i.e., traditional surgical or catheter based) to be tailored to the specific type of lesion and the patient's characteristics. The endovascular surgeon is thereby ideally suited to care for the patient with peripheral vascular disease.

Further reading

Adar R, Critchfield GC, Eddy DM. A confidence profile analysis of the results of femoropopliteal percutaneous transluminal angioplasty in the treatment of lower extremity ischemia. *Journal of Vascular Surgery* 1989; **10**: 57–67. [A summary of 12 studies of angioplasty for infrainguinal disease.]

Ahn SS, Hiyama DT, Rudkin GH, Fuchs GJ, Ro KM, Concepcion B. Laparoscopic aortobifemoral bypass. *Journal of Vascular Surgery* 1997; **26**: 128–32. [The first report of laparoscopic aortobifemoral bypass in a human.]

Arnesen H, Heilo A, Jakobson E. A prospective study of streptokinase and heparin in the treatment of deep vein thrombosis. *Acta Medica Scandinavica* 1978; **203**: 457–63.

Arnesen H, Hoiseth A, Ly B. Streptokinase or heparin in the treatment of deep vein thrombosis: follow-up results of a prospective study. *Acta Medica Scandinavica* 1982; **211**: 65–8. [A mean 6.5 year follow-up streptokinase group had improved venograms and clinical outcome.]

Beebe HG, Stark R, Johnson ML, Jolly PC, Hill LD. Choices of operations for subclavian–vertebral arterial disease. *American Journal of Surgery* 1980; **139**: 616–23. [Describes the transition from intra- to extrathoracic procedures.]

Bergeron P, Chambran P, Bianca S, Benichou H, Massonat J. Traitment endovasculaire des artères à destinée cérébrale (échec et limité). *Journale des Maladies Vasculaire* 1996; **21** (Suppl. A): 123–31. [Early experience with angioplasty of supraaortic vessels details the risk with carotid angioplasty.]

Brewster DC, Moncure AC, Darling RC, Ambrosino JJ, Abbott WM. Innominate artery lesions: problems encountered and lessons learned. *Journal of Vascular Surgery* 1985; **2**: 99–112. [Suggests extrathoracic procedures are less durable than intrathoracic bypass.]

Brewster DC *et al.* Long term results of combined iliac balloon angioplasty and distal surgical revascularization. *Annals of Surgery* 1989; **210**: 324–30. [Demonstrates the utility of combined iliac angioplasty with more distal bypass.]

Brothers TE, Greenfield LJ. Long term results of aortoiliac reconstruction. *Journal of Vascular and Intervascular Radiology* 1990; **1**: 49–55. [Decreases long-term results with both endarterectomy and bypass.]

Bull PG, Mendel H, Hold M, Schlegl A, Denck H. Distal popliteal and tibioperoneal transluminal angioplasty: long-term follow-up. *Journal of Vascular Surgery* 1992; **3**: 45–53. [Describes lesions most and least amenable to angioplasty in the tibial vessels.]

Capek P, McLean GK, Berkowitz HD. Femoropopliteal angioplasty. Factors influencing long-term success. *Circulation* 1991; **83** (Suppl. I): I70–80. [Provides data for femoropopliteal angioplasty with mean 7 year follow-up.]

Chen MHM, Murphy EA, Halpern V, Faust GR, Cosgrove JM, Cohen JR. Laparoscopic-assisted abdominal aortic aneurysm repair. *Surgical Endoscopy* 1995; **9**: 905–7. [Details 10 patients with laparoscopic aortic surgery.]

Cherry KJ, McCullough JL, Hallett JW, Pairolero PC, Gloviczki P. Technical principles of direct innominate artery revascularization: a comparison of endarterectomy and bypass grafts. *Journal of Vascular Surgery* 1989; **9**: 718–24. [Provides evidence for trans-sternal repair as the procedure of choice.]

Childers JM, Hatch KD, Tran AN, Surwit EA. Laparoscopic para-aortic lymphadenectomy in gynecologic malignancies. *Obstetrics and Gynecology* 1993; **82**: 741–7. [Details a series of GI patients.]

Childers JM, Surwit EA. Laparoscopic para-aortic lymph node biopsy for diagnosis of a Hodgkin's lymphoma. *Journal of Surgical Laparoscopy and Endoscopy* 1992; **2**: 139–42. [Single case non-report for laparoscopy of para-aortic disease.]

Crowe MT, Davies CH, Gaines PA. Percutaneous management of superior vena cava occlusions. *Cardiovascular and Interventional Radiology* 1995; **18**: 367–72. [Provides good results with SVC occlusion due to malignancy.]

Dean RH, Hansen KJ. Renovascular hypertension. In: Moore WS, ed. *Vascular surgery. A comprehensive review*, pp. 521–41. Saunders, Philadelphia, 1998. [Details the diagnosis and management of renovascular hypertension.]

Dion YM, Gracia CR, Demalsy MC. Laparoscopic aortic surgery. [Letter] *Journal of Vascular Surgery* 1996; **23**: 539.

Dion YM, Katkhouda N, Rouleau C, Aucoin A. Laparoscopy-assisted aortofemoral bypass. *Journal of Journal of Surgical Laparoscopy and Endoscopy* 1993; **3**: 425–9. [First case report for this

procedure.]

Dion YM, Levesque C, Doillon CJ. Experimental carbon dioxide pulmonary embolization after vena cava laceration under pneumoperitoneum. *Surgical Endoscopy* 1995; **9**: 1065–9. [Report on surgically created IVC venotomy under pneumoperitoneum with 827 having no gas embolus.]

Dotter CT, Judkins MP. Transluminal treatment of atherosclerotic obstruction: description of a new technique and preliminary report of this application. *Circulation* 1964; **30**: 645–70. [Landmark paper describing this new technique.]

Farina C *et al.* Percutaneous transluminal angioplasty versus surgery for subclavian artery occlusive disease. *American Journal of Surgery* 1989; **158**: 511–14. [Long-term patency was better with bypass.]

Glovicki P *et al.* Safety, feasibility and early efficacy of subfascial endoscopic perforator surgery: a preliminary report from the North American registry. *Journal of Vascular Surgery* 1997; **25**: 94–105. [A preliminary report from the North American registry. J Vasc Surg, 1997; **25**: 94–105. Describes the technique and rationale for a minimally invasive approach to perforator ligation.]

Hebrang A, Maskovic J, Tomac B. Percutaneous transluminal angioplasty of the subclavian arteries: long-term results in 52 patients. *American Journal of Roentgenology* 1991; **156**: 1091–4. [Initial success is better with subclavian stenoses than occlusion.]

Johnston, KW. Femoral and popliteal arteries: reanalysis of results of balloon angioplasty. *Radiology* 1992; **183**: 767–71. [Discusses long-term patency and factors determining success with femoral and popliteal angioplasty.]

Johnston KW. Iliac arteries: reanalysis of results of balloon angioplasty. *Radiology* 1993; **186**: 207–12. [Angioplasty of common ilial arteries provide better long-term patency compared with external iliac arteries.]

Jordan WD, Schroeder PT, Fisher WS, McDowell HA. A comparison of angioplasty with stenting versus endarterectomy (CEA) for the treatment of carotid artery stenosis. *Annals of Vascular Surgery* 1997; **11**: 2–8.

Kachel R, Basche S, Heerklotz I, Grossmann K, Endler S. Percutaneous transluminal angioplasty (PTA) of supra-aortic arteries especially the internal carotid artery. *Neuroradiology* 1991; **33**: 191–4. [Provides over 700 cases of angioplasty for lesions of the supraaortic vessels.]

Kakkar VV, Flanc C, Howe CT. Treatment of deep vein thrombosis. A trial of aspirin, streptokinase and arvin. *British Medical Journal* 1969; **1**: 806–10.

Kakkar VV, Lorenz D. Hemodynamic and clinical assessment after therapy for deep vein thrombosis: A prospective study. *American Journal of Surgery* 1985; **140**: 54–63. [Shows no physiologic improvements at 2 years using streptokinase vs. heparin.]

Kotb MM, Kadir S, Bennett JD, Beam CA. Aortoiliac angioplasty: is there a need for other types of percutaneous intervention? *Journal of Vascular and Intervascular Radiology* 1992; **3**: 67–71. [Provides evidence that atherectomy provides no advantage over angioplasty.]

Krepel VM, van Andel GJ, van Erp WF, Breslau PJ. Percutaneous transluminal angioplasty of the femoropopliteal artery: initial and long-term results. *Radiology* 1985; **156**: 325–8. [Supports angioplasty for both stenoses and short occlusions of the femoropopliteal segment.]

Lowman BG, Queral LA, Holbrook WA, Estes JT, Bayly B. The treatment of innominate artery stenosis by intraoperative transluminal angioplasty. *Surgery* 1981; **89**: 565–8. [Details an early case report of innominate angioplasty via carotid exposure.]

Lumsden AB, Eaves FF, Ofenloch JC, Jordan WD. Subcutaneous, video-assisted saphenous vein harvest: report of the first 30 cases. *Cardiovascular Surgery* 1996; **4**: 771–6. [Highlights initial experience with a novel technique.]

Marder VJ *et al.* Quantitative venographic assessment of deep vein thrombosis in the evaluation of streptokinase and heparin therapy. *Journal of Laboratory and Clinical Medicine* 1977; **89**: 1018–29. [Uses a novel approach to gauge the efficacy of treatment for both thrombolysis and anticoagulation with DVT.]

Marin M, Veith FJ, Lyon RT, Cynamon J, Sanchez LA. Transfemoral endovascular repair of iliac aneurysms. *American Journal of Surgery* 1995; **170**: 179–82. [Describes a minimally invasive approach to a difficult surgical problem.]

Marin M *et al.* Transluminally placed endovascular stented graft repair for arterial trauma. *Journal of Vascular Surgery* 1994; **20**: 466–73. [Highlights the role for endovascular repair in the traumatic in the traumatized vessel.]

Marin M *et al.* Endovascular aortoiliac grafts in combination with standard infrainguinal arterial bypasses in the management of limb-threatening ischemia: preliminary report. *Journal of Vascular Surgery* 1995; **22**: 316–25. [Reports the utility of the combined approach using endovascular and traditional bypass.]

Marin M *et al.* Initial experience with transluminally placed endovascular grafts for the treatment of complex vascular lesions. *Annals of Surgery* 1995; **222**: 449–69. [Reports the early experience with the technique of endovascular repair.]

Martin LG, Cork RD, Kaufman SL. Long-term results of angioplasty in 110 patients with renal artery stenosis. *Journal of Vascular and Intervascular Radiology* 1992; **3**: 619–26. [Gives support to angioplasty in the management of renal artery osteal stenoses.

Matsuura JH, Dattilo JB, Poletti LF, Cohen IK, Posner MP, Makhoul RG. Minimally invasive vein harvesting: a comparison of endoscopic versus traditional open saphenectomy. *Vascular Surgery* 1997; **31**: 519–22. [Describes the advantages of endoscopic vein harvest.]

Motarjeme A, Keifer JW, Zuska AJ, Nabawi P. Percutaneous transluminal angioplasty for treatment of subclavian steal. *Radiology* 1985; **155**: 611–13. [Details the technique and success of angioplasty for subclavian steal.]

Parodi JC, Palmaz JC, Barone HD. Transfemoral intraluminal graft implantation for abdominal aortic aneurysms. *Annals of Vascular Surgery* 1991; **5**: 491–9. [A landmark paper describing the authors' initial experience.]

Porter JM, Seaman AJ, Common HH, Rosch J, Eidemiller LR, Calhoun AD. Comparison of heparin and streptokinase in the treatment of venous thrombosis. *American Surgeon* 1975; **41**: 511–19. [Compared both groups using venograms.]

Potkin BN *et al.* Coronary artery imaging with intravascular high-frequency ultrasound. *Circulation* 1990; **81**: 1575–85. [Describes the utility of IVUS in the coronary circulation.]

Reul GJ *et al.* Innominate artery occlusive disease: surgical approach and long-term results. *Journal of Vascular Surgery* 1991; **14**: 405–12. [Details 54 patients over 10 years undergoing innominate devascularization.]

Ringelstein B, Zeumer H. Delayed reversal of vertebral artery blood flow following percutaneous transluminal angioplasty for subclavian steal syndrome. *Neuroradiology* 1984; **26**: 189–8. [Documents the time course of low reversal with angioplasty for subclavian steal.]

Rosenthal D *et al.* Infrainguinal endovascular *in situ* saphenous vein bypass: ongoing results. *Journal of Vascular Surgery* 1994; **20**: 389–94. [Describes results with this novel approach to lower extremity bypass.]

Rosenthal D *et al.* Extraluminal endoscopic-assisted ligation of venous tributaries for infrainguinal *in situ* saphenous vein bypass: a preliminary report. *Cardiovascular Surgery* 1996; **4**: 512–14. [Initial experience for side branch occlusion from an extraluminal approach.]

Salam TA, Lumsden AB, Smith RB. Subclavian artery revascularization: a decade of experience with extrathoracic bypass procedures. *Journal of Surgical Research* 1994; **56**: 387–92. [Provides evidence for using the ipsilateral common carotid as the inflow for bypass.]

Scott RAP, Chuter TAM. Clinical endovascular placement of bifurcated graft in abdominal aortic aneurysm without laparotomy. *Lancet* 1994; **343**: 413. [Early experience with bifurcated stent grafts for AAA treatment.]

Seeger JM, Abela GS. Angioscopy as an adjunct to arterial reconstructive surgery: a preliminary report. *Journal of Vascular Surgery* 1986; **4**: 315–20. [Highlights the role of angioscopy in arterial surgery.]

Stanley B, Teague B, Raptis S, Taylor DJ, Berce M. Efficacy of balloon angioplasty of the superficial femoral artery and popliteal artery in the relief of leg ischemia. *Journal of Vascular Surgery* 1996; **23**: 679–85. [Describes poor initial success and patency at 24 months.]

Tabbara M, Kopchock G, White R. *In vitro* and *in vivo* evaluation of intraluminal ultrasound in normal and atherosclerotic arteries. *American Journal of Surgery* 1990; **160**: 556–60. [Describes the use of IVUS in the periphery.]

Tegtmeyer CJ, Selby JB, Hartwell GD, Ayers C, Tegtmeyer V. Results and complications of angioplasty in fibromuscular disease. *Circulation* 1991; **83** (Suppl. I): I55–61. [Looks specifically at fibromuscular disease of the renal artery with good results found.]

Theron J, Courtheoux P, Alachkar F, Bouvard G, Maiza D. New triple coaxial catheter systm for carotid angioplasty with cerebral protection. *American Journal of Neuroradiology* 1990; **11**: 869–74. [Describes a new technique for carotid angioplasty.]

Theron J, Raymond J, Casasco A, Courtheoux F. Percutaneous angioplasty of atherosclerotic and postsurgical stenosis of carotid arteries. *American Journal of Neuroradiology* 1987; **8**: 495–500.

[Highlights the use of an occlusion balloon as an adjunct to carotid angioplasty.]

Tsapogas MJ, Peabody RA, Wu KT, Karmody AM, Devaraj KT, Eckert C. Controlled study of thrombolytic therapy in deep vein thrombosis. *Surgery* 1973; **74**: 973–84. [An early experience using thrombolysis with DVT.]

Utley JR *et al.* Preoperative correlates of impaired wound healing after saphenous vein excision. *Journal of Thoracic and Cardiovascular Surgery* 1989; **98**: 147–9. [The first prospective look at saphenectomy wound healing following CABG.]

Veith FJ *et al.* Changing arteriosclerotic disease patterns and management strategies in lower limb-threatening ischemia. *Annals of Surgery* 1990; **212**: 402–12. [Provides evidence for an aggressive policy of bypass for limb salvage.]

Vroegindeweij D, Idu M, Buth J, Nillesen C, Schol FP, Tielbeek AV. The cost-effectiveness of treatment of short occlusive lesions in the femoropopliteal artery: balloon angioplasty versus endarterectomy. *European Journal of Vascular and Endovascular Surgery* 1995; **10**: 40–50. [Looks at cost-patency ratio with equivalency of the two methods despite worse long-term patency of angioplasty.]

Weibull H, Bergqvist D, Jonsson K, Hulthen L, Mannhem P, Bergentz SE. Long-term results after percutaneous transluminal angioplasty of renal artery stenosis: the importance of intensive follow-up. *European Journal of Vascular Surgery* 1991; **5**: 291–301. [Showed only 554 of patients needed a single angioplasty with remainder requiring repeat angioplasty or surgery.]

White GH, White RA, Kopchok GE. Intraoperative video angioscopy compared with arteriography during peripheral vascular operations. *Journal of Vascular Surgery* 1987; **6**: 488–95. [Highlights the advantages of angioscopy over angioplasty.]

Ziomek S, Quinones-Baldrich WJ, Busuttil RW, Baker JD, Machleder HI, Moore WS. The superiority of synthetic arterial grafts over autologous veins in carotid-subclavian bypass. *Journal of Vascular Surgery* 1986; **3**: 140–4. [Provides proof of improved patency with synthetic grafts over saphenous vein for carotid subclavian bypass.]

17.14 Thrombolytic therapy

Farzin Adili and Glenn M. LaMuraglia

Thrombolytic agents
Streptokinase
Urokinase
Tissue plasminogen activator
Indications for clinical use
Cardiac indications
Venous indications
Arterial indications
Intraoperative indications
Monitoring thrombolytic therapy
Complications of thrombolytic therapy
Future trends of thrombolytic therapy
Further reading

Inhibitors of coagulation shift the equilibrium of the clotting cascade, retarding the formation of thrombus with resulting slow resorption and organization of clot. However, their efficacy is limited for the treatment of thromboembolic disorders for which the fibrinolytic pathway has a greater role. Since identifying the fibrinolytic system, relentless efforts have been made to utilize safe but rapid activation of the fibrinolytic pathway to improve the treatment of thrombotic disorders. The main reaction of the fibrinolytic pathway is the activation of plasminogen to plasmin by the plasminogen activators, tissue plasminogen activator and urokinase (Fig. 1). Fibrin is the major substrate of plasmin and it also helps regulate fibrinolysis. Streptokinase, a bacterial product, not normally found in the body, is another potent activator of plasminogen and was among the first substances to induce fibrinolysis therapeutically.

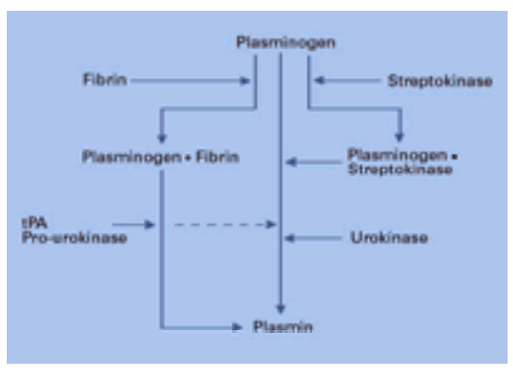


Fig. 1. The action of thrombolytic agents in the activation of plasminogen to plasmin.

Thrombolytic agents

Streptokinase

Streptokinase is a non-enzymatic single-chain polypeptide produced by Lancefield group C strains of b-hemolytic streptococci. It cannot directly transform plasminogen to plasmin, but forms a 1:1 stoichiometric complex with plasminogen (streptokinase–plasminogen activator complex) that activates another plasminogen molecule to form plasmin (indirect two-phase activation). Streptokinase complexes consist of both circulating and fibrin-bound plasminogen for conversion to plasmin. Initially, circulating plasmin is neutralized by a_2-antiplasmin, which is consumed, and the result is that plasmin exerts its proteolytic effect on plasma coagulation proteins, including fibrinogen and factors V and VIII. This consumption of fibrinogen and coagulation factors contributes to the bleeding risk of streptokinase. The half-life of streptokinase, which represents the least expensive of the thrombolytic drugs, is short (23 min); however, the effects of plasmin activation may persist for longer. Sensitized individuals bear antistreptococcal antibodies which inactivate streptokinase; the drug therefore has to be administered following a 100 000 to 250 000 IU loading dose. Intravenous doses of streptokinase average 100 000 IU/h, while local intra-arterial administration usually requires 5000 to 10 000 IU/h. Following an antithrombotic therapy with streptokinase, antibody titers rise and reach a peak after 7 to 10 days. Therefore, this therapy should be limited to 4 to 6 days and not be repeated before 9 to 12 months.

Urokinase

Urokinase is a double-chain glycoprotein and a trypsin-like serine protease isolated from human urine, where it is found in a concentration of approximately 6 IU/ml. It is now obtained from tissue cultures of human embryonic kidney cells and occurs in two molecular forms, S1 (32 kDa) and S2 (54 kDa), the former being a proteolytic degradation product of the latter. The low-molecular-weight form causes direct thrombolysis, whereas the high-molecular-weight form activates circulating plasminogen. Commercially available urokinase is a mixture of both forms. Its fibrinolytic potency is inferior to streptokinase. Since urokinase is an endogenous human protein, it is non-antigenic. It directly activates plasminogen and generates plasmin by cleaving the Arg560–Val561 bond of the plasminogen molecule. Urokinase is not specific for fibrin and activates fibrin-bound as well as circulating plasminogen. Thus, it also induces a systemic activation of the fibrinolytic system as soon as the a_2-antiplasmin level is reduced. The half-life of urokinase is about 16 min, so that the persistence of thrombolysis may be less than that for streptokinase. Intravenous administration is commonly started by a bolus of 4400 IU/kg followed by infusion of 440 IU/kg.h. Local intra-arterial doses vary between 1000 and 6000 IU/min.

Tissue plasminogen activator

To minimize the bleeding complications associated with the use of thrombolytic drugs, efforts have been undertaken to find plasminogen activators which are much more efficient in the presence of fibrin. This would mean that the activity of the thrombolytic drug could be limited to the thrombotic process, and little systemic fibrinolysis would occur. The benefits of the drug's clinical activity could thus be optimized, while hemorrhagic complications could be reduced.

Tissue plasminogen activator is the most important physiologic fibrinolysis activator in human tissues. It is synthesized predominantly by endothelial cells and represents a trypsin-like serine protease. In the absence of fibrin, tissue plasminogen activator is a poor plasminogen activator; however, in the presence of fibrin it activates plasminogen several hundredfold more rapidly, which makes it a more clot-specific fibrinolytic agent. The underlying mechanism is that both plasminogen and tissue plasminogen activator bind to fibrin and form a ternary complex. Thus fibrin serves as a cofactor for plasminogen activation. As a consequence of these characteristics, plasmin is predominantly generated on the fibrin surface. Despite high hopes of tissue plasminogen activator being a very specific drug for thrombus dissolution, initial clinical trials have been disappointing. In addition, plasma half-life of the drug is very short (approximately 5 min). Recombinant tissue-type plasminogen activator is mainly single chain and the only one currently on the market for clinical use.

Indications for clinical use

Thrombolytic therapy is indicated for the treatment of venous thrombosis, peripheral and cerebral arterial occlusion, and myocardial ischemia.

The clinical aim of thrombolytic therapy is to achieve a faster and more thorough dissolution of thrombus than can be achieved by anticoagulation alone. Most clinicians agree that anticoagulation is primarily of use in preventing thrombus propagation, not in thrombolysis.

Several important factors need to be considered in patients prior to thrombolytic therapy. First, there must be unequivocal evidence that a thrombus is causing the clinically important vascular occlusion. Clot lysis is most successful when the thrombus is fresh or only a few days old. When occlusions are arterial, the ischemia must not be severe enough to require emergency surgical intervention. There must be no contraindications to the use of the thrombolytic drug (Table 1): even with judicious

use of these agents complications occur and these risks need to be weighed against the potential benefits of thrombolytic therapy.

Risk of bleeding	*Peripheral and systemic bleeding*	*Intracranial bleeding*
High (absolute contraindication)	Surgery or biopsy (10 days) Severe trauma (6 weeks) Gastrointestinal bleeding (6 months) Active bleeding Aortic dissection Blood dyscrasia Pericarditis	Intracerebral tumor Neurosurgery Stroke (6 months) Head injury (1 month)
Medium (relative contraindication)	Puncture of a non-compressible blood vessel Cardiac life support, closed chest compression (>10 min) Anticoagulation Aneurysmal disease Atrial fibrillation Postpartum or pregnancy	Uncontrolled hypertension Stroke Anticoagulation
Low (no definite contraindications)	Diabetic retinopathy Cardiac life support, closed chest compression (<10 min)	Age > 70 years Hypertension

Table 1 Contraindications to thrombolytic therapy

Cardiac indications

Early reperfusion offers excellent potential for the salvage of ischemic myocardium early with acute transmural myocardial infarction. Necrosis can begin to occur in 20 to 30 min with total deprivation of blood flow or may be delayed for several hours because of residual blood flow via collateral circulation. Thrombolytic drugs may be used to re-establish principal blood flow as quickly as possible and improve collateral blood flow, preferably within 2 to 4 h after onset of symptoms of acute infarction. The clinical goal is for immediate clot lysis with reperfusion of the ischemic cardiac tissue to limit the extent of the infarction. Candidates for thrombolysis are patients presenting with acute myocardial infarct less than 6 h after the onset of symptoms with characteristic chest or myocardial ischemic pain lasting at least 20 min and accompanied by typical ECG changes in two or more contiguous precordial leads. The optimal method by which the goal of early reperfusion can be achieved is by the intravenous administration of thrombolytic drugs shortly after the onset of cardiac symptoms. Minimizing the delay between onset of symptoms and administration of the drug increases the likelihood of limiting the infarct: the thrombolytic drugs that hold the greatest appeal for this application are those with clot or fibrin ‘selectivity'. Four possible intravenous thrombolytic strategies are given in Table 2. Thrombolytic drugs have also been administered very successfully via the intracoronary route at the time of acute cardiac catheterization.

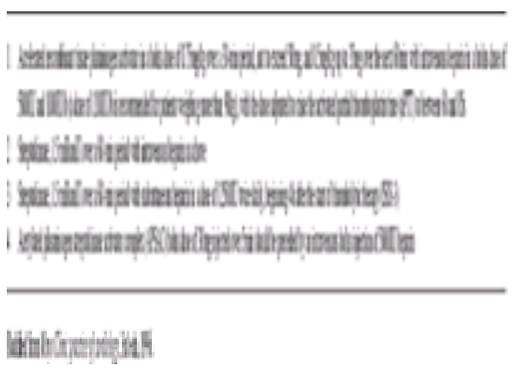


Table 2 Strategies for intravenous thrombolysis for cardiac indications

In the 70 to 80 per cent of patients in whom coronary flow is successfully re-established, anticoagulation diminishes the risk of reocclusion while the stenotic lesion that precipitated the original thrombosis is treated.

Venous indications

Venous thrombosis or thromboembolism is a common and potentially life-threatening problem. In cases of deep venous thrombosis, the objectives of therapy are inhibition of clot propagation and embolism and minimizing the venous injury precipitated by the thrombus, thus reducing the morbidity associated with venous insufficiency and the post-phlebitic syndrome. In the long term it is estimated that over 60 per cent of patients develop symptoms of the post-thrombotic syndrome following acute deep venous thrombosis. This syndrome, associated with obstruction or damage to the valves with subsequent venous hypertension, encompasses a range of symptoms and signs from pain and swelling to skin pigmentation and venous ulceration. The standard therapy of anticoagulants, bed rest, and elevation of the extremities effectively achieves the objective of inhibiting clot propagation and embolism. It does not, however, minimize the venous inflammation, scarring, and valve injury with subsequent venous hypertension.

Thrombolysis is an attractive treatment alternative to anticoagulation as it has the potential to lyse thrombus rapidly and may preserve valve function, thereby preventing the post-thrombotic syndrome in the long term. The effectiveness of thrombolytic drugs and anticoagulation in patients with deep venous thrombosis of the lower extremity has been compared in several studies. The pooled results indicate that thrombolytic therapy lyses 47 per cent of clots compared with 6 per cent lysis achieved with anticoagulation alone. The mode of acute treatment did not affect the incidence of pulmonary emboli. The incidence of bleeding complications was four times higher in patients treated with thrombolytic therapy than in those treated with anticoagulation alone. Although it is clear that thrombolytic drugs are useful and effective for the removal of acute thrombus from veins, the long-term benefit in reducing venous valve incompetence with thrombolytic drugs is more controversial. Several studies failed to demonstrate any significant benefit in patients treated with thrombolytic therapy after a follow-up of 2 years. With the advent of new thrombolytic agents such as urokinase and recombinant tissue plasminogen activator, however, lower dosing regimens for systemic infusion, and catheter-directed techniques of administration improved success and decreased complication rates were obtained. In 45 per cent of patients receiving recombinant tissue plasminogen activator, complete thrombolysis was achieved, compared with 5 per cent receiving heparin. Most encouraging have been recent reports of catheter-directed techniques which deliver the thrombolytic agent directly to the site of the thrombus. Approaching the thrombus from the internal jugular vein, complete lysis of iliofemoral deep venous thrombosis was achieved in 72 per cent and partial lysis in a further 20 per cent of cases with no major complications. Newer approaches from the posterior tibial and popliteal veins for catheter-directed thrombolysis are being investigated. However, the risk of pulmonary embolism during catheter-directed thrombolysis has so far not been established. One trial investigating the effectiveness of a new temporary vena cava filter in patients undergoing thrombolysis demonstrated clot within the filter in 25 per cent of patients. Until the risk of pulmonary embolism is properly determined it seems prudent to use a temporary filter in all cases.

The use of thrombolytic therapy in patients with a pulmonary embolus is controversial. The mainstay of therapy for acute minor pulmonary emboli remains heparin anticoagulation followed by several months of oral anticoagulation, and treatment of the precipitating cause of the venous thrombosis. In patients with acute massive pulmonary emboli, with significant hemodynamic compromise and pulmonary hypertension, thrombolytic therapy can cause a more rapid improvement in hemodynamics and in the angiographic appearance of the emboli. There is no evidence that its use improves mortality rates, and at 7 days there were no differences in lung scans of patients treated with heparin alone and those treated with thrombolytic drugs. Again, bleeding complications are more common in patients treated with thrombolytic drugs. In our practice we use thrombolytic therapy in those few patients without contraindications who have symptomatic pulmonary emboli with dyspnea and hypoxia, but not in patients with severe right heart failure, who should have surgical embolectomy. Most patients are therefore treated with heparin, followed by coumadin anticoagulation as clinically indicated.

Subclavian or axillary vein thrombosis is a specific type of venous thrombosis that can be successfully treated with thrombolytic therapy; its use in this setting is less controversial (Fig. 2). These thromboses can be precipitated by local intimal injury or effort thrombosis, thoracic outlet obstruction with venous compression, and by chronic indwelling venous catheters. Thrombolytic therapy has been used more frequently and with a higher degree of success in proximal upper extremity thrombosis for several reasons. First, the clot is usually localized to the axillary–subclavian distribution, without distal extension down the arm veins. There is not usually a large number of venous tributaries present, in contrast to the situation when thrombus is present in the lower extremities. Another important factor is that catheters can readily be inserted into the axillary–subclavian venous thrombus, allowing the use of high-dose local infusions that yield a higher lysis rate with a lower incidence of

complications. Once recanalization of the vein has been established, the precipitating factor for the venous thrombosis can be ascertained and appropriately treated.

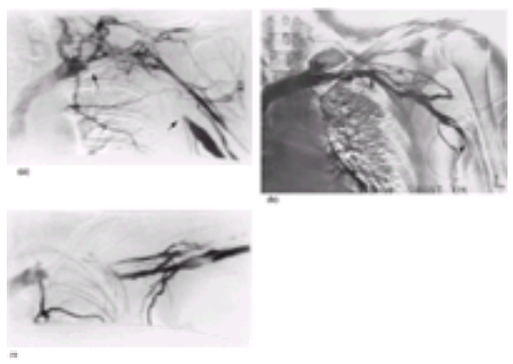


Fig. 2. Case demonstration of subclavian vein thrombolysis. (a) Thrombosed left axillary/subclavian vein (between arrows). (b) Recanalized axillary/subclavian vein through catheter injection (arrow). (c) Abduction venogram demonstrating thoracic outlet compression of vein.

Arterial indications

The use of thrombolytic therapy in patients with an acute peripheral arterial thrombosis or embolus is controversial. It can, however, be very useful in the properly selected patient with an acute presentation, in whom ischemic symptoms are not severe, where thrombus diffusely affects run-off, and when the risks for surgical intervention are high.

Before describing the criteria by which patients should be selected for arterial thrombolytic therapy, it is important to discuss its mode of administration. Historically, the drug was given intravenously to achieve a systemic thrombolytic state. Because the initial results were promising only for localized and acute artery occlusions, and were uniformly poor for thrombosed grafts, regimens of administration were altered to permit recovery of endogenous plasma levels of plasminogen. One of these protocols was ‘burst' therapy, which called for a high-dose intravenous administration of the thrombolytic drug for several hours followed by a recovery time of 12 to 24 h and a repeat of the cycle several times.

These intravenous methods have been largely replaced by intra-arterial infusion of the thrombolytic drug directly into the thrombus. Local injection of the drug into the thrombus can activate intraclot, or local plasminogen, increasing the rapidity of lysis without totally depleting systemic concentrations of plasminogen. Lower doses of thrombolytic agents can also be used with this route of administration. Technical success can also be predicted during such local administration by the ability to pass a guidewire through the occlusion in question. After thrombolysis, sheaths are already in place for balloon angioplasty, if the lesion is suitable for such treatment (Fig. 3).

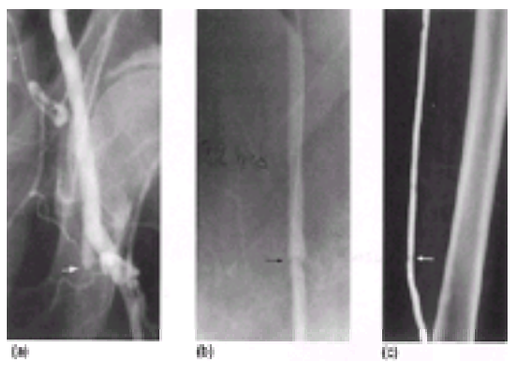


Fig. 3. Case demonstration of femoral–popliteal vein graft thrombolysis. (a) Left groin angiogram demonstrating proximal portion of the thrombosed vein graft (arrow). (b) After 22 h of thrombolytic therapy a vein graft stenosis is identified (arrow). (c) The vein graft after balloon angioplasty of the vein graft stenosis (arrow).

Although the technique by which thrombolytic therapy is administered is important, the aspect crucial to success is proper selection of patients. As well as recognizing an appropriate indication, it is important to determine that there are no contraindications to lytic therapy. Only patients who would otherwise be candidates for surgery because of their clinical presentations should be considered. This includes patients with an occluding thrombus precipitating a significant clinical end-organ ischemia and a clinical presentation that would have a high likelihood of benefiting from its use (Table 3).

Acute occlusions during balloon angioplasty or other interventional technique
Acute occlusion of native artery, especially with loss of run-off (e.g. popliteal aneurysm)
Acute occlusion of vein bypass graft
Acute to chronic occlusion of prosthetic bypass graft
Patient with multiple vascular reoperations
Acute thrombosis in the setting of a prohibitive surgical risk

Table 3 Cases likely to benefit from intra-arterial thrombolytic therapy

Recent arterial occlusions are clearly more responsive to thrombolytic therapy than well-organized, older clot. Therefore, a new thrombus that forms during an interventional technique or manipulation is the most likely to lyse. Acute occlusions of bypass grafts can be recanalized with thrombolytic therapy, followed by identification and treatment of the precipitating cause of the thrombosis: preoperative identification of the precipitating factor facilitates the subsequent procedure. Long-standing prosthetic bypass occlusions can be recanalized since they do not scar down as do thrombosed vein grafts but retain organized thrombus at either anastomosis with resorbed thrombus in their mid-portion. Thrombolytic therapy is often useful in acute arterial or graft occlusion in the presence of loss of outflow vessels (Fig. 4). Surgical thrombectomy of small or partially diseased arteries that have occluded as part of a graft or arterial thrombosis is often unsuccessful, and thrombolytic therapy can re-establish outflow before correction of the precipitating problem is attempted. In deciding whether to use thrombolytic therapy, other considerations include general surgical risk, complex anatomy, and scarring secondary to multiple operations. The Thrombolysis or Peripheral Arterial Surgery trial (TOPAS), the latest randomized, multicenter trial conducted at 113 North American and European centers compared vascular surgery, for instance thrombectomy or bypass surgery, with thrombolysis by catheter-directed intra-arterial recombinant urokinase in patients with an acute (less than 14 days) thrombotic or embolic occlusion of a leg. Despite its association with a higher frequency of hemorrhagic complications, intra-arterial infusion of urokinase reduced the need for open surgical procedures with no significantly increased risk of amputation or death. However, several problems are associated with this study. The study included patients with very different forms of acute limb-threatening ischemia, such as arterial embolism, native vessel thrombosis, prosthetic graft thrombosis, and vein graft thrombosis which may not be comparable with each other. The primary end point, which included an amputation-free survival of 6 months, implies a necessity for amputation if thrombolysis was not performed. This assumption can certainly not be regarded as evident since many of the patients' ischemic limbs may have also been saved without the use of either thrombolysis or surgery. Moreover, inclusion of the criterion ‘survival' as a primary end point seems problematic. Many premorbid patients with terminal illnesses which have also been included in the TOPAS trial also suffered from limb ischemia. In a significant number of patients survival in all likelihood may be

more closely related to the patient's condition at the entry into this study than to any treatment variable. In conclusion, the TOPAS trial did not conclusively demonstrate an improved clinical outcome of patients with thrombolytic therapy as compared with primary surgery.

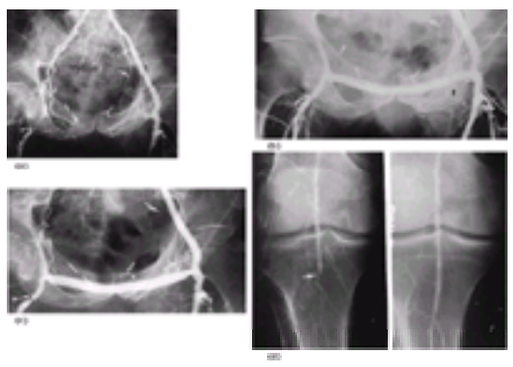


Fig. 4. Case demonstration of femoral–femoral prosthetic graft thrombolysis. (a) Thrombosed transpubic graft. (b) Partially recanalized transpubic graft with proximal thrombus (arrow) and right femoral stenosis (right arrow). (c) Totally recanalized graft after further thrombolysis. (d) Popliteal embolus (arrow) which occurred during thrombolysis, and was lysed after further drug administration.

Besides the usual contraindications to thrombolytic therapy, there are specific contraindications for patients with arterial occlusions. Since implementation of the procedure and infusion to achieve clinical success can take up to 48 h, patients in whom end-organ viability is seriously threatened should not undergo thrombolytic therapy, but should immediately be taken to the operating room. Patients with early postoperative occlusion of a bypass graft should not be treated with thrombolytic therapy because of the high likelihood of substantial bleeding from the surgical site. An embolus in a surgically accessible vessel is not a good indication, especially when the added potential risk of further fragmenting an intracardiac thrombus and precipitating further emboli is considered, though unlikely. The presence of knitted Dacron grafts is considered a relative contraindication because of the reported incidence of bleeding through the graft wall if thrombolysis is performed before the graft is incorporated by the surrounding tissue (2 to 4 weeks).

Results of thrombolytic therapy for arterial or bypass occlusion vary with the technique, drug, and route of administration, but probably depend most of all on selection of patients. Experience from several units suggested that intravenous administration of thrombolytic therapy resulted in successful lysis in 38 per cent of 340 patients, with an incidence of major complications of 8 per cent. Intra-arterial administration of thrombolytic drugs does not change the complication rate (13 per cent), but more than doubles the success rate, to 78 per cent in 478 patients reviewed. These studies and their methods differed significantly and they include many early studies, performed when the techniques were still evolving.

We exclusively use intra-arterial thrombolytic drugs for recanalizing occluded peripheral arteries or bypass grafts. Attempts at recanalizing native arterial occlusions are undertaken in patients with recent onset of symptoms and minimal atherosclerotic disease in the adjacent vessels, suggesting focal disease. Attempts at intra-arterial thrombolysis are more frequently performed in patients with recent occlusions of infrainguinal bypass grafts. These distal reconstructions are more likely to benefit from thrombolysis which allows the precipitating cause of the thrombosis to be determined, distal run-off that may have been lost at the time of occlusion to be re-established, and balloon angioplasty for correction of the underlying problem to be used. Suprainguinal bypass grafts, such as thrombosed limbs of aortofemoral grafts, present a different problem. Invariably, the occlusion results from a problem with the distal anastomosis, revision of which can be performed at the time of surgical thrombectomy of the limb of the aortofemoral graft. Thrombolysis is therefore of little benefit in this setting.

Recent trials of thrombolytic therapy in acute ischemic stroke have given conflicting results. The National Institute of Neurological Disorders and Stroke (**NINDS**) trial suggests that thrombolysis with tissue plasminogen activator given within 3 h of the stroke significantly reduced poor functional outcome. Based on this trial, the United States Food and Drug Administration has recommended the licensing of tissue plasminogen activator for use in ischemic stroke under the criteria of the NINDS trial. However, in a recent systematic review of 12 completed randomized trials which included 3435 patients, the risks of thrombolytic therapy for acute ischemic stroke were considered substantial and the benefit uncertain. The data appear to indicate that if early intervention of thrombolysis can be achieved in patients with intracranial (e.g. middle cerebral or basilar artery, Fig. 5) emboli or thrombosis, the prognosis is better than in attempts to open larger extracranial arteries such as the internal carotid or vertebral arteries.

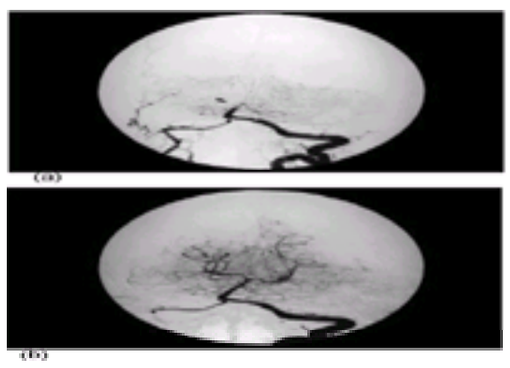


Fig. 5. Case demonstration of basilar artery thrombolysis. (a) Thrombosed basilar artery (arrow) by anteroposterior arteriography; note the distal midbrain and posterior cerebral vessels filling by collaterals. (b) Recanalized basilar artery after catheter-based thrombolysis

Intraoperative indications

Intraoperative thrombolytic therapy has been used when residual thrombus has been identified on intraoperative arteriography or when there is evidence of distal run-off loss. The reported series are small and hard to interpret since they are anecdotal, with no control group. Success of thrombolytic therapy is also variably defined, from angiographic criteria to clinical restoration of pedal pulses. Arterial flow in these patients was invariably restored after administration of the drug, and it is difficult to know how many of the ‘thrombotic blockages' were thin thrombus films that would have opened in response to arterial pressure. Our indications for intraoperative thrombolytic therapy are the angiographic presence of retained thrombus after thrombectomy or known loss of distal run-off. Reported success averages 69 per cent, with a 14 per cent incidence of bleeding complications, and an amputation rate of approximately 15 per cent. Both streptokinase and urokinase are effective, but we have used a regimen of two slow injections of 100 000 IU of urokinase into the clamped distal circulation separated by 10 to 15 min.

Monitoring thrombolytic therapy

Monitoring of thrombolytic therapy follows three specific goals. Before implementation of therapeutic thrombolysis the state of the coagulation system as well as exclusion of underlying coagulopathies is mandatory. Once thrombolytic therapy is begun, a lytic state has to be ensured, and hemorrhagic complications minimized.

Before thrombolytic therapy is instituted, a standard hematologic screen, including a complete blood count, prothrombin time, partial thromboplastin time, platelet count, thrombin time, and fibrinogen level should be performed. If streptokinase is used, quantification of the antistreptolysin titer is important since high levels of antibodies can inactivate the drug and make the therapy ineffective. There are no good laboratory tests which can predict lysis or hemorrhagic problems. Nevertheless, during thrombolytic therapy, a complete blood count, partial thromboplastin time, and thromboplastin time should be repeated at least every 8 to 12 h for quick diagnosis of thrombocytopenia and hemorrhage. The thrombin time, which measures the clotting time of plasma after the administration of thrombin, is an indirect measure of the fibrinogen and fibrin split products: these can now be measured directly.

Some evidence suggests that low fibrinogen levels are associated with a higher incidence of hemorrhagic complications and that the patients who have serious problems are those with fibrinogen levels below 0.05 g/dl. It is therefore recommended that fibrinogen levels arechecked and maintained above 0.05 g/dl, either by

slowing the infusion of thrombolytic agent or by the administration of cryoprecipitate.

Despite these published data, we have not used clinical laboratory testing during thrombolytic therapy other than a routine hematologic screen to exclude a coagulopathy. The results of these tests are difficult to obtain rapidly, and careful clinical evaluation in an intensive care setting for evidence of bleeding has resulted in minimal complications. In this setting, it is also recommended that patients who undergo thrombolysis should have urine analysis for early diagnosis of hematuria.

Complications of thrombolytic therapy

Some complications of thrombolytic therapy are common to all of the drugs, while others are specific to an individual drug (Table 4). Immunologic reactions such as pyrexia, anaphylaxis, serum sickness, or a rash are limited to streptokinase because of its antigenic source: such complications are treated by stopping the infusion, with supportive care to treat the symptoms.

Hemorrhage
Embolism
Pseudoaneurysm
Transgraft extravasation
Stroke
Compartment syndrome
Phlebitis
Myocardial infarction
Hepatic/splenic rupture
Loss of nails
Anaphylaxis, pyrexia, rash, or serum sickness (streptokinase only)

Table 4 Complications of thrombolytic therapy

The most common complication is hemorrhage. Because bleeding is unpredictable, its potential occurrence is the major criterion for exclusion of patients from thrombolytic therapy. Bleeding can occur at the site of drug administration, from other puncture sites, or from areas of recent surgery as well as spontaneously in other anatomic areas. Treatment consists of stopping the infusion of the thrombolytic drug, treating the hemorrhagic area as warranted, and, if necessary, transfusing blood products to replenish coagulation factors. If bleeding is serious, cryoprecipitate, the only source of fibrinogen, must be replenished along with fresh frozen plasma. Administration of e-aminocaprionic acid is not usually needed since the half-life of the thrombolytic agents is very short.

The administration of heparin during thrombolytic therapy to inhibit the formation of thrombus around catheters or keep the thrombus from propagating has been said to increase the incidence of hemorrhagic complications. However, with the use of intra-arterial catheters, heparin decreases catheter thrombosis and does not appear to increase the incidence of hemorrhagic complications.

Embolism is a potentially serious complication, which has been inconsistently reported, but may have an incidence as high as 50 per cent. It can occur as a result of thrombus formation around the catheters used in intra-arterial infusion, the partial dissolution of thrombus and distal embolization in the same vessel, or as a wash-over from the thrombus into a more proximal vessel secondary to the break-up of clot and infusion into an occluded artery. These emboli may be clinically silent; however, they can temporarily produce worsening of ischemia during the procedure. If emboli are identified, they are treated with continued thrombolytic therapy to the embolus, but if there is no clinical improvement and limb-threatening ischemia persists, surgical intervention may be required.

Future trends of thrombolytic therapy

As thrombolytic therapy has been used for over two decades, advances continue. New drugs may be more selective in activating clot-specific plasminogen without causing a systemic activation of the thrombolytic system. Most recently, novel gene transfer approaches to the prevention and therapy of arterial thrombosis have been developed. The potential of gene transfer for therapy of arterial thrombosis is suggested by *in vitro* data from studies in which endothelial cells were transduced with genes with antithrombotic activities and by experimental *in vivo* data showing both transgene expression in the vessel wall and successful inhibition of thrombosis. Local expression of a gene with antithrombotic activity provides a unique opportunity to decrease thrombus deposition or promote thrombolysis at specific arterial sites, while avoiding the systemic side-effects of conventional anticoagulant.

Further reading

Anderson HV, Willerson JT. Current concepts: thrombolysis in acute myocardial infarction. *New England Journal of Medicine* 1993; **329**: 703–9.

Armon MP, Hopkinson BR. Thrombolysis for acute deep vein thrombosis. *British Journal of Surgery* 1996; **83**: 580–1.

Arneson H, Hoiseth A, Ly B. Streptokinase or heparin in the treatment of deep vein thrombosis: follow-up results of a prospective study. *Acta Medica Scandinavica* 1982; **211**: 65.

Belkin M, Belkin B, Bucknam CA, Straub JJ, Lowe R. Intra-arterial fibrolytic therapy: efficacy of streptokinase vs urokinase. *Archives of Surgery* 1986; **121**: 769–73.

Bookstein JJ, Flemish B, Roberts A, Vallejo K, Davis G, Matched T. Pulsed-spray pharmacomechanical thrombolysis: preliminary clinical results. *American Journal of Roentgenology* 1989; **152**: 1097–100.

Durham JD *et al.* Regional infusion of urokinase into occluded lower-extremity bypass grafts: long-term clinical results. *Radiology* 1989; **172**: 83–7.

Goldhaber SZ *et al.* Randomised controlled trial of recombinant tissue plasminogen activator versus urokinase in the treatment of acute pulmonary embolism. *Lancet* 1988; **ii**: 293–8.

Guiliani ER, Gersh BJ, McGoon MD, Hayes DL, Schaff HF, eds. *Mayo Clinic practice of cardiology*, 3rd edn. Mosby, St. Louis, 1996.

Gurewich V, Pannell R. A comparative study of the efficacy and specificity of tissue plasminogen activator and pro-urokinase: demonstration of synergism and of different thresholds of non-selectivity. *Thrombosis Research* 1986; **44**: 217–28.

GUSTO Investigators. An international randomised trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 1993; **329**: 673–82.

Kakkar VV, Lawrence D. Hemodynamic and clinical assessment after therapy for acute deep vein thrombosis. A prospective study. *American Journal of Surgery* 1985; **150**(Suppl 4A): 54–63

LaMuraglia GM, Ortu P, Fillmore D, Obremski S, Athanasoulis C, Abbott WM. Selective laser photoablation enhances thrombolytic graft recanalization. *Surgery Forum* 1990; **41**: 363–4.

Maizel AS, Bookstein JJ. Streptokinase, urokinase, and tissue plasminogen activator: pharmacokinetics, relative advantages, and methods for maximising rates of consistency of lysis. *Cardiovascular and Interventional Radiology* 1986; **9**: 236–44.

McNamara TO, Fischer JR. Thrombolysis of peripheral arterial and graft occlusions: improved results using high-dose urokinase. *American Journal of Roentgenology* 1985; **144**: 769–75.

Meissner AJ *et al.* Hazards of thrombolytic therapy in deep vein thrombosis. *British Journal of Surgery* 1987; **74**: 991–3.

Ouriel K, Veith FJ, Sasahara AA. A comparison of recombinant urokinase with vascular surgery as initial treatment for acute arterial occlusion of the legs. *New England Journal of Medicine* 1998; **338**: 1105–11.

Parent FN, Bernhard VM, Pabst TS, McIntyre KE, Hunter GC, Malone JM. Fibrinolytic treatment of residual thrombus after catheter embolectomy for severe lower limb ischemia. *Journal of Vascular Surgery* 1989; **9**: 153–60.

Perrsson AV *et al.* Burst therapy: a method of administering fibrinolytic agents. *American Journal of Surgery* 1984; **147**: 531.

Rauwerda JA, Bakker FC, van den Broek TAA, Dwars BJ. Spontaneous subclavian vein thrombosis: a successful combined approach of local thrombolytic therapy followed by first-rib resection. *Surgery* 1988; **103**: 477–80.

Semba CP, Dake MD. Iliofemoral deep vein thrombosis: aggressive therapy with catheter-directed thrombolysis. *Radiology* 1994; **191**: 487–94.

Sharma GVRK *et al.* Thrombolytic therapy. *New England Journal of Medicine* 1982; **306**: 1268–76.

The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischaemic stroke. *New England Journal of Medicine* 1995; **333**: 1581–7.

Thery C *et al.* Use of a new removable vena cava filter in order to prevent pulmonary embolism in patients submitted to thrombolysis. *European Heart Journal* 1990; **11**: 334–41.

van Breda A *et al.* Local thrombolysis in the treatment of arterial graft occlusions. *Journal of Vascular Surgery* 1984; **1**: 103–12.

Van De Werf F *et al.* Coronary thrombolysis with tissue type plasminogen activator in patients with evolving infarction. *New England Journal of Medicine* 1984; **310**: 609–13.

Vassalli G, Dichek DA. Gene therapy for arterial thrombosis. *Cardiovascular Research* 1997; **35**: 459–69.

Wardlaw JM, Warlow CP, Counsell C. Systematic review of evidence on thrombolytic therapy for acute ischaemic stroke. *Lancet* 1997; **350**: 607–14.

17.15 Neointimal hyperplasia, thrombophilia, and graft thrombosis

Changyi Chen and Alan B. Lumsden

Neointimal hyperplasia
Clinical significance
Cellular and molecular mechanisms
Prevention and treatment
Thrombophilia
Definition and clinical features
Pathogenesis
Prevention and management
Graft thrombosis
Incidence
Etiologic considerations
Surveillance and prevention
Management
Further reading

Neointimal hyperplasia

Neointimal hyperplasia is a chronic structural change in the blood vessel that leads to formation of a thickened fibrocellular layer between the endothelium and the inner elastic lamina or on the luminal surface of prosthetic vascular grafts. It often appears as a response to blood vessel injury and can become hemodynamically important in situations of luminal narrowing, reduced blood flow, or thrombotic occlusion.

Clinical significance

Neointimal hyperplasia is a significant clinical problem that threatens every known vascular reconstructive procedure, including balloon angioplasty, atherectomy, endarterectomy, and the insertion of vascular grafts, and is responsible for 20 to 50 per cent of the clinical failures of all vascular interventions. Within 3 weeks to 6 months after what seemed to be a successful dilation, percutaneous transluminal coronary angioplasty can be followed by restenosis from 25 to 50 per cent of primary lesions. The development of eointimal hyperplasia in patients who undergo carotid endarterectomy has been well documented, with restenosis appearing in 10 to 30 per cent of those patients during the first year. Neointimal hyperplasia is also the primary etiology in approximately one-third of the failures that occur during the first 12 to 18 months following vein graft operations. In synthetic grafts, significant neointimal hyperplasia develops at the anastomoses or, just beyond, in the distal vessels. Expanded polytetrafluoroethylene (**ePTFE**) is the most frequently used prosthetic grafts for both middle-sized artery reconstruction and the creation of arteriovenous conduits for hemodialysis access. Neointimal hyperplasia has also been identified as the cause of failure in 21 per cent of ePTFE grafts. The failure of arteriovenous ePTFE grafts occurs, on average, 18 months following implantation and is often a consequence of venous anastomotic neointimal hyperplasia (Fig. 1).

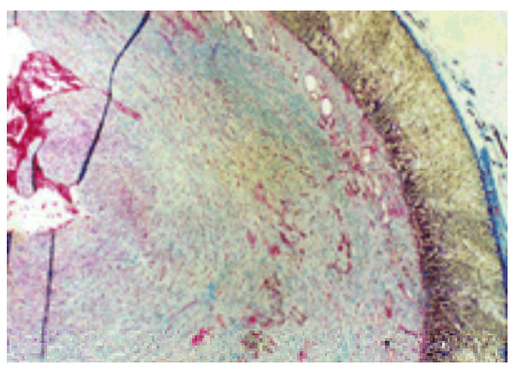


Fig. 1. Neointimal hyperplasia formation at the venous anastomosis of a human arteriovenous ePTFE graft for hemodialysis access. The original lumen was nearly completely occluded by the neointimal hyperplastic tissue. Widespread vascularization by capillary-sized vessels was seen within the neointima. Verhoeff–Massoon's stain: blue represents collagen staining, black represents elastin staining, and pink represents all other components. Original magnification was ×100.

Cellular and molecular mechanisms

The precise mechanisms which initiate the formation of neointimal hyperplasia are unknown. Much of our understanding of the cellular and molecular biology of neointimal hyperplasia has been gained from studies performed in animal models such as balloon denudation, angioplasty, endarterectomy, and the insertion of vein or prosthetic grafts. The balloon injury in the rat or rabbit common carotid artery has been studied extensively (Fig. 2). However, large animal models such as pigs, dogs, and baboons may be more clinically relevant in studies of neointimal hyperplasia, since their vessel structure and hemodynamics more closely resemble those of humans (Fig. 3, Fig. 4). In general, the development of neointimal hyperplasia after vascular injury involves medial smooth muscle cell proliferation (first wave), medial smooth muscle cell migration into the intima (second wave), and intimal smooth muscle cell proliferation and extracellular matrix production (third wave), with the final result being vessel wall thickening and luminal narrowing.

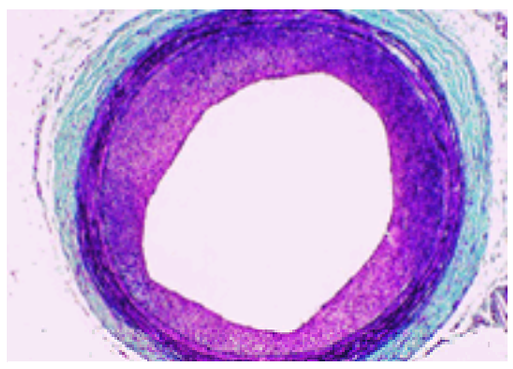


Fig. 2. Neointima hyperplasia formation in the rat carotid balloon injury model. The rat underwent balloon denudation on the left common carotid artery. The specimen was harvested at 2 weeks after surgery and subjected to histology examination. The cross-section of the artery shows a significant neointimal tissue development in this model. Verhoeff–Massoon's stain. Original magnification was ×100.

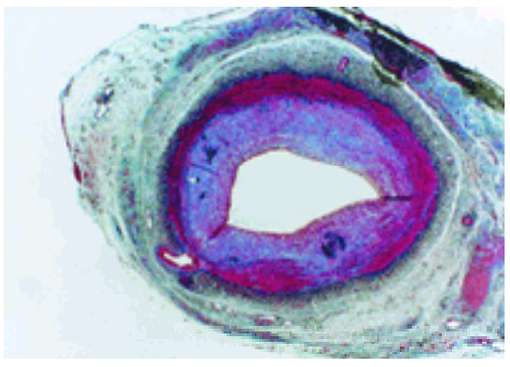


Fig. 3. Neointimal hyperplasia formation in the dog carotid endarterectomy injury model. A 1-cm endarterectomy was performed on the dog carotid artery by fine forceps. This procedure involved mechanical removal of the normal intima and a partial thickness of the media. The specimen was harvested at 11 weeks after surgery. Histology examination shows the significant neointimal hyperplasia development at the endarterectomy site. Verhoeff–Massoon's stain. Original magnification was ×20.

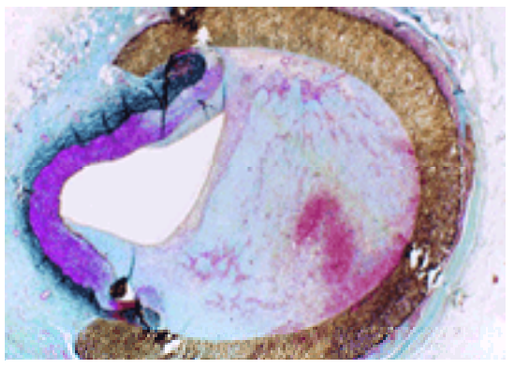


Fig. 4. Neointimal hyperplasia formation in the dog femoral artery ePTFE bypass graft model. The standard ePTFE graft (6 mm internal diameter and 6 cm long) was implanted into the dog femoral artery by end-to-side anastomosis. The intervening segment of the artery between the anastomoses was ligated. The specimen was harvested at 11 weeks after graft implantation. Histology examination was made at the midanastomosis. Massive neointimal tissue was seen in the luminal side of ePTFE graft. Verhoeff–Massoon's stain. Original magnification was ×20.

Vascular injury can be a consequence of biochemical substances (such as cholesterol, nicotine, or homocysteine), hemodynamic alter-ations (such as low shear stress or turbulent flow), and surgical interventions (such as angioplasty, endarterectomy, or vascular grafting). After vascular injury has occurred, the activation of smooth muscle cells is associated with a shift from a so-called contractile to synthetic phenotype, initiating a sequence of proliferation, migration, and synthesis of extracellular matrix. Smooth muscle cell proliferation begins immediately after the initial injury and may last for weeks or months.

The alteration of endothelial morphology and functions is usually associated with neointimal hyperplasia. In contrast to normal endothelial cells, the injured or regenerated endothelial cells grow as a sheet with close cell-to-cell contacts, are no longer aligned with blood flow, and are both polygonal in shape and irregular in size with cytoplasm budding toward the lumen. They also display an aberrant gene expression including upregulation of adhesion molecules such as intercellular adhesion molecule-1 and vascular cell adhesion molecule-1, and the reduction of endothelial-derived relaxing factors such as nitric oxide, which result in increased inflammatory cell infiltration and impaired endothelial-dependent relaxation. The increased production of extracellular matrix by neointimal smooth muscle cells is a very significant mechanism of neointimal expansion and lumen narrowing during the later stage of neointimal hyperplasia formation, with extracellular matrix component usually constituting about 80 to 90 per cent of neointimal volume.

Many factors have been implicated as regulators of the formation of neointimal hyperplasia. Platelet adhesion and aggregation at the site of vascular injury results in the release of the constituents of the platelets' a-granules. These granules contain numerous mitogenic substances such as platelet-derived growth factor (**PDGF**). Endothelial cells, macrophages, and smooth muscle cells may themselves secrete PDGF after arterial injury, and PDGF may be critical in the second wave of medial smooth muscle cell migration.

Basic fibroblast growth factor (**bFGF**) is another important component in accelerating neointimal hyperplasia formation. It lacks a classical signal sequence that targets it for secretion from cells after its synthesis. For this reason, bFGF accumulates within uninjured cells rather than being released immediately after synthesis. bFGF is released only after vessel injury, at which point it mediates the first wave of medial smooth muscle cell proliferation. Additional growth factors that may play a role in neointimal hyperplasia include epidermal growth factor, transforming growth factor, vascular endothelial growth factor, insulin growth factor type 1, angiotensin II, and endothelins.

A wide variety of hemodynamic factors, including high- and low-flow velocities, high and low shear stress, and mechanical compliance mismatch, have all been implicated in the formation of neointimal hyperplasia. These forces may either damage the endothelial cells or alter the gene expression and interaction between components of the vessel wall and elements of circulating blood.

Prevention and treatment

The multifactorial nature of the development of neointimal hyperplasia has led to a diversity of approaches, all of which attempt to prevent the undesirable vessel response to injury. By striving to minimize injury to blood vessels in vascular reconstruction procedures, the surgeon may be better able to achieve a positive clinical outcome. Experimental evidence has suggested that a wide variety of agents (such as warfarin, hirudin, heparin, antiplatelet drugs, angiotensin-converting enzyme inhibitors, calcium antagonists, steroids, cyclosporin, angiopeptin, lipid-lowering drugs, fish oils, and antioxidants) hold promise as inhibitors of neointimal hyperplasia, but none has consistently been shown to produce clinically significant results.

Unlike the human situation, most animal models involve arteries that are otherwise free of atherosclerosis. The target of arterial reconstruction in humans is the pre-existing, complex, atherosclerotic lesion, containing a variety of cell types including phenotypically modulated smooth muscle cells, monocyte–macrophages, and T lymphocytes. Unfortunately, such a scenario is poorly reproduced in animal models. In addition, the extent and depth of arterial injury may differ not only among the various animal models but also between animal and human. Effective pharmacologic doses may also vary between species due to differences in absorption, clearance, and end-organ effects. The results in one species may not be applicable across species types, and this lack of applicability may extend from animal models to human beings. Despite such limitations, studies using animal models are likely to remain an important tool for the testing of future therapeutic strategies against restenosis.

At present, there is no agent, substance, or technique clinically proved to prevent or reduce restenosis in humans. Future efforts of research will focus on gaining a better understanding of the mechanisms of human neointimal hyperplasia formation, and establishing new strategies to control neointimal hyperplasia. As more techniques become available that permit study of the human lesion as it evolves, we will learn as much about arterial injury in humans as we already know about arterial injury in animals.

Many potential strategies need to be evaluated in both clinically relevant animal models and human clinical trials. Various combinations of several agents or several classes of agents should be tested for their possible efficacy in overcoming the cascade of cellular events that accompany vascular injury. Local drug delivery techniques may bring higher concentrations of therapeutic agents to the site of vascular injury in order to achieve a maximum therapeutic effect, while minimizing the systemic side-effects. Photodynamic therapy utilizes the cytotoxic properties of light-excitable photosensitizers to destroy smooth muscle cells that would otherwise be involved in the proliferative neointimal lesion. These photosensitizers, which have no adverse effects in the absence of light, become cytotoxic when activated by an appropriate wavelength of light. Intraluminal irradiation may also prove to be an adjunctive treatment option for the prevention and/or treatment of the neointimal hyperplasia that occurs after vascular reconstruction procedures. Nitric oxide is now recognized as a principal component for the maintenance of vascular tone and has been proved to have multiple effects on vessel response to injury, including inhibition of platelet aggregation and leukocyte adhesion, and cytotoxicity of smooth muscle cells and macrophages. Such capabilities make it a logical and potentially powerful agent for controlling the process of neointimal hyperplasia formation. Many of the monoclonal antibodies which react against PDGF, bFGF, and adhesion molecules may be able to reduce neointimal hyperplasia formation by directly inhibiting smooth

muscle cell proliferation, migration, and inflammatory cell infiltration.

As we learn more about molecular events induced by vascular injury, it appears that the development of molecular-based strategies to inhibit neointimal hyperplasia formation may also be useful. These strategies include: recombinant chimeric toxin therapy, which is designed to target proliferating smooth muscle cells for cytotoxic molecules; antisense strategy, which is intended to inhibit expression of specific gene products necessary for cell proliferation; and gene therapy, which is developed to transfer exogenous genes into cells of the vascular wall.

Thrombophilia

Definition and clinical features

Thrombophilia is a hypercoagulable state associated with an increased risk of thrombosis and can be acquired and congenital. The thrombosis associated with thrombophilia may be venous or arterial. Although thrombosis occurs most commonly in the deep veins of the lower limb, atypical sites such as adrenal, axillary, hepatic, renal, and ocular veins are also reported. In cases of arterial thrombosis, the coronary, retinal, mesenteric, peripheral, and intracranial vessels can be involved. Arterial thrombosis affecting intracranial vessels presents with transient ischemic attacks or cerebrovascular accident and, if recurrent, multi-infarction dementia or psychosis.

Another common clinical feature of thrombophilia in women is recurrent miscarriage. The fibrinoid necrosis and inflammation of decidual blood vessels that often results from a series of miscarriages may cause thrombosis and inadequate maternal blood supply through the diseased vessels. If such conditions occur during the first trimester, the physiologic vasodilator adaptation of the spinal arteries could also fail.

On occasion, thrombophilia is also associated with thrombocytopenia, hypertension, pulmonary hypertension, post-phlebitic syndrome, leg ulcers, and warfare-induced skin necrosis.

Pathogenesis

Acquired thrombophilia can be caused by antiphospholipid syndrome, autoimmune diseases such as systemic lupus erythematosus and ulcerative colitis, endocrine conditions such as diabetes mellitus and Cushing's syndrome, and myeloproliferative disorders such as polycythemia vera. It may also be observed in other conditions, including nephrotic syndrome, liver disease, heparin-induced thrombocytopenia, malignancy, pregnancy, and combined oral contraceptives (especially those containing third-generation progestogens).

The antiphospholipid syndrome, which may be considered to be the combination of clinical features found in the presence of an antiphospholipid antibody, is the most common cause of acquired thrombophilia. The pathogenesis of thrombosis in this syndrome remains poorly understood, but may involve some of the following processes. As indicated by the increased release of platelet products, antiphospholipid antibodies may be associated with platelet activation, although a direct link with thrombosis is not demonstrated and the specific epitope(s) is/are uncertain. Prostacyclin (PGI_2) is a potent vasodilator produced by cells using arachidonic acid from cell membrane phospholipids, a process that requires phospholipase A_2. Antiphospholipid antibodies may be able to impair release of arachidonic acid from the phospholipid membrane through the inhibition of phospholipase A_2. Antiphospholipid may also cause platelet destruction, resulting in a secondary thromobocytopenia.

Protein C is a naturally occurring anticoagulant which is activated by thrombin in association with the endothelial receptor thrombomodulin. In combination with the cofactor protein S, activated protein C inactivates the active clotting factors Va and VIIIa, thereby limiting the coagulation cascade.

Antiphospholipid antibodies may also affect the protein C–protein S system by either binding to the thrombomodulin required for protein C activation or by binding to the phospholipid sites required for protein S-dependent anticoagulation activity of activated protein C, thereby blocking the inactivation of the active coagulation factors.

Antithrombin III is another naturally occurring anticoagulant that is activated by glycosaminoglycans on the surface of endothelial cells. Various studies have shown that certain antiphospholipid antibodies found in up to 11 per cent of patients bind to these structures, consequently preventing the glycosaminoglycan-dependent activation of antithrombin III.

Congenital thrombophilia is associated with a heritable deficiency of either antithrombin, protein C, protein S, or activated protein C resistance. Antithrombin deficiency is inherited as an autosomal-dominant trait which has equal expression in males and females; the estimated prevalence in the general population is between 1 in 2000 and 1 in 5000 individuals. Heterozygous patients have antithrombin levels between 30 and 70 per cent of normal. Homozygous antithrombin deficiency is lethal *in utero*. Two major types of heritable antithrombin deficiency are recognized. Type 1 defects are characterized by the quantitative reduction of qualitatively (functionally) normal antithrombin, with further laboratory testing revealing that there is a concordant reduction in functional activity and protein antigen assay results. In type 2 defects, however, functional and protein antigen assay results are discordant because the production of qualitatively abnormal antithrombin protein creates a protein antigen level that is higher than the functional activity.

Like antithrombin deficiency, protein C deficiency has also been classified into type 1 (quantitative) and type 2 (qualitative) abnormalities. This genetic defect has been characterized in a large number of patients with protein C deficiency, with both clinically recessive and dominant forms of protein C deficiency being recognized, the clinically recessive type is the more common of the two.

Three types of protein S deficiency are recognized: type 1, low plasma total and free protein S; type 2, functional defect; and type 3, low free protein S. At least 32 causative mutations in the protein S gene have been reported, a relatively small number in comparison to antithrombin and protein C mutations which owes to the complexity of the protein S gene. The vast majority of patients with familial activated protein C resistance have the same mutation in their factor V gene (1691 G to A) which destroys one of the activated protein C cleavage sites in activated factor V by the replacement of Arg 506 by Gln. This is the so-called factor V Leiden mutation.

Prevention and management

Prevention of thrombosis

Other risk factors such as smoking, obesity, immobility, and the combining of oral contraceptive pills should definitely be avoided. In the short term, prevention involves the use of graduated elastic stockings and, at times of increased risk, subcutaneous unfractionated or low-molecular-weight heparin. For long-term prevention of thrombosis, patients with thrombophilia with previous episodes of spontaneous thrombotic events may be treated with oral anticoagulants. The benefit of these agents has to be balanced against inconvenience, cost, and most importantly, the risk of bleeding whilst on warfarin.

Management of acute thrombosis

The management of acute thrombosis in the patient with congenital thrombophilia is usually identical to that of patients with acquired thrombophilia. Following an initial bolus of 5000 U intended to maintain the APTT ratio at 1.5 to 2.5, unfractionated heparin is infused at 1300 U/h. Warfarin is introduced on the second day of heparin treatment; in order to avoid the risk of warfarin-induced skin necrosis, a loading dose smaller than the usual daily 10 mg is given. Heparin is administered for at least 5 days or until the International Normalized Ratio is greater than 2.0. The therapeutic range for first thrombotic events is around 2.0 to 3.0. There are rare occasions where patients with antithrombin deficiency require very high doses of heparin to achieve adequate anticoagulation, a phenomenon known as heparin resistance. Using a therapeutic ratio of around 2.0 to 3.0, anticoagulation should be induced for 6 months in patients with thrombophilia and a first thrombosis. Life-long anticoagulation should be offered to patients with more than one thrombosis.

Thromboprophylaxis in pregnancy

Clearly, low-dose aspirin is safe for use during pregnancy and is also an effective antiplatelet therapy in the prophylaxis of thromboembolism in a surgical setting. It is the preferred treatment modality for low-risk patients. Without more rigorous thromboprophylaxis, however, a pregnant woman who has already experienced a thrombotic event is at high risk of recurrence. Therefore, subcutaneous heparin is recommended in addition to low-dose aspirin; either 10 000 U of unfractionated heparin can be self-administered subcutaneously twice a day, or a low-molecular-weight heparin such as enoxaparin at a dosage of 40 mg can be self-administered subcutaneously once a day. Heparin does not cross the placenta or pass into breast milk, making it safe for use in pregnancy as far as the fetus is concerned. Women taking heparin for more than 6 weeks should be warned, however, that there is a risk of thrombocytopenia and osteopenia.

Graft thrombosis

Incidence

Graft thrombosis is the ultimate complication of arterial reconstruction procedures, although other complications such as graft sepsis or distal thromboembolism may have more dire consequences. The success of a bypass procedure is traditionally measured by graft patency rates, and the graft patency rates in peripheral reconstructions are summarized in Table 1. Usually, autologous saphenous vein grafts have better outcomes than prosthetic grafts. Many factors have been reported to be involved in graft failure, with the factors changing as the length of time from implantation increases.

Conduit		*Five-year patency rate (%)*
Saphenous vein		
	Above-knee popliteal	72–77
	Below-knee popliteal	61–82
	Femorotibial	49–69
ePTFE		
	Above-knee popliteal	38–65
	Below-knee popliteal	54–60
	Femorotibial	12–22
Dacron		
	Above-knee popliteal	57–71
	Below-knee popliteal	32–57
	Femorotibial	NA

Table 1 Long-term patency rate of vascular grafts.

Etiologic considerations

Operative technique and patient selection

Graft occlusion within 1 week of operation and particularly within 2 to 3 days is generally attributed to technical error and poor patient selection. Causes include deficient suture line construction, twisted grafts, improper graft placement and tunneling, faulty judgment of adequate outflow or inflow, inferior saphenous vein (too small, thick walled, varicosed, or phlebilic), and improper branch ligature placement. A poorly constructed suture line that obstructs flow is the most frequent technical problem resulting in thrombosis. Occlusions that occur within the post-procedure span of the first few days through the first month may well be due to more subtle technical imperfections, persistent underlying disease, or a surface thrombogenicity of the graft that exceeds its thrombotic threshold. Grafts with flow rates lower than 100 ml/min have been shown to have a lower patency than those with higher flow rates. Resistance measurements have also been correlated with graft patency.

Graft surface thrombogenicity

Properly harvested autologous veins slough their endothelial surface for only a brief period, making their increased surface thrombogenicity rather transient. In contrast, the surfaces of prosthetic grafts reduce their thrombogenicity after a period of weeks or months, though unfortunately they never achieve a reduced thrombogenicity comparable with that of autologous veins. Abnormal platelet consumption on prosthetic bypass grafts rarely subsides until 6 months after implantation and may never return to normal. In humans, such grafts do not undergo significant endothelialization but are lined with a layer of compact fibrin. Although this layer has a negative charge and is much less thrombogenic than the initial prosthetic surface, it does not compare with a healthy endothelial layer in resisting thrombosis.

Neointimal hyperplasia and progressive atherosclerosis

Coronary saphenous vein grafts develop two fairly distinct lesions. Within 1 to 18 months after grafting, neointimal hyperplasia is often present at grafts as well as anastomoses. After 12 to 18 months, atheromatous lesions appear superimposed upon the neointimal thickening. These lesions contain large concentrations of T cells and macrophages in the subendothelial space, as well as foam cells, calcification, and regions of necrosis. Peripheral saphenous vein grafts display the same pathologic and physiologic alterations as coronary grafts. Prosthetic grafts often develop neointimal hyperplasia at the anastomoses. Many factors, including foreign body response to the prosthetic material, platelet activation and release, compliance mismatch, and flow disturbances at the perianastomotic regions, are responsible for this anastomotic lesion formation. Aggressive neointimal hyperplasia is a particular problem at the venous anastomosis of the arteriovenous graft. In addition, progressive atherosclerosis in the native vessels may occur either below the distal anastomosis or beyond the proximal anastomosis of the graft, thereby reducing blood flow and promoting graft thrombosis to a significant degree.

Graft structural abnormalities

The study of autologous vein femoropopliteal grafts concluded that approximately 33 per cent of these grafts developed various structural abnormalities during the 10-year follow-up period, including neointimal hyperplasia and fibrotic valves in the intermediate period and atheromatous narrowing and aneurysm formation in the much later stages. The low incidence of structural abnormalities in prosthetic grafts is negatively balanced by the high level of fabric failure, particularly in the porous and ultralightweight grafts popular in the early 1980s. Bovine heterografts and modified human umbilical vein grafts have been observed to undergo aneurysmal changes, and the considerable mural thrombus which can develop in these dilated biologic grafts often leads to graft thrombosis or distal embolism.

Obstructive venous disease

Venous obstruction, both acute and chronic, can increase arterial outflow resistance and is an underappreciated factor that adversely affects graft patency. Following the onset of acute proximal deep vein thrombosis, patients with symptomatic obliterative arterial disease may develop rapidly progressive ischemic symptoms. Unrecognized acute deep vein thrombosis can lead to early graft failure due to high outflow resistance.

Other risk factors

Although more patients with graft thrombosis are suspected of harboring predisposed clotting abnormalities than are documented, hyper-coagulability (acquired or congenital thrombophilia) must be considered among the important causes of graft occlusion. Any patient with a thrombosed graft, not identified to have an inflow, outflow, or stenotic graft problem, should be suspected as having a hypercoagulatable state. Hypotension, hypothermia, and acidosis are common features after major transabdominal aortic reconstruction, and all can contribute to periods of abnormal coagulability during the perioperative period. Furthermore, the type, length, and diameter of bypass grafts combined with patient conditions can also be shown to correlate with patency.

Surveillance and prevention

Bypass graft surveillance and intervention for the failing graft will avert a significant number of graft occlusions. As discussed earlier, careful attention to signs and symptoms of impending graft failure, such as recurrent or progressive intermittent claudication, decreasing ankle pressure or pulse waveform, or diminished flow velocity with routine arterial duplex surveillance, can identify lesions. Long-term patency can be significantly improved by correcting these lesions before graft thrombosis.

Early mechanical graft failure can be prevented by identifying anastomotic abnormalities with routine intraoperative angiography, and any discovered abnormalities can then be immediately revised. Grafts must be carefully placed so that they will not twist through the tunnel. Synthetic grafts are marked with a line that the surgeon observes during graft placement so as to be aware of potential twisting. Ethylene blue can be used to mark saphenous vein grafts that are being used in the reversed fashion. Low flow can cause early postoperative thrombosis, so the avoidance of hypotension by hemodynamic monitoring has become an important component of basic patient care. Pharmacologic intervention is the most common adjunctive method of preserving graft patency. Various clinical trials have concluded that drugs such as heparin, warfarin, aspirin, dipyridamole, and Dextran 40 can be beneficial contributors to improved graft patency. However, use of any of these methods must be balanced by an awareness of the efficacy of each and their potential for complications.

Management

The appropriate management of graft occlusion depends on numerous considerations, including the likely cause of the occlusion, the degree of ischemia, the patient's ability to tolerate reoperation, graft type, the original indications for operation, current indications for revascularization, the condition of proximal and distal arteries, and the likelihood of success (and complications) of intervention.

If graft thrombosis appears within 30 days after implantation, the original indications for surgery still apply. When a stage of reversible ischemia is encountered in combination with a patient's general unchanged condition, immediate reoperation with thrombectomy, intraoperative arteriography, and correction of any technical errors must be considered as the most viable means of management.

With graft occlusion that occurs more than 30 days after implantation, the surgeon has several options. The first consideration is limb viability. Irreversible ischemia leaves primary amputation as the only option. The threatened but reversible ischemic extremity must be promptly evaluated and treatment planned accordingly. In patients with severe motor and sensory loss dysfunction, immediate thrombectomy with operative arteriography is indicated, with need for further revision being determined by the findings on the operative arteriogram. Patients with less severe ischemia may be candidates for catheter-directed thrombolytic therapy. If the patient is not a candidate for thrombolysis, either because of contraindications to lytic therapy or because of the inability to pass a guidewire or to position the catheter properly, operative reconstruction should be offered. Late revisions and secondary reconstructions can add significantly to the overall success rate in such patients.

Further reading

Abbott WM, Vignati JJ. Prosthetic grafts: when are they a reasonable alternative? *Seminars in Vascular Surgery* 1995; **8**: 236–45.

Andersen BS *et al.* Third generation oral contraceptives and heritable thrombophilia as risk factors of non-fatal venous thromboembolism. *Thrombosis and Haemostasis* 1998; **79**: 28–31.

Angelini GD. Saphenous vein graft failure: etiologic considerations and strategies for prevention. *Current Opinion in Cardiology* 1992; **7**: 939–44.

Bailey SR. Local drug delivery: current applications. *Progress in Cardiovascular Diseases* 1997; **40**: 183–204.

Bauters C, Meurice T, Hamon M, McFadden E, Lablanche JM, Bertrand ME. Mechanisms and prevention of restenosis: from experimental models to clinical practice. *Cardiovascular Research* 1996; **31**: 835–46.

Bennett MR, Schwartz SM. Antisense therapy for angioplasty restenosis. Some critical considerations. *Circulation* 1995; **92**: 1981–93.

Chen C *et al.* Recombinant mitotoxin basic fibroblast graft factor-saporin reduces venous anastomotic intimal hyperplasia in the arteriovenous graft. *Circulation* 1996; **94**: 1989–95.

Chen C *et al.* Boundary layer infusion of nitric oxide reduces early smooth muscle cell proliferation in the endarterectomized canine artery. *Journal of Surgical Research* 1997; **67**: 26–32.

Colburn MD, Moore WS. Reoperative approach for failed aortofemoral, axillofemoral, and femorofemoral bypasses. *Seminars in Vascular Surgery* 1994; **7**: 139–51.

Esmon NL, Smirnov MD, Esmon CT. Thrombogenic mechanisms of antiphospholipid antibodies. *Thrombosis and Haemostasis* 1997; **78**: 79–82.

Feldman LJ, Tahlil O, Steg PG. Adenovirus-mediated arterial gene therapy for restenosis: problems and perspectives. *Seminars in Interventional Cardiology* 1996; **1**: 203–8.

Girling J, de Swwiet W. Acquired thrombophilia. *Bailière's Clinical Obstetrics and Gynaecology* 1997; **11**: 447–62.

Hartmann JR. Urokinase recanalization of chronically occluded aortocoronary vein grafts. *Coronary Artery Disease* 1996; **7**: 641–8.

Hofstra L, Bergmans DC, Leunissen KM, Hoeks AP, Kitslaar PJ, Tordoir JH. Prosthetic arteriovenous fistulas and venous anastomotic stenosis: influence of a high flow velocity on the development of intimal hyperplasia. *Blood Purification* 1996; **14**: 345–9.

Kapur RK, Mills LA, Spitzer SG, Hultin MB. A prothrombin gene mutation is significantly associated with venous thrombosis. *Arteriosclerosis, Thrombosis and Vascular Biology* 1997; **17**: 2875–9.

Libby P, Tanaka H. The molecular bases of restenosis. *Progress in Cardiovascular Diseases* 1997; **40**: 97–106.

Makris M, Rosendaal FR, Preston FE. Familial thrombophilia: genetic risk factors and management. *Journal of Internal Medicine* 1997; **740** (suppl.): 9–15.

Mills JL. Mechanisms of vein graft failure: the location, distribution, and characteristics of lesions that predispose to graft failure. *Seminars in Vascular Surgury* 1993; **6**: 78–91.

Nikkari ST, Clowes AW. Restenosis after vascular reconstruction. *Annals of Medicine* 1994; **26**: 95–100.

Oluwole BO, Du W, Mills I, Sumpio BE. Gene regulation by mechanical forces. *Endothelium* 1997; **5**: 85–93.

Pickering JG, Takeshita S, Feldman L, Losordo DW, Isner JM. Vascular applications of human gene therapy. *Seminars in Interventional Cardiology* 1996; **1**: 84–8.

Rai R, Regan L. Obstetric complications of antiphospholipid antibodies. *Current Opinion in Obstetrics and Gynecology* 1997; **9**: 387–90.

Rubin BG, Santoro SA, Sicard GA. Platelet interactions with the vessel wall and prosthetic grafts. *Annals of Vascular Surgery* 1993; **7**: 200–7.

Schwartz RS. The vessel wall reaction in restenosis. *Seminars in Interventional Cardiology* 1997; **2**: 83–8.

Sharafuddin MJ, Hicks ME. Current status of percutaneous mechanical thrombectomy. Part I. General principles. *Journal of Vascular and Interventional Radiology* 1997; **8**: 911–21.

Sharafuddin MJ, Hicks ME. Current status of percutaneous mechanical thrombectomy. Part II. Devices and mechanisms of action. *Journal of Vascular and Interventional Radiology* 1998; **9**: 15–31.

Sluiter W, de Vree WJ, Pietersma A, Koster JF. Prevention of late lumen loss after coronary angioplasty by photodynamic therapy: role of activated neutrophils. *Molecular and Cellular Biochemistry* 1996; **157**: 233–8.

Sottiurai VS. Biogenesis and etiology of distal anastomotic intimal hyperplasia. *International Angiology* 1990; **9**: 59–69.

van den Belt AG, Sanson BJ, Simioni P, Prandoni P, Buller HR. Recurrence of venous thromboembolism in patients with familial thrombophilia. *Archives of Internal Medicine* 1997; **157**: 2227–32.

Walker ID. Congenital thrombophilia. *Bailière's Clinical Obstetrics and Gynaecology* 1997; **11**: 431–45.

Walsh K, Perlman H. Molecular strategies to inhibit restenosis: modulation of the vascular myocyte phenotype. *Seminars in Interventional Cardiology* 1996; **1**: 173–9.

Welch HJ, O'Donnell TF Jr. Role of thrombolytic, antiplatelet, antithrombotic, and other drugs in the management of failed arterial grafts. *Seminars in Vascular Surgery* 1994; **7**: 201–9.

17.16 Vascular access and other techniques for dialysis and chemotherapy

Derek W. R. Gray
The author acknowledges the contribution of Helen Hamilton to the section on Total parenteral nutrition.

[The development of vascular access](#)
[Anatomic and physiological aspects](#)
[Planned use of access sites](#)
[Access for haemodialysis](#)
- [Anaesthesia](#)
- [Preoperative preparation](#)
- [Postoperative management](#)
- [Techniques for acute access](#)
- [Techniques for long-term access](#)
- [Long-term haemodialysis access in children](#)

[Vascular access for total parenteral nutrition](#)
- [Access for short-term total parenteral nutrition](#)
- [Access for long-term total parenteral nutrition](#)

[Vascular access for chemotherapy](#)
[Access for peritoneal dialysis](#)
[Further reading](#)

The development of vascular access

Necessity is the mother of invention, and in the field of surgery vascular access is surely the best example of this truism, for a whole specialty has grown to meet the requirements of new fields such as dialysis and chemotherapy. The earliest vascular access was achieved by the introduction of intravenous glass cannulas in the early 1900s; these were replaced in the 1950s by plastic cannulas which allowed prolonged intravenous infusions. Attempts to extend the duration of infusion and to use more concentrated (and therefore irritant) solutions for intravenous feeding in the 1950s led to the development of central venous cannulation, made possible by the development of longer catheters and less thrombogenic plastics. This technology advanced in the 1970s with the introduction of Teflon- and Silastic-coated catheters, new tunnelling techniques, and the use of porous cuffs to anchor the catheters beneath the skin and perhaps provide a bacteria-proof seal. The central veins can now be used for long-term, high flow access sufficient for intravenous feeding, chemotherapy, and even haemodialysis, although their frequent use for haemodialysis is relatively recent.

The advent of haemodialysis, pioneered by Kolff in 1944, introduced a new requirement for repeated high volume blood flow both into and out of the circulation. Initially this was obtained by adapting the technique of intravenous cannulation to allow repeated catheterization of the femoral artery and vein, made safer by the introduction of the Seldinger guide wire technique. However, this approach rarely allowed dialysis for more than a few weeks. Permanent cannulation of arteries and veins always resulted in thrombosis, until the introduction of the Quinton–Scribner shunt using Silastic tubing in association with Teflon-coated vessel tips. The Quinton–Scribner shunt was the mainstay of haemodialysis for many years, but required relatively frequent revision and was prone to thrombosis and sepsis. A major advance was the realization that the formation of an artificial arteriovenous fistula resulted in massive enlargement of veins sufficient to allow repeated cannulation, and the Brescia–Cimino forearm fistula, first described in 1966, remains the mainstay of haemodialysis access today.

Anatomic and physiological aspects

Vascular access is only ever required to allow some form of medical treatment, and the surgery should be performed in such a way as to make the medical treatment as simple as possible. The first requirement of good access is prolonged patency or absence of thrombosis. The factors important in preventing thrombosis are the use of high flow vessels, non-thrombogenic materials, and reduced blood clotting factors. The second requirement is minimal morbidity including a low rate of infection, and lack of disability caused by subsequent loss of the vessel.

For many patients these requirements are best met by using cannulas placed in large veins. Catheters inserted into the subclavian and internal jugular veins, with the catheter tip in the superior vena cava or right atrium, have high flow rates, and loss of the vessel due to subsequent thrombosis usually causes relatively little disability. Furthermore, the good blood supply of the skin of the neck helps minimize infection. The leg veins and inferior vena cava have a slower flow and so thrombose more easily. Thrombosis often causes leg swelling or embolization and cannulas are at greater risk of infection.

In other patients a form of arteriovenous fistula is the best solution. Here the connection of the radial artery to the cephalic vein in the forearm is the preferred technique, giving high blood flow, low infection rate, and low thrombosis rate, with little morbidity if the vessels are lost. One arm is disabled during cannulation of the fistula, but this drawback is slight compared with the higher thrombosis and infection rate seen when leg vessels are used. Similar arguments apply to the use of cannulas (shunts) to connect arteries to veins.

Planned use of access sites

It is now possible to keep patients with endstage renal failure alive and well for many years, reliant on vascular access for regular therapy. The strengths and limitations of various forms of access in different patient groups have been learnt over the years, allowing the adoption of a flow diagram, which varies somewhat depending on the procedure planned and from unit to unit. A typical flow diagram for dialysis access is shown in Fig. 1. Forward planning is the key to good access surgery, allowing procedures to be performed on routine lists with sufficient time for fistulas to mature before use is required. The object of access surgery should be to avoid using all of the ‘easy' access sites during the lifetime of the patient. Provision of a lifetime of access has been a particular problem for the young patient on dialysis, but the recent success of renal transplant programs has meant that many patients can be relieved of the need to dialyse for large portions of their lives, and judicious planning of access is possible so as to maximize this benefit. Thus a patient who is likely to gain a rapid transplant should be given an early transplant rather than a fistula which is likely to thrombose if the transplant is successful. One of the most difficult messages to get across is the importance of preserving the cephalic veins, particularly so because often patients arrive via other services.

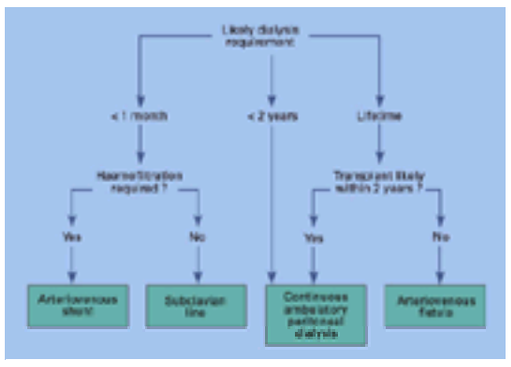


Fig. 1. Flow diagram to illustrate the options available for dialysis access for renal failure.

Access for haemodialysis

Anaesthesia

Patients with renal failure, both acute and chronic, are often a daunting prospect for a general anaesthetic when they first present. Fortunately, local anaesthesia is adequate for most access procedures, although general anaesthesia is recommended for open internal jugular vein exposure in nervous patients. Administration of 1

per cent xylocaine with epinephrine is recommended to minimize oozing.

Preoperative preparation

It is important that patients are neither fluid overloaded nor dehydrated. There is a tendency to dehydrate recently dialysed patients, and it is best to have a drip running, dextran being a good choice because of its added anticoagulant properties. If a limb is to be used it should be shaved preoperatively, painted with Betadine solution, and wrapped in cotton wool to keep it warm. A light premedication is often helpful to calm the nervous patient.

Postoperative management

The danger time for all forms of vascular access is during anaesthesia for another procedure, when dehydration, hypotension, and inadvertent local pressure may result in thrombosis. Great care should be taken to avoid these risk factors. Patients with long-term indwelling cannulas must be scrupulous about covering the exit site and cannula ends with a sterile dressing, careful sterilization of the cannula prior to connection, and regular flushing of the cannula if not in use.

Techniques for acute access

Subclavian and jugular vein catheterization

The subclavian veins or internal jugular veins can be used to provide access for central venous monitoring, intravenous feeding, or chemotherapy, when a single lumen catheter is usually adequate, or for haemodialysis, when a double lumen catheter is required (Fig. 2). Triple lumen catheters are now available that give sufficient flow to allow both uses to be combined. The catheter can usually be placed under local anaesthesia, with or without a short subcutaneous tunnel. The subclavian veins have the advantage of convenience, since neck lines tend to be obtrusive, but jugular line insertion is safer and also avoids the risk of later subclavian vein stenosis, which may create problems for forearm fistula formation. A number of commercial catheters are available. Most now use a smaller bore needle and syringe to locate the vein initially, followed by introduction of a flexible guide wire into the vein, withdrawal of the needle, and then advancement of the final catheter over the guide wire (Seldinger technique). There are several variations of techniques for the catheter insertion: those most commonly used are illustrated in Fig. 3 and described in Table 1. Use of a Doppler can be very helpful for identification of the internal jugular vein for the less experienced.

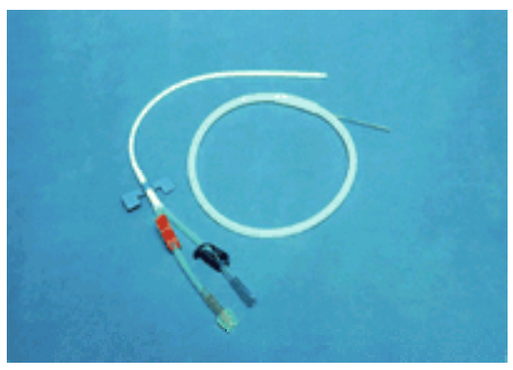


Fig. 2. Double lumen catheter with insertion guide wire. Note the clamps for temporary control of each lumen and caps for sealing between dialyses.

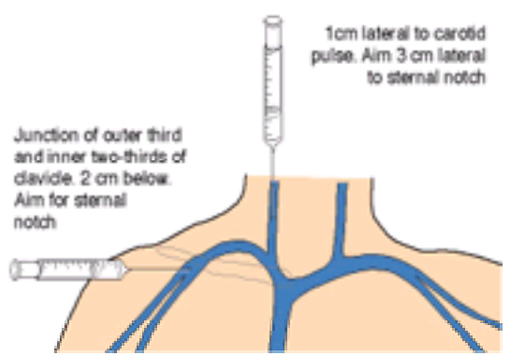


Fig. 3. Landmarks for subclavian and internal jugular line insertion.

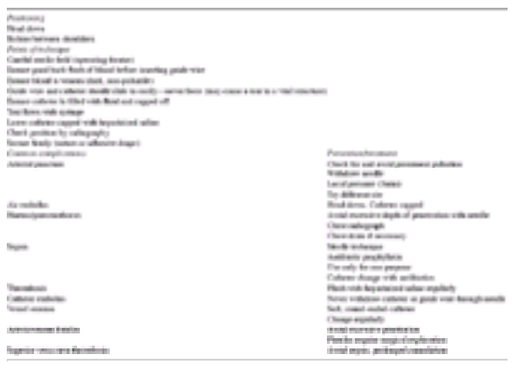


Table 1 Internal jugular and subclavian vein catheterization

Removal of a catheter can usually be achieved by simple withdrawal, with local pressure to the puncture wound for 5 min.

Femoral vein catheterization

The femoral vein is now normally used only to provide access for single haemodialysis in the emergency situation and for venovenous haemofiltration. In this circumstance the femoral approach has the advantage over subclavian vein catheterization because the patient does not need to be placed head down (which can be difficult in patients with fluid overload) and it is more easily performed at the bedside, with less risk of complications, provided prolonged repeated dialysis is not attempted. The actual technique of cannulation is very similar to that employed for the subclavian vein, namely a Seldinger technique using a guide wire and double lumen catheter (Table 2). If arterial pressure is required for haemofiltration then the femoral artery can be cannulated using a similar technique.

Positioning	
Patient lying as flat as possible	
Legs at 30° with slight external rotation	
Points of technique	
Locate femoral pulse just below inguinal ligament	
Puncture vertically 1cm medial using needle and syringe	
Insertion otherwise as for subclavian vein	
Radiograph to check position not necessary	
Common complications	*Prevention/treatment*
Catheter displacement	Secure with adhesive drape
Arterial puncture	Pressure to artery for 2min
Retroperitoneal bleeding	Avoid puncture above inguinal ligament
Groin sepsis/thrombosis	Limit catheterization to under 7 days

Table 2 Technique of femoral catheterization

The catheter can usually be removed by simple withdrawal followed by local pressure to the puncture site for 5 min. Prolonged bleeding may sometimes necessitate longer pressure, and it should be remembered that patients with renal failure have defective platelet function, which it is largely possible to reverse using desmopressin acetate (DDAVP). Occasionally a large haematoma or false aneurysm formation may require exploration, although local pressure using a duplex ultrasound probe to identify and seal the leak now makes surgical intervention rarely necessary.

Quinton–Scribner shunt

The Quinton–Scribner shunt was once the anchor of dialysis access, and can be used for both acute and long-term access, with the vessels of one limb lasting many years in some cases. Unfortunately, minor revisions are frequently required and in addition the device is inconvenient to dress and keep covered. However, the technique still has advantages for the patient with acute renal failure because it allows reliable access with less risk than subclavian puncture, particularly in patients with a bleeding tendency (as is often the case). In addition the arterial pressure can be used to ‘drive' diafiltration systems. Either the radial artery/cephalic vein (Fig. 4) or posterior tibial artery/saphenous vein can be used, but the forearm cannula is usually more practical once the patient is more mobile. This can also be converted to a fistula if needed later. Separate arterial and venous cannulas are placed, which are joined together by a Teflon connector when not dialysing (Fig. 5, Fig. 6(a),Fig. 6(b) and Fig. 6(c)). Local anaesthesia is usually adequate, and a so-called ‘straight' Silastic cannula is preferred (Fig. 5) because it allows simple thrombectomy using a short Fogarty-type embolectomy catheter. The largest possible Teflon vessel tip should be used to cannulate the vessel (Table 3; Fig. 6(a),Fig. 6(b)). Another variety of cannula that has gained some popularity is the Buselmeier shunt, which has the advantage of keeping the shunt loop intact during dialysis, taking the flow to the machine off a side access port. In the event of clotting occurring in the dialyser the shunt will then keep working. The basic principles of insertion are similar.

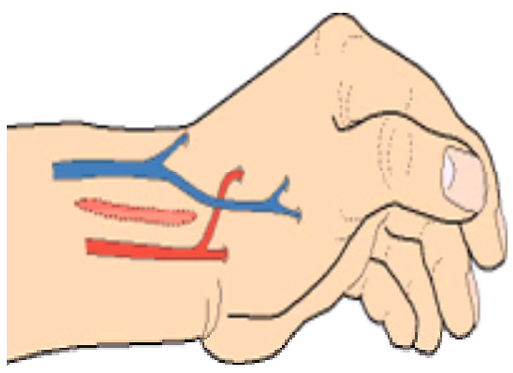


Fig. 4. Incision for exposure of cephalic vein and radial artery for either shunt insertion or fistula formation.

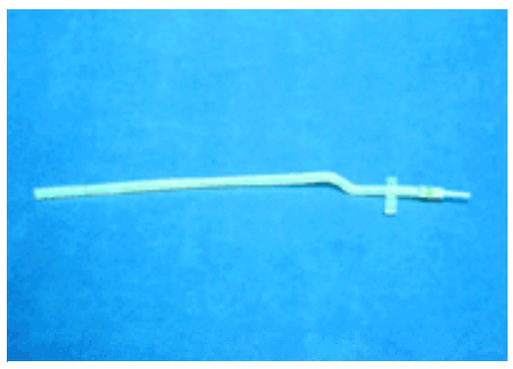


Fig. 5. A so-called ‘straight' shunt cannula, which actually has a kink to allow easy passage through the skin. The side wings are intended to prevent rotation, but may erode through the skin if not placed deeply, and are removed by many surgeons prior to insertion. A Teflon vessel tip is also shown.

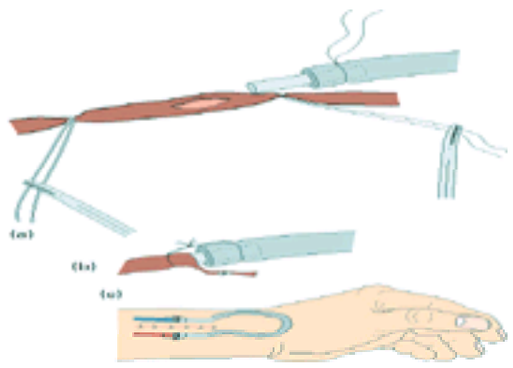


Fig. 6. (a) Insertion of cannula with vessel tip into vessel. (b) Cannula secured in vessel with silk ligatures. (c) Completed shunt with cannulas connected.

Points of technique	
1. Ensure other limb entry (and alternate path) or clear entry proved by palpation or Doppler ultrasound. Allen test	
2. Look for signs of subclavian vein stenosis/thrombosis (especially if previous subclavian catheter)	
3. A single incision between the two vessels is best: avoid parallel incisions	
4. Dissect subcutaneous cuff at least 5 cm from the vein	
5. Ensure vein is patent and accepts a good flow of heparinized saline	
6. Dissect artery, ligating all side branches	
7. Cannulate and tunnel catheters ensuring no kinks or twists	
8. Division of the distal vessel (to clamp) is later removed	
Common complications	Prevention/treatment
Thrombosis	Choose good vessels
	Ensure good hydration and blood pressure
Impaired thrombosis	Exclude mechanical cause
	Shunt angiogram
	Anticoagulation
Sepsis (usually caused by *Staphylococcus aureus*)	Prevent with prophylactic antibiotics and meticulous asepsis
	Antibiotic treatment sometimes effective

Table 3 Quinton–Scribner shunt insertion

Removal of cannulas may be required because of repeated thrombosis, sepsis that cannot be cleared, or when the shunt is no longer required. Provided clotting is normal and in the absence of sepsis, the shunt can simply be clamped off for a period of at least 4 and preferably 7 days. Local anaesthetic is then infiltrated around the catheter tips and the cannulas removed by avulsion. In the presence of clotting disorders or sepsis it is wiser to remove the cannulas under direct vision and ligate the vessels with an absorbable ligature such as polyglycolic acid.

Techniques for long-term access

Arteriovenous fistulas

The connection of an artery to a vein results in steady widening of the vein and its tributaries, with increased rates of flow. Veins which were originally difficult to cannulate and unable to deliver the rate of flow sufficient for dialysis become ideally suited for repeated cannulation (Fig. 7). Once developed fully the fistula can remain patent indefinitely and provide permanent access. The ideal site, originally described by Brescia *et al.*, is the radiocephalic fistula, but other sites are possible. Most fistulas can be formed under local anaesthesia, but maintenance of adequate hydration and a warm limb are particularly important for successful fistula formation. A recent innovation has been the use of a glyceryl trinitrate patch on the distal limb to encourage vasodilatation. For patients with a particular fear of needles, especially children, the use of **EMLA** (eutectic mixture of local anaesthetic) cream can allow painless cannulation.



Fig. 7. Dilated veins after radiocephalic fistula formation.

For the creation of a radiocephalic fistula the incision and dissection should follow guidelines 1 to 5 (Table 3) outlined for shunt formation, although a slightly longer incision is needed (Fig. 4) and enough vessel should be dissected free to bring the two vessels side by side over approximately 2 cm without tension (Fig. 8(a),Fig. 8(b); Table 4). The cephalic vein is usually the best to use, but occasionally other nearby tributaries will have suffered less trauma from previous infusions. Soft Silastic slings are used to control and approximate the vessels. The vessels are opened for a distance of 7 to 10 mm side by side, and the vein patency and size tested by injecting heparinized saline. The vessels are then anastomosed using continuous 6/0 or 7/0 Prolene suture, taking particular care not to narrow the lumen at the proximal end. A more precise anastomosis can be achieved using magnification loupes, but adequate anastomosis is possible without. Some surgeons prefer to tie off the distal vein or perform an end-to-side anastomosis in order to prevent the possibility of distal venous hypertension, but this is an uncommon complication that is relatively simple to deal with and leaving the distal vein initially may allow better run-off in the critical postoperative period. On releasing the slings a palpable thrill should be evident. The expected patency rate is variously reported at between 60 and 90 per cent at 1 year and 60 to 75 per cent at 5 years.

Common complications	Management/treatment
Early thrombosis	Avoid: unsuitable veins (assessed surgeon early), compliant, recruitment of veins)
	Check for technical error
	Correct deficiency
	Early thrombectomy
Thrombosis	Carefully figure all vessel tributaries
Failure to develop fully	Vein too small / <3 mm; or previously damaged by intravenous infusions
Early failure to mature/develop	Infection/vein wall damage/thrombosis
	Patients with previous subclavian catheter/vein stenosis have a characteristic symptom
Distal venous hypertension (distal to shunt)	Tie off distal end of saphenous vein
Aneurysm formation (see Fig. 9)	Avoid: repeated needling in same place. May require elective surgery
Late thrombosis	Poor needling technique causing perivascular haematoma
	Aneurysm or pseudoaneurysm required
High-output symptoms	Venous stenosis: angiogram
	Revision required
Heart failure	Early: with high-output failure (20–30 per cent of cardiac output) unusual with radial fistula
	Treatment: narrow fistula
Distal ischaemia (rare)	Digital arterial steal: revision/ligate fistula

Table 4 Brescia–Cimino fistula formation

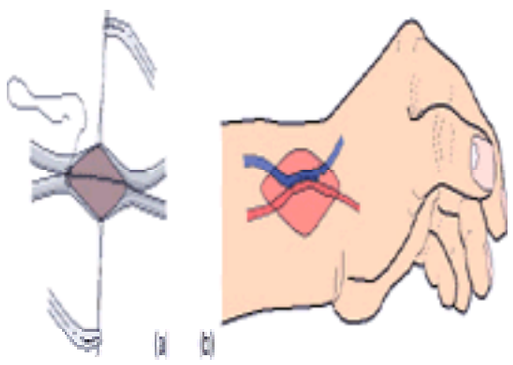


Fig. 8. (a) Anastomosis of artery to vein. (b) Completed radiocephalic fistula.

The general rule in fistula formation is that the non-dominant arm and the most distal site possible are used. It is possible to locate the radial artery distally in the

anatomic snuffbox and at this point the artery lies close to the cephalic vein. However, the access for the anastomosis is more limited, making the procedure technically demanding and, should the cephalic vein prove to be unsuitable, alternative veins are not available. If the radial artery is occluded or absent the ulnar artery can be used, together with a cephalic tributary or the basilic vein (less ideal as it tends to lie deep in an awkward position), but an occlusion test and Doppler ultrasound studies should be used to confirm another arterial supply to the hand in case of thrombosis. If a forearm fistula fails it is possible to make a fistula between the cephalic (or median cephalic) vein and the brachial artery at the elbow, and still have sufficient dilated veins develop in the upper arm to allow satisfactory dialysis. In addition to the complications mentioned for radiocephalic fistula, the major complications of using this site are distal ischaemia due to stealing of blood and the development of cardiac failure due to massive fistula development. Both these complications can be minimized by limiting the diameter of the anastomosis to 6 mm.

In general a functioning fistula should not be closed, even if a kidney transplant makes the fistula unnecessary, since it can never be certain that later dialysis will not be required. In fact many fistulas thrombose after successful transplantation. Indications for fistula closure include cosmetic appearance, extremely large flows leading to heart failure (see above) and, occasionally, sepsis and distal embolism. It is usually possible to dissect the fistula free, control the feeding vessels, and take down the anastomosis. Often the artery can be reconstituted using fine Prolene suture, but the veins are best ligated.

Long-term central venous catheterization

Although a forearm fistula is the procedure of choice for long-term dialysis access, it does have disadvantages (Table 4; Fig. 9). The fistula cannot be used for dialysis for at least 4 weeks and often longer, since time is required for dilatation of the veins to occur. However, in some cases full development of the veins does not result, and thus a fistula may be a less attractive option in patients who present in renal failure and who require early dialysis. The waiting period can be covered by temporary dialysis via a subclavian line (see above), but the insecurity of this access makes most clinicians prefer inpatient treatment, and subclavian thrombosis and sepsis become more frequent with prolonged subclavian cannulation using temporary lines. In elderly patients with a limited life expectancy this waiting period may be a significant proportion of their remaining life. Some patients also have a horror of needles or have a phobia of needling themselves. Lastly, connection of an indwelling catheter is certainly less technically demanding than the insertion of two dialysis catheters into a fistula, and this may be an important factor in considering home dialysis in some patients.



Fig. 9. Aneurysm formation in a radiocephalic fistula.

The development of more efficient membranes has meant that venovenous dialysis can now be a long-term option, even though recirculation of 7 to 15 per cent of the blood may occur, and hence dialysis via a long-term central venous catheter represents another chronic access option (Table 5).

Catheter thrombosis	Frequent flushing Fill catheter with urea sodium heparin 5000 U/ml Slow low-dose infusion of urokinase 10 000 units/h may clear partial blockages Use flexible guide wire to clear abdominal occlusions
Persistent occlusion	Change catheter over two guide wires (one each channel)
Intermittent occlusion of outflow	Reposition catheter tip in right atrium, chest side away from atrial wall
Infection	Prophylactic antibiotics on insertion Use line only for dialysis Obturational sterile connection technique Change catheter over guide wire using new tunnel and antibiotics Remove catheter
Superior vena cava thrombosis	Avoid positioning tip too high in superior vena cava
Atrial thrombus formation	Avoid positioning tip too low in atrium on tricuspid valve Both may respond to lytic therapy Surgery occasionally required

Table 5 Complications of long-term central venous dialysis catheterization

Although it is possible to use the subclavian vein, stenosis or occlusion is a common eventual outcome. The jugular veins are a better option, since loss of the external jugular veins or one internal jugular vein leads to no disability, and both internal jugular veins can be occluded with surprisingly little disability, provided there is a gap of some months between procedures.

The catheter must be robust and made of Silastic or other non-thrombogenic material with a cuff to allow tissue ingrowth; most have a double lumen with well designed caps and temporary occlusion clips (Fig. 10), although a single lumen (Francis) catheter based on an adapted peritoneal dialysis catheter, has also been used with success. Although the catheter can be placed in the external jugular vein of some patients, this vein is often too small and even if cannulation is successful malposition is common. The internal jugular vein is more certain to allow satisfactory cannulation.



Fig. 10. Catheter for long-term venovenous dialysis. Note the dual lumen with staggered distal opening to minimize recirculation, temporary occlusion clamps, Dacron cuff, and caps for closure between dialyses.

At one time the only option for placement of this form of catheter in the internal jugular vein was formal surgical exposure, but the introduction of semiflexible dilators and ‘peel-away' sheaths (Fig. 11) means that even cuffed cannulas can be successfully placed using an adaptation of the percutanous approach to the internal jugular vein, already described under the techniques for temporary access. Indeed, the ability to place a central venous catheter percutaneously with potential for more long-term use is changing the previous fairly clear demarcation between temporary access procedures, which were often performed by non-surgically trained personnel, and long-term access, which was usually a surgical procedure. Furthermore, the major difficulty with placing cannulas in central veins is no longer getting

the catheter in, but rather ensuring correct placement, and for this reason the involvement of interventional radiologists is a logical development.

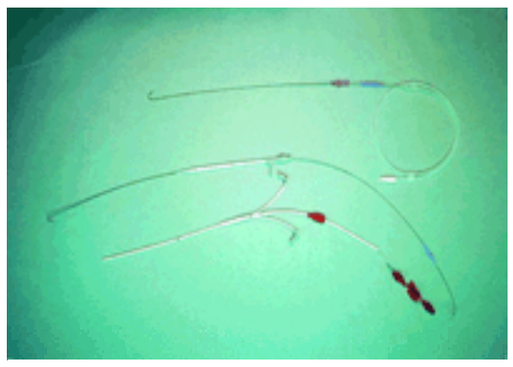


Fig. 11. Peel-away' sheath and dilators for percutaneous tunnelled line insertion

However, a number of patients still require surgical placement of central venous catheters. The internal jugular vein is approached at the lower end of sternomastoid, separating the sternal and clavicular heads (Fig. 12). Care must be taken in mobilizing the vein as small tributaries may be present, and avulsion may cause haemorrhage. The internal jugular vein must be carefully separated from the vagus behind and medial to it. The vein can be controlled by two double-looped slings and a purse-string suture of 4/0 Prolene inserted. The catheter is prefilled with saline, tunnelled to the wound, and the patient tipped head-down; the catheter is then inserted into the vein via an incision in the centre of the purse-string, which is then tied snugly around the catheter. A radiograph to check the position of the catheter should always be taken immediately and, ideally, the tip should be placed at the entrance to the right atrium. After closing the wound the patient should be nursed as upright as possible to minimize hematoma formation. Patency rates of 60 per cent at 2 years are claimed. Early failure of the catheter is usually due to development of a fibrin sheath around the catheter tip, and recently a technique for stripping the fibrin sheath from the catheter via a percutanous wire snare under radiological control has been described, with success reported in 80 per cent of cases, but complications are reported. Problems with catheter clotting can often be treated by instillation of thrombolytic agents: recombinant tissue plasminogen activator is fast and effective.



Fig. 12. Incision used for exposure of internal jugular vein. The anterior border of sternomastoid is marked by a dotted line.

Catheter removal is usually required for intractable thrombosis or sepsis (Table 5) and involves reopening of the original wound to free the Dacron cuff by sharp dissection. It is usually best to locate and divide the purse-string suture. The catheter is removed with the patient in the head-down position; bleeding can be controlled by a finger, whilst the track is closed with absorbable sutures. For percutaneously placed cuffed and tunnelled lines removal involves making a small incision above the Dacron cuff, lying in the subcutaneous tunnel. Careful, sharp dissection will reveal the cuff, whereupon the catheter immediately above the cuff can be located with an aneurysm hook. The catheter can then be safely hooked, removed, and the cuff gently dissected agay from the fibrous tissue tethering the catheter. Finally the incision may be sutured and dressed.

Salvage procedures for patients with access problems

Expertise in the more complex access salvage techniques is not a skill that should be acquired easily, since it represents failure of the simpler procedures. Nevertheless, despite the greatest care, some patients eventually exhaust the simple access sites. This problem is most common in small women, where the vessels are naturally tiny, patients with vascular disease or Raynaud's phenomenon, patients with diabetes, and patients with abnormal susceptibility to sepsis for some reason. A large number of ingenious techniques have been described for more complex access: only those that are commonly used will be described.

Venous transposition

A fistula may be functioning but unusable due to the deep-seated position or inaccessibility of the veins, for example the basilic vein. Mobilization of the vein to a more superficial position may prevent loss of an access site.

Interposition grafts

A vascular graft (either vein or prosthesis) inserted subcutaneously between a suitable artery and vein will allow repeated vascular access. Autogenous saphenous vein is probably the conduit of choice, since extensive experience in vascular surgery has shown this to have a lower natural thrombogenicity and infection rate with repeated needling. Saphenous vein is usually sufficiently strong to resist dilatation by persistent arterial pressure, a feature that is advantageous for vascular bypass grafts, but not ideal for access where needling of the graft is required. For this reason autogenous saphenous vein should only be used if a large vein is available. Patency rates of 70 per cent at 1 year and 65 per cent at 5 years are quoted. The alternative is a synthetic graft, with purpose-made polytetrafluoroethane (PTFE) grafts with loop reinforcement probably being preferred over other synthetic grafts (Fig. 13). Patency rates of 58 to 95 per cent at 1 year and 40 to 80 per cent at 2 years are claimed by various authors. Alternatives such as human umbilical vein, bovine artery, and human saphenous vein allografts have shown a higher complication rate from sepsis and aneurysm formation. An interesting technical modification is the addition of a connection port to the conduit to allow simple connection for dialysis without the need for cannulation (Haemasite). Two-year patency rates of 50 per cent with low infection rates are claimed for this prosthetic graft and thrombosis is easily cleared by opening the port.

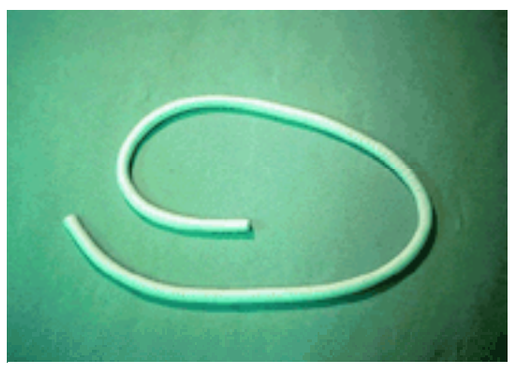


Fig. 13. PTFE interposition graft. Note the reinforced rings designed to prevent kinking or compression of the loop apex.

In most patients the forearm and lower leg vessels have been used previously for either shunt or fistula formation, and the jugular veins may also have been used. The first requirement is location of a suitable artery and vein for anastomosis. For some patients a number of options are open: the radial artery may be patent, but since there are no suitable veins in the forearm a longitudinal graft needs to be run from the radial artery and connected to a cubital vein. In other patients the brachial artery and cubital veins are available and a forearm loop can be performed (Fig. 14). Many of these patients have had previous attempts at formation of fistulas for dialysis but no vein suitable for subsequent dialysis use has been found. In such patients it may still be possible to perform a fistula anastomosis to veins that may not be suitable for subsequent dialysis directly (for example venae comitantes round the brachial artery), and the author recommends performing such a fistula whenever the opportunity arises during an otherwise ‘failed' fistula formation attempt, because it allows development of veins that can subsequently be used for interpostion graft placement. It is possible to cross the elbow joint to anastomose to the cephalic vein above the elbow, provided a reinforced graft is placed with the rings across the joint line. If these options are not possible a loop can be formed in the thigh, most simply by connecting the mobilized long saphenous vein as a loop anastomosed directly on to the common femoral artery. Other grafts are usable if anatomically possible: axillary artery to femoral vein and grafts between femoral vessels of opposite legs have been used. In these cases the basic technique is similar. The points of importance are listed in Table 6.

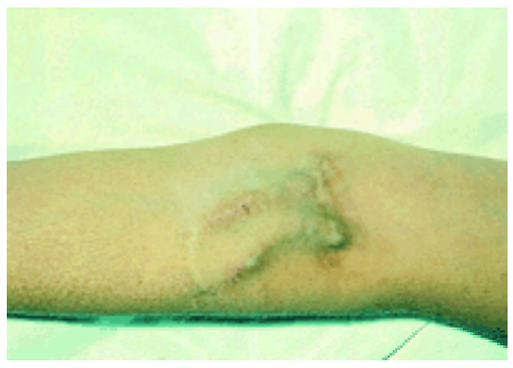


Fig. 14. Forearm saphenous vein loop between brachial artery and cephalic vein.

Careful preoperative preparation, hydration, and anesthesia
Prophylactic antibiotics
Choose widely patent, preferably unstenosed, vessels
Confirm proximal vessel patency preoperatively (radiology if necessary)
Use grafts at least 6 mm in diameter
Meticulous anastomosis technique avoiding vessel narrowing and intimal flaps
Careful tunnelling of grafts in the subcutaneous position (do not place too deeply—difficult to use)
Keep loops short and carefully avoid twists and kinks
Confirm rapid flow before closure
Early re-exploration of graft occlusions
On-table radiology if graft occludes
Late graft stenoses must be investigated by radiology and treated by percutaneous dilatation or early revision

Table 6 Points of importance in formation of interposition grafts for dialysis access

Long-term haemodialysis access in children

Although the general principles outlined for adults also apply to children, the problems are increased by the small size of vessels available and the importance of psychological factors such as needle fear. Fortunately the introduction of EMLA cream allows the painless insertion of needles, making a fistula a far more attractive option than was previously the case. Since one must plan for a lifetime of renal replacement therapy, decisions regarding the use of access sites are particularly important. General anaesthesia is recommended for all access-forming procedures.

Acute access for dialysis is a particular problem in children, since the subclavian veins are relatively difficult to cannulate percutaneously and tend to thrombose once cannulated. In neonates and many older children the best option may be peritoneal dialysis, which is discussed elsewhere. If this option is not available the best solution is usually to insert an internal jugular line under direct vision, using a cannula of as large a bore as can be inserted, with the tip of the cannula in the right atrium. The principles of insertion are similar to those described for adults. In many children the next option chosen would be early transplantation rather than fistula formation, in order to preserve the forearm site for future use.

In children in whom early transplantation is not an option the procedure of choice is again a radiocephalic fistula. Although the surgical principles are similar to the technique used in adults, experience of microsurgical techniques and special instruments are required.

Vascular access for total parenteral nutrition

The requirements of access for parenteral nutrition differ from those of haemodialysis in that only a single line is required, and the flow rates needed are relatively small. However, the nutrients being delivered are highly irritant to vessels and it is thus not possible to use peripheral veins. The haemoglobin level of these patients is often near normal, as is platelet function, and so formation of an arteriovenous connection is contraindicated because of the likelihood of thrombosis, although this is a technique that has been used occasionally for total parenteral nutrition. The best solution is to use central venous cannulation, where the concentrated nutrients can be diluted rapidly enough to prevent vessel inflammation. The catheter used can be small bore, but must be made of non-irritant material such as silicone, pure polyethylene, or a polyethylene/silicone blend and in the case of permanent nutrition must be sufficiently robust for long-term use. Inclusion of a Dacron cuff is felt to be an advantage for prevention of track infection. A number of commercial catheters are now available (Fig. 15) and a major advance has been the development of totally implantable catheters with ports containing self-sealing membranes that allow intermittent cannulation (Fig. 16). The number of cannulations possible is still rather limited for the purposes of long-term total parenteral nutrition, but it is certainly an option that is worth considering in children and for relatively short-term outpatient management. The sepsis rate is generally low, the major contraindication being methicillin-resistant staphylococcal (MRSA) skin infection or colonization.

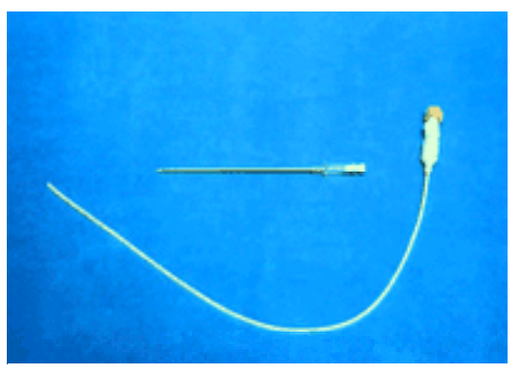


Fig. 15. Silastic catheter suitable for total parenteral nutrition, with needle for percutaneous insertion. The hub is detachable to allow removal of the needle and formation of a subcutaneous tunnel.

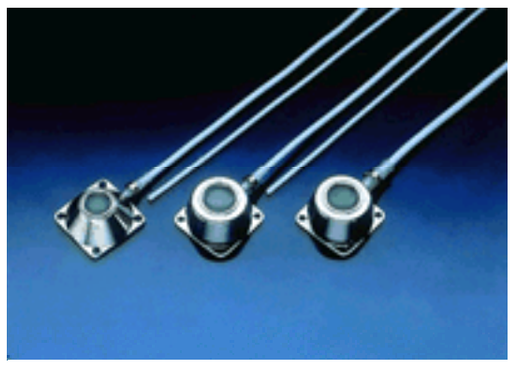


Fig. 16. Three varieties of implantable cannulas and ports for repeated access to venous, arterial, and peritoneal sites.

Access for short-term total parenteral nutrition

Percutaneous insertion of the catheter into the subclavian vein under local anaesthesia is the usual approach because the catheter can then be conveniently kept out of the way by strapping to the chest. The cannulation technique is similar to that described earlier for haemodialysis catheter insertion, but the proximal line should also be tunnelled out from the puncture site for a distance of about 5 cm. The complications include those noted for subclavian dialysis catheter placement, but sepsis is the most frequent problem. This usually arises from poor connection technique, particularly if the catheter is used for any purpose other than total parenteral nutrition; since many of these patients are in an intensive care unit the temptation to allow use of the line for blood sampling and drug delivery must be firmly avoided. Delivery of the daily requirement premixed in one bag also limits the number of bag changes and reduces the chance of infection.

Lines suspected of being infected may be cultured by taking blood from the line: a negative result is reliable, provided the patient has no antibiotics in the bloodstream at the time. Infected lines are best managed by removal and insertion of a fresh line on the other side, and occasionally infection can be cleared by changing the catheter over a guide wire, using a new subcutaneous tunnel, combined with intensive antibiotic treatment, although this is rarely possible with the smaller modern tunnelled lines.

Access for long-term total parenteral nutrition

The catheters used for this purpose need to be considerably more robust and are consequentially larger. Insertion is often best undertaken by cannulation of the cephalic vein under local anaesthesia (Fig. 17), and percutaneous insertion into the jugular or subclavian vein is possible even for cuffed catheters by the use of ‘peel-away' insertion sheaths. The direction of the catheter is often incorrect at the first attempt when inserted by the percutaneous route and radiographic screening should be undertaken, making the interventional radiologist the logical person to perform the procedure, although the requirement for a higher standard of sterility than prevails in routine vascular interventional radiology must be understood. Again the catheter should be tunnelled for at least 6 cm from the cannulation site below the clavicle, minimizing entry of bacteria directly into the vascular system. Although the presence of a tunnel does not seem to influence the sepsis rate, it is more comfortable and convenient for the patient. The inclusion of a Dacron felt cuff is beneficial since the tissue ingrowth holds the catheter in place and may effectively seal the track to bacteria. The catheter tip must be smoothly rounded off and carefully positioned to lie free just at the entrance to the right atrium, since there have been descriptions of catheter perforation and thrombosis of the heart and great vessels. Sepsis is the main complication and attention to aseptic technique during bag changes must be meticulous (Table 7 gives a comprehensive list of complications and their management).

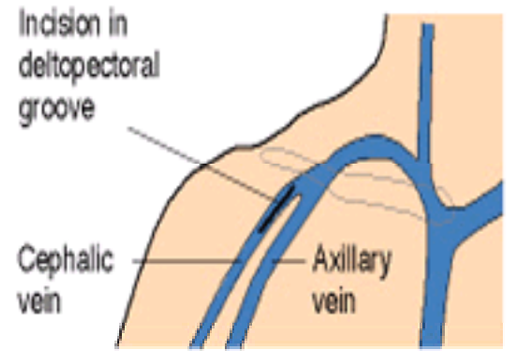


Fig. 17. Incision for cannulation of the cephalic vein.

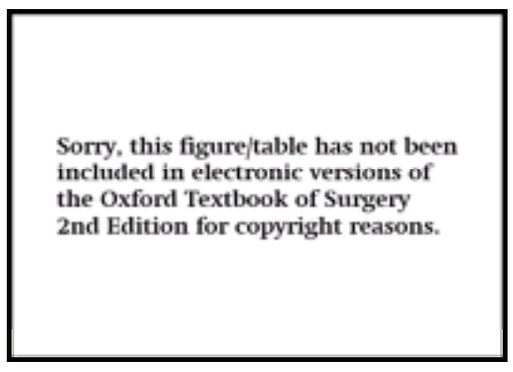


Table 7 Catheter related complications of parenteral nutrition

Home total parenteral nutrition is mainly indicated in patients with short bowel syndrome. Meticulous training in catheter and bag changing technique is vital to prevent sepsis, which is the main complication. Attention to detail with aseptic technique in every aspect of the administration of parenteral nutrition is essential. Suspected sepsis may be confirmed by finding bacterial counts at a ratio of greater than 5:1 in blood samples cultured from the catheter and a peripheral vein, respectively. The most common causative organism is *Staphyloccus aureus*. Sepsis may be cleared by administration of urokinase and appropriate antibiotics via the catheter for 5 days. Occlusion may also occur due to lipid or calcium phosphate encrustation, which may respond to local instillation of ethanol or 0.1 mol hydrochloric acid, respectively (into the catheter only and not systemically).

Vascular access for chemotherapy

The treatment of a number of malignancies has been transformed by the advent of effective chemotherapeutic regimens, but the drugs used must be given intravenously at regular intervals, often for many weeks. Many of the therapeutic agents are highly irritant to peripheral veins and need to be delivered into high flow vessels. It has been suggested that constant infusion of chemotherapeutic agents may increase their efficacy and lower toxicity, but firm evidence for this point is lacking. Many chemotherapeutic drugs have toxic effects on the bone marrow and other organs and numerous blood samples need to be taken on a daily basis to monitor toxicity. Perhaps the most demanding example of intensive chemotherapy is that required for treatment of leukaemia, particularly when bone marrow transplantation is also required. There is, therefore, a need for reliable access to the central veins over a period of weeks to months and, in addition, a catheter that will allow both drug delivery and blood sampling at the same time is most useful. The most convenient catheter is a double lumen catheter of the Hickman variety (Fig. 18), which can be inserted under local anaesthesia into the sublavian, cephalic, or external jugular veins.

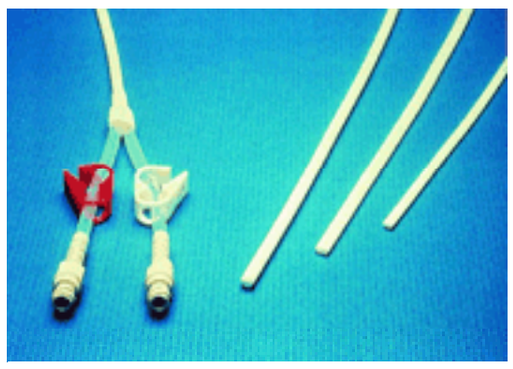


Fig. 18. Hickman-style catheter for medium- to long-term venous access. Three sizes of double lumen catheter are shown. Note the clamps for temporary control of each lumen.

Patients suffering from leukaemia are often thrombocytopenic and neutropenic, and insertion of a line should only be attempted under full coagulation factor and platelet replacement with antibiotic prophylaxis. The line should normally be tunnelled a minimum of 10 cm to minimize infection, although in patients with absent platelets the tunnel must be kept very short to avoid bleeding and haematoma formation.

A variety of alternative routes for delivery of drugs to allow treatment of apparently localized tumours have been described including the recanalized umbilical vein for hepatic perfusion, hepatic artery cannulation, and carotid artery and femoral artery perfusion via fine bore catheters. None has reached the status of accepted treatment and will not be considered further here. Some preliminary studies have examined the use of intraperitoneal delivery of drugs for ovarian cancer, and for this purpose a standard Tenckhoff catheter is ideal (see below).

Similar principles to those used in adults apply to insertion of lines for chemotherapy in children, but it is usually necessary to use the internal jugular vein for line insertion, although successful use of the long saphenous vein to cannulate the inferior vena cava has been described. General anaesthesia is normally required.

Access for peritoneal dialysis

The concept of using the peritoneum as a membrane for dialysis arose in the 1920s. It was initially employed to provide temporary renal replacement therapy, for example in cases where recovery of renal function was expected. The catheters used for this purpose were inserted using a ‘blind' direct puncture technique with a trochar. Acute peritoneal dialysis catheters are still available (Trocath) but are probably used more for diagnostic or therapeutic peritoneal lavage than for actual dialysis. The development that allowed long-term peritoneal dialysis was again the introduction of Silastic non-irritant catheters, with the addition of cuffed tubes and use of a long subcutaneous tunnel. A variety of commercial catheter designs have been marketed, but the original Tenckhoff design is still the most popular and very effective, although the addition of a curl to the end of the catheter may help prevent displacement (Fig. 19). The requirements are for a catheter that allows rapid inflow and outflow of fluid, without leakage from the peritoneum and without allowing the introduction of agents which may cause sepsis. Complications of peritoneal dialysis catheter insertion are listed in Table 8.

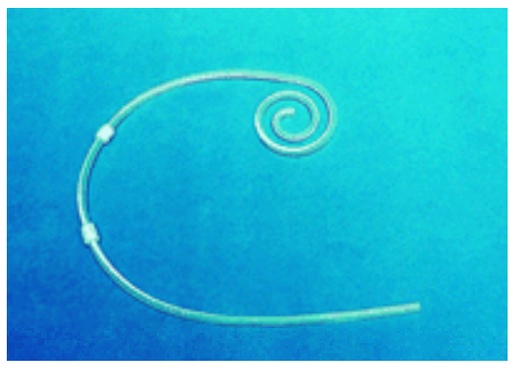


Fig. 19. Curled Silastic catheter for peritoneal dialysis with double cuff.

Leakage of fluid	Small peritoneal incision Careful insertion with easy ligatures Avoid usage till possible (at 2 weeks especially in the elderly)
Infection (peritonitis usually caused by *Staphylococcus aureus*, especially in diabetics)	Good surgical technique Prophylactic antibiotics Meticulous connector sterility Regular antiseptic cleaning of exit site
Inadequate dialysis	
(i) catheter displacement	Carefully place catheter tip in pelvis Suture mesenteric fold/omentum/pexing sutures Replacement under direct vision (i.e. suture of tip)
(ii) obstruction by omentum	Carefully place catheter tip in pelvis Omentectomy
Bleeding/bowel perforation	Avoid forcible insertion Minimal bleeding usually stops Direct vision placement in patients with adhesions (i.e. adhesion lysis)
Groin or scrotal oedema External Dacron cuff erosion	Usually short-term in noninfectious hydrothorax. Stop dialysis and repair hernia Place cuff > 2cm from exit site (>1) Shave off cuff using scalpel

Table 8 Complications of peritoneal dialysis catheter insertion

Continuous ambulatory peritoneal dialysis has the advantage of being relatively cheap, simple to set up, and relatively easy to learn to run. The drawbacks are the relatively high readmission rate for complications, a tendency to obesity due to glucose absorption, and a short useful life, with approximately 50 per cent of catheters failing and requiring removal by 2 to 3 years. It is therefore not usually used for long-term dialysis of young, fit patients, but can be used as an option for patients where transplantation is expected within 2 years. It is also a favoured option for elderly patients, where survival is expected to be limited, and in patients with diabetes and other vasculopathies, where vessels suitable for haemodialysis may not be easily available. In addition, peritoneal dialysis using a smaller version of the Tenckhoff catheter is a good way of providing short- to medium-term dialysis for children, including neonates, and may provide sufficient time for successful transplantation to be performed. For chronic dialysis in children a double cuff catheter is not recommended as growth may cause puckering of the abdominal wall.

The catheter can be inserted by percutaneous puncture using a peritoneoscope, a technique that has firm advocates, but most surgeons prefer insertion by an open operative technique. Insertion under local anaesthesia is possible, but is only a comfortable procedure if no difficulties are encountered. General anaesthesia with relaxation is probably wiser in most patients. Prophylactic antibiotics are advisable. The catheter should be placed so that the subcutaneous tube and exit site is located either above or below the belt line, preferably on the opposite side to the side on which the patient normally sleeps (Fig. 20(a)). The peritoneum is best reached via the lower rectus sheath. A vertical or transverse incision can be used, opening the anterior rectus sheath. The rectus muscle can either be split or swept laterally to allow access to the posterior rectus sheath/transversalis fascia. A small incision in the peritoneum is then made, and a purse-string suture inserted. The catheter is then lubricated with sterile glycerin, loaded on to a blunt-ended introducer and gently inserted along the anterior abdominal wall, over the bladder (and uterus in the female) and then slid down in front of the rectum. Some surgeons prefer to make a bigger peritoneal incision and insert the catheter under direct vision using sponge-holding forceps. A larger peritoneal incision also allows omentectomy to be performed safely, an addition that may reduce catheter failure due to occlusion by omentum. Omentectomy has been described as particularly important for peritoneal dialysis in children, although doubt has recently been cast on this dictum. The purse-string suture is then tied and stitched to the inner Tenckhoff Dacron cuff (Fig. 20(b)).

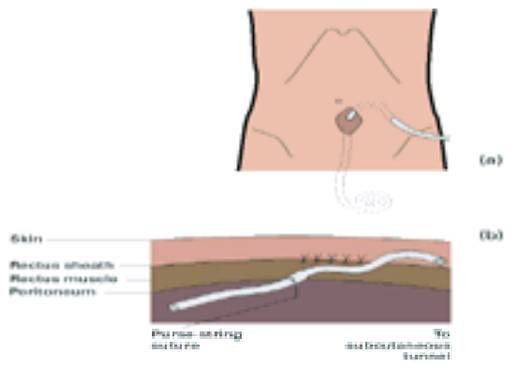


Fig. 20. (a) Incision and placement of peritoneal dialysis catheter. (b) Detail of peritoneal dialysis catheter insertion to show purse-string suture attaching catheter cuff to peritoneum and also the method of closure of the rectus sheath in order to keep the catheter pointing inferiorly.

The peritoneum can then be pulled up to allow a non-perforating, encircling ligature below the purse-string that helps prevent leakage of fluid from the stitch holes. The catheter is then tunnelled to allow a gentle curve away from the peritoneum so that the skin exit site is facing downwards, which discourages accmulation of debris at the exit site. The use of catheters with a second cuff has not been proved to have any advantage over single cuffed catheters for adults (although there may be some advantage in children), but where a second cuff is used it must lie at least 3 cm from the skin exit to prevent later extrusion. The rectus muscle and sheath should be closed snugly around the catheter so as to keep the catheter pointing into the pelvis (Fig. 20(b)), followed by the skin closure. On completion the catheter flow and absence of leakage should be tested by connection to a peritoneal dialysis bag and exchanging 1 litre of saline rapidly. A catheter that is subsequently found to be malpositioned can often be repositioned by an enthusiastic interventional radiologist using a stiff guidewire.

The indications for catheter removal are repeated episodes of peritonitis, an infected catheter track unresponsive to antibiotics, loss of effective dialysis due to peritoneal thickening, and return of renal function or a successful transplant rendering the catheter redundant. It has been shown that infection is associated with bacterial colonization of a mucoid fibrin sheath laid down on the catheter itself, which gives a rationale for the several reports of successful removal and replacement of the catheter in the one procedure. Removal can usually be accomplished under local anaesthesia. It is necessary to reopen the original insertion wound, freeing both the internal and external Dacron cuffs (where present) using sharp dissection. After removing the catheter the rectus sheath defect should be closed using a non-absorbable suture if there is no infection, to avoid later incisional hernia.

Further reading

Historical references

Brescia MJ, Cimino JE, Appel K, Hurwich BJ. Chronic hemodialysis using venipuncture and a surgically created arteriovenous fistula. *New England Journal of Medicine* 1966; **275**: 1089–92.

Buselmeier TJ, Kjellstrand CM, Santiago EA, Simmons RL, Najarian JS. A new subcutaneous arteriovenous shunt. *Surgery* 1973; **73**: 512–20.

Kolff WJ, Berk HJ, Welle M, Van der Ley AJW, Van Dijk EC, Van Noordwijk J. The artificial kidney: a dialyser with a great area. *Acta Medica Scandinavica* 1944; **147**: 121–34.

Quinton W, Dillard D, Scribner BH. Cannulation of blood vessels for prolonged haemodialysis. *Transactions of the American Society of Artificial Organs* 1969; **15**: 104–13.

Seldinger SI. Catheter replacement of the needle in percutaneous arteriography. *Acta Radiologica* 1953; **39**: 369–76.

Tenckhoff H, Schechter H. A bacteriologically safe peritoneal access device. *Transactions of the American Society of Artificial Organs* 1968; **14**: 181–7.

Useful textbooks and guides

Bell PRF, Wood RFM. *Surgical aspects of haemodialysis*, 2nd edn. Churchill Livingstone, Edinburgh, 1983.

Bengmark S, ed. *The peritoneum and peritoneal access for dialysis*. Wright, London, 1989.

Fine RN, Alexander SR, Warady BA, eds. *CAPD/CCPD in children*. Boston: Kluwer Academic, 1998.

Nissenson AR, Fine RN, Gentile DE, eds. *Clinical dialysis*. 3rd edn. London: Prentice-Hall, 1995.

Sommer BG, Mitchell LH. *Vascular access for hemodialysis*. Pluribus Press, Chicago, 1989.

17.17 Limb amutation

Derek W. R. Gray

[The development of amputation techniques](#)
[Indications for amputation](#)
- [Trauma](#)
- [Ischaemia](#)
- [Malignancy](#)
- [Congenital deformity](#)
- [Other indications for amputation](#)

[The choice of amputation site](#)
- [Clinical assessment](#)
- [Other assessment techniques](#)

[Techniques of amputation](#)
- [Anaesthesia and preparation](#)
- [Surgical technique](#)
- [Lower limb amputations](#)
- [Upper limb amputations](#)
- [Postoperative management](#)
- [Complications of amputation surgery](#)

[Further reading](#)

The development of amputation techniques

For many centuries limb injuries have been common as a result of man's fascination with war. The outcome of even minor limb injuries inflicted on the battlefield was frequently fatal, in strong contrast to the minor importance accorded to such injuries in modern historical drama! The cause of these deaths was presumably sepsis from contaminated wounds containing devitalized tissue, but it was not until the last century that the important role of debridement was fully appreciated. The signs of increasing limb pain, swelling, and toxaemia were recognized as harbingers of death, and amputation the only cure. Despite the absence of aseptic technique, removal of the limb by incisions through healthy tissue with free drainage was often followed by recovery and healing. Although there are several descriptions of amputation dating back to Hippocrates, the technique was popularized by Ambroise Paré around 1575 and was rapidly taken up by other military surgeons and then transferred to civilian practice. In the absence of true anaesthesia it was not possible to perform other than a guillotine-type amputation and healing was always by secondary intention, although the later addition of a short flap by Yonge and Lowdham in 1679 made closure easier. To minimize the horrific experience surgeons became adept at rapid amputation, and apocryphal tales describe amputations performed in a few seconds, sometimes with the loss of the assistant's fingers.

The desire to replace an amputated limb with a prosthesis is as old as the procedure itself. There are many examples of false legs, mainly of wood, in ancient drawings and the use of a hook to replace the amputated forearm is also well known. The wooden peg-leg is common to many cultures and relied on the user kneeling on the weight-bearing surface. However there are also many examples of ingenious and complex artificial legs dating from the time of Paré.

The modern era of amputation surgery was heralded by the advent of anaesthesia and later by aseptic surgical techniques. It became possible to perform amputation using flaps fashioned to allow primary closure at a variety of levels. The huge number of amputees that followed the First World War provided a stimulus to the development of improved limb prosthetics, and many of the modern services originated at this time. Over the next few decades it was slowly appreciated that not all indications for amputation could be treated identically, and that patients presenting with infection or gangrene associated with loss of limb pulses had to be treated more cautiously in terms of amputation than patients undergoing amputation for trauma. Thus, Homans (1939) was voicing contemporary surgical opinion about the outcome of amputation in patients with vascular disease when he wrote ‘amputations below the knee can almost never be expected to offer a healthy stump'. Fortunately, this view has been proved incorrect, but the technical demands of amputations performed for limb ischaemia are emphasized by this quotation.

Indications for amputation

As a consequence of increasing specialization in surgery, limb amputations are performed by different specialists, depending on the indications. Thus orthopedic surgeons most often perform amputations for trauma, and vascular surgeons usually amputate for vascular disease. Although the techniques used for different indications are broadly similar there are some individual factors that must be considered.

Trauma

Since traumatic vascular injury is usually repairable, limb amputation is only indicated when there has been massive destruction of tissue, by either crush or blast injuries. The common causes are either motor accident or gunshot wounds in civilian practice. Ideally, the decision to amputate should be made immediately—a late amputation dictated by developing sepsis in an inadequately debrided limb represents a failure of management. Because the vascularity and health of the tissue proximal to the injury is usually good, the amputation can be placed as distally as possible, within the limitation imposed by the need for excision of all devitalized tissue. Contamination with soil and dirt is common and in these cases the wound should be left open and closed secondarily after several days. Even if the early swelling of skin flaps prevents full closure by secondary suture, the basic health of the tissue will usually allow good healing by granulation and amputation at a higher level should be considered only if serious sepsis ensues.

Ischaemia

Amputation of the limb of a patient with rest pain, sepsis, or gangrene with the aim of primary wound healing is perhaps one of the greatest challenges to a surgeon. Several factors conspire against the healing process. First, the tissue at the line of incision has often only slightly better vascular perfusion than the critically ischaemic or frankly gangrenous tissue distally, and the lack of oxygen and nutrients delivered to this tissue means that the ability to form new vessels in granulation tissue and the metabolic processes involved in healing are greatly impaired, as is the ability to fight bacterial contamination. The lack of perfusion may be made worse by the frequent coexistence of poor cardiac function, respiratory disease, and anaemia. Furthermore, many patients have sepsis already within the limb and there may be considerable oedema as a result of this and other coexisting conditions such as heart failure. Many patients are diabetic, further reducing resistance to infection.

It therefore behooves the surgeon to pay attention to these factors and correct those that are correctable, wherever possible. Heart failure, anaemia, chest infection, and limb infection should all be treated, since this may make the crucial difference between success and failure. Unfortunately, delay is not always humane or wise, since the toxaemia from an infected ischaemic leg may make factors such as heart failure worsen despite treatment until the limb is removed, and opiates required to treat severe rest pain may make a chest infection untreatable. The judgement to wait or proceed is often difficult and is probably best taken by the surgeon since other physicians may not appreciate the role of the ischaemic limb itself in dictating the general condition of the patient.

Malignancy

Amputation for a tumour affecting a limb is now a relatively rare occurrence, since most tumours can be controlled locally by a combination of radiotherapy and cytotoxic therapy, even if excision is not possible. When amputation becomes necessary it is always a planned procedure, usually in a limb with good vascularity without sepsis and the amputation is usually, therefore, technically quite straightforward. Psychological effects are likely to be the greatest problem in this group, since there is ample opportunity to brood on the forthcoming operation and the subsequent disability is compounded by the fear of recurrence of the tumour.

Congenital deformity

At one time amputation was considered the best option for a number of congenital deformities, on the basis that this gave superior cosmesis and would ultimately lead to better acceptance by society. However, it has now been realized that even patients with severely deformed limbs usually can obtain better function from the deformed limb, albeit with a number of prosthetic additions, than by amputating the limb and attempting to fit a conventional prosthesis. Furthermore, such patients are unlike normal limb amputees, in that they have never developed a body image of themselves with a full-size limb and do not adjust easily to a full-size prosthesis. A limb remnant represented by a single digit may be used to activate a control for a mechanized transport or prosthesis and should not be removed.

The only common exception to this rule is the deformity caused by congenital absence of the fibula, resulting in shortening of the lower leg. Conversion to a Syme's amputation produces an end-bearing stump which may be fitted to a foot prosthesis that can result in remarkably normal function in young children.

Other indications for amputation

Amputation is rarely required for a limb that is severely deformed, useless, or painful from acquired disease. Examples of conditions that sometimes require amputation are neurological diseases leading to a painful flail limb and chronic osteomyelitis, where renal failure due to subsequent amyloidosis is a danger. These indications have become increasingly rare since the following points have been understood. First, even a severely deformed limb can usually function better than a prosthesis. Second, a paralysed limb can still provide support following arthrodesis of appropriate joints or by the use of splints and gives better cosmesis. Third, persistent severe pain of central neurological origin may make a limb useless but, unfortunately, removing the limb often does not result in relief since the pain persists in the ‘phantom' limb.

The choice of amputation site

The site of amputation is influenced by several factors. The overriding general principle is that the least mutilating procedure possible should always be chosen, a principle that is particularly important in the lower limb, where the retention of the knee joint results in much greater mobility after rehabilitation. However other factors are also important, the most obvious being the indication for amputation. In the case of amputation for malignancy the amputation must be performed at a site that is sufficiently proximal to ensure clearance of the malignant process and minimize the chances of recurrence. This usually means removing the limb at a level that includes the joint proximal to the lesion. In the case of amputation for trauma the amputation site should be as distal as possible, subject to the possibility of fitting a usable prosthesis. In the case of amputation for ischaemia the choice of amputation site is a balance between using as distal a site as is compatible with removing dead or near-dead tissue and a proximal site sufficient to ensure healing of the wound.

In the past the ease of fitting the limb for a prosthesis was also a factor. For instance, above- or below-knee prostheses were easier to fit and functioned better than through-knee prostheses. However the steady advance in materials and design of prostheses has meant that this factor is less important than previously, although it is still difficult to get a good cosmetic result from a jointed prosthesis for long thigh amputations such as the Gritti–Stokes.

Clinical assessment

The most important requirement for successful amputation is skin healing, preferably by first intention. The skin that is to form the amputation flaps must at least be viable at the outset. Signs of demarcation, fixed staining, or anaesthesia of the flap skin are obvious contraindications to its use. The presence of excoriation, cellulitis, or frank ulceration are relative contraindications, although the likelihood that the phenomenon is entirely localized and that infection has been treated adequately must be considered. The tissues beneath the skin must also be healthy. The presence of deep tenderness and loss of function may signal necrosis of deep muscles or sepsis, which make an amputation unlikely to heal. The skin temperature is more a reflection of the ambient temperature, or the presence of sepsis or systemic illness and is not a reliable guide to subsequent healing. The absence of a pulse at the next most proximal palpation point has been said to be a contraindication to more distal amputation and vice versa, but this is certainly an unreliable sign. The decision about the level and type of amputation should also not be taken during the operation, since the degree of bleeding noted on incision of tissues at operation is a very poor guide to subsequent healing, being greatly affected by anaesthetic agents and transient hypotension.

The decision to undertake a more distal amputation is inevitably associated with a greater risk of failure to heal and the need for reamputation at a higher level. Furthermore, the exact magnitude of risk has been shown to vary considerably from centre to centre, and to make a rational decision it is necessary to know one's own results: a valuable contribution of continuous audit. The final decision should always take into account all of the factors influencing outcome including the physical and mental capacity of the patient to withstand a reamputation and whether the patient is likely to make use of the increased mobility that a joint-saving amputation may give, to name just two.

Other assessment techniques

A variety of techniques for assessing the blood supply to the amputation flaps, including thermography, transcutaneous oxygen measurement, cutaneous perfusion measurement, xenon-133 clearance, Doppler ultrasound, and digital plethysmography have been described. There is no doubt that in the hands of enthusiasts these techniques can allow the correct choice of amputation site in a larger proportion of cases than is probably achieved by clinical judgement alone. However, all of the techniques require some degree of expertise in their performance and some are not readily available in many hospitals. No single technique has been shown to be sufficiently superior to the others to attain widespread acceptance, although transcutaneous oxygen measurement is probably the most valuable.

Techniques of amputation

Anaesthesia and preparation

Although general anaesthesia is preferred by most centres it is possible to use local anaesthesia, nerve block, epidural, or spinal anaesthesia in appropriate situations. Some surgeons object to the use of techniques other than general anaesthesia on the grounds that the noise and motion associated with amputation is psychologically distressing to the patient, even in the absence of pain. However, the use of amnesic sedative agents during the procedure is highly effective and most patients remember nothing if such agents are used.

Local anaesthesia is only useful for digital amputation and has the potential disadvantage of causing local swelling of the tissues which may make closure less easy. There is also the theoretical (but in practice, remote) disadvantage of spreading infection. Local anaesthetic combined with epinephrine must never be used as this may further prejudice the blood supply.

Nerve block techniques have the advantage of avoiding the dangers of general anaesthesia and hypotension, but unfortunately tend to be unreliable, and conversion from nerve block to general anaesthesia in the middle of an amputation is distressing not only to the patient but also the theatre staff.

Epidural or spinal anaesthesia is usually very adequate for amputation of the lower limb, producing vasodilation as a by-product which certainly does no harm from the point of view of the procedure. A tourniquet should never be used for patients with vascular disease and is usually unnecessary even in patients with normal vessels. Broad-spectrum antibiotics, such as penicillin, with activity against *Clostridium* spp. must always be given preoperatively.

After careful identification of the patient and the correct limb, the limb should be carefully painted with skin-prep prior to starting surgery, wrapping any gangrenous or infected portion in a plastic bag, sealed to the skin above by adherent non-porous tape. The incision lines for flap formation must always be marked out using a sterile marker.

Surgical technique

In modern surgical practice in Europe and the United States approximately 80 per cent of amputations are performed for ischaemia secondary to vascular disease. The life expectancy of this generally elderly group of patients is limited, mainly due to concomitant cardiovascular and cerebrovascular disease. Approximately 50 per cent of vascular amputees will die within 3 years. The main consideration in performing the amputation is to obtain stump healing. In non-vascular amputations, such as those for trauma or tumour, the patient is often young and a lifetime of prosthetic use must be anticipated. In these patients more advanced procedures to obtain maximal function may be indicated. Since the majority of surgeons perform lower limb amputations for ischaemia, most of the techniques described below will be those recommended for this indication. Modifications and additional procedures applicable to amputation in non-ischaemic limbs will be considered as addenda to each technique. In those amputations indicated mainly for non-ischaemic disease, such as those in the upper limb, the reverse order will apply.

The amputation technique should be governed not only by the disease process and anatomic principles but also by the prosthesis available to replace the severed limb. A comment on prosthetics is therefore a necessary adjunct to the description of each amputation technique.

Lower limb amputations

Toe amputation

A ‘racquet'-shaped incision is recommended (Fig. 1). The tissues should be severed from the incision line directly down to the shaft of the phalanx and dissection of tissues should then continue proximally as close to the bone as possible, to avoid the possibility of damaging the digital vessels laterally and thus compromising the blood supply to the flaps and other digits. It is not usually wise to amputate just the distal phalanges for vascular disease and the dissection should continue up to the metacarpophalangeal joint, which is disarticulated to allow removal of the digit. The head of the metatarsal should then be removed by the careful use of nibbling forceps preventing further damage to the surrounding tissues. The bone should be nibbled back until the tissues can be seen to be falling in to cover approximately half of the cut end of the bone. Tendons should be pulled down with forceps, cleanly severed as short as possible, and allowed to retract. The incision should not be sutured closed but is best packed lightly with ribbon gauze, to limit bleeding, and allowed to heal by granulation. Prostheses are not usually required.

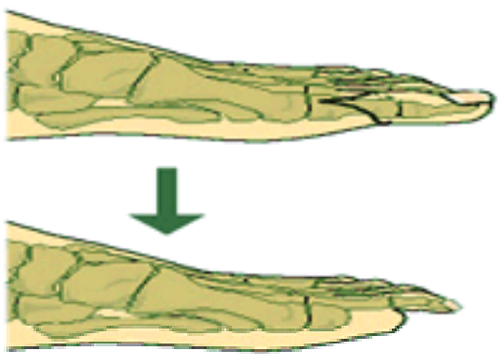


Fig. 1. ‘Racquet' incision for toe amputation.

The usual indication for toe amputation in non-ischaemic limbs is for trauma. The wounds can be closed primarily provided that there is no danger of contamination. Amputation of distal phalanges may be contemplated for cosmetic reasons.

Transmetatarsal amputation

This amputation performed in an ischaemic limb allows surprisingly good function, despite the loss of the great toe, but is undoubtedly the amputation with the highest failure rate. The important requirement is an adequately vascularized flap of sole skin, allowing application of the technique shown in Fig. 2. It is uncommon to have an adequate vascular supply in the presence of gangrenous toes, and for this reason the amputation is not often performed for vascular disease. However the correct indications do occur sometimes in diabetic patients who have localized digital gangrene in the presence of foot pulses. If insufficient vascularized sole skin is available to ensure primary closure the amputation should not be performed as secondary closure often fails in patients with vascular disease. The dorsal transverse incision is first made over the metatarsals as shown (Fig. 2), incising through the dorsal tendons down to bone. The metatarsals are divided using bone cutting forceps, or more satisfactorily, a power saw, and then the distal bones are elevated to allow careful dissection posteriorly close to the periostium and eventual separation and removal of the forefoot, leaving behind the sole, which should be kept as long as possible at this stage. Each metatarsal shaft should then be nibbled back for a distance of 1.5 cm using bone nibbling forceps to allow the tissues to fall in over the bone ends. The exposed tendons should then be removed from the sole flap by pulling down with forceps and then cleanly severing the tendon as short as possible. After ensuring haemostasis, the flap is brought forward over the metatarsals and cut to the appropriate length to allow closure without tension, using deep interrupted absorbable sutures and an atraumatic skin closure (either Steri-Strips or subcutaneous Prolene inserted without instruments is preferred by the author).

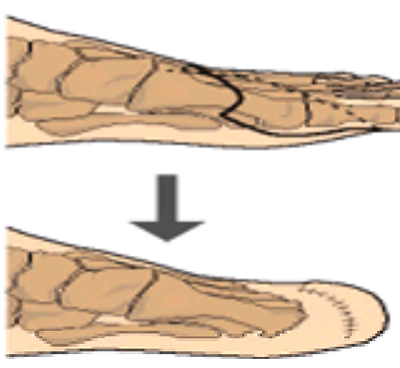


Fig. 2. Incision for transmetatarsal amputation, requiring a long viable flap of sole skin.

In patients without ischaemic disease it is possible to leave the flaps open if gross contamination has occurred due to trauma, and use skin grafts later if full closure cannot be obtained.

A simple shoe filler is usually sufficient to provide a good functional result.

Transtarsal amputations (Chopart, Lisfranc)

Owing to the poor healing and inevitable equinus deformity caused by unopposed contraction of the ankle flexors these amputations are not recommended in patients with or without vascular damage.

Syme's operation

Syme's operation is the best amputation for patients with extensive non-ischaemic damage to the forefoot. The stump that is produced has the advantage of allowing considerable weight bearing, and thus allows considerable independence in the home without a prosthesis. It may suffice alone in areas where prostheses are not available. The defects of the procedure used to be difficulty in fitting a foot prosthesis because of lack of clearance from the ground, which has now been overcome, and poor cosmesis of the prosthesis in females, due to the thickness at the ankle: this is still a problem. There is a difference between the technique that can be used in children and in adults although the incisions used are similar (Fig. 3). In children the amputation uses the same heel flap to cover the end of the stump, but the amputation itself can consist of simple disarticulation of the talus, followed by removal of the malleoli, leaving behind most of the joint surface. In adults the incision is made as shown in Fig. 3. The anterior incision is then deepened through the anterior tendons to allow disarticulation of the talus from the joint. The Achilles tendon is divided posteriorly and the talus and calcaneum are then dissected free from the heel by sharp dissection, staying close to the periosteum to avoid damage to vascular structures. The dissection is particularly difficult over the thin skin at the back of the heel, and there is a danger of ‘buttonholing' the skin, which can again be avoided by staying close to the periostium. A variant operation described by Boyd is to leave a slice of the calcaneum attached to the heel skin, which can then be attached to the proximal cut bones in a similar manner to the Gritti–Stokes operation. This may lessen the trauma to the tissue and lengthens the stump.

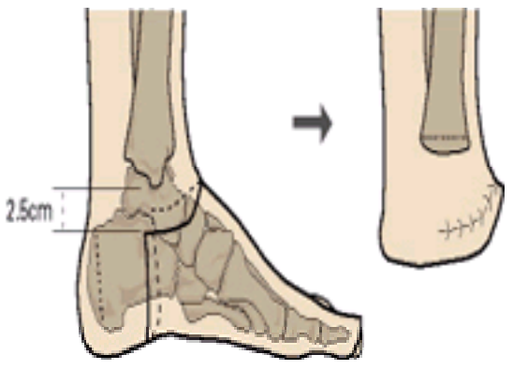


Fig. 3. Incision for Syme's amputation.

Once the posterior flap has been dissected free the distal foot is removed, leaving the heel flap long at this stage. The exposed distal tibia and fibula are dissected free of adherent tissues staying on the periosteum. A transverse cut is then made approximately 1 cm above the ankle mortice. It is important that this cut is truly transverse, to prevent uneven pressure later when weight bearing. It is also important to leave some element of the malleolar prominence on each side, since this allows stabilization of the prosthesis and prevents rotation. After cutting tendons short and allowing them to retract and ensuring haemostasis the flap is brought over the cut end of the bone and trimmed to a suitable length without tension. The wound is closed by suturing the deep tissues of the heel pad to the deep fascia anteriorly with a suitable atraumatic skin suture.

Syme's amputation requires a relatively good blood supply to the heel pad and most patients that need to lose the foot from vascular disease do not have sufficiently good blood supply to allow this procedure. However, some diabetics with forefoot infection or gangrene may have normal ankle pulses and a Syme's amputation can be successful.

An available prosthesis is shown in Fig. 4.



Fig. 4. Bivalved plastic prosthesis for Syme's amputation. Note the poor cosmetic result.

Below-knee amputation

This is the most common amputation performed for vascular disease, and possibly the most demanding technically. Preserving the knee joint is of such major benefit in terms of rehabilitation that below-knee amputation should usually be undertaken in all patients with non-reconstructible ischaemia distal to the calf, unless there are clear indications that the operation is certain to fail. Such contraindications would include necrosis of the potential posterior flap skin, or a patient who is certain not to benefit from the advantages of a below-knee prosthesis, for example where the knee has been arthrodesed. The operation described is a variant of the Burgess long posterior flap technique. A more complex technique, based on similar principles, is the skew-flap technique (see Further reading), which probably has similar healing rates but produces a more ideal conical stump.

The important principles of the technique are shown in Fig. 5, Fig. 6, Fig. 7 and Fig. 8. The stump should be as short as possible to maximize the chances of healing but long enough to allow optimal use of a below-knee prosthesis: 9 cm from the knee joint line is ideal in thin legs, 11 cm in fat legs, and 7.5 cm can be regarded as the absolute minimum. A long posterior flap is used which should initially be kept as long as possible, and only trimmed prior to final suturing. The flap is based on half the circumference of the leg to maximize the blood supply from above. After outlining the flaps the skin incision is deepened through fascia, marking the tibia anteriorly to denote the later line of bone section. The muscle of the anterior and lateral compartments is divided to expose the fibula, ligating the anterior tibial vessels. The fibula is divided with a lateral bevel 1.5 cm proximal to the tibial line. The tibia is divided at the marked line with an anterior bevel and the lower leg removed by dissecting the posterior flap away from the bones. Bone dust should be washed away to prevent spur formation. Bleeding can be controlled by folding the flap forward and applying pressure. The bulk of the flap should be reduced by complete removal of the soleus muscle (Fig. 6, Fig. 7), which does not contribute to the blood supply of the flap. If the lateral portions of the gastrocnemius muscle bellies are particularly prominent these should also be trimmed to reduce ‘dog-ears'. The posterior tibial and peroneal vessels are individually ligated whilst removing soleus. The flap is then brought forward to the tibia, cut to a length that will allow closure without tension and sutured, deep fascia to periostium and deep fascia, over a vacuum drain. Skin closure should be performed with meticulous care of the skin. The author's preference is for subcuticular Prolene inserted by hand needle, without the use of instruments, and covering Steri-Strips (Fig. 8).

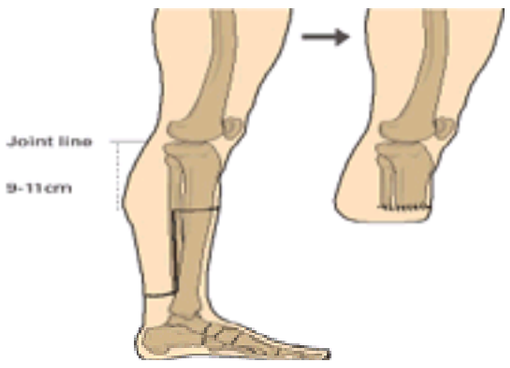


Fig. 5. Incision for below-knee amputation. The posterior flap should be left as long as possible and only cut to an appropriate length prior to suturing.



Fig. 6. Separation of the plane between soleus and gastrocnemius.



Fig. 7. Muscle bulk of soleus to be excised.



Fig. 8. Below-knee amputation before dressing.

A longer stump may be of considerable benefit in young patients with non-ischaemic limbs: 15 cm is ideal. Short anterior or lateral flaps may be used to keep the scar away from the prosthetic-bearing surfaces. More complex myodesis procedures may also allow better function. A fibula bone bridge can be inserted between the tibia and fibula to allow the Achilles tendon to be brought over and sutured to holes in the anterior tibia. The anterior and posterior tibial muscles can be sutured together over the interosseus membrane.

Most below-knee prostheses rely on a patellar tendon-bearing socket with either strap or suction suspension (Fig. 9, Fig. 10). A variety of ankle joints with varying degrees of ankle and toe flexion are available.

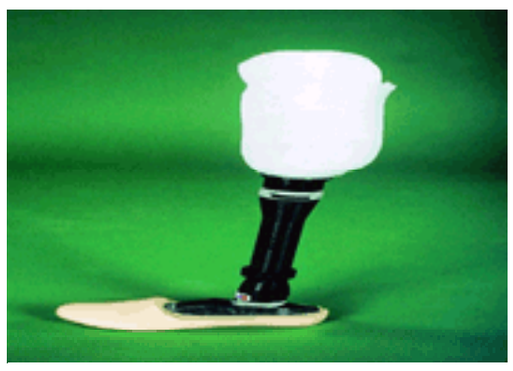


Fig. 9. Below-knee patella tendon-bearing ‘Endolite' prosthesis without cosmesis.

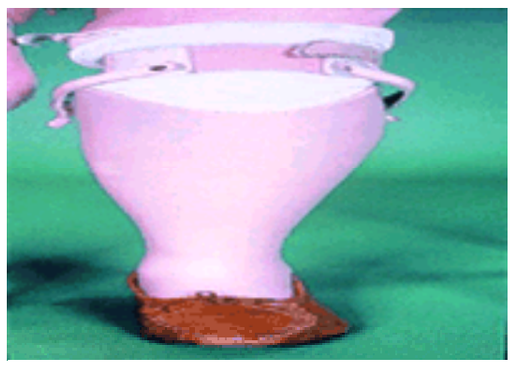


Fig. 10. Below-knee patella tendon-bearing prosthesis with cosmesis and cuff suspension.

Through-knee amputation (knee disarticulation)

In non-ischaemic limbs where the knee joint is non-functional or when the limb is damaged in such a way that a below-knee amputation is not possible this operation has several advantages: it is quick and easy to perform, using the lateral flaps shown in Fig. 11; the stump is end-bearing, has good leverage, and the bulbous bony

shape can be used to suspend the prosthesis. Modern polycentric knee prostheses have overcome many of the difficulties of limb fitting and the cosmetic result is reasonable.



Fig. 11. Incision used for through-knee amputation (knee disarticulation) and final position.

The flaps required to cover the bulbous end of the femur are of considerable length and use less-well vascularized lateral skin. Healing may thus be poor in patients with vascular disease. A below-knee amputation will heal as well and if the ischaemia is considered inadequate for a below-knee amputation a knee disarticulation is unlikely to heal either. This operation is not usually performed in patients with vascular disease, except where retention of the knee joint is contraindicated and a fast bloodless amputation is required.

Prostheses which are available tend to be rather bulky and of similar design to those used after Gritti–Stokes amputation with external joints (see Fig. 13). However the bulbous end can be used to allow suspension without a thigh corset and the socket can be end-bearing.



Fig. 13. Prosthesis for Gritti–Stokes amputation: an external knee joint must be used.

Gritti–Stokes amputation

The Gritti–Stokes amputation is a variety of through-knee amputation that has the advantage of using relatively well vascularized anterior skin, and so is useful for patients with or without vascular disease. The operation is useful for the few patients with ischaemic limbs who appear to have good perfusion below the knee but do not have the posterior flap of skin available to allow below-knee amputation (perhaps because of local infection or ulceration) and in those patients where retention of the knee joint is of no advantage. Gritti–Stokes amputation gives a long lever with partial end-bearing, but not as good as that obtained by knee disarticulation, and although the end is bulbous, direct suspension of the prosthesis using the bulbous end is often not possible. Prostheses are good functionally but cosmetically inferior to those available for patients with above-knee amputation.

The basic points of the technique are shown in Fig. 12. The key point is the fixation of the patella remnant to the cut end of the femur. Some authorities recommend drilling the femur and using wire sutures for this purpose, but the author has found that six strong absorbable sutures between the strong tissues at the periphery of the patella to the periosteum and surrounding tissues of the femur are adequate to ensure bony union. It is also helpful to make the cut through the femoral condyles proximal enough to ensure there is relatively little tension and with a slight backward slope that encourages the patellar remnant to stay attached (Fig. 12).

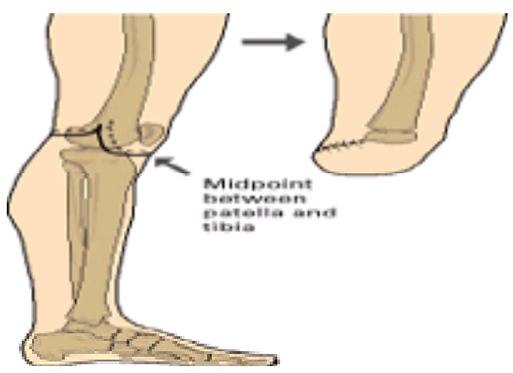


Fig. 12. Incision and principles of Gritti–Stokes amputation.

An available prosthesis is shown in Fig. 13.

Above-knee amputation

The principles of the technique for patients with ischaemic limbs are shown in Fig. 14. Important points to note are that although a long femoral stump has advantages for leverage of a prosthesis this should not be at the expense of poor healing, as is often the case in excessively long stumps where the bone may be inadequately covered with soft tissues and the skin is stretched. In any case the site of division of the femur should not be more than 12 cm from the knee joint line to allow room for the knee mechanism in the prosthesis. The minimum length of stump that can be used to fit an above-knee prosthesis is 7.5 cm below the adductor muscle insertion. If the amputation must be shorter than this and it is hoped eventually to fit a prosthesis, a hip disarticulation should be performed, as this allows better fitting of modern prostheses and reduces weight.

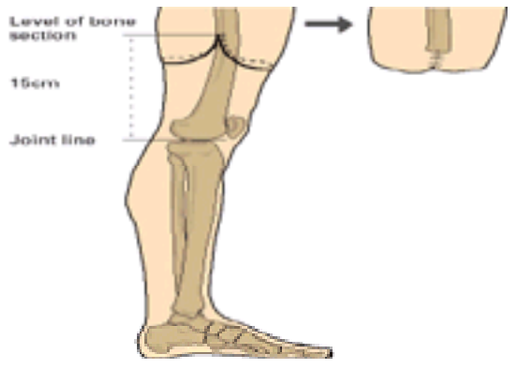


Fig. 14. Incision for above-knee amputation.

The bone end should be rounded off and covered by suturing the deep fascia of the posterior flap, with its attached muscle, to the same structures anteriorly, enclosing a vacuum drain. The author prefers a subcuticular polypropylene suture with additional adhesive strips for closure.

Where amputation is performed in young patients with good vasculature the addition of more complex myodesis techniques to suture and overlap opposing groups of muscles using drill holes in the distal femur produces a stump that is a conical shape with better function. However these techniques do not work well in ischaemic limbs and may result in devitalization and failure of healing.

Prostheses available are shown in Fig. 15 and Fig. 16.



Fig. 15. Sophisticated knee joints are available to allow as near-normal gait as possible. Illustrated is a stabilized knee with pneumatic swing-phase control.

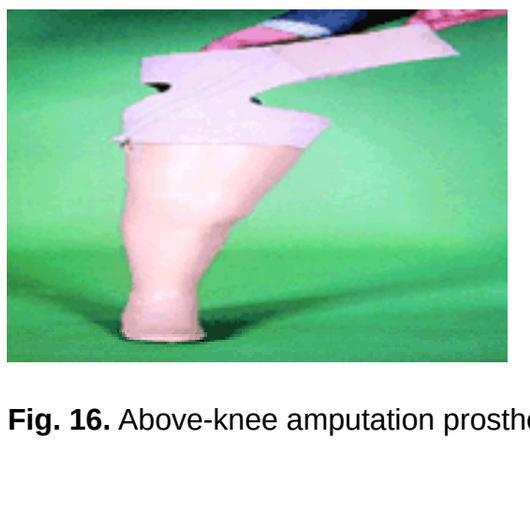


Fig. 16. Above-knee amputation prosthesis with cosmetic outer covering and soft neoprene suspension.

Hip disarticulation and hemipelvectomy

There is insufficient space to describe in detail the dissection required for these operations, which are usually performed in non-ischaemic patients, and the reader is directed to the references at the end of the chapter. The principle is to remove the required extent of the limb leaving a posterior flap with a layer of viable muscle that can be used to cover and cushion the deep structures (Fig. 17).

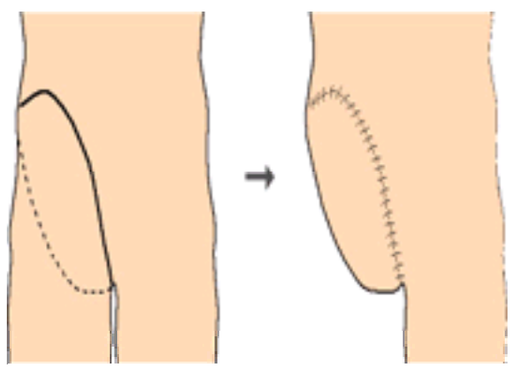


Fig. 17. Incision for hip disarticulation using a long posterior flap of skin and muscle to cover the defect.

Hip disarticulation and/or hemipelvectomy are rarely required for ischaemia, and when they are required infective gangrene is the usual cause. It is often not possible to produce the ideal flaps described for elective operations, and it is a question of using the best vascularized non-infected flap of skin available to cover the amputation site. Vascularized free flaps may be necessary. Mortality is high and serious consideration should be given in the elderly to non-intervention to allow death with dignity.

An available prosthesis is shown in Fig. 18.

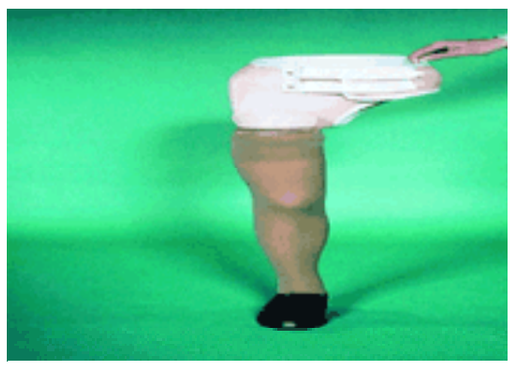


Fig. 18. Prosthesis for through-hip disarticulation. Note the anterior hinge joint.

Upper limb amputations

Finger and partial hand amputation

The usual indication is trauma and the general principle is to preserve as much viable tissue as possible, particularly preserving palmar surface skin and as much of the thumb as is practicable. Multiple injuries require specialist attention. Medial and lateral flaps should be used, which may be closed unless there has been gross contamination. If all the fingers and the palm of the hand are lost the wrist should be preserved since useful grip function can be obtained by use of an opposition plate.

Ischaemic fingers should be amputated as for ischaemic toes (see above), leaving the wound open.

Cosmetic fingers are available but most patients choose to do without, since the prosthesis tends to get in the way.

Forearm amputation

The usual indication is for trauma. Equal length extensor and flexor flaps are fashioned as shown in Fig. 19, allowing bone section at around mid-forearm, if possible. Longer stumps han this have little advantage and tend to heal poorly. The minimum length that can work a forearm prosthesis is 5 cm below the biceps insertion. The extensor and flexor muscles are then sutured together under gentle tension over the bone ends, including the deep fascia with each bite.

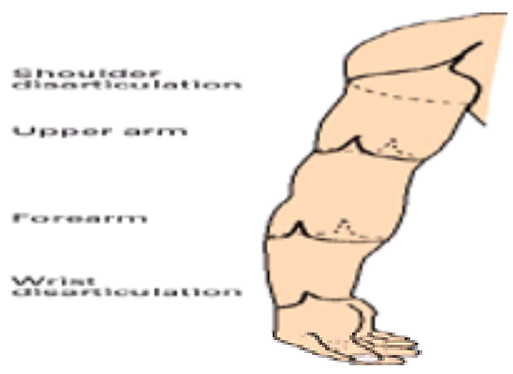


Fig. 19. Incisions used for amputations of the upper limb.

A variety of forearm prostheses are available (Fig. 20, Fig. 21 and Fig. 22). Some use mechanical hands with the grip mechanism worked by a shoulder strap and cable (Fig. 20), or an electrically driven grip activated by muscle depolarization (Fig. 21). However, many patients find these cosmetic prostheses both heavy and functionally poor. The most functional prosthesis remains a split hook (Fig. 22) with special tools which can be fitted for specific tasks.



Fig. 20. Below-elbow amputation prosthesis with mechanical hand worked by a shoulder strap and cord.



Fig. 21. Below-elbow prosthesis with electrically operated hand grip and myoelectric activator (child's).

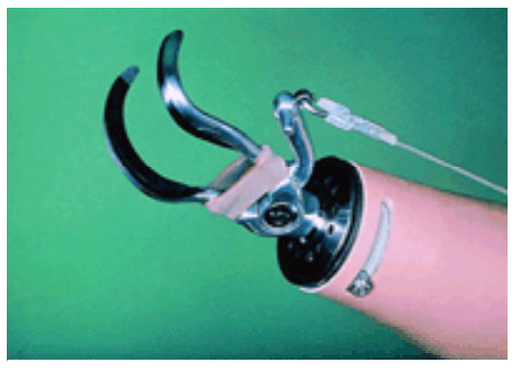


Fig. 22. Split hook prosthesis with wrist rotary action.

Elbow disarticulation

This amputation is not favoured, as fitting a useful prosthesis is difficult.

Above-elbow amputation

Equal length anterior and posterior flaps are used (Fig. 19). The bone should ideally be divided no closer than 9 cm above the medial epicondyle, to allow room for an elbow mechanism. The minimum length of humerus that allows an above-elbow prosthesis is 5 cm below the anterior axillary fold. However, provided that there are no overriding reasons such as clearance of malignancy, even a non-functional amputation across the humeral neck should be retained in preference to disarticulation, since it gives better cosmesis. The anterior muscle flap should be sutured over the bone end to the posterior muscle flap, with bites that include the deep fascia.

Although cosmetically acceptable prostheses are available, some with active grip using a shoulder-activated cable (Fig. 23), the functional result is poor and these limbs are rarely used functionally if the patient has a normal contralateral arm.

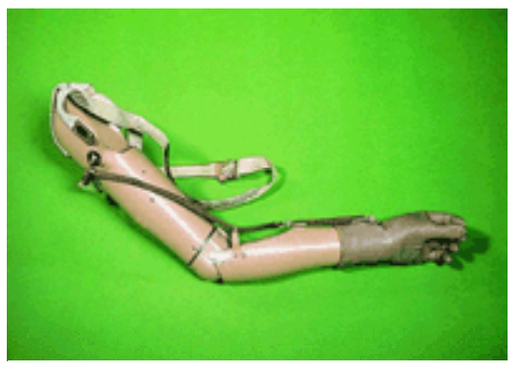


Fig. 23. Above-elbow prosthesis with mechanical hand and Bowden operating cable.

Shoulder disarticulation and forequarter amputation

Virtually the only indication for these procedures is amputation for malignancy. There is insufficient space to discuss the details of the dissection required, but the incision for both is placed so as to allow a predominantly posterior flap to be brought forward to cover the amputation site, with an extension of the incision anterosuperiorly to allow access for proximal ligation of the limb vessels (Fig. 19). The operation results in considerable deformity which is difficult to disguise with prostheses that allow normal fit of clothing and retain function.

Although functional prostheses are available (see below) they are weighty and have poor function, so that many patients opt for a cosmetic shoulder filler only.

Auxiliary procedures for upper limp amputees

A number of complex operations have been described in the past to allow better function of arm prostheses. Most of these have now been abandoned and those that are still performed are usually not worth considering in a patient who has one normal arm. The formation of a skin-lined tunnel in the biceps muscle (‘kineplastic tunnel') can sometimes be used to activate the terminal device of a forearm prosthesis. Patients that have lost both forearms can be given grip function by Krukenberg's operation, which separates the radius and ulna to act as separate jaws of a pincer. Although surprisingly good function can be obtained, the cosmetic penalties are considerable and are accepted by few patients.

Postoperative management

There are considerable differences in opinion on the dressing, bandaging, and prosthetic fitting of amputation stumps. It is the author's opinion that these differences stem from the experiences of different specialties dealing with either predominantly amputations performed for vascular disease or for non-vascular conditions, and the management of these groups should probably be different.

Amputations for vascular disease should be dressed with a loose-fitting stump bandage, without compression. The bandage will tend to fall off, but compression may produce local necrosis of skin that is balanced on the edge of ischaemia, and a tourniquet effect, which is easily produced by inexpert bandaging, can be disastrous. The author recommends the use of meticulously applied longitudinal straps of adhesive tape, running from groin to the end of the bandage, placed at four points around the limb. Although the dressing is best left undisturbed if all is well, the stump must be viewed if there is any question of infection. The use of a plaster cast, suggested by some authors, is not routinely recommended since it impedes the ease of access to the stump, but can be useful to prevent ‘picking' by a confused patient. Gentle active exercise should be encouraged by the physiotherapist, within the limits of pain. Some authorities recommend immediate fitting of a prosthesis and encourage rapid progression to walking within a few days. In the author's opinion this method may be possible in a centre with prosthetists able to attend at all times and the staff able to supervise walking properly, but for most hospitals such facilities are not available and this approach is potentially dangerous. Older patients require continuous reminders that they have lost a limb during initial mobilization, and a fall on to the stump causing disruption is a common event leading to non-healing if supervision during mobilization is inadequate. The first goal is wound healing and the author recommends that the stump be left until primary healing of the wound is well under way, usually 2 weeks or more, before trying a few steps under supervision in a pneumatic pylon to regain bipedal balance. Subsequent progression to an adjustable training prosthesis encourages early learning of correct gait with moulding and shrinkage of the stump and the final prosthesis can often be fitted by 6 weeks.

Refinements such as preoperative physiotherapy and counseling, postoperative compressive stump bandages, and plaster casts with fitted prostheses to allow immediate mobilization are applicable to young amputees, particularly those undergoing amputation for malignancy, and remarkable results can be obtained in this group.

Most patients, particularly the elderly, need to use a wheelchair some or all of the time around the house, and the home must be visited at an early stage to assess the modifications necessary for a wheelchair to be used before the atient returns.

Complications of amputation surgery

Early complications

Primary haemorrhage should be a rarity if blood vessels are ligated, but venous oozing is common and should be prevented from collecting and causing swelling and stump breakdown by insertion of a vacuum drain in all amputations. Infection is common and is usually due to *Staphylococcus aureus*, but other organisms such as *Clostridium* spp. may occur, particularly in ischaemic limbs. Strict aseptic technique, avoidance of excessive dressings, and use of appropriate preventative antibiotics preoperatively and for at least 5 days postoperatively is recommended. Extensive infection despite these measures usually denotes inadequate blood supply and reamputation to a higher level is needed. Small areas of skin necrosis in the suture line may remain uninfected and subsequently heal, helped by wedge excision surgery if necessary, but extensive necrosis with pain should not be left until it becomes infected: early reamputation is a better option.

Late complications

Provided that skin healing is obtained, most stumps can be fitted with a usable prosthesis and the site of the scar is usually immaterial. The exception is a scar that is adherent to a bony prominence since this will eventually ulcerate. Sometimes a thick skin graft or myoplastic flap can be swung in to cover the area, usually requiring the assistance of the plastic surgery department. Otherwise this problem may prevent satisfactory prosthesis use.

Some stumps with redundant tissue develop epidermoid cystic change at pressure points that often become infected. These can be excised, but unless the fundamental problem of chafing and maceration by the prosthesis is cured recurrence is inevitable. Terminal chronic oedema with verrucose hyperplasia occurs in stumps that are not supported sufficiently at the distal end (Fig. 24).

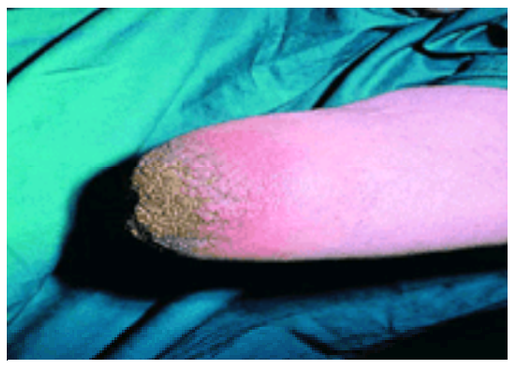


Fig. 24. Below-knee amputation stump with terminal verrucose hyperplasia due to inadequate distal support.

Stump neuromas, sometimes up to golf-ball size, are common, but unless they lie under a point of pressure and a trigger point is demonstrable they rarely need excision. Persistent stump or phantom pain is a problem in some patients. Sometimes ischaemia is the obvious cause but in others this is not the case. Phantom pain is a major problem in some patients. Its severity is related to the duration of preoperative pain experienced by the patient, and is one good reason why amputation should not be delayed in patients with rest pain due to non-correctable ischaemia. A claim that preoperative regional anaesthesia could reduce postoperative phantom pain has not been supported by recent studies. Although phantom limb pain will usually lessen with time, this is not always the case. Carbemazepine and amitryptylline may be useful in some patients. Early referral of patients with severe persistent phantom pain to a specialist pain management unit is recommended, where techniques such as ultrasound therapy and transcutaneous electrical nerve stimulation may be helpful, combined at times with psychiatric support and antidepressive therapy.

Further reading

Burgess EM, Romano RL. The management of lower extremity amputees using immediate postsurgical prostheses. *Clinical Orthopaedics and Related Research* 1968; **57**: 137–56. [Original description of below knee amputation using long posterior flap.]

Burgess EM, Romano RL, Zettl JH, Schrock RD. Amputations of the leg for peripheral vascular insufficiency. *Journal of Bone and Joint Surgery* 1971; **53A**: 874–90. [Early description of results of below knee amputation.]

Gray DWR, Ng R. Anatomical aspects of the blood supply to the skin of the posterior calf: technique of below-knee amputation. *British Journal of Surgery* 1990; **77**: 662–4. [Details of blood supply relevant to below knee amputation.]

Greenhalgh RM, Jamieson CW, Nicolaides A, eds. *Limb salvage and amputation for vascular disease*. Saunders, Philadelphia, 1988. [A comprehensive textbook.]

Kostuik JP, Gillespie R, eds. *Amputation surgery and rehabilitation: the Toronto experience*. Churchill Livingstone, New York, 1981. [An account of efficient amputation management.]

Moore WS, Malone JM, eds. *Lower extremity amputation*. Saunders, Philadelphia, 1989. [A useful textbook.]

Murdoch G, Donovan RG, eds. *Amputation surgery and lower limb prosthetics*. Blackwell Scientific, Oxford, 1988. [Good description of prostheses.]

Murdoch G, Bennett Wilson Jr. A, eds. *Amputation: surgical practice and patient management*. Oxford: Butterworth-Heinemann, 1996. [Up-to-date textbook.]

Robinson KP, Hoile R, Coddington T. Skew flap myoplastic below-knee amputation: a preliminary report. *British Journal of Surgery* 1982; **69**: 554–7. [Original description of a useful technique.]

Royal College of Surgeons. Report of a symposium on amputations and prosthetics held at the Royal College of Surgeons of England. *Annals of the Royal College of Surgeons* 1967; **40**: 203–88. [A useful summary article.]

Sherman RA. *Phantom pain*. New York: Plenum Press, 1997. [Multi-author text on latest management.]

Vitali M, Robinson KP, Andrews BG, Harris EE. *Amputations and prostheses*, 2nd edn. Baillière Tindall, London, 1986. [A comprehensive textbook.]

18.1 Venous disorders, vascular malformations, and chronic ulceration in the lower limbs*

David J. Tibbs and John H. Scurr

Normal anatomy and physiology
Essential properties of lower limb veins
Venous physiology
Clinical symptoms and signs of venous disease
Clinical examination and special investigations in patients with venous disease
Clinical examination
Clinical tests
Venous investigations
Clinical patterns of venous disorder
Primary varicose veins
Syndromes of valve deficiency and weak veins
Valveless syndrome or primary valve deficiency
Weak vein syndrome
Deep vein thrombosis and post-thrombotic (or post-phlebitic) syndrome
Investigations and treatment
Prophylaxis
Venous ulceration
Surgical treatment of post-thrombotic syndrome
Perforator veins
Congenital venous disorders
Venous anomalies in the lower limb
Further reading

The veins form part of a complex pumping mechanism returning blood to the heart against the force of gravity. If active thrombosis is excluded, it is some form of failure in this mechanism that underlies nearly all venous disorders in the lower limb. The essentials of normal anatomy and physiology will briefly be described here, particularly those aspects that may undergo change and cause venous problems. Note that the term ‘leg' is used in the anatomical sense, referring to the portion of lower limb below the knee.

Normal anatomy and physiology

Essential properties of lower limb veins

Structure

The veins are specifically designed to allow flow in one direction. The presence of numerous valves prevents venous reflux. The valves are supported by the vein wall and their integrity depends on the vein having sufficient strength to prevent dilatation. In the upright position the veins distend, but will collapse when the person lies flat.

The vein is lined with endothelium, providing a non-thrombo-genic surface. The endothelial layer may be damaged as a result of long periods of stasis, hypoxia, and an interaction with the white cells. Such damage will enhance thrombogenesis, but the veins themselves have a protective mechanism resulting in the production of fibrinolysins capable of dissolving clot.

Arrangement of deep veins

The deep veins lie beneath the deep fascia; the superficial veins lie superficial to the deep fascia. Some of the veins act as conduits and others, venous sinuses located within the muscles, form an important part of the pumping mechanism (Fig. 1, Fig. 2).

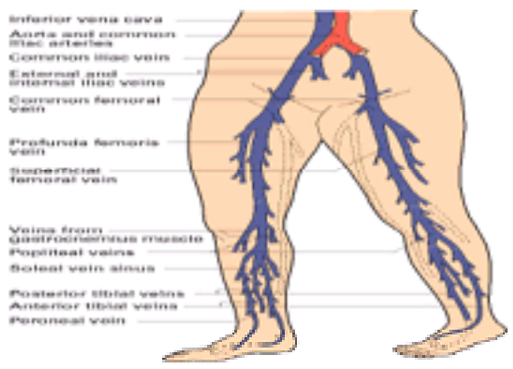


Fig. 1. No caption available.



Fig. 2. No caption available.

Veins as pumping chambers

Venous pumps are described in the foot, the calf, and the thigh (Fig. 3, Fig. 4). On walking and on muscle contraction, the veins are compressed. The valves ensure the blood is moved towards the heart and this mechanism is important in assisting venous return. Damage to the joints, the muscles, or the valves will interfere with this mechanism.

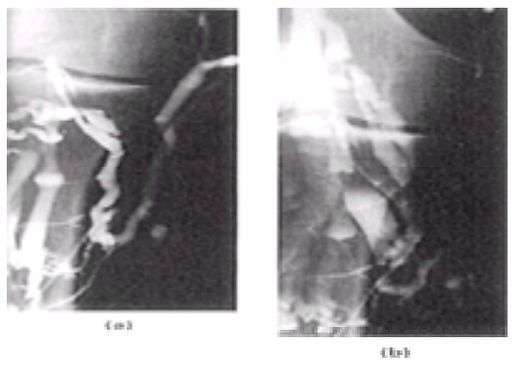


Fig. 3. No caption available.

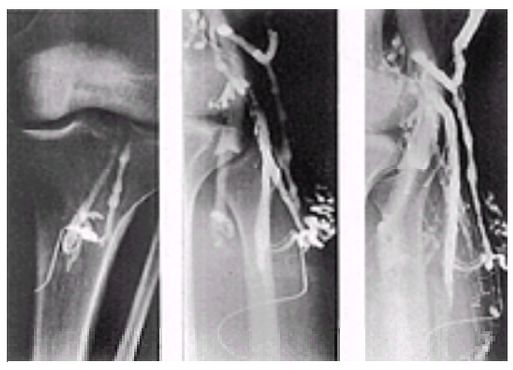


Fig. 4. No caption available.

With an impaired venous pump, normal venous return does not take place; venous pressure, instead of falling, will remain the same during exercise. If there is evidence of venous obstruction, then during exercise the venous pressure will rise. This has pathophysiological consequences—the development of skin changes and ulceration.

Superficial and perforating veins

Superficial

The superficial veins (superficial to the deep fascia) act as conduits taking blood from the surface to the deep veins via perforating veins. Two clearly identifiable systems with free interconnection between them (Fig. 5) can be identified. One the long saphenous vein draining the inner leg to the groin, the other the short saphenous vein draining the back of the calf to the popliteal vein behind the knee. It is dilatation of these veins that plays an important part in temperature regulation. Gross dilatation will lead to the development of varicose veins. Both systems have valves along their length, becoming more numerous in the lower leg (Fig. 6).

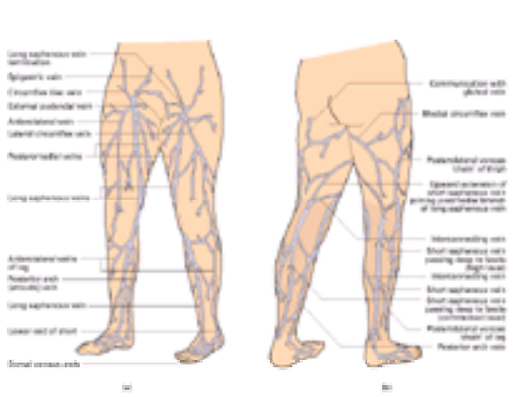


Fig. 5. No caption available.

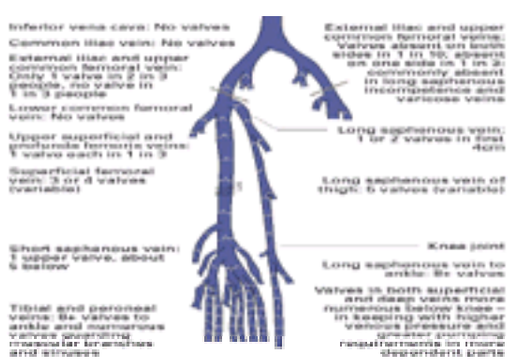


Fig. 6. No caption available.

Perforating

Perforating veins connect the superficial veins to the deep veins. In addition to the connections between the long saphenous and common femoral vein, and the short saphenous and the popliteal vein (Fig. 7), there are estimated to be over 60 other sites of communication. The perforating veins allow flow from the superficial to the deep system. In some perforating veins, valves can be identified; in others the inward flow seems to be regulated by muscle contraction.

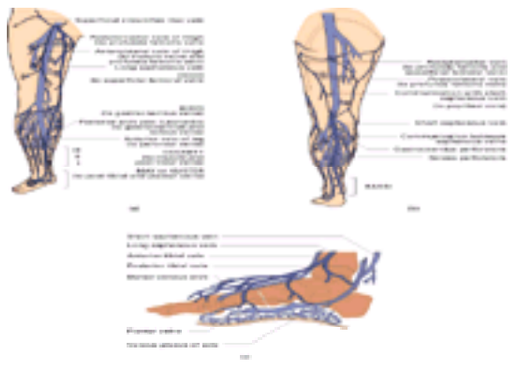


Fig. 7. No caption available.

The role of perforating veins in the causation of venous problems remains controversial.

Collateral flow

Collateral veins will develop when other veins become obstructed; occlusion of the femoral vein may result in dilatation of the superficial veins providing an alternative drainage mechanism. Patients who develop occlusion of the iliac vein following thrombosis get dilated superficial veins in the region of the groin.

Venous physiology

Four-fifths of the blood in the circulation is in the veins. The veins have an important role in regulating the capacity of the circulation and contraction of the main veins provides the body's initial response to blood loss. Paralysis of the veins resulting in venous pooling can lead to postural hypotension and fainting.

Blood flows through the veins as a result of arterial pressure across the capillary bed, muscular venous pumps, and the effect of gravity (Fig. 8).

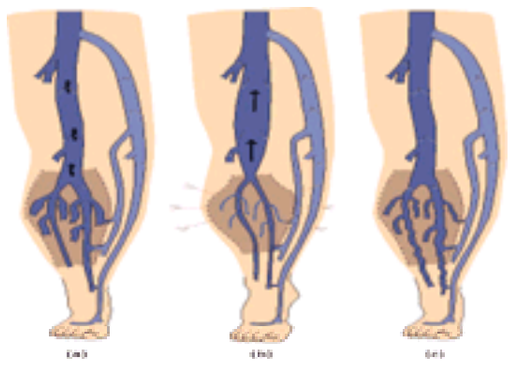


Fig. 8. No caption available.

The pressure measured in the veins [mmHg (millimetres of mercury)] will depend on the position of the patient. With them lying and the foot slightly elevated, the pressure in the veins will be zero. When standing, the pressure measured in the veins is equivalent to the weight of a column of blood extending from the foot to the heart.

The pressure measured in the foot with a patient standing still may be equivalent to 100 mmHg. If the patient then exercises, the pressure will fall to under 20 mmHg (Fig. 9). This fall is caused by the venous muscle pumps and depends on normally functioning veins, active muscles, and normal joints. Damage to the system will result in no fall in venous pressure on exercise.

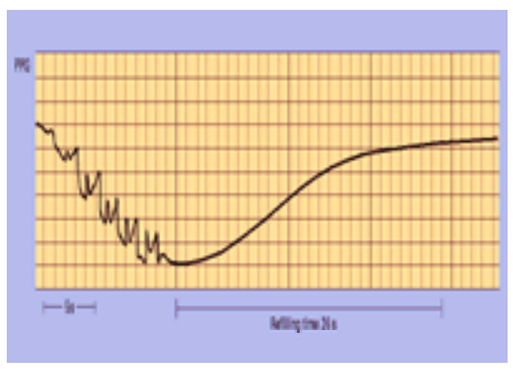


Fig. 9. No caption available.

Venous occlusion following thrombosis of the major axial vein (femoral iliac vein) may result in an increase in venous pressure on exercise, due to venous outflow obstruction.

Clinical symptoms and signs of venous disease

Patients presenting with venous disease may have some or all of the expected signs and symptoms. Severe signs may be accompanied by relatively few symptoms and, conversely, severe symptoms may be accompanied by relatively few signs. The classical symptoms of venous disease include aching, heaviness, tingling, numbness, restlessness, and bursting pain.

The signs of venous disease may change with position. It is always better to examine a patient with venous disease standing, and then only after they have been standing for a period of 2 min. Firstly, inspection of the leg may reveal colour changes, eczematous change, and ulceration. Dilatation of the veins may be apparent either in the distribution of the long saphenous vein or the short saphenous vein, or sometimes both; tortuosity of the veins with saccular varicosities may be seen; or the presence of smaller reticular veins and even smaller telangectasia or dermal flares may be observed.

Palpation of the veins may reveal areas of hardness with superficial hrombophlebitis, the presence of a fluid thrill and a cough impulse, and, in the presence of an arteriovenous malformation, a palpable thrill.

Alterations in temperature when compared between the two legs, between the top of the leg and the foot, and over localized areas of inflammation may also be apparent.

The causes of oedema in the legs are listed in Table 1 and the factors influencing the movement of tissue fluid in relation to oedema are outlined in Fig. 10.



Table 1 Causes of oedema in lower limbs (see also Fig. 10)

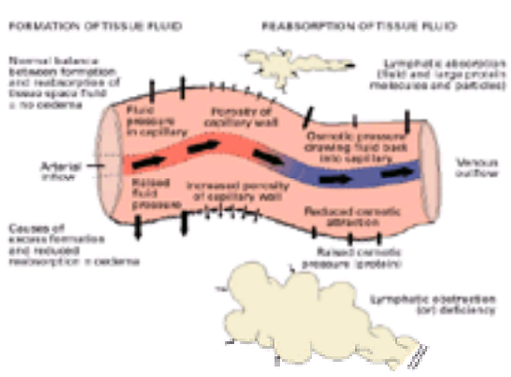


Fig. 10. No caption available.

Clinical examination and special investigations in patients with venous disease

Clinical examination

A clinical examination with a proper history still forms an important part of the management of patients with venous disorders.

History

The history of the presenting complaint and the duration of symptoms should be listed. It is important to inquire about family history and about any previous history of venous disorders, in particular of venous thrombosis. Venous thrombosis may accompany major surgical procedures but remain symptomless. Any patient who has undergone major surgery or who has sustained injury to the lower limbs may have had a silent, deep vein thrombosis. In a young patient with a family history of venous thrombosis, the possibility of an underlying thrombophilia should be considered.

Examination

Ask the patient to stand and be prepared to wait for up to 2 min for the veins to fill. Examine the patient for signs of colour change, deformity, swelling, and scars. Assess whether the veins fall within the distribution of the long or short saphenous system. Palpate the leg, feeling for differences in temperature gradient and areas of localized heat. Determine whether there is any oedema and whether this is chronic with induration. Palpate the veins to determine whether there are any thrills, particularly on coughing or on tapping. Palpation along the length of the vein will reveal apparent defects, but these seldom correspond to perforator sites.

Clinical tests

The features so far described are the observable changes that may be present in venous disorders. These give indirect evidence of abnormal function in veins but few are diagnostic of a specific state. Of the clinical tests relying solely on the examiner's senses, only two are capable of giving a specific diagnosis. In both the following tests the use of an encircling rubber band to occlude superficial veins is avoided, partly because it may cause artefacts by constricting the deep veins and also because it does not localize the superficial veins at fault.

The Trendelenburg (selective occlusion) test

This test is based upon demonstration that temporary selective occlusion of one or more superficial veins by localized finger pressure will delay the filling of varicose veins when the patient first stands after the veins have been emptied by elevation of the limb (Fig. 11). It shows that the varicose veins fill by downflow in the vein selected for occlusion and, therefore, its valves must be incompetent. This observation is of great diagnostic value and accurately identifies simple (primary) varicose veins. When positive it is virtual proof of this, but no great reliance can be placed on a negative result in the exclusion of incompetence in a superficial vein. Although this test is still widely used in clinical practice, we rely on non-invasive venous testing in routine clinical management.

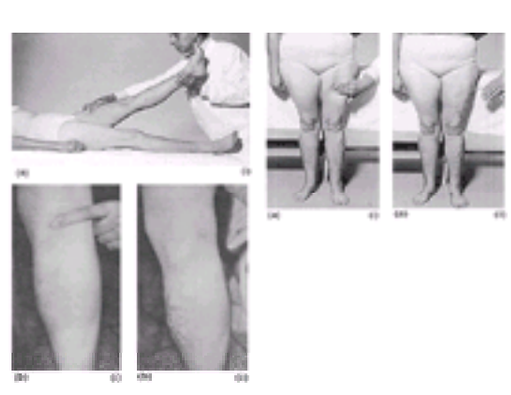


Fig. 11. No caption available.

Perthes' test

This test depends upon the change in the distended varicose veins of an upright patient. With a superficial vein occluded, if the patient exercises, contracting the calf

muscles, the superficial veins should empty if the normal mechanism of the deep veins is functioning (Fig. 12).

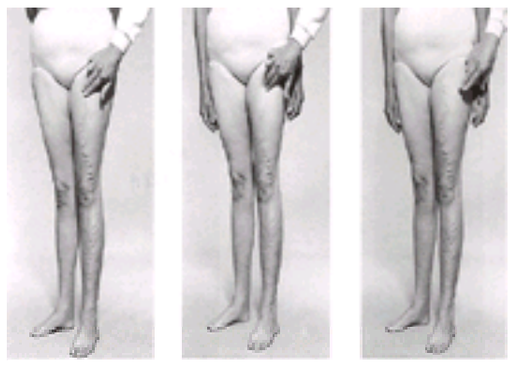


Fig. 12. No caption available.

Venous investigations

Non-invasive tests

Doppler ultrasonography

A Doppler ultrasound probe will detect flow; a bidirectional Doppler probe will detect both forward and backward flow. If the probe is placed over a vein, the flow is normally too slow to be detected. By compressing the calves, augmented flow through the veins can be detected. Using this technique the patency of a vein can be determined and, by listening to the presence of reverse flow, the presence of venous reflux detected. A simple hand-held device can be placed over the veins, looking specifically for patency, augmented venous flow, and reflux. If the veins in the lower limb are systematically examined from the groin, throughout the thigh and behind the knee, venous patency can be established and the presence or absence of venous reflux noted.

Plethysmography (Fig. 13, Fig. 14, Fig. 15)

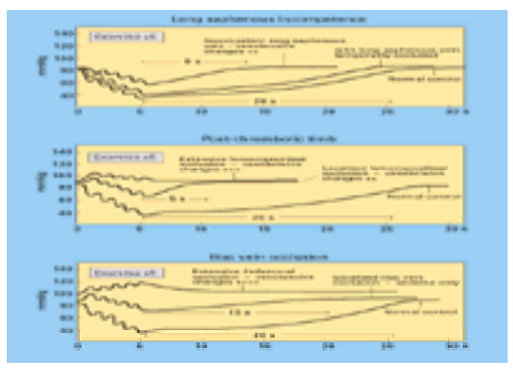


Fig. 13. No caption available.

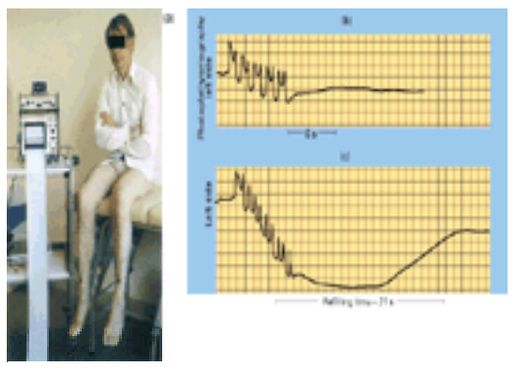


Fig. 14. No caption available.

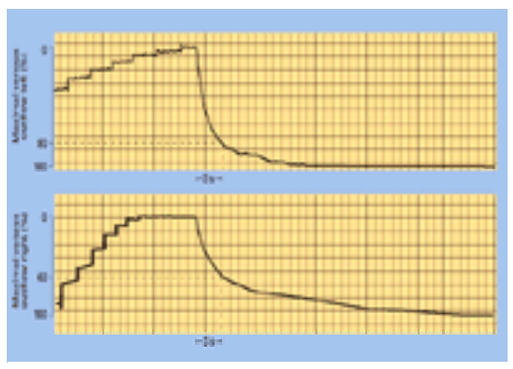


Fig. 15. No caption available.

Photoplethysmography A light-emitting diode transmits and then receives infrared light. The amount of light reflected bears a direct relation to the amount of blood in the superficial tissue. In a normal limb with normal venous function, dorsiflexion of the foot and activation of the calf muscle pump will empty blood from the superficial tissues, producing a change in the amount of light reflected. When dorsiflexion stops, refilling occurs slowly from arterial inflow. In the presence of superficial venous insufficiency, blood returns rapidly to the limb, with short refilling times. By applying tourniquets at the groin, above the knee or below the knee, the level of the superficial venous insufficiency can be determined.

Strain-gauge plethysmography By measuring the conductance of a mercury-filled tube applied to the limb, changes in electrical activity reflecting changes in limb volume can be determined. When a tourniquet is applied to the upper thigh and inflated to 80 mmHg to block venous outflow, the limb will swell. The swelling will result in a change in electrical activity, with a progressive rise in the baseline until it reaches a plateau. Once the plateau has been achieved, if the tourniquet is released there would be a rapid outflow of blood with the electrical activity returning to normal. In the presence of venous outflow obstruction, for example a venous thrombosis,

the venous outflow may be reduced, giving a marked delay in the return to normal conductance.

Air plethysmography Air plethysmography involves the application of an air-filled chamber to the lower limb. Changes in venous volume, venous outflow, and venous refilling times can be measured. Air plethysmography has the advantage that the changes can be quantified following calibration and measured in cc/100 ml.

Duplex ultrasound imaging (Fig. 16)

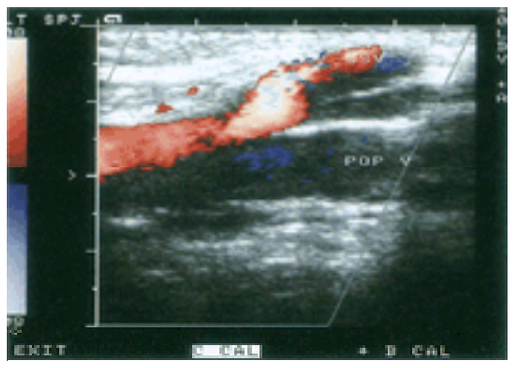


Fig. 16. No caption available.

This technique is now the method of choice for investigating patients with venous disease. By using B-mode ultrasound imaging the veins and arteries can be mapped in the lower limbs. The application of Doppler to B-mode imaging (Duplex) allows the flow of blood to be detected within the tissues. This flow can now be coloured red or blue, depending on whether the flow is in a forward or reverse direction. Duplex colour-flow imaging will detect the patency of the veins, and the presence/absence of valves; it will also detect evidence of previous thrombosis and scarring in the deep veins. The technique has many advantages over simple Doppler procedures, as the veins being insonated can be identified: reflux demonstrated on Doppler behind the knee may be in the short saphenous vein or in the popliteal; with Duplex ultrasound imaging the site of reflux can be accurately localized.

Venography

The role of venography has been reduced by the application of non-invasive venous tests. Venography involves an injection of non-ionic contrast material into a dorsal foot vein. The application of tourniquets around the ankle allows identification of the deep veins. Failure to opacify may indicate thrombosis. The present of perforating veins and reflux can be demonstrated with venography, but this involves tipping the patient from a horizontal to a standing position. Functional phlebology is still carried out, particularly in the assessment of patients undergoing deep venous reconstruction.

Clinical patterns of venous disorder

Primary varicose veins

Primary varicose veins are the most common cause of venous disease (see, for example, Fig. 17). By definition they are not associated with any underlying problems of deep veins. They may be associated with a familial tendency and may represent a primary failure of the vein wall, resulting in dilatation and valvular incompetence. Primary varicose veins may be isolated segments or they may be related to truncal incompetence in the distribution of either the long or the short saphenous systems. In addition to truncal incompetence there may be junctional incompetence, with incompetence of the saphenofemoral or saphenopopliteal junctions (Fig. 18, Fig. 19 and Fig. 20).

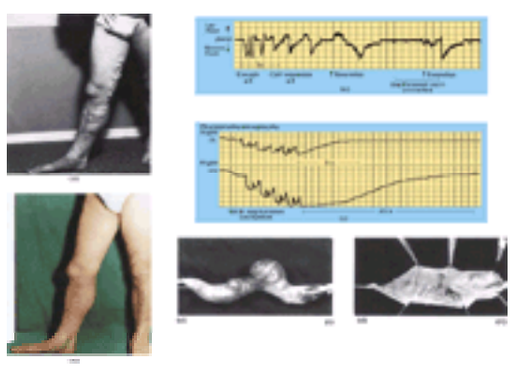


Fig. 17. No caption available.

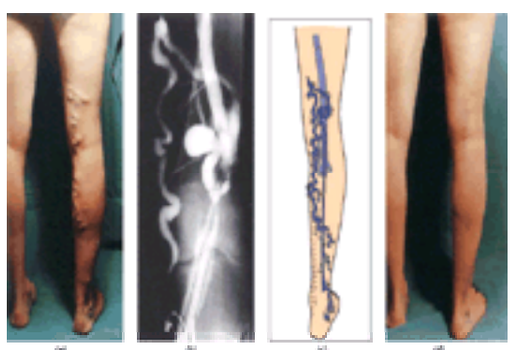


Fig. 18. No caption available.

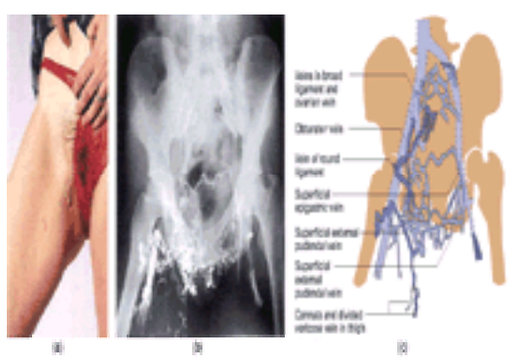


Fig. 19. No caption available.

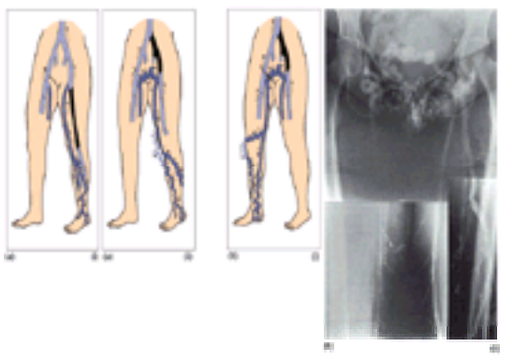


Fig. 20. No caption available.

Presentation

Patients with varicose veins may present with a cosmetic disability, a functional disability or, more commonly, a combination of both. Varicose veins occur in both men and women, with women more prone to incompetence of superficial veins than men, in the ratio of 2:1. In women, veins often first appear during pregnancy, with a strong hormonal influence. Women often complain that the veins become more prominent just before menstruation. Women are four times more likely to present with varicose veins than men because of the troublesome cosmetic affects. Increasing age is also an important determining factor.

Management

Following a full history, a clinical examination, and appropriate non-invasive venous tests, treatment can be undertaken. Treatment for varicose veins involves surgery, sclerotherapy, and new forms of laser and light treatments. Patients with junctional incompetence, significant truncal reflux, and large varicosities are better treated surgically. Patients with isolated segments of varicosities can be treated by compression sclerotherapy. Patients with reticular varices and dermal flares can be treated by microsclerotherapy; with the laser and high-intensity light treatment being reserved for special cases.

Compression sclerotherapy

The three essentials in compression sclerotherapy (Fig. 21) are an effective sclerosant, injection into an empty vein, and maintained compression with exercise. The aim of effective sclerotherapy is to produce an endothelial disobliteration and not a thrombophlebitis. If the end result is a thrombophlebitis, the varicosities will recur as the thrombophlebitis resolves. With an endothelial disobliteration the vein is destroyed forever. The strength of the sclerosant required will depend on the size of the vein. Three per cent sodium tetradecyl sulphate is used for injecting larger veins, but this should be diluted to 0.5 per cent when injecting reticular varices.

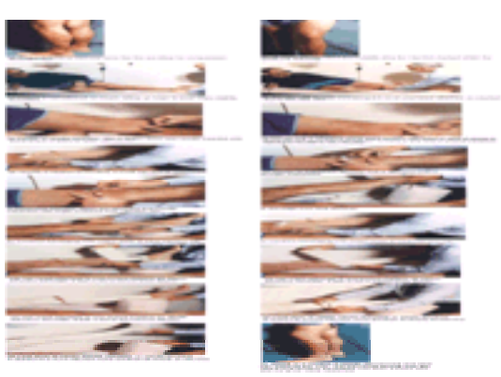


Fig. 21. No caption available.

The technique involves positioning the patient sitting on a couch. A needle with the syringe containing the sclerosant is inserted into a vein. The limb is then elevated to empty the vein; care must be taken to ensure that the needle does not get displaced. If there is any doubt about the position of the needle, no injection should be carried out, the needle should be removed, and the procedure started again. The volume of sclerosant injected will range from 0.5 to 1 cc, depending on the size of the vein. Having injected the sclerosant into an empty vein, pressure pads are applied, followed by compression bandaging. If the process starts above the ankle, once compression has been applied the limb can then be placed dependent, the next injection site identified, and the procedure repeated. Multiple injections can be given on any one occasion. Exercise is important after injection sclerotherapy and patients should be encouraged to walk vigorously for several hours a day. The duration of compression remains controversial: in the original description of this technique, compression was applied for 6 weeks; most people have reduced the compression to 3 weeks, some to 1 week, and more recently to a matter of days. With an effective sclerosant, endothelial disobliteration occurs rapidly and there is little evidence that prolonged compression assists this. There are no good comparative studies looking at the effectiveness of sclerotherapy in relation to the duration of bandaging.

A modification of this technique involves the use of Duplex ultrasound imaging (echosclerotherapy). In some countries with a strong phlebological representation amongst medical practitioners, the indications for sclerotherapy have been extended to include patients with truncal and junctional incompetence. By using ultrasound guidance and catheter techniques for injecting, junctional and truncal varices may be obliterated.

One advantage of sclerotherapy is that the technique can be repeated frequently without the need for any form of anaesthesia.

Warnings/complications Sclerotherapy is very safe, but patients should be warned to seek medical advice if they experience pain after the procedure. The limb may swell and the bandages may become too tight. An inadvertent extravascular injection can lead to ulceration. Extravasation down the needle track will not result in ulceration. Redness may occur at an injection site, and when this happens the patient should be given antibiotics. Infection is more likely to cause a small ulcer than extravasation.

An inadvertent, large extravascular injection close to the skin will inevitably lead to a large ulcer. These complications should be avoidable with a proper technique.

Surgical treatment (Fig. 22, Fig. 23)

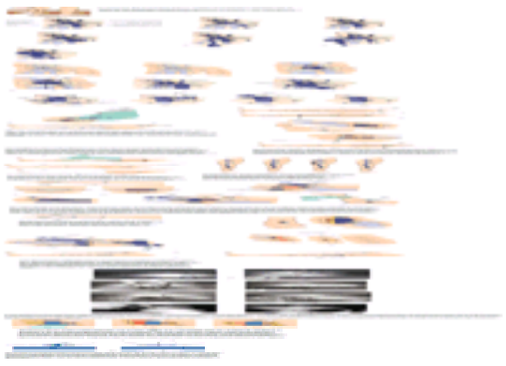


Fig. 22. No caption available.

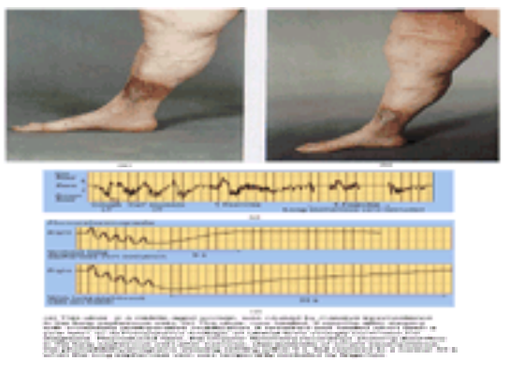


Fig. 23. No caption available.

The principle of surgical treatment is to separate the superficial from the deep system of veins at sites of junctional incompetence, to remove truncal varices, to interrupt communications between the truncal varices and the deep system via perforating veins, and finally to remove any residual varicosities through ministab incisions (avulsions). The key to successful surgical treatment is a proper preoperative assessment. The best opportunity to treat varicose veins is the first operation, as surgery for recurrent varicose veins is difficult and associated with an increased incidence of complications.

P>***Ligation of the saphenofemoral junction*** Through a small, 2-cm incision the long saphenous, superficial femoral, and common femoral vein are identified. All the tributaries arising from the saphenofemoral junction are ligated and divided, the saphenofemoral junction being ligated flush with the common femoral vein. Having ligated the saphenofemoral junction the long saphenous vein should be removed to just below the knee. If the long saphenous vein is left in place, small communicating/perforating veins will result in persistent reflux and a recurrence rate of greater than 50 per cent. Removing the long saphenous vein to the ankle is associated with an unacceptably high instance of saphenous nerve injury. By removing the long saphenous vein to just below the knee, the saphenous nerve is protected and the recurrence rate reduced significantly.

Several stripping techniques have been described involving the use of the pin stripper or the flexible stripper. By inverting the long saphenous vein during the stripping procedure we may reduce the size of the haematoma.

Ligation of the saphenopopliteal junction Before ligating the saphenopopliteal junction, accurate preoperative ultrasonographic marking should be done. A transverse incision 1 cm below the junction should be made. The vein is deep to the fascia. The deep fascia can then be split vertically and retracted laterally. The short saphenous vein is easily identifiable with preoperative marking and followed to the junction. Branches including the Giacomini vein and gastrocnemius veins may be seen. Once the short saphenous vein has been identified and divided, the short saphenous vein can be removed using an inverting pin-stripping technique to the mid-calf. Stripping to the mid-calf avoids significant injury to the sural nerve. Patients should be warned that after this procedure small areas of numbness and altered sensation may occur over the lower leg. These changes usually resolve, but may be permanent.

If a vertical incision of the deep fascia is made this reduces the incidence of hernias developing behind the knee and facilitates simple closure. The fascia must be closed before closing the skin to avoid an unsightly cosmetic bulge behind the knee.

Ministab avulsions By using a 1.5-mm incision positioned in the direction of Langer's lines and fine hooks (Boersch or Muller), dilated varicose veins can be brought to the surface and then avulsed.

It is not necessary to close these incision and they heal without leaving scars.

Warnings/complications

Before surgery for varicose veins, patients should be warned that there is a risk of small areas of altered sensation and numbness. These areas are usually associated with the ministab incisions and are often transient. Larger areas of numbness and altered sensation occur when haematomas surround major nerves or the nerves themselves have been damaged during the procedure. Large areas of numbness can be distressing, with most patients complaining when the nerves start to recover. Patients should be warned that the bruising will be extensive, but this usually resolves within 2 to 6 weeks.

Recurrent varicose veins (Fig. 24, Fig. 25)



Fig. 24. No caption available.

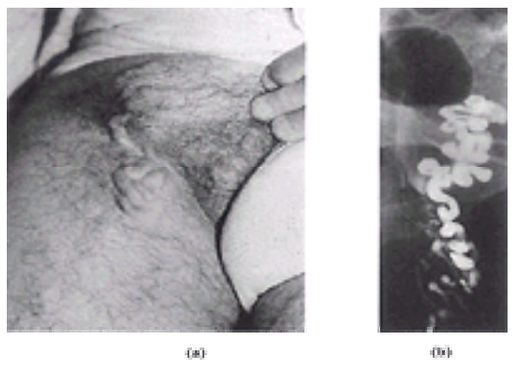


Fig. 25. No caption available.

Recurrent varicose veins occur commonly and often with clearly defined reasons. The reasons why patients present with recurrent varicose veins may be:

- failure at the first operation to deal with junctional incompetence and truncal veins
- failing to appreciate the presence of junctional incompetence, particularly the saphenopopliteal junction
- pudendal vein incompetence filling residual thigh varices
- failure to appreciate the significance of underlying deep vein problems.

All patients presenting with recurrent varicose veins should undergo a formal, non-invasive venous assessment using Duplex ultrasound imaging. In those patients with superficial venous insufficiency alone, further surgery or surgery combined with sclerotherapy may be appropriate. Where there is a significant deep venous component, surgery to the superficial veins or to perforating veins will depend on the relative contribution of the deep venous problem.

Accurate localization, particularly of the saphenopopliteal junction, is essential and can be done with ultrasound imaging. Re-exploration of the saphenofemoral and saphenopopliteal junctions can be difficult. If a patient has undergone surgery in that region before, a lateral approach to the common femoral vein, identifying the artery, is extremely useful. When re-exploring the saphenopopliteal junction, in most cases the surgeon will have failed originally to identify the short saphenous vein and religation of the junction is relatively straightforward. Incompetence in the gastrocnemius veins can contribute to the superficial venous insufficiency and these veins can be ligated in the same procedure.

Having successfully completed the junctional ligation, residual truncal veins can be removed, followed by multiple ministab avulsions. Sclerotherapy is then effective in dealing with any residual varicosities.

Patients should be warned that after surgery for recurrent varicose veins there is a risk that they will develop further veins and therefore long-term follow-up is probably advisable.

Complications of varicose veins

Haemorrhage

Varicose veins lying directly beneath the skin commonly become adherent to it and may so stretch it that it no longer provides adequate protective covering. Then, only the thinnest layer of skin and fragile, attenuated vein wall retain the blood, which shows inky blue through the membranous covering. This state is most likely to develop in varicose veins on the foot, ankle, and lower leg, particularly in elderly patients. The vein may burst with minor trauma or spontaneously when the patient is up and about. The ensuing haemorrhage can be copious but is easily stopped by finger pressure, or when the patient lies down with the foot elevated and a firm pad and bandage is applied. This can be an alarming experience for the patient and, although the aperture will be temporarily plugged with clot, it will soon bleed again (Fig. 26(a)). There is little natural tendency for it to heal because the underlying vein remains open and the unsupported; devitalized skin lacks the vascular base necessary for repair processes. Treatment must be completed by elimination of the affected varicosity and in an elderly person often this can be most simply achieved by compression sclerotherapy. However, the problem is usually more than a local one and whenever possible it is better to treat the accompanying incompetence of the superficial vein surgically and at the same time to excise the point of haemorrhage and its underlying varicose vein.

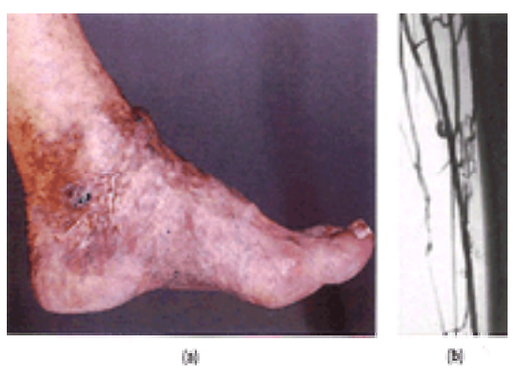


Fig. 26. No caption available.

Veins liable to this complication can often be recognized and treated before haemorrhage occurs. Patients with such varicosities should be given clear instructions how to control haemorrhage if it should occur before treatment.

Superficial thrombophlebitis (Fig. 26)

Thrombosis in varicose veins is rather common and rarely has any serious implications. Occasionally, spontaneous superficial thrombophlebitis may be associated with malignancy of the pancreas or bronchus, leukaemia, polycythaemia, or vascular disease (thrombo-angiitis obliterans).

Thrombosis in superficial varicose veins is often associated with tenderness, swelling, and redness. These changes are due to inflammation and only rarely is there an infective component. Superficial thrombosis usually remains localized and responds to anti-inflammatory drugs. Occasionally, superficial thrombophlebitis will rapidly extend, occasionally into the deep veins. When rapid extension occurs, ligation of the saphenofemoral junction and treatment with anticoagulants should be considered. Usually anti-inflammatory agents and compression bandaging is all that is required, with elevation to the leg and rest until the condition starts to recede. Anticoagulants are only used in extreme cases and similarly, surgical ligation of the saphenofemoral junction is reserved for extensive superficial thrombophlebitis with evidence of progression.

The condition is painful and in patients with extensive superficial thrombophlebitis, often associated with varicose veins, urgent treatment of the varicose veins and expression of the superficial thrombosis can be carried out.

Syndromes of valve deficiency and weak veins (Fig. 27)

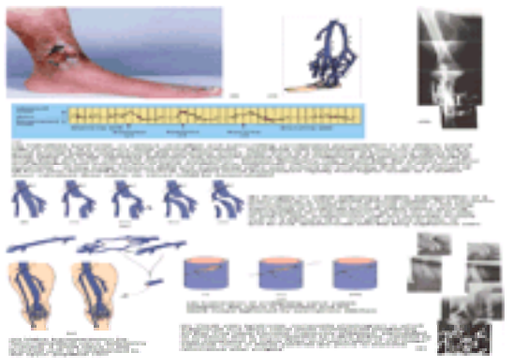


Fig. 27. No caption available.

Valveless syndrome or primary valve deficiency

The majority of patients with incompetence in the superficial veins, described in the preceding section, fall into a clearly define group that responds well to treatment. However, there is a group, about 8 per cent of venous disorders, in which a more widespread defect is present and treatment far less satisfactory. There is no sharp demarcation between the two groups but, rather, a spectrum of defect with, at one end, well-defined incompetence of superficial veins and, at the other end, a widespread deficiency of functioning valves in both superficial and deep veins. This deficiency of valves in the deep veins must not be confused with the post-thrombotic state (described later) because the patients give no history suggesting previous deep venous thrombosis and, on phlebography, the deep veins are widely open without any evidence of post-thrombotic deformity or occlusion.

In the spectrum of valve deficiency, the ‘well-defined' states at one end have a normal complement of valves in superficial veins but these are incompetent; moving along the spectrum, valves in both superficial and deep veins become increasingly deficient in number. The absence of valves without any evidence of preceding thrombosis suggests an inborn error in the development of valves. This suggestion is supported by the not uncommon finding of incompetent superficial veins in early teenage boys and girls, and the occurrence of valveless deep and superficial veins in well-recognized states of congenital venous abnormality, such as Klippel–Trenauney syndrome. Deficiency of valves in the deep veins leads to ineffective venous pumping so that venous hypertension easily develops, particularly if the superficial veins also have gross incompetence. Patients with well-valved deep veins have a robust pumping mechanism that is not easily overwhelmed, even by substantial incompetence of superficial veins, and this may explain why many patients have large varicose veins without any evidence of venous hypertension. By contrast, other patients with similar varicosities show obvious venotensive changes and ulceration, possibly because they lie halfway along the spectrum of change, with poorly valved deep veins and a relatively weak pumping mechanism, easily overwhelmed by superficial incompetence; surgery to the superficial veins can restore a precarious balance with healing of the ulcer, but requiring some care to keep it so. At the far end, the predominant failure is in the deep veins and causes a pump insufficiency too severe to be remedied by removing incompetent superficial veins. These patients suffer from intractable ulceration, not amenable to any form of surgery to the superficial veins [Fig. 27(a)(i)]. For descriptive purposes this has been named here as the valveless syndrome but it is also known as primary valve deficiency and is, of course, one variety of chronic venous insufficiency. The main features of this state are summarized thus:

- there is no history of a previous deep venous thrombosis
- gross saphenous incompetence is often present, or incompetence may come from multiple sources; obvious varicose veins are not necessarily present
- the selective Trendelenburg test does not control enlarged superficial veins because of the incompetence in deep veins
- severe venotensive changes, often with ulceration, are present
- Doppler flowmetry to enlarged veins shows surge back and forth with little purposeful movement in either direction [Fig. 27(a)(ii)]
- photoplethysmography shows a short recovery time in keeping with an inadequate venous pump and incompetent valves; this is not improved by any manipulation of superficial veins
- on phlebography the deep veins are widely open throughout with no evidence of previous deep venous thrombosis and few, if any, functioning valves can be demonstrated in them [Fig. 27(a)(iii)].
- ultrasonography will confirm substantial reflux and lack of valves in the deep veins.

Clinical presentation of these cases is very similar to post-thrombotic syndrome and the diagnosis is usually made from phlebography or ultrasonography.

Treatment

Provided that phlebography has shown the deep veins to be widely open but valveless, it is permissible to remove enlarged and valveless superficial veins. In borderline cases, this reduction in the load on the pumping mechanism may restore it to an adequate performance. In more severe examples it brings little or no benefit and treatment will have to rely upon the conservative measures of elevation whenever possible and the use of external support by elastic stockings.

Surgical restoration of valve function

As incompetence in deep veins is the main problem, it may be possible to improve this by one of the following methods.

Transposition of a deep vein into a neighbouring vein that is well valved, for example, implanting the upper end of a superficial femoral into the side of a valved profunda femoris vein nearby (Fig. 27(b)). This may be useful in congenital states of valve deficiency including Klippel–Trenauney syndrome.

Transplantation of a suitable venous valve taken from elsewhere in the body and inserted as an autograft into an incompetent deep vein, for example, using a valve from the brachial vein, or a saphenous vein of the opposite side, to transplant into the upper popliteal vein (Taheri's operation) (Fig. 27(c)). Again, this may be of help in inborn states of valve deficiency.

Repair of prolapsing valve cusps by valvuloplasty (Kistner's operation): this procedure tightens up selected valve cusps when incompetence is caused by excessive length with one cusp prolapsingbeneath the other (Fig. 27(d)). A small number of patients with severe venous insufficiency have this form of incompetence in the femoropopliteal deep veins and may be suitable for the procedure. The operation has yet to gain general acceptance, although successful cases are reported.

Use of a tendon or Silastic sling around the popliteal vein to act as a substitute external valve (Psathakis): this procedure has yet to be evaluated independently, and there is some indication thatdeep venous thrombosis and pulmonary embolism may be aproblem.

Weak vein syndrome

This is a different state from that just described, although both may coexist. The patients often have a strong family history of varicose veins and develop unusually large, incompetent superficial veins leading down to equally large varicosities (Fig. 27(e)). With surgical treatment there is a strong tendency for further large, recurrent varicosities to appear within a year or two. Gross examples are seen following inadequate surgery, for instance when the long saphenous vein has been ligated at too low a level or has not been stripped and the source of the recurrent varicosities is from surviving branches in a saphenous stump, perforators in the thigh, or interconnections between the saphenous systems. It is possible that extrinsic factors play a part: for example, some of the most severe cases are seen in people whose work involves prolonged standing and high consumption of alcohol, as in publicans and hoteliers. Clinical examination and phlebography give a strong impression that the veins lack inherent strength and easily overexpand. This, indeed, may be the primary problem because weak vein walls will allow valve cusps to separate and become incompetent; deficiency of valves does not seem to be a cause because an adequate number may be present but grossly incompetent. Histological studies commonly find some degenerative changes in the vein walls of patients with varicose veins and in the state just described the primary defect may be an extreme form of this.

Diagnosis and treatment

If there has been no previous treatment all the usual features of simple incompetence will be found, including good control with a Trendelenburg test, although the veins are unusually large. This does not contraindicate surgery but is rather an encouragement for conscientious surgery; saphenous ligation must not leave any branches at the upper end, stripping of the long saphenous vein must be at least to upper calf, and all major varicosities must be removed.

If the patient presents with recurrent varicose veins, the massive size of these and their unusual pattern may be daunting. However, with careful mapping out and use of the Doppler flowmeter, the source and pathway of incompetence may be identified. Further surgery must be carefully planned and this will require detailed phlebography to guide the surgeon. A good long-term result may be obtained, but, in the very nature of the condition, further recurrence is possible and, as a last resort, strong elastic stockings will be required to control the varicose veins as a long-term policy.

The two states just described, the valveless syndrome and the weak vein syndrome, are ill defined and ill understood. However, it would be unrealistic in describing the venous disorders to give the impression that all categories are clear-cut and separate from each other. Certainly, most patients can be fitted into a well-defined category and rational treatment based upon this. In others, one of the ill-defined varieties just described may baffle the surgeon, or, it is possible for two separate categories of venous disorder both to be present in the same limb and give a very confusing picture. In these circumstances, clinical expertise alone is not sufficient and the extra information provided by ultrasonographic imaging and phlebography must be called upon to make sure that the best decision is made for the patient. Care must be taken to exclude the next category, the post-thrombotic syndrome.

Deep vein thrombosis and post-thrombotic (or post-phlebitic) syndrome (Fig. 28, Fig. 29, Fig. 30)

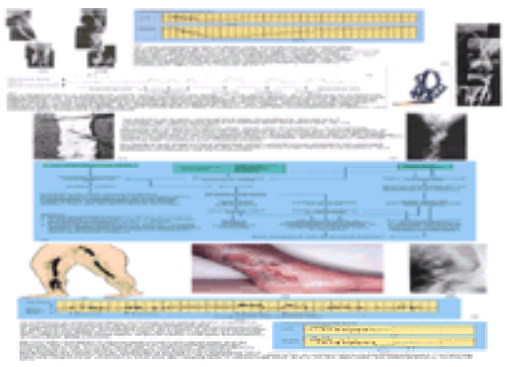


Fig. 28. No caption available.



Fig. 29. No caption available.



Fig. 30. No caption available.

Thrombosis in the deep veins of the lower limb is a common complication of serious illness, pregnancy, following surgical operations and after severe injuries, especially with fractures of the lower limb and pelvis. It may be localized to small areas or extend massively throughout both lower limbs. In many instances there are no signs or symptoms, the first manifestation of deep venous thrombosis being a pulmonary embolism. With extensive iliofemoral deep venous thrombosis the classical signs of deep vein thrombosis are more likely to occur. These signs include swelling of the leg, dilatation of the superficial veins, warmth, pain and tenderness, often associated with a low-grade pyrexia. Following a deep vein thrombosis the body's own fibrolytic system will attempt to remove the clot. Over a period of time, dissolution of the clot occurs, but there is a failure to restore valvular function and inevitable scarring and stenosis of the deep veins. Although some patients make a complete recovery from a deep venous thrombosis, others have marked changes in the deep vein giving rise to severe deep venous insufficiency.

In patients with severe deep venous changes, there is a failure of the calf muscle pump, which results in the development of post-thrombotic changes. The post-thrombotic changes result in sustained high venous pressure unrelieved by exercise. This causes venous congestion, oedema, pigmentation and induration of the superficial tissues, lipodermatosclerosis, and eventually ulceration near the ankle. The long-term consequences of deep vein thrombosis are known as the post-thrombotic or the post-phlebitic syndrome. Although this syndrome is a common cause for chronic venous insufficiency and venous ulceration it must be remembered that gross incompetence of the superficial veins alone, curable by surgery, can produce the same signs.

Investigations and treatment

Any patient with a suspected deep vein thrombosis should undergo Duplex ultrasound imaging or failing that venography. Having established the diagnosis of deep venous thrombosis the patient should be fully heparinized and started on warfarin. Warfarin should be administered to maintain an international normalized ratio of between 2 and 3 for a period of between 3 and 4 months. The duration of treatment will depend on how quickly the deep vein thrombosis resolves. The resolution can be monitored by repeating the non-invasive venous assessment.

Thrombectomy is rarely performed. The primary indication is acute venous ischaemia due to massive venous thrombosis. There is little evidence to show that venous thrombectomy produces a better long-term result than anticoagulants.

Prophylaxis (see Chapter 18.2)

The results of treating deep vein thrombosis are unpredictable. Although some patients show complete resolution, the majority will show post-thrombotic changes. The best treatment will involve preventing the formation of deep venous thrombosis in the first place. Prophylactic administration of low-dose heparin, low molecular-weight heparin or mechanical methods of preventing deep vein thrombosis to patients at risk now form part of routine clinical practice. The principal risk factors include age, a past history of venous thromboembolic disease, prolonged surgery, underlying malignancy, serious medical conditions, and specific surgical procedures. All patients

now being admitted to hospital should be assessed for the risk of developing deep venous thrombosis and appropriate prophylaxis applied. When there is concern about the development of a deep vein thrombosis, non-invasive venous testing using Duplex ultrasound imaging should be carried out and appropriate treatment instituted.

Venous ulceration (Fig. 31)



Fig. 31. No caption available.

Pathophysiology

Before concluding that an ulcer on the lower limb is venous, other causes of ulceration need to be excluded. Some ulcers may have a mixed aetiological background, including an arterial and a venous component. Although there are many causes of ulceration of the lower leg, most ulcers are venous, arterial or diabetic. Exclusion of diabetic and arterial ulcers probably indicates an underlying venous cause.

Venous ulceration is due to sustained venous pressure that does not reduce on exercise. Failure to reduce venous pressure on exercise indicates a failure of the calf muscle pump, either due to direct damage to the deep valves or due to incompetence feeding the superficial system of veins. Venous ulceration can be due to deep venous insufficiency, superficial insufficiency, or a combination of both. Superficial insufficiency responds to surgical treatment and in those patients with superficial insufficiency alone, surgical treatment of the superficial veins will result in rapid ulcer healing and no recurrence. In patients with mixed superficial and deep venous insufficiency, improvement will occur with early ulcer healing, although the prognosis remains guarded. In those patients with deep vein problems the ulcers remain difficult to heal, with no realistic prospect of surgical intervention.

As a result of maintained venous pressure and a reduced pressure gradient across the capillary bed, white-cell trapping and aggregation occurs. The white cells interact with the capillary endothelium, resulting in their activation. Activation and the release of free radicals results in local tissue destruction and ulceration. A number of markers of white-cell activation (including CD11B) are elevated in patients with chronic venous insufficiency.

It is the venous stagnation and extravasation of cells, the breakdown of red blood cells and the deposition of fibrin that lead to the classical changes of pigmentation (lipodermatosclerosis), leading to ulceration.

Management (Fig. 32)

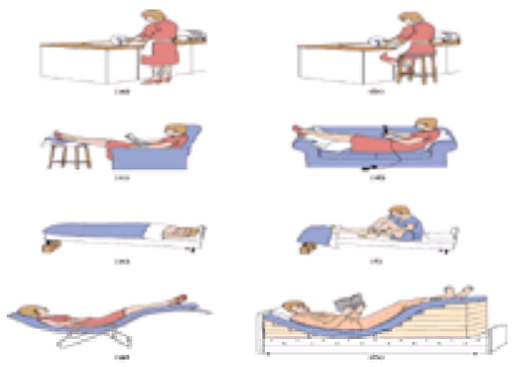


Fig. 32. No caption available.

It is important to determine the cause of the ulcer, excluding arterial and diabetic causes. Having determined that an ulcer is venous, those patients with superficial venous insufficiency should be identified and treated surgically. Those with primarily deep venous insufficiency or a combination of deep and superficial venous insufficiency are treated by surgical debridement and elastic compression. Superficial infection results in pain. Effective debridement, with antibiotics reserved for those patients with a spreading cellulitis, should be employed. Compression can be applied by bandaging the limb or by the application of compression stockings. Bandaging using four-layer techniques to achieve good levels of compression results in ulcer healing between 6 and 12 weeks. Rarely is it necessary to admit a patient to hospital to undergo surgical debridement and split-skin grafting.

Although split-skin grafting is effective in reducing the area of ulceration and on occasions achieving complete healing, the underlying pathophysiology remains unchanged and recurrence is inevitable.

Differential diagnosis: other causes of leg ulcers

The list of causes for leg ulceration given below is by no means complete but does serve the need to be aware of other causes of ulceration.

1. *Ischaemic ulcers (arterial insufficiency)*. These ulcers are common and should not be mistaken for venous ulcers. These ulcers are usually painful, situated on the toes or foot, and likely to occur over bony prominences. The ulcer is typically deep and may involve tendons. Evidence of a circulatory disturbance, such as intermittent claudication or rest pain, may be present. Arterial ulcers require very specific treatment.
2. *Diabetic neuropathic ulcers*. Diabetes is the most common cause of neuropathic ulceration, but other causes include neurological conditions such as spina bifida, tabes dorsalis, syringomyelia, spinal cord injury, and leprosy. The cause of ulceration is the loss of protective reflex resulting in local trauma to the skin. The ulcers commonly occur over bony prominences, such as the metatarsal head, the calcaneum or the malleoli. The diagnosis should be suspected in a diabetic patient. Steps should be taken to improve the circulation and control the diabetes. Local amputation may be required to control infection.
3. *Neoplastic ulcers* [see Fig. 31(l)]. A primary neoplastic ulcer of the skin, such as a malignant melanoma or epithelioma, can occur. Basal-cell carcinomas are uncommon on the lower limb. Malignant change in a long-standing venous (Marjolin) ulcer can also occur.
4. *Tropical ulcers*. Tropical ulcers include cutaneous leishmaniasis and fungus infections, and may be seen in travellers returning from temperate climates.
5. *Specific infections*. Ulceration, including that of tuberculosis and syphilis, may be caused either as part of a specific illness or as a localized infection. The possibility of a lesion due to acquired immune deficiency syndrome should not be overlooked.
6. *Blood dyscrasias*. Any severe anaemia, sickle-cell anaemia, thalassaemia, hereditary spherocytosis, or leukaemia can provide obscure forms of chronic leg ulceration. The diagnosis is made by a routine blood examination only in a patient presenting with chronic leg ulcers.
7. *Nutritional and metabolic disturbances*. Vitamin and nutritional deficiencies, uraemia, and other metabolic disorders may cause or aggravate chronic ulceration.
8. *Skin sensitivity or allergy*. Skin sensitivity or allergy to materials at work, including a wide variety of drugs, may present with skin reactions and eventual ulceration.
9. *Trauma*. Trauma as a single episode commonly sets off an ulcer in the presence of venous stasis. The patient who is currently on corticosteroids has fragile skin

and is particularly vulnerable.

10. *Necrosis by injection of chemical, insect bite, or radiation*. Misplaced injections during sclerotherapy can cause skin necrosis; inadvertent interarterial injection (Fig. 31(m)) may result in extensive skin loss. Insect bites injecting necrotoxins followed by a severe vasculitis may result in the development of an ulcer. Radiation, including the sun, may give rise to an ulcer.
11. *Rheumatoid arthritis*. Patients with rheumatoid arthritis may develop intractable ulceration on the legs or feet caused by vasculitis and this is sometimes mistaken for venous ulceration.
12. *Systemic, autoimmune, and microvascular disease*. Disorders such as systemic lupus erythematosus and polyarteritis can form lesions on the legs; these resemble the eczematous changes seen in venous disorder and eventually ulcerate.

Surgical treatment of post-thrombotic syndrome

Following venous thrombosis there may be complete resolution, leaving valvular damage scarring and stenosis of the veins, or resolution may be partial, leaving a complete obstruction to the more proximal deep veins.

Surgical treatment to replace damaged valves has not proved successful when the valvular damage is related to thrombosis. In primary valvular insufficiency, valve repair procedures have been described and shown to be effective (Kistner operation). Surgical bypass procedures designed to overcome occlusions or divert blood from the deep veins through competent superficial veins are more effective.

Palma operation (Fig. 33(a))

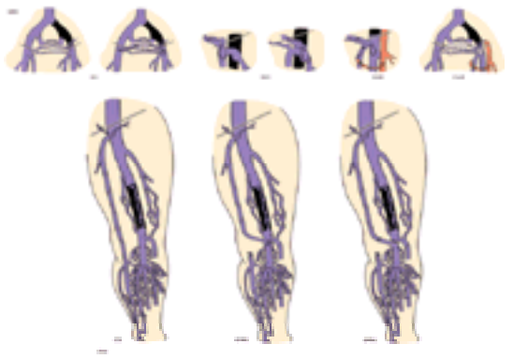


Fig. 33. No caption available.

In patients with iliofemoral thrombosis and complete occlusion, blood from the deep veins can be diverted into the deep veins of the opposite leg by routing the long saphenous vein from the opposite leg, suprapubically, and then anastomosing it into the common femoral vein on the affected side. As an alternative to the saphenous vein a polytetrafluoroethylene, externally supported, vascular graft can be used.

Popliteal femoral vein bypass (May–Husni operation)(Fig. 33(b))

By ligating the superficial femoral vein and diverting blood from the superficial femoral vein via the saphenous vein with competent valves, effects of deep venous insufficiency can be offset.

Vein transfers

The transfer of axillary vein valves to the popliteal segment in an attempt to restore competence to the popliteal vein has been tried. Whilst these procedures may work, the results are unpredictable and at best this surgery should be considered experimental.

Perforator veins (Fig. 34)



Fig. 34. No caption available.

Perforating veins connect the superficial to the deep veins. The role of these veins is to convey blood from the superficial to the deep system. In patients with superficial venous insufficiency the perforator veins may enlarge to transfer more blood from the superficial to the deep system. They only become incompetent when blood is transferred from the deep to the superficial system.

Some perforating veins have valves to prevent flow from deep to superficial; other perforating veins do not appear to have valves but remain competent due to muscular activity.

The role of a perforating vein in ulceration remains controversial. Perforating veins can become incompetent following deep venous thrombosis and perforating veins are often seen in association with ulceration. The transmission of high deep venous pressure to the subcutaneous tissues is undoubtedly an important factor in the changes that take place in the skin and subcutaneous tissues resulting in skin changes followed by ulceration.

The perforators can now be identified using Duplex ultrasound imaging. Perforating veins can be interrupted either by a direct surgical approach or indirectly with subfascial endoscopic perforator vein surgery (SEPS).

The direct approach to perforating veins using the posterior or posteromedial incision described by Dodd and Linton can rarely be justified. This procedure involves making an incision down to and including the fascia, reflecting it forwards, identifying the perforating veins, and ligating them. The problems with wound healing and extensive scarring have made this an unpopular surgical procedure with little evidence that it results in maintained ulcer healing.

More recently, an endoscopic technique using an endoscope to enter the subfascial space to identify the perforator veins and direct surgical ligation has been developed. Studies are being undertaken to assess the efficacy of this procedure in terms of ulcer healing and recurrence. The technique involves making a small incision beneath the knee, inserting an endoscope, and creating a space using gas, saline or, more recently, inflatable devices. An inflatable balloon placed in this space when inflated would create a space through which the perforating veins can be seen to traverse. These veins are then ligated under direct vision with clips, the

balloon and endoscope retrieved, and the small scar in normal skin closed.

Congenital venous disorders (Fig. 35)

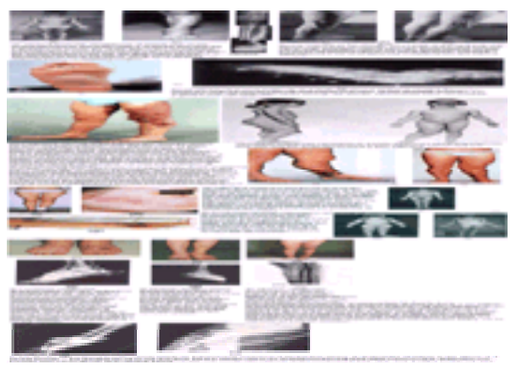


Fig. 35. No caption available.

Venous anomalies in the lower limb

The venous system is subject to many minor aberrations such as duplicated deep veins, differing levels of short saphenous termination, or the presence of unusually large interconnecting branches between long and short saphenous systems in the thigh. All these variants originate in embryo during development of the venous system and, if the valves are inadequate, may account for unusual sources of incompetence, but most are harmless and need not be detailed here. One particular abnormality involving the persistence of the lateral vein of the lower limb forming a massive, valveless channel running superficially up the outer side of the leg and terminating either in the profunda femoris or by running with the sciatic nerve into the pelvis to join the internal iliac vein (Klippel–Trenauney syndrome) (Fig. 35). This is an important abnormality because it may be associated with an abnormal deep venous system. Surgical removal of this vein may cause considerable embarrassment to venous return.

Abnormal development of capillaries and veins

Localized vascular malformations

These take the form of angiomas of capillary or cavernous varieties. The former are made up of multiple enlarged capillaries and venules to give a dark-red blemish (strawberry nevus), often present at birth and disappearing within a year or two. The cavernous angioma forms a protuberant swelling composed of large, irregular blood-filled spaces. It usually proves persistent and may show apparent enlargement in childhood. Often it affects no more than skin and subcutaneous tissues but may extend deeply, lying in and amongst muscle (Fig. 35). The boundaries of these lesions are not clearly defined but may be sufficiently limited to allow surgical removal. Surgery may require considerable skill as these angiomas cross anatomical boundaries and involve important structures that will require separating from the main mass by sharp dissection; the fringe of angioma that has to be left behind is usually unimportant. A small angioma may be treated by sclerotherapy or by laser.

Extensive venous angiomatosis

Cavernous angiomatosis may occur as a congenital defect extensively over a lower limb or elsewhere in the body (Fig. 35). It is most obvious when the skin is involved, as a dark-purple, irregular, compressible swelling but it often pervades the underlying layers extending into muscle and bone beneath. In the worst examples, this grossly abnormal venous state may be accompanied by severely defective deep veins that are either absent or lacking valves. A massive persistent lateral vein, referred to above, may be present. In such limbs, venous return against gravity is likely to be severely defective and by the time adult life is reached venous hypertension causes a strong tendency to ulceration.

The significance of bony overgrowth in a limb (limb hypertrophy)

Overlengthening of a limb may arise without any other abnormality and will then usually be seen in an orthopaedic clinic, but it is often found as a previously unnoticed accompaniment of gross vascular abnormality, as in Klippel–Trenauney syndrome, by exposed vascular abnormalities (Fig. 35). Even if obvious vascular changes are not present, the possibility of abnormality in deep veins, perhaps with large, valveless channels causing venous hypertension and eventually ulceration in the limb, must be borne in mind. The association with a large superficial lateral venous channel and defective or absent deep veins has already been mentioned. Another cause for overlengthening of bones is the presence of multiple arteriovenous fistulas (Parkes–Weber syndrome), but this condition should be regarded as separate from the venous states just described. When bony overgrowth is found in a venous clinic, it is important to place the young patient under orthopaedic care at an early stage.

Relation between congenital arteriovenous fistulas and Klippel–Trenauney syndrome

It is well recognized that extensive congenital arteriovenous fistulas will cause overgrowth of bone in the vicinity. This leads to speculation that occult arteriovenous fistulas may account for bony overgrowth in Klippel–Trenauney syndrome. This has not been demonstrated reliably and, although there are cases where both states undoubtedly coexist, the current view is that the two conditions are not directly related.

Congenital multiple arteriovenous fistulas in a limb (Parkes–Weber syndrome)

Congenital arteriovenous fistulas may be localized to, for example, part of the foot or leg (Fig. 35), or multiple fistulas may extend over the whole limb and perhaps even the adjoining pelvis (Fig. 35). There is a corresponding overgrowth of neighbouring bone, with enlargement of the bony structure of the foot if localized there, or, if the entire limb is involved, 5 cm or more increase in its length.

The communications between artery and vein are through sponge-like vascular formations, each supplied by a single, large artery that breaks into multiple branches connecting with a leash of enlarged veins running from it. An obvious pulsatile swelling may be present, with discoloration of the overlying skin and large veins radiating from it; with a stethoscope, a machinery-pumping bruit can be heard extensively over the limb. Massive arteriovenous shunting may cause an increased pulse rate that drops back to normal when the main artery to the limb is temporarily occluded (Branham's sign). The constant struggle to maintain blood pressure and the premature return of a large volume of blood may cause considerable cardiac enlargement with eventual heart failure. Brown pigmentation and ulceration of the skin in the distal limb is often present, due to the sustained venous hypertension characteristic of this condition.

Management of venous anomalies in the lower limb

The key to the management of these conditions involves a proper venous assessment. This is now carried out using Duplex ultrasound imaging supplemented by venography if appropriate.

Small lesions may be treated by laser, slightly larger lesions by sclerotherapy and, if appropriate, surgical interruption or surgical removal. In some patients, surgical is inappropriate and strong elastic compression may bring relief.

Following careful evaluation in those patients where an arteriovenous malformation is suspected, arteriography and in particular digital subtraction angiography may be carried out. Where there are significant feeding vessels, transluminal angioplasty with occlusion is effective.

---

*Please note that the Figures for this chapter are collected together at the end of its text.

Further reading

The following authoritative works give detailed accounts of the venous disorders and their treatment, together with comprehensive references to the most significant publications upon this subject.

Bergan JJ, Yao JST. *Surgery of the veins*. Grune and Stratton, New York, 1985.

Bergan JJ, Yao JST. *Venous disorders*. Saunders, Philadelphia, 1991.

Bradbury AW, Murie JA, Ruckley CV. Role of the leucocyte in the pathogenesis of vascular disease. *British Journal of Surgery* 1993; **80**: 1503–12.

Browse NL, Burnand KG, Lea Thomas M. *Diseases of the veins: pathology, diagnosis and treatment*. Arnold, London, 1988.

Dodd H, Cockett FB. *The pathology and surgery of the veins of the lower limbs*. Churchill Livingstone, Edinburgh, 1976.

Gardner AMN, Fox RH. *The return of blood to the heart: venous pumps in health and disease*. Libbey, London, 1989.

Lea Thomas M. *Phlebography of the lower limb*. Churchill Livingstone, Edinburgh, 1982.

May R. *Surgery of the veins and the pelvis*. Saunders, Philadelphia and Thieme, Stuttgart, 1979.

Negus D. *Leg ulcers*. Butterworth-Heinemann, Oxford, 1991.

Nicolaides A, Christopoulos D, Vasdekis S. Progress in the investigation of chronic venous insufficiency. *Annals of Vascular Surgery* 1989; **3**: 278–92.

Nicolaides AN, Sumner DS. *Investigation of patients with deep vein thrombosis and chronic venous insufficiency*. Med-Orion, London, 1991.

Tibbs DJ. *Varicose veins and related disorders*. Butterworth-Heinemann, Oxford, 1992.

Tibbs DJ, Scurr JH, Sabiston DC, Davies MG, Mortimer PS. *Varicose veins, venous disorders, and lymphatic problems in the lower limbs*. Oxford University Press, Oxford, 1997.

18.2 Deep vein thrombosis and pulmonary embolism

Lazar J. Greenfield and Mary C. Proctor

Introduction
Pathophysiology
Prophylaxis
Diagnosis
Treatment
Further reading

Thrombosis is the end product of two interrelated processes, activation of platelets and of the blood coagulation pathway. Thrombin is the key to both of these events. Homeostatic mechanisms keep these processes in balance but alterations in internal or external factors may result in disequilibrium. In the following, we will consider the pathophysiology, prevention, diagnosis, treatment, and future areas of investigation regarding this historical problem.

Introduction

As far back as ancient Greece, the sequelae of deep vein thrombosis have plagued humans. In the late 1800s, Virchow developed a simple model which laid the foundation for our understanding of the initiating factors of thromboembolism: stasis, endothelial damage, and disturbances in the coagulation pathway. It is generally assumed that the majority of thrombi develop in the lower extremities and that a high percentage of them arise in the calf veins. With the increasing use of central venous catheters for long-term nutritional or antibiotic therapy in the upper extremities, there has been an increase in the incidence of thrombosis originating at that level. A second source of upper extremity thrombosis is associated with compression of the subclavian vein from structural narrowing in the region of the thoracic outlet. Complications of deep vein thrombosis are common, and pulmonary embolism represents the worst outcome for thrombi originating from either the upper or lower extremity. Thrombi also may arise in the right side of the heart.

Regardless of the source, pulmonary embolism can lead to death if not adequately treated. While the diagnosis may be elusive, the physician who maintains a high level of clinical suspicion for pulmonary embolism may prevent significant morbidity and mortality in patients who otherwise would have been under- or misdiagnosed. Thromboembolism is a significant health problem not only for the individual but for society as well. The annual incidence is estimated at 250 000. With respect to the individual, a deep vein thrombosis may prolong the period of hospital stay, require 3 to 6 months of therapy, and in up to 40 per cent of patients result in lifetime problems with the post-thrombotic syndrome of chronic venous insufficiency. This condition results in symptoms of pain and chronic lower extremity edema that in some patients leads to ulceration. From a public health perspective, venous thromboembolism is a significant problem. Post-thrombotic changes affect 6 to 7 million people and ulceration is found in 400 000 to 500 000, converting them from independent to dependent members of society. A large number of individuals are affected but there is a remarkable lack of interest and investigation in this area in contrast to more highly visible and readily treated conditions.

Treatment of deep vein thrombosis during a hospital stay for an underlying condition adds to the length of stay and use of health care resources. It is estimated that the incremental cost of treatment is $4000 during a hospital stay alone. For those individuals who develop the post-thrombotic syndrome, lifetime care is required. When ulceration develops, home care can run as high as $2500 per month with an annual national cost of greater than $1 billion. The tragedy is that, in the majority of cases, this sequence and outcome could have been prevented with an adequate plan for prophylaxis.

Pathophysiology

Thrombosis develops when the balance between clotting and fibrinolysis is shifted to favor coagulation. The thrombotic process can develop through either an intrinsic or extrinsic pathway. The extrinsic pathway begins with local cell injury which leads to the release of tissue factor and exposure of the collagen cell matrix which promotes platelet aggregation. Factor VII becomes activated as well as factors IX and X. Coagulation proteins assemble on the platelet membrane surface. Adhesion of the platelets is stimulated by von Willebrand's factor while platelets become adherent to each other in the presence of fibrinogen. This results in the formation of a platelet plug. In the presence of the prothrombinase complex (factors Xa and Va, calcium, and prothrombin), thrombin is catalyzed resulting in the cleavage of fibrin peptides A and B and the activation of factor XIII, which in turn catalyzes the cross-linking of fibrin monomers. The net result is a firm clot in the presence of activated platelets and factors Va and VIIIa.

Coagulation develops along the intrinsic pathway through contact activation, when factor XI is converted to XIa, which in turn catalyzes the activation of factor IX to IXa and activates the sequence converting factor X to Xa. Acting together on platelets, factors VIII, IXa, X, and calcium catalyze the activation of factor X to Xa and merge with the prothrombinase complex.

Several anticoagulant mechanisms balance the clotting. Antithrom-bin III stops the cleavage of fibrinopeptides A and B, stops the activation of factors V and VIII, and inhibits platelet aggregation and activation as well as factors IXa, Xa, and XIa. Activated protein C inactivates factors Va and VIIIa and reduces the acceleration of the rate of thrombin formation. Heparin cofactor II regulates thrombin formation. Extrinsic pathway inhibitor is a rapid acting agent also known as tissue factor pathway inhibitor which inactivates tissue factor VIIa activation of factor X; however, it does not affect factor IX.

Plasmin is the major fibrinolytic enzyme, whose substrates are fibrin, fibrinogen, and coagulation factors which act to interrupt platelet adhesion. Plasmin degrades both circulating and clot-bound fibrin creating two fragments: E and D. It is the D fragment that is measured in the D-dimer ELISA test and which serves as a marker of fibrinolysis and thrombosis. Plasmin is the product of the interaction of plasminogen and tissue plasminogen activator. Plasminogen is activated exogenously by streptokinase or urokinase, indigenously by tissue plasminogen activator, and by the intrinsic factors. The balance within this exquisite system depends on thrombus formation, thrombus inhibition, and fibrinolysis.

From the historical perspective, Virchow described the risk of developing thromboembolism as due to some imbalance in these factors favoring coagulation. Coon conducted an extensive autopsy study at the University of Michigan in the 1950s to identify premorbid factors that resulted in an increased risk of deep vein thrombosis among patients admitted to hospital. This work was modified when the study was repeated at a later time. Among the factors found to be associated with an increased risk are heart disease, cancer, trauma, and obesity. While the risk increased with age, other factors biased this finding and age was not considered to be an independent factor. Surgery was found to have an additive effect on the risk. These findings and their additive effects are also supported by the work of others.

The importance of understanding the factors resulting in an increased incidence of deep vein thrombosis is the ability to provide cost-effective prophylactic interventions for individual patients. However, a great deal of poorly developed data exist regarding these factors. While the work cited earlier is carefully developed using sound epidemiologic principles, not all work meets this standard. Many reports have been drawn from limited populations and lack statistical power. Well validated lists of risk factors for specific patient populations are sorely needed (Table 1).

Obesity	Increasing age	Immobilization
Varicose veins	Prior deep vein thrombosis/embolism	Sepsis
Malignancy	Thrombophilias	Stasis
Stroke	Birth control pills	Pregnancy
Trauma or operation > 2h	Postpartum	

Table 1 Risk factors for thromboembolic disease documented by consensus reports

In general, risk factors can be divided into three categories: disease specific, iatrogenic, and hereditary. Patients with malignancy, sepsis, and lower extremity fractures

are examples of risk directly linked to a disease or condition. Iatrogenic factors include those interventions applied during treatment such as central venous lines, immobility, or surgical intervention. The final category includes the inherited defects of coagulation which increase the individual's risk of deep vein thrombosis. These include problems with antithrombin III, proteins C and S, factor V resistance to activated protein C, lupus anticoagulant, and the dysfibrinogenemias. Patients who have a strong family history of deep vein thrombosis or who have recurrent episodes of deep vein thrombosis at a young age are appropriate candidates for laboratory screening. These should include aPTT and platelets, ATIII antigen and antibody tests, protein C and S levels, mixing studies for lupus anticoagulant, fibrinogen level, plasminogen activity, and platelet aggregation. In some cases, for instance protein C and S, the testing must be delayed until anticoagulation has been discontinued.

Prophylaxis

Once the patient's risk of thromboembolism has been assessed, an appropriate method of prophylaxis can be selected. Beginning with the National Institute of Health Consensus Conference in 1986 and the European conferences in 1992, multiple guidelines have been suggested based on expert opinion. More recently, Clagett and coworkers developed guidelines based upon levels of evidence found in the literature. The methods and techniques of prophylaxis can be categorized as mechanical or pharmacologic. They can be used alone or in combination, based upon the needs of the individual.

The mechanical methods include antithromboembolism hosiery (TEDs), compression devices, foot pumps, and vena caval filters. The major advantage of these devices is that they are not associated with an increased risk of bleeding. Among the questions remaining to be answered are the value of combining elastic hosiery with compression devices, whether or not compression devices effect a systemic lytic response, and their functional mechanism of action—whether peak flow velocity or volume is associated with reducing the risk of deep vein thrombosis. Major questions remain concerning the efficacy of these devices with respect to pulmonary embolism. Currently, vena caval filters are being used predominantly for prophylactic indications in patients with multiple trauma or a malignancy (Fig. 1 and Fig. 2).

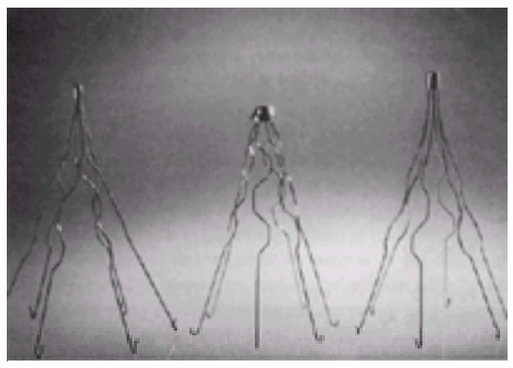


Fig. 1. Currently marketed Greenfield filters: the titanium, the original stainless steel, and the over-the-wire stainless steel for percutaneous placement.

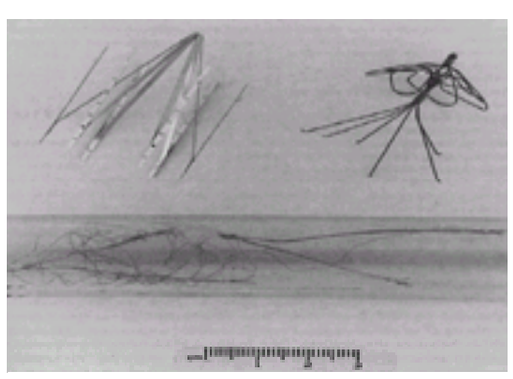


Fig. 2. The VenaTech, Simon Nitinol, and Bird's Nest filters are effective in trapping emboli but have limited long-term outcome data.

Pharmacologic prophylaxis includes both standard unfractionated and low-molecular-weight heparin, antiplatelet agents such as aspirin, and oral anticoagulation with warfarin. Unfractionated heparin given in divided doses has proved effective in general surgery and for medical indications. It is effective against both deep vein thrombosis and pulmonary embolism. The major concern is the risk of bleeding complications. In a review of the literature, Palmer and coworkers rated the probability of a major or minor bleeding complication at 0.05 and 0.19 per cent, respectively. Low-molecular-weight heparin is also effective as prophylaxis, with the greatest body of evidence in orthopedic surgery. A meta-analysis demonstrated a decreased relative risk of deep vein thrombosis and pulmonary embolism with minor bleeding favoring heparin of low molecular weight over unfractionated. There was also a reduced risk of deep vein thrombosis compared with warfarin. With respect to major bleeding, the risk with low-molecular-weight heparin is higher than with unfractionated heparin or warfarin.

Diagnosis

When prophylaxis fails, early accurate diagnosis of deep vein thrombosis and pulmonary embolism is important in limiting the sequelae. Appropriate sensitivity to the possibility of thromboembolism aids early diagnosis. Unilateral edema, pain, and tenderness in a lower extremity should initiate an objective diagnostic test; preferably duplex ultrasound which combines Doppler with B-mode ultrasound. MRI or contrast venography are complementary and appropriate under some circumstances. The traditional Homans' sign and physical assessment are not reliable since more than half of patients with ‘classic' symptoms and findings will have negative objective tests. When thrombosis is suspected on the basis of physical findings in the upper extremities, venography provides highly reliable information and potential catheter access for treatment.

Pulmonary embolism is difficult to detect on the basis of physical findings as the signs and symptoms are common to other pulmonary and cardiac conditions. Typical presentation may include chest pain, cough, dyspnea, tachypnea, and anxiety (Table 2). When the burden of thrombus is large and the patient has had major embolism, objective findings may include tachycardia, increased pulmonary second sound, cyanosis, prominent jugular veins, and varying degrees of hypotension.

Symptoms	*Prevalence (%)*	*Signs*	*Prevalence (%)*
Dyspnea	80	Tachypnea	88
Apprehension	60	Tachycardia	63
Pleural pain	60	Accentuated P2	60
Cough	50	Rales	51
Hemoptysis	27	S3 or S4	47
Syncope	22	Pleural rub	17

Table 2 Clinical manifestations of major pulmonary embolism

If the patient is hemodynamically stable, electrocardiography, chest radiography, arterial blood gases, central venous pressure, and ventilation/perfusion scanning can all prove useful in establishing the diagnosis (Table 3). When the patient is hypotensive, requiring mechanical ventilation and/or inotropic support, the first study should be a pulmonary angiogram which can be followed immediately by open or transvenous pulmonary embolectomy. The latter procedure allows extraction of recent

emboli without the risks associated with general anesthesia and a major operative procedure requiring bypass. While this procedure is not effective in those with chronic pulmonary embolism, it is highly effective with respect to extraction of thrombus (76 per cent) and improved mortality (70 per cent) in patients with major or massive pulmonary embolism. Table 4 provides assistance in determining the severity of pulmonary embolism based on objective criteria.

Test	Function
Chest radiograph	Rules out pneumothorax, adult respiratory distress syndrome, pulmonary edema, aspiration
Arterial blood gas	PO_2 ¿ 80mmHg with normal PCO_2 excludes pulmonary embolism
Electrocardiogram	Rules out myocardial infarction and pericarditis

Table 3 Preliminary diagnostic evaluation of hemodynamically stable patients with new onset of chest pain and dyspnea

Category	Hemodynamics	Signs and symptoms	Blood gas (mmHg)	Pulmonary artery occlusion (%)
Minor	Tachycardia	Anxiety, hyperventilation	PaO_2 < 80 $PaCO_2$ < 35	20–40
Major	↑ CVP, PA > 20, Responds to resuscitation	Dyspnea, collapse	PaO_2 < 65 $PaCO_2$ < 30	30–50
Massive	↑ CVP, PA > 25, Requires pressors and inotropes	Dyspnea, shock	PaO_2 < 50 $PaCO_2$ < 30	>50
Chronic	↑ CVP, PA > 40, Fixed low cardiac output	Dyspnea, syncope	PaO_2 < 70 $PaCO_2$ < 30–40	>50

The severity of a pulmonary embolism can be categorized by considering hemodynamic factors, presenting signs and symptoms, degree of pulmonary occlusion, and the arterial blood levels of oxygen and carbon dioxide.
CVP, central venous pressure; PA, mean pulmonary artery pressure in mmHg.

Table 4 Categorization of patients with pulmonary embolism

Treatment

Intravenous heparin is the treatment of choice for patients with thromboembolism. This is followed with several months of oral anticoagulation with warfarin. Heparin should be adjusted to keep the patient at a therapeutic level with the aPTT between 1.5 and 2.5 times control. It is extremely important that this be achieved within the first 24 h to prevent recurrence or extension. The total daily dose may be in the range of 30 000 to 40 000 u/24 h, and warfarin can be started concurrently. The therapeutic goal is an International Normalized Ratio between 2.0 and 3.0. Heparin's major effect on coagulation is the inhibition of thrombin-induced activation of factors V and VIII. Standard unfractionated heparin requires close laboratory monitoring as its pharmacokinetics are not linear and the anticoagulation response rises disproportionately with the dose. Evidence is accumulating that demonstrates the safety and efficacy of substituting low-molecular-weight heparin in the treatment of thromboembolism. The major advantage is the improved bioavailability and linear kinetics of this form, which eliminates the need for frequent laboratory monitoring and allows for once or twice daily dosing. This leads to significant cost savings as a large percentage of patients can be treated in the outpatient setting.

The side-effects of heparin treatment include major and minor bleeding, thrombocytopenia, osteoporosis, alopecia, and hypersensitivity. Bleeding is also the major complication of warfarin along with skin necrosis and the teratogenic effects which make it unsuitable for use during pregnancy. Patients who receive therapeutic heparin and/or warfarin have a 1.6 to 7.1 per cent risk of major bleeding and a 4.7 to 7.1 per cent incidence of thromboembolic recurrence.

Goldhaber and coworkers have recommended the use of urokinase or tissue plasminogen activator for treating major pulmonary embolism. However, there are significant contraindications to its use, especially in patients postoperatively or following trauma, and it often requires several hours to achieve a therapeutic effect (Table 5). Additionally, while there is more rapid clearing of the embolus, the effect at 1 week is comparable with standard treatment with heparin and there is no effect on mortality. This is in contrast to catheter embolectomy which can be performed in virtually all patients unless they require closed chest compression. For the deteriorating patient, the effects of embolectomy are immediately obvious with recovery of cardiac output allowing vasopressors to be discontinued.

Absolute
Active or recent internal bleeding
Hemorrhagic stroke
Intracranial neoplasm
Recent cranial surgery or trauma
Relative
Bleeding diathesis
Cardiopulmonary resuscitation
Non-hemorrhagic stroke
Surgery within 10 days
Uncontrolled, severe hypertension

Table 5 Contraindications to lytic therapy

In those patients with a major contraindication to anticoagulation (adverse events or recurrent embolism while on anticoagulants), a vena caval filter is the standard of care. There are four currently approved by the Food and Drug Administration as permanent implants. The oldest is the Greenfield filter with a cone geometry which balances the need to trap major emboli with the opposing need to maintain unimpeded flow within the IVC (Fig. 1). Initially designed for operative placement, it has been redesigned to allow percutaneous insertion from a femoral or jugular approach. Over a 25-year period, the rate of recurrent pulmonary embolism is between 2 and 4 per cent and long-term caval patency is 96 to 98 per cent. The incidence of adverse events such as caval penetration with clinical sequelae and filter fracture is less than 1 per 10 000. The Simon Nitinol, VenaTech, and Bird's Nest filters are all effective in trapping emboli, but the Nitinol and VenaTech devices have 15 to 20 per cent lower long-term patency rates (Fig. 2).

Surgical thrombectomy is reserved for those cases with threat of limb loss. Results depend upon the ability to remove adherent thrombus and maintain the patency of the vessel. At some centers, an arteriovenous fistula is created to increase blood flow and patency has been prolonged. Another technique to improve flow is to place a wire stent to correct areas of stenosis. Long-term outcome data for these two procedures are lacking.

Several areas require future study. The question of long-term outcomes for patients having thrombolytic therapy for deep vein thrombosis is important as the cost and risks of this treatment are high. Larger numbers of patients need to be studied over a 5- to 10-year period to answer the question of whether it is possible to preserve valvular function and avoid the post-thrombotic syndrome.

In addition, methods of identifying patients at risk for thrombosis need to be developed and validated. As many as 25 per cent of all patients admitted to hospital are receiving prophylaxis while 20 per cent of those who develop thromboembolism receive no protection. The role of coagulation markers is one potential method of

identifying a persistent thrombotic state.

A final area under current investigation is the appropriateness of outpatient ambulatory treatment of deep vein thrombosis with low-molecular-weight heparin, which offers the prospect of a considerable reduction in the cost of treatment. Selecting patients and establishing algorithms for treatment are pressing needs.

Venous disease has long been neglected in the field of vascular surgery but new insights into the biochemical and molecular pathophysiology have awakened interest in this important area.

Further reading

Anonymous. Prevention of venous thrombosis and pulmonary embolism. NIH Consensus Development. *Journal of the American Medical Association* 1986; **256**: 744–9. [Report of a consensus panel on assessment of risk factors for thromboembolism and appropriate prophylaxis.]

Anonymous. Prevention of venous thromboembolism: European Concensus Statement. *International Angiology* 1992; **11**: 151–9. [Updated consensus report on thromboembolism.]

Boneu B, Bes G, Pelzer H, Boccalon H. D-dimers, thrombin antithrombin III complexes and prothrombin fragments 1+2: diagnostic value in clinically suspected deep vein thrombosis. *Thrombosis and Haemostasis* 1991; **65**: 28–31. [Description of laboratory test for detection of thromboembolism.]

Cho KJ *et al.* Evaluation of a new percutaneous stainless steel Greenfield filter. *Journal of Vascular and Interventional Radiology* 1997; **8**: 181–7. [Presents outcome of clinical trial with the most recent Greenfield filter.]

Clagett GP, Anderson FA Jr, Levine MN, Salzman EW, Wheeler HB. Prevention of venous thromboembolism. *Chest* 1992; **102**(Suppl.): 391S–407S. [Recommendations are provided for thromboembolism prophylaxis.]

Clagett GP, Anderson FA Jr, Heit JA, Levine MN, Wheeler HB. Prevention of venous thromboembolism. *Chest* 1995; **108**: 312S–334S. [This is an update of the 1992 article.]

Collins R, Scrimgeour A, Yusuf S, Phil D, Peto R. Reduction in fatal pulmonary embolism and venous thrombosis by perioperative administration of subcutaneous heparin. *New England Journal of Medicine* 1988; **318**: 1162–73. [Description of the efficacy and safety of perioperative heparin prophylaxis.]

Coon W. Risk factors in pulmonary embolism. *Surgery, Gynecology and Obstetrics* 1976; **143**: 385–90. [Elaboration on the risk factors associated with thromboembolism.]

Coon W, Coller F. Clinicopathologic correlation in thromboembolism. *Gynecology and Obstetrical Investigation* 1959; **109**: 487. [Landmark autopsy study of clinical pathologic factors associated with thromboembolism.]

Coon W, Willis P, Keller, J. Venous thromboembolism and other venous disease in the Tecumseh community health study. *Circulation* 1973; **48**: 839–46. [Identification of risk factors for thromboembolism based upon epidemiologic data.]

Crochet *et al.* VenaTech-LGM filter: long-term results of a prospective study. *Radiology* 1993; **188**: 857–60. [Well documented study of 100 patients with the Vena-Tech filter demonstrating clinical adverse events.]

De Takats G. The urokinase pulmonary embolism trial. *American Journal of Surgery* 1973; **126**: 311–13. [Summary of the major clinical trial evaluating the use of urokinase in treatment of pulmonary embolism.]

Esmon CT. Cell mediated events that control blood coagulation and vascular injury. *Annual Review of Cell Biology* 1993; **9**: 1–26. [Good review of cellular events effecting blood coagulation.]

Goldhaber SZ. Contemporary pulmonary embolism thrombolysis. *Chest* 1995; **107**: 45S–51S. [Discussion of thrombolytic therapy to treat massive pulmonary embolism.]

Goldhaber SZ. Pulmonary embolism thrombolysis. Broadening the paradigm for its administration. *Circulation* 1997; **96**: 716–18. [Presentation of argument for extension of thrombolytic therapy for treating pulmonary embolism.]

Greenfield LJ *et al.* Results of a multicenter study of the modified hook titanium Greenfield filter. *Journal of Vascular Surgery* 1991; **14**: 253–7. [Report of the PDA trial demonstrating clinical function of the titanium Greenfield filter.]

Greenfield LJ, Proctor MC, Williams D, Wakefield T. Long-term experience with transvenous catheter pulmonary embolectomy. *Journal of Vascular Surgery* 1993; **18**: 450–8. [Case series of 36 patients with massive or major PE treated with transvenous catheter pulmonary embolectomy.]

Greenfield LJ, Proctor MC. Inferior vena caval filters. *Current Opinion in Cardiology* 1994; **9**: 606–11. [Review of vena caval filter devices.]

Greenfield LJ, Proctor MC. Venous interruption. In: Haimovici H, Ascer E, Hollier LH, Strandness DE, Towne JB, eds, *Vascular surgery principles and techniques*, pp. 1210–19. Blackwell Science, Cambridge, 1996. [Technical presentation of caval interruption with the Greenfield filter.]

Hirsh J. Heparin. *New England Journal of Medicine* 1991; **324**: 1565–74. [Excellent review of unfractionated heparin.]

Hull RD, Pineo GF. Low-molecular-weight heparins for the treatment of venous thromboembolism. *Annals of Medicine* 1993; **25**: 457–62. [Discussion of the use of low molecular weight heparin for treatment of thromboembolism.]

Hull, RD, Pineo GF. Low molecular weight heparin treatment of venous thromboembolism. *Progress in Cardiovascular Diseases* 1994; **37**: 71–8. [Describes differences between LMWH and standard heparin making it effective in treatment.]

Hull RD, Raskob GE, Brant RF, Pineo GF, Valentine KA. The importance of initial heparin treatment on long-term clinical outcomes of antithrombotic therapy. *Archives of Internal Medicine* 1997; **137**: 2317–21. [Justification of the use of dosing algorithms when using heparin for treatment of thromboembolism.]

Hume M. Venous ulcers, the vascular surgeon, and the medicare budget. *Journal of Vascular Surgery* 1992; **16**: 671–3. [Excellent presentation of the public health cost of treatment of venous ulcers.]

Kakkar V, Adams P. Preventive and therapeutic approach to venous thromboembolic disease and pulmonary embolism—Can death from pulmonary embolism be prevented? *Journal of the American College of Cardiology* 1986; **8**: 146–58. [A standard reference for the use of prophylaxis.]

Kelsey J, Wood P, Charnley J. Prediction of thromboembolism following total hip replacement. *Clinical Orthopaedics and Related Research* 1976; **114**: 247–58. [A description of the problem of thromboembolism in patients undergoing elective orthopaedic procedures.]

Palmer AJ, Koppenhagen K, Kirchhof B, Weber U, Bergemann R. Efficacy and safety of low molecular weight heparin, unfractionated heparin and warfarin for thrombembolism prophylaxis in orthopaedic surgery: a meta-analysis of randomised clinical trials. *Haemostasis* 1997; **27**: 75–84. [The meta-analysis evaluates efficacy and safety of various prophylactic interventions in the orthopaedic literature.]

Pineo GF, Hull RD, Rosenbloom D, Raskob GE, Gatni A. A cost-effectiveness analysis using actual costs of low-molecular-weight heparin compared with continuous intravenous heparin in the treatment of proximal-vein thrombosis. *International Angiology* 1994; **13**(Suppl. 1): 87. [An initial attempt to compare costs of standard and low molecular weight heparin.]

Rogers FB, Shackford SR, Wilson J, Ricci MA, Morris CS. Prophylactic vena cava filter insertion in severely injured trauma patients: indications and preliminary results. *Journal of Trauma* 1993; **35**: 637–42. [A justification of the prophylactic use of vena caval filters in multiply injured patients.]

Rogers FB, Shackford SR, Ricci MA, Wilson JT, Parsons S. Routine prophylactic vena cava filter insertion in severely-injured trauma patients decreases the incidence of pulmonary embolism. *Journal of the American College of Surgeons* 1995; **180**: 641–7. [Review of the author's experience with prophylactic Greenfield filter use in trauma patients.]

Sarasin FP, Eckman MH. Management and prevention of thromboembolic events in patients with cancer-related hypercoagulable states: a risky business. *Journal of General Internal Medicine* 1993; **8**: 476–86. [Discussion of the problems related to DVT prophylaxis in patients with cancer.]

Sigel B, Ipsen J, Felix W. The epidemiology of lower extremity deep venous thrombosis in surgical patients. *Annals of Surgery* 1974; **179**: 278–90. [Presentation of the risk factors for thromboembolism.]

THRIFT Consensus Group. Risk of and prophylaxis for venous thromboembolism in hospital patients. *British Medical Journal* 1992; **305**: 567–74. [Recommendations by a panel of European physicians for assessing the risk of thromboembolism and methods of prevention.]

Virchow R. Uber die Vertpfung der Lungenarterie. *Neue Notizen und Geb d Natur Heilk* 1846; **36**: 26. [Early works by Virchow highlighting the triad of factors related to thromboembolism.]

Virchow R. Die cellular Pathologie in ihrer Begrundung auf physiologische and pathologische Gewebelehre. A. Hirschwald, Berlin, 1858. [Early works by Virchow highlighting the triad of factors related to thromboembolism.]

Wakefield TW. Hemostasis. In: Greenfield LJ, Oldham K, Zelenock G, Mulholland M, eds. *Surgery*. Philadelphia: Lippincott, 1997: 39–108. [Summary of current thinking on the processes related to coagulation, and fibrinolysis with an insight into the importance of the inflammatory response.]

Wilson JT, Rogers FB, Wald SL, Shackford SR, Ricci MA. Prophylactic vena cava filter insertion in patients with traumatic spinal cord injury: Preliminary results. *Neurosurgery* 1994; **35**: 234–9. [Describes the se of prophylactic Greenfield filters in a particularly thrombogenic group of trauma patients.]

19 Abnormalities of the lymphatic system

Kevin G. Burnand and Catharine L. McGuinness

Introduction
Physiology
Lymphoedema
Secondary lymphoedema
Primary lymphoedema
Differential diagnosis
Investigations
Management
Lymphoedema of other special sites
Genital lymphoedema
Eyelids and upper limb
Liposuction
Chylous reflux
Lymphangioma circumscriptum
Cystic hygroma
Mesenteric cysts
Further reading

Introduction

The lymphatic system has two main functions. The first is to remove macromolecules and fluid from the interstitial space. Large molecules that escape into the tissue fluid have considerable difficulty in re-entering the vascular compartment. Proteins such as albumin, globulins, and fibrinogen that enter the interstitial fluid are usually returned to the plasma through the lymphatics. A number of coagulation factors and fibrinolytic activators also enter the lymph. Between 2 and 4 litres of interstitial fluid are returned to the vascular compartment each day by the lymphatics (Fig. 1 and Table 1.)

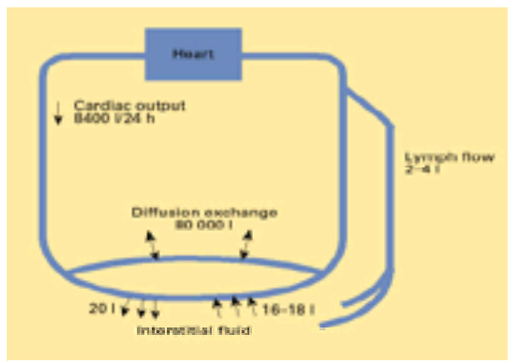


Fig. 1. The circulation passing through the interstitial fluid showing the contribution of the lymph circulation (2–4 l/day).

Content	Filtrate	Absorbate	Lymph
Fluid (l)	20	16–18	2–4
Protein (g)	80–200	0–5	75–195

Table 1 The volume and protein content of the three compartments shown in Fig. 1

The second major function of the lymphatics is to allow the circulation of lymphocytes from the lymph nodes into the bloodstream. Most exogenous antigens are presented to the central lymphoid system for the first time via the lymphatics. Recognition of antigens, with subsequent proliferation of specific clones of lymphocytes, takes place in the lymph nodes. Activated lymphocytes then pass into the circulation and thus to the other lymphatic tissues throughout the body.

The lymphatic system develops from four cystic spaces that appear on either side of the neck and in both groins. These large cisterns develop communications (lymphatic vessels or lymphangioles) which allow most of the lymph from the lower limbs and abdomen to be channelled through the cisterna chylae into the thoracic duct, which passes up on the left side of the bodies of the thoracic vertebrae before entering the internal jugular vein in the left-hand side of the neck. A separate lymphatic trunk which drains lymph from the right upper limb and right side of the head and neck enters the right internal jugular vein. Lymph nodes develop as condensations along the course of these lymphatic pathways.

Abnormalities of development result in lymphatic aplasia, cystic hygroma, lymphatic and nodal hypo- and hyperplasia, and lymphangiomas.

Physiology

The interstitial space has a negative pressure and this, in combination with the hydrostatic pressures of the capillaries, encourages fluid to escape from the vascular compartment, overcoming the oncotic pressure of the plasma proteins which acts to suck fluid back into the circulation. The intraluminal pressure of the lymph system is similar to that of the interstitial fluid, and lymph capillaries must therefore actively absorb proteins through their pores. The mechanism by which this is achieved is unknown, but appears to require energy.

The lymphatic capillaries have large pores to allow large molecules to enter the lumen, and many valves that prevent the reflux of lymph and encourage its onward passage more proximal. Lymphatics have some circular smooth muscle in their wall and are capable of contraction. The combination of inherent contractility and valves ensure that lymph is propelled along the lymphatics and into the veins. Lymphagioles are less than 0.5 mm in diameter and can only be seen with ease down an operating microscope when they are full of chyle or dye. Other factors that may encourage lymphatic drainage include compression from surrounding arteries, and the negative pressure of the thoracic cavity sucking lymph upwards into the thorax from the abdomen.

Lymphoedema

This is defined as the excessive accumulation of interstitial fluid as a result of defective lymphatic drainage. There is also impaired transport of autologous and foreign proteins. The production of cytokines by the parenchymal and immune cells may be responsible for the proliferation of fibroblasts and epithelial cells, which cause

subsequent sclerotic changes of the skin and subcutaneous tissues.

The condition may be further subdivided into primary and secondary lymphoedema, secondary lymphoedema being the most common. Primary lymphoedema is three times more common in women than in men. It has no known cause but roughly a third of patients have a family history of the condition, and studies of family trees suggest a large genetic element (Fig. 2); secondary lymphoedema is the result of some recognized pathological processes disrupting the lymphatic drainage.

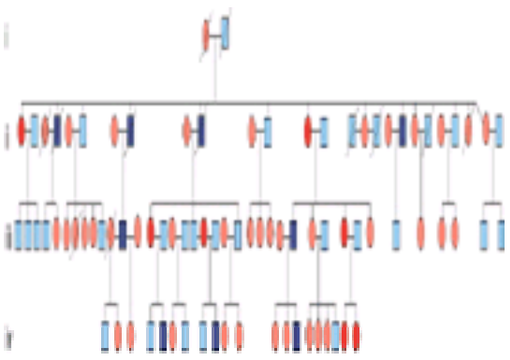


Fig. 2. Family tree that shows a number of relatives affected by lymphoedema (purple).

Further subdivisions of primary lymphoedema have been made on the basis of the anatomical lymphatic abnormalities that are present. The lymphatic channels may be absent or severely hypoplastic (Fig. 3), being few in number and petering out more proximally. They may also be excessive in number (Fig. 4), though defective in function; such lymphatic hyperplasia is usually associated with excessive numbers of lymph nodes. The lymphatics may be dilated and ectatic (megalymphatics) (Fig. 5) and this abnormality is often associated with chylous ascites, chylothorax, and lymphatic reflux. Finally the lymphatics may be obstructed (Fig. 6), and in primary lymphoedema this is often associated with fibrosis within the lymph nodes. Primary lymphoedema is associated with other congenital abnormalities including vascular deformities, gonadal dysgenesis, congenital heart disease, and cleft palate.



Fig. 3. A lymphangiogram showing severe hypoplasia of lymphatic channels in the left leg compared to normal numbers in the right leg. Proximal obstruction probably coexists in the nodes of the groin and left iliac region.

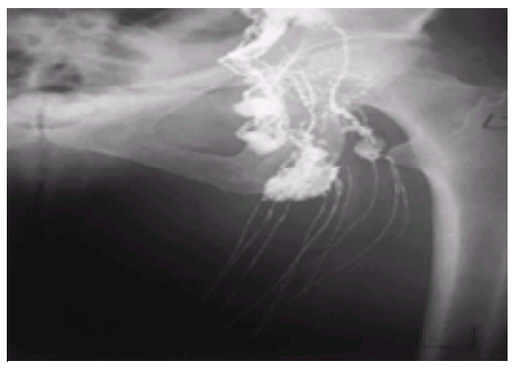


Fig. 4. A lymphangiogram demonstrating hyperplasia of the lymphatic channels with excessive numbers of lymphatics with enlarged lymph nodes in the groin.

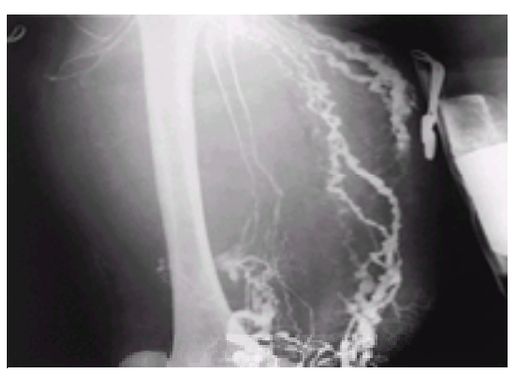


Fig. 5. Dilated ectatic lymphatics are present in the thigh without any competent valves.

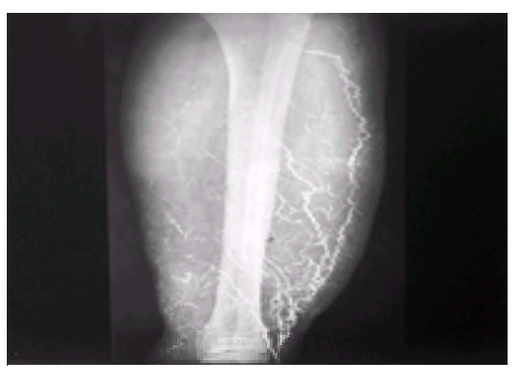


Fig. 6. Obstructed lymphatics in the calf with many extra lymphatics filling because of the proximal obstruction. The lymphatics are tortuous and fade out proximally.

Secondary lymphoedema

All patients presenting with lymphoedema must have a possible ‘cause', excluded by careful examination and special tests where necessary.

Filariasis

This helminthic infection is a major global cause of lymphoedema. The three filarial worms that cause ‘lymphatic filariasis' are *Wuchereria bancrofti, Brugia malayi*, and least importantly, *Brugia timori*. Infection is a result of a bite from an infected arthropod. The larvae mature into adult worms within the human host and the female produces microfilariae that are then transmitted to biting insects, thus completing the lifecycle. A recent World Health Organization report has shown that 3287 million people live in countries that are endemic for filariasis and around 751 million of these are in areas where the transmission is known to occur. Approximately 80 million people in 76 countries are infected with filarial parasites. *Wuchereria bancrofti* accounts for about 90 per cent of infections and *Brugia malayi* for most of the remaining cases. Globally, about two-thirds of infected people live in China, India, and Indonesia.

The worm enters the lymphatics and lodges in the lymph nodes, where a severe fibrotic reaction causes obstruction to the lymphatic pathways, which are often grossly dilated (Fig. 7). This results in severe swelling of the limbs (usually the lower) called ‘elephantiasis'.

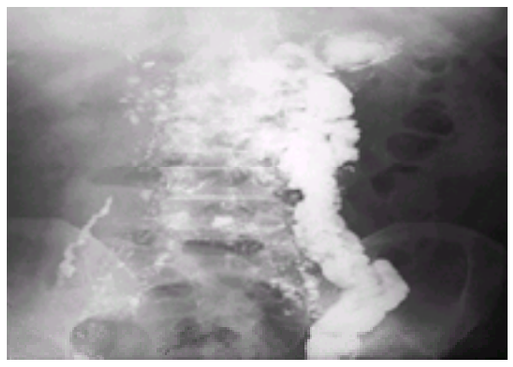


Fig. 7. A lymphangiogram showing massively dilated retroperitoneal lymphatic channels in a patient with filariasis

The diagnosis is confirmed by finding microfilariae, which enter the blood in large numbers at night. To ensure the maximum possibility of detecting filariae, a blood sample should be taken at midnight. Thick blood smears using 0.02 ml of blood stained with Giemsa should be sufficient to diagnose infections where microfilariae densities are between 250 to 20000/ml, but light infections may require concentration techniques. A strongly positive complement fixation test suggests active or past filariasis.

Treatment with diethylcarbamazine destroys the filariae but does not reverse established lymphoedema; progression of the disease may, however, be slowed or prevented. Established lymphoedema is treated by the same methods as those used to treat primary lymphoedema.

Non-filarial elephantiasis

Podoconiosis is a form of endemic non-filarial elephantiasis, which has been reported in certain parts of East Africa and Ethiopia where filariasis does not exist. It is thought that the condition is a result of an obstructive lymphopathy caused by aluminosilicate and silica absorbed from soil through the soles of the feet. The silica causes a dense fibrotic reaction in the inguinal nodes of the barefoot tribesmen.

Malignancy

Any malignant process that spreads to the lymph nodes can cause secondary lymphoedema, but this is more common after surgical resection or radiotherapy directed against nodal deposits of tumour. Hodgkin's disease and the non-Hodgkin's lymphomas can present with lymphoedema; this may also occur in patients with malignant melanomas and testicular seminomas (Fig. 8). The mass effect of large tumours may obstruct lymph flow.



Fig. 8. A lymphangiogram showing massive nodal deposits in the inguinal nodes in a patient with secondary malignant melanoma.

Surgical block dissection

This operation is invariably carried out to treat malignancies affecting lymph nodes, although in many cases it forms part of a staging or prophylactic procedure. The carcinomas commonly requiring block dissections are those of the breast and uterus. Malignant melanoma and penile and testicular tumours are also often treated by block dissection or irradiation (Fig. 9(a), Fig. 9(b)).

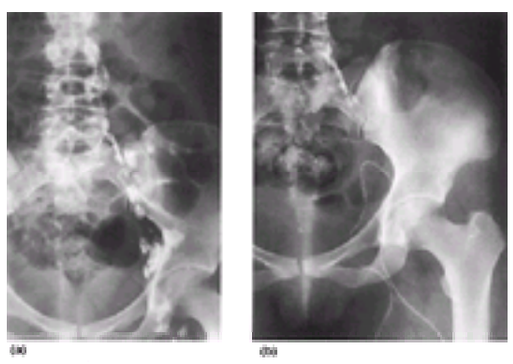


Fig. 9. (a) The lymphogram of a patient presenting with secondary deposits from a malignant melanoma within the iliac nodes. (b) The same patient after a

lymphadenectomy showing that all the lymph nodes have been removed. A redivac drain is visible.

Radiotherapy

Radiotherapy is a common cause of secondary lymphoedema of the upper limb in patients with breast carcinoma, especially when surgical block dissection has also been performed. Such combination therapy carries a higher risk of lymphoedema than either treatment in isolation. Radiotherapy results in nodal fibrosis, which can also cause obstruction of the lymphatic vessels. Recurrent tumour in an irradiated field may be responsible for lymphoedema developing some years after treatment of the primary disease.

Trauma

Severe trauma occasionally causes loss of tissue which includes lymph nodes or lymphatic channels. This is particularly seen after severe degloving injuries.

Chronic infection

Although tuberculosis has often been cited as a cause of lymphoedema, it is uncommon.

Chronic inflammation

At the St Thomas' lymphoedema clinic one or two patients are seen every year with severe rheumatoid disease (Fig. 10) or long-standing chronic eczema who develop mild lymphoedema. Chronic stimulation of the lymph nodes in these patients results in fibrosis and mild obstruction to the lymphatic drainage.



Fig. 10. A patient with lymphoedema of the upper limbs secondary to rheumatoid arthritis.

Acute infection

Severe cellulitis can occasionally damage the local subcutaneous lymphatics and cause mild lymphoedema. Patients suffering from subclinical primary lymphoedema may also develop a secondary cellulitis and the two presentations can be difficult to distinguish.

Self-induced

This quite common form of Munchausen's syndrome is produced by repeated tight application of a tourniquet. Total disuse of a limb can also cause swelling: this form of self-induced lymphoedema should be suspected if the limb cannot be moved passively (Fig. 11). Lymphograms are usually normal or only mildly unusual. The cause should be suspected if there is a sharp cut-off to the lymphoedema and a rut due to application of the tourniquet. Patients should be told of the doctor's suspicions and referred for psychiatric advice.



Fig. 11. A patient with ‘self-induced' lymphoedema caused by disease of the arm. Notice the arm is held fixidly flexed at the elbow.

Primary lymphoedema

Aetiology

Although by definition the cause of primary lymphoedema is obscure, some factors clearly influence its development. The small number of babies who have lymphoedematous limbs at birth usually have ‘aplastic' and truly absent lymphatics. This condition (Milroy's disease) has a clear genetic predisposition in some families. How and why the lymphatic system is damaged or malformed is not known, but other congenital abnormalities such as yellow nails, distichiasis, and Pierre–Robin syndrome may coexist. Some degree of inheritance can be demonstrated in about one-third of all patients with primary lymphoedema. Congenital absence of the lymphatics does not explain why lymphoedema develops relatively late in most patients, the most common age of onset being between 10 and 25 years of age.

It is possible that the constituents of the lymph draining through the lymphatics may damage the lymphangioles or lymph nodes. Lymph may contain large amounts of fibrinogen under certain circumstances, and this may coagulate and block the lymphatics; abnormal lymph may also cause nodal fibrosis which may lead to ‘die-back', or disappearance of the lymphangioles. The primary disease may therefore be in the node, and this may cause the lymphatic hypoplasia. Anticoagulant therapy has been reported to produce improvement in patients with lymphoedema, lending some support to the concept that hypercoagulability of lymph may be harmful. The benzopyrone group of drugs, which have been reported to reduce lymphoedema clinically, increase proteolysis by macrophages and the number of macrophages within lymphatics. Molecules as yet unrecognized within lymph may also be harmful to nodes and lymphatic vessels. None of these explanations accounts for the development of hyperplastic or dilated lymphatics.

Clinical features

Primary lymphoedema is much more common in the lower limbs than in the upper limbs: although this is partly explained by the influence of gravity, anatomical

abnormalities are rarely present in the lymphatics of the upper limb.

The swelling may affect one or both lower limbs, the lower abdomen, the genital region, one or both upper limbs, and, rarely, the face or chest. In the lower limbs swelling usually develops around the ankle and on the dorsum of the foot, and spreads proximally (Fig. 12). In the majority of patients the oedema does not spread above the knee, but severe oedema of the whole limb including the buttock (Fig. 13) suggests a proximal nodal lymphatic occlusion. There are exceptions to this rule, however, and patients with proximal lymphatic occlusions may have no oedema of the ankle or foot. Because lymphatic oedema is chronic and has often been present for many years, it stimulates a fibrotic reaction in the subcutaneous tissues, making them more resistant to deformation than is the case in acute oedema associated with cardiac failure or hypoproteinaemia. However, prolonged digital pressure will always produce a ‘pit'. If such pitting cannot be demonstrated, the diagnosis of lymphoedema must be questioned and another cause for the swelling should be sought.

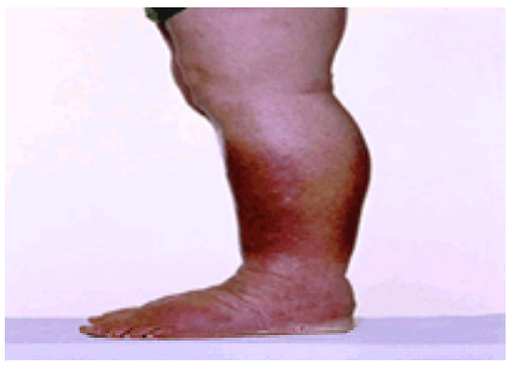


Fig. 12. Lymphoedema primarily affecting the dorsum of the foot.

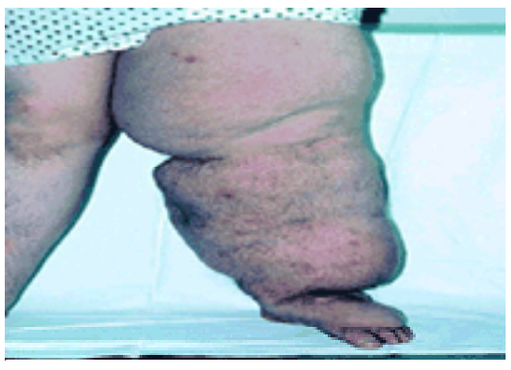


Fig. 13. Whole limb lymphoedema.

The onset of the swelling is usually insidious, and it may fluctuate. Even when lymphoedema becomes fixed, all patients report that the swelling decreases during sleep and is maximal at the end of the day. Onset can occasionally be sudden, and progression rapid. This is often associated with cellulitis, which can be both a cause and a result of lymphoedema. Patients with sudden severe swelling usually have a proximal lymphatic occlusion and may have an underlying cause for the condition. Patients with malignant obstruction of both the veins and lymphatics often develop a severe brawny oedema of relatively rapid onset that can cause intractable pain which may be very difficult to alleviate. Some patients develop marked cutaneous thickening which can progress to lymphatic warts (condylomas) and multiple coarse papillae. Other patients are troubled by repeated attacks of cellulitis. The infecting agent may enter through the hyperkeratotic skin or through cracks in the interdigital clefts which occur in athletes' foot (Fig. 14).



Fig. 14. Athletes' foot.

Occasionally patients develop numerous vesicles in the skin that may leak clear lymph or, occasionally, chyle. These vesicles usually indicate that there are megalymphatics (Fig. 15) and lymphatic reflux. Vesicles usually arise over the upper thighs or on the external genitalia, and they may also act as a portal of entry for bacteria. Occasionally the lymph leakage from these vesicles is severe enough to be a major source of embarrassment and irritation.



Fig. 15. A vesicle can be seen on the base of the scrotum. The penis and scrotum are swollen indicating genital lymphoedema.

Severe oedema of the male genitals is both embarrassing and uncomfortable, interfering with work and sexual relationships. Penetration may be impossible and urination may be difficult when the penis is almost hidden inside a grossly swollen foreskin and scrotum. Leaking vesicles and recurrent attacks of cellulitis often complicate the condition. Women with genital oedema usually have fewer problems but massive labial swelling can be uncomfortable and embarrassing (Fig. 16).

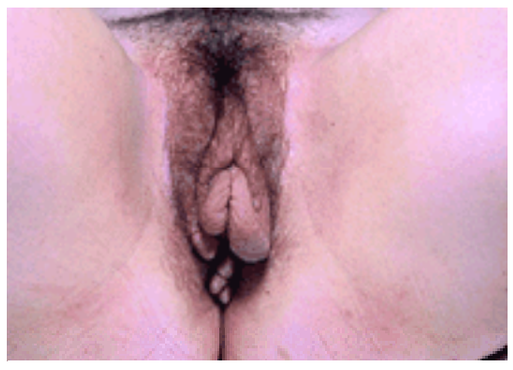


Fig. 16. Massive labial swelling due to lymphatic obstruction.

Lymph can leak into both the abdominal and pleural cavities, causing chylous ascites and pleural effusions. Patients present with abdominal distension, dyspnoea, or both. These problems are usually the result of leakage from refluxing megalymphatics; chyluria and chylous leakage from the vagina (chylometrorrhoea) are rare complications. Protein-losing enteropathy can cause severe weight loss (Fig. 17) and chylous leakage from the serosal surface of the bowel may increase the ascites.



Fig. 17. A barium meal showing thickened mucosal folds in a patient with protein-losing enteropathy.

Lymphoedema may occasionally affect the upper limb, including the fingers (Fig. 18), and unilateral pectoral swelling can also occur. Oedema of the face usually presents as swelling of the eyelids, which are the most lax tissues in this region.



Fig. 18. Lymphoedema of the arm that was the result of secondary tumour within the axillary nodes from an unknown primary presumed to be bronchial.

Patients with lymphoedema usually seek advice because they want to know the cause of the swelling, which causes major cosmetic embarrassment, even though it is often only of nuisance value. Severe swelling may make it impossible to buy shoes, and if the limb continues to swell its weight interferes with normal walking. Recurrent attacks of cellulitis may also be a major problem. Lymphangiosarcoma (Stewart Treves syndrome; Fig. 19) can infrequently develop in limbs following long-standing lymphoedema, but is more often a problem in patients with secondary lymphoedema.



Fig. 19. A patient with chronic lymphoedema of the upper limb after mastectomy who developed lymphangiosarcoma in the affected limb — Stewart Treves syndrome.

A detailed past and family history should exclude secondary causes of lymphoedema and should suggest a possible inheritance of the primary condition. Apart from confirming the presence of pitting, physical examination excludes other causes of limb swelling and reveals any associated abnormalities. ‘Square toes' often result when footwear prevents toe expansion. It is important to inspect the web spaces between the toes for athletes' foot, which is especially common in lymphoedematous limbs and is an important portal of entry for the bacteria that cause recurrent cellulitis. Yellow nail syndrome is occasionally associated with lymphoedema, as is distichiasis (two layers of eyelashes). Occasionally Pierre–Robin syndrome (micrognathia) and other skeletal abnormalities are present, and congenital cardiac anomalies are found also in patients with lymphatic hyperplasia.

The skin should be carefully examined for papillae and vesicles, and the limbs should be measured at standard levels, above and below fixed bony points, to allow both the natural history and the results of therapy to be evaluated. The length of the limbs should also be measured and the presence of any abnormal veins must be recorded. The abdomen and chest should be carefully examined for ascites or effusions. The groins and axillae should be palpated for pathologically enlarged lymph nodes. Rectal and vaginal examinations are indicated if a pelvic malignancy is suspected.

The physical examination should provide indications of whether the patient has primary or secondary lymphoedema, but unequivocal confirmation of the diagnosis is

desirable. If the swelling is not the result of lymphoedema, other investigations are necessary.

Differential diagnosis

Venous oedema can sometimes be difficult to differentiate from lymphoedema, especially if it is caused by the iliac vein compression syndrome, when there is often little in the way of superficial venous engorgement. The presence of dilated collateral veins or varicose veins normally suggests venous oedema, as does a past history of deep vein thrombosis. Other causes of bilateral limb oedema include cardiac disease, nephrotic syndrome, hypoproteinaemia, fluid overload during intravenous therapy, and chronic liver disease. These disorders can be excluded by a careful physical examination and appropriate blood tests including urea and electrolytes and protein estimation.

Klippel Trenauney and Parkes–Weber syndromes also cause limb enlargement: the former is occasionally associated with lymphoedema. True gigantism (Robertson's giant limb) also occurs, when all the tissues (muscles and bones) are hypertrophied. Abnormal fat deposition (lipoidosis or lipodystrophy) can be excluded by the fact that fat does not pit (Fig. 20). Periodic oedema experienced by women before menstruation is often difficult to separate from lymphoedema. Rapidly growing soft tissue sarcomas rarely cause diagnostic problems. Before making a diagnosis of one of these rare conditions, for which venography, CT scanning, and arteriography may be required, it is often simpler to exclude lymphoedema by isotope lymphography.

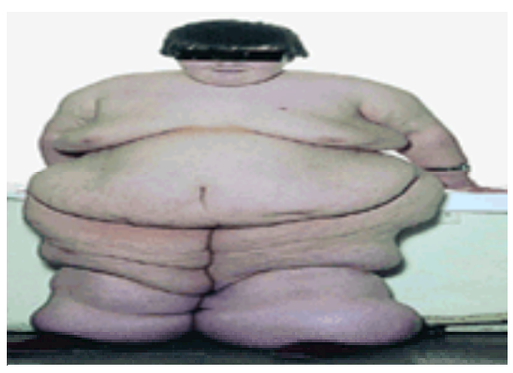


Fig. 20. A boy with Prader–Willi syndrome who has massive fat deposition (lipodystrophy), not lymphoedema.

Investigations

A full blood count, erythrocyte sedimentation rate, and chest radiographs are usually requested, but these investigations can be omitted. Measurement of serum protein, blood urea, creatinine, and electrolyte levels and liver function tests should be obtained in all patients.

Isotope lymphography

Isotope lymphography has largely replaced contrast lymphangiography as the primary diagnostic technique. Rhenium sulphur colloid is specifically taken up by lymphatics and allows the presence of lymphoedema to be confirmed by a simple outpatient investigation. Normally, 0.3 per cent of the injected dose arrives in the groin within 30 min, and more than 0.6 per cent arrives within 1 h. In patients with venous oedema there is an excessive uptake, often above 3 per cent at 30 min, and this test can therefore distinguish between venous and lymphatic oedemas. g-Camera pictures provide information that the isotope is reaching the lymph node of the groin and delayed images may show a failure of progression, indicating proximal obstruction (Fig. 21). This should be confirmed by contrast lymphography, as should any equivocal findings. An attempt was made to produce a contrast media which was selectively taken up by the lymphatics from a subcutaneous injection, but this has not yet found its way into routine clinical practice.

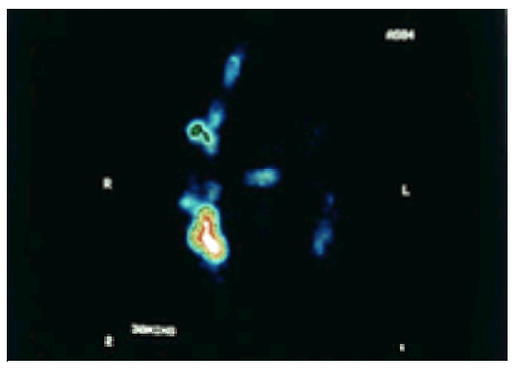


Fig. 21. A g-camera image of the groin and pelvis after injection of rhenium sulphur colloid showing uptake in the nodes of the right groin while activity is not reaching the nodes of the left groin.

Ultrasound imaging, CT scanning, and magnetic resonance scanning can all show enlarged lymph nodes, and guided biopsies of lymph nodes can be taken if malignancy is suspected. Needle or true-cut biopsies are probably safer as a preliminary procedure, since removal of large solitary fibrotic nodes may worsen pre-existing lymphoedema. A calcium chloride test may be helpful when protein-losing enteropathy is suspected

Contrast lymphangiography

This is indicated to confirm the diagnosis when the isotope lymphogram is equivocal, and to determine, if possible, the type of lymphatic abnormality that is present. This investigation originally required direct infusion of contrast into lymphatics that were visualized by subcutaneous injection of patent blue-green into the web spaces (Fig. 22); patients had to be admitted to hospital for the investigation, and general anaesthesia was usually necessary as few patients could keep still for the time required to obtain the radiographs. Such patients often require bed rest and leg elevation in hospital for several days to minimize foot oedema and make lymphatic cannulation easier. Lymphangiography does, however, provide precise information on the presence of hypoplasia, megalymphatics, and obstruction (including its site), and it still remains the ‘gold standard' against which other techniques are judged. It is still essential before certain therapeutic options are considered, and is of some value in assessing prognosis.

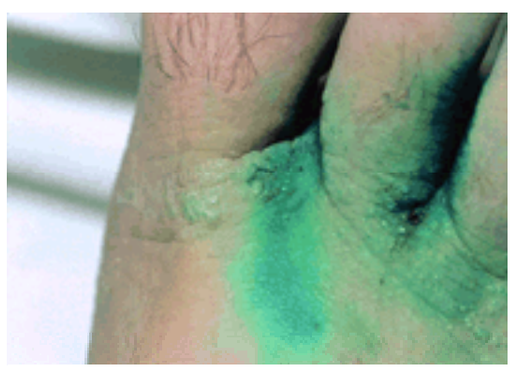


Fig. 22. Patent blue green injected into the web spaces to outline the lymphatics.

Management

Many surgeons are content to diagnose mild lymphoedema on the basis of the history and physical findings without investigating the patient further. However, this clinical diagnosis can be incorrect and isotope lymphography should be obtained if possible. This not only provides a firm diagnosis, but also allows an assessment to be made of prognosis and of possible problems that may be present in the contralateral limb.

Young women with mild lymphoedema of gradual onset usually have distal lymphatic hypoplasia: there is prolongation of the time taken for the isotope to reach the groin nodes, but normal onward passage. This type of lymphoedema is often inherited (at least 30 per cent), and rarely becomes severe or extends above the knee. It is often bilateral, but one limb may be affected several years before the other. Rarely, the upper limbs are also involved.

Physical methods

Patients with mild lymphoedema rarely require surgery. They should be given advice on limb elevation, especially in the evenings and at night, and some benefit from regular massage or mechanical compression combined with wearing of graduated elastic compression stockings. Pneumatic massaging devices are obtainable, and sequential segmental machines such as the lymphopress™ are probably more effective than single chamber boots (Fig. 23). These may be worn in the evenings or in bed at night, although they may interfere with sleep. Correctly graduated strong compression stockings (30 to 50 mmHg at the ankle, decreasing up the limb) only need to reach below knee level if the lymphoedema is distal in distribution. Elastic compression stockings do not cure lymphoedema, but they reduce fluid accumulation and often produce considerable symptomatic relief. They are poorly tolerated in warm climates and young women tend to be conscious of their appearance. Weight reduction is often beneficial and physical exercise is never harmful; concentrated compression therapy and massage may also reduce the size of lymphoedematous limbs.

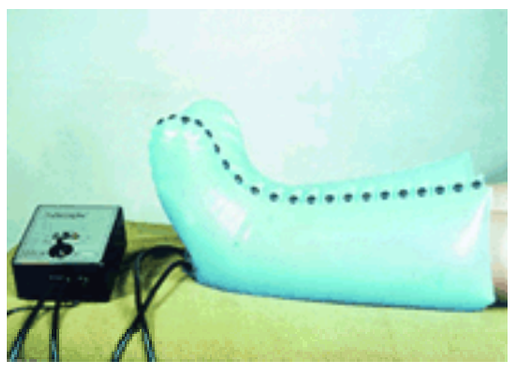


Fig. 23. Flowtron boots are less expensive but less effective than a multicompartment lymphopress.

Drug therapy

Diuretics are of little value in removing fluid from the whole body and may cause a number of problems when used unnecessarily for many years. Paroven (hydroxyrutosides) has some anecdotal support, as do the benzopyrones, but although these drugs produce statistically significant reductions in limb size these are not ‘clinically' significant, that is, they are relatively minor.

Antibiotics (flucloxacillin, amoxycillin, or one of the cephalo-sporins) should be prescribed for cellulitis and may be given in a low dose prophylactically if patients are troubled by repeated attacks. Athletes' foot must be eradicated by appropriate antifungal medication and creams. Careful attention to drying and powdering of the feet prevents infection and avoids an important portal of entry for virulent bacteria.

Surgery

Surgical reduction of severe whole-limb lymphoedema that interferes with mobility or causes severe deformity (Fig. 24) may be appropriate. In a small proportion of patients preoperative contrast lymphangiography discloses a proximal lymphatic obstruction in the ileoinguinal region with normal distal limb lymphatics (Fig. 25). These patients (1 to 3 per cent of all those seen) can be expected to benefit from some form of lymphatic bypass operation.



Fig. 24. This lymphoedematous limb is so large that it is interfering with mobility and should have its size reduced.



Fig. 25. A lymphogram showing lymphatic obstruction in the right inguinal region with contrast entering groin nodes but not passing up into the iliac chain although it does so on the left side.

Lymphatic bypass

A number of methods have been used to reunite obstructed lymphatics with the venous system. Many of these techniques, such as the skin bridge devised by Gillies and the omental pedicle, both of which were sutured to the obstructed lymph nodes in the groin, are only of historical interest. Direct anastomosis of lymph nodes to veins was originally performed by Niebulowitz, but fibrosis and low flow rates resulted in a high failure rate. Degni used a specially designed needle to insert lymphatics into the lumen of the vein, but the imprecise nature of this procedure has prevented its widespread acceptance. The advent of the operating microscope has made it possible to divide obstructed lymphatics and directly anastomose them into the side of the vein. However, the results have generally been disappointing. At least three of four lymphatics should be attached to the femoral vein in the groin in the hope that one or two anastomoses will remain patent.

Kinmonth and his associates developed the mesenteric bridge as an alternative to direct lymphovenous anastomosis. This operation uses the copious submucosal lymphatic plexus and the mesenteric lymphatics to drain the lymph from obstructed nodes in the ileoinguinal region. About 5 cm of the terminal ileum is resected on its mesenteric pedicle, as for an ileal conduit, taking great care to maintain the lymphatic drainage. The small bowel is reanastomosed behind the pedicle. The isolated segment is then opened along its antimesenteric border and the mucosa is stripped off the submucosa by a combination of sharp and blunt dissection after injection of a solution of adrenaline in saline (1:400 000). The isolated pedicle is then brought down to the first normal group of lymph nodes below the level of the obstruction, and sutured over them after they have been bivalved. Connections develop between the divided nodes and the submucosal plexus and lymph drains up the pedicle into the mesenteric lymph nodes, and eventually into the thoracic duct (Fig. 26).

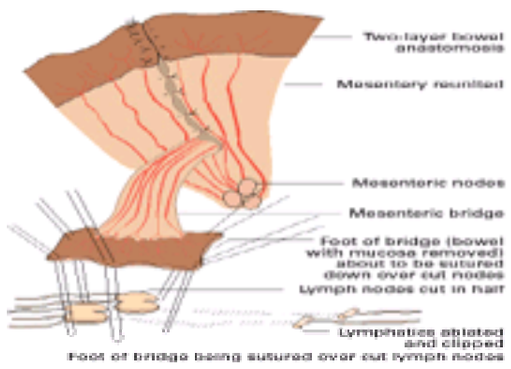


Fig. 26. A diagram to show the principles of the enteromesenteric bridge bypass operation. The pedicle of small bowel denuded of its mucosa is sutured over the bivalved lymph nodes below the level of obstruction.

This operation has been performed on over 50 patients at St Thomas' Hospital, London and has produced good results in more than half. Unfortunately there is no way of predicting those individuals who will benefit from the procedure, although the careful exclusion of unsuitable patients prevents inevitable failure. Young patients appear to fare better, and the distal limb lymphatics must still be functioning if a successful result is to be achieved. Resolution is poor if limbs are too swollen, but the swelling must be severe enough to justify major abdominal surgery. Perhaps for this reason surgery is appropriate for relatively few patients.

Reduction operations

Four types of excisional operation have been described to reduce the size of the limb. The Sistrunk operation involves excision of a large wedge or ellipse of skin and subcutaneous tissue which is then closed primarily. Homans elevated skin flaps from the subcutaneous fat, excising the underlying subcutaneous tissue before resuturing the skin flaps in place (Fig. 27). Thompson modified the Homans' operation by suturing one of the skin flaps to the deep fascia. Denudation of the superficial layers of the flap stops hair growth and prevents pilonidal sinus formation. The second flap is then sewn over the top of the denuded skin area (Fig. 28). This operation has largely been abandoned: it leaves unsightly scars, it is often complicated by pilonidal sinus formation, and the results appear to be no better than those of the simpler Homans' procedure.

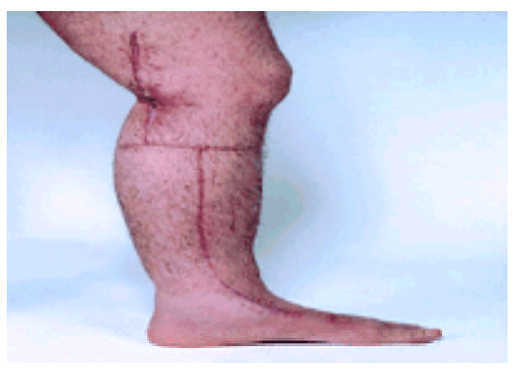


Fig. 27. A limb that has had a Homans' reduction of the lower leg and a Sistrunk excision of the thigh.

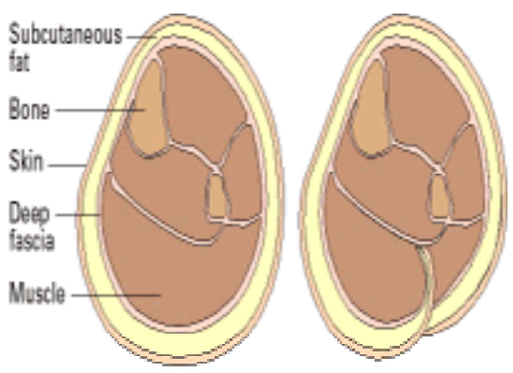


Fig. 28. The principle of the Thompson buried dermal flap operation — now largely obsolete.

Both Homans' and Thompson's operations can be complicated by skin flap necrosis and poor healing, particularly at the corners of the flaps. Great care needs to be taken to maintain the blood supply of the flaps, which must not be cut too thin. Flap reduction of the calf and foot is normally combined with a Sistrunk operation on the thigh if the whole limb is to be reduced in size (Fig. 27).

Charles invented an operation to remove the severely thickened skin in patients with filariasis. He excised all the diseased skin and the waterlogged subcutaneous tissue down to, and often including, the deep fascia from just above the ankle to just below the knee. The periosteum over the tibia was left intact and split skin grafts were then taken from normal donor skin (the opposite normal limb, or the abdomen, back, and buttocks) and used to cover the deep fascia or muscle. This operation produces the best reduction in limb size, but often at the expense of cosmesis. The ankle and knee area have to be carefully tailored to avoid a pantaloon effect (Fig. 29), and thigh reduction is also often necessary. Some patients have a poor acceptance of split skin grafts and require multiple operations to achieve complete healing. Other patients develop severe hyperkeratotic scars with warty excrescences, which produce severe deformity in the operated limb (Fig. 30). This can be treated by shaving off the warty nodules and thickened scars with a scalpel or skin graft knife; additional skin grafts are occasionally required. The final results are, however,

usually very satisfactory, especially in a grossly enlarged limb with very abnormal calf skin.



Fig. 29. A limb after a Charles reduction: the thigh has not been tapered enough, resulting in a pantaloon appearance.



Fig. 30. A foot showing a number of warty excrescences which may need to be shaved off or skin grafted.

Summary

The majority of patients can be managed conservatively. Few patients are suitable for bypass surgery: when this type of surgery is indicated an enteromesenteric bridge is probably the best form of bypass, having a spectacular effect in about half of the patients. Bypass surgery should be reserved for patients with gross limb swelling that interferes with limb function. Patients with really gross limb swelling and severe skin changes are best treated by a Charles reduction, combined with a local excision of enlarged thigh tissue. Homans' operation should be reserved for those with a moderate to severe degree of swelling. Patients with secondary lymphoedema caused by malignancy often have associated venous oedema. The results of reduction surgery under these circumstances are extremely poor.

Lymphoedema of other special sites

Genital lymphoedema

Minor scrotal and penile lymphoedema can be tolerated without specific treatment, although support garments may be helpful. Severe scrotal oedema is best treated by excisional reduction surgery in which a large central segment is excised from the scrotum, preserving the spermatic cords and testicles. The flaps are then primarily sutured using an absorbable material polydioxine sulphate and the scrotum is drained. Mobilization of the testes with gentle abrasion of their surfaces may encourage adhesions to form, allowing lymph to drain via the testicular lymphatics, aiding the scrotal reduction.

The penis may be reduced by simple excisional procedures, combined with circumcision if necessary. Alternatively, the skin and subcutaneous tissue can be stripped off the deep fascia and split skin grafts applied (a Charles operation of the penis). Both scrotal and penile reduction operations produce gratifying results for the surgeon and patient.

Massive labial swelling can also be treated by excisional procedures.

Eyelids and upper limb

Eyelid swelling can be treated by lid reduction. Arm swelling can be treated by a Homans' type of limb reduction, which can be performed on both the inner and outer sides of the upper limb. Patients with postmastectomy oedema must be assessed carefully to ensure that the venous drainage is satisfactory and to be certain that there is no evidence of recurrent axillary nodal disease. Both venous obstruction and recurrent malignancy are contraindications to arm reduction. Postoperatively an elasticated sleeve should be worn to try to prevent recurrent swelling.

Liposuction

Liposuction has been used to remove subcutaneous fat in patients with mild lymphoedema. Anecdotal successes have been achieved but the cosmetic results are variable and the procedure should be used with caution.

Chylous reflux

Some patients have dilated (almost varicose) valveless megalymphatics which allow the reflux of lymph (often chyle) against the expected direction of flow. These dilated lymphatics often end in cutaneous vesicles which are visible in the skin or which may rupture into body cavities such as the pleura (see Chapter 41.9), peritoneum, kidney (see Chapter 47.2.7), bladder, uterus, and vagina. Rupture results in the accumulation of lymph or chyle in the relevant cavity (chylothorax, hydrothorax, chylous ascites, chyluria) and chylous discharge on to the skin surface or mucosa can also occur. Accumulation of chyle in the pleural and peritoneal cavities produces severe symptoms, and patients often become dyspnoeic and very distended. Patients with megalymphatics often also have a protein-losing enteropathy which can cause weight loss and exacerbate accumulationof fluid in the body cavities and tissues. This results from leakage of lymph from the mucosal surface of the bowel; associated lymphatic leakage from the serosal surface may exacerbate the accumulation of ascites.

The diagnosis of chylous ascites or chylothorax must first be confirmed by aspiration of the fluid, which is then tested for chylomicrons. The condition may be suspected if there is pre-existing lymphoedema of the extremities and it is especially likely if vesicles and lymphatic leakage are present. In quite a few patients, however, the condition develops *de novo*. Chylothorax and chylous ascites must be distinguished from malignant ascites or a malignant effusion: cytological examination of the aspirate may help to exclude or confirm the presence of malignant cells. CT scan and ultrasound can demonstrate the presence of moderate or severe enlargement of the abdominal or mediastinal lymph nodes which suggests the possibility of a lymphoma or secondary malignant spread. Guided biopsy, laparoscopy, or laparotomy may be necessary to confirm these diagnoses. Contrast lymphography demonstrates lymphadenopathy, filling defects, or the presence of megalymphatics and is indicated if the diagnosis remains in doubt. Contrast lymphography may also demonstrate a lymphatic leak which can be surgically sealed.

Lymphoedema associated with megalymphatics rarely requires reduction surgery, but the complications of lymphatic vesicles, recurrent infections, lymphatic discharge on to the skin, chylous ascites, chyluria, and chylothorax often demand treatment. Leakage of chyle or lymph may be prevented by ligating or underrunning the dilated lymphatic channels, but this carries the risk of lymphatic obstruction which will worsen the limb swelling. Despite this many patients benefit from ligation of dilated lymphatics, and sealing off of any obvious site of fistulation.

If a patient with chylous ascites or chylothorax has no obvious leak on the lymphangiogram, chromium chloride studies and a barium study of the small bowel may provide useful information before a laparotomy is performed. At laparotomy the posterior abdominal wall over the main lymphatic pathways must be carefully inspected for the presence of lymphatic leakage, and the whole of the intestine should be examined. If the surface of the small bowel is grossly abnormal and leaking lymph, the involved or most abnormal segment should be resected. Consideration must be given to shunting the ascites back into the venous system using a LeVeen or Denver shunt if this simple approach fails. Although these shunts often work well in patients with refractory ascites, chyle often blocks the plastic tubing, or the valve, and produces an early occlusion of the shunt. Many patients improve with simple avoidance of fat and prescription of medium-chain triglycerides combined with diuretics.

A chylothorax may respond to aspiration but often recurs and is best prevented by surgical pleurodesis by pleural stripping. Some patients die from water- or lymph-logged lungs after this procedure as the lymphatics draining the lung become obstructed when they are no longer able to empty into the pleural cavity. Nevertheless many patients with severe problems as the result of megalymphatics can be helped by some of the procedures outlined above. Cutaneous vesicles may be simply excised or touched with the diathermy or cautery, but they tend to recur. Recurrent infections should be treated by a prolonged course of broad-spectrum antibiotics.

Lymphangioma circumscriptum

These lesions are either considered as hamartomas or as localized abnormalities of the cutaneous lymphatic drainage. They present as a number of clear or slightly haemorrhagic cutaneous vesicles, often associated with subcutaneous thickening in the underlying fat (Fig. 31). Whimster thought that a lymphangioma circumscriptum was the result of defective lymphatic drainage from the subcutaneous tissue where a number of cisterns ‘pump' lymph back into the overlying skin. These areas should be excised if they are unsightly or painful. They often occur on the trunk and it is important to excise a generous amount of subcutaneous tissue well beyond the ellipse of skin bearing the vesicles in order to remove the subcutaneous bladders described by Whimster. It is often quite difficult to excise all the skin lesions and they have a propensity to recur: excisional surgery is only required if they are symptomatic.

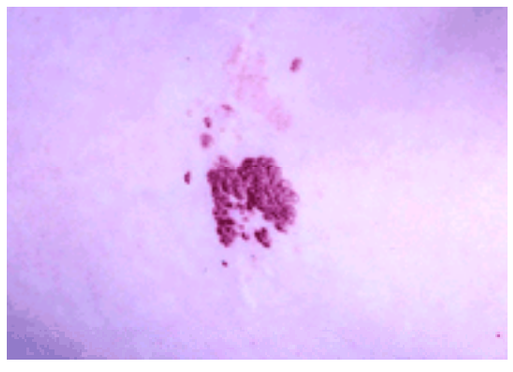


Fig. 31. Lymphangioma circumscriptum with a number of vesicles containing some blood-stained material.

Cystic hygroma

In this developmental abnormality of the lymphatic system, lymphatic fluid collects in a cystic space which is often multilocular and situated in the base of the neck. Cystic hygroma commonly appears in childhood and presents as a soft, brilliantly translucent swelling in the base of the neck. Aspiration and injection of sclerosant may be attempted, but the swellings often recur and may require excision. Cystic hygromas must be dissected with great care as a number of important structures lie adjacent to them.

Mesenteric cysts

These localized lymphatic cysts within the mesentery appear as well-circumscribed mobile lumps within the abdomen. The diagnosis can be confirmed by ultrasound or CT scanning. They are treated by resection, often in association with the overlying area of small bowel. Although harmless, they may reach a considerable size if left untreated.

Further reading

Browse NL, Stewart G. Lymphoedema: pathophysiology and classification. *Journal of Cardiovascular Surgery* 1985; **26**: 91–106.

Casley-Smith JR, Wang CT, Casley-Smith JR, Zi-hai Cui. Testament of filarial lymphoedema and elephantiasis with 5,6-benzo-alpha-pyrone (coumarin). *British Medical Journal* 1993; **307**: 1037–41.

Charles RH. In: Latham A, English TC, eds, *A system of treatment*, vol. 3, pp. 516. Churchill, London, 1912.

Degni M. New techniques of lymphatico-venous anastomosis for the treatment of lymphoedema. *Journal of Cardiovascular Surgery* 1978; **19**: 577–80.

Evans AL, Brice G, Sotirova V, *et al.* Mapping of primary congenital lymphedema to the 5q35.3 region. *American Journal of Human Genetics* 1999; **64**: 547–55.

Kinmonth JB. Lymphangiography in man. *Clinical Science* 1952; **11**: 13–20.

Kinmonth JB. The lymphatics: surgery, lymphography and diseases of the chyle and lymph systems, 2nd edn. Arnold, London, 1982.

Kinmonth JB, Wolfe JH. Fibrosis in the lymph nodes in primary lymphoedema. *Annals of the Royal College of Surgeons of England* 1980; **62**: 344–54.

Kinmonth JB, Patrick JH, Chilvers AS. Comments on operations for lower limb lymphoedema. *Lymphology* 1975; **8**: 56–61.

Newman G. Which patients with arm oedema are helped by intermittent external pneumatic compression therapy? *Journal of the Royal Society of Medicine* 1988; **81**: 377–9.

Nielubowicz J, Olsewski W. Surgical lympho-venous shunts. *British Journal of Surgery* 1968; **55**: 440–2.

O'Brien BM. Micro lymphatico-venous surgery for obstructive lymphoedema. *Australian and New Zealand Journal of Surgery* 1977; **47**: 284.

Ottesen EA, Duke BO, Karam M, Behbani K. Strategies and tools for the control/elimination of lymphatic filariasis. *Bulletin of the World Health Organization* 1997; **75**: 491–503.

Sistrunk WE. Further experiences with the Kondoleon operation of elephantiasis. *Journal of the American Medical Association* 1918; **71**: 800.

Stewart G, Gaunt J, Croft DN, Browse NL. Dynamic and static lymphoscintography in the evaluation of chronic limb oedema. *European Journal of Nuclear Medicine* 1983; **8**: A40.

Thompson N. Surgical treatment of chronic lymphoedema of the arm and leg. *British Journal of Hospital Medicine* 1971; **1**: 395–408.

Thornton A, Pickering D. Abdominal lymphoscintography: an effective substitute for lymphography. *British Journal of Radiology* 1985; **58**: 603–10.

Woolfe JHN. The prognosis and possible cause of severe primary lymphoedema. *Annals of the Royal College of Surgeons of England* 1984; **66**: 251–7.

Zelikovsky A. Treatment of lymphedema with multicompartmental pneumatic device. *Journal of the American Academy of Dermatology* 1990; **23**(5): 951–2.

Zelikovski A *et al.* Lympho-presse: a pneumatic device for the treatment of lymphoedema of the limbs. *Lymphology* 1980; **13**: 68–73.

20.1 Genetic aspects of endocrine disease

Terry C. Lairmore and Samuel A. Wells

Genetics of the multiple endocrine neoplasia syndromes
Multiple endocrine neoplasia type 1(MEN 1)
The multiple endocrine neoplasia type 2 syndromes (MEN 2)
The *RET* proto-oncogene
Early thyroidectomy in patients with MEN 2 based on genetic testing
Genetic defects associated with sporadic thyroid neoplasms
Overview of the tumorigenesis of differentiated thyroid neoplasms
Follicular adenoma and carcinoma
Papillary thyroid carcinoma
Anaplastic thyroid carcinoma
Medullary thyroid carcinoma
Genetic defects associated with parathyroid tumors
Activation of the *PRAD1* proto-oncogene by chromosomal rearrangement
Mutations in the *MEN1* tumor suppressor gene in sporadic parathyroid adenomas
Other genetic syndromes associated with parathyroid neoplasia
Genetic abnormalities in adrenal neoplasms
Familial pheochromocytomas
Sporadic pheochromocytomas
Adrenocortical adenoma and carcinoma
Conclusion
Further reading

Cancer is a genetic disease. Neoplastic transformation of a cell results from a stepwise accumulation of genetic damage that renders the cell unresponsive to normal growth signals. The majority of cancer-causing mutations are somatic and accumulate in individual cells to result in a malignancy. However, the relatively rare human familial cancer syndromes result from the Mendelian inheritance of a mutated gene in the germline DNA.

Mutations in a variety of proto-oncogenes and tumor suppressor genes have been described in endocrine neoplasms. Mutation in one homologous copy of a proto-oncogene, the normal cellular counterpart of the gene, results in an oncogene that encodes a protein product with abnormal or inappropriate expression. Oncogene mutations result in a gain of function that is deleterious for the regulation of cellular growth, and the mutated allele thus confers oncogenic potential in a dominant fashion. In contrast, tumor suppressor genes encode protein products that normally exert a negative control on cellular growth and division, or are integrally involved in programmed cell death (apoptosis). Neoplastic transformation requires inactivation or loss of both alleles of the tumor suppressor gene that results in a loss of function and therefore unregulated cellular growth and proliferation.

Knowledge of the specific genetic defects associated with endocrine tumors has in some cases provided clinically applicable diagnostic or prognostic information. The recent identification of the genes mutated in patients with the multiple endocrine neoplasia type 2 (**MEN 2**) and multiple endocrine neoplasia type 1 (**MEN 1**) syndromes has allowed the development of direct DNA testing for individuals at risk. In patients with medullary thyroid carcinoma arising in the setting of the MEN 2 syndromes, early thyroidectomy may be performed based on genetic testing at a time when the medullary thyroid carcinoma is likely confined to the thyroid gland and therefore curable. Finally, the molecular defects that occur in tumors arising in the familial endocrine neoplasia syndromes have been found to play an important role in tumorigenesis of sporadic neoplasms arising in the same endocrine tissues.

Genetics of the multiple endocrine neoplasia syndromes

Multiple endocrine neoplasia type 1(MEN 1)

Multiple endocrine neoplasia type 1 is characterized by the development of parathyroid hyperplasia, neuroendocrine tumors of the pancreas and duodenum, and adenomas of the anterior pituitary gland. In addition, affected patients develop bronchial and thymic carcinoids, benign thyroid and adrenocortical tumors, subcutaneous lipomas, and ependymomas of the central nervous system with increased frequency. Ninety-five per cent of patients inheriting an *MEN1* gene mutation will develop parathyroid hyperplasia and hyperparathyroidism. Depending on the method of study, 35 to 75 per cent of mutation carriers develop neuroendocrine tumors of the pancreas and duodenum that are frequently malignant. The neuroendocrine tumors in patients with MEN 1 result in symptoms either due to excess secretion of a specific hormone product or the effects of the tumoral process itself. The malignant duodenopancreatic tumors and intrathoracic tumors account for almost all of the disease-related morbidity and mortality in patients with MEN 1.

The familial occurrence of MEN 1 was first recognized by Wermer in 1954. He described the features of MEN 1 in several members of a single kindred, and correctly reasoned that the disease was caused by an autosomal dominant trait with high penetrance. In 1988, the *MEN1* gene was originally localized to chromosome 11q13 by tumor deletion studies and genetic linkage mapping. The observation that chromosome 11 sequences are deleted in tumor DNA suggested that the *MEN1* gene is a tumor suppressor gene. The features of tumor development in patients with MEN 1 (young age at presentation, multiple primary tumors, presence of tumors in multiple target tissues) is consistent with the ‘two-hit' model of tumorigenesis associated with a classic tumor suppressor gene. The first defect is an inherited mutation in the germline DNA, and the second defect is a somatic event that inactivates or deletes the remaining wild-type copy (inherited from the unaffected parent) of the tumor suppressor gene. All of the cells in the body carry the inherited ‘first hit', and each separate ‘second hit' results in a tumor within the target tissues.

The recently identified *MEN1* tumor suppressor gene encodes a predicted 610-amino acid product termed MENIN. The *MEN1* gene contains 10 exons, with the first exon being untranslated. The 2.8-kilobase mRNA transcript is expressed in a wide variety of tissues (lymphocytes, thymus, pancreas, thyroid, testis, ovaries). MENIN has no significant sequence similarity to other known proteins. Evidence has been provided for the presence of two nuclear localization signals in the 3' portion of MENIN and subcellular localization of the normal protein in the nucleus of a cell. The specific role of MENIN in the regulation of cell growth remains to be elucidated.

The *MEN1* mutations are evenly distributed throughout the coding sequence and include missense, nonsense, frameshift, and mRNA splicing defects (Fig. 1). Missense mutations are point mutations in which a DNA base pair change results in the substitution of an erroneous amino acid for the appropriate amino acid at a specific residue within the protein. Nonsense mutations are base pair changes that result in a codon that does not specify an amino acid, and therefore result in premature stopping of translation. Frameshift mutations change the reading frame and therefore usually cause premature termination of translation downstream. The majority of the *MEN1* mutations that have been described result in truncation of the C-terminus of the MENIN protein, which contains the two putative nuclear localization signals.

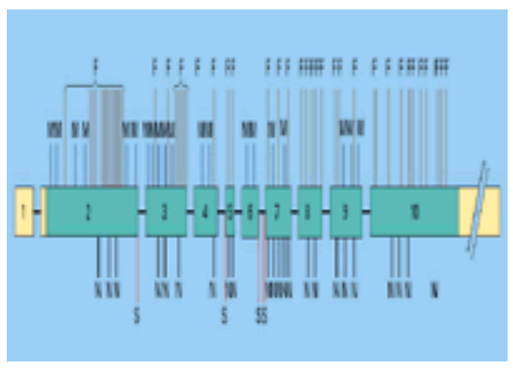


Fig. 1. Distribution of germline mutations along the *MEN1* gene. Missense (M) mutations and frameshift (F) mutations are denoted above the gene, and nonsense (N) mutations and splice (S) defects are denoted below the gene.

The multiple endocrine neoplasia type 2 syndromes (MEN 2)

The multiple endocrine neoplasia type 2 syndromes are a group of clinically and genetically related, autosomal dominant, familial cancer syndromes in which the predominant feature is medullary thyroid carcinoma. Multiple endocrine neoplasia type 2A (MEN 2A) is associated with the development of medullary thyroid carcinoma, pheochromocytomas, and parathyroid hyperplasia or adenomas. Virtually 100 per cent of mutation carriers develop medullary thyroid carcinoma which metastasizes early to cervical lymph nodes, and later to distant sites, and may be lethal.

The hormone calcitonin, which is secreted by the thyroid C-cells, serves as a sensitive tumor marker for the detection of medullary thyroid carcinoma or C-cell hyperplasia. Biochemical screening of at-risk individuals by measuring stimulated levels of plasma calcitonin after the sequential intravenous injection of calcium gluconate and pentagastrin has previously been the most sensitive method for early detection of medullary thyroid carcinoma. However, the provocative test will only be positive after a C-cell neoplasm has developed. Furthermore, the test needs to be performed annually and is associated with transient unpleasant side-effects. Stimulation with calcium and pentagastrin remain the most sensitive way to detect persistent or recurrent disease following thyroidectomy for medullary thyroid carcinoma.

The MEN 2A syndrome is associated with missense germline mutations in the *RET* proto-oncogene, a receptor tyrosine kinase that maps to the centromere region of chromosome 10. In over 95 per cent of families with MEN 2A, the missense mutations involve one of five codons in exons 10 and 11 (codons 609, 611, 618, 620, 634), each of which specifies a highly conserved cysteine residue in the extracellular portion of the receptor immediately adjacent to the cellular membrane (Fig. 2). The mutations result in replacement of a cysteine residue with any of several inappropriate amino acids. Mulligan and coworkers studied the relationship between genotype and the spectrum of target tissue phenotypic expression in patients with MEN 2A and demonstrated that a correlation exists between any mutation at *RET* codon 634 and the development of pheochromocytomas. In addition, a significant correlation exists between the specific codon 634 mutation TGC to CGC (cysteine to arginine) and the development of parathyroid disease.

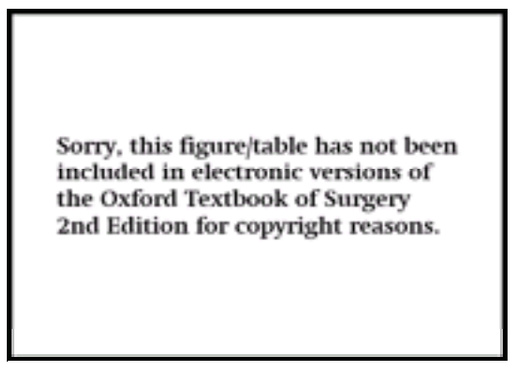


Fig. 2. Schematic representation of the *RET* proto-oncogene structure. Locations of the MEN 2A-associated and FMTC-associated mutations that affect conserved cysteine residues in the portion of the extracellular domain immediately adjacent to the transmembrane segment are indicated by numbers in the yellow oval-shaped circles. The single MEN2B-associated mutation that replaces methionine with threonine in a critical position within the tyrosine kinase catalytic domain is indicated by the number in the diamond-shaped box. The numbers within the symbols indicate the codon number.

The rare multiple endocrine neoplasia type 2B (MEN 2B) syndrome may occur as a *de novo* mutation, and is then transmitted as an autosomal dominant trait in subsequent generations. The MEN 2B syndrome is characterized by a slightly different complex of abnormalities. These patients have a recognizable appearance (Fig. 3) that includes mucosal neuromas (thick, bumpy lips), prognathism, and a ‘marfanoid' body habitus. Patients affected with MEN 2B develop an aggressive form of medullary thyroid carcinoma (often developing in the first decade), pheochromocytomas which are frequently bilateral, skeletal abnormalities, and intestinal ganglioneuromatosis. Parathyroid disease is not a component of the MEN 2B syndrome. Ninety-four per cent of patients with MEN 2B have an identical missense mutation at codon 918 in the tyrosine kinase catalytic domain of *RET* that results in substitution of a threonine for a methionine (Fig. 2). The *RET* codon 918 mutation in patients with *de novo* MEN 2B arises virtually exclusively on the paternally derived allele, and is associated with advanced paternal age. These findings suggest that a *RET* allele may be more susceptible to mutation when inherited from a father as opposed to from a mother. Recently, two patients with clinical features indistinguishable from patients with MEN 2B and a codon 918 mutation were found to have a two base-pair substitution in codon 883 within the catalytic domain resulting in the substitution of phenylalanine for alanine.



Fig. 3. Characteristic appearance of multiple mucosal neuromas on lips, tongue, and buccal mucosa of a patient with MEN 2B.

Familial medullary thyroid carcinoma (FMTC) denotes the autosomal dominant inheritance of medullary thyroid carcinoma without any associated endocrinopathies. The medullary carcinoma in patients with FMTC has a later age of onset and a characteristically indolent biologic behavior that results in almost no disease-related deaths. The clinically distinct syndrome of familial medullary thyroid carcinoma is associated with germline mutations in either the intracellular or extracellular portions of *RET.*

The *RET* proto-oncogene

The *RET* proto-oncogene encodes a member of the transmembrane receptor tyrosine kinase family of signal transduction molecules. The structure of the *RET* protein includes an extracellular ligand-binding domain with homology to the cadherin family of cell adhesion molecules, a transmembrane domain, and an intracellular tyrosine kinase catalytic portion split into two separate domains (Fig. 2). The *RET* gene consists of at least 20 exons spanning 30 kilobases of genomic DNA. The glial-derived neurotrophic factor is a ligand for *RET*.

Embryologic studies in mice suggest a role for *RET* in the migration and development of neural crest cells and in the induction of kidney development. Mice that are homozygous for a targeted mutation in *RET* have renal agenesis, failure of development of enteric ganglia, and die shortly after birth. In keeping with the lack of development of enteric neurons in the mouse model, a subset of patients with familial Hirschsprung's disease have been shown to have loss of function mutations in the *RET* proto-oncogene.

The MEN 2A-associated missense mutations that confer the substitution of a novel amino acid for a conserved cysteine in the RET extracellular domain result in ligand-independent receptor dimerization and autophosphorylation of the protein and lead to tyrosine kinase activation. The mutations associated with MEN 2B occur in the tyrosine kinase catalytic domain and are postulated to alter either substrate specificity or the catalytic properties of the tyrosine kinase domain, resulting in the clinically distinct expression of the syndrome.

Early thyroidectomy in patients with MEN 2 based on genetic testing

Early detection of the thyroid cancer is the key component of effective clinical management for patients affected with one of the MEN 2 syndromes. Medullary thyroid carcinoma is the only consistently malignant feature of MEN 2 and is responsible for almost all of the disease-related deaths. The ability to detect germline *RET* mutations in patients at risk for the MEN 2 syndromes represents a unique model for directed surgical intervention based on the results of a molecular diagnostic test. The inheritance of a *RET* mutation confers the virtual certainty of developing a malignant thyroid neoplasm, providing the rationale for early surgical extirpation of the target organ with the intent of complete removal of early neoplastic C-cell changes while they are still confined to the thyroid gland, and more importantly, for preventing regional nodal or distant metastases of the medullary thyroid carcinoma.

Total thyroidectomy may be performed as early as 6 years of age and is associated with very low morbidity and no mortality. The surgical treatment of patients with a *RET* proto-oncogene mutation should include a total thyroidectomy and removal of the lymph nodes from the central compartment of the neck (from the hyoid bone to the sternal notch and between the carotid sheaths). A modified dissection of the lateral compartment of either side of the neck is indicated when clinically suspicious nodes are present. The appropriate management of the parathyroid glands in patients undergoing thyroidectomy for inherited medullary thyroid carcinoma remains controversial. Our group has recommended total parathryoidectomy with heterotopic autotransplantation of parathyroid tissue into the non-dominant forearm, especially in patients undergoing early thyroidectomy based on genetic testing. The rationale for performing total parathyroidectomy and autotransplantation is to minimize permanent postoperative hypoparathyroidism (the blood supply to the parathyroid glands may be inadvertantly compromised during radical total thyroidectomy), and to prevent the need for neck re-exploration, which carries substantial morbidity, in the small subset of patients who will develop hyperparathyroidism during the lifetime period of risk. Other experts argue that selective parathyroidectomy is effective in almost all patients and that autotransplantation increases the rate of permanent postoperative hypoparathyroidism.

Cervical lymph node metastases are infrequent in patients with MEN 2 who undergo early thyroidectomy in the first or second decade of life on the basis of direct DNA testing. However, microscopic cervical lymph node metastases have been reported in patients as early as 10 years of age, even in the absence of abnormally elevated preoperative stimulated calcitonin levels. These findings consitute the strongest argument for thyroidectomy in the first decade of life in patients with an MEN 2-associated mutation. Patients with MEN 2B should undergo thyroidectomy as soon as the disease is recognized owing to the early onset and aggressive nature of the medullary thyroid carcinoma in this syndrome. Long-term follow-up will be required to determine the frequency of recurrent medullary thyroid carcinoma and the rate of cure in patients with MEN 2 who have been treated with early thyroidectomy based on DNA testing. Direct genetic testing for *RET* mutations provides the best chance for surgical cure of medullary thyroid carcinoma, and obviates the need for continued clinical surveillance in patients who test negative for a mutation in the *RET* gene.

Genetic defects associated with sporadic thyroid neoplasms

Overview of the tumorigenesis of differentiated thyroid neoplasms

The thyroid follicular cell is a differentiated epithelial cell characterized by the ability to concentrate iodine and synthesize thyroid hormones. In the adult, the population of follicular cells exhibits ‘conditional renewal' growth properties. Follicular cells have a very low rate of both active proliferation and programmed cell death in the basal state, but are capable of a significant proliferative response when the appropriate signals are presented. Thyroid-stimulating hormone is the principal tropic factor and regulator of differentiated function of the follicular cell, via binding to a specific cell surface receptor and stimulation of a cAMP-mediated intracellular signal transduction pathway. Thyroid-stimulating hormone stimulates iodine transport, RNA and protein synthesis, and reuptake of colloid with secretion of thyroid hormone. Proliferation of the follicular cell is also influenced by a variety of growth factors (insulin-like growth factor 1, epidermal growth factor) and cytokines (interleukin 1, interleukin 8, transforming growth factor-b), as well as by the amount of iodine in the diet. Finally, other environmental and genetic factors are known to influence significantly the rate of neoplastic transformation of thyroid follicular cells. Exposure to radiation in childhood is a well-established risk factor for the subsequent development of benign and malignant thyroid nodules. The incidence of adenomatous thyroid nodules derived from the follicular cell is also increased in several hereditary cancer syndromes including familial adenomatous polyposis coli, Cowden's disease, and multiple endocrine neoplasia type 1.

Thyroid neoplasms comprise a broad spectrum of tumor types with variable biologic and clinical behavior. Although both papillary carcinoma and follicular carcinoma are derived from the differentiated follicular cell, these two tumors have distinct clinical characteristics. Follicular carcinomas typically are solitary, metastasize hematogenously to the lungs and bones, and occur with increased frequency in areas of iodine deficiency. Conversely, papillary carcinomas are frequently multifocal, metastasize by lymphatic spread to regional nodes, and are associated with previous radiation exposure. Importantly, the basic categories of molecular genetic defects that have been characterized in each of these tumor phenotypes are also distinct. Follicular thyroid tumors are associated with *RAS* mutations as well as mutations in the thyroid-stimulating hormone receptor and *gsp* (the a-subunit of the G_s intracellular signal transducer), while papillary thyroid carcinomas are associated with activation of receptor tyrosine kinases including *RET* and *NTRK1*.

The pathogenesis of differentiated thyroid neoplasms likely involves a combination of genetic and environmental factors. The local growth properties of a follicular cell reflect the summation of the specific positive and negative influences exerted at any one time. A proliferative advantage may be provided to a thyroid follicular cell by genetic mutational events, environmental stimuli, the local influence of growth factors or cytokines, or a combination of these factors. Genetic mutational events may result in an increased susceptibility to further genetic damage. Nodule formation results from cumulative factors that result in loss of differentiated function and unregulated growth, thereby conferring a proliferative advantage to a cell and subsequent clonal expansion. The current understanding of the molecular genetic events and environmental factors associated with thyroid neoplasia are in keeping with the multistep model of tumorigenesis proposed for other epithelial cells. Progression of benign adenomatous nodules to differentiated carcinoma is speculatory at present, but ample evidence exists for the development of undifferentiated or anaplastic carcinomas from pre-existing differentiated thyroid neoplasms, especially papillary carcinoma. A schematic model of the proposed events in thyroid follicular cell tumorigenesis is shown in Fig. 4.

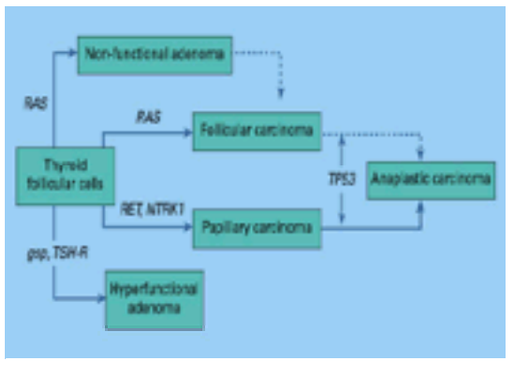


Fig. 4. Flow diagram of events in thyroid follicular cell tumorigenesis. Mutations in the *RAS* oncogene are early events in the development of follicular adenoma and carcinoma. Chromosomal rearrangement and activation of the *RET* and *NTRK1* receptor tyrosine kinases is specific to papillary thyroid carcinomas. Mutations in the *TP53* tumor suppressor gene are associated with malignant progression to undifferentiated thyroid cancers.

Follicular adenoma and carcinoma

Thyroid neoplasms with a follicular histology are associated with a high frequency of activating point mutations in all three members of the *RAS* oncogene family (K-*RAS*, N-*RAS*, and H-*RAS*). The *RAS* family of oncogenes encode 21-kDa small G proteins that are involved in signal transduction from a diverse array of growth factor receptors. *RAS* mutations are thought to represent an early event in thyroid follicular cell tumorigenesis, and have been detected in follicular adenomas, follicular carcinomas, and anaplastic thyroid cancers. By contrast, *RAS* mutations are a rare event in papillary thyroid carcinomas. It appears that *RAS* mutations are not sufficient for transformation of the thyroid cell and that additional genetic events are required. *RAS* mutations in association with follicular thyroid carcinomas are more frequent in areas of iodine deficiency. Moreover, the presence of mutations in a *RAS* oncogene appears to correlate with the incidence of distant (particulary bone) metastases.

Molecular defects strongly associated with hyperfunctioning or ‘hot' thyroid nodules include mutations affecting receptors or intermediates along the adenylate cyclase/cAMP intracellular signaling cascade mediated by the binding of thyroid-stimulating hormone to its cell surface receptor on the thyroid follicular cell. A subset of hyperfunctioning adenomas have been shown to harbor somatic mutations in the gene encoding the thyroid-stimulating hormone receptor that result in constitutive

activation of downstream events. Mutations in the G protein intracellular mediator of adenylate cyclase (G_S) have been detected in 25 per cent of hyperfunctioning thyroid nodules. Therefore, a unifying molecular theme in the pathogenesis of hyperfunctioning nodules is inappropriate activation of the thyroid-stimulating hormone signal transduction pathway resulting from specific gene mutations altering a key mediator of an otherwise well balanced cascade.

Papillary thyroid carcinoma

At the molecular level, a significant proportion of papillary thyroid carcinomas are associated with activation of the receptor tyrosine kinases (RTKs), including the *RET* proto-oncogene on chromosome 10 and the *NTRK1* proto-oncogene on chromosome 1. The members of the receptor tyrosine kinase family normally function as transmembrane signal transduction molecules, in which activation of the intracellular tyrosine kinase catalytic domain and initiation of intracellular signaling pathways results from the binding of specific ligands to the extracellular receptor portion of the molecule.

Activation of receptor tyrosine kinases in papillary thyroid carcinomas usually results from chromosomal rearrangements that fuse sequences encoding the tyrosine kinase catalytic portion of the gene to the upstream or 5' sequences of one of several different ‘activating' genes. This rearrangement of chromosomal sequences results in a chimeric transforming gene (oncogene) in which the tyrosine kinase enzymatic activity is inappropriately or constitutively ‘turned on'. In the majority of cases, the rearrangement is an intrachromosomal inversion that juxtaposes the receptor tyrosine kinase with an inappropriate ‘activating' partner gene located on the same chromosome. Chromosomal translocations that combine the receptor tyrosine kinase with activating gene sequences from a separate chromosome are less frequent events. A hot spot or ‘breakpoint cluster region' for these recombinational events has been demonstrated to occur within both the *RET* and *NTRK1* genes.

The *RET* (*re*arranged during *t*ransfection) proto-oncogene was first discovered based on its ability to transform mouse NIH 3T3 fibroblasts in culture. As discussed above, point mutations or deletions in the *RET* proto-oncogene on chromosome 10q11.2 are associated with the development of medullary thyroid carcinoma in patients with MEN 2A and MEN 2B as well as a small subset of cases of Hirschsprung's disease. By contrast, activation of *RET* by chromosomal rearrangement leading to chimeric transforming sequences is a distinct genetic mechanism associated with the development of thyroid carcinomas of the papillary type. Three distinct forms of activating rearrangement of *RET* have been described and termed *RET/ptc1, RET/ptc2*, and *RET/ptc3. RET/ptc1* and *RET/ptc3* result from intrachromosomal inversions, whereas *RET/ptc2* results from a reciprocal translocation between chromosomes 10 and 17. The most frequent activated chimeric gene (*RET/ptc1*) results from a paracentromeric inversion of chromosome 10q which fuses D10S170 (H4) sequences with the tyrosine kinase domain of *RET* (Fig. 5). The frequency of *RET* rearrangements in papillary carcinomas varies considerably in different geographic areas, possibly due to environmental factors or variability in the genetic backgrounds of the population. The occurrence of *RET* rearrangement has been associated with exposure to ionizing radiation, a well-known risk factor for the development of papillary thyroid carcinoma.

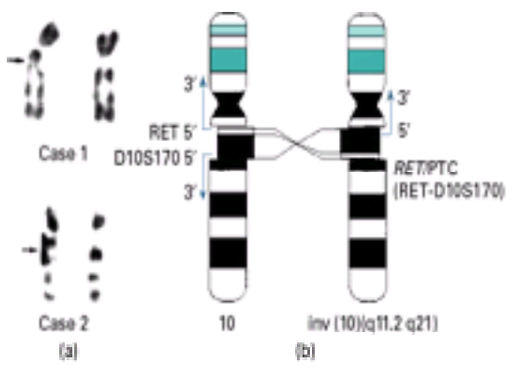


Fig. 5. Chromosome 10q inversion in papillary thyroid carcinoma. (a) Two representative chromosome 10 homologs from tumor cells of patients 1 and 2 showing inv(10)(q11.2q21) (arrows). (b) Schematic view of the paracentric inversion of chromosome 10q generating the transforming sequence *RET/ptc1*.

The *NTRK1* (tropomyosin receptor kinase) proto-oncogene maps to chromosome 1q22 and encodes a component of the high-affinity cell surface receptor for nerve growth factor. In a manner analogous to activating *RET* rearrangements, chromosomal recombinational events that activate *NTRK1* have also been detected in some papillary carcinomas. The *TRK-T1* oncogene is generated by an inversion of chromosome 1q that juxtaposes the tyrosine kinase domain of *NTRK1* and the 5' sequences of TPR (translocated promoter region). Additional intrachromosomal rearrangements (*TRK-T2, TRK-T4*) and a translocation with chromosome 3q (*TRK-T3*) have also been described. Activation of receptor tyrosine kinases by the common mechanism of gene rearrangement that brings the tyrosine kinase domain under inappropriate upstream regulators derived from any of several ‘activating genes' appears to be specific for the transformation of follicular cells into papillary thyroid carcinoma.

Anaplastic thyroid carcinoma

Mutations in the *TP53* tumor suppressor gene on chromosome 17p have been described in a wide variety of human tumors. The *TP53* gene encodes a nuclear phosphoprotein termed p53 that functions as a transcriptional regulator and plays a pivotal role in the regulation of DNA repair, cell cycle arrest, and cellular apoptosis. Disruption of these protective functions appears to be relevant to the progession of thyroid neoplasms to an aggressive, undifferentiated phenotype. Mutations in the *TP53* gene are frequently detected in highly undifferentiated or anaplastic thyroid carcinomas, but occur only rarely in more differentiated neoplasms. The role for *TP53* mutations in the malignant progression of thyroid tumors is supported by the hypothesis that cells with disabled *TP53* functions may incur rapid accumulation of further genetic damage resulting in degeneration to more aggressive growth properties.

Medullary thyroid carcinoma

In contrast to the differentiated neoplasms that arise from the thyroid follicular cell (papillary and follicular carcinomas), medullary thy-roid carcinoma arises from the parafollicular C-cells that are dispersed between the thyroid follicles. Thyroid C-cells are neuroendocrine cells that stream from the neural crest early in embryologic development to populate the thyroid gland along with the follicular cells.

Although medullary carcinoma is an uncommon histopathologic variant of sporadic cases of thyroid cancer (5 to 8 per cent of cases), medullary thyroid carcinoma is the defining feature in patients with the MEN 2 syndromes. Therefore, the diagnosis of medullary thyroid carcinoma in a patient with a thyroid nodule should prompt a thorough family history and measurement of the urinary excretion of catecholamines and metabolites to search for the presence of a pheochromocytoma. Pathologic features that suggest familial disease include multiple, bilateral foci of medullary thyroid carcinoma and the presence of a C-cell hyperplasia in areas of the thyroid distant from microscopic or gross foci of medullary thyroid carcinoma.

The molecular mechanisms that are responsible for hereditary medullary thyroid carcinoma have been found also to play a role in the tumorigenesis of sporadic cases of medullary thyroid carcinoma. Whereas the transmission of the *RET* mutation from parent to offspring in the germline results in a familial MEN 2 syndrome, the rare somatic occurrence of a *RET* mutation within a thyroid C-cell may give rise to a sporadic medullary thyroid carcinoma. Therefore, testing for *RET* mutations in normal and tumor DNA is useful to differentiate sporadic from familial cases of medullary carcinoma of the thyroid. Patients with MEN 2 inherit one mutated copy of *RET* in the germline (constitutional) DNA. This inherited mutation confers the risk of tumor formation in the entire population of thyroid C-cells, and one mutated allele of *RET* is therefore present in both tumor DNA and constitutional DNA. In sporadic cases of medullary thyroid carcinoma, one mutated copy of *RET* is present in the tumor DNA, whereas both copies of *RET* in the germline are wild type.

Approximately 33 to 67 per cent of sporadic cases of medullary thyroid carcinoma have been found to harbor a somatic *RET* mutation in the tumor DNA. The most common *RET* mutation identified in sporadic medullary thyroid carcinomas is the codon 918 missense mutation resulting in the substitution of threonine for methionine within the catalytic domain. This is the identical mutation that is present in the germline in patients with MEN 2B. Mutations involving the five conserved cysteine codons in the extracellular domain of *RET* that are frequently mutated in the germline of patients with MEN 2A are infrequently found in tumor DNA from sporadic medullary thyroid carcinomas. When detected, these mutations are often found to be present in the germline as well, indicating the index case of a patient with familial MEN 2 or a *de novo* germline mutation.

Activating mutations of *RAS* or other oncogenes are very infrequent in tumor DNA from medullary thyroid carcinomas. Although a higher than expected rate of chromosomal loss for 3q and 22q has been observed in medullary thyroid carcinomas, a strong role for tumor suppressor genes in the development of sporadic medullary thyroid carcinoma has not been established. In summary, with the exception of mutations in the *RET* proto-oncogene, few molecular events have been

consistently demonstrated to contribute to the initiation or progression of medullary thyroid carcinoma.

Genetic defects associated with parathyroid tumors

Parathyroid adenomas are typically solitary, benign tumors characterized by both an increase in cellular mass (hyperplasia) as well as an abnormality in the ‘set point' of parathyroid hormone release in response to the extracellular calcium concentration. Primary hyperparathyroidism due to a parathyroid adenoma occurs with increased frequency in postmenopausal women and individuals with a history of exposure to neck irradiation. Hyperparathyroidism due to multiglandular parathyroid disease occurs in association with a variety of genetically distinct familial endocrine neoplasia syndromes. Three broad categories of molecular genetic abnormalities have been associated with parathyroid tumors, including: (1) activation of a proto-oncogene, (2) inactivation of a tumor suppressor gene, and (3) mutations in the gene encoding the cell surface calcium-sensing receptor.

Activation of the *PRAD1* proto-oncogene by chromosomal rearrangement

In a small subset (5 per cent) of parathyroid adenomas, a pericentromeric inversion of chromosome 11 has been identified that results in activation of the *PRAD1* (*par*athyroid *ad*enoma1) proto-oncogene on chromosome 11q13. The breakpoint and rejoining of this intrachromosomal event places the transcriptional regulatory elements of the parathyroid hormone gene immediately upstream of the *PRAD1* gene (Fig. 6). The consequence of this erroneous association is inappropriate or unregulated expression of the *PRAD1* protein product. The novel influence of the parathyroid hormone regulatory elements results in a parathyroid tissue-specific overexpression of product of the activated oncogene. *PRAD1* encodes a 295-amino acid product, also termed cyclin D1, that functions as a gatekeeper in the regulation of progression through the cell cycle by facilitating advancement of a cell through the critical G1/S phase checkpoint.

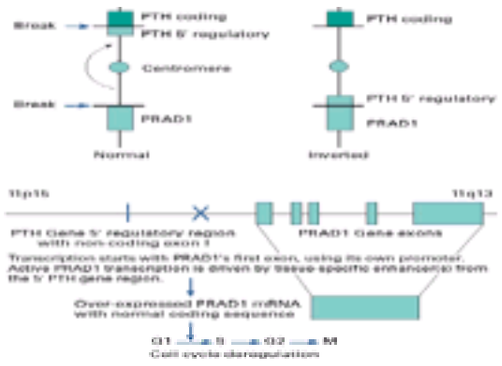


Fig. 6. (Above) Schematic diagram illustrating the pericentromeric inversion of chromosome 11 responsible for recombination between the parathyroid hormone (PTH) gene and the *PRAD1* proto-oncogene in a subset of parathyroid adenomas. (Below) Diagram of the molecular stucture of the PTH/*PRAD1* DNA rearrangement and its functional consequences. X represents the chromosome breakpoint between the PHT gene regulatory region, plus PTH non-coding exon 1 (solid, light, vertical bar) and part of its first intron from 11p15 (left), and the intact promoter and five exons of the *PRAD1* proto-oncogene from 11q13.

The parathyroid tissue-specific overexpression of cyclin D1 presumably results in unregulated cellular growth without conferring malignant transformation or invasive potential, in keeping with the benign phenotype of parathyroid adenomas. Parathyroid adenomas associated with a *PRAD1* proto-oncogene rearrangement may be larger or more often symptomatic when compared with other sporadic parathyroid adenomas.

Mutations in the *MEN1* tumor suppressor gene in sporadic parathyroid adenomas

Inactivation of both copies of the *MEN1* tumor suppressor gene on chromosome 11q13 is associated with the development of multiglandular parathyroid neoplasms in greater than 95 per cent of individuals affected with the familial MEN 1 syndrome. In keeping with the established theme of common molecular mechanisms for familial tumors and their sporadic counterparts, it is now recognized that a significant subset of sporadic parathyroid adenomas arises from the occurrence of sequential somatic events that result in inactivation of both copies of the *MEN1* tumor suppressor gene within the DNA of a parathyroid cell. In these instances, the combination of targeted genetic defects presumably results in clonal expansion and the development of a parathyroid adenoma.

Chromosome deletions encompassing the *MEN1* tumor suppressor gene on chromosome 11q13 have been detected in approximately one-third of sporadic parathyroid adenomas. A recent study confirmed the presence of somatic mutations in one allele of the *MEN1* gene in tumor DNA from 7 of 33 parathyroid adenomas. In each of these tumors, a chromosome deletion was detected that eliminated the remaining wild-type copy of the *MEN1* gene in tumor DNA. These findings suggest that a significant subset of sporadic parathyroid adenomas are caused by a combination of somatic genetic events within a parathyroid cell that inactivate both copies of the *MEN1* tumor suppressor gene.

Other genetic syndromes associated with parathyroid neoplasia

Parathyroid neoplasia is a component of several additional infrequent genetic syndromes. Familial benign hypercalcemia, or familial hypercalcemic hypocalciuric hyperparathyroidism (**FHHH**), is an autosomal dominant syndrome usually associated with mutations in the gene on chromosome 3q that encodes the membrane bound calcium-sensing receptor on parathyroid cells. FHHH is characterized by mild hypercalcemia, low urinary excretion of calcium, and mild or absent complications of hyperparathyroidism. Surgical intervention is unsuccessful in the treatment of FHHH. Other rare genetic syndromes associated with parathyroid tumors include the association of parathyroid adenoma or carcinoma with ossifying jaw fibromas and familial isolated hyperparathyroidism.

Genetic abnormalities in adrenal neoplasms

Familial pheochromocytomas

Pheochromocytomas may occur as a component of several autosomal dominant hereditary syndromes, including multiple endocrine neoplasia type 2A (MEN 2A) or type 2B (MEN 2B), von Hippel–Lindau syndrome, and neurofibromatosis type 1. In the familial setting, pheochromocytomas may occur bilaterally. Approximately 30 to 50 per cent of patients with MEN 2A or MEN 2B develop adrenal medullary neoplasms. A preneoplastic proliferation of the chromaffin cells of the adrenal medulla (adrenal medullary hyperplasia) precedes and coexists with the development of discrete pheochromocytomas in this familial cancer syndrome. Extra-adrenal and malignant pheochromocytomas are rare in the setting of MEN 2. Although less frequent, pheochromocytomas are also associated with neurofibromatosis type 1 and the von Hippel–Lindau syndrome.

The most consistent cytogenetic abnormality detected in familial pheochromocytomas is a chromosome deletion involving chromosome 1p. The finding of frequent 1p chromosome deletions, especially in pheochromocytomas arising in the MEN 2 syndromes, suggests the existence of a tumor suppressor gene in this region which may play a facilitative role in the initiation or progression of pheochromocytomas in patients who have inherited a germline mutation in the primary cancer predisposition locus, the *RET* proto-oncogene on chromosome 10. Interestingly, chromosome 1p deletions have been detected in other tumors of neuroectodermal origin, including neuroblastomas and melanomas.

Sporadic pheochromocytomas

The same genetic abnormalities implicated in the pathogenesis of familial pheochromocytomas have been detected, with decreased frequency, in sporadic pheochromocytomas. Chromosome 1p deletions have been described in approximately 30 per cent of sporadic pheochromocytomas. Somatic mutations in exons 10, 11, and 16 of the *RET* proto-oncogene are present in 10 to 21 per cent of sporadic pheochromocytomas.

Adrenocortical adenoma and carcinoma

Inactivating mutations or deletion of the *TP53* tumor suppressor gene are found with significant frequency in adrenocortical carcinomas, but not in benign adrenocortical adenomas or hyperplasia. These observations suggest a role for *TP53* mutations in the progression of adrenocortical neoplasms to the malignant phenotype. It is of interest to note that adrenocortical carcinomas occur with increased frequency in patients with the Li–Fraumeni syndrome, who have been shown to

harbor germline mutations in *TP53*.

Conclusion

A variety of genetic abnormalities in both oncogenes and tumor suppressor genes have been described in endocrine tumors. The unraveling of the basic genetic defects responsible for the development of multiple endocrine tumors in specific target tissues in the familial endocrine neoplasia syndromes has provided valuable insight into the mechanisms of tumorigenesis in the rarer sporadic neoplasms arising in endocrine tissues. The tumor suppressor gene which is mutated in patients with the MEN 1 syndrome has recently been identified. Defects in the *MEN1* gene are associated with the development of a subset of sporadic parathyroid adenomas. The ability to detect *RET* proto-oncogene mutations by direct DNA testing in patients at risk for the multiple endocrine neoplasia type 2 syndromes has made early thyroidectomy possible at a time when the medullary carcinoma may be confined to the thyroid gland and therefore amenable to cure by surgical resection.

Further reading

Arnold A. Genetic basis of endocrine disease 5. Molecular genetics of parathyroid gland neoplasia. *Journal of Clinical Endocrinology and Metabolism* 1993; **77**: 1108–12. [Comprehensive up-to-date summary of the molecular pathogenesis of parathyroid neoplasms.]

Carlson KM *et al.* Single missense mutation in the tyrosine kinase catalytic domain of the *RET* protooncogene is associated with multiple endocrine neoplasia type 2B. *Proceedings of the National Academy of Sciences of the USA* 1994; **91**: 1579–83. [One of the original articles describing the RET mutation associated with the MEN 2B syndrome.]

Chandrasekharappa SC *et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 1997; **276**: 404–7. [The original article describing the successful positional cloning of the MEN1 tumor suppressor gene.]

Donis-Keller H *et al.* Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. *Human Molecular Genetics* 1993; **2**: 851–6. [One of the original articles describing germline RET mutations associated with multiple endocrine neoplasia type 2.]

Edery P *et al.* Mutations of the *RET* proto-oncogene in Hirschsprung's disease. *Nature* 1994; **367**: 378–80. [One of the original articles describing RET mutations in a subset of children with Hirschsprung's disease.]

Farndon JR *et al.* Familial medullary thyroid carcinoma without associated endocrinopathies: A distinct clinical entity. *British Journal of Surgery* 1986; **73**: 278–81. [Original description of familial medullary thyroid carcinoma as a distinct clinical entity.]

Foley TP Jr, Harrison HC, Arnaud CD, Harrison HE. Familial benign hypercalcemia. *Journal of Pediatrics* 1972; **81**: 1060–7. [Comprehensive summary of the clinical characteristics of familial benign hypercalcemia.]

Grieco M *et al.* PTC is a novel rearranged form of the *ret* proto-oncogene and is frequently detected *in vivo* in human thyroid papillary carcinomas. *Cell* 1990; **60**: 557–63. [Original article describing the *RET* gene rearrangement associated with a subset of papillary thyroid carcinomas.]

Guru SC *et al.* Menin, the product of the *MEN1* gene, is a nuclear protein. *Proceedings of the National Academy of Sciences of the USA* 1998; **95**: 1630–4. [Description of the presumed nuclear localization of the menin protein encoded by the *MEN1* tumor suppressor gene.]

Hofstra RMW *et al.* A mutation in the *RET* proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature* 1994; **367**: 375–6. [One of the original articles describing the RET mutation associated with the MEN 2B syndrome.]

Jing S *et al.* GDNF-induced activation of the Ret protein tyrosine kinase is mediated by GDNFR-a, a novel receptor for GDNF. *Cell* 1996; **85**: 1113–24. [Article describing the role of GDNFR-a in the activation of RET receptor tyrosine kinase.]

Lairmore TC, Frisella MM, Wells SA Jr. Genetic testing and early thyroidectomy for inherited medullary thyroid carcinoma. *Annals of Medicine* 1996; **28**: 401–6. [This study reports the results of early thyroidectomy based on genetic testing in a large cohort of patients with MEN 2.]

Larsson C, Skogseid B, Öberg K, Nakamura Y, Nordenskjöld M. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. *Nature* 1988; **332**: 85–7. [Original article mapping genetic locus for MEN 1 to chromosome 11.]

Lemoine NR *et al.* High frequency of ras oncogene activation in all stages of human thyroid tumorigenesis. *Oncogene* 1989; **4**: 159–64. [Description of the frequency of RAS mutations in follicular thyroid neoplasms.]

Mulligan LM, Ponder BAJ. Genetic basis of endocrine disease: multiple endocrine neoplasia type 2. *Journal of Clinical Endocrinology and Metabolism* 1995; **80**: 1989–95.

Mulligan LM *et al.* Germ-line mutations of the *RET* proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* 1993; **363**: 458–60. [Excellent review of the genetic basis for the MEN 2 syndromes.]

Mulligan LM *et al.* Specific mutations of the *RET* proto-oncogene are related to disease phenotype in MEN 2A and FMTC. *Nature Genetics* 1994; **6**: 70–4. [This article reports the findings of associations between genotype and phenotype for the RET mutations involved in patients with MEN 2.]

Pollak MR *et al.* Mutations in the human Ca^{2+}-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell* 1993; **75**: 1297–303. [Original article describing mutations in the Ca^{2+}-sensing receptor gene in patients with familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism.]

Santoro M *et al.* Activation of *RET* as a dominantly transforming gene by germline mutations of MEN2A and MEN2B. *Science* 1995; **267**: 381–3. [Elegant study demonstrating that mutated *RET* acts as a dominantly tranforming oncogene *in vitro*.]

Sipple JH. The association of pheochromocytoma with carcinoma of the thyroid gland. *American Journal of Medicine* 1961; **31**: 163–6. [First description of the association of thyroid carcinoma and pheochromocytoma, and the clinical syndrome which would become known as multiple endocrine neoplasia type 2 or Sipple's syndrome.]

Wells SA Jr, Baylin SB, Linehan WM, Farrell RE, Cox EB, Cooper CW. Provocative agents and the diagnosis of medullary carcinoma of the thyroid gland. *Annals of Surgery* 1978; **188**: 139–41. [Definitive paper describing the use of provocative agents and the superiority of calcium and pentagastrin for the early diagnosis of medullary thyroid carcinoma.]

Wells SA Jr *et al.* Predictive DNA testing and prophylactic thyroidectomy in patients at risk for multiple endocrine neoplasia type 2A. *Annals of Surgery* 1994; **220**: 237–50. [Detailed early report of prophylactic thyroidectomy for patients with MEN 2 based on the results of direct genetic testing.]

Wermer P. Genetic aspects of adenomatosis of the endocrine glands. *American Journal of Medicine* 1954; **16**: 363–71. [One of the first comprehensive descriptions of the clinical features of MEN 1, which the author correctly proposed an autosomal dominant mode of inheritance.]

Williams ED, Pollack DJ. Multiple mucosal neuromata with endocrine tumours: A syndrome allied to von Recklinghausen's disease. *Journal of Pathology and Bacteriology* 1966; **91**: 71–80. [Early description of the MEN 2B syndrome.]

20.2.1 Physiology of the thyroid gland

J. A. Franklyn

Thyroid hormone synthesis and secretion
Thyroid hormone metabolism, transport, and action
Biochemical assessment of thyroid function
Tests for the investigation of suspected hyperthyroidism
Tests for the investigation of suspected hypothyroidism
Other investigations in patients with suspected thyroid dysfunction
Hyperthyroidism
Pathogenesis
Clinical features of hyperthyroidism
Management of hyperthyroidism
Hypothyroidism
Pathogenesis
Clinical features of hypothyroidism
Management of hypothyroidism
Further reading

Thyroid hormone synthesis and secretion

The major function of the follicular (epithelial) cells of the thyroid is to trap, transport, and oxidize iodide, leading ultimately to the synthesis of iodothyronines (Fig. 1). The major synthetic products of the thyroid gland are thyroxine (**T4**) and tri-iodothyronine (**T3**), which are stored in the colloid within thyroid follicles in association with the large glycoprotein thyroglobulin and released in response to stimulation of the thyroid gland by thyrotrophin (thyroid-stimulating hormone, **TSH**).

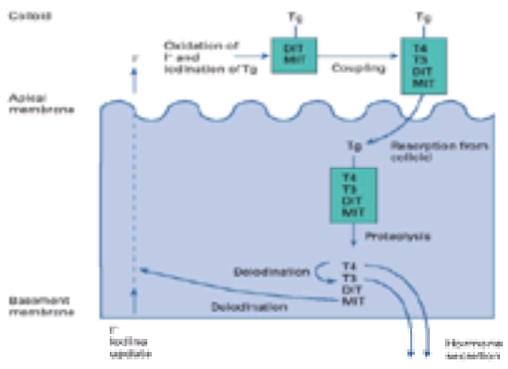


Fig. 1. Diagrammatic representation of thyroid hormone synthesis and release from a thyroid follicular cell. Processes of organification of iodide (I^-) and iodination of thyroglobulin (Tg) to form di-iodotyrosine (DIT) and mono-iodotyrosine (MIT) are illustrated, together with coupling of molecules of MIT and DIT to form thyroxine (T_4) and triiodothyronine (T_3).

The synthesis of thyroid hormones is dependent upon the availability of inorganic iodine in the diet. The requirement for dietary iodine is 100 to 200 µg/day, iodine being obtained largely from milk and other dairy products, as well as from seafoods. Intake of iodine varies considerably depending upon geographical location, the United Kingdom and United States together with Japan being iodine replete areas. Worldwide, dietary deficiency of iodine (leading to goitre and hypothyroidism, see below) represents a major public health problem with millions of affected subjects, iodine deficiency being especially prevalent in mountainous regions. Programmes of iodine repletion are underway in such areas and involve iodine supplementation of foodstuffs such as salt and bread or injection of iodized oils.

Ingested iodides are absorbed in the stomach and upper gastrointestinal tract and inorganic iodide is removed from the plasma at a rate of 10 µg/h through a process of active transport into thyroid follicular cells. Inorganic iodide attaches to a specific thyroid peroxidase at the apical (luminal) aspect of the follicular cell which also represents a binding site for tyrosine residues on thyroglobulin. Both the iodide and tyrosine are oxidized by the peroxidase enzyme to form mono-iodotyrosine and di-iodotyrosine. On the thyroglobulin molecule, coupling of mono-iodotyrosine and di-iodotyrosine molecules occurs through the action of peroxidase and transaminase resulting in the synthesis of T_3 (mono-iodotyrosine + di-iodotyro-sine) or T_4 (di-iodotyrosine + di-iodotyrosine).

Thyroglobulin with its associated thyroid hormones is reabsorbed from the colloid by endocytosis into the apical aspect of the thyroid follicular cell. Thyroglobulin is then hydrolysed by peptidases within lysosomes and free thyroid hormones released by diffusion into capillaries from the basal surface of the cell. Small quantities of thyroglobulin are also released during this process while mono-iodotyrosine and di-iodotyrosine molecules are deiodinated and degraded to release iodine and tyrosine which are recycled.

These processes of iodine trapping and incorporation, and thyroid hormone synthesis and release, are tightly regulated by the glycoprotein hormone TSH released from the anterior pituitary (Fig. 2). TSH acts via surface membrane receptors on the follicular cell. The synthesis and release of TSH is, in turn, regulated by the hypothalamic peptide thyrotrophin-releasing hormone (**TRH**), which reaches the anterior pituitary via the capillary bed of the pituitary stalk. Circulating concentrations of free T_4 and free T_3 are normally tightly controlled by a negative feedback loop, with increases in T_4 and T_3 leading to inhibition of TSH synthesis and release from the pituitary, and to a lesser extent inhibition of hypothalamic TRH release. Conversely, a fall in circulating thyroid hormone concentrations is associated with reduced negative feedback inhibition of the pituitary and hypothalamus and a rise in TSH synthesis and release.

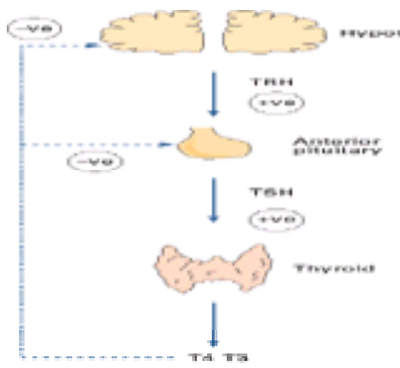


Fig. 2. Representation of the hypothalamic–pituitary–thyroid axis. Thyrotrophin-releasing hormone (TRH) stimulates the release of thyrotrophin (thyroid-stimulating hormone, TSH) from the anterior pituitary which in turn stimulates the synthesis and release of thyroid hormones T_4 and T_3 from the thyroid gland. T_4 and T_3 then exert a negative feedback effect upon TSH production and release so that thyroid homeostasis is maintained.

Thyroid hormone metabolism, transport, and action

T_4 is produced exclusively by the thyroid gland but the majority (80 per cent) of T_3 is produced by peripheral conversion (deiodination) of T_4, which takes place largely in

the liver and kidneys. Most of the T_3 in the body is intracellular and in some tissues, especially the pituitary, it is the local rate of deiodination of T_4 to T_3 which controls the supply of hormone to intracellular receptors to a greater extent than the circulating concentration of T_3. Depending upon which iodine atom is removed, T_4 can also be deiodinated to the inactive metabolite reverse T_3 (**rT3**) (Fig. 3). The function of the deiodinase enzymes, and hence the rate of production of T_3 and rT_3 in the peripheral tissues, can be affected by a variety of factors including drugs and illnesses. Serious illnesses such as renal failure, cirrhosis, and major surgery typically result in inhibition of the enzyme 5' monodeiodinase (which catalyses conversion of T_4 to T_3) and hence a fall in the circulating concentration of T_3 and an associated rise in rT_3.

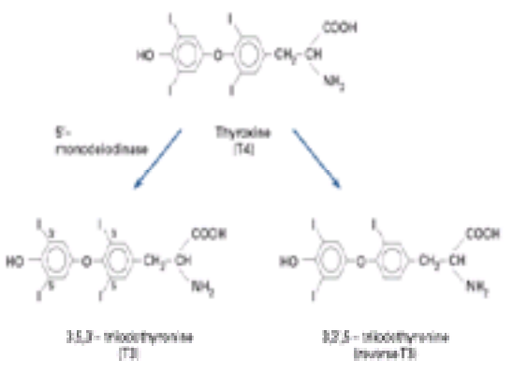


Fig. 3. Molecular structures of T_4, T_3, and reverse T_3.

Thyroid hormones in the circulation are largely bound to binding proteins with only a tiny fraction (< 0.5 per cent) being in the ‘free' or non-protein-bound form. The most important binding protein is thyroxine-binding globulin (**TBG**) which binds T_4 with a higher affinity than T_3. A much smaller proportion of T_4 and T_3 is bound to thyroxine-binding prealbumin (transthyretin), as well as to albumin, which both have a low affinity but high capacity for binding thyroid hormones in serum. A variety of factors affect serum TBG concentrations, the most marked changes being associated with pregnancy (rise in TBG) and in congenital deficiency or excess. These changes affect total (protein-bound) concentrations of T_4 and T_3 but do not influence free thyroid hormone concentrations in serum.

Free or non-protein-bound T_4 and T_3 enter cells largely by passive diffusion from the circulation, although there may be some active transport as well. The intracellular actions of thyroid hormones are mediated by binding to specific receptors within the nucleus, although it is suggested that there may be direct extranuclear effects, especially upon mitochondria. There exist two major classes of thyroid hormone receptors within the nucleus, termed a and b, and various specific isoforms within those classes. These receptors bind T_3 with a 10-fold higher affinity than T_4, determining that the role of T_4 is that of a prohormone, the biologically active thyroid hormone being T_3. Thyroid hormone receptors together with T_3 (and in the presence of other nuclear transcription factors such as retinoid X-receptors) bind to specific regions of DNA with high affinity to influence the rate of transcription of specific genes directly and hence the rate of protein synthesis (Fig. 4). Thyroid hormone receptors are present in almost every tissue of the body, a fact reflected in the diverse physiological actions of thyroid hormones and marked clinical consequences of thyroid hormone excess or deficiency.

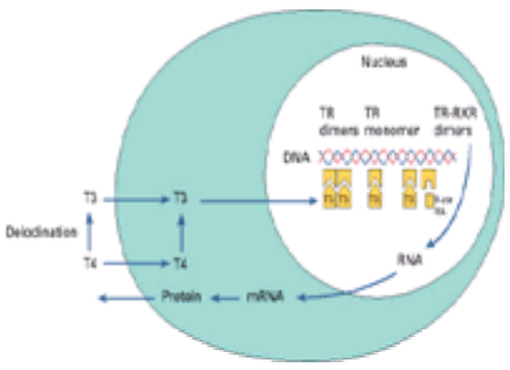


Fig. 4. Diagrammatic representation of the cellular effects of T_3. Once inside the cell T_3 is transported to the nucleus where it binds with high affinity to specific nuclear receptor proteins (TRs). TRs bind as monomers or dimers, either in association with each other or with retinoid X-receptors (RXRs), to DNA upstream of coding sequences for genes regulated by thyroid hormones. Binding of the hormone receptor complex (together with other nuclear transcription factors) to DNA stimulates or inhibits the transcription of target genes to mRNA and hence the rate of protein synthesis.

Biochemical assessment of thyroid function

It is essential that any clinical suspicion of thyroid dysfunction is investigated by assessment of thyroid function since symptoms and signs can be non-specific or misleading. Central to the biochemical assessment of thyroid function is understanding of the hypothalamic–pituitary–thyroid axis described above.

Measurements of serum free T_4 and free T_3 concentrations together with TSH currently represent the most useful tests of thyroid function. Methods for measurement of free thyroid hormones have largely superseded those for measurement of total circulating concentrations of T_4 and T_3 since they provide a more direct index of the biologically active fraction of thyroid hormones and because they are unaffected by changes in levels of TBG. Rarely, however, spuriously abnormal results can be obtained in assays of free thyroid hormones secondary to the presence of binding proteins with abnormal affinity (familial dysalbuminaemic hyperthyroxinaemia) or autoantibodies which lead to assay interference. Sensitive methods for measurement of TSH (which are able to distinguish low concentrations associated with thyrotoxicosis from normal values in euthyroid patients) have now largely superseded examination of the TSH response to TRH (the TRH test).

Tests for the investigation of suspected hyperthyroidism

Hyperthyroidism is confirmed biochemically by elevation of serum free T_4 concentrations and suppression of serum TSH to below normal. If serum TSH is low but free T_4 concentration normal, hyperthyroidism can be indicated by elevation of serum free T_3 (‘T_3-toxicosis'). Suppression of serum TSH alone is not indicative of overt hyperthyroidism and may reflect a number of factors including the presence of goitre, drug therapy (e.g. with glucocorticoids), or non-thyroidal illness. The finding of elevated free thyroid hormone concentrations with a normal or even elevated serum TSH is unusual and raises the possibility of assay interference, or less commonly TSH-secreting tumours of the anterior pituitary or hereditary syndromes of thyroid hormone resistance.

Tests for the investigation of suspected hypothyroidism

Hypothyroidism is confirmed biochemically by a rise in serum TSH concentration together with a fall in serum free T_4. Measurement of free T_3 is not useful in the investigation of hypothyroidism because circulating T_3 levels are often maintained within the normal range even in marked thyroid failure. A rise in serum TSH in the absence of a fall in free T_4 (termed subclinical hypothyroidism) is usually indicative of early thyroid failure and warrants biochemical follow-up, although a transient modest rise in TSH can occur during the recovery phase of a serious illness or after partial thyroidectomy. Falls in serum free T_4 and free T_3 in the absence of a rise in serum TSH are usually indicative of non-specific influences of drug therapy (e.g. anticonvulsants, antidepressants) or non-thyroidal illness, but may also indicate hypothyroidism secondary to pituitary/hypothalamic disease.

Other investigations in patients with suspected thyroid dysfunction

Measurement of antibodies to thyroid peroxidase (also known as microsomal antibodies) and to thyroglobulin may indicate the underlying pathology in patients with thyroid dysfunction. A high titre of such antibodies is found in patients with autoimmune thyroid failure (Hashimoto's thyroiditis), and a low titre in patients with autoimmune hyperthyroidism (Graves' disease). Such antibodies are also found frequently in subjects without obvious thyroid disease and serve as a marker for increased risk of progression to thyroid dysfunction. Antibodies to the TSH receptor explain the pathogenesis of Graves' hyperthyroidism and are found in most such patients at presentation. Their measurement is not routinely available in most centres.

Measurements of isotope (iodine or technetium) uptake into the gland are not routinely indicated because they give an imprecise measure of thyroid function. Measurement of iodine uptake is useful in distinguishing Graves' hyperthyroidism (high uptake) from hyperthyroidism secondary to thyroiditis (in which uptake is low) or cases secondary to iodine ingestion (uptake low). Isotope imaging of the thyroid (usually with technetium-99m) is not indicated routinely but will distinguish Graves' hyperthyroidism from toxic nodular disease if that distinction cannot be made clinically.

Hyperthyroidism

Pathogenesis

Graves' disease is the commonest cause of hyperthyroidism in the United Kingdom. This is an autoimmune disorder typically affecting females in the first four decades of life. It is associated with antibodies to TSH receptors on thyroid follicular cells. These antibodies have been recognized for some years and were formerly termed ‘long-acting thyroid stimulators' because of their effect of prolonged and potent stimulation of thyroid cell function. Thyroid histology shows lymphocytic infiltration, proliferation of follicular cells, and increased vascularity. The next most common cause of hyperthyroidism is toxic nodular hyperthyroidism in which one or more nodules of the thyroid functions autonomously leading to excess thyroid hormone production. If isotope imaging reveals a single ‘hot' nodule to be responsible, this is often called ‘solitary toxic nodule' or ‘solitary adenoma'. Toxic nodular hyperthyroidism, in contrast to Graves' hyperthyroidism, typically affects females of 60 years or over and there is often a preceding history of long-standing nodular goitre.

An infrequent cause of hyperthyroidism is subacute or silent thyroiditis. Subacute (DeQuervain's) thyroiditis is typically associated with a history of viral illness, thyroid tenderness, and raised erythrocyte sedimentation rate. The condition is characterized by low or absent uptake of isotopes of iodine into the thyroid because the process is a destructive one associated with release of preformed thyroid hormones (and hence biochemical hyperthyroidism) and not increased synthesis of thyroid hormones found in Graves' disease. Silent thyroiditis is a similar condition but is not associated with local symptoms or signs in the neck. This may occur sporadically but also commonly occurs in the first 12 months post-partum. It is important to differentiate cases of thyroiditis from Graves' or toxic nodular hyperthyroidism because standard antithyroid therapies with drugs or radioiodine are ineffective and hence inappropriate.

Globally, iodine intake is an important determinant of thyroid function. Nodular goitre is more common in areas of relative iodine deficiency (including some parts of Western Europe such as Germany). Starting iodine replacement therapy precipitates hyperthyroidism in a significant proportion of patients with nodular goitre. Hyperthyroidism also complicates the use of iodine-containing radiographic contrast agents, ingestion of kelp, or various drug therapies, especially the iodine-containing antiarrhythmic amiodarone.

Clinical features of hyperthyroidism

The typical clinical features of hyperthyroidism are illustrated in Table 1. Features such as those listed are characteristically seen in younger patients with hyperthyroidism, older patients more typically presenting with less specific features such as tiredness, apathy, worsening heart failure, or angina. Lid lag and lid retraction can be seen in hyperthyroidism of any cause, but other eye signs including periorbital oedema, proptosis, and diplopia are features ofthyroid-associated ophthalmopathy and diagnostic of Graves' disease (Fig. 5).

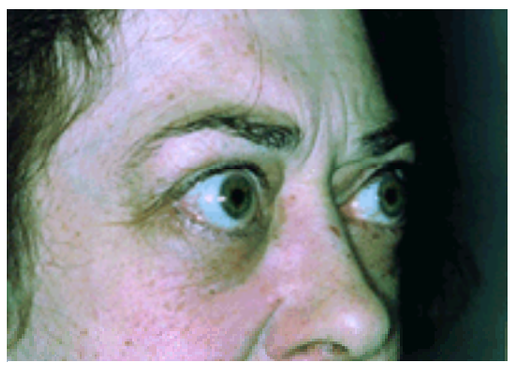


Fig. 5. An example of Graves' ophthalmopathy illustrating lid retraction, conjunctival injection, and forward protrusion of the eyes (proptosis).



Table 1 Clinical features of hyperthyroidism

Management of hyperthyroidism

Three treatment options exist for the management of Graves' or toxic nodular hyperthyroidism—medical therapy with a thionamide drug, radioiodine, or surgery. Each of these treatments has advantages and disadvantages and none can be guaranteed to result in permanent euthyroidism. Recent surveys of thyroid specialists have indicated a growing trend for the use of radioiodine and a lessening role of surgery. Standard practice in the United Kingdom is to treat young patients (age less than 40 years) with a first episode of Graves' hyperthyroidism medically in the first instance but to consider older patients with Graves' disease at presentation, those with relapsed Graves' disease, and those with toxic nodular hyperthyroidism for definitive treatment with radioiodine or, less commonly, surgery. Even in young patients with Graves' disease there is a trend for increasing use of radioiodine early in the disease because of low long-term remission rates (< 50 per cent) with antithyroid drugs alone.

Thionamide drug therapy

Antithyroid drugs can be used short term (up to 6 months) to prepare patients for definitive therapy with radioiodine or surgery, or medium term (18 months) in the hope of inducing long-term remission of hyperthyroidism. It is important to distinguish those with Graves' disease from those with toxic nodular hyperthyroidism since long-term remission of the latter disease is never achieved with antithyroid drugs and definitive treatment should be planned.

The most commonly used antithyroid drug in the United Kingdom is carbimazole. It is rapidly metabolized to methimazole after ingestion, methimazole being the most commonly prescribed thionamide in the United States. Propylthiouracil represents an alternative, all of these drugs acting through inhibition of the uptake and

organification of iodine into the thyroid. Propylthiouracil has an additional effect of inhibition of peripheral deiodination of T_4 to T_3, but this is probably not clinically relevant. Carbimazole is typically prescribed at a starting dose of 20 mg per day (as a single daily dose), doses being titrated at roughly monthly intervals against measurements of free T_4. After 2 to 3 months it is usually possible to reduce carbimazole prescription to maintenance levels of 5 to 10 mg per day. An alternative is to continue carbimazole at high dose and add T_4 replacement therapy when biochemical hypothyroidism begins to ensue; there is little evidence that such an approach has any advantages in terms of rates of long-term remission in Graves' disease.

Hyperthyroidism in pregnancy is nearly always due to Graves' disease and should be managed medically if possible. The lowest dose of thionamide which restores circulating thyroid hormone concentrations to normal should be used. In most cases hyperthyroidism can be controlled with ease because of the immunosuppressive effect of pregnancy upon autoimmune diseases. High-dose thionamide therapy should be avoided since this may lead to goitrogenesis and hypothyroidism in the fetus. Propylthiouracil is often regarded as the drug of choice in pregnancy because of the possible link between carbimazole therapy and the very rare congenital abnormality of aplasia cutis. Relatively small amounts of carbimazole or propylthiouracil are excreted in breast milk so that breast feeding is not contraindicated unless high doses of thionamide are required. Congenital or neonatal hyperthyroidism is a rare complication of maternal hyperthyroidism, so obstetricians should be aware of any maternal diagnosis of Graves' disease. Neonatal hyperthyroidism reflects transplacental passage of antibodies to the TSH-receptor and is a self-limiting condition which resolves after 6 to 8 weeks when maternal immunoglobulins are destroyed. Thionamide and b-adrenergic blockade may be required in the meantime.

In general, thionamide drugs are well tolerated although may result in itchy skin rashes. The most important side-effect is induction of agranulocytosis, which if unrecognized may be fatal. This complication is rare (less than 1 in 1000) and typically, but not exclusively, occurs in the early weeks of treatment and in association with high doses. It is therefore mandatory to warn all patients treated with thionamides to stop medication immediately and seek an urgent full blood count should they develop a sore throat or any other bacterial infection. Other rare, but serious, side-effects are hepatitis and lupus-like syndromes; these are contraindications to further thionamide therapy.

Radioactive iodine

Radioactive iodine (iodine-131) is administered orally and after ingestion is rapidly concentrated in the thyroid gland. Through its action as an emitter of b-radiation it destroys functioning thyroid cells. Despite attempts to titrate doses for individual patients (according to thyroid size, measures of isotope uptake, or the degree of hyperthyroidism) in order to guarantee euthyroidism, this degree of precision cannot be achieved. Arbitrary doses of 200 to 600 MBq are therefore usually administered. Most patients with Graves' and toxic nodular hyperthyroidism are cured by one or two doses, although a maximal effect is not seen for up to 6 months after treatment. Adjunctive therapy with a thionamide is sometimes given in the meantime. Patients with toxic nodular hyperthyroidism are cured by similar doses of radioiodine to patients with Graves' disease, although higher doses are required in some patients with large ‘solitary' toxic nodules. It is essential to have a mechanism for long-term follow-up and biochemical testing in place since hypothyroidism occurs in the majority of patients treated with radioiodine and can ensue more than 20 years after radioiodine treatment.

Fears regarding long-term risks of cancer and infertility after radioiodine treatment have proved groundless so that restrictions based upon patient age and sex were withdrawn more than 10 years ago. Radioiodine can thus be administered to women of reproductive age, especially for Graves' hyperthyroidism which has relapsed after a course of medical treatment or for those with Graves' disease who do not wish to consider a full course of thionamides. Pregnancy represents an absolute contraindication to radioiodine therapy since iodine-131 crosses the placenta and may render the fetus hypothyroid; women of reproductive age are also warned to avoid conception for a period of 4 months after treatment.

The presence of Graves' ophthalmopathy represents a relative contraindication to radioiodine treatment because of evidence that radioiodine may lead to deterioration in eye disease. Patients with active ophthalmopathy should therefore be treated medically or considered for surgery. The presence of a large diffuse goitre or multinodular goitre is not a contraindication to radioiodine since thyroid size is reduced considerably by this treatment and fears regarding short-term increase in goitre size have proved groundless. Side-effects after radioiodine (apart from hypothyroidism) are uncommon, although a transient thyroiditis with soreness of the neck and short-term worsening of hyperthyroidism may occur.

Surgical treatment of hyperthyroidism

Indications for surgery are similar to those for radioiodine—relapsed Graves' hyperthyroidism, or desire for rapid cure of thyrotoxicosis. It is essential that patients treated surgically are rendered euthyroid with thionamides before surgery because of the rare but serious occurrence of thyroid storm. Preparation with b-adrenergic blocking agents alone is insufficient and preparation with Lugol's iodine and b-adrenergic blockers is no longer recommended.

Hypothyroidism is a common complication of partial thyroidectomy for hyperthyroidism, with rates being similar to those for radioiodine treatment. In those treated by partial thyroidectomy, a rise in TSH in the first 6 months after surgery may be transient and is not necessarily indicative of the need for thyroxine replacement therapy. Tests of thyroid function should be checked annually after the first 12 months to look out for late hypothyroidism. A more serious problem is recurrence of hyperthyroidism, determining that near-total or total thyroidectomy (with inevitable hypothyroidism) represent the surgical treatment of choice in some centres. Other surgical complications include hypoparathyroidism, recurrent laryngeal nerve palsy, and bleeding into the neck.

Hypothyroidism

Pathogenesis

Autoimmune (Hashimoto's) thyroiditis represents the commonest cause of thyroid failure in the United Kingdom, together with treatment for hyperthyroidism with antithyroid drugs, radioiodine, or surgery. Less common causes include therapy with drugs such as lithium or amiodarone and congenital problems of thyroid hormone synthesis or thyroid agenesis/dysgenesis. All babies born in the United Kingdom are screened for the presence of congenital hypothyroidism by measurement of serum TSH or T_4 within the first week of life so that treatment can be instituted promptly (in which case the prognosis is good). Iodine deficiency represents a major cause of hypothyroidism worldwide, deficiency *in utero* secondary to maternal iodine deficiency being accompanied by largely irreversible delay in neurological and physical development.

Clinical features of hypothyroidism (Fig. 6)



Fig. 6. An example of hypothyroidism. There is marked periorbital swelling and puffiness of the face. The patient had a moderate-sized, firm, diffuse goitre characteristic of Hashimoto's thyroiditis.

The typical clinical features of hypothyroidism are listed in Table 2. Goitre is often but not invariably present at the time of diagnosis. In Hashimoto's disease the thyroid is characteristically moderately enlarged, firm, and non-tender. Histological examination reveals a diffuse lymphocytic infiltrate which in time is replaced by fibrosis.



Table 2 Clinical features of hypothyroidism

Management of hypothyroidism

Patients with symptomatic hypothyroidism require treatment with thyroxine, which is available as tablets of 25, 50, and 100 µg. A dose of 100 to 150 µg is effective in most patients. Symptomatic improvement is usually seen in 2 to 3 weeks, though longer periods may be required for full recovery especially if hypothyroidism is biochemically severe or protracted.

Treatment should be started in the majority of patients at a full replacement dose of 100 to 150 µg per day. A more cautious approach should be adopted in the elderly or in those known to have ischaemic heart disease in whom starting doses of 25 to 50 µg per day should be considered. The aim of treatment is to restore circulating concentrations of free T_4 and TSH to the normal range. Tests of thyroid function should be checked about 6 weeks after a dose change until biochemical euthyroidism is restored. Thyroxine therapy is generally required lifelong, although short-term treatment may be indicated in the first 12 months after thyroid surgery and for treatment of the hypothyroid phase of a thyroiditis. Compliance with treatment should be checked by measurement of serum TSH every year. Dose requirements for thyroxine do not generally change, although an increase in dose may be indicated by a rise in TSH in pregnancy.

Further reading

Bartalena L *et al.* Relation between therapy for hyperthyroidism and the course of Graves' ophthalmopathy. *New England Journal of Medicine* 1998; **338**: 73–8. [Study examining effect of antithyroid treatment on ophthalmopathy.]

Franklyn JA. The management of hyperthyroidism. *New England Journal of Medicine* 1994; **330**: 1731–8. [Review of management options.]

Franklyn JA, Daykin J, Drolc Z, Farmer M, Sheppard MC. Long-term follow-up of treatment of thyrotoxicosis by three different methods. *Clinical Endocrinology (Oxford)* 1991; **34**: 71–6. [Paper defining risks of hypothyroidism.]

Lazarus JH. Hyperthyroidism. *Lancet* 1997; **349**: 339–43. [Review of disease management.]

Reinwein D *et al.* A prospective randomized trial of antithyroid drug dose in Graves' disease therapy. *Journal of Clinical Endocrinology and Metabolism* 1993; **76**: 1516–21. [Study of antithyroid drug doses.]

Roti E, Minelli R, Salvi M. Management of hyperthyroidism and hypothyroidism in the pregnant woman. *Journal of Clinical Endocrinology and Metabolism* 1996; **81**: 1679–82. [Review of management in pregnancy.]

Toft AD. Thyroxine therapy. *New England Journal of Medicine* 1994; **331**: 174–80. [Review of treatment of hypothyroidism.]

Torring O, Tallstedt L, Wallin G *et al.* Graves' hyperthyroidism: treatment with antithyroid drugs, surgery, or radioiodine—a prospective, randomized study. *Journal of Clinical Endocrinology and Metabolism* 1996; **81**: 2986–93. [Study of risk of ophthalmopathy.]

Vanderpump MPJ *et al.* The incidence of thyroid disorders in the community: a 20-year follow-up of the Whickham survey. *Clinical Endocrinology (Oxford)* 1995; **43**: 55–68. [Large study examining thyroid disease incidence.]

Weetman AP. Hypothyroidism: screening and subclinical disease. *British Medical Journal* 1997; **314**: 1175–8. [Review of the debate on treating mild thyroid disease.]

20.2.2 The thyroid gland

Gregory P. Sadler and Nicholas Dudley

Surgical anatomy
General orientations
Development of the thyroid
Blood supply
Lymphatic drainage
Important anatomical relations
Physiology
Pathophysiology
Non-toxic goitre
Toxic goitre
Special goitres
Investigation for thyroid disorders
Tests of circulating thyroid hormone
Tests of hypothalamic–pituitary function
Dynamic tests of thyroid function
Tests of thyroid dysfunction
Tests for suspected thyroid malignancy
Flow cytometry
Signs and symptoms
Symptoms
Signs
Indications for surgery
Non-toxic goitre
Thyrotoxicosis
Special goitres
Carcinoma of the thyroid
Operative surgery on the thyroid
Preoperative measures
Standard surgical approach to the thyroid
Specific thyroid surgical procedures
Postoperative complications
Haemorrhage
Recurrent laryngeal nerve damage
Superior laryngeal nerve paresis
Hypoparathyroidism
Hypothyroidism and recurrent hyperthyroidism
Tracheal collapse
Thyroid crisis
Cervical sympathetic damage
Wound complications
Further reading

Surgical anatomy

General orientations

The thyroid gland is situated in the anterior triangle of the neck, weighs approximately 20 g, and consists of two lateral lobes (right and left), joined together by a midline isthmus. The small pyramidal lobe of Lalouette, of variable size, commonly joins the isthmus at its junction with the left lateral lobe by a fibrous band or strand of muscle fibres known as the levator glandulae thyroideae. The lobes measure approx. 5 × 3 × 1.5 cm (slightly larger in women) and extend from the middle of the thyroid cartilage above to the sixth tracheal ring below. Each lobe fills the space between the trachea and oesophagus medially and the carotid sheath laterally. A strong condensation of vascular connective tissue, known as the suspensory ligament of Berry, binds the gland firmly to each side of the cricoid cartilage; it is this ligament, together with the pretracheal fascia, that splits to invest the gland, which makes the thyroid move up and down on swallowing. The fascia (false or surgical capsule) sends fibrous septa into the gland substance, dividing it into numerous lobules. Each lobule consists of 30 to 40 follicles that contain colloid; these are the main secretory and storage elements.

Development of the thyroid

The thyroid gland develops from two distinct embryological structures, the primitive pharynx and the neural crest. A median pharyngeal downgrowth migrates between the first- and the second-arch components of the tongue and descends in a caudal direction along a line extending from the foramen caecum at the back of the tongue to the pyramidal lobe of the thyroid. In so doing, the track passes ventral to the hyoid bone and then loops behind it (Fig. 1). The track usually becomes obliterated but part occasionally persists, giving rise to a thyroglossal cyst or fistula. Rarely the thyroid bud fails to descend but develops *in situ* at the back of the tongue (lingual thyroid). Conversely, it may descend too far, causing a primary mediastinal or retrosternal goitre. Even less commonly the thyroid bud may fail to divide in two and appear as one lateral lobe, the left usually being absent. The parafollicular or C cells scattered between the cuboidal epithelial cells that line the thyroid follicles are derived from the neural crest. They first migrate to the ultimobranchial bodies of the fourth and fifth branchial pouches, and then to the thyroid (Fig. 2). These are the cells that in later life have the potential to undergo hyperplastic and malignant change, resulting in calcitonin-producing medullary carcinoma of the thyroid.

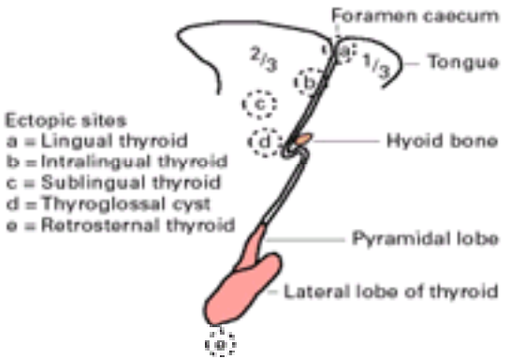


Fig. 1. Migration of the thyroid via the thyroglossal duct, and possible ectopic sites of development and duct remnants.

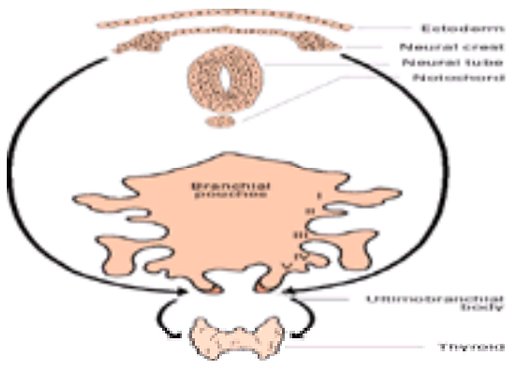


Fig. 2. Migration of the parafollicular C cells.

Blood supply (Fig. 3)

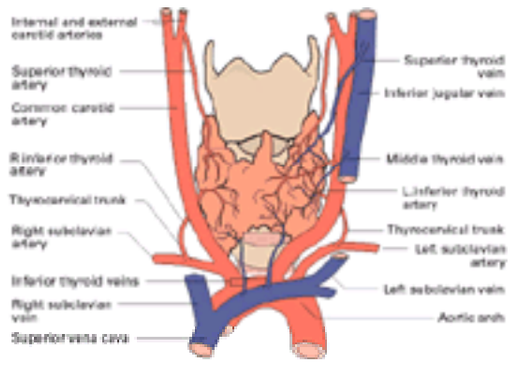


Fig. 3. Vascular supply to the thyroid.

The vascular supply of the gland is impressive and becomes more so in hyperactive thyroid states. The main supply is via two paired arteries; a third vessel occasionally supplies the lower pole of one or other lobe. The superior thyroid artery, the first branch of the external carotid, runs downward on the inferior constrictor to reach the apex of the lateral lobe, where it divides into a large anterior branch and a usually smaller, but important, posterior branch. Occasionally a tributary leaves high on the left to supply the pyramidal lobe near the midline. The inferior thyroid artery is generally much larger than the superior thyroid artery but is less constant, being absent or duplicated on one side or the other in 10 per cent of individuals. It arises from the thyrocervical trunk and passes upwards for a variable distance before looping down, running medially behind the carotid sheath to reach the posterolateral aspect of the gland at the junction of the middle and lower thirds. Numerous unnamed accessory arteries arise from the oesophagus and trachea, but the most frequently encountered is the thyroidea ima (Neubauer's artery), which courses up anteriorly on the trachea to reach the isthmus or one of the lower poles and originates from the aorta or brachiocephalic artery. In the absence of the inferior thyroid artery on one side, the thyroidea ima may be the principal source of blood supply to the lobe and therefore substantial. The named thyroid veins, although three in number like the arteries, are subject to greater variation. The superior thyroid vein, formed by a confluence of vessels from the upper pole, crosses the common carotid artery high in the neck to drain into the internal jugular. The middle thyroid vein, which overlies the inferior thyroid artery, also ends in the internal jugular vein after crossing the common carotid. The inferior thyroid veins descend from the isthmus and inferior poles of the lateral lobes to join the internal jugular or brachiocephalic veins in the anterior mediastinum and are intimately associated with the thyrothymic ligaments, which expand inferiorly as the lobes of the thymus.

Lymphatic drainage

The thyroid is generously supplied with lymphatics and a rich network ramifies throughout the gland. They drain primarily into mediastinal nodes inferiorly, tracheo-oesophageal nodes laterally, and the midline delphian nodes superiorly. Studies following injection of dye suggest that the majority of lymph from the thyroid returns to the thoracic duct without passing through the deep cervical lymph-node chain or the nodes of the posterior triangle, although these pathways may open up secondarily (Fig. 4). This has implications for the assessment of patients with carcinoma of the thyroid, who may develop lymph-node deposits outside the primary drainage areas, even on the contralateral side.

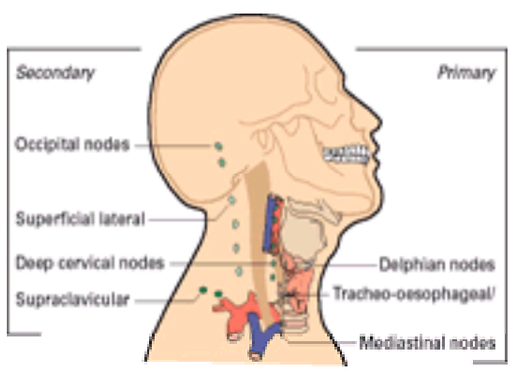


Fig. 4. Primary and secondary zones of lymphatic drainage of the thyroid.

Important anatomical relations

Recurrent laryngeal nerves (Fig. 5)

There are several structures closely related to the gland with which a surgeon must be familiar. The most important of these is the recurrent laryngeal nerve on each side, which is a branch of the vagus. The vagus, having entered the mediastinum, gives off the recurrent nerve, which returns to the neck having circled around the arch of the aorta on the left and the right subclavian artery on the right. It ascends in the tracheo-oesophageal groove and has a variable relation to the inferior thyroid artery on each side (Fig. 6). Occasionally the nerve itself divides early and branches around the artery (10 per cent of individuals). In approximately 0.25 per cent of individuals the recurrent laryngeal nerve on the right is non-recurrent and passes directly from the vagus to the cricothyroid muscles. As it takes the same course as the inferior thyroid artery, it is particularly vulnerable if its presence is unrecognized when this vessel is routinely ligated laterally. Whichever course the nerve takes, it ultimately enters the larynx posterior to the cricothyroid articulation, passing under or through Berry's ligament. The nerve supplies all the intrinsic muscles of the larynx, together with some sensory supply to the mucosa below the vocal cords. The principal effect of division of this nerve is paralysis of the vocal cord on that side.

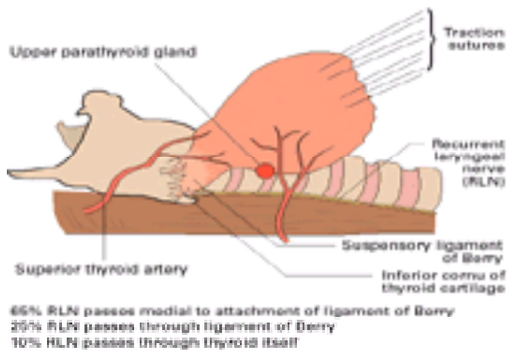


Fig. 5. Course of the recurrent laryngeal nerve in the neck.

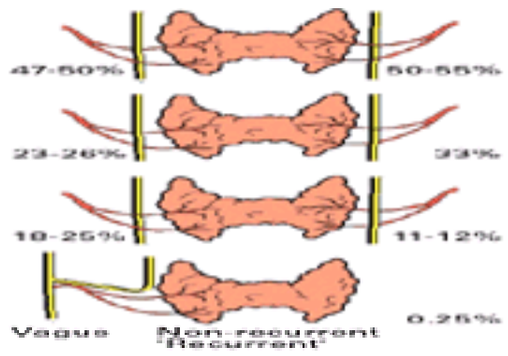


Fig. 6. Relation of the recurrent laryngeal nerve to the inferior thyroid artery (right and left).

The superior laryngeal nerve

This also arises from the vagus (inferior ganglion) and divides at the level of the hyoid bone into a large internal laryngeal nerve and a smaller external laryngeal nerve. The external runs close to the superior thyroid artery but at a deeper plane, immediately above the superior pole of the thyroid. It terminates as the nerve supply to the cricothyroid muscle, which acts as a tensor of the vocal cords on the same side.

The cervical sympathetic chain

This underlies the carotid sheath just medial to the vagus on the prevertebral fascia and is in close proximity to the inferior thyroid artery as it arches around medially.

Parathyroid glands

There are normally four parathyroid glands, the upper pair of which lies in close proximity to the dorsal aspect of the thyroid. They are usually found just above and medial to where the recurrent laryngeal nerve crosses the inferior thyroid artery, frequently tucked round behind its branches (Fig. 7). The lower parathyroid gland on each side is situated within a 2-cm radius of the lower pole of the thyroid, typically on its surface anterolaterally and at a level below and medial to where the recurrent laryngeal nerve crosses the inferior thyroid artery (Fig. 8).

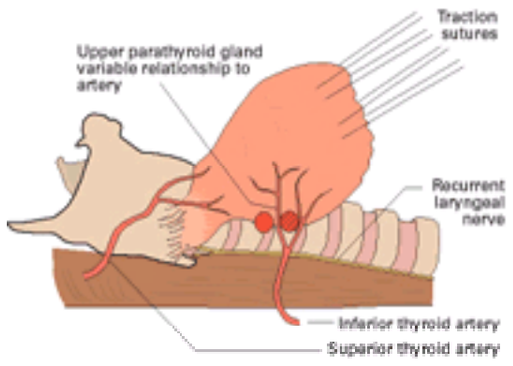


Fig. 7. Normal site of the upper parathyroid gland.

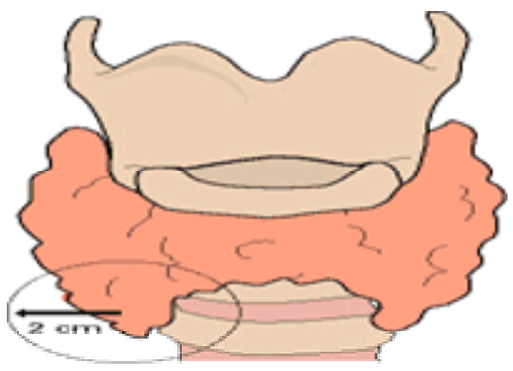


Fig. 8. Normal site of the lower parathyroid gland.

Physiology

This is discussed in detail in Chapter 20.2.1.

The role of calcitonin in normal thyroid physiology has not been established in man, but it may be involved in the regulation of plasma calcium and phosphate concentration. However, thyroidectomy, which removes nearly all the parafollicular C cells, causes no disturbance of calcium homeostasis. Medullary thyroid carcinoma, a tumour of the C cells that frequently results in gross hypercalcitoninaemia, also rarely disturbs the plasma calcium. The rise in plasma calcitonin that occurs during pregnancy and lactation appears to have no effect on the maternal skeleton, but calcium resorption may be prevented by a concomitant increase in the circulating cholecalciferol.

Pathophysiology

Nearly all disorders of the thyroid result in some swelling of the gland itself and the non-specific term ‘goitre' embraces them all. In clinical practice a working classification based on whether the gland is toxic or not and the nature of the enlargement is helpful. This enables a diagnosis to be made and appropriate action taken in the majority of patients (Fig. 9).

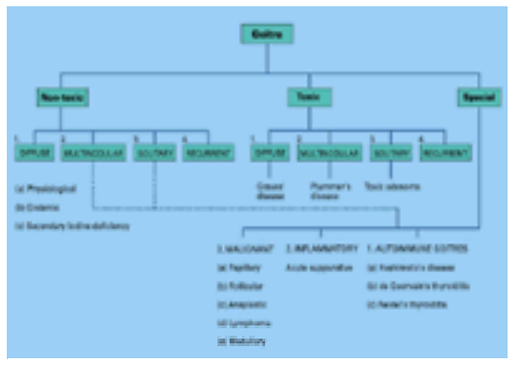


Fig. 9. Classification of goitres.

Non-toxic goitre

Diffuse/simple goitre

Physiological goitre

Enlargement of the thyroid is common during puberty and pregnancy, and at the menopause. This may be the result of increased physiological demand for thyroid hormone or as a response to growth hormone and a variation in oestrogen levels. Increased concentrations of thyroid-stimulating hormone (**TSH**) are believed to be involved in the process but are not readily detectable in euthyroid patients.

Primary iodine-deficiency/endemic goitre

The majority of ‘endemic' goitres are due to low dietary intake of iodine (less than 100 µg/day). The worldwide geographical distribution of endemic goitre corresponds closely to alpine areas where glacial action has leached iodine from the soil and carried it away to the sea: endemic goitres are rarely seen in coastal areas. The response of high-risk populations to the addition of iodine to table salt or drinking water, or to a depot injection of intramuscular lipiodol, suggests that some patients with endemic goitres also have a metabolic defect in thyroxine synthesis that only declares itself when iodine is scarce.

Secondary iodine-deficiency goitre

Dietary Some endemic goitres are due to substances in the diet (goitrogens) that interfere with the trapping of iodine or the synthesis of thyroxine. These include the thioureas in uncooked brassicas—turnips, swedes, cabbages, Tasmanian weeds (used for cow feed)—and soya beans. Excess dietary fluoride has been incriminated in the cause of goitre in the Punjab, while calcium excess has been implicated in Columbia, Cape Province, Burma, and Western China.

Drugs The drugs most commonly responsible for thyroid enlargement of this type are the thioureas, especially when overused in the treatment of thyrotoxicosis. Other drugs, including thiocyanates, iodine, *p*-aminosalicylic acid, resorcinol, lithium, and phenylbutazone, may all cause goitre if administered over a long period.

Genetic defects/dyshormonogenetic goitre In a small proportion of patients the susceptibility to goitre formation is due to an autosomal, non-sex-linked, partially recessive gene defect. Five distinct biochemical defects resulting from the inheritance of a single-gene abnormality have been described:

1. inability to concentrate iodine in the gland;
2. inability to bind iodine, which if total results in congenital hypothyroidism—in some cases congenital deafness may be associated (Pendred's syndrome);
3. inability to couple iodotyrosines to form iodothyronine (peroxidase and dehalogenase deficiency);
4. inability to retain iodine in iodotyrosines due to the lack of the deiodinase enzyme—this results in loss of iodine in its organic form in the urine;
5. abnormal protein binding of iodine in the plasma, which is then unavailable to the normal process of hormone synthesis.

Although of academic interest, it is not usually practical or helpful to investigate which of the above defects is responsible for a congenital goitre. All of the above causes of dyshormonogenesis cause compensatory hypertrophy of the gland in an attempt to maintain physiologically appropriate serum concentrations of thyroid hormone. Under the influence of increased TSH, the acini multiply in an even fashion throughout the gland, producing what is sometimes described as a parenchymatous goitre. When the process is controlled or the demand for thyroid hormone ceases, colloid may collect in previously hyperplastic acini, resulting in a colloid goitre. This process is typically patchy, causing nodularity.

Multinodular non-toxic goitre

This type of thyroid enlargement most commonly affects middle-aged women and occurs sporadically: it may represent an acquired enzyme defect due to ageing. Many multinodular goitres develop from simple goitres, especially if iodine intake or availability are compromised. The initially diffuse hyperplastic process then becomes localized to one or several areas of disorganized thyroid metabolism in which the hyperplastic acini undergo colloid involution while others show haemorrhage, cystic degeneration, or necrosis. Fibrosis and calcification may supervene: a typical multinodular goitre shows all these macroscopic features.

Solitary nodular non-toxic goitre

Fifty per cent of goitres assessed clinically to be of this type are found to be multinodular on scanning or at surgery. Truly solitary nodules may be cystic or solid. The pathogenesis of solitary cysts is uncertain, but a minority are due to degeneration of a papillary carcinoma. A solitary solid nodule may be an adenoma, a degenerative nodule, a carcinoma, or occasionally secondary to thyroiditis. Four times as many females as males are affected, with a peak incidence in middle age. Ninety per cent of solitary nodules in this age group are benign, but at the extremes of life over 50 per cent are malignant.

Recurrent nodular goitre

This may occur after surgery for multinodular goitres or solitary nodules and usually represents progression of the original underlying process. It may be modified or prevented by giving thyroxine to stop TSH release and suppress gland function.

Toxic goitre

Diffuse toxic goitre—Graves' disease/primary toxic goitre (see also Chapter 20.2.1)

The increase in size of the gland seen in patients with Graves' disease is typically due to epithelial proliferation, an increase in stromal vascularity, and lymphocytic infiltration. This may be due to a focal thyroiditis, reflecting the production of autoantibodies. Occasionally this condition progresses, with increasing fibrosis, to become Hashimoto's disease. Graves' disease is not simply the result of the action of increased TSH secondary to a primary defect in the hypothalamus or pituitary, as was once thought. Genetic predisposition and psychological trauma have a role, and seasonality of the disease may result from variation in dietary intake of iodine. In populations with a high dietary iodine intake such as in the fish-eating Japanese, Grave's disease is more common. There is now general agreement that the

hyperthyroidism and goitre of Graves' disease is caused by antibodies directed against the TSH receptor on the thyroid membrane: these increase the effect of TSH acting via the adenyl cyclase–cAMP system. The thyrotrophin-receptor antibody originally thought to be responsible for the condition was first isolated in 1956 from the serum of patients with Grave's disease as a substance that, on injection into guinea-pigs pretreated with ^{131}I, caused prolonged stimulation of radioiodine release. The nature of its action resulted in the name long-acting thyroid stimulator; this has subsequently been identified as a 7*s* IgG molecule produced by lymphocytes. Recently, however, it has become apparent that a number of different antibodies are responsible for the condition; these are all linked under the term thyroid-receptor antibodies. Thyroid-stimulating immunoglobulins or antibodies attach and stimulate the TSH receptor, whereas TSH-binding, inhibiting immunoglobulins or antibodies block the TSH receptor. There is strong evidence of a genetic predisposition associated with the *HLA-A1, -B8*, and *-DR3* haplotype in Caucasians.

Toxic multinodular goitre/Plummer's disease/secondary thyrotoxicosis

Patients with long-standing nodular goitres often develop thyrotoxicosis. Excess thyroid hormones may be produced by the nodules themselves, by the paranodular tissue, or by a combination of both. Eye signs are usually slight, and cardiac arrhythmias and cardiac failure more common than in Graves' disease; the reason for this is not known.

Solitary nodular toxic goitre/toxic adenoma/‘hot' autonomous nodule

The distinction between multinodular and solitary nodular toxic goitres may be unnecessary, but in approx. 5 per cent of thyrotoxic patients a single nodule consisting of hyperplastic epithelium and surrounded by acini in the resting phase is found. There is good correlation between the size of the nodule and the degree of hormone overproduction: overproduction of **T3** (tri-iodothyronine) but not $\mathbf{T_4}$ (tetra-iodothyronine) is common (T_3 toxicosis). Women are more likely than men to be affected with this type of goitre, and its maximum prevalence occurs between the ages of 40 and 60 years, although children may also be affected.

Recurrent nodular toxic goitre

Nodules associated with hormone overproduction may occur after thyroidectomy, suggesting that the causal factors are still prevailing.

Special goitres

Autoimmune thyroiditis

This term embraces a group of conditions that have in common the presence of circulating antithyroid antibodies; these may or may not have a causal relation to the thyroiditis. Glandular enlargement due to lymphocytic infiltration is common.

Hashimoto's disease (lymphadenoid goitre)

This characteristically affects middle-aged women: the female:male ratio is approx. 15:1. Antibodies against thyroglobulin are present in 100 per cent of patients, whereas antibodies directed against thyroid peroxidase, the membrane-bound enzyme involved in thyroid hormone synthesis, are found in about 50 per cent of patients. Previously these were termed antimitochondrial antibodies. High antibody titres are usually associated with thyroid failure. The degree of lymphocytic infiltration of the thyroid also correlates well with titres of these antibodies, especially in the fibrous variant, suggesting long-standing hyperimmunization. Microscopically the epithelial cells are enlarged and eosinophilic; these so-called Askenazy cells have been compared to hepatocytes. Lymphocytic infiltration of the stroma is intense, with the formation of lymphoid follicles. It seems probable that sensitization of these lymphocytes to thyroglobulin, mitochondria, and microsomes leads to the destructive fibrosis. A subacute variety of the condition causes transient pain and tenderness of the thyroid, but in the majority of patients the thyroid enlargement changes over time from a rubbery consistency to stony hardness, associated with progressive hypothyroidism. The risk of lymphoma and primary thyroid neoplasia developing in a Hashimoto's gland is now considered to be low. Hashimoto's disease is not associated with HLA-DR3 but is weakly associated with HLA-DR5; antithyrotrophin-receptor antibodies are sometimes present and are likely to be responsible for the thyroid hyperplasia that is commonly seen. Some of these antibodies are biologically active, causing hormone overproduction.

de Quervain's thyroiditis/subacute thyroiditis

This condition is rare in Europe but is becoming more common in North America. The slightly tender thyroid enlargement is frequently preceded by an infection of the upper respiratory tract or by a viral illness such as mumps or coxsackie virus infection, suggesting an infective background. Women in the 20- to 50-year age group are most commonly affected. Disruption of epithelial cells and extrusion of nuclei causes a pseudo-giant cell appearance. Inflammatory infiltration of the stroma by polymorphs, mononuclear cells, and lymphocytes can result in microabscesses and later fibrosis. Subsequent hypothyroidism is rare.

Riedel's thyroiditis/woody thyroid/ligneus thyroiditis

This is an extremely rare condition, and some even doubt its existence as a separate entity distinct from Hashimoto's and de Quervain's thyroiditis, both of which can produce a very hard and fibrotic gland. If the fibrosis is particularly dense and extends beyond the thyroid, tethering it to the trachea and strap muscles, the diagnosis of thyroid carcinoma must be considered. The cause is unknown but is probably one of a group of conditions, including fibrosing mediastinitis, retroperitoneal fibrosis, sclerosing cholangitis, and orbital pseudotumour, all of which are characterized by multifocal midline fibrosis.

Acute suppurative thyroiditis

This condition is now rarely seen, owing to the widespread use of antibiotics. The thyroid is usually infected by *Streptococcus pyogenes* or *Staphylococcus aureus*, originating from the bloodstream or from adjacent structures. The gland is enlarged, exquisitely tender, and there is surrounding induration. If not treated promptly with antibiotics the patient becomes acutely ill and an abscess will form. There is no disturbance of thyroid dysfunction, normal thyroxine and autoantibody concentrations being maintained.

Carcinoma of the thyroid

Malignancy of the thyroid is rare but the incidence appears to be rising: in the United Kingdom 25 cases occur per million of the population each year, with a death rate of six per million. Areas with a high incidence of endemic goitre (notably Columbia) also report a high prevalence of thyroid carcinoma, but there is no clear aetiological relation. External radiotherapy to the head and neck in children (used in the past for the treatment of acne and tonsillitis) undoubtedly increases the risk of subsequent thyroid carcinoma (notably papillary), with an average latent interval of 10 years. Radioiodine therapy in children has also been associated with the subsequent development of low-grade malignant nodules. Papillary carcinoma is the most common malignancy in individuals exposed to modest radiation doses following atomic explosions and industrial catastrophes, e.g. Chernobyl. Familial thyroid carcinomas are rare. Intake of large amounts of iodine predisposes to papillary carcinoma; and some races are more at risk. Follicular carcinoma, on the other hand is more common in areas where iodine intake is low; it appears to be TSH-induced. The existence of an association between Hashimoto's thyroiditis and thyroid neoplasia is controversial; an important prospective study has not established a high incidence of the two conditions coexisting. Malignant thyroid tumours may be classified as primary and secondary, and their prevalence, major features, and prognosis are summarized in Table 1.

Type	Prevalence (%)	Average age at presentation (years)	Spread	10-year survival
Primary				
Papillary	60	40	Via lymph nodes	80
Follicular	25	50	Via bloodstream	60
Anaplastic and differentiated	10	60	Local and distant	0
Malignant lymphoma	1	60	Local	?
Medullary	5	50	Local	50
Secondary				
Gastrointestinal illness	?	60		0

Table 1 Malignant thyroid tumours

Papillary carcinoma

Papillary carcinomas account for 80 per cent of thyroid tumours in patients under the age of 40 years, reflecting its tendency to affect teenagers and young adults with a 2:1 female:male ratio.

Typically, primary tumours are unencapsulated, pale, and rather homogeneous; cystic degeneration is not uncommon. Infiltration into lymphatics in the thyroid gland and to local cervical nodes is the usual mode of spread. Multifocality is therefore common, and, if the whole of the thyroid gland is examined, micrometastases are present in 90 per cent of affected individuals. Multiple macroscopic deposits are found at operation in 20 per cent of patients, with spread to the regional lymph nodes occurring in 50 per cent.

Histologically, tumours are characterized by a ‘papillary' pattern, resembling the fronds of a fern. Variations do, however, occur. Tumours may adopt a ‘follicular' pattern and be mistakenly diagnosed as such, or occasionally a mixed pattern (papillary/follicular structures). The diagnosis is established by characteristic cellular features. Cells are cuboidal with pale, abundant cytoplasm, crowded nuclei, nuclear ‘grooving', and intranuclear cytoplasmic inclusions (‘orphan Annie' cells). A characteristic fibrovascular stroma with calcium deposits (psammoma bodies) may be present. Tumours exhibiting papillary features, such as mixed tumours or the follicular variant of papillary carcinoma, behave biologically as papillary carcinomas.

Three macroscopic types of papillary carcinoma are currently recognized. (1) Minimal: defined as tumours of 10 mm or less in size, with no evidence of local invasion through the thyroid capsule and not associated with lymph-node metastases. Typically they are non-palpable and are thus usually incidental findings at surgical, histological, or autopsy examination. Formerly the tumours were termed occult or microcarcinomas and were classified as such if up to 15 mm in size, but this terminology has now been abandoned. (2) Intrathyroidal tumours: defined as greater than 10 mm but confined to the thyroid gland, with no evidence of extrathyroid invasion. (3) Extrathyroidal tumours: locally advanced tumours with invasion through the thyroid capsule into adjacent structures. At the microscopic level, several variants of papillary carcinoma have been described (tall cell, columnar, diffuse sclerosing, clear cell, trabecular, Hürthle cell, and poorly differentiated); these have been recognized to run a more aggressive course and are thus associated with a poorer long-term prognosis. There is a case for more aggressive surgical management in these tumours.

Follicular carcinoma

This tumour affects three times as many women as men and exhibits a tendency to present in an older age group than papillary carcinoma (mean age at presentation 50 years compared to 35 years for papillary carcinoma). It is associated with lack of iodine, and is therefore probably induced by TSH; this accounts for the fact that the tumour is more common in areas with a high incidence of endemic goitre. Macroscopically, the tumour is usually unifocal and encapsulated (90 per cent), with a pinkish cut surface that bulges.

Two types of follicular carcinoma are recognized. (1) Minimally invasive tumours: in which there is evidence of invasion into but not through the surrounding tumour capsule at one or two sites at maximum. In the past these tumours were often reported as atypical adenomas. There are, however, recorded instances of metastases in association with these tumours and they should therefore be regarded as low-grade carcinomas. (2) Frankly invasive tumours: in these invasion of vessels and/or the tumour capsule is evident. Tumour infiltration and invasion may be apparent at surgery, with tumour thrombus in the middle thyroid and/or jugular veins. Cystic degeneration in follicular carcinomas is rare.

The malignant potential of follicular carcinoma is dependent on the extent to which its capsule has been breached and the blood vessels invaded. This can only be assessed by examining the whole lobe microscopically in paraffin sections. Spread is via the bloodstream, and the most common distant sites to be affected are the lung, bones, and brain. Lymph-node involvement is unusual.

Anaplastic carcinoma

This tumour is noted for its speed of growth, local invasiveness, and early dissemination, but is fortunately rare. Its follicular-cell origin is supported by the finding of small areas of differentiated thyroid carcinoma in a proportion of anaplastic carcinomas and probably accounts for its high prevalence in endemic goitrous areas. The histological appearance is varied, with spindle-cell and giant-cell patterns. If a patient diagnosed as having a anaplastic thyroid carcinoma survives for more than 1 year the histological diagnosis should be reviewed: it will almost invariably be revised to lymphoma.

Malignant lymphoma

This occurs mainly in elderly women and grows rapidly. Its association with long-standing Hashimoto's disease is now discounted. Histologically, malignant lymphoma may be confused with small round-cell anaplastic carcinoma of the thyroid; this is resolved by monoclonal antibody studies.

Medullary thyroid carcinoma

This tumour has attracted considerable interest since it was first described in 1959, and now that clinical awareness has been aroused the diagnosis is more frequent. It arises from the parafollicular C cells, which are distributed throughout the gland but are present in highest concentration in the upper poles, where most medullary thyroid carcinomas occur. If the tumour is solitary the disease is likely to be sporadic (80 per cent); if multiple the familial form is more likely (20 per cent). The familial type of medullary carcinoma shows a predominantly autosomal pattern of inheritance and men and women are equally affected; successive generations tend to develop tumours at a progressively earlier age. An abnormality affecting chromosome 10 has been identified in individuals with the familial form and should prove helpful in identifying affected individuals in the future. Familial medullary thyroid carcinoma is also associated with tumours of the adrenal medulla (phaeochromocytoma) and parathyroid adenomas—an endocrine triad categorized as multiple endocrine neoplasia (**MEN**) type IIA, or Sipple's syndrome. A distinct group of patients has a phenotypically different disease, characterized by medullary thyroid carcinoma and phaeochromocytoma but without parathyroid disease. This syndrome, MEN type IIB, is also associated with characteristic facies, marfanoid habitus, and submucosal neurofibromas of the tongue, eyelids, and lips (Fig. 10). This type is more aggressive in its behaviour than IIA and often rapidly fatal. The C cells of both sporadic and medullary carcinoma produce a near-specific tumour marker, calcitonin, which helps in its diagnosis and management.

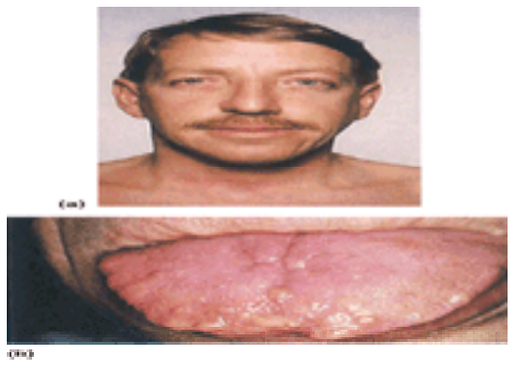


Fig. 10. Characteristic facies in MEN type IIb showing submucosal neurofibromas.

Secondary thyroid carcinoma

Very rarely breast, renal, ovarian, and gastrointestinal cancers metastasize to the thyroid, where they present as a solitary nodule. It is surprising that metastases to the thyroid are not more common, considering the high percentage of the cardiac output relative to unit volume that circulates through the very vascular gland, which must therefore be considered a protected site.

Molecular biology of thyroid tumours

Recently, attention has been focused on the role of oncogenes in the genesis of thyroid tumours. Oncogenes are cellular genes that give rise to tumour-forming genes when they are altered or when their expression is modified. The modified genes may encode for a mutated cell-surface receptor, resulting in an abnormal ‘growth signal' being transmitted to the nucleus. This may be either through overexpression of the receptor or the receptor being continually ‘activated' by a ligand. Inappropriate cell division results and may lead to malignant transformation.

Several oncogenes appear to be involved in the formation of thyroid tumours, both benign and malignant. In toxic thyroid nodules, mutations in the TSH-receptor and G-protein genes activate the adenylate cyclase–protein kinase A signal-transduction pathway and result in a well-differentiated thyroid tumour. Other defects associated with benign thyroid tumour formation are the allelic loss of genes on the short arm or chromosome 11 leading to the development of follicular adenomas, and mutations in the *ras* gene occurring in thyroid adenomas and multinodular goitres

In malignant thyroid tumours a significant role in the development of papillary thyroid cancer is the part played by the *RET* oncogene. Rearrangements of this gene have been demonstrated in papillary thyroid cancer, particularly in childhood thyroid cancers. The *RET* gene is located on chromosome 10 and its rearrangement by fusion with heterologous genes creates transforming oncogenes implicated in the pathogenesis of papillary thyroid cancer. These genes are identified as *RET/PTC1, RET/PTC2*, and *RET/PTC3*. The *RET* proto-oncogene encodes for a tyrosine kinase receptor on the cytoplasmic membrane, the receptor ligand being glial cell line-derived neurotrophic factor. Not all cases of papillary thyroid cancer express *RET* and regional differences in expression are quite marked. It is particularly associated, however, with radiation-induced thyroid cancers, *RET/PTC1* and *RET/PTC3* being expressed in two-thirds of patients undergoing thyroidectomy for cancer following the disaster in which the Chernobyl nuclear power station emitted large amounts of radioactive material into the atmosphere. Point mutations involving *RET* have also been demonstrated in patients with familial medullary thyroid cancer, MEN type 2A and 2B. It is now possible, therefore, to screen family members of patients presenting with medullary thyroid cancer in an attempt to identify those at risk for developing such cancer at an early age. Total thyroidectomy performed before that development may then be offered to affected members. This test has now replaced the pentagastrin stimulation test formerly used to identify such at-risk individuals.

In patients with anaplastic thyroid cancer a mutated *p53* gene is commonly found. The gene *p53* is a tumour-suppressor gene encoding for a protein (p53) acting as a transcriptional regulator. Mutations in the *p53* gene result in protein that lacks its primary function (regulation of genomic integrity) and a proliferation of malignant cells may ensue.

Investigation for thyroid disorders (see also Chapter 20.2.1)

There is no substitute for good history taking and careful clinical examination in the assessment of thyroid disorders. None of the available tests is infallible and misleading results may be obtained, especially if the patient is taking medication or has altered physiology, for example in pregnancy. Nevertheless, major advances in the accurate measurement of hormones and antibodies either produced by or acting upon the thyroid have resulted in the development of an impressive array of tests to help confirm the clinical diagnosis and elucidate the more difficult diagnostic problems.

Tests of circulating thyroid hormone

Serum thyroxine (T_4) (normal range 55–150 nmol/l)

This measures the total protein-bound thyroxine and, provided that the patient is not on any drug or in a condition that might affect the serum concentrations of binding proteins, offers a good basic test for thyroid function, especially when myxoedema is suspected. Low concentrations are seen in patients with the nephrotic syndrome and falsely high concentrations occur in pregnancy and in those taking oral contraceptives. Some drugs, notably salicylates, phenytoin, and phenylbutazone, compete with T_4 for protein binding, resulting in falsely low concentrations.

Serum tri-iodothyronine (T_2) (normal range 1.2–3.1 nmol/l)

Serum concentrations of T_3, are also affected by changes in the thyroid-binding proteins and are therefore subject to the same limits of interpretation. Serum T_3 falls more profoundly than T_4 in elderly people and in patients with non-thyroidal diseases. Enhanced secretion of T_3 is seen when the thyroid is deprived of iodine, making this assay unhelpful for the assessment of patients treated with drugs that block iodine uptake. The test is most useful in the confirmation of hyperthyroidism, especially when the clinical picture strongly supports a diagnosis of thyrotoxicosis but the concentrations of serum T_4 are normal and patients with solitary ‘hot' nodules—T_3 toxicosis.

Free thyroxine (normal range 8–26 pmol/l)

Since the introduction of a radioimmunoassay in kit form, it is now possible to measure the biologically active unbound circulating T_4 with precision and at relatively low cost. Drugs such as oral contraceptives or phenytoin have no effect on the results since protein binding is not a factor. Concentrations of free T_4 do not appear to vary with age in healthy individuals.

Free tri-iodothyronine (normal range 3–9 pmol/l)

The radioimmmunoassay for free T_3 has the same merit as the free T_4 assay, and is the best single test in the assessment of hyperthyroidism, provided the patient is not suffering from severe non-thyroidal illness. It is not helpful in the diagnosis of hypothyroidism.

Tests of hypothalamic–pituitary function

Thyroid-stimulating hormone (normal range 0.5–5.0 mu/l)

Routine assays for TSH lack sensitivity and cannot identify patients who may have subclinical hypothyroidism and might benefit from replacement therapy. New, sensitive, immunoradiometric assays probably represent the most helpful confirmatory test for both hypothyroidism when concentrations are high and hyperthyroidism when concentrations of TSH are undetectable. In severely ill patients or those in early pregnancy, however, falsely low TSH concentrations may be recorded, and hyperthyroidism can be missed.

Thyrotrophin-releasing hormone test

Following the intravenous administration of 200 mg of thyrotrophin the patient's TSH concentrations are then measured in blood samples taken after 0, 20, and 60 min. An exaggerated response producing concentrations greater than 20 mU/l at 20 min is seen in patients with hypothyroidism. Little or no response (less than 1.8 mU/l) is seen in hyperthyroid patients. The test is being superseded by the immunoradiometric assay for TSH where this is available.

Dynamic tests of thyroid function

Radioisotope uptake

Uptake measurements combined with thyroid scanning provide information about thyroid activity and also show graphic information about the size and extent of the gland. This method is particularly helpful in showing the retrosternal extent of the gland (Fig. 11) and identifying which part is hyperactive (hot nodules) (Fig. 12). In North America, where radioiodine therapy is more commonly employed for the treatment of thyrotoxicosis, a radioisotope uptake study may identify clinically thyrotoxic patients with silent thyroiditis but low radioiodine uptake. Ablative therapy should be avoided in such patients.

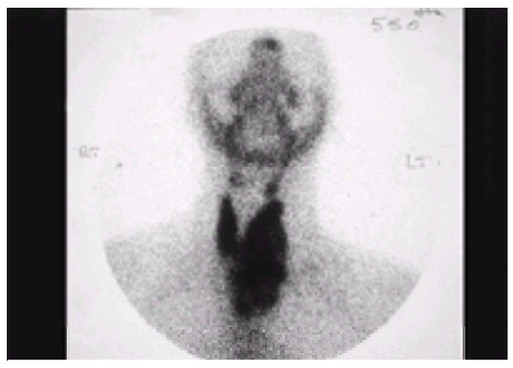


Fig. 11. ^{131}I scintiscan showing retrosternal extension of a goitre.

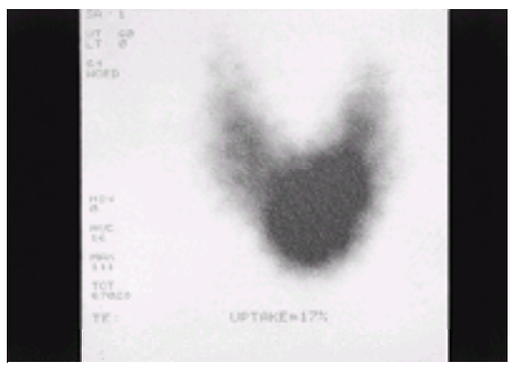


Fig. 12. ^{131}I scintiscan showing a solitary hyperactive ‘hot' nodule.

Triiodothyronine suppression test

T_3 administration normally suppresses pituitary TSH secretion and therefore reduces radioiodine uptake in normal patients. Radioiodine uptake is not so reduced in patients with Graves' disease, in whom the thyroid is being stimulated by antibodies rather than by TSH. The test is useful in distinguishing high radioiodine uptake associated with iodine deficiency from that of hyperthyroidism, but can precipitate cardiac failure in elderly individuals. It has largely been superseded by the thyrotrophic-releasing hormone test or immunoradiometric TSH assay.

Tests of thyroid dysfunction

Antithyroid antibodies (antithyroglobulin and antimicrosomal antibodies)

Very high titres of antithyroglobulin antibodies are seen in patients with Hashimoto's thyroiditis, especially in those with long-standing disease, when their presence strongly supports the diagnosis. Surgery should not usually be undertaken in patients with Hashimoto's thyrotoxicosis. The presence of antithyroglobulin antibodies in a hyperthyroid patient without eye signs suggests the diagnosis of Graves' disease. The presence of antimicrosomal antibodies indicates autoimmune thyroid disease of Hashimoto's type, and when found in patients who have already received thyroxine makes a presumptive initial diagnosis of thyroid failure more likely. This antibody is also seen in patients with thyroid malignancy.

Erythrocyte sedimentation rate (ESR)

The ESR is often raised (over 90 mm/h) in many patients with myxoedema and Hashimoto's thyroiditis, and is markedly elevated in de Quervain's thyroiditis.

Tests for suspected thyroid malignancy

Scintigraphy

^{123}I has a half-life of 13 h and is a g-ray emitter; it gives a low radiation dose and is the isotope of choice. Areas of high uptake (hot nodules) are almost invariably benign; however, more than 50 per cent of nodules have low uptake rates (cold nodules) but only 10 per cent are subsequently shown to be malignant. The test therefore discriminates poorly between benign and potentially malignant nodules and is thus hard to justify. It has now been replaced by fine-needle aspiration in the diagnosis of thyroid cancer. The principal benefits of isotope scanning are in confirming the presence of a ‘hot/toxic' nodule in the thyroid gland as a thyrotoxic patient, and in identifying metastases or residual local disease after total thyroidectomy for carcinoma (Fig. 13).

Fig. 13. Total-body scintiscan showing metastases from follicular carcinoma post-thyroidectomy.

Ultrasonographic scanning

In patients who have a solitary thyroid nodule, the incidence of thyroid cancer in the nodule is about 15 per cent. This incidence falls to 5 per cent when the nodule is ‘dominant' in a multinodular gland. Ultrasound may therefore be helpful in identifying impalpable thyroid nodules, thus decreasing the likelihood of malignancy. There are, however, no published reports in which ultrasonography has been able to distinguish reliably benign from malignant nodules. Its use, therefore, in the management of thyroid disease is limited and is no substitute for histological examination of the nodule after fine-needle aspiration.

Fine-needle aspiration or aspiration biopsy cytology

This technique, promoted by the Karolinska Institute in Sweden for over 40 years, has now gained acceptance as the procedure of choice in evaluating thyroid nodules, having replaced radionuclide and ultrasonographic scanning. A 25G disposable needle on a 10-ml syringe enables cells to be aspirated from any suspicious area of the thyroid (Fig. 14). The procedure can be performed quickly and painlessly in the outpatient department without the need for local anaesthetic, providing a specimen that

is smeared, fixed, and stained on a microscope slide.

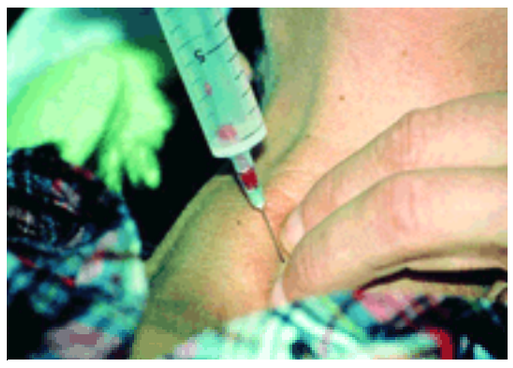


Fig. 14. Fine-needle aspiration from a thyroid nodule.

In the hands of an experienced cytopathologist, 90 per cent of nodules can be categorized into the following groups: benign 65 per cent, suspicious 15 per cent, malignant 5 per cent, and non-diagnostic 15 per cent. False-positive results should be about 1 per cent and false-negative about 5 per cent.

Though fine-needle aspiration cytology is the preferred investigation of choice in patients with thyroid nodules it does have some limitations in the diagnosis of follicular or Hürthle-cell neoplasms. It is unable to demonstrate whether capsular or vascular invasion is present, the criteria needed to establish malignancy. Thus, when follicular cells are demonstrated on fine-needle aspiration from a thyroid nodule, 20 per cent will be malignant and 80 per cent benign; thyroid lobectomy to obtain definitive histological diagnosis is recommended in these patients. The needle technique is also less reliable in patients who have had irradiation to the head and neck or have a family history of thyroid cancer, as many of these patients have multifocal lesions.

Serum calcitonin (normal range <0.08 µg/l)

Elevated basal concentrations of this hormone produced by the parafollicular C cells are diagnostic of medullary thyroid carcinoma; serial measurements are also useful in predicting complete removal of the tumour or recurrence. Patients with MEN type II who have occult or premalignant medullary thyroid carcinoma (C-cell hyperplasia) can be identified by a provocation test using a combination of pentagastrin (0.5 µg/kg body wt) and calcium gluconate (2.5 mg/kg body wt) infused intravenously over 30 s. Blood samples are drawn for calcitonin estimation before the infusion and at 2, 5, 10, and 15 min after in a fasted individual. The test is safe and can be carried out in the outpatient department.

Serum thyroglobulin (normal range <1–35 µg/l)

In patients who have undergone total thyroidectomy with or without radioiodine therapy for differentiated thyroid carcinoma, concentrations of thyroglobulin above 50 µg/l indicate probable residual or recurrent tumour. Concentrations above 100 µg/l strongly suggest the presence of pulmonary or skeletal metastases.

Flow cytometry

Measuring the nuclear DNA content by flow cytometry offers the prospect of identifying diploid tumours, which have a good prognosis, and aneuploid tumours, which have a poor prognosis. Though data accumulated from patients with papillary tumours correlate well with predicted outcome, the technique has not been widely employed by surgical centres throughout the world.

Signs and symptoms

Symptoms

There are two broad categories of symptoms related to thyroid disease: those occurring as a result of the enlargement of the gland itself and those related to its disordered endocrine activity. The history will establish whether one or both classes of symptoms are present, and examination then aims to elicit the relevant physical signs.

Neck symptoms

A lump in the neck

Nearly all goitres grow slowly and are painless; the patient only visits the clinician when the mass becomes a cosmetic problem. A rapid change in size of all or part of the thyroid may be caused by haemorrhage into a necrotic nodule, a fast-growing carcinoma, or by one of the varieties of thyroiditis. When haemorrhage is the cause of a lump its appearance is invariably sudden (within 24 h) and it is usually painful. Malignant tumours only cause pain when local structures are involved. This occurs within weeks of development of an anaplastic carcinoma or within months or years for a papillary or follicular tumour.

Discomfort on swallowing

Thyroid swellings make swallowing uncomfortable but rarely obstruct the oesophagus, which easily moves out of the way or becomes stretched. Thyroiditis causes extreme pain on swallowing, radiating up into the jaw and ears. True dysphagia as a result of thyroid disease is a sinister symptom indicative of aggressive anaplastic carcinoma.

Dyspnoea

This most commonly arises as a result of deviation and compression of the trachea caused by asymmetrical thyroid enlargement. It may be accompanied by stridor when the patient has the head flexed forward or laterally to the side of the enlarged lobe.

Hoarseness

This symptom in association with a goitre is sinister and until proven otherwise should be considered secondary to a malignant process of the thyroid involving one of the recurrent laryngeal nerves. The hoarseness seen in hypothyroid patients derives more from mucus on the vocal cords than an inability to phonate.

Symptoms of hyper- and hypothyroidism

All bodily systems can be affected by altered thyroid activity. In hyperthyroidism, increased metabolism is usually reflected in heat intolerance with a preference for cold weather coupled with increased sweating. The principal symptoms and signs of hyperthyroidism are listed in Table 1, while in many instances the clinical features of hypothyroidism are the exact opposite (see Table 2).

Antithyroid drugs	*Cautions* [13] *Y*	*Surgery*
Requires compliant and highly motivated patient to cope with variable drug regimen	Prospect of radiation therapy may be unacceptable to the patient	Prospect of neck surgery may be unacceptable to the patient
Needs close clinical and laboratory supervision over 18–24 months	Uncertainty about long-term carcinogenic effects of low radiation doses, known to be more oncogenic than high doses	Unsuited to patients of advanced age or poor medical status
Danger of overtreatment leading to hypothyroidism and thyroid enlargement	Uncertainty about teratogenic effects which obstructs reproduction for (at least) 2–3 months after therapy	Scar inevitable with possibility of keloid formation
Risk of side-effects from drugs: rashes, itchiness, and potentially dangerous marrow depression	Slow response to therapy (3–6 months): especially large multinodular goitres	Results are highly operator-dependent
Poor response with large multinodular goitres	Not appropriate for cosmetically unacceptable toxic goitres	Small but definite risk of serious complications in skilled hands: permanent recurrent laryngeal nerve damage 0.1%; permanent hypoparathyroidism 0.5%
Not appropriate for cosmetically unacceptable toxic goitres High relapse rate	Cumulative (high) hypothyroidism rate requiring long-term follow-up	Modest rate of hypothyroidism

Table 2 Disadvantages of antithyroid drugs, radioactive iodine, and surgery in the management of hyperthyroidism

Signs

Since examination traditionally follows functional enquiry the clinician will usually have a fairly strong impression of the patient's thyroid status and will know whether attention needs to be directed primarily towards the thyroid itself. In those with thyroid dysfunction, more specific signs need to be elicited.

Overall appearance

Euthyroid patients appear normal, apart from possible signs in the neck. Hyperthyroid patients are frequently noisy, agitated, and clearly nervous, but may be apathetic. Hypothyroid patients show little response to their environment, being subdued, slow in their movements, and frequently overweight. They typically have thickened and expressionless features, and a ‘peaches and cream' complexion (Fig. 15).



Fig. 15. Characteristic facies in myxoedema.

Examination of the hands

The hands provide a wealth of clues about a patient, and this is certainly true in thyroid disease. The hands of a hyperthyroid patient usually show a fine tremor, which is accentuated by placing a sheet of paper on the back of the hand with the arms outstretched. Patchy depigmentation of the skin (vitiligo) surrounded by areas of increased pigmentation may be apparent (Fig. 16). Patchy subperiosteal deposition of new bone, simulating clubbing (thyroid acropathy) can occur. On shaking hands with the patient one is aware of increased temperature and sweating. Pulses at the wrist and at the bases of the fingers are easily felt, due to increased systolic pressure; this is often accompanied by tachycardia greater than 80/min and irregularity.



Fig. 16. Vitiligo in thyrotoxicosis.

The hypothyroid patient has cool, dry hands, and the palms are often pale yellow in colour. The increased bulk of the subcutaneous tissues cause the hands to look puffy and spade-like, and rings become tight. The nails are often fissured and cracked, becoming separated from the nailbed (onycholysis or Plummer nails). Bradycardia and a low-volume pulse, difficult to detect, are characteristic of hypothyroidism.

Examination of the thyroid

Inspection

Confirmation that a swelling in the neck is arising from the thyroid is based on two observations: the site is anatomically correct and the swelling moves up and down on swallowing. Offering the patient a drink of water is helpful at this stage and subsequently aids more detailed examination by palpation. Much can be learned by simple inspection with the patient's neck properly exposed and with a good light source on one side. Prominent neck veins may indicate superior vena caval obstruction secondary to retrosternal extension of the goitre. Puckering of the skin, accentuated on swallowing, is seen with locally invasive thyroid cancer. Erythema of the skin is often present in patients with de Quervain's thyroiditis or suppurative thyroiditis. Deviation of the thyroid cartilage suggests asymmetrical enlargement of the thyroid.

Palpation

Palpation should be undertaken with the patient sitting comfortably in an upright chair with all clothes removed from around the neck. With the clinician standing behind the patient with thumbs placed on the patient's occiput to flex the head forward slightly, the clinician's hands encircle the neck and the right hand draws the sternomastoid muscle on the same side laterally whilst the fingertips of the other hand scan the surface of the underlying lobe (Fig. 17). The position of the hands is then reversed to evaluate the left lobe. The normal trachea should be in the midline of the suprasternal notch and not obscured by the thyroid isthmus. If there is any doubt about the relation between anatomical features and any lump, or difficulty feeling below the lower pole of the lobe, the patient is asked to swallow. This will confirm whether the lump is part of the thyroid and if retrosternal extension is present. If the head and neck becomes suffused, with or without the onset of stridor, when the patient raises his or her arms above the head (Fig. 18), retrosternal extension is confirmed. In a patient with thyrotoxicosis, careful palpation may detect a thrill. The overall neck circumference and precise measurements of any nodules should be recorded to provide a baseline if longer-term surveillance is indicated. The lymph

nodes draining the thyroid should also be checked for any enlargement (see Fig. 4).



Fig. 17. Position of the hands for palpation of the thyroid—Lahey's grip.



Fig. 18. Demonstration of superior caval obstruction, due to retrosternal extension of a goitre, accentuated by raising the arms.

Percussion

Testing for retrosternal extension of the thyroid by this method is considered to be of no value.

Ausculatation

A systolic bruit may be heard in Graves' disease and is an important sign of toxicity. It should not be confused with a transmitted systolic murmur secondary to aortic or pulmonary stenosis or from a venous hum that can be eliminated by pressure on the internal jugular vein at the level of the hyoid bone.

Special signs in hypothyroidism

The skin is characteristically dry and flaking, with hyperkeratosis over the flexures. Coarsening of features and subcutaneous thickening, notably across the back of the neck and hands, are due to deposition of hygroscopic hyaluronic acid. This mimics oedema but does not pit on pressure. The hair is coarse, brittle, and falls out easily; loss of the outer one-third of the eyebrows is an unreliable sign of hypothyroidism, and is often seen in normal perimenopausal women. Macroglossia is often present and impairs speech. The reflexes are sluggish and their relaxation period is prolonged.

Special signs in hyperthyroidism

There are two main groups of physical signs, other than those affecting the thyroid: those confined to the eyes, and more general signs.

Eye signs

Swelling of the eyelids without exophthalmos is an indicator of dangerous congestive ophthalmopathy. Conjunctivitis and chemosis (Fig. 19) may be induced by a reduction in blinking rate and by an inability fully to close the eyelids whilst asleep. The conjunctivae are inflamed and prominent vessels are seen, especially at the lateral canthi. When oedema occurs (chemosis) the conjunctiva becomes thickened and wrinkled, and may actually prolapse between the lids. Excessive watering and photophobia occur.



Fig. 19. Conjunctivitis and chemosis in hyperthyroidism.

Exophthalmos is caused by the eyeball being pushed forward by an increase in retro-orbital contents. In Graves' disease this is due to an increase in fat and bulkiness of the extrinsic muscles secondary to deposition of water and mucopolysaccharides combined with lymphocytic infiltration. The eyes may be affected asymmetrically. The single most crucial observation relates to the appearance of the white sclera below the inferior limbus of the cornea on forward gaze (Fig. 20). Accurate measurement of the degree of exophthalmos requires an exophthalmometer: an absolute reading of 18 mm or more or a difference of 2 mm between the eyes is considered significant. Looking down directly over the patient's forehead and observing the cornea in front of the supraorbital ridges is an easy bedside means of assessment but only reveals gross exophthalmos. The same applies when the patient can look upwards without wrinkling the forehead.

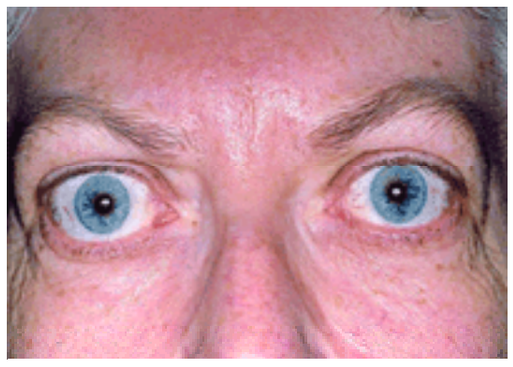


Fig. 20. Appearance of the eyes in a patient with exophthalmos, showing sclera visible beneath the iris.

Lid retraction is the result of spasm of the levator palpebrae superioris muscles, which causes the white sclera to appear between the upper lid and upper limbus of the cornea in all positions of gaze (Fig. 21). It prevents apposition of the eyelids during sleep, and is a cause of conjunctival irritation and ulceration. This same spasm allows the sclera to become visible between the upper lid and cornea (lid lag) as a patient's eye follows the examiner's finger downwards (ideally an arm's length away from the patient's head). This test of lid lag is poorly reproducible and unreliable.

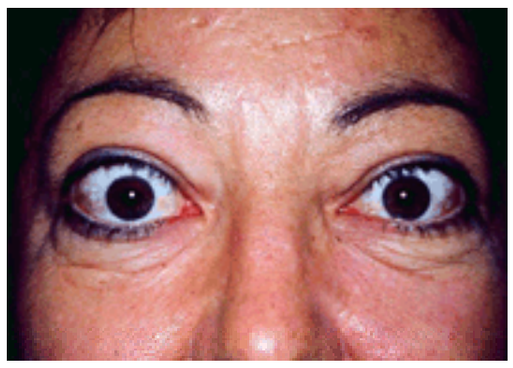


Fig. 21. Appearance of the eyes in a patient with lid retraction, showing sclera visible above the iris.

Disturbance of vision occurs as a result of oedema and lympho-cytic infiltration affecting the extrinsic muscles of the eye, notably the superior rectus and inferior oblique. Diplopia commonly occurs on upward outward conjugate gaze, but upward inward gaze may also be affected. Both of these movements should be checked by asking the patient to follow the examiner's finger appropriately.

General signs

1. Proximal myopathy, producing wasting of the muscles of the shoulder girdle and pelvic girdle, may result in difficulty getting up from a chair or inability to carry a heavy shopping bag.
2. Hyper-reflexia.
3. Fine tremor of the tongue may also be present.
4. Pretibial myxoedema is seen in approx. 5 per cent of patients with Graves' disease, especially those with eye signs, and is due to deposition of mucopolysaccharides and hyaluronic acid. The skin is thickened, livid, with *peau d'orange* (orange peel) appearance, typically extending over the front of the shins and occasionally over the dorsum of the feet (Fig. 22).

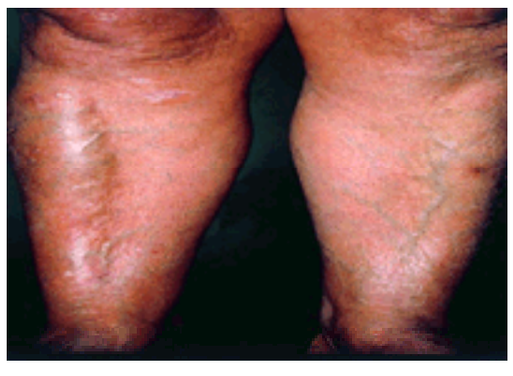


Fig. 22. Pretibial myxoedema seen in Graves' disease.

5. Splenomegaly is rare, but may be associated with generalized lymphadenopathy.

Indications for surgery

Non-toxic goitre

Diffuse non-toxic goitre

The incidental finding of a small, soft goitre, especially if ‘physiological' (see under Pathophysiology above), requires no treatment. Goitres occurring as a result of primary or secondary iodine deficiency rarely present until the condition is well established, but those detected at a small size may regress in response to iodine or eradication of the offending goitrogen in the diet. This process may be accelerated by giving thyroxine, which depresses the TSH drive to the gland. If there is no response, partial thyroidectomy may be indicated on cosmetic grounds.

Multinodular non-toxic goitre

Large goitres of this type may well cause cosmetic disfigurement, pressure symptoms, and discomfort. All of these are good indications for surgery, the extent of which is dependent on the size and number of nodules present. Following surgery a small dose of thyroxine (0.1 mg daily) reduces the risk of recurrence in the residual thyroid remnant.

Solitary nodular non-toxic goitre

Because of the high prevalence of solitary nodules (and even though over 50 per cent of those thought to be solitary are in fact part of a multinodular process), a policy of removing all of them surgically, although sound, is simply not practical. Fine-needle biopsy (aspiration biopsy cytology; described above) has helped to reduce the incidence of surgical exploration of thyroid nodules by 25 per cent. It allows benign and malignant lesions to be identified with a high degree of specificity and sensitivity, and has the merit of being quick, painless, and easily performed as an outpatient procedure. All biopsy techniques are limited by the skill and ability of the

operator to obtain a truly representative sample and by the experience and accuracy of the cytopathologist examining the specimen. The only serious limitation of this technique is failure to differentiate between a follicular adenoma and a follicular carcinoma: this can be done only by studying the overall histological pattern and, in particular, the presence or otherwise of capsular and vascular invasion. Unless the surgeon can depend on a cytology service that has a very low false-negative rate, total lobectomy (see Fig. 23(a)) is advocated for all cases, especially in children and men under 40 years of age. Fine-needle biopsy quickly identifies a cystic nodule or simple cyst and, if less than 2 cm in diameter, drainage may be curative. Ideally the cyst wall should also be biopsied to ensure that a papillary carcinoma is not missed. Recently, the tendency has moved to total thyroidectomy for glands that have widespread disease.

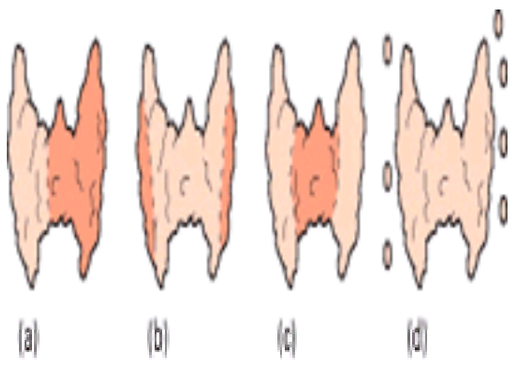


Fig. 23. Schematic representation of thyroid operative procedures: (a) thyroid lobectomy; (b) subtotal/near-total thyroidectomy; (c) isthmusectomy; (d) total thyroidectomy and node clearance.

Thyrotoxicosis

There are three methods of treatment for this condition: antithyroid drugs, radioiodine, and surgery (discussed in more detail in Chapter 20.2.1). Over 50 years, fashions have waxed and waned but a marked preference for radioiodine persists in North America, while its use is more selective in the United Kingdom. The different treatments have their own advantages and disadvantages (Table 2). Ultimately, it is for the patient to decide which treatment is most acceptable.

Surgery is most usually appropriate for the following conditions.

Large toxic multinodular goitre/Plummer's disease

The thyroid function of patients with large toxic multinodular goitre or classical Graves' disease is often very labile during treatment with antithyroid drugs, and relapse rates following withdrawal of medication are greater than 50 per cent. Unfortunately, there is no reliable test to identify patients likely to relapse, although attempts to do so have been made on the basis of HLA status and concentrations of thyroid-stimulating antibody. Radioactive iodine treatment is least suited for Plummer's disease as drug uptake is variable, there is minimal impact on the size of the gland, and response is slow. Surgery is best carried out sooner rather than later to avoid compromise of the airway and the cardiovascular system. In the past, surgical procedure has been near-total thyroidectomy, with the removal of at least 85 per cent of the gland (see Fig. 23(b)), notably those parts shown to be ‘active' on radioiodine scanning. Recently, however, it has been the authors' tendency to favour total thyroidectomy in these patients.

Large diffuse toxic goitre (Graves' disease)

Patients affected with this condition are likely to have florid disease and cosmetic embarrassment. Prompt elective surgery should be performed after a course of treatment with antithyroid drugs and b-blockers. It is more debatable whether surgery should be undertaken when the gland is small, since the risks of hypothyroidism are greater than when a bulkier gland is the target.

In women, however, especially those in their reproductive years who have relapsed after 12 to 18 months' treatment with antithyroid drugs, or in those in whom medication has proved impractical due to poor compliance or unacceptable side-effects, surgery is the first choice, providing it is acceptable to the patient and the necessary surgical skills are available. A subtotal thyroidectomy should be performed, which aims to leave one-eighth of the total mass (approximately 4 g) of thyroid tissue on each side.

In patients with thyrotoxic eye disease, florid thyrotoxicosis, or where medical treatment has proved difficult to control the disease, the authors have favoured total thyroidectomy and thyroxine replacement therapy as the treatment of choice.

Toxic solitary nodular goitre/toxic adenoma/autonomous ‘hot nodule'

Surgical resection of the physiologically hyperactive part of a thyroid gland (demonstrable on radioiodine uptake scan) is so straightforward that the other methods of treatment need not be seriously considered unless the overall medical state of the patient makes surgery inadvisable. Antithyroid drugs have the disadvantage of having to be taken for life, and radioiodine therapy runs the risk of exposing surrounding normal tissue to irradiation. In addition, the nodule usually persists. The operative procedure indicated is subtotal lobectomy or isthmusectomy (see Fig. 23(c)), depending on the location of the active nodule.

Childhood Graves' disease

Children with hyperthyroidism are prone to relapse after withdrawal of medical therapy and radioiodine therapy is contraindicated owing to the high incidence of benign and malignant nodules that arise as a result of exposure to radiation. Unless medical therapy is effective quickly, surgery is the treatment of choice. Hypertrophy of thyroid remnants and recurrent thyrotoxicosis is common in children, and surgery therefore needs to be more radical. A surgeon experienced in operating on the thyroid gland in children can perform the operation to avoid high rates of myxoedema and complications. Long-term follow-up is essential.

Thyrotoxicosis in pregnancy

This is rare, affecting 0.2 per cent of pregnancies, but can present a difficult management problem. Antithyroid drugs given to the mother cross the placental barrier and may cause hypothyroidism and goitre in the fetus (iatrogenic cretinism). Excretion of antithyroid drugs in the milk is at a concentration too low to cause comparable effects. Radioiodine is never given in view of its possible teratogenic effects, especially in the first trimester or at the time when the fetal thyroid is trapping iodine. If maternal thyrotoxicosis is moderate or severe, and especially if cardiac symptoms herald possible cardiac failure late in pregnancy, elective subtotal thyroidectomy should be undertaken in the middle trimester, when it is safe and effective.

Special goitres

Hashimoto's disease/lymphadenoid goitre

A small goitre of this type in a euthyroid patient may require no treatment. As the goitre enlarges symptoms of pressure on the trachea and oesophagus are common; these may be relieved by exogenous thyroxine (0.1–0.2 mg/day) and prednisolone (20 mg/day). If there is still no response, removal of the affected lobe or isthmus is indicated. Such goitres are relatively avascular and easily separable from the false capsule. Total lobectomy or total thyroidectomy is indicated in any patient with Hashimoto's disease in whom malignancy is suspected on clinical or cytological grounds.

De Quervain's thyroiditis

The acute pain associated with this condition can be alleviated by aspirin and a b-blocker. If this fails, 20 to 30 mg of prednisolone daily in divided doses for 1 month usually provides rapid relief. Thyroxine replacement may be required if hypothyroidism develops, but this is usually transient and treatment should be stopped after 6 months. Surgery is only required if the diagnosis is uncertain or equivocal on fine-needle biopsy.

Riedel's thyroiditis

Surgery is almost invariably required to exclude malignancy and to relieve the severe constriction of the airway. Relief can be obtained by dividing the thyroid isthmus, which is hard and brittle and can literally be snapped off the trachea with the fingers. Hypothyroidism frequently arises as the fibrosis progresses and needs to be treated indefinitely with thyroxine.

Thyroid cyst

Cysts less than 2 cm in diameter are treated by aspiration: the fluid should be examined cytologically, and the patient's condition monitored. Larger cysts are likely to recur and may well be cosmetically unacceptable. These carry a risk of sudden airway obstruction due to intracystic haemorrhage and are, therefore, best dealt with electively by thyroid lobectomy or isthmusectomy, according to site.

Carcinoma of the thyroid

The indications for, and extent of, surgery in carcinoma of the thyroid depend primarily on the histological type, but the age of onset and the mode of presentation will also influence management.

Papillary carcinoma

There is controversy about the extent of surgery required in patients with unifocal tumours with no macroscopic lymph-node involvement. In 1987, Dr Hay at the Mayo Clinic reviewed their experience of 860 patients with papillary thyroid cancer and devised a prognostic scoring system for the risk of dying from that cancer. Factors taken into account were *A*ge, pathological tumour *G*rade, *E*xtent of disease, and *S*ize of tumour (the AGES system). Two distinct groups of patients were identified: 85 per cent were categorized as low risk (young patients, well-differentiated tumours, no metastases, and small primary lesions) and 15 per cent as high risk (older patient, poorly differentiated tumours, local invasion, distant metastases, and large primary lesions). Of the two groups the cancer-risk mortality at 25 years in the low-risk group was 2 per cent, total or near total thyroidectomy conferring no survival advantage over total lobectomy in this group. Subtotal lobectomy, however, was associated with a poorer survival. The cancer-risk mortality in the high-risk group was 46 per cent. In this group, however, total or near total lobectomy conferred a very significant survival advantage over just thyroid lobectomy. In 1993, Dr Hay and colleagues published a further refinement of their scoring system, the MACIS scale. The factors assessed are: distant *M*etastases, *A*ge at presentation (under or over 40 years), *C*ompleteness of original surgical resection, extrathyroidal *I*nvasion, and *S*ize of original lesion (cm). Again, two groups of patients are categorized, low risk (score below 4) and high risk (score above 4).

The minimum conservative approach, therefore, for patients with a solitary macroscopic tumour with no lymph-node disease, categorized as being ‘low risk', is a total lobectomy on the side of the tumour, isthmusectomy, and removal of the pyramidal lobe. The prognosis for the majority of these patients is excellent, especially if the TSH is then permanently suppressed below 0.1 mU/l by thyroxine (usually 0.1 mg/day). Total thyroidectomy is reserved for patients categorized as ‘high risk', i.e., those with macroscopic multifocal disease, vascular invasion, extrathyroidal invasion, and/or demonstrable lymph-node disease. Whether total thyroidectomy provides advantage over near-total thyroidectomy (see Fig. 23(d)) has not been established.

Proponents of total thyroidectomy argue the advantages of this approach are that multifocal tumour throughout the gland is always removed, the risk of anaplastic change in residual foci is excluded, and postoperative surveillance may be possible through serial thyroglobulin estimation. Undoubtedly, however, the greatest advantage of total thyroidectomy is in facilitating the use of radioiodine screening for micrometastases of thyroid cancer by the removal of all thyroid tissue in the neck. Clearly the risk of damage to the recurrent laryngeal nerve and of permanent hypoparathyroidism are greater than with near-total or subtotal thyroidectomy. Surgery of this type should therefore be embarked on by trained thyroid surgeons with a specialist interest in thyroid cancer.

Surgery for papillary carcinoma should always include clearance of nodes in the primary lymphatic drainage zone. The thymus need not be disturbed unless obviously involved in the disease process, as ectopic parathyroids are frequently located in the thyrothymic ligament. Disease affecting lymph nodes in the secondary drainage areas may be suspected by their blue/black discoloration and often fleshy or cystic appearance. Clearance of all such nodes should be by a modified neck dissection, only sacrificing the internal jugular vein and sternomastoid muscle if these are directly involved. The diseased nodes rarely progress above the level of the hyoid bone: the classical radical block dissection has been shown to confer no benefit in terms of reduced tumour recurrence rates or improved survival and leaves the patient disfigured and at risk of lymphoedema of the face. When disease recurs in lymph nodes, often years later (fewer than 10 per cent of cases treated by total lobectomy), the nodes are well encapsulated and can easily be excised ‘berry picking'. Even patients treated in this way for successive recurrences do not appear to have a compromised survival. When local invasion or distant spread not amenable to surgery occur, radioiodine may show uptake in approx. 20 per cent of papillary metastases and then the option of therapeutic radioiodine can be considered.

Follicular carcinoma

This tumour typically presents as a non-toxic, solitary nodule. Distant metastases rarely occur without a clinically impressive primary lesion. Most cases will be suspected on fine-needle biopsy and require surgery in the form of thyroid lobectomy. If there are unequivocal malignant features at operation such as extrathyroidal spread of tumour, total thyroidectomy is the treatment of choice. Where extrathyroidal invasion is not present, frozen-section histology of the tumour following total lobectomy may be obtained at the time of operation. If capsular or vascular invasion suggesting malignancy are clearly present, total thyroidectomy should be performed. Where the pathological findings are not so clear-cut (malignant features are easily missed because of sampling error), it is best to wait until definitive paraffin sections are available.

With minimally invasive lesions, total lobectomy and suppression of TSH with thyroxine are all that is required. If, however, overt vascular or capsular invasion is seen, reoperation is strongly advised: total thyroidectomy should ideally be carried out within 7 to 10 days; beyond that time surgery becomes difficult due to new vessel growth and fibrosis. Although multifocal disease is not common, removal of the main ‘iodine trap' enables distant metastases to be identified and ablated using radioiodine. The usual routine is to withhold thyroxine after thyroidectomy until the patient becomes hypothyroid, usually within 3 to 6 weeks. The increased TSH drive then makes the secondaries avid for iodine and therefore more likely to be identified on scintigraphy.

Anaplastic carcinoma

Surgery has little to offer for most patients, who are usually elderly, female, and infirm. The rare middle-aged patient may benefit from total thyroidectomy followed by external radiotherapy, although tumour cells may be implanted in the surgical incision. The surgeon's role is therefore mainly to establish the diagnosis beyond doubt. In view of the bulky nature of most anaplastic tumours a good representative sample can be obtained using a Trucut needle under local anaesthesia. Pressure on the trachea can sometimes be achieved by removal of the isthmus, but local radiotherapy often offers the only prospect of worthwhile palliation. The 12-month survival rate is zero and chemotherapy has so far failed to improve the dismal prognosis.

Malignant lymphoma

Some of these tumours, which mimic anaplastic carcinoma in their presentation and histological appearances, arise in pre-existing Hashimoto's disease or as part of a generalized lymphomatous disease. Most are radiosensitive and regress after external-beam radiotherapy; good 5-year survival figures have been reported. Chemotherapy is reserved for patients with disseminated disease. The role of surgery is therefore diagnostic and attempts to remove the lymphoma are rarely appropriate.

Medullary thyroid carcinoma

Total thyroidectomy is the procedure of choice in view of the high incidence of multicentric lesions in patients with both familial and sporadic disease. The upper poles, and in particular the primary lymphatic drainage areas, must be completely resected. The resection of lymph nodes in the central compartment should extend from the hyoid bone to the innominate vessels, since up to 75 per cent of them will be found to contain metastatic disease. Lymph nodes in the secondary drainage areas should be sampled and, if positive, removed in accordance with the modified dissection described for papillary carcinoma. When the diagnosis is suspected from the family history or confirmed by fine-needle biopsy, a coexisting phaeochromocytoma or parathyroid tumour needs to be excluded: failure to attend to these endocrine tumours first may be catastrophic. A serial rise in serum calcitonin after thyroidectomy is a strong indicator of recurrent disease; this is detectable by pentavalent DMSA scanning. Metastases may be treated by surgery, radiotherapy, radioiodine, or chemotherapy. Results are, however, often disappointing and calcitonin concentrations rarely return to normal.

Secondary thyroid carcinoma

Although rare, these most commonly arise from breast, kidney, ovary, or colon. Thyroidectomy may be indicated if the secondary deposits are confined to the thyroid and if the primary tumour is under control.

Operative surgery on the thyroid

Preoperative measures

Preparation is directed to ensure safe induction of anaesthesia and a trouble free intra- and postoperative course. Haemoglobin estimation, chest radiology (Fig. 24(a)), and an electrocardiogram are mandatory. Blood transfusion is rarely required: grouping and saving of serum is all that is required. If thyroid enlargement is massive or retrosternal, and if the patient shows clinical signs of respiratory embarrassment or superior vena caval obstruction, a CT scan of the neck and thoracic inlet (Fig. 24(b)) will indicate the possible need to enter the chest and potential problems that may be encountered on intubation. The vocal cords should always be examined by indirect laryngoscopy; this is especially important when the voice is compromised, malignancy is suspected, or when previous thyroid surgery has been undertaken. A small proportion of patients have unsuspected palsy of the recurrent laryngeal nerve; it is of clinical and medicolegal importance to both the patient and the surgeon to determine before surgery whether or not the vocal cords are moving normally. Thyrotoxic patients need to be rendered euthyroid or the peripheral effects of high circulating concentrations of thyroxine blocked. The majority of thyrotoxic patients referred for surgery have already received one or more courses of antithyroid drugs, but their condition remains unstable. Where residual toxicity is modest and surgery can be undertaken quickly, it is standard practice in many centres to stop antithyroid drugs 10 days before operation and switch to oral propranolol in a dose of 30 to 120 mg every 6 to 8 h. The dose is adjusted to keep the patient's sleeping pulse at around 70 beats/min. Since the effective duration of action of propranolol is approx. 6 h it is important to administer medication right up to induction of anaesthesia and to continue thereafter, especially if the patient develops tachycardia. b-blockers are contraindicated in patients with bronchial asthma, sinus bradycardia, or congestive heart failure. Patients with severe thyrotoxicosis who require relatively early surgery should receive a 6- to 8-week course of carbimazole (Neo-Mercazole® 10–15 mg 8-hourly). In the event of an adverse drug reaction, propylthiouracil (100 mg three times daily) may be substituted for carbimazole. Extended use of thiourea drugs may cause prothrombin deficiency (hence the advisability of stopping the drug 10 days before surgery), leucopenia, or profound bone-marrow depression. Full blood count and the international normalized ratio should be checked if reduced resistance to infection or impaired haemostasis are suspected.

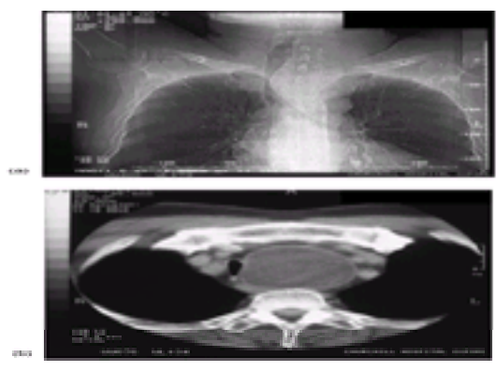


Fig. 24. Radiological assessment of a retrosternal goitre by (a) conventional chest radiograph and (b) CT scanning.

Standard surgical approach to the thyroid

The patient should be placed comfortably on the operating table and the indifferent electrode of the surgical diathermy attached securely to the thigh before placement of the surgical drapes. Good access to the anterior compartment of the neck is facilitated by placing a pillow under the shoulders to extend the cervical spine while supporting and stabilizing the head on a rubber ring or horseshoe. Failure to do so may result in severe postoperative pain and headache, especially in elderly patients with cervical spondylitis. In patients with a short, stocky neck or a large goitre it is helpful to depress the shoulders by exerting gentle traction on the arms alongside the body and securing their position with foam wedges (Fig. 25). The table is then tilted 15° head-up to reduce engorgement of the neck veins. A flat board or magnetic pad supported by pillows placed on the lower chest and upper abdomen provides a convenient instrument tray. The skin is prepared with a suitable antibiotic solution and the operative field draped after gauze packs have been pushed well down into the recesses between the head support and the shoulder pillow on each side. These absorb blood loss at the lateral extent of the surgical excision and provide a good anchor for towel clips. The skin to be incised is now rarely infiltrated with 1:1000 adrenaline in saline solution, since much of the bleeding at the skin edge stops spontaneously. More persistent points of loss are best recognized and dealt with immediately, since extensive bruising and wound haematoma can develop after the vasoconstrictive effects of adrenaline have worn off.

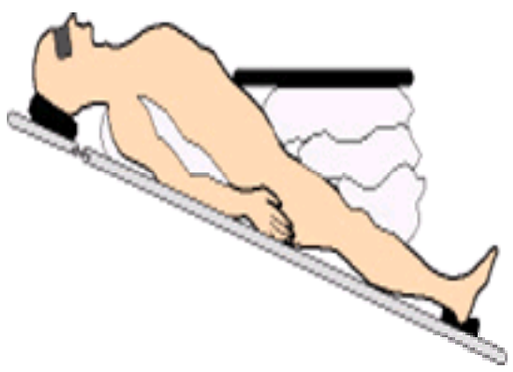


Fig. 25. Position of patient for operations on the thyroid gland.

The majority of thyroid procedures can be carried out through a Kocher collar incision that is placed in one of the natural skin creases (Langer's lines), approximately two finger breadths above the supraclavicular joint (Fig. 26). This may conveniently be marked before incision with a length of silk held taut against the skin around the convexity of the neck. The incision should be symmetrical: when the collagen in the scar matures it may contract unevenly if the stress lines are unequal, leading to a poor cosmetic result. Symmetry can be checked by the surgeon standing at the head of the table and looking down on the proposed incision directly from above. Matching the skin flaps at the end of the operation is rarely a problem for an experienced surgeon, and the practice of cross-hatching the incision line with a needle, which carries the risk of provoking a keloid reaction, is undesirable. Deepening the incision through the subcutaneous fat and platysma muscle discloses the deep cervical fascia, investing the strap muscle centrally and the sternomastoid laterally. The anterior jugular veins course beneath the fascial layer on the anterior surface of the thyroid and, provided that the surgeon is careful and keeps to the plane between the platysma and fascia, blood loss is minimal. Mobilization is extended superiorly to the prominence of the thyroid cartilage and inferiorly to the suprasternal notch. The diamond-shaped surgical field is held open with a self-retaining retractor and the deep cervical fascia is then incised through the midline raphe from top to bottom. Defining the planes beneath the strap muscles that extend over the surface of the thyroid lobes laterally requires care: the sternothyroid muscle may be extremely thin due to compression by a greatly enlarged thyroid. Adherence of these layers is a feature of the inflammatory process seen in autoimmune thyroiditis and when thyroid malignancy extends outside the gland. Further exposure of the thyroid requires delivery of each lobe into the wound after division and ligation of the middle thyroid vein on each side. In most patients the strap muscles present no impediment to the forward dislocation of the lobes; if they do they should be divided since division or removal causes no detectable functional disability. In practice, a pair of long, straight artery forceps inserted under the strap muscles at the junction of their upper one-third and lower two-thirds can define their lateral margin against the sternomastoid, allowing identification and preservation of the nerve of supply (the ansa hypoglossi) before the muscle is divided. Stay stitches to the upper and lower muscle flaps hitched over the joints of the self-retainer can act as a convenient retractor (Fig. 27).

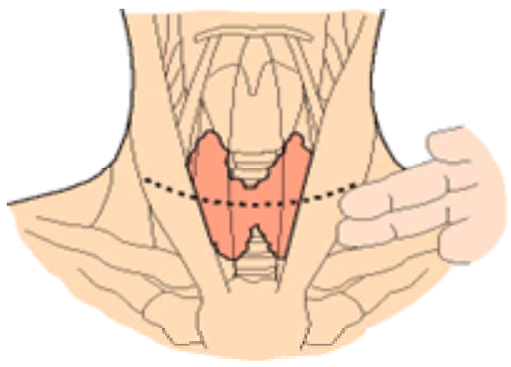


Fig. 26. Standard incision for operations on the thyroid gland.

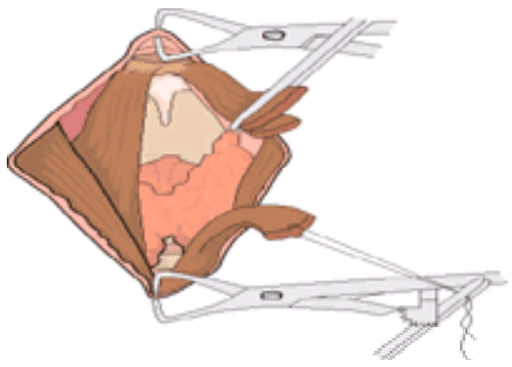


Fig. 27. Division and mobilization of strap muscles.

Specific thyroid surgical procedures

Subtotal thyroidectomy for thyrotoxicosis

The aim of this operation is to remove sufficient thyroid tissue to cure hyperthyroidism yet preserve a posterior remnant of the gland on each side sufficient to maintain the patient in a euthyroid state. Most surgeons prefer to stand on the opposite side of the table to the lobe being delivered. Following mobilization of the lobe the vascular pedicle of the upper pole is identified below the adherent sternothyroid muscle. A thyroid director is then passed between the larynx and the thyroid pedicle just above the suspensory ligament. Damage to recurrent laryngeal nerve and the superior laryngeal nerve is avoided if the point of the director is aimed upwards and laterally; fingertip pressure is sufficient to ensure that the point of the director is passing behind all of the upper lobe tissue, but nothing else (Fig. 28). With the director in place the pedicle is doubly ligated with a transfixion, non-absorbable suture. The lobe is then grasped with a pair of tenaculum forceps and rotated medially. The inferior thyroid artery, identified by blunt dissection of the space between the trachea and common carotid artery, is under-run cleanly and precisely with a ligature mounted on an aneurysm needle. If this is placed laterally it is ready for tying in continuity, avoiding involvement of the recurrent laryngeal nerve and parathyroids. Gentle traction on this ligature often throws the recurrent laryngeal nerve into prominence as it runs up at an acute angle from the mediastinum to reach the tracheo-oesophageal groove before assuming its intimate relations to the branches of this artery. When the nerve is not evident an aberrant course should be suspected; this may be lateral or anterior to the trachea, or even non-recurrent. The nerve can confidently be identified by its white colour, fine longitudinal surface vein, lack of pulsation, and lack of elasticity. To proceed with the operation without having identified the nerve is hazardous. The thyroid lobe is freed inferiorly by isolating and dividing between ligatures the inferior thyroid veins, and the thyroidea ima artery where present. During this manoeuvre the recurrent laryngeal nerve should be kept in view and avoided at all times. The inferior parathyroid gland may be seen at this stage, especially if ectopic, lying in the thyrothymic ligament. This and its blood supply will be preserved if the veins are swept medially and secured close to the gland. In this form of thyroidectomy the parathyroids are not specifically identified; attempting to do so may jeopardize their blood supply. If a parathyroid gland seems to have been inadvertently excised or devascularized it can be diced into 1-mm cubes and autotransplanted into the adjacent sternomastoid muscle. The surgeon then decides how much thyroid tissue to leave behind. The empirical formula of resecting seven-eighths and leaving one-eighth of the gland renders a large number of patients euthyroid. The merit of this approach is that it requires no modification for varying sizes of gland, but it has the disadvantage that it does not take into account the age of the patient (generally the younger the patient the more radical the resection needs to be) or the presence of antithyroid autoantibodies (which call for less radical excision). Attempts to standardized the size of the thyroid remnant to 3 to 4 g of tissue on each side, assessed by linear measurement or dental-wax impressions, are not only highly inaccurate but meaningless. Having decided on the size of the thyroid remnant, small artery forceps are placed along the line of the incision on the posterolateral aspect of the surgical capsule of the gland. Placement above the concentration defined by the anterior surface of the trachea allows injury to the recurrent laryngeal nerve and parathyroids to be minimized. The gland is then incised with a scalpel blade directed obliquely down towards the trachea (Fig. 29). This procedure is then performed on the opposite side of the neck. Both lobes and isthmus having been freed the gland now remains attached only by the pyramidal lobe or fibrous remnant of the thyroglossal tract. This requires careful dissection upwards so that no additional thyroid tissue or blood supply is overlooked. The thyroid remnants are then sutured to the pretracheal fascia with a continuous absorbable suture in a herringbone fashion, picking up the surgical capsule and rolling the gland towards the midline and away from the recurrent laryngeal nerves (Fig. 30). As well as having a haemostatic effect, this ensures that the fibrotic reaction around the thyroid is well away from important structures if the neck requires further exploration at a later date. The thyroid bed should usually be drained on each side: where the dead space is modest a suction drain is ideal. After flexing the head, the strap muscles, if divided, are rejoined by interrupted sutures and then approximated in the midline. Finally, the platysma is closed and then the skin either with clips or a subcuticular suture.

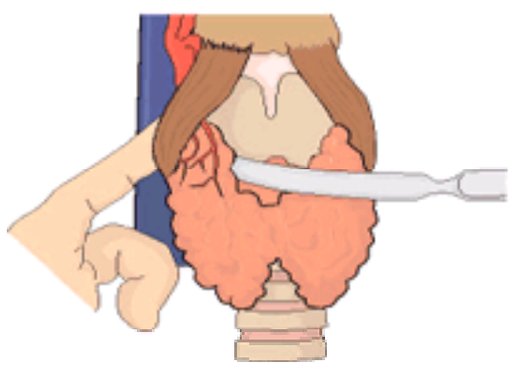


Fig. 28. Mobilization of superior pole of the thyroid gland.

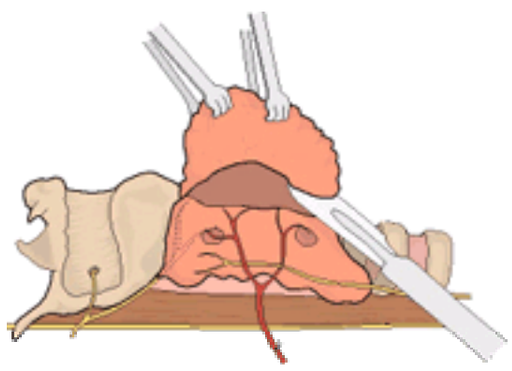


Fig. 29. Removal of thyroid lobe by sharp dissection, preserving parathyroid glands.

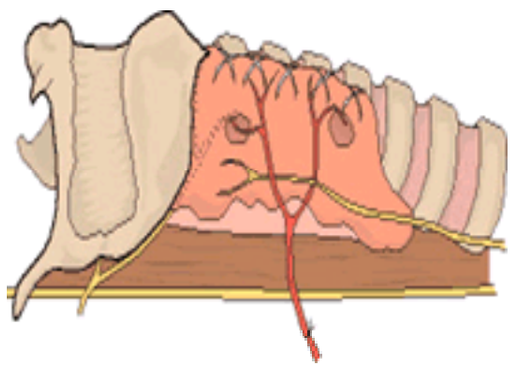


Fig. 30. Haemostatic suture of residual thyroid.

Subtotal thyroidectomy for multinodular goitre

This procedure is appropriate when a multinodular goitre needs to be reduced in size. The standard exposure of the thyroid is performed as described above. A total lobectomy is done on the most affected side and a partial lobectomy on the other side, leaving as much normal tissue as possible. The isthmus and pyramidal lobe should be removed, even if apparently normal, since compensatory hypertrophy at these sites is particularly unacceptable cosmetically. The rationale for total lobectomy on one side is that should further surgery be necessary for subsequent hypertrophy in the remaining remnant, only one side will need to be operated on. The possibility of damage to the recurrent laryngeal nerve, always a significant risk with reoperative thyroid surgery, is therefore reduced. Where large multinodular glands with diffuse disease are encountered at surgery, however, it is the authors' policy to perform a total thyroidectomy and place the patient on thyroxine replacement therapy.

Removal of a retrosternal goitre

Nearly all retrosternal goitres can be removed via a cervical incision, which should be placed higher than the conventional incision already described. However, if preoperative symptoms and investigations indicate that the chest may need to be opened, then the anterior chest wall is prepared and draped accordingly. In this instance the services of an experienced anaesthetist are imperative, especially if the airway is severely compromised. It is often safer to withhold sedative premedication and to rely on rapid intubation of the conscious patient, after spraying the vocal cords with local anaesthetic. Wholly or partially intrathoracic goitres derive their arterial supply from the superior and inferior arteries in the neck, and these should be secured first. Once the upper pole has been freed an attempt should be made to dissect the retrosternal portion of the lobe with the index finger, generally easing it upwards. Minimal force should be used to avoid damage to the recurrent laryngeal nerves, which cannot always be visualized. If this manoeuvre fails, intracapsular enucleation of the retrosternal goitre may be appropriate; if retrosternal thyroid malignancy is suspected or if surgery is being performed for a recurrent goitre, formal splitting of the sternum is appropriate. An incision is made in the midline down to the periosteum, with cutting diathermy from the suprasternal notch to the xiphisternum (Fig. 31). The parietal pleura is freed in the midline and deviated laterally; the retrosternal goitre is then removed (Fig. 32) along routine lines and the chest closed by accurately reapproximating the sternum using wire sutures. A retrosternal drain is brought out superiorly and attached to an underwater seal to control any small pneumothorax that might be present.

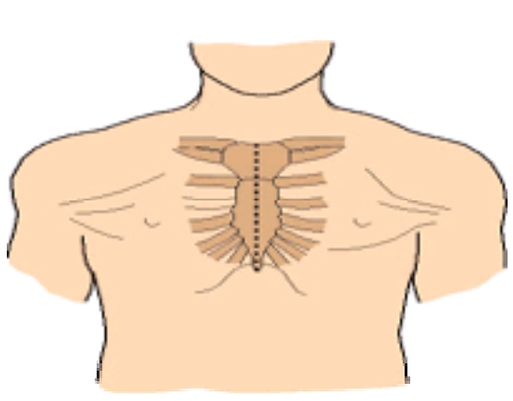


Fig. 31. Incision for partial or complete splitting of the sternum to gain access to a large retrosternal goitre.

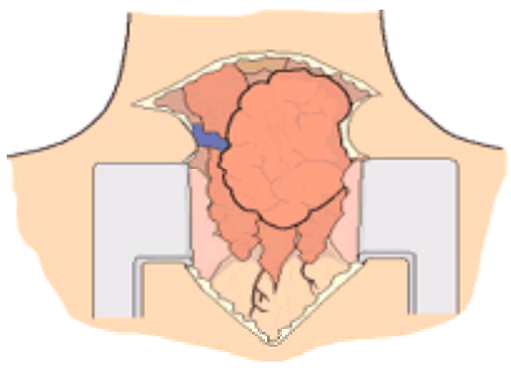


Fig. 32. Exposure of retrosternal goitre following splitting of the upper sternum.

Total thyroid lobectomy or total thyroidectomy

Thyroid lobectomy should always be used for the treatment of a non-toxic solitary nodule, which may prove to be malignant. If malignancy is confirmed on histological examination of frozen section, total thyroidectomy may be appropriate, depending on cell type. The approach to the thyroid is standard but the subtotal excision technique described earlier differs in several important ways. At no stage is the gland grasped with tissue forceps for traction: this may rupture a malignant focus, with spillage of tumour cells. Multicentric tumour deposits or capsular invasion must be assumed to be present. The entire lobe, isthmus, and the midline portion of the contralateral lobe should therefore be removed, although if the recurrent laryngeal nerve or parathyroids may be compromised, it is better to leave a small amount of thyroid tissue and ablate this later with radioiodine. Blood is supplied to each parathyroid via an end artery arising from the inferior thyroid artery, but collaterals to this artery are probably picked up on the thyroid capsule from the oesophageal and tracheal vessels. It is better, therefore, not to ligate the inferior thyroid artery in continuity but to trace out each of its terminal branches and individually clip and divide these between mosquito forceps as they enter the false capsule. Concurrently, the recurrent laryngeal nerve is traced out along its course. As it traverses Berry's ligament to gain access to the larynx, damage will be avoided by dissecting the nerve free under direct vision. Local invasion may prevent complete removal of thyroid malignancy, notably on the side of the larynx and oesophagus. In these circumstances the extent of the residual tumour should be marked with small clips to assist the radiotherapist in planning subsequent treatment fields.

Modified block dissection for thyroid carcinoma

The standard Kocher incision is extended on one or both sides (Fig. 33) to allow routine exposure of the thyroid gland. If the malignancy involves the strap muscles or sternomastoid on one side these are resected *en bloc* with the thyroid, as described previously. The plane between the sternomastoid and strap muscles is opened up and the carotid sheath exposed. Involvement of lymph nodes of the internal jugular chain, extending to the posterior cervical and supraclavicular groups, can be detected if the sternal and clavicular heads of the sternomastoid are detached, and the muscle rotated upwards. This provides excellent exposure yet preserves the blood supply (occipital artery) and nerve supply (accessory). The muscle can be reattached on completion of the nodal dissection. Great care is needed when removing the upper and lower deep cervical nodes and their surrounding fat from the surface of the internal jugular vein. However, if the internal jugular is heavily affected it too will require excision. The vagus and phrenic nerves are easily identified and preserved; however, the thoracic duct on the left and the main lymphatic duct on the right are easily damaged, and if this occurs the affectedduct should be tied off rather than repaired.

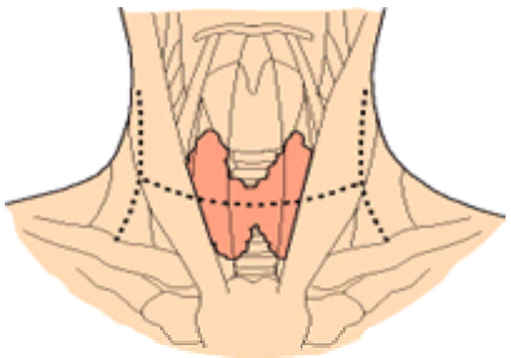


Fig. 33. Modification of thyroid incision to include bilateral lymph-node dissection.

Postoperative complications

In competent hands, thyroid surgery is associated with few complications and no fatality. There are, however, local and specific complications that can severely compromise the outcome and be life threatening. The majority are avoidable with sound surgical technique and good preparation, particularly in thyrotoxic patients.

Haemorrhage

This is typically reactionary and is a potential problem in the first 24 h after surgery. Failure to secure the superior thyroid vessels efficiently, preferably with a transfixion suture, is associated with the risk of serious blood loss. Inadequate control of the inferior and middle thyroid vein may also have serious consequences. Major haemorrhage deep to the strap muscles must be recognized quickly: this causes pressure on the airway within a confined space and may lead to rapid laryngeal and subglottic oedema. Medical and nursing staff need to be aware of the significance of pallor, respiratory difficulty, stridor, and swelling of the wound. The absence of blood loss from drains is not a reliable indicator of haemostasis since these can block easily. Any haematoma should be evacuated immediately, and intubation or tracheostomy may be necessary to avert a potentially life-threatening situation. Clip removers and a pair of artery forceps should be readily available at the bedside of all patients after thyroidectomy. These will allow the incision and the strap muscles to be opened. If stridor persists, skilled anaesthetic help is needed to perform intubation, which may be difficult in the presence of oedema. In the absence of such help a minitracheostomy or large Medicut needle and cannula (No.12 blue) inserted percutaneously through the cricothyroid membrane or between the tracheal rings should stabilize the patient until haemorrhage can be arrested.

Recurrent laryngeal nerve damage

The incidence of permanent and transient damage to the nerve is low (less than 0.1 per cent and 2–4 per cent, respectively). It is largely avoidable if the surgeon routinely seeks to identify the nerve on each side during all operations on the gland. Loss of vocal power and huskiness is often evident for 2 or 3 days after surgery; this is most likely to be due to oedema and is relieved by local anaesthetic lozenges and/or humidified air. Persistence of symptoms may indicate neuropraxia, caused by stretching or crushing of the nerve; this is reversible and recovers over several weeks or months. Permanent damage will result if the nerve is divided or ligated and is more likely to occur when the anatomy is distorted, for example with recurrent or malignant goitres. Unilateral injury may be asymptomatic and pass undetected due to compensatory hyperadduction of the unaffected cord unless routine postoperative laryngoscopy is performed. Symptomatic unilateral paralysis improves if the affected cord is stabilized in adduction by the submucous injection of Teflon under direct laryngoscopy. The effects of bilateral nerve injury are likely to be temporary but pose an immediate problem when the patient is extubated at the end of surgery since the unopposed adductor action of the cricothyroid muscles closes the glottis to such an extent that the least exertion results in obstruction. The patient must be promptly reintubated, paralysed, and ventilated whilst hydrocortisone is given, 100 mg three times daily, to combat the oedema and inflammatory response. The patient can usually be successfully extubated within 48 h; if this extubation fails tracheostomy is required. The permanent nature of laryngeal damage should not be accepted unless it lasts for more than 9 months. Exploration and resuture of the nerve, with grafting when necessary, is now feasible, as is the anastomosis of the hypoglossal and recurrent nerves.

Superior laryngeal nerve paresis

The true incidence of this injury is unknown, owing to the lack of any objective test of function until recently. A change in the voice, such as loss of pitch and inability to make explosive sounds, makes damage likely. Voice analysis using a Visipitch oscilloscope will help to confirm this damage, which may occur in up to 25 per cent of patients. The majority will recover, since the nerve has only been stretched. If no improvement is evident after 3 months, it is unlikely to occur. Bilateral damage results in a very flat, hoarse voice that tires easily.

Hypoparathyroidism

The serum calcium concentration should be monitored postoperatively as patients without overt hypoparathyroidism may develop vague lethargy and depression, insidious cataracts, mental deterioration, and psychosis. Hypocalcaemia due to parathyroid deficiency will usually be evident within 1 week of operation and should be suspected if the patient appears unduly agitated or depressed, or hyperventilates. Circumoral tingling is generally the first and most sensitive indicator of a low serum calcium; paraesthesia in the fingers and toes preceding frank tetany is seen when hypocalcaemia is profound. Tapping over the facial nerve will cause contraction of the facial muscles (Chvostek–Weiss sign); however, this phenomenon may also be observed in 10 to 15 per cent of normal individuals. Carpopedal spasm, provoked by occlusion of the circulation to the arm (Trousseau's sign), indicates severe hypocalcaemia: intravenous infusion of 10 ml of 10 per cent calcium gluconate (given slowly to avoid cardiac arrest in systole) is required. This infusion may need to be repeated every 4 to 6 h. Oral effervescent calcium, 4 to 6 g daily, should also be administered, depending on response. If hypocalcaemia persists, vitamin D (calciferol 25 000–100 000 units) and 2 to 3 g oral calcium per day are given until a normocalcaemic state is achieved. Signs and symptoms of hypocalcaemia will recur in these patients at times of metabolic stress, such as pregnancy or the menopause.

Hypothyroidism and recurrent hyperthyroidism

The ability of the thyroid to produce sufficient thyroxine after thyroidectomy reflects not only the size of the remnant but also the pre-existing pathological processes within the gland. Hypothyroidism is inevitable after total thyroidectomy or malignancy, but is less predictable after, for example, a thyroid lobectomy to remove a benign solitary nodule. Avoiding hypothyroidism is one of the main challenges when operating for thyrotoxicosis. Factors affecting postoperative thyroid function include the severity of disease before surgery; the age of the patient; the presence of high concentrations of antithyroid autoantibodies before surgery; the size of the gland and histological evidence of lymphoid infiltration; all of these influence how much of the gland is to be removed. An experienced surgeon can expect resultant hypothyroidism rates of 10 to 15 per cent and persistent hyperthyroidism below 5 per cent. Hypothyroidism with rising TSH concentrations should be allowed to develop over 6 weeks after total thyroidectomy for malignancy (notably follicular lesions). A radioiodine scan will then identify possible distant metastases, which can be ablated. Thereafter, T_3 (50–100 µg/day) is administered in preference to T_4 as its shorter biological half-life (1 week) enables repeat scans to be performed with minimal delay. Once isotope ablation of residual disease has been achieved, conversion to T_4 is appropriate. The dose required varies but the majority of patients only require 0.1 to 0.2 mg/day. Nearly all patients undergoing surgery for thyrotoxicosis become biochemically hypothyroid for 2 to 3 months after surgery; no correction is necessary as the majority will then stabilize in a euthyroid state. Clinical assessment should be maintained for at least 2 years, during which serum thyroxine and TSH concentrations should be monitored: a small percentage of patients will become clinically hypothyroid and require T_4 supplementation (0.1 to 0.2 mg daily). Routine administration of T_4

(0.1 mg daily) is recommended for all patients undergoing surgery for non-toxic, diffuse, or multinodular goitres since failure to suppress TSH drive can result in recurrent goitre, even if hypothyroidism is subclinical.

Tracheal collapse

Collapse of the trachea, the wall of which has become softened due to chondromalacia, occasionally occurs after removal of a long-standing goitre, especially if retrosternal. Such patients usually also suffer laryngeal oedema and reduced movement of the vocal cords, following difficult intubation and delivery of the lobes. Whenever the trachea is markedly soft and narrow, an elective tracheostomy should be performed.

Thyroid crisis

This is now very rare, with the improved methods of control of thyrotoxicosis, but when fully expressed is characterized by high fever, tachycardia (atrial fibrillation), extreme restlessness, and delirium. High doses of antithyroid drugs [carbimazole (Neo-Mercazole®) 30 mg immediately and then 15 mg 8-hourly], plus 1 g of sodium iodide intravenously should be given promptly. Propranolol (2 mg) is slowly infused intravenously, with electrocardiographic control. Fluid replacement, ice-pack cooling, and sedation may help to abort the crisis.

Cervical sympathetic damage

This rare complication results from deep, forceful retraction on the carotid sheath, producing Horner's syndrome. This is notable by the absence of the vascular dilatation component. The resulting myosis and ptosis are frequently permanent.

Wound complications

The deposition of excessive collagen in the scar to form a keloid is an unpredictable complication, but is said to be more prevalent in blacks, redheads, and in those undergoing surgery during pregnancy. Unless the scar can be excised and adapted to conform more readily to Langer's lines, reoperation is unlikely to confer any improvement, but topical steroids and low-dose irradiation may prevent recurrence. Infection is an uncommon complication and when it occurs a foreign body should be suspected; the most common offender is non-absorbable suture material. Rarely, sensitivity to the nickel clips used for skin closure results in blistering and breakdown.

Further reading

Bäckdahl M, Wallin G, Aner G, Lundell G, Löwhagen T, Cranber PO. Cellular DNA content in thyroid tumours. A reliable factor for grading and prognosis. *Progress in Surgery* 1988; **19**: 40–53.

Barnes HV, Gann DS. Choosing thyroidectomy in hyperthyroidism. *Surgical Clinics of North America* 1974; **54**: 289.

Beahrs OH, Sakulsky SB. Surgical thyroidectomy in the management of exophthalmic goitre. *Archives of Surgery* 1968; **96**: 512–15.

Black EG *et al.* Serum thyroglobulin in thyroid cancer. *Lancet* 1981; **ii**: 443–5.

Crile G Jr. The fallacy of the conventional neck dissection for papillary carcinoma of the thyroid. *Annals of Surgery* 1957; **145**: 317.

De Visschar M, ed. *The thyroid gland*. Raven, New York, 1980.

Duffy BJ Jr, Fitzgerald PJ. Cancer of the thyroid in children. *Journal of Clinical Endocrinology and Metabolism* 1950; **10**: 1296–308.

Gharib H, Goellner JR. Fine-needle aspiration biopsy of the thyroid: an appraisal. *Annals of Internal Medicine* 1993; **118**(4): 282–9.

Gharib H, Goellner JR, Johnson DA. Fine-needle aspiration cytology of the thyroid. A 12-year experience with 11,000 biopsies. *Clinics in Laboratory Medicine* 1993; **13**(3): 699–709.

Hall R, Besser M, ed. *Fundamentals of clinical endocrinology*, 4th edn. Churchill Livingstone, Edinburgh, 1989.

Hay ID, Grant CS, Taylor WF, McConahey WM. Ipsilateral lobectomy versus bilateral lobar resection in papillary carcinoma. A retrospective analysis of surgical outcome using a novel prognostic scoring system. *Surgery* 1987; **102**: 1088–95.

Hay ID, Bergstralh EJ, Goellner JR, Ebersold JR, Grant CS. Predicting outcome in papillary thyroid carcinoma: development of a reliable prognostic scoring system in a cohort of 1779 patients surgically treated at one institution during 1940 through 1989. *Surgery* 1993; **114**(6): 1050–7.

Hedley AJ *et al.* The effect of remnant size on the outcome of subtotal thyroidectomy for thyrotoxicosis. *British Journal of Surgery* 1972; **59**: 559–63.

Hollingshead WH. Anatomy of the endocrine glands. *Surgical Clinics of North America* 1952; **32**: 1115–40.

Johnson IDA, Thompson NW, ed. *Endocrine surgery*. Butterworth, London, 1983.

Kaplan EL, ed. Surgery of the thyroid and parathyroid glands. *Clinical surgery international, Vol. 6.* Churchill Livingstone, Edinburgh, 1983.

Kark AE *et al.* Voice changes after thyroidectomy: the role of the external laryngeal nerve. *British Medical Journal* 1984; **289**: 1412–15.

Löwhagen T, Granberg PO, Lundell G, Skinnari P, Sunblad R, Willeum JS. Aspiration biopsy cytology (ABC) in nodules of the thyroid gland suspected of being malignant *Surgical Clinics of North America* 1979; **59**: 3–18.

Mazzaferri EL. Management of a solitary thyroid nodule. *New England Journal of Medicine* 1993; **328**(8): 553–9.

Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. *American Journal of Medicine* 1994; **97**(5): 418–28.

McGregor AM, Rees-Smith B, Hall R, Petersen MM, Millar M, Dewer PJ. Prediction of relapse in hyperthyroid Graves' disease. *Lancet* 1980; **i**: 1101.

Miccoli P, Antonelli A, Iacconi P, Alberti B, Gambuzza C, Baschieri L. Prospective, randomized, double-blind study about effectiveness of levothyroxine suppressive therapy in prevention of recurrence after operation: result at the third year of follow-up. *Surgery* 1993; **114**(6): 1097–101.

Moley JF. Medullary thyroid cancer. *Surgical Clinics of North America* 1995; **75**(3): 405–20.

Moreno A, Rodriguez JM, Sola J, Soria T, Parrilla P. Encapsulated papillary neoplasm of the thyroid: retrospective clinicopathological study with long term follow up. *European Journal of Surgery* 1996; **162**(3):177–80.

Ozata M, Suzuki S, Miyamoto T, Liu RT, Fierro-Renoy F, DeGroot LJ. Serum thyroglobulin in the follow-up of patients with treated differentiated thyroid cancer. *Journal of Clinical Endocrinology and Metabolism* 1994; **79**(1): 98–105.

Neil HB III, Townsend GL, Devine KD. Bilateral vocal cord paralysis of undetermined aetiology: clinical course and outcome. *Annals of Otology, Rhinology and Laryngology* 1972; **81:** 514–19.

Peden NR, Gunn A, Browning MCK. Nadolol and potassium iodine in combination in the surgical treatment of thyrotoxicosis. *British Journal of Surgery* 1982; **69**: 638–41.

Reeve TS *et al.* Thyroid cancers of follicular cell origin. *Progress in Surgery* 1988; **19**: 78–88.

Rose N. Investigation of post-thyroidectomy patients for hypoparathyroidism. *Lancet* 1963; **i**: 124–7.

Russell CF *et al.* The surgical management of medullary thyroid carcinoma. *Annals of Surgery* 1983; **197**: 42–8.

Russell WO *et al.* Thyroid carcinoma classification. Intraglandular dissection and clinicopathological study based upon whole organ section of 80 glands. *Cancer* 1963; **16**: 1425–60.

Sadler GP and Wheeler MH. Parathyroid and thyroid cancer. In: Taylor I, Cooke TG, Guillou P, ed. *Essential general surgical oncology*, Ch. 18, pp. 189–202. Churchill Livingstone, London, 1996.

Samaan NA *et al.* The results of various modalities of treatment of well differentiated thyroid carcinomas: a retrospective review of 1599 patients. Journal of Clinical *Endocrinology and Metabolism* 1992; **75**(3): 714–20.

Sanders LE, Rossi RL. Occult well differentiated thyroid carcinoma presenting as cervical node disease. *World Journal of Surgery* 1995: **19**(4): 642–6.

Santoro M, Grieco M, Melillo RM, Fusco A, Vecchio G. Molecular defects in thyroid carcinomas: role of the RET oncogene in thyroid neoplastic transformation. *European Journal of Endocrinology* 1995; **133**(5): 513–22.

Scheumann GF, Gimm O, Wegener G, Hundeshagen H, Dralle H. Prognostic significance and surgical management of locoregional lymph node metastases in papillary thyroid cancer. *World Journal of Surgery* 1994; **18**(4): 559–67.

Shaha AR, Loree TR, Shah JP. Prognostic factors and risk group analysis in follicular carcinoma of the thyroid. *Surgery* 1995; **118**(6): 1131–6.

Smith BR, Hall R. Thyroid stimulating immunoglobulins in Grave's disease. *Lancet*, 1974; **ii**: 427–30.

Soh EY, Clark OH. Surgical considerations and approach to thyroid cancer. *Endocrinology and Metabolism Clinics of North America* 1996; **25**(1): 115–39.

Sozzi G *et al.* Cytogenetic and molecular genetic characterization of papillary thyroid carcinomas. *Genes, Chromosomes and Cancer* 1992; **5**(3): 212–18.

Sugg SL, Zheng L, Rosen IB, Freeman JL, Ezzat S, Asa SL. *ret/PTC-1, -2*, and *-3* oncogene rearrangements in human thyroid carcinomas: implications for metastatic potential? *Journal of Clinical Endocrinology and Metabolism* 1996; **81**(9): 3360–5.

Sweeney DC, Johnston GS. Radioiodine therapy for thyroid cancer. *Endocrinology and Metabolism Clinics of North America* 1995; **24**(4): 803–39.

Takahashi M. Oncogenic activation of the ret protooncogene in thyroid cancer. *Critical Reviews in Oncogenesis* 1995; **6**(1): 35–46.

Udelsman R, Dudley NE. Medullary carcinoma of the thyroid: management of persistent hypercalcitonaemia utilizing [^{99m}Tc] (v) dimercaptosuccinic acid scintography. *British Journal of Surgery* 1989; **76**: 1276–81.

Wade JSH. The management of malignant thyroid tumours. *British Journal of Surgery* 1983; **70**: 253–5.

Woolner LB, Beahrs OH, Black BM, McConahey WM, Keating FR. Classification and prognosis of thyroid carcinoma. *American Journal of Surgery* 1961; **102**: 354–87.

20.3.1 Parathyroid disorders: molecular genetics and physiology

R. V. Thakker

[Introduction](#)
[PTH gene structure and function](#)
[Hypercalcaemic diseases](#)
- [Parathyroid tumours](#)
- [Disorders of the calcium-sensing receptor](#)
- [Jansen's disease](#)
- [Williams' syndrome](#)

[Hypocalcaemic disorders](#)
- [Hypoparathyroidism](#)
- [Calcium-sensing receptor abnormalities](#)
- [Pseudohypoparathyroidism](#)

[Further reading](#)

Introduction

The parathyroid glands secrete parathyroid hormone (**PTH**) at a rate that is appropriate to, and dependent upon, the prevailing extracellular calcium ion concentration. Parathyroid gland dysfunction results in either hypercalcaemia or hypocalcaemia, and these disorders can be classified according to whether they arise from an excess of PTH, or from a deficiency or insensitivity to PTH (Table 1). The application of recent developments in molecular biology to the study of these parathyroid disorders has enabled characterization of some of the mechanisms involved in the regulation of parathyroid development, proliferation, and PTH secretion. Thus, mutations in the gene for the calcium-sensing receptor have been reported in patients with familial benign hypercalcaemia and autosomal-dominant hypocalcaemia, and the role of the oncogene *PRAD1*, which encodes a novel cyclin, and of the multiple endocrine neoplasia type 1 (*MEN1*) gene, have been revealed in the pathogenesis of some parathyroid tumours (Fig. 1). In addition, mutations in the *PTH* gene and the mitochondrial genome have been associated with some forms of hypoparathyroidism, mutations in the *G*a gene have been demonstrated in some patients with pseudohypoparathyroidism, and candidate genes for the DiGeorge syndrome have been identified. The location of other susceptibility genes, for example for the hereditary hyperparathyroidism and jaw tumours syndrome, has been identified on chromosome 1q21–q31, and genetic heterogeneity in the familial benign hypercalcaemia and DiGeorge disorders has been established. These molecular genetic studies have provided a unique opportunity to elucidate the pathogenesis of some diseases of the parathyroids, and these advances, together with the structure and function of the *PTH* gene, will be reviewed in this chapter.

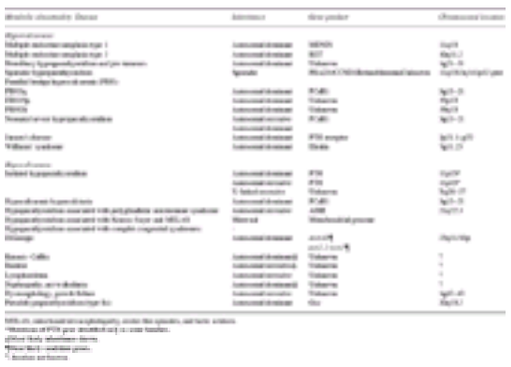


Table 1 Parathyroid disease and their chromosomal locations

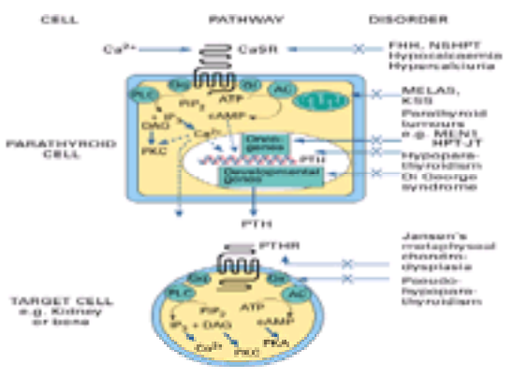


Fig. 1. Schematic representation of some of the components involved in calcium homeostasis. Alterations in extracellular calcium are detected by the calcium-sensing receptor (CaR), which is a 1078 amino acid G-protein-coupled receptor. The PTH-receptor (PTHR) is also a G-protein-coupled receptor. Thus, both Ca^{2+} and PTH involve G-protein-coupled signalling pathways. Ca^{2+} and PTH bind to their specific receptors, which leads to activation of Gs, Gi and Gq. Gs stimulates adenyl cyclase (AC), which catalyses the formation of cAMP from ATP. Gi inhibits AC activity. cAMP stimulates phosphokinase (PK) A (PKA), which phosphorylates cell-specific substrates. Activation of Gq stimulates phospholipase C (PLC), which catalyses the hydrolysis of the phosphoinositide (PIP_2) to inositol triphosphate (IP_3), which increases intracellular calcium, and diacylglycerol (DAG), which activates PKC. These proximal signals modulate downstream pathways, which result in specific physiological effects. Abnormalities in the genes and encoding proteins in these pathways have been identified as associated with specific disorders of calcium homeostasis (see Table 1).

PTH gene structure and function

The *PTH* gene is situated on the short arm of chromosome 11 (11p15) and consists of three exons (transcribed regions) separated by two introns. Exon 1 of the *PTH* gene is 85 bp in length and is untranslated (Fig. 2), whereas exons 2 and 3 code for the 115-amino acid, prepro-PTH peptide. Exon 2 is 90 bp in length and encodes the initiation (ATG) codon, the prehormone sequence, and part of the prohormone sequence. Exon 3 is 612 bp in size and encodes the remainder of the prohormone sequence, the mature PTH peptide, and the 3'-untranslated region. The 5'-regulatory sequence of the human *PTH* gene contains a vitamin D-response element 125 bp upstream of the transcription start site, which downregulates PTH mRNA transcription in response to vitamin D-receptor binding. *PTH* gene transcription (as well as PTH peptide secretion) are also dependent upon the extracellular calcium concentration, although the presence of a specific upstream ‘calcium response element' has not yet been demonstrated. The mature PTH peptide is secreted from the parathyroid chief cell as an 84-amino acid peptide, but when the PTH mRNA is first translated it is as a prepro-PTH peptide. The ‘pre-' sequence consists of a 25-amino acid signal peptide (leader sequence), which is responsible for directing the nascent peptide into the endoplasmic reticulum to be packaged for secretion from the cell. The ‘pro-' sequence is six amino acids in length and, although its function is less well defined than that of the ‘pre-' sequence, it is also essential for the correct processing and secretion of PTH. After the 84-amino acid, mature PTH peptide has been secreted from the parathyroid cell it is cleared from the circulation, with a short half-life of about 2 min, via non-saturable hepatic uptake and renal excretion. PTH shares a receptor with PTH-like hormone (formerly known as PTH-related peptide) and the PTH/PTH-like hormone receptor (Fig. 1) is a member of a subgroup of the G-protein-coupled receptor family. The gene for the PTH/PTH-like hormone receptor is localized to the short arm of chromosome 3 (3p21–p24) and is expressed in kidney and bone, where PTH is its predominant agonist. The PTH/PTH-like hormone receptor is also expressed in the brain, heart, skin, lung, liver, and testis, where it mediates the actions of PTH-like hormone. Mutations involving the genes that encode these proteins and receptors in this calcium-regulating pathway (Fig. 1) are associated with hypercalcaemia and hypocalcaemic disorders (Table 1).

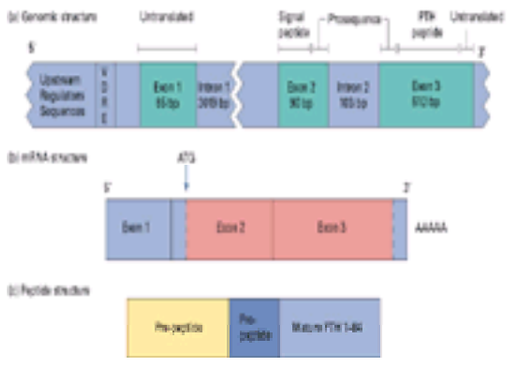


Fig. 2. Schematic representation of the *PTH* gene, PTH mRNA, and PTH peptide. The *PTH* gene consists of three exons and two introns. Exon 1, the 5' end of exon 2, and the 3' end of exon 3 are untranslated, and the PTH peptide is encoded by exons 2 and 3. The PTH peptide consists of the prepeptide, the propeptide, and the mature PTH peptide. Only the mature PTH peptide, which is 84 amino acids in size, is secreted from the parathyroid cell.

Hypercalcaemic diseases

Parathyroid tumours

Parathyroid tumours may occur as an isolated and sporadic endocrinopathy, or as part of inherited tumour syndromes such as the multiple endocrine neoplasias or hereditary hyperparathyroidism with jaw tumours, or in response to chronic overstimulation as in uraemic hyperparathyroidism. The aetiology of parathyroid tumours involves activation or overexpression of the proto-oncogenes *PRAD1* (*par*athyroid *a*denoma *1*) and *RET*, and the inactivation of tumour suppressor genes such as the *MEN1* gene, the retinoblastoma (*Rb*) gene, and a locus on chromosome 1p.

PRAD1 gene

Studies of the organization of the *PTH* gene in sporadic parathyroid adenomas revealed that some adenomas had abnormally sized restriction fragment length polymorphisms with a DNA probe for the 5' part of the *PTH* gene, indicating that there was an acquired disruption to this part of the gene. Further molecular studies of the tumour DNA demonstrated that the first exon of the *PTH* gene (Fig. 2) was separated from the fragments containing the second and third exons, and that a rearrangement had occurred juxtaposing the 5' PTH-regulatory elements with ‘new' non-PTH DNA. Investigation of this rearranged DNA sequence localized it to chromosome 11q13, and detailed analysis revealed that it was highly conserved in different species and expressed in normal parathyroids and in parathyroid adenomas. This expressed sequence from the rearrangement, which was designated *PRAD1*, was demonstrated to encode a 295-amino acid member of the cyclin-D family of cell-cycle regulatory proteins. Cyclins were initially characterized in the dividing cells of budding yeast, where they control the G_1–S transition of the cell cycle, and in marine molluscs, where they regulate the mitotic (M) phase of the cell cycle. Cyclins have also been identified in man and have an important role in regulating many stages of cell-cycle progression. Thus, *PRADI*, which encoded a novel cyclin referred to as D1, is an important cell-cycle regulator, and overexpression of *PRADI* may be an important event in the development of at least 15 per cent of sporadic parathyroid adenomas. Interestingly, a transgenic mouse overexpressing *PRAD1* under the control of a mammary tissue-specific promoter has recently been investigated; two-thirds of these mice were found to develop breast carcinoma in adult life but they have not been found to develop parathyroid tumours.

MEN1 gene

The multiple endocrine neoplasia syndrome, which consists of several subtypes (Table 2), is characterized by the occurrence of tumours involving two or more endocrine glands within a single patient. This disorder has previously been referred to as multiple endocrine adenopathy and the pluriglandular syndrome. However, glandular hyperplasia and malignancy may also occur in some patients and the term multiple endocrine neoplasia is now preferred. There are two major forms of multiple endocrine neoplasia, type 1 (Wermer's syndrome) and type 2 (Sipple's syndrome); each form is characterized by the development of tumours within specific endocrine glands (Table 2). These syndromes are uncommon, but, because they are inherited as autosomal-dominant disorders, the finding of multiple endocrine neoplasia in a patient has important implications for other family members; first-degree relatives have about a 50 per cent risk of developing the disease. Occasionally, these syndromes may arise sporadically (i.e., without a family history). It may be difficult to make the distinction between sporadic and familial forms; in some cases the family history may be absent because the parent with the disease may have died before developing any manifestations. Thus, biochemical and genetic screening may have important roles in patients with multiple endocrine neoplasia and their families.

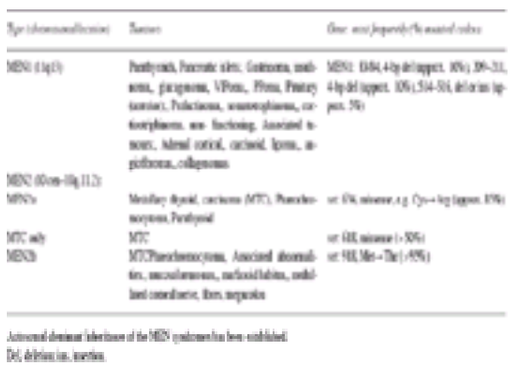


Table 2 The multiple endocrine neoplasia (MEN) syndromes, their characteristic tumours, and associated genetic abnormalities

Multiple endocrine neoplasia type 1 is characterized by the combined occurrence of tumours of the parathyroids, pancreatic islet cells, and anterior pituitary. Parathyroid tumours occur in 95 per cent of type 1 patients, resulting in the hypercalcaemia of primary hyperparathyroidism. Parathyroid tumours detected by hypercalcaemia are the first manifestation of type 1 syndrome in about 90 per cent of patients. Pancreatic islet-cell tumours occur in 40 per cent of type 1 patients. Gastrinomas are the most common type and insulinomas are the second most common. Glucagonomas and VIPomas are rare, whereas pancreatic polypeptidomas (PPomas) are more common but remain asymptomatic. Gastrinomas, leading to the Zollinger–Ellison syndrome, are the most important cause of morbidity and mortality in patients with type 1 multiple endocrine neoplasia. Anterior pituitary tumours occur in 30 per cent of type 1 patients: most (60 per cent) are prolactinomas, somatotrophinomas are the next most common (20 per cent), and corticotrophinomas and non-functioning tumours represent less than 15 per cent of these pituitary tumours. Associated tumours, which occur less commonly in multiple endocrine neoplasia type 1, include adrenal cortical tumours (5 per cent), carcinoids (4 per cent), lipomas (1 per cent), and collagenomas (less than 1 per cent). Facial angiofibromas have recently been recognized to be a common associated feature of MEN1.

The gene causing multiple endocrine neoplasia type 1 was localized to chromosome 11q13 by genetic mapping studies that investigated the associated tumours for loss of heterozygosity and by segregation studies in type 1 families. The results of these studies, which were consistent with Knudson's model for tumour development, indicated that the *MEN1* gene represented a putative tumour-suppressor gene. Further genetic mapping studies defined a less than 300-kb region as the minimal critical segment that contained the *MEN1* gene; characterization of genes from this region led to the identification of the *MEN1* gene, which consists of 10 exons with a 1830-bp coding region that encodes a novel, 610-amino acid protein, referred to as ‘**MENIN**'. Mutations of the *MEN1* gene have been identified: of the approximately 140 germ-line mutations, about 25 per cent are nonsense, about 45 per cent are deletions, about 15 per cent are insertions, less than 5 per cent are donor-splice mutations, and about 10 per cent are missense. Thus, the majority (more than 80 per cent) of these mutations are inactivating, and are consistent with those expected in a tumour-suppressor gene. The mutations are not only diverse in their types but are also scattered throughout the 1830-bp coding region of the *MEN1* gene, with no evidence for clustering. However, some of the mutations occur several times in unrelated families, and the three deletional and insertional mutations involving codons 83 and 84, 209–211, and 514–516 account for approximately 25 per cent of all the germ-line *MEN1* mutations, and thus may represent potential ‘hot' spots (Table 2). Correlations between the *MEN1* mutations and the clinical manifestations of the disorder appear to be absent. For example, a detailed study of five unrelated families with the same 4-bp deletion (CAGT) in codons 210 and 211 revealed a wide range of multiple endocrine neoplasia type 1-associated tumours. More than 10 per cent of the *MEN1* mutations arise *de novo* and may be transmitted to subsequent generations. It is also important to note that between 5 to 20 per cent of type 1 patients may not harbour mutations in the coding region of the *MEN1* gene, and that these individuals may have mutations in the promoter or untranslated regions, which remain to

be investigated. These, together with the wide diversity of mutations in the 1830-bp coding region of the *MEN1* gene and an apparent lack of genotype–phenotype correlations, which contrasts with the situation in multiple endocrine neoplasia type 2 (see below), will make mutational analysis for diagnostic purposes in type 1 time-consuming and expensive.

MEN1 mutations in sporadic, non-MEN1 tumours

Parathyroid, pancreatic islet-cell, and anterior pituitary tumours may occur either as part of MEN1 or more commonly as sporadic, non-familial tumours. Tumours from MEN1 patients harbour the germ-line mutation together with a somatic loss of heterozygosity involving chromosome 11q13, as expected from Knudson's model and the proposed role of the *MEN1* gene as a tumour suppressor. However, loss of heterozygosity involving chromosome 11q13, which is the location of the *MEN1* gene, has also been observed in 5 to 50 per cent of sporadic endocrine tumours, implicating the *MEN1* gene in their aetiology. Mutational analyses of sporadic parathyroid tumours, gastrinomas, insulinomas, anterior pituitary adenomas, carcinoids, and lipomas have revealed somatic *MEN1* mutations in 0.03 to 36 per cent, with the occurrence of *MEN1* mutations in sporadic parathyroid tumours, gastrinomas, and carcinoids being greater than that in anterior pituitary adenomas. The sporadic tumours harbouring a somatic *MEN1* mutation all had loss of heterozygosity at chromosome 11q13 as the other genetic abnormality or ‘hit', consistent with Knudson's hypothesis. These studies indicate that, although inactivation of the *MEN1* gene may have a role in the aetiology of some sporadic endocrine tumours, the involvement of other genes is highly likely.

Function of MENIN

Analysis of the predicted amino acid sequence encoded by the *MEN1* gene did not reveal homologies with any other proteins, sequence motifs, signal peptides, or consensus nuclear-localization signals, and thus the putative function of the protein (MENIN) could not be deduced. Studies based on immunofluorescence, Western blotting of subcellular fractions, and epitope tagging with enhanced green fluorescent protein have revealed that MENIN is located primarily in the nucleus. Furthermore, deletional constructs of enhanced green fluorescent protein-tagged MENIN have identified at least two independent nuclear localization signals in the C-terminal quarter of the protein. Interestingly, none of the *MEN1* germ-line missense or in-frame deletions alters either of these putative nuclear-localization signals. However, all of the truncated proteins that would result from the nonsense and frameshift mutations of *MEN1*, if expressed, would lack these signals. The nuclear localization of MENIN suggests that it could act in the regulation of transcription, and recent studies have confirmed this. MENIN interacts with the N-terminus of the AP1 (apurinic/apyridonimic) transcriptional factor JunD to repress transcriptional activity.

The *MEN2* gene (*c-ret*)

Multiple endocrine neoplasia type 2 describes the association (Table 2) of medullary thyroid carcinoma, phaeochromocytomas, and parathyroid tumours. Three clinical variants of MEN2 are recognized: MEN2a, MEN2b, and medullary thyroid carcinoma only (Table 2). MEN2a is the most common variant, and the development of medullary thyroid carcinoma is associated with phaeochromocytomas (50 per cent of patients), which may be bilateral, and parathyroid tumours (20 per cent of patients). MEN2b, which represents 5 per cent of all MEN2 cases, is characterized by the occurrence of medullary thyroid carcinoma and phaeochromocytoma in association with a marfanoid habitus, mucosal neuromas, medullated corneal fibres, and dysfunction of intestinal autonomic ganglia leading to multiple diverticulae and megacolon. Parathyroid tumours do not usually occur in MEN2b. MTC-only is a variant in which medullary thyroid carcinoma is the sole manifestation of the syndrome. The gene causing all three variants of MEN2 was mapped to chromosome 10cen–10q11.2, a region containing the *c-ret* proto-oncogene, which encodes a tyrosine kinase receptor with cadherin-like and cysteine-rich extracellular domains, and a tyrosine kinase intracellular domain. Specific mutations of *c-ret* have been identified for each of the three variants of MEN2 (Table 2). Thus, in 95 per cent of patients, MEN2a is associated with mutations of the cysteine-rich extracellular domain and mutations in codon 634 account for 85 per cent of MEN2a mutations. However, a search for *c-ret* mutations in sporadic, non-MEN2a parathyroid adenomas revealed no mutations in codon 634. MTC-only is also associated with mutations in the cysteine-rich extracellular domain and most mutations are in codon 618. However, MEN2b is associated with mutations in codon 918 of the intracellular tyrosine kinase domain in 95 per cent of patients. Interestingly, the *c-ret* proto-oncogene is also involved in the aetiology of papillary thyroid carcinoma and Hirschsprung's disease. Mutational analysis of *c-ret* to detect mutations in codons 609, 611, 618, 634, 768, and 804 in MEN2a and MTC-only, and codon 918 in MEN2b, has been used in the diagnosis and management of patients and families with these multiple endocrine neoplasistic disorders.

Rb gene

The *Rb* gene, which is a tumour-suppressor gene located on chromosome 13q14, is involved in the pathogenesis of retinoblastoma and a variety of common, sporadic human malignancies including ductal breast, small-cell lung, and bladder carcinomas. Allelic deletion of the *Rb* gene has been demonstrated in all of six parathyroid carcinomas and in four of 40 parathyroid adenomas studied. An abnormal staining pattern for the Rb protein was also found in three of the parathyroid carcinomas studied but not in any of the parathyroid adenomas. This allelic deletion of the *Rb* gene in parathyroid carcinomas, coupled with abnormal Rb protein staining in the tumours, demonstrates the important role of the *Rb* gene in the development of parathyroid carcinomas, and this may help in the histological distinction of parathyroid adenoma from carcinoma. However, the findings of extensive deletions of the long arm of chromosome 13 (including the *Rb* locus) in some parathyroid adenomas and carcinomas, and similar findings in pituitary carcinomas, suggest that other tumour-suppressor genes on chromosome 13q may also have a role in the development of such tumours.

Gene on chromosome 1p

Allele loss of chromosome 1p32-pter has been documented in 10 of 25 (40 per cent) sporadic parathyroid adenomas. This region is estimated to be about 110 cM, equivalent to 110 million base pairs (Mbp) of DNA, but recent studies have narrowed the interval containing this putative tumour suppressor gene(s) to a 4-cM (i.e., 4-Mbp) region.

Autosomal-dominant hyperparathyroidism syndromes

The hereditary hyperparathyroidism jaw tumour syndrome is an autosomal-dominant disorder characterized by the occurrence of parathyroid adenomas and carcinomas in association with fibro-osseous mandibular or maxillary jaw tumours and occasional Wilms' tumours or adult nephroblastomas. Genetic linkage studies of five families with this syndrome have mapped the gene causing this disorder to chromosome 1q21–q31. The lack of allelic loss over chromosome 1q21–q31 in eight parathyroid tumours from affected family members examined suggests that this gene (designated *HRPT2*) is a proto-oncogene rather than a tumour suppressor.

Familial isolated primary hyperparathyroidism has been reported in several kindreds, and family linkage studies have been pursued to localize the genes causing this disorder in two kindreds. A study of one large British family with parathyroid carcinomas has excluded linkage between this disorder and genetic markers from 11q13. However, a study of one large Danish kindred has established linkage between familial isolated primary hyperparathyroidism and genetic markers from chromosome 11q13, and it remains to be established whether this is associated with abnormalities at the *MEN1* or the *PRAD1* loci.

Hyperparathyroidism in chronic renal failure

Chronic renal failure is associated with a form of secondary hyperparathyroidism that may subsequently result in the hypercalcaemic state of ‘tertiary' hyperparathyroidism. The presence of a parathyroid proliferative response in this condition led to the proposal that the autonomous parathyroid tissue might have undergone hyperplastic change and therefore be polyclonal in origin. However, studies of X-chromosome inactivation in parathyroids from 11 patients on haemodialysis with refractory hyperparathyroidism have demonstrated that seven patients (64 per cent) had at least one monoclonal parathyroid tumour. Furthermore, one of these parathyroid tumours had allele loss involving several loci on chromosome Xp11, thereby suggesting the involvement of a tumour-suppressor gene from this region in their pathogenesis. Interestingly, none of the parathyroid tumours from these 11 patients with chronic renal failure had allele losses involving loci from chromosome 11q13. This unexpected finding of monoclonal parathyroid tumours in the majority of patients with ‘tertiary' hyperparathyroidism suggests that a possible high parathyroid-cell turnover in secondary hyperparathyroidism may render the parathyroid glands more susceptible to mitotic non-disjunction or other mechanisms of somatic deletions, which may involve loci other than those located on chromosome 11q13.

Disorders of the calcium-sensing receptor

Two hypercalcaemic disorders due to mutations of the calcium-sensing receptor (CaR) have been reported: familial benign hypercalcaemia, which is also referred to as familial hypocalciuric hypercalcaemia (FHH), and neonatal severe hyperparathyroidism (NSHPT). Mutational analyses of the human calcium-sensing receptor, which is a G-protein-coupled receptor located on chromosome 3q13–q21, have revealed 21 different mutations that result in a loss of function of the calcium-sensing receptor in patients with familial benign hypercalcaemia and neonatal severe hyperparathyroidism. Many of these mutations cluster around the aspartate- and glutamate-rich regions (codons 39–300) within the extracellular domain of the receptor, and this has been proposed to contain low-affinity calcium-binding sites, based on similarities to that of calsequestrin, in which the ligand-binding pockets also contain negatively charged amino acid residues. Approximately two-thirds of the kindreds with familial benign hypercalcaemia investigated have unique heterozygous mutations of the calcium-sensing receptor, and expression studies of these mutations have demonstrated a loss of the receptor function whereby there is an increase in the calcium ion-dependent set-point for PTH release from the parathyroid cell. Neonatal

severe hyperparathyroidism occurring in the offspring of a consanguineous family with familial benign hypercalcaemia has been shown to be due to homozygous mutations in the calcium-sensing receptor. However, some patients with sporadic neonatal hyperparathyroidism have been reported to be associated with *de novo* heterozygous mutations of the calcium-sensing receptor, thereby suggesting that factors other than mutant gene dosage (for example, the degree of set-point abnormality, the bony sensitivity to PTH, and the maternal extracellular calcium concentration) may also all play a part in the phenotypic expression of a mutation in the calcium-sensing receptor in the neonate. The remaining one-third of families with familial benign hypercalcaemia in whom a mutation within the coding region of the calcium-sensing receptor has not been demonstrated may either have an abnormality in the promoter of the gene or a mutation at one of the two other familial benign hypercalcaemia loci that have been revealed by family linkage studies. One of these loci for familial benign hypercalcaemia is located on chromosome 19p and is referred to as FBH_{19p}. However, similar linkage studies of one kindred with familial benign hypercalcaemia from Oklahoma, which also suffered from progressive elevations in PTH, hypophosphataemia and osteomalacia, demonstrated that this variant (designated FBH_{Ok}) was not linked to chromosomes 3q13–q21 or 19p, but to 19q13, thereby revealing a third locus for familial benign hypercalcaemia whose location remains to be defined.

Jansen's disease

Jansen's is an autosomal-dominant disease characterized by short-limbed dwarfism due to metaphyseal chondrodysplasia, and severe hypercalcaemia and hypophosphataemia despite a normal or undetectable serum PTH. These abnormalities are associated with an activated mutation of the PTH receptor (Fig. 1), and thus this represents a PTH-independent activation of that receptor. Two different mutations of the receptor have been identified, and these involve codon 223 (His¬Arg) and codon 410 (Thr¬Pro). Expression of the mutant receptors in COS-7 cells resulted in constitutive, ligand-independent accumulation of cAMP, while the basal accumulation of inositol phosphates was not increased. These findings provide a likely explanation for the abnormalities observed in mineral homeostasis and growth-plate development in this disorder.

Williams' syndrome

Williams' syndrome is an autosomal-dominant disorder characterized by supravalvular aortic stenosis, elfin-like facies, psychomotor retardation, and infantile hypercalcaemia. The underlying abnormality of calcium metabolism remains unknown but abnormal metabolism of 1,25-dihydroxy vitamin D_3 or decreased production of calcitonin have been implicated, although none has been consistently demonstrated. Recent studies have demonstrated hemizygosity at the elastin locus on chromosome 7q11.23 in nine children with Williams' syndrome and this finding has been confirmed in over 90 per cent of patients with the classical Williams' phenotype. Only one of these patients was found to have a large deletion detectable by conventional cytogenetic methods, thereby indicating that Williams' syndrome is most often due to a microdeletion of 7q11.23. This microdeletion has been reported to involve another gene, designated LIM-kinase, that is expressed in the central nervous system. The calcitonin receptor gene has recently been localized to chromosome 7q21, close to the region deleted in Williams' syndrome. However, the calcitonin receptor gene was not involved in the deletion found in four patients with Williams' syndrome, thereby indicating that it is unlikely to be implicated in the hypercalcaemia of such children. While the involvement of the elastin and LIM-kinase genes in the deletions in patients with Williams' syndrome can explain its cardiovascular and neurological features, respective, it seems possible that another, as yet uncharacterized, gene within this contiguously deleted region is likely to be involved to explain the abnormalities of calcium metabolism.

Hypocalcaemic disorders

Hypoparathyroidism

Hypoparathyroidism may occur as part of a pluriglandular autoimmune disorder or as a complex congenital defect, as for example in the DiGeorge syndrome. In addition, it may develop as a solitary endocrinopathy—isolated or idiopathic hypoparathyroidism. Familial occurrences of isolated hypoparathyroidism with autosomal-dominant, autosomal-recessive, and X-linked recessive inheritances have been established.

PTH gene abnormalities

DNA-sequence analysis of the *PTH* gene, which is located on chromosome 11p15 and consists of three exons, from one patient with autosomal-dominant isolated hypoparathyroidism has revealed a single-base substitution (T¬C) in exon 2. This resulted in the substitution of arginine (**C**GT) for cysteine (**T**GT) in the signal peptide (Fig. 2); the presence of a charged amino acid in the midst of the hydrophobic core of the signal peptide would impede the processing of the mutant prepro-PTH. This block has been demonstrated by *in vitro* studies, which revealed that the mutation impaired the interaction with the nascent protein and the translocation machinery, and that cleavage of the mutant signal sequence by solubilized signal peptidase was ineffective. Another abnormality of the *PTH* gene, involving a donor splice site at the exon 2–intron 2 boundary, has been identified in one family with autosomal-recessive isolated hypoparathyroidism. This mutation involved a single-base transition (g¬c) at position 1 of intron 2; the effects of this alteration in the invariant gt dinucleotide of the 5' donor splice-site consensus on mRNA processing were assessed by detecting the non-tissue-specific transcription of the normal and mutant *PTH* genes. The extent of the non-tissue-specific expression of genes is estimated to be one molecule of correctly spliced mRNA per 1000 cells; the physiological relevance of, and mechanisms involved in, this low level of ectopic transcription are not known. However, the demonstration of non-tissue-specific transcription is of clinical importance as it enables the use of easily accessible peripheral blood lymphocytes for the detection of abnormalities in mRNA processing and thereby avoids the requirement for expressing tissue that may only be obtainable by biopsy. Advantage was taken of these methods to demonstrate that the donor splice-site mutation resulted in exon skipping, in which exon 2 of the *PTH* gene was lost and exon 1 was spliced to exon 3. The lack of exon 2 would lead to a loss of the initiation codon (ATG) and the signal peptide sequence (Fig. 2), which are required for the commencement of PTH mRNA translation and for the translocation of the PTH peptide, respectively.

X-linked recessive hypoparathyroidism

X-linked recessive hypoparathyroidism has been reported in two multigenerational kindreds from Missouri. Only males were affected and they suffered from infantile onset of epilepsy and hypocalcaemia, which was due to an isolated defect in parathyroid gland development. Linkage studies utilizing X-linked restriction fragment length polymorphisms in these families localized the mutant gene to chromosome Xq26–q27 by establishing linkage between hypoparathyroidism and the locus *DXS98*.

Pluriglandular autoimmune hypoparathyroidism

Hypoparathyroidism may occur in association with candidiasis and autoimmune Addison's disease, and the disorder has been referred to as either the autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (**APECED**) syndrome or the polyglandular autoimmune type 1 syndrome. This disorder has a high incidence in Finland, and a genetic analysis of 58 patients from 42 Finnish families indicated autosomal-recessive inheritance. In addition, the disorder is reported to have a high incidence among Iranian Jews, although the occurrence of candidiasis was less common in this population. Linkage studies of 14 Finnish families mapped the *APECED* gene to chromosome 21q22.3. Positional cloning studies isolated a novel gene from chromosome 21q22.3. This gene, referred to as *AIRE* (*a*uto*i*mmune *re*gulator), encodes a 545-amino acid protein that contains motifs suggestive of a transcriptional factor and includes two zinc-finger motifs, a proline-rich region, and three LXXLL motifs. Six *AIRE* mutations have been reported in the APECED families and a codon-257 (Arg®Stop) mutation was the predominant abnormality in 82 per cent of the Finnish families. The identification of the genetic defect causing APECED will not only facilitate genetic diagnosis but will also enhance the elucidation of the mechanisms causing autoimmune disease.

Mitochondrial disorders associated with hypoparathyroidism

Hypoparathyroidism has been reported to occur in two disorders associated with mitochondrial dysfunction: the Kearns–Sayre syndrome and the **MELAS** syndrome. Kearns–Sayre syndrome is characterized by progressive external ophthalmoplegia and pigmentary retinopathy before the age of 20 years, and is often associated with heart block or cardiomyopathy. The MELAS syndrome consists of a childhood onset of *m*itochondrial *e*ncephalopathy, *l*actic *a*cidosis, and *s*troke-like episodes. In addition, varying degrees of proximal myopathy can be seen in both conditions. Both the Kearns–Sayre and MELAS syndromes have been reported to occur with insulin-dependent diabetes mellitus and hypoparathyroidism. A point mutation in the mitochondrial gene tRNA leucine (UUR) has been reported in one patient with the MELAS syndrome who also suffered from hypoparathyroidism and diabetes mellitus. A large deletion, consisting of 6903 bp and involving 39 per cent of the mitochondrial genome, has been reported in another patient who suffered from Kearns–Sayre syndrome, hypoparathyroidism, and sensorineural deafness. The role of these mitochondrial mutations in the aetiology of hypoparathyroidism remains to be elucidated.

DiGeorge syndrome

Patients with the DiGeorge syndrome suffer from neonatal hypoparathyroidism, immunodeficiency, congenital heart defects, and deformities of the ear, nose, and mouth. The disorder arises from a congenital failure in the development of the derivatives of the third and fourth pharyngeal pouches, with resulting absence or hypoplasia of the parathyroids and thymus. Most cases of DiGeorge syndrome are sporadic, but an autosomal-dominant inheritance has been observed, and an association between the syndrome and an unbalanced translocation and deletions involving 22q11.2 has also been reported. In some patients, deletions of another

locus on chromosome 10p have been observed in association with DiGeorge syndrome. Many expressed sequences [i.e., genes: *ZNF74, T10, COMT, LZTR-1, GPIB*b, *TUPLE-1* (*HIRA*), *LAN, DGCR2*, and *DGS-A* to *DGS-J*] have been described as mapping to the DiGeorge syndrome-deleted region on chromosome 22q11.2. However, the *T10, COMT, LZTR-1, GPIB*b, and *TUPLE-1* (*HIRA*) genes have been found to be outside a recently defined, 250-kb minimal critical region and these have thereby been excluded as candidate genes for DiGeorge syndrome. The cloning of the translocation breakpoint on 22q11.21 from a patient with DiGeorge syndrome has revealed that there are probably two genes (*rnex40* and *nex2.2–nex3*), transcribed in opposite directions, which are disrupted by this breakpoint. The coding region of one of these genes, designated *rnex40*, has homology to the mouse and rat androgen receptors and contains a leucine zipper motif, suggesting that the candidate gene for DiGeorge syndrome may be a DNA-binding protein. Eleven nucleotides of the *rnex40* gene are deleted at the translocation junction, making it likely that loss of function of this gene is responsible for at least part of the DiGeorge phenotype. Another partial transcript, referred to as *nex2.2–nex3*, was also identified from this breakpoint. Both *rnex40* and *nex2.2–nex3* are deleted in all patients with DiGeorge syndrome with 22q11 deletions; studies aimed at assessing the presence of hemizygosity and mutations in these two genes in patients with the syndrome who do not have detectable 22q11 deletions are required to demonstrate the role of these genes in its aetiology.

Additional familial syndromes

Several familial syndromes have been reported in which hypoparathyroidism forms a component of a complex, multisystem, developmental disorder that is unique to that kindred (Table 1). The inheritance of the disorder has, in some instances, been established and abnormalities of the *PTH* gene excluded. The susceptibility gene for one of these complex, familial, multisystem syndromes has been localized to chromosome 1q42–43.96. This syndrome, which has been identified in families of Bedouin origin, consists of congenital hypoparathyroidism, growth retardation, developmental delay, and dysmorphic features. Homozygosity and linkage-disequilibrium mapping studies in six families have mapped this gene associated with hypoparathyroidism to chromosome 1q42–43.

Calcium-sensing receptor abnormalities

Mutations of the calcium-sensing receptor that result in a loss of function are associated with familial benign (hypocalciuric) hypercalcaemia and it was speculated that mutations which resulted in a gain of function may lead to hypocalcaemia with hypercalciuria. Investigation of 12 kindreds with autosomal-dominant forms of hypocalcaemia have identified such mutations of the calcium-sensing receptor. The hypocalcaemic individuals generally had normal serum concentrations of intact PTH and hypomagnesaemia, and treatment with vitamin D or its active metabolites to correct the hypocalcaemia resulted in marked hypercalciuria, nephrocalcinosis, nephrolithiasis, and renal impairment. Ten missense mutations of the calcium-sensing receptor resulting in a functional gain have been identified, and eight of these are located within the extracellular domain. This suggests that these mutations of the extracellular domain may increase the affinity of the receptor for calcium binding.

Pseudohypoparathyroidism

Patients with pseudohypoparathyroidism are characterized by hypocalcaemia and hyperphosphataemia due to PTH resistance rather than PTH deficiency. A complete resistance to a PTH infusion, demonstrated by a lack of increase in urinary cAMP and urinary phosphate excretion, is referred to as pseudohypoparathyroidism type I. The occurrence of type I with the somatic features of Albrights's hereditary osteodystrophy is referred to as type Ia, whereas the sole occurrence of the biochemical features is referred to as type Ib. The occurrence of Albrights's hereditary osteodystrophy without any biochemical abnormalities is referred to as pseudopseudohypoparathyroidism. The absence of a normal rise in urinary excretion of cAMP after an injection of PTH in pseudohypoparathyroidism type Ia and pseudopseudohypoparathyroidism indicated a defect of the hormone-sensitive adenylate cyclase system, which is regulated by at least two G-proteins, one of which stimulates (Gsa) and another inhibits (Gia) the activity of the membrane-bound enzyme that catalyses the formation of the intracellular second messenger, cAMP. Mutations of the *Gsa* gene, which is located on chromosome 20q12–q13.2, have been identified in patients with pseudohypoparathyroidism type Ia and pseudopseudohypoparathyroidism. *Gsa* mutations have not been detected in type Ib, which has been considered to be due to a defect of the PTH/PTH-related peptide receptor. However, a study of 17 patients with pseudohypoparathyroidism type Ib did not detect abnormalities of the PTH/PTH-related peptide receptor gene and the aetiology of this form of pseudohypoparathyroidism remains to be elucidated.

Further reading

Agarwal *et al.* MENIN interacts with API transcriptional factor JunD and represses Jun-activated transcription. *Cell* 1999; **96**: 143–52.

Arnold A, Horst SA, Gardella TJ, Baba H, Levine MA, Kronenberg HM. Mutations of the signal peptide encoding region of preproparathyroid hormone gene in isolated hypoparathyroidism. *Journal of Clinical Investigation* 1990; **86**: 1084–7.

Budarf ML *et al.* Cloning a balanced translocation associated with DiGeorge syndrome and identification of a disrupted candidate gene. *Nature Genetics* 1995; **10**: 269–78.

Chandrasekharappa SC *et al.* Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 1997; **276**: 404–7.

European Consortium on MEN1. Identification of the multiple endocrine neoplasia type 1 (MEN1) gene. *Human Molecular Genetics* 1997; **6**: 1177–83.

Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc finger domains. *Nature Genetics* 1997; **17**: 399–403.

Gagel RF, Cotes GJ. *Ret* protooncogene mutations in multiple endocrine neoplasia type 2. In: Bilezikian JP, Raisz LG, Rodan GA, ed. *Principles of bone biology*, pp. 799–807. Academic Press, San Diego CA, 1996.

Juppner H *et al.* A G protein-linked receptor for parathyroid hormone and parathyroid hormone-related peptide. *Science* 1991; **254**: 1024–6.

Lloyd SE, Pannett AAJ, Dixon PH, Whyte MP, Thakker RV. Localisation of the Oklahoma variant of familial benign hypercalcaemia (FBH_{Ok}) to chromosome 19q13. *American Journal of Human Genetics* 1999; **64**: 189–95.

Moraes CT *et al.* Mitochondrial deletions in progressive external ophthalmoplegia and Kearns-Sayre syndrome. *New England Journal of Medicine* 1989; **320**: 1293–9.

Motokura T *et al.* A novel cyclin encoded by a bcl1-linked candidate oncogene. *Nature* 1991; **350**: 512–15.

Mulligan LM, Ponder BAJ. Multiple endocrine neoplasia type 2. *Journal of Clinical Endocrinology and Metabolism* 1995; **80**: 1989–95.

Mulligan LM *et al.* Germline mutations of the *RET* proto-oncogene in multiple endocrine neoplasia type 2a (MEN2a). *Nature* 1993; **363**: 458–60.

Nagamine K *et al.* Positional cloning of the APECED gene. *Nature Genetics* 1997; **17**: 393–8.

Parkinson DB, Thakker RV. A donor splice site mutation in the parathyroid hormone gene is associated with autosomal recessive hypoparathyroidism. *Nature Genetics* 1992; **1**: 149–53.

Pearce SHS, Bai M, Quinn SJ, Kifor O, Brown EM, Thakker RV. Functional characterisation of calcium-sensing receptor mutations expressed in human embryonic kidney cells. *Journal of Clinical Investigation* 1996; **98**: 1860–6.

Pearce SHS *et al.* A familial syndrome of hypocalcaemia with hypocalciuria due to mutations in the calcium-sensing receptor gene. *New England Journal of Medicine* 1996; **335**: 115–22.

Pollak MR *et al.* Mutations in the human Ca^{2+}-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell* 1993; **75**: 1297–303.

Pollak MR *et al.* Autosomal dominant hypocalcaemia caused by a calcium-sensing receptor gene mutation. *Nature Genetics* 1994; **8**: 303–7.

Schipani E, Kruse K, Juppner H. A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia. *Science* 1995; **268**: 98–100.

Thakker RV. Multiple endocrine neoplasia type 1 (MEN1). In: de Groot LJ, Besser GM, Burger HC, ed. *Endocrinology 1*, pp. 2815–31. Saunders, Philadelphia, 1995.

Thakker RV. Molecular genetics of parathyroid disease. *Current Opinion in Endocrinology and Diabetes* 1996; **3**: 521–8.

Thakker RV. Multiple endocrine neoplasia—syndromes of the twentieth century. *Journal of Clinical Endocrinology and Metabolism* 1998; **83**: 2617–20.

Weinberg RA. Tumor suppressor genes. *Science* 1991; **254**: 1138–45.

20.3.2 Surgery of the parathyroid glands

Gregory P. Sadler and Nicholas Dudley

Surgical anatomy
Size and number
Embryological development
Blood and nerve supply
Macroscopic and microscopic appearance
Pathophysiology
Primary hyperparathyroidism
Secondary hyperparathyroidism
Tertiary hyperparathyroidism
Pseudohyperparathyroidism
Clinical features of hyperparathyroidism
Bones
Stones
Abdominal groans
Psychic moans
Physical signs
Investigation of hyperparathyroidism
Localization of parathyroid tumours
Selective venous sampling with PTH assay
Thallium/technetium subtraction scanning
Sestamibi scanning
Ultrasonography
Computed tomography (CT) and nuclear magnetic resonance imaging (MRI)
Indications for surgery
Asymptomatic primary hyperparathyroidism
Symptomatic primary hyperparathyroidism
Secondary and tertiary hyperparathyroidism
Operative surgery on the parathyroids
Preoperative preparation
Exposure and technique of exploration
Surgical strategy for primary hyperparathyroidism
Surgical strategy for secondary and tertiary hyperparathyroidism
Surgery for parathyroid carcinoma
Surgery for MEN syndromes
Endoscopic parathyroidectomy and minimally invasive, radioguided parathyroidectomy
Postoperative results, complications, and management
Strategy after failed cervical exploration for hyperparathyroidism
Further reading

Surgical anatomy

Size and number

There are normally four parathyroid glands in man but developmental anomalies can give rise to a smaller or larger number. Gilmore's much quoted autopsy study of 428 cadavers found two glands in 0.2 per cent, three in 6.1 per cent, five in 6 per cent, and six in 0.5 per cent of individuals. In other words, in only 87 per cent of the population is there the normal complement of four glands. It is possible that amongst the 6.2 per cent of individuals with fewer than four glands, some are instances of fused upper and lower parathyroid tissue or of failed identification. Normal parathyroids also vary in size, shape, and location. Typically they measure 3 to 5 × 2 to 4 × 0.5 to 2 mm. The combined weight of the four glands is approximately 120 mg, the superior glands usually being smaller (between 20–40 mg) and the inferior ones from 30 to 50 mg.

Embryological development

The upper parathyroids arise from the dorsal endoderm of the fourth branchial pouch, the ventral part of which is fused laterally to the developing thyroid gland on the floor of the primitive pharynx. Thus some 92 per cent of the upper parathyroids remain in close contact with the posterolateral aspect of the thyroid lobes, just above and behind the level at which the recurrent laryngeal nerve crosses the inferior thyroid artery (Fig. 1).

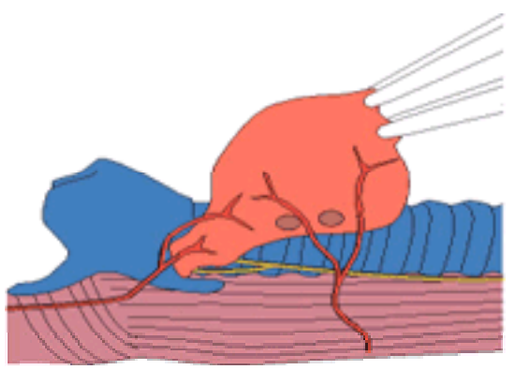


Fig. 1. Normal site of the upper parathyroid gland above or behind branches of the inferior thyroid artery.

When the upper parathyroid glands are ectopic they become progressively more dorsally displaced. As a result, the upper parathyroids may, in a small number of individuals, come to lie between the thyroid and oesophagus or behind the oesophagus in the upper posterior mediastinum (Fig. 2). It has been suggested that parathyroid enlargements favour dorsal displacement due to the forces of deglutition and negative intrathoracic pressure. Aberrant superior glands may also be found rarely within the carotid sheath.



Fig. 2. Yellow shading indicates area where ectopic upper parathyroids may be found; black dots indicate where the author has found ectopic upper parathyroid tumours.

The lower parathyroid glands develop from the dorsal endoderm of the third branchial pouch, which also gives rise to the thymus. In contrast to the fourth-pouch derivatives, which remain fairly static during embryological development, the structures developing in association with the third pouch undergo caudal migration, which is often excessive. The third pouch leap-frogs over the fourth and divides, leaving a discrete mass of parathyroid tissue on each side, usually within 2 cm of the lower pole of the thyroid (Fig. 3), and the thymus as a bilobed structure in the anterior mediastinum overlying the great vessels. Excessive fragmentation produces thymic rests and accessory parathyroid glands. Failure to fragment results in intrathymic parathyroid tissue, which accounts for the location of 20 per cent of parathyroids, half of which are bilateral. The area where ectopic lower parathyroids may be found is shown in Fig. 4.

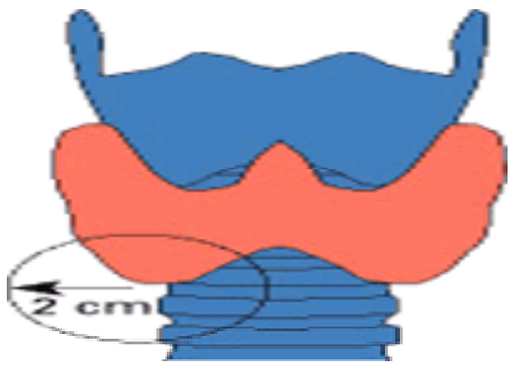


Fig. 3. Normal site of the lower parathyroid gland.

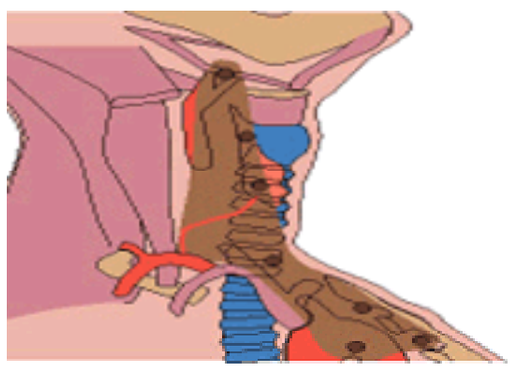


Fig. 4. Brown shading indicates area where ectopic lower parathyroids may be found; black dots as for Fig. 2.

Blood and nerve supply

The arterial supply to the upper and lower parathyroids is almost exclusively from the inferior thyroid artery on each side; indeed, tracing out the terminal branches of this artery can aid recognition of the gland. Although the final arterial blood supply to each gland is an end artery, collateral circulation occurs proximal to the hilum, this being derived principally from the superior thyroid artery but also from oesophageal and tracheal vessels. The nerve supply arises either directly from the sympathetic chain (superior and middle cervical ganglia) or indirectly by a plexus on the back of the thyroid.

Macroscopic and microscopic appearance

The parathyroid glands vary from a tan to yellow or reddish-brown colour, depending on the fat content and vascularity. The fat content increases after puberty and may comprise up to 80 per cent of the total volume in elderly individuals. Unlike lymph nodes and fat lobules, with which parathyroids are easily confused, the consistency is softer (like blancmange) and they are typically tongue shaped. The base comprises the vascular hilum and there is often a sharp leading edge to one side. Normal and abnormal glands are frequently found in a fatty envelope, within which they are freely mobile and can be displaced by probing. Microscopically, two main cell types are seen in parathyroid tissue, chief and oxyphil cells. The chief cells are of three types: light chief cells rich in glycogen with few secretory granules, dark chief cells with little glycogen but with many secretory granules, and water clear cells, which are large (up to 40 µm diameter) and have clear, vacuolated cytoplasm rich in glycogen. The oxyphil cells are variable in size, appear after puberty, and have pink cytoplasm and small dense nuclei.

Pathophysiology

Much of this topic has already been discussed in detail in Chapter 20.3.1, but some aspects of the pathophysiology bear repeating from the surgical perspective.

Primary hyperparathyroidism

Aetiology

The cause of primary hyperparathyroidism is unknown but the prevalence with increasing age, especially in females, is so high that it almost seems to be part of the ageing process and may be influenced by the menopause. Radiation to the head-and-neck area in childhood increases the risk of adenoma and hyperplasia of the parathyroids developing later in life. Hyperplasia is a notable feature of the multiple endocrine neoplasia (**MEN**) syndromes described in previous sections (20.1 and 20.2).

Incidence

Long-term studies of large, well-defined populations in Sweden and the United States indicate an incidence of 10 cases/100 000 population per year in those aged under 40 years. Over the age of 60 years, however, there is a steep rise in incidence: 92/100 000 per year for men, 188/100 000 for women. Symptomatic hyperparathyroidism tends to be diagnosed as a result of illness rather than screening, and is highest in the third to fifth decades.

Pathology (Table 1)

In over 80 per cent of cases, primary hyperparathyroidism is due to a single, chief-cell adenoma affecting one parathyroid gland. The pathologist distinguishes an adenoma from a normal parathyroid gland by finding a rim of compressed normal parathyroid tissue to one side of the adenoma (Fig. 5). Within the cells of the adenoma itself there is nuclear pleomorphism and the cytoplasm contains little fat. Rarely the adenoma consists of oxyphil or water clear cells. Multiple adenomas are well documented but occur infrequently. Hyperplasia of all four parathyroids is typical of secondary hyperparathyroidism. However, hyperplasia is responsible for the primary disease in about 5 to 10 per cent of cases and may affect two or more glands. The chief cells are again most frequently involved, but oxyphil- and water clear-cell hyperplasia may be seen. In contrast to an adenomatous gland, a hyperplastic parathyroid shows no rim of normal tissue and the overall appearance is uniform, with a marked increase in parenchymal cells and few lipocytes. Nodular hyperplasia is a feature strongly associated with the MEN syndromes.

Pathological cause	*Oxford series (%)*	*Worldwide series (%)*
Single adenoma	84.0	70–93
Multiple adenoma	2.0	2–24
Hyperplasia	14.0	3–26
Carcinoma	1.5	1–4

Table 1 Pathological entities responsible for primary hyperparathyroidism

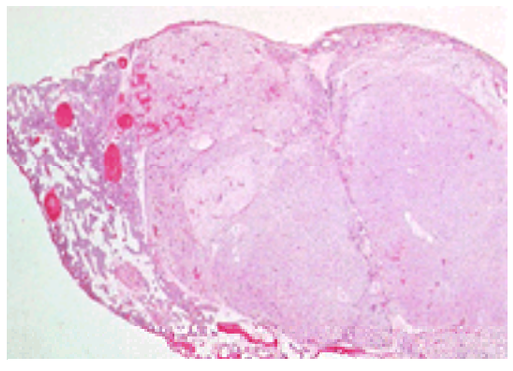


Fig. 5. Parathyroid adenoma showing compressed rim of normal parathyroid tissue around the adenoma.

Parathyroid carcinoma is the rarest cause of primary hyperparathyroidism (fewer than 3 per cent of cases) and its recognition depends more on the surgeon than on the pathologist because the diagnostic criteria are more clinical than histological. Typically a carcinoma is firm to hard, grey in colour, and adherent to adjacent tissues. The capsule of the gland eventually becomes invaded, and metastases develop frequently in the local lymph nodes. Distant metastases in liver and bone are a late feature and indicate terminal disease.

Disturbance of physiology

This arises because of increases in both the circulating parathyroid hormone (**PTH**) and calcium. The production of PTH is said to have become ‘autonomous' and fails to switch off as the plasma calcium rises. The hypercalcaemia causes the renal tubular absorption of calcium to be increased, so that at any given filtered load more is lost in the urine and reflected in an increase in the 24-h urinary calcium. Bone turnover is also increased, with reabsorption exceeding deposition, whilst more calcium is absorbed from the intestine because of stimulated calcitriol production in the kidneys.

Secondary hyperparathyroidism

Aetiology

This condition results from prolonged hypocalcaemia and is caused most frequently by chronic renal failure. Vitamin D deficiency and gluten-sensitive enteropathy are other provocative states.

Incidence

No precise figures are available, but it is inevitable that patients with endstage renal failure will develop secondary hyperparathyroidism unless prophylactic therapy with vitamin D analogues is given.

Pathology

Typically all four glands are enlarged and histologically they show the same features seen in hyperplasia due to primary hyperparathyroidism. The chief cells are the predominant cell type affected.

Disturbance of physiology

When the kidney starts to fail the reduction in glomerular filtration causes an increase in plasma phosphate. This in turn produces a reduction in the plasma ionized calcium to which the parathyroids respond by undergoing hyperplastic change.

Tertiary hyperparathyroidism

This condition affects those patients with long-standing secondary hyperparathyroidism who develop the same sort of autonomous gland function described in primary hyperparathyroidism. The physiological sequelae are similar, with hypercalcaemia coexisting with a raised PTH. It is most frequently seen in patients on long-term dialysis for chronic renal failure and after renal transplantation. The glands are hyperplastic but may additionally produce adenomas, sometimes referred to as quarternary hyperparathyroidism.

Pseudohyperparathyroidism

This condition is mentioned because it gives rise to biochemical disturbances similar to those seen in primary hyperparathyroidism. Oat-cell and squamous carcinoma of the lung, head, and neck, and carcinoma of the kidney and ovaries may all rarely produce parathyroid-like, adenolate cyclase-stimulating proteins. These interact with the PTH receptors and can mimic some of the effects of primary hyperparathyroidism.

Clinical features of hyperparathyroidism

The widespread use of multichannel autoanalysers for routine biochemical screening has brought to light an increasing number of individuals with hypercalcaemia, who, on further investigation, are shown to have hyperparathyroidism. Frequently the diagnosis is made so early that signs and symptoms have not had time to develop. Mildly symptomatic or asymptomatic patients identified in this way now account for more than 50 per cent of the total in most large series. In the 12 years preceding 1980, 54 per cent of the 200 patients coming to surgery in Oxford had originally presented with significant bone and renal disease. In the following decade a further 200 patients underwent surgery but the number of those with symptomatic bone and renal disease had fallen to 35 per cent. The time-honoured rhyme ‘bones, stones, abdominal groans, and psychic moans' is still a useful prompt for the symptoms associated with hyperparathyroidism.

Bones

Disturbance of calcium metabolism causes vague bony aching and arthralgia, especially in secondary hyperparathyroidism. It becomes acute and site-specific if the

weakened bone fractures. Very severe arthralgia is a feature of chondrocalcinosis (pseudogout), in which the surface of the articular cartilage becomes the site of metastatic calcification.

Stones

Long-standing, severe hypercalcaemia may present with renal pain. The increased filtered load of calcium and the passage of alkaline urine lead to the formation of renal calculi. Less frequently, nephrocalcinosis occurs, with calcification of the renal parenchyma, i.e., outside the collecting system. The stones predispose to renal colic, urinary infection secondary to obstructive uropathy, and renal failure. Once established, renal damage caused by hyperparathyroidism is irreversible and the stones tend to persist even after calcium homeostasis is restored by successful surgery.

Abdominal groans

Mild abdominal pain is often due to large-bowel colic related to the chronic constipation brought on by dehydration and hypercalcaemia. More severe epigastric pain arises in 5 to 10 per cent of patients with hyperparathyroidism, owing to peptic ulceration. This disorder may be causally related if the hypothesis is true that hypercalcaemia stimulates gastrin production and thereby increases basal gastric acid secretion. Certainly, ulcer symptoms usually remit after successful parathyroid surgery and studies have shown that the gastric acid secretion returns to normal. There is a rare association, in patients with MEN type I, with a Zollinger–Ellison-type syndrome, which probably represents an extreme gastrin response. Because surgical removal of the diseased parathyroid glands in such patients results in cure without the need for gastric surgery, the term pseudo-Zollinger–Ellison syndrome has been proposed. The link between hyperparathyroidism and pancreatitis, both acute and chronic, is well described. It may be explained simply on the basis of metastatic calcification within the common bile duct causing acute pancreatitis (common channel theory) or throughout the pancreas in the chronic form. Alternatively, it has been postulated that hypersecretion of glucagon by the a-cells of the inflamed pancreas causes hypoglycaemia, which stimulates the parathyroids to become hyperplastic.

Psychic moans

Mental symptoms range from subtle mood change and behavioural disorder to marked organic psychosis and dementia. Loss of concentration, mild depression, and lassitude are frequent.

Hypercalcaemia *per se* induces anorexia, nausea, vomiting, and thirst, and is associated with polydipsia and polyuria. The most extreme form of hypercalcaemic syndrome is seen in acute hypercalcaemic crisis—a fulminating hyperparathyroidism state in which all the above symptoms are exaggerated. Unless this medical and surgical emergency is recognized and treated quickly the outcome can be fatal.

Physical signs

There are few signs associated with primary hyperparathyroidism but if the calcium is markedly elevated a parathyroid tumour may be palpable in the neck. Band keratopathy is occasionally visible on slitlamp examination of the cornea (Fig. 6). Proximal myopathy with related wasting and motor weakness is a rare finding. Diffuse osteitis fibrosa cystica with gross skeletal deformity and loss of height, first described by von Recklinghausen in 1891, is unlikely ever to be seen again in developed countries, due to earlier diagnosis. However, focal swellings of bone, notably of the lateral end of the clavicles due to the presence of ‘brown tumours', are occasionally encountered (Fig. 7). In patients with secondary hyperparathyroidism, soft-tissue calcification is common, especially in arteries, muscles, and subcutaneous tissues particularly around joints, and scratch marks due to pruritus are frequent. The physical signs of tertiary hyperparathyroidism include all the foregoing, pruritus being even more severe, with skin necrosis due to subcutaneous calcium deposition occasionally seen.

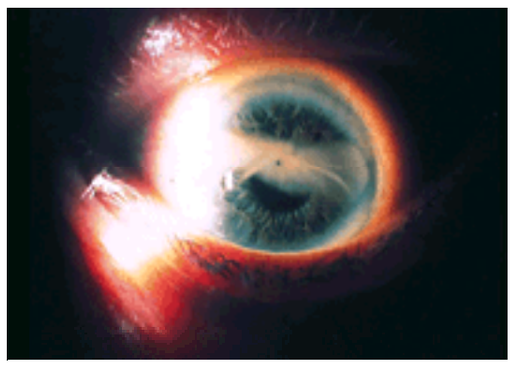


Fig. 6. Corneal calcification/band keratopathy in hyperparathyroidism.

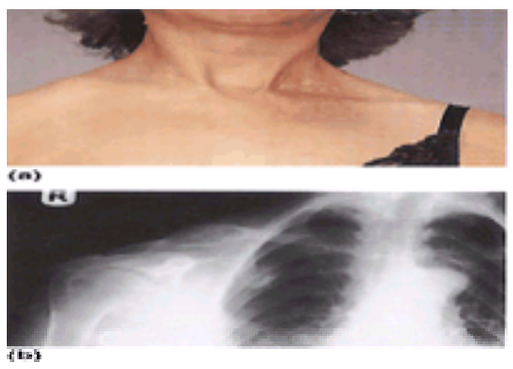


Fig. 7. Swelling of the clavicle due to a ‘brown' tumour; note previous radiotherapy to the neck, which may have induced hyperparathyroidism.

Investigation of hyperparathyroidism

The single most important criterion for a diagnosis of primary hyperparathyroidism is a persistent and significant increase in the plasma calcium concentration. However, primary hyperparathyroidism accounts for only approximately 20 per cent of all symptomatic patients with hypercalcaemia, and a careful history and physical examination are necessary to exclude other causes, notably metastatic bone disease (primary breast, lung, etc.) and multiple myeloma. A more comprehensive list of causes of hypercalcaemia is summarized in Table 2. Distinguishing primary hyperparathyroidism from these other causes rests on three key biochemical tests.

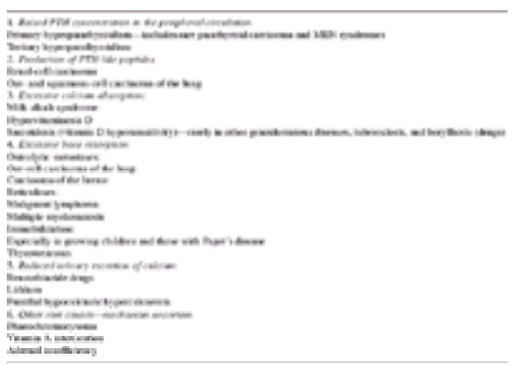


Table 2 Causes of hypercalcaemia (some conditions affect the plasma calcium by several mechanisms)

1. *Plasma calcium (normal range 2.35–2.55 mmol/l)* A minimum of three estimations should be performed, and where hypercalcaemia is modest it becomes more important to draw an uncuffed venous sample when the patient has been fasted. Normocalcaemic hyperparathyroidism is an accepted entity and tends to affect patients with recurrent renal-stone formation. This small group exhibits the rare biochemical profile of a calcium concentration usually towards the top of the normal range but with an inappropriately high PTH.
2. *Plasma albumin (normal range 35–50 g/l)* This should be estimated at the same time and under the same circumstances as for plasma calcium. Albumin is the main calcium-binding protein in the plasma, so alternations in the albumin concentration have to be taken into account and the plasma calcium corrected appropriately. A simple and reliable method is the addition of 0.02 mmol/l of calcium for every 1 g/l of albumin below 40 and the subtraction of 0.02 mmol/l for every 1 g/l above.
3. *Immunoradiometric intact PTH assay (normal range 0.9–5.4 pmol/l)* This should be estimated at the same time as the two foregoing tests. Two-site immunoradiometric assays measuring the intact PTH molecule in the plasma have only recently become standardized and more readily available in kit form. In the past, radioimmunoassays using antisera raised either against the mid-, carboxy-, or the N-terminal fragments of the PTH molecule could erroneously measure similar PTH polypeptides produced by some occult malignancies. Squamous and oat-cell carcinoma of the lung and renal-cell carcinoma, as mentioned earlier, are notable examples of tumours with this ability to elaborate PTH-like provoking hypercalcaemia in the absence of bony metastases.

Complementary tests that support the diagnosis and identify patients who have associated deleterious effects of hyperparathyroidism are as follows.

1. *Plasma phosphate (normal range 0.8–1.45 mol/l)* As the tubular reabsorption of calcium increases, that for phosphate decreases, with more phosphate lost in the urine relative to calcium. The plasma concentration falls, notably in patients with advanced disease. Conversely, elevated plasma phosphate in the presence of hypercalcaemia and normal renal function suggests bony metastases or vitamin D intoxication.
2. *Plasma creatinine (normal range 70–150 µmol/l)* This is a useful indicator of renal function and if clearance falls significantly in a patient with hyperparathyroidism (by more than 30 per cent) who has otherwise been asymptomatic, it signals renal damage and the need for surgical intervention.
3. *Alkaline phosphatase (normal range 80–250 iu/l)* This is a useful indicator of bone disease; if in the normal range, significant skeletal involvement is ruled out and there is no need for extensive radiological investigations of the skeleton.
4. *Plasma magnesium (normal range 0.75–1.05 mmol/l)* Hypomagnesaemia has been reported in up to 15 per cent of patients with primary hyperparathyroidism and may need to be corrected especially after the patient has been operated upon (see Chapter 20.2.1).
5. *Urinary calcium (normal range 2.5–7.5 mmol/l per 24 h)* An increase in the 24-h urinary calcium supports the diagnosis of hyperparathyroidism but a markedly elevated urinary calcium above 9 mmol/l characterizes non-PTH-mediated hypercalcaemia. Perhaps the most important value of this test is to identify the rare patient with familial hypocalciuric hypercalcaemia. This condition frequently presents at a young age with symptoms of hypercalcaemia, but the 24-h urinary calcium is less than 2.5 mmol/l and the patient has no parathyroid pathology. Mention of a relative who has undergone a failed neck exploration should alert one to the strong possibility of this condition.

The foregoing biochemical tests diagnose and quantify primary hyperparathyroidism in 95 per cent of suspected cases but the following tests are sometimes required to reinforce the diagnosis when the data are marginal or conflicting.

1. *Renal tubular absorption of phosphate (normal range 0.8–1.35)* This is a sophisticated method of identifying the effect of excessive PTH on the renal tubular absorption of phosphate. Simultaneous measurements of urinary phosphate, plasma phosphate, plasma creatinine, and urinary creatinine enable the relation between filtered load and renal excretion of phosphate, i.e., phosphate clearance, to be compared with creatinine clearance. It is expressed as a ratio, referring to a standard nomogram. In primary hyperparathyroidism the ratio is below 0.8.
2. *Urinary cAMP (normal range < 2.5 nmol/100 ml of glomerular filtrate)* Raised concentrations have been claimed in nearly 90 per cent of patients with primary hyperparathyroidism in some series. False-positive and false-negative results reported by others limit the usefulness of the test.
3. *Hydrocortisone suppression/Dent test* This test has largely been rendered obsolete since the introduction of immunoradiometric assays of the intact PTH molecule. Administering high doses of hydrocortisone (40 mg three times a day for 10 days) does not usually lower the hypercalcaemia produced by primary hyperparathyroidism, but if the corrected plasma calcium falls on day 4, 7, and 10 into the normal range or more than 0.25 mmol/l, the hypercalcaemia is likely to be due to causes other than primary hyperparathyroidism.

Radiological findings are rarely in themselves diagnostic of hyperparathyroidism and more usually complement the biochemical data. They may reinforce the need to offer surgery. The most frequent radiographic changes seen in both primary and secondary hyperparathyroidism are those affecting the skeleton.

1. *Bony radiographic changes* The overall picture is that of bone density loss (osteopenia) and subperiosteal reabsorption of bone. These are seen especially in the middle phalanges of the hand on the radial side, together with loss of the terminal phalangeal tufts (Fig. 8). Subperiosteal absorption is seen in many other bones, especially the lateral ends of clavicles, upper tibia, pubic symphysis, and sacroiliac joints. Localized areas of osteoclastic bone resorption with marrow fibrosis are seen in advanced bone disease. In the most severe form, bone cysts and ‘brown' tumours consisting of massive aggregations of osteoclasts (‘osteoclastoma') together make up osteitis fibrosa cystica (von Recklinghausen's disease), which, as remarked earlier, is now largely a medical curiosity. Generalized bone loss at a variety of sites has attracted some colourful descriptive nomenclature such as rugger-jersey spine (Fig. 9) and pepper-pot skull (Fig. 10), all of which aptly describe the osteoporosis and small osteolytic lesions. Absorption of the lamina dura around the teeth is reported but is an unconvincing radiological sign.



Fig. 8. Radiograph of the hand showing subperiosteal absorption of bone and loss of terminal phalangeal tufts.

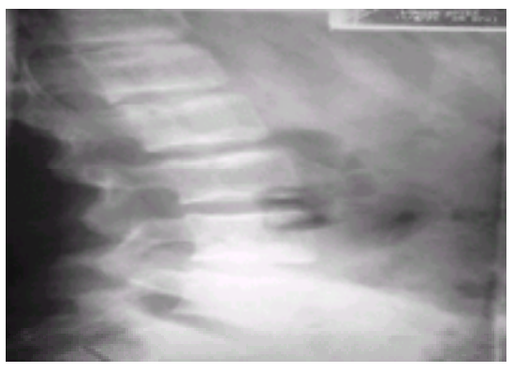


Fig. 9. Radiograph of ‘rugger-jersey' spine.



Fig. 10. Radiograph of ‘pepper-pot' skull.

2. *Renal radiographic changes* Discrete renal stones or diffuse calcification throughout the renal parenchyma (nephrocalcinosis) may be found and should be looked for on plain abdominal radiographs in patients with hyperparathyroidism, even in the absence of symptoms. Sometimes the findings are unexpectedly dramatic (Fig. 11). An ultrasonographic examination may be indicated in the assessment of upper-tract obstruction and overall renal function.

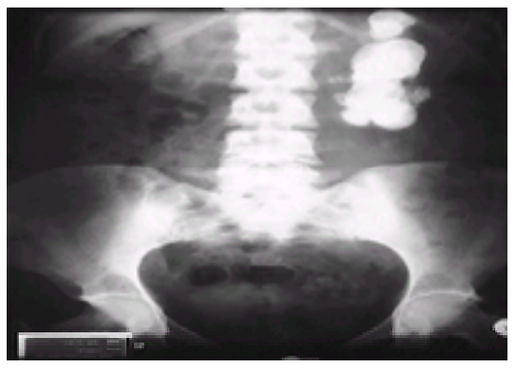


Fig. 11. Plain radiograph showing massive nephrolithiasis secondary to hyperparathyroidism.

3. *Metastatic calcification* Mention has been made of the calcification that may be felt or seen radiologically in the skin and around blood vessels and joints. This phenomenon is readily confirmed on radiography of the relevant part.

Localization of parathyroid tumours

Opinion is divided about the need to attempt localization of parathyroid tumours before the neck is explored surgically for the first time. In that situation an experienced surgeon can identify the pathology without any assistance in over 90 per cent of patients. None of the localization procedures to be described approaches such a degree of accuracy and some are no better than 50 per cent accurate. Unfortunately, localization techniques are least helpful when most needed, i.e., when the tumour is small (under 1 cm in diameter and 100 mg in weight), ectopic (close to the heart and great vessels), or when the patient has a bulky neck due to obesity or coexistent thyroid enlargement. In view of the time and cost involved in such studies, they are hard to justify except for patients who have had a previous unsuccessful neck exploration. The following localizing procedures are of proven value.

Selective venous sampling with PTH assay

This is the single most reliable method. It is based on the measurement of PTH concentration in samples obtained from the superior vena cava, innominate vein, internal jugular, and the multiple small veins draining the thyroid and mediastinum. The sampling catheter is introduced via the femoral vein and each sample numbered with the site noted precisely. A map can then be constructed, plotting the samples, to look for the vein(s) draining the source of highest PTH concentration on radioimmunoassay (Fig. 12). Although good at lateralization in single-gland disease (concentra-tions twice normal to that found in the background peripheral circulation being diagnostic), it cannot distinguish between an adenoma in the neck or one that has descended into the mediastinum leaving its venous drainage behind. Raised PTH concentrations in all the thyroid veins suggest multigland hyperplasia. The drawback to this technique is that it is time-consuming, expensive, and highly operator-dependent.

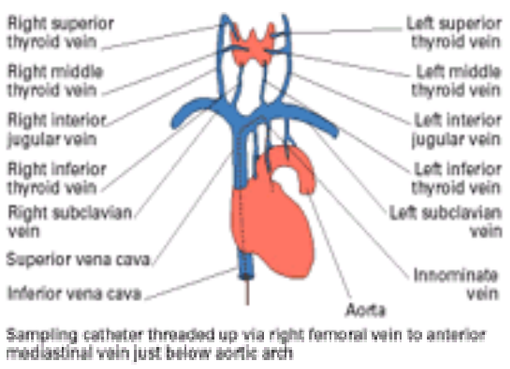


Fig. 12. ‘Map' obtained by selective venography and venous sampling showing site of highest PTH value found under aortic arch (see Table 3, sample 8).

Sample no.	*Concentration (pmol/l)*
1	17
2	400
3	18
4	20
5	20
6	21
7	16
8	1500
9	415
10	12
11	14
12	16
13	18
14	10

Table 3 PTH intact assay

Thallium/technetium subtraction scanning

A double-tracer scintigram is performed using technetium-99m as sodium pertechnetate for imaging the thyroid and then thallium-201 (as thallous chloride) for imaging the thyroid and parathyroid glands simultaneously. A gamma-camera with on-line computer facilities subtracts the pertechnetate image from the scatter-corrected

thallium image. Abnormal parathyroid tissue may then be demonstrated very convincingly (Fig. 13). However, the success rates reported from various centres have varied widely. False-positive results may be produced by thyroid adenomas and abnormal lymph nodes. False-negative results arise in cases of parathyroid hyperplasia and tumours close to the heart and great vessels that are lost in the normal high concentration of isotope in the central mediastinum. Recently, thallium/technetium subtraction scanning has given way to scanning with sestamibi.

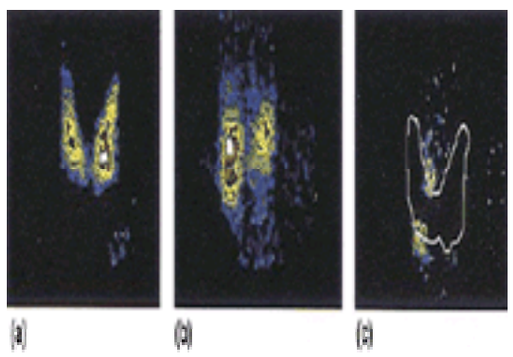


Fig. 13. (a) A technetium scan; (b) a thallium scan; (c) subtracted image showing parathyroid adenoma.

Sestamibi scanning

Developed in the 1990s, sestamibi scanning has rapidly gained acceptance as being markedly superior to thallium/technetium subtraction scanning in imaging diseased parathyroid glands. Sestamibi is a small protein labelled with technetium-99m that localizes specifically in overactive parathyroid glands. It is sensitive in identifying parathyroid adenomas (90 per cent) but less so in identifying multigland disease such as hyperplasia or double adenoma (Fig. 14). Sensitivity can be increased by single photon-emission computed tomography, which is especially helpful in the precise location of mediastinal disease. In the United States the use of sestamibi to perform radioguided, minimally invasive parathyroidectomy has been reported. It is, however, an expensive test, and its use in the United Kingdom has so far been confined to localization of ectopic parathyroid tissue or ‘re-do' parathyroid operations where a focused operation is deemed more suitable.

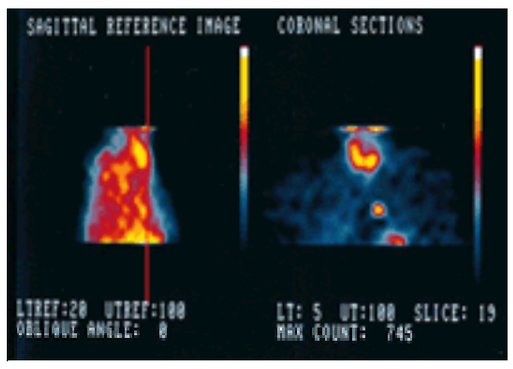


Fig. 14. [$^{99}Tc^{m}$] sestamibi scan demonstrating a mediastinal parathyroid adenoma.

Ultrasonography

This technique has the merit of being quick and inexpensive but cannot scan behind the sternum and is confined to identification of parathyroids that may or may not be in the neck. It lacks the resolution to pick up small tumours (less than 0.5 cm) but can raise suspicion with intrathyroidal tumours, those in the carotid sheath, and those between the trachea and oesophagus.

Computed tomography (CT) and nuclear magnetic resonance imaging (MRI)

There is no consensus on which of these imaging techniques is best, but used in conjunction with selective venous sampling one or other may improve localization for the difficult reoperative case (Fig. 15).



Fig. 15. MRI of the neck showing mediastinal adenoma.

Indications for surgery

Asymptomatic primary hyperparathyroidism

This is one of the controversial areas of hyperparathyroidism management and opinion is divided. Unfortunately, only limited information is available about the natural history of primary hyperparathyroidism and, although 20 per cent of asymptomatic patients in Purnell's series reported from the Mayo Clinic came to surgery over a 10-year follow-up period, not all were operated upon because of the adverse effects of the disease. Over 50 per cent, indeed, showed no deterioration over 5 years. The longer-term effects on renal function, bone density, hypertension, and survival itself remain unclear but there is increasing evidence from Sweden of higher mortality from cardiovascular disease in untreated patients with primary hyperparathyroidism. Meanwhile, asymptomatic patients are being diagnosed with increasing frequency, as mentioned before, and in the absence of prospective trials randomizing patients to surgery or observation alone (attempted but failed in the United Kingdom), a policy has emerged in the United States and the United Kingdom that identifies those patients for whom surgery is most clearly indicated (Table 4).

1. Persistent hypercalcaemia: corrected plasma calcium > 2.9mmol/l
2. Osteopenia (bone densitometry) or alkaline phosphatase (bone origin) > 350iu/l
3. Renal calculi or nephrocalcinosis on abdominal radiograph
4. Renal impairment: creatinine clearance down 30 per cent
5. Patients under the age of 50 in whom long-term follow-up is impractical

Table 4 Indications for treatment in (mild) asymptomatic primary hyperparathyroidism

Symptomatic primary hyperparathyroidism

There is good evidence that treatment is effective. The incidence of renal stones and infection decreases, osteitis fibrosa cystica improves, subperiosteal resorption resolves, and patients with bone and joint pain are dramatically relieved. Where peptic ulceration is present it regresses, and psychological disturbance may well resolve. Hypertension, which is found in 40 per cent of sufferers, is not in itself an indication for treatment, and indeed deterioration after parathyroid surgery can occur. However, there should be no hesitation in advising surgery unless there are overwhelming medical contraindications. Alternative but less satisfactory options include embolization and CT- or ultrasound-guided laser or alcoholic destruction of an adenoma, if it can be confidently localized.

Secondary and tertiary hyperparathyroidism

The majority of patients whose parathyroids undergo hyperplastic change in response to chronic renal failure, or rarely chronic intestinal malabsorption, can be controlled non-operatively. Correction of the hypocalcaemia with vitamin D, which enhances calcium absorption from the gut, a high calcium intake, restricted dietary phosphate, and the use of phosphate-binding agents such as aluminium hydroxide are effective standard medical measures. Where the patient is undergoing dialysis, secondary hyperparathyroidism can be controlled by increasing the concentration of calcium in the dialysate, having first lowered the serum phosphate to avoid ectopic calcification. Surgery is indicated when these measures fail and the patient has intractable bone pain, pruritus, or when soft-tissue deposition has led to ischaemic skin necrosis and joint pain. Tertiary or autonomous hyperparathyroidism tends to be less responsive to the foregoing treatment; where the patient is not responding and is unlikely to receive a renal transplant, total parathyroidectomy is gaining momentum as the best option. An alternative strategy is either to perform a subtotal parathyroidectomy and accept a high recurrence rate or to remove all the parathyroids in the neck and cryopreserve all but 60 mg of tissue, which is then implanted into the forearm (see below for precise details of this technique). The hypercalcaemia seen after a successful renal transplant can usually be reversed by oral phosphate supplements and is usually slowly self-limiting.

Operative surgery on the parathyroids

Preoperative preparation

In general this is identical to the broad recommendations for thyroid patients (see Chapter 20.2.2) but the following special measures apply. Those with primary hyperparathyroidism and raised alkaline phosphatase, and most sufferers from renal bone disease (secondary hyperparathyroidism), should receive 2 to 4 mg per day of 1a-hydroxycholecalciferol 2 to 3 days before surgery. This practice helps to combat the profound hypocalcaemia and tetany that is unleashed when the parathyroid disease is removed and the ‘hungry bones' then take up all available calcium. Grossly hypercalcaemic patients, some of whom may present in crisis with coma, delirium, anorexia, vomiting, and abdominal pain, pose a particularly difficult diagnostic challenge. Where rapid investigation excludes other possible causes of hypercalcaemia, fluid and electrolyte balance must be corrected as a matter of urgency. If large volumes of intravenous saline fail to reverse the dehydration oliguria and azotaemia provoked by the vomiting and polyuria, then a vigorous diuresis stimulated by frusemide (furosemide) or ethacrynic acid is appropriate. If renal function is severely compromised, then peritoneal or haemodialysis may need to be considered.

Preoperative localization studies have already been discussed and are only advocated in those patients who have undergone a previous neck exploration. However, intraoperative localization is greatly assisted by giving the patient a vital dye, methylene blue, which selectively stains parathyroid tissue and makes recognition easier. It is given in a dose of 5 mg/kg body wt, diluted in 500 ml dextrose saline, infused over an hour preceding neck exposure. A larger dose up to 10 mg/kg may be indicated in obese patients, based on lean muscle mass, but care is needed to avoid cardiotoxic effects. The intensity of staining is most impressive in four-gland hyperplasia due to secondary hyperparathyroidism (Fig. 16), but many adenomas are equally recognizable, appearing navy blue/purple depending on the degree of subcapsular haemorrhage (Fig. 17). Normal glands have a subtle, pale greenish tinge.

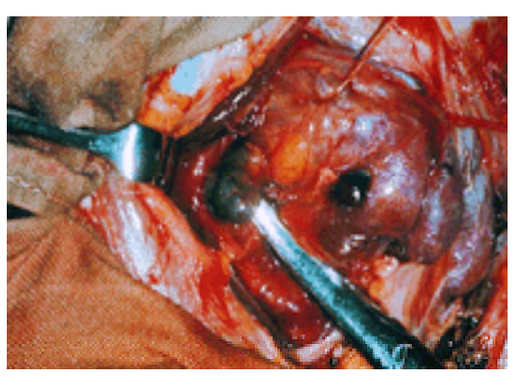


Fig. 16. Methylene blue staining of hyperplastic parathyroids.

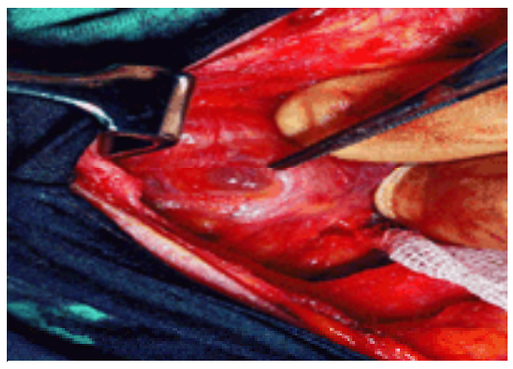


Fig. 17. Methylene blue staining of an adenoma.

Exposure and technique of exploration

This is identical to that described for the thyroid (Chapter 20.2.2), but is refined to meet the objective of identifying the four parathyroids and, where they are all normal,

a supernumerary abnormal gland. In meeting this objective there is no substitute for a patient, systematic routine of neck exploration unconstrained by time. Meticulous haemostasis is crucial, especially when the thyroid is freed from the strap muscles and the space is opened between the gland and carotid sheath. Where present, the middle thyroid veins must be divided to allow full mobilization of the thyroid lobes forwards and medially. Finger retraction on a gauze swab is preferable to the use of grasping forceps, which can traumatize the thyroid and lead to bleeding. If blood extravasates it stains the local tissues and makes identification of the parathyroids much more difficult. The inferior thyroid artery and recurrent laryngeal nerve are identified and gently retracted with linen or Silastic vascular slings. These not only reduce the risk of damage to the nerve, which may be adherent to the capsule of a parathyroid adenoma, but also help in the identification of the parathyroids, which, as stated earlier, frequently derive their blood supply from one of the branches of this artery. Exposure of the superior parathyroid gland may require incision of the fascia binding the thyroid posteromedially to the trachea. If this is extended up to the level of the superior thyroid vessels, the whole upper pole can be rolled forward. This manoeuvre also provides excellent access to ectopic superior parathyroid sites without the need to divide the superior thyroid pedicle. Some 90 to 95 per cent of upper parathyroids, however, are closely related to the inferior thyroid artery as it breaks up on the posterior surface of the gland, i.e., within a 2-cm radius of this vessel. Not infrequently it is tucked round behind the upper branches (see Fig. 1). The inferior parathyroid glands are less constant in position, but 80 per cent are still found within a 2-cm radius of the lower pole of the thyroid and many are quickly identified on the anterolateral aspect (see Fig. 3).

By this stage the experienced surgeon will usually have identified one or more normal parathyroids, which narrows down the field of search for the missing pathological gland(s). Close inspection of the thyrothymic ligament, which includes the inferior thyroid veins, of the lower polar fat, and of the thymic horns as they extend into the anterior mediastinum will reveal the vast majority of these ectopic lower glands (20 per cent). If not, the next step is to search between the tracheo-oesophageal groove and behind the oesophagus. If a lower parathyroid is still elusive, then it may lie within the thyroid (1–2 per cent of cases). Embryologically this is only possible if the parathyroid has become indented into a cleft in the thyroid, so close inspection and palpation should detect a ‘bulge' and opening up a cleft or incising the lower pole anterolaterally will display the missing gland without the need to do a blind thyroid lobectomy. Failing that, ectopic parathyroid may reside within the carotid sheath, which should be opened right up to the carotid bifurcation near the angle of the jaw, especially if a suspected abnormal upper parathyroid is still missing. A transcervical thymectomy is the last manoeuvre to be performed and may reveal the missing parathyroid when the thymus is sectioned. The entire routine described may have taken several hours, but, if unsuccessful (only 1–2 per cent of patients), a formal mediastinal exploration via a median sternotomy is not in the patient's best interest at this stage. When hypercalcaemia is severe it can always be moderated until urgent localization studies are performed. The basic requirement after sternotomy is to clear all remaining thymic and fatty tissue, notably around the innominate veins, great vessels, and within the aortopulmonary window.

Surgical strategy for primary hyperparathyroidism

In the majority of patients (80 per cent) a single adenoma is responsible for the condition and removal will be curative. The tumour is removed with considerable care, since there is a real danger of autotransplantation and recurrent disease if the capsule is ruptured and cells are spilt, resulting in parathyromatosis. When the pedicle can be located easily it is clipped between mosquito forceps and the gland then gently lifted up and dissected clear from the surrounding connective tissues with iris scissors. The pedicle is ligated with a non-degradable suture. Frozen-section microscopy confirms the diagnosis. Two main strategies are then possible. (1) Identify the three other parathyroids, which will involve exploration of the contralateral side, and, relying on clinical judgement alone, conclude, if they look normal, that the patient has single-gland disease and do nothing more. This strategy can be justified if the surgeon is experienced and acknowledges the fact that even if a macroscopically normal parathyroid appears hypercellular on biopsy, this is of doubtful clinical importance. The majority of such patients are cured by removing the adenoma alone and more aggressive resection runs the risk of an increased incidence of hypoparathyroidism. If other macroscopically enlarged glands are found, this raises the possibility of multigland disease, and biopsy helps identify a second adenoma in 2 per cent of patients or genuine two- to four-gland hyperplasia in 10 to 15 per cent of patients with primary hyperparathyroidism. (2) Identify the other ‘normal' gland on the ipsilateral side and submit it to frozen section. If normal the contralateral side of the neck is left undisturbed (Tibblin strategy). This strategy chooses to ignore the 1 per cent chance of another adenoma being present on the other side of the neck and neglects the chance to double-check whether the patient might have multigland hyperplasia. For these reasons this strategy cannot be recommended.

When all four glands are enlarged due to hyperplasia, a subtotal (three-and-a-half) parathyroidectomy is recommended. Initially the two largest glands are removed, and then one-half of each of the remaining glands is excised, ensuring that the half left behind has an intact blood supply. After a brief interval these two halves are inspected and the one that appears less viable is removed and the other is left *in situ*, clearly marked with a titanium clip. If the patient becomes hypercalcaemic at a later date, the clip facilitates recognition for further resection.

Surgical strategy for secondary and tertiary hyperparathyroidism

The challenge with these patients who require surgery is to identify all the hyperplastic glands (up to 6 per cent may have a supernumerary fifth gland), and excise sufficient parathyroid tissue to relieve the symptoms of renal osteodystrophy but leave an adequate amount to preserve normal parathyroid function. There is controversy about how this goal is best achieved. There are those who advocate subtotal parathyroidectomy and others who promote total parathyroidectomy. The disadvantage of subtotal parathyroidectomy, especially for patients with tertiary hyperparathyroidism whose calcium and PTH are both elevated, is the high recurrence rate. If renal transplantation is likely in due course, which would slowly correct the patient's hyperparathyroidism, then subtotal parathyroidectomy avoids the risk of unmasking vitamin D-resistant osteomalacia, which is sometimes encountered after total parathyroidectomy and is very difficult to treat. An alternative strategy that theoretically overcomes both these problems is to remove all the parathyroid tissue from the neck, implant 50 to 60 mg in the forearm, and to cryopreserve the rest. If the forearm transplant is too generous and hypercalcaemia persists or recurs, then the surgeon has easy access to reduce the volume of transplanted tissue. If the patient becomes hypocalcaemic, then the cryopreserved tissue is brought out of storage and some more is implanted. Whilst freshly transplanted parathyroid cells function well in over 90 per cent of patients, implanted cryopreserved tissue functions in no more than 70 per cent of patients. The precise technique is to remove all parathyroid tissue from the neck, confirm this histologically, and then dice one parathyroid on a flat surface into 1-mm cubes with a scalpel. Some 50 to 60 mg is transplanted into three or more separate muscle pockets in the brachioradialis muscle of the non-dominant forearm (Fig. 18). This is a delicate task requiring fine instruments to avoid crushing the fragments and haematoma, which compromise the viability of the transplant. Each pocket is closed with a fine, non-absorbable suture such as Prolene, which prevents extrusion and is easily recognized if removal is required later. The rest of the parathyroid tissue from the other three or four glands is cut into 1 × 1 × 3 mm slices and placed in polypropylene vials containing tissue-culture medium, antibiotics, dimethyl sulphoxide, and autologous serum. Freezing to -80°C is commenced quickly in a controlled manner after harvesting and the tissue is then stored in a liquid nitrogen freezer. The process is reversed in a 37 °C water bath and successful transplantation has been reported as long as 18 months after storage.

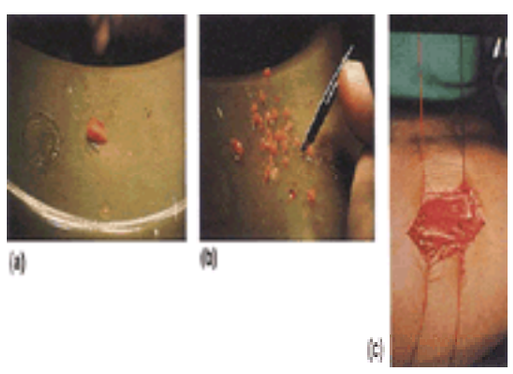


Fig. 18. Preparation of parathyroid tissue for autotransplantation to non-dominant forearm.

Surgery for parathyroid carcinoma

The surgeon should be alerted to the possibility of a parathyroid carcinoma if the plasma calcium is appreciably elevated above 3.5 mmol/l, especially if recorded in the presence of a palpable neck mass. It is at operation, however, that suspicion should be raised if the affected parathyroid is firmer than normal, grey in colour, and adherent to the local tissues. Lymph-node involvement is frequent. The surgical strategy is to perform an *en bloc* resection of the tumour and surrounding soft tissue, which usually implies an ipsilateral thyroid lobectomy and nodal dissection. Local recurrence may be eliminated by postoperative, localized external-beam radiotherapy.

Surgery for MEN syndromes

The MEN type I syndrome is characterized by asymmetrical hyperplasia of the parathyroids with a high percentage of supernumerary glands. Nearly all of these patients develop hyperparathyroidism eventually. Partial thyroidectomy and cervical thymectomy is therefore advocated to reduce the chance of recurrent hyperparathyroidism. The MEN type IIA syndrome, by comparison, only leads to hyperparathyroidism in approximately 30 per cent of sufferers and is due to adenomatous change in one or two glands. Removal of just one or two of these obviously enlarged glands is necessary to secure a normocalcaemic state.

Endoscopic parathyroidectomy and minimally invasive, radioguided parathyroidectomy

Recently, two new techniques of parathyroidectomy have been described. Endoscopic parathyroidectomy utilizes small ports in the neck (5 mm) and a 5- or 10-mm camera, to identify and remove parathyroid adenomas. The technique requires preoperative localization of the tumours by either MRI or sestamibi imaging and at present is in its infancy, with only a relatively small number of cases being reported. Whether this technique is a realistic and safe alternative to open parathyroidectomy remains to be proved in controlled clinical trials.

Minimally invasive, radioguided parathyroidectomy is a technique that has been pioneered in the University of South Florida. Patients with primary hyperparathyroidism undergo localization of their tumours with sestamibi imaging. When tumours are clearly localized, a small incision is made under local anaesthetic over the tumour site and a unilateral exploration is performed using a hand-held, a-particle-detecting probe. The probe enables the operator to focus directly in on the tumour location. The technique has grown in popularity recently in the United States as a primary procedure for primary hyperparathyroidism, which can thus be offered on an outpatient basis. It appears, however, to offer the greatest advantage in performing a focused re-operation in patients with recurrent or persistent parathyroid disease.

Postoperative results, complications, and management

The care of a patient after exploration for hyperparathyroidism is identical to that described for a routine thyroidectomy in Chapter 20.2.2. The same potential complications can arise but the risk of haemorrhage and damage to the recurrent laryngeal nerve is much less. Temporary hypocalcaemia is a frequent occurrence, especially in those patients with metabolic bone disease. As already described in the section on preoperative preparation, this can be treated expectantly with 1a-hydroxycholecalciferol, which modifies the fall in the plasma calcium.

However, calcium supplementation may be required in the early days after surgery, especially if a total parathyroidectomy has been performed for tertiary hyperparathyroidism. Permanent hypoparathyroidism is a rare problem and affects 1 per cent of patients. In primary hyperparathyroidism an experienced surgeon will at the first operation successfully correct the hypercalcaemia in 95 per cent of patients with no recurrence over a 2-year period. Operative mortality approaches zero (0.2 per cent in the author's series). Eight per cent of patients undergoing subtotal parathyroidectomy for secondary and tertiary hyperparathyroidism, i.e., multigland involvement, have persistent or recurrent hyperparathyroidism. This figure rises to 30 per cent in those affected with the MEN I syndrome, who also suffer a greater risk of permanent hypoparathyroidism (4 per cent). Total parathyroidectomy with immediate autotransplantation of fresh parathyroid tissue achieves a functioning graft in over 90 per cent of patients. When the initial graft proves inadequate to maintain the plasma calcium, cryopreserved tissue can be successfully grafted after 18 months storage, but 35 per cent of these grafts fail.

Strategy after failed cervical exploration for hyperparathyroidism

Frequently, patients in this category have had their primary surgery performed in a non-specialized unit. The validity of the diagnosis should be challenged, with review of the original biochemical data. Whether there are discrepancies or not, biochemical results will need to be rechecked. The possibility of familial hypocalciuric hypercalcaemia should be borne in mind, as this syndrome has been implicated in 9 per cent of patients with failed neck exploration. The disease likewise should be reviewed in case multigland disease has been overlooked. Assuming the diagnosis is confirmed, the next crucial information to be checked is in the operation notes. These should have recorded the precise location and appearance of all parathyroid tissue. When three normal glands have been found, the side of the missing fourth is established and the field of search narrowed. If four normal glands have been identified, the missing fifth is likely to be mediastinal. If there is incomplete documentation then a methodical re-exploration of the neck may be necessary unless localization studies have clearly shown the missing parathyroid(s) to be in the chest. The choice of localization study will depend on the local availability and expertise.

Further reading

Bruining HA *et al.* Results of operative treatment of 615 patients with primary hypoparathyroidism. *World Journal of Surgery* 1981; **5**: 85–9.

Brunt LM, Jones DB, Wu JS, Quasebarth MA, Meiniger T, Soper NJ. Experimental development of an endoscopic approach to neck exploration and parathyroidectomy. *Surgery* 1997; **122**(5): 893–901.

Clark OH *et al.* Localization studies in patients with persistent or recurrent hyperparathyroidism. *Surgery* 1985; **98**: 1083–94.

Coakley AJ, Kettle AG, Wells CP, O'Doherty MJ, Collins RE. $^{99}Tc^{m}$ sestamibi—a new agent for parathyroid imaging. *Nuclear Medicine Communications* 1989; **10**(11): 791–4.

Dudley NE. Methylene blue for the rapid identification of the parathyroids. *British Medical Journal* 1971; **3**: 680.

Dunlop DAB *et al.* Parathyroid venous sampling: anatomic considerations and results in 95 patients with primary hyperparathyroidism. *British Journal of Radiology* 1980; **53**: 183.

Edis AJ, Beahrs OH, van Heerden JK, Akwari OE. ‘Conservative' versus ‘liberal' approach to parathyroid neck exploration. *Surgery* 1977; **82**: 466–73.

Gilmour JR. The gross anatomy of the parathyroid glands. *Journal of Pathology and Bacteriology* 1938; **46**: 133–49.

Heath H III, Hodgson SF, Kennedy MA. Primary hyperparathyroidism. Incidence, morbidity and potential economic impact in a community. *New England Journal of Medicine* 1980; **302**: 189–93.

Higgins RM, Richardson AJ, Retcliffe PJ, Woods CG, Oliver DD, Morris PJ. Total parathyroidectomy alone or with autograft for hyperparathyroidism. *Quarterly Journal of Medicine* 1991; **79**: 323–32.

Ishibashi M, Nishida H, Hiromatsu Y, Kojima K, Tabachi E, Hayabuchi N. Comparison of technetium-99m-MIBI, technetium-99m-tetrofomin, ultrasound and MRI for localization of abnormal parathyroid glands. *Journal of Nuclear Medicine* 1998: **39**(2): 320–4.

Kaplan RA *et al.* Metabolic effects of parathyroidectomy in asymptomatic primary hyperparathyroidism. *Journal of Clinical Endocrinology and Metabolism* 1976; **42**: 415–26.

Lafferty FW, Hubay CA. Primary hyperparathyroidism: a review of the long term surgical and nonsurgical morbidities as a basis for a rational approach to treatment. *Archives of Internal Medicine* 1989; **149**: 789–96.

Law WM Jr, Heath H III. Familial benign hypercalcaemia (hypocalciuric hypercalcaemia): clinical and pathogenetic studies in 21 families. *Annals of Internal Medicine* 1985; **102**: 511–19.

Lilliemoe K, Dudley NE. Parathyroid carcinoma-pointers to successful management. *Annals of the Royal College of Surgeons of England* 1985; **67**: 222–4.

McGeown MG. Effect of parathyroidectomy on the incidence of renal calculi. *Lancet* 1961; **i**: 586–7.

Miccoli P *et al.* Minimally invasive, video-assisted parathyroid surgery for primary hyperparathyroidism. *Journal of Endocrinological Investigation* 1997; **20**(7): 429–30.

Naitoh T, Gagner M, Garcia-Ruiz A, Heniford BT. Endoscopic endocrine urgery in the neck. An initial report of endoscopic subtotal parathyroidectomy. *Surgery–Endoscopy* 1998; **12**(3): 202–5.

Norman J, Albrink MH. Minimally invasive videoscopic parathyroidectomy: a feasibility study in dogs and humans. *Journal of Laparoendoscopy and Advanced Surgical Technique* (A) 1997; **7**(5): 301–6.

Norman J, Chheda H. Minimally invasive parathyroidectomy facilitated by intraoperative nuclear mapping. *Surgery* 1997; **122**(6): 998–1003.

Proye C. Exploration parathyroidienne pour hyperparathyroidie. *Journale de Chirugie* (Paris) 1978; **115**: 101.

Rothmund M, Wagner PK. Total parathyroidectomy and autotransplantation of parathyroid tissue for renal hyperparathyroidism. A one to six year follow-up. *Annals of Surgery* 1983; **197**: 7–16.

Russel CF, Edis AJ. Surgical primary hyperparathyroidism. *British Journal of Surgery* 1982; **69**: 244.

Scholz DA, Purnell DC. Asymptomatic primary hyperparathyroidism: 10 year perspective study. *Mayo Clinic Proceedings* 1981; **56**: 473–8.

Shane E, Bilezikian JP. Parathyroid carcinoma: a review of 62 patients. *Endocrinology Reviews* 1982; **3**: 218–26.

Tibblin S, Boulesion AG, Ljungberg O. Unilateral parathyroidectomy due to a single adenoma. *Annals of Surgery* 1982; **195**: 245–52.

van Heerden JA *et al.* Primary hyperparathyroidism in patients with multiple endocrine neoplasia syndromes: surgical experience. *Archives of Surgery* 1983; **118**: 533–5.

Wang CA. Surgical management of primary hyperparathyroidism. *Current Problems in Surgery* 1985; **22**: 1–50.

Wells SA *et al.* Transplantation of the parathyroid glands in man. Clinical indications and results. *Surgery* 1975; **78**: 33–44.

Wilder WT, Frame B, Haubrich WS. Peptic ulcer in primary hyperparathyroidism. An analysis of 52 cases. *Annals of Internal Medicine* 1961; **55**: 885–93.

Yeung GH, Ng JW. The technique of endoscopic exploration for parathyroid adenoma of the neck. *Australia and New Zealand Journal of Surgery* 1998: **68**(2): 147–50.

Young AE *et al.* Location of parathyroid adenomas by thallium-201 and technetium 99m subtraction scanning. *British Medical Journal* 1983; **286**: 1384.

20.4.1 The adrenal gland

Justin A. Roake

Introduction
Adrenal anatomy and physiology
Disorders of the adrenal cortex
Adrenocortical hypofunction
Adrenocortical hyperfunction
Defective steroid biosynthesis
Adrenocortical tumours
'Incidentalomas' and non-functioning adrenal masses
Disorders of the adrenal medulla
Neuroblastoma
Phaeochromocytoma
Surgery of the adrenal gland
Further reading

Introduction

One consequence of increasing subspecialization is that few surgeons now require a detailed knowledge of the disorders of the adrenal gland. Adrenal pathology is not only relatively uncommon but also tends to present with functional abnormalities due to excess or insufficient hormone secretion for which, most often, a physician will be consulted initially. Surgical involvement is usually limited to the management of conditions such as adrenocortical neoplasia or phaeo-chromocytomas at a stage when much of the diagnostic work-up has been completed. With this background in mind, this section has been written to provide core knowledge for the surgeon in training and a useful reference for the surgeon who may occasionally be called upon to manage specific disorders of the adrenals.

Adrenal anatomy and physiology

The adrenal glands are paired structures located medial to the upper pole of each kidney (Fig. 1). Normally, each weighs between 4 and 5 g but after a prolonged illness they may increase substantially as a result of adrenocorticotrophic hormone (ACTH) stimulation. Rarely, adrenal tissue may develop at an ectopic site which bears an embryological relationship to the urogenital ridge and its derivatives. Ectopic tissue has most frequently been documented in the testis or spermatic cord but it may be located anywhere in the retroperitoneum from the diaphragm to the pelvis; in most cases it consists almost entirely of cortex but it can include medulla, especially when it is located in the region of the coeliac ganglion. The importance of ectopic adrenal tissue is that it may undergo hyperplasia in Cushing's disease or, rarely, it may undergo neoplastic change.

Fig. 1. Anatomic relations and blood supply of the adrenal glands.

The adrenal cortex and medulla are anatomically and functionally discrete units. The steroid-secreting cortex originates from the mesoderm of the urogenital ridge and gains a distinctive yellow colour from its content of steroid precursors, which include free and esterified cholesterol, triglycerides, and phospholipids. The catecholamine-secreting medulla, on the other hand, is derived from the neural crest and is closely related to the sympathetic nervous system and the extra-adrenal paraganglia. Primitive medullary cells develop along two lineages: chromaffin cells which become phaeochromocytes and neuroblasts which mature into ganglion cells.

At birth, the cortex consists predominantly of a wide fetal zone but after several months three clearly defined, cortical zones can be identified (Fig. 2). The subcapsular zona glomerulosa, which accounts for 10 to 15 per cent of the adult cortex, is the source of mineralocorticoids, the most important of which is aldosterone. The intermediate zona fasiculata, which accounts for about 80 per cent of the adult cortex, and the inner zona reticularis (5–10 per cent) produce glucocorticoids (predominantly from the zona fasiculata), the most active of which is cortisol, androgens (of which testosterone is the most significant), and small quantities of oestrogens. Cortisol and testosterone secretion is regulated by ACTH whereas aldosterone secretion by the zona glomerulosa is primarily regulated by the renin–angiotensin system and is largely independent of ACTH.

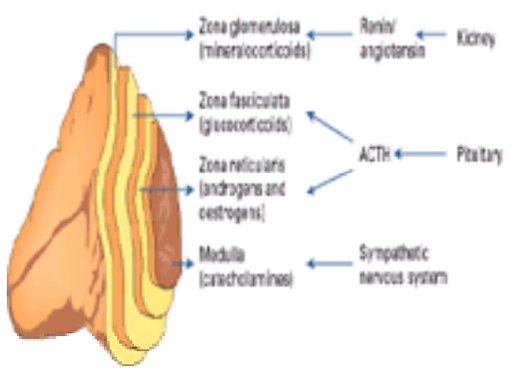


Fig. 2. Internal organization of the adrenal gland and functional relationships.

The adrenal medulla is extremely vascular and consists largely of a reticular network of catecholamine-secreting chromaffin cells with a closely related plexus of venous sinusoids intertwined amongst them. This facilitates release of catecholamines into the circulation. The gland has a rich nerve supply derived mainly from the coeliac and renal plexuses. The nerve endings terminate directly on the chromaffin cells which contain numerous electron-dense granules involved in the synthesis and storage of catecholamines. The predominant catecholamine produced and stored in the adrenal medulla is, as its name suggests, adrenaline but smaller quantities of noradrenaline are also produced.

Disorders of the adrenal cortex

Clinical disorders resulting from abnormal adrenocortical function can be considered to be of four types (Table 1). Hypofunction is most often a consequence of iatrogenic suppression or autoimmune destruction of the gland. Hyperfunction results from excessive production of adrenal hormones secondary to a functioning adrenal tumour or, more commonly, to overproduction of ACTH. Inborn errors of steroid synthesis may produce a mixed clinical picture of hyper- and hypofunction.

Finally, non-functioning adrenal masses may present as an incidental finding or with symptoms unrelated to adrenal function.

Adrenal cortex		
Hypofunction	Congenital hypoplasia Primary chronic adrenal insufficiency Primary acute adrenal insufficiency Secondary adrenal insufficiency	
Hyperfunction	Cushing's syndrome	Overproduction of ACTH Pituitary microadenoma Ectopic ACTH Adrenal tumour Hyperplasia Iatrogenic
	Conn's syndrome	Adrenal tumour Hyperplasia
	Virilization/feminization	Adrenal tumour
Defective steroidbiosynthesis	Congenital adrenal hyperplasia	
Non-functioning adrenal tumours	Cyst Myelolipoma Adenocortical adenoma Adenocortical carcinoma Metastasis Adrenal haemorrhage	
Adrenal medulla		
Medullary neoplasia	Neuroblastoma Ganglioneuroma Phaeochromocytoma	Benign Malignant

Table 1 Disorders of the adrenal cortex and adrenal medulta

In this section the clinical syndromes associated with these disorders will be considered followed by an account of the diagnosis and management of adrenocortical tumours and the management of adrenal masses discovered incidentally during investigation of an unrelated condition.

Adrenocortical hypofunction

Congenital adrenal hypoplasia is a rare, developmental anomaly which presents in the neonatal period. Two forms are recognized. The anencephalic type is associated with other serious congenital anomalies, usually including anencephaly, and these infants are usually stillborn or do not survive for more than a few days. The cytomegalic type may be genetically determined and exhibits a recessive pattern of inheritance. If recognized and treated appropriately with steroid replacement, long-term survival may result.

In later life, adrenocortical hypofunction can be split into three categories: primary chronic adrenal insufficiency, primary acute adrenal insufficiency, and secondary adrenal insufficiency. To the surgeon, secondary insufficiency is probably of greatest importance, and is certainly the most likely to be encountered. Most often it is a consequence of adrenal suppression resulting from chronic steroid therapy for a variety of conditions, including autoimmune diseases, or in the recipients of organ transplants. Alternatively, it may be encountered after treatment of Cushing's syndrome by adrenalectomy or following removal of a functional adrenal adenoma. Secondary adrenal insufficiency may present with an ‘adrenal crisis' especially during times of stress such as acute illness or following major surgery, unless adequate precautions are taken to prevent its occurrence (see below).

Primary chronic adrenal insufficiency (Addison's disease) is an uncommon condition which today is usually the result of autoimmune destruction of the adrenal cortex (previously tuberculosis was the most common cause). It is more common in women than men (M : F ratio approximately 3 : 1) and antiadrenocortical antibodies are demonstrable in the majority of cases, particularly in women. It may be associated with other autoimmune conditions and antibodies against the thyroid, parietal cells, or intrinsic factor are commonly found.

Primary acute hypoadrenalism may occur on a background of chronic adrenal insufficiency, for instance during ‘adrenal crises' of addisonian patients, or following too rapid withdrawal of steroids from patients with suppressed adrenals, or as a result of failure to increase replacement steroids in times of stress. More rarely it may result from massive haemorrhagic adrenal infarction. In neonates, in whom the adrenal is relatively large, this may follow birth trauma and/or hypoxia. In adults, when it is known as Waterhouse–Friederichsen syndrome, it is often related to acute bacteraemic infection, most commonly with the meningococcus, but it may also occur during pneumococcal, staphylococcal, or *Haemophilus influenzae* infections. It may also occur as a preterminal, agonal event.

The symptoms and signs of primary and secondary adrenal insufficiency are similar except that primary chronic insufficiency may be associated with abnormal pigmentation, especially in the skin creases, scars, and inside the lips and cheeks. This is a consequence of chronic stimulation of the pituitary and overproduction of ACTH and b–melanocyte stimulating hormone which stimulate increased pigmentation.

Chronic cortisol deficiency commonly results in weight loss which is often associated with episodes of colicky non-specific abdominal pain, vomiting and diarrhoea, general malaise and lack of energy, and there may be marked postural hypotension. The features of ‘adrenal crises' due to acute cortisol insufficiency include hypotension, shock, and hyponatraemia often associated with muscle cramps, myalgia, or unexplained fever.

The biochemical characteristics of hypoadrenalism are hyponatraemia, hyperkalaemia, and an elevated blood urea (8–15 mmol/l) all of which may occur in either primary adrenal insufficiency, in which both cortisol and aldosterone are deficient, or secondary insufficiency, where cortisol alone is deficient. These abnormalities tend to be most marked during acute adrenal crises. Proof of the diagnosis is not always easy since normal urinary cortisol excretion may be below the limits of detection and some urinary cortisol may be detected in hypoadrenalism. Likewise, plasma cortisol levels in adrenal insufficiency are often in the low–normal range. A high plasma cortisol, however, especially during acute illness, effectively eliminates the diagnosis. Definitive proof of hypoadrenalism usually rests on demonstration that the plasma ACTH is disproportionately high compared to plasma cortisol levels or that the adrenal fails to produce a normal secretory response following a challenge with exogenously-administered ACTH.

A hypotensive patient with acute adrenal insufficiency requires immediate volume replacement with normal saline solution and steroid replacement, for example hydrocortisone 100 mg intramuscularly or intravenously. This should be administered before confirmation of the presumed diagnosis and improvement should be expected within 4 to 6 h if the diagnosis is correct. The patient should then receive regular steroid parenterally, for example hydrocortisone 100 mg intramuscularly or intravenously every 6 h for 2 or 3 days, and may subsequently require oral maintenance steroid replacement.

In primary chronic adrenal insufficiency both cortisol and aldosterone require replacement. This requires the equivalent of about 30 mg of cortisol per day (two-thirds administered in the morning and one-third in the evening) and the mineralocorticoid (for example fludrocortisone) dosage adjusted according to blood pressure and plasma renin. In secondary adrenal insufficiency only cortisol needs to be replaced. Patients should be advised to double the daily glucocorticoid dose during intercurrent illness, for example fever above 38°C, systemic infection, or trauma, and if vomiting occurs for more than 24 h parenteral replacement should be given. Prior to surgery patients should receive replacement therapy as for acute adrenal insufficiency, commencing 1 h before anaesthesia and converting to oral replacement therapy once reliable oral intake has been re-established.

Adrenocortical hyperfunction

Cushing's syndrome

Cushing's syndrome encompasses a group of disorders presenting with the symptoms and signs of exposure to excess glucocorticoids. The onset of symptoms is usually insidious and the typical picture of the patient with all the features of the syndrome develops slowly. As a result, a relatively late clinical diagnosis is common.

Some features are more important than others diagnostically. These include skin atrophy producing skin which is palpably thin and the development of spontaneous purpura due to capillary fragility. Wasting affecting the proximal muscles, especially those of the pelvic girdle, and weakness which may be aggravated by hypokalaemia are also particularly characteristic. Osteoporosis, mainly affecting the axial skeleton and leading to spontaneous fractures of the ribs or vertebrae, and growth arrest in children are also important diagnostic clues. Other features of the syndrome which tend to be of less diagnostic value include central obesity, purple striae, poor wound healing and paper thin scars, facial hirsuitism, acne and plethora, oedema, moderate hypertension, glucose intolerance, amenorrhoea, and psychiatric manifestations such as depressive psychosis. Obesity tends to affect predominantly the face (‘facial mooning'), neck, and trunk and may contribute to the classical ‘buffalo hump' appearance (Fig. 3).



Fig. 3. The moon face of a patient with Cushing's syndrome.

Any of several underlying conditions of diverse aetiology may lead to Cushing's syndrome. Probably the most common cause today is iatrogenic Cushing's syndrome secondary to chronic administration of glucocorticoids to organ transplant recipients or patients with immune disorders. Cushing's syndrome resulting from primary adrenal disease or excessive exposure to ACTH is less common but collectively these conditions are of great importance because of the diagnostic challenge they pose and the potential for curative surgical intervention. Of these cases, between 65 and 75 per cent are attributable to an ACTH-secreting pituitary microadenoma which is often referred to as Cushing's disease as a tribute to Harvey Cushing who first described the condition in 1932. A further 10 to 15 per cent are due to ectopic sources of ACTH such as carcinoma of the lung (particularly small-cell carcinoma), malignant pancreatic tumours, benign or malignant thymomas or (less commonly) carcinoids, phaeochromocytomas, medullary carcinoma of the thyroid, and rarely other primary carcinomas. Primary adrenal disease such as adenoma, adenocarcinoma, or micronodular adrenal hyperplasia account for the remaining 10 to 20 per cent.

Certain clinical features may give some indication as to the aetiology of Cushing's syndrome. Ectopic ACTH production by a rapidly growing malignant primary tends to produce very high levels of cortisol secretion with rapid development of a Cushing's syndrome in which features such as hypokalaemia and oedema predominate in addition to the clinical features of the primary disease. Cushing's syndrome resulting from ACTH secretion by benign tumours tends to have a more indolent course with gradual development of the clinical manifestations. Malignant tumours of the adrenal cortex may synthesize cortisol inefficiently and this may lead to excessive production of androgen precursors, which in turn may produce hirsuitism and virilization, neither of which tends to be a feature of Cushing's syndrome induced by adrenal adenomas.

Investigation of patients suspected of having Cushing's syndrome is conducted in two phases: confirmatory screening or diagnostic tests followed by investigations to determine the underlying aetiology (Table 2). Probably the most useful screening test for hypercortisolism is the 24-h urinary free (unconjugated) cortisol excretion (the sensitivity and specificity of this investigation are approximately 95 per cent) but raised cortisol excretion may also be seen in obesity, during periods of stress, during pregnancy, or if oral contraceptives are being taken. The level of steroid over-production may also give some guidance to the aetiology. Levels of more than 5000 nmol/day are commonly found in Cushing's syndrome due to ectopic ACTH or adrenal carcinoma but only rarely in other cases. Urinary 17-hydroxysteroid and 17-ketosteroid excretion reflect basal cortisol production but are less discriminating than free cortisol excretion and there is greater overlap with normal subjects. Plasma cortisol levels are more difficult to interpret because under normal circumstances there is great variability. However, loss of the normal diurnal variation with failure of the plasma cortisol level to fall at night is a useful and constant feature of Cushing's syndrome. Elevation of the midnight plasma cortisol is a sensitive test but it may also occur during pregnancy, stress, or in subjects taking oral contraceptives. A somewhat more definitive test for Cushing's syndrome depends upon the relative resistance to suppression by exogenous steroids. Dexamethasone is a potent steroid which normally suppresses pituitary-derived ACTH and thus lowers plasma cortisol but in subjects with Cushing's syndrome suppression of cortisol secretion is resistant to low doses of dexamethasone and urinary excretion of 17-hydroxysteroids remains elevated. The single dose (overnight) dexamethasone suppression test, in which the morning plasma cortisol is measured after administration of 1 to 2 mg of dexamethasone at midnight, is associated with a low false-negative rate but a false-positive result may occur in the obese patient, in subjects taking oestrogens or phenytoin, or in the chronically ill. The so-called low-dose dexamethasone suppression test in which 24-h urinary 17-hydroxysteroids are measured during administration of 0.5 mg of dexamethasone 6-hourly may be more reliable.

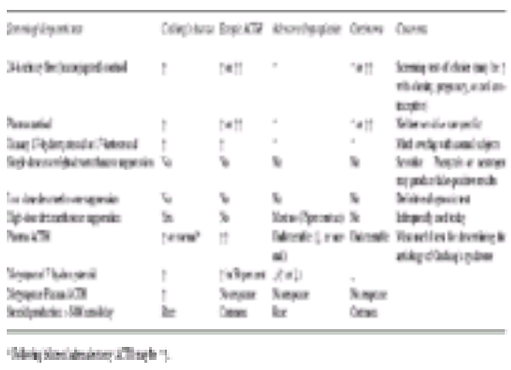


Table 2 Screening and diagnostic tests for Cushing's syndrome

The high-dose dexamethasone suppression test, in which 2 mg of dexamethasone is administered 6-hourly for 2 days and the 24-h urinary 17-hydroxysteroid excretion is measured during the second day, has been used to determine the underlying aetiology of Cushing's syndrome. In nearly all patients with Cushing's disease the 17-hydroxysteroid excretion is suppressed to less than 40 per cent of the baseline measurements but nearly all patients with adrenal tumours, most with adrenal hyperplasia, and all with an ectopic ACTH source fail to show suppression. However, because suppressibility with dexamethasone is a matter of degree and subject to errors induced by natural biological variability, other, more reliable, diagnostic tests have tended to supersede the dexamethasone suppression tests. Determination of the plasma ACTH by radioimmunoassay of blood samples taken in the morning, when ACTH is usually detectable in normal subjects, is the most useful investigation for determining the cause of Cushing's syndrome. In primary adrenal disease, the levels are undetectable (except occasionally in patients with micronodular adrenal hyperplasia in whom ACTH may be normal) whereas in Cushing's syndrome secondary to pituitary ACTH production the level is almost always elevated or normal. In cases of ectopic ACTH secretion plasma ACTH levels are always elevated and levels greater than 200 pg/l are virtually diagnostic. Note, however, that after adrenal ablation ACTH levels may be grossly elevated and this may be associated with Nelson's syndrome (see below).

Metyrapone is also useful in the determination of the aetiology of Cushing's syndrome. Metyrapone inhibits adrenal 11b-hydroxylation of 11-deoxycortisol to cortisol. This leads to reduced cortisol secretion, a rise in ACTH, and production and excretion of cortisol precursors. In cases secondary to primary adrenal adenoma, adrenal carcinoma, or ectopic ACTH production the pituitary is suppressed and no rise in ACTH or cortisol precursors occurs following administration of metyrapone but in 90 per cent of cases of Cushing's disease the pituitary does respond by increasing ACTH production.

Cushing's syndrome of non-adrenal origin

Diagnostic imaging and treatment of pituitary microadenomas is dealt with in detail elsewhere and will therefore be considered only briefly here. Untreated Cushing's disease commonly leads to death from the complications of excessive exposure to cortisol including hypertension, stroke, and ischaemic heart disease. Although spontaneous remission has been documented this is an exception rather than the rule and therefore surgical treatment is generally recommended.

In the past, permanent relief of hypercortisolism was achieved by total adrenalectomy but this was associated with a high operative morbidity and mortality, mainly as a consequence of chronic excessive exposure to steroids, and left the individual dependent upon life-long replacement therapy. Furthermore, many patients developed very high ACTH levels, enlarged pituitary glands, and hyperpigmentation (Nelson's syndrome) often necessitating pituitary ablation. Today total adrenalectomy is reserved for the occasional patient in whom the pituitary lesion cannot be treated directly by surgical removal or irradiation.

In Cushing's syndrome secondary to ectopic ACTH secretion the ACTH secreting tumour should be sought and if possible removed. In the case of benign tumours, for example thymomas and some carcinoids, this may effect a surgical cure. When it is not possible to remove the source of the ACTH, excessive cortisol secretion can usually be controlled with adrenolytic drugs such as metyrapone or aminoglutethimide which inhibits the conversion of cholesterol to D5-pregnenolone, thereby inhibiting production of cortisol, mineralocorticoids, and androgens. Occasionally, surgical adrenalectomy may be considered if the ACTH source cannot be removed

and the patient would otherwise have a good prognosis following relief of the hypercortisolism.

Primary aldosteronism (Conn's syndrome)

Primary aldosteronism most commonly results from excessive and autonomous secretion of aldosterone by an adenoma of the adrenal cortex (Fig. 4) but less commonly may be due to bilateral micronodular hyperplasia involving the zona glomerulosa. Although the aetiology of the latter is unknown it has been suggested that it may represent tertiary aldosteronism, in which excessive renin secretion (subsequently suppressed) was the original stimulus. This remains entirely hypothetical.



Fig. 4. A primary aldosterone-secreting tumour (Conn's tumour) of the adrenal cortex.

In 1955, Conn was the first to describe a patient with hypertension, neuromuscular symptoms, and renal potassium wastage which was associated with elevated aldosterone secretion and an adrenocortical adenoma. Excessive secretion of aldosterone produces hypertension, predominantly as a result of intravascular volume expansion secondary to aldosterone-induced salt and water retention, but overall, as a cause of hypertension, it is rare, and accounts for less than 0.5 per cent of cases. Primary aldosteronism is suspected when hypokalaemia is associated with alkalosis in the presence of hypertension but it must be distinguished from other (more common) causes of hypokalaemia associated with hypertension, particularly that induced by loop or thiazide diuretics, and by renin-induced secondary hyperaldosteronism. The latter is seen in hypertension due to renovascular disease and in severe essential hypertension. Primary aldosteronism without hypokalaemia is uncommon. Patients presenting with hypertension in whom studies to exclude Conn's syndrome should be undertaken include those with spontaneous hypokalaemia (<3.5mmol/l), moderately severe hypokalaemia (<3.0 mmol/l) while taking diuretics, or difficulty maintaining a normal potassium despite oral supplements or potassium sparing diuretics, and patients with hypertension refractory to treatment with no specific evidence of a secondary cause.

Confirmation of suspected primary aldosteronism can sometimes be difficult but most often the diagnosis is easily established by demonstrating elevated plasma aldosterone, or increased urinary excretion of aldosterone, despite adequate salt loading (confirmed by measurement of urinary sodium excretion) over several days. Hypokalaemia, inappropriate kaliuresis, and depressed plasma renin activity that fails to rise in response to salt and water depletion or the upright posture are supportive features but their absence does not exclude the diagnosis. In primary aldosteronism the plasma sodium is usually normal or elevated whereas, in contrast, hyponatraemia is often a feature of secondary aldosteronism. If the diagnostic tests are equivocal, multiple measurements of aldosterone during salt loading may be needed.

It is important to distinguish between adrenal hyperplasia and an adenoma because the treatment of each condition differs. An adenoma is said to produce more pronounced biochemical abnormalities (higher aldosterone, and lower renin and potassium) than does hyperplasia but this is of little value in distinguishing the two. Diagnostic imaging is required. Large adenomas (greater than 1 cm in diameter) may be detected reliably by computed tomography (CT) (Fig. 5) or magnetic resonance imaging (MRI). These investigations provide little or no information on the function of any abnormal tissue detected but this can be provided by radionuclide imaging or selective venous sampling which may be required to distinguish an adenoma from hyperplasia. Selective adrenal sampling for aldosterone and cortisol levels is an extremely valuable test in confirming the diagnosis and localizing a tumour. The cortisol levels allow the validity of the adrenal vein sampling to be established in comparison with peripheral samples (see adrenocortical tumours below).

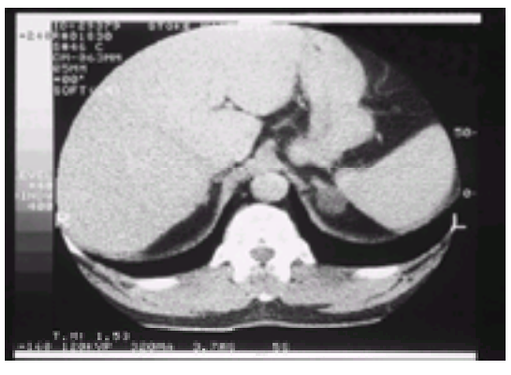


Fig. 5. A CT scan showing a Conn's tumour of the left adrenal gland.

Primary aldosteronism from an adenoma is treated by surgical excision (adrenalectomy) once the hypertension and hypokalaemia have been corrected with spironolactone or amiloride. Surgical treatment of micronodular hyperplasia is much less satisfactory and the treatment of choice is long-term control with spironolactone or amiloride.

Virilizing or feminizing tumours

In children, virilizing tumours may present with precocious puberty, phallic enlargement, and development of secondary hair growth on the face, axillae, and pubes. The differential diagnosis includes congenital adrenal hyperplasia (see below), idiopathic hirsuitism, and other causes of precocious puberty. In women, virilizing tumours present with deepening of the voice, hirsuitism, amenorrhoea, enlargement of the clitoris, increase in muscle bulk, and increased libido, whereas in men they may be relatively silent and may only be recognized as an incidental finding or, if malignant, by the development of metastases.

Virilizing adrenal tumours are most commonly diagnosed by finding an enlarged adrenal or adrenal mass on CT scan in association with appropriate biochemical abnormalities. Both adrenal and gonadal causes of virilization produce increased plasma androgens but elevated dehydroepiandrosterone (DHEA) which is secreted almost exclusively by the adrenal gland, suggests an adrenal cause for virilization. DHEA may also be elevated in some cases of idiopathic hirsuitism but this may be distinguished from an adrenal tumour biochemically by the dexamethasone suppression test: 17-ketosteroid secretion by adrenal tumours is not normally suppressed. In patients with adrenocortical carcinoma producing virilization, plasma DHEA and 17-ketosteroid levels tend to be very high.

Feminizing tumours are very rare and are almost always malignant. In men they may present with gynaecomastia and impotence before they have metastasized and curative resection may be possible but in women they tend to be advanced when first detected.

Defective steroid biosynthesis

Congenital adrenal hyperplasia (androgenital syndrome)

This is a group of genetically controlled congenital disorders which have in common a complete or partial defect in the control of steroid biosynthesis producing elevated ACTH, excessive production of androgens and/or mineralocorticoids, and striking enlargement of the adrenals which reach 10 to 20 times their normal weight. Inheritance is by recessive genes with variable penetrance and the incidence is between 1 : 10 000 and 1 : 50 000 live births. Any one of several enzyme defects may

be responsible but the most common is 21-hydroxylase deficiency in which cortisol and aldosterone production is defective. Excessive secretion of pituitary ACTH then results in high levels of adrenal androgen production which may produce virilization at birth or sexual precocity in boys. Less commonly an 11b-hydroxylase deficit produces excess 11-deoxycortisol, a cortisol precursor with mineralocorticoid activity. The net result may be hypertension in addition to increased androgen secretion. Other, less common, defects include deficiencies of 17-hydroxylase, 18-hydroxylase, and 3b-hydroxysteroid dehydrogenase.

There is a wide spectrum of clinical presentations but the diagnosis should be suspected in babies or children with abnormal genitalia (for example cryptorchidism, hypospadias, or ambiguous genitalia), clitoromegaly with or without pubic hair development in girls, or precocious puberty in boys, or in children with hypertension. A family history of neonatal deaths, or siblings with hirsuitism, or congenital adrenal hyperplasia helps to identify individuals at increased risk.

The diagnosis is confirmed by laboratory investigations including 24-h urinary oxogenic- and oxosteroids and plasma cortisol (which maybe normal), ACTH, testosterone, and 17a-hydroxyprogesterone. Treatment is directed at correcting metabolic abnormalities (for example salt deficiency) followed by replacement of mineralocorticoid and cortisol, to suppress ACTH and reduce androgen overproduction. This allows emergence of normal gonadotrophic function at puberty and permits normal growth. Surgery for cryptorchidism or genital abnormalities may be required, often as staged procedures, and the perioperative period should be covered with increased steroid replacement as for adrenal insufficiency.

Adrenocortical tumours

Clinically-detected adrenocortical neoplasms are roughly equally divided between benign and malignant. Most benign tumours secrete either glucocorticoids, mineralocorticoids, androgens, or oestrogens. In contrast, carcinomas of the adrenal often secrete multiple, biologically active and inactive steroids in large amounts and about half of them produce clinical features attributable to steroid secretion. Cushing's syndrome is the most common clinical presentation of adrenocortical tumours but they may also present with aldosteronism, virilization, or, very rarely, feminization. A mixed picture is characteristic of adrenocortical carcinoma and reflects the spectrum of steroids and biologically active steroid precursors produced by the tumour.

Diagnostic imaging and treatment of adrenocortical tumours

After biochemical confirmation of a clinically-suspected functional adrenal tumour, localization and staging is required before planning surgical intervention. CT, the most commonly used imaging modality in the assessment of adrenal pathology, can identify almost all lesions greater than 1 cm in diameter and it has largely replaced investigations such as intravenous pyelography, retroperitoneal pneumography, nephrotomography, selective arteriography, or selective venography. Some adrenal masses have characteristic CT scan features that permit a precise diagnosis. Myelolipoma can be distinguished confidently because of the high fat content. Acute haemorrhage often shows characteristic, high CT attenuation especially on unenhanced scans. Cysts, however, can be difficult to distinguish from adenomas and CT cannot reliably distinguish between adenoma, carcinoma, or lymphoma. A CT scan may detect evidence of local invasion of the kidney or other adjacent structures and it may also demonstrate lymph node or liver metastases. Tumours more than 6 cm in diameter, as assessed by CT, should be suspected of malignancy and in these cases selective arteriography or venography may be useful to determine the degree to which local structures have been invaded, information which may be crucial in determining whether or not the tumour is operable. MRI is as sensitive as CT for detecting adrenal lesions but it may also be used to assist in characterisation of masses. MRI features of benign and malignant adrenal lesions have been extensively investigated. Adrenal adenomas have a significant lipid content which is typically absent from adrenal metastases and other non-adenomatous adrenal masses and can be detected by chemical shift MRI. Further studies, especially of adrenocortical carcinomas, are required before the reliability of this mode of assessment can be determined. MRI does, however, demonstrate the relationship of adrenal pathology to adjacent vascular structures more clearly than CT and has therefore proved to be useful in the assessment of the local extent of suspected malignant tumours. CT and MRI should be treated as complementary investigations.

Isotope scans provide both anatomical and functional information and may be a useful adjunct to CT or MRI. In Cushing's syndrome the pattern of uptake of ^{131}I–6b-iodomethyl–19-norcholesterol (NP59) is dependent upon the underlying aetiology. Four distinct patterns are recognized. Lateralized increased uptake with non-visualization of the other side is virtually pathognomonic of an adrenocortical adenoma. However, there have been several reports of well differentiated cortisol secreting adrenocortical carcinomas which demonstrated intense uptake of NP59 and were initially thought to be benign on the basis of the NP59 scan. Adrenocortical carcinomas are usually unable to incorporate enough NP59 to be visualized by NP59 scintigraphy because of inefficient production of cortisol, and since pituitary ACTH is suppressed by the excessive cortisol production the contralateral gland is also suppressed, resulting in bilateral non-visualization. Thus, bilateral non-visualization in the presence of Cushing's syndrome is considered to be diagnostic of carcinoma. Androgen or oestrogen-secreting tumours do not usually suppress ACTH and in these cases the pattern is one of bilateral uptake with marked asymmetry since the tumour itself is not visualized but the remainder of the ipsilateral gland and the entire contralateral gland are. In contrast, in Cushing's syndrome due to pituitary or ectopic ACTH the uptake is increased bilaterally and symmetrically. Micronodular hyperplasia of the adrenal may produce asymmetric bilateral increased uptake.

In primary aldosteronism radionuclide imaging, using radiolabelled cholesterol during dexamethasone suppression of cortisol secretion, can be used to detect functional tissue and may detect tumours as small as 5 mm in diameter. Selective adrenal vein catheterization and measurement of aldosterone on each side should reveal unilateral increased secretion with suppression of the other side if a functional adenoma is present but bilaterally increased secretion if aldosteronism is due to hyperplasia or is secondary to increased renin secretion. When radionuclide imaging is combined with selective adrenal vein catheterization a clear distinction between adenoma and hyperplasia should be achieved in most cases. Occasionally, however, a case for diagnostic surgical exploration can be made.

The preferred treatment for a functioning adrenocortical adenoma is surgical excision because the results of surgery are generally good and often produce a complete cure. Most adenomas coming to surgery are less than 4 cm in diameter and in selected cases it may be possible to preserve uninvolved adrenal tissue. Operations to resect adrenal adenomas causing Cushing's syndrome usually have excellent results but in the long term many patients remain obese and hypertensive. Surgical intervention in this group is associated with a relatively high morbidity because of chronic exposure to excess glucocorticoids specifically resulting in poor wound healing, increased risk of infection, and a relatively high risk of pulmonary thromboembolism. However, the development of endoscopic adrenalectomy has led to a major reduction in morbidity and is now the preferred approach (see 20.4.2). Perioperative glucocorticoid supplementation is required because of chronic suppression of the contralateral gland and this must be continued, with graduated dosage reduction, for several months.

After surgical removal of an aldosterone-producing adenoma, hypertension is cured in 70 to 80 per cent of cases and in the remainder surgery usually renders the hypertension more responsive to medical therapy. Several weeks before operation, hypertension and metabolic abnormalities should be corrected medically. The hypertension is salt and water dependent and is best treated by sodium deprivation and potassium sparing diuretics. Postoperatively there may be a sodium diuresis and retention of potassium and close monitoring of plasma electrolytes is required in the first week. In some cases suppression of the contralateral gland results in postoperative aldosterone deficiency which may produce hyponatraemia with hyperkalaemia necessitating mineralocorticoid replacement until the remaining gland recovers.

Adrenocortical carcinoma

Carcinoma of the adrenal cortex is usually highly aggressive and has a poor prognosis. Fortunately it is a rare tumour with an incidence of about 1 per 1.5×10^6 population per year. It can occur in all age groups but most commonly presents in the fifth to seventh decades and it occurs roughly equally in both sexes.

The tumour may attain a very large size (>20 cm) but overall the median size at presentation is about 10 to 12 cm. Relatively few (<10 per cent) are less than 6 cm at presentation. Traditionally, adrenocortical carcinomas are classified as functioning or non-functioning. Non-functioning tumours fail to produce clinical evidence of steroid synthesis but in most cases detailed studies indicate that these tumours do produce precursor steroids with little or no biological activity. The non-functioning tumours, which account for about 50 per cent of all adrenal carcinomas, often attain a very large size before detection and frequently present as an abdominal mass or with abdominal pain, weight loss, or fatigue. In some cases haemorrhagic necrosis of a large tumour may result in severe pain, fever, and even hypovolaemic shock.

Functioning tumours may also be large at presentation because of inefficient steroid production by the malignant tissue and therefore relatively late development of endocrine manifestations. Commonly they present with symptoms and signs of cortisol, androgen, or oestrogen excess, often with a mixed picture, but clinical evidence of aldosterone excess is rarely a presenting feature and exclusive production of aldosterone by the tumour is very rare. Metastases to the local lymph nodes, liver, lungs, or, less commonly bone, are found in 70 to 75 per cent of patients at presentation and these may produce the initial manifestations of the disease.

Features typical of adrenocortical carcinoma on CT scan or MRI include:

1. size more than 6 cm;
2. evidence of local invasion (particularly of the IVC) or distant metastases;
3. central areas of low attenuation from tumour necrosis; and
4. calcification which is seen in about one-third of cases.

Adrenocortical carcinomas usually appear lobulated and encapsulated and they frequently erode through their capsule and invade local structures such as the

pancreas, kidney, and bowel. On cutting, areas of haemorrhagic necrosis are commonly found. Histologically it can be difficult to distinguish between a benign adenoma and a well differentiated carcinoma on the basis of cellular characteristics alone and it is often necessary to fall back upon the presence of local invasion and metastases as hallmarks of malignancy to make the diagnosis. The cellular DNA content may be useful diagnostically: it has been suggested that aneuploid tumours metastasize whereas euploid tumours do not. As CT-guided percutaneous fine-needle aspiration cytology is used more frequently in the assessment of adrenal masses, incorporating measurement of the DNA content may prove to be a useful diagnostic approach.

Surgical eradication of the tumour offers the only chance of cure but even if complete resection is not possible because of invasion of vital structures, debulking of the tumour is important for palliation especially if the tumour is functional. In patients without demonstrable remote metastases, a radical approach is generally advocated. Adrenocortical carcinomas are approached by a transabdominal or thoracoabdominal route and the tumour and adjacent involved organs are excised *en bloc*. In some cases even involvement of the inferior vena cava may not prevent complete resection provided the facilities for cardiopulmonary bypass are available. Lymph node metastases are excised with the tumour and liver metastases accessible for wedge resection should be removed. The operation must be covered with perioperative steroid replacement but in the case of non-functioning tumours this can be tapered relatively quickly over the first 1 to 2 weeks after the operation.

Chemotherapy may be used for inoperable or recurrent carcinomas or as an adjuvant to maintain remission following apparently complete excision. The agent with which there is the greatest accumulated experience in the treatment of these tumours is mitotane (*o,p*'-DDD), a derivative of DDT which produces selective necrosis of the zona fasiculata and zona reticularis. The effectiveness of mitotane is dependent upon the bulk of the tumour present and it is therefore important that surgical debulking is followed by early postoperative chemotherapy. Cortisol replacement therapy is invariably required and in some patients mineralocorticoid deficiency may also require treatment.

The clinical response to mitotane is rather variable and the side-effects (nausea, anorexia, neurotoxicity, ataxia, papilloedema, skin rashes, and cystitis) may lead to poor patient acceptance. However, the alternatives are limited as there is relatively little experience with other agents even though short-term responses to alkylating agents and doxorubicin have been reported. Treatments to reduce steroid production with aminoglutethimide or metyrapone may afford significant palliation but they do not produce tumour regression. Radiotherapy has not been systematically examined in the treatment of adrenocortical carcinoma but it has never been shown to be effective as a primary treatment or as an adjuvant for residual disease. However, it does appear to be effective for the relief of pain from bony metastases.

The prognosis correlates with the size of the tumour and the functional status, but overall the outlook is poor. Patients presenting with anaplastic tumours without laboratory evidence of steroid production have a median survival of about 6 months and the response to therapy is very poor. Differentiated carcinomas with hormone production are associated with a greater median survival (about 2 years) and in about 50 per cent there is objective evidence of a response to chemotherapy. Even in cases where surgical excision has been apparently curative, long-term follow-up is required because late recurrence may occur.

'Incidentalomas' and non-functioning adrenal masses

Discovery of an incidental adrenal mass (an ‘incidentaloma') during investigation for an unrelated condition has become an increasingly common clinical problem. It has been estimated that approximately 0.35 to 1.3 per cent of upper abdominal CT scans performed for a variety of reasons will detect a clinically silent adrenal lesion. Although most solid lesions will be found to be non-functioning cortical adenomas, they all require careful evaluation directed towards ruling out malignancy and detecting tumours with hormonal activity which require excision. Other conditions such as myelolipoma usually remain clinically silent, producing neither symptoms nor signs, but occasionally complications, such as bleeding into a cyst, may bring them to clinical attention.

The CT appearance may be diagnostic. Simple cysts, myelolipomas, and adrenal haematomas can usually be identified and managed appropriately on the basis of the CT image alone. Most cystic masses are benign endothelial-lined (lymphangiomatous or angio-matous) cysts or pseudocysts secondary to haemorrhage into a normal adrenal. Occasionally, however, cortical adenomas, adenocarcinomas, or phaeochromocytomas may undergo cystic degeneration. Following biochemical evaluation cystic masses should be aspirated, under ultrasonic or CT guidance, and the aspirate sent for cytological evaluation to confirm their benign nature. Suspicion of malignancy is an indication for operation. Adrenal haemorrhage may result from bleeding into an adrenal tumour and it should therefore be followed by a repeat CT scan.

Rarely, the CT appearances will be diagnostic of primary adrenal carcinoma on the basis of size, lack of homogeneity, calcification, irregular borders, local infiltration, and evidence of metastases. Primary adrenal carcinomas are rarely found to be less than 6 cm in diameter and it has been estimated that in the absence of CT characteristics suggestive of malignancy fewer than 1 in 10 000 incidentally discovered lesions below 6 cm in diameter will be adrenocortical carcinomas. The likelihood that solid lesions 6 cm or more in diameter will be malignant is more difficult to estimate but it is probably between 10 to 30 per cent. Metastatic carcinoma of the adrenal is also very uncommon but in the presence of a known primary malignancy elsewhere, particularly of the lung, kidney, or colon, this becomes the most likely explanation for an incidentally-detected solid adrenal lesion.

Careful clinical assessment is required to detect evidence of clinical or subclinical function but the extent of testing is the subject of considerable debate. According to Ross and Aron, in the absence of certain clinical features of a functional tumour, biochemical exclusion of the diagnosis is unnecessary. For instance, virtually all patients with primary hyperaldosteronism have hypertension and in its absence biochemical exclusion may not be required. Furthermore, in hypertensive patients with an incidentally discovered adrenal mass the absence of spontaneous hypokalaemia has a 95 per cent negative predictive value for an aldosterone-producing adenoma and further biochemical exclusion is therefore probably not required. It is reasonable, then, to limit the initial studies directed towards the exclusion of function to plasma potassium, and 24-h urinary cortisol, catecholamine, and vanillylmandelic acid (VMA) excretion unless specific clinical evidence of function is present but in countries where specific tests of adrenal function are readily available a more complete biochemical workup is often advocated. In doubtful cases, isotope scans such as NP59 and ^{131}I-MIBG * scans may also help to distinguish truly non-functional tumours from those with subclinical function. Evidence of hormone hypersecretion is an indication for operation.

The management of solid non-functional adrenal masses remains controversial. It seems generally agreed that tumours 6 cm or more in diameter should be resected because they carry a high risk of malignancy and most surgeons would also agree that tumours less then 3 cm in diameter without biochemical evidence of function or any identifiable features of malignancy may be observed by CT. Evidence of growth should be an indication for operation but if they remain unchanged after repeated studies they are most likely to be benign, since carcinomas proliferate rapidly, and further studies are unnecessary. For tumours between 3 and 6 cm some surgeons advocate routine resection or selective resection based upon patient age (<60 years for example) and operative risk but this policy involves morbidity and expense with few patients benefiting from an earlier diagnosis of malignancy and it has not found general acceptance. Others routinely observe by repeated CT scans because the probability of malignancy is low and adenomas are not known to become malignant. MRI may help to distinguish adrenal metastases, phaeochromocytomas, and carcinomas from adrenal adenomas, lipomas, and cysts depending upon differences in signal intensities and may thus be helpful in determining whether operation should be advised. There is little evidence that MRI substantially improves diagnostic accuracy but it does increase the costs of investigation. The Swedish MFR group has suggested a protocol for management of incidentalomas in an attempt to achieve a degree of consensus and this is presented in modified format in Fig. 6. They have selected a size threshold of 4 cm.

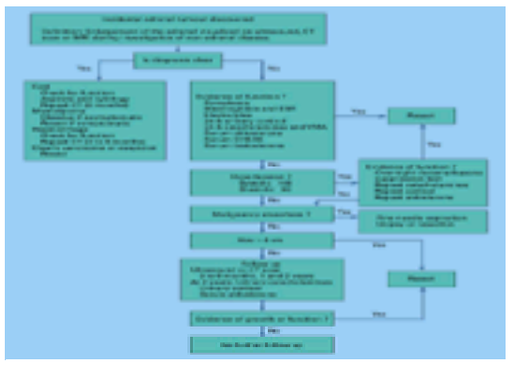


Fig. 6. A proposed management of adrenal ‘incidentalomas'.

The role of surgery in the management of adrenal metastases is controversial but non-surgical management is known to be associated with poor survival. Adrenalectomy, where there is a clinically solitary and completely resectable lesion, can result in prolonged survival especially when there has been a disease-free interval of more than 6 months.

Disorders of the adrenal medulla

The clinically important disorders of the adrenal medulla are tumours; the most important of these are neuroblastomas and phaeochromocytomas.

Neuroblastoma (see also Chapter 42.10)

Neuroblastomas occur almost exclusively in children and are the most common malignant tumours in neonates and infants (less than 1 year old). The overall incidence is approximately 8 per million population per year with 50 per cent presenting within the first 2 years of life and about 75 per cent in the first 5 years. They arise from the neuroblasts of the adrenal medulla in roughly 50 per cent of cases or from the prevertebral or paravertebral ganglia from the neck to the pelvis.

The tumours have a variable appearance and may be irregular or nodular and range in size from 3 to 4 cm to massive retroperitoneal masses. They are very vascular, friable tumours and areas of haemorrhage, necrosis, and calcification are commonly found in larger examples. They tend to invade their capsule early and then infiltrate along tissue planes and perineural pathways. Frequently, however, the plane between an adrenal neuroblastoma and the kidney parenchyma remains intact and although structures such as the ureters, major vessels, and gut are often surrounded by tumour the lumen is rarely invaded. Neuroblastomas metastasize early to the prevertebral lymph nodes and by haematogenous spread to bone, bone marrow, and liver but pulmonary metastases are relatively rare. A peculiarity is that the bony orbit is frequently involved and this may produce proptosis and ‘bruising' around the orbit.

A neuroblastoma may present with a wide variety of symptoms and signs related to the varied sites of the primary and metastases. Abdominal pain may be a presenting symptom but it is often vague and difficult to assess in children. Sudden recognition of an abdominal mass, which is usually firm to hard, knobbly and irregular, and not particularly tender, is probably the most common mode of presentation. Some tumours produce biologically-active peptides, for example vasoactive intestinal peptide (VIP), which produce specific symptom clusters. The so called ‘dumbbell' tumours may produce paralysis or other neurological signs in the lower limbs as a result of spinal cord compression or invasion. Metastases are present in over 50 per cent of cases at presentation and may produce symptoms including bone pain, or respiratory distress secondary to massive hepatomegaly.

In investigation and staging of the tumour, the minimum set of investigations should include liver and renal function tests, a chest radiograph and skeletal survey with orbital views, a bone scan, CT of the abdomen and pelvis, and a bone marrow aspirate. CT is highly sensitive for abdominal masses and demonstration of lymph node involvement. MRI defines vessel encasement better than CT and provides superior imaging of the vertebral canal in the assessment of ‘dumbbell' tumours. ^{131}I-MIBG scanning is also highly sensitive and is especially useful in the demonstration of metastases or recurrent tumour. Most tumours produce catecholamines and in 85 to 90 per cent of cases abnormal levels of urinary catecholamines or their metabolites may be detected. It should be noted, however, that benign ganglioneuromas can also raise the urinary catecholamine excretion.

Neuroblastomas exhibit a wide range of clinical virulence which reflects great biological heterogeneity. Small tumours presenting in infancy have a fairly high incidence of spontaneous regression or maturation into a benign ganglioneuroma but equally the tumours may be particularly aggressive especially when presenting after the first year of life. A number of genetic abnormalities that play a role in determining tumour clinical behaviour have been identified but the molecular events underlying the variability in tumour growth and outcome are largely unknown.

Clinical and biological prognostic factors can be identified and are summarized in Table 3. The most important clinical factor is stage at diagnosis. Age is the only other independent clinical prognostic factor identified so far. For all stages, age less than 1 year at diagnosis carries a significantly better prognosis than tumours of similar stage in older children. In fact, most infants with disseminated disease can be cured whereas advanced disease in children is usually fatal despite aggressive multimodality therapy. Several staging systems have been used and that proposed by Evans in 1971 has probably been the most widely used. However, this did not take into account the feasibility of surgical excision and in 1986 an international consensus group met to establish a neuroblastoma staging system based on clinical, surgical, and pathological findings. This was revised in 1993 and has now been implemented world-wide as the INSS (International Neuroblastoma Staging System) (Table 4).

Factor	Good prognosis	Poor prognosis
Age	< 1 year	> 1 year
Stage	I and II all ages	III if > 1 year
	III and IVs if < 1 year	IV all ages
Histology (see text)	Favourable	Unfavourable
MYCN oncogene	Single copy	Amplified
Chromosome 1p	Normal	Deletion
DNA content	Hyperploidy	Diploidy
Urinary catecholamine metabolites	Low HVA/VMA	High HVA/VMA
Tumour markers	Normal	Elevated
Neurotrophin receptors	Normal	High
Multidrug resistance protein	Normal	High
CD44 expression	Normal	Low

Table 3 Prognostic factors in neuroblastoma

Stage I	Localized tumour confined to area of origin Complete gross excision with or without microscopic residual disease Identifiable ipsilateral and contralateral lymph nodes negative microscopically
Stage IIa	Unilateral tumour with incomplete gross excision Identifiable ipsilateral and contralateral lymph nodes negative microscopically
Stage IIb	Unilateral tumour with complete or incomplete gross excision with positive ipsilateral regional nodes Identifiable contralateral lymph nodes negative microscopically
Stage III	Tumour infiltrating across the midline with or without regional lymph node involvement or Unilateral tumour with contralateral lymph node involvement or Midline tumour with bilateral lymph node involvement
Stage IV	Dissemination tumour to distant lymph nodes, bone, bone marrow, liver, and/or other organs
Stage IVs	Localized primary tumour as defined for stage I or IIa with dissemination limited to liver, skin, or bone marrow

Table 4 International neuroblastoma staging system (INSS)

Important biological prognostic factors are summarised in Table 3. Tumours are classified histologically as favourable or unfavourable depending upon:

1. the degree of neuroblast differentiation;
2. stroma content;
3. mitosis–karyorhexis index;
4. mitotic ratio (number of mitoses per 10 high power fields); and
5. presence or absence of tumour calcification.

Unfavourable histology is associated with amplification of the *MYCN* oncogene and clinically aggressive disease. *MYCN*-amplifica-tion is itself associated with advanced disease at diagnosis and rapid tumour progression, and is a strong predictor of an adverse outcome. Deletion of the short arm of chromosome 1 (possibly the site of a neuroblastoma suppressor gene) is the most common cytogenetic abnormality seen in neuroblastoma but it is associated with *MYCN*-amplification and may not be an independent predictor of outcome. In contrast to other tumours, in which an abnormal cellular DNA content is often associated with a poor outcome, in neuroblastomas hyperploidy is a favourable marker. Poorly differentiated tumours, especially those said to be stroma-poor and those with abnormal nuclear morphology, have a poor prognosis as do those with evidence of immature catecholamine synthesis as indicated by a high urinary homovanillic acid (HVA) to VMA ratio. Other adverse biological factors include elaboration of circulating tumour markers (elevated serum ferritin, lactate dehydrogenase, or neurone-specific enolase (NSE)), high expression of neurotrophin receptors trkA and trkC, high level expression of the multidrug resistance-associated protein (MRP), expression of

angiogenesis factors (evidenced by high tumour vascularity), and low expression of CD44 (a cell adhesion molecule which has a role in tumour metastasis).

Treatment of a neuroblastoma usually requires a combined approach by surgeons and oncologists which is well planned and co-ordinated according to extent of disease, age of the patient, and tumour biology. Total surgical removal of localized disease is the principal aim but when incomplete removal only is possible, the tumour must be controlled with chemotherapy or radiotherapy. In disseminated disease, surgery may initially be limited to establishing the diagnosis through biopsy of the primary and or metastases, although this may also be achieved by needle core biopsy. Delayed surgery may then be undertaken to assess the effectiveness of chemotherapy or radiotherapy and to excise residual tumour.

Patients are classified as low, intermediate, or high risk (Table 5) and this determines the therapy offered. Most low-risk patients can be cured by surgery alone and 5-year survival rates in excess of 95 per cent are reported. Intermediate-risk patients are generally treated with a combination of surgery and moderate intensity multiagent chemotherapy and this is associated with reported survival rates of up to 90 per cent. High-risk patients have a poor long-term outlook (survival 10 to 30 per cent) despite intensive multimodality therapy which may include myeloablative chemotherapy or chemoradiotherapy followed by autologous bone marrow transplantation.

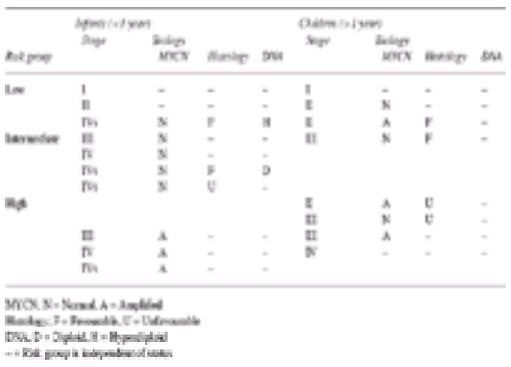


Table 5 Prognostic risk groups in neuroblastoma

Phaeochromocytoma

Phaeochromocytomas are uncommon tumours of the chromaffin cells that secrete excessive quantities of catecholamines into the circulation. They are named from the Greek *phaeos* (dusky) and *chromos* (colour) for the dark brown staining of the tumour cells with chromium salts. Their incidence is approximately 1.5 to 2 per million population per year and although most commonly presenting between the ages of 20 and 50 years they may present at any age and they occur with roughly equal frequency in both sexes. Their importance, apart from the risk of malignancy, is that they cause secondary hypertension which if diagnosed early may be cured by surgery but if left untreated may produce serious and sometimes lethal complications. Unfortunately, up to one-third of cases are not diagnosed during life and in many of these cases death is associated with general anaesthesia for an unrelated procedure.

Phaeochromocytomas may be found anywhere from the neck to the pelvis, in the distribution of the neural crest-derived sympathetic/adrenal system. However, 80 to 90 per cent arise in the adrenal medulla and most extra-adrenal tumours (also known as paragangliomas) are intra-abdominal and arise in the region of the aortic bifurcation from the organ of Zuckerkandl. Overall, only about 3 per cent of phaeochromocytomas are extra-abdominal and most of these are found in the paravertebral region in the chest but very rarely they may arise in the neck from the cervical ganglia.

The tumour may range in size from a few millimetres to a large cystic mass weighing approximately 3 kg but most tumours weigh less than 100 g, are between 3 and 5 cm in diameter, and are roughly spherical in shape. About 10 to 20 per cent are bilateral and this is more common in the familial varieties. Overall, 5 to 10 per cent are malignant and this is also more likely in familial phaeochromocytomas. In children, phaeochromocytomas are less frequently malignant but a greater proportion are extra-adrenal, more are bilateral or multiple, and there is a greater chance of familial disease or an association with multiple endocrine neoplasia (MEN) syndromes.

Overall, between 10 and 20 per cent of phaeochromocytomas are familial and may be associated with some well known syndromes. These cases are more likely to present in childhood and are frequently (>50 per cent) multiple or bilateral. In some, there is just a simple familial predisposition with an autosomal dominant pattern of inheritance without other associations. Others are associated with Sipple's syndrome also known as MEN type IIa (phaeochromocytoma, medullary carcinoma of the thyroid and C cell hyperplasia, and parathyroid hyperplasia or adenoma) or with MEN type IIb (phaeochromocytoma, medullary carcinoma of the thyroid and C cell hyperplasia, mucosal neuromata, and marfanoid features) both of which have an autosomal dominant pattern of inheritance. Approximately 5 per cent of patients with phaeochromocytoma have neurofibromatosis (von Recklinghausen's syndrome) that may be familial or sporadic but relatively few patients (approximately 1 per cent) with neurofibromatosis develop a phaeochromocytoma.

Phaeochromocytomas can present in many ways and this, combined with their relative rarity, explains why the diagnosis is often missed. Hypertension is the most consistent feature and is present in more than 90 per cent of cases. Two patterns are recognized: sustained hypertension with or without episodic rises, which occurs in about 50 per cent of adults with phaeochromocytoma, and paroxysmal attacks with normal blood pressure during the intervals. Most commonly, both the systolic and diastolic pressures are elevated but if adrenaline is the major secretory product systolic hypertension and tachycardia may be associated with a normal or low diastolic pressure. A consistently normal blood pressure in the presence of a functioning tumour occurs in approximately 10 per cent of cases. It is not cost effective to screen all patients with hypertension since less than 1 per cent will have a phaeochromocytoma but hypertension in association with one or more of the other clinical features of phaeochromocytoma, or hypertension in children or young adults, should lead to appropriate screening tests.

The triad of headache, often sudden in onset and described as pounding, diaphoresis (sweating), and tachycardia in a patient with hypertension has high sensitivity and specificity (each in excess of 90 per cent) for phaeochromocytoma but is relatively uncommon. Tremor, palpitations, chest pain, and feelings of apprehension or anxiety are more common and the episodes may be associated with facial pallor or a mottled appearance to the skin. Attacks may occur several times per day or as infrequently as once a week. Most are short lived, lasting from 15 min to 1 h, and may be followed by feelings of exhaustion, weakness, and muscle aches. Various precipitating factors have been described which are often associated with a change in pressure on the tumour. In the case of abdominal phaeochromocytomas, straining at stool, lying in a particular position, bending over, or abdominal palpation have all been described as precipitating factors. Some drugs, for example metoclopramide, tricyclic antidepressants, or naloxone may precipitate dangerous hypertensive crises.

Other cardiovascular abnormalities are common and may be the first manifestation of a phaeochromocytoma. Patients may present with symptoms and signs of acute myocardial infarction, with myocarditis, with pulmonary oedema, or with arrhythmias that may be precipitated by anaesthesia for unrelated surgery. Cardiovascular crises under anaesthesia, especially unexplained hypertension, tachycardia, oedema, or shock, should lead to performance of screening tests for phaeochromocytoma.

Neurological or psychiatric symptoms may also be presenting features as may strokes or seizures due to hypertension. The anti-insulin effect of catecholamines may produce glucose intolerance or diabetes, and other endocrine abnormalities may occasionally dominate the clinical picture. Some phaeochromocytomas secrete parathyroid hormone, ACTH, or other biologically active peptides (for example VIP) each of which may produce an appropriate cluster of symptoms and signs. Some patients present with acute abdominal pain which is thought to be due to bowel ischaemia. Occasionally, however, abdominal pain may be due to haemorrhagic necrosis of the tumour which may precipitate marked hypertension followed by hypotension, shock, and even death as a consequence of sudden withdrawal of catecholamines resulting in arterial and venous dilatation in the presence of a shrunken intravascular volume.

Physical signs of phaeochromocytoma other than hypertension are relatively uncommon. In approximately 50 per cent of cases, hypertension is associated with postural hypotension that is probably a consequence of impaired orthostatic autonomic reflexes resulting from chronic overexposure to catecholamines. A large phaeochromocytoma may be palpable as an abdominal mass. Café-au-lait patches and neurofibromata may be associated with phaeochromocytoma and the thyroid should be examined for carcinoma.

Evaluation of suspected phaeochromocytoma

The most useful diagnostic tests for phaeochromocytoma are measurement of 24-h urinary excretion of noradrenaline, adrenaline, and their metabolites (normetadrenaline, metadrenaline, and vanillylmandelic acid). Noradrenaline is generally the major secretory product but occasionally adrenaline, and rarely dopamine, may predominate. Although it has been suggested that adrenaline secretion may indicate an adrenal tumour it is clear that extra-adrenal phaeochromocytomas can also secrete adrenaline and thus the spectrum of catecholamine secretion is not a reliable guide to the site. In most cases excessive excretion of at least one

catecholamine or metabolite will be detected but the pattern of abnormality is quite variable and assays of any one of them (for example vanillylmandelic acid) to the exclusion of the others may be misleading. In some cases repeated assays may be required to make the diagnosis especially when the symptoms are sporadic or if clinical suspicion is high in the presence of a negative or equivocal result. Ideally the urine collection should be begun immediately after an attack.

The place of assays for plasma catecholamines is disputed but there is evidence that small tumours may release predominantly unmetabolized catecholamines into the circulation and may be best diagnosed by direct measurement of circulating catecholamines. Overall, assays for plasma catecholamines, urinary metadrenaline and normetadrenaline, or urinary vanillylmandelic acid each have a very high positive predictive value but combinations of tests give the greatest diagnostic accuracy. False-positive results may occur in patients taking certain drugs (phenothiazines, methyldopa, or monoamine oxidase inhibitors, for example) or with excessive intakes of tea, coffee, or chocolate and rebound hypertension with high levels of circulating and urinary catecholamines may occur following withdrawal of clonidine.

Occasionally, it may be necessary to distinguish between patients with a phaeochromocytoma and equivocally elevated plasma or urinary catecholamines and patients without a phaeochromocytoma who have excessive activity of the sympathetic nervous system. If clinical suspicion is high provocation or suppression tests may be performed but provocation tests must be undertaken with considerable caution because dangerous hypertensive crises may be precipitated. In the glucagon stimulation test, which is relatively free of adverse effects, a positive result is indicated by a clear increase in plasma catecholamines 1 to 3 min after administration of the drug. The clonidine suppression test relies upon the capacity of clonidine to inhibit neurogenically-mediated catecholamine release. In patients with a phaeo-chromocytoma, clonidine fails to suppress plasma catecholamine levels but the test should be administered with caution because clonidine suppression may result in hypotension and it should not be undertaken on patients with volume depletion or those already treated with b-blockers.

After the diagnosis has been established biochemically, localization of the tumour is important in order to confirm the diagnosis and to assist in the planning of the surgical approach. Modern non-invasive imaging can safely detect and localize virtually all tumours. CT of the abdomen and pelvis is the preferred modality since it will detect the majority of tumours. Whether it is positive or negative, however, a ^{131}I-MIBG scan provides useful additional anatomical and functional information and may locate occult second tumours or metastases to bone, liver, or lung. Preoperative CT and ^{131}I-MIBG together will locate more than 95 per cent of tumours. The place of MRI is less clear but it does have certain advantages over CT scan in relation to *in vivo* tissue characterization. Phaeochromocytomas produce a high T_2-weighted signal intensity unlike other tumours in which the signal intensity is low and it may be possible to distinguish a phaeochromocytoma from other adrenal masses. Unfortunately malignant and benign phaeochromocytomas have the same MRI signal and cannot be distinguished. MRI can, however, distinguish tumour from surrounding vascular structures without the need for the administration of contrast and, because there is no exposure to X-rays, MRI is the investigation of choice in pregnancy. Intravenous administration of iodinated ionic contrast material is not advised because it may precipitate an idiosyncratic rise in adrenaline and a dangerous hypertensive crisis. Older methods of localization, such as intravenous pyelography or venous catheterization and selective sampling at sites along the inferior and superior vena cavae, are rarely used today. Both are invasive and can provoke release of catecholamine and development of dangerous hypertensive crises. Occasionally, venous sampling is used in subjects in whom clinical and biochemical tests strongly suggest phaeochromocytoma but radiological studies fail to identify the tumour. Several days before embarking upon this investigation, adequate a- and b-adrenergic receptor blockade should be established.

Treatment

In most cases the appropriate treatment for phaeochromocytoma is surgery after medical therapy to prevent cardiovascular complications and reduce the risks of operative intervention. The medical preparation and anaesthesia is a critical aspect of surgery for phaeochromocytoma and is discussed in detail in Chapter 20.4.2.

The results of surgery are generally good: 75 to 90 per cent of patients can be expected to be cured of their hypertension and free of the risk of recurrent disease. However, the surgery is risky and close attention to management in the perioperative period is essential. In skilled centres the operative mortality is about 3 per cent. Several postoperative complications require special mention. Hypotension may result from marked arterial and venous dilatation following the sudden withdrawal of catecholamines and may be compounded by inadequate volume loading. The primary treatment here should be volume replacement rather than pressor agents. Hypoglycaemia may occur as a result of withdrawal of the anti-insulin effect of catecholamines and this may manifest as postoperative hypotension resistant to volume replacement and vasopressors. Finally, if the arterial blood pressure does not return to normal within about 2 weeks of surgery evidence of residual phaeochromocytoma or metastases should be sought by checking urinary or plasma catecholamines. Persistent hypertension may be due to hypertensive vascular disease or coexistent essential hypertension and if the catecholamine levels remain elevated the clonidine suppression test should distinguish those cases in which the hypertension is under neurogenic control. Patients should generally be followed for life with annual blood pressure checks and catecholamine studies.

Malignant phaeochromocytoma

The diagnosis of malignancy depends upon clinical or macroscopic pathological evidence of local invasion or the presence of metastases (usually in the liver, lung, or bone) rather than on histological criteria. Histology may be misleading because the usual features of malignancy (nuclear pleiomorphism, giant cells, frequent mitotic figures, and capsular invasion) may all occur in tumours that subsequently follow a benign course. DNA ploidy, however, may prove to be a useful discriminatory test: benign phaeochromocytomas tend to have diploid DNA content whereas malignant tumours are hyperdiploid. High plasma or urinary dopamine levels are suggestive, but not diagnostic, of malignancy.

Malignant phaeochromocytomas tend to be slow growing tumours and evidence of recurrence, or of malignancy following excision of an apparently benign tumour, may occur many years after surgery. Following excision some patients survive for more than 20 years but the overall 5-year survival is about 45 per cent.

Whenever possible surgical excision or debulking of the tumour should be attempted. Some patients will be cured, and debulking facilitates the medical control of symptoms since the level of catecholamine secretion is proportional to the size of the tumour. If there is residual catecholamine secretion then medical therapy aims to control the effects of circulating catecholamines by a- and b-adrenergic receptor blockade (often supplemented with calcium channel blockade), and to inhibit the synthesis of catecholamines using a-methyltyrosine. This ‘false' catecholamine precursor inhibits tyrosine hydroxylase which is the rate limiting enzyme for catecholamine synthesis.

Although malignant phaeochromocytomas respond poorly to chemotherapy or radiotherapy, combination chemotherapy with cyclophosphamide, vincristine, and dacarbazine may produce temporary control in some patients and administration of ^{131}I-MIBG to ablate residual primary or secondary deposits has produced short-term remission or palliation in some cases but evidence of a significant long-lasting benefit is lacking.

Surgery of the adrenal gland

The history of adrenal surgery begins in 1889 when Thornton resected a large adrenal tumour *en bloc* with the kidney in a young woman with hirsuitism. In 1914, Sargent performed the first planned operation to remove an adrenal tumour causing Cushing's syndrome and in 1926 César Roux in Lausanne and Charles Mayo in Rochester independently removed phaeochromocytomas. In the 1950s, after the discovery of cortisone, subtotal or total bilateral adrenalectomy was performed for adrenal hyperplasia and for pituitary-dependent Cushing's syndrome. However, this procedure was associated with high surgical morbidity, 2 to 5 per cent operative mortality, and the need for life-long replacement therapy, and has now been largely abandoned. Occasionally, surgical ablation of the adrenals is still used in Cushing's disease or in the treatment of hormonally-dependent tumours such as breast carcinoma.

Today the indications for adrenalectomy are largely restricted to resection of the various adrenal neoplasia. With modern imaging, exploratory operations have become virtually obsolete and a carefully planned approach to adrenal surgery is almost always possible. There are several operative approaches to the adrenal glands and each has its own particular indications. The choice of approach depends upon the adrenal pathology and indications for operation, the size and shape of the patient, and the experience and familiarity of the surgeon with the various options.

The adrenal appears to be well suited to a laparoscopic approach because of its small size, low incidence of malignancy, and the morbidity associated with conventional operative approaches. Laparoscopic adrenalectomy was first performed in 1992 and is rapidly gaining favour as the procedure of choice for apparently benign adrenal masses where operative removal is indicated. This approach has not been recommended for adrenal carcinoma or for large tumours (more than about 6 cm). Operative time is generally longer than for an open approach but advantages claimed for laparoscopic adrenalectomy are:

1. decreased operative morbidity;
2. decreased recovery time and length of hospital stay;
3. reduced postoperative analgesic requirements;
4. better cosmesis; and
5. technically better operative visualization.

There is, however, a lack of randomized studies to support these claims.

Anterior and lateral transabdominal approaches have been described and are the most widely utilized, but a posterior or lateral retroperitoneal approach, both of which

are more technically demanding, may offer advantages under some circumstances. The posterior retroperitoneal approach seems particularly suited to patients needing bilateral adrenalectomy for relatively small lesions but in Cushing's syndrome retroperitoneal adiposity may compromise visibility and a transabdominal approach may be more appropriate. The indications for each technique have yet to be defined but familiarity with a variety of approaches is likely to continue to be an advantage.

Despite the interest in laparoscopic adrenalectomy, there is still a place for open adrenalectomy. It is generally agreed that malignant tumours and large benign masses are best removed by conventional techniques and laparoscopic adrenalectomy may not gain widespread acceptance outside tertiary referral centres because of greater technical difficulty than for other laparoscopic techniques and the small numbers of patients requiring adrenalectomy.

The open posterior extraperitoneal approach was first described over 50 years ago and is still used for selected patients. The original ‘hockey stick' incision, however, is now rarely used and an oblique incision through the bed of the 11th or 12th rib is usually employed. The advantages of this approach are that it is direct, and relatively atraumatic, which minimizes postoperative pain, and both glands can be exposed simultaneously. The main disadvantage is that the operative field is relatively restricted which limits its usefulness to resection of small lesions such as adenomas less than 5 cm in diameter. It is not suitable for removal of large adrenal lesions (>5 cm) or potentially malignant tumours. Retroperitoneal exposure of the adrenal through the flank is rarely indicated but it may be the preferred route in obese patients with Cushing's syndrome in whom only one gland needs to be explored.

The open transabdominal approach, by an extended subcostal or bilateral subcostal incision, provides excellent exposure of both glands, if necessary, as well as the abdominal organs and the retroperitoneum. This approach is indicated for large lesions or potentially malignant tumours and it is mandatory for the resection of a phaeochromocytoma since it allows for early vascular control and for the possibility of bilateral, multiple, or extra-adrenal tumours. A midline approach may be used for extra-adrenal tumours in the organ of Zuckerkandl or the pelvis.

A thoracoabdominal approach provides outstanding exposure of the adrenal, but exposure of the contralateral gland is more difficult than by the anterior transabdominal route, and it is generally reserved for very large tumours that cannot be removed by the anterior approach.

Further reading

Adrenal surgery. *Urology Clinics of North America* 1989; **16**: 417–606. [This is a good general account of adrenal gland pathology and surgery including anatomy and physiology of the cortex and medulla together with articles on each of the major pathological conditions.]

Brossard J, Bernstein ML, Lemieux B. Neuroblastoma: an enigmatic disease. *British Medical Bulletin* 1996; **52**: 787–801. [A review on the impact of developments in knowledge of the genetics and biology of neuroblastoma on understanding of the natural history, prognosis, and treatment of this cancer.]

Castlebury RP. Neuroblastoma. *European Journal of Cancer* 1997; **33**: 1430–7. [A summary of the emerging biology of neuroblastoma and its relationship to current treatment and future strategies for management.]

Gajraj H, Young AE. Adrenal incidentaloma. *British Journal of Surgery* 1993; **80**: 422–6. [A concise and balanced review of the evaluation and management of adrenal incidentalomas from a British perspective.]

Graham DJ, McHenry CR. The adrenal incidentaloma: guidelines for evaluation and recommendations for management. *Surgical Oncology Clinics of North America* 1998; **7**: 749–64. [A detailed account of guidelines for evaluation and recommendations for management of adrenal incidentalomas.]

Jacobs JK, Goldstein RE, Geer RJ. Laparoscopic adrenalectomy. *Annals of Surgery* 1997; **225**: 495–502. [A review of a single centre experience of laparoscopic adrenalectomy and discussion of its role in management of adrenal masses.]

Korobkin M, Francis IR. Imaging of adrenal masses. *Urology Clinics of North America* 1997; **24**: 603–22. [A detailed review of the place of modern imaging modalities in the assessment of adrenal masses.]

Ross NS, Aron DC. Hormonal evaluation of the patient with an incidentally discovered adrenal mass. *New England Journal of Medicine* 1990; **323**: 1401–5. [A useful review giving the basis for a rational, evidence-based, approach to evaluation of an incidentally discovered adrenal mass.]

Seregni E. Radionuclide imaging of neuroendocrine tumours: biological basis and diagnostic results. *European Journal of Nucleic Medicine* 1998; **25**: 639–58. [A review of radionuclide imaging of neuroendocrine tumours including phaeochromocytomas and paragangliomas.]

20.4.2 Surgery of the adrenal gland

Collin J. Weber and C. Daniel Smith

Surgical approaches to adrenal tumors
Preoperative preparation of patients with pheochromocytoma
Intraoperative management of pheochromocytoma
Anterior approach for adrenalectomy
Right adrenalectomy
Left adrenalectomy
Posterior approach for adrenalectomy
Laparoscopic adrenalectomy
Transabdominal lateral approach
Complications of adrenalectomy
Further reading

Surgical approaches to adrenal tumors

Operative resection is the primary treatment for adrenal masses. The adrenal glands are positioned high in the retroperitoneum, at the junction of the chest and abdomen, and are surrounded by such important structures as the inferior vena cava, aorta, pancreas, kidneys, liver, spleen, and diaphragm. As a result, several approaches have been devised to expose the glands sufficiently for intraoperative evaluation and safe resection. Each approach has specific advantages in terms of exposure of the gland, the ability to sufficiently explore the entire abdomen, and the degree of invasiveness. The procedure chosen depends on the suspected pathology, the size of any adrenal mass, the potential for malignancy, occurrence of extra-adrenal or bilateral disease, and invasion of adjacent structures, as well as individual patient characteristics.

Preoperative preparation of patients with pheochromocytoma

Pharamacologic preparation of patients with pheochromocytoma is mandatory, since severe hypertension may ensue during general anesthesia and/or tumor manipulation at the time of resection. A hypertensive crisis can be extremely difficult to control intraoperatively and may result in stroke or cardiac arrest. The drug of choice for initial blockade is phenoxybenzamine. This drug is effective when administered orally and has a long biologic half-life of approximately 8 h. Treatment is usually given for at least 7 to 14 days prior to surgery when tachycardia is present. Phenoxybenzamine has a number of side-effects including gastrointestinal distress, excessive sedation, fatigue, and tachycardia. A b-antagonist such as propranolol or atenolol may be administered cautiously in the 2 to 3 days prior to surgery when tachycardia is present. Preoperative blockade with phenoxybenzamine offers several advantages, the most significant being the opportunity to correct the relative hypovolemia that is universally present in patients with pheochromocytoma, in order to prevent significant reductions in blood pressure following intraoperative resection of the tumor. b-Blocking agents are contraindicated in patients who are not being treated with b-blockade, because vasoconstriction mediated by high concentrations of catecholamines would then be unopposed by vasodilating b_2-receptors, potentially leading to exacerbations of hypertension, pulmonary edema, and negative inotropic effects on the heart.

Some idea of the adequacy of a-blockade can be obtained by sequential measurement of the hemotocrit, which fall as the blood volume expands. However, the most reliable method of assessing the adequacy of both a- and b-blockade is simply to chart the patient's pulse and blood pressure for several days. The blood pressure should be recorded both lying and standing. In addition, the pulse and blood pressure should be recorded if the patient has any symptomatic episode. The aim with effective treatment is a chart showing a steady pulse rate in the range of 60 to 70 beats/min, with no episodes of tachycardia, and a blood pressure also devoid of surges, with a small postural drop. Because of the slow onset of adequate a-blockade and its associated increase in blood volume, at least a week is required for effective preoperative preparation. In some cases, effective control of the blood pressure may be resistant to adrenergic-blocking drugs alone and other antihypertensive drugs, for example calcium channel-blocking agents and angiotensin-converting enzyme inhibitors, may be added. a-Methyltyrosine, which inhibits tyrosine hydroxylase, the enzyme involved in the conversion of tyrosine to dopa [an essential step in the synthesis of both adrenaline (epinephrine) and noradrenaline (norepinephrine)] may help stabilize blood pressure in difficult cases.

Intraoperative management of pheochromocytoma

Successful anesthetic management requires skillful control of acute hypertensive episodes, hypotension, and arrhythmias. It is important that the surgeon maintain a dialogue with the anesthesiologist throughout the operative procedure. Anesthetic induction is effectively accomplished with thiopental sodium. Anesthesia may then be maintained by use of a number of agents such as halothane, nitrous oxide, or enflurane. Cyclopropane anesthesia should be avoided, since it may sensitize the myocardium to the actions of epinephrine. Likewise, curare should not be used.

Sodium nitroprusside is an indispensable drug for intraoperative use in the resection of pheochromocytoma. This direct peripheral vasodilator is the drug of choice for the management of significant intraoperative hypertension. Nitroprusside is preferable to phentolamine, a short-acting a-blocking agent, since its action is prompt and its effects end abruptly when the infusion is stopped. In spite of adequate preoperative a-adrenergic blockade, many patients with pheochromocytoma become hypotensive and require large volumes of crystalloid fluid immediately after the main adrenal vein draining the pheochromocytoma has been ligated. This hypotension is due to the drop in circulating catecholamines and the continuing effects of previously administered a-adrenergic-blocking agents. Intraoperative arrhythmias are not uncommon during resection of pheochromocytomas, and may be treated with lidocaine (lignocaine).

Anterior approach for adrenalectomy

The anterior approach is commonly used for pheochromocytomas, suspected cancerous lesions or bilateral lesions, and for adrenal masses more than 6 cm in diameter. It facilitates complete exploration of the entire abdomen and provides very good exposure of both glands.

The patient is placed supine on the operating table, over a roll positioned transversely at the level of the kidneys. This positioning displaces the retroperitoneal structures more anteriorly, providing easier access to the adrenals. Either a standard midline celiotomy incision or a bilateral subcostal incision is used. Very large adrenal masses, such as pheochromocytomas or invasive adrenocortical carcinomas, require greater exposure for adequate intraoperative assessment and complete resection. In these cases, the patient may be placed in the lateral decubitus position with the affected side superiorly, and a thoracoabdominal incision made (Fig. 1).

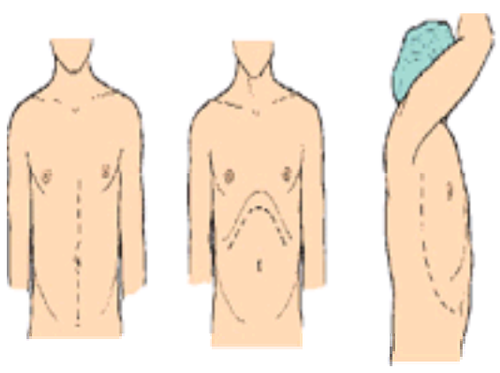


Fig. 1. Incisions used for anterior approach to the adrenals.

When an anterior approach is used to resect a pheochromocytoma, thorough exploration of the abdomen is advisable. Extra-adrenal pheochromocytomas (paragangliomas) most commonly occur in the organ of the Zuckerkandl, a sympathetic ganglion located near the origin of the inferior mesenteric artery just to the left of the midline. Pheochromocytomas also arise in the renal hila, the periaortic ganglia, and less commonly, the urinary bladder.

Right adrenalectomy

The right adrenal gland is best approached anteriorly by mobilizing the right lobe of the liver medially and entering the retroperitoneum behind the liver. The triangular, coronary, and hepatocolic ligaments of the right hepatic lobe are divided with scissors or cautery, freeing the liver from the diaphragm and colon. The entire right lobe of the liver is retracted anteriorly and medially. The right adrenal gland is identified just superior to the kidney. The hepatic flexure of the colon may also be mobilized from the retroperitoneum and retracted medially to provide increased exposure (see Fig. 2).

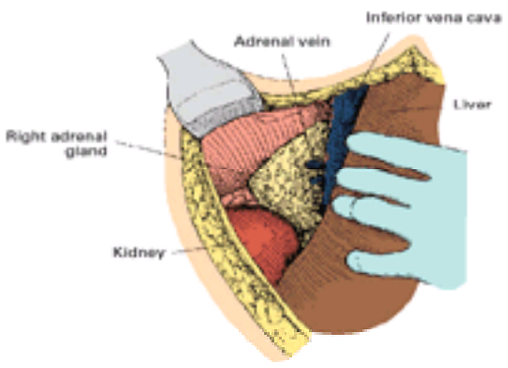


Fig. 2. Exposure of the right adrenal gland via anterior approach, with medial mobilization of the liver and exposure of the posterolateral border of the inferior vena cava.

Before dissecting a large adrenal mass or a pheochromocytoma, it is important to gain control of the venous drainage. The right adrenal vein is short and stubby, approximately 5 mm long, and drains directly into the inferior vena cava at its posterolateral aspect. Torsion of the vein, which may occur during the mobilization of a large adrenal mass, may tear the vessel and considerable, life-threatening hemorrhage may ensue. Furthermore, manipulation of a pheochromocytoma may cause catecholamine discharge and labile intraoperative blood pressure. Therefore, it is preferable to isolate and ligate the adrenal vein early in the operation.

If a large adrenal mass precludes identification of the right adrenal vein from the superior aspect of the gland, mobilizing the duodenum using the Kocher maneuver and retracting it medially will expose the vena cava inferior to the gland. Dissection upward along the posterolateral cava may then allow identification of the right adrenal vein from its inferior aspect. The vein may be divided immediately after ligation or after the adrenal gland has been dissected free. Early division prevents traction on the vein and vena cava during subsequent dissection of the gland.

Once the right adrenal vein has been ligated, an avascular plane may be developed underneath the gland from its lateral aspect, allowing palpation of the adrenal gland and tumor. Numerous arteries typically penetrate the superior, medial, and inferior perimeter of the gland, arising from the inferior phrenic artery, the aorta, and the right renal artery, respectively. These arteries are sequentially ligated as they are encountered during the dissection of the gland.

Adrenal tumors, such as adrenocortical carcinoma, that invade adjacent structures often require *en bloc* resection of all involved tissues. If nephrectomy is contemplated, the function of the contralateral kidney should always be confirmed. If preoperative studies indicate compression or involvement of the inferior vena cava by an adrenal mass, preoperative evaluation of the cava with intravenous contrast or ultrasonography may be warranted to assess blood flow and further define the anatomy. Intraoperative ultrasound to evaluate tumor involvement of the inferior vena cava may also be helpful. If tumor is identified within the cava, it can usually be extracted by venotomy, with prior control of the proximal and distal cava. Occasionally, cardiopulmonary bypass is necessary to avoid tumor embolization during resection.

Left adrenalectomy

The left adrenal gland is most clearly exposed form the anterior approach by mobilizing the left abdominal viscera medially. The splenic flexure of the colon is mobilized medially. The spleen, stomach, and tail of the pancreas are then lifted and rotated medially *en bloc*, exposing the left gland (see Fig. 3). Occasionally, with small, left adrenal tumors, elevation of the body of the pancreas may provide sufficient exposure to allow safe tumor resection.

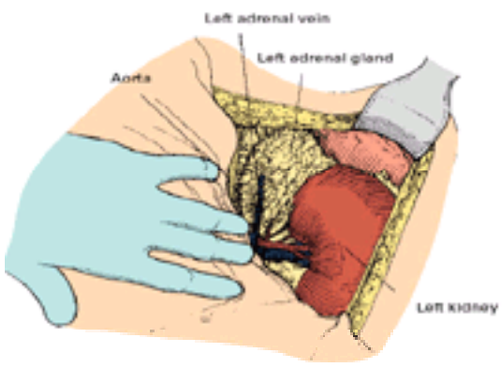


Fig. 3. Anterior approach to the left adrenal, with medial mobilization of the splenic flexure of the colon, spleen, stomach and tail of pancreas.

The vein is identified and ligated first. The left adrenal vein is longer and located near the interior pole of the gland and more anteriorly; it courses medially to drain into the left renal vein. Once the adrenal vein has been ligated, the arteries are clipped circumferentially, and divided. Rarely, the left adrenal vein may drain directly into the inferior vena cava, in which case it may not be visualized until the adrenal gland is completely mobilized.

Posterior approach for adrenalectomy

Small, isolated adrenal tumors and those that appear benign may be resected using a posterior approach. Although exposure of the gland is more limited than from the anterior approach, the peritoneal cavity is not violated. Postoperative ileus is rarely a problem, and the postoperative rehabilitation period may be shorter. The patient is placed in the prone jack-knife position on two vertically placed rolls. This positioning allows access to either side of the back and displacement of the abdominal contents away from the retroperitoneum. The skin incision begins three fingerbreadths from the midline at the tenth rib (see Fig. 4). The incision continues inferiorly and is curved increasingly more laterally to a maximum of four fingerbreadths from the midline at its most inferior extent, the superior aspect of the posterior iliac crest.

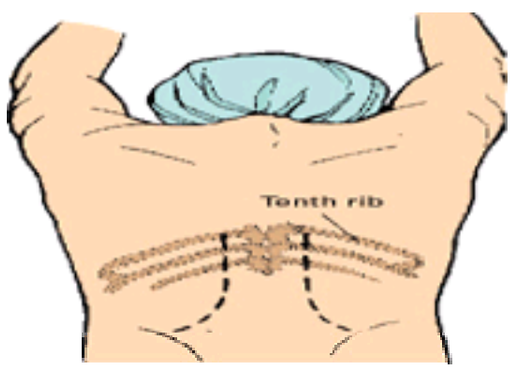


Fig. 4. Position of incisions for posterior approach to the adrenals.

The incision is carried down through the subcutaneous tissue and latissimus dorsi muscle until the posterior lamella of the lumbodorsal fascia is encountered. A vertical incision is made in this fascia to uncover the sacrospinalis muscle, which is retracted medially. Its attachments to the eleventh and twelfth ribs are divided. The lumbar vessels and cutaneous nerves are ligated and divided. The periosteum of each rib is dissected off the bone, and the intercostal vessels are ligated and divided. The intercostal nerve is carefully retracted superiorly. The eleventh and twelfth ribs are each cut as far medially as possible using a bone cutter. Removal of both these ribs allows better exposure of the adrenal gland, which typically lies slightly more superiorly. The middle and anterior lamellae of the lumbodorsal fascia are then opened. Entrance into the retroperitoneum is evident by visualizing Gerota's fascia and retroperitoneal fat.

Gentle traction on the kidney exposes the suprarenal area and the adrenal gland. Exposure is increased by dividing the diaphragm to the level of the tenth rib using cautery, taking care not to enter the pleural space. The parietal pleura is bluntly dissected superiorly from the diaphragm. If the pleural cavity is entered, which is not uncommon, the opening is easily repaired at closure as the pneumothorax is evacuated via a small rubber catheter during hyperinflation of the lung.

From the posterior aspect, the adrenal arteries lie superficial to the adrenal veins. The gland is, therefore, initially dissected free circumferentially from its lateral to medial margins; any arteries encountered in the process are clipped and divided. The inferior border of the adrenal gland is left attached to the kidney to allow inferior retraction of the gland.

As one dissects deeper into the retroperitoneal cavity, the adrenal vein is identified, ligated, or clipped and divided. The right adrenal vein should be divided before extensive mobilization of the right adrenal gland, since it is quite short and easily torn (see Fig. 5). The left adrenal vein lies anterior to the left adrenal gland. This makes isolation and ligation of the left adrenal vein difficult from the posterior aspect. Typically, therefore, the left adrenal vein is ligated at the very end of the adrenalectomy (see Fig. 6).

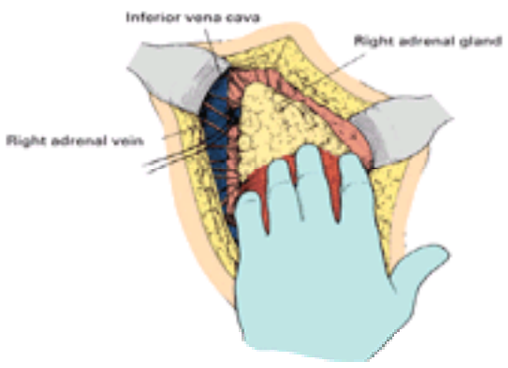


Fig. 5. Exposure of the right adrenal via a posterior approach.

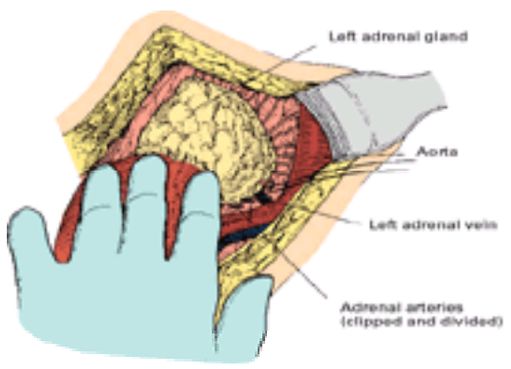


Fig. 6. Exposure of the left adrenal via a posterior approach.

Laparoscopic adrenalectomy

Laparoscopic adrenalectomy offers several advantages, including more cosmetically acceptable incisions, decreased postoperative pain, and a shorter postoperative rehabilitation period. The laparoscopic approach may be appropriate for small (< 5 cm) adrenocortical tumors (Cushing's syndrome, aldosteronomas, ‘incidentalomas') and for some pheochromocytomas, which are solitary, confined to the adrenal, secreting little or no norepinephrine, and with confirmation by computed tomography and magnetic resonance imaging that there are no synchronous extra-adrenal paragangliomas. As a surgeon's experience with this approach increases, a greater percentage of adrenal tumors may be safely resected laparoscopically. If an adrenocortical carcinoma, malignant pheochromocytoma, or paraganglioma is suspected based on tumor size (> 5 cm) or imaging characteristics (adherence or invasion of adjacent structures), an open approach is wise.

Transabdominal lateral approach

The transabdominal lateral approach to adrenalectomy as originally described by Gagner places the patient in lateral decubitus position (see Fig. 7) to allow a gravity-facilitated exposure of the adrenals. In this way, tissue and organs overlying the adrenals do not need to be manipulated with laparoscopic instruments, and the complications and bleeding associated with such manipulation are avoided.

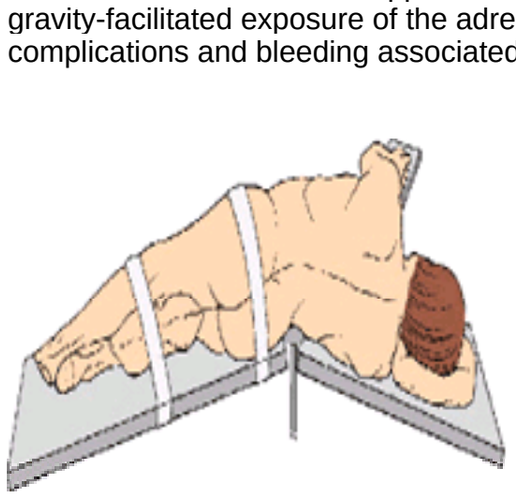


Fig. 7. Patient positioning for lateral-approach adrenalectomy (note kidney rest and table break at iliac crest).

Left adrenalectomy

After induction of general, endotracheal anesthesia, a Foley catheter and an orogastric tube are placed and the patient is placed in the lateral decubitus position with the left side up (Fig. 7). The patient's iliac crest is positioned immediately over the table's kidney rest and the kidney rest elevated and table extended. This maximizes the distance between the iliac crest and the costal margin in the midaxillary line for subsequent cannula insertion. Using a bean-bag on the operating table facilitates stabilization of the patient in this position. The left arm is positioned over the chest on a sling.

The patient is prepared and draped so that either laparoscopy or open surgery can be performed. Cannula sites and uses are shown in Fig. 8. Preincisional local anesthesia is used, along with moderate reverse Trendelenburg positioning. Carbon dioxide pneumoperitoneum to 15 mmHg is initiated with a Veress needle inserted at the midclavicular line below the left costal margin (cannula site for the camera). A fourth cannula in the posterior axillary line is optional. This fourth cannula has been avoided in nearly 50 per cent of our patients. A 30°-angled laparoscope is used.

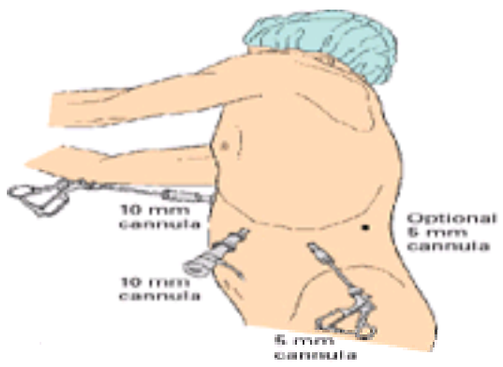


Fig. 8. Cannula sites and uses for lateral-approach, laparoscopic left adrenalectomy.

The first step of the operation is to mobilize the left abdominal viscera medially as in the anterior approach to open adrenalectomy; this is accomplished by incising the splenorenal ligament and mobilizing the spleen laterally (Fig. 9). The decubitus positioning facilitates this dissection and mobilization. With gravity pulling the spleen medially and away from the anterior surface of the kidney, the spleen and tail of the pancreas are dissected away from the retroperitoneum and the superior pole of the kidney as the adrenal is exposed. This dissection plane is relatively avascular. If excessive bleeding is encountered, the wrong plane of dissection is being developed. It is important to continue this mobilization up to the diaphragm and close to the greater curve of the stomach and short gastric vessels. The dissection along the anterior surface of the kidney and the adrenal continues until the inferior pole and medial border of the adrenal are exposed (see Fig. 10). Depending on the adrenal pathology and amount of retroperitoneal fat, the lateral and anterior surface of the adrenal gland will become visible during this dissection. It is important to avoid mobilizing the adrenal gland along its lateral edge too early in the exposure. If this mistake is made, gravity will allow the mobilized adrenal to fall medially and prevent visualization and access to its medial and inferior edges, where the left adrenal vein is most likely encountered.



Fig. 9. Initial lines of dissection along splenorenal and splenocolic ligaments.

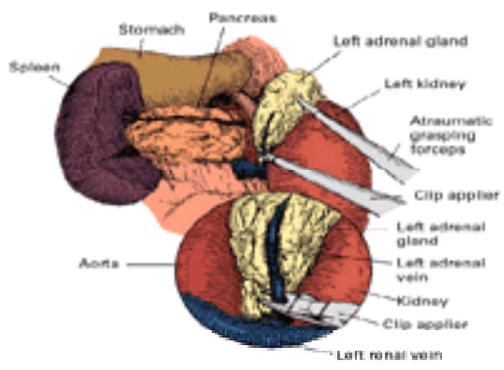


Fig. 10. Exposure of the left adrenal and isolation and control of the left adrenal vein.

When retroperitoneal fat prevents clear visualization of the gland, its medial and inferior edge can be localized by palpation, since the adrenal is ballotable in the retroperitoneal tissue along the anterior surface of the kidney. In even the most obese patients or those with Cushing's, this technique allows identification of the dissection plane between the anterior surface of the kidney and the inferior border of the adrenal. Once this cleavage plane has been identified, careful dissection with a hook cautery eventually exposes the inferior and medial edge of the gland. In these difficult situations, intraoperative laparoscopic ultrasonography may be helpful to locate the gland, but is rarely needed.

The next step is isolation of the left adrenal vein. With small tumors (< 5 cm) this is most easily accomplished by first dissecting the inferior and medial aspect of the adrenal, staying close to the gland until the vein is isolated and clipped (see Fig. 10). A right-angle dissector greatly facilitates this exposure and isolation. Risk of injuring the left renal vein is minimized by staying close to the adrenal gland during this dissection. The adrenal vein is then clipped with a medium-large Ligaclip. Once the vein has been transected, dissection continues from inferior and medial to superior and lateral, following the anterior surface of the kidney. For large tumors, early identification of the adrenal vein may be difficult. In these cases, the gland may be mobilized laterally and inferiorly to identify the inferior border of the gland and the adrenal vein.

As in open adrenalectomy, final dissection and gland excision progresses from medial to lateral and inferior to superior, using an ultrasonic dissector or monopolar electrocautery with a hook dissector. The inferior phrenic artery is frequently encountered along the superior edge of the adrenal and should be sought and ligated with Ligaclips.

Before specimen extraction, the operative field is carefully inspected for hemostasis. Once hemostasis has been ensured, the adrenal is placed in a specimen retrieval sac that has been inserted through the medial 10-mm cannula. The bag must be stout enough not to rupture during extraction. The fascia of the cannula site of extraction may need to be stretched with a Kelly clamp to facilitate removal. For large tumors, the entire incision may need to be extended. The adrenal should not be morcellated; its histologic architecture must be preserved for pathologic analysis.

The operation is completed by closing the fascia of the 10-mm incisions with absorbable suture. If a significant amount of irrigation was used during the procedure, the

fluid will tend to disperse to the lower and right abdomen, out of the reach of suction. Aspiration of this irrigant may be difficult, and when the patient is again supine, it will tend to drain from the lateral cannula site. In such a circumstance, a soft Silastic drainage catheter positioned in the left upper quadrant and exiting the lateral-most cannula site may help control and evacuate this fluid during the first 12 hours postoperatively. This drain is removed on the first postoperative morning. When irrigation is minimal (< 500 ml) no drain is necessary.

Right adrenalectomy

The patient is placed in the lateral decubitus position with the right side up. The remainder of patient positioning is identical to that described for left adrenalectomy. Pneumoperitoneum to 15 mmHg is initiated with a Veress needle inserted at the midclavicular line below the right costal margin (cannula site for the camera). After establishing pneumoperitoneum, the remaining cannula sites are marked (see Fig. 11). A fourth cannula in the epigastrium is necessary for a retractor to elevate the right lobe of the liver. This cannula site should be several centimeters below the costal margin, allowing the angle of retractor insertion to be parallel to the undersurface of the right lobe of the liver. Too cephalad an insertion will create an acute angle to the undersurface of the liver, making positioning of the liver retractor difficult. The liver retractor can be held by the first assistant/camera operator. Alternatively, a table-mounted liver retractor may be used, allowing the first assistant to stand at the surgeon's left side.

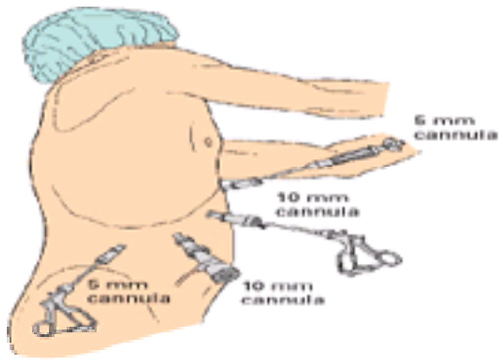


Fig. 11. Cannula sites and uses for lateral approach, laparoscopic right adrenalectomy.

With the liver retractor positioned, the anterior surface of the right kidney and the lateral edge of the interior vena cava can be clearly seen. The dissection commences by creating a hockey-stick shaped incision along the retroperitoneal attachment of the right lobe of the liver and the medial border of inferior vena cava (see Fig. 12). This mobilizes the right lobe of the liver posteriorly and allows exposure of the anterior surface of the adrenal as the liver is pushed cephalad. Except with exceptionally large or high-riding tumors, the triangular ligament of the liver does not need to be incised. The lateral border of the inferior cava is carefully exposed, looking for the right adrenal vein, remembering that it is typically broad, short, and enters the cava slightly posteriorly. A blunt-tipped, right-angle dissector is useful for this dissection. Once the adrenal vein has been isolated, it is ligated (see Fig. 13) with three medium-large Ligaclips proximally and two distally. Due to the short length of the vein, the proximal-most clip should be immediately at the edge of the cava.

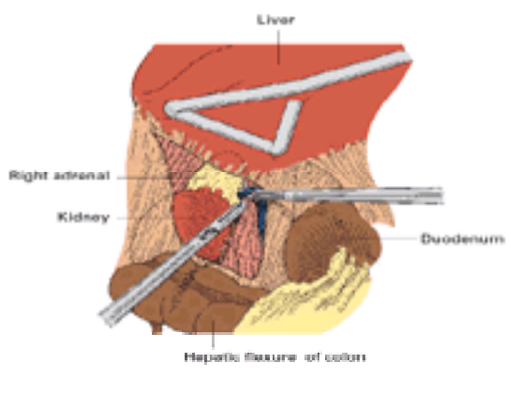


Fig. 12. Hockey-stick incision for initial dissection of right adrenal gland, and exposure and isolation of the right adrenal vein using a blunt-tipped, right-angle dissector.

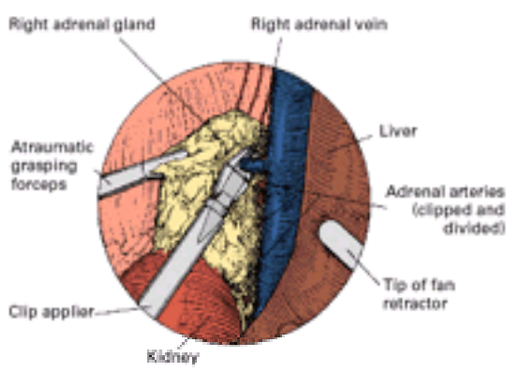


Fig. 13. Ligation of right adrenal vein with Ligaclips.

Once the right adrenal vein has been controlled and divided, dissection continues to expose and mobilize the medial edge of the adrenal. As with the left adrenal, the lateral edge should not be dissected too early, since gravity will cause the laterally mobilized gland to hang over the medial edge, making visualization difficult. By dissecting from medial to lateral and inferior to superior, the superior pole of the kidney can be used as a dissection plane as the dissection progresses in a direction away from any anatomic concerns (inferior vena cava and renal vein). Once excised, the adrenal is placed in a specimen retrieval sac and removed intact through the medial-most 10-mm cannula site.

Bilateral adrenalectomy

Bilateral adrenalectomies are performed in the manner already described for each individual side. Since right adrenalectomy has a higher risk of conversion to open adrenalectomy due to the immediate consequences of injury to an adrenal vein/cava, it is desirable to perform left adrenalectomy first. In this way, the patient will have the greatest likelihood of benefiting from a laparoscopic approach. Before repositioning for right adrenalectomy, the entire left adrenalectomy is completed, including wound closure and abdominal desufflation (minimizing duration of CO_2 pneumoperitoneum).

Complications of adrenalectomy

The most significant complications that may occur intraoperatively during adrenalectomy include laceration of the renal vein or inferior vena cava, ligation of a renal artery, damage to the ureter, or injury to an adjacent organ such as the tail of pancreas, the liver, or the colon. Postoperative monitoring of the adrenalectomy patient should include frequent measurements of vital signs, urine output, electrolytes, and hematocrit.

Complications following adrenalectomy for Cushing's syndrome are frequent and concern poor wound healing, impaired resistance to wound infections and

thromboembolism, all related to long-standing high concentrations of corticosteroids. Furthermore, perioperative exogenous steroid therapy must be carefully monitored. Perioperative corticosteroids may be administered intramuscularly and/or intravenously and should include at least 40 mg of methylprednisolone intravenous on call to the operating room and every 6 h thereafter on the day of surgery. Prednisolone therapy may be tapered over the first 3 days postoperatively to approximately 40 mg intravenous each 8 h. After resumption of oral intake, therapy may be switched to oral prednisone, approximately 10 mg daily given in divided doses. This is continued for at least the first month after surgery and thereafter tapered under supervision to the equivalent of approximately 25 mg cortisone/day as chronic maintenance therapy, with great attention paid to increased cortisone or prednisone therapy given immediately upon diagnosis of any infection or stress, for the lifetime of the patient.

In 1960, Nelson *et al.* described a patient who developed a pituitary tumor after bilateral adrenalectomy for Cushing's disease. The patient suffered increased skin pigmentation, amenorrhea, visual problems, and high concentrations of adrenocorticotrophic hormone were demonstrated in serum. This characteristic clinical presentation, now designated Nelson's syndrome, has been described in approximately 10 per cent of patients following bilateral adrenalectomy. Careful follow-up is required to identify this syndrome since transphenoidal pituitary resection may be required.

Further reading

Auda SP, Brennan MF, Gill JR. Evolution of the surgical management of primary aldosteronism. *Annals of Surgery* 1980; **191**: 1–7. [A review of the history of adrenal procedures for hyperaldosteronism]

Bellantone R *et al.* Role of reoperation in recurrence of adrenal cortical carcinoma: results from 188 cases collected in the Italian National Registry for Adrenal Cortical Carcinoma. *Surgery* 1997; **122**: 1212–18. [A large experience with surgery for adrenocortical carcinoma.]

Bravo EL, Gifford RW. Pheochromocytoma: diagnosis, localization and management. *New England Journal of Medicine* 1984; **311**: 1298–303. [A landmark description of diagnostic approaches to pheochromocytoma.]

Bravo E, Gifford R, Manager W. Adrenal medullary tumors: pheochromocytoma. In: Mazzaferri E, Samaan N, ed. *Endocrine tumors*, pp. 426–47. Blackwell Scientific, Boston, 1993. [An important summary of successful management of adrenal tumours.]

Brunt LM *et al.* Laparoscopic adrenalectomy compared to open adrenalectomy for benign adrenal neoplasms. *Journal of the American College of Surgeons* 1996; **183**: 71–3. [Alternative surgical approaches to adrenalectomy.]

Cohn N, Gottesman L, Brennan M. Adrenocortical carcinoma. *Surgery* 1986; **100**: 1170–7. [The Memorial-Sloan-Kettering experience with adrenal carcinoma.]

Conn JW. Primary aldosteronism, a new clinical syndrome. *Journal of Laboratory and Clinical Medicine* 1955; **45**: 3–17. [The first description of hyperaldosteronism.]

Duh QY *et al.* Laparoscopic adrenalectomy. Comparison of the lateral and posterior approaches. *Archives of Surgery* 1996; **131**: 875–6. [The pros and cons of two accepted approaches to laparoscopic adrenalectomy.]

Fernandez-Cruz L *et al.* Laparoscopic unilateral and bilateral adrenalectomy for Cushing's syndrome. Transperitoneal and retroperitoneal approaches. *Annals of Surgery* 1996; **224**: 734–6. [Application of laparoscopy to adrenalectomy in Cushing's syndrome.]

Gagner M *et al.* Laparoscopic adrenalectomy: lessons learned from 100 consecutive procedures. *Annals of Surgery* 1997; **226**: 246–7. [A large early experience with laparoscopic adrenalectomy.]

Lee JE *et al.* Unknown primary cancer presenting as an adrenal mass: frequency and implications for diagnostic evaluation of adrenal incidentalomas. *Surgery* 1998; **124**: 115–22. [Important considerations in management of non-functioning adrenal tumours.]

Nelson DH, Meakin JW, Thorn GW. ACTH-producing pituitary tumors following adrenalectomy for Cushing's syndrome. *Annals of Internal Medicine* 1960; **52**: 560–9. [An historical account of pituitary complications of adrenalectomy.]

Orchard T, Grant CS, van Heerden JA, Weaver A. Pheochromocytoma—continuing evolution of surgical therapy. *Surgery* 1993; **114**: 1153–9. [A review of adrenal surgery.]

Smith D, Weber C, Amerson J. Laparoscopic adrenalectomy: the new gold standard. *World Journal of Surgery* 1999; **23**; 389–96. [A recent description of the laparoscopic approach to the adrenal glands.]

Thompson G *et al.* Laparoscopic versus open posterior adrenalectomy: a case–control study of 100 patients. *Surgery* 1997; **122**: 132–6. [A comparison of results with open versus laparoscopic adrenalectomy.]

Well SA *et al.* Therapeutic controversy. The role of laparoscopic surgery in adrenal disease. *Journal of Clinical Endocrinology and Metabolism* 1998; **83**: 3041–9. [The views of clinical endocrinologists regarding laparoscopic adrenalectomy.]

21.1 Benign conditions of the breast

Michael J. Greenall

Anatomy and physiology of the breast
Development
The anatomy of the adult breast
Physiological changes in the breast
Breast milk
Benign breast disease
Classification
Symptoms
Breast pain (mastalgia)
Benign breast lumps
Fibroadenomas and associated conditions
Breast cysts
Cyclical nodularity
Galactocele
Sclerosing adenosis
Fat necrosis
Lipoma, adenolipoma
Chronic abscess
Normal structures
Disorders of the nipple and periareolar tissue
Infection of the breast
The relation of benign to malignant breast disease
Descriptive terms
Studies of risk
Further reading

The demand for specialist treatment of patients with breast disease is increasing for several reasons. Firstly, the subject has become more complex, and an integrated approach involving not only surgeons but also radiologists, radiotherapists, pathologists, and medical oncologists is necessary. Secondly, the scope of the subject has increased dramatically. Not long ago the majority of surgeons who dealt with breast disease perceived their role as little more than differentiating benign conditions from malignancy and treating the latter. Patients now demand specific management of benign disorders that were previously ignored. Increased understanding of the relation between benign and malignant disease has increased awareness of conditions previously dismissed as ‘benign changes', mastitis, or fibroadenosis. Finally, the development of new therapeutic and diagnostic methods, such as screening, breast conservation, and adjuvant chemotherapy, has increased the complexity of the management of breast cancer itself.

Many institutions have now developed specialized breast disease units with surgeons having a special interest and training.

Anatomy and physiology of the breast

Development

The breasts are modified sweat glands in that they are embryologically derived from a downward growth of ectoderm into the underlying mesenchyme. The first stage of development occurs at 6 or 8 weeks of gestation, when two strips of thickened ectoderm, the mammary ridges, grow in a line extending from the embryonal axilla to the inguinal region. In many animals breasts develop along the whole length of this ridge, but in humans true breast tissue occurs only in the pectoral region.

Branching epithelial cords appear as 15 to 20 buds, which eventually become lactiferous ducts and associated alveoli. Each cord becomes surrounded by a stroma of mesenchymally derived fat, connective tissue, and vascular tissue. These cords form the basis of the segmental pattern of the adult breast.

Towards the end of gestation the lactiferous ducts become canalized and open on to a pit in the epidermis. At the same time, mesenchymal proliferation beneath the epidermis allows nipple development; failure to do so results in inversion.

Lobular development within the breast occurs particularly at puberty, with the ultimate development of 15 to 20 lobes, each of which drains into a single lactiferous duct; true secretory alveoli develop only during pregnancy and lactation.

Few breast changes occur in the early years of childhood. However, at the age of 10 years there is growth of mammary tissue beneath the areola, producing a characteristic protuberance on the chest wall called the breast bud or mound. True nipple development occurs at about the age of 12 years, followed 2 or 3 years later by further subareolar growth and the formation of the bulk of the breast tissue. Finally, there is areolar recession, resulting in the classical shape of the adult breast. These changes at puberty are a result of the action of follicle-stimulating and luteinizing hormones, produced by hypothalamic stimulation of the pituitary gland on oestrogen production from the ovary.

The anatomy of the adult breast

Gross anatomy

Although the adult breast varies greatly in size, its base is fairly consistent anatomically, extending from the second to the sixth rib in the midclavicular line and overlying the pectoralis major, serratus anterior, and external oblique muscles. Medially, the breasts reach the sternal edge and laterally the midaxillary line. The pyramidal axillary tail extends into the axilla.

The fascial covering of the breast is of importance to surgical technique. As it develops from the skin the breast is invested with superficial fascia, which divides into two layers. The anterior layer provides a plane of dissection subcutaneously between the relatively small, subcutaneous fat lobules and the larger lobules of mammary fat. The posterior layer of superficial fascia abuts the deep fascia derived from the pectoralis major and serratus anterior, thus producing a potential space; this retromammary space is easily located surgically (Fig. 1).

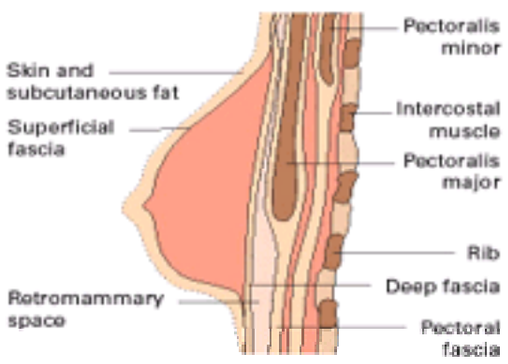


Fig. 1. The fascial coverings of the breast.

Between the two layers of superficial fascia there are condensations of fibrous tissue. These form the suspensory ligaments of Cooper, which both divide the breast into lobes and act as a supportive framework.

Blood supply

The breast has a rich vascular supply from the internal mammary artery and from the thoracoacromial, subscapular, and lateral thoracic branches of the axillary artery. The main supply is via the second perforating branch of the internal mammary and lateral thoracic arteries; the former should be preserved during subcutaneous mastectomy.

The regional venous drainage forms a rich subareolar plexus that drains via the intercostal, internal mammary, and axillary veins.

Lymphatic drainage

The lymphatic drainage of the breast has obvious implications in the management of breast cancer and has been well documented in an elegant series of studies by Turner-Warwick.

Approximately 75 per cent of the lymphatic vessels drain to the 30 or so regional lymph nodes in front of and below the axillary vein. These nodes can conveniently be subdivided into three main groups according to their relation to the pectoralis minor muscle: nodes at level 1 lie below the muscle, level 2 lymph glands lie behind it, and those of level 3 are in the apex of the axilla above the muscle. Most lymph drains from nodes at level 1 sequentially to those at levels 2 and 3; a small amount drains to the subscapular and interpectoral (Rotter's) groups of nodes. Extension of tumour to these latter nodes may imply retrograde spread and heavy involvement of axillary nodes by metastases (Fig. 2). A small amount of lymph drains from the superior aspect of the breast directly to the apical nodes. These pathways account in part for the well-documented, but relatively unusual, findings of nodal disease at level 3 but negative nodes in the lower axillary chain.

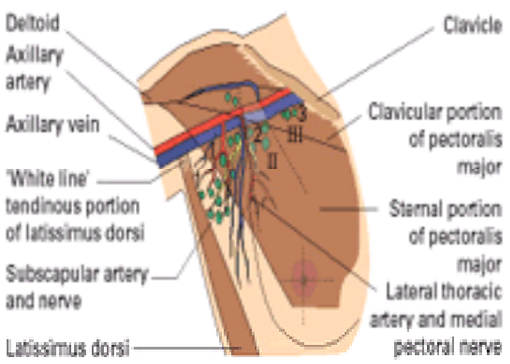


Fig. 2. The anatomy of the axilla showing lymph-node groups, major vessels, and nerves: (1) most lateral border of level 1 at medial border of latissimus dorsi; (2) central portion of level 2; (3) most medial border of level 3 at Halstead's ligament.

About 25 per cent of lymph drains to the internal mammary nodes in the second, third, and fourth intercostal spaces. This drainage is from the whole breast, although the majority of these pathways arise in its medial half. Some lymph drains to the opposite breast and down the rectus sheath. Drainage to the contralateral axilla and to intercostal nodes via the pectoral fascia is, however, minor. Similarly, there are few direct lymphatic communications between the breast and the nodes above the axillary vein.

The distribution of major lymphatics accompanies the blood supply. The original description by Sappey in 1885 of a subareolar plexus that receives lymph centripetally and then redirects it to the axilla has been discounted.

Innervation

The nervous supply to the breast is primarily by sensory and sympathetic nerves; little parasympathetic innervation has been demonstrated. The nipple has a rich sensory supply, whereas most of the sympathetic innervation is to the breast parenchyma.

Microscopic anatomy

The microscopic anatomy of the breast is a complex subject that has been the source of much controversy and confusion. Basically, there is a system of major ducts arranged in a segmental and radial pattern; these lead to the secretory component of the breast, the terminal ductal–lobular unit. The breast is thus subdivided into lobes, although these are not precisely defined anatomically. The interlobar fascia (the most prominent of which are Cooper's ligaments) is dense and fibrous, whereas the periductal and intralobular connective tissue is much more loose and vascular. The connective tissue round the lobule is particularly loose, allowing expansion during pregnancy and lactation (Fig. 3).

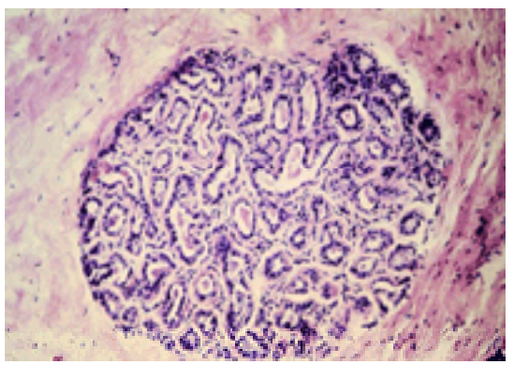


Fig. 3. A normal lobule composed of multiple acini with surrounding loose connective tissue (by courtesy of Dr D. Tarin).

The ‘terminal ductule' of the terminal ductal–lobular unit has two components (Fig. 4): one lies outside the lobule and is known as the extralobular terminal ductule; the other lies within the lobule (intralobular terminal ductule). The terminal ramifications of the intralobular terminal ductule form the secretory units of the breast (terminal ductules or acini), with morphological similarity to the terminal units of respiratory alveoli in the lung.

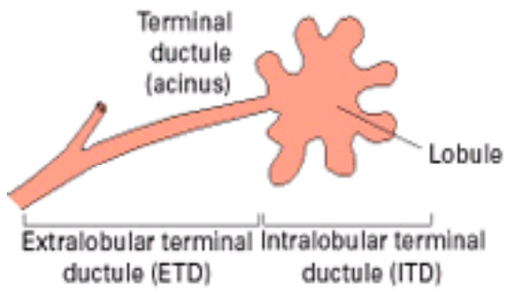


Fig. 4. The structure of the terminal ductal lobular unit (TDLU).

The whole subject is complex because of difficulties with nomenclature. The term ‘terminal duct' has been used for both the smallest epithelial unit in the lobule and the largest duct that opens on to the nipple surface. The same term has been used interchangeably with ‘tertiary duct', ‘ductule', and ‘terminal ductule'. The problem is compounded by some authorities wishing to retain the term ‘acinous' only for those terminal units during breast-feeding.

Histologically, that part of the duct system adjacent to the skin of the nipple is lined with stratified squamous epithelium. However, a sudden change to a double layer of columnar or cuboidal epithelium characterizes the remainder of the duct system. The terminal ductules (acini) can undergo secretory changes and therefore have a role of both transport and secretion.

Between the epithelial cell layers and the basement membrane is a network of myoepithelial cells; these respond to oxytocin and are responsible for milk ejection during lactation.

Physiological changes in the breast

Changes during the menstrual cycle

Despite widespread anecdotal belief to the contrary, there is little evidence of substantial histological change in the breast throughout the different phases of the menstrual cycle. Retention of fluid may occur during the luteal phase, but this does not appear to be associated with morphological change.

Changes during pregnancy and lactation

The main histological change that occurs in the breast during pregnancy is lobular–alveolar growth and the development of new secretory units; this gives rise to the characteristic microscopic description of ‘adenosis of pregnancy' (Fig. 5). It is characterized histologically by alveolar dilatation and conversion of the resting two-layer epithelium to a monolayer that demonstrates secretory changes. Colostrum formation, capillary growth, vascular engorgement, and myoepithelial cell hypertrophy are also apparent as pregnancy progresses. There is a doubling of breast weight from about 200 to 400 g, much of which is due to fluid retention. These changes occur under the influence of increased concentrations of luteal and placental sex hormones, placental lactogens, and chorionic gonadotrophins.

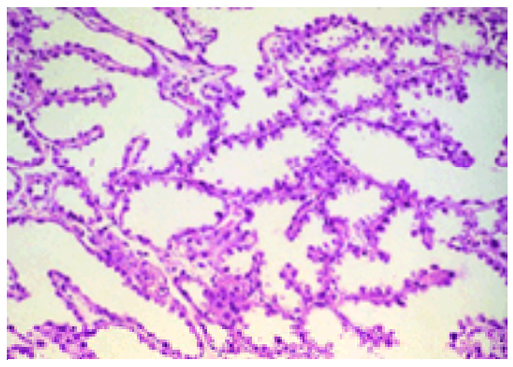


Fig. 5. Adenosis of pregnancy: there is alveolar dilation and additional secretory units (by courtesy of Dr D. Tarin).

Although some colostrum is formed within the breast during pregnancy, true milk is not produced until about 2 days after parturition; this occurs due to the high post-partum concentrations of prolactin, which are maintained despite falling concentrations of other sex hormones that normally inhibit lactogenesis. During lactation the alveoli are distended with milk, the cells become cuboidal, and there is a resultant diminution of the intralobular space.

On cessation of breast-feeding there is gradual return to the non-pregnant state; this may take as little as 3 months, but in some individuals may take much longer. This process, known as postlactational involution, is characterized histologically by lymphocytic infiltration and hyalinization of the lobules (Fig. 6). There is, however, little or no reduction in the number of ducts and lobules present.

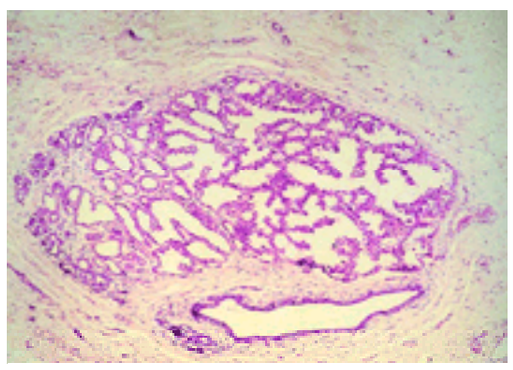


Fig. 6. Postlactational change: in one area of the lobule is adenosis of pregnancy; elsewhere there is a decrease in the number of acini, an increased amount of fibrous tissue, and lymphocytic infiltration (by courtesy of Dr D. Tarin).

Changes at the menopause

Involutional changes occur from about the age of 35 years, with regression of glandular tissue and its replacement by fat and fibrosis. Before the age of 50 years this process is characterized by loss of some lobular tissue; in older women, its progression results in the almost complete replacement of lobular tissue by collagen and fat. The end result is that, although major duct systems are visible, few lobules are seen. This is in contrast to postlactational involution, which is characterized by minimal loss of lobular units.

Changes at the menopause have two important clinical implications. Firstly, fat infiltration of the breast produces the low-density appearance of the parenchyma seen

on mammography, and thus makes this technique more successful in older woman. Secondly, aberrations of this involutional change may explain some of the benign disorders that occur in this age group.

Breast milk

It is not the purpose of the present discussion to elaborate on the advantages and disadvantages of breast-feeding. However, breast milk from a healthy mother allows normal development of an infant for 6 months or longer without need for any supplementation. In many underdeveloped countries, breast-feeding provides a major source of nutrition for the infant for up to 2 years. Breast-feeding will continue despite a poor nutritional state of the mother.

Breast milk has a relatively low protein content; that present is in the form of casein and lactoglobulin. Carbohydrate is present as lactose, and fat as triglyceride. Human breast milk is rich in vitamins C and D, and is a source of IgA. It is therefore important in helping provide an immune mechanism for the newborn.

A number of drugs are transmitted in breast milk; this should always be taken into account when prescribing for breast-feeding women. Of particular importance are alcohol, barbiturates, anticoagulants, tetracycline, and metronidazole.

Benign breast disease

Until recently, benign disorders of the breast were regarded as relatively unimportant; far more attention was focused on breast cancer. As a result, many patients with benign breast disease received rather scant attention from clinicians. There has been relatively little academic investigation into this complex subject. Benign breast disease has also suffered from the major disadvantage of a hopelessly confusing terminology, inadequate classification, and poor correlation between clinical, radiological, and pathological features.

During the past decade there has been increasing interest in benign breast disease for a number of reasons. Firstly, patients demand investigation and treatment for symptoms of benign disease, which, in turn, has increased the number of women referred to specialist breast disease units; these have participated in scientific studies on the classification and treatment of their condition. Secondly, there is the question of premalignant disorders and histological features that may imply an increased risk of breast cancer. Increasing understanding of these conditions may prove important in understanding the pathogenesis of breast cancer and in defining high-risk groups in whom regular surveillance may be beneficial. Finally, the recently introduced breast screening programmes are likely to present pathologists and clinicians with as yet ill-defined histological entities that may be of importance in understanding the development of invasive cancer and its eventual treatment.

Classification

There is no completely satisfactory classification of benign breast disease. Previous attempts have been based on a number of different factors such as clinical symptoms, patient age, histological features, or that part of the secretory system in which the abnormality has arisen. They all have inherent disadvantages. Firstly, there is poor correlation between clinical, pathological, and radiological features in any particular case. Secondly, benign breast disorders encompass a wide spectrum of clinicopathological features, ranging from near normality to severe disease. Finally, the breast must be regarded as a physiologically dynamic structure in which cyclical variations are superimposed on changes of development and involution throughout the woman's life. These physiological changes may themselves be so extensive that they fall outside what is regarded as the normal spectrum. The histological features of an individual abnormality must therefore be evaluated within the context of a wide range of normality.

It has therefore been suggested that the broad concept of benign breast disease should be reconsidered. Many so-called diseases of the breast might now be regarded more accurately as disorders that are based on aberrations of the processes of development, cyclical change, and involution (aberrations of normal development and involution; **ANDI**). This does not mean that benign breast disease does not occur, but that the term should be reserved for disorders of such severity that they are frankly abnormal. This concept is, of course, rather ill defined and will depend on interpretation and perception by both patient and clinician.

Aberrations of normal development and involution can account for many, if not all, benign breast disorders. A simplified system based on the various stages of physiological change (development, cyclical change, pregnancy, lactation, and involution) is shown in Table 1. It follows that most conditions listed under ‘benign breast disorders' can be regarded as minor aberrations of normal development or involution. Many patients with these conditions require reassurance rather than specific treatment, and explanation that they do not have a disease process.

Physiological state of the breast	Normal	Benign disorder	Benign disease
Development	Duct development	Nipple inversion	Mammary fistula
	Lobular development	Fibroadenoma	Giant fibroadenoma
	Stromal development	Adolescent hypertrophy	
Cyclical change	Hormonal activity	Mastalgia and nodularity	
	Epithelial activity	Benign papilloma	
Pregnancy and lactation	Epithelial hyperplasia	Blood-stained discharge	
	Lactation	Galactocele	
Involution	Ductal involution	Duct ectasia, nipple retraction	Periductal mastitis with suppuration
	Lobular involution	Cyst, sclerosing adenosis	
	Involutional epithelial hyperplasia	Hyperplasia and microcalcification	Lobular and ductal hyperplasia with atypia

Severe forms of the benign disorders may indicate a disease.

Table 1 Classification of the pathogenesis of non-malignant breast disease based on the concept of aberration of normal development and involution

This approach is well demonstrated in patients with cyclical mastalgia and nodularity. The vast majority of premenopausal women experience a degree of breast discomfort and increasing nodularity before menstruation. For most, this amounts to little more than an inconvenience and is regarded as a normal physiological process. About 2 or 3 per cent of women are referred to a clinic with cyclical mastalgia, the symptoms of which are more severe, with distressing discomfort lasting a week or more. Despite exhaustive investigation there is little evidence of any specific histological or endocrinological abnormality in these women with more severe cyclical breast pain, and if abnormal microscopic features are observed they correlate poorly with clinical features. Thus, these women with severe cyclical mastalgia are regarded not as having a disease process but as representing an extreme, or aberration, of the normal developmental and involutional processes that occur with each menstrual cycle. Only in exceptional cases may the term ‘disease' be cautiously adopted for patients with breast pain. It follows that cyclical mastalgia and nodularity, like other benign breast conditions, must be regarded in its broadest terms.

Symptoms

Despite the complexity of its classification there are relatively few presenting symptoms of benign breast disease. Symptoms fall into three main groups: breast pain, lumps, and disorders of the nipple and periareolar region. Infection of the breast causes further symptoms. As has been previously stated, attempts to correlate pathological and radiological findings with clinical features are a cause of much confusion and should be discouraged. A working clinical classification of benign breast disease is shown in Table 2, although it must be stressed that this does not imply whether the process is physiological or pathological, and in no way does it attempt to correlate symptoms with pathological features.

Breast pain (mastalgia)
Cyclical
Primary non-cyclical
Musculoskeletal
Sclerosing adenosis
Postoperative
Chest wall pain
Breast lumps
Fibroadenoma
Cyclical nodularity
Cysts
Galactocele
Sclerosing adenosis
Fat necrosis
Lipoma
Chronic abscess
Normal structures (prominent rib, edge of previous breast biopsy, margin of breast tissue, etc.)
Disorders of the nipple and periareolar region
Discharge
Retraction
Sepsis
Breast infection
Lactational
Non-lactational

Table 2 Symptoms of benign breast disease

Finally, it must be remembered that the reason for many referrals is anxiety on the part of either the patient or her doctor, which accounts for the increasing number of patients seen with non-breast conditions, including skin rashes and chest-wall pain. An important function of any breast clinic is not only to treat symptoms but also to allay the patient's fears of breast cancer.

Breast pain (mastalgia)

Pain is the most common reason for referral to a breast clinic and accounts for up to 50 per cent of patients seen. It is, however, the least understood of all breast symptoms, and the one whose management causes the most controversy.

Reports on breast pain are often anecdotal, uncontrolled, and of poor quality. Nomenclature poses a major problem: a number of different terms, such as mastitis, mastodynia, and mazoplasia, have been used to describe breast pain. Mastalgia has also been correlated with specific histological criteria, resulting in its description as ‘fibrocystic disease', although this has lost favour for reasons described above.

It must be stressed that mastalgia is a symptom and does not imply any specific pathological process, any more than does pain in other sites of the body.

Classifications

Attempts to classify breast pain are surprisingly recent. There are two distinct group of patients with these symptoms (Table 2): one group has symptoms that bear a definite relation to the menstrual cycle (cyclical mastalgia); in the remainder there is no such correlation (non-cyclical mastalgia). Non-cyclical mastalgia has recently been reclassified to distinguish pain in the breast from that originating in surrounding tissues such as the chest wall.

Cyclical mastalgia

This is the most common type of breast pain, accounting for 40 per cent of all cases referred to a breast clinic. There is an important correlation with the menstrual cycle, with discomfort lasting for a varying period of time before menstruation. Because of this cyclical relation, mastalgia is generally a condition of premenopausal women, with a median presenting age of about 35 years. Characteristically, the features of cyclical mastalgia wax and wane. Episodes of discomfort may last for some months; there may then be years of freedom until symptoms begin again.

The pain of cyclical mastalgia is frequently, but not always, bilateral and is usually located in the upper, outer quadrants. It is poorly localized and may radiate across the chest wall into the axilla or down the inside of the arm. The breasts are frequently described as being ‘heavy' as if pregnant, and many patients describe marked nodularity at the time of the discomfort.

There is a wide spectrum of symptoms in cyclical mastalgia. The majority of patients have only mild discomfort lasting 2 or 3 days before menstruation and are not unduly concerned. Such individuals are therefore best classified as having a breast ‘disorder' (ANDI) rather than a disease. The small minority who have severe symptoms lasting throughout the cycle with relief only during menstruation are those to whom the term ‘disease' may be applied.

There are no mammographic or pathological characteristics of cyclical mastalgia: indeed this lack of correlation between clinical, radiological, and histological findings is one of the major characteristics of the condition. The mammograms shown in Fig. 7 are from patients with severe breast pain. In one there is almost complete replacement with fat, giving a translucent appearance, whereas the other is dense and nodular (Wolfe DY pattern).

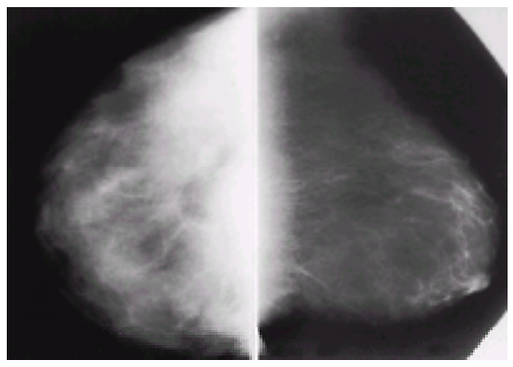


Fig. 7. Mammograms in patients with breast pain: in one the breast is radiologically translucent whereas the other is dense and nodular, the Wolfe DY pattern (by courtesy of Dr B. Shepstone).

Aetiology

The fact that symptoms of cyclical mastalgia correlate with the menstrual cycle implies a hormonal background. Early investigations suggested that hormonal imbalance was the cause, the fundamental abnormality being relative hyperoestrogenism due to either increased oestrogen secretion or deficient progestogen production. However, the vast majority of studies have failed to demonstrate either abnormality in women with mastalgia.

More recently, abnormal prolactin secretion has been incriminated as an aetiological factor in cyclical mastalgia. Although both random and basal concentrations of prolactin are normal in women with cyclical mastalgia, there is some evidence of impaired hypothalamic control of the release of this hormone in patients with severe symptoms. It should be appreciated, however, that the control of prolactin release is extremely complicated and that our current knowledge of its physiology is rather rudimentary.

The belief that cyclical mastalgia has a hormonal basis resulted in the suggestion that there would be an associated effect on fluid retention. However, despite a widespread belief that breast pain is due to water retention, this has never been scientifically confirmed.

Other aetiological factors, including excessive caffeine ingestion or inadequate essential fatty acid intake, have been suggested. The latter is of particular interest as there is good evidence that unsaturated fatty acid supplements can reduce the symptoms of cyclical mastalgia. The actual mechanism is unclear, but it may relate to correcting the amounts of prostaglandin E_1 in epithelial cell membranes and the subsequent effect of prolactin on breast sensitivity.

Psychoneurosis has been widely incriminated as an important factor. However, there is no evidence of excessive anxiety, depression, or phobia in these patients when they are evaluated against appropriate control groups.

Treatment

More than 80 per cent of women attending a breast clinic with cyclical mastalgia require no treatment other than simple reassurance, particularly that such symptoms do not imply any form of neoplastic process. Fear of cancer drives many women to seek specific advice about breast pain, and the importance of such reassurance cannot be overemphasized.

About 5 to 10 per cent of patients with cyclical mastalgia experience pain despite all reassurance. For them, specific drug therapy may be considered. There is a sound theoretical basis for the use of most, but not all, agents that have been tried, despite the fact that no constant physiological or pathological changes have been

identified in this condition. Furthermore, no drug satisfies the criteria of being universally effective, free from side-effects, and freely available for use in patients suffering from benign breast disease. A large number of studies evaluating the efficacy of these drugs have been performed. However, because of the placebo effect of such treatment the results of many studies are inherently flawed, and reliance can only be placed on prospective, randomized, placebo-controlled trials. A further problem of many studies is that they do not take into account the natural history of mastalgia. As a result a false impression of benefit may occur merely from natural remission, such as occurs in pregnancy or at the menopause.

Table 3 shows some of the agents widely used for the treatment of cyclical mastalgia, their possible modes of action, and common side-effects. Their overall efficacy is shown in Table 4.

Class of agent	Mode of action	Drug and dosage	Incidence of side-effects (%)	Most common side-effects
Diuretic	Reduction of body water	Dextrothiazide, daily	Rare	Metabolic disorders, gout
Progestogen	Correction of luteal insufficiency	Medroxyprogesterone acetate, 20 mg daily	20	Premenstrual symptoms, weight gain
Antioestrogen	Correction of hyperoestrogenism	Tamoxifen, 10–20 mg daily	10	Hot flushes, weight gain, menorrhoea, nausea, interference with anticoagulants
Dopamine agonist	Correction of hyperprolactinaemia	Bromocriptine, 2.5 mg twice daily	20	Nausea, dizziness, headaches, postural hypotension
Antigonadotrophin	Suppression of FSH and LH	Danazol, 200–400 mg daily	25	Weight gain, acne, amenorrhoea, hirsutism, voice change
Unsaturated fatty acid	Correction of deficiency	Evening primrose oil, 6 caps. daily	1	Nausea

FSH, follicle-stimulating hormone; LH, luteinizing hormone.

Table 3 Drugs used for treatment of cyclical mastalgia

	Response rate (%)
Tamoxifen*	90
Danazol	70
Bromocriptine	50
Evening primrose oil	45

* One study only.
Pyridoxine, diuretics and progestogens show no benefit compared to placebo.

Table 4 Efficacy of drugs used for cyclical mastalgia

Unproved therapies Diuretics have been widely prescribed, although there is little rational basis for their use, and it is widely believed that much of their efficacy is due to a placebo effect. Similarly, there is no rationale for using antibiotics. The concept of relative hyperoestrogenism as a result of luteal deficiency has stimulated the evaluation of progestogens in cyclical mastalgia; the results have been generally disappointing in placebo-controlled trials. Other widely adopted treatments of cyclical mastalgia, such as vitamins A or B_6, have failed to show any effect on cyclical mastalgia. Oral contraceptives may reduce the symptoms of cyclical breast pain, but total success can by no means be guaranteed. A possible metabolic effect produced by xanthine administration in the form of caffeine has resulted in a vogue for recommending decaffeinated drinks for patients with breast pain, this approach has not been validated in clinical studies.

Tamoxifen Relative hyperoestrogenism can also be treated with the antioestrogen drug tamoxifen, which was shown to be of benefit in randomized trials but as yet only has a licence for use in malignant disease. Pain relief is provided by 10 mg tamoxifen daily; there seems little benefit in increasing the dose. Side-effects such as weight gain and hot flushes are troublesome, and, as tamoxifen is contraindicated in pregnancy, appropriate contraception is necessary. Despite these potential problems, tamoxifen is increasingly used for patients with mastalgia.

Danazol Danazol probably acts as an antigonadotrophin by its effect on the pituitary–ovarian axis. At a dose of 200 to 400 mg daily it depresses production of follicle-stimulating and luteinizing hormones, and ovarian function. It significantly reduces breast pain, but is associated with side-effects in 20 per cent of patients: these include weight gain, acne, amenorrhoea, masculinization with hirsutism, voice changes, and reduction of breast size. Adequate non-hormonal contraception is necessary.

Bromocriptine The suggestion of abnormal prolactin concentrations in patients with cyclical mastalgia, and the possibility that prolactin stimulates glandular breast tissue, led to hopes that the specific prolactin-lowering agent bromocriptine would be useful in the treatment of cyclical mastalgia. This, a dopamine agonist, significantly reduced symptoms of cyclical mastalgia in benign breast disease. However, as with danazol, its use has been curtailed by side-effects, such as nausea, postural hypotension, vomiting, and dizziness, which occur in up to 20 per cent of patients.

Evening primrose oil Although it is regarded as homeopathic by many patients and their doctors, there is, in fact, a pharmacological basis for the use of evening primrose oil. As discussed above, prostaglandin E_1 appears to influence the sensitivity of breast epithelium to hormones such as prolactin; it is synthesized via the following pathway:

Linoleic acid
↓ => Rate-limiting step; inhibited by saturated fat and, perhaps, by stress and caffeine
Gamolenic acid
↓
Dihydrogamolenic acid
↓
Prostaglandin E_1

Interest in the relation between fat intake and breast cancer risk resulted in studies evaluating the effect of dietary fat in benign disease. Women with breast pain have significantly higher concentrations of circulating saturated fatty acids than controls. Furthermore, the saturated to unsaturated fatty acid ratio is inverted in favour of saturated fat in patients with breast pain. According to the above pathway, this results in inhibition of prostaglandin E_1 synthesis. The addition of unsaturated fatty acid to the diet in the form of gamolenic acid from evening primrose oil reverses the imbalance of saturated to unsaturated fat and is associated, in randomized controlled trials, with a significant reduction in breast pain. Side-effects are uncommon. A daily therapeutic dose of 320 mg of gamolenic acid is required; many non-prescription preparations of evening primrose oil fail to provide this dose.

Summmary of treatment Patients with cyclical mastalgia should be treated initially with evening primrose oil, followed by danazol for those refractory to treatment. Bromocriptine is a third choice, with activity similar to that of evening primrose oil but a significant incidence of side-effects. The response rates tend to be lower for second and third lines of treatment. Tamoxifen has the drawback that it is not strictly registered for use in benign disease, but is increasingly used as second-line therapy after evening primrose oil.

Responses to treatment are relatively short lived, usually of the order of 6 months. When using danazol, bromocriptine or tamoxifen it is our policy to treat for 3 months and then to see whether relapse occurs on cessation of therapy. Any relapse is an indication for restarting treatment, perhaps at a lower dose than originally used, or for a change in therapy if the initial response has been poor. Satisfactory treatment is particularly difficult in young women, in whom mastalgia is often resistant to treatment, whose potential for breast pain may span several decades, and whose fertility must be considered. Bromocriptine and danazol are potentially teratogenic;

they also require the use of a barrier contraceptive because they interfere with oral contraceptives. Many younger women also dislike the amenorrhoea induced by tamoxifen and danazol.

Occasionally, women with a long history of mastalgia unresponsive to all medical treatment demand consideration of mastectomy to release them from their discomfort. Although occasional patients may benefit from subcutaneous mastectomy, the general impression is that it should be avoided if at all possible. Case selection is all important if considering surgery.

Non-cyclical mastalgia

Non-cyclical mastalgia (Table 2) is even less well defined than its cyclical counterpart. It occurs in both pre- and postmenopausal women, with a median age of presentation of 45 years. As well as having no close relation to the menstrual cycle, non-cyclical mastalgia tends to be more chronic, unilateral, and located in the medial quadrants of the breast or the periareolar regions. It is not associated with lumpiness to the same degree as cyclical mastalgia, and the pain is frequently described as ‘burning' or ‘dragging' rather than a ‘heavy feeling'. The mastalgia is sometimes well localized; the term ‘trigger spot zone' has been used in these individuals.

Attempts to classify non-cyclical mastalgia have been compromised by the dubious inclusion of other conditions that may cause breast pain, the best example of which is duct ectasia/periductal mastitis. The inclusion of this condition was based on the mammographic appearances of many patients with non-cyclical mastalgia, which showed widespread, coarse calcification throughout he substance of the breast, also a feature of duct ectasia/periductal mastitis. However, the use of this term for patients with non-cyclical mastalgia has fallen into disfavour because of the principle of not mixing symptomatology with pathology, and because of lack of evidence that the pathological changes of duct ectasia correlate with breast pain.

Up to 50 per cent of patients with non-cyclical mastalgia have pain that arises not from the breast but from surrounding musculoskeletal structures (Table 2). A careful history and examination will identify these patients, who, unlike those with true non-cyclical mastalgia, can be provided with relatively simple and effective therapy.

Aetiology

The aetiology is unclear, but some of the factors associated with cyclical mastalgia may also be associated with non-cyclical breast pain.

Treatment

The management of true non-cyclical mastalgia is unsatisfactory. Many principles in the treatment of cyclical mastalgia may be applied to non-cyclical breast pain, but overall response rates to various drug therapies are only about 50 per cent of those observed in patients receiving treatment for cyclical mastalgia. Both bromocriptine and evening primrose oil have shown particularly disappointing results. Response rates to the various drugs improve if patients with true non-cyclical mastalgia are differentiated from those with musculoskeletal pain.

Some success has been ascribed to surgical excision of ‘trigger spot zones', but this approach has not been widely adopted.

Other causes

Musculoskeletal pain This had previously been included as a cause of non-cyclical mastalgia, but was recently shown to be a separate entity. It is the most common cause of apparent pain in the breast originating from other sites. Musculoskeletal pain is often unilateral and localized along the lateral chest wall or the costochondral junction (Tietz syndrome). Injection of local anaesthetic or steroids into the affected area produces good relief of symptoms.

Sclerosing adenosis This may be found microscopically either as a single entity or in association with other abnormalities. It may be a minimal histological change or a macroscopically obvious condition. It is classified as an ANDI; microscopically it is characterized by proliferation of terminal duct lobules, myoepithelial cell proliferation, increased number of acini, and fibrous stromal change (Fig. 8). Multifocal and nodular types are described; either may be painful and have been documented as a cause of mastalgia.

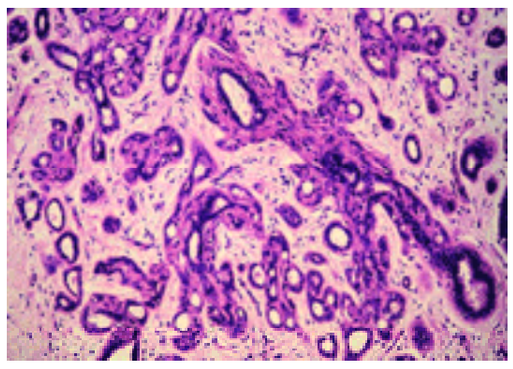


Fig. 8. Sclerosing adenosis: note proliferation of duct lobules, increased number of acini, and fibrosis (by courtesy of Dr D. Tarin).

The pain associated with sclerosing adenosis may be due to perineural invasion, and this may account for some patients with ‘trigger spot zones'. The main importance, however, is that macroscopically sclerosing adenosis may have a stellate appearance and may calcify: it may thus mimic carcinoma, both clinically and radiologically. The increased cellularity associated with sclerosing adenosis has been confused with carcinoma histologically, especially when examining frozen sections.

Previous surgery A small number of patients continue to complain of pain after biopsy for benign breast disease. There is no clear reason for this, although if the biopsy was performed in an area already subject to mastalgia it is likely that the original process will continue and be painful.

Cancer Cancer is an uncommon cause of breast pain, although women do occasionally complain of discomfort, which may be more pronounced before menstruation, at the site of a tumour. The author has seen one young, intelligent woman complaining solely of breast pain, while failing to notice that the breast had been largely replaced by tumour.

Referred cervical root pain This cause of pain in the breast should be considered in elderly patients in whom no other cause for mastalgia can be found.

Benign breast lumps

Approximately 40 per cent of all patients attending a breast clinic have a benign breast lump (Table 2). In the past there was a tendency to excise all lumps, and an excessive amount of unnecessary surgery was performed for benign disease. The main problem from the woman's point of view is fear that such a lump may be cancer. The clinician must therefore provide a high degree of diagnostic accuracy while at the same time ensuring that an excessive biopsy rate is prevented. It is now easier to exclude cancer with the development of diagnostic aids such as mammography, ultrasonography, and aspiration cytology.

The surgeon in the breast clinic has two important tasks when confronted with a patient with a breast lump: firstly, to decide whether the lump is truly an abnormality, that is, different in consistency from the surrounding breast, or whether it can be regarded as being within the spectrum of normality; secondly, if the lump is a true abnormality, to determine whether it is malignant.

The history is of particular importance. The lump's duration, pain, change in size, and relation to the menstrual cycle should be the subject of detailed enquiry. The presence of menstrual irregularity and a previous history of similar problems should be sought. It is also important to enquire into the patient's risk factors for breast cancer, such as her age, number of children, age at first pregnancy, family history, and other potential predisposing factors such as hormone replacement therapy or

oral contraceptive use.

Having established the risk or otherwise for breast cancer, the clinical impression must be confirmed by careful examination. Tethering of the skin, distortion of the breast, and nipple retraction are common features of malignancy, but they can also occur as a result of benign change. Mobility of the lump should be assessed: this is characteristic of a fibroadenoma and also quite obvious in cysts. Cancers tend to be more fixed, but are occasionally quite mobile.

Finally, the surface of the lump should be assessed. Fibroadenomas have a lumpy, bosselated surface whereas cysts are usually smooth and tense. Cancerous lumps are usually, but not always, hard. While a cancerous lump is likely to be hard, irregular, and fixed, it is not uncommon to see malignant tumours that are firm, quite regular, and have a degree of mobility.

If there is any doubt, a pathological diagnosis or biopsy is necessary; this should always be undertaken in any woman over the age of 25 years with a solid, discrete lump. In the majority of patients, aspiration cytology is sufficient, but if this is in any way unsatisfactory, or if any doubt remains, then biopsy is mandatory. It is a daring clinician who does not remove a discrete breast lump from a 40-year-old nulliparous woman with a family history of breast cancer even if fine-needle aspiration cytology is benign and mammography is reassuring.

Fibroadenomas and associated conditions

Fibroadenoma and related tumours are derived from the breast lobule and are characterized by proliferation of both connective tissue and epithelium. They encompass a wide spectrum of conditions, ranging from the totally benign simple fibroadenoma to locally invasive and, rarely, frankly malignant tumours. There has been great confusion over their pathogenesis, particularly at the more malignant end of the spectrum. Such are now commonly known as phyllodes tumours, but were previously described as cystosarcoma or phyllodes sarcoma. Those descriptions are inappropriate because they imply a totally mesenchymal stromal origin whereas all of these tumours, whether benign, locally invasive, or malignant, also have an epithelial component. The major features of fibroadenomas and associated conditions are summarized in Table 5.

	Usual age at presentation (years)	Size	Characteristics of fibrous stroma & elements	%	Cellular atypia and proliferation
Benign simple fibroadenoma	16–40	< 5 cm	Hypocellular		Absent
Giant fibroadenoma (including juvenile fibroadenoma)	15–20/65–75	> 5 cm	Mild epithelial hyperplasia		Not prominent
Phyllodes tumour	35–50	Variable	Hypercellular		Marked

Table 5 Features of fibroadenoma and associated conditions

Benign simple fibroadenoma

Fibroadenomas are benign tumours showing evidence of both connective tissue and epithelial proliferation. They originate from the breast lobule and can be regarded as an aberration of normal lobular development rather than a true benign tumour. Their origin explains why fibroadenomas are common in young women at the time of lobular development, and why they are occasionally seen in combination with lobular carcinoma. The aetiology of a fibroadenoma is unknown; hypersensitivity to oestrogen within a lobule has been suggested.

The most important pathological aspect of a fibroadenoma is its connective tissue stroma. In the past, great importance was attached to whether this stroma compressed adjacent ducts to form curved, slit-like structures (intracanalicular pattern) or whether it simply surrounded a duct circumferentially (pericanalicular pattern). The fibrous stroma of fibroadenoma has low cellularity and a regular cytology (Fig. 9). Occasionally there is histological evidence of fat, smooth muscle, squamous metaplasia, and infarction. The epithelial proliferation may be quite hyperplastic, but this is of no prognostic importance.

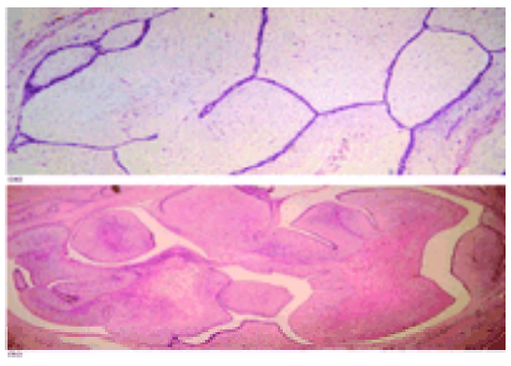


Fig. 9. (a) Fibroadenoma: note the low cellularity and regular cytology of the benign fibroadenoma; (b) phyllodes tumour: note the hypercellularity, atypia, and abundant mitoses (by courtesy of Dr D. Tarin).

If the fibrous stroma shows a marked increase in both cellularity and atypia, then the locally invasive and occasionally metastatic phyllodes tumour should be considered (see below). This entity can be regarded as an extreme end of the disease spectrum, the simple fibroadenoma representing the other end.

Clinical features

Most fibroadenomas present in young women 16 to 24 years old. However, with the use of pathological examination in the diagnosis of breast lumps in older women, the overall median age of presentation is nearer 30 years. Fibroadenomas decrease in incidence after the menopause, when they undergo involution. During this time they may calcify and thus become apparent on mammography. As a result, they are commonly identified on mammographic screening programmes.

Fibroadenomas are smooth or slightly lobulated structures, usually measuring about 2 or 3 cm in diameter. With the exception of those adjacent to the nipple, they are characteristically mobile. In young women the term ‘breast mouse' is thus aptly applied. With increasing age the degree of mobility lessens because of the restraining effects of surrounding, involuting fibrotic tissue. In the elderly woman they can still present as a small, hardened mass that is still quite mobile.

About 10 per cent of all patients have multiple fibroadenomas on presentation, and occasionally one sees young women in whom the whole breast seems to be replaced by fibroadenotic tissue. Others present with multiple recurrent fibroadenomas; this occurs particularly among black and oriental individuals.

Diagnosis

Up to the age of 25 years a clinical diagnosis will suffice, although diagnosis by aspiration cytology is preferable. Thereafter, pathological confirmation is required because of the need to exclude carcinoma, tubular cancers in particular being a source of clinical confusion. Fine-needle aspiration cytology provides an accurate method of diagnosis (Fig. 10), although hyperplastic epithelial cells can occasionally be mistaken for neoplasia.

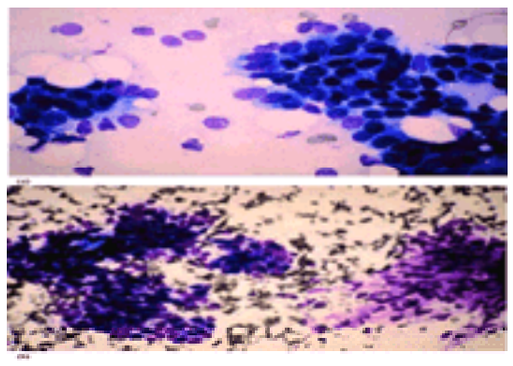


Fig. 10. (a) Fine-needle aspiration cytology of fibroadenoma: note the regular groups of benign epithelial cells with associated bare nuclei; (b) fine-needle aspiration cytology of phyllodes tumour: the aspirate shows gross cellularity, a hypercellular stroma, and atypia (by courtesy of Dr I. Buley).

As fibroadenomas usually present clinically in younger women, mammography has little place in routine diagnosis. In older patients, fibroadenomas appear mammographically as a solitary, smooth lesion with a density similar to or slightly greater than that of the surrounding tissue. With increasing age, stippled calcification becomes apparent (Fig. 11).

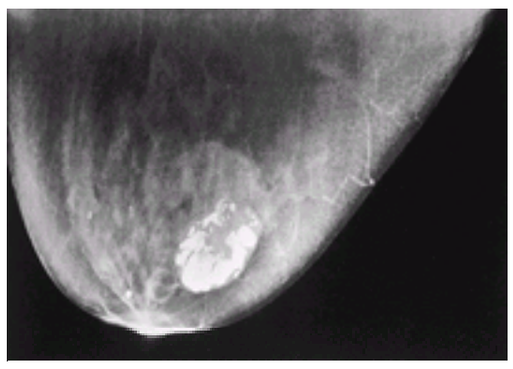


Fig. 11. Calcified fibroadenoma on mammography: note the smooth opacity with coarse calcification within its structure (by courtesy of Dr B. Shepstone).

Management

The practice of surgically removing all fibroadenomas has now been condemned because of greater understanding of the natural history of this condition. If left untreated, most fibroadenomas will slowly increase in size from 1 to 3 cm in diameter over a period up to 5 years. The active growth phase is about 6 to 12 months, during which time there is a doubling of size. Thereafter they remain static or may (in up to one-third of cases) gradually become smaller.

In women under the age of 25 years, routine removal is unnecessary. This conservative approach may be recommended for women under 35 years old, provided that cytological examination, repeated after 3 months, rules out malignancy. Such a policy may, however, result in a small number of cancers being missed and removal of fibroadenomas is recommended by some authorities after the age of 25 years. Excision can be done under either local or general anaesthetic.

Recurrence of a fibroadenoma after removal is uncommon. If there is recurrence, it may be due to a number of factors. Firstly, a truly metachronous fibroadenoma may develop. Secondly, the original tumour may have been incompletely removed or missed at operation; and, finally, it may be the mode of presentation of a previously undiagnosed phyllodes tumour.

Giant fibroadenoma

Giant fibroadenoma has a bimodal age of presentation at the extremes of reproductive life; those occurring in the younger age group have been described as juvenile fibroadenomas. Giant fibroadenomas thus occur particularly in the 14- to 18- and the 45- to 50-year age groups, and are characterized by rapid growth to a large size. They are, by definition, bigger than the common type of fibroadenomas, being at least 4 or 5 cm in diameter, and sometimes achieving a diameter of 10 cm or more (Fig. 12).



Fig. 12. A giant fibroadenoma 12 cm in size (by courtesy of Dr D. Tarin).

Histologically, giant fibroadenomas contain the typical hypocellular stromal and epithelial components showing variable, though usually mild, degrees of hyperplasia and atypia; mitoses are uncommon. These features are different from those of phyllodes tumours, which generally exhibit much more cellularity, pleomorphism, and mitotic activity, but there is some overlap of microscopic appearances between these two conditions (Table 5).

Clinical features

Giant fibroadenomas are more common in black and oriental individuals. Clinically, patients present with pain in the breast associated with a rapid increase in size. On examination the breast is enlarged and the nipple may be displaced. The overlying skin frequently has a characteristic shiny appearance, and dilated veins may be apparent. In extensive, neglected cases, skin necrosis can occur.

Occasionally, young women may present with unilateral breast enlargement and the fact that a mass is the cause is not appreciated. Giant fibroadenomas may be confused with virginal hypertrophy, although this bilateral and not associated with cutaneous or venous changes.

Treatment

Management is by enucleation through an appropriately cosmetic incision. While this treatment initially results in some discrepancy in breast size, the remaining breast

tissue expands to virtual normality within a few months. Wide excision or mastectomy are contraindicated.

Although some giant fibroadenomas can appear somewhat aggressive histologically and may even be confused with phyllodes tumour, their clinical behaviour is completely benign. There is no evidence that they recur locally with any frequency or metastasize.

Phyllodes tumour

Phyllodes tumours have been the cause of much misunderstanding and argument, partly related to their varied nomenclature. They have been described as phyllodes sarcoma, cystosarcoma, cystosarcoma phyllodes, and benign cystosarcoma. They have also been equated with giant fibroadenoma, but this is also misleading as it understates the malignant potential of the phyllodes tumour and implies a similar histological appearance. Conversely, terms such as cystosarcoma overstate the malignant potential and imply a false correlation with true mesodermally derived sarcomas. Phyllodes tumours show a wide spectrum of activity, varying from an almost benign condition to a locally aggressive, and sometimes metastatic, tumour.

Phyllodes tumours appear well circumscribed but are characterized by irregular surface projections that may be cut during surgical excision, predisposing to recurrence. The cut surface is soft, brown in colour, and may exhibit cysts, necrosis, or haemorrhage. Histologically, both epithelial and fibrous stromal elements are present, with the stroma showing hypercellularity, much atypia, and numerous mitoses (Fig. 9). Grading of phyllodes tumours is based on their mitotic index and degree of pleomorphism. Low-grade tumours have five or fewer mitoses per 10 high-power fields; high-grade phyllodes have a mitotic count of 10 or more. Grading may relate to liability to local recurrence.

Clinical features

Phyllodes tumours occur in premenopausal women. They are usually seen in the 30- to 50-year age group, but are not uncommon in women aged about 20 years. They have the features of a common fibroadenoma but can grow rapidly to a large size and may involve much of the breast. The overlying skin may become reddened and, in advanced cases, can become frankly ulcerated. However, a degree of mobility is retained, even by large tumours. Axillary lymphadenopathy is uncommon; if it occurs it indicates an extremely aggressive form of the disease.

Despite the tendency to grow rapidly to a large size it is not uncommon for phyllodes tumours to present as a much smaller mass that is clinically indistinguishable from a simple fibroadenoma.

Treatment

Phyllodes tumours occurring in young women under the age of 20 years are said to represent the benign end of the spectrum of this condition. As such, simple enucleation has been recommended, although the author prefers to excise the area more widely.

Older patients require wider excision with a 1-cm margin of normal breast tissue. Vary large tumours or those with aggressive histological features may merit even wider excision, with quadrantectomy or even simple mastectomy and reconstruction for the largest.

Even with an aggressive surgical policy of wide excision, approximately 25 per cent of phyllodes tumours recur over a 10-year period. Local recurrences should be widely excised. If recurrent tumours develop persistently, mastectomy with reconstruction should be considered. The tendency for recurrence is greater for large and histologically higher-grade tumours. A major worry with persistently recurrent phyllodes tumours is that they may metastasize, although this is a rare occurrence, being described in less than 5 per cent of patients. The most common site of metastasis is the lung.

Adjuvant radiotherapy and chemotherapy have no proven role in the management of phyllodes tumour.

Breast cysts

Breast cysts are among the more common reasons for referral to a breast clinic. They are frequently confused with more extensive cyclical nodularity. The description of cyclical nodularity as fibrocystic disease compounds the problem, and results in the false hope that many patients with cyclical nodularity can be treated by simple cyst aspiration.

True breast cysts are very common: up to 7 per cent of women develop a clinical cyst at some time during their lives. Autopsy studies show that a further 20 per cent of women have evidence of subclinical cysts in the breast, although many of these are only 2 or 3 mm in diameter.

As is the case with fibroadenomas, breast cysts can be regarded as an aberration of normal lobular physiology. The specific cause of this aberration is unknown, although there is some weak evidence that cyst formation may relate to hyperoestrogenism, such as may result from hormone replacement therapy. The pathogenesis of breast cysts is similarly unclear. Early workers suggested that they might simply be distended ducts or that they may result from cystic lobular involution. During this process, lobules develop microcysts, which eventually coalesce to become larger cysts; this change is potentiated by obstruction of lobular outflow and replacement of surrounding stroma by fibrous tissue.

More recent investigations have suggested that the aetiology of breast cysts is more complex than previously believed. There appear to be two distinct populations of macrocyst defined by their microscopic appearance, their biochemical profile, and clinical features (Table 6).

	Simple	*Apocrine*
Lining epithelium	Simple cuboidal	Apocrine
Na^+/K^+ ratio	> 3	‘3
pH	‘7.4	> 7.4
Tendency to recurrence	Low	Moderate
Association with cancer	Not proven	Possible

Table 6 Comparison of two types of breast cyst

Aspirated fluid from simple, uncomplicated cysts has a high Na^+:K^+ ratio (greater than 3), similar to that found in plasma. The pH of this fluid is less than 7.4 and it is likely that the flat, rather featureless epithelium of these cysts acts as a simple membrane through which interstitial fluid passively diffuses. Simple cysts tend to be single, not recurrent, and are unlikely to be associated with an increased risk of cancer.

The second type of cyst is lined by apocrine epithelium, characterized by large, columnar cells resembling those found in apocrine sweat glands. The Na^+:K^+ ratio is less than 3, and similar to that of interstitial fluid. The pH of apocrine cysts is higher than that of simple cysts and their lining membrane secretes substances such as androgen conjugates. These observations suggest that apocrine epithelium actively secretes potassium into the cyst fluid. Apocrine cysts tend to recur, as the balance between secretion of fluid and its reabsorption is in favour of reaccumulation. They may also be associated with an increased risk of breast cancer, although the overall evidence is weak on this issue.

Other studies have shown that in the early stages of cyst development, microcysts are of the apocrine secretory type. It is only when macrocysts develop that differentiation into simple cysts occurs.

Clinical features

Cysts are classically seen in perimenopausal women between the ages of 45 and 52 years, although they frequently occur outside this age range, especially in individuals receiving hormone replacement therapy. They are usually single at presentation, but it is not uncommon to see multiple cysts in a breast. In the most extreme examples the whole breast seems to be composed of a number of cysts.

A characteristic of breast cysts is that they suddenly appear, even if they are quite large. The reason for this is that the cyst exists in a flaccid, subclinical state before its presentation as a lump. The accumulation of a relatively small amount of fluid causes a disproportionate effect on intracystic pressure, according to La Place's equation *(P = 2T/r)*.

Cysts may be uncomfortable and are often frankly painful. There may be a vague association between the discomfort and the menstrual cycle, with increasing pain before menstruation, although this is not a pronounced feature.

Cysts are frequently visible. However, their most characteristic feature is their smooth, tense nature on palpation. They have a degree of mobility but this is not as pronounced as that of fibroadenomas. The classic clinical appearances may be masked if the cyst is situated deep in the breast. Normal nodular breast tissue overlying the cyst may hide its classic smoothness on palpation.

The diagnosis of a simple breast cyst is straightforward, as aspiration confirms the diagnosis. The amount of fluid aspirated is variable: it is usually about 6 or 8 ml, although occasionally cysts containing 60 or 80 ml of fluid are encountered. Cyst fluid varies in colour, ranging from pale yellow to almost black; sometimes the aspirate appears translucent whereas on other occasions it is thick and turgid.

Mammography and ultrasonography may aid diagnosis (Fig. 13) but these investigations are not essential in symptomatic patients. Little important information is gained from cytological examination of cyst fluid unless it is blood-stained.

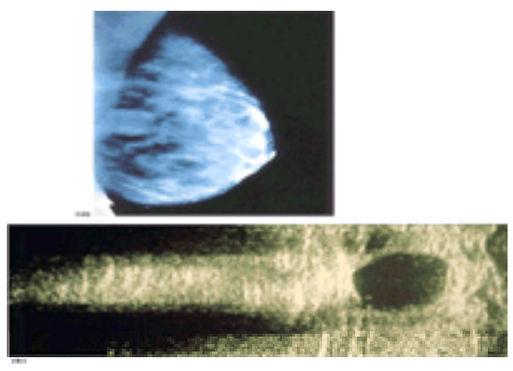


Fig. 13. (a) Breast cyst on mammography: note the well-delineated, homogeneous opacity; (b) breast cyst on ultrasonogram: note the well-delineated- homogeneous opacity (by courtesy of Dr B. Shepstone).

Treatment

Breast cysts were previously treated by surgical excision, but such treatment is no longer recommended as simple aspiration will normally suffice. After aspiration the cyst remains as a lax, impalpable structure that may still be seen on mammography. However, there must be no clinical evidence of a mass remaining after aspiration. If a mass does remain, further investigation with fine-needle aspiration cytology or biopsy is indicated.

There are two main indications for surgical excision of the cyst. If the aspirate is bloodstained (as long as this is not due to direct trauma from the needle), the rare intracystic carcinoma may be present. The second indication, which is perhaps more contentious, is cyst recurrence. This may be due simply to inadequate aspiration and further such treatment may be attempted before excision. However, if the cyst recurs rapidly and more than twice, its excision is recommended.

Patients who develop persistently recurrent cysts throughout their breasts can present a difficult management problem. Recurrence is often at a different site from the presenting cyst. Up to 15 per cent of patients develop further cysts over 5 to 10 years, although the majority will only have one or two recurrences. However, a small number of women develop recurrences regularly and may attend the breast clinic every 2 or 3 months for cysts to be drained. In the past some of these patients were treated by subcutaneous mastectomy. We now recommend danazol or tamoxifen, although evidence in favour of this treatment is sparse and there are side-effects and limitations associated with the use of these drugs.

Although it does not aid diagnosis, mammography should be performed as a routine screening procedure on women over the age of 35 years presenting with cysts, but the yield of occult cancers is low. As there is some evidence of increased cancer risk, patients with recurrent apocrine cysts may benefit from continued mammographic surveillance. Patients with a solitary simple cyst certainly do not require regular mammographic monitoring.

Cyclical nodularity

Many patients are referred to breast clinics with a lump that is really a manifestation of cyclical nodularity, but in a more localized and clinically discrete form. It is the mass, rather than the pain, that is the predominant feature. A careful history will frequently reveal that the lump has been present for some time and that its size varies with the menstrual cycle. Many patients will also admit to discomfort or pain in the lump when it is most prominent.

A variation of this presentation is seen, particularly in teenagers and occasionally in older women approaching the menopause. A large, diffuse, and frequently uncomfortable swelling develops suddenly, often, but not always, in the upper outer quadrant. Examination discloses a diffuse, nodular swelling that may be somewhat tender. These changes usually resolve with the next menstrual cycle. Aspiration cytology with mammography are indicated in older women, and in the young if there is any diagnostic doubt.

As long as malignancy is excluded, patients with cyclical nodularity presenting as a breast lump can simply be reassured.

Galactocele

Galactocele is well documented in older texts on breast disease, but is perhaps less common than previously thought. Classically the galactocele presents as a cyst in a woman who has recently stopped breast-feeding, although they occasionally occur during lactation. Aspiration shows breast milk and is usually successful in resolving the problem.

The pathogenesis of galactocele is unclear: it may simply be a pre-existing simple cyst that fills with milk.

Sclerosing adenosis

Sclerosing adenosis is an uncommon cause of a breast lump. Clinically, it usually presents as a smooth, relatively mobile mass in the 30- to 50-year age group. It is frequently painful and can occasionally be a cause of mastalgia rather than of a mass.

Sclerosing adenosis can be regarded as an ANDI, characterized by lobular enlargement and distortion associated with fibrous stromal change.

Fat necrosis

Fat necrosis is another condition that has attracted much attention, as it has previously been a cause of diagnostic difficulty in women with breast disease. In particular, delayed diagnosis of cancer in patients with a history of trauma, and the occasional mastectomy performed in patients with fat necrosis thought to be a tumour, have been recorded.

A history of trauma is readily provided by the patient but this is a trap for the unwary, as many patients with cancer will try to attribute the lump to a previous injury.

By the time the patient presents at the clinic, any external evidence of injury has frequently disappeared. The lump can be small and hard, and clinically may easily be confused with a carcinoma. Sometimes there is associated inflammation, tethering, and oedema.

Fine-needle aspiration cytology is usually diagnostic (Fig. 14). Areas of fat necrosis are occasionally cystic, and aspiration can partially resolve the condition. However, in our unit we review patients diagnosed as having fat necrosis after 6 weeks, when the lump is excised if it has not disappeared. Mammography must be interpreted with care as fat necrosis can have similar radiological features to those of cancer (Fig. 15).

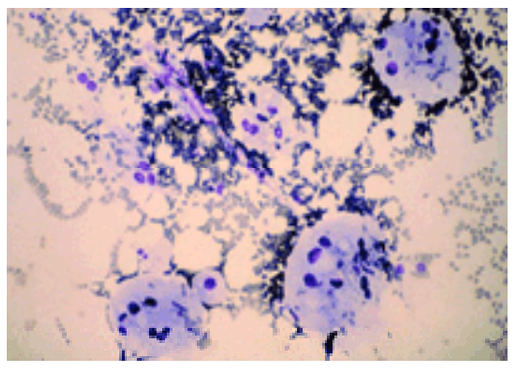


Fig. 14. Fat necrosis of the breast: fine-needle aspiration showing characteristic foamy, fat-laden macrophages with few epithelial cells (by courtesy of Dr I. Buley).

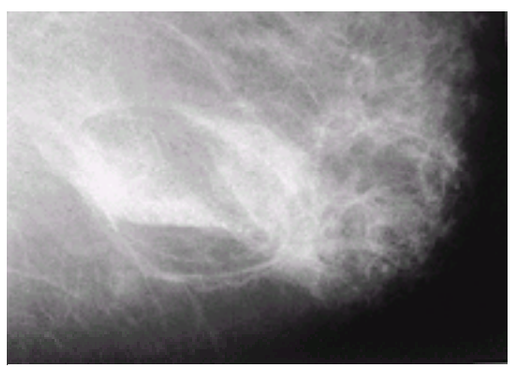


Fig. 15. Fat necrosis of the breast: mammogram showing a speculated opacity with calcification and features similar to a carcinoma (by courtesy of Dr B. Shepstone).

Lipoma, adenolipoma

Lipomas occur quite frequently in the breast, but not to the extent that might have been thought in view of the amount of fat that is present. They produce a soft mass, which is best excised if there is any doubt as to its nature. The adenolipoma is a variation of lipoma. It sometimes has a marked fibrotic component and is best regarded as a hamartoma.

Chronic abscess

The increasing use of antibiotics as a treatment for inflammatory breast conditions occasionally results in a chronic, sterile abscess. Treatment is by aspiration, or by open drainage with excision of the wall if recurrence develops.

Normal structures

Patients may present with the impression of a breast lump that may be due to either a normal structure, such as a rib, or a prominent area of breast tissue, which may be made more obvious by the defect from a previous biopsy or from weight loss. Fatty replacement at the time of breast involution can also give the impression of a lump: at operation, fibrotic fatty tissue is found.

Disorders of the nipple and periareolar tissue

Disorders of the nipple are no less controversial than other aspects of benign breast disease. There are the usual reasons for these difficulties, such as lack of consensus over terminology and poor correlation between clinical features, mammography, and pathological findings. The symptoms of disorders of the nipple and periareolar tissues are discharge, retraction, and the effects of periareolar sepsis. There are multiple causes for each of these symptoms but one condition, duct ectasia/periductal mastitis, is of paramount importance.

Duct ectasia/periductal mastitis

Duct ectasia and periductal mastitis has been described for more than a century but until recently it has been relatively ignored by both clinicians and pathologists. In the past it has been regarded as part of the spectrum of fibrocystic disease, and because of its histological appearance it has been given a variety of names such as plasma cell mastitis, mastitis obliterans, and granulomatous mastitis. The clinical features of duct ectasia/periductal mastitis are extremely varied and, as histological or cytological confirmation is not always possible, diagnosis must sometimes be made on clinical grounds alone.

Pathogenesis

In 1951, Haagensen suggested that the primary change in patients with duct ectasia/periductal mastitis was simple dilatation of the larger periareolar ducts (duct ectasia; Fig. 16). It is unusual for all ducts to become dilated but the changes are frequently bilateral. The dilated ducts fill with a stagnant, thick, green or creamy secretion (grumous). These stagnant secretions lead to loss of duct epithelial lining, and associated ulceration can cause further discharge, with bleeding from the nipple. There may also be a chronic inflammatory response (periductal mastitis) in the periareolar tissues because of leakage of the secretions through the damaged duct walls. Periductal mastitis may produce a painful mass or even a frank periareolar abscess; repeated inflammatory processes cause fibrosis and result in nipple retraction.

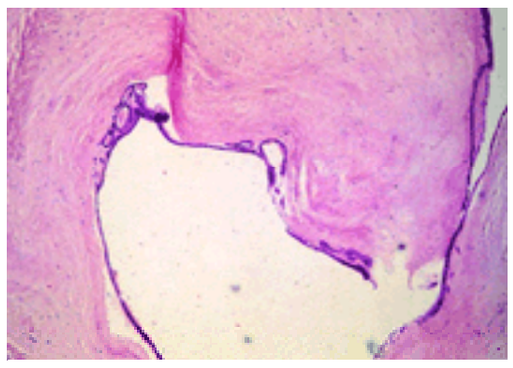


Fig. 16. Duct ectasia (by courtesy of Dr D. Tarin).

The various pathogenetic processes of duct ectasia explain many of its symptoms. A difficult question to answer is what causes the initial duct dilatation. Suggested possibilities include hormonally induced muscular relaxation of the duct wall, inadequate absorption of secretions, or obstruction of the system by squamous debris. Unfortunately, there is little scientific basis for these suggestions. An alternative theory is that the periareolar inflammation rather than the duct ectasia is the primary process. Periareolar inflammation may result in secondary duct dilatation and the consequent discharge, fibrosis, and nipple retraction. This alternative theory explains why younger women tend to suffer the inflammatory complications of duct ectasia, whereas nipple inversion and discharge occur in older women.

Duct ectasia/periductal mastitis is, therefore, a complex disorder of uncertain aetiology. Even smoking has been incriminated as an aetiological factor in younger women with periductal mastitis. The wide variety of clinical symptoms can best be explained and understood by appreciating that there is more than one process in its pathogenesis. Indeed, because inflammatory complications tend to occur in younger women whereas the effects of duct dilation are seen in the more elderly, some authorities believe that periductal mastitis and duct ectasia are two separate diseases.

The more severe symptoms of duct ectasia/periductal mastitis, such as abscess formation, can be regarded as a true benign breast disease. However, in all but its most severe form, duct ectasia is best classified as a breast benign disorder originating from the normal process of duct involution with fibrosis.

Nipple discharge

Nipple discharge is a common symptom presenting to breast clinics. The patient may fear cancer and the discharge may in itself be a cause of social embarrassment and annoyance. Treatment is therefore directed not only at diagnosing the cause of the discharge accurately but, if necessary, stopping the discharge itself.

Patients presenting with nipple discharge should be questioned about whether it has been bloodstained, whether it is unilateral or bilateral, and whether it is associated with a lump. If a lump is present, then diagnosis of the cause of the discharge becomes secondary to investigation of the lump.

Although inspection may reveal the source of discharge as one or more ducts, most patients have little visible abnormality. Excessive crusting may occasionally be seen on the nipple: this may simply be the dried products of secretion. Sometimes a skin disorder such as eczema may be found, the associated serous exudate producing the impression of nipple discharge.

Routine palpation of the breast must be followed by firm but gentle pressure around the areola. This will determine whether the discharge is unifocal or multifocal and whether it occurs in one or both breasts. It should also determine the nature of the discharge. Particular attention should be paid to whether the discharge has the grey–creamish, shiny characteristic of human milk (galactorrhoea in the non-pregnant or non-lactating patient), whether it is watery, serous or, as is often the case, of a thick, opalescent nature that may vary in colour from cream to almost black. Of paramount importance is whether the discharge is spontaneous or contains blood; this may be bright red and fresh or much darker in colour due to the presence of degradation products.

Investigations

Our own experience in investigating nipple discharge radiologically and with cytology has been disappointing, although others have reported success. All women aged 35 years or more require mammography. Particular features of note on the mammogram include the intraductal microcalcifications of carcinoma *in situ* and the dilated ducts associated with secretory granules that indicate duct ectasia (Fig. 17). Ductography may demonstrate the presence of duct papillomas, but this investigation tends to be painful, insensitive, and technically demanding.

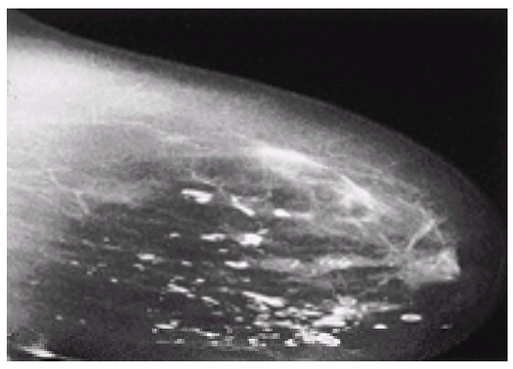


Fig. 17. Mammographic appearances of duct ectasia, showing coarse calcifications of secretory granules in dilated ducts (by courtesy of Dr B. Shepstone).

A number of centres recommend cytological examination of the discharge; this occasionally produces a positive finding, although false-negative results are common.

Causes

There are a number of causes of nipple discharge (Table 7). Unfortunately, it is difficult to correlate accurately the nature of the discharge with the cause, although certain features do act as a diagnostic guide. The presence of bright red blood in the discharge, a watery discharge, or spontaneous drainage from a single duct are all factors that indicate a potentially serious cause and require thorough investigation.

Nipple discharge
Physiological
Galactorrhoea
Duct ectasia
Papilloma and carcinoma
Cysts
Nipple inversion
Congenital
Nipple retraction
Duct ectasia
Cancer
Previous surgery

Table 7 Disorders of the nipple and periareolar region

Physiological Discharge of milk is a normal phenomenon during pregnancy and lactation. During pregnancy, blood-staining is occasionally observed: this is of no significance and reassurance is all that is required. A milky nipple discharge may occur transiently in the neonate due to transplacental passage of luteal and placental hormones from the mother's circulation.

Galactorrhoea Galactorrhoea is secretion of milk not related to pregnancy or lactation, although in itself it can be a primary physiological process. Physiological causes include mechanical stimulation of the breast and stress. It may occur for some years after cessation of breast-feeding. Under these physiological circumstances, simple reassurance is all that is required.

Galactorrhoea may also be a secondary phenomenon, occurring as a side-effect of drugs that enhance dopamine activity, such as chlorpromazine, haloperidol, metoclopramide, and methyldopa. Hyperprolactinaemia due to a primary prolactin-secreting tumour or from a secondary source, such as bronchogenic cancer, is an uncommon cause of galactorrhoea, but should be excluded if symptoms persist in the absence of an obvious cause.

Duct ectasia Duct ectasia is a common cause of nipple discharge. Characteristically it causes a multifocal, bilateral discharge that is thick and opalescent, and of variable colour. However, the discharge of duct ectasia can be unifocal and frankly bloodstained, particularly in the perimenopausal and older age groups.

Duct papillomas Papillomas are one of the more important causes of nipple discharge. The discharge is often from a single duct; it is frequently serous or serosanguinous, and is frankly bloodstained in 50 per cent of cases. Papillomas also account for many of the relatively unusual cases of a watery discharge.

Most papillomas are solitary and are not considered to be premalignant. Occasionally, two or three papillomas may be found in a single duct, and they too are unlikely to have neoplastic potential (Fig. 18). Multiple papillomas, however, especially those occurring in the periphery of the breast and affecting more than one duct, carry an increased risk of malignant change. In one series, 15 of 39 patients with multiple papillomas developed carcinoma. These peripheral lesions are more likely to cause a breast mass than a nipple discharge.

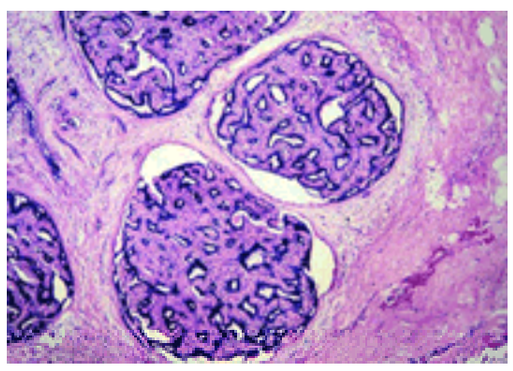


Fig. 18. Multiple intraduct papilloma (by courtesy of Dr D. Tarin).

Carcinoma As discussed below, intraductal carcinoma, and even invasive cancer, can present as nipple discharge. This is usually from a single duct, is frequently watery or serous, and is often frankly bloodstained.

Cysts Cysts may be more common as a cause of nipple discharge than is generally appreciated. Some regard cysts simply as dilated ducts and patients can occasionally produce a discharge simply by compressing an area of cystic change.

Idiopathic Despite all attempts at establishing a diagnosis, even with biopsy, about 10 per cent of patients have no obvious pathological entity ascribed to the discharge. If the discharge has been blood-stained it must be considered that the cause has been missed. However, if it was serous and not bloodstained, simple reassurance with careful follow-up will suffice.

Management

If the nipple discharge is associated with a mass, then appropriate treatment for the lump is instituted. If no mass is felt, the management depends on the nature of the discharge. Indications for exploration would include a single-duct discharge, especially if watery or frankly blood-stained. The most common causes for a watery or bloody discharge from a single duct are duct ectasia or papilloma, although carcinoma may rarely present in this way. A serous discharge from a single duct is likely to be due to duct ectasia. Discharge from multiple ducts, particularly if bilateral, is also due to duct ectasia and may be managed conservatively unless the symptoms are cosmetically embarrassing.

Bloody discharge superimposed on a multifocal discharge is usually due to duct ectasia. If this symptom causes the patient distress, then exploration is indicated; if a conservative approach is chosen, the patient should be reviewed after 3 months.

There has been a recent trend to reduce the amount of duct surgery performed because of the relatively low yield of unsuspected malignancy and because of the attendant complications from wound infection, and distortion and numbness of the nipple. Two surgical options are possible. The more conservative approach is microdochetomy, which has the advantage that the woman should be able to breast-feed after the operation. It is therefore probably the treatment of choice in younger individuals with a single-duct discharge. However, as many patients with a single-duct discharge will have a more widespread disorder, simple microdochetomy may be followed by recurrence of symptoms. A more appropriate operation for older women is major duct excision, which also allows treatment of potentially multifocal disease.

Nipple inversion and retraction

The terms nipple inversion and retraction have been used interchangeably, and the distinction is somewhat arbitrary. Nipple inversion describes a congenital failure of eversion during development, whereas retraction relates to a secondary process, usually due to duct ectasia or carcinoma (Table 7).

Congenital nipple inversion

Congenital nipple inversion is a common problem occurring, to a greater or lesser extent, in up to 20 per cent of all women. Frank inversion causing difficulty with breast-feeding is much less common: a degree of flattening of the nipple does not seem to interfere with the breast-feeding process. This is hardly surprising because the nipple itself probably plays a relatively small part in anatomical aspect of suckling; the infant creates a ‘teat' from surrounding breast tissue as well as from the nipple.

Congenitally inverted nipples tend to be bilateral and patients can generally be reassured that there are few long-term sequelae. They should certainly be encouraged to breast-feed. Women with congenitally inverted nipples have a higher incidence of duct ectasia/periductal mastitis.

Young women with congenitally inverted nipples often request surgical eversion. If possible this should be resisted, as the only satisfactory way of everting the nipple is to divide all the underlying ducts, which will prevent subsequent breast-feeding. Furthermore, after such cosmetic surgery the nipples still have a rather flattened appearance rather than being protuberant.

Nipple retraction

The three main causes of nipple retraction are duct ectasia, carcinoma, and the effects of injudicious previous surgery (Table 7).

Duct ectasia Nipple retraction due to duct ectasia is characterized by bilateral changes and a characteristic, linear transverse defect in its early stages. At this time it is possible to evert the nipple, but as the process progresses digital eversion becomes more difficult. Other features of duct ectasia, such as a creamy, multifocal discharge or stigmata of previous periareolar abscesses, are often present. An earlier belief that these changes often begin during pregnancy has not been confirmed.

Carcinoma Retraction due to carcinoma is characterized by a shorter history and is unilateral. It is frequently associated with a mass. The effects of nipple retraction itself and any associated inflammation can make clinical assessment of the retroareolar area difficult. If there is any doubt, mammography and biopsy by major duct excision should be performed.

Postsurgical Injudicious surgery with inadequate care in reconstituting the breast can result in nipple inversion. This is often associated with an ugly periareolar defect and is difficult to rectify. It should be avoided by due attention to surgical technique.

Other disorders of the nipple and periareolar region

While discharge and retraction account for the majority of patients presenting with nipple disorders, a number of other problems are occasionally encountered. Skin disorders, especially eczema, often affect the nipple and periareolar regions. Eczema must be differentiated from Paget's disease and is characterized by a long, intermittent history, bilaterality, and its presence at other cutaneous sites. Features such as itching, serous discharge, and the nature of the periphery of the lesion are of little diagnostic significance. If there is any diagnostic doubt, biopsy is required; a blind trial of topical steroids is to be condemned as Paget's disease can resolve temporarily in response to this treatment.

Fibroepithelial polyps present in adolescence and in young women. Excision, if required, can be performed under local anaesthetic. Chronic sebaceous cysts and retention cysts arising from Montgomery's tubercles are occasionally seen.

Nipple adenomas produce an uncomfortable mass beneath the nipple. They can cause ulceration and may be confused with Paget's disease. Simple excision will suffice; the condition is completely benign.

Pain and hypersensitivity in the nipples is difficult to explain. Sometimes this has a cyclical nature and enquiry will show a traumatic cause. Other patients describe hypersensitivity to cold, as in a Raynaud's type of phenomenon.

Infection of the breast

The majority of breast infections can be subdivided into those occurring during lactation and those that are a complication of duct ectasia/periductal mastitis. The two have entirely different causes, pathogenesis, and treatment, and may be considered completely separately.

Lactational breast abscess

Lactational breast abscesses occur during breast-feeding, are generally somewhat peripherally situated, and are the result of infection by *Staphylococcus aureus*. They tend to occur at the commencement of breast-feeding when an inexperienced mother develops cracked nipples. They also occur at weaning when engorgement results from incomplete drainage of breast milk.

Cracked nipples resulting from the trauma of breast-feeding are seen in the first week after delivery and again after about 6 months when the child's teeth first develop. There is acute pain in the nipple and examination reveals a linear fissure, which may become secondarily infected.

Clinical features

The patient initially complains of discomfort in the breast, followed by painful swelling. The overlying skin may appear red and in extreme cases may undergo necrosis. Constitutional symptoms are common, and by the time the patient presents she may have a high fever and be systemically unwell.

If ignored, breast abscesses, like those elsewhere, will point and spontaneously discharge on to the skin surface.

Treatment

If lactational breast infection is seen before frank abscess formation, antibiotic treatment alone is often successful. Aspiration should be attempted to ensure that no pus is present. As many cases, especially those occurring in hospital, are due to penicillinase-producing staphylococci, treatment must be with a penicillinase-resistant antibiotic such as a second-line penicillin or a cephalosporin. Such treatment will be successful in about 90 per cent of early cases.

If, on aspiration, pus is found, or if other systemic features of an abscess are present, drainage is necessary. However, some recent authorities have recommended repeated daily aspiration under antibiotic cover rather than formal surgical intervention. If formal drainage is performed, however, further antibiotic therapy is no longer required. The incision used to drain the abscess should limit any cosmetic deficit and allow drainage of pus under the influence of gravity. The majority of surgeons leave the wound open and pack it daily with antiseptic-soaked ribbon gauze, although primary closure under antibiotic cover has also been recommended.

An area of confusion is whether breast-feeding should continue after treatment of a breast abscess. There is no place for suppression of lactation. The infant should be encouraged to feed from the contralateral breast while the affected side should be emptied either manually or with a breast pump. Cracked nipples should be washed gently and dried by dabbing rather than by rubbing. If there is evidence of nipple infection, a nystatin-containing ointment should be applied topically.

Non-lactational breast abscesses

Non-lactational breast abscesses are entirely different from those occurring during breast feeding. They arise in the periareolar tissues, frequently recur, and the infecting organisms (if successfully cultured) are a mixture of bacteroides, anaerobic streptococci, and enterococci. Non-lactational abscesses are a manifestation of duct ectasia/periductal mastitis and are usually seen in the 30- to 60-year age groups.

Clinical features

Patients with non-lactational breast abscesses often have a history of previous infective episodes in the periareolar region. Abscesses often begin as a slightly tender, periareolar mass. Spontaneous resolution is common but they often progress, with associated reddening of the overlying skin and increasing tenderness until the features of a frank abscess are present. Systemic upset is less pronounced than in patients with lactational abscesses. Scarring and distortion arising from treatment of similar episodes may be present, and there may be other manifestations of duct ectasia/periductal mastitis, such as nipple retraction.

Treatment

If a non-lactational abscess is suspected, the inflammatory mass should be aspirated and sent for both aerobic and anaerobic culture. Initial treatment with metronidazole and flucloxacillin may be successful, although repeated aspiration may be required. Many patients will require drainage, although open drainage should be avoided if possible. If drainage is necessary, it should be performed through the smallest possible incision.

Because of the association between smoking and periductal mastitis, patients with chronic relapsing symptoms should be urged to stop smoking.

Definitive treatment requires duct excision and possible nipple eversion, which should be performed under appropriate antibiotic prophylaxis and only when the condition is quiescent. The majority of surgeons leave the wound open to heal secondarily, although there is an increasing vogue for primary closure under antibiotic

cover if local conditions are suitable.

Surgical or spontaneous discharge of an abscess from periductal mastitis may result in a mammary fistula with intermittent drainage of pus or serous fluid on to the areola; this may be superimposed on further episodes of periductal sepsis. Characteristically, the breast demonstrates multiple incisions for drainage of previous abscesses, distortion, nipple retraction, and a fistula at the edge of the areola (Fig. 19). The basic cause is, again, duct ectasia/periductal mastitis.

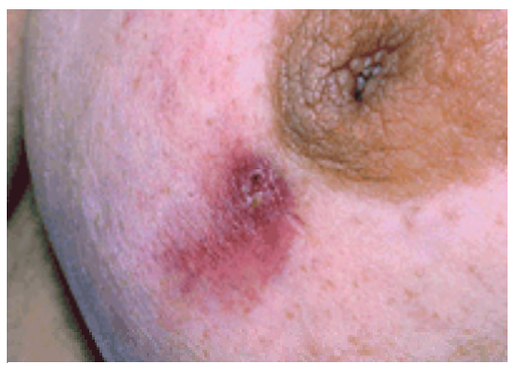


Fig. 19. Mammary fistula: note the fistula opening on to the border of the areola.

Treatment is by fistulectomy, excision of the offending duct, and nipple eversion if required. If there is extensive periareolar disease or the nipple is grossly retracted, a major duct excision should be performed. Most surgeons prefer to leave the wound open, but again there is a growing trend for primary closure.

Fistula recurs in about 5 per cent of patients. The cause is variable, but includes continuation of duct ectasia/periductal mastitis in adjacent ducts and persistence of the proximal duct adjacent to the nipple because of an inadequate surgical technique.

Other causes of breast infection

Postoperative wound infections are relatively common after breast surgery, especially following surgical treatment of duct ectasia/periductal mastitis. Reasons for this include the relatively poor blood supply to fat, ischaemia resulting from deep sutures, and an accumulation of serum in the wound itself. Prophylactic antibiotics and drainage seem to have little impact in reducing the incidence of such problems; they are best minimized by meticulous technique.

The incidence of wound infection after surgery for duct ectasia/periductal mastitis has been so high in some centres that antibiotic prophylaxis with drugs such as combination metronidazole and flucloxacillin is now widely used.

Breast abscesses occasionally occur in neonates due to infection of milk induced by the transplacental passage of maternal hormones. If antibiotics do not help this condition, great care must be taken during surgical drainage as damage to the breast disc at this age may lead to distortion in later life.

Tuberculosis, syphilis, hidradenitis, and pilonidal abscess have all been described as causing inflammation in the breast.

The relation of benign to malignant breast disease

General risk factors for breast cancer are discussed in the following chapter. Some pre-existing benign breast lesions are considered to be a risk factors, although the literature on the subject is conflicting. Studies have generally been based on selected populations who have had incomplete follow-up for an inadequate amount of time. A major problem has been lack of consensus between pathologists in ascribing individual terms to the various microscopic features and an obsession with the cancer risk associated with fibrocystic disease.

Descriptive terms

The following terms require definition before further discussion.

Hyperplasia

This is cellular proliferation (Fig. 20). In the breast the normal ducts are lined by two layers of cells above a basement membrane. Hyperplasia is therefore defined as the presence of three or more layers of cells, although individual cells may fill up, or protrude into, epithelial-lined spaces. It is synonymous with the terms papillomatosis and epitheliosis.

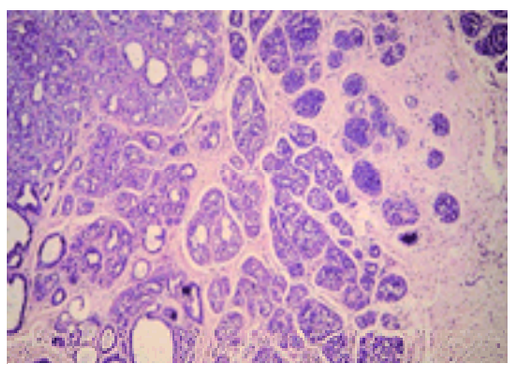


Fig. 20. Hyperplasia: note the benign epithelial proliferation; normal ducts are filled with hyperplastic cells (by courtesy of Dr D. Tarin).

Atypia

This occurs when hyperplastic cells exhibit bizarre or unusual features, either in the pattern of their cellular relations or in the appearance of their nuclei (Fig. 21). The extent of atypia can be graded from 1 to 3, although this classification is empirical and subject to variation between pathologists.

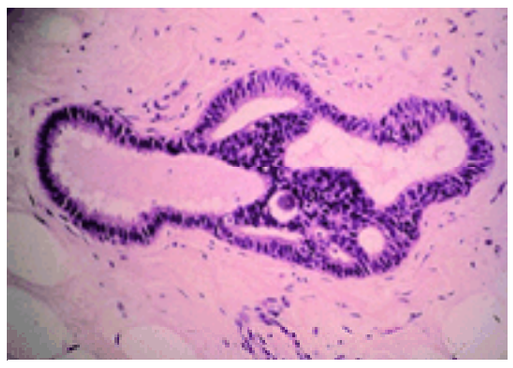


Fig. 21. Atypical ductal hyperplasia: cells lining the duct show loss of polarity, nuclear pleomorphism, and atypia (by courtesy of Dr D. Tarin).

Adenosis

This is an increase in the number of glandular elements. There is a normal relation between the cells and the basement membrane (Fig. 22).

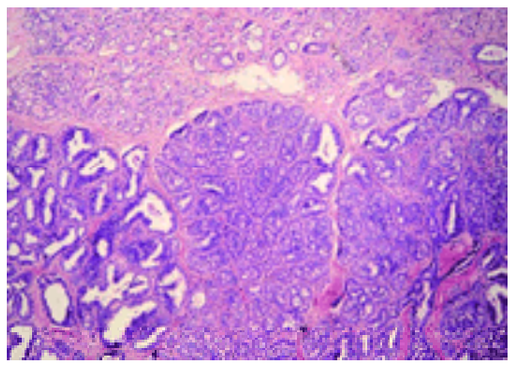


Fig. 22. Nodular adenosis (by courtesy of Dr D. Tarin).

Epitheliosis

This controversial term has found greater acceptance in the United Kingdom than in the United States. It has been used to describe the solid or semisolid, benign epithelial proliferation that is found predominantly in the small ducts, ductules, and lobules. It has been implied that severe forms of epitheliosis amount to carcinoma *in situ*, although this point of view has been widely criticized. There has been a recent tendency to replace the term epitheliosis with hyperplasia.

Papillomatosis

This is also a controversial term, which has found favour in North America. It has been criticized as it implies a derivation from the term papilloma yet the vast majority of examples bear no resemblance to a papillary pattern. Hyperplasia is being used instead of papillomatosis.

Studies of risk

The first study to use accurate data to assess the risk of cancer in patients with benign breast disease was that of Page *et al.*. Its results formed the basis of an American Cancer Society consensus statement on the risk of breast cancer in patients with benign disease, which was first published in 1986 (Table 8). Cancer risk is defined as a liability to develop breast cancer in the ensuing 10 to 20 years, compared with development in age-matched women who have had no breast biopsy. It should be noted that these are not lifetime risks.

No increased risk
Sclerosing adenosis
Apocrine change
Duct ectasia
Mild hyperplasia
Fibroadenoma
Cysts
Apocrine metaplasia
Slight increased risk (1.5–2 times)
Moderate or florid hyperplasia
Papilloma with fibrovascular core
Moderate increased risk (4–5 times)
Atypical ductal hyperplasia
Atypical lobular hyperplasia
High risk (8–10 times)
Ductal carcinoma *in situ*
Lobular carcinoma *in situ*

Table 8 Relation between benign and malignant breast disease

In their original study, Dupont and Page ascribed a slightly increased cancer risk to patients with sclerosing adenosis. Other studies did not confirm this finding and sclerosing adenosis is therefore currently considered as having no increased cancer risk.

The American consensus statement maintained that breast cysts were not associated with an increased cancer risk. Other studies have confirmed this opinion, but there is an increasing body of data suggesting that macroscopic cysts, especially those of the apocrine type, may be associated with an increased risk of breast cancer. Further work is required to clarify this question.

The inclusion of *in situ* cancer may be criticized, as this diagnosis implies a neoplastic rather than a benign process. Conversely, multiple papillomatosis, not included in the above discussion, has been clearly demonstrated to be associated with an increased cancer risk. On the other hand, recent data do indicate a slight long-term risk from fibroadenoma, especially if of a histologically complex type.

The relative risk of developing invasive breast cancer following benign disease implies the existence of the classical pathway from normality to invasive cancer:

Normal ® hyperplasia ® atypia (grades 1–3) ® carcinoma *in situ*® microinvasive cancer ® clinically invasive cancer.

Care must be taken in interpreting this rather crude and perhaps naive pathway. There is no evidence that the progression from normality to invasive cancer needs to pass through all the above stages. Furthermore, once a certain point in the pathway has been reached there does not have to be further progression; in theory, regression may occur.

Further reading

Ader DN, Shriver LTC. Update on clinical and research issues in cyclical mastalgia. *The Breast Journal* 1998; **4**: 24–32. [An important review on the modern approach towards breast pain integtrating research results with pharmacologic therapies.]

Anderson TJ, Page DL. Risk assessment in breast cancer. In: Anthony PP, MacSween RNM, Lowe DG, eds. *Recent Advances in Histopathology*. Edinburgh: Churchill Livinstone,1997; 69–91. [An updated account of the most important work on the relation between benign and malignant breast disease.]

Bundred NJ, Dover MS, Aluwihare N, Faragher EB, Morrison JM. Smoking and periductal mastitis. *British Medical Journal* 1993; **307**: 772–3. [The first account of the correlation between smoking and duct ectasia.]

Dixon JM, Scott WN, Miller WR. Natural history of cystic disease: the importance of cyst type. *British Journal of Surgery* 1985; **72**: 190–2. [A review of the two types of breast cyst which compares features of simple and apocrine cysts.]

Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *New England Journal of Medicine* 1985; **312**: 146–51. [Seminal paper on relation between benign and malignant breast disease.]

Fentiman IS, Caleffi M, Hamed H, Chaudar MA. Dosage and duration of tamoxifen treatment for mastalgia: a controlled trial. *British Journal of Surgery* 1988; **75**: 845–6. [The initial randomized trial of tamoxifen in patients with mastalgia.]

Gateley CA *et al.* Plasma fatty acid profiles in benign breast disorders. *British Journal of Surgery* 1992; **79**: 407–9. [An initial account of the relation between mastalgia and plasma fatty acids.]

Haagensen CD. Mammary duct ectasia—a disease the may simulate cancer. *Cancer* 1951; **4**: 749–61. [A historically important account of duct ectasia.]

Hughes LE, Mansel RE, Webster DJT. Aberrations of normal development and involution (ANDI): a new perspective on pathogenesis and nomenclature of benign breast disorders. *Lancet* 1987; **ii**: 1316–19. [The initial concept of ANDI and its importance in classification of benign breast disease.]

Hughes LE, Mansel RE, Webster DJT. *Benign disorders and diseases of the breast: concepts and clinical management*. Baillière Tindall, London, 1989. [The most comprehensive account of benign breast disease.]

Maddox PR, Harrison BJ, Mansel RE, Hughes LE. Non-cyclical mastalgia: an improved classification and treatment. *British Journal of Surgery* 1989; **76**: 901–4. [The attempt to classify mastalgia according to symptomology.]

Mansel RE, Dogliotti L. European multicentre trial of bromocriptine in cyclical mastalgia. *Lancet* 1990; **335**: 190–3. [An important randomized trial in management of mastalgia.]

Mansel RE, Wisbey JR, Hughes LE. Controlled trial of the antigonadotrophin danazol in painful nodular benign breast disease. *Lancet* 1982; **i**: 928–31. [An important randomized trial in management of mastalgia.]

Page DL, Anderson TJ. *Diagnostic histopathology of the breast*. Churchill Livingstone, Edinburgh, 1987. [A comprehensive account of breast pathology.]

Thomas WG, Williamson RN, Davies JD, Webb AJ. The clinical syndrome of mammary duct ectasia. *British Journal of Surgery* 1982; **69**: 423–5. [An account of the outcome of patients with duct ectasia.]

Turner-Warwick RT. The lymphatics of the breast. *British Journal of Surgery* 1959; **46**: 574–82. [The most important experimental evaluation of the lymphatic drainage of the breast.]

Winchester DP. American College of Pathologists consensus statement. The relationship of fibrocystic disease to breast cancer. *American College of Surgeons Bulletin* 1986; **71**: 29–31. [An important landmark in the relation between benign and malignant breast disease.]

21.2 Cancer of the breast

Michael J. Greenall and William C. Wood

Demographic features
Risk factors
Major risk factors
Intermediate risk factors
Minor and controversial risk factors
Pathological features of breast cancer
Ductal carcinoma of the breast
Lobular carcinoma of the breast
Natural history of breast cancer
The spread of breast cancer
The clinical presentation of breast cancer
The diagnosis of breast cancer
Evaluation of patients with breast cancer
Staging
Treatment of carcinoma of the breast
The management of cancer of the breast
Management of the breast
Management of axillary nodes
Treatment of potential micrometastases
Special problems in the management of breast cancer
Paget's disease of the nipple
Breast cancer in the elderly
Breast cancer during pregnancy and lactation
Locally advanced cancer (stage III)
Carcinoma *in situ*
Lobular carcinoma *in situ*
Screen-detected cancer
Follow-up of women with breast cancer
The prognosis of breast cancer
Prognostic factors in breast cancer
Metastatic breast cancer
Assessment of patients with metastatic disease
Prognostic indicators in metastatic disease
Treatment of metastases
Management of metastatic disease at specific sites
Psychological sequelae following treatment of breast cancer
Prevention of breast cancer
Further reading

Breast cancer is a major and important form of malignant disease in the Western world. During the second half of the twentieth century there has been a massive increase in the recorded incidence of breast cancer, although in some countries the number of cases is gradually diminishing. In North America, breast cancer has been the most common malignancy among women, accounting for 27 per cent of all female cancers. Eighteen per cent of all cancer deaths are currently due to cancer of the breast, although since the mid-1980s the mortality from lung cancer has equalled and subsequently exceeded that of breast cancer as a cause of death in women. In the United States of America, 180 000 new cases of breast cancer were diagnosed in 1997 and there were about 44 000 deaths from the disease. The equivalent figures from the United Kingdom are 30 000 and 15 000, respectively.

The risk of a woman developing breast cancer is a controversial subject and may be expressed in a variety of ways. In the Western world the cumulative risk (the proportion of a fixed group of women developing breast cancer over a set period of time) is about 7 per cent up to the age of 70. Thus, 1 in 14 women can expect to develop breast cancer. In the United States it is 1 in 10.

Demographic features

There is remarkable variation in the incidence of breast cancer between different countries. The rates in the United States and Canada are six times higher than those in Asia or black Africa. Rates nearly equal to those seen in North America occur in north European countries and in New Zealand. Intermediate rates occur in eastern and southern European countries, South America, and the Caribbean. Japan has a low incidence of breast cancer, although it is becoming more common.

The difference in breast cancer rates is not simply a function of genetic susceptibility. The incidence of breast cancer in black Americans parallels that of white Americans rather than that of black Africans, and the cancer incidence of offspring of migrants to the United States of America from Japan is similar to that of native Americans.

Risk factors

The cause of breast cancer is unknown. However, epidemiological data indicate well-defined factors that show an increased likelihood of developing the disease. Such risk factors for breast cancer fall into three main groups: genetic, endocrine, and environmental; each may be of major, intermediate, or minor importance (Table 1). Many minor factors are the source of continued debate, but as evidence accumulates, some of these factors assume a more important role in the aetiology of breast cancer. Recently, this has been particularly true of oral contraceptive and hormone replacement therapies.

Major factors
Gender
Age
Previous breast cancer
Family history and genetic predisposition (BRCA 1 or 2 mutations)
Intermediate factors
Alcohol and diet
Endocrine factors
Early menarche
Late menopause
Oral contraceptive and hormone replacement therapy
Nulliparity
Irradiation
Benign proliferative breast disease (e.g. multiple papillomatosis)
Benign breast disease (e.g. hyperplasia with moderate or severe atypia)
Minor or controversial factors
Body size
Stress
Benign breast disease (e.g. hyperplasia with moderate or mild atypia)

Table 1 Risk factors for breast cancer

Major risk factors

Some major risk factors are self-evident but require inclusion for the sake of completeness.

Gender

Breast cancer is 100 times more common in women than in men. In strict epidemiological terms, therefore, female sex is a major risk factor for breast cancer, although

it is often forgotten as such.

Age

As for other epithelial cancers the incidence of breast cancer increases with age. Breast carcinoma is only occasionally seen in the late teens but thereafter there is a rapid rise in age-specific rates. Up to the age of 40 years, the increase in age-specific cancer rate is very steep; the rate of increase then slows dramatically, although the overall breast cancer rate continues to rise until old age (Fig. 1). The cumulative risk of developing breast cancer between the ages of 20 and 40 is 0.5 per cent, whereas between 50 and 70 it is 5 per cent. This accounts for the fact that the majority of patients presenting with breast cancer are over the age of 50 years.

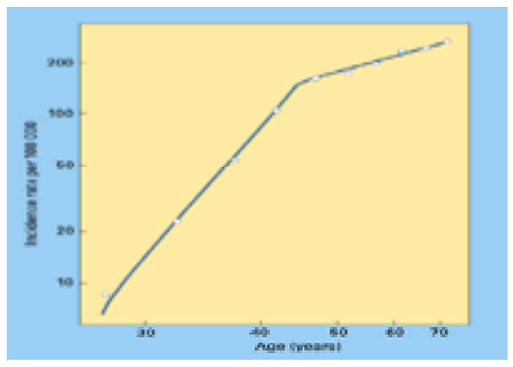


Fig. 1. Age-specific breast cancer incidence. Note the rapid increase in age-specific rate up to age 40.

Previous breast cancer

The development of a second breast cancer may be a clinical manifestation of multifocal origin of the first cancer or may be an entirely new cancer. The risk of developing metachronous tumours has been addressed in a number of studies which have evaluated the development of cancer in the remaining breast after partial mastectomy. The incidence of such second cancers depends on whether or not synchronous bilateral cancers are included, but there appears to be an overall increased risk of 0.5 to 1 per cent per year. Thus, the relative risk of developing a second, non-synchronous primary 20 years after initial diagnosis of breast cancer is 1.2 to 1.5. This risk appears greatest in young women if their initial breast cancer is diagnosed before the age of 40.

Family history and genetic predisposition

A family history of breast cancer is associated with an increased risk of the disease. The risk is greatest in patients with first-degree relatives (mother or sister) affected, especially if under the age of 50 when the disease developed. The relative risk of developing breast cancer is 1.7 to 2.5 in women with a history of breast cancer in a first-degree relative, and 1.5 among those with an affected second-degree relative. Multiple family members with breast cancer, the existence of bilateral disease, or the identification of an affected male all indicate excessive risk—especially if in association with ovarian cancer among other relatives.

There is probably an inherited genetic factor involved in the development of breast cancer in about 5 per cent of all patients. Because all cancer is genetically controlled, it is not surprising that a susceptibility to breast cancer can be inherited. Up to 20 per cent of women diagnosed with breast cancer have at least one affected family member and about 5 per cent of breast cancers appear attributable to inherited gene mutations that are dominant and highly penetrant. There may be a higher risk of inheriting a genetic susceptibility for a lower penetrance of breast cancer that is at present indistinguishable from sporadic breast cancer.

In 1990, gene linkage studies of families with breast cancer identified a locus on chromosome 17q21 and the putative gene was named BRCA 1. The next year the relationship of BRCA 1 to ovarian cancer was appreciated and by 1994 the gene had been cloned. BRCA 2 was located on chromosome 13q12 that same year. By 1996 commercial DNA testing was available for both genes. Mutations of these two genes appear to account for over half of all the identifiable breast cancer families.

Although testing is available, its value is related to the desires of a person for possible intervention should she have inherited such a mutation. Oophorectomy seems appropriate, especially by minimally invasive techniques, for a mutation-positive woman who has completed her family. Although primary peritoneal carcinoma may still occur, the difficulty in detecting early ovarian carcinoma would seem to warrant recommending prophylactic surgery. Oral contraceptive agents taken for 5 years cut the risk of ovarian carcinoma in half and should also be recommended.

In women at increased risk, 5 years of tamoxifen has been shown to reduce the overall risk of breast cancer by about half. Unlike ovarian carcinoma, data specific to mutation-positive women are lacking. The lifetime risk of breast cancer appears to be about 60 per cent, but may vary by the specific mutation and other, presently unknown, modulators of penetrance. Bilateral mastectomy will reduce the incidence by about 90 per cent, but is not to be undertaken lightly, despite the possibility of immediate reconstruction. Breast cancer is more easily detected early than is ovarian cancer, and the possibility exists of better methods of prevention arising in the next decade. Consequently, the consideration for prophylactic mastectomy should be couched in the risk over the next 10 years and the risk of ultimate mortality from breast cancer, rather than focusing on the issue of lifetime incidence.

Genetic testing should be performed in a setting of genetic counselling and with informed consent, because issues of survivor guilt in mutation-negative family members and anxiety about insurability of mutation-positive persons needs to be considered, as well as the need for continued careful screening of persons who do not have BRCA 1 or 2 mutations but may have an inherited susceptibility that has not yet been defined. Families with early-onset breast cancer in several first-degree relatives, especially with any history of ovarian cancer or male breast cancer, or who have Ashkenazi Jewish or Icelandic backgrounds may be at sufficient risk to consider testing. A paternal history carries the same risk as a maternal family history.

Breast cancer arising in persons with BRCA 1 mutations appears to carry the same risk as sporadic cancer of the same characteristics and stage. It is twice as likely to be hormone receptor negative, however, and the development of contralateral breast cancer is at least four times as frequent. It might be assumed that breast conservation therapy would be inappropriate for women with BRCA 1 or 2 mutations, but early data do not support such a contention. This agrees with data that family history does not correlate with local failure after breast sparing.

Benign breast disease

Benign disease is not usually recognized as a major risk factor, although multiple papillomatosis may be regarded as such.

Intermediate risk factors

Diet and alcohol intake

These factors are important as they are to an extent under the control of individual women. Weight does correlate with breast cancer risk although a causal relationship between dietary fat or cholesterol intake has not been clearly demonstrated. Evidence for an association between alcohol consumption and an increased likelihood of developing breast cancer is becoming stronger. The relative risk for one unit of alcohol, such as a glass of red wine, per day is 1.1 and increases to 1.3 to 1.5 if intake increases to two glasses a day.

Many intermediate factors have an endochrinological basis and may relate to the number of menstrual cycles to which the breasts are exposed.

Endocrine factors

Endochrinological factors may relate to the number of menstrual cycles to which the breasts are exposed.

Nulliparity

Nulliparity removes a protective effect against breast cancer. Single and nulliparous married women have a relative risk of 1.4 compared with parous women, although this effect is almost completely confined to the protective influence of early age at first birth. Women who give birth to their first child before the age of 20 have a relative risk of 0.5 compared with nulliparous women; for those whose first birth occurred after the age of 30 there appears to be virtually no protective effect, with a relative risk of 0.94. Some evidence even suggests that women whose first birth is over the age of 35 may have an increased risk of breast cancer.

If the age at first birth is taken into account, subsequent pregnancies appear to have less influence on the risk of developing breast cancer, although they do offer some protection. The protective effect occurs only if the pregnancy continues to full term.

Any protective effect of breast feeding is hard to assess as it is difficult to differentiate from the effects of pregnancy. However, recent data indicate an independent protective effect of breast feeding, although not all studies confirm this finding.

Age of menarche and menopause

Women whose menarche occurs before the age of 12 have a relative risk of 2.30 compared with those starting menstruation after this age. The risk decreases as the age of onset of menstruation increases. The reduction in the age of the menarche over the past 100 years, especially in the Western world, probably results from improved nutrition and general health and may be important in the demographic variations in incidence of breast cancer.

The risk of developing breast cancer also relates to the age of the menopause. The relative risk of developing breast cancer is 0.5 per cent in those who cease to menstruate before the age of 45, compared with women who continue menstruating beyond age 55. Artificial menopause by oophorectomy or irradiation also reduces the risk of breast cancer.

Oral contraceptive and hormone replacement therapy

The importance of these two factors is becoming increasingly appreciated; until relatively recently there was disagreement as to whether they had any significant role in the aetiology of breast cancer.

Meta-analysis has demonstrated the relative risk of developing breast cancer while taking oral contraceptives is 1.24. On stopping therapy this increased risk diminishes to 1.01 over the ensuing 10 years. Thus, there is no lifetime risk of breast cancer from oral contraceptive use as had previously been feared by some authorities.

Hormone replacement therapy has similarly been shown by meta-analysis to be associated with an increased risk of breast cancer, although this risk is not excessive until 5 and perhaps as long as 10 years of use. After this time it may be associated with approximately 6 additional cases per 1000 women (relative risk of 1.3); this increases to 12 extra cases per 1000 women after 15 years of use.

Irradiation

An increased risk of breast cancer has been demonstrated in survivors of atomic explosions, women treated by radiation for postpartum mastitis, and patients receiving multiple chest radiographs during assessment of tuberculosis. This increased risk becomes apparent after a latent period of 10 to 15 years: the effect is most obvious in women exposed to irradiation when under the age of 35; there is little increased risk in women exposed after the age of 40.

Benign breast disease

Severe atypia with hyperplasia is associated with a moderately increased risk of developing breast cancer. This association is greatest in women with a family history of breast cancer and may identify those individuals who have an unidentified genetic mutation.

Minor and controversial risk factors

Body size

There is a minor relationship between body size and breast cancer but this is dependent on age and whether height or body mass is considered. There may be an association between body fat and risk from hormone replacement therapy.

Stress

There is no evidence that stress may lead to the development of breast cancer.

Benign breast disease

As indicated above, some pathological entities, such as multiple papillomatosis and hyperplasia with gross atypia are certainly associated with an increased risk of breast cancer (3.0). The risk is lower with lesser degrees of atypia. Patients with recurrent macroscopic apocrine cysts may also have a slightly increased risk of breast cancer, but convincing evidence of this relationship is lacking.

The relationship between benign breast disease and risk of cancer is bedevilled by ascribing cancer risk to fibroadenoma and ‘fibrocystic change'. There is no increased cancer risk for these two entities

Pathological features of breast cancer

The histological assessment of breast cancer is of paramount importance in establishing the diagnosis of the tumour. It also helps determine the patient's prognosis and allows a greater understanding of the biology of the disease in any one particular case. Even relatively recently, pathologists were satisfied in ascribing the term ‘breast carcinoma' to all breast tumour specimens; the complexity of the subject is such that a more precise histological diagnosis is now required.

There are many methods of pathologically classifying breast cancer; most are based on whether the tumour is invasive or non-invasive and whether it is derived from the duct system or the lobule (Table 2). However, it should be remembered that most tumours arise from the terminal ductules. This fact also explains the common occurrence of mixed tumours with both lobular and ductal components existing within the same breast cancer.

	Approximate incidence in symptomatic population (%)
Ductal carcinoma	
In situ	5
Invasive	
Invasive (NOS)[a]	65
Medullary	5
Tubular	3
Mucinous	3
Papillary	3
Cribriform	3
Others (e.g. signet ring, clear cell, inflammatory)	3
Lobular carcinoma	
In situ	1
Invasive	10

[a]NOS, not otherwise specified.

Table 2 Pathological classification of breast cancer

Ductal carcinoma of the breast

This is the most common form of breast cancer accounting for 85 to 90 per cent of all cases. It can conveniently be subdivided into *in situ* and invasive types.

Ductal carcinoma *in situ*

Ductal carcinoma *in situ*, or intraductal carcinoma, is a preinvasive form of breast cancer. It is characterized by a proliferation of malignant breast epithelial cells, is confined to the duct system, and does not invade the basement membrane or surrounding tissues. In the unscreened population, ductal carcinoma *in situ* accounts for less than 5 per cent of all cases of breast cancer; with screening programmes the incidence increases to 15 to 20 per cent because it is associated with mammographically visible microcalcification. Postmortem studies show that the incidence of ductal carcinoma *in situ* is about 6 per cent in apparently normal breasts, but is as high as 40 per cent in mastectomy specimens associated with invasive cancer. The most important aspect of ductal carcinoma is its malignant potential. The little that is known of the natural history of this entity comes from retrospective evidence of treating ductal carcinoma *in situ* by local excision alone. Such studies demonstrate a 30 to 50 per cent incidence of ipsilateral invasive cancer, usually in the same quadrant, after an interval of some 10 to 15 years. The risk of developing invasive cancer depends on the extent of ductal carcinoma *in situ* and its grade. The traditional classification of ductal carcinoma *in situ* was based entirely on morphological grounds. Thus, traditionally, five main histological types of intraductal carcinoma have been identified. The ducts, the terminal duct lobular units, and the lobules may be filled with malignant cells (solid type), the lesions may undergo central necrosis (comedo type), or they may have a sieve-like appearance (cribriform type). They may produce papillary projections (papillary or micropapillary types) or they may cling to the duct wall (clinging type). Comedo and papillary forms (Fig. 2) are the most common and are both associated with multicentric disease; comedo carcinoma is the most likely to become invasive and has the greatest expression of both DNA aneuploidy and of the C-*erb*2 oncogene.

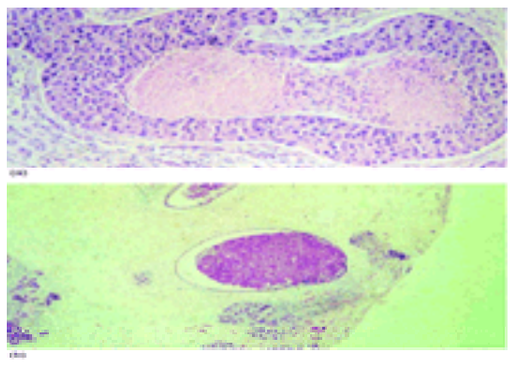


Fig. 2. Ductal carcinoma *in situ.* Duct distension with cancer cells and central necrosis (comedo type). (b) Papillary intraductal carcinoma distending the duct. In neither case is there evidence of invasion through the basement membrane. (By courtesy of Dr D. Tarin.)

The above classification, however, suffers from its relative lack of clinical correlation and therefore there have recently been attempts to classify ductal carcinoma *in situ* according to the likelihood of local recurrence or development of invasive disease. Such classification relies not only on morphology but also on nuclear grade, the presence of necrosis or mitotic figures, and loss of cell polarization. A number of different classifications exist; that recommended by Lagios and his colleagues has been most widely adopted. In this classification, ductal carcinoma *in situ* is categorized as high, intermediate, or low grade according to size, nuclear diameter, the distribution of chromatin within the nucleus, the number of nucleoli, the mitotic rate, and absence or presence of necrosis. Modifications of this system, such as used by Silverstein and colleagues, have been shown to correlate with local recurrence and probability of developing invasive disease.

Invasive ductal cancer

Once intraductal carcinoma has invaded the basement membrane of the duct it has demonstrated the most important prerequisite of all malignant tumours—the ability to infiltrate into surrounding tissue. The term ‘invasive ductal cancer' therefore refers to the majority of tumours arising from the duct system. A large number of different morphological types of invasive duct cancer is apparent to the pathologist. Some have prognostic importance. However, they can be basically subdivided into those with specific histological features, and those in which there is no characteristic microscopic appearance. The majority of invasive ductal cancers fall into the latter group. These are described as infiltrating ductal carcinoma, not otherwise specified (**NOS**).

Infiltrating ductal carcinoma (NOS)

This is essentially a diagnosis for exclusion, but it accounts for about 65 per cent of all invasive mammary cancers. No specific gross or microscopic features allow recognition of this type of cancer: it is simply the lack of specific and consistent histological features that characterize lobular carcinoma and other special types of ductal cancer (Fig. 3). In a system of histological grading, infiltrating ductal carcinoma, not otherwise specified, would usually be classified as intermediate or high grade and, within the spectrum of invasive breast cancer, it has a relatively poor prognosis. However, other features such as tumour size and axillary nodal involvement are of more prognostic importance than histological features alone.

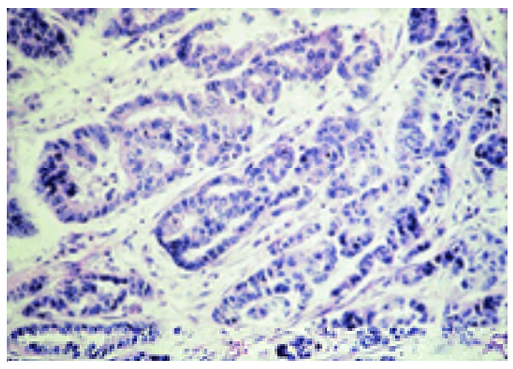


Fig. 3. Invasive ductal cancer (not otherwise specified). Malignant cells and poorly differentiated duct structures can be seen invading the surrounding stroma. (By courtesy of Dr D. Tarin.)

Special types of infiltrating ductal carcinoma

Medullary carcinoma Medullary carcinoma is a well recognized entity for about 6 per cent of all invasive mammary cancers (Fig. 4). Macroscopically they tend to be soft, well circumscribed, and have a uniform consistency. They have a smooth contour and may even appear to have a pseudocapsule. Their most obvious histological characteristic is a prominent lymphocytic infiltration, a feature not seen in ductal carcinoma (NOS).

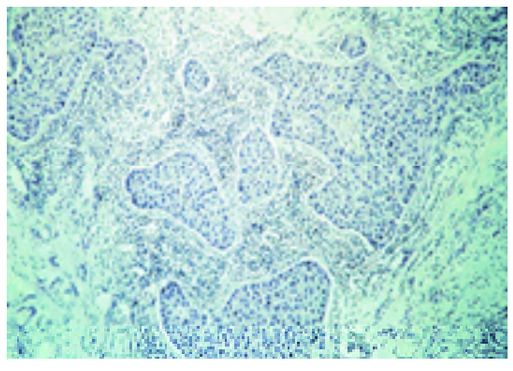


Fig. 4. Medullary carcinoma. Note associated lymphocytic infiltrate. (By courtesy of Dr D. Tarin.)

Medullary carcinoma is less frequently associated with lymph node metastases than other types of breast cancer and is therefore associated with a better prognosis.

Tubular carcinoma This type of breast cancer is uncommon, accounting for only about 3 per cent of all infiltrating cancers of the breast (Fig. 5). Tubular carcinoma is well differentiated and previously had often been classified inappropriately in the not otherwise specified group.

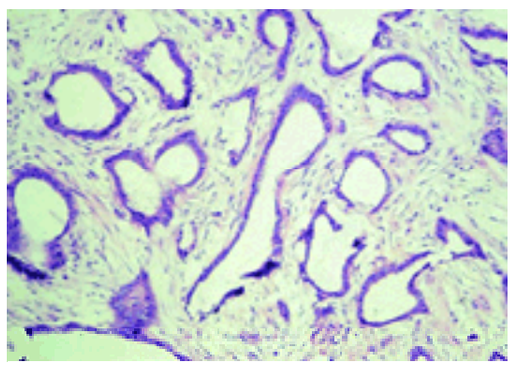


Fig. 5. Tubular carcinoma. Well-differentiated tubular structures are apparent. (By courtesy of Dr D. Tarin.)

Most tubular cancers are small, being about 1 cm in diameter. They are hard and on section have a radial appearance. Histologically they form tubular structures with an open central space lined by a single layer of epithelium.

Recent interest in tubular carcinoma has been heightened by the impact of the breast screening programme. Not only is tubular carcinoma more frequently seen in screened individuals than in the symptomatic population, but also other abnormalities detected on mammography can cause diagnostic confusion. Of particular importance is the differential diagnosis between tubular cancer, sclerosing adenosis, and radial scar.

Patients with tubular carcinoma have a good prognosis and a 10-year survival rate in excess of 75 per cent. Nodal metastases, even when present, tend to be limited to one or two nodes only. It is uncommon for pure, well-differentiated tubular carcinomas to metastasize to distant sites.

It is unclear whether special types of tumour such as tubular carcinoma dedifferentiate into more aggressive types of cancer as they progress (phenotypic drift). Such questions are of paramount importance when assessing the value of breast screening programmes.

Mucinous (mucoid) carcinoma Mucinous carcinoma (Fig. 6) is characterized by a small rounded tumour that histologically demonstrates pools of mucous material in which are aggregates of cancer cells. This tumour accounts for about 2 or 3 per cent of all invasive cancer, but it has been associated with a survival rate which is appreciably better than that of ductal carcinoma (NOS). These tumours are said to occur in older women and, like tubular carcinoma, rarely metastasize to lymph nodes.

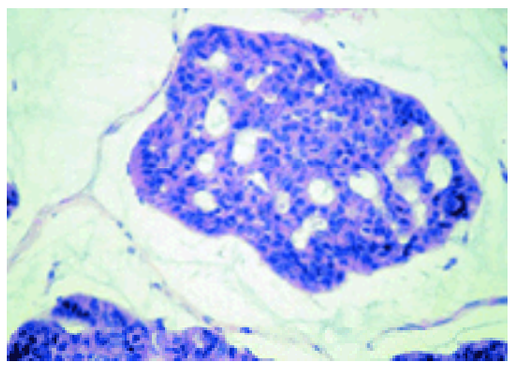


Fig. 6. Mucinous carcinoma. Tumour cells are lying in pools of mucinous material. (By courtesy of Dr D. Tarin.)

Papillary carcinoma Unlike other subtypes of ductal carcinoma, papillary carcinoma can be subdivided easily into *in situ* and invasive types. The non-invasive papillary carcinoma has the fronded structure of a papilloma. It is usually well delineated from surrounding breast tissue by a fibrous covering, hence the essentially inaccurate term of intracystic papillary carcinoma. As long as there is no extracapsular evidence of *in situ* carcinoma, simple excision of these ‘intracystic papillary carcinomas' will suffice.

Invasive papillary carcinoma accounts for only about 2 per cent of all breast cancers. Microscopically they are usually well circumscribed and histologically demonstrate papillary formation. The prognosis is generally good.

Invasive cribriform carcinoma This has only recently been recognized as a special type of invasive ductal carcinoma. It accounts for about 3 per cent of all breast cancers and may be more common in the screened population. It classically presents with a firm mass, but microscopy reveals characteristic stromal invasion with a cribriform pattern. It tends to be well differentiated and have a good prognosis.

Other types of invasive ductal carcinoma

The presence of signet ring cells in breast cancer is a rare but important phenomenon. Most of these tumours are diffuse and poorly defined. Histologically, for reasons which are unclear, they are often associated with lobular carcinoma and are characterized by signet ring cells with a peripherally placed nucleus and marked cytoplasmic vacuolization. Their prognosis is poor; they selectively metastasize to the peritoneal cavity and it becomes difficult to differentiate metastatic signet cell carcinoma from those originating in the gastrointestinal tract.

Clear cell tumours also have a poor prognosis. Their appearance is characteristic on haematoxylin and eosin staining; further selective stains demonstrate either a

lipid-rich or glycogen-rich pattern.

Secretory carcinomas are rare. They are frequently small and are characterized by abundant intra- and extracellular areas of vacuolization. These tumours frequently occur in younger women and generally have a good prognosis; extremely rare breast cancers occurring in childhood are often of this type.

Inflammatory carcinoma has previously been ascribed to tumours occurring in pregnancy and during lactation (mastitis carcinoma-tosis), although it is now apparent that it occurs in all age groups. This type of breast cancer has previously been diagnosed on the pathological observation of tumour emboli in dermal lymphatics, although it is now defined by its clinical appearance with more than 50 per cent of the breast being red and warm in association with a brief history. This condition is associated with a poor prognosis, whether or not the dermal lymphatics contain tumour. Some studies have indicated an improved survival, however, if dermal lymphatics contain tumour (‘pathologically defined inflammatory breast cancer') but the clinical picture is absent.

Other rare types of infiltrating ductal carcinoma include those with carcinoid, adenocystic, and squamous features.

Lobular carcinoma of the breast

Lobular carcinoma can also be conveniently subdivided into *in situ* and invasive forms, depending on whether the basement membrane of the lobule has been invaded by tumour.

Lobular carcinoma *in situ*

Lobular carcinoma *in situ*, like its intraductal counterpart, is also a non-invasive form of breast cancer. The term ‘lobular carcinoma *in situ*' was used in its original description by Foote and Stewart in 1941; other terms, such as intra-acinous carcinoma and lobular neoplasia, have also been widely used.

The microscopic criteria for diagnosing lobular carcinoma *in situ* are well defined (Fig. 7). Within the lobule there must be a uniform proliferation of cells and, as a result, there should be no interstitial spaces and expansion of at least half of the acini in the lobular unit. If these criteria are not completely fulfilled then the term ‘atypical lobular hyperplasia' is appropriate (Fig. 8). The term ‘lobular neoplasia' was used by Haagensen to describe both atypical lobular hyperplasia and lobular carcinoma *in situ*, although its use has been criticized as it implies that atypical lobular hyperplasia is a neoplastic process.

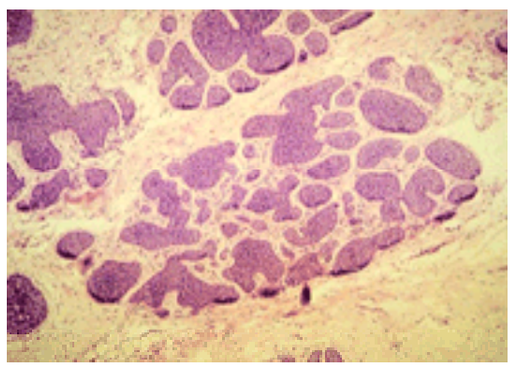


Fig. 7. Lobular carcinoma *in situ.* The acini are distended with a uniform population of malignant cells. There are no interstitial spaces. (By courtesy of Dr D. Tarin.)

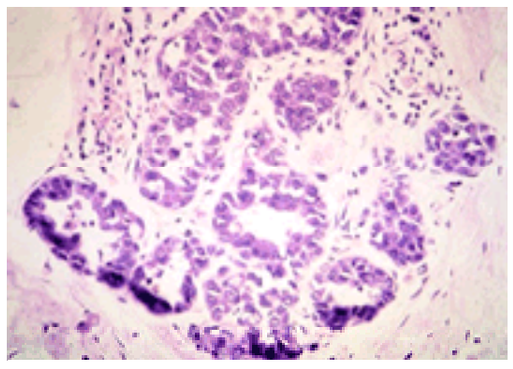


Fig. 8. Atypical lobular hyperplasia. The criteria for lobular carcinoma *in situ* are not met. Only some acini are distended and interstitial spaces remain. (By courtesy of Dr D. Tarin.)

The most important aspect of lobular carcinoma *in situ* is its potential for becoming invasive. The risk of invasive lobular carcinoma developing following simple biopsy for lobular carcinoma *in situ* is only about 25 per cent over 20 years: this is a lower risk than that of ductal carcinoma *in situ.* Unlike ductal carcinoma *in situ*, lobular carcinoma *in situ* rarely expresses the C-*erb*2 oncogene.

The biggest single study evaluating the malignant potential of lobular carcinoma *in situ* is that of Haagensen. As well as identifying the risk of frank malignant change, Haagensen made a number of important observations (Table 3). It is now generally agreed that as the risk of invasive cancer after diagnosis of lobular carcinoma *in situ* (relative risk, 10) relates equally to both breasts, it should be regarded as a risk factor for tumour development rather than a direct precursor. Atypical lobular hyperplasia carries a fourfold increased risk for development of breast cancer.

Predominately premenopausal
Multifocal and bilateral
Does not metastasize
Is not palpable
Does predispose to invasive cancer
50% of invasive cancers develop in the contralateral breast

After Haagensen 1978.

Table 3 Features of lobular carcinoma *in situ*

In practice, lobular carcinoma *in situ* is usually discovered by chance. It has no mammographic features and is usually found on biopsy undertaken for other reasons. As in the case of ductal carcinoma *in situ* there have been recent attempts to grade lobular carcinoma *in situ*, although such experience with this latter condition is limited.

Invasive lobular cancer

Invasive lobular cancer accounts for about 10 per cent of all cases of breast carcinoma, although its incidence varies quite widely. Up to 10 per cent of lobular cancers have a coexisting ductal component. Five main subtypes are described: classic, solid, alveolar, mixed, and pleomorphic. Their main pathological features are summarized in Table 4. The most common type, classical lobular carcinoma, is characterized by the 'Indian filing' of invading malignant cells, often in a 'targetoid' pattern (Fig. 9). The 'monotonous' features of individual cells along with lack of a specific histological architecture has resulted in difficulty in grading lobular carcinoma, although most pathologists now attempt to do so.

Type	*Pattern*	*Cell type*
Classic	Indian filing, targetoid diffuse invasion	Small non-cohesive, regular, monotonous cell population
Solid	Sheet-like	
Alveolar	Globular aggregates	
Mixed	Mixture of above	
Pleomorphic	Diffuse infiltrate	Nuclear pleomorphism

Table 4 Pathological features of invasive lobular cancer

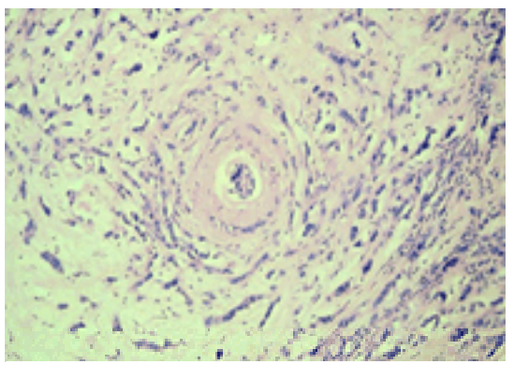


Fig. 9. Invasive lobular cancer. Indian filing of individual cells with a targetoid pattern is apparent. (By courtesy of Dr D. Tarin.)

Invasive lobular carcinoma is important with respect to its tendency to bilaterality and the fact that its clinical diagnosis may be difficult because of the diffuse infiltrative nature which frequently produces distortion of the breast rather than a lump. The prognosis of classic invasive lobular carcinoma is similar to that of invasive ductal cancer. The other types of lobular carcinoma, especially its pleomorphic variant, have a worse prognosis than the classic form of this disease.

Natural history of breast cancer

The lay public, and indeed a number of medical practitioners, believe that transition from normality to breast cancer and subsequently to metastatic disease is a rapid process. The reverse is, in fact, the case. The natural history of breast cancer is normally characterized by long duration, but shows extreme heterogeneity between patients. Breast cancer is one of the more slowly growing tumours, which therefore renders it suitable for screening programmes. The evolution of breast cancer, especially in its preclinical phases, can be measured in years or even decades, although there are exceptions to this rule in which the disease takes on a more aggressive form.

Our understanding of the preclinical phases of breast cancer is based on indirect observations of the morphological changes from normality to hyperplasia, atypia, carcinoma *in situ*, and finally to invasive cancer. It should be remembered, however, that this progression is by no means inevitable and, in theory, stages may be missed out. It is also possible that a given stage may be permanent or may even regress to a more normal state. Histological characteristics may be correlated with other biological factors, such as labelling index, doubling time, and growth fractions. Serial mammography has indirectly indicated that the doubling time for human breast cancer is usually of the order of 100 to 300 days, although exceptions to this are frequently seen. It is also assumed that if a tumour of 1 cm size contains about 10^9 cells it would require about 30 doublings for a malignant cell to produce this volume. This does, of course, assume that there is no cell loss, that there is a linear growth rate, and that the tumour is initially derived from a single cell. Although these assumptions can be challenged, they do indicate that the preclinical phase of breast cancer would still be of the order of about 7 years and often much longer. Clinical observations of patients treated by biopsy alone for hyperplasia or atypia give some support for these figures.

The heterogeneity of breast cancer can also be appreciated from the percentage of cells that are undergoing cell division at any particular time (labelling index). Tumours with high labelling indices are associated with short doubling times and are therefore more aggressive. Overall, the median labelling index is about 3 per cent; the range extends from less than 1 per cent to approximately 40 per cent. High labelling indices are associated with poor prognostic factors such as premenopausal status and with a poor overall clinical outcome.

The natural history of clinically apparent breast cancer is also surprisingly slow. Historical observations on untreated breast cancer as recorded by the Middlesex Hospital, London, and the Connecticut Tumor Registry show an average survival of about 2.5 to 3 years without treatment. Such data can be criticized as they are historical and probably relate to a group of patients presenting with locally advanced rather than metastatic disease. They do, however, indicate the need for prolonged monitoring in patients with clinical breast cancer.

The spread of breast cancer

Once breast cancer is clinically apparent it will, by definition, tend to spread locally and to regional or distant sites. The extent and rate of spread varies greatly from one individual to another and is characteristic of the heterogeneity of the disease. Some tumours spread rapidly to regional nodes or distant sites while remaining small in the breast; others grow slowly to a large size without metastasizing. Clearly, those with the greatest metastatic potential have the worst prognosis.

Local spread within the breast

Within the breast there are three main mechanisms of spread. The most important mechanism is local spread by direct infiltration into the surrounding parenchyma. This occurs by the ramifying projections that give the characteristic macroscopic stellate appearance of breast cancer. If uncontrolled, direct infiltration of overlying skin or the underlying fascia occurs.

A second mode of local spread is by direct infiltration along ducts, although it is unclear whether this represents actual tumour growth or whether it reflects a field change of pre-existing *in situ* disease. The presence of widespread *in situ* cancer explains the phenomenon of multifocality in patients with breast cancer. Multifocality, or cancer within the same quadrant as the primary tumour, should be distinguished from multicentricity, which is defined as the presence of carcinoma outside the quadrant containing the primary tumour. In one study, 43 per cent of 246 patients with invasive breast cancer less than 4 cm in diameter demonstrated multifocality. In 43 per cent of these atients, affected areas were more than 2 cm from the primary tumour; in 9 per cent additional foci of cancer were found more than 4 cm away. Such multifocal tumours are confined to the duct system in 66 per cent of patients, but the remainder demonstrate invasion into surrounding tissues. The incidence and extent of multifocality depends on the size of the primary tumour. These findings have obvious implications when treating breast cancer by conservative surgery.

Local lymphatic and vascular spread within the breast are also of prognostic importance. Of particular importance are those lymphatic pathways extending into the pectoral fascia and subareolar regions.

Regional spread of breast cancer

The regional spread of breast cancer is defined as that to the axillary, internal mammary, and supraclavicular nodes.

Axillary nodal spread

The axillary nodes represent the most important site of regional spread from breast cancer. Spread to axillary nodes is the most important prognostic indicator of breast cancer. Previously, about 45 per cent of all patients have had nodal disease at presentation, although this figure is now as low as 25 or 30 per cent because of earlier clinical presentation and screening. The likelihood of axillary nodal spread is a function of the size of the breast primary, but even when correction is made for tumour size, disease in the axillary nodes remains the most important prognostic variable. For tumours less than 2 cm in diameter, the incidence of axillary nodal spread is less than 20 per cent. The incidence of axillary nodal disease is 35 per cent in those with tumours 2 to 5 cm in diameter, and 50 per cent in patients with tumours greater than 5 cm in size.

The clinical assessment of axillary nodes is notoriously unreliable. About 30 per cent of palpable and apparently diseased nodes are found to be histologically free of metastases; conversely, up to 30 per cent of apparently normal axillary nodes demonstrate histological evidence of metastatic disease.

The extent of nodal disease reported depends to a great extent on the pathological evaluation of the specimens. Not only is there a great variation from one centre to another in the surgical removal of nodes at axillary sampling or clearance, but there are also marked differences among pathologists in retrieving and examining all the nodes provided in the surgical specimen. Techniques such as clearing the axillary fat with xylene to increase nodal yield and more thorough sectioning of the nodes themselves increase the positivity rate. However, few pathologists have adopted these methods.

The relationship between axillary nodal spread and prognosis depends on three factors: the number of nodes affected, the level of axillary nodal disease, and the extent of disease within the nodes themselves. Accurate appraisal of the nodes is therefore dependent on close liaison between the surgeon and pathologist with respect to orientation of the specimen at the time of axillary exploration.

Number of nodes involved Survival rates based on the number of diseased lymph nodes show remarkable unity between studies. Results assessed by the American College of Surgeons have shown that the number of nodes affected by cancer provides valuable prognostic information. Documentation of nodal status as negative or positive for metastatic disease yields only qualitative data and fails to provide the prognostic information that can be obtained from full analysis of the number of nodes affected at axillary clearance. The number of negative nodes recovered on axillary clearance is of no prognostic importance, although some authorities claim that to ensure the axilla is clear of metastases at least four, and possibly up to 10, negative nodes must be isolated. After 5 years, women treated by mastectomy have an 82 per cent relative survival rate if no lymph nodes are involved and a 60 per cent survival rate if one or two lymph nodes are involved at the time of initial treatment. Women with five or six nodes affected have a 47 per cent survival rate; this is reduced to only 31 per cent if 11 or 12 nodes are involved. Women with more than 20 nodes involved have only an 8 per cent survival rate at 5 years.

Thus, according to studies made by the National Surgical Adjuvant Breast Project, within the group of women with positive nodes, it is apparent that if one to three nodes are affected by the tumour the prognosis, while not good, can be relatively optimistic. However, women with four to 12 positive lymph nodes should be given a more guarded prognosis, and those with more than 13 axillary nodes affected have a decidedly poor outcome.

Important issues relating to the evaluation of nodal involvement are adequacy of both surgical technique in exploring the axilla and in pathological examination of the resulting specimen. Failure to retrieve at least 10 nodes may be associated with higher local relapse rates and decreased survival.

Level of disease As previously described, the axillary nodes are conveniently divided into three groups depending on their relationship to the pectoralis minor muscle. Prognosis relates to the level of axillary node affected, although this is a less powerful factor than the total number of nodes affected by the tumour. Thus, Shottenfeld and colleagues showed that patients with disease in level I nodes had a 5-year survival rate of 65 per cent, those with disease at level II had a 5-year survival rate of 31 per cent, and no patients with disease affecting level III nodes survived for 5 years.

Patients with disease in level III nodes usually have a large number of affected lymph nodes. So-called skip metastases, in which metastatic deposits are found at levels II or III, but not at level I, are found in only 2 per cent of patients. Thus, a dissection limited to level I alone is effective in determining the presence or absence of nodal disease but will tend to underestimate the degree of spread in many patients and will not provide accurate prognostic information.

The extent of disease in individual axillary nodes The extent of disease within individual axillary nodes may also be of importance although, as is the case for level of involvement, there is a greater correlation with the number of nodes involved. Deposits less than 2 mm in size (micrometastases) can be detected only by histological methods. Such deposits usually occur in one node only; occasionally they are found in two, but their presence in a third is exceptional in the absence of macroscopic metastases. The prognostic significance of occult micrometastases is the subject of wide debate. They are probably far more common than is suggested by routine pathological examination of axillary nodes. The likelihood of detecting a micrometastasis during routine single section of an axillary node is about 6 per cent; this increases to 36 per cent if six equally spaced sections are taken through each lymph node.

The work involved in detecting micrometastases is extensive, and the question still remains as to their overall prognostic importance. It is likely that the presence of a single micrometastasis implies a slightly worse prognosis than for node-negative disease, but improved survival compared with patients with a single pathologically obvious macrometastasis. This prognostic importance, however, is only applicable after at least 10 years; in practice the prognosis for patients with a single micrometastasis can be regarded as similar to those with node-negative disease.

In contradistinction to micrometastases, Haagensen and Stout down-staged their patients if nodal metastases were greater than 2.5 cm in diameter. Spread of metastatic disease outside the confines of the axillary nodes into surrounding fat is also a bad prognostic sign.

Internal mammary nodal spread

The internal mammary chain of lymph nodes lies at the anterior end of the intercostal spaces adjacent to the internal thoracic artery. Metastases in the internal mammary nodes are more commonly associated with medial or periareolar tumours and the overall incidence of such involvement is as high as 20 per cent. Disease affecting the internal mammary lymph nodes is rare in the absence of axillary nodal spread; only 8 per cent of patients have such disease if the axilla is clear.

Metastatic disease in the internal mammary nodes alone has the same prognostic implication as axillary nodal disease. However, if both the internal mammary and axillary nodes are affected the outlook is very poor, with a 25 per cent 10-year survival rate.

Because of the poor prognosis of patients with disease of the internal mammary nodes, various attempts at treatment have been adopted, including both surgical excision (in extended radical mastectomy) and radiation therapy. In some studies this aggressive approach has been associated with an improvement in survival of some subgroups of women with such metastatic disease. However, the routine excision of internal mammary nodes in all patients with breast cancer has not been associated with an overall improvement in survival.

It is unclear whether intramammary nodal disease, like supraclavicular nodal spread, should be regarded as regional or metastatic disease.

Supraclavicular nodes

Metastatic disease in the supraclavicular nodes implies extensive involvement of the internal mammary or axillary nodes. The frequent association with widespread metastatic disease means that supraclavicular nodal disease is associated with a poor prognosis.

Spread to distant sites

No site is immune to spread of the tumour, either at the time of presentation or later in the course of the disease. The most commonly affected sites are the bone, liver,

and lung, but metastases to the brain, skin, and peritoneum are by no means uncommon. Peritoneal spread is particularly associated with lobular carcinoma.

Micrometastases in the bone marrow have been recognized for many years in patients with metastatic breast cancer. It is now becoming increasingly clear that such spread also occurs in up to one-third of patients with early, operable, breast cancer. The presence of micrometastases in bone marrow may correlate with other indicators of poor prognosis such as axillary lymph node metastases, tumour size, and oestrogen receptor status, but there is no clear evidence that they indicate a poor prognosis. In short, their relevance is unclear although there is increasing evidence that bone marrow micrometastases at presentation are associated with a worsened prognosis.

The clinical presentation of breast cancer

The majority of women presenting with breast cancer complain of a lump. Classically this is hard, painless, immobile, and demonstrates a degree of fixity to surrounding tissues, overlying skin, or the underlying pectoral muscle. The lump may be visible on inspection and may cause distortion of the breast on elevation of the arms or when tensing the pectoral muscles. Many tumours fail to demonstrate these classic characteristics, however. Not all breast cancers are hard: they are often described as being only firm. The degree of mobility can also vary: small tubular cancers can have a degree of mobility almost characteristic of a fibroadenoma. Finally, not all cancers are painless and occasionally patients describe the disconcerting symptoms of increased discomfort in the lump prior to menstruation. It is, therefore, advisable to undertake histological and cytological examination of any discrete lump in a woman over the age of 25. A diagnosis of a fibroadenoma or ‘cystic change' without such confirmation is extremely dangerous.

Up to 15 per cent of women with breast cancer present with a more diffuse process within the breast, in which a lump is not necessarily the presenting feature. This is particularly common in younger patients with lobular carcinoma. The diffuse nature of the tumour produces distortion, puckering, and the eventual feeling of heaviness as the tumour extends through the breast. Later there may be nipple retraction and discomfort. A similar process occurs in patients who describe a thickening in the breast. Such symptoms can be difficult to differentiate clinically from those of benign breast disease.

Changes in the skin may be the sole presenting symptom; alternatively, they may be associated with other features of the cancer. Puckering may be pronounced and is often associated with the rather scirrhotic tumours of the elderly. However, this clinical sign is commonly observed in younger patients. *Peu d'orange* is a feature of advanced cancer. The cause of this characteristic clinical sign is oedema of the skin: this is not due to direct infiltration of the skin by tumour but represents lymphatic obstruction of the breast as a result of axillary metastasis. It must be differentiated from inflammatory breast cancer, in which cutaneous lymphatics contain tumour emboli. In advanced, untreated cases the skin may be broken, and tumour ulcerates through the skin, with associated haemorrhage and odour. Surrounding skin nodules are not uncommon in seriously neglected cases.

Presentations with changes at the nipple are not uncommon, but these primary tumours may be difficult to diagnose as they are often small and may be easily missed by mammography. Nipple distortion and inversion are not uncommon presenting features in patients with breast cancer. A unifocal or bloodstained nipple discharge is also indicative of malignant disease, often an intraductal carcinoma developing within the major duct system beneath the nipple. The discharge associated with intraductal cancer is characteristically watery and bloodstained, in contrast to the milk discharge of galactorrhoea and multifocal, tenacious, and coloured discharge of duct ectasia.

Paget's disease (Fig. 10) is usually a presenting feature of breast cancer in the elderly, although it is occasionally seen in other age groups. It usually represents an underlying intraductal carcinoma that may be quite extensive in the breast and which may also exhibit an invasive component. The origin of the characteristic, large, clear Paget cells remains an enigma. Whether they represent migration of tumour cells from the primary cancer or originate from the skin of the nipple itself remains unclear.



Fig. 10. Paget's disease of the nipple. The nipple is replaced by the classic exudative Paget's disease. (By courtesy of Dr D. Tarin.)

Finally, it should be remembered that breast cancer frequently presents with metastatic disease. This may be either to regional lymph nodes, such as those in the axilla or supraclavicular fossa, or to distant sites such as bone, lung, liver, or brain.

The diagnosis of breast cancer

There is no indication for a purely clinical diagnosis of breast cancer. The medicolegal consequences of an unnecessary mastectomy are such that there should be no doubt as to the diagnosis before embarking on definitive surgery.

Confirmatory diagnosis is by pathology and radiology, the former being by far the most important.

Pathological diagnosis

Fine-needle aspiration cytology

This method has the advantages of being performed as an outpatient procedure and producing almost immediate results. The technique of aspiration is not difficult after an initial learning period, and it is relatively atraumatic. Its disadvantage is that it requires expert and specialized pathological interpretation. The false-negative rate is about 5 per cent, and false-positive results occur in about 2 per cent of cases. False-negative results are due either to sampling errors as a result of missing the tumour at the time of aspiration or from failing to aspirate a particularly scirrhous acellular carcinoma. False positives are uncommon but may be due to confusion with a hypercellular fibroadenoma or the effects of undisclosed hormone therapy, pregnancy, or lactation on normal breast tissue.

Aspiration cytology cannot differentiate between *in situ* or invasive cancer but can distinguish ductal from lobular carcinoma (Fig. 11).

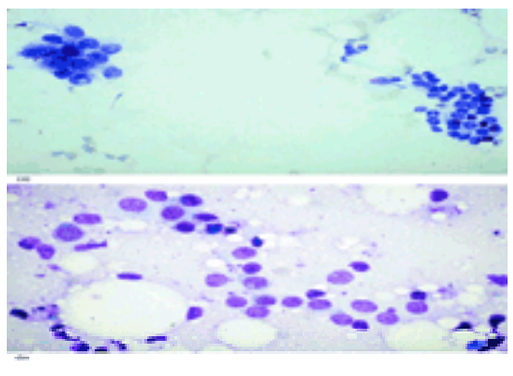


Fig. 11. Fine-needle aspiration cytology of breast cancer. (a) Classic invasive ductal cancer; the cells are large, heterogenous, and have irregular, dark-staining nuclei.

A group of normal cells is also present. (b) The onotonous appearance of tumour cells typical of lobular carcinoma. (By courtesy of Dr I. Buley.)

Core biopsy

A variety of instruments can be used to provide a core of tissue as an outpatient procedure; the recent development of spring-loaded ‘biopsy guns' has increased enthusiasm for the technique such that it has replaced aspiration cytology in many departments. This technique has to be performed under local anaesthetic, but it has the advantage that it produces a histological rather than a cytological specimen. Thus, *in situ* can be differentiated from invasive disease and ductal distinguished from lobular carcinoma. Grading and identification of oestrogen receptor is also possible, although sampling errors can occur with small specimens.

Open surgical biopsy

Biopsy is required in patients who clinically are suspected to have a cancer yet in whom fine-needle aspiration cytology or core biopsy fails to demonstrate malignant disease. It has the disadvantage of requiring hospital admission, although the majority of patients can be treated and discharged the same day. Its advantage is that it provides a definitive method of proving or excluding malignant disease. Open biopsy can occasionally be performed under local anaesthetic, but it is performed more easily under general anaesthesia.

Open surgical biopsy and frozen section

Frozen section evaluation of an excised specimen at the time of definitive surgery has become less common with the more widespread use of fine-needle aspiration cytology. There is a false-positive rate of about 1 to 2 per cent: sclerosing adenosis can cause particular confusion with tumour.

A further reason for the decline in popularity of frozen section techniques is the very reasonable change in the surgical approach to women with breast cancer. Modern practice should avoid the outmoded approach of performing mastectomy on the basis of frozen section while the patient is anaesthetized and unable to discuss various treatment options determined by pathological findings.

Mammography

The last decade has witnessed a huge improvement in imaging methods for breast disease, and a high level of sensitivity and specificity is now available using modern mammographic techniques. The classic features of breast cancer on mammography are tissue asymmetry, mass effect, microcalcification, skin thickening, and nipple inversion (Fig. 12). One or more of these features may be present, the most reliable being a combination of mass effect with localized microcalcification.

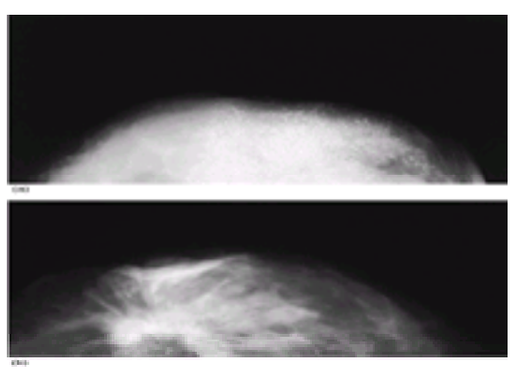


Fig. 12. Mammographic features of breast cancer. (a) Widespread microcalcification indicative of ductal carcinoma *in situ.* (b) A mass effect associated with distortion, skin thickening, and retraction. This is a classic feature of invasive ductal cancer. (By courtesy of Dr B. Shepstone.)

Other methods of imaging such as ultrasonography, xerography, and thermography have also been successfully used in the diagnosis of breast cancer. However, for symptomatic masses which are amenable to fine-needle aspiration cytology, good-quality mammography alone should suffice as a diagnostic method.

Preoperative mammography is not only of diagnostic importance but also detects potential multicentricity and acts as a baseline for radiological evaluation of the breast in the years after initial diagnosis.

Evaluation of patients with breast cancer

Once the diagnosis of breast cancer has been made, certain investigations are of value in subsequent management. All patients with breast cancer should undergo chest radiography and routine evaluation of haemoglobin, full blood count, electrolytes, and liver function tests. The value of other more sophisticated tests is open to doubt. There has been a vogue for performing bone, liver, and even brain scanning for the detection of asymptomatic metastases in all patients who were felt to have early operable disease. In practice, however, there is a low incidence of true positive scans in ‘operable' cancer. The positive yield for bone scanning in patients with apparently localized disease is about 2 per cent, and approximately 50 per cent of these cases will be false positives as a result of previous trauma or coexisting arthritis. There is an even lower positive yield with liver and brain scanning, some studies demonstrating a 100 per cent incidence of negative results following such investigations.

Although there is some disagreement over the usefulness of routine preoperative scanning, the overall incidence of positive results is low. However, patients with clinical stage III disease and those with symptoms suggestive of metastases should undergo further investigation. Brain scanning is of little value in patients with apparently operable breast cancer and the current recommendation is that if screening investigations are to be performed they should be limited to bone scintigraphy and ultrasonography of the liver.

Staging

Staging relates to the classification of breast cancer according to the anatomical extent of disease, each stage serving to aggregate cases having an approximately similar prognosis.

In essence, staging is somewhat simplistic as every variable in breast cancer is classified under a few simple categories. There is therefore loss of precision and wide variation within the stages. There is also the question of relating clinical to pathological staging. As discussed above there is a 30 per cent error rate in clinical evaluation of axillary nodes and the assessment of tumour size is similarly inaccurate. Pathological assessment of nodes and tumour size allows a greater accuracy of staging. However, since pathological staging can only be undertaken after surgical treatment it has obvious limitations in determining management.

Many staging systems have been proposed; none has been shown to be significantly better than others, although some have the advantage of simplicity. More commonly used systems are described below.

The Manchester system (1940)

Stage I. Tumour confined to breast. Any skin involvement covers an area less than the size of the tumour.

Stage II. Tumour confined to breast. Palpable, mobile axillary nodes.

Stage III. Tumour extends beyond the breast tissue because of skin fixation in an area greater than the size of the tumour or because of ulceration. Tumour fixity

underlying fascia.

Stage IV. Fixed axillary nodes, supraclavicular nodal involvement, satellite nodules or distant metastases.

The TMN classification (UICC)

In 1954 the International Union against Cancer (Union Internationale Contre Cancere) attempted to classify breast cancer with a description based on the primary tumour (T), the regional lymph nodes (N), and distant metastases (M). This system has been modified on a number of occasions and has led to confusion and recent criticism. A simple modification is as follows.

T0—no evidence of primary tumour.

T1—tumour less than 2 cm in diameter.

T2—tumour 2 to 5 cm in diameter.

T3—tumour greater than 5 cm in diameter.

T4—tumour fixation to chest wall or skin.

N0—axillary nodes not considered to contain tumour.

N1—mobile involved axillary nodes.

N2—fixed axillary nodes.

N3—supraclavicular lymph node involvement or arm swelling.

M0—no distant metastases.

M1—distant metastases present.

The designation ‘X' after T, N, or M indicates inability to assess the clinical status of the tumour, nodes, or metastases. The correlation of the TNM system with stage is summarized in Fig. 13. This system is universal in current literature.



Fig. 13. A system of combining T, M, and N staging subsets with clinical stage.

The Columbia classification (Haagensen, Cooley, and Stout 1943, 1969)

Stage A—no skin oedema, ulceration, or fixation to chest wall; axillary nodes not clinically involved.

Stage B—clinically involved axillary nodes less than 2.5 cm in diameter and not fixed.

Stage C—grave signs of comparatively advanced carcinoma: oedema of skin, skin ulceration, fixation to chest wall, massive axillary involvement with nodes greater than 2.5 cm in diameter, and axillary fixation.

Stage D—advanced carcinoma including two or more signs in Stage C, and in addition satellite skin nodules, supraclavicular nodes, inflammatory cancer, arm oedema, or distant metastases.

Treatment of carcinoma of the breast

There has been a massive change in the treatment of breast cancer over the past 100 years. The past 20 years have witnessed fundamental changes in our approach, with decreasing reliance on radical surgical excision of the primary tumour and associated regional lymph nodes. This change of emphasis has resulted from a fundamental alteration in our philosophy of the spread of breast cancer, a greater appreciation of the systemic aspects of the disease, and increased realization that variations in local regional treatment are unlikely to alter prognosis.

Theories relating to the spread of breast cancer

Early concepts regarding the spread of breast cancer were proposed by Halsted. His opinion dictated the treatment of breast cancer for almost a century. He believed that breast cancer spread by direct permeation rather than embolization to the regional nodes. The nodes acted as a barrier to the further spread of cancer until they themselves were overladen with tumour, whereupon bloodborne metastases occurred. Such a concept of the spread of disease was the basis for performing *en bloc* radical resection of the entire region in radical mastectomy and for the even more radical operations which followed that of Halsted.

A major flaw in Halsted's theory was, of course, that many deaths from metastatic disease occurred after radical surgery for small localized cancers. It thus became apparent that embolization, rather than permeation, of tumour cells is the predominant mode of simultaneous spread to both regional nodes and distant sites. It is now widely believed that spread from tumour to lymph nodes and distant sites represents two independent but correlated processes. According to this alternative theory of breast cancer, spread to axillary nodes represents the risk of metastatic disease rather than its determinant. This theory also suggests the existence of an inherent systemic component to breast cancer and implies a limitation of local regional treatment alone on prognosis. It also suggests that some patients may well have systemic metastases at the time of clinical presentation even in the absence of axillary nodal metastasis.

The pros and cons of these two conflicting theories on breast cancer have been fiercely argued. Support for Halsted's theory is based on the fact that a significant proportion of patients with disease affecting axillary nodes are apparently cured after radical surgery. Furthermore, there is generally an orderly spread of involvement up the axillary lymph nodes, skip metastases being relatively uncommon, and there is little evidence for systemic spread accompanying nodal involvement. Free circulating tumour cells have not been demonstrated and imaging techniques demonstrate subclinical metastases in only a very small proportion of patients with apparently operable breast cancer.

The advocates of the alternative theory of breast cancer base their argument on both clinical and experimental evidence. There is no evidence to suggest that treatment of disease in axillary nodes is of any significance with respect to survival, although it does substantially improve the local control of tumour in that area.

Further evidence for the alternative theory of breast cancer comes from experimental studies demonstrating failure of lymph nodes to filter tumour cells and from the existence of direct communications between afferent lymphatics and efferent veins around lymph nodes. Finally, studies of the natural history of breast cancer indicate that many tumours exist for several years before they become clinically apparent. Thus, there is ample opportunity for metastases to develop during this protracted period of occult growth. It has therefore been implied by some authorities that clinically apparent invasive breast cancer is never really curable.

The alternative theories of breast cancer are summarized in Table 5.

Halstedian	*Alternative*
1. Tumours spread in orderly manner based on mechanical considerations	1. There is no orderly pattern of tumour dissemination
2. Tumour cells traverse lymphatics to nodes by direct extension	2. Tumour cells traverse lymphatics by embolization
3. Positive lymph nodes indicate tumour spread and are the instigator of disease	3. Positive nodes indicate a tumour/host relationship which permits development of metastases rather than being their instigator
4. Regional nodes are barriers to passage of tumour cells	4. Regional lymph nodes are ineffective barriers to passage of tumour cells
5. The bloodstream is of little significance in tumour dissemination	5. The bloodstream is of considerable significance in tumour dissemination
6. The tumour is autonomous of the host	6. The host and tumour have a complex relationship
7. Operable breast cancer is a local-regional disease	7. Operable breast cancer is often a systemic disease
8. The extent of surgical operation determines patient outcome	8. Variations in local-regional treatment do not substantially affect survival

Table 5 Opposing theories of breast cancer

The management of cancer of the breast

The management of breast cancer requires a complex multidisciplinary approach involving surgeons, radiotherapists, medical oncologists, and pathologists, as well as other professionals such as counsellors and breast care nurses. For the sake of convenience the treatment of breast cancer will be discussed in three sections: the management of the breast itself, the axillary nodes, and potential micrometastases. However, it must be stressed that these three approaches have a final common aim—the successful treatment of patients with breast cancer. A properly co-ordinated approach by the various health professionals is required if optimal results are to be achieved.

Management of the breast

The importance of the management of the breast itself is to provide local tumour control and reduce any potential for metastasis to either local or distant sites. If this can be achieved using a surgical technique which allows conservation of the breast there is also the added bonus of improved cosmesis and, hopefully, a reduction in psychological trauma.

The traditional surgical treatment for breast cancer involved total removal of the breast. The various types of mastectomy describe those structures resected in association with the entire breast (Table 6). Up until the early 1970s breast cancer was treated by total mastectomy, in one or other of the forms described in Table 6. The most commonly performed operation was the Halsted radical mastectomy, although in Europe there had been a vogue for simple mastectomy with adjuvant radiotherapy since the early 1950s.

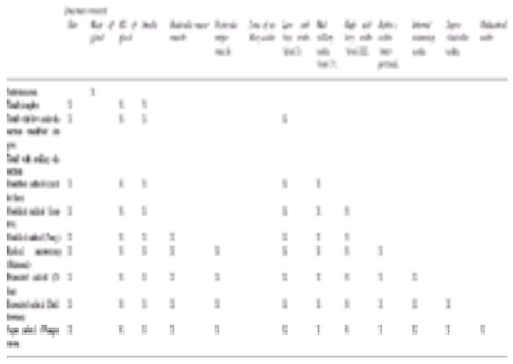


Table 6 Anatomical structures removed in various types of mastectomy

The second half of the twentieth century witnessed increasing disillusionment with radical and mutilating forms of surgery for breast cancer. As a result the trend towards breast conservation has increased since the mid-1960s, although a number of centres had adopted this approach since before the Second World War. Again there are a number of different descriptions relating to breast conservation which have caused confusion. Tumourectomy, lumpectomy, tylectomy, segmental mastectomy, and quadrantectomy are all synonymous with a therapeutic procedure in which the primary tumour is removed and the breast is preserved. Unfortunately, these terms are not precisely defined, although they imply the removal of varying amounts of normal breast tissue in association with a primary tumour. The terms ‘lumpectomy', ‘tumourectomy', and ‘tylectomy' imply removal of the tumour with a minimal or no margin of normal breast tissue around it. Segmental mastectomy implies excision of the tumour with a rim of associated normal breast tissue. However, this term is also somewhat misleading as it implies that the breast is anatomically a segmental organ and that tumours occur in a localized segment. This is clearly not the case. The term ‘quadrantectomy' denotes removal of a breast quadrant, and implies wider excision of normal breast tissue than segmental mastectomy. In practice, however, there is little distinction between these terms and although a number of authorities have recommended the adoption of a uniform nomenclature, none has found universal favour.

Once the questions regarding definition of terms and nomenclature have been addressed the simple, yet fundamentally important, question which remains is whether breast conservation provides results as reliable in the treatment of breast cancer as total mastectomy. Furthermore, is there an additional benefit in terms of cosmetic and emotional adjustment? Finally, if breast conservation is justified, in which patients is this appropriate?

In terms of management of the breast the simplest approach would be to remove the tumour itself, preferably with a margin of normal tissue around it. In theory the more limited procedures of tumourectomy or lumpectomy are likely to be followed by a good cosmetic result but are more likely to be followed by local recurrence because of the likelihood of failure to excise the tumour completely. More extensive forms of conservative surgery such as quadrantectomy are more likely to provide good tumour control but are more liable to be followed by a less satisfactory cosmetic result because of the amount of breast tissue excised.

Studies evaluating simple excision of the tumour without adjuvant radiotherapy have produced somewhat disappointing results, with a local recurrence rate of approximately 30 per cent within 7 years. Local recurrence may result from positive margins of excision as a result of inadequate surgery or from development of further tumours either *de novo* or from pre-existing multifocal disease.

Failure of local excision alone to provide adequate tumour control in the breast led to the use of adjuvant radiotherapy to the remaining breast tissue. This combination of conservative surgery with adjuvant radiotherapy has now become a standard procedure for patients with breast cancer. A large number of studies have provided retrospective analysis of conservative treatment with surgery and irradiation for the treatment of this disease. The results have been generally encouraging with local recurrence rates of under 10 per cent, although in some centres 20 per cent of patients have developed locally recurrent disease after periods of 7 years of more.

Unfortunately, as the majority of studies relating to conservative surgery with adjuvant radiotherapy are retrospective, there are doubts about case selection. A more accurate appraisal can only be obtained from analysis of prospective randomized trials. There have been a number of studies comparing conservative treatment with mastectomy from both Europe and the United States. Three such studies demand detailed discussion.

The Guy's Hospital studies

In the first study on breast conservation Atkins and his colleagues at Guy's Hospital, London, randomized patients over the age of 50 years with clinical stage I or II disease to either extended tylectomy or classic radical mastectomy. Both operations were followed by axillary, supraclavicular, and internal mammary irradiation and the patients treated by tylectomy also underwent radiation treatment to the residual breast tissue. Some patients received adjuvant thiotepa. Unfortunately, the dose of radiation used in this study was somewhat suboptimal.

After 10 years there was a significantly higher local recurrence rate in patients treated by breast conservation for both stage I and II disease, although this had no impact on survival for patients with stage I disease. Patients with stage II disease treated by breast conservation had a worse prognosis.

A follow-up study from the same institution by Haywood and his colleagues randomized a further 258 patients with clinical stage I to the same regimens, but again with suboptimal radiation doses. On this occasion conservative treatment resulted not only in relatively poor local control but also decreased survival. Thus, the Guy's group provided inconsistent results following breast conservation, even in patients with stage I disease.

The Milan studies

A second important randomized study was that of Veronesi and his colleagues from Milan. A total of 701 women with tumours less than 2 cm in diameter and without palpable axillary nodes were randomized to either radical mastectomy or quadrantectomy, axillary dissection, and breast irradiation. Adjuvant chemotherapy was given to those patients with histologically proved lymph node metastases. Overall 15-year survival was 66 per cent in each group. Local recurrence rates were similar in patients treated by radical mastectomy or breast conservation, being 3 and 6 per cent, respectively.

In a second study the same group randomized 705 patients to quadrantectomy or tumourectomy. All patients had radiotherapy to the breast and axillary clearance. Survival after 79 months was similar at 85 per cent in both groups. However, local recurrence occurred in 20 per cent of patients treated by the lesser procedure of tumourectomy compared with only 5 per cent after quadrantectomy.

The NSABP (protocol B–06) study

The third study evaluating breast conservation in the treatment of breast cancer was the protocol B–06 conducted by the National Surgical Adjuvant Breast Project (**NSABP**). In this trial 1843 women were randomized to mastectomy or to segmental resection (lumpectomy), with or without radiotherapy. All patients underwent axillary clearance, and those found to have positive nodes were given adjuvant chemotherapy. Patients found to have positive margins after breast conservation were converted to the mastectomy group—perhaps introducing some bias into the study.

After 8 years there was no significant difference in survival between patients treated with breast conservation and those with mastectomy. However, those patients treated by lumpectomy without radiotherapy suffered a greater incidence of local recurrence in the ipsilateral breast.

Thus, two of these randomized studies clearly demonstrated that patients suitable for breast conservation have an equally good prognosis when treated by a conservative procedure with adjuvant radiotherapy rather than mastectomy. All the above studies may be criticized in that case selection was still liable to exclude patients with more aggressive tumours, who may not have been selected for randomization. However, the other major criticism of these studies—that there has been inadequate follow-up—can now be answered, in that assessment after 8 to 15 years is now available.

Indications for breast conservation treatment

Patient choice is important. If a patient prefers to be treated by total mastectomy rather than breast conservation then, after appropriate counselling, her wishes should be adhered to. It is widely believed that women with breast cancer prefer to be treated by partial mastectomy. This is often, but not always, the case: many women who actually have the disease feel more secure after a total mastectomy, and other prefer to avoid radiotherapy if at all possible. Previous studies from the United Kingdom showed that 50 per cent of patients suffering from breast cancer actually preferred the concept of total mastectomy, although there is now much greater acceptance of breast conservation.

The other main indications for conservative therapy relate to tumour characteristics, the position of the cancer in the breast, and the nature of the breasts themselves. Tumour size is of paramount importance; most authorities recommend mastectomy if the tumour is more than 3 or 4 cm in diameter. The NSABP recommends partial mastectomy only if the tumour is less than 4 cm in diameter and when there is no fixation to the underlying muscle or overlying skin. Other surgeons pursue a more conservative approach and prefer to reserve breast conservation for tumours of less than 3 cm in diameter. Much will, of course, depend on the size of the breast. In small breasts a tumour 4 cm in diameter may require mastectomy, and conservation with good cosmetic results may only be a reality if the tumour is 2 cm or less in size. On the other hand, breast conservation may be permitted for tumours greater than 4 cm in diameter if the breasts are large, although it must be remembered that heavy pendulous breasts may provide poor cosmetic results after conservation because of the oedema and fibrosis which may follow irradiation.

A further contraindication to breast conservation is multicentricity. Patients with a second tumour in the breast, whether it is clinically apparent or seen on mammography, should be considered for mastectomy if the tumours cannot be excised in continuity. Mammography is therefore an essential preoperative investigation to determine the presence of multicentric tumours.

Multifocality, as opposed to multicentricity, is present in up to 50 per cent of patients with breast cancer. Its extent depends on tumour size and it is this multifocality which may result in the unacceptable incidence of local recurrence if breast conservation is performed without adjuvant radiotherapy. It is yet to be determined whether adjuvant radiotherapy truly sterilizes multifocal disease or whether it simply delays recurrence to the very long term (10 to 20 years). Widespread multifocal disease certainly relates to local recurrence—its presence is a relative contraindication to breast conservation.

The question of multifocality is particularly important in patients with lobular carcinoma. It may be felt that this potentially multifocal and multicentric disease should be treated by mastectomy. However, as long as criteria relating to tumour size and type are adhered to there is no evidence that this particular type of tumour is any more likely to recur than is ductal carcinoma.

Centrally located tumours are a relative contraindication to breast conservation. Such tumours are often quite diffuse and they can be difficult to remove by segmental mastectomy. Furthermore, the end cosmetic result is often disappointing as many women perceive the nipple as being a fundamentally important part of the breast—both physiologically and psychologically. It is not uncommon for women to hold the view that if the nipple has to be removed then the treatment may as well be mastectomy. Despite this opinion many women are, in fact, pleased with the end cosmetic result, even though the nipple has to be removed in the treatment of a centrally located tumour. This is particularly true for those women with larger breasts, in whom a nipple reconstruction or prosthesis may be used.

Poor tumour differentiation is also a relative contraindication to breast conservation, although many authorities no longer regard it as such. Moreover, a complete assessment of tumour grade is not possible until the cancer has actually been removed. Many poorly differentiated tumours are somewhat ill-defined and this specific clinical feature may be a further relative contraindication to breast conservation.

There have been doubts as to whether breast conservation is appropriate for patients with node-positive disease. As the presence of disease in axillary nodes indicates a higher risk of tumour dissemination, there is, in fact, an argument in favour of breast conservation in these circumstances.

Despite the various contraindications to breast conservation, this type of treatment has now become the management of choice for the majority of patients with breast cancer. However, there is still much variation from centre to centre, with some authorities recommending total mastectomy in up to 70 per cent of cases, whereas others recommend breast conservation for virtually all their patients.

Local recurrence after treatment of breast cancer

Local recurrence after either total mastectomy or breast conservation surgery is a major clinical problem. Locally recurrent disease must be distinguished from regional recurrence; the two are often confused. Local recurrence is that seen in the breast after conservative therapy or on the chest wall after mastectomy; regional recurrence denotes recurrent disease in regional lymph nodes. Local recurrence may be either localized or widespread. Both types, but particularly the latter, may be associated with widespread metastatic disease and it is therefore mandatory to stage all patients with either local or regional recurrence for distant metastases.

In general, the more radical the treatment the lower the incidence of local recurrence either on the chest wall after mastectomy or in the remaining breast tissue

following breast conservation therapy. Thus, very minimal treatment such as lumpectomy alone is associated with a higher incidence of local recurrence than is partial mastectomy or quadrantectomy, particularly if associated with adjuvant irradiation. Similarly, there is a higher incidence of local recurrence after partial than total mastectomy. The lowest local recurrence rate (2 per cent for stage I disease) has been recorded after extended radical mastectomy.

The overall incidence of local recurrence after breast conservation with adjuvant radiotherapy should be less than 10 per cent, although some authorities have recorded figures in excess of 20 per cent after 7 years or more. Local recurrence after total mastectomy has previously occurred in about 5 to 8 per cent of patients. However, in modern practice this incidence may be even greater because of selection of large, prognostically unfavourable tumours for this type of treatment.

Factors predisposing to local recurrence

Recurrent disease, either in the breast after partial mastectomy or on the chest wall after total mastectomy, is more common in patients with positive axillary nodes. It has therefore been suggested that as well as giving radiotherapy to all patients after partial mastectomy, the chest wall should be irradiated after total mastectomy if disease has spread to the nodes. Meta-analysis, however, has shown no improved survival in patients given adjuvant radiotherapy, although it reduces the local recurrence rate on the chest wall. Adjuvant orthovoltage irradiation after mastectomy is associated with a small increased overall mortality because of an increased probability of myocardial infarction and lung cancer. More recent megavoltage techniques are not, however, associated with this increased probability.

Tumour size has also been shown in some studies to correlate with incidence of local recurrence after both total mastectomy and breast conservation. Some authorities therefore recommend adjuvant irradiation after both partial and total mastectomy if the primary tumour is large. Again, there is little evidence that therapy improves survival, although it does reduce local recurrence.

Positive surgical margins following partial mastectomy are an important risk factor for local recurrence. Although positive surgical margins may relate to inadequate surgical technique, they are also likely to occur in patients with extensive *in situ* disease. It therefore follows that widespread multifocality is an important factor predisposing to local recurrence.

Other factors such as high tumour grade, lymphatic or vascular invasion, and young age are also associated with recurrent disease.

Treatment of local recurrence after breast conservation

Local recurrence after breast conservation therapy can be treated by salvage mastectomy. Such treatment has been claimed to produce a 5-year survival rate of up to 50 per cent. This may well be so for those patients with localized recurrence and those with a long disease-free interval since initial treatment. Further breast conservation is sometimes feasible in these women. Unfortunately, however, up to 20 per cent of those who develop local recurrence in the residual breast tissue after breast conservation have advanced or unresectable disease. This is because local recurrence after breast conservation with adjuvant radiotherapy can occur as a ‘field-change' throughout the irradiation field. Positive surgical margins may occur despite extensive surgical resection, musculocutaneous flaps, and skin grafts. Further local recurrence is not uncommon in these individuals, who face terminal illness with severe physical and psychological morbidity from uncontrolled and prolonged chest wall recurrence. Systemic treatment with tamoxifen or chemotherapy may alleviate symptoms but responses are often quite short lived when recurrence is widespread.

Patients who develop local recurrence within 2 years of breast conservation are liable to experience a rapid metastatic relapse indicating the inherent biological aggressiveness of the primary tumour.

Treatment of local recurrence after total mastectomy

Local recurrence after total mastectomy can present a difficult problem. Small focal areas of local recurrence can be surgically excised. If irradiation has not been previously used, radiotherapy may prevent or delay further local recurrences. Similarly, adjuvant drug treatment with either chemotherapy or endocrine therapy may help reduce the risk of further recurrence.

Local recurrence after total mastectomy may present as a widespread area of recurrent disease. This is particularly difficult to treat. Surgical excision with associated transposed musculocutaneous flaps or skin grafting is often associated with positive margins, further local recurrence, and the rapid development of metastatic disease. Radiotherapy may be attempted if this has not already been used; chemotherapy or endocrine treatment may produce some regression of this somewhat remorseless process.

Reconstruction of the breast after total mastectomy

All women undergoing total mastectomy should be offered reconstruction, even if they have advanced disease. Although it was hoped that this would reduce the psychological sequelae of breast cancer, this is not always so, but reconstruction can improve an otherwise miserable existence for some women. The increasing trend towards breast conservation should reduce the number of women requiring reconstruction after total mastectomy; if reconstruction is required after breast conservation therapy, there must be doubts as to whether the original surgical operation was appropriate.

The two main questions relating to breast reconstruction are the timing and the technique. Breast reconstruction can be performed either at the time of initial surgery or delayed for, say, 1 to 2 years after diagnosis. The obvious advantage of immediate reconstruction is that women do not undergo a period of loss of the breast. Furthermore, only one surgical procedure is required. Previous arguments against immediate reconstruction included worries about unfavourable cosmetic comparisons against the original breast and doubts about the ability to detect local recurrence, but these have not been realized. Immediate reconstruction should therefore now be regarded as standard treatment following mastectomy in women desiring reconstructive surgery.

Methods of reconstruction

There are numerous methods of reconstruction: no single method is perfect. Appropriate counselling is of particular importance before embarking on any reconstructive surgery. The patient should see photographs of both good and bad results to ensure a realistic approach. If possible she should talk to women who have previously undergone reconstruction. It is essential to stress that although reconstruction may provide a satisfactory cosmetic appearance, the reconstructed breast will not have sensation or mobility like the original. Further counselling is required about nipple reconstruction, as this aspect is often forgotten.

The reader is advised to consult specialized texts on plastic surgery for methods of reconstruction. However, for the sake of completeness some of the more widely used techniques are briefly discussed.

Reconstruction with subcutaneous or subpectoral implant

This is the easiest technique to perform, although skin necrosis and implant extrusion may occur, especially after irradiation and if a subcutaneous position is used. Most surgeons prefer to place the implant in a subpectoral position; the subcutaneous technique is mentioned only to be condemned as it frequently causes skin necrosis, extrusion, and a poor cosmetic result (Fig. 14).

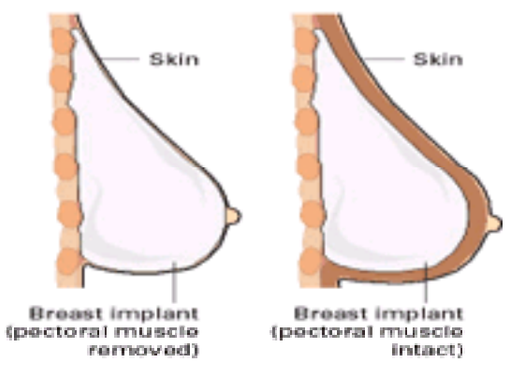


Fig. 14. Subcutaneous and subpectoral implants.

When inserting the implant subpectorally, care must be taken to ensure that the prosthesis is in an appropriate position. Since there is a tendency for such implants to migrate superiorly, it is necessary to divide the tight fascial bands between the insertion of the pectoralis muscle and the rectus abdominus to allow appropriate positioning.

Two main cosmetic problems occur with this technique. The first is capsule formation, which is unsightly and also imparts a very hard texture to the breast. Capsule formation is less common if ‘textured' implants are used. If excessive, the capsule can be surgically fractured or excised to improve appearance and to soften the reconstructed breast.

The second cosmetic problem with simple implants is lack of ptosis which occurs with the normal breast, especially with increasing age. The use of tissue expansion techniques can, to some extent, overcome this problem (see below), especially if performed in conjunction with one of the various plastic surgical procedures which enhance ptosis.

Most implants in current use are made of silicone gel. This material has the advantages that it is inert and has a relatively natural feel. Its use in the United States has, however, been curtailed because of worries that it may be associated with autoimmune disorders—especially if leakage should occur. However, a recent government enquiry in the United Kingdom into these risks found no clear evidence of their existence and condoned the continued use of silicone implants.

The alternative type of implant using saline rather than silicone gives a rather soft, striated, and ‘flabby' result. Furthermore, leakage of saline can occur in the long term.

A major difficulty relating to this and other types of reconstruction is the fact that a nipple has to be recreated (see below). This problem can be avoided by performing a subcutaneous mastectomy, although many oncologists feel this is an inappropriate cancer operation because 2 to 10 per cent of the original breast tissue may be left behind following this technique.

Tissue expansion

Simple implants have limitations with respect to lack of natural ptosis and complications in irradiated skin. Both these problems can, to some extent, be overcome by gradual tissue expansion. Traditionally, this has been achieved by a tissue expander followed by insertion of a permanent prosthesis. The introduction of a combined tissue expander/permanent prosthesis has greatly simplified matters. The Becker expander/mammary prosthesis consists of an outer lumen filled with silicone gel and an inner chamber connected to a removable, self-sealing, side port (Fig. 15). This inner chamber is filled with saline and is gradually expanded to the desired volume over a number of weeks. The expansion eventually provided is substantially greater than the volume of the opposite breast. Once expansion is complete the saline chamber is aspirated to allow creation of a reconstructed breast of similar volume to that on the opposite side, while at the same time providing some ptosis. The self-sealing injection port can then be removed under local anaesthetic.

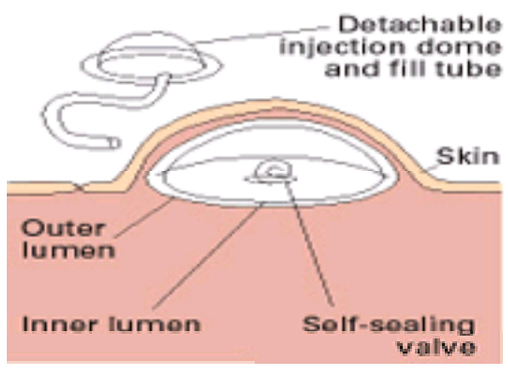


Fig. 15. The Becker combined expander/mammary prosthesis.

Pedicled and free myocutaneous flaps

The most commonly adopted technique has been the pedicled latissimus dorsi myocutaneous flap, because of its proximity, good blood supply, and ease of creation (Fig. 16). The bulk of muscle transposed, however, is often quite small and the technique frequently has to be combined with an implant. Some authorities report good results after immediate reconstruction with this technique, although it leaves a scar on the back, is associated with skin mismatch, and is often complicated by seroma formation. A modification of this technique is the latissimus dorsi miniflap. This utilizes the latissimus alone, without overlying skin, on its neurovascular pedicle. Its tendinous insertion into the humeral head is divided to allow mobility for the muscle bulk to be used to fill a defect in the lateral part of the breast after partial mastectomy.

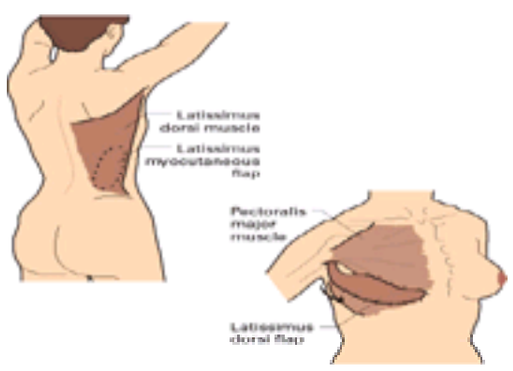


Fig. 16. The latissimus dorsi pedicled flap. The myocutaneous flap is swung on its axis anteriorly preserving its blood supply. The resulting flap can be sewed to cover a defect after mastectomy for advanced disease or may be employed (in association with an implant) for reconstruction.

A more recent innovation has been the rectus abdominus myocutaneous flap. A longitudinal or transverse incision can be used. In the former a pedicle graft is swung on its axis to provide tissue bulk on the chest wall. Most authorities now prefer to use a transverse rectus abdominus flap in which a horizontal, elliptical, subumbilical incision is used for access (Fig. 17). Because of the rather precarious blood supply to this region, many surgeons now recommend a free transverse rectus abdominus flap with microvascular anastomosis between epigastric and thoracodorsal vessels. There is a graft failure rate of up to 8 per cent with this technique.

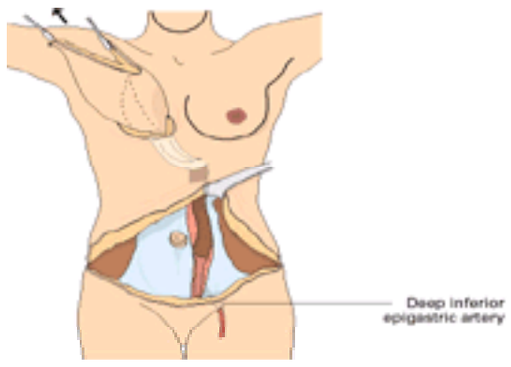


Fig. 17. The transverse rectus abdominus pedicled flap. The contralateral rectus muscle with associated skin is used to cover a defect on the chest wall. An alternative method adopts a free flap with microvascular anastomosis.

The cosmetic results of the free transverse rectus abdominus flap are such that they should be regarded as the ‘gold standard' by which other methods of reconstruction are judged. Unfortunately, it is a time-consuming procedure and requires specific expertise with microvascular anastomosis. It is also relatively contraindicated in patients who are very obese or very thin, in those who have a pre-existing lower abdominal scar, in those who are heavy smokers, and in those who are above age 55. In these circumstances other methods of reconstruction, such as using a gluteal flap, may have to be considered.

All the above forms of reconstruction may require reduction mastopexy in the contralateral breast if equality is to be achieved.

Nipple reconstruction

For many women the nipple is the most important part of the breast, yet the majority of techniques described above fail to provide appropriate nipple reconstruction. Only subcutaneous mastectomy routinely preserves the nipple–areola complex; the disadvantages of this method are described above. Several methods of nipple reconstruction are available, using either any excessive skin on the reconstructed breast or a free graft from a distant site, such as the groin or thigh. An alternative approach is to use an adhesive prosthesis.

A difficulty with nipple reconstruction relates to pigmentation. If the reconstructed nipple is pale in comparison to that on the opposite side then simple tattooing may provide an appropriate effect.

Management of axillary nodes

The management of the axillary nodes is a controversial issue, and practice varies widely from centre to centre. However, the management of axillary nodes should provide certain well-defined end points. First, it should act as a guide to staging and prognosis, although the dominating importance of axillary nodal status has lessened somewhat as other biological markers of disease aggression have been identified. Second, it dictates the need for adjuvant radiotherapy, chemotherapy, or hormonal therapy, although choice of systemic treatment is now based on many factors other than nodal status alone. Third, it must provide good local control of disease in the axilla as untreatable axillary metastases can cause much morbidity.

Despite the importance of axillary nodal status there is, however, no evidence that good therapeutic control of the axilla correlates with survival. As discussed above this is because axillary nodal involvement is a marker of likelihood of systemic spread rather than its determinant.

The major argument relating to management of axillary nodes concerns the extent of their surgical excision and the impact that this may have on staging, prognosis, and the subsequent control of axillary disease. The extent of surgical excision of the axillary contents can range from that of non-intervention to routine block clearance of all nodes in all patients. All approaches have their supporters and detractors.

The policy of non-intervention

Failure to assess axillary nodes has been rightly condemned. This approach has unfortunately been relatively common and has resulted in subsequent inability to prescribe rational local and systemic adjuvant therapy. Radiotherapy then has to be given blindly, resulting in the needless irradiation of disease-free nodes. Similarly, it is impossible to provide a rational programme of systemic hormonal or chemotherapy if nodal status is not known.

Recently, however, there has been a resurgence of this rather negative approach to axillary nodes. This has resulted from appreciation that nodal control fails to correlate with prognosis and the wider use of adjuvant therapy. Supporters of non-intervention claim that if all patients receive systemic adjuvant therapy and careful monitoring, axillary nodal relapse can be treated as it occurs clinically. In fact, such axillary nodal relapse is relatively uncommon—a further point in favour of this approach. Whether this view will become more widespread is unclear, although recent recommendations suggesting wider use of chemotherapy and tamoxifen in patients with breast cancer might support this philosophy as indications for this treatment are no longer based on axillary status alone.

The policy of axillary nodal sampling

Axillary nodal sampling depends on the philosophy of surgically removing lower axillary nodes at level 1 only to determine the presence or absence of metastatic deposits at this site. The technique is dependent on the fact that skip metastases to higher nodes are rare (about 2 per cent) and that patients with positive glands can be successfully treated with adjuvant radiotherapy. Axillary sampling has been widely practised in the United Kingdom although, like the policy of non-intervention, it has been criticized. A number of studies have failed to demonstrate adequate node retrieval using this technique, and staging is claimed not to be as accurate as with clearance. There are also doubts as to whether radiotherapy is as good as clearance in controlling axillary nodal disease.

A major problem relating to sampling is that it is poorly defined. There is no agreement as to the extent or boundaries of axillary surgery when sampling, although it is important to emphasize the necessity of removing enlarged or palpably hard nodes. At least four and if possible as many as 10 glands should be removed to ensure an appropriate sample.

The supporters of axillary node sampling have answered all the above criticisms. First, those studies which failed to demonstrate adequate nodal retrieval used inappropriate anatomical landmarks to define the axillary glands. Thus, the extent of dissection had been demarcated by landmarks such as the axillary tail of the breast or the intercostal–brachial nerve. When the axillary fat has been entered using appropriate anatomical dissection, sampling provides a qualitative assessment of the axilla. It must be admitted, however, that sampling fails to provide as much quantitative prognostic information as a full nodal assessment which results from complete dissection of the axilla.

The supporters of sampling also claim that, if done properly, patients with positive nodes can be specifically targeted for adjuvant radiotherapy, whereas a policy of full axillary dissection in all patients needlessly overtreats the 60 per cent of women with negative nodes. Patients who are node negative after sampling do not benefit from adjuvant radiotherapy. The ability of axillary sampling with selective radiotherapy to provide adequate control has been well demonstrated in a recent randomized trial of this technique compared with clearance by Forrest and his colleagues. In this study there was no difference in the axillary recurrence rates between the two techniques. Indeed, local relapse was relatively rare in the axilla, being less than 5 per cent after either option.

The whole question relating to sampling depends on how well it is done. When inappropriately performed it is rightly condemned. However, if done properly, it provides qualitative data nearly as good as those obtained from clearance, allows patients with positive nodes to be selectively treated with radiotherapy, and is a rational basis for helping to decide on appropriate systemic treatment.

A recent innovation has been sentinel node biopsy. Any single area in the breast has primary lymphatic drainage into one or two regional lymph nodes. These sentinel nodes can be identified by injecting a colloidal dye (isosulphan blue) or a radiolabelled colloid (technetium–sulphur colloid) into the edge of the tumour. Using a protocol appropriate to the dye used, one to four sentinel lymph nodes may be identified and excised for biopsy. If these nodes do not contain tumour metastases, it is highly unlikely that an axillary dissection will identify involved lymph nodes. This preserves the benefits of axillary lymph node clearance for those with involved lymph nodes, and spares the patients with negative nodes from the morbidity of this procedure. Because the procedure depends upon the identification of a limited lymphatic drainage area, it relates to small tumours with no palpably enlarged or suspicious regional lymph nodes. Contraindications to this technique are: large tumours (> 4 cm)

or biopsy cavities, axillary or tail of breast tumours, suspicious lymph nodes, multiple tumours, or very medial lesions without a prior lymphoscintigram to identify the area of primary drainage. If the sentinel lymph node/s can be identified and are free of tumour metastases, the likelihood of other nodes being involved by tumour appears too small to warrant axillary dissection. Clinical trials continue to refine both technique and margin of error. Until a surgeon has demonstrated the ability to identify reliably sentinel nodes that reflect the results of axillary dissection, the dissection should not be abandoned when its results will influence therapy.

The policy of axillary nodal clearance

Many of the criticisms relating to sampling are answered by axillary clearance. Block dissection of axillary nodes accurately stages the axilla, both qualitatively and quantitatively, and has the advantage that it provides a good mechanism for tumour control. Adjuvant radiotherapy is not required after clearance as this combination does not improve local control or prognosis but is associated with unacceptable limb oedema.

The main criticism of clearance is that the 60 per cent of patients with negative nodes are overtreated by this treatment. There are also doubts as to the efficacy of clearance obtained with differing operative techniques. In some centres the pectoralis minor muscle is routinely removed or divided; in others it is merely retracted. This may explain the wide variation in nodal count between centres undertaking axillary clearance. The technical aspects of clearance may be compounded when it has to be performed through a separate axillary incision distant from the primary tumour site: in these circumstances access to the apex of the axilla may be limited.

Critics of clearance also suggest it may be followed by a higher incidence of arm lymphoedema than is sampling. However, as long as axillary clearance is not combined with radiotherapy there is little evidence that arm oedema after this more radical procedure is any worse than following sampling with selective adjuvant radiotherapy—infact, shoulder stiffness may be less of a problem.

Overall, there is little to choose between well-performed axillary sampling, with adjuvant radiotherapy treatment for patients with diseased nodes, and axillary clearance. Axillary clearance has the disadvantage that 60 per cent of patients undergo unnecessary dissection of their axilla, whereas radiotherapy is both time-consuming and expensive postoperatively. In one author's unit (M.G.) reliance is generally placed on axillary sampling followed with adjuvant radiotherapy for node-positive patients. However, if there is bulk nodal disease with metastases greater than 2 cm in diameter, we prefer full surgical axillary clearance.

Treatment of potential micrometastases

Systemic metastatic disease in patients with breast cancer is a serious development, with a median survival after diagnosis of only about 14 months. Such metastases are particularly distressing when they occur in patients originally thought to have a relatively good prognosis by virtue of tumour size and negative axillary nodes. Earlier expectations of improved survival with either more radical surgery or adjuvant radiotherapy have not been realized. The systemic component of breast cancer, as described above, indicates that either the patient is cured by local treatment or that death occurs from metastatic disease at about the same time that she would have died without local intervention. This is because, according to the non-Halstedian theory of breast cancer, metastases are established prior to diagnosis and are outside the scope of even the most extensive local treatment.

The limitations of local treatment provided the initial rationale for the use of adjuvant systemic therapy at the time of, or following, initial surgical treatment of breast cancer. Such an approach was first adopted for breast cancer in 1948, when oophorectomy was suggested as an adjuvant treatment. Early studies produced conflicting results but indicated the need for a wider appraisal of the systemic approach, which gradually developed over a 30-year period as more active chemotherapeutic and hormonal agents became available. As a result of numerous studies, adjuvant systemic therapy is now regarded as an integral party of the primary treatment of breast cancer.

Adjuvant chemotherapy

The first large-scale studies evaluated single agent chemotherapy such as thiotepa, phenylalanine mustard, and cyclophosphamide. After a 12-year study, Nissen-Meyer and his colleagues demonstrated a significant survival advantage in patients receiving a brief perioperative course of cyclophosphamide compared with untreated controls. Other experience with single agents was also encouraging, especially in patients with positive nodes, although there was some scepticism because of stratification within the trials and variations in treatment methods. Subsequent studies have evaluated a large number of chemotherapeutic regimens using a variety of agents and varying durations of treatment. A combination of cyclophosphamide, 5-fluorouracil, and methotrexate (**CMF**) has been most widely used, because of its known activity in patients with overt metastatic disease and its relative ease of administration.

The initial study evaluating this combination therapy in breast cancer showed a 30 per cent reduction in mortality in patients receiving 12 cycles of treatment, but also showed that the effect was most clearly apparent in premenopausal women with one to three positive nodes. These results were regarded as a major advance in treatment of breast cancer, although some similar studies subsequently failed to confirm these findings. This discrepancy was compounded by occasional studies that even showed a worse prognosis in patients receiving adjuvant CMF, although the majority of investigators demonstrated some benefit in selected groups of patients.

Because of the conflicting results obtained by various trials of adjuvant chemotherapy, there was initially very little agreement as to the value of this form of treatment in patients with breast cancer. The matter was finally settled by meta-analysis of all randomized trials evaluating adjuvant chemotherapy for breast cancer. The results of this analysis were initially published in 1985 and 1988 but were subsequently updated in 1992 and 1998 when 15-year monitoring was available.

More than 30 trials evaluating adjuvant polychemotherapy are now available for study, with a follow-up of up to 15 years. Initial results with meta-analysis demonstrated a significant survival benefit in node-positive women aged less than 50, with only marginal effects in other subgroups. With longer follow-up the benefit extended to patients over age 50 and those who were node negative, although the greatest effect has always been demonstrated in those under 50 with positive nodes.

It is now apparent that adjuvant chemotherapy produces a relative reduction in recurrence of breast cancer of approximately 24 per cent and of death of approximately 15 per cent. This reduction may achieve only marginal benefit in patients with very favourable tumours, but the relative reduction in risk produces a significant absolute difference in survival in those with greater risk of recurrence and death (Table 7).

Age (years)	Reduction in risk of recurrence (%)	Reduction in risk of death (%)
<40	37	27
40–49	34	27
50–59	22	14
60–69	18	8
70+	15	10

ns, no significant effect.
* Derived from Early Breast Cancer Trialists' Collaborative Group (EBCTCG)

Table 7 Effect of combination of cytotoxic chemotherapy*

The actual benefit may be estimated by multiplying the projected risk of either failure or death over the ensuing 10-year period, and applying the appropriate factor of risk reduction to see what the absolute benefit will be for any specific subpopulation of women with breast cancer. For example, the latest analysis in 1998 demonstrated an absolute 10-year survival benefit of 11 per cent in node-positive women aged less than 50; in node-negative patients of the same age group the advantage was 7 per cent. The impact on women aged 50 to 69 was less, being 3 per cent in those with positive nodes and only 2 per cent amongst those with negative nodes. The reduction of recurrence following adjuvant chemotherapy occurs mainly within 5 years of initial treatment, but mortality improvement continues for at least 10 to 15 years.

As a result of the meta-analysis, adjuvant chemotherapy is now recommended for all node-positive women who are less than 50 years of age. Many women without lymph node involvement will still relapse, particularly those with large tumours (node-negative women with 3 cm tumours have a 30 per cent failure rate at 10 years). Those with smaller tumours may be content with anti-oestrogen therapy if they are hormone receptor positive, but desire chemotherapy if receptor negative.

Chemotherapy that is anthracycline based has been shown to be marginally superior to CMF polychemotherapy in both reduction in recurrence and overall survival.

Many women elect to have the two-drug regimen of doxorubicin and cyclophosphamide that may be given in four cycles that are 3 weeks apart, completing all therapy in 10 weeks. The benefit of adding taxanes to this combination and the study of various schedules and doses continue to be the objects of clinical trials that are rapidly changing this field.

Patients with a very poor prognosis, such as those with a large number of diseased axillary nodes, may be considered for even more aggressive polychemotherapy or even high-dose chemotherapy with bone marrow rescue. Such treatments are currently being extensively investigated but as yet there is no clear evidence as to their efficacy over standard regimens.

Adjuvant hormonal therapy

Hormonal manipulation has been associated with almost as much controversy as adjuvant chemotherapy. However, its relative ease of administration and lesser toxicity has resulted in its more widespread application. Meta-analysis has now demonstrated very definite indications for its use.

Early studies evaluating hormonal manipulation adopted adjuvant oophorectomy in premenopausal women: these failed to show consistently any survival advantage. The introduction of tamoxifen produced an easily administered, relatively non-toxic form of oestrogen blockade, and this was, therefore, the natural agent to assess as an adjuvant treatment. The initial trials produced conflicting results, although meta-analysis has again clarified issues. As in the case for chemotherapy, meta-analysis of all randomized trials of adjuvant tamoxifen was first reported in 1988 and subsequently updated in 1992 and 1998. Initially tamoxifen was demonstrated to be most effective in node-positive patients aged more than 50, although subsequent analyses identified survival advantages in all groups other than node-negative women under age 50 who were also oestrogen receptor negative. The 1998 analysis showed an 11 per cent absolute 10-year survival improvement for node-positive women and a 6 per cent advantage for those who were node negative. This effect was independent of age, menopausal status, and whether chemotherapy had been given. Women with tumours that were oestrogen receptor poor, however, had no survival benefit (Table 8 and Table 9).

Age (years)	*Reduction in risk of recurrence (%)*	*Reduction in risk of death (%)*
'50	45	32
50–59	37	11
60–69	54	33
70+	54	34

[1] Derived from EBCTCG.

Table 8 Effect of 5 years of tamoxifen therapy*Effect of 5 years of tamoxifen therapy*

Oestrogen receptor	*Reduction in risk of recurrence*	*Reduction in risk of death*
Negative	0%	0%
Positive	50%	28%

[1] Derived from EBCTCG.

Table 9 Effect of 5 years of tamoxifen therapy*

As for adjuvant chemotherapy, the absolute improvement in recurrence rates was greatest within the first 5 years of tamoxifen administration, but survival advantage progressed for at least 10 years after initial diagnosis. An important observation was that adjuvant tamoxifen also reduced the incidence of contralateral cancer by up to 47 per cent. This has obvious implications with respect to prophylaxis in high-risk groups of patients.

Previous concerns about tamoxifen centred on duration of treatment and worries about its side-effects. Both direct and indirect comparisons suggest 5 years of tamoxifen is more effective than 2 years of treatment. There are currently a number of trials comparing 5 years of tamoxifen against longer periods of administration, but reliable long-term results are not yet available. Therefore, 5 years treatment is currently accepted as appropriate.

Tamoxifen has troublesome side-effects including the effects of oestrogen withdrawal such as flushing, vaginal dryness, and discharge. It may also cause weight gain, nausea, and occasional disturbances of nail and hair growth. Such side-effects are more common in younger women and can result in cessation of therapy in up to 30 per cent of individuals. The development of endometrial carcinoma has been a particular worry, although the most recent overview analysis demonstrated an excess mortality of only 1 to 2 per 1000 from this cause. There is certainly little evidence that women taking tamoxifen should be routinely screened for endometrial cancer.

In contradistinction to its side-effects there are certain theoretical advantages to tamoxifen administration resulting from its oestrogenic effects. It reduces low-density lipoprotein cholesterol by 20 per cent and may therefore lessen deaths from coronary artery disease. It also may stabilize bone density and protect against the effects of osteoporosis.

A surprising finding of the 1992 overview analysis was a 24 per cent reduction in mortality in node-positive patients aged less than 50 following oophorectomy. The effect was as great as that of adjuvant chemotherapy and, as discussed above, indicated that some of the effect of CMF was provided by a pharmocological oophorectomy. However, it is also likely that this drug treatment has inherent chemotherapeutic activity. As yet there appears to be no clear benefit of oophorectomy, as opposed to chemotherapy or tamoxifen, as an adjuvant treatment in premenopausal breast cancer. It is also unclear whether oophorectomy, if so used, should be performed surgically or whether reliance can be placed on either irradiation or administration of luteinizing hormone-releasing hormone antagonists such as Zoladex to provide this effect. If performed surgically, it is probable that oophorectomy could be performed using laparoscopic techniques.

Special problems in the management of breast cancer

Paget's disease of the nipple

The majority of patients with Paget's disease are elderly, although the author has seen one patient presenting with this condition in her twenties. The clinical presentation of Paget's disease with an itching, vesicular eruption involving the nipple is confirmed by biopsy and the subsequent demonstration of large cells with pale cytoplasm (Paget's cells) in the epidermis (Fig. 18). The origin and nature of Paget's cells are unknown, although they reflect an underlying ductal carcinoma. The ductal cancer may be either *in situ* or invasive and may or may not be associated with a mass. There is sometimes quite extensive disease within the breast.

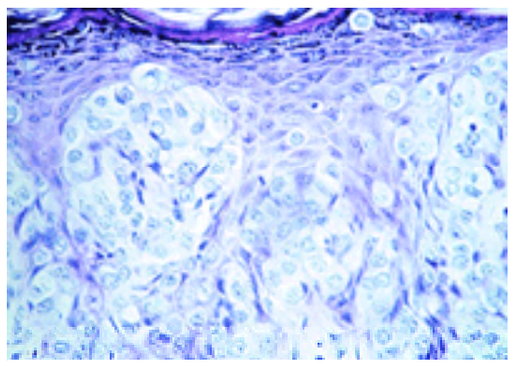


Fig. 18. Paget's cells. The large, clear Paget's cells are seen invading the epidermis. (By courtesy of Dr D. Tarin.)

The differential diagnosis is with eczema, which is usually bilateral and associated with disease elsewhere. Paget's disease of the nipple should also be distinguished from a frank tumour that is simply eroding the skin of the nipple and areola.

Paget's disease of the nipple has traditionally been treated by mastectomy with lower axillary clearance and is thought to have an excellent prognosis. If the cancer is of the *in situ* type, this is certainly true. Patients with more extensive disease associated with invasive cancer have a relatively worse prognosis.

Because so many patients with Paget's disease are elderly, there has been a recent trend to treat this condition by wide local excision of the nipple and underlying ducts. As with other breast conserving techniques, it is important to obtain clear margins. The nipple may be reconstructed. Treatment with tamoxifen or simple radiotherapy alone has not been adequately evaluated.

Breast cancer in the elderly

Breast cancer is a common condition in very elderly women. Some, but not all, authorities claim that tumours in this age group are slow growing, are associated with a high oestrogen receptor status, and have a good prognosis. Treatment with the anti-oestrogen tamoxifen has therefore been advocated after lumpectomy alone. The necessity of adjuvant radiotherapy in the elderly is under study.

The conservative approach using tamoxifen alone should be reserved for the very frail. Randomized trials have shown that, in the elderly, survival is similar after mastectomy compared with tamoxifen alone, without surgery, despite an increased local recurrence rate with the latter treatment. It should also be remembered that not all elderly patients have tumours with a good prognosis. Presentation with axillary nodal involvement and the subsequent development of visceral metastases is common.

Breast cancer during pregnancy and lactation

Breast cancer during pregnancy and lactation is perhaps the subject *par excellence* in which opinion is based on anecdotal rather than proved scientific facts. It is fortunately rare, accounting for less than 1 per cent of all breast cancer cases. However, it is this rarity, along with difficulty in evaluating the pregnant or lactating breast, which makes diagnosis difficult and delayed treatment common.

The traditional view of breast cancer during pregnancy and lactation was that it was an inflammatory type of tumour with cutaneous erythema from intradermal lymphatic involvement (mastitis carcinomatosis). However, although inflammatory cancers do occur in pregnancy, they are more common in other age groups. The majority of tumours seen in pregnant women are similar to those occurring in the non-pregnant state.

Those who believe breast cancer in pregnancy to be of an inflammatory type have also suggested that these tumours have a very bad prognosis. Anecdotal evidence implies that the prognosis is poor, although proper scientific appraisal of the available literature on the subject fails to confirm this belief.

Breast cancer occurring during pregnancy should be treated according to standard principles. Each case must be judged from its own merits regarding the indications or otherwise for breast conservation, adjuvant irradiation, and chemotherapy.

Previous recommendations regarding termination of pregnancy have been revised in the light of lack of evidence that it improves survival. However, as irradiation and adjuvant chemotherapy are contraindicated during pregnancy, a first trimester termination would be indicated for patients with locally advanced disease in whom these adjuvant treatments are necessary. The pregnancy can be continued, if the patient wishes, in those with apparently local disease. The treatment of choice would be mastectomy and axillary clearance, as this would avoid the need for adjuvant irradiation. If nodes were found to be positive, then consideration would have to be given to termination of the pregnancy and provision of adjuvant chemotherapy.

If cancer is detected in the last trimester, surgical treatment can be provided using normal criteria. Adjuvant chemotherapy or irradiation can be delayed until after delivery.

For tumours occurring in the second trimester the choices of management are more difficult. A full discussion with the patient and her partner is required and a treatment policy formulated which takes into account not only the state of the disease but also the patient's wish regarding the pregnancy.

Breast cancers occurring during lactation present special problems. Lactation may be suppressed with bromocriptine and treatment of the cancer based on standard principles.

One question asked by younger women previously treated for breast cancer is what the effect of a subsequent pregnancy may be on their well being. Many authorities recommend that after the diagnosis of breast cancer pregnancy should be avoided for 2 and preferably 5 years. Such advice is again based on somewhat anecdotal evidence and is not widely supported in the literature. Indeed, a recent review from Scandinavia failed to demonstrate any harm resulting from pregnancy after breast cancer. However, it is important to give honest advice to these patients so that the consequences of an untimely pregnancy can be avoided.

Locally advanced cancer (stage III)

Locally advanced tumours are defined as those more than 5 cm in diameter or those with fixed axillary nodes in which there is no evidence of distant metastases. The management of these cancers has been somewhat disappointing because of poor local tumour control and the high incidence of subsequent metastatic disease, resulting in an overall survival rate of 20 per cent or less.

Induction chemotherapy for stage III breast cancer has been associated with a survival in excess of 50 per cent at 5 years, though it has not been established that the treatment is totally responsible for this improvement over historical results. Improved radiological staging has led to the finding of occult metastatic breast cancer in many women who would previously have been considered stage III. The results of induction chemotherapy have been so dramatic, however, that no randomized trial in stage III breast cancer has been undertaken. Two randomized trials considered the use of mastectomy or irradiation following shrinkage of the tumour by systemic chemotherapy; both found survival to be unaffected by the choice but also demonstrated that local control was inferior to that achieved by clinical trials of both irradiation and mastectomy following induction chemotherapy. This has become standard therapy for women who present with inoperable stage III breast cancer. All three modalities are not required for operable stage III tumours.

For women with stage III breast cancer who desire breast conserving therapy but have large tumours relative to the size of their breast, induction chemotherapy has been tried in prospective randomized trials. Several have demonstrated that induction chemotherapy allows a higher rate of breast conservation and achieves identical survival.

Carcinoma *in situ*

Ductal carcinoma *in situ*

Ductal carcinoma *in situ* accounts for only about 5 per cent of all symptomatic breast cancers. It presents either with a mass or with nipple discharge, although it is occasionally a chance finding on biopsy for other reasons. In screening programmes, however, the incidence of ductal carcinoma *in situ* may be as high as 20 per cent and its relative importance has therefore increased over the last 10 years. Indeed, it is now a common clinical problem in routine breast practice.

The natural history of ductal carcinoma *in situ* has been discussed. Most available data relate to symptomatic patients rather than those detected by screening, but it appears that the risk of developing invasive cancer in the former group is about 30 to 50 per cent over a 15-year period. The comedo, high-grade variant of the tumour has the greatest malignant potential. Further evidence relating to the malignant potential of ductal carcinoma *in situ* is available from studies which have found that this cancer occurs in the contralateral breast in up to 50 per cent of women previously treated for breast cancer. However, the risk of developing a frank invasive cancer is only about 1 per cent per year after the diagnosis of the original primary tumour.

One unexplained feature of ductal carcinoma *in situ* is its high expression of the C-*erb*2 oncogene. This is present in approximately 80 per cent of all such tumours, but in only 40 per cent of invasive tumours.

The treatment of ductal carcinoma *in situ* is controversial. When it is extensive in the breast, mastectomy is usually recommended. Such treatment provides local control and survival rates of almost 100 per cent. Treatment of smaller tumours, such as those detected by screening, is the subject of controversy. While occasional surgeons still prefer mastectomy for these smaller tumour with perhaps immediate reconstruction, most now recommend breast conservation by wide excision. A problem with this treatment is the difficulty in ensuring negative margins after resection because of the diffuse, multifocal nature of the disease. Furthermore, subsequent ductal carcinoma *in situ* may develop in surrounding tissues with the passage of time because of an inherent likelihood of malignant change.

It is therefore of great importance that local recurrence be minimized, especially in view of the fact that 50 per cent of all ductal carcinoma *in situ* recur in association with invasive disease. Many studies have now clearly demonstrated the importance of radiotherapy in reducing recurrent disease. The NSABP B-17 study reported an event-free survival of 84 per cent in patients having wide excision with irradiation and 74 per cent in those treated by surgery alone. These significant findings were criticized because they failed to address the question of whether some patients with ductal carcinoma *in situ* could be safely treated without irradiation. This randomized, prospective trial did, however, confirm results of other retrospective surveys on the influence of irradiation on ductal carcinoma *in situ* after wide excision.

In order to correlate risk of recurrence with pathological features and treatment and therefore define the need for adjuvant radiotherapy, Silverstein and his colleagues devised an index dependent on major risk factors for local recurrence, namely nuclear grade, comedo histology, and the extent of surgical clearance. This Van Nuys Prognostic Index (**VNPI**) is summarized in Table 10.

	Score		
	1	2	3
Size	‘1.5 mm	1.6–4.0 mm	> 4.1 mm local recurrence
Margin	> 10 mm	1–9 mm	‘1 mm
Pathology	Non-high grade –necrosis	Non-high grade +necrosis	High grade ± necrosis

Table 10 The Van Nuys Prognostic Index

This index provides a score from 3 to 9 indicating the probability of local recurrence. When this index was applied to 394 patients with ductal carcinoma *in situ*, those with a VNPI of 3 or 4 demonstrated no benefit from adjuvant irradiation. Patients with VNPI scores of 5, 6, or 7 had a statistically significant, 13 per cent benefit in being free of local recurrence when treated with radiotherapy. Patients with scores of 8 or 9 demonstrated most benefit from irradiation but had recurrence rates of 60 per cent at 6 years.

The conclusions from the Van Nuys study are that patients with scores of 3 or 4 may be treated by wide excision alone. Those with intermediate scores of 5, 6, or 7 should have adjuvant irradiation, whereas patients with scores of 8 and 9 should be considered for mastectomy with reconstruction. Thus, this system allows specific targeting of irradiation to those patients who will benefit most from its administration.

As the risk of axillary metastases in ductal carcinoma *in situ* is less than 4 per cent, and in modern screen-detected series approaches zero, there is no place for surgical clearance of the axilla in its management. Sampling may be performed but most authorities now agree that even this limited option is not really necessary.

Lobular carcinoma *in situ*

This has generated less interest than ductal carcinoma *in situ* in recent years because of the importance of the latter in screening programmes. Furthermore, its malignant potential is less than that of ductal carcinoma: only about 25 per cent of patients demonstrating invasion within 20 years of diagnosis. The risk of developing an invasive cancer is as great in the contralateral breast as in the side in which the original diagnosis was made. Lobular carcinoma *in situ* is now regarded, therefore, as a mark of cancer risk rather than a true premalignant tumour. It does not express the C-*erb*2 oncogene.

There are no clinical or radiological features of lobular carcinoma *in situ*, which is a chance histological finding following biopsy for some other condition. Recommendations for treatment vary from simple monitoring to mastectomy with contralateral mirror-image biopsy because of its predisposition to bilaterality. The author suggests, somewhat arbitrarily, that wide excision should be performed. If the specimen obtained shows only a limited amount of fully excised tumour, routine observation is sufficient. If disease is extensive, and particularly if there are other risk factors for breast cancer, mastectomy with or without reconstruction must be considered. There is little merit in contralateral biopsy as positive findings are uncommon.

Screen-detected cancer

Breast cancer is a common condition with a long natural history and there is evidence, perhaps controversial, that treatment at an earlier stage improves prognosis. In theory, therefore, breast cancer lends itself to earlier detection by mammography. The guidelines for successful screening have been laid down by the World Health Organization, and most of these criteria are fulfilled by breast screening programmes (Table 11). Areas of theoretical difficulty regarding breast cancer screening relate to a number of issues. For example, there is no clear consensus as to the overall treatment of breast cancer. Mammography can be applied to populations of women, although it is technically sophisticated and may create anxiety. Whether breast cancer screening is truly cost beneficial in terms of overall health care has been the source of much controversy, although there is an increasing amount of data indicating that it is.

The condition must have a high prevalence
Its natural history must be known
It must have a well-defined early stage
There must be evidence that early treatment provides better results than treatment of later disease
There must be a suitable test
The test must be applicable
There must be facilities available for treatment of abnormality
The screening must do no psychogenic harm
It must be cost efficient

Table 11 World Health Organization guidelines for screening

There have been five main studies evaluating breast cancer screening: the Health Insurance Plan (**HIP**) study from New York, two studies from Sweden (the Two Counties and the Malmo studies), and one study each from the United Kingdom and Canada. In addition there are case–control studies from Holland and Italy.

These studies provide conflicting results. The HIP study demonstrated a 30 per cent reduction in mortality in the screened group of patients, although this was not detected until after 7 years. This study has been criticized in various ways, although it was a remarkable feat when it is considered that it was initiated in the early 1960s, when mammographic techniques were rudimentary. This may explain the small impact provided by mammography in detecting cancers in the screened group. Two hundred and ninety-nine cancers were detected in the study group; 167 of these were interval cancers in that they became clinically apparent between screenings, and 88 were clinically obvious at the time of screening. In only 44 patients was a mammographic abnormality found which was not clinically apparent.

The Two Counties study from Sweden was based on randomization by geographical area and there are a number of statistical considerations which may be criticized. However, on balance, it is a large, well-constructed study which provides good data in favour of screening.

The second Swedish study from Malmo failed to show a significant mortality difference between the screened and unscreened populations, although this was a relatively small trial and a true effect may have been missed. The United Kingdom study also failed to demonstrate a mortality difference between the screened and unscreened populations at initial observation times, although after 8 years a significant difference did become apparent.

The Canadian study was characterized by demonstrating a worsening prognosis in women aged 40 to 49 undergoing screening; this has never been adequately explained. This study did, however, show a positive effect in older women.

Meta-analysis of the above studies in conjunction with case–control studies and interim results from other trials indicates that screening may reduce the mortality of breast cancer by about 30 per cent. The effect is most obvious in women aged over 50. Few data support screening in women younger than 50, although a significant reduction in mortality did occur in this age group after prolonged observation in the HIP study. The question of routine screening under age 50 remains vexed. Meta-analysis of all such screening data fails to demonstrate a significant effect, mainly because of the very negative effect of the Canadian study.

National breast cancer screening programmes have now been initiated in a number of countries, particularly in northern Europe. The age range of women screened and the screening interval depends on the programme concerned. Generally, however, women between 45 and 65 are screened approximately every 2 years. In the United Kingdom, screening is undertaken over the age range 50 to 65, every 3 years. These criteria have been established to provide the maximum reduction in mortality for minimum cost. Screening in the United Kingdom is generally provided by a single oblique mammographic view; if an abnormality is found, the patient is recalled for clinical examination and two-plane mammographic evaluation. Many programmes do not rely on clinical palpation as an initial mode of screening because of lack of evidence as to its efficacy.

The aims and results of national, centrally co-ordinated screening programmes must be differentiated from those populations encouraged to undergo breast screening as part of an overall policy of health promotion, as the latter have varying screening intervals and protocols.

Breast cancer screening has its critics. The reduction in mortality of 30 per cent translates to an improvement in survival of approximately 10 per cent. If these relative percentages are converted to absolute figures, the number of patients in a given population whose lives may be extended by screening may be small. Furthermore, the results of the studies on screening may have been influenced by bias. Lead-time bias is the time by which the diagnosis is advanced through screening: only the HIP study considered this factor. Lag-time bias is the tendency of screening programmes to detect tumours of low biological potential. Other biases, such as selection bias for inclusion in screening programmes and overdiagnosis bias (the diagnosis of an abnormality which might not otherwise have been diagnosed in the lifetime of the patient) should also be considered.

Critics of screening claim that its expense and relatively small effect on survival are not cost-effective. They claim that any apparent improvement in survival may result from lead-time bias and from increased diagnosis of biologically favourable tumours such as ductal carcinoma *in situ* and those tumours of special type. In routine clinical practice, ductal carcinoma *in situ* accounts for only 5 per cent of all cancers detected; in the screened population, it accounts for about 20 per cent of all tumours diagnosed. Similar figures pertain to ‘special' tumours with a good prognosis, such as tubular and medullary cancers. Whether these criticisms are justified cannot be easily addressed, although when a correction has been made for lead-time bias there is still improved survival in the screened population.

The management of screen-detected cancers requires special consideration. Expertise in guided-wire localization is mandatory. An understanding of the natural history of breast cancer is important as are the issues involved in treating ductal carcinoma *in situ* and small tumours of special type. As these patients are asymptomatic the importance of their psychological needs cannot be understated and there is need for appropriate expertise to be readily available to provide adequate support.

Despite worries about its efficacy the advent of screening has been associated with a reduction in tumour size in overall clinical practice. More tumours are of low grade and overall node positivity has lessened from 50 per cent to as low as 30 per cent in some centres. It is therefore likely that screening is contributing to the overall reduction in breast cancer mortality noted in both Europe and the United States during the past decade.

Questions remain about national screening programmes such as the appropriate target age of the screened population, the duration between screenings, and the number of views taken at mammography. If breast cancer screening is to be successful it must be centrally controlled and funded, performed by specially trained personnel, and be associated with a quality assurance programme. Adequate public information is also of prime importance. If screening is performed in less than 70 per cent of the target population its impact is likely to be small. If performed properly with due attention to quality assurance, screening is likely to be of benefit to women in late middle-age.

Follow-up of women with breast cancer

The follow-up of women treated for breast cancer is generally somewhat haphazard with varying intervals of review, no protocol for investigation, and an arbitrary time of discharge—usually 5 years. This confusion is compounded by evidence that prognosis is no different in intensive, investigation-orientated, hospital environments compared with that of the primary care physician.

Patients with breast cancer should receive emotional and psychological support as well as physical evaluation. The latter is directed towards the detection of local regional recurrence as well as distant metastases, although it should be appreciated that two-thirds of recurrences are symptomatic and noted by the patients between hospital visits.

The evaluation of local regional recurrence should include detection of new primary breast cancers. Mammography is therefore recommended every 1 to 2 years after initial diagnosis. Local recurrence in the breast or in regional lymph nodes can be treated or at least palliated with further surgery, irradiation, or chemotherapy.

Investigation for early, asymptomatic distant metastases is much less efficacious. Bony metastases are the most frequent type of relapse, and some authorities have recommended bone scintigraphy at yearly intervals, or less often. The yield of such investigations is low, the expense is considerable, and there is little evidence that detection and treatment of such asymptomatic recurrence improves the prognosis over that obtained when treatment is delayed until the patient complains of symptoms. Furthermore, false-positive investigations provide a constant source of uncertainty. The argument for routine tests such as chest radiographs and liver scanning is even less convincing than that for bone scintigraphy.

A common practice is to review patients at 3-monthly intervals for the first year. Patients with stage I disease are then seen 6-monthly for a year and yearly thereafter. Patients with stage II or III disease are seen 3-monthly for 2 years, 6-monthly for the third year, and then yearly. This is somewhat arbitrary. The time of discharge is also empirical, but most patients are discharged after 5 years—for research purposes monitoring may continue for 10 years, although the accumulation of large numbers of patients may well limit this ideal.

The prognosis of breast cancer

The arbitrary description of the prognosis of breast cancer as a survival rate after an indeterminate amount of time is somewhat naïve. First, the natural history of breast cancer is uncertain and survival, even after the development of metastases, may be prolonged without treatment. Second, recurrence and subsequent death

may occur as long as 25 years after the original diagnosis and, finally, if survival is prolonged, death from other causes must be taken into account. Despite these limitations the analysis of survival demonstrates that about 80 per cent of patients with stage I disease are alive after 5 years, whereas at 10 years this figure falls to 50 per cent. For stage II disease survival rates are 50 per cent and 35 per cent, respectively. Overall survival, including those with stage III and stage IV disease at presentation, is likely to be less than 25 per cent after 10 years

Critics of this method of survival analysis recommend an alternative method (Fig. 19). Curve A describes the survival characteristics in an age-matched population without breast cancer. Curve D is that of the population with breast cancer. Initially the mortality from breast cancer is excessive and curve D diverges from A. With the passage of time deaths from breast cancer become less common and the two curves become more parallel. True parallelism indicates that the risk of death in the population indicated by curve D is no greater than that of A and the group, as a whole, can be considered cured. Such analysis with long-term follow-up, however, shows an increased death rate from breast cancer even 30 years after initial diagnosis, although some of this long-term mortality is due to death from second breast cancers and the cardiopulmonary effects of irradiation.

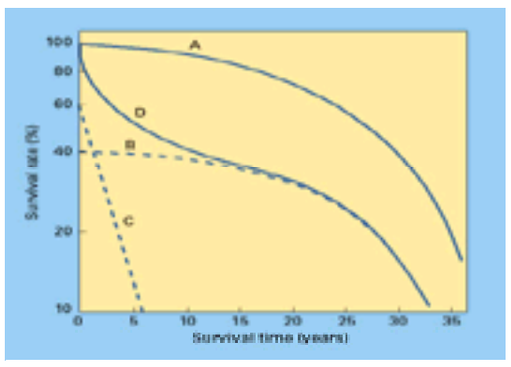


Fig. 19. Long-term survival in breast cancer (see text).

Curve D describes the survival of two distinct groups. Patients in curve B have an intrinsically good prognosis. Curve C, however, represents the poor prognosis group. Unless their outlook can be improved it may be that the overall survival of patients with symptomatic breast cancer will remain unaltered.

There is now increasing evidence of an improved prognosis of breast cancer in Europe and the United States. A number of factors may be involved, such as greater awareness by women and better clinical diagnosis. The main reason for this improvement, however, is adjuvant treatment with chemotherapy and tamoxifen. Screening is likely to produce further benefit in due course.

Prognostic factors in breast cancer

Prognostic factors have important implications for both the patient with breast cancer and for her medical attendant. Not only do they act as a guide to overall prognosis, but they may also determine the need for adjuvant treatment. The latter is of current interest in that it may define a subgroup of patients with apparently favourable node-negative tumours who none the less have a poor prognosis and to whom adjuvant therapy may be specifically targeted.

The major arguments relating to prognostic factors are their interrelationships and relative importance. The relative importance of clinical as opposed to pathological factors is also controversial: there is inherent inaccuracy in clinical assessment of features such as tumour size and axillary nodal involvement. Overall clinical stage is, therefore, only a relatively poor prognostic indicator, and may be less important than specific pathological and certain biochemical parameters. Other clinical factors, such as patient age, menopausal status, obstetric history, and site of tumour in the breast have not been shown to be of much prognostic importance.

Prognostic factors can be determined by multivariate analysis of the various clinical and pathological features of a group of patients with breast cancer. There is some discrepancy relating to the relative importance of prognostic features between studies, although most show that axillary nodal disease, tumour size, and differentiation are the most important (Table 12). It should be noted that these prognostic factors are of importance only for the first 5 to 10 years of follow-up. It is unclear what factors determine longer term prognosis.

Axillary nodal status
Tumour size
Histological grade
Hormone receptor status
Others (menopausal status, age, vascular invasion)

Table 12 Major prognostic determinants

The question of interdependence is difficult to address. An obvious example is tumour differentiation, which may be regarded purely as an expression of the inherent biology of the tumour. Thus a poorly differentiated tumour has unfavourable biological characteristics and is likely to be large at presentation, to be associated with axillary nodal involvement and negative receptor status, and to have a poor prognosis. On the other hand, many studies have shown that tumour differentiation is an independent prognostic variable in its own right and acts as a determinant independent of factors such as tumour size and nodal disease.

Lymph node metastases

Nearly all studies show that the presence of axillary lymph node metastases is the most important prognostic determinant. Prognosis is a function of the number of lymph node metastases, although other factors such as level of disease, extranodal spread, or size of tumour metastases may be of some minor significance. Thus, isolated tumour cells seen in axillary nodes using histochemical techniques represent the most minor degree of nodal involvement. The significance of these features is unclear. A single micrometastasis less than 2 mm in diameter is associated with an overall prognosis similar to that of node-negative disease. Disease affecting one node is also associated with a prognosis similar to that of node-negative disease, although after 10 to 12 years such spread may be associated with a worsening prognosis. Disease affecting up to three lymph nodes classifies the patient in a relatively favourable subgroup of stage II. Those with four to 12 nodes affected form an intermediate group; the prognosis is decidedly poor if more than 12 lymph nodes exhibit metastatic disease.

Tumour size

Tumour size is the second most important prognostic determinant. Although tumour size and lymph node metastases are closely related, they are also independent of each other in terms of prognosis. The influence of tumour size is most obvious when lymph nodes contain metastases and is of greatest importance in the first 5 years after the initial diagnosis.

Histological grade

The grading of invasive breast cancer has been regarded as notoriously difficult because of varying interpretation by pathologists. Recent experience indicates that specialists in this field can, however, obtain a remarkable degree of agreement—this is due mainly to the creation of specific guidelines to standardize grading by

pathologists. Grade is classified as I, II, or III on the basis of tubule formation, nuclear pleomorphism, and most importantly, mitotic rate.

About 20 per cent of cases are classified as well differentiated (grade I), although there is some variation between centres. Twenty per cent of patients are classified as intermediate (grade II), and the remainder demonstrate poor differentiation (grade III). Investigations such as those of the Cancer Research Campaign and the Nottingham Tenovus Breast Cancer Study group clearly demonstrate a relationship between grade and prognosis, which is independent of other variables.

A difficult problem relating to histological grade is its correlation with the morphological type of the tumour, as there is a relationship between tumour type and prognosis. Special types of breast cancer such as pure tubular, mucinous, and invasive cribriform carcinoma have an excellent prognosis while infiltrating lobular and medullary carcinoma appear to have an intermediate prognosis between the first group and the common infiltrating ductal carcinoma of no special type. It could be argued that histological grading is appropriate only to invasive ductal cancer of no special type, but this would result in important prognostic information being unavailable for up to 25 per cent of patients.

The grading of classic lobular cancer has also been difficult because of the nature of the tumour cells and lack of histological pattern in this particular type of cancer. However, grading is now routinely performed even for this tumour type.

The importance of axillary nodal disease, tumour size, and differentiation has resulted in a number of successful attempts to define a ‘prognostic index' using this information.

Hormonal receptor status

It has been known for many years that certain patients with breast cancer respond to endocrine therapy and have a more indolent form of the disease. Such endocrine-sensitive tumours can be identified by measurement of oestrogen or progesterone receptors within the cancer itself. The oestrogen receptor is a cytosol protein that can be identified using a variety of methods. Previously, the most common method of detection involved the incubation of the supernatant fluid of a tissue homogenate with radiolabelled oestrogen. The unbound label is then removed by dextran-coated charcoal or sucrose density centrifugation and an assay performed. More recent methods use monoclonal antibodies against the receptor; these methods can be performed on fixed tissue and are not biased by the presence of oestrogen or anti-oestrogen.

When cancer cells with intact and functional oestrogen receptors are exposed to oestrogen the latter is transported to the site of mRNA production in the nucleus. The result of this interaction includes the formation of additional oestrogen and progesterone receptors and other growth factors that help dictate cell characteristics.

The presence of oestrogen or progesterone receptors within the cell has been associated with an improved short-term prognosis in some, but not all, studies. Although they provide some prognostic information, they are somewhat weak indicators of overall prognosis and not as important as tumour size, nodal status, or degree of differentiation.

The relative importance of oestrogen and progesterone receptors is difficult to assess. Some studies indicate that the presence of progesterone receptor is an inferior prognostic indicator compared with the presence of oestrogen receptor. On the other hand, some authorities have demonstrated progesterone receptor to be as good a prognostic indicator as node status or tumour size. If oestrogen and progesterone receptors are both present, the prognostic information is likely to be more powerful than if only one of these factors is present.

Other controversial factors

Other histological features such as vascular or lymphatic invasion in the breast, perineural invasion, tumour necrosis, or mucin production may be of prognostic importance although their clinical significance is open to doubt.

While not a tumour characteristic, there has been interest as to whether the time of the menstrual cycle at which surgery—or other physical manipulation—of the breast occurs may have a bearing on prognosis. The basis of this suggestion is the hypothesis that the unopposed oestrogen secreted in the follicular phase of the menstrual cycle (days 3 to 12) adversely affects some breast cancers leading to increased risk of viable cells being shed at the time of surgery. Oestrogens may also stimulate tumour cell proliferation and induce protease synthesis leading to increased probability of invasion and establishment of micrometastases. In the luteal phase of the menstrual cycle, progesterone production negates these effects.

The initial evaluation of any prognostic influence of the menstrual cycle phase in which excisional surgery was performed was in 249 patients at Guy's Hospital, London. In this study 58 per cent of patients survived 10 years if operated on in the follicular phase, but 84 per cent survived if they had surgery at other times in the cycle. Some other studies confirmed these findings, whereas other authorities failed to demonstrate an effect or even found that operating in the folliclar phase was associated with improved outlook. One meta-analysis of 21 publications on the subject demonstrated a 16 per cent reduction in mortality in patients who had surgery during the luteal phase. Prospective studies are addressing this question.

At present there is no prognostic factor, other than histological grade, that has been shown to add significantly to the combination of TNM staging. Attention is turning from prognostic factors to the search for predictive factors that will identify the tumour most likely to benefit from a specific therapy; the first example being oestrogen and progesterone receptor levels that predict a response or lack of response to tamoxifen. The newest example is the identification of overexpression of HER2-neu (also called *erb*-B2), which is found in a minority of tumours. It appears to indicate tumours that derive less benefit from some adjuvant therapies than from others. A humanized monoclonal antibody directed against it has been shown to have effect in clinical trials on those who overexpress this marker. The next few years will further define what significance this predictive factor has for clinical decision-making.

The use of a panel of monoclonal antibodies or a pooled polyclonal antibody directed against breast cancer antigens to detect micrometastases in the bone marrow of patients with breast cancer has been shown to have prognostic significance in several institutional series. At present, the expense and morbidity of this assay have not been shown to be offset by sufficient clinical utility to recommend it in other than research settings.

Some have suggested that the identification of BRCA1 or 2 mutations should influence the choice of primary therapy in favour of mastectomy. Although no prospective study exists, anecdotal data do not support such an inference from the presence of such mutations.

Many prognostic factors have been identified that do relate to outcome. To have any clinical utility, a marker must have added benefit to tumour size, nodal status, histological grade, and hormone receptor status. A new factor will deserve inclusion as standard care only if multivariate analysis of a prospective study shows that it exhibits clinical utility beyond the standard parameters.

Metastatic breast cancer

Recurrence of breast cancer is of ominous significance. It is incurable, with a median survival of about 15 months, although there is a wide range lasting from only a few weeks to many years. The role of treatment in metastatic disease is twofold: to alleviate symptoms and to prolong survival. Despite a plethora of new drugs and therapeutic options there is little evidence that overall survival has been much improved once metastases have occurred, although the duration and quality of life in patients responding to therapy have improved.

The management of metastases demands a multimodality approach by surgeons, radiotherapists, and medical oncologists. The role of careful diagnostic radiology in assessing the extent of disease cannot be overemphasized: appropriate counselling is particularly important, especially for mothers with young families, whose overall outlook must be frankly discussed.

Assessment of patients with metastatic disease

Once a patient has presented with recurrence it is necessary to determine the full extent of metastatic spread. This requires full and careful examination of the site of the original primary tumour and its regional lymph nodes. A full series of staging investigations, including chest radiography, abdominal ultrasound, and bone scintigraphy are also necessary to determine the presence of other sites of occult metastatic disease. Brain scanning is not recommended: the likelihood of this investigation detecting asymptomatic recurrence is small.

There is little evidence that tumour markers such as carcino-embryonic antigen have any significant role in assessment of metastatic breast cancer or the evaluation of response to treatment.

A full series of staging investigations helps to determine the appropriate therapy. It also allows a more accurate appraisal of response to treatment.

Prognostic indicators in metastatic disease

The most important prognostic indicator in metastatic disease is disease-free interval after diagnosis of the primary tumour. The median survival of women who relapse in a single site within 1 year of diagnosis is about 11 months; those who experience a disease-free interval of more than 5 years have a survival of up to 40 months.

The site of metastatic disease is also important: prognosis is more optimistic in patients with local regional or bony metastases than in those with recurrence in the liver or central nervous system. The number of metastatic sites is also a factor: survival of patients with three sites of metastases is only 50 per cent of those with one affected site.

Some prognostic information may be obtained from knowledge of oestrogen and progesterone receptor status, although this is not as important as the factors discussed above. However, such information may be important in determining treatment: patients positive for such receptors are more likely to respond to endocrine manipulation.

Treatment of metastases

The treatment of metastatic disease at individual sites is discussed below. However, as the majority of patients with metastases can be regarded as suffering from widespread systemic disease, the use of systemic therapy is nearly always appropriate. Breast cancer is relatively unusual in being responsive not only to radiotherapy and chemotherapy but also to various endocrine manipulations.

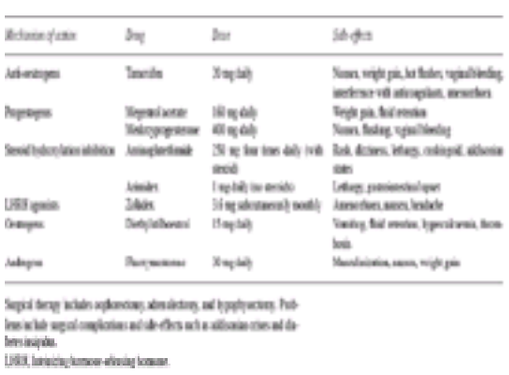
Endocrine therapy

The mechanism of response to endocrine therapy is unclear. The simplistic explanation is that breast cancer requires oestrogen for growth and that inhibition of this hormone's action will produce a response. This explanation, however, fails to explain why a given breast tumour can respond to both oestrogen withdrawal and addition. It also fails to explain how breast cancer may respond to other therapeutic manipulations involving progestogens, androgens, and corticosteroids, when there is little evidence that they are involved in tumour growth.

Despite the difficulties in understanding the mechanisms of endocrine therapy it appears that its effect is mediated by receptors, especially those for oestrogen. The overall response to endocrine therapy in breast cancer is about 30 per cent. However, it may reach 60 per cent in those positive for oestrogen receptors, or be as low as 10 per cent in those negative for such receptors. The higher the level of oestrogen receptors the greater the likelihood of response. If the level of progesterone receptors is also elevated the response rate is likely to be increased.

The choice of endocrine therapy

A wide range of endocrine therapies is available (Table 13). Some are pharmacological agents; others require surgical manipulation. The latter were frequently adopted up to the 1960s, but in the last 20 years oral agents have been used with increasing frequency. Apart from their ease of administration, pharmacological agents have a much broader application, as many of the surgical options were limited to premenopausal patients only.



Mechanism of action	Drug	Dose	Side effects
Anti-oestrogens	Tamoxifen	20 mg daily	Nausea, weight gain, hot flushes, vaginal bleeding, interference with anticoagulant, amenorrhoea
Progestogens	Megestrol acetate	160 mg daily	Weight gain, fluid retention
	Medroxyprogesterone	400 mg daily	Nausea, flushing, vaginal bleeding
Steroid hydroxylase/aromatase inhibition	Aminoglutethimide	250 mg four times daily + with steroids	Rash, dizziness, lethargy, cerebrospinal, addisonian states
	Arimidex	1 mg daily (no steroids)	Lethargy, gastrointestinal upset
LHRH agonists	Zoladex	3.6 mg subcutaneously monthly	Amenorrhoea, nausea, headache
Oestrogens	Diethylstilboestrol	15 mg daily	Vomiting, fluid retention, hypercalcaemia, thrombosis
Androgens	Fluoxymesterone	30 mg daily	Masculinization, nausea, weight gain

Surgical therapy includes oophorectomy, adrenalectomy, and hypophysectomy. Problems include surgical complications and side-effects such as addisonian crises and diabetes insipidus.
LHRH, luteinizing hormone-releasing hormone.

Table 13 Endocrine therapy in breast cancer

In general there is little evidence that response rates to or survival benefit from one type of endocrine therapy are superior to those of any other, although androgens have fallen into disfavour becauseof their lower activity and side-effects. The choice of endocrine therapy is therefore based on ease of administration and freedom from side-effects. As a result the anti-oestrogen, tamoxifen, has become the treatment of choice for first-line endocrine therapy. Younger patients in particular, however, are prone to suffer hot flushes and weight gain with this treatment. Randomized trials have shown tamoxifen to be as effective as oophorectomy, aminoglutethimide, and progestogens.

Progestogens are as active as tamoxifen as first-line agents. They are also good second-line treatment after failure of response to tamoxifen in both premenopausal and postmenopausal patients. Side-effects, particularly weight gain, fluid retention, and nausea, can be a problem.

Aminoglutethimide, an aromatase inhibitor, is as active as tamoxifen as a first-line treatment and, in some studies, has been shown to be particularly effective against bone metastases. Dosage should be slowly increased, but at the normal therapeutic dose (250 mg four times daily) supplemental hydrocortisone (30 mg daily) and even mineralocorticoid is required. Cushingoid and addisonian states can therefore accompany aminoglutethimide therapy. Other side-effects such as rashes, nausea, and drowsiness somewhat limit its use, although it can be useful in premenopausal patients after failure of tamoxifen and progestogens.

A major advance has been the development of a series of selective aromatase inhibitors which have fewer side-effects than aminoglutethimide because they do not interfere with cortisol synthesis. The best known is Arimidex, which at a dose of 1 mg daily has been shown to be as efficacious as progesterone but without the side-effects. It is likely to become the second-line hormonal therapy after tamoxifen.

Several authorities have recently described success with luteinizing hormone-releasing hormone agonists such as Zoladex. It is likely that this group of drugs will be used more often in premenopausal patients with metastatic breast cancer.

The success of the various oral agents has reduced enthusiasm for surgical endocrine therapy: adrenalectomy and hypophysectomy have almost disappeared from use. Some authorities still recommend oophorectomy with enthusiasm, and recent data relating to its potential role as an adjuvant therapy in premenopausal stage II disease suggests that its future may be assured. Whether oophorectomy will be replaced by the newly introduced luteinizing hormone-releasing hormone agonists remains to be seen. The suggestion that a combination of endocrine therapies may improve survival has not been demonstrated in randomized studies, although higher response rates may be seen.

Tamoxifen is, therefore, the ideal first-line endocrine therapy, with an overall response rate of about 30 per cent (up to 60 per cent in those positive for oestrogen receptors). Failure to respond is an indication for abandoning endocrine therapy. Relapse following an initial response can be treated with a progesterone or, in postmenopausal women, one of the newer selective aromatase inhibitors such as Arimidex, although the response rate to such second-line treatment is only about half that seen with first-line therapy. In the unlikely event of a meaningful response to second-line hormonal therapy then a third hormonal agent may be introduced. For example, if Arimidex has been used as second-line treatment then a progesterone may be introduced or vice versa. In postmenopausal women oestrogens may be considered as second-or third-line treatment. Oestrogens are not used as much as previously and care must be taken in patients with bone metastases because of the risk of hypercalcaemia. Androgens and surgical endocrine therapy such as oophorectomy or adrenalectomy are no longer considered appropriate treatments.

A characteristic, but relatively unusual, feature of response to hormonal therapy is ‘flare', the mechanism of which is unclear. It is characterized by an exacerbation of

symptoms associated with erythema and swelling after the initiation of endocrine treatment. The symptoms resolve within 1 month.

Chemotherapy

About 60 per cent of patients with previously untreated metastatic breast cancer respond to appropriate chemotherapeutic protocols, although complete resolution of disease is seen in fewer than 20 per cent. The median duration of response is about 1 year, although this depends on the site of metastatic disease: better responses are seen in soft tissue and bone than in the liver or central nervous system. The time of response is also variable, depending on the site of disease. Cutaneous or lymphatic metastases may respond within 3 to 6 weeks; response may not be seen until after 4 or 5 months if metastases are in bone.

Chemosensitivity of breast cancer

Overall response rates to the more commonly used chemotherapeutic agents used in breast cancer are shown in Table 14. The most active agents are anthracyclines such as Adriamycin. In some, but not all, studies Adriamycin alone has produced a response rate approaching 60 per cent; in other investigations it has been shown to be as active as combination chemotherapy with cyclophosphamide, 5-fluorouracil, and methotrexate.

Doxorubicin	45
Taxanes	45
Cyclophosphamide	33
Mitozantrone	33
Methotrexate	25
5-Fluorouracil	25
L-Pam	20
Vincristine	20

Table 14 Percentage response rates to chemotherapeutic agents used for breast cancer as single agents

Vincristine has been widely used in combination therapy for metastatic breast cancer, but no advantage over regimens in which it is lacking has been demonstrated. Its advantage is that, unlike other commonly used agents, it is not myelosuppressive.

Mitomycin C demonstrates substantial activity against metastatic breast cancer and has even been shown to induce responses in tumours resistant to Adriamycin and cyclophosphamide. It is most commonly used in combination with methotrexate and mitozantrone (MMM) as a second-line treatment.

Combination chemotherapy has enhanced activity, albeit at the expense of a higher incidence of side-effects. The most commonly used regimen is the combination of cyclophosphamide, methotrexate, and 5-fluorouracil (CMF), which has an overall response rate of up to 60 per cent. The addition of vincristine and prednisone to this combination does not significantly enhance the response obtained. However, the substitution of methotrexate with Adriamycin (CAF) is associated with an increased response rate of prolonged duration, although with more side-effects.

Adriamycin, either alone or in combination with cyclophosphamide and 5-fluorouracil, has been used mainly as a second-line treatment following failure of initial chemotherapy, or in the management of younger patients with a poor prognosis, such as those with liver metastases. Its toxicity precludes its use in older patients.

A more recently adopted combination is based on mitomycin C, methotrexate, and mitozantrone. This has potential value both in primary management of metastatic breast cancer and as a second-line treatment.

Hopes that newer agents such as Taxol might be associated with enhanced responses in patients with metastatic breast cancer have unfortunately not been realized. Similarly, high-dose chemotherapy with bone marrow rescue has not demonstrated significant survival advantage over standard treatments.

Combined chemoendocrine therapy

There has been a natural tendency to combine chemotherapy with endocrine treatment. Although response rates and duration of response may be enhanced, the overall benefit is small and as a result such combinations are not recommended. In theory at least, the endocrine treatment may slow down the cell cycle, thus reducing the efficacy of the chemotherapy.

Management of metastatic disease at specific sites

Local regional recurrence

Recurrence in the remaining breast tissue after partial mastectomy or on the chest wall after total mastectomy has been discussed above. Local recurrence must be distinguished from regional recurrence: the latter relates to recurrent disease in the adjacent lymphatic field. Regional recurrence is always of clinical importance and implies a worsening of prognosis. However, it is particularly important if it occurs in a treated rather than an untreated field. The management of regional recurrence depends on the site of relapse. If it is in the internal mammary or supraclavicular areas, radiation can be used, followed by endocrine or cytotoxic therapy as appropriate. Surgery has little part to play in the treatment of relapse at these sites.

Treatment of axillary relapse is by full surgical clearance, if this has not already been performed. Clearance of a previously irradiated axilla is likely to cause arm lymphoedema. Relapse occurring after clearance is treated by radiotherapy or systemic treatment.

Metastases in the gastrointestinal tract

The liver is a common site of metastatic spread of breast cancer. Such metastases are characteristically multiple and surgical resection has little role in their management. Resection is not even indicated in the unusual situation of a single hepatic metastasis. The prognosis of hepatic metastases is poor, and aggressive therapy is indicated. Hepatic metastases do not respond well to hormone therapy. In young patients with good performance status, combination chemotherapy, including Adriamycin, is appropriate as first-line therapy. Older patients may only be suitable for endocrine therapy, although this is not likely to be so efficacious.

Peritoneal metastases may cause ascites or intestinal obstruction. Such metastatic spread is said to be particularly common in patients with lobular carcinoma. Systemic combination chemotherapy is required if it can be tolerated. Surgical intervention should be avoided if possible; it may be required to relieve obstruction, or for insertion of a peritoneal–systemic shunt if ascites becomes intractable.

Occasionally, isolated metastases are seen in the ovary or the adrenal gland. If the retroperitoneum is involved, however, it is usually a manifestation of widespread carcinomatosis including the peritoneal cavity. Renal failure due to obstructive uropathy is a common cause of death in these patients.

Pulmonary metastases

Metastatic disease in the lungs is often multiple and not amenable to surgical resection or radiotherapy. Systemic combination chemotherapy is indicated, although endocrine treatment is more appropriate for older patients with poor performance status. Metastatic spread causing lymphangitis can cause distressing dyspnoea. The diagnosis can be difficult in its early stages because it produces few changes on chest radiographs. Combination chemotherapy should be instituted if possible, with the addition of steroids for symptomatic relief.

Widespread pleural metastases may cause effusions; if recurrent and symptomatic they should be tapped. Pleurodesis is indicated if pleural effusions are persistent.

Bony metastases

Bony metastases are the most common form of secondary spread of breast cancer. These are usually at multiple sites, although it is not unusual for a solitary tumour metastasis to cause symptoms. Pain and pathological fracture are the most common forms of presentation.

Bisphosphonates have been demonstrated to reduce dramatically both the subjective symptoms of bone metastases and the incidence of bone fractures. At the first sign of bony or soft-tissue metastases, oral or intravenous bisphosphonates should be instituted.

A short course of radiotherapy will relieve the symptoms of bony metastases. Systemic therapy is indicated at the same time because of the high risk of developing recurrences at multiple sites, which may not be clinically or radiologically apparent. Endocrine therapy with tamoxifen is appropriate first-line treatment, especially in older patients; failure to respond is an indication for chemotherapy. Oestrogen therapy is contraindicated in patients with widespread bony metastases because of the risk of hypercalcaemia; conversely, aminoglutethimide and Arimidex have been shown to be particularly active.

Pathological fractures require appropriate orthopaedic evaluation and surgical fixation where necessary. Widespread bony metastases, even if asymptomatic, may cause hypercalcaemia, with symptoms of fatigue, weakness, vomiting, and eventually coma with renal failure. Initial treatment involves rehydration with forced diuresis and administration of steroids to enhance calcium excretion in the urine. Resistant cases can be treated with either mithramycin or one of the diphosphonate drugs; both reduce serum calcium by inhibiting osteoclastic bone reabsorption.

Central nervous system metastases

Metastases may affect the brain itself, the cord, the meninges, or the epidural space, where they may cause cord compression. The symptoms resulting from such spread are therefore very varied, and central nervous system metastases must be considered in any patient with neurological signs or symptoms. Their treatment relies on radiation therapy. Surgery has a relatively small role, except for patients with cord compression: excision of brain metastases has no advantage over radiation therapy alone. Systemic therapy has not been widely used as most agents fail to cross the blood–brain barrier, although recent experience suggests that combination chemotherapy with steroids may have some activity in these patients.

Bone marrow metastases

The significance of occult bone marrow metastases is unclear. If symptomatic they are often associated with bony deposits. Treatment is complex because of potential leucopenia and thrombocytopenia resulting from both marrow replacement and the effects of chemotherapy. The choice of endocrine or chemotherapy is made on the basis of performance status, age, and ease of administration.

Ocular metastases

Acute presentation to the ophthalmologist with sudden painless blindness is not uncommon and may be due to a choroidal metastasis. Treatment is with local radiotherapy.

Other ocular symptoms may be produced as a result of metastatic disease to the bony orbit, the orbital contents, the optic nerve, or to the optic pathways within the brain itself.

Psychological sequelae following treatment of breast cancer

Women suffer psychiatric morbidity for at least 1 year after the diagnosis of breast cancer. Approximately 20 per cent of women demonstrate a significant departure from their normal mood 12 months after diagnosis compared with only 8 per cent of women who have undergone biopsy for benign disease.

The major symptoms relate to a combination of depression and anxiety, along with associated sleep disturbance, impaired concentration, fatigue, and various somatic symptoms such as headache and palpitations. About 1 in 10 of these patients have symptoms severe enough to warrant inpatient treatment.

Preoperative counselling and appropriate support in both the perioperative and postoperative periods is mandatory in the management of breast cancer. Specially trained health care professionals, including breast care nurses, social workers, and clinical psychologists should provide an integrated support network in conjunction with nursing and medical personnel. This is of particular importance if radiotherapy or chemotherapy are to be considered. The routine provision of such support results in improved quality of life; it may even favourably influence survival. Physiotherapy while in hospital reduces symptoms due to arm stiffness and may well speed rehabilitation. It is important to ensure that patients undergoing mastectomy without reconstruction are also introduced to surgical appliance specialists so that a prosthesis may be fitted while in hospital.

There is some controversy over the impact of a formal programme of psychological support in patients with breast cancer. Some controlled studies have failed to demonstrate a reduction in psychiatric morbidity following perioperative counselling. There is even some evidence that excessive counselling may lead to ‘sensitization' of the patient and increased anxiety or depression, at least in the short term.

One area of particular interest relates to the psychological sequelae of mastectomy itself. It was hoped that partial mastectomy, with breast conservation, would reduce psychological morbidity. However, this does not appear to be the case: there is a similar incidence of psychological problems after both total and partial mastectomy. Fear of mastectomy is inextricably linked to that of the cancer itself, and variations in treatment have relatively little impact in reducing psychological morbidity. However, it must be remembered that mastectomy, as opposed to breast conservation, may devastate the lives of a substantial proportion of women.

Prevention of breast cancer

Breast cancer is nearly always a sporadic event; a true genetic factor is present in less than 10 per cent of cases. Therefore, the majority of individuals with risk factors do not develop breast cancer and the majority of patients with the disease are not at excessive risk. This makes prevention of breast cancer difficult.

Demographic studies show breast cancer to be particularly a disease of the Western world, and perhaps a result of the lifestyle adopted by those individuals in whom there is a high incidence. It is clearly difficult to change lifestyle factors such as age of having children, diet, and use of exogenous hormones, and attempts to reduce the incidence of breast cancer by influencing these factors may fail.

Two areas of prevention require special mention. It has been suggested that the low incidence of breast cancer in Japan may be secondary to the very low-fat diet eaten in that country. In theory, reducing the amount of fat eaten by populations at risk may reduce the cancer incidence. However, it is very difficult for people of Western culture to reduce the amount of fat in their diet to that of indigenous Japanese people. Animal studies and dietary studies in the West have failed to confirm any effect of fat intake on breast cancer.

Studies on adjuvant tamoxifen have shown a significant reduction in the incidence of contralateral breast cancer. Prophylactic chemosuppression with tamoxifen, especially for ‘high-risk' patients, has therefore been suggested, and this is the subject of a number of trials currently in progress. Benefit has been shown in one American study, although interim results from three European trials have failed to demonstrate an effect. An alternative method of reducing the incidence of breast cancer might be that of chemoprevention with agents such s the retinoids or selenium.

Occasionally, very high-risk patients will ask their medical attendant for prophylactic mastectomy, with or without reconstruction, to avoid breast cancer. There is undoubtedly the occasional case where this is justified, but if possible this rather aggressive approach should be avoided as there remains doubt as to the efficacy of the procedure.

Further reading

Badwe RA *et al.* Timing of surgery during menstrual cycle and survival of premenopausal women with operable breast cancer. *Lancet* 1991; **337**: 1261–4. [The initial publication relating breast cancer

outcome to timing of surgery in the menstrual cycle.]

Basade MM, Gulati SC. High dose chemotherapy in metastatic breast cancer *Lancet* 1998: **351**: 386–7. [A review on the role of high dose chemotherapy in patients with metastatic breast cancer.]

Baum M. The curability of breast cancer. *British Medical Journal* 1976; **i**: 439–42. [A philosophical viw on the curability of different types of breast cancer.]

Brinkley D, Haybittle JL. Long term survival of women with breast cancer. *Lancet* 1984; **i**: 1118. [Discussion of long term mortality.]

Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Clinical Trial. *Journal of the National Cancer Institute* 1988; **80**: 1125–32. [Long-term results of HIP screening trial.]

Cuzick J *et al.* Cause specific mortality in long term survivors of breast cancer who participated in trials of radiotherapy. *Journal of Clinical Oncology* 1994; **12**: 447–53.[Overview analysis on role of radiotherapy in management of breast cancer.]

DeVita VT. Breast cancer therapy: exercising all our options. *New England Journal of Medicine* 1989; **320**: 527–9.

Early Breast Cancer Trialists' Collaborative Group. Ovarian ablation in early breast cancer: overview of the randomised trials. *Lancet* 1996; **348**: 1189–96. [Overview on effect of ovarian ablation on breast cancer mortality.]

Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; **351**: 1451–67.[Overview on mortality reduction using adjuvant tamoxifen.]

Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998; **352**: 930–42. [Overview on mortality reduction using adjuvant chemotherapy.]

Fentiman IS, Mansel RE. The axilla: a no-go zone. *Lancet* 1991; **337**: 221–3. [An argument for radical axillary surgery.]

Fentiman IS, Gregory WM, Richards MA. Effect of menstrual cycle phase on surgical treatment of breast cancer. *Lancet* 1994; **344**: 402. [An overview analysis on the timing of breast surgery with respect to the time within the menstrual cycle.]

Fisher B *et al.* Eight year results of a randomised clinical trial comparing total mastectomy and lumpectomy with or without irradiation in the treatment of breast cancer. *New England Journal of Medicine* 1989; **320**: 822–8. [An evaluation of the influence of the extent of axillary nodal metastases on prognosis.]

Fisher B *et al.* Lumpectomy compared with lumpectomy and radiation therapy for treatment of intraductal breast cancer. *New England Journal of Medicine* 1993; **328**: 1581–6. [A pathologic assessment of breast cancer screening trials.]

Fisher ER *et al.* Pathologic findings from the NSABP (Protocol No.4). V. Significance of axillary nodal micro- and macrometastases. *Cancer* 1978; **42**: 2032–8. [Long-term results of randomized trial of breast conservation therapy.]

Fisher ER, Sass R, Fisher B. Pathologic findings from the National Surgical Adjuvant Project for Breast Cancers (Protocol No.4). *Cancer* 1984; **53**: 712–23. [Randomized trial demonstrating the effect of radiotherapy in reducing local recurrence after breast conservation for DCIS.]

Haagensen CD *et al.* Treatment of early mammary carcinoma: a co-operative international study. *Annals of Surgery* 1969; **170**: 875–78. [Historic review on breast cancer treatment.]

Haagensen CD, Lane N, Lattes R, Bodian C. Lobular neoplasia (so-called lobular carcinoma *in situ)* of the breast. *Cancer* 1978; **42**: 737–69. [Original and seminal paper on lobular carcinoma *in situ.*]

Harris JR, Lippman ME, Veronesi U, Willett W. Medical progress. Breast cancer. *New England Journal of Medicine* 1992; **327**: 319–27. [Review on breast conservation in breast cancer.]

Hayward JL. The Guy's trial of treatments of early breast cancer. *World Journal of Surgery* 1977; **1**: 314–16. [Long-term results on an early breast conservation trial.]

Holland R, Veling SHJ, Mravunac M, Hendriks JHCL. Histologic multifocality of Tis, T1–2 breast carcinomas. *Cancer* 1985; **56**: 979–90. [An important paper establishing multifocality in invasive breast cancer.]

Kerlikowske K *et al.* Efficacy of screening mammography. A meta-analysis. *Journal of the American Medical Association* 1995; **273**: 149–54. [A meta-analysis of screening for breast cancer.]

Lagios MD. Ductal carcinoma *in situ*: Controversies in diagnosis, biology and treatment. *Breast Journal* 1995; **1**: 68–78. [A surgical–pathological review of DCIS.]

McGuire WL. Adjuvant therapy of node negative breast cancer. *New England Journal of Medicine* 1989; **320**: 525–7. [An analysis of the role of adjuvant chemotherapy in lower risk patients.]

Neven P, Vergote I. Should tamoxifen users be screened for endometrial lesions? *Lancet* 1998; **351**: 155–7. [A review of the need for evaluating risk of endometrial cancer in patients receiving tamoxifen.]

Page DL, Anderson TJ. *Diagnostic histopathology of the breast*. Churchill Livingstone, Edinburgh, 1987. [Classical account of pathology of breast disease.]

Peto R. Five years of tamoxifen—or more? *Journal of the National Cancer Institute* 1996; **88**: 1791–3. [A discussion of the optimal duration of tamoxifen administration.]

Roberts MM *et al.* Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* 1990; **335**: 241–6. [Results of a randomized trial of breast cancer screening.]

Schnitt SJ, Silen W, Sadowsky NL, Connolly JL, Harris JR. Ductal carcinoma *in situ* (intraductal carcinoma) of the breast. *New England Journal of Medicine* 1988; **318**: 898–903. [An early review on management of DCIS.]

Silverstein MJ *et al.* A Prognostic Index for ductal carcinoma *in situ* of the breast. *Cancer* 1996; **77**: 2267. [A description of a clinicopathological index that determines treatment of DCIS.]

Tabar L *et al.* Reduction in mortality from breast cancer after mass screening with mammography. *Lancet* 1985; **i**: 829–32. [Early results of the first Swedish trial of breast cancer screening.]

Tabar L, Fagerberg G, Duffy S, Day N. The Swedish Two County Trial of mammographic screening for breast cancer: recent results and calculation of benefit. *Journal of Epidemiology and Community Health* 1989; **43**: 107–14. [Long-term results of the first Swedish trial of breast cancer screening.]

Veronesi U *et al.* Breast conservation is the treatment of choice in small breast cancer: long term results in a randomised trial. *European Journal of Cancer* 1990; **26**: 668–70. [Randomized trial of breast conservation in management of breast cancer.]

Veronesi U *et al.* Quadrantectomy versus lumpectomy for small size breast cancer. *European Journal of Cancer* 1990: **26**: 671–3. [Trial of breast conservation in treatment of breast cancer.]

Wilson RE *et al.* The 1982 National Survey of carcinoma of the breast in the United States by the American College of Surgeons. *Gynecology and Obstetrics* 1984; **159**: 309–18. [Correlation between nodal positivity and prognosis in patients with breast cancer.]

21.3 Abnormalities of the male breast

P. Jane Clarke and Linda Hands

Gynaecomastia
Clinical and histological appearance
Physiological (primary) gynaecomastia
Pathological (secondary) gynaecomastia
Differential diagnosis
Investigations
Treatment options
Breast carcinoma
Presentation
Histology
Aetiology
Diagnosis
Treatment
Prognosis
Further reading

Gynaecomastia

True gynaecomastia is defined as the enlargement of ductal and stromal tissue of the male breast. Although a common condition, often with a physiological basis, it is important to consider the possibility of underlying disease. The ultimate mechanism for breast development is usually an increase in the oestrogen/androgen ratio, due either to a relative increase in oestrogen levels or to a relative decrease in circulating androgens. In pseudogynaecomastia the visual appearance may suggest the presence of breast tissue but palpation and histological examination reveals the swelling to consist of fat.

Clinical and histological appearance

In order to make a clinical diagnosis of gynaecomastia most authorities state that a firm disc of breast tissue at least 2 cm in diameter should underly the nipple. Approximately 90 per cent of cases are bilateral, although one side may be affected several months before the other.

The pathological abnormalities present will depend on the duration of the condition. Histological examination shows that the enlargement is due to proliferation of the periductal stroma with a variable degree of ductal maturation, and lobular development can also occur. In gynaecomastia of many months duration fibrous tissue replaces most of the breast tissue.

Physiological (primary) gynaecomastia

Physiological gynaecomastia is usually classified as neonatal, pubertal, or senescent, although the condition is best viewed as a continuum through life, with a pubertal peak, a fall in late teenage years, and then a steady increase with advancing age (Table 1).

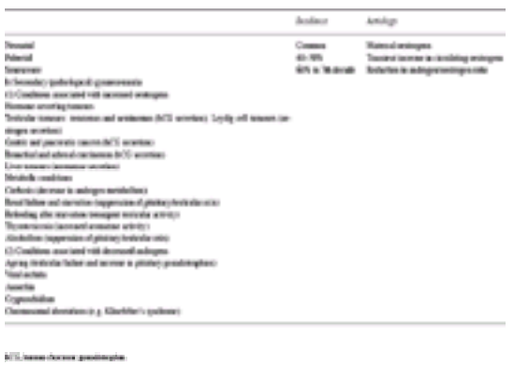


Table 1 Hormone inbalance in gynaecomastia

The normal testis produces testosterone and small quantities of oestrogens. The adrenal gland also produces weak androgens, the bulk of which are metabolized by the liver. Excess adrenal androgens and some testosterone are aromatized to oestrogens in the peripheral tissues, particularly fat. At puberty a surge of gonadotrophins induces testicular activity. Oestrogen production reaches adult levels before that of testosterone, thus increasing the oestrogen/androgen ratio, sometimes enough to stimulate breast development. This usually lasts a matter of months and in almost 75 per cent of cases it subsides within 2 years. Physiological gynaecomastia later in life is due to waning testicular function with a consequent increase in pituitary gonadotrophin, which favours relatively greater oestrogen production by the testis. This disposition to the development of gynaecomastia with ageing is exacerbated by an increase in body fat, leading to greater peripheral aromatization of androgens to oestrogens.

Pathological (secondary) gynaecomastia

Drugs

Drug-induced breast development is the most common cause of gynaecomastia, especially in those over 50 years of age (Fig. 1). Some drugs may have an oestrogen-like or antiandrogen effect, but in many cases the mechanism behind drug-associated gynaecomastia is unknown. Use of drugs such as heroin and cannabis is also known to cause gynaecomastia, as is anabolic steroid abuse in body builders. A list of commonly used drugs known to be associated with gynaecomastia is shown in Table 2.



Fig. 1. Gynaecomastia due to drug treatment.

Cardiovascular drugs	*Antimicrobial drugs*	*Psychotropic drugs*	*Others*
Atenolol	Isoniazid	Diazepam	Cimetidine
Captopril	Ketoconazole	Tricyclic	Cannabis
Digoxin	Metronidazole	antidepressants	Heroin
Enalapril	Miconazole		Oestrogens
Methyldopa			Cyproterone
Nifedipine			
Spironolactone			
Verapamil			

Table 2 Drugs associated with gynaecomastia

Hormonal inbalance

Tumours (especially testicular teratomas), metabolic disease, and hypogonadism can all result in gynaecomastia, hence the importance of investigating the patient for underlying disease when the condition has no obvious physiological or drug-related cause (Fig. 2). Relevant conditions and the probable mechanisms behind breast development are shown in Table 1.



Fig. 2. Gynaecomastia due to a feminizing tumour of the testicle.

Differential diagnosis

In the elderly patient with unilateral gynaecomastia, carcinoma of the breast must always be considered in the differential diagnosis. Pseudogynaecomastia due to local fat deposition with no breast tissue development results in a soft breast lacking the firm, well-defined appearance of true gynaecomastia. It occurs in obese individuals and is not associated with the potential implications already discussed. Very rarely, conditions relatively more common in the female breast, such as fibroadenoma, fibrocystic change, and phylloides tumour, have been described in men and may need to be included in the differential diagnosis.

Investigations

Other than neonatal gynaecomastia, breast development before puberty is always pathological and warrants investigation. In pubertal gynaecomastia, if growth indices are appropriate then no investigations are indicated. Approximately 50 per cent of the remaining patients with gynaecomastia will have a pathological or iatrogenic cause. A careful drug history is essential and a careful clinical examination should be performed to assess for evidence of liver and thyroid dysfunction. The testes should be examined for atrophy or tumour, and suggested investigations are outlined in Table 3. In elderly patients, malignancy should be excluded in unilateral disease by mammography and fine needle aspiration biopsy, especially in the absence of a clear aetiological factor and if the swelling is firm and eccentric. If any residual doubt remains, open biopsy should be performed. A chest radiograph is also important in this age group to exclude a bronchial neoplasm.

Pubertal	None if growth otherwise normal
Under 40 years	Liver function tests
	Thyroid function tests
	Chest radiograph
	Testicular ultrasound
	Hormone profile (hCG, LH, testosterone, and oestrogen)
Elderly patients	Chest radiograph
	Liver function tests
	Mammography and fine needle aspiration to exclude neoplasia

hCG, human chorionic gonadotrophin; LH, luteinizing hormone.
*The clinician will be guided by history and clinical signs as to which tests are appropriate.

Table 3 Investigation of gynaecomastia*

Treatment options

Medical treatment

Gynaecomastia requires no treatment unless it causes serious discomfort or embarrassment and often the patient with physiological gynaecomastia needs only reassurance. Any underlying pathology must be treated and responsible drugs stopped if possible, although this does not guarantee resolution of breast development if it has been present for some time.

The results of hormone manipulation have been disappointing and, in general, studies have involved small numbers of patients and have been poorly controlled, making them of limited value in a condition which may resolve spontaneously.

In cases of profound androgen deficiency, exogenous testosterone may be of benefit, but in less severe cases this may be converted to oestrogen. Androgens which cannot be aromatized have been developed to address this problem. Tamoxifen (20 mg daily) hasbeen reported to be of help in some cases of pubertal and middle-aged breast development, and clomiphene (100 mg twice daily) has also been used in pubertal breast development. Both drugs are thought to act locally as anti-oestrogens.

The use of danazol (300 to 600 mg/day) has been more widely reported. Some studies report its main benefit to be in reduction of pain and discomfort, whereas others have reported reduction in breast size with fewer patients requesting surgery at the end of a 3-month period of treatment.

Surgical treatment

As medical treatment is of limited value in the majority of patients, surgery remains the mainstay of treatment for symptomatic or unsightly gynaecomastia. A grading system that is frequently used to assess the extent of breast enlargement has been developed (Table 4). As any surgical intervention is essentially for cosmetic reasons it is important to leave the nipple/areola in the correct position, symmetrical with the other side, with minimal obvious scarring. Most patients with grade I and II disease have relatively modest breast development and a satisfactory subcutaneous mastectomy can be performed through an incision following the inferior margin of the areola (Fig. 3(a)). It is important to leave a smooth contour to the breast and avoid a central crater under the areola as this can be unsighted. This can be achieved by leaving a subareolar disc of breast tissue. Minor skin redundancy is not a problem as there is a tendency for skin contraction to occur. This technique can be modified in cases of unilateral breast swelling in which there is areola asymmetry by performing a circumareolar incision, de-epithelializing a concentric ring of peripheral areola on the affected side, and carefully resuturing to take account of the excess circumference which results. This approach can also be used to give better access for the mastectomy in a larger breast (Fig. 3(b)).

Grade I	Minor breast enlargement, no skin redundancy
Grade IIa	Moderate breast enlargement, no skin redundancy
Grade IIb	Moderate breast enlargement with minor skin redundancy
Grade III	Gross breast enlargement with skin redundancy so as to simulate a female pendulous breast

*Simon et al., Plastic and Reconstructive Surgery (1973); 51: 48–52.

Table 4 Classification of gynaecomastia*

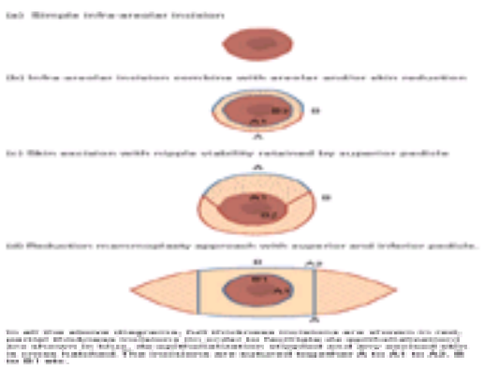


Fig. 3. Surgical approaches to gynaecomastia. (a) Simple infra-areolar incision. (b) Infra-areolar incision combined with areolar and/or skin reduction. (c) Skin excision with nipple viability retained by superior pedicle. (d) Reduction mammoplasty approach with superior and inferior pedicle. In all the above diagrams, full thickness incisions are shown in red, partial thickness incisions (in order to facilitate de-epithelialization) are shown in blue, de-epithelialized skin is stippled, and any excised skin is cross-hatched. The incisions are sutured together A to A1 to A2, B to B1 etc.

With grade III disease, breast development is so marked that excess skin has to be removed. Several techniques have been developed to do this while preserving the areola in position with an intact blood supply. The technique described above can be further modified and the mastectomy performed through a circumareolar incision with removal of the appropriate amount of skin. Nipple viability is achieved by de-epithelialization of a superior pedicle (Fig. 3(c)). More complex procedures are based on reduction mammoplasty techniques, initially developed for the female breast, whereby a horizontal elliptical incision is made and the skin triangles medial and lateral to a central, nipple-containing strip are removed, leaving a vertical nipple flap with excellent viability (Fig. 3(d)). The disadvantage of this technique is that it leaves a horizontal incision either side of the areola which may be conspicuous in the man without a covering of chest hair.

Liposuction can be a useful adjunct in the above operations, especially in the fatty breast, in the attempt to achieve an optimal chest profile. Its use has been reported as a technique in its own right in male breast reduction, but as it removes only fat and not breast tissue it cannot be the complete surgical answer in true gynaecomastia.

Breast carcinoma

About 1 per cent of all breast cancers occur in men. The incidence is slightly higher in the Middle East and Africa, and the male/female ratio is higher among black than among white populations. Overall, breast malignancy accounts for approximately 0.5 per cent of all male cancers.

Presentation

Male breast cancer occurs most often in the seventh decade. Seventy-five per cent of patients present with a breast mass, the others with a bloodstained nipple discharge, nipple distortion, ulceration, or axillary lymphadenopathy. Despite the ease of examining the breast in men in comparison with women, these tumours tend to present late, probably because of lack of concern regarding male breast symptoms. Inevitably most of these tumours arise beneath the areola, often infiltrate both skin and muscle, and axillary spread is common by the time of presentation.

Histology

Most male breast cancers are ductal adenocarcinomas of no special type. Lobular tumours are very rare; papillary tumours are also uncommon but occur more often than in women. Over 80 per cent of male breast cancers are oestrogen- or progesterone-receptor positive. Secondary deposits can occur from tumours elsewhere, especially from the prostate.

Aetiology

There is strong evidence for a hormonal basis to the disease, with several studies showing significantly raised serum levels of oestrogens in men with breast carcinoma. Being overweight in early adulthood is a risk factor for developing the condition. Obesity is linked to increased aromatase activity which increases the conversion of adrenal androgens to oestrogens. Male breast tumours are very common in patients with Klinefelter's syndrome, a condition in which there is impaired testicular function. Other conditions associated with an increase in the ratio of oestrogen to testosterone include long-standing hypopituitarism and hypogonadism, mumps orchitis in adulthood, and testicular injury; breast cancer is seen in association with these diseases. Although similar influences have been described in the development of gynaecomastia, there is no evidence that gynaecomastia itself predisposes to the development of breast carcinoma. There is a positive correlation between male breast cancer and prostate cancer.

Two gene defects have been associated with a predisposition to the development of male breast cancer, namely BRCA2 and the androgen receptor (AR) gene. As in female familial breast cancer, the cases tend to present at an earlier age than the sporadic disease. A higher proportion of male breast cancers seem due to recognized germ line mutations compared with female breast cancer (20 per cent compared with less than 5 per cent, respectively). Previous radiotherapy to the chest wall has occasionally been associated with the development of male breast cancer.

Diagnosis

The diagnosis of breast cancer in the male is reached in the same way as for a female patient, using triple assessment (clinical examination, cytology or histopathology, and radiology).

Treatment

The proximity of the tumour to the chest wall and the frequency of axillary spread have led to treatment by either radical surgery or simple mastectomy and radiotherapy. Unfortunately, the rarity of this tumour has precluded a prospective study comparing different surgical treatments.

In view of the high proportion of tumours which are oestrogen-receptor positive, anti-oestrogen therapy is now the most frequently used first-line adjuvant treatment and in many centres this has replaced orchidectomy. Tamoxifen is the drug of choice, and this has been shown to result in a significant improvement in disease-free survival when compared with historical controls (unfortunately there are no randomized studies due to the rarity of the disease). Side-effects are similar to those in the female, with reduced libido being the most common, with weight gain and hot flushes also occurring. On recurrence a further response can still be elicited by administration of antiandrogens, luteinizing hormone-releasing hormone (LHRH) analogues, or progestogens. Many patients also respond to therapy with cytotoxic agents, but the use of these drugs is less frequent than in women because of the older age of the affected population.

Prognosis

Overall about 50 per cent of patients die of their disease in the 5 years following diagnosis. Stage for stage, the prognosis seems similar to that in women. If no axillary nodes are involved then about 80 per cent survive 5 years, but only 30 to 40 per cent of those with axillary spread survive this long. As in the female disease, the prognostic indicators of most value are nodal status, grade, and tumour size.

Further reading

Braunstein GC. Gynecomastia. *New England Journal of Medicine* 1993; **328**: 490–5. [An excellent review and reference list.]

Carlson HE. Gynecomastia. *New England Journal of Medicine* 1980; **303**: 795–9. [A good review.]

Memon MA, Donohue JH. Male breast cancer. *British Journal of Surgery* 1997; **84**: 433–5. [Good leading article.]

Ribeiro G. Male breast carcinoma. *British Journal of Cancer* 1985; **51**: 115–19. [Good review and discussion.]

Ward CM, Khalid K. Surgical treatment of grade III gynaecomastia. *Annals of the Royal College of Surgeons of England* 1989; **71**: 226–8. [Good explanation of operative technique.]

22.1 Dysphagia

J. Shapiro

Physiology
Oral phase
Pharyngeal phase
Esophageal phase
Pathophysiology
Oral phase dysfunction
Pharyngeal phase dysfunction
Aspiration
Factors contributing to dysphagia
Disorders associated with dysphagia
History
Physical examination
Videofluoroscopy
Barium swallow
Endoscopy
Other techniques
Treatment
Further reading

Physiology

For purposes of study, the swallow has been divided into three phases: oral, pharyngeal, and esophageal.

Oral phase

Before the onset of the oral phase, the major portion of a food bolus is formed into a cohesive mass and is held between the anterior tongue and the hard palate. The soft palate is pulled anteriorly and rests against the back of the tongue, which is slightly elevated, closing the oral cavity. The oral phase begins when the tongue moves the bolus posteriorly and ends when the bolus passes the anterior tonsillar pillars. During this phase, the tongue elevates in an anterior to posterior direction. An anterior labial seal and a lateral buccal seal keep the bolus in proper position (Table 1).

Begins:	tongue moves bolus posterior
Ends:	bolus passes anterior pillars
Time:	~1 s
Components:	tongue elevates anterior to posterior
	tongue forms central groove
	labial seal anterior
	buccal seal lateral
Control:	peripheral nervous system, voluntary

Table 1 Oral phase

Pharyngeal phase

The pharyngeal phase begins when the food bolus passes the anterior tonsillar pillars and ends when the bolus passes through the upper esophageal sphincter into the esophagus. During this phase the palate elevates and retracts, closing the nasopharynx. The laryngeal valves close and the suprahyoid muscles elevate the larynx under the base of the tongue. The bolus is propelled through the pharynx by the backward motion of the base of the tongue and contraction of the pharyngeal constrictor muscles. When the upper esophageal sphincter opens it creates a negative pressure in the hypopharynx, facilitating passage of the bolus into the esophagus. During the pharyngeal phase respiration is inhibited. The exact trigger for initiation of this phase of the swallow is unknown; it may be related to the position of the base of the tongue (Table 2).

Begins:	bolus passes anterior pillars
Ends:	bolus passes through upper esophageal sphincter into esophagus
Time:	~1 s
Components:	velum elevates and retracts—nasal passage closed
	bolus propelled through pharynx
	larynx closes and elevates
	respiration inhibited
	upper esophageal sphincter relaxes
Control:	peripheral nervous system, involuntary

Table 2 Pharyngeal phase

Esophageal phase

The esophageal phase begins when the bolus enters the esophagus and ends when it passes through the lower esophageal sphincter into the stomach. A sequential peristaltic wave in the esophagus propels the bolus. The lower esophageal sphincter relaxes as the bolus passes into the stomach (Table 3).

Begins:	bolus enters esophagus
Ends:	bolus passes through lower esophageal sphincter into stomach
Time:	8–30 s
Components:	sequential peristaltic wave propels bolus
	relaxation of lower esophageal sphincter
Control:	autonomic nervous system, involuntary

Table 3 Oesophageal phase

The upper esophageal sphincter is a zone of high pressure between the pharynx and the esophagus that prevents air from entering the esophagus and helps prevent esophageal contents from refluxing into the pharynx. The muscles which form the upper esophageal sphincter are the cricopharyngeus and either the lower fibers of the inferior pharyngeal constrictor or the upper fibers of the cervical esophagus.

At rest the sphincter is closed, and closure is maintained by both tonic contraction of the intrinsic sphincter muscle and passive elastic forces. Inhibition of the intrinsic musculature contraction of the sphincter and contraction of the suprahyoid muscles creates anterosuperior displacement of the cricoid cartilage, thereby opening the sphincter (Fig. 1). This bolus itself also distends the upper esophageal sphincter.

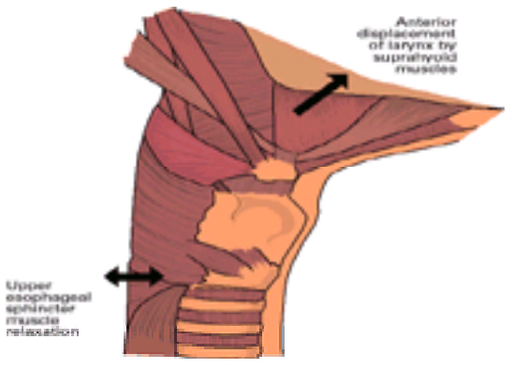


Fig. 1. Opening of the upper esophageal sphincter requires relaxation of the upper esophageal muscles and anterior displacement of the larynx by contraction of the suprahyoid muscles.

Pathophysiology

Dysphagia, or difficulty in swallowing, is a common complaint. For purposes of discussion, abnormalities in swallowing can be divided into dysfunction in each of the separate phases. In reality, these phases are so inter-related that the abnormalities are often in more than one phase.

Oral phase dysfunction

If the tongue holds the bolus too far forward on the palate the result is anterior leakageof the bolus. Weakness of the labial or cheek muscles leads to malpositioning of the bolus either out of the oral cavity or into the lateral sulci. Any decreased sensation in the oral cavity makes it much more difficult to manipulate the bolus into proper position. Tongue dysfunction may be due to either tethering of the tongue or decreased tongue movement. Many patients have an incoordination of the normal anterior to posterior tongue motion. This incoordination can be a devastating problem because it leads to difficulty in initiating the pharyngeal phase of the swallow.

Pharyngeal phase dysfunction

In the pharyngeal phase, limitation in palatal motion leads to nasal regurgitation of the bolus. A frequent abnormality is decreased propulsion of the bolus by the pharynx. This may be secondary to poor function of the tongue, decreased pharyngeal constrictor muscle activity, or dysfunction in opening of the upper esophageal sphincter. Decreased laryngeal elevation may compromise the opening of the upper esophageal sphincter; this may also result in poor laryngeal protection due to reduced epiglottic deflection or because the larynx is not positioned properly under the base of the tongue. Failure of the vocal cords to close during the swallow will also leave the larynx unprotected as will a delay in the onset of the pharyngeal phase of the swallow.

Aspiration

Aspiration is a symptom of dysphagia; it can result from dysfunction in either the oral or pharyngeal phases. Aspiration itself is divided into three components: that occurring before, during, or after the swallow. Aspiration before the swallow is caused by presentation of the bolus to the pharynx before the larynx has closed and elevated, probably because of either poor oral control, causing spillover of the bolus into the pharynx or delay in the onset of the pharyngeal phase of the swallow. Aspiration during a swallow occurs because of abnormalities in the pharyngeal phase such as poor laryngeal elevation or closure, or both. Aspiration after a swallow results from inability of the pharynx to clear the bolus during the pharyngeal phase. This leaves residue in the pharyngeal recesses that is subject to spillover into the pharynx at a time when the larynx is unprotected.

Factors contributing to dysphagia

Insertion of a tracheotomy tube exacerbates dysphagia by limiting laryngeal elevation and decreasing normal vocal fold reflexes. Hyposalivation makes it difficult to initiate a swallow. Radiation therapy often causes hyposalvation and may also lead to tissue edema and fibrosis, affecting both the oral and pharyngeal phases of the swallow. The effect of medications on swallowing function is not well known but may play an important part in dysphagia.

Disorders associated with dysphagia

Many disorders can cause dysphagia (Table 4). Neurologic deficits may be associated with cerebral vascular accidents, head trauma, or cranial nerve neuropathies, and often result in gross swallowing incoordination. Muscular disorders such as polymyositis can affect the laryngopharyngeal musculature, resulting in the severely reduced ability to propel a bolus. Oculopharyngeal muscular dystrophy is a rare genetic disorder which affects people of French Canadian descent. The two symptoms, bilateral ptosis and dysphagia, usually present after age 50. There is progressively poor bolus propulsion and aspiration.

Table 4 Disorders associated with dysphagia

Zenker's diverticulum is an outpouching of the pharynx, usually between the upper border of the cricopharyngeus muscle and the lower border of the inferior constrictor muscle (Fig. 2). Impaired opening of the upper esophageal sphincter may contribute to its etiology. In this condition, the bolus collects in the diverticulum and spills over

into the unprotected larynx, causing aspiration.

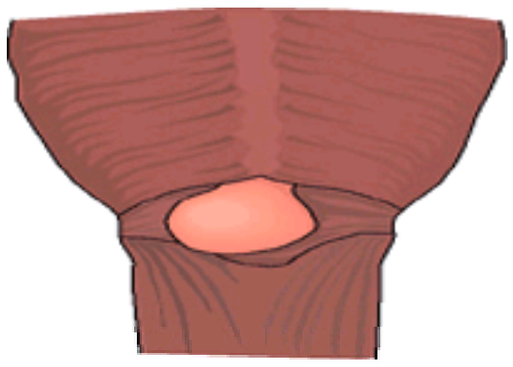


Fig. 2. Zenker's diverticulum. Most common location is between the upper border of the cricopharyngeus and the lower border of the inferior constrictor.

Cricopharyngeal achalasia is a rare entity defined as isolated incomplete relaxation of the cricopharyngeus muscle without dysfunction of the other pharyngeal constrictor muscles (Fig. 3). Some experts doubt that there is isolated upper esophageal sphincter dysfunction. In most instances there is other evidence of pharyngeal muscle weaknesses seen in patients with impaired upper esophageal sphincter opening.

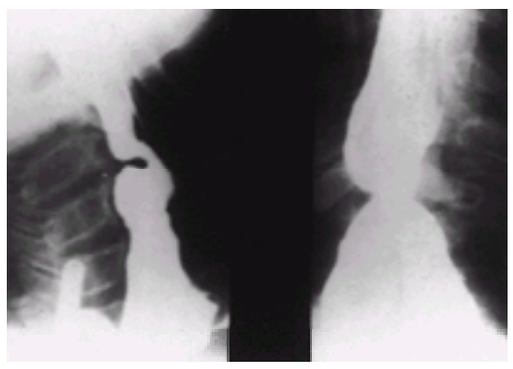


Fig. 3. Prominence of the cricopharyngeal muscle seen on anterior–posterior and lateral views of the barium swallow.

Structural lesions, such as esophageal webs, rings, or strictures, often cause solid food dysphagia.

History

A precise history is essential when evaluating a patient with dysphagia. The distinction between dysphagia and odynophagia must be made: odynophagia is pain on swallowing and is often associated with either neoplasia or infection. The presence of referred pain, such as otalgia, may be caused by a hyopharyngeal lesion.

When a patient complains of dysphagia, specific symptoms must be known and the consistencies of food or liquid which exacerbate the symptom must be defined. Patients may describe a bolus as being ‘stuck' or ‘caught'. The location of this sensation and the kind of maneuver used by the patient to dislodge the bolus are all helpful points. Inability to pass the bolus requiring regurgitation of the bolus is common in patients with a structural esophageal lesion.

Aspiration in a patient with normal laryngotracheal sensation causes coughing; the timing of this cough can help to indicate the timing of the aspiration. For example, patients with Zenker's diverticulum often aspirate anywhere from several minutes to some hours after the meal, when the contents of the filled diverticulum spill over into the unprotected pharynx. Nasal regurgitation and voice change may indicate problems with palatal and vocal cord function, respectively. Change in diet and weight loss can reflect the degree of severity of the swallowing abnormalities.

Because dysphagia is often associated with other systemic disorders, a thorough clinical history must be obtained. Neurologic abnormalities such as change in mental status, dysarthria, and diplopia are all potentially relevant. Gastrointestinal symptoms might include dyspepsia or acid regurgitation. Rheumatologic manifestations include diffuse muscle weakness or skin disorders. Psychosocial factors that may have either precipitated or exacerbated the dysphagia should be investigated.

Physical examination

Physical examination should include the muscles of facial expression, tongue mobility and strength, and palatal elevation and sensation. The correlation between the gag reflex and dysphagia is not known and may not be clinically important. A full examination of the nasopharynx, hypopharynx, and larynx is important to determine whether there are any structural or neurologic lesions. Particularly important is the presence of pooled secretions in the vallecula or piriform sinuses and the mobility of the vocal cords.

Videofluoroscopy

Videofluoroscopy is a radiologic technique used for evaluating dysphagia. The patient is seated in an upright position and given small quantities of barium in liquid, paste, and solid forms. A camera is focused on the oral and pharyngeal regions throughout the swallow, recording the images on videotape. If a swallowing abnormality is seen, various maneuvers such as change in bolus texture, head position, and breathing, are used to attempt to correct the abnormality. The specific events evaluated during fluoroscopic examination include oral transit time, pharyngeal transit time, the presence and degree of residue in the pharyngeal recesses, the extent of laryngeal elevation, the presence and timing of aspiration, degree of upper esophageal sphincter opening, the effect of various bolus textures and position on the swallow, and the integrity of esophageal peristalsis.

Swallowing videofluoroscopy is indicated for almost all symptomatic dysphagia patients who are able to co-operate to some degree. The small quantity of barium used carries minimal risks, even for a patient with gross aspiration. There have been no reported complications from this test. The test is also particularly helpful in any patient in whom aspiration is suspected, whether or not the patient is subjectively aware of the problem. Videofluoroscopy does not accurately demonstrate structural abnormalities, and subtle esophageal abnormalities may be missed.

Barium swallow

The barium swallow is a cineradiographic study of the esophagus. Following administration of a large quantity of liquid barium to the supine patient, the fluoroscopy camera follows the bolus from the upper esophageal sphincter to the lower esophageal sphincter. Suspected reflux can be accentuated by procedures such as the Valsalva maneuver.

A barium swallow should be performed in any patient with odynophagia. If the laryngoscopic examination is normal, the barium study may reveal the source of the odynophagia. If the laryngoscopy reveals any infection, such as candidiasis or a tumor, the extent of esophageal disease needs to be assessed. Any patient with dysphagia for solid foods should also undergo a barium swallow as a structural esophageal lesion may not be seen on videofluoroscopy. A standard barium swallow is often used as one method for detecting gastroesophageal reflux. Patients whose symptoms localize to the lower neck or sternum are likely to have esophageal disease and they should also undergo a barium swallow. Any unexplained dysphagia in a patient in whom a swallowing videofluoroscopy is normal should be further studied

with a standard barium swallow.

Endoscopy

Endoscopic examination performed under anesthesia might include direct laryngoscopy, nasopharyngoscopy, oropharyngoscopy, esophagoscopy, and bronchoscopy. In patients with head and neck cancer it is important to delineate the exact site of the lesion and also to rule out any other synchronous primary lesions. Flexible fiberoptic esophageal endoscopy is indicated for patients with suspected esophageal disease. Endoscopy may be combined with a dilatation in patients with a known esophageal stricture.

Other techniques

Techniques such as scintigraphy, sonography, and electromyography are being developed and refined so that they may have more applicability in the clinical sphere.

Treatment

Dysphagia is sometimes relieved by treating the underlying disorder with specific drugs. For example, Parkinson's disease may respond to levodopa and polymyositis may respond to oral steroids. Esophageal lesions such as strictures or webs are treated by dilatation.

The indications for upper esophageal sphincterotomy (often called a cricopharyngeal myotomy) are varied. Diffuse pharyngeal muscle dysfunction such as is seen in oculopharyngeal muscular dystrophy responds relatively well to a cricopharyngeal myotomy. The benefit of a myotomy for a patient with neurogenic dysphagia where there is swallowing incoordination is much less clear.

The surgical approach to treating a Zenker's diverticulum can be either external or endoscopic. The external approach is performed through a left cervical incision and involves either a diverticulopexy (i.e. suturing the diverticular sac superiorly to the posterior pharynx) or a diverticulectomy. In either case, an upper esophageal sphincterotomy, also called a cricopharyngeal myotomy, must be performed or the diverticulum will recur. The endoscopic approach involves sectioning the cricopharyngeus muscle with either a laser or an endoscopic stapler.

Patients with disorders that do not respond to medication or surgery may benefit by swallowing therapy. After the swallowing abnormalities are identified on videofluoroscopy, various maneuvers are tried to see which facilitates the most normal swallow; for example, patients with a supraglottic laryngectomy, who often aspirate postoperatively, must be taught a ‘supraglottic swallow', which involves flexing the neck, holding the breath on inspiration, swallowing, coughing, and then exhaling.

If none of the above treatment options results in the ability to swallow without aspiration and to maintain nutritional status, the patient requires a non-oral feeding route, such as a gastrostomy tube or a jejunostomy tube.

Further reading

Logman J. *Evaluation and treatment of swallowing disorders*. College-Hill Press, Inc., San Diego, 1983. [Discusses the approach to a dysphagia patient from the perspective of a swallowing therapist. Particular emphasis is placed on the interpretation of swallowing videofluoroscopy.]

McConnel FMS. Analysis of pressure generation and bolus transit during pharyngeal swallowing. *Laryngoscope* 1988; **98**: 71–8. [Details the various forces acting on the passage of a bolus from the oropharynx into the esophagus.]

Shapiro J, Goyal R. Disorders of the upper oesophageal sphincter. In: Fried MP, ed. *The larynx*, pp. 293–317. Little Brown and Co., Boston, 1988. [Details the anatomy, physiology, and pathophysiology of the pharyngoesophageal segment (PES or UES) as well as the possible therapeutic options for patients with dysphagia secondary to PES dysfunction.]

Sokol EM, Heitman P, Wolfe BS. Simultaneous cineradiographic and manometric study of the pharynx, hypopharynx and cervical esophagus. *Gastroenterology* 1966; **51**: 960–74. [One of the first descriptions of the correlation between manometry and videofluoroscopy.]

Wouters B, van Overbeek JJM. Endoscopic treatment of the hypopharyngeal (Zenker's) diverticulum. *Hepatogastroenterology* 1992; **39**: 105–8.

22.2.1 Reflux disease and hiatus hernia

Reginald V. N. Lord and Tom R. Demeester

Gastroesophageal reflux disease
Clinical features
Pathogenesis
Diagnostic evaluation
Treatment
Techniques of antireflux operations
Summary guidelines for the treatment of gastroesophageal reflux disease
Hiatus hernia
Definition and classification
Physiologic herniation
Prevalence of pathologic herniation
Pathophysiology
Symptoms
Diagnosis
Treatment
Further reading

Gastroesophageal reflux disease

Clinical features

Reflux of small amounts of gastric juice into the esophagus is a normal physiologic event. Gastroesophageal reflux disease occurs when this reflux becomes excessive. This may be manifested by symptoms, by endoscopic or histologic esophagitis, or by measurement of increased esophageal acid exposure on 24-h pH monitoring.

Typical symptoms

The principal typical symptom of gastroesophageal reflux disease is heartburn (pyrosis), which is a retrosternal burning discomfort or pain. Heartburn may be precipitated by bending over or lying flat, especially if the lower esophageal sphincter is mechanically defective. Regurgitation of either gastric juice or food is another typical symptom of gastroesophageal reflux disease. Acid regurgitation is typically sudden and effortless, may be associated with belching, and tastes sour or bitter (‘acid brash'). Forceful regurgitation, in contrast, usually consists of bland-tasting food which has not entered the stomach, and is typical of an esophageal body motor disorder or esophageal obstruction.

Both heartburn and regurgitation are more common within the first hours after meals. Refluxogenic stimuli include large, fatty, or spicy meals, chocolate, onions, peppermint, garlic, nicotine, fruit juices, and drinks that are carbonated or contain alcohol or caffeine. Rapid production of saliva in response to a reflux episode may be perceived as a volume of slightly salty or tasteless fluid in the mouth, a symptom known as ‘waterbrash'. Reflux dyspareunia, or heartburn and regurgitation during intercourse, is not uncommon in females with gastroesophageal reflux disease. In children, the adoption of unusual postures can be a symptom of reflux and is referred to as Sandifer's syndrome.

Dysphagia is a typical symptom of advanced gastroesophageal reflux disease. It can be caused by tissue edema associated with acute severe esophagitis, a distal esophageal stricture, or by loss of esophageal compliance or motility due to chronic inflammatory changes in the esophageal wall. The symptoms of heartburn and regurgitation may improve after the development of a stricture if the stricture significantly interferes with reflux. Dysphagia may also be due to the presence of a large (more than 5 cm) hiatus hernia. Malignancy or a named esophageal body motility disorder needs to be excluded in patients with dysphagia, even if a diagnosis of gastroesophageal reflux disease is likely. Dysphagia due to gastroesophageal reflux disease usually develops slowly, and patients may unconsciously adjust their eating habits to account for the difficulty with swallowing. Consequently, it is important to take a detailed swallowing history in patients with suspected gastroesophageal reflux disease, questioning them as to whether they cut food into small pieces, always finish eating after others, or avoid eating in company. In general, patients with obstructive dysphagia localize the site of obstruction either at, or proximal to, the anatomic level of the obstruction. Pain on swallowing (odynophagia) occurs in up to 50 per cent of patients with reflux esophagitis. Many patients initially complain of ‘indigestion', in which case further enquiry is needed to determine more precisely the specific symptom or symptoms experienced.

Atypical symptoms

Atypical and extraesophageal symptoms of gastroesophageal reflux disease include chest pain, epigastric pain, bloating, and pulmonary symptoms. In some patients, these are the only symptoms of gastroesophageal reflux disease. Non-cardiac chest pain caused by gastroesophageal reflux disease has been found in up to 50 per cent of patients with chest pain and normal coronary angiography. Although the pain can be induced by exercise, thus masquerading as angina pectoris, the lack of a relationship to exercise helps differentiate most cases of reflux-induced chest pain from true angina.

Airway symptoms of gastroesophageal reflux disease include chronic hoarseness (the Cherry–Donner syndrome), chronic cough, and symptoms of asthma such as wheezing and shortness of breath. Episodic or chronic aspiration can cause pneumonia, lung abscess, and interstitial pulmonary fibrosis. In 6 to 10 per cent of patients with chronic cough, gastroesophageal reflux disease is the underlying cause. A chest radiograph is usually normal in these patients. Asthma-like symptoms are more likely to be caused by gastroesophageal reflux disease in patients with adult-onset, non-allergic asthma, and in patients who have typical symptoms of gastroesophageal reflux disease as well as asthma symptoms. At least 50 per cent of patients with asthma who have typical symptoms of gastroesophageal reflux disease have abnormal esophageal acid exposure on 24-h pH monitoring. Gastroesophageal reflux can be secondary to asthma in many of these patients, because the intrathoracic pressure alterations caused by the increased work of breathing increase the peritoneopleural pressure gradient, thus promoting reflux from the stomach into the esophagus.

Patients who reflux into the proximal esophagus above the level of the upper esophageal sphincter (laryngopharyngeal reflux) may have a hoarse voice, voice fatigue, chronic throat clearing, dental erosions, and reflux laryngitis. Laryngeal pathology in these patients commonly includes erythema and edema of the vocal cords, but ulceration of the cords and arytenoids, vocal nodules, granulomas, and even carcinoma may result from laryngopharyngeal reflux. Direct reflux of gastric fluid into the lungs is not necessary for the development of airway symptoms related to gastroesophageal reflux disease, because the symptoms may be provoked by distal esophageal acid exposure via an esophagotracheobronchial cough reflex.

Airway symptoms due to gastroesophageal reflux disease occur commonly at night, when they can be distressing or even frightening for the sufferer. Nocturnal reflux, with choking, coughing, shortness of breath, and wheezing, will often wake the patient from sleep. In the morning, the patient may complain of a foul taste in the mouth, halitosis, a hoarse voice, or tiredness.

Physical examination is not usually helpful in the diagnosis of gastroesophageal reflux disease, but is important for detecting signs of other disease. Examination of the teeth can show erosions and inspection of the arytenoids and vocal cords may show erythema, with or without ulceration.

Pathogenesis

Prevalence and natural history

Gastroesophageal reflux disease is a common disease that accounts for the great majority of esophageal pathology. Approximately 10 per cent of adults in Western societies experience heartburn daily, and more than one-third have occasional heartburn. Gastroesophageal reflux disease is an episodic disease in some patients, but in most patients it is a chronic disease, requiring long-term medication.

Risk factors for gastroesophageal reflux disease

A recent study from this department investigated risk factors for gastroesophageal reflux disease in 1147 patients with symptoms suggestive of reflux disease. No patient had a history of a previous foregut operation, and all underwent distal esophageal 24-h pH monitoring and completed a detailed questionnaire. Significant epidemiologic or lifestyle risk factors for reflux disease were male sex, older age, white non-Hispanic ethnic group, higher body mass index, and tobacco smoking. Alcohol use was a risk factor in those with an average consumption of more than 30 g of alcohol per day. The most significant risk factors were physiologic factors indicating a structurally defective lower esophageal sphincter, as will be discussed later in this chapter. This study and studies from other institutions found a significant inverse association between gastroesophageal reflux disease and the presence of gastric *Helicobacter pylori* infection, indicating that colonization with this organism may have a protective effect for the development of gastroesophageal reflux disease. The mechanism by which *H. pylori* infection has this effect is unclear, but presumably involves a neutralizing effect on gastric acidity.

The normal mechanisms that protect the esophagus from excessive exposure to injurious gastric contents are illustrated schematically in Fig. 1. As shown, the antireflux mechanism can be classified as having pump, valve, and reservoir components. Gastroesophageal reflux disease results from a defect in one or more of these components, with the risk of complications progressively increasing with the loss of each additional component.

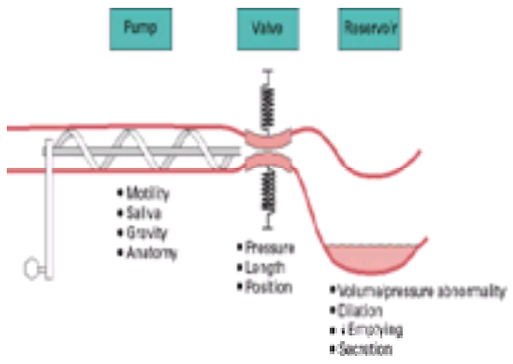


Fig. 1. Mechanical model of the antireflux mechanism, showing the esophageal body acting as a pump, the lower esophageal sphincter as a valve, and the stomach as a reservoir.

Valve

The most important antireflux component is the valve, or lower esophageal sphincter. This can be identified manometrically as a 2- to 4-cm asymmetric zone of high pressure in the distal esophagus. It is normally closed, opening to allow ingested solids, liquids, and saliva to pass into the stomach. It also opens normally for short periods after meals and at other times to allow venting of gas from the distended stomach, and it opens for vomiting. The common event for virtually all episodes of gastroesophageal reflux, whether physiologic or pathologic, is the loss of the normal resistance imposed by the sphincter to the flow of gastric juice from an environment of higher pressure, the stomach, to an environment of lower pressure, the esophagus. In normal subjects or patients with mild disease this loss is transient. In patients with severe disease this loss is usually permanent and is manifested by a reduced or non-existent high-pressure zone.

An anatomically distinct sphincter at the gastroesophageal junction has been difficult to demonstrate in humans, but the increased thickness of the muscle layers and the specialized arrangement of muscle fibers in this region are considered to be the anatomic manifestation of the lower esophageal sphincter. Compared with the esophageal body, the sphincter has a higher density of neuronal complexes and a more diffuse arrangement of neurons. The intrinsic tone of the sphincter is mediated by myogenic, neurogenic, and mechanical factors. Vagal cholinergic innervation is largely responsible for maintaining resting sphincter pressure in humans. Relaxations of the sphincter are mediated by cholinergic vagal fibers, with nitric oxide and vasoactive intestinal peptide acting as the major postsynaptic inhibitory neurotransmitters. The afferent stimuli causing sphincter relaxation have not been confirmed, but transient common cavity events with equalization of the gastric and esophageal pressure can result from mechanical shortening of the lower esophageal sphincter caused by gastric distension. Transient loss of sphincter pressure is the probable mechanism for the loss of sphincter resistance in patients with early or mild disease. In contrast, permanent loss of sphincter resistance occurs in patients with advanced gastroesophageal reflux disease, resulting in chronic severe symptoms, erosive esophagitis, or Barrett's esophagus. Most of these patients have a persistently hypotensive, mechanically defective sphincter. Consequently, the pressure gradient between the positive-pressure abdominal cavity and the negative-pressure thoracic cavity results in the free flow of gastric juice from the stomach into the esophagus, with prolonged exposure of the esophageal mucosa to acid, pepsin, pancreatic enzymes, and bile.

The pinch-cock action of the crural diaphragm, which closes the esophageal hiatus during short periods of diaphragmatic contraction and raised intra-abdominal pressure, such as occurs during straining or coughing, adds to the effectiveness of the gastroesophageal antireflux barrier. The presence of a hiatus hernia impairs the sphincter function of the crural diaphragm. In addition, the additive pressure effect created by the superimposed lower esophageal sphincter and crural antireflux components is lost in patients with larger hernias, in whom the lower esophageal sphincter and crura are separated.

Pump

The pump antireflux mechanism includes the peristaltic function of the esophageal body, the effect of gravity in the upright position, and the neutralizing effect of saliva. The pump is important for clearance of refluxed material from the esophagus. The peristaltic wave associated with swallowing and relaxation of the upper esophageal sphincter is called primary peristalsis. Secondary peristaltic waves are localized to the esophagus, are independent of oropharyngeal events, and are stimulated by esophageal distension or the presence of air, solids, or fluid in the esophagus. The peristaltic wave is caused by contraction of the esophageal muscle, and closes the esophageal lumen behind a swallowed bolus. In the upright position, a swallowed fluid bolus thus moves down the esophagus ahead of the peristaltic wave.

The esophageal body functions as a worm-drive propulsive pump because of the helical arrangement of fibers within the circular muscle layer. The peristaltic wave generates an occlusive pressure of 30 to 160 mmHg, and the speed of peristalsis is normally 2 to 5 cm/s. Average pressures at the junction of the upper and middle thirds of the esophageal body, where there is a blending of triated and smooth muscles, are lower than in the proximal striated muscle or the distal smooth muscle segments.

Primary peristaltic waves are responsible for volume clearance of refluxed acid. Any acid that remains in the esophagus after volume clearance is neutralized by the alkaline saliva. Patients with xerostomia, impaired esophageal peristalsis, or reflux during sleep, when primary peristaltic waves are infrequent, are more likely to have mucosal injury from gastroesophageal reflux disease because of defective pump function. Peristaltic waves with contraction amplitudes of at least 20 mmHg are required for efficient clearance of supine reflux episodes.

Esophageal body function may also be abnormal because of defective organization of the contraction waves. Abnormal waveforms can be dropped (incomplete), interrupted (skipped segments), or excessively rapid (appearing as simultaneous contractions at more than one level). Esophageal clearance can be impaired by the presence of a hiatus hernia due to the loss of abdominal anchoring of the esophageal body. In one study, complete esophageal emptying was found in only 32 per cent of test swallows in patients with a non-reducing hiatus hernia.

Reservoir

Abnormalities of the gastric reservoir that predispose to gastroesophageal reflux disease include gastric distension, impaired gastric emptying, and gastric acid hypersecretion. Approximately 40 per cent of patients with gastroesophageal reflux disease have delayed gastric emptying. Gastric distension can be caused by ingestion of excessive quantities of air or of fatty and fried foods that delay gastric emptying. Gastric distension causes a mechanical shortening of the lower esophageal sphincter, resulting in a transient loss of the high-pressure zone and reflux episodes. Delayed gastric emptying may also be due to myogenic abnormalities caused by diabetes, diffuse neuromuscular disorders, postviral gastroparesis, or anticholinergic medications. Non-myogenic causes are vagotomy, antropyloric dysfunction, and duodenal dysmotility. A combination of delayed gastric emptying and increased intragastric pressure may be found in some patients with gastric outlet obstruction. Up to 28 per cent of patients with gastroesophageal reflux disease have abnormally increased gastric acid secretion, which may contribute to the development or severity of reflux.

Composition of the refluxate

Acid and pepsin

The gold standard test for the diagnosis of gastroesophageal reflux disease is 24-h esophageal pH monitoring, which measures the amount and duration of acid exposure in the distal esophagus. The importance of measuring acid reflux is demonstrated by the following observations. (1) The reflux symptom of heartburn is produced by the reflux of gastric juice at low pH into the esophagus, and this symptom can be reproduced by the Bernstein acid perfusion test. Furthermore, acid-suppressant drugs effectively relieve heartburn. (2) The severity of mucosal injury correlates well with the duration of esophageal acid exposure. Patients with complicated gastroesophageal reflux disease, defined by the presence of esophagitis, stricture, esophageal body hypomotility or shortening, or Barrett's esophagus, have greater amounts of acid reflux than patients with uncomplicated gastroesophageal reflux disease.

These findings suggest that gastric juice is the main injurious component of the refluxed fluid. Animal studies, however, have shown that hydrochloric acid alone has a relatively innocuous effect on the esophageal mucosa, unless the acid is at very high concentration (pH 1.0 to 1.3), or acid exposure is prolonged. The combination of acid with the gastric proteolytic enzyme pepsin is significantly more injurious than either acid alone or pepsin alone and the concentration of acid required to cause mucosal injury is reduced by the presence of pepsin. The lack of significant injury after acid exposure alone is probably due to the effectiveness of the normal mucosal defense mechanisms. These defense mechanisms are reviewed by Orlando (see Further reading). The effect of pepsin probably results from its ability to digest intercellular proteins in the esophageal mucosa. Pepsin and other agents which increase acid-induced injury may thus act as ‘barrier breakers', allowing acid to overcome the normal esophageal epithelial defenses.

In humans, it has been difficult to quantify the relative importance of acid compared with pepsin in causing mucosal injury because acid is most injurious at low pH, and the optimum pH for pepsin activity is also at low pH. Esophageal aspiration studies have demonstrated, however, that both acid and pepsin concentrations are increased in patients with esophagitis compared with healthy control individuals.

Duodenal fluid

Some reflux of duodenal juice into the stomach occurs normally, especially in the postprandial and supine periods, but duodenal juice rarely reaches the esophagus. Reflux of duodenal fluid into the esophagus is significantly more common in patients with gastroesophageal reflux disease. As discussed in the next section, measurement of esophageal bilirubin concentration is used to detect duodenogastroesophageal reflux. Measurement of alkaline reflux correlates with reflux of duodenal contents only if very prolonged exposure to a pH above 7 (for more than 15 per cent of the study time) is detected. In a large study of patients with gastroesophageal reflux disease, 57 per cent had reflux of both gastric acid, measured by exposure to a pH of less than 4, and duodenal juice, measured by exposure to bilirubin, whereas only 34 per cent had reflux of gastric acid alone. Nine per cent had sufficient duodenogastric reflux to neutralize all gastric acid, resulting in the detection of only duodenal juice in the esophagus. Reflux of duodenal contents in the absence of gastric acid reflux is rare, but occurs more commonly in patients after gastrectomy and in those with severe atrophic gastritis.

Reflux of both gastric and duodenal contents is especially frequent in patients with mucosal injury (Fig. 2 and Fig. 3). The most noxious components of duodenal juice are thought to be the bile acids, although pancreatic enzymes may also be important. Esophageal aspiration studies have demonstrated the presence of bile acids at toxic levels in the esophagus in patients with gastroesophageal reflux disease and esophagitis. Patients with erosive esophagitis had a tenfold increase in bile acid concentration compared with patients with no mucosal injury, but the esophageal acid exposure was similar in both groups. Patients with reflux-related strictures or Barrett's esophagus have even greater amounts of both gastric acid and bile reflux than patients with esophagitis or those with gastroesophageal re-flux disease without esophagitis. Of interest, Barrett's esophagus has been reported to occur in patients after total gastrectomy. In animal models, exposure to acid alone, or duodenal juice alone at concentrations typically found in humans, is not particularly harmful. In contrast, combinations of acid and duodenal juice have a significantly greater noxious effect.

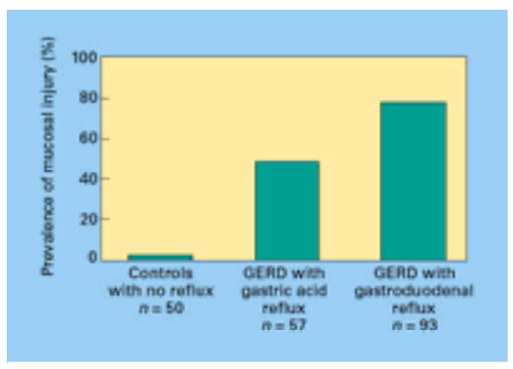


Fig. 2. The relationship between the composition of the gastric juice refluxed and the prevalence of mucosal injury. Gastric acid reflux is present when there is an increased esophageal exposure to a pH of less than 4 on 24-h pH monitoring. Gastroduodenal reflux is present when there is gastric acid reflux and an increase in bilirubin exposure in the distal esophagus, indicated by an absorbance above a threshold value of 0.2 at a wavelength of 453 nm.

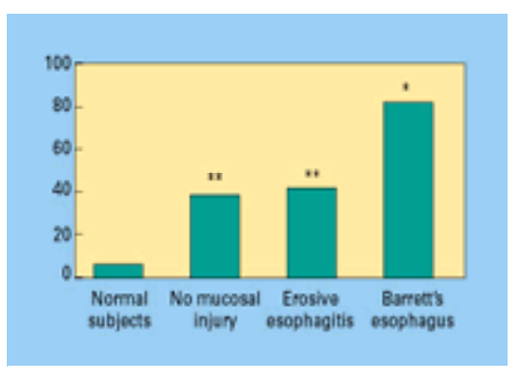


Fig. 3. Prevalence of abnormal esophageal bilirubin exposure in healthy subjects and in patients with gastroesophageal reflux disease with varied degrees of mucosal injury. (*$P < 0.03$ compared with all other groups;**$P < 0.03$ compared with normal subjects.)

Bile acids can act as ‘barrier breakers' similar to pepsin, allowing entry of hydrogen ions into the esophageal cells, with consequent cell damage. The mechanism for this is unclear, but may involve the known detergent ability of bile acids to dissolve lipid cell membranes, thereby increasing acid entry into the cell. Alternatively, the lipophilic nature of the bile acids, which allows their entry into the cell, may result in intracellular injury and membrane damage.

The chemistry of bile acids suggests that interactions between hydrochloric acid and bile acids in the gastric juice have an important contribution to the noxious properties of the fluid refluxed into the esophagus. Bile acid conjugation, solubility, and ionization state are all affected by pH. The damaging effect of mixed reflux at a pH of less than 4 can be shown experimentally to be due to acid and pepsin, or to conjugated (especially taurine-conjugated) bile acids. At a pH of less than 2, bile acids normally precipitate irreversibly and are no longer noxious to the mucosa.

Above pH 3, bile acids are soluble, either as bile salt and hydrogen ions or as unionized bile acids. At a pH of around 5, most unconjugated bile acids are soluble and unionized. Above pH 6, more than 90 per cent of bile acids are ionized. In theory, the reflux of gastric and duodenal juice with a pH above 3 results in the entry of lipophilic, unionized bile acids into the mucosal cells. Above pH 5, unionized, unconjugated bile acids in particular are able to enter the intracellular compartment. It is perhaps significant that a gastric pH range of 3 to 5 is present in patients with mixed reflux of acid and duodenal juice, and in a high proportion of patients on standard-dose acid suppression therapy. A benefit of antireflux surgery is that it maintains the esophageal pH at its normal value of 6, irrespective of the pH of the

gastric juice.

Diagnostic evaluation

Important components for the diagnosis of gastroesophageal reflux disease include the patient's history, endoscopic findings, radiologic abnormalities, and the measurement of increased esophageal exposure to gastric acid or duodenal juice on pH or bilirubin monitoring. These components are complementary, and the presence of more than one is usually necessary for a confident diagnosis. The use of symptoms alone, even if they are predominantly typical symptoms of gastroesophageal reflux disease, will result in misdiagnosis in some patients. Prior to considering aggressive medical or surgical therapy, it is important to establish that gastroesophageal reflux disease is indeed the cause of the patient's complaints. Errors in the treatment of patients with suspected gastroesophageal reflux disease commonly result from failure to correctly perform or interpret the diagnostic tests. Diagnostic tests are performed to confirm the presence of acid or bile in the esophagus, to assure that the symptoms are due to reflux, to assess if mucosal injury is present, and to identify if the loss of esophageal function that is usually associated with advanced disease has occurred. Preoperative investigations are also important to identify esophageal shortening or a large hiatal hernia, as these factors may influence the operative approach.

Gastroesophageal reflux disease can reliably be diagnosed if a patient has typical symptoms and at least one piece of objective evidence of reflux, such as an abnormal esophageal acid exposure on 24-h pH monitoring or visible esophagitis. A biopsy showing inflammatory cells infiltrating the mucosa on histology provides weaker objective evidence. For patients with atypical symptoms, the clinician should be more conservative, requiring two pieces of objective evidence to establish the diagnosis of gastroesophageal reflux disease. In patients with atypical respiratory symptoms, endoscopic laryngitis can be counted as one piece of evidence.

Upper gastrointestinal endoscopy

Endoscopy with mucosal biopsy should be performed in patients with the following conditions: (i) a history of heartburn or regurgitation lasting longer than 3 months, (ii) atypical symptoms, (iii) a poor response to medical treatment, and (iv) the presence of ‘warning symptoms' such as dysphagia, bleeding, weight loss, respiratory symptoms, or chest pain. In practice, virtually all patients referred for a surgical opinion for gastroesophageal reflux disease will require endoscopy. This should ideally be performed by the surgeon. Endoscopy is usually the first diagnostic procedure performed, unless the primary symptom is dysphagia, in which case a barium swallow is the first study.

Examination of the esophagus by endoscopy should include inspection of the vocal cords for reflux laryngitis changes, the esophageal mucosa for inflammatory or metaplastic changes, and the gastroesophageal junction for a hernia and mucosal abnormalities. The gastroesophageal junction is viewed in both antegrade and retroflexed positions. The location of the squamocolumnar junction, the diaphragmatic crura, and the anatomic gastroesophageal junction should be noted. The squamocolumnar junction is the visible irregular transition line (the ‘Z-line') where the pale esophageal mucosa becomes gastric mucosa. The crural location is obtained by asking the patient to sniff and noting the level at which crural indentation occurs. The anatomic gastroesophageal junction is identified as the proximal extent of the vertical gastric rugal folds, which is where these folds are replaced by the smooth mucosa of the esophageal ampulla. A hiatus hernia is present when the anatomic gastroesophageal junction is more than 2 cm above the crura (Fig. 4).

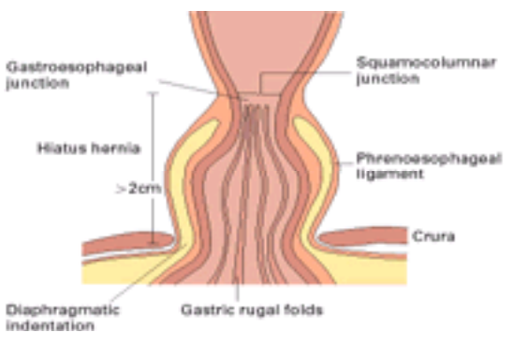


Fig. 4. Schematic diagram showing the endoscopic criteria for the diagnosis of a sliding hiatus hernia. The distance between the diaphragmatic crura, identified by having the patient sniff, and the endoscopic gastroesophageal junction, identified as the proximal extent of the gastric rugal folds, is at least 2 cm.

Reflux esophagitis

Endoscopy is the procedure of choice for diagnosing reflux esophagitis and Barrett's esophagus. Reflux esophagitis is commonly graded endoscopically according to the new Savary–Miller classification, shown in Table 1. Reflux esophagitis may be present histopathologically even if the mucosa appears normal endoscopically. The microscopic criteria for the diagnosis of esophagitis are controversial, but include the presence of neutrophils and lymphocytes, intraepithelial eosinophils, thickening of the basal cell layer, and elongation of the papillae. A minority of patients with esophageal symptoms can have endoscopic esophagitis that is not caused by reflux. These individuals usually have either chemical esophagitis due to drug ingestion or infective esophagitis. Infective esophagitis is most commonly caused by *Candida albicans*, herpes simplex virus, or cytomegalovirus infection. Infective esophagitis does not usually resemble reflux esophagitis, but the appearance of pill-induced esophagitis may be similar.

Grade 1	Single or multiple erosions in a single fold; erosions may be erythematous or exudative
Grade 2	Multiple erosions affecting multiple folds; erosions may be confluent
Grade 3	Multiple circumferential erosions
Grade 4	Ulcer, stenosis, or esophageal shortening
Grade 5	Barrett's epithelium. Columnar metaplasia in the form of circular or non-circular islands or tongues/extensions

Table 1 New Savary–Miller five-grade classification of reflux esophagitis

Barrett's esophagus

Barrett's esophagus is the condition in which the normal squamous epithelium of the distal esophagus is replaced with columnar epithelium in response to chronic exposure to gastric juice. Several different definitions for Barrett's esophagus are currently in use. The usual definition of Barrett's esophagus requires the presence of a macroscopically visible length of columnar mucosa, along with microscopic evidence of intestinal metaplasia. Intestinal metaplasia is characterized morphologically by the presence of mucin-containing goblet cells and a villiform epithelial surface. If the length of a macroscopically visible segment of intestinal metaplasia is greater than or equal to 3 cm, it is called long-segment Barrett's esophagus, and if less than 3 cm, short-segment Barrett's esophagus.

The detection of intestinal metaplasia on microscopic examination only, with no macroscopic columnar mucosa seen at endoscopy, is termed ultrashort-segment Barrett's or intestinal metaplasia of the cardia. Cardiac mucosa is a simple non-intestinalized epithelium with no parietal or chief cells. If small numbers of these cells are present, the epithelium is termed oxyntocardiac or mixed-type mucosa, and if, as is usual, there is an inflammatory infiltrate present in cardiac or oxyntocardiac mucosa, the term carditis is used. Carditis is found at the gastroesophageal junction in most adults over the age of 20. The etiology, and hence the importance, of non-endoscopically visible cardiac mucosa, both with and without intestinal metaplasia, is controversial. In particular, there is a dispute as to whether cardiac epithelium at the gastroesophageal junction is present at birth in all humans, or is acquired. Cardiac mucosa can be an acquired lesion, as shown by the development of cardiac mucosa in place of squamous epithelium in the cervical esophagus of patients who have undergone esophagectomy with esophagogastrostomy. In these patients,

reflux of gastric juice into the cervical esophagus from the pulled-up stomach is the cause of the development of cardiac mucosa. Studies have suggested that carditis at the gastroesophageal junction is an early marker for reflux disease. Further studies have shown that the prevalence of intestinal metaplasia in cardiac mucosa is related to the extent of esophageal acid exposure and the mechanical competence of the lower esophageal sphincter. Other investigators have emphasized that pangastritis due to *H. pylori* infection may also contribute to the development of intestinalized cardiac mucosa in some patients.

Barium swallow study

A barium swallow examination is helpful for investigating patients with known or suspected gastroesophageal reflux disease, especially if an operation is planned. A barium study provides information about the handling of liquid and solid boluses and about the presence and reducibility of a hiatus hernia. This information is often not gained from the other tests. More information is obtained if the study is video recorded (‘video-esophagram') than if only a stationary plain film record is obtained. A hiatus hernia can often only be seen in the prone position. The presence of a fixed non-reducible hernia in both the prone and upright positions can be a sign of esophageal shortening. In patients who present with dysphagia, a contrast study provides an important ‘road map' with information about the site and severity of any strictures, diverticula, or tumors that may be present. This information is useful for planning subsequent investigations such as endoscopy and ultrasound.

The barium study also provides information about gastric emptying. Absent antropyloric waves in the stomach with markedly delayed gastric emptying suggests a diagnosis of gastric atony. If the barium swallow study is suggestive of a gastric motility problem, gastric emptying can be further assessed by stationary or ambulatory radionuclide studies. Other tests of gastric function include gastric acid analysis for detecting gastric acid hypersecretion, and tests of duodenogastric reflux (discussed below). Tests of gastric function may be helpful in the evaluation of patients with proven gastroesophageal reflux disease who have mechanically normal measurements of the lower esophageal sphincter and normal acid clearance mechanisms. Many of these patients will have abnormal gastric motility, indicating the important influence of gastric function in the pathogenesis of gastroesophageal reflux disease. The rate of gastric emptying is usually increased after antireflux surgery.

Manometric examination of the lower esophageal sphincter

Esophageal manometry is performed in patients with suspected gastroesophageal reflux disease to assess the structural competence of the lower esophageal sphincter, to identify esophageal body motility disorders, and to locate accurately the upper border of the lower esophageal sphincter so that the pH electrode for 24-h pH monitoring can be positioned correctly. Stationary manometry is performed in the fasting supine patient by passing a catheter, usually through the nose, into the esophagus and stomach. The catheter used can be either a multilumen, low-pressure, water-perfused catheter attached to pressure transducers, or one with miniature solid-state transducers mounted directly in the catheter. The pressure is measured at multiple levels along the catheter. A pressure-sensing belt around the lower thorax records respiration, and swallows are recorded by a microphone placed around the neck. Initially all the pressure-sensing ports are placed in the stomach. As the catheter is withdrawn into the esophagus by either manual or motorized methods, a pressure and length profile of the lower esophageal sphincter is obtained (Fig. 5). The distal border of the lower esophageal sphincter is identified as the level at which the pressure rises by more than 2 mmHg above the gastric baseline pressure. The proximal border of the lower esophageal sphincter is identified asthe point at which the pressure falls from the high-pressure zone of the lower esophageal sphincter to the resting negative intrathoracic pressure of -4 mmHg. The overall length of the lower esophageal sphincter is the distance between its distal and proximal borders (Fig. 5).

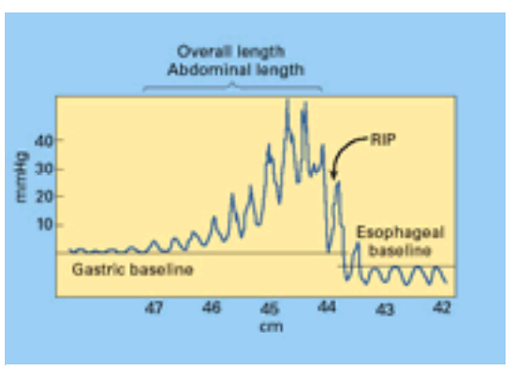


Fig. 5. Pressure tracing obtained as a transducer is pulled across the lower esophageal sphincter, showing the pressure, the overall and intra-abdominal lengths, and the RIP (respiratory inversion point).

The lower esophageal sphincter is normally positioned at the esophageal hiatus, with part of the sphincter within the abdominal cavity. This length of lower esophageal sphincter exposed to intra-abdominal pressure is important for resisting increases in abdominal pressure that might otherwise produce reflux, but the resting sphincter pressure is not dependent on exposure to the abdominal pressure, being measurable even when the abdominal cavity is open at laparotomy. The preservation of the normal geometry of the gastroesophageal junction, especially an acute angle of His, is also critical for normal lower esophageal sphincter function. By altering the manner by which gastric wall tension is transmitted to the lower esophageal sphincter, loss of normal geometry, such as occurs with a hiatus hernia, tends to allow the sphincter to be pulled open.

Measurement of the intra-abdominal length and the resting pressure of the lower esophageal sphincter require an appreciation of the variation in pressure that occurs with respiration. With inspiration, the descent of the diaphragm causes an increase in the intra-abdominal pressure and a decrease in intrathoracic pressure. The lower esophageal sphincter pressure can be measured at different locations over the pressure profile, but is commonly measured at the point at which the inspiratory rise in pressure that occurs within the abdomen first inverts to a fall in pressure within the chest. This point, termed the respiratory inversion point, is the level at which the pressure-recording port passes from the positive-pressure environment of the abdominal cavity into the negative-pressure environment of the thoracic cavity (Fig. 5). Anatomically, this corresponds to the point of insertion of the phrenoesophageal membrane or ligament into the lateral wall of the sphincter.

The resistance of the lower esophageal sphincter to the flow of gastric juice into the esophagus is a function of both its pressure and the length over which that pressure is exerted. The resting pressure, the overall sphincter length, and the intra-abdominal sphincter length are the three manometric characteristics of the lower esophageal sphincter that maintain its mechanical function as an antireflux barrier. The lower limits of normal for these parameters, as determined in our laboratory from the fifth percentile values for a group of 50 normal volunteers, are an overall length of 2 cm, an intra-abdominal length of 1 cm, and a resting pressure measured at the respiratory inversion point of 6 mmHg. If any of the parameters are below these values, the sphincter is considered to be permanently and structurally defective. The most common cause of a perman-ently defective sphincter is an inadequate pressure, but the effectiveness of a sphincter with normal pressure can be nullified by an abnormally short overall or abdominal length.

The sphincter lengths are measured with the stomach empty. Under these conditions, each patient's sphincter has a reproducibly constant overall and abdominal length. The sphincter length can be shortened, however, by filling of the stomach (Fig. 6). This shortening is analogous to the neck of a balloon shortening as the balloon is inflated. If the overall length of the sphincter is abnormally short when the stomach is empty, then with minimal gastric distension there will be insufficient sphincter length for the existing pressure to maintain sphincter competency, and reflux will occur.

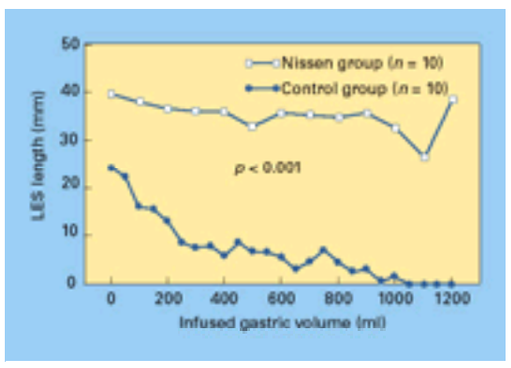


Fig. 6. Graph showing the relationship between increasing gastric volume and overall length of the lower esophageal sphincter. The lower esophageal sphincter length is maintained after Nissen fundoplication, even after large volumes of fluid have been infused into the stomach. (Modified from Mason et al. 1997.)

It is possible to measure the integrated effects of radial pressure exerted over the entire length of the high-pressure zone by forming a three-dimensional computerized image of the sphincter. The volume of this image reflects the sphincter's resistance and is called the sphincter pressure vector volume. A calculated volume less than the fifth percentile of normal subjects is another indication of a structurally defective sphincter.

For the clinician, the finding of a structurally defective sphincter has several implications. It is almost always associated with esophageal mucosal injury, it indicates that lifelong therapy will be necessary to control the patient's symptoms, that it will be difficult to manage the patient medically, and that antireflux surgery is likely to be needed for consistent, long-term symptom control. A structurally defective sphincter is permanent and irreversible, even if the associated esophagitis is healed. A structurally defective sphincter is commonly associated with reduced esophageal body function and, if uncorrected, the progressive loss of effective esophageal acid clearance can result in severe esophageal mucosal injury, including Barrett's esophagus, and in severe pulmonary fibrosis from repetitive aspiration.

Esophageal body motility study

The second part of the stationary manometry examination is the assessment of esophageal body function, the most important component of the pump antireflux mechanism. For this, the catheter is positioned so that the pressure ports span the length of the esophageal body. The peristaltic response to 10 swallows of 5-ml water boluses is analyzed. The measured features of individual contractions are the amplitude, duration, propagation speed, and morphology (i.e. whether single-, double-, or triple-peaked). Non-transmitted, interrupted, spontaneous, and simultaneous waves are also noted. The finding of a named esophageal body motility disorder suggests that this is the primary cause of the patient's symptoms, but non-specific abnormalities, such as distal esophageal hypocontractility, may result from long-standing gastroesophageal reflux disease. In these patients, recovery of some body contractility may occur after abolition of ongoing reflux and healing of esophagitis. If achalasia or scleroderma are found on manometry, an antireflux procedure, if indicated, should be a partial fundoplication type to avoid adding excessive outflow resistance to an already failed esophageal body. Failure to diagnose achalasia in a patient thought to have reflux disease is disastrous if fundoplication is performed without myotomy.

Ambulatory manometry study

Patients with equivocal findings on stationary manometry, or with non-obstructive dysphagia and a normal stationary manometry study, can be evaluated in more detail with ambulatory 24-h manometry. The ambulatory study uses a solid-state catheter containing four miniature transducers connected to a portable recording device. The earliest sign of a motility disorder on the ambulatory examination is loss of the normal organization of peristaltic waves during a meal. Dysphagia is likely when more than 50 per cent of the waves during meals are non-peristaltic or of insufficient amplitude to be effective in the transport of a bolus.

Ambulatory esophageal pH study

The gold standard for diagnosing gastroesophageal reflux disease is 24-h pH monitoring of esophageal acid exposure. Previous studies, such as the Bernstein test, are now rarely performed. Monitoring 24-h pH provides a quantitative measurement of reflux severity, and allows an objective assessment of the effectiveness of treatment. An important feature of the test is that patients keep a diary of symptoms experienced during the monitored period; this allows the patient's symptoms to be correlated with reflux events. Unless a patient has both typical reflux symptoms and definite endoscopic esophagitis, the diagnosis of gastroesophageal reflux disease based on symptoms alone is unreliable. Patients with esophagitis without typical symptoms may have another cause for the esophagitis, as discussed previously. There is also a large group of patients who have typical symptoms of gastroesophageal reflux disease and abnormal esophageal acid exposure, but who do not have endoscopic esophagitis. They constitute a population of patients referred to as those with non-erosive gastroesophageal reflux disease, and may be as large as the group of patients who have erosive disease. As a result of these observations, 24-h pH monitoring is a routine investigation for patients with suspected gastroesophageal reflux disease.

The pH monitoring study is usually performed after the esophageal manometry study, so that the pH probe can be positioned exactly 5 cm above the manometrically determined proximal border of the lower esophageal sphincter. The probe is connected to a data recorder worn on the patient's waist. The gastric pH is measured at the start and end of the test to ensure that it is within the normal range (pH 1 to 2).

All acid-suppressant medications must be stopped for a fortnight before the study, with only antacid use permitted during this period. A false-negative test may be caused by patients limiting their food intake and activity during the study, partly as a result of the discomfort caused by the test itself. Patients are thus encouraged to eat normal quantities of food and be physically active. If the patientsare allowed their normal diet, reflux events during the meal period are excluded from the analysis. Patients are given a list of foods and drinks that should be avoided during the monitored period. In particular, the ingestion of tea, coffee, and acidic foods, which can give false-positive results, should be limited. Abstinence from alcohol and tobacco use for the duration of the study is also required.

The normal pH of the esophagus ranges from 5 to 7. Reflux episodes are defined as periods when the esophageal pH is less than 4. The pH limit of 4 has been adopted arbitrarily, but the rationale for using this value is that the enzyme pepsin is active below this level, and there is a good correlation between measurement of an intraesophageal pH of less than 4 and patients' reports of heartburn. Several studies which examined the use of other pH levels for diagnosing gastroesophageal reflux disease concluded that, although a threshold of pH 4 may result in the severity of gastroesophageal reflux disease being underestimated, this pH value should continue to be used. In normal asymptomatic volunteers, the pH in the esophagus is less than 4 for approximately 2 per cent of a 24-h period, with the upper limit (95th percentile) being 4.4 per cent (Table 2). Physiologic reflux to a pH of less than 4 occurs more often when normal individuals are upright, especially during the postprandial period. In contrast, reflux uncommonly occurs when normal subjects are supine. For the 24-h pH study, patients are asked to avoid lying down except for the night-time sleep. Patients with abnormal reflux in both the upright and supine periods (bipositional reflux) are more likely to have gastroesophageal reflux disease complicated by mucosal injury, and are less likely to be managed adequately with standard medical therapy.

	Mean	SD	Median	Minimum	Maximum	95th Percentile
% Total time pH < 4	1.5	1.4	1.2	0	6.0	4.5
% Upright time pH < 4	2.2	2.3	1.6	0	9.3	8.4
% Supine time pH < 4	0.6	1.0	0.1	0	4.0	3.5
No. of episodes	19.0	12.8	16.0	2.0	56.0	40.9
No. of episodes > 5 min	0.8	1.2	0	0	5.0	3.5
Longest episode (min)	4.7	7.9	4.0	0	46.0	19.8
Composite score	6.0	4.4	5.0	0.4	18.0	14.7

Table 2 Values of 24-h esophageal pH monitoring in 50 healthy volunteers for a pH of less than 4

The normal values for the total, upright, supine, and postprandial periods, and for the total number of reflux episodes, the number of episodes longer than 5 min, and the longest reflux episode, are shown in Table 2. A composite score incorporates the six parameters shown in this table in a single value. Most commercially available pH software analysis programs automatically calculate the score. The score is particularly valuable in distinguishing between patients who have similar total percentage time values but different patterns of reflux. For example, if the total reflux times are equal, a smaller number of reflux episodes of long duration are more injurious than a larger number of brief, rapidly cleared episodes. This difference is not detected by the total percentage time for which the pH is less than 4, but is identified by the composite score value. In a multivariate analysis, the composite score was the most accurate predictor of the likelihood of successful operative treatment in patients with reflux symptoms.

Detection of proximal esophageal reflux

Dual pH probe catheters, with a distal probe in the usual position 5 cm above the lower esophageal sphincter and a proximal probe 20 cm above the lower esophageal sphincter, are used to investigate patients with respiratory or laryngeal symptoms. A single catheter containing probes 15 cm apart is available, so that there is no additional discomfort to the patient. In normal individuals, reflux episodes causing a fall to below pH 4 in the proximal probe are rare, but are commonly detected in patients with airway disease due to gastroesophageal reflux disease. The finding of a normal amount of proximal acid exposure does not exclude the possibility that gastroesophageal reflux disease is the cause of a patient's respiratory symptoms because airway symptoms such as cough and wheezing may be triggered, via a

reflex mechanism, by reflux of acid into the distal esophagus.

Detection of duodenogastroesophageal reflux

Abnormal duodenogastroesophageal reflux is diagnosed using a commercially available probe (Bilitec probe™) that detects bilirubin as a marker for duodenal juice. The Bilitec probe™ is a portable spectrophotometric device that continuously measures bilirubin concentration. A fiberoptic cable is connected to a light source and data recorder that the patient wears at the waist. The light source emits light at the wavelength of maximal absorbance of bilirubin (453 nm). The light is transmitted across a 2-mm space to a white Teflon reflector which reflects the non-absorbed light back to the probe. The degree of absorbance of light at this wavelength is directly related to the concentration of bilirubin. Falsely high recordings may result from ingestion of foodstuffs with yellow colourings, such as butter, and the signal is lost if small particles of food become lodged between the probe and the reflector. The patient's diet must thus be limited to a narrow selection of plain soft foods for optimum conduct of the study. There is some additional discomfort for the patient if both the Bilitec probe™ and a pH probe are in the esophagus, but more information is obtained if the two tests are performed simultaneously (Fig. 7).

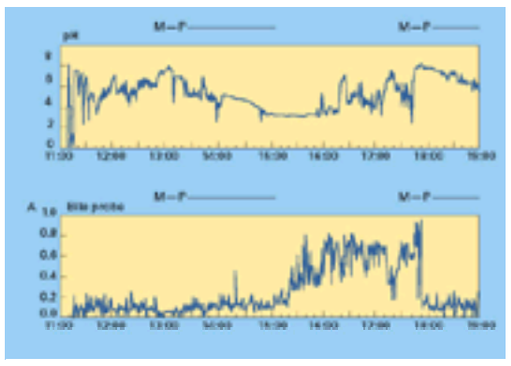


Fig. 7. Simultaneous display of an 8-h monitoring period from a patient with increased distal exposure to both acid (top) and bilirubin (bottom). M, meal period; P, postprandial period.

Bilirubin rarely refluxes into the esophagus of control subjects, with the upper limit of normal being less than 2.2 per cent of the total monitoring time. Measurement of bilirubin exposure above an absorbance threshold of 0.2 correlates well with exposure to bile acids.

Measurement of reflux of duodenal contents into the stomach, rather than into the esophagus, may be helpful in the evaluation of patients with atypical symptoms, especially if the 24-h esophageal pH study is normal. Duodenogastric reflux can be detected using a gastric Bilitec probe™ for gastric bilirubin exposure, a gastric pH probe for measuring exposure above pH 2, or a cholescintigraphy study.

Treatment

The prevalence of heartburn in the general population suggests that gastroesophageal reflux disease is very common. This high prevalence partly explains why many people, including some physicians, regard this disease as minor. Patients with gastroesophageal reflux disease have a significantly reduced quality of life, however, and one study reported that patients with untreated reflux disease had worse quality of life scores than patients with mild heart failure. Furthermore, because gastroesophageal reflux disease is the main identified risk factor for the development of the premalignant condition Barrett's esophagus, patients with reflux disease are at increased risk for developing Barrett's-associated adenocarcinomas of the esophagus or gastroesophageal junction.

Most patients with mild symptoms self-medicate with over-the-counter drugs such as antacids, alginic acid, or histamine H_2-receptor antagonists, whereas patients with more severe or persistent symptoms seek medical attention. Lifestyle modifications to reduce reflux include elevation of the head of the bed, weight loss, restriction of alcohol, caffeine, fatty and fried foods, avoidance of large meals and meals late at night, and the elimination of tobacco smoking. Lifestyle modifications receive little emphasis now that potent alternative treatments for gastroesophageal reflux disease are available, but they remain an important part of the treatment of this disease. For many patients, a 2- to 3-month course of antacids or alginic acid, combined with these lifestyle modifications, will result in adequate symptom relief.

Medical therapy

Patients with persistent symptoms will usually be treated initially with medical therapy. The aim of medical therapy is to reduce the acidity of the refluxed gastric juice. Histamine H_2-receptor antagonists are less effective in relieving symptoms and healing esophagitis than H^+, K^+ -ATPase/proton pump inhibitors, which are the mainstay of current medical treatment. Standard dose regimens of proton pump inhibitors, such as omeprazole at 20 mg/day, lansoprazole at 15 to 30 mg/day, or pantoprazole at 20 mg/day, can reduce gastric acid production by up to 90 per cent, and double-dose regimens can reduce acid output by more than 95 per cent. These dosages will heal esophagitis in most patients, but up to 20 per cent of patients will have persistent esophagitis. Severe esophagitis may be resistant to conventional dosage regimens and, in these patients, long-term studies have shown that dose escalation is often needed to maintain healing.

Proton pump inhibitors relieve heartburn in most patients, but they are less effective for relieving regurgitation or pulmonary symptoms. Relief of heartburn does not mean that reflux has been adequately controlled. Studies of 24-h esophageal pH have demonstrated that even in patients taking high doses of proton pump inhibitors, nocturnal gastric acid breakthrough occurs and commonly results in abnormal esophageal acid exposure. Patients who become asymptomatic often discontinue medications or take medications only when symptoms return, thus allowing ongoing reflux and mucosal damage to occur.

The aim of promotility agents such as metoclopramide or cisapride is to increase lower esophageal sphincter tone, increase esophageal body peristalsis, and promote gastric emptying. These drugs have a beneficial effect in some patients, especially those with mild disease. The use of metoclopramide has been limited by its frequent central nervous system side-effects. Cisapride has been shown to be more effective than placebo, and has an approximately equivalent effect on healing of esophagitis as H_2 blockers. However, the combination of cisapride with some drugs that are metabolized by the cytochrome P450 system has been associated with deaths due to cardiac dysrhythmia. Other promotility agents, such as domperidone, bethanecol, and erythromycin, have been disappointingly ineffective.

Unfortunately, within 6 months of discontinuation of any formof medical therapy for gastroesophageal reflux disease, 80 per cent of patients will have a recurrence of symptoms. Most patients with reflux disease will thus require long-term, and perhaps lifelong, proton pump inhibitor therapy. The long-term effects of maintenance therapy with these agents remain unknown. The development of atrophic gastritis in patients receiving these drugs has been reported, and the clinical significance of the hypergastrinemia which results from proton pump inhibitor use is not known. Overall, however, serious proven side-effects are remarkably infrequent. Of concern is the potential detrimental effect of elevating the gastric pH from a level of 2 or below to 3 or above on a chronic basis. The possibility of inducing phenotypic changes in the gastric and esophageal mucosa as a consequence of this pH shift is currently being investigated.

Surgical therapy

The aim of surgical therapy is to stop the increased esophageal exposure to gastric juice, and as a consequence to relieve the symptoms of gastroesophageal reflux disease, heal esophagitis, and prevent the development of complications, without the need for medication or unrealistic lifestyle restrictions. Three randomized trials have compared medical and surgical treatment for gastroesophageal reflux disease. All three reported that open Nissen fundoplication was significantly better than conventional medical treatment (including up to 40 mg omeprazole per day in one study). As advised by the American College of Gastroenterology Practice Parameters Committee, properly performed antireflux surgery is a maintenance option for the patient with well-documented gastroesophageal reflux disease. All patients with objectively proven reflux disease should be informed that antireflux surgery is a valid alternative to lifelong medication use, with advantages over medical therapy in many instances. If patients are properly informed in this way about surgical treatment, many will prefer this option to medication use. Patients' preference for surgical therapy has significantly increased as a result of the development of minimally invasive operations for gastroesophageal reflux disease.

Although the cost to the patient of long-term medical therapy varies in each country, it is an important factor for patients with an average income who have to pay for their medications. Most cost–benefit analyses show that surgery becomes less expensive than proton pump inhibitor therapy at around 3 to 5 years after operation. Cost factors are likely to be more important in younger patients. Patients under 50 years of age can reliably be assured that surgery is less costly than continuous medical treatment. Some patients seek surgical treatment for lifestyle and convenience reasons. These individuals do not want to be dependent on lifelong daily

medication, and they wish to live without the anxiety and inconvenience associated with constant observance of dietary, alcohol, and other lifestyle restrictions.

Indications for surgery

The indications for surgical therapy for gastroesophageal reflux disease have widened recently as a result of a better understanding of the pathogenesis of disease, clarification of risk factors for disease progression despite medical therapy, and the widespread acceptance that laparoscopic antireflux surgery provides equivalent results to open surgery, with significantly less morbidity. Important findings that encourage consideration of operative treatment are the presence of an incompetent sphincter, severe esophagitis at the patient's initial endoscopy, and breakthrough of symptoms while on medical therapy.

It was thought previously that antireflux surgery was only beneficial for patients with a structurally defective sphincter, but it is now known that antireflux surgery will also prevent reflux in patients with a structurally normal lower esophageal sphincter. The mechanism for this effect is that surgery reduces reflux episodes caused by transient loss of the high-pressure zone due to shortening of the lower esophageal sphincter with gastric distension. The presence of a structurally defective sphincter is an indication, however, that medical therapy is likely to be less effective over the long term, and that surgery will eventually be required for disease control.

In the past, some clinicians emphasized that surgical treatment should only be considered if esophagitis was present. It is now recognized that many patients with significant symptoms and gastroesophageal reflux disease proved by pH test have a normal-appearing esophageal mucosa. These patients are candidates for operation if they would otherwise require long-term medical therapy. The presence of severe esophagitis at the initial examination is associated, however, with a greater likelihood of failure of medical therapy and is thus an encouragement for surgery. A recent study compared outcomes at approximately 4 years after the initiation of medical or surgical therapy in 35 725 patients with erosive esophagitis treated by the US Department of Veterans Affairs medical system. There were reassuringly few complications after either type of treatment in patients with uncomplicated esophagitis, but in those who had severe esophagitis on their initial encounter, the prevalence during follow-up of esophagitis, esophageal ulcer, and stricture were all significantly lower in patients treated with surgical rather than medical therapy.

It was formerly thought that surgery should be reserved for patients who receive little or no benefit from acid-suppressant drugs. It is now evident that these patients are unlikely to have gastroesophageal reflux disease, and are not good candidates for antireflux surgery unless the diagnosis of reflux disease is definite. A good response to medical therapy identifies patients who are likely to have a good result with surgery. Other factors that predict the likelihood of a successful outcome after antireflux surgery are the presence of typical symptoms and an abnormal composite score or percentage time for which the pH is less than 4 on 24-h pH testing. Patients with atypical symptoms who have a good response to acid-suppression therapy are also good candidates for surgery.

Additional indications for surgical rather than medical treatment for gastroesophageal reflux disease are the presence of duodenogastroesophageal reflux or Barrett's esophagus. Although there is some evidence that medical therapy has a limited effect in reducing duodenogastroesophageal reflux, prevention of the reflux of duodenal fluid into the esophagus is more effectively achieved with antireflux surgery. As discussed above, reflux of duodenal fluid is particularly important in patients with complications of gastroesophageal reflux disease. Several studies indicate that surgery provides more effective disease control in patients with Barrett's esophagus. The few longitudinal studies reported have shown that new cases of Barrett's esophagus are more likely to develop in patients with gastroesophageal reflux disease treated medically than in those treated surgically. Antireflux surgery may also reverse the process of intestinalization within cardiac mucosa in some cases, especially if the intestinalized cardiac mucosa is confined to microscopic disease at the gastroesophageal junction.

Principles of antireflux operations

Successful antireflux surgery prevents reflux of gastric juice into the esophagus but preserves the patient's ability to swallow normally. This can be achieved if several principles are adhered to. First, the operation should restore the pressure, the overall length, and the length exposed to abdominal pressure of the lower esophageal high-pressure zone to normal levels. This not only augments the resistance to reflux in patients who had a defective lower esophageal sphincter preoperatively, but also prevents unfolding of the sphincter during periods of gastric distension. Second, the operation should correct a hiatus hernia if present and restore the normal geometry of the gastroesophageal junction, with a normal angle of His and at least 1.5 cm of abdominal sphincter length. This maintains competency of the antireflux barrier during periods of increased intra-abdominal pressure. Third, to avoid troubling, long-term dysphagia, the fundoplication should not be made too tight. This can be achieved by constructing the fundoplication over a 60 French bougie, by using both the anterior and posterior fundic walls, and by ensuring that the fundoplication is no longer than 2 cm. It is also important for avoidance of postoperative dysphagia that the fundoplication is not twisted, and is not positioned around the stomach below the gastroesophageal junction. The latter is likely to occur if the esophagus has been foreshortened by disease. Improper placement of the fundoplication below the gastroesophageal junction will interfere with vagally mediated relaxation of the sphincter on swallowing, resulting in prolonged postoperative dysphagia. Postoperative dysphagia has been associated with significant impairment of relaxation, and is reflected by elevation of the residual relaxation pressure. Dysphagia may also result if the resting pressure of the lower esophageal sphincter is too high, indicating that the fundoplication was constructed too tight. A fundoplication that is too loose, on the other hand, can result in continued reflux.

Choice of operation

The choice of repair and operative approach (transabdominal or transthoracic) is influenced by the function and length of the esophageal body, by the presence of a large crural separation, as is commonly found with a large hiatal hernia, and by whether there has been a previous repair.

A Nissen fundoplication produces a greater degree of outflow obstruction than the partial fundoplication operations. This observation led some surgeons to advocate that a partial fundoplication should be performed in patients with distal esophageal contraction amplitudes less than 30 mmHg. Recent studies have shown that most patients with this level of contraction amplitude can undergo complete fundoplication without long-term or troubling dysphagia. The minimum contraction amplitude required for the esophageal body to overcome the resistance of a Nissen fundoplication has not been determined, but our experience has suggested that a mean contraction amplitude below 20 mmHg in the mid- and distal esophagus is the likely lower limit.

Recent studies, including some randomized controlled trials, have helped clarify the relative merits of partial and total fundoplication procedures. The prevalence of complications such as dysphagia, excessive flatus, and inability to belch is higher after Nissen fundoplication, but these complications are generally far less troubling on long-term follow-up than the patients' preoperative reflux symptoms. Most investigators agree that the total Nissen fundoplication operation provides better long-term control of reflux than partial fundoplication. In patients with severe symptoms, a mechanically defective sphincter, or extensive mucosal disease, it is essential to restore an effective and durable antireflux barrier. For this reason, the Nissen operation is our preferred procedure, with partial fundoplication reserved for patients with absent or severely disordered peristalsis or severe esophageal body hypomotility, as evidenced by contraction amplitudes less than 20 mmHg in the distal esophagus. Patients with moderate body hypomotility are at greater risk of severe, progressive disease, and they should thus be treated with complete fundoplication if possible. The indications for partial fundoplication are still evolving, and it may become the procedure of choice for patients with early disease without mucosal injury, a manometrically normal lower esophageal sphincter, and normal peristalsis. In these patients, effective control of reflux without risk of troubling dysphagia may be obtained with a laparoscopic partial fundoplication.

Esophageal shortening

Although the recognized prevalence of anatomic shortening of the esophagus in patients with gastroesophageal reflux disease varies between centers, some patients undoubtedly have a short esophagus, and the operative approach may need to be adjusted in these patients. Shortening is caused by chronic reflux injury with fibrosis and contraction in the esophageal wall. A short esophagus is suspected preoperatively by the presence of a sliding hiatus hernia that measures more than 5 cm between the crural indentation and the gastroesophageal junction on endoscopy, or a hernia that fails to reduce completely on a barium swallow performed in the upright position. We prefer a thoracic approach for these patients to maximize esophageal mobilization. Inability to bring the gastroesophageal junction of the mobilized esophagus below the diaphragm without tension identifies the truly short esophagus. Moderate degrees of esophageal shortening may be managed by transthoracic mobilization, but for severe degrees of shortening a gastroplasty should be performed.

Failure to recognize and manage properly the shortened esophagus is a reason for herniation into the chest of either an intact or a disrupted repair, and is a reason for the occurrence of the so-called ‘slipped' Nissen. A slipped Nissen usually results from incorrect intraoperative placement of the fundoplication around the proximal stomach below the gastroesophageal junction, rather than around the distal esophagus above the gastroesophageal junction. This error commonly occurs when the esophagus is short. Seventy per cent of patients with a shortened esophagus have poor esophageal body contractility and disordered waveforms due to inflammatory damage to the esophageal muscle. If these findings are severe or there is a history of several failed antireflux operations, and the patient has persistent dysphagia despite dilatation, an esophageal resection is likely to be necessary to provide relief.

The distance between the crura is rarely more than 4 cm, but in some patients with large hiatus hernias the crura are separated by this distance or more. In this situation, which is probably due to improper development, the crura cannot be approximated easily and the right crus is often atretic, allowing sutures to tear throughand resulting in crural separation and reherniation. We approach these patients through a thoracic incision in order to expose the diaphragm in a way that will

allow for an effective closure.

Techniques of antireflux operations

Laparoscopic Nissen fundoplication

Nissen fundoplication is the most commonly performed antireflux operation. The procedure can be performed via laparotomy or thoracotomy, but the laparoscopic approach is preferred. The basic elements of the operation are: (i) exposure of the esophageal hiatus, (ii) crural closure, (iii) fundic mobilization, and (iv) creation of a short, loose circumferential fundoplication.

The patient is placed in a steep reverse Trendelenburg position. Five 10-mm ports are used (Fig. 8). The left lateral segment of the liver is retracted towards the anterior abdominal wall. To expose the esophageal hiatus, an incision is made in the gastrohepatic omentum above and below the hepatic branch of the anterior vagus nerve and in the underlying peritoneum covering the anterior aspect of the right crus (Fig. 9). Injury to an aberrant left hepartic artery in the gastrohepatic omentum should be avoided. On the left side, the gastric fundus is separated from the lateral surface of the left crus and both crura are dissected inferiorly along the diaphragmatic hiatus until the V-shaped decussation anterior to the aorta is reached. The esophagus is mobilized by dissection of the anterior and posterior soft tissues within the hiatus. The anterior and posterior vagi must be identified and preserved. Dissection of the left crus is assisted by encircling the esophagus with a Penrose drain and retracting the fundus and the esophagus to the right side. Dissection of the esophagus within the mediastinum should be undertaken with care to avoid opening a pleural cavity. With the esophagus retracted anteriorly and to the left, the crura are closed from below upwards with two to four interrupted figure-of-eight sutures, using a strong, non-absorbable suture material (Fig. 10). The fundus is completely mobilized by dividing the anterior short gastric vessels along the upper third of the greater curvature and the posterior short pancreaticogastric vessels located in the retroperitoneal area near the cephalad tip of the gastric fundus and spleen. The fundoplication is constructed by bringing the posterior fundic wall behind the esophagus to the right side and by bringing the anterior fundic wall anterior to the esophagus above the Penrose drain (Fig. 11). The anterior and posterior fundic lips are manipulated to allow the fundus to envelop the esophagus without twisting the gastroesophageal junction or distorting the normal plane of the stomach (Fig. 12 and Fig. 13). A lubricated 60-Fr Maloney bougie is passed through a well-lubricated esophagus into the stomach. This assists proper placement of the fundoplication above the gastroesophageal junction, reduces the stomach into the abdomen, and stretches the stomach out to allow proper sizing of the fundoplication around the lower esophagus.

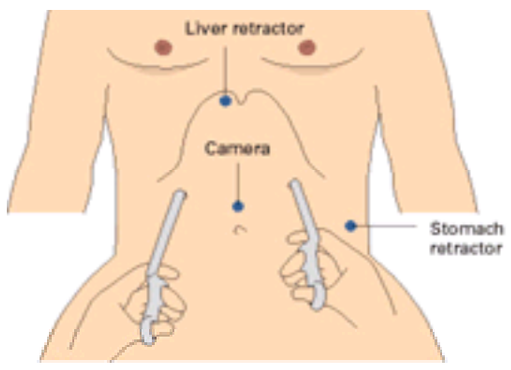


Fig. 8. Laparoscopic Nissen fundoplication. (1) Port placement.

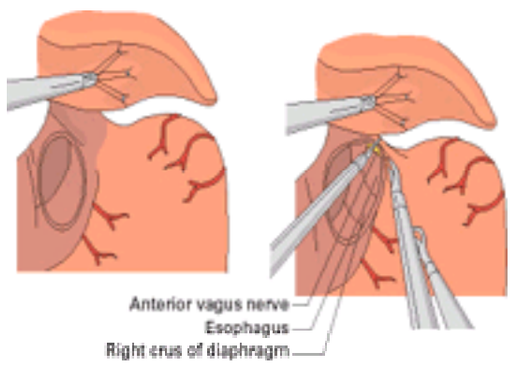


Fig. 9. Laparoscopic Nissen fundoplication. (2) The lesser omentum is incised to expose the right crus.

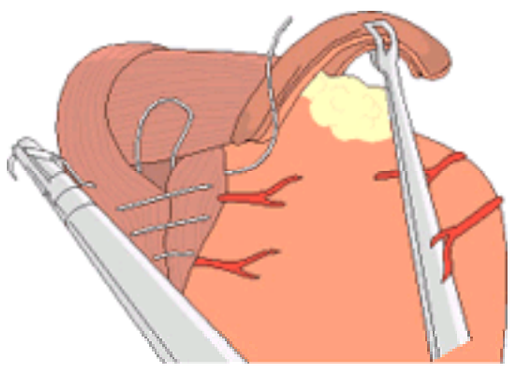


Fig. 10. Laparoscopic Nissen fundoplication. (3) The crura are closed from below upwards, with the esophagus retracted anteriorly and to the left.

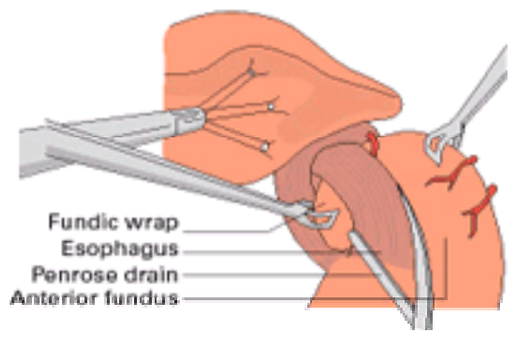


Fig. 11. Laparoscopic Nissen fundoplication. (4) The posterior fundus is brought behind the esophagus.

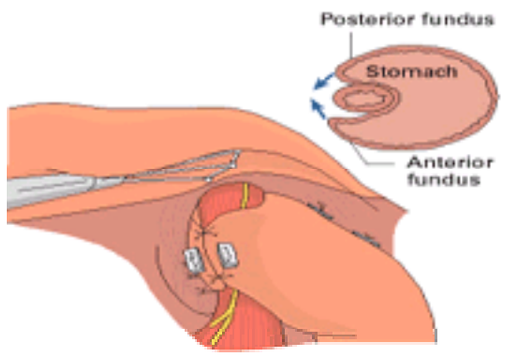


Fig. 12. Laparoscopic Nissen fundoplication. (5) The completed 360° fundoplication.

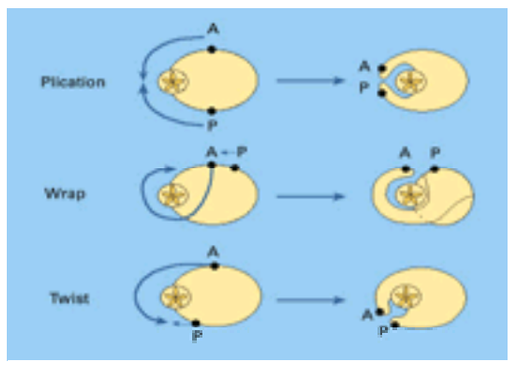


Fig. 13. Diagram showing possible orientations of a Nissen fundoplication. The top method is preferred, as the lower two methods result in twisting of the fundoplication.

The fundoplication is formed with a single temporary holding stitch and its construction is checked for proper position and size. If acceptable, it is permanently secured with a single U-stitch of 2–0 Prolene buttressed with 1.5 × 0.5 cm felt pledgets, with the stitch including the muscular layer of the esophageal wall. Two anchoring sutures of 3–0 silk are placed above and below the U-stitch to bring the fundoplication to a 1.5 to 2 cm length. The suture line of the fundoplication should be in the right lateral position at the end of the procedure.

A popular modification of the Nissen fundoplication is the Nissen–Rosetti operation, in which the short gastric vessels are not divided and the anterior fundus is brought behind and around the esophagus and fixed to the remaining gastric wall. The Nissen–Rosetti operation is slightly quicker to perform but risks producing torsion of the wrap, with consequent postoperative pain and/or dysphagia. Dividing the short gastric vessels enhances the ability to perform a tension- and torsion-free fundoplication.

Open Nissen fundoplication

An open transabdominal Nissen fundoplication is indicated when other open procedures are combined with an antireflux procedure or when the patient has had previous upper abdominal surgery. The open transthoracic approach for a Nissen fundoplication is preferred if the esophagus is judged preoperatively to have a high probability of being short, but after intrathoracic mobilization there is sufficient length to perform a repair without tension. This approach is also used if there is associated pulmonary pathology requiring operation, if the patient is severely obese, or if the patient requires a complex redo operation.

The transthoracic procedure is performed through a left posterolateral thoracotomy made in the sixth intercostal space (Fig. 14). When entrance into the abdomen is necessary, the diaphragm is incised circumferentially 2 to 3 cm from the lateral chest wall for a distance of approximately 10 to 15 cm. The esophagus is mobilized from the level of the diaphragm to underneath the aortic arch and freed from the borders of the esophageal hiatus. The fundus and part of the body of the stomach are brought up through the hiatus into the chest. The gastroesophageal junction is marked with a suture. Endoscopy may be needed to identify the junction. The stomach is then replaced into the abdomen and the position of the marking suture in relation to the crura is noted. If the suture is 2 cm or more below the crura, a gastroplasty is not needed and a transthoracic Nissen fundoplication can be performed. Crural sutures are then placed to close the hiatus (Fig. 15) and the fundoplication is constructed in a manner similar to that described for the transabdominal operation (Fig. 16, Fig. 17, Fig. 18, and Fig. 19). Closure of the crura is facilitated by the transthoracic approach because an excellent purchase of the stout diaphragmatic tissue, starting from the aorta and moving anteriorly, can be obtained. Prior to closing the hiatus, the fundoplication is placed into the abdomen by compressing the fundic ball manually and maneuvering it through the hiatus.

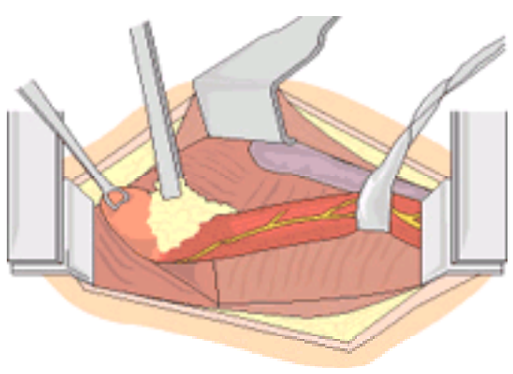


Fig. 14. Transthoracic Nissen fundoplication. (1) Through a left posterolateral thoracotomy, the esophagus is mobilized and the gastroesophageal junction is dissected from the diaphragmatic hiatus. The fundus of the stomach is drawn into the chest using a Babcock clamp. The forceps are on the vascular fat pad at the gastroesophageal junction.

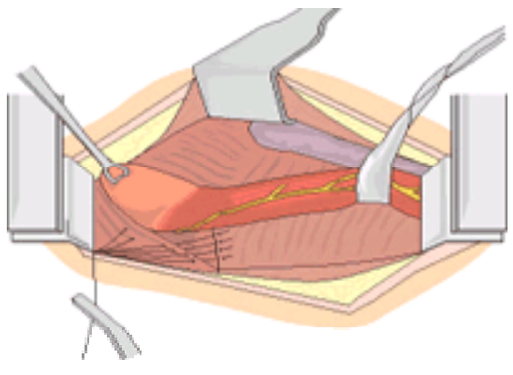


Fig. 15. Transthoracic Nissen fundoplication. (2) The vascular fat pad has been excised, and the esophagus is retracted anteriorly for posterior placement of sutures for closure of the hiatus.

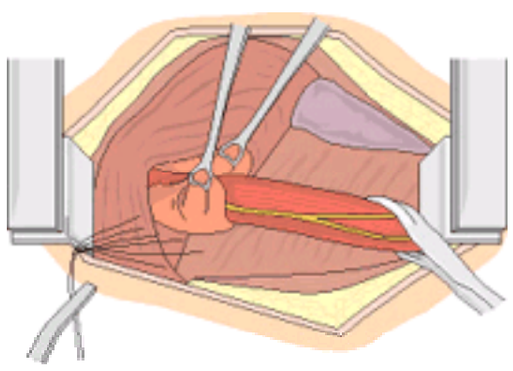


Fig. 16. Transthoracic Nissen fundoplication. (3) The posterior and anterior fundic walls are brought through the hiatus into the chest and are plicated around the lower esophagus.

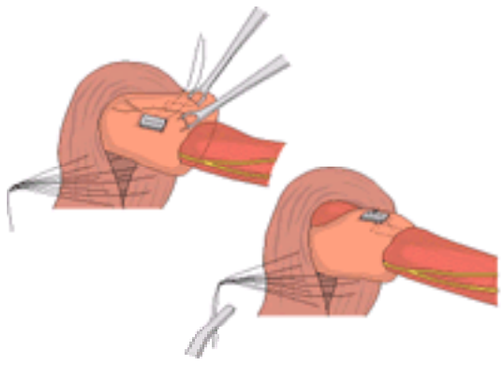


Fig. 17. Transthoracic Nissen fundoplication. (4) After passing a 60-Fr bougie into the stomach for accurate sizing, the 360° fundoplication is constructed using a pledgeted U-stitch.

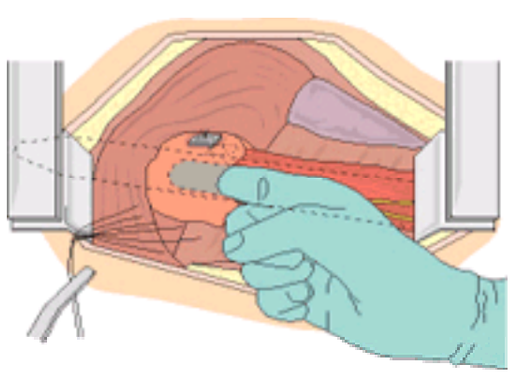


Fig. 18. Transthoracic Nissen fundoplication. (5) The tip of the surgeon's finger should be able to be inserted between the fundoplication and the esophagus containing the 60-Fr bougie. The U-stitch is located on the right anterolateral wall of the esophagus.

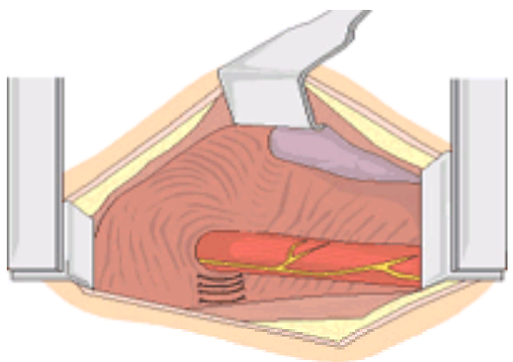


Fig. 19. The completed transthoracic Nissen fundoplication. The repair is in the abdomen and the right and left crura have been approximated by tying the previously placed sutures.

Belsey Mark IV operation

The Belsey Mark IV operation is also performed through the left chest. It is the prototype of the partial fundoplication operations and consists of an anterior 240° gastric fundoplication (Fig. 20, Fig. 21, Fig. 22, Fig. 23 and Fig. 24). The incision and dissection for this operation is the same as that described for the transthoracic Nissen operation. Sutures are placed in the crura, but are not tied until the fundoplication is complete. The fundoplication is constructed by rolling the mobilized fundus anteriorly on to the esophagus and fixing the fundoplication in place with two rows of three, non-absorbable, horizontal mattress sutures. The three sutures in each row are placed at the 4, 8, and 12 o'clock positions, with the 12 o'clock position corresponding to the anterior apex of the esophageal hiatus. The stitch should include the seromuscular layers of the stomach and the muscular wall of the esophagus. The first row of sutures approximates the fundus 1.5 cm from the gastroesophageal junction, with the tubular esophagus a similar distance proximal to the gastroesophageal junction (Fig. 20). The second row of sutures is made at the same 4, 8, and 12 o'clock positions, 1.5 to 2.0 cm above the first row (Fig. 21). The tails of the tied sutures in the second row are now passed through the diaphragm 0.5 cm from the edge of the hiatus in the 4, 8, and 12 o'clock positions (Fig. 22). After positioning the reconstructed cardia without tension in the abdomen, the diaphragmatic and crural sutures are tied to secure the repair in the abdomen and close the hiatus around the esophagus (Fig. 23 and Fig. 24).

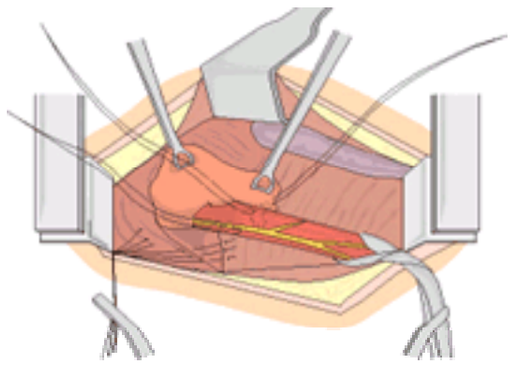


Fig. 20. Performance of the Belsey 240° gastric fundic wrap. (1) The first row of sutures is placed 1.5 cm above the gastroesophageal junction. The sutures are placed at the 4, 8, and 12 o'clock positions on an imaginary clockface oriented with the 12 o'clock position anterior. Particular attention must be given to placement of the right lateral (4 o'clock) suture.

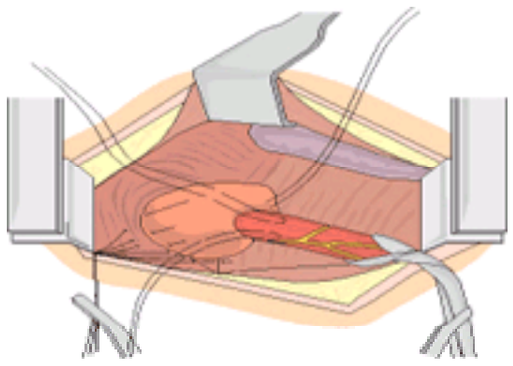


Fig. 21. Performance of the Belsey 240° gastric fundic wrap. (2) The second row of sutures is placed 1.5 to 2.0 cm distal and proximal to the first row.

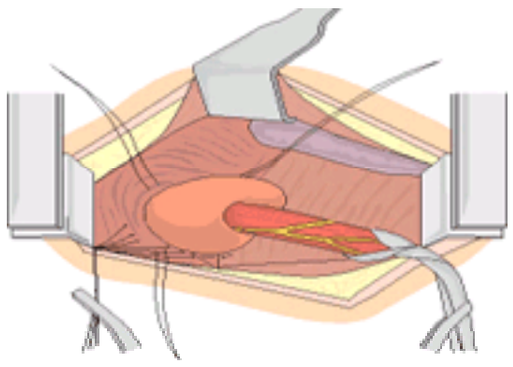


Fig. 22. Performance of the Belsey 240° gastric fundic wrap. (3) The tails of the previously tied second row of sutures are passed through the diaphragm, 0.5 cm apart and 1.0 to 1.5 cm from the edge of the hiatus.

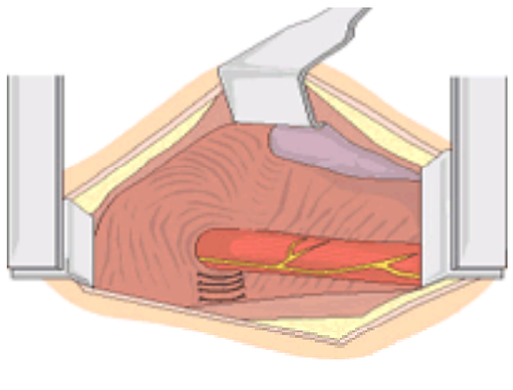


Fig. 23. Performance of the Belsey 240° gastric fundic wrap. (4) The completed Belsey 240° gastric fundic wrap, showing the right and left crura approximated by tying the previously placed sutures.

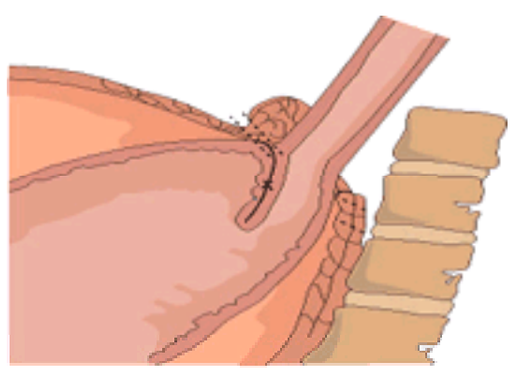


Fig. 24. Performance of the Belsey 240° gastric fundic wrap. (5) Sagittal section of the complete repair showing posterior sutures in the crus and the two rows of sutures used to construct the fundoplication. The second row of sutures passes through diaphragm, stomach, and esophagus.

Collis gastroplasty

A Collis gastroplasty is indicated in patients with a markedly shortened esophagus. This operation is probably best performed through the chest, although abdominal open and laparoscopic techniques have been described. A 4- to 5-cm gastric tube or ‘neo-esophagus' is created over a 48-Fr bougie placed along the lesser curvature of the stomach. This can be accomplished with a single application of a 55-mm linear cutting stapler placed along the fundic side of the bougie (Fig. 25, Fig. 26 and Fig. 27). To achieve a uniform diameter of the gastric tube throughout its length, gentle traction is exerted on the greater curvature before closure of the jaws of the stapler (Fig. 26). A Belsey Mark IV or Nissen fundoplication is then performed around the gastric tube. The neo-esophagus may be inert, and patients with a short esophagus often have reduced peristaltic contraction amplitudes. The combination of a Nissen operation and a Collis gastroplasty may therefore result in dysphagia because of an inability of the primary peristaltic contractions to overcome the resistance of the 360° fundoplication. The Collis operation is therefore usually combined with a Belsey procedure. A problem with the Collis operation is that some patients will secrete gastric acid from the neo-esophagus. If the acid-secreting area is above the antireflux barrier, the esophagus will be exposed to acid. Consequently, if a 24-h pH study performed after Collis gastroplasty shows increased esophageal acid exposure, it may not mean that the repair has failed.

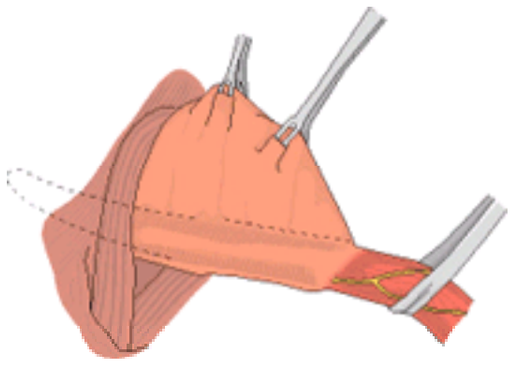


Fig. 25. Construction of a Collis gastroplasty. (1) A 48-Fr bougie is passed into the stomach. The dotted line indicates the line of division of the gastric wall for construction of the gastric tube.

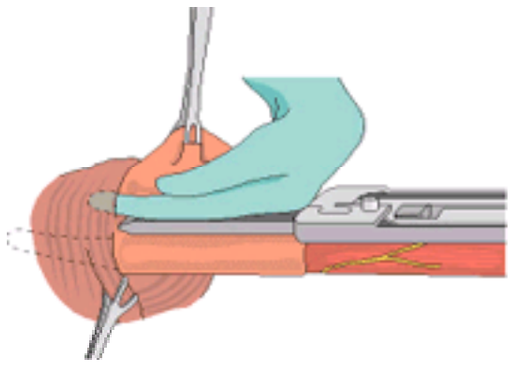


Fig. 26. Construction of a Collis gastroplasty. (2) The stomach is divided using a GIA stapler. Traction is exerted on the greater curvature side of the fundus before closing the jaws of the stapler. This ensures that the gastric tube closely approximates the diameter of the 48-Fr bougie throughout its length.

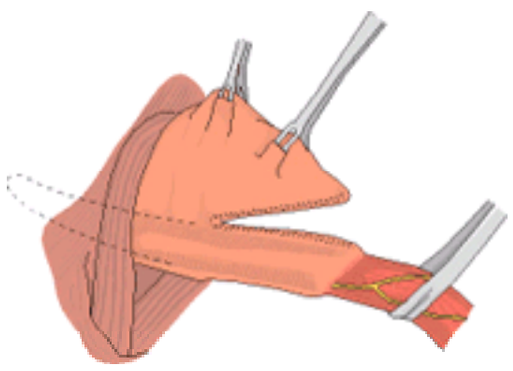


Fig. 27. Construction of a Collis gastroplasty. (3) After stapling and division of the stomach, a 5-cm gastric tube is formed along the proximal portion of the lesser curvature. This effectively lengthens the esophagus and allows the fundoplication to be constructed without tension.

Other partial fundoplication and cardiopexy operations

Partial fundoplication and cardiopexy operations are usually performed laparoscopically. The Toupet operation is a 180° or 270° posterior partial fundoplication (Fig. 28). The esophageal hiatus exposure, crural closure, and fundic mobilization stages are as described for the Nissen operation. The mobilized gastric fundus is brought behind the esophagus and sutured first to the right crus and then to the anterior right esophagus using interrupted non-absorbable sutures. The fundoplication is completed by suturing the left limb of the wrap to the left side of the esophagus (Fig. 28).

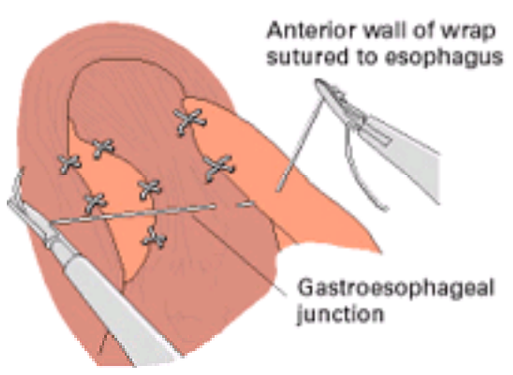


Fig. 28. The Toupet posterior 270° partial fundoplication.

The Dor operation is a left anterolateral 180° hemifundoplication that is often added to laparoscopic Heller's myotomy for achalasia to prevent postoperative reflux. When performed correctly, it also augments the angle of His. The medial anterior fundic wall is sutured along the distal esophagus just anterior to the body of the left crus, with the most cephalad suture placed through the apex of the left crus. A portion of the anterior fundic wall is rolled over to the right and sutured to the apex of the right crus and down the esophagus. If the Dor operation is performed in combination with a left anterolateral myotomy, the anterior medial fundus is sutured to the edges of the myotomy. The Dor operation does not require extensive dissection of the esophagus or hiatus.

The Hill repair fixes the lower esophageal sphincter within the abdo-men by anchoring the phrenoesophageal ligament along the lesser curvature of the gastroesophageal junction to the median arcuate ligament. The Hill repair, as originally described, does not include a fundoplication. Its effectiveness thus demonstrates, as first recognized by Allison and by Barrett, the importance of reducing the hernia, maintaining a length of intra-abdominal esophagus, and preserving the angle of His between the esophagus and the fundus of the stomach.

Complications of antireflux surgery

Collective experience with the open transabdominal Nissen fundoplication shows an operative mortality of 0.1 per cent. This low rate seems to have been maintained in the laparoscopic era. Morbidity after laparoscopic operations is also similar to that reported for open surgery, averaging 5 to 15 per cent. The conversion rate to open operation is generally less than 5 per cent. The main intraoperative complications are pneumothorax or pneumomediastinum, perforation of the esophagus or stomach, splenic injury, and bleeding. Pneumothorax and pneumomediastinum are caused by excessive hiatal and mediastinal dissection. Perforations occur most often during hiatal and circumferential dissection of the esophagus. If recognized, they can usually be repaired laparoscopically, but an unrecognized perforation is a life-threatening

complication. The incidence of intraoperative complications is related to the surgeon's experience, and decreases with increasing experience.

In the early postoperative period, pulmonary atelectasis or collapse and ileus are the commonest complications. Other less common complications include cardiac arrhythmia, pulmonary embolism, urinary infection, and port site hematoma. The incidence of transient dysphagia after laparoscopic surgery was initially higher than that for open surgery, but has returned to acceptable levels with increasing familiarity with the laparoscopic techniques. Within the first 3 months after laparoscopic Nissen fundoplication, up to 50 per cent of patients experience some dysphagia. After 3 months, however, the overall incidence of *de novo* postoperative dysphagia requiring dietary modification should be less than 2 per cent. Many patients report an increase in flatus, epigastric bloating, and inability to vomit after an antireflux procedure. Some patients may have substernal or epigastric discomfort in the immediate postoperative period, which they may believe is attributable to recurrent reflux. These patients do not have recurrent gastroesophageal reflux disease on pH testing and the symptom is probably due to transient esophageal distension due to outflow resistance through the edematous and indurated repair. These symptoms typically resolve over 8 to 12 weeks following the operation.

Effectiveness of antireflux operations

Follow-up studies of open antireflux operations show that the Nissen and Belsey Mark IV procedures reduce 24-h esophageal acid exposure to normal, heal esophagitis, and alleviate the symptoms of reflux in more than 80 per cent of patients at 5 or more years after surgery. The full Nissen fundoplication appears to provide significantly better antireflux control than the partial Belsey fundoplication. Successful Nissen fundoplication results in less esophageal acid exposure on 24-h pH testing than healthy controls.

Similar follow-up data for laparoscopic antireflux procedures have not yet been produced, but relief from typical reflux symptoms has been reported in more than 90 per cent of patients at 6 to 30 months after operation. Some follow-up studies after laparoscopic antireflux operations have included postoperative 24-h pH testing, with the longest follow-up at 21 months. These studies have reported abolition of reflux in more than 85 per cent of patients. In general, follow-up studies of patients after laparoscopic antireflux surgery suggest that the good results obtained with the open operations will be duplicated or even improved upon with the laparoscopic procedure, although the long-term durability of the laparoscopic repair is still unknown.

Reoperation for failed antireflux surgery

Failure of an antireflux procedure is manifested by the patient's inability to swallow normally, persistent epigastric or substernal discomfort during or after meals, or the recurrence of reflux symptoms. Identifying the cause of failure after antireflux surgery and correcting it can be challenging, and patients will need thorough investigation before reoperation. Placement of the fundoplication around the stomach, or ‘slipped' Nissen, is the commonest cause for failure after open antireflux surgery, while herniation of the repair into the chest is the commonest problem after laparoscopic operation. In both situations, an unrecognized short esophagus can be the cause of the failure. During laparoscopic surgery the pneumoperitoneum elevates the diaphragm and facilitates the repair, but after the pneumoperitoneum is deflated and the crura relax, tension from the short esophagus can lift the repair above the diaphragm. Partial or complete disruption of the fundoplication is another cause of failure of an antireflux procedure, and can occur after both open and laparoscopic repairs.

It may be possible to perform a laparoscopic reoperation to correct a previous failed antireflux repair if only one operation has been performed and the initial operation was done laparoscopically. In other situations, a left thoracotomy with a peripheral circumferential incision in the diaphragm is the preferred approach for reoperation. This approach allows safe dissection of the previous repair above and below the diaphragm. The best results for reoperation are obtained for recurrent reflux symptoms in patients without dysphagia or an esophageal body motility disorder. A good result can also be expected following reoperation for dysphagia caused by a correctable technical error. When dysphagia is associated with deterioration of esophageal motility after Nissen fundoplication, improvement may result from taking down the full fundoplication and substituting a partial fundoplication. In patients with chronic severe motility abnormalities of the esophageal body and a history of multiple failed antireflux operations, the benefits of reoperation are reduced with each succeeding operation, and consideration should be given to esophageal resection and replacement. The vagal-sparing method of esophagectomy is most suitable in this situation, but is commonly unable to be performed due to difficulty in identifying the vagi in an area of extensive scarring, or because the nerves have been injured.

Summary guidelines for the treatment of gastroesophageal reflux disease

A suggested approach to the management of patients with gastroesophageal reflux disease symptoms is to start with a trial of lifestyle modifications and empirical symptomatic treatment for 4 to 8 weeks. Antacids, alginic acid, or histamine H_2-receptor antagonists, or a combination of these can be used at this stage. Patients who have complete relief of symptoms with these measures, with no return of symptoms after cessation of the medications, need no further investigation or treatment. Failure to control symptoms, or return of symptoms after medication cessation, suggests that either the diagnosis is incorrect or that the patient has relatively severe disease. Endoscopic examination, manometric evaluation of the lower esophageal sphincter and esophageal body, and 24-h pH studies are then warranted. Other indications for endoscopy are given in the section on diagnostic evaluation.

Features which predict an increased likelihood of a poor response to medical therapy are a permanently defective lower esophageal sphincter, bipositional (both upright and supine) reflux, duodenogastroesophageal reflux, poor esophageal body contractility, and severe esophagitis or Barrett's esophagus. Patients with these features should have the option of antireflux surgery as the initial modality of treatment. Patients without these features may be treated according to their preference with either medical or surgical treatment. Patients with mild esophagitis who prefer medical treatment should be treated with an 8-week course of standard-dose proton pump inhibitors, after which they can be given the option of continuing or stopping the medications. Most patients who stop medical therapy at this point will relapse, but as many as 30 per cent will be able to stop medical treatment indefinitely. Patients with severe esophagitis treated medically should be given an 8- to 12-week course of double-dose proton pump inhibitors, perhaps with an additional evening dose of H_2-receptor antagonists, with confirmation of healing by endoscopy at the end of the course. After healing has been obtained, these patients need long-term, continuous, standard-dose proton pump medication.

In general, the option of surgical therapy should be considered if continuous medical treatment is needed, if escalating doses of proton pump inhibitors are required to control symptoms, if endoscopic healing does not occur, if the cost of medications is a financial burden, or if the patient is young. Patients with Barrett's esophagus treated medically or surgically need regular surveillance endoscopy to detect the development of dysplasia or invasive malignancy.

Hiatus hernia

Definition and classification

Hiatus hernia is an extension of the peritoneal cavity into the chest through the esophageal hiatus of the diaphragm. The hernia consists of a peritoneal sac which usually contains part or all of the stomach but can rarely contain other abdominal viscera such as the colon or spleen. Hiatus hernias are classified according to the position of the gastroesophageal junction in relation to the diaphragm. The type I or sliding hiatus hernia is characterized by proximal herniation of the gastroesophageal junction though the esophageal diaphragmatic hiatus into the posterior mediastinum (Fig. 29). In a type II hernia, also known as a rolling or pure paraesophageal hernia, the gastroesophageal junction is anchored in its normal subdiaphragmatic position, and the gastric fundus herniates alongside the esophagus into the chest (Fig. 30). A type III hernia, also known as a combined sliding/rolling, or mixed hernia, is usually a large hernia and is characterized by proximal herniation of both the gastroesophageal junction and the gastric fundus through a widely open diaphragmatic hiatus (Fig. 31). Type III hernias are commonly but inappropriately classified as type II paraesophageal hernias, but a pure type II hernia is a rare finding. The endstage of a type I or II hernia is a large type III hernia, which occurs when the entire stomach migrates up into the chest by rotating 180° around its longitudinal axis, with the gastroesophageal junction and pylorus as fixed points (Fig. 32). The greater curvature is thus uppermost in the chest. In this situation, the abnormality is usually referred to as an intrathoracic stomach with organoaxial volvulus (Fig. 33). In a type IV hernia, the hiatal defect is very large and has allowed other organs such as the colon, spleen, small intestine, and pancreas to enter the chest. Type IV hernias develop from type III hernias, and are exceedingly uncommon. For this chapter, type IV hernias are classified as type III hernias, so that only three types of hernia will be discussed.



Fig. 29. Barium esophagram showing a type I (sliding) hiatus hernia.

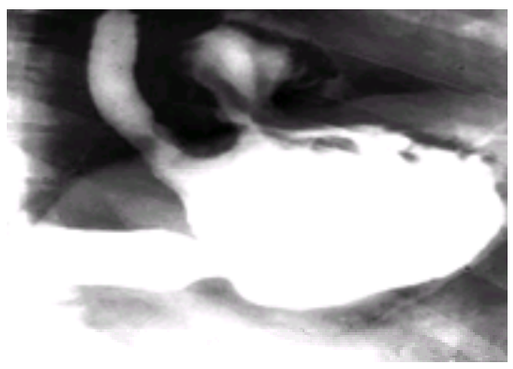


Fig. 30. Barium esophagram showing a type II rolling or pure paraesophageal hernia. In this case the fundus of the stomach herniated into the chest through a separate defect in the diaphragm adjacent to the esophageal hiatus, but the fundus may also herniate alongside the esophagus through the hiatus.

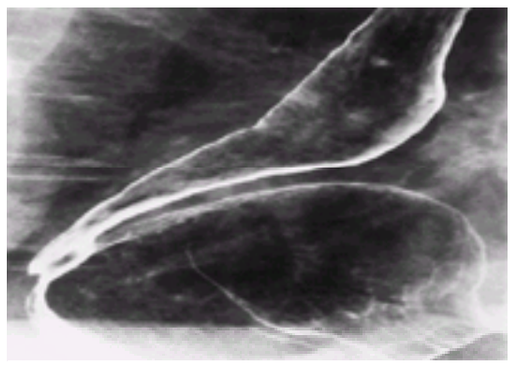


Fig. 31. Barium esophagram showing a type III (mixed sliding/rolling) hiatus hernia.



Fig. 32. Barium esophagram showing a very large hiatus hernia with an intrathoracic stomach.

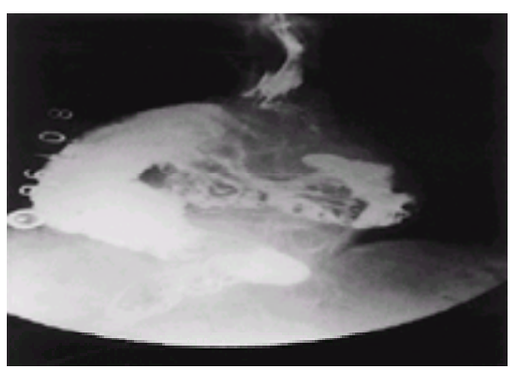


Fig. 33. Barium esophagram showing a large type III hernia with organoaxial volvulus.

Physiologic herniation

In normal individuals, the esophagus is fixed to the diaphragm and significant herniation through the hiatus is impossible. During swallowing, the gastroesophageal junction and the gastric cardia move cephalad to lie above the diaphragm for several seconds. The gastric cardia should be distinguished from the cardia, which refers to the gastroesophageal junction. The mechanism for swallow-induced physiologic herniation is that, with swallowing, contraction of the pharyngeal strap muscles and the longitudinal muscles in the esophageal wall shortens the esophagus and elevates the gastroesophageal junction. This elevation is accompanied by stretching of the elastic phrenoesophageal membrane. With relaxation of the strap and longitudinal muscles, the elastic recoil of the phrenoesophageal membrane returns the gastroesophageal junction to its usual subdiaphragmatic position. On a barium study, the widened area below the tubular esophagus that herniates with swallowing is termed the phrenic ampulla, and it contains the gastric cardia and the intra-abdominal portion of the esophagus.

Prevalence of pathologic herniation

The majority of hernias, probably more than 95 per cent, are type I sliding hernias. The exact prevalence of hiatus hernia, both in the general population and in patients with foregut symptoms, remains unclear. Reasons for this uncertainty are that most hernias are small sliding hernias which reduce in the upright position, most are asymptomatic, and the detection rate varies with the method used for diagnosis and with the age of the population studied. The reported prevalence of hiatus hernia detected by barium contrast studies in the general population ranges from 10 per cent to more than 70 per cent. It is likely that some of these studies, especially if they are based on static rather than videotaped imaging, classified some physiologic hernias as pathologic hernias. A reasonable estimate of the prevalence of pathologic herniation in the general adult population is 20 per cent, with the prevalence of hernias in patients with gastroesophageal reflux disease being 80 per cent. The mean age of patients with a type I hernia is approximately 50 years and for those with a type III hernia is approximately 60 years. Type III hernias are more common in females, with a female:male ratio of 4:1.

Pathophysiology

A pathologic hiatus hernia is an acquired condition. Anatomic alterations found in patients with hiatus hernia are widening of the hiatus and thinning and lengthening of the phrenoesophageal membrane. The phrenoesophageal membrane consists of an upper fascial layer of loose connective tissue and a lower fascial layer of thicker, stronger, elastic tissue (Fig. 4). The upper layer is formed by a supradiaphragmatic continuation of the endothoracic fascia. The more important lower layer is formed by a subdiaphragmatic continuation of the transversalis fascia. The phrenoesophageal membrane functions as an elastic barrier. In comparison with the other diaphragmatic openings, the esophageal hiatus is relatively incompletely filled by the esophagus, and the barrier function of the phrenoesophageal membrane is needed to close the gap between esophagus and diaphragm. The elastic function of the phrenoesophageal membrane allows it to stretch and contract with swallowing and respiration. There are normally more than 1000 swallows per day, and the diaphragm moves up and down between 16 000 and 35 000 times per day. The lengthening and loss of elasticity of the phrenoesophageal membrane that is found in a hiatus hernia is thought to be a gradual age-related wear and tear effect, but obesity, pregnancy, and sudden traumatic elevations in the intra-abdominal pressure may all place additional pressure on the phrenoesophageal membrane.

Mixed or type III hernias are thought to be complications of type I hernias, and are more likely in patients who have an abnormally wide esophageal hiatus. Pure paraesophageal or type II hernias are thought to arise by the same mechanism as sliding hernias, with the difference that the gastroesophageal junction is fixed below the hiatus by firm posterior fixation to the preaortic fascia and the median arcuate ligament. Another difference is that the herniation may occur either through a widened esophageal hiatus or through a defect adjacent to the esophageal hiatus. The fundus of the stomach rolls up alongside the esophagus through the widened hiatus or the adjacent defect, but the gastroesophageal junction remains in its normal intra-abdominal position.

A hiatus hernia predisposes to the development of gastroesophageal reflux by four mechanisms. First, contraction of the crural components of the diaphragm augments the antireflux mechanism by buttressing the lower esophageal sphincter, adding the pressure produced by crural contraction to the underlying pressure in the lower esophageal sphincter. The presence of a hiatus hernia interferes with this buttressing because the normal alignment between the crura and the lower esophageal sphincter is lost. This encourages reflux during periods of increased intra-abdominal pressure. Second, the formation of a hiatus hernia results in loss of the acute angle of His. Third, the clearance of refluxed acid from the esophagus is also impaired in patients with hiatal hernias, especially non-reducing hernias, due to loss of anchoring of the distal esophagus. Fourth, the presence of a hiatus hernia interferes with transhiatal flow of gastric juice. This occurs in particular in patients with a non-reducing hernia, in whom the hernia remains above the diaphragm during inspiration. Fluid trapped within the supradiaphragmatic stomach is prevented from flowing into the subdiaphragmatic stomach by closure of the crura during inspiration, and thus refluxes freely into the negative-pressure thoracic esophagus.

Symptoms

Symptoms in patients with a sliding hernia are mostly those of the associated gastroesophageal reflux. The likelihood of reflux increases with increasing hernia size. Dysphagia and intermittent vomiting can occur in patients with a large sliding hernia, due to diaphragmatic contraction obstructing the passage of a bolus through the herniated stomach. Dysphagia is common in patients with type II hernias, due to compression and angulation of the esophagus by the herniated gastric fundus. In large type III hernias with gastric volvulus, dysphagia can be caused by twisting of the gastroesophageal junction. Recurrent bleeding or anemia may be caused by ulceration or erosive gastritis secondary to ischemia of the herniated gastric mucosa. Crural contractions may also cause ‘riding ulcers' at the level of the diaphragmatic indentations. Respiratory symptoms such as dyspnea may occur from compression of the lungs by a very large hernia, and cough, choking, and pneumonia may result from aspiration of regurgitated gastric juice into the airways.

Type II or III hernias can cause acute life-threatening events in up to 20 per cent of patients due to acute obstruction from gastric torsion or volvulus. Swallowed air causes gastric distention, which compresses the blood supply passing to the intrathoracic stomach along the margins of the diaphragmatic hiatus (Fig. 34). This can result in gastric ischemia, perforation, and sepsis. The triad of epigastric pain, inability to vomit, and difficulty in the ability to pass a nasogastric tube is an indication for immediate endoscopic decompression or operative intervention.

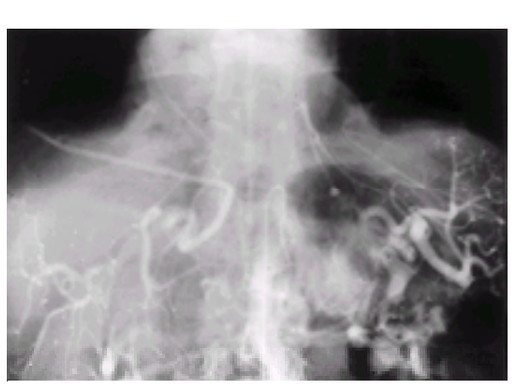


Fig. 34. Angiogram showing occlusion of the arterial blood supply to the stomach in a patient with an incarcerated type III hernia.

Diagnosis

Radiology

Large hernias may be detected on an upright or lateral plain chest radiograph by the presence of a gas-filled viscus behind the heart. Smaller hernias require a barium swallow study for diagnosis. Barium studies are diagnostic in more than 90 per cent of patients. The barium swallow study should include examination in both upright and supine positions, to detect sliding hernias that reduce when upright. Video-esophagram studies demonstrate that most sliding hernias empty during expiration.

Endoscopy

A hiatus hernia is present endoscopically when the anatomic gastroesophageal junction, identified as the proximal extent of the gastric rugal folds, is more than 2 cm above the crura, identified by having the patient sniff (Fig. 4). Small hernias are best detected endoscopically if the endoscope is retroflexed within the stomach, allowing inspection of the gastroesophageal junction from below. The geometry of the gastroesophageal junction seen on retroflexed view can be classified using the Hill grading system (Fig. 35). This classification system emphasizes that development of a sliding hernia is accompanied by loss of the angle of His, signified by the loss of the normal frenulum over the endoscope on retroflexed viewing. A true paraesophageal hernia is identified on retroflexed view by noting a hernia orifice adjacent to the gastroesophageal junction. A type III hernia appears on retroflexed view as a sliding hernia with the gastroesophageal junction entering on the side of the sac, rather than at the apex of the sac as in a type I hernia.

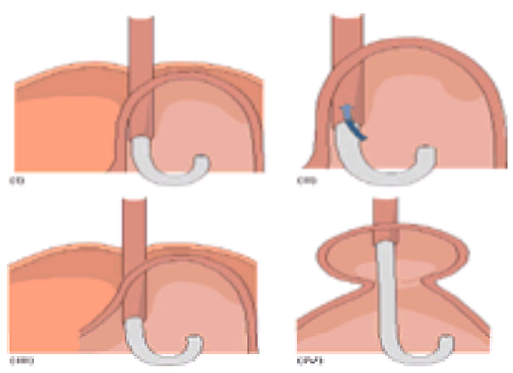


Fig. 35. The Hill grading system for classifying the competency of the gastroesophageal flap valve. (I) Retroflexed view of a grade I gastroesophageal valve. A prominent fold of tissue grips the shaft of the endoscope and extends 3 to 4 cm along the lesser curve at the entrance of the esophagus into the stomach. (II) The fold of tissue is less prominent in a grade II gastroesophageal valve, and there are occasional periods of opening and rapid closing around the endoscope with respiration. (III) There is no prominent fold at the entrance of the esophagus into the stomach for a grade III gastroesophageal valve, and the endoscope is not gripped tightly by the tissues. (IV) An obvious hiatus hernia is present. There is essentially no fold of tissue at the gastroesophageal junction, and the squamous epithelium in the

esophagus can be seen from within the stomach.

Manometry

Stationary manometry studies are not used for the diagnosis of hiatus hernia, but an awareness of manometric features of hiatus hernia is helpful since many patients undergoing manometry for the investigation of reflux disease have a hiatus hernia. A ‘double hump' motility pattern is characteristic of a sliding hernia. As the manometry catheter is withdrawn from the stomach into the esophagus, pressurized areas or ‘humps' are recorded, with the distal hump corresponding to the crural compression of the stomach and the proximal hump corresponding to the lower esophageal sphincter. Between the two humps is a plateau region in which the pressure is slightly above the gastric baseline pressure. The plateau region corresponds to the hernia sac. The pressure fluctuations that occur with respiration indicate that the hernia sac may be exposed to both intra-abdominal and intrathoracic pressure conditions, depending upon whether the crura compartmentalize the herniated proximal stomach from the distal stomach and peritoneal cavity.

Treatment

Sliding hernias do not require treatment unless they produce symptoms specific to the presence of the hernia. In patients with gastroesophageal reflux disease, reduction of the hernia and closure of the hiatus is an essential part of operative treatment, and as a consequence the hernia is repaired. Patients with large hernias undergoing antireflux surgery are likely to have a shortened esophagus requiring special operative techniques, as discussed previously. The high rate of serious or life-threatening complications in patients with type II or III hernias indicates that operative repair should be performed in all patients with these hernias, regardless of the presence or severity of symptoms or the size of the hernia. Type II or III hernias progressively worsen unless repaired by operation. There is an approximately 20 per cent mortality for emergency repair of an incarcerated paraesophageal hernia, compared with less than 1 per cent mortality for elective repair of a non-incarcerated hernia.

Repair of type II or III hernias by laparoscopic methods can be difficult and is associated with a higher rate of recurrence. We prefer an open approach to these large hernias. An open transthoracic approach allows full esophageal mobilization, resection of the sac, and proper closure of the wide hiatus. An open approach also facilitates performance of a Collis gastroplasty if this is needed to gain adequate esophageal length to reduce tension on the repair. The transabdominal open approach facilitates reduction of a volvulus, but it may be difficult to obtain an adequate view for dissection in the posterior mediastinum, the risk of vagus nerve injury is high, and stout closure of the widened hiatus may be compromised. At a median follow-up of 17 months following repair of large type III hernias at our institution, a recurrent hernia was found on video-esophagram examination in 9 of 22 (41 per cent) patients who underwent a laparoscopic repair. This rate of reherniation was significantly higher than the 15 per cent (3 of 20 patients) found at a median 35 months after open repair. This result suggests that the open approaches are preferable to laparoscopic operation for the treatment of pure paraesophageal or mixed hiatus hernias.

Further reading

Campos GMR *et al.* Multivariate analysis of the factors predicting outcome after laparoscopic Nissen fundoplication. *Journal of Gastrointestinal Surgery* 1999; **3**: 292–300. [Use of 24-h pH monitoring provides the best single outcome predictor for the success of laparoscopic Nissen fundoplication.]

Costantini M, Zaninotto G, Anselmino M, Boccu, C, Nicoletti L, Ancona E. The role of a defective lower esophageal sphincter in the clinical outcome of treatment for gastroesophageal reflux disease. *Archives of Surgery* 1996, **131**: 655–9. [Patients with a defective sphincter who decline surgery require lifelong medical therapy.]

DeMeester SR *et al.* The impact of an antireflux procedure on intestinal metaplasia of the cardia. *Annals of Surgery* 1998; **228**: 547–56. [Loss of intestinal metaplasia after antireflux surgery is rare in patients with macroscopically visible intestinal metaplasia (Barrett's esophagus), but occurs in patients with non-macroscopic intestinal metaplasia at the gastroesophageal junction.]

DeMeester TR, Bonavina L, Albertucci M. Nissen fundoplication for gastroesophageal reflux disease—evaluation of primary repair in 100 consecutive patients. *Annals of Surgery* 1986; **204**: 9–20. [The open Nissen fundoplication operation can provide good long-term symptom control.]

DeMeester TR, Peters JH, Bremner CG, Chandrasoma P. Biology of gastro-esophageal reflux disease: pathophysiology relating to medical and surgical treatment. *Annual Review of Medicine* 1999; **50**: 469–506. [A review of the mechanical basis of gastroesophageal reflux disease.]

DeVault KR, Castell DO. Updated guidelines for the diagnosis and treatment of gastroesophageal reflux disease. *American Journal of Gastroenterology* 1999; **94**: 1434–42. [The current American College of Gastroenterology practice guidelines.]

Isolauri J *et al.* Long-term comparison of antireflux surgery versus conservative therapy for reflux esophagitis. *Annals of Surgery* 1997; **225**: 295–9. [At a median follow-up of 10.9 years, better subjective and objective results were found after surgical than medical treatment.]

Johnsson F, Holloway RH, Ireland AC, Jamieson GG, Dent J. Effect of fundoplication on transient lower oesophageal relaxation and gas reflux. *British Journal of Surgery* 1997; **84**: 686–9. [Fundoplication inhibits the triggering of transient lower esophageal sphincter relaxation and prevents the complete ablation of pressure at the gastroesophageal junction during lower esophageal sphincter relaxation.]

Kauer WKH, Peters JH, DeMeester TR, Heimbucher J, Ireland AP, Bremner CG. A tailored approach to antireflux surgery. *Journal of Thoracic and Cardiovascular Surgery* 1995; **110**: 141–7. [Excellent symptomatic control is possible if surgical treatment of gastroesophageal reflux disease is tailored to the patient's anatomic and functional status.]

Lagergren J, Bergstrom R, Lindgren A, Nyren O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. *New England Journal of Medicine* 1999; **340**: 825–31. [There is a strong association between recurrent symptoms of gastroesophageal reflux and the development of esophageal adenocarcinoma.]

Mason RJ *et al.* Nissen fundoplication prevents shortening of the sphincter during gastric distension. *Archives of Surgery* 1997; **132**: 719–24. [By preventing sphincter shortening, a Nissen fundoplication improves competency of the lower esophageal sphincter to progressive degrees of gastric distension.]

Öberg S *et al.* Inflammation and specialized intestinal metaplasia of cardiac mucosa is a manifestation of gastroesophageal reflux disease. *Annals of Surgery* 1997; **226**: 522–30. [The findings of cardiac mucosa, carditis, and intestinal metaplasia at the gastroesophageal junction may be the earliest signs of gastroesophageal reflux disease.]

Orlando RC. Review article: oesophageal mucosal resistance. *Alimentary Pharmacology and Therapeutics* 1998; **12**: 191–7. [A review of the structure and functions of the esophageal mucosa that contribute to its defense against injury caused by exposure to refluxed gastric acid.]

Peters JH, Heimbucher J, Kauer WK, Incarbone R, Bremner CG, DeMeester TR. Clinical and physiologic comparison of laparoscopic and open Nissen fundoplication. *Journal of the American College of Surgeons* 1995; **180**: 385–93. [The laparoscopic and open operations provide equivalent early control of gastroesophageal reflux.]

Peters JH *et al.* The treatment of gastroesophageal reflux disease with laparoscopic Nissen fundoplication. *Annals of Surgery* 1998; **228**: 40–50. [Laparoscopic Nissen fundoplication provides an excellent symptomatic and physiologic outcome in patients with proven gastroesophageal reflux and typical symptoms.]

Ritter MF *et al.* Outcome after laparoscopic fundoplication is not dependent on a structurally defective lower esophageal sphincter. *Journal of Gastrointestinal Surgery* 1998; **2**: 567–72. [Relief of typical reflux symptoms after antireflux surgery is excellent, regardless of the resting mechanical status of the lower esophageal sphincter.]

Sontag SJ *et al.* Most asthmatics have gastroesophageal reflux with or without bronchodilator therapy. *Gastroenterology* 1990; **99**: 613–20. [More than 80 per cent of adult with asthma have abnormal gastroesophageal reflux.]

Spechler SJ. Comparison of medical and surgical therapy for complicated gastroesophageal reflux disease in veterans. The Department of Veterans Affairs Gastroesophageal Reflux Disease Study Group. *New England Journal of Medicine* 1992; **32**: 786–92. [Surgery was significantly more effective than medical therapy in improving the symptoms and endoscopic signs of esophagitis in men with complicated gastroesophageal reflux disease.]

22.2.2 Benign oesophageal strictures

Reginald V. N. Lord and Tom R. DeMeester

Reflux strictures
Symptoms
Investigations
Treatment
Dilatation
Surgery for reflux strictures
Pill-induced injury
Corrosive injury
Infection
Iatrogenic
Crohn's disease
Further reading

Strictures are fixed fibrous narrowings of the esophageal lumen. Most occur in the distal esophagus as a complication of gastroesophageal reflux disease and are termed reflux or peptic strictures. Other causes of stricture are listed in Table 1.

Gastroesophageal reflux disease
Corrosive ingestion
Medications (pill-induced stricture)
Iatrogenic
Rings, including Schatzki's ring
Infection
Crohn's disease
Eosinophilic esophagitis
Congenital
Webs
Epidermolysis bullosa
Sarcoidosis

Table 1 Etiology of benign esophageal strictures

Reflux strictures

Between 2 and 10 per cent of patients with gastroesophageal reflux disease develop a reflux stricture. This incidence seems to have decreased in the last decade, while other complications of reflux, such as Barrett's esophagus and Barrett's-associated adenocarcinomas, have markedly increased in incidence. It may be that more patients with gastroesophageal reflux disease are being investigated and treated at an earlier stage in their disease, and that the potent acid-suppression drugs introduced in the late 1970s are effectively preventing the development of strictures. In contrast, these drugs presumably have less effect, or even an adverse effect, on the development of Barrett's esophagus and its complications.

Reflux strictures develop as a result of reflux esophagitis with chronic inflammation and fibrosis, which in advanced cases involves the full thickness of the esophageal wall. Most are less than 1 cm in length, but longer strictures may result from a combination of reflux disease and pill-induced injury. The principal antireflux mechanisms are the mechanical barrier provided by the lower esophageal sphincter, the clearance of refluxed fluid by the peristaltic contractions of the esophageal body, and a gastric reservoir that empties effectively. These elements are more frequently and severely impaired in patients with a reflux stricture than in patients with gastroesophageal reflux disease without stricture. More than 80 per cent of patients with stricture have a mechanically defective lower esophageal sphincter (Fig. 1), and up to 64 per cent have motility disorders of the esophageal body. Reflux strictures are common in patients with scleroderma, due to the loss of both the lower esophageal sphincter and esophageal body motility. Although the role of impaired gastric emptying in the development of reflux strictures is not well understood, defective gastric emptying increases both the frequency and the severity of reflux episodes.

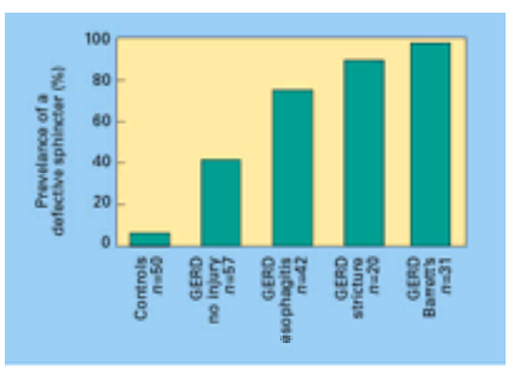


Fig. 1. The relationship between various degrees of mucosal injury and the prevalence of a mechanically defective lower esophageal sphincter. The majority of patients with mucosal injury have a permanently defective sphincter. More than 80 per cent of patients with a stricture have a defective sphincter.

The proportion of a 24-h period that the distal esophagus is exposed to acid at pH less than 4 is the standard measurement for the diagnosis of gastroesophageal reflux disease, and is significantly higher in patients with strictures than in those with either no mucosal injury or esophagitis but no stricture (Fig. 2). The median percentage time that the pH is less than 4 in patients with reflux stricture is approximately 15 per cent, which is more than three times the upper limit of acid exposure in normal individuals. Patients with strictures also have significantly more duodenogastroesophageal reflux than do patients with gastroesophageal reflux disease without strictures.

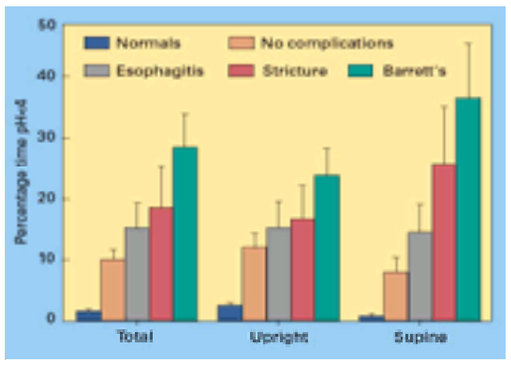


Fig. 2. Acid exposure in the distal esophagus expressed as a percentage of total time and time in the upright and supine position during which pH was less than 4. (From Zaninotto *et al.*; see Further reading.)

Symptoms

Dysphagia for solids is the typical presenting symptom in patients with a stricture, and when advanced, dysphagia for liquids may also occur. The likelihood of dysphagia is related to the esophageal diameter. Radiologic studies by Schatzki and Gary showed that dysphagia is usual in patients with a diameter less than 13 mm, and may occur in those with a diameter between 13 and 20 mm. Dysphagia usually develops slowly, and patients may unconsciously adjust their eating habits to account for the difficulty with swallowing. It is thus important to take a detailed swallowing history, questioning patients as to whether they cut food into small pieces, whether they always finish eating after others, or if they avoid eating in company. Dietary modifications usually allow patients with strictures to maintain their weight, but significant weight loss may be present in those with tight strictures. Malignancy or a named motility disorder of the esophageal body needs to be excluded in all patients with dysphagia, and early endoscopy with biopsy is essential.

Other symptoms of gastroesophageal reflux disease such as heartburn and regurgitation may improve after the development of a stricture if it provides a significant barrier to reflux. A complete medication history, including medications taken at any time, is important to identify the possibility of a superimposed pill-induced stricture.

Investigations

A barium esophagram should be the first investigation performed in patients with dysphagia or a suspected stricture (Fig. 3 and Fig. 4). The barium study provides information about the number, location, tightness, and length of strictures, and it may suggest the presence of cancer, achalasia, or another disorder. This information is useful for subsequent endoscopy, allowing the endoscopist to plan for the use of fluoroscopy, dilatation, or endoscopic ultrasonography. Furthermore, the barium study is more sensitive than endoscopy for detecting strictures, especially if the stricture diameter is relatively wide and if a narrow (9 mm) endoscope is used. The accuracy of a barium study is increased by asking the patient to swallow a barium hamburger bolus, a 13-mm barium tablet, or a barium marshmallow, in which ase obstruction of the bolus at the stricture may occur.



Fig. 3. Barium esophagram showing a reflux-induced stricture. The stricture is at the gastroesophageal junction.



Fig. 4. Barium esophagram showing a long, complex, distal esophageal stricture.

Endoscopy is necessary to determine the cause of a stricture and detect additional pathology. Biopsies should be taken of both the stricture and the surrounding mucosa. Reflux strictures always involve the squamocolumnar junction. If a stricture is above the squamocolumnar junction, a non-reflux cause for the stricture must be sought. Reflux strictures usually have a smooth surface, and the presence of mucosal irregularity may be due to an underlying occult malignancy. Strictures can interfere with reflux and should be dilated before a 24-h pH test is performed to confirm the diagnosis of gastroesophageal reflux disease.

A stricture is often present in patients with Barrett's esophagus. In these patients, the squamocolumnar junction, and thus the reflux-induced stricture, is at the upper limit of the Barrett's columnar segment. In the past, most patients with Barrett's esophagus also had a stricture, but the current prevalence of stricture in patients with Barrett's esophagus is approximately 30 per cent.

A concentric ring, termed a lower esophageal or Schatzki's ring, may be detected radiologically or endoscopically at the squamocolumnar junction (Fig. 5). These rings show varying amounts of inflammatory cells and fibrosis in the mucosa and in some instances in the submucosa. They are commonly found in association with a hiatus hernia, in which case the ring marks the upper limit of the hernia. Although Schatzki's rings are not usually classified as strictures, it can be impossible to distinguish between the localized esophageal narrowing produced by a ‘true' reflux stricture and that produced by a Schatzki's ring. Both occur at the squamocolumnar junction, most patients with Schatzki's rings have objective gastroesophageal reflux disease on 24-h pH testing, and rings may develop into fixed strictures with transmural fibrosis. Dilatation is needed for symptomatic rings, and patients with associated gastroesophageal reflux disease need treatment for their reflux.



Fig. 5. Barium esophagram showing a Schatzki's ring. The ring is a circumferential narrowing at the squamocolumnar junction above a sliding hiatus hernia.

Treatment

The natural history of untreated reflux strictures is that they persist as chronic lesions that progressively narrow the esophageal lumen, causing gradually worsening dysphagia. Treatment involves dilatation of the stricture and either medical or surgical management of the reflux disease.

Patients with reflux-induced strictures already have severe reflux disease, and aggressive antireflux therapy is therefore indicated. Lifestyle and dietary modifications should be encouraged. Most patients are treated by a combination of dilatation and acid suppression using proton pump inhibitor drugs such as omeprazole and lansoprazole. Randomized trials that measured symptom control and the need for dilatation have demonstrated that proton pump inhibitors are more effective than H_2-receptor antagonists for the treatment of reflux strictures. After initial dilatation, between 21 and 41 per cent of patients treated with proton pump inhibitors will need redilatation within 12 months.

Dilatation

Dilatation of a benign esophageal stricture is indicated to relieve dysphagia or incipient dysphagia, to permit a diagnostic 24-h pH study to be performed, or to allow passage of the endoscope through a tight stricture. Even small degrees of lumenal dilatation can produce significant symptomatic relief. Most long-term studies have reported that the majority of patients need repeat dilatation. Adequate control of gastroesophageal reflux disease lessens the need for repeat dilatations. Neither the degree of stenosis nor the length of the stricture is significantly associated with the need for repeat dilatations.

There are two types of dilators: axial dilators or bougies, which fracture the fibrous narrowing as they are forcibly passed through the stricture, and radial or balloon dilators, which are positioned across the stricture and then inflated until stricture disruption occurs (Fig. 6). The ‘rule of threes' for performing dilatation states that no more than three dilatations that meet resistance should be attempted at one session. To relieve dysphagia reliably, the stricture should be dilated to a diameter of more than 40 French for a bougie, or more than 15 mm for a balloon dilator.

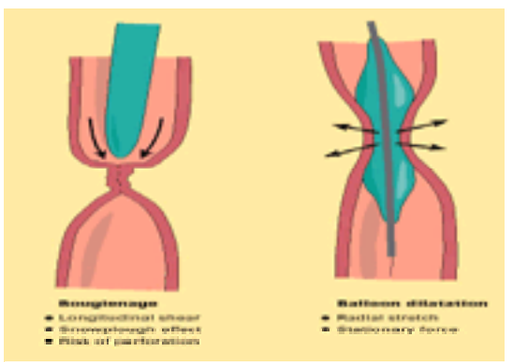


Fig. 6. Schematic representation of the different forces generated by bougienage compared with balloon dilatation.

Mercury-filled rubber bougies are easy, economical, and safe to use. The Maloney mercury-filled bougies have a tapered tip, whereas the Hurst bougies have a blunt tip. These bougies may be passed with only topical gel anesthesia, and patients who require long-term regular dilatations can be taught to use them at home. In an awake patient, mercury-filled bougies should be passed with the patient sitting in the upright position.

The Savary–Gilliard, Eder–Puestow, and American dilators consist of a guidewire and different size bougies which are passed over the guidewire. The guidewire is threaded through the stricture either at endoscopy or with fluoroscopic assistance. Guidewire dilators are preferred to mercury-filled dilators for tight, long, or complicated strictures.

Through-the-scope or TTS balloon dilators are passed through the biopsy channel of an endoscope. They thus allow visualization of the stricture during the dilatation procedure. They are easy to use, but are not reusable and are thus expensive. Randomized trials have found no significant difference in the measured relief of dysphagia and need for repeat dilatation between bougie and balloon dilators.

The main complication of dilatation is esophageal perforation, which occurs in approximately 0.5 per cent of procedures. If perforation is suspected, a Gastrografin contrast study should be obtained prior to discharging the patient from the endoscopy suite.

Surgery for reflux strictures

Patients with gastroesophageal reflux disease complicated by stricture have severe reflux disease. These patients almost always have a mechanically defective lower esophageal sphincter, along with an esophageal body dysmotility. Most patients are initially treated by a combination of multiple dilatations and acid suppression therapy.

The principal indication for surgery in these medically treated patients is the demonstrated need for frequent dilatations whilst on an adequate dose of proton pump inhibitor medication. An operation may also be required in patients with esophageal perforation complicating dilatation. Esophageal shortening is often present in patients with strictures, and unless a Collis gastroplasty procedure is done to restore adequate esophageal length before the construction of an antireflux valve, failure of surgical therapy is not uncommon. A recent comparison of medical and surgical treatments for patients with esophagitis treated by the United States Veterans Affairs medical system showed that patients with stricture treated surgically had less need for repeat dilatation than did patients treated medically. Although approximately 20 per cent of patients need repeat dilatation after antireflux surgery, typically only one or two dilatations are required.

Persistent dysphagia in a patient with a non-dilatable stricture, or in a patient with a dilatable stricture and either a history of more than one previous failed antireflux operation or the presence of severely disordered esophageal body motility, are indications for esophageal resection. The ideal operation in these patients is a limited vagal-sparing esophagectomy with colon interposition between the cervical esophagus and intact stomach. Although the stricture is often short, limited resections with an intrathoracic anastomosis are likely to be complicated by severe postoperative reflux esophagitis and Barrett's esophagus. Medical therapy is these patients is not applicable, and an antireflux procedure is likely to fail, because of persistent dysphagia and the need for repetitive dilatations.

Pill-induced injury

Tablet or capsule medications may lodge and dissolve in the esophagus rather than passing into the stomach. If the contents of the medication are noxious to the esophageal mucosa, inflammation will result. Stricture, perforation, mediastinitis, and hemorrhage are complications of severe or repetitive pill-induced injury. Lodgement of pills in the esophagus is more common if the tablets or capsules are large, if they are taken without water or in the supine position, or if the patient has esophageal pathology. Pill-induced injury can also occur in normal individuals. Sustained or delayed-release medications are more harmful than standard-release preparations of the same medications.

The most common site of lodgement is at approximately 25 cm from the incisor teeth, where the mixture of smooth and striated muscles in the muscularis propria results in relatively low peristaltic wave contraction amplitudes, and the esophagus may be extrinsically compressed by the aortic arch. The second commonest site for pill-induced strictures is in the distal esophagus, where they are often misdiagnosed as reflux strictures. Pill-induced injury may occur in association with reflux strictures if noxious medications become lodged above the stricture. The endoscopist should consider the possibility of a superimposed drug injury if a reflux stricture is difficult to dilate. Pill impaction may also occur at the level of the left atrium in patients with left-sided heart failure, or in patients who have had cardiac surgery, in whom the esophagus may be displaced posteriorly and compressed by the heart.

Certain medications are considered more likely to cause esophageal injury (Table 2). More than 20 different antibiotics and antiviral agents have been reported to cause injury, with tetracyclines, especially doxycycline, most frequently implicated. Aspirin and almost all the other non-steroidal anti-inflammatory drugs (NSAIDs) have been reported to cause esophageal injury. It is unclear whether the injury results from the drugs themselves or from the increased reflux-induced injury secondary to NSAID-mediated inhibition of the normal prostaglandin epithelial defense mechanisms. The bisphosphonates alendronate and pamidronate, which inhibit bone resorption and are commonly prescribed for the treatment of osteoporosis, are also associated with a relatively high frequency of esophageal injury. Sodium hydroxide-containing tablets such as Clinitest tablets and denture cleaning tablets can cause a burn injury similar to that produced by ingestion of alkaline corrosive

fluids.

Antibiotics, especially doxycycline and other tetracyclines
Potassium chloride
Quinidine
Aspirin and other NSAIDs
Ferrous sulfate and ferrous succinate
Alendronate and pamidronate
Ascorbic acid (vitamin C)
Phenytoin
Alprenolol
Theophylline and aminophylline
Oral contraceptives

Table 2 Medications which can cause esophageal injury

Esophageal stricture is particularly common after injury caused by four drugs. These drugs are potassium chloride, quinidine, ferrous sulfate or ferrous succinate, and the b-blocker alprenolol. In some series, one or a combination of these agents was involved in the development of more than half of the esophageal strictures studied, and death due to perforation and hemorrhage has been reported in at least seven patients.

Treatment of pill-induced injury involves cessation of the offending medication, analgesia for acute injury if needed, and the treatment of complications. Endoscopy is indicated to assess the extent of injury and to exclude other pathology. The typical acute endoscopic appearance after pill-induced injury is of ulceration at one of the usual sites of pill impaction, sometimes with the pill or parts of the pill still present at the site. In severe cases, there may be circumferential ulceration and a marked exudate, in which case stricture formation is more common. If symptomatic, patients with pill-induced strictures need dilatation. Dilatation can be extremely difficult in some patients because of the extensive fibrosis induced. In these patients, the risk of splitting the stricture is high, and surgical resection is preferred. The ideal operation in these patients is a vagal-sparing esophagectomy with colon replacement.

Corrosive injury

Ingestion of caustic substances such as corrosive cleaning fluids can result in esophageal injury, with stricture development if the injury involves the submucosa or deeper layers of the esophageal wall. Strictures can develop as soon as 2 weeks after injury. Corrosive strictures are often present at more than one level, are commonly irregular, and are often more than several centimeters in length. Most caustic injuries occur in children from accidental ingestion of alkaline household bleaches or detergents. In adults, caustic ingestion is usually associated with a suicide attempt, but accidental ingestion is not uncommon amongst alcoholics. Some pill-induced injuries are also mediated by release of corrosive acids or alkalis. It has been estimated that there are more than 5000 caustic ingestions per year in the United States.

The extent of injury depends on the nature and concentration of the ingested fluid and the duration of exposure. Alkaline fluids are most harmful at concentrations around pH 14, but injury with secondary stricture formation has been reported to occur with ingestion of fluids with lower pH. Attracted to their shiny appearance, children may swallow small disk batteries. If the batteries lodge and dissolve in the esophagus, they can cause both an alkaline and an electrical injury.

Acid corrosives are commonly used as liquid or granular home toilet cleaners. In contrast to the liquefaction necrosis caused by alkaline ingestion, acids cause a coagulation necrosis of the tissues, with formation of a protective eschar. Esophageal injury from acid ingestion is much less prevalent than from alkali ingestion, and it is commonly thought that acids tend to injure the stomach preferentially, leaving the esophagus relatively unaffected. However, acids can cause severe esophageal injury, with all the complications associated with alkaline injury, including perforation and death.

When a caustic injury is suspected, the time of ingestion and the agent involved should be determined. This may require retrieval of the corrosive by a parent after ingestion by a young child, or by someone other than the patient after a suicide attempt. If necessary, the pH of the fluid can be tested. There may be few symptoms or signs of injury in the case of a mild injury, but pain and odynophagia are almost always present in severe cases. Other symptoms and signs are hoarseness, oropharyngeal burns, dysphagia, and chest, abdominal, and back pain. In severe cases, there may be evidence of airway obstruction, aspiration injury, or esophageal perforation with pneumothorax, pneumomediastinum, and septic shock. Chest radiographs should be performed at presentation, and a water-based contrast material should be given prior to the barium swallow study if there is a possibility of perforation.

Endoscopy should be performed within the first 48 h, preferably within the first 24 h. Because of the risk of perforation in acute severe injuries, many endoscopists do not advance the endoscope through a severely injured segment. In this case, the stomach is not inspected and the treating doctor must be vigilant for signs of gastric injury. The extent of injury can be graded endoscopically as first, second, or third degree. First-degree injury is characterized by mild erythema and edema. Second-degree injury is characterized by ulceration and severe edema. Third-degree injury is recognized by deep ulceration, severe edema, lumenal obstruction, and dusky necrosis. Strictures can develop after either second- or third-degree injuries.

The acute management of caustic injuries involves maintaining an airway, fluid resuscitation, pain control, and antibiotic coverage for second- or third-degree injuries. Nutritional support may be needed later. A randomized controlled trial found no benefit for corticosteroid use in children with corrosive injury. Dilatation for acute edematous stricture is no longer performed at most institutions because of the risk of perforation. Most patients with chronic strictures secondary to corrosive injury can be adequately managed with dilatation, although multiple dilatations are often required for continued relief of dysphagia. Some patients need chronic or lifelong intermittent dilatations, in which case they can be taught to pass a mercury-filled bougie in the upright sitting position at home. A late risk of squamous cell carcinoma development has been reported, but this is rare. Esophageal resection is indicated for patients with persistent dysphagia and a chronic stricture or a narrowed non-compliant esophagus. The ideal operation for this is the vagal-sparing esophagectomy with a colon interposition, which provides the best functional reconstruction. If severe fibrosis is present, the nerves and mediastinal tissue may become adherent to the esophagus, and a standard esophageal resection may be required with the colon or stomach used as a substitute, provided that the latter was unaffected by the corrosive injury. Hypopharyngeal strictures pose a special problem and usually need a pharyngoplasty with an esophagectomy to regain acceptable swallowing. Late complications of corrosive injury with stricture formation are esophageal shortening, hiatus hernia, and gastroesophageal reflux disease.

Infection

Infections are rare causes of stricture, but the incidence of this etiology may have increased due to the increased numbers of immunosuppressed patients with HIV infection or other immunosuppressive conditions. *Candida albicans* is the most common organism found, accounting for nearly half the cases. Other pathogens in immunocompromised patients are herpes simplex virus, cytomegalovirus, mucormycosis, histoplasmosis, and tuberculosis.

In the acute phase, infectious strictures are caused by edema, and treatment consists principally of antimicrobial drug therapy. Dilatation is performed rarely, and is accompanied by a significant risk of perforation because of the lack of the fibrotic wall thickening that is found in reflux, drug, and corrosive strictures. Dilatation, and sometimes esophageal resection, are more often required for strictures caused by drug-resistant organisms, such as may occur with mucormycosis, histoplasmosis, and tuberculosis infections. Rarer infectious causes of an esophageal stricture are actinomycosis, syphilis, and blastomycosis.

Iatrogenic

Anastomotic strictures after esophagectomy are more common if the stomach is used for esophageal replacement (18 per cent incidence) than if the colon is used (less than 5 per cent incidence). Anastomotic strictures are treated by dilatation. Strictures may develop in patients after prolonged nasogastric intubation. The mechanism for this probably involves a combination of local trauma and an increase in gastroesophageal reflux around the tube through the open lower esophageal sphincter. Stricture development, presumably secondary to localized fibrosis, has been reported in between 4 and 50 per cent of patients after injection sclerotherapy for esophageal varices. Esophageal strictures may also be caused by radiation therapy for esophageal or other chest or chest wall tumors. These strictures can develop several years after the administration of the therapy, and severe strictures of this type often require resection and reconstruction.

Crohn's disease

Crohn's disease of the esophagus is unusual, with a reported incidence in large series of patients with Crohn's disease ranging from zero to 6.5 per cent. The mid- or distal esophagus is most commonly involved, and patients with esophageal Crohn's disease almost always have ileocolic disease as well. Dysphagia due to a Crohn's stricture may be severe. The radiologic and histopathologic findings are similar to those found for strictures in the small intestine. The endoscopic appearance ranges from mild erosions, to ulceration, to a long tubular stricture. Dilatation and medical therapy provide sufficient relief of dyphagia in almost all patients, and esophageal resection is rarely required.

Further reading

Anderson KD, Rouse TM, Randolph JG. A controlled trial of corticosteroids in children with corrosive injury of the esophagus. *New England Journal of Medicine* 1990; **323**: 637–40. [This randomized trial found no benefit from the use of steroids to treat children with a caustic injury. Stricture development was related only to injury severity.]

Bonavina L, DeMeester TR, McChesney L, Schwizner W, Albertucci M, Bailey RT. Drug-induced esophageal strictures. *Annals of Surgery* 1987; **2**: 173–83. [Pill-induced strictures are not uncommon.]

Bonavina L, Fontebasso V, Bardini R, Baessato M, Peracchia A. Surgical treatment of reflux stricture of the oesophagus. *British Journal of Surgery* 1993; **80**: 317–20. [Results of operative treatment for 65 patients.]

El-Serag HB, Sonnenberg A. Outcome of erosive esophagitis after Nissen fundoplication. *American Journal of Gastroenterology* 1999; **94**: 1771–6. [Fundoplication improved the clinical outcome of patients with esophagitis complicated by stricture.]

Hands LJ, Papavramidis S, Bishop H, Dennison AR, McIntyre RL, Kettlewell MGW. The natural history of peptic oesophageal strictures treated by dilatation and antireflux surgery alone. *Annals of the Royal College of Surgeons* 1989; **71**: 306–10. [Evaluation of a large experience.]

Karasick S, Lev-Toaff AS. Esophageal strictures: findings on barium radiographs. *American Journal of Roentgenology* 1995; **165**: 561–5. [A review of the imaging features.]

Kikendall JW. Pill-induced esophageal injury. In: Castell DO, Richter JE, eds, *The esophagus*, 3rd edn, pp. 527–37. Lippincott Williams & Wilkins, Philadelphia, 1999. [A comprehensive review.]

Klingler PJ *et al.* Laparoscopic antireflux surgery for the treatment of esophageal strictures refractory to medical therapy. *American Journal of Gastroenterology* 1999; **94**: 632–6. [Antireflux surgery in patients with medically refractory esophageal strictures results in a good clinical outcome with minimal complications.]

McDougall NI, Johnston BT, Kee F, Collins JS, McFarland RJ, Love AH. Natural history of reflux oesophagitis: a 10 year follow up of its effect on patient symptomatology and quality of life. *Gut* 1996; **38**: 481–6. [This recent study found that nearly three-quarters of patients with reflux esophagitis still had significant morbidity related to gastroesophageal reflux disease more than 10 years after diagnosis, but only 2 per cent developed strictures.]

Mercer CD, Hill LD. Surgical management of peptic esophageal stricture: twenty-year experience. *Journal of Thoracic and Cardiovascular Surgery* 1986; **91**: 371–8. [An analysis of factors that are associated with successful surgical treatment in 160 patients with peptic strictures.]

O'Connor JB, Richter JE. Esophageal strictures. In: Castell DO, Richter JE, eds. *The esophagus*, 3rd edn, pp.473–83. Lippincott Williams & Wilkins, Philadelphia, 1999. [A comprehensive review.]

Payne WS. Surgical management of reflux-induced oesophageal stenoses: results in 101 patients. *British Journal of Surgery* 1984; **71**: 971–3. [A large experience from a single surgeon.]

Ritter M *et al.* Treatment of advanced gastroesophageal reflux disease with Collis gastroplasty and Belsey partial fundoplication. *Archives of Surgery* 1998; **133**: 523–9. [The presence of a stricture causing preoperative dysphagia was associated with a poor outcome after a Collis–Belsey operation.]

Smith PM *et al.* A comparison of omeprazole and ranitidine in the prevention of recurrence of benign esophageal stricture. Restore Investigator Group. *Gastroenterology* 1994; **107**: 1312–18. [Omeprazole, 20 mg once daily, was more effective than ranitidine, 150 mg twice daily.]

Spivak H, Farrell TM, Trus TL, Branum GD, Waring JP, Hunter JG. Laparoscopic fundoplication for dysphagia and peptic esophageal stricture. *Journal of Gastrointestinal Surgery* 1998; **2**: 555–60. [Laparoscopic antireflux surgery is effective therapy for symptomatic reflux strictures.]

Zaninotto G, DeMeester TR, Bremner CG, Smyrk TC, Cheng SC. Esophageal function in patients with reflux induced strictures and its relevance to surgical treatment. *Annals of Thoracic Surgery* 1989; **47**: 362–70. [More than 80 per cent of patients with stricture had a mechanically defective sphincter. The indications for antireflux and resectional surgery, and their results, are presented.]

22.3.1 Oesophageal diverticula

John C. Wain

Introduction
Pharyngoesophageal diverticula
Epiphrenic diverticula
Traction diverticula
Further reading

Introduction

Diverticula are acquired, focal pouches of oesophageal wall consisting of mucosa and variably attenuated muscular coat. They may occur anywhere in the esophagus between the pharyngoesophageal junction and the lower esophageal sphincter. Adults are usually affected, the incidence increasing with age.

Two different types of diverticula are recognized on the basis of their pathogenesis. Pulsion diverticula form as localized herniations due to increased luminal pressure in patients with motility disorders or chronic esophageal obstruction, and provide an ‘escape route' for pressure and swallowed food. Traction diverticula develop at the level of the tracheobronchial lymph nodes as a result of granulomatous inflammation with secondary involvement of the adjacent esophagus and subsequent contracture of scar tissue. Diverticula occur predominantly in three anatomic locations. Zenker's or pharyngoesophageal diverticulum is a pulsion defect of the posterior mucosa between the inferior pharyngeal constrictor muscle and the cricopharyngeal sphincter due to simultaneous incoordinate contraction. Mid-esophageal diverticula can be either traction diverticula located adjacent to subcarinal, paratracheal, or hilar lymph nodes or pulsion diverticula with an etiologic esophageal motility disorder. Epiphrenic diverticula are found in the lower esophagus above the diaphragm and are also of the pulsion type. The upper thoracic esophagus is rarely affected by diverticula. Pulsion defects are more likely to interfere with esophageal function and produce symptoms. As a result, symptoms occur regularly in Zenker's diverticula, often in epiphrenic diverticula, and rarely in traction diverticula.

Pharyngoesophageal diverticula

Some 70 to 80 per cent of all esophageal diverticula requiring surgical therapy are Zenker's diverticula. The principal physiologic defect leading to formation of a pharyngoesophageal diverticulum is a loss of co-ordination during the second stage of swallowing. The oblique course of the inferior pharyngeal constrictor muscle andthe horizontal direction of the cricopharyngeal sphincter create an unsupported triangular region devoid of muscle in the posterior wall (Fig. 1). Normally, the cricopharyngeal sphincter relaxes during contraction of the pharyngeal constrictors, allowing propulsion of food. In patients with Zenker's diverticulum the sphincter is closed during the pharyngeal contraction. Although not every swallow may demonstrate this discoordinate pattern, its cumulative effects expose the muscle-free area to high pressures, eventually leading to the formation of a pouch. It is unclear whether this disorder arises from a primary or secondary dysfunction of the cricopharyngeal sphincter. Gastroesophageal reflux, although suspected to cause reflex contraction of this muscle, is not associated with increased cricopharyngeal pressures. Biochemical studies of the cricopharyngeal muscle have shown increased collagen to elastin ratios, suggesting a primary dysfunction as the etiology for the disease.

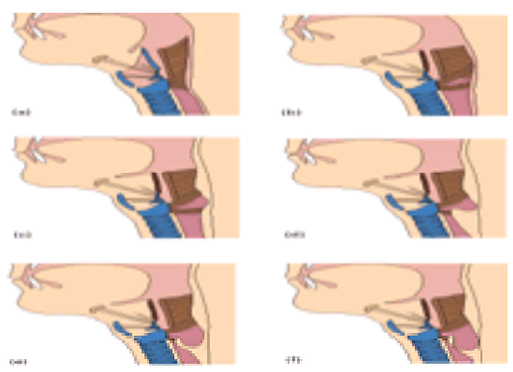


Fig. 1. Development of a pharyngoesophageal diverticulum. (a) and (b) Normal movements of larynx and pharynx during swallowing. (c) to (f) Formation of a pouch. (a) Respiration with descended larynx and relaxation of its sphincters. The cricopharyngeal sphincter closes the esophagus with forward displacement of the posterior pharyngeal wall. The pharyngeal constrictors are relaxed. (b) During deglutition, the larynx is pulled upward and forward to the hyoid. The pharyngeal constrictors begin a peristaltic wave, and opening of the cricopharyngeus allows food to pass into the esophagus. (c) Failure of the cricopharyngeus to relax appropriately delays passage of the food bolus. (d) A small pouch is formed in the bare area and is passively enlarged by food. (e) Further enlargement of the pouch separates the esophagus from the prevertebral fascia and aligns pharynx and diverticulum in a vertical axis, while the esophageal opening is moved to an oblique position. (f) Pressure on the esophagus by the large pouch causes further obstruction.

Symptoms occur early in the course of pouch formation and are progressive. Patients complain of dysphagia and gurgling noises during swallowing. As food is retained in the pouch, fetor oris and spontaneous regurgitation of its undigested contents develop. Respiratory complications are frequent, including aspiration, asthma, pneumonia, and occasionally, lung abscess. Pressure of the pouch on the recurrent laryngeal nerve may cause hoarseness. As the pouch increases in size, a soft mass may be palpated in the neck. Neglect of symptoms may result in weight loss, cachexia, and respiratory compromise. Perforation is rare and usually due to iatrogenic injury. One-third of patients have associated esophageal functional disorders, such as hiatal hernia, diverticula elsewhere in the esophagus, achalasia, or diffuse spasm. Nevertheless, radiographic contrast studies (Fig. 2) are usually sufficient to arrive at the diagnosis and to delineate the rest of the esophagus; motility studies are generally not required. Esophagoscopy is indicated when there is complete obstruction or if malignancy is suspected, since cancers occasionally occur in the pouch.

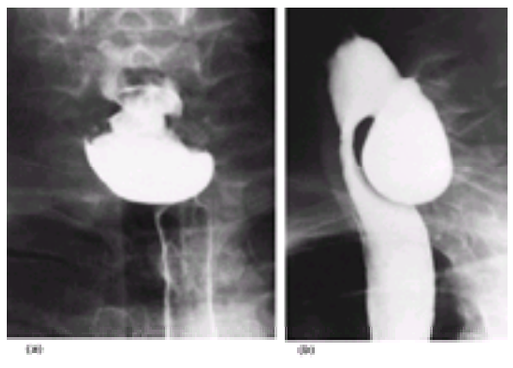


Fig. 2. Barium swallow in a patient with Zenker's diverticulum. Sagittal (a) and lateral (b) view of a large pharyngoesophageal pouch.

Zenker's diverticulum requires treatment regardless of diverticulum size. Symptoms do not improve with other forms of therapy, and the hazards of regurgitation and aspiration increase with pouch enlargement. Myotomy and diverticulectomy are the two elements of successful surgical therapy. If the pouch is small, cricopharyngeal myotomy alone relieves symptoms. The cricopharyngeal muscle is divided vertically in the posterior midline, dissecting the mucosa over half the circumference to prevent recurrent obstruction. If the pouch is well developed, diverticulectomy is performed during the same operation to remove the cause of regurgitation (Fig. 3). Care must be exercised to avoid overexcision of esophageal mucosa and creation of a defect too large to close without tension or stenosis. Some surgeons prefer diverticulopexy with fixation to the prevertebral fascia to obliterate the lumen of the sac.

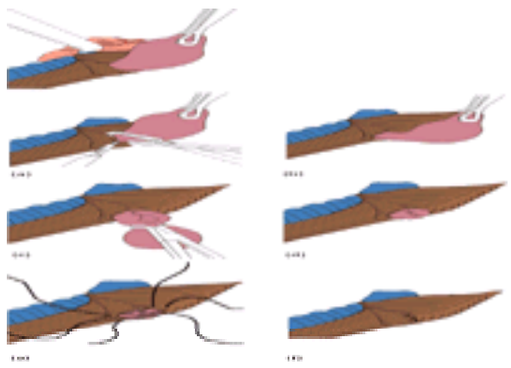


Fig. 3. Technique of cricopharyngeal myotomy and diverticulectomy. The esophagus is exposed through an incision anterior to the sternocleidomastoid muscle. (a) and (b) The myotomy is performed in the posterior midline. (c) and (d) The diverticulum is clamped and excised, closing the esophagus transversely with inverted, interrupted sutures. (e) and (f) The esophageal closure is protected with transverse closure of the muscularis. If the sac is small and only a myotomy is performed, the muscle layer is left open.

Cricopharyngeal myotomy, with diverticulectomy or diverticulopexy, produces excellent results. In a follow-up study from the Mayo Clinic, over 90 per cent of patients were asymptomatic. Operative mortality was 1.2 per cent. Complications, including recurrent laryngeal nerve palsy, usually transient, and an esophagocutaneous fistula, which closed spontaneously with adequate drainage, affected 5 per cent of patients.

Epiphrenic diverticula

These are relatively uncommon, representing 10 to 20 per cent of all esophageal diverticula. The male to female ratio is 2:1. Epiphrenic diverticula occur usually in the lower esophagus within 10 cm of the diaphragm and project as solitary lesions into the right chest; left-sided or multiple lesions may be encountered. The resulting pouch is from 2 to over 10 cm in diameter and often has a narrow neck. Most epiphrenic diverticula are associated with functional motor disturbances, and some cause organic obstruction. In rare cases, carcinoma can arise within a diverticulum. Symptoms due to the diverticulum, however, develop in fewer than half of all patients. Patients present with dysphagia, regurgitation, and epigastric or chest pain. Occasionally, examination is prompted by aspiration with choking or coughing. Because these complaints are frequently associated with other benign esophageal disorders, symptomatic patients should undergo complete evaluation including barium swallow (Fig. 4), motility studies, and endoscopy.

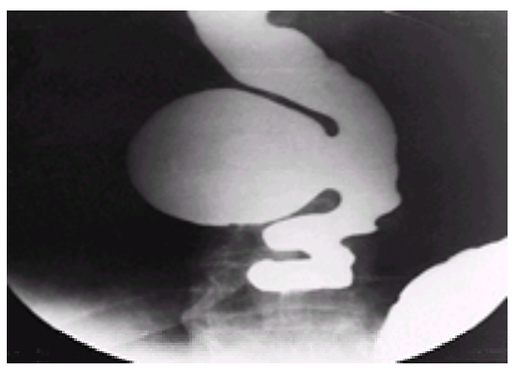


Fig. 4. Barium swallow in a patient with an epiphrenic diverticulum.

Appropriate therapy in patients with symptoms consists of surgical resection of the diverticulum and alleviation of underlying motor or organic disorders (Fig. 5 and Fig. 6). Access to the esophagus is provided by posterolateral thoracotomy, usually on the left. The diverticulum is exposed and resected, closing the defect in the esophagus. Distal obstruction should be excluded by preoperative endoscopy because it jeopardizes the closure of the esophagus and may result in recurrence of the diverticulum. In most cases, an esophageal myotomy extending from the site of the diverticulum distally to the lower esophageal sphincter is required. If motility studies demonstrated motor dysfunction due to diffuse spasm, a long myotomy extending from the lower esophageal sphincter to the aortic arch is performed. Some authors suggest that a long myotomy should be performed routinely. Any associated hiatal hernia is repaired, albeit cautiously because of the risk of recurrent obstruction.

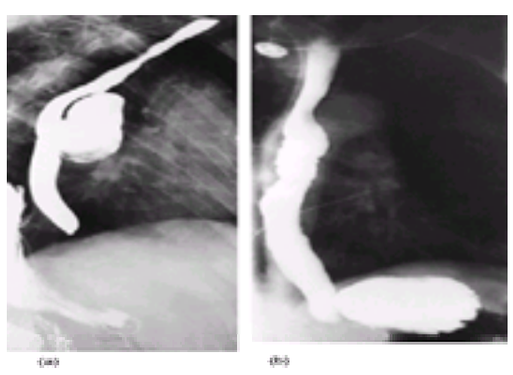


Fig. 5. Mid-esophageal pulsion diverticulum. (a) Preoperative view.(b) Postoperative view after excision of diverticulum and distal esophageal myotomy.

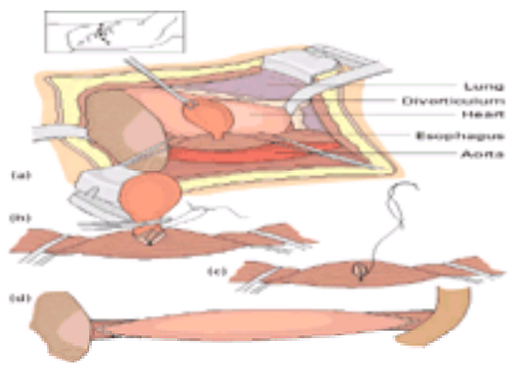


Fig. 6. Technique for excision of an epiphrenic diverticulum. (a) Surgical approach via left thoracotomy. The esophagus is turned around its axis to expose the right-sided diverticulum. (b) The diverticulum is clamped and excised, closing the mucosa with inverted, interrupted sutures. (c) Closure of the muscularis. (d) Long myotomy opposite to the esophageal closure.

Although pulsion diverticula in the mid-esophageal or epiphrenic region can be treated using thoracoscopic or laparoscopic methods, the complexity of the requisite myotomy suggests that it is best performed as an open procedure. The individualized operative approach to each of these conditions emphasizes the need for a thorough preoperative understanding of the pathophysiology based on motility and other functional studies. Postoperatively, the esophagus remains decompressed with a nasogastric tube until a Gastrografin study after 1 week has confirmed the absence of a leak.

Traction diverticula

These diverticula occur as a result of granulomatous necrosis of tracheobronchial lymph nodes due to tuberculosis or histoplasmosis. The esophagus is involved because of its proximity to the inflammation. Fixation of a transmural segment of esophageal wall, followed by cicatricial contraction of the involved nodes and peristaltic extension of the surrounding esophagus results in the formation of a true diverticulum. The sac is typically located on the left side, has a broad neck, and points upward. For this reason, it is rarely filled with food particles and rarely causes symptoms. Fistulization to the airway and hemorrhage are rare complications. Bleeding into the esophagus originates most often from erosion of small bronchial or esophageal vessels, but communication with the superior vena cava and exsanguination have been reported. Symptomatic patients should undergo local excision of the diverticulum with closure of the esophagus and separation from the tracheobronchial tree with a vascularized tissue flap of intercostal muscle or pericardial fat.

Further reading

Allen TH, Clagett OT. Changing concepts in the surgical treatment of pulsion diverticula of the lower oesophagus. *Journal of Thoracic and Cardiovascular Surgery* 1965; **50**: 455–62. [The first paper to define the importance of distal myotomy in managing epiphrenic diverticula.]

Altorki NK, Sunagaawa M, Skinner DB. Thoracic esophageal diverticula. Why is operation necessary? *Journal of Thoracic and Cardiovascular Surgery* 1993; **105**: 260–4. [Another view of management supporting surgical resection of all patients.]

Benacci JC, Deschamps C, Trastek VF, Allen MS, Daly RC, Pailorlero PC. Epiphrenic diverticulum: results of surgical treatment. *Annals of Thoracic Surgery* 1993; **55**: 1109–13. [A recent series from the Mayo Clinic emphasizing individualized treatment of patients.]

Ellis FH, Schlegel JF, Lynch VP, Payne WS. Cricopharyngeal myotomy for pharyngo-oesophageal diverticulum. *Annals of Surgery* 1969; **170**: 340–9. [The classic description of the technique of myotomy.]

Ferraro P, Duranceau A. Esophageal diverticula. *Chest Surgery Clinics of North America* 1994; **4**: 741–67. [An excellent overview of the pathophysiology and clinical aspects of esophageal diverticula.]

Huang BS, Payne WS, Cameron AJ. Surgical management for recurrent pharyngoesophageal (Zenker's) diverticulum. *Annals of Thoracic Surgery* 1984; **37**: 189–91. [A review of the results of surgical therapy.]

Orringer MB. Epiphrenic diverticula: fact and fable. *Annals of Thoracic Surgery* 1993; **55**: 1067–8. [An editorial accompanying the previous paper that outlines the author's rationale for a long esophageal myotomy in all patients.]

Rosati R, Fumagalli U, Bona S, Bonavina L, Peracchia A. Diverticulectomy, myotomy and fundoplication through laparoscopy. A new option to treat epiphrenic esophageal diverticula? *Annals of Surgery* 1998; **227**: 174–8. [A description of laparoscopic diverticulectomy.]

Schima W *et al.* Association of mid-oesophageal diverticula with oesophageal motor disorders. Videofluoroscopy and manometry. *Acta Radiologica* 1997; **38**: 108–14. [An important demonstration that pulsion diverticula can also occur in the mid-esophagus.]

Schultz SC, Byrne DM, De Cunzo P, Byrne WB. Carcinoma arising within epiphrenic diverticula. A report of two cases and review of the literature. *Journal of Cardiovascular Surgery* 1996; **37**: 649–51. [A review of the literature regarding this rare topic.]

Venturi M *et al.* Biochemical markers of upper esophageal sphincter compliance in patients with Zenker's diverticulum. *Journal of Surgical Research* 1997; **70**: 46–8. [Biochemical evidence that the defect in Zenker's may be primary to the cricopharyngeus muscle.]

22.3.2 Achalasia

John C. Wain

Introduction
Epidemiology
Pathogenesis
Evaluation
Differential diagnosis
Treatment
Pharmacologic therapy
Pneumatic dilatation
Esophagomyotomy
Sigmoid megaesophagus
Long-term follow-up
Further reading

Introduction

Achalasia is a motility disorder of the esophagus characterized by failure of the lower esophageal sphincter to relax in response to swallowing and by an absence of peristalsis in the esophageal body. The disease is most often progressive, leading to gradual dilatation of the esophagus above the sphincter. Patients typically present with dysphagia, regurgitation, and substernal chest pain. Disruption of the lower esophageal sphincter relieves the obstruction and effectively restores swallowing.

Thomas Willis in 1674 treated a man with perpetual vomiting and provided the earliest available description of the disease. Willis fashioned a sponged-tipped rod made of whalebone that the patient used to force open his cardia after meals for the following 15 years. Russell first reported treatment of the constricted sphincter by inflation of a silk-covered rubber balloon in 1887. Von Mikulicz introduced the term cardiospasm in 1904 and described dilatation of the gastroesophageal junction through a gastrostomy. In 1913 Heller performed surgical division of the muscular narrowing by means of two myotomies on the anterior and posterior surface of the lower esophagus, through an abdominal incision. Plummer developed a successful method of forcefully dilating the sphincter with hydrostatic pressure using a rubber bag passed over a thread; he reported his personal experience with 301 cases in 1921. Modifications of the procedures introduced by Heller and Plummer are the mainstay of effective therapy today.

Epidemiology

The incidence rate of radiologically confirmed cases of achalasia in population studies from Great Britain and the United States averages between 0.4 and 0.6 cases/10^5 population per year, although peak rates as high as 2 cases/10^5 have been observed. The disease is known in most races and on all continents. In Europe and the United States, men and women are affected in equal proportions, although some series show a slight male predominance. Most patients present between the third and fifth decade of life, and less than 5 per cent of patients have the infantile type of disease. In South America, particularly in Brazil, megaesophagus is most frequently a manifestation of Chagas' disease, caused by *Trypanosoma cruzi* infection. Chagas' disease resembles achalasia because of the diffuse ganglion cell involvement and similar abnormalities of esophageal motility. Patients with Chagas' disease, however, develop disease of multiple organs, including cardiomyopathy, megacolon, and megaureter. It can be diagnosed by specific serologica tests.

Pathogenesis

The primary abnormality causing chronic constriction of the lower esophageal sphincter and loss of peristalsis remains unknown. Potential etiologies are a primary disease of the esophageal wall, a disorder affecting the peripheral or central autonomic motor neurons, and autoimmune or inflammatory disease. Various experimental procedures, including selective ischemic or toxic injury to the esophagus and bilateral cervical vagotomy, result in conditions similar, but not identical, to achalasia.

The earliest documented histologic finding in achalasia is degeneration of ganglion cells in the myenteric plexus of Auerbach. A reduction in number or absence of ganglion cells, predominantly in the esophageal body but also in the distal narrowed esophageal segment, is noted. These changes are generally more pronounced in advanced disease. Recent studies have found a selective impairment of postganglionic inhibitory neurons in the circular layer of the lower esophageal sphincter. As a result, cholinergic agonists produce an accentuated tightening and the lower esophageal sphincter relaxant cholecystokinin produces a paradoxical contraction of the sphincter. Neural damage occurs at two further levels. Abnormalities similar to wallerian degeneration occur in both myelinated and non-myelinated fibers of the vagus nerve. Nerve cells in the dorsal motor nucleus of the vagus nerve in the brainstem are diminished in number or absent.

Ultrastructural studies of esophageal muscle fail to demonstrate a specific injury in achalasia. In advanced disease, gross observation of the esophagus demonstrates a 1.5 to 4.5 cm distal narrowing with normal wall thickness. The wall of the esophageal body is thickened due to an increase of the circular muscle layer. The lumen is enlarged, elongated, and tortuous. Diverticula or pseudodiverticula of the mid-esophagus occur occasionally. The mucosa demonstrates typical lesions of chronic stasis, including esophagitis and ulceration.

Evaluation

A gradual onset of symptoms over months to years is characteristic of achalasia. Over 90 per cent of all patients have symptoms for longer than 1 year, and half are symptomatic for more than 3 years. Some may have symptoms for several decades. In addition to obstruction of swallowing, patients complain of regurgitation or vomiting of undigested food. Ellis and Olsen described how a typical patient copes with his eating problem.

> He shuns company while eating and is unwilling to dine in public. He learns quickly which foods will pass most easily and hence often avoids cold foods or solids such as meats or apples. He knows that he must not eat before retiring for fear of regurgitation, choking, and coughing. Some patients with advanced achalasia develop techniques for forcing food through the cardia. Merely drinking large quantities of water will sometimes overcome the resistance at the cardia […]. Others learn to execute a Valsalva maneuver thus increasing intrathoracic pressure until the lower sphincter gives way. In the process the patient usually stands in a crouched position, often holding on to a mantelpiece or stair rail.

Weight loss is present in 60 per cent of patients. Substernal chest or epigastric pain occurs in one-third of all patients. Nocturnal regurgitation may lead to aspiration, and pneumonia is present in 10 per cent of patients. In infants, respiratory problems may be the first symptom to prompt evaluation and diagnosis.

Typical findings on plain chest radiograph include widening of the mediastinum and a posterior mediastinal air–fluid level. Chronic aspiration may cause abnormalities of the lung fields. Barium swallow demonstrates the characteristic dilatation of the esophagus with a narrowing at the gastroesophageal junction (bird's beak deformity) in over 85 per cent of patients (Fig. 1). The size of the esophageal body on contrast studies provides an approximation of disease severity: a diameter of less than 4 cm is termed mild disease, 4 to 6 cm moderate, and greater than 6 cm severe (‘megaesophagus'). Further indicators of disease severity are the amount of retained food and degree of peristalsis observed on fluoroscopy.

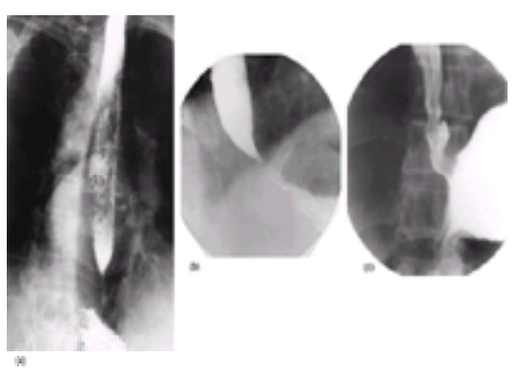


Fig. 1. Barium swallow in a patient with achalasia. (a) Retained food, moderate esophageal dilatation, and bird's beak deformity of distal esophagus. (b) The

gastroesophageal junction before esophagomyotomy. (c) Same region after esophagomyotomy.

Esophageal manometry confirms the clinical diagnosis (Fig. 2). Four criteria are characteristic of achalasia: elevated lower esophageal sphincter pressure, incomplete or absent relaxation of the lower esophageal sphincter in response to swallowing, absence of peristalsis in the esophageal body, and elevated intraesophageal pressure. Absence of co-ordinated peristalsis is required for diagnosis, but weak simultaneous contractions are often observed early in the course of the disease. A subgroup of patients have ‘vigorous achalasia', a hyperdynamic discoordinate motility pattern with simultaneous, high-amplitude contractions of the esophageal body exceeding 40 mmHg. Esophageal diverticula may be found in 5 per cent of patients with achalasia.

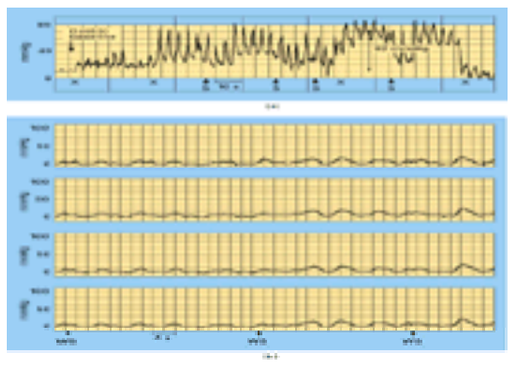


Fig. 2. Esophageal motility tracing of a patient with achalasia. (a) The catheter is moved through the maximum pressure zone. The lower esophageal sphincter is hypertensive with incomplete relaxation during swallowing. x, catheter moved up 0.5 cm; S, swallow. (b) The esophageal body demonstrates total absence of peristalsis. Contractions are simultaneous, of low amplitude, and resemble respiratory artifact. WS, wet swallow.

A complete endoscopic examination of esophagus and stomach should be part of the initial evaluation, to detect disease complications, such as ulceration, and to rule out carcinoma. The esophagus is often patulous and contains food particles and secretions. The mucosa exposed to these contents is frequently inflamed. The lower sphincter is closed, but provides little or no resistance to the endoscope, which is easily advanced into the stomach.

Differential diagnosis

Two conditions mimic the radiologic and manometric features of achalasia. Chagas' disease occurs uncommonly outside South America and generally also affects other organs. Pseudoachalasia is the rare presentation of a carcinoma at the gastroesophageal junction with gradually worsening dysphagia, narrowing at the lower sphincter, and manometric findings characteristic of achalasia. Patients typically have symptoms for less than 1 year, and weight loss predominates. Inhalation of amyl nitrate leads to sphincter relaxation in achalasia, but not in pseudoachalasia. Endoscopic examination and biopsy make the diagnosis of pseudoachalasia. Intraluminal findings are subtle, but endoscopic ultrasonography may be helpful in delineating an intramural neoplasm.

Diffuse esophageal spasm may present with symptoms similar to those of early achalasia. Manometric studies of diffuse spasm, however, demonstrate segmental contraction of the esophagus and relaxation of the lower esophageal sphincter with swallowing. A peptic stricture that occurs as a complication of scleroderma may imitate achalasia on contrast studies, but systemic manifestations of scleroderma usually indicate the diagnosis.

Treatment

Successful therapy consists of disruption of the lower esophageal sphincter by forceful dilatation or esophagomyotomy to allow gravity drainage of food while avoiding chronic gastroesophageal reflux. Therapy does not restore the ormal anatomy of the sphincter or co-ordinated peristalsis of the esophageal body but does relieve dysphagia and regurgitation in most patients.

Pharmacologic therapy

Calcium channel blockers and nitrates decrease lower esophageal sphincter pressure transiently. Transient improvement of swallowing is noted in about 50 per cent of patients with mild and moderate symptoms. However, these drugs do not produce lasting relief. Botulinum toxin, type A, can be injected transmucosally into the lower esophageal sphincter. Botulinum toxin is effective in 85 per cent of patients, more commonly in patients older than 60 years and those with vigorous achalasia. By 6 months 50 per cent will have recurrence of symptoms. Pharmacologic therapy should therefore be used for extended periods only in patients unable to undergo standard treatment.

Pneumatic dilatation

Pneumatic dilatation is the most effective non-surgical therapy for achalasia. The dilator is an expandable polyethylene balloon that is positioned endoscopically, over a guide wire or under radiographic control. Initial endoscopic evaluation to rule out cancer or other gastroesophageal pathology is a prerequisite. The dilator is placed through the narrowed segment and the balloon is rapidly inflated with air to obliterate the waist of the segment for 60 s. A pressure of 7 to 10 psi is required. If the result is unsatisfactory, repeated attempts or a larger balloon are required. The patient can leave the hospital on the same day. If the first treatment produces no improvement, another dilatation may be considered after 2 months.

Approximately 70 per cent of patients improve symptomatically after the procedure, although 30 per cent of patients with symptom relief may have objective evidence of persistent obstruction. Advantages of pneumatic dilatation are the avoidance of general anesthesia and an incision. Treatment cost is lower. Further dilatations are required in approximately 15 per cent of patients; a myotomy is required in 8 per cent. Disadvantages of pneumatic dilatation include a 20 per cent incidence of gastroesophageal reflux and a 2 to 10 per cent risk of esophageal perforation. Treatment of acute perforation consists of thoracotomy, primary repair of the tear, and myotomy on the contralateral side of the esophagus.

Esophagomyotomy

Esophagomyotomy for achalasia is derived from the duplicate myotomy performed by an abdominal approach initially described by Heller in 1914. However, many modifications of esophagomyotomy have evolved with the aim of reducing operative morbidity and postoperative reflux and recurrence. Good to excellent results are found in 83 per cent of patients, with an operative mortality of less than 1 per cent.

Thoracic or abdominal access

Thoracotomy offers direct access to the lower sphincter without obligatory injury to the diaphragmatic attachments and risk of reflux (Fig. 3). A routine fundoplication is not necessary. The incidence of reflux is 10 per cent after a transthoracic esophagomyotomy. A trans-abdominal approach requires division of the phrenoesophageal ligament to expose the lower esophagus; a routine loose Nissen fundoplication is therefore necessary. The incidence of reflux is 20 per cent with this method.

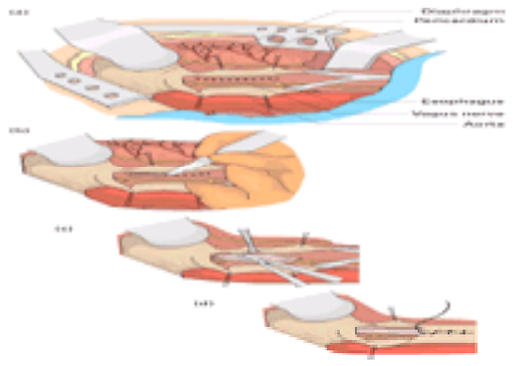


Fig. 3. Esophagomyotomy via left thoracotomy. (a) Operative exposure. Dotted line indicates extent of myotomy. (b) Incision of the muscularis. (c) Separation of mucosa and muscularis over half the esophageal circumference. (d) Suture narrowing of the hiatus in the presence of a hernia.

Minimally invasive approaches, laparoscopy or thoracoscopy, for esophagomyotomy are associated with a reduced length of hospital stay. Thoracoscopic procedures are associated with an 85 per cent rate of symptomatic improvement, but a 50 per cent rate of reflux. Laparoscopic myotomy, performed in conjunction with a Toupet or a Dor fundoplication, is associated with a 94 per cent rate of symptomatic improvement and an 11 per cent rate of reflux. The Toupet fundoplication is preferred unless the esophagus is markedly dilated.

The extent of myotomy

The longitudinal incision into the muscle must divide all circular fibers on the lower esophagus and reach proximally to the dilated esophagus. Any remaining constriction will lead to persistent dysphagia. Dissection of the muscle from the submucosa assures complete division of the sphincter muscle. When performed via a thoracotomy, extending the myotomy too far on to the gastric wall can disrupt the hiatal mechanism and result in reflux. The myotomy should continue for less than 1 cm on to the stomach, where the presence of transverse cardiac veins is first noted. Exact execution of this step is the most critical determinant of postoperative results with the thoracotomy approach.

Addition of antireflux procedure

While those who perform only a thoracic myotomy restrict a fundoplication to specific indications, for instance the presence of a hernia, others consider it routine. They argue that a widely dissected hiatus provides improved exposure and believe that the incidence of postoperative reflux is reduced. However, the rate of failure after fundoplication is 10 to 15 per cent, which equals the overall incidence of postoperative reflux after a thoracic myotomy and raises doubt regarding the benefit of adding this procedure in all patients. If reconstruction of the hiatus is required, the lack of esophageal peristalsis often results in recurrent obstructive symptoms that require esophageal dilatation and long-term observation. Total fundoplication is therefore avoided, and ‘loose' Belsey or Nissen repairs are preferred. As noted, when a trans-abdominal or laparoscopic approach is used, fundoplication is routine.

Sigmoid megaesophagus

Patients presenting with extensive dilatation of the esophageal body or unrelenting post-treatment reflux may respond poorly to conventional therapy. Esophagectomy and replacement with the stomach or an intestinal interposition should be considered to alleviate symptoms.

Long-term follow-up

In a cumulative review of studies comparing pneumatic dilation and esophagomyotomy, the success rates were 67 per cent for pneumatic dilation and 83 per cent for myotomy with a mean follow-up of 4 years. In the only prospective comparison study of the two methods, no difference in efficacy was found with a follow-up of 2 years. Patients who are low surgical risk can therefore have either dilatation or myotomy, while those who cannot tolerate surgery are candidates for botulinum toxin therapy. Overall, laparoscopic myotomy and fundoplication appears to produce the best symptomatic relief with the least risk of postoperative reflux disease.

Patients with achalasia have an increased risk of dying from squamous cell esophageal cancer. Cancer usually develops in patients with advanced achalasia. Successful palliation of dysphagia does not prevent malignancy. Treatment during the early stage of the disease may thus afford a relative protection from esophageal cancer. Most frequently located in the middle-third of the esophagus, tumors develop in 4 per cent of patients, typically many years after diagnosis or treatment. Routine endoscopic examination with brush cytology or biopsy should continue indefinitely, even after effective palliation.

Further reading

Abid S, Champion G, Richter JE, McElvein R, Slaughter RL, Koehler RE. Treatment of achalasia: the best of both worlds. *American Journal of Surgery* 1994; **89**: 979–85. [A prospective study of modern pneumatic dilation versus achalasia.]

Csendes A, Braghetto I, Henriquez A, Cortes C. Late results of a prospective randomised study comparing forceful dilatation and oesophagomyotomy in patients with achalasia. *Gut* 1989; **30**: 299–304. [A long-term study of pneumatic dilation versus achalasia.]

Ellis FH, Olsen AM. *Achalasia of the esophagus*. WB Saunders, Philadelphia, 1969. [A classic treatise on the disease, its pathophysiology, and its treatment.]

Ellis FH, Crozier RE, Watkins E. Operation for esophageal achalasia. Results of esophagomyotomy without an antireflux operation. *Journal of Thoracic and Cardiovascular Surgery* 1984; **88**: 344–51. [A large series of transthoracic myotomy without fundoplication.]

Hunter JG, Richardson WS. Surgical management of achalasia. *Surgical Clinics of North America* 1997; **77**: 993–1015. [An excellent review with superb illustrations of the laparoscopic technique of esophagomyotomy.]

Kahrilas PJ, Kisk SM, Helm JF, Dodds WJ, Harig JM, Hogan WJ. Comparison of pseudoachalasia and achalasia. *American Journal of Medicine* 1987; **82**: 439–46. [A concise summary of the characteristics of each entity.]

Little AG, Soriano A, Ferguson MK, Winans CS, Skinner DB. Surgical treatment of achalasia: results of esophagomyotomy and Belsey repair. *Annals of Thoracic Surgery* 1988; **45**: 489–94. [A report from a group favoring routine fundoplication with transthoracic myotomy.]

Okike N, Payne WS, Neufeld DM, Bernatz PE, Pairolero PC, Sanderson DR. Esophagomyotomy versus forceful dilation for achalasia of the esophagus: results in 899 patients. *Annals of Thoracic Surgery* 1979; **28**: 119–25. [The classic comparative report from the Mayo Clinic suggesting a long-term benefit for esophagomyotomy.]

Ott DJ, Hodge RG, Chen MYM, Wu WC, Gelfand DW. Achalasia associated with esophageal diverticula: prevalence and potential implications. *Journalof Clinical Gastroenterology* 1994; **18**: 343–6. [A review of the finding of diverticula in 120 patients with achalasia, emphasizing the risk of perforation in these patients during dilation.]

Shimi S, Nathanson LK, Cuschieri A. Laparoscopic cardiomyotomy for achalasia. *Journal of the Royal College of Surgeons of Edinburgh* 1991; **36**: 152–4. [One of the first reports of successful laparoscopic myotomy.]

Spiess AE, Kahrilas PJ. Treating achalasia: from whalebone to laparoscope. *Journal of the American Medical Association* 1998; **280**: 630–42. [A review of the results of therapy from the English literature from 1966 to 1997.]

Strietz JM Jr, Ellis FH Jr, Gibb SP, Heatley GM. Achalasia and squamous cell carcinoma of the esophagus: analysis of 241 patients. *Annals of Thoracic Surgery* 1995; **59**: 1604–9. [One of the largest reports in the literature on the subject.]

Vaezi MG, Richter JE. Current therapies for achalasia: comparison and efficacy. *Journal of Clinical Gastroenterology* 1998; **27**: 21–35. [An excellent discussion of techniques for pneumatic dilation and comparison of published results for all methods of treating achalasia.]

22.4 Oesophageal perforation, Boerhaave's syndrome, and Mallory–Weiss syndrome

Joseph I. Miller, Jr.

Introduction
Surgical anatomy
Etiology of esophageal perforation
Boerhaave's syndrome
Mallory–Weiss syndrome
Diagnosis of esophageal perforation
Treatment of esophageal perforation
Conservative management
Operative therapy for cervical esophageal perforations
Operative therapy of thoracic esophageal perforations
Surgical therapy of Mallory–Weiss syndrome
Further reading

Introduction

The esophagus is located in three strategically located anatomic regions: the neck, the thorax, and the abdomen. Because of its divergent location and lack of serosal covering, it is an unforgiving organ whether it be due to disease, iatrogenic injury, or spontaneous emetic perforation or bleeding.

Esophageal perforation can be caused by external trauma or any instrument, device, or foreign body reaching the esophagus. Postemetic lacerations or rupture of the lower esophagus can also occur, all presenting similar pictures. Since the original description of Boerhaave, the term ‘spontaneous rupture of the esophagus' has been used routinely to include all perforations involving the full thickness to the esophageal wall and associated with emesis. Mallory and Weiss described another syndrome of acute postemetic laceration of the gastric cardia as a source of major hemorrhage following ethanolic intoxication with emesis.

The diagnosis and management of any of these conditions is determined by the cause, physiologic status, and medical condition of the patient and the duration of symptoms (i.e. early or late). For purposes of definition, ‘early' injuries to the esophagus are those caused within 24 h and ‘late' injuries are those presenting to the physician more than 24 h after the onset of symptoms or diagnosis. Injuries of the esophagus present some of the most complex management problems in thoracic disease.

Surgical anatomy

The esophagus extends from the oropharynx in the neck to the esophagogastric junction in the abdomen. Its blood supply is from the inferior thyroid arteries and short esophageal branches off the aorta. The vagi provide motor innervation of the esophagus. From the surgical standpoint, there are four anatomic landmark levels that establish the various anatomic regions of the esophagus and frequently correspond to levels at which injury occurs to the esophagus. These levels are as depicted in Fig. 1 and Table 1. The anatomic level of the cricopharyngious muscle at 15 cm is the most frequent site of iatrogenic perforation and the second most frequent site of iatrogenic perforation is in the lower one-third. The most frequent site of Boerhaave's syndrome and Mallory–Weiss syndrome is in the lower one-third of the esophagus just above the esophagogastric junction.



Fig. 1. Anatomic regions of the esophagus.

Anatomic level	*Distance from incision*
Cervical esophagus	0–20cm
Upper 1/3 of thoracic esophagus	20–25cm
Middle 1/3 of thoracic esophagus	25–35cm
Lower 1/3 of thoracic esophagus	35–40cm

Table 1 Anatomic regions of the esophagus

Etiology of esophageal perforation

The numerous etiologies of esophageal perforation are as listed in Table 2. A review of 511 perforations by Jones revealed the following frequency of causes. Causes of esophageal perforation include: (a) endoscopic instrumentation; (b) periesophageal surgery; (c) trauma; (d) barotrauma; (e) tumors; and (f) infection.

I.	Endoscopic instrumentation • esophagoscopy • Bougienage • pneumatic dilatation • sclerosis of esophageal varices • endoscopic prosthesis • traumatic intubation
II.	Periesophageal surgery
III.	Trauma • penetrating • foreign body • caustic ingestion
IV.	Barotrauma
V.	Tumors
VI.	Infection

[a] Modified from Bell (1995).

Table 2 Causes of esophageal perforation

Endoscopic instrumentation is by far the most frequent cause of esophageal perforation. The most frequent site of instrumental perforation is at the cricopharyngeus muscle. In a review in 1988, the perforation rate was 0.11 per cent of 35 412 endoscopies. The second most frequent site of perforation is in the lower third just above the esophagogastric junction. This occurs most frequently just anterior to the left vagus nerve. Perforation can also occur with esophageal dilatation for strictures or tumor, but occurs less frequently when done over a guidewire under fluoroscopy. Pneumatic dilatation carries a risk of about 5 per cent incidence of esophageal perforation.

Perforation may also be caused by surgery in the peri-esophageal area. It has been most frequently association with mediastinoscopy, pulmonary resections, particularly the right lower lobe, anterior spinal surgery, and radical lymphadnectomy. It can also go undetected at the time of antireflux surgery.

Perforation may likewise result from penetrating trauma, foreign body, or ingestion of a caustic substance. It may also be due to baratrauma which is most frequently postemetic. Tumors of the lung and esophagus can likewise cause perforation by erosion through the esophageal wall..

Boerhaave's syndrome

On October 29, 1723, Herman Boerhaave was summoned to the home of Baron Jan von Wassenerner, Grand Admiral of Holland's fleet. Following a glutinous feast of beer and duck, the Admiral complained of abdominal discomfort and vomited quite violently. He gave a horrifying cry and stated he felt if something had been torn. He complained with severe left upper gastric and thoracic discomfort. Death occurred 18 h later. At autopsy, Boerhaave noted emphysema of the abdominal and thoracic wall and a tear of the posterior lateral wall of the esophagus, 3 in above the diaphragm. He observed particles of duck, beer, and olive oil in the left pleural space. His 70-page autopsy protocol titled *History of a grievous disease not previously described* published in 1724 won him admiration; hence, the eponym Boerhaave's syndrome has been applied to postemetic rupture of the normal esophagus (Boerhaave, 1955). Barrett performed the first successful repair in 1947.

All cases of barotrauma resulting in esophageal rupture may be considered as variants of Boerhaave's syndrome as the pathology and clinical course are similar.

Mallory–Weiss syndrome

Mallory–Weiss syndrome is similar to Boerhaave's in that it frequently follows an alcoholic emetic binge. It results in laceration of the vessels in the region of the gastroesophageal junction, thus, causing gastrointestinal bleeding.

Diagnosis of esophageal perforation

A high index of suspicion is the key to early diagnosis of esophageal perforation. A history of recent esophageal instrumentation or emesis should alert the physician to the potential of an esophageal injury. When injury occurs in the neck, there is the onset of pain, fever, dysphagia, crepitus and subcutaneous emphysema within a short period of time. When the perforation occurs intrathoracically the symptoms are more insidious. There is generally left pleuritic chest discomfort and upper gastric tenderness. Fever and a general feeling of not being well ensue. The patient may develop marked abdominal pain and tenderness. When left undiagnosed, shock and death can occur within 48 h, if massive contamination is present. Thoracic esophageal perforation can also simulate an acute myocardial infarction, thus causing further delay.

Once the diagnosis is suspected a gastrogaffin swallow should immediately be performed. If negative, it should be followed by a thin barium study. If either of these is positive (i.e. a leak demonstrated), then appropriate therapeutic intervention should be taken. Computed tomography of the mediastinum after contrast can occasionally be useful in detecting very small leaks in the thorax. A lateral view of the neck may be helpful in demonstrating air in cervical esophageal perforation; air may be seen in the pre-vertebral fascial planes. The computed tomography scan may likewise demonstrate an abscess. A chest radiograph may show developing fluid and air–fluid levels within the pleural space.

Endoscopy may be performed, particularly if a foreign body is suspected as the cause or if the patient has hemetemesis. Once the diagnosis has been confirmed, a definitive course of action and intervention should be planned.

Treatment of esophageal perforation

Once a diagnosis has been made, a definite course of action should be undertaken. Conservative management may be appropriate in some cases of esophageal perforation, but most frequently surgical intervention will be required.

Conservative management

Conservative management may applicable in two clinical situations. The first is with a small cervical perforation without evidence of systemic sepsis or marked clinical symptoms. The second situation is when there is a small contained leak in the thoracic esophagus that does not rupture into the free pleural space and re-enters into the esophagus. The incidence of contained intrathoracic leak is extremely rare. The author has seen only two such cases while operating on approximately 35 cases during a 25-year practice experience. In general, a thoracic leak should be treated with an aggressive open surgical approach.

Principles of conservative theory are: (a) nothing by mouth; (b) broad-spectrum antibiotics to cover aerobic and anaerobic organisms; and (c) intravenous hydration and hyperalimentation combined with nasogastric tube suction. At the earliest indication that conservative therapy is not working, surgical intervention should be performed.

Operative therapy for cervical esophageal perforations

Operative therapy for cervical esophageal perforations is simple, effective, and associated with a low morbidity and mortality. A brief description of our operative technique is as follows: an incision is made parallel to the anterior border of the sternocleidomastoid muscle. The carotid sheath and its contents are retracted laterally and the thyroid is retracted medially. It should be noted that only finger retraction is used on the thyroid in order to avoid traction injury to the recurrent laryngeal nerve. The pre-vertebral fascial plane is entered and the presence of pus, barium, or esophageal leak is identified. It is mandatory that the above be identified in order to ensure that one is in the correct anatomic plane. If less than 24 h old, it may be possible to close the leak directly if it can be identified. Our choice of suture would be with interrupted 3-O silk or prolene with the knots tied inside. If not easily identified, one should place drains inferiorly into the mediastinum and superiorly along the postvertebral gutter. Most cervical leaks will close in 2 to 3 weeks when the patient has nothing by mouth, is well drained, and receives hyperalimentation.

Operative therapy of thoracic esophageal perforations

The operative management of a thoracic esophageal perforation depends upon the medical status of the patient and the timing of the perforation whether early or late. Needless to say, the earlier the diagnosis is made and appropriate therapy instituted, the better the outlook for the patient. Basic management principles of all thoracic esophageal perforations are as outlined in Table 3. These are: (a) control of the esophageal leak; (b) eradication of mediastinal and pleural sepsis; (c) re-expansion of the remaining lung; (d) prevention of gastric reflux; (e) appropriate nutritional and pulmonary support; (f) appropriate antibiotics; and (g) adequate postoperative drainage of the involved anatomic area.

1.	Control of esophageal leak
2.	Eradication of mediastinal and pleural sepsis
3.	Re-expansion of the lung
4.	Prevention of gastric reflux
5.	Nutritional and pulmonary support
6.	Appropriate antibiotics
7.	Postoperative drainage of resident septic foci

Table 3 Management principles of thoracic perforation

Surgical methods utilized in dealing with thoracic perforations include the following as listed in Table 4. These include: (a) primary closure with buttress or patch; (b) exclusion and diversion; (c) T-tube fistula; (d) thoracic drainage and irrigation; (e) resection; and (f) decompression gastrostomy and feeding jejunostomy.

1.	Primary closure with buttress or patch
2.	Exclusion and diversion
3.	T-tube fistula
4.	Thoracic drainage and irrigation
5.	Resection
6.	Decompression gastrostomy and feeding jejunostomy

Table 4 Methods utilized in thoracic perforations

The specific judgement as to how an esophageal perforation should be treated is related to the timing of the perforation, the medical and physiologic status of the patient, the degree of contamination and the experience of the operating surgeon. As a basic tenet, esophageal resection should not be performed on an emergency basis except when there is gross obstruction from tumor and the perforation is early with minimal contamination. This is otherwise associated with a high morbidity and mortality.

For most thoracic perforations, the following narrative description is given as a guideline. The patient is prepared preoperatively with attention to the ABCs of resuscitation with the addition of broad-spectrum antibiotics to cover both aerobic and anerobic micro-organisms. Once adequately stabilized and with a confirmed diagnosis of esophageal perforation, the patient is explored through a left sixth intercostal space incision. Only if the perforation presents to the right side should one explore the patient through the right chest.

Once the thoracic cavity is entered, there is usually a large amount of pleural reaction and fluid in the mediastinum. The mediastinal pleura should be opened widely and the entire thoracic esophagus exposed from the aortic arch to the esophageal hiatus. Once this exposure is obtained, the tear or perforation of the esophagus is generally easily identified. Once identified, and if less than 24 h old and in some even 48 h, the perforation should be closed in two layers of interrupted 3-O silk. The first layers are generally mucosal approximation and the second layer gives muscular approximation. If definitive closure is obtained, a pleural based Grillo flap is mobilized from the surrounding parietal pleura and wrapped around the esophagus in a 360-degree fashion. A Jackson Pratt drain is placed down to the area of the perforation and two large chest tubes are left in place. The patient is allowed anything orally and hyperalimented for 10 days and then restudied.

If the perforation cannot be repaired definitely, a T-tube is placed or an attempt is made to close the injury definitely. A pleural-based Grillo flap is then placed around the esophagus. In this latter incidence a decompression gastrostomy and feeding jejunostomy are performed. In the past 25 years, the author has treated approximately 35 patients with duration of perforation from 4 h to 5 days with attempt at primary closure with a pleural wrap. Widespread mediastinal debridement and drainage is combined with hyperalimentation. Gastrostomy and jejunostomy have rarely been performed in the past 5 years. This technique has resulted in only two hospital deaths in the last 10 years.

While highly publicized and published in the late 1970s and 1980s, diversion and exclusion is rarely performed today. If done, the esophageal exclusion is generally a side diverting esophagotomy and the exclusion is generally a 0-Vicryl or PDS wrapped around the gastroesophageal junction and brought out through the lateral thoracic skin.

Thoracic esophageal perforation continues to carry a high morbidity and mortality. In general, when diagnosis is made within 24 h, there is only a 5 to 10 per cent mortality. When diagnosed after24 h, mortality increases exponentially. Some series report a greater than 50 per cent death rate when diagnosed after 48 h.

Thoracic esophageal perforation continued to be a major problem for the general thoracic surgeon. However, when diagnosis is made early and treated appropriately, the results can be quite acceptable.

Surgical therapy of Mallory–Weiss syndrome

The majority of patients with Mallory–Weiss syndrome can be treated conservatively with endoscopy, laser, no oral intake, and blood replacement. An occasional patient will require surgical treatment because of continued bleeding. These patients should be explored trans-abdominally through a high vertical gastrotomy incision. The bleeding site can be located and stitched appropriately. The gastrostomy should be closed in two layers and hyperalimentation instituted.

Surgical therapy of these three conditions presents some of the most challenging surgery in the field of thoracic surgery. The experienced surgeon with early intuition and appropriate therapy can lead to a successful outcome in the majority of these patients who are frequently so desperately ill.

Further reading

Barrett NR. Spontaneous perforation of the esophagus. *Thorax* 1946; **1**: 48–54.

Boerhaave H. Atrocis, nec descripti prius, morbi historia. Verbatum English translation (Derbes V, Mitchel R). *Bulletin of the American Library* 1955; **43**: 217.

Fell SC. Esophageal perforation in esophageal surgery. In: Pearson FG, Deslauriens J, Ginsberg RJ, Hiebert CA, McKneally MF, Vrschel HC, eds. *Esophageal surgery*, pp. 495–515. Churchill Livingstone, New York, 1995.

Grillo HC, Wilkens EW Jr. Esophageal repair following late diagnosis of intrathoracic perforation. *Annals of Thoracic Surgery* 1975; **20**: 387–97.

Michel LA, Collard JM. Perforation, Boerhaave's syndrome, and Mallory–Weiss syndrome. In: Morris, PJ, Sir, ed. *Oxford textbook of surgery*, 1st edn, Vol. 2, pp. 868–73. Oxford Medical Publications.

Vrschel HC, Razzok MA, Wood RE. Improved management of esophageal perforation: exclusion and diversion in continuity. *Annals of Surgery* 1974; **179**: 587–91.

22.5 Benign and malignant tumours

Cameron D. Wright, Henning A. Gaissert and Douglas Mathisen

Esophageal carcinoma
Introduction
Epidemiology
Risk factors
Pathology
Biology
Staging
Symptoms
Signs
Laboratory data
Radiology
Endoscopy
Treatment options
Unusual esophageal neoplasms
Benign tumors of the oesophagus
Leiomyoma
Benign polyp
Further reading

Esophageal carcinoma

Introduction

Esophageal cancer is an uncommon malignancy that is uniformly lethal when untreated. There are no early symptoms and no anatomic barriers to its spread. Lymph-node involvement is common and indicates a poor long-term prognosis. Early stages of the disease are only found serendipitously or during screening for precursor lesions. As a result, the typical patient presents with locally advanced disease.

Surgical resection provides the best chance of cure. Combined-modality treatment including radiotherapy and platinum-based chemotherapy has raised hope for improvement of survival with promising preliminary data.

Epidemiology

There is considerable global variation in the incidence of squamous-cell esophageal cancer. The incidence ranges from 5 per 1 000 000 in the United States of America to 100 per 100 000 in Linxian county, China (Table 1). There is also a marked difference in death rates within each country. The peak age of occurrence is between 50 and 70 years, with a median age of death of 66 years in the United States. In Western nations, esophageal carcinoma is much more common (about fivefold) in men than women. It is threefold more common in black (African American) than white (caucasoid) individuals in the United States.

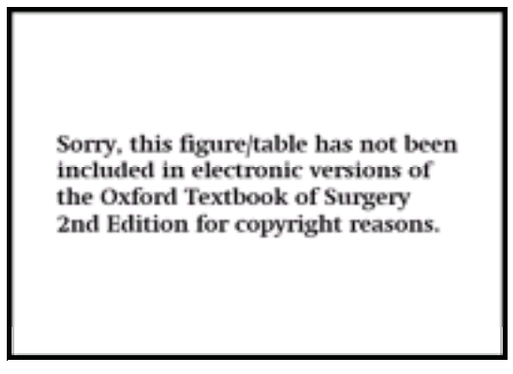


Table 1 No caption available.

Adenocarcinoma of the esophagus separate from the gastric cardia was once an exceedingly rare tumor. Forty years ago, two North American pathology reviews found adenocarcinoma in less than 3 per cent of all cases of esophageal cancer. Recent reports from industrialized nations indicate a rise in incidence such that adenocarcinoma accounts for more than half of all new carcinomas. A epidemiologic survey by Blot in the United States showed that incidence rates between 1976 and 1987 were fairly stable for squamous-cell carcinoma, but increased more than 100 per cent for adenocarcinoma among men. From 1984 to 1987, adenocarcinoma accounted for 34 per cent of all esophageal cancers among white men, while the corresponding figures for black men, white women, and black women were 3, 12, and 1 per cent, respectively. The rate of increase during the 1970s and 1980s surpassed that of any other cancer. An important observation is the epidemiologic similarity between carcinoma of the gastric cardia and the esophagus; the increase in incidence and the sex and ethnic predilections apply to both tumors, whereas the incidence of cancer of the gastric fundus and antrum continues to decline.

Risk factors

Alcohol

Many epidemiologic studies confirm that heavy alcohol users have an increased risk of squamous-cell carcinoma and that the risk is substantial in very heavy drinkers. The risk increases incrementally when alcohol and tobacco abuse are combined. Alcohol does not appear to be a risk factor for esophageal adenocarcinoma.

Tobacco use

Smoking leads to a dose-dependent increased risk of squamous-cell carcinoma. Moderate smokers have a fivefold increase in risk; this is true for cigar and pipe smoking as well as for cigarette smoking. Smoking approximately doubles the risk for esophageal adenocarcinoma.

Nutritional deficiency

The geographic variability in the incidence of squamous esophageal cancer correlates strongly with areas where poor nutrition is prevalent. The nutritional factors (other than malnutrition) implicated include deficiencies in zinc, molybdenum, magnesium, and iron. The clearest association is with molybdenum deficiency in the soil: this leads to an accumulation of nitrates and nitrites in plants, which in turn may lead to the production of nitrosamines, a known esophageal carcinogen in experimental animals. Plummer–Vincent (Patterson–Kelly) syndrome is associated with carcinoma of the upper third of the esophagus and occurs not only with iron-deficiency anemia but also with vitamin B deficiencies. Celiac disease (non-tropical sprue) is associated with esophageal cancer, perhaps due to the malabsorption syndrome that leads to multiple nutritional deficiencies.

Socioeconomic status

Both squamous-cell and adenocarcinoma of the esophagus are associated with lower socioeconomic status, specifically with an education of fewer than 12 years and

an income below US$15 000 per year.

Chronic esophageal irritation

Gastroesophageal reflux may lead to loss of squamous mucosa and its replacement with columnar epithelium. Columnar-lined esophagus is observed in per 15 cent of patients with endoscopic evidence of oesophagitis; there is a 30- to 40-fold increase in the incidenceof adenocarcinoma of the esophagus compared to that in the general population. The presence of intestinal metaplasia and dysplasia increases the incidence further, while risk is not related to the length of the columnar replacement.

Esophageal disorders associated with an increased incidence of carcinoma include chronic lye strictures, achalasia, and perhaps diverticula. Long-term use of spicy and hot drinks may also predispose to chronic irritation and is implicated in the pathogenesis of esophageal carcinoma. Tylosis is a rare, autosomal-dominant disease associated with an increased risk of carcinoma of the esophagus.

Pathology

Three gross patterns of growth are commonly seen. The ulcerating type is a well-defined ulcer with raised, irregular edges. The fungating type has a large intraluminal growth component with an irregular surface. The infiltrative type has extensive intramural growth, often circumferential, with minimal ulceration. There is often a significant submucosal extension of tumor far from the visible edge at the mucosal surface (longitudinal spread), more so in a cephalad than a caudad direction. In one study, cancer was present 6 cm cephalad from the primary tumor in 22 per cent of patients and 9 cm from it in 11 per cent; caudad growth was rarely more than 5 cm distal. These data emphasize the importance of wide resection of the esophagus and frozen-section analysis of resection margins to ensure complete removal of the tumor.

For purposes of classification the esophagus may be divided into three anatomic areas, the upper, middle, and lower thirds. The upper third (cervical esophagus) extends from the cricopharyngeal sphincter to the thoracic inlet, the middle third from the thoracic inlet to 10 cm above the gastroesophageal junction, and the lower third from 10 cm above that junction to the cardia of the stomach. About 15 per cent of squamous-cell cancers occur in the upper third, 50 per cent in the middle third, and 35 per cent in the lower third.

Eighty per cent of adenocarcinomas are in the lower third and 20 per cent are in the middle third. Patients who present with dysphagia have neoplasms that are usually transmural with regional lymph-node metastases. There is extensive submucosal spread of the tumor similar to squamous-cell carcinoma. Dysplasia is present in the majority of patients who undergo resection for Barrett's adenocarcinoma, lending to support to the suggestion that dysplasia leads to high-grade dysplasia and hence to invasive carcinoma. Dysplasia is graded using criteria established by the Inflammatory Bowel Disease–Dysplasia Morphology study group. Low-grade dysplasia consists of mildly dysplastic nuclei confined to the basal half of the epithelium; high-grade dysplasia is synonymous with carcinoma *in situ.*

Transmural tumor penetration (vertical growth) is present in the majority of cases that come to surgery. Lack of a serosal covering of the esophagus may explain the early invasion of mediastinal structures. Those commonly invaded include the trachea and left main bronchus, aorta, pericardium, and pleura. A malignant fistula between the airway and esophagus occurs more commonly than an aortoesophageal fistula.

Lymph-node metastases are common and correlate with the depth of tumor invasion. For lesions limited to the submucosa (stage T1) there is only a 14 per cent incidence of positive nodes (for discussion of staging, see below.) The incidence of nodal metastases rises from 30 per cent for lesions limited to the muscularis propria (T2), 50 per cent for lesions up to the adventitia (T3), and 75 per cent when the tumor invades the periesophageal tissues (T4). Lymph-node metastases may be found distant from the primary because of the extensive longitudinal spread of tumor via the submucosal lymphatic vessels. The lymphatic drainage of the esophagus is mostly longitudinal rather than segmental, further complicating the surgical treatment of esophageal cancer. The Japanese surgeon Akiyama has convincingly demonstrated that lymph nodes are often positive at the opposite end of the esophagus from the carcinoma, regardless of whether it is in the upper or lower esophagus (Table 2). Visceral metastases involving the lung, liver, bone, kidneys, pleura, and brain are common at death.

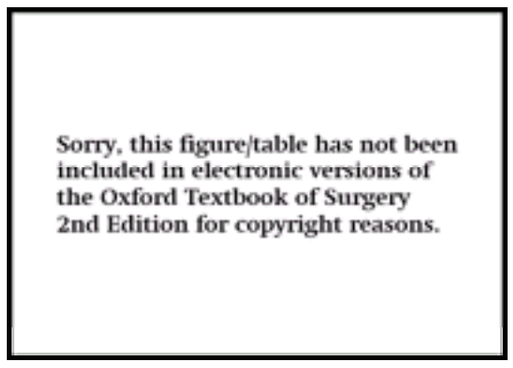


Table 2 No caption available.

Biology

Squamous-cell carcinoma

The natural history of esophageal cancer may be inferred from the results of extensive mass screening done in China on high-risk populations. Screening was primarily by abrasive cytology utilizing a swallowed balloon that is inflated and pulled back through the esophagus, thus collecting a representative sample of epithelial cells. Dysplastic cells were almost always found in association with early carcinoma. Furthermore, dysplasia could progress to carcinoma, although not invariably so. In one study of over 1500 patients followed for up to 12 years, 15 per cent developed carcinoma if severe dysplasia was present. Only 1 per cent with mild dysplasia developed esophageal cancer, and no carcinomas were found in the cytologically normal control group. In a separate study of 14 000 individuals, the average age of those with severe dysplasia was 52 years, whereas the average age of those with carcinoma was 57 years, suggesting a 5-year lag time in developing carcinoma once there is severe dysplasia. Not all patients with dysplasia will develop carcinoma: in one study, 45 per cent of cases with mildly dysplastic cells and 40 per cent with severely dysplastic cells had returned to normal over several years. In patients with early, cytologically diagnosed carcinoma who refused treatment and were followed, about one-half remained asymptomatic in the early stages of the disease for a mean period of 75 months. The other one-half developed symptomatic late carcinoma an average of 55 months after diagnosis. Thus the development of symptomatic carcinoma takes many years and allows ample time for early diagnosis. Treatment of asymptomatic cancers in China provides a greater than 90 per cent survival at 5 years, emphasizing the importance of early diagnosis.

Once overt symptoms occur, the average survival without treatment is 9 months. The most common cause of death is bronchopneumonia. Patients usually procrastinate in reporting their symptoms, leading to a further delay in diagnosis, often of 3 to 4 months. Metastases to lymph nodes and visceral organs as well as mediastinal invasion have often already occurred, precluding the chance of a surgical cure. Mass screening programs in areas of low incidence are not cost-effective.

Adenocarcinoma

Most esophageal adenocarcinomas are associated with a columnar-lined (Barrett's) esophagus. Other possible sites of origin include the superficial mucosal glands at either end of the esophagus, the deep submucosal glands distributed throughout the esophagus, and areas of ectopic gastric mucosa. Barrett's esophagus is thought to be an acquired condition due to severe gastroesophageal reflux. The development of adenocarcinoma in the columnar-lined mucosa was first described in 1952 and the tendency for this type of epithelium to develop adenocarcinoma was first reported by Naef in 1975. Barrett's mucosa can become dysplastic and when high-grade dysplasia develops, there is a substantial risk for developing adenocarcinoma. Estimates of risk of malignant change in patients with simple Barrett's esophagus are about one case per 300 patient years from the time of diagnosis.

Patients with Barrett's esophagus should undergo surveillance endoscopy approximately every year to screen for the development of dysplasia or carcinoma. Four-quadrant biopsies should be taken every l to 2 cm until normal squamous epithelium is reached. Special attention should be paid to the subtle alterations in the appearance of the epithelium, as these areas are more likely to show advanced disease. If low-grade dysplasia is diagnosed, surveillance should be increased to every 3 to 6 months and intensive medical therapy instituted; dysplasia has regressed with medical therapy in the majority of patients so treated, but carcinoma has

developed during intensive medical therapy for low-grade dysplasia. There is no proof that an antireflux operation can prevent adenocarcinoma from developing. Once high-grade dysplasia has developed, esophagectomy should be performed. Approximately one-half of patients have invasive carcinoma in the resected specimen (Table 3). The majority so treated are either stage 0 or I: 5-year survival approaches 100 per cent in stage 0 patients and 80 to 90 per cent in stage I patients who undergo resection for high-grade dysplasia.

Author	Year	No. of patients	Adenocarcinoma (%)	Stage I	> Stage I
Altorki	1991	8	4	3	1
Pera	1992	18	9	6	3
Rice	1993	16	6	6	0
Heitmiller	1996	30	13	8	5
Ferguson	1997	15	8	7	1
TOTAL		87	40 (46%)	30 (75%)	10 (25%)

Table 3 Esophagectomy for high-grade dysplasia in Barrett's esophagus

Staging

The current staging system was developed by the American Joint Committee on Cancer in conjunction with the International Union Against Cancer (Table 4). The primary determinants of survival in surgically treatable esophageal cancer are the depth of invasion and the presence of lymph-node metastases; the staging system incorporates these two critical variables. Survival is stage dependent. Although postsurgical staging can be done accurately, presurgical staging is still suboptimal because computed tomographic scanning does not reliably determine the depth of invasion or the status of the regional nodes. Endoscopic ultrasonography is assuming an increasingly important role, as it determines tumor depth and lymph-node metastasis with high accuracy.

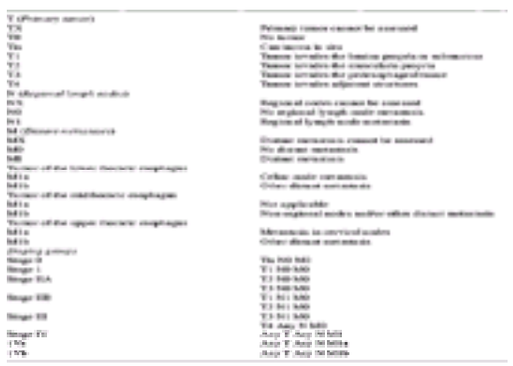


Table 4 Staging of esophageal carcinoma

Symptoms

Dysphagia, first to solids then liquids, is the most common presenting symptom. This is a late symptom and usually indicates that two-thirds of the circumference is involved. Patients often adjust their diet to soft solids and flush their foods through with liquids to accommodate their dysphagia. Weight loss is common and at times disproportionate to the duration of dysphagia. Odynophagia (painful swallowing) is often seen. Steady, deep chest pain often indicates mediastinal invasion.

Cough or hoarseness is common in upper esophageal tumours as a result of a direct invasion of the airway or the recurrent laryngeal nerve. Aspiration pneumonia may be present either as a result of overflow obstruction or due to a malignant fistula between esophagus and airway. Patients with a fistula usually have severe paroxysms of coughing immediately after swallowing liquids.

Signs

Physical examination is usually unrewarding. Cervical adenopathy is rarely palpable. Evidence of weight loss is usually apparent.

Laboratory data

Laboratory tests are rarely helpful in making the diagnosis, although they may be of some benefit in staging (i.e., elevated alkaline phosphatase for bone or liver metastases). Laboratory data also help in assessing the risk of esophagectomy (i.e., renal and hepatic function). Pulmonary function testing and arterial blood-gas measurement are helpful in quantifying the extent of chronic obstructive pulmonary disease. Patients with underlying, symptomatic coronary arterial disease should have an echocardiogram to assess left ventricular function and a stress test to assess the extent of ischemic myocardium.

Radiology

Chest radiographs

Chest radiographs are usually unremarkable. Possible abnormalities include an esophageal air–fluid level, infiltrates suggesting aspiration, pleural effusion suggesting pleural dissemination, mediastinal widening (due to tumor or lymphadenopathy), and pulmonary nodules suggesting lung metastases.

Barium esophagram

The barium swallow is the mainstay of radiologic diagnosis (Fig. 1). Abnormality of the esophageal axis (angulation of a long axis) strongly suggests mediastinal invasion and thus helps predict unresectability.

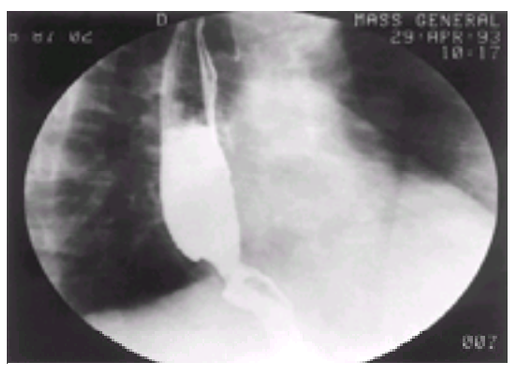


Fig. 1. Barium swallow demonstrates the typical annular defect of the lower esophagus in a patient with an adenocarcinoma.

Computed tomography of the chest and abdomen

Computed tomography (**CT**) displays the esophageal wall and mediastinal structures clearly, and, in addition, allows evaluation of the common metastatic sites (liver, lymph nodes, adrenal glands, lungs, kidneys) (Fig. 2). All patients should undergo CT evaluation, but CT does not accurately predict invasion. Gross distortion of a mediastinal structure indicates invasion; the presence of a fat plane between the tumor and the structure predicts the absence of invasion. Evaluation of tumors at the gastroesophageal junction is very difficult, although it is improved by distension and contrast opacification of the stomach. Evaluation of lymph nodes is suboptimal: enlargement does not always predict metastases and most excised, positive lymph nodes are of normal size.



Fig. 2. Computed axial tomogram demonstrates a large retrocardiac mass in a patient with an adenocarcinoma of the distal esophagus.

Radionucleotide scans

Bone scans are performed on clinical suspicion of bone metastases based on either a history of bone pain or an elevated alkaline phosphatase. Liver scans are inferior to CT in determining the presence of liver metastases.

Endoscopy

Esophagoscopy

Esophagoscopy is required in the assessment of all patients with dysphagia and allows histologic (or cytologic) confirmation of suspected carcinoma. It is important to measure the length of the lesion and the distance from the incisors for staging and treatment planning. Typical tumors are friable and bleed easily. Multiple biopsies should be taken from suspicious areas; the ability to make a histologic diagnosis is increased by taking multiple biopsies. Brush cytology is often very helpful. About 10 per cent of biopsies are non-diagnostic; repeat esophagoscopy and biopsy is then required. The stomach, pylorus, and duodenum should also be examined to assess any additional pathology and to evaluate the stomach for use as an esophageal substitute.

Bronchoscopy

Bronchoscopic examination of the airway is necessary in all upper- and middle-third carcinomas. Transmural spread of tumor can be confirmed by visual inspection and biopsy. Bulging of the airway indicates abutment, but does not usually signal direct invasion of the airway.

Laparoscopy

Laparoscopy can allow reliable assessment and biopsy of the superficial aspect of the liver, perigastric lymph nodes, and the peritoneum, if indicated by previous staging examinations.

Thoracoscopy

Thoracoscopy can be used to confirm invasion into local structures or to sample lymph nodes for staging. The role of video-assisted thoracoscopy for esophagectomy needs to be defined.

Endoscopic ultrasonography

Endoscopic ultrasound improves the ability to determine wall penetration and abnormal lymph nodes; thus it may improve preoperative staging. Five distinct wall layers can be identified that correspond to the mucosa, lamina propria, muscularis mucosa, muscularis propria, and adventitia. Carcinoma appears as an irregular, hypoechoic mass (Fig. 3).

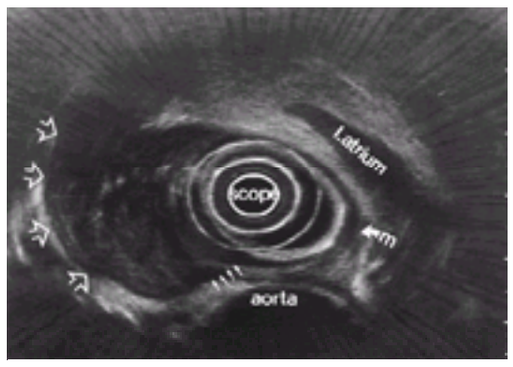


Fig. 3. Endoscopic ultrasonographic image of a mass in the mid-esophagus; note the adjacent structures in the mediastinum, aorta, and the left atrium. The wall structure of the esophagus is also evident; the thickest hypoechoic layer is the muscularis propria (m). The tumor has extended beyond the muscularis (small arrows) and into the mediastinum (large arrows).

Depth of penetration of the wall can be assessed accurately. Regional lymph nodes can be identified and metastases predicted on the basis of their size and appearance. The use of endoscopic ultrasound is limited to those neoplasms that allow passage of the probe, but the presence of a tight malignant stricture is an excellent (90 per cent) predictor of a stage III tumor. The exact role of endoscopic ultrasound in the management of esophageal cancer is uncertain. Its greatest use may be in improved staging for patients undergoing preoperative adjuvant therapy. It appears to be more accurate than CT for determining both the T and N status of esophageal cancer.

Treatment options

The goal of treatment in carcinoma of the esophagus is twofold: palliation of dysphagia and cure of the cancer. The standard of therapy is esophageal resection. Resection quickly restores swallowing to normal and palliates dysphagia. As a single therapy, surgery offers the greatest chance for cure. Cure rates vary from institution to institution. A cumulative review of all published reports in English for surgical resection of squamous-cell carcinoma of the esophagus by Muller demonstrates the magnitude of the problem and summarizes the experience with 76 900 patients. Only 56 per cent of patients had resectable disease at first presentation. Resection was associated with an operative mortality of 13 per cent. Survival was 27 per cent at 1 year, 12 per cent at 2 years, and only 10 per cent at 5 years. Recent reports from selected institutions have shown a decrease in operative mortality and a increase in 5-year survival. Many centers have reported mortality rates of under 5 per cent and 5-year survivals around 20 per cent with surgery alone for squamous-cell carcinoma of the esophagus.

Resection

There are many surgical approaches available for esophageal resection. The main determinants of the operation chosen are the surgeon's preference and the level of the tumor. The three most common approaches currently in use are the Sweet left thoracoabdominal, the Lewis laparotomy and right thoracotomy, and the Grey–Turner transhiatal esophagectomy as popularized by Orringer. Excellent clinical results have been obtained with each technique in centers experienced in their use. There is little difference in morbidity or mortality between any of these approaches. Other techniques include the radical *en bloc* esophagectomy, an exclusive left thoracic approach (the stomach mobilization is done through a diaphragmatic incision), a three-hole esophagectomy (right thoracotomy, followed by laparotomy and neck incision), and the right thoracoabdominal approach.

The Sweet approach

The Sweet left thoracoabdominal approach is best employed for carcinomas of the gastroesophageal junction or low esophagus; tumours 35 cm or more from the incisors are ideally suited to this approach. A double-lumen endotracheal tube is used to block and deflate the left lung, enhancing exposure. The patient is placed in the right lateral decubitus position. An oblique, left upper quadrant laparotomy is performed to explore the gastroesophageal junction and the liver. If there is no contraindication to resection, a thoracoabdominal incision is made through the sixth or seventh interspace. The diaphragm is incised circumferentially to avoid injury to branches of the phrenic nerve. The stomach is mobilized on the right gastric and gastroepiploic arteries. The omentum is divided, preserving the right gastroepiploic artery; the left gastric artery is double ligated. The gastrohepatic omentum is divided, with care taken to identify accessory arteries to the left lobe of liver. The hiatus is dissected. A pyloromyotomy or pyloroplasty is then done, according to the surgeon's preference. The esophagus is then dissected *en bloc* from the pericardium to aorta. The level of anastomosis should be at least 5 cm above the gross tumor to ensure adequate margins. If a higher level of intrathoracic anastomosis is desired, dissection can be easily performed below the arch of the aorta and then above it, allowing a supra-aortic anastomosis to be made lateral to the aortic arch. If exposure is limited, a second interspace thoracotomy can be made, usually through the fourth interspace, allowing better access for anastomosis. The main advantage of this approach is the excellent exposure of the gastroesophageal junction and the ease of mobilizing the stomach, especially in obese patients. It offers the best exposure for the most complete abdominal lymphadenectomy. The limiting factor to this exposure is the position of the heart and aortic arch while doing the anastomosis; this can be partially overcome by choosing a second higher interspace, as discussed. The largest negative factor about this approach is the fact that the aortic arch is in the way of the surgeon, thus limiting access to the total thoracic esophagus and encouraging an anastomosis below the aortic arch.

Lewis approach

The Lewis right thoracotomy and laparotomy is best for mid- or lower-third lesions. An upper midline laparotomy is performed and the upper abdomen explored. If there are no contraindications to resection, the stomach is mobilized as previously described. It is important to enlarge the hiatus to prevent compression of the stomach when it is brought into the chest. As much of the lower esophagus is mobilized as can be accomplished from the abdomen, since this can be difficult through a high right thoracotomy. A table-mounted retractor facilitates the dissection of the hiatus and lower esophagus. The patient is then positioned for a right thoracotomy. The right lung is collapsed, and the azygos vein ligated and divided. There is excellent exposure for the entire thoracic esophagus, which is dissected from pericardium to vertebral body. Care should be taken to ligate tissues in between the aorta and esophagus to avoid injury to the thoracic duct. The stomach is then elevated up into the chest and a high intrathoracic anastomosis is made at the apex of the right chest. Care should be taken to avoid pulling too much of the stomach into the chest: if pulled too tightly a relative obstruction can be created at the hiatus; excess stomach will tend to fall into the costophrenic sulcus and impair emptying. The advantage to this approach is excellent exposure of the thoracic esophagus and ease of anastomosis. It is easy to obtain a wide margin on the tumor because of the excellent exposure of the superior aspect of the esophagus. The main difficulty encountered with the Lewis approach is lack of exposure of the gastroesophageal junction and hiatus, especially in obese patients.

Transhiatal esophagectomy

Transhiatal esophagectomy is best used to remove upper-third or lower-third neoplasms; it is relatively difficult to remove stage III, mid-esophageal tumors with this approach. The operation is done with the patient in a supine position and with a single-lumen endotracheal tube. A laparotomy is performed first and the abdomen explored. The stomach is prepared as an esophageal substitute. It is helpful to open the hiatus anteriorly, as described by Pinotti; this facilitates exposure of the distal esophagus almost to the level of the carina. Vessels can be ligated under direct vision. The side of left neck is opened and the esophagus exposed; care is taken to avoid retractor injury to the recurrent laryngeal nerve in the tracheoesophageal groove. The upper third of the esophagus can be dissected under direct vision. If a transmural tumor is present in this area and if additional exposure is required, a partial upper sternal split can be performed, as first described by Scannell, which facilitates exposure in this area. The area around the carina is often difficult to expose and usually has to be done ‘blind'. The esophagus is removed and the stomach brought up through the posterior mediastinal esophageal bed; a cervical anastomosis is then performed.

Perceived advantages to transhiatal esophagectomy are the avoidance of a thoracotomy and the presence of a cervical anastomosis. Anastomotic leaks occur more commonly in the cervical area, but can be easily drained and are rarely associated with the devastating complications of thoracic leaks. Although avoidance of a thoracotomy allows greater tolerance of the procedure as compared with transthoracic techniques, any advantages are difficult to prove. The major drawbacks are the limited *en bloc* resection, the ‘blind' area around the carina, and the possibility of an increased incidence of injury to nearby structures (thoracic duct, azygous vein, trachea, recurrent laryngeal nerves). No randomized studies have compared transhiatal with transthoracic esophagectomy. Nonrandomized studies have failed to demonstrate less morbidity, mortality, or hospital stay with the transhiatal technique (Table 5).

Author (year)	*Operative mortality (%)*	*Respiratory complications (%)*	*Anastomotic leaks (%)*
Shiraishi (1986)	THE (30) 13.3	21.4	15.4
	TTE (65) 6.2	16.9	1.8
Hankins (1989)	THE (26)	8.0	54.0
	TTE (52) 6.0	38.0	5.9
Stewart (1991)	THE (87)	5.9	
	TTE (43) 2.3		2.3
Daniel (1992)	THE (77)	8.0	
	TTE (24) 8.0		4.0

Number of patients in parentheses.

Table 5 Comparison of transhiatal (THE) and transthoracic (TTE) esophagectomy

Fate of the pylorus

Most surgeons favor a drainage procedure with esophagectomy. A small number of controlled trials have compared the results of no pyloric drainage with using a drainage procedure; no clear advantage was shown. The incidence of gastric-outlet obstruction with no drainage has been under 10 per cent, and the need for reoperation has been under 2 per cent. Dumping is a recognized side-effect of a drainage procedure and is one reason that some surgeons prefer not to perform one. If postoperative obstruction of the gastric outlet fails to respond to conservative measures, reoperation can be difficult. This is the main reason why surgeons prefer a pyloromyotomy or pyloroplasty.

Anastomotic technique

In 1942, Churchill and Sweet described a triple-layer technique of esophagastrostomy and conventional *en bloc* resection of the cancer and adjacent lymph nodes. Five years later, Richard Sweet published his initial experience with surgical management of carcinoma of the esophagus in 141 patients. In a time when there were no sophisticated postoperative monitoring devices, mechanical ventilation, or broad-spectrum antibiotics, his results were remarkable: an operative mortality of 15 per cent, anastomotic leaks in 1.4 per cent of patients, and overall 5-year survival of 11 per cent. This served as a standard for many years. Much of the success is directly attributed to the reliability of the anastomosis and remarkably low rates of anastomotic leakage. Sweet emphasized the details of technique and warned against factors predisposing to anastomotic leakage. The lack of a serosal layer and the segmental blood supply of the esophagus make esophageal anastomosis more demanding than other intestinal anastomoses. Atraumatic handling of the tissues, preservation of the blood supply of both the esophagus and stomach, avoidance of the use of crushing clamps, lack of tension on the anastomosis, use of fine interrupted sutures, cutting with a knife or other sharp instrument, and tying sutures gently but firmly to avoid cutting tissues are all important details in the performance of an anastomosis. Few modifications have been made from the technique Churchill and Sweet proposed 45 years ago. Preservation of the blood supply is crucial when mobilizing the stomach and esophagus. The blood supply of the stomach will be from the right gastric and right gastroepiploic arteries. The esophagus should not be mobilized a few centimeters above the proposed level of the anastomosis to avoid interference with the segmental blood supply. A circle approximately 2 cm in diameter is scored on a portion of the serosa of the stomach (Fig. 4). The circular defect should be 2 cm away from the stapled edge of the stomach to avoid compromising the blood supply. Individual vessels are identified and ligated with fine silk sutures; this minimizes bleeding while the anastomosis is performed and allows for precise placement of sutures.

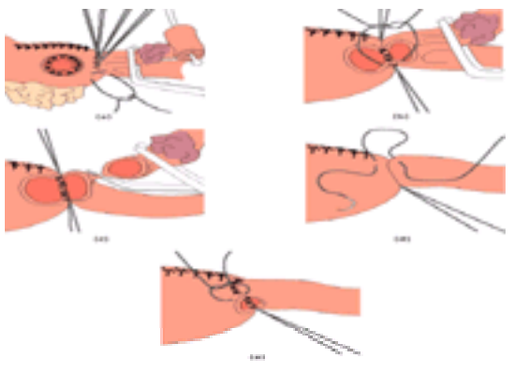


Fig. 4. Technique of two-layer esophagogastric anastomosis. (a) A 2-cm circle of gastric serosa has been circled and submucosal vessels ligated. A distance of at least 2 cm is kept between this gastric button and the suture line to prevent ischemia. A back row of horizontal mattress sutures is completed. (b) The gastric button is excised and the esophageal lumen opened. The back row of the inner layer is placed, approximating gastric and esophageal mucosa. (c) The back row of the inner layer is completed and the oesophagus transected. (d) The front row of the inner layer is completed, inverting the last opening with a Connell stitch. A nasogastric tube is advanced under direct vision into the stomach. (e) The outer layer is nearly finished.

Interrupted horizontal mattress sutures of fine material (we use 4–0 silk) are used to construct the back row of the anastomosis. Corner stitches are placed first, and the remaining sutures are evenly spaced between them. The sutures on the stomach involve the seromuscular layers and those on the esophagus the longitudinal and circular muscle layers. The esophageal sutures should be deep enough to include both the longitudinal and circular muscles. The sutures should not be tied too tight, to avoid necrosis or cutting through the muscle.

The esophagus is opened sharply from one corner stitch to the other. The circular button of stomach is removed. The inner layer is completed with simple sutures, including just the mucosa of the esophagus and the full thickness of the stomach. The knots are on the inside, thereby allowing inversion or turning-in of the mucosa of both the esophagus and stomach; this is accomplished for the entire circumference of the anastomosis. A nasogastric tube is passed into the stomach under direct vision before a single Connell stitch is placed for closure of the final opening. Healing of the inverted mucosa is an important feature in preventing leakage, and the location of the knots on the luminal side minimizes foreign-body reaction within the tissues of the anastomosis. The outer row is completed using horizontal mattress sutures as described for the back row of the outer layer.

The omentum mobilized with the stomach is placed over the anastomosis anteriorly to provide an additional layer of coverage. The posterior part of the anastomosis lies between the esophagus and the more proximal stomach. A few sutures are placed between the stomach and the mediastinal pleura to avoid tension on the anastomosis when the patient is upright, particularly if the stomach is full.

The viability of tissues on each edge of the anastomosis is best maintained by avoiding surgical trauma. The edges are never crushed with clamps and, indeed, are handled with forceps as little as possible. Once the first stitch has been placed and tied, traction on it permits placement of the next without the need for grasping the mucosa with instruments. The sutures are tied by positioning the index finger cephalad to the anastomosis, lifting the stomach to the esophagus, and avoiding pulling down on the fixed and more fragile esophagus. Delicacy in suturing is especially important for the outer layer of the anastomosis because the esophagus lacks a peritoneal surface.

A nasogastric tube passed through the anastomosis for a short time avoids distraction at the suture line from a distended stomach. Gentle, periodic irrigation of the tube ensures its patency. Temporary gastric decompression more than compensates for any potentially deleterious effect of an intraluminal foreign body lying against the suture line for a short period.

We have published our experiences with this technique in a consecutive series of 104 patients. There were three postoperative deaths (2.9 per cent). Two deaths were attributable to pneumonia and respiratory failure (patients aged 59 and 73 years); both patients smoked at least one pack of cigarettes per day preoperatively. The third death was also due to pneumonia and respiratory failure. This 76-year-old patient had an emergency operation for massive gastrointestinal bleeding. Dilatation was necessary in five patients for anastomotic stricture 3 to 6 weeks postoperatively. One-to-three dilations were required for the successful resolution of dysphagia. Delayed anastomotic stricture was not apparent in this group of patients. All patients had postoperative barium swallows. There were no anastomotic leaks, even of localized type. Similar results have been reported by others using a similar two-layer technique.

Postoperative management

Patients are usually returned to the intensive care unit extubated. Thoracic epidural anesthesia with a combination of narcotics and local anesthetic is routinely used and greatly ameliorates postoperative pain. Pulmonary arterial catheters are not routinely used unless clinically indicated. Fluid status is carefully monitored and an indwelling bladder catheter is placed for several days. All patients receive perioperative antibiotics. Nasogastric tubes are routinely used re removed when bowel function returns. Chest physical therapy and nebulized aerosols are routinely administered to improve pulmonary toilet. Jejunostomy tubes are placed in high-risk patients or those who are nutritionally depleted. Enteral feeding is usually begun on the second or third postoperative day. A gastrografin swallow is obtained to check the anastomosis when bowel function returns. An oral diet is then begun and advanced to a soft solid diet.

Morbidity and mortality

Pulmonary complications are the most common postoperative complication encountered regardless of the approach used (Table 6).

	Transhiatal (%)	Transthoracic (%)
	(n = 8866)	(n = 2436)
Atelectasis	21	26
Pneumonia	27	13
Respiratory insufficiency	23	10

Table 6 Pulmonary complications after esophagectomy

The most worrying complication of esophagectomy is an anastomotic leak. Reported leakage rates as high as 10 to 15 per cent are common, but many series report rates of less than 5 per cent, usually associated with a two-layer anastomosis. Cervical anastomotic leaks usually present themselves with fever and a painful, swollen neck incision. The neck incision should be reopened and adequate drainage established. Healing will usually occur on provision of drainage, antibiotics, and adequate nutrition. A stricture may develop, requiring several dilatations to restore normal swallowing. Cervical anastomotic leaks are rarely fatal. Intrathoracic anastomotic leaks are much more serious. Mortality rates of 50 per cent or higher are still reported following free intrathoracic leaks, which demand prompt and aggressive treatment. If the leak occurs in the first few days postoperatively, re-exploration is necessary to exclude the possibility of gastric necrosis. In that case, the stomach should be debrided, closed, and returned to the abdomen. A cervical esophagostomy should be performed, and wide mediastinal and pleural drainage carried out. Reconstruction of the gastrointestinal tract can be done at a later date. Leaks detected after 1 week are best treated by immediate dependent drainage. Drainage should be converted to open drainage with a rib resection at an appropriate time. The critical issue is to establish adequate drainage of all infected material immediately after the diagnosis is made. Failure to achieve adequate drainage may lead to erosion of the airway or mediastinum. Small, asymptomatic, contained leaks can be seen as well. If they are indeed contained, treatment with antibiotics, withholding oral intake, and adequate nutrition eventually lead to healing. At least 1 week should pass before the patient is subjected to further studies. Anastomotic leaks remain the source of greatest morbidity and mortality. They require great judgement and immediate action to avoid a fatal outcome. These facts underscore the need for meticulous attention to detail in performing the esophagogastric anastomosis.

Survival following esophagectomy

At the Massachusetts General Hospital, before multimodality treatment of esophageal carcinoma became available, survival following resection in all patients was as follows: 2 years, 31 per cent; 3 years, 24 per cent; 5 years, 21 per cent. In a large series of Japanese patients who were grouped according to stage, 5-year survivals were as follows: stage I, 60 per cent; stage II, 30 per cent; stage III, 20 per cent; stage IV, 5 per cent. There is no proof of any difference in survival between radical *en bloc* esophagectomy, transthoracic esophagectomy, or transhiatal esophagectomy, even though they obviously involve removal of different amounts of esophagus and surrounding tissues as well as different degrees of lymphadenectomy (Table 7).

Technique	Author	Year	Number	Three-year survival (%)
Historical	Sweet	1952	284	22
Transthoracic	Kaifie	1982	142	24
Transhiatal	Orringer	1986	147	23
Radical (en bloc)	Skinner	1986	80	24

Table 7 Survival following esophagectomy

Radiation therapy

Radiation therapy remains as one of the possible primary treatments for squamous-cell carcinoma of the esophagus. Numerous large studies have been published, facilitating comparison with surgical treatment. Unfortunately, 5-year statistics in many large series in which patients were treated with curative intent with high-dose radiation therapy show a survival of only 5 to 10 per cent. The best results obtained with primary radiation therapy are in cervically located carcinomas. Local control and survival are greatest in patients with early stage I or stage II lesions and poorest with advanced lesions. Radiation therapy does not usually control the local disease. There is an approximately 70 per cent failure rate after radical radiotherapy at the local site. The majority (75 per cent) of local failures occur in the first year, thus prohibiting effective palliation for the majority of these patients. However, temporary palliation can be achieved with lower doses of radiation therapy, which can be of help in patients who are not expected to live long.

Adjuvant radiotherapy

Hancock has recently reviewed the results of preoperative radiation therapy in over 1000 patients. Unfortunately, the average 5-year survival was only 6 per cent overall, which is not significantly different from the results of either surgical treatment or radiation therapy alone. In a highly selected group of patients who completed both radical radiotherapy and resection, the 5-year survival rate was approximately 14 per cent, not significantly different from that reported for surgery. Additionally, some groups have reported inferior results with preoperative radiation therapy, primarily because of the toxicity associated with radical radiotherapy leading to greater rates of postoperative complications. There have been three well-conducted, prospective trials involving preoperative radiation therapy (Table 8). In none was a significant difference noted in median or in 5-year survival compared with surgical treatment alone. Preoperative treatment with radiation therapy is not recommended. One well-designed, randomized trial of postoperative radiation therapy following esophagectomy again showed no significant difference in long-term survival. Routine use of postoperative radiation therapy is therefore not recommended.

Author	Year	No. of patients	Regime	Mortality (%)	Survival (%)
Fernandez	1993	43	CTRT	7	34.5Y
Huff	1993	68	CTRT	7	51.2Y
Jones	1997	66	CTRT	11	32.3Y
Wright	1997	40	CTRT	10	61.2Y

CT, chemotherapy; RT, radiotherapy; Y, years.

Table 8 Results of induction therapy for esophageal cancer

Adjuvant chemotherapy

Two recent randomized trials have suggested a significant survival advantage to induction chemoradiotherapy followed by esophagectomy for esophageal carcinoma. Walsh reported on a randomized trial that involved preoperative treatment with 5-fluorouracil and cisplatin with radiation therapy followed by esophagectomy as compared to esophagectomy alone in patients with adenocarcinoma. One of the criticisms of this trial has been the lack of routine pretreatment staging with CT scans of the chest and abdomen, possibly leading to imbalance in the treatment arms. The surgery-alone arm did very poorly, especially in comparison to modern American series, whereas the induction arm had a 3-year survival of 32 per cent as compared to 6 per cent with surgery alone; this difference was statistically significant. Interestingly, the survival difference only became statistically significant at 3 years. Urba reported (in an abstract for the American Society for Clinical Oncology) on the University of Michigan trial. One hundred patients were randomized to transhiatal esophagectomy alone or to induction therapy followed by transhiatal esophagectomy. Induction therapy was with cisplatin, vinblastine and 5-fluorouracil, and twice-daily radiation therapy. Seventy-five per cent of the patients had adenocarcinoma. There was no difference in survival at 2 years, but at 3 years the survival was doubled in the induction therapy group (32 as compared to 15 per cent); this difference was statistically significant. Numerous phase II trials of induction therapy for esophageal cancer carried out in a single institution have been reported; their results seem to show improved survival when compared to institutional historic controls, but historic comparisons are not statistically valid. In general, induction therapy is safe and usually results in about a 25 per cent complete response rate at the pathologic level if a combination of chemotherapy and radiotherapy is given. Almost all series show a survival advantage if patients are converted to a pathologic complete response. Induction therapy does prolong total treatment time for the cancer and occasionally results in severe complications or death. Surgical complications do not appear to be greatly increased by induction treatment. The optimal induction program is unknown. A preliminary (abstract) report by Kelsen of the intergroup trial of induction chemotherapy compared with surgery alone for resectable carcinoma of the esophagus has appeared. No radiation was used. Induction was with cisplatin and 5-fluorouracil for three cycles. Fifty-five per cent of the patients had adenocarcinoma. At 2 years there was no survival advantage to the induction chemotherapy. Late results will be very important as induction therapy will only improve survival if it helps control micrometastatic disease, which will not usually present in the first year or two but later. The full results of this trial are awaited with great interest. An intergroup trial that has just opened again involves induction chemoradiotherapy with cisplatin and 5-fluorouracil along with 50 Gy of radiation therapy.

Wright recently completed a phase I/II induction trial centered around the use of Paclitaxel and twice-daily, high-dose radiation therapy in an attempt to maximize complete response rates. Paclitaxel is an exciting new chemotherapeutic agent with high activity in both adenocarcinoma and squamous-cell carcinoma of the esophagus, as reported by Ajani. Response rates are higher with Paclitaxel than with cisplatin alone. Forty patients with esophageal cancer were treated with a combination of cisplatin, 5-fluorouracil, and Paclitaxel with concurrent hyperfractionated radiation therapy (45 Gy to the mediastinum, 58.5 Gy to the tumor). Eighty per cent of the patients had adenocarcinoma. Seventy per cent of the patients had locally advanced tumors (T3, stage IIIB, or greater). The most common toxicity was severe esophagitis requiring nutritional support. The maximum tolerated dose of Paclitaxel in this regimen was 100 mg/m^2. Two patients died during induction treatment. Ninety per cent of the patients underwent resection. The median length of hospital stay following esophagectomy was 10 days. Two patients died after esophagectomy complications. Fourteen of 36 patients (39 per cent) had a pathologic complete response. Patients who received all the prescribed chemotherapy had a higher complete response rate (50 per cent) at the pathologic level than did patients who required dose reduction (17 per cent). The 2-year survival rate was 61 per cent. The high complete response rate at the pathologic level in this study was encouraging, but the toxicity was substantial. Only long-term follow-up will determine whether the substantial toxicity was worth the cost of the induction treatment.

Palliation of dysphagia

Dysphagia is the primary clinical symptom of patients with locally advanced esophageal carcinoma requiring palliation. Although palliative resection is recommended by many, it is associated with higher postoperative morbidity and mortality than in those patients who undergo ‘curative' esophagectomies. Esophageal bypass is considered by some surgeons to provide appreciable palliation of dysphagia in those for whom resection is not possible. Morbidity rates of 50 per cent and mortality rates of 25 to 30 per cent are commonly reported. Median survival following palliative bypass is usually less than 6 months; therefore most surgeons have abandoned this approach. Use of the neodymium–yttrium aluminum garnet laser may relieve dysphagia, but several treatment sessions are usually necessary to restore an adequate lumen. Repeat treatments are not usually necessary as patients usually die before the tumor regrows. There is a low incidence of perforation with this approach and the ability to swallow is usually quickly restored.

There has been a renewed interest in esophageal bypass tubes. The older rigid tubes, usually inserted by open techniques, were associated with high mortality rates and frequent complications. The newer Silastic tubes are inserted by dilating the esophagus and pushing the tube in place. The incidence of complications, perforation, and fatality have been low in most series. The distal end is ideally situated above the gastroesophageal junction to avoid reflux. These prostheses allow swallowing of liquids and some soft solids. This approach allows quick restitution of swallowing and satisfactory palliation. Average median survival is 3 to 6 months following tube insertion. Expandable metallic stents are now available for endoscopic insertion and appear to be at least as effective as plastic tubes.

Unusual esophageal neoplasms

Small-cell carcinoma

About 150 cases of esophageal small-cell carcinoma with striking morphologic and biologic similarities to small-cell carcinoma of the lung have been reported. These tumors contain cells with characteristic cytoplasmic granules (neurosecretory type) that are argyrophilic. Because both the lung and esophagus are derived from the embryonic foregut, argyrophilic neurosecretory cells can be expected in the esophagus.

The clinical course is remarkably similar to that of small-cell carcinoma of the lung. The tumors are typically fungating and polypoid with surface ulcerations, and are commonly found in the middle and lower third. Widespread metastases are common. Multimodality therapy with chemotherapy and radiation is the treatment of choice, with little role for surgery. There are a few long-term survivors.

Melanoma

Melanoblasts occur in the esophageal mucosa and are typically scattered throughout the esophagus. Oesophageal melanosis is a benign condition seen in approximately 5 per cent of people with apparently normal esophaguses, but is associated with malignant melanoma in approximately one-third of all reported cases. Melanoma of the esophagus is a rare tumor and fewer than 150 cases have been reported. The average age is about 60 years and men are mostly affected. Melanomas typically are polypoid and often can grow to a large size. The tumor may be black, brown, or grey; most are ulcerated. Patients present with dysphagia and when first seen appear to have localized, resectable disease. Regional lymph-node metastases are rather common. Surgical resection is the treatment of choice, although long-term survival is rare. Chemotherapy and radiation therapy have not proved effective.

Leiomyosarcoma

Leiomyosarcoma is the most common sarcomatous tumor of the esophagus. Men mostly are affected and the typical age at presentation is 60 years. Most occur in the middle or lower thirds of the esophagus and have a characteristic pedunculated appearance on radiologic examination. The tumor often achieves a large size before obstructive symptoms occur. Myosarcomas are typically well localized and resectable, and in general do not invade adjacent mediastinal structures until later in their course. Surgical resection is the treatment of choice and has produced long-term survivors. Although radiation therapy may provide palliation, surgery appears to offer superior cure rates.

Benign tumors of the oesophagus

Leiomyoma

Autopsy studies confirm that these are extremely rare tumors. Reported occurrences have ranged from 1/1000 to 2/36 000 autopsy examinations. Leiomyomas of the esophagus represent about 10 per cent of all gastrointestinal leiomyomas. They rarely occur in the cervical region, but are equally distributed between the middle and lower thirds. Less than 5 per cent of the leiomyomas are multiple. Leiomyomas are usually found in men (sex ratio 2:1 male:female). There is a wide age distribution between 20 and 70 years. The tumors are usually intramural and well circumscribed. Unusual configurations with a horseshoe pattern are not uncommon. Leiomyomas rarely undergo malignant transformation. Their growth rate appears rather slow as the duration of reported symptoms often can be rather long. Dysphagia and odynophagia are the most common presenting symptoms. They are frequently found as incidental findings during assessment for other gastrointestinal complaints. A barium esophagram typically shows a smooth, semilunar defect with sharp borders and an intact mucosa. Horseshoe leiomyomas characteristically produce obstruction. Esophagoscopy should be performed, but typically only a bulging mass is seen. The overlying mucosa is usually intact, and should not be biopsied. All symptomatic leiomyomas should be removed. Small, asymptomatic tumors should be left in place. For mid-third lesions a right thoracotomy is chosen, whereas lower-third lesions are best approached through a left thoracotomy. The tumor can be enucleated from the esophageal wall. Care must be taken to avoid entering the esophageal mucosa. The incision in the esophageal musculature is closed with interrupted sutures.

Benign polyp

Benign polyps of the esophagus are rare, but are notable for their sometimes dramatic presentation with regurgitation of the polyp into the mouth, which can produce airway obstruction. They are usually solitary, and are frequently rather long and cylindrical. Because of constant peristaltic action, elongation frequently occurs, which then can permit regurgitation. Marked dilatation of the esophagus can occur because of gradual enlargement of the polypoid mass. Polyps are typically composed of vascular, fibroblastic connective tissue, covered by normal mucosa. Dysphagia is the predominant symptom, but occasionally the patient will relate a history of intermittent regurgitation of a mass into the mouth. The barium esophagram is often diagnostic, showing a long intraluminal filling defect with a rounded lower border. Esophagoscopy confirms the results of barium studies. These lesions usually occur in older men and usually arise from the cervical esophagus. Treatment is always surgical, both to relieve symptoms and to rule out malignancy. Occasionally, small polyps have a base that can be readily seen endoscopically so that the stalk can be divided and the base cauterized. Larger polyps require esophagotomy on the side opposite to the tumor: this can be done through a cervical approach for high lesions or upper sternotomy or lateral thoracotomy for lower lesions; the esophagus is closed in layers.

Further reading

Aikiyama H *et al.* Principles of surgical treatment for carcinoma of the esophagus. Analysis of lymph node involvement. *Annals of Surgery* 1981; **194**: 438–46.

Altorki NK *et al.* High-grade dysplasia in the columnar-lined esophagus. *American Journal of Surgery* 1991; **161**: 97–100. [Early report dealing with incidence of invasive cancer in high-grade dysplasia.]

Blot WJ *et al.* Rising incidence of adenocarcinoma of the esophagus and gastric cardia. *Journal of the American Medical Association* 1991; **265**: 1287–9.

Burt M *et al.* Malignant esophagorespiratory fistula: management options and survival. *Annals of Thoracic Surgery* 1991; **52**: 1222–9. [Large series documenting the grim prognosis in patients with this serious complication.]

Cusumano A *et al.* Push-through intubation: effective palliation in 409 patients with cancer of the esophagus. *Annals of Thoracic Surgery* 1992; **53**: 1006–9.

Daniel TM *et al.* Transhiatal esophagectomy: a safe alternative for selected patients. *Annals of Thoracic Surgery* 1992; **54**: 686–90.

Delarue NC, Wilkins E Wassel, Wong J, ed. *International trends in general thoracic surgery*, Vol. 4. *Esophageal cancer*. Mosby, St. Louis, 1988. [Excellent review of topic from worldwide authorities.]

Earlom R, Cunha-Melo JR. Oesophageal squamous cell carcinoma. I. A critical review of surgery. *British Journal of Surgery* 1980; **67**: 381–90.

Earlom R, Cunha-Melo JR. Oesophageal squamous cell carcinoma. II. A critical review of radiotherapy. *British Journal of Surgery* 1980; **67**: 457–61. [Excellent cumulative review of English literature.]

Ferguson MK, Naunheim KS. Resection for Barrett's mucosa with high-grade dysplasia: implications for prophylactic photodynamic therapy. *Journal of Thoracic and Cardiovascular Surgery* 1997; **114**: 824–9.

Forastiere AA *et al.* Preoperative chemoradiation followed by transhiatal esophagectomy for carcinoma of the esophagus: final report. *Journal of Clinical Oncology* 1993; **11**: 1118–23. [Excellent report on the use of neoadjuvant therapy.]

Gammon MD *et al.* Tobacco, alcohol and socioeconomic status and adenocarcinoma of the esophagus and cardia. *Journal of the National Cancer Institute* 1997; **89**: 1277–84.

Gignoux M *et al.* The value of pre-operative radiotherapy in esophageal cancer: results of a study of the EORTC. *World Journal of Surgery* 1987; **11**: 426–32.

Hancock SL, Glatstain E. Radiation therapy of esophageal cancer. *Seminars in Oncology* 1984; **11**: 144–58.

Hankins JR *et al.* Carcinoma of the esophagus. A comparison of the results of transhiatal versus transthoracic resection. *Annals of Thoracic Surgery* 1989; **47**: 700–5.

Heitmiller RF *et al.* Barrett's esophagus with high-grade dysplasia: an indication for prophylactic esophagectomy. *Annals of Surgery* 1996; **224**: 66–71. [Excellent report dealing with invasive cancer in high-grade dysplasia.]

Hoff SJ *et al.* Preliminary results with neoadjuvant therapy in resection for esophageal carcinoma. *Annals of Thoracic Surgery* 1993; **56**: 282–7.

Jones DR *et al.* Induction chemoradiotherapy followed by esophagectomy in patients with carcinoma of the esophagus. *Annals of Thoracic Surgery* 1997; **64**: 185–92.

Launois B *et al.* Pre-operative radiotherapy for carcinoma of the esophagus. *Surgery, Gynecology Obstetrics* 1981; **153**: 690–2.

Lewis I. The surgical treatment of carcinoma of the oesophagus. *British Journal of Surgery* 1946; **34**: 18–31.

Mathisen DJ. *Seminars in Thoracic and Cardiovascular Surgery—Esophagus*. 1992; **4**.

Mathisen DJ *et al.* Transthoracic esophagectomy: a safe approach to carcinoma of the esophagus. *Annals of Thoracic Surgery* 1988; **45**: 137–43. [Report dealing with anastomotic technique resulting in zero leak rate.]

Moon BC, Woolfson IK, Mercer CD. Neodymium: yttrium–aluminum–garnet laser vaporization for palliation of obstructing esophageal carcinoma. *Journal of Thoracic and Cardiovascular Surgery* 1989; **98**: 11–15.

Muller JM *et al.* Surgical therapy of oesophageal carcinoma. *British Journal of Surgery* 1990; **77**: 845–57.

Pera M *et al.* Barrett's esophagus with high-grade dysplasia: an indication for esophagectomy? *Annals of Thoracic Surgery* 1992; **54**: 199–204. [Mayo Clinic experience with invasive cancer in high-grade dysplasia.]

Rice TW *et al.* Esophageal carcinoma: esophageal ultrasound assessment of pre-operative chemotherapy. *Annals of Thoracic Surgery* 1992; **53**: 972–7. [Early report on use of ultrasound in staging esophageal cancer.]

Rice TW *et al.* Surgical management of high-grade dysplasia in Barrett's esophagus. *American Journal of Gastroenterology* 1993; **88**: 1832–6.

Rice TW *et al.* Esophageal carcinoma: depth of tumor invasion is predictive of regional lymph node status. *Annals of Thoracic Surgery,* 1998; **65**: 787–92.

Roth, JA, Ruckdeschel JC, Weisenburger TH, ed. *Thoracic oncology*. Saunders, Philadelphia, 1989.

Shahian DM *et al.* Transthoracic versus extrathoracic esophagectomy: mortality, morbidity and long term survival. *Annals of Thoracic Surgery* 1986; **41**: 237–46. [Retrospective comparison of morbidity and mortality of two common approaches to esophageal cancer.]

Skinner DB. *En bloc* resection for neoplasms of the esophagus and cardia. *Journal of Thoracic and Cardiovascular Surgery* 1983; **85**: 59–71. [Early report on radical esophagectomy.]

Steiger Z *et al.* Eradication and palliation of squamous cell carcinoma of the esophagus with chemotherapy, radiotherapy and surgical therapy. *Journal of Thoracic and Cardiovascular Surgery* 1981; **82**: 713–19.

Streitz JM, Jr *et al.* Adenocarcinoma in Barrett's esophagus: clinicopathologic study of 65 cases. *Annals of Surgery* 1991; **213**: 122–5.

Streitz JM, Williamson WA, Ellis FH. Current concepts concerning the nature and treatment of Barrett's esophagus and its complications. *Annals of Thoracic Surgery* 1992; **54**: 586–91.

Sweet RH. Surgical management of carcinoma of the midthoracic esophagus. *New England Journal of Medicine* 1945; **233**: 1–7. [Classical report of historical significance.]

Turner GG. Excision of thoracic oesophagus for carcinoma, with construction of extra-thoracic gullet. *Lancet*, 1933; **ii**: 1315–16.

Walsh TN *et al.* A comparison of multimodal therapy and surgery for esophageal adenocarcinoma. *New England Journal of Medicine* 1996; **335**: 462–7. [Randomized study of neoadjuvant therapy versus surgery alone.]

Wright CD *et al.* Induction therapy for esophageal cancer with Paclitaxel and hyperfractionated radiotherapy: a Phase I and II study. *Journal of Thoracic and Cardiovascular Surgery* 1997; **114**: 811–16. [Massachusetts General Hospital pilot study of neoadjuvant chemotherapy and high-dose irradiation for esophageal cancer.]

23.1 Peptic ulcer—stomach and duodenum

Alan G. Johnson

[Historical perspective](#)
[Worldwide prevalence and incidence](#)
[Classification of peptic ulcers](#)
[Pathophysiology of peptic ulceration](#)
[Physiology of gastric secretion](#)
[Mucosal defence](#)
[The role of Helicobacter in the pathogenesis of peptic ulcer](#)
[Duodenal juice and gastric mucosal defence](#)
[The natural history of untreated peptic ulcer](#)
[Symptoms and signs of peptic ulcer](#)
[Diagnosis](#)
[Treatment of uncomplicated peptic ulcers](#)
[Elimination of H. *pylori*](#)
[Prevention of ‘stress' ulcers](#)
[Should a patient be given treatment for suspected peptic ulcer disease without a diagnosis being made?](#)
[Indications for elective surgery](#)
[Which elective operation for duodenal ulcer?](#)
[Elective operations for gastric ulcer](#)
[Complications of peptic ulcer](#)
[Perforation](#)
[Haemorrhage](#)
[Stenosis](#)
[Fistulas](#)
[Malignant change in peptic ulcers](#)
[Surgical techniques](#)
[Billroth I partial gastrectomy](#)
[Pylorus-preserving gastrectomy (Maki)](#)
[Billroth II (Polya) partial gastrectomy](#)
[Vagotomy and antrectomy](#)
[Truncal vagotomy and pyloroplasty or gastrojejunostomy](#)
[Proximal gastric vagotomy](#)
[Postoperative management](#)
[Alcohol](#)
[Postvagotomy and postgastrectomy syndromes](#)
[Dumping](#)
[Diarrhoea](#)
[Bile ‘reflux'](#)
[Malignant change](#)
[Delayed gastric emptying](#)
[Stomal ulcer](#)
[Nutritional problems](#)
[Conclusion](#)
[Further reading](#)

Historical perspective

The first description of gastric ulcer is attributed to the Italian physician Marcello Donati in 1586, and the first case of perforated gastric ulcer was recorded by Christopher Rawlinson in England in 1727. Duodenal ulcer was first described by Georg Hamberger in Germany in 1746, and Jacopo Penada from Italy recorded a perforation in 1793. In the second part of the nineteenth century, duodenal ulcers were still rare, but benign gastric ulcer was increasingly recognized. Once general anaesthesia (1846) and antisepsis (1867) had made surgery acceptable and safe, surgeons became involved in the treatment of peptic ulcer. In 1881, Rydigier performed a gastric resection for a benign gastric ulcer and the patient survived. However, he received wide criticism, and the Editor who published the paper, which was entitled ‘The first gastric resection for gastric ulcer', added the footnote ‘and hopefully the last'! In 1892, Ludwig Heusner in Germany successfully sutured a perforated gastric ulcer. It was probably not until 1893 that the first operation for duodenal ulcer was performed, a gastrojejunostomy by Codivilla. This rapidly became the treatment of choice for both duodenal and gastric ulcers. On the other hand, 20 years later, von Haberer and Finsterer were championing extensive gastrectomies and their view prevailed when an increasing number of gastrojejunal ulcers were seen after simple gastroenterostomy.

Resection held sway until the 1940s, when Dragstedt re-introduced vagotomy, which had, in fact, been used by Latarjet in 1922 and Skiassi in 1925; but Dragstedt had the experimental evidence that placed the operation on a sound physiological footing. Over the following 40 years, vagotomy became refined and prospective randomized trials were performed to determine the best operation. However, until H_2-receptor antagonists became available in 1976, surgery was the only way to alter the natural history of the disease, and even then the antagonists, and the later proton-pump inhibitors, were only effective while they were being taken. The major change in approach to the treatment of peptic ulcer was the recognition in 1982 that *Campylobacter* (*Helicobacter*) *pylori* was an important factor in the pathogenesis, and that the natural history of the disease could be altered by eliminating this micro-organism.

So surgery was the definitive, elective treatment of peptic ulcer for exactly 100 years. Today the surgeon's major contribution is in the treatment of life-threatening complications.

Worldwide prevalence and incidence

It is difficult to establish the true prevalence of peptic ulcer, because no one has gastroscoped a large, representative sample of the population with and without dyspeptic symptoms. Probably about 10 per cent of adults in the United States of America and Europe suffer from an ulcer at some time in their lives. Among individuals presenting to open-access endoscopy clinics with symptoms, between 10 and 19 per cent are found to have peptic ulcer. The ratio of duodenal ulcer to gastric ulcer is 4:1. However, there is evidence that the incidence and severity of the disease is decreasing in the West, and that this decrease started before the introduction of powerful acid-suppressing drugs. In Sweden, elective hospital admissions for peptic ulcer diminished from 64 per 100 000 of the population in 1950 to 11 per 100 000 in 1986. In contrast, for the 10 years after the introduction of H_2-receptor antagonists in the United Kingdom, whereas the elective admissions diminished considerably, emergency admissions for bleeding or perforation remained the same overall but actually increased in males and females over 65 years of age (Fig. 1). In the early 1990s, there were still 40 000 admissions and 4700 deaths annually in the United Kingdom. The reduction in pyloric stenosis is notable and may be due to earlier detection and effective treatment, by whatever means. However, when fibreoptic endoscopy replaced the barium meal as the main method of diagnosis, many small and acute ulcers were seen that were not recognized before.

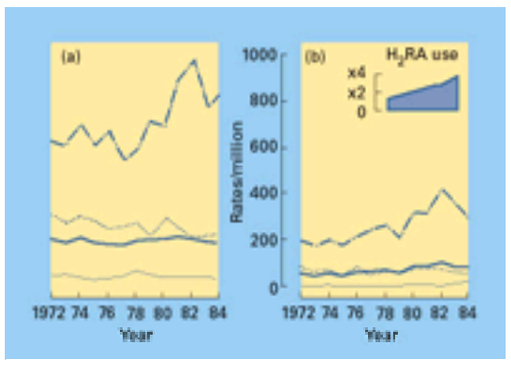


Fig. 1. Admission rates for (a) men and (b) women with bleeding duodenal ulcer, subdivided by age, in the Trent Regional Health Authority, 1972–84. The total volume

of H_2-receptor antagonists (H_2RA) used is shown only in the graph on the right, and is not subdivided by sex. – • – > 65 years; … 35–64 years; – < 35 years; — all ages.

In Eastern and African countries the pattern varies and is changing. In Singapore, for example, the changes vary in different racial groups. In Japan, where gastric cancer is so common, benign peptic ulcer has a lower prevalence than in the West. In some African countries there are high local prevalences due to specific dietary habits, and migration to the towns in many developing countries has been accompanied by an increase, or increased diagnosis, of peptic ulceration.

Classification of peptic ulcers

Ulcers are classified either by their state, acute or chronic, or by their site (Table 1). Acute ulcers are those that have not had time to produce any fibrotic reaction around them and they may be very superficial. When they are multiple and widespread they are often referred to as ‘erosions'. A useful classification of gastric ulcers is based on that of Johnson (Table 2). Nearly all chronic duodenal ulcers occur in the bulb (or ‘cap'), and those occupying a more distal site raise the suspicion of Zollinger–Ellison syndrome (or an ulcerating carcinoma of the pancreas). Acute ulcers tend to perforate easily because there is no fibrotic reaction, and they may bleed from superficial mucosal vessels rather than from the major vessels, such as the splenic or pancreaticoduodenal arteries, that are eroded by chronic ulcers.

Gastric
Duodenal:
Bulbar
Post bulbar
Combined:
Gastric type II
Zollinger–Ellison syndrome
Anastomotic

Table 1 Classification of gastroduodenal peptic ulcer by site

Acute superficial:
Single or multiple (‘erosions’)
Chronic:
Type I, usually lesser curve
Type II, combined with duodenal ulcer
Type III, prepyloric
Type IV, proximal stomach ‘2cm from oesophageal junction

Table 2 Classification of benign gastric ulcers (after Johnson, 1957)

The classification is important because ulcers differ in their pathogenesis and their management. The management of a patient with multiple, recurrent, aggressive chronic ulcers from Zollinger–Ellison syndrome is very different from that of a patient with acute gastric erosions from non-steroidal anti-inflammatory drugs.

Pathophysiology of peptic ulceration

‘No acid, no ulcer' is an old adage, but modern drug therapy and surgery have shown it is true. When acid secretion is effectively suppressed, nearly all ulcers heal and do not recur so long as the suppression continues. It was William Prout in 1824 who discovered that the acid in the stomach was hydrochloric.

Physiology of gastric secretion

Acid secretion

Acid is secreted from parietal cells, which are found in the fundus and body of the stomach, whereas the antrum, the muscular ‘pump', contains gastrin- and mucus-secreting cells. It is this anatomical demarcation that makes parietal-cell (proximal gastric) vagotomy without drainage possible. The secretion of acid is by stimuli originating from (a) the brain, (b) mechanical (distension), and (c) chemical stimulation of the stomach. Secretion from these cells is regulated by neural–hormonal and paracrine pathways that involve (a) direct cholinergic stimulation of acid secretion and (b) cholinergic inhibition of somatostatin secretion, (c) direct gastrin stimulation of parietal cells, and (d) gastrin-induced stimulation of histamine release from enterochromaffin-like cells. In the antrum, cholinergic nerves and gastrin-releasing peptide stimulate gastrin secretion, which is accompanied by cholinergic inhibition of somatostatin. These mechanisms together both stimulate and remove inhibitory influences on gastrin release and thus parietal-cell activity. There are also calcium sensory receptors on gastrin cells, which may explain the increased gastric acid secretion in patients with a high serum calcium (hyperparathyroidism). Figure 2 summarizes the main mechanisms.

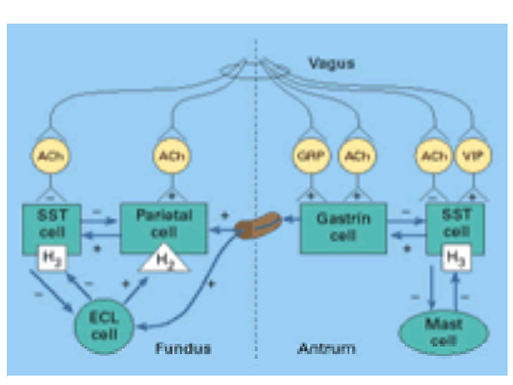


Fig. 2. Model illustrating the neural, paracrine, and hormonal regulation of acid secretion in the fundus, and gastrin secretion in the antrum of the stomach. ACh, acetylcholine; GRP, gastrin-releasing peptide; VIP, vasoactive intestinal peptide; SST, somatostatin; ECL, enterochromaffin-like cell.

Drugs act by blocking either the H_2 receptors on the parietal cells (H_2-receptor antagonists) or the secretory mechanism within the cells (proton-pump inhibitors). Vagotomy blocks the cholinergic stimulation of the parietal cells, which desensitizes them to the other stimuli. If the antrum is also vagotomized, the secretion of

somatostatin is increased, which blocks gastrin release and so reduces acid secretion by a second mechanism.

Pepsinogen secretion

Pepsinogen, the precursor of pepsin, is secreted via the gastric chief cells and is stimulated by two pathways, one involving hormones such as vasoactive intestinal peptide (secretin) and prostaglandins (E_2), the other involving carbachol, cholecystokinin, and other peptide hormones. It is most active in protein digestion between pH 2 and 4.

Mucosal defence

Defence involves resistance and repair, and if small ulcers occur frequently, the rapid repair mechanism is crucial in preventing them becoming chronic, perforating, or bleeding. The gastric mucosa is highly efficient at protecting itself against its luminal acid. Growth factors such as epidermal and platelet-derived promote ulcer healing. The stomach also has the ability to adapt to harmful stimuli whereby initial exposure to a damaging agent, such as aspirin, makes the mucosa more resistant to subsequent doses. Possibly transferring growth factor-a plays a part in this process.

The duodenum has efficient neutralizing mechanisms if acid from the stomach is delivered to it at a reasonable rate. However, if gastric emptying is particularly rapid, gastric secretion particularly high, or duodenal mixing of bile and pancreatic juice defective, then the defence can be overcome: hence most duodenal ulcers are associated with an increased secretion of gastric acid. The proton-pump inhibitors omeprazole and lansoprazole enhance duodenal bicarbonate secretion as well as reducing gastric acid.

Mucosal blood flow

Increase in mucosal blood flow protects the mucosa, provides oxygen and nutriments to the epithelium, and dilutes and neutralizes back diffusion of acid and toxins. If there is acidaemia centrally, then this neutralizing mechanism may fail.

Non-steroidal anti-inflammatory drugs

Ulceration by non-steroidal anti-inflammatory drugs remains a problem. In the United States of America, 70 million prescriptions for non-steroidals are given annually, and there are 76 000 related hospital admissions and 7600 deaths. The discovery of two isoforms of cyclo-oxygenase has led to the hope that non-steroidals that only inhibit cyclo-oxygenase 2 will be ‘gastrointestinal safe'. At present, patients with a history of peptic ulcers should not be started on these drugs without added protection, and they should be stopped if ulcers develop.

The role of Helicobacter in the pathogenesis of peptic ulcer

How does Helicobacter pylori influence acid secretion?

The precise pathway is not known, but one possible mechanism is that it decreases antral somatostatin secretion (which inhibits gastrin release), leading to prolonged hypergastrinaemia. It may also act by synthesizing a selective H_3-receptor agonist. However, patients with duodenal ulcer have a greater sensitivity to gastrin than *H. pylori*-affected controls and also an increased maximum acid secretory capacity.

***H. pylori* and mucosal defence: microcirculation**

Extracts of *H. pylori* from human infection produce platelet activation and aggregation in the mucosal microcirculation of rats, with leakage of albumin. These changes would predispose to ulceration and delay healing. It is not yet clear in what other ways, if any, *H. pylori* inhibits defence mechanisms. The publication of the entire genome of *H. pylori* in 1997 has opened new possibilities for understanding and combating the infection.

Duodenal juice and gastric mucosal defence

For surgeons the effect of duodenal juice on the gastric mucosa is important. Not only does this action occur before surgery in certain conditions, but gastric operations such as gastroenterostomy, partial gastrectomy, and vagotomy and pyloroplasty all increase the duodenal reflux into the stomach. For many years, duodenal juice has been implicated in damaging gastric mucosa in such a way as to make it more susceptible to acid damage. The evidence comes from clinical studies, which show a higher concentration of bile and pancreatic juice in the stomach of patients with gastric ulcer. Exposure time is obviously important and delayed gastric emptying compounds the problem of duodenal gastric reflux; there are a number of different factors leading to the final ulceration.

The risk factors for gastric ulceration include smoking, excessive alcohol, non-steroidal anti-inflammatory drugs, and corticosteroids. Additional risk factors for duodenal ulcer are liver cirrhosis, atheroma (particularly coronary arterial disease), renal transplantation, and other states of immune suppression.

The natural history of untreated peptic ulcer

In the days when the only treatment for ulcers was to try to neutralize the acid after it had been secreted, the natural history of the disease could be observed. Peptic ulcer is a cyclical disease with a sequence of healing and exacerbation over the years. In some countries, ulcers tend to become reactivated in the winter months, but chronic ulcer can be present for years. The concept of a ‘burnt out' ulcer is attractive, but even an ulcer that has been quiescent for many years can be reactivated at a time of stress, such as a surgical operation or new drug therapy, and it is important that prophylaxis be started when a patient with a past history of ulceration is undergoing such stress. It was always thought that acid secretion diminished with age, and the most recent work has confirmed this view, once other diseases have been excluded. However, the important message is that any untreated ulcer is potentially lethal from perforation or bleeding, and it is impossible to predict in any individual patient when this might happen. The only safe policy is to keep ulcers healed.

Before the importance of *Helicobacter* was appreciated, clinicians had noted that bismuth-containing ulcer treatments reduced the recurrence rate after healing. During the years when H_2-receptor antagonists were the mainstay of treatment, recurrence was high after stopping treatment (80 per cent within 18 months) and did not depend on how long the treatment had been used. The only two treatments that altered the natural history were continuous, full-dose, H_2-receptor antagonists or surgery (e.g. vagotomy).

Symptoms and signs of peptic ulcer

Many ulcers are symptomless and a severe complication may be the first sign of their presence. It used to be thought that peptic ulcers gave very characteristic symptoms, and that duodenal and gastric ulcers could be distinguished by the timing of the pain in relation to food, duodenal ulcer being characterized by pain when starving, relieved by meals, whereas gastric ulcer pain was brought on by food, so that the patient avoided meals and lost weight. There is some truth in these assertions, but symptoms are often not so clear-cut. When all the many symptoms attributed to duodenal ulcer are carefully analysed, the most discriminating (but by no means 100 per cent) is the patient being woken with localized epigastric pain in the middle of the night and it being relieved by taking antacids or milk. Various forms of dyspepsia, such as fullness and slight nausea, are common, but vomiting usually only occurs if there is obstruction. In gastric ulcer the symptoms occur sooner after food, and vomiting and weight loss are more common, whereas they are rare with uncomplicated duodenal ulcer. Continuing pain radiating to the back suggests a penetrating ulcer, but it must be differentiated from carcinoma of the pancreas. A stomal ulcer at a gastrojejunostomy will give a similar type of pain felt in the periumbilical region rather than the epigastrium, because the ulcer affects the mid- rather than the foregut.

Symptoms may be confusing because of the coexistence of other digestive conditions such as oesophageal reflux, with or without hiatus hernia, and gallstones. Anaemia is an important symptom because both types of ulcers can bleed slowly. The other, most likely cause of anaemia without rectal bleeding or melaena is carcinoma of the caecum or right colon. The most important differential diagnosis of peptic ulcer is carcinoma of the stomach. Patients over the age of 50 years with new dyspepsia must be investigated by endoscopy as soon as possible. Perhaps the most useful, practical guide in younger patients is that a man with intermittent upper abdominal pain is likely to have a duodenal ulcer, whereas a woman is likely to have gallstones. In older people, the sex difference for peptic ulcer is not so marked.

Physical examination is useful for confirmation if there is localized tenderness in the epigastrium, and to exclude other conditions and detect clinical anaemia. It does

not, however, make the diagnosis.

Diagnosis

Fibreoptic endoscopy is the principal method of diagnosis, as it enables biopsies to be taken to exclude malignancy in gastric ulcers, and to do a rapid urease test or histological examination for *Helicobacter*. The ^{14}C breath test is used to confirm eradication to avoid a further gastroscopy.

A double-contrast barium meal is rarely used as a primary means of diagnosis, but is useful in patients with a morbid fear of endoscopy. On open-access endoscopy for ‘dyspepsia', about 12 per cent of patients have duodenal ulcer and 3 per cent gastric ulcer, but the numbers and ratios vary with age. Every gastric ulcer should be suspected of being malignant until proved benign by multiple biopsies and complete endoscopic healing.

A frequent clinical problem is the patient who is referred for gastroscopy having been started on acid-inhibitory drugs several weeks before. If no ulcer is seen, it is often impossible to know whether there was an acute ulcer that has healed, or whether the diagnosis was not peptic ulceration. In these circumstances, it is best to stop the treatment and arrange to endoscope the patient as soon as the pain recurs. Ideally, patients should not be put on powerful acid-reducing drugs, which heal ulcers rapidly, before a diagnosis is made: otherwise they may be condemned to long periods of drug therapy or to *Helicobacter* elimination for the wrong reasons.

Gastric acid secretion tests

These have no place in the routine diagnosis of ulcers, but a serum gastrin of several hundred pmol (normal < 40 pmol/l) is strongly suggestive of a gastrinoma in Zollinger–Ellison syndrome, provided that the patient is not achlorhydric with pernicious anaemia (in which case any ulcer would be malignant), is not on acid-inhibiting drugs, nor has had a vagotomy. H_2-receptor antagonists should be stopped for 48 h and proton-pump inhibitors for at least 1 week before a sample for serum gastrin is taken.

Treatment of uncomplicated peptic ulcers

The aim of treatment is (i) to relieve symptoms; (ii) to heal the ulcer as quickly as possible; (iii) to prevent recurrence; (iv) to prevent complications.

A full dose of H_2-receptor antagonists will heal about 90 per cent of duodenal ulcers in 8 weeks, and proton-pump inhibitors will have the same effect in about 4 weeks. Pain relief occurs in a few days, there being little difference between the two groups of drugs at 2 weeks. Rapid pain relief has led to patients not completing the full healing course, or using powerful acid-suppressing drugs as ‘indigestion tablets', taking them when they have an attack of pain. It is important to stress that the risk of complications is not over until the ulcer is completely healed and stays healed. Proton-pump inhibitors are the most effective drugs in healing benign gastric ulcers, with a healing rate of 70 to 80 per cent at 4 weeks and 80 to 90 per cent at 8 weeks.

Elimination of H. *pylori*

It is important to eliminate, not just suppress, *H. pylori*, and many regimens have been used and compared. The classic triple therapy of ‘bismuth', metronidazole, and amoxycillin (amoxicillin) for 2 weeks eliminates *H. pylori* in 90 per cent of patients. However, this was not very palatable and new regimens have superseded it, based on ranitidine–bismuth citrate or proton-pump inhibitors with an antibiotic. A proton-pump inhibitor, with metronidazole 400 mg and amoxycillin 500 mg three times a day, results in eradication rates of 85 to 95 per cent in 1 week. If patients are allergic to amoxycillin, clarithromycin may be substituted; this antibiotic combined with ranitidine–bismuth citrate heals 93 per cent of patients. For the 5 to 15 per cent that do not respond to the proton-pump inhibitor, bismuth should be added and the course extended for 2 weeks. Antibiotic resistance is an increasing problem, both primary resistance to metronidazole (41 per cent) and secondary resistance to clarithromycin (in 25 per cent of patients whose ulcer failed to heal on dual therapy). The elimination of *H. pylori* should be limited to patients with peptic ulceration: there is no evidence that universal testing and elimination is justified. Elimination of *H. pylori* may make gastro-oesophageal acid reflux worse.

Prevention of ‘stress' ulcers

Acute peptic ulcers are a complication of any serious trauma, particularly if the patient is in an intensive care unit. Prophylaxis with H_2-receptor antagonists (e.g. ranitidine) intravenously was associated with only 1.7 per cent clinically significant bleeds, but acid suppression leads to bacterial colonization of the stomach, and many intensive care units prefer mucosal protection with sucralfate. It is also important to avoid a systemic acidosis by careful monitoring. For those patients who cannot come off non-steroidal anti-inflammatory drugs, concurrent treatment with the prostaglandin E_1 analogue misoprostol, or with effective acid-suppressive drugs, helps to prevent ulcers and treat their complications. If a patient has presented with a bleed or perforation, these protective measures are essential. Eradication of *Helicobacter* before treatment with non-steroidals reduces the incidence of gastroduodenal ulceration, but this is not possible acutely.

Should a patient be given treatment for suspected peptic ulcer disease without a diagnosis being made?

The answer to this question is ‘no', because of the danger of masking a carcinoma in the older patient, and of submitting a younger patient to intensive drug treatment without knowing the diagnosis. However, young patients who relapse with exactly the same symptoms as when the diagnosis was first made may reasonably be given a second course of acid-suppressing drugs. In most circumstances, *H. pylori* would now have been treated to prevent recurrence.

Indications for elective surgery

With the introduction of H_2-receptor antagonists in 1976, the frequency of elective hospital admissions for peptic ulcer started to decline, but overall the frequency of emergency surgery stayed the same, and, as mentioned above, actually increased in elderly men and women over the following 7 years (Fig. 1). Elective surgery was the alternative to lifelong drug treatment or for the 10 per cent or so of patients whose ulcers failed to heal. With very few treatment failures with proton-pump inhibitors, and the ability to eliminate *H. pylori* in most cases, the need for elective surgery is even less. The present indications can be summarized as follows:

- The occasional patient whose ulcer fails to heal with medical treatment or who relapses frequently and is *H. pylori* negative.
- Prevention of recurrent complications when the patient is at high risk.
- Possibly as a protection in patients who cannot be without non-steroidal anti-inflammatory drugs.
- For financial reasons, either from the patient's or supplier's viewpoint: in some countries the patient has to pay for drugs but the operation is free, and, in many countries, surgery is considerably cheaper than long-term drug therapy.

Which elective operation for duodenal ulcer?

There is a variety of operations, all of which aim to produce a permanent reduction in gastric acid secretion: truncal or proximal gastric vagotomy, partial gastrectomy, or truncal vagotomy and antrectomy. The classic, Leeds–York, prospective, randomized trial for duodenal ulcer compared three of these, truncal vagotomy and drainage, subtotal gastrectomy, and vagotomy and antrectomy. It found that the recurrence rate was zero with the most radical operation, vagotomy and antrectomy, but this had significant side-effects, whereas the vagotomy and gastroenterostomy had the highest recurrence (4.2 per cent). However, the side-effects of dumping and diarrhoea were still significant, although in different proportions for the different operations, and surgeons realized that it was as much the drainage procedure as the vagotomy that produced these. The next logical step was to do a selective and then highly selective or proximal gastric vagotomy where just the parietal mass was denervated leaving the antral muscular pump to empty the stomach. Therefore, a drainage procedure was not required, as gastric emptying, particularly of solids, would be relatively normal. Prospective, randomized trials comparing this new procedure with truncal vagotomy and pyloroplasty for duodenal ulcers showed a similar recurrence rate (9 per cent) but significantly fewer side-effects after proximal gastric vagotomy. Surgeons have been obsessed, over the years, with the recurrence rate of operations, whereas a transient ulcer recurrence may be of no consequence. Much more important is persistence of ulcer and the life-threatening complications. If the recurrence of ulcers is classified in a modified Visik grading, then proximal gastric vagotomy has been extremely effective in the elective treatment of duodenal ulcer, with very few complications.

When proximal gastric vagotomy was a standard, long-term treatment for duodenal ulcer, it was being done before there was an effective medical treatment or as an alternative to long-term, full-dose H_2-receptor blockers. Although it was suggested that it was not so effective in patients whose ulcer had not healed with H_2-receptor blockers, further studies did not confirm this. Nowadays, operation is only being considered for a few patients with very resistant ulcers and it is not known whether this selected subgroup will respond as well to proximal gastric vagotomy. In addition, the new generation of gastric surgeons have little experience with proximal gastric vagotomy, which is an operation that is technically demanding if the vagotomy is to be complete. Therefore, unless the surgeon has had much experience in the past of proximal gastric vagotomy, the right operation is probably truncal vagotomy and antrectomy because, in the selected patients, it is absolutely essential that the ulcer is

healed and remains so. Surgeon and patient would have to accept the small but definite incidence of side-effects (see below).

Elective operations for gastric ulcer

The great difference between gastric and duodenal ulcers is the gastric ulcer's malignant potential. Therefore, whichever operation is contemplated, excision of the ulcer is always advisable. A Billroth I partial gastrectomy has stood the test of time as a standard operation for benign gastric ulcer. Proximal gastric vagotomy with ulcer excision did have a vogue and was tested in trials; overall, it was found not to be as satisfactory. In addition, it is technically difficult if there is a lot of scarring around the lesser-curve gastric ulcer. Emergency surgery for bleeding gastric ulcer is a different issue and will be discussed below. The recurrence rate of gastric ulcer after Billroth I in different studies varies between 0 and 16 per cent.

Complications of peptic ulcer

Complications of peptic ulcer account for 4500 deaths a year in the United Kingdom. The surgeon, nowadays, is mainly concerned with the complications of peptic ulcer rather than elective treatment. Unfortunately, so many of the patients are elderly with other serious conditions, such as atrial fibrillation treated with anticoagulants, which makes emergency surgery more hazardous.

Perforation

Acute perforation may be the first sign of peptic ulcer, and fatality in elderly people can be as high as 20 per cent. The classic symptoms and signs of severe epigastric pain, board-like rigidity, with air under the diaphragm on the chest radiograph, often make the diagnosis easy in 80 per cent of patients. But perforation is not always so straightforward. A gastric ulcer perforated into the lesser sac (Fig. 3) can give misleading symptoms and be confused with pancreatitis (the serum amylase may be slightly raised because of the absorption of pancreatic juice from the peritoneal cavity). Alternatively, the fluid from a perforated duodenal ulcer may track down the right paracolic gutter and present with tenderness in the right iliac fossa, mimicking acute appendicitis. More than one surgeon has made a small appendectomy incision only to find bile-stained fluid emerging from the peritoneal cavity! Perforation is particularly difficult to diagnose in the patient who is on a high dose of steroids, because the symptoms and signs (although not the radiograph) are masked. The mortality has a direct relation to the delay before treatment. The common differential diagnosis of peritonitis with air under the diaphragm is a perforated diverticulum of the colon.



Fig. 3. Erect radiograph of abdomen showing air in lesser sac from perforated gastric ulcer.

Management

The first stage of management is restoration of fluids and correction of electrolyte abnormalities, and the start of nasogastric suction. The aspirate commonly contains small flecks of fresh or altered blood. There is a long-standing tradition in surgical management of the acute stage that no analgesia should be given until a firm diagnosis has been made. This is unkind and unnecessary, especially for a patient who has to travel some distance with every bump in the road exacerbating the peritoneal pain. A short-acting analgesic such as pethidine may be given as long as the receiving room in the hospital is made aware that this has been done.

Is there a place for non-operative treatment?

Although surgery is normally the correct treatment for perforated duodenal ulcer, the whole patient and the comorbidity need to be taken into account. Perforations may seal themselves by adherence to liver, gallbladder, or omentum. If an elderly, unfit patient presents with a 3-day history suggesting perforation but does not show the signs of generalized peritonitis, and is otherwise a very poor operative risk, then non-operative treatment is entirely reasonable. This would mean nasogastric suction, intravenous fluids and antibiotics, and may be undertaken despite the presence of air under the diaphragm on the radiograph. A careful watch must be kept to ensure the patient does not deteriorate further. Abscesses caused by the perforation can often be drained percutaneously later.

A water-soluble contrast radiograph can help to confirm that the perforation is, in fact, sealed, but gastroscopy must be done with great care, with little gastric and duodenal distension, because a lightly sealed omental patch can be blown off by enthusiastic insufflation! Mortality rates of less than 10 per cent are claimed for non-surgical treatment, but these figures have little meaning unless the selection criteria are carefully scrutinized. A follow-up gastroscopy is required to ensure any ulcer has healed and to biopsy a gastric ulcer.

Operative techniques—simple suture or patch?

An acute, perforated duodenal ulcer may be closed by simple suturing and reinforced with an omental patch over the top. However, it is impossible to close a large, chronic ulcer when there is fibrosis, which holds the edges apart. In this case, an (attached) omental patch is sutured into the defect, with sutures passing across the perforation over the patch. If they are tied too tight, the omentum will become infarcted. Long-lasting but absorbable sutures are ideal for this, whereas non-absorbable sutures remain in the duodenal wall and can act as the site for further ulcerations.

Perforated gastric ulcers may be excised, the new edges brought together with interrupted sutures, and reinforced by an overlapping layer or some omentum. Normally there is plenty of ‘spare stomach' on the lesser curve or posterior surface to allow the edges to come together without tension. The excised ulcer must be sent for histological examination to exclude malignancy. In patients on steroids, special care and extra reinforcement are necessary because of slow healing. The peritoneal cavity is washed out thoroughly with saline and a drain inserted only if there was a collection of infected material that might re-accumulate.

Laparoscopic repair of perforated ulcer has been done, the aim being to reduce surgical trauma. However, washing out gastric contents, which is an important part of treatment, is not so easy through a laparoscope, and peas and pieces of chipped potatoes may be left behind! A well-conducted, randomized, controlled trial compared open and laparoscopic suture and glue; it was concluded that the laparoscopic approach had no advantage over the open procedure and took longer to perform, which is not surprising, considering that an open repair of perforation and wash-out is a quick and simple procedure. The use of glue is more controversial and has little advantage in the open procedure, although it is attractive if it is being done laparoscopically. But more work needs to be done before this can be recommended as a routine.

Should an ulcer-curing operation be added?

An important decision at operation is whether or not to add a vagotomy to simple ulcer closure (or perform a partial gastrectomy for a perforated gastric ulcer). Before the days of powerful acid inhibitors or the recognition of *H. pylori*, surgeons based their decisions on whether or not there was chronic ulcer disease and on the risks of prolonging the operation. Now that the natural history of peptic ulcer can be altered, either by continuous drug therapy or elimination of *Helicobacter*, the need to do a definitive procedure has greatly diminished. The surgeon must aim to close the perforation as quickly and effectively as possible, especially in poor-risk patients. The exception may be when a patient is likely to have to stay on non-steroidal anti-inflammatory drugs or steroids and may already be on a protective dose of acid-reducing drugs. Although vagotomy does not protect completely from ulceration by these drugs, it acts synergistically with misoprostol or long-term H_2-receptor antagonists. For gastric ulcer, there is still debate whether Billroth I for perforation is safer in terms of making sure the ulcer actually heals. Again, this judgement has to be made against the experience of the surgeon and the general condition of the patient. If a drainage procedure is used to accompany a truncal vagotomy, then a gastrojejunostomy is preferable to a pyloroplasty across the perforation for duodenal ulcer. This is because it can be closed later, should dumping or diarrhoea be a problem (see below). A

gastrojejunostomy may also be required when the repair of a very scarred duodenal ulcer might produce some gastric-outlet obstruction.

Haemorrhage

All haemorrhage from peptic ulcers is potentially lethal and temporary cessation of bleeding is no cause for complacency. Bleeding peptic ulcer accounts for 70 per cent of all upper gastrointestinal bleeds, and the general mortality is 5 to 10 per cent. Thirty per cent of patients who bleed from peptic ulcers have no previous definite history or previous diagnosis of ulcer disease. The mortality among those who need surgery is higher, around 20 per cent, because their bleeding is the most severe. Sadly, the mortality has not improved over the last 50 years because, despite fibreoptic endoscopy and other advances, the patients who bleed are increasingly elderly with many concurrent illnesses. Important points in their history include previous peptic ulcer, alcohol abuse, drugs (particularly non-steroidals), and anticoagulants, and cardiovascular and pulmonary disease. Important signs are the stigmata of liver disease, which point to oesophageal varices as the cause of the bleeding, but do not, it must be emphasized, exclude peptic ulcer disease.

The management of haemorrhage consists of five phases: (i) resuscitation; (ii) diagnosis; (iii). immediate control of bleeding; (iv) prevention of re-bleeding; (v) prevention of recurrent bleeding in the future. Only when an ulcer is healed and remains healed has the risk of haemorrhage passed. There is increasing world-wide evidence that patients with upper gastrointestinal haemorrhage are best treated in a dedicated ‘bleeding unit' jointly staffed by surgeons and gastroenterologists, so that they can be closely observed and agreed protocols adhered to so that intervention can be prompt. In such units, a mortality of less than 5 per cent can be achieved. Bleeding presents as haematemesis, melaena, or both. Massive bleeding can lead to only slightly altered blood being passed per rectum, but in this situation the patients are severely shocked because the bleeding is so severe.

Resuscitation

This consists of an intravenous line, cross-matching of blood, and a full blood count. The haemoglobin measured immediately afterwards does not give any indication of the severity of the bleed as the concentration will not drop until there has been dilution with replacement fluids. A low haemoglobin at the time of the bleed indicates previous chronic bleeding. There is continued transatlantic debate about whether crystalloids or colloids should be given while awaiting the blood.

Diagnosis

If there has been a significant bleed with the blood pressure low and other signs of shock, patients should be gastroscoped as soon as they are stable. The purpose of the gastroscopy is to determine the site of bleeding and judge the risk of recurrent haemorrhage; it does not often help in determining when the surgeon should operate. The most important differential diagnoses are the results of portal hypertension—oesophageal or gastric varices and congestive gastropathy (see Chapter 32.1). The stigmata of recent haemorrhage are shown in Table 3 and have some value in predicting re-bleeding. An intravenous injection of metaclopromide before endoscopy helps to clear the stomach of blood and, incidentally, control temporarily the bleeding from oesophageal varices.

Sign	*Risk of rebleed* (%)*
Pulsatile bleeding or oozing	85
Fresh clot in ulcer crater	85
Adherent clot on ulcer	50
Visible vessel in base: i.e., dark spot, red spot, elevated tubular structure, defect in base of ulcer	40

*Or of continued bleeding

Table 3 Endoscopic signs of recent haemorrhage

Immediate control of bleeding

The simplest method of controlling bleeding is the injection of 0.5 to 1 ml volumes of 1 in 100000 adrenaline (epinephrine) into the ulcer base around the bleeding point. Other methods include neodymium–yttrium aluminium garnet lasers and bipolar electrocoagulation, as well as the injection of sclerosants with and without adrenaline. Prospective, randomized trials support simple injection techniques. Indications for immediate surgery are as follows:

1. very severe bleeding in the absence of signs of liver disease when endoscopy would cause delay;
2. the failure to be able to see the bleeding point in someone who is still bleeding (oesophageal varices having been excluded);
3. failure to control a spurting vessel by injection;
4. rare lesions, such as an aortoenteric fistula, that invariably re-bleed.

Prevention of re-bleeding

Re-bleeding in hospital is the major cause of death, and so it is very important that the patient should not be allowed to have a third bleed. Most elderly patients stand an operation better than a recurrence of severe bleeding. Continued oozing, although not dramatic, can be an unrecognized danger. The signs of re-bleeding are:

1. fresh haematemesis or melaena in a stabilized patient (old melaena may be present for some time after the initial bleed);
2. a fall in blood pressure;
3. a steady rise in pulse rate in a stabilized patient.

Indications for surgery are as follows:

1. one severe re-bleed;
2. transfusion requirements of 4 units in 24 h in a patient over 50 years or 6 units in a patient under 50 years of age;
3. shortage of whole blood.

It is normal to start acid-suppressing drugs, but it is naive to think that they would prevent re-bleeding in a large ulcer with an exposed vessel. It takes many days, if not weeks, for the ulcer to heal to the extent that eliminates the risk of further bleeding.

Prevention of recurrent ulcer

The secondary aim of treatment is to prevent recurrent ulcer in the future, but this must be considered as part of the overall treatment plan. It would determine the type of surgery.

Operation

As with a perforated ulcer, the decision is whether to do definitive surgery with an ulcer-curing operation, such as Billroth I partial gastrectomy for a gastric ulcer, or truncal vagotomy and pyloroplasty for duodenal ulcer, or whether to use medical treatment to prevent future ulcers. Theoretically, undersewing of the bleeding vessel together with medical treatment, such as the elimination of *H. pylori*, is an option. Unfortunately, the international randomized trial designed to answer this question did not control for surgical techniques and had to be stopped because of a high mortality after minimal surgery. The decision depends on the patient's general condition and the experience of the surgeon. Figure 4 summarizes the management of bleeding peptic ulcer.



Fig. 4. Algorithm for management of bleeding peptic ulcer.

Stenosis

Stenosis of the body of the stomach, creating two chambers like an hourglass, is rare nowadays because ulcers are treated effectively before they reach this stage. Pyloric (or, more accurately, duodenal) stenosis still occurs because fibrosis is the result of healing. However, duodenal oedema is more common than true fibrosis and may clear with medical treatment. When a benign gastric ulcer is secondary to pyloric stenosis (type 2), the stenosis needs treating. Occasionally, a benign gastric ulcer can be secondary to malignant gastric-outlet obstruction. A non-operative option for established stenosis is balloon dilatation at endoscopy, and this may be successful when the main problem is oedema rather than true stenosis. However, effective dilatation of a stenosis can lead to rupture of the duodenal wall, or to re-stenosis rather soon afterwards. For established fibrosis, a vagotomy and pyloroplasty or gastroenterostomy is necessary, with excision of a gastric ulcer if present. The advantage of a gastroenterostomy is that, if the stenosis does improve and there are side-effects from the drainage procedure (such as dumping or bilious vomiting), the gastroenterostomy may be closed later. If the stenosis is sufficiently distal to the pylorus, a duodenoplasty may be used, thus preserving the pyloric function. Operative dilatation of the pylorus followed by a proximal gastric vagotomy, although an attractive option, was not always successful, because of leakage and re-stenosis.

Fistulas

Gastrocolic

A benign gastric ulcer of the greater curve may occasionally cause a gastrocolic fistula, although this is more common (10–20 per cent of cases) when a jejunal ulcer occurs at the anastomosis of a retrocolic gastroenterostomy; these should strictly be called gastrojejunocolic fistulas. The unpleasant side-effects are diarrhoea, weight loss, and faecal vomiting. Surgeons used to recommend a proximal colostomy first, before doing a resection of the ulcer, closing the colon, and adding an antrectomy or truncal vagotomy depending on the previous operation. Nowadays a one-stage operation is feasible after the patient's nutrition and electrolyte disturbances have been corrected. Duodenocolic fistulas are rare and can result from a duodenal ulcer or carcinoma of the hepatic flexure.

Duodenobiliary

A fistula may form between the duodenum and the gallbladder or the common bile duct. Cholecystoduodenal fistulas are usually due to gallbladder disease and may lead to gallstone ‘ileus' when a large gallstone passes through the fistula and becomes impacted in the terminal ileum. Fistulas are particularly common when there is xanthogranulomatous cholecystitis. Choledocoduodenal fistulas are due to posterior penetration of a duodenal ulcer. If there is a narrow opening or other cause of biliary stenosis, then cholangitis results. The fistula may heal when the ulcer does, but if not it is excised, and the duodenum and bile duct closed over a T-tube if stenosis is a risk. Occasionally a choledocojejunostomy is required if there is severe narrowing of the bile duct.

Malignant change in peptic ulcers

See Chapter 23.2.

Surgical techniques

Billroth I partial gastrectomy

This is the standard operation for gastric but not for duodenal ulcer. The distal stomach and pylorus, including the ulcer, is excised and an end-to-end gastroduodenal anastomosis is made to restore continuity (Fig. 5). The benign ulcers are commonly on the lesser curve and therefore the resection line may be taken high. A new lesser curve is made to narrow the stoma of the gastric remnant to the size of the duodenum. A specially angled clamp makes this easier (Fig. 5(a)); stapling devices are quicker, but far more expensive than sutures. Dividing the lateral peritoneal attachments of the duodenum (Kocherization) allows the duodenum to be mobilized medially to prevent any tension on the anastomosis. The physiological advantage is that the food progresses in the normal direction.

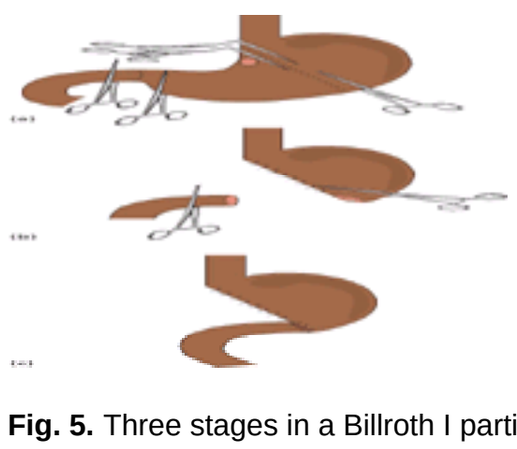


Fig. 5. Three stages in a Billroth I partial gastrectomy.

Pylorus-preserving gastrectomy (Maki)

This variant was devised to retain pyloric function and therefore prevent the rapid gastric emptying and dumping that may occur when the pylorus is excised. About 1.5 cm proximal to the pylorus is preserved with the pyloric branches of the vagus nerve that come down from the liver (Fig. 6). This may lead to initial delayed emptying compared with Billroth I, but both clinical assessment and isotope emptying studies show that it retains a more normal emptying pattern long term and therefore may be appropriate for benign disease.

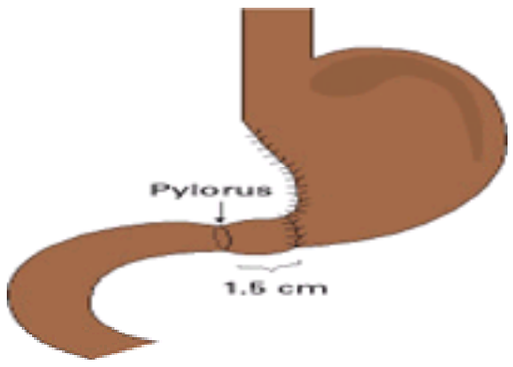


Fig. 6. Pylorus-preserving gastrectomy (Maki).

Billroth II (Polya) partial gastrectomy

This can be used for both duodenal and gastric ulcers, and involves resecting two-thirds of the stomach, closing the duodenum distal to the pylorus (usually beyond the ulcer), and an end-to-side gastrojejunal anastomosis. In the past there was much discussion of whether the jejunum should be retrocolic or antecolic and whether the loop should be isoperistaltic to the stomach or retroperistaltic (Fig. 7). For a benign disease, this probably does not matter, but for malignant disease the jejunum is brought in front of the colon so that it is less likely to be involved in any local recurrence. Closing off the duodenum takes the ulcer-bearing area out of the acid stream, but the physiology of digestion is more disturbed than after a Billroth I.

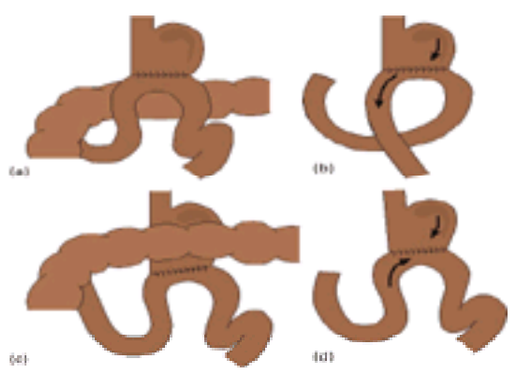


Fig. 7. Gastrojejunostomies: (a) antecolic, (b) isoperistaltic, (c) retrocolic, (d) antiperistaltic.

The closing of the duodenum can be very difficult if there is much scarring. It may have to be closed distal to a posterior penetrating ulcer, and this may involve a longitudinal rather than transverse suture line (Fig. 8). Sometimes the ulcer can be excluded and the duodenum closed, leaving the ulcer posteriorly on to the pancreas. With very large, fibrosed ulcers, as a last resort, a drainage tube may be put into the duodenum and the duodenum sutured around it to provide a formal duodenal fistula, but this is not ideal. Another technique is to anastomose a jejunal Roux limb over the duodenum instead of trying to close the duodenum when there is only a little tissue available. It is the leakage from the duodenal stump that provides the main risk of Polya partial gastrectomy, and that is why vagotomy and drainage is a safer operation.

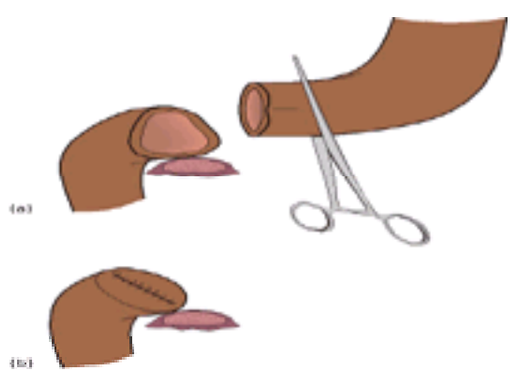


Fig. 8. Longitudinal closure of duodenum distal to a posterior penetrating duodenal ulcer.

Vagotomy and antrectomy

If resection of the antrum (gastrin-secreting area) is used in addition to a truncal vagotomy, the resected area is smaller, about 6 cm proximal to the pylorus on the lesser curve. The re-anastomosis may be gastroduodenal or gastrojejunal, the gastroduodenal being preferable if the duodenum is not too scarred (Fig. 9).

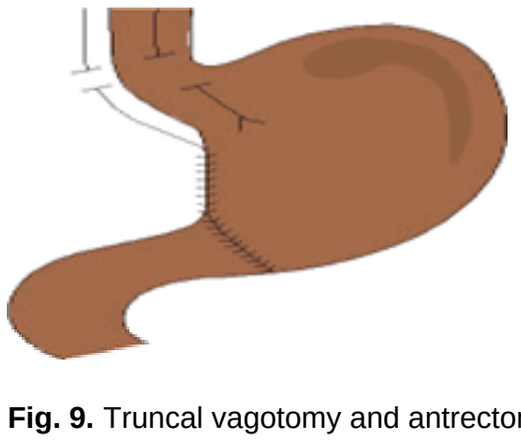


Fig. 9. Truncal vagotomy and antrectomy with gastroduodenal anastomosis.

Truncal vagotomy and pyloroplasty or gastrojejunostomy

Truncal vagotomy is carried out as high as possible on the abdominal oesophagus above the area where the vagi start to branch. It is really a perhiatal procedure. The peritoneum over the anterior surface of the oesophagus is divided and the phreno-oesophageal ligament identified and elevated, exposing the anterior vagal trunk. The posterior trunk lies on the right posterior aspect of the oesophagus, often well away from the oesophageal wall, and the encircling finger around the oesophagus may pass medial to it and therefore it may be missed. A 1-cm piece is taken out of each of the trunks and the surgeon must carefully search the oesophageal wall for small

branches. A 4 cm-long segment of the oesophagus should be denuded of its attachments. The habit of sending a piece of suspected vagus nerve for histological examination does not help completeness because the trunks may branch high and small branches be left. Some surgeons recommend re-anastomosing the peritoneum over the oesophagus, but it does not seem that this is important. When a patient has perforated and the peritoneal cavity is contaminated, there is a theoretical possibility of a mediastinitis but this is not a problem in practice.

Pyloroplasty

These are of several different types, for example Heineke–Mikulicz, Finney (gastroduodenostomy), and ulcer-excluding (Fig. 10). They all bypass or destroy the pyloric mechanism and so affect gastric emptying and duodenogastric reflux; once this has been destroyed, a good-sized pyloroplasty is probably better than a very small one, which could delay emptying. Pyloromyotomy, as is sometimes used in conjunction with oesophagectomy, is not usually appropriate when there is a marked duodenal ulcer.

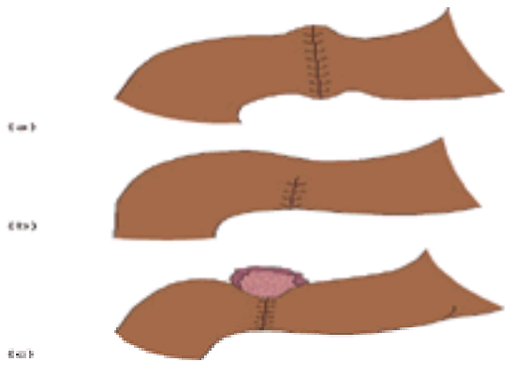


Fig. 10. Pyloroplasties: (a) Heineke–Mikulicz, (b) Finney,(c) ulcer-excluding.

Proximal gastric vagotomy

This is also referred to as parietal-cell or highly selective vagotomy, and is a modification of vagotomy that leaves the antrum innervated and thus avoids the need for a drainage procedure by leaving gastric emptying as normal as possible. It is more demanding technically because only small nerve branches are divided close to the lesser curve of the stomach. The only places that residual innervation can be left are at the lower and upper limits of dissection. Unless special intraoperative tests are used, the distal (antral) dissection should be less than 6 cm proximal to the pylorus on the lesser curve, leaving only the most distal branch from the nerve of Latarjet. The proximal dissection is carried 6 cm up the oesophagus and 6 cm out along the fundus towards the spleen. This may or may not entail dividing the first short gastric vessel (Fig. 11).

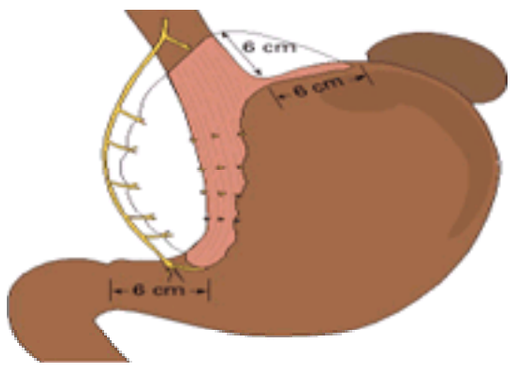


Fig. 11. Completion of proximal gastric vagotomy showing denervation 6 cm up oesophagus, 6 cm out along fundus and 6 cm proximal to pylorus.

Anterior seromyotomy and posterior truncal vagotomy

A modification in order to speed the operation was to cut the muscle of the lesser curve just in from the edge, and divide the nerve branches as they were just plunging in through the muscle. As this is difficult to do on the posterior surface, a posterior truncal vagotomy was performed. The operation is based on the principle that only the anterior nerve to the antrum need be preserved to ensure antral function. However, more recent anatomical dissections have found that, in about 3 per cent of patients, the only supply to the antrum is from the posterior nerves, such that they would get delayed gastric emptying. This operation was effective in the hands of those who were enthusiasts but did not gain general acceptance; although it is quicker to perform, it probably is not as sure in dividing all the branches as the more tedious standard dissection that divides all structures entering the lesser curve (Fig. 12).

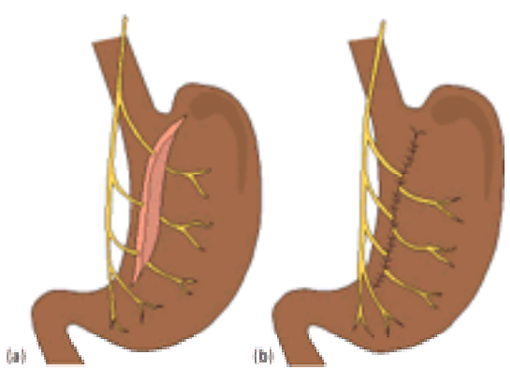


Fig. 12. Lesser-curve seromyotomy: (a) extent of seromyotomy;(b) completed overlap repair.

Laparoscopic procedures

Truncal vagotomy and gastroenterostomy has been performed laparoscopically and several small series have been reported of the more selective procedures. Seromyotomy is probably the most suitable for the laparoscopic approach. It is still an open question how much the method of access alters the metabolic response and there are not enough elective operations being done to mount a randomized, controlled trial. However, pre- and postoperative acid-secretion studies should be used to assess completeness of vagotomy, as this is essential to prevent recurrence.

Postoperative management

Apart from the general management of any patient after a major abdominal operation, the question the patients wish to have answered is when they can start to eat and drink! All gastric operations cause some degree of initial gastric stasis and it used to be the practice to keep a nasogastric tube in for some days afterwards. There is good evidence that this does no good and, in fact, probably does some harm. Very occasionally, acute dilatation of the stomach occurs after both vagotomy and

resection, and presents as shock and effortless bilious vomiting. This must be recognized and a nasogastric tube passed immediately, while fluids are replaced intravenously.

After proximal gastric vagotomy, gastric emptying is nearly normal as there is no antral denervation, so the patients may drink in a few hours and take a soft diet the following day. All other gastric operations involve a suture line and after a gastrojejunal anastomosis the stomach tends to fill with bile. Sometimes a nasogastric tube is helpful in relieving nausea during this period, but otherwise the patient can start to increase fluid gradually from 48 h. When a significant proportion of the stomach has been removed, meals have initially to be little and often until the gastric remnant adapts, but no longer should patients be put on strict ‘gastric diets'; they should be encouraged to try a full, mixed diet in small quantities and just to avoid any particular food that upsets them.

Alcohol

After all operations for peptic ulcer except proximal gastric vagotomy, the more rapid emptying of liquids means that alcohol will be absorbed very rapidly into the circulation; therefore, even a small amount of alcohol quite soon gives high blood concentrations and the patient must be warned, ideally in writing, about the risk of driving or working machinery.

Postvagotomy and postgastrectomy syndromes

Dumping

This is associated with rapid emptying from the stomach. It is traditionally divided into early or late, but often the differentiation is not so clear-cut. It consists of a group of cardiovascular and gastrointestinal symptoms, such as faintness, sweating, tachycardia, bloating, nausea, and cramping abdominal pain.

Early dumping

Gastric emptying is normally regulated by duodenal osmoreceptors, but if the pylorus is divided or bypassed, hypertonic fluids can be ‘dumped' into the upper small intestine. This leads to an outpouring of fluid into the small intestine to dilute the bowel contents, thereby reducing the blood volume. Whether or not a particular patient experiences cardiovascular symptoms may depend on how sensitive he/she is to slight changes in plasma volume. Gastrointestinal symptoms are due to the sudden release of gastrointestinal peptides such as cholecystokinin and motilin. Symptoms severe enough to interfere with normal activity occur in about 5 per cent of patients after vagotomy and drainage or partial gastrectomy, but another 10 per cent have milder symptoms. Symptoms tend to improve with the time from the operation as the upper gastrointestinal tract adapts.

Treatment is initially dietary by avoiding high-osmotic foods and separating drinking and eating. In severe cases, somatostatin injected subcutaneously just before meals may give dramatic release by blocking the output of peptides or hormones. Although this can be a nuisance and is expensive, it is no more troublesome than that experienced by diabetics who have regularly to self-administer insulin. Long-acting forms of somatostatin analogue are now available.

Late dumping

This is due to hypoglycaemia occurring about 2 h after a meal because of a large initial secretion of insulin in response to the high sugar load. The insulin and glucose concentrations can be measured and the phenomenon confirmed. The syndrome is less common than early dumping, and the dietary changes should be the same. However, the patient can also carry a glucose sweet, which can be taken as soon as the symptoms start, to prevent a severe hypoglycaemia.

Surgical treatment

If the patient has a gastroenterostomy and a patent, intact pylorus, then just taking down the gastroenterostomy will probably solve the problem. (One year after a truncal vagotomy the drainage procedure is no longer needed because the vagal reflexes within the stomach have adapted.) Repair of a pyloroplasty is possible, and in over 60 per cent of cases improves the symptoms sufficiently for them not to interfere with the patient's normal activity. The operation must be precise over an 10 to 12 mm bougie (Fig. 13) (12 mm is the maximum resting diameter of the normal pylorus). The insertion of a reversed jejunal segment between the stomach and jejunum may give temporary relief, but after a time the antiperistaltic activity appears to cease.

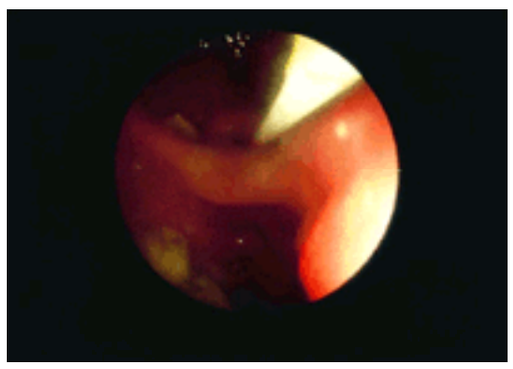


Fig. 13. Gastroscope view of reconstituted pylorus.

Diarrhoea

‘Postvagotomy' diarrhoea is the most common postoperative symptom. Some change in bowel habit (often for the better!) occurs in 70 per cent of patients, but severe diarrhoea may affect 10 per cent of patients after truncal vagotomy and drainage, but only 1 per cent after proximal gastric vagotomy. It is partly due to the removal of the parasympathetic influence on the function of the fore- and midgut, but the drainage procedures are also important. The increase in small-bowel fluid load and rapid release of gastrointestinal peptides affecting the colon also play their part (exaggerated duodenocolic reflex). In addition, the hypoacidity of the stomach removes the barrier to bacterial entry into the bowel, which can cause enteritis. The role of bile salts is important, and if a patient has had a cholecystectomy and a truncal vagotomy and drainage at the same time, or separately, the incidence of diarrhoea is 40 per cent.

Treatment

Fortunately, we have good antidiarrhoeal drugs such as codeine phosphate or loperamide, but, despite this, 1 per cent of patients find the condition disabling. Bacterial overgrowth should be looked for using the breath test and treated if present. Cholestyramine, which binds bile salts, does have an effect on diarrhoea, although not on reflux symptoms. Surgery is occasionally needed and many types of reversed jejunal loops have been tried. Perhaps the most successful is the distal reversed ileal onlay graft (Cuschieri).

Bile ‘reflux'

The pylorus not only controls emptying but also prevents reflux of duodenal juice. When it is removed, destroyed or bypassed, there is reflux into the stomach (remnant); although bilirubin is the obvious culprit because of its colour, pancreatic juice is also important. When bile and pancreatic juice mix, lysolecithin is produced from lecithin. This, together with the bile salts, is damaging to the gastric mucosa. The symptoms are nausea, burning epigastric pain, bilious vomiting, and weight loss. As with the other syndromes, many people have the physiological changes without the symptoms and it is not known why some are ‘sensitive' to duodenal juice in the stomach and others are not. Persistent reflux can damage the gastric mucosa causing foveolar hyperplasia, even in the absence of *H. pylori*. Duodenal juice can also reflux into the oesophagus and may be particularly dangerous there: it is probably linked with the intestinal metaplasia of Barrett's oesophagus, with its malignant potential. Nowadays, bile in the stomach and oesophagus can be verified by a semiquantitative bile probe (Bilitech) or it can be inferred from the sodium concentrations, detected by a sodium-sensitive probe. There is no direct correlation between symptoms and the degree of reflux, but mucosal damage can occur in the absence of symptoms.

Treatment

Cholestyramine seems to have little effect in binding bile salts in the stomach (as opposed to the colon) and this may relate to pH. However, prokinetic drugs such as metoclopramide or cisapride may help to clear the stomach and oesophagus, particularly if the antrum is still present.

Revisional surgery

This should only be done once the presence of significant reflux has been established, either by a history of repeated bilious vomiting or direct measurement. Pyloric reconstruction or the closure of a gastrojejunostomy are the first surgical measures if there has been no resection. After a Polya (Billroth II) gastrectomy, a Roux-en-Y reconstruction or Tanner Roux 19 is relatively easy to do without refashioning the anastomosis (Fig. 14). If the Roux limb is at least 40 cm, reflux of duodenal juice is eliminated from the stomach, but sometimes the Roux limb empties poorly and patients must be warned that all symptoms may not go.

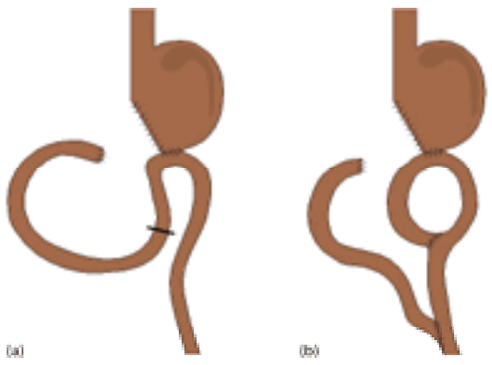


Fig. 14. Tanner Roux 19 reconstruction after Polya (Billroth II) partial gastrectomy.

The dilemma for the surgeon is whether to add a vagotomy to prevent recurrent ulceration of the stoma in the unprotected Roux limb, or whether to rely on long-term, low-dose, acid-reducing drugs. Adding a vagotomy to the gastric remnant can make any delayed emptying worse. As the vagotomy can be technically difficult after a previous operation, drug treatment, particularly in the elderly patient, is probably best.

Malignant change

Twenty years after a gastric resection for benign disease, a patient has a 3.7-fold increased risk of developing carcinoma of the gastric remnant. When the operation is for gastric ulceration, the risk is 5.5, rising to 8.6 if the original procedure was a Billroth II (Polya). The frequency does not justify regular endoscopic follow-up, but any new symptom or anaemia long after a gastrectomy should be taken very seriously. The risk to the stomach and other organs may relate to the production of nitrosamines by bacteria in the relatively hypoacidic stomach remnant or due to long-term bile damage to the gastric mucosa leading to metaplasia and dysplasia.

Delayed gastric emptying

Delayed gastric emptying of solids can coexist with rapid emptying of liquids and persists in a few patients long after the early postoperative period. Solid emptying depends on antral contractions, whereas the early liquid emptying is produced by fundal pressure. After vagotomy, especially if there has been some obstruction of the antral outlet, emptying may take many months to resolve. Patients, therefore, are advised to keep their meals as dry as possible and drink between meals, and to bite their meals up well.

Treatment

Further surgery does not help unless there is actual narrowing or obstruction at the outlet. Adding a gastrojejunostomy to a pyloroplasty merely adds more bile reflux to the poorly emptying stomach. Prokinetic drugs are helpful, for example cisapride, metoclopramide or erythromycin, and have even been found to give some benefit on the gastric remnant when the antrum has been removed.

Stomal ulcer

Ten years after vagotomy and gastroenterostomy or pyloroplasty, there is a 10 per cent chance of stomal ulcer, and after Polya (Billroth II) gastrectomy about 5 per cent where a vagotomy has not been added. After a vagotomy and antrectomy, recurrent ulcer is very rare. The persistent use of non-steroidal anti-inflammatory drugs correlates with this risk, whereas *Helicobacter* disappears after a drainage procedure when there is increased bile reflux, e.g. Polya partial gastrectomy.

Nutritional problems

The classic two-thirds partial gastrectomy leads to loss of weight in populations where the food is bulky and not high in calories because there is not enough space for sufficient food. Vagotomy, which left the stomach intact, was a significant improvement.

Late problems include iron, folate and vitamin B_{12} deficiency as well as hypocalcaemia. Vitamin B_{12} is usually absorbed if a fair gastric remnant has been left. It is important to check that the patients are eating a balanced diet. Persistent diarrhoea as steatorrhoea can lead to hypocalcaemia and malabsorption of fat and fat-soluble vitamins, especially when the duodenum is bypassed and the mixing of food with bile and pancreatic secretion is poor. It is prudent to check the calcium several years after peptic ulcer surgery.

Conclusion

For 100 years, surgery was the definitive treatment for chronic peptic ulcer. Over those years, operations were refined to make them more physiological and improvements in perioperative care made them increasingly safe and successful. But surgeons were often treating the secondary cause, that is, hypersecretion of acid, rather than the primary cause.

Today, gastric surgeons have the challenge of difficult decisions and difficult operations in the treatment of life-threatening emergencies affecting an increasingly elderly population. Surgeons must also be alert to early gastric malignancy associated with peptic ulcer, or its previous surgical treatment, as well as to the long-term complications of peptic ulcer surgery, which was performed frequently up to 10 to 15 years ago.

However, preliminary reports in the United States of America suggest that, now, 50 per cent of gastroduodenal ulcers are not associated with *H. pylori*. This and emerging antibiotic resistance by *Helicobacter* is a warning to the surgical community not to lose the experience and skills so painfully learnt in the long years before *H. pylori* was recognized.

Further reading

Axon ATR, Tytgat GNJ, ed. Educational Training Workshop in *Helicobacter pylori*, Bologna 1997. *Gut* 1998; **43**: Suppl. 1. [This is a thorough analysis of all aspects of *Helicobacter pylori*.]

Casas AT, Gadacz TR. Laparoscopic management of peptic ulcer disease. *Surgical Clinics of North America* 1996; **76**: 515–22. [A review of laparoscopic techniques.]

Caygill C, Hill MJ, Kirkham JS, Northfield TC. Mortality from gastric cancer following gastric surgery for peptic ulcer. *Lancet* 1986; **i**: 929–31. [A classic long-term follow-up of large numbers of patients.]

Cook DJ, Bown SG. Endoscopic treatment for bleeding peptic ulcers: randomised comparison of adrenaline injection and adrenaline injection and Nd:YAG laser photocoagulation. *Gut* 1991; **32**:

1100–3. [One of a number of prospective randomized trials.]

Cook D *et al.* A comparison of sucralfate and ranitidine for the prevention of upper gastrointestinal bleeding in patients requiring mechanical rehabilitation. *New England Journal of Medicine* 1998; **338**: 791–7. [A prospective trial showing that sucralfate was effective.]

Cuschieri A. Surgical management of severe intractable post-operative diarrhoea. *British Journal of Surgery* 1986; **73**: 981–4. [A detailed description of probably the best technique.]

Dragstedt LR, Harper PV, Tovee EB, Woodward ER. Section of the vagus nerve to the stomach in the treatment of peptic ulcer. *Annals of Surgery* 1947; **126**: 687–708. [The classic paper which led to the re-introduction of vagotomy into clinical practice.]

Goligher JC *et al.* Several standard operations for duodenal ulcer: ten to sixteen year clinical results. *Annals of Surgery* 1979; **189**: 18–24. [The famous prospective randomized Leeds–York trial whichwas a model for future surgical trials.]

Hawkey CJ *et al.* Omeprazole compared with misoprostol for ulcers associated with nonsteroidal anti-inflammatory drugs. *New England Journal of Medicine* 1998; **338**: 727–34. [An important paper dealing with the problem of preventing non-steroidal ulcers.]

Hopkins RJ, Girardi LS, Turney EA. Relationship between *Helicobacter pylori* eradication and reduced duodenal and gastric ulcer recurrence: a review. *Gastroenterology* 1996; **110**: 1244–52. [An excellent detailed review of the long-term effects of *Helicobacter* eradication.]

Johnson HD. Etiology and classification of gastric ulcers. *Gastroenterology* 1957; **33**: 121–3. [A classic paper differentiating between various types of gastric ulcer.]

Johnston D, Wilkinson AR. Highly selective vagotomy without a drainage procedure in the treatment of duodenal ulcer. *British Journal of Surgery* 1970; **57**: 289–95. [The key paper that introduced highly selective vagotomy into clinical practice.]

Labenz J, Blum AL, Bayerdörffer E, Meining A, Stolte M, Börsch G. Curing *Helicobacter pylori* infection in patients with duodenal ulcer may provoke reflux oesophagitis. *Gastroenterology* 1997; **112**: 1442–7. [A warning that curing *Helicobacter* may have other effects.]

Lau WY *et al.* A randomised study comparing laparoscopic vs. open repair of perforated peptic ulcer using suture or sutureless techniques. *Annals of Surgery* 1996; **224**: 131–8. [A well conducted, prospective study with quite small numbers showing no overall advantage to the laparoscopic technique.]

Lindsetmo RO, Johnsen R, Revhaug A. Abdominal and dyspeptic symptoms in patients with peptic ulcer treated medically or surgically. *British Journal of Surgery* 1998; **85**: 845–9. [A wider look at symptoms and not just ulcer healing.]

Mansi C, Mela GS, Savarino V, Mele MR, Valle F, Celle G. Open access endoscopy: a large scale analysis of its use in dyspeptic patients. *Journal of Clinical Gastroenterology* 1993; **16**: 149–54. [This gives the proportion of different diagnoses in patients attending open access endoscopy.]

Marshall BJ, Warren, JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* 1984; **i**: 1311–15. [The seminal papers that identified campylobacter (*Helicobacter*) and its possible relationship to ulcer disease.]

Marshall BJ, McGeckie DB, Rogers PA, Glancy RJ. Pyloric campylobacter infection and gastroduodenal disease. *Medical Journal of Australia* 1985; **142**: 439–44.

Peura DA. The report of the Digestive Health Initiative International Conference on *Helicobacter pylori. Gastroenterology* 1997; **113**: S4–8. [An extensive review on *Helicobacter*.]

Schubert ML, Makhlouf GM. Neural and paracrine regulation of gastrin and gastric acid secretions. *Gastroenterology* 1996; **111**: 837–8. [A detailed analysis of the different stimuli to gastric-acid secretion.]

Seely SF, Campbell D. Non-operative treatment of perforated peptic ulcer. *Surgery, Gynecology and Obstetrics* 1956; **102**: 435–46. [An old paper following non-operative treatment for perforated peptic ulcer showing surprisingly good results.]

Sekine T, Sato T, Maki T, Shiratori T. Pylorus-preserving gastrectomy for gastric ulcer—one to nine year follow up study. *Surgery* 1975; **77**: 92–9. [The paper that gives the description and results of the Maki procedure.]

Stoddard CJ, Johnson AG, Duthie HL. The four to eight year results of the Sheffield trial of elective duodenal ulcer surgery: highly selective or truncal vagotomy? *British Journal of Surgery* 1984; **71**: 779–82. [One of the definitive trials of vagotomy.]

Tanner NC. Result of operations for post-gastrectomy symptoms. *Gastroenterologia (Basel)* 1959; **92**: 146–56. [A description of Roux 19 and other techniques from one of the leading gastric surgeons of this century.]

Taylor TV, Gunn AA, Macleod DA, MacLennan I. Anterior lesser curve seromyotomy and posterior truncal vagotomy in the treatment of chronic duodenal ulcer. *Lancet* 1982; **ii**: 846–9. [A detailed description of the anterior lesser curve seromyotomy.]

Tomb JF *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori. Nature* 1997; **388**: 539–47. [An important paper for future study and control of *Helicobacter*.]

Wilkinson JM, Hosie KB, Johnson AG. Long term results of highly selective vagotomy: a prospective study with implications for future laparoscopic surgery. *British Journal of Surgery* 1994; **81**: 1469–71. [Long-term follow-up showing excellent results of highly selective vagotomy and modified Visick grading.]

Wyman A, Stuart RC, Ng EKW, Chung SCS, Li AKC. Laparoscopic truncal vagotomy and gastroenterostomy for pyloric stenosis. *American Journal of Surgery* 1996; **171**: 600–3. [Paper from Hong Kong showing that laparoscopic gastric surgery is possible.]

Yeomans ND *et al.* A comparison of omeprazole with ranitidine for ulcers associated with nonsteroidal anti-inflammatory drugs. *New England Journal of Medicine* 1998; **338**: 791–26. [Another paper looking at the problem of non-steroidal anti-inflammatory drug induced ulcers and their prevention.]

23.2 Carcinoma of the stomach

Hugh Barr and Michael J. Greenall

Epidemiology and aetiology
Site of gastric carcinoma
Aetiological factors
Pathology
Macroscopic appearances
Microscopic appearances
Molecular biological factors
Patterns of spread and metastasis
Staging for gastric carcinoma
Clinical presentation
Investigation
Laboratory
Radiography
Endoscopy and biopsy
Cytology
Laparoscopy and laparotomy
Population screening
Treatment and outcome
Historical considerations
Oncological approach
Radical and attempted curative surgery
Palliative surgery
The role of radiotherapy
The role of chemotherapy
Antimicrobial therapy and gastric lymphoma
Further reading

The ancient Egyptian papyrus Ebers describes a patient with dysphagia in whom the stomach had the appearance of a shrivelled fetal face. This description by Egyptian physicians may well be the first of a gastric carcinoma. Today, the disease is a major health-care and surgical problem.

Epidemiology and aetiology

As with many tumours, there is a widespread variation in the incidence of gastric adenocarcinoma around the world. The tumour occurs in almost epidemic proportions in Japan, with an incidence of 70 per 100 000 males; the cumulative risk of developing the disease by the age of 75 years is 11 per cent. It is said to be the ‘national disease' and in 1960 accounted for nearly 52 per cent of deaths from malignancy in men and 38 per cent in women. Other countries with a high incidence (more than 30 per 100 000 males) include Chile, Costa Rica, Hungary, Poland, Portugal, Iceland, Romania, Indonesia, and Italy. There is a particularly low incidence in some Central African countries (for example Uganda).

Marked regional variations have also been reported. In Columbia, an incidence of 50 per 1 000 000 males occurs in the inland city of Cali compared with 12 per 100 000 males on the coast at Cartagena. The rural population in Poland is at greater risk than the urban, in contrast to the risk for other cancers.

In the United States and the United Kingdom the overall incidence has fallen over the last 50 years. In the United States the age-adjusted mortality for men fell from 28 per 100 000 in 1935 to 9.7 in 1967 and 8.2 in 1986. This decline has occurred in all racial/ethnic groups and in both sexes, and reflects the change in incidence and not earlier diagnosis, better treatment, or changes in definition. Although a falling incidence has been noted in most countries, there is no decline in Poland or amongst Japanese men. In both men and women in Portugal there is an increase in the number of deaths reported.

Carcinoma of the stomach is closely associated with age, occurring predominantly between the fifth and seventh decades of life and in people in the lower socioeconomic groups (social classes III and IV). In places of high incidence the peak incidence tends to earlier than in low-risk areas. Stomach cancer is universally more common in men than women. In the United Kingdom the male:female ratio is 1 for young adults, rising to 2:1 in the sixth decade and to 1.5:1 in older people. The variations in geographical incidence are not irrevocable, as environmental factors play an important part. First-generation migrants sustain the risk rate of their country of origin, but the mortality rate in subsequent generations of Polish and Japanese immigrants to the United States, for example, falls to an intermediate level.

Site of gastric carcinoma

Together with the fall in the incidence of gastric carcinoma in the United States and United Kingdom, the tumour is moving proximally, with a rising incidence of adenocarcinoma of the cardia. In one series from in the United Kingdom that compared a period from 1951 to 1955 with that from 1981 to 1985, the number of carcinomas reported in the upper third of the stomach rose from 17 per cent (1951–55) to 39 per cent (1981–85). This was accompanied by a reduction in middle-third (from 31 to 19 per cent) and lower-third (from 51 to 47 per cent) tumours. Similarly, in Japan, proximally located cancer has increased from 10 to 40 per cent of diagnosed tumours. There is some evidence that tumours in the fundus are more aggressive, with a greater tendency to submucosal invasion regardless of the histological type. The gross morphological difference in the fundus that may explain this is the thinner muscularis mucosae. In addition, signet-ring carcinoma (see below) is more common in the fundus. In general, proximally placed tumours have a worse prognosis than those sited distally. This phenomenon may be as a result of the features described above or because proximally placed tumours appear to be at a more advanced stage at presentation.

Aetiological factors

Genetic

There is clear evidence of familial clustering of carcinoma of the stomach, the most famous being that of the Bonapartes: Napoleon, his father, his grandfather, brother, and three sisters died of gastric carcinoma. Some 10 per cent of patients have a family history of the disease, and there is slightly greater disease correlation between identical rather than fraternal twins.

Familial gastric cancer was first reported in New Zealand, and recent investigations from that country have identified significant germline mutations. E-cadherin, a cell-adhesion molecule, was poorly expressed in those patients who developed early, poorly differentiated, diffuse gastric cancer. Sequencing of the gene revealed a G–T nucleotide substitution leading to a defective gene product. Loss of the E-cadherin function disrupts normal cell adhesion, which may initiate cell proliferation. The molecule is transmembrane and the cytoplasmic portion modulates Wnt signalling, inhibiting b-catenin in a similar fashion to that of the *APC* gene.

Aird, in 1953, described an association between blood-group A and gastric carcinoma. The relative risk over patients with blood group O is 1.2 times. This difference has been related to the nature of mucopolysaccharide secretion in the stomachs of group-A individuals, and greater susceptibility to ingested carcinogens. The tumour that develops is of the diffuse type, there being no increased incidence of the intestinal type (these types are described below).

Environmental and dietary

Although genetic factors may be important, environmental and dietary agents are of far greater importance. Various foodstuffs have been implicated aetiologically, predominantly low-quality diets poor in milk, animal protein, and vitamins but rich in starch. Heavily salted pickles favoured in Japan (*tsukemono*) have been implicated, as have the smoked fish and meats eaten in Iceland and Scandinavia. These smoked foods contain polycyclic hydrocarbons (1,4-benzpyrene), which are the probable carcinogens. The decline in the incidence of the tumour in the United States has been attributed to the widespread use of refrigeration, with a decline in food preservation by smoking or salting.

Much interest has been devoted to the role of nitrosamines, which experimentally are carcinogenic to the gastric mucosa in animals. It is suggested that in man ingested nitrates and nitrites present in some high-protein diets, as food preservatives, or in water and soil may be converted to nitrosamines (*N*-nitroso compounds) by

the action gastrointestinal bacteria. The presence of atrophic gastritis and the associated achlorhydric stomach might predispose to the production of *N*-nitroso carcinogens. High gastric pH encourages bacterial overgrowth in the stomach; the organisms are able to reduce dietary nitrates and convert the nitrites to carcinogenic nitrosamines in the presence of dietary protein. This proposed sequence may only be one factor, and defects in the mucosal barrier associated with atrophic gastritis are likely to facilitate the penetration of carcinogens.

Other ingested agents implicated are the neat spirits favoured in some Scandinavian countries, also Japanese *saki* and contaminated whisky. Cigarette smoking, and elevated concentrations of zinc and lead in drinking water have been implicated, as have talc and asbestos in the atmosphere. Several groups of workers have been shown to be at high risk, including those in the metal, ceramic and clay industries, painters, printers, and fishermen.

There is a clear link between high consumption of fruit and a low incidence of gastric cancer. It is felt that the it is vitamin C which may be the protective agent. The normal stomach can concentrate vitamin C in gastric juice, but infection with *Helicobacter pylori* depletes vitamin C. Ascorbic acid, formed in the stomach by the reduction of vitamin C, is a remarkable antioxidant. Large amounts are required in the stomach, which is an area of great oxidative stress. Reactive oxygen species produce disruption of DNA strands, point mutations, and chromatid exchanges, which can result in the activation of oncogenes or the inactivation of suppressor genes. Currently the World Health Organization (**WHO**) recommends an increase in fruit and vegetable consumption in the general population.

Premalignant conditions

In this group are included conditions that, if untreated, become malignant, as well as disorders of the stomach that may predispose to gastric cancer. In particular, patients with hypogammaglogulinaemia (50-fold excess) and pernicious anaemia (3-fold excess) are at high risk.

Helicobacter pylori, chronic atrophic gastritis, intestinal metaplasia

Intestinal metaplasia of the gastric mucosa is a common finding in association with gastric cancers, and epidemiological studies have confirmed that populations with a high incidence of carcinoma of the stomach also have a high incidence of intestinal metaplasia. Mucosal atrophy and intestinal metaplasia are common phenomena, their incidence increasing with increasing age, and so they are particularly common in elderly people. It is becoming clear that the intestinal type of gastric cancer arises in gastric mucosa that has undergone a sequence of mutations and defined histopathological changes that may start in the first decade of life. The first lesion is atrophic gastritis, followed by progressive intestinalization of the mucosa to intestinal metaplasia, then dysplasia, and finally carcinoma. *Helicobacter pylori* is the predominant cause of chronic and atrophic gastritis. Atrophic gastritis is a major risk factor for gastric cancer.

The link between *H. pylori* infection and gastric cancer is now convincing. The WHO Consensus Committee has concluded that there is sufficient evidence of the aetiological connection. Virtually all infected patients will develop gastritis, leading to destruction and atrophy of gastric glands and other specialized cells, which are replaced by atrophic mucosa with intestinal metaplasia. The prevalence of atrophic gastritis increases by 2 per cent annually in patients with *Helicobacter* infection, compared with 0.3 per cent in the uninfected. There are many confounding factors, since treatment with long-term acid suppression of patients who are infected causes an increase in gastritis and intestinal metaplasia. The hypothesis is that acid suppression changes the pattern of infection, with an antral gastritis being transformed to a gastric-body gastritis, further suppressing acid secretion, increasing inflammation, and leading to atrophic gastritis. A large case-control study in rural Japan showed, by unconditional logistic-regression analysis, that *H. pylori* infection was associated with an increased odds ratio of 11 for men and 7 for women for the development of atrophic gastritis. In this study, diet, smoking, and lifestyle were not related to atrophic gastritis. *H. pylori* infection also causes increased proliferation of gastric cells and impaired secretion of vitamin C, factors that are improved after eradication of the infection.

The particular subtype of *H. pylori* infection is becoming important. Infected individuals who had CagA antibodies were 5.8-fold more likely to develop gastric cancer, whereas those without CagA antibodies had an odds ratio of 2.2. The most important strain has been characterized further as being CagAvacAS1a-positive. These bacteria increase cellular proliferation and also decrease apoptosis.

In the context of *H. pylori* and gastric malignancy, it is appropriate to mention primary low-grade lymphoma of the mucosal-associated lymphoid tissue (**MALT**) of the stomach. MALT lymphoma is rare, accounting for 1 to 5 per cent of stomach malignancies (see below). This tumour is strongly associated with *H. pylori* infection, and will regress completely or be cured by the eradication of the bacterium.

Patients with hypogammaglobulinaemia and pernicious anaemia have chronic atrophic gastritis. Certainly, achlorhydria results from the chronic atrophic gastritis and 75 per cent of patients with gastric carcinoma are achlorhydric. Strickland has divided chronic atrophic gastritis into two subgroups. Type A, which is associated with pernicious anaemia, predominantly affects the fundus and body, and is autoimmune in origin. Type B affects the antrum and is related to environmental factors. It is also found in the stomach some years after gastrectomy for benign peptic-ulcer disease. It may be regarded as a failure of the gastric mucosa to respond to repeated injury. Both types predispose to cancer formation.

Benign gastric and peptic ulcers

The relation between benign gastric ulcer and gastric cancer is controversial, the issue being whether benign chronic gastric ulcers have malignant potential or not. The suggestion is that the regenerating mucosa around an ulcer is prone to become malignant. The issue is further clouded by the fact that some ulcerating gastric cancers can closely mimic benign gastric ulcers, sometimes healing in response to medical treatment. Approximately 4 to 10 per cent of all gastric cancers behave in this way. In particular, 70 per cent of early gastric cancers may heal as part of the ‘lifecycle of the malignant ulcer'. The great improvements in endoscopy with biopsy have improved detection of both benign and malignant disease, and it now appears that the ‘ulcer cancer' (carcinoma developing in the edge of a benign gastric ulcer) is rare. There is no clear description of a proven be-nign ulcer turning malignant. The stable incidence of gastric ulcer and decline in incidence of gastric carcinoma supports the view that gastric ulcer is not a premalignant condition. Similarly, the location of benign and malignant disease is different, with most benign gastric ulcers (50–70 per cent) occurring on the lesser curve. It is now accepted that the most important question on finding a gastric ulcer is to decide whether it is benign or malignant from the outset. A history of duodenal ulcer has emerged as having a significantly protective effect for gastric cancer.

Gastric polyps

Gastric polyps are found in 0.5 per cent of autopsies. Most (65–90 per cent) are hyperplastic polyps that are regenerative, non-neoplastic lesions and usually smaller than 2 cm. Only two patients have been reported in which a carcinoma was found in association with a hyperplastic polyp. There is therefore no strong association with gastric carcinoma and these polyps should not be regarded as premalignant.

In contrast, adenomatous polyps are truly premalignant. They are often larger (80 per cent greater than 2 cm), and are tubulovillous or villous on microscopic examination. The frequency of malignant change increases with increasing size. In a large series, 38 per cent of patients with gastric adenomatous polyps had gastric carcinoma. Similarly, 34 per cent of postgastrectomy specimens for gastric cancer contain adenomatous polyps, and severe dysplasia with carcinoma *in situ* has been found in over 20 per cent of removed polyps.

Previous gastric surgery

As early as 1922, Balfour reported a gastric cancer occurring in the residual stomach after surgery for benign peptic-ulcer disease. The term ‘stump cancer' was soon used, as carcinoma seemed to occur more frequently after Billroth-I and -II gastrectomy than after vagotomy with pyloroplasty or gastroenterostomy. A large autopsy study demonstrated that gastric cancer was less frequent in patients who had undergone gastric surgery 15 years before death, but six times greater in those who had had surgery 25 years earlier. In a historical, prospective, cohort study in Denmark from 1955 to 1982, it was confirmed that the risk of gastric cancer immediately, and for 15 years after, surgery for peptic ulcer was less than expected. This was explained by the fact that patients have less gastric mucosa to be exposed to carcinogen after gastrectomy. However, the risk of cancer was 2.1 times greater than in the general population 25 years after surgery. The greatest risk (3.2 times) was in men who had had a Billroth-II gastrectomy. The patients who had simple suture of a perforated ulcer had no increased risk, indicating that peptic-ulcer disease was not a risk factor. The pathogenesis of gastric ‘stump cancer' has been shown experimentally to be related to operations that promote duodenogastric bile reflux, achlorhydria, and atrophic gastritis. Overall, the risk of developing gastric cancer following gastrectomy is reportedly between 3 and 10 per cent. The decline in surgery for peptic-ulcer disease means that there is every prospect that in 20 to 30 years ‘stump cancer' will have become a rarity.

Menetrier disease and hyperplastic gastropathy

Gastric carcinoma has been described as a complication of Menetrier disease, but the magnitude of risk is unknown. In Menetrier disease there is giant hyperplasia of the gastric mucosal folds and the condition can be difficult to distinguish from gastric polyposis or lymphoma. The mucosal abnormality associated with Menetrier disease results in hyperplasia of mucous glands, whereas the parietal cell mass falls. Thus the secretion is rich in protein and mucus but often hypochlorhydric. Hypersecretory conditions, including Zollinger–Ellison syndrome, may be associated with hyperplastic rugal folds and excessive acid secretion without increased risk of

gastric cancer.

Pathology

Macroscopic appearances

Advanced gastric cancer

There are very marked variations in the gross appearance of advanced gastric cancers. In the Western literature they have been described by the appearance of operative and excised specimens, and by the Japanese using the endoscopic descriptions of predominantly early gastric cancers. Approximately 10 per cent of tumours are polypoid fungating, with a nodular polypoid surface with superficial ulceration (Fig. 1). This group has a relatively good prognosis after aggressive surgical management, often being well-differentiated adenocarcinomas. Ulcerating or penetrating cancers occur in more than 50 per cent of patients (Fig. 2). The tumour is sessile and may look like a benign gastric ulcer. Superficial spreading carcinoma is more unusual (6 per cent of tumours): it is diffusely infiltrative over a wide area, and, although it is predominantly confined to the mucosa and submucosa, 50 per cent have metastasized to the perigastric lymph nodes at the time of presentation and operation. A further subgroup of diffusely infiltrative cancer is the linitis plastica (leather bottle) carcinoma (Fig. 3), characterized by extensive infiltration of the submucosa and muscular layers with a marked fibroblastic/desmoplastic reaction around columns of malignant cells. Endoscopic recognition of this tumour can be difficult and superficial, mucosa-only biopsy may prove negative. Extension into the oesophagus and mesentery may occur, and in advanced tumours the entire stomach can be involved. Superficial erosions are common but deep ulceration is unusual. Although it has been taught that spread of gastric cancer beyond the pyloric ring is rare, up to 25 per cent of diffusely infiltrative tumours may involve the first part of the duodenum; this pattern carries a particularly poor prognosis.

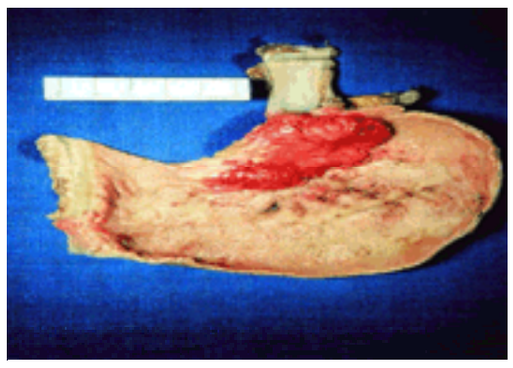


Fig. 1. Polypoid carcinoma of the gastric cardia causing oesophageal obstruction.

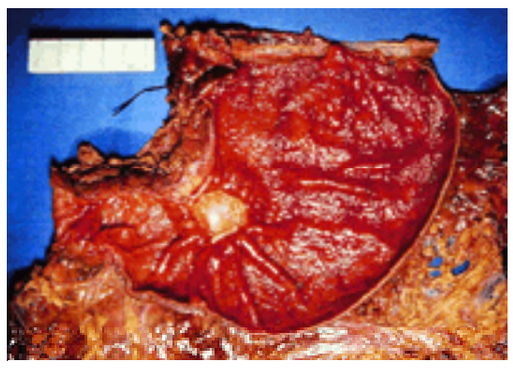


Fig. 2. Ulcerating cancer of the gastric antrum.



Fig. 3. Linitus plastica carcinoma of the stomach.

Precise description of the morphology of advanced gastric cancer by the Japanese has defined three types according to strict morphovolumetric types. The ratio of the amount of muscle invasion to mucosal involvement gives (i) a funnel type (mucosal greater than muscle involvement; ratio less than 0.75), (ii) a column type (equal involvement; ratio 0.75 to 1.25), and (iii) a mountain type (muscle greater than mucosal involvement; ratio greater than 1.25). The metastatic characteristics of these tumours are variable. The funnel type has the best prognosis, with 62 per cent lymph-node involvement, followed by the column type (80 per cent lymph-node metastasis), and finally the mountain type (lymph-node metastasis in 85 per cent).

Early gastric cancer

The great interest in the detection of early cancer by endoscopy has produced a separate classification based on endoscopic appearances. The term ‘early gastric cancer' is used to describe tumours confined to the gastric mucosa and submucosa, irrespective of nodal status, and was elaborated in 1962 by the Japanese Society of Gastroenterological Endoscopy. The close correlation between the depth of invasion at microscopic examination and postoperative survival has defined this entity. Early carcinoma of the stomach is divided into three main groups: type I, protruding, type II, superficial, and type III, excavated (Fig. 4). Type II is divided into three subgroups: (i) elevated, (ii) flat, and (iii) depressed. There is undoubtedly overlap between protruded (I) and superficial elevated (IIa); also between excavated (III) and superficial depressed (IIc). As well as these basic types there are three that exhibit combined features of two different types: type I and IIa (III), type IIa and IIc, and type IIc and III.

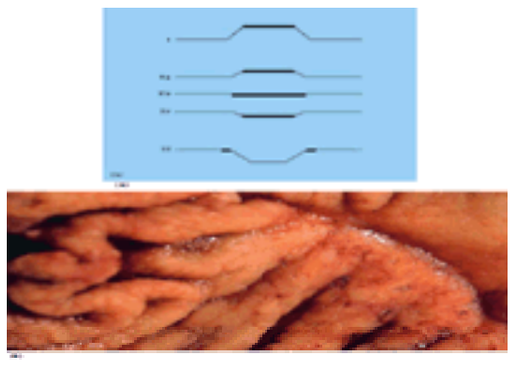


Fig. 4. (a) Schematic diagram of the Japanese classification of early gastric cancer; (b) resected early gastric cancer of the depressed type III, indicating the difficulty in detecting these tumours, and the importance of precise endoscopy.

The precise classification of early gastric cancer by the Japanese and the use of vigorous investigation has increased its detection from 5 to 40 per cent of all gastric cancers over the past 20 years. In Western series, early gastric cancer accounts for 9 per cent of all cancers detected and 19 per cent of resected tumours. The rate of detection does not appear to be increasing in the West.

Comparison of the clinical, morphological, and histological features of early gastric cancer reveals remarkable similarity in the disease between Japan, the United Kingdom, and the United States. The estimated median duration of early gastric cancer before becoming advanced is 37 months. Poor prognostic features include lymph-node metastasis and invasion of the muscularis mucosae. The surgical outcome is only marginally poorer in Europe than Japan. In Japan, gastric carcinoma is diagnosed on nuclear and structural criteria even when invasion is absent. Presence of invasion is the most important diagnostic criterion for most Western pathologists. An appreciation of this difference is important for the interpretation of cancer statistics. A united effort is required to reach consensus.

Recently, the definition of early gastric cancer has been challenged, since the 5-year survival is 99 per cent if there are no lymph-node metastases, but falls to 73 per cent if the lymph nodes are involved. It is suggested that tumours with lymph-node involvement should not be called early gastric cancers.

Multiple synchronous gastric cancer

Multiple synchronous gastric cancer was described by Moertel, who defined strict criteria for diagnosis. Each lesion must be proved histopathologically malignant, all lesions must be separated by normal gastric wall, and the possibility that the lesions represent a local extension or metastasis must be ruled out beyond reasonable doubt. It is sometimes difficult to verify this last criterion, and emphasis is laid on the extent of venous and lymphatic invasion surrounding each tumour. If there is extensive invasion a diagnosis of multiple gastric cancer is inappropriate. The incidence was originally thought to be 2 per cent of all gastric cancers, but with advances in diagnostic techniques it appears to be between 6 and 9 per cent.

Microscopic appearances

The great majority of gastric cancers are adenocarcinoma; other tumours are very rare. Classification is a difficult task, primarily because different histological features may coexist in one tumour mass. The WHO classification based on morphology divides gastric cancer into five types: adenocarcinoma, adenosquamous carcinoma, squamous-cell carcinoma, undifferentiated carcinoma, and unclassified carcinoma. The adenocarcinomas are divided into four patterns, papillary, tubular, mucinous, and signet ring, which may divided further by their degree of differentiation. In tumours presenting a mixed picture with varying degrees of differentiation in the same lesion, classification is based on the predominant type.

A simpler classification by Lauren recognized an intestinal type, similar to colon adenocarcinoma, and a diffuse type that did not form glands. The intestinal type is defined by its cellular architecture, and the diffuse by its pattern of growth. Polypoid and superficial spreading carcinomas were of the intestinal type, whereas linitis plastica was diffuse. Ulcerative tumours fell into either group. The intestinal type carries the more favourable prognosis and is more common in areas with a high incidence of the disease. In general, the intestinal type is said to arise within intestinalized mucosa and the diffuse type from normal epithelium without evidence of a preceding lesion. Synchronous cancers are more commonly of the intestinal type. Some carcinomas fall into a mixed or unclassified category.

A similar classification for advanced carcinomas invading beyond the muscularis propria was proposed by Ming in 1977. They were divided into an expanding type, which produces nodules that compress adjacent tissue (similar to Lauren's intestinal type), and an infiltrative type, which does not form masses (Lauren's diffuse type). The best concordance for tumour classification by different pathologists examining the same specimens is with Lauren's scheme and this is the most widely used.

Considerable interest surrounds the prognostic significance of histological evidence of host defence against the tumour. Infiltration of the tumour by macrophages, polymorph leucocytes, lymphocytes, and plasma cells is a favourable sign. Similarly, evidence of a host response in draining lymph nodes (sinus histiocytosis and follicular hyperplasia) is a good sign associated with extended survival and correlates with lymphoid infiltration of the primary tumour.

The microscopic appearance of early gastric cancer is strongly correlated with the macroscopic appearance (Lauren classification). Types I and IIa are of the intestinal type whereas types IIc and III are all diffuse. This is not surprising since intestinal-type carcinomas have an expansile growth pattern leading to polypoid growth. The prognosis of early gastric cancer is predominantly related to the depth of invasion and this greatly supersedes any histological classification.

One of the characteristics of gastric cancer is its tendency to spread intramurally via lymphatic channels. Thus the surgical resection margins must be at some distance from the palpable edge of the tumour. The extent of intramural spread has been related to the depth of invasion of the tumour: if the tumour was confined to the muscularis propria, its lateral extension was not greater than 3 cm, but if this layer was penetrated, significant numbers of tumours had intramural spread beyond 3 cm but never greater than 6 cm.

Gastric lymphoma and ‘maltoma'

Between 2 to 8 per cent of gastric malignancies are lymphomas. The initial stage preceding lymphoma is the excessive development of MALT (‘maltoma'). The concept of MALT was introduced to distinguish mucosal from nodal lymphoid tissue. The condition is related to *H. pylori* infection and treatment by eradication of the organism is successful in producing complete histological remission. MALT is an immunological defence system to control the local infection. It is clear that the preneoplastic cells are proliferating in response to an antigenic stimulus that is T-cell dependent. The proliferation is composed of reactive T cells, plasma cells, and B cells, and mimics the appearance of lymphoid follicles in lymph nodes. Transformation from this low-grade tumour to high-grade lymphoma occurs when the mass becomes dominated by large blast cells.

Molecular biological factors

The transformation of normal to malignant invasive cells is associated with multiple gene abnormalities. In gastric cancer, oncogenes coding for tyrosine kinase receptors (c-*met*, c-*erb*B2), epidermal growth factor, intracellular signal transduction (*ras*), and tumour suppressor genes (*p53, APC*) have been identified as of potential significance. Tumours also have altered expression of the adhesion molecule E-cadherin, which may be of critical importance in the dysplasia–carcinoma sequence. Nuclear and non-cell-surface localization of catenins occurs in tumours where E-cadherin is reduced and cytoplasmic (Fig. 5). Inhibited expression of this molecule may be mediated by c-*erb*B2. In addition, the *bcl-2* gene, which codes for a cell-membrane protein that blocks apoptotic cell death, is associated with poor prognostic features in gastric cancer.

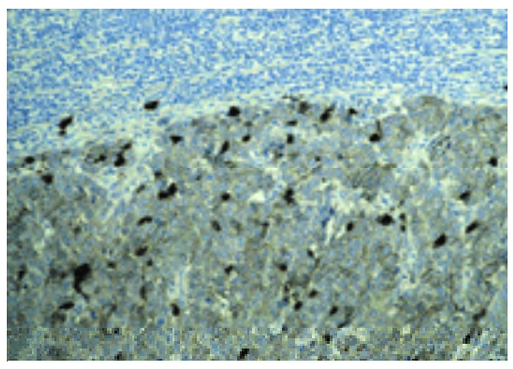


Fig. 5. Immunohistochemical staining of adenocarcinoma to demonstrate disordered localization of catenins (brown stain). Nuclear localization of catenins occurs when the expression of E-cadherin is reduced and cytoplasmic.

Patterns of spread and metastasis

In Western series, between 40 to 60 per cent of patients have obviously incurable, often disseminated, disease at presentation. In Japan, the number who can be offered curative resection with excision of all macroscopic tumour is much higher (75–80 per cent). The poorly differentiated mucinous and signet-ring tumours are more invasive, although in series from the United Kingdom most tumours are large and advanced at presentation, and have spread, so that the degree of differentiation is of little relevance. In Japan, up to 40 per cent of patients have early gastric cancer with tumour confined to the mucosa and submucosa; those with poorly differentiated carcinomas are more prone to metastasis and have decreased survival.

Penetration of the gastric serosa

The depth of invasion of the cancer through the gastric wall has, as outlined above, a marked effect on the prognosis. If the muscularis propria and then the serosa are breached, the prognosis following treatment becomes considerably worse. If the serosa is not penetrated, lymph-node metastasis occurs in 18 per cent of patients and over 50 per cent survive for 5 years after resection. If the serosa is penetrated, 80 per cent of patients have lymph-node metastases, with a correspondingly poorer prognosis. Once the serosa of the stomach has been breached the tumour is likely to spread by transcoelomic implantation of shed cells. If the area of serosa involved is less than 10 cm^2, viable malignant cells can be found free intraperitoneally in 22 per cent of patients, compared with 72 per cent if the area involved is greater than 20 cm^2. These cells characteristically become implanted in the ovaries, producing Krukenberg tumours (bilateral ovarian tumours from a signet-ring carcinoma), or in the pelvis, producing a shelf-like mass palpable rectally (Blummer's shelf). It must be emphasized that the penetration of the gastric serosa is the most important prognostic indicator.

Gastric tumours may spread by direct extension and invasion of adjacent structures including liver, pancreas, and spleen. In autopsy studies from the United States, the peritoneum, mesentery, and omentum were invaded in over 40 per cent.

Lymphatic drainage and lymph-node involvement

The involvement of lymph nodes and the pattern of nodal spread are particularly important for gastric cancers. In Western series, some form of lymph-node involvement is present in 70 per cent of resected tumours. Even in Japanese series, intramucosal cancer is associated with a 2 to 4 per cent incidence of lymph-node involvement, rising to 14 to 20 per cent if the submucosa is involved. The anatomical classification of lymph nodes varies in the West and Japan, and confuses the interpretation of the results of surgery. In the Western literature, four zones of lymph drainage corresponding to the blood supply are identified. Zone 1 is located in the gastrocolic omentum along the right gastroepiploic vessels, draining the pyloric portion of the greater curve, and draining to the pylorus and then to coeliac and aortic nodes. Zone 2 is in the gastrocolic and gastrosplenic omentum around the left gastroepiploic vessels, draining from the upper half of the greater curve. From here drainage is to the pancreaticosplenic lymph nodes and to the aortic nodes. Zone 3 has efferent channels from the proximal two-thirds of the stomach and the upper lesser curve, and surrounds the left gastric artery. Some lymph from this area drains into perioesophageal lymph nodes. The distal portion of the lesser curve and pylorus drains to zone 4 situated above the pylorus, going to the hepatic artery and para-aortic lymph nodes. However, according to Western data, the lymphatic drainage of the stomach is unpredictable and radical resection must included lymph nodes from all zones.

The Japanese have adjusted and stratified the classification of lymph drainage (Fig. 6). Group I (N1) are perigastric lymph nodes; group II (N2) are nodes along and at the roots of the major vessels; group III (N3) are lymph nodes at the root of the superior mesenteric artery, in the hepaticoduodenal ligament, and behind the pancreas; and group IV (N4) are distant lymph nodes. It is clear that this classification may not be adequate for planning surgical resection along strict oncological principles, since a perigastric node (N1) at the antrum is a distant node (N4) if the cancer is in the cardia. Thus the lymph nodes can be given station numbers (Table 1), and the node status and surgical excision planned according to the node station in relation to the site of the primary tumour (Table 2). The Japanese have found that there is an orderly spread to regional lymph nodes clearly related to the position of the D. This finding has not been confirmed in Western studies. The difference may, in part, relate to the difference between the advanced presentation of gastric cancer in the West and the greater incidence of potentially curable patients identified in Japan. However, skip lesions to distant lymph nodes do undoubtedly occur occasionally: in 11 per cent of patients, N2 nodes are involved without N1 (perigastric) nodal involvement.

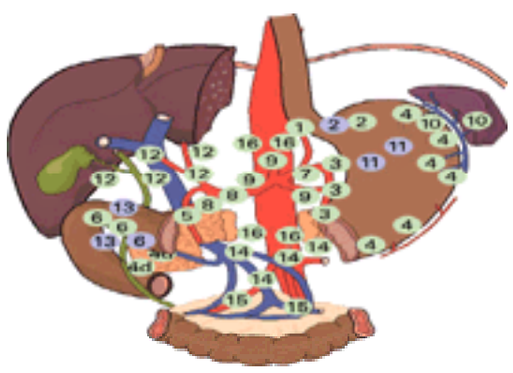


Fig. 6. Lymph-node stations defined by the Japanese Research Society for gastric cancer.

Station no.	*Anatomical location*
1, 2	Adjacent to the cardia (perigastric)
3, 4	Adjacent to lesser and greater curve
5	Suprapylorus (right gastric artery)
6	Infrapyloric
7	Left gastric artery
8	Common hepatic artery
9	Coeliac artery
10	Hilum of the spleen
11	Splenic artery
12	Hepaticoduodenal ligament
13	Behind pancreatic head
14	At the root of the mesentery (superior mesenteric artery)
15	Middle colic artery
16	Para-aortic

Table 1 Numbering of the gastric and upper abdominal node stations

Node number	Location of gastric cancer			
	Antral/body	Body	Cardia	Total stomach
1	N2	N1	N1	N1
2	N3	N2	N1	N1
3, 4	N1	N1	N1	N1
5, 6	N1	N1	N2	N1
7–9	N2	N2	N2	N2
10, 11	N3	N2	N2	N2
12–14	N3	N3	N3	N3
15, 16	N4	N4	N4	N4

Table 2 Node stations related to location of the tumour to define nodal status and extent of nodal dissection

It is important to remember that carcinoma at the cardia presents a distinct problem, for it may involve lymph nodes in the mediastinum. In a review of over 400 patients, involvement of paracardial nodes occurred in 48 per cent and of paraoesophageal nodes in 37 per cent of patients.

In general, any involvement of the lymph nodes is a poor prognostic indicator, as is the number affected. Invasion of four or more lymph nodes was prognostically less favourable than involvement of lesser numbers, though this conclusion implies that an adequate lymphadenectomy has been performed to sample these nodes. The number of lymph nodes to be found at each station is very variable. If stations 1 to 11 are removed (D_2 gastrectomy; see below), an average of 27 nodes should be removed. If stations 1 to 16 are cleared (D_3 gastrectomy; see below), the mean number rises to 43.

Distant metastasis

The most common sites for distant metastatic spread are the liver (49 per cent), lung (33 per cent), ovary (14 per cent), bones (11 per cent), and cervical and supraclavicular (Virchow's) nodes (8 per cent). Patients with distant metastases have a very poor prognosis: 95 per cent of patients with liver involvement will die within 1 year irrespective of the treatment of the primary cancer.

Staging for gastric carcinoma

As with all tumours a uniform staging system is required to allow the results of treatment to be compared. The TNM system of staging can be applied to gastric cancer and is summarized in Table 3. Another, widely used, staging system based on the results following excision and pathological examination is somewhat simpler (Table 4). In tumours that are locally resectable the most important prognostic sign that can be assessed at laparotomy is penetration of the wall of the stomach by the carcinoma.

T stage	*Tumour status in relation to penetration of gastric wall*
T1	Confined to mucosa
T2	Involving all layers to serosa
T3	Penetrating serosa, with or without direct invasion of adjacent tissues and organs
T4	Diffuse infiltration of gastric wall
TX	Degree of involvement unknown
N stage	*Lymph-node involvement*
N0	No lymph nodes involved
N1	Perigastric nodes adjacent to tumour involved
N2	Metastasis to nodes on both curvatures or regional areas
NX	Lymph nodal status unknown
M stage	*Distant involvement*
M0	No metastasis
M1	Distant metastasis, distant lymph-node involvement beyond regional lymph nodes

Table 3 Summary of TNM system for the staging gastric carcinoma

Stage	Definition
Ia	Tumour confined to gastric mucosa only (T1, N0, M0)
Ib	Tumour extending beneath muscularis mucosae (T2, N0, M0)
Ic	Tumour extending through gastric serosa with or without local invasion (T3, N0, M0)
II	Diffuse involvement of gastric wall (T4, N0, M0), or any involvement of gastric wall in association with tumour deposits in the perigastric lymph nodes (T1–3, N1, M0)
III	All diffuse of gastric wall with involvement of perigastric lymph nodes (distant from the tumour or to wide gastric resections) (T1–3, N2, M0)
IV	Tumour metastasis

Table 4 Staging system for gastric cancer

Clinical presentation

The major problem with gastric carcinoma is that the symptoms are often vague, non-specific, and attributed to dietary indiscretion or indigestion. By the time the diagnosis has been made, the tumour is often incurable or irremovable. This problem has been elegantly illustrated by Theodor Storm (1888), in his poem ‘Beginn des Endes' describing his death from advanced gastric carcinoma: ‘Tis a prick, ‘tis scarce a pain… indeed ‘tis naught;... it is too late'. Unfortunately, over 100 years later, this experience is still common. Definite symptoms do not usually occur until the tumour is large enough to obstruct the lumen, cause disordered gastric function by invading a large segment of the wall, or bleeds. Over 70 per cent of patients have had some symptoms for longer than 6 months before seeking advice. The most common symptom is vague indigestion or upper abdominal/epigastric pain, followed by weight loss, nausea and vomiting, haematemesis and melaena, profound anorexia, early satiety, and flatulence. The pain may mimic angina pectoris, exhibit periodicity, and be relieved by food mimicking benign gastric disease. Interestingly, true postprandial pain is relatively unusual in patients with gastric carcinoma. If the tumour is at the cardia, over 60 per cent of patients will present with dysphagia, indicating a greater than 80 per cent obstruction of the oesophageal or gastric cardia lumina. Weight loss may be insidious, and, although most patients have lost significant amounts, few complain of this directly.

Early gastric cancer presents with similar symptoms of dyspepsia. Indeed, purely mucosal gastric cancers are symptomatic in 50 per cent of patients, and early endoscopy is advised for patients over the age of 40 years with persistent dyspeptic symptoms. If these patients have dysplastic changes, then regular follow-up endoscopy is necessary. Some recommend resection if there are severe dysplastic changes in young, otherwise healthy, patients. The duration of symptoms before surgery in those with early gastric cancer ranges from 3 to 72 months; surprisingly this is little different from those who present with advanced cancers. There are generally few physical signs in patients with early gastric cancer, although epigastric tenderness is present in 10 per cent, with the remainder having a normal physical examination. It is not possible to be highly specific about important symptoms. It is clear that there is no correlation between any symptom and the extent of the disease. The presence of an epigastric mass is a poor prognostic sign, but is not in itself an indication of inoperable disease.

Approximately 10 per cent of patients in Western series present with palpable cervical lymph nodes, ascites, jaundice, and a palpable abdominal or pelvic mass. Sister

Joseph's nodule, a visible and palpable secondary deposit at the umbilicus due to spread along the lymphatics around the falciform ligament, is a not uncommon presentation of advanced disease. In patients with this sign, the primary tumour is most commonly in the ovary, followed by the stomach and colon. It is a poor prognostic sign and the mean survival if the primary lesion is gastric is only 3.5 months. Troisier's sign of an enlarged lymph node in the left supraclavicular fossa (Virchow's node) indicates lymphatic spread via the thoracic duct. Armand Trousseau (1801–1867), a Paris physician, first suspected he had a gastric cancer when he developed superficial thrombophlebitis on the legs (Trousseau sign), but this is also associated with pancreatic cancer.

Investigation

Laboratory

Routine biochemical and haematological tests may reveal anaemia, which is common in these patients. It is usually of an iron-deficiency type, microcytic and hypochromic, reflecting blood loss. Anaemia is not only a feature of advanced disease; 20 per cent of patients with early gastric cancer will be anaemic. Testing of the stools for occult blood may well be positive. Approximately two-thirds of patients with gastric cancer will have achlorhydria, but this is of little diagnostic relevance. The concentration of carcinoembryonic antigen in the serum is elevated in 30 per cent of patients with advanced tumours but testing for this antigen fails to detect early disease. Similarly, oncofetal antigen is elevated in some benign inflammatory gastric disorders as well as in malignancy.

Radiography

The mainstay of diagnosis for many years was barium contrast studies of the upper gastrointestinal tract. In Japan, the double-contrast, air–barium study has been perfected and used for mass screening. Air and barium are introduced together to coat the mucosa with a thin layer of barium and enhance mucosal detail. The technique can be further refined by using high-density barium, carbon dioxide, and simethicone for gas dispersion, and glucagon to induce gastroparesis. In Western countries, endoscopy is more sensitive than radiography, but in some Japanese reports using modern photographic techniques the diagnostic accuracy of these methods is similar. It is clear that endoscopy and careful examination of the gastric mucosa is essential for early diagnosis. Patients with persistent dyspepsia should have endoscopy. With schemes for open-access endoscopy the availability of this investigation is increasing. It will be some time before any impact of this service becomes clear.

Computed tomography (**CT**), ultrasonography, and magnetic resonance imaging are most helpful in the diagnosis of metastatic disease. These methods identify enlarged lymph nodes and tumour deposits, but do not differentiate between tumour and benign changes. In general the presence of hepatic metastatic deposits should be confirmed by biopsy. In early gastric cancer, metastases are usually confined to the perigastric lymph nodes, which are difficult to detect. Attempts to improve the preoperative accuracy of detecting lymph-node involvement include immunolocalization with monoclonal targeted isotopes, endoscopic lymphography, endoluminal ultrasonography, and dynamic CT.

Endoscopic ultrasonography will visualize perigastric lymph nodes greater than 3 mm in diameter in 70 per cent of patients. If the examination is preceded by the oral administration of an oil-in-water emulsion, uninvolved nodes can be identified by echo enhancement at their margin; metastatic nodes show no enhancement. The sensitivity of this test in Japanese hands is as high as 92 per cent, with a specificity of 100 per cent for nodes greater than 3 mm. Endoscopic lymphography following submucosal injection of contrast can identify metastatic filling defects in involved lymph nodes. Dynamic CT has only slightly improved the diagnostic accuracy of conventional CT, and there are some lymph-node regions that cannot be seen clearly.

Endoscopy and biopsy

The development of fibreoptic, flexible, forward-viewing endoscopes with a controllable tip has been a major advance. It allows direct visualization of the stomach and accurate biopsy of any lesion identified. Endoscopy is observer-dependent, and it has been clearly shown that visual endoscopic interpretation, even in experienced hands, is often incorrect in deciding on the nature of a lesion in over 10 per cent of cases. If up to 10 biopsies were taken from each lesion, the diagnostic accuracy was 100 per cent. The accuracy of endoscopic biopsy diagnosis is now clearly shown to be related to the number of biopsies taken. The working recommendation is that four to six biopsies be taken from each lesion, and from the inner border of the edge of any ulcer.

Improvements in endoscopic techniques include dye-spraying, fluorescence endoscopy, magnified endoscopy, electronic endoscopy, and endoscopic ultrasonography. This last technique can differentiate early gastric cancer from advanced tumours in 80 per cent of patients, but has so far failed to allow preoperative differentiation of mucosal from submucosal cancer because of the fibrosis in the muscularis propria and submucosa that may surround ulcerated, early gastric cancers.

Cytology

Cytological study of gastric aspirate for exfoliated malignant cells has been used in patients with advanced disease, but has a variable accuracy of between 40 to 90 per cent. Methods of collecting the cytological specimen include gastric washings, washings with the addition of a mucolytic agent, and the passage of a balloon. Cytological analysis of brushings directly collected from suspicious lesions can alone be accurate in 81 per cent, but when combined with biopsy has an improved accuracy of 91 per cent. The principal limitation of both cytology and biopsy is failure to obtain an adequate sample. Problems of interpretation may be posed by the atypical cells associated with regeneration around an ulcer. Immunocytochemical stains such as fetal sulfaglycoprotein may be useful for cytological preparations.

Laparoscopy and laparotomy

Laparoscopy can be valuable as an initial operative assessment to exclude widespread disease, since hepatic, peritoneal, ovarian, or extensive local disease may be assessed. Laparotomy is necessary in all patients with local disease, as it is often the only method that will detect peritoneal involvement before laparotomy. The presence of extensive nodal disease excludes patients from curative surgery. Laparoscopy of the lesser sac is rather difficult and time-consuming but may reveal advanced, incurable disease. Some advocate biopsy and frozen section of a lymph node from the infracolic, periaortic region before commencing radical resection. In Japan, at initial laparotomy, peritoneal washings are taken for cytology, and lymph-node sampling by frozen section is performed before deciding whether a curative resection is feasible. Locally invasive disease may be excised curatively only if it is minimal. Adequate staging of the extent of a carcinoma of the stomach is important not only to minimize unnecessary laparotomy but also to select patients for possible adjuvant therapy.

Population screening

As the postsurgical survival of patients with early gastric cancer is greater than 90 per cent, screening of the asymptomatic population has been implemented in Japan, Chile, and Venezuela. In Europe, screening has been focused on high-risk groups, while others advocated a low threshold for endoscopy in patients with dyspepsia. In Japan, government-subsidized mass screening has been in place for some years. The methods used are predominantly radiological, with photofluorographic radiology performed at designated centres or by mobile units. Approximately 5 million people are screened per year. After 25 years of screening, deaths from gastric cancer have decreased from over 50 to 33 per cent of all deaths caused by malignant disease in men. Deaths in women have been reduced from 38 to 28 per cent. The incidence/detection of early gastric cancer has risen from under 2 per cent in 1945 to 63 per cent in some centres. The overall 5-year survival from gastric cancer has risen to more than 50 per cent. There are clear shortcomings in using survival rate in evaluating the effectiveness of screening. In particular there is concern over the definition of early gastric cancer and its distinction from dysplasia or ‘worrying mucosal appearance'. Also, most studies report 5-year survival and there may well be a lead-time bias. Recently, a detailed follow-up has demonstrated that two-thirds of patients with screening-detected cancer in Osaka, Japan, were cured of their disease 15 years after resection.

Screening of asymptomatic patients has not been considered feasible elsewhere. In Japan, 25 per cent of cancers are identified by screening programmes. Despite the wider use of the flexible endoscope in the West, the detection of early (stage I) cancers has not increased. However, it is still recommended that patients with dyspepsia have access to early and prompt endoscopy as the only method for the detection of early, and potentially curable, gastric cancer.

Screening of certain high-risk groups is feasible. Some centres have introduced programmes to allow early endoscopy of dyspeptic patients over the age of 40 years, and serial endoscopy until healing of all patients with gastric ulcer. Patients with atrophic gastritis, dysplasia, or adenomatous polyps are offered endoscopy on an annual basis. Among 2659 patients examined in this scheme, 57 cancers were identified; 20 per cent were early gastric cancers but overall only 60 per cent were suitable for attempted curative surgery.

The excellent results of surgery in Japan have fuelled debate over whether the disease in Japan is different from that in the West. In all countries, stage-1 disease is associated with a relatively good prognosis. Yet the overall 5-year survival in Europe, for this stage of disease, is reportedly 70 per cent, compared with 84 per cent in Japan (Table 5), suggesting differences in the biological behaviour of the tumour. Clarification of this issue arose from a detailed review of tumour classification. The British Society of Gastroenterology reviewed 319 patients from 41 hospitals who had histopathological findings of dysplasia and early gastric cancer. There was good agreement between pathologists on the difference between dysplasia and cancer, but over a third of patients thought to have early gastric cancer were found to have more advanced disease on reassessment. The true 5-year survival rate of the group redefined as having early gastric cancer was 90 per cent, compared with 75 per

cent in the group initially so defined. Thus European survival rates for the disease approach those in the East if the strict Japanese criteria are used. Unfortunately, the issue cannot yet be regarded as completely resolved. In Japan there is a higher incidence of tumours with a better prognosis and some tumours do behave differently in the East and West; but when direct comparisons are made between like tumours the results appear similar.

Stage	UK			Japan		
	Year			Year		
	1	2	5	1	2	5
I	86	82	72	95	93	84
II	72	55	32	90	85	65
III	52	26	10	80	60	35
IV	10	5	1	20	15	5

Table 5 Comparison of the 1-, 2-, and 5-year survival (per cent) in patients treated in the United Kingdom and Japan with a ‘curative' resection for stage I, II, and III

Treatment and outcome

Historical considerations

Historical references to what appears to be gastric carcinoma can be found in the writings of Hippocrates, Galen, and Avicenna. However, it was not until the nineteenth century that Jean Cruveilhier attempted to distinguish between benign and malignant gastric lesions, and gastric cancer was initial referred to as ‘la maladie de Cruveilhier' by French physicians. A treatise in 1839 by Bayle clearly described the detailed pathology of gastric cancer and suggested treatments. Resection was attempted by Jules Pean (France) in 1879 and one year later by Ludwig von Rydygier (Poland). Reconstruction was by direct anastomosis using catgut to form a gastroduodenostomy. Both operations were unsuccessful, the patients dying shortly after as a result of peritonitis from gastric leakage. Using the magnifying glass of retrospective examination, it would appear that an inadequate number of sutures was used. It has been reported that a French surgeon in Arras attempted gastric resection one year before Pean but details are not available. Christian Albert Theodor Billroth (1829–1894) believed a safe technique could be developed for the excision of gastric cancer and dispatched Carl Gussenbauer and Alexander Winwater to the autopsy room and the laboratory to investigate the possibility. They found, after examining the records of the Vienna Pathological Institute, that over 40 per cent of pyloric tumours were amenable to surgical resection and did not appear to have distant metastases. In the dog they developed the procedure of the Billroth-I gastrectomy; two of seven dogs survived for a prolonged period and two others died of anastomotic dehiscence.

Billroth's first gastrectomy for gastric cancer was performed on Frau Therese Heller in 1881. The procedure was done under chloroform anaesthesia using antiseptic precautions. The gastroduodenal anastomosis was made with carbolized silk sutures. Unfortunately, a few hazelnut- sized mesenteric lymph nodes were found and histological examination confirmed tumour involvement. The patient survived for 4 months, and died of recurrent disease.

Immediately, after the initial success of the Billroth-I operation, Wolfler (Billroth's assistant) was instructed by his superior to devise a palliative operation to relieve gastric-outlet obstruction for irremovable tumours. Thus the technique of bypass gastroenterostomy was devised in 1881, and used clinically later that year. In January 1885, Billroth performed a resection of the distal stomach, closed the transected duodenum and stomach, and restored continuity by a gastroenterostomy to the posterior wall of the stomach, the patient surviving for 18 months. Thus the Billroth-II operation was born (Fig. 7).

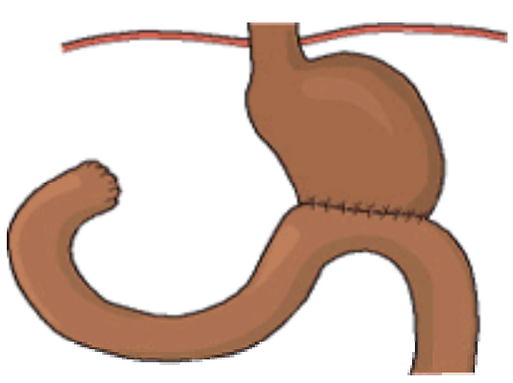


Fig. 7. The reconstruction by a Billroth-II or Pólya operation after resection of an antral carcinoma of the stomach.

By 1890, 41 resections for gastric cancer had been performed at the Billroth clinic, with a successful outcome in 19; by 1894 the overall mortality was reported as 55 per cent. These results prompted Welch in 1885 to write that no instances of prolonged recovery or cure followed operation for gastric cancer despite the ‘great sensation' produced by Billroth's operative achievements. Extended surgery was explored as a means of improving survival, with Carl Schlatter (1897) performing total gastrectomy with end-to-side oesophagojejunal anastomosis. Brigham (1898) also excised the entire stomach in a patient who survived for a remarkable 18 years and died of other causes. The fact that this patient survived longer than 3 years in an era before the identification of gastric intrinsic factor (in 1929 by W.B. Castle) when vitamin B_{12} supplements were not available could indicate that the gastrectomy might not have been total.

In Billroth's initial gastroduodenal anastomosis with partial closure of the gastric lesser curve, it was quickly identified that leaks occurred at the junction of the lesser curve with the duodenal anastomosis. This area was called the *jammer ecke*' (crying corner) and subsequently a further inverting triangular suture was often inserted in this area. It is appropriate to acknowledge the work of the Budapest surgeon Eugene Pólya (1876–1944), which was first recognized by William Mayo. Pólya devised the retrocolic anastomosis of the entire width of the gastric segment to the jejunum after gastrectomy (Fig. 7).

There are now many distinguished names associated with various forms of gastrectomy. There was an extended period of debate in the 1940s as to whether the jejunal anastomosis should be isoperistaltic, retroperistaltic, antecolic, retrocolic, with a long or short afferent loop, to the entire gastric stoma, or a whether a valve was necessary. Various modifications allowed the attachment of an eponym to modifications of gastrectomy, often with a strongly held, anecdotal belief in its superior merits. At present most surgeons favour the partial closure of the gastric stoma described by Hofmeister, with an antecolic anastomosis for reconstruction and a short afferent loop after a cancer resection. It appears that the precise form of reconstruction is of little relevance.

Oncological approach

Factors affecting prognosis

Since Billroth's first resection, gastrectomy has been the mainstay of therapy. However, the optimal surgical management is still a subject of intense debate. The controversy lies between those (predominantly Western) who believe that the pathological stage of disease is the crucial prognostic factor and others (Eastern) who are convinced that the extent and nature of the resection combined with adjuvant therapy are major factors in improving the results. There are undoubtedly predetermined factors that govern survival, such as the site of the tumour, and the extent of nodal and, in particular, serosal involvement, but the manner of execution of the resection can have a significant impact on the prognosis. Thus one view must not be exclusive of the other. The importance of early detection is now universally accepted, and, in the absence of distant metastasis, aggressive resection of the carcinoma is justified.

As outlined above, poor prognostic factors in relation to the presentation of the tumour are serosal invasion, presence of lymph-node metastasis, presence of free carcinoma cells in the peritoneum, Lauren's classification (intestinal type better than diffuse type), cardial tumour, histological evidence of the invasion of lymph

vessels, tumour stage, tumour depth and size, and the patient's age. The poor prognostic variables that relate to surgery are positive resection margins, inadequate extent of lymphadenectomy, and the need for an associated splenectomy. The single most important prognostic factor is serosal invasion. It is now clearly demonstrated that in the presence of serosal invasion, the number of lymph nodes affected or the extent of resection will not affect the prognosis.

The basic oncological approach for resection of mucosal cancers is wide excision of the primary tumour with *en bloc* removal of the draining network of lymph nodes. This approach has undergone re-evaluation for some tumours, with lesser resection in combination with other means of treatment providing a more conservative and equally as effective option. At present our understanding of the biological behaviour of gastric carcinoma, in particular with regard to the very limited effectiveness of radiotherapy and chemotherapy, means that radical excisional surgery probably offers the best chance of cure.

Radical and attempted curative surgery

The problems of definition of radical excision in the treatment of gastric cancer confuse the interpretation of the results of surgery. The main areas that must be addressed are the extent of gastric and of lymph-node resection.

Endoscopic local excision

The Japanese in particular have pioneered endoscopic local excision of early gastric cancer. The technique involves the use of an endoscopic snare, with pathological assessment of the adequacy of excision. As already indicated, some of the areas excised would be classified as dysplasia by Western pathologists. If the lesion is not amenable to removal by snare, the injection of saline beneath it may allow its removal. Laser ablation is an alternative, but fails to allow histological analysis; it can be used to treat tumours that penetrate deep into the submucosa. In a selected group of 145 patients with a median follow-up of 3 years, a survival of 98 per cent was reported. There are other locally ablative techniques, such as photodynamic therapy, that are of predominantly experimental interest.

Extent of gastric resection

Adequate gastrectomy implies surgical margins in the stomach free of tumour; thus gastrectomy may be partial or subtotal if the tumour is distal, or total if it is more proximal. The definition of adequate resection margins is an 8- to 10-cm clearance proximally and distally in the unstretched stomach. Failure to resect the stomach widely with microscopically clear margins is highly detrimental to survival. If the resection margin is not confirmed free of microscopic disease, the prognosis of a stage-II tumour falls to that of a stage IV. The specific operation for a gastric carcinoma depends somewhat on its site. There is a widespread belief that subtotal gastrectomy is associated with poor local control. However, there appears to be no advantage for total gastrectomy as opposed to adequate partial gastrectomy for local tumour. A prospective, controlled study comparing total and subtotal gastrectomy (with the same node dissection) for carcinoma of the gastric antrum demonstrated no difference in survival between the two. At the Memorial Sloan-Kettering Cancer Center, total gastrectomy was found to be detrimental to survival, not because of the extent of resection but mostly because splenectomy was associated with this operation. It is suggested that an exclusive policy of elective total gastrectomy when a tumour can be widely resected by a subtotal gastrectomy is incorrect. For patients with synchronous multiple gastric cancers, logic suggests that a total gastrectomy is essential. Paradoxically, their survival after total gastrectomy does not exceed that after partial gastrectomy, and it has been suggested that small synchronous lesions may regress after resection of the main tumour. There is no evidence for this. In general, multiple cancers should be treated by total gastrectomy where possible.

Lesions in the body and fundus of the stomach present different problems. Total gastrectomy is often advocated in these circumstances because the small amount of stomach that remains has little reservoir capacity. Both procedures require an oesophagoenteric anastomosis, and the operative mortality for both total and proximal gastric resection is 5 per cent. Theoretically, total gastrectomy carries a lower risk of recurrence or of a second gastric cancer in long-term survivors. Total gastrectomy with a Roux-en-Y reconstruction is also superior to proximal gastrectomy since these patients will be less prone to alkaline and biliary reflux. Tumours of the body and fundus are often far advanced at presentation because they reach considerable size before producing symptoms. They have a less favourable prognosis than tumours in the antrum.

The preferred method of reconstruction after total gastrectomy is as a Roux-en-Y with a 60-cm Roux to prevent bile reflux. The creation of a pouch to act as a reservoir to prevent early satiety has been advocated. The Hunt–Lawrence pouch (Fig. 8) is created from a long Roux loop with an enteroenteric anastomosis at 80 cm. It is simply folded on itself and a 10 cm, side-to-side anastomosis created below the oesophageal anastomosis. A partial gastrectomy is best reconstructed as an antecolic, Pólya type of gastrectomy (see Fig. 7).

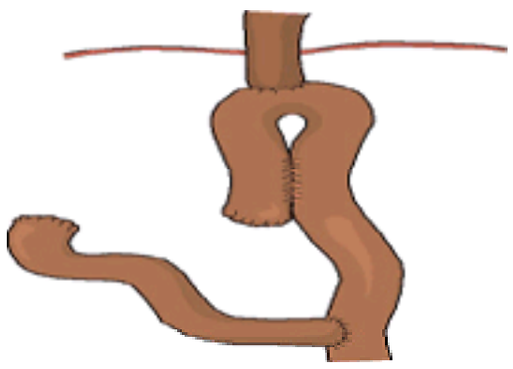


Fig. 8. Following total gastrectomy a Hunt–Lawrence pouch is created from a long Roux loop with an enteroenteric anastomosis at 80 cm.

Cancer at the cardia represents a special problem of management. At presentation only 10 per cent of patients have disease restricted to the stomach, and only 2 per cent are early cancers. The tumour may infiltrate the lower oesophagus, and a 10-cm oesophageal clearance is advised to be certain of clear resection margins. Thus the operation involves principles of gastric as well as oesophageal surgery, and in certain respects it should be regarded as a separate entity.

Extent of lymphadenectomy

The benefit of lymphadenectomy for gastric cancer has been emphasized by the Japanese. This concept is not new in oncology. In some Western series, wide lymphadenectomy encompassing one or more echelons of uninvolved lymph nodes has been found to improve survival in the small number of patients with limited nodal involvement. The Japanese Research Society for Gastric Cancer defines a curative resection as one in which patients without peritoneal, serosal, or hepatic involvement have a gastric resection with a lymph-node dissection one level beyond that of pathological lymph-node involvement.

The Japanese classification of the type of gastrectomy corresponds to how radical the operation is at removing the groups of lymph nodes according to the Japanese classification (N0–N4). D_0 gastrectomy does not remove any lymph-node group, D_1 gastrectomy removes those nodes in group I (N1), predominantly perigastric nodes, but leaves a large portion of the greater omentum. No formal lymphadenectomy is performed during this resection. A hybrid form of this resection, a traditional radical gastrectomy including omentectomy and partial lymphadenectomy, is the one mostly performed in the West. The Japanese results suggest that in most circumstances the radical D_2 gastrectomy gives the best chance of prolonged survival and this is the operation mostly performed in that country.

D_2 gastrectomy has the same criteria for adequate gastric removal but includes lymphadenectomy to remove group-II (N2) nodes *en bloc* with the stomach. The precise tissues removed depends on the site of the cancer. In general the entire greater omentum is removed, with the superior leaf of the transverse mesocolon, pancreatic capsule, and lesser omentum. The lymph-node dissection starts by removing the nodes along the gastroduodenal artery to its origin at the hepatic artery. It is continued laterally to the porta hepatitis along the common hepatic artery. The nodes are cleared medially along the common hepatic to the coeliac axis. This is cleared and continued along the splenic artery to the hilum of the spleen. Clearance in this area is facilitated by mobilization of the distal pancreas. D_3 gastrectomy attempts to remove nodes in group III (N3), and involves pancreatic and splenic resection. The Japanese results overall demonstrate a corrected 5-year survival following D_0, D_1, D_2, and D_3 resections of 26, 42, 50, and 40 per cent, respectively. Thus the Japanese have adopted the D_2 gastrectomy as the usual form of treatment for operable gastric cancer. It is recommended as the principal form of treatment for both early gastric and advanced cancer. The justification for the use of this radical operation in early tumours is the finding that N1 lymph nodes are involved in 1 to 5 per cent of mucosal and 11 to 19 per cent of submucosal cancers. The rate of metastasis in N2 lymph nodes is 0 to 2 per cent for mucosal and 2 to 9 per cent for submucosal cancers. Metastasis to group-N3 and -N4 nodes is unusual. A selective policy might be adopted if the depth of the primary could be accurately assessed preoperatively. Indeed there is no difference in survival after D_1 and D_2 gastrectomy

for mucosal cancer without lymph-node metastasis, but a significant survival advantage for D_2 resection if lymph nodes are involved. The radical operations of D_2 and D_3 gastrectomy are performed remarkably safely by the Japanese surgeon, with an operative mortality of 1 to 3 per cent.

Resistance to, and failure to adopt, the D_2 gastrectomy in the West have arisen because a clear advantage over the traditional Western radical resection has not been demonstrated, chiefly because of varying experience. In the West, the D_2 gastrectomy carries a longer operation time, greater blood loss, a longer postoperative stay, and greater morbidity. However, by avoiding pancreatectomy and splenectomy the morbidity can be reduced. The Japanese surgeons have demonstrated no difference between postoperative mortality and the level of nodes dissected. The D_2 gastrectomy remains the mainstay of their treatment for gastric cancer.

Lesions at the gastric cardia tend to metastasize to all the regional lymph nodes, and extended total gastrectomy (D_3) is required to encompass all of them, in addition to mediastinal dissection. This carcinoma has such a dismal prognosis that almost any surgical procedure is palliative and often a more simple, proximal gastric resection is most appropriate.

The treatment of ‘stump cancer' should follow the same principles as for radical excision. The pattern of lymph-node spread is different. Often, following the previous resection, there are no lymph channels along the left gastric vessels, and lymph drainage from the greater curve and into the jejunal mesentery is more important. Radical excision usually implies total gastrectomy and excision to the origin of the jejunal artery supplying the gastroenterostomy.

The mortality and morbidity of the operation are related to the nature of resection, with distal partial gastrectomy carrying the lowest rates, followed by subtotal gastrectomy and then total gastrectomy. The factors that are significantly related to postoperative complications, apart from operative procedure, are age, sex, and splenectomy. Men, and patients of more advanced age, are more prone to complications, as are those who have a splenectomy. In general the mortality from surgical resection of gastric cancer in Western countries has fallen from over 10 per cent in the 1970s to 2 per cent in the 1990s, and is approaching the outcome obtained in Japan.

The results of surgery are summarized in Table 5 above, comparing series from Japan and the United Kingdom; overall the survival results are substantially better in Japan.

Palliative surgery

In the West, very few patients present with early gastric cancer: in one large series there were only 90 patients fulfilling the criteria among 13 000 with gastric cancer. Approximately 55 per cent of patients with gastric cancer are suitable for laparotomy, with curative resection in only 21 per cent. The remainder will have no procedure (15 per cent), palliative resection (10 per cent), or bypass (9 per cent). Unfortunately, laparotomy and palliative surgery carry a significant mortality (10–30 per cent) and morbidity. Thus, before embarking on palliative excision, it is important to define the goals of this form of surgery. Symptoms due to obstruction, such as dysphagia, vomiting and obstructive pain, may be relieved by surgical resection or bypass (Fig. 9). Bleeding can be relieved by resection. If, at laparotomy, curative resection is impossible, the alternatives are a palliative resection, a bypass, or no procedure. The consensus is that palliative, non-radical resection, if it can be performed, may provide the best chance of relief of symptoms. Prolonged survival in some patients (6 per cent) has occurred following palliative resection. Palliative bypass procedures do not increase survival times and are only necessary if there is obstruction. The benefit for the quality of life is highly questionable. The mean survival for patients unsuitable for, or not receiving, a resection is 5 months. There are various methods of non-operative palliation, including laser therapy and intubation for dysphagia, and interstitial laser therapy for bleeding gastric cancers.

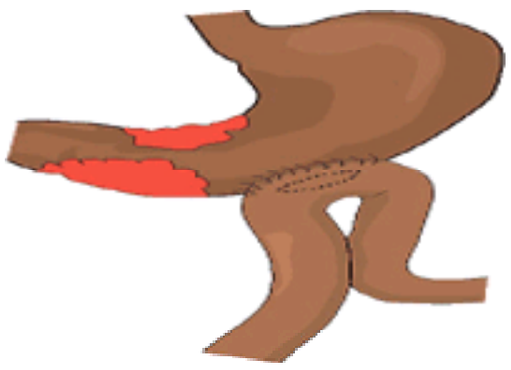


Fig. 9. Palliative gastroenterostomy to bypass an obstructing distal carcinoma of the gastric antrum.

The role of radiotherapy

For some years it has been apparent that locoregional recurrence arises in 80 per cent of patients with distant metastases, and at death 16 to 22 per cent have locally recurrent disease only. Also, a ‘second look' laparotomy in asymptomatic patients without evidence of distant metastasis revealed local recurrence in 69 per cent. These findings have been used to suggest a role for adjuvant radiotherapy. Gastric adenocarcinoma has generally been regarded as radioresistant. Although less radiosensitive than squamous-cell carcinoma, a useful response and tumour shrinkage have been achieved in patients with adenocarcinoma irradiated for palliation of malignant dysphagia. The major limitation to radiotherapy has been the problem of dosimetry involved in sparing adjacent normal tissue, including the liver and small intestine. Intraoperative radiotherapy may overcome some of these problems. In some Japanese centres, after attempted curative resection, intraoperative radiotherapy is administered to the coeliac axis and the tumour bed. In comparison with historical controls there is some evidence that patients with stage-III and -IV disease have a prolonged survival with this technique. Radiotherapy delivered postoperatively to patients following palliative and curative resection seems to have little influence on survival.

It is well established at other anatomical sites that the benefits of radiotherapy are best if bulk disease is controlled by other means. Unfortunately, patients with gastric cancer may have large and unresectable tumours at presentation. It has been suggested that preoperative radiotherapy might be used to shrink the tumour and allow subsequent resection. Unfortunately, accurate planning of radiotherapy is difficult preoperatively, and there is little evidence that preoperative radiotherapy is beneficial.

The role of chemotherapy

Systemic

The rationale for the use of adjuvant chemotherapy in gastric cancer is clear, since surgery is only potentially curative, and 70 per cent of patients with cancer confined to the gastric wall on pathological examination will die of the disease. Therefore, although the disease is apparently localized, it has in fact extended beyond surgical resection or is already disseminated, and systemic therapy is required in any attempt to eradicate it.

Gastrointestinal cancers are, in general, relatively unresponsive to chemotherapy, but gastric cancers appear to be more sensitive than most and, in particular, respond better than colorectal cancers. Despite evidence that combined chemotherapy is better, most studies have used single agents. Mitomycin C, doxorubicin, 5-fluoruracil, and nitrosoureas will produce tumour shrinkage in up to 30 per cent of patients with advanced disease. **FAM** (5-fluorouracil, doxorubicin, and mitomycin C) at present represents the most effective regimen for advanced gastric cancer, with 40 per cent of patients obtaining a partial response. Unfortunately, only 5 per cent will have a complete response. The translation of a response in advanced cancer into an effective adjuvant regimen in less advanced disease has proved difficult. Again, there is differing experience between Western series and the positive Japanese studies. In the extensive European and American trials, no survival benefit has been demonstrated with combinations of fluoruracil, mitomycin and methyl CCNU, and FAM, used after potentially curative surgery. The Japanese have demonstrated some survival advantage in patients treated with mitomycin C alone or in combination when given intraoperatively or in the very early postoperative period. The trials have been criticized because the survival benefit was demonstrated by the retrospective formation of subgroups. Overall there is no convincing evidence that cytotoxic chemotherapy is of use after resection for gastric cancer. The general application of adjuvant chemotherapy cannot be recommended. The role of preoperative (neoadjuvant chemotherapy) has to be established, but it might increase the resectability rate. The combination of postoperative chemotherapy with radiotherapy has also failed to improve survival in this disease.

Regional

Because malignant cells are shed into the peritoneum by the tumour, various local methods have been developed to eradicate localized intraperitoneal disease. Local instillation of cytotoxics, or their hyperthermic peritoneal perfusion, initially appeared effective but failed to demonstrate a survival benefit when included in a randomized clinical trial. Palliation of advanced tumours, with endoscopic regression and some long-term survival, can be achieved by local arterial infusion of cytotoxics, but this method remains inappropriate for most patients.

Antimicrobial therapy and gastric lymphoma

The association between *H. pylori* infection and gastric lymphoma and maltoma is clear. Complete regression can occur if the bacterium is eradicated in 70 per cent of patients and reinfection can result in a relapse. It is clear that the stage of the disease is important. Most patients with lymphoma restricted to the mucosa or submucosa as assessed by endosonography will show a complete response. More invasive tumours respond, but rarely show complete regression. After transformation, gastric lymphoma can be treated with chemotherapy and radiotherapy to conserve the stomach, but gastrectomy is still necessary for bulky disease.

Further reading

Abe S, Shiraishi M, Nagaoka S, Yoshimura H, Dhar DK, Nakamura T. Serosal invasion as the single prognostic indicator in stage IIA (T3N1M0) gastric cancer. *Surgery* 1991; **109**: 582–8. [Serosal invasion is among the most important prognostic indicators following surgical excision.]

Allgayer H, Heiss MM, Schidberg FW. Prognostic factors in gastric cancer. *British Journal of Surgery* 1997; **84**: 1651–64. [The Western experience of treatment of gastric cancer with an emphasis on prognostic factors.]

Bailey T, Biddlestone L, Shepherd N, Barr H, Warner P, Jankowski J. Altered cadherin and catenin complexes in the Barrett's esophagus–dysplasia–adenocarcinoma sequence. *American Journal of Pathology* 1998; **152**: 1–10. [The molecular sequence that occurs to disrupt cell adhesion in Barrett's oesophagus leading to cancer. This paper needs to be read in context with Guilford P *et al.* (see below). The nuclear expression of b-catenin is described (see also Fig. 5).]

Everett SM, Axon ATR. Early gastric cancer in Europe. *Gut* 1997; **41**: 142–50. [An important discussion of the European experience of early gastric cancer. This disease is more commonly detected in Japan.]

Guilford P *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature* 1998; **392**: 402–5. [An important paper from New Zealand demonstrating significant mutations in E-cadherin (a cell-adhesion molecule) in patients with familial gastric cancer.]

Kampschoer GHM, Nakajima T, van de Velde CJH. Changing patterns in gastric cancer. *British Journal of Surgery* 1989; **76:** 914–16. [The distribution of gastric cancer is changing. The incidence of distal gastric cancer is falling and the incidence of cancer in the cardia is rising.]

Maruyama K, Gunven P, Okabayashi K, Sasako M, Kinoshita T. Lymph node metastases of gastric cancer. General pattern in 1931 patients. *Annals of Surgery* 1989; **210**: 596–602. [A clear description of the patterns of lymph node metastases following gastric cancer resection. There is an orderly progression of lymph metastases in this series.]

Molloy RM, Sonnenberg A. Relation between gastric cancer and previous peptic ulcer disease. *Gut* 1997; **40**: 247–52. [The relationship between gastric and duodenal ulcers and the devlopment of gastric cancer. A history of duodenal ulcer is a significantly protective factor for the development of gastric cancer.]

O'Toole P, Lombard M. Vitamin C and gastric cancer: supplements for some or fruit for all. *Gut* 1996; **39**: 345–7. [A clear description of the role of vitamin C as a protective agent and antioxidant in the stomach. The stomach is an area of high oxidative stress.]

Parsonnet AJ, Friedman GD, Orentreich N, Vogelman H. Risk of gastric cancer in people with CagA positive or CagA negative *Helicobacter pylori* infection. *Gut* 1997; **40**: 297–301. [The importance of *Helicobacter pylori* infection and subtype infection for the risk of gastric carcinoma. Infected individuals who have antibodies to CagA *Helicobacter pylori* are 5.8 times more likely to develop gastric cancer.]

Roukos DH. Extended lymphadenectomy in gastric cancer: when, for whom and why. *Annals of the Royal College of Surgeons of England* 1998; **80**: 16–24. [The debate on the risks and benefits of extended resection and lymphadenectomy for gastric cancer.]

Schlemper RJ *et al.* Differences in diagnostic criteria for gastric carcinoma between Japanese and Western pathologists. *Lancet* 1997; **349**: 1725–9. [The emphasis of Western pathologists is on invasion to diagnose cancer. The cytological appearances are more important to the Japanese.]

Taal BG, Boot H, van Heerde P, de Jong D, Hart AAM, Burgers JMV. Primary non-Hodgkin lymphoma of the stomach: endoscopic pattern and prognosis in low versus high grade malignancy in relation to MALT concept. *Gut* 1996; **39**: 556–61. [The recognition of gastric lymphomas associated with *Helicobacter pylori* infection. The treatment of lymphomas is discussed.]

Takeshita K *et al.* Endoscopic treatment of early oesophageal cancer or gastric cancer. *Gut* 1997; **40**: 123–7. [Early gastric cancer can be treated by local mucosal therapy and resection.]

23.3 ‘Gastritis'

Alan G. Johnson

Introduction
Gastritis or gastropathy?
Classification
Importance for surgeons
Intestinal metaplasia
Pseudopyloric metaplasia
Foveolar hyperplasia
Endocrine cell hyperplasia
Biopsy sites
Clinical management of diffuse types of gastritis
Acute gastritis
Chronic gastritis
Chemical chronic gastritis
NSAIDs
Alcohol induced
Postgastrectomy (bile) gastritis
Rare forms of gastritis
Radiation gastritis
Infective gastritis
Granulomatous gastritis
Vascular gastropathies
Gastritis and symptoms
Ménétrier's disease
Treatment
Conclusion
Further reading

Introduction

Gastritis is a confusing word. To the endoscopist, it is a visual appearance; to the histopathologist, a very variable pathological entity; and to the family doctor and patient, a useful label for unexplained dyspepsia. The recognition of the importance of *Helicobacter pylori* has added a new dimension (and thousands of references) to the debate. The endoscopic appearance of the gastric mucosa can be very misleading, except where there is florid erosive gastritis or severe atrophy. The degrees of redness can, in these days of video-endoscopy, be altered at the press of a button. When endoscopists have predicted gastritis and histopathologists (blinded to the endoscopy findings) have reported on the histology, correlation has been very poor.

Gastritis or gastropathy?

The term gastritis implies inflammation and inflammatory cell infiltration on histology. So when there are visual and histological changes without inflammation, the term gastropathy has been used, as in ‘portal hypertensive gastropathy', where there are dilated submucosal and subepithelial vessels without associated excess of inflammatory cells.

Classification

Any classification must take into account aetiological, topographical, and morphological information. In 1990, a working party met, at the World Congress of Gastroenterology in Sydney, to establish guidelines for the classification and grading of gastritis—‘the Sydney System'. Four years later, an international group of gastrointestinal pathologists met in Houston to reappraise the Sydney System and to establish an agreed terminology of gastritis—‘the Up-dated Sydney System', which has been widely used by surgical pathologists since then. The three broad categories are acute, chronic, and special (or distinctive) forms. Table 1 shows the detailed classification of chronic gastritis. It is important to note that *H. pylori*, for example, may play a role in more than one type, and that more than one aetiology can coexist in the same patient, for instance chemical gastritis (due to non-steroidal anti-inflammatory drugs) and *Helicobacter*-associated chronic gastritis. Therefore, the presence of *Helicobacter* does not necessarily mean it is the cause of the changes.

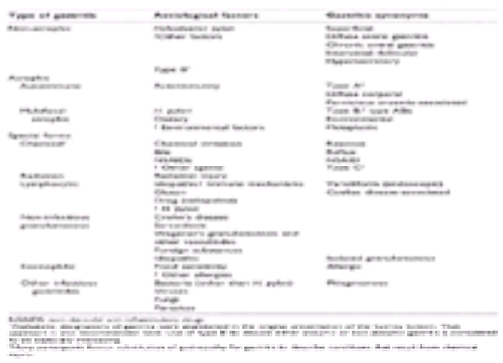


Table 1 Classification of chronic gastritis based on topography, morphology, and aetiology (after Dixon *et al.*, 1996)

The three common types, namely non-atrophic (*H. pylori*), multifocal atrophic, and autoimmune atrophic, show different distributions throughout the mucosa as follows.

1. Non-atrophic is predominantly antral or uniformly over the antrum and body of the stomach. Neutrophil activity is universal and disappears within days of eradicating *H. pylori* (Fig. 1(a)).

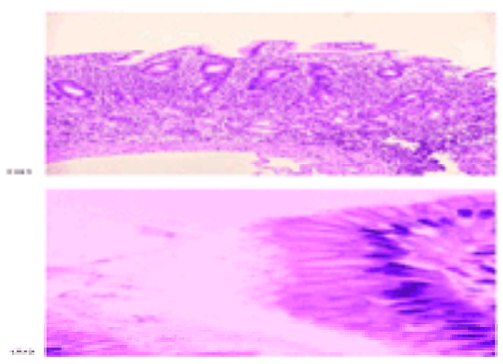


Fig. 1. (a) Photomicrograph of *H. pylori*-associated gastritis (haematoxylin and eosin stain ×40). (b) Same photomicrograph ×400 to show *H. pylori*.

2. Multifocal atrophic gastritis shows less intense inflammation and similar antral and body changes with atrophic areas particularly around the incisura and transitional zones. Atrophy implies lack of glandular tissue.
3. Autoimmune atrophic gastritis is almost always restricted to the body of the stomach.

Importance for surgeons

Surgeons are not primarily interested in histological subclassifications but they do need to know the answer to two questions.

1. Is the gastritis predictive of serious complications, either acute or long term? Examples are portal hypertensive gastropathy and chemical gastropathy, which may lead to acute bleeding; and autoimmune atrophic gastritis, which may lead to malignant change.
2. Should the gastritis be treated? Whether or not to treat *Helicobacter* infection, when it is accompanied by gastritis but not duodenal ulcer, is a very relevant question that will be discussed later.

Intestinal metaplasia

Although atrophy may occur alone, intestinal metaplasia is common in all forms of chronic gastritis. It is characterized by goblet cells and absorptive cells and can be subdivided into either small intestinal or colonic types by mucin histochemistry. It is important because it is premalignant, but there are some studies which show that the risk is small in the complete (type 1) form but up to five times greater in the large intestinal form (type 3 or incomplete).

Pseudopyloric metaplasia

This is a form of metaplasia which occurs when the body mucosa changes to resemble antral mucosa, but without the accompanying G cells. This difference is used in differentiation. The pseudopyloric cells are of a repair cell lineage with implications of previous damage including ulceration.

Foveolar hyperplasia

This is particularly pronounced in chemical gastritis, such as bile reflux after gastric surgery. It is recognized by increased length and tortuosity of the foveolas and increase in nuclear size compared with the depleted cytoplasm (Fig. 2).

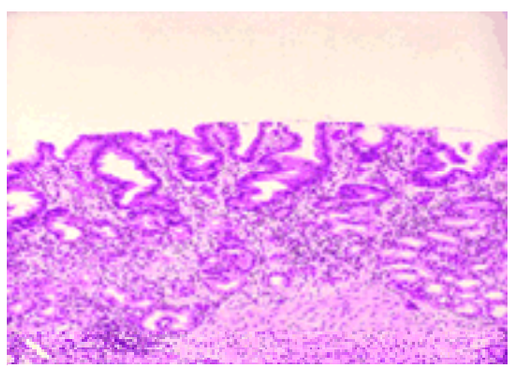


Fig. 2. Photomicrograph of reactive chemical gastritis showing foveolar hyperplasia (haematoxylin and eosin stain ×40).

Endocrine cell hyperplasia

This is most prominent in autoimmune atrophic gastritis. Excessive hyperplasia of the enterochromaphin-like cells in the oxyntic glands may progress to carcinoid tumour. The hyperplasia is due to reduced acid secretion leading to hypergastrinaemia from the antrum which in turn stimulates the enterochromaphin-like cells. This development was one of the anxieties when patients were first treated with long-term acid-suppressing drugs, but has not, so far, been a problem in humans.

Biopsy sites

At endoscopy, a biopsy should be taken from any areas of the gastric mucosa with localized changes; but to assess a diffuse gastritis, five biopsies should be taken: two from the body of the stomach (one lesser and one greater curve), two from the antrum (one lesser and one greater curve), and one from the incisura, as this is the area of maximal atrophy in intestinal metaplasia and dysplasia (Fig. 3).

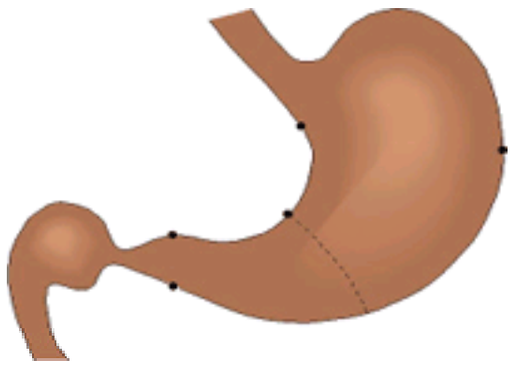


Fig. 3. Biopsy sites (•) for the diagnosis of diffuse gastritis.

Surgeons must work closely with pathologists in giving as much clinical data as possible and also understanding the significance of their reports.

Clinical management of diffuse types of gastritis

Acute gastritis

This must be distinguished from active chronic gastritis and is in two forms: acute haemorrhagic or erosive gastritis, and acute phlegmonous or suppurative gastritis.

Acute haemorrhagic or erosive gastritis

This is usually due to chemical injury with non-steriodal anti-inflammatory drugs (**NSAIDs**), alcohol, or other chemicals. A similar picture also occurs with the stress of major trauma and is seen in patients in intensive care units.

Treatment is to withdraw the harmful substance and treat with intravenous acid-suppressing drugs. If there is continued significant bleeding, in extreme cases, a total gastrectomy is required. Truncal vagotomy may temporarily stop the bleeding by opening submucosal vascular shunts, but is not reliable enough to be recommended. In preventing erosive gastritis in patients in intensive care units, a surface protecting agent such as sucralfate is preferred to acid suppression, since the hypoacidic

stomach is quickly colonized by bacteria which can form a source of lung and other infections.

Acute phlegmonous or suppurative gastritis

This is rare and usually fatal. Pyogenic organisms spread through the stomach from the submucosa causing oedema and pus formation accompanied by septicaemia. Broad-spectrum antibiotics may be given, but are usually too late.

Chronic gastritis

H. pylori can be diagnosed by antral biopsy for histology or the rapid urease test, or by urea breath tests or anti-*H. pylori* antibodies in the blood. If endoscopy is being performed, then biopsy for CLO or histology is very reliable, but to check elimination after treatment, a breath test avoids a further endoscopy, whereas antibody tests are not useful after treatment because they remain raised for some time.

Should *H. pylori* be treated?

There are four situations in which treatment is recommended:

i. when there is also a peptic ulcer and the organism is being treated to reduce the risk of recurrence (see Chapter 23.1);
ii. in mucosa-associated lymphoma there is increasing evidence that elimination of *H. pylori* treats the lymphoma;
iii. for patients on NSAIDs who have had erosions, the elimination of *H. pylori* may reduce, but not eliminate, the risk of further problems; and
iv. after resection of early gastric cancer.

For all patients where the organism is an incidental finding, even though associated with a gastritis, elimination is not indicated or is even contraindicated for two reasons: the chances of antibiotic resistance would increase and gastro-oesophageal reflux of acid may be made worse.

The question of whether to eradicate *Helicobacter* when intestinal metaplasia is present, to prevent malignant changes, is still very controversial. Whereas hyperproliferation reverses when *H. pylori* is eradicated, intestinal metaplasia, where the oncoprotein P21 is expressed, does not reverse. The P53 mutation may also explain persistence of metaplasia. Other studies have shown that there is an inverse correlation between the extent of intestinal metaplasia and the anti-*H. pylori* titre in the blood. Further studies are needed before the elimination of *H. pylori* to prevent atrophic gastritis and carcinoma can be recommended, but there are theoretical reasons why suppression of acid long term without elimination could lead to an increase in atrophic gastritis. Therefore, patients who need to be on long-term acid suppression, for example because they have to continue NSAIDs or cortiosteroids, should probably have the infection treated. However, the drop in incidence in gastric cancer in the Western world started before any therapy would have been effective.

Chemical chronic gastritis

NSAIDs

See Chapter 23.1.

Alcohol induced

Alcohol is both absorbed and metabolized in the stomach and oxidation generates the toxic metabolite acetaldehyde. Chronic alcohol abuse also favours colonization by *H. pylori*. Therefore, chronic gastritis is very common in these patients and has more than one cause. Other drugs can also affect the gastric metabolism of alcohol. Treatment is by stopping the alcohol or, if that is not possible, *Helicobacter* elimination may be tried if recurrent acute bleeding is the problem. However, there are two important warnings for doctors treating these patients: symptoms and gastritis may not correspond, and inflammatory gastritis must be distinguished from portal hypertensivegastropathy caused by cirrhosis of the liver in the late stages of overconsumption of alcohol.

Postgastrectomy (bile) gastritis

Old duodenal ulcer operations, such as partial gastrectomy (Billroth I and II) or vagotomy and drainage, all lead to reflux of duodenal contents into the stomach or gastric remnant. Only proximal gastric vagotomy does not, because the pylorus is neither bypassed nor destroyed. Although the term ‘bile reflux' is often used (because a patient can identify green bile in the vomit), other constituents of duodenal juice such as lysolecithin and pancreatic enzymes can also damage the gastric mucosa. The characteristic histological changes are of foveolar hyperplasia (Fig. 2). *Helicobacter* is absent after partial gastrectomy, but not after vagotomy and pyloroplasty and proximal gastric vagotomy, even in patients whose ulcers have been healed for years.

Treatment

If the reflux is severe and causing loss of weight and bilious vomiting, revisional surgery should be considered. The standard operation is Roux-en-Y conversion, but a Roux-en-Y bile diversion has also been used to try to prevent the Roux stasis syndrome. Intestinal metaplasia commonly occurs in the postoperative stomach and fluctuates with time from operation. There is an increased risk of malignancy in the gastric stump after many years, but the indication to divert bile and to eliminate *H. pylori* to prevent this has not been established and regular endoscopic surveillance of the postoperative stomach is probably not cost-effective.

Rare forms of gastritis

Eosinophilic gastritis is thought to be part of a generalized gut allergic reaction and is particularly seen in patients with a history of asthma or atopic eczema. However, an increase in eosinophils can also be seen in Crohn's disease and parasitic disorders. It may be due to a food sensitivity, but this needs to be diagnosed with care.

Radiation gastritis

Acute changes manifest as necrosis of fundal glands and late changes are fibrosis in the gastric wall.

Infective gastritis

Fungi, mycobacteria, bacteria, viruses, and parasites can all affect the gastric mucosa, but particularly do so in those patients with reduced immunity as occurs in HIV infection. Fig. 4 shows a gastritis associated with giardia infection.

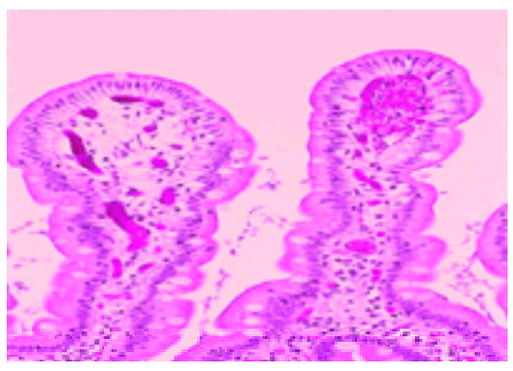


Fig. 4. Photomicrograph of infective gastritis due to giardiasis (arrowed) (haematoxylin and eosin stain ×200).

Granulomatous gastritis

Conditions such as Crohn's disease, tuberculosis, and sarcoidosis are covered elsewhere under the particular pathology.

Vascular gastropathies

Portal hypertensive gastropathy was first recognized as a distinct entity in the 1980s. The gastroscopic appearances are distinctive and can be demonstrated endoscopically. Biopsy is not advisable in patients where there has been recent bleeding, and the endoscopist must be aware that a gastric varix may not be far beneath the surface.

The endoscopic appearances are classified as follows:

i. ‘snake skin': a mild form characterized by a fine white rectangular pattern separating areas of red oedematous mucosa;
ii. ‘cherry red spots': discrete red spots, as seen in the oesophagus with portal hypertension; and
iii. diffuse haemorrhagic.

The last two represent severe forms of the condition.

It is a significant cause of bleeding in patients with portal hypertension, especially those whose oesophageal varices have been sclerosed, and it is most pronounced in the proximal stomach. The diagnosis is important as it responds to b-blockers (or portal decompression) and the risk of bleeding can be reduced significantly.

Gastric vascular ectasia is a distinct syndrome occurring in the antrum and associated with gastric atrophy. It can be diagnosed on histological examination.

Gastritis and symptoms

Gastritis is not a symptom; it is a histological diagnosis. In patients on NSAIDs there is no correlation between mucosal histology and symptoms. The symptoms of flatulent dyspepsia may be caused by motility disorders and the gastritis may be secondary. On the other hand, an active gastritis may give symptoms with other stimuli which have not caused it. Elimination of *H. pylori* has little effect on dyspepsia symptoms in the absence of peptic ulceration or when an ulcer has healed.

Ménétrier's disease

Ménétrier's original description was of massively enlarged gastric folds in patients with gastric carcinoma. It is obvious, endoscopically (Fig. 5), but there is debate about what else is required to make the diagnosis. Some authors insist on hypoproteinaemia or normal or decreased acid secretion or both, apart from the foveolar hyperplasia which is seen on histology. The hypoproteinaemia is due to protein loss from the stomach. Recently, a group at the Mayo Clinic have analysed a series of patients diagnosed clinically as having Ménétrier's disease, and have found that there are two distinct pathologies. The first, which they have named hypertrophic gastritis, has markedly increased numbers of intraepithelial lymphocytes and is considered to be an extreme form of lymphocytic gastritis. The second is a hypertrophic gastropathy with a thicker mucosa and greater mucosal oedema, but no inflammatory process and it is, for this group, that the term Ménétrier's disease should be reserved. Both have a low frequency of *H. pylori* and, in both, the histological changes affect the antrum, although it is the body and fundus that show the dramatic gastric folds.

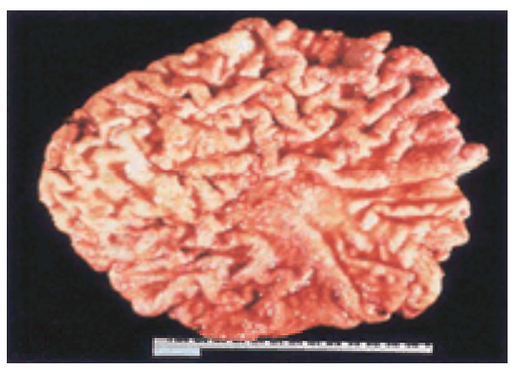


Fig. 5. Massive enlargement of rugal folds in the body and fundus of the stomach in Ménétrier's disease.

Treatment

If the hypoproteinaemia cannot be controlled by a high protein diet, then a total gastrectomy is justified. Although the initial description of Ménétrier's disease was associated with gastric carcinoma, there is now doubt about whether it is a premalignant condition: prophylactic gastrectomy to prevent carcinoma may not be justified.

Conclusion

Surgeons may encounter patients with acute erosive gastritis or portal hypertensive gastropathy in the emergency room and a clear differential diagnosis from peptic ulcer or bleeding varices is essential. Surgeons see chronic gastritis in the endosopy suite and need to take five good biopsies to establish a diagnosis. They have a particular role in evaluating the postoperative stomach, but the most pressing research needs in this field are predictors of malignant change in colonic-type intestinal metaplasia whether or not there has been previous surgery. This is likely to come from the identification of more specific combinations of oncogenes and proteins. If patients at high risk can be identified, then prophylactic gastrectomy or close surveillance may be appropriate. Meanwhile, much work needs to be done on the relationship of *H. pylori* to premalignant changes and whether in some patients these changes can be reversed as has been seen with mucosa-associated lymphomas.

Further reading

Belair PA, Metz DC, Faigel DO, Firth EE. Receiver operator characteristic analysis of endoscopy as a test for gastritis. *Digestive Diseases and Sciences* 1997; **42**: 2227–33. [A comparative study between endoscopic appearances and histology with pathologists blinded to the endoscopic findings and showing poor correlation between the two.]

Bernstein DE, Phillips RS. Portal hypertensive gastropathy. *Gastrointestinal Endoscopy Clinics of North America* 1996; **6**: 697–708. [A detailed description of the present knowledge of gastropathy.]

Blaser MJ, ed. *Helicobacter pylori* infection: atrophic gastritis and gastric cancer. *Alimentary Pharmacology and Therapeutics* 1998; **12**: Suppl. 1. [This is the proceedings from a round table meeting discussing all aspects of the relationship between Helicobacter and gastric cancer.]

Chan FKL *et al.* Randomised trial of eradication of *Helicobacter pylori* before non-steriodal anti-inflammatory drug therapy to prevent peptic ulcers. *Lancet* 1997; **350**: 975–9. [This showed that Helicobacter helps to prevent non-steroidal anti-inflammatory drug ulceration but does not eliminate it completely.]

Dixon MF *et al.* Classification and grading of gastritis. *American Journal of Surgical Pathology* 1996; **20**: 1161–81. [This is an up-dating of the Sidney Classification by an expert group with practical recommendations about biopsy sites, etc.]

European *Helicobacter pylori* Study Group. Current European concepts in the management of *Helicobacter pylori* infection. The Maastricht consensus. *Gut* 1997; **41**: 8–13.

Faigel DO, Childs M, Furth EE, Alavi A, Metz DC. New non-invasive tests for *Helicobacter pylori* gastritis: a comparison with tissue-based gold standard. *Digestive Diseases and Sciences* 1996; **41**: 740–8. [This evaluates the new non-evasive test which can avoid endoscopy and biopsy.]

Haggitt RC, Rubin CE. Endoscopic mucosal biopsy. In: Yamada T, Alpers D, Owyang C, Powell D, Silverstein F, eds, *Textbook of gastroenterology*, 3rd edn., pp. 2836–82. Lippincott, Philadelphia, 1998. [Discusses details of the techniques of biopsy and their interpretation.]

Ierardi E, Francavilla R, Panella C. Effect of *Helicobacter pylori* eradication on intestinal metaplasia and gastric epithelium proliferation. *Italian Journal of Gastroenterology and Hepatology* 1997; **29**: 470–5. [This is a useful contribution to the discussion on whether Helicobacter pylori is related to dysplasia and malignant change.]

Khakoo S, Lobo AJ, Shepherd NA, Wilkinson SP. Histological assessment of the Sydney classification of endoscopic gastritis. *Gut* 1994; **35**: 1172–5. [A useful review of the application of the Sidney Classification in practice.]

Kuipers EJ. *Helicobacter pylori* and the risk and management of associated diseases: gastritis, ulcer disease, atrophic gastritis and gastric cancer. *Alimentary Pharmacology and Therapeutics* 1997; **11**: 71–88. [This takes a practical approach to the management of Helicobacter and the risks of premalignant and malignant changes in the mucosa.]

Laine L. Nonsteroidal anti-inflammatory drug gastropathy. *Gastrointestinal Endoscopy Clinics of North America* 1996; **6**: 489–504. [An excellent review of the gastric problems associated with non-steroidal anti-inflammatory drugs]

Lieber CS. Gastric ethanol metabolism and gastritis: interactions with other drugs, *Helicobacter pylori*, and antibiotic therapy (1957–1997)—A review. *Alcoholism, Clinical and Experimental Research* 1997; **21**: 1360–6. [A review of the mechanisms involved in alcohol damage of the mucosa.]

McCormack TT *et al.* Gastric lesions in portal hypertension: inflammatory gastritis or congestive gastropathy? *Gut* 1985; **26**: 1226–42. [This is one of the early reports highlighting the changes in portal hypertension and showing they are congested, not inflammatory.]

Peura DA. The report of the Digestive Health Initiative International Conference on *Helicobacter pylori. Gastroenterology* 1997; **113**: S4–S8. [Another thorough review of all aspects of *Helicobacter pylori*.]

Wolfsen HC, Carpenter HA, Talley NJ. Ménétrier's disease: a form of hypertrophic gastropathy or gastritis? *Gastroenterology* 1993; **104**: 1310–19. [A detailed analysis of the two types of Ménétrier's disease.]

23.4 Gastric volvulus and acute gastric dilatation

Hugh Barr

Gastric volvulus
Introduction
Anatomy, aetiology, and presentation
Predisposing factors
Diagnosis
Treatment
Acute gastric dilatation
Presentation
Treatment
Further reading

Gastric volvulus

Introduction

Gastric volvulus is an abnormal rotation of the whole or part of the stomach. The terminology does not always distinguish between ‘torsion', which is simple rotation, and true volvulus, which implies luminal obstruction. Little attention is paid to the difference, although volvulus is more dangerous because of the risk of necrosis and perforation. Berti first described the condition in 1866. He reported the autopsy findings in a 60-year-old female of ‘an entire mass of organs making two complete horizontal turns, the oesophagus and the duodenum were interlaced'. Subsequently, Berg described the successful operative treatment of gastric volvulus in two patients in 1895 and 1896. The detailed review and observations in 1930 by Buchanan clarified the anatomical variants associated with this rare condition, and aetiological factors were most clearly addressed by Tanner in 1968.

Anatomy, aetiology, and presentation

The stomach is maintained in its normal position by four ligaments. The lesser curve and liver are joined by the gastrohepatic ligament, the greater curve is attached to the spleen and transverse colon by the gastrosplenic and gastrocolic ligaments, and the cardia is held fixed by the phrenico-oesophageal ligament. Some form of ligamentous abnormality (extreme laxity, absence, or disruption) is essential to allow rotation. The direction of rotation is determined by which ligaments are lax and which points remain relatively fixed. The condition can be broadly classified into two groups: organoaxial and mesenteroaxial.

Organoaxial volvulus (Fig. 1)

The pylorus and oesophagogastric junction remain relatively fixed, and rotation occurs around a line between the two. There are two subgroups. In the first (posterior organoaxial) the stomach rotates through 180° from left to right such that the anterior surface is facing backwards. The posterior surface presents under the abdominal wall and is covered by the mesocolon, when, as is usual, the transverse colon has participated in the rotation. In some patients the colon has remained inferior, owing to extreme laxity or rupture of the gastrocolic ligament. Thus two further subgroups of posterior organoaxial volvulus can be defined, the infracolic and the supracolic, depending upon the position of the stomach in relation to the colon. The spleen and pancreas may also be displaced with the stomach.

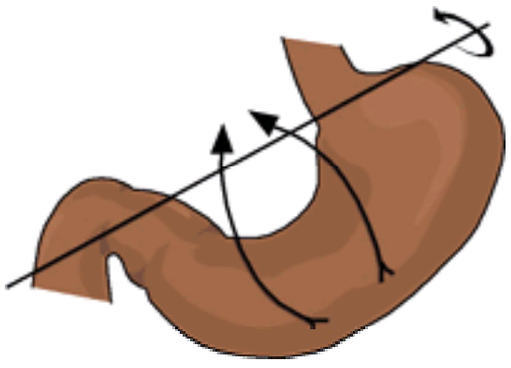


Fig. 1. Organoaxial volvulus with axis of rotation.

The second, more unusual, type is the anterior organoaxial volvulus. The greater curve passes backwards and presents in the lesser sac; thus the anteriorly facing posterior gastric wall is covered by the gastrohepatic ligament and usually cannot proceed beyond 180°. This type of volvulus is often partial. In a number of patients the cardia has remained in place and only the antral portion of the stomach has rotated. This abnormality may be seen in association with an hourglass stomach in which a large antral pouch had formed that is able to pivot on its margins.

Mesenteroaxial volvulus (Fig. 2)

Rotation occurs from right to left around a vertical line at right angles to the axis of rotation of organoaxial volvulus. The cardia remains in position and a mobile pylorus rotates anteriorly or posteriorly until it is juxtaposed to the cardia. The anterior wall is sharply kinked and folded on itself, while the posterior wall is covered by a lax gastrocolic ligament. The right half of the transverse colon is carried up in front of the volvulus towards the splenic flexure. The remainder of the abdominal organs remain undisplaced.

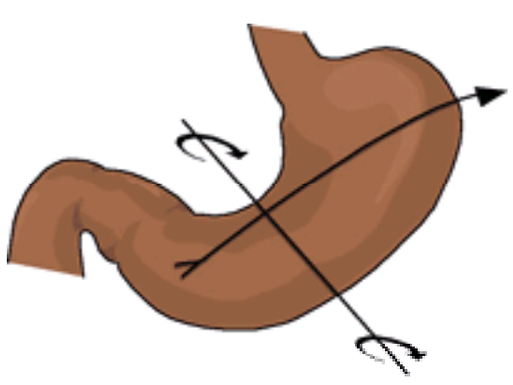


Fig. 2. Mesenteroaxial volvulus with axis of rotation.

In adults, organoaxial volvulus is the most common, accounting for 50 per cent of reported cases. Mesenteroaxial volvulus accounts for 29 per cent and a combination of the two, 2 per cent. The remainder have not been classified. However, in infants and children, mesenteroaxial volvulus accounts for 53 per cent of reported cases,

organoaxial for 35 per cent, and the combination for 5 per cent.

Predisposing factors

Although ligamentous laxity must be present, there are a number of conditions that are associated with the development of volvulus. These can be divided into three groups:

1. abnormalities of the stomach;
2. abnormalities of the surrounding viscera;
3. rotation of the stomach to fill an abnormal space.

Abnormalities of the stomach

These include conditions that produce acute or chronic distension. Pyloric stenosis and duodenal obstruction produce elongation of the ligaments and gastric ptosis. A heavy bolus lying in the lower part of the greater curve may act as a lower fixed point around which rotation (mesenteroaxial) can occur. In infants, the absence or attenuation of the ligaments may be a result of failure of fusion of fetal visceral mesenteries, and congenital bands may be predispose to volvulus.

Abnormalities of the surrounding viscera

Splenomegaly, producing elongation of the gastrosplenic ligament, has been cited as an aetiological factor. Other associated conditions include volvulus of the transverse colon and midgut, and dislocation and hypoplasia of the left lobe of the liver.

Rotation of the stomach to fill an abnormal space

There are three circumstances in which the stomach may enter an abnormal space: first, in association with a paraoesophageal hernia; secondly, with other forms of hiatus or diaphragmatic hernia; and finally, with congenital or acquired eventration of the diaphragm. Approximately 65 per cent of children with gastric volvulus have an associated eventration or hernia of the left hemidiaphragm. In affected infants below 12 months of age, 81 per cent had one of these anomalies. In adults the most common association is an organoaxial volvulus in a large paraoesophageal hernia. In one series of 138 surgically treated paraoesophageal hernias, 15 per cent contained an intrathoracic gastric volvulus.

Diagnosis

The clinical presentation of gastric volvulus is entirely dependent on whether it is acute with complete obstruction and/or strangulation, or chronic and associated with partial obstruction and no ischaemia. An acute event may occur in a stomach that has had a chronic volvulus.

Acute volvulus

In 1904, Borcharat and Lenormont described a diagnostic triad of vomiting followed by non-productive retching, localized epigastric pain, and the failure to pass a nasogastric tube. The fact that the nasogastric tube is not in the stomach is often not fully appreciated. An initial, gentle attempt at intubation of the stomach is allowable, but forcible, repeated attempts can be deleterious. Aspiration of frothy oesophageal contents rather than gastric material occurs when the tube has not traversed the gastro-oesophageal junction.

These symptoms indicate that the gastro-oesophageal junction is obstructed. True vomiting is unlikely to occur; instead frothy, white material and saliva is regurgitated. In the initial phase there are likely to be few abdominal physical signs. The patient may be profoundly compromised haemodynamically, requiring vigorous resuscitation. The differential diagnosis is often between a major cardiac or respiratory emergency. Gastric distension may occur, and signs of peritonitis and perforation arise after strangulation. It is important to establish the diagnosis before these events happen. Signs of peritonitis may be late, since perforation is often contained in the thoracic cavity.

The classic triad is often obscured in infants because pain and retching are non-specific and the failure to pass a nasogastric tube may be even more difficult interpret. In older patients the condition can be difficult to distinguish from myocardial ischaemia. An electrocardiogram is often very helpful. The appearance on straight abdominal and erect chest radiographs is usually dramatic in both adults and children. There is a hugely dilated stomach with a double fluid level on the erect film. The fluid level and distended stomach may be hidden behind the cardiac image, and detailed examination is vital. In patients with eventration or diaphragmatic hernias the inverted stomach may be seen in the chest. Contrast studies are often unhelpful, since retching renders the study dangerous and will only demonstrate obstruction with no contrast entering the stomach.

Chronic volvulus

Chronic or recurrent volvulus presents a clinical picture which may be mistaken for that of gallbladder disease, gastritis, or peptic ulceration. The patients often have a degree of respiratory compromise if the volvulus is associated with a paraoesophageal hernia. Pain is often mild and episodic, although bouts of upper abdominal colic and vomiting can occur. Dysphagia may be present if the oesophagogastric junction is distorted, and eructation of swallowed air can be difficult. After meals, gastric peristalsis may be noisy and cause embarrassment, a condition that is relieved by lying down. There are no characteristic physical signs. Unlike acute volvulus, contrast radiographs may demonstrate the abnormality. Often the patient is referred for upper gastrointestinal endoscopy. The endoscopist usually finds the procedure difficult, with clearly distorted anatomical features. It is important to recognize the abnormality and request a contrast study to confirm the endoscopic suspicion. Endoscopy can be hazardous, particularly if endoscopic retrograde pancreatography is being performed for suspected biliary disease, since persistent and forcible attempts to intubate the duodenum can be traumatic.

Gastric volvulus in children

Children present a particular problem because, as described above, the classical presentation is usually missing. The most common finding is of upper abdominal distension and colicky abdominal pain preceding the vomiting of bile or blood. A succussion splash may be present but is difficult to elicit in a distressed child.

Treatment

Acute volvulus requires immediate preoperative resuscitation followed by urgent laparotomy. The stomach must be derotated, gangrenous areas resected, and then fixed, with repair of any associated defects. Although passage of a nasogastric tube will deflate the stomach and occasionally produce spontaneous reduction, it can result in perforation of the oesophagus or stomach and persistent attempts in infants are particularly hazardous. The treatment of chronic volvulus can proceed more slowly, with careful preoperative evaluation and assessment of the risks of surgery.

Decompression and derotation

An upper midline incision provides adequate access in adults, and a transverse incision in children. If the stomach is massively distended, then decompression with needle, trochar or gastrotomy aspiration is necessary before reduction. If a gastrotomy is used, it should be closed immediately since the gastric opening may migrate to an inaccessible position on reduction. Careful inspection of the stomach for areas of ischaemia, including the posterior wall through the lesser sac, is essential,. Resection of the stomach and surrounding organs, in particular the transverse colon, is required if they are non-vital. Once the stomach has been reduced the remainder of the procedure is to prevent a recurrence.

Fixation and prevention of recurrence

Conditions predisposing to volvulus should be dealt with at once, and this may be all that is required. In other patients without a predisposing cause, or where the causative defect does not lend itself to surgical correction, or if this is inappropriate in an ill, frail patient, some form of gastropexy should be performed. The simplest involves the formation of a gastrostomy, and this technique is particularly suitable in children, tethering the stomach to the anterior abdominal wall. However, this simple fixation by suturing the stomach to the anterior abdominal wall is very prone to recurrence. It must be remembered that most infants have a diaphragmatic defect that requires correction. Gastroenterostomy and partial gastrectomy should be restricted to patients with peptic ulcer disease, who require definitive surgery for this. In patients with eventration of the left diaphragm, Tanner recommended gastropexy with colonic displacement.. The colon was detached from the stomach and placed

under the left hemidiaphragm to fill the space. The lesser curve of the stomach was then sutured to the edge of the liver and falciform ligament. This form of gastropexy carried the lowest recurrence rate. An operation devised by Opolzer of a side-to-side, fundus-to-antral gastrogastrostomy is not to be recommended since it allows the volvulus to remain and oesophageal and pyloric obstruction can still occur. Prolonged gastric stasis can occur after surgery and a gastrostomy is useful to maintain decompression.

Minimally invasive surgery: laparoscopic and endoscopic approaches

The use of minimal invasive approaches is very appropriate in the treatment of this condition. In the management of acute volvulus it is imperative to ensure that the stomach is viable, and endoscopy to examine the gastric mucosa is important. A combined laparoendoscopic approach is very attractive. After reduction of the volvulus and confirmation that there is viable gastric mucosa at endoscopy, the stomach may be fixed with gastropexy sutures placed laparoscopically. In addition, an endoscopically placed gastrostomy tube helps with the fixation and decompression of an atonic stomach. Laparoscopic repair of any hiatal defect and an antireflux procedure can be performed during endoscopic surgery.

Acute gastric dilatation

In modern surgical practice this is an unusual, but not a rare, disorder. It can produce profound circulatory disturbance and death, particularly in the postoperative period. It is akin to paralytic ileus and the same factors are often the cause. It may occur after gastric or abdominal surgery, from trauma, or with retroperitoneal haematoma, hypoxia, and electrolyte disturbances. It appears to be slightly more common in women and children, and in some instances occurs in sedated patients receiving oxygen through nasal catheters. The use of the laryngeal mask, particularly in children, has created some concern, but recent investigations have indicated that the airway seal around the mask is well maintained during surgery and gastric dilatation does not usually occur. However, caution must be exercised if positive-pressure ventilation is applied. The pressures should be kept below the leak pressure of the mask to reduce the risk of inadvertent inflation the stomach. The surgeon and anaesthetist should be aware of this possibility.

Acute idiopathic gastric dilatation has been described in patients with Prader–Willi syndrome. This predisposition may be associated with gastric necrosis and related to genetic abnormalities of gastric homeostasis. These patients are able to consume large quantities of food without vomiting.

Other patients with eating disorders can also be similarly afflicted. Gastric dilatation occurs with the superior mesenteric artery syndrome, which is in turn associated with profound weight loss.

Presentation

The clinical presentation is sufficiently distinct to allow differentiation from paralytic ileus. The patient develops tachycardia, and may become shocked. Hiccuping is common, and may be followed by belching and then enormous and effortless vomiting. The vomitus has a characteristic appearance that was described by Hamilton Bailey as the colour of ‘storm water of a peat-laden stream'. There may be upper abdominal distension. The condition can be suspected clinically and often the distended stomach can be outlined by percussion. The diagnosis is confirmed by the passage of a nasogastric tube and the aspiration of copious volumes of fluid. A straight abdominal or chest radiograph will demonstrate a hugely distended stomach. The condition can be fatal, since vomitus may be aspirated, and the hypovolaemia from fluid loss is considerable. Vigorous fluid resuscitation is often essential.

Treatment

The stomach is decompressed by nasogastric suction and the fluid and electrolyte loss replaced intravenously. An underlying cause should be sought and corrected if present, but generally this is not the case, and following decompression for several days spontaneous recovery of gastric function occurs.

Further reading

Buchanan J. Volvulus of the stomach. *British Journal of Surgery* 1930; **18**: 99–108. [An early literature review and case history with clear illustrations.]

Cole BC, Dickinson SJ. Acute volvulus of the stomach in infants and children. *Surgery* 1971; **70**: 707–15. [A presentation of the condition as it affects children from 5 months to 12 years.]

Haas O, Rat P, Christophe M, Friedman S, Favre JP. Surgical results of intrathoracic gastric volvulus complicating hiatal hernia. *British Journal of Surgery* 1990; **77**: 1379–81. [Over a 7-year period 138 patients were treated surgically for hiatal hernia; in 21 patients there was an associated intrathoracic gastric volvulus. The results indicate the need for elective intervention when intrathoracic gastric volvulus is diagnosed.]

Haller JA, Bohrer SL. Gastric volvulus, gastric fistulas and gastric trauma. In: Scott WH, Sawyers JL, ed. *Surgery of the stomach, duodenum and small intestine*, pp. 489–98. Blackwell Scientific, Oxford, 1987. [A well-illustrated description of gastric volvulus, with surgical techniques described.]

Newman RM, Newman E, Kogan Z, Stien D, Falkenstein D, Gouge TH. A combined laparoscopic and endoscopic approach to acute primary gastric volvulus. *Journal of Laparoendoscopic and Advanced Surgical Techniques* 1997; **7**: 177–81. [A clearly described minimally invasive method to deal with acute gastric volvulus, also applicable to chronic volvulus.]

Tanner NC. Chronic and recurrent volvulus of the stomach. *American Journal of Surgery* 1968; **115**: 505–15. [The aetiology and clinical presentation is clearly described, with illustrations of surgical techniques.]

Wastell C, Ellis H. Volvulus of the stomach; a review with a report of 8 cases. *British Journal of Surgery* 1971; **58**: 557–62. [A description of a personal series accumulated over 6 years (8 patients).]

23.5 Foreign bodies and bezoars

Michael N. Margolies

Foreign bodies
Bezoars
Further reading

Foreign bodies

Ingested foreign bodies occur most often in children under the age of 3 years, in patients with psychiatric disorders, in patients who use excess alcohol and drugs, among prisoners, and in denture wearers. More than 99 per cent of ingested foreign bodies are asymptomatic: the principal risk associated with foreign body ingestion in children is esophageal impaction and injury. More than 95 per cent of objects that reach the stomach pass through the gastrointestinal tract without ill effect, and guilt and anxiety of the parents often require more attention than does the child who has swallowed an object. In children, coins represent the most frequently swallowed object. The role of the surgeon is to identify the occasional patient in whom surgical treatment is needed. The history may or may not be reliable, depending upon parental observation of objects held by the infant. The majority of potentially harmful objects are radio-opaque. Radiographs should encompass the entire gastrointestinal tract from pharynx to anus. Non-radio-opaque objects may be further localized using contrast studies.

Foreign bodies in asymptomatic patients for the most part should be allowed to progress through the gastrointestinal tract spontaneously, monitored by serial radiographs. Foreign bodies are frequently missed on stool examinations, which are not usually sufficiently assiduous. Unless a particular object has a known capacity for causing mischief, or outpatient observation is not reliable, children need not be kept in hospital.

All symptomatic patients with gastric foreign bodies should undergo early endoscopy. The indications for operative removal of gastric foreign bodies include: signs of obstruction, perforation, or bleeding; accumulation of multiple foreign bodies in the stomach; certain large (> 2.5 cm diameter) or long sharp objects; objects known to result in toxicity; failure of the foreign body to leave the stomach, failure of endoscopic retrieval, and impaction.

Gastric outlet obstruction due to a foreign body is not encountered as frequently as is small bowel obstruction. Hemorrhage may occur following local mucosal ulceration at the site of impaction. Ulceration secondary to prolonged foreign body retention usually heals once the foreign body is removed. Perforation of the stomach and duodenum often presents as a fistula into adjacent organs, including the liver in the case of gastric perforations and the right kidney or inferior vena cava in the case of sharp objects in the adjacent duodenum. Late perforations complicating chicken or fish bone or toothpick ingestion are insidious and associated with much morbidity; they are rarely identified preoperatively.

Smooth objects are innocuous in the stomach, and in asymptomatic patients considerable patience should be exercised to permit their passage. Even remarkably large objects such as scissors or tableware can pass through the gastrointestinal tract in the adult. Gastric retention of foreign bodies may be related to pre-existing pyloric stenosis. If after 4 weeks the object remains, or if the object is larger than the duodenal loop, or long and sharp, removal using fibreoptic endoscopy is indicated. A variety of grasping forceps, baskets, and snares have been devised for removal. Endoscopic removal carries the risk of esophageal damage or impaction upon withdrawal; endoscopic devices should be equipped with a protective sheath drawn over the object at the time of withdrawal. One might argue that if a foreign body can be safely extracted by endoscopy it can usually be left to pass naturally.

Objects likely to cause perforation or impaction include those that are long, sharp, and pointed, such as open safety pins or hair grips in children under 2 years, toothpicks, needles, and toothbrushes, as well as objects with a configuration likely to prevent passage through the pylorus or through the second portion of the duodenum. Of all gastrointestinal tract perforations due to foreign body one-quarter to one-half occur in the stomach or duodenum. Alkaline disc batteries are innocuous unless there is radiographic evidence of disruption of the battery case, which may result in leakage and caustic damage to the stomach wall. In the case of narcotic packet ingestion for purposes of smuggling, endoscopic removal is contraindicated as intragastric breakage of the package may prove fatal (Fig. 1).



Fig. 1. Abdominal radiograph of a 35-year-old male who swallowed multiple heroin packets.

If endoscopic removal fails or there is evidence of obstruction, perforation, or bleeding, surgical gastrotomy should be performed. It is important to obtain an immediate preoperative film to be sure that the object has not migrated. Removal of non-impacted duodenal foreign bodies may be simplified by manually replacing the object into the stomach; otherwise duodenotomy is necessary. If multiple foreign objects are present, the small bowel should always be examined.

Bezoars

Bezoars are concretions of ingested material, originally described in animals. They are of considerable historical and now clinical interest. Bezoar is a transliteration of the Arabic *badzehr* or the Turkish *panzehr*, meaning antidote. The oriental bezoar is found in the stomach and intestine of the bezoar goat *Capra aegagyrus* or that of the gazelle *Antelope dorcas*. Bezoar stones were prized until the eighteenth century as cures for a variety of diseases—a large bezoar stone set in gold was included in the inventory of the crown jewels at the time of the ascension of James I to the English throne in 1662.

The classification of bezoar into four types is somewhat arbitrary in that the term is usually reserved for concretions of ingested food or chemicals; high fibre food boluses are close relatives.

Trichobezoars consist of a mass of ingested hair combined with other fibres, and usually occur in young women (90 per cent), including some with psychiatric difficulties. The hair forms a black cast of the stomach with a glistening mucoid appearance and a foul odour (Fig. 2).



Fig. 2. Gastric trichobezoar at gastrotomy in a young girl.

The most common type of phytobezoar, a concretion of botanic origin, occurs following the ingestion of persimmons (*Diospyros virginiana*). In the south-eastern United States this bezoar occurs typically in hunters who eat this fruit when unripe. Related members of the same genus are found in Israel, Japan, Korea, India, and Zimbabwe. Unripe persimmons contain large amounts of a soluble phlobatannin which coagulates in the stomach to form a tenacious glue that entraps fibres. This type of bezoar is also known as a ‘diospyrobezoar'. A second common type of phytobezoar follows ingestion of citrus fruits, although an immense variety of other vegetable matter has also been implicated. Incompletely chewed citrus segments are hygroscopic and expand in the small bowel, sometimes causing obstruction. Trichophytobezoar is a combination of the two forms mentioned above.

Bezoars may also be formed by concretions of medications or chemicals. The shellac bezoar occurs in furniture workers who imbibe an alcoholic solution of shellac; the shellac precipitates in the stomach, particularly when this cocktail is followed by water. The relative incidence of neonatal lactobezoar and antacid bezoar has increased, the former being associated with prematurity, and the latter with intensive antacid (aluminum hydroxide) therapy in high-risk hospital inpatients, particularly those in renal failure. Other common medication bezoars or ‘pharmacobezoars' include enteric-coated aspirin, psyllium, sucralfate, and extended-release nifedipine. Like phytobezoars, medication bezoars are associated with impaired gastric motility and emptying, typically after gastric surgery. Although the symptoms and treatment of medication bezoars are analogous to those for other bezoars (see below), an additional consideration is toxicity due to drug release.

Factors predisposing to bezoar formation include both diet and pre-existing gastroduodenal pathology. In addition to the binge eating of persimmons or citrus fruits, monotonous high fibre diets during famine or following periods of religious fast or at harvest time in the tropics may produce bezoars. During the past several decades the most common form of bezoars in developed countries has been found in patients who have undergone gastric surgery. These phytobezoars, 90 per cent of which are composed of citrus fruits, occur several months to many years after any surgical procedure with the potential for altering gastric emptying. They occur in 5 to 14 per cent of patients following gastrectomy. Postgastrectomy bezoars occur in patients with poor dentition, high fibre intake, reduced gastric acid production, and gastric stasis with or without partial gastric outlet obstruction. An increased incidence of bezoar may also occur in patients with diabetic gastroparesis or other neurological conditions affecting the stomach, such as autonomic neuropathy and myotonic dystrophy.

Gastric bezoars may be asymptomatic and found only incidentally. When symptoms occur they are due to the size of the bezoar and its complications—most commonly obstruction, and less often bleeding or perforation. Large masses may cause a sensation of epigastric fullness and early satiety. Nausea, vomiting, and abdominal pain are common due to the size of the mass or obstruction, or associated ulceration. Gastric outlet obstruction is uncommon in phytobezoar but can certainly cause partial outlet obstruction after gastrectomy. However, small bowel obstruction due to migration or fragmentation of a gastric bezoar is common. Obstruction may follow attempts at endoscopic fragmentation. Intestinal obstruction due to bezoar is seen in patients with an intact gastrointestinal tract, as well as in those who have undergone gastric surgery.

Patients with a large trichobezoar may suffer weight loss, anemia, gastrointestinal bleeding, and obstructive symptoms. Trichobezoars may extend from the stomach into and through the entire small intestine, the so-called ‘Rapunzel syndrome'. A history of trichophagia may be elicited. The physical findings include palpable epigastric mass, alopecia, and halitosis. ‘Daughter' trichobezoars may result in intestinal obstruction. Gastric ulceration with hemorrhage or frank perforation due to trichobezoar carries significant mortality.

Bezoars may be visible on plain films as mottled densities in the left upper quadrant. Phytobezoars are frequently missed (75 per cent) during contrast studies: they appear as a mobile foreign body of varying size that may become infiltrated with barium. Trichobezoars form a cast of the stomach outlined with barium. Small bowel bezoar is occasionally seen directly on plain films, but more commonly the findings are those of small bowel obstruction; the bezoar may be identified on antegrade contrast studies of the small bowel.

The principal differential diagnosis of gastrobezoar is neoplasm, particularly when the bezoar is fixed in position. Endoscopy can serve to differentiate the two; bezoars are often discovered incidentally at endoscopy performed during evaluation of symptoms following gastrectomy. Endoscopy should be undertaken in all cases of suspected bezoar to detect associated abnormalities, as gastroduodenal ulceration is present in 10 to 40 per cent of patients.

Uncomplicated gastric phytobezoar can usually be managed without operation. Digestion of the bezoar using repeated doses of oral cellulase has greater reported success rates (> 83 per cent) than the use of the proteolytic enzyme papain or the mucolytic agent acetylcysteine. Gastric emptying may be promoted by the use of metoclopramide. Occasionally, simple gastric lavage is sufficient. If these measures are unsuccessful, bezoars may be fragmented and removed endoscopically using biopsy forceps or stone baskets, or disrupted using streams of water. Ulcerations at the site of an impacted bezoar usually heal following removal of the foreign body, although bezoars can occur in patients with pre-existing ulceration at the gastric outlet. Surgical gastrotomy is reserved for patients with large phytobezoars or those that are symptomatic. In contrast, trichobezoar should always be managed surgically (Fig. 2), as conservative measures and endoscopy are of no avail; untreated trichobezoar carries a significant morbidity and mortality. Prophylactic antibiotics should be given; trichobezoars are typically putrefied.

Small bowel bezoars are treated surgically if obstruction supervenes. At laparotomy, attempts should be made to advance the bezoar into the colon manually. If these efforts are unsuccessful, enterotomy and extraction are necessary. One must guard against the not infrequent occurrence (4 to 17 per cent) of multiple bezoars by examining the stomach and the entire small bowel at laparotomy. Preoperative endoscopy is of value in cases of small bowel obstruction due to bezoar in order to identify unsuspected gastric or duodenal bezoar and extract or fragment these if possible, as they may be readily missed upon attempted palpation at laparotomy when there has been previous gastric surgery.

After successful bezoar removal, attention must be directed towards prevention of recurrence. Intake of high fibre foods, especially citrus fruits, should be avoided and adequate dentition assured. Chronic prophylactic oral enzyme therapy and the use of metoclopramide may reduce the incidence of recurrence. Tachyphylaxis to metoclopramide is rapid, however. The paradox of the bezoar is that while they were treasured in antiquity, much effort now is spent in ridding ourselves of them.

Further reading

DeBakey M, Ochsner A. Bezoars and concretions. A comprehensive review of the literature with an analysis of 303 collected cases and a presentation of 8 additional cases. *Surgery* 1938; **4**: 934–63, 1939; **5**: 132–60. [An exhaustive exegesis on the historical importance of bezoars, their classification and clinical presentation. Later reviews are by comparison incremental.]

Diettrich NA, Gau, PC. Postgastrectomy phytobezoars—endoscopic diagnosis and treatment. *Archives of Surgery* 1985; **120:** 432–5. [Management of bezoars occurring following procedures that affect gastric emptying.]

James AH, Allen-Mersh TG. Recognition and management of patients who repeatedly swallow foreign bodies. *Journal of the Royal Society of Medicine* 1982; **75**: 107–10. [Emphasizes conservative management of purposeful ingestion of foreign bodies in adults.]

Lee J. Bezoars and foreign bodies of the stomach. *Gastrointestinal Endoscopy Clinics of North America* 1996; **6**: 605–19. [A more recent review of the pathogenesis and management of gastric bezoars and gastric foreign bodies.]

Stack PE, Thomas E. Pharmacobezoar: an evolving new entity. *Digestive Diseases* 1995; **13:** 356–64. [Review of bezoars resulting from concretions of ingested drugs.]

Suita S, Ohgami H, Nagasaki A, Yakabe S. Management of pediatric patients who have swallowed foreign objects. *American Surgeon* 1989; **55**: 585–90. [Clinical aspects of management of ingested foreign bodies specific to children.]

24.1 Small bowel obstruction

Stephanie Rohovsky and Ronald Bleday

Introduction
Etiology and pathophysiology
Diagnosis
Management
Non-operative management
Operative management
Prevention
Special considerations
Postoperative small bowel obstruction
Metastatic disease
Radiation enteritis
Summary
Further reading

Introduction

Small bowel obstruction is one of the most common problems which surgeons encounter. Patients present with a variety of complaints; the difficulty often lies in making the correct diagnosis. Untreated complete bowel obstruction can lead not only to the need for bowel resection but to significant associated morbidity and mortality as well. Ever since Dr Owen Wangensteen placed the first nasoenteric tube for bowel decompression, we have had two management modalities at our disposal: non-operative and operative intervention. The challenge lies in making the correct diagnosis and implementing an appropriate treatment strategy.

Etiology and pathophysiology

Obstruction as defined for the purposes of this chapter relates to a mechanical blockage which precludes intestinal contents from moving in the usual oral to anal progression. This may be due to torsion at the point of an adhesion, incarceration in a hernia sac, intraluminal masses, or extrinsic compression from a neighboring inflammatory process. Obstruction can be complete, meaning nothing can pass beyond the point of obstruction, or partial, meaning that transit occurs through a significantly narrowed lumen. Strangulation implies compromized blood flow to the involved segment with impending necrosis. Intestinal contents back up from the point of obstruction creating dilatation of small bowel proximal to the site of mechanical obstruction. The resulting intraluminal distention can lead to venous engorgement, ischemia, and eventual intestinal wall necrosis. Additionally, the large bowel is deprived of its usual absorption of fluid, since the enteric contents are prevented from reaching the colon. These two factors set up a cycle of fluid shifts from the intravascular space to the lumen and interstitium of the small intestine. Hemoconcentration ensues, followed by electrolyte abnormalities, worsening intestinal necrosis, and eventual hemodynamic collapse. Complete obstruction requires operative intervention to relieve the mechanical blockage and to prevent strangulation. Partial obstruction, however, may respond to non-operative management. Thus the goal of evaluation is to determine which patients will require operation and which can be safely managed non-operatively.

Intestinal obstruction can result from a variety of etiologies (Table 1). Adhesions are the most common reason and account for over 60 per cent of patients admitted for small bowel obstruction. Thus previous abdominal surgery is a risk factor for development of obstruction. Hernias are the second most common reason for small bowel obstruction and may arise in a number of anatomic sites including inguinal, femoral, ventral, and intra-abdominal locations. Neoplasms account for the third largest group of patients presenting with obstruction. This may be more commonly due to metastatic disease or less commonly due to primary neoplasms of the small intestine. Finally, intussusception, inflammatory bowel disease, strictures, radiation enteritis, and local inflammatory processes such as appendicitis can also cause intestinal obstruction.

Adhesions	Postoperative Inflammatory Radiation
Hernias	
Malignancy	Primary tumors Metastatic disease
Intussusception	Neoplasm Meckel's diverticulum
Volvulus	
Stricture	Inflammatory bowel disease Radiation
Gallstones, bezoars	

Table 1 Etiology of small bowel obstruction

Diagnosis

Obtaining a detailed history and physical examination are essential in diagnosing intestinal obstruction. Patients invariably present with abdominal pain. This is usually crampy or colicky in nature in the early stages and can progress to severe, constant pain as the process progresses. Nausea and vomiting are also quite common findings, as is the absence of flatus. Much emphasis has been placed on the nature of the emesis suggesting that bilious emesis occurs with proximal obstruction while feculent emesis occurs in distal obstruction, but this is not without exception. Intestinal obstruction may present with an acute onset of pain (less than 1 day) such as those with proximal obstruction or a more indolent course (a few days) such as those associated with metastatic disease. Acuity is related to type (complete or partial) rather than the location of obstruction. A history should also include past surgical procedures, new onset of hernias, and relevant past medical history. Physical examination must include assessment of vital signs. Tachycardia and orthostasis may signify intravascular volume depletion from fluid losses into the bowel lumen and interstitium. Fever may be due to intestinal compromise. Abdominal examination may reveal distension, evidence of prior surgery, or incarcerated hernias. Abdominal auscultation may reveal high-pitched bowel sounds with obstruction, though the presence or absence of bowel sounds is not an absolute indicator of abdominal pathology. Abdominal pain may be elicited with or without signs of peritoneal irritation. A rectal examination should be performed not only to check for occult blood but also to rule out an obstructing rectal mass.

All laboratory and radiographic tests must be interpreted in the context of the history and physical findings. Laboratory tests are helpful in determining the severity of illness, though they are not specific for small bowel obstruction. A complete blood count and chemistry are routinely obtained. Patients can manifest leukocytosis with bandemia in cases of intestinal ischemia. Hemoconcentration can occur with severe volume depletion. Electrolyte abnormalities commonly occur, especially hypokalemia from protracted vomiting. Abnormal levels of serum calcium, phosphate, and magnesium may be present with both ileus and partial small bowel obstruction. An elevated level of blood urea nitrogen (**BUN**) signifies intravascular volume depletion. Urine specific gravity and osmolarity may be elevated for the same reasons. Mild serum amylase elevations are also associated with small bowel obstruction due to amylase release from enterocytes, though amylase is not specific for small bowel ischemia. These tests may suggest that obstruction should be included in the differential diagnosis, though they are not absolute indicators.

Plain film radiographs of the abdomen can be helpful in diagnosing small bowel obstruction. Three films are usually obtained. An upright chest radiograph is best for ruling out free air under the diaphragm. Free air characteristically accumulates over the liver or just below the diaphragm. A chest film also allows for evaluation of diaphragmatic hernias and other intrathoracic pathology. An abdominal plain film is usually performed in the upright position. Signs of small bowel obstruction include bowel dilatation proximal to the site of obstruction, air–fluid levels, paucity of large bowel gas, bowel wall thickening, a fixed loop, and ground glass appearance signifying intraluminal fluid. In early small intestinal obstruction, however, there may still be gas in the large bowel due to incomplete evacuation of contents distal to the point of obstruction. Air–fluid levels may suggest small bowel obstruction in a patient with a consistent history, though this finding can be present in any illness which decreases bowel motility resulting in ileus. Plain films can also appear normal in the setting of small bowel obstruction. They have been shown to be diagnostic in 50 to 60 per cent of cases, which leaves 40 to 50 per cent either equivocal or normal. This lack of sensitivity and specificity has led to the use of other radiographic modalities in diagnosing obstruction.

Upper gastrointestinal studies with small bowel follow through are often utilized in diagnosing small bowel obstruction. Barium or gastrograffin are administered with timed plain films to evaluate intraluminal transit. This study can show the point of obstruction, the degree of narrowing in the case of a partial small bowel obstruction, and associated mucosal abnormalities. It involves an initial bolus of enteral contrast with subsequent filming to document transit through the small bowel to the colon. When contrast does not reach the colon after several hours, a complete obstruction must be postulated. In patients with stomas, a retrograde contrast study through the stoma can delineate parastomal herniation causing complete obstruction. Enteroclysis has also been advocated in diagnosing small bowel obstruction. This involves placement of a long nasoenteric tube followed by contrast and air administration. It has been shown to be accurate in approximately 85 per cent of cases, and like upper gastrointestinal studies with small bowel follow through, allows one to gauge the severity of obstruction. Complete as opposed to partial obstruction is often discernible, especially with delayed films, as is the level of obstruction. Enteroclysis, however, is uncomfortable for the patient, difficult to perform in an emergency setting, evaluates only the intestinal lumen, and requires skilled interpretation at the time the study is being performed.

Computed tomography (**CT**) is playing a growing role in the diagnosis of intestinal obstruction (Fig. 1). Initial studies indicated a greater than 90 per cent accuracy in defining both the site and nature of obstruction. Proximal obstructions are more easily diagnosed by CT with an accuracy approaching 95 per cent. Distal obstructions have a lower sensitivity and specificity ranging from 60 to 75 per cent. High-grade obstructions are also more accurately diagnosed by CT scan than low-grade obstructions. Signs of obstruction by CT scan include proximal dilatation with transition point and closed-loop obstruction with a ‘beak' sign. Small bowel strangulation can be shown as circumferential thickening of the bowel wall, increased small bowel attenuation, pneumatosis, and ‘target sign' secondary to thickening. CT scans are easily obtainable in an emergency setting, require less technical expertise than enteroclysis, are non-invasive and quick, and provide imaging of the entire abdomen. Contrast administration is helpful, though fluid-filled loops of small bowel often act as their own contrast medium. Rectal contrast is useful in ruling out large bowel obstruction as the etiology of small bowel obstructive symptoms. Current recommendations include utilizing CT scan in cases where plain films are non-diagnostic, there is a disparity between clinical and radiographic findings, there is postoperative small bowel obstruction, and cases where neoplasms are suspected.



Fig. 1. CT scan of a small bowel obstruction. Dilated bowel (D) is seen proximal to the transition point with a thickened wall and fluid-filled loops. A transition point (T) is seen followed by collapsed, non-dilated bowel (C). Oral contrast can be used but in the acute setting enteric content in the dilated bowel acts as a contrast medium. CT scans are useful in the diagnosis of complete as opposed to partial small bowel obstruction in the postoperative period, with metastatic disease, in inflammatory bowel disease, where a bowel malignancy is suspected, and in patients with a history of a chronic partial bowel obstruction.

Management

Once the diagnosis has been made, one must decide whether the patient can be managed non-operatively or whether immediate operative intervention is required. The difficulty lies in trying to determine if bowel ischemia or necrosis has occurred. The morbidity and mortality associated with strangulation range from 20 to 40 per cent. Thus non-operative management may not always be conservative management. Studies have extracted four variables which, if present, should suggest the need for operative intervention. Tachycardia, fever, significant abdominal pain, and leukocytosis are all associated with strangulation. If one or more of these variables are present upon evaluation, the risk of compromised bowel increases dramatically. Additionally, any patient with a virgin abdomen and no history of malignancy should be considered for operative intervention. Signs of strangulation on CT scan can also be helpful in determining bowel compromise, though if the bowel appears normal this does not rule out strangulation. Thus patients who manifest one or more of these findings should strongly be considered for operation as should those with evidence of complete obstruction. Alternatively, if a patient presents with none of the above signs, has minimal tenderness, or has evidence of a partial obstruction, a course of non-operative management with careful examinations would be warranted.

Non-operative management

Non-operative management includes therapeutic interventions aimed at counteracting the pathophysiology fueling intestinal obstruction. Volume resuscitation is paramount and can be achieved with fluid boluses in the emergency room followed by a generous hourly intravenous infusion. Estimates of needed volume should be based upon the duration of illness, amount of vomiting, heart rate, and orthostasis. A minimum urine output of 0.5 ml/kg per hour should be achieved; Foley catheterization is usually necessary to assure adequate hourly urine outputs. Nasogastric decompression is also required to decompress air and enteric contents proximal to the site of obstruction. Serial abdominal examinations are imperative, as development of severe abdominal pain should trigger operative intervention. Repletion of electrolytes, especially potassium, is also vital in non-operative management. Occasionally, a patient may present as a partial small bowel obstruction, but in fact have an ileus. Repletion of potassium, magnesium, phosphate, and calcium can help correct the ileus. Complete blood counts should also be checked, as leukocytosis may signify need for operation. In an effort to reduce gastric secretions, histamine-2 (H_2) blockers should be administered. If prolonged nil-by-mouth status (i.e. seven days or more) is needed or expected due to persistent symptomatology, nutrition support should be instituted. This may be instituted earlier if there is a history of significant weight loss or evidence of malnutrition on presentation. Nutrition support can help not only in electrolyte repletion but in minimization of nitrogen losses as well.

Resolution of obstruction is signified by a number of things. Abdominal distension improves. Patients may begin passing flatus or even move their bowels. Nasogastric outputs will decrease. Nasogastric tubes may be clamped for several hours followed by suction to determine residual volumes. If residuals are low, the nasogastric tube can be removed. If other signs of improvement (such as decreased distension and passage of flatus) are present, nasogastric tubes may be removed. Intravenous fluids can be reduced and diets can be gradually resumed. If patients do not develop recurrent symptoms and tolerate diet advancement, they have successfully completed non-operative management. Recurrent symptoms with diet advancement, however, should prompt further study with either CT scan or upper gastrointestinal studies with small bowel follow through.

Operative management

Patients who fail non-operative therapy or who present with a complete bowel obstruction by history, physical examination, and radiologic assessment should be taken to the operating room. Expedient volume resuscitation and electrolyte repletion should be undertaken in the emergency room prior to laparotomy. Preoperative antibiotics to cover bowel and skin flora should be administered. A midline incision allows for best visualization of the entire abdominal cavity. Preoperative physical examination provides clues as to the etiology of obstruction (i.e. adhesions, hernias) and can direct initial exploration. Full exploration of the abdomen is then undertaken. The entire small bowel from the ligament of Treitz to the ileocecal valve should be examined for a transition point. Care in lysis of adhesions is imperative to avoid inadvertent enterotomy. Adhesions contributing to obstruction should be lysed. Bowel involved in obstruction is carefully examined for viability. If frankly necrotic, the bowel should be resected. Primary anastamosis can then be performed, either with stapled or hand-sewn techniques. If there is an extensive length of bowel involved with questionable areas of viability, a conservative resection can be performed followed by second-look laparotomy within 24 to 48 h. The determinination of bowel viability can also be aided by intraoperative inspection after either a waiting period or fluoroscein injection with Woods lamp detection. An intestinal bypass may be necessary in cases of diffuse metastatic disease. Once the bowel has been resected, the mesenteric defect should be closed to avoid postoperative internal herniation. Irrigation is then used to clean the abdominal cavity. Drains are normally unnecessary unless a discrete abscess cavity is encountered at the time of exploration. Volume resuscitation and electrolyte repletion should be aggressively pursued postoperatively, as should nutrition support in patients with a prolonged preoperative obstructive history. Postoperative complications include atelectasis, wound infection, fascial dehiscence, intra-abdominal abscess, and postoperative small bowel obstruction. The incidence of wound infection and dehiscence increase dramatically if an enterotomy is made at the time of exploration, even in the presence of antibiotics. Thus, careful dissection techniques should be stressed intraoperatively.

Laparoscopy currently does not play a large role in operative management of a complete intestinal obstruction. If there is no evidence of pathology with laparoscopy, one would still be obliged to perform a laparotomy to locate the problem. If necrotic bowel is encountered, a laparotomy would normally be required for adequate resection. Laparoscopy, however, may be more helpful with partial small bowel obstructions where preoperative work-up has revealed the site of obstruction.

Laparoscope-assisted ileocolic or small bowel resection can be performed with good results in this setting.

Prevention

Efforts at prevention of intestinal obstruction have been aimed at elective herniorraphy and adhesion prevention. Intraoperative strategies to minimize adhesion formation focus on minimal handling of the bowel and general careful surgical technique. Washing gloves to remove powder and avoidance of peritoneal closure are also thought to aid in adhesion prevention. Adhesions are one of the mechanisms the body utilizes to heal sites of injury. Thus each laparotomy places the patient at risk for development of new adhesion formation.

A new anti-adhesion product has been developed for use in patients undergoing general surgery. This product is a sodium hyaluronate-based bioresorbable membrane. It has been shown in a prospective, randomized, double-blinded, multicenter trial to reduce the incidence, severity, and extent of postoperative adhesions to the abdominal incision. In addition, it can reduce operative time in reoperative cases owing to reduced need for adhesiolysis. Current work is underway to link this reduction in adhesion formation to a reduction in intestinal obstruction.

Special considerations

Postoperative small bowel obstruction

A clinical conundrum often exists with postoperative delay of intestinal function. A delay of several days is normal after laparotomy, is not mechanical, and is termed an ileus. Diet advancement is then instituted in a step-wise fashion. The problem arises when a patient recovers temporary bowel function and then develops signs of obstruction. Persistent nausea and vomiting usually require reinsertion of a nasogastric tube. Management strategies involve those utilized for non-operative management of small bowel obstruction (NGT, intravenous fluid, Foley catheter, etc.). After a period of observation, however, one must decide if continued non-operative management is warranted in the face of no clinical improvement. Some argue for up to 12 weeks of intravenous nutrition, nil by mouth, and electrolyte repletion with the thinking that the obstruction would resolve. Obstruction due to edema may resolve with time, but an obstruction due to ischemic stricture or adhesion may not. Studies have shown that after a period of non-operative intervention which can range from 3 to 10 days, there is little added benefit from continued observation. CT scans can be helpful in distinguishing a mechanical bowel obstruction from a functional one in the postoperative period and can help in planning further operative intervention. If a complete obstruction with transition point is evident on CT scan after a period of non-operative management in a postoperative patient, operative intervention should be considered. This would allow for diagnosis and correction of the etiology of obstruction, earlier discharge from the hospital, and less hospital-related morbidity.

Metastatic disease

Malignant obstruction can be due to either primary neoplastic or metastatic disease. Small bowel tumors are much less prevalent than large bowel tumors, though they can still be responsible for mechanical obstruction. Gastric and pancreatic tumors can cause extrinsic compression on the small intestine and a mechanical obstruction as well. Diffuse metastatic disease or peritoneal studding can also act as a nidus for adhesion formation or direct invasion to cause a mechanical obstruction. Regardless of the etiology, the principle remains the same in terms of treating a small bowel obstruction, though a longer time trial of non-operative management may be considered if diffuse metastatic disease is likely. Rather than resection, a small bowel bypass may be a quicker, safer alternative if endstage disease is present. This will shorten operative time and dissection thus reducing the chances of enterotomy.

Radiation enteritis

Radiation effects on small intestine include a vasculitis-like reaction which can lead to extensive scar and adhesion formation. Bleeding due to inflammation is also quite common. Patients with radiation enteritis commonly complain of symptoms suggestive of partial small bowel obstruction. In cases of complete obstruction, laparotomy is still necessary. In cases of partial obstruction, however, every effort should be made to treat these patients non-operatively, as operative intervention may worsen the reactive process.

Summary

Small bowel obstruction can present in a number of different ways depending on both the etiology and degree (partial or complete) of obstruction. There is no one laboratory test or radiographc study that is always positive in cases of bowel obstruction. The diagnosis can only be made in the context of a detailed history and physical examination coupled with directed adjunctive tesing. Special attention should be paid to predisposing factors such as prior abdominal surgery, hernias, or malignancy. The most important decision following diagnosis of obstruction is whether operative intervention is warranted. Risks of non-operative management include bowel necrosis, abscess, and multisystem organ failure. Thus the proper management strategy is paramount to patient outcome. Complete obstruction usually necessitates laparotomy. If, however, a trial of non-operative management is undertaken owing to signs of partial obstruction, vigilance in serial examinations and electrolyte correction is important. If there is no improvement in clinical symptomatology after several days or if deterioration occurs, prompt operative intervention should be undertaken. Special considerations such as postoperative obstruction or metastatic disease may warrant earlier imaging studies. With technological advances, the incidence of small bowel obstruction may decline. For now, we are left with existing tools to try to determine the proper course of therapy.

Further reading

Becker JM *et al.* Prevention of postoperative abdominal adhesions by a sodium hyaluronate-based bioresorbable membrane: a prospective, randomized, double-blind multicenter study. *Journal of the American College of Surgeons* 1996; **183**: 297–306.

Frager DH, Baer ZW, Rothpearl A, Bossart PA. Distinction between postoperative ileus and mechanical small bowel obstruction. *American Journal of Roentgenology* 1995; **164**: 891–94.

Frager DH, Baer JW. Role of CT in evaluating patients with small bowel obstruction. *Seminars in Ultrasound, CT and MR* 1995; **16**: 127–40.

Francois Y *et al.* Small intestinal obstruction complicating ileal pouch–anal anastomosis. *Annals of Surgery* 1989; **209**: 46–50.

Maglinte DD, Balthazar EJ, Kelvin FM, Megibow AJ. The role of radiology in the diagnosis of small bowel obstruction. *American Journal of Roentgenology* 1997; **168**: 1171–80.

Pickelman J, Lee RM. The management of patients with suspected early postoperative small bowel obstruction. *Annals of Surgery* 1989; **210**: 216–19.

Ray NF, Denton WG, Thamer M, Henderson SC, Perry S. Abdominal adhesiolysis: inpatient care and expenditures in the United States in 1994. *Journal of the American College of Surgeons* 1998; **186**: 1–9.

Resnick J, Greenwald DA, Brandt LJ. Delayed gastric emptying and postoperative ileus after nongastric abdominal surgery: part II. *American Journal of Gastroenterology* 1997; **92**: 934–40.

Seror D *et al.* How conservatively can postoperative small bowel obstruction be treated? *American Journal of Surgery* 1993; **163**: 121–6.

Stewardson RH, Bombeck CT, Nyhus LM. Critical operative management of small bowel obstruction. *Annals of Surgery* 1978; **187**: 189–93.

Stewart RM, Page CP, Brender J, Schwesinger W, Eisenhut D. The incidence and risk of early postoperative small bowel obstruction. *American Journal of Surgery* 1987; **154**: 643–7.

Sykes PA, Schofield PF. Early postoperative small bowel obstruction. *British Journal of Surgery* 1974; **61**: 594–600.

Wangensteen OH. The early diagnosis of acute intestinal obstruction with comments on pathology and treatment: with a report on successful decompression of three cases of mechanical small bowel obstruction by nasal catheter siphonage. *Western Journal of Surgery, Obstetrics and Gynecology* 1932; **40**: 1–17.

24.2.1 Crohn's disease of the small intestine

D. P. Jewell and J. Satsangi

Definition
Epidemiology
Genetics
Aetiology
Infective agents
Environmental factors
Immunopathogenesis
Pathology
Clinical features
Symptoms
Physical signs
Acute presentation
Diagnosis
Differential diagnosis
Complications
Extraintestinal manifestations
Assessment of disease activity
Medical management
Nutrition and diet
Treatment of mild to moderate disease
Treatment of severe disease
Chronic active disease
Symptomatic treatment
Maintenance of remission
Prognosis
Crohn's disease in pregnancy
Crohn's disease in children
Further reading

Strictures of the small intestine, not obviously due to tuberculosis, have been recognized since at least the sixteenth century. In the late nineteenth century, amongst the case histories of patients with supposed ulcerative colitis, there are those who also had gross ileal involvement. In 1913 Dalziel recognized that ileal inflammation could occur as a distinct disease entity. He described the macroscopic appearances at surgery and the histological features of a chronic granulomatous inflammation predominant in the submucosa, which he distinguished from tuberculosis. However, little attention was given to this description until Crohn, Ginsberg, and Oppenheimer published their classic description of ‘regional ileitis' in 1932. Later, in the 1930s, it became apparent that the disease often affected the right colon in continuity with terminal ileal disease, but primary disease of the colon was not accepted until the description of Lockhart-Mummery and Morson in 1960.

The presence of disease in the colon made the term regional ileitis or terminal ileitis inappropriate. Subsequently, regional enteritis or granulomatous enteritis has been used, but the eponymous term ‘Crohn's disease' is preferable, since it does not imply an anatomic site nor the presence of granulomas.

Definition

Crohn's disease may affect any part of the alimentary canal from mouth to anus, although ileal, ileocolonic, or colonic involvement are the most common patterns. The disease is characteristically discontinuous, or in other words, the disease may affect a number of segments with relatively normal intestinal mucosa in between. The inflammation is often transmural but is predominantly submucosal and is associated with a chronic inflammatory infiltrate, fissures, ulcers, and fibrosis, which may lead to stricturing. Granulomas, with giant cell formation, are the hallmark of the disease but are not always present.

Epidemiology

Crohn's disease has a worldwide distribution, although there is considerable variation in its actual incidence. In northern Europeand the United States of America, the annual incidence is about 5to 8 per 100 000 of the population, with a prevalence of about 60 to 80 per 100 000. In Japan the disease appears to be about 10 times less common, and it is rarer in developing countries. However, diagnostic difficulties are considerable in those parts of the world where intestinal tuberculosis is still common.

The incidence of the disease in Britain, Scandinavia, and Israel increased during the 1950s and 1960s (Table 1), but it may now have reached a plateau. The incidence of the disease may be increasing in children. Ethnic differences also influence the incidence, with particularly high incidence rates in Ashkenazi Jews and a low incidence in Americans of African descent. Some series show a slight female preponderance, but this is not marked. The peak age of onset is in young adults (20 to 40 years of age), but the disease may present at virtually any age.

Country	*Time period*	*Incidence*
USA		
Baltimore	1960–63	1.7
	1977–79	3.1
Olmsted County	1955–75	13.5
	1976–82	3.9
UK		
Aberdeen	1955–57	1.2
	1964–67	2.8
	1967–69	4.5
	1970–75	2.6
Scandinavia		
Stockholm	1955–59	1.5
	1965–69	3.6
	1970–74	4.5
	1975–79	4.1
Copenhagen County	1964–69	1.3
	1970–78	2.7
Israel	1971–75	0.5
	1976–80	1.8

Table 1 Incidence rates (per 100 000 population) of Crohn's disease.

The pattern of the disease appears to be similar wherever it occurs in the world, although extraintestinal manifestations are probably more common in Caucasians. Ileal disease is particularly interesting in Japan, since it is often associated with very extensive longitudinal ulcers. These can be seen in Western patients but they are rare.

Genetics

Although the prevalence of familial disease varies in different ethnic groups, studies from northern Europe and the United States have consistently demonstrated a high prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. First degree relatives, particularly siblings, are most commonly affected in multiply affected families with Crohn's disease and considerable concordance has been reported for disease extent, severity, and extraintestinal manifestations. Spouses and adopted family members appear to be at no increased risk, compared with the general population. In Oxford, familial prevalence of the disease was studied in relatives of 433 patients with Crohn's disease. The relative risk in siblings of patients with Crohn's disease as compared with the general population was 35.6 for Crohn's disease and 16.6 for ulcerative colitis. These data are consistent with a strong genetic component. However, no simple mendelian mode of inheritance was attributable to Crohn's disease in this study or in others. Such data, implicating both genetic and environmental factors, were supplemented by the reported concordance rates in twin pairs. In the most systematic analysis, derived from the Swedish registry of twin births, 8 out of 18 monozygotic twins were concordant, compared with only 1 of 26 dizygotic twins.

These data, the variability in disease prevalence in different ethnic groups, and the association between Crohn's disease and other disorders with known important

genetic influences (primary sclerosing cholangitis, Turner's syndrome, ankylosing spondylitis) underline the contribution of inherited susceptibility. The model of disease inheritance which appears most pertinent to inflammatory bowel disease is that Crohn's disease and ulcerative colitis are related polygenic diseases, sharing some but not all susceptibility factors.

Progress is being made to determine the location of these susceptibility genes. Techniques which allow the whole genome to be searched for linkage to specific diseases has identified linkage between Crohn's disease and the pericentromeric region on chromosome 16. This linkage has been confirmed in various populations and studies on this region are now underway to determine gene identification. Susceptibility loci have also been described on chromosomes 3, 7, and 12 which are shared for both Crohn's disease and ulcerative colitis. Studies are in progress to narrow the regions of linkage in order to begin to identify specific genes. In all of these regions of linkage there are genes which encode for cytokines, cytokine receptors, growth factors, and mucins, any of which may be relevant for the pathogenesis of chronic inflammation. However, for Crohn's disease in contrast to ulcerative colitis, genes of the histocompatibility complex (HLA class I or II) do not seem to be important determinants.

Aetiology

Infective agents

Mycobacterium paratuberculosis

The presence of granulomas suggested a mycobacterial aetiology to the clinicians of the 1930s, but no evidence for such an organism could be found.

However, in 1984, isolates of *Mycobacterium paratuberculosis* from Crohn's tissue were reported and the isolated organisms induced intestinal inflammation when given, by mouth, to newborn goats. Since then, several more isolates have been made in different countries and, using DNA probes, most of these have been shown to be identical with *M. paratuberculosis*. Nevertheless, the significance of these isolates remains very doubtful, since less than 10 per cent of tissues cultured have proved positive. In addition, the clinical response of Crohn's disease to antimycobacterial therapy has not been overwhelmingly successful and, in controlled trials, little effect has been seen. Attempts to detect specific DNA sequences of *M. paratuberculosis* in tissue using the polymerase chain reaction (**PCR**) have again met with mixed results. Where positive results in about two-thirds of patients with Crohn's disease have been reported, about 10 per cent of healthy individuals and even a higher percentage with ulcerative colitis are also positive. Most studies have failed to find any evidence for *M. paratuberculosis* DNA in tissue. However, the organism can be found in milk and water and it is therefore possible that it may be a secondary invader of previously inflamed tissue. Even if this were the case, the organism may still be relevant in so far as it may influence the inflammatory response, for example by inducing granuloma formation.

Measles

The possibility that the granulomatous vasculitis seen in Crohn's disease might represent a persistent measles infection was suggested by some preliminary epidemiological and molecular biology observations. However, these have not been confirmed and measles RNA cannot be found in Crohn's tissue using PCR techniques. Thus, at the present time, there is no irrefutable evidence to implicate measles in the pathogenesis of Crohn's disease.

Other infective agents

Many other infectious agents have been implicated (viruses, cell-wall-deficient pseudomonads) but no convincing evidence for them has accumulated. However, several lines of evidence suggest that luminal factors may be important. Active disease often responds to the use of elemental diets, which appear to be almost as effective as corticosteroids. Colonic disease, resistant to medical treatment, frequently subsides following defunctioning of the colon by means of a split ileostomy. Challenging the defunctioned colon with ileostomy effluent will trigger an inflammatory response, although this is not seen if an ultrafiltrate of the effluent is used. Thus, luminal bacteria may have some influence on the course of the disease, although the effects of elemental diets and defunctioning may equally well implicate a biochemical or metabolic alteration in the luminal environment.

Environmental factors

Many studies have now shown that patients with Crohn's disease are more likely to be smokers, and this is in sharp contrast with ulcerative colitis, which is found more often in non-smokers. Thus smoking confers a relative risk of 3 to 4 for Crohn's disease and smokers tend to have more aggressive disease and an earlier postoperative relapse than non-smoking patients with Crohn's disease. The reasons for this are completely obscure. Smoking is known to affect the synthesis of colonic mucus and mucosal blood flow. It also affects arachidonic acid metabolism and some immunological effector mechanisms. However, whether any of these mechanisms can explain the opposite associations between smoking and ulcerative colitis or Crohn's disease remains speculative.

There may be an increased incidence of Crohn's disease in young women receiving oral contraceptives. However, the association, if it exists, is weak and largely disappears when the data are adjusted for smoking habits and social class.

Events in early childhood may be risk factors for Crohn's disease later in life. These include early weaning and a higher socio-economic environment as shown by lack of crowding and the presence of running hot water in the home. Some series, but not all, have suggested that the disease is more prevalent in ‘white-collar' than in ‘blue-collar' groups, which would again fit with the hypothesis that ‘cleaner' environments may predispose to the disease.

Immunopathogenesis

There is considerable evidence to suggest that immunological mechanisms play a role in the pathogenesis of mucosal inflammation. Thus, intestinal macrophages and T cells are activated within the inflamed intestine and release a wide variety of cytokines as well as inflammatory mediators. These soluble mediators have profound local and systemic effects, which include altering epithelial cell permeability and ion flux, initiating fibrosis by activating fibroblasts, damaging endothelium resulting in local ischaemia, and stimulating an acute-phase response (Table 2).

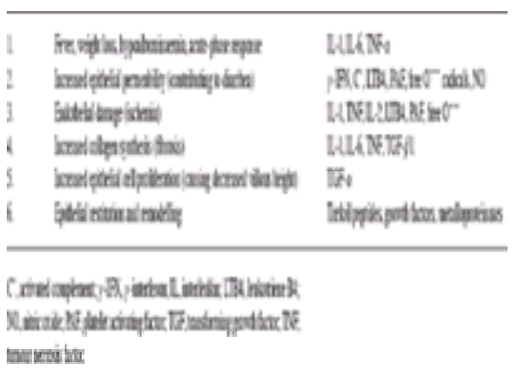


Table 2 Pathogenesis of Crohn's disease—the potential role of soluble mediators released during immune activation.

Animal models, generated by the specific deletion of particular immunoregulatory genes, have provided insight into the contribution of cytokines in disease pathogenesis. It is of particular interest that only certain targeted cytokine deletions result in inflammatory bowel disease in animals. Deletion of the genes encoding interleukin 2 (**IL-2**) and interleukin 10 (**IL-10**) results in chronic intestinal inflammation. In contrast, deletion of the genes encoding interleukin-1b (**IL-1b**), tumour necrosis factor-a (**TNF-a**), interleukin 4 (**IL-4**), g-interferon, and interleukin 6 (**IL-6**) do not result in intestinal disease. Furthermore, it is apparent from the effects of germ- and pathogen-free conditions on these models that the bacterial flora of the gut are necessary to perpetuate the disease.

The more difficult question is whether Crohn's disease represents a specific immunological disturbance, or whether the changes noted represent an appropriate immunological response to an environmental pathogen. For example, does the chronic inflammatory nature of the disease reflect an abnormality in the regulation of the mucosal immune response? So far, evidence to support such a hypothesis is lacking. Assays for T-cell suppressor or helper function, whether using peripheral blood or mucosal lymphocyte populations, have produced variable results. However, there are data suggesting that patients with Crohn's disease, in common with patients with

ulcerative colitis, may have an impaired ability to induce suppressor cells to specific antigens. This inability is present in patients whose disease is in remission and on no therapy, suggesting that there may, indeed, be an underlying problem in immunoregulation. Disturbances in immunoregulation could be mediated by changes in the subsets of regulatory T cells. Some studies have suggested that Crohn's disease may be a 'TH_1-mediated' disease, meaning that T-helper cells producing IL-2, g-interferon, and IL-12 may predominate over the TH2 population of T-helper cells that are associated with the production of IL-4, IL-5, and IL-10. However, there is no uniform agreement between studies and the issue is currently unresolved. Nevertheless, the TH_1 cells are undoubtedly important since g-interferon is essential for granuloma formation.

Deficient immunoregulation would allow an immune response in the intestinal mucosa to proceed unchecked and might explain why these patients exhibit humoral and cellular immune responses to a wide variety of luminal and epithelial antigens. A particularly interesting finding is that, although patients with active ulcerative colitis or Crohn's disease may show a rise in total serum IgG concentration, the patients with Crohn's disease show a predominant rise in the IgG2 subclass whereas those with ulcerative colitis have a predominant increase in IgG1. This difference could indicate that different antigens are involved in each disease, since protein antigens are associated with an IgG1 response whereas carbohydrate or bacterial antigens stimulate an IgG2 response. Equally, it could represent a genetic difference between the two diseases.

Autoantibodies are infrequently found in patients with Crohn's disease, in contrast to ulcerative colitis. Antineutrophil cytoplasmic antibodies with a perinuclear staining pattern (**pANCA**) are found in over 60 per cent of patients with ulcerative colitis but in less than 15 per cent of patients with Crohn's disease. Nevertheless, pANCA are present in patients with Crohn's disease affecting the left colon—the 'UC-like' Crohn's disease. This is another pointer to clinical and genetic heterogeneity within 'inflammatory bowel disease'. More recently, Crohn's disease has been associated with serum antibodies (**ASCA**) to baker's yeast (*Saccharomyces cerevisiae*), which are rarely found in ulcerative colitis. The detection of pANCA/ASCA is currently being explored as a means of differential diagnosis between the two diseases.

Pathology

Crohn's disease affects the small intestine in at least 70 per cent of all patients with the disease. Of these, about half have ileal involvement alone and the rest will have associated colonic disease—usually affecting the right colon. However, 30 per cent of patients with ileal disease alone will show a proctitis on sigmoidoscopy.

Macroscopically, the affected segment of small intestine is thickened due to oedema and fibrosis, which can be severe enough to cause a stricture. The mucosa is inflamed and may have a 'cobblestone' appearance, due to a combination of submucosal oedema and fissuring ulcers (Fig. 1). In mild disease, small aphthoid ulcers may be visible on the mucosa. In more severe disease, these ulcers become larger and coalesce. Ulceration is often serpiginous, especially on the mesenteric border. On the serosal surface, Crohn's disease may have a fairly characteristic appearance, with creeping fat and marked vascularity.

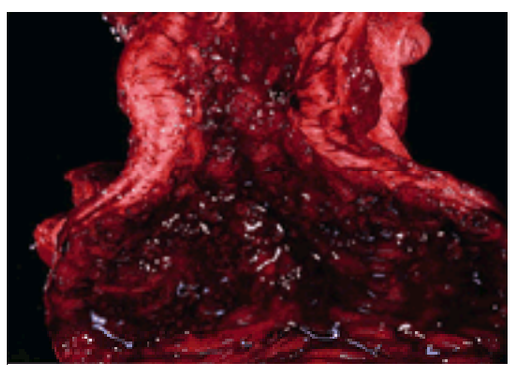


Fig. 1. Macroscopic appearances of ileal Crohn's disease, showing thickening of the wall and extensive ulceration.

The predominant histological feature is a chronic inflammatory infiltrate of lymphocytes and plasma cells in the lamina propria and submucosa (Fig. 2). There is usually shortening of villous height with increased crypt depth, and there may be evidence of 'pyloric metaplasia'.

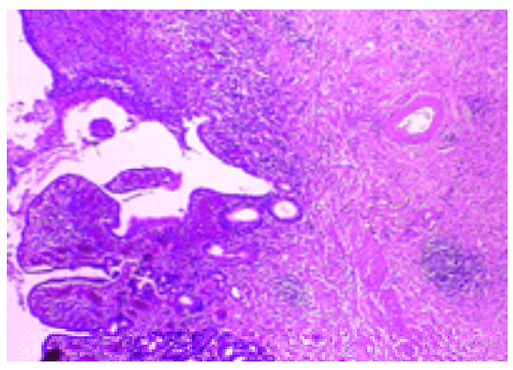


Fig. 2. Low-power views of ileal Crohn's disease, showing expansion and inflammation of the submucosa and a fissure ulcer in the mucosa (by courtesy of Dr B.F. Warren).

This last finding has been investigated recently and represents a novel lineage of cells budding from the adjacent crypts. This appears to be part of a normal mucosal repair mechanism and may develop under the influence of epidermal growth factor.

When the Crohn's disease is active there is also an increase in neutrophils and eosinophils, and there is usually ulceration of the surface epithelium. One characteristic feature is the fissure ulcer, which often penetrates deep into the mucosa (Fig. 2). These fissures are lined by macrophages and lymphocytes and may be associated with granulomas and multinucleated giant cells (Fig. 3). Fibrosis is another prominent feature, especially in the submucosa. The inflammation of Crohn's disease is often transmural and therefore it is common to see chronic inflammatory changes with granuloma formation on the serosal surface.

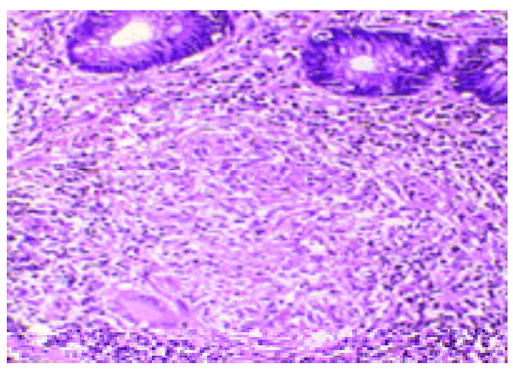


Fig. 3. A granuloma containing a multinucleated giant cell in the submucosa of the ileum (by courtesy of Dr B.F. Warren).

Lymphoid follicles are prominent throughout the intestinal wall including the serosa. However, there may be small, localized areas of ulceration in the overlying epithelium. It has been postulated that these lesions represent the earlier lesion of Crohn's disease, but this remains a speculative, though attractive, hypothesis.

Granulomas occur with variable frequency, partly dependingon the number of sections examined and the time spent searching for them. However, they are probably present in no more than 65 per cent of patients with ileal disease, and some series quote only 30 per cent. Certainly there appear to be fewer granulomas in ileal disease than in colonic Crohn's disease, the frequency increasing as the disease becomes more distal. Granulomas are often associated with blood vessels (Fig. 4) or lymphatics and experiments, using operative specimens injected with intra-arterial resins, have shown that the granulomas are often within the walls of arterioles. Although the presence of granulomas has been claimed to indicate a good prognosis, subsequent studies have not confirmed this.

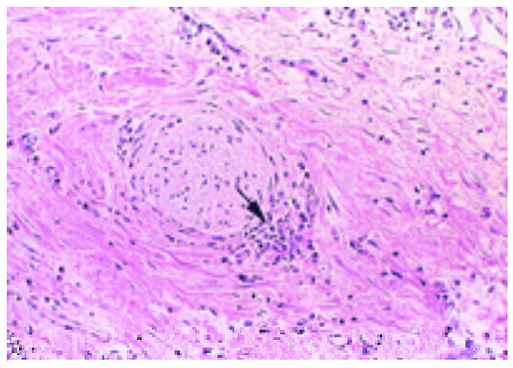


Fig. 4. Granuloma (arrow) in the wall of a submucosal arteriole (by courtesy of Dr B.F. Warren).

Neuronal hypertrophy is another histological feature that is frequently present. Nerve endings, in particular, become prominent and stain strongly for vasoactive intestinal polypeptide, but whether this is specific for Crohn's disease, or whether it is merely associated with the neural hypertrophy known to occur proximal to obstructive lesions, is not known.

Clinical features

Symptoms

The most common symptoms of small intestinal Crohn's disease are diarrhoea (90 per cent), abdominal pain (55 per cent), anorexia, nausea, and weight loss (22 per cent). Malaise, lassitude, and the symptoms of anaemia are frequently present.

The diarrhoea usually consists of the passage of frequent loose or watery stools. Nocturnal diarrhoea may occur. Frank blood in the stool is not usually seen, although small amounts of blood may be passed in patients with an accompanying proctitis. A few patients may have a steatorrhoea but the passage of pale, bulky stools that are difficult to flush away is uncommon.

The pain may either be a dull ache, often located to the right iliac fossa, or a periumbilical colicky pain. Both types of pain can radiate into the back.

Some patients present with general symptoms of vague ill health and fever, with few, if any, abdominal symptoms; this usually leads to a considerable delay in diagnosis and is seen particularly in children, in whom failure to thrive or growth failure are common presentations.

The symptoms of small intestinal Crohn's disease may be caused by active inflammation, intestinal obstruction, or both. Patients with previous ileal resections may have a watery diarrhoea due to malabsorption of bile salts, or their symptoms of anaemia may be due to vitamin B12 deficiency. Partial obstruction from stricturing disease may cause bacterial overgrowth, which may cause steatorroea, bloating, excess wind, weight loss, and a general malaise.

Thus, the symptomatology of Crohn's disease is complex, and an awareness of the differing causes of symptoms is important in making an accurate clinical assessment.

Physical signs

Many patients look well, are well nourished, and have few abnormal physical signs. However, evidence of weight loss, anaemia, and signs of iron deficiency and clubbing of the nails are common. Peripheral oedema may occur if hypoalbuminaemia is present. In very sick patients, there is often marked cachexia, proximal myopathy (from hypokalaemia or vitamin D deficiency), easy bruising, and pigmentation.

Abdominal examination may be normal, but right iliac fossa tenderness is common. Thickened loops of ileum or a frank abdominal mass may be palpable. Visible peristalsis usually denotes obstruction but, in cachectic patients, may just be normal small-bowel movement. Perianal disease, which includes fleshy skin tags, fissures, or fistulas, can occur in up to 14 per cent of patients with ileal disease. Rectal examination may demonstrate tender pelvic loops of inflamed intestine.

Acute presentation

Although the majority of patients complain of symptoms which have been present for weeks or months, a few (less than 10 per cent) present with a more acute history. This often consists of right iliac fossa pain, fever, and malaise over a few days. These patients are frequently diagnosed as having acute appendicitis, and the true diagnosis is only revealed at operation. Nevertheless, a careful history often reveals more long-standing symptoms, and investigation often shows a mild anaemia or a lowish albumin. Thus, however classic the clinical picture for acute appendicitis appears to be, a full history, together with documentation that signs of chronic disease are absent, documentation of a normal anus, and a normal haemoglobin, should be essential before proceeding to appendectomy.

Diagnosis

For patients presenting with non-specific abdominal symptoms, in whom a diagnosis of irritable bowel syndrome seems likely, documented weight loss and nocturnal diarrhoea needs careful investigation. A full blood count, erythrocyte sedimentation rate, and levels of C-reactive protein, albumin, serum iron, folate, and B12 should be requested—any abnormality should trigger more extensive investigation.

The diagnosis of small intestinal Crohn's disease is made by barium contrast radiology. When available, the technique of enteroclysis (small-bowel enema) should be used, as this allows much better definition of the mucosal pattern than a conventional barium meal and follow through. Thus, scattered aphthoid ulcers are much more likely to be seen on a small-bowel enema, and it also allows a much better assessment of strictures. More advanced disease will be associated with the appearance of fissure ulcers, mucosal oedema, and narrowing of the intestinal lumen (Fig. 5 and Fig. 6). Fistulas to other loops of bowel or to other organs may be visualized (Fig. 7 and Fig. 8).

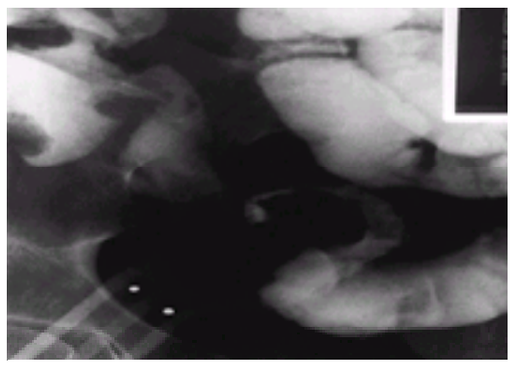


Fig. 5. Small-bowel enema, showing Crohn's disease of the terminal ileum. The ileum is thickened, strictured, and shows mucosal oedema. There is proximal dilatation, indicating obstruction.

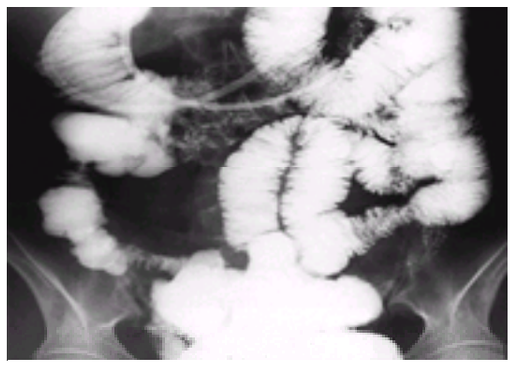


Fig. 6. Ileal Crohn's disease, shown by a small-bowel enema, illustrating fissure ulcers and characteristic asymmetry.

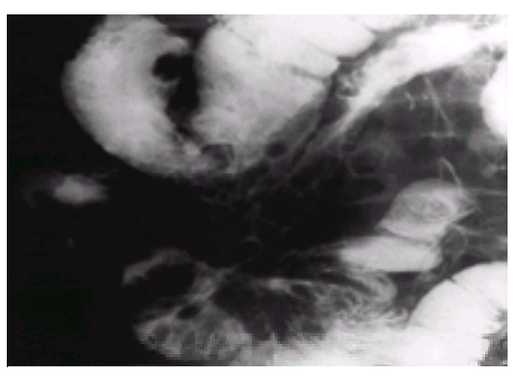


Fig. 7. Severe ileal Crohn's disease with multiple fistulas into neighbouring loops.

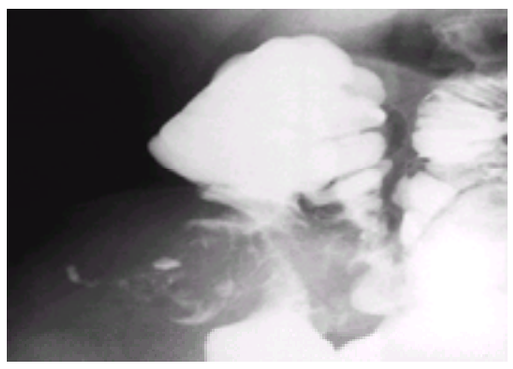


Fig. 8. Crohn's disease of the terminal ileum, associated with a large abscess resulting from fistula formation and perforation.

When a small-bowel enema confirms Crohn's disease, a full examination of the colon should always be made. It may be sufficient to perform a sigmoidoscopy with rectal biopsy and a double-contrast barium enema. Nevertheless, the barium study will miss scattered aphthoid lesions in the colon and so, for a full assessment, colonoscopy should be performed. The importance of taking biopsy specimens cannot be overemphasized since focal inflammation and even granulomas can be seen histologically in 25 to 30 per cent of patients in whom the macroscopic appearance of the mucosa is normal. Colonoscopy may also allow biopsy specimens of the terminal ileum to be obtained, which can be helpful in cases where the radiological appearances are suggestive but not diagnostic.

White-cell scans (using indium- or technetium-labelled granulo-cytes) may also be helpful in determining the extent of the disease (Fig. 9). A major use is in assessing patients where symptoms are more severe than would be expected from the barium radiology and in excluding local perforation with abscess formation.

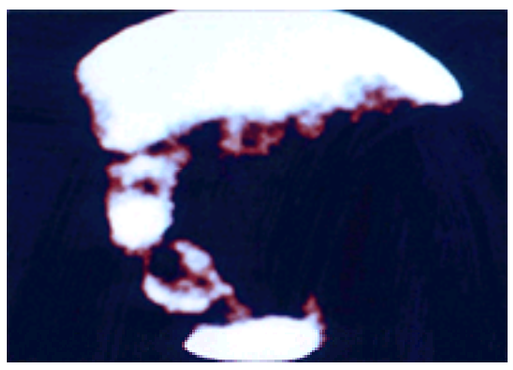


Fig. 9. ^{99}Tcm-labeled white-cell scan in a patient with ileal Crohn's disease. The liver is seen clearly, but, in addition, the white cells have accumulated in loops of the ileum and the ascending and transverse colon, indicating active disease in these segments.

Computed tomography (**CT**) scans, especially following contrast medium, may demonstrate thickening of the intestinal wall, but they are of low specificity. CT scans

and magnetic resonance imaging can be very useful in the assessment of fistulas or pelvic abscesses.

Differential diagnosis

As already mentioned, an acute ileitis frequently mimics an acute appendicitis. Nevertheless, only a small proportion of patients with an acute ileitis have Crohn's disease, on the grounds that only about 10 per cent will have recurrent disease on prolonged follow-up. The remainder are assumed to have an infective cause; *Yersinia enterocolitica* or *Y. pseudotuberculosis* account for some of these. These organisms can be isolated from stool cultures, but a rise in serum antibody titer is the most reliable method of making the diagnosis.

For more chronic disease, the differential diagnosis includes lymphoma, Behçet's disease, tuberculosis, actinomycosis, and tumours (adenocarcinoma, carcinoid). A small-intestinal lymphoma may be difficult to distinguish from Crohn's disease, except following laparotomy. The presence of focal colonic disease at colonoscopy, and the appearances on an abdominal CT scan, may be helpful. Behçet's disease of the small intestine is primarily a vasculitic process and virtually always occurs when there is evidence of disease elsewhere (for example, ocular or genital ulceration, deep vein thrombosis, neurological involvement). Ileal perforation is not uncommon in Behçet's disease but is very rare in Crohn's disease of the small intestine. Radiation enteritis may also mimic Crohn's disease, but a history and radiological appearances make the diagnosis clear.

The most difficult differential diagnosis is intestinal tuberculosis when Crohn's disease is suspected in patients at high risk from tuberculosis (for example Asian patients). Stool culture and serum antibody titers are unhelpful. Radiologically, tuberculosis often affects a very short segment of the ileum and is not usually associated with cobblestoning or asymmetry, which are so characteristic of Crohn's disease. Skip lesions are uncommon in tuberculosis. Colonoscopy with multiple biopsies may be helpful, as caseating granulomas containing acid-fast organisms can be found in a few patients. Finally, laparoscopy may show serosal or peritoneal tubercules. Obviously, active tuberculosis elsewhere (such as lungs and kidney) makes the diagnosis of the intestinal disease, and hence the management, straightforward.

Complications

Perforation, acute dilatation, and massive haemorrhage may all occur in small-intestinal Crohn's disease but they are rare, being much less common than in Crohn's disease of the colon. Obstruction due to fibrous strictures is the most common local complication and the strictures may be multiple. Fistulas to neighbouring loops of small or large bowel or to other organs such as the bladder or vagina occur in 10 to 17 per cent of patients.

Gallstones are more frequently found in patients with ileal Crohn's disease, particularly if an ileal segment has been excised, than in a healthy population. The mechanism is the interruption of the enterohepatic circulation of the bile salts, with eventual depletion of bile salts.

Renal stones are also more commonly present in patients with Crohn's disease. They are usually oxalate stones, secondary to hyperoxaluria. This occurs when there is steatorroea as the fat binds intraluminal calcium. Thus the normal precipitation of calcium oxalate does not occur, leaving free oxalate to be absorbed from the colon and excreted through the kidneys. Another renal complication is the involvement of the right ureter in the inflammatory process occurring in the terminal ileum and cecum. This may give rise to a recurrent pyelonephritis or, more commonly, a right hydronephrosis.

Amyloidosis is a rare but well-recognized complication of long-standing Crohn's disease. When it occurs, it is often present throughout the intestinal mucosa and may also affect the kidney, leading to nephrotic syndrome and renal failure. If renal amyloid occurs, it is important to get the Crohn's disease into remission (that usually means resection) as the amyloid has been documented to regress subsequently.

Adenocarcinoma complicating small-intestinal Crohn's disease is very rare but is well recognized. A review of the 58 reported cases up to 1982 showed that the majority occurred in the ileum; 18 occurred in a bypassed loop. Most patients died within a year of diagnosis, presumably indicating the difficulty of making the diagnosis in a patient who already has Crohn's disease and abnormal radiological appearances of the small intestine.

Extraintestinal manifestations

Small-intestinal Crohn's disease, in contrast to colonic disease, is rarely associated with erythema nodosum, uveitis, or an acute arthropathy, which occur in about 5 per cent of patients with disease confined to the ileum. If these extraintestinal manifestations represent deposition of immune complexes, and the evidence for this is only suggestive, then it is possible that the greater antigenic load within the colonic lumen may explain their association with colonic disease. The peripheral arthropathy can be classified into a large joint pauciarticular arthropathy which normally goes along with active disease and a small joint symmetrical arthropathy which is much more persistent. These two types have different HLA associations suggesting that a patient's genotype will determine susceptibility to develop the complications.

Sacroiliitis occurs radiologically in 10 to 12 per cent of patients, although only a small proportion will complain of low back pain. Unless the patient has the HLA B27 haplotype, sacroiliitis does not develop into a full-blown ankylosing spondylitis. This latter condition occurs in about 1 per cent of patients with Crohn's disease, who frequently have the HLA B27 haplotype. It is more commonly seen in patients with colonic involvement than in those with only ileal disease. Ankylosing spondylitis may present long before intestinal disease is suspected and is characterized by progressive stiffness of the back, leading ultimately to a rigid spine. The natural history of the back condition is independent of the Crohn's disease and, therefore, the presence of disabling back symptoms is not an indication for extensive surgical resection of the Crohn's disease.

Pyoderma gangrenosum may also complicate active Crohn's disease but, again, it is very rare when there is no colonic involvement.

Fatty liver may occur in any patient who is severely ill from Crohn's disease and is reversible once remission is achieved. During these severe attacks there may be transient rises in serum concentrations of aminotransferases or alkaline phosphatase, but they have no clinical significance.

Isolated granulomas may occur within the liver. Primary sclerosing cholangitis seems to be much less common in Crohn's disease than in ulcerative colitis, and is usually diagnosed on the basis of a persistently raised serum alkaline phosphatase. Liver biopsy may reveal the classic periductular fibrosis with chronic portal inflammation, but the appearances range from a chronic active hepatitis picture with portal inflammation and piecemeal necrosis, to gross fibrosis with obliteration of the bile ducts. The definitive diagnostic test is endoscopic retrograde cholangiopancreatography, which will show the distribution of the disease—intrahepatic, extrahepatic, or a combination. Occasionally, there is a predominant stricture in the extrahepatic bile duct which can be worth dilating or stenting if the patient has an obstructive jaundice. Cholangiocarcinoma, a well-recognized complication of sclerosing cholangitis in patients with ulcerative colitis, is very rare in Crohn's disease and very few properly documented cases have been described. Whether this relates just to the low incidence of sclerosing cholangitis in Crohn's disease or whether other mechanisms are operating is unclear.

Assessment of disease activity

Since patients with Crohn's disease may have symptoms relating to active inflammation, bacterial overgrowth, obstruction, or as a result of previous surgery, assessment of the activity of the disease is difficult and a clinical assessment by itself can be misleading. Laboratory data frequently help and should include the haemoglobin, platelet count, albumin, and erythrocyte sedimentation rate. When available, serum concentrations of the acute-phase proteins, C-reactive protein and orosomucoid are particularly good indicators of active inflammation.

A large number of activity indices have been devised in order to standardize activity assessment. The Crohn's Disease Activity Index is the most widely used, as well as the Harvey–Bradshaw simplification of this index. However, the calculation of activity indices is subject to wide variation and their use is mainly for clinical trials rather than in routine clinical practice.

Labeled white-cell scans are often useful in patients when there is still doubt about whether symptoms are primarily inflammatory or obstructive. The distinction is obviously important for management decisions.

Medical management

Since Crohn's disease cannot be cured, the role of the clinician is to control the inflammation, to correct nutritional deficiencies, and to ameliorate symptoms. These aims will frequently involve surgery and, indeed, 70 to 75 per cent of patients will require at least one operation during their lifetime. Thus, management of these patients requires close co-operation between physicians and surgeons. In general, medical treatment should be dictated by symptoms. Large doses of corticosteroids or immunosuppression are not indicated in asymptomatic patients who have active disease shown only by laboratory data.

Nutrition and diet

Patients with severe disease are often malnourished and will require both nutritional replacement and support during the acute illness. Nutritional support may be parenteral or enteral, the latter being administered through a fine-bore nasogastric tube.

Parenteral nutrition by itself has no primary effect on the disease process, but several trials have now shown that elemental diets are as effective as corticosteroids in obtaining remission. However, meta-analyses of all the comparative trials have shown a significantly higher remission rate with steroids than elemental or polymeric diets. The unpalatability of elemental diets makes for poor patient compliance and there is some evidence that relapses occur more quickly after dietary therapy than following corticosteroids. Nevertheless, dietary therapy can be useful, especially in patients who are intolerant of steroids and in children. The diet can be given either by nasogastric tube or by a percutaneous endoscopically placed gastric tube which allows the disease to settle without risking further growth retardation. In the long term, patients should be advised to eat a normal well-balanced diet, avoiding only those foods which they know upset them.

A few patients with extensive small-bowel disease may have hypolactasia and will benefit from a lactose-free diet. The use of elimination diets is still unproven. However, a low-residue diet is advisable for patients with stricturing disease.

For most patients who have intermittent relapses of distal ileal disease, vitamin supplements are unnecessary, provided a serum vitamin B12 concentration is measured regularly (an annual measurement should be sufficient unless the value is progressively falling). However, patients with more extensive disease may benefit from folic acid, iron, B vitamins, and calcium. Magnesium, zinc, and selenium deficiencies may also occur, and serum concentrations should be measured in patients with extensive disease that is chronically active. Oral iron is frequently poorly tolerated by these patients, and a total dose of iron given by slow intravenous infusion is often indicated. There is a potential danger of anaphylaxis but this occurs very rarely. Nevertheless, it is a wise precaution to have epinephrine and hydrocortisone readily available and the patient closely monitored while the infusion is in progress.

Treatment of mild to moderate disease

Early clinical trials suggested that patients with mild symptoms associated with only minimal changes in their inflammatory markers (platelet count, C-reactive protein, erythrocyte sedimentation rate) might respond to 5-aminosalicylic acid; sulphasalazine seemed to be particularly effective for colonic disease. However, there are now a large number of trials (not all of which are published) and meta-analysis of published data shows only minimal benefit. The most optimistic trial suggested that 4 g mesalazine daily was significantly better than placebo, but two subsequent studies by the same group, which were well designed with large numbers of patients to obtain adequate power, have failed to show a difference between this high dose and placebo. Furthermore, budesonide (see later) has been shown to be significantly better than 4 g mesalazine for treating active disease.

Thus, mild to moderate disease is best treated with corticosteroids. This usually implies starting with 40 to 60 mg prednisolone daily, tapering to 40 mg daily over a period of 2 to 3 weeks. Patients are then treated for a further 4 weeks before the prednisolone is tailed off completely. If the disease is going to respond to corticosteroids, that is usually achieved by 6 weeks of therapy. In formal clinical trials, there has been no advantage in combining corticosteroid therapy with sulphasalazine. In France, an initial dose of 1 mg/kg (up to a maximum of 70 mg) is recommended and, on this dose, over 80 per cent of patients go into clinical remission.

Recently, a controlled ileal release form of budesonide has been introduced. Budesonide is a highly potent steroid that, following absorption, undergoes 90 per cent first-pass metabolism in the liver. Thus, it might be expected to have a powerful, local effect in the diseased ileum, but with very little systemic adverse reaction. Clinical trials have shown that 9 mg budesonide is virtually as effective as oral prednisolone (in the doses described above) for treating ileal or ileocolonic disease and the side-effect profile is much better than prednisolone; it is nevertheless not completely free of steroid side-effects.

The role of antibiotics for treating active Crohn's disease is controversial. One controlled clinical trial suggested that metronidazole was better than placebo and a further study showed that it was equivalent to sulphasalazine. In clinical practice, metronidazole and particularly ciprofloxacin are often used in North America to treat active Crohn's disease, but the evidence to support this practice is weak.

Treatment of severe disease

A patient with severe Crohn's disease has symptoms of active disease (diarrhoea, abdominal pain, anorexia, weight loss) and evidence of systemic disturbance (tachycardia, fever, anaemia). An abdominal mass may be present. These patients are best admitted to hospital, and they respond well to intravenous corticosteroids (for example, hydrocortisone, 100 mg every 6 h), intravenous fluids, and electrolyte and blood replacement. Intravenous metronidazole (500 mg every 8 h) is also often given, although there is no firm evidence to support its use. The majority of patients settle rapidly on this regimen and can be started on a light diet after a few days. For those who fail to respond, the addition of cyclosporin (4 mg/kg by continuous intravenous infusion) may be helpful.

For patients with severe ileal disease complicated by stricturing or fistulas, intravenous corticosteroids and metronidazole should be given initially to allow much of the inflammation to settle before proceeding to surgery.

Those severely ill patients who are not responding well to corticosteroids might now be considered for an infusion of anti-TNF antibody (Infliximab). This has now been approved by the United States Food and Drug Administration but, in Europe, is not yet licensed. Another therapeutic possibility is the administration of intravenous azathioprine, which appears to produce a rapid response and is followed by oral azathioprine.

Parenteral nutrition, while not indicated for all patients requiring intravenous corticosteroid therapy, should be considered for each patient, especially if complications such as fistulas are present and if surgery seems a likely outcome.

Chronic active disease

Some patients go into a satisfactory remission but become symptomatic when their prednisolone dosage falls below 15 or 10 mg daily. Yet other patients rapidly relapse once prednisolone is withdrawn completely. These patients can be considered to have chronic active disease and it is this group that may benefit from immunosuppressive therapy.

In Europe azathioprine (2 mg/kg) is usually used, but in the United States 6-mercaptopurine is more commonly used (azathio-prine is metabolized to 6-mercaptopurine in the liver). Some patients respond well to these drugs and it is then possible to withdraw the corticosteroids. The exact mode of action of these drugs is not known but their beneficial effect, if it occurs, is seen over a period of several weeks. If a patient responds well, the problem then is to know how long to continue treatment. Current anecdotal data suggest that positive benefit continues over 4 to 5 years and potential side-effects of long-term administration have not been seen (e.g. lymphoma, carcinoma).

Side-effects are fairly uncommon, with nausea and diarrhoea being the most frequent. However, regular blood counts (every 1 to 2 months) should be done, as marrow suppression occasionally occurs. This complication is usually seen in patients who are deficient in the thiopurine methyl transferase enzyme, which occurs in about 1 in 500 of the normal population. Minor elevations of aminotransferase and an acute pancreatitis are other potential adverse reactions.

Cyclosporin A has been used for chronic active disease at a dose of 5 mg/kg but three of four trials have failed to show benefit. It is not recommended for general use at present because adverse reactions, in particular nephrotoxicity, are common. Methotrexate, intramuscularly at a dose of 25 mg weekly, has been shown to be of benefit in resistant disease and allows corticosteroids to be withdrawn. Anti-TNF antibody might also be useful in this resistant group of patients.

Symptomatic treatment

Diarrhoea in patients with Crohn's disease is best treated by dealing with the cause (active disease, bacterial overgrowth, etc.). Patients with a bile salt-induced diarrhoea following ileal resection may benefit from cholestyramine, but this must be used sparingly if more than 100 cm of terminal ileum has been resected as it will exacerbate fat excretion. However, a few patients will continue to complain of diarrhoea when all treatable causes have been identified and corrected. Antidiarrhoeal drugs (loperamide, codeine phosphate, diphenoxylate hydrochloride) may then have a role.

Maintenance of remission

In contrast to ulcerative colitis, there is little evidence that 5-aminosalicylic acid (mesalazine, olsalazine, sulphasalazine) is useful for the maintenance treatment of

Crohn's disease. Although some trials have been positive, the latest meta-analysis of all published trials has shown only a marginal benefit. Nevertheless, there is more convincing data that high-dose mesalazine (3 to 4 g daily) may reduce postoperative recurrence in patients undergoing ileal or ileocaecal resection. Thus, this group of patients should be treated for at least 5 years following the resection in order to achieve a 15 to 20 per cent reduction in relapse rate.

Dietary trials have shown that, although patients with Crohn's disease as a group eat more sugar and less fiber than healthy individuals or patients with ulcerative colitis, the subsequent course of the disease is not affected by altering the diet to contain less sugar and more fiber. However, the course of the disease is improved and the postoperative recurrence rate is reduced in patients who stop smoking. This is probably the most effective intervention that can be made to ensure a more favorable long-term course.

Prognosis

Crohn's disease is a disease of high morbidity and low mortality. Nevertheless, despite recurrent relapses or, indeed, surgical operations, the majority of patients lead full and active lives if there has been good nutritional support, regular follow-up, and judicious use of medical and surgical treatment.

Following ileal resection, at least 45 per cent of patients will have a recurrence within 5 to 10 years, which tends to be less than the recurrence rate following ileocolonic resection. Of those that do relapse, approximately half will require further surgery. However, the risk of requiring further resections is probably not increased by previous resections, although this point is debated.

Recently, endoscopic examination of the ‘neoterminal ileum' has shown recurrent inflammation in 65 to 80 per cent of patients during the 6 months following ileocolonic resection. Endoscopic appearances range from diffuse inflammation to aphthoid ulceration to confluent ulceration. However, it is not yet entirely clear whether the severity of these endoscopic recurrences predicts the subsequent clinical course.

Survival curves drawn by actuarial methods have shown that mortality tends to increase with the length of history. For the first 15 years of the disease, mortality is no different from a matched control population, but then tends to increase above that expected.

Crohn's disease in pregnancy

Female patients with Crohn's disease are as fertile, in general, as a control population. However, patients should avoid conceiving when the disease is active as there tends to be a high incidence of miscarriage. Apart from this, pregnancy seems to have little effect on the Crohn's disease and therefore patients should not be discouraged from having children if they wish. Should a relapse occur during pregnancy, it should be treated as vigorously as if the patient was not pregnant. Corticosteroids and mesalazine should be used as indicated, and should in no way be withheld. Patients conceiving while on treatment should be reassured that the developing fetus is not at risk, and the drugs should only be withdrawn as dictated by the activity of the disease. Some patients conceive while on azathioprine. Once again, there is no hard evidence that this will harm the fetus. If the Crohn's disease has been difficult to control and has only recently gone into remission with the introduction of azathioprine, the drug should be continued. A recent audit on the outcome of pregnancies in women with Crohn's disease who were receiving azathioprine has not shown an increased incidence of congenital abnormalities. Methotrexate is absolutely contraindicated in pregnancy.

Crohn's disease in children

The disease can occur rarely within the first few months of life, but the frequency then increases after the age of 2 years. Epidemiological studies in the United Kingdom suggest that the incidence of Crohn's disease in children is increasing at a time when it has levelled out, or even fallen, in the adult population. The anatomic distribution of the disease is similar to that seen in adults. Symptomatically, the presentation may also be similar to that in adults, but some children will present with growth failure as the only manifestation. There is frequently a considerable delay in diagnosis, either because of the absence of gastrointestinal symptoms or because the children are labelled as having ‘toddler's diarrhoea', ‘growing pains', milk allergy, or even a functional syndrome. Once the diagnosis is suspected, it should be confirmed with full gastrointestinal assessment as described for adults.

Treatment should follow the same guidelines as those outlined for adults, with appropriate modification in the dosage of the drugs. Many paediatricians are using elemental or polymeric diets as first-line therapy in order to prevent possible growth retardation induced by corticosteroids. For children who require corticosteroids, alternate day therapy should be used as this protects, at least to a degree, against growth retardation. Growth charts should be in the hospital notes, and recordings of height and weight should be made at each visit. A dietitian should document the child's dietary intake at regular intervals and supplements may be necessary to maintain an intake of at least 2000 kcal with 15 g nitrogen daily.

Further reading

Allan RN, Keighley MRB, Alexander-Williams J, Hawkins C, eds. *Inflammatory bowel diseases*, 3rd edn. Churchill Livingstone, Edinburgh, 1997.

Alstead EM, Ritchie, JK, Lennard-Jones JE, Farthing MJG, Clark ML. Safety of azathioprine in pregnancy in inflammatory bowel disease. *Gastroenterology* 1990; **99**: 443–6.

Camma C, Giunta M, Rosselli M, Cottone M. Mesalamine in the maintenance treatment of Crohn's disease: a meta-analysis adjusted for confounding variables. *Gastroenterology* 1997; **113**: 1465–73.

Kirsner JB, Shorter RG, eds. *Inflammatory bowel disease*, 4th edn. Lea & Febiger, Philadelphia, 1997.

Parkes M, Satsangi J, Jewell DP. Mapping susceptibility loci in inflammatory bowel disease: why and how? *Molecular Medicine Today* 1997; **3**: 546–53.

Prantera C, Korelitz BI, eds. *Crohn's disease*. Marcel Dekker Inc., New York, 1996.

24.2.2 Surgery for Crohn's disease of the small intestine

Timothy A. Cook and Neil Mortensen

Introduction
The aims of surgery
Crohn's disease at specific sites in the small intestine
Gastroduodenal
Jejunum and ileum
Surgery
General principles
Ileocolic disease
Multisite small intestinal disease
Fistula and abscess
Appendicectomy and fistula
Free perforation
Fistula and abscess: investigation
Treatment of abscesses and fistulas
Further reading

Introduction

Opinion on the nature of surgery for small-bowel Crohn's disease has varied since the earliest operations described by Crohn and colleagues. At first wide resection was the treatment of choice, but high relapse rates led to a period when bypass procedures were popular. In due course these patients appeared to have a high incidence of septic and fistula complications from their bypassed segments and there was a return to radical surgery once more. Since Crohn's disease is a diffuse intestinal problem, and the majority of patients undergo surgery at some time during the course of their disease, there has been a move towards minimal or conservative surgery. The rate of recurrence increases with the length of follow-up, and there is a group of patients who require repeated resections at increasingly frequent intervals. It is this group who are at greatest risk of developing a surgically induced short-bowel syndrome and malnutrition. Recurrence of Crohn's disease is independent of the presence of microscopic disease at the resection margins and, as yet, there is no successful adjuvant medical therapy to decrease the incidence of recurrence following surgery. Smoking does however seem to be associated with increased recurrence and patients should thereforebe encouraged to stop.

Strictly speaking, there is no such thing as a surgical recurrence of Crohn's disease, but rather a recrudescence of a previously quiescent focus of Crohn's disease at a different site.

The aims of surgery

Since Crohn's disease cannot be cured by wide surgical excision, surgical management is reserved for the complications of Crohn's disease. The main indications are to:

1. remove or relieve an area of stenosis causing persistent or recurrent obstructive symptoms;
2. treat enterocutaneous or enterovesical fistulas;
3. drain an intra-abdominal abscess—this role is now being achieved increasingly by percutaneous radiological techniques;
4. control acute or chronic blood loss, though this is rare; and
5. treat free perforation, but this is exceedingly uncommon.

Since the risk of recrudescence of Crohn's disease is sufficient to require the treatment of approximately 6 per cent of cases per year, many patients will need a number of operations during a lifetime. It is important when operating to deal only with the specific complication which has been the indication for the procedure, and bowel should not be removed just because it is affected by Crohn's disease. Surgery should be as conservative as possible, but stenosed bowel, even in a defunctioned segment, should not be left untreated. The greatest care must be taken with surgical technique, ensuring well-vascularized, tension-free anastomoses and the avoidance of inadvertent damage to adjacent loops of the bowel.

Crohn's disease at specific sites in the small intestine

Gastroduodenal

Gastroduodenal disease is rare and almost always occurs in association with disease elsewhere. It can cause bleeding or stenosis, and disease in the duodenum most frequently produces clinical symptoms (Fig. 1). At endoscopy it can be difficult to distinguish between peptic ulcer and Crohn's disease. Peptic ulcers almost always respond to H_2-receptor antagonists or proton pump inhibitors, whereas Crohn's disease does not. A Crohn's ulcer often responds to medical therapy and undegoes a relapsing course, but stricture development is unlikely at that site.

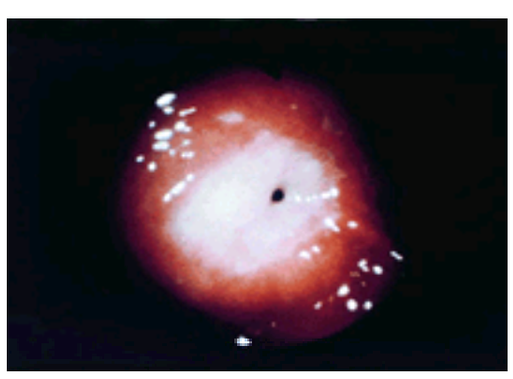


Fig. 1. Endoscopic appearance of Crohn's disease at the pylorus. There is pinpoint gastric outlet with surrounding ulceration.

Resection of the duodenum is not feasible and therefore this is the one area of the small bowel in which bypass is the preferred treatment. At one time gastrojejunostomy was thought to be the treatment of choice. However, in some patients who have already had extensive small-bowel resection, gastric hypersecretion may result, and it has been customary to add a vagotomy to reduce the high risk of stomal ulceration. This vagotomy may, in turn, compound the patient's problem with diarrhoea.

A selective vagotomy has the theoretical advantage that diarrhoea is less likely to occur. With the advent of H_2-receptor antagonists and proton pump inhibitors it is possible to control gastric secretion in patients with Crohn's disease. Where possible, a strictureplasty should be carried out on pyloric or duodenal disease to save the patient the complications of a gastroenterostomy. There have been some recent reports of the use of balloon dilatation for duodenal stenosis caused by Crohn's disease, but it is uncertain whether this is appropriate therapy.

Jejunum and ileum

The most commonly affected site is the terminal ileum. However, there may be multiple foci throughout the small bowel (Fig. 2) and here it is often difficult to tell exactly which lesion is responsible for the patient's symptoms. The most common indication for surgery in Crohn's disease of the small bowel is recurrent intestinal colic, provoked by solid food. Presentation is often insidious or subacute and may or may not be associated with diarrhoea. Weight loss is a major symptom. The patient may

be anorexic, or be literally frightened to eat for fear of pain.

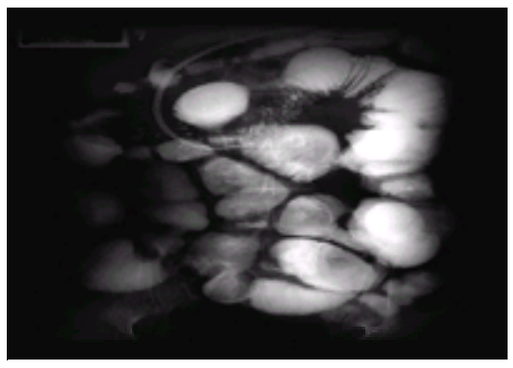


Fig. 2. Small-bowel enema demonstrating multiple strictures and dilated areas in diffuse small-bowel Crohn's disease.

Surgery

General principles

It is wise to give the patient full bowel preparation just in case there is any colonic involvement. Routine antiembolus measures are taken, including stockings and low-dose subcutaneous heparin. Since there is good evidence that patients with Crohn's disease have organisms in lymphatic vessels and lymph nodes outside the bowel wall, antibiotic prophylaxis with a cephalosporin and metronidazole is given for between 2 and 5 days. In this situation the antibiotics are probably being used for established infection rather than to prevent inoculation of a clean abdominal operation site. We usually give intravenous steroid cover (hydrocortisone 100 mg thrice daily) for the operation, introducing oral steroids when the patient is taking fluids by mouth. This is reduced to a dose of 10 mg, on which the patient is maintained until review. Although many surgeons would operate with a patient flat on the operating table, ileosigmoid or ileorectal fistulas in Crohn's disease are common and, in complicated cases, access to the rectum may be necessary. Therefore we usually operate on patients in the Lloyd-Davis position used for colorectal surgery. A full laparotomy is carried out, carefully examining the whole length of the intestine for external macroscopic features of Crohn's disease.

Ileocolic disease

This is best treated with a conservative ileocaecal resection to include a short margin of macroscopically normal gut on each side of the resection specimen (Fig. 3). In patients with severe disease, however, the presence of minor macroscopic evidence of Crohn's disease at the anastomotic site is not important and the emphasis should be on preserving gut length. There has been some controversy over the use of various suture materials and the precise type of anastomosis used, and their possible effect on disease recrudescence. There is no definite evidence of an effect of either of these; we usually carry out a straight end-to-end ileocolic anastomosis, using absorbable suture material (3–0 Vicryl). It is helpful to mark the site of the anastomosis with metal clips so that future radiological investigations can localize new disease.



Fig. 3. Resection specimen of ileocolic Crohn's disease.

In recent years laparoscopically assisted ileocolic resections have been introduced. The terminal ileum and ascending colon can be mobilized laparoscopically prior to extracorporeal anastomosis. It has potential advantages in reducing postoperative pain and length of hospital stay and may have a role in the management of new ileocolic disease.

Results of surgery for ileocolic disease

Actuarial methods are now widely used for the effective comparison of results of treatment between various centres because they correct for the varying lengths of follow-up between individual patients.

The cumulative operation rate for patients with distal ileal disease at 5 years from the time of diagnosis is 80 per cent. Cumulative reoperation rate after the first resection for distal ileal Crohn's disease at 5 years after the first operation is 20 per cent and at 10 years this rises to 35 per cent. If an ileocolic anastomosis is inspected carefully by colonoscopy, aphthous ulceration can be seen on the ileal side of the anastomosis as early as 6 months after surgery. There is no evidence that age at the time of initial diagnosis has any effect on recurrence rate.

Although recurrent disease is therefore very common, surgical treatment of distal ileal disease rapidly restores the patient with chronic persistent symptoms to good health. On average, patients have an operation once every 10 years, so they could expect to have three or four operations over a lifetime.

Multisite small intestinal disease

Lee pioneered the concept of treating Crohn's disease strictures by strictureplasty. He operated on patients with intestinal obstruction and malnutrition in whom excisional surgery was thought to be contraindicated because of diffuse small-bowel involvement or a short-gut syndrome resulting from previous excisions. Lee proposed the concept of minimal surgery to overcome sites of obstruction. In some places this would be a mini-resection, and in others strictureplasty. In a subsequent report of 15 years' experience of strictureplasty for obstructive Crohn's disease, including the first patients operated on by Lee in Oxford, 52 patients had 241 strictureplasties in 76 separate operations. There were no deaths or fistulas, septic complications occurred in only two, and patients made positive weight gains in the postoperative period. Nineteen patients (36 per cent) required a second operation for Crohn's disease between 1 and 57 months after initial strictureplasty. Most symptomatic recurrence was caused by new areas of disease and recurrence was found at only nine strictureplasty sites (4 per cent). Seven patients (13 per cent) required a third operation.

The Cleveland Clinic performed 698 strictureplasties in 162 patients and found a cumulative 5-year reoperative recurrence of 28 per cent following strictureplasty. Patients undergoing resection and strictureplasty at the first operation were no less likely to require subsequent surgery than those undergoing strictureplasty alone.

Alexander Williams carried out 198 strictureplasties in 64 patients. In addition to using the technique in short stenoses, he has used a Finney pyloroplasty-like method for stenotic areas of more than 10 cm in length. Major complications appeared in five patients who developed enterocutaneous fistulas, a clinical leak rate of 2.5 per cent. At 6 months 80 per cent had an excellent symptomatic response, being free of pain with improvement in general well being.

Another study looked at the site-specific recurrence rate in 41 patients treated by strictureplasty, and reported that no recurrence occurred in those patients. A control

group continued to be treated by small-bowel resection and there was no difference between the two groups. There is no evidence as yet that disease remaining at the strictureplasty site precipitates early recrudescence.

Operative technique

In multifocal disease it is essential to identify every stricture. Even though a stricture may not look very narrow from the outward appearance, this can be misleading (Fig. 4). We usually choose the middle stricture and open it on the antimesenteric border, along the long axis of the small intestine. A 20 French gauge Foley balloon catheter is inflated with saline and volumes are calibrated to give balloon diameters of 15, 20, and 25 mm (Fig. 5). The intestine is then characterized from the duodenojejunal flexure to the caecum by passing the catheter along the gut lumen and then inflating the balloon, allowing it to impact at each stricture site as it is slowly withdrawn. This ensures that no strictures are undetected and prevents the dangerous scenario of obstruction downstream from a suture line. The diameter is recorded and its position marked with a suture. A stenosis of less than 20 mm in diameter is usually treated by strictureplasty. The stricture site is held in stay sutures and a longitudinal incision made by diathermy (Fig. 6). The bowel is opened out, and, if the appearance is suspicious, a biopsy should be taken to exclude the possibility of cancerous change, although we do not do this routinely. The stricture site is then closed transversely in the manner of a Heinecke–Mickulicz pyloroplasty (Fig. 7), although some slightly longer strictures have been anastomosed in the manner of Finney. We use a continuous absorbable suture in a single layer. In our experience the median length of the strictureplasty has been 3 cm. Simultaneous intestinal resections have been undertaken for longer sections of disease, but we make these mini-resections as far as possible. Multiple strictureplasties can be carried out in one session and the strictureplasty sites are usually marked with metal clips for future reference. A stapling technique has been described using a linear stapler, but the resulting anastomosis looks more like a side-to-side than a strictureplasty. There is no evidence that suture material or the type of anastomosis makes a differences to the rate of recurrence.



Fig. 4. A typical short segment of ileal Crohn's disease. Note the narrowing and fat wrapping. The external appearance of a stricture like this can be misleading with much more luminal narrowing than expected.

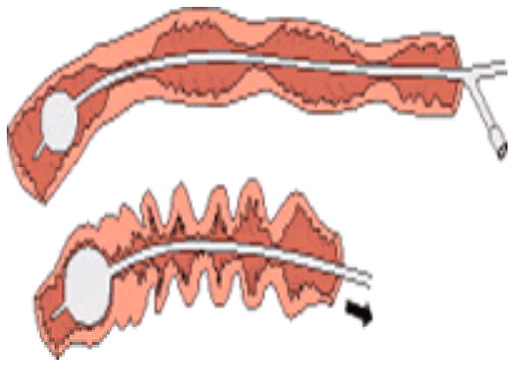


Fig. 5. Balloon catheter characterization of small-bowel Crohn's disease strictures.

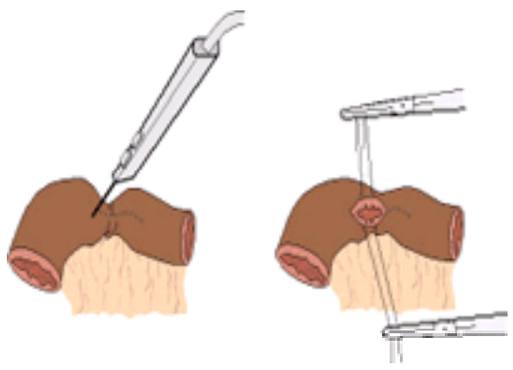


Fig. 6. Strictureplasty—a longitudinal incision is made between stay sutures with diathermy.

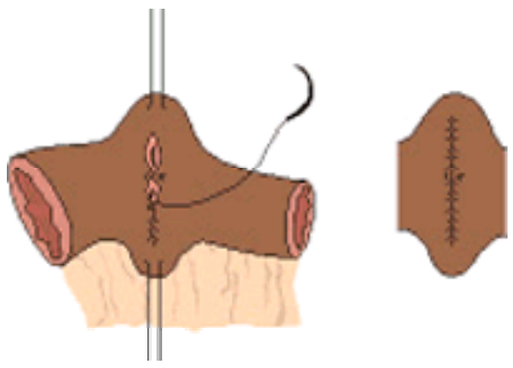


Fig. 7. Strictureplasty—the stricture site is sutured transversely.

What happens at the strictureplasty site?

There is some evidence that disease may heal, or at least not progress. Removing the obstructive element, even though the original triggering factors are likely to be present, may modify the progression of the disease. Once obstruction has occurred, a chain of events is precipitated resulting in mucosal leakiness, transmural fissuring, and, eventually, abscess and fistula formation.

Fistula and abscess

Small-bowel Crohn's disease can be complicated by enteroenteric, enterovesical, or enterocutaneous fistulas. It is important to emphasize that postoperative fistulas may be due to an anastomotic breakdown, and not associated with any residual or recurrent active Crohn's disease.

The incidence of external fistulas is 17 to 22 per cent and of internal fistulas, 9 to 25 per cent. Preoperative abscess as a complication of Crohn's disease occurs in some 10 per cent of patients and, following resection, postoperative abscess is found in 14 per cent, of which a third occur in patients with a preoperative abscess. The organisms involved are usually enteric, including *Escherichia coli, Bacteroides fragilis*, and enterococci. Wound infection and intra-abdominal abscesses are common problems in patients with Crohn's disease.

These abscesses may be primary or secondary. They probably occur at the site of a transmural fissure and may be the first stage in the development of a fistula as the abscess tracks to another loop of gut or the abdominal wall. The common sites for an abscess are the pelvis, abdominal wall, and psoas sheath, or around the segment of affected bowel. Abscesses penetrating posteriorly into the psoas produce symptoms of fever, weight loss, and a flexion deformity of the hip. They can point at the groin or buttock (Fig. 8). Drainage of the abscesses invariably results in a fistula. Secondary abscesses are more common and follow surgical procedures, often in malnourished septic patients. They may also occur slowly and present many months later.



Fig. 8. Sinogram through a sinus in the right inguinal region showing a psoas abscess secondary to ileocaecal Crohn's disease.

Most external fistulas pass through the site of a previous surgical incision, and a spontaneous fistula through the abdominal wall is rare.

Appendicectomy and fistula

An external fistula following appendicectomy usually arises from a segment of ileum actively involved with Crohn's disease rather than the appendix stump. If Crohn's disease is found at a laparotomy for presumed ‘appendicitis', the appendix should be removed. This will ensure that future attacks of pain can be managed without confusion. Where there is extensive caecal involvement, an ileocaecal resection is preferred. There is some evidence indicating that the majority of patients with terminal ileal Crohn's disease found at appendicectomy require resection within 3 years. This would suggest that a local resection should be performed when Crohn's disease is found at the initial operation for presumed appendicitis. However, it is sometimes difficult to distinguish Crohn's disease confidently from other causes of terminal ileitis.

Free perforation

This is surprisingly rare from small-bowel Crohn's disease. Sudden onset of pain in a patient with an exacerbation of Crohn's disease treated with steroids should be regarded as a possible perforation until proved otherwise.

Fistula and abscess: investigation

Fistula and abscess go hand in hand, and a fistula cannot be treated successfully without the elimination of any concurrent abscess.

History and examination

Fever, weight loss, diarrhoea, and abdominal pain are the usual symptoms of an internal fistula, although some enteroenteric fistulas can be entirely asymptomatic. A discharge sinus is the hallmark of an external fistula, although only about 50 per cent may discharge faeces, the rest producing gas or pus (Fig. 9). There may be a mass or area of tenderness to suggest an abscess.



Fig. 9. An enterocutaneous fistula complicating Crohn's disease pointing through an old laparotomy scar.

Investigations

Nutritional and biochemical assessment should include haemoglobin, serum lbumin, body weight, lymphocyte count, and liver function tests. Barium contrast studies, especially a small-bowel enema, are best for demonstrating areas of small-bowel disease, a fistula, and even an abscess (Fig. 10). Fistulography provides important information about the complexity of a track. Ultrasound and computed tomography (**CT**) scans are the best methods for demonstrating an abscess, and CT-guided percutaneous drainage of abscesses is becoming increasingly successful (Fig. 11). Endoscopy will help to define the extent of mucosal disease and the presence and nature of a stricture, but rarely identifies a tiny fistula track, for example at an ileocolic anastomosis.



Fig. 10. A small-bowel enema showing ileal loops separated by a mass—an abscess.

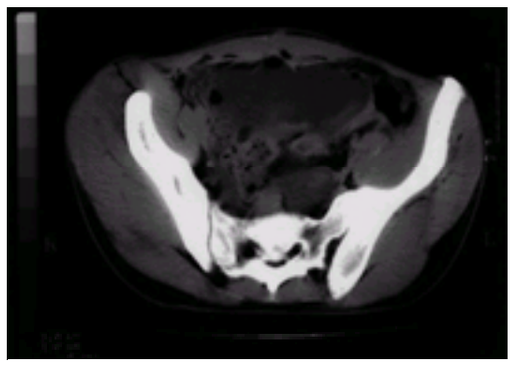


Fig. 11. CT demonstrating an intra-abdominal abscess.

Treatment of abscesses and fistulas

Many internal fistulas are asymptomatic and often a chance finding at laparotomy. Any area of active small-bowel Crohn's disease adherent to an adjacent organ should be regarded as a fistula until proved otherwise. Disease is usually on the small-bowel side of an enterocolic fistula. This can be pinched off and the hole in the colon closed. Formal resection is seldom necessary. Enterocutaneous fistulas usually arise at an anastomosis or in a segment of non-involved bowel damaged at surgery. The management of these can be very challenging and requires a team approach. Within the first 24 h, fluid and electrolyte problems must be corrected and the skin protected with stomadhesive and a wound-management appliance. Intravenous nutrition should next be established through a dedicated central venous line. If there is an abscess it must be drained, preferably by a percutaneous technique. Broad-spectrum antibiotics are given for a defined period and are not a substitute for adequate drainage. Then the fistula anatomy can be defined. It is reasonable to try a period of conservative management, and some 60 to 70 per cent of postanastomotic fistulas will heal spontaneously.

Spontaneous enterocutaneous fistulas rarely heal. Although medical trials with azathioprine, 6-mercaptopurine, and cyclosporin have been reported, surgical exploration and resection of the fistula with anastomosis is usually necessary.

Surgical closure

Operative treatment can be successful in experienced hands, provided there is not extensive sepsis or profound hypoalbuminaemia, and distal obstruction must be carefully excluded. If the patient is grossly hypoalbuminaemic or septic, the procedure of choice is resection and exteriorization of both bowel ends or, less optimally, a proximal bypass leaving the disease and fistula in place. When the patient is well and the sepsis drained, a second procedure to restore continuity is permissible. It may be necessary to wait for many weeks until a patient is fit enough for a further operation. Where the patient is well, resection and primary anastomosis are indicated (Fig. 12).

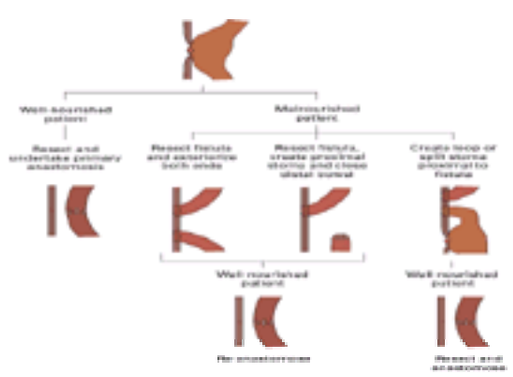


Fig. 12. Diagram showing the strategy for managing an enterocutaneous fistula complicating Crohn's disease.

Abscess

An abscess pointing at the skin surface should be drained but is usually followed by a fistula. The introduction of ultrasound and CT have not only improved diagnosis and localization of abscesses, but percutaneous drainage techniques have largely replaced open laparotomy with all its hazards in a sick patient. The drainage of an intra-abdominal abscess is an absolute priority and undrained sepsis can be lethal. Antibiotics should be given in short effective courses, taking into account bacteriological reports on any swabs taken from an abscess.

Enterovesical fistula

Repeated urinary tract infections, turbid urine, and pneumaturia suggest an enterovesical fistula. This is not an absolute indication for surgery but they seldom heal and eventually come to surgery. The small bowel can usually be ‘pinched off' the bladder and resected (Fig. 13, Fig. 14). The bladder itself only requires curettage and careful closure. A catheter is left in the bladder for at least a week.

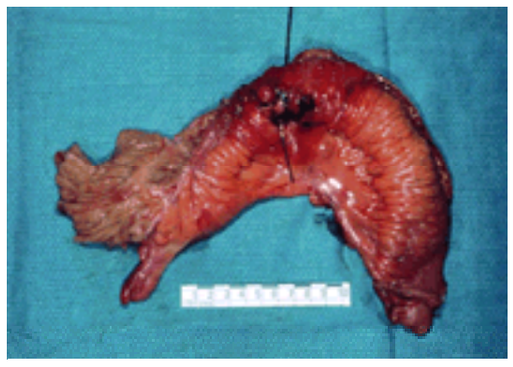


Fig. 13. Ileocolic resection specimen from a patient with an enterovesical fistula. The probe indicates the site of the fistula which has been ‘pinched off' the bladder.



Fig. 14. Mucosal surface of the same specimen showing the discrete ulceration characteristic of Crohn's disease. The large ulcer in the middle was the internal opening of the enterovesical fistula.

Further reading

Allan RN, Rhodes JM, Hanauer SB, Keighley MRB, Alexander Williams J, Fazio VW, eds. *Inflammatory bowel diseases*, 3rd edn. Churchill Livingstone, Edinburgh, 1997. [Textbook on medical and surgical aspects of inflammatory bowel disease.]

Coutsoftides T, Fazio VW. Small intestine cutaneous fistulas. *Surgery, Gynecology and Obstetrics* 1979; **149**: 333–6. [Outcome of 174 patients with small intestine cutaneous fistulas.]

Greenstein AJ *et al.* Perforating and non perforating indications for repeated operations in Crohn's disease. *Gut* 1988; **29**: 588–92. [Retrospective analysis comparing reoperation rates in perforating and non-perforating Crohn's disease.]

Keighley MRB, Eastwood D, Ambrose NS, Allan RN, Burdon DW. Incidence and microbiology of abdominal and pelvic abscess in Crohn's disease. *Gastroenterology* 1982; **83**: 1271–5. [Study of the incidence of abdominal and pelvic abscesses in 111 patients.]

Lee ECG, Papaioannou N. Minimal surgery for chronic obstruction in patients with extensive or universal Crohn's disease. *Annals of the Royal College of Surgeons of England* 1982; **64**: 229–33. [Original description of strictureplasty for small bowel Crohn's disease.]

Ozuner G, Fazio VW, Lavery IC, Milson JW, Strong SA. Reoperative rates for Crohn's disease following strictureplasty. Long-tem analysis. *Diseases of the Colon and Rectum* 1996; **39**: 1199–203. [Analysis of cumulative reoperation rates in 162 patients undergoing 692 strictureplasties.]

Sardinha TC, Wexner SD. Laparoscopy for inflammatory bowel disease; pros and cons. *World Journal of Surgery* 1998; **22**: 370–4. [Review of published data on laparoscopic resection for Crohn's disease.]

Sayfan J, Wilson DAL, Allan A, Andrews H, Alexander Williams J. Recurrence after strictureplasty or resection for Crohn's disease. *British Journal of Surgery* 1989; **76**: 335–8. [Retrospective comparison of reoperation interval following strictureplasty or small bowel resection.]

Stebbing JF, Jewell DP, Kettlewell MGW, Mortensen NJMcC. Long-term results of recurrence and reoperation after strictureplasty for obstructive Crohn's disease. *British Journal of Surgery* 1995; **82**: 1471–4. [Results of strictureplasty in 52 patients over 15 years.]

Williams JG, Wong WD, Rothenberger DA, Goldberg SM. Recurrence of Crohn's disease after resection. *British Journal of Surgery* 1991; **78**: 10–19. [Review article on recurrence after surgery for small and large bowel Crohn's disease.]

24.3 Fistulas

Gordon Carlson and Miles Irving

Definitions
Classification
Aetiology
Management
Control of sepsis and skin care
Provision of nutritional and metabolic support
Anatomic assessment
Performing the definitive procedure to effect fistula closure
Further reading

Definitions

A fistula is an abnormal communication between two epithelial surfaces. The communication frequently takes the form of a tract lined with granulation tissue but may be completely lined with epithelium, particularly in the case of internal fistulas, external fistulas with short tracts, and those of congenital origin. This has important implications for the likelihood of spontaneous closure (see below). This chapter will deal with fistulas arising from the duodenum, jejunum, or ileum.

Classification

Small intestinal fistulas can be classified in several ways. These classifications are valuable not only for descriptive purposes but also for allowing comparison of different management strategies and determining the optimal treatment in individual cases. There are two principal classifications: anatomic and physiological.

Anatomic classifications (Fig. 1(a),Fig. 1(b) and Fig. 1(c)) distinguish between internal (e.g. duodenocolic) and external (e.g. iliocutaneous) fistulas. The fistulas can also be described in terms of the anatomic surfaces involved in the origin and termination of the tracts (e.g. iliovesical, iliovesicocutaneous). The orientation of fistula tracts can be described in relation to the long axis of the small intestine. Lateral fistulas arise from the side of a segment of small intestine that is otherwise in continuity and end fistulas arise within the long axis of the small intestine (e.g. the duodenal stump after a Polya gastrectomy or an anastomosis after complete separation of the anastomotic ends). Fistulas may also be described as simple if a direct single tract connects the two involved epithelial surfaces, or complex if there are associated abscess cavities, multiple tracts, or combinations of internal and external components.

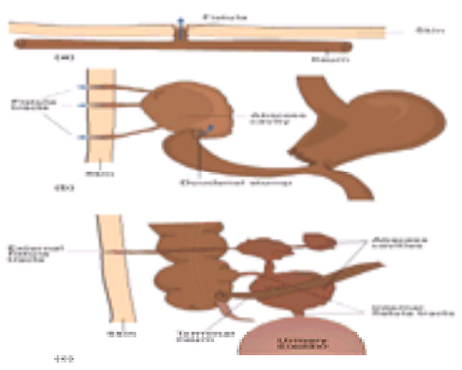


Fig. 1. (a) Simple lateral ileocutaneous fistula with a single tract. (b) Complex end duodenocutaneous fistula with associated abscess cavity and multiple tracts. (c) Complicated lateral ileocolovesicocutaneous fistula with multiple tracts and abscess cavities connecting terminal ileum, colon, urinary bladder, and abdominal wall.

Physiological classifications principally relate to the output of intestinal contents from enterocutaneous fistulas and the presence of associated sepsis. Fistulas are classified as high output if the total daily fistula loss exceeds 500 ml per day and low output if losses are less than this. In general a fistula output less than 500 ml per day is associated with a significantly smaller metabolic disturbance than a greater loss and this is usually reflected in the reduced need for total parenteral nutrition in the management of such patients.

Aetiology

With the exception of the patent vitellointestinal duct, small intestinal fistulas are acquired rather than congenital. Over 80 per cent of small intestinal enterocutaneous fistulas develop as early complications of surgical operations (Fig. 2).



Fig. 2. Typical postoperative fistula.

They occur most frequently because of failure of healing of suture lines either at anastomoses or at sites of enterotomy repair. Occasionally, during difficult and tedious dissections, small enterotomies may be overlooked and present later as fistulas. Failure of a suture line to heal may be due to local or systemic factors. Local factors include technical error, poor blood supply to the bowel ends, tension at the anastomosis, distal obstruction, and disease at the anastomotic site (e.g. malignancy, radiation enteritis). Systemic factors such as sepsis or malnutrition may also impair healing at an intestinal anastomosis although the exact mechanisms are unclear.

The remainder of external, and virtually all internal, fistulas develop as a result of injury to healthy bowel or perforation of diseased viscera. Closed injury may result from compression of fluid- or gas-filled loops of small intestine during abdominal compression, such as from seat belt trauma. The initial injury may lead to the development of an intramural haematoma and localized perforation which remains unrecognized until the resulting abscess develops and discharges through the abdominal wall. Localized perforation may develop as a result of Crohn's disease, peptic ulcer disease, malignancy, or radiation enteritis (Fig. 3).



Fig. 3. Typical Crohn's fistula.

In such cases, an abscess cavity forms, walled off by adjacent viscera and/or the abdominal wall and eventually perforates into them, leading to the development of internal, external, or complex fistulas. The small intestine may not only be the site of the primary disease, but may also become secondarily involved in disease processes in adjacent viscera such as colonic diverticular disease, appendicitis, malignancy, or gallstone disease.

Management

Management of patients with small intestinal fistulas is frequently complex and labour intensive. It requires the support of a dedicated multidisciplinary team including surgeons, gastroenterologists, nurses, stomatherapists, pharmacists, dietitians, and biochemists. Many patients require extensive psychological support, although this may be better provided by the above staff than by psychologists.

The goal of treating patients with intestinal fistulas is to achieve permanent closure of the fistula in the most rapid, cost-effective, and safe manner. This requires the development of an individually tailored management plan for each patient. The management plan is constructed according to a number of key principles and can be divided into four phases, conveniently summarized by the acronym ‘SNAP':

1. sepsis control;
2. nutrition and metabolic support;
3. anatomic assessment; and
4. performing definitive procedure (where necessary).

Control of sepsis and skin care

Uncontrolled sepsis remains a major cause of death in patients with an intestinal fistula. Detection and treatment of sepsis is therefore the priority. A high index of clinical suspicion should be maintained for the diagnosis of sepsis in patients with intestinal fistulas. Many patients have been in hospital for a considerable length of time and sepsis may have arisen from chest, urinary, wound, or central line infection. Most cases of sepsis in association with an intestinal fistula are, however, attributable to intra-abdominal infection. In established cases, characteristic indicators of infection such as pyrexia, leucocytosis, and tachycardia may be absent and the presence of failure to progress, cachexia despite the provision of an apparently satisfactory amount of calories and nitrogen, and hypoalbuminaema should point to the presence of untreated sepsis. Although ultrasonography and radiolabelled white-cell scanning may be of value, the most useful investigation for the assessment of sepsis in association with intestinal fistulas is computed tomography (**CT**). CT scanning with oral, rectal, and intravenous contrast is not only of benefit in delineating the anatomy of intra-abdominal collections, but frequently allows simultaneous percutaneous radiological drainage (Fig. 4(a) and Fig. 4(b)).

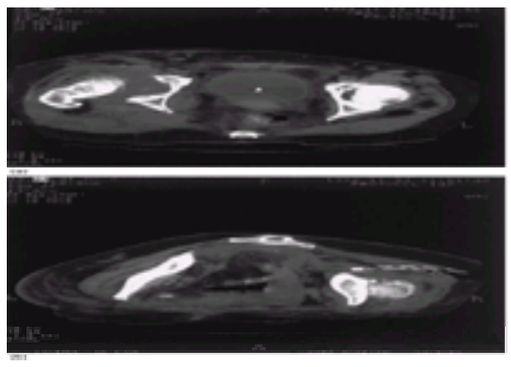


Fig. 4. (a) CT scan showing intra-abdominal abscess. (b) CT scan showing intra-abdominal abscess, treated by percutaneous drainage.

Percutaneous drainage may convert a complex fistula into a simple tract which may heal spontaneously, given appropriate anatomic considerations and optimal metabolic and wound care. Where abscesses are multiloculated or associated with complete separation of anastomotic ends, early surgical treatment is likely to be required in order to deal adequately with sepsis. Under these conditions, exteriorization of the intestinal ends is required because the risks associated with reanastomosis in the presence of active sepsis and localized abscess formation are considerable.

Teaching point—when dealing with anastomotic dehiscence associated with an abscess cavity do not anastomose, exteriorize.

Provision of satisfactory wound care represents an additional challenge in patients with external fistulas, particularly if high output. In such cases the abdominal wall may be rapidly digested by the fistula output which may be strongly acidic or alkaline (depending upon the site of the fistula and the presence of gastric hypersecretion) and composed primarily of digestive enzymes. Fistulas within large abdominal wall defects can be managed using a combination of large stoma bags which can be cut to size and suction catheters (Fig. 5).



Fig. 5. Multiple fistulas within laparostomy wound, treated by ‘Eakin' bag and suction catheters.

Life for patients with very high output fistulas can be made tolerable by the use of ‘megostomy' tubing which connects the stoma bag to a 24-h collection bottle,

obviating the need for frequent emptying of the stoma bag (Fig. 6).



Fig. 6. Megostomy device and tubing.

The assistance of an experienced stomatherapist in the selection of the appropriate appliances is invaluable. Further measures to make fistula losses more manageable such as thickening the effluent and reducing its volume (see below) may also be useful.

Provision of nutritional and metabolic support

There is current evidence to indicate that the efficacy of nutritional support is markedly impaired in the presence of sepsis and for this reason nutritional support is accorded a lower priority than treatment of sepsis. Once sepsis has been identified and is being treated, however, early consideration should be given to the need for the metabolic care of patients with small intestinal fistulas.

High output fistulas

Fistula losses may be very dramatic when the fistula arises from the proximal small intestine. Losses from a lateral duodenal fistula, for example, may exceed 6 litres each day. In general, daily fistula losses from the small intestine in excess of 500 ml should be replaced with normal saline with 20 mmol per litre of potassium chloride. In the case of proximal small intestinal fistulas, total parenteral nutrition is almost always required except occasionally, when access to the small intestine distal to the site of the fistula can be obtained. In such cases, the distal intestine can be intubated and enteral feeding initiated provided that radiological studies have confirmed the remaining bowel to be intact. Attempts at feeding via the distal limb of such fistulas are fraught with difficulty and frequently unsuccessful.

Simultaneous steps to reduce the fistula output should be taken, including minimizing oral intake. In proximal small intestinal fistulas there may be marked gastric hypersecretion as a result of hypergastrin-aemia because of diversion of the gastric output away from the small intestine. The evidence to suggest that this is a major factor in the majority of cases is poor and longitudinal studies of gastric hypersecretion in patients with short bowel syndrome suggest that the phenomenon is short lived. Nevertheless, oral or intravenous proton pump inhibitors may prove useful in reducing fistula output in selected patients. Similarly, administration of analogs of somatostatin may reduce fistula output in individual patients. There is no clear evidence on the basis of randomized controlled trials that such treatments promote spontaneous closure of fistulas. Although oral intake should be minimized in patients in whom conservative treatment is being undertaken, to expedite local healing and to simplify wound management by reducing the fistula output, it should be stressed that withholding oral intake and providing total parenteral nutrition *per se* will not close a fistula unless local conditions are already conducive to fistula closure. The purpose of metabolic and nutritional support is primarily to maintain the patient's nutritional state until the fistula has healed spontaneously or a definitive procedure carried out and oral feeding can be reintroduced.

Randomized controlled trials comparing parenteral nutrition with surgery for Crohn's fistulas of the small intestine have shown that while a small but significant percentage of Crohn's fistulas will heal on total parenteral nutrition, resolution is generally short lived and the vast majority of patients ultimately require surgical treatment.

Total parenteral nutrition should be provided using a dedicated, tunnelled central venous feeding line. Despite advances in the design of peripheral lines and formulation of feeding solutions with lower tonicity and balanced pH, rates of thrombophlebitis with peripheral lines remain too high for them to represent a satisfactory alternative to central venous lines, provided strict asepsis can be maintained. Line care should be undertaken by a team of specialized nurses trained in the aseptic techniques required for successful and infection-free management of such lines. Provided strict aseptic protocols can be developed and adhered to the rates of line infection can be kept to less than 6 per cent per annum. The nature and composition of parenteral nutrition formulations required for the management of patients with intestinal fistulas are beyond the scope of this chapter. Provision of adequate nutritional, fluid, and electrolyte support should not be assumed without careful monitoring. Accurate daily fluid balance should be charted, together with daily measurement of plasma electrolyte concentrations until it is clear that the patient is in a stable metabolic state. Urinary electrolyte measurements are often more useful because they give an indication of sodium and potassium balance before changes in plasma electrolyte concentrations occur. Patients should be weighed twice weekly and estimations of serum albumin and other major nutritional proteins undertaken weekly. Progressive wasting or persistent hypoalbuminaema should lead to a renewed search for sepsis.

Low output fistulas

The metabolic requirements of patients with low-output (generally distal) small intestinal fistulas can frequently be met by enteral feeding. Low residue diets are more palatable than elemental diets and there is little evidence to suggest that spontaneous fistula closure is less likely if they are used. Where patients are unable to take an oral diet, consideration should be given to insertion of a fine-bore nasoenteric feeding tube, or, if longer term feeding is thought likely to be required, a percutaneous gastrostomy, which is associated with a lower incidence of oesophageal stricture and aspiration pneumonia.

Anatomic assessment

Once sepsis has been treated and adequate nutritional and metabolic support initiated, a series of studies should be undertaken to delineate the anatomy of the fistula.

Anatomic assessment of external fistulas begins with a careful clinical assessment of the nature of fistula openings for signs of exposed mucosa. If there is obvious mucocutaneous continuity at the site of an enterocutaneous fistula, spontaneous closure is most unlikely. Anatomic assessment should provide all the information required to prepare a full description of the fistula according to the anatomic classification outlined above and to formulate a plan for definitive treatment. The principal information required from such studies is the precise origin, size, number, orientation, and termination of the fistula tracts, the nature of the disease process which has led to the fistula (where appropriate), and the presence of distal obstruction. Except in cases of high output fistulas, the most useful single investigation for delineating the anatomy of an external fistula is a fistulogram using a water-soluble contrast medium. In cases of high output fistulas, a small bowel enema is generally more effective than fistulography and also shows the level at which the fistula originates. Additional studies such as barium enema, intravenous urography, or even cystography may be required, depending upon individual circumstances. Once the patient is in a stable metabolic state, with sepsis controlled and nutritional support established, the provision of information regarding the anatomy and aetiology of the fistula will allow the completion of the management plan and the safe conduct of definitive surgical treatment where appropriate.

Performing the definitive procedure to effect fistula closure

Spontaneous closure

Spontaneous closure occurs in up to 60 per cent of external small intestinal fistulas within 4 to 6 weeks of conservative treatment, provided sepsis has been controlled and nutritional status maintained. In general, such fistulas are lateral fistulas arising from small anastomotic leaks or enterotomy sites within otherwise normal bowel. In contrast, spontaneous healing of an enterocutaneous fistula secondary to Crohn's disease occurs in less than 5 per cent of cases. Failure of a fistula to close generally indicates one of the causes given in Table 1.

1.	A fistula arising from a diseased segment of intestine (e.g. Crohn's disease, radiation enteritis, malignancy)
2.	An end fistula (except for the duodenal stump) or total discontinuity of the bowel ends at an anastomosis
3.	Distal obstruction
4.	A persistent abscess cavity at the site of the leak
5.	A foreign body
6.	Microcutaneous continuity
7.	Continued sepsis or malnutrition

Table 1 Causes of failure of an external small intestinal fistula to close spontaneously

Internal fistulas, in which there is generally either direct mucosal continuity or an intervening small abscess cavity, rarely, if ever, heal spontaneously. They may be asymptomatic or may present as a result of intestinal obstruction. Jejunocolonic fistulas may give rise to severe diarrhoea and weight loss as a result of bypass of the distal small bowel. These features are occasionally seen with ileocolonic fistulas and result from bacterial overgrowth in the small intestine. Temporary improvement may be obtained with metronidazole treatment while the patient is prepared for surgery.

Surgical treatment

Early surgical treatment may be required to deal with life-threatening sepsis (for example, drainage of abscesses), enabling satisfactory skin care (a well-sited proximal defunctioning loop jejunostomy may be much easier for a stomatherapist to control than multiple fistulating small intestinal loops in a deep abdominal wall cavity) or providing conditions which will allow spontaneous closure (for example, insertion of a gastrostomy above a lateral duodenal fistula).

Where definitive surgery is required to produce closure of the fistula, careful consideration should be given to the scope and timing of surgery. In postoperative fistulas a delay of at least 3 months is essential before embarking upon further surgery, to allow the maturation of intra-abdominal adhesions. Where a laparostomy has been used to deal with severe intra-abdominal sepsis and this has resulted in the development of multiple enterocutaneous fistulas, 6 months is usually required. Protrusion of the fistulating bowel through the fistula site is a useful guide to re-establishment of the neoperitoneal cavity (Fig. 7).

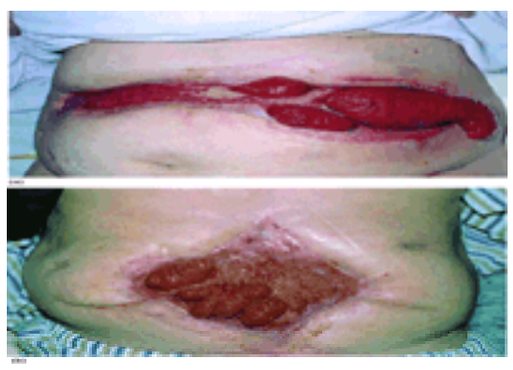


Fig. 7. (a) and (b) Prolapse of fistulating bowel, signifying development of neoperitoneal cavity.

Local repair of fistulas is virtually always associated with recurrence. The minimum required is a thorough laparotomy, with a tedious dissection of adhesions and resection of the fistula. In Crohn's disease, adjacent and intrinsically normal viscera may become secondarily involved (for example, the sigmoid colon or bladder in terminal ileal disease) and can simply be pinched off the inflammatory mass rather than resected. Bypass surgery, especially for Crohn's disease, should be avoided if at all possible because the continued presence of the fistulating mass leads to further complications, ultimately necessitating more extensive resection and leading to a risk of short bowel syndrome. Intestinal reanastomosis should only be attempted if the patient is free from sepsis and well nourished. An anastomosis should not be performed within or close to an abscess cavity.

In certain cases, anatomic considerations dictate a different approach to segmental resection and anastomosis. Duodenal defects may require drainage into a Roux loop or a serosal patch repair using a loop of small bowel brought up to and sutured carefully over the defect. In complex cases such as these, diversion of gastric and biliary contents with a gastrostomy and cholecystostomy inserted at the time of surgery may help protect the fistula repair until healing has occurred, while insertion of a feeding jejunostomy may obviate the need for postoperative total parenteral nutrition.

Further reading

Alexander-Williams J, Irving M. *Intestinal fistulas*. Bristol: Wright, 1982. [A comprehensive overview of the management of all forms of intestinal fistulas, with detailed accounts of aetiology, pathogenesis, and strategies for operative management.]

Bernini A, Spencer MP, Wong WD, Rothenberger DA, Madoff RD. Computed tomography-guided percutaneous abscess drainage in intestinal disease. *Diseases of the Colon and Rectum* 1997; **40**: 1009–13. [A study of the efficacy of radiologically-guided drainage of intra-abdominal collections and factors which may limit the usefulness of these techniques.]

Carlson GL, Irving MH. Recognition and management of infection in surgical patients. In: Hanson G, ed., *Critical care of the surgical patient*., pp. 273–90. London: Chapman and Hall Medical, 1997. [A detailed and up-to-date review of present methods of diagnosis and management of intra-abdominal infection.]

Edmunds LH, Williams GM, Welch CE. External fistulas arising from the gastrointestinal tract. *Annals of Surgery* 1960; **152**: 445–71. [An early paper detailing presentation, aetiology, and pathogenesis of enterocutaneous fistulas, with an account of prognosis and management in the era prior to the advent of total parenteral nutrition.]

Hill GL. *Disorders of nutrition and metabolism in clinical surgery*. Edinburgh: Churchill Livingstone, 1992. [A comprehensive text which provides detailed accounts of nutritional and metabolic disease in surgical patients. Gives an extensive appraisal of methods used for nutritional assessment and therapy.]

Irving M. Assessment and management of external fistulas in Crohn's disease. *British Journal of Surgery* 1983; **70**: 233–6. [Review of methods for delineating the anatomy of enterocutaneous fistulas in Crohn's disease and surgical treatment.]

Meguid MM, Campos ACL. Surgical management of gastrointestinal fistulas. *Surgical Clinics of North America* 1996; **76**: 5. [A detailed review of operative and non-operative strategies for the management of both internal and enterocutaneous fistulas.]

Muller JM, Keller HW, Erasmi H, Pichlmaier H. Total parenteral nutrition as the sole therapy in Crohn's disease. A prospective study. *British Journal of Surgery* 1983; **70**: 40–3. [A prospective trial which demonstrates that parenteral nutrition alone fails to produce sustained closure of Crohn's fistulas.]

Nightingale JM, Lennard-Jones JE, Walker ER, Farthing MJ. Jejunal efflux in short bowel syndrome. *Lancet* 1990; **336**: 765–8. [Analyses the metabolic consequences of intestinal fluid output in patients after loss of much of the small intestine and examines factors which contribute to high fluid losses.]

Pettit SH, Irving MH. The operative management of fistulous Crohn's disease. *Surgery, Gynecology, and Obstetrics* 1988; **167**: 223–8. [A review of strategies for the surgical management of fistulas arising from Crohn's disease and its treatment.]

Scott NA, Finnegan S, Irving MH. Octreotide and postoperative enterocutaneous fistulae: a controlled prospective study. *Acta Gastroenterologica Belgica* 1993; **56**: 266–70. [Controlled trial showing that octreotide therapy does not alter the natural history of enterocutaneous fistulas.]

Scripcariu V, Carlson GL, Bancewicz J, Irving MH, Scott NA. Reconstructive abdominal surgery after multiple laparotomies for abdominal sepsis. *British Journal of Surgery* 1994; **81**: 1475–8. [This review, from a specialized intestinal failure unit, explores the consequences for the peritoneal cavity of healing by secondary intention after laparostomy and provides an account of strategies and timing for intestinal reconstruction.]

Streat SJ, Beddoe AH, Hill GL. Aggressive nutritional support does not prevent protein loss despite fat gain in intensive care patients. *Journal of Trauma* 1987; **27**: 262–6. [A prospective study in critically-ill patients which demonstrates that even aggressive parenteral nutrition, with calorie and nitrogen administration in excess of predicted requirements does not maintain protein economy in the presence of sepsis.]

24.4 Diverticular disease of the small bowel

Michael N. Margolies

Meckel's diverticulum
Associated anomalies
Radiologic diagnosis
Clinical syndromes due to Meckel's diverticulum
Peptic ulceration and hemorrhage
Obstruction
Diverticulitis
Enteroliths and foreign bodies
Neoplasms
Management of the incidental Meckel's diverticulum
Acquired diverticula of the small bowel
Diagnosis
Complications of small-bowel diverticulosis and their treatment
Further reading

Meckel's diverticulum

Meckel's diverticulum is a congenital diverticulum arising from failure of embryonic obliteration of the omphalomesenteric (vitelline) duct connecting the fetal gut to the yolk sac, which normally occurs during the fifth to seventh week of gestation. In contrast to other small-bowel diverticula, Meckel's diverticulum contains all intestinal layers and is antimesenteric. It is the most common congenital anomaly of the small intestine, found in 0.5 to 3.0 per cent of the population at postmortem examination. It is most often found within 1 m of the ileocecal valve and rarely up to 180 cm from the valve; the length of the diverticulum is usually less than 12 cm but can vary from 0.5 to 56 cm.

Meckel's diverticulum has a blood supply independent of that of the contiguous ileum, distinguishing it from ileal duplication; in addition, ileal duplication usually arises from the mesenteric border. The embryonic blood supply of the vitelline duct is by paired vitelline arteries, of which the right forms the superior mesenteric artery, while the left artery involutes. The blood supply to the diverticulum is derived from a remnant of one of these arteries. A mesodiverticular band may connect the diverticulum to the ileal mesentery; the contained vessel is often obliterated.

Associated anomalies

Simple Meckel's diverticulum represents the most common end-result among a set of omphalomesenteric duct anomalies, in which the entire duct is obliterated and absorbed, with the exception of several centimeters attached to the small intestine (Fig. 1). Other associated abnormalities of duct obliteration include the following:



Fig. 1. Meckel's diverticulum in an 11-year-old male with rectal bleeding. Note the pale area near the top, signifying underlying heterotopic tissue.

1. the entire vitelline duct persists, resulting in a fistula between the ileum and the umbilicus;
2. the distal end of the duct persists as a sinus opening at the umbilicus or as an umbilical polyp;
3. the duct lumen is obliterated but not absorbed, forming a fibrous band attached to the umbilicus, to other viscera, or with a free end;
4. a segment of the duct remains patent, forming a cyst or ‘enterocystoma' along its course;
5. the mesodiverticular band described above;
6. heterotopic tissue, usually consisting of gastric epithelium, is found in 30 to 50 per cent of cases, but may include pancreatic (5 per cent), colonic, jejunal, or duodenal elements;
7. the diverticulum may be buried in an intramesenteric position or have a separate mesentery.

Radiologic diagnosis

Meckel's diverticulum can be identified preoperatively by radionuclide scanning, barium contrast studies, arteriography, and occasionally by computed tomographic scanning, with varying degrees of success. The pertechnetate anion ($^{99}Tc^{m}$) is selectively taken up and is secreted by gastric mucosal parietal cells in their normal and aberrant locations, as well as by thyroid and salivary glands and the choroid plexus. Thus, the demonstration of Meckel's diverticulum by abdominal scanning (Fig. 2) as an extragastric localized area of uptake, which increases temporally in parallel with gastric mucosal uptake, depends on the presence of ectopic gastric mucosa. Scanning is most successful in cases of bleeding due to associated peptic ulceration in children. In the most experienced centers, the sensitivity of detection is 85 per cent and specificity is 95 per cent. However, technetium-99m scanning may be less accurate than generally believed, particularly in adults, where the sensitivity is lower. A negative scan at any age does not exclude the diagnosis. Moreover, false-positive scans occur, ascribed to such unrelated conditions as small-bowel tumors, small-bowel obstruction, hydronephrosis, enteric duplication, and arteriovenous malformations.

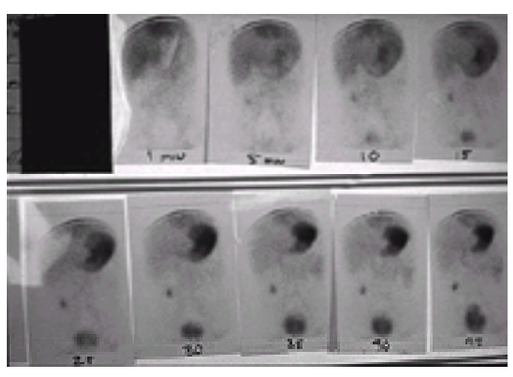


Fig. 2. Serial technetium-99m scans in an 11-year-old male with rectal bleeding. Note that the isotope uptake at a site in the right side of the abdomen corresponding to a Meckel's diverticulum increases simultaneously with gastric uptake. The large round area of uptake at the bottom represents the urinary bladder.

The diagnosis of Meckel's diverticulum is occasionally suggested on plain abdominal films on the basis of an enterolith or radio-opaque foreign body in the right lower quadrant or a persistent right lower quadrant or subumbilical gas collection. On antegrade small-bowel barium studies, the diverticulum is sometimes directly visualized, or suspected by virtue of the mass effect on the adjacent ileum or frank obstruction. Because the diverticulum may only fill transiently, careful fluoroscopy is necessary. The diagnostic yield is increased by use of a small-bowel enema or enteroclysis, in which contrast is given following nasoduodenal or nasojejunal intubation. Barium enema examination may identify ileocolic or even ileo-ileal intussusception as the basis for obstruction. During episodes of major gastrointestinal bleeding due to Meckel's diverticulum, barium studies should not be done. Occasionally, extravasation of contrast dye is seen on mesenteric arteriography in actively bleeding patients, but the diagnostic yield is low.

Clinical syndromes due to Meckel's diverticulum

The manifestations of Meckel's diverticulum are protean. The pathologic complications of the diverticulum are related directly to the particular pattern of accompanying vitelline duct abnormalities. Thus they can mimic many intra-abdominal conditions. Meckel's diverticulum should be considered as a possible cause of any intra-abdominal disease in which the diagnosis is not evident.

The complications of Meckel's diverticulum include bleeding, obstruction, diverticulitis, enterolith, and tumors, in addition to associated anomalies. The diagnosis is seldom made preoperatively (less than 4 per cent of cases), except in cases of lower gastrointestinal bleeding in the pediatric population. The pain or tenderness from inflammatory or obstructive complications of the diverticulum do not necessarily occur in the right lower quadrant, but may be located in other abdominal quadrants or vary in location due to small-bowel motility and the position of the diverticulum relative to the ileocecal valve. Sixty per cent of Meckel's diverticula become symptomatic before the age of 10 years. The pattern of complications is also age related, with bleeding most common in children (75 per cent of complications before 10 years of age) and rare after age 30. Small-bowel obstruction may occur at any age, but in children intussusception of the diverticulum is a more likely cause of obstruction. In children, obstruction and bleeding account for approximately equal proportions (30 to 35 per cent) of symptomatic cases. Diverticulitis accounts for 20 per cent of cases, with a peak incidence in older children. Persistent umbilical fistula is seen in neonates. Neoplasms and the hernia of Littré occur in older adults. The incidence of complications of the diverticulum is also related to gender. Although the anomaly is found equally among males and females, complications are observed two to three times more frequently in males.

Peptic ulceration and hemorrhage

Meckel's diverticulum is the most frequent cause of painless, major, lower gastrointestinal bleeding in a previously healthy infant. One-half of cases occur before 2 years of age. More than 95 per cent of bleeding Meckel's diverticula contain ectopic gastric mucosa, causing mucosal ulceration either in the diverticulum near the heterotopia or in the adjacent ileal mucosa. Bleeding may be intermittent or continuous, appearing as melena or exsanguinating bleeding with frank blood and clots. The bleeding has sometimes been described as ‘brick red'. The bleeding is occasionally associated with pain, presumably due to the peptic ulceration. Spontaneous cessation of bleeding usually occurs. There may be chronic bleeding of lesser magnitude. The differential diagnosis includes juvenile polyps, arteriovenous malformation, intestinal hemangiomas, and blood dyscrasias. Diagnostic maneuvers include hematologic examination, sigmoidoscopy, and when appropriate, contrast radiologic studies, which are useful only to exclude other sources of bleeding. The sodium technetium-99m scan, if positive, is usually diagnostic (Fig. 2).

Treatment choices include surgical resection of the diverticulum or of the involved small bowel. Care must be taken not to miss a diverticulum adherent to the ileum. If diverticulectomy is elected, attention should be directed to removing any peptic ulcer in the adjacent ileum. At present the complication rates from small-bowel resection and diverticulectomy with closure are likely to be similar; in earlier reports diverticulectomy was assumed to be safer.

The diagnosis of peptic ulceration of Meckel's diverticulum in the absence of bleeding (‘dyspepsia Meckelii') is uncommon. There may be episodes of abdominal pain in the right lower quadrant or in the region of the umbilicus with a temporal pattern similar to gastroduodenal peptic disease, but not directly relieved by alkali. Pain may be relieved by food intake in a delayed fashion.

Obstruction

The precise preoperative diagnosis of small-bowel obstruction due to Meckel's diverticulum is seldom made in the absence of a characteristic antecedent history of episodes of intermittent obstruction manifested by abdominal pain and vomiting and of peptic symptoms occurring in the absence of demonstrable upper gastrointestinal ulceration. Meckel's diverticulum may cause small-bowel obstruction through several different mechanisms.

1. Two per cent of cases of intussusception are due to Meckel's diverticulum. Intussusception occurs more frequently in children with heterotopic tissue acting as the leading point. The intussuscepting mass may also be due to diverticulitis or contained enterolith. Symptoms include abdominal pain, vomiting, and abdominal tenderness. In addition, rectal bleeding may occur and an abdominal mass may be appreciated. Obstruction may be intermittent, suggesting a reduction of the intussusception. In adults with intussusception due to the diverticulum, neoplasm in the diverticulum is a more likely cause; symptoms may be chronic and recurrent.
2. A persistent, obliterated vitelline duct attached to the umbilicus may serve as a fixed point for volvulus, or cause entrapment as an internal hernia.
3. The diverticulum may be attached by a band to another viscus, resulting in an internal hernia or serving as a fulcrum for volvulus.
4. Similarly, obstruction may occur due to herniation of the small bowel beneath a mesodiverticular band.
5. The mesodiverticular band can cause obstruction by direct ileal compression.
6. Adhesive obstruction may arise secondary to prior inflammation of the diverticulum.
7. A diverticulum which becomes enlarged due to diverticulitis, or by obstruction of the mouth of the diverticulum, may cause obstruction by compression of the adjacent ileum.
8. The diverticulum may become inverted into the ileal lumen, causing an obturating obstruction.
9. The inguinal or femoral hernia described by Littré, mainly in the elderly, contains a strangulated diverticulum. Up to one-quarter of cases occur in umbilical hernias.
10. Rarely, an axial volvulus of the diverticulum occurs, resulting in infarction.
11. In one case of a very long diverticulum, obstruction was due to the formation of a true knot involving another viscus.
12. Obstruction may occur in neonates due to extrusion or prolapse of the ileum through the umbilicus via the patent vitelline duct.

Surgical treatment of obstruction found to be due to Meckel's diverticulum includes division of offending bands if present, and small-bowel resection or simple diverticulectomy, as well as resection of non-viable segments that may occur owing to volvulus or intussusception.

Diverticulitis

Acute diverticulitis may result from peptic ulceration associated with gastric heterotopia, contained enteroliths or foreign body, or obstruction or narrowing of the mouth of the diverticulum. The symptoms of abdominal pain, fever, and vomiting with signs of abdominal tenderness, sometimes in the right lower quadrant, are usually indistinguishable from the presentation seen in acute appendicitis. Abdominal tenderness may be located elsewhere than in the right lower quadrant. Meckel's diverticulitis is potentially more serious than appendicitis, because the process is not contained as often. Perforation and generalized peritonitis may ensue. When the inflammatory process involves an adjacent viscus, a fistula may form. Surgical resection and administration of antibiotics are dictated by the same considerations that apply to appendicitis.

Enteroliths and foreign bodies

Enteroliths are rare complications, forming in narrow-necked diverticula where there is stasis (see Chapter 24.10). True foreign bodies may lodge in the diverticulum. Both enteroliths and foreign bodies may be associated with diverticulitis, may result in bowel obstruction, or can cause hemorrhage from local pressure necrosis. Plain abdominal films that show calculi as well as small-bowel obstruction may cause diagnostic confusion with gallstone ileus.

Neoplasms

Benign or malignant neoplasms account for 1 per cent of the complications of Meckel's diverticulum, occurring more often in men. In contrast to the small intestine *per se*, where adenocarcinoma is the most common histologic type, carcinoid (34 per cent) and leiomyosarcoma (18 to 44 per cent) are most commonly identified in Meckel's diverticulum, while 12 to 20 per cent of tumors are adenocarcinomas. Among benign tumors, leiomyoma is the most frequent. Although the majority of carcinoids are incidental findings, 20 per cent are associated with metastases. Neoplasms of Meckel's diverticulum become apparent when they produce obstruction, sometimes as the leading point for intussusception, or when they cause hemorrhage (sometimes occult) or inflammation.

Management of the incidental Meckel's diverticulum

Whether or not to remove a Meckel's diverticulum when it is found incidentally at laparotomy remains a source of controversy. Early series of cases, based on autopsy or pathologic materials or on retrospective clinical studies that were often anecdotal, resulted in the conclusion that up to 34 per cent of patients with a diverticulum would suffer a complication. Surgery for complications of Meckel's diverticulum appeared to result in a higher incidence of mortality and morbidity than did incidental resection when an uncomplicated Meckel's was found at laparotomy, particularly in the elderly. Opinions regarding incidental diverticulectomy have been based on:

1. the incidence of the anomaly in the population;
2. the likelihood in that population of complications that may be related to age and gender or to anomalous anatomic findings at laparotomy;
3. the morbidity and mortality attributed to resection of a non-diseased diverticulum found incidentally compared with that for a complicated Meckel's diverticulum.

Whether the incidence in a population is 1 or 2 per cent, and whether the complication rate is determined to be 5 or 10 per cent, would result in opposing recommendations.

A substantially large percentage of Meckel's diverticula remain innocuous. Soltero and Bill concluded that at birth there is a 4.2 per cent risk of complications from Meckel's diverticulum over the life time of a patient, and that such risk was age related, falling to 1 per cent at the age 50 years and to near zero in the elderly. Most recent reports are conflicting, in which the incidence of complications of Meckel's diverticulum is instead evenly distributed by age. The current mortality of surgery for complicated Meckel's diverticula is lower than in the past. On the other hand, the mortality for excision of incidental Meckel's diverticulum in youth is virtually nil, while the morbidity associated with small-bowel resection or enterotomy remains.

None the less, certain guidelines may be offered. The most important factor is the age of the patient. The majority of surgeons would remove an incidental Meckel's diverticulum in children under 2 years of age, but most would not interfere with this lesion in patients over 30 to 40 years of age in the absence of complicating pathology. Relative indications for resection include diverticula in men and the nature of the pathologic process that prompted the laparotomy, such as the presence or absence of appendicitis or peritonitis. Thus, in children or youths undergoing elective surgery, removal of an incidental Meckel's diverticulum is reasonable. Since ectopic tissue is found in 40 to 50 per cent or more of patients with complications of Meckel's diverticulum (compared with 6 to 10 per cent of patients with bland diverticula), the presence of palpable heterotopia may be an indication for resection. However, up to one-half of heterotopia are not appreciated on palpation, although the larger ones usually are. There is insufficient evidence to argue that diverticula with a narrow orifice should be removed. Longer diverticula may be more likely to cause difficulties. Findings at surgery indicating prior diverticulitis, including scarring or adhesions, would dictate excision. The presence of a band to the umbilicus or other viscus would require, as a minimum, division of the band at any age.

Complicated or symptomatic Meckel's diverticulum requires ileal resection rather than stapling across the base of the diverticulum with diverticulectomy, as all the heterotopic gastric mucosa and any associated ileal ulceration must be removed. Simple diverticulectomy can suffice for some diverticula found incidentally. The diagnosis and treatment of Meckel's diverticulum through a laparoscopic approach has been reported. Liberalization of indications for excision of incidental Meckel's diverticulum simply because of the availability of laparoscopic approaches and stapling devices is not justified.

Acquired diverticula of the small bowel

In contrast to Meckel's diverticulum, small-bowel diverticula, like colonic diverticula, are acquired; they occur along the mesenteric border and usually lack muscular layers. The reported incidence of small-bowel diverticulosis varies between 0.2 and 4.6 per cent in autopsy series, no doubt dependent on the age of the population under study and the degree of enthusiasm of the prosector for their detection, using bowel insufflation. The mean age for detection is in the seventh decade. The diverticula may be single, but are usually multiple, and are confined to the jejunum in 80 to 90 per cent of cases, the remainder being in both jejunum and ileum or confined only to the ileum (Fig. 3). Jejunal diverticulosis is associated (in 33 to 75 per cent of cases) with diverticula elsewhere in the gastrointestinal tract, particularly in the colon.



Fig. 3. Operative specimen of a segment of jejunum inflated to show three large diverticula which were the source of gastrointestinal bleeding in a 53-year-old man. This patient had melena, requiring 13 units of blood, with negative preoperative endoscopic and angiographic examinations.

Small-bowel diverticulosis has been associated with disorders of intestinal motility in which there is ‘intestinal pseudo-obstruction' due to myopathy or neuroenteric disorders; these pulsion diverticula, due to increased intraluminal pressure acting at weak points in the lumen, are proposed to represent the endstage of a chronic intestinal motility disorder. At surgery the jejunal musculature may be thickened, with relative dilatation of the proximal bowel.

Diagnosis

Because of an aging population, the diagnoses of small-bowel diverticulosis and its complications are being made more frequently. Vague epigastric or periumbilical pain, bloating, and early satiety have been ascribed to small-bowel diverticulosis. However, these symptoms allow a specific diagnosis only by exclusion. The triad of anemia, epigastric distress, and air–fluid levels on plain abdominal films may be a clue to the diagnosis. Small-bowel diverticulosis is most often discovered as a result of more dramatic complications. Except in the case of malabsorption, the cause is not often identified preoperatively.

Jejunal diverticula may be detected by barium studies of the small bowel, particularly on delayed films. Jejunal dyskinesia can be seen during fluoroscopy, where barium is not propelled normally, but moves in and out of the diverticula, associated with partial obstruction. The diagnostic yield is enhanced by small-bowel enteroclysis, in which the duodenum/jejunum is intubated and the bowel is insufflated and examined by the radiologist during compression.

Complications of small-bowel diverticulosis and their treatment

Pathologic consequences of small-bowel diverticulosis include diverticulitis, hemorrhage, obstruction, and malabsorption. The formation of enteroliths, fistulas to surrounding organs, asymptomatic pneumoperitoneum, and malignant tumors are less common complications.

Diverticulitis

The pathogenesis of small-bowel diverticulitis parallels that of colonic diverticulitis. The common presenting complaints are pain, nausea, vomiting, signs of sepsis, and abdominal tenderness. Occasionally a mass is palpable. The diagnosis is rarely made preoperatively, as the process may be localized in virtually any abdominal quadrant. Perforation may occur into the mesentery, can be walled off by adjacent organs, or may result in free perforation with generalized peritonitis. Older anecdotal reports of high mortality rates for perforated diverticulitis treated by surgical resection are no longer valid.

Obstruction

In addition to the ‘pseudo-obstruction', or motility disorder, associated with small-bowel diverticulosis, frank mechanical obstruction may occur from diverticulitis, adhesions associated with the inflammatory process, enteroliths arising in a diverticulum, and volvulus about surrounding adhesions. Relief of the obstruction should

include resection of the segment involved with diverticulitis. Enteroliths, usually composed of choleic acid, form in diverticula where there is stasis, and may lead to obstruction due to local diverticulitis or to obturation in more distal small bowel. Enterotomy to remove enteroliths is sufficient treatment, although resection of diverticula, if localized, may be advocated. When jejunal diverticulosis is found incidentally at laparotomy, resection should be carried out if there are signs of pseudo-obstruction of the involved segment (thickened bowel wall and proximal dilatation). There is little evidence at present to indicate that asymptomatic diverticulosis of the jejunum or ileum should be resected.

Hemorrhage

Hemorrhage from jejunal diverticulosis presents as massive rectal bleeding in more than two-thirds of instances, occasionally accompanied by hematemesis; bleeding is often recurrent. Initial endoscopy serves to exclude other sources. Arteriography is the procedure of choice thereafter for localization of the bleeding site, although sometimes the tagged red blood cell scan is useful. When preoperative localization of the bleeding site fails and laparotomy is still required, cross-clamping of the bowel above and below the bleeding site may be useful to define the segment which fills with blood, thus dictating resection. When bleeding cannot be localized in this way and the patient, usually elderly, has concurrent colonic diverticulosis, thought preoperatively to be a likely source, both colon resection and resection of the small-bowel segment involved with diverticulosis should be done.

Malabsorption

Small-bowel diverticulosis is one possible cause of the ‘blind loop' or ‘stagnant bowel syndrome', in which there is stasis of intestinal contents, resulting in bacterial overgrowth. The metabolic consequences of bacterial overgrowth include megaloblastic anemia, due to bacterial uptake of vitamin B_{12}, and to steatorrhea, due to decreased bile salts as a result of bacterial hydrolysis of conjugated bile salts, with impaired micelle formation. In addition, there may be diarrhea, weight loss, neuropathy, hypoproteinemia, and mild abdominal pain associated with jejunal dyskinesia. In addition to laboratory evidence for malabsorption, stasis is also indicated by the presence of colonic bacteria in duodenal aspirates and by radiologic detection of the diverticula. Initial treatment includes vitamin B_{12} parenteral supplementation and broad-spectrum antibiotics. Twenty-five per cent of patients do not respond to this treatment, and in these cases resection of the segment containing the diverticula is required. Because diverticula are usually in the jejunum, extended resections are relatively well tolerated.

Further reading

Meckel's diverticulum

Case Records of the Massachusetts General Hospital. *New England Journal of Medicine* 1985; **313**: 680–8. [Reviews the differential diagnosis of radionuclide Meckel's scan.]

Dowse JLA. Meckel's diverticulum. *British Journal of Surgery* 1961; **48**: 392–9. [Review of 100 cases of Meckel's diverticulum, detailing presentation and description of morphological types.]

Mackey WC, Dineen P. A fifty year experience with Meckel's diverticulum. *Surgery, Gynecology and Obstetrics* 1983; **156:** 56–64. [Summary of modes of presentation in 402 patients; describes the management of the incidental diverticulum.]

Cullen JJ, Kelly KA, Moir CR, Hodge DO, Zinsmeister AR, Melton LJ. Surgical management of Meckel's diverticulum: an epidemiologic, population-based study. *Annals of Surgery* 1994; **220**: 564–9.

Editorial. Meckel's diverticulum: surgical guidelines at last? *Lancet* 1983; **ii:** 438–9.

Leijonmarck CE. Meckel's diverticulum in the adult. *British Journal of Surgery* 1986; **73**: 146–9.

Peoples JB, Lichtenberger EJ, Dunn MM. Incidental Meckel's diverticulectomy in adults. *Surgery* 1995; **118:** 649-52.

Soltero MJ, Bill AH. The natural history of Meckel's diverticulum and its relation to incidental removal. *American Journal of Surgery* 1976; **132**: 168–73.

Williams RS. Management of Meckel's diverticulum. *British Journal of Surgery* 1981; **68:** 477–88.

[The above six papers address the risk due to Meckel's diverticulum and suggest guidelines for managing the incidentally discovered Meckel's diverticulum.]

Acquired diverticula of the small bowel

Akhrass R, Yaffe MB, Fischer C, Ponsky J, Shuck JM. Small-bowel diverticulosis: perceptions and reality. *Journal of the American College of Surgeons* 1997; **184**: 383–8.

Wilcox RD, Shatney CH. Surgical implications of jejunal diverticula. *Southern Medical Journal* 1988; **81**:1386–91. [The above two papers are retrospective reviews of experience with jejunal diverticular disease.]

Maglinte DDT, Chernish SM, DeWeese R, Kelvin R, Kelvin FM, Brunelle RL. Acquired jejunoileal diverticular disease: subject review. *Radiology* 1986; **158:** 577–80. [Detection of small bowel diverticula by radiologic enteroclysis and review of the complications of acquired diverticula.]

24.5 Malabsorption syndromes

D. P. Jewell

Physiology of digestion and absorption
Carbohydrates
Protein
Fat
Minerals and vitamins
Intestinal adaptation
Clinical features of malabsorption
Features of general ill health
Diarrhoea
Abdominal symptoms
Nutritional deficiencies
Investigation
Documentation of malabsorption
Investigations to determine the cause
Causes of malabsorption
Failure of intraluminal digestion
Failure of mucosal absorption
Short bowel syndrome
Management of malabsorption
Further reading

The daily dietary intake of an adult consists, on average, of 70 g protein, 100 g fat, and 300 g carbohydrate. The average fluid intake is 1.5 litres in 24 h and, in addition, about 7 litres of endogenous secretions enter the small intestine. The digestive and absorptive capacity is such that only 500 to 700 ml passes into the colon from the ileum during a 24-h period and the average stool weight on a Western diet is 150 to 200 g (Fig. 1).

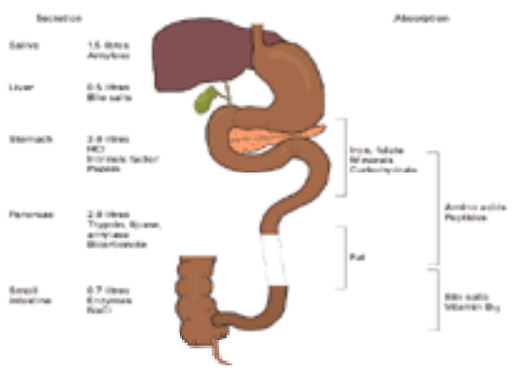


Fig. 1. Sites of secretion and absorption in the upper gastrointestinal tract. Volumes represent 24-h outputs.

Physiology of digestion and absorption

Carbohydrates

The major dietary carbohydrates comprise starch, disaccharides (sucrose and lactose), and monosaccharides (glucose and fructose). Salivary and pancreatic amylases split the 1–4a-glucosidic linkages of the starch to liberate maltose. Further digestion and absorption of the disaccharides is achieved by the action of the oligosaccharidases (lactase, sucrase, maltase), which are globular proteins within the brush border of the enterocyte. The glucose and galactose released by these enzymes are transported into the cell by Na^+-dependent specific transporter proteins. Fructose transport occurs by facilitated diffusion and is not Na^+ dependent.

Protein

Protein digestion begins in the stomach, where secreted pepsinogen becomes activated by an acid pH, and continues in the duodenum and jejunum by pancreatic proteases. These are again secreted in their inactive form (trypsinogen, chymotrypsinogen) and are activated by enterokinase, an enzyme secreted by the duodenal mucosa. This intraluminal digestion results in the release of small peptides containing two to four amino acids. Further digestion occurs by the action of peptidases located in the brush border. Dipeptides can be absorbed by a specific transport system following hydrolysis by dipeptidases, whereas the absorption of free amino acids is achieved by a number of mechanisms, all of which are Na^+ dependent and hence require energy.

Fat

Most of dietary fat is in the form of triglycerides, which contain three long chains of fatty acid linked to glycerol. The first step in fat digestion is hydrolysis by pancreatic lipase which splits two of the fatty acid chains from the a-positions of the glycerol molecule leaving a b-monoglyceride. The second step depends on the formation of micelles by the bile salts which carry the long-chain fatty acids and the monoglyceride into the water phase and hence to the lipophilic surface of the enterocyte. At this point, the micelles disaggregate and the free fatty acids and the monoglyceride pass through the brush border into the cytoplasm by mechanisms which are poorly understood. Once in the cytoplasm, a series of reactions involving cholesterol leads to the synthesis of chylomicrons and very low-density lipoproteins.

The presence of adequate concentrations of bile salts in the small intestinal lumen is therefore crucial for fat absorption. In man the primary bile acids are cholic acid (a trihydroxy bile acid) and two dihydroxycholic acids, chenodeoxycholic and deoxycholic acid, which are conjugated to taurine or glycine to form the bile salts. The majority of bile salts secreted into the bile are reabsorbed from the ileum and recycled—this occurs 10 times on average in a 24-h period. The total pool of bile salts is about 2 to 4 g and the 24-h synthesis rate is about 600 mg: this makes up for faecal losses. It follows, therefore, that fat malabsorption may occur as a result of either cholestasis or ileal disease or resection which causes increased bile salt loss. When the liver synthesizes less bile salts than are lost, the intestinal concentration falls to a point below the minimal micellar concentration and fat malabsorption results. Two other mechanisms will potentially lower the bile salt concentration in the lumen. First, at normal jejunal pH, the bile salts are present in ionized form. If the pH falls (for example, in the Zollinger–Ellison syndrome) the bile salts become non-ionized and are then absorbed by passive non-ionic diffusion. Second, bacteria (especially anaerobes) are able to deconjugate bile salts, leading to the steatorrhoea which occurs in association with bacterial overgrowth of the small intestine.

Minerals and vitamins

Most minerals and vitamins are absorbed from the proximal jejunum. Vitamins A, D, and K are fat soluble and dependent on the normal mechanism of fat absorption; malabsorption therefore occurs in the presence of steatorrhoea. Iron is absorbed from the duodenum by a carrier-mediated process that is controlled by the body's iron stores and which involves mechanisms that are not clearly defined; thus, iron absorption increases in the presence of iron deficiency. Folic acid is present as polyglutamates in the diet and these are hydrolyzed to monoglutamates by folate deconjugase, another brush-border enzyme. Calcium, like iron, is absorbed in the duodenum and proximal jejunum by a carrier-mediated process involving calcium-binding protein and a calcium-activated ATPase. The synthesis of the calcium-binding protein is under the control of 1,25-dihydroxy cholecalciferol (1–25 $(OH_2)D_3$).

Vitamin B_{12} is only absorbed from the terminal ileum and in order for it to combine with its specific receptor on the enterocyte surface it has to be complexed with intrinsic factor, a glycoprotein synthesized by the gastric parietal cells. Deficiency of intrinsic factor, either due to subtotal gastric resection or to autoimmune

mechanisms as in pernicious anaemia, causes vitamin B_{12} malabsorption. Vitamin B_{12} malabsorption also results from ileal disease (such as Crohn's disease) or from ileal resection, due to loss of enterocyte receptors for the vitamin B_{12}–intrinsic factor complex.

Intestinal adaptation

The small intestine has considerable reserve capacity since, as previously mentioned, absorption mainly occurs in the jejunum with only fat, bile salts, and vitamin B_{12} being absorbed from the ileum. Studies in experimental animals, as well as observations in humans following small intestinal resection, have shown that the ileum can undergo hypertrophy, which increases its absorptive capacity very considerably. This contrasts with the proximal small intestine, which has only a limited capacity to adapt. Thus, resection of the jejunum may not necessarily cause malabsorption whereas ileal resection inevitably causes malabsorption of vitamin B_{12} and bile salts. If more than 100 cm of ileum is resected, bile salt malabsorption becomes sufficiently severe for hepatic synthesis to fail to maintain the circulating pool and fat malabsorption occurs.

Clinical features of malabsorption

Features of general ill health

These include weight loss, malaise and lassitude, anorexia, and symptoms of anaemia. There may also be evidence of hypoalbuminaemia with dependent oedema and skin changes. Finger clubbing is common in patients with long-standing malabsorption such as coeliac disease. Aphthous ulcers in the mouth are frequently present in patients with coeliac disease or Crohn's disease.

Diarrhoea

This usually, although by no means always, has the features of a steatorrhoea. The stools become bulky, pale, and offensive. Their high fat content causes them to float and they are difficult to flush away. Stool frequency may be increased. However, many patients with malabsorption may not complain of an altered bowel habit.

Abdominal symptoms

Symptoms of excess wind, abdominal distension, borborygmi, and varying degrees of discomfort are common. If the diarrhoea is not typically a steatorrhoea, these symptoms may be diagnosed as an irritable bowel syndrome. The combination of abdominal distension and wasting may give rise to a marked ‘pot-belly', particularly in children.

Nutritional deficiencies

The physical signs of iron deficiency are common and include a glossitis, angular stomatitis (Fig. 2), brittle nails, and a dry skin. Neurological complications of vitamin B_{12} deficiency (dementia, peripheral neuropathy, subacute combined degeneration of the spinal cord) are rarely seen except in patients with a severe pernicious anaemia. Proximal myopathy may result from osteomalacia or hypokalaemia. Bone pain and positive Chovstek's and Trousseau's signs may indicate osteomalacia and the last two signs may also be caused by hypomagnesaemia. Bruising may indicate vitamin K deficiency. Night blindness as a result of vitamin A deficiency is rare but is probably more common than is recognized. Pigmentation is common.

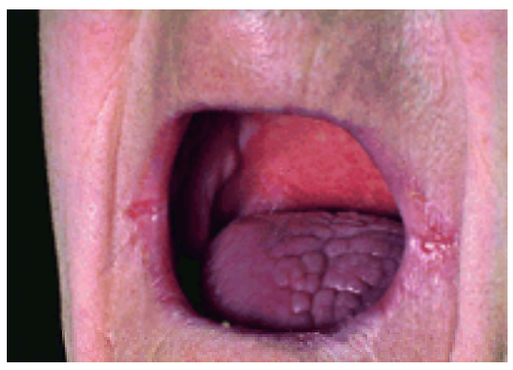


Fig. 2. Angular stomatitis and glossitis in a patient with iron deficiency.

Investigation

Investigation of a patient with suspected malabsorption is directed towards documenting the nature and severity of the malabsorption and defining the cause.

Documentation of malabsorption

Table 1 lists the investigations that will allow the degree of malabsorption to be assessed. The presence of fat malabsorption, and therefore of vitamins A, D, and K, but with normal serum concentrations of iron and folate, is highly suggestive of a pancreatic aetiology rather than a mucosal cause. This can be confirmed with a D-xylose test: this test, performed on a 5-h urine collection and a 1-h serum sample following the ingestion of 25 g of xylose, will detect patients with mucosal abnormalities causing malabsorption with a high degree of specificity and sensitivity. An abnormal test result may be due, however, to inadequate urinary collection, delay of gastric emptying, or interference by drugs such as aspirin, neomycin, or non-steroidal anti-inflammatory drugs. Thus, it is not widely used in clinical practice.

Haemoglobin
Mean cell volume
Blood film
Total white cell count, platelets
Serum iron and total iron binding capacity
Serum folate and vitamin B_{12}
Serum calcium, phosphate, alkaline phosphatase
24-h urinary calcium
Serum K^+, Mg^{2+}
Bilirubin, aminotransferases
Prothrombin time
Total plasma protein, albumin
Faecal fat excretion
D-Xylose absorption tests
Breath tests—lactose, lactulose, ^{14}C-triolein
Secretin test

Table 1 Investigations to document the presence of malabsorption.

Estimation of faecal fat excretion is crucial in order to document steatorrhoea, and this should be performed on a 3-day stool collection. Unfortunately an increasing number of clinical laboratories are refusing to provide this assay despite the recent development of closed systems for faecal collection and homogenization. Staining faecal smears with appropriate stains (such as Oil Red O) can detect an excess of fat globules and this qualitative test has shown good correlation with the quantitative assay. However, the qualitative test is unreliable in patients with mild steatorrhoea (<10 g/24 h). The ^{14}C-triolein breath test can also be used to assess fat absorption, but this suffers from the disadvantages of an 8-h duration and the need to administer a test meal containing 60 g of fat.

Investigations to determine the cause

Small bowel biopsy

This is the definitive test for most causes of mucosal malabsorption. It is essential for the diagnosis of coeliac disease, tropical sprue, and intestinal lymphangiectasia and is diagnostic for malabsorption caused by amyloidosis, Whipple's disease, a-chain disease, and parasitic infestations (giardiasis, cryptosporidiosis). Endoscopic biopsies from the distal duodenum are increasingly being used and they appear to be as useful as jejunal biopsies taken with the Crosby capsule. For multiple biopsies from differing sites, the Quinton hydraulic machine has to be used (Fig. 3).



Fig. 3. Instruments for jejunal biopsy: the Quinton hydraulic suction tube (left), Medi-Tech steerable catheter (middle), Crosby capsule (right).

Barium radiology

Contrast examination of the small intestine is best performed using a small bowel enema. It is essential for detecting anatomic abnormalities (such as jejunal diverticulosis), strictures (as in Crohn's disease, radiation enteritis, ischaemia, systemic sclerosis, lymphomas, carcinoma) or motility disorders (pseudo-obstruction, systemic sclerosis).

Ultrasound, CT scanning

These imaging modalities are primarily useful for detecting biliary, pancreatic, or hepatic disease. CT scans may also detect infiltration of the small intestinal wall (such as lymphoma) and enlarged mesenteric nodes.

Breath tests

Hypolactasia is best diagnosed by biochemical assay of lactase activity in a jejunal biopsy specimen. However, this assay is not widely available and the lactose breath test provides a simple, rapid, and non-invasive alternative. The hydrogen concentration in a sample of end-expiratory air is measured before and 2 h after a 50 g dose of lactose given to a fasting patient. A rise of greater than 20 p.p.m. of H_2 indicates hypolactasia.

Bacterial overgrowth of the small bowel can be detected using the lactulose breath test. Lactulose is a non-absorbed disaccharide which usually passes into the colon where it undergoes bacterial fermentation. Much of the released hydrogen is absorbed and expired and in a normal subject there is a rise in breath hydrogen after 90 to 150 min, which represents small intestinal transit time. Bacterial overgrowth of the small bowel is associated with an early rise of breath hydrogen (at 15 to 45 min) (Fig. 4). Obviously, the test can be misleading in patients with altered transit times. The ^{14}C-glycocholate breath test may also be used and relies on the ability of bacteria to deconjugate the bile salt. Aspiration of jejunal contents for microbiology can be used but meticulous technique is required if this is to be reliable.

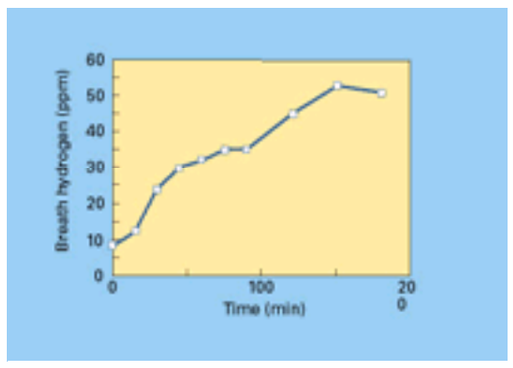


Fig. 4. The results of a lactulose breath test in a patient with jejunal diverticulosis. There is an early rise in breath hydrogen (expressed as parts per million) beginning at 10 min. The colonic peak occurs at 150 min.

Immunological tests

Measurement of serum immunoglobulins is essential for the diagnosis of hypogammaglobulinaemia. The presence of antibodies to reticulin or gliadin supports the diagnosis of coeliac disease, but is not diagnostic. However, a positive serum antibody to endomysium has at least a 95 per cent specificity and sensitivity for coeliac disease. Most laboratories measure an IgA antibody, thus the test is meaningless in the 5 per cent or so of patients with coeliac disease who also have an IgA deficiency. The antigen is tissue transglutaminase but it is not yet clear how this is linked to the pathogenesis of coeliac disease.

Antibody to intrinsic factor in serum is highly suggestive of pernicious anaemia. Selective immunoelectrophoresis of saliva, serum, urine, or jejunal aspirate is required to diagnose the free a-chains characteristic of a-chain disease.

Causes of malabsorption

There are numerous causes of malabsorption and it is convenient to divide them into those involving a failure of intraluminal digestion and those where there is a failure of mucosal absorption. Table 2 lists some of the major causes.

Failure of intraluminal digestion	
	Cholestasis
	Pancreatic insufficiency
	Poor mixing
	Bacterial overgrowth
	Drugs
Failure of mucosal absorption	
	Coeliac disease
	Tropical sprue
	Whipple's disease
	Hypogammaglobulinaemia
	Lymphoma, α-chain disease
	Infections
	Radiation damage
	Mesenteric ischaemia
	Enzyme deficiencies
	Crohn's disease
	Short bowel syndrome

Table 2 Causes of malabsorption

Failure of intraluminal digestion

Cholestasis

This may be due to intrahepatic or extrahepatic causes: the former include primary biliary sclerosis, alcohol, drugs, or viral hepatitis; the latter include benign or malignant biliary strictures, ampullary or pancreatic tumours, and bile-duct stones. Chronic pancreatitis may also result in obstruction of the lower end of the common bile duct, giving rise to a typical ‘rat-tail' stricture which can be seen at endoscopic retrograde cholangiopancreatography (ERCP). Primary sclerosing cholangitis can involve both intrahepatic and extrahepatic ducts. Whatever the cause of cholestasis, malabsorption is caused by bile salt insufficiency, and patients therefore develop steatorrhoea and the accompanying malabsorption of the fat soluble vitamins (A, D, K).

Pancreatic insufficiency

This is usually caused by chronic pancreatitis or by proximal obstruction to the pancreatic duct by tumour, and results in steatorrhoea and malabsorption of the fat-soluble vitamins. The secretory capacity of the pancreas has to be reduced to less than 10 per cent of normal before malabsorption develops.

Poor mixing

Inadequate mixing of food with bile and pancreatic secretions may occur after a Polya gastrectomy, gastroenterostomy, or a Roux-en-Y procedure. This may give rise to a degree of maldigestion but, if frank malabsorption occurs, other factors such as bacterial overgrowth may also be involved.

Bacterial overgrowth

The mechanisms involved in fat malabsorption caused by bacterial overgrowth are discussed above. Bacteria may also compete with the host for vitamin B_{12}, resulting in the development of a megaloblastic anaemia. The proximal small intestinal flora is usually less than 10^4 organisms/ml: bacterial counts greater than this constitute bacterial overgrowth. This is found in many small bowel diseases, including Crohn's disease, jejunal diverticulosis, strictures, autonomic neuropathies, or other causes of pseudo-obstruction. Bacterial overgrowth will also complicate a ‘blind loop' such as the afferent limb of a Polya gastrectomy or an end-to-side anastomosis (Fig. 5), or internal enteric fistulas (especially gastrocolic or ileocolic). It is recognized now that elderly patients may develop bacterial overgrowth and malabsorption in the absence of any demonstrable anatomic abnormality. In some patients with bacterial overgrowth (for example, those with jejunal diverticulosis) the small bowel mucosa may show some loss of villous height and a chronic inflammatory infiltrate, but this is never as severe as the mucosal changes seen in coeliac disease.



Fig. 5. A large blind loop as a result of an end-to-side anastomosis in the small intestine (by courtesy of Dr D. J. Nolan).

Jejunal diverticulosis (Fig. 6) occurs in the elderly and is usually asymptomatic unless bacterial overgrowth is sufficient to cause steatorrhoea. When this occurs antibiotics should be given, as for any cause of bacterial overgrowth. Metronidazole is the drug of choice, but since many of these patients will require repeated courses, which may induce resistant organisms, doxycycline and neomycin can be used in rotation with metronidazole. The complications of jejunal diverticulosis, in addition to malabsorption, are a silent perforation (which needs no specific treatment) and volvulus (which requires resection). Only very rarely does an acute diverticulitis develop; this should, in the first instance, be treated with intravenous fluids and antibiotics.



Fig. 6. Small bowel enema demonstrating jejunal diverticulosis (by courtesy of Dr D. J. Nolan).

Autonomic neuropathy of the intestine is usually seen in association with diabetes mellitus but other causes include vincristine, spinal cord injury, Shy–Drager syndrome, amyloidosis, and systemic sclerosis. Other signs of autonomic neuropathy, such as postural hypotension and a fixed R–R interval on a Valsalva manoeuvre, are nearly always present. Apart from controlling bacterial overgrowth, management consists of symptomatic control of diarrhoea and nutritional support.

Drugs

The drug in common usage which may rarely cause malabsorption is neomycin, which appears to cause steatorrhoea by precipitating bile salts. Cholestyramine can also increase fat excretion by binding bile salts.

Failure of mucosal absorption

Coeliac disease

Coeliac disease predominantly affects children and young adults, although it may occur at any age. The incidence varies considerably, being present in about 1 in 500 to 1000 of the population in the West but being exceptionally rare in other parts of the world. This distribution may be explained by the very close association of coeliac disease with HLA antigens. Approximately 95 per cent of patients with coeliac disease are HLA DQ2 and over 80 per cent have the haplotype B8, DR3, DQ2. Thus, the aetiology of coeliac disease represents an interplay between genetic factors and the ingestion of gluten. Nevertheless, the fact that there is 30 per cent discordance amongst monozygotic twins suggests that other factors are important in initiating the disease. Recently, there has been considerable interest in the role of adenovirus 12 as an initiating factor because there is sequence homology between an epitope in an ‘early protein' (E1b) of the virus and a surface epitope of a gliadin fraction (gliadin is prepared from gluten by alcohol extraction). Patients with coeliac disease show cellular immune responses to both viral and gliadin epitopes and the gliadin epitope has been shown to be toxic in *in vivo* challenge studies. Hence, it is possible that exposure to this virus in a person with the genetic susceptibility to develop coeliac disease may break oral tolerance to gluten and allow a harmful mucosal immunological reaction to occur.

Although patients with coeliac disease may still present with the classic clinical picture of weight loss, anorexia, steatorrhoea, and the signs of nutritional deficiency, the majority of patients are now diagnosed at a much earlier stage: unexplained iron deficiency anaemia or the symptoms of an irritable bowel syndrome are the most frequent presenting features. Any patient who presents with a variable bowel habit, abdominal pain, distension, and wind, who has lost weight or who also complains of aphthous ulcers of the mouth should undergo a small intestinal biopsy. This will show severe villous atrophy with crypt hyperplasia in affected individuals. Treatment consists of a gluten-free diet. Patients should be instructed by dietitians and, ideally, should also become members of a national coeliac society, which continually provides their members with information on the gluten status of new food products and disseminates knowledge about the disease. A repeat biopsy should be obtained after 3 to 4 months of dietary treatment to check that the mucosa has returned towards normal. The strict definition of coeliac disease requires a gluten challenge with repeat biopsy, but this is usually reserved for patients in whom the diagnosis may be in doubt. Patients will need to be on a gluten-free diet for life and should adhere to it strictly. A visit to a specialized clinic is recommended at least once a year.

Adherence to a strict diet not only achieves maximal histological remission, but may avoid the long-term complications of coeliac disease, namely an ulcerating jejunoileitis or a mucosal T-cell lymphoma. Carcinomas of the gastrointestinal tract, especially the oesophagus, are also associated with coeliac disease, but whether strict dietary control influences the incidence of carcinoma is less clear.

Patients who have clear evidence of nutritional deficiencies (e.g. iron, folate, vitamins K or D) will need replacement therapy initially, but once the jejunal mucosa returns to near normal there should be no need for long-term supplements. If the coeliac disease is severe, affecting a substantial part of the small intestine (indicated by a marked steatorrhoea), patients may also benefit from lactose withdrawal until the brush border of the enterocytes has recovered on the gluten-free diet.

Some patients (<10 per cent) who appear to have coeliac disease do not show histological recovery on gluten withdrawal even after many months. The usual reason is lack of dietary compliance. However, if this has been excluded, other conditions should be considered. These include collagenous sprue (the biopsy shows >12 µm subepithelial collagen band and, usually, hypoplastic crypts), lymphoma, hypogammaglobulinaemia, amyloidosis, tropical sprue, chronic ischaaemia, or Whipple's disease. Other conditions such as Crohn's disease, tuberculosis, systemic sclerosis, and radiation enteritis may also be associated with a flat mucosal biopsy, but there is usually obvious evidence of these diseases on barium radiology.

Hypogammaglobulinaemia

IgA deficiency is the most common immunodeficiency disorder, and may be symptomless or associated with recurrent infections of the upper respiratory tract, sinusitis, and skin sepsis. It also affects about 5 per cent of patients with coeliac disease.

Common variable hypogammaglobulinaemia, in which there is a variable depression of serum IgA, IgG, and IgM, is also associated with a flat small intestinal mucosa which does not usually respond to gluten withdrawal. Many of these patients develop giardiasis or cryptosporidiosis, and bacterial overgrowth is common. Treatment involves eradication of infection and patients may benefit from a lactose-free diet.

AIDS is associated with a lack of mucosal CD4+ cells in parallel with the fall in circulating CD4+ lymphocytes. Hence, degrees of mucosal abnormalities are common in these patients and opportunistic infections may contribute to the mucosal damage and subsequent malabsorption.

Lymphoma

Intestinal lymphomas are of three main types.

1. The so-called Western lymphoma is usually a focal tumour (Fig. 7). It is a B-cell lymphoma, mainly centrocytic or centroblastic in type.



Fig. 7. Lymphoma of the ileum as seen on a small bowel enema (by courtesy of Dr D. J. Nolan).

2. A diffuse lymphoma (Mediterranean type). This usually causes ‘a-chain' disease, in which the abnormal clone of B cells gives rise to plasma cells which synthesize and release free a-chains of IgA without the accompanying light chain. The disease causes profound malabsorption. Diagnosis is made by a small intestinal biopsy and by the demonstration of free a-chains in serum, saliva, jejunal aspirate, or urine. In the early stages the condition may respond to antibiotics, especially tetracycline.
3. T-cell lymphoma may also complicate coeliac disease.

Short bowel syndrome

This usually results from massive surgical resection following mesenteric infarction or from repeated resections for Crohn's disease. Malabsorption occurs because there is loss of absorptive surface. However, there may be bacterial overgrowth or there may be a degree of mucosal abnormality leading to hypolactasia.

Management involves correction of bacterial overgrowth, a lactose-free diet (if necessary, and diet may need to be low in disaccharides generally if there is severe brush-border damage), and full dietary supplements, together with intramuscular administration of vitamin B_{12} (250 µg every 3 months). Some patients will require home parenteral nutrition to maintain body weight and an adequate intake of minerals and vitamins.

Management of malabsorption

The primary management is to treat the cause of the malabsorption: pancreatic enzymes must be given for pancreatic insufficiency, a gluten-free diet for coeliac

disease, antibiotics for bacterial overgrowth. In these instances, there is usually no need for long-term nutritional supplements provided the malabsorption is corrected adequately. However, in the acute stage, replacement with vitamins, iron, calcium, magnesium, and zinc may be necessary.

For patients in whom the cause of malabsorption cannot be corrected, management may require oral supplements of B vitamins, iron, and folate. All extensive mucosal atrophy of the small bowel, and those with a short bowel syndrome, will suffer from malabsorption of vitamins A, D, and K. Monthly intramuscular injections of these fat-soluble vitamins may be needed (e.g. vitamin A 10 000 U, vitamin D 100 000 U, vitamin K 10 mg). Dietary management will include a low fat diet, for those with prolonged steatorrhoea, and supplements of carbohydrate and protein. However, if there is extensive villous atrophy, these supplements will need to be in the form of glucose and tri- or dipeptides which can be absorbed directly without requiring enzymatic digestion. Medium-chain triglyceride should also be given as a way of providing fat in a steadily absorbable form.

Further reading

Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Rieckon EO. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nature Medicine* 1997; **3**: 797–801.

Godkin A, Jewell DP. The pathogenesis of coeliac disease. *Gastroenterology* 1998; **115**: 206–10.

Maki M. Tissue transglutaminase as the autoantigen of coeliac disease. *Gut*, 1997; **41**: 565–6.

Riley SA, Turnberg LA. Maldigestion and malabsorption. In Sleisenger M, Fordtran JS, eds. *Gastrointestinal disease*, 5th edn, pp. 1009–26. WB Saunders, Philadelphia, 1993.

Romano TJ, Dobbins JW. Evaluation of the patient with suspected malabsorption. *Gastroenterology Clinics of North America* 1989; **18**: 467–83.

24.6 Short bowel syndrome

A. Bianchi

Definition
Aetiology
Pathophysiology
Intestinal adaptation response
Presentation and prevention
Medical management
Surgical management
Timing of autologous gastrointestinal reconstruction
Long-term outlook
The future
Further reading

Definition

The short bowel syndrome refers to a state of intestinal failure which may be temporary or permanent, wherein bowel length and absorptive capability are inadequate to sustain survival, growth, and development on enteral nutrition. The minimum intestinal cell mass is influenced by the patient's age at the time of measurement, the type and condition of the residual bowel, the absorptive capability and immunocompetence of the mucosa, the presence of the ileocaecal valve, and the length of the colon.

Aetiology

In adults total parenteral nutrition can be anticipated if the remaining small bowel length is less than 60 to 80 cm, that is 15 to 20 per cent of normal. The commonest causes are mesenteric infarction, Crohn's disease, radiation enteritis, tumours, and trauma. In children, despite individual reports of survival with jejunal lengths as short as 7 cm, realistically an average of 30 to 40 cm of jejunum (some 10 per cent of the total small bowel length) measured at first operation along the antemesenteric border of the collapsed bowel, with an ileocaecal valve and all the colon, is necessary for survival and growth in the full-term neonate with no other bowel disorders. The most common aetiological factors are midgut or segmental volvulus, often occurring antenatally in association with bowel non-fixation or gastroschisis. The commonest cause of short bowel postnatally is extensive necrotizing enterocolitis.

Pathophysiology

The length of the small bowel increases progressively during fetal life, with more rapid growth occurring during the third trimester. Thus the small bowel virtually doubles its length between the 19th to 27th week and the 35th week, eventually averaging between 250 and 330 cm in the full-term normal child. Thus an equivalent length of bowel in a premature child at 33 weeks gestation represents a significantly greater intestinal cell mass than for a full-term child.

Loss of small bowel occurring antenatally within a sterile uterine environment is followed by steady resorption of the necrotic bowel and residual features of intestinal atresia (Fig. 1). The sealed distal bowel is effectively defunctioned and fails to develop further. The obstructed proximal bowel undergoes massive luminal dilatation and hypertrophy of its wall. This process of natural tissue expansion increases the residual small bowel surface area, but renders the dilated loop propulsively ineffective. Postnatally intraluminal stasis leads to bacterial overgrowth, inflammatory mucosal injury, and bacterial translocation with adverse consequences for mucosal adaptation, progressive liver injury, and reduced prospects of survival.

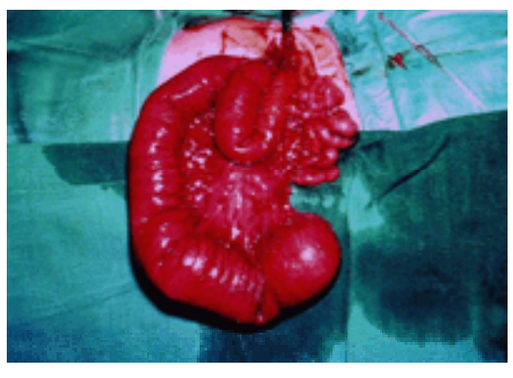


Fig. 1. Typical features of intestinal atresia with dilated proximal segment, collapsed distal segment, and V-shaped mesenteric defect.

The most significant defects associated with massive small bowel loss are a severely reduced absorptive enterocyte mass and a markedly diminished volume of gut-associated immunocompetent tissue because of loss of ileal Peyer's patches and the mesenteric lymph nodes. There is therefore an impact not only on the ability to absorb but also on immunological competence, particularly at the mucosal barrier. Short bowel animal models have been shown to have increased bacterial translocation when compared with normal. Such a heavy bacterial onslaught is likely to play a major role in the development of the progressive, often fatal, liver injury which frequently complicates neonatal short bowel syndrome. Liver injury commences early with frequent septic episodes and a steadily deteriorating cholestatic jaundice. Less than perfect parenteral solutions, such as phytosterols in fat solutions, may further aggravate the problem.

The most obvious effect of the short bowel state is the severe fluid, electrolyte, and global nutrient deficits because of the insufficient absorptive enterocyte mass. In addition there is the loss of enteral secretions normally produced into the proximal bowel and largely resorbed with nutrients during intestinal passage. The loss of specific binding sites leads to failure of absorption of vitamin B_{12}, fat-soluble vitamins, and bile salts. Interference with the enterohepatic circulation and consequent abnormal bile increases the tendency to gallstone formation. Passage of proximal small bowel secretions containing unbound bile salts into the colon results in fermentation, osmotic diarrhoea, and excessive absorption of unbound oxalates with consequent high urinary oxalate excretion and a risk of stone formation.

The short bowel state is associated with major alterations in gut hormone levels. There is a significant hypergastrinaemia and a markedly increased gastric acid secretion creating an acid milieu that is less favorable to enzymatic activity and accounts for the high risk of duodenal ulceration. Less clinically obvious are the raised levels of enteroglucagon, cholecystokinin, motilin, and vasoactive intestinal peptide amongst others. It is fortuitous and interesting that correction occurs spontaneously with increasing intestinal adaptation.

Intestinal adaptation response

Long-term survival and growth following massive small bowel loss is dependent on spontaneous intestinal adaptation within the residual bowel. Once adaptation is complete, absorptive capability can increase by about 400 times. An increase in bowel length is not part of the natural adaptation response such that there is little actual increase in length beyond that expected from normal bowel growth. The adaptation process is one of villus hypertrophy and mucosal cell hyperplasia such that the elongated villi support a larger number of enterocytes and immunologically competent cells. There is therefore a larger brush border and a greater absorptive surface area. It is interesting that although actual absorption per cell is decreased, the overall effect of the greater number of cells is an increase in absorption per unit area of mucosa. Concomitant with the increasing enterocyte mass is a similar enhancement of gut-associated lymphoid tissue on and in the hypertrophic mucosal villi with increased effectiveness of the mucosal barrier. Thus as intestinal adaptation progresses there is less bacterial translocation, a reduced incidence of systemic sepsis, and an improvement in liver function, probably reflecting improved overall immunocompetence.

The intestinal adaptation response is complex and multifactorial. It is dependent on constant mucosal contact with natural nutrients, the presence of pancreaticobiliary secretions, and various as yet incompletely understood hormonal factors. Significantly, animals and patients with a short bowel naturally develop hyperphagia. In comparison those denied enteral nutrients and maintained solely on intravenous feeding demonstrate a poor adaptation response and tend to develop mucosal atrophy. Enteral nutrition is an essential component in the management of the patient with a short bowel. Several hormones including enteroglucagon and epidermal growth factor have positive influences in stimulating mucosal cell proliferation and enhancing absorption. Thus it is likely that epidermal growth factor plays a role in the

crypt cell proliferation seen in response to the short bowel state. In future such hormones may come to play an important role in the treatment of patients with a short bowel.

Presentation and prevention

The child with a short bowel frequently first comes to attention at antenatal fetal scan when dilated loops of bowel raise the possibility of high intestinal atresia. This is even more significant if there is an associated gastroschisis, when midgut volvulus is the most likely cause. The child's inability to absorb liquor, frequently leads to hydramnios and premature labour at 33 to 37 weeks. Vomiting causes the liquor to be bile stained. Commonly these children have no other associated anomalies. Essential to antenatal parent counseling is a clear exposition of the seriousness of the short bowel state, the long-term dependence on parenteral nutrition, and the possibility of a poor outcome with early death from severe liver injury. At this time ‘further management' may be a termination of the pregnancy.

At birth the diagnosis is obvious when the only viable bowel consists of a short dilated loop of jejunum in association with gastroschisis. The necrotic bowel may be absent or may persist as an amorphous mass of tissue (Fig. 2). The child vomits copious amounts of bilious fluid and is unable to feed. Children presenting with intestinal atresia show the same clinical features of a high small bowel obstruction. Abdominal radiography shows a paucity of air-filled loops and one very large fluid level in the dilated segment just proximal to the atresia. The rest of the abdomen is featureless and clinically flat (Fig. 3). At operation the majority of the midgut is absent and the dilated proximal jejunum, representing some 10 per cent of the normal small bowel length, is usually the only remaining absorptive bowel (Fig. 4). The ileocaecal valve is often absent and the defunctioned residual colon commonly runs from mid-transverse or splenic flexure distally.

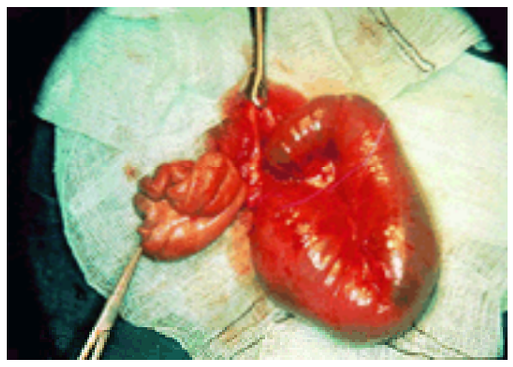


Fig. 2. Gastroschisis with necrotic midgut and dilated proximal atretic jejunum.



Fig. 3. Typical radiograph of high jejunal atresia with a single large fluid level and an otherwise featureless abdomen.



Fig. 4. Short bowel following antenatal midgut volvulus with residual 35-cm dilated jejunal loop.

Short bowel syndrome occurring postnatally may be secondary to volvulus because of non-fixation and malrotation of the midgut or may follow extensive necrotizing enterocolitis. Midgut volvulus presents with sudden onset of copious bile-stained vomiting in a previously well child, with a flat abdomen and with minimal abdominal features on radiography. This typical and alarming history demands an immediate contrast study of the upper gastrointestinal tract which demonstrates the abnormal obstructed S-shaped duodenum with the small bowel passing towards the right abdomen (Fig. 5). Instant laparotomy is indicated, with derotation and return of blood supply to the midgut.

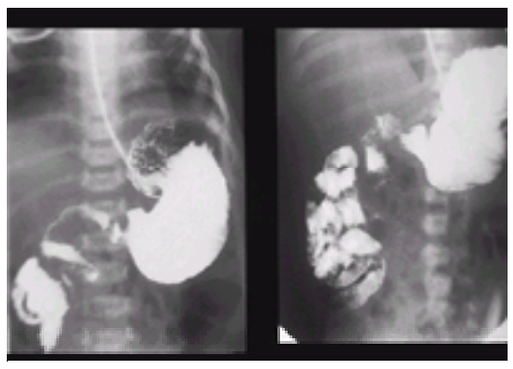


Fig. 5. S-shaped duodenum and small bowel passing to right side of abdomen typical of malrotation and midgut non-fixation.

Children presenting the typical features of necrotizing enterocolitis are frequently premature, often requiring mechanical ventilation for idiopathic respiratory distress syndrome. Others are ‘stressed and at risk' because of congenital heart disease or abnormal pregnancy, for example a mother with diabetes or intrauterine growth retardation. There is commonly a history of early feeding with products based on cow's milk. The child is ill, with features of sepsis relating to intra-abdominal

inflammatory bowel disease. The condition is variable in extent but may involve an extensive length of the small and large bowel. Abdominal radiography reveals a typical granular appearance with static dilated loops of variable size, many demonstrating intramural gas bubbles indicative of severe mucosal ulceration. Intravascular gas may outline the portal tracts within the liver. Localized or generalized perforation may be present with release of gas and septic bowel contents into the peritoneal space.

There is still controversy over the best form of management. Whenever possible, prevention is the obvious choice. Feeding regimens in premature and ‘at risk' babies should be introduced very slowly, preferably with breast milk or predigested formulas rather than products based on cow's milk. At the first sign of abdominal distension, and particularly intraperitoneal inflammation, surgical consultation becomes crucial. It is inappropriate to await the traditional but late indications of localized or generalized peritonitis or abdominal abscess formation prior to considering surgery. These children require early and aggressive treatment. Control of the generalized septicaemia with broad-spectrum antibiotics; cardiovascular support with blood, clotting factors, and platelets; correction of acid–base balance; and respiratory assistance with mechanical ventilation underpin the resuscitative process. The peritoneal cavity should be drained with a large dialysis-type catheter and washed out intermittently with warm dialysate solution. These measures will often improve the child's condition over a period of several hours, following which urgent laparotomy is indicated. The bowel is exteriorized and the peritoneal cavity thoroughly cleaned. Any clearly non-viable bowel is resected, but a strict conservative approach should be adopted, retaining all possibly viable tissue regardless of the number of anastomoses or stomas which may be necessary. The whole of the gastrointestinal tract from mouth to anus is washed out with warm saline, cleaning out all intraluminal septic material and collapsing the entire bowel. Extensive drainage of the peritoneal cavity is provided. Such aggressive early management when the child is still in reasonable condition affords a better opportunity to overcome sepsis and preserves maximal bowel. Further surgery should be avoided for as long as possible allowing bowel recovery. The traditional ‘early second look' is detrimental and may lead to further unnecessary bowel loss. Despite all measures there will be a small group of children with extensive bowel necrosis which will lead to short bowel syndrome. In common with postnatal midgut volvulus these children do not initially have dilated bowel, making management even more complex.

Medical management

The child with established short bowel syndrome and an inability to absorb sufficient fluid, electrolyte, and nutrients from enteral feeds requires long-term parenteral nutrition for survival and growth. The nature of parenteral nutrition is tailored to the child's specific fluid, electrolyte, and nutrient requirements. In view of hepatotoxicity, particularly from phytosterols in fat solutions, the long-term total parenteral nutrition regimen is best modified to a high protein and reduced fat mix, limiting fats to 1 g/kg/day to provide some essential fatty acids. The pattern of delivery should incorporate as a major consideration the child's developmental need for a close-to-normal lifestyle within the family environment. It should be appreciated that many children with a short bowel will have relatively short lives, such that maximal family contact within the home and from an early stage is especially precious. Intravenous feeding should be cycled during the night, thus leaving the child free during the day to enjoy ‘quality time' and normal family activities. Hospital management should rapidly give way to home-based care with support from a unit specialized in the management of the child with this condition.

Enteral feeding, essential to bowel adaptation, should commence early and be sustained with normal or predigested foods which should be administered to bowel tolerance levels. Feeding should be oral, following a normal feeding pattern and allowing the child to learn to appreciate food and taste. Continuous nasogastric or gastrostomy tube feeding, while achieving improved absorption, is less satisfactory for the child and the family. Drug therapy such as codeine, loperamide, cholestyramine, and ursodeoxycholic acid and gut hormones such as epidermal growth factor have variable effects but may be beneficial. As a general principle the long-term outcome depends on sustained enteral feeding and parenteral nutrition until bowel adaptation becomes maximal over a 1- to 2-year period. The smaller the available small bowel mass, the less effective is the adaptation process, with poorer absorption and reduced mucosal immunocompetence.

Surgical management

Placement and care of the central venous feeding catheter and scrupulous preservation of the limited venous access sites, the child's life line to long-term survival, is a major surgical responsibility. Thus infected or blocked feeding lines should be replaced, following antibiotic control of sepsis, at open operation retaining the same venous access site for as long as possible. It cannot be overstressed that loss of venous access is an avoidable iatrogenic complication of grave consequence for the child's future.

Simple anastomosis of the massively dilated proximal small bowel to the narrow defunctioned distal segment leads to poor isoperistaltic propulsion with a ‘to and fro' movement of intraluminal contents. Stasis and bacterial overgrowth lead to mucosal inflammation and heavy bacterial translocation both to the liver and systemically. Effective bowel-related surgery consists of a redistribution of available autologous bowel, or of instant provision of an adequate length of bowel by transplantation. Presently a lack of appropriate donors, major difficulties with immunosuppression, and a high risk of immunoproliferative disease make transplantation a final port of call for those in whom all alternatives have been exhausted and who face death from endstage liver failure or loss of venous access. The alternative to bowel transplanation is autologous gastrointestinal reconstruction which redistributes available autologous bowel. Table 1 outlines presently available options, of which longitudinal intestinal lengthening and tailoring has received increasing application. Longitudinal lengthening divides the dilated bowel along the mesenteric and antemesenteric borders, maintaining blood supply to each hemisegment which is then rolled into a loop of half diameter. The hemiloops are anastomosed isoperistaltically thus increasing bowel length (Fig. 6). Effective isoperistaltic propulsion reduces intraluminal stasis and bacterial translocation and longer mucosal contact over a greater bowel length enhances absorption and stimulates the intestinal adaptation process. Despite experimental evidence of disrupted peristalsis because of circular muscle division, the summation of peristaltic activity is isoperistaltic propulsion. Complications have been few, with occasional reports of small leaks from the extensive suture lines, stenosis at the hemiloop anastomosis, and loss of one hemiloop from poor blood supply. Most reports have indicated increased enteral absorption with a significant proportion of children achieving sole enteral nutrition. Longitudinal lengthening can be combined with other techniques such as single or multiple reversed segments. Multistage techniques (Table 1) allow reconstruction of greater bowel lengths. Undilated bowel can have the tissue expanded using the Georgeson semi-obstructive valve and can then be lengthened by a combination of the Kimura isolated bowel segment ‘Iowa' model and longitudinal lengthening to achieve better bowel redistribution and even greater length. With a similar purpose, composite bowel tubes constructed by grafting small bowel mucosa to peristaltic colonic or gastric muscle achieve further bowel length. Thus present-day techniques are complex but versatile and allow individualized therapy tailored to maximize autologous bowel potential. It is evident that for best results such children should be referred to specific units specializing in the management of the child with short bowel.

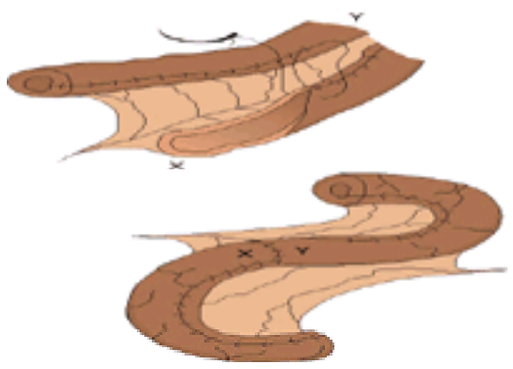


Fig. 6. Longitudinal intestinal lengthening and tailoring.

Prolonged transit time	Vagotomy and pyloroplasty
	Reversed segments—single or multiple
	Colonic transposition—iso- or uniperistaltic
Tailoring procedures	Plication
	Antemesenteric resection
Tailoring and lengthening	Longitudinal intestinal lengthening and tailoring (*LILT*)
	Combined techniques, e.g. LILT + reversed segments
	Composite bowel tubes
	Isolated bowel segment 'Iowa' models
	Bowel expansion and sequential intestinal lengthening
Neomucosal growth	Stem cell–submucosal graft complexes (experimental)

Table 1 Autologous gastrointestinal reconstruction

Timing of autologous gastrointestinal reconstruction

Death in neonatal short bowel syndrome is largely due to relentless progressive liver injury presenting as cholestatic jaundice of early onset with rapid progression to endstage liver failure over some 8 to 18 months. The condition is noticeably aggravated by frequent sepsis and bacterial translocation, which may be partly related to the reduced gut-associated lymphoid mass and to diminished immunocompetence. Imperfect parenteral nutrition, such as phytosterols in fat solutions, given over a prolonged period may aggravate the situation. The appropriate timing for autologous gastrointestinal reconstruction is still an open question. It is evident that self-selected survivors who have come through the hazardous neonatal phase with minimal liver injury will benefit most from resolution of their residual bowel-related problems. Indeed longitudinal intestinal lengthening and tailoring has proven most effective for children with more than 40 cm of dilated jejunum who also have an ileocaecal valve and all the colon. Even more serious, however, are those children with less than 40 cm of small bowel and a reduced length of colon. An 18-year audit of early bowel reconstruction to reduce stasis and bacterial translocation and to enhance bowel adaptation has shown this policy to be associated with approximately 50 per cent mortality from relentless hepatic dysfunction commencing in the neonatal phase. Present management concepts, therefore, place the major emphasis on liver management and avoidance of heptic toxicity by specific high-protein and reduced fat total parenteral nutrition regimens and bowel sterilization with enteral tobramycin (10 g four times daily with feeds). Meanwhile enteral ‘sham' feeding with normal or predigested foods allows brain learning for taste and swallowing, enhances mucosal adaptation, and allows bowel dilatation (tissue expansion). Early bowel surgery is avoided or at best limited, for example tube jejunostomy. Bowel reconstruction is only undertaken once liver dysfunction is controlled and liver function has returned to normal, usually at 4 to 6 months of age and more than 5 kg weight.

Long-term outlook

Neonatal short bowel syndrome is a serious condition of grave prognosis. Presently some 50 per cent of newborn children with an established short bowel will die largely from relentless progressive liver injury. Survivors may achieve full enteral nutrition. Others experience an improvement in quality of life because of improved absorption following autologous reconstruction, but still require partial parenteral support. Home parenteral nutrition that is cycled during the night or intermittently during the week allows a close to normal, good quality life often of lesser risk than after bowel transplantation. Enterally adapted long-term survivors are usually able to eat a normal unrestricted diet but require specific replacement of vitamin B_{12} and fat-soluble vitamins because of loss of appropriate binding sites. Some require a low oxalate diet because of increased colonic absorption and high urinary excretion with a risk of stone formation. The disrupted enterohepatic circulation of bile salts and consequent abnormal bile make long-term monitoring for gallstones mandatory. Initially high stool frequency reduces with bowel adaptation to two to four motions of semisolid consistency each day. Faecal continence is usually normal, even at night. In essence, experience has shown that children with a short bowel who achieve full intestinal adaptation will have an unrestricted, good quality, normal life. These children demonstrate unimpaired physical and mental growth and normal development.

The future

Improved immunosuppression regimens, greater donor availability, and a better understanding of chimerism and tolerance will make bowel transplantation a more attractive proposition. Advances in genetic engineering may result in autologous organ cloning. Further understanding of the part played by gut hormones in intestinal adaptation will clarify their therapeutic role within medical management of the short bowel syndrome. Combinations of presently available autologous gastrointestinal reconstructive techniques already create additional lengths of absorptive small bowel which is the patient's own and is not rejected. Further development of the, as yet experimental, autologous stem cell–submucosal graft complexes (Table 1) offers the enticing possibility of new absorptive small bowel mucosa, thus enhancing the potential for enteral independence and offering the patient with a short bowel a better chance for good quality, long-term survival. The last word in the management of the child with short bowel syndrome has yet to be written. In the meantime it remains a fruitful field for imaginative innovation.

Further reading

Barrena MJ *et al.* Lymphocyte subpopulations after extensive small bowel resection in the rat. *Journal of Pediatric Surgery* 1995; **30**: 1447–9. [Evidence for reduced gut-associated lymphoid tissue in the short bowel.]

Bianchi A. Autologous gastrointestinal reconstruction. *Seminars in Pediatric Surgery* 1995; **4**: 54–9. [Review of options in bowel reconstruction.]

Bianchi A. Longitudinal intestinal lengthening and tailoring: results in 20 children. *Journal of the Royal Society of Medicine* 1997; **90**: 429–32. [Long-term experience with longitudinal lengthening.]

Collins JB, Georgeson KE, Vincente Y, Kelly DR, Figueroa R. Short bowel syndrome. *Seminars in Pediatric Surgery* 1995; **4**: 60–73. [Informative and detailed review.]

Figueroa-Colon R, Harris PR, Birdson E, Franklin FA, Georgeson KE. Impact of intestinal lengthening on the nutritional outcome for children with short bowel syndrome. *Journal of Pediatric Surgery* 1996; **31**: 912–16. [Documents absorption following bowel lengthening.]

Iyer KR, Spitz L, Clayton P. New insight into mechanisms of parenteral nutrition-associated cholestasis: role of plant sterols. *Journal of Pediatric Surgery* 1998; **33**: 1–6. [Of major relevance to long-term total parenteral nutrition.]

Kelly DA, Buckels JAC. The future of small bowel transplantation. *Archives of Disease in Childhood* 1995; **72**: 447–51. [Reviews status of bowel transplantation.]

Moore TC. Necrotizing enterocolitis. In:*Challenges in Pediatric Surgery*, pp. 83–97. RG Landes, Austin, 1994. [Informative, innovative, different, and stimulating.]

Reyes J, Tzakis AG, Todo S, Nour B, Starzl T. Small bowel and liver small bowel transplantation in children. *Seminars in Pediatric Surgery* 1993; **2**: 289–300. [Experience from the major centre for bowel transplantation in children.]

Shanbhogue LKR, Molennar JC. Short bowel syndrome: metabolic and surgical management. *British Journal of Surgery* 1994; **81**: 486–99. [Useful additional details and critical surgical review.]

Uroa M, Teitelbaum DH, Drongowski RA, Coran AG. The association of gut-associated lymphoid tissue and bacterial translocation in the newborn rabbit. *Journal of Pediatric Surgery* 1996; **31**: 1482–7. [Development of normal immunocompetence.]

Vanderhoof JA, Languas AN, Pinch LW, Thompson JS, Kaufman SS. Short bowel syndrome: a review. *Journal of Pediatric Gastroenterology and Nutrition* 1992; **14**: 359–70. [Detailed review.]

24.7 Radiation enteritis

Christopher G. Willett and Marcelo Mester

Introduction
Acute complications
Long-term complications
Summary
Further reading

Introduction

Radiation therapy induces early and late effects in the gastrointestinal tract. In the pelvis, the small bowel is the most radiosensitive and, therefore, dose-limiting organ. Acute symptoms, consisting of diarrhea and abdominal cramps, occur in the majority of patients treated with a fractionated course of pelvic radiotherapy. These symptoms are generally transient and respond to medication. Late small bowel symptoms usually occur in the first year, but latency up to 7 years is reported. These complications are often irreversible and difficult to treat.

Small and large bowel complications (both acute and long-term) are closely related to radiation therapy techniques. Acute complications are correlated with volume of bowel irradiated and dose per fraction, while long-term complications are related not only to the volume of bowel treated and dose per fraction size but also to total dose of radiation administered. Limiting small bowel irradiation to 45 to 50.4 Gy in 1.8 Gy fractions and large bowel irradiation to 55 to 60 Gy in 1.8 Gy fractions, with the use of carefully designed fields and technique, should limit complications to less than 5 per cent of patients. Other reported predisposing factors to long-term radiation injury include obesity, diabetes mellitus, hypertension, pelvic inflammatory disease, prior abdominal or pelvic surgery, and the use of concurrent chemotherapy during radiation therapy. Acute and long-term complications occur with distinct clinical courses and pathologic manifestations.

Acute complications

Gastrointestinal symptoms are common during the course of radiation treatment to the pelvis and/or abdomen. Large bowel symptoms consist of an acute proctocolitis with diarrhea and tenesmus. Sigmoidoscopsy during treatment frequently shows an inflamed, edematous, and friable mucosa. Acute small bowel reactions may result in nausea, abdominal cramps, and watery diarrhea, primarily due to the depletion of actively dividing cells in what is otherwise a stable cell renewal system. In the small bowel, loss of the mucosal cells results in malabsorption of various substances including fat, carbohydrate, protein, and bile salts. In the large bowel, failure of fluid absorption results in watery diarrhea.

Although occasionally severe, these symptoms are usually self-limited and managed by the use of antispasmodic and anticholinergic medications. Some patients with symptoms refractory to these drugs may respond to a bile sequestrating agents (such as cholestyramine), although such reports are anecdotal.

Non-steroidal anti-inflammatory drugs have been used to prevent acute enteritis. In a double-blind placebo-controlled trial, buffered acetysalicylate was effective in improving gastrointestinal symptoms in patients who had undergone curative radiotherapy for malignant gynecologic disease. Three subsequent trials have not confirmed this finding and it is unlikely that anti-inflammatory drugs are useful in preventing or ameliorating acute radiation enteritis. A placebo-controlled study has shown a decrease in both early and late symptoms of radiation enteritis in patients who received oral sucralphate during pelvic radiotherapy, compared with controls. These symptoms refer to both diarrhea and proctitis, defecation incontinence, and loss of blood and mucus due to rectal mucosa damage. By forming a viscous coagulum on mucosal surfaces, sucralphate provides protection against bile and enzyme effects for several hours after ingestion. The prevalence and severity of radiation-induced diarrhea may also be reduced by using an elemental diet or total parenteral nutrition during therapy.

Long-term complications

Long-term complications include obstruction, perforation, bleeding, malabsorption, and fistulas. The latent period between completion of radiation therapy and the subsequent development of these complications is usually 6 to 24 months, but chronic radiation enteritis can occur many years after the original treatment. Colicky abdominal pain is the most common symptom, and generally precedes bloody diarrhea, tenesmus, steatorrhea, and weight loss. Less commonly, patients present with an acute obstruction or perforation of the small or large bowel.

Pathologic study of bowel affected by chronic radiation enteritis demonstrates obliterative endarteritis of the small vessels in the intestinal wall, submucosal fibrosis, and lymphatic dilatation. Perhaps the injury sustained by endothelial and connective tissue in the intestinal wall leads to progressive ischemia of the intestinal wall. Ischemia then leads to mucosal ulceration and necrosis of the intestinal wall with subsequent bleeding, perforation, stricture, and fistula formation, alone or in combination.

Radiation enteritis can also be progressive: in a series of 51 patients reported by Galland, only 24 patients remained symptom free after the first manifestation and treatment, whereas 20 patients developed second, third, or more gastrointestinal problems. Bleeding appeared to be a less ominous presentation: no new complications followed a bleeding episode, while 33 per cent of those with stricture and 89 per cent of those with perforation or fistula later manifested new lesions. Small bowel lesions occur most frequently in the ileum. Ileal injury by itself proved to be a poor prognostic factor in a retrospective series of 57 patients who required surgical intervention for any radiation-induced bowel injury. Forty-four per cent of patients with isolated ileal lesions died due to radiation enteritis, and patients with ileal lesions often had other lesions.

The medical and surgical treatment of chronic radiation enteritis is complex and requires individual management. Isolated damage to the rectum or rectosigmoid junction can often be managed with a low residue diet and steroid suppositories, but damage to the large intestine is frequently associated with significant injuries to the distal part of the small intestine. In the management of lesions of the small intestine, antispasmodics, anticholinergics, broad-spectrum antibiotics, cholestyramine, and salicylazosulphapyridine have been used, but responses to these agents have been anecdotal and based on a small number of patients. The use of a low fat diet, a low residue diet, or a gluten- and lactose-free diet with low fat as well as elemental diets have all been reported to be useful in selected patients with chronic radiation enteritis. A randomized study comparing total parenteral nutrition, with or without methylprednisone, with enteral nutrition for patients with severe chronic small bowel radiation enteritis found the former to be beneficial, with improvement in nutritional, radiologic, and clinical parameters. The efficacy of the administration of total parenteral nutrition appeared to be enhanced by use of methylprednisone.

Before definitive surgical treatment is embarked upon, the status of the original malignancy and the extent of radiation damage should be assessed. Sepsis, malnutrition, and biochemical abnormalities should be corrected. Although wide resection and primary anastomosis may be appropriate for patients with single, discrete areas of radiation-induced pathology manifested as obstruction, perforation, or bleeding, the majority of patients have more generalized radiation damage. Those who present with radiation-induced perforation of the small bowel should be treated as emergencies, with avoidance of primary anastomosis. Exteriorization of the involved segments may be the optimal procedure in this situation, even though a further operation will be necessary to reconstitute the alimentary tract. An enteric fistula can sometimes be managed conservatively but such management is rarely successful. Gastrointestinal bypass without extensive resection of the involved bowel has less mortality and a higher success rate. Surgery for any type of radiation-induced bowel injury is associated with a serious high complication rate (such as anastomotic leakage) and perioperative mortality. Furthermore, approximately 40 per cent of patients may need second or even third procedures for more injuries since vascular-induced injury progresses in time.

Summary

Abdominal and pelvic radiotherapy is increasingly used in the curative treatment of many abdominal and pelvic malignancies. Bowel tolerance is a major limitation to this treatment. Long-term complications occur less frequently than acute complications but are substantially more serious. Efforts must be made to minimize long-term bowel complications: these include the use of modern radiation treatment planning and techniques. Surgical measures to exclude the small bowel from the radiation field—absorbable mesh, ectopic implant, cystopexy, and omental flap—have shown encouraging results in reducing the amount of small bowel in the radiation field and thus reducing the incidence of small bowel injury.

Further reading

Baughan CA *et al.* A randomized trial to assess the efficacy of 5-aminosalicylic acid for the prevention of radiation enteritis. *Clinical Oncology* 1993; **5**: 19–24.

Covens A *et al.* The prognostic importance of site and type of radiaton-induced bowel injury in patients requiring surgical management. *Gynecologic Oncology* 1991; **43**: 270–4

Devereux DF, Eisenstat T, Zinkin L. The safe and effective use of postoperative radiation therapy in modified Astler-Coller stage C3 rectal cancer. *Cancer* 1989; **63**: 2393–6.

Galland RB, Spencer J. Natural history and surgical management of radiation enteritis. *British Journal of Surgery* 1987; **74**: 742–7.

Gunderson LL, Martenson JA. Gastrointestinal tract radiation tolerance. *Frontiers of Radiation Therapy and Oncology* 1989; **23**: 277–98.

Gunderson LL, Russell AH, Llewellyn HJ, Doppke KP, Tepper JE. Treatment planning for colorectal cancer: radiation and surgical techniques and value of small-bowel films. *International Journal of Radiation Oncology, Biology, Physics* 1985; **11**: 1379–93.

Henriksson R, Franszen L, Littbrand B. Effects of sucralfate on acute and late bowel discomfort following radiotherapy of pelvic cancer. *Journal of Clinical Oncology* 1992; **10**: 969–75.

Hoskins RB *et al.* Adjuvant postoperative radiotherapy in carcinoma of the rectum and rectosigmoid. *Cancer* 1985; **55**: 61–71.

Kinsella TJ, Bloomer WD. Tolerance of the intestine to radiation therapy. *Surgery, Gynecology and Obstetrics* 1980; **151**: 273–84.

Letschert JG. The prevention of radiation-induced small bowel complications. *European Journal of Cancer* 1995; **31A**: 1361–5.

Loiudice TA, Lang JA. Treatment of radiation enteritis. *Journal of Gastroenterology* 1983; **78**: 481–7.

Martenson JA *et al.* Olsalazine is contraindicated during pelvic radiaton therapy: result of a double blind randomized trial. *International Journal of Radiation Oncology, Biology, Physics* 1996; **35**: 299–303.

Mennie AT *et al.* Treatment of radiation-induced gastrointestinal distress with acetylsalicylate. *Lancet* 1975; **ii**: 942–3.

Minsky BD, Cohen AM. Minimizing the toxicity of pelvic radiation therapy in rectal cancer. *Oncology* 1988; **2**: 21–5.

Piver MS, Barlow JJ. High dose irradiation to biopsy confirmed aortic node metastases from carcinoma of the uterine cervix. *Cancer* 1977; **39**: 1243–6.

Resbeut M *et al.* A randomized double blind placebo controlled multicenter study of mesalazine for the prevention of acute radiation enteritis. *Radiotherapy and Oncology* 1997; **44**: 59–63.

Sher ME, Bauer J. Radiation-induced enteropathy. *Gastroenterology* 1990; **85**: 121–8.

Smith DH, DeCosse JJ. Radiation damage to the small intestine. *World Journal of Surgery* 1986; **10**: 189–94.

Taylor Whartor J, Facog WJ III, Facog TGD, Facog FNR, Fletcher GH. Preirradiation cellotomy and extended field irradiation for invasive carcinoma of the cervix. *Obstetrics and Gynecology* 1976; **49**: 333–8.

Vigliotti A, Richard TA, Romsdahl MM, Withers HR, Oswald MJ. Postoperative adjuvant radiotherapy for adenocarcinoma of the rectum and rectosigmoid. *International Journal of Radiation Oncology, Biology, Physics* 1986; **13**: 999–1006.

Yeoh EK, Horowitz M. Radiation enteritis. *Surgery, Gynecology and Obstetrics* 1987; **165**: 373–9.

24.8 Small-bowel tumours

T. R. Worthington and R. C. N. Williamson

General considerations
Presentation
Investigation
Pathological classification
Benign neoplasms
Intermediate neoplasms
Malignant neoplasms
Non-neoplastic lesions
Further reading

General considerations

Considering that the small intestine comprises 80 to 90 per cent of the luminal surface of the entire alimentary canal, it is remarkable that enteric neoplasms account for only 1 to 2 per cent of alimentary tumours. Specifically, carcinoma is at least 100 times less frequent than in the stomach, oesophagus, or colorectum.

The relative sterility and fluidity of chyme, its short transit time, and enzymatic detoxification of ingested carcinogens are all suggested as explanations for the rarity of cancer of the small intestine. These factors do not, however, account for the strikingly uneven distribution of adenocarcinoma within the small bowel, with a proximal concentration that decreases towards the terminal ileum. Perhaps this apparent resistance in the ileum might reflect a local protective mechanism, possibly immunological in nature and related to the high IgA content of the mucosa. In addition, different molecular mechanisms could be involved in the carcinogenic processes of the various tumours.

Presentation

Tumours of the small bowel predominantly present in the sixth and seventh decades of life. In contrast to malignant lesions, which almost all become symptomatic with time, over 50 per cent of benign tumours are found incidentally at laparotomy or autopsy. The mean time from the onset of symptoms to diagnosis is 6 to 9 months, and this delay may be due to several factors: vague symptoms, absence of clinical signs, the difficulty in investigating much of the small bowel, and the rarity of these tumours. Symptoms vary from non-specific complaints, such as nausea, dyspepsia, epigastric discomfort, fatigue, bloating and weight loss, to haemorrhage or obstruction.

Obstruction, although often partial and recurrent, can be sudden and complete. Intussusception is a frequent cause of obstruction, with the tumour acting as the apex of the intussusceptum. In other instances, obstruction may be due to a mass effect of the lesion or to volvulus. Rarely, stenosis of the bowel occurs, due to the local effects of vasoactive peptides secreted by a carcinoid tumour.

Haemorrhage is frequently occult and causes anaemia, but major bleeding can lead to both haematemesis and melaena. The small bowel must therefore be thoroughly investigated in patients whose cause of gastrointestinal bleeding remains undetermined. Occasionally, mucocutaneous pigmentation (in Peutz–Jegher syndrome), *café-au-lait* spots (neurofibromatosis), or telangiectasia (Osler–Rendu–Weber syndrome) will alert the clinician to the presence of small-bowel lesions.

Tumours in the periampullary region may present with obstructive jaundice or acute pancreatitis. Other unusual presentations include a palpable abdominal mass, perforation, fistula formation, or intraperitoneal haemorrhage. Tumour products can give rise to systemic complications, for example gastrin causing peptic ulceration and vasoactive kinins causing the carcinoid syndrome.

Investigation

Detection of duodenal lesions is more straightforward than for those in the jejunum and ileum. Upper gastrointestinal endoscopy can readily detect tumours in the first and second part of the duodenum, and it allows biopsy of tissue. Beyond the papilla, enteroscopy can be carried out by introducing a 5-mm endoscope transnasally and allowing peristalsis to carry the instrument into the small intestine, but these examinations tend to be time-consuming and often incomplete. Access can also be achieved by using a percutaneous fistula to enter the stomach or caecum.

Radiological contrast studies are invaluable when done carefully. Hypotonic duodenography can identify small mucosal lesions, and the jejunum and ileum can be visualized by a small-bowel enema.

In patients with active haemorrhage, visceral angiography is useful in locating lesions when blood loss is as low as 0.5 ml/min. Isotope scanning techniques are said to be sensitive for bleeding rates as low as 0.1 ml/min. Although accurate, these tests should not delay laparotomy, as most actively bleeding lesions are readily seen or felt. Occasionally the source of bleeding is not found during inspection of the bowel and additional procedures may be needed. Cross-clamping of the bowel at various levels may localize the source, and, lastly, operative enteroscopy with transillumination of the bowel wall may outline more obscure lesions (Fig. 1).



Fig. 1. On-table enteroscopy: transillumination of the small-bowel wall shows a small angiodysplastic focus that had been causing intermittent haemorrhage.

In addition to its use in locating actively bleeding lesions, angiography can characterize tumour blood supply, portal venous compression or occlusion, aberrant arterial anatomy, and actual tumour type, since lesions such as neuroendocrine tumours are characteristically highly vascular.

Increasingly, computed tomography (**CT**) and magnetic resonance imaging are yielding important information on tumour site and size, and associated metastatic disease. Other techniques undergoing assessment include both endoscopic and laparoscopic ultrasonography, which in experienced hands can give both specific and sensitive information on tumour site and size.

Pathological classification

The classification of tumours of the small bowel is not straightforward and thus it is not surprising that the subject is contentious. Firstly, whether some lesions should be classified as benign tumours as opposed to hamartomas has given rise to debate. A hamartoma is defined as ‘an excessive focal overgrowth of mature normal cells and tissues in an organ composed of identical cellular elements'. Some authorities consider that Brunner's gland adenomas, certain neurogenic tumours, haemangiomas, and lymphangiomas are all hamartomas, whereas others classify them as benign tumours. In this chapter, these lesions are considered as benign tumours. Secondly, the malignant nature of some lesions is controversial and therefore the concept of ‘intermediate neoplasms' has arisen for tumours that have a

definite malignant potential but can be difficult to classify histopathologically. These tumours will be considered separately.

Benign neoplasms

Tubular adenoma

Tubular adenomas usually present as polypoid lesions, whereas the villous variety has a sessile appearance. Some adenomas have features of both types and can be termed tubulovillous, akin to those in the large bowel. Tubular adenomas are relatively common in the duodenum but less so in the distal small bowel. They are usually single but can be multiple. Symptoms depend not only on the size but also on the position of the lesion. Abdominal pain, obstruction, and occult (or overt) haemorrhage are possible presentations. Obstructive jaundice may occur with tumours near the ampulla, but it is often absent or minimal. Malignant change does occur and mimics the adenoma–carcinoma sequence seen in the colon. The incidence of malignant change increases with the size, site, and number of lesions as well as with the histological type. In one series, the mean diameter of benign polyps was 2.65 cm compared with 3.7 cm for those containing both benign and malignant elements. Interestingly, tumours situated in the periampullary region are more prone to malignancy, as are multicentric or villous lesions.

To obtain an accurate histological diagnosis and remove the risk of malignant change, tumours must be excised in their entirety. The preferred treatment for jejunal and ileal polyps is unclear: either local enterotomy with excision or more formal segmental resection. Smaller duodenal polyps (less than 2 cm) may be removed endoscopically (Fig. 2), but larger lesions require formal laparotomy and duodenotomy.

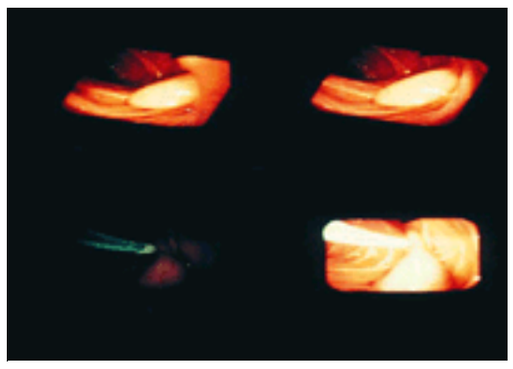


Fig. 2. Endoscopic snaring of a tubular adenoma in the duodenum (by courtesy of Mr J. Spencer).

Villous adenoma

Villous adenomas again tend to arise in the duodenum (Fig. 3), although they are larger (mean diameter 4 cm) than tubular adenomas (1–3 cm). Villous tumours have a strong propensity for malignant change, carcinoma being reported in 35 to 63 per cent of cases (Fig. 4). Management of these tumours therefore requires accurate detection and assessment of malignant potential. Preoperative assessment may be useful: CT scanning may indicate hepatic metastases or local extension into surrounding tissues, and endoscopic biopsies may show evidence of dysplasia or frank malignancy. As these tumours tend to be large, small areas of carcinomatous change are often overlooked so that there is a strong risk of a sampling error: endoscopic biopsy has a false-negative rate of 25 to 65 per cent. In view of this fact, a definitive diagnosis can often only be made after laparotomy and examination of the entire lesion.

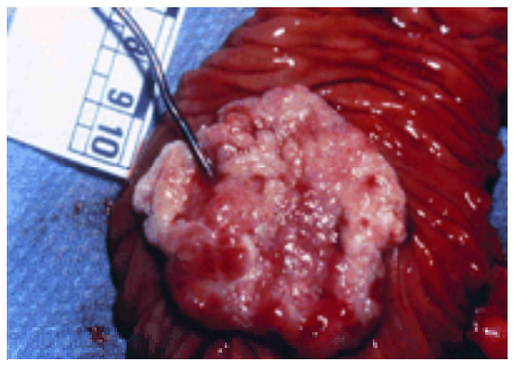


Fig. 3. Villous adenoma of the duodenum: the patient presented with haemorrhage; the lacrimal probe illustrates that the papilla is patent despite the size of the tumour.

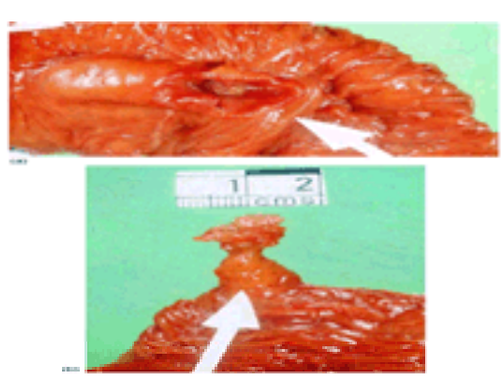


Fig. 4. Villous adenoma and carcinoma of the duodenum: (a) ulcerating adenocarcinoma of the duodenum (arrowed) arising at the papilla in a 34-year-old man who presented with obstructive jaundice; (b) origin of the carcinoma from a villous adenoma is strongly suggested by the fact that a concomitant (benign) villous adenoma was found in the third part of the duodenum in the resection specimen.

At laparotomy for duodenal lesions, the tumour and possible sites of metastatic disease should be thoroughly assessed. Frozen-section biopsy may indicate malignant change, in which case radical resection in the form of pancreatoduodenectomy should be considered. If carcinomatous change is not evident, duodenotomy can be performed to permit direct inspection of the tumour and submucosal resection. Local excision of the ampulla with reimplantation of the ducts or sphincteroplasty has been described for a lesion that lies close to the ampulla. Since a local recurrence rate of 46 per cent has been reported with this procedure, however, it has been suggested that radical resection should be undertaken in fit patients.

In unfit individuals, alternative procedures to be considered include endoscopic removal or laser ablation of tumour tissue, but histological examination is difficult thereafter. In tumours encroaching on the ampulla, a stent may be inserted to avert biliary or pancreatic obstruction.

The follow-up of these patients should be determined by the final histological diagnosis and the extent of resection. Locally excised lesions must be followed up by regular upper-gastrointestinal endoscopy and biopsy.

Polyposis syndromes

In familial polyposis coli, autosomal-dominant inheritance of the mutated *APC* (adenomatous polyposis coli) gene on chromosome 5q21 typically results in thousands of adenomas in the colorectum. Adenocarcinoma usually develops in only a few of these adenomas, typically within the left colon. During the past 20 years it has become increasingly clear that, in patients carrying the mutant gene, polyps can occur throughout the gastrointestinal tract; hence the term familial adenomatous polyposis has been introduced to describe the condition. Polyps can occur in the stomach, these tending to be hyperplastic, and throughout the small bowel.

Adenomatous polyps occur within the small intestine in 24 to 93 per cent of patients with familial adenomatous polyposis. They are usually multiple and clustered within the duodenum in the periampullary region. The many reports of periampullary carcinoma developing in familial adenomatous polyposis suggest that a proportion of these polyps will become malignant. Although most patients with familial adenomatous polyposis will develop duodenal adenomas, only 2 to 12 per cent develop duodenal cancer, indicating that the natural history of these tumours is unclear. A recent report indicated no progression of duodenal lesions in 14 of 17 patients over an observation period of 7 years, yet in one other patient a 1.7-cm polyp was overlooked and 22 months later a carcinoma was found. Other studies have advocated regular endoscopic evaluation together with biopsy of the periampullary region as microscopic adenomas can occur in the ‘normal' duodenal mucosa. In addition to this, it has been suggested that polyps greater than 0.5 cm in diameter should be removed and smaller polyps be kept under review or destroyed by endoscopic ablation.

The optimal approach to jejunal and ileal tumours in familial adenomatous polyposis is also unclear. It would seem reasonable to excise those lesions with an increased risk of malignant change, such as large or villous tumours.

Brunner's gland adenoma

These rare lesions are restricted to the duodenum and account for between 10 to 30 per cent of benign duodenal tumours. They are almost always benign. Three forms have been described: polypoid, and isolated or diffuse nodular hyperplasia. Adenomas usually arise proximal to the papilla and are less than 1 cm in diameter; lesions as large as 12 cm in diameter have been described, however. Tumours are generally found incidentally on barium-meal examination or endoscopy, but they may present with symptoms varying from dyspepsia and nausea to diarrhoea, jaundice, obstruction, and haemorrhage. As they are easily accessible, smaller lesions may be removed endoscopically, but larger lesions require surgical operation. Concomitant microcarcinoids have been reported in some resected specimens.

Both forms of nodular hyperplasia, isolated and diffuse, have sessile projections on the surface. The isolated variety is restricted to the first part of the duodenum, whereas the diffuse form can involve the whole duodenum. Although the underlying reasons are unclear, the incidence of nodular hyperplasia is higher in patients with chronic renal failure. These lesions tend to be asymptomatic, and if endoscopy and biopsy can confirm the diagnosis then no treatment is needed.

Lipoma

The incidence of lipomas is probably somewhat underestimated as these tumours seldom become large enough to cause symptoms that warrant surgical intervention. Intussusception or (rarely) bleeding from surface ulceration can supervene, however. Some reports indicate that lipomas are more common in the ileum than in the duodenum or jejunum.

Lipomas almost certainly have no malignant potential, although a solitary case of ileal liposarcoma has been described. Symptomatic lesions should be removed endoscopically or surgically. There is also a case for removing asymptomatic duodenal lipomas because of their potential for causing obstruction, obstructive jaundice, and haemorrhage.

Neurogenic tumours

These tumours are all derived from nervous tissue, either the nerve sheath, paraganglia, or ganglia. Neurofibromas are the most common form. Gangliocytic paragangliomas, paragangliomas, and ganglioneuromas are other rarer variants.

Neurofibromas can occur alone or in association with neurofibromatosis (Fig. 5); the gastrointestinal tract is involved in 12 to 25 per cent of patients with neurofibromatosis. These patients have tumours, sometimes several in number, that are primarily located in the jejunum. Characteristically the tumours appear along the antimesenteric border of the bowel and are smooth, firm, and round, ranging in size from 1 to 20 cm. They can present with pain, haemorrhage, intussusception, or volvulus. Histologically, most neurofibromas are benign, but 10 to 15 per cent show sarcomatous degeneration. Fortunately, spread to lymph nodes is uncommon; treatment is therefore by segmental resection.

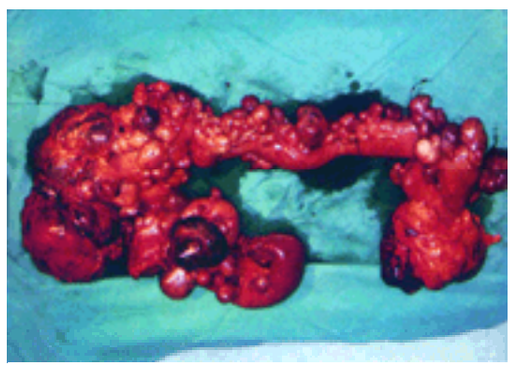


Fig. 5. Multiple neurofibromas of the small bowel. A 65-year-old man presented with gastrointestinal haemorrhage and peritonitis On clinical examination, cutaneous neurofibromas and *café-au-lait* lesions were evident. At laparotomy, multiple tumours of the jejunum were present and resection was undertaken (by courtesy of Mr A. Isla).

Gangliocytic paragangliomas preferentially arise in the second part of the duodenum (Fig. 6) but can occur in the ileum and jejunum. They appear as a polypoid mass. In a review of 39 cases, the mean age of presentation was 52 years; symptoms included haemorrhage (62 per cent) and abdominal pain (21 per cent), but a few tumours were asymptomatic (13 per cent). Paragangliomas and ganglioneuromas are rare.

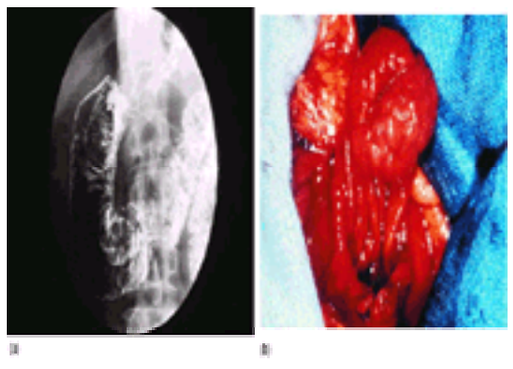


Fig. 6. Duodenal paraganglioma: (a) barium-meal examination undertaken for mild dyspepsia in a man of 64 years showed a tumour in the third part of the duodenum; (b) at operation a circumscribed lesion was found at this site and locally excised.

Fibroma

These rare lesions arise in the wall of the small intestine and are usually asymptomatic, although they may present with intussusception and obstruction. In the past, some may have been misdiagnosed as smooth-muscle or neurogenic tumours. Treatment is by local excision or segmental resection.

Vascular tumours

These comprise less than 10 per cent of small-bowel tumours. Three types are described: capillary haemangiomas (Fig. 7), cavernous haemangiomas (Fig. 8), and multiple telangiectasia. Cavernous hae-mangiomas tend to occur singly and can affect patients from the second decade onwards, while capillary haemangiomas may be multiple and arise in infancy. The most common presentation of capillary and cavernous haemangiomas is haemorrhage, but large tumours may cause obstruction. Multiple telangiectasia, which may be sporadic or hereditary (Osler–Weber–Rendu disease), is characterized by numerous haemangiomas, 1 to 5 mm in diameter, occurring in the ileum or less commonly in the jejunum. This condition is usually symptomless. Regardless of type, these tumours all arise in the submucosa. Malignant change is extremely rare.

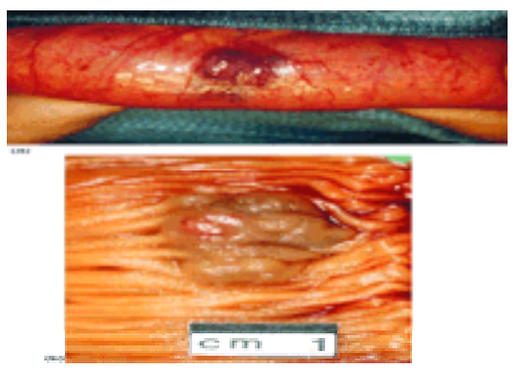


Fig. 7. Capillary haemangioma of the small bowel in a woman of 62 years who had persistent anaemia from occult gastrointestinal haemangioma:(a) serosal aspect of the tumour; (b) mucosal aspect showing ulceration over the lesion

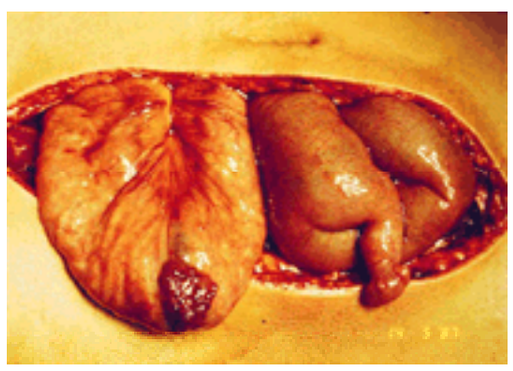


Fig. 8. Cavernous haemangioma of the small bowel; note coincident Meckel's diverticulum (by courtesy of Mr J. Spencer).

Angiography is useful in identifying vascular tumours, especially if the lesion is actively bleeding. Additionally, laparotomy with transillumination of the bowel lumen can demonstrate small lesions, but intraluminal blood can be of some hindrance. Intraoperative endoscopy may also be helpful. Single lesions are best treated by simple excision, but with multiple lesions affecting large segments of bowel a conservative approach is adopted, only actively bleeding tumours being excised.

In addition to these vascular neoplasms, various other lesions should be considered in the differential diagnosis of gastrointestinal bleeding and these have variously been termed arteriovenous malformations, angiodysplasia, and ectasia. For the purpose of simplicity these will be termed angiodysplasias. A system has been proposed classifying these lesions into three main types: type 1, lesions situated in the right side of the colon in patients over 55 years of age; type 2, those situated in the small intestine in patients under 55 years of age; and type 3, those associated with Osler–Weber–Rendu syndrome.

Again, angiography is useful in making a diagnosis. Findings such as rapidly filling afferent arterioles, berry-like vascular tufts, and the early filling of engorged veins are highly suggestive. Intraoperative endoscopy and transillumination of the bowel wall may also localize abnormal areas. A histological diagnosis of angiodyplasia cannot easily be made by microscopic examination of a biopsied specimen, or even of the resected intestine. Additional techniques have to be employed, therefore. Postoperative radiographs of the resected specimen that contains vessels into which contrast material is injected have been useful in confirming the abnormality, and a water-soluble contrast medium and a silicone rubber dye can be injected to aid in the macroscopic and microscopic identification of these lesions.

Intermediate neoplasms

Carcinoid tumours

Carcinoid tumours (Fig. 9) originate in enterochromaffin cells, which form part of a system of neuroendocrine cells scattered throughout the body. These cells are located in the mucous membrane of the gut near the base of crypts. The overall incidence of abdominal carcinoids has been estimated as 0.7 per 100 000 of the population. These tumours may occur in the foregut (including the duodenum), the midgut (including the jejunoileum), and the hind gut. Midgut carcinoids characteristically secrete large amounts of 5-hydroxytryptamine (**5-HT**; serotonin), whereas foregut carcinoids secrete small amounts of this peptide. In addition to serotonin, tumour metabolites include histamine, substance P, and neuropeptide K. Urinary excretion of 5-hydroxyindoleacetic acid, a breakdown product of 5-HT, is used as a diagnostic test.



Fig. 9. Ileal carcinoid tumour causing luminal stenosis: multiple carcinoids were found with nodal metastases, and side-to-side bypass of the small bowel has been performed (by courtesy of Mr J. Spencer).

Carcinoid tumours occur most commonly in the appendix, followed by the jejunum, rectum, and duodenum. Duodenal tumours are concentrated in the first and second parts, whereas ileal lesions tend to arise near the ileocaecal valve. Multicentric tumours are seen in about one-third of patients. Macroscopically, tumours are yellow in colour and appear in a submucosal or serosal position. They tend to spread outwards but can also ulcerate. Jejunoileal lesions may cause fibrosis and stenosis as a result of local release of vasoactive peptides, and this can lead to obstruction, which is often subacute.

Carcinoid tumours of the small intestine are slow-growing tumours that may be present for years without giving rise to symptoms. Presentation is usually during the fifth decade, and both sexes are affected equally. Some 70 to 80 per cent per cent of jejunal and ileal tumours are asymptomatic, and the symptoms that do arise are often non-specific; obstruction and bleeding can occur, however. Duodenal tumours may cause ulceration, obstruction, and jaundice.

All carcinoid tumours have a malignant potential, but the risk of developing metastases depends largely on the site of the primary. Thus liver metastases occur in 35 per cent of small-bowel carcinoids but in only 2 per cent of carcinoids of the appendix. Size is also a determinant of prognosis: once primary tumours exceed 2 cm in diameter, there is an 80 to 90 per cent chance of metastasis developing. Wide local excision of the primary tumour is therefore preferable. Obviously the site of the tumour influences the type of resection. Lesions in the jejunum and ileum can be managed by wide local excision, including the draining mesenteric lymph nodes. Terminal ileal tumours may require a right hemicolectomy. Smaller duodenal neoplasms may be dealt with by local excision, whereas larger lesions necessitate a pancreatoduodenectomy.

The carcinoid syndrome, characterized by flushing, diarrhoea and wheezing, occurs when tumour metabolites are secreted directly into the systemic circulation from hepatic metastases or a primary lesion in the bronchus or ovary. Not all hepatic metastases will produce the carcinoid syndrome, however. While approximately one-third of patients with small-bowel carcinoids will develop liver metastases, only a quarter of them will develop the carcinoid syndrome.

Hepatic metastases may be amenable to either formal liver resection or surgical debulking by means of multiple enucleations. Liver transplantation has been carried out in a number of young patients with disease confined to the liver. Unfortunately, tumour recurrence is frequent, probably due to the associated long-term immunosuppression. Less invasive methods such as selective hepatic arterial embolization can cause tumour necrosis and provide symptomatic relief. Although experience is limited, cryotherapy can reduce tumour volume (as assessed by CT scan) and alleviate the patient's symptoms. Pharmacological strategies for the treatment of the carcinoid syndrome involve the use of the somatostatin analogue octreotide, an antagonist of tumour metabolites. Ketanserin, a specific 5-HT antagonist, is especially effective in reducing diarrhoea and flushing. Tumour response to cytotoxic chemotherapy has proved to be disappointing, though interferon-a can reduce peptide secretion for up to 2 years.

Patients with local lesions have a 5-year survival rate of 75 to 100 per cent after resection, compared to 20 per cent for those with hepatic metastases.

Gastrinoma

Until recently the vast majority of gastrinomas were thought to arise within the pancreas, but a detailed study of failed operations for Zollinger–Ellison syndrome has highlighted the fact that small gastrinomas may occur in the duodenum. Duodenal gastrinomas may account for up to 40 per cent of all gastrinomas. They can occur in association with multiple endocrine neoplasia syndrome type 1, and they themselves tend to be multiple. Lesions, often less than 1 cm in size, may only be detected by duodenotomy and eversion of the mucosa to allow adequate visualization and palpation. Approximately 50 per cent of duodenal gastrinomas are associated with lymphnode metastases, which may therefore necessitate a pancreatoduodenectomy.

Gut stromal tumours

These include a group of intramural intestinal tumours formerly known as leiomyoma and leiomyosarcoma because their light-microscopic appearance mimicked smooth muscle. It has emerged from recent pathological research that these tumours may not arise exclusively from smooth muscle and hence the change in nomenclature. The tumour may originate from smooth muscle, nerve, mesenchyme, or just the stroma. The most common sites involved are the jejunum and ileum. Tumours arise from either the muscularis propria or the muscularis mucosae. They expand towards the bowel lumen, the serosa, or in both directions (dumb-bell shaped). Intraluminal lesions (Fig. 10) usually present with haemorrhage secondary to mucosal ulceration, but obstruction can arise from intussusception or mass effect. Extraluminal lesions (Fig. 11) can become enormous and undergo central necrosis with a risk of haemorrhage. Larger lesions may be palpable. Other unusual presentations include volvulus, perforation, and fistula formation.

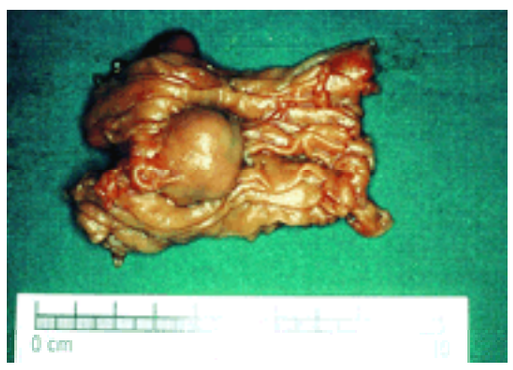


Fig. 10. Intraluminal stromal tumour of the small bowel treated by local excision (mucosal aspect).

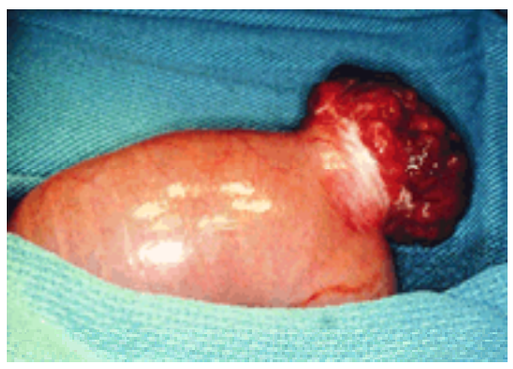


Fig. 11. Large extraluminal stromal tumour of the jejunum.

The histopathological distinction between benign and malignant (Fig. 12) stromal tumours is difficult, and the classification of malignancy is often unreliable. The strongest pathological predictors of malignant behaviour are size (generally larger than 5 cm) and mitotic count (generally more than five mitoses per 10 high-power fields). Small lesions with an unremarkable, benign appearance may behave in malignant fashion, however. In one recent series, tumour recurred locally in four of 34 patients, although two of these lesions were originally classified as histologically benign.

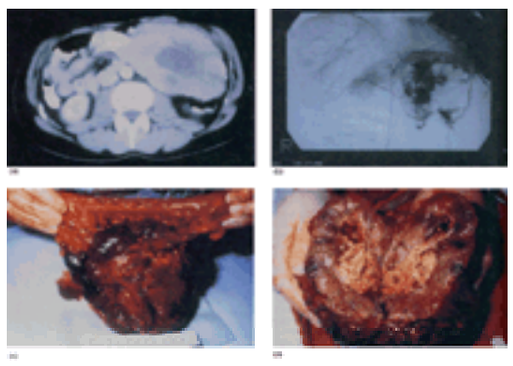


Fig. 12. Malignant stromal tumour of the fourth part of the duodenum. A 27-year-old man presented with melaena and a palpable abdominal mass. (a) CT scan showed a large tumour arising in close proximity to the tail of the pancreas. (b) Visceral angiography clearly demonstrates the neovascularization associated with the tumour. (c) At operation, a large tumour arising from a narrow pedicle on the fourth part of the duodenum was invading the transverse mesocolon; thus a small-bowel and colonic resection were carried out. (d) The cut surface of the tumour showed central necrosis.

Metastases are most commonly found in the liver and peritoneum, implying haematogenous and local spread. Since tumour is seldom found in regional lymph nodes, there does not appear to be a need for regional nodal dissection. These tumours are usually resistant to radiotherapy and chemotherapy, and treatment should therefore involve adequate excision to provide a negative histological margin. Some suggest that patients with very large tumours (over 5 cm) should have lifelong follow-up, including serial CT scans, due to the risk of recurrence or metastasis.

Malignant neoplasms

Primary disease

The most common malignant tumour of the small bowel is carcinoid (see above), followed by carcinoma (Fig. 13 and Fig. 14), lymphoma, and malignant stromal tumours (see above).

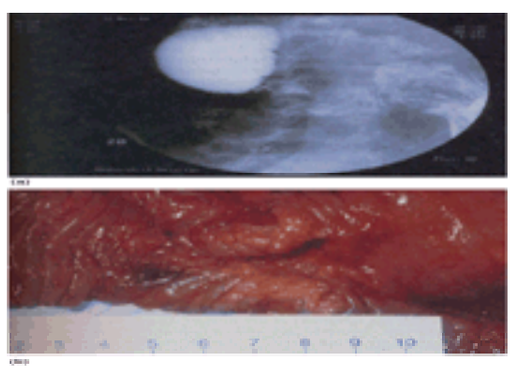


Fig. 13. Obstructing carcinoma of the duodenum. (a) A barium-meal examination was performed on a 73-year-old woman who presented with gastric-outlet obstruction; a concentric stenosis of the third part of the duodenum is illustrated. (b) The resection specimen with the stenosing carcinoma.

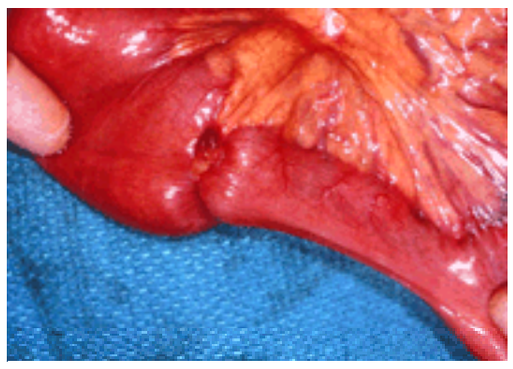


Fig. 14. A resection specimen demonstrates an obstructing carcinoma of the jejunum in a 80-year-old woman.

Adenocarcinoma

There is a markedly uneven distribution of adenocarcinoma throughout the small bowel, with a relative excess proximally that diminishes towards the terminal ileum. The vast majority of malignant duodenal tumours are adenocarcinomas (between 80 and 90 per cent). Within the duodenum itself, tumours tend to be situated in the periampullary region; as in the large intestine, adenomas have a particular predisposition to malignant change. Indeed, the pathogenesis may resemble that of colorectal carcinoma, which is thought to be a multistep process that takes place over a prolonged period of time. This sequence probably involves a small neoplastic polyp that enlarges, through a dysplastic stage, into an invasive carcinoma. The molecular abnormalities underlying this progression are thought to involve the activation of oncogenes and the inactivation of tumour-suppressor genes. Specifically, a high frequency of K-*ras* mutations has been found in human colorectal carcinoma. In similar investigations of the genetic abnormalities found in small-bowel cancer, one study has shown an increased prevalence of K-*ras* mutations in carcinomas of the duodenum, but not of the jejunum or ileum; the researchers speculated that different molecular mechanisms might be involved in the carcinogenesis of the various tumours, and this phenomenom could account for the disparity seen in tumour incidence throughout the small bowel.

Macroscopically, tumour appearance ranges from ulcerative and infiltrative (the predominant type) to polypoid. Multiple lesions are present in 15 to 25 per cent of cases, and the adjacent duodenal mucosa may show evidence of dysplasia. Microscopically, virtually all tumours appear to be moderately differentiated, mucin-secreting adenocarcinomas. Few tumours are diagnosed at an early stage, over 50 per cent of patients having metastases at the time of presentation. Lymph nodes are generally involved, but transperitoneal spread and hepatic metastases can also occur.

Clinical symptoms arising from carcinoma of the duodenum are many and varied: haemorrhage and anaemia, abdominal pain, nausea and vomiting, weight loss, acute pancreatitis, duodenal obstruction (Fig. 13), and jaundice. Many patients present with non-specific symptoms, however. A personal series of 22 duodenal carcinomas showed an equal sex incidence, with a median age at presentation of 60 years. Many patients reported vague symptoms of pain, anaemia, and intermittent vomiting, but 12 were jaundiced. A pre-existing villous adenoma was seen in 11 patients and adjacent duodenal dysplasia in 13, seemingly supporting the proposed adenoma–carcinoma sequence. The site of origin was mostly the second part of the duodenum, but the third and fourth parts were also involved; the first part was spared. In this group, 17 patients underwent curative resection with one hospital death at 25 days; the 5-year survival rate thereafter was 40 per cent.

In general, periampullary lesions require pancreatoduodenectomy, but small lesions in the fourth part of the duodenum may be amenable to duodenal resection with duodenojejunostomy. Irresectable lesions should be treated by bypass, with gastroenterostomy and biliary bypass if necessary.

Unlike the duodenum, the jejunum and ileum have approximately equal numbers of carcinomas, carcinoid tumours, and lymphomas. The mean age of presentation is during the seventh decade. Carcinoma of the jejunum (Fig. 14) may present with subacute intestinal obstruction and/or chronic intestinal blood loss, as the tumours

tend to be annular and ulcerating. Perforation and peritonitis are uncommon complications. Jejunal and ileal tumours are best treated by segmental resection including the regional lymph nodes. The overall 5-year survival rate for jejunoileal carcinomas is 20 to 30 per cent. If there is no nodal involvement at operation, survival is increased to 50 to 70 per cent. Adjuvant therapy for adenocarcinoma of the small bowel has been of little value; chemotherapy has shown no appreciable effect, and these tumours are also resistant to radiotherapy.

It is of interest that most carcinomas of the ileum occur in association with Crohn's disease and these cancers can be regarded as a distinct clinical entity. Crohn's disease of the small intestine is associated with malignant change in 0.3 to 0.6 per cent of cases, and tumours arise exclusively in segments of intestine that show evidence of long-standing Crohn's disease. The mean interval between diagnosing regional enteritis and enteric carcinoma is 18 years. The ileum is involved in 70 per cent and the jejunum in 30 per cent, while this ratio is reversed in those with ‘ordinary' carcinoma. Crohn's carcinoma occurs some 20 years earlier than usual, and there is a male preponderance of about 70 per cent. Histologically there is a pattern of diffuse and extensive carcinomatous infiltration, with adjacent areas of epithelial dysplasia. Symptoms may be confused with the underlying disease, and cancers affecting bypassed loops (30 per cent) are particularly difficult to diagnose. Thus prognosis is very poor, the mean survival being 8 months from diagnosis.

Enteric carcinoma is occasionally associated with neurofibromatosis, Lynch syndrome 2 (a variant of hereditary non-polyposis colorectal cancer), and radiation enteritis.

Gastrointestinal lymphoma

Primary gastrointestinal lymphoma, first described by Theodor Billroth in 1871, is a rare malignancy that accounts for 1 to 4 per cent of all primary gastrointestinal cancers. Within the gastrointestinal tract, 50 to 55 per cent of tumours occur in the stomach (Fig. 15), 30 to 32 per cent in the small bowel, and 8 to 18 per cent in the large intestine. Although the gastrointestinal tract is the most common site of extranodal lymphoma, secondary involvement of the small bowel is 10 times more common than primary disease. It is important therefore to distinguish primary gastrointestinal lymphoma from the secondary manifestations of nodal lymphoma. Patients must be thoroughly assessed before the diagnosis can be entertained: the presence of regional and mediastinal lymphadenopathy (by means of CT scan), bone-marrow involvement (by means of a normal blood film and bone-marrow aspirate), and hepatic and splenic involvement (except by direct invasion from the bowel), must all be excluded before the diagnosis is made.

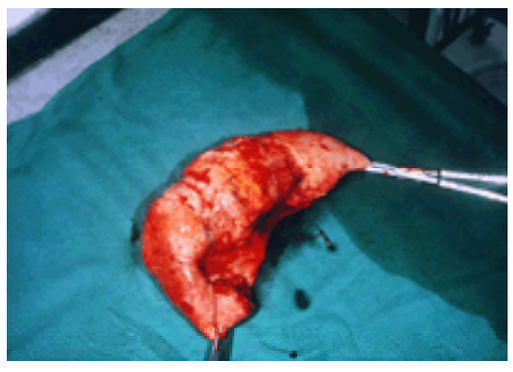


Fig. 15. A 60-year-old woman presented with peritonitis; at laparotomy a perforated lymphoma of the small bowel was discovered, and resection undertaken.

Many systems have been used to classify these tumours, but in short there are three main types:

1. The ‘Western'-type lymphoma, seen in Europe and North America, in which a discrete lesion arises from small intestine that was previously free of mucosal disease.
2. Primary lymphoma associated with either the villous atrophy of coeliac disease or lymphoid nodular hyperplasia.
3. The ‘Mediterranean' type of lymphoma, which arises in association with a-chain disease as part of the entity of immunoproliferative small-intestinal disease.

Western lymphoma This tumour tends to occur either in those under 10 or over 50 years of age. There is a slight male preponderance. Risk factors include immunosuppressive therapy (steroids, azathioprine, and cyclosporin), Crohn's disease, autoimmune disorders, and acquired immune deficiency syndrome (**AIDS**).

The lesions of Western lymphoma are generally single but sometimes multiple (8–20 per cent). They occur throughout the small bowel in adults, but tend to be clustered in the ileocaecal region in children. Macroscopically, lesions are generally localized and annular, but a more diffuse, infiltrative form is also recognized. Spread occurs early to regional lymph nodes and later to the spleen, liver, and other viscera. Histologically the tumours are B-cell lymphomas, mainly centrocytic or centroblastic in type.

Typical symptoms include abdominal pain, nausea and vomiting, weight loss, and anorexia. An abdominal mass may be present in 30 to 40 per cent of cases, and bleeding (leading to anaemia) has been reported in 50 per cent. Tumours may ulcerate, perforate or bleed, and they can present acutely with intestinal obstruction (due to intussusception or stenosis) or peritonitis.

The diagnosis of lymphoma may be confirmed by endoscopy for more proximal lesions, or by contrast studies for more distal lesions. Radiological appearances suggestive of lymphoma include mucousal nodularity and ulceration. In addition, an abdominal CT scan can be diagnostic. A histological diagnosis may be obtained with the use of endoscopic biopsy or a Crosby capsule, but a laparotomy is often required to make a definitive diagnosis.

As lymphatic spread occurs early, localized tumours should be treated by en bloc resection of the tumour and regional lymph nodes. Ideally, a margin of normal tissue should be included in the resection specimen. In patients with disseminated disease, the primary lesion should be removed if possible, since both radiotherapy and chemotherapy can lead to perforation or haemorrhage of the tumour. As most tumours are advanced at the time of diagnosis, patients usually receive adjuvant chemotherapy and/or radiotherapy. At present, randomized trials are being carried out to determine the optimal treatment.

The overall 5-year survival is 10 to 40 per cent, although children often do worse. The prognosis depends on the stage of the disease and the histological type of tumour. Advanced tumours and those that perforate (and thus disseminate) do badly. Curative resection results in a 5-year survival rate of approx. 50 per cent.

Enteropathy-associated T-cell lymphoma This tumour accounts for approximately one-third of small-intestinal lymphomas and commonly arises in the sixth and seventh decades in patients with known coeliac disaese. It has also been described in children with coeliac disease. Tumours affect the jejunum more frequently than the ileum, reflecting the underlying pattern of the enteropathy. Macroscopically, the tumours may be single or multiple, often extending into the mesentery and mesenteric lymph nodes.

Initial symptoms, often insidious, are weakness, loss of weight, nausea, and diarrhoea. If a tumour is allowed to progress, it can cause obstruction, haemorrhage, and perforation. Physical signs include fever, anaemia, cachexia, lymphadenopathy, clubbing, and abdominal distension. In patients with coeliac disease, the diagnosis of malignancy must be suspected if they continue to detiorate despite adhering to a gluten-free diet. In such patients, endoscopic biopsies and radiological studies are often inadequate; diagnostic laporotomy is indicated in cases of doubt.

Treatment involves local resection of tumour and combination chemotherapy and/or radiotherapy. Prognosis is very poor, with few patients surviving a year.

Mediterranean lymphoma Mediterranean lymphoma arises mainly in young males in the Mediterranean region and the Middle East. Patients tend to come from lower socioeconomic groups; poor hygiene and parasitic intestinal infections are possible aetiological factors. The lymphoma arises in association with diffuse plasma-cell infiltration of the small intestine (known as immunoproliferative small-intestinal disease). Typically there is diffuse thickening of the small intestine with multiple nodules and early lymph-node involvement. Infiltrating plasma cells synthesize an abnormal a-immunoglobulin heavy chain that is often detectable in serum and urine.

Characteristic symptoms include diarrhoea, steatorrhoea, abdominal cramps, and weight loss. On examination, finger clubbing may be present and an abdominal mass may be palpable. Acute presentations can also occur, with obstruction and intussusception, bleeding or perforation. Intestinal parasites such as *Giardia lamblia* and *Ascaris lumbricoides* are detectable in 50 per cent of patients. Other investigations may reveal megaloblastic anaemia due to vitamin B_{12} and folate deficiency, and

electrophoresis may demonstrate an abnormal protein.

Immunoproliferative small-intestinal disease is found throughout the small intestine, although the duodenum and jejunum are most commonly involved. A diagnosis therefore can be achieved in most patients by means of upper gastrointestinal endoscopy and biopsy. Radiological contrast studies are useful in showing mucosal thickening or polypoid tumours in patients with jejunal and ileal disease.

A staging system has evolved that is useful in planning treatment and gauging prognosis. Stage A is characterized by infiltration with plasma cells, which may be slightly dystrophic. In stage C, an immunoblastic lymphoma infiltrates the small intestine or forms distinct tumours. Stage B is intermediate between stage A and stage C. Stage A may respond to eradication of parasites if these are present. In addition, antibiotic therapy can be given in the form of tetracycline or metronidazole plus ampicillin, and this therapy can lead to cure in 40 per cent of patients. In stages B and C, and in cases of antibiotic failure, combination chemotherapy and/or radiotherapy is used. In patients with advanced disease and in those who fail antibiotic therapy, response to combination chemotherapy is limited; at best 60 per cent of patients will respond. The prognosis in stages B and C is poor, most patients dying within 2 years.

Secondary malignancy

Secondary involvement of the small bowel may occur through direct local spread from a tumour in close proximity to the affected segment of bowel, or by transcoelomic, haematogenous, or lymphatic spread from a more distant site. The duodenum can be involved by direct local extension of colonic, gastric, pancreatic, or renal tumours, and by metastatic disease from primary cancer of the breast, kidney, lung, and testis. Secondary deposits of malignant melanoma can also occur.

Symptoms of duodenal involvement include anaemia, abdominal pain, obstruction of the gastric outlet, and jaundice. It is common for patients with such duodenal disease to have secondary disease elsewhere, but isolated metastases are occasionally seen and may be amenable to potentially curative treatment. In one report, patients with both local extension and isolated metastases were treated by pancreatoduodenectomy. In this group, primary tumours included colonic, renal-cell, breast, lung, and stomach carcinomas and malignant melanoma. Eighteen patients were identified and a perioperative mortality rate of 5 per cent, a median survival of 40 months, and a 5-year survival rate of 35 per cent. reported It was concluded that in the absence of other metastatic disease, pancreatoduodenectomy should be considered for locally advanced tumours or isolated metastases.

The ileum and jejunum may be invaded by tumours of the colon, uterus and ovaries, and in addition can become involved in malignant mesenteric lymph nodes. The serosal aspect of the small bowel is often affected by carcinomatosis peritonei, but occasional mucosal deposits are seen in patients with carcinoma of the lung or breast and in those dying of melanoma. Resection of such tumours should be undertaken if technically feasible, but palliation by means of bypass is often the only option.

Secondary spread of nodal lymphoma to involve the gut is approximately 10 times commoner than primary gastrointestinal lymphoma. As nodal lymphoma progresses, the gastrointestinal tract becomes increasingly involved; in the terminal stages of non- Hodgkins and Hodgkins lymphoma the bowel is involved in 50 per cent and 20 per cent of cases respectively. Treatment of the disease usually consists of combined chemotherapy and radiotherapy, with operation being reserved for complications such as obstruction and perforation. Prognosis is poor, with a mean survival of 9 months.

Malignancy associated with AIDS

Malignant disease is commonly associated with AIDS. Indeed, Kaposi sarcoma, which was previously rare, is the second most common mode of presentation of the syndrome after *Pneumocystis carinii* pneumonia. Clinically, Kaposi sarcoma presents as violaceous lesions scattered on the skin. If upper and lower gastrointestinal endoscopy is performed on patients at the time of presentation, gastrointestinal lesions will be detected in 40 per cent. At autopsy, 70 per cent of patients have gastrointestinal involvement. Tumours can appear throughout the gut, but the oropharynx, oesophagus, and rectum are particularly affected; the duodenum and the jejunoileum are affected to a lesser degree. Tumours range from flat, telangiectatic lesions to larger, polypoid lesions, the former proving difficult to diagnose with contrast radiology. Endoscopy is the main diagnostic tool. A tissue diagnosis can be difficult to obtain as less than a quarter of suspicious lesions prove to be positive on biopsy; the diagnosis is often made on the presence of mucocutaneous involvement, therefore.

Although most patients with AIDS do have gastrointestinal involvement, complications seldom arise. Haemorrhage can be overt or occult, and patients can also present with a syndrome that mimics ulcerative colitis. Although oral lesions may respond to radiotherapy, chemotherapy is the mainstay of treatment. As patients are already immunocompromised, chemotherapy must be carefully tailored to avoid the risk of opportunistic infections. Regimens using vincristine and bleomycin alone or in combination produce a response in 75 per cent of patients, with a low risk of immunosuppression.

Along with Kaposi sarcoma, lymphomas make up 95 per cent of tumours affecting those with AIDS, and these tend to be non-Hodgkin lymphomas of B-cell type. Extranodal involvement is common, the central nervous system, bone marrow and gut can all be involved. The clinical features are abdominal pain and diarrhoea, but perforation can occur. Treatment options include radiotherapy and chemotherapy; operative treatment is reserved for complications. The prognosis is generally poor, but it depends on the histological characteristics of the tumour; large, non-cleaved-cell tumours have the best outlook.

Non-neoplastic lesions

Hamartomatous lesions

Peutz–Jegher syndrome

This syndrome is inherited in an autosomal-dominant fashion with variable, incomplete penetrance. Patients have hamartomatous polypsthroughout the gastrointestinal tract and these may be pedunculated or sessile. Most polyps occur in the small bowel, but colonic and gastric polyps are also well recognized. Polyps are generally between 0.5 to 1 cm in size, but larger lesions are sometimes seen. They tend to be multiple, and groups of polyps can appear in different areas sequentially.

Malignant change does occur in hamartomatous polyps, as a recent study indicated that cancer developed in 6 per cent of lesions in the small bowel; patients with duodenal lesions had a greater risk of malignant change. In addition to this predisposition, patients with Peutz–Jegher syndrome are at an increased risk of dying from carcinoma of the colon and other apparently unrelated tumours. The mechanisms involved in this phenomenon are unclear.

Peutz–Jegher syndrome can present as early as infancy and the vast majority of patients are diagnosed before 30 years of age. Alert clinicians may notice the characteristic pigmentation of the lips and buccal mucosa. Symptoms may range from recurrent colicky abdominal pain to haemorrhage, anaemia, and intussusception. If complications arise, conservative resection is recommended, as these lesions tend to be recurrent and repeat procedures are often required. Peroperative endoscopy and snaring of polyps have been recommended to reduce the need for early reoperation and undue sacrifice of normal bowel.

Other hamartomatous conditions

These conditions are rare. Juvenile polyposis can affect the whole gastrointestinal tract and may present with haemorrhage, obstruction, or intussusception. Malignant change has not been reported.

The Cronkhite–Canada syndrome is a non-familial acquired condition in which multiple hamartomatous polyps occur throughout the gastrointestinal tract. Patients present with diarrhoea and steatorhoea leading to weight loss and malnutrition. Associated ectodermal changes include alopecia, nail dystrophy, and skin hyperpigmentation.

Heterotopic tissue

This is defined as normal tissue situated where it is not usually found. In the case of the small bowel, it can be either localized or disseminated.

Localized

In the gastrointestinal tract, the most common tissues involved are gastric mucosa and exocrine pancreas. Gastric heterotopia can occur throughout the gastrointestinal tract from mouth to anus, but it occurs predominantly in the small bowel, most commonly in a Meckel's diverticulum. The incidence of gastric heterotopia in these diverticula has been estimated at between 20 and 50 per cent, and this figure rises to 80 per cent in symptomatic diverticula. Presentation usually follows ulceration, which leads to bleeding or perforation; indeed, the abnormally sited mucosa may undergo the full range of histopathological changes seen in

orthotopic gastric tissue. Symptomatic lesions should be excised by Meckelian diverticulectomy, taking care to avoid narrowing the lumen of the adjacent ileum.

Ectopic pancreatic tissue may be found in the stomach, duodenum, jejunum, or Meckel's diverticulum (Fig. 16). The appearance is usually that of a nodule within the submucosa or muscularis, although polypoid forms have been described. Most of these lesions are asymptomatic. Periampullary lesions can cause biliary obstruction, and (very rarely) duodenal obstruction and haemorrhage. Symptomatic lesions should be locally excised, as should incidental lesions discovered at laparotomy if the surgeon is in diagnostic doubt. Very occasionally, a combination of heterotopic gastric and pancreatic tissue may be found at one site.

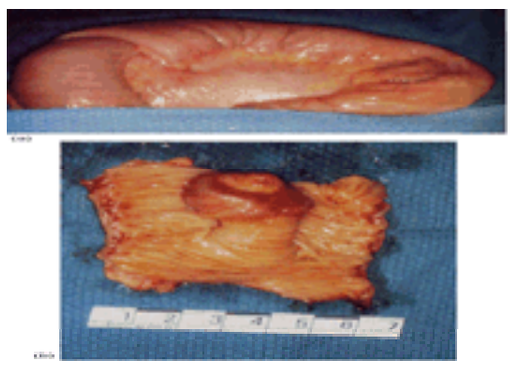


Fig. 16. Invaginated Meckel's diverticulum containing heterotopic pancreatic tissue. Barium enema had shown a filling defect defect in the terminal ileum in 47-year-old man with a short history of diarrhoea. (a) At laparotomy an intussuscepting tumour of the ileum was found; (b) an ulcerated nodule of ectopic pancreas at the tip of an invaginated Meckel's diverticulum formed the apex of the intussusception.

Disseminated

Endometriosis may also be considered a heterotopic condition and it can affect the small bowel, albeit rarely. The gastrointestinal tract is affected in 12 per cent of patients, the small bowel being the site of disease in 7 per cent of these. The ileum is most often affected. Endometrial tissue occurs on the serosal surface of the small bowel, and cyclical engorgement and sloughing may give rise to haemorrhage. In response to this haemorrhage and subsequent inflammation, adhesions may develop and in turn cause intestinal obstruction. Symptoms include nausea, vomiting, and colic, but these lesions are often symptomless. Thus endometriosis may be discovered as an incidental finding during laparoscopy or laparotomy for other conditions.

If symptomatic, the affected segment of bowel should be resected. If asymptomatic lesions are present and malignancy can be excluded, current developments in the management of endometriosis, including endocrine therapy or laser ablation, may preclude the necessity for enteric resection.

Duplications

Duplications of the small bowel are rare developmental abnormalities. The jejunum and ileum are affected in approx. 50 per cent of cases, the duodenum less often so. Macroscopically they appear as a cystic swelling in the mesentery, communicating proximally or distally with the lumen of the bowel. The lumen of the duplication is lined with intestinal epithelium plus part or all of the normal muscle components. Gastric heterotopia occurs in 15 to 36 per cent of cases. Presentation generally occurs in infancy with an abdominal mass or intestinal obstruction. Less commonly, there is bleeding, perforation, ileus, or acute pancreatitis. The diagnosis may be suggested by plain abdominal radiographs and contrast studies. With improvements in ultrasonography, some duplications have been discovered *in utero*. Characteristic appearances include an anechoic, fluid-filled lumen with a double-layered surrounding wall. Peristaltic contractions may be seen within the lumen, and this appearance makes the diagnosis of duplication highly likely.

If the abnormality is limited, treatment involves resection of the duplication along with the associated bowel. If extensive, the common blood supply between the duplication and native bowel must be carefully protected to avoid undue sacrifice of normal bowel. If this manoeuvre is not possible, options include removal or destruction of the mucosa, with or without partial resection of the abnormality, or anastomosis of both ends of the duplication into the normal bowel lumen.

Further reading

Arnold JC. Liver transplantation for primary and secondary apudomas. *British Journal of Surgery* 1989; **76**: 248. [Series of four patients undergoing liver transplantation for primary and secondary ‘apudomas'.]

Ben-Ayed F. Treatment of alpha-chain disease: results of a prospective study in 21 Tunisian patients by the Tunisian-French Intestinal Lymphoma Study Group. *Cancer* 1989; **63**: 1251–6. [Prospective study describing diagnosis, treatment, and outcome of patients with varying stages of alpha-chain disease.]

Billroth T. Multiple lymphome Erfolgreiche Behandlung mit Arsenik. *Weiner Medizinische Wochenschrift* 1871; **21**: 1066. [Historical reference describing gastrointestinal lymphoma and its possible relationship to arsenic consumption.]

Cooper M.J, Williamson RCN. Enteric adenoma and carcinoma. *World Journal of Surgery* 1985; **9**: 914–20. [Review of enteric adenoma and carcinoma.]

Cozzi P. Cryotherapy treatment of patients with hepatic metastases from neuroendocrine tumours. *Cancer* 1995; **76**: 501–9. [Study showing useful alleviation of symptoms in six patients with neuroendocrine metastases.]

Cscerni G. Gastric pathology in Meckel's diverticulum. A review of cases between 1965–1995. *American Journal of Clinical Pathology* 1996; **106**: 782–5. [In 140 cases of Meckel's diverticulum examined 25 showed gastric heterotopia. In every case of gastric heterotopia, gastritis was present in the mucosa.]

Damron T. Gangliocytic paraganglioma in association with a duodenal diverticulum. *American Journal of Gastroenterology* 1989; **84**: 1109. [A case of gangliocytic paraganglioma in a duodenal diverticulum is described.]

Galadiuk S. Villous tumours of the duodenum. *Annals of Surgery* 1988; **207**: 234. [A series of 32 patients who either underwent resection or surgical bypass for adenoma and carcinoma of the duodenum is described.]

Harrison L. Pancreaticoduodenectomy for nonperiampullary primary tumours. *American Journal of Surgery* 1997; **174**: 392–5. [Series of 18 patients undergoing pancreatoduodenectomy for non-periampullary tumours.]

Hermann R. Gastrointestinal involvement in non-Hodgkin's lymphoma. *Cancer* 1980; **46**: 215–22. [Case series reporting primary gastrointestinal involvement in 71 of 813 patients with non-Hodgkin's lymphoma.]

Kurtz RC. Upper gastrointestinal neoplasia in familial polyposis. *Digestive Diseases and Sciences* 1987; **32**: 459. [Report of a high incidence of upper gastrointestinal involvement in asymptomatic patients with FAP.]

Levine AM. Non-Hodgkin's lymphomas and other malignancies in the acquired immune deficiency syndrome. *Seminars in Oncology* 1987; **14**: 34–9. [Review suggesting possible aetiologies to account for the development of non-Hodgkin's lymphoma in AIDS patients.]

Ludwig DJ. Gut stromal tumours and their clinical behaviour. *American Journal of Surgery* 1997; **173**: 390–4. [Review of 39 cases of gastrointestinal stromal tumour, including diagnostic considerations and surgical treatment.]

McGowan I. AIDS and the gut. In: Misiewicz J, Pounder R, Venables C, ed. *Diseases of the gut and pancreas*, pp. 989–1006. Blackwell Scientific, Oxford, 1994. [Chapter of book outlining gastrointestinal pathology in AIDS patients.]

Maglinte D. Brunner's gland adenoma; diagnostic considerations. *Journal of Clinical Gastroenterology* 1982; **4**: 127. [Report outlining diagnosis, histological features, and treatment of Brunner's gland adenomas.]

Melin M. A review of primary tumours of the small intestine. *American Journal of Gastroenterology* 1994; **89**: 1888–90. [Extensive review of the literature on small bowel tumours.]

Mittal VK. Primary malignant tumours of the small bowel. *American Journal of Surgery* 1980; **140**: 396–9. [Review of 39 cases of primary malignant tumours of the small intestine.]

Mohandras D. Liposarcoma of the ileum with secondaries in the liver. *American Journal of Gastroenterology* 1972; **58**: 172. [Case report describing this uncommon tumour.]

Perzin KH. Adenomas of the small intestine: a clinicopathological review of 51 cases and a study of their relationships to carcinomas. *Cancer* 1981; **48**: 799. [Study of 51 patients with small bowel tumours that contained adenomatous epithelium. The report showed that large tumours, those with a villlous architecture, and those situated in the periampullary region had a higher incidence of malignant features.]

Ranzi T. Gastric and duodenal polyps in familial polyposis coli. *Gut* 1981; **22**: 766. [Study reporting adenomatous change in macroscopically normal duodenal mucosa in patients with FAP.]

Reiner MA. Primary malignant neoplasms of the small bowel. *Mount Sinai Journal of Medicine* 1976; **43**: 274–80. [Report describing personal experience of small bowel tumours. Review of literature also included.]

Schulten M. Villous adenoma of the the duodenum. A case report and review of the literature. *American Journal of Surgery* 1976; **132**: 90. [Case report and review of literature.]

Scott-Coombes DM, Williamson RCN. Surgical treatment of primary duodenal carcinoma: a personal series. *British Journal of Surgery* 1994; **81**: 1472–4. [Personal series describing clinical features, investigation, and treatment of 22 duodenal cancers.]

Spigelman AD. Cancer and the Peutz–Jegher's syndrome. *Gut* 1989; **30**: 1588. [Study describing follow-up of patients with Peutz-Jegher's syndrome. Twenty-two per cent of patients developed malignant tumours of the small and large bowel after prolonged follow-up.]

Spotsswood SE. Ante-natal diagnosis of gastrointestinal abnormalities. *Paediatric Radiology* 1994; **24**: 344–5. [Report describing antenatal diagnosis of duplications of the gastrointestinal tract.]

Studley JGN, Williamson RCN. Small bowel tumours. *Surgery (Oxford)* 1991; **95**: 2263–7.

Thompson NW. Microgastrinomas of the duodenum. A cause of failed operations for the Zollinger–Ellison syndrome. *Annals of Surgery* 1989; **209**: 396. [Report describing microgastrinomas of the duodenum, emphasizing that these tumours may undetectable by all means other than duodenotomy and eversion of the mucosa.]

Woods H. Abdominal carcinoid tumours in Sheffield. *Digestion* 1990; **45**: 17. [Paper describing incidence, presentation, and treatment of abdominal carcinoid tumours in the Sheffield region.]

Younes C. *Ras* mutations in duodenal carcinomas. *Cancer* 1997; **79**: 1804–7. [Paper suggesting that mutations of the *K-ras* oncogene may play a significant role in the development of duodenal adenocarcinoma.]

24.9 Foreign bodies

Michael N. Margolies

True foreign bodies
Repeated and multiple ingestion of metallic objects
Miniature battery ingestion in children
Other objects
Narcotic packet ingestion
Bezoars
Food bolus obstruction
Gallstones and enteroliths
Concretions of medications or chemicals
Intestinal parasites
Further reading

The variety of intestinal ‘foreign bodies' is enormous and is limited only by the imagination and appetites of patients. Although ingestion of foreign bodies is common, the symptoms from ingestion are not: the diagnosis is only made by history, if available, or following the appearance of a complication. The definition of foreign body is problematic, as certain ‘foreign bodies' such as a food bolus or phytobezoar would be considered food, whereas other ‘foreign bodies' such as enteroliths and gallstones originate within the digestive system of the host (Table 1). However, all these are included as foreign bodies by virtue of the complications that may arise from their presence in the gut lumen.

True foreign bodies: metal, wood, plastic, etc.
Narcotic packets
Bezoars
Food bolus
Enteroliths and gallstones
Concretions
Intestinal parasites

Table 1 Foreign bodies of the small intestine.

The complications of small intestinal foreign bodies requiring surgical treatment include obstruction, perforation, and bleeding. In the absence of a history of unusual ingestion or of the finding of a radio-opaque foreign body, the specific cause is rarely identified preoperatively. The clinical presentation and the indications for surgery in patients with complications of foreign body ingestion are indistinguishable from other causes of obstruction, perforation, and bleeding.

Only 1 to 2 per cent of all cases of acute small bowel obstruction are due to obturation by a foreign body. The site of obstruction is most commonly in areas of narrowing: the distal ileum, the ileocecal valve, or sites of pre-existing inflammatory disease with stricture, neoplasm, or diverticula. Choices of operative treatment include manual displacement of the intraluminal object, if safe, into the colon, removal via enterotomy, or resection if the object is imbedded or has caused transmural necrosis. Perforation due to presence of a foreign body may result in diffuse peritonitis, but more commonly it presents itself as contained sepsis with a localized abscess or as fistulization into an adjacent viscus or parenchymatous organ. Treatment includes, in addition to antibacterial agents, removal of the foreign body, closure of the enterotomy or resection as dictated by the local findings, and drainage of abscess if it is present. Bleeding from a small bowel foreign body arising through mucosal pressure necrosis and ulceration is infrequent.

The management of patients with an ingested foreign body in the absence of the above complications is dictated by the nature of the foreign body. Once objects have left the stomach more than 95 per cent of them proceed through the small intestine and colon unimpeded and without untoward effects. The population of patients suffering foreign body ingestion includes children, those with psychiatric disorders, alcoholics, prisoners, and denture wearers who have diminished palatal sensation. A relative, although controversial, indication for surgery is failure to pass a foreign body as evidenced by lack of progress on serial radiographs. In children, the average time for passage of an object through the small intestine is 5 days. If their parents are reliable, children can be managed as outpatients to await passage confirmed by observation of stools or occasional radiographs.

Certain categories of foreign objects have special features that may dictate surgical removal.

True foreign bodies

Repeated and multiple ingestion of metallic objects

Certain patients with severe psychiatric disorders or incarcerated felons repeatedly ingest an astonishing variety of objects, including pins, eating utensils, bedsprings, razor blades, bolts, and scissors. Evidence of self-mutilation and drug use is common. Once these objects reach the small bowel they are likely (more than 90 per cent) to traverse the remainder of the gastrointestinal tract safely. Perforation or obstruction occurs in only 0.5 per cent of instances. Inasmuch as these patients have a remarkable propensity for eating their environment, as manifested in particular by ingesting parts of hospital beds, if surgery becomes indicated, radiographs should be obtained immediately prior to exploration.

Miniature battery ingestion in children

Although alkaline disc batteries can potentially leak their caustic contents in the stomach, other types of batteries (mercury, lithium, silver) are innocuous; the risk of mercury poisoning is low. Once such smooth round batteries reach the small intestine virtually all will pass. The presence of a battery in the esophagus, however, requires urgent endoscopy.

Other objects

In contrast to metallic foreign bodies, pointed wooden objects such as splinters or toothpicks, and fish or chicken bones are more hazardous owing to their length and sharp ends, which make them more likely to perforate the bowel wall. Up to one-third of such objects which reach the small bowel may cause perforation. These patients most often present with an acute abdomen, radiographs and history being unrevealing. Perforation usually results in intra-abdominal abscess or fistula: this may occur at any site in the bowel wall but in particular in the ileocecal region, including Meckel's diverticulum, where the differential diagnosis includes more common inflammatory lesions. Toothpick perforations are particularly prone to produce fistulas with late septic complications.

Narcotic packet ingestion

A popular method of narcotic smuggling involves the placement of drugs, commonly cocaine or heroin, in latex packages or condoms into body cavities or by swallowing, their retention being favoured by the use of constipating agents. In addition to the risk of obstruction (6 per cent), acute narcotic toxicity and death has occurred. The fatal oral dose of cocaine is 1 to 3 g; a single packet of cocaine contains 3 to 12 g, and carriers ingest many packets. The likelihood of ruptured packets and resultant mortality has decreased, however, as smugglers have grown more sophisticated, and immediate surgery is no longer mandatory. Mild cathartics may be used and patients may be discharged when proved free of packets by radiologic and serial stool examinations. Older types of packets (containing loose cocaine covered by two to four layers of condoms or latex) are more prone to rupture. If this type of packaging can be identified by history or by recovery of packets from the

rectum or stool, or if toxicity is present, surgical removal is indicated. Packets are frequently visible on plain films.

Bezoars

Small intestinal obstruction due to bezoar is a common form of obturating obstruction, usually occurring as a late sequela of gastric surgery. Bezoars consisting of concretions of vegetable fibres (phytobezoar) reach the small bowel either directly through the pylorus or through a gastroenterostomy, or after fragmentation consequent upon attempted gastroscopic removal. Caution must be exercised at laparotomy not to overlook other bezoar fragments in addition to the one provoking the obstruction. Small bowel bezoars may also originate in duodenal or jejunal diverticula, as well as in Meckel's diverticulum. Trichobezoar may involve the small bowel by extension for a considerable distance distal to a large gastric hairball.

Food bolus obstruction

Although not true foreign bodies, large boluses of certain foods may cause small intestinal obturating obstruction. These include masses of citrus fruit fibres, desiccated fruits, any high fibre food such as coarse bread at times of famine or following religious fasts, turtle eggs, and grasshoppers. Food bolus obstruction in non-temperate zones can often be managed without laparotomy, as the obstruction is usually partial. Additional factors that predispose to food bolus obstruction include prior gastric surgery and inadequate dentition. In cases of obstruction by food bolus or bezoar, postoperative education of patients to avoid binge eating, modification of fibre intake, and fitting of dentures may be useful in preventing recurrence. The operative management of obstruction by food bolus or by bezoar is similar, with attempts to squeeze the mass or fragment into the colon. When firmly impacted, an enterotomy may be needed.

Gallstones and enteroliths

Enteroliths or intestinal calculi are presumed to form *de novo* in the bowel lumen. They consist of calcium or magnesium salts, or contain cholesterol and bile acids. They were, in earlier reports, considered to be common ‘foreign bodies'. Their appearance is, however, rare, and a specific diagnosis preoperatively is rarer still. They may form in the stagnant milieu of duodenal or jejunal diverticula, in blind loops, and in areas of Crohn's disease or tuberculosis.

Concretions of medications or chemicals

Small bowel obstruction may result from medications, including antacids and boluses of hydrophilic colloid laxatives, which absorb water and swell into a gelatinous mass. Antacid obstruction is seen particularly in patients undergoing hemodialysis. Concretions in the small bowel can also occur in those with more bizarre forms of ingestion, such as drinking alcoholic solutions of shellac, and cement powder used for producing mortar or concrete.

Intestinal parasites

Infestation by the round worm *Ascaris lumbricoides* accounts for 10 to 15 per cent of cases of intestinal obstruction in endemic tropical areas in Africa and Asia. Following appendicitis, ascariasis is the second most common cause of acute abdomen in children in those areas. Massive ascariasis can cause low-grade intestinal obstruction, which can often be managed conservatively with fluids, gastric drainage, and repeated doses of anthelmintics. However, acute complete obstruction may occur, sometimes following the use of a vermifuge. The dead or dying worms impacted in the terminal ileum and right colon are capable of causing tissue necrosis and perforation. In addition to the presenting symptoms of small bowel obstruction the masses of worms may be palpable or visible on plain films. Worms may be present in the vomitus or stool. Although intestinal obstruction may be due to obturation, volvulus, or intussusception, irritative intestinal spasm due to the worms may contribute. At surgery, if the obstructing bolus of worms cannot be advanced distally, resection may be necessary, particularly if local gangrene, volvulus, or intussusception has occurred.

Further reading

James AH, Allen-Mersh TG. Recognition and management of patients who repeatedly swallow foreign bodies. *Journal of the Royal Society of Medicine* 1982; **75**: 107–10. [Emphasizes conservative management of purposeful ingestion of foreign bodies in adults.]

Lancashire MJ R, Legg PK, Lowe M, Davidson SW, Ellis BW. Surgical aspects of international drug smuggling. *British Medical Journal* 1988; **296:** 1035–7. [Describes types of drug packaging used by ‘body packers'.]

Spitz L. Management of ingested foreign bodies in childhood. *British Medical Journal* 1971; **4**: 469–72. [Review of 660 pediatric cases with discussion of endoscopic and surgical approaches.]

Stack LB, Munter DW. Foreign bodies in the gastrointestinal tract. *Emergency Medical Clinics of North America* 1996; **14:** 493–521. [Comprehensive recent review of gastrointestinal foreign body management.]

24.10 Pneumatosis cystoides intestinalis

Andrew Mitchell

Further reading

Pneumatosis cystoides intestinalis is an uncommon condition characterized by the presence of submucosal and/or subserosal gas-filled cysts in the intestinal wall. These cysts, which may reach several centimetres in diameter, contain largely nitrogen or hydrogen and occur in the small bowel more commonly than in the colon and rectum.

Pneumatosis cystoides intestinalis was first reported as a post-mortem observation by Du Vernoi in the eighteenth century. The condition is commonly asymptomatic and discovered incidentally at laparotomy, laparoscopy, or on radiological investigation of an unrelated symptom, but it can also produce abdominal pain, diarrhoea, subacute intestinal obstruction, intussusception, mucus discharge, or rectal bleeding. Its aetiology remains obscure. Favoured theories include the intramural penetration of gas-producing organisms, excessive intraluminal fermentation, and the forcing of gas under high pressure into the bowel wall, either from the lungs via the mediastinum and mesentery in patients with pulmonary disease, or from the intestinal lumen after surgery or biopsy has left a mucosal breach. Each of these theories is supported by clinical and experimental observations. The condition is well recognized in pigs and some other animals.

Pneumatosis cystoides intestinalis is most common in men between the ages of 30 and 50 years, and has an association with peptic ulceration and pyloric stenosis, intestinal bypass surgery for morbid obesity, and chronic lung disease. A plain abdominal radiograph may show a linear arrangement of gas collections, or pneumoperitoneum, without clinical signs of peritonitis. Contrast studies distinguish this condition from polyposis by demonstrating radiolucent defects in the bowel wall which extend outside the column of barium (Fig. 1).

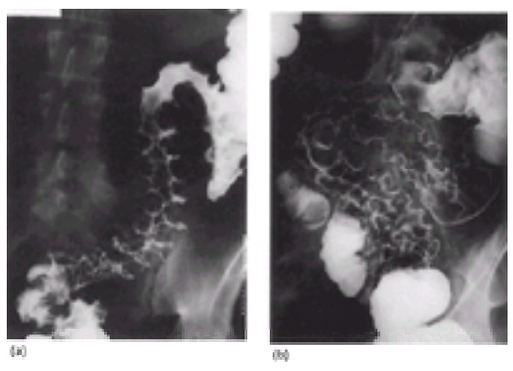


Fig. 1. (a) and (b) Barium enema appearance of pneumatosis coli.

Accurate diagnosis prevents unnecessary surgery. The occasional familial occurrence of the condition increases the risk of confusion with familial polyposis coli and, therefore, of inappropriate colectomy. Biopsy should be performed, when the cyst will sometimes rupture with a hissing sound. Histology is diagnostic (Fig. 2 and Fig. 3).

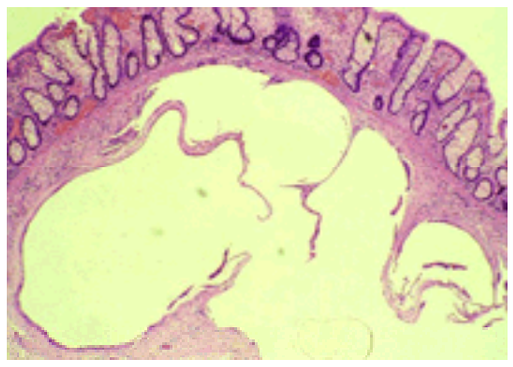


Fig. 2. Rectal biopsy showing normal mucosa with multiple cysts in the submucosa.

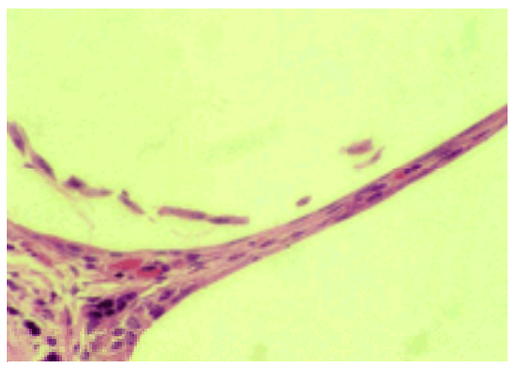


Fig. 3. High-power view of rectal biopsy showing chronic inflammation with a foreign body giant cell in the cyst lining.

No treatment is necessary for the 85 per cent of patients who are asymptomatic. Oxygen therapy has been used successfully for symptomatic patients and is thought to work by inducing the exchange of inert gases from the cysts for oxygen, which is later absorbed. Good results have been reported after inhalation of 70 per cent oxygen, achieving a PO_2 of 250 mmHg for 5 days and after hyperbaric oxygen therapy at 2.5 atmospheres (252.5 kPa) for 150 min on each of 3 successive days. Dietary treatment has been successful in selected patients.

Intestinal resection should be reserved for patients with complications or severe refractory symptoms. The results of surgery in these cases are good.

Further reading

Boerner RM, Fried DB, Warshauer DM, Isaacs K. Pneumatosis intestinalis. Two case reports and a retrospective review of the literature from 1985 to 1995. *Digestive Diseases and Sciences* 1996; **41**: 2272–85.

Case WG, Hall R. Surgical treatment of pneumatosis coli. *Annals of the Royal College of Surgeons of Engand* 1985; **67**: 368–9.

Elberg JJ. Oxygen therapy for pneumatosis coli. *Acta Chirurgica Scandinavica* 1985; **151**: 399–400.

Gagliardi G, Thompson IW, Hershman MJ, Forbes A, Hawley PR, Talbot IC. Pneumatosis coli: a proposed pathogenesis based on a study of 25 cases and review of the lierature. *International Journal of Colorectal Disease* 1996; **11**: 111–18.

Galandiuk S, Fazio VW. Pneumatosis cystoides intestinalis. A review of the literature. *Diseases of the Colon and Rectum* 1986; **29**: 358–63.

Heng Y, Schuffler MD, Haggitt RC, Rohrmann CA. Pneumatosis intestinalis: a review. *American Journal of Gastroenterology* 1995; **90**: 1747–58.

Spigelman AD, Williams CB, Ansell JK, Rutter KRP, Phillips RKS. Pneumatosis coli: a source of diagnostic confusion. *British Journal of Surgery* 1990; **7**: 155.

van der Linden W, Marsell R. Pneumatosis cystoides coli associated with high H2 excretion: treatment with an elemental diet. *Scandinavian Journal of Gastroenterology* 1979; **14**: 173–4.

25 Surgery for obesity

D. Michael Grace

Introduction
Definition and prevalence
Cause
Significance
Abdominal compartment syndrome
Alternatives to operation
Patient selection
Surgery
- Preparation for operation
- Operation
- Jejunoileal bypass
- Gastric bypass
- Gastroplasty
- Biliopancreatic diversion
- Gastric banding
- Laparoscopic procedures
- Complications
- Results
- Conclusions

Further reading
Selected overviews

Introduction

Obesity is epidemic worldwide and a serious threat to health. It is a paradox that obesity increases with national wealth but decreases with personal wealth and social standing. The dilemma is magnified by mass media that advertise high calorie, high fat foods while portraying the ideal human shape as thin and fit. The result is an obsession with dieting which feeds an enormous weight-loss industry. The desperation of the obese for thinness leads to extreme measures. Surgery for obesity is one of the most drastic. The risks of operation are significant but the benefits to mental and physical health and appearance can be dramatic. This field has evolved over 30 years but is changing quickly as laparoscopic procedures are adapted to very obese patients. Treatment alternatives for this group are limited but will appear as research on hunger and appetite progresses and new drugs and diets become available.

Definition and prevalence

Obesity is usually defined as weight more than 20 per cent above ideal. Body mass index (**BMI**) is commonly used to express degrees of obesity and is calculated as weight in kilograms divided by height in meters squared (wt in kg/(ht in m)^2). Normal values are 20 to 25. A 70-kg man with a height of 175 cm would have a body mass index of 23. A body mass index of 25 to 30 represents preobesity, 30 to 35 is mild or class I obesity, 35 to 40 is moderate or class II obesity, and greater than 40 is class III or morbid obesity. A separate category of extreme or superobesity with a body mass index greater than 50 is sometimes recognized. Candidates for the surgical treatment of obesity usually have a body mass index greater than 40 and a weight more than 45 kg above ideal.

Recent data from the United States showed a stable population of adult citizens with preobesity at about 32 per cent but a large increase in obesity BMI > 30) from 14.5 to 22.5 per cent between the surveys of 1976–1980 and 1988–1994. The prevalence of severe or class III obesity (BMI > 40) has been difficult to determine because such patients may be socially isolated and unwilling to respond to surveys. In some minority groups such as American black women the prevalence may exceed 10 per cent in the fourth and fifth decades. The prevalence is increasing most in those with the least education.

Cause

Both environmental and genetic causes contribute to obesity. The simple explanation is that energy intake exceeds expenditure, but the control system is complex and still incompletely understood. Modern high-fat fast foods combined with limited exercise and excessive television viewing are significant environmental factors. Obesity runs in families, which may show a reduced rate of energy expenditure in comparison with thinner subjects. Studies on twins show a strong genetic component to the BMI of adults, but weight changes with time are strongly influenced by the environment. There are no characteristic emotional changes in obesity. Depression seems more the result than the cause of obesity.

Significance

The health and cost implications of obesity are significant. Severe obesity increases the mortality rate of men and women, especially from cardiovascular disease. Other cardiovascular problems include hypertension, thrombophlebitis, and venous stasis ulcers. Obstructive sleep apnoea and obesity hypoventilation syndrome are severe complications of obesity which can lead to significant disability, sudden death, and motor vehicle accidents. Arthritis in weight-bearing joints limits exercise and mobility in the obese. Obesity is a major and treatable cause of type II diabetes mellitus. Hyperlipidaemia is common in the obese and contributes to the risk of vascular disease. Abnormal liver function and cholelithiasis are common complications of obesity. Increased intra-abdominal pressure contributes to reflux oesophagitis, stress incontinence, and benign intracranial hypertension. Depression may contribute to abnormal eating and obesity, but often results from obesity. Central or truncal obesity increases the risk of diabetes, cardiovascular disease, and early death. Table 1 summarizes the complications of obesity.



Table 1 Complications of severe obesity

The annual health and dieting costs of obesity may be as high as $100 000 000 000 per year in the United States. European data suggest that obesity contributes between 2 and 7 per cent of total health care costs. Quality of life is impaired in the severely obese. Prejudice among the general public and health care workers is severe. The obese have low social standing, are often unemployed or do menial work, and are often disabled. Even minimal exertion may be difficult. Clothing is difficult to find. Normal furniture and seats on public transportation may be too small.

Abdominal compartment syndrome

Acute increases in abdominal pressure have recently been recognized in patients subject to trauma or recent surgery and can impair renal, cardiac, and respiratory function. Sugerman and colleagues have described a chronic increase in abdominal pressure as measured by bladder catheter which correlates with poor ventilation, gastro-esophageal reflux, venous stasis, stress incontinence, and a high risk of postoperative incisional hernias. These conditions are relieved or prevented by

significant weight loss and decrease in abdominal pressure. Benign intracranial hypertension may have the same cause and cure in the morbidly obese.

Alternatives to operation

An astonishing range of methods have been used in an attempt to lose weight but there is virtually no long-term success. Behavior modification has been popular but of limited success. Fad diets and commercial plans have become a major industry but late success is rare. Appetite suppressants such as amphetamines and thyroxine were abandoned years ago. Fenfluramine and dexfenfluramine looked promising but were associated with a risk of rare but devastating primary pulmonary hypertension and a higher risk of cardiac valvular defects. Both have been taken off the market. New appetite suppressants and fat substitutes have been introduced but safety and long-term effectiveness are still to be established.

Patient selection

Patient demand for obesity surgery and weight loss is high because non-operative methods are ineffective. The problem is to select the patients who will benefit most, lose the most weight, and do so safely. Selection criteria are summarized in Table 2. Selection and follow-up should be carried out by a team including surgeon, internist, dietitian, and psychologist or psychiatrist. Assessment by a social worker, physiotherapist, nurse, or anaesthetist may be needed on occasion. Figure 1 shows the usual assessment process. Some studies show no significant psychological changes in morbidly obese people while others show an increase in mood disorders and depression, anxiety, bulimia, and tobacco dependence. Psychological criteria for patient selection are not well defined. Poor patient motivation demonstrated by skipped appointments or failure to change lifestyle suggest an unsuitable patient. In my experience teenagers do not have the discipline to follow postoperative diets. Patients as old as 65 years may occasionally be candidates when faced with a wheelchair or institution because of severe arthritis.

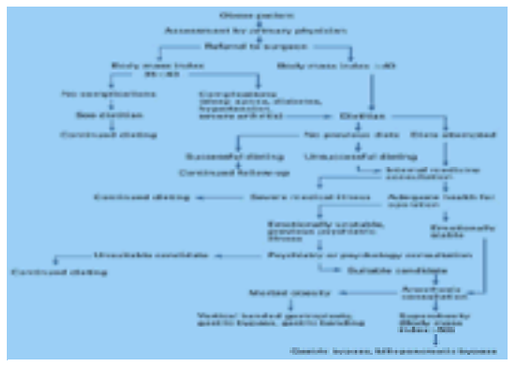


Fig. 1. Decision process for severely obese patients.

Weight > 45kg above ideal
Body mass index > 40
Dieting attempts made
Age 18 to 50
Emotionally stable and co-operative
Healthy enough for major operation
Complications of obesity:
Sleep apnoea, type II diabetes mellitus, hypertension, reflux oesophagitis, stress incontinence, infertility, osteoarthritis

Table 2 Criteria for patient selection

The selection process also applies to the bariatric surgeon and centre. An experienced surgeon working as part of a team is required. Most operations are uncomplicated but facilities to handle major problems in high-risk patients are needed. Intubation may be difficult and respiratory support may be needed, especially in patients with sleep apnoea. Because of the potential for late nutritional problems and bone disease, a commitment to long-term follow-up is essential.

Surgery

Table 3 outlines many of the surgical methods used to achieve weight loss. Only the more common procedures will be reviewed. Operations such as jejunoileal bypass cause weight loss by malabsorption and bypass of most of the small bowel. Gastroplasty is classed as a restrictive procedure since it reduces intake by reduction in gastric capacity. Gastric bypass causes both malabsorption and restriction. Gastroplasty is safe and quick with moderate chance of poor weight loss and obstruction. Gastric bypass is a more difficult and dangerous operation with a low rate of failure but potential for dumping and late nutritional problems.

Operation	Year begun	Advantages	Disadvantages	Present status
Jejunoileal bypass	1954	Good weight loss	Diarrhea, liver failure, renal stones	Rarely used
Jejunocolic bypass	1963	Dramatic weight loss	Concretions (diarrhea), liver failure	Dangerous—never used
Gastric bypass	1966 Mason	Excellent weight loss with no small pouch	Iron and B_{12} deficiency, dumping, metabolic bone disease	Widely used
Jaw wiring	1973	Simple and safe	Inability to vomit, weight regain when wires are removed	Rarely used
Gastric banding	1976	Simple	Variable weight loss, mesh erosion, pouch enlargement	Silastic adjustable band still used
Biliopancreatic diversion	1979 Scopinaro	Best long-term weight loss	Complex operation, protein malnutrition, stomal ulcer, metabolic bone disease	Still prescribed at choice in a few centres
Gastric balloon	1979	Simple	Balloon migration or deflation, unreliable	Not used
Proximal gastroplasty	1979	Simple, good early weight loss	Weight regain in most patients, stress ulcers	Not used
Vertical banded gastroplasty	1980 Mason	Simple, good early weight loss with small pouch, normal food absorption	Low weight loss than gastric bypass, weight regain, micronutrient deficiencies	Widely used
Laparoscopic gastric banding	1992	Quick patient recovery	Difficult operation, especially in superobese	Increasing, evaluation continues
Laparoscopic gastroplasty and gastric bypass	1994	Quick recovery	Technically difficult, same problems as open procedure	Still being developed

Table 3 Operations for obesity

Preparation for operation

Selection of motivated patients and careful explanation of the risks and benefits of operation assist in patient preparation. Preoperative meetings with other gastroplasty patients helps to ease fears. The dietitian should review postoperative meals. Smoking should stop at least 6 weeks preoperatively to decrease pulmonary complications. Preoperative consultation with internal medicine can help in the assessment and management of problems like hypertension, cardiovascular disease, sleep apnoea, and diabetes. Even very obese patients can be admitted on the day of operation. Subcutaneous heparin and prophylactic antibiotics are started preoperatively.

Operation

An experienced anesthetist is needed because intubation may be difficult in a massive patient with thick neck. Measures to prevent aspiration are necessary in patients with high abdominal pressure, large residual gastric volume, and low pH. Ventilation and exposure are assisted by elevation of the head of the bed and by mechanical

retractors. Careful closure of the abdominal wall is important. The thick layer of fat and high intra-abdominal pressure contributes to an increased risk of wound infection, disruption, and incisional hernia formation. Arterial lines are rarely needed but may help monitor blood pressure when a massive arm makes values inaccurate. To encourage early ambulation we avoid urinary catheters and use an orthopaedic frame with trapeze handle. Nasogastric tubes are used for 24 h or avoided entirely. A large supportive dressing or external support encourages early ambulation.

Jejunoileal bypass

This operation was first performed in the early 1960s and was very popular in the 1970s. In the most common procedure 35 cm of proximal jejunum was anastomosed to the side of the terminal ileum 10 cm from the ileocecal valve (Fig. 2). Another procedure used a similar end-to-end anastomosis. The bypassed intestine was not removed but was anastomosed to the transverse or sigmoid colon and, more recently, the stomach. Theoretically, this technique alters flora in the bypassed bowel and decreases postoperative problems like diarrhea and liver failure. Weight loss with these procedures was good and sustained, but diarrhea and flatulence were severe with significant loss of potassium and aggravation of hemorrhoids. Fat malabsorption contributed to hypocalcemia and oxaluria with renal stones. Severe liver failure causing death or requiring liver transplantation occurred in some patients. Normal eating was tolerated although large meals increased diarrhea. Only a few centres still perform this procedure with gastric drainage of the bypassed intestine. Parenteral nutrition may be needed prior to intestinal reconstitution. Rapid weight gain may result so gastroplasty may be considered for the patient who does not have liver disease.

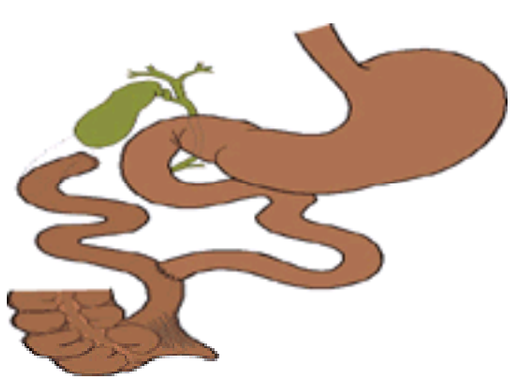


Fig. 2. Jejunoileal bypass attaching 35 cm of proximal jejunum to the terminal ileum 10 cm from the ileocecal valve.

Gastric bypass

Mason, working at the University of Iowa, developed the gastric operations for obesity. He observed that patients who had a Billroth II gastrectomy for management of peptic ulcerations often lost weight. He designed and meticulously documented a series of gastric bypass and gastroplasty operations that gave rise to modern procedures. Although he has favoured the vertical banded gastroplasty since the early 1980s, he and Ito reported the first gastric bypass operations in 1966.

Randomized trials have shown that gastric bypass gives better weight loss than gastroplasty but at the expense of more early technical problems and late iron and vitamin B_{12} deficiency and osteoporosis. A small upper gastric pouch is drained by gastrojejunostomy. In early years the stomach was divided and sutured, while in recent years staplers have been used to separate upper and lower stomach. Upper pouches may be horizontal (Fig. 3) or vertical (Fig. 4), separated (Fig. 4) or in continuity (Fig. 3), drained by loop or Roux-en-Y gastrojejunostomy with antecolic or retrocolic and antegastric or retrogastric anastomosis. The pouch is constructed small (< 30 ml) and may be surrounded by a band (Fig. 4) to limit dilatation. The vertical lesser curve pouch allows excellent blood supply and drainage and a short retrocolic and retrogastric route for anastomosis, but revision may be difficult if problems occur. Weight loss is excellent with these procedures and sustained.

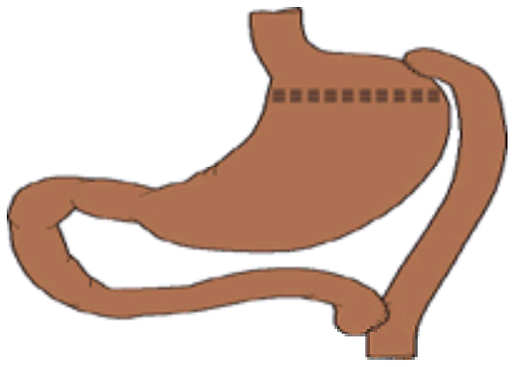


Fig. 3. Horizontal gastric bypass in which the proximal stomach is drained by a 45- to 75-cm Roux-en-Y limb of jejunum.

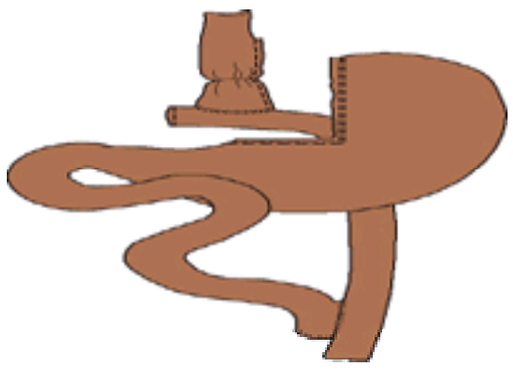


Fig. 4. Vertical gastric bypass with complete separation of the gastric pouch. A Silastic ring may be used to limit pouch dilatation.

Dumping will inhibit sugar intake for obese patients who like sweet foods. Stomal ulcers have been rare with a small pouch. Assessment of the distal stomach is difficult if abdominal pain is undiagnosed. Acute dilatation of the lower pouch postoperatively is fortunately rare because diagnosis is difficult. Severe nausea, tachycardia, elevated liver enzymes and amylase suggest this diagnosis. Gastric air–fluid level is absent so ultrasound or early operation may be needed and lifesaving. Small bowel obstruction may produce unusual symptoms.

Gastroplasty

Eating is restricted by a small proximal gastric pouch but there is no alteration of digestion or absorption. In the early 1980s horizontal staple lines were favoured. Early weight loss was good but weight regain was common due to staple line disruption or stoma dilatation. The modern operations followed the vertical banded gastroplasty of Mason (Fig. 5). Silastic ring gastroplasty is similar but does not require removal of any gastric tissue. Mason has achieved good early and late results by careful techniques using a very small and carefully measured pouch with a 5-cm ring of Marlex. He favours the procedure because it is simple, quick and safe, gives good weight loss, and allows easy inspection of the pouch, stoma, and distal stomach by endoscopy or barium studies. In comparison with patients who have had a gastric bypass, vitamin B_{12} and iron deficiency and osteoporosis are rare.

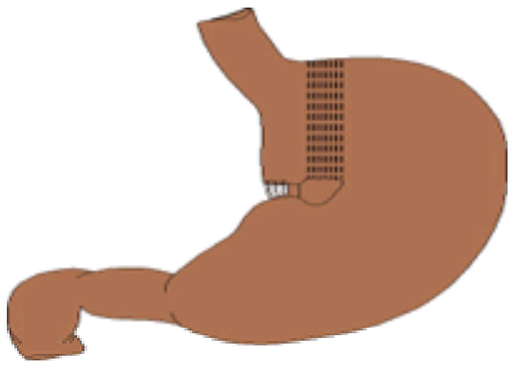


Fig. 5. Vertical banded gastroplasty of Mason.

Other surgeons have had difficulty duplicating his results. Failure to follow his exact technique may explain some of the differences. Staple line disruptions have been common and a cause of late failure. The small rigid stoma does inhibit ingestion of meat, raw fruit and vegetables, and encourages a ‘soft calorie' syndrome. Ice cream, soft drinks, and potato chips are readily tolerated and may contribute to late weight gain.

Biliopancreatic diversion

Scopinaro is another of the pioneers of obesity surgery. After a series of dog experiments he developed an operation with some superficial resemblance to the gastric bypass. However, there are significant differences. The proximal gastric pouch is large (200 to 500 ml) and adjusted to patient size. The gastroenteric anastomosis allows ingestion of regular food. The distal stomach is removed. The small bowel is divided 250 cm from the ileocecal valve and a gastroileostomy performed. Bile and pancreatic juice are diverted into the terminal ileum 50 cm from the ileocecal valve (Fig. 6).

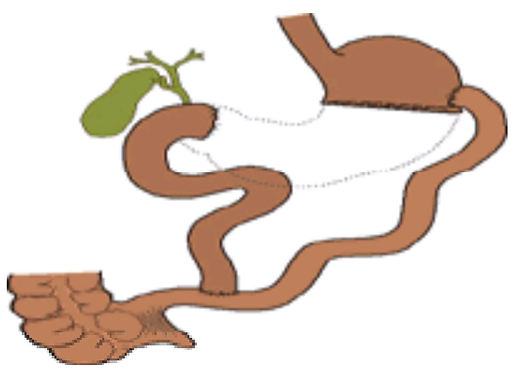


Fig. 6. Biliopancreatic diversion in which ileum 250 cm from the ileocaecal valve is anastomosed to a large gastric pouch and proximal bowel joined 50 cm from the ileocecal valve.

Scopinaro has described a series of 1610 patients of whom 916 had the above operation between 1984 and 1991. Average weight loss was 70 per cent of excess weight (weight above ideal) after 2 years and this was maintained for 12 years. Operative mortality of 7 patients (0.8 per cent) is higher than for other operations. However, weight loss is better than that achieved with any other procedure and complications of obesity are reversed. Poor fat absorption contributes to frequent bowel movements and flatulence. Fat-soluble vitamins are poorly absorbed. Calcium deficiency and osteoporosis may occur. In a few patients significant protein–calorie malnutrition develops.

Recent modification preserving the duodenum as a switch and lengthening the common limb may maintain weight loss and decrease complications. The magnitude of this operation combined with the complications have frightened most surgeons. In the hands of Scopinaro and a few North American surgeons excellent long-term weight loss has been achieved.

Gastric banding

Application of a prosthetic band to the proximal stomach is a simple procedure in comparison with other operations. Early Marlex or Dacron bands gave variable weight loss with frequent stoma obstruction, mesh erosion, or poor emptying of the gastric pouch. Kuzmak developed a more effective silicone band. In 1986 he started to use an adjustable silicone band with subcutaneous chamber. Weight loss of 43 kg (32 per cent) after 4 years has been achieved with few complications. Band removal is easy in the event of late problems. Laparoscopic gastric banding is the direct result of this procedure.

Laparoscopic procedures

Laparoscopic surgery has developed quickly. Adjustable gastric banding has been the easiest operation to adapt to laparoscopy. The band is placed very high on the stomach (Fig. 7) and a small, measured gastric pouch is created. The band is sutured in position to prevent it from slipping. Vertical banded gastroplasty and gastric bypass have been carried out by laparoscopy. The operations are difficult and still in the development stages. Small series are now being reported with good safety and satisfactory weight loss. The procedure is most easily learned by surgeons who carry out laparoscopic hiatus hernia repair. Surgeons or physicians familiar with the selection and follow-up standards for obesity surgery must be involved. These procedures may be difficult or impossible in very obese patients with a body mass index greater than 50.

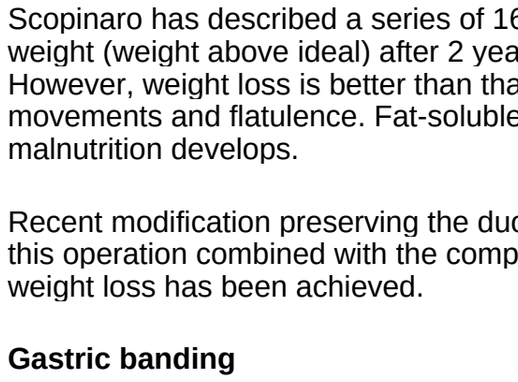


Fig. 7. Laparoscopic silicone band (Lap-Band®) (reproduced with permission of BioEnterics Corporation).

Complications

Mortality should be less than 0.5 per cent for these operations. Large series with no postoperative deaths have been reported. The most common causes of death are leak with sepsis and pulmonary embolism. Leak can occur from a staple line or anastomosis, beneath a prosthetic band, or from distended or ischaemic gastric

pouches. Severe abdominal or left shoulder pain raise concern. Clinical examination of severely obese patients is difficult so postoperative fever and tachycardia are important signs. Abdominal ultrasound, CT scan, and soluble contrast studies may be needed but are difficult in the morbidly obese. Obstruction or leak in the distal gastric pouch after gastric bypass is particularly difficult to identify. Reoperation may be lifesaving. Pulmonary embolism may be minimized by same day admission, subcutaneous heparin, compression stockings, and early ambulation.

Wound infection has been rare with prophylactic antibiotics, short hospital stays, and quick operations. Incisional hernias are common and may be related to high intra-abdominal pressures. Pulmonary complications decrease if smoking stops 6 weeks preoperatively.

Patients who see the surgeon and dietitian regularly have better weight loss and weight maintenance. Thiamine deficiency and Wernicke's encephalopathy may develop quickly in patients with persistent vomiting. Late vitamin B_{12} and iron deficiency are common after gastric bypass. Folate replacement is especially important in women who may become pregnant. Dietary instruction and reduced sugar intake may prevent the dumping syndrome, especially after gastric bypass. Gallstone formation is common during rapid weight loss. Some surgeons perform prophylactic cholecystectomy. Others use ursodiol to prevent gallstones.

Results

Success includes weight loss of 25 per cent or more, absence of major complications, and reversal of obesity-related diseases like type II diabetes mellitus and sleep apnoea. With most procedures the average weight loss of 33 per cent has been achieved 1 year after operation. Best results occur with gastric bypass and biliopancreatic diversion. Late weight regain was common with horizontal gastroplasty and is more common after vertical banded gastroplasty than gastric bypass. Results 5 years after operation are needed to judge any procedure, but are rare. A few excellent long-term results have been published after gastric bypass and biliopancreatic diversion.

Conclusions

The National Institutes of Health Consensus Development Conference on the Surgical Treatment of Obesity in 1991 concluded that surgery for obesity was reasonable for patients who had failed conservative treatment, met strict selection criteria and had operations performed by experienced surgeons working in a group setting. Information is still needed on the natural history of severe obesity. Better long-term data are needed that compare complications and results of different operations. Two large databases have been established: the Swedish Obese Subjects trial and the International Bariatric Surgery Registry in Iowa City. Favourable early results are available but late results are not yet published.

Further reading

Abenhaun L. *et al.* Appetite-suppressant drugs and the risk of primary pulmonary hypertension. *New England Journal of Medicine* 1996; **335**: 609–16.

Black DW, Goldstein RB, Mason EE. Prevalence of mental disorder in 88 morbidly obese bariatric clinic patients. *American Journal of Psychiatry* 1992; **149**: 227–34.

Cleator IGM, Gourley RH. Ileogastrostomy for morbid obesity. *Canadian Journal of Surgery* 1988; **31**: 114–16.

Connolly HM. *et al.* Valvular heart disease associated with fenfluramine-phentermine. *New England Journal of Medicine* 1997; **337**: 581–8.

Favretti F. *et al.* Laparoscopic adjustable silicone gastric banding (Lap-band®): how to avoid complications. *Obesity Surgery* 1997; **7**: 352–8.

Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL. Overweight and obesity in the United States: prevalence and trends, 1960–1994. *International Journal of Obesity* 1998; **22**: 39–47.

Grace DM. Patient selection for obesity surgery. *Gastroenterology Clinics of North America* 1987; **16**: 399–413.

Korkeila M, Kaprio J, Rissanen A, Koskenvuo M. Consistency and change of body mass index and weight. A study on 5967 adult Finnish pairs. *International Journal of Obesity* 1995; **19**: 310–17.

Kuzmak LI. Stoma adjustable silicone gastric banding. *Problems in General Surgery* 1992; **9**: 298–317.

Lagacé M. *et al.* Biliopancreatic diversion with a new type of gastrectomy: some previous conclusions revisited. *Obesity Surgery* 1995; **5**: 411–18.

Lévy E, Lévy P, LePen C, Basdevant A. The economic cost of obesity: the French situation. *International Journal of Obesity* 1995: **19**: 788–92.

MacLean LD, Rhode BM, Forse RA. Late results of vertical banded gastroplasty for morbid and super obesity. *Surgery* 1990; **107**: 20–7.

Mason EE. Vertical banded gastroplasty for obesity. *Archives of Surgery* 1982; **117**: 701–6.

Mason EE, Ito C. Gastric bypass in obesity. *Surgical Clinics of North America* 1966; **47**: 1345–51.

National Institutes of Health Consensus Development Conference. Draft statement on Gastrointestinal Surgery for Severe Obesity, 25–27 March 1991. *Obesity Surgery* 1991; **1**: 257–65.

Payne JH, DeWind L, Schwab CE, Kern WH. Surgical treatment of morbid obesity: sixteen years of experience. *Archives of Surgery* 1973; **106**: 432–7.

Pietmen P, Vartiainen E, Männisto S. Trends in body mass index and obesity among adults in Finland from 1972 to 1992. *International Journal of Obesity* 1996; **20**: 114–20.

Pories WJ. *et al.* Who would have thought it? An operation proves to be the most effective therapy for adult-onset diabetes mellitus. *Annals of Surgery* 1995; **222**: 339–59.

Scopinaro N, Gianetta E, Civalleri D, Bonalumi V, Bachi V. Bilio-pancreatic bypass for obesity: II. Initial experience in man. *British Journal of Surgery* 1979; **66**: 618–20.

Scopinaro N. *et al.* Bilio-pancreatic diversion for obesity. *Problems in General Surgery* 1992; **9**: 362–79.

Sjostrom L, Narbro K, Sjostrom D. Costs and benefits when treating obesity. *International Journal of Obesity* 1995; **19**: S9–S12.

Stevens J, Cai J, Pamuk ER, Williamson DF, Thun MJ, Wood JL. The effect of age on the association between body-mass index and mortality. *New England Journal of Medicine* 1998; **338**: 1–7.

Stunkard AJ, Stinnett JL, Smoller JW. Psychological and social aspects of the surgical treatment of obesity. *American Journal of Psychiatry* 1986; **143**: 417–29.

Sugerman HJ, Starkey JV, Birkenhauer R. A randomized prospective trial of gastric bypass versus vertical banded gastroplasty for morbid obesity and their effect on sweets versus non-sweets eaters. *Annals of Surgery* 1987; **205**: 613–22.

Sugerman HJ. *et al.* A multicenter, placebo-controlled, randomized, double-blind, prospective trial of prophylactic ursodiol for the prevention of gallstone formation following gastric-bypass induced rapid weight loss. *American Journal of Surgery* 1995; **169**: 91–7.

Sugerman H, Windsor A, Bessos L, Wolfe L. Intra-abdominal pressure, sagittal abdominal diameter and obesity comorbidity. *Journal of Internal Medicine* 1997; **241**: 71–9.

Sullivan M. *et al.* Swedish obese subjects (SOS)—an intervention study of obesity. Baseline evaluation of health and psychosocial functioning in the first 1743 subjects examined. *International Journal of Obesity* 1993; **17**: 503–12.

Wolf AM, Colditz GA. Current estimates of the economic cost of obesity in the United States. *Obesity Research* 1998; **6**: 97–106.

Selected overviews

Chagnon YC, Perusse L, Bouchard C. The human obesity gene map: the 1997 update. *Obesity Research* 1998; **6**: 76–92. [A good summary of research in the field.]

Deitel M. *Surgery for the morbidly obese patient*. Lea & Febiger, Philadelphia, 1989. [An excellent review of the field up to the year 1989.]

Mason EE. *Surgical treatment of obesity*. WB Saunders, Philadelphia, 1981. [A classic work describing the first 25 years of surgery for obesity, written by the man who started the gastric procedures.]

Zittel TT. *et al.* High prevalence of bone disorders after gastrectomy. *American Journal of Surgery* 1997; **174**: 432–8. [A warning that late bone disease could result from gastric operations for obesity.]

26.1 Diverticular disease: diverticulitis, bleeding, and fistula

Bruce D. George

Introduction
Pathology
Epidemiology and aetiology
Clinical features and investigation
- **Diverticulosis**
- **Acute diverticulitis**

Treatment
- **Diverticulosis**
- **Acute, uncomplicated diverticulitis**
- **Acute diverticulitis with localized abscess (Hinchey 1 and 2)**
- **Diverticulitis associated with generalized peritonitis (Hinchey stage 3 and 4)**
- **Diverticulitis associated with fistulas**
- **Diverticular stricture**

Surgery for diverticular disease
- **Preoperative preparation**
- **Operative treatment**
- **Postoperative care**

Special conditions
- **Diverticular disease in young patients**
- **Diverticular disease in the immunocompromised**
- **Diverticulitis after previous resection**
- **Right-sided and caecal diverticulitis**
- **Giant diverticulum**

Diverticular haemorrhage
- **Continued active bleeding**
- **Patients who bleed slowly or intermittently**
- **Patients who stop bleeding**
- **Surgery for diverticular haemorrhage**

Summary
Further reading

Introduction

Diverticula are outpouchings from the lumen of a viscus and may occur throughout the gastrointestinal tract. True diverticula, such as a Meckel's diverticulum, contain all layers of the bowel wall and represent a congenital abnormality. False diverticula are acquired herniations of mucosa through the muscle layer and occur most frequently in the colon.

Diverticular disease of the colon is a common condition, especially in the developed world, accounting for about 200 000 hospital admissions per year in the United States of America. The clinical spectrum ranges from asymptomatic to generalized peritonitis with multisystem failure. Optimal management depends on accurate diagnosis, an estimation of the likely natural history, and tailoring treatment accordingly.

The term diverticulosis is used to describe colonic diverticula with no associated inflammation. Diverticulitis indicates associated inflammation. The term diverticular disease may be applied to all stages of the disorder from diverticulosis to complicated diverticulitis.

There was little mention of colonic diverticular disease in the nineteenth century, although Habershon gave a remarkably accurate account in 1857:

> … pouches of the colon which sometimes becomes a considerable size … the orifices of these small sacs are bounded by the hypertrophied circular and longitudinal fibres and their contents remain almost shut off. These pouches are the result of constipation, muscular fibres become hypertrophied, but their effort to propel onward their contents lead to these mini-hernial protrusions.

Pathology

Diverticular disease (see Table 1) affects the sigmoid colon in over 95 per cent of cases and this is the exclusive site in over 50 per cent (Fig. 1). Total colonic involvement occurs in 5 per cent of cases (Fig. 2). Predominantly right-sided disease is unusual in Western populations, but is the dominant pattern amongst Oriental peoples.

Condition	Comment
Pre-diverticular state(myochosis)	Muscular thickening predates appearance of diverticula
Diverticulosis:	
Asymptomatic	
Symptomatic	Difficult to be certain that symptoms are due to diverticula
Diverticulitis:	
Uncomplicated	
Acute complicated	Hinchey classification
Fistula formation	
Stricture formation	
Diverticular haemorrhage	Usually no associated inflammation

Table 1 Spectrum of diverticular disease

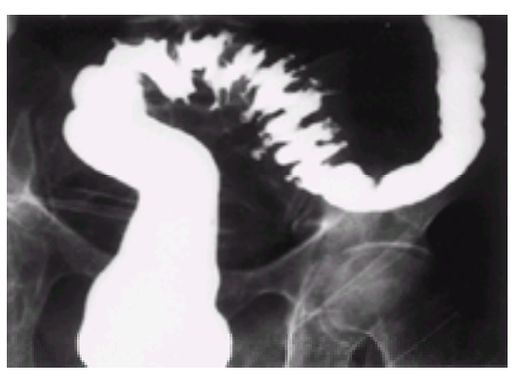


Fig. 1. Barium enema showing diverticulosis localized to the sigmoid colon.

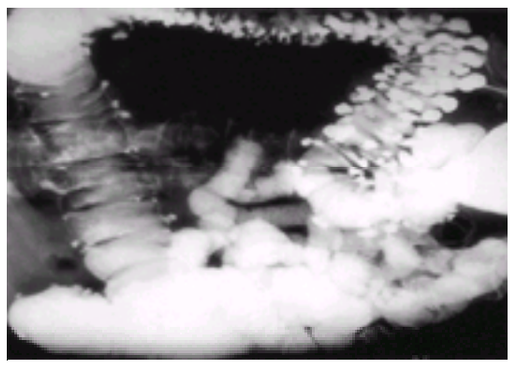


Fig. 2. Barium enema showing diffuse diverticulosis of the colon.

Diverticula usually occur between the single mesenteric taenia and one of the antimesenteric taenia, at sites where blood vessels penetrate the muscle wall. They consist of mucosa, including muscularis mucosa, projecting through the circular smooth muscle and lying in the pericolic fat or appendices epiploica (Fig. 3). They are usually filled with faeces or mucin. Diverticula are invariably associated with thickening of the colonic smooth muscle. This muscular thickening, termed myochosis, is believed to predate the appearance of diverticula. The term ‘pre-diverticular state' has been used to describe myochosis without recognizable diverticula.

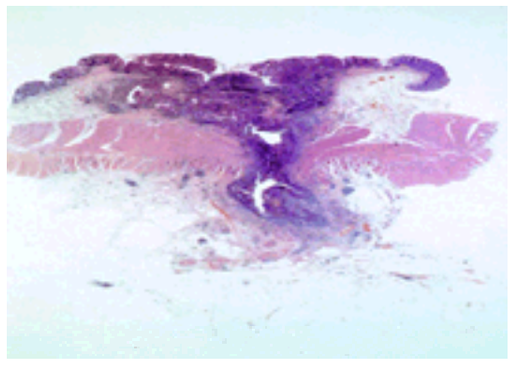


Fig. 3. Section of colon affected by diverticulosis showing false diverticulum. Note the associated muscle wall thickening and prominent mucosal folds.

Macroscopically, diverticulosis is characterized by the triad of false diverticula, muscular thickening, and redundant folds of mucosa

The main complications of diverticulosis are diverticulitis and bleeding. Acute diverticulitis is characterized by swelling and erythema of the affected segment of colon. Persistent localized inflammation may result in a thickened, firm segment to which the term phlegmon may be applied. Localized diverticulitis, but without abscess formation, fistulation, free perforation or ischaemia, is termed uncomplicated diverticulitis. Perforation may lead to the formation of a pericolic abscess, other contained intra-abdominal abscesses, or generalized peritonitis. A useful grading system for the degree of perforation has been devised by Hinchey (Table 2).

Stage	
1	Diverticulitis with associated pericolic abscess
2	Diverticulitis associated with distant abscess (retroperitoneal or pelvic)
3	Diverticulitis associated with purulent peritonitis
4	Diverticulitis associated with faecal peritonitis

Table 2 Hinchey staging of complicated diverticulitis

Fistulas occur when a contained abscess, often around the apex of the sigmoid loop, finds a route of drainage into an adjacent organ such as bladder, vagina, small bowel, or skin.

Recurrent episodes of diverticulitis may eventually produce stricturing of the colon.

Haemorrhage, which may be massive, results from rupture of one of the colonic blood vessels that runs alongside the diverticulum. It usually occurs in the absence of inflammation.

Epidemiology and aetiology

Diverticular disease is a disease of the modern elderly ‘Western' population. Numerous studies have shown a progressive increase in prevalence throughout the twentieth century. Diverticular disease occurs in about 5 to 10 per cent of people in their 40s, increasing to about 70 per cent of people in their 80s. The prevalence is much greater in the developed world than in less industrialized nations. Studies of migrant groups, for example Japanese immigrants to Hawaii, have shown a dramatic increase in diverticular disease with the shift to a ‘Western' environment.

The widely held view is that lack of dietary fibre is the major contributing factor in the development of colonic diverticula. The steep increase in prevalence in England has been attributed to the introduction of roller milling of wheat flour, in about 1880, resulting in a lower intake of crude cereal grains. Major differences in dietary fibre intake probably explain the observed geographical differences. One study demonstrated that the prevalence of symptomless diverticulosis, as demonstrated by barium enema, was lower in vegetarians (12 per cent) than non-vegetarians (33 per cent). Dietary fibre intake was approximately double in the vegetarian group.

Formation of diverticula is likely to be due to a combination of high intraluminal pressures and areas of weakness in the colonic wall. The colon is unusual in that the longitudinal muscle layer is condensed into three bundles, the taenia coli. The areas between the taenia, especially at the sites of entry of blood vessels, are natural sites of weakness. The sigmoid colon is the narrowest part of the colon and therefore (Laplace's Law) generates the highest intraluminal pressures. Arfdwidsson has reported higher intraluminal pressures in the sigmoid colon in patients with diverticular disease than in those with no diverticula. Painter, using a combination of pressure recordings and cine-radiology, demonstrated segmentation of the colon, with localized regions of high pressure, which he argued contributed to pulsion diverticula. Diminished stool bulk from diets deficient in fibre results in altered gastrointestinal transit time and increased intraluminal pressure.

The relative importance of weakness of the colonic wall in the formation of diverticula is uncertain. Several studies have shown progressive changes in elastin and collagen in the colon with increasing age. Abnormalities in elastin content have been reported in diverticular disease. Reported associations with Marfan syndrome, Ehlers–Danlos syndrome, and polycystic renal disease support the concept of a connective tissue abnormality predisposing to the formation of diverticula.

There is recent evidence from case–control studies that patients with complicated diverticulitis are more likely to be taking non-steroidal anti-inflammatory drugs than controls and that their consumption is associated with a more severe form of the disease.

Clinical features and investigation

Diverticulosis

The majority of people with diverticulosis are entirely asymptomatic. Diverticula may be discovered by barium enema examination, colonoscopy (Fig. 4), computed tomography (**CT**) (Fig. 5), laparoscopy, or laparotomy performed for unrelated reasons.

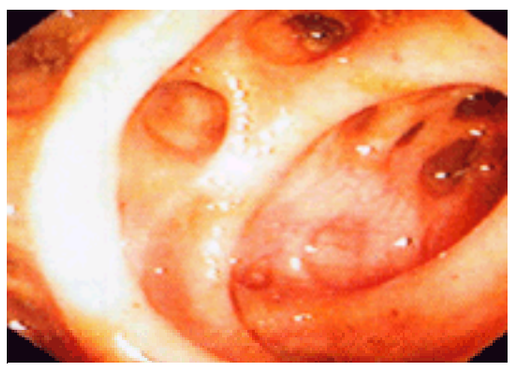


Fig. 4. Colonoscopic view showing multiple diverticula.

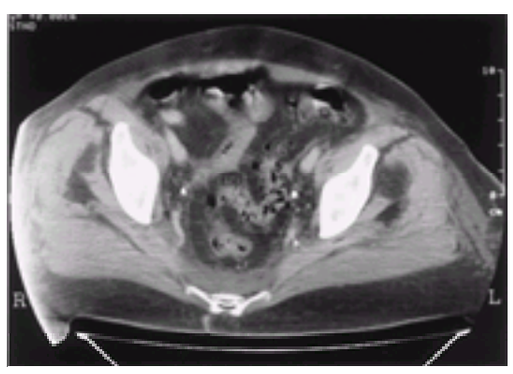


Fig. 5. CT scan showing diverticula in the sigmoid colon

The natural history of diverticulosis is uncertain. Horner followed 503 patients for an average of 5.6 years and suggested that the incidence of diverticulitis was 17 per cent. Kubo and colleagues followed 1124 patients with diverticular disease shown on barium enema; over 15 years, 2.4 per cent developed diverticulitis. Since both of these populations probably included symptomatic patients, the risk of developing diverticulitis in truly asymptomatic diverticulosis is probably substantially smaller.

It is difficult to know what symptoms may be attributed to the mere presence of diverticula (without inflammation). Stimulation of visceral afferent pain fibres by stretching or increased tension would be expected to cause abdominal pain. Given that intraluminal pressures are raised in patients with diverticular disease, it is possible that visceral-type pain may occur in diverticulosis. Visceral pain from the descending or sigmoid colon, embryologically part of the hindgut, would be expected to be distributed diffusely in the lower abdomen. Visceral pain tends to be associated with nausea, vomiting, or sweating. However, very similar symptoms may occur in irritable bowel syndrome, and in urinary or gynaecological disease. In summary, diverticulosis may cause symptoms, but in any individual patient it is very difficult to be certain that their symptoms are due to diverticula alone.

Acute diverticulitis

Acute diverticulitis affects the sigmoid or descending colon in at least 85 per cent of cases. The clinical features resemble ‘left-sided appendicitis'. Pain starts in the lower abdomen, tending to localize in the left iliac fossa. Associated anorexia, nausea, fever, and altered bowel habit are common features. If the inflamed segment of colon is adjacent to the bladder, urinary symptoms occur. If the inflamed sigmoid is ‘flopped' over to the right side, the clinical presentation is difficult to distinguish from appendicitis.

On examination, patients are flushed, pyrexial, and tachycardic. Abdominal examination reveals tenderness and guarding over the affected segment of colon, usually in the left iliac fossa. If the inflamed loop of sigmoid is situated deep in the pelvis, tenderness may only be elicited by rectal or vaginal examination. In patients who have had symptoms for several days, it is not uncommon to feel a tender mass, comprising the thickened, inflamed segment of colon and adherent omentum. Elderly and immunocompromised patients tend to present with more non-specific features, with few features to localize the condition to the abdomen.

Basic investigations include urinalysis, routine haematological and biochemical tests, and plain abdominal and erect chest radiographs. A leucocytosis supports the diagnosis but is not invariably present. Routine blood tests are required to exclude other pathology such as pancreatitis and to detect significant comorbidity, such as anaemia and renal or hepatic dysfunction. Chest radiography may demonstrate important comorbidity, an alternative explanation for left-sided abdominal pain such as pneumonia of the left lower lobe or free air below the diaphragm indicative of free perforation. Plain abdominal radiography rarely adds support to the diagnosis, although a few dilated loops of small bowel adjacent to the inflamed colon are not unusual.

In all patients with suspected acute diverticular disease (except those with mild disease), it is important to confirm the diagnosis objectively. Water-soluble contrast enema (e.g. gastrografin) or CT are the best two means of doing so. The use of barium enema in the acute setting is contraindicated because of the risk of producing barium peritonitis, which has a high mortality. Criteria for the diagnosis on water-soluble contrast enema include diverticula, mass effect, intramural mass, sinus track, and extravasation. As diverticulitis is predominantly a extramucosal disease, contrast radiology may underestimate the severity of the problem. Recently, contrast CT has become the investigation of choice in the acute assessment of patients with suspected acute diverticulitis. Localized thickening of the colonic wall (over 4 mm) in association with diverticula and inflammation of adjacent pericolic fat and mesentery support the diagnosis (Fig. 6). CT has the advantage over water-soluble contrast enema of more accurately assessing complications such as abscess formation or detecting other conditions that may be causing the patient's symptoms.

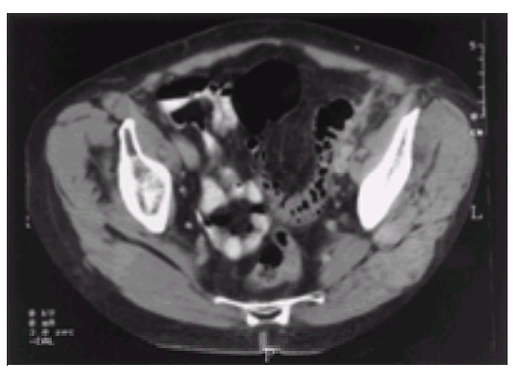


Fig. 6. CT scan showing acute diverticulitis: note thickening of wall of sigmoid colon, presence of diverticula, and streaking of adjacent mesentery.

Complicated acute diverticulitis

Abscess

Abscess formation in association with acute diverticulitis may occur adjacent to the affected colon (pericolic abscess), in the adjacent mesentery (mesocolic abscess), or separately in the peritoneal cavity, such as in the pelvis (Fig. 7) or subphrenic areas.

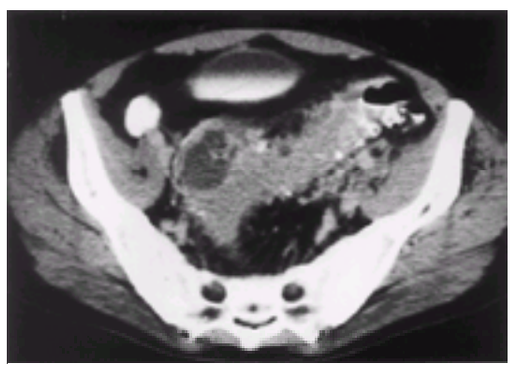


Fig. 7. CT scan showing sigmoid diverticulitis and pelvic abscess.

The clinical features of a pericolic or mesocolic abscess are very similar to those of uncomplicated diverticulitis. Patients tend to be more unwell systemically and may have a high, swinging fever. The diagnosis is usually made by contrast CT. A pelvic abscess may be diagnosed by the finding of a tender ‘boggy' swelling in the pelvis on rectal examination and confirmed by CT or ultrasonography. A subphrenic abscess is difficult to locate on purely clinical grounds. Hiccoughing, shoulder-tip pain, or an associated pleural effusion are clues, although again the diagnosis is usually made by radiological imaging.

Generalized peritonitis

Generalized peritonitis may occur due to free perforation of a diverticulum, resulting in faecal peritonitis or rupture of an initially contained abscess leading to purulent peritonitis.

Patients will complain of severe, generalized abdominal pain, usually of sudden onset, exacerbated by unguarded movements. On examination the patient will look unwell, will be lying still with shallow respirations, and will have a tachycardia and probably pyrexia. Generalized peritonitis, especially faecal, may be associated with features of multiorgan failure such as hypotension, breathlessness, and oliguria. On inspection the abdomen will not be moving with respiration; palpation reveals generalized tenderness and guarding; bowel sounds are absent.

The diagnosis of generalized peritonitis is usually clinical. Laboratory investigations are helpful in assessing the degree of systemic upset and comorbidity. An erect chest radiograph usually demonstrates free gas, indicative of perforation. In unequivocal cases more detailed radiological imaging is not required. In equivocal cases, particularly the elderly or immunocompromised individual, CT may be helpful.

Retroperitoneal perforation

Occasionally, sepsis from acute diverticulitis spreads posteriorly into the retroperitoneal tissues. The patient may have features of diverticulitis plus back or loin pain. Retroperitoneal sepsis may spread inferiorly, for example along the psoas muscle, presenting as an inflamed area or abscess in the groin. Sometimes the retroperitoneal features may overshadow the abdominal signs and symptoms. Tenderness, redness or swelling in these areas, in the context of a patient with possible diverticulitis, should alert the clinician to the possibility of retroperitoneal sepsis. Pain on hyperextension of the hip suggests inflammation in the region of the psoas muscle. Radiologically, gas may be seen in the soft tissues of the retroperitoneum on plain abdominal radiographs. The diagnosis is usually confirmed by CT (Fig. 8).

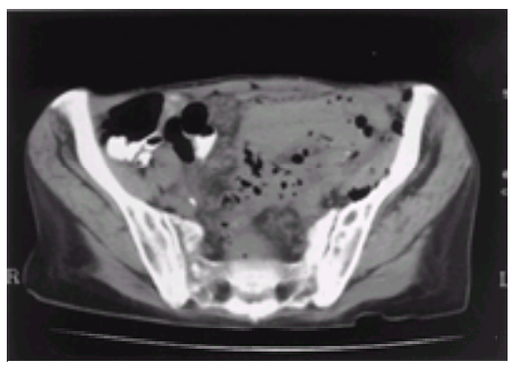


Fig. 8. CT scan showing diverticulitis with retroperitoneal gas within iliacus muscle.

Fistulas

A fistula is an abnormal communication between two epithelialized surfaces. An abscess in association with diverticulitis may find a spontaneous route of drainage into another organ, most commonly the bladder, vagina, small bowel, or skin. Having done so a fistula is formed. The clinical features depend on the site of the fistula, as follows.

Colovesical Bubbles in the urine (pneumaturia) are diagnostic of a colovesical fistula. Overt faecaluria is rare, although ‘bits' or sediment in the urine are suggestive of a fistula. Recurrent urinary-tract infections, especially if faecal organisms are grown, should suggest a colovesical fistula. Radiologically, gas in the bladder (in the absence or recent catheterization or instrumentation) on plain radiography, ultrasound or CT suggests a fistula. The precise site of a fistula is delineated by contrast studies, usually barium enema (Fig. 9).



Fig. 9. Barium enema showing colovesical fistula.

Colovaginal This usually presents with an obvious faeculent vaginal discharge. It occurs more commonly in women who have had a hysterectomy. The fistula is usually demonstrated by barium enema, although occasionally a contrast study per vagina is diagnostic. Fistulation into the uterus or fallopian tubes is rare (Fig. 10).

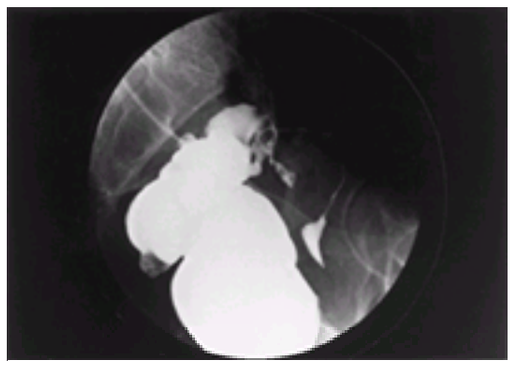


Fig. 10. Barium enema showing fistula from sigmoid colon to fallopian tube, uterus, and vagina (by courtesy of Dr D. Nolan, Oxford).

Enterocolic A fistula into the small bowel usually presents with diarrhoea due to large-volume small-bowel contents passing directly into the colon. The degree of symptoms will depend on the level of small bowel (and colon) affected. The diagnosis is made by contrast radiology, either a small-bowel or a barium enema.

Colocutaneous This usually starts as a subcutaneous abscess, which once drained (either surgically or spontaneously) drains faecal material (Fig. 11). Alternatively, a colocutaneous fistula may develop spontaneously at the site of a previous surgical scar or drain site.

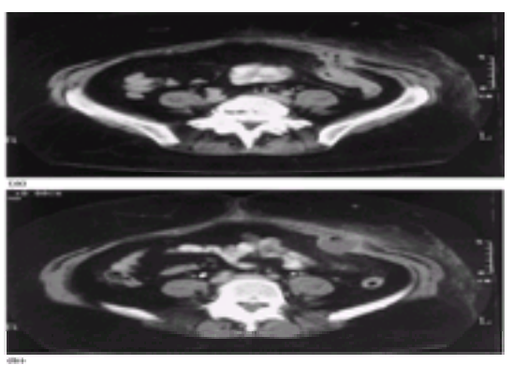


Fig. 11. (a) and (b) CT scans showing subcutaneous abscess related to underlying diverticulitis.

Colonic stricture

Repeated episodes of mild diverticulitis may result in progressive fibrosis and narrowing of the affected segment. This may present gradually with a change in bowel habit, crampy abdominal pain, and distension. Alternating constipation and diarrhoea is characteristic of a sigmoid stricture. The major differential diagnosis is of colonic malignancy. A stricture is usually identified by barium enema or colonoscopy. Radiologically, a diverticular stricture has a tapered appearance, without the shouldered ends suggestive of malignancy. Colonoscopically, diverticular strictures appear as a narrowed segment without obvious mucosal irregularity. Malignancy is difficult to rule out entirely. Multiple biopsies and brushings for cytology are helpful.

Alternatively a diverticular stricture may present as an emergency with complete large-bowel obstruction. A water-soluble contrast enema is important to confirm complete mechanical obstruction and rule out pseudo-obstruction. Differentiation between diverticular and malignant large-bowel obstruction may, however, be difficult before resection.

Treatment

Diverticulosis

Patients with diverticulosis are advised to take to a high-fibre diet, although evidence from randomized, controlled trials that this reduces the risk of diverticulitis is lacking. Several uncontrolled studies and one small, randomized, controlled trial by Brodribb suggest that a high-fibre diet reduces symptoms in patients with symptomatic diverticular disease. One recent, multicentre, randomized, controlled trial has suggested significant benefit from the poorly absorbed antibiotic rifaximin in symptomatic diverticular disease, although further studies are needed to substantiate this.

A recent prospective study of over 47 000 American men aged 40 to 75–years, without known colonic disease, suggested that increased physical activity (in addition to a high-fibre diet) protects against the development of symptomatic diverticular disease.

Anticholinergic drugs, such as propantheline bromide (15 mg three times a day), and spasmolytic agents, such as mebeverine hydrochloride (135 mg three times a day), are commonly prescribed for symptomatic diverticulosis, although their efficacy is unproved.

Acute, uncomplicated diverticulitis

Acute diverticulitis should be treated by bowel rest and antibiotics. Outpatient treatment with clear fluids and antibiotics orally is appropriate for patients with mild disease. Patients with more severe disease or significant comorbidity require inpatient treatment with intravenous fluids and antibiotics. Antibiotics with adequate Gram-negative and anaerobic cover are required. Cefuroxime (750 mg three times a day) and metronidazole (500 mg three times a day) is a frequently used

combination.

Early confirmation of the diagnosis by CT (as discussed above) is helpful in the management of these patients. The vast majority with uncomplicated acute diverticulitis will settle with conservative treatment.

Once the acute episode has settled, usually 4 to 6 weeks later, it is important to visualize the colon by colonoscopy or barium enema to rule out a coexisting neoplasm. Subsequent treatment is with a high-fibre diet, which may reduce the risk of further acute episodes. Surgery is required occasionally when the acute episode fails to respond to conservative treatment and should be considered after recurrent episodes of acute diverticulitis (see below).

Acute diverticulitis with localized abscess (Hinchey 1 and 2)

All patients should be treated with intravenous fluids, appropriate antibiotics, and general supportive medical care in the first instance. Small abscesses may respond to such treatment alone. All other abscesses require drainage. In general it is preferable to drain abscesses percutaneously, under radiological guidance (Fig. 12), followed by elective resection, rather than undertake urgent surgery. Approximately 70 to 90 per cent of diverticular abscesses that are amenable to CT-guided, percutaneous drainage are successfully treated in this way. Acute surgical intervention is indicated if the patient does not respond to conservative therapy and percutaneous drainage, or if the abscess is inaccessible or too multiloculated to be adequately drained percutaneously.



Fig. 12. CT scan showing percutaneous drainage of pelvic abscess.

Once the acute episode has subsided the colon should be visualized by colonoscopy or barium enema.

Elective surgery should be considered after resolution of an acute episode of complicated diverticulitis.

Diverticulitis associated with generalized peritonitis (Hinchey stage 3 and 4)

Patients with generalized peritonitis require immediate resuscitation followed by prompt surgery. Surgical options are discussed below.

Diverticulitis associated with fistulas

Patients with diverticular-associated fistulas should be thoroughly investigated to determine the site and complexity of the fistula and to rule out other fistula-causing diseases such as malignancy or Crohn's. Most patients may be operated on in an elective setting, after medical optimization and full bowel preparation. The general principle is to resect the diverticular segment of bowel and to repair the defect in the contiguous structure. A primary anastomosis is usually appropriate.

Diverticular stricture

Complete large-bowel obstruction due to a diverticular stricture requires rapid resuscitation followed by prompt surgery. The standard surgical options are discussed below. Recently, experience with expandable metal stents to relieve acute large-bowel obstruction has been reported. This technique may obviate the need for emergency surgery and allow planned surgery to be undertaken when the patient is in a more favourable condition.

A diverticular stricture presenting less acutely can usually be operated on in a planned fashion after thorough investigation and full bowel preparation.

Surgery for diverticular disease

The indications to operate as an emergency, urgently, or soon for complications of diverticular disease are fairly clear (Table 3). Elective surgery after resolution of acute complications is more controversial. Clearly this must take account of the patient's biological age and comorbidity, the number and severity of attacks of diverticulitis, their speed of response to medical treatment, and the magnitude of symptoms between attacks. There is some evidence that recurrent attacks of acute diverticulitis become increasingly less likely to respond to medical therapy. A prospective audit of 120 patients treated for complicated diverticular disease in the United Kingdom demonstrated that 70 (58 per cent) were asymptomatic at 5 year follow-up and 50 (42 per cent) were symptomatic; 39 developed serious complications, of whom 10 died as a result. The investigators advocated elective surgical resection for patients who have survived an acute episode of complicated diverticular disease managed conservatively, provided that they are fit for elective surgery. Similarly, the American Society of Colon and Rectum Surgeon's (**ASCRS**) guidelines (1995) are that elective resection is indicated after two proven episodes of acute, uncomplicated diverticulitis or after one episode of complicated acute diverticulitis.

Emergency	Generalized peritonitis
	Large-bowel obstruction with distended caecum
	Massive haemorrhage
Urgent	Acute diverticulitis unresponsive to medical treatment (rare)
	Acute diverticulitis with localized abscess unresponsive to medical/radiological treatment
	Large-bowel obstruction
Soon	Diverticular-associated fistula
	Diverticular stricture
Elective	Two episodes of proven acute diverticulitis treated conservatively
	One episode of acute, complicated diverticulitis treated conservatively

Table 3 Indications for surgery in diverticular disease

Preoperative preparation

Patients should be prepared for surgery as thoroughly as the clinical situation permits. Patients with generalized peritonitis require correction of dehydration and major electrolyte abnormalities, such as hypokalaemia, as rapidly as safely possible. Cardiac arrhythmias causing circulatory compromise, such as fast atrial fibrillation, should be corrected. All patients should be given broad-spectrum antibiotics and prophylaxis for deep venous thrombosis preoperatively.

In less acute circumstances, patients should undergo full bowel preparation. They should be seen by a stomatherapist and optimal stoma sites marked preoperatively.

Operative treatment

Operative treatment of diverticular disease may be difficult; sepsis, scarring, foreshortening, thickening and narrowing of the bowel, associated abscesses, obstruction, and fistulas present challenges that require both technical expertise and mature judgement.

Patients should be positioned in the modified Lloyd-Davies position. A midline incision is the most conventional, although some surgeons prefer a left paramedian incision or a curved oblique incision in the left iliac fossa. Laparoscopic and laparoscopically assisted resections are becoming increasingly popular, although learning such procedures is difficult and to date there are no data from randomized trials that there is any benefit compared to laparotomy.

There is a wide spectrum of surgical procedures to chose from when operating for diverticular disease (Table 4). A major skill of the surgeon is to select the correct procedure for the individual patient.

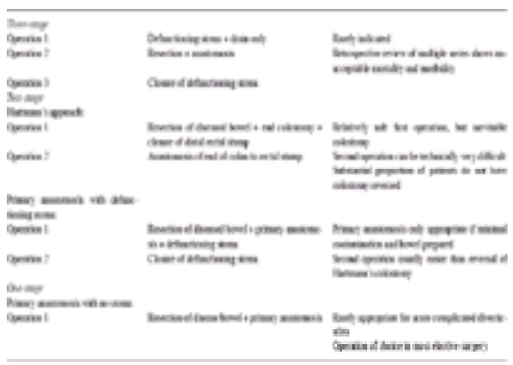


Table 4 Surgical options in diverticular disease

Most elective operations for diverticular disease involve resection and primary anastomosis (one-stage approach). The operation of sigmoid myotomy, popularized by Reilly in the 1970s, is no longer performed.

The choice is more difficult in an emergency or urgent situation.

The three-stage approach was once favoured for diverticulitis associated with generalized peritonitis. However, this approach has fallen into disrepute: the septic focus is not removed; multiple operations in elderly patients result in high cumulative mortality, morbidity, and total hospital stay, and many patients fail to complete all three stages. A major review of the operative management of acute, complicated diverticular disease by Krukowski and Matheson emphasized the need to resect the diseased segment of colon. However, their conclusions were largely based on retrospective studies, which may be prone to consistent biases. One small, randomized trial by Kronborg favoured transverse colostomy and suture of the colonic perforation over acute resection without a primary anastomosis for purulent peritonitis. Despite this, most surgeons (and the ASCRS guidelines) favour resection of the diseased segment of colon at the initial operation. The most frequent surgical decision is whether or not to do a primary anastomosis or whether to exteriorize the proximal colon and close off the distal rectal stump (Hartmann's procedure). Opinions differ in this matter: proponents of the Hartmann's approach argue about the safety of avoiding an anastomosis in the presence of sepsis; opponents emphasize the difficulty of reversing a Hartmann's procedure and point out that up to one-third of patients never have their colostomy reversed. No randomized trials have addressed this question.

The ASCRS Standards Task Force (1995) favours a Hartmann's procedure for diverticulitis with free perforation (Hinchey stage 3 and 4). Acute surgery for diverticulitis associated with a localized abscess is more debatable. If contamination is minimal and adequate bowel preparation can be achieved, resection and primary anastomosis, with or without a defunctioning stoma, is appropriate. Otherwise a Hartmann's procedure should be performed. If a defunctioning stoma is utilized, the evidence favours loop ileostomy over a transverse colostomy.

When operating for acute large-bowel obstruction, the Hartmann's procedure is often appropriate. A one-stage approach with on-table lavage and primary anastomosis is favoured by some surgeons, particularly in the United Kingdom, although there are no trials comparing these techniques. In the very frail or unstable patient, a preliminary transverse colostomy (three-stage approach) may be appropriate. If the ileocaecal valve is competent and the caecum critically distended, a subtotal colectomy with ileosigmoid or ileorectal anastomosis may be undertaken. Although this usually permits a primary anastomosis with no defunctioning stoma, the increased bowel frequency afterwards may be troublesome, especially in the elderly individual.

The spleen and left ureter are at particular risk during resections for diverticular disease. Accidental splenic injury and subsequent splenectomy increase the risk of venous thromboembolic complications and permanently impair immunity. The left ureter may be adherent to the sigmoid mesentery or bound in dense scar tissue in the region of the pelvic brim, sometimes with partial ureteric obstruction. Preoperative placement of a left ureteric stent may help to identify the ureter in an area of inflammation or dense scarring. Ureteric injury can be avoided by dissecting the mesentery of the descending colon from Gerota's fascia, identifying the ureter near the pelvic brim, then dissecting downward through the difficult area, keeping the ureter in a posterior position and always in view.

The length of colon resected depends upon the extent of the muscular abnormalities in the gut wall as well as the extent of the inflammatory changes. Resection for sigmoid diverticular disease should extend distally to the upper rectum.

Anastomoses may be hand-sewn or stapled, depending on the surgeons' preference. Comparative trials show no significant difference.

Postoperative care

Patients should receive standard postoperative care. Where there is established infection, antibiotics should be continued for 5–days; otherwise, prophylactic antibiotics only should be given perioperatively. Thromboprophylaxis should be continued in all patients until they are fully mobile.

Postoperative complications are not uncommon, particularly when surgery is undertaken acutely in elderly patients. Reported mortality rates from faecal and purulent peritonitis are 48 per cent and 27 per cent, respectively. General complications include respiratory infection, cardiac events, thromboembolic disease, urinary-tract infection or retention, renal impairment, and strokes. Specific complications include haemorrhage, intra-abdominal abscesses, ileus, small-bowel obstruction, wound infection, and stoma-related problems. Ureteric or splenic damage is rare. The most feared postoperative problem is sepsis: this may be due to continuation of pre-existing infection or may occur because of an anastomotic leak postoperatively. Several studies have shown that the mortality rate from complicated diverticulitis is proportional to the degree of peritonitis at operation. Patients with generalized peritonitis are at high risk of developing multisystem failure, which requires support under intensive care. Problems related to anastomotic leakage typically present 5 to 7–days postoperatively. The clinical features vary from obvious peritonitis to non-specific malaise or confusion (Table 5). It is important to have a high index of suspicion for septic complications following surgery for complicated diverticulitis. CT scanning, or contrast enema if a recent anastomosis has been fashioned, are the principal investigations. Early reoperation with peritoneal lavage, stoma formation, and drainage is usually required. In cases of extremely severe intra-abdominal contamination, leaving the abdomen completely open as a laparostomy is an option.

Generalized peritonitis
Intra-abdominal abscess
Prolonged postoperative ileus
Spike in temperature at day 5–7 postoperatively
Non-specific failure to progress well postoperatively
Confusion

Table 5 Clinical presentations of postoperative abdominal sepsis

Special conditions

Diverticular disease in young patients

Diverticulosis occurs in about 6 to 9 per cent of the population under 40 years of age. Approximately 11 to 30 per cent of proven cases of acute diverticulitis occur in the same age group. Some reports suggest an association with obesity in the young. Acute diverticulitis tends to be misdiagnosed in the young, particularly as appendicitis or pelvic inflammatory disease in women. Many retrospective reviews suggest that acute diverticulitis in the young is a more aggressive disorder than in older patients. Eusebio and Eisenberg reported that two-thirds of 181 patients presenting with diverticular disease under 40–years ultimately required surgery. However, this may be a false impression, given that acute diverticulitis in the young (particularly milder episodes) tends not to be diagnosed accurately and younger patients are less likely to require hospital admission for comorbidity. The management principles are similar to those for older patients, except that many surgeons would recommend elective resection after one episode of proven, uncomplicated diverticulitis in the young.

Diverticular disease in the immunocompromised

Immunocompromised patients are encountered increasingly in modern hospital practice. Common causes include congenital or acquired immune disorders, uraemia, haematological malignancies, and immunosuppressant drug therapy, especially steroids. The mortality of diverticulitis in immunocompromised patients is much greater than in immunocompetent patients. Review of 90 immunocompromised patients with acute diverticulitis (from eight published series) demonstrated a surgical mortality of 40 per cent. Reasons for this include delayed diagnosis due to masked clinical features, higher rates of free perforation due to impaired ability to localize sepsis, impaired immune function, and coexistent medical problems. Management therefore requires a higher index of suspicion of complications than usual. A lower threshold for investigation, particularly CT and repeat CT, is required. Surgery in the acute phase should not involve an anastomosis. First-line antibiotic therapy should be the same as in immunocompetent patients, although fungal or cytomegalovirus infection should be considered if the clinical state is not improving.

Patients with polycystic kidney disease have a very high frequency of diverticulosis, 33 per cent of whom develop diverticulitis at some stage. There may be a case for elective resection for diverticulosis in this defined group.

Diverticulitis after previous resection

Recurrent diverticulitis following previous resection is fortunately rare. The most important cause is retention of some distal sigmoid colon after a sigmoid colectomy. Recurrent diverticulitis is reported in 12.5 per cent of patients following sigmoid colectomy when the distal sigmoid is retained, compared to 6.7 per cent when the anastomosis is to the upper rectum.

In cases of assumed recurrent diverticulitis after previous resection, it is important to confirm both the original and current diagnoses. The original pathology specimen should be reviewed, particularly to rule out Crohn's disease. The current diagnosis should be confirmed by CT in the acute phase and subsequently by colonoscopy once the acute phase has subsided. Elective re-resection should be considered after a complicated episode or repeated, uncomplicated attacks. Repeat surgery may be technically difficulty and ureteric stenting is recommended.

Right-sided and caecal diverticulitis

Caecal and ascending diverticula account for the majority of diverticula in Oriental populations. Right-sided diverticula tend to occur at a younger age than left-sided disease. Interestingly, although migration of Oriental populations to a ‘Western' environment is associated with an increase in prevalence of diverticular disease, the disease is still predominantly right-sided. This suggests that both genetic and environmental factors influence the development of diverticular disease.

Right-sided diverticula tend to be multiple and false, although a solitary true diverticulum is a well-recognized entity.

Caecal diverticulitis presents with clinical features similar to acute appendicitis, the diagnosis usually being made intraoperatively. A locally inflamed, easily identifiable caecal diverticulum is best treated by diverticulectomy, drainage, and antibiotics. The more common operative finding is of an inflammatory mass in the caecum that is difficult to distinguish from a locally perforated neoplasm. A right hemicolectomy is the usual treatment. Generalized peritonitis due to a perforated, right-sided diverticulum is usually treated by a right hemicolectomy, with or without a primary anastomosis, depending on the clinical situation.

Giant diverticulum

Giant diverticula of the colon are extremely rare, but sizes of up to 30 to 40 cm have been reported. They present chronically with bloating and vague discomfort or acutely as a result of perforation, torsion, or infarction. Plain abdominal radiography shows a solitary, gas-filled, cyst-like structure. Confirmation of the diagnosis requires barium enema (Fig. 13) or CT. Treatment is resection, either as a one-stage (elective presentation) or two-stage (acute presentation) procedure.

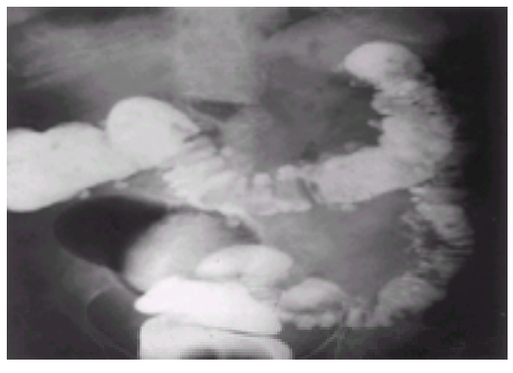


Fig. 13. Barium enema showing giant diverticulum.

Diverticular haemorrhage

It has been estimated that significant bleeding occurs in about 5 per cent of patients with diverticulosis at some stage in their life. Bleeding is characteristically of sudden onset, painless, and severe but may be intermittent or minimal. The main differential diagnoses are angiodysplasia, ischaemic colitis, and neoplasia.

Patients presenting with lower gastrointestinal bleeding should be rapidly assessed and resuscitated as appropriate. A history shouldbe taken, concentrating on the colour and volume of blood loss, other gastrointestinal symptoms, medical comorbidity, and drug therapy. Examination should focus on general condition, features of hypovolaemia, coexisting medical conditions, and clues on the source of bleeding. All patients should have blood sent for routine biochemistry and haematology, including a clotting screen and cross-matching. Reliable venous access should be established and appropriate monitoring commenced. All patients should undergo rectal examination, proctoscopy, and rigid sigmoidoscopy to rule out major anorectal pathology responsible for the bleeding. It should be remembered that approx. 10 to 15 per cent of apparently lower gastrointestinal bleeding is in fact from an upper gastrointestinal (or small bowel) source. A raised urea:creatinine ratio is suggestive of upper rather than lower gastrointestinal bleeding. A nasogastric tube draining bile-stained fluid is a popular but not entirely reliable method of excluding an upper gastrointestinal source. If any doubt exists an upper gastrointestinal endoscopy is mandatory.

Patients with lower gastrointestinal bleeding will tend to fall into one of three categories: those who clearly continue to bleed; those who stop, and those who slow down but continue to trickle or bleed intermittently.

Continued active bleeding

These patients clearly require active resuscitation and simultaneous investigation. Mesenteric angiography (Fig. 14) may identify bleeding at rates of 0.5 ml/min or greater. The pathognomonic finding of diverticular haemorrhage is pooling of contrast in the diverticulum. The main disadvantage of this technique is its invasiveness and that the patient must be actively bleeding at the time of the study. Review of 972 published cases of mesenteric angiography for lower gastrointestinal bleeding showed that a bleeding site was identified in 56 per cent of cases. Angiographic localization allows a limited colonic resection to be undertaken with a very low risk of rebleeding. Angiography may also allows access for therapeutic measures such as the infusion of vasoconstrictive agents (e.g. vasopressin) or for selective embolization of the bleeding vessel. Therapeutic angiography may have a role in stabilizing patients before surgery or in avoiding the need for surgery altogether in selected patients. Although the immediate response rate with vasopressin is good, the risk of rebleeding is high. Selective embolization of the bleeding vessel has been reported, although the poor collateral vasculature of the colon means that ischaemic complications may occur. Recent reports of superselec-tive embolization suggest lower risks of bleeding and fewer complications. In most hospitals however, whenever angiography identifies a bleeding site, surgical resection is felt to be appropriate. If angiography is negative but major bleeding continues, surgical intervention is required; intraoperative colonoscopy after on-table lavage is recommended.

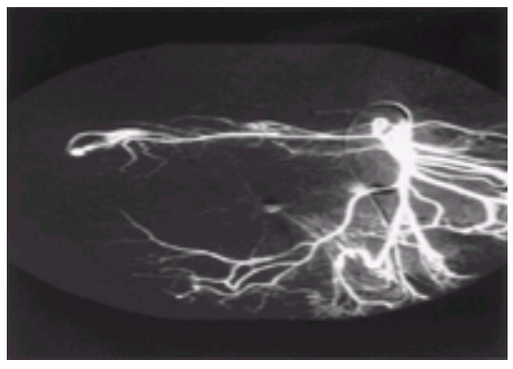


Fig. 14. Angiogram showing bleeding from vessel associated with diverticula at the hepatic flexure.

Patients who bleed slowly or intermittently

Patients who continue to bleed intermittently or who trickle at a slow rate such that angiography is negative are difficult to manage well. Colonoscopy after full bowel preparation is probably the most useful investigation. Scintigraphy using technetium-labelled red blood cells has been reported to identify bleeding at rates as low as 0.05 to 0.1 ml/min, although the overall results from published series are less encouraging. The key to this group of patients is to investigate very thoroughly and repeat investigations if necessary. Colonoscopy, angiography, scintigraphy, upper gastrointestinal endoscopy, small-bowel enema, and abdominal CT are the mainstays of such investigation. If all of these are negative but bleeding persists, laparotomy with panendoscopy is necessary.

Patients who stop bleeding

Patients who present with a lower gastrointestinal bleed who settle rapidly should be investigated electively by colonoscopy, looking principally for diverticula, neoplasia, or angiodysplasia. Approximately 25 per cent of patients who have a diverticular bleed managed conservatively will have another bleed. If recurrent bleeding occurs, resection should be considered, although it is critical to confirm accurately the site of bleeding before laparotomy. Although most diverticula are left-sided, right-sided diverticula account for about 50 per cent of bleeding diverticuloses.

Surgery for diverticular haemorrhage

The majority of patients with diverticular haemorrhage do not require surgical intervention. Patients who require four or more units of blood during the first 24 h of treatment have a 50 per cent chance of needing operative intervention. Preoperative localization of the bleeding point permits a limited rather than extensive colonic resection. Many studies have demonstrated a dramatically reduced mortality and morbidity from such an approach. If the bleeding site cannot be identified preoperatively or operatively, transillumination of the bowel and colonoscopy are recommended to find the source. If no precise bleeding point can be identified and rectal, upper-, and small-bowel causes are excluded, a subtotal colectomy is indicated.

Summary

Diverticular disease is a disorder of the elderly population of the modern world. Lack of dietary fibre is thought to be the dominant cause. The clinical spectrum ranges from no symptoms to generalized peritonitis with multisystem failure. Patients with diverticulosis should be advised to eat a high-fibre diet and to increase regular physical exercise.

Acute diverticulitis and its complications require thorough evaluation; CT is the investigation of choice in acute diverticulitis.

Most cases of acute, uncomplicated diverticulitis and those complicated by localized abscess formation respond to conservative therapy, including percutaneous abscess drainage if necessary. Elective surgery is indicated after recurrent episodes of uncomplicated diverticulitis or after one episode of complicated diverticulitis.

Patients with generalized peritonitis require immediate resuscitation followed by prompt surgical intervention. The diseased segment of colon should be resected at the time of primary surgery.

Diverticular haemorrhage is unusual but may be massive. Patients require immediate assessment and appropriate resuscitation. Mesenteric angiography is the key investigation in the acute phase to identify the precise site of haemorrhage.

Further reading

Aldoori WH *et al.* Prospective study of physical activity and the risk of symptomatic diverticular disease in men. *Gut* 1995; **36:** 276–82. [Prospective cohort study of 47 678 American men, suggesting that physical activity, along with a high-fibre diet, may be an important factor in the prevention of symptomatic diverticular disease.]

Almy TP, Howell DA. Diverticular disease of the colon. *New England Journal of Medicine* 1980; **302**: 324–31. [Detailed review of epidemiology, natural history, and pathogenesis of diverticular disease.]

Ambrosetti P, Grossholz M, Becker C, Terrier F, Morel P. Computed tomography in acute left colonic diverticulitis. *British Journal of Surgery* 1997; **84:** 532–4. [Study of 423 patients with proven diverticulitis demonstrating 97 per cent sensitivity of CT.]

Arfwidsson S. Pathogenesis of multiple diverticula of the sigmoid colon in diverticular disease. *Acta Chirurgica Scandinavica Supplementum*, 1964; **342**: 1–68.

Brodribb AJ. Treatment of symptomatic diverticular disease with a high-fibre diet. *Lancet* 1977; **i**: 664–6. [Small, randomized, controlled trial suggesting that a high-fibre diet reduces symptoms in patients with symptomatic diverticular disease.]

Gear JSS *et al.* Symptomless diverticular disease and intake of dietary fibre. *Lancet* 1979; **i**: 511–14. [Study showing that vegetarians, who eat almost twice as much fibre as non-vegetarians, have a much lower prevalence of asymptomatic diverticular disease.]

Habershon, SO. *Pathological and practical observations on disease of the alimentary canal*. London: Churchill, 1857: 296.

Hinchey EJ, Schaal PG, Richards GK. Treatment of perforated diverticular disease of the colon. *Advances in Surgery* 1978; **12**: 85–109. [Important classification of complicated diverticulitis.]

Krukowski ZH, Matheson NA. Emergency surgery for diverticular disease complicated by generalised and faecal peritonitis: a review. *British Journal of Surgery* 1984; **71**: 921–7. [Major review of the surgical management of peritonitis due to diverticular disease emphasizing the importance of resecting the affected segment of colon at the primary operation.]

Painter NS. Diverticular disease of the colon. A disease of this century. *Lancet* 1969; **ii**: 586–8. [This paper argues that diverticular disease is caused by lack of roughage in the diet.]

Papi C, Ciaco A, Koch M, Capurso L. Efficacy of rifaximin in the treatment of symptomatic diverticular disease of the colon. A multicentre double-blind placebo-controlled trial. *Alimentary Pharmacology and Therapeutics* 1995; **9**: 33–9. [Trial suggesting beneficial effect from the poorly absorbed antibiotic rifaximin in patients with symptomatic diverticular disease.]

Standards Task Force. American Society of Colon and Rectal Surgeons. Practice parameters for sigmoid diverticulitis. *Diseases of the Colon and Rectum* 1995; **38**: 125–32. [Guidelines prepared for the ASCRS addressing the management of acute sigmoid diverticulitis.]

Stevenson ARL, Stitz RW, Lumley JW, Fielding GA. Laparoscopically assisted anterior resection for diverticular disease. *Annals of Surgery* 1998; **227**: 335–42. [Experience of laparoscopically assisted anterior resection in 90 patients, demonstrating acceptable morbidity and mortality rates.]

Welch JP, Cohen JL, Sardella WV, Vignati PV. *Diverticular disease. Management of the difficult surgical case*. Williams and Wilkins, Baltimore, 1998. [Recently published book covering wide range of difficult clinical problems related to diverticular disease.]

Williams NS, Nasmyth DG, Jones D, Smith AH. De-functioning stomas: prospective controlled trial comparing loop ileostomy with loop transverse colostomy. *British Journal of Surgery* 1986; **73**: 566–70. [Randomized trial demonstrating superior results from loop ileostomy compared to loop transverse colostomy in patients undergoing elective colorectal resections.]

26.2.1 Pathology of non-specific chronic idiopathic inflammatory bowel disease

Bryan F. Warren

Ulcerative colitis
Pouchitis
Crohn's disease
Mucosal biopsy and non-specific inflammatory bowel disease
Histopathology in severe colitis
Further reading

The only two pathological features distinguishing absolutely between Crohn's disease and ulcerative colitis are the presence of small-bowel disease other than backwash ileitis and the finding of deep granulomas or granulomatous vasculitis in Crohn's disease. Macroscopically, the presence of fat wrapping is also extremely helpful (see below), but is sometimes difficult to assess in the distal colon. The distinction between ulcerative colitis and Crohn's colitis is based on consideration of the macroscopic disease distribution and the microscopic pattern of the inflammation.

Ulcerative colitis

Ulcerative colitis is a mucosal disease that almost invariably involves the rectum and extends proximally. The cut off between the continuous area of mucosal ulceration and the proximal normal colon is characteristically abrupt. Skip lesions are not generally a feature of ulcerative colitis but they are recognized to occur in two rare and special circumstances, in the appendix and in the caecum, though a degree of caution is needed in differentiating this from Crohn's disease.

The severity of the disease may fluctuate, giving recognizable periods of activity and remission. Remission usually follows treatment. Inflammation is limited to the mucosa, except in fulminant colitis where the mucosa is lost by ulceration and inflammation ex-tends into the underlying muscularis propria. This inflammation is diffuse, with associated destruction of the muscle (myocytolysis). There may be little visible abnormality on the serosal aspect of the bowel on gross inspection; in particular, there is no fat wrapping in ulcerative colitis. The mucosa is usually congested and friable, with a velvety, slightly granular texture. There may be varying degrees of ulceration. Pseudopolyps (inflamed areas of mucosa between ulcers or adjacent to the sites of previous ulcers) may be seen; they have no malignant potential. The left side of the colon is usually affected and improvement following steroid enemas can produce sparing of the rectum or sigmoid. This explanation for any sparing is not, of course, present in all cases; true rectal sparing is rare and is to be diagnosed with caution.

The microscopic changes are also diffuse. They show an irregular mucosal surface with a diffuse pattern of architectural distortion and branching of crypts. Long-standing disease can result in substantial shortening of crypts, identified by the considerable distance between the crypt base and the muscularis mucosae, which are close together in normal mucosa.

There is a diffuse infiltration of chronic inflammatory cells limited to the mucosa (Fig. 1), with prominent basal lymphoid aggregates. The amount of acute inflammation reflects the disease activity, and also tends to be diffuse. Goblet-cell mucin depletion is also usually a reflection of severity of activity in ulcerative colitis. However, in fulminant colitis there may be considerable preservation of goblet-cell mucin between areas of ulceration. Long-standing disease may be associated with dysplasia of the epithelium, which is graded low or high. Dysplasia (a precancerous change) may be seen in flat mucosa or in raised, irregular lesions that are known as dysplasia-associated lesions or masses (**DALM**s). High-grade dysplasia is frequently associated with cancer elsewhere in the colon, and, when detected in rectal or colonoscopic biopsies, proctocolectomy must be considered. Multifocal, low-grade dysplasia or low-grade dysplasia within a DALM may also be an indication for considering proctocolectomy. The diagnosis of dysplasia is histological. Interobserver variation is reasonably low between pathologists in the diagnosis of high-grade dysplasia, but can be quite high for low-grade dysplasia. The reason for this discrepancy is that regenerative changes following inflammation and ulceration can closely simulate the appearances of dysplasia. The presence of dysplasia should ideally be confirmed by the opinions of two pathologists with experience of this condition.

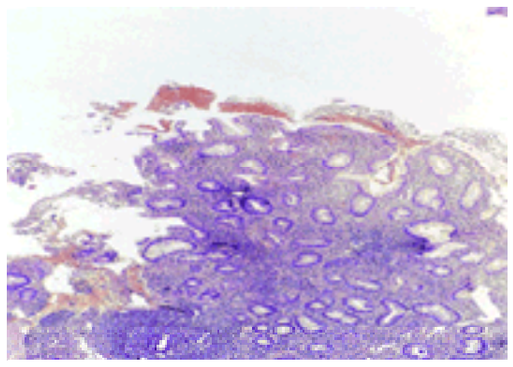


Fig. 1. Photomicrograph showing typical appearance of ulcerative colitis, with an irregular mucosal surface and ulceration.

Pouchitis

Pouchitis affects patients with ulcerative colitis who have had a panproctocolectomy and a pelvic ileal reservoir constructed. It does not usually affect patients with pouches created in the treatment of other conditions such as familial adenomatous polyposis. Pouchitis has three components, clinical, endoscopic, and histological. The clinical symptoms are of diarrhoea and systemic upset. Endoscopic inflammation is seen within the pouch mucosa and is accompanied histologically by severe acute inflammation and ulceration. The cause is unknown but is seems likely that pouchitis represents recurrence of ulcerative colitis within the pouch. Many of the minor histological changes within pouch biopsies are attributable to adaptive changes within the mucosa and care must be taken to score the amount of acute inflammation and ulceration in pouch biopsies to ensure it is of a severity known to be associated with the clinical symptoms and endoscopic appearances of pouchitis. Adaptive changes seen in the mucosa with time include varying degrees of villous atrophy, and chronic and acute inflammation. Small-bowel enzyme activity is preserved; this is therefore termed colonic phenotypic change rather than colonic metaplasia. The pouch functions as a neorectum, and consequently may develop several other inflammatory conditions familiar in the rectum, including mucosal prolapse. These have to be distinguished histologically from pouchitis.

Crohn's disease

Crohn's disease of the large intestine has been recognized as a separate entity since the definitive description by Lockhart-Mummery and Morson in the early 1960s. It is restricted to the colon in some 20 per cent of patients; another 60 per cent have ileocolonic involvement. It is usually a right-sided disease but limited left-sided disease can occur, especially associated with perianal disease. In the colon, Crohn's disease is characteristically a granulomatous condition with transmural aggregates of chronic inflammatory cells and deeply penetrating, fissuring ulceration (Fig. 2). The bowel wall is thickened and the serosa opaque and frequently covered by fat wrapping. The thickening seen in Crohn's disease results from connective tissue changes in all layers of the wall, including thickening and disruption of the muscularis mucosae, muscularization and fibrosis of the submucosa, and neuromuscular hyperplasia of the muscularis propria and Meissner's and Auerbach's plexuses, all accompanied by fat wrapping. Fat wrapping indicates diseased segments at laparotomy but may not reflect exactly the length of the diseased segment. Strictures, fistulas, and abscesses occur. Fistulas may be seen between bowel segments, or between bowel and other organs, or between bowel and skin. Large, fleshy lymph nodes are found in the mesentery or pericolic fat. The mucosa is oedematous and characteristically traversed by linear ulcers, giving a cobblestone appearance to the surface. Areas of uninvolved bowel separate the disease segments (skip lesions), although there may be cases with diffuse disease. Careful inspection of normal-looking mucosa may reveal tiny aphthoid ulcers, which start as ulcerated mucosa overlying lymphoid follicles; these represent the initial lesion in the disease. Rectal sparing and perianal disease are common features. The key distinguishing features on microscopy are patchy transmural inflammation in the form of lymphoid aggregates, focal ulceration, and focal acute and chronic inflammation. Granulomas, when present, are scattered throughout the bowel wall, together with fissuring and ulceration. The crypt pattern and goblet-cell population remain relatively well preserved despite considerable inflammation. Granulomas are only found in 60 per cent of patients with Crohn's disease and 50 per cent of those with isolated Crohn's colitis, however, and fissures are seen in only 30 per cent. Adjacent to sites f ulceration, ulcer-associated cell lineage (previously known as pyloric metaplasia or pseudopyloric metaplasia) may be seen. The function of ulcer-associated cell lineage is to produce epidermal growth factor to stimulate epithelial restitution at sites of ulceration. This cell lineage is not specific to Crohn's disease and may be seen adjacent to ulceration from any cause. Granulomatous vasculitis is a specific but rare finding in Crohn's disease, as are intralymphatic granulomas. Occasional, small,

poorly formed granulomas within lymph nodes are not specific for Crohn's disease. Inflammation associated with diverticular disease may provide good mimicry of inflammatory bowel disease, particularly Crohn's.

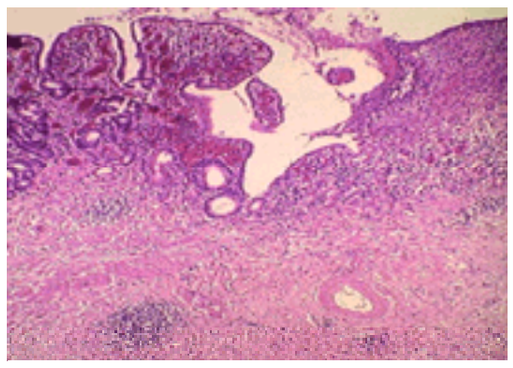


Fig. 2. Photomicrograph showing typical appearance of Crohn's colitis, with deep fissures and focal, transmural inflammation in the form of lymphoid aggregates.

Mucosal biopsy and non-specific inflammatory bowel disease

Rectal biopsy is often more helpful in establishing a diagnosis in ulcerative colitis than in Crohn's disease, because of the problems of sampling error. A normal rectal biopsy may be used as evidence against the diagnosis of active ulcerative colitis, but it must be stressed that normal appearances on sigmoidoscopy do not imply normal histology. It is also now recognized that in a number of patients the biopsy appearances in ulcerative colitis may return to normal after treatment. In Crohn's disease the chance of obtaining an abnormality on biopsy increases as the diseased segment becomes closer to the anal canal, but even when disease is restricted to the ileum, abnormal rectal biopsies can be found in up to 12 per cent of patients. These abnormalities include isolated, well-formed granulomas, distinctly focal acute and chronic inflammation with or without microgranulomas, or cryptolytic granulomas (focal granulomatous destruction of crypts). A microgranuloma has been defined as five epithelioid macrophages in aggregation. Multiple colonoscopic biopsies make it possible to document precisely the extent and distribution of colonic disease, particularly in those with mild ulcerative colitis and Crohn's disease.

Histopathology in severe colitis

Distinction between Crohn's disease and ulcerative colitis is most difficult when the disease is severe. Ulcerative colitis can become transmural, with deep ulceration. The irregularity in goblet-cell depletion is then less obvious and the histopathological appearances can be confusingly similar to those of Crohn's disease. Occasionally the distinction between Crohn's and ulcerative colitis cannot be made, even after study of the operative specimen. The histopathological diagnosis has to remain as ‘fulminant colitis, indeterminate'. The diagnosis may be resolved on subsequent proctectomy, since the usual effect of diversion of the faecal stream is to make the inflammation of ulcerative colitis more severe, whereas that of Crohn's disease will tend to resolve.

Further reading

Lockhart-Mummery HE, Morson BC. Crohn's disease (regional enteritis) of the large intestine and its distinction from ulcerative colitis. *Gut* 1960; **1**: 87: 105. [The classical description of colonic Crohn's disease.]

Morson BC, Dawson IMP. *Gastrointestinal pathology*, 2nd edn. Blackwell, Oxford, 1979. [The comprehensive textbook of gastrointestinal pathology.]

Price AB. Overlap in the spectrum of non-specific inflammatory bowel disease colitis indeterminate. *Journal of Clinical Pathology* 1978; **3**: 567–77. [One of the most important papers in inflammatory bowel disease. The initial description of ‘colitis indeterminate'.]

Warren BF, Shepherd NA. The role of pathology in pelvic ileal reservoir surgery. Intestinal pathology: pouchitis and diversion colitis. In: Underwood JCE, Lowe DG. *Recent advances in histopathology (18).* Churchill-Livingstone, Edinburgh, 1999. [A recent review chapter in which we try to offer some practical diagnostic advice.]

26.2.2 Colonic inflammation—diagnosis and management

D. P. Jewell

Ulcerative colitis
Diagnosis
Management
Crohn's colitis
Medical management
Nutritional and drug therapy
Social prognosis
Radiation colitis
Ischaemic colitis
Aetiology
Clinical features
Investigation
Differential diagnosis
Treatment
Microscopic colitis
Drug-induced colitis
Antibiotic colitis
NSAIDs
Miscellaneous inflammatory disorders
Solitary rectal ulcer syndrome
Colitis cystica profunda
Pneumatosis coli
Malakoplakia
Diversion colitis
Further reading

There are many causes of colonic inflammation, many of which are listed in Table 1. The diagnostic process begins when the patient presents with symptoms, but these can be highly variable. Change of bowel habit is usual and although increased frequency of loose or liquid stool is often the presenting symptom of colonic inflammation, it is not uncommon for patients to complain of constipation if only the rectum is inflamed. Bleeding usually triggers early investigation but many causes of colonic inflammation are frequently not associated with overt bleeding; for example, Crohn's disease of the colon. Rapid onset of symptoms is normally suggestive of an infective cause but ulcerative colitis can present for the first time with a very short history. Abdominal pain is very characteristic of *Campylobacter* infections and acute ischaemic colitis, whereas it is frequently not a prominent symptom of severe ulcerative colitis. Thus, timing of onset of symptoms, change of bowel habit, passage of blood or mucus, and abdominal pain may provide pointers to the diagnosis but are clearly insufficient by themselves. Anorexia, malaise, fever, and weight loss are additional vital pieces of clinical information. Urgency, tenesmus, and incontinence often indicate rectal inflammation but can also be associated with other pathologies. A history of previous radiotherapy to the abdomen or pelvis is not usually overlooked but a drug history (especially for non-steroidal anti-inflammatory drugs, **NSAIDs**) is often not recorded. Therefore, the diagnosis of colonic inflammation is dependent on investigation, in particular stool examination and culture, and visualization of the colon. This can be performed either by radiology or colonoscopy but, as many of the diseases listed in Table 1 are dependent upon histological appearances for accurate diagnosis, colonoscopy is preferable. Multiple biopsy specimens must be obtained even if the colon is macroscopically normal in order to diagnose microscopic colitis.

Infective
Bacteria
Viruses
Parasites
Ulcerative colitis
Crohn's disease
Radiation colitis
Ischaemic colitis
Microscopic colitis
Drug-induced colitis
Miscellaneous
Pneumatosis cystoides intestinalis
Colitis cystica profunda
Solitary ulcers
Cap polyposis
Malakoplakia
Diversion colitis
Neutropenic typhlitis

Table 1 Causes of colonic inflammation

Ulcerative colitis

Diagnosis

The majority of patients present with an insidious onset of symptoms; initially there is loosening of stool and increased frequency and then subsequently there is passage of mucus and blood mixed with the stool. Patients often complain of tenesmus and urgency and, when the inflammation is severe, there may be incontinence. Abdominal discomfort is fairly common but patients rarely complain of frank pain. Severe disease is associated with symptoms of systemic illness, such as weight loss, anorexia, vomiting, and fever. Patients with severe disease usually pass very loose stool containing pus and altered blood—there is no frank blood and thus patients can deny passing blood at all. Patients with a very limited proctitis usually present with constipation with the passage of fresh blood and mucus, often separate from or coating the stool, which is easily misdiagnosed as a problem with haemorrhoids unless sigmoidoscopy is performed.

On examination, there are usually no abnormal physical signs although the colon may be tender. For patients with severe disease, there may be tachycardia, fever, signs of anaemia and iron deficiency, and a hypoproteinaemic oedema.

The clinical diagnosis of ulcerative colitis is confirmed on sigmoidoscopy and rectal biopsy. The endoscopic appearances are those of diffuse inflammation, which vary in severity from loss of vascular pattern in a mild colitis to ulceration and spontaneous haemorrhage in severe disease (Table 2, Fig. 1). There is considerable observer variation in describing the changes of mild inflammation (hyperaemia, loss of vascular pattern, granularity), but much less for grades 2 to 4. Thus biopsy specimens are essential to determine disease severity as well as for sorting out the differential diagnosis. The extent of disease is determined by a full colonoscopy with multiple biopsy. Ulcerative colitis is usually a diffuse inflammation extending proximally from the rectum for a variable extent but never involving the small intestine. However, a few caveats need to be made. Partially treated colitis can have a rather patchy distribution and there may be relative rectal sparing in patients who have been given topical treatment with foams or suppositories. Patients with a severe, total colitis can have a ‘backwash' ileitis which is a very non-specific inflammation of the terminal ileum and is not indicative of Crohn's disease. Finally, a small proportion of patients with a left-sided or distal colitis can have localized areas of inflammation in the right colon (especially in the caecum—a ‘caecal patch'). The significance of this ‘skip lesion' is not known but it does not automatically signify Crohn's disease. If the diagnosis of ulcerative colitis seems unequivocal, examination of the small intestine may not be essential but it is certainly required to exclude small intestinal Crohn's disease if there is any doubt. About 30 per cent of patients with ileal Crohn's disease have a proctitis which may be the presenting problem and thus the ileal disease can be missed under these circumstances.

Grade	Appearance
0	Normal
1	Loss of vascular pattern, granularity
2	Granular mucosa with contact bleeding
3	Spontaneous bleeding
4	Ulceration

Table 2 Endoscopic grades of severity for ulcerative colitis

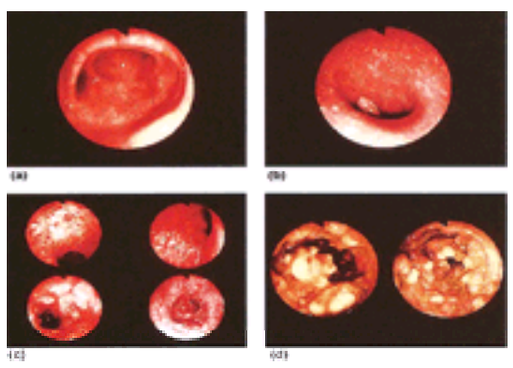


Fig. 1. Colonoscopic appearance in ulcerative colitis. (a) Mild colitis. (b) Moderate colitis. (c) Severe colitis. (d) Pseudopolyps in long-standing colitis.

Haematological, biochemical, or immunological investigations are not involved in the diagnostic process but are, of course, essential for the overall assessment of nutritional status and disease severity. Stool examination for amoebae, schistosomes, and other parasites is essential if there is any potential for exposure to these pathogens, and should be coupled with careful examination of biopsy material. Stool cultures for the common pathogens causing a bloody diarrhoea are essential (*Campylobacter, Salmonella, Shigella, Clostridium difficile*) as well as cell cultures for *C. difficile* toxins. For patients with an acute onset of bloody diarrhoea associated with severe abdominal pain, the possibility of infection with *Escherichia coli* 0157 must be considered even if there is no evidence of diffuse intravascular coagulation or renal impairment. Many laboratories do not routinely exclude all these pathogens unless specifically asked to do so.

Management

Medical therapy is directed towards controlling active disease (corticosteroids, 5-aminosalicylic acid) and maintaining remission by prevention of relapse (5-aminosalicylic acid), and is based on the results of many randomized, controlled, clinical trials which have provided clear guidelines for management. The details of therapy depend not only on severity but also extent of disease.

Proctitis

Disease limited to the rectum is ideally suited to topical therapy. Twice daily administration of corticosteroids or 5-aminosalicylic acid in the form of foams or suppositories can be highly effective. How long treatment should be continued is less clear but should be for at least 4 to 6 weeks or until there is histological reduction of the inflammation. An oral 5-aminosalicylic acid compound should also be given for long-term maintenance therapy. Many patients with a proctitis are frequently constipated above the inflamed rectum and many of their symptoms (bloating, distension, abdominal discomfort, flatus) are more related to the motility disorder than the inflamed rectum. Thus, clearing the colon is often of considerable benefit and can be done using a combination of a bulking agent and lactulose, or an osmotic purge (e.g. Picolax, Golytely, CleanPrep).

Mild disease

Patients who have up to four bloody stools daily, but who are systemically well, are said to have mild disease. In placebo-controlled trials, many of these patients make a satisfactory response to a 5-aminosalicylic acid drug, especially in high dose. The only comparative trial with oral corticosteroids, however, showed that prednisolone was significantly better than oral sulphasalazine for inducing a rapid response. The route of administration is largely determined by extent of disease. Patients with extensive or total disease require oral therapy (which is often combined with topical therapy), but distal disease can be treated with retention enemas or foam preparations. If oral prednisolone is used, 20 mg daily is effective for the majority of patients. For topical therapy, meta-analyses of trials comparing steroid with 5-aminosalicylic acid preparations have suggested that 5-aminosalicylic acid is more effective than steroids. However, in practice, steroid enemas and foams are often better tolerated than mesalazine (5-aminosalicylic acid) preparations and are considerably cheaper. Steroid enemas containing prednisolone metasulphobenzoate or budesonide are associated with minimal blood levels of steroids and should not give rise to steroid-induced complications. The optimal duration of therapy is not defined but it should continue for 6 to 8 weeks, and should certainly be continued well beyond the point at which the patient becomes asymptomatic. This occurs long before healing is visible by endoscopy and treatment should probably continue until there is histological resolution although this is not a commonly used end-point. It should be remembered that persistent histological inflammation predicts early relapse if treatment is withdrawn.

Moderate disease

Moderate disease activity is denoted by patients having more than four bloody stools daily, frequently with nocturnal diarrhoea, but they are systemically well. If the disease is very distal (proctitis or proctosigmoiditis), topical therapy can be used. However, most patients will require oral therapy. High-dose mesalazine may be tried (e.g. 3 to 4 g daily) but prednisolone is usually given to obtain a rapid response. A regimen of prednisolone which is effective in the majority of cases is 40 mg daily for 1 week, 30 mg daily for 1 week, 20 mg daily for 1 month, and then tapering to zero over 3 to 4 weeks.

Severe disease

Severe attacks are defined as more than six bloody stools daily with evidence of systemic illness as witnessed by tachycardia (more than 100 beats/min), fever (more than 37.5°C), anaemia (haemoglobin less than 10.0 g/dl), or hypoalbuminaemia. If the diarrhoea is associated with at least two of these features, the patient should be admitted to hospital for intravenous therapy, even though they may feel well. Frequently, nausea, vomiting, anorexia, and a tender colon on abdominal examination are additional features.

These patients need hospital admission. A blood count, biochemical screen (creatinine, electrolytes, liver enzymes, albumin), erythrocyte sedimentation rate, C-reactive protein, a plain abdominal radiograph, and a stool sample for microscopy and culture are the minimum of tests that are urgently required. Intravenous replacement of fluids, electrolytes, and blood is instituted and intravenous corticosteroids are begun (e.g. hydrocortisone 100 mg four times daily). Rectal hydrocortisone (100 mg in 100 ml) is often given twice daily although there is no trial to support this practice. Assessment by both physician and surgeon is essential. If patients are clinically malnourished, parenteral nutrition is indicated, but this is usually unnecessary. Most patients, however, prefer to be on clear fluids for the first few days although the outcome of the attack is probably not influenced by complete ‘bowel rest'. Pulse, temperature, bowel frequency, and daily inflammatory markers (especially C-reactive protein) are monitored. Plain abdominal radiographs should be taken every 2 days if the patient starts to improve, but should be done more frequently if there is a failure to improve or the patient deteriorates.

Seventy to eighty per cent of patients will respond to this regimen after 5 to 7 days and can be converted to oral therapy (prednisolone 40 mg daily together with maintenance 5-aminosalicylic acid therapy) prior to discharge. Patients often ask for dietary advice but there are no specific diets known to alter the course of disease. However, it is worth remembering that lactose-free or milk-free diets can occasionally help—approximately 11 per cent of patients with severe attacks of colitis are

hypolactasic. However, the general advice is for a well-balanced diet.

For patients who deteriorate during the first few days of steroid therapy, colectomy should be performed without delay. Deterioration may be shown by increasing diarrhoea, malaise or abdominal pain, by a rise in pulse or temperature, by increasing colonic tenderness or loss of bowel sounds, or by dilatation of the colon on a plain radiograph.

However, some patients will improve but clinically are not going into remission. Thus, mild anorexia, a tachycardia, mild intermittent pyrexia, persistent abdominal tenderness, continuing diarrhoea (more than four stools daily), or persistently raised inflammatory markers all indicate failure to make a satisfactory response. Clinical trials of higher-dose steroids (e.g. 1 g methyl prednisolone intravenously daily), or the addition of antibiotics (metronidazole, gentamicin, vancomycin, tobramycin) have not proved effective. However, continuous intravenous infusion of cyclosporin (4 mg/kg) has shown benefit in patients failing to respond to 10 days of intravenous corticosteroids. In a small double-blind trial of such patients, 9 of 11 receiving cyclosporin went into remission compared with 0 of 9 receiving a placebo. Subsequently, clinical experience suggests that the response rate is about 60 per cent in the short term. Nevertheless, especially for patients in their first attack, this is a useful form of treatment as it allows the patient to come to terms with the disease and to learn about its nature and the consequences of surgery even though they may relapse and come to surgery within a year of their initial attack.

Side-effects of cyclosporin during intravenous administration are few although it should be avoided in patients with serum cholesterol concentrations of less than 3.0 mmol/l—the cremaphor in which the cyclosporin is suspended can cross the blood–brain barrier with these low cholesterol levels and can cause fits. The major threat is that of opportunistic infection. This is rarely seen f cyclosporin is introduced early (i.e. after 5 to 7 days of steroid therapy) and is used for short periods—if the patient is going to respond this is usually clear after 2 to 3 days of treatment. However, prolonged usage in patients who have had 10 days or more of intravenous steroid may be associated with pneumocystis or cytomegalovirus infection. Deaths have been reported. A recent audit in Oxford (John Radcliffe Hospital) has shown that perioperative morbidity and mortality for urgent colectomy is not increased by the use of cyclosporin. This was based on a protocol whereby the drug was introduced 3 to 7 days following intravenous steroids and only allowing 3 to 4 days of treatment before deciding on colectomy.

Many questions remain unanswered concerning the use of intravenous cyclosporin. Anecdotal data suggest that 2 mg/kg may be as effective as 4 mg/kg but this needs formal evaluation. Intravenous cyclosporin as monotherapy, without corticosteroids, may be effective. Finally, it is not at all clear how to proceed once a patient has gone into remission on cyclosporin. Many clinicians will continue oral cyclosporin (5 mg/kg) until steroids have been tapered off. Others will use azathioprine during the steroid taper, while others will use oral cyclosporin in combination with azathioprine. If oral cyclosporin is used, side-effects are common but mostly reversible (gum hypertrophy, paraesthesia, hirsutism, increase in serum aminotransferases); however, the fall in glomerular filtration rate may not always go back to pretreatment values.

How to predict the outcome of a severe attack

Retrospective studies have shown that there is a high chance of colectomy (greater than 55 per cent) if in the first 24 h, the pulse rate is more than 100 beats/min, there is a fever of greater than 38°C, or a stool frequency of more than 12 stools/24 h. The presence of deep ulceration on a plain radiograph (the ‘mucosal island' sign) or on flexible sigmoidoscopy also predicts the need for colectomy in 75 per cent of cases. A recent prospective study involving 37 clinical, laboratory, and radiological variables in 50 severe attacks has shown that more than 80 per cent of patients will require colectomy if, after 3 days of intravenous steroids, the stool frequency is still more than 12/day or 5 to 8/day with a C-reactive protein greater than 45 mg/l (normal: less than 8.0 mg/l).

Timing of diagnostic procedures

All patients with a severe attack require stool examination and culture. A plain abdominal radiograph accurately predicts the extent of disease to within one colonic segment but it is advisable to examine the rectum with a rigid or flexible sigmoidoscope and take biopsy specimens. Amoebiasis can very rarely occur in patients who have not travelled to endemic areas. The role of colonoscopy is controversial but adds very little more if all the above information is available and can be frankly dangerous.

Management of complications

Patients with a dilated colon on presentation (transverse colon more than 5.5 cm) should be started on medical therapy. About 50 per cent will respond but failure to do so within 2 to 3 days, or the development of a dilatation during intravenous steroid therapy, are absolute indications for colectomy. If there is free gas on a plain film, absence of bowel sounds, or other reasons to suspect perforation, emergency colectomy is also indicated after 2 to 3 h of resuscitation with intravenous fluids, electrolytes, and antibiotics. Especially if the patient is on steroids, the classic signs of perforation may not be present. Massive bleeding from the inflamed colon is rare. It usually responds to transfusion and intravenous steroids but may occasionally require colectomy.

Maintenance of remission

This is achieved using 5-aminosalicylic acid drugs. The original drug, sulphasalazine, contained 5-aminosalicylic acid linked to a sulphonamide (sulphapyridine) by an azo bond. Only about 10 per cent of the whole compound is absorbed but bacterial action in the colon splits the azo bond to release the two moieties. Sulphapyridine is rapidly absorbed, acetylated in the liver, and excreted in the urine. The majority of the 5-aminosalicylic acid (60 to 70 per cent) is excreted in the stool. The remainder is absorbed, metabolized (mostly by acetylation), and about 25 per cent of the oral dose is excreted in the urine. Studies in Oxford showed that 5-aminosalicylic acid is the active ingredient of sulphasalazine and it was already known that most of its side-effects were related to sulphapyridine (Table 3). Thus, many drugs are now available which deliver 5-amino-salicylic acid into the colon. The 5-aminosalicylic acid cannot be given alone as it is rapidly absorbed and excreted and, although its exact mode of action is unknown, it appears to act topically—therefore, a high luminal concentration within the colon is required. Table 4 lists the nature of the 5-aminosalicylic acid drugs available. Mesalazine is now the generic name for 5-aminosalicylic acid and can be formulated either as delayed-release preparations or as a prodrug which depends on bacterial action to split the azo bond and release the active compound.

Dose-related	*Idiosyncratic*
Headaches	Hypersensitivity rashes
Nausea	Stevens–Johnson syndrome
Diarrhoea	Male infertility
Heinz body haemolytic anaemia	Agranulocytosis
	Hepatitis
(b) Side-effects of 5-aminosalicylic acid	
Nausea	
Diarrhoea	
Headaches	
Urticaria	
Interstitial nephritis	
Colitis	

Table 3 (a) Side-effects of sulphasalazine

	Recommended maintenance dose
Mesalazine	
Asacol	800mg twice daily
Salofalk	500mg three times daily
Pentasa	500mg three times daily
Prodrugs	
Sulphasalazine	1g twice daily
Olsalazine	500mg twice daily
Balsalazide	1.5g twice daily

Table 4 The new 5-aminosalicylic acid drugs

Sulphasalazine has been shown to reduce the relapse rate of ulcerative colitis fourfold—from approximately two attacks per year to one every 2 years. This effect appears to exist over many years and, therefore, treatment should continue. All the new 5-aminosalicylic acid compounds are as effective as sulphasalazine, with far fewer side-effects, but are considerably more expensive. Few comparative trials have been performed but the prodrugs may be slightly more effective than the mesalazine preparations. Table 4 also gives recommended dosages for maintenance therapy or for treating active disease.

Chronic active disease

Some patients relapse soon after coming off oral steroids or topical therapy or develop symptoms once the dose of oral prednisolone falls below 15 to 10 mg daily. Many of these patients will respond well to azathioprine (2.0 to 2.5 mg/kg) or 6-mercaptopurine and can be weaned off steroids without relapsing. These immunosuppressives seem to act slowly and may take 6 to 12 weeks before maximum benefit is achieved. Weekly blood counts should be obtained during the first 2 to 3 months of therapy to detect bone-marrow suppression. However, this is rare and severe marrow suppression probably only occurs in patients who have low activity of the thiopurine methyl transferase enzyme (approximately 1 in 300). Patients who are intolerant of azathioprine (nausea, diarrhoea, headache, myalgia) may tolerate lower doses or a change to 6-mercaptopurine. For patients who are intolerant of these drugs, or who fail to respond, an alternative is methotrexate at 25 mg weekly. Although, there is some anecdotal experience to support this, there is no randomized controlled trial of this dosage in ulcerative colitis.

Colectomy is often a better option for patients with chronic active disease, especially if they have had the disease for more than 10 years and have extensive disease. Nevertheless, resistant proctitis can be very disabling and may impair quality of life considerably. Provided the possibilities of an irritable bowel, dietary intolerance, or drug-induced colitis are excluded as contributing factors to the symptoms, colectomy should not be denied to these patients with very limited disease.

Crohn's colitis

Inflammation confined to the colon occurs in about 20 per cent of patients with Crohn's disease but 35 to 45 per cent will have an ileocolitis. Presenting symptoms are usually diarrhoea, malaise, anorexia, weight loss, and abdominal pain. Frank bleeding is often absent despite extensive and deep ulceration in the colon. Perianal disease is frequently associated with colonic disease and may vary from fleshy skin tags, fissures, or perianal abscesses to extensive fistulization and abscess formation. Diagnosis is best made by colonoscopy with multiple biopsy (see Chapter 26.2.1) and this should be performed even if a barium enema has already suggested the diagnosis. At diagnosis, all patients should have both small bowel and colon visualized to assess the full extent of disease. Labelled white-cell scans will document inflammatory activity throughout the entire length of the intestine but are clearly insensitive in detecting strictures or fistulas. The use of CT scans or MRI is being explored for assessing the presence of affected gut but is not yet established for clinical use. However, CT scans as well as labelled white-cell scans are useful for detecting intra-abdominal abscesses and MRI provides the best modality for assessing per-rectal fistulas and abscesses.

Medical management

Medical treatment is aimed at relieving symptoms rather than treating endoscopic or radiological features, or laboratory tests. Thus, the cause of the symptoms has to be identified before a treatment plan can be formulated, which requires assessment of the following.

Nutrition

Symptoms of malaise, tiredness, and weight loss may be due to nutritional deficiencies. Deficiencies of iron, folic acid, vitamin B_{12}, calcium, magnesium, or zinc need to be excluded, even if there is no obvious small intestinal disease. Hypoalbuminaemic oedema is common. The low albumin is a result of reduced dietary protein, protein loss through the inflamed colon, and the switching off of albumin synthesis in the liver by proinflammatory cytokines. Patients with long-standing colonic Crohn's disease, particularly those with additional small bowel disease, frequently have low levels of selenium although this is rarely obvious clinically.

Strictures

Symptoms of diarrhoea, pain, and weight loss can be due to partial obstruction secondary to a stricture; nausea is common but vomiting may not necessarily occur until obstruction is virtually complete. Strictures in the right colon or at the ileocaecal valve often lead to small bowel stasis and hence bacterial overgrowth. This, together with symptoms of obstruction, may occur secondary to fibrotic strictures with very little active inflammation—treatment is directed towards reducing the bacterial population with antibiotics (e.g. metronidazole, ciprofloxacin) and relieving the obstruction (low-residue diet, surgery).

Previous surgery

Patients who have previously had an ileal resection are at risk from a bile-salt induced diarrhoea, the effects of which are exacerbated if there has also been a colonic resection or there is active disease in the remaining colon. If it can be tolerated, cholestyramine is the treatment of choice. Previous bypass operations (e.g. ileotransverse colostomy) can lead to bacterial overgrowth.

Active Crohn's disease

Bleeding usually indicates active inflammation, but it is a relatively infrequent symptom even if the colon is extensively ulcerated. Fever, tachycardia, and abdominal tenderness or a palpable mass are clear indicators of active inflammation. Active perianal disease is usually obvious on clinical examination and consists of reddening and swelling of the perianal skin and/or skin tags, discharging fistulas, and anal canal ulceration.

Complications

Recurrent urinary tract infections, cloudy urine, or pneumaturia indicate a fistula to the bladder. Fistulas to the vagina usually cause a faecal vaginal discharge. Fistulas to the skin are often related to previous surgical incisions. Carcinoma of the colon occurs in 3 to 5 per cent of patients with Crohn's colitis and, as in ulcerative colitis, there may be delay in diagnosis because the symptoms are compatible with active disease. Rarely, amyloid may complicate Crohn's disease, especially if chronic fistulating disease is present. It is often diffusely present throughout the intestinal mucosa and can then cause diarrhoea even though there is no obviously active Crohn's disease. Nephrotic syndrome can occur if renal amyloid is sufficiently severe.

Nutritional and drug therapy

This is similar to that already described for ulcerative colitis (see above) and for Crohn's disease of the small intestine (see Chapter 24.2.1). Active disease is usually treated with corticosteroids or by elemental or polymeric diets. The combined evidence that 5-aminosalicylic acid drugs are effective is weak, even when high doses (4 g daily) are used. Steroid-dependent or resistant disease often responds to the addition of immunosuppressive drugs (azathio-prine, 6-mercaptopurine, methotrexate). Infusions of monoclonal antibodies to tumour necrosis factor may offer an effective alternative therapy.

In randomized controlled trials, about two-thirds of patients with chronic active disease resistant to steroids and immunosuppressives will respond well to a single infusion (5 mg/kg) infliximab (Romicade). A similar proportion of patients with fistulas will also respond; usually these infusions are given at 0, 2, and 6 weeks. However, patients will subsequently relapse. Side-effects are few but long-term complications (e.g. lymphoma) are a potential problem.

Maintenance of remission is also unsatisfactory. 5-Aminosalicylic acid drugs are again not very effective although in high dose may prolong the time to relapse following resectional surgery.

Antibiotics are clearly indicated when there is frank infection—abscesses, fistulas, or bad perianal disease. However, there is considerable controversy concerning the effectiveness of antibiotics in controlling active disease. Ciprofloxacin, metronidazole, clarithromycin, or combinations of antibiotics (usually designed to eradicate atypical mycobacteria) have all been proposed but, where randomized controlled trials have been reported, the benefits have been marginal and the number of patients studied very few.

As with small intestinal disease, patients should be strongly advised to give up smoking as this habit predicts a worse natural history than is seen in non-smokers. For the majority of patients, a healthy, well-balanced diet is recommended. Patients with strictures will need a low-residue diet and some patients will benefit from exclusion

diets.

Social prognosis

Crohn's disease has a high morbidity but low mortality. Seventy to eighty per cent of patients will have at least one operation during their lifetime and for patients with colonic or perianal disease, there is a considerable chance of an ileostomy. Nevertheless, the majority of patients maintain full employment and, apart from the 10 to 20 per cent with chronic active disease, have good quality of life when in remission. There are still major problems with life insurance but the situation is improving. Infertility is rarely a problem unless there has been extensive pelvic surgery or infection. There is an increased risk of early abortion when the disease is active but otherwise pregnancies should be successful. Apart from methotrexate, all the drugs used for controlling disease activity are safe during pregnancy.

Radiation colitis

This may occur early after radiotherapy or as a late phenomenon, often occurring years later. Nowadays, radiation colitis occurs mainly in the rectum and sigmoid and it is rare to see extensive colonic and ileal damage because the irradiation is much better focused. However, pelvic irradiation for gynaecological cancers and irradiation of the prostate still frequently affects the sigmoid and rectum. Colitis may occur if the total radiation dose exceeds 4500 rad and proctitis if the dose exceeds 5500 rad.

Early symptoms, which usually develop within 1 to 4 weeks of radiotherapy, include nausea, diarrhoea, tenesmus, mucus, and bleeding. Sigmoidoscopy shows an oedematous, inflamed mucosa which only rarely ulcerates. The mucosa may appear dusky red in colour. Treatment is needed if the symptoms are sufficiently severe and it is usual to use topical steroids or 5-aminosalicylic acid. There is no evidence from randomized clinical trials to support their use but they seem effective in clinical practice.

Late radiation damage includes telangiectasia, chronic inflammation, stricture formation, and fistulas. Rectal bleeding is the usual symptom in patients with telangiectasia, which is well seen at flexible sigmoidoscopy. Treatment can be very difficult, although for minor bleeding, reassurance is all that is necessary. For more marked bleeding leading to anaemia, electrocoagulation or endoscopic laser coagulation has been used with occasional success. Oestrogen therapy can also be used, as for angiodysplasia, but there is no hard evidence demonstrating efficacy for any of these approaches. Frank inflammation causing diarrhoea, mucus, bleeding, and tenesmus is treated with steroids and/or 5-aminosalicylic acid. Topical therapy is usually sufficient. Rectal strictures can be dilated but sigmoid strictures or fistulas usually require surgical management. Unfortunately, operating on ischaemic colon from radiation damage is often associated with major complications and should therefore be avoided if at all possible.

Ischaemic colitis

Ischaemic colitis is an inflammatory condition produced by interruption of the blood supply to the colon insufficient to cause full thickness tissue death. It most commonly affects those in the sixth to the eighth decades of life and is thus being seen with increasing frequency in our progressively elderly population.

Aetiology

Ischaemic colitis may be caused by occlusion of a major artery, small vessel disease, venous obstruction, ‘low flow' states, or intestinal obstruction (Fig. 2). In each case, the mucosa and sub-mucosa are predominantly affected, the extent of injury being determined by the severity and longevity of the insult. Ischaemia reduces the integrity of the mucosa and allows invasion by pathogenic organisms such as clostridia, which are normal constituents of colonic flora. These processes produce inflammation and mucosal ulceration which may resolve completely. Alternatively the insult can result in permanent injury with healing by fibrosis and subsequent stricture formation. Rarely, necrotizing colitis develops, which can spread to affect areas of the colon which are not ischaemic.

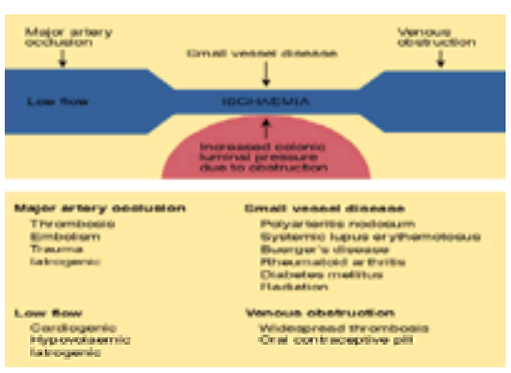


Fig. 2. Aetiological factors responsible for ischaemic colitis.

Although any part of the colon can be affected, the splenic flexure is particularly susceptible to ischaemic injury because it is the site of the watershed between the superior mesenteric artery, supplying the transverse colon, and the inferior mesenteric artery which supplies the descending colon. These vessels are linked by a marginal artery, but this is frequently absent or poorly developed at the splenic flexure. Occlusion of either major artery or their feeding branches (middle colic artery from the superior mesenteric artery and left colic artery from the inferior mesenteric artery) can therefore result in ischaemia. This point is of particular relevance during aortic and colorectal surgery if the inferior mesenteric artery is ligated. During aortic surgery it is important to confirm pulsatile flow in the superior mesenteric and marginal arteries prior to ligating a patent inferior mesenteric artery. If this is absent or if doubt exists then the in-ferior mesenteric artery should be reimplanted into the graft.

Clinical features

History

A typical patient is 50 years of age or more and complains of left-sided abdominal pain which is acute in onset and started in the left iliac fossa. Loose stools, which characteristically contain dark blood as well as clots, may be passed. There may be a history of previous similar episodes, or of peripheral or cardiovascular disease, or collagen vascular disease, especially if the symptoms are atypical.

Examination

As ischaemic colitis is predominantly a disorder of colonic mucosa and submucosa it is not usually associated with a major systemic upset, but a low grade pyrexia and tachycardia should be expected. On abdominal examination, the affected colon is tender and may be palpable. Dark blood will be present *per rectum*. Signs of peripheral vascular disease or other associated conditions should be sought.

Investigation

It is important to first establish the diagnosis and then determine the presence of any treatable aetiological factors.

Radiological investigations

A plain abdominal radiograph and a contrast enema are the most useful investigations in the initial stages of this disorder. ‘Thumb-printing' is diagnostic and is most often seen at the splenic flexure (Fig. 3). It is present at an early stage (from 3 days) and is the result of submucosal oedema and haemorrhage which produce swellings that project into the bowel lumen. These are clearly seen in contrast studies.



Fig. 3. A barium enema showing the typical ‘thumb printing' sign of ischaemic colitis. Note that this is most marked at the splenic flexure. The appearances are the result of oedematous mucosal folds caused by ischaemia.

Later, mucosal ulceration and irregularity may develop and these can resemble the appearances of ulcerative colitis or Crohn's disease. However, ulcerative colitis invariably affects the rectum and there is loss of the normal colonic haustral pattern while in Crohn's disease deep ulcers resemble ‘rose thorns' and areas of affected colon are separated by normal bowel. These features are not seen in ischaemic colitis.

Although many of the features of ischaemic colitis are reversible, stricture formation, if it occurs, is not and causes further diagnostic problems. Ischaemic strictures are often long, uniform and have smooth, gradual beginnings and ends, an appearance called ‘funnelling'. However, these findings do not exclude carcinoma; this diagnosis should be considered, particularly if only a short segment of colon is affected. The role of angiography is not established. Although it can be valuable in isolated cases where significant, symptomatic occlusive lesions are revealed, there is generally no correlation between the appearance of vessels at angiography and the integrity of the colonic blood supply.

Endoscopy

Ischaemic lesions are usually beyond the reach of the rigid sigmoidoscope, but colonoscopy can be used to visualize and biopsy affected colon. In the early stages of ischaemia, the mucosa will be heaped up, oedematous, and bluish purple (the ‘thumb-prints' seen radiologically). It will bleed on contact with the endoscope or other instruments. Later, ulceration as well as strictures may be seen.

Differential diagnosis

It should be noted that some of these diagnoses, for example carcinoma, are also possible aetiological factors for ischaemic colitis.

Treatment

This will be determined by the mode of presentation, which in turn reflects the underlying stage of the ischaemic process. Conservative management is the mainstay of treatment for those seen with acute symptoms. The patient is rested in bed and given intravenous fluids. Broad-spectrum antibiotics are often administered, although there is no conclusive evidence to suggest that they influence outcome. There is no place for anticoagulation or steroid administration unless this is indicated by an underlying disorder such as vasculitis.

It is very rare for ischaemic colitis to progress to frank colonic gangrene, but all patients should be monitored frequently to assess progress. If the injury is transient then resolution occurs after a few days to a week. More severe insults lead to stricture formation. These require investigation and treatment if they produce symptoms or if there is diagnostic doubt. Excision followed by end-to-end anastomosis is safe, although it is essential to ensure the viability and vascularity of the resection margins. If malignancy is excluded, then the resection can be limited to the affected segment but a more radical excision should be performed if there is continuing diagnostic uncertainty.

Microscopic colitis

Microscopic colitis is most commonly seen in women over the age of 50 and presents as a painless, watery diarrhoea. Some mucus may be passed but bleeding is very rare. Weight loss does not occur and there are usually no abdominal physical signs. Furthermore, sigmoidoscopy and/or colonoscopy reveals an apparently healthy colonic mucosa with a normal vascular pattern. Thus, the diagnosis can only be made if biopsy specimens are taken (see Chapter 26.2.1).

Most pathologists now subdivide microscopic colitis into three categories: lymphocytic, collagenous, and ‘non-lymphocytic, non-collagenous'. In lymphocytic colitis, there is a pronounced increase in the number of intraepithelial cells and the lamina propria shows a mild increase in lymphocytes, plasma cells, and frequently eosinophils. Collagenous colitis is characterized by a subepithelial collagen band greater than 10 µm thick. It is usually maximal in the right colon and its thickness diminishes distally—there may be rectal sparing which can provide a diagnostic pitfall if only rectal biopsies have been taken. Collagenous colitis is also associated with a mild increase in chronic inflammatory cells within the lamina propria.

Treatment of either lymphocytic or collagenous colitis can be difficult. 5-Aminosalicylic acid drugs can be helpful but many patients require oral steroids. However, 10 to 20 per cent of patients are still symptomatic. Other causes of the colitis must be excluded—coeliac disease can be associated with a microscopic colitis and ingestion of NSAIDs can also be associated with the condition. A wide variety of other treatments have apparently been useful in anecdotal cases and these include dietary exclusion, cholestyramine, and mepacrine.

One interesting report from Sweden showed that a split ileostomy to defunction the colon can be a very effective way of treating refractory collagenous colitis. Furthermore, the collagen band tended to disappear, but most patients relapsed when colonic continuity was restored.

Drug-induced colitis

Antibiotic colitis

This is caused by toxins produced by *C. difficile*. The organism is a Gram-positive, anaerobic bacillus that can be found in otherwise healthy individuals but under certain circumstances is able to multiply, produce toxin (toxins A and B), and cause a pseudomembranous colitis. This form of colitis may occur spontaneously without antibiotic exposure, in the elderly, associated with colonic obstruction, and in immunosuppressed patients. However, the majority of cases occur 1 to 3 weeks following antibiotic usage. Virtually all antibiotics have been implicated but the commonest have been clindomycin, lincomycin, ampicillin, amoxycillin, and the cephalosporins.

Severe watery diarrhoea, abdominal pain, and fever are common symptoms. Extraintestinal manifestations can occur (e.g. an acute arthropathy, erythema nodosum) and an acute dilatation of the colon can also occur. There is usually a marked leucocytosis, grossly elevated C-reactive protein, and hypoalbuminaemia. The diagnosis is made by demonstrating the presence of toxin in stool and growing the organism from stool cultures. Sigmoidoscopy shows an acutely inflamed rectum, often with whitish plaques adherent to the epithelium. These are the pseudomembranes, but they are not invariably present and may actually spare the rectum.

Treatment includes correction of fluid and electrolyte losses and antibiotics directed towards *C. difficile*. Metronidazole is usually used as first-line treatment (400 mg three times daily) as it is effective and cheap and is equally as effective as vancomycin. Treatment should continue for at least 7 days. However, the relapse rate is high (20 to 25 per cent). Should this occur, vancomycin can then be used (125 mg four times daily). For resistant cases, combination therapy with both metronidazole and vancomycin can sometimes be effective. Anion exchange resins such as cholestyramine or probiotic therapy have also been used with apparent benefit but are not used widely.

Occasionally, colectomy may be required if a toxic dilatation occurs, although some surgeons will merely defunction the colon with an ileostomy.

NSAIDs

The non-steroidal anti-inflammatory drugs are well known for their ability to cause gastric erosions, ulcers in the stomach and duodenum, and upper gastrointestinal bleeding. Their effects on the intestine are well described but often poorly recognized. In the colon, they may cause thin fibrous strictures or ‘diaphragms', and a colitis which is very similar to ulcerative colitis. The diaphragms are very rare and are often associated with similar lesions in the ileum. Patients present with post-prandial colic, nausea, anorexia, and weight loss—that is, symptoms of obstruction. However, the diaphragms are often difficult to diagnose, especially on barium radiology; they are more readily visible at colonoscopy. Surgical resection is usually required as they do not resolve even if the NSAIDs are withdrawn.

NSAID-induced colitis causes diarrhoea, usually with blood and mucus. Sigmoidoscopic appearances are identical to a mild ulcerative colitis and, histologically, the two forms of colitis can be difficult to distinguish. A marked increase in apoptosis of epithelial cells is said to be a distinguishing feature of NSAID-induced colitis. 5-Aminosalicylic acid drugs can cause a similar lesion, albeit rarely, which can be particularly confusing as ulcerative colitis is usually the reason for their usage. Intramuscular gold injections for the treatment of rheumatoid arthritis can also induce a non-specific colitis. Treatment requires stopping the NSAID.

Miscellaneous inflammatory disorders

Solitary rectal ulcer syndrome

This occurs predominantly in young patients (20 to 40 years) who present with rectal bleeding, mucus, tenesmus, and difficulty in evacuation. On questioning, many patients will admit to straining and, in severe cases, there may be rectal prolapse. Some patients resort to digital evacuation but the bowel function is frequently normal. Solitary rectal ulcers usually occur on the anterior wall and may often be felt as an irregular area on rectal examination. On sigmoidoscopy, there may be a range of appearances. Early lesions often appear as localized inflammation usually 5 to 10 cm from the anal canal, but frank ulceration is often present. This can consist of single or multiple ulcers which are irregular and may be surrounded by polypoid mucosa. The lesion can sometimes be extensive and virtually circumferential. Thus the initial impression may be one of carcinoma. Biopsy specimens will often reveal the characteristic muscle fibres that extend upwards from the muscularis mucosae between the glands.

Solitary ulcers of the rectum are often refractory to treatment. Anorectal physiological measurements and a defaecating proctogram are useful investigations to determine exactly what is going on. Medical therapy is aimed at keeping the stool consistency soft (e.g. lactulose, bulking agents) and using topical steroids or 5-aminosalicylic acid preparations to minimize the inflammatory response. Spontaneous remissions can occur. Surgical management is usually only required if there is a troublesome prolapse.

A variant of the solitary rectal ulcer syndrome is cap polyposis. Symptoms are similar but on sigmoidoscopy small ‘volcanic' lesions are seen on the mucosal folds of the rectum. They appear as very small polyps often with an ulcer at the tip. Most patients with cap polyposis have a history of straining or prolapse. Very few cases have been described but the condition is probably more frequent than the literature suggests.

Colitis cystica profunda

This disorder consists of mucus-filled cysts which are deep to the muscularis mucosae, although they often communicate into the lumen of a gland via a defect in the muscularis. The condition often causes marked irregularity of the mucosa which can become polypoid. Localized areas of colitis cystica can occur in association with solitary rectal ulcers and adjacent to adenocarcinomas. In this situation, the histological changes are thought to result from aberrant mucosal repair secondary to ischaemic damage induced by straining, prolapse, or obstruction. A more diffuse form has been described in association with ulcerative colitis.

Symptoms are usually rectal bleeding, mucus, and tenesmus. Diarrhoea, abdominal colic, and rectal discomfort may also occur. Endoscopically, the lumen can be narrowed by an abnormal polypoid mucosa or there may be plaque-like lesions. Definitive diagnosis can only be made histologically.

Treatment may not be necessary if symptoms are minimal. Nevertheless, a careful search through serial sections of multiple biopsies is required to exclude a carcinoma. If the condition occurs in association with a solitary rectal ulcer or a prolapse, management is directed appropriately. Occasionally, the submucosal cysts are sufficiently large to cause bulky polypoid lesions. Segmental resection for colonic lesions or transanal excision for rectal involvement may be required.

Pneumatosis coli

This is another rare cause of colonic inflammation and consists of gas-filled cysts in the submucosa and subserosa. They can occur anywhere in the gastrointestinal tract but are most frequently seen in the colon where they may be segmental (usually rectum and sigmoid) or diffuse. Symptoms are very non-specific and consist of watery diarrhoea, mucus, abdominal cramps, and distension. Bleeding can occur. Very rarely, pneumatosis may cause a volvulus of intussusception. The diagnosis can be made endoscopically—the mucosa shows a polypoid appearance due to the smooth, rounded cysts protruding into the lumen.

A plain abdominal radiograph may be diagnostic. A barium enema or CT scan may also be diagnostic. Histologically, the walls of the cysts are lined with cuboidal epithelium and are often surrounded by chronic inflammatory cells and granulomas.

Asymptomatic pneumatosis needs no treatment. However, it can be very difficult to treat once symptoms develop. Oxygen therapy is usually the first-line therapy. Patients are usually admitted to hospital and given 60 to 70 per cent oxygen by nasal catheter. Symptoms often settle but many patients relapse soon after stopping oxygen therapy. Hyperbaric oxygen may then be helpful, if available. Elemental diets and antibiotics (e.g. metronidazole) are also used with variable success. Surgical resection is required for severe or relapsing symptoms as well as for the complications of intussusception or volvulus.

Malakoplakia

This is rare and is a chronic inflammatory process that most often affects the bladder but can also involve the gastrointestinal tract. In the large bowel, it usually affects the rectosigmoid and presents with rectal bleeding, diarrhoea, and occasionally fever. It is diagnosed endoscopically, often as an incidental finding, and may appear as a mucosal plaque, a polypoid lesion, or even a stricture. Thus, it may be mistaken for a carcinoma. The diagnosis is essentially a histological one. The characteristic lesion consists of collections of tissue macrophages which may contain inclusion bodies—Michaelis–Guttman bodies. These inclusions are usually Gram positive and are thought to be glycolipid, possibly representing breakdown products of micro-organisms.

Malakoplakia may occur in association with a carcinoma but can also occur secondary to malnutrition or immunosuppression. For colonic disease, there is no satisfactory treatment.

Diversion colitis

This form of colitis occurs in colon which has been defunctioned, for example, for the management of complicated diverticular disease or to protect an anastomosis. It is usually asymptomatic but can be a cause of a rectal discharge of blood and mucus, and patients may complain of lower abdominal pain. Endoscopically, there is a diffuse inflammation looking similar to ulcerative colitis. Mild lesions are associated with loss of vascular pattern and granularity, but friability, spontaneous bleeding, and ulceration can occur. Histologically, the appearances may also suggest ulcerative colitis but may show the characteristic follicular lymphoid hyperplasia (see Chapter 26.2.1).

The colitis responds rapidly if intestinal continuity is restored. However, this is not always possible. Topical corticosteroids or 5-aminosalicylic acid may be helpful. If there is no benefit, enemas of sodium butyrate (40 mmol) or enemas of mixed short-chain fatty acids (butyrate, propionate, and acetate) have been reported to be very useful, but they have to be made up specially and have a distinctively unpleasant smell. However, it is a logical form of therapy since short-chain fatty acids are the preferred fuel substrate for epithelial cells of the distal colon, and it is thought that diversion colitis represents an energy-deficient disorder. In the intact colon, the short-chain fatty acids are generated by luminal bacteria from non-absorbable carbohydrates. When the faecal stream is diverted, there is insufficient substrate to generate sufficient fatty acids to maintain epithelial integrity.

Further reading

Allan RN, Keighley MRB, Alexander-Williams J, Rhodes JM, Hanauer SB, Fazio VW. *Inflammatory bowel diseases*, 3rd edn. Churchill Livingstone, New York, 1997.

Edwards CM, George B, Warren BF. Diversion colitis—new light through old windows. *Histopathology* 1999; **34**: 1–5.

Hyde GM, Thillainayagam AV, Jewell DP. Intravenous cyclosporin as rescue therapy in severe ulcerative colitis: time for a reappraisal? *European Journal of Gastroenterology and Hepatology* 1998; **10**: 411–13.

Lichtiger S, Present DH, Kornbluth A, Gelernt I. Cyclosporin in severe ulcerative colitis refractory to steroid therapy. *New England Journal of Medicine* 1994; **330**: 1841–5.

Marshall JK, Irvine EJ. Rectal aminosalicylate therapy for distal ulcerative colitis: a meta-analysis. *Alimentary Pharmacology and Therapeutics* 1995; **9**: 293–300.

Marston A. Vascular disease of the colon. In: Marston A, ed. *Vascular disease of the gut*. Edward Arnold, London 1986: 158–70.

Present DH, Rutgeerts P, Targan S, *et al.* Infliximab for the treatment of fistulas in patients with Crohn's disease. *New England Journal of Medicine* 1999; **340**: 1398–405.

Sutherland LR, May GR, Shaffer EA. Sulphasalazine revisited: a meta-analysis of 5-aminosalicylic acid in the treatment of ulcerative colitis. *Annals of Internal Medicine* 1993; **118**: 540–9.

Sutherland LR, Roth DE, Beck PL. Alternatives to sulphasalazine: a meta-analysis of 5ASA in the treatment of ulcerative colitis. *Inflammatory Bowel Diseases* 1997; **3**: 65–78.

Targan SR, Hanauer SB, van Deventer SJ, *et al.* A short-term study of chimeric monoclonal antibody cA2 to tumor necrosis factor alpha for Crohn's disease. Crohn's Disease cA2 Study Group. *New England Journal of Medicine* 1997; **337**: 1029–35.

26.2.3 Surgical management of ulcerative colitis

Timothy A. Cook and Neil Mortensen

Indications for operation
Acute illness
Chronic illness
The risk of carcinoma
Surgical options
Complete proctocolectomy
Colectomy with the formation of a Kock continent ileostomy
Colectomy with ileorectal anastomosis
Restorative proctocolectomy
Crohn's disease and indeterminate colitis
Emergency surgery for severe acute colitis
Restorative proctocolectomy
Historical background
The plane of rectal dissection
Pouch design
The ileal pouch–anal anastomosis
Is a temporary ileostomy always needed?
Postoperative management and course
Functional results
Long-term mucosal changes
Other long-term considerations
Further reading

Indications for operation

Although most patients with ulcerative colitis escape surgery, 2 per cent of those with distal colitis and up to 33 per cent of those with extensive disease will require an operation.

Acute illness

Severe acute colitis is characterized by the passage of more than six loose, bloody motions per day, together with systemic signs, including tachycardia, fever, and hypoalbuminaemia. The crucial points in management are early recognition, aggressive medical therapy, and regular review by a joint medical and surgical team. Stool is cultured to rule out a specific bacterial cause or the rare case of amoebic colitis. Corticosteroids are administered intravenously (400 mg hydrocortisone or 64 mg methylprednisolone daily) and rectally (100 mg hydrocortisone in 100 ml normal saline, infused into the rectum over 15 to 20 min). The patient is allowed only sips by mouth; fluids are given intravenously.

This treatment is successful in 75 per cent of patients, whose illness resolves within 5 to 7 days. They will feel well, passing two to three semiformed, bloodless motions each day. The remaining 25 per cent either deteriorate despite treatment or fail to recover completely; they may suffer a relapse when a normal diet is resumed. In both of these instances urgent surgery is indicated, since these patients will all require surgery in the subsequent few weeks and there seems little point in delay. A surgical procedure that does not result in a permanent ileostomy is also more acceptable to anxious and ill patients. The administration of antibiotics or steroids at a higher dosage has no therapeutic advantage and is not recommended. Intravenous cyclosporin is effective at reducing the need for surgery in the short term but so far it has not been shown to confer an advantage in the longer term.

If within the first 24 h of admission the patient has a pulse rate greater than 100 beats/min, a temperature over 38°C, or passes more than nine stools per day, or if mucosal islands are visible on a plain abdominal radiograph (Fig. 1), surgery will probably be required. Perforation and colonic dilatation (the upper normal limit of colon diameter is 5 cm) that fails to resolve rapidly when medical treatment is started or that develops during treatment is an indication for emergency surgery. Massive colorectal bleeding, although less common, may also require emergency surgical intervention. Intraoperative colonic lavage and colonoscopy can help locate the source of bleeding and may obviate the need for proctocolectomy.



Fig. 1. Plain abdominal radiograph in a patient with severe ulcerative colitis showing dilatation of the transverse colon and mucosal islands.

Chronic illness

The main indication for elective surgery is chronic illness that responds poorly to medical treatment or that is punctuated by frequent episodes of severe acute colitis. The threshold for referral by physicians and the degree of debility which many patients are prepared to tolerate has been reduced with the advent of restorative proctocolectomy. Physical, social, and employment factors should all be considered in the evaluation of a patient, as these can be affected by symptoms of the colitis or by the medication used in its treatment. The patient's age and the duration of illness are also important considerations when deciding on the appropriateness of surgical intervention.

Elective surgery is also performed in adolescents with growth retardation secondary to steroid treatment and in patients with long-standing disease and malignant transformation. Benign colorectal strictures produced by ulcerative colitis, and some extraintestinal manifestations such as extensive pyoderma gangrenosum (Fig. 2) and iritis (but not sclerosing cholangitis or ankylosing spondylitis), are rare indications for surgery.



Fig. 2. Pyoderma gangrenosum on the lower leg of a patient with ulcerative colitis. This healed after proctocolectomy.

The risk of carcinoma

Patients with ulcerative colitis have an increased risk of developing colorectal carcinoma (Fig. 3). This situation is particularly true for those who have chronic total colitis which started as a severe first attack in childhood. A precancerous stage is indicated by the finding of dysplasia in colorectal biopsy specimens. Surveillance programmes have been established to examine the large bowel at colonoscopy every 2 years when the disease has been present for 10 years. Biopsy specimens are taken either of specific lesions or from random sites of the mucosal field, as dysplasia in one area can indicate the presence of an unsuspected malignancy elsewhere.

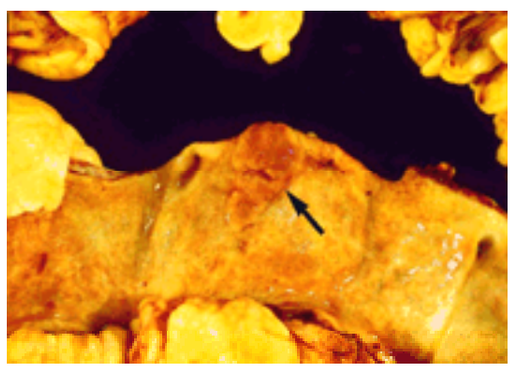


Fig. 3. Post-fixation colectomy specimen from a patient with a small carcinoma in the sigmoid colon (arrowed) complicating ulcerative colitis.

Results from these programmes suggest that the probability of developing a carcinoma is 3 per cent at 17 years, 5 to 7 per cent at 20 years, and 9 to 13 per cent at 25 years. Very few, if any, cancers develop before a 10-year history of colitis.

Unfortunately, dysplasia is not an ideal marker of malignancy. Its detection is subject to observer variation, and cancer can develop in the absence of changes at a distant site. Identifying dysplastic changes, however, is the only currently available method for detection of a precancerous state, and is most reliable when severe or high-grade changes are found in a biopsy from a villous or polypoidal lesion (Fig. 4). A similar degree of dysplasia from an area of flat mucosa is less likely to be associated with synchronous carcinoma. The finding should be verified by a second, independent, pathologist and colectomy should be advised only if dysplastic change is present at two sites in the colon. Surgery is also recommended when low-grade dysplasia is found in a raised lesion; its presence in flat mucosa is an indication for increased vigilance only.

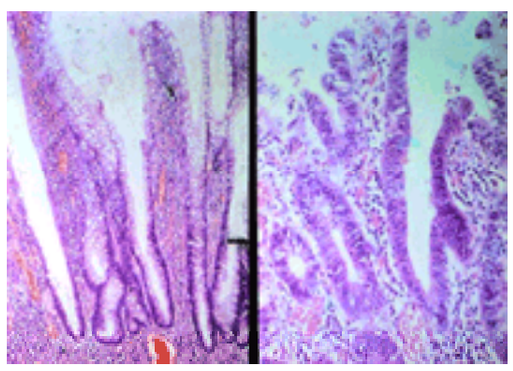


Fig. 4. Photomicrograph of a section from a colonic biopsy showing severe dysplasia complicating ulcerative colitis.

Until a more sensitive and specific marker of premalignancy is found, the true value of screening for colorectal cancer in ulcerative colitis remains uncertain.

Surgical options

Until quite recently, complete proctocolectomy (or panprocto-colectomy) was the best operation for patients with ulcerative colitis who required surgery. This position has, however, been challenged by the advent of restorative proctocolectomy which, with few exceptions, is now considered the first-choice elective surgical treatment for this condition. Restorative proctocolectomy has also forced a re-evaluation of the recommended management for those undergoing urgent or emergency operations. The various options available and their indications are shown in Table 1.

Advantages	Curative; non-operative	Sphincter saving and, if successful, avoids stoma and ileal/ileo-rectal dysfunction	Curative; excellent
Disadvantages	Ileostomy; retained (excising) rectum carries risk of cancer; 4–6 stages per day	Non-curative; cancer risk persists; proctectomy for cancer or dysplasia 5–10 yrs on	Stoma; long-term, small risk of cancer 4–6 stages per day
Complications	Stoma—stricture 10–15% per year; prolapse 5–10% per year; small bowel obstruction 5–15% per year Trouble-free function minimal, even in elderly patients	Small bowel obstruction in 5–15% per year; synchronous cancer in 1–2% at first check	Bowel irritation, sepsis, stenosis, small bowel obstruction in 5–15% per year; pouchitis, debilitating dysfunction minimal, even in elderly patients; small risk of cancer
Contraindications	None	Sphincter incontinence; severe rectal disease; rectal dysplasia or cancer	Crohn's disease; rectal cancer; fecal fluid

Table 1 Comparison of the three operations most commonly used in the treatment of ulcerative colitis

Complete proctocolectomy

Until the development of the Brooke ileostomy and advanced materials for stoma appliances, surgery was a last resort. In the 1950s, however, proctocolectomy became the established procedure.

This operation has several advantages: all diseased tissue is excised at one operation, it removes the risk of cancer, it is a well tried and usually uncomplicated procedure, and the patient can return to normal activities expeditiously. Unfortunately, the patient is left with a permanent ileostomy, which may have serious social and psychological consequences. In the long term the ileostomy may require revision and small bowel obstruction may develop due to adhesions. Chronic perineal sinus occurs in 5 to 10 per cent of cases and delayed healing of the perineal wound is not uncommon. Sexual dysfunction arising from damage to pelvic autonomic nerves occurs in 0.5 to 1 per cent of male patients. Dyspareunia may be a result of adhesions in female patients.

The advent of restorative surgery has led to a choice of procedures. We currently recommend complete proctocolectomy only in elderly patients, particularly those with weak anal sphincters, and in those unwilling to subject themselves to a more complicated operation with the possibility of additional morbidity and a longer hospital stay. Some patients feel restorative surgery would not be compatible with their lifestyle. Overall in Oxford, one-quarter of our patients elect to have a complete proctocolectomy; the majority choose the restorative procedure.

Procedure

With the patient in the Lloyd-Davis position, the abdomen and perineum are prepared. The rectum is washed out with an antiseptic and a close perianal purse string is placed. The previously marked ileostomy site is trephined. The abdomen is opened through a long mid-line incision, the colon mobilized, and the vessels divided comfortably near the bowel, from the caecum to sigmoid. The pelvic dissection can be carried out in the close rectal or mesorectal plane (Fig. 5). The perineal operator meanwhile makes a close perianal incision and develops the intersphincteric plane. Infiltration with dilute adrenaline may aid this dissection. The perimuscular dissection and intersphincteric technique may improve perineal wound healing and reduce the risk of perineal sinus. The specimen is removed through the perineal wound. The perineum is closed in layers with absorbable sutures. Suction drains are placed from above.

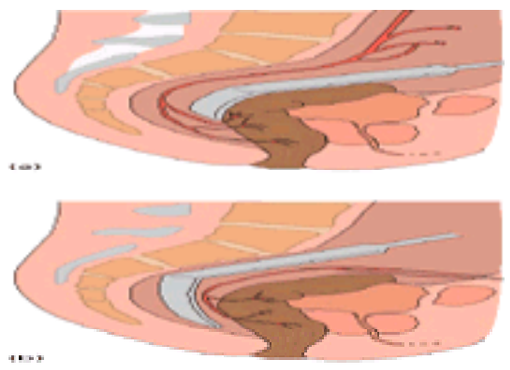


Fig. 5. Diagram to show the difference between the close rectal (a) and the mesorectal (b) plane used in pelvic dissection.

The terminal ileum is delivered through the stoma trephine and everted to make a 2- to 3-cm spouted (Brooke) ileostomy (Fig. 6). The lateral space is closed to prevent small bowel herniation.

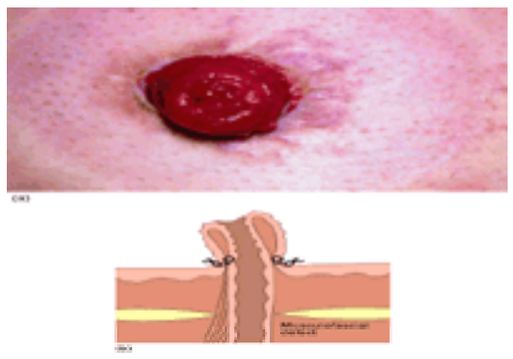


Fig. 6. (a) Spouted Brooke ileostomy and (b) diagram showing its formation. The spout should be long enough to protrude into the bag and should ideally be angled slightly downwards. Care must be taken when forming the musculofascial defect. If the defect is too large there is a risk of parastomal herniation and there is a chance of retrostomal stenosis if the defect is too small.

Complications include pelvic bleeding and intra-abdominal sepsis. Small bowel adhesion obstruction occasionally requires laparotomy. Stoma protraction or retraction and parastomal herniation requiring revisional surgery occurs in 5 to 10 per cent of patients.

Stomatherapy

Specialist nurse support in stomatherapy is essential. The stomatherapist will mark the appropriate stoma site and provide information on the management of ileostomy preoperatively. Careful construction of the ileostomy is essential to avoid complications. The spouted Brooke ileostomy overhangs the stoma appliance, allowing effluent to be discharged into the bag. Retraction of the stoma such that it becomes flush with the skin predisposes to leakage and subsequent skin excoriation. Passage of large amounts of flatus sometimes presents a problem, but ileostomy bags with gas filters have recently been introduced to overcome this. The stomatherapist is vital for postoperative care in stoma management and for psychological help in adjusting to an altered body image.

Colectomy with the formation of a Kock continent ileostomy

This procedure retains the advantages of a complete proctocolectomy and replaces the spouted Brooke ileostomy with a flush, continent stoma and an intra-abdominal reservoir constructed of ileum (Fig. 7). Patients do not have to wear a stoma bag and pass a catheter into the pouch in order to empty it. Early technical problems, particularly with the nipple valve, have been resolved, but the advent of restorative proctocolectomy has virtually abolished the indications for this procedure. It could be considered in a patient who has already had a complete proctocolectomy and who expresses a strong wish to be rid of the ileostomy.

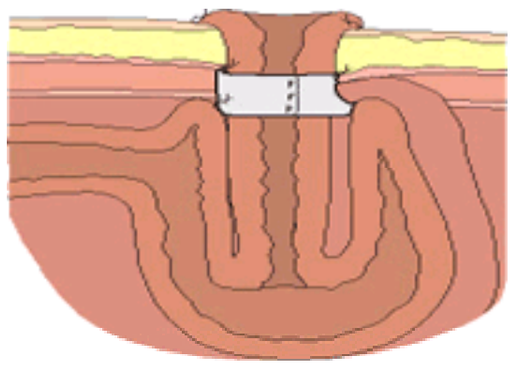


Fig. 7. Kock-pouch continent ileostomy showing a Marlex mesh collar reinforcing the nipple valve.

Colectomy with ileorectal anastomosis

This procedure is attractive because it is relatively quick and straightforward, and it can be associated with a good functional outcome. However, its use remains controversial because, by leaving behind the rectum, it is associated with a continuing risk of carcinoma. It might be considered in a patient with a compliant rectum which is relatively spared from disease. The presence of a functioning sphincter is essential, and the patient must be reliable as lifelong monitoring is required. Coexisting sclerosing cholangitis or portal hypertension, which would make rectal dissection hazardous, are also indications for this operation. It is also an option, in children, allowing them to pass through adolescence and complete their schooling without a stoma or until a pouch is indicated.

Restorative proctocolectomy

This is the procedure of choice for the elective treatment of ulcerative colitis. It is not indicated in the presence of Crohn's disease, a lower-third rectal cancer (a high early cancer does not exclude restorative proctocolectomy if local tumour clearance can be achieved without sphincter excision), or poor anal sphincter function. Preoperative work-up should therefore include anal manometry and anal ultrasound. Current operations involve the creation of an ileal reservoir of various designs, followed by an anastomosis to the anal canal.

Crohn's disease and indeterminate colitis

Diagnostic difficulties between ulcerative colitis and Crohn's disease are well recognized. Distinguishing features are shown in Table 2. Considerable overlap may occur making it impossible for the pathologist to distinguish between the two diseases. Indeterminate colitis represents a spectrum of disease and further information should be gathered with repeat biopsies from the rectum. When radiological, histopathological, and clinical features are considered together, patients with indeterminate colitis can still be judged to incline more to ulcerative colitis or Crohn's disease. Patients with indeterminate colitis can still be offered reconstructive surgery but should be warned of the greater potential for failure and its consequences. Reconstruction in the presence of Crohn's disease is discussed in Chapter 26.2.4.

	Ulcerative colitis	*Crohn's disease*
Macroscopic		
Distribution	Rectum and colon	Whole gastrointestinal tract
Rectum	Involved	Often spared
Anal involvement	Rare	Common
Risk of malignancy	10 per cent at 25 years	Probably similar to colitis
Strictures	Very rare	Common
Fistula	Never	Common
Microscopic		
Bowel wall involvement	Mucosa and submucosa	Full thickness
Granulomas	None	60–70 per cent
Crypt abscess	Common	Rare
Fissuring	Absent	Common

Table 2 Histopathological distinction between ulcerative colitis and colonic Crohn's disease

Emergency surgery for severe acute colitis

Both complete proctocolectomy and subtotal colectomy, with formation of an ileostomy and mucous fistula, have been used in the treatment of acute colitis. However, with the possibility of later restorative surgery, subtotal colectomy has become the procedure of choice. Timing of surgery is obviously critical, and with careful management the ‘late case' should be avoidable. Turnbull introduced the principle of ‘blow hole' colostomy, in which the thin, friable, perforating colon was managed by multiple ostomies to the abdominal wall. This avoided the hazardous resection of a difficult colon, but it is rarely required nowadays. The splenic flexure is usually the most dangerous area in an emergency colectomy, and this part of the procedure must be undertaken with great care, packing off the area in case of an inadvertent perforation. If omentum is adherent to the colon it should be left in place and should not be preserved.

The best way to manage the mucous fistula remains undecided. It can be sewn to the skin at the lower end of the laparotomy wound should the rectum break down or, less conventionally, closed and placed subcutaneously. This will avoid creation of a second stoma but allow for discharge of blood and pus through the wound should the rectum break down. In either case, subsequent identification of the rectum is straightforward. Closure of the rectum at the peritoneal reflection (as in Hartmann's procedure) carries the risk of possible perforation of the inflamed stump and abscess formation. Conservative proctocolectomy which includes removal of the rectum to the level of the pelvic floor is not advised. Proctocolectomy is associated with a high degree of morbidity and may make any subsequent pelvic dissection technically difficult. It is only indicated in the rare patient with colitis and rectal bleeding. The various options for managing the rectum after subtotal colectomy are shown in Fig. 8.

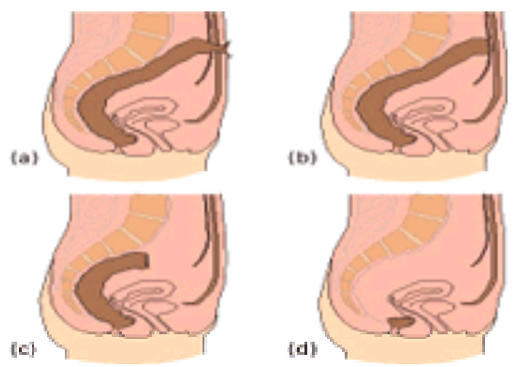


Fig. 8. Options for the management of the retained rectum after subtotal colectomy for ulcerative colitis. (a) Conventional mucous fistula. (b) Subcutaneous closure. (c) Hartmann's procedure. (d) Conservative proctocolectomy.

Restorative proctocolectomy

Historical background

In 1947, Ravitch and Sabiston demonstrated that in both animals and humans it was possible to remove the rectal mucosa and leave denuded but functional rectal muscle. This mucosal proctectomy could be followed by creation of an ileoanal anastomosis, pulling the ileum through the rectal muscle cuff and sewing it to the modified skin of the anal canal. Ravitch performed this operation in two young adults with ulcerative colitis, who were subsequently continent. Although the technique was tried by a number of other surgeons over the following decade, the procedure never gained widespread acceptance because of operative complications, incontinence, intolerable frequency, and perianal dermatitis. However, the considerable success of mucosal proctectomy and pull-through procedures for children with Hirschsprung's disease tempted some paediatric surgeons to try a similar approach in young patients suffering from ulcerative colitis, with encouraging results.

A better understanding of the mechanisms of normal continence and defecation, together with the introduction of the Kock pouch, were the stimuli for the development of the pelvic ileal reservoir. The use of a single narrow tube of ileum as a replacement for rectum resulted in very poor reservoir function. Parks and colleagues introduced a pelvic triple-loop ileal reservoir, which was sewn to the anus after mucosal proctectomy. Parks' original operation has undergone considerable modifications since it was first described. Several of these, together with other important issues surrounding restorative proctocolectomy, are discussed below.

The plane of rectal dissection

Many patients suffering from ulcerative colitis are young and sexually active. Preservation of the pelvic nerves, and in particular the nervi erigentes, is of great importance. Early proctocolectomy was associated with a 10 per cent incidence of impotence, but adoption of a perimuscular plane for rectal dissection reduced the incidence of this complication to below 0.5 per cent. Similar results have been achieved by dissection in the avascular plane around the mesorectum posteriorly and laterally but close to the bowel wall anteriorly (Fig. 5). This technique is reported to be quicker and associated with less blood loss; it is widely used in North America. Formal comparison of the two methods and information on the relative incidences of sexual dysfunction are not available.

Pouch design

P>The pouch originally described by Parks and Nicholls was a triple-loop S pouch with a long efferent limb which emptied poorly and frequently required intubation to effect complete evacuation. This problem has been overcome by shortening the length of the efferent segment. A two-loop J pouch, four-loop W pouch, and lateral isoperi-staltic or H pouch have also been described. These pouches can be hand-sewn or stapled and use approximately 40 to 50 cm of terminal ileum. The various designs are illustrated in Fig. 9; their function is compared below. Figure 10 shows a triplicated S pouch *in situ* in the pelvis.

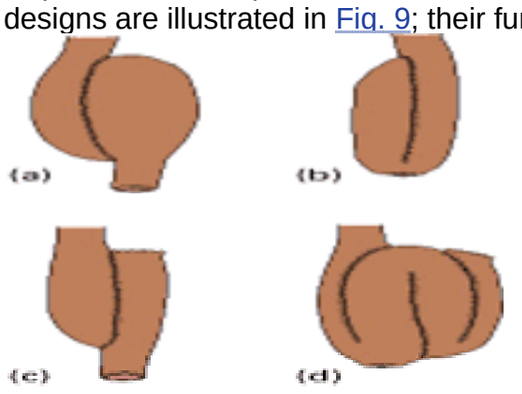


Fig. 9. The four pouch designs. The J and W pouches are most often used now. (a) Triplicated or S. (b) Duplicated or J. (c) Lateral or isoperistaltic. (d) Four-loop or W.

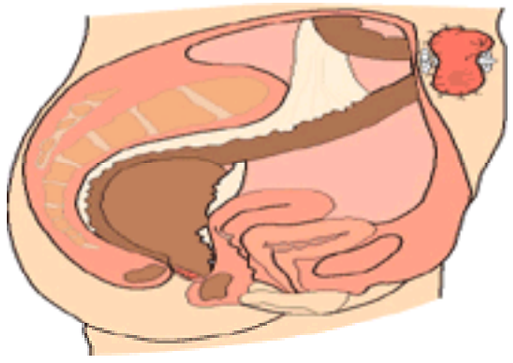


Fig. 10. Triplicated S pouch in the pelvis with a covering loop ileostomy. Care has to be taken not to make the efferent limb too long in this design (see Fig. 9(a)).

The ileal pouch–anal anastomosis

Is mucosectomy necessary?

When restorative proctocolectomy was first introduced, the rectum was transected 10 cm above the pelvic floor and the mucosa striped off the underlying muscle by a tedious endoanal dissection up from the dentate line. This achieved the aim of excising all potentially diseased mucosa, but also removed the anal transitional zone, an area of cuboidal rather than columnar mucosa which extends for approximately 1.5 cm above the dentate line and has a rich sensory nerve supply.

The rectal muscle cuff has not been shown to be essential for the appreciation of pouch fullness; indeed it may increase the incidence of pelvic sepsis if left in place. On the other hand, the anal transitional zone is thought to play an important role in sensory discrimination as well as in continence. The postoperative functional results of restorative proctocolectomy are better if this mucosa is preserved. However, there is a long-term risk from leaving behind potentially diseased mucosa and retention of the anal transition zone is not universally accepted. Pathological changes suggestive of ulcerative colitis have been found in anal canal mucosa from patients undergoing proctocolectomy. Even more worrying are reports of dysplasia and adenocarcinoma in this zone. In Oxford, no dysplasia has been found in a review of over 50 consecutive specimens, and clinical carcinoma arising in this area in patients with long-standing ulcerative colitis is virtually unheard of.

Based on the current balance of information, mucosectomy can be avoided in the absence of rectal mucosal dysplasia. However, mucosectomy should be carried out if rectal dysplasia is diagnosed preoperatively or on subsequent histological examination of the surgical specimen which must, for this reason, be meticulous. Figure 11 illustrates the possible levels of ileoanal anastomosis.

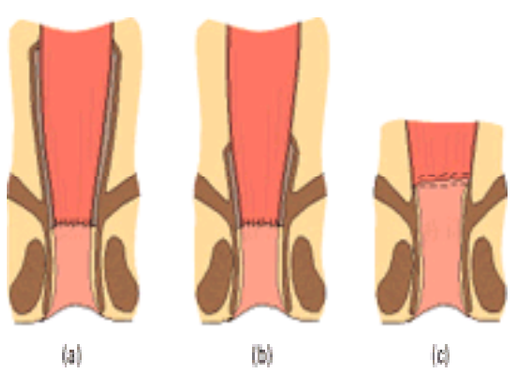


Fig. 11. The evolution of the ileoanal anastomosis from a long muscle cuff and mucosectomy (a), short muscle cuff and mucosectomy (b), and flush transection at the

pelvic floor without mucosectomy and preserving the anal transitional zone (ATZ) (c).

A sutured or stapled anastomosis?

The question of whether the ileoanal anastomosis should be hand-sutured or stapled is less controversial. Minor defects in continence are not uncommon after manual endoanal anastomosis. Such deficiencies may be caused by retraction and dilatation injury to the anal sphincters at the time of operation. Certainly, impaired internal sphincter function has been noted. It was hoped that circular stapling devices would improve continence, as they produce minimal mechanical sphincter damage. However, objective evidence suggests that a significant fall in internal anal sphincter activity still occurs after stapled anastomosis, although the cause of this problem remains obscure. There are fewer septic complications following stapled anastomosis, but the only demonstrable improvement in pouch–anal function after stapling the anastomosis is a reduction in nocturnal evacuation. However, stapling devices do have the advantage of speed and can produce a great saving in time if used throughout the operation (Fig. 12). Four-loop W reservoirs cannot be constructed easily with staplers.

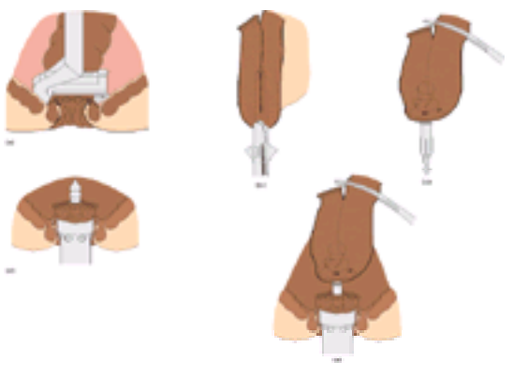


Fig. 12. Steps in the stapled construction of an ileoanal reservoir. (a) The rectum is transected at the top of the anal canal. (b) The two-loop reservoir is constructed. (c) An opening is made at the most dependent point in the pouch and the anvil shaft complex placed inside with a purse string. (d) The anastomosing gun is placed in the anal remnant and the trocar advanced. (e) The two parts of the anastomosing gun are docked and the stapled anastomosis completed.

Is a temporary ileostomy always needed?

Until quite recently, all authors agreed that the reservoir and ileoanal anastomosis should be temporarily defunctioned with a loop ileostomy. This accepted practice has now been challenged. The risk of leak at the pouch–anal anastomosis is increased in patients over 40 years, men, and those on high-dose steroids preoperatively. There are no randomized controlled trials assessing the effect of omitting a defunctioning stoma. However, if the surgeon is familiar with the technique of pouch surgery, the operation is technically straightforward, and none of the risk factors are present, the ileostomy can be omitted as, based on present evidence, this does not seem to be associated with additional morbidity. This avoids a subsequent surgical procedure which has its own attendant complications and which requires hospital admission.

Postoperative management and course

If a defunctioning ileostomy is formed, a period of 6 to 8 weeks is allowed to elapse before closure. A contrast study of the pouch prior to ileostomy closure excludes small leaks. If an ileostomy is omitted, the pouch is intubated with a wide-bore balloon catheter for at least 5 days to divert the faecal stream. The perianal region should be protected by barrier cream and a careful record must be kept of fluid balance. Patients can be allowed home 10 to 14 days postoperatively if they are well and confident. Early review is sensible in order to monitor and advise on pouch function. Constipating medication will reduce stool frequency in the first few months.

Very few deaths have been reported after restorative proctocolectomy, no doubt as a result of careful patient selection. However, there is appreciable morbidity which declines as the surgeon becomes increasingly familiar with the techniques involved. Pelvic sepsis, with or without anastomotic breakdown, develops in 8 to 25 per cent of cases and can be difficult to manage as it cannot drain freely. Healing by fibrosis can impair the eventual functional result. Adhesion obstruction also occurs frequently and further laparotomy is needed in up to 15 per cent of patients. Strictures are seen in 10 per cent of patients, but usually respond to simple dilatation. Fluid and electrolyte loss from the defunctioning ileostomy can produce a dehydration syndrome which may be exacerbated by steroid withdrawal.

Overall, 75 per cent of patients recover uneventfully and 25 per cent experience serious morbidity.

Functional results

Following restorative proctocolectomy, most patients defecate spontaneously about five to seven times every 24 h, and will be able to defer defecation without urgency. Few patients suffer frank incontinence, but minor imperfections such as spotting or soiling occur in up to 33 per cent during the day and 56 per cent at night. Some 20 per cent of patients use codeine phosphate or loperamide hydrochloride.

There are no prospective trials assessing the functional outcome of different pouch designs. The best clinical results are associated with a large capacity, compliant pouch which empties completely and which sits above normally active anal sphincters. Total daily stool volume, postoperative pelvic sepsis, and pouchitis are also important. Some evidence suggests that three- and four-loop pouches produce the lowest stool frequency because they are more capacious than those formed from two loops of ileum. However, this is disputed. It is recognized that the pouch functions in part by modifying terminal ileum motility, but the relationship between this, capacity, and compliance remains unexplained.

Outright failure which requires pouch excision increases with time and occurs in up to 10 per cent of patients after 10 years. Principal causes of failure are persistent pelvic sepsis, undiagnosed Crohn's disease, unacceptable stool frequency, or pouchitis.

Long-term mucosal changes

The long-term health of the pouch mucosa is of more concern. Early studies on Kock pouches showed chronic inflammation as well as villous atrophy, and similar changes are seen in most pelvic reservoirs. Colonic metaplasia has been noted on histological examination and is supported by studies using histochemical and immunological markers. The cause of these changes is not known, but probably involves bacteriological and immunological factors. The risk of future dysplasia and even frank malignancy cannot be ignored but, to date, no such occurrence has been reported in either a pelvic or a Kock pouch.

More acute inflammatory changes also occur and can produce pouchitis—a condition which is characterized by diarrhoea in the presence of endoscopic and histological features of acute inflammation (Fig. 13). It has only been diagnosed unequivocally in patients who had ulcerative colitis, affecting up to 30 per cent of those undergoing restorative proctocolectomy. Pouchitis is no more common in patients who had extensive disease than in those whose colitis was left-sided or distal. The pathogenesis of pouchitis is not known and is almost certainly multifactorial. Contact between the pouch mucosa and ileal contents with stasis and bacterial overgrowth are essential features. The importance of genetic predisposition, the presence of specific bacterial strains, epithelial defects, and immunological abnormalities is uncertain, but these have all been implicated as causative agents. Outlet obstruction, pouch ischaemia, and Crohn's disease are important differential diagnoses which should be excluded.

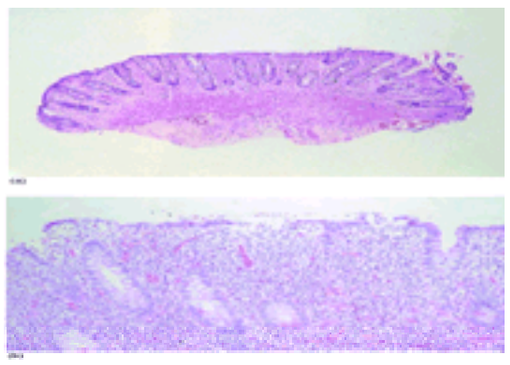


Fig. 13. (a) and (b) Photomicrographs showing the typical change which occurs in the pouch mucosa.

In the absence of a controlled therapeutic trial, the treatment of pouchitis is empirical. Metronidazole is probably the most commonly used agent and may function as an immunosuppressive agent as well as an antibiotic. Enemas containing steroid or 5-aminosalicylic acid derivatives, oral sulphasalazine, and short courses of oral steroids can also be used. Most episodes of pouchitis will resolve when treated with one or a combination of these agents. Very rarely, pouchitis is unresponsive and must be treated by the formation of a defunctioning ileostomy or even pouch excision.

Other long-term considerations

Almost all patients undergoing restorative proctocolectomy are satisfied with the outcome of their operation and over 90 per cent prefer the pouch to a permanent ileostomy because of increased self-confidence, cleanliness, sexual self-image, lack of interference with social and sports activities, and ease at work. Impaired sexual function is related to the proctocolectomy rather than the pouch.

No serious long-term nutritional sequelae have emerged after restorative proctocolectomy, but a mild microcytic anaemia associated with a low serum iron concentration is seen in up to 30 per cent of patients. Serum vitamin B_{12} levels may be marginally lowered but serum folate concentration is normal, and there is no abnormality of fat and fat-soluble vitamin absorption or of liver function.

Further reading

Berry AR, De Campos R, Lee ECG. Perineal and pelvic morbidity following perimuscular excision of the rectum for inflammatory bowel disease. *British Journal of Surgery* 1986; **73**: 675–7. [Report of 115 patients undergoing perimuscular proctectomy.]

Cohen Z, McLeod RS, Stephen W, Stern HS, O'Connor B, Reznick R. Continuing evolution of the pelvic pouch procedure. *Annals of Surgery* 1992; **216**: 606–12. [Analysis of risk factors for anastomotic leak following restorative proctocolectomy.]

Everett WG, Pollard SG. Restorative proctocolectomy without temporary ileostomy. *British Journal of Surgery* 1990; **77**: 621–2.

Farouk R, Pemberton JH. Surgical options in ulcerative colitis. *Surgical Clinics of North America* 1997; **77**: 85–94. [Review article on surgical management of ulcerative colitis.]

Hewett PJ, Stitz R, Hewett MK. Comparison of the functional results of restorative proctocolectomy for ulcerative colitis between the J and W configuration ileal pouches with sutured ileoanal anastomosis. *Disases of the Colon and Rectum* 1995; **38**: 567–72. [Retrospective analysis of function of 35 patients with W-pouches and 52 with J-pouches.]

Lennard-Jones JE *et al.* Precancer and cancer in extensive ulcerative colitis: findings among 401 patients over 22 years. *Gut* 1990; **31**: 800–6. [Assessment of cumulative risk of developing carcinoma in 401 patients with ulcerative colitis over a 22-year period.]

Nicholls RJ, Lubowski DZ. Restorative proctocolectomy: the four loop (W) reservoir. *British Journal of Surgery* 1987; **74**: 564–6.

Parks AG, Nicholls RJ. Proctocolectomy without ileostomy for ulcerative colitis. *British Medical Journal* 1978; **ii**: 85–8. [Original description of pelvic pouch procedure.]

Samarasekera DN, Stebbing JF, Kettlewell MGW, Jewell JP, Mortensen NJMcC. Outcome of restorative proctocolectomy with ileal reservoir for ulcerative colitis: comparison of distal colitis with more proximal disease. *Gut* 1996; **38**: 574–7. [Retrospective comparison of morbidity and functional outcome following restorative proctocolectomy in patients with distal *vs* total colitis.]

Thompson-Fawcett MW, Mortensen NJMcC. Anal transition zone and columnar cuff in restorative proctocolectomy. *British Journal of Surgery* 1996; **83**: 1047–55. [Review article on the anal transition zone and columnar cuff.]

Ziv Y, Fazio VW, Church JM, Lavery IC, King TM, Ambrosetti P. Stapled ileal pouch anal anastomoses are safer than handsewn anastomoses in patients with ulcerative colitis. *American Journal of Surgery* 1996; **171**: 320–3. [Retrospective comparison of stapled *vs* handsewn ileoanal anastomoses.]

26.2.4 Crohn's disease of the colon

Neil Mortensen and Timothy A. Cook

Clinical features
Investigation
Medical management
Indications for surgery
Specific surgical treatment for colonic disease
Elective surgery
Perianal disease
Medical management
Surgical management
Rectovaginal fistula
Haemorrhoids
Further reading

Clinical features

These depend upon the site affected. There are three typical patterns. Small bowel disease usually affects the ileocaecal segment. The patient presents with colicky abdominal pain, diarrhoea, and weight loss. Colonic disease may present as colitis with bloody diarrhoea, urgency, and frequency, similar to that of ulcerative colitis. Discontinuous disease with fibrosis and stenosis may, however, cause diarrhoea without bleeding or colonic obstructive symptoms. Fistulation to adjacent organs may give rise to the distinct clinical feature of a colovesical or rectovaginal fistula. Ileocolic fistula between the ileum and the sigmoid colon is usually due to ileal disease. Colonic disease is secondary and may cause an increase in diarrhoea or no symptoms at all. Perianal disease is a particular feature of Crohn's disease: a chronic anal fissure may be the first presenting symptom.

Investigation

Barium enema may show areas of discontinuous disease. In patients with stricture (Fig. 1), in whom the diagnosis may be in doubt, a colonoscopy will show mucosal changes, and information from multiple biopsies will allow the exact distribution of disease to be documented (Fig. 2). Colonic stricture should be carefully investigated since it may indicate occult malignancy. Blood tests, including measurement of erythrocyte sedimentation rate or viscosity and C-reactive protein levels will give an indication of disease activity. Serum albumin levels indicate nutritional status.

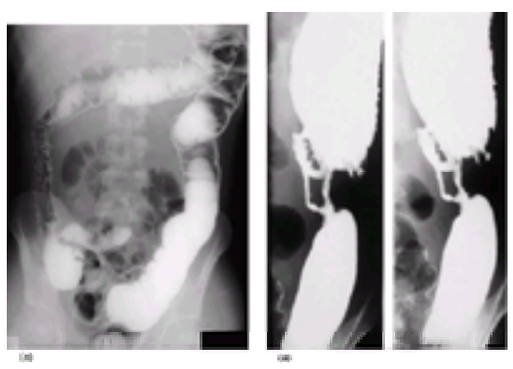


Fig. 1. Barium enema showing (a) a segment with Crohn's disease in the right colon with normal mucosa elsewhere, (b) views of the left colon in a patient with colonic Crohn's disease. There is a stricture with fistulation.

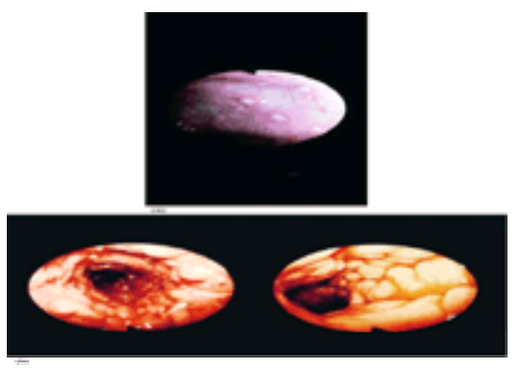


Fig. 2. (a) Colonoscopic appearance of early Crohn's disease with aphthous ulceration. (b) Colonoscopic appearance of severe Crohn's colitis.

Medical management

There are two clinical pictures which must be distinguished in the management of colonic Crohn's disease. Disease flares, which give rise to mucosal ulceration and oedema, can be successfully treated by administration of steroids, and occasionally azathioprine. This treatment is similar to that used for ulcerative colitis. Fibrosis and thickening, with obstruction or formation of a fistula and an abscess often requires surgical treatment, however. Before surgery is undertaken nutritional status must be assessed. Patients with extensive gut disease or sepsis may be severely nutritionally depleted: serum albumin level and weight loss history are a rough guide to the degree of the nutritional problem. Intravenous nutrition or, in some suitable cases, nasoenteric feeding may be necessary.

Indications for surgery

As with small intestinal disease, the management of Crohn's disease of the colon is medical unless a specific complication is present (Table 1).

Dangerous acute complications
Perforation, dilatation
Severe haemorrhage
Chronic subacute complications
Abscess
Fistula
Obstruction
The failure of a severe acute attack of Crohn's colitis to respond to drug therapy
Chronic disability or ill health
Risk of developing carcinoma, though this is less common than in ulcerative colitis

Table 1 Complications of Crohn's disease.

Specific surgical treatment for colonic disease

Although Crohn's disease of the colon may behave like ulcerative colitis, surgical treatment is different in a number of respects.

Emergency surgery

The usual indication is acute fulminating Crohn's colitis with bleeding, toxic dilatation, or perforation. It is important to note that perforation can occur without toxic dilatation. The procedure of choice is usually a subtotal colectomy with formation of an end ileostomy. A mucous fistula may be constructed initially if there is relative rectal sparing. Oversewing of the rectal stump at the level of the peritoneal reflection is advised only if the rectum is normal, since in the presence of active Crohn's disease there is the risk of breakdown and intra-abdominal sepsis. The formation of a mucous fistula also has the advantage of allowing topical steroid irrigation of the defunc-tioned distal rectum.

There is a limited place for primary resection and ileorectal anastomosis as an emergency procedure, for example in patients who are fit, in whom there is no pre-existing sepsis, and when the rectum is spared. If the operative field is contaminated, the risk of dehiscence, leakage, or fistula formation is high and it would be wise to employ a proximal loop ileostomy in these cases. The indications for a primary proctocolectomy in emergency situations are also limited. If there is severe bleeding from the rectum there may be no choice, but secondary rectal excision is best performed at a later date, when the patient's general condition has improved.

Use of an emergency defunctioning loop or split ileostomy should be considered. Defunctioning of diffuse or multiple site colonic disease may allow it to resolve, and in about one-third of patients gut continuity can be restored without resection. A loop ileostomy has the advantage that closure does not require a laparotomy, but the disadvantage that, unless constructed properly, overspill occurs and defunction of the distal bowel is not complete. Unfortunately up to 60 per cent of patients relapse when continuity is restored.

Following an emergency operation for Crohn's colitis the patient is usually left with a mucous fistula or oversewn rectal stump. The diagnosis of Crohn's disease can be confirmed on the colectomy specimen. Restorative proctocolectomy is contraindicated and patients in whom restoration of intestinal continuity may be possible have to be carefully selected. An ileorectal anastomosis is only advised where there is rectal sparing, minimal small bowel disease, and quiescent anal disease. They must therefore be carefully investigated by a rectal examination, proctoscopy, anal manometry, and anal ultrasound. Continuing inflammation in the rectal stump can be due to defunction colitis as well as Crohn's disease, and this should be borne in mind when interpreting rectal biopsies.

Elective surgery

Segmental colectomy

In some patients with colonic Crohn's disease where a single localized segment is affected, causing either a stricture, fistula, or abscess, it is reasonable to perform a limited resection with an immediate colocolonic anastomosis. This preserves macroscopically normal colonic tissue and may have a functional advantage in preserving colonic water handling. Allan *et al.* have reported their experience with 36 patients treated by segmental colectomy. The 10-year reoperative rate for recurrent disease was 66 per cent, compared with 53 per cent among patients undergoing subtotal colectomy and ileorectal anastomosis. There was no clinical evidence of an anastomotic leak in 29 patients, suggesting that the procedure can be safe. Data from the Mayo Clinic show recurrence rates of 54 per cent at 10 years with similar low complication rates.

Subtotal colectomy and ileorectal anastomosis

This is indicated in patients with severe diffuse colonic disease and rectal sparing. It is particularly indicated in younger patients, since it allows them to complete their education, start a career, and begin a family without the risk of sexual dysfunction or the disability associated with a stoma and perianal wound. This operation is contraindicated if the anal sphincter has been damaged as a result of previous perianal surgery or severe perianal disease, or if the patient has extensive rectal disease. In the presence of extensive small bowel disease, recurrence rates after ileorectal anastomosis may be high.

The majority of patients have a good functional result with an ileorectal anastomosis, provided that the reservoir function of the rectum can be maximized by anastomosis at the rectosigmoid junction. Most patients will have less than six bowel actions a day and troublesome diarrhoea is unusual. The worst functional outcome is seen in patients who have obvious macroscopic disease of the rectum.

The major early complication is anastomotic dehiscence, which occurs in between 5 and 30 per cent of patients. A covering loop ileostomy may reduce the incidence of this complication, although if this is thought necessary it may not be wise to perform the ileorectal anastomosis at all. The operative mortality ranges from 0 to 5 per cent. Recurrence rates are shown in Table 2: these seem to be higher than those in similar patients undergoing proctocolectomy and ileostomy.

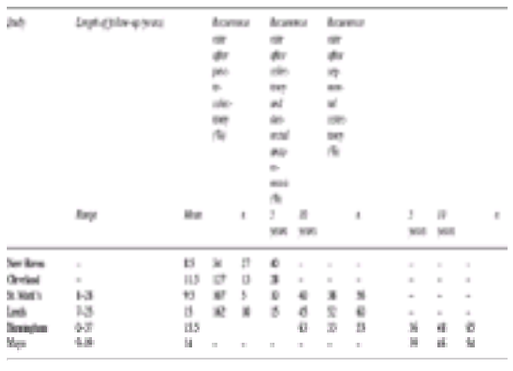


Table 2 Recurrence rates after resection of colonic Crohn's disease.

Recurrence of disease does not always mean that a proctocolectomy and ileostomy is inevitable. Recurrence at the ileorectal anastomosis can be dealt with medically or by a further resection and ileorectal anastomosis.

Restorative proctocolectomy

Crohn's disease is usually regarded as a contraindication to an ileoanal pouch procedure. Complication rates are higher when compared with surgery for ulcerative colitis or familial adenomatous polyposis. However, restorative operations such as segmental resection or ileorectal anastomosis are part of accepted surgical practice for Crohn's disease despite higher recurrence rates than are seen following proctocolectomy alone. One concern is that pouch failure will lead to excessive loss of small bowel, but experience with Kock pouches suggests this is unlikely to be the case. Pouch surgery in Crohn's disease should be considered in the context of other restorative procedures and not compared with surgery for completely different conditions such as ulcerative colitis. In the few patients without anal or ileal disease, restorative proctocolectomy should not be dismissed out of hand.

Total colectomy with formation of an end ileostomy and mucous fistula

Total colectomy with formation of an end ileostomy and a mucous fistula or oversewing of the rectal stump has the advantage of safety and reduces the risk of recurrence proximal to an anastomosis. The recurrence rates in the ileum and ileostomy are similar to those seen after a proctocolectomy. The main indication is as an emergency procedure for severe colitis, but subsequent restoration of continuity is often not possible. This operation can also be used for the treatment of severe

proctitis or severe perianal disease in patients likely to suffer delayed perianal wound healing or persistent perineal sinus. Continuing sepsis in the perineum, however, often means that the rectal stump has to be removed later. Operative mortality is of the order of 8 per cent and recurrence proximal to the stoma in the ileum is 13 per cent over 10 years.

Total proctocolectomy

This is indicated for extensive colonic disease involving the rectum, with or without perianal disease. It is also indicated in patients with severe anorectal disease, even in the presence of an apparently normal upstream colon. Recurrence rates in the proximal colon and the problem of a liquid flush end colostomy are considerable. A permanent Brooke end ileostomy is therefore preferable (see Chapter 26.2.3.)

The incidence of stomal dysfunction is high. Complications associated with the stoma include retraction, prolapse, fistula formation, and obstruction. Most of these can be managed by local revision, but in the presence of recurrent Crohn's disease, laparotomy and resection is often necessary.

There is a high incidence of delayed perianal wound healing with resulting long-term morbidity and a persistent perianal sinus. Healing rates are slower if the wound is left open than is the case in patients treated with suture and suction drainage. Healing rates range from 63 per cent by 12 weeks to 33 per cent by 6 months. If there is faecal contamination and the risk of sepsis is high, the perineal wound should be left open. Perianal wounds which show delayed or non-healing are associated with high fistula *in ano*, faecal contamination, and postoperative perineal wound infection.

Management of the unhealed perineal wound

Around 10 per cent of perineal wounds fail to heal within a year. Management should be conservative initially, but if after a few months the perineal wound is recurrently discharging or has exuberant granulation tissue around the sinus it should be explored. Primary causes of long-term failure to heal include a foreign body such as a stitch, a pilonidal sinus, hydradenitis, an enteroperineal fistula, retained rectal mucosa, and malignancy. A large number of procedures have been suggested, such as curettage, suture and excision, or reconstruction with a musculocutaneous graft; all are associated with a variable success rate.

Other problems

A significant number of patients develop urinary and sexual dysfunction as a result of nerve injury and perineal fibrosis. The incidence can be reduced by a perimuscular excision of the rectum and an intersphincteric excision of the anus.

Proctocolectomy is associated with the lowest recurrence rate but it has an operative mortality of between 3 and 9 per cent. The postoperative stay in hospital is often long, while the patient learns to care for the stoma, and also due to delayed perineal wound healing. Recurrence rates following proctocolectomy are about half of those seen after restorative surgery (see Table 2). Recurrent disease is usually located in the distal ileum.

Proctectomy

In a small proportion of patients with Crohn's disease only the rectum is involved, often associated with severe perianal disease. If the disease cannot be controlled medically, a permanent end colostomy may be suitable for selected patients. Particular indications for this procedure include a high fistula or rectovaginal fistula, together with narrow anorectal strictures. Problems of perineal wound healing should be borne in mind, together with the difficulty in managing a liquid colostomy experienced by some patients.

External faecal diversion

Faecal diversion (loop ileostomy or split ileostomy), originally advocated by the Oxford group and now more widely practised, has been used as an alternative to conventional surgical management in the following circumstances.

1. To achieve colonic healing and allow intestinal continuity to be restored without resection where the disease is diffuse but not severe enough to warrant a proctocolectomy.
2. To facilitate major resection in those with poor health.
3. To limit resection in patients with diffuse disease.
4. To avoid growth retardation in children with diffuse colitis requiring colectomy.
5. To protect or avoid a primary anastomosis, and after small bowel resection in patients with persistent colonic disease.
6. In the management of refractory perianal disease as a means of delaying or even preventing proctocolectomy.

It is associated with a high incidence of early disease remission.

A limited number of patients are suitable for this form of management. The percentage of patients in whom intestinal continuity can be restored varies from over 60 per cent to less than 30 per cent, and relapse after restoration of continuity ranges from 28 to 60 per cent. Nevertheless, a proximal loop ileostomy is a relatively minor procedure and in the debilitated patient may buy time before definitive surgery is contemplated.

Perianal disease

Over 50 per cent of patients with Crohn's disease have anal lesions (Table 3); these are most common in those with rectal disease. Fissure *in ano* is the most common lesion and may be asymptomatic. Such a fissure can heal with medical management or may become chronic and the site of subsequent fistula formation, probably by distorting anal glands. These fissures are often painless, but if there is an associated submucosal or intersphincteric abscess, pain is so severe that proximal defunction is necessary (Fig. 3(a)). Anal surgery for fissure should be avoided as far as possible and 0.2 per cent glyceryl trinitrate paste has proved useful in healing chronic anal fissures including those due to Crohn's disease. A fistula *in ano* may develop either directly as a result of Crohn's disease or secondary to its effect upon the anal gland anatomy; fistulas are sometimes multiple and complex (Fig. 3(b)). Associated abscesses have to be drained but every attempt should be made to conserve any sphincter muscle. MRI, particularly when used with an anal coil, is useful at identifying obscure tracts and collections in complicated cases (Fig. 4). The key note to management is to be conservative.

[illegible]	[illegible]	[illegible]
Anal fissure	Skin tags	Piles
Undermined oedematous pile	Anal/rectal stricture	Perianal abscess/fistula
Crohn's ulcer	Anovaginal/rectovaginal fistula	Cryptitis
	Carcinoma	Hidradenitis suppurativa

Table 3 Perianal lesions in patients with Crohn's disease.



Fig. 3. Typical appearances of perianal Crohn's disease. (a) Deep perianal ulceration. (b) Perineal fistulas. (c) Large oedematous skin tags.

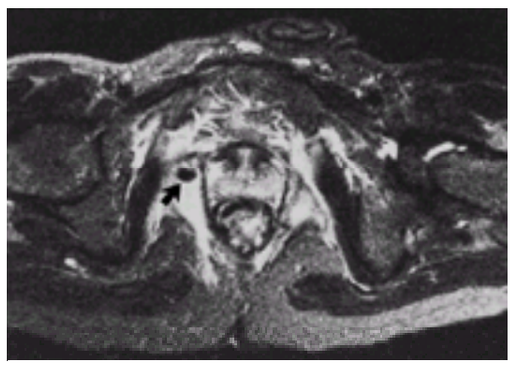


Fig. 4. MRI scan of perineal area showing gas (arrowed) within an abscess cavity extending into the ischiorectal fossa.

Medical management

Since many of these fistulas are asymptomatic or only intermittently symptomatic, treatment with metronidazole, steroids, and occasionally azathioprine can have dramatic effects on control of progress of disease, provided that no undrained pus is present.

Surgical management

Proper assessment is difficult in the clinic and symptomatic patients should undergo an examination under anaesthesia. Fistula tracts are carefully probed and curetted; if associated abscesses cannot be drained without division of muscle a seton can be left in place, often for many weeks or months, to control symptoms. A long-term indwelling seton functioning as a drain can often prevent or delay the need for proctectomy. We usually use a 2-mm diameter, coloured, Silastic vessel loop tied loosely rather than tightly (Fig. 5). If the disease is progressive or fails to respond to conservative management despite repeated examinations under anaesthesia, it is worth considering a proximal diversion (see above). Proctectomy is necessary for the treatment of severe disease unresponsive to conservative therapy. Bear in mind that long-standing fistula *in ano* may undergo malignant change in patients with Crohn's disease.



Fig. 5. Perianal disease managed with a seton.

Rectovaginal fistula

These may be quiescent, but can be very troublesome (Fig. 6). Asymptomatic patients need no treatment, and low fistulas can be laid open. A chronic indwelling seton can be used for drainage, as for perianal fistula, and where there is no severe rectal disease or proximal disease, repair of a fistula by a vaginal or rectal advancement flap can be successful. A temporary defunctioning ileostomy may be formed to protect the perineal operation. Patients with intractable disease and those developing incontinence require a proctectomy.

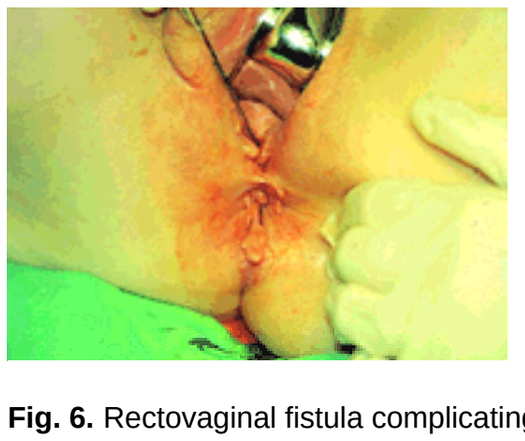


Fig. 6. Rectovaginal fistula complicating perianal Crohn's disease.

Haemorrhoids

Great care must be taken in treating haemorrhoids in patients with Crohn's disease. A haemorrhoidectomy should be avoided as far as possible; symptoms will generally settle spontaneously or following local steroid applications.

Further reading

Allan A, Andrews H, Hilton CJ, Keighley MRB, Allan RN, Alexander Williams J. Segmental colonic resection is an appropriate operation for short skip lesions due to Crohn's disease in the colon. *World Journal of Surgery* 1989; **13**: 611–16. [Outcome of 36 patients undergoing segmental resection for colonic Crohn's disease.]

Fazio VW, Wu JS. Surgical therapy for Crohn's disease of the colon and rectum. *Surgical Clinics of North America* 1997; **77**: 197–210. [Review article of the the management of colonic and rectal Crohn's disease.]

Phillips RKS. Ileal pouch–anal anastomosis for Crohn's disease. *Gut* 1998; **43**: 303–8. [Leading article on restorative proctocolectomy in Crohn's disease.]

Prabhakar LP, Laramee C, Nelson H, Dozois RR. Avoiding a stoma: role for segmental or abdominal colectomy in Crohn's colitis. *Diseases of the Colon and Rectum* 1997; **40**: 71–8. [Outcome of 49 patients with Crohn's disease undergoing colectomy or segmental resection.]

Williams JG, Wong WD, Rothenberger DA, Goldberg SM. Recurrence of Crohn's disease after resection. *British Journal of Surgery* 1991; **78**: 10–19. [Review article on recurrence after surgery for small and large bowel Crohn's disease.]

26.3 Colorectal tumours

Alfred M. Cohen

Epidemiology
Etiology and risk factors
Diet
Familial risk factors
Hereditary non-polyposis colonic cancer
Personal risk factors
Molecular biology of colorectal neoplasia
Familial adenomatous polyposis
Hereditary non-polyposis colon cancer
p53
DCC
ras
Prevention
Diet
Chemoprevention
Polypectomy
Surgical resection
Screening
Average-risk patients
Family history positive
Adenomatous polyps
Inflammatory bowel disease
Pathology and staging of colorectal neoplasia
Histopathologic grading and typing
Prognosis
Treatment of precancerous disease
Adenoma
Villous adenoma of the rectum
Familial adenomatous polyposis
Ulcerative colitis
Hereditary non-polyposis colon cancer family members
Clinical presentations
Cecal and right-sided carcinoma
Left-sided and sigmoid lesions
Rectal cancer
General issues in colorectal cancer surgery
Preoperative assessment
Barium enema and colonoscopy
Ultrasonography and computed tomography
Magnetic resonance imaging
Blood tumor markers
Bowel preparation
Antibiotics
Contiguous organ involvement
Management of obstruction or perforation
Perforation
Oophorectomy
Anastomotic techniques
Management of cancer in polyps
Laparoscopy
Surgical treatment of colon cancer
Specific preoperative assessment
Surgical resection
Adjuvant therapy of colon and rectosigmoid cancer
End-results with surgical resection alone
Portal-vein chemotherapy
Systemic chemotherapy
Immunotherapy
Recommendations
Surgical treatment of rectal cancer
Preoperative assessment
Local treatments
Radical resections
Adjuvant therapy of rectal cancer
End-results with surgical resection
Adjuvant therapy with local excision
Adjuvant therapy and radical surgery
Recommendations for adjuvant therapy
Follow-up
Surgical treatment of locally recurrent rectal cancer
Recurrence after limited local therapy
Suture-line recurrences
Recurrence after radical surgery
Surgical treatment of metastatic colorectal cancer
Ovarian metastases
Hepatic metastases
Pulmonary metastases
Benign colorectal tumors
Carcinoid tumors
Rectal sarcomas
Further reading

The past decade has produced dramatic increases in our understanding of the biology of colorectal cancer, as well as incremental improvements in its cure through the use of multimodality treatments. Screening strategies are leading to earlier diagnoses for many patients. Changes in lifestyle and the use of chemopreventative agents may lead to a marked reduction in incidence in the future. However, at present, cancer of the colon and rectum remains the second most common cancer in countries where a ‘Western' diet is eaten, and surgical resection remains the mainstay of curative treatment. It is a disease that affects men and women equally.

Advances in understanding of the molecular biology of colorectal cancer and the use of adjuvant chemotherapy and radiation do not obviate the need for high-quality surgical technique as the most important component of treatment. In this chapter we will define the appropriate preoperative evaluation of patients with colorectal cancer, as well as issues concerning the quality of surgical treatments—this is of considerable importance in rectal cancer: local treatment options such as local excision and radiation are playing an increasingly important part in the treatment of early cancers; in patients with more advanced rectal cancers, sharp mesorectal excision with preservation of the autonomic nerves maximizes cure, local control, sphincter preservation, and bladder and sexual function. Controversial issues concerning laparoscopic-assisted colectomy and prophylactic resection in patients with familial cancer syndromes will be discussed.

The treatment section of this chapter separates colon from rectal cancer. Surgical technique and issues of locoregional recurrence are of greater importance in the management of rectal cancer. Multimodality strategies for rectal cancer include radiation therapy, rarely required in the treatment of colon cancer. The medical literature varies in its definition of rectal cancer; the mobile rectosigmoid may be included in reports from some centers. The pivotal distinction is the peritoneal reflection: tumors with a lower edge below that reflection are at increased risk for locoregional recurrence, and are appropriately treated by sharp mesorectal excision

with selective adjuvant chemoradiation; tumors above the reflection, whether in the rectosigmoid or ‘intraperitoneal rectum' (10–15 cm above the anal verge) should receive adjuvant therapy based on the treatment paradigm of colon cancer.

Epidemiology

The incidence of large-bowel cancer varies greatly amongst countries, with considerable increases in populations associated with emigration from a low-risk to high-risk countries. These data suggest an environmental etiologic component, most commonly suggested to be dietary factors such as high fat and low fiber. This concept is supported by the observation that within countries, dramatic differences in colorectal cancer exist amongst various ethnic groups, generally correlated with dietary behavior. Table 1 and Table 2 demonstrate some of these data.

USA	35
Western Europe	40
South America	15
Asia	15
Africa	2

Table 1 Approximate annual incidence of colorectal cancer per 100 000 people

Race/heritage	*Males*	*Females*
Japanese	62	37
White American	60	42
Black	56	46
Chinese	49	33
New Mexico	10	9

Table 2 Approximate annual incidence of colorectal cancer per 100 000 people in the United States

In the absence of familial risk factors the peak incidence of colorectal cancer is in the seventh decade of life, general around the age of 65 years. Its incidence is about equal in the two sexes, although it is frequently perceived by the public as a cancer of men. Rectal cancer is more common in men, with women having a higher incidence of more proximal cancers. In Western societies there has been an increase in more proximal cancers, with fewer rectal cancers. No longer are half of large-bowel cancers within the reach of the rigid sigmoidoscope. However, in many parts of the world, the rectum remains the predominant site for large-bowel cancer.

Etiology and risk factors

Diet

Since the epidemiologic evidence suggests that, in the large majority of patients with large-bowel cancer, environmental factors predominate, we will defer discussion of familial risk factors to later in this section. Epidemiologic studies have explored not only dietary patterns, but also exercise, and the impact of aspirin and other putative chemopreventative agents. Dietary studies are complex, since in a disease of long latency, a lifestyle adopted over decades must be analysed retrospectively. Dietary factors include total calories, calories from fat, different types of fat, fiber, different types of fiber, and the many micronutrients in our diets. Fat and fiber are likely the most important factors. Table 3 delineates these issues.

Component	Effect	Data
Fat*	Linear correlation	Dietary fat by country/colorectal cancer incidence; migrant studies
Fiber	Inverse correlation	Fiber by country/colorectal cancer incidence
Calories	Direct relation	Total calories correlate with risk by country

*Data on fat are the most compelling.

Table 3 Diet and risk

Fat

Rates of colorectal cancer in various countries are strongly correlated with the presence of animal fats and meat in the diet. In countries such as Japan, in which traditional low-fat diets have been replaced with more Westernized diets, the incidence of colorectal cancer over the past 50 years has increased 2.5-fold. Fat intake in Westernized diets is currently 35 to 40 per cent of total calories. Case-control studies have explored the complex interrelations of fat intake and total caloric intake, specific fat types, and the general level of exercise. Data from such studies are less conclusive, although the regular eating of red meat appears to be the most important determinant of colorectal cancer risk.

Dietary fiber

Burkitt's report on the epidemiology of appendicitis, diverticular disease, gallstones, coronary arterial disease, and colorectal cancer in relation to dietary fiber intake has generated 25 years of epidemiologic and animal studies. Although a large body of epidemiologic literature supports the relation between low fiber intake and colorectal cancer, our lack of understanding of the various types of dietary fiber, both soluble and insoluble, makes such studies problematic.

Dietary fiber is a group of plant components that are generally resistant to human digestive enzymes. The amounts and type of such fiber vary with the type and age of plant, as well as among its various parts. Most cellulose- and lignin-containing fibers are insoluble and non-fermentable; they increase water absorption and reduce transit time, which will reduce mucosal exposure to intraluminal carcinogens. Pectins, gums, and many hemicellose fibers are water-soluble and fermentable; this may

prevent colorectal cancer by binding intraluminal carcinogens and by reducing the intraluminal pH (by production of short-chain fatty acids).

Micronutrients

Epidemiologic studies suggest that populations with abundant intake of fresh fruits and vegetables have a lower risk of colorectal cancer. In such foods are the well-known antioxidants b-carotene, selenium, and vitamins C and E. Growth inhibition may be associated with calcium and vitamins A and D. However, prospectively randomized, placebo-controlled studies have failed to show the benefit of such supplements in the prevention of colorectal cancer.

Familial risk factors

In addition to the obvious genetically related familial diseases such as familial adenomatous polyposis, there are considerable data demonstrating the relation between other familial cancer sites and colorectal cancer. In addition, a family history of colorectal polyps is associated with increased risk of colorectal cancer in other family members. However, at least 80 per cent of patients have ‘sporadic' colorectal cancer, with no familial risk factor identified.

Polyposis syndromes

Familial syndromes take the form of familial adenomatous polyposis syndromes and the familial cancer syndromes (or hereditary non-polyposis colonic cancer; Lynch syndromes). Familial adenomatous polyposis syndromes account for as little as 0.5 per cent of all colonic cancers but they are important as a biologic model for other cancers and for constructing algorithms of therapeutic management based on sound biologic principles (Table 4). Familial adenomatous polyposis is an inherited propensity to develop numerous adenomas throughout the colon, some of which in time become malignant. The condition is inherited as an autosomal dominant with high penetrance but variable expression, and approximately half the children of the index case inherit the condition. The polyps appear at around puberty, and nearly all patients manifest polyps by their early thirties. There are a few reports of polyps presenting later, but the chance of being affected if a sigmoidoscopy is normal at the age of 35 years is less than 1 per cent.

A. Familial adenomatous polyposis
1. Includes Gardener, Oldfield, and Turcot syndromes
2. All have related 5q gene mutations
3. Extracolonic lesions:
(a) Duodenal adenoma/cancer
(b) Desmoids
(c) Osteomas
(d) Intracranial tumors
(e) Sebaceous cysts
(f) Congenital hypertrophy of retinal pigment epithelium
B. Peutz–Jegher syndrome
1. Small-bowel and other gastrointestinal polyps/hamartomas
2. Perioral freckles
3. Increased risk of gastrointestinal cancer
C. Juvenile polyposis
1. Hamartomatous polyps
2. Increased risk of cancer
D. Cowden disease
1. Orocutaneous and gastrointestinal hamartomatous lesions
2. Increased risk of breast and gastrointestinal cancer
E. Muir–Torr syndrome
1. Sebaceous cysts
2. Increased colon cancer

Table 4 Polyposis syndromes

Recent work has shown that the genetic abnormality in familial adenomatous polyposis is a mutation on the long arm of chromosome 5. The *FAP* gene has been cloned on 5q. The absence of this gene or cluster of genes seems responsible not only for the colonic polyps but also neoplasms in other parts of the gastrointestinal tract, the mesenchyme, and occasionally the brain. There appears to be a general derangement of cell protein formation common to many tissues. Gardner syndrome, with osteomas, desmoids, gastric hamartomas, and a propensity for gastric, pancreatic, and small-bowel tumors, is the most severe form of the syndrome. Patients with the more common familial adenomatous polyposis seldom have clinically obvious extraintestinal manifestations. The least severe manifestation, Turcot syndrome, is characterized by fewer colonic polyps, but affected patients also have intracranial tumors, which are more lethal. While it is likely that Turcot cases are part of the spectrum of familial adenomatous polyposis and therefore exhibit the same DNA abnormality, it has yet to be proved. There is an ‘attenuated' form of familial adenomatous polyposis with large numbers of adenomas, but usually less than 100, found in middle-aged patients.

Peutz–Jegher syndrome and juvenile polyposis are not strictly part of the familial adenomatous polyposis syndromes but for convenience they may be considered with them. These conditions are rarer and the polyps fewer in number, but the trait is inherited in a dominant fashion. The polyps are hamartomas that often have adenomatous changes around them. This change predisposes to malignant transformation and increases the risk of colon cancer. It is likely that the genetic abnormality of these two conditions is different from that of familial adenomatous polyposis.

Hereditary non-polyposis colonic cancer

This is more common but less obvious than familial adenomatous polyposis. The inheritance is again autosomal dominant, but the appearance of the cancer is clinically similar to that of sporadic cases of colorectal cancer, although it occurs at a younger age (Table 5). Two subtypes of the syndrome have been recognized: site-specific colonic cancer, where individuals of a family are susceptible to col-onic cancer but not cancers of other organs (Lynch type 1), and cancer family syndromes where female members of the family are prone to breast and uterine cancers as well as colonic tumors (Lynch type 2). The tumors may all develop in the same individual but, more commonly, the three types of cancer appear in different members of the family. Occasionally, tumors of other organs are found (the Muir–Torre syndrome). The small bowel and bladder are at greatest risk. It is interesting that the cancers in patients with Lynch type 1 and 2 syndromes develop in early middle life rather than in the seventh and eighth decades as do sporadic cases. The cancers occur predominantly on the right side of the colon and are of low malignancy, again in contrast to the sporadic cases.

A. Lynch type I
1. Autosomal dominant
2. Proclivity for proximal colon
3. Younger age at onset
B. Lynch type II
1. Autosomal dominant
2. Proclivity for young age-onset proximal colon cancer
3. Increase cancer risk for cancer of:
(a) Endometrium
(b) Ovary
(c) Small bowel
(d) Urinary tract
C. 'Amsterdam criteria' (currently less stringent)
1. Three or more relatives with colorectal cancer, one of whom is a first-degree relative of other two
2. Colorectal cancer involving at least two generations
3. One or more cancers developing before age 50 years

Table 5 Hereditary non-polyposis colon cancer

Personal risk factors

The major personal risk factors for developing colorectal cancer are a prior history of adenomatous polyps, prior colorectal cancer, or the presence of inflammatory bowel disease. A personal history of endometrial, ovarian, or breast cancer may also produce some increase in relative risk, but these tumors are primarily associated with familial risk factors and germline genetic mutations.

History of adenomas

Pathologic, epidemiologic, and molecular genetic data support the concept that in almost all cases, invasive adenocarcinomas begin as adenomas. Carcinoma is rare in polyps of less than 1 cm. The larger the polyp, the greater the likelihood that microscopic carcinoma will be found. The greater the villous rather than tubular

component, and the larger the number of polyps, will influence the likelihood of subsequent development of colorectal cancer.

Prior colorectal cancer

Between 3 to 5 per cent of patients with sporadic colorectal cancer will develop a metachronous colorectal cancer. It is assumed that follow-up colonoscopy and polypectomy will minimize the risk of a subsequent cancer.

Inflammatory bowel disease

Both ulcerative and Crohn's colitis are associated with an increased risk for colorectal cancer. Ulcerative colitis involving the entire large bowel for at least a decade is associated with a 1 to 2 per cent per year incidence of cancer. The magnitude of the risk in Crohn's colitis has not been as clearly defined. The majority of those at risk for invasive cancer appear to have severe dysplasia seen in biopsies that antedate the development of cancer; this provides the opportunity for selective prevention, which will be discussed later.

Molecular biology of colorectal neoplasia

Since the findings that adenomas are precursors of colorectal cancer, the availability of tissues from normal mucosa, low-risk (tubular adenoma) and high-risk (villous adenoma) polyps, invasive cancers, and hepatic metastases has allowed the molecular genetic components of this disease to be extensively studied. Many data are available on multistep carcinogenesis. Combinations of germline and acquired genetic mutations and deletions are involved in many patients. Only the highlights of the genetic changes will be given here (Table 6): ‘p' refers to the short arm of a chromosome, and ‘q' the long arm; numbers that precede indicate the chromosome number and those that follow refer to chromosomal banding locations.

APC	Gene mutation
DNA	Mismatch repair gene mutations
p53	Deletion or mutation
DCC	Deletion
ras	Mutation

Table 6 Molecular changes in colorectal cancer

Familial adenomatous polyposis

Positional cloning techniques from affected family kindreds identified chromosomal site 5q21, which led to the identification and cloning of this gene (*APC*) in 1991. Many different mutations of this gene have been identified, and there may be some correlation between the specific sites of mutation and the familial phenotype. One-half to two-thirds of patients with sporadic colorectal cancer also have such mutations. Studies of adenomas indicate that this is one of the earliest genetic perturbations in colorectal neoplasia. Genetic testing is usually done on the gene-product protein, truncated in the presence of these mutations.

Hereditary non-polyposis colon cancer

Molecular studies on tumors in hereditary non-polyposis colon cancer showed that alterations in dinucleotide and trinucleotide microsatellite repeats (‘MSI' – microsatellite instability) are ubiquitous; these are referred to as ‘replication errors', or RER, and have been found in all hereditary non-polyposis colon cancers. This finding rapidly led to the realization that such abnormalities in yeast and bacteria were associated with defects in ‘DNA mismatch repair' genes. Germline mutations in human specimens were confirmed, with several mutations identified, the most common being *MLH1* and *MSH2*. Penetrance of some of these mutations may approach 80 per cent. This gene is required for the continuing repair of DNA fidelity following replication. Of note is that tumors in patients with hereditary non-polyposis colon cancer are general diploid, and have few of the additional deletions and mutations to be discussed below. Stage for stage, such tumors have a better prognosis than sporadic cancers, perhaps related to these limited genetic changes. The mechanism of carcinogenesis in such patients appears to be an alteration in the growth-inhibitory pathway of transforming growth factor-b.

p53

Deletion or mutation of this gene on 17p is a common finding in many human solid tumors, and a frequent earlier event in colorectal cancer transformation. Mutation or deletion is rare in adenoma. Evidence for this being a tumor-suppressor gene is compelling. Patients with a germline mutation (Li–Fraumeni syndrome) have a high risk of cancer; *p53* ‘knockout mice' develop multiple cancers. *In vitro*, the introduction of a ‘wild-type' (normal) *p53* gene converts transformed cells toward normal. The presence of this mutation is associated with a more malignant phenotype, even with comparable stages.

DCC

This portion of 18q is deleted in 70 to 80 per cent of colorectal cancers, half of large adenomas with dysplasia, and rarely in small tubular adenomas. *DCC* is an acronym for ‘deleted in colon cancer'. Evidence suggests that this is a tumor-suppressor gene, and deletion is also associated with a worse prognosis.

ras

This is a member of a family of ‘proto-oncogenes' that are involved with tyrosine kinase function. This mutated gene is present in 50 per cent of colorectal cancers and large dysplastic polyps, and only rarely in small, innocuous adenomas. This mutation appears to be important in maintaining the malignant phenotype. Of interest is that certain single-codon mutations result is a malignant phenotype seemingly incapable of metastasizing.

Prevention

Prevention strategies involve lifestyle issues, the use of putative chemopreventative agents, and the selective use of prophylactic operations. Based on epidemiologic data, a diet of relatively restricted total calories, with a total fat intake of less than 25 per cent of total calories, including high fiber and fresh fruits and vegetables, in association with moderate exercise will likely reduce the risk of developing colorectal cancer. Such a lifestyle will likely also minimize the risk of breast and prostate cancer and well as cardiovascular disease; hence there are no negative aspects to such a regimen. Colonoscopy and polypectomy will have a major impact on the incidence of invasive cancer.

Diet

Epidemiologic data suggest that diet influences the risk of developing colorectal cancer, but prospective, randomized or population-based trials are required to demonstrate efficacy. Most studies use the ‘adenoma population' as a surrogate for such decade-long studies. The subsequent formation of adenomas following endoscopic polypectomy allows trials to be completed over a 5- to 10-year time-frame. At present, there is a North American trial evaluating low fat and high fiber following endoscopic polypectomy: the goal was to reduce dietary fat to 20 per cent of calories; only 25 per cent fat has been accomplished. Many dietary trials also combine changes in macronutrients with added micronutrients.

Micronutrient trials

Studies are evaluating the antioxidants such as b-carotene, vitamins C and E, as well as calcium and w-3 fatty acids. A large double-blind trial of American physicians using b-carotene for 12 years was not effective. An Australian trial combining diet, reduced fat, and b-carotene suggested the risk of subsequent larger (higher-risk) adenomas was reduced. The Polyp Prevention Study Group in North America evaluated b-carotene, vitamin C and vitamin E, alone or in combination, without demonstrable effect.

Chemoprevention

Aspirin and other nonsteroidals are under study. Epidemiologic evidence of the efficacy of aspirin is compelling, yet in a randomized prospective trial of low-dose aspirin, prevention could not be demonstrated. Laboratory data suggest that inhibition of the cyclo-oxygenase II pathway is the most important aspect of tumor inhibition; such specific nonsteroidals (‘COX II' inhibitors) are available and in clinical trial as putative chemopreventative agents.

Polypectomy

Data that endoscopic polypectomy reduces the incidence of subsequent cancer by 50 to 90 per cent are highly compelling; based on the polyp–cancer association, this would seem to be intuitive, yet of course requires confirmation by clinical trial. The National Polyp Study Workgroup, analyzing 1418 patients following polypectomy with a median follow-up of 6 years, compared cancer incidence with a number of control cohort groups; their data suggest a 75 to 90 per cent cancer reduction.

Surgical resection

Patients at high risk because of germline mutations, such as the various familial cancers, or because of personal risk factors such as inflammatory bowel disease, may be offered prophylactic surgical resection, as discussed later.

Screening

From a public-health perspective, screening strategies based on risk factors in defined populations are selected. In an individual, such recommendations are modified, based on personal risk factors as well as personal preferences. Obtaining stool specimens for occult blood determination from large cohorts of patients is feasible through primary-care physicians, but screening colonoscopy requires a highly commited patient population. Both have considerable financial implications.

Numerous professional societies including the American Cancer Society, the American College of Gastroenterology, the American Gastroenterology Association, the American Society of Colon and Rectal Surgeons, and the American Society for Gastrointestinal Surgeons (amongst others) have endorsed a set of guidelines for colorectal screening. The reader is referred to the 1997 publication and associated references for details. In considering screening options, one should not analyze individual modalities (such as fecal occult blood), but consider options suitably based on the patient's personal and family risk factors. The following are general guidelines.

Average-risk patients

Screening should begin at the age of 50 years. Annual tests for fecal occult blood, flexible sigmoidoscopy every 5 years, and possible colonoscopy every 10 years are advised. If occult blood or sigmoidoscopy are positive, a more complete evaluation by colonoscopy is recommended. Air-contrast barium enema is satisfactory, but not ideal.

Family history positive

Hereditary non-polyposis colon cancer

Genetic counseling is important, and yearly or every 3 years' colonoscopy beginning in the third decade is appropriate, with yearly colonoscopy from the age of 35 to 40 years.

Familial polyposis

Genetic counseling, and genetic testing if a parent is positive; flexible sigmoidoscopy yearly from the age of 12 years.

Family history of cancer or polyps

Same recommendations as the average-risk population, except to begin at the age of 40 years.

Adenomatous polyps

Colonoscopy until the entire colon is cleared, then every 3 years thereafter.

Inflammatory bowel disease

Patients with ulcerative colitis and Crohn's are at risk. If prophylactic surgery is not performed, patients with pancolitis are advised to have screening colonoscopy with serial multiple biopsies every year for the assessment of dysplasia, beginning 8 years after diagnosis or symptom development.

Pathology and staging of colorectal neoplasia

Colorectal cancers, in common with most epithelial tumors, are polyclonal, with clones of cells exhibiting differing degrees of ‘malignancy'. The more undifferentiated or more ‘malignant' clones are more likely to spread and metastasize. A tumor's biologic behavior is the main determinant of the its propensity to spread locally and to metastasize, and therefore indicates the ultimate prognosis, which in turn is reflected to some extent by the histopathologic features. There is little doubt that, in the near future, other phenotypic characteristics, identified by molecular biologic methods, will provide even more accurate prognostic information. For example, tumor cells that stimulate angiogenesis or have fewer surface-adhesion molecules are more likely to metastasize and therefore to have a worse prognosis. Metastases, which are often derived from selected clones of a polyclonal primary, are unlikely to behave in the same way as the primary and will generally be more malignant.

Tumor biology is very much reflected by the stage of the disease, and therefore the amount of spread, at presentation. It is no surprise that there is a significant inverse relation between the length of history and the stage of the disease at diagnosis. Quicker-growing tumors usually present with a short history and are advanced at the time of treatment. Although the clinical and pathologic stage is a ‘snapshot' in the life of a tumor, it provides the most accurate prognostic index; this may be refined further by the histopathologic features.

Several staging methods are in use throughout the world, and each has strengths and weaknesses. The most commonly used are the Dukes' classification and its derivatives, and the American Joint Committee on Cancer (**AJCC**) and the Union Internationale Contre Cancer (**UICC**) TNM classification. Dukes' has the advantage of great simplicity but considerable disadvantages from lack of precision: it does not reflect accurately the depth of tumor penetration, the extent of spread outside the bowel, the number of lymph nodes affected by tumor, or the presence or absence of metastasis, all of which have an important bearing upon prognosis. Derivatives such as the Astler–Coller and Australian classifications refine the Dukes' staging but do not provide the flexibility of the TNM method, which enables useful division into subsets without being unduly complex. It is therefore most appropriate that surgeons adopt the AJCC/UICC TNM classification as a suitable international standard (Table 7(a),Table 7(b)). Fig. 1 presents common ‘T stage' examples.

Primary tumor (T)				
TX: Primary tumor cannot be assessed				
T0: No evidence of primary tumor				
Tis: Carcinoma in situ: intraepithelial or invasion of lamina propria*				
T1: Tumor invades submucosa				
T2: Tumor invades muscularis propria				
T3: Tumor invades through the muscularis propria into the subserosa, or into nonperitonealized pericolic or perirectal tissues				
T4: Tumor directly invades other organs or structures, and/or perforates visceral peritoneum**				
Regional lymph nodes (N)				
NX: Regional lymph nodes cannot be assessed				
N0: No regional lymph node metastasis				
N1: Metastasis in one to three regional lymph nodes				
N2: Metastasis in four or more regional lymph nodes				
Distant metastasis (M)				
M0: Distant metastasis cannot be assessed				
M0: No distant metastasis				
M1: Distant metastasis				

*Note: Tis includes cancer cells confined within the glandular basement membrane (intraepithelial) or lamina propria (intramucosal) with no extension through the muscularis mucosae into the submucosa.
**Note: Direct invasion in T4 includes invasion of other segments of the colon/rectum by way of the serosa, or into vagina/prostate in the rectum.

Table 7(a) AJCC/UICC TNM pathologic staging, 1997 version

AJCC/UICC				Dukes'
Stage 0	Tis	N0	M0	
Stage I	T1	N0	M0	A
	T2	N0	M0	
Stage II	T3	N0	M0	B*
	T4	N0	M0	
Stage III	Any T	N1	M0	C*
	Any T	N2	M0	
Stage IV	Any T	Any T	M1	D

*Dukes' B is a composite of better (T3N0M0) and worse (T4N0M0) prognostic groups, as is Dukes' C (any TN1M0 and any TN2M0).

Table 7(b) Stage grouping

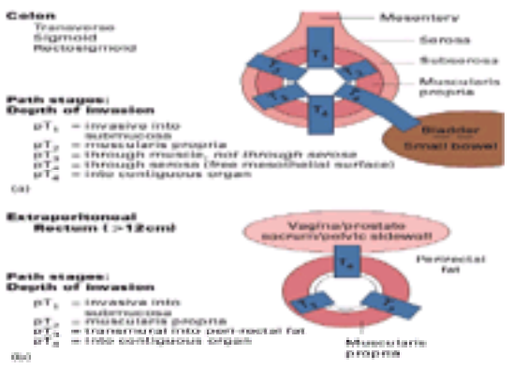


Fig. 1. Common ‘T stage' penetration in colon and rectal (no serosa) cancers.

The anatomic site of the cancer has an influence on the stage of the disease and an independent effect upon prognosis. Right-sided colonic cancers tend to be at a more advanced pathologic stage at the time of presentation, but their prognosis is generally no worse than that of left-sided tumors. Stage for stage, right-sided lesions have a better prognosis; this may reflect subtle etiologic differences. Rectal tumors have a generally worse prognosis despite earlier presentation and easier accessibility.

Staging gives information about prognosis in general, but particularly indicates the probability of occult hepatic metastasis, which is the major factor affecting survival: patients with Dukes' C tumors are more likely to have such metastases than those with Dukes' B tumors, while patients with Dukes' A lesions are most unlikely to have them. Occult hepatic metastases account for the majority of deaths from colonic cancer: only about 20 per cent of patients die from local spread, which is also reflected in the clinical stage. Age generally has little effect on the behavior or prognosis of colonic tumors, except for patients under 40 years old, who appear to have a particularly poor prognosis.

Histopathologic grading and typing

Grading depends upon subjective interpretation of the degree of tumor differentiation at histologic examination. Various grading systems have been proposed, but division into two broad groups, low- to average-grade tumors, which are well to moderately differentiated, and high grade or undifferentiated, reduces the variation between observers while at the same time providing useful prognostic information. Patients with high-grade cancers fare worse than those with well-differentiated lesions after taking account of the tumor stage. Typing, on the other hand, reflects the cellular characteristics. Mucinous, signet-cell, and small-cell tumors are variants of the more common adenocarcinoma and the frankly undifferentiated cancers. Again, typing may give some useful additional prognostic information. Signet-cell and small-cell tumors have a worse prognosis than adenocarcinoma, while mucinous lesions tend to recur locally. Occasionally, rectal cancers turn out to be squamous-cell types, which are more responsive to chemotherapy and irradiation. Melanomas, which have a particularly poor prognosis, are found rarely in the rectum. Both squamous tumors and melanomas are, however, more common at the anus. Histologic features such as vascular, lymphatic, or perineural invasion are prognostically unfavorable. In contrast, lymphocytic infiltration of the tumor and a histiocytic reaction in the regional lymph nodes are minor favorable prognostic features.

Identification of surface tumor antigens such as carcinoembryonic antigen, oncogene expression, and DNA ploidy adds potential refinement, but is not yet in routine use. Full characterization of the molecular biologic features of cancer cells will in the future probably provide more important prognostic and therapeutic information than histologic typing and grading. It may appear academic at present to stage patients and their tumors, but it will become important clinically as the place of adjuvant chemotherapy and irradiation becomes clearer, particularly when chemotherapy improves. In the mean time, it is important to collect pathologic data accurately to allow clinical studies and audit of surgical performance; the latter cannot be judged in the absence of good histopathologic information. Figure 2, Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 show some of the pathologic features described above.

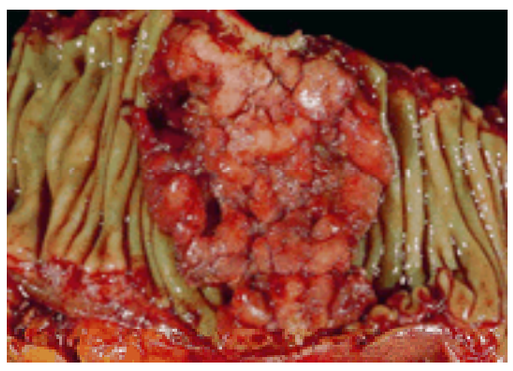


Fig. 2. Gross specimen of a resected rectal cancer.

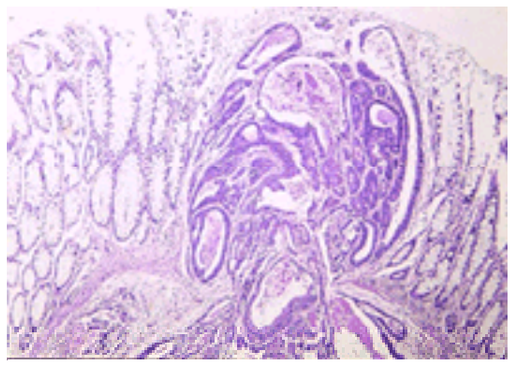


Fig. 3. Histopathology of focal cancer within a tubulovillous adenoma.

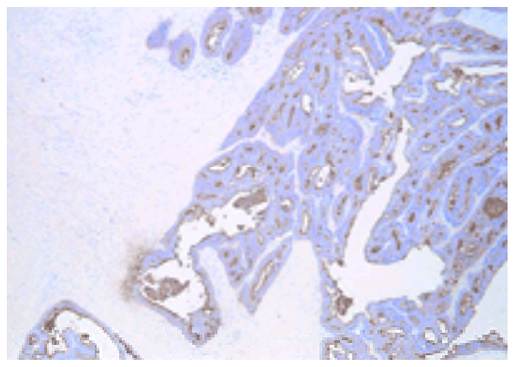


Fig. 4. A well-differentiated adenocarcinoma invading muscularis propria; immunoperoxidase stain for carcinoembryonic antigen is brown.

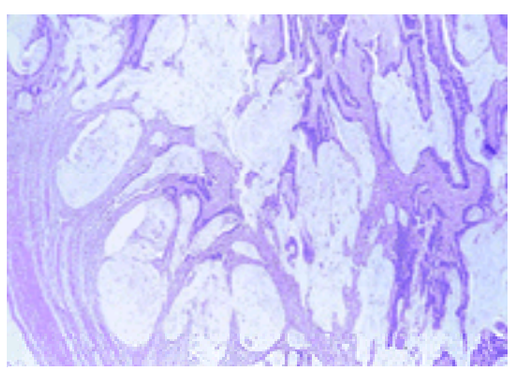


Fig. 5. Mucinous adenocarcinoma with a colloid (extracellular) pattern.

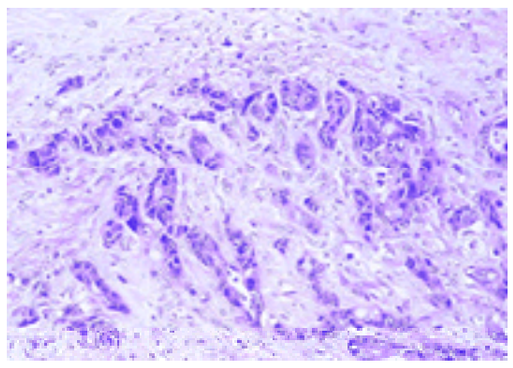


Fig. 6. Poorly differentiated adenocarcinoma.

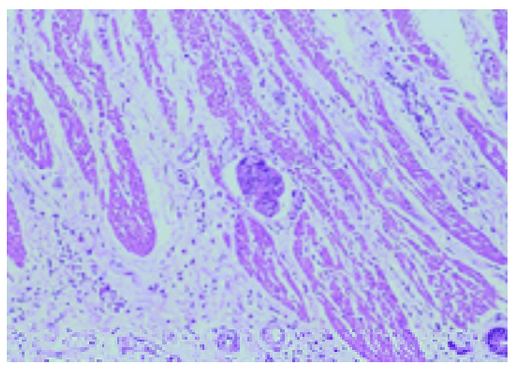


Fig. 7. Lymphatic vessel invasion.

Prognosis

Stage remains the most important indicator of prognosis (Table 8). The prognosis of patients with adequately treated, stage 1 cancers is little different from that of an otherwise healthy population of the same age; 95 to 100 per cent live 5 years or more after resection. Patients whose cancer has spread through the serosa only have a 40 to 60 per cent chance of living 5 years, although the prognosis is more favorable if the tumor is only just through the serosa and is correspondingly worse if adjacent structures are invaded. Lymph-node metastasis further adversely affects prognosis, with only about 30 to 40 per cent of patients surviving 5 years. Subclassification is useful: 60 per cent of N1 (1–3 positive nodes) patients may live 5 years.

TNM stage		Percentage 5-year survival		
	Rectal cancer	Colon cancer		
T1	N0	All	>95	>95
T2	N0	All	90	100
T3	N0	All	85	90
T4	N0	All	70	70
any T	N1	All	55	65
	N2	All	30	40
	any N	All	<10	<10

Table 8 Five-year survival by stage

Treatment of precancerous disease

Adenoma

The development of flexible fiberoptic endoscopy revolutionized the treatment of patients with adenomatous polyps. Prior to this advance, laparotomy with colotomy and polypectomy was required, with multiple operations over the decades for many patients. Currently, the vast majority of polypoid lesions are amenable to excision by endoscopic snare, which serves as an excisional biopsy. The risks of polypectomy are bleeding and perforation. Perforation may be related to the instrument (pushing ahead blindly), or more commonly to the cautery. Perforations may present several hours to a day after polypectomy, when the full-thickness cautery burn injury perforates. The incidence of perforation after endoscopic polypectomy is less than 1 per cent. Not all patients with a perforation require laparotomy for closure. Those who have had scrupulous mechanical bowel preparation and then have a cautery-related perforation may be treated with antibiotics and clinical observation. Laparotomy is only essential for instrument-related perforations and those with progressive symptoms and signs of diffuse peritonitis. When laparotomy is performed, the pathologic status of the polypectomy specimen should be obtained in order to consider resection, if appropriate, based on the pathology and the operative extent of the inflammation.

Almost all pedunculated lesions are removed with cautery/snare. Many sessile lesions can be excised piecemeal, sometimes over two or three sessions. Injection of saline into the submucosa to elevate sessile lesions from the muscularis propria is a useful technique. Argon-beam coagulation of bleeding from a flat polyp bed can also help. However, sessile lesions not amenable to such approaches, even if benign on biopsy, may require a laparotomy or laparoscopic-assisted resection. India-ink tattooing of the site avoids the difficulty of intraoperative detection of the exact location of these non-palpable polyps. Intraoperative colonoscopy is another approach.

Villous adenoma of the rectum

Management of these is frequently a vexing problem. Excision of a large, soft, villous adenoma without areas of induration (excluding biopsy sites) demonstrates cancer in 25 per cent, but two-thirds of that group have only *in situ* or intramucosal cancer, not requiring additional treatment. Only 10 per cent of all patients have invasive cancer in a clinically benign, large adenoma. Hence, the initial treatment strategy is complete excisional biopsy at the level of the submucosa. Small lesions are amenable to transanal excisional biopsy by using submucosal saline/epinephrine (adrenaline) injection to facilitate the resection at the appropriate level. Many large lesions, including circumferential adenomas, are also excised in this manner, with prolapse and mobilization of upper rectal mucosa. Plication of the muscular tube may facilitate reconstruction. A massive adenoma filling the rectum, even although entirely benign, may require abdominoperineal resection or total proctectomy and coloanal reconstruction. Laser or radiation therapy should not be used.

Familial adenomatous polyposis

In families with unambiguous familial adenomatous polyposis and visible polyps in children and young adults, colorectal cancer is almost inevitable by the age of 40 years. Such families should have their affected offspring evaluated for prophylactic surgery as teenagers or young adults. Unless bleeding is bothersome, surgery is appropriate between the ages of 18 to 25 years. The two resective options are abdominal colectomy with ileorectal anastomosis or total proctocolectomy with ileal-pouch anal reconstruction. Patients suitable for rectal preservation should have no or only a few polyps in the rectum; in addition, they must be committed to regular endoscopy every 6 months and realize that they may still develop rectal cancer despite surveillance and fulguration of polyps. With improving results for restorative proctocolectomy and little long-term ‘pouchitis' in patients with familial adenomatous polyposis, total proctocolectomy with pouch reconstruction is frequently recommended. For young women anxious to maximize their ability to become pregnant, the ileorectal anastomosis is still preferable. In the absence of extensive bleeding, large polyps, or dysplasia, elective surgery may be deferred to a convenient time for the patient.

It is important to appreciate that total proctocolectomy does not ‘cure' the problem. Patients with familial adenomatous polyposis saved from fatal colorectal cancer frequently succumb to abdominal desmoids or adenocarcinoma of the duodenum, ampulla of Vater, or bile ducts.

Ulcerative colitis

Patients with pancolitis of 8 years' duration have a subsequent 1 to 2 per cent per year risk of developing colorectal cancer. Prophylactic proctocolectomy is recommended in those with severe dysplasia on screening colonoscopy or who present with invasive cancer. The activity of the colitis is not a reliable indicator of risk. Duration and extent are the most valuable predictors. A subtotal colectomy and ileorectal anastomosis is not adequate surgical prophylaxis in such cases. If a stapled-pouch anal reconstruction is performed, the anal canal still requires surveillance.

Hereditary non-polyposis colon cancer family members

The management of colon cancer in hereditary non-polyposis individuals will be discussed in a subsequent section. There are family members of such patients with germline mutations who have a lifetime risk of over 80 per cent for developing colorectal cancer. As described previously, the current recommendation is their frequent endoscopic screening. However, prophylactic resections may be considered as our understanding of the various phenotypes and genetic penetrance in hereditary non-polyposis colon cancer becomes clearer. Highly selected patients may wish to consider abdominal colectomy and ileorectal anastomosis as well as hysterectomy in women.

Clinical presentations

Patients with colonic and rectal cancer have a broad range of clinical presentations that can be subclassified according to the anatomic site of the primary. Cecal and right-sided tumors account for about 20 per cent of large-bowel cancers, 70 per cent occur distal to the splenic flexure, and about 45 per cent are at or below the rectosigmoid junction. Right-sided cancers are becoming more common, particularly in women, and generally have rather ‘silent' symptoms.

Cecal and right-sided carcinoma

As mentioned, right-sided tumors are often remarkably silent and many patients present with only the symptoms and signs of iron-deficiency anemia from protracted occult blood loss. Rarely, the blood loss is copious, particularly in patients who are receiving anticoagulants. The feces entering the cecum are liquid and obstruction is a relatively late presentation. As the lumen becomes narrowed the patient complains of intermittent colic, centrally or in the right iliac fossa, which is often postprandial, stimulated by the gastrocolic reflex. The pain is often followed by the onset of intermittent diarrhea, possibly the consequence of fecal fermentation and the accumulation of bacterial toxins within the bowel lumen. Typical distal ileal obstruction occurs if the tumor blocks the ileocecal valve, or if the ileocecal valve becomes incompetent in the face of complete cecal obstruction. Waves of central abdominal colic occur, with progressive central abdominal distension and borborygmus. Visible peristalsis, feculent vomit, and dehydration are late manifestations. Not infrequently a palpable mass is the presenting symptom.

Patients occasionally present with symptoms and signs of acute appendicitis when the carcinoma occludes the appendicular orifice and produces acute inflammation, or from a perforated carcinoma. The diagnosis may not be obvious at the time the appendix is removed and must be sought subsequently by barium enema or preferably colonoscopy. The tumor may penetrate the bowel wall posteriorly, producing a sealed perforation and an abscess in the psoas muscle. Such patients present with the symptoms and signs of infection accompanied by a painful mass in the right iliac fossa. The pain may radiate down into the leg or hip, particularly if the

femoral nerve is involved in the abscess, or if the abscess tracks down below the inguinal ligament to appear in the femoral triangle. Similarly, the pain may radiate posteriorly if the abscess erodes into the lumbar muscles. Occasionally an anterior tumor may cause a free perforation producing acute peritonitis with severe generalized abdominal pain, a silent abdomen, and rebound tenderness on percussion.

Occasionally, right-sided colon cancers present with general symptoms of malaise and lack of well-being, sometimes with a pyrexia of unknown origin. These symptoms are either the result of a small occult abscess or are due to tumor burden, usually from metastases. The symptoms and signs of metastases are legion, but are usually accompanied by pain and tenderness over the liver, which is the most common site of metastasis. These symptoms are usually produced by rapid growth of the metastases distending the liver capsule. The metastases may also outgrow their blood supply, partially infarct and undergo necrosis. Secondary hemorrhage into the necrotic metastasis may then occur. Fever related to tumor necrosis is usually associated with an elevated serum lactic dehydrogenase.

Left-sided and sigmoid lesions

The stool dehydrates and becomes harder as it reaches and passes through the left colon to be stored in the rectosigmoid before defecation. Patients with a left-sided colonic cancer commonly present with a change in bowel habit, often constipation alternating with diarrhea, usually accompanied by lower abdominal colic, possibly distension, and a desire to defecate. The symptoms tend to become progressively severe, and this may serve to distinguish cancer from diverticular disease or colonic irritability. The irritable colon syndrome is usually seen in younger adults; if a middle-aged or older patient presents with a change in bowel habit the symptoms should be assumed to be caused by a colon cancer until proved otherwise. Progressive constipation or diarrhea are less common changes of bowel habit, either of which may end in complete obstruction of the large bowel.

Change in bowel function is often accompanied by the passage of altered blood, and sometimes mucus, in the stool or on its surface, particularly in the case of distal sigmoid lesions. Occasionally, patients present with colonic bleeding as an isolated symptom. The loss is usually intermittent, with small amounts of dark blood, but may be brisk, a symptom more usually associated with diverticular disease. Brisk bleeding from a colonic cancer is more likely to occur in a patient treated with anticoagulants.

A few patients present with a pain or mass in the left iliac fossa, but a mass is often palpable in the abdomen on physical examination. A palpable carcinoma of the splenic flexure must be distinguished from an enlarged spleen or kidney.

Some patients, surprisingly, have remarkably few symptoms until they present with massive abdominal distension due to complete obstruction of the large bowel. In these circumstances the cecum becomes very distended. Unless distension is recognized and treated rapidly, or unless the ileocecal valve becomes incompetent, cecal perforation leads to fecal peritonitis. If the ileocecal valve is incompetent, the obstructed large bowel decompresses into the ileum to produce a mixed clinical picture of large- and small-bowel obstruction. Occasionally the tumor itself will perforate, causing sudden acute abdominal pain and peritonitis. More commonly the tumor becomes attached to adjacent organs and may invade them. A sigmoid cancer may invade the lateral abdominal wall and form an abscess, or invade a loop of small bowel and produce either an ileocolic fistula with severe diarrhea or obstruction of the small bowel. Neoplasms of the splenic flexure or descending colon invading the jejunum sometimes present with severe intestinal hemorrhage. Sigmoid cancers commonly invade either the uterus, ovaries, or bladder. Colonic cancer is the second most common cause of colovesical fistulas after diverticular disease, and patients usually present with hematuria and recurrent urinary-tract infections initially, and later with pneumaturia or fecaluria. Similarly, a sigmoid cancer fixed in the pelvis may fistulize into the vagina to produce a malodorous, irritating discharge.

Rectal cancer

The accessibility of rectal lesions should enable the diagnosis to be made at an early and favorable stage. However, the diagnosis is often needlessly delayed while the symptoms are attributed by the patient or by the physician to hemorrhoids or an anal fissure.

Most patients with rectal cancer present with bleeding and/or a change in bowel habits. While the blood is often dark and mixed with the stool or coating the surface, it may be bright and separate from the feces. For this reason the symptoms are often attributed to hemorrhoids. Minor changes in bowel habit, such as increased frequency of defecation, mucus with the stool, or mucous diarrhea are also quite common. Mucous diarrhea is particularly associated with large villous adenomas that have often become malignant. The mucus is rich in potassium and may be sufficiently profuse to produce dehydration and coma. Tenesmus, the continuous urge to defecate, is a grave symptom produced by an advanced rectal tumor inducing a permanent sense of fullness. Tenesmus may give way to continual sacral pain, sometimes radiating down the perineal and thighs, as the tumor invades the sacrum and the sacral nerve plexus. Anal pain, initially on defecation and later continuous, may occur as a low rectal cancer invades the anal canal. Incontinence supervenes when the anal sphincters are destroyed. Proctalgia fugax (fleeting perineal pain) is a rare presenting symptom and therefore when it occurs *de novo* in later life, rectal cancer must be considered as a possible diagnosis. Few patients present with disseminated disease, though this may occur in younger people with rapidly progressive tumors or in those whose rectal symptoms are ignored for a long time. Similarly, obstruction of the large bowel is a rare and late mode of presentation of rectal cancer. Bright-red blood on the toilet tissue only may be evaluated in a young patient by proctosigmoidoscopy; all other types of bleeding warrant a more complete evaluation.

General issues in colorectal cancer surgery

Preoperative assessment

The importance of a good history and a careful physical examination cannot be overstressed. The completion of the physical examination should always be a digital rectal examination to feel for a mass, to assess its mobility and position, and to detect enlarged, extrarectal lymph nodes. A careful pelvic examination in women and prostate assessment in men are essential. For patients with rectal cancer, a proctosigmoidoscopic examination of the rectum and anus should follow.

Both the rigid and flexible instruments can be used easily in the left lateral position, most popular in Great Britain, or in the prone jack-knife position using a special tilting table, which is the custom in the United States. The distance of the tumor from the anal verge, and its position in relation to the anterior/posterior/lateral, are best determined with a rigid ‘scope. The configuration, size, and extent of circumferential involvement are noted. Mobility and tethering to surrounding structures are ascertained. Rigid sigmoidoscopy requires some care to avoid rectal perforation, but needs less experience than the fiberoptic method to obtain useful diagnostic information. Both instruments must only be advanced when the lumen is seen clearly. The investigation should be painless or at worst produce only mild discomfort. Blind intubation is a recipe for disaster ranging from pain to perforation. It is possible to employ either method of sigmoidoscopy selectively to use the patient's and clinician's time most efficiently. When a patient presents with symptoms and signs of rectal pathology, an unprepared rigid proctosigmoidoscopy is preferable. If, however, the history suggests that the problem lies more proximally, flexible sigmoidoscopy is more practical.

Barium enema and colonoscopy

There is much futile debate about the relative merits of barium enema and colonoscopy as the better method of investigating the remainder of the colon, for each provides different but complementary information. The barium enema gives good anatomic and topographic information that not only may be sufficient to diagnose a polyp or carcinoma but also demonstrates the site and configuration of the lesion, and the presence or absence of diverticulosis. The anatomic position of a cancer is clearly of great importance to a surgeon planning an operation. Discrimination for small lesions and mucosal abnormalities is considerably enhanced by the double-contrast technique rather than the single-contrast enema. The air insufflated after the barium shows clearly the mucosal destruction from an ulcerated carcinoma or the mucosal coating of both adenomatous polyps and polypoid carcinomas (Fig. 8, Fig. 9, Fig. 10). Only double-contrast assessments are appropriate in the assessment of the cancer patient. In the presence of partial obstruction, a soluble-contrast evaluation with gastrografin may be more suitable.

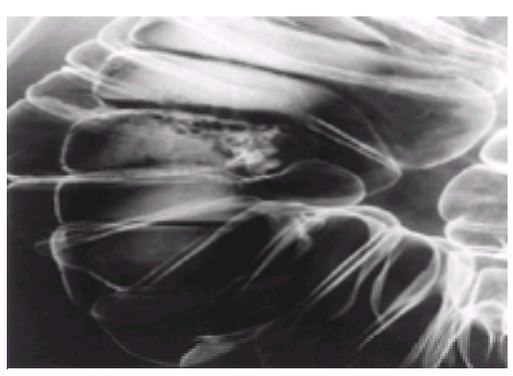


Fig. 8. High-quality double (air)-contrast barium enema demonstrating a small villous adenoma.

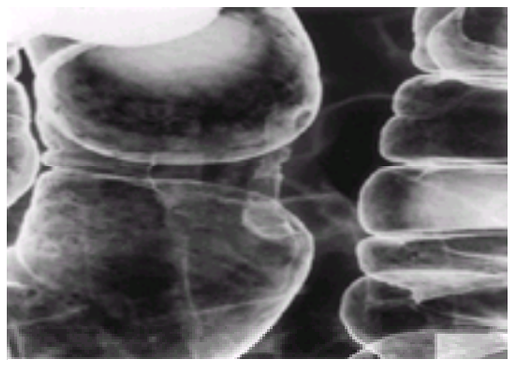


Fig. 9. Air-contrast barium enema with small colon cancer.

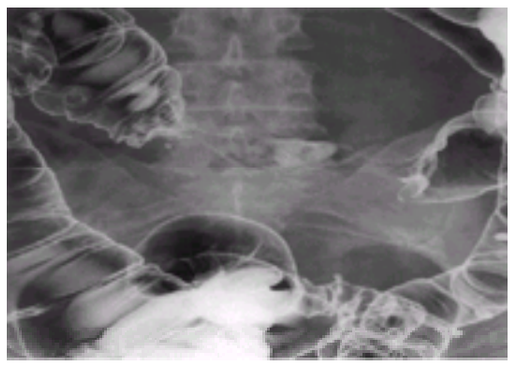
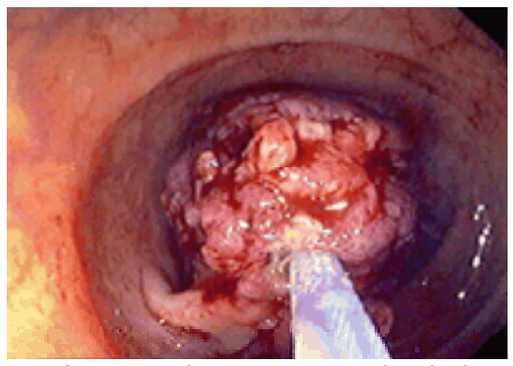


Fig. 10. Air-contrast barium enema with ‘apple core' appearance of a transverse colon cancer.

Colonoscopy requires the same meticulous bowel preparation as a double-contrast barium enema, but the patient also needs some sedation. It is therefore more invasive, and the patient needs both time to recover and somebody to take them home. There is also a small risk, about 1:1500, of colonic perforation during endoscopy. Nevertheless, colonoscopy enables a more detailed study of the mucosa, visualizing lesions of less than 0.5 cm. The principal advantage over radiology is that lesions can be biopsied, or removed by snare cautery if they prove to be adenomatous polyps or a small polypoid carcinoma (Fig. 11, Fig. 12). As a general rule, all patients will colorectal cancer should have a preoperative colonoscopy or as least an air-contrast barium enema. In patients with acute symptoms of obstruction an early postoperative colonoscopy with be needed.



P>**Fig. 11.** Adenomatous polyp being removed by colonoscopic snare cautery.

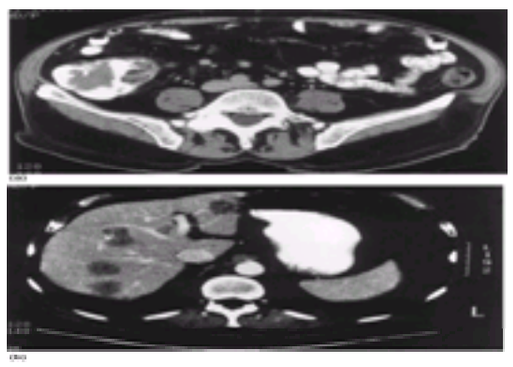


Fig. 12. Abdominal computed tomogram revealing cecal cancer (a) with synchronous hepatic metastases (b).

Ultrasonography and computed tomography

Once the diagnosis of carcinoma of the colon or rectum has been made, additional investigations are necessary to plan and organize treatment. For patients with rectal cancer, preoperative detection of metastasis is clearly crucial to allow treatment to be modified. The need for such tests in patients with colon cancer is less clear. Tests should be efficient and yield maximal information at minimal cost. The liver is the most common site for metastasis, followed by lung, retroperitoneum, ovary, peritoneal cavity, and, rarely, adrenal glands; abdominal ultrasonograms and a plain chest radiographs are both useful and economic. However, ultrasonography is very operator dependent and tends to vary in quality. The hard copy of the study is not as helpful as the real-time examination, and so review is not particularly useful. Abdominal computed tomography (**CT**) provides better discrimination for smaller hepatic lesions and can distinguish angiomas from metastases as small as 1.0 cm in diameter after contrast enhancement. CT also provides more information about lymphadenopathy and invasion of surrounding structures but at greater cost (Fig. 12, Fig. 13). Either investigation is excellent for detecting small volumes of ascites, ovarian enlargement, and hydronephrosis from retroperitoneal spread. CT- or ultrasound-guided biopsy or fine-needle aspiration cytology may be particularly useful in confirming or excluding the presence of metastasis. Since intraoperative management is occasionally altered by CT results, and a ‘baseline' CT of the liver is helpful, the routine use of CT in all patients with colorectal cancer is encouraged, excluding only the very elderly and those with very early cancers, such as within a polyp or rectal T1.



Fig. 13. Pelvic CT scan indicating a rectal cancer with suspicion of pelvic nodal metastases.

Endorectal ultrasonography

Small endoluminal ultrasound instruments have recently added refinement to the diagnosis and preoperative staging of tumors, particularly of the rectum (Fig. 14, Fig. 15, Fig. 16). Ultrasound probes mounted on a colonoscope are expensive and seldom used, but rigid endorectal ultrasonography is now an established means of assessing the depth of penetration of the bowel wall by the tumor. Enlarged lymph nodes can also be identified, but confirmation of nodal metastasis is problematic. This improved diagnostic information is essential when considering local treatment for a rectal cancer, and when choosing between an abdominoperineal excision of the rectum or a low anterior resection as well as for selective use of preoperative radiation therapy.

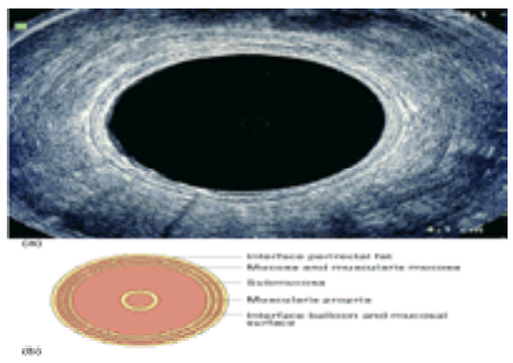


Fig. 14. (a) and (b) Normal rectal anatomy on endorectal ultrasound.

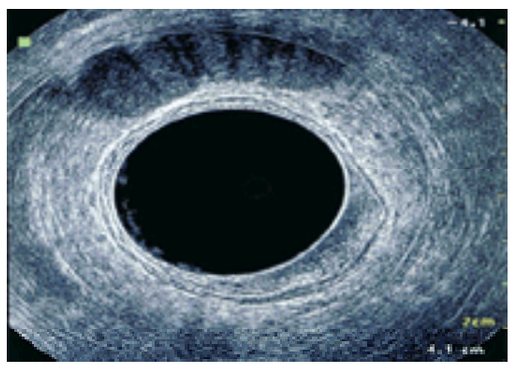


Fig. 15. Endorectal ultrasound of an early cancer limited to the submucosa (T1).

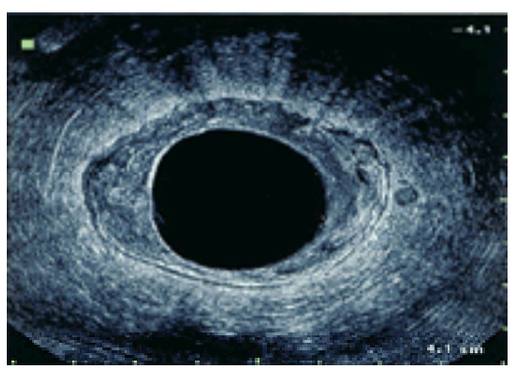


Fig. 16. A transmural (T3N1) rectal cancer on endorectal ultrasound (it is important to remember that there is no serosa in the lower two-thirds of the rectum).

Magnetic resonance imaging

Magnetic resonance imaging (**MRI**) is the most sensitive method of evaluating the liver, but is generally used only prior to planned major hepatectomy. Pelvic MRI with enhanced endorectal coils is a complex approach and generally not superior to endorectal ultrasonography. The main use of pelvic MRI is to assess the local extent of pelvic recurrence before embarking on salvage surgery.

Blood tumor markers

Recent studies suggest a baseline measure of carcinoembryonic antigen is helpful. In node-negative colon cancer, additional prognostic data provided by the preoperative carcinoembryonic antigen is of value. Patients with a preoperative carcinoembryonic antigen of greater than 10 have a 20 per cent reduced survival, and may benefit from adjuvant regimens normally limited to node-positive patients.

Bowel preparation

Mechanical bowel preparation combined with oral antibiotics is appropriate, except for patients with clinical obstruction. Patients with partial obstruction will usually tolerate cathartics. Oral phosphasoda- or polyethylene glycol-based cathartics are both effective. Enemas are rarely necessary.

Antibiotics

Oral antibiotics the afternoon and evening before surgery are effective in reducing the risk of superficial wound infection. Aerobic and anaerobic coverage is needed, with neomycin/metronidazole or neomycin/erythromycin being common combinations. Oral antibiotics should be supplemented with intravenous antibiotics given prior to the skin incision, usually 30 min before. A single dose of a long-acting aerobic/anaerobic agent such as cefotetan suffices to minimize the risk of postoperative, superficial wound infections in clean-contaminated colorectal cases. In patients requiring cardiac and bowel prophylaxis, combination therapies such as ampicillin, gentamicin, and either metronidazole or clindamycin are effective. A recent study suggests that maintenance of intraoperative normothermia may play an important part in minimizing the risk of wound infection.

Contiguous organ involvement

Both colon and rectal cancers may invade contiguous organs. These large T4 lesions are frequently node-negative with a high cure rate. Contiguously involved organs should be resected in continuity with the tumor if their are no other indications of distant spread. Half of such cases will demonstrate inflammatory adhesions rather than malignant penetration, but violating the tumor to obtained frozen sections before committing to the additional resection will likely lead to peritoneal seeding. In colon cancer, small bowel is the most common additional organ involved by adhesion. The dome of the bladder, or involvement of the female pelvic organs, stomach, and tail of the pancreas all may be resected easily. Duodenal invasion is more problematic, since a pancreaticoduodenectomy in the presence of a small bile and pancreatic duct has considerable morbidity. If feasible, a more limited duodenectomy with a gastrojejunostomy should be performed. Contiguous spread of a rectal cancer to the posterior vagina is amenable to rectal resection with posterior vaginectomy. A complete hysterectomy is not always necessary, and the retained uterus may be helpful in excluding the small bowel from the pelvis. If a vaginectomy is done in association with a sphincter-preserving rectal resection, a rectus abdominus flap, with or without the overlying skin, is useful in the absence of a large omental flap. Direct invasion of the prostate in men generally requires an exenterative procedure, since a rectal resection with a partial or even complete prostatectomy and bladder sparing is rather difficult with high morbidity. Adherence to Denonvilliers' fascia is best managed with preoperative radiation therapy rather than exenteration.

Management of obstruction or perforation

Management of obstructing tumors of the right or transverse colon is usually straightforward. In almost all cases, a primary resection is appropriate. The decision to proceed with a primary anastomosis is made by the operating surgeon, based on the overall condition of the patient and the quality of the bowel.

Left colonic and rectal obstruction presents more of a problem and surgical strategies are still evolving. It used to be considered that emergency left colonic resection in obstructed, unprepared bowel carried too high a mortality rate and that the obstruction was best relieved by a cecostomy or transverse-loop colostomy. The patient was then restored to better health, the bowel prepared by wash-outs, and the left colon resected and anastomosed electively 2 or 3 weeks later. The colostomy was sometimes closed at the same time as the resection but more usually later when the patient had recovered fully from the resection, so completing a three-stage operation. The cumulative mortality and morbidity of the three stages is high, particularly in elderly patients. The length of hospital stay is also long, which substantially increases the costs of treatment. Multistage resections are now seldom performed, but an emergency colostomy still has a place if the patient is ill. If at all possible, diverting colostomy should allow palpation of the liver to rule out metastases, and visualization of the cecum to ensure the absence of incipient perforation. Hence a small, upper abdominal, midline incision is ideal, which facilitates the exploration as well as the location of the colostomy.

Over the last 20 years, accumulated evidence supports definitive, single-stage treatment for the obstructed left colon. The present choice is between an extended or subtotal colectomy with ileosigmoid or ileorectal anastomosis, removing all the obstructed colon, or a more limited colonic resection, with on-table colonic lavage to decompress and clean the obstructed colon, and a colocolic, or colorectal, anastomosis. There are no convincing comparative data to decide the issue. Early evidence suggested that limited colonic resection and anastomosis was liable to leak because the bowel was unprepared, there was great disparity in size between the ends to be joined, and the vascular supply to the obstructed colon was suspect; therefore, resection of all the obstructed colon became fashionable. Mortality and morbidity were certainly lower but at the expense of poorer long-term bowel function. With on-table colonic lavage, systemic antibiotics, and better suture materials, anastomotic techniques and postoperative support, the single-stage procedure is now generally applicable.

An intermediate approach is the two-stage Hartmann procedure. The obstructing mass (tumor or diverticulitis) is resected, with a proximal end-colostomy and either a distal stoma (mucous fistula) or a closed distal bowel left within the abdomen. The ‘colostomy closure' is not a minor operation, and a one-stage approach, even with a diverting proximal-loop diversion, is generally preferable. The Hartmann approach is best reserved for palliation when pelvic peritoneal metastases would make an anastomosis hazardous or risk recurrent obstruction.

Perforation

Perforation is less common than is obstruction, occurring in about 5 per cent of patients. The site of perforation is usually within the tumor and is not associated with obstruction but is the consequence of tumor necrosis. Occasionally, left-sided obstructive cancers will perforate the cecum. The patient presents with severe, often sudden, abdominal pain and signs of peritonitis. Rapid cardiovascular collapse and endotoxemic shock usually signify a major leak and fecal peritonitis rather than a small perforation with purulent peritonitis. When the perforation occurs through the tumor, malignant cells are disseminated throughout the peritoneum, making cure less likely. Good palliation is therefore the paramount aim and long-term survival is an unanticipated bonus. The patient requires pain relief, rapid fluid and colloid resuscitation, and systemic antibiotic therapy before moving to the operating theatre. Delay in diagnosis and definitive treatment greatly increases the risk of subsequent multiorgan failure and death. The surgical judgements are similar to those required for obstruction, with the caveat that as few tissue planes as possible should be opened in order to reduce the risk of abscess formation. Resections should be conservative rather than radical, and the ends of the bowel should either be brought out as a stoma and mucous fistula or, if a primary anastomosis is fashioned, a proximal diversion should be made because an anastomosis is more likely to leak in the presence of severe sepsis. In addition the peritoneal cavity should be well lavaged with saline alone or containing antibiotics effective against enteric organisms. Systemic antibiotic therapy should be continued for at least 5 days. It is wise to leave the skin of the abdominal incision open, with delayed closure or secondary healing.

Oophorectomy

Reports indicate that 2 to 8 per cent of women with colorectal cancer have synchronous ovarian metastases and 1 to 7 per cent of those who undergo potentially curative resections develop subsequent ovarian metastases. Hence, one can argue that prophylactic oophorectomy may be useful in their overall management. Such an approach would reduce the risk for primary ovarian cancer, which is about 1 per cent for women over the age of 40 years and is perhaps higher in patients with colorectal cancer. However, ovarian metastases are almost always part of a wider dissemination. It remains unclear whether prophylactic removal of grossly normal ovaries containing micrometastatic colon cancer actually increases the cure rate. The risk of micrometastases in grossly normal ovaries removed at colectomy is only 1 to 2 per cent. If the ovaries are found to be grossly abnormal, they should be removed. If metastases are present, a hysterectomy is not required. If primary ovarian cancer is found, more complete surgical staging and a full hysterectomy is appropriate.

Anastomotic techniques

Ischemia is a major cause of anastomotic failure and subsequent leakage. Tension on the anastomosis and overtight sutures are other important, avoidable technical errors that may lead to failure of the anastomosis. The material with which the anastomosis is made does not seem to be of critical importance. Colocolic anastomoses were, until fairly recently, fashioned in two layers, an outer seromuscular layer with a non-absorbable suture and an inner, absorbable suture layer. The outer layer was usually interrupted, while the inner layer was often a continuous suture (Fig. 17). Randomized studies have demonstrated that a single layer of interrupted, full-thickness suture is as good or better. Interrupted sutures are probably safer than a continuous suture in the left colon, where the stool is solid, because they cause less stenosis before they have disintegrated. Occasional leakage after ileocolic or colocolic anastomosis has in general been considered ‘acceptable'. In the absence of obstruction, peritonitis, or the use of high-dose corticosteroids, leaks should be considered a technical error and great attention to detail when placing the individual sutures is required to avoid such problems. The strongest layer of the bowel is the submucosa, which must be included in all deeply placed sutures.

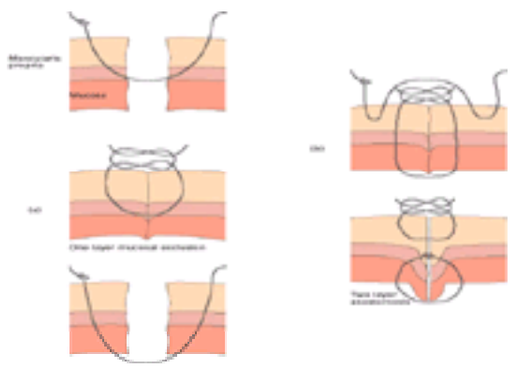


Fig. 17. Examples of various suture placements suitable for colorectal anastomoses. It is important always to remember that the major strength of the bowel wall resides in the submucosal collagen fibers.

Colorectal anastomoses, with or without preoperative radiation therapy, are associated with a small rate of leakage. Ischemia related to the proximal and distal mobilization, the use of ‘total mesorectal excision', tension, and pelvic sepsis add to the increased technical difficulty in performing a deep pelvic anastomosis. Diverting stomas do not reduce the leakage rate, but perhaps the morbidity and mortality of such leaks.

Stapling instruments, first developed in Russia, are now widely used to perform anastomoses (Fig. 18). Construction of a side-to-side anastomosis using linear stapling devices is suitable if the bowel can be delivered on to the abdominal wall. End-to-end anastomosis using a circular stapler is most useful for low pelvic anastomoses. The instruments may shorten operating time, particularly when performing a low anterior resection of the rectum, but they are much more expensive than sutures. Controlled studies do not show any convincing differences in leakage rates between hand-sutured and stapled anastomoses, provided the surgeon is equally competent with either technique. Selection of method is one of surgeon preference and economics. Concern has been expressed about steel staples potentiating recurrence of the cancer at the anastomosis, for which there is some experimental support. However, multivariate controlled clinical comparisons show that stapled anastomoses are not associated with increased risk of recurrence.

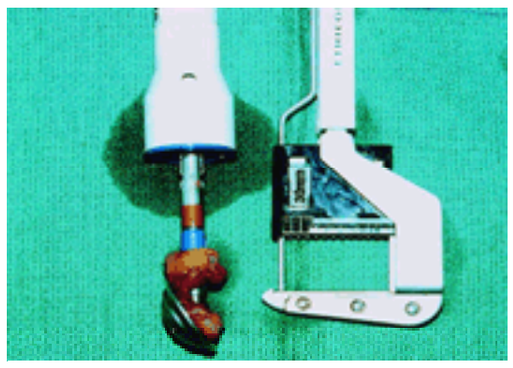


Fig. 18. Examples of surgical staplers: a linear stapler and an intraluminal (EEA) type.

Management of cancer in polyps

Approximately one out of every 20 ‘polyps' removed endoscopically will have cancer present on microscopic analysis. Cancers invasive only to the level of the muscularis mucosae do not have access to the lymphatic pathways and can be cured by endoscopic or surgical polypectomy. After endoscopic polypectomy, the major issue is the risk for residual localized or nodal cancer compared with the risk for definitive colectomy.

Fig. 19 demonstrates the various levels of invasion. In addition to that level, additional histopathologic data to be taken into account include the degree of differentiation (grade), the presence of lymphatic or blood vessel invasion, and adequacy of endoscopic resection (margin). Patients at high risk for local residual or nodal metastatic cancer have polyps containing one or several of the following features: poorly differentiated cancer, lymphatic vessel invasion, tumor invasive to level 3 or 4, or a positive or close polypectomy margin.

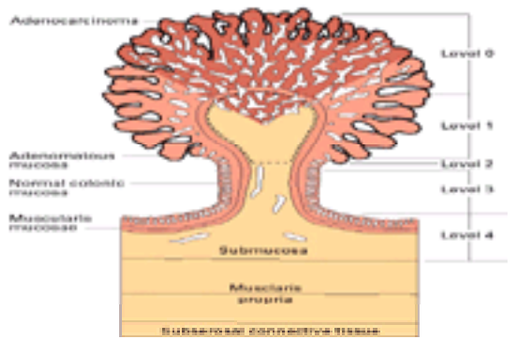


Fig. 19. Levels of invasion of early cancer within a pedunculated polyp.

The major issue is the risk for lymph-node metastases. Colacchio and colleagues have documented what they believe is an unacceptable risk associated with conservative (nonresective) management of most of these patients; they report an overall risk for nodal metastases of 10 per cent. Nivatvongs looked only at patients who subsequently underwent surgical resection and defined subsets based on the configuration of the polyps and gross extent of invasion (Table 9a,Table 9b,Table 9c and Table 9d); in the favorable group of patients the incremental benefit of surgical resection is likely only 1 to 2 per cent, and in many situations may be nil. It is strongly recommended that all patients with cancer in a polyp undergo immediate repeat endoscopy and India ink marking, in order to identify the site if resection is advised. In patients not undergoing resection the polypectomy site should be examined endoscopically in 4 to 6 months to confirm the absence of mucosal recurrence.

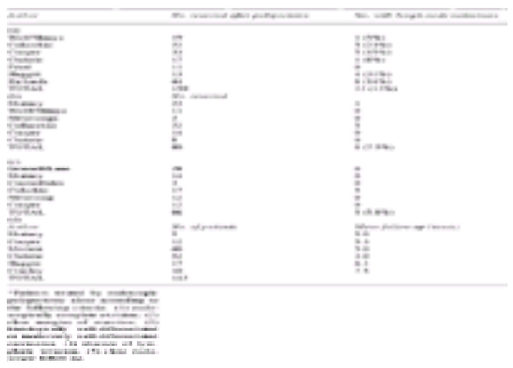


Table 9 Incidence of lymph-node metastases (a) in sessile and pedunculated colorectal polyps with invasive carcinoma; (b) in pedunculated polyps with invasive carcinoma; (c) in pedunculated polyps with invasive carcinoma limited to the head of the polyp; (d) malignant colorectal polyps treated by polypectomy alone*

Not all polyps can be removed endoscopically. Large, sessile lesions, particularly of the thin-walled ascending colon, pose a major therapeutic challenge. If endoscopic removal is attempted, piecemeal removal with the snare cautery over several sessions is required. Submucosal infiltration with saline facilitates removal by snare. Such patients should best be observed as inpatients for 1 to 2 days to rule out a perforation. Large, benign, sessile adenomas may require surgical resection as the least

morbid approach to removal. Laparoscopic-assisted resection of such right colonic tumors may be considered.

Laparoscopy

Rapid technologic advances and the clinical success of laparoscopic cholecystectomy have expanded the indications for minimally invasive surgery to include a broad range of general surgical procedures. Recently, laparoscopic techniques have been employed in the management of numerous benign colorectal diseases. The feasibility of laparoscopic-assisted colectomy has been well documented. Reports suggest short-term gains of reduced pain, shortened time in hospital, and improved cosmesis. However, the initial experience of laparoscopic colorectal procedures has identified an increased incidence of unrecognized injuries to the small bowel and ureters, anastomotic leaks, postoperative bowel obstruction, and port-site herniation. Overall, the incidence of major complications for a laparoscopically-assisted colectomy, based on early published reports, ranged as high as 20 per cent. However, in expert hands the operative morbidity for open and laparoscopic colectomy groups was similar. The likelihood of anastomotic leaks may be increased as surgeons struggle to minimize the incision size for the extracorporeal anastomosis, particularly after right colectomy.

In using laparoscopically assisted approaches in potentially curable cases, any putative short-term benefits from laparoscopic colorectal surgery, such as reduced pain, shortened hospital stay, accelerated activity, cosmesis, and possible cost reductions, must be balanced against the possibility of reduced cure rates. However, data on the extent of lymphadenectomy and margins of resection suggest that an oncologic, laparoscopic resection is identifiable. Laparoscopic rectal dissection may, in fact, allow for a more precise assessment of excision margins.

A major concern with laparoscopically assisted resections of colorectal cancer has been the apparent increased risk of recurrences in wound and trocar sites. Although the true incidence of recurrences at trocar sites is unknown, it appears to be greater than the 1 per cent incidence of recurrence in the wound described for patients undergoing open resection of colorectal cancer. The apparently increased risk for recurrences at trocar sites may be related to technical difficulties leading to spillage and inoculation of tumor cells or biologic aspects related to the pneumoperitoneum.

In order to clarify the risks and benefits of laparoscopically assisted colectomy, a prospective, randomized multicenter trial involving 1200 patients is currently under way in North America. An identical trial has begun in the United Kingdom. The principal aim is to examine whether disease-free and overall survival are similar among patients who undergo laparoscopically assisted or open colectomy for colorectal cancer. A secondary aim will be to determine the early and late complications and operative mortality of these two surgical approaches. The third aim is to determine if laparoscopically assisted colectomy is a cost-effective approach leading to better quality of life relative than the traditional operation.

Surgical treatment of colon cancer

Specific preoperative assessment

In patients who are generally fit medically, in the absence of malignant ascites or malignant hepatomegaly, surgical resection is appropriate as the first component of treatment. A baseline chest radiograph and abdominal ultrasound or CT are recommended, but not obligatory.

Surgical resection

The extent of surgical resection for colon cancer is primarily determined by the pattern of nodal spread. The regional lymph nodes at risk generally follow defined arterial arcades (Fig. 20).

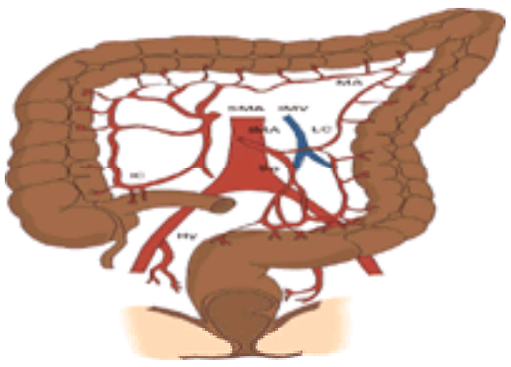


Fig. 20. Extent of surgical resection is defined by the arterial-associated nodal distribution. The normal arterial anatomy is shown here. The superior hemorrhoidal (superior rectal) artery is the continuation of the inferior mesenteric artery. It gives off several sigmoidal branches before entering the extraperitoneal rectum. IMA = inferior mesenteric artery; SMA = superior mesenteric artery; IC = ileocecal artery; MA = mesenteric artery; LC = left colic artery; IMV = inferior mesenteric vein; Hy = hypogastric artery;Sg = sigmoid artery.

Right colon cancers

Right hemicolectomy is the standard operation for cancers of the cecum and ascending colon, removing as little of the terminal ileum and as much of the ileocolic artery as is possible while still maintaining the blood supply to the terminal ileum (Fig. 21). The ‘space of Treves' is medial to the ileocolic artery, with an absence of lymphatics. Hence, taking more than 10 to 15 cm of ileum will not add to the nodal clearance. The right colic artery is also ligated at its origin. It is a variable vessel, branching from the ileocolic, superior mesenteric artery directly, or from the right branch of the middle colic. Some patients do not have any definable ‘right colic' artery. The distal extent of colon removed depends upon the site of the tumor. The resection may need to extend to the mid-transverse colon for cancers near the hepatic flexure, which always requires entering the lesser sac with clearance along the right gastroepiploic vessels and head of the pancreas. The terminal ileum is then anastomosed to the transverse colon in one or two layers with sutures or staples. The anastomosis may be end to end, end-to-side, or side-to side. In rare patients, primarily the elderly individual, the cecum may be preserved when resecting tumors at the hepatic flexure and a colocolic anastomosis fashioned, thus preserving the reabsorptive capacity of the cecum and reducing the risk of postoperative diarrhea (Fig. 22).

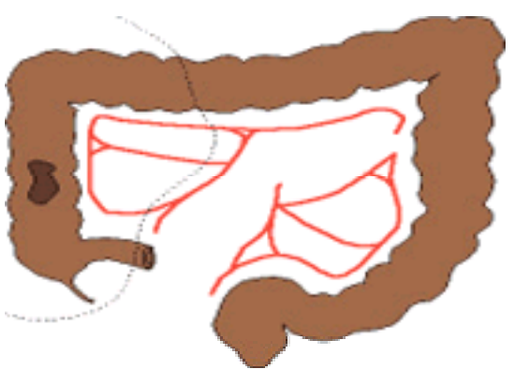


Fig. 21. Extent of resection for a cecal or ascending colon cancer; the ileocolic, right colic, and right branch of the middle colic vessels are divided for adequate lymphadenectomy.

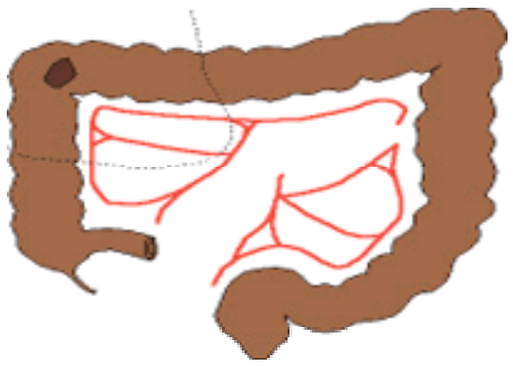


Fig. 22. Cecal-sparing resection in selected patients with hepatic-flexure cancers.

In younger and slender patients, lymphadenectomy may be carried out along the superior mesenteric vein to the uncinate process of the pancreas. It is important to remember that the superior mesenteric vein is to the right of the superior mesenteric artery, and may be damaged during aggressive mobilization and dissection. It is a thin-walled structure, not easily repaired. Bleeding at the root of the mesentery should not be dealt with by deep sutures placed ‘blindly'. If bleeding occurs, the vessels should be dissected and repaired under direct vision. A thrombosed or occluded superior mesenteric vein at this location will lead to infarction of the entire small bowel. The duodenum and pancreas are also at risk during mobilization of the hepatic flexure. If tumor abuts the duodenum, a full Kocher maneuver posterior to the pancreas and duodenum will facilitate visualization of the problem and allow an appropriate decision on duodenal clearance or resection.

Transverse colectomy

The lymphatic drainage for cancers of the transverse colon is along the middle colic vessels to the root of the superior mesenteric artery. Resection therefore removes most of the colon supplied by the middle colic artery and in particular the splenic flexure, where the pericolic vascular arcade may be incomplete (Fig. 23). The ascending colon is often resected to facilitate reconstruction and avoid a mesenteric ‘trap'. The omentum is removed *en bloc* with the tumor, leaving the gastroepiploic vessels. The most common procedure involves a right and transverse colectomy, with ileal reconstruction to the descending colon. Selected patients with early cancers may undergo a segmental resection of the involved colon. Resection of part of the stomach or jejunum may be required if the colonic tumor is invading either organ. The resection is completed by an end-to-end colocolic anastomosis.

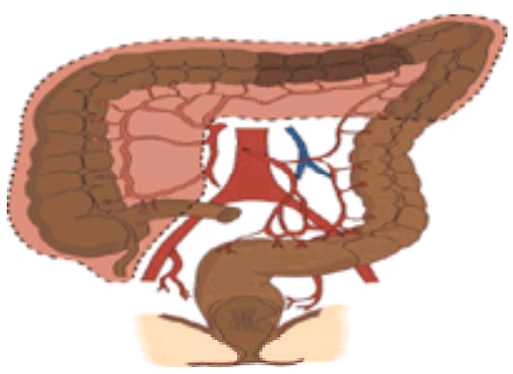


Fig. 23. Right and transverse colectomy for cancers in the mid-transverse colon or splenic flexure.

Left hemicolectomy

For cancers at the splenic flexure, descending colon, or proximal sigmoid colon, a left hemicolectomy is performed (Fig. 24). Nodal drainage from the splenic flexure is more common toward the left colic than the middle colic. Generous exposure is necessary to provide good access to the splenic flexure, which is often high and contiguous with the spleen. The lesser sac must be opened. If the splenic capsule is torn, the use of argon-beam coagulation as well as topical hemostatics may be successful in avoiding splenectomy. The proximal resection should include the splenic flexure because the ascending left colic artery is ligated at its origin and the pericolic anastomosis is often insufficient to ensure a good blood supply to the anastomosis. The transverse colon is mobilized sufficiently to perform a tension-free anastomosis. The distal resection extends into the sigmoid colon, preserving one of the distal sigmoid arteries and the pericolic vascular arcade. The inferior mesenteric artery is seldom ligated at its origin, but should it be necessary to take this vessel because of nodal metastases, the distal resection should be extended to the proximal rectum, which is adequately supplied from the middle rectal artery.

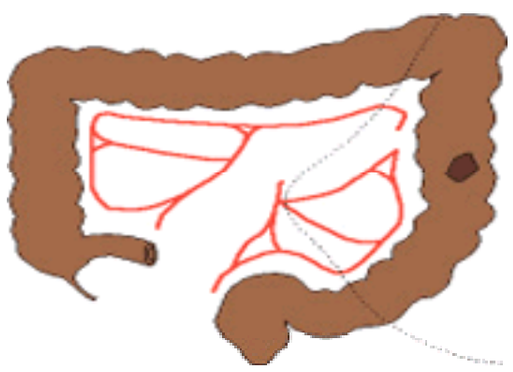


Fig. 24. Left hemicolectomy resection for a proximal sigmoid or descending colon cancer. If the inferior mesenteric artery (with superior rectal) is included in the resection, the anastomosis must be to the upper rectum to allow adequate vascularization from pelvic branches.

Sigmoid and rectosigmoid colectomy

Cancers of the mid or low sigmoid colon are resected with either a sigmoid or full left hemicolectomy. In either case, resection of the superior hemorrhoidal (rectal) artery is essential. Resection of the inferior mesenteric artery with full mobilization of the splenic flexure and left colectomy is based on the size and exact location of the patient's tumor, the redundancy of the sigmoid, nodal findings, age, general medical condition, and body habitus. Routine radical left hemicolectomy is not appropriate for every cancer of the sigmoid colon, but neither is a segmental sigmoid resection with removal of only the paracolic nodal drainage. Patients with positive nodal metastases along the inferior mesenteric artery are rarely cured.

Sigmoid cancers often invade surrounding structures such as the abdominal wall, ovary, uterus, bladder, or small bowel. In such cases, removal of the invaded organs in continuity is necessary. Care must be taken to identify and preserve the ureter during these dissections. Ureteral stents are primarily useful for recurrent rectosigmoid cases. Reimplantation of a resected ureter by bladder mobilization and ‘psoas hitch' is preferable to a ureteral anastomosis.

Adjuvant therapy of colon and rectosigmoid cancer (Table 10)

Agent	Efficacy
Systemic 5-fluorouracil (5FU)	None to marginal
Portal-vein chemotherapy (5FU)	Marginal; probably a systemic effect
5FU and levamisole (12 months)	One-third reduction in cancer mortality in node-positive patients
5FU and leucovorin (6 months)	One-third reduction in cancer mortality in node-positive patients, probably also node-negative
Monoclonal antibody (17-1A)	One-third reduction in cancer mortality in node-positive patients

Table 10 Adjuvant Strategies in colon cancer

End-results with surgical resection alone

For patients undergoing a ‘complete' curative resection, the overall 5-year survival rate is between 55 and 75 per cent. A further 10 per cent of patients will die of recurrence after 5 years. In the absence of microscopic or gross, local or distant, residual cancer, the most important prognostic factor is the presence and number of affected regional lymph nodes. For node-negative cancers treated by surgery alone the 5-year survival rate is 80 to 90 per cent. For node-positive cancers the 5-year survival ranges from 59 per cent (one positive node) to 27 per cent (six or more positive nodes), with an overall survival rate of 40 to 50 per cent. Other prognostic factors include T stage, grade, and bowel obstruction or perforation at the time of presentation. Site of colon tumor, age, and sex are not strong prognostic factors.

The prognosis with T1N0M0 cancers is 90 per cent, and almost 100 per cent in the subset with early invasive cancers. Intramucosal lesions may be limited to above the lamina propria and are true ‘*in situ*' cancers; these are non-invasive, and are frequently described as ‘severe dysplasia' to avoid undue anxiety in the patient and unnecessary recommendations for surgical resection by the physician. Invasive cancers that penetrate the lamina propria but remain above or at the muscularis mucosae are essentially all cured by surgical resection. Risk of spread to regional lymph nodes or distant sites is negligible until the muscularis mucosae is penetrated.

Defining high-risk, node-negative patients remains a challenge. With increasing data to support the benefits of adjuvant chemotherapy in node-positive patients, determining ‘high-risk, node-negative' in recommending adjuvant chemotherapy has become a practical, not just an academic, exercise. Transmural penetration (T4) involving the free mesothelial surface of the colon reduces the survival by 20 per cent. Invasion of a major microscopic vein in the bowel wall or mesentery also reduces survival. High-grade or lymphovascular invasion appear to have only minor impact on survival. Signet-cell histologic appearances are rare but a definite adverse factor. Mucinous histologic appearances, so-called colloid cancers, have little prognostic significance.

The presence of aneuploidy on flow cytometry of nuclear DNA from fixed tissue and the presence of an elevated serum carcinoembryonic antigen appear to have adverse prognostic power, and are being used by some clinicians as indicators for adjuvant chemotherapy. Certain molecular determinants, such as the concentration of thymidine kinase, mutation *p53*, and loss of the *DCC* oncogene (allelic loss on 18q) show promise as prognostic determinants.

The rationale for adjuvant chemotherapy stems from animal models which demonstrate that, in the presence of microscopic residual cancer, chemotherapeutic regimens effective, but not curative, for widespread cancer can lead to the eradication of micrometastases. Hence, clinical trials have explored the use of adjuvant treatment for high-risk patients compared with the traditional approach of surgical resection alone, and treatment only if recurrence subsequently becomes apparent. Since many patients in adjuvant trials will already have been cured by the surgical resection, such treatments must have a low treatment-related mortality to be justifiable. In colon cancer, this is the case.

Portal-vein chemotherapy

The rationale for adjuvant perioperative infusion of chemotherapeutic agents into the portal vein is based on the observation that cells embolize the portal venous system both preoperatively and intraoperatively, seeding the liver. There is considerable clinical and autopsy evidence that the liver may be the most frequent, and sometimes the only, site of metastasis. Approximately one-third of patients dying from their colon cancer have isolated hepatic metastases. Although established metastases are fed primarily by the hepatic artery, micrometastases are still likely to depend on the portal venous blood.

In 1975, Taylor and colleagues began a randomized study of intraportal 5-fluorouracil, using 1 g/day with 5000 units of heparin infused continuously for 7 days immediately postoperatively for patients with Dukes' A, B, and C lesions. About 250 patients were studied; by 1985, with a median follow-up of greater than 50 months, 80 deaths had occurred. There was a survival benefit in patients with node-negative colon cancer. This exciting study stimulated several randomized trials of similar design. The Australia and New Zealand Trial evaluated 372 patients with colon cancer randomly allocated to observation alone or to immediate postoperative chemotherapy with 5-fluorouracil, 600 mg/m^{2}per day for 7 days, given intravenously or intraportally. Compared with the other two groups, portal-vein infusion of 5-fluorouracil resulted in a highly significantly superior disease-free and overall survival in node-positive patients (in conflict with the Taylor trial). A similar study was conducted by the Swiss Group for Clinical Cancer Research: 505 patients with all stages of tumor were randomly allocated to a control or to a chemotherapy group that received immediate postoperative intraportal 5-fluorouracil, 500 mg/m^{2}per day with 5000 units of heparin, given by continuous infusion for 7 days. On the first day of therapy, a 10 mg/m^2 bolus of mitomycin C was also administered. All patients were followed for at least 5 years. The 5-year disease-free survival rate was 57 per cent for treated patients compared with 48 per cent for untreated controls (p = 0.05), and the overall 5-year survival rate was 66 per cent for treated patients and 55 per cent for controls (p = 0.026). Like the previous trial, the major benefit was in the node-positive patients.

Wereldsma and colleagues report the results of a multi-institutional Dutch study: 317 patients were randomized to portal-vein infusion of 5-fluorouracil and heparin as compared with portal-vein infusion of urokinase (10 000 units over 24 h) and no further therapy. In a median follow-up of 44 months, a significant decrease in liver metastases was demonstrated in the patients receiving 5-fluorouracil with heparin (7 per cent versus 23 per cent in controls; p = 0.01), but there was no improvement in overall survival after intraportal treatment.

The Mayo Clinic/North Central Cancer Treatment Group trial randomly assigned 219 patients with Dukes' B2 or C colorectal cancer to no further therapy after surgery or to receive portal-vein infusion of 5-fluorouracil, 500 mg/m^{2}per day with 5000 units of heparin per day for 7 days. In a median follow-up of 5.5 years, the incidence of hepatic metastases and estimated 5-year survival were essentially identical in the two arms of the study. The largest randomized study conducted was reported by the National Surgical Adjuvant Breast and Bowel Project in 1990 and updated in 1994: 1158 patients with Dukes' A, B, or C colon cancer were randomized to either no further treatment after surgery or to postoperative portal-vein infusion of 5-fluorouracil, 600 mg/m^{2}per day with heparin 5000 units/day for 7 days. A comparison between the two groups at 5 years revealed an improvement in disease-free survival (68 versus 60 per cent; p = 0.01) and overall survival (76 versus 71 per cent; p = 0.03) in favor of the group receiving chemotherapy, but there was no reduction in the incidence of liver metastases in the chemotherapy group compared with the control group. The failure of the portal-vein infusion to reduce the incidence of liver metastases while producing a modest survival advantage was attributed to a direct systemic effect of the chemotherapy.

The use of portal-vein infusion chemotherapy in the immediate postoperative recovery period is attractive because of its convenience, modest toxicity, and relative lack of added expense. However, the clinical results are equivocal. A meta-analysis of nine randomized trials on more than 3000 individual patients has revealed a statistically significant reduction in risk of death of 13 ± 6 per cent and an improvement in disease-free survival of 14 ± 5 per cent in favor of portal-vein infusion therapy. It remains uncertain whether these beneficial effects result from the unique route of drug administration employed or from the systemic effects of the treatment. It should be stressed that no portal-vein chemotherapy program has been long term, for example 6 months. Short-term portal-vein chemotherapy may be of value in conjunction with long-term systemic chemotherapy, but this awaits results of a clinical trial in North America, EST 1292.

Systemic chemotherapy

Combination chemotherapy

In the 1970s and 1980s, trials of combination chemotherapy were begun. Six studies used a surgery-only control (the exception was the Eastern Cooperative Oncology Group 2276 study). Six studies used a combination of methyl-1(2-chloroethyl)-3-(cyclohexyl)-1-nitro-sourea (**methyl-CCNU**) and 5-fluorouracil with (MOF) or without (MF) vincristine. Combination chemotherapy (MOF or MF) was assumed to be more active than 5-fluorouracil alone in advanced disease. The empiric application of nonspecific immunotherapy with bacillus Calmette–Guérin (**BCG**), BCG-MER (BCG with a methanol extraction residue of BCG), or levamisole was studied. More than 4000 patients were evaluated. A trend favoring chemotherapy was noted in several studies, but statistically significant benefit was detected only in the National Surgical

Adjuvant Breast and Bowel Project C01 and the North Central Cancer Treatment Group 78–48–52 studies.

Levamisole

Levamisole is a synthetic phenylimidothiazole that has been in use for many years as an antihelminthic agent. Interest in the use of levamisole in cancer therapy developed after the observation that it enhanced the immune response of mice vaccinated against *Brucella*. Subsequent studies have demonstrated that levamisole has a broad range of immunomodulatory properties, including enhancement of specific immune responses in normal hosts and restoration of immunity in immune-deficient hosts. Levamisole has been evaluated alone and in combination with 5-fluorouracil as therapy for metastatic colorectal cancer without additional benefit.

A study of single-agent levamisole used as an adjuvant was reported by Verhaegen in 1974. Sixty patients only were randomized either to levamisole or to no further therapy after resection of Dukes' A, B, or C colon and rectal cancer. Five-year overall survival was 69 per cent for patients receiving levamisole compared with 37 per cent for those in the control group. Two subsequent prospective, randomized, placebo-controlled trials have been completed and have failed to demonstrate a survival advantage for treatment with single-agent levamisole.

Windle and colleagues randomized 131 patients with colon or rectal cancer to no further therapy, 5-fluorouracil alone, or the combination of 5-fluorouracil and levamisole. With a minimum follow-up of 5 years, the combined 5-fluorouracil and levamisole arm demonstrated a significant superiority to 5-fluorouracil alone in overall survival. These data have subsequently been confirmed in two large trials, which led to a change in patient care in the United States. The North Central Cancer Treatment Group/Mayo Clinic reported the results of a randomized study comparing levamisole alone with a combination of levamisole and 5-fluorouracil and with no further therapy after surgery. A total of 401 eligible patients was randomized; almost all had colon cancer. With a median follow-up in excess of 7 years, the levamisole/5-fluorouracil combination demonstrated a significant improvement in disease-free survival compared with no further therapy ($p = 0.02$). A subset analysis demonstrated a significant disease-free survival advantage for 5-fluorouracil plus levamisole only in stage C patients. A major United States Intergroup trial enrolled 1296 patients with colon cancer. For stage C patients, levamisole alone produced no effect on recurrence rate or overall survival. The combination of levamisole and 5-fluorouracil resulted in a statistically significant reduction in tumor recurrence (disease-free survival 63 per cent as compared to 47 per cent for surgery only) and improved survival (3-year overall survival of 71 per cent as compared to 55 per cent for surgery only). Longer follow-up of this study continues to demonstrate a one-third reduction in cancer mortality.

Toxic effects attributable to the combination of 5-fluorouracil and levamisole are generally mild and reversible. Leucopenia, nausea, diarrhea, and stomatitis occur most commonly. Toxicity attributed primarily to levamisole has included nausea, diarrhea, dermatitis, metallic taste, fatigue, arthralgias, fever, seizures, tremors, agitation, and confusion.

Leucovorin

Because of its superior activity in metastatic colon cancer, the combination of 5-fluorouracil and leucovorin has been studied extensively as an adjuvant in colon cancer patients. The National Surgical Adjuvant Breast and Bowel Project CO3 trial evaluated several adjuvant regimens. With a median follow-up of 4 years, significantly better disease-free and overall survival were observed for the group receiving 5-fluorouracil/leucovorin. The IMPACT study, a cooperative effort of the National Cancer Institute of Canada, Gruppo Italiano Valutazione, and Foundation Francaise Cancerologie Digestive, represented pooled data from three independently conducted studies comparing surgery alone to surgery followed by 5-fluorouracil (400 mg/m^2per day) and leucovorin (20 mg/m^2per day) given five times daily every 4 weeks for 6 months. The study enrolled 309 patients and demonstrated a disease-free survival advantage for the chemotherapy group. Finally, an Italian study compared 5-fluorouracil plus leucovorin chemotherapy administered for 1 year to surgery alone in 236 patients with Dukes' stage B2 and C colon cancer. Once again there was a survival advantage for the chemotherapy group compared to surgery alone. In all of the 5-fluorouracil/leucovorin studies, the major benefit has been in the node-positive patients, and the overall extent of cancer-related mortality reduction is approximately one-third. Overall toxicity has been ‘acceptable', with very few treatment-related deaths.

Several North American cooperative group studies have now compared the various 5-fluorouracil-based adjuvant regimens in node-positive patients. None of these trials had a surgery-only control arm. The use of levamisole, leucovorin, levamisole plus leucovorin, and various treatment durations were studied. Initial reports indicate that 5-fluorouracil/leucovorin by intravenous bolus injection for 6 months provides major survival benefit with modest toxicity, and this has rapidly become the ‘standard' for patients not being treated in clinical trials. It should be stressed that none of the completed studies of adjuvant chemotherapy has yet demonstrated a major benefit for chemotherapy in patients with stage B2 disease, likely related to the excellent prognosis of such patients with surgery alone. Current colon adjuvant trials are comparing fluorouracil/leucovorin with or without CPT II or oxaliplatin.

Immunotherapy

Although attempts at non-specific stimulation of the immune system with agents such as BCG or levamisole have not proved effective as adjuvants, various vaccine and antibody strategies have been assessed. Hoover and colleagues randomized 80 patients with stage B2 or C colon or rectal cancer to receive autologous tumor vaccine or no further therapy. Autologous vaccine (irradiated autologous tumor cells plus BCG) was given by intradermal injection weekly for three consecutive weeks. With a median follow-up of 6.5 years, there was a statistically significant improvement in both disease-free and overall survival for colon cancer patients receiving immunotherapy. However, an Eastern Cooperative Oncology Group study of irradiated autologous tumor cells combined with BCG vaccine therapy given to 412 patients with resected Dukes' stage B and C colon cancer failed to confirm a beneficial effect.

Another strategy is passive-specific immunotherapy using murine monoclonal antibody. A study by Reithmuller *et al.* demonstrated that postoperative adjuvant therapy with monoclonal antibody 17–1A resulted in improved disease-free and overall survival among patients with Dukes' C colon cancer compared to a control group treated with surgery alone. After a median follow-up of 5 years, antibody treatment reduced the overall death rate by 30 per cent and decreased the recurrence rate by 27 per cent; these results are similar to the benefits of adjuvant cytotoxic chemotherapy. Confirmatory studies are under way in the United States and Europe; antibody alone or in combination with chemotherapy is being evaluated.

Recommendations

Generally healthy patients, following potentially curative resection of node-positive colon cancer, should be offered participation in a clinical trial or advised to received 6 months of 5-fluorouracil/leucovorin or 12 months of 5-fluorouracil/levamisole. Node-negative patients deemed to be at high risk because of clinical, tumor-related, or biologic factors should also be entered into clinical trials or offered adjuvant chemotherapy. They should be aware that there is no unequivocal proof that survival is improved with such treatments.

Surgical treatment of rectal cancer

Preoperative assessment

As discussed previously, patients with rectal cancer warrant a more complete assessment of the extent of their disease, since local treatment options are suitable for some patients with early cancers, and preoperative radiation for others with locally advanced cancers. Hence, chest radiographs, abdominal pelvic CT scans, and endorectal ultrasonograms are obtained in most patients.

Local treatments

Local treatment of early rectal cancer is feasible by transanal local excision, suprasphincteric posterior proctectomy (the Kraske procedure), trans-sphincteric posterior approaches (Bevan or York–Mason procedure), transanal endoscopic microsurgery, fulguration, or local/contact radiation therapy (Papillon approach). For all of these, correct selection of the patient is pivotal to successful treatment. Some important issues are the patient's age, overall medical condition, body habitus, and personal preferences. Tumor-related features include size, location, configuration (exophytic or ulcerated), mobility and stage by digital rectal examination and intrarectal ultrasound, and histologic features such as grade and lymphovascular invasion. The goal is to select patients with minimal risk of transmural and regional lymph-node spread. A mobile, well-differentiated, exophytic T1 lesion less than 2 cm in size is ideal. Larger, deeper tumors and those with high-grade or lymphovascular invasion are treated by local excision as a ‘compromise' strategy only, primarily in medically unfit patients. Surgical excision for cure should always involve a full-thickness resection, ideally attempting to remove a few perirectal lymph nodes if feasible. The ‘harmonic scalpel' is a helpful technical addition.

Transanal excision

This is the technique most frequently used for local excision of rectal carcinomas. However, because of limitations in exposure it should be reserved principally for

relatively small tumors (less than 3 cm in diameter) within 6 to 8 cm of the anal verge and generally limited to one quadrant of the rectal circumference. Although a 1-cm margin is ideal, a 5-mm margin of grossly normal mucosa beyond the edge of the tumor is more realistic. The rectal wall is excised to the level of perirectal fat, with emphasis placed on a perpendicular dissection in order to prevent undermining and premature ‘coning-in'. If feasible, perirectal fatty tissue should be included in the excision specimen in order to attempt to detect nodal spread. Proper orientation of the specimen is maintained as it is pinned on a flat surface to facilitate accurate evaluation of marginal status and depth of invasion by the pathologist. The wound should be irrigated frequently to reduce the potential risk of implanting malignant cells.

Exposure and lighting are the most problematic. The patient may be positioned in lithotomy, prone jack-knife, or lateral. Use of various self-retaining retractors is helpful. A fiberoptic headlight is essential. Cautery is used to define the proposed excision margins, then begun inferiorly (caudad) until perirectal fat is identified. Anteriorly, in females, fingers should be placed in the vagina to control the depth of dissection at the level of the rectovaginal septum. In men, dissection anteriorly should be at the level of Denonvilliers' fascia. Extensive undermining at the level of the perirectal fat is important prior to excision of the lesion; this will greatly facilitate closure, which is generally done with a single layer of full-thickness absorbable sutures.

Trans-sphincteric and Kraske suprasphincteric approaches

These have been used for lesions that are not, because of their proximal or anterior location, suitable for a standard transanal approach. In husky males, transanal techniques are problematic and these posterior approaches are advantageous. In addition, lymph-node excision may be facilitated by such procedures. The suprasphincteric or Kraske operation facilitates wide resection of posterior or lateral lesions (Fig. 25). It is done with the patient in the prone jack-knife position. Midline, parasacral, and horizontal curvilinear incisions have been described. I prefer a direct midline incision. Methylene-blue tattooing facilitates appropriate alignment for skin closure. Once the suprasphincteric midline fascia (anococcygeal ‘ligament') has been divided, the pelvis is entered by dividing the levator muscle in the midline. The coccyx and a portion of S5 sacrum is removed with a rongeur. For a posterior lesion, the mesorectum is mobilized superiorly and laterally. With a finger in the rectum defining the tumor location, the mesorectum is divided down to the muscle. The rectal wall with overlying fat is then excised, with an intraluminal finger providing control of the margins. After copious irrigation, the defect is closed, preferably in a transverse fashion. A suction drain is used for 2 to 3 days. Diverting colostomy is rarely needed.

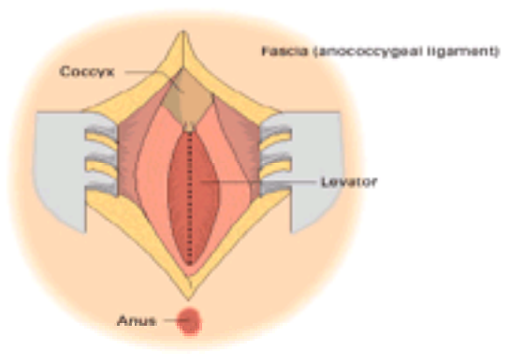


Fig. 25. Kraske posterior proctotomy: after dividing the midline fascia (anococcygeal ligament), the coccyx and a portion of S5 may be removed to facilitate exposure.

York–Mason or Bevan operations

In these a posterior proctotomy is combined with sphincter division. Scrupulous mechanical and antibiotic bowel preparation is essential. The procedure begins in an identical fashion to the Kraske approach. Once the rectal wall is exposed, the anoderm is divided with cutting cautery. A right-angle clamp exposes portions of the sphincter mechanism, and is ‘tagged' with figure-of-eight sutures. The muscle is divided between sutures. Matching clamps must be used to allow matching of appropriate muscle during closure. Exposure of anterior-wall lesions is excellent. Closure of the proctotomy and sphincter must be meticulous. Diverting colostomy is not required.

Transanal endoscopic microsurgery

This new technique, developed in the 1980s, incorporates a high-quality binocular operating system and pressure-regulated insufflation with continuous suction, which provides access to sessile polyps and small cancers from the mid and upper rectum. The results from a recent tri-institutional report suggest that this technique is suitable in the management of sessile adenomas located in the mid and upper rectum, and of carefully selected carcinomas.

Fulguration and intracavitary radiation

Alternative local treatment options include fulguration and local radiation therapy. Fulguration can be safely performed completely through the bowel wall. A ball-tip cautery with frequent curettage of charred material is required. Late, delayed hemorrhage as the eschar sloughs occurs in 20 per cent of patients. Fulguration should not be used near the vagina in women or the urethra in men. Local or contact radiation is available at relatively few centers. Low-energy (50 kV) X-ray is applied through an operating (4 cm) proctoscope in doses of 3000 cGy every 3 to 4 weeks for three to four sessions. A small, mid-posterior lesion is ideal, but these are usually amenable to excision.

Radical resections

Most rectal cancers require radical resection because 70 to 80 per cent present with disease beyond the rectal wall, either by direct extension or lymphatic spread. Optimal oncologic and functional results require a technically precise surgical approach selectively integrated with adjuvant chemoradiation therapy.

Since surgery for rectal cancer is a local therapy, its oncologic efficacy is based principally on its rate of local control. The pelvis is the most frequent site of tumor recurrence, a major cause of morbidity and death. Salvage therapy in most cases is limited, offering incomplete and only temporary palliation. Therefore, the attitude that a pelvic recurrence is best prevented should prevail and help guide the choice of operation and the pelvic dissection technique.

The major tumor-related risk factors for local recurrence are the number of involved regional lymph nodes, the extent of transmural penetration, and tumor grade. Two observations also strongly implicate inadequate surgical resection as a major cause of pelvic recurrence. First, involvement of the lateral or circumferential margin of the resection strongly correlates with subsequent local recurrence. Traditional rectal resection yields a positive lateral margin in 25 per cent of cases, with approximately 80 per cent of these patients developing a local recurrence. Second, studies from the United Kingdom and Germany demonstrate that the frequency of local recurrence varies among individual surgeons from less than 10 per cent to more than 50 per cent. Therefore, the surgeon's operative technique and ability to achieve a negative circumferential margin are strong determinants of local control.

Mesorectal excision

Heald and colleagues have advocated total mesorectal excision in conjunction with a low anterior or abdominoperineal resection as the optimal surgical treatment for rectal cancer. This technique involves removal of the entire rectal mesentery, including that distal to the tumor, as an intact unit. Total mesorectal excision requires precise dissection in an areolar plane outside the visceral fascia that envelopes the rectum and its mesentery. Hence, a better term is sharp mesorectal excision. In patients with mid to low rectal cancer, sharp mesorectal excision combined with total mesorectal excision is an optimal surgical strategy. In contrast to conventional blunt dissection techniques, sharp mesorectal excision facilitates nerve preservation, enables complete hemostasis, and emphasizes gentle handling to avoid tearing or disruption of the smooth outer surface of the mesorectum. In a large personal series of patients treated with rectal cancer resection and total mesorectal excision, Mr Heald has reported a 5 per cent local failure rate without the use of radiotherapy. These outstanding results have been attributed to improved lateral clearance, enhanced removal of potential tumor deposits in the mesentery, and decreased risk of tumor spillage from a disrupted mesentery; this last is important since tumor deposits have been detected in the mesentery distal to the gross tumor. Although these explanations for success are not fully proved, there is increasing evidence to support the conclusion that sharp/total mesorectal excision does improve local control. In comparison to the only 80 per cent negative lateral margins obtained with blunt dissection, total mesorectal excision has been shown to achieve a negative circumferential margin in 93 per cent of resected specimens. Most importantly, other surgeons using similar total mesorectal excision techniques have reported local failure rates of less than 10 per cent for transmural and/or node-positive rectal cancers. Although no randomized trial of total mesorectal excision has been performed, it has been evaluated prospectively in Sweden, where it has been introduced via a

formal, preceptorship-based, training program. A 5-year prospective audit reveals a local recurrence rate of 7 per cent following the addition of total mesorectal excision compared to a historic control rate of 23 per cent.

Details of the pelvic dissection

Whether for a sphincter-preserving operation or an abdominoperineal resection, the essence of an optimal operation for mid or low rectal cancer is to dissect sharply, using scissors and/or cautery, in defined anatomic planes outside the mesorectum. Although the mesorectum is not a true mesentery, it is the term in common usage to describe the node-bearing fatty tissue that surrounds the extraperitoneal rectum. It is covered by visceral fascia. There is an areolar plane between the mesorectal fascia and the posterior pelvis, which is bone, and the lateral pelvis, which is the parietal fascia overlying the pyriformis muscles and, inferiorly, the levator ani. The visceral and parietal fascia fuse into the ‘rectosacral fascia', which tethers the posterior rectum into the sacral hollow. Anteriorly, perirectal fat abuts Denonvilliers' fascia, which envelops the seminal vesicles and fuses with the most caudal portion of the prostatic capsule. In women, perirectal fat abuts the rectovaginal septum. The posterior–superior portion of the mesorectum at the level of the sacral promontory encompasses the superior hemorrhoidal vessel and nodes. In a separate posterior plane is Waldeyer's fascia, within which are the sympathetic nerve trunks. The main parasympathetic nerve trunk for male erections is S3, which exits in the mid-pelvis, just distal to the pyriformis muscle, and lies just medial to the parietal pelvic fascia.

Low anterior resection with autonomic nerve preservation

As described above, the essence of the pelvic dissection for mid and distal rectal cancer involves sharp dissection, using scissors and cautery in clearly defined planes peripheral to the visceral fascia enveloping the mesorectum. Ideally, this dissection allows identification and preservation of the main autonomic nerve trunks within the pelvis. Surgery is performed with the patient in the modified lithotomy position. The rectum is aspirated, and antibiotics given, prior to the skin incision. In women, the vagina is prepared as well. After general abdominal exploration, the inferior extent of the tumor is determined. If the major tumor mass is at or below the peritoneal reflection/pouch of Douglas, then a total mesorectal resection is performed. If the tumor is above the peritoneal reflection, the mesorectum may be divided 5 cm below the tumor.

The sigmoid and descending colon are mobilized, and the ureters are identified and preserved. The tissue overlying the sacral promontory is left in place as the superior hemorrhoidal vessel is lifted. The left and right sympathetic nerve trunks are identified and dissected along their medial edges and pushed laterally. Waldeyer's fascia between the sympathetic nerves is divided, and the distal dissection proceeds along the bony descending pelvis. The pyriformis muscles are cleared. Strong traction on the distal divided sigmoid colon with retraction in reverse Trendelenburg, with the benefit of a Parks' or similar retractor and a fiberoptic headlamp, allow scissor spreading and cautery division within the posterior areolar plane. In the mid and distal sacral hollow the rectosacral fascia is encountered (Fig. 26), and entered with scissor spreading and cautery. The levators are then cleared by sharp dissection under direct vision. The anterior dissection proceeds along the rectovaginal septum in women. In the absence of a prior hysterectomy, a heavy traction suture in the uterus provides the necessary countertraction. If hysterectomy has already been performed, sharp dissection along the rectovaginal septum is done with several fingers in the vagina. In men, the anterior dissection is generally anterior to Denonvilliers' fascia; the exception is a relatively early posterior rectal cancer. Cautery is used to open the peritoneum just anterior to the peritoneal reflection, exposing the seminal vesicles. With strong traction on the rectum, scissor and cautery dissection clears the anterior aspect of Denonvilliers' fascia. Sharp scissor separation of Denonvilliers' fascia from the distal prostatic capsule is required. At this point in the operation, sharp dissection is used to define the levators anterolateral to the rectum, meeting the previously cleared posterolateral levators.

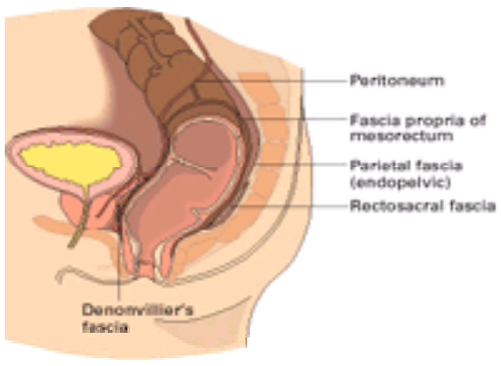


Fig. 26. The fascial layers of the rectum in the sagittal view.

After the anterior and posterior dissection is complete, the lateral or side-wall dissection proceeds. A bulky asymmetric tumor may require sacrifice of the parasympathetic nerve trunks on the affected side. Otherwise, the main S3 trunk may be preserved. Just distal to the pyriformis the S3 nerve is identified posterolaterally, and dissected from posterior to anterior. The nerve is most vulnerable anterolaterally where it merges with the sympathetic trunk and arches anteromedially toward the prostate or vagina. Careful use of dissecting scissors, clips, cautery, and the endo-GIA vascular laparoscopic stapler facilitates this portion of the dissection. At this point the mesorectum should be intact, with total mobilization to the anorectal ring. A purse-string or staple line is placed for a stapled colorectal reconstruction, or a peranal mucosectomy for a coloanal reconstruction. If at all feasible, the rectum should be copiously irrigated prior to division to minimize the potential for tumor cell implants in the suture or staple line. Stapled or sutured reconstruction is performed, with or without mobilization of the splenic flexure, and with or without a J-pouch reconstruction of the colon based on the tumor location, quality of the proximal colon, and the patient's body habitus.

Splenic flexure mobilization

Sufficient proximal length to allow the colorectal anastomosis not only to be tension free but also to recreate the 90° anorectal angle by having the colon lie against the sacrum (sacralization) is important for subsequent bowel control. Studies do not support routine ‘high ligation' with division of the inferior mesenteric artery and inferior mesenteric vein on oncologic grounds; such division is necessary for adequate length and mobilization of the proximal bowel. The ‘arc of Riolan' is an ascending branch from the inferior mesenteric artery along the inferior mesenteric vein and thence to the bowel near the splenic flexure. If present, vessels should be carefully assessed prior to division because of concern about adequate collaterals along the marginal artery of Drummond. Full mobilization of the left colon from the retroperitoneum above Gerota's fascia, off the retroperitoneum along the pancreas to the left branch of the middle colic artery maximizes bowel length. However, it also provides a denervated bowel, at risk for subsequent spasm and dysfunction following reconstruction. In addition, the spleen is at risk during the procedure. Overall, it is preferable to mobilize the splenic flexure not only for length, but to allow the more proximal, higher quality, more compliant colon to be used for reconstruction.

Sutures or staples

Adequate length, vascularity, exposure, hemostasis, and lighting are the most important aspects in achieving a low pelvic anastomosis. A fiberoptic headlight can be essential. Whether using sutures or staplers, a the distal rectum should be washed out, particularly in the nonradiated patient, in order to minimize tumor cell implants. Since the advent of circular stapling devices in the 1970s, stapled colorectal procedures have evolved to become as safe as sutured anastomoses. A variety of methods have been developed to facilitate construction of the anastomosis. However, the double-stapled technique developed by Knight and Griffin in 1980 is most frequently used in creating a distal anastomosis following resection of a low-lying rectal cancer. This approach minimizes intraoperative fecal contamination, and effectively deals with the discrepancy in size between the mid-rectum and the proximal colon. Although the initial experience with stapled anastomosis suggested an increased number of complications, recent studies report low rates of leakage and mortality with stapled colorectal anastomosis while obtaining increased rates of sphincter saving with diminished needs for temporary fecal diversion.

In order to best utilize surgical staplers, the patient must be positioned in modified lithotomy to allow peranal placement ofthe stapler. Colotomy and placement of the circular stapler from the proximal bowel should be avoided if at all possible. The proximal bowel is divided initially, using a clamp, staple line, or purse-string staple applicator. At the completion of the pelvic dissection, a right-angle clamp or staple line is placed below the tumor, and the rectal wash-out is performed. Another staple line (Fig. 27) or purse-string clamp is placed below (distal) and the specimen is removed.

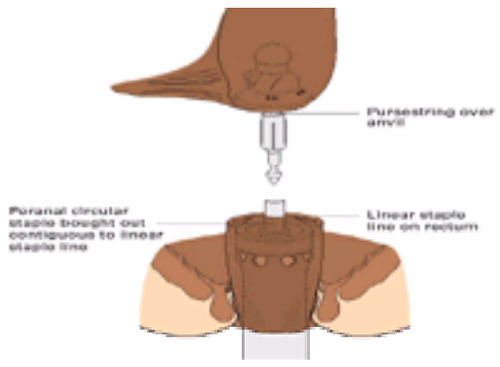


Fig. 27. The ‘double staple' technique for rectal reconstruction avoids a distal purse-string suture and obviates concerns about the discrepancy in size between the colon and the capacious mid-rectum.

The size of the circular stapler is determined by the size ofthe proximal colon. If the double purse-string approach is used, the stapler is placed per anus with the head (anvil) attached. For the ‘double-staple' approach, the anvil is removed and replaced withthe barbed point. When placed per anus, care must be taken not to tear the anus, and not to suddenly ‘pop through' and rip the residual rectal pouch. Once in the rectum, a fist around the stapler shaft allows pressure on the perineum, producing improved visualization from the abdominal exposure. The pin is advanced and punctures made just to either side of the linear staple line, but not directly through it. The electrosurgical cutting current along the stapler pin is ideal to avoid rectal tearing. When removing the pin, care should be taken not to damage the presacral veins when the pin ‘pops out'.

Placement of the head/anvil into the proximal bowel requires traction sutures and adequate stretching. Water-soluble jelly on the anvil is helpful. After tying the proximal purse-string, mesenteric fat with blood vessels should be cleared very gently from the end of the bowel. The use of a long clamp to place the anvil pin into the stapler is important. Hands should not be used on the fragile bowel. Prior to closing and firing, great care is taken to exclude fat and the vaginal wall from the anastomosis. The operating surgeon should place one hand on the anastomosis and hold the (contaminated) stapler in the other hand to facilitate its removal after firing; this allows gentle and controlled pulling of the stapler through the anastomosis. Air insufflation is then used to confirm the completeness of the anastomosis. Palpation and endoscopy may also be helpful, but all of these anastomoses are fragile, and excessive inspection may cause more problems than it solves. Two complete ‘donuts' and airtight to insufflation are adequate reassurance.

End-to-end or side-to-end

The side-to-end Baker anastomosis was commonly used for many years. It facilitated exposure and management of the size discrepancy between the colon and mid-rectum. There are some data to suggest the incidence of anastomotic leakage may be less with a side-to-end approach (see J-pouch discussion next). Most stapled colorectal reconstructions are done end-to-end. In the presence of extensive diverticulosis, a stapled or sutured side-to-end approach may be helpful.

J-pouch

It is assumed that a narrow-caliber, poorly compliant colon used for a low colorectal or coloanal anastomosis will result in frequency and urgency. A number of investigators have implemented the construction of a colon J-pouch to create a neorectal reservoir (Fig. 28). Two groups from France were the first to report the use of colon J-pouches as part of the Parks' coloanal reconstruction. Out of 20 patients with coloanal pouch anastomoses, Lazorthes and associates in Toulouse report that 87 per cent had one to two bowel movements per day at 1 year, compared to 33 per cent in a group of 33 patients with straight coloanal anastomoses. Bowel frequency was correlated with reservoir size. Parc and colleagues in Paris utilized this approach in 31 patients, with a resultant mean daily bowel movement of 1.1. However, 25 per cent of patients were unable to evacuate spontaneously, requiring the use of an enema every other day. The importance of the pouch in reducing stool frequency has been confirmed by Nicholls and associates from St. Mark's Hospital, London, and Kusunoki from Hyogo, Japan. The use of an 8-cm colon J-pouch has also been reported by Drake and coworkers from the Mayo Clinic and Cohen from the Memorial Sloan-Kettering Cancer Center.

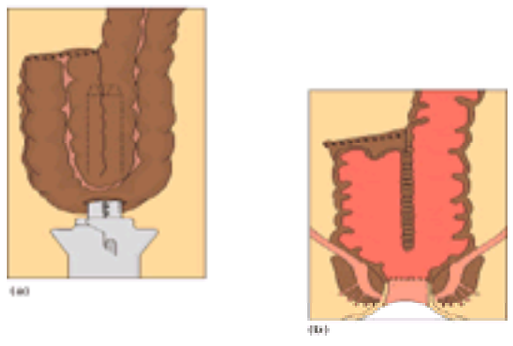


Fig. 28. Colon J-pouch should be 6 to 8 cm in the maximum length; the apex can be stapled or sutured to the distal colon or anus.

Hallbrook, Pahlman, and Wexner have completed a randomized trial of colon J-pouch reconstruction performed primarily in Sweden and at the Cleveland Clinic, Florida. Ninety-seven patients were randomized. All had total mesorectal excision and most a stapled colorectal or high coloanal reconstruction. There were fewer anastomotic leaks in the pouch group, perhaps related to the ‘side-to-end' anastomosis. There was a significant reduction in frequency of bowel movements, urgency, and nocturnal bowel movements in the pouch group in the first year following the operation.

J-pouch construction

Following mobilization of the entire left colon, and placement of a linear staple line, the antemesenteric tenea are approximated with two parallel rows of fine silk sutures. The pouch is made 5 to 8 cm in length. Longer, more capacious pouches will not evacuate adequately, and will render the patient enema-dependent. The apex of the J is cleared of the mesentery and opened, and four holding sutures placed. Several applications of the GIA-type stapler are required to form the pouch. After the first application, Allis clamps are used to evert the staple lines to check for bleeding and allow placement of the second firing at the appropriate ‘crotch' in the initial staple apex. These small colonic pouches fit rather well into the pelvis, and are suitable for the endoanal sutured or the coloanal stapled reconstructions. If a circular stapled reconstruction is to be used, the anvil is placed through the open apex of the J-pouch, and a purse-string placed. If a sutured coloanal reconstruction is planned, traction sutures are placed in the open pouch apex to facilitate positioning into the anal canal.

Coloanal reconstruction

Although low colorectal anastomoses, either hand-sewn or stapled, are feasible in many patients with low-lying rectal cancers, a coloanal reconstruction is another alternative to a permanent colostomy. Since fewer than 10 per cent of rectal cancers demonstrate distal mural spread beyond the gross edge of the tumor and only 2.5 per cent have histologic evidence of spread beyond 2 cm, margins of 1 cm may be adequate in most low-lying rectal cancers with the exception of poorly differentiated or bulky lesions. It is estimated that in approximately 5 to 10 per cent of patients with low-lying rectal cancers, a sphincter preservation is made possible by a coloanal anastomosis. In these cases, a distal rectal cancer is adequately resected using a total mesorectal excision. However, due to the pelvic anatomy, usually a narrowed male pelvis with an enlarged prostate, intrapelvic anastomosis may be unsafe or technically impossible. In these cases, a coloanal reconstruction is done via a stapled peranal anastomosis to the anal rectal ring or a hand-sewn peranal approach (Parks' procedure). The use of the term ‘coloanal reconstruction' should be restricted to peranal sutured reconstructions (Fig. 29).

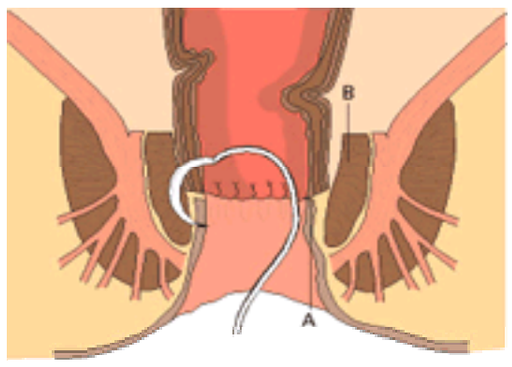


Fig. 29. The ‘peranal' sutured coloanal reconstruction. Following mucosectomy, the distal anoderm (point A) may prolapse distally. The bowel anastomosis should capture the anoderm, then the high anal-canal internal sphincter (point B) before passing through the colon.

Recently, it has been shown that preoperative radiation therapy can down-stage the primary tumor, enabling the surgeon to change the planned procedure from an abdominal perineal resection to a low anterior resection with a coloanal anastomosis; this is important in low-lying transmural tumors. In our experience and that at Thomas Jefferson University, approximately 90 per cent of patients with tumors palpable on rectal examination may undergo sphincter-preserving resection using preoperative radiation therapy and total proctectomy with coloanal reconstruction.

For coloanal reconstruction, the total mesorectal excision as described above is used. Incising the rectosacral fascia to allow full mobilization to the anorectal ring from above is essential. The surgeon must be comfortably seated for the perineal surgery, which requires considerable elevation of the table with the patient in Trendelenburg. A fiberoptic headlight is essential. The anus is gently dilated, no more than two fingers. The perianal skin, sphincter muscle, and anoderm just below the dentate line are included in the deep traction sutures, which are tied to a 10-cm ring retractor. A ‘Lonestar' or similar device may also be used. Using a 10-cc syringe with a 25-gauge needle, 15 to 20 ml of epinephrine (adrenaline) 1:200 000 is injected in 2-ml aliquots into multiple submucosal sites circumferentially just above the dentate line. To stretch the lower rectal mucosa, the Parks' or similar bivalve retractor is used. The cutting cautery is used to incise the anoderm just cephalad to the dentate line over a width of 10 to 15 mm. Scissor dissection lifts the mucosa for 2 cm until above the anorectal ring. The Parks' retractor is rotated, and the same process continued circumferentially. It is best to start posteriorly and work up both asides of the anus toward the anterior mid-point to avoid having the operative field obscured by blood. The extent of the mucosectomy should be limited, no more than 2 cm, since in such cases the cancer is likely within view.

The mural wall is incised with cautery over a small area anteriorly and posteriorly; a rubber catheter is passed into the pelvis through one opening and out the perineum through the second. The assistant pulls the rectum cephalad, and the perineal operator uses the Parks' retractor and the rubber catheter to divide the lateral rectal wall at the level of the anorectal ring with cautery. By cauterizing directly on the rubber catheter, a total proctectomy is performed, leaving only the anal canal with the internal sphincter. A frozen section may be helpful if clearance of the cancer is in doubt.

A single-layer endoanal anastomosis is made with interrupted, absorbable sutures. Each bite includes full-thickness colon, a portion of the internal sphincter, and anoderm; this should result in an anastomosis 1.5 to 2.0 cm above the anal verge, important in avoiding mucous discharge. To minimize the risk of ‘side-to-side' healing, a roll of Vaseline gauze is placed into the anal canal for several days. All of these patients should undergo temporary fecal diversion for 6 to 8 weeks. If a colostomy rather than an ileostomy is used, care must be taken not to injure the marginal artery either during the formation or the closure of the colostomy.

Proximal diversion

Anastomotic leakage after low colorectal reconstruction that leads to a pelvic abscess and peritonitis can be fatal. Proximal diversion with loop ileostomy or transverse colostomy may reduce the morbidity and mortality of a leak. However, leaks occur in only approximately 5 per cent of low reconstructions and a policy of selective diversion would seem appropriate. Intraoperative contamination, poor bowel preparation, a technically problematic anastomosis, and pinhole air bubbles on insufflation are all considerations for diversion. Preoperative radiation therapy by itself does not justify diversion. Patients with peranal-sutured coloanal reconstructions should undergo diversion, since initial mucous and fecal leakage with burning of perineal skin is unavoidable.

Functional results after low anterior resection with nerve preservation

Sexual function

Conventional resection for rectal cancer in men is associated with postoperative erectile impotence and/or retrograde ejaculation in 25 to 75 per cent of cases. Postoperative sexual dysfunction is due to injury to the pelvic autonomic nerves and can be correlated with the extent of lateral pelvic dissection. Impotence after nerve-sparing dissections has been reduced to 10 to 30 per cent, and is primarily age-related. Sexual function adequate for intercourse, with orgasm and ejaculation, can be obtained in 90 per cent of men of less than 60 years of age. The long-term impact of adjuvant radiation therapy on sexual function has not been defined.

Bowel function

Physiologic studies of patients treated by low anterior resection and colorectal or coloanal anastomosis have demonstrated that anal sphincter tone, sensation of rectal distension, and local reflexes regulating anal sphincter tone are generally preserved. Hence, fecal continence is maintained in the majority of patients. Despite general continence, urgency, frequency and irregular evacuation are common. ‘Cluster' bowel movements are somewhat bothersome, with episodic multiple movements over several hours several times a week. Impaired long-term function has been associated with low level of anastomosis, anastomotic leakage or stricture, and the use of postoperative pelvic radiotherapy.

Rectal reconstruction using a colon J-pouch to increase the capacity of the neorectum appears to be beneficial with coloanal or very low colorectal reconstructions, particularly if the caliber and compliance of the colon used for the anastomosis are poor. Overall, long-term bowel function continues to improve over several years. Dietary fiber supplements and selective use of antispasmodics improve the quality of life for patients that have such dysfunction.

Abdominoperineal resection

Many patients with cancers in the distal third of the rectum, and in particular those with poor preoperative anal function, still may require an abdominoperineal resection. The original operation described by Miles required sequential positioning. The rectum is mobilized down to the pelvic floor, the bowel is divided, and the distal sigmoid brought out as a colostomy in the left iliac fossa (Fig. 30). The distal colon is oversewn and folded down into the pelvis before closing the pelvic peritoneum over the rectum. The abdominal wound is then closed and the colostomy sutured. The patient is then turned on to their left side or placed in lithotomy for the perineal dissection. An elliptical perianal incision is made to mobilize and deliver the anus and distal rectum. The most common error in the perineal dissection is to dissect too close to the rectum, and therefore to the tumor, defeating the prime objective of controlling the local disease. To avoid this mistake the initial dissection is carried well into the ischiorectal fossa, dividing the pudendal vessels supplying the anus before dividing the levator ani muscles laterally. The tip of the coccyx may also have to be removed to give adequate clearance of posterior cancers. Anteriorly, the dissection in males is constrained by the urethra and prostate. A large urethral catheter facilitates identification and protection of the urethra. Mobilization is much easier in females, for the back wall of the vagina can be removed with the rectum, providing a particularly useful maneuver for clearing anterior tumors.

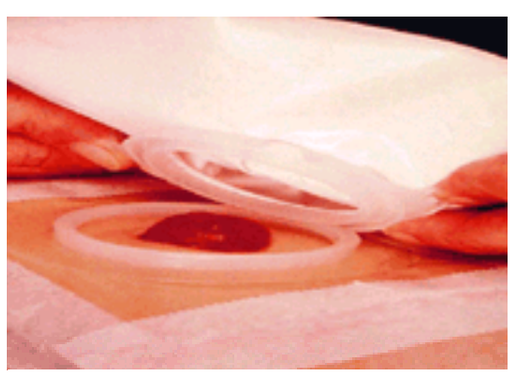


Fig. 30. Example of an end-sigmoid colostomy: appropriate location is through the rectus muscle on a convex portion of the abdomen (best determined preoperatively with the patient sitting).

Abdominoperineal resection is now more commonly done with the patient in the modified lithotomy Trendelenburg position and two surgeons operating simultaneously, or a single surgeon sequentially; this facilitates the difficult perineal dissection and avoids having to transect the rectum in the two-stage procedure. The anus should be carefully sutured with a double purse-string to avoid intraoperative contamination. Following clearance of the pudendal vessels and ischiorectal fossa, placement of a deep retractor (Beckman or Lace) facilitates the completion of the procedure. A major feature of the synchronous approach is the increased safety in defining the posterior and anterior planes of dissection. The posterior pelvis should be entered first. An assistant should remove any pelvic packs, place a hand on the mesorectum, and lift anteriorly; this avoids inadvertent perforation of the rectum. The levators posteriorly and laterally may then be safely divided with cautery. The anterior dissection then clears the superficial transverse perineal muscle. At this point, the assistant places a hand anterior to the specimen, with a finger along the distal prostate and the urethra, which will guide the perineal surgeon to enter safely the pelvis anteriorly. The anterolateral pedicles, which include the puborectalis sling, contain major blood vessels and should be clamped, divided, and suture ligated.

Management of the perineum and the possibility of small-bowel prolapse into the empty pelvis must be dealt with by the operating surgeon. In the absence of fecal contamination the perineum can be closed primarily in almost all patients. The levators should have been widely excised, so cannot be approximated. The ischiorectal fat can be approximated, followed by absorbable vertical mattress sutures to the perineal skin. In non-irradiated patients the pelvis can be left open and packed, with rapid granulation and secondary closure within 2 months. Irradiated patients should all be closed primarily if at all possible. Occasionally, difficulty in obtaining hemostasis requires packing.

In patients who received preoperative radiation therapy the bowel may be allowed to prolapse into the pelvis without consequence. However, in the remaining patients an effort to exclude the small bowel from the pelvis is warranted, although not always feasible, in case the microscopic findings dictate a benefit for adjuvant radiation. In women the uterus may be retroverted into the pelvis with a single figure-of-eight suture to the mid-sacrum. In men and some women the peritoneum may be reapproximated; this is usually ‘easier said than done' and the very thin peritoneum may leave small defects, leading to a subsequent herniation and obstruction of the small bowel. An omental pedicle flap is ideal, but not always feasible. Absorbable mesh may be sutured with fine silk over the true pelvic inlet. It should be stressed that the goal is not to exclude the entire small bowel from radiation therapy, but allow the dose through the ‘PA fields' (see later) to be minimized. Additional radiation given through lateral fields will exclude the small bowel if mesh or omentum keeps the bowel out of the true deep pelvis.

Adjuvant therapy of rectal cancer

End-results with surgical resection

After local excision

The potential benefits of local excision for rectal cancer include reduced perioperative complications and preservation of anorectal, bladder, and sexual function. Data from nine published series indicate that when strict selection criteria are observed, local excision can also provide acceptable oncologic results. In these series nearly all carcinomas selected for local excision were T1 or T2. The local failure rate ranged from 0 to 26 per cent, with an overall rate of 20 per cent. The overall survival rate was 73 per cent. Approximately half of all local recurrences are amenable to salvage surgery, usually by abdominoperineal resection. In a review of 16 published series, Graham and colleagues have shown that a positive surgical margin, poor histologic differentiation, and transmural (T3) extension are the major risk factors for local recurrence. A recent study demonstrated similar 5-year actuarial recurrence-free survivals (87 and 91 per cent) and tumor control (96 and 91 per cent) for patients with favorable (well or moderately differentiated without venous/lymphatic invasion) T1 and T2 rectal cancers following a local excision or abdominal perineal resection.

Since T2 and T3 rectal cancers have an increased risk of lymph-node metastases, it has been proposed that such lesions in patients not able to tolerate a rectal resection may be managed in a sphincter-saving manner by combining local excision with postoperative chemo-/radiation therapy.

After radical resection

Four large retrospective series utilizing conventional low anterior or abdominoperineal resection prior to the widespread use of adjuvant therapy report an approximately 50 per cent 5-year survival. The overall local recurrence rate was 30 per cent, with the pelvis being the most common site of relapse. Risk of pelvic recurrence increased (from 10 to 20–35 per cent) while the probability of survival decreased (from 80 to 60–35 per cent) with increasing stage of disease (stages I, II and III, respectively). Experience from many centers has demonstrated that sphincter-preserving resections for mid and selected distal rectal cancers provide excellent results, equivalent to those achieved with abdominoperineal resection. Consequently, low anterior resection has replaced abdominoperineal resection as the principal operation for rectal cancer.

Baseline data on local/pelvic control as defined from the information presented above is at odds with results reported by experienced surgeons performing optimal pelvic dissections (as discussed above under total/sharp mesorectal excision). The adjuvant trials that will be presented here used the traditional end-results' data as a justification. The integration of more optimal pelvic surgery with selective radiation is currently under trial and will be discussed later in this chapter.

Adjuvant therapy with local excision

Several institutions have reported the results of local excision followed by postoperative radiation therapy (Table 11). Local recurrences ranged from 0 to 20 per cent, with overall survival rates from 70 to 100 per cent. Since the incidence of local failure was directly related to the T stage (T1 = 3 per cent, T2 = 10 per cent, T3 = 24 per cent), T stage is considered amongst the most reliable variables used to predict local failure and disease-free survival following local excision. After transanal excision of a T1 tumor with adverse pathologic features, or of a T2 tumor, postoperative radiation therapy is recommended. Concurrent 5-fluorouracil is commonly used with the radiation.

	T1	T2	T1/T2
New England Deaconess			4/48
Massachusetts General	0/10	2/11	
MD Anderson	0/16	1/15	
Memorial Sloan-Kettering	0/4	2/12	

*5-Fluorouracil used in some patients.

Table 11 Local failure after local excision and radiation therapy*

Adjuvant therapy and radical surgery

Radiation–surgery sequencing

Adjuvant radiation therapy has been utilized preoperatively, postoperatively, and both (sandwich). Preoperative therapy has the advantage of increased efficacy in a well-vascularized tumor bed, perhaps less acute toxicity, and the ability to resect and remove irradiated tissues. Tumor shrinkage may permit sphincter salvage rather than abdominoperineal resection by allowing local excision or better lateral clearance of low-lying tumors, with low stapled reconstruction or a sutured coloanal procedure. Postoperative radiation therapy allows improved patient selection: avoiding the early cancers at lowest risk, allowing total mesorectal excision surgery alone

for medium risk, and selecting patients who are high risk, node positive or have close radial margins for postoperative chemoradiation. In addition, patients with metastatic disease in the liver found at laparotomy will be excluded from radiation therapy. With the use of endorectal ultrasound to stage rectal cancer and high-quality CT scans to assess the liver, selection issues for preoperative therapy are not as problematic as in the past.

Adjuvant preoperative radiation

There have been many randomized trials of preoperative radiation therapy (without chemotherapy) for resectable rectal cancer; they are summarized in Table 12. Caveats in interpreting these data involve understanding the radiation dose and patient selection. In general, the eligibility criteria in these trials allowed nearly all patients to be entered. Contemporary studies would limit entry to patients with a negative CT or hepatic ultrasound for distant disease, and an endorectal ultrasound confirming local transmural disease. Secondly, the radiation fraction size varies greatly. The use of 2000 rads in 200 rad per fraction is a relatively low-dose treatment, unlikely to be beneficial. The use of 2500 rads in 500 rad per fraction is the biologic equivalent of at least 4000 rads in 180- to 200-rad fractions. Some studies show a decrease in local failure, and in four of the more recent reports (Stockholm I, Stockholm II, Swedish Rectal Cancer Trial, and the British Rectal Cancer Group) this difference reached statistical significance. The Stockholm I trial (Stockholm-Malmo) showed a significant advantage in disease-free survival. The Stockholm II trial revealed an advantage in survival in some subsets of patients. The most recent report from a repeat Swedish multicenter trial records a reduction in local failure as well as improved overall survival.

Series	*Radiotherapy dose*/fraction a.c.r*	*Local recurrence (%)*	*Overall 5-year survival (%)*
Princess Margaret	500/500	NA	35
	Surgery only	NA	35
British MRC	500/500	45	55
	Surgery only	43	51
British MRC	5000/300	47	53
	Surgery only	45	57
EORTC	3450/230	22	61
	Surgery only	36	62
Stockholm	2500/500	14	36
	Surgery only	28	36
Swedish trial	2500/500	11	58
	Surgery only	27	48

*Rads or cGY.
NA, not available.
EORTC, European Organization for Research on Treatment of Cancer; MRC, Medical Research Council; VA, Veterans Administration.

Table 12 Randomized trials of preoperative adjuvant radiation for rectal cancer

Adjuvant postoperative radiation

Three randomized trials have evaluated the use of adjuvant postoperative radiation therapy in stage T3–4N1–2M0 (B2, C) rectal cancer in which there was a postoperative radiation-alone arm and a surgery control arm; these are shown in Table 13. In the series from Odense University, 494 patients were randomized to postoperative radiation therapy (4500–5000 cGy) compared to surgery alone. In patients with stage T2–3N0 disease there was no difference in the incidence (6 per cent) or mean time to local failure between the arms. In patients with stage T1–3N1–2 disease there was no difference in the incidence of local failure (6 and 9 per cent) but the mean time to local failure was significantly longer in patients who received radiation therapy compared with the surgical control arm (19 months and 6 months, respectively; $p = 0.01$). In the Gastrointestinal Tumor Study Group series, 58 patients underwent surgery alone and 50 received postoperative radiation therapy (4000–4800 cGy); there were no significant differences in either local failure or survival between these two arms. The only randomized trial to suggest improved local control with postoperative adjuvant radiation alone (without chemotherapy) is from the National Surgical Adjuvant Breast and Bowel Project, and this reached only marginal statistical significance.

Series	*Radiotherapy dose (cGy)*	*Local recurrence (%)*	*5-year survival (%)*
Odense	4500–5000	0% node-negative	NA
		0% node-positive	
	Surgery alone	0% node-negative	
		9% node-positive	
Gastrointestinal Tumor Study Group	4000–4800 Surgery alone	20, 24	44, 41
National Surgical Adjuvant Breast and Bowel Project	4500 Surgery alone	16, 25	41, 43

Table 13 Randomized trials of postoperative adjuvant radiation for rectal cancer

Preoperative or postoperative radiation?

The only randomized trial of preoperative compared to postoperative radiation therapy for resectable rectal cancer was reported by Pahlman and Glimelius; it is a very important and well-performed clinical trial. In this multicenter, randomized study from Sweden, referred to as the Uppsala Trial, 471 patients were randomized to receive either 2550 cGy preoperatively (in 1 week) or 6000 cGy (split course) postoperatively. Postoperative radiation therapy was limited to patients with stage T3 and/or N1–2 disease. Those with stage T1–2N0 disease who were randomized to the postoperative radiation arm did not receive radiation and were observed. Patients who received preoperative radiation therapy had a significant decrease in local failure (13 as compared 22 per cent; $p = 0.02$) but there was no difference in 5-year survival (42 and 38 per cent). Although there was no increase in immediate radiation-related complications or postoperative mortality, there was a significant increase of perineal wound sepsis in the preoperative group (33 compared to 18 per cent; $p < 0.01$). As in other randomized trials of preoperative radiation therapy, the excessively large fraction size (510 rad/day) may have contributed to this complication. Despite the increased incidence of acute toxicity, the long-term toxicity was lower in patients receiving preoperative radiation therapy. The incidence of small-bowel obstruction was 5 per cent in patients receiving preoperative radiation therapy and 11 per cent in patients receiving postoperative radiation therapy.

Adjuvant chemoradiation

Chemotherapy and radiation therapy may be combined both preoperatively and postoperatively. Combined-modality therapy is the term most commonly used. There is only one randomized trial of preoperative combined-modality therapy. The European Organization for Research on Treatment of Cancer was unable to show a benefit with combined treatment: their study of preoperative adjuvant therapies compared irradiation (3450 cGy in 230-cGy fractions) to the same radiation regimen combined with an intravenous bolus only of 5-fluorouracil (375 mg/m^2) on the first 4 days of irradiation. Among the 247 patients followed, overall survival was better with radiation therapy only.

There are several randomized trials of postoperative combined-modality therapy that suggest local control and overall survival are maximized by such treatment in patients following resection of transmural and/or node-positive cancers; these are summarized in Table 14. Not all such trials have had surgery-alone control arms. Lastly, the figures for baseline local failure are far larger than in single-institution reports with more optimal surgical technique.

Series	Regimen	Local failure (%)	5-year survival (%)
Gastrointestinal Tumor Study Group	5FU/MeCCNU/XRT Surgery only	12/43	59/43
North Central Cancer	XRT/5FU/MeCCNU	13	
Treatment Group	XRT	25	
North Central Cancer	XRT/5FU bolus		
Treatment Group	XRT/5FU infusion		38/6
Intergroup 0114	XRT/5FU	9–13 m	Nothing
	XRT/5FU/Arm reserve	all	better than
	XRT/5FU/Arm reserve	groups	5FU/XRT
	XRT/5FU/Arm reserve/levamisole		

5FU, 5-fluorouracil; MeCCNU, methyl-CCNU; XRT, radiotherapy

Table 14 Postoperative adjuvant chemoradiation for rectal cancer

Gastrointestinal Tumor Study Group study 7175 was begun in 1975 and allocated patients with completely resected stage B2 or C tumors to one of four treatment groups: surgery only, chemotherapy for 18 months, radiation therapy employing 4000 to 4800 cGy to the pelvis, compared to radiation therapy using 4000 to 4400 cGy with concomitant 5-fluorouracil (500 mg/m^2) on the first three and last three days of radiation plus subsequent 5-fluorouracil and methyl-CCNU (the same as in the second group). At an 80-month median follow-up the control population had a 55 per cent recurrence rate, but the rate was 33 per cent for the radiation therapy plus chemotherapy group ($p < 0.009$). There were 14 local recurrences in the 32 control patients and five in the 46 patients treated by combination therapy. A subsequent analysis at a median follow-up of 94 months showed an even larger margin of benefit for the combination therapy. In addition to a disease-free survival advantage, combination chemoradiation therapy was associated with a statistically significant overall survival benefit ($p = 0.005$). There was an approximately 20 per cent superiority in survival rates at 6 years for the 96 patients at risk. Although these data support an aggressive multimodality approach to patients with rectal cancer, the morbidity of combined radiation therapy was considerable. Severe or life-threatening acute toxic effects occurred in 18 per cent of patients in the chemoradiation arm. There were three late deaths, two from enteritis in the combined treatment group and one from acute nonlymphocytic leukemia in patients treated with chemotherapy alone.

Another major trial of adjuvant therapy for rectal cancer was initiated by the North Central Cancer Treatment Group (study 79–47–51). A total of 204 patients were randomly assigned to postoperative radiation therapy only (4500 cGy with a 540-cGy boost) or to an integrated program of methyl-CCNU/5-fluorouracil. At a median follow-up of more than 7 years there were 62 recurrences in the group treated with radiation alone and only 40 among those treated with combination therapy ($p < 0.0025$). Even after adjustment for known prognostic factors, the risk of relapse for patients receiving combination treatment was reduced by 47 per cent compared with postoperative radiation therapy alone. The local failure rate was 13.5 per cent in the combination treatment group and 25 per cent in those treated with postoperative irradiation only. An overall survival advantage was demonstrable for the combination therapy. The risk of cancer-related death was reduced by 36 per cent ($p = 0.0071$) and the overall death rate by 29 per cent ($p = 0.025$). The therapy was well tolerated and long-term toxicity was not apparent.

The largest study was performed by the National Surgical Adjuvant Breast and Bowel Project (R01). A total of 555 patients with B2 or C lesions were randomized to one of three arms: postoperative observation; postoperative pelvic radiation therapy of 4700 cGy with a boost to 5300 cGy maximum; or chemotherapy with methyl-CCNU, vincristine, and 5-fluorouracil (**MOF**). IN a 64-month mean follow-up time there was a statistically significant disease-free survival advantage for MOF chemotherapy ($p = 0.05$) and an overall survival improvement for selected subsets of patients receiving MOF, particularly men and patients younger than 65 years.

These three studies permit the conclusion that combined radiation therapy and chemotherapy is superior to surgery alone, but the exact role of irradiation and the optimal protocol for chemotherapy are not defined. These studies led to the United States National Institutes of Health Consensus Conference statement indicating that patients with transmural or node-positive rectal cancer should be offered postoperative adjuvant chemoradiation therapy.

Two subsequent studies addressed the relative value of the addition of methyl-CCNU to 5-fluorouracil as part of a combined-modality irradiation plus chemotherapy treatment. The Gastrointestinal Tumor Study Group study 7180 evaluated 210 patients. In the 3-year follow-up data there was no evident superiority in recurrence or survival rates for the addition of methyl-CCNU. The North Central Cancer Treatment Group study 86–47–51 compared bolus to infusional 5-fluorouracil. All patients received pelvic radiation therapy and concomitant 5-fluorouracil in the sandwich sequence of chemotherapy–irradiation chemotherapy. Half the patients received methyl-CCNU and 5-fluorouracil chemotherapy as in the previous North Central Cancer Treatment Group trial (79–47–51) and the other half received 5-fluorouracil only. With a median follow-up of 46 months, patients who received a protracted infusion of 5-fluorouracil during radiation had a significantly increased disease-free survival compared to those receiving bolus 5-fluorouracil during radiation. The addition of methyl CCNU to 5-fluorouracil did not result in any improvement in survival.

Recent trials of combined-modality therapy

The are two recent cooperative group randomized trials in the United States. The National Surgical Adjuvant Breast and Bowel Project R02 study compared its standard MOF chemotherapy regimen with a 5-fluorouracil/leucovorin regimen. Half the patients received radiation therapy in conjunction with chemotherapy. Because of preliminary evidence of qualitative therapeutic interactions in specific subgroups, women received 5-fluorouracil/leucovorin with or without radiation therapy and men were randomly assigned to the four options. The large Intergroup Study (INT-0114) was initiated in 1990 to evaluate more than 1300 patients with rectal cancer. All patients underwent complete resection and then received one cycle of chemotherapy followed by 5000 to 5400 cGy of pelvic irradiation with chemotherapy and an additional cycle of therapy. All patients received 5-fluorouracil alone or with leucovorin or levamisole or both. The data do not demonstrate improved efficacy with the three-drug combination concurrent with radiation therapy. At present, postoperative chemoradiation therapy with infusional 5-fluorouracil alone or combined with leucovorin are both acceptable.

Adjuvant therapy with total mesorectal excision

The incremental benefit of adjuvant radiation or chemoradiation therapy in the setting of more optimal surgical technique is the focus of several current or proposed clinical trials. A multicenter trial in The Netherlands is evaluating by random assignment the efficacy of short-course/high fraction-size radiation therapy in the presence of total mesorectal excision and pelvic dissection. A study under development through the American College of Surgeons will assess the incremental benefit of postoperative chemoradiation therapy after total mesorectal excision for T3N0M0 (B2) cancers.

Adjuvant therapy of locally advanced rectal cancer

Patients with fixed or tethered rectal tumors who undergo initial resection will generally have gross or microscopic residual cancer. Anterior adhesion can be dealt with by extending the scope of the operation to include posterior vaginectomy, posterior exenteration (in women), or total pelvic exenteration. More problematic are cases with adhesion or invasion of the pelvic side-wall or sacrum. These patients should receive preoperative radiation therapy (generally with infusional 5-fluorouracil or bolus 5-fluorouracil/leucovorin), with operative resection 5 to 8 weeks after the completion of such treatments. Although this approach produces tumor clearance in 75 per cent of patients, local recurrence remains high at 30 per cent.

Intraoperative radiation therapy appears to be beneficial in such cases. Reports from three centers in the United States provide identical data. The Massachusetts General Hospital and the Mayo Clinic combine external-beam radiation therapy with resection and an intraoperative electron-beam boost. The Memorial Sloan-Kettering Cancer Center uses high dose-rate, iridium-192, intraoperative radiation therapy. Data from all three centers suggest that this additional radiation reduces the local recurrence rate to 10 per cent; there are no applicable randomized data. The intraoperative electron-beam approach is complex and rather expensive. The high dose-rate system is simple to use and modest in cost, with increasing interest in the United States and Europe.

Recommendations for adjuvant therapy

Patients with high-risk colon cancer, like other cancers, should be considered for entry into clinical trials. Off-protocol, the highest-risk patients should be evaluated for preoperative combined-modality therapy. These are patients with fixed or tethered cancers, and those with gross transmural cancers. Endorectal ultrasound, CT, and digital rectal examination help in defining such patients. Optimal surgical resection of circumferential, low-lying cancers in husky males is very problematic. In patients undergoing primary local excision who have muscle invasion or lymphovascular permeation, adjuvant radiation or chemoradiation is recommended. Patients undergoing primary radical resection who have pathologic indications of high risk should be evaluated for postoperative combined-modality treatment. Even those patients treated by total mesorectal excision who have close or positive radial margins, or positive lymph nodes, are at high risk for local and distant recurrence and should also be evaluated for chemoradiation.

Follow-up (Table 15)

Strategy	Goal	Value
Interval history and physical examination	Detect asymptomatic recurrences	Rare suture-line recurrence
	Advise on bowel dysfunction	Helpful
	Advise on colostomy care	Helpful
	Reassurance	Variably reassures some; in others increases anxiety
Sigmoidoscopy	Detect recurrence	Helpful, in identifying anastomotic strictures and suture-line recurrence
Colonoscopy	Serve polyps or cancers	Helpful, every 3 years
Mammography	Screen for breast cancer	Helpful, yearly
Chest radiograph	Detect lung cancer or metastases	Variable
Carcinoembryonic antigen	Screen for liver metastases	Possibly helpful
Liver function tests	Screen for liver metastases	Not helpful
CT/ultrasound	Screen for metastases or new primary	Variable

Table 15 Postoperative follow-up after curative resection

The major purpose of follow-up after cancer treatment has been the detection of asymptomatic recurrence. Since this rarely translates into improved survival, other considerations are important. Maximizing the quality of life by management of treatment-related problems is extremely important in patients with colorectal disease. The psychologic benefits of a ‘clear' check-up seem obvious. However, in some patients, follow-up visits produce more anxiety, not less.

Dietary modifications may be necessary for control of bowel function. Initially, a low-roughage diet may be helpful. After sphincter-saving rectal surgery, fecal urgency and frequency may be lessened with stool-bulking agents. Antispasmodic drugs may be required. Ultimately, a low-fat, high-fiber diet should be recommended. Anastomotic strictures may require dilation or division by cautery. Small-bowel obstruction, a late-occurring problem related to surgical scarring or radiation enteritis, necessitates surgery in 5 per cent of patients. The need for continued advice about colostomy management, irrigation techniques, and prevention of skin irritation may necessitate consultation with an enterostomal therapist.

Surgical castration or pelvic irradiation in premenopausal women lead to menopause, with vaginal dryness with dyspareunia and general symptoms. Local or systemic estrogens may be helpful. Male impotence due to psychologic and organic factors may require intervention. Papaverine injections, oral Viagra, urethral prostaglandin drugs, and surgical implants can improve erectile function in impotent men.

The most important aspect of follow-up is ensure that the patient is undergoing regular follow-up colonoscopy to identify and remove additional adenomas. As discussed in the section on screening, it is appropriate that siblings and children of patients with colorectal cancer should be counseled on appropriate prevention and screening strategies. The patient and family members should be actively counseled about stopping smoking, regular mammography, prostate assessment, and other cancer and general health issues.

Patients with a resected rectal cancer should have regular proctoscopy and pelvic examination. This inexpensive part of the routine physical assessment occasionally provides an opportunity for cure of an early suture-line or pelvic recurrence. The most controversial aspect of follow-up concerns the detection of asymptomatic recurrent cancer, and specifically the role of serial estimations of serum carcinoembryonic antigen. Considerable data support the use of this serial testing in the early detection of recurrence, but attempts to show that this translates into an overall survival benefit have failed. The British randomized trial, the Mayo retrospective analysis, and an analysis from the Memorial Sloan-Kettering Cancer Center suggest that any survival advantage is small or nil. However, carcinoembryonic antigen is elevated in 90 per cent of patients with hepatic metastases. As discussed below, liver resection can cure a subset of such patients. ‘Symptomatic' liver metastases are almost never resectable. Hence, the use of serial assays for carcinoembryonic antigen in follow-up is unlikely to be ‘cost-effective', yet is unequivocally effective in selecting patients for follow-up ultrasound or CT scans of the liver. Whether this approach is justifiable to help one to two patients out of each 100 is not easily resolved. Issues of patient age and general health, as well as the risk of developing recurrence, should be included in the assessment of a follow-up program.

Surgical treatment of locally recurrent rectal cancer

Recurrence after limited local therapy

After treatment of rectal cancer by local excision, intracavitary radiation, or fulguration, some patients with limited local recurrences can be cured by radical surgery. Low anterior or abdominoperineal resections can be performed for patients without distant disease, with a salvage rate of 25 to 50 per cent. If not used previously, chemoradiation therapy as an ‘adjuvant' either before or after such resections is appropriate.

Suture-line recurrences

Limited suture-line recurrence in a patient with metastatic cancer may not require treatment. Prophylactic colostomy should be avoided, with partial obstruction managed by dietary restriction and stool softeners. The neodymium:yttrium aluminum garnet laser can be used to keep the lumen patent. In patients without distant metastases, long-term disease-free survival can be achieved with abdominoperineal resection, but with only an expected 25 per cent 5-year survival. Once again, if radiation therapy has not been previously used, ‘adjuvant' chemoradiation should be added to resection.

Recurrence after radical surgery

Although isolated, resectable local recurrences are encountered, most patients with pelvic ‘failure' have diffuse locoregional recurrences. It is important to differentiate between pelvic and perineal recurrences after abdominoperineal resection. Isolated perineal recurrences can be palliated by wide local excision, but the disease is rarely cured.

Patients with central pelvic recurrences, particularly those with minimal symptoms, are amenable to salvage pelvic exenteration. Urinary reconstruction with an ileal conduit or using the right colon for an ‘Indiana pouch' is feasible. In patients with posterior pelvic recurrences, near or adherent to the sacrum, even more radical surgery is required. Abdominoperineal resection with in-continuity sacrectomy or pelvic exenteration with in-continuity sacrectomy (composite or abdominosacral exenteration) may help highly selected, younger patients.

The major experience with these extended operations for recurrent rectal cancer has been reported from three centers: the Memorial Sloan-Kettering Cancer Center, the University of Virginia, and the University of Colorado. Wanebo and Marcove first reported the inte-gration of the technique of primary sacral resection through the posterior approach with transabdominal visceral exenteration. The two-part approach begins with a thorough abdominal exploration to rule out extrapelvic cancer. The hypogastric vessels are preferentially ligated distal to the superior gluteal branch. The proximal line of sacral transection is determined, and the rectum or rectum and bladder are then mobilized. After complete anterior and partial lateral dissection, the abdomen is closed, and the surgery continues with the patient in the prone position. The sciatic nerves are identified and preserved, and the pelvis is entered. A laminectomy exposes and preserves the uppermost nerve roots, and the sacrum is transected with removal in continuity of visceral contents, pelvic side-wall soft tissue, and the sacrum. Wanebo has now reported on over 40 patients, with seemingly good palliation and excellent pain control, and 20 to 25 per cent likelihood of long-term survival. These operative procedures have a formidable morbidity and also mortality from hemorrhage or infection. Pearlman and associates resected 21 patients with recurrent rectal cancer, and 18 had prior radiation. Twelve patients had complete abdominosacral exenteration. Of the 16 patients who had potentially curative surgery, eight were free of recurrence at 6 to 48 months after the procedure.

If tumor is adherent only to the distal sacrum, I have preferred to use a chisel and remove the anterior cortex, which is fairly simple from S4 down, rather than do a full-thickness sacrectomy. The use of argon-beam coagulation provides additional tumor destruction along the periosteum without the need for sacrectomy. Lastly, intraoperative radiation therapy may replace or be added to the bone resections for posterior pelvic recurrences.

Surgical treatment of metastatic colorectal cancer

The palliative treatment of recurrent or metastatic colorectal cancer is beyond the scope of this chapter. However, there are selected sites of recurrence amenable to

surgical intervention that can potentially lead to long-term survival and possible cure. Three patterns of failure amenable to surgical intervention will be reviewed.

Ovarian metastases

As many as 7 per cent of women will recur with ovarian metastases, frequently symptomatic. Issues of prophylactic oophorectomy have been discussed above. Most patients with gross ovarian metastases will have widespread cancer or additional peritoneal metastases. How-ever, a subset of such patients will not have liver or lung metastases, and will have tumor limited to the ovaries with or without concurrent peritoneal disease. The serum concentrations of carcinoembryonic antigen and CA-125 may not be helpful in separating colorectal recurrence from primary ovarian cancer. In the absence of other disease on chest radiography and CT scanning, a laparotomy may be beneficial. The extent of recurrence will be ascertained. If primary ovarian cancer is present, then aggressive debulking followed by systemic chemotherapy will palliate, and possibly cure, the patient. If metastatic colorectal cancer is present on frozen section, then a bilateral salpingo-oophorectomy (without hysterectomy) is performed. Residual peritoneal disease should be resected or fulgurated as feasible. Postoperative chemotherapy may be beneficial, particularly in those women with complete debulking and only microscopic residual disease. General performance status, prior chemotherapeutic treatments, and the extent of residual cancer will influence this decision. I prefer intraperitoneal chemotherapy with floxuridine and leucovorin. In order to be prepared for such complex treatment strategies, all such women should be evaluated preoperatively by surgeons, gynecologists, and medical oncologists in order to agree on a treatment algorithm suitable for the individual patient.

Hepatic metastases

Autopsy data indicate that as many as one-third of patients dying from colorectal cancer have isolated hepatic metastases, which explains why resection is potentially curative in the treatment of patients with colorectal metastases clinically isolated to the liver. Although prospective, randomized clinical trials have not been conducted, retrospective comparative studies provide compelling evidence that liver resection prolongs survival and cures a small subset of patients. In appropriately selected patients, liver resection is associated with a 30 to 40 per cent 5-year survival and a 20 per cent long-term disease-free survival. In experienced centers this can be accomplished with an operative mortality of less than 5 per cent. (Table 16)

Study, year	Number resected	Mortality	5-year survival (%)
Foster, 1978	168	5	20
Adson, 1984	141	2	25
Hughes, 1986	607	185	33
Schlag, 1990	122	4	30
Doci, 1991	100	5	30
Scheele, 1991	219	6	39
Rosen, 1992	280	4	25
Gayowski, 1994	204	0	32
Scheele, 1995	469	4	39
Fong, Cohen, 1996	577	4	38

Table 16 Liver resection for colorectal metastases

Prognostic features influencing long-term results for patients following potentially curative liver resection have been extensively examined. Age, sex, primary tumor grade, and primary tumor location are not major factors. Patients resected after presenting with incidental disease (for example, at laparotomy for the primary tumor or for subsequent bowel obstruction), or following detection of an elevated carcinoembryonic antigen, fare better than those presenting with symptoms. It is important to know the number of metastatic lesions and whether multiple lesions are bilateral only slightly reduces the end-results of liver resection. Surgery should not be limited to those with a single lesion. However, the presence of satellitosis in association with major lesions or positive portal lymph nodes rarely results in long-term survival. If tumor-free margins are obtained by resection of the liver, long-term survival and cure are possible.

Preoperative assessment

Preoperative investigation for resection of metastatic colon cancer is directed at delineating the extent of the liver tumor and to rule out the presence of gross extrahepatic disease. Liver function tests provide minimal information unless associated cirrhosis is a consideration. Jaundiced or malnourished patients, or those with fatty livers, should not be resected. Scans are obtained not only to define the size and location of hepatic metastases but also to place these lesions in the context of feasible hepatic resection. Adherence to the major portal inflow in both the right and left liver or to all major hepatic veins would preclude resection, even if considerable normal liver is apparent.

CT scans give valuable preliminary information. Peripherally placed lesions never involve the hilar structures or vena cava, and, under these circumstances, CT scanning may obviate the necessity for other studies. More centrally situated lesions may involve major vessels close to the hilus and CT scanning with dynamic portography or fast-spin echo MRI are helpful. Scans should assess the hilar, celiac, and peripancreatic nodal drainage, since positive nodes in this location are a major and perhaps absolute contraindication to resection. Doppler ultrasonography can provide considerable cost-effective assessment of the liver, but is highly operator dependent and does not provide nice images for use in the operating theater. The use of hepatic angiography is rarely of benefit to the planning or conduct of hepatic resections.

Extrahepatic disease is evaluated by a series of studies including colonoscopy, study of CT images of the abdomen, by chest radiographs, and by CT of the chest if the radiograph is abnormal. Routine chest CT is not required. Bone scans are not helpful.

Right hepatectomy as a paradigm of liver resection

The portal vein and hepatic artery divide into major right and left branches outside the liver substance below the hilus. For many years, all major liver resections involved an initial dissection of such structures, with ligation and division of appropriate vessels. This is time-consuming and can be rather difficult. Couniaud demonstrated that the branches of the hepatic artery and portal vein together with the accompanying bile ducts form a triad that, as they pass into the liver tissue, takes the continuation of Glisson's capsule with them so that within the liver these structures are contained together in a well-formed sheath. Intrahepatic division by ‘pedicle ligation' greatly expedites such surgery and provides much safer control of vascular inflow and the bile ducts.

The most important aspect of right hepatectomy is full mobilization of the liver from the retroperitoneum and diaphragm. The suprahepatic vena cava is cleared and the right hepatic vein is dissected outside the liver. After this maneuver, short branches from the inferior vena cava to the liver are ligated or clipped and divided. Any liver or capsular tissue around the right lateral inferior vena cava is divided. The right hepatic vein can now be encircled by careful dissection from below and above. Since tearing the vena cava or the hepatic veins at this time can be fatal, if the operating surgeon is having difficulty or is relatively inexperienced, complete encircling of the right vein is not crucial. Enough should be dissected from above to allow placement of a vascular clamp during parenchymal dissection.

The gallbladder is now removed and the right hepatic artery ligated. A noncrushing clamp (Pringle maneuver) controls vascular inflow, and clamping of the right hepatic vein, or preferably, its division with a laparoscopic vascular stapler (Fig. 31) allows parenchymal dissection to proceed in a relatively bloodless field. In the absence of cirrhosis the liver will tolerate the 30 to 45 min necessary for liver transection with continuous Pringle clamping. However, if there is concern about liver function, or if the residual liver is relatively small, the Pringle clamp may be removed every 10 to 20 min for 2 min. The method of parenchymal dissection is a personal preference. The ultrasonic dissector is slow but effective. Most experienced surgeons prefer just suction dissection or clamp crushing; both separate the soft normal parenchyma, leaving vascular and biliary structures exposed for clipping, cautery, or ligation. Once the parenchyma has been divided down to the intrahepatic right ‘sheath', the posterior liver at the caudate lobe is divided. The ‘pedicle' is then divided; again some type of vascular stapler is ideal. Branches from the middle hepatic vein are problematic late in the liver transection, and should be clearly identified and ligated.

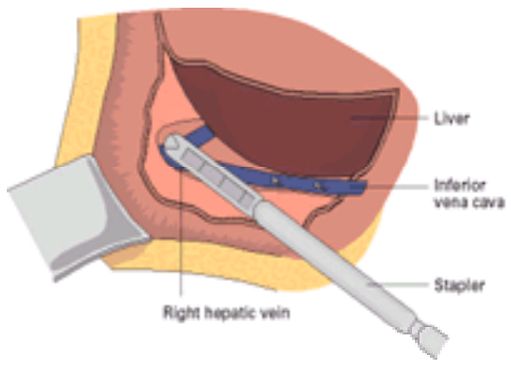


Fig. 31. Division of the right hepatic vein using a laparoscopic vascular stapler.

Following specimen removal, complete hemostasis is essential. Ball-tip cautery or the argon-beam coagulator are used, as well as delicate sutures. ‘Liver sutures' are not to be used for elective liver resections, since they leave considerable ischemic or dead liver within the suture, which leads to infection. In the absence of bile leakage at the completion of the resection, a randomized trial has demonstrated that routine drainage does not reduce the risk of subphrenic abscess or ‘biloma' formation.

Cryosurgery or cryoablation

In situ tumor destruction by rapid freezing of tissues results in crystal formation and produces significant cellular damage and cell death. This cytotoxic effect is the basis of cryoablation as a potential tool in the treatment of cancer. Two technical developments have resulted in renewed interest in cryoablation for liver tumors. First, the development of insulated cryoprobes cooled by liquid nitrogen delivery systems allows precise, rapid-controlled freezing of tumors even at great depths within the liver. Secondly, high-frequency intraoperative ultrasound devices facilitate not only precise placement of the cryoprobes but also accurate delineation of tumor location and freeze margins (Fig. 32, Fig. 33). Studies are now seeking to determine the role of this modality in the treatment of colorectal metastases, but definitive data comparing cryoablation with other treatments are not yet available and the exact role for cryosurgery in surgical treatment of colorectal hepatic metastases remains under investigation.

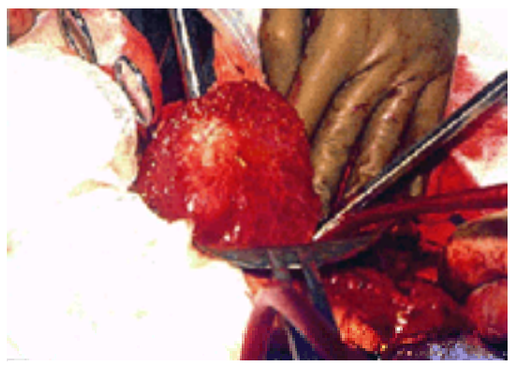


Fig. 32. Hepatic cryosurgery under ultrasound control.

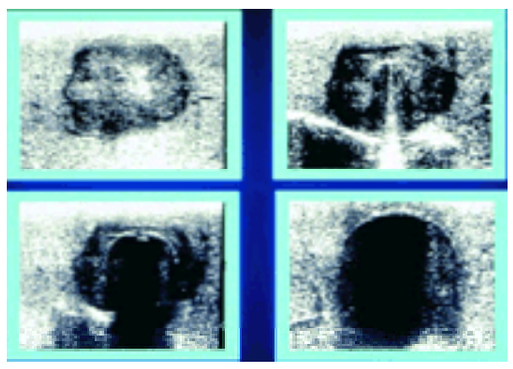


Fig. 33. Cryoprobe with ‘ice-ball'.

Hepatic arterial infusional chemotherapy

The rationale of hepatic arterial infusional chemotherapy is well established and is twofold: first, the predominant blood supply to normal liver is from the portal vein, while 90 to 95 per cent of the supply to liver tumors is from the hepatic artery; secondly, short-acting drugs used for hepatic arterial infusion are rapidly cleared by the normal liver, resulting in several logs of increased drug exposure to hepatic metastases treated by infusion rather than by the systemic/intravenous route.

Many randomized trials have attempted to evaluate and confirm the putative advantages of hepatic arterial infusion in the treatment of isolated but unresectable hepatic metastases. One of the initial and detailed trials was the Memorial Sloan-Kettering Cancer Center study, which compared hepatic arterial infusion with systemic infusion, applying the same chemotherapeutic agent (fluorodeoxyuri-dine), drug schedule, and method of administration (an implanted Infusaid pump). All patients underwent exploratory laparotomy to assess the percentage of liver involvement and to rule out extrahepatic disease. Both groups had hepatic arterial catheters placed, generally through the gastroduodenal artery. Patients randomized to the systemic group had the arterial catheter connected to an access port and the infusion pump connected to an additional catheter placed in the cephalic vein; this allowed a cross-over to intrahepatic therapy by a minor surgical procedure in the event of tumor progression on systemic therapy.

This study demonstrated a significantly higher response rate (greater than 50 per cent reduction in measurable disease) with hepatic therapy, 50 per cent as compared to 20 per cent for systemic infusion ($p = 0.001$). Thirty-one of the systemic patients crossed over to hepatic arterial infusional therapy after tumor progression. Twenty-five per cent had a partial response, 22 per cent stabilization of disease, and 50 per cent a drop in carcinoembryonic antigen on intrahepatic therapy. The toxicity was different between the two groups. In the infused group the most frequent side-effect was elevated hepatic enzyme activities. In four patients (8 per cent) there were changes in the bile ducts that resembled biliary sclerosis; in three of the four the process was reversed by discontinuing therapy. Another frequent side-effect was gastritis or ulcer disease. The most frequent toxicity from systemic therapy was diarrhea, occurring in 70 per cent of patients and requiring admission for intravenous hydration in 9 per cent.

Many other randomized and many more non-randomized trials have been performed. When using hepatic arterial infusion with fluorodeoxyuridine, leucovorin and dexamethasone, response rates in excess of 70 per cent can be obtained. However, a clear-cut improvement in overall survival has not been reproducibly demonstrable despite over 15 years of trials. For that reason a multicenter trial is under way in North America in which patients are randomized to hepatic arterial infusion or systemic treatment, both with leucovorin-modulated fluoropyrimidines. Cross-over from the systemic to the infusion is not allowed, although patients receiving the infusion whose tumors progress outside the liver may receive the systemic regimen.

Placement of an hepatic arterial catheter

All patients should undergo hepatic arterial angiography, with injections of the celiac axis and the superior mesenteric artery; this will detect the variants of ‘standard anatomy', as well as variations of the normal relevant to hepatic catheter placement. At laparotomy, the abdomen should be explored as well as the portal and associated hepatic lymph nodes, as if the operation were a liver resection. As a general rule, extrahepatic disease is a contraindication to hepatic arterial infusional chemotherapy. The hepatic arterial system is dissected and all branches encircled. Branches back to the stomach and duodenum, the right gastric particularly, must be ligated and divided to prevent chemotherapy-induced gastritis. The gastroduodenal artery is generally ligated as part of the catheter-access approach demonstrated in Fig. 34.

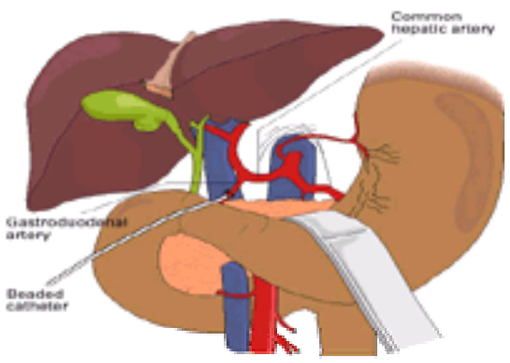


Fig. 34. Hepatic arterial catheter placement for regional chemotherapy in patients with ‘normal anatomy'.

Pulmonary metastases

As with surgical resection of hepatic metastases, patient selection is crucial and leads to bias in end-results' reporting. Patients with short disease-free intervals and synchronous lung and liver metastases do poorly. However, fit patients with limited lung metastases amenable to resection should be evaluated for surgery. A negative physical examination and a recent colonoscopy and negative abdominal/pelvic CT are essential. The number of metastatic lesions has only a relative impact on the end-results of resection. A single metachronous lung lesion, particularly in a cigarette smoker, is often a primarylung cancer; hence all single lesions should be resected. The same applies to a single lesion in association with a second, small, subpleural nodule, which is often just a benign lymph node.

Multiple lung metastases may be amenable to video-assisted thoracoscopy, but this approach will often miss additional lesions. Thoracotomy with wedge or lobar resection is most often required. Bilateral lesions in younger patients may be resected with staged posterolateral thoracotomies, or simultaneously through a transverse or ‘clamshell' incision. Median sternotomy is less complicated, but the exposure of certain parts of the lungs is rather problematic.

Reports from major centers indicate that 5-year survivals of 15 to 25 per cent can be obtained with resection in appropriately selected patients.

Benign colorectal tumors

Carcinoid tumors

These neuroendocrine tumors are most commonly an incidental finding in the appendix or a submucosal nodule in the rectum. Appendiceal carcinoids are ‘midgut' in their origin, and rarely metastasize unless they are 2 cm or greater in diameter. Associated hepatic metastases do result in the ‘carcinoid syndrome'. Simple appendectomy is adequate for tumors of 1 cm or less, and right colectomy for tumors of 2 cm or more. Management of patients with tumors of 1 to 2 cm should be individualized.

Rectal carcinoids are ‘hindgut' in origin. They are rarely associated with the carcinoid syndrome, even in the presence of extensive metastases. Rectal tumors of less than 2 cm are treated by local excision, with cure to be expected. Larger rectal tumors behave like high-grade adenocarcinomas, with aggressive locoregional spread and metastatic behavior.

Rectal sarcomas

These are almost always leiomyosarcoma, with rare cures. The treatment is radical resection, whether colon or rectal site. Despite complete resection, cure in low-grade tumors is less than one-third. High-grade tumors rapidly metastasize. Although usually fatal, low-grade tumors, completely resected, have a median survival of 5 years. High-grade tumors have a median survival of only 1 to 2 years.

Because of the small likelihood of cure, treatment of rectal sarcomas has attempted to avoid abdominoperineal resection. Small tumors may be adequately treated by wide local excision and iridium-192 brachytherapy.

Further reading

General

Cohen AM, Minsky BD, Schilsky RL. Colon cancer. In: Devita VTJ, Hellman S, Rosenberg SA, ed. *Principles and practice of oncology*, 5th edn, pp. 1144–97. Lippincott, Philadelphia, 1997.

Cohen AM, Minsky BD, Friedman M. Rectal cancer. In: Devita VTJ, Hellman S, Rosenberg SA, ed. *Principles and practice of oncology*, 5th edn, pp. 1197–1234. Lippincott, Philadelphia, 1997.

Jessup JM, Menck HR, Fremgen A, Winchester DP. Diagnosing colorectal carcinoma: clinical and molecular approaches. *CA-A Cancer Journal for Clinicians* 1997; **47**(2); 70–89.

Etiology and risk factors

Hardcastle JD. Colorectal cancer. *CA-A Cancer Journal for Clinicians* 1997; **47**(2); 66–9.

Nugent KP, Spigelman AD, Phillips RKS. Risk of extracolonic cancer in familial adenomatous polyposis. *British Journal of Surgery* 1996; **83**: 1121–2.

Willet WC. Cancer prevention: diet and risk reduction. In: Devita VT, Hellman S, Rosenberg SA, eds. *Cancer: principles and practice of oncology*, 5th edn, pp. 559–607, 1997.

Winawer SJ *et al.* Risk of colorectal cancer in the families of patients with adenomatous polyps. *New England Journal of Medicine* 1996; **334**: 82–7.

Molecular biology

Cho KR, Vogelstein B. Genetic alterations in the adenoma–carcinoma sequence. *Cancer* 1992; **70**: 1727–31.

Giardiello F *et al.* Phenotype expression of disease in families that have mutations in the 5' region of the adenomatous polyposis. *Annals of Internal Medicine* 1997; **126**: 514–19.

Giardiello F *et al.* The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *New England Journal of Medicine* 1997; **336**: 823–7.

Harris CC, Hollstein M. Clinical implications of the *p53* tumor-suppressor gene. *New England Journal of Medicine* 1993; **329**: 1318–27.

Lynch HT, Smyrk T. Hereditary nonpolyposis colorectal cancer (Lynch syndrome)—an updated review. *Cancer* 1996; **78**: 1149–67.

Lynch HT, Smyrk T, Lynch JF. Overview of natural history, pathology, molecular genetics and management of HNPCC (Lynch syndrome). *International Journal of Cancer* 1996; **69**: 38–43.

Powell SM *et al.* Molecular diagnosis of familial adenomatous polyposis. *New England Journal of Medicine* 1993; **329**: 1982–7.

Rosen N. Molecular biology of gastrointestinal cancers. In: Devita VT, Hellman S, Rosenberg SA, eds. *Cancer: principles and practice of oncology*, 5th edn, pp. 971–9. Lippincott, Philadelphia, 1997.

Yao J, Goh HS, Smith DR. p53 mutations in primary colorectal adenocarcinomas and liver metastases. *British Journal of Surgery* 1996; **83**: 1245–6.

Screening

Hung LM HJ, Armitage NC. Colonoscopy for all first-degree relatives of patients with colorectal cancer. *British Journal of Surgery* 1995; **82**: 1644.

Lang CA, Ransohoff DF. On the sensitivity of fecal occult blood test screening for colorectal cancer. *Journal of the National Cancer Institute* 1997;**89**: 1392–3.

Weinberg DS, Strom BL. Screening for colon cancer: a review of current and future strategies. *Seminars in Oncology* 1995; **22**: 433–47.

Winawer S *et al.* Colorectal cancer screening: clinical guidelines and rationale. *Gastroenterology* 1997; **112**: 594–642.

Prognosis and staging

Deans GT *et al.* Jass' classification revisited. *Journal of the American College of Surgeons* 1994; **179**: 11–17.

Fleming ID *et al. Colon and rectum* (AJCC cancer staging manual), 5th edn, pp. 83–90. Lippincot-Raven, Philadelphia, 1997.

Sobin LH, Wittekind C. Colon and rectum. In: *TNM classification of malignant tumours*, 5th edn, pp. 66–9. Wiley, New York, 1997.

Preoperative assessment

Anderson BO, Hann LE, Enker WE, Dershaw DD, Guillem JG, Cohen AM. Transrectal ultrasonography and operative selection for early carcinoma of the rectum. *Journal of the American College of Surgeons* 1994; **179**: 513–17.

Kerner BA, Oliver GC, Eisenstat TE, Rubin RJ, Salvati EP. Is preoperative computerized tomography useful in assessing patients with colorectal carcinoma? *Diseases of the Colon and Rectum* 1993; **36**: 1050–3.

Phang PT, Wong WD. The use of endoluminal ultrasound for malignant and benign anorectal diseases. *Current Opinion in Gastroenterology* 1997; **13**: 47–53.

Thaler W *et al.* Preoperative staging of rectal cancer by endoluminal ultrasound vs. magnetic resonance imaging: preliminary results of a prospective, comparative study. *Diseases of the Colon and Rectum* 1994; **37**: 1189–93.

Polyps

Gordon MS, Cohen AM. Management of invasive carcinoma in pedunculated colorectal polyps. *Oncology* 1989; **3**: 99–105.

Markowitz AJ, Winawer SJ. Management of colorectal polyps. *CA-A Cancer Journal for Clinicians* 1997; **47**(2); 112.

Volk E, Goldblum J, Petras RE, Carey WD, Fazio V. Management and outcome of patients with invasive carcinoma arising in colorectal polyps. *Gastroenterology* 1995; **109**: 1801–7.

Waye JD. Endoscopic treatment of adenomas. *World Journal of Surgery* 1991; **15**: 14–19.

Williams WB, Hunt RH, Loose H, Riddell RH, Sakai S, Swarbrick ET. Colonoscopy in management of colon polyps. *British Journal of Surgery* 1974; **61**: 673–82.

Winawer SJ *et al.* Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. *New England Journal of Medicine* 1993; **328**: 901–6.

Surgical treatment

Binderow SR, Cohen SM, Wexner SD, Nogueras JJ. Must early postoperative oral intake be limited to laparoscopy? *Diseases of the Colon and Rectum* 1994; **37**: 584–589.

Enker WE. Potency, cure and local control in the operative treatment of rectal cancer. *Archives of Surgery* 1992;**127**: 1396–402.

Guillem JG, Paty PB, Cohen AM. Surgical treatment of colorectal cancer. *CA-A Cancer Journal for Clinicians* 1997; **47**(2): 113–28.

Havenga K, DeRuiter MD, Enker WE, Welvaart K. Anatomical basis of autonomic nerve-preserving total mesorectal excision for rectal cancer. *British Journal of Surgery* 1996; **83**: 384–8.

Havenga K, Enker WE, McDermott K, Cohen AM, Minsky BD, Guillem J. Male and female sexual and urinary function after total mesorectal excision with autonomic nerve preservation for carcinoma of the rectum. *Journal of the American College of Surgeons* 1996; **182**: 495–502.

Heald RJ. The ‘Holy plane' of rectal surgery. *Journal of the Royal Society of Medicine* 1988; **81**: 503

MacFarlane JK, Ryall RDH, Heald RJ. Mesorectal excision for rectal cancer. *Lancet* 1993; **341**: 457–60.

Ramos JM, Beart RWJ, Goes R, Ortega AE, Schlinkert RT. Role of laparoscopy in colorectal surgery. A prospective evaluation of 200 cases. *Diseases of the Colon and Rectum* 1995; **38**: 494–501.

Scotia Study Group. Single-stage treatment for malignant left-sided colonic obstruction: a prospective randomized clinical trial comparing subtotal colectomy with segmental resection following intraoperative irrigation. *British Journal of Surgery* 1995; **82**: 1622–7.

Stewart J, Diament RH, Brennan TG. Management of obstructing lesions of the left colon by resection, on-table lavage, and primary anastomosis. *Surgery* 1993; **114**: 502–5.

Adjuvant therapy

Cohen AM *et al.* Adjuvant therapy for colorectal cancer. *Current Problems in Surgery* 1998; **22**(1): 1–80.

Dube S, Heyen F, Jenicek M. Adjuvant chemotherapy in colorectal carcinoma: results of a meta-analysis. *Diseases of the Colon and Rectum* 1997; **40**: 35–41.

Hyams DM *et al.* A clinical trial to evaluate the worth of preoperative multimodality therapy in patients with operable carcinoma of the rectum. A progress report of the National Surgical Adjuvant Breast and Bowel Project Protocol R-03. *Diseases of the Colon and Rectum* 1997; **40**: 131–9.

International Multicenter Pooled Analysis of Colon Cancer Trials Investigators. Efficacy of adjuvant fluorouracil and folinic acid in colon cancer. *Lancet* 1995; **345**: 939–44.

MacDonald JS. Adjuvant therapy for colon cancer. *CA-A Cancer Journal for Clinicians* 1997; **47**(4): 243–56.

Minsky B *et al.* Preoperative 5-FU, low-dose leucovorin, and radiation therapy for locally advanced and unresectable rectal cancer. *International Journal of Radiation Oncology* 1997; **37**: 289–95.

Moertel CG FT, MacDonald JS. Intergroup study of fluorouracil plus levamisole as adjuvant therapy for stage II/Dukes' B2 colon cancer. *Journal of Clinical Oncology* 1995; **13**: 2936–43.

Stockholm Colorectal Cancer Study Group. Randomized study on preoperative radiotherapy in rectal carcinoma. *Annals of Surgical Oncology* 1996; **3**: 423–30.

Swedish Rectal Cancer Trial. Improved survival with preoperative radiotherapy in resectable rectal cancer. *New England Journal of Medicine* 1997; **336**: 980–7.

Tepper J *et al.* Adjuvant postoperative fluorouracil-modulated chemotherapy combined with pelvic radiation therapy for rectal cancer: initial results of Intergroup 0114. *Journal of Clinical Oncology* 1997; **15**: 2030–9.

Follow-up

Burke AB *et al.* Recommendations for follow-up care of individuals with an inherited predisposition to cancer. *Journal of the American Medical Association* 1997; **277**: 915–19.

Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen C. An evaluation of the carcinoembryonic antigen (CEA) test for monitoring patients with resected colon cancer. *Journal of the American Medical Association* 1994: 943–7.

Safi F, Link KH, Beger HG. Is follow-up of colorectal cancer patients worthwhile? *Diseases of the Colon and Rectum* 1993; **36**: 636–44.

Wolf RF, Cohen AM. The minuscule benefit of serial CEA monitoring after effective curative treatment for primary colorectal cancer. *Journal of the American College of Surgeons* 1997; **185**: 60–4.

Zeng Z, Cohen AM, Urmacher C. Usefulness of carcinoembryonic antigen monitoring despite normal values in node-positive colon cancer patients. *Diseases of the Colon and Rectum* 1993; **36**:

1063–8.

Liver metastases

Fong Y, Blumgart LH, Cohen AM, Fortner J, Brennan MF. Repeat hepatic resections for metastatic colorectal cancer. *Annals of Surgery* 1994; **220**(5): 657–62.

Fong Y *et al.* Liver resection for colorectal metastases. *Journal of Clinical Oncology* 1997; **15**: 938–46.

Launois B, Jamieson GG. The importance of Glisson's capsule and its sheaths in the intrahepatic approach to resection of the liver. *Surgery, Gynecology and Obstetrics* 1992; **174**: 7–10.

26.4 Large-bowel obstruction

John P. Welch

Pathophysiology
Clinical features
Laboratory investigations
Etiology
- Carcinoma of the colon
- Diverticulitis
- Volvulus
- Others

Management principles
- Colon carcinoma
- Diverticulitis
- Volvulus
- Other causes

Further reading

Intestinal obstruction involves the colon 20 to 40 per cent of the time. Colonic obstruction is associated with potentially serious complications such as perforation, and the timing and selection of appropriate operative procedures are important. Symptoms can develop slowly and progressively or fulminantly. Among adults, elderly people are usually affected. The sigmoid colon is the usual site: this portion of the intestine is thick walled, not particularly distensible, and comparatively narrow.

Pathophysiology

Large-bowel obstruction is associated with increased fluid secretion into the lumen, and, in later stages, with decreased intestinal motility. As the colon dilates, the functioning of the ileocecal valve is important. When the valve is competent (thus preventing reflux back into the ileum), a closed-loop obstruction develops. Because of a correlation between tension in the colon wall and the diameter of the bowel, based on LaPlace's law, the cecum dilates more rapidly than the remainder of the colon. Because it is thin walled, the cecum is prone to perforate, especially when it reaches a diameter of 10 to 12 cm. Initially the serosa splits between the teniae, and patchy sites of hemorrhage and perforation appear. Progressive peritoneal contamination occurs, although the obstructed colon may not fully decompress. Perforation may also occur in a more distal segment, usually just proximal to an obstructing lesion; ischemic changes have been described proximal to obstructing carcinomas.

Clinical features

The manifestations of an obstructed colon can occur insidiously or less often rapidly, superimposed on chronic complaints. Frequently the abdomen distends gradually and there is progressive constipation and finally obstipation. The severe, crampy abdominal pain characteristic of small-bowel obstruction is not a common feature; most patients have dull, lower abdominal cramps, which may radiate to the hypogastrium if the ascending or transverse colon is involved. The distended abdomen is tympanitic, and high-pitched tinkles or more prolonged, low-pitched peristaltic sounds may be present. There is local tenderness in the right lower quadrant over the cecum if perforation is impending. The rectum tends to feel empty and capacious; rarely, a rectal tumor is palpable on digital rectal examination. Traces of blood suggest a possible tumor or bowel ischemia. The patient is carefully examined for the presence of hernias.

The clinical presentation can be more fulminant when there is complete obstruction and/or perforation of the colon. Abdominal distension is frequently pronounced; abdominal tenderness, guarding and rebound are found. Hypovolemia can lead to hypotension and oliguria.

Laboratory investigations

Abdominal films provide useful information: they may suggest or confirm the diagnosis and site of obstruction, and the degree of cecal distension can be assessed from them. Pneumoperitoneum suggests perforation of the colon. Although the cecum may dilate greatly in chronic obstruction without rupturing, perforation should be suspected whenever this portion of the colon is greater than 10 cm in diameter. Urgent surgical intervention is indicated if obstruction appears to be complete without decompression into the small bowel. If the point of obstruction is near the ileocecal valve, films may suggest small-bowel obstruction alone. Radiographs of patients with fecal impaction mimicking obstruction show the mottled density of fecal matter in the rectum.

The site of obstruction and its severity can be detected with a retrograde contrast study (Fig. 1 and Fig. 2); the diagnosis of pseudo-obstruction is also ruled out. Computed tomographic (**CT**) scans are of value: they may illustrate the transitional area of colonic obstruction, as well as extracolonic abnormalities and more subtle degrees of pneumoperitoneum (Fig. 3).

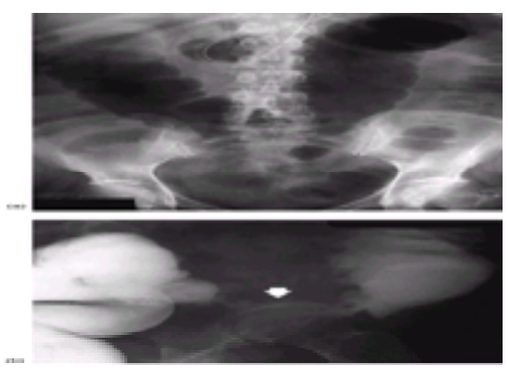


Fig. 1. (a) Plain film of a patient with large-bowel obstruction caused by a diverticular stricture. (b) The point of narrowing (arrow) and the dilated proximal colon are well demonstrated in this spot view from a contrast study.



Fig. 2. This patient had an annular obstructing carcinoma at the junction of the sigmoid and descending colon; the point of narrowing (arrows) is seen well with this barium enema.



Fig. 3. CT scan of a patient with a colonic stricture caused by diverticulitis: the obstructed colon is dilated; note the descending colon (curved arrow) and ascending colon (straight arrow).

Etiology

Carcinoma of the colon

Carcinoma is the most frequent cause of large-bowel obstruction in developed countries (at least 55 per cent of cases). The left colon is the most likely site of obstruction and the extraperitoneal rectum the least. Signs of partial obstruction progress to those of complete obstruction when the narrowed colonic lumen is occluded by a fecal bolus. Since the right colon has semiliquid contents and a relatively wide lumen, obstruction occurs late in this segment and may be acute in its presentation, especially if the ileocecal valve is competent. The operative risk is increased considerably when perforation is present.

Diverticulitis

Mechanical obstruction by diverticular disease tends to be chronic from stricture formation. Patients have a history of diverticulosis and possibly symptomatic episodes of diverticulitis. Acute obstruction occurs less frequently, with obturation of a strictured, edematous segment. Partial or complete small-bowel obstruction may occur simultaneously, related to adherence of small-bowel loops to the inflamed sigmoid colon. Differentiation from carcinoma may be impossible short of resection.

Volvulus

Volvulus occurs in the sigmoid colon approximately. 60 per cent of the time. Typically it is a disease of elderly, institutionalized individuals. The condition is precipitated by twisting of the long mesentery, with possible resulting strangulation. Radiographs show a massively dilated colon and retrograde contrast studies demonstrate a smooth ‘bird's beak' at the point of obstruction.

Cecal volvulus involves a floppy cecum, at times attached to adhesions. Plain films suggest the diagnosis, with a ‘coffee-bean' appearance of the dilated colon in the left upper quadrant together with evidence of small-bowel dilatation.

Volvulus of the transverse colon and splenic flexure are rare.

Others

Rare causes of large-bowel obstruction include hernias, metastatic malignancies originating in the ovary or pancreas, obturation due to factors other than fecal impaction, radiation strictures, endometriosis, intussusception, or adhesions (Fig. 4) (Table 1).



Fig. 4. CT scan demonstrating a colonic lipoma (arrow) that led to intussusception and large-bowel obstruction.

Strictures:
Radiation
Ischemic
Infections:
Schistosomiasis
Lymphogranuloma
Actinomycosis
Tuberculosis
Amebiasis
Obturation:
Gallstone
Bezoar
Extrinsic:
Pancreatitis
Endometriosis
Metastatic carcinoma
Inflammatory adhesions

Table 1 Unusual causes of colon obstruction

Management principles

If possible, large-bowel obstruction should be treated before perforation occurs. As few operations as possible are done, although the exact procedures depend on the degree of preparation of the colon, the patient's overall medical condition, and the urgency. If an anastomosis is contemplated, certain factors should be present, including an adequate blood supply, an operative field free of contamination, and lack of tension between the bowel segments. An anastomosis between distended and normal colon is more risky than one between distended ileum and normal colon.

Retrograde contrast studies are useful to point out the site of obstruction before surgery, and they can serve to distinguish carcinoma from volvulus or a diverticular stricture. Barium should not be refluxed above the point of obstruction in great amounts since it could later impact; gentle infusion of a water-soluble agent is preferred if there is any suspicion of perforation. CT adds the dimension of visualizing extracolonic inflammation with diverticulitis, extracolonic spread with colon cancer, and subtle degrees of pneumoperitoneum possibly missed within plain films. Endoscopy may be attempted to visualize the site of obstruction, but has the disadvantage of

introducing air into the distended colon.

Since many patients are elderly and fluid-depleted, hemodynamic monitoring may be necessary, along with nasogastric suctioning and measurement of urinary output. Electrolyte imbalance and acidosis are infrequent.

Most patients ultimately require surgical treatment. Three major types of operations are done: (1) decompressive procedures such as loop ileostomy; (2) colon resections; and (3) bypass procedures. At times, different types of procedures are combined. Over the past few decades there has been a tendency toward a higher proportion of resections and less use of stomas such as colostomies and ileostomies. These developments have led to an improved quality of life in affected patients. Resections are safest done under the following circumstances: when (1) the ileocecal valve is incompetent; (2) the obstructed colonic segment can be removed in its entirety; and (3) the bowel can be decompressed to a normal size without vascular compromise.

Colon carcinoma

Tumors lying in the right colon are treated by right colectomy. Even when the colon is distended a curative resection is possible and the ileum can be safely sutured to the collapsed distal colon. If perforation and contamination are present, an ileostomy and mucous fistula are constructed rather than an anastomosis. A lesion in the mid-transverse colon can be managed by an extended right colectomy (to the region of the splenic flexure) and one in the splenic flexure by subtotal colectomy. A segmental resection and direct anastomosis using distended, unprepared proximal bowel is not recommended.

Operations for obstructing tumors in the distal sigmoid are particularly difficult. Using a distended, fecal-loaded colon for anastomosis is dangerous. In this situation, a resection with end colostomy and mucous fistula is a popular alternative. With rectal tumors the distal segment may not reach the abdominal wall and this is closed rather than exteriorized (Hartmann procedure). On-table lavage techniques are becoming increasingly popular: these permit preparation of the bowel lumen and treatment of bowel distension, thus allowing construction of a safe anastomosis. The colon is irrigated with fluid introduced into the ileum or the base of the appendix. Fluid passing through the distal colon (proximal to the bowel to be resected) is drained into a floor receptacle through anesthetic tubing. Although such maneuvers lengthen the operative procedure, they allow construction of a safe anastomosis. Perhaps a less satisfactory choice is subtotal colectomy since troubling diarrhea can result, a problem especially important in elderly patients with limited control of their anal sphincters.

The cecum should always be visualized during the operation, since there is risk of perforation, especially when the ileocecal valve is competent. Preoperative radiographs will have given an indication of marked cecal distension if present; cecal decompression is needed if the diameter is greater than 9 to 12 cm in the presence of mechanical obstruction. An extended resection may be necessary in these circumstances. Minor degrees of cecal ischemia can be managed with a cecostomy tube for decompression; perforation or advanced ischemia require resection.

Diverticulitis

A symptomatic diverticular stricture may be very difficult to differentiate from carcinoma, and the correct diagnosis may not be established until a laparotomy is done. Strictures can usually be operated on electively and it is possible to prepare the colon satisfactorily. This in turn allows a safe, one-stage resection and anastomosis.

In cases of acute diverticulitis, colonic ileus may accompany the inflammatory process. The symptoms can resolve after antibiotic therapy and bowel rest. A picture of high-grade mechanical obstruction is less common, resembling obstruction by carcinoma.

On-table lavage has potential application in these patients. It allows early resection and a safe anastomosis. The most common operation for acute complicated diverticulitis today is resection with end colostomy and closure of the distal segment (Hartmann procedure). If a large abscess lies adjacent to the bowel, preliminary percutaneous drainage may lead to resolution of the obstructive symptoms and allow a one-stage resection and anastomosis a few weeks later.

Volvulus

A sigmoid volvulus can usually be decompressed successfully with a sigmoidoscope. A tube is inserted through the rigid instrument and left in place to allow decompression to continue. Usually the colon is resected following bowel preparation, permitting a safe primary anastomosis. In a very debilitated patient, decompression alone can be opted for, although the risk of recurrent volvulus is high. Rarely the colon is gangrenous or perforated; emergency laparotomy is then mandatory. An anastomosis is not recommended in this setting; the Hartmann procedure or construction of an end colostomy and mucous fistula are the procedures of choice.

Cecal volvulus is managed differently. Occasionally reduction can occur after colonoscopy, but the risk of recurrence or of bowel compromise is great. Thus the treatment of choice is early laparotomy. The anterior aspect of the cecum and ascending colon can be sutured to the peritoneum (cecopexy) or a tube cecostomy is done. In most good-risk patients, however, primary resection and anastomosis are preferable to avoid recurrence. If significant peritoneal contamination from perforation of the bowel is present, resection with ileostomy and later re-anastomosis should be considered.

Volvulus of the transverse colon is rare; it is managed in the same fashion as cecal volvulus with early operative intervention.

Other causes

Other anatomic and pathologic entities causing large-bowel obstruction are listed in Table 1. The cause of the obstruction needs to be managed individually. Thus a hernia is repaired once the obstructed colon is reduced, whereas an intussusception is reduced, with resection of the involved colon if necessary. Because it is so common, fecal impaction should be ruled out in institutionalized or debilitated elderly patients. The cecum must be visualized to assess its viability. If an anastomosis is to be performed, the criteria for a good outcome must be met.

Further reading

Ballantyne GH, Brandner MD, Beart RW, Jr, Ilstrup DM. Volvulus of the colon. Incidence and mortality. *Annals of Surgery* 1985; **202**: 83–92.

Deans GT, Krukowski ZH, Irwin ST. Malignant obstruction of the left colon. *British Journal of Surgery* 1994; **81**: 1270–6.

Feldman A., Welch JP. Diverticular stricture. In: Welch JP, Cohen JL, Sardella WV, Vignati PV, ed. *Diverticular disease. Management of the difficult surgical case*, pp. 194–205. Philadelphia: Williams and Wilkins, 1998.

Gibney EJ. Volvulus of the sigmoid colon. *Surgery, Gynecology and Obstetrics* 1991; **173**: 243–55.

Hughes LE. Complications of diverticular disease: inflammation, obstruction and bleeding. *Clinics in Gastroenterology* 1975; **4**: 147–70.

Koruth NM, Krukowski ZH, Younger GG, Hendry WS, Logir JR, Jones PF. Intra-operative colonic irrigation in the management of left-sided large bowel emergencies. *British Journal of Surgery* 1985; **72**: 708–11.

Radcliffe AG, Dudley HA. Intraoperative antegrade irrigation of the large intestine. *Surgery, Gynecology and Obstetrics* 1983; **156**: 721–3.

Saini S, Mueller PR, Wittenberg J, Butch RS, Rodkey GV, Welch CE. Percutaneous drainage of diverticular abscess. An adjunct to surgical therapy. *Archives of Surgery* 1986; **121**: 475–8.

Welch JP. *Bowel obstruction. Differential diagnosis and clinical management*. Philadelphia: WB Saunders, 1990.

Welch JP, Warshaw AL. Isolated small bowel obstruction as the presenting feature of colonic disease. *Archives of Surgery* 1977; **112**: 809–12.

26.5 Volvulus of the colon

Alan R. Berry

Introduction
Volvulus of the sigmoid colon
Incidence and pathogenesis
Presentation and clinical features
Radiology
Treatment
Volvulus of the caecum
Presentation and clinical features
Radiology
Treatment
Other types of colonic volvulus
Further reading

Introduction

Acute volvulus of the colon is a surgical emergency and, although accounting for not more than 5 per cent of all cases of obstruction of the large bowel in the West (Fig. 1), is a much more common cause of that obstruction eleswhere in the world. The incidence of the various types of colonic volvulus is shown in Fig. 2. Sigmoid volvulus is by far the most common, accounting for about two-thirds of all cases. Caecal volvulus accounts for most of the remaining cases, while other forms, including of the tranverse colon and splenic flexure, are very rare indeed. A simultaneous volvulus of the sigmoid and right colon has been reported. Together, sigmoid and caecal volvulus represent over 90 per cent of cases of colonic volvulus.

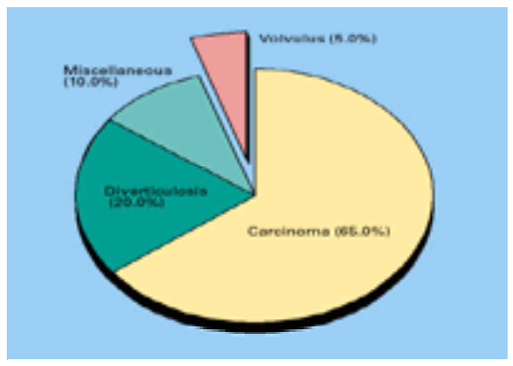


Fig. 1. The causes of colonic obstruction.

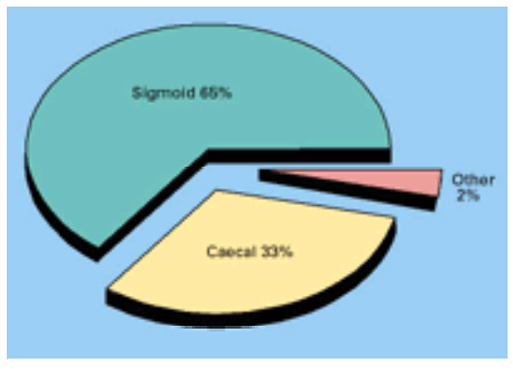


Fig. 2. The types of colonic volvulus.

The presence of an elongated mesentery about which the bowel can rotate is fundamental to the pathogenesis of colonic volvulus. Although a mesentery is a normal feature of the sigmoid colon, a caecal volvulus can only occur if the usual fixation of the caecum in the right iliac fossa has not occurred during development. Although many of the features of the different kinds of colonic volvulus are similar, important differences exist and so they are best considered separately.

Volvulus of the sigmoid colon

Incidence and pathogenesis

Sigmoid volvulus is common in India, Africa, parts of South America, and in Eastern Europe, where it has been shown to be responsible for up to 50 per cent of all cases of intestinal obstruction. In Western Europe and North America, it is uncommon, accounting for about 5 per cent of all cases of large-bowel obstruction. The incidence in one large reported series in New York was 1.3 per 10 000 hospital admissions.

It is thought that the major contributing factor in those parts of the world where there is a high incidence is the high dietary fibre content. Other factors that are recognized to be important are a long redundant colon, an acquired megacolon, chronic constipation, and a narrowly based mesentery.

The peak age incidence of sigmoid volvulus is in the eighth decade, with 70 per cent of patients presenting after 70 years of age and up to 40 per cent being over 80 years old. A high proportion of patients presenting with acute sigmoid volvulus (up to 60 per cent) come from institutional care. The condition is particularly common in elderly patients suffering from psychiatric and chronic neurological diseases such as stroke or multiple sclerosis. Other conditions with which it tends to be associated include cardiovascular disease and diabetes.

Presentation and clinical features

The condition usually presents with acute, colicky, abdominal pain, almost invariably associated with distension and, in one-half of all patients, with constipation. These features, in an elderly person with a chronic psychiatric or neurological complaint, should immediately raise the possibility of sigmoid volvulus. On examination, the most striking finding is of a tensely distended, tympanitic (‘drum-like') abdomen. The rectum is empty of stool. Bowel sounds are often increased, but signs of peritoneal inflammation such as rebound tenderness or guarding are unusual. When these signs are present they suggest that colonic infarction or gangrene has occurred. Signs of dehydration may be apparent if presentation has been delayed and these should always be sought.

Radiology

The most useful investigation of patients suspected of having a sigmoid volvulus is a plain supine abdominal radiograph; this alone may be diagnostic in 70 to 80 per cent. The typical appearance is that of a single, grossly distended loop of colon arising out the pelvis and extending towards the diaphragm. Haustral markings are

usually lost (Fig. 3).



Fig. 3. Plain supine radiograph showing sigmoid volvulus.

Investigation by a contrast enema, such as dilute barium or a water-soluble contrast medium, will increase the diagnostic yield of radiology to over 90 per cent. If gangrenous bowel or perforation are suspected, a water-soluble contrast must be used rather than barium, which can produce a severe peritonitis. The pathognomonic sign on a contrast enema is described as a ‘bird's beak' or ‘ace of spades' appearance, produced as the upper end of the barium column tapers into the spirally twisted, distal sigmoid colon. A computed tomographic scan can show a characteristic whirl pattern and may be helpful in difficult cases.

Treatment

Most patients can be treated initially by non-operative means. Careful rigid sigmoidoscopy and the passage of a flatus tube via the sigmoidoscope is successful in up to 90 per cent of cases, and is well worth trying in the first instance. Protective clothing is recommended as the results of a successful decompression may be explosive. A mortality of between 1 and 4 per cent is reported for the treatment, and great care is essential if colonic perforation is to be avoided. Following a successful deflation, the flatus tube should be left in place for at least 48 h, and some would recommend as long as 5 days. Unless this is done the likelihood of an early recurrence is very high (50–90 per cent).

The alternative to endoscopic deflation is emergency surgery, but this is associated with a mortality of up to 40 per cent. Mortality is higher for sigmoid resection (52 per cent) than for operations aimed at fixing the mobile colon (colopexy), which has a reported 2.8 per cent mortality. However, the 28 per cent incidence of recurrence reported to follow sigmoid colopexy militates against it. When the disease is confined to the sigmoid alone, late recurrence after resection is unusual (6 per cent). When, however, sigmoid volvulus is associated with megacolon, late recurrence is as high as 82 per cent and subtotal colectomy is recommended. In the minority of patients who have signs of infarction and peritonitis, emergency surgery should not be delayed, but in most cases early treatment should be non-surgical, surgery being reserved for those in whom this fails.

The emergency operation of choice is sigmoid exteriorization and resection, using a modification of the Paul–Mickulicz technique (Fig. 4). After resuscitation with intravenous fluids, antibiotic prophylaxis (such as a cephalosporin and metronidazole) is given before inducing general anaesthaesia. The twist in the colon is reduced through a midline incision. It is helpful to remove gas from the colon by puncturing the bowel wall with a small (19 gauge) intravenous needle attached to the suction apparatus. This makes the bowel easier to handle and less likely to be torn or damaged during handling. A second, smaller skin incision is made in the left iliac fossa through which the collapsed, freely mobile sigmoid colon can easily be exteriorized. At this stage the future afferent and efferent loops of colon should be opposed, preparing the bowel for what will become a ‘double-barrelled' colostomy. The midline abdominal incision is closed. The sigmoid colon is excised outside the abdomen and the double-barrelled colostomy is completed. Care must be taken when dividing the mesentery at this point to secure and ligate all of the blood vessels, as the large sigmoid branches of the inferior mesenteric artery can easily slip back into the abdominal cavity causing troublesome haemorrhage. This problem can be avoided by dividing the mesentery at the level of the abdominal wall, inside the abdomen, before closure. Subsequently, the two lumens of bowel are transected and the mucosa is sutured to the skin to form the colostomy.

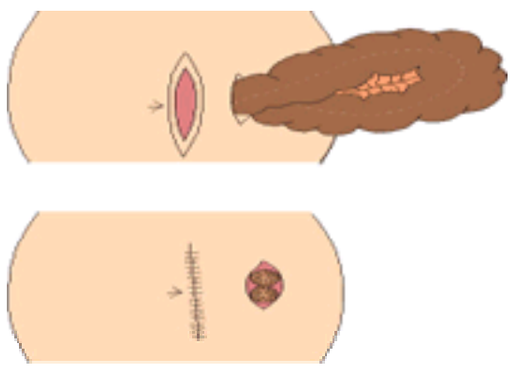


Fig. 4. Sigmoid exteriorization and resection with a double-barrelled colostomy.

Unlike the original Paul–Mickulicz operation, an enterotomy clamp is not used to close the colostomy, but a single, one-stage closure is performed approx. 6 to 8 weeks after the original operation. Because of the close approximation of the afferent and efferent limbs, a full laparotomy is not necessary in order to do this.

An alternative operation in an emergency is the Hartmann procedure, in which the distal bowel end is closed within the pelvis after sigmoid colectomy. This operation has no advantages in the management of sigmoid volvulus when the distal and proximal ends of bowel can be approximated easily. It has the disadvantage that the second operation to reanastamose the bowel requires a full laparotomy and can be very difficult. Sigmoid colectomy and primary anastomosis is not usually recommended in an emergency, but may be an alternative for an experienced colonic surgeon, using ‘on-table', antegrade colonic irrigation via caecal intubation to clean the bowel of faeces before anastomosis.

Patients who develop recurrent volvulus after non-operative, or rarely after emergency operative, treatment should be offered elective surgery, as there is a reported mortality of over 20 per cent for those who continue to be managed conservatively. The operation of choice is sigmoid colectomy and primary anastomosis after a suitable preoperative bowel preparation. This can be achieved by giving clear fluids only by mouth for 48 h before surgery and two doses of sodium picosulphate on the day before operation.

Mesosigmoplasty has been advocated as an alternative to sigmoid resection either as an elective procedure after successful conservative treatment or as an emergency. The operation involves plicating the mesentery by incising the peritoneum over both sides of the mesosigmoid vertically, mobilizing the peritoneum, and resuturing it transversely. In a large series reported from India, Subrahmanyam has shown low mortality, morbidity, and recurrence rates.

Other techniques advocated for treating this condition are extraperitonealization of the sigmoid and laparoscopically assisted resection.

An algorithm for the management of patients with sigmoid volvulus is shown in Fig. 5.

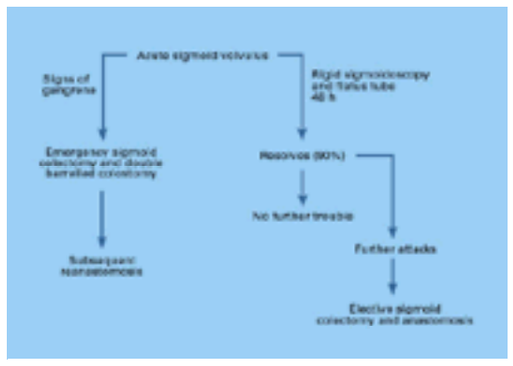


Fig. 5. Algorithm for the management of sigmoid volvulus.

Volvulus of the caecum

Caecal volvulus is much less common than volvulus of the sigmoid colon, accounting for less than 1 per cent of all cases of intestinal obstruction and up to 40 per cent of all cases of colonic volvulus. In a series from Scandinavia the incidence was 3 to 6 per million per year. It carries a mortality of 20 per cent. In this condition the caecum remains mobile and shares a common mesentery with the ileum. It is therefore free to rotate, and can do so either in a clockwise or anticlockwise direction out of the right iliac fossa to the middle or left side of the upper abdomen (Fig. 6), producing a closed-loop obstruction of the caecum and ascending colon. A variant of caecal volvulus where the caecum ends up in a more superior and anterior position is known as caecal bascule.

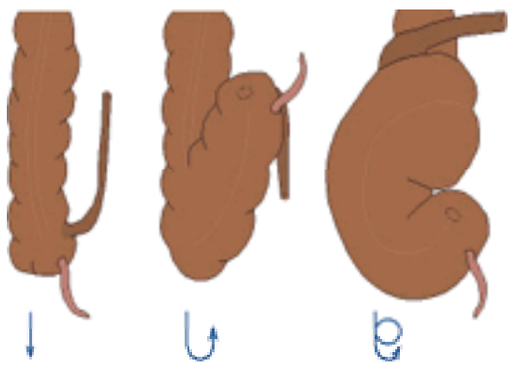


Fig. 6. One mechanism of caecal volvulus recently encountered.

Presentation and clinical features

Caecal volvulus may present as a fulminant condition with intestinal strangulation secondary to mesenteric torsion or, less dramatically, with features of subacute intestinal obstruction. Rarely it is a chronic, intermittent condition. The peak age of presentation is 30 to 40 years, much younger than that for sigmoid volvulus. It is more common in females. There often appears to have been a triggering event, such as a recent laparotomy, and it is well recognized following gynaecological procedures. Occasionally, it can occur secondarily to an obstructing colonic carcinoma and, like sigmoid volvulus, it may be related to a high fibre intake. The presenting symptoms are usually non-specific. Abdominal pain is almost invariably present and nausea, vomiting, constipation, and distension will occur in about one-third of patients. In thin patients it may be possible to palpate the resonant, distended caecum in the central or upper abdomen while the right iliac fossa is empty. Not infrequently the diagnosis is only made at diagnostic laparotomy in a patient with continuing subacute obstruction that does not resolve.

Radiology

The key to the diagnosis of caecal volvulus, as with sigmoid volvulus, is the plain abdominal radiograph. The caecum typically assumes a gas-filled, ‘comma shape' facing inferiorly and to the right (Fig. 7). In a retrospective analysis, the diagnosis was evident on the plain abdominal radiograph in 40 out of 45 patients with caecal volvulus. The diagnosis is usually made in around 50 per cent of patients, however. Other radiological appearances are of non-specific colonic or small-bowel obstruction. A barium enema may be useful to exclude any other predisposing colonic lesion and on occasions it has been effective in reducing the volvulus.



Fig. 7. Plain supine radiograph showing caecal volvulus.

Treatment

The mainstay of treatment for this condition is surgery and, unlike sigmoid volvulus, endoscopic deflation has little place. Prompt caecal resection is mandatory in those with caecal perforation or infarction, and this is usually possible with a primary ileocolic anastomosis. In severely ill patients, however, it may be expedient to exteriorize the proximal and distal bowel ends after resection, as an ileostomy and mucous fistula. These can then be anastomosed at a second operation 6 weeks later. The mortality in patients with gangrenous bowel is as high as 40 per cent.

The management of patients who present with caecal volvulus without evidence of strangulation is less clearly defined. Fixation of the caecum (caecopexy) to prevent recurrent torsion has its advocates. This is best achieved by combining suture fixation to the lateral abdominal wall with stripping of the parietal and visceral peritoneum to create raw surfaces that will stick firmly and permanently together. Unfortunately, the technique has a recurrence rate of up to 20 per cent. The addition of a tube caecostomy in addition to caecopexy has been advocated, and has been shown to reduce the recurrence rate. However, this addition is associated with an increase in surgical morbidity, mainly of septic problems from leakage, such as cellulitis, wound infection, and occasionally necrosis of the abdominal wall. The caecostomy tube can be removed after 1 week and the tract will normally close spontaneously within a few days. Occasionally, however, discharge of ileal contents persists, requiring further surgery formally to close the hole or resect the caecum.

The definitive treatment of caecal volvulus, even when the bowel is viable, is caecal resection and primary ileocolic anastomosis. In the author's opinion this is the preferred management. Whichever method is chosen when the diagnosis is made, prompt treatment is essential to prevent bowel strangulation.

Other types of colonic volvulus

Volvulus of the splenic flexure is extremely rare, accounting for less than 1 per cent of colonic volvuli; it results from a congenital absence of one or more of the ligaments that normally fix that part of the bowel. The patients present with acute intestinal obstruction, and usually give a history of chronic constipation. The diagnosis is made by diagnostic enema. Endoscopic deflation is often possible, but the patient should undergo definitive bowel resection to prevent recurrence.

Volvulus of the transverse colon is also very rare. It occurs when a degree of malrotation exists and can also result from chronic constipation, which leads to elongation of the mesentery. The mechanisms are similar to those involved in sigmoid volvulus. The clinical presentation is usually subacute or chronic, but patients can present with an acute, fulminant condition requiring emergency surgery. For this reason, when the diagnosis is suspected a barium enema should be performed and prompt elective resection undertaken. Isolated volvulus of the descending colon leading to gangrene is rare but has been described.

Further reading

Anderson JR, Welch GW. Acute volvulus of the right colon. *World Journal of Surgery* 1986; **10**: 336–42. [Describes a combined experience from Scotland and Northern Ireland of 69 patients with acute caecal volvulus, and reports a mortality of 40 per cent when the diagnosis is delayed and the colon becomes gangrenous.]

Bak MP, Boley SJ. Sigmoid volvulus in elderly patients. *American Journal of Surgery* 1986; **151**: 71–5. [Reports high success in managing the initial volvulus by non-operative means and recommends early elective surgery subsequently to maximize survival in elderly patients.]

Subrahmanyam M. Mesosigmoplasty as a definitive operation for sigmoid vulvolus. *British Journal of Surgery* 1992; **70**(7): 683–4. [This paper from India reports the successful use of mesosigmoplasty either as an elective procedure following successful non-operative treatment or as an emergency operation, when non-operative treatment has failed, claiming low mortality and recurrence.]

Tejler G, Jiborn H. Volvulus of the caecum. *Diseases of the Colon and Rectum* 1981; **31**: 445–9. [Describes the anatomical basis and types of caecal volvulus, reports the authors' own experience, and presents a literature review, concluding that right hemicolectomy is the treatment of choice.]

26.6 Chronic constipation in adults

Neil Mortensen and Timothy A. Cook

Introduction
Assessment
Adult Hirschsprung's disease
Idiopathic megarectum and megacolon
Severe idiopathic slow transit constipation
Chronic idiopathic intestinal pseudo-obstruction
Evacuatory disorders
Further reading

Introduction

An increasing number of adult patients with severe constipation who have not responded to the usual dietary measures and laxatives are being seen in surgical practice. This is not a trivial symptom and the patient's life both at home and at work is often severely restricted by abdominal pain, distension, a poor sense of well being, and decreased bowel frequency. The symptom of constipation cannot be precisely defined. Patients have a problem with defaecation and may be worried by infrequent and/or hard stools which are difficult to pass. Different types of severe constipation have a different cause and require a different therapeutic approach.

Assessment

Patients with minor, easily treated symptoms do not require sophisticated investigation. A careful history of the exact nature of the constipation, together with any history of sexual or psychiatric problems, is essential. Abdominal examination will reveal the presence of a palpable faecal mass. Weakness of the pelvic floor is indicated by extreme descent of the perineum on straining. Laxity of the anal sphincter and perianal soiling is usually associated with gross faecal impaction in the rectum.

Patients can generally be divided into those with a colon of normal width and those with an abnormally wide or capacious rectum and/or colon: a barium enema defines into which of these two groups the patient falls. In those patients with rectal impaction, an unprepared radiograph using a water-soluble contrast material will illustrate the rectum and the distal colon with the impacted stools still in place. The rate of transit of contents through the colon can be estimated by a single abdominal radiograph taken 5 days after the patient has swallowed 20 radio-opaque markers. In specialist units, information about evacuation disorders can be obtained by recording the appearance of the rectal and anal canal on video tape or fast film sequences as the patient expels a barium gel or paste. Simple manometric studies can be used to test rectal sensation and the rectoanal inhibitory reflex by distending a balloon in the rectal ampulla.

Adult Hirschsprung's disease

A small proportion of patients with Hirschsprung's disease do not present until adolescence or young adult life. Elliot and Todd reported 39 such patients treated at St Mark's Hospital, London, between 1965 and 1984. There were 26 males and 13 females with a mean age of 23 years. The Duhamel operation produced good results, with complete continence, and daily evacuation, in 36 patients; the remaining three patients had poorer results, with occasional incontinence. Investigation of this disorder should include a barium enema which will reveal the characteristic junction between the dilated proximal colon and the aganglionic segment (Fig. 1). Anorectal physiology studies will show an absent rectosphincteric reflex, and full thickness rectal biopsy will demonstrate the absence of ganglion cells in the myenteric plexus. A frozen section biopsy of the normal colon during the operation will ensure that normally innervated bowel is brought down behind the aganglionic rectum.

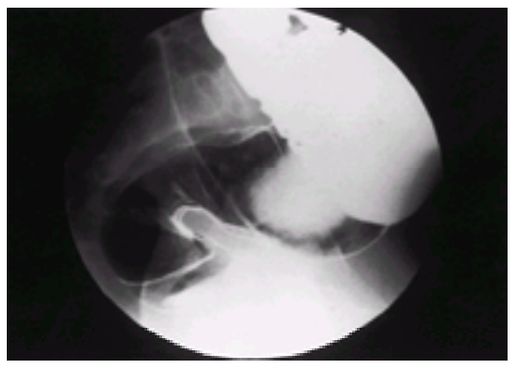


Fig. 1. Lateral radiograph of the pelvis. Barium enema demonstrating the narrowed area in Hirschsprung's disease at the junction with the aganglionic segment.

Idiopathic megarectum and megacolon

Once Hirschsprung's disease and metabolic and other secondary causes of a dilated colon have been excluded there remains a group of patients with a dilated large bowel of unknown cause. This condition affects males and females equally and it may commence in childhood or adult life. Patients with idiopathic megarectum present earlier than those with idiopathic megacolon. Recurrent faecal impaction is the usual presentation of idiopathic megarectum, though it may present with soiling. Abdominal pain and distension are predominant features in idiopathic megacolon and patients may develop sigmoid volvulus. The nerve plexuses and muscle coats appear grossly normal on histological assessment and the aetiology of this condition remains unknown. The rectal diameter is usually greater than 6.5 cm. In two-thirds of cases the rectum alone is affected, but in one-third of cases the dilation extends into the sigmoid colon (Fig. 2).

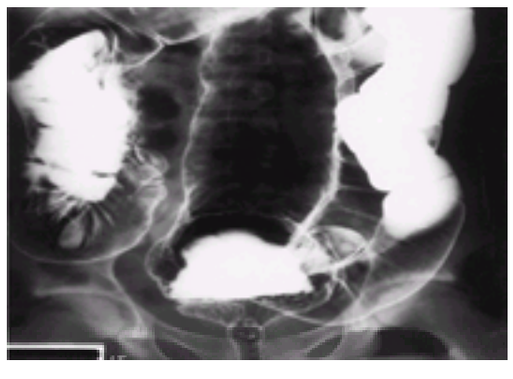


Fig. 2. Barium enema. Megarectum and megasigmoid.

Around two-thirds of patients will respond to medical treatment; a combination of enemas, suppositories, and osmotic laxatives which will keep the rectum empty. If laxatives are poorly tolerated or are ineffective and it seems unlikely that the grossly dilated bowel will recover, surgery is indicated. It is not clear which is the best procedure to offer these patients. The alternatives include a preliminary loop ileostomy followed by either a Duhamel procedure or a resection and coloanal anastomosis. In those with a moderately dilated colon a colectomy with ileorectal anastomosis may offer a reasonable compromise. In a recent series of 40 patients undergoing a colectomy for this condition 80 per cent achieved normal bowel frequency and most were relieved of the need for laxatives. A third of the patients,

however, continued to experience some abdominal pain. Ileorectal anastomosis produced results superior to those of a caecorectal anastomosis or a sigmoid resection: these last operations had a higher incidence of persistent constipation. In patients with a grossly dilated rectum the most commonly performed procedure was the Duhamel operation. It seems, however, that results in this situation are less satisfactory than those obtained in patients with Hirschsprung's disease. In a series of 20 patients reported by Stabile only half achieved a normal bowel frequency. Seven patients remained constipated and of these five required further surgery. The other alternative is a resection of the grossly dilated rectum and a coloanal anastomosis. This usually has to be performed using a hand suturing technique because of the thickness of the rectal stump. There are no reliable data on the outcome of this procedure.

The patient who remains constipated despite colectomy may be treated by creation of a stoma, bringing out proximal bowel of normal diameter. If there is progressive dilatation of the proximal colon an ileostomy may be necessary and, in very selected cases, a restorative proctocolectomy would be an alternative approach.

Severe idiopathic slow transit constipation

Preston and Leonard Jones (1986) reported a series of 64 women complaining of severe constipation. The patients passed about one stool weekly with the aid of laxatives and were greatly troubled by abdominal pain, bloating, malaise, and nausea to the extent that their symptoms were an increasing social disability. A decrease in bowel frequency and other symptoms were often noticed around the age of puberty and slowly became worse until they were at their most severe by the third decade. In a few patients the symptoms appeared suddenly after an abdominal operation or an accident.

A barium enema in these patients shows a normal rectal and colonic diameter; on large bowel transit studies, less than 80 per cent of the 20 shapes or radio-opaque markers ingested on the first day have been passed on the fifth day (Fig. 3). The rectosphincteric reflex is normal, and a full thickness rectal biopsy is not normally required, though in doubtful cases it is sensible to exclude Hirschsprung's disease.

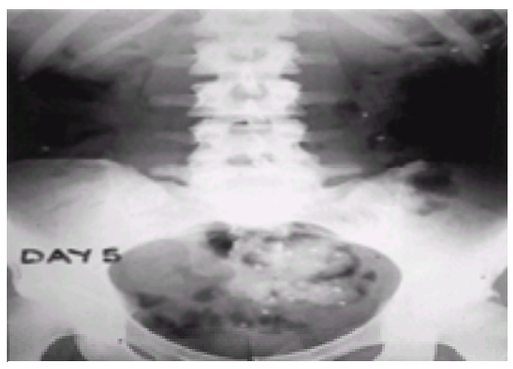


Fig. 3. Plain radiograph 5 days after the ingestion of 20 radio-opaque markers. More than 80 per cent are still present, indicating slow transit.

As is the case in patients with megarectum, it is as well to institute medical therapy for as long as possible, aimed at emptying the rectum with osmotic laxatives, suppositories, or enemas. These patients have usually tried the whole range of laxatives and many of them become increasingly constipated despite quite high doses of concurrent laxative medication. Interestingly, high fibre diets often make their symptoms worse. The condition in some patients with intractable constipation may be improved by using biofeedback where they are taught to relax the pelvic floor muscles during defaecation.

The surgical options include a segmental colectomy, a colectomy and caecorectal anastomosis, a colectomy and ileorectal anastomosis, a proctocolectomy and ileoanal reservoir, or an ileostomy. It is now generally believed that segmental colectomies have nothing to offer: patients tend to become progressively constipated within a year or so of this operation and this may reflect the presumed panintestinal nature of the disorder. The most widely used operation is colectomy and ileorectal anastomosis. Lubowski et al. (1996) reported a median bowel frequency of four per 24 h in 52 patients following ileorectal anastomosis. Five patients used antidiarrhoeal medication and six were incontinent postoperatively. One patient required an ileostomy for recurrent constipation. Overall, 90 per cent were satisfied with outcome of the operation. A worrying feature was the fact that 52 per cent of the patients had continuing abdominal pain, but the there was improvement in the degree of pain. Kamm et al. (1988) reported persistent abdominal pain in 71 per cent of patients. In this study 10 of 44 patients needed psychiatric treatment for severe psychological disorders. While many of these patients are completely normal there is a small subgroup with severe and often occult psychological problems who need to be very carefully investigated and managed prior to surgery. If a colectomy and ileorectal anastomosis fails there is the possibility of creating a stoma.

Chronic idiopathic intestinal pseudo-obstruction

This is a rare group of patients who have severe constipation and a dilated colon, but also dilatation of the upper intestine. In one-third of patients the disorder shows Mendelian inheritance. Histological studies have demonstrated abnormalities of nerve plexuses and smooth muscle in most patients. Treatment is usually with prokinetic agents to increase intestinal motility. A number of patients also require nutritional support.

Evacuatory disorders

Some patients will give a clear history of an evacuatory disorder. Despite what seems to be a normal propulsive effort by the colon with stool in the rectum, the patient is unable to expel it. Evacuatory disorders have been given a number of names including outlet obstruction, anismus, pelvic floor spasm, the solitary ulcer syndrome, rectal intussusception, and the descending perineum syndrome.

The patient gives a characteristic history of evacuatory difficulty. This may involve long periods of time sitting straining at stool, manipulation of the anal margin, pressure on the perineal body, and, in female patients, pressure on the posterior wall of the vagina. In some patients the distress becomes so extreme that the anal canal is digitated and the faecal material manually removed.

The cause of this condition is not known: there may be a number of underlying problems ranging from an inappropriate contraction of the pelvic floor to an occult rectal intussusception. Physiological studies usually show a normal anal sphincter and rectum but the key investigation is the evacuation proctogram. A mixture of some form of starch and barium or a dilute barium paste is inserted into the rectum and the patient is then asked to strain as at stool. The time taken to expel the simulated stool is noted and any abnormalities in the architecture of the rectal wall recorded (Fig. 4).

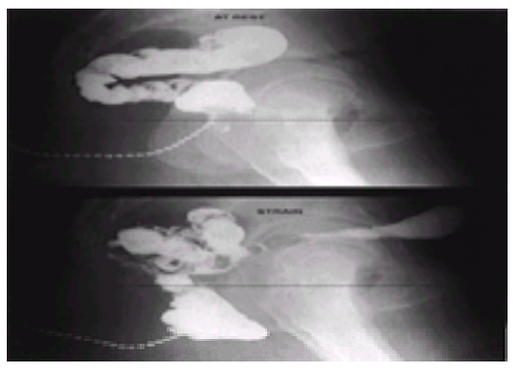


Fig. 4. Evacuation proctogram. The patient is straining and unable to pass any barium paste from the rectum.

The surgical options include rectal prolapse repair, anorectal myectomy, lateral division of the puborectalis, repair of a rectocele, and colostomy. The demonstration of spasm of the pelvic floor led Kamm et al. (1988) to use lateral division of the puborectalis muscle for the treatment of 15 patients with this condition: four of them improved while three had mild incontinence. The operation has now largely been abandoned. Pinho et al. (1989) used a variation of a procedure advocated for patients

with Hirschsprung's disease. A long myectomy of the internal anal sphincter and the rectal smooth muscle wall was made in the midline, posteriorly. There is, however, no evidence that these patients have an abnormality of the smooth muscle of either the anal sphincter or the rectum. In 57 patients with a 24-month follow-up there was no functional improvement in 67 per cent and 10 per cent of the patients developed mild incontinence. When an occult rectal intussusception is found on evacuation proctogram it is tempting to think that a rectopexy rather as for a rectal prolapse procedure would solve the problem. However, in a group of 17 patients reported by Orrom et al. (1991), only four treated by rectopexy required no further treatment, five went on to require a subtotal colectomy, two had an ileostomy, and one a sphincterotomy. Only two patients had a successful result. This is an extremely difficult group of patients to manage satisfactorily. Often a careful explanation of the problem and reassurance that digitation, for example, does not do any harm will suffice. When it is impossible to distinguish between slow transit constipation and an evacuation disorder, a loop ileostomy will allow time for further evaluation. If it is quite clear that there is an evacuatory disorder it is worth creating a colostomy and using colostomy irrigation as a semipermanent solution.

Further reading

Chrotakakou-Faliakou E, Kamm MA, Roy AJ, Storrie JB, Turner IC. Biofeedback provides long term benefit for patients with intractable, slow and normal transit constipation. *Gut* 1998; **42**: 517–21. [Outcome of 100 patients undergoing biofeedback for idiopathic constipation.]

Elliot MS, Todd IP. Adult Hirschsprung's disease: results of the Duhamel procedure. *British Journal of Surgery* 1985; **72**: 884–5. [Results of Duhamel operation in 39 adults with Hirschsprung's disease.]

Gattuso JM, Kamm MA. Clinical features of idiopathic megarectum and idiopathic megacolon. *Gut* 1997; **41**: 93–9. [A study highlighting differences between idiopathic megarectum and idiopathic megacolon.]

Hinton JM, Lennard Jones JE, Young AC. A new method for studying gut transit times using radio opaque markers. *Gut* 1969; **10**: 842–7.

Kamm MA, Hawley PR, Lennard Jones JE. Outcome of colectomy for severe idiopathic constipation. *Gut* 1988; **29**: 969–75. [Results of colectomy in 44 women with idiopathic constipation.]

Kamm MA, Hawley PR, Lennard Jones JE. Lateral division of the puborectalis muscle in the management of severe constipation. *British Journal of Surgery* 1988; **75**: 661–3. [Outcome of 18 patients undergoing lateral division of puborectalis.]

Lubowski DZ, Chan FC, Kennedy ML, King DW. Results of colectomy for severe slow transit constipation. *Diseases of the Colon and Rectum* 1996; **39**: 23–9. [Outcome of colectomy in 59 patients with slow transit constipation.]

Orrom WJ, Bartolo DC, Miller R, Mortensen NJ, Roe AM. Rectopexy is an ineffective treatment for obstructed defaecation. *Diseases of the Colon and Rectum* 1991; **34**: 41–6. [Effect of rectopexy in 17 patients with obstructed defaecation.]

Pinho M, Yoshioka K, Keighley MRB. Longterm results of anorectal myectomy for chronic constipation. *British Journal of Surgery* 1989; **76**: 1163–4. [Results of 57 patients undergoing myectomy for constipation.]

Preston DM, Lennard Jones JE. Severe chronic constipation of young women: ‘idiopathic slow transit constipation'. *Gut* 1986; **27**: 41–8. [Comparison of symptoms in 64 women with constipation with age-matched controls.]

Stabile G, Kamm MA, Hawley PR, Lennard Jones JE. Results of the Duhamel operation in the treatment of idiopathic megarectum and megacolon. *British Journal of Surgery* 1991; **78**: 661–3. [Results of Duhamel procedure in 20 patients with idiopathic megarectum or idiopathic megacolon.]

Stabile G, Kamm MA, Hawley PR, Lennard Jones JE. Colectomy for idiopathic megarectum and megacolon. *Gut* 1991: **32**: 1538–40. [Outcome of colectomy in 40 patients with idiopathic megarectum or idiopathic megacolon.]

26.7 Angiodysplasia

Bruce D. George and Paul C. Shellito

Introduction
Pathology
Aetiology
Clinical presentation and course
Diagnosis
Management
Further reading

Introduction

Angiodysplasia is an acquired submucosal arteriovenous malformation which may cause lower gastrointestinal bleeding in elderly patients. It is often difficult to diagnose, and many aspects of the disease are unclear; reported statistics about incidence and results of treatment differ widely. Angiodysplasia has been reported to account for between 2 and 60 per cent of adult patients with lower gastrointestinal bleeding; it probably causes the majority of incidents of major lower gastrointestinal bleeding in older patients, with diverticulosis accounting for most of the remainder. There is also little agreement about the prevalence of angiodysplasia in adults; estimates range from less than 1 to 30 per cent, but the true figure is probably 1 to 5 per cent.

Pathology

Typical angiodysplastic lesions are 0.5 to 1.0 cm in diameter, bright red, flat or slightly raised, and covered by very thin epithelium (Fig. 1). Between 70 and 90 per cent of the lesions appear in the right colon, although they may also be found in the distal colon, and can also occur, less often, anywhere else in the gastrointestinal tract. Most patients have two or three lesions. Macroscopic pathological examination will usually miss angiodysplasia unless special injection techniques are used. The mesenteric vessels of a resected specimen may be injected with barium followed by radiography, or the vessels may be perfused with silicone rubber followed by tissue clearing with methyl salicylate (Fig. 2), so that only the vascular network remains opaque. Microscopically, angiodysplasia is characterized by thin-walled submucosal veins communicating with clusters of dilated mucosal capillaries (Fig. 3).

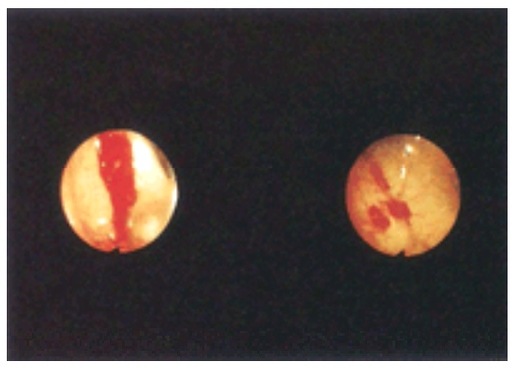


Fig. 1. Typical angiodysplasia lesions in the caecum. At right, prompt bleeding after slight surface trauma is seen.

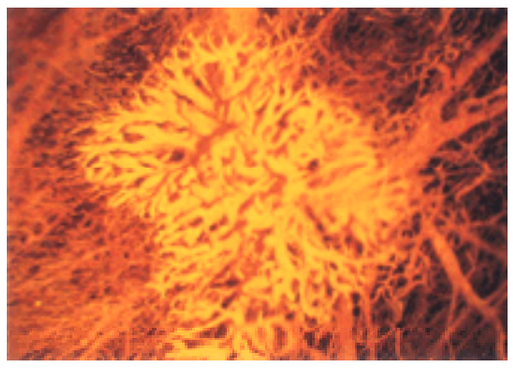


Fig. 2. Angiodysplasia after silicone injection and tissue clearing. A dense collection of abnormal submucosal vessels is demonstrated.

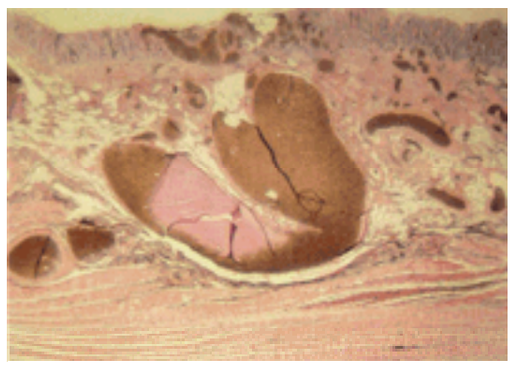


Fig. 3. Histopathology of angiodysplasia. Barium has been injected intravascularly, which stains dark brown in this slide. Submucosal veins are seen, as well as an overlying smaller plexus of mucosal vessels

Aetiology

The aetiology of angiodysplasia is unknown. It is generally agreed to be a degenerative disease as it is rarely encountered in young patients. High tension in the colonic wall and weak supportive connective tissue have been implicated as contributing factors. It is argued that high colonic wall tension leads to obstruction to submucosal venous outflow, causing dilatation first of submucosal veins, and then capillaries and precapillary sphincters, leading to small arteriovenous malformations. Since the right colon has the widest diameter in the large bowel, the law of LaPlace predicts that wall tension is greatest there. Thus, the intramural venous obstruction would also be maximal in that location, which could account for the right-sided predilection of angiodysplasia. Alternatively, the high right colonic wall tension may produce relatively low mucosal perfusion leading to chronic submucosal shunting.

The thin wall of the right colon may provide minimal support for the microscopic vasculature, leading to foci of dilatation which degenerate into arteriovenous malformations. Recently, a deficiency of collagen type IV has been reported in mucosal vessels in angiodysplasia.

Clinical presentation and course

Angiodysplasia may be asymptomatic or may cause gastrointestinal bleeding. The amount of bleeding may vary from major to recurrent minor bleeds or positive faecal occult blood tests only. Gross haemochezia or maroon stools is a typical presentation. The natural history of angiodysplasia is uncertain. Because the bleeding occurs from tiny capillaries and venules covered by very thin epithelium, episodes tend to be self-limiting, but chronic and recurrent. Massive haemorrhage with hypotension is unusual. In contrast, diverticular bleeding, which comes from one or two larger eroded arterioles, is more likely to be acute and massive. After a clot has formed around the responsible diverticular vessel, recurrent bleeding is less likely than with angiodysplasia. Patients with angiodysplasia may have a long history of recurrent idiopathic gastrointestinal bleeding, with multiple unrevealing prior investigations, possibly including laparotomy. The arteriovenous malformations are too small to cause haemodynamic changes (in the absence of bleeding), pain, or perforation.

Patients are usually over 60 years of age. Associations with cardiovascular disease, especially aortic stenosis and chronic renal failure have been reported but are unproved. Angiodysplasia is similar but unrelated to inherited gastrointestinal arteriovenous malformations such as hereditary haemorrhagic telangiectasia (Osler–Weber–Rendu syndrome).

Diagnosis

Angiodysplasia has largely been recognized and characterized only since the advent of colonoscopy and mesenteric angiography. Because the lesions are small, flat, and not transmural, they are overlooked by barium enema radiography or inspection of the serosal surface at laparotomy. Endoscopic biopsies are rarely helpful, because the mucosal lesion is variable, and the biopsy forceps do not usually reach as deep as the abnormal submucosal veins.

The appearances at angiography reflect the theory of pathogenesis. In 90 per cent of cases a large, tortuous, slowly draining submucosal vein is found; a mucosal vascular tuft is seen in 70 per cent; 60 per cent of cases show an early draining dilated vein, which correlates with the final stage of formation of the arteriovenous malformation. Small angiodysplasias can be quite inconspicuous on angiography, unless active haemorrhage is taking place and extravasation of contrast is demonstrated. Unfortunately, because bleeding from angiodysplasia is typically episodic, extravasation is not often demonstrable at angiography.

The colonoscopic appearance is pathognomonic (Fig. 1). The lesions are flat or slightly raised, bright red, and 0.5 to 1.0 cm in diameter. Because of the magnification afforded by the colonoscope, the tiny dilated mucosal vessels in and around the lesion can frequently be seen on close inspection, although these small malformations can be overlooked on endoscopy unless the colonic mucosa is quite clean. Polyethylene glycol gut lavage solutions clean out the colon more effectively than older laxative and enema preparations, and make discovery of angiodysplasia more likely. Care must be taken not to confuse angiodysplasia with traumatic ecchymoses (which are seen on withdrawal but not insertion of the colonoscope), and prominent veins (which are bluish, not bright red). Angioplastic lesions may be almost hidden behind a mucosal fold in the right colon, or just proximal to the ileocaecal valve. The diagnosis can be made 80 to 90 per cent of the time using either colonoscopy or angiography. Endoscopy is probably the more sensitive and specific technique, but opinions differ, depending upon local expertise, and upon which method is accepted as the standard by which to judge the other. Both angiography and colonoscopy may produce false-positive results. Colonoscopy is less invasive than angiography, and also has the advantage of providing biopsies if desired. Furthermore, the entire colon and sometimes the terminal ileum can be meticulously assessed to exclude other diseases or detect synchronous angiodysplasia.

Management

The treatment of angiodysplasia depends on the amount of bleeding and the extent and site of the lesions. Aymptomatic cases do not need any treatment. At the opposite end of the spectrum, massive haemorrhage requires resuscitation and thorough investigation, including angiography. If an angiodysplastic lesion is identified as the cause of massive bleeding, appropriate surgical resection is indicated. When bleeding from angiodysplasia is slow or intermittent, selection of optimum treatment is more difficult. A variety of options are available (Table 1), although there are no comparative trials. Surgical resection of the affected segment of bowel, usually the right colon, is usually successful in stopping bleeding. Rebleeding occurs in 15 to 25 per cent of surgically treated patients, usually because of overlooked arteriovenous malformations in other areas of the bowel, or an error in diagnosis. If preoperative evaluation has reasonably convincingly demonstrated right-sided angiodysplasia, and left-sided diverticulosis is found at laparotomy, it is still acceptable to limit the resection to the right colon. If no specific treatment is undertaken, at least half of the patients with angiodysplasia will rebleed. The major disadvantage of resection is the associated morbidity and mortality risk, particularly in an elderly population with coexisting medical diseases. Endoscopic coagulation therapy, either laser photocoagulation (Fig. 4) or electrocoagulation (Fig.), is an attractive alternative unless the lesions are large or numerous. If a laser is to be used, argon is preferable to neodymium–yttrium aluminium garnet (**Nd:YAG**). The light wavelength from an argon unit is preferentially absorbed by red pigment and also penetrates more superficially than does light from a Nd:YAG machine. Both characteristics are nicely suited to treatment of superficial blood vessel malformations. Whether using laser photocoagulation or hot biopsy electrocoagulation, it is best to start at the periphery of the lesion and progress toward the centre. The mucosa should be cauterized until it is white, not black. Coagulation of large angiodysplasias should be carried out in stages, several weeks apart, to minimize the risk of perforation. The colon and terminal ileum should be searched carefully for synchronous lesions. Rebleeding occurs in only 10 to 30 per cent of patients, and this is usually from lesions overlooked or incompletely coagulated, making early follow-up colonoscopy prudent. Nevertheless, even in patients who rebleed after endoscopic treatment of arteriovenous malformation of the gastrointestinal tract, the frequency of haemorrhage episodes and number of transfusions declines. Perforation follows in up to 7 per cent of treatments, usually when a Nd:YAG laser is used.

1. Surgery	Resection
2. Endoscopic	Electrocoagulation, Injection therapy, Contact probes, Laser
3. Angiographic	
4. Pharmacological	Oestrogen–progesterone, β-Blockers, Somatostatin

Table 1 Treatment options in angiodysplasia

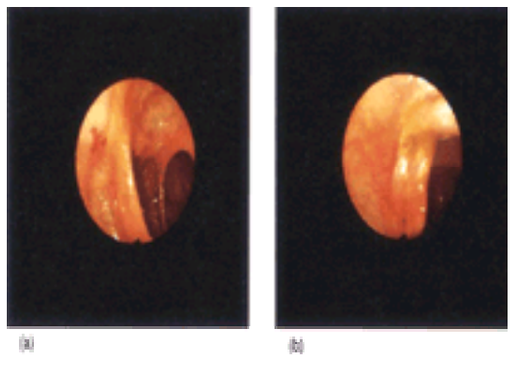


Fig. 4. Nd:YAG laser photocoagulation of angiodysplasia. (a) Right colon angiodysplasia. (b) After cauterization.

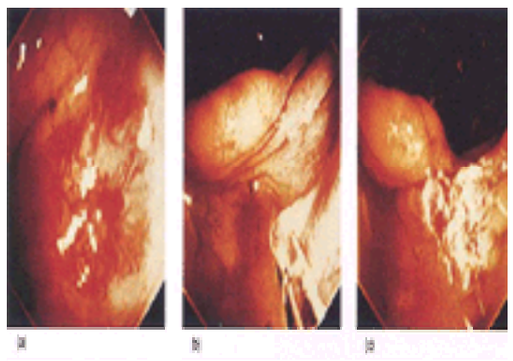


Fig. 5. Hot biopsy coagulation of angiodysplasia. (a) Close up view of the lesion. (b) The angiodysplasia has been grasped and pulled up by the biopsy forceps, while electrocoagulation is carried out. (c) The mucosa has been released (without biopsy), leaving white coagulated mucosa visible. The ileocaecal valve is just above and to the left of the coagulated area.

When angiodysplastic lesions are very extensive or not amenable to endoscopic therapy, drug treatment may be effective. Oestrogen–progesterone preparations, b-blockers, and somatostatin have been reported to reduce bleeding in a small number of cases. One small randomized trial of ethinyloestradiol (0.05 mg) and norethisterone (1 mg) showed a benefit compared with placebo in the treatment of bleeding gastrointestinal vascular malformations, although most of the cases had Osler–Weber–Rendu disease rather than angiodysplasia.

Further reading

Foutch PG, Rex DK, Lieberman DA. Prevalence and natural history of colonic angiodysplasia among healthy asymptomatic people. *American Journal of Gastroenterology* 1995; **90**: 564–7. [Review of screening colonoscopies in 964 asymptomatic people suggesting prevalence of angiodysplasia of 0.83 per cent.]

Gupta N, Longo WE, Vernava AM. Angiodysplasia of the lower gastrointestinal tract: an entity readily diagnosed by colonoscopy and primarily managed non-operatively. *Diseases of the Colon and Rectum* 1995; **38**: 979–82. [Recent review of 32 patients suggesting that most patients do not require surgical intervention.]

Imperiale TF, Ransohoff DF. Aortic stenosis, idiopathic gastrointestinal bleeding and angiodysplasia: is there an association? *Gastroenterology* 1988; **95**: 1670–6. [Detailed review of literature casting doubt on any association between aortic stenosis and angiodysplasia.]

Richter JM *et al.* Angiodysplasia: natural history and efficacy of therapeutic interventions. *Digestive Diseases and Sciences* 1989; **34**: 1542–6. [Report of 101 patients with angiodysplasia treated either surgically, endoscopically, or with supportive medical care only suggesting superior results with surgery.]

Van Cutsem E, Rutgeerts P, Vantrappen G. Treatment of bleeding gastrointestinal vascular malformations with oestrogen–progesterone. *Lancet* 1990; **335**: 953–55. [Small randomized trial (*n* = 10) suggesting benefit of oestrogen–progesterone over placebo in the treatment of bleeding gastrointestinal vascular malformations.]

26.8 Prolapse of the rectum

Timothy A. Cook and Neil Mortensen

Incidence
Aetiology
Partial prolapse
Complete prolapse
Clinical features
Prolapse in children
Prolapse in adults
Differential diagnosis
Complications
Treatment
Prolapse in children
Rectal prolapse in adults
Descending perineum syndrome
Aetiology
Clinical features
Management
Solitary ulcer syndrome
Clinical features
Treatment
Further reading

The term ‘rectal prolapse' implies a full-thickness circumferential descent of the rectum through the anus. A prolapse of only the rectal mucosa is called incomplete or mucosal prolapse. A variant of complete rectal prolapse has been described in which the upper rectum prolapses into the middle or lower rectum without actually reaching the anal canal. This is called an internal prolapse, or intussusception of the rectum.

Incidence

Rectal prolapse occurs at the extremes of life. Complete rectal prolapse is found chiefly in elderly female patients: 85 per cent of adults with full-thickness rectal prolapse are women and the incidence is maximal in the fifth decade and upwards. Many patients are of very advanced age, being in their eighties or nineties. In men, though the incidence is much lower, rectal prolapse presents throughout the age range or may be more common in the second and third decades of life. Mucosal prolapse is most common in young children.

The increased incidence in female patients might imply an effect of childbirth on the pelvic floor, but in many series half of the patients are childless. Furthermore, uterine and rectal prolapse only rarely occur together.

Aetiology

Partial prolapse

In children there is usually some promoting factor such as diarrhoea, constipation, or bad defecatory habits. A chronic cough or severe coughing attacks can cause a rectal prolapse. In adults prolapsed mucosa is associated with third-degree haemorrhoids, and in older patients it may occur as a result of a weak anal sphincter. The descending perineum syndrome has also been described in which the anterior wall of the rectum causes an evacuatory disorder.

Complete prolapse

Anatomical findings

Complete rectal prolapse is associated with a collection of distinct anatomical changes (Table 1). The most striking of these is the abnormally deep rectovaginal, or in males rectovesical, pouch. This gave rise to the earliest theory of the cause of rectal prolapse, namely that it was a form of sliding hernia, the pouch of Douglas being the hernial sac pressing the anterior rectal wall into the rectal lumen.

Diastasis of the levators
Deep pouch of Douglas
Redundant rectosigmoid
Lax lateral and posterior attachments
Loss of horizontal lie
Elongated mesorectum

Table 1 Anatomical findings in complete rectal prolapse

Radiological studies have suggested that an intussusception was the lead point of a rectal prolapse found at 6 to 8 cm from the anal verge. In 60 per cent of patients this was anterior, in 32 per cent annular, and in 8 per cent posterior. An enterocele developed late, if at all (Fig. 1).

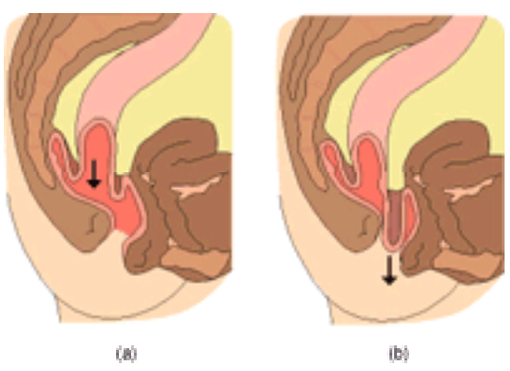


Fig. 1. Diagram showing a rectal intussusception (a) passing into the lower rectum (arrow). The anterior portion (b) passes through the anal canal (arrow) to become a full-thickness rectal prolapse.

Most patients with complete prolapse have a weak atonic anal sphincter, and many also have a weak pelvic floor. It is not clear whether this is a cause or an effect of the prolapse: some patients have a large prolapse, but a normal anal sphincter once the prolapse has been reduced, while others with prolonged marked changes in the pelvic floor may have had a rectal prolapse for only weeks or months.

Although rectal prolapse may occur in patients with cauda equina lesions, most patients with rectal prolapse have no obvious neurological abnormality. Studies of anorectal physiology show a reduction of resting anal pressure and maximum voluntary contraction pressure, impaired rectal sensation, absence or abnormalities of the rectosphincteric reflex, an increase in pudendal nerve latency, and histological evidence of muscle denervation in the external anal sphincter. These changes are most apparent in those patients who have a combination of prolapse and incontinence. Denervation of the voluntary sphincter muscles may be due to stretching of the pudendal nerves during descent of the rectum in the course of prolapse. The prolapse is thought to result from an intussusception which starts well above the pelvic floor, and therefore these changes are probably a result of the rectal prolapse rather than the cause of it. Around 20 per cent of people with anterior mucosal prolapse develope a full-thickness rectal prolapse within 10 years.

The aetiology of rectal prolapse is still not clear. Known initiating factors include diarrhoea, constipation, and disorders of the pelvic floor which, acting together with predisposing factors such as a deep pouch of Douglas, a redundant sigmoid colon, and a weak pelvic floor, bring about intussusception as the first stage in the genesis of a rectal prolapse.

Clinical features

Prolapse in children

Rectal prolapse in a child usually becomes apparent during defecation. It may reduce itself spontaneously sometimes, while at other times it may need to be replaced digitally. There is sometimes discomfort and a slight discharge of mucus or blood, but the child is usually perfectly well and normally continent.

On examination the prolapse consists of a ring of mucosa projecting 2 to 4 cm beyond the perianal verge. Gentle palpation between finger and thumb usually reveals that the prolapse consists of only two layers of mucosa. Very little pressure is required to reduce the prolapse and the anal sphincter appears normal. If there is only a history of prolapse, it is important for the surgeon to see the prolapse occurring in the clinic. Rarely the prolapse may be a complete rectal prolapse. Differential diagnosis includes a prolapsed rectal polyp or the apex of an intussusception. In the latter case it is possible to pass the tip of an examining finger along the slit between the intussuscepted bowel and the wall of the anal canal.

Prolapse in adults

Symptoms include those due to the prolapse itself and those due to the weakness of the anal sphincter. The prolapse first becomes apparent during evacuation or during episodes of raised intra-abdominal pressure. When the bowel wall is prolapsed mucus and blood may soil underclothes. There is also a degree of incontinence of faeces and flatus, giving rise to urgency and soiling episodes. Patients often have coincident constipation which aggravates the prolapsed condition.

The anus is usually patulous and there is a degree of mucosal prolapse evident at the anal orifice. Distraction on the anal margin causes the anus to open widely and the patient will often be able to push down a prolapse if it is not immediately apparent (Fig. 2). On palpation the anal sphincter is usually very weak. Voluntary contraction is usually poor and there is blunting of anal and rectal sensation. The mucosa on the apex of the prolapse may show some granularity or even superficial ulceration from repeated trauma caused by contact with underclothing.



Fig. 2. A full-thickness rectal prolapse.

When the prolapse is produced with the patient straining, palpation between finger and thumb reveals the double thickness of tissue; this is especially evident anteriorly where the deep pouch of Douglas and any possible contained loops of small intestine may add to the bulk of the prolapsing rectum. Most prolapses over 5 cm in length are complete, but it can sometimes be difficult to tell whether a shorter prolapse is simply mucosa or a complete procidentia. If the patient is unable to produce a prolapse on the examination couch, the diagnosis can be made on allowing the patient to adopt the usual squatting position secluded in a toilet. Some patients may find it easier to produce the prolapse in these circumstances than in the clinic.

On reducing the prolapse it is important to carry out a sigmoidoscopic examination to exclude any other abnormality within reach of the instrument.

Differential diagnosis

Large third-degree haemorrhoids, a large polypoid tumour of the rectum, sigmoid colon prolapsing and emerging at the anus, or a purely mucosal prolapse must all be differentiated from a complete procidentia.

Complications

Most patients can reduce the prolapse readily, although urgent admission may be required if the bowel becomes oedematous and cannot be reduced. It may be possible to reduce such a prolapse manually in the clinic with the patient lightly sedated. In the rare situation of a gangrenous rectal prolapse an emergency perineal recto-sigmoidectomy may be necessary. Proctitis, ulceration, and rarely severe haemorrhage can occur, but these are not usually clinically important.

Treatment

Prolapse in children

Rectal prolapse in children is a self-limiting disease and usually responds to conservative measures. Correction of bowel habit and small doses of laxatives may be all that is necessary. Occasionally a submucosal injection of phenol or alcohol may be required to secure fixation of the prolapse.

Rectal prolapse in adults

Partial prolapse

Minor mucosal prolapse can be treated in the same way as third-degree haemorrhoids, either by conservative measures or surgical excision. In older patients, in whom a partial prolapse complicates a weak anal sphincter, a postanal repair operation may be necessary.

Complete prolapse

Innumerable methods for fixing a rectal prolapse have been described in the surgical literature, and they can be divided into perineal, sacral, and abdominal procedures. A procedure should be chosen with consideration for the age and fitness of the patient. There is perhaps no single ideal operation, and the art of prolapse management is to match the procedure to the patient. Trials are currently being performed in the United Kingdom and United States to determine the best operation.

Perineal procedures—partial excision of the rectum through the anus

These operations are compromise procedures. They are less invasive than abdominal procedures but remove the reservoir function of the rectum which must have considerable physiological effects.

Delorme's operation This operation is becoming increasingly popular. With the bowel fully prolapsed a circular incision is made through the mucosa of the prolapse 1 cm from the dentate line (Fig. 3). Infiltration with dilute adrenaline solution helps indicate the submucosal plane. The mucosa is then dissected from the underlying muscle coat as a sleeve until the apex of the prolapse is reached, and the dissection is carried on down into the prolapse as far as possible (Fig. 4). This leaves the outer aspect of the prolapse without any mucosal covering. The underlying muscle coat is imbricated with a series of longitudinal sutures to bunch up or reef it and bring the edges of the mucosa together (Fig. 5). The mucosa is then sutured with absorbable sutures and the prolapse gently reduced. The procedure has been used successfully in elderly patients. It has a particular place in this group and can also be used in patients with rectal prolapse complicating chronic ambulatory peritoneal dialysis.



Fig. 3. Delorme's operation. An incision is made in the mucosa 1 cm above the dentate line with the rectum fully prolapsed.

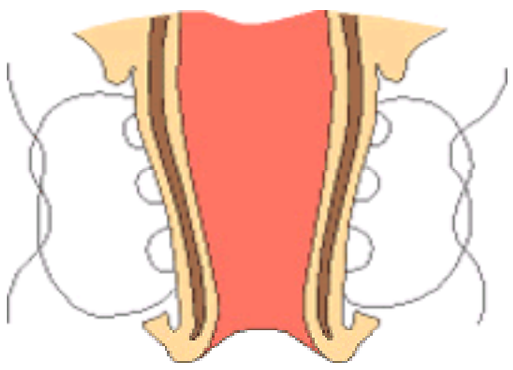


Fig. 4. Delorme's operation. The mucosa is stripped off the underlying muscle as far up as possible and the muscle coat is imbricated with a series of longitudinal sutures.

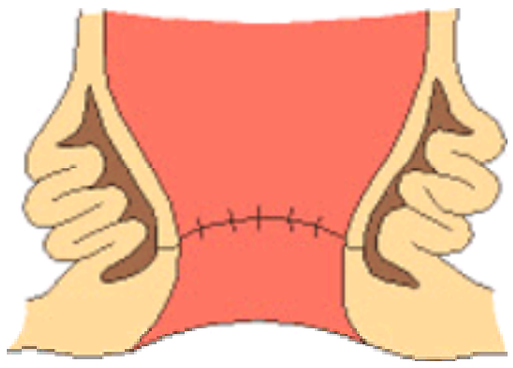


Fig. 5. Delorme's operation. The imbricating sutures are tied and the mucosal edges sutured.

Where there is coincident anal sphincter weakness it is possible to carry out a postanal repair at the same time. Only a small series of cases have been reported. Delorme's procedure is a compromise operation and recurrences occur in 10 to 25 per cent of cases. However, morbidity is low and the procedure can be repeated should a recurrence develop. It may also have a further advantage in that it does not cause constipation or an evacuation disorder, and there is no danger of pelvic nerve damage.

Rectosigmoidectomy With the patient in the lithotomy or jack-knife position the prolapse is pulled down, and the outer of the two tubes of rectal wall is divided circumferentially just above the dentate line. The inner tube of rectum can then be drawn down bringing the distal sigmoid to the anal canal, where it is divided and sutured to the anal remnant with absorbable sutures (Fig. 6). The specimen resected should be 15 to 20 cm in length, leaving no slack rectum or distal colon to allow further prolapse (Fig. 7). Popularity for this procedure has waned in the United Kingdom, but it is still very popular in the United States.

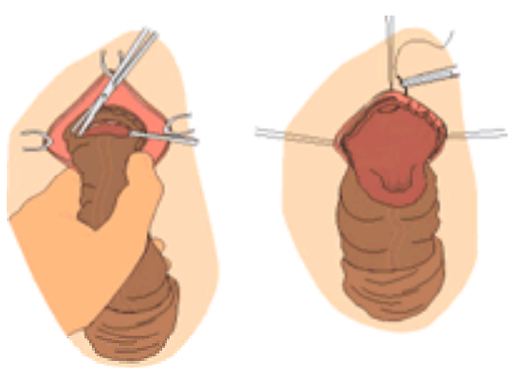


Fig. 6. Perineal rectosigmoidectomy. The inner tube of rectum can be brought down bringing the distal sigmoid to the pelvic floor.

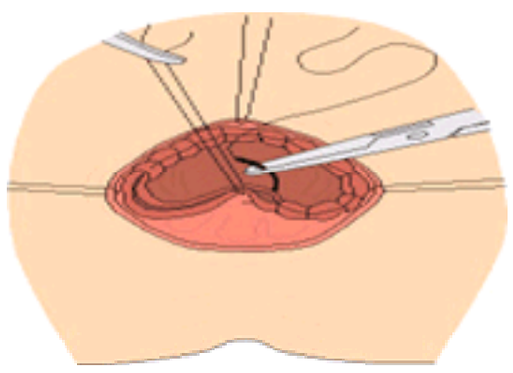


Fig. 7. Perineal rectosigmoidectomy. The rectosigmoid is divided and sutured to the anal remnant leaving no slack for further prolapse.

Experience in the United Kingdom in the 1960s was disappointing with a high recurrence rate. The procedure has been modified to include suturing of the levator muscles anterior to the rectum. This is a plication of the puborectalis sling which attempts to improve sphincter function postoperatively and to reduce the incidence of recurrence. Williams *et al.* described recurrences in 11 of 104 patients with dramatic improvements in continence if levator plication was added to the procedure. However, in a small randomized study, Deen *et al.* found better functional results following resection rectopexy than after rectosigmoidectomy

Encircling the anal orifice with foreign material

Variations of this procedure have been popular because of their simplicity.

The Thiersh operation Incisions are made in front of and behind the anal margin to allow the passage of wire, stout nylon, or even Silastic around the anal sphincter. It is usually overlapped and sewn together anteriorly to provide the right amount of tension (Fig. 8). The procedure can be carried out under regional or local anaesthesia in elderly frail patients. In principle, the technique works by supporting the reduced prolapse and causing a local reaction which induces fibrosis and stenosis of the anal canal.

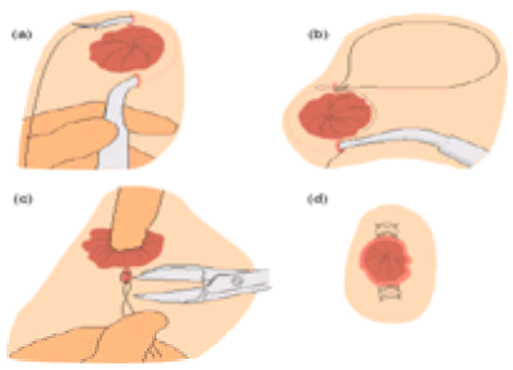


Fig. 8. The Thiersch operation. Incisions are made in front of and behind the anus and an encircling suture placed under correct tension.

In our experience this procedure can work quite well for small prolapses, but if the prolapse is large the wire or nylon tends to migrate and cut through the skin. Recurrences are common and the Thiersh operation cannot be regarded as anything more than a temporary solution.

Variations on the technique have been described, including placing a Silastic band around the top of the external sphincter below the levator muscles, but there is no evidence that this produces better results.

Postanal repair and perineorrhaphy In patients with a weak anal sphincter and marked symptoms of incontinence it is tempting to use a sphincter plication procedure to hold in the prolapse. Unfortunately this is rarely successful for long and the technique can only be used for small prolapses.

Sacral procedures

A number of operations have been described which exploit the Kraske approach alongside the coccyx and behind the sacrum. With the patient in the jack-knife position an incision is made over the coccyx and parasacral area, giving access to the presacral and postrectal space. The rectum is mobilized, shortened by imbricating sutures, and foreign material is placed in the presacral space. None of these procedures has stood the test of time and become particularly popular.

Abdominal procedures

The general principle of these operations is to repair the prolapse by mobilizing the rectum and fixing it to the sacrum. Repair of the pelvic floor is sometimes performed simultaneously.

The Pemberton–Stalker operation In this procedure the mobilized rectum is lifted out of the pelvis and the sigmoid colon is then fixed in an elevated position to the peritoneum of the anterior abdominal wall or the pelvic bone, and in female patients to the uterus. The procedure was popularized at the Mayo Clinic in the 1940s and gave a recurrence rate of 35 per cent.

The Roscoe–Graham procedure Variations on this procedure have been described by a number of surgeons. The rectum is fully mobilized but in addition the pelvic floor is plicated either in front of or behind the rectum. The aim of this part of the procedure is to maintain the rectum in an elevated position and to improve postoperative continence. However, it has not gained widespread acceptance and it has been suggested that the operation provided no more than a temporary buttress which after several months could no longer be palpated.

The Frykman Goldberg procedure Here the rectum is fully mobilized preserving the lateral ligaments which are pulled up and made taught. These tight bands are then sutured to the presacral fascia to keep the bowel up out of the pelvis. After mobilizing the descending colon the sigmoid loop is resected. Watts *et al.* reported 138 patients treated in this way, with a recurrence rate of 1.9 per cent.

Ripstein's operation The mobilized rectum is fixed to the sacral hollow by means of a sling of Teflon mesh 5 cm wide. This is passed around the rectum and the ends are sutured behind it to the fascia on the front of the sacrum, just below the promontory. In addition, a few sutures are passed between the edges of the Teflon and the anterior and lateral rectal walls. Roberts *et al.* reported 135 patients treated in this way at the Lahey clinic. There was one perioperative death, and a 9.6 per cent recurrence rate. In a postal survey of members of the American Society of Colon and Rectal Surgeons the recurrence rate was 2.3 per cent and significant problems with constipation and faecal impaction occurred in 6.7 per cent of patients. Stricturing at the site of the sling occurred in 2 per cent of patients.

Ivalon rectopexy Following early experience with the use of polyvinyl alcohol sponge in the repair of hernias, Wells used a 3-mm thick sheet of Ivalon measuring 10 ×

15 cm. After a full mobilization of the rectum the Ivalon is placed behind it, partially wrapping the full circumference of the rectum, and resutured, fixing it and the sponge to the sacral hollow. In 101 cases reported by Penfold and Hawley from St Mark's Hospital there were no operative deaths but four patients developed postoperative sepsis and the Ivalon had to be removed in one of these: sepsis in the implanted foreign body is a recognized hazard of this procedure. After 2 to 10 years, three patients had developed a recurrence and 31 suffered mucosal prolapse. Continence was improved in about 30 per cent of the patients postoperatively. Reasonable results have also been reported in a group of young patients with rectal prolapse treated by the same operation. Ivalon sponge induces a fibrotic reaction and this can lead to reduced compliance of the rectum. Ivalon sponge has largely been supeseded by Marlex or Mersiline mesh. Recurrence rates following posterior rectopexy are around 2 per cent.

Presacral rectopexy and simple suture Another alternative to the use of an implant of either Teflon or Ivalon is simply to mobilize the rectum and fix it to the sacral promontory. In a small prospective study of 63 patients, Novell *et al.* found suture rectopexy to be equivalent to Ivalon sponge at a median of 47 months follow-up. It was suggested that the Ivalon rectopexy could be abandoned in favour of the simpler suture technique.

Anterior resection Some surgeons favour a straight anterior resection. Occasionally if the anal sphincter is very weak, an abdominoperineal excision would be justified.

Technique

There are a number of abdominal rectopexy techniques, but the following is perhaps the most widely used.

Patients should undergo a full bowel preparation in case the rectum is damaged during mobilization. Prophylactic intravenous antibiotics should be given. The patient is placed in Lloyd-Davis stirrups, a urinary catheter is passed, and a lower abdominal incision is made. The small bowel is packed off, and the uterus is hitched up with stay sutures. Peritoneal cuts are then made to mobilize the rectum, starting on the right side of the base of the mesosigmoid but a little way up the mesentery so that the peritoneum can be preserved and closed over the repair. The incision is carried down to the bottom of the often deep pouch of Douglas, crossing anteriorly and joining a similar cut on the left side. The presacral space is opened, identifying and preserving the presacral nerves posterolaterally and the left ureter. The posterior dissection can now be carried right down to the pelvic floor (Fig. 9). Anteriorly, the plane between the posterior vaginal wall and the rectum is opened and dissected with a combination of blunt and sharp dissection. There is debate over whether the lateral ligaments should be divided: some evidence suggests that bilateral division results in abnormal rectal function.

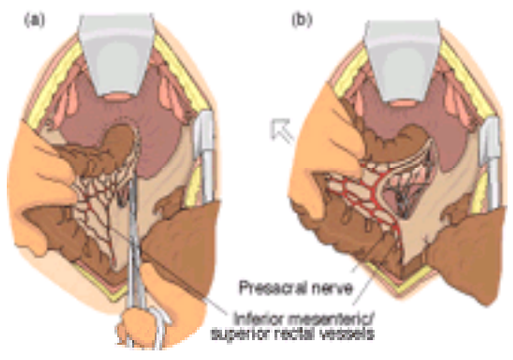


Fig. 9. (a and b) Abdominal rectopexy. Mobilization of the rectum in the mesorectal plane down to the pelvic floor. The lateral ligaments have not been divided.

A 15 × 10 cm sheet of implantable material (Ivalon sponge, Marlex, Mersiline, or Teflon) is fixed to the presacral fascia, care being taken not to puncture middle sacral vessels or pelvic veins (Fig. 10). The use of several sutures increases the risk of damage to sacral veins: two or three probably suffice. Once the implanted material has been attached to the sacral promontory and presacral fascia, either the same sutures or a new set of sutures can be used to attach this in turn to the fully mobilized and stretched up mesorectum, taking in a section of the serosa of the rectal wall at the same time. The implant is then partially wrapped around the lateral sides of the mobilized rectum (Fig. 11), leaving one-third of the anterior circumference of the rectal wall uncovered to minimize the risk of stenosis or functional constipation. Sutures are then placed to fix this in position. The pelvic peritoneum can then be covered over all of this, leaving two suction drains in the cavity to prevent a haematoma. Variations on this procedure include the use of a transverse abdominal incision and limited peritoneal incisions gaining access to the presacral space and the mesorectal plane.

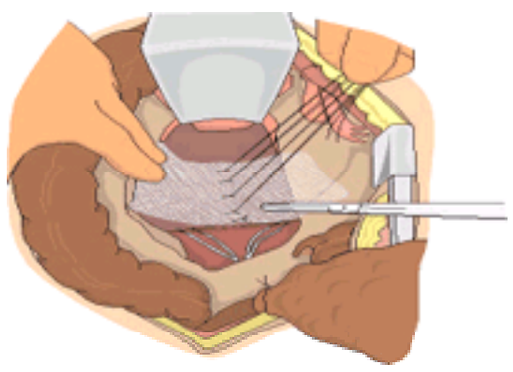


Fig. 10. Abdominal rectopexy. The mobilized rectum is drawn up and a sheet of implantable material is sutured to the sacral promontory.

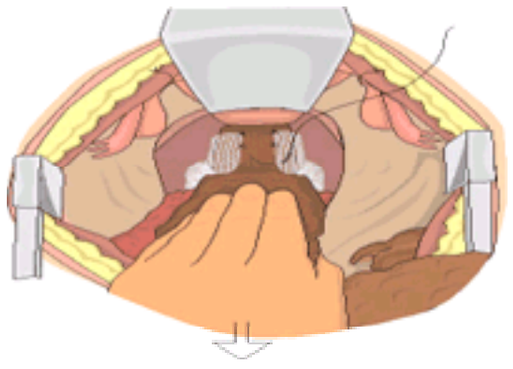


Fig. 11. Abdominal rectopexy. A partial wrap around half the posterior circumference of the rectum fixes it to the sacrum.

The procedure is usually well tolerated, even by elderly patients. Intravenous fluids should be given for 48 h since some patients suffer marked postoperative paralytic ileus. A mild laxative is prescribed from the second or third postoperative day to prevent constipation and straining at stool. There is often an irregularity of bowel habit and continence in the immediate 2 or 3 weeks postoperatively. Postoperative constipation may be a problem in the longer term but combining rectopexy with sigmoid resection may reduce the incidence of this complication.

Laparoscopic rectopexy Over the last few years, small personal series of laparoscopic rectopexy have been published. The principles of the technique are similar to the open operation. The procedure can be performed entirely laparoscopically or the rectum can be mobilized laparoscopically and an open resection performed.

Operating times are longer but hospital stays shorter compared with open operations. However, the long-term outcome needs to be evaluated before its widespread adoption.

Descending perineum syndrome

During straining the anal canal should not descend more than 2 cm below a line joining the ischial tuberosities. In patients with descending perineum syndrome the anal canal at rest lies at a lower level; straining causes the perineum itself to balloon well below the lower margin of the bony pelvis. It is more common in women and may occur at any age, although it is rare before the third decade.

Aetiology

It is probable that excessive straining weakens the pelvic floor muscles and stretches the pudendal nerves supplying the pelvic floor, which then bulges and allows the anterior wall of the rectum to prolapse into the upper anal canal (Fig. 12). This anterior mucosal prolapse results in a feeling of incomplete evacuation, to which the patient will respond by further straining efforts. Gross perineal descent on its own does not always cause symptoms and some patients remain continent.

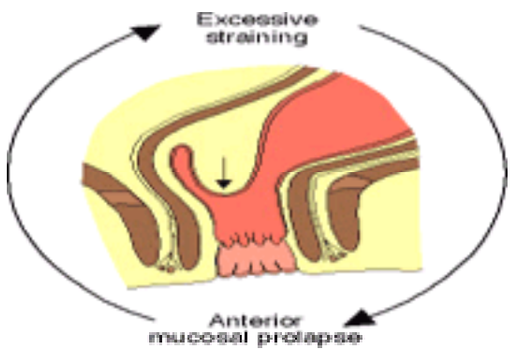


Fig. 12. The vicious circle causing anterior mucosal prolapse and solitary rectal ulcer.

Clinical features

The patient's main symptoms are of intractable tenesmus, but there may also be the passage of mucus and blood. There is often a past history of haemorrhoidectomy, inappropriately performed for what were thought to be haemorrhoidal symptoms. There may be anorectal incontinence, and other patients have a dragging perineal pain. Physical examination usually reveals reduced resting anal tone and descent of the perineum at rest and on straining. Proctoscopy may reveal an anterior rectal mucosal prolapse, and in intractable cases a solitary ulcer of the rectum may develop (see below).

Management

Treatment consists of avoiding further straining by alterations in diet and the administration of bulking agents, suppositories, and enemas. Anterior mucosal prolapse or prolapsing haemorrhoids can be treated by injection therapy or band ligation. Great care must be taken with surgical excision, since the haemorrhoidal tissue may be an important component of the patient's remaining continence.

Solitary ulcer syndrome

The solitary ulcer syndrome is often, but not always, associated with the descending perineum syndrome. It rarely takes the form of a single large depressed ulcer, and is more commonly an area of mucosal change on the anterior wall of the lower rectum. This may be ulcerated in places, but the mucosa is oedematous, heaped up, and bleeds easily on contact. Sometimes these appearances can extend around the whole rectal circumference. Histologically, there is replacement of the lamina propria by fibroblasts and smooth muscle cells, typically arranged at right angles to the muscularis mucosa.

Clinical features

It is now thought that these changes are the result of mucosal trauma. The lower, anterior ulcers are usually associated with the mucosal prolapse, while higher ulcers are associated with intussusception of the rectum. The mucosal changes probably result from a straining effort in the presence of a contracting sphincter muscle which then pinches and damages the prolapsing mucosa. This gives rise to the common symptoms of the passage of mucus and blood, and the thick, ridged, oedematous mucosa gives a sense of anal obstruction and pain, resulting in prolonged and excessive straining. There may be multiple fruitless attempts at defecation.

These symptoms and the appearance of a rectal ulcer can be easily mistaken for a carcinoma and it is important to obtain a biopsy sample.

Treatment

This is usually conservative. Straining efforts are discouraged. If symptoms become intractable they may be improved by an abdominal rectopexy, fixing both the anterior and posterior walls of the rectum.

Further reading

Deen KI, Grant E, Billingham C, Keighley MRB. Abdominal resection rectopexy with pelvic floor repair versus perineal rectosigmoidectomy and pelvic floor repair for full-thickness rectal prolapse. *British Journal of Surgery* 1994; **81** : 302–4. [Small randomized study comparing abdominal resection rectopexy with perineal rectosigmoidectomy.]

Eu KW, Seow-Choen F. Functional problems in adult rectal prolapse and controversies in surgical treatment. *British Journal of Surgery* 1997; **84** : 904–11. [Review article on current management of rectal prolapse.]

Novell JR, Osborne MJ, Winslet MC, Lewis AA. Prospective randomized trial of Ivalon Sponge versus sutured rectopexy for full-thickness rectal prolapse. *British Journal of Surgery* 1994; **81** : 904–6. [Prospective randomized trial of 31 patients treated with Ivalon sponge compared with 32 patients treated by sutured rectopexy.]

Penfold JCB, Hawley PR. Experiences of Ivalon-sponge implant for complete rectal prolapse at St Mark's Hospital. *British Journal of Surgery* 1972; **59** : 846–8. [Outcome of 101 patients undergoing Ivalon rectopexy.]

Roberts PL, Schoetz DJ, Coller JA, Veidenheimer MC. Ripstein procedure. Lahey clinic experience: 1963–1985. *Archives of Surgery* 1988; **123** : 554–7. [Results of the Ripstein procedure in 135 patients.]

Stevenson AR, Stitz RW, Lumley JW. Laparoscopic-assisted resection rectopexy for rectal prolapse: early and medium follow-up. *Diseases of the Colon and Rectum* 1998; **41** : 46–54. [Single centre experience of laparoscopic rectopexy in 34 patients.]

Tobin SA, Scott IH. Delorme operation for rectal prolapse. *British Journal of Surgery* 1994; **81** : 1681–4. [Results of Delorme's procedure in 49 elderly patients.]

Watts JD, Rotherburger DA, Buls JG, Goldberg SM, Nivatvongs S. The management of procidentia. *Diseases of the Colon and Rectum* 1985; **28** : 96–102. [Results of Frykman-Goldberg operation in 138 patients.]

Williams JG, Rothenberger DA, Madoff RD, Goldberg SM. Treatment of rectal prolapse in the elderly by perineal rectosigmoidectomy. *Diseases of the Colon and Rectum* 1992; **35** : 830–4. [Analysis of outcome in 104 patients following perineal rectosigmoidectomy.]

Yoshioka K, Heyen F, Keighley MRB. Functional results after posterior abdominal rectopexy for rectal prolapse. *Diseases of the Colon and Rectum* 1989; **32** : 835–8. [Results of marlex mesh rectopexy in 165 patients.]

27.1 Acute appendicitis

Charles M. Ferguson

Introduction
History
Physical findings
Laboratory examinations
Differential diagnosis
Diagnosis in difficult circumstances
Retrocecal appendicitis
Appendicitis in the elderly
Appendicitis in pregnancy
Management
Management of the appendiceal mass
Antibiotics
Further reading

Introduction

The diagnosis of appendicitis can be difficult, occasionally taxing the skills of even the most experienced surgeon. Likewise, decisions on the management of patients with appendiceal inflammation or abscess can be difficult. The patient with appendicitis must first recognize that he or she has an episode of pain that is unique, and then present to a physician who recognizes the condition. Delays in diagnosis arise from errors on the part of either patient or physician, and all delays complicate the illness.

History

Patients with appendicitis generally present with a typical pattern of distress. The most common initial symptom is pain, typically diffuse, epigastric, or periumbilical in distribution and gnawing in character. Some have no pain, but feel unwell, with a severe upset stomach. Shortly after the initial pain, anorexia, nausea, and occasionally vomiting develop. Anorexia is the most constant symptom of appendicitis. If it is absent, the diagnosis should be questioned. Vomiting, if present, occurs early in the attack, usually several hours after the initial pain. Vomiting occurring before the onset of pain makes a diagnosis of appendicitis questionable, and a viral illness more likely.

The initial pain is due to the early obstruction, dilation, and infection of the appendix. As this visceral pain is mediated through visceral pain fibres, it is poorly localized. When the infection in the appendix becomes established and transmural, the serosa of the appendix and the parietal peritoneum are involved, causing a localized pain from somatic pain fibres in the abdominal wall at the area of the appendix. Because the appendix is generally located one-third of the way from the anterior, superior iliac spine to the umbilicus (McBurney's point), the area of greatest pain is generally in the right lower quadrant.

Patients may present with localized pain in the right upper quadrant from a long appendix, in the left lower quadrant if malrotation is present, and in the anterior wall of the rectum if the appendix is located in the pelvis. The most common location of ‘atypical' somatic pain is the right flank in patients with a rectrocecal appendix. Somatic pain reflects the location of the appendix; if the development of symptoms and signs suggest appendicitis, an atypical location of maximal pain does not rule out the diagnosis.

Many patients with appendicitis have pain on motion and pain present during the trip to hospital may abate when motion ceases. These events do not mean that their illness is resolving: pain on motion is caused by movement of the inflamed appendix against the peritoneum. The patient presenting after perforation of the appendix complains of generalized pain. The history is usually a sequence of generalized pain that became localized to the right lower quadrant and later became generalized.

Less common symptoms include diarrhea, which may occur early or late in the course of appendicitis. Early in the course, patients may have one or two loose bowel movements, or they may have an episode of massive evacuation of normal stool. This sequence represents a response to visceral pain and is usually limited to one or two episodes, rather than the persistent diarrhea caused by viral or bacterial infection. Later in the course of appendicitis, diarrhea may return because of irritation of the rectum by an inflamed pelvic appendix. This diarrhea is mucoid and persistent; it is accompanied by tenesmus and can easily be mistaken for gastroenteritis if proper attention is not paid to the history. Testicular pain or retraction of the testes may occur at any time in the course of appendicitis, the appendix and testicle both being innervated by the tenth thoracic spinal segment.

The concept of recurrent appendicitis is gradually being accepted in the United States, and patients often describe previous episodes of pain that were the same as the present in all aspects except severity. Surveys suggest that this sequence occurs in about 25 per cent of patients.

Physical findings

Tenderness over the site of the appendix is the hallmark of appendicitis. However, tenderness may be absent early in the course of the illness or unelicitable in obese individuals. Tenderness over the appendix is due to inflammation of the serosa of the appendix and the overlying parietal peritoneum; so early in appendicitis there may not be enough inflammation to cause this diagnostic finding, although it is unusual for patients to present this early in the course of their illness. More commonly, difficulty in eliciting tenderness over the appendix arises from its retrocecal location. Patients with a retrocecal appendix may experience some mild right-sided or right-flank tenderness. In extremely obese people, localizing the tenderness is difficult, simply because palpation of the abdomen is cushioned by subcutaneous fat. In addition, the entire panniculus of the abdomen may be caudal to the peritoneal cavity. When the abdomen is palpated in such a patient, the normal anatomical landmarks must be ignored.

While much has been made of tenderness at McBurney's point, this is not as reliable as the presence of tenderness at some point in a patient with a suggestive history. Classically, the area of maximal tenderness will be one-third of the way from the anterior superior iliac spine to the umbilicus, but in fact it will be wherever the appendix is in the individual patient.

Many surgeons rely on the demonstration of local muscular rigidity to make a diagnosis of appendicitis. This rigidity is produced by inflammation of the parietal peritoneum overlying the appendix, and thus takes longer to develop than local tenderness. Many patients with early appendicitis have little or no rigidity of the abdominal wall. Subtle rigidity can be demonstrated by palpation of the left lower quadrant while talking to the patient. The palpating hand is slowly moved toward the right lower quadrant (or area of maximal pain), which is gently palpated. This sequence helps to differentiate true rigidity from the voluntary spasm that occurs from nervousness. Similarly, Rovsing's sign, or pain in the right lower quadrant upon palpation of the left lower quadrant, is a sign of local peritonitis. Severe, well-established rigidity of the abdominal wall is a sign of well-established local peritonitis and of impending (or previous) perforation.

Another useful sign in establishing the presence of local peritonitis is the shake test. Most surgeons perform this by grasping the iliac wings and shaking the pelvis from side to side. The patient complains of pain at the site of the appendix if local peritonitis is present. It is more helpful to kick or push the stretcher gently while watching the patient's face: a grimace is a sure sign of local peritonitis.

The elicitation of signs of local peritonitis may be difficult in patients with a retrocecal or pelvic appendix. The psoas sign, pain caused by extending the thigh to stretch the psoas muscle, is generally positive in patients with a retrocecal appendix and local peritonitis. If the appendix is adjacent to the obturator internus, stretching of this muscle by external rotation of the hip elicits severe pain and spasm. With a true, inflamed pelvic appendix, the only signs of local peritonitis may be found on rectal examination in men or vaginal examination in women. It is often impossible to differentiate pelvic appendicitis from pelvic inflammatory disease by physical examination alone, but an accurate history will generally solve the dilemma.

Patients who present late with appendicitis may have only general tenderness and rigidity. This is a sure sign of perforation, but without an accurate history the diagnosis remains obscure.

Hyperaesthesia or dysesthesia may occur in patients with non-perforated appendicitis. Although these are inconstant findings, when present they occur on the right side of the abdomen, in the distribution of the tenth, eleventh, and twelfth thoracic nerves. Paresthesia is best elicited by light scratching with a sharp, sterile needle.

Fever is a late physical finding in appendicitis. Before perforation, the body temperature is usually no more than 39 to 39.5°C, but with perforation may rise to 40 to 41°C. If fever has been present since the onset of the illness, consider other causes.

Laboratory examinations

Laboratory examinations are rarely helpful in the diagnosis of appendicitis. Leukocytosis is common, usually in the range of 11 000 to 17 000 mm^3. However, this increase occurs once appendicitis is well established, and the illness should generally be diagnosed before it develops. A leukocytosis of over 20 000 mm^3 suggests perforation of the appendix or another diagnosis. Urinalysis is unhelpful in the diagnosis or exclusion of appendicitis. Mild pyuria may occur due to irritation of the bladder by a pelvic appendix or irritation of the ureter by a retrocecal appendix. Thus, in patients with symptoms suggestive of either appendicitis or urinary-tract infection, urinalysis is not diagnostic of either condition.

Certain radiographic techniques have been recommended in the evaluation of patients with possible appendicitis. Plain abdominal films demonstrate a fecalith in the area of the appendix in 10 per cent of patients with appendicitis. This finding is highly suggestive of appendicitis in a patient with a compatible history and physical examination. Plain abdominal films may reveal other abnormalities, such as localized ileus in the right lower quadrant, soft-tissue density in the right lower quadrant, or free intraperitoneal air. These are so non-specific as to be of no value in the diagnosis of appendicitis.

Barium enema has long been recommended for the evaluation of possible appendicitis. Findings suggestive of appendicitis include spasm of the terminal ileum or cecum, external compression of the cecum, and non-filling or partial filling of the appendix. While diagnostic accuracy rates of 90 per cent are reported in several series, review of these studies discloses that about one-quarter of patients with an abnormal barium enema have a normal appendix at the time of exploration. Probably the best case that can be made for barium enema is that if the entire appendix is filled, appendicitis is virtually ruled out.

Ultrasonography and computed tomography (**CT**) have recently been used in the evaluation of patients with suspected appendicitis. Ultrasonography is performed with high-resolution linear-array transducers and gentle but thorough compression of the abdomen, with the goal of eliminating any gas-filled bowel between the transducer and the appendix. Ultrasonographic criteria for the diagnosis of appendicitis are a non-compressible appendix, surrounded by a hypoechoic, thickened wall more than 2 mm in diameter. In addition, the maximal diameter of the visualized appendix should exceed 6 mm. Using these criteria, sensitivity rates of 75 per cent and specificity rates of 100 per cent have been reported. The sensitivity is usually considerably lower in patients with perforated appendicitis, as the rigidity of the abdominal wall prevents its adequate compression. Realistically, a sensitivity of only 75 per cent is too low to be acceptable in a diagnostic test. Ultrasonography has not, therefore, become widely used in the diagnosis of appendicitis, although a totally normal appendix visualized by ultrasound makes appendicitis unlikely. CT has been touted by radiologists but has not been widely accepted by surgeons for the diagnosis of appendicitis. CT findings suggestive of appendicitis include pericecal inflammation manifested by increased density, appendiceal thickening, pericecal adenopathy, and pericecal collections of fluid. Various imaging techniques have been described, some using oral and/or rectal contrast, but helical scanning seems to be the most accurate.

Sensitivity, specificity, and accuracy are reportedly in the 95 per cent range, and much is made of the ability to diagnose other conditions that mimic appendicitis (mesenteric adenitis, pelvic inflammatory disease, perforated cecal carcinoma, etc.). However, the vast majority of patients may be diagnosed with standard clinical techniques, making a CT scan unnecessary. The real value of CT in appendicitis appears to be in the diagnosis of truly obscure cases, and the differentiation of appendiceal abscess from appendiceal phlegmon.

Differential diagnosis

In evaluating a patient with abdominal pain suggestive of appendicitis, other possible causes of abdominal pain must be considered. From a management viewpoint, the various diagnoses can be divided into those that require surgery and those that do not. The former include perforated carcinoma of the right colon, perforated right colonic diverticulitis, perforated ulcer (with tracking of visceral contents along the right gutter to the right lower quadrant), sigmoid diverticulitis (especially with a mobile sigmoid colon), Meckel's diverticulitis, ectopic pregnancy, and ovarian torsion. Any of these disorders may be indistinguishable from appendicitis at presentation, but as each usually requires surgery for resolution, their differentiation from appendicitis is not crucial.

Surgery has limited value in the management of many other disorders that are almost indistinguishable from appendicitis. These include pelvic inflammatory disease, *Mittleschmerz* pain (from ruptured ovarian follicular cyst), ‘mesenteric adenitis', viral gastroenteritis, bacterial gastroenteritis, acute Crohn's ileitis, and typhoid. Pelvic inflammatory disease is more likely than appendicitis to occur during the proliferative phase of the menstrual cycle. It has a longer duration of symptoms, higher fever, greater leukocytosis, and less well localized pain, with more pelvic pain and cervical motion tenderness than appendicitis. *Mittleschmerz* pain can usually be diagnosed by its occurrence at ovulation and by the fact that most patients have experienced painful ovulation before. Although fever is uncommon, tenderness and leukocytosis may occur. The diagnosis is usually made by exclusion: the patient is observed and the pain resolves. ‘Mesenteric adenitis' is also a diagnosis of exclusion: the patient has all the typical features of appendicitis, but at exploration is found to have a normal appendix and some enlarged lymph nodes in the mesentery of the distal ileum. The cause of the illness is obscure. Viral and bacterial gastroenteritis are indistinguishable clinically, and the diagnosis of bacterial gastroenteritis is made only when the results of stool cultures become available (usually after the patient has been treated or the illness has resolved spontaneously). Viral and bacterial gastroenteritis may usually be differentiated from appendicitis by the massive diarrhea present in typical gastroenteritis. Diarrhea associated with appendicitis is rarely massive or prolonged. Most patients with gastroenteritis present with diffuse abdominal pain that rarely becomes well localized; if any tenderness develops, it is usually mild and generalized. It is most unusual to find a patient with true tenderness of the right lower quadrant due to gastroenteritis.

Diagnosis in difficult circumstances

Retrocecal appendicitis

The symptoms and signs of appendicitis are altered when the appendix is retrocecal. Pain is generally not as severe as that in patients with an abdominal or pelvic appendix, and pain rarely becomes well localized to the right lower quadrant. Many patients with a retrocecal appendicitis have a history of generalized abdominal pain, fairly constant for several days, never localizing, and never really interfering with normal activities. Occasionally, the pain becomes localized to the right flank or to the right upper abdomen. Patients with retrocecal appendicitis have tenderness in the area of the appendix, provided that the area can be palpated. In true retrocecal appendicitis, the area of tenderness is in the flank or costovertebral angle, simulating pyelonephritis. Occasionally the tenderness is subcostal, because the tip of the appendix is located at or near the hepatic flexure. In an obese patient, tenderness can often not be elicited because the retrocecal appendix is so well encased in retroperitoneal fat that it cannot be felt.

The combination of abdominal pain, more or less localized to the right side, with some aspect of nausea or anorexia and no other suggestive diagnosis, should raise the suspicion of retrocecal appendicitis. The presence of fever or leukocytosis makes the diagnosis of retrocecal appendicitis even more likely. Unfortunately, most such patients are dismissed by physicians as having gastroenteritis, a most unusual occurrence in the absence of significant diarrhea. Eventually, they present with a retrocecal abscess due to perforation of the appendix (Fig. 1).

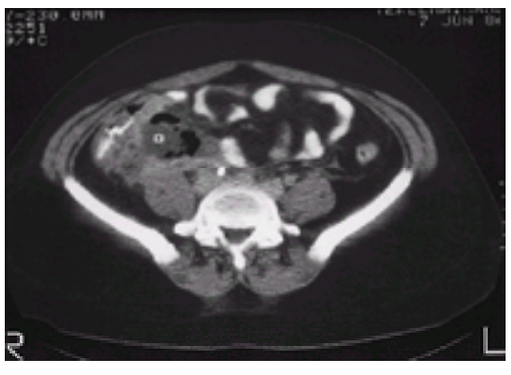


Fig. 1. CT scan demonstrating retrocecal appendiceal abscess (marker).

Appendicitis in the elderly

Elderly patients tend to present late in the course of appendicitis and with less well-defined symptoms; the incidence of perforation is therefore higher. Most elderly patients with appendicitis have a history of several days of poorly defined abdominal pain, anorexia, and fever. Rarely is the pain well localized to McBurney's point; rather, it is generally described as being in the right side of the abdomen. Most old patients have fever and abdominal tenderness at the time of presentation with appendicitis. Tenderness is either in the right lower quadrant, the right flank, or is diffuse from the effects of free perforation. The psoas and obturator signs are unhelpful in old patients because almost every manoeuvre is equally uncomfortable.

Appendicitis in pregnancy

The diagnosis of appendicitis in pregnancy is difficult because the appendix is displaced by the gravid uterus. Early in the course of pregnancy the appendix remains in its normal position, and diagnosis is routine. By the middle of the second trimester, however, the appendix becomes displaced superiorly, attaining a position in the right upper flank or epigastrium. Appendicitis may be easily mistaken for pyelonephritis or cholecystitis.

The abdominal wall is lifted from the appendix by the gravid uterus, and muscular laxity occurs: the abdominal findings associated with peritoneal irritation by the inflamed appendix may therefore be fewer than one might expect in the non-pregnant woman. Leukocytosis is a normal physiological response of pregnancy (up to 12 500 leukocytes/mm^3) and cannot be relied upon to help confirm the diagnosis of appendicitis. White blood-cell counts as high as 25 000 leuko-cytes/mm^3 are not unusual in pregnant women with appendicitis.

Management (Fig. 2)

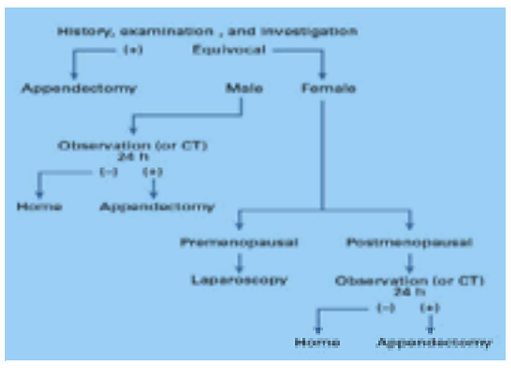


Fig. 2. Algorithm for management of possible appendicitis.

The management of non-perforated appendicitis is surgical removal. If the diagnosis is clear, an incision should be made over the point of maximal tenderness, generally at McBurney's point. Many surgeons favour a true McBurney incision, with an oblique skin incision at McBurney's point and splitting of the external and internal oblique muscles in the line of their fibres; others prefer a transverse skin incision with muscle splitting. The skin incision need be only 3 to 6 cm long (depending on the patient's build), and can be easily extended by cutting the rectus fascia and retracting the rectus muscle medially. The tenia of the colon are followed to the base of the appendix, and blunt dissection is used to free the appendix from its surrounding inflammatory tissue. The mesoappendix is divided between clamps and ligated with an absorbable suture. The base of the appendix is divided and is also ligated with absorbable suture material. The base of the appendix may be inverted using either a purse-string suture or a ‘Z-stitch' (Fig. 3), although there are no firm data to suggest that inversion of the stump produces better healing than simple ligation. The muscle layers of the abdominal wall are closed with absorbable suture; the skin is closed with a monofilament suture that should be placed deep enough to eliminate dead space in the subcutaneous tissue (Fig. 4).

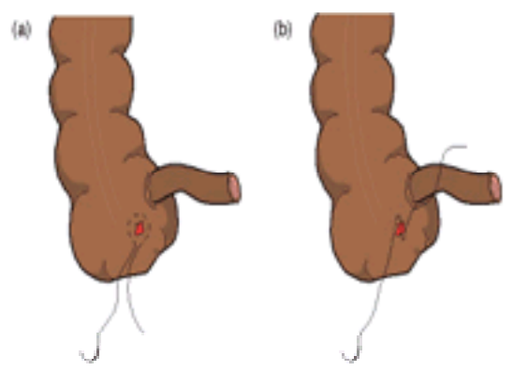


Fig. 3. (a) Purse-string stitch at base of appendix; (b) ‘Z stitch' used for closing base of appendix.

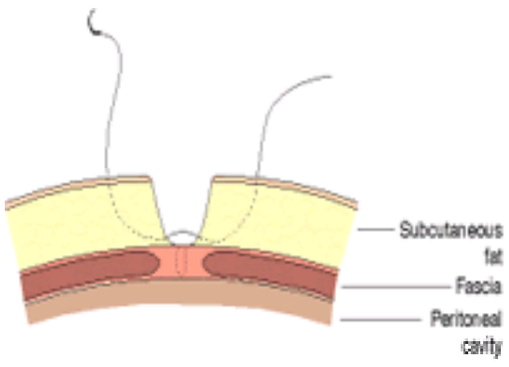


Fig. 4. Placement of cutaneous stitch to eliminate dead space.

For patients in whom the diagnosis of appendicitis is suspected, but not definite, several options are available. The most attractive of these is simply to admit them to the hospital for further observation. Over the course of 12 to 24 h one should be able to characterize the course of the illness well enough to determine whether an exploration for possible appendicitis, further tests (for example, computerized tomography), or discharge are in order. Although there is some risk of perforation of the appendix during this period of observation, if the patient is re-examined frequently, one should be able to avoid this complication. The risk of perforation is low in the first 24 h of symptomatic appendicitis; it is therefore safe to observe patients early in the course of their illness, when physical findings may not be well developed.

Another option is laparoscopy. This examination is commonly used in patients with symptoms and signs suggestive of both appendicitis and pelvic inflammatory disease. To exclude the diagnosis of appendicitis by laparoscopy, the entire appendix must be seen and must be normal. If the appendix is retrocecal, mobilization ofthe right colon and retraction of the appendix will be necessary. In a patient with an unclear diagnosis, laparoscopy may be used as both a diagnostic and a therapeutic manoeuvre. The technique for laparoscopic appendectomy is straightforward. A high flow-rate laparoscopic insufflator, video camera and monitor, and adequate assistance are required. The abdomen is filled with carbon dioxide, a large trocar (10, 11, or 12 mm) is placed infraumbilically and the laparoscope with camera is placed through it. A large trocar is placed through the left rectus sheath, avoiding the inferior epigastric vessels, halfway between the pubis and the

umbilicus. An additional (5 mm) trocar is placed suprapubically. Through these trocars, the appendix is freed from surrounding tissues. Once free, it is usually helpful to place an additional small trocar in the right upper quadrant to hold the appendix up toward the abdominal wall. The mesoappendix is dissected and bleeding is controlled with clips or sutures. The base of the appendix may be controlled with pretied suture loops, sutures, or endoscopic staples. The laparoscope is moved to the left-sided large trocar, and the appendix is removed through the umbilical trocar by widening the fascial opening in the abdominal wall.

What the real value of laparoscopic appendectomy may be remains unclear. Randomized, prospective trials fail to show any dramatic improvement in recovery from surgery in the general population treated with laparoscopic rather than open surgery, though costs are higher with laparoscopic surgery. There does appear to be a slight increase risk of wound infection in patients treated with laparoscopic appendectomy, though this risk can be eliminated with careful wound protection. Laparoscopic appendectomy appears to be of greatest value in obese patients who would otherwise require a large incision.

Management of the appendiceal mass

Patients seen late in the course of appendicitis often have a palpable mass in the right lower quadrant. These account for about 3 per cent of patients, and traditionally they have been managed with antibiotics and bed rest, reserving surgery for those who do not respond to such conservative therapy. Non-operative therapy is successful in 80 to 90 per cent of patients; causes of failure include sepsis, unresolved abscess, and small-bowel obstruction from adhesions to the inflammatory mass. The rate of failure of non-operative therapy is higher in patients with a well-defined abscess at presentation than in those with diffuse inflammation, and abscess can be reliably diagnosed by CT. If an abscess is found, operative or CT-directed drainage is indicated. Percutaneous, radiologically guided drainage is successful in 85 per cent of patients and avoids the need for operative intervention in patients with active sepsis. Most patients treated in this way can be discharged from the hospital in 7 to 10 days.

Traditionally, all patients with an appendiceal mass underwent appendectomy 6 to 8 weeks after their acute illness. However, only 20 per cent of patients develop another episode of appendicitis following treatment of an appendiceal abscess or diffuse inflammation, prompting some surgeons to abandon interval appendectomy. The morbidity associated with interval appendectomy is so low, however, that there is a case for it in patients for whom surgery does not pose a risk because of coexistent medical illnesses. If an interval appendectomy is not done, patients beyond young adulthood should undergo a barium enema examination or colonoscopy to exclude the possibility of a locally perforated carcinoma of the right colon.

Antibiotics

All patients with suspected appendicitis should receive broad-spectrum antibiotics preoperatively. The duration of antibiotic therapy is the subject of some controversy, but it should be governed by the severity of the appendicitis. In the patient with early, acute appendicitis, without purulence or gangrenous changes, 24 to 48 h of antibiotic therapy should suffice. In those with perforated appendicitis, appendiceal abscess, or gangrene of the appendix, 7 to 10 days of antibiotic therapy seems prudent. The choice of antibiotics seems to be less important in the prevention of wound infection than the timing of their administration. Effective regimens include cefoxitin alone, ampicillin–sulbactam, ampicillin–clavulanic acid, cefotetan, cefoperazone, an aminoglycoside with metronidazole, aminoglycoside and clindamycin, and imipenem.

Further reading

Balthazar EJ, Birnbaun BA, Yee J, Megibow AJ, Roshkow J, Gray C. Acute appendicitis: CT and US correlation in 100 patients. *Radiology* 1994;**190**: 31–5.

Bauer T, Vennits B, Holm B. Antibiotic prophylaxis in acute nonperforated appendicitis. *Annals of Surgery* 1989; **144**: 338–40.

Berry J, Malt RA. Appendicitis near its centenary. *Annals of Surgery* 1984; **200:**, 567–75.

Burns RP, Russell WL. Appendicitis in mature patients. *Annals of Surgery* 1985; **201:**, 695–704.

Crabbe MM, Norwood SH, Robertson HD. Recurrent and chronic appendicitis. *Surgery Gynecology Obstetrics* 1986; **163**: 11–13.

Ferzli GE, Ozuner G, Davidson PG. Barium enema in the diagnosis of acute appendicitis. *Surgery Gynecology Obstetrics* 1990; **171**: 40–2.

Fisher KS, Ross DS. Guidelines for therapeutic decisions in incidental appendectomy. *Surgery Gynecology Obstetrics* 1990; **171**: 95–8.

Frazee RC *et al.* A prospective, randomized trial comparing open versus laparoscopic appendectomy. *Annals of Surgery* 1994; **219**: 725–8.

Gotz F, Pier A, Bacher C. Modified laparoscopic appendectomy in surgery. *Surgical Endoscopy* 1990; **4**: 6–9.

Horowitz MD, Gomez GA, Santiesteban R. Acute appendicitis during pregnancy. *Archives of Surgery* 1985; **120**: 1362–7.

Puylaert JBCM, Rutgers PH, Lalisang R. I. A prospective study of ultrasonography in the diagnosis of appendicitis. *New England Journal of Medicine* 1987; **317:**, 666–9.

Rao PM, Rhea JT, Novelline RA, Mostafavi AA, McCabe CJ. Effect of computed tomography of the appendix on treatment of patients and use of hospital resources. *New England Journal of Medicine* 1998; **338**: 141–6.

Silen W. *Cope's early diagnosis of the acute abdomen*. Oxford University Press, 1991.

27.2 Primary appendiceal malignancies

Charles A. Staley

Introduction
Benign tumors
Malignant tumors
Carcinoid tumors
Mucinous cystadenocarcinoma
Adenocarcinoma
Adenocarcinoid
Further reading

Introduction

Because the appendix is derived embryologically from the large intestine, both benign and malignant conditions of the large intestine can also occur in the appendix. Malignant tumors of the appendix are rare, but merely reflect the small size of the organ. The four main types of appendiceal malignancies are carcinoid, mucinous cystadenocarcinoma, adenocarcinoma, and adenocarcinoid tumors (Table 1). In general, these tumors clinically present with symptoms of acute appendicitis, are rarely diagnosed preoperatively or intraoperatively, and are usually found on pathologic examination of the removed appendix. Knowledge of these malignancies is important since each has different clinical presentation, prognosis, and surgical treatment.

Type of tumor	*Frequency (%)*
Carcinoid	85
Mucinous cystadenocarcinoma	8
Adenocarcinoma	4
Adenocarcinoid	2
Others: sarcoma, paraganglioma, and granular cell tumors	1

Table 1 Primary appendiceal malignancies

Benign tumors

Benign tumors are also rare and comprise two groups, polyps and adenomas. Appendiceal polyps are similar to those in the rest of the colon and may therefore have varying degrees of malignant transformation. On pathologic examination, an adenoma of the appendix tends to be diffuse and villous, unlike its colorectal counterpart. Excessive production of mucus by an adenoma causes a large sausage-shaped cystic mass referred to as a cystadenoma.

Benign lesions of the appendix are usually aymptomatic and are usually found incidentally at exploration or pathologic examination. Cystadenomas may present with acute appendicitis or a palpable mass. Preoperative ultrasonography, computed tomography (**CT**), or magnetic resonance imaging (**MRI**) of patients with a cystadenoma should reveal a fluid-filled, variable-shaped, thin-walled structure in the right lower quadrant containing low density contents. Cystadenomas can occasionally rupture, but the mucus is usually contained in the right iliac fossa (localized pseudomyxoma peritonei).

Benign tumors are cured by appendectomy provided that the resection margin is negative. Localized pseudomyxoma peritonei usually resolves after appendectomy and excision of local mucin deposits. Similar to patients with colonic polyps, those with appendiceal polyps should undergo colonoscopy to rule out synchronous colorectal adenomas or carcinoma.

Malignant tumors

Carcinoid tumors

The most common tumor of the appendix is a carcinoid, occurring in up to 0.5 per cent of appendectomy specimens and accounting for 85 per cent of all appendiceal tumors. At least 50 per cent of all carcinoid tumors originate in the appendix. They are common in young women and are detected incidentally at abdominal exploration or present with signs of acute appendicitis. Spread to locoregional lymph nodes or metastatic disease is rare although patients with carcinoid syndrome usually have liver or retroperitoneal metastases. Increased urinary excretion of 5-hydroxy-indoleacetic acid (5-HIAA) is useful in monitoring disease progression.

Pathology

Carcinoids are small, yellow-white firm, well-circumscribed, but unencapsulated tumors derived from neuroectodermal origin and thus classified as part of the amine precursor uptake and decarboxylation (APUD) neoplasms. Many carcinoids are unrecognized grossly being seen only on microscopic section. About 75 per cent of the tumors occur at the tip, 15 per cent at the middle, and 10 per cent at the base of the appendix. Eighty per cent of appendiceal carcinoids measure less than 1 cm in diameter, 14 per cent measure 1 to 2 cm, and 6 per cent measure more than 2 cm. Two main histologic patterns are recognized, one morphologically identical to classic (midgut) carcinoid, the other to rectal (hindgut) carcinoid. However, the histologic patterns are of little practical importance because there is little correlation between the microscopic patterns and the biologic behavior of the tumor. Although mitoses are infrequent, lymphatic invasion is often seen. The mesoappendix is frequently involved, but spread to the regional lymph nodes and distant sites is rare. Metastasis to the liver may be associated with the carcinoid syndrome.

Treatment

The most important determinant of surgical treatment of appendiceal carcinoids is the tumor size (Table 2). Many studies have shown that distant metastases and death occur only in patients with tumors greater than 2 cm. Moertel *et al.* found no recurrences and no metastases among 108 patients with a tumor size of less than 1 cm. However, there was a 30 per cent incidence of metastases if the tumor size was greater than 2 cm. Because of these findings, standard treatment of lesions less than 1 cm has been appendectomy. For lesions greater than 2 cm, a right hemicolectomy is recommended. Adequate therapy for lesions from 1 to 2 cm remains controversial. In the long-term study by Moertel *et al.*, no patients with this sized lesion had tumor recurrence after simple appendectomy with a follow-up period of more than 25 years. However, metastases in patients with appendiceal carcinoids from 1 to 2 cm have been reported. In a review of 46 reported metastasizing appendiceal carcinoids, Thirlby *et al.* found five patients with a primary tumor size from 1 to 2 cm. The correlation of positive regional lymph nodes in patients with vascular/lymphatic invasion and mesoappendix involvement has influenced authors to recommend a right hemicolectomy in patients with these pronostic factors.

Size	Percentage of patients	Treatment
<1 cm	80	Appendectomy
1.0–2.0 cm	14	
	Without mesoappendix and lymphatic invasion	Appendectomy
	With mesoappendix and lymphatic invasion	Right hemicolectomy
>2.0 cm	6	Right hemicolectomy

Table 2 Carcinoid tumors

Prognosis

The prognosis in carcinoid of the appendix with rare exceptions is favorable with 5-year survival rates of 90 to 100 per cent. Recurrences are found only in patients with large tumors or those who had intra-abdominal spread at the original surgery. In patients in whom all recurrent disease can be resected the 5-year survival rate is about 75 per cent. Radiation therapy and chemotherapy have not been helpful in the treatment of carcinoids. Metastatic disease in the liver is rare but may be associated with carcinoid syndrome. Treatment with serotonin antagonists, interferon-a, or hepatic artery embolization may help with the symptoms of carcinoid syndrome in these patients with liver metastases.

Mucinous cystadenocarcinoma

Although mucinous cystadenocarcinoma is included with colonic adenocarcinoma as adenocarcinoma, it is considered separately here because of its different clinical behavior. Cystadenocarcinoma is the second most common appendiceal malignancy and one that many times can be diagnosed preoperatively. Patients usually present with symptoms of acute appendicitis or a right lower quadrant mass. A barium enema may show a non-filling appendix with a globular mass. CT demonstrates a mass near water density with calcium in its wall. About 50 per cent of these patients will have intra-abdominal metastases and pseudomyxoma peritonei.

Pathology

Mucocele is a term used to describe an appendix dilated full of mucus due to an obstruction of the appendiceal lumen by mucosal hyperplasia, mucinous cystadenomas, or mucinous cystadenocarcinomas. Mucinous cystadenomas and cystadenocarcinomas resemble each other both grossly and microscopically. Cystadenocarcinomas show abundant extracellular mucin production and tend to be well differentiated. The distinction between malignant mucinous cystadenocarcinoma and cystadenoma is based on two histologic features: the first is invasion of the appendiceal wall by atypical glands; the second is the identification of epithelial cells in any intraperitoneal areas. As the intraluminal mucus production continues, the appendiceal wall becomes thin, ulcerated, and calcified. Almost half of those with cystadenocarcinoma will rupture leading to diffuse pseudomyxoma peritonei.

Pseudomyxoma peritonei is a rare condition in which diffuse collections of gelatinous fluid are associated with mucinous implants on the peritoneal surfaces and omentum. It occurs more in women and in most cases is found unexpectedly at surgical exploration. Pseudomyxoma peritonei has been associated with cystic tumors of both the appendix and the ovary; however, many authors believe that this condition arises from the appendix and secondarily spreads to the ovary. Patients with pseudomyxoma peritonei can present with nausea, vomiting, fatigue, abdominal pain, distention, or a mass in the abdomen. CT scans demonstrate abundant mucinous material within the peritoneal cavity similar in density to fat and with a heterogeneous appearance.

Treatment

Cystadenomas are cured by a simple appendectomy. Similar to colonic adenocarcinoma, appendiceal cystadenocarcinoma is best treated with a right hemicolectomy. The optimal treatment for pseudomyxoma peritonei is somewhat controversial. Most surgeons would agree that thorough surgical debulking, resecting all gross disease, is the mainstay of treatment for this condition. Sugarbaker, who has published extensively on pseudomyxoma peritonei, advocates an ultraradical surgical approach with intraperitoneal chemotherapy. Sugarbaker's most recent series of 130 patients with appendiceal cancer with pseudomyxoma peritonei demonstrates a 3-year survival of 73 per cent with a short mean follow-up of 24 months. Depending on the extent of disease, Sugarbaker advocates that up to six peritonectomy procedures may be necessary to resect all gross disease. These operations took a mean of 11 h, had a 1700 ml blood loss, and involved a mean of 2.5 intestinal anastomoses. The overall complication rate was 36.1 per cent and the mean duration of ileus was 19.3 days. Critics point out that survival rates with this radical surgical procedure may be similar to more conservative operative debulking as reported by the Mayo Clinic. Gough *et al.* reviewed 56 patients with pseudomyxoma peritonei. Of these patients, 52 per cent had primary carcinomas of the appendix. Their surgical approach involved aggressive surgical debulking, appendectomy, bilateral oophorectomy, and omentectomy at the initial procedure. Their 2- and 5-year survival rates were 86 and 53 per cent, respectively. It is not clear that the increased complexity of Sugarbaker's surgical peritonectomy along with significant complication rates extend long-term survival rates compared with more conservative but aggressive debulking procedures. Both studies have shown a benefit of adjuvant intraperitoneal chemotherapy with 5-fluorouracil-based treatment.

Prognosis

Because mucinous adenocarcinomas have a slowly progressive clinical course, patients tend to have a better prognosis than those with adenocarcinoma. Patients undergoing a right hemicolectomy have 5- and 10-year survival rates of 70 and 65 per cent, respectively. Aggressive surgical debulking of pseudomyxoma peritonei results in a 50 per cent 5-year survival rate.

Adenocarcinoma

Adenocarcinoma of the appendix is much less common than carcinoids and accounts for 0.1 per cent of all performed appendectomies. The mean age of presentation is 50 years and men are affected more than women. Approximately 50 per cent of patients have symptoms of acute appendicitis, frequently with an appendiceal abcess. Rarely is the diagnosis made pre- or intraoperatively and confirmation is found in the pathology report.

Pathology

Adenocarcinoma is identified histologically by invasion of the appendix wall by neoplastic cells typical of a poorly differentiated colonic adenocarcinoma. Most tumors are near the base of the appendix and appear as a firm nodularity. Lymph node metastases are found in 25 per cent of the cases. Like other appendiceal tumors, adenocarcinoma also tends to metastasize to the ovaries.

Treatment

The lymphatic drainage of the appendix is identical to the cecum. A right hemicolectomy is the treatment of choice for all lesions with invasion beyond the mucosa. For lesions confined to the mucosa (*in situ* carcinoma) some authors suggest there is no survival advantage in performing a right hemicolectomy over appendectomy alone. For patients who have had an appendectomy and are found to have invasive adenocarcinoma, a secondary right hemicolectomy should be performed. Bilateral oophorectomy should be done in postmenopausal women. Surveillance for synchronous or metachronous tumors is important since 20 per cent of patients will develop a second gastrointestinal cancer.

Prognosis

The prognosis of patients with adenocarcinoma of the appendix is worse than those with cystadenocarcinoma and carcinoid tumors. The outlook is similar to colonic adenocarcinoma. Overall, 5-year survival after a right hemicolectomy is 60 per cent. Survival falls off if regional lymph nodes are involved. Adjuvant therapy with 5-fluorouracil-based chemotherapy is justified for patients with positive lymph nodes.

Adenocarcinoid

In the 1960s, an appendiceal tumor with both carcinoid and adenocarcinoma histologic features was described. Because these tumors behaved worse than pure carcinoids and better than adenocarcinomas, adenocarcinoid seemed an appropriate classification. Other names used for this tumor include goblet cell carcinoid, mucinous carcinoid, and crypt cell carcinoma. Patients with adenocarcinoid tumors are usually symptomatic presenting with acute appendicitis, an abdominal mass, or an ovarian mass. Most patients are 40 to 50 years old.

Pathology

Most of the appendices show a fibrin purulent reaction and discrete tumors are not obvious. Most tumors are less than 2 cm and half are in the proximal appendix and the other half in the tip of the appendix. Metastasis to the ovaries is common (20 per cent) and may or may not be the same histologic type as the primary tumor.

Histologically, adenocarcinoids are a mixture of tubuloglandular structures and goblet cells that arise from the base of the glands and spare the luminal mucosa. Vascular and perineural invasion are commonly seen in these tumors. Debate exists about the histogenesis of adenocarcinoid tumors. Adenocarcinomas derive from primordial endodermal elements, while carcinoids develop from neural crest cells. Recently a subepithelial neurosecretory cell identical to the Kulchitsky cell has been identified. These cells contain neuron-specific enolase and acid mucins. This subepithelial neurosecretory cell may be a plausible progenitor for adenocarcinoid tumors.

Treatment

Unlike carcinoids, tumor size is not a predictable indicator of clinical behavior. Therefore, because we cannot predict which adenocarcinoids will behave innocently, a right colectomy is recommended for patients with localized disease. Because in most cases the diagnosis is made postoperatively, a right colectomy is performed as a secondary procedure along with a bilateral oophorectomy in most cases. Some patients will undergo an exploratory laparotomy and will be found to have a Krukenburg tumor(s) with no obvious gastrointestinal source. An appendectomy should be done to avoid overlooking a small primary appendiceal adenocarcinoid tumor.

Prognosis

The prognosis of adenocarcinoid tumors is in between that of carcinoids and adenocarcinomas. Those with a histologic tubular pattern have a better prognosis than those with a goblet cell pattern. The overall 5-year survival is 75 to 80 per cent. Patients with intra-abdominal spread and ovarian metastases have a poor prognosis.

Further reading

Carr NJ, McCarthy WF, Sobin LH. Epithelial noncarcinoid tumors and tumor-like lesions of the appendix. *Cancer* 1995; **75**: 757–68. [Prognostic factors in appendiceal cancers.]

Deans GT, Spence AJ. Neoplastic lesions of the appendix. *British Journal of Surgery* 1995; **82**: 299–306. [General review of benign and malignant tumors of the appendix.]

Edmonds P, Merino MJ, Livolsi VA, Duray PH. Adenocarcinoid of the appendix. *Gastroenterology* 1984; **86**: 302–9. [Review of adenocarcinoid tumors of the appendix.]

Gough DB *et al.* Pseudomyxoma peritonei. *Annals of Surgery* 1994; **219**: 112–19. [Mayo Clinic review of pseudomyxoma peritonei.]

Hananel N, Powsner E, Wolloch Y. Primary appendiceal neoplasms. *Israel Journal of Medical Sciences* 1993; **29**: 733–4. [Retrospective review of adenocarcinoma of the appendix.]

Higa E, Rosai J, Pizzimbono CA, Wise L. Mucosal hyperplasia, mucinous cystadenoma, and mucinous cystadenocarcinoma of the appendix. *Cancer* 1973; **32**: 1525–41. [A re-evaluation of appendiceal mucocele.]

Hinson FL, Ambrose NS. Pseudomyxoma peritonei. *British Journal of Surgery* 1998; **85**: 1332–9. [General review of pseudomyxoma peritonei.]

Loftus JP, van Heerden JA. Surgical management of gastrointestinal carcinoid tumors. *Advances in Surgery* 1995; **28**: 317–36. [Review of the surgical treatment of gastrointestinal carcinoids.]

Moertel CG, Weiland LH, Nagorney DM, Dockerty MB. Carcinoid tumors of the appendix: treatment and prognosis. *New England Journal of Medicine* 1987; **317**: 1699–701. [Treatment and prognosis of appendiceal carcinoid tumors.]

Roggo A, Wood WC, Ottinger LW. Carcinoid tumors of the appendix. *Annals of Surgery* 1993; **217**: 385–90. [Retrospective review of carcinoid tumors of the appendix.]

Stinner B, Kisker O, Zielke A, Rothmund M. Surgical management for carcinoid tumors of small bowel, appendix, colon, and rectum. *World Journal of Surgery* 1996; **20**: 183–8. [General review of gastrointestinal carcinoid tumors.]

Sugarbaker PH. Patient selection and treatment of peritoneal carcinomatosis from colorectal and appendiceal cancer. *World Journal of Surgery* 1995; **19**: 235–40. [Prognostic factors associated with long-term survival with pseudomyxoma peritonei.]

Sugarbaker PH, Jablonski KA. Prognostic features of 51 colorectal and 130 appendiceal cancer patients with peritoneal carcinomatosis treated by cytoreductive surgery and intraperitoneal chemotherapy. *Annals of Surgery* 1995; **221**: 124–32. [Prognostic factors in peritoneal carcinomatosis from colorectal and appendiceal cancers.]

Sugarbaker PH, Zhu BW, Sese GB, Shmookler B. Peritoneal carcinomatosis from appendiceal cancer: results in 69 patients treated by cytoreductive surgery and intraperitoneal chemotherapy. *Diseases of the Colon and Rectum* 1993; **36**: 323–9. [Results of cytoreductive surgery for pseudomyxoma peritonei.]

Warkel RL, Cooper PH, Helwig EB. Adenocarcinoid, a mucin-producing carcinoid tumor of the appendix. *Cancer* 1978; **42**: 2781–93. [A review of adenocarcinoid tumors.]

28.1 Haemorrhoids or piles

W.Hamish Thomson

Introduction
Anatomy of the anal lining
The anal cushions
Blood supply
Support
Function
The nature of piles
Pathology
Classification
Symptoms
Bleeding
Prolapse
Pain
Itching
Anorectal dysfunction
Soiling
Examination
Signs
Differential diagnosis
Anal tags
Fibroepithelial polyp
Sentinel pile
Fissure
Dermatitis
'Perianal haematoma'
Rectal prolapse
Rectal tumour
Miscellaneous
Treatment
Conservative treatment
Interventional treatment
Complications
Ephemeral
Permanent
Which treatment?
Management of infarcted ('strangulated') piles
Conclusion
Further reading

Introduction

Seldom can prevalence have been so paradoxically matched by incomprehension, or ignorance embarrassed by accessibility as in piles, for the disease is both common and easy to inspect, and yet unnecessary confusion exists. There are two reasons: a legacy of misleading terms related in turn to a misconceived pathological explanation, and an inadequacy of anatomic description. Thus disarmed by a stylized and simplistic image of anal anatomy and disabled by both erroneous terminology and time-honored myths, the qualifying doctor can be excused an inaccurate mental picture of what is in essence a simple complaint. Given a true picture of the morphology of the anal lining, however, with an understanding of its pertinent microscopic features one can readily perceive both what piles are, how they occur, and what they would look like, and be able to predict and envisage the complications to which they are prone.

The names 'haemorrhoids' and 'piles' are essentially synonymous though differently derived from the two main—and only certain—symptoms, respectively bleeding and protrusion. The disease, though, is not the result of varicosity nor, as the reader will appreciate, are itching and pain its necessarily expected accompaniments. All this, and their logical management, will become clear from the anatomic account.

Anatomy of the anal lining

The anal canal is usually shown as a 4-cm long tube lined in its upper two-thirds by insensible mucosa and below by a hairless, glandless cuff of highly sensitive squamous epithelium, the anoderm. The mucosa is seen to be thrown into longitudinal folds (named after Morgagni) which are sometimes—incorrectly—depicted as containing vertically ascending columns of serpigious veins. This oversimplified description derives from the usual method of preservation and dissection and provides no clue either to the intricacies of its infrastructure or the almost inevitable vagaries of its existence (Fig. 1).

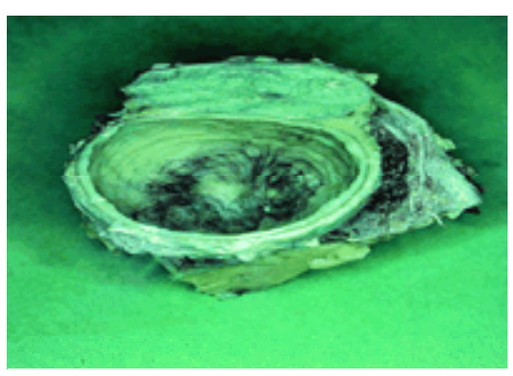


Fig. 1. The view from above: on looking down from the rectal ampulla and into the anal canal (closed at the anal verge) the horizontal folds of rectal mucosa can be seen to give way to the vertically disposed 'columns of Morgagni' lining the upper anal canal. The transected prostate, staining light blue, is seen adjoining the rectum anteriorly.

If, instead, the vessels of a fresh anorectum are filled (retrogradely, from the superior rectal vein) prior to inflation and fixation, the interior of the anal canal is seen, on transection of the rectum above, to look quite different (Fig. 2). Now the anal lining bulges into the lumen as three (occasionally four) pads—each more or less impressed by the vertical mucosal folds—which have been called the anal cushions. Their presence and curious architecture are the key to piles.

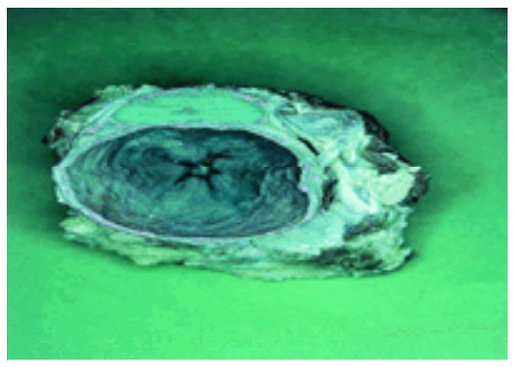


Fig. 2. The same view as Fig. 1, now with the veins filled with a coloured suspension of barium sulphate demonstrating the shape, position, and existence of the anal cushions conferring a stellate cross-section to the anal lumen. The finer folds of Morgagni have been more or less erased, and subdivide the anal cushions. Again the prostate is seen anteriorly.

The anal cushions

Nineteenth-century anatomists gave a more accurate and detailed description of the anal lining. One feature they recognized was the thickness and rich vascularity of the anal submucosa. It was in fact likened to cavernous tissue and reckoned to assist anal closure. The observation of its discontinuous grouping into three main and constantly sited masses, however, explained the nature of piles. We called them the anal cushions. There is often a fourth midline posterior one and their substance extends above and below the dentate line. When their component vasculature is filled they confer an appearance to the anal lining which belies the standard anodyne description (Fig. 3).

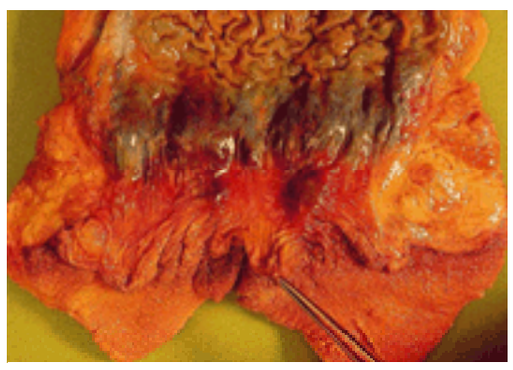


Fig. 3. An anal canal, prepared to demonstrate the anal cushions as in Fig. 2, slit open to show the interior. The anal cushions can be seen dark in their mucosal part with blood pushed down ahead of the barium sulphate suspension injected into the superior rectal vein.

Blood supply

The anal cushions receive a rich intercommunicating supply from the superior, middle, and inferior rectal (synonymously called haemorrhoidal) arteries. From five to eight branches of the superior rectal artery pass from the mesorectum (Fig. 4(a)) through muscular ‘button holes' in the rectal ampullary wall to descend into the anal submucosa (Fig. 4(b)), there to anastomose freely with contributions from the middle and inferior vessels. Local mucosal excisional procedures will inevitably encounter substantial bleeding from any one or all three sources. Part of this supply is carried straight into the venous system by direct arteriovenous shunts and probably provides for the mechanical function (described below) without which the wealth of vasculature is not fully explained.

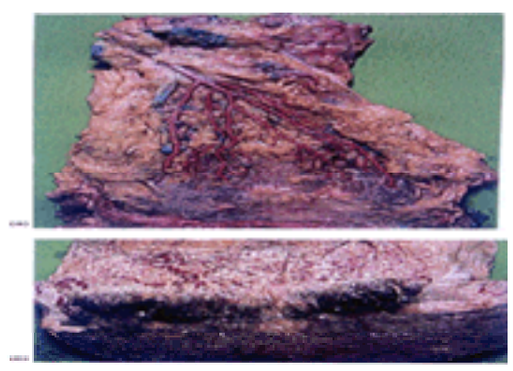


Fig. 4. (a) A dissection of the superior rectal artery (injected with red latex) branching within the mesorectum of an anteriorly opened anorectal specimen. (b) Branches of the superior and middle rectal arteries (injected with red latex) dissected as they emerge into the anal submucosa.

The veins of the anal submucosa are particularly notable, exhibiting one of the two entirely unique local morphological characteristics. They are distinguished by discrete dilations along their course, particularly subanodermally. These vein sacs, resemclover root, were once thought to be pathological, and in fact the underlying fault in piles. John Hunter (1728–1793) described them in a haemorrhoidectomy specimen (Spec. 1277, Hunterian Museum, Royal College of Surgeons of England) and noted the great curiosity of their being separated by segments of normal vein (Fig. 5). They are, however, normal. They occur in all adults (Fig. 6(a), Fig. 6(b)) and are found at birth (Fig. 7). The subanodermal sacs look like petals of a blue daisy through the skin of a baby's distracted anal verge, and can also be demonstrated in the adult (Fig. 8). They drain mainly cephalad into the portal system but also through the sphincter and below it into the systemic circulation; a route, though, that dissections suggest becomes increasingly tenuous with age (Fig. 6(a), Fig. 6(b)), which might explain the postdefaecatory anal verge engorgement and oedema that troubles some patients, and the oedema and discomfort of some prolapsed piles.

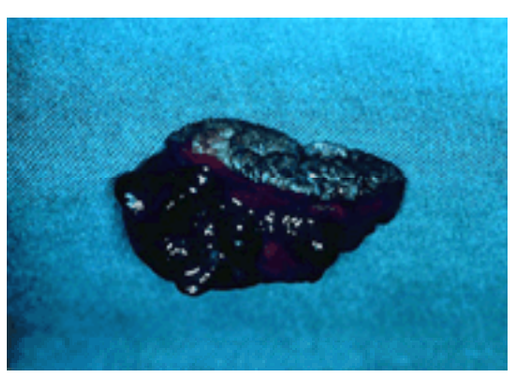


Fig. 5. A haemorrhoidectomy specimen clearly showing the vein sacs of the ‘haemorrhoidal' venous plexus, here clotted.

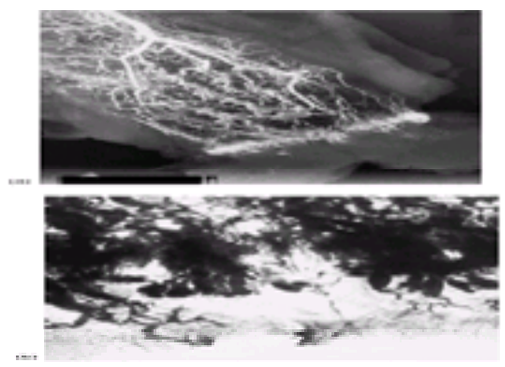


Fig. 6. (a) Radiograph of an anorectal specimen prepared by filling the superior rectal vein with a warm barium/gelatin mixture which on cooling sets in the veins. The specimen has been slit open. The venous saccules are easily filled from above but drain poorly inferiorly into the subcutaneous perianal tributaries of the inferior rectal vein, a feature often noted in adult specimens. (b) Close-up of the slit open anal canal from an adolescent, viewed with transmitted light. The veins have been filled with latex and the tissues rendered translucent. Drainage into the inferior rectal tributaries is freer than in later life.

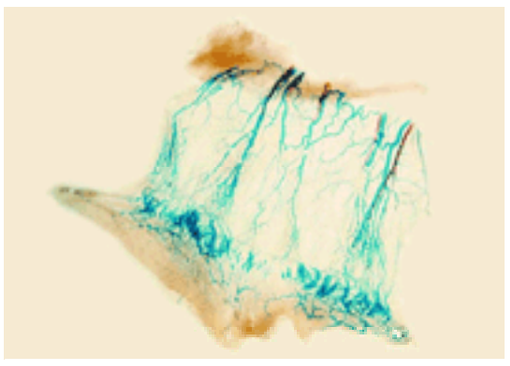


Fig. 7. Venous saccules in an infant anorectum demonstrated by (blue) latex injection and a tissue clearing technique. In babies drainage into the perianal veins below the anal sphincter is more readily demonstrable than in adult specimens.



Fig. 8. Engorged normal vein sacs in the adult visible through the distracted anoderm of a patient undergoing examination (under general anaesthetic) for a fistula (dilute methylene blue is being injected into it through a fine cannula, lower right, and appearing in the anal canal). Anodermatitis, perhaps from local medication, is manifest as readily visible punctate excoriations and splitting.

Support

The cushions are held against the shearing, extruding force of defaecation by smooth muscle, the musculus submucosae ani, and by elastic tissue. This muscle, the second unique anatomic feature—nowhere else is muscle to be found in the submucosa—was discovered by Treitz (1853) and has been observed and drawn by others since. It descends from the internal sphincter in separate bundles (Fig. 9(a)) which coalesce subanodermally to form a dense supporting stroma around the vein sacs (Fig. 9(b)). A longitudinal section (Fig. 10) shows its full extent and demonstrates how the looser upper part of the cushions is supported by the tough more strongly secured lower component, and how the muscle's contraction, which occurs during defaecation, both flattens the cushions and braces them against the internal sphincter.

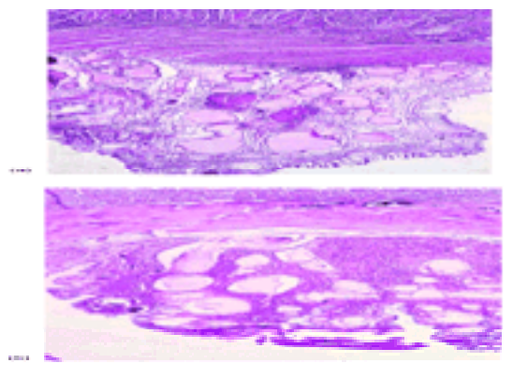


Fig. 9. (a) Transverse section of anal cushion above the dentate line, stained with haematoxylin and eosin. The vein sacs are distended with faintly pink appearing barium sulphate suspension. Two separate bundles of deep pink staining smooth muscle of the musculus submucosae ani are seen amongst the venous saccules in the submucosa. (b) Transverse section of an anal cushion below the dentate line where the vein saccules are now surrounded by dense deep pink staining smooth muscle. There is also abundant elastic tissue and collagen, not demonstrated by this stain, all providing much firmer support for the subanodermal part of the anal cushions.

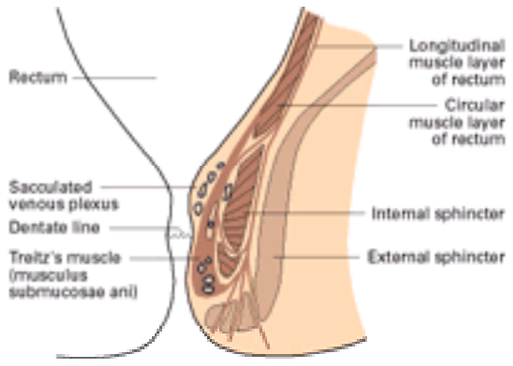


Fig. 10. Diagrammatic longitudinal section of anal canal illustrating the derivation and distribution of Trietz's muscle (muscularis submucosae ani).

Function

There can be no doubt that the anal cushions contribute to anal closure. The spongy substance and variable volume conferred by the vein sacs, with their direct arterial communications, imparts an appropriate texture—firm, too, below and floppy above—upon which the sphincter can ‘squeeze' to complete closure. Of interest, an inner tube of an inflated tyre falls naturally into three parts if a band around its perimeter is tightened (David Tibbs, John Radcliffe Hospital—personal communication).

The nature of piles

The anal cushions can as we have seen be viewed from above in an appropriately prepared specimen (Fig. 2). They can also be seen anoscopically (Fig. 11), in transverse histological section (Fig. 12(a), Fig. 12(b)), and by holding the dissected-free anal lining up to the light (Fig. 13). By whatever means of demonstration they are found constantly to occupy the left lateral (3 o'clock), right posterior (7 o'clock), and right anterior (11 o'clock) sectors of the anal circumference, and not infrequently the posterior midline, which of course is where piles present. It is logical to conclude, therefore, that the condition we call piles or haemorrhoids results from the internal disruption and downward displacement of the anal cushions, a conclusion supported by both their macroscopic (Fig. 14) and microscopic appearance (Fig. 15(a), Fig. 15(b)). Varicosity of the anal veins, when it (rarely) occurs (Fig. 16), looks quite different.

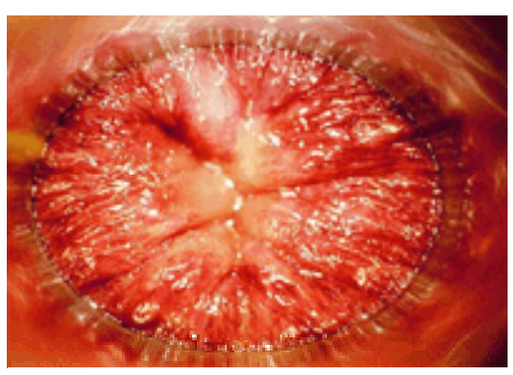


Fig. 11. The proctoscopic appearance of normal anal cushions.

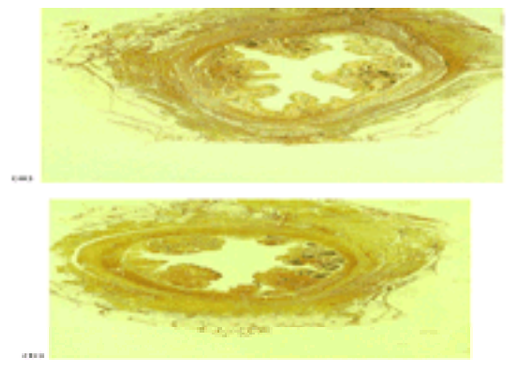


Fig. 12. (a) Transverse section of the anal canal prepared to demonstrate the cushions above the dentate line. (b) Transverse section of the anal canal prepared to demonstrate the cushions below the dentate line.



Fig. 13. The anal lining dissected from the internal sphincter and held up to the light.



Fig. 14. Prolapsed piles.

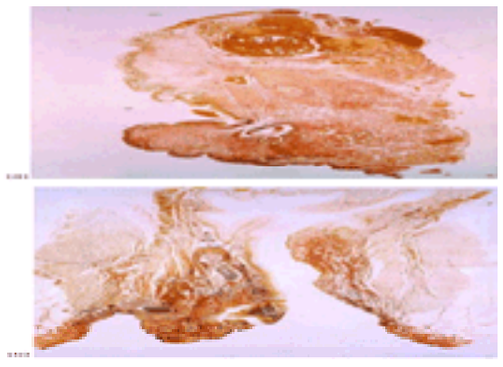


Fig. 15. (a) Longitudinal section of a haemorrhoidectomy specimen showing smooth muscle, connective tissue, and vein sacs. (b) Longitudinal section of anal canal from an autopsy specimen showing a prolapsed pile. Hypertrophied Treitz's muscle bundles are clearly seen.

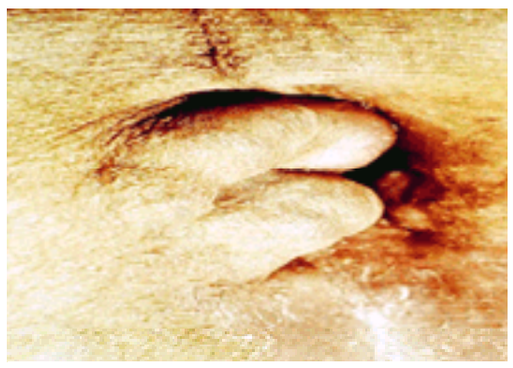


Fig. 16. Anal varices in portal hypertension.

Pathology

The anal cushions are disrupted to produce piles by the forces of defaecation. For many sufferers defaecatory habits and stool consistency are probably to blame. The Valsava effect of excessive straining engorges the cushions, which have lost the support of the external sphincter as it relaxes. The shearing force of hard stools will increase the damage. In other patients who claim a lifetime of regular easy bowel actions, the anal cushions may be structurally deficient. Weakness arising from the influence of progesterone on smooth muscle and elastic tissue may explain the predisposition to haemorrhoids in pregnancy, though an increase in pelvic vascularity may contribute. Many women date their haemorrhoids not to actual pregnancy but to parturition, when the supporting tissues of the anal cushions may be stretched and torn.

Histological examination often shows larger vascular spaces than normal and more prominent connective tissues but no changes not accounted for by the effects of disruption.

Classification

It is customary to classify haemorrhoids by degree: first degree, only bleeding announces their presence; second degree, spontaneously reducing prolapse at defaecation; third degree, prolapse requiring manual replacement; fourth degree, permanent prolapse. However, while a classification is required for the purpose of comparing different treatment techniques scientifically, the degree of any particular patient's piles may in fact vary with time.

The terms ‘internal' and ‘external' piles add little.

Symptoms

Although the underlying lesion in piles—disruption of the supporting and anchoring tissues of the cushions—means that prolapse is inherent in their nature, bleeding is more worrying and is the usual reason for seeing a doctor. Prolapse is, however, the other unequivocal symptom. Pain, itching, and anal dysfunctional effects are less reliable diagnostic criteria.

Bleeding

The capillaries of the lamina propria are only protected by a single layer of epithelial cells, and little trauma is required to breach them. Since it is the more lax-textured, upper part of the anal cushion which mainly prolapses, dragging the mucosa to the outside, trauma due to wiping or contact with clothes often occurs. Repeated trauma produces a chronic inflammatory response, making the damaged mucosa a brighter red, and granular (Fig. 17) and so more friable and likely to bleed.

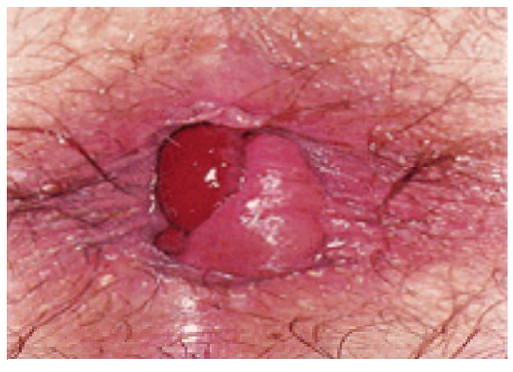


Fig. 17. Inflamed permanently prolapsed pile.

A great deal of unnecessary investigation, which is costly, inconvenient, uncomfortable, and occasionally even hazardous for the patient, can be avoided by time spent unravelling exactly what is meant by bleeding. Patient and courteous attention to detail in taking a history is always amply repaid, but never more so than here: haemorrhoids are very common, and yet bleeding may also indicate a more serious condition. First- and second-degree piles, which remain intra-anal except at defaecation, bleed with the bowel movement. Being capillary blood it is bright red. If enquiry reveals that it occasionally drips, an anal origin is certain, because the anus remains closed by tonic contraction of the sphincter except at the moment of defaecation. Blood that drips into the pan, after passage of the stool, must originate

either from extruded anal mucosa, or from a fissure in the anoderm. The only other, and most uncommon, possibility is a rectal polyp on a long enough stalk. Similarly, bleeding into clothing is almost certainly of anal origin. Blood smeared on the stool in the pan is ominous and unlikely to be coming from piles, since freshly shed blood ought to disperse into the water. The fact that it remains on the stool suggests either that it has congealed there, or is mixed with mucus, indicating a higher lesion.

Passage of clotted blood also demands exculpation of a colorectal source, and a careful history may provide a useful clue. Piles may still be the explanation if questioning reveals that the clots were only seen on the paper; such clotting can have occurred in freshly shed blood lying at the anal verge. It is very rare for a large pile to bleed back into the rectum and proclaim itself by passage of older clots at stool.

Prolapse

Many patients have not tried manual replacement of their piles after defaecation, having been ‘afraid to', and therefore put up with more discomfort than they need. Others replace them promptly only to be demoralized and inconvenienced by their messy extrusion on exertion later.

Pain

Pain is a contentious issue in pile symptomatology. Although claimed to be a prominent and attributable problem, there seems to be no good reason why a disrupted anal cushion should actually be painful. When trapped outside the closed anus, distortion combined with oedema and congestion from lymphatic and venous impairment may well cause discomfort. In many cases pain on defaecation is due to an easily overlooked fissure. None the less, some patients do experience relief from what they had thought of as pain from successful treatment of their uncomplicated piles, and the wise clinician allows for some hyperbole, perhaps, in description.

Episodes of painful irreducible swelling which last a week or so can be most unpleasant. Often called ‘strangulated' piles or an ‘attack of the piles', they are usually due to greater or lesser degrees of infarction resulting from obstruction of venous drainage by thrombosis and consecutive clotting in the sacculated venous plexus. Infarction is used here in its proper sense, denoting an intravascular and interstitial ‘stuffing with blood', and not in its common contemporary misusage implying necrosis. Although necrosis would supervene if circulatory impairment by venous blockage were sufficient, complete obstruction of venous return is in fact very rare and the usual outcome is spontaneous resolution as the clot shrinks and lyses and venous circulation is restored.

Itching

When the patient's main concern is itching, piles are seldom to blame. A local skin condition is usually responsible. Although treatment of coexisting piles may procure relief, it is unwise to encourage a patient greatly bothered by pruritus to believe that the answer is at hand. Mucus discharge from a prolapsed pile, however, causes an alleviable irritation in some patients.

Anorectal dysfunction

Defaecatory derangement can be excited by disrupted anal cushions causing a sensation of incomplete evacuation, particularly when further engorged by fruitless straining. Of course a feeling of unsatisfied defaecation—tenesmus—may have a more serious explanation.

Soiling

Blood and serum from the exposed inflamed mucosal part of a pile dries dark on underclothing and may be thought fecal. Only very rarely, however, do third- and fourth-degree haemorrhoids allow minor conduction of rectal contents to the surface. Mucus may also exude from the exteriorized mucosa of piles and can be the presenting symptom.

Examination

When a meticulous history suggests piles and the findings agree, examination can be confined to the anorectum. The only equipment then required is proctoscope (anoscope), rigid sigmoidoscope (rectoscope), light source, and biopsy forceps. Many, however, would disagree and argue for routine adjunctive fiber-optic inspection of the distal colon, at least in those over 40 years of age for rectal bleeding, however described and whatever the anoscopic findings. Some workers in this field indeed advocate full colonoscopy in patients aged 40 or over presenting with rectal bleeding even when bright red, on the basis of the frequency of finding right-sided pathology in those of middle age and older, but in the author's view theirs is more an argument—still unresolved—for screening.

Signs

There are several dynamic influences on a pile's presentation—the vigorous arterial supply, the presence and possibly changing diameter of the arteriovenous shunts, the variability of cushion bulk due to the capacity of the venous saccules, and the effects of cushion displacement and anal sphincter contraction on venous and lymphatic drainage. As a result, not only does the appearance change from time to time in the same patient, but the same symptom may have different causes. For instance, whereas most people complaining of prolapse have simple displacement of the anal cushion(s) (Fig. 14, Fig. 17), a ‘lump' felt by others may be due to engorgement of the subanodermal veins from, one presumes, impaired drainage (Fig. 6, Fig. 18) or transient but most uncomfortable postdefaecatory anodermal oedema. (Fig. 19).

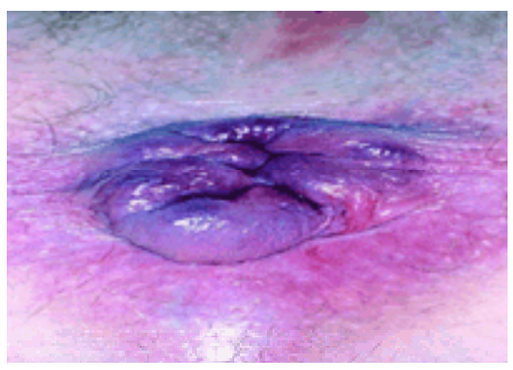


Fig. 18. Engorgement of the subanodermal veins masquerades as prolapse in some patients. Paucity of drainage of the venous saccules trapped outside by sphincteric closure (Fig. 6(a)) may explain this phenomenon.



Fig. 19. Postdefaecation oedema also occasionally contributes greatly to the bulk of ‘prolapse', and must result in some way from impairment of either venous or lymphatic drainage.

Piles that are transiently displaced suffer little trauma, but when the mucosal part is frequently exposed it becomes inflamed (Fig. 17). Thrombosis and clotting in the vein sacs also influence the appearance of the pile, but as an indication there will be associated discomfort or, depending on the extent of clotting and consequent infarction, frank pain. Clotting of a small part of the venous plexus causes an uncomfortable attack of swelling of the pile, with oedema but little infarction. Greater degrees of obliteration of venous drainage embarrass the circulation accordingly (Fig. 20). However, even the fully infarcted pile (Fig. 21) will, despite its appearance, resolve. Many patients who seek medical attention because of such an attack of saccular clotting, and who graphically describe the severity of the condition, have recovered by the time of specialist consultation. The term ‘strangulated piles' may be misapplied to this condition, causing inappropriate and inevitably unsuccessful efforts at supposedly therapeutic replacement.



Fig. 20. Early venous clotting in an anal cushion. Individually clotted saccules can be seen through the subanodermal oedema.



Fig. 21. Later effects of clotting of the sacculated venous plexus with infarction of the pile.

A disordered cushion may, therefore, present in one of several ways as a lump at the anal verge. Commonly, however, external inspection provides no clue to their presence, and nothing abnormal is found on anal digitation, since uncomplicated piles are impalpable. A nodular induration is felt if clotting has occurred. In most patients, the diagnosis is suggested by the history and confirmed with the anoscope. Interpretation of the appearance is not straightforward. Since anal cushions are normal structures (Fig. 11), their distinction from piles (Fig. 22) is only one of degree. Bright red granularity of the mucosal part of a cushion is certain evidence of its disruption, and the extent to which the cushions bulge into the instrument's end on straining and follow it out on withdrawal, provide a valuable guide. The beginner will often miss the anoscopic diagnosis of piles from failing to ask the patient to push down as if defaecating as the instrument is gradually withdrawn. A previously unimpressive anal cushion may then demonstrate its obvious disruption.

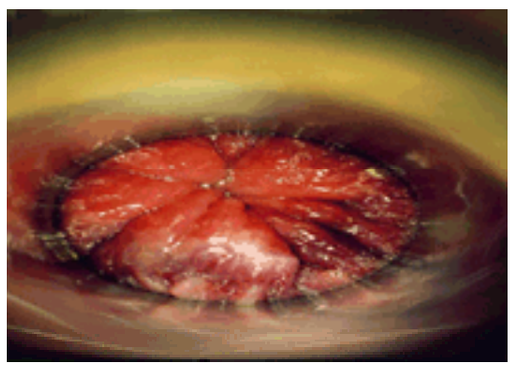


Fig. 22. Proctoscopic appearance of a disrupted anal cushion.

Rectoscopic (rigid ‘sigmoidoscopic') exclusion of rectal disease is an essential part of the establishment of the diagnosis. Because piles are common, finding them does not rule out another condition higher in the rectum causing the symptoms. There is, however, no evidence for the claim still occasionally made that haemorrhoids can result from rectal carcinoma or pelvic masses.

Differential diagnosis

Anal tags

Many patients mistake anal tags for piles, and indeed the disrupted anodermal part of a cushion may have a similar appearance. Anal tags are cutaneous protruberances at the junction of the anoderm and perianal skin. They are of uncertain origin, but possibly result from local derangement of lymphatic drainage—as their occasional disarming partial reformation soon after excision suggests. They can be solitary and discrete, or form a circumferential irregular fringe (Fig. 23).



Fig. 23. Skin tags.

Fibroepithelial polyp

These are club-like protruberances from the dentate line and seem to be hypertrophied anal papillas, again possibly due to lymphatic obstruction (Fig. 24).



Fig. 24. A fibrous anal polyp, in this case surmounting a pile which consists, unusually, entirely of the disrupted anodermal component of the cushion.

Sentinel pile

This misnomer is given to a skin tag marking—and often containing within it—the distal end of an anal fissure, found usually in the posterior midline (Fig. 25).



Fig. 25. A midline oedematous tag, or ‘sentinel pile', at the lower end of a posterior fissure. Note also the perianal dermatitis with punctate excoriations, perhaps in this case iatrogenously arising from a reaction to locally applied local anaesthetic cream.

Fissure

A patient described as having ‘painful itchy piles' may well be suffering from an anal fissure, particularly when associated with a sentinel tag masquerading as a pile. The deep burning pain of a fissure on and after defaecation and the associated itching are quite unlike the discomfort appropriate to a pile. Typically, too, the pain of a fissure will start some 30 min after defaecation and continue for 2 to several hours.

Dermatitis

Because of the widespread belief, both in the lay and the medical mind, that itching and soreness mean piles, in many patients referred for a surgical opinion the problem is in fact dermatological. Hyperkeratosis (seen as pale, slightly soggy or glazed skin), erythema, punctate excoriations, and multiple hairline radiating cracks—often more pronounced in the anterior or posterior midlines—will suggest the correct diagnosis (Fig. 8, Fig. 25). While the occasional patient may have psoriasis and quite a few are prone to itchy rashes elsewhere, in most, in the author's experience, the problem is confined to the anus. The pain at defaecation of anodermatitis usually lasts only a matter of seconds.

‘Perianal haematoma'

The term ‘pile' might easily have been inspired by this condition (Latin: *pila* = ball) so spherical and usually singular is it (Fig. 26(a), Fig. 26(b)), and its other name, ‘thrombosed external pile', is not entirely inappropriate. However, to keep our terminology exact we should use ‘pile' to denote a disrupted anal cushion. These often painful lesions of sudden onset and, when not relieved by incision, of usually self-limiting nature (either by rupture or by absorption) are not in fact the ruptured blood vessel the term suggests (nor are they strictly speaking ‘perianal', arising as they do subanodermally) but vein sacs distended by clot. The term ‘clotted vein sac' describes them accurately but though preferable is unlikely to supersede its time-honoured alternative.



Fig. 26. (a) So-called ‘perianal haematoma'; in fact a single venous saccule greatly distended with clot. (b) Histological section of so-called ‘perianal haematoma' showing a vein sac distended with clot.

Rectal prolapse

Early rectal prolapse may be confused with piles when the patient is unable to describe the size of the protrusion and is inhibited from straining sufficiently at the

examination to produce it. Treatment for piles may then be instituted with later disappointment, but no substantial harm done.

Rectal tumour

Rectal tumour can easily be missed by impatient history taking and unreflective digital examination, for it is not so much the length of the finger which matters as the amount of thought behind it. Because of the rectum's curvature, even upper-third tumours may be palpable. Even when nothing is felt or seen, if the patient's symptoms do not accord with the findings further investigation is required. Ominous symptoms are old blood, particularly if slimy or clotted, tenesmus, altered bowel behaviour, deep discomfort, and ‘wet'—messy—flatus.

Miscellaneous

It is a curious fact that almost any discomfort in or irregularity of the anorectum may be attributed to piles. Thus proctalgia fugax, proctitis, solitary rectal ulcer, fistula-in-ano, and warts may also all masquerade.

Treatment

Since piles may be blamed for almost any anal condition the first step is to decide whether they could be responsible for the symptoms. When itching is the main complaint piles are unlikely to be the cause, and actual pain—rather than the discomfort of protrusion or episodes attributable to attacks of thrombosis—is more likely to be due to a fissure. Chronic anal pain, of course, is never due to piles. With prolapse and bleeding, however, we are on firmer ground.

The management will be either conservative or interventional.

Conservative treatment

Although piles are by definition disrupted anal cushions, symptoms from them are partly determined by their size, which can alter greatly depending on the state of engorgement of their constituent vein sacs, which in turn is affected by straining and the state of tone of the surrounding anal sphincter. Simple avoidance of prolonged straining at stool may achieve sufficient symptomatic relief. An increase in dietary fiber, therefore, and desistance from reading in the lavatory, together with advice to ignore the false signal suggesting the need for a greater straining effort imparted sometimes by prolapsing piles may be enough.

Interventional treatment

Surgical measures work by reducing the bulk of the disrupted anal cushion (not only has its attachment loosened but its internal structure as well), and inducing adhesion of the remainder. Since the mucosal part causes most of the symptoms—all of the bleeding and mucus discharge, and most of the discomfort—its reduction will be sufficient for the majority of patients. Even a large cutaneous component may be pulled up and so be flattened and tidied by tissue reduction above. It is therefore fortunate that the mucosa is insensitive and so allows quick and effective treatment in the outpatient clinic or office. Various techniques are available.

Rubber band ligation

In ‘banding', as it is called, a ‘polyp' of the pile centred sufficiently cephalad to prevent involvement of the sensitive anoderm is pulled into a ligating device passed down the anoscope and strangulated by displacing from the ring-mounting a small rubber O-ring previously rolled there from a cone loader. The incorporated tissue withers and falls away within 2 or 3 days leaving a small ulcer which is usually healed in a month.

The size of the ‘polyp' banded will depend on the toughness of the tissue, the volume of redundant loose cushion, and the traction force applied by the operator. Previous intervention—banding, injection, and so on—may cause such dense scarring as to prevent the incorporation of sufficient tissue to help. Usually, however, with good grasping forceps a polyp the size of an average raspberry can be banded, and sometimes, gratifyingly, the size of a large one. Because the anatomy of the anal canal and the mechanics of banding make it impossible to ablate enough tissue to be damaging, the author takes the view that the greater the volume of the mucosal part of the pile which can be ensnared the better. The pile is therefore first carefully ‘sized up' down the anoscope to choose the site to be seized by the graspers which will ensure the strangulation of the greatest part of the pile without involving the anoderm. If the amount of tissue obtained seems less than expected, a further application, this time firmly grasping and pulling on the previously banded polyp, will often achieve more (Fig. 27(a)). If on the other hand the band looks to have been placed too high or too low, a further adjoining application is worthwhile (Fig. 27(b)).

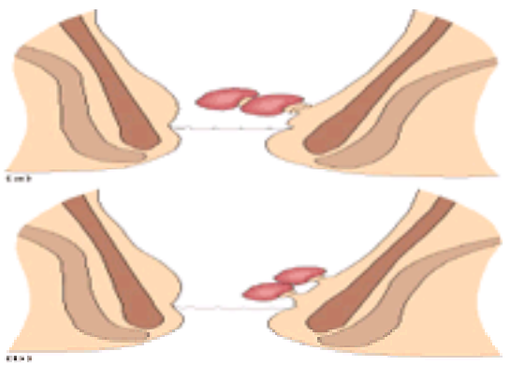


Fig. 27. (a) A second banding attempt will occasionally improve on the achievement of the first. (b) If the first ‘polyp' is insufficient but has been banded too low (too near the dentate line) for enlargement, an adjoining cephalad band can be placed.

Even permanently prolapsed irreducible piles can be cured by banding. The band can be applied to the exposed mucosa outside the anal canal with complete symptomatic relief, even the cutaneous/anodermal element being improved, and anyway seldom much of a nuisance (Fig. 28).

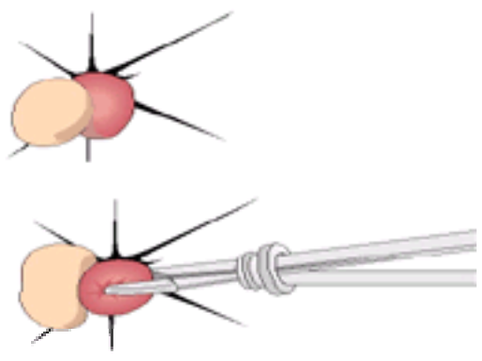


Fig. 28. Banding the exposed mucosal part of an irreplaceable, permanently prolapsed pile.

Several banding instrument models are available and although they work on the same broad principle there are two main categories, single-operator and assistant-required. As well as their self-evident attraction, single-operator banders confer the advantage that in holding the proctoscope throughout the surgeon can keep it exactly on the selected target. There are three types. One, a suction device (which can be driven by the Venturi effect of a running tap), uses a metal sucker tube—with the banding rubber ring mounted at its end and displaceable by a trigger—to suck the pile to which it is applied (under vision down any proctoscope) into it.

Although straightforward to use, the need for suction may create practical difficulties, and the impossibility of knowing how much tissue has been sucked into the end for banding and how much suction can safely be applied without causing avulsion are deterring factors. Another consists simply of a small-diameter proctoscope with the rubber O-ring, again displaceable by a releasing mechanism, mounted at the operating end. A theoretical disadvantage to its use may be the problem of inadequacy of view. The ‘One-man bander' seems to overcome both these disadvantages (Fig. 29, Fig. 30). Several loaded ones are kept ready so that even multiple banding can be the work of a moment. It is designed for use down the Naunton-Morgan ‘rectal speculum', a proctoscope, in fact, whose wide diameter allows a more certain diagnosis of piles than those of smaller bore (which may prevent their appreciation through providing too narrow a view). It is certainly the author's experience that piles can easily be missed down narrow anoscopes.

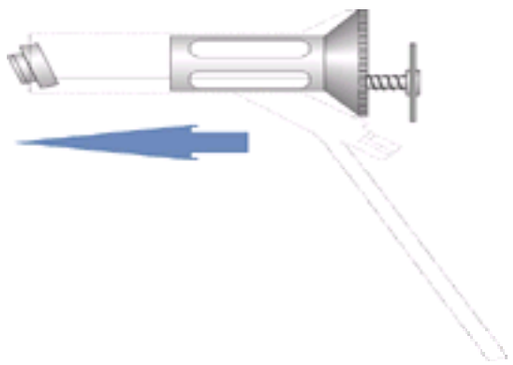


Fig. 29. Thomson One-man bander used within the Naunton-Morgan anoscope. Cone loader illustrated in line profile to show overlapping flange which prevents band from slipping between cone and bander when loading.

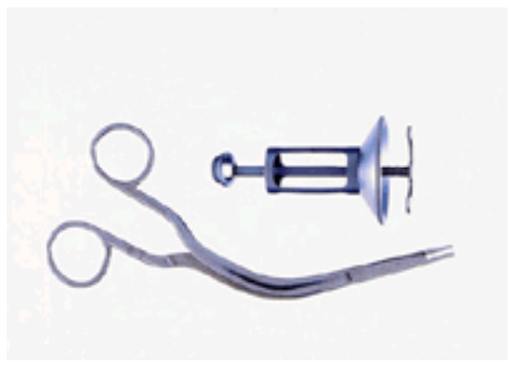


Fig. 30. Irvin Moore's nasal conchal forceps for grasping the pile for banding, shown with banding instrument.

For grasping the pile, Irvin-Moore's nasal conchal forceps seem ideally suited. They open more widely within the ring of the ligator than at least one purpose-designed model, and grip the tissue with less danger of tearing (Fig. 30).

On a note of caution, it has been the author's experience that combining pile banding with an anal stretch procedure may precipitate clotting in the residual sacculated venous plexus and so cause severe anal cushion infarction with its associated pain and swelling.

An empty rectum making banding easier and certainly more agreeable if not safer, it is best to have the patient self-administer a mini-enema or glycerin suppository prior to the banding. In addition, in case any pain afterwards would otherwise inhibit defaecation, it is sensible to advise patients to start a week's course of ispaghula husk or similar the day before. Finally, to make the procedure and its immediate aftermath as comfortable as possible the patients are also given analgesic tablets to take an hour before, and a supply to use afterwards.

Infrared photocoagulation and bipolar diathermy

Both of these cause tissue destruction by heat. The instruments are applied through an anoscope to coagulate a predictable volume of adjoining tissue. The mucosal part is treated so no anaesthesia is required.

In bipolar diathermy the mucosa is simply grasped by the instrument and the intervening tissue destroyed by an electric current passed across from one electrode to the other.

Sclerotherapy

An irritant chemical solution, usually 3 ml of 5 per cent phenol in arachis oil, is injected into the submucosa of the mucosal part of each pile. When the varicose vein theory of piles prevailed it was thought to act by inducing fibrosis which constricted the superior rectal venous drainage, so, it was thought, deterring the causative—as they thought (from the human's erect posture)—transmission of the supposed high pressures from the portal system. In fact it probably causes shrinkage of tissue by necrosis, and adhesion as a result of the ensuing inflammatory reaction.

Cryotherapy

A liquid nitrogen probe is placed against the pile for 3 min, and causes cold necrosis. Local anaesthesia, at least, is said to be needed.

Laser treatment

Laser evaporation has also been used for pile excision, with good initial reports. Its theoretical attraction lies in the reduction to the minimum of damage to residual tissue. However, it is an elaborate way of performing a simple task and uses expensive and fallible equipment in a forgiving area of great blood supply and margin for error in which excellent results can be obtained by simpler means. Were it to prove significantly less painful than scissor excision, and perhaps to allow more rapid healing, neither of which has so far been shown, then it might find widespread use.

Haemorrhoidectomy

When the troublesome cutaneous part of a pile is too large or immobile for patient relief simply from addressing the mucosal moiety above, scissor excision—haemorrhoidectomy—is required. However, although not ‘ambulant', it is therapy that can often be done under local anaesthetic and seldom requires much more than an overnight stay. The excessively painful and prolonged experience of folk memory resulted mainly from the well intentioned use of wide-bore rubber drains or packs inserted in theater against the risk of haemorrhage. Mistaken, too, was the much encouraged belief that three-sector excision was always required (‘the operation's over when it looks like a clover'), a practice unfounded in either trial or systematic experience. It arose with little doubt from the appearance of even normal anal cushions after anal manipulation in the lithotomy position, particularly when a patient is ‘light' and strains under the anaesthetic. In fact the surgeon should be guided by the disease and will therefore sometimes excise in only one place, although in other cases in four. Always, however, adequate bridges of healthy tissue must be left between excision sites because not only is the anoderm of great functional importance in continence—‘sampling' the rectal contents before release—but its excessive removal will result in a stricture, an avoidable tragedy. The other half of the mnemonic is certainly true, ‘if it looks like a dahlia it'll end up a failure'.

The patient's rectum should be emptied prior to surgery by means of an enema administered about 2 h before. Whereas in the United Kingdom the patient is usually put ‘in lithotomy' for the procedure, the prone jack-knife position is probably preferable for reasons both of access and local conditions, the anal canal's axis then being in a line with the surgeon's eye rather than, as in the lithotomy position, at an angle to it, and the vasculature not being unnaturally suffused with blood. However, the

lithotomy position and the lateral position with buttocks taped apart, also useful, are easier to arrange.

Haemorrhoidectomy can be performed either open or closed and is done with scissors or diathermy (or for that matter, laser). Either way, the essential principle, emphasized by the reminder that surgery is the replacement of one lesion by another, is to keep the damage to the minimum required for symptomatic relief. Remember, too, that piles are just overlarge, floppy, and displaced parts of a structure which almost certainly contributes to anal control, so excision should leave sufficient behind. Above all, ample ‘bridges' should be left. Finally, the midline anoderm, particularly posteriorly, being so unstable and prone to fissure, it is as well to stay away from it. Therefore, a symptomatic midline posterior pile, as can occur, may be better treated by band ligation—so above the dentate line—with acceptance of any residual cutaneous irregularity there.

Open—excision/ligation—haemorrhoidectomy was popularized by Milligan and his colleagues at St Marks Hospital, London. Forceps are applied to the cutaneous part of the pile at the anal verge to reveal the mucosal moiety. A second pair then grasps the main pile mass which is thereby lifted from its surroundings. It is then snipped from the adjoining skin in a racket-shaped cut whose wide handle, directed cephalad above the dentate line, contains the superior rectal contribution to its blood supply (Fig. 31(a)). The elliptical part of the cut overlying the intersphincteric groove passes through the subanodermal part of the haemorrhoidal venous plexus to reveal the discrete vein sacs within. It is deepened to reveal the lower border of the internal sphincter. The pile is then dissected from the internal sphincter for a few millimetres (emerging fibres of Treitz's muscle will be encountered and divided)—both by identifying to safeguard it and to allow ligation above the dentate line (so of insensitive tissue)—so creating a thick pedicle which is transfixed and ligated (Fig. 31(b)). The ligated remnant—most is excised—falls away in a few days to allow the open wound to heal by secondary intention. Despite that much diathermy coagulation may be required at the edges to achieve haemostasis, the wound usually contracts rapidly and will be found almost closed in as little as 10 days.

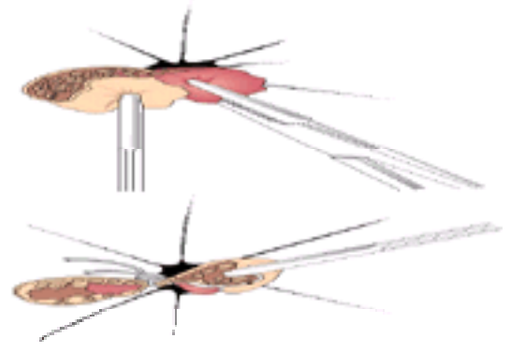


Fig. 31. (a) The first step in open—excision/ligation—haemorrhoidectomy. (b) Dissection complete and the pedicle ligated, in open haemorrhoidectomy.

The closed method, with or without submucosal excision of adjoining ‘haemorrhoidal' tissue, was designed for faster healing and less pain. (The term ‘haemorrhoidal' here illustrates the confusion which misleading terminology engenders. The spongy cavernous-like submucosa of the anal canal serves a valuable function in continence. Calling it ‘haemorrhoidal' suggests disease and so encourages excision). It has not, however, been shown to achieve either aim, neither pain relief nor healing being improved. (Confoundingly, haemorrhoidectomy was also no more comfortable after a smooth muscle relaxant or internal sphincter division (lateral sphincterotomy)).

By operating down a wide-diameter slotted speculum the pile can be excised in its anatomic position without distortion, a technique which recommends itself from first principles alone. The pile exposed in the slot is excised within an ellipse, exposing the internal sphincter in its base, narrow enough to allow closure without tension (Fig. 32(a), Fig. 32(b)). It is a technique which has been found most satisfactory.

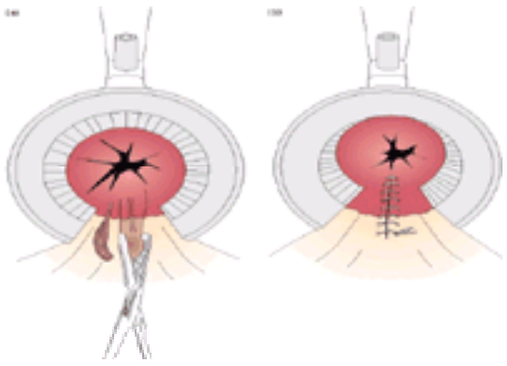


Fig. 32. (a) and (b) Closed haemorrhoidectomy performed within the exposing slot of an operating anoscope.

After haemorrhoidectomy by whichever method, light dressings are placed over the anus, perhaps kept in place by elasticated pants for easy changing. If general anaesthesia has been used the anal wounds should be thoroughly infiltrated with a long-acting local anaesthetic. The anus should not be packed. Warm baths are comforting, with or without salt, and measures are taken to keep the stools soft and regular. There is no need for the patient to stay in hospital until the bowels have moved but if they have not by the third day the patient should report back for an enema. Antibiotic cover is not required unless there is a particular predisposition to sepsis.

Manual anal dilation

This was widely practised as a treatment for piles in the last century. The nineteenth-century French surgeon, Verneuil, again reconciling its effect with the varicose vein theory, thought it improved venous drainage by stretching the rectal muscular button holes which convey the anal tributaries of the superior rectal vein. When reintroduced some years ago, it transiently displaced the surgical standby of the time, haemorrhoidectomy, but was later shown to have limited application. It probably helps patients who have discomfort and difficulty in defaecation by easing the effort of evacuation, so reducing the congestion of the cushions from excessive straining.

Pile stitching

This has been advocated. Absorbable sutures are placed above the dentate line to attach the cushion back to the internal sphincter. Obliteration of its blood supply also reduces bulk. It has a place when large piles demand amelioration yet there is a substantial risk from haemorrhage—in the presence of a cavernous haemangiomatous deformity, for instance, or when a patient must not discontinue anticoagulants.

Complications

Ephemeral

Vasovagal

A few people—mainly young males in the author's experience—faint after banding of piles, and even injection. It is therefore sensible to warn all patients to arrange to be driven home.

Pain

Some discomfort follows all the tissue ablative procedures but amounts in some patients to severe pain lasting as long as a week. Again, prior knowledge will prevent much needless worry and also allow the patient to plan in advance.

Haemorrhage

Tissue destructive or excisional techniques will also leave a well vascularized, raw wet surface and so carry the irreducible risk of secondary haemorrhage. It is rare, but when it occurs, alarming, but once more if the patient is told of the possibility beforehand there will be less anxiety and disruption. In the author's experience only once in 20 years has continued bleeding required suture for haemostasis. In the other dozen or so instances the bleeding stopped spontaneously and without resort even to attempted tamponade with a balloon catheter. A micronized flavonidic fraction of diosmin and hesperidin (Daflon) significantly reduced secondary haemorrhage in one trial, and in an experience of 12 patients with secondary haemorrhage after haemorrhoidectomy a submucosal injection of 1:10 000 epinephrine through a proctoscope under sedation secured haemostasis each time.

Infection

Sepsis is an uncommon complication and will be treated on its merits. Fatal sepsis even after simple banding has been reported in immunocompromised patients. Stories are handed down of portal pyemia following haemorrhoidectomy, and of necrotizing fasciitis complicating it. Both no doubt have occurred and will again but must be excessively rare. Neat surgery with the least trauma will surely reduce the possibility as will patient selection and judicious drainage rather than skin closure when the tissues have been compromised.

Urinary

Haemorrhoidectomy is notorious for causing transient difficulty in voiding in men. Banding may also induce curious bladder symptoms. Inadvertent intra- or periprostatic injection of sclerosant may have serious sequelae, even impotence being reported.

Anal cushion thrombosis

Occasionally, excessive pain seems to be attributable to thrombosis/clotting in the residual anal cushion after pile banding, the complication being detectable by the digital discovery of knobbly induration within the anal canal.

Permanent

Impairment of continence

A thorough study of 172 patients who had undergone haemorrhoidectomy at least 3 years previously (at a time when three sector excision was standard) found that 26 per cent had experienced a consequent impairment of control. Given the function of the anal cushions and the fact that haemorrhoidectomy may have been relegated to the end of the operating list to be performed by an unsupervised trainee, this outcome is not greatly surprising. However, if the surgeon confines tissue removal to the clearly redundant and leaves what Waldeyer called the ‘corpus cavernosum recti' substantially intact, no functional impairment need be feared.

Oedema and tags

Occasionally, however careful the excision, the intervening skin bridges may swell uncomfortably with oedema to remain afterwards as tags. Though lymphatic impairment, perhaps from diathermy damage, is presumably responsible, the aetiology of anal tags is obscure; furthermore, when the entire anal lining has dislocated to form a circular curtain at the anus, the intervening bridges will of necessity remain as tags. They can be improved later under local anaesthesia if required.

Stricture

The anal lining is extraordinarily forgiving; elastic, robust, and quick healing. Only inexcusably excessive excision will result in stenosis, or some other extraordinary circumstance.

Which treatment?

A generation ago the choice was limited; for bleeding there was injection, and for prolapse haemorrhoidectomy. Nowadays open surgery is seldom necessary. The invention of outpatient tissue-reducing techniques has not only increased the possibilities but improved the prospect. At the same time the introduction of prospective randomized clinical trials has for the first time enabled scientific evaluation.

Conservative treatment has been compared with band ligation; bulk purgatives or high fibre diet with sclerotherapy, stretching, sphincterotomy, band ligation, and freezing. Banding has also been compared with sclerotherapy, haemorrhoidectomy, stretching, and infrared photocoagulation. Photocoagulation has been compared with bipolar diathermy, and one type of haemorrhoidectomy has been judged against another. Excellent and praiseworthy though such trials are they suggest a ‘rivalry', when in fact no one treatment is best for all patients. Their greater merit is in showing prospectively and with careful control and monitoring that benefit can be derived from each measure, so allowing us an informed choice of the possible treatments.

However, while infrared- and electrocoagulation achieve the same objective with less pain and time off work, band ligation requires significantly fewer treatment sessions, and the instrument used is both cheaper and more robust. Multiple banding having been shown to be no more painful than a single application per visit, in the author's experience the great majority of patients are put right at one attendance. A recent overview, in fact, concluded that rubber-band ligation was the most effective ‘ambulant' measure available.

Haemorrhoidectomy though requiring admission, anaesthesia, and a longer recovery, still retains its place when the cutaneous part of the pile is causing the problem. Careful and targeted surgery will be rewarded by an excellent result and a most satisfied, relieved patient.

Cryodestruction is an anatomically less accurate means of tissue ablation and leaves a smelly, weeping wound slow to heal. Laser surgery offers attractive qualities but, given the rapidity of both doing and recovering from simple scissor excision and the cost of the equipment required, it seems unlikely to be preferred.

Manual anal dilation may have the occasional adjunctive place in the management of piles but would no longer be advocated as primary treatment.

Management of infarcted (‘strangulated') piles

When one reflects on the necessarily turbulent flow in the sacculated venous plexus it seems odd how seldom thrombosis and clotting occurs. Afflicting a pile it causes considerable swelling and discomfort, and invites attempts at immediate amelioration. However, there have not been, and are unlikely ever to be, randomized trials of conservative and surgical treatment. In the author's experience the infarction is only rarely so dense as to cause insupportable pain and lead to necrosis, so most patients can be reassured that natural thrombolysis will restore the circulation and result in resolution in about 10 days (Fig. 21), (Fig. 33). Nitroglycerin ointment is reported to provide dramatic relief from the pain. In severe cases, however, there is no doubt that debridement haemorrhoidectomy, excising the worse afflicted tissue, is rewarded by rapid deliverance from misery (Fig. 34).

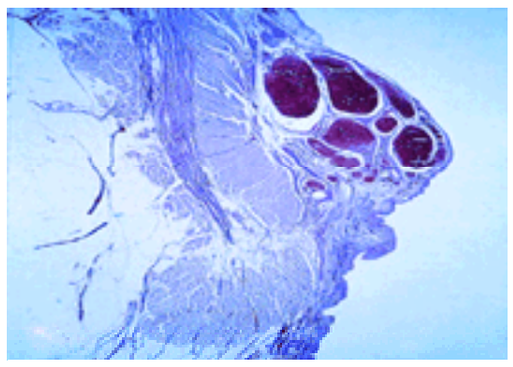


Fig. 33. Longitudinal section of an infarcted pile (chance finding at autopsy).

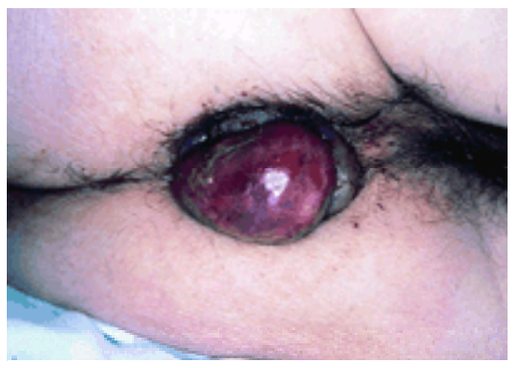


Fig. 34. Gross, extremely painful, swelling in a densely infarcted and now completely necrotic pile. Debridement haemorrhoidectomy provided dramatic relief.

For the majority sufficient relief will be obtained by rest, analgesics, and hot baths, and by the medical attendant's confident prediction that the worst will be over in a week or at the most 10 days. Meanwhile they will be best advised to ensure soft regular stools by the daily consumption of ispaghula husk or its alternatives.

Conclusion

Once one understands the detailed anatomy of haemorrhoids and their pathological possibilities, their management is straightforward, and the treatment effective, acceptable, and reasonably trouble free. The important part is diagnosis. The patient's preconception that piles are responsible for their symptoms must not cloud the clinician's mind, for it does not take long experience in a busy anorectal disorder practice to realize that many patients sailing under the pennant of piles have something functional, dermatological, or otherwise idiopathic causing their symptoms. Uppermost must be the question whether reducing the bulk of the anal cushions and mooring them more firmly will logically address the presenting problem. If not, then in a nineteenth-century surgeon's words, ‘in such cases prudence equally forbids the rash interposition of unavailing art, and the useless indulgence of delusive hope'.

Further reading

Bennett RE, Freidman MHW, Goligher JC. Late results of haemorrhoidectomy by ligature and excision. *British Medical Journal* 1963; **ii**: 216–19. [A functional study of 172 patients who had undergone haemorrhoidectomy 3 or more years previously. Some 26 per cent had experienced consequent impairment of control.]

Bullock N. Impotence after sclerotherapy of haemorrhoids: case reports. *British Medical Journal* 1997; **314**: 419. [A report of three men who became impotent after sclerotherapy for haemorrhoids. It is suggested that the mechanism is paraprostatic autonomic nerve plexus damage from misplaced injection.]

Gorfine SR. Treatment of benign anal disease with topical nitroglycerin. *Diseases of the Colon and Rectum* 1995; **38**: 453–6. [Five patients with thrombosed piles enjoyed ‘dramatic relief of pain' from 0.5 per cent nitroglycerine ointment.]

Ho YH, Foo CL, Seow-Choen F, Goh HS. Prospective randomized controlled trial of a micronized flavonidic fraction to reduce bleeding after haemorrhoidectomy. *British Journal of Surgery* 1995; **82**: 1034–5. [A prospective randomized controlled trial in 228 patients of the effect of daflon after haemorrhoidectomy found a significant reduction in the rate of secondary haemorrhage.]

Ho YH, Seow-Choen F, Low JY, Tan M, Leong AP. Randomized controlled trial of trimebutine (anal sphincter relaxant) for pain after haemorrhoidectomy. *British Journal of Surgery* 1997; **84**: 377–9. [A prospective randomized controlled trial in 160 patients of the anal muscle relaxant trimebutine after standard haemorrhoidectomy found no reduction in pain despite a lowered mean resting anal pressure.]

Ho YH, Seow-Choen F, Tan M, Leong AFPK. Randomized controlled trial of open and closed haemorrhoidectomy. *British Journal of Surgery* 1997; **84**: 1729–30. [A prospective randomized comparison of open and closed haemorrhoidectomy in 67 patients showed faster healing after the open method but no difference in pain or complications.]

MacRae HM, McLeod RS. Comparison of hemorrhoidal treatment modalities. A meta-analysis. *Diseases of the Colon and Rectum* 1995; **38**: 687–94. [A meta-analysis of all randomized controlled comparisons of treatment modalities found that rubber band ligation was best for first to third degree piles.]

Mathai V, Ong BC, Ho YH. Randomized controlled trial of lateral internal sphincterotomy with haemorrhoidectomy. *British Journal of Surgery* 1996; **83**: 380–2.[A randomized controlled trial in 33 patients found not only no significant reduction in postoperative pain by adding lateral internal sphincterotomy to haemorrhoidectomy but an increased risk of incontinence.]

Milligan ETC, Morgan CN, Jones LE, Officer R. Surgical anatomy of the anal canal and the operative treatment of haemorrhoids. *Lancet* 1937; **ii**: 1119–24. [The first description of excision/ligation (‘open') haemorrhoidectomy, which had been originated by Salmon at St Mark's Hospital, London, in the nineteenth century.]

Nivatvongs S. Hemorrhoids. In: Gordon PH, Nivatvongs S, eds. *Colon, rectum, and anus*. Quality Medical Publishing Inc., Missouri, 1992. [A well-illustrated description of Fansler's intra-anal ‘anatomic' haemorrhoidectomy.]

Nyam DC, Seow-Choen F, Ho YH. Submucosal adrenaline injection for posthemorrhoidectomy hemorrhage. *Diseases of the Colon and Rectum* 1995; **38**: 776–7. [Complete haemostasis was obtained by submucosal 1:10 000 epinephrine injection under sedation in 12 of 12 patients suffering secondary haemorrhage after haemorrhoidectomy.]

Perez-Miranda M, Gomez-Cedenilla A, Leon-Colombo T, Pajares J, Mate-Jimenez J. Effect of fiber supplements on internal bleeding hemorrhoids. *Hepato-Gastroenterology* 1996; **43**: 1504–7. [A prospective randomized controlled study in 50 patients with bleeding internal haemorrhoids found the addition of fibre to the diet (compared with placebo) to result in both an objective and a subjective improvement.]

Pfenninger JL. Modern treatment for internal haemorrhoids. *British Medical Journal* 1997; **314**: 1211–12. [A leading article concludes from an overview of published data that ‘scalpel surgery is now rarely needed'.]

Thomson WHF. The nature of haemorrhoids. *British Journal of Surgery* 1975; **62**: 542–52. [A detailed anatomical study of the structure and blood supply of the anorectum in 95 cadaveric specimens, 10 them from infants.]

Thomson WHF. The one-man Bander: a new instrument for the elastic ligation of piles. *Lancet* 1980; **ii**: 1006–7. [A new device is presented.]

Thomson WHF. The real nature of ‘perianal haematoma'. *Lancet* 1982; **ii**: 467–8. [A pathological study of so-called ‘peri-anal haematoma'.]

28.2 Pruritus ani

C. G. Marks

Aetiology
Generalized medical diseases
General dermatoses
Examination
Treatment

Pruritus is defined as itching of the skin, especially without visible eruption (*The Oxford English Dictionary*). When this is confined to the perianal area it is known as pruritus ani (Fig. 1). Individuals vary in their capacity to feel itching: some hardly notice an extensive dermatitis while others suffer greatly from an eruption that does not usually cause itching. Within a single individual the threshold of itching also varies with the state of concentration, tiredness, or boredom.



Fig. 1. Pruritus ani.

Aetiology (Fig. 2)

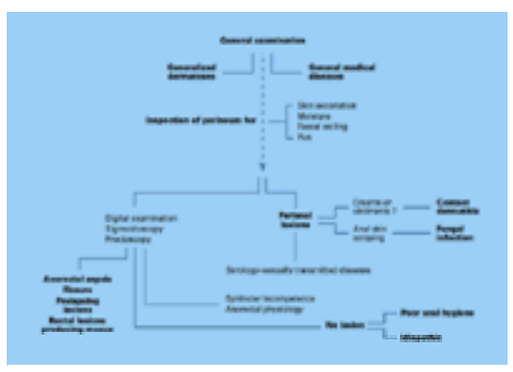


Fig. 2. Diagnosis of pruritus ani.

Generalized medical diseases

Generalized itching is a common complaint. In the initial investigation, parasitic diseases such as scabies and infestation with body lice must be excluded. A number of organic diseases, including anaemia, uraemia, diabetes, liver disease, and reticuloses, may present with itching. Clinical, biochemical, and haematological examinations will exclude most of these conditions. In the absence of organic disease or local skin disorders, a diagnosis of psychogenic irritation may be made. One of the most common sites for intractable localized itching is the anogenital region.

General dermatoses

On inspection of the perineum, the anal lesions responsible for pain, discharge, and difficulty in cleaning the perineum may be seen.

Local applications, particularly local anaesthetic ointments, are a common cause of contact dermatitis. This may be exacerbated by excessive trauma caused by toilet paper (bottom polishing!), or contact with perfumes in tissues, soaps, and biological detergents.

Fungal infection may be primary or secondary to existing lesions. Recent treatment with broad-spectrum antibiotics may encourage growth of yeasts (*Candida* spp.) Dermatophytic ringworm should be suspected if there are active lesions between the toes, which must always be inspected. Threadworm infestation is often found in children.

Sexually transmitted disease such as herpes, anal warts, and human immunodeficiency virus (**HIV**) infection, must be excluded, particularly in the male with poor anal tone, a finding that suggests the possibility of homosexuality and the risk of HIV infection.

Sphincter incompetence will be seen in a small group of patients in whom a sphincter repair may be appropriate.

Examination

The patient must be undressed completely for a full examination and inspection of the skin. The perineum should be examined carefully for moisture and soiling, skin maceration and excoriation, a skin eruption, perianal lesions, and prolapse. Separation of the buttocks will demonstrate the presence of a gaping anus due to poor anal tone, and digital examination will confirm this. Digital examination together with sigmoidoscopy and proctoscopy, will also identify sphincter incompetence, fistulas, haemorrhoids, and rectal lesions.

If fungal infection is suspected, skin scrapings should be taken for microbiological examination. Serological tests must be requested if sexually transmitted disease is suspected. Microscopy and culture are necessary to exclude specific infections by agents responsible for diarrhoea. Threadworm infestation can be diagnosed by the presence of ova on microscopic examination of a swab from the perianal skin. Rectal lesions should be biopsied and anorectal physiological studies made in patients with sphincter incompetence.

Treatment

In severe pruritus a vicious circle becomes established (Fig. 3): the pruritus causes scratching that leads to secondary bacterial or fungal infections, which are treated

by a multitude of topical applications, often containing local anaesthetic, which in turn produces more sensitization and more scratching, making matters worse.

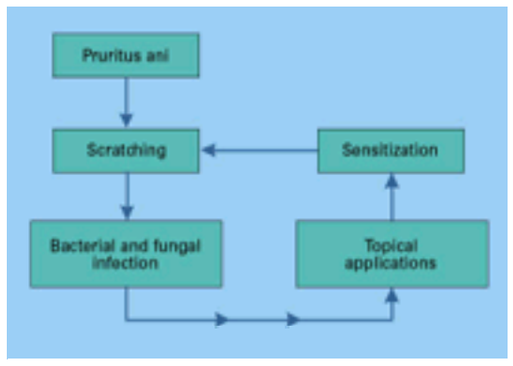


Fig. 3. The vicious circle.

The first step is to control infection and then to stop all topical applications. A cream containing low-dose hydrocortisone (1 per cent) combined with an antifungal agent (miconazole nitrate 2 per cent) should be applied to the perianal area. If severe inflammation and especially yeast infection are present, a stronger steroid is used, for example an ointment containing triamcinolone acetonide 0.1 per cent, gramicidin 0.025 per cent, neomycin 0.25 per cent, and nystatin 1 000 000 units/g. These ointments should be used for approximately 2 weeks, and sparingly thereafter. In the long term a patient with very dry skin may be helped by moisturising cream or lotion (Johnson's Baby Lotion for example); conversely, a powder will dry excessively moist skin.

Despite the topical application of medications, itch at night may continue to be a major problem. An antihistamine (promethazine hydrochloride, 10–25 mg) may control this, although it may cause drowsiness. Anal lesions that require surgery should be excised after the pruritus has improved. With this approach the majority of patients with pruritus will be cured, although most understand that a lasting cure requires a great deal of effort on their own behalf. At the end of the first consultation, the patient should be given general advice on anal hygiene, which is reinforced by a leaflet giving a list of do's and don'ts. This leaflet is a good way of ending the consultation and providing the obsessional patient with a well-defined strategy.

28.3 Anal fissure

Emin A. Carapeti and R. K. S. Phillips

[Aetiology and pathogenesis](#)
[Clinical features](#)
[Diagnosis](#)
 [Differential diagnosis](#)
[Management](#)
 [Conservative measures](#)
 [Operative measures](#)
 [Pharmacological treatments](#)
[Follow-up and prevention](#)
[Further reading](#)

Anal fissure is an elongated split in the squamous epithelium of the distal anal canal, which usually extends from the anal margin to a point below the dentate line. Most idiopathic fissures occur in the midline posteriorly, but 10 per cent of fissures in women and 1 per cent in men occur anteriorly. Fissures situated eccentrically around the anus are not usually idiopathic. They may arise in some patients with evacuation difficulty who anally digitate, in inflammatory bowel disease (particularly Crohn's disease), or from specific infection or cancer. They should be approached with caution and biopsy is often wise. Treatment of idiopathic anal fissures remained largely unchanged during most of the twentieth century until recently, when novel pharmacological treatments allied to concerns about minor but permanent incontinence in a minority of surgically treated patients have led to a complete reappraisal of the management of this common condition.

Aetiology and pathogenesis

Current thinking favours the ‘ischaemic ulcer' hypothesis. An initial traumatic event leads to tearing of the pecten and pain. A constipated stool is often implicated, but there can be many other causes. The pain leads to associated anal spasm and it is this that then impedes the local blood supply, interfering with healing. Over time the fissure becomes an ischaemic ulcer. There have been many speculations on why the posterior midline is so commonly the site of an anal fissure, including the supposed lack of anatomical support to this area favouring tearing. Studies of local arterial anatomy and blood supply demonstrate a vascular watershed in the midline posteriorly that makes this a particular site for poor healing.

Fissures may be seen in infants, where they are usually associated with constipation rather than sexual abuse. Acute fissures are very common and superficial, with little associated oedema or inflammation. They are small, fresh, linear cuts and often occur in children or young adults, usually healing spontaneously. A ‘classic' chronic fissure can mostly be diagnosed by inspection, as there is an associated oedematous skin tag at the verge, often obscuring the fissure itself, or a fibroepithelial polyp proximally in the anal canal (Fig. 1). In chronic fissure the ulcer is deep and fibres of the internal sphincter may be visible at the base. The edges are indurated, and the tag of oedematous skin known as a ‘sentinel' pile invites the clinician to look more closely. Sepsis may complicate the pathology with an intersphincteric abscess or a subsequent fistula. Despite their frequency it is not known why some fissures persist and become chronic, although, as mentioned above, the ‘ischaemic ulcer' theory is currently popular. Pain results in fear of defaecation, passage of hard stool, and further pain and sphincter spasm, thus completing a vicious cycle.

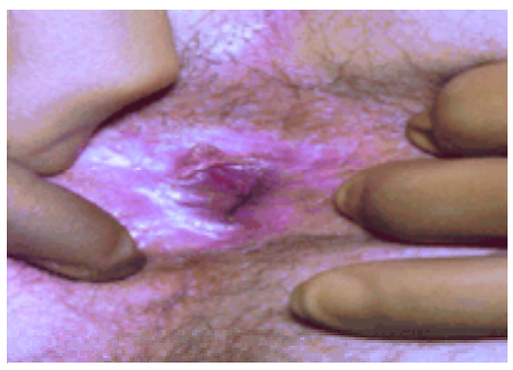


Fig. 1. Posterior idiopathic anal fissure. Not the ‘sentinel pile'.

Clinical features

Fissures can occur at any age, although they are most common between the ages of 30 and 50 years. They are also rather common in childhood as a cause of anal pain and bleeding. Fissures were previously considered to be more common in women, but in fact slightly more men are affected.

Symptoms are often out of proportion to the underlying pathology. The prominent symptom is severe anal pain, usually on defaecation and lasting between a few minutes to a few hours afterwards. Bleeding is common (80 per cent), characteristically small in quantity, bright red, and usually only occurring during straining or defaecation. In chronic cases there may be mucoid discharge with associated perianal irritation.

Diagnosis

There is usually a typical history. On examination, the patient may be in pain and a sentinel tag may be seen in chronic cases. The anal orifice is seen to be puckered, drawn in, and tightly closed, necessitating firm elevation of the buttock in order to obtain an adequate view. Traction permits slight eversion of the anus in the midline posteriorly and allows inspection of the pecten, which is where anal fissures lie. Diagnosis can usually be made on history and inspection alone. Digital palpation and proctoscopy (a paediatric sigmoidoscope is often kinder) are indicated to confirm the presence of a fissure if careful inspection has failed to reveal it. These examinations can be facilitated by prior application of local anaesthetic jelly to the fissure itself using a cotton bud or equivalent, but if there is intense anal spasm they may be impossible to perform. An examination under anaesthesia may then be necessary.

It is imperative to exclude other, more serious causes of anorectal pain and bleeding, and at some point in the course of the management of each patient a rigid sigmoidoscopy must be performed. Contrast enema or colonoscopy are indicated on an individual basis, dictated by the clinical features.

Differential diagnosis

Anal fissure must be distinguished from superficial cracks in the perianal skin seen with severe pruritus ani, but here there is no sentinel pile or sphincter spasm. Idiopathic anal stenosis is seen in elderly individuals and can be misleading. Atypical perianal ulceration may be due to syphilis, tuberculosis, human immunodeficiency virus (Fig. 2), or ulcerative proctocolitis and Crohn's disease, as well as anal carcinoma. These diagnoses must be considered and any suspicion excluded by biopsy before definitive treatment of the fissure. However, it should be emphasized that fissures associated with these conditions are rare while idiopathic anal fissure is common.

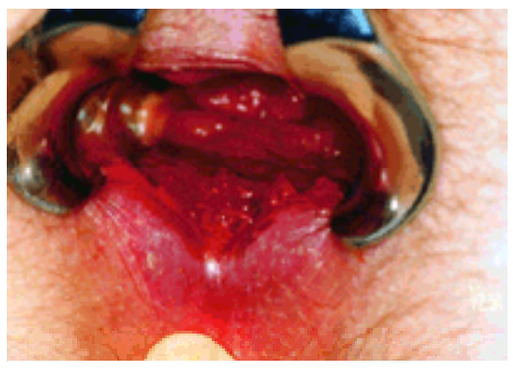


Fig. 2. Posterior anal ulcer on pecten in an HIV-positive patient. Note features not characteristic of an anal fissue: florid, poorly-defined edge, no sentinel pile.

Management

Conservative measures

Most acute fissures and up to 40 per cent of chronic fissures will heal spontaneously aided by conservative measures to relieve pain. Constipation should be treated with bulk laxatives and dietary-fibre supplements in order to avoid straining, which can cause persistent trauma and bleeding. Regular perineal bathing and hygiene are advised, and pain can be eased by application of local anaesthetic cream such as 2 per cent or 5 per cent lignocaine. Topical steroids such as 1 per cent hydrocortisone cream have been used to reduce inflammation, and healing of a significant proportion of acute anal fissures has been reported after 3 weeks' application of Proctosedyl (cinchocaine + hydrocortisone ointment), although most acute fissures heal spontaneously in any case. The use of local anaesthetic agents can occasionally lead to skin sensitization, and topical steroids should be used sparingly and for a short course only. The use of an anal dilator is no longer regarded as necessary.

Operative measures

Surgical management is indicated if the fissure fails to heal with conservative measures or when severe pain precludes them.

Manual anal dilatation

Manual dilatation of the anus achieves pain relief and variable healing rates, but it is uncontrolled and may impair anal continence permanently. Recurrence rates vary from 2 to 57 per cent, but there may be flatus incontinence or soiling of the underwear in as many as 40 per cent of patients and even frank faecal incontinence has been reported in up to 16 per cent. Anal endosonography has demonstrated that manual anal dilatation produces disruption and fragmentation of the internal sphincter. More controlled dilatation with an anal speculum or a balloon are proposed to be safer but this claim has not been substantiated. The procedure is no longer recommended.

Internal sphincterotomy

This is the surgical division of the lower third of the internal sphincter muscle, which is usually done under general anaesthesia. The earlier technique of fissurectomy and posterior sphincterotomy is now almost entirely superseded by lateral sphincterotomy, dividing the sphincter in the ‘3 o'clock' or ‘9 o'clock' position. This is the surgical procedure of choice for chronic anal fissures associated with sphincter spasm.

Method

The patient is examined under a general anaesthetic and rigid sigmoidoscopy performed. The fissure should first be pinched between finger and thumb in order to determine whether there is any pea-like induration in association with it. If there is, then it is likely there is an intersphincteric abscess, which rahter commonly opens into the fissure base. In these circumstances a dorsal sphincterotomy through the base of the fissure is still indicated in order to lay open and drain the underlying abscess. Usually there will be no induration and so a lateral sphincterotomy can be performed (Fig. 3). Inserting and opening Parks' anal retractor puts the internal sphincter on stretch. Pulling down the anal retractor whilst stretching the anal skin makes the internal sphincter very obvious. It is then a simple matter to make a small incision in the skin overlying the intersphincteric groove and to pass a pair of scissors submucosally as far as the dentate line and then intersphincterically for the same distance. The internal sphincter is now isolated. Crushing it momentarily with artery forceps before cutting it will avoid sometimes rather troublesome bleeding. The sphincter should be cut to the same level as the apex of the fissure. The Parks' anal retractor can now be rotated to supply tamponade to the operation site while any sentinel skin tag or fibroepithelial polyp is removed. It is usually unnecessary to stitch the small skin incision.

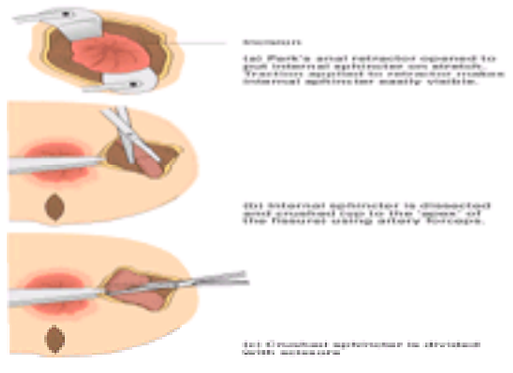


Fig. 3. Open internal sphincterotomy.

Closed and open sphincterotomy techniques are practised, both being equally effective. Healing rates in excess of 95 per cent can be expected within 4 to 6 weeks. Postoperative complications such as haematoma, bleeding, perianal abscess, and occasional urinary retention are uncommon. If an inadvertent breach of the anal mucosa occurs operatively, then the whole wound should be laid open or a superficial fistula may result.

Anal incontinence has been reported following internal anal sphincterotomy. Faecal soiling or impairment of flatus control may occur in approx. 5 per cent of patients overall, although higher rates have been reported in some series. Permanent major impairment of continence is a serious complication of internal sphincterotomy that can occur in a very small proportion of patients, and may be a result of overzealous division of the sphincter, especially in women who have a shorter anal canal. This is why it is now recommended to tailor the length of the sphincterotomy to the proximal extent of the anal fissure. One should be wary in performing a sphincterotomy in women who have had a vaginal delivery, as up to 30 per cent of them may have an occult sphincter injury. If the external sphincter is already compromised, then, given the short internal sphincter in women, division of it may result in significant faecal incontinence. It is a simple matter to take an obstetric and continence history. If there is any doubt, then an anal ultrasonographic examination will determine if there is already an occult injury to the anal sphincter, in which case sphincterotomy would be ill advised.

Despite these caveats, lateral internal sphincterotomy is a good operation. It is quick, simple, provides immediate relief of pain in uncomplicated cases of anal fissure, and results in satisfied patients—so long as they are informed of the potential problems mentioned above.

Anal advancement flaps

Most classic ‘high pressure' fissures can be treated successfully with lateral sphincterotomy. However, patients who have had surgery or trauma to the anal sphincter, or those with recurrent fissures, are more difficult to manage. Further division of the internal sphincter may impair continence. In recurrent cases after prior sphincterotomy it often pays to perform an anal ultrasound to check that the internal sphincter has indeed been divided. In such patients, healing of fissures can be achieved using an advancement flap of perianal skin after excision of the fissure. These flaps cover the fissure with perianal skin in an attempt to achieve primary healing. Broad-based V-Y or rhomboid advancement flaps have been used successfully (Fig. 4). They are indicated for the treatment of chronic fissures in the presence of normal or low resting anal-sphincter pressure, persistent fissures following obstetric delivery, and for patients in whom there is concern over sphincter function and continence. They are obviously more intricate technically than sphincterotomy and may fail, sometimes leaving rather a large wound that can take a long time to heal. They should therefore be used advisedly.

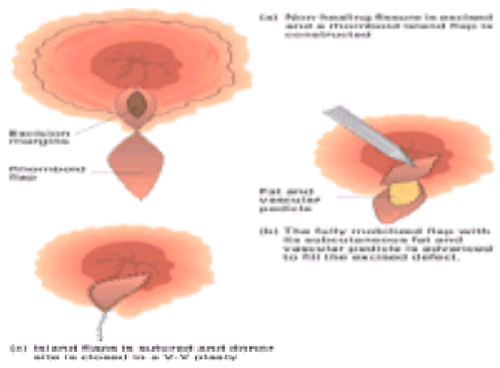


Fig. 4. V-Y (rhomboid) advancement flap for persistent fissure.

Pharmacological treatments

Nitric oxide donors

Following the recent demonstration of nitric oxide as the most important biological mediator of the rectoanal inhibitory reflex, it has been shown that topical application of a nitric oxide donor, such as glyceryl trinitrate, can lower the sphincter pressure and heal anal fissures. Over two-thirds of chronic anal fissures can be treated with an 8-week course of 0.2 per cent glyceryl trinitrate ointment [Percutol® (Dominion Pharma UK); 2 per cent glyceryl trinitrate, diluted to 0.2 per cent in yellow paraffin) applied topically three times a day. Similarly, 1 per cent isosorbide dinitrate ointment applied five to six times a day results in healing of 80 per cent of chronic anal fissures. Limitations of these novel pharmacological agents are significant headaches experienced by over 50 per cent of patients, as well as early recurrence of symptomatic fissures in over one-third of patients after initial healing. Many of the recurrent fissures can be successfully treated with a further course of topical nitrates and these agents are now considered a good first-line treatment for chronic anal fissures, avoiding the need for a surgical sphincterotomy in a significant proportion of patients.

Botulinum toxin

Purified botulinum toxin, a potent biological toxin that inhibits presynaptic release of acetylcholine from cholinergic nerve endings and causes temporary paresis of striated muscle, has recently been shown to lower resting anal-sphincter pressure when injected into the smooth muscle of the internal sphincter. In a randomized, controlled trial, injection of 20 units of this toxin into the internal anal sphincter resulted in healing of 73 per cent of chronic fissures within 8 weeks. Limitations of this treatment are its cost, occasional severe perianal thrombosis, and unknown long-term efficacy. At present a few centres have adopted its use, mainly in the research setting. However, the advent of active creams has probably now limited its appeal.

New agents

The calcium-channel blockers nifedipine and diltiazem reduce resting anal pressure when given orally. Similarly, topical application of 2 per cent diltiazem cream, and of the cholinergic agonist bethanechol 0.1 per cent cream, have recently been shown to produce a ‘chemical sphincterotomy', reducing the anal-sphincter pressure by 28 and 22 per cent, respectively. In a preliminary, open, pilot study, 2 per cent diltiazem cream resulted in healing of anal fissures in 7 out of 10 patients while 0.1 per cent bethanechol healed chronic fissures in 6 out of 10 patients. In addition, combination of these two agents in a topical cream produces a greater decrease in the resting anal-sphincter pressure than each cream individually, and this combined cream is currently under evaluation in a controlled trial for the treatment of chronic anal fissures. The advantage of these new agents over glyceryl trinitrate is the lack of any side-effects, particularly headaches, which results in high compliance.

Follow-up and prevention

Irrespective of the method of treatment, some fissures may recur, but less often after lateral sphincterotomy than pharmacological or conservative treatments. It is important that patients are given adequate advice about taking a high-fibre diet after active treatment is complete, particularly when non-operative treatments have been used, since, although it may not be possible to avoid all recurrences, certain measures may decrease their likelihood. Regular perineal hygiene should be encouraged and excessive application of topical preparations and local anaesthetic creams and steroids avoided in order to prevent perianal skin irritation, excoriation, and pruritus ani. These simple measures are often sufficient to prevent recurrence of most idiopathic anal fissures. Failure to heal and multiple recurrences should raise the suspicion of another underlying aetiological factor or of failure of compliance.

Further reading

Carapeti EA, Kamm MA, Phillips RKS. Diltiazem lowers resting anal sphincter pressure—a potential low side-effect alternative to glyceryl trinitrate for fissures. *Gut* 1998; **42** (Suppl. 1): A97.

Carapeti, EA., Kamm MA, Phillips RKS. Glyceryl trinitrate heals anal fissures, higher doses are not more effective, and there is a high recurrence rate. *Gut* 1999; **44**: 727–30.

Farouk R, Monson JR, Duthie GS. Technical failure of lateral sphincterotomy for the treatment of chronic anal fissure: a study using endoanal ultrasonography. *British Journal of Surgery* 1997; **84**: 84–5.

Loder PB, Kamm MA, Nicholls RJ, Phillips RK. ‘Reversible chemical sphincterotomy' by local application of glyceryl trinitrate. *British Journal of Surgery* 1994; **81**: 1386–9.

Lund JN, Scholefield JH. Aetiology and treatment of anal fissure. *British Journal of Surgery* 1996; **83**: 1335–44.

Lund JN, Scholefield JH. A randomised, prospective, double-blind, placebo-controlled trial of glyceryl trinitrate ointment in treatment of anal fissure. *Lancet* 1997; **349**: 11–14.

Maria G, Cassetta E, Gui D, Brisinda G, Bentivoglio AR, Albanese A. A comparison of botulinum toxin and saline for the treatment of chronic anal fissure. *New England Journal of Medicine* 1998; **338**: 217–20.

Nielsen MB, Rasmussen OO, Pedersen JF, Christiansen J. Risk of sphincter damage and anal incontinence after anal dilatation for fissure-in-ano. An endosonographic study. *Diseases of the Colon and Rectum* 1993; **36**: 677–80.

Nyam DC, Wilson RG, Stewart KJ, Farouk R, Bartolo DC. Island advancement flaps in the management of anal fissures. *British Journal of Surgery* 1995; **82**: 326–8.

Schouten WR, Briel JW, Auwerda JJ. Relationship between anal pressure and anodermal blood flow. The vascular pathogenesis of anal fissures. *Diseases of the Colon and Rectum* 1994; **37**: 664–9.

Schouten WR, Briel JW, Boerma MO, Auwerda JJ, Wilms EB, Graatsma BH. Pathophysiological aspects and clinical outcome of intra-anal application of isosorbide dinitrate in patients with chronic anal fissure. *Gut* 1996; **39**: 465–9.

Sultan AH, Kamm MA, Hudson CN, Thomas JM, Bartram CI. Anal-sphincter disruption during vaginal delivery. *New England Journal of Medicine* 1993; **329**: 1905–11. [See comments.]

Walker WA, Rothenberger DA, Goldberg SM. Morbidity of internal sphincterotomy for anal fissure and stenosis. *Diseases of the Colon and Rectum* 1985; **28**: 832–5.

28.4 Anorectal abscesses and fistula-in-ano

Julio Garcia-Aguilar and David A. Rothenberger

Surgical anatomy
Anorectal abscesses
- Etiology and pathogenesis
- Bacteriology
- Clinical examination
- Investigations and differential diagnosis
- Treatment
- Special considerations

Fistula-in-ano
- Classification
- Diagnosis
- Examination under anesthesia
- Treatment

Further reading

Most anorectal abscesses and fistula-in-ano represent the acute and chronic manifestations of the same pathologic condition, an infection that originates in the glands of the anal canal. Their diagnosis and treatment require understanding not only of the etiology and pathogenesis but also of the regional anatomy and potential routes of spread. Surgical misadventures following imprecise diagnosis and misunderstanding of the relation of the infectious process to the sphincter mechanism may lead to incomplete eradication of the infection and/or permanent impairment of anorectal function.

Surgical anatomy

The dentate line, an embryologic landmark that represents the junction of the hindgut with the proctodeum, contains eight to twelve crypts, each one draining an anal gland. The anal glands, originally described by Chiari in 1878, penetrate the submucosa of the anal canal and branch into the anal sphincter, usually reaching the intersphincteric space. They are tubular glands, lined by stratified squamous epithelium, with a direct opening into the base of an anal crypt. Parks found that two-thirds of the anal glands send branches to the internal sphincter, and one-half cross the internal sphincter completely to end in the longitudinal muscle. No branch was observed to penetrate the external sphincter.

The inner muscular layer of the anal canal, known as the internal sphincter, represents the thickened distal continuation of the circular muscular layer of the rectum (Fig. 1). As is true of the rest of the musculature of the gut, the internal sphincter is innervated by the autonomic nervous system. Outside the internal sphincter the longitudinal muscular coat of the rectum is joined by fibers of the levator ani and puborectalis muscles to form the conjoined longitudinal muscle. The internal sphincter and conjoined longitudinal muscle are surrounded by a tube of skeletal muscle with somatic innervation known as the external sphincter (Fig. 1). The proximal portion of the external sphincter is fused superiorly with the puborectalis and levator ani muscles (Fig. 1). The levator ani is a broad plate of muscle that consists of the iliococcygeus and pubococcygeus, and forms the greater part of the pelvic floor (Fig. 1). The impression of this muscle in the rectal wall, along with the upper borders of the internal and external sphincters, was called the anorectal ring by Milligan and Morgan. This anatomic landmark is very important during the treatment of abscesses and fistulas. The most distal portion of the external sphincter lies underneath the skin, slightly distal and lateral to the inferior edge of the internal sphincter. A depression palpable between these two muscles is known as the intersphincteric groove.

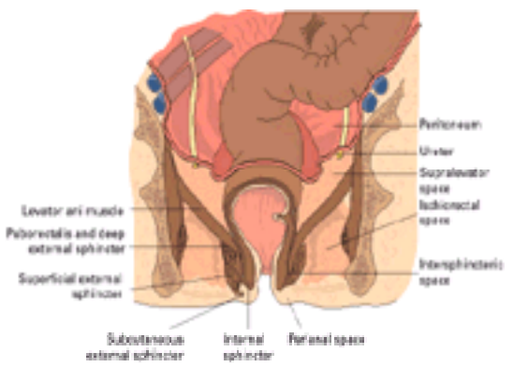


Fig. 1. Anal sphincter and associated spaces (frontal view).

The anatomic structures of the pelvis and anorectal region demarcate several potential spaces that are normally filled with fat or areolar tissue (Fig. 1). Perianal infections can spread to these spaces, producing well-defined abscesses. The intersphincteric space lies between the internal and external sphincter (Fig. 1). It contains the conjoined longitudinal muscle, which theoretically facilitates the spreading of infection to the perianal and supralevator spaces. The perianal space is the potential space surrounding the anal verge that communicates with the intersphincteric plane and is continuous with the fat of the buttocks (Fig. 1). The ischiorectal space is limited superiorly by the levator ani, medially by the anal sphincter, laterally by the obturator internus, obturator fascia, and the ischium, and inferiorly by the skin of the perineum (Fig. 1). Both ischiorectal spaces communicate with each other posteriorly behind the anal canal via the so-called postanal space. The anococcygeal ligament divides the postanal space into superficial and deep compartments. The supralevator space is situated around the rectum, above the levator ani, and is bounded superiorly by the pelvic peritoneum and laterally by the pelvic wall (Fig. 1). Both sides are connected posteriorly around the rectum.

Anorectal abscesses

Etiology and pathogenesis

Most anorectal abscesses are secondary to a suppurative process that starts in the anal glands. The cryptoglandular theory suggests that obstruction of the glandular ducts by feces, foreign material, or trauma will result in stasis and secondary infection located in the intersphincteric space (Fig. 2(a)). From this site the infectious process may extend distally along the fibers of the conjoined longitudinal muscle to emerge subcutaneously as a perianal abscess, or it may spread laterally through the longitudinal muscle and external sphincter producing an ischiorectal abscess (Fig. 2(b)). Although the vast majority of cryptoglandular abscesses are perianal or ischiorectal, other spaces can become infected. Upward extension in the intersphincteric space gives rise to a high intersphincteric abscess. It may then break through the longitudinal muscle into the supralevator space, causing a supralevator abscess (Fig. 2(b)). After the abscess is drained, either spontaneously or surgically, the abnormal communication that often persists between the anal canal and the perianal skin is called a fistula-in-ano.



Fig. 2. Pathways of infection in perianal spaces.

In addition to tracking upward, downward, and laterally, the suppurative process may spread circumferentially around the anus. This type of spread can occur at three tissue planes, the ischiorectal fossa, the intersphincteric space, and the supralevator space. This circumferential spread is known as ‘horseshoeing' (Fig. 3).

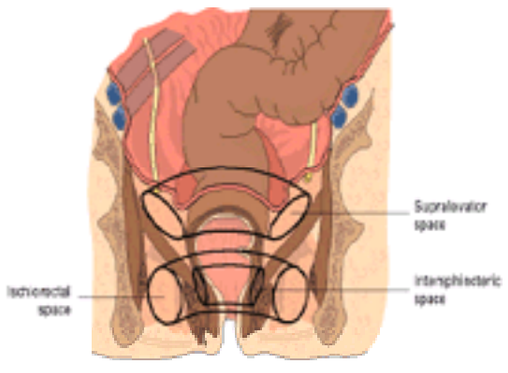


Fig. 3. ‘Horseshoe' extension of anorectal spaces.

Anorectal abscesses can also be secondary to other pathologic conditions such as inflammatory bowel disease, tuberculosis, actinomycosis, lymphogranuloma venereum, carcinoma, lymphoma, infectious complications of infection with the human immunodeficiency virus (**HIV**), trauma, surgery, and radiation.

Bacteriology

Typical cryptoglandular anorectal abscesses contain a bowel-derived flora, particularly *Escherichia coli* and *Bacteroides fragilis*, whereas perianal abscesses resulting from secondary infection of blocked apocrine glands contain a skin-derived flora, particularly *Staphylococcus aureus*. Although bacterial smears and cultures may help differentiate between these two types of abscess, culture is not routinely used in clinical practice to guide their treatment. Microbiologic investigation becomes important in immunosuppressed patients or those with atypical abscesses.

Clinical examination

Pain is the most common complaint of patients with anorectal abscesses. It is generally described as throbbing or as a dull ache. It is usually intense and may be severe. Pain described as a pressure sensation up inside the rectum that is aggravated by sitting, walking, and sudden increases of abdominal pressure produced by coughing and sneezing is highly suggestive of an intersphincteric abscess. General symptoms include malaise and fever. Partial drainage through the crypt of origin into the anal canal can result in minor purulent discharge. Sudden onset of purulent discharge followed by symptomatic improvement suggests spontaneous drainage of the abscess.

Physical examination usually reveals localized swelling, redness, and tenderness. Fever may be present. Fluctuation tends to be a late finding and may be difficult to elicit because of local tenderness. Its absence should not delay the diagnosis or treatment. With intersphincteric and supralevator abscesses, there is usually no visible swelling or induration in the perianal region. Anal and rectal examinations are often extremely painful and generally ill advised, but in a few cases an anal mass or sensation of fullness may be palpable. A tender pelvic mass diagnosed by rectal or vaginal examination, together with lower abdominal pain and tenderness, is characteristic of a supralevator abscess.

Investigations and differential diagnosis

Pain generally precludes anoscopy, proctoscopy, flexible sigmoidoscopy, or other tests. Though usually not needed, anal ultrasonography can be helpful in the diagnosis of an intersphincteric abscess, while computed tomography (**CT**) of the pelvis can assist in the diagnosis of supralevator abscesses. The differential diagnosis of perianal abscess includes hydradenitis suppurativa, pilonidal abscess, thrombosed hemorrhoid, anal fissure, abscess of the Bartholin gland, and folliculitis of the perianal skin.

Patients with significant anal pain but without abnormal perianal findings should be examined under anesthesia. Most are then found to have a condition amenable to surgical treatment, usually an intersphincteric abscess or a fissure. If a patient with a classic history suggestive of an intersphincteric abscess is examined under anesthesia and found to have no obvious abnormalities, one should strongly consider creating a small internal sphincterotomy in the posterior midline. This maneuver is often rewarded with the confirmation of an intersphincteric abscess. Neglect of an abscess will only allow it to extend into adjacent spaces and can result in life-threatening sepsis.

Treatment

At the time of diagnosis, most anal and perianal infections are in the suppurative phase and therefore unlikely to respond to antibiotic therapy alone. The treatment of anorectal abscess is immediate, adequate drainage. Antibiotics may be useful as an adjuvant to drainage in diabetic and immunocompromised patients, and in those with an endovascular prosthesis or extensive cellulitis.

The clinical distinction between perianal and ischiorectal abscesses is more academic than practical. Small and superficial perianal abscesses and moderate-sized ischiorectal abscesses can often be drained under local anesthesia in the office or emergency unit. The skin is gently prepared with an antiseptic solution. A 2-cm area of skin over the point of maximal tenderness is infiltrated with 1 per cent lidocaine (lignocaine) with 1:200 000 epinephrine (adrenaline) solution. A cruciate incision in the anesthetized area allows immediate drainage of the pus (Fig. 4). The skin edges are excised to prevent early sealing of the incision and recurrence of the abscess. Except when temporarily required for hemostasis, wounds are not packed. Tight packing occludes the incision, impeding the drainage of pus, and is unnecessarily painful.

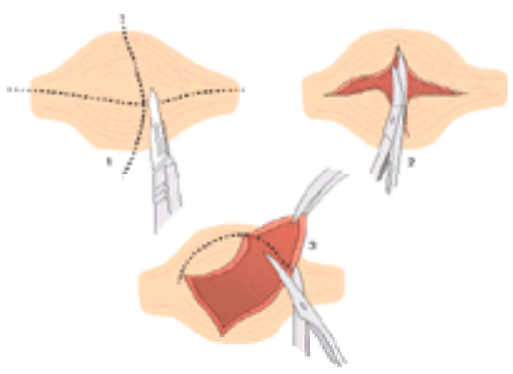


Fig. 4. Incision and drainage of perianal abscess.

Larger abscesses and those with horseshoe extensions are best drained in the operating room with the patient under regional or general anesthesia. Induration on digital examination may help localize an intersphincteric abscess. Inspection of the anal canal with a bivalve speculum may demonstrate pus coming from the anal gland responsible for the abscess. Deep, hard-to-palpate abscesses may be localized by aspiration with a needle and a syringe. As in more superficial abscesses, a cruciate incision should be made over the area of fluctuance. In general, when draining perianal and ischiorectal abscesses, the skin incision should be made as close as possible to the anus, thereby minimizing the amount of tissue to be incised in case a fistula develops after resolution of the abscess. After making the cruciate incision the skin corners should be excised and the abscess cavity irrigated. Extensive debridement with the finger or forceps is not necessary, and risks creating a false passage. A gauze dressing may be left in the wound, but extensive, tight packing is unnecessary and painful.

Alternatively, the abscess can be drained through a small stab incision using a No. 11 scalpel blade, and a 10F to 16F mushroom catheter inserted into the abscess cavity. The external portion of the catheter is shortened to leave 2 or 3 cm outside the skin, and the catheter is left in place until the drainage has ceased, the cavity has closed around it, and the inflammation has subsided. At the time of removal of the catheter, the patient is examined carefully to exclude an associated fistula. If present, a fistulotomy should be performed.

Ischiorectal abscesses with horseshoe extensions usually start as a posterior intersphincteric abscess that extends to the deep postanal space, and from there to both ischiorectal fossas. The crypt of origin is almost always in the posterior midline, and therefore the operation is usually started by inserting a crypt hook through the primary opening in the direction of the coccyx. Using the probe as a guide, the postanal space is entered by splitting the fibers of the anococcygeal ligament. The lateral ischiorectal extensions of the abscess are drained through separate para-anal incisions (Fig. 5). Anterior horseshoe abscesses with their internal opening in the anterior midline crypt are far less common but are treated similarly.

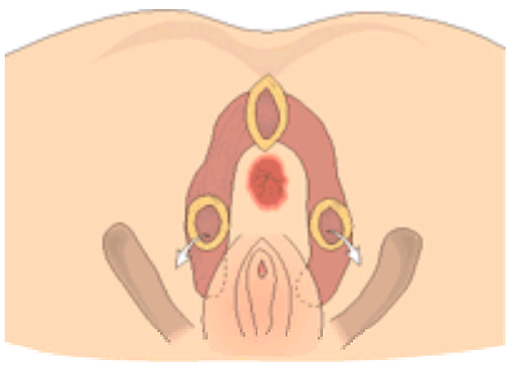


Fig. 5. Drainage of horseshoe abscess: the deep postanal space is entered, incising the anococcygeal ligament; a counter-drainage is made of each limb of the ischiorectal space.

Intersphincteric abscesses are best treated under regional or general anesthesia. Once localized they should be drained by lateral internal sphincterotomy over the area of fullness. The internal sphincter should be divided from its lower edge to the dentate line or higher if the abscess has a high extension.

The appropriate treatment of a supralevator abscess requires understanding of its etiology. Supralevator abscesses secondary to pelvic pathology such as perforated diverticulitis, Crohn's disease, or appendicitis can be drained through the rectum, through the ischiorectal fossa, or percutaneously with the help of CT scan. Permanent eradication of the abscess usually requires treatment of the abdominal process. Supralevator abscesses secondary to upward extension of high intersphincteric abscesses should be drained through the rectum, because drainage through the ischiorectal fossa will result in a suprasphincteric fistula (Fig. 6). On the other hand, supralevator abscesses secondary to upward extension of ischiorectal abscesses should be drained through the ischiorectal fossa because drainage through the rectum will result in an extrasphincteric fistula (Fig. 6). Supralevator abscesses are fortunately rare, as their origin is not always clear and their treatment can sometimes be somewhat empiric.

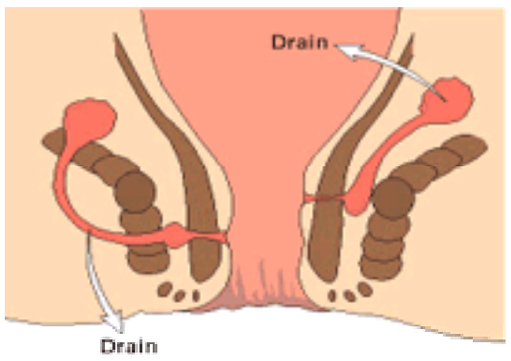


Fig. 6. Proper route of drainage in supralevator abscess.

A point of controversy in the management of perianal and ischiorectal abscesses is whether a fistulotomy should be done at the time of the initial incision and drainage. Proponents of primary fistulotomy argue that many abscesses treated by incision and drainage alone will result in the development of a fistula. They feel that the originating crypt can be identified easily during the acute phase, and that a fistulotomy done at the time of draining the abscess will eliminate the risks of developing a fistula. Opponents to primary fistulotomy believe that a significant percentage of anorectal abscesses heal without a fistula. They find routine primary fistulotomy unnecessary, and even dangerous, because overzealous attempts to find the fistulous tract in infected, friable tissue may result in the creation of a false passage, thus increasing the risks of persistent sepsis and anal incontinence. Several prospective, randomized trials report a 0 to 2.9 per cent incidence of fistula-in-ano after primary fistulotomy and a 25 to 49 per cent incidence after incision and drainage alone; these data can be used to support both positions. We do not attempt primary fistulotomy when draining abscesses under local anesthesia in the office or emergency room. We perform primary fistulotomy in patients with a fistulous abscess who require drainage under regional or general anesthesia when there is an unequivocal internal opening at the dentate line and an easy-to-identify fistulous tract. Because most abscesses can be drained under local anesthesia, primary fistulotomy has a small part to play in our practice.

Special considerations

Neglected perianal infections, particularly in diabetic and immunocompromised patients, can spread along fascial planes causing synergistic, necrotizing, perineal infections. Spread along the Colles' fascia produces a severe necrotizing infection of the perineum and scrotum known as Fournier's gangrene. Extensive edema or cellulitis, systemic symptoms of sepsis, and skin necrosis suggest a necrotizing infection. Treatment requires broad-spectrum antibiotics, wide debridement of all nonviable tissue, tetanus prophylaxis, and nutritional support. Suprapubic cystostomy may be necessary in patients with urethral disruption, while colostomy may be required in patients with necrotizing infections causing destruction of the anal sphincters.

Perianal infections are common in patients with hematologic malignancy receiving chemotherapy and usually present with severe anal pain. The risk of infection is related to the degree of the granulocytopenia, and therefore symptoms of systemic involvement predominate. Suppuration with local tenderness, swelling, and erythema develop only after the white blood-cell count increases. Fluctuance is a late and uncommon finding. The more commonly isolated micro-organisms are *Bacteroides, E. coli, Klebsiella, Proteus*, and *Streptococcus faecalis. Pseudomonas aeruginosa* seems to be more common in children. The difficulty in distinguishing anorectal abscesses from perianal leukemic infiltrates and the concern that surgical instrumentation of the anorectal region may result in spread of the infection and impaired wound healing are responsible for the controversy about surgical management of these lesions. Most of these patients can be treated with precautionary measures such as sitz baths, stool softeners, analgesics, and antibiotics, while avoiding digital rectal examinations, suppositories, or enemas. As the white count returns to normal, most areas of inflammation and infection resolve. Each case should be treated individually, depending on the underlying malignant disease, the patient's general condition, and the degree of immunosuppression. Appropriate surgical intervention can relieve the pain, and prevent local spread of the infection and the development of septicemia. Antibiotics, either a third-generation cephalosporin plus an antianaerobic agent, or the combination of an extended-spectrum penicillin,

an aminoglycoside, and an antianerobic agent, should be used in these patients. If a necrotizing infection is suspected, immediate examination under anesthesia and aggressive debridement will be necessary. Patients with associated thrombocytopenia usually require platelet transfusion at the time of debridement.

The chronic depression of CD4+ lymphocyte counts and the alterations in the natural barriers to infection, such as the loss of integrity of the skin and rectal mucosa and granulocytopenia, predispose patients infected with HIV to anorectal infections. However, the life-threatening infections described at the beginning of the HIV epidemic, when the disease had a fulminant course that was invariably lethal, are exceptional today. The change in the manifestations of the disease caused by the introduction of the new antiretroviral therapies has also decreased the incidence and severity of the anorectal complications. However, anorectal abscesses in patients with HIV infection have some unique features. Although most abscesses have a cryptoglandular origin, some may be secondary to anal ulcers that erode the internal sphincter. Some abscesses may contain opportunistic bacteria such as *Pseudomonas* or mycobacteria. The frequent use of antibiotics in these patients may mask the symptoms of anorectal sepsis and delay the diagnosis. Many patients have associated pathology, particularly sexually transmitted diseases. The differential diagnosis of anorectal abscess is larger in patients with HIV infection than in the general population, because entities such as Kaposi sarcoma, perianal lymphoma, and even squamous-cell carcinoma can be confused with anorectal abscesses. Wound healing is normal in asymptomatic patients with normal CD4+ counts; it may be impaired in patients with symptomatic opportunistic infections or a CD4+ count lower than 200 cells/mm^3. With these considerations in mind, the surgical management of anorectal abscesses in patients with HIV infection should generally follow the same principles observed in the general population. In order to reduce the risk of cross-contamination, anorectal abscesses in patients known to have HIV should always be drained in the operating room under regional or general anesthesia. Prophylactic antibiotics should be used systematically. The presence of suppuration should be confirmed by needle aspiration of the indurated area, and a sample of pus sent for culture. The abscess should be drained through an adequate incision, avoiding unnecessary soft-tissue trauma. A sample of the abscess wall should be sent for pathologic examination.

Fistula-in-ano

A fistula-in-ano is an unnatural connection between the epithelium of the anal canal and the epidermis of the perianal skin.

The cryptoglandular hypothesis, as outlined above, postulates that a fistula-in-ano represents the late sequela of a drained anorectal abscess originating in the anal glands. Fifty per cent of anorectal abscesses will heal after surgical or spontaneous drainage. In the other half, a chronic inflammatory process will persist in the track connecting the anal gland with the perianal skin. Spreading of the intersphincteric abscess to the different anorectal spaces produces different types of abscesses (Fig. 3(b)). Drainage of each one of these abscesses may lead to the development of different types of fistula (Fig. 7). Anorectal abscesses with other etiologic backgrounds, such as Crohn's disease, can also result in the development of anal fistulas. Superficial fistulas observed in patients with anal fissure do not have a cryptoglandular cause. Their internal opening is usually located in the base of the fissure in the posterior midline and does not involve the internal sphincter.

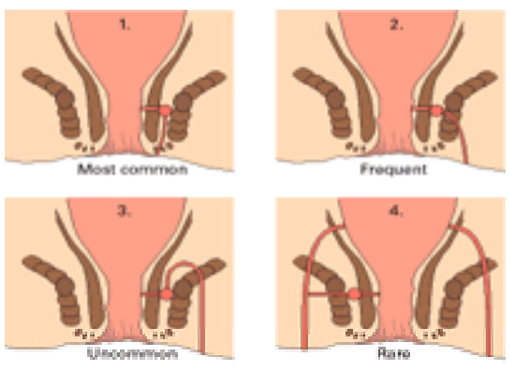


Fig. 7. The four main anatomic types of fistula: (1) intersphincteric,(2) trans-sphincteric, (3) suprasphincteric, and (4) extrasphincteric.

Classification

The surgical treatment of fistula-in-ano requires an understanding of the relation of the fistulous track to the anal sphincters. Fistulas are classified into four major groups.

Intersphincteric

This fistula represents the sequela of a drained perianal abscess with the track passing from the primary abscess in the intersphincteric plane straight downwards to the anal verge (Fig. 7). The intersphincteric fistula may have a high blind track extending upwards in the intersphincteric plane all the way to the supralevator space.

Trans-sphincteric

This type of fistula, the second most common variety, is the result of an ischiorectal abscess. The fistulous track passes from the intersphincteric plane, through the external sphincter, into the ischiorectal fossa, and then to the skin (Fig. 7). The level at which the tract passes across the external sphincter varies, and a trans-sphincteric fistula may involve almost the entire external sphincter or only its most superficial portion. Trans-sphincteric fistulas can also have a high blind tract that may reach the apex of the ischiorectal fossa.

Suprasphincteric

This uncommon type of fistula results from a supralevator abscess that has drained downwards through the levator ani muscle into the ischiorectal fossa and finally through the skin (Fig. 7). The fistula tracks upwards in the intersphincteric plane to a point above the puborectalis muscle, then laterally over this muscle, and finally downwards through the levator ani muscle into the ischiorectal fossa.

Extrasphincteric

In this type of fistula, the least common variety, the tract lies outside of the sphincter complex. The track passes from the perineal skin, to the ischiorectal fossa, through the levator ani muscles, and finally into the rectum (Fig. 8). Extrasphincteric fistulas of cryptoglandular origin are often the result of iatrogenic opening into the rectum by vigorous probing while attempting to find the internal opening of a trans-sphincteric fistula. Extrasphincteric fistulas may also be secondary to penetrating perineal injuries, inflammatory bowel disease, cancer, or pelvic abscesses drained through the perineum.

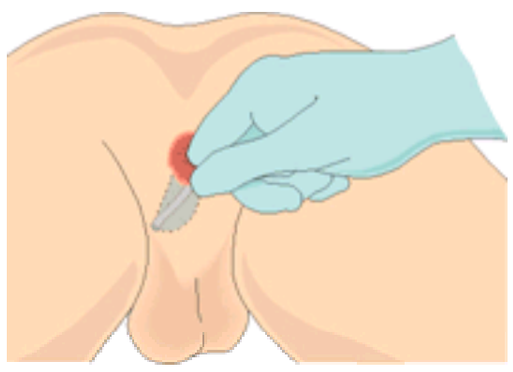


Fig. 8. Palpation of fistulous track.

As with anorectal abscesses, anal fistulas can extend in a horseshoe fashion in the different perirectal spaces.

Diagnosis

Most patients with anal fistula have a previous history of an anorectal abscess that drained spontaneously or was surgically drained, and complain of intermittent purulent drainage, a sensation of wetness, and pruritus ani. Recurrent abscesses in the same site are also a frequent manifestation of fistula-in-ano.

Inspection of the perianal area, anal digital examination, anoscopy, and sigmoidoscopy are mandatory in every patient with a history suggestive of fistula-in-ano. The external opening or openings are usually found at the bottom of a depressed area or surrounded by granulation tissue. It may heal temporarily, leaving only an indurated scar corresponding to the site of previous drainage. An indurated cord extending from the external opening towards the anal canal can sometimes be palpated in patients with intersphincteric fistula. Simultaneous digital palpation of the anal canal and the external fistulous opening can provide valuable information about the direction and depth of the fistulous track (Fig. 8). The internal opening can be felt as an area of induration or visualized as a dimple at the level of the dentate line. Probing of the fistula in the office, even when performed gently, is uncomfortable for the patient and carries the risk of creating a false passage. Except in very superficial fistulas, the chances of probing the entire tract and finding the internal opening are very low.

Given the limitations of the preoperative clinical evaluation, several diagnostic procedures have been used to assist the surgeon in defining the anatomy of the fistula. These tests are unnecessary in the initial evaluation of most anal fistulas, and are only indicated in complex or recurrent fistula.

CT is useful in identifying deep anorectal abscesses, but is rarely used in the preoperative evaluation of fistula-in-ano. The CT scan has relatively poor soft-tissue resolution, and the levator muscles and the sphincter musculature are not well identified on axial sections.

Fistulography, like the CT scan, is not helpful in defining the relation of the fistulous tracts to the normal anatomic structures. The presence of granulation tissue and purulent material in the fistulous tracts often obstructs the flow of contrast to some fistula extensions or creates a false image.

Endoanal ultrasonography provides excellent images of the anal region and is very accurate at identifying fluid collections and the fistulous tract. However, it can rarely identify the internal fistulous opening and, even after enhancement with hydrogen peroxide, it frequently misses secondary tracts.

Magnetic resonance imaging has good tissue resolution and multiplanar capabilities, and is very accurate at identifying the internal opening and the fistulous tracts. Although it seems to be superior to endoanal ultrasonography in the evaluation of fistula-in-ano, ultrasound is cheaper and can be used in the operating room at the time of surgery.

Preoperative documentation of the status of continence is important in all patients undergoing surgery for fistula-in-ano, because significant functional defects in the sphincter mechanism may influence the selection of the surgical procedure.

Examination under anesthesia

In most patients, the anatomy of the fistula and its relation to the sphincter complex can only be defined in the operating room. We generally perform fistula operations under general or regional anesthesia, with the patient in the prone jack-knife position. The need for good illumination, ideally with a headlight, cannot be overemphasized. The buttocks are spread apart with adhesive tape. Infiltration of the perianal area with local anesthesia may cause edema and hematomas that may distort the anatomy of the fistula. Therefore, we defer injection of local anesthetic solution until the end of the procedure.

The perianal area should be inspected, searching for the external opening(s), and palpated to feel the induration of the fistulous tracts. The anal canal should be inspected visually with the help of a bivalve speculum. The internal opening, often marked by an area of induration or a dimple at the level of the dentate line, can occasionally be localized by a drop of pus expressed out of the crypt by pressing on the perianal skin over the suspected fistulous track.

The location of the external fistulous opening in relation to the transverse anal line has been used to predict the direction of the track and the location of the internal opening. Anal fistulas with an external opening posterior to the transverse anal line have a curved track towards an internal opening located in the posterior midline. Anterior fistulas are supposed to have a straight tract toward an internal opening located in the same radial line as the external one. However, the internal opening of most anterior anal fistulas is also located in the anterior midline.

The fistulous track should be gently probed, starting from the external opening and advancing the probe along the fistulous tract until the internal opening at the dentate line is reached, or until it cannot be easily advanced. At this point the injection of a diluted solution of hydrogen peroxide through the external opening may help localize the internal opening by the small bubbles emerging at the dentate line. Alternatively, the tract can be opened over the probe by ‘following the granulation tissue', starting at the external opening until a change in direction of the fistulous tract allows advancing the probe to the internal opening.

Horseshoe fistulas should be treated by probing the primary trans-sphincteric tract between the posterior midline crypt and the postanal space. Management of this primary tract should follow the same principles as for any trans-sphincteric fistula. Any secondary tract can then be treated by curettage or the ‘lay-open' technique (see below).

When the internal opening cannot be clearly identified despite all these maneuvers, the correct approach is to stop the operation without dividing any muscle, and then either study the patient with magnetic resonance imaging and endoanal ultrasonography to define the anatomy of the fistula or refer the patient to a surgeon with more experience in the management of anorectal fistula.

Treatment

The difficulty of the treatment of fistula-in-ano is reflected by the diversity of techniques developed over the years (Table 1). ‘Lay open', with division of all the tissues distal to the primary tract, is the most effective way of eliminating the fistula. However, this effectiveness should be balanced against the risk of disturbing anal continence. The dogmatic statement that sphincter muscle can be divided with impunity provided that the puborectalis muscle is preserved has probably been responsible for many cases of iatrogenic incontinence. The lay-open technique carries a risk of incontinence that is proportional to the amount of sphincter muscle involved by the fistula. Procedures that attempt to obliterate the tract and close the internal fistulous opening are less likely to cause anal incontinence, but have a higher recurrence rate. Therefore, the single most important criterion in the selection of the surgical treatment of fistula-in-ano is the relation of the fistulous tract to the sphincter complex. The prediction of the functional consequences of dividing the portion of the sphincter complex distal to the fistula is difficult and requires expertise, consideration of the patient's existing continence, and knowledge of the anatomy of the fistula.

I.	Operations that divide the sphincter complex:
	Fistulotomy and marsupialization
	Two-stage seton fistulotomy
	Cutting seton
II.	Core out of fistula and occlusion of the internal opening:
	Endorectal advancement flap
	The intersphincteric approach
III.	Other techniques:
	Draining seton
	Fistulotomy and sphincter reconstruction
	Fibrin glue
	Island-flap anoplasty

Table 1 Alternatives in the surgical treatment of fistula-in-ano

Intersphincteric and low trans-sphincteric fistulas are effectively treated by fistulotomy and marsupialization (Fig. 9). Once the probe is passed into the internal opening, the tissue encircled by the track is divided over a grooved probe director. The granulation tissue is curetted and sent for biopsy if circumstances suggest an atypical

cause. Hemostasis is obtained with electrocautery. Any secondary tract should also be opened. To prevent immediate postoperative bleeding and to shorten the healing time the skin edges are marsupialized to the edge of the track with a running absorbable suture. A loose cotton dressing may be applied over the open area.

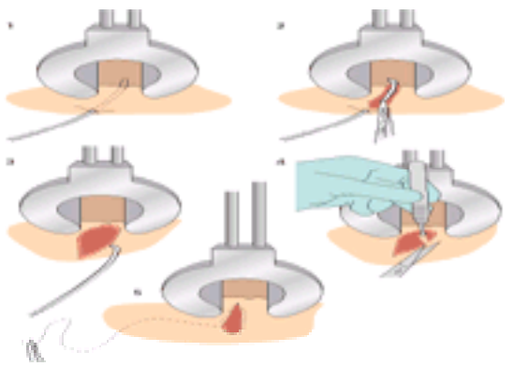


Fig. 9. Fistulotomy: (1) the probe is in the fistulous track; (2) unroofing of the fistulous track over the probe; (3) laid open fistulous track; (4) ‘redundant' skin is excised; (5) the wound is marsupialized.

The lay-open technique eliminates the fistula in more than 90 per cent of the patients. It causes minor disturbances of continence in a significant number, but most of them are satisfied with the results of the procedure. However, fistulas that involve more than one-third of the sphincter complex, anterior fistulas, particularly in women, and fistulas in patients with compromised continence, should not be treated by fistulotomy.

The use of setons is based on the assumption that the chronic inflammatory reaction and fibrosis induced by the seton around the fistulous tract prevent sphincter retraction once the portion of the muscle encircled by the fistula has been divided. Before placing a seton the portion of the track outside the sphincter mechanism and any lateral secondary tracks are laid open (Fig. 10). The granulation tissue is removed from the fistulous track with the help of a curette. The anoderm and perianal skin overlying the portion of the sphincter complex to be encircled by the seton are incised, and the intersphincteric space is drained by division of the internal sphincter. A high blind track, when present, is managed by proximal extension of the internal sphincterotomy. The edges of the anoderm and skin are marsupialized with a 3–0 catgut. Two types of setons have been used in the surgical management of high anal fistula. Most often, a large nonabsorbable suture is loosely tied around the remaining portion of the sphincter complex. The seton is removed 6 to 8 weeks later at the time of performing the secondary fistulotomy. Alternatively, a rubber band is passed through the fistulous track and tied loosely around the sphincter mechanism. The seton is tightened every 2 weeks until it cuts through the muscle. Both seton techniques are effective at eradicating the fistula, but more than 50 per cent of patients with high anal fistulas are left with symptoms of incontinence.



Fig. 10. Placement of seton and marsupialization of the wound.

We believe that the fibrosis that is assumed to prevent the sphincter retraction and the development of major anal incontinence creates a deformity in the anal canal, which may be responsible for the anal seepage described by many of these patients. Therefore we no longer recommend the traditional use of a cutting seton if the fistula involves half or more of the external sphincter. Instead, a Silastic seton is left in place for several months until the secondary extensions have healed. The seton is then removed without dividing the muscle. With this technique we try to convert a complex fistula with multiple tracts into a straight, simple, although still high, tract. In different studies, healing rates of 44 and 78 per cent are reported with this approach. For those that persist, the intermittent drainage secondary to this tract may be better tolerated than the anal incontinence that often follows division of a major part of the sphincter complex. The patients can still undergo more complex procedures if the drainage of a persistent fistula is not well tolerated or new abscesses develop.

Several surgical procedures have been developed over the years to treat the fistula by closing the internal opening and eliminating the tract without dividing the anal sphincter (Table 1). Excision (core-out) of the fistulous tract and closure of the internal fistulous opening with an endorectal advancement flap (Fig. 11), or closure of both internal and external sphincters exposed by an intersphincteric approach (Fig. 12), have been used successfully for the surgical management of high anal fistulas. Recurrence rates with these techniques are higher than with the lay-open techniques, but continence is less disturbed.

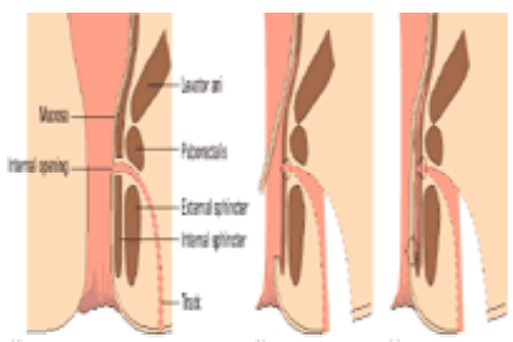


Fig. 11. Endorectal advancement flap: (a) the fistulous tract is completely excised; (b) a flap including mucosa, submucosa, and part of the internal sphincter is advanced over the internal opening already closed with interrupted reabsorbable stitches: (c) the advancement flap is sutured in place.

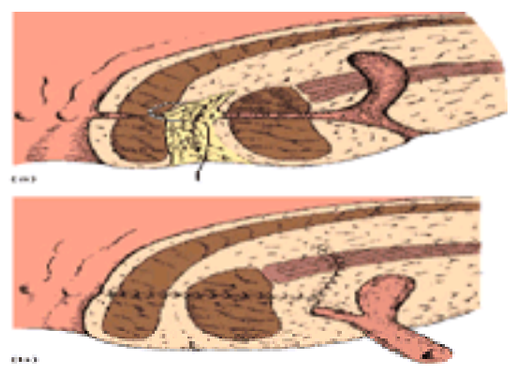


Fig. 12. Intersphincteric approach for the surgical treatment of fistula-in-ano: (a) the internal opening is closed from the intersphincteric spaces; (b) the tract is completely excised and the local defect in the external sphincter primarily closed.

Other surgical alternatives such as fistulotomy with immediate sphincter reconstruction, occluding the tract with fibrin glue, and island-flap anoplasty have shown promising results, but the experience is still preliminary.

Further reading

Abel ME, Chiu YSY, Russell TR, Volpe PA. Autologous fibrin glue in the treatment of rectovaginal and complex fistulas. *Diseases of the Colon and Rectum* 1993; **36**: 447–9. [Autologous fibrin glue may be a viable alternative for the treatment of complex fistula.]

Barker PG, Lunniss PJ, Armstrong P, Reznek RH, Cottam K, Phillips RK. Magnetic resonance imaging of fistula-in-ano: technique, interpretation and accuracy. *Clinical Radiology* 1994; **49**: 7–13. [MRI is an accurate imaging method in the assessment of anal fistula.]

Christiansen J, Ronholt C. Treatment of recurrent high anal fistula by total excision and primary sphincter reconstruction. *International Journal of Colorectal Diseases* 1995; **10**: 207–9. [Excision of the fistula is based on the principles of the trans-sphincteric approach.]

Corfitsen MT, Hansen CP, Christensen TH, Kaae HH. Anorectal abscesses in immunosuppressed patients. *European Journal of Surgery* 1992; **158**: 51–3. [The authors propose an interventionist approach to perirectal abscess in immunocompromised patients, but minimize the extent of the operation and use antibiotics.]

Deen KI, Williams JG, Hutchinson R, Keighley MRB, Kumar D. Fistulas in ano: endoanal ultrasonographic assessment assists decision making for surgery. *Gut* 1994; **35**: 391–4. [Anal endosonography is accurate in identifying fistulous tracks, fluid collections, and horseshoe extensions, but not the internal openings.]

Del Pino A, Nelson RL, Pearl RK, Abcarian H. Island flap anoplasty for treatment of transsphincteric fistula-in-ano. *Diseases of the Colon and Rectum* 1996; **39**: 224–6. [Closure of the internal fistulous opening by using an advancement flap of perianal skin.]

Eisenhammer S. A new approach to the anorectal fistulous abscess based on the high intermuscular lesion. *Surgery, Gynecology and Obstetrics* 1958; **106**: 595–9. [A classic article ascribing the origin of anorectal abscesses to an infection in the intersphincteric space.]

Garcia-Aguilar J, Belmonte C, Wong WD, Goldberg SM, Madoff RD. Anal fistula surgery. Factors associated with recurrence and incontinence. *Diseases of the Colon and Rectum* 1996; **39**: 723–9. [Recurrence is related to the complexity of the fistula and the experience of the surgeon. Incontinence after fistulotomy is proportional to the amount of muscle divided.]

Garcia-Aguilar J, Belmonte C, Wong WD, Goldberg SM, Madoff RD. Cutting seton versus two-stage seton fistulotomy in the surgical management of high anal fistula. *British Journal of Surgery* 1998; **85**: 243–5. The cutting seton and the two-stage seton fistulotomies are successful in treating the fistula but both techniques cause significant incontinence.]

Golub RW, Wise WE, Jr, Kerner BA, Khanduju KS, Aguilar PS. Endorectal mucosal advancement flap: the preferred method for complex cryptoglandular fistula-in-ano. *Journal of Gastrointestinal Surgery* 1997; **1**: 487–91. [Endorectal advancement flap heals most fistulas and preserves continence.]

Grace RH, Harper IA, Thompson RT. Anorectal sepsis: microbiology in relation to fistula in ano. *British Journal of Surgery* 1982; **69**: 401–3. [Anorectal abscess growing bowel-derived flora are associated with fistula-in-ano. Those growing skin-derived flora are not.]

Guillaumin E, Jeffrey RB, Shea WJ, Asling CW, Goldberg HI. Perirectal inflammatory diseases: CT findings. *Radiology* 1986; **161**: 153–7. [Excellent review of the pararectal spaces in CT scanning.]

Ho Y-H, Tan M, Chui S-H, Leong A, Eu K-W, Seow-Choen F. Randomized controlled trial of primary fistulotomy with drainage alone for perianal abscesses. *Diseases of the Colon and Rectum* 1997; **40**: 1435–8. [Only 25 per cent of patients benefit from primary fistulotomy.]

Hussain SM, Stoker J, Schouten WR, Hop WCJ, Lameris JS. Fistula-in-ano: endoanal sonography versus endoanal MR imaging in classification. *Radiology* 1996; **200**: 475–81. [Endoanal MR imaging is more accurate than endoanal ultrasound in depicting the anatomy of fistula-in-ano.]

Kennedy HL, Zegarra JP. Fistulotomy without external sphincter division for high anal fistulae. *British Journal of Surgery* 1990; **77**: 898–901. [Seventy-eight per cent primary healing with minimal incontinence in a high anal fistula treated by fistulotomy without external sphincter division.]

Kuijpers HC, Schulpen T. Fistulography for fistula-in-ano. Is it useful? *Diseases of the Colon and Rectum* 1985; **28**: 103–4. [Fistulography is not accurate in defining the anatomy of the fistula.]

Law PJ, Talbot RW, Bartram CI, Northover JMA. Anal endosonography in the evaluation of perianal sepsis and fistual in ano. *British Journal of Surgery* 1989; **76**: 752–5. [Initial description of the use of ultrasound in the evaluation of patients with fistula-in-ano.]

Matos D, Lunniss PJ, Phillips RKS. Total sphincter conservation in high fistula in ano: results of a new approach. *British Journal of Surgery* 1993; **80**: 802–4. [The authors describe their intersphincteric approach to the fistulous track and provide their results in 13 patients.]

Ozuner G, Hull TL, Cartmill J, Fazio VW. Long-term analysis of the use of transanal rectal advancement flaps for complicated anorectal/vaginal fistulas. *Diseases of the Colon and Rectum* 1996; **39**: 10–14. [Fistula recurrence in one-third of patients with endoanal advancement flap.]

Parks AG, Gordon PH, Hardcastle JD. A classification of fistula-in-ano. *British Journal of Surgery* 1976; **63**: 1–12. [Classic description of the different types of anal fistulas.]

Schouten WR, van Vroonhoven TJMV. Treatment of anorectal abscess with or without primary fistulectomy: results of a randomized trial. *Diseases of the Colon and Rectum* 1991; **34**: 60–3. [More than half of the patients will have no further problems after simple drainage of perirectal abscess.]

Seow-Choen F, Ho JMS. Histoanatomy of anal glands. *Diseases of the Colon and Rectum* 1994; **37**: 1215–18. [Anal glands are fairly evenly distributed in the anal circumference at the dentate line.]

Thompson JPS, Ross AH. Can the external anal sphincter be preserved in the treatment of transsphincteric fistula-in-ano. *International Journal of Colorectal Diseases* 1989; **4**: 247–50. [Forty-four per cent primary healing of a complex trans-sphincteric fistula with a draining seton without fistulotomy.]

Williams JG, MacLeod GA, Rotherberger DA, Goldberg SM. Seton treatment of high anal fistulae. *British Journal of Surgery* 1991; **78**: 1159–61. [Successful treatment of anal fistula with the seton technique.]

28.5 Faecal incontinence

B. James Mander and Norman S. Williams

Introduction
Anatomy and physiology of normal continence
Aetiology of incontinence
Clinical features
History
Examination
Investigations
Treatment
Conservative measures
Medical therapy
Surgical options
Conclusions
Further reading

INTRODUCTION

Faecal incontinence may be defined as the inability to defer the passage of faeces or flatus until socially acceptable. It may be classified as minor (the occasional passage of flatus or soiling of underwear) or major, although its severity depends on its effect on any given patient. Incontinence is a stigmatizing symptom that sufferers often conceal, which results in social isolation and debilitating psychological sequelae. Reliable epidemiological data are not available, but faecal incontinence is estimated to affect 0.1 per cent of the general public, though the prevalence rises to more than 1 per cent in the over-65 age group and may be nearer 60 per cent in long-stay institutions.

Normal continence utilizes a feedback loop consisting of the reservoir capacity of the rectum, sensory receptors in the colon, anal canal and pelvic floor, an integrator in the brain and spinal cord, and the muscular end-organs, the sphincters. Disruption of any part of this system may result in incontinence. Even if anorectal function is normal, it may be overwhelmed by the presence of large quantities of liquid stool such as occurs in infective gastroenteritis or inflammatory bowel disease. Treatment in these cases is directed at the underlying cause. In contrast, the coloproctologist is usually involved in treating incontinence arising from the congenital absence, surgical removal, or pathological damage of the anorectum or its neurological pathways. The cost of incontinence is substantial both in terms of social isolation and in its treatment (incontinence supplies; nursing and medical care).

Anatomy and physiology of normal continence

The anorectum develops from the fusion of the hindgut with the internal cloaca at around 4 weeks of gestation. Over the following 2 weeks, mesoderm aggregates on the surface of the perineum to form anal tubercles. These fuse to produce a patent anal orifice and sphincter complex by week 8 of intrauterine development. Development of the anorectum parallels the development of the urogenital system. Abnormalities of fusion result in a variety of congenital fistulas and atresias. Development of the neuromuscular structures essential to continence mirrors the formation of the rectum and anus.

The preservation of continence requires specialized sensory receptors integrated via an intact nervous system with a functioning sphincter complex. Sensory input arises from proprioceptive nerve endings in the colon and rectum, and from the densely innervated modified endoderm of the anal canal. The sphincter complex comprises the smooth muscle of the internal anal sphincter and the striated musculature of the levator ani and external sphincter. The latter is traditionally divided into subcutaneous, superficial, and deep components. The smooth muscle of the colon, anorectum, and internal sphincter responds to local factors and to reflexes mediated by the autonomic nervous system. The internal sphincter is innervated by the autonomic nervous system, with sympathetic stimulation being excitatory and parasympathetic causing sphincter relaxation. The striated muscle of the voluntary sphincter apparatus is under the control of spinal and cerebral centres via somatic efferent and afferent connections. The striated muscle is innervated by the pudendal nerve (S2, 3) and by the perineal branch of S4.

Control of continence is the normal state and develops during infancy. The process is largely unconscious, but it is subject to modification by local reflexes and the conscious will. Maintenance of continence requires the interaction of a number of factors. These include stool consistency and volume, colonic transit, rectal compliance, anorectal sensation, anal sphincter function, anorectal reflexes, the central nervous system, and intact neurological pathways.

The physiology of normal continence is not well understood, partly because any physiological studies are biased by significant measurement artefact. Numerous theories abound and there is little consensus. The importance of the sphincter muscles and the high-pressure zone they produce is not disputed. If the pressure within the anal canal is maintained above that in the rectum, then continence will not be threatened. At rest the rectum is normally empty of stool but the sphincters are tonically activated, with the internal smooth-muscle sphincter contributing 80 per cent of resting tone. At times of increased intrarectal pressure, such as with coughing or straining, when continence is threatened the external sphincter contracts, mediated by spinal reflexes. The rectum distends as faeces or flatus are introduced by the peristaltic activity of the sigmoid colon, producing a conscious awareness of rectal filling. The site of the specialized sensory receptors has not been fully elucidated but is likely to be within the rectal wall. Rectal distension characteristically provokes relaxation of the internal anal sphincter (the rectoanal inhibitory reflex). This opens the upper part of the anal canal and allows the contents of the rectum to come into contact with the richly innervated sensory epithelium of the upper anal canal. This is known as the ‘sampling' reflex and is postulated as the mechanism that allows for accurate discrimination between faeces and flatus. As the internal sphincter relaxes the external sphincter reflexly contracts to preserve continence. The external sphincter is capable of generating considerable closing pressures but only for relatively short periods of time. When evacuation is socially acceptable, then the voluntary sphincters may be inhibited by the central nervous system, allowing the anal canal to open. Rectal pressure rises by either a mass contraction of the rectal musculature or by raising intra-abdominal pressure (straining). When it exceeds anal-canal pressure, rectal contents are extruded (defaecation).

Aetiology of incontinence

The causes of incontinence are outlined in Table 1.



Table 1 Causes of incontinence

The incidence of congenital anorectal anomalies is approximately 1:5000 live births. They are classified as high, intermediate or low, depending on the position of the agenesis relative to the pelvic floor. There is a slight preponderance of males affected, particularly with the debilitating high lesions. The vast majority are associated with developmental abnormalities of the urogenital system, with fistula formation occurring frequently. Incontinence may also arise directly as a result of other congenital conditions such as spina bifida, or secondary to surgical correction of an underlying defect such as Hirschsprung disease.

Any condition that causes persistent or profuse diarrhoea may overwhelm the continence mechanism even if the sphincter complex is normal. The situation may be

exacerbated by the loss of the normal rectal reservoir capacity either pathologically or iatrogenically. Numerically, the most common cause of incontinence is faecal impaction, and this is especially common in the long-term institutionalized and geriatric populations. Here, incontinence arises from the failure of the external sphincter to respond to relaxation of the internal sphincter. A combination of diminished anorectal sensation and persistent relaxation of the internal sphincter produced by rectal distension from the faecal bolus are implicated. Frequently the situation is exacerbated by the prescription of aperients. Faecal leakage associated with a full-thickness rectal prolapse arises from a similar mechanism.

Disruption at any level of the nervous system can lead to incontinence. Characteristically, patients with an upper motor-neurone deficit from high spinal lesions have better function than those with a lower defect. The most severe problems are experienced by those with a mixed defect. Damage to the somatic afferent and efferent neurones between the anorectum and the spinal cord results in a lower motor-neurone lesion. Incontinence associated with diabetes is multifactorial, related primarily to autonomic neuropathy but exacerbated by diminished rectal sensation and frequently by diarrhoea. ‘Idiopathic faecal incontinence' is the most common cause of primary incontinence in middle age and arises from pudendal neuropathy. The pudendal nerve is particularly vulnerable to neuropraxia-type injuries because it is tethered at Alcock's canal. Factors associated with denervation include straining at stool, excessive perineal descent, prolonged second stage of labour, the use of forceps, and the delivery of high birthweight infants.

Traumatic damage to the anal sphincter complex is most commonly associated with obstetric tears or inadvertent surgical division. One or both sphincters will be torn in approx. 2 per cent of vaginal deliveries. Classically the defect is situated anteriorly and is frequently undiagnosed at the time of delivery. Forceps delivery, primiparous delivery, long duration of second stage of labour, birthweight of more than 4 kg, and occipitoposterior presentation are all associated risk factors. The damage occurs despite the presence of a posterolateral episiotomy in up to half the women.

Incontinence may arise after surgery to the anorectum. The two procedures most frequently complicated by unintentional postoperative incontinence are fistulotomy and maximal anal dilatation. It is the author's view, fistulotomy should be performed only by specialist coloproctologists, and anal stretch has no place in modern surgical practice. Other procedures less commonly implicated include haemorrhoidectomy (usually extensive) and overzealous lateral sphincterotomy. Many patients undergoing anorectal surgery will have pre-existing abnormalities of continence. A continence history should be taken before surgery in all high-risk cases and it is wise to test anorectal physiology. Accidental disruption of the sphincters may be associated with pelvic fractures, impalement injuries, gunshot or blast wounds, and the insertion of foreign bodies and deviant sexual practices. In these instances there is frequently concomitant damage to the non-sphincteric components of the continence mechanism, making these patients especially challenging to treat.

Clinical features

History

The assessment of patients with faecal incontinence relies heavily on a carefully taken history. A confidential setting and sympathetic atmosphere will help to overcome the reluctance of many patients to reveal the true nature of their symptoms. The degree, character, and frequency of leakage should all be determined. A history of urgency, the consistency of stool, and the presence or absence of awareness should be obtained. The severity of the incontinence, particularly in relation to its impact on the patient's lifestyle, is of fundamental significance. Past medical history, obstetric history, and previous surgery should be explored. The presence of concomitant pathology of the pelvic floor (rectal prolapse, concurrent urinary incontinence, and vaginal and/or uterine prolapse) should be ascertained. A thorough history coupled with a detailed examination should highlight any general conditions that require treatment before the management of the patient's incontinence. Objective assessment of severity of incontinence and its social impact can be made using one of a number of incontinence scores.

Examination

A holistic approach to every patient presenting with incontinence is essential because faecal leakage is often the presenting symptom of systemic disease. Particular attention is directed towards the examination of the back, abdomen, lower limbs, and perineum. Evidence of soiling, perineal scarring, fistulous openings, rectocele, or any deformity should be noted. The anocutaneous junction should be assessed at rest and then upon straining to determine the degree of perineal descent. Perineal sensation should be tested by using a pinprick technique or by evoking the anocutaneous reflex. Digital examination should be thorough and should include an appraisal of the rectal walls, the rectal contents, the symmetry of the anal canal, and the integrity of the perineal body. The integrity of the two components of the sphincter mechanism can be assessed in the clinic by determining the resting sphincteric tone and the voluntary squeeze increment (elicited by getting the patient to cough). Proctoscopy is essential to exclude the presence of haemorrhoids; fissure, fistula, abscess or mucosal prolapse, and sigmoidoscopy should always be undertaken to rule our rectal pathology.

Investigations

All patients should have their urine screened for glycosuria. If a neurological lesion is being considered, conventional radiology, including computed tomography and magnetic resonance imaging, of the chest, spine and pelvis should be undertaken to assess and/or exclude any skeletal deformities, soft-tissue lesions or disc pathology. Contrast studies of either the large or small bowel, and colonoscopy, are not routinely necessary but may be helpful in selected cases. The role of anorectal physiological tests in the management of patients with incontinence remains controversial. In experienced hands a carefully taken history and thorough physical examination are usually sufficient to elucidate the cause of the incontinence and the components of the continence mechanism that are abnormal. None the less, it is our opinion that anorectal physiological tests, including endoanal ultrasonography, are essential in the assessment of the incontinent patient. They facilitate diagnosis and provide objective data, which assists surgical management. In addition, they give useful indicators to the prognosis of different treatments and are useful research tools.

Assessment of the sphincters

The integrity of the sphincter complex can be assessed both anatomically and in terms of its motor function. A number of different methods can be used to image the sphincters. The advantages and disadvantages of each are summarized in Table 2.

Technique	Component identified	Advantages	Disadvantages
Endoanal ultrasound	Internal sphincter External sphincter, Puborectalis	Widely available No ionizing radiation, Relatively inexpensive Radiologist unneeded	Limited resolution Difficult to reconstruct, 3-D images
MRI	Internal sphincter External sphincter Pelvic floor	Excellent resolution 3-D images No ionizing radiation Pelvic floor abnormalities visualized	Expensive Limited availability Radiologist required
CT scanning	Internal sphincter External sphincter Pelvic floor	High resolution 3-D images	Ionizing radiation Relatively expensive Radiologist required

Table 2 Comparison of methods used to image the anal sphincters

The development of endoanal ultrasonography in the late 1980s has provided an invaluable tool for visualizing both the internal and external sphincter. The technique allows the size and distribution of defects to be mapped. It also allows the completeness of sphincter repair to be determined following surgery. Although dedicated equipment is required, this is relatively cheap and the technique can be readily learned with a little application. In time the introduction of spiral computed tomography and intra-anal magnetic resonance imaging may supersede anal ultrasonography. These newer techniques afford much higher degrees of resolution and associated developments in computer software allow three-dimensional images of the sphincter complex to be generated with ease.

The functional integrity of the sphincter complex is tested by measurement of the pressures generated by the individual components. The resting pressure within the anal canal is predominantly a function of tonic activation of the internal anal sphincter. The rise in pressure associated with a maximum squeeze arises from the recruitment of the external anal sphincter. Repeating the pressure measurements while simultaneously inflating a balloon in the rectum can be used to determine the integrity of the anorectal inhibitory reflex. This reflex is mediated by the submucosal and myenteric nerve plexuses, and is absent in conditions affecting them, for example Hirschsprung and Chagas diseases. Vector manometry affords three-dimensional assessment of sphincteric motor function and will reveal if any functional

asymmetry exists within the sphincter complex.

The external anal sphincter exhibits continual electrical activity as recorded by conventional concentric needle electromyography. Routine electromyographic sphincter mapping has been superseded by less painful and more accurate techniques. The technique is still useful in situations where the sphincter is anatomically intact but pressure measurements are low. Electromyogram traces may reveal characteristic denervation patterns (as seen in idiopathic faecal incontinence) or areas of electrical silence.

Studies of neurological function

While anatomical and physiological assessments are useful in the diagnosis of dysfunction in either the internal or external anal sphincters, they provide no insight into the underlying cause of the deficit. The pathophysiology may involve disruption, denervation, or myopathy. Treatment will depend on which of these predominates and further tests are required to differentiate between them.

There are no tests currently available to elucidate isolated defects in the innervation of the internal sphincter. The external sphincter is supplied via the somatic nervous system (the puborectalis muscle by S3 and S4 branches of the sacral plexus and the external sphincter from the pudendal nerve (S2, 3, 4). Nerve-conduction studies can be used at various sites to determine the integrity of innervation. The most clinically useful test is the pudendal nerve terminal motor latency. This is measured using a modified fingerstall that incorporates a stimulating electrode at its tip and a recording electrode at its base. By stimulating the pudendal nerve as it passes around the ischial spine and recording the electromyographic response, the latency of conduction can be calculated. Prolonged latencies are associated with trauma and diabetes, are found in 80 per cent of patients with idiopathic faecal incontinence, and indicate damage to the pudendal nerves. While there are currently no treatments that directly correct pudendal neuropathy, it remains an important diagnosis to make. This is particularly the case if there is a concomitant sphincter defect that is amenable to surgical correction. In the presence of pudendal neuropathy, repair of the anatomical defect alone is unlikely to enhance sphincter activity or continence, and other techniques have a higher failure rate in this group of patients.

Rarely, sphincter hypotonia occurs without any anatomical disruption of the sphincters or pudendal neuropathy. In these cases a muscle biopsy may sometimes be helpful to determine if there is any sphincter myopathy.

Tests of anorectal sensation

Normal anorectal sensation is essential to normal bowel function, alerting higher centres to impending incontinence and allowing the discrimination between flatus and faeces via the sampling reflex. While the anal canal is sensitive to the modalities of touch, temperature, pain and movement, these are difficult to quantify. Sensation is routinely measured using a constant-current electrode inserted into the anal canal with the current increased until the patient perceives a tingling sensation. The threshold to electrostimulation is increased in many patients with incontinence. Rectal sensation is measured by means of a balloon inserted into the rectum and slowly inflated. Three volumes are recorded: threshold volume of first sensation, volume when defaecatory desire becomes persistent, and maximum tolerated volume. Rectal sensation may be abnormally attenuated in conditions of proctitis, leading to urge incontinence, or dulled in patients with impaction and overflow.

Tests of continence and evacuation

Tests of continence may be useful in equivocal cases, enabling continence to solids or liquids to be quantified by determination of the force required to pull an inflated balloon through the anus at rest and maximum squeeze, or by the ability to retain infused saline. Difficulties in evacuation are frequently associated with incontinence and must be determined before treatment as excretory problems may be compounded. Evacuation is assessed by video evacuatory proctography using a mixture of dilute contrast medium, usually barium and porridge, inserted into the rectum. Lateral views of the pelvis are then taken as the patient attempts defaecation. This technique is useful in highlighting any underlying structural abnormality, such as a rectocele, intussuception or rectal prolapse, that may be amenable to surgical correction. This imaging method allows for the quantification of perineal descent and the anorectal angle. It can also be simultaneously combined with measurement of intrarectal pressure and sphincter electromyography to produce an integrated, dynamic image of anorectal emptying.

Test of rectal compliance and colonic motility

Occasionally, incontinence arises as the result of abnormally low rectal compliance or because of abnormalities in colonic motility. Measurement of compliance requires the simultaneous measurement of rectal pressure, using an intraluminal pressure sensor, while the rectum is distended at a constant rate. Colonic motility is most commonly assessed by means of colonic transit studies using radiopaque markers. Further information on rectal motility can be achieved by using ambulatory anorectal manometric techniques.

Treatment

Faecal incontinence can be treated by conservative measures, medical therapy, or surgery (Table 3). The strategy used depends on determining the underlying cause and assessing the severity of the symptoms.

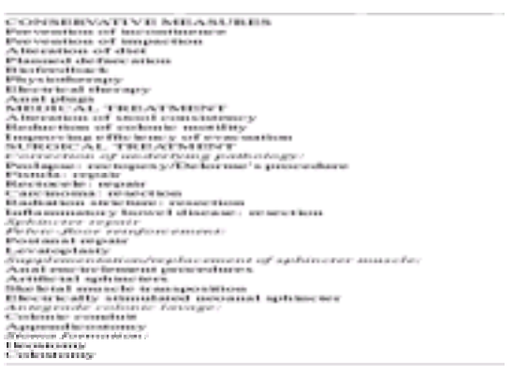


Table 3 Treatment of faecal incontinence

Conservative measures

The vast majority of incontinence arises from pudendal neuropathy, sphincter trauma, and faecal impaction. A prevention strategy aimed at these conditions will minimize the debilitating sequelae that ensue. Attempts to reduce the incidence of pudendal neuropathy focus on the early treatment of constipation, the avoidance of defaecatory straining, and alleviation of prolonged and difficult labour with the use of early episiotomy and caesarean section. Appreciation of the factors predisposing to obstetric-related sphincter injuries, the use of caesarean section in high-risk cases, and the early detection and careful repair of any injuries will reduce the morbidity from this source. The trend towards subspecialization should ensure that anal surgery is performed only by competent coloproctologists, thereby minimizing iatrogenic damage. Similar reductions can be achieved by the early treatment of anal fistula, careful assessment of patients with complex and recurrent conditions, and the abandonment of procedures such as anal stretch. The treatment of faecal impaction is directed both at its immediate relief and the prevention of its recurrence. Impaction should be treated with suppositories or enemas but, if severe, manual evacuation may be required. By recognizing at-risk individuals a regimen of regular defaecation can be introduced, with or without aperients.

Patients with incontinence related to idiopathic diarrhoea will often be improved by simple dietary measures without recourse to pharmacological agents. The fibre content of the diet should be modified and spicy foods, caffeine, alcohol, citrus fruits, and dairy products should be avoided. Planned defaecation is a useful tool in many patients suffering from incontinence arising from neurological causes. Often it is helpful to get patients to attempt evacuation after meals using the increase in colonic motility mediated by the gastrocolic reflex.

Biofeedback is the process of operant conditioning whereby patients learn to enhance sphincter activity in response to rectal distension. The technique requires considerable time and nursing resources, and is only appropriate for patients with some residual sphincter activity and a degree of anorectal sensation. Although response has been reported in up to 80 per cent of patients with mild to moderate symptoms, the technique is not proven to be of long-term benefit. Physiotherapy aimed specifically at the pelvic floor may improve sphincteric tone and has been used in conjunction with electrostimulation of the external sphincter either with an

anal-plug electrode or using the pudendoanal reflex. Anal plugs in the form of an anal tampon may temporize in some patients with minor leakage.

Medical therapy

Where diarrhoea is the predominant aetiological factor, treatment should be aimed at correcting this. Stool consistency may be enhanced with constipating agents, such as codeine phosphate, or by reducing colonic motility with loperamide. Inflammatory bowel disease should be treated with the appropriate medication. In patients whose incontinence is related to inadequate faecal emptying the judicious use of glycerine or bisacodyl suppositories or enemas will often result in complete resolution of their symptoms. In some patients, powerful laxatives, such as sodium picosulphate, or regular colonic irrigation are necessary.

Surgical options

For the majority of patients, simple conservative and dietary measures will suffice to ensure either complete resolution of their symptoms or reduction to an acceptable level. Unless clinical assessment points to an underlying condition that requires intervention, conservative measures should be attempted in all patients. The treatment of incontinence related to underlying coloproctological disease such as malignancy should be aimed primarily at that condition and oncological safety should not be compromised in an attempt to preserve continence. Resection may be required in cases of radiation proctitis, inflammatory bowel disease, or megacolon. If a rectocele or rectal prolapse are implicated, these require correction. These measures alone may be sufficient to restore continence. For those whose symptoms persist a suggested strategy for surgical correction is given in Fig. 1.

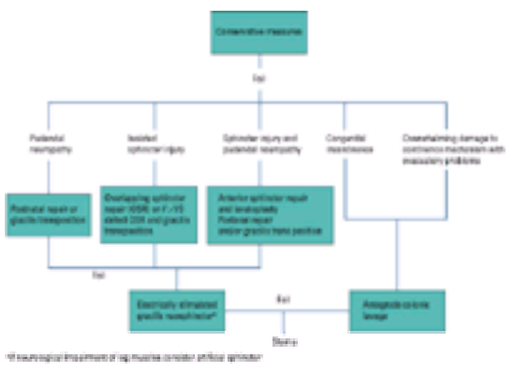


Fig. 1. Suggested algorithm for the management of patients with incontinence.

Sphincter repair

In cases of isolated sphincter damage, be it complete or partial, a direct sphincter repair is the treatment of choice. Direct sphincter injuries should be assessed with imaging and associated neuropathy excluded before surgery. The injury is explored and the disrupted ends identified and brought together in an overlapping repair (Fig. 2). Preservation of the scarred ends of the sphincter facilitates the apposition of the two ends. Successful results have been reported in 80 to 90 per cent of cases (Table 4), although we believe that this drops to nearer 50 per cent in the long term. Good success rates are reported in elderly people, and even when there has been a long interval between injury and repair. Effective repair is associated with a rise in voluntary squeeze pressures. The need for a defunctioning colostomy in the perioperative period remains contentious, with many surgeons opting for ‘medical defunctioning' for the first postoperative week. Defects greater than one-third of the anal diameter are less successfully treated by this technique alone. In such cases the repair can be combined with either a pelvic-floor repair or gracilis transposition. If function is poor after transposition, the gracilis can be stimulated at a later date.

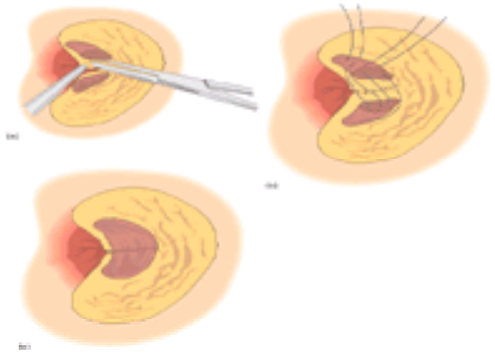


Fig. 2. Direct sphincter repair: (a) the divided ends of the sphincter are dissected; (b) the divided ends are sutured by an overlapping technique; (c) the completed overlapping repair.

Study	*No. of patients studied*	*Percentage continent of solids*
Hagihara and Griffen (1976)	7	85
Slade *et al.* (1977)	30	97
Cuzzo and Pittman (1978)	19	89
Rudd *et al.* (1982)	21	100
Keighley and Fielding (1983)	34	88
Fang *et al.* (1984)	75	90
Browning and Motson (1984)	85	92
Christiansen and Pedersen (1987)	23	95
Pezim *et al.* (1987)	40	85

Table 4 Results of anal sphincter repair

Pelvic-floor reinforcement

Sir Alan Parks devised postanal repair for the treatment of idiopathic faecal incontinence (Fig. 3). In such patients, incontinence is associated with pudendal neuropathy. The procedure was believed to work by reconstituting the anorectal angle and lengthening the anal canal following plication of the levator muscles. Good initial results were reported in up to 80 per cent of patients, but in all series the majority of patients deteriorate back to preoperative continence with only 20 to 30 per cent showing persistent benefit. Enthusiasm for the technique has been limited because of doubts over its clinical effectiveness and uncertainty over its physiological effects. The combination of postanal and anterior sphincter repair appears to provide better results. Anterior plication techniques have been described but have not been widely practised.

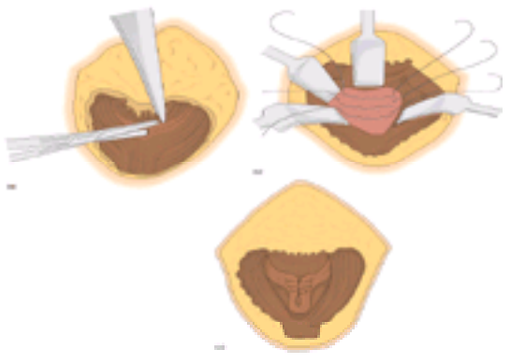


Fig. 3. Pelvic-floor repair: (a) posterior perineal dissection via the intersphincteric plane; (b) once the levators have been identified they are plicated; (c) the completed repair.

Supplementation/replacement of sphincter muscle

Patients with congenitally absent or grossly disrupted sphincters, and those in whom conventional repair techniques fail, present a greater challenge. Attempts have been made to fashion new sphincters from artificial materials, autologous tissues, or a combination of the two. Encirclement techniques using materials such as stainless-steel wire (the Thiersch technique) have been used to treat incontinence associated with rectal prolapse. A variety of other materials such as nylon, silicon bands, Silastic rods, and collagen injections have all been tried. While these may be effective in creating a passive barrier to the egress of faeces, their use has been limited because of complications with wound breakdown and sepsis. Prosthetic inflatable sphincter devices, successful in treating urinary incontinence, have recently been modified for use in patients with faecal incontinence. An inflatable cuff surrounds the anal canal, with a balloon reservoir placed in the anterior abdominal wall and a control pump in the scrotum or perineum. Early reports suggest that the technique may be useful in some patients, although mechanical failure, infection, and erosion are likely complications.

In order to minimize infective complications, attention has focused on using autologous tissue to reconstruct the sphincter complex. Of a plethora of skeletal muscles that have been tried, only two have gained lasting favour: the gluteus maximus and the gracilis. The results using these muscles when unstimulated are at best moderate and often short-lived. The transplanted muscle does not work as a dynamic sphincter and patients often have to perform awkward movements to achieve imperfect continence. Attempts to improve on the results of native muscle have centred on the application of chronic, low-frequency electrical stimulation to it once transposed. The technique of the electrically stimulated gracilis neoanal sphincter or dynamic graciloplasty (Fig. 4) involves placing a stimulating electrode on the main nerve to the gracilis, which is then freed from its distal attachment and transposed around he anal canal.

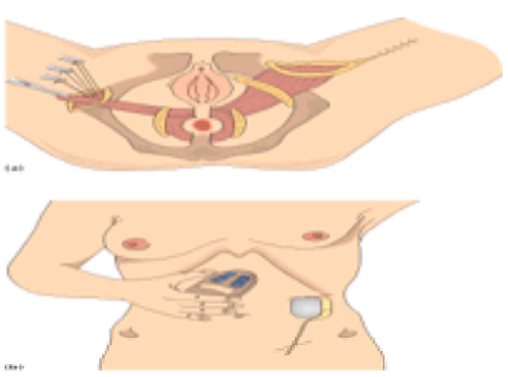


Fig. 4. (a) The gracilis muscle is harvested from the leg and transposed around the anal canal; (b) the muscle is stimulated via a totally implanted stimulator, which can be controlled by the patient.

The electrode is then connected to a totally implanted stimulator that can be switched off by the patient when defaecation is desired. The use of electrical stimulation not only provides a true dynamic sphincter, controlled by the patient, but also effects phenotypic conversion of the muscle, making it more appropriate for sphincteric function. Substantial improvements in continence are reported in 60 to 70 per cent of patients refractory to other treatment options. Failures are usually related to technical sequelae, either infective or mechanical. In a proportion of patients the neosphincter preserves continence too effectively and they experience evacuatory difficulties. The long-term efficacy of the procedure will be determined by current clinical trials.

Antegrade irrigation techniques

Another approach to the patient with absent or irreparably damaged sphincters is to regular cleanse the colon using bowel lavage. By achieving complete colonic emptying the patient experiences long and predictable periods without leakage. Lavage may be instilled in a retrograde or antegrade direction. Retrograde irrigation, using a rectal tube, is difficult to perform effectively and is not acceptable to many patients. Antegrade irrigation is performed by intubating the colon with a catheter via a surgically constructed fistula. The two techniques commonly performed use an appendicostomy or a continent, intussuccepted valve in the colon as a means of access to the transverse colon. The patient irrigates with tap water until evacuation is complete and is then continent for the rest of the day. Although an appendicostomy may be adequate for children, it is frequently absent or unsuitable for use in adults. The diameter of the appendix limits the size of catheter used and may be too small to allow for the necessary flow to produce effective irrigation. For these reasons a colonic conduit (Fig. 5) is preferred in adults. These techniques have produced substantial improvements in quality of life for patients with irreparable sphincters and intractable symptoms.



Fig. 5. Continent colonic conduit.

Some patients will be unsuitable for major surgical intervention or resistant to other treatment options. In this group a permanent stoma may be the only way to manage their symptoms. Whether an ileostomy or colostomy is constructed will depend on patient-related factors and the surgeon's preference. This should be considered as a last resort as considerable morbidity, both psychological and physical, is associated with life with a stoma.. Nonetheless, a carefully constructed stoma may afford the patient sufficient control over evacuation to greatly improve their quality of life.

Conclusions

Faecal incontinence is a common symptom that incurs considerable misery and social costs. Its causation is multifactorial and patients require careful investigation to determine the best form of treatment. Conservative measures will ameliorate symptoms in many. A number of surgical options exist for the minority resistant to these.

Further reading

Browning GCP, Motson RW. Anal sphincter injury. Management and results of Park's sphincter repair. *Annals of Surgery* 1984; **199**: 351–6. [Results of primary overlapping sphincter repair in 85 patients.]

Castro AF, Pitman RE. Repair of the incontinent sphincter. *Diseases of the Colon and Rectum* 1978; **21**: 183–7. [Results of 19 patients undergoing primary sphincter repair.]

Christiansen J, Pedersen IK. Traumatic anal incontinence, results of surgical repair. *Diseases of the Colon and Rectum* 1987; **30**: 189–91. [Results of delayed primary sphincter repair in traumatic damage to the spincter in 23 patients.]

Christiansen J, Lorentzen M. Implantation of an artificial sphincter for anal incontinence. Report of five cases. *Diseases of the Colon and Rectum* 1989; **32**: 432–6. [First reports of use of an artificial anal sphincter.]

Fang DT *et al.* Overlapping sphincteroplasty for acquired anal incontinence. *Diseases of the Colon and Rectum* 1984; **27**: 720–2. [Results of series of overlapping sphincter repairs.]

Fleshman JW *et al.* Anal sphincter reconstruction: anterior overlapping muscle repair. *Diseases of the Colon and Rectum* 1991; **34**: 739–43. [[Primary sphincter repair in 55 patients.]

Hagihara PF, Griffin WO. Delayed correction of anal incontinence due to anal sphincteric injury. *Archives of Surgery* 1976; **111**: 63–6. [Early report of sphincter repair.]

Keighley MR. Faecal incontinence. In: Keighley MRB, Williams NS, ed. *Surgery of the anus, rectum and colon*, Vol. 1, pp. 516–608. Saunders, London, 1993. [Comprehensive study on aetiology, pathophysiology, investigation, and treatment of faecal incontinence.]

Keighley MR, Fielding JW. Management of faecal incontinence and results of surgical treatment. *British Journal of Surgery* 1983; **70**; 463–8. [Results of sphincter repair in 34 patients.]

Mander BJ, Williams NS. The electrically stimulated gracilis neoanal sphincter. *European Journal of Gastroenterology and Hepatology* 1997; **9**: 435–41. [Review of the use of autologous muscle, stimulated or unstimulated, to replace anal sphincters.]

Mann CV, Glass RE. *Surgical treatment of anal incontinence*. Springer-Verlag, London, 1997. [Comprehensive surgical atlas on the techniques for treating faecal incontinence.]

Penninckx F. Fecal incontinence: indication for repairing the anal sphincter. *World Journal of Surgery* 1992; **16**: 820–5.[Review on the indications, investigation, and techniques of sphincter repair.]

Pezim ME *et al.* Sphincter repair for faecal incontinence after obstetric or iatrogenic injury. *Diseases of the Colon and Rectum* 1987; **30**: 521–5. [Results of sphincter repair in 40 patients.]

Rogers J, Mander BJ. Anorectal physiology in surgical management. In: Taylor I, Johnson CD, ed. *Recent advances in surgery*, Vol. 20, pp. 89–102. Churchill Livingstone, London, 1997. [Comprehensive review of anorectal physiology tests.]

Rudd WH *et al.* Anal incontinence (symposium). *Diseases of the Colon and Rectum* 1982; **25**: 90–107. [Results of primary overlapping sphincter repair.]

Schoetz DJ. Operative therapy for anal incontinence. *Surgical Clinics of North America* 1985; **65**: 35–46. [Review of surgical options in treatment of incontinence.]

Setti-Carraro P, Kamm MA, Nicholls RJ. Long term results of post-anal repair for neurogenic faecal incontinence. *British Journal of Surgery* 1994; **81**: 140–4. [Assessment of the efficacy of postanal repair.]

Slade MS *et al.* Sphincteroplasty for acquired anal incontinence. *Diseases of the Colon and Rectum* 1977; **20**: 33–5. [Results of overlapping sphincter repair in 30 patients.]

Vaizey CJ, Kamm MA, Nicholls RJ. Recent advances in the surgical treatment of faecal incontinence. *British Journal of Surgery* 1998; **85**: 596–603. [Recent review of the surgical management of faecal incontinence.]

Williams NS, Hughes SF, Stuchfield B. Continent colonic conduit for rectal evacuation in severe constipation. *Lancet* 1994; **343**: 1321–4. [Introduction to the technique of antegrade colonic lavage via a continent intussusscepted valve.]

28.6 Cancer of the anus

Michael J. Greenall

Introduction
Epidermoid cancer of the anus
Anatomical considerations
Pathology
Clinical features
Diagnosis
Treatment
Malignant melanoma
Adenocarcinoma of the anus
Basal-cell carcinoma
Anal intraepithelial neoplasia (AIN)
Paget's disease
Further reading

Introduction

The most common type of cancer arising in the anus is squamous or epidermoid carcinoma (Table 1). This is, however, a rare tumour, and is 20 to 30 times less common than colorectal cancer. Of the other tumours that occasionally arise in the anus, only malignant melanoma is seen with any frequency; this is eight times less common than epidermoid carcinoma. Other tumours are so rare that there is only anecdotal experience of their treatment.

1. Epidermoid anal cancer:
(a) at the anal margin
(b) in the anal canal
2. Melanoma
3. Adenocarcinoma
Either arising de novo from anal glands or fistula, or from down-ward spread of rectal adenocarcinoma
4. Basal-cell carcinoma
5. Bowen's disease
6. Paget's disease

Table 1 Anal cancer

Epidermoid cancer of the anus

Anal epidermoid cancer is sensitive to both chemotherapy and irradiation. This has led to reappraisal of its surgical treatment, with a trend away from abdominoperineal resection to a much more conservative surgical approach that allows preservation of the anal sphincter mechanism.

Much of the early information on anal cancer was derived from small, retrospective studies that used a variety of pathological terms and staging systems. Other studies combined their results with those for other types of anal cancer such as melanoma and adenocarcinoma. Data were therefore conflicting. In addition, there has never been any agreement about the anatomical boundaries that distinguish the anal canal from its margin; epidermoid cancers arising at these two sites have, therefore, not been accurately differentiated. As epidermoid cancer arising in the canal differs fundamentally in its clinical presentation, pathological features, treatment, and prognosis from that at the margin, distinction between the two is essential.

Anatomical considerations

Anatomists regard the anal canal as that part of the alimentary tract distal to the rectal ampulla. However, this definition does not correlate with the clinical and pathological characteristics of anal cancer. Although there is general agreement that the proximal end of the anal canal corresponds to the anorectal ring, its distal limit is more poorly defined; both the dentate (pectinate) line and the anal verge have been used (Fig. 1). Similarly, there is no clear definition of the lateral border of the anal margin, although one authority has recommended that it lies within a 5 cm radius of the verge itself.

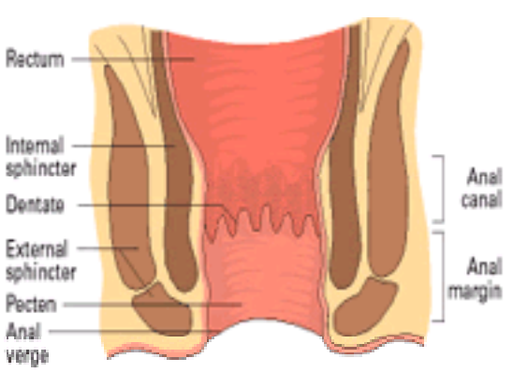


Fig. 1. The anatomy of the anal canal and anal margin.

The definition of the distal limit of the anal canal as the dentate line or the anal verge is of paramount importance since the anatomical boundary used to differentiate between the anal canal and its margin determines the relative incidence of tumours at these two sites. When the anal verge is used as the boundary, fewer than 15 per cent of cancers are ascribed to the margin, whereas those authorities using the more proximal dentate line claim that 30 per cent of tumours occur at this site.

The dentate line itself is a cause of confusion. It corresponds to the site of the anal valves and represents the junction of the postallantoic gut with the proctodeum. It is described as the ‘anal transitional zone' because of its transitional-cell appearance with characteristics of both the cuboidal rectal mucosa above and the modified squamous epithelium of the pecten below. The transitional zone does not, however, correspond to the macroscopic limits of the dentate line. Transitional epithelium, with islands of squamous cells within it, extends up to 2 cm above the line, explaining how tumours of this epithelial type occur in the more proximal anal canal.

Pathology

There has been much recent interest in the pathology of anal cancer, owing to increased understanding of the anatomy of the anal region and because of interest in premalignant changes in that area. ‘Anal intraepithelial neoplasia' (**AIN**) describes dysplasia in the squamous epithelium, with loss of cellular maturation, nuclear hyperchromasia, pleomorphism, and abnormal mitoses. Such features are similar to those seen in cervical and vulvar squamous carcinomas, and just as in those carcinomas, anal squamous cancer is aetiologically linked with human papillomavirus type 16. Similarily, the extent of the cellular changes identified in AIN are classified according to genital intraepithelial neoplasia. In AIN I, changes are restricted to the lower third of the epithelium, whereas in AIN II and AIN III the lower

two-thirds and full thickness of the epithelium are affected, respectively.

Both AIN I and AIN II have minimal malignant potential and may even regress spontaneously. However, AIN III may be associated frank anal squamous cancer as a result of malignant transformation. AIN III may correspond to ‘Bowen's disease', although this term should be avoided if possible.

Invasive epidermoid carcinoma of the margin accounts for only about 15 to 30 per cent of all squamous anal cancers. Some 85 per cent are well differentiated and produce keratin. Basaloid features are rare and presumably result from the development of an invasive squamous component within a pre-existing basal-cell carcinoma.

Invasive epidermoid cancers of the anal canal are usually more poorly differentiated, and only about 30 per cent produce keratin. Basaloid (cloacogenic or transitional) features, found in about 40 per cent of tumours, imply derivation from the anal transitional zone. When basaloid features are present there is usually a squamous component as well: classification as squamous or basaloid is based on the predominant cell type, although this is somewhat academic and subject to differences in interpretation.

Clinical features

The most common presenting features of anal cancer are pain, bleeding, change in bowel habit, and pruritus ani. Less common presentations include inguinal lymphadenopathy from secondary spread to that site, and as an unexpected histological finding in a haemorrhoidectomy specimen.

About 60 per cent of patients with epidermoid cancer at the anal margin have associated perianal conditions, including condylomas, chronic fistula, leukoplakia, or the effects of previous irradiation. Such features are found in less than 10 per cent of patients with tumours arising in the anal canal.

The relation between anal cancer and sexually transmitted diseases is an area of particular interest. Earlier studies suggested an association with lymphogranuloma venereum and syphilis, although this was not subsequently confirmed. However, more recent epidemiological data have provided evidence in support of the hypothesis that a sexually transmitted agent is involved in the pathogenesis of anal squamous cancer, especially in the male homosexual population. Human papilloma virus type 16 DNA has been detected in more than 50 per cent of patients with anal cancer, indicating that this may be the transmissible agent involved.

Diagnosis

Epidermoid anal cancer is diagnosed on the basis of examination of biopsy specimens. The differential diagnosis includes conditions such as rectal adenocarcinoma, Bowen's disease (AIN III), Paget's disease (of skin), condyloma acuminata, leukoplakia, Crohn's disease, and certain intrinsic skin disorders such as lichen sclerosus et atrophicus. At the time of diagnosis it is necessary to determine accurately the stage of the disease. Difficulties encountered in staging on a clinical basis have been overcome by the increasing use of local ultrasonography to assess muscle invasion. Computed tomographic or magnetic resonance imaging may also help determine the presence of local invasion or metastatic disease.

Treatment

Invasive epidermoid cancer of the anal margin

Local excision is sufficient treatment in the majority of patients with epidermoid cancer of the anal margin, giving a 5-year survival rate in excess of 80 per cent, although large tumours may require skin grafting. The prognosis may be determined by the depth of invasion.

Abdominoperineal resection has not been widely adopted for patients with tumours at the anal margin. When it has been used it has been usually reserved for deeper infiltrative tumours or for patients with persistent recurrence; the survival rate of about 50 per cent is therefore somewhat less than that obtained after simple local excision. A potential advantage of abdominoperineal resection is that inferior mesenteric nodal metastases will be resected, although such spread is rare in cancers arising at the margin.

There are few data on primary irradiation or combined chemotherapy and irradiation of epidermoid cancer of the anal margin. Although responses have been observed, there seems little need for this approach as simple local excision will suffice in the vast majority of patients.

If recurrence occurs after local excision of anal-margin cancer, further simple excision may be attempted. If this is not possible, radiotherapy, using either external-beam irradiation or an implant, may be considered. Occasional patients require abdominoperineal resection, although good local tumour control may be difficult to achieve if there is persistent recurrence.

The good survival following local excision of epidermoid cancer of the anal margin results from the relatively benign characteristics of the tumour. These relatively well-differentiated, keratinizing squamous tumours therefore behave rather like other primary skin cancers arising elsewhere on the body.

Invasive epidermoid cancer of the anal canal

Epidermoid cancer of the anal canal has a worse prognosis and requires more complex treatment than its relatively benign counterpart at the margin.

A wide variety of treatments have been adopted for patients with tumours in the anal canal, ranging from local excision to preoperative neoadjuvant chemotherapy and radiotherapy followed by abdominoperineal resection. It is its potential for chemo- and radiosensitivity that has caused much interest in the past few years.

Local excision

Simple local excision alone is only rarely applicable to patients with epidermoid cancer of the anal canal. When performed the prognosis has been relatively good, with a 5-year survival of up to 65 per cent: this presumably reflects the early stage of disease for which the operation was performed. It is probably only suitable for those superficially invasive tumours that are less than 2 cm in diameter, and is not generally recommended.

Abdominoperineal resection

Abdominoperineal resection has been the standard treatment for epidermoid cancer of the anal canal, with a 5-year survival rate of 50 per cent. A wide perineal phase of the operation is recommended; if inguinal lymph nodes are involved, inguinal lymphadenectomy is required as a secondary procedure. Survival depends on tumour size, histological grade, depth of invasion, and overall staging. Earlier suggestions that predominantly basaloid tumours have a better prognosis than the more usual squamous type have not been substantiated.

Primary radiotherapy

Primary irradiation has been used in the treatment of epidermoid anal cancer for over 50 years. In the past the various radiotherapeutic techniques were inadequate and produced disappointing results in terms of both tumour control and local complications. However, more modern megavoltage X-ray therapy has produced encouraging results, with 5-year survival rates in excess of 50 per cent, and with the added advantage of allowing preservation of the anal sphincter. Interstitial techniques have also been used with success in some centres.

Primary irradiation therapy may sometimes be associated with severe local complications and patients may require a colostomy because of anal necrosis or stenosis.

Neoadjuvant chemoirradiation

The original, somewhat empirical, observation by Nigro and his colleagues in 1974 of the effects of preoperative treatment of epidermoid anal cancer with combination chemotherapy and irradiation attracted much attention. Their use of 5-fluorouracil and mitomycin C was somewhat arbitrary, although both agents had demonstrated cytotoxic activity against a wide range of gastrointestinal tumours and also exhibited properties of radiosensitization.

The chemoirradiation regimen now currently used varies between institutions. Most give mitomycin C on day 1 of the treatment, followed by infusion of 5-fluorouracil

over a 5- to 7-day period. Irradiation starts either concurrently with chemotherapy or 3 days after the mitomycin C has been given. The average radiation dose is about 30 to 50 Gy, given over 3 to 7 weeks, a lower dose than is usually recommended for patients treated by radiation alone. These regimes are associated with side-effects such as radiation-induced proctitis and dermatitis, leucopenia, thrombocytopenia, stomatitis, and diarrhoea.

In most centres, patients are re-examined about 4 to 6 weeks after completion of radiotherapy: About 70 per cent will show a complete response to their treatment and in only 10 per cent is there no effect. Those patients who fail to respond require abdominoperineal resection. The major question is whether, after a complete response, simple excision of residual scar tissue will suffice, or whether these patients should be treated by abdominoperineal resection. Current data provide no real answer, but do indicate that neoadjuvant chemoirradiation may reduce the need for abdominoperineal resection. Whether it improves survival is less clear. The UKCCR study on 585 patients comparing chemoirradiation with radiotherapy alone in a randomized, multicentre trial has demonstrated improved local control with combined therapy. Only 36 per cent of patients receiving chemoirradiation had local failure compared to 59 per cent after irradiation. There was no significant survival advantage for either group, but the results do imply that combination therapy should be standard treatment for anal squamous cancer and that surgery be reserved for treatment failures

Malignant melanoma

Malignant melanoma of the anus is rare and has a poor prognosis. Whereas one can usually expect a 50 to 70 per cent 5-year survival for cutaneous melanoma occurring elsewhere in the body, the average 5-year cure rate reported in the literature for malignant melanoma in the anal region is only 6 per cent.

Malignant melanoma of the anal region usually presents with bleeding or a mass. It often appears as a slightly pigmented lesion, although amelanotic melanoma is recognized at this site. Unfortunately, in many patients the disease is already involving the inguinal lymph nodes at the time of presentation.

Because of its rarity, virulence, and poor prognosis the treatment of this tumour has varied from simple local excision to radical abdominoperineal resection with bilateral prophylactic groin dissection. Adjuvant chemotherapy and radiation treatment seem to provide little benefit.

The majority of long-term survivors have been treated by abdominoperineal resection. There seems to be little benefit of prophylactic inguinal-node dissection, although staging by nodal biopsy at the time of abdominoperineal resection may be useful clinically.

Adenocarcinoma of the anus

Adenocarcinoma of the anal canal may result from downgrowth of a primary rectal tumour or from carcinoma developing *de novo* in an anal gland or fistula. Such tumours are rare, their prognosis is poor, and treatment is by abdominoperineal resection.

Basal-cell carcinoma

Basal-cell carcinoma is another rare anal condition. Experience is anecdotal but full-thickness local excision appears to assure a cure.

Anal intraepithelial neoplasia (AIN)

AIN appears as a slowly growing, minimally elevated, erythematous, plaque-like lesion. Multiple biopsies allow classification as AIN I, II or III, and exclude an association with invasive cancer.

As the malignant potential of AIN I and II is low, simple follow-up will suffice, with repeat biopsy after 6 months to a year. The treatment of AIN III is wide local excision, with skin grafting if necessary: special attention must be paid to the margin of excision. Long-term follow-up is necessary because of the possibility of recurrent disease in an area predisposed to the development of this condition.

Paget's disease

Extramammary Paget's disease is an extremely rare intraepithelial neoplasm. Unlike its counterpart in the breast it is not always associated with frank underlying malignancy. The clinical appearance of Paget's disease is that of a pale grey, crusting, scaly lesion. The diagnosis is confirmed by demonstrating the characteristic Paget cells with their large, pale cytoplasm that stains positively with aldehyde-fuscin. Any associated invasive cancer arises from a skin appendage such as a sweat gland rather than from rectal mucosa.

If no frank cancer is present, then simple excision is suitable treatment for Paget's disease. Negative margins of resection must be obtained or recurrence is likely. If there is underlying malignancy, then a deeper, wider excision is needed, but the outlook is generally poor.

Further reading

Fenger C. The anal transitional zone. Location and extent. *Acta Pathologica, Microbiologica et Immunologica Scandinavica* (A) 1979; **87**: 379–86. [This paper remains the definitive description of the scientific basis of anatomy of the anal canal.]

Greenall MJ, Quan SHQ, DeCosse JJ. Epidermoid cancer of the anus. *British Journal of Surgery* 1985; **72** (Suppl.): S97–103. [A major review of the historical development of treatment of anal cancer.]

Greenall MJ, Quan SHQ, Stearns MW, Urmacher C, DeCosse JJ. Epidermoid cancer of the anal margin: pathologic features, treatment and clinical results. *American Journal of Surgery* 1985; **149**: 95–101. [The largest series of anal margin cancers.]

Nigro ND, Vaitkevicius VK, Considine B. Combined therapy for cancer of the anal canal: a preliminary report. *Diseases of the Colon and Rectum* 1974; **17**: 354–56. [The original paper demonstrating the chemosensitivity of anal cancer.]

Palmer JG *et al.* Anal cancer and human papillomavirus. *Diseases of the Colon and Rectum* 1989; **32**: 1016–22. [The definitive demonstration of an infective aetiology in anal cancer.]

Schofield J *et al.* Anal intraepithelial neoplasia: part of a multifocal disease process. *Lancet* 1992; **340**: 1271–3. [The first definitive paper describing a scientific classification of premalignant changes in anal cancer.]

UKCCR Anal Cancer Trial Working Party. Epidermoid anal cancer: results from the UKCCR randomised trial of radiotherapy alone versus radiotherapy, 5 fluouracil and mitomycin. *Lancet* 1996; **348**: 1049–54. [The only randomized trial evaluating chemoirradiation in anal cancer.]

29 Sacrococcygeal pilonidal sinus

Jan Rakinic and Robert Fry

Treatment
Pilonidal abscess
Chronic pilonidal sinus
Recurrences and unhealed wounds
Curettage and saucerization
Re-excision and closure
Pilonidal sinus and carcinoma
Summary
Further reading

Pilonidal sinus afflicts 0.7 per cent of postadolescent young adults. It is rare in childhood, and generally becomes asymptomatic by about the age of 40 years. Males are afflicted 2.2 to 4 times as frequently as females. There is a higher incidence among hirsute individuals.

The diagnosis of pilonidal sinus is confirmed by the presence of primary midline openings, or pits, in the intergluteal cleft, approximately 5 cm posterior to the anus (Fig. 1). Hairs disconnected from the skin may be seen protruding from the pits. The sinus tract is lined with squamous epithelium, and extends for a variable distance. Secondary tracts or abscess cavities may branch off the primary tract. Secondary openings are marked by elevations of granulation tissue and discharge of seropurulent material.

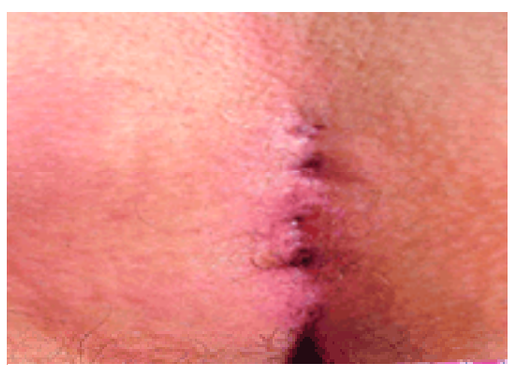


Fig. 1. Pilonidal sinus (reproduced by kind permission of R. G. Souter).

Most tracts run in a cephalad direction; less than 10 per cent run caudad and may be difficult to distinguish from anal fistula or hidradenitis. Careful examination will reveal one or more midline pits, although these may occasionally be obscured by edema. Differential diagnosis includes skin furuncle, syphilitic or tubercular granuloma, osteomyelitis with draining sinuses, and actinomycosis.

Treatment

Pilonidal abscesses are polymicrobial, with anaerobes (especially *Bacteroides* spp. and anaerobic cocci) predominating. However, cultures from chronic pilonidal sinuses grow predominantly aerobes. Antibiotics have no benefit in relation to wound complications, time to complete healing, or recurrence rate. Perioperative antibiotics are not indicated, except in selected conditions such as the diabetic or otherwise immunocompromised patient.

Pilonidal abscess

This is the most common presentation of pilonidal disease. The incision to drain a pilonidal abscess should be placed lateral to the midline, as wounds in the intergluteal cleft heal slowly. Drainage can almost always be accomplished under local anaesthesia in the office or emergency room. Hair in the cavity should be removed and the wound curetted. Postoperatively the wound is cleansed daily in the shower, and surrounding hair is shaved every 2 weeks for at least 3 months after complete healing. This will cure pilonidal disease in about three-quarters of patients.

Chronic pilonidal sinus

Incision/marsupialization

Tracts are opened in the midline and curetted. The fibrous tissue of the tract is not excised but sutured to the wound edges to produce a smaller, shallower wound. This can be done as an outpatient procedure. Daily wound cleansing and frequent changes of dressing are required; removal of surrounding hair is felt to be essential. Healing is usually complete within 6 weeks. Recurrence is approximately 10 per cent.

Conservative excision (Fig. 2)

Lord and Millar first advocated excision of the midline pits and curettage of the sinus cavity with a small brush through the midline pits without excision of the sinus. Bascom employs a similar approach, with a lateral incision used to enter and curette the sinus cavity. The midline pits are excised with ‘rice-grain'-sized specimens. The sinus cavity itself is not excised. Bascom advocates closing the small midline wounds, leaving the lateral wound open to drain. In this author's (J. R) experience the midline wounds usually open postoperatively, so the lateral wound has been closed primarily without problems. Postoperative shaving of the area surrounding the wound is very important. Cure rates are as high as 84 per cent. Advantages include small wounds, outpatient surgery, minimal dressing changes, an acceptable recurrence rate, and rapid healing, usually within 3 weeks. This is the author's (J. R) technique of choice for chronic pilonidal sinuses, and can be used for suitable recurrences.

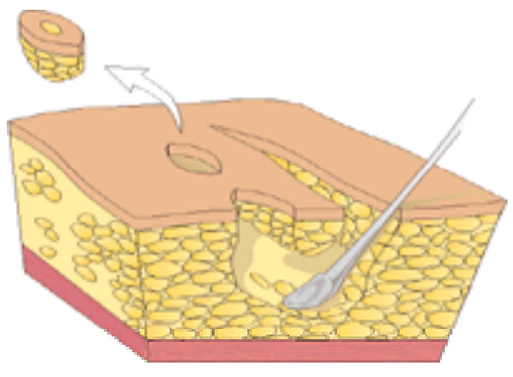


Fig. 2. An incision is made lateral to the midline, through which the cavity is curetted of hair and debris; the midline pits are excised separately or *en bloc*.

Wide local excision

All sinus tracts and a 5-mm rim of normal tissue are excised *en bloc* down to the sacrococcygeal fascia. The gluteal fascia is not entered. Wounds heal more rapidly if closed primarily, but recurrence rates can be as high as 38 per cent. Complete healing by secondary intention takes an average of 2 months, with frequent changes of dressing required. This technique offers no advantage over marsupialization.

Flap procedures

Flap procedures, such as Z-plasty, V–Y advancement, and other techniques require extensive dissection into normal tissue, and failure can result in significant loss of tissue. This approach should be reserved for management of the recurrent or unhealed wound.

Phenol treatment

An injection of 1 to 2 ml of 80 per cent phenol is made into the tract and repeated every 4 to 6 weeks as necessary. This method destroys the epithelium, sterilizes the tract, and removes embedded hair. Because of the intense local inflammation, most patients should stay in hospital overnight. Afterwards, shaving of the region is absolutely necessary. Cure rates as high as 91 per cent have been reported, but 70 per cent is more common. Median time to complete healing is 1 to 2 months. This approach has not gained wide acceptance.

Recurrences and unhealed wounds

Rates of recurrence vary widely (0–40 per cent) among published series. Causes are poorly understood. Theories include infection of the scar, causing abscess; and a suction effect from motion, coupled with anatomical characteristics of the intergluteal cleft, drawing hair into the cavity and resulting in a chronic infection. Extensive recurrent disease and unhealed wounds can be extremely challenging problems, often complicated by loss of tissue from previous attempts at excision. The principles of treatment for recurrent disease are the same as for primary disease: therapy should be tailored to the extent of disease; and as midline wounds in the intergluteal cleft heal poorly, they should be avoided.

Curettage and saucerization

All surrounding hair is shaved. Granulation tissue is completely debrided. The wound is saucerized to avoid premature closure and the accumulation of discharge. Meticulous postoperative wound care and frequent changes of dressing are required, together with the removal of surrounding hair. Healing is likely to be protracted.

Re-excision and closure

An extensive wound that persists after conservative therapy should be re-excised. The unhealed wound, scar, and granulation tissue are excised to the normal fat and sacrococcygeal fascia. Loss of tissue generally requires a flap procedure to achieve closure. The following techniques are useful, depending on the availability of tissue locally and the surgeon's preference. Reported recurrence rates vary from 6 to 20 per cent

Z-plasty

Excision of the sinus is carried out down to the subcutaneous tissues. The limbs of the Z are cut to form a 30° angle with the midline. Full-thickness flaps are raised, transposed, and sutured; closed suction drains are placed beneath the flaps. Recurrence is reportedly less than 5 per cent. This technique is extensive, and unsuitable for the outpatient setting.

Advancing flap (Karydakis' procedure)

An elliptical incision is made over the sinuses, but centered off the midline, and carried down to the sacrococcygeal fascia. Sinuses are excised *en bloc* with the ellipse of overlying skin. A full-thickness flap is mobilized on the side opposite the excision and advanced contralaterally for closure, thus avoiding a midline wound. The wound is closed in several layers over a suction drain. In a series of 1687 patients, hematoma or infection developed in 8.5 per cent, and recurrence in 1.3 per cent.

Cleft closure

This is Bascom's modification of Karydakis' flap procedure, differing largely in that Karydakis' flap includes fat, whereas Bascom's is a skin flap over *in situ* fat. After excision of the sinus tract, a skin flap is mobilized over the *in situ* fat, flattening the intergluteal cleft. The final wound is lateral to the midline. This can be done on an outpatient basis. Recurrence is reported as 3.3 per cent.

Other advancement flaps

Many other techniques have been described, including the rhomboid, gluteus maximus myocutaneous, and V–Y advancement flaps. These procedures require admission to hospital and extensive dissection into normal tissues.

Pilonidal sinus and carcinoma

There are 44 reported cases of carcinoma arising in a chronic pilonidal sinus. The gross appearance is an ulcer with friable, fungating margins. Most are well-differentiated squamous-cell carcinomas. Treatment is wide excision with reconstruction by skin grafting, or local advancement or rotational flaps. If inguinal nodes contain tumor, as reported in 14 per cent of cases, the prognosis is poor. Recurrence rates approach 50 per cent and 5-year survival is 50 per cent. Chemoradiation has recently been proposed as adjunct treatment to help minimize local recurrence.

Summary

The treatment of pilonidal disease can cause postoperative disability out of proportion to the severity of the original disease. Conservative approaches to treatment appear to be gaining favor. Proponents of minimal operative therapy point out that pilonidal disease commonly resolves by the age of 40 or so regardless of treatment. These physicians favor shifting the focus of care from the operating theater to the clinic, and from ‘cure' to ‘management' of the disease.

Simple drainage of pilonidal abscesses and shaving of the region results in cure of 60 to 80 per cent of cases. For those who develop a chronic pilonidal sinus, minimal excisional techniques emphasizing incisions lateral to the midline with primary closure appear to provide more rapid wound healing, lower recurrence rates, fewer days as an inpatient (many of the procedures can be done in the outpatient), and less time lost from work or school. Extensive recurrent disease and unhealed wounds remain challenges, and are most appropriately treated by re-excision and flap closure.

Further reading

Armstrong JH, Barcia PJ. Pilonidal sinus disease. The conservative approach. *Archives of Surgery* 1994; **129**: 914–17. [Excellent report of longitudinal follow-up after nonexcisional therapy.]

Bascom J. Pilonidal disease: origin from follicles of hairs and results of follicle removal as treatment. *Surgery* 1980; **87:**, 567–72. [Classic in nonexcisional therapy.]

Bascom J. Pilonidal disease: long-term results of follicle removal. *Diseases of Colon and Rectum* 1983; **26**: 800–7. [Classic in nonexcisional therapy.]

Davis KA, Mock CN, Versaci A, Lentrichia P. Malignant degeneration of pilonidal cysts. *American Surgeon* 1994; **60**: 200–4. [Largest review of malignant degeneration of pilonidal cysts.]

Jensen SL, Harling H. Prognosis after simple incision and drainage for a first-episode acute pilonidal abscess. *British Journal of Surgery* 1988; **75**: 60–1. [Landmark paper describing high cure rate after simple incision and drainage.]

Karydakis GE. Easy and successful treatment of pilonidal sinus after explanation of its causative process. *Australian and New Zealand Journal of Surgery* 1992; **62**: 385–9. [Interesting approach,

although extensive for use in primary disease.]

Kelly SB, Graham WJH. Treatment of pilonidal sinus by phenol injection. *Ulster Medical Journal* 1989; **58**: 56–9. [Good summary of expectations for phenol therapy.]

Kronborg O, Christensen K, Zimmerman-Nielsen C. Chronic pilonidal disease: a randomized trial with complete 3-year follow-up. *British Journal of Surgery* 1985; **72**: 303–4. [Primary wound closure versus closure by secondary intention.]

Lord PH, Millar DM. Pilonidal sinus: a simple treatment. *British Journal of Surgery* 1965; **52**: 298–300. [Clasic paper, describing nonexcisional therapy.]

Sondenaa K, Nesvik I, Andersen E, Natas O, Soreide JA. Bacteriology and complications of chronic pilonidal sinus treated with excision and primary closure. *International Journal of Colorectal Disease* 1995; **10**: 161–6. [Interesting reading regarding basis (or lack therof) for antibiotic therapy in this disease.]

Toubanakis G. Treatment of pilonidal sinus disease with the Z-plasty procedure (modified). *American Surgeon* 1986; **52**: 611–12. [Elegant description of Z-plasty technique.]

30.1 Surgical anatomy of the liver and biliary tree

Adrian Savage and Julian Britton

Development
Surgical anatomy of the liver
Segmentation of the liver
The intrahepatic bile ducts
The gallbladder
Calot's triangle
The bile ducts
Further reading

Development

The liver and biliary tree develops as a hollow endodermal bud, the hepatic diverticulum, from the distal foregut in the 3-week embryo. The rapidly proliferating cells of the bud penetrate the septum transversum and eventually develop into the liver, while the connection between the hepatic diverticulum and the foregut is preserved to form the bile duct. A ventral outgrowth of the bile duct gives rise to the gallbladder and cystic duct. With the rotation of the gut, the opening from the bile duct into the intestine migrates to a posterior position and the common bile duct comes to lie behind the duodenum and pancreas.

Surgical anatomy of the liver

The external appearance of the mature liver shows its division into two lobes by the umbilical fissure and falciform ligament. Further subdivisions are made on other superficial features. The quadrate lobe is a subdivision of the right lobe and lies to the left of the gallbladder fossa and to the right of the umbilical fissure. The transverse hilar fissure forms the posterior boundary of the quadrate lobe and divides it from the caudate lobe posteriorly (Fig. 1).

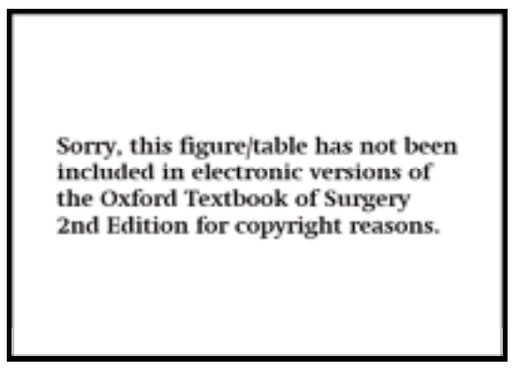


Fig. 1. Divisions of the liver by its external features.

The internal architecture of the liver bears only a superficial relation to its external appearance. Cast studies of the biliary tree and portal venous radicles show that the liver is divided into right and left halves, according to the territories of drainage of the right and left hepatic ducts and the areas of supply of the right and left branches of the portal vein and hepatic artery. This principal division is called Cantlie's line, after its first description in 1898, but it is not readily visible on external examination. It runs from the medial edge of the gallbladder fossa to the inferior vena cava posteriorly. The nomenclature of hepatic anatomy has become confused by the use of the term ‘lobe', which has been applied to both the division of the liver by its external features and the territories of drainage of the right and left hepatic ducts.

Glisson's capsule, a peritoneal and fibrous covering, invests the liver. The reflections of the capsule on to the right hemidiaphragm form the coronary ligament and right triangular ligament, and the reflection from the left liver on to the left hemidiaphragm forms the left triangular ligament. Glisson's capsule is also reflected over the falciform ligament. The structures at the hilum of the liver are invested in dense fibrous tissue continuous with Glisson's capsule; here this covering is known as the hilar plate. The hilar plate is continuous with the peritoneal layers investing the common hepatic and common bile duct, cystic duct, and gallbladder.

The liver is supplied by blood 80 per cent of which comes via the portal vein and 20 per cent via the hepatic artery. Venous drainage is by three, large, short, hepatic veins that pass posteriorly to the inferior vena cava, which lies on the posterior surface of the liver. Drainage of bile occurs from the left and right hepatic ducts to the common hepatic and bile duct, and then to the second part of the duodenum.

The portal vein is formed by the confluence of the superior mesenteric vein and the splenic vein in front of the inferior vena cava and behind the neck of the pancreas. The portal vein runs behind the pancreas to the free border of the lesser omentum, where it traverses to the hilum of the liver in the hepatoduodenal ligament behind the common bile duct and to the right of the hepatic artery. At the hilum of the liver, the portal vein divides into left and right branches. The vein, with its accompanying branches of the biliary tree and hepatic artery, is invested in a fibrous sheath continuous with the hilarplate.

The common hepatic artery usually arises from the coeliac axis and travels across the posterior abdominal wall to lie just above the pylorus. Here it gives off the gastroduodenal artery before continuing as the hepatic artery proper, which then runs in the gastroduodenal ligament medial to the common bile duct and anterior to the portal vein to the hilum of the liver. The hepatic artery divides into the left and right hepatic artery well below the hilum of the liver. Sixteen per cent of individuals have an aberrant right hepatic artery arising from the superior mesenteric artery that runs in the groove to the right of the portal vein and common bile duct. Less commonly, the arterial supply to the left half of the liver comes from the left gastric artery.

The venous radicles in the liver give rise to three hepatic veins, the right, middle, and left, which are short and large. The middle hepatic vein usually joins the left hepatic vein before entering the inferior vena cava. In addition, a number of unnamed short veins enter the inferior vena cava directly. These arise in the caudate lobe, which, because of its embryological development form the dorsal mesogastrium, has a different venous drainage

Segmentation of the liver

The three hepatic veins divide the liver into four sectors, each of which is further subdivided into two segments. The whole liver is therefore divisible into eight segments: four are in the right half, and three in the left half (Fig. 2). The remaining segment is the caudate lobe, which should be considered separately because of its different embryological origin, variable blood supply, and venous drainage. Two differing descriptions of the segmentation of the liver are in common use, that of Couinaud and that of Goldsmith and Woodburne. These differ mainly in nomenclature, and the description of Couinaud will be used here.

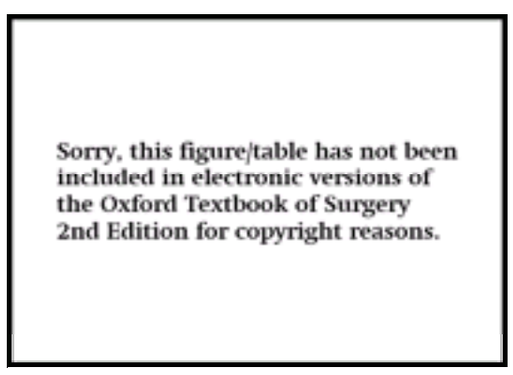


Fig. 2. Schematic representation of the segmentation of the liver.

The segments are numbered anticlockwise I to VIII, starting with the caudate lobe (Fig. 3). Each segment is supplied by a named portal venous radicle and is drained by a segmental bile duct, forming the smallest anatomical unit of hepatic resection. Removal of segments II to IV is described as ‘left hepatectomy' and removal of segments V to VIII, ‘right hepatectomy'. Removal of segment IV (the quadrate lobe) in addition to right hepatectomy is described as extended right hepatectomy. The use of this nomenclature avoids the confusion inherent in the use of the terms ‘hepatic lobectomy' and ‘trisegmentectomy'.

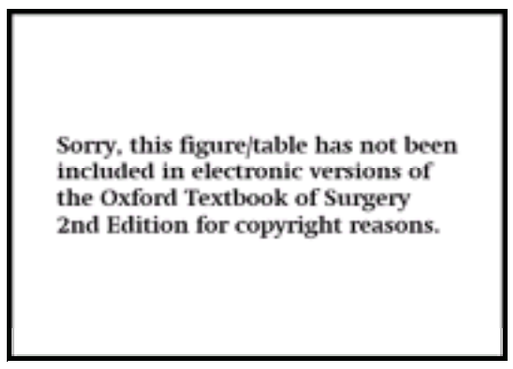


Fig. 3. The segments of the liver, showing the segmental biliary tree and venous drainage.

The intrahepatic bile ducts

The interlobular bile canaliculi join to form segmental bile ducts that eventually drain into the right or left hepatic ducts. On the right, ducts from segments VI and VII join to form the right posterior sectoral duct, which runs horizontally across the gallbladder fossa, where it is surgically accessible after localization by needle puncture or intraoperative ultrasonography. The right anterior sectoral duct runs more vertically and is formed by the confluence of the ducts from segments V and VIII.

Segmental ducts from segments II, III, and IV merge to form the left hepatic duct at the base of the umbilical fissure. Although there are variations in the exact anatomy of this confluence of bile ducts, these are of little clinical relevance. The duct from segment III is surgically accessible by dissection in the groove to the left of the umbilical ligament, where it lies anterior to its accompanying branch of the portal vein and hepatic artery. The left hepatic duct runs from the base of the umbilical fissure to the hilum in the transverse hilar fissure, invested by the fibrous tissue of the hilar plate with the left portal vein lying posterior and the left hepatic artery lying inferior. The left duct is surgically accessible by division of the peritoneal fold under the quadrate lobe (segment IV), a procedure known as lowering the hilar plate.

At the hilum of the liver, the right and left hepatic ducts join to form the confluence of the bile ducts. Anatomical variations of both the intrahepatic and extrahepatic biliary tree are so common that a ‘normal' pattern is seen in less than 60 per cent of individuals (Fig. 4). In 57 per cent, the right anterior and posterior sectoral ducts join to form a right hepatic duct, whereas in the remainder, the right anterior and posterior sectoral ducts join the confluence individually. One important variation is the presence of an anomalous subvesical duct, the duct of Luschka, which runs in the gallbladder fossa. It is found in 12 to 50 per cent of individuals, drains a variable portion of the right liver, and is potentially vulnerable during cholecystectomy.

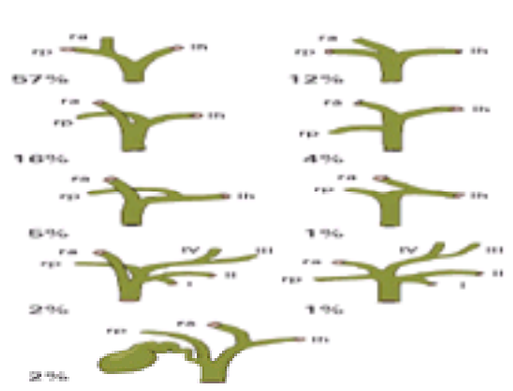


Fig. 4. Anomalies of the confluence of the bile ducts: ra, right anterior sectoral duct; rp, right posterior sectoral duct; lh, left hepatic duct; roman numerals refer to hepatic segments.

The gallbladder

The gallbladder, a pear-shaped reservoir 5 to 12 cm in length, lies in a fossa on the lower surface of the liver. Four parts of the gallbladder are described: the fundus, the body, the infundibulum, and the neck. In addition, a Hartmann's pouch often develops as a pathological feature in the neck and infundibulum of the gallbladder in the presence of gallstones. Various congenital abnormalities have been described, including double, bilobed, and intrahepatic gallbladder, and congenital absence. The occasional presence of a long mesentery is of significance since it may allow torsion. The gallbladder drains by the cystic duct to the junction of the common hepatic duct and common bile duct. The wall of the cystic duct contains muscle fibres that form the sphincter of Lutkens, while the mucosa of the cystic duct forms crescentic folds known as the spiral valve of Heister.

Calot's triangle

Calot's triangle is formed by the common hepatic duct to the left and the cystic duct below. Although the original description of this area gave the cystic artery as the superior border, the inferior surface of the liver is now accepted as this border. The cystic artery usually arises from the right hepatic artery behind the common hepatic duct and runs behind the right hepatic duct and through Calot's triangle to the gallbladder. In 20 per cent of individuals the cystic artery arises from a right hepatic artery that runs anterior to the common hepatic duct, and the right hepatic artery forms a loop or ‘caterpillar hump' with the cystic artery originating from the apex in 7 per cent of individuals. In the latter case, the right hepatic artery may be mistaken for the cystic artery during cholecystectomy. In 10 per cent of individuals the cystic artery arises proximally from the right hepatic artery and runs anterior to the common hepatic duct, while the right hepatic artery runs posterior to this duct.

There are no discrete veins draining the gallbladder and bile duct, although all arteries are normally accompanied by a small vein or venous plexus. Some veins drain directly from the gallbladder into the liver. Lymphatic drainage is first to the cystic lymph node, which is usually seen adjacent to the cystic artery during cholecystectomy, and thence to the retroduodenal lymph nodes. Some lymphatic channels from the fundus drain to the lymphatic channels in the liver capsule. Motor and sensory sympathetic nerves from the coeliac plexus reach the gallbladder along the hepatic artery, and the parasympathetic motor supply comes from the right and left vagus nerves.

Two other major anomalies may be encountered during the course of dissection in Calot's triangle for cholecystectomy. An aberrant right hepatic artery from the superior mesenteric artery occurs in 16 per cent of individuals, running in the groove between the common hepatic duct and the portal vein. It can be seen in the medial border of Calot's triangle in 90 per cent of these individuals. The right posterior or anterior sectoral ducts may also run through Calot's triangle and may be mistaken for the cystic duct.

The bile ducts

From the confluence of the bile ducts, the common hepatic duct runs for some 2.5 to 3.5 cm down to its confluence with the cystic duct, resulting in the formation of the common bile duct. This junction is variable. In 5 per cent of people the cystic duct spirals behind the common hepatic duct and enters the bile duct low down within the pancreas. In 2 per cent the cystic duct opens directly into the confluence of the bile ducts (Fig. 5). The common bile duct is normally 10 to 12 cm in length and about 6 mm in diameter in anatomical specimens. In life, the upper limit of normal measured by ultrasonography is 7 mm, whilst on direct cholangiography, when the duct is deliberately distended, it may be 10 mm in diameter.

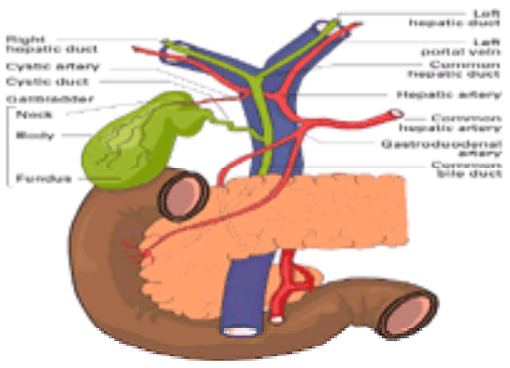


Fig. 5. The relations of the extrahepatic biliary tree.

The common hepatic and bile ducts are supplied by adjacent arteries that supply two axial arteries that run at 3 o'clock and 9 o'clock along the bile duct wall. Other small arteries run in the mesentery around the bile duct to form a plexus. About 60 per cent of blood flow to the duct arises inferiorly from the retroduodenal and gastroduodenal arteries, while 38 per cent comes from the cystic and hepatic arteries superiorly.

The common bile duct passes behind the first part of the duodenum. It may be exposed by division of the peritoneal fold over the superior aspect of the first part of the duodenum and by drawing the duodenum downwards. It then runs either in a groove in the back of the head of the pancreas or in the loose areolar tissue behind the head of the pancreas. Here it may be exposed by Kocher's manoeuvre, that is, division of the peritoneum lateral to the duodenum and reflection of the duodenum and head of the pancreas medially. It curves to the right to enter the medial duodenal wall about 2 cm below the duodenal cap, where it is joined by the main pancreatic duct of Wirsung to form the sphincter of Oddi, which discharges into the duodenum through the ampulla of Vater.

Some 2 cm of the terminal portion of the common bile duct lies within the wall of the duodenum, where it is surrounded by the smooth muscle fibres of the sphincter of Oddi. The pancreatic duct may be closely applied to the common bile duct at this point and may similarly be invested in smooth muscle of the sphincter of Oddi (Fig. 6). The exact anatomy of the terminal common bile duct and pancreatic duct follows one of three patterns. They may unite outside the wall of the duodenum and traverse the duodenal wall to the papilla as a common channel; they may join within the duodenal wall and have a short common terminal channel; and separate orifices have been described.

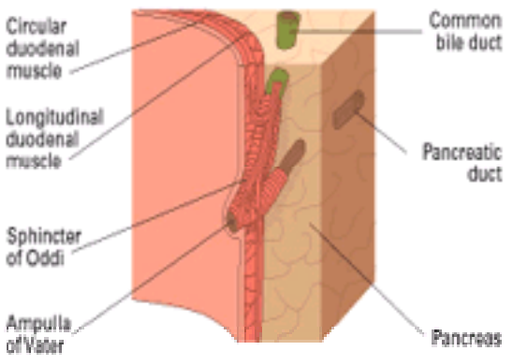


Fig. 6. Anatomy of the sphincter of Oddi.

Further reading

Bismuth H. Surgical anatomy and anatomical surgery of the liver. *World Journal of Surgery* 1982; **6**: 3–9.

Couinaud C. Lobes et segments hépatiques. Notes sur lárchitecture anatomique et chirurgicale du foie. *Presse Medicale*, 1954; **62**: 709–12. [Classic description of hepatic segmentation.]

Goldsmith NA, Woodburne RT. The surgical anatomy pertaining to liver resection. *Surgery, Gynecology and Obstetrics* 1957; **195**: 310–18.

Northover JMA, Terblanche J. A new look at the arterial supply of the bile duct in man and its surgical implications. *British Journal of Surgery* 1979; **66**: 379–84.

Smadja C, Blumgart LH. The biliary tract and the anatomy of biliary exposure. In: Blumgart LH, ed. *Surgery of the liver and biliary tract*, Vol. 1, pp. 11–22. Churchill Livingstone, Edinburgh, 1988.

30.2 Hepatic trauma

David V. Feliciano

Mechanism of injury
Diagnosis
Nonoperative management
General principles of operative management
Simple techniques of hemostasis
Advanced techniques of hemostasis
Extensive hepatorrhaphy
Hepatotomy with selective vascular ligation
Viable omental pack
Resectional debridement with selective vascular ligation
Absorbable mesh compression
Formal resection
Selective hepatic artery ligation
Intrahepatic balloon tamponade
Perihepatic packing
Atriocaval shunt
Drainage
Complications
Mortality
Further reading

Mechanism of injury

Compressive injuries to the liver from the overlying ribs occur most frequently in frontal motor vehicle crashes in which the victim has an impact with the lower rim of the steering wheel or the dashboard. Compression against a shoulder belt restraint may be a cause, as well, particularly if the device is worn improperly under the right upper extremity. In ‘T-bone' side impacts, the front seat passenger is at significant risk for a hepatic injury.

Patients with penetrating wounds to the right thoracoabdominal area (nipples to costal margin and medial to right anterior axillary line) are at risk of a hepatic injury if the diaphragm is penetrated. This occurs in approximately 15 per cent of patients with penetration of the body wall by a stab wound and in 45 to 48 per cent of those with gunshot wounds.

Diagnosis

In hypotensive patients who have suffered blunt abdominal or multisystem trauma, either surgeon-performed ultrasound or a standard infraumbilical diagnostic peritoneal lavage is appropriate. Using a 3.5 mHz transducer in the right midaxillary line between ribs 10 and 11, the visualization of fluid (blood unless ascites is present) in Morison's pouch mandates a laparotomy in the absence of other overt sites of hemorrhage. An experienced surgeon-sonographer may visualize a hepatic injury, also. When no fluid is present in Morison's pouch, the ultrasound probe is moved to image the left subphrenic area/splenorenal recess and the pelvis. A diagnostic peritoneal tap that yields 10 to 20 ml of gross blood or a formal lavage whose effluent is cloudy enough to obscure the print on the bag of intravenous fluids mandates laparotomy in the hypotensive patient, also. In any patient undergoing emergency laparotomy after suffering blunt abdominal trauma, the most likely sources of hemorrhage are injuries to the liver, spleen, or mesentery.

A patient who is hemodynamically stable and without peritonitis after suffering blunt abdominal trauma is evaluated by a spiral contrast CT if the physical examination is equivocal or compromised or if there is intra-abdominal fluid on the preliminary ultrasound. The volume of intraperitoneal fluid (blood), magnitude of injury to the liver or other organ, and the presence or absence of active hemorrhage on the contrast CT will determine whether nonoperative or operative management is chosen in the stable patient.

Penetrating wounds to the abdomen in patients with peritonitis, hypotension, or significant evisceration mandate laparotomy. Stab wounds to the right thoracoabdominal area in patients without fluid in the right subphrenic space or Morison's pouch on ultrasound undergo in-hospital serial physical examinations for 24 h after admission. An occasional stable patient with a gunshot wound to this area and minimal tenderness may be evaluated by a contrast spiral CT to determine the magnitude of hepatic and pulmonary injuries.

Nonoperative management

Approximately 80 to 85 per cent of all patients with hepatic trauma are stable upon arrival in the emergency center, and, in the absence of other indications for an emergency laparotomy, nonoperative management is appropriate after a contrast spiral CT. Contraindications to nonoperative management include a period of hypotension in the field or in the emergency center, persistent significant tachycardia despite aggressive resuscitation, the presence of active hemorrhage from the liver, spleen, or kidney on the contrast CT, or the presence of another organ injury mandating laparotomy. Patients are kept at bed rest, and their vital signs are monitored in the surgical intensive care unit if a significant injury is present (American Association for the Surgery of Trauma—Organ Injury Scale Grades III, IV, V) (Table 1). A falling hematocrit or continuing need for transfusion during the nonoperative period should prompt an emergency hepatic arteriogram or laparotomy (Fig. 1 and Fig. 2). New onset peritonitis and hypotension are followed by an emergency laparotomy. In stable patients a repeat spiral CT is appropriate at 5 to 7 days following injury to determine whether progression of the injury or some healing has occurred. With some healing, discharge to a home situation in which a family member is available to the patient is indicated if a Grade III, IV, or V injury was present. Return to vigorous physical activity or contact sports is prohibited until a late follow-up spiral CT shows healing.

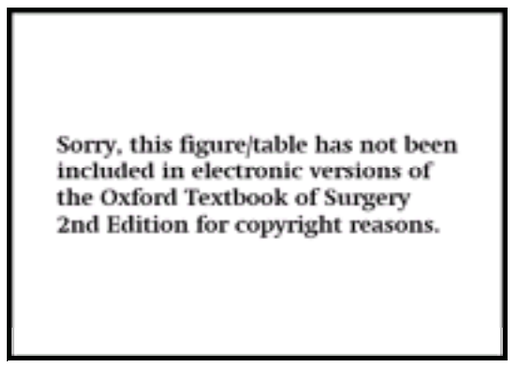


Table 1 Liver injury scale (1994 Revision), American Association for the Surgery of Trauma*



Fig. 1. Small subcapsular hematoma in posterior aspect of right lobe of liver was managed nonoperatively.

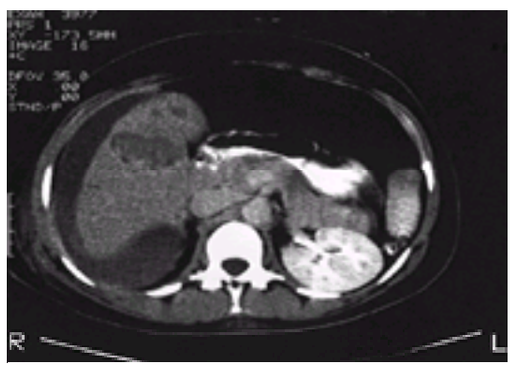


Fig. 2. Same patient as in Fig. 1. On the sixth day of nonoperative management, a repeat abdominal CT demonstrated enlargement of the subcapsular hematoma and extension around the right lobe. Hepatic arteriography and selective embolization was used to control a bleeding vessel.

Nonoperative management fails in approximately 2 to 7 per cent of patients with blunt hepatic injuries. The hepatic injury, itself, will be the cause in 50 to 75 per cent of the failures, and 65 to 85 per cent of the hepatic failures will be in patients with Grade IV or V injuries on the original CT.

Nonoperative management of gunshot wounds of the liver is practiced in a similar fashion. The success rate is similar to that described above for blunt trauma as missile tracks from civilian handguns are significantly smaller than many of the Grade IV or V hepatic injuries presently undergoing nonoperative management.

General principles of operative management

A midline incision is used, and blood and clots are evacuated manually or with a suction device. A vascular clamp is applied to the porta hepatis (Pringle maneuver) if a significant (Grade III, IV, V) hepatic injury is present. The injured lobe is compressed between laparotomy pads in the hands as the surgeon informs the anesthesiologist about the need to contact the blood bank. Also, the surgeon should request that an upper hand retractor, various sizes of metal clips, O-chromic sutures on blunt needles, and a 36–38 French thoracostomy tube be available in the operating room. When blood and appropriate equipment is available in the operating room, the packs around the liver are removed and the hepatic injury is inspected. Posterior lobar injuries or Grade III, IV, or V injuries are best visualized by division of the ipsilateral triangular ligament and the anterior coronary ligament at the edge of the liver (Fig. 3). Folded dry laparotomy pads are then placed beneath the injured lobe to elevate it into the midline incision. In obese patients or in those with a high likelihood of an injury to the extrahepatic veins or retrohepatic vena cava (dark venous hemorrhage as the injured lobe is mobilized), a median sternotomy is also performed.

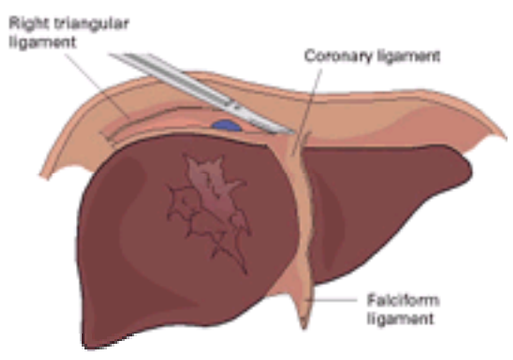


Fig. 3. Mobilization of the right lobe of the liver by division of the triangular and anterior coronary ligaments. (Reproduced with permission from Baylor College of Medicine.)

Simple techniques of hemostasis

Approximately 90 per cent of penetrating injuries and 60 per cent of blunt injuries can be managed with 5 min of compression, the application of topical hemostatic agents, or simple suture hepatorrhaphy. Currently available topical hemostatic agents include oxidized regenerated cellulose, microfibrillar collagen hemostat, and fibrin sealant. Fibrin sealant, only recently available in the United States, contains human fibrinogen and thrombin, aprotinin, and calcium chloride. Five minutes of compression is performed after the application of a topical agent. After releasing compression, the electrocautery is used for any remaining bleeders when only Grade I or Grade II hepatic injuries are present. Suture hepatorrhaphy with O-chromic material is appropriate for Grade II and Grade III injuries. An interrupted or continuous suture technique is used, with the caveat that crushing sutures cause postoperative hepatic necrosis and ‘liver fever'. Drainage is not necessary in the absence of further hemorrhage or obvious leakage of bile.

Advanced techniques of hemostasis

Advanced techniques are necessary in 10 per cent of penetrating wounds and in 40 per cent of blunt hepatic injuries. These patients have Grade III, IV, or V injuries that will require the use of one or more of the following techniques: (1) extensive hepatorrhaphy; (2) hepatotomy with selective vascular ligation; (3) viable omental pack; (4) resectional debridement with selective vascular ligation; (5) absorbable mesh compression; (6) formal resection; (7) selective hepatic artery ligation; (8) intrahepatic balloon tamponade; (9) perihepatic packing; and (10) atriocaval shunt. The most important adjunct to these techniques is avoiding intraoperative hypothermia using the maneuvers listed in Table 2.

Increase operating room temperature above 85°F
Infuse crystalloid solutions and blood through a warming device such as the Level 1 Fluid Warmer*
Cover patient's head with a radiant or warming device
Cover body parts out of the operative field with a warming device such as the Bair Hugger† or Life-Aid‡ system
Irrigate nasogastric and thoracostomy tubes with warm saline during laparotomy
Irrigate open pericardial cavity, pleural cavities, and peritoneal cavity during simultaneous sternotomy or thoracotomy and laparotomy

*Level 1 Technologies Inc., Rockland, MA.
†Augustine Medical, Inc., Eden Prairie, MN.
‡Allied Healthcare Products Group, Marshall, MN.

Table 2 Maneuvers to prevent or reverse hypothermia during ‘damage control operations'

Extensive hepatorrhaphy

Extensive hepatorrhaphy is indicated in ‘damage control' situations in which intraoperative hypothermia (<34–35°C), metabolic acidosis (pH < 7.1–7.2), and/or a coagulopathy (PT or PTT > 50 per cent normal) mandate a rapid operation. Large figure-of-eight sutures or a continuous O-chromic suture is used to reapproximate the sides of hepatic lacerations in the hope that hemorrhage from small hepatic arteries and low pressure hepatic veins or portal veins will be controlled by compression. Extensive postoperative hepatic necrosis is likely when such sutures are tied too tight in the presence of a prolonged Pringle maneuver.

Hepatotomy with selective vascular ligation

Gaining further exposure of a deep hepatic laceration or connecting the entrance and exit wounds of a penetrating wound with the finger fracture technique or the electrocautery is known as hepatotomy. Once completed, large Deaver or Harrington retractors are used to maintain visibility in the depths of the hepatotomy as selective vascular clipping or suture ligation of injured vessels is performed. This technique should be utilized prior to the onset of hypothermia and only by surgeons with sufficient experience in elective or traumatic hepatic surgery.

Viable omental pack

The gastrocolic omentum mobilized off the transverse colon with its blood supply intact is used to fill Grade III, IV, or V hepatic injuries or hepatotomy sites (Fig. 4). Intrahepatic omentum is effective in controlling venous hemorrhage, managing dead space, and in bringing mobile macrophages to the site of injury. While it does not appear to aid healing, postoperative bleeding and drainage of bile are much decreased in the experience of most trauma surgeons. The viable omental pedicle is held in place by compressing hepatic sutures tied under moderate tension.

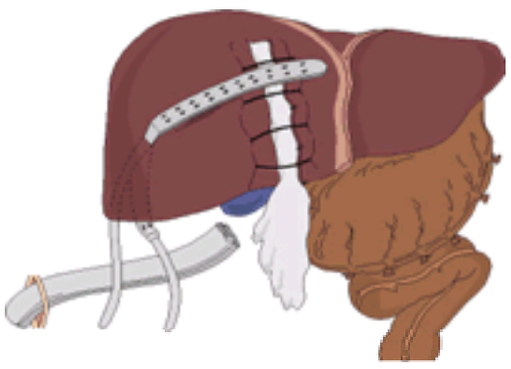


Fig. 4. Viable omental pack used to fill hepatic laceration after selective vascular ligation. (Reproduced with permission from Baylor College of Medicine.)

Resectional debridement with selective vascular ligation

With disrupted hepatic tissue on the edge of an injured liver, the finger fracture technique or the electrocautery should be used to create a new fresh edge of the liver around the area of injury. Vessels and biliary ducts can then e clipped or suture ligated where they are intact, and all disrupted tissue outside this new line is then debrided (Fig. 5). The application of a viable omental pedicle to this new raw surface is controversial, though this is appropriate when a coagulopathy makes hemostasis difficult.



Fig. 5. Resectional debridement of portion of segments II and III of liver with selective vascular ligation using clips and sutures.

Absorbable mesh compression

Wrapping an injured hepatic lobe in which all fragments are viable with a large sheet of absorbable mesh tailored around the porta hepatis and inferior vena cava has been used in some centers. The technique is time-consuming, but eliminates the need for reoperation as when perihepatic packs are used for compression.

Formal resection

Anatomic lobectomy is used in approximately 3 per cent of patients undergoing operative management. No dissection is performed in the porta hepatis, and the lobectomy is performed with a Pringle maneuver in place using finger fracture or electrocautery and metal clips. The large right hepatic vein can usually be controlled inside the liver as the lobectomy is completed. Anatomic segmentectomy is much more commonly utilized, especially with extensive lacerations beneath the falciform ligament mandating resection of Couinaud's segments II and III (left lateral segment) (Fig. 6).



Fig. 6. Resection of segments II and III after avulsion injury beneath falciform ligament.

Selective hepatic artery ligation

Selective hepatic artery ligation is used in about 1 per cent of patients undergoing operative management. It is indicated when arterial hemorrhage in a deep hepatic laceration cannot be directly controlled, but stops whenever a Pringle maneuver is applied. Extrahepatic ligation of the artery to the injured lobe in the porta hepatis will fail to control hemorrhage when the wrong artery is ligated or when intrahepatic or retrohepatic venous hemorrhage is present.

Intrahepatic balloon tamponade

The passage of a Foley or Fogarty balloon catheter into the hepatic track of a knife or missile may allow for balloon compression of the site of parenchymal hemorrhage. This technique is particularly useful when the novice trauma surgeon has little experience in completing an extensive hepatotomy through one or both lobes. The inflated balloon catheter is passed through the body wall away from the midline incision at the completion of the first laparotomy. After 48 to 72 h of balloon compression, the balloon is eflated and removed through the body wall in the surgical intensive care unit. Rebleeding is extraordinarily rare when a parenchymal track has been tamponaded for this period of time.

Perihepatic packing

The insertion of folded dry laparotomy pads over and, occasionally, below an injured hepatic lobe is used in approximately 5 per cent of patients undergoing operative management (Fig. 7). Packs should be used to tamponade minor hepatic injuries or subcapsular hematomas when a damage control procedure is performed. They are also useful for any major hepatic parenchymal injury when advanced techniques of hemostasis fail secondary to intraoperative hypothermia or a coagulopathy. The use of a plastic sheet beneath the packs to prevent sticking to raw edges of parenchyma has been useful in the author's experience. Packs are removed at a reoperation 48 to 72 h after the original laparotomy when hypothermia, acidosis, and any coagulopathy are corrected and the cardiovascular, respiratory, and renal systems are stable. Perihepatic packs have also been used with success in patients with unruptured retrohepatic hematomas from presumed injuries to the retrohepatic vena cava.

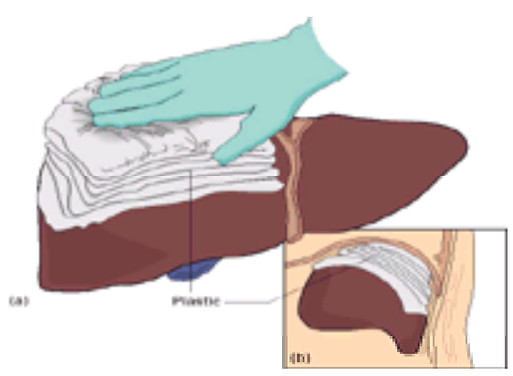


Fig. 7. Perihepatic packing with folded dry laparotomy pads over plastic sheet (Reproduced with permission from Baylor College of Medicine.)

Atriocaval shunt

A no. 36 French thoracostomy tube or no. 8 endotracheal tube inserted through the right atrial appendage into the infrarenal inferior vena cava is an atriocaval shunt. An extra hole at the level of the right atrium is cut before insertion. By pulling circumferential umbilical tape tourniquets tight around the shunt at the suprarenal inferior vena cava and intrapericardial inferior vena cava, venous return from the lower body and renal veins is diverted into the shunt (Fig. 8 and Fig. 9). This causes a 40 to 60 per cent decrease in hemorrhage from an injury in the retrohepatic vena cava and should allow for a rapid repair.

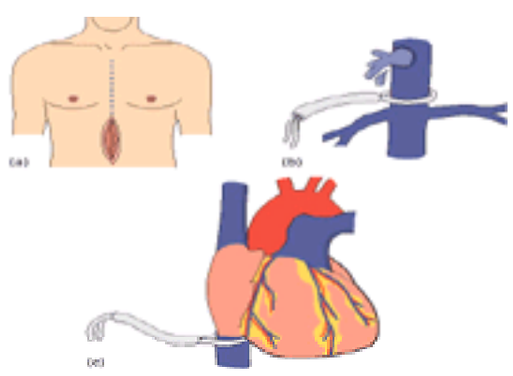


Fig. 8. After decision to insert atriocaval shunt is made, a median sternotomy is performed. Also, the suprarenal infrahepatic inferior vena cava (B) and intrapericardial inferior vena cava (C) are looped with umbilical tapes to act as tourniquets. (Reproduced with permission from Baylor College of Medicine.)

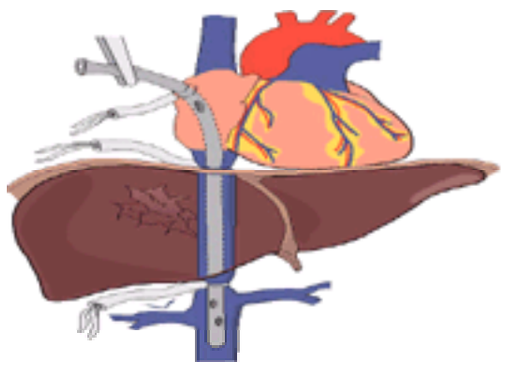


Fig. 9. A no. 36 French thoracostomy tube is inserted through right atrial appendage to act as atriocaval shunt. Blood in the infrahepatic inferior vena cava is diverted

into the shunt by tightening the umbilical tape tourniquets. (Reproduced with permission from Baylor College of Medicine.)

When there has not been a preoperative or intraoperative cardiac arrest from exsanguination, use of the atrocaval shunt has resulted in a 33 to 50 per cent survival in the modern era. Alternative approaches for injuries to the retrohepatic vena cava include direct approach behind an injured lobe, total hepatic vascular isolation, and deep hepatotomy.

Drainage

Closed suction drains above and below an injured lobe are used when an intraoperative coagulopathy or the extent of hepatic repair suggests that postoperative drainage of blood and bile is likely (Fig. 4).

Complications

Postoperative hyperpyrexia occurred in nearly two-thirds of patients with Grade III, IV, or V injuries in one review. Early postoperative coagulopathies occur in 15 per cent of patients, while reoperations for persistent or late hemorrhage used to be necessary in 3 to 7 per cent of patients. Self-limited biliary fistulas occur in 8 to 10 per cent of patients, while intra-abdominal abscesses develop in 4 to 10 per cent.

The embolization of disrupted intrahepatic arteries or pseudoaneurysms to control postoperative bleeding or late hemobilia by the interventional radiologist has caused a significant decrease in reoperations for hemorrhage. Reoperations for perihepatic abscesses have essentially disappeared in the modern era for the same reason (Fig. 10).

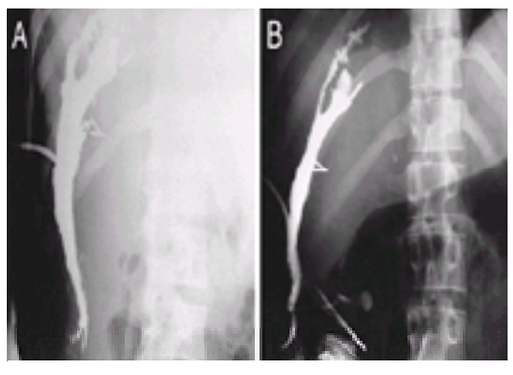


Fig. 10. Percutaneous drainage of right subphrenic abscess (A) led to significant decrease in size over 7 days (B).

Mortality

The liver-related mortality for patients undergoing nonoperative management has been 0 to 0.5 per cent in recent reports. Overall mortality in such patients is 8 to 9 per cent and is primarily due to associated intracranial injuries.

In patients undergoing operation for blunt hepatic injuries, the overall mortality is 15 to 20 per cent. The mortality after operation for stab wounds and gunshot wounds to the liver is 2.5 and 10 per cent, respectively.

Further reading

Burch JM, Feliciano DV, Mattox KL. The atriocaval shunt. Facts and fiction. *Annals of Surgery* 1988; **207**: 555–68.

Cogbill TH, Moore EE, Jurkovich GJ, Feliciano DV, Morris JA, Mucha P. Severe hepatic trauma: a multi-center experience with 1,335 liver injuries. *Journal of Trauma* 1988; **28**: 1433–8.

Cohn SM, Cross JH, Ivy ME, Feinstein AJ, Samotowka MA. Fibrin glue terminates massive bleeding after complex hepatic injury. *Journal of Trauma* 1998; **45**: 666–72.

Feliciano DV, Mattox KL, Burch JM, Bitondo CG, Jordan GL Jr. Packing for control of hepatic hemorrhage. *Journal of Trauma* 1986; **26**: 738–43.

Feliciano DV, Mattox KL, Jordan GL Jr, Burch JM, Bitondo CG, Cruse PA. Management of 1000 consecutive cases of hepatic trauma (1979–1984*). Annals of Surgery* 1986; **204**: 438–45.

Feliciano DV, Pachter HL. Hepatic trauma revisited. *Current Problems in Surgery* 1989; **26**: 453–524.

Feliciano DV. Continuing evolution in the approach to severe liver trauma. *Annals of Surgery* 1992; **216**: 521–3.

Feliciano DV, Pachter HL. Trauma to the liver vasculature, aneurysm and arteriovenous fistula. In: Blumgart LH, ed. *Surgery of the liver and biliary tract*. Edinburgh: Churchill Livingstone, 1994:1243–57.

Pachter HL, Hofstetter SR. The current status of nonoperative management of adult blunt hepatic injuries. *American Journal of Surgery* 1995; **169**: 442–54.

Pachter HL, Liang HG, Hofstetter SR. Liver and biliary tract trauma. In: Mattox KL, Feliciano DV, Moore EE, eds. *Trauma.* New York: McGraw-Hill, 2000: 633–80.

Pachter HL, *et al.* Status of nonoperative management of blunt hepatic injuries in 1995: a multicenter experience with 404 patients. *Journal of Trauma* 1996; **40**: 31–8.

Pachter HL, Spencer FC, Hofstetter SR, Liang HC, Coppa GF. The management of juxtahepatic venous injuries without an atriocaval shunt: Preliminary clinical observations. *Surgery* 1986; **99**: 569–75.

Pachter HL, Spencer FC, Hofstetter SR, Liang HG, Coppa GF. Significant trends in the treatment of hepatic trauma. Experience with 411 injuries. *Annals of Surgery* 1992; **215**: 492–502.

Stone HH, Lamb JM. Use of pedicled omentum as an autogenous pack for control of hemorrhage in major injuries of the liver. *Surgery, Gynecology and Obstetrics* 1975; **141**: 92–4.

30.3 Abscesses—pyogenic and amoebic

Robert H. Rubin

[Pyogenic abscesses](#)
[Epidemiology](#)
[Etiology and pathogenesis](#)
[Bacteriology](#)
[Clinical presentation](#)
[Diagnostic evaluation](#)
[Treatment](#)
[Amebic abscesses](#)
[Epidemiology](#)
[Etiology and pathogenesis](#)
[Clinical presentation](#)
[Diagnostic evaluation](#)
[Treatment](#)
[Further reading](#)

Pyogenic abscesses

Epidemiology

The incidence of pyogenic liver abscess in developed countries has not changed appreciably over the past 50 years, being estimated at 8 to 16 cases/100 000 admissions, with a prevalence at autopsy of 0.3 to 1.5 per cent. A slight male predominance of cases has remained during this period, and no ethnic group appears to be at increased risk (Table 1).

	Pyogenic abscesses	*Amebic abscesses*
Incidence/100 000 hospital admissions		
USA	10	2
India	30	90
Change in incidence, 1900s compared with present	None	Increase
Age (years)	43–60	28–48
Sex ratio (male:female)	1.4:1	Up to 19:1
Risk factors	Biliary tract disease	Immigrants and travelers from endemic regions
	Malignancy	Homosexual males
	Diabetes mellitus	Institutionalized patients
	Cirrhosis	
	Immunosuppressed patients	

Table 1 Liver abscesses—epidemiology

What has changed over this time is the age of the patients, the underlying cause of the liver infection, and the increasing role of malignancy in the pathogenesis of the abscess. Fifty years ago, the majority of patients were under the age of 40 years and appendicitis was the leading cause of the disease; today the average age is between 43 and 60 years, with an increasing proportion of patients over the age of 60. This change corresponds to the finding that appendicitis has been replaced by biliary tract disease as the most common underlying etiology. The increasing age of patients has been associated with an increased incidence of malignancy in patients with liver abscesses, with malignancy playing a role in a number of ways: tumors invading or compressing the biliary tract; secondary infection of primary or metastatic tumors within the liver; and gastric and intestinal malignancies providing a portal of entry for infection to gain access to the liver. Indeed, with the improvement in management of these patient that has occurred over the past 50 years, the presence or absence of underlying cancer is now the most important prognostic determinant.

Etiology and pathogenesis

Pyogenic liver abscesses may be divided into two general categories, based upon the size and distribution of the focal sites of inflammation, the acuity of clinical presentation, and the nature of the therapy that is required. Macroscopic abscesses are usually restricted to one lobe of the liver, are frequently single or confluent, present subacutely with symptoms of several days to weeks' duration, and require some form of primary drainage. Microscopic abscesses are multiple, widely distributed throughout the hepatic parenchyma, usually manifest themselves acutely over a few days, and require primarily medical therapy, with any surgery that is carried out being aimed at the underlying process, rather than the hepatic parenchymal inflammation.

Focal infection within the liver can be divided into six general categories, based upon the pathogenetic route by which infecting organisms were introduced into the liver (Table 2).

Biliary tract disease
Portal vein pylephlebitis
Hepatic arterial infection, secondary to bacteremia
Post-traumatic, from blunt or penetrating injuries
Direct extension from a contiguous source of infection
Miscellaneous (including cryptogenic)

Table 2 Etiology of pyogenic liver abscesses

Biliary tract disease

Hepatic abscess may arise due to cholangitis whenever bile flow is obstructed. In general, total obstruction is associated with elevated pressure within the biliary tree, an acute septic course, and miliary microabscesses throughout the hepatic parenchyma—a process that has been termed ‘acute suppurative cholangitis'. Infection associated with less complete obstruction is associated with normal biliary tract pressure, a subacute course, and macroscopic abscesses. In older patients, malignancy of the liver, biliary tree, or pancreas is the most common cause of biliary obstruction, with a resulting dismal prognosis for such patients.

Portal vein pylephlebitis

Liver abscess may arise because of suppurative thrombophlebitis in the portal venous system that is secondary to such intra-abdominal inflammatory processes as

appendicitis (the classical cause), diverticulitis, infected hemorrhoids, or any other cause of intra-abdominal or pelvic infection that impacts upon the portal venous system. In the older population, colonic malignancy is particularly common as the portal of entry, often with a clinical presentation of diverticulitis. A clinical curiosity which remains unexplained is that although portal vein bacteremia is common in patients with inflammatory bowel disease, hepatic abscess is uncommon.

Hepatic arterial infection

Two subcategories of liver abscess should be considered. Systemic bacteremias may occasionally seed the liver, resulting in either macro-scopic abscesses (usually in the setting of some form of antimicrobial therapy that permits the individual to survive, although not to be cured of the systemic process) or, more commonly, miliary microabscesses. Typically, these infections occur in children with such underlying conditions as chronic granulomatous disease, leukemia, or other disturbances of granulocyte number or function, and are caused by such organisms as *Staphylococcus aureus*. In other patients the primary process affecting the hepatic artery is thrombosis, with hepatic injury then becoming superinfected with micro-organisms of local or systemic origin. With the increasing prevalence of liver transplantation, particularly in young children in whom the hepatic arterial anastomosis is especially vulnerable, this entity is becoming more common.

In recent years there have been increasing attempts to embolize hepatic tumors as part of treatment programs. Secondary infection of the resulting infarcted tumor and surrounding tissue can occur.

Post-traumatic

Both penetrating and non-penetrating trauma to the liver can result in liver abscess formation. The common denominator is hepatic necrosis, intrahepatic hemorrhage, and intraparenchymal bile extravasation. Such areas of devitalized tissue commonly become infected, even if they are initially sterile, resulting in macroscopic abscesses. Prevention of this form of hepatic abscess is dependent upon an aggressive surgical approach to devitalized hepatic tissue, wide excision of such necrotic tissue being essential.

Direct extension of infection

Contiguous sites of infection involving the gallbladder, subphrenic space, or pleural space and disease processes in which gallbladder, gastric, or intestinal perforation occur directly into the liver can result in macroscopic hepatic abscesses. Malignancy is commonly found in these tissues as the cause of the perforation that results in liver abscess formation.

Miscellaneous causes

The etiology of the liver abscess remains obscure in about 5 per cent of patients, even after extensive evaluation. Presumably, a minor injury to the liver renders such individuals susceptible to seeding by a transient bacteremia. Other unusual causes of liver disease, such as cysts (including polycystic liver disease), intrahepatic malignancy (including hepatic injury from cryosurgery of tumors), amebic abscesses, and hydatid disease, may become secondarily infected, resulting in a pyogenic liver abscess. The common denominator, as in the post-traumatic cases, is the microbial seeding of a *locus minoris resistentiae* within the liver.

Bacteriology

The majority of the organisms that invade the liver to cause hepatic abscesses are derived from the gastrointestinal tract: gastrointestinal flora account for more than 75 per cent of these abscesses (Table 3). Polymicrobial infection occurs in 22 to 64 per cent of cases, usually involving both facultative and anaerobic flora of gastrointestinal origin. *Escherichia coli* is the most common facultative organism isolated from liver abscesses, being demonstrated in specimens from 35 to 45 per cent of patients, with *Klebsiella pneumoniae* being the second most frequent isolate in this category. In general, infection with *K. pneumoniae* tends to be more severe, be associated with gas formation in the involved liver or portal vein, have concomitant infection at such sites as the eye (endophthalmitis), and be particularly common in diabetic individuals. Other facultative or aerobic Gram-negative bacteria, including *Proteus* spp., *Enterobacter cloacae, Citrobacter* spp., *Pseudomonas aeruginosa, Morganella morganii, Serratia marsecens*, and *Acinetobacter* and *Eikenella* spp. may also be isolated, usually in association with other gut flora. These are particularly common in patients with biliary tract disease.

	Percentage of all cultures
Single organism	22–64
Enteric Gram-negative bacteria	
Escherichia coli	35–45
Klebsiella pneumoniae	8–27
Others	8–40
Anaerobic bacteria	
Bacteroides fragilis	Up to 60
Others	2–36
Gram-positive bacteria	
Staphylococcus aureus	Up to 25
Streptococcus pyogenes group A	2–3

Table 3 Bacteriology of pyogenic liver abscesses

Anaerobic and microaerophilic organisms, either alone or in conjunction with aerobic organisms, are isolated from up to 60 per cent of pyogenic liver abscesses. *Bacteroides fragilis* (particularly sp. *fragilis)* is the most common anaerobe isolated, but such others as other *Bacteroides* spp., *Fusobacterium* spp., anaerobic streptococci, *Clostridium* spp., and *Actinomyces* spp. may be found on occasion. An important group of organisms are the microaerophilic streptococci, particularly *Streptococcus milleri*. If appropriate microbiologic techniques are employed (especially the provision of an environment enriched with carbon dioxide), microaerophilic streptococci may be found to be the most common causes of pyogenic liver abscess. *S. milleri* is especially virulent and likely to cause suppuration of the liver and other organs, even in the absence of the usual predisposing anatomic abnormalities commonly associated with hepatic abscess formation.

Other Gram-positive organisms account for less than 25 per cent of isolates. *Staphylococcus aureus* and group A streptococci occur most commonly after trauma and in children with the previously delineated granulocyte disorders. Seeding of the liver in these individuals usually is secondary to a systemic bacteremia.

Focal candidal infection of the liver and/or spleen has been reported in an increasing number and variety of patients, most notably in those undergoing chemotherapy for leukemia or liver transplantation. In the patient with leukemia, in particular, a subacute-chronic entity that has been termed hepatosplenic candidiasis has been defined. This is characterized by persistent fevers and macroscopic abscesses due to *Candida* spp., most notably *C. albicans* and *C. tropicalis*. This process is usually initiated when the patient is neutropenic due to chemotherapy, presumably due to the entrance of the yeast into the portal vein through gut mucosal ulcerations induced by the chemotherapy; alternatively, systemic candidemia from infected intravascular access devices could have the same result. The abscesses and the clinical symptoms persist even after hematological remission has been achieved, and require prolonged antifungal therapy.

A variety of organisms, once they reach the bloodstream, can seed the liver, becoming unusual causes of hepatic abscess. The most frequent are those whose portal of entry is the gastrointestinal tract, most commonly *Salmonella* spp., but also including *Yersinia enterocolitica, Campylobacter jejuni, Listeria monocytogenes*, and, rarely, *Brucella* spp. Rarely, mycobacterial infection can present as a liver abscess. Finally, it is important to point out that in developing areas of the world, such parasitic infections as schistosomiasis, ascariasis, and toxocariasis can cause liver injury sufficient to enhance the localization of bacteria, and increase the risk of liver abscess.

Clinical presentation

Patients with microscopic liver abscesses usually have an acutely septic clinical presentation, with fever, rigors, and, not uncommonly, hypotension, as well as right upper quadrant discomfort that can be quite severe. Other manifestations depend upon the underlying condition producing the microabscesses: rapidly progressing jaundice in the presence of biliary tract disease, congestive heart failure if the systemic sepsis is associated with endocarditis.

In contrast, the clinical presentation of macroscopic liver abscesses is more subacute, developing over several days to weeks, with fever, night sweats, anorexia,

weight loss, and malaise far more common than rigors and hypotension. Fever is present in 90 per cent of these patients with macroscopic liver abscesses; nausea, vomiting, and abdominal pain occur in 50 to 75 per cent; and symptoms such as pleurisy, diarrhea, dyspnea, and cough are seen in 5 to 25 per cent of patients. Uncommonly, intra-abdominal rupture of the liver abscess will occur, transforming the illness into an acute one, characterized by manifestations of septic shock, peritonitis, and increasing jaundice.

Other than fever, abdominal tenderness, usually localized to the right upper quadrant, is the most common physical finding, being demonstrable in 50 to 75 per cent of affected individuals. Hepatomegaly is demonstrable in approximately 50 per cent of patients with macroscopic liver abscesses. Jaundice is uncommon, unless biliary obstruction is present.

Almost all patients with pyogenic hepatic abscesses have abnormal hematologic and liver function tests. Leukocytosis, usually of a moderate extent, is noted in 70 to 80 per cent, an elevated erythrocyte sedimentation rate in at least 90 per cent, and anemia in 50 to 65 per cent of patients with liver abscess. The most characteristic liver function test abnormality observed in patients with hepatic abscess is an elevated alkaline phosphatase level, which is observed in more than 75 per cent of these individuals. An elevated serum bilirubin level is seen in 40 per cent of patients, with elevated aminotransferases being found in approximately 30 per cent. Other abnormalities that are not uncommon include a prolonged prothrombin time and a raised serum vitamin B_{12} level. Laboratory abnormalities associated with a poor prognosis include an elevated bilirubin and a serum albumin level of 2 g/dl or less.

Positive blood cultures are obtained in approximately 50 per cent of all patients. Although up to 65 per cent of abscesses are polymicrobial when aspirated pus is cultured from the liver, it is unusual to retrieve more than one organism from the blood cultures.

Chest radiographs are abnormal in about 50 per cent of patients, usually because the inflammatory process within the liver impinges on the diaphragm, producing a variety of ‘sympathetic' responses. These include a right-sided pleural effusion, right lower lobe atelectasis and pneumonitis, and elevation of the right hemidiaphragm. Occasionally, if a gas-forming organism is present in the abscess, air-fluid levels are discernible on chest or abdominal radiographs. Rarely, a liver abscess presents by discharging itself into the chest, with both clinical symptoms (cough, chest pain, hemoptysis, dyspnea) and chest radiographic appearances which reflect this.

Diagnostic evaluation

Clinical suspicion of a liver abscess (either pyogenic or amebic) is aroused when a patient has fever, right upper quadrant abdominal pain and tenderness, and abnormal liver function tests (Fig. 1). The differential diagnosis includes acute cholecystitis, cholangitis, subphrenic or subhepatic abscess, malignancy, and hepatitis.



Fig. 1. Diagnostic scheme for suspected liver abscess.

Ultrasonography is the initial procedure of choice to assess a suspected liver abscess because it is non-invasive, 80 to 90 per cent accurate, and capable of delineating liver lesions as small as 2 cm in diameter. Ultrasound is more useful than computed tomography (**CT**) for distinguishing solid masses from cystic lesions. However, ultrasonography may miss lesions in the dome of the right liver lobe or multiple microscopic abscesses. Fatty infiltration can produce an echogenic liver, making detection of small abscesses difficult. Although some features of amebic abscesses differ on ultrasound from those of pyogenic origin, the differences are not sufficient to permit a specific diagnosis.

Abdominal CT can detect intrahepatic collections as small as 0.5 cm in diameter and can be particularly useful in identifying multiple small abscesses or abscesses located near the hemidiaphragm. The diagnostic accuracy of CT is 90 to 95 per cent. Another advantage of CT is that it may identify other abdominal pathology responsible for the pyogenic liver abscess.

Although radionuclide scans of the liver have a sensitivity comparable with that of ultrasonography in detecting liver abscesses, nuclide scans have largely been replaced by ultrasonography and by CT, at least in part because either sonography or CT allows the clinician to proceed directly to percutaneous aspiration for either diagnosis or therapy. Any material obtained by aspiration should be examined microscopically after Gram staining, cultured aerobically and anaerobically, and, if there is any suspicion clinically or epidemiologically, submitted for examination for *Entamoeba histolytica* trophozoites. Fungal and mycobacterial cultures should also be carried out, particularly in immunosuppressed patients.

Treatment

The traditional approach to the therapy of pyogenic liver abscesses has been open surgical drainage of the abscess, correction, whenever possible, of the underlying pathology that led to the abscess, and a 4- to 6-week course of parenteral antibiotics (Fig. 2). Such antibiotics have usually included a b-lactam, an aminoglycoside, and either metronidazole or clindamycin (aimed at the anaerobic organisms), but treatment could be modified on the basis of microbiologic results. In recent years, such drugs as pipercillin-tazobactam, imipenem, and meropenem have largely replaced aminoglycosides in the therapy of these patients. Over the past decade, the majority of patients with macroscopic pyogenic abscesses have been managed with antibiotics and percutaneous drainage, thus avoiding more surgery in typically debilitated patients. Recently, laparoscopic drainage has been introduced as a ‘middle level' alternative to open surgery for those patients not amenable to percutaneous drainage.

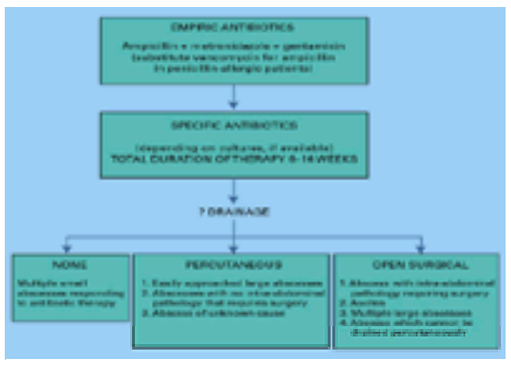


Fig. 2. Treatment of pyogenic liver abscesses.

Percutaneous drainage is carried out under CT or ultrasound guidance, with the insertion of a pigtail catheter using the Seldinger technique. Samples are then withdrawn for microbiologic examination, the abscess cavity is gently irrigated with saline, and the catheter is left in place to provide continuing drainage. More than one catheter may need to be placed to provide complete drainage. Such percutaneous aspiration and drainage does not correct the problem in 10 to 30 per cent of patients, and open drainage is then necessary. Failure to achieve drainage may be due to poor catheter placement, the presence of a multiloculated abscess, excessive viscosity of the abscess contents causing plugging of the drainage catheters, thick abscess walls that do not collapse with drainage, and inadequate

anatomic localization of the abscess. Follow-up ultrasonography or CT scanning is necessary to ensure complete resolution of the process.

Percutaneous drainage is less likely to be of value when there are multiple abscesses, a known intra-abdominal source of infection that requires surgical correction, an abscess of unknown cause, ascites, or when the abscess requires a transpleural drainage route.

Patients with biliary tract disease, diverticulitis, and appendicitis as the source for their liver abscesses are better treated with open surgical drainage, than by percutaneous drainage. Guidelines other than those mentioned to recommend percutaneous or an open surgical approach to these patients are still being formulated.

Mortality rates of patients with macroscopic liver abscesses reported in the 1960s and early 1970s were 65 to 79 per cent. Recent studies have noted a marked improvement, with mortality rates being as low as 11 per cent. Such improvement is due to the widespread availability of ultrasonography and CT scanning and hence earlier diagnosis, and the utility of the percutaneous approach to drainage, especially in debilitated patients, who tolerate conventional surgery poorly. The major determinants of mortality are the nature of the underlying process causing the abscess, the anatomy of the intrahepatic infection, and the presence or absence of such comorbidity factors as cancer, diabetes, heart disease, and renal failure.

Amebic abscesses

Epidemiology

Entamoeba histolytica infection affects an estimated 10 per cent of the world's population, with the great majority of such infections occurring in people living in sub-Saharan Africa, the Indian subcontinent, Asia, and parts of Central and South America. In these endemic areas approximately 50 per cent of the population is infected, with 90 per cent or more being asymptomatic cyst passers. In more developed countries, amebic infection occurs predominantly in immigrants or travelers returning from endemic areas, in sexually active male homosexuals, in residents of Indian reservations, and in people institutionalized for mental or emotional disability. The common denominator in these last groups is an increased opportunity for person-to-person spread via the fecal-oral route.

Amebic liver abscess occurs in less than 10 per cent of individuals infected with this organism. Whereas amebic infection of the liver is far less common than is pyogenic infection in the United States, in other areas of the world such as India, amebic abscesses are three to five times as frequent as pyogenic liver abscesses. The average age of patients with amebic liver abscess is between 28 and 48 years, which is significantly younger than patients with pyogenic infection (Table 1). Although infection rates are similar in men and women, there is a striking male predominance (up to 20:1) in patients who develop hepatic abscesses from amebiasis. Particularly severe invasive disease occurs in patients with compromised cellular immunity, in young infants, in the malnourished, in pregnant women, andin patients receiving corticosteroids.

Etiology and pathogenesis

Amebiasis is initiated by the ingestion of *E. histolytica* cysts. Once these cysts reach the small intestine, motile trophozoites are released and migrate to the colon, where they proliferate along with the resident bacterial flora. It is now apparent that the *E. histolytica* complex is made up of two distinct species: *E. histolytica*, which is capable of invading the colonic mucosa and causing extraintestinal disease; and *E. dispar*, which remains a gut commensal, neither invading nor causing extraintestinal manifestations. *E. histolytica* differs strikingly from *E. dispar*: genetically, resistance to complement-mediated lysis, and the presence of cysteine proteinases, which are essential for virulence. Other important determinants of invasiveness probably include diet, the constituents of the bacterial flora of the gut, and both humoral and cell-mediated host resistance.

Once intestinal infection is established, amebas may be carried to the liver via the portal vein. The ability to adhere to endothelial cells, and to bind to laminin, a component of the extracellular matrix, is an important virulence factor in the pathogensis of an amebic liver abscess. Within the liver these organisms multiply and block small intrahepatic portal radicles, causing focal infarction of hepatocytes. A proteolytic enzyme produced by the invading trophozoites causes coalescence of the invaded areas and abscess formation. Necrosis rather than apoptosis appears to be the primary form of liver injury induced, and polymorphonuclear leukocytes are the first line of host defense. Some authorities differentiate between so-called amebic hepatitis and amebic hepatic abscess, depending upon whether or not a macroscopic abscess has formed. We regard this differentiation as a difficult one, as this process might best be regarded as a continuum rather than as two separate entities.

At the time of clinical presentation, approximately 80 per cent of amebic liver abscesses are solitary; 83 per cent of them are located in the right lobe of the liver, characteristically high in the dome subjacent to the diaphragm. The propensity for this site reflects the fact that venous return from the right side of the colon (amebic infection having a particular impact on the cecum and right colon) into the portal vein is predominantly delivered to the right lobe of the liver. The juxtaposition of these abscesses to the diaphragm explain the common occurrence of thoracic symptoms in these patients. Discharge of amebic hepatic abscesses into the subphrenic, pleural, and even into pericardial spaces is not uncommon, with frank rupture into the lung as an uncommon complication.

Amebic abscesses can become extremely large, containing several liters of fluid that is classically described as ‘anchovy sauce' but may be yellow or green. This fluid is primarily necrotic liver tissue and blood, with a paucity of inflammatory cells unless bacterial superinfection has occurred. Because bile appears to have a deleterious effect on amebas, infection of the gallbladder and bile ducts does not occur.

Clinical presentation

Amebic liver abscess has a more subacute presentation than that of pyogenic liver abscess in the majority of patients. Symptoms typically evolve over a few weeks to months (as opposed to several days to weeks for a pyogenic process) before medical attention is sought. Initial symptoms are non-specific: fever, anorexia, night sweats, malaise, nausea and vomiting, and weight loss. As the disease becomes established, right upper quadrant abdominal pain becomes a dominating symptom in at least two-thirds of these patients. Approximately 25 per cent of patients exhibit thoracic symptoms, such as pleurisy, non-productive cough, right shoulder pain, and/or hiccups as an important part of the symptom complex. Characteristically, simultaneous intestinal complaints such as dysentery or diarrhea are not present. Uncommonly, patients may have a more fulminant presentation, suggesting an acute abdominal surgical emergency.

The patients typically appear chronically ill, with fever, abdominal tenderness, and hepatomegaly. Chest findings (rales, decreased breath sounds, dullness to percussion, impaired diaphragmatic movement) are observed in 50 per cent of patients. Jaundice is rare (less than 15 per cent of individuals).

Anemia and an elevated erythrocyte sedimentation rate are present in at least 80 per cent of patients, as well as a moderate leukocytosis in 60 to 75 per cent. More extreme white blood cell responses suggest the presence of bacterial superinfection. Eosinophilia is not observed in patients with amebic liver abscesses; if present, another explanation should be sought. Normal liver function tests do not exclude the diagnosis of an amebic abscess, although slight to moderate elevations of the alkaline phosphatase, reduction in the serum albumin, and minimal changes in aminotransferase values are generally observed.

Immune complex glomerulonephritis has been reported in the setting of an amebic liver abscess. Prognostic factors connoting a high risk for mortality include size of abscess, the presence of multiple abscesses, a bilirubin level of more than 3.5 mg/dl, and hypoalbuminuria (serum albumin less than 2.0 g/dl). Positive blood cultures are obtained in up to 20 per cent of patients with amebic abscess with secondary bacterial infection. Aspirated abscess fluid may also be culture positive for bacteria. Stool examination may reveal trophozoites or cysts in up to 25 per cent of patients.

The typical location of an amebic abscess in the dome of the right lobe of the liver produces not only thoracic symptoms, but also abnormalities on chest radiograph: an elevated right hemidiaphragm, pleural effusion, and atelectasis are the most common. Uncommonly a bronchohepatic fistula develops, with the characteristic expectoration of sputum that resembles ‘anchovy sauce'. Left hepatic lobe abscesses can produce not only left-sided pleuropulmonary signs and symptoms, but also can rupture into the pericardium, producing tamponade and/or mediastinal infection.

Diagnostic evaluation

In areas of the world free of endemic amebiasis, serologic testing is extremely useful in evaluating the patient for an amebic liver abscess (Fig. 1). The indirect hemagglutination test is positive in 90 to 95 per cent of patients with an amebic abscess, and in areas of low prevalence a positive result strongly suggests the presence of acute infection. False-negative tests are usually obtained early in the disease course. Levels of hemagglutinating antibodies remain elevated for many years after an episode of invasive disease, and the indirect hemagglutination test is therefore less useful in diagnosing acute disease in endemic regions, where 50 per cent of the general population may be seropositive. Newer ELISA assays appear to be at least as sensitive and specific. Levels of antibodies detectable by counterimmunoelectrophoresis and indirect immunofluorescence usually become undetectable within 6 months of acute infection; they may be more useful in evaluating patients in endemic areas.

Since amebic abscesses, unlike most pyogenic liver abscesses, respond to antimicrobial therapy without drainage, a non-invasive approach to diagnosing amebic

infection is needed. In the patient with subacute disease in whom the problem is a diagnostic dilemma rather than a therapeutic emergency, amebic serology should be performed and a therapeutic trial is initiated if the patient has an appropriate epidemiologic history. If the patient is unstable, if there is reason to suspect a pyogenic component to the illness on the basis of clinical or epidemiologic findings, if the amebic serology is negative, or if the patient has failed to respond clinically to several days of antiamebic therapy, a percutaneous needle aspiration should be performed.

Treatment

Most amebic abscesses are cured with a regimen of metronidazole 750 mg orally or intravenously three times per day, for 10 days, followed by treatment with an agent that is effective in eradicating the cysts which may persist in the intestine after treatment with metronidazole (Fig. 3). Agents effective in the treatment of such luminal disease include iodoquinol 650 mg orally, three times per day, for 20 days; diloxanide furoate 500 mg orally, three times per day, for 10 days; or paramomycin 25 to 30 mg/kg per day orally, in three divided doses, for 7 days.

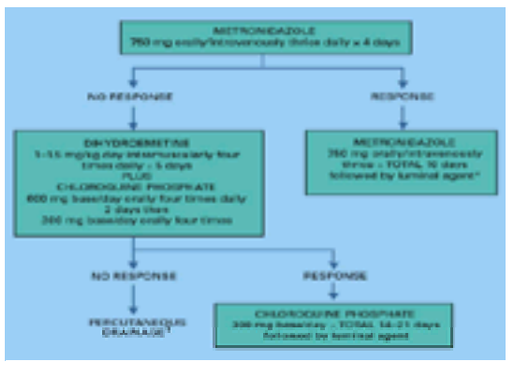


Fig. 3. Treatment of amebic liver abscesses. *Luminal agents include: iodoquinol 650 mg orally, thrice daily, for 20 days; diloxanide furoate 500 mg orally, thrice daily, for 10 days; or paramomycin 25 to 30 mg/kg per day orally, in three divided doses, for 7 days. +Other indications for percutaneous drainage are abscess in the left lobe of the liver and a possible large abscess. Indications for open surgical drainage are perforation to peritoneum, possible bacterial superinfection, and if the diagnosis is uncertain and percutaneous drainage is impossible.

Most patients show a prompt therapeutic response to metronidazole, with defervescence and decreased abdominal pain within 3 or 4 days. This response is useful for differentiating amebic and pyogenic abscesses in situations where serologic testing is unavailable or uninterpretable. Fewer than 10 per cent of patients with amebic liver abscess fail to respond to metronidazole therapy. Treatment in the non-responders is with dihydroemetine (1 to 1.5 mg/kg per day, maximum dose 90 mg/day, intramuscularly for 5 days) plus chloroquine phosphate (600 mg base/day for 2 days, followed by 300 mg of chloroquine base orally daily for 2 to 3 weeks). This regimen should be followed by treatment with an agent active against luminal disease.

Considerable controversy surrounds the role of percutaneous drainage in the management of amebic liver abscesses. Proposed indications for such drainage include liver abscesses in the left lobe of the liver, because of the risk of rupture into the pericardial sac; drainage of large abscesses to facilitate more rapid healing with chemotherapy; and lack of response to metronidazole therapy. However, because of the risk of introducing bacteria during catheter drainage, we reserve a drainage procedure for patients in whom the differentiation between pyogenic and amebic infection is unclear, for acutely ill patients, for patients who have failed to respond to therapy, and for patients whose infection has spread into adjoining structures.

Treatment of an amebic hepatic abscess is usually successful, mortality being associated only with delayed diagnosis or complications such as bacterial superinfection, or rupture into adjoining structures.

Further reading

Barnes PF, DeCock KM, Reynolds TN, Ralls PW. A comparison of amebic and pyogenic abscess of the liver. *Medicine* 1987; **66**: 473–83.

Bertel CK, van Heerden JA, Sheedy PF II. Treatment of pyogenic hepatic abscesses (surgical vs. percutaneous drainage). *Archives of Surgery* 1986; **121**: 554–8.

Bowers ED, Robison DJ, Doberneck RC. Pyogenic liver abscess. *World Journal of Surgery* 1990; **14**: 128–32.

Chen C *et al.* Clinical and microbiological features of liver abscess after transarterial embolization for hepatocellular carcinoma. *American Journal of Gastroenterology* 1997; **92**: 2257–9.

Chou FF, Sheen-Chen SM, Chen YS, Less TY. The comparison of clinical course and results of treatment between gas-forming and non-gas-forming pyogenic liver abscess. *Archives of Surgery* 1995; **130**: 401–5.

Chou FF, Sheen-Chen SM, Lee TY. Rupture of pyogenic liver abscess. *American Journal of Gastroenterology* 1995; **90**: 767–70.

Chou KM, Fan ST, Lai EC, Lo CM, Wong J. Pyogenic liver abscess. An audit of experience over the past decade. *Archives of Surgery* 1996; **131**: 148–52.

Cohen JL, Martin FM, Rossi RL, Schoetz DJ. Liver abscess. (The need for complete gastrointestinal evaluation.) *Archives of Surgery* 1989; **124**: 561–4.

Conter RL, Pitt HA, Tompkins RK, Longmire WP Jr. Differentiation of pyogenic from amebic hepatic abscesses. *Surgery, Gynecology, and Obstetrics* 1986; **162**: 114–20.

Farges O, Leese T, Bismuth H. Pyogenic liver abscess: an improvement in prognosis. *British Journal of Surgery* 1988; **75**: 862–5.

Gyorffy EJ, Frey CF, Silva Jr J, McGahan J. Pyogenic liver abscess (diagnostic and therapeutic strategies). *Annals of Surgery* 1987; **206**: 699–705.

Huang CJ *et al.* Pyogenic hepatic abscess. Changing trends over 42 years. *Annals of Surgery* 1996; **223**: 600–7.

Klatchko BA, Schwartz SI. Diagnostic and therapeutic approaches to pyogenic abscess of the liver. *Surgery, Gynecology, and Obstetrics* 1989; **168**: 332–6.

Li E, Stanley SL Jr. Protozoa. Amebiasis. *Gastroenterology Clinics of North America* 1996; **25**: 471–92.

Maharaj B, Bhoora IG, Patel A, Maharajh J. Ultrasonography and scintigraphy in liver disease in developing countries. *Lancet* 1989; **ii**: 853–6.

Miedema BW, Dineen P. The diagnosis and treatment of pyogenic liver abscesses. *Annals of Surgery* 1984; **200**: 328–35.

Navarro-Garcia F, Chavez-Duenas L, Tsutsumi V, Posadas del Rio F, Lopez-Revilla R. *Entamoeba histolytica*: increase of enterotoxicity and of 53- and 75-kDa cysteine proteinases in a clone of higher virulence. *Experimental Parasitology* 1995: **80**: 361–72.

Pineiro-Carrero VM, Andres JM. Morbidity and mortality in children with pyogenic liver abscess. *American Journal of Diseases of Childhood* 1989; **143**: 1424–7.

Prasad M, Arora A. Problem of pyogenic liver abscess in a tropical country. *American Journal of Gastroenterology* 1989; **84**: 1466.

Rajak CL, Gupta S, Jain S, Chawla Y, Gulati M, Suri S. Percutaneous treatment of liver abscesses: needle aspiration versus catheter drainage. *American Journal of Roentgenology* 1998; **170**: 1035–9.

Ralls PW. Focal inflammatory disease of the liver. *Radiologic Clinics of North America* 1998; **36**: 377–89.

Ralls PW, Barnes PF, Radin DR, Colletti P, Halls J. Sonographic features of amebic and pyogenic liver abscesses: a blinded comparison. *American Journal of Roentgenology* 1987; **149**: 499–501.

Rardin JI. Amebiasis. *Clinical Infectious Diseases* 1995; **20**: 1453–64.

Riley DK, Babinchak TJ, Zemel R, Weaver ML, Rotheram EB. Infectious complications of hepatic cryosurgery. *Clinical Infectious Diseases* 1997; **24**: 1001–3.

Rintou R, O'Riordain MG, Laurenson IF, Crosbie JL, Allan PL, Garden OJ. Changing management of pyogenic liver abscess. *British Journal of Surgery* 1996; **83**: 1215–18.

Seeto RK, Rockey DC. Pyogenic liver abscess. Changes in etiology, management, and outcome. *Medicine* 1996; **75**: 99–113.

Sharma MP, Dasarathy S, Verma N, Saksena S, Shukla DK. Prognostic markers in amebic liver abscess: a prospective study. *American Journal of Gastroenterology* 1996; **91**: 2584–8.

Singh JP, Kashyap A. A comparative evaluation of percutaneous catheter drainage for resistant amebic liver abscesses. *American Journal of Surgery* 1989; **158**: 58–62.

Tay KH, Ravintharan T, Hoe MN, See AC, Chng HC. Laparoscopic drainage f liver abscesses. *British Journal of Surgery* 1998: **85**: 330–2.

Thompson JE, Forlenza S, Verma R. Amebic liver abscess: a therapeutic approach. *Review of Infectious Diseases* 1985; **7**: 171–9.

Yeh TS *et al.* Pyogenic liver abscesses in patients with malignant disease: a report of 52 cases treated at a single instutition. *Archives of Surgery* 1998; **133**: 242–5.

30.4 Cancers of the liver

George W. Daneker, Jr. and Mark S. Talamonti

Introduction
Anatomy
Physiology
Patient evaluation
- **Clinical**
- **Laboratory**
- **Tumor markers**
- **Imaging studies**
- **Diagnostic laparoscopy**

Primary liver tumors
- **Hepatocellular carcinoma**
- **Cholangiocarcinoma**
- **Hepatoblastoma**
- **Epithelioid hemangiendothelioma**
- **Hepatic sarcomas**
- **Hepatic adenomas**
- **Hemangiomas**

Liver metastases
- **Colorectal liver metastases**
- **Neuroendocrine (including carcinoid) metastases**
- **Liver metastases from other primary tumors**

Further reading

Introduction

The liver is a common site for the development of primary malignancies as well as a favorite target organ for metastatic cancers. Despite its common involvement by cancers, the liver has remained a formidable object of respect to surgeons. Some apprehension in dealing with the liver undoubtedly stems from our earliest references to hepatic surgery. Prometheus, who incurred the wrath of Zeus following his creation of men from clay and the theft of fire, was chained to a rock and underwent daily hepatectomy by a vulture. Although this ritual was described as torture, Prometheus showed that the liver has amazing regenerative powers and could be a beneficial object of surgical attention. The modern era of liver surgery began in 1888 when Langenbuch performed the first left hepatic lobectomy. Tiffany was credited with performance of the first liver resection for a solid tumor 2 years later. Since that time, the surgical treatment of liver cancers, including resection and transplantation, has become commonplace and can be performed with low morbidity and mortality.

The emergence of successful treatments for cancers of the liver has resulted from improvements in imaging and tumor staging, in surgical techniques as a result of a better understanding of hepatic anatomy, in anesthesia and critical care, and in the ability to assess hepatic tolerance and maximize hepatic reserve. Selection of the appropriate therapy in each clinical situation is based on a sound understanding of the biologic behavior of the tumor, familiarity with the indications for and limitations of each therapeutic modality, and a thorough knowledge of hepatic anatomy.

Surgeons have long had a predominant role in the management of liver cancers. This position is a result of two factors: (i) surgery, including transplantation, is the sole modality that can produce a long-term disease-free survival or cure from primary and metastatic liver cancers; and (ii) surgeons have had a discriminating appreciation for the appropriate treatment modality, even non-surgical, due to their understanding of hepatic anatomy and procedural morbidity and mortality. At the present time, we are witnessing a proliferation of new treatments. While some treatments are surgical, the majority are not, and are primarily under the control of our non-surgical colleagues. In order to maintain our influence on the treatment of liver cancers, it is essential that we remain fully informed of all the latest developments. Equally important, we should then critically assess and integrate promising treatments, including non-surgical, into the overall multimodality management of liver cancers. An essential component of this integration is our participation in clinical trials. In this way, we can maintain our role as experts in the overall treatment of patients with liver cancers, not just experts on surgery for liver cancers.

Anatomy

A thorough knowledge of anatomy is essential for the hepatic surgeon. Armed with this knowledge, intelligent decisions can be made about resectability, surgical planning, and operative conduct. Modern hepatic surgery should be efficient, controlled, and with limited blood loss. Failure to adhere to anatomic principles portends an opportunity for massive blood loss and a poor surgical outcome.

The liver, which is the largest gland in the body, principally occupies the right subcostal and epigastric regions with an extension into the left subcostal region and the right lumbar region. The liver lies directly beneath diaphragm and is hidden behind a cage of ribs over the majority of its anteriolateral surface. A small portion of the liver's anterior surface is in contact with the anterior abdominal wall. Only that portion of the anterior surface of the liver is normally palpable, the remainder the liver is hidden behind its protective rib cage and away from the inquiring hand.

The anteriosuperior surface of the liver is convex and shaped by the overlying diaphragm. In a similar fashion, the inferior surface of the liver is imprinted by and conforms to the surrounding intra-abdominal organs. The liver's inferior surface is covered by a peritoneal continuation of the gastrohepatic omentum. The peritoneum envelops the liver and extends from the anterior, anteriosuperior, and lateral surfaces to reflect on to the diaphragm. This hepatic envelope, which is reinforced with fibrous tissue, is called Glisson's capsule. The reflections of Glisson's capsule off the liver and on to the diaphragm leave most of the posterior surface of the right lobe and a strip of the posterior surface of the left lobe in direct contact with the diaphragm. This ‘bare area' of the liver is contained within these peritoneal reflections, called coronary ligaments, and the lateral fusion of the coronary ligaments called triangular ligaments. Of importance, the inferior vena cava and hepatic veins are fully contained within the bare area. The liver's lateral surface lies beneath the costal margin at the midaxillary line and is covered by a portion of the thoracic wall comprised of the seventh to eleventh ribs.

The liver is functionally divided into lobes, sectors, and segments based on the arterial and portal venous blood supply, hepatic venous drainage, and biliary drainage (Fig. 1). The main anatomic division of the liver is into the right and left lobe. Branching of the proper hepatic artery, portal vein, and common bile duct into major right and left branches define each lobe. The division between these lobes follows a plane running vertically, through the liver, connecting the anterior border of the gallbladder fossa to the left side of the inferior vena cava posteriorly. This plane is known by an eponym as ‘Cantlie's line'.

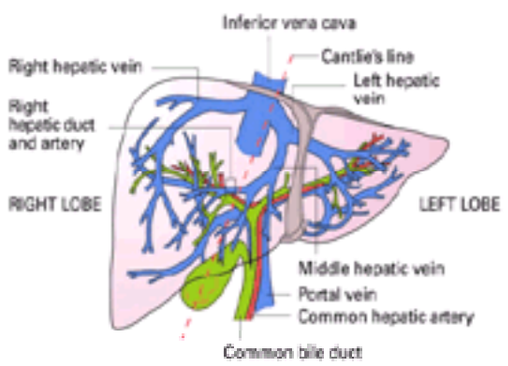


Fig. 1. Overview of intrahepatic vascular and biliary anatomy.

The right and left lobes can be further subdivided into sectors based on the distribution of portal pedicles and hepatic veins (Fig. 2). Using a plane of division along

each of the three main hepatic veins, the liver is divided into four sectors, known as the right anterior, right posterior, left medial, and left lateral sectors. Each sector is supplied by a vascular and biliary pedicle comprised of the major lobar hepatic artery branch, portal vein branch, and bile duct branch. No clear topographic features or morphologic boundaries exist between the sectors in the right lobe. Within the left lobe, the umbilical fissure and falciform ligament defines a plane of division between the medial and lateral sectors.

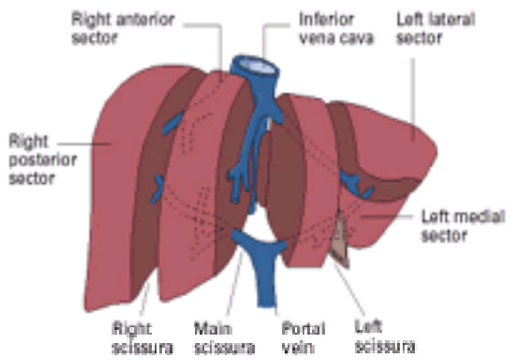


Fig. 2. Sectoral anatomy of the liver

Each hepatic sector can be further subdivided into numbered segments as described by Couinaud (Fig. 3). This segmentation is one step further beyond the division of liver into sectors and is based on the bifurcation of portal pedicles within the sectors. The liver's segmental anatomy can be remembered as follows: in a clockwise fashion beginning at the vena cava (12 o'clock) are segments 2, 3, and 4, which comprise the left lobe of the liver. Segment 4, the portion of liver between the falciform ligament and Cantlie's line, can be further divided into a superior half (segment 4a) and an inferior half (segment 4b). Continuing clockwise, the right lobe comprises segments 5, 6, 7, and 8. No clear boundaries exist between the segments of the right lobe. The caudate lobe is assigned segment 1 by this system. The caudate lobe is considered an autonomous segment from the standpoint of functional anatomy. It receives branches from both hepatic arteries and portal veins, although the majority of its blood supply comes from the left. Its venous drainage is not into a hepatic vein but rather through a variable number of bridging veins directly into the inferior vena cava.

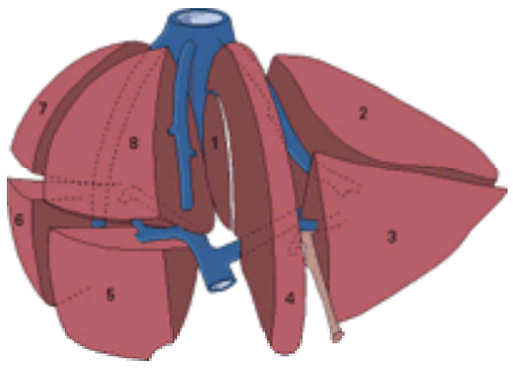


Fig. 3. Segmental anatomy of the liver (after Couinaud).

The common bile duct branches into the main right and left hepatic ducts on the right side of the liver hilum, anterior to the portal vein bifurcation, and overlying the origin of the right main portal trunk. The right hepatic duct is short and ascends into the parenchyma after the bifurcation. In 28 per cent of cases, one of the main right segmental ducts crosses to join the main left hepatic duct. Due to its length, orientation, and the variable anatomy of the biliary tree, the right duct is more vulnerable to injury then the more horizontally oriented left duct.

The hepatic veins arise from the inferior vena cava as it emerges from the liver immediately below the diaphragm. The suprahepatic inferior vena cava has a very short course before it passes through the diaphragm into the chest. In a similar fashion, the hepatic veins have a short extrahepatic course before they enter the liver. The right hepatic vein is usually single (61 per cent) and drains the anterior and posterior sectors of the right lobe. In 61 per cent of patients there are an additional one or two large veins draining from the posterior or posterioinferior right lobe directly into the inferior vena cava. The middle hepatic vein runs along Cantlie's line and drains the right anterior and left medial sector. The left hepatic vein drains the left lateral sector and a small portion of the left medial sector. In 84 per cent of patients the left and middle hepatic veins arise from a common trunk. The bifurcation of this common trunk usually occurs near its origin from the vena cava and makes extrahepatic division of both the left and middle hepatic veins potentially hazardous.

Physiology

The hepatic surgeon must remain cognizant of the liver's normal and pathophysiologic functions. This information is not just of academic interest, but is also practically important because the sequelae of hepatic dysfunction must be addressed during the perioperative evaluation and management, particularly for patients with chronic liver diseases. A more in-depth discussion of hepatic function is not within the scope of this chapter and the reader is referred to other sources for that information.

In contrast to its relatively formless and bland appearance, the liver is amazing in its performance of a plethora of critical functions. Underscoring this claim is the fact that the liver constitutes only 5 per cent of the total body weight but expends approximately 20 per cent of the body's energy and consumes up to 25 per cent of the body's total oxygen.

In general, the liver's main functions include anabolic and catabolic metabolism, vitamin and mineral storage, reticuloendothelial clearance, and maintenance of plasma volume and electrolyte concentrations. Within the context of anabolic function, the liver is responsible for synthesis of a number of integral proteins including serum albumin and a-globulin. In addition, the liver synthesizes 11 proteins critical for hemostasis involving both the intrinsic and extrinsic coagulation pathways. Additional critical anabolic functions include maintenance of serum glucose by glycogenolysis and gluconeogenesis, as well as lipoprotein and cholesterol metabolism. The liver has a key role in the metabolism of vitamins A, B, C, D, E, and K. Among the most significant vitamin deficiency syndromes associated with hepatic dysfunction is coagulopathy resulting from vitamin K deficiency. In addition to its effects on the extrinsic pathway, chronic liver disease may produce coagulopathy due to insufficient absorption of vitamin K as a result of bile salt deficiency.

The liver also has a primary role in the metabolism and detoxification of drugs and other metabolic by-products. Drug catabolism occurs principally by the cytochrome P-450 system. The most significant metabolic by-product produced by the liver is bilirubin, a breakdown product of hemoglobin that is excreted almost entirely in the bile.

The liver can perform immunologic functions by means of the fixed reticuloendothelial Kupffer cells as well as through the activities of the sinusoidal endothelial cells.

Among the amazing synthetic functions of the liver is its regeneration after partial hepatectomy. The liver can routinely reconstitute its normal mass following a hepatectomy of up to 75 per cent, provided the regenerating parenchyma is normal. In humans, the time-frame for complete regeneration is from weeks to several months. Studies have shown that hepatic regeneration results from a combination of compensatory hypertrophy and hyperplasia of the cells within the hepatic remnant. A number of growth factors are involved, but the exact role of these growth factors is not well known. The anatomy of regeneration is such that the new parenchyma expands and deforms the residual liver. After a right hepatectomy, the regenerated liver is centered to the left near the epigastrium. After a left hepatectomy, the bulk of the regenerative liver is in the right subcostal region. The physiologic importance of regeneration is the reserve it supplies after a cancer is resected. As a neoplasm grows, it displaces the normal liver tissue but does not destroy it. In the unlikely event that parenchyma is damaged by the neoplasm, regeneration can compensate. Enormous neoplasms can therefore be removed from the liver with little compromise of the hepatic mass.

H3>Patient evaluation

Clinical

Routinely, all patients seen for evaluation of a primary or metastatic liver tumor undergo an initial complete history and physical examination. The history and physical examination are important for two reasons: an evaluation of the patient's general health and comorbid conditions which impact on the choice of proposed treatment, and the initial clinical staging of the patient with preliminary selection of the appropriate treatment option(s). Although simple, these evaluations are important because the detection of advanced comorbid disease, such as cirrhosis, or more extensive tumor spread, usually extrahepatic metastases, will often eliminate the patient from further consideration of surgery as the principal therapy.

Laboratory

Patients presenting with a liver tumor should also have blood taken for laboratory studies. These studies should include a complete blood count, platelet count, serum electrolytes, and ‘liver function tests'. The ‘liver function tests' should include aminotransferases, alkaline phosphatase, bilirubin, and lactate dehydrogenase; these studies will provide information as to the overall health of the liver. In addition to these classic liver function tests, measurement of the albumin level, prothrombin time, and partial thromboplastin time are included. These tests are important because they reflect hepatic synthetic function. In patients with primary liver tumors or a history of hepatitis, levels of anti-hepatitis B antibody, hepatitis B surface antigen, and anti-hepatitis C antibody should be measured. If these hepatitis B and C screening studies are positive, they should be confirmed by quantitative assessment of viral DNA or RNA by polymerase chain reaction.

Tumor markers

Tumor cells may also synthesize and secrete substances, usually glycoproteins, which can be clinically useful in screening for the tumor, or for monitoring tumor status following treatment. In the case of primary and metastatic liver tumors, the markers a-fetoprotein and carcinoembryonic antigen are the most useful.

Measurement of a-fetoprotein levels (normal range 1 to 10 ng/ml) has been used extensively in the management of hepatocellular carcinoma. Despite its widespread use, this test is beset with both false-positive and false-negative results. At present, elevation of serum a-fetoprotein has a sensitivity of 68 per cent and a specificity of only 20 per cent in detecting small early hepatocellular carcinomas (less than 3 cm in diameter). Even with these limitations, measurement of a-fetoprotein remains an essential component of screening high-risk patients for hepatocellular carcinoma. Currently, the recommendations are for measurement of serum a-fetoprotein levels every 4 months, along with trans-abdominal hepatic ultrasound every 16 months in patients with chronic viral hepatitis, especially those with established cirrhosis. Overall, elevation of a-fetoprotein above normal has a true-positive rate of 80 per cent and a false-negative rate of 20 per cent, when all hepatocellular carcinomas are taken into consideration. False-negative elevations of a-fetoprotein usually result from chronic active hepatitis or cirrhosis. If the cut-off for an abnormally elevated a-fetoprotein is raised to 500 ng/ml, the sensitivity is decreased to 60 per cent but specificity for hepatocellular carcinoma is increased to 100 per cent. In addition, a sharp steady rise in serum a-fetoprotein levels is highly diagnostic for hepatocellular carcinoma. As with most tumor markers, rising a-fetoprotein levels after definitive treatment are usually a sensitive indicator of recurrence, provided that the primary tumor could secrete a-fetoprotein. Another tumor marker useful in hepatocellular carcinoma is abnormal prothrombin or PIVKA-II. Because PIVKA-II levels are elevated in 65 per cent of patients with hepatocellular carcinoma, and can be elevated when a-fetoprotein levels are normal, this marker may be useful for diagnosis and monitoring of therapy in patients with normal serum a-fetoprotein levels.

Carcinoembryonic antigen (normal range 1 to 5 ng/ml) is one of the longest known and most extensively studied of all the tumor markers. Since most (65 to 90 per cent) colorectal cancers are able to synthesize and secrete carcinoembryonic antigen, the measurement of carcinoembryonic antigen levels is most commonly used in the management of colorectal cancer, even though this antigen may be elevated in other cancers such as breast and lung. Overall, serum carcinoembryonic antigen levels are an indirect reflection of tumor mass, although the threshold for detection of the antigen in the serum is high. This fact is reflected in the finding that up to 25 per cent of early stage tumors (Duke's A and B) have an elevated carcinoembryonic antigen level, while 85 to 90 per cent of patients with hepatic metastases (Duke's D) have elevated antigen levels. In addition, some investigators believe that carcinoembryonic antigen production is associated with a more aggressive biologic behavior and that cells producing this antigen have a greater ability to become both locally invasive and metastatic than cells not producing the antigen.

The high threshold for detection of carcinoembryonic antigen elevation makes it poorly suited for colorectal cancer screening. This antigen's main usefulness is as an initial staging study and, more importantly, as a marker of response to therapy (for both primary colorectal cancer and metastases) and of recurrence following therapy. The serum carcinoembryonic antigen level, collected at the time of an initial evaluation, will give a general indication of prognosis based on the degree of elevation. Many investigators have reported that higher serum levels are associated with an increased incidence of recurrence following treatment, but correlations between specific values and outcome are controversial. The measurement of carcinoembryonic antigen levels is far more helpful in the management of a patient with an elevated level on initial screening, whose level returns to normal following resection of the primary tumor. In this patient, a progressive rise of the antigen level is virtually diagnostic of recurrence and will serve as the first indicator of recurrence in approximately two-thirds of patients. However, on a cautionary note, an isolated elevation in carcinoembryonic antigen may be falsely positive and should be confirmed by an additional two or three measurements at separate times (usually during the following 2 to 3 months). Since serum carcinoembryonic antigen levels are often an indirect reflection of tumor burden, measurement of these levels is an indicator of response to therapy independent of imaging studies. In particular, for those patients undergoing chemotherapy, 90 per cent of those who respond will have a decrease in their carcinoembryonic antigen levels, while 90 per cent of those with progression of their tumor will have increasing levels.

Recently developed tumor markers, which include CA19–9 and CA 125, have also been used in the management of colorectal cancer. Although these markers can be helpful, particularly in patients who are negative for carcinoembryonic antigen, both their sensitivities and specificities are inferior to carcinoembryonic antigen. Because of these factors, plus the easy commercial availability of the test for the carcinoembryonic antigen, routine use of these markers cannot be recommended at this time.

Imaging studies

The importance of imaging studies in the management of patients with liver malignancies cannot be overemphasized. These studies provide the most specific and detailed staging information and are indispensable for clinical decision making. While improvements in the understanding of hepatic anatomy, surgical technique, and perioperative care have made major contributions to the current success of liver-directed therapy, the single most important factor in improved survival is better patient selection due to the accuracy of modern imaging studies.

When used for evaluating the patient contemplated for liver-directed therapy, imaging studies have the following goals: (i) determination of the number and distribution of the liver lesion(s); (ii) anatomic and functional characterization of liver lesion(s); (iii) delineation of the lesion(s) relationship to significant vascular and biliary structures; and (iv) detection of extrahepatic and extra-abdominal tumor.

Nuclear scans

Normal liver tissue contains Kupffer cells, which are able to take up $^{99}Tc^{m}$-sulfur colloid, thereby displaying the liver on liver–spleen scans. Tumors, on the other hand, lack Kupffer cells, and are visualized as photopenic defects on liver–spleen scans. Although these differences are theoretically useful, in practicality the details provided by liver–spleen scans are inadequate for staging and this procedure was largely abandoned for staging of the liver with the advent of computed tomography (**CT**) in the 1980s. For the evaluation of liver lesions, the most commonly used nuclear scan, at present, is the tagged red blood cell scan. The study is used to determine whether a liver lesion, usually detected by another imaging modality, is a hemangioma and contains many tagged red blood cells or a solid lesion with few red blood cells. On tagged red blood cell scans, a hemangioma is imaged as a photodense area in the liver while, in contrast, a tumor is imaged as a photopenic area within the liver.

Ultrasound

Percutaneous ultrasound is the most readily available, least expensive, and least invasive of the imaging modalities for the liver. Unfortunately, the consistency of results from percutaneous ultrasound is limited, in up to 50 per cent of cases, by dependence on operator expertise and the inability to obtain a complete study due to interference from air in the lung or gastrointestinal tract, patient body habitus, and motion artifact. In addition, small tumors, either primary hepatocellular or metastatic carcinoma, can be difficult to detect and differentiate from regenerating nodules in the cirrhotic liver (this can be a difficult problem with CT as well). None the less, percutaneous ultrasound remains the most popular liver screening modality worldwide. The United States is an exception, where an abundance of CT scanners has caused CT to be favored over ultrasound. In the hands of experts, results from percutaneous ultrasound are as accurate as those obtained by CT or magnetic resonance imaging (**MRI**) with respect to lesion size, distribution, and relationship to intrahepatic anatomy. In addition, greater detail about intratumoral anatomy and tissue characteristics is provided by ultrasound when compared with CT. Percutaneous ultrasound has also had a valuable therapeutic role when used in the ablation of liver lesions, particularly in alcohol ablation for patients with hepatocellular carcinoma. The advantage of ultrasound over CT or other imaging modalities during therapeutic interventions is the real-time information obtained by ultrasound, particularly during placement and positioning of intrahepatic wires or needles.

Intraoperative ultrasound is an invaluable tool for the hepatic surgeon. In this type of ultrasound, the probe is placed directly on the parenchymal surface of the liver following its mobilization. Because the probe is in direct contact with the liver surface, transducers (5 to 10 MHz) with a shorter depth of penetration but capable of displaying greater anatomic detail are used. In general, the higher the transducer frequency the greater the anatomic detail, but at the expense of a reduced depth of

penetration. Dedicated intraoperative ultrasound probes are usually small, easily hand-held, and configured in either an ‘I' or ‘T' shape. During hepatic examination by intraoperative ultrasound, the following information is obtained: definition of portal vascular and biliary anatomy, definition of hepatic vein anatomy, identification of occult and non-palpable tumors, and definition of tumor(s) and its relationship to intrahepatic vascular and biliary anatomy. Once a surgeon has become experienced with the techniques of intraoperative ultrasound, it should become a routine part of any liver procedure. Overall, ultrasound can improve the sensitivity for detection of tumors larger than 1 cm in size to over 95 per cent. Ultrasound findings may alter the planned operation or lead to abandonment of the procedure in up to 30 per cent of patients.

Computed tomography

Since its development in the 1980s, CT has been the most widely used modality for definitive imaging of liver tumors (Fig. 4(a)). Initially, CT sensitivity was poor, owing to similarities in density between tumor and the surrounding liver. To overcome this obstacle, techniques were developed that used timed scanning following rapid administration of intravenous contrast. Despite the use of intravenous contrast, the sensitivity for detection of liver metastases by planar CT remained between 40 and 75 per cent. In the 1990s, a major advance in CT technology occurred with the ntroduction of spiral (helical) CT. This technology allows for continuous data acquisition from rapid uninterrupted scanning during a single breath hold. The helical data set results in improved visualization of the lesion and the ability to reconstruct overlapping images at variable intervals. Studies have shown that an additional 20 per cent of liver lesions could be identified, when compared with planar CT. The greater sensitivity of spiral CT was mostly a result of detecting additional lesions of less than 2 cm. As a result of reduced scan times, newer CT techniques have focused on optimization of the intravenous contrast bolus and timing of image acquisition, so that the arterial and portal venous phase of liver enhancement can be captured. Using this ‘biphasic' technique, up to 40 per cent of additional smaller liver lesions may be detected. A further benefit of spiral CT is the ability to merge the overlapping slices retrospectively to produce three-dimensional reconstructions. These reconstructions can be used to image anatomically complex areas, such as the porta hepatis, with more clarity and precision.

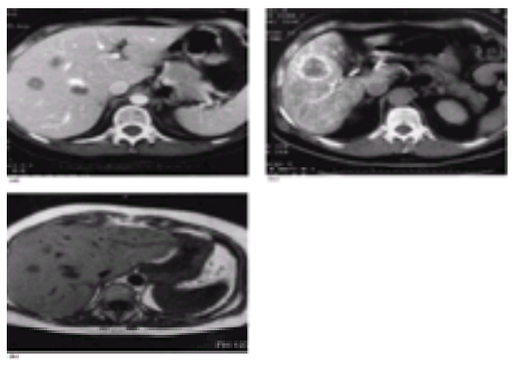


Fig. 4. (a) Biphasic spiral CT of the liver showing colorectal metastases. (b) Angiographically assisted CT of the liver showing a colorectal metastasis. (c) MRI of the liver showing colorectal metastases.

The liver is unique in that it receives vascular inflow from both the hepatic artery (30 per cent) and the portal vein (70 per cent), while metastases receive their blood flow almost exclusively from the hepatic artery. Angiographically assisted CT was developed to capitalize on this differential perfusion of liver metastases (Fig. 4(b)). In this technique, CT is performed during transcatheter infusion of contrast material into the hepatic artery (CT hepatic angiography) or during portal venous enhancement following infusion of contrast material into the superior mesenteric or splenic artery (CT angioportography). Since its introduction, angiographically assisted CT has consistently demonstrated an 85 to 90 per cent sensitivity for detection of intrahepatic lesions and has become the ‘gold standard' imaging study for the assessment and planning of surgical resections. The greater sensitivity of angiographically assisted CT has also primarily resulted from improved detection of lesions less than 2 cm in size. Angiographically assisted CT not only improves lesion detection, but it also very clearly depicts relevant intrahepatic vascular anatomy critical to operative planning. Although it remains the current ‘gold standard' study, criticisms of this technique include a high false-positive rate (up to 20 per cent) from benign perfusion defects and more frequent technical problems than CT or MRI. Furthermore, the angiographically assisted CT is an invasive examination with risks and significant expense. An additional important feature is the delineation of hepatic arterial anatomy provided by celiac and superior mesenteric artery angiograms. Angiographically assisted CT still remains the definitive preoperative liver staging study. As spiral CT and MRI technology have improved, the need for angiographically assisted CT has diminished and will, in all likelihood, eventually disappear.

For patients with hepatocellular carcinoma, the contrast agent lipiodol can be delivered during hepatic angiography. Lipiodol, which is an ethiodized oil emulsion, is unique in that it is retained (indefinitely) within tumors but not normal or cirrhotic liver. While useful radiographically, this property of lipiodol can also be exploited for therapeutic purposes. Lipiodol is mixed with chemotherapeutic drugs so that these drugs will have both a higher concentration and duration of action when retained within the tumor (see section below on embolization and chemoembolization).

Magnetic resonance imaging

MRI has not achieved the same widespread popularity as CT for routine imaging of the liver. Historically, this has largely been due to problems with motion artifact, inferior spatial resolution when compared with enhanced CT, and lack of suitable intravenous and intestinal contrast agents for MRI. As MRI technology has improved, in particular with the routine use of higher field strength (1.5 T) magnets, improved gradients, improved pulse sequences, and refinements in motion suppression techniques, many of these criticisms are no longer relevant (Fig. 4c). Unenhanced hepatic MRI has been shown to have a sensitivity for detection of hepatic lesions that is equal to, or slightly better than, contrast-enhanced CT. MRI does offer excellent delineation of lesion morphology and characteristics as a result of using multiple pulse sequences. In fact, the ability to characterize reliably benign liver lesions, such as cysts or hemangiomas, from malignant lesions is a major advantage of MRI. Also, MRI can depict the relationship of lesions to major vascular structures without the use of contrast agents. Because of rapid technologic advances, differences between various MRI manufacturers, and the continued evolution of pulse sequences, hepatic MRI techniques vary widely between institutions. Despite these differences, a combination of T_1- and T_2-weighted sequences are routinely used for lesion detection and characterization. The T_2-weighted sequences, with tumors appearing hyperintense, appear to be the most efficient for detecting intrahepatic tumor, although the combination of T_1- and T_2-weighted images usually provides complementary information. In an effort to improve sensitivity for lesion detection, contrast agents for hepatic MRI have been developed. The efficacy of these agents, including gadolinium and ferumoxide, has yet to be shown conclusively; although recent studies have shown the superiority of lesion detection for ferumoxide-enhanced over unenhanced MRI. Despite improvements in hepatic imaging, questions about the ability of MRI to detect extrahepatic tumor still remain. An additional criticism of hepatic MRI stems from the current image quality. Although MRI has continued to improve in clarity with technologic advances, it generally remains inferior to CT with respect to depiction of intrahepatic vascular and biliary anatomy. As noted above, CT holds the current principal role for definitive staging of liver.

Positron emission tomography (PET)

The alteration of biochemical processes within tumors usually precedes their detection by gross anatomic changes. Positron emission tomography is one of the functional modalities that images by exploiting tumor-related biochemical processes, in contrast to the anatomic and structural information provided by most imaging modalities. Because of its emphasis on biochemical changes, PET scanning is theoretically more sensitive than CT or MRI for detection of occult or low-volume cancers. Enhanced glycolysis and glucose retention in tumor cells was first described by Warburg in the 1930s and is the basis for ^{18}F-fluorodeoxyglucose PET imaging in the diagnosis, staging, and follow-up of patients with colorectal cancer. In recurrent colorectal cancer, the presence of extrahepatic metastases has a profound negative impact on surgical curability. The limitations of CT and MRI for detection of extrahepatic tumor recurrence is well recognized and, at best, these studies may detect only 60 per cent of extrahepatic recurrences. In early studies on patients with metastatic colorectal cancer, there was a greater than 90 per cent sensitivity and specificity for detection of both intra- and extrahepatic recurrences by whole-body PET scans. In addition, these studies showed that surgical management was altered in up to 30 per cent of patients according to the findings on PET scan. Before PET can be recommended as a routine part of staging for patients with recurrent colorectal cancer, the impressive results of these series will have to be duplicated. Unfortunately, the widespread use of PET may be limited by the prohibitive costs associated with the production of radiolabeled metabolic substrates.

Diagnostic laparoscopy

Because of the aforementioned limitations of CT and MRI for detection of extrahepatic tumor recurrences, diagnostic laparoscopy has been considered as an initial step in surgical exploration. Small older series have shown that up to one-third of patients with metastatic colorectal cancer or hepatocellular carcinoma will be deemed unresectable by laparoscopy due to the detection of unrecognized intrahepatic tumor spread or peritoneal seeding. Although promising, contributions from the routine use of laparoscopy must be weighed in light of modern CT, MRI, and PET. In addition, adequate laparoscopic staging in the setting of a previous laparotomy may

necessitate a time-consuming lysis of adhesions to examine the entire abdomen fully. At present, diagnostic laparoscopy cannot be recommended as a routine staging study in patients with recurrent colorectal cancer or hepatocellular carcinoma.

Primary liver tumors

Hepatocellular carcinoma

Hepatocellular carcinoma is the most common liver malignancy worldwide, causing an estimated one million deaths annually. The geographic distribution of hepatocellular carcinoma shows distinct variation with areas of high, intermediate, and low incidence in the general population. In endemic ‘hot spots,' including sub-Saharan Africa, South-East Asia, and southern China, incidence rates may run as high as 150 per 100 000 persons. At the other end of the spectrum, North America, northern Europe, and Australia have incidence rates of less than 2 per 100 000 persons. The high mortality from hepatocellular carcinoma is largely due to both patient presentation at a late stage of the disease and poor hepatic function/reserve resulting from cirrhosis.

Etiology

The two major groupings of etiologic factors for hepatocellular carcinoma are environmental and viral. The best studied and most potent natural carcinogen is a product of *Aspergillus* fungus called aflatoxin B1. In humid parts of the world, where grain is stored in unrefrigerated conditions, fungal overgrowth produces contamination with aflatoxin. Epidemiologic studies in hyperendemic areas have shown the close correlation between aflatoxin exposure and the development of hepatocellular carcinoma. It is unclear whether aflatoxin exposure alone is adequate for the development of hepatocellular carcinoma, or whether it acts as a cocarcinogen with hepatitis B virus (**HBV**) infection. In low-incidence areas, alcohol-induced hepatic cirrhosis has a major etiologic association with hepatocellular carcinoma and 10 per cent of patients dying from alcoholic cirrhosis will have hepatocellular carcinoma. The exact role of alcohol is not known, but evidence supports its role as a cocarcinogen rather than a direct carcinogenic agent. Other environmental factors include mycotic toxins, plant alkaloids, anabolic steroids, and pollutants including pesticides and insecticides.

The most prominent factor worldwide linked to the development of hepatocellular carcinoma is chronic hepatitis from HBV or hepatitis C virus (**HCV**) infection. These infections result from contact with contaminated blood or bodily fluids. In particular, one common source of HCV infections in developed countries has been transfusion with contaminated blood in the era prior to routine HCV testing. Chronic viral infection contributes to the high incidence of cirrhosis (75 per cent) or chronic hepatitis (10 per cent) in the majority of patients with hepatocellular carcinoma. The annual cumulative risk of developing hepatocellular carcinoma in the setting of chronic viral hepatitis without cirrhosis is approximately 1 per cent, while the risk of developing hepatocellular carcinoma with cirrhosis ranges from 3 to 10 per cent. This significant risk has prompted aggressive screening programs in high-incidence countries, like Japan, where patients with cirrhosis are evaluated by ultrasonography and measurement of a-fetoprotein levels every 3 months. In addition, a campaign of immunization against HBV may have a significant impact on the development of hepatocellular carcinoma in endemic areas. Differences in hepatocellular carcinoma due to HBV compared with HCV include: a 10-year earlier age of onset (50 instead of 60 years) for HBV, a shorter interval from infection to development of hepatocellular carcinoma with HCV, and more advanced cirrhosis and poorer hepatic function with HCV.

Clinical presentation

Presenting symptoms and signs are related to the tumor stage and are an indirect reflection of the intensity of screening and ease of access to medical care. Common symptoms include abdominal pain, shoulder pain, abdominal swelling, weight loss, malaise and weakness, and anorexia/nausea/vomiting. Common physical findings include hepatomegaly, abdominal pain and tenderness, splenomegaly, ascites, jaundice, fever, hepatic bruit, muscle wasting, and signs of portal hypertension. In areas with limited medical care, patients are more likely to present with abdominal pain, abdominal mass or swelling, and hepatomegaly. Patients in screening programs are more likely to be asymptomatic from the tumor *per se* and will be predominantly symptomatic from cirrhosis and hepatic dysfunction. Up to 10 per cent of patients, particularly in Asia, will present with atraumatic acute hemoperitoneum from bleeding into a necrotic tumor. Jaundice is commonly due to hepatic dysfunction, however it may also be due to tumor compression of the main intrahepatic or extrahepatic biliary tree. A large number of paraneoplastic syndromes, usually biochemical in nature, are seen with hepatocellular carcinoma. These include hypoglycemia, hypercalcemia, hypercholesterolemia, hypertriglyceridemia, porphyria, pseudohyperparathyroidism, sexual changes, hypertrophic pulmonary osteoarthropathy, and carcinoid syndrome. Associated hematologic abnormalities include erythrocytosis, plasmacytosis, dysfibrinogenemia, and antifibrinolysis. Abnormalities in serum levels of globulins, haptoglobin, ceruloplasmin, and a-antitrypsin and increased thyroid-binding globulin are also seen.

Diagnosis

As noted above, a-fetoprotein was the first and is the most widely used serologic assay for hepatocellular carcinoma. By using radioimmunoassay, 70 to 90 per cent of patients from Asia and 75 per cent of patients from the United States with hepatocellular carcinoma will have significant a-fetoprotein elevations. This number decreases for patients with hepatocellular carcinoma in non-cirrhotic livers. In general, a sharp, steady rise in serum a-fetoprotein, especially in a patient at risk for hepatocellular carcinoma, is highly diagnostic. It is noteworthy, however, that serum a-fetoprotein is normal in 40 per cent of patients with tumors smaller than 3 cm. Ultrasonography is particularly useful in covering the weaknesses of a-fetoprotein measurements, especially in screening programs. Unfortunately, percutaneous ultrasound is rarely able to detect lesions smaller than 1 cm in diameter.

Although ultrasound is useful for screening, it is supplanted by biphasic spiral CT as the principal imaging study for hepatocellular carcinoma. Because of the distorted intrahepatic anatomy resulting from cirrhosis, MRI, particularly with ferumoxide enhancement, can be helpful for distinguishing between small hepatocellular carcinoma and adenomatous hyperplastic nodules when CT is inconclusive in answering this question. Unfortunately, the accuracy for detection of hepatocellular carcinoma with MRI falls off dramatically for tumors less than 2 cm in diameter. Angiographically assisted CT can be difficult to interpret because of perfusion abnormalities from cirrhosis and is not routinely obtained. Angiography can be helpful as a route to deliver lipiodol into the hepatocellular carcinoma. Lipiodol retention within small tumors may improve their detection by CT. PET imaging has enjoyed only modest success in hepatocellular carcinoma and is not recommended as a preoperative screening study. Ultrasound or CT have also been useful for directing a core biopsy of suspicious liver lesions, but routine biopsy for at-risk patients with an elevated a-fetoprotein and a clearly defined liver mass(es) is not indicated. In addition, a core liver biopsy in patients with cirrhosis and the resultant coagulopathy poses a risk of severe bleeding, in addition to the mostly theoretical risk of abdominal contamination by tumor cells.

Patient evaluation and selection

In contrast to other patients with primary and metastatic liver malignancies, the prognosis of patients with hepatocellular carcinoma is determined not only by the tumor's stage, but also by the functional status of the patient's liver. Often, liver function will loom as an important, if not the most important, factor in treatment decisions. In order to deal with the difficulties imposed by the tumor biology and hepatic dysfunction of hepatocellular carcinoma, a number of treatment modalities have been developed and are detailed below.

Patients evaluated for treatment of hepatocellular carcinoma undergo a history, physical examination, laboratory studies, and imaging studies as detailed above. The studies, particularly those grouped together in Child's classification of liver function (Table 1), give an excellent initial idea of treatment limitations imposed by hepatic dysfunction. The chest should be imaged by routine posterioanterior and lateral chest radiographs. If these are abnormal, they should be followed by chest CT. Further pretreatment quantitative assessment of liver function is usually required. A number of quantitative assessments have been developed that measure the disappearance of a test substrate from the blood, principally as an indirect reflection of hepatic function. Unfortunately, the kinetics involved with these tests are dependent on, and affected by, hepatic blood flow. The studies include indocyanine green clearance, galactose elimination, and the radiolabeled aminopyrine breath test. Indocyanine green clearance has been shown to be an accurate indicator of hepatic reserve and is currently the most favored test in clinical use.

Score	1	2	3
Ascites	Absent	Slight to moderate	Severe
Encephalopathy	Absent	Slight to moderate	Severe
Serum albumin	> 3.5 g/dl	3-3.5 g/dl	< 3 g/dl
Serum bilirubin	< 2 mg/dl	2-3 mg/dl	> 3 mg/dl
Prolongation of prothrombin time	< 4 s	4-6 s	> 6 s
Score of 5 to 6 corresponds to Child–Pugh class A.			
Score of 7 to 11 corresponds to Child–Pugh class B			
Score of 12 to 15 corresponds to Child–Pugh class C			

Table 1 Child–Pugh classification of liver function

Accurate information about the tumor, obtained principally from laboratory and imaging studies, is critically important for the selection of appropriate treatment and prognosis (Table 2). Adverse tumor factors include multicentricity, bilobar distribution, size greater than 5 cm, absence of tumor capsule, capsular invasion, portal vein thrombosis, extrahepatic metastases, and vascular invasion or thrombosis. The ominous significance of vascular invasion cannot be overemphasized; many studies have demonstrated this factor as the most important determinant of survival following resection. Although biopsy is not routinely recommended for hepatocellular carcinoma, biopsy information can be useful for modifying surgical aggressiveness. In particular, the fibrolamellar variant of hepatocellular carcinoma has a much better prognosis than other histopathologic types. This cancer usually occurs in patients in their 20s and is not associated with elevated serum a-fetoprotein levels, chronic viral hepatitis, or cirrhosis. Survival following surgical resection is significantly better than for other types of hepatocellular carcinoma, even in patients with extrahepatic metastases.



Table 2 TNM staging of hepatocellular carcinoma

Surgical resection

Surgical extirpation, either by resection or liver transplantation, is the only curative treatment for hepatocellular carcinoma. At present, 5-year survival rates following resection average from 20 to 30 per cent worldwide, while survival rates as high as 50 per cent have been reported by several groups from Asia, Europe, and North America (Table 3). In both American and European experiences, there has been reported a 75 per cent 3-year survival for stage II disease, 50 per cent 3-year survival for stage III disease, and 10 to 20 per cent 3-year survival for stage IVa (non-metastatic) disease. These data probably reflect differing populations, with the best results in patients with a lower incidence of cirrhosis and smaller tumors. Overall, the survival rates for patients with cirrhosis are approximately one-half that of patients without cirrhosis. As noted above, surgical selection criteria not only include the tumor stage, but also the functional status of the liver. In general, patients selected for resection should meet the following criteria: a solitary hepatocellular carcinoma less than 10 cm in diameter or up to three smaller hepatocellular carcinomas in a surgically accessible location(s); no vascular invasion of the lobar portal trunk, main portal trunk, or main hepatic vein(s); no extrahepatic tumor (with the exception of fibrolamellar hepatocellular carcinoma in highly selected patients); Child's class A or well compensated B; confirmation of adequate liver function by quantitative studies; intact performance status; and absence/excellent control of significant comorbid medical illnesses, particularly diabetes mellitus and renal insufficiency. The operative goal is as limited a resection as possible to achieve a 1-cm margin of normal tissue. The resection is limited to prevent the increased morbidity and mortality, mostly from hepatic failure in patients with cirrhosis, associated with progression from a non-anatomic resection to a segmental resection to a formal lobectomy (without a clear increase in disease-free survival). Operative mortality following liver resection is directly related to underlying liver function and ranges from 3 per cent for patients without cirrhosis up to 25 per cent for patients with cirrhosis. Despite resection of all detectable disease, intrahepatic recurrence rates can run a high as 80 per cent within 5 years of resection. These data strongly argue for more effective adjuvant chemotherapy rather than condemning surgical therapy as ultimately ineffective.

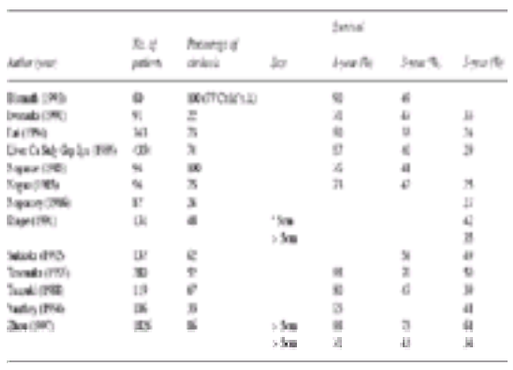


Table 3 Results of surgical resection for hepatocellular carcinoma

Principles of surgical resection

The initial operative step is a thorough exploration for extrahepatic metastases through a limited right subcostal incision. If no extrahepatic metastases are found, or limited extrahepatic metastases are found for a fibrolamellar hepatocellular carcinoma, the incision is then extended down the right subcostal margin to the lateral peritoneal reflection and down the left subcostal margin to the lateral rectus sheath. For large and/or posteriorly situated tumors, a midline extension to the xiphoid improves access to the vena cava and hepatic veins. Self-retraining retractors are then placed to retract the lower rib cage in anterior, cephalad, and lateral directions. The falciform, triangular, and coronary ligaments are divided and the appropriate lobe is mobilized. The liver is then thoroughly examined by intraoperative ultrasound (see above). If the patient is to undergo a formal lobectomy, the caudate lobe is mobilized from the vena cava by ligation of the bridging veins. The portal pedicle is then divided, taking great care to avoid encroachment on the bile duct and portal vein bifurcations. Attempts are then made at extrahepatic isolation and division of the lobar hepatic vein. This maneuver is potentially dangerous, and intrahepatic division of the hepatic veins should be pursued if any trouble is encountered. The parenchyma is then cautiously transected. A number of transection techniques are available, but use of the Cavitron Ultrasonic Surgical Aspirator is preferred.

A partial lobectomy or non-anatomic resection is frequently used, particularly in patients with hepatic dysfunction from cirrhosis. After the liver has been mobilized, vascular inflow is occluded by intermittent clamping of the main portal triad. No formal division of lobar pedicles or lobar hepatic veins is required. The parenchyma is then divided using the Cavitron Ultrasonic Surgical Aspirator device. An attempt is made to limit inflow occlusion to less than 60 min for cirrhotic livers. Regardless of technique, the principal goal during parenchymal transection is to maintain more than a 1-cm margin between the tumor and the transection margin.

Liver transplantation

Orthotopic liver transplantation has potential advantages over liver resection, including correction of the underlying liver dysfunction/cirrhosis (at least transiently in the case of viral hepatitis) and removal of the entire organ at risk for development of metastatic or recurrent primary hepatocellular carcinoma. Results from transplantation have rivaled those of resection, with survival rates ranging from 40 to 50 per cent in series with 3-year follow-up data and 15 to 35 per cent in series with 5-year data (Table 4). Recurrence rates within the transplanted liver are approximately one-half of those seen with liver resection. The results of transplantation eclipse those of resection in the patient with a small T1 (80 per cent 3-year survival) or T2 (65 per cent 3-year survival) hepatocellular carcinoma and cirrhosis. This improvement in survival over resection is due to the abovementioned factors. Overall, a number of comparative studies have shown a significant survival advantage of transplantation over resection for patients with smaller tumors in conjunction with cirrhosis. This advantage is not seen for patients without cirrhosis. An additional indication for transplantation, as treatment of choice, is in patients with technically unresectable or recurrent hepatocellular carcinoma. Direct comparison of results from

transplantation with resection is complicated by the use of aggressive multiagent chemotherapy in the liver transplantation group. The efficacy of chemotherapy in this setting is promising, but lacks confirmation in prospective randomized trials. More widespread adoption of liver transplantation, solely as a means of treating hepatocellular carcinoma, is limited by the high cost, 'poor' results of transplantation for hepatocellular carcinoma when compared with transplantation for benign diseases, and the scarcity of donor livers. A current practice is to transplant patients with hepatocellular carcinoma and cirrhosis for hepatic dysfunction if the following conditions are met: the hepatocellular carcinoma is solitary and 5 cm or less in size or there are up to three tumor nodules (each up to 3 to 5 cm). Overall, liver transplantation for hepatocellular carcinoma remains an unsettled issue. In the future, if improved immunomodulatory agents or genetically engineered xenografts become readily available, there may be an expansion of transplants for the treatment of hepatocellular carcinoma.

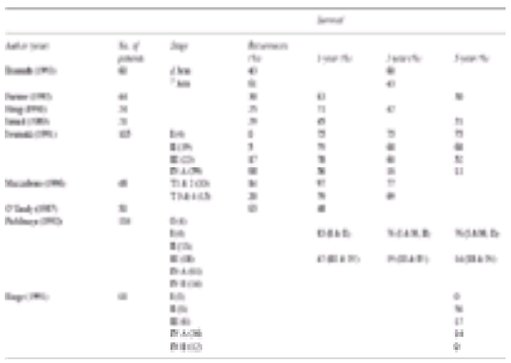


Table 4 Results of liver transplantation for hepatocellular carcinoma

Ablation

Ablative techniques to treat hepatocellular carcinoma range from simple (alcohol injection) to exotic (interstitial radiofrequency), but have as a common goal the direct mechanical destruction of all cells within the ablated field. Hepatocellular carcinoma ablation by percutaneous ethanol injection has gained widespread popularity, especially in Europe and Asia, due to its excellent efficacy coupled with simple administration and low cost. Most investigators report excellent disease-free survival following treatment of tumors of 3 cm or less, with results only surpassed by surgical resection for patients with excellent liver function. While more than 90 per cent necrosis may be obtained with these small tumors, the efficacy decreases with increasing tumor size, due to the exponential increase in the volume of ethanol needed for tumor treatment [$V = 4/3\ p\ (r + 0.5)^3$] and the heterogeneous dispersion of ethanol within the tumor. In the United States, enthusiasm for percutaneous ethanol injection has been dampened by diminishing patient compliance with repeated biopsies and sequential injections; the inability of radiologic studies to distinguish accurately between scar, necrotic tumor, and viable tumor; and a complication rate of up to 35 per cent (major and minor). Currently, percutaneous ethanol injection is useful for tumors of less than 4 cm in patients with Child's class B or C liver function.

Cryoablation is currently available in clinical practice and used widely for treatment of unresectable liver metastases. This technique uses super-cooled liquid nitrogen to create a freeze zone surrounding a probe placed into the hepatic parenchyma under ultrasound guidance (see details in the cryoablation and radiofrequency ablation section of colorectal liver metastases). The advantage of cryoablation over percutaneous ethanol injection is that much larger tumors (up to 6 cm) can be treated with good local control. A recent study from China reported a 5-year survival of 38 per cent for patients with tumors of less than 5 cm. Because multiple treatment modalities were used in this group of patients, the contribution of cryotherapy alone is difficult to assess. Unfortunately, cryoablation requires a laparotomy for safe conduct of the ablation. The need for open laparotomy may be averted with the developing technique of radiofrequency ablation. Radiofrequency ablation uses microwave energy radiated from a probe placed in the parenchyma of the liver, also under ultrasound guidance. The radiofrequency technique lends itself to a minimally invasive (laparoscopic) or percutaneous approach. Because of its ease of use and significantly lower cost, when compared with cryoablation, and its potential use for larger tumors inadequately treated by percutaneous ethanol injection, radiofrequency ablation may prove to be superior to the other ablative techniques, however more experience with this technique is required before it can be recommended further.

Embolization and chemoembolization

The vascular supply to a hepatocellular carcinoma, in significant contrast to the normal surrounding liver, is almost completely derived from the hepatic artery. This fact can be exploited to produce ischemic necrosis of the hepatocellular carcinoma following occlusion of its arterial branches. Occlusion is achieved by transcatheter embolization of segmental feeding vessels using gelfoam or metal coils and enhanced by concurrent administration of lipiodol, which adheres preferentially to (or is unable to be cleared from) the tumor endothelium within the hepatocellular carcinoma. Embolization can be repeated on a monthly schedule based on the degree of lipiodol retention in the tumor and according to the tumor response. The addition of chemotherapeutic agents, including doxorubicin, cisplatin, and mitomycin C, to lipiodol was shown to improve the response to embolization in several non-randomized trials. In contrast, a recent prospective randomized French trial and meta-analysis of smaller trials failed to reveal a statistical benefit of chemoembolization over simple embolization (or supportive therapy). One reason for the lack of effect in the French study was the high incidence of hepatic failure after chemoembolization. Unfortunately, the incidence of liver failure following chemoembolization, as reported in the literature, is widely variable (4 to 60 per cent) and the threat of this post-procedural complication is difficult to assess. Despite the results of the French trial and meta-analysis, many investigators still feel that the addition of chemotherapy to embolization may enhance its effects. Further prospective randomized trials, stratified for tumor stage and patient performance characteristics, will be needed if accurate conclusions are to be drawn. Side-effects from the embolization are common and include abdominal pain and tenderness, anorexia, nausea and vomiting, and fever. These symptoms and signs usually resolve within 2 to 3 weeks, although lasting effects may linger for several months. Patients with significant intrahepatic venous shunting can have an extrahepatic distribution of chemotherapy and may experience side-effects for this reason. Patients who may benefit from embolization include those with moderate liver dysfunction (Child's class B) who are not candidates for resection/cryoablation because of liver dysfunction or percutaneous ethanol injection because of tumor size. Patients who have occlusion of their main portal vein or main lobar trunks are not candidates for embolization.

Chemotherapy

The results of systemic chemotherapy for hepatocellular carcinoma have been disappointing. Doxorubicin has been shown to be the most effective single agent against hepatocellular carcinoma and the majority of randomized trials have used it alone or in combination with other agents. These trials have shown response rates in the 20 per cent range with little increase in median survival. The addition of other agents to doxorubicin has not significantly improved the response rates or the duration of response. There was a brief flurry of interest over estrogen-receptor blockade by tamoxifen, however the benefit of this drug was not supported by follow-on randomized controlled trials. At present, there is little justification for the use of chemotherapy for patients with hepatocellular carcinoma outside the setting of a clinical trial.

Radiotherapy

Although used in the treatment of hepatocellular carcinoma, external beam radiotherapy is limited by the lack of differential radiosensitivity between tumor and normal liver. Several small studies have shown poor results with the use of external beam radiotherapy for palliation. The administration of transcatheter radiopharamaceuticals including ^{131}I lipiodol or ^{90}Y microspheres has been described, but these series are small and recommendations are still pending.

Gene therapy

With the expansion of molecular biology techniques, the presence of specific intercellular genetic abnormalities within the hepatocellular carcinoma cells has been identified. At present, the most frequently identified genetic abnormality is the loss of p53, a tumor suppressor gene. The percentage of hepatocellular carcinomas with p53 deletions was shown to be increased in areas with higher incidence rates. *In vitro* studies have shown that when p53 is reintroduced into hepatocellular carcinoma cells, the growth rate is slowed and programmed cell death, called apoptosis, will occur. The potential use of replacement gene therapy is currently being explored for patients with hepatocellular carcinoma.

Cholangiocarcinoma

Cholangiocarcinomas account for 10 per cent of all primary hepatic malignancies and are second to hepatocellular carcinoma in incidence. They may occur anywhere in the biliary tree; however, cholangiocarcinoma within the liver parenchyma (known as peripheral cholangiocarcinoma) accounts for up to 5 per cent of these tumors. Peripheral cholangiocarcinoma develops in the small intrahepatic bile ducts and is associated with etiologic factors including hepatobiliary parasites, particularly flukes,

and intrahepatic lithiasis. Peripheral cholangiocarcinoma is 10 times more common in Asia than in Western countries.

Although cholangiocarcinoma is usually considered one entity, peripheral cholangiocarcinoma has different clinical characteristics and tends to have a more aggressive biologic behavior than hilar cholangiocarcinoma. While hilar cholangiocarcinoma presents with obstructive jaundice, usually due to a small tumor within the common duct or lobar ducts, peripheral cholangiocarcinoma often presents with jaundice from multilevel intraparenchymal ductal obstruction late in the disease course. When compared with hepatocellular carcinoma or liver metastases, peripheral cholangiocarcinoma is far more diffusely infiltrative with a higher incidence of intrahepatic metastases. These characteristics are significant factors accounting for the low resectability rate of peripheral cholangiocarcinomas.

The evaluation of patients with a peripheral cholangiocarcinoma is similar to that described for patients with hepatocellular carcinoma. There are no tumor markers specific for peripheral cholangiocarcinoma. Accurate intrahepatic staging of peripheral cholangiocarcinoma can be extremely difficult, even with modern CT and MRI. Part of the difficulty lies with differentiation between infiltrative peripheral cholangiocarcinoma and hepatic congestion and fibrosis due to chronic biliary obstruction from peripheral cholangiocarcinoma.

Surgical resection is the mainstay of treatment for peripheral cholangiocarcinoma. The surgeon should be aware that intraoperative findings may require expansion or abandonment of the originally planned procedure in up to one-half of patients. Operative mortality rates range from 0 to 14 per cent. Operative morbidity is elevated due to hepatic fibrosis with atrophy and segmental cholangitis with abscess formation. Five-year survival rates range from 30 to 40 per cent. Poor outcome is associated with multiple tumors and incomplete resection. Liver transplantation can also be performed for peripheral cholangiocarcinoma. Five-year survival rates in the largest series range from 27 to 40 per cent. Poor outcome after transplantation was seen in patients with positive margins, multiple tumors, and lymph node involvement. Currently, transplantation is reserved for highly selected patients with a small single peripheral cholangiocarcinoma without associated malignant adenopathy.

Results from the palliative treatment of patients with chemotherapy have been disappointing. 5-Fluorouracil has been the mainstay of treatment with response rates ranging from 10 to 24 per cent. Use of other agents alone or in combination with 5-fluorouracil has shown a minimal effect or additive benefit. The combination of radiation with chemotherapy may have some promise for hilar cholangiocarcinoma but cannot be used for peripheral cholangiocarcinoma because of hepatic toxicity.

Hepatoblastoma

Hepatoblastoma is the most common primary hepatic malignancy in children. This rare tumor (1 per 100 000 children) is usually diagnosed before the age of 3 years. Males are affected twice as often as females. Common presenting symptoms include abdominal swelling or mass, anorexia, nausea, weight loss, and pain. Hepatomegaly, due to a mass, is commonly seen on physical examination. The serum a-fetoprotein level is elevated in 75 to 95 per cent of patients. Abdominal CT scans will show an isolated enhancing hepatic tumor in the majority of patients. Surgical resection is the mainstay of therapy. Complete resection is possible in up to 65 per cent of children and is associated with cure for 30 to 70 per cent. Preoperative chemotherapy with cisplatin, 5-fluorouracil, and vincristine has been shown to be effective for improving resectability and decreasing recurrence rates. Adjuvant chemotherapy following liver resection has shown promise, however the results are difficult to interpret because of small series size. Small series of liver transplantations for hepatoblastoma have also been promising, with 5-year survival rates of 50 per cent. Present recommendations are for surgical resection of early stage tumors. More advanced tumors should be considered for resection or transplantation and strong consideration for preoperative adjuvant chemotherapy should be entertained. Late-stage tumors should be treated initially with chemotherapy and a recommendation for resection or transplantation pending the tumor response.

Epithelioid hemangiendothelioma

Epithelioid hemangiendothelioma is a rare tumor arising from intrahepatic endothelium. These tumors are relatively indolent, with a clinical behavior between hemangioma and angiosarcoma. Epithelioid hemangiendothelioma may arise in a wide age range, with the average age at presentation of 50. There is a slight female predominance. Patients may have a history of vinyl chloride exposure. The preferred treatment is surgical resection. Patients with extensive epithelioid hemangiendothelioma may undergo liver transplantation. The survival of a small group of patients, undergoing transplantation, was 40 per cent at 5 years.

Hepatic sarcomas

Angiosarcomas are interesting not because they are rare and incurable, but mainly because their etiology is usually specific and obvious. They usually arise following injection of Thorotrast decades ago, following arsenic exposure of agricultural workers (especially in the vineyards of the Moselle Valley), following testosterone in boys with Fanconi syndrome, and after exposure to vinyl chloride (the monomer of many plastics). The peak incidence of angiosarcoma occurs in patients in their 60s and 70s. There is a pronounced male predominance. Most patients present with abdominal pain and hepatomegaly. Hepatic angiosarcomas are extremely aggressive with a median survival of 6 months. The results of surgical resection have been disappointing, as has the use of chemotherapy and radiation. Because most patients will present with advanced tumors, transplantation has been used in a small series of patients. Although the 2-year survival rate was 15 per cent, no patient survived more than 28 months after transplantation. Other rare hepatic sarcomas including leimyosarcomas and rhabdomyosarcomas may arise within the liver. They are commonly treated by surgical resection.

Hepatic adenomas

Hepatic adenomas are benign, but large adenomas may rupture and bleed thereby coming to surgical attention. Contemporary publicity has now alerted physicians and patients to the role of oral contraceptives in fostering the development of adenomas, although both pristine liver cell adenomas and the ruptured variety are infrequent. It is not yet determined whether hepatic adenomas progress to carcinoma. Although the risk is small, some adenomas, without doubt, undergo carcinomatous transformation. The majority of hepatic adenomas are situated within the right liver. No one can predict which adenomas will rupture, nor why. When adenomas in the right liver rupture, they are often deep seated and immediate partial hepatectomy, with its attendant risk, is sometimes required.

From a clinical viewpoint, any source of hormonal stimulation should be withdrawn from a man or woman receiving hormone treatment who is thought to have a hepatic adenoma. If the neoplastic mass regresses by 50 per cent over the next 3 months, no treatment is likely to be needed. Adenomas that do not regress upon stopping hormonal therapy should be removed. Because of the possibility of carcinomatous transformation, albeit small, imaging studies over 6-month intervals for a total of 5 years, then yearly *ad infinitum* should probably be pursued to assess unresected adenomas for malignant growth.

Hemangiomas

Although benign cavernous hemangiomas are frequent, they are rarely important. Those less than 8 cm in diameter seldom need treatment, except when an ill-advised biopsy provokes hemorrhage. Intervention is required only when the hemangioma (usually best diagnosed by MRI) causes symptoms of a mass lesion, displacing or irritating other organs, or if it ruptures. Leaks or rupture are uncommon, fewer then 50 cases of ruptured hemangioma have been recorded. All of those that ruptured were over 8 cm in diameter, although a few smaller ones have leaked. The common hemangioma is resectable along non-anatomic lines. Most, including giant hemangiomas, can and should be enucleated rather than resected. There is no cause for concern if the resection margin contains evidence of hemangioma, since they do not recur.

Liver metastases

Successful management of the patient with liver metastases is dependent on matching the appropriate treatment to the tumor and its stage. Liver metastases may arise as an isolated site of regional failure or herald the development of generalized metastatic spread. The key element is selection of that biologically privileged subset of patients with favorable isolated liver metastases who can be cured by aggressive therapy. For as yet undetermined reasons, liver-directed therapy, particularly resection, of isolated liver metastases from only a limited number of primary tumor sites has been shown to be beneficial. These sites include the colon, rectum, neuroendocrine (pancreas) and carcinoid, and highly selected patients with metastases from other primary tumors. Unfortunately, liver metastases from more common primary sites, such as the lung, breast, or pancreatic adenocarcinoma, are usually only a part of more widespread metastases and herald a poor prognosis. Aggressive surgical treatment for an apparent isolated liver metastasis from these primary sites is considered only in highly selected patients, as discussed below.

In general, survival rates following resection of liver metastases have improved over the preceding 10 years. The most significant reason for this improvement is better patient selection. Better selection is a direct reflection of more accurate tumor staging, resulting from technologic improvements in imaging studies and refinement of preoperative clinical selection criteria.

Colorectal liver metastases

The most frequent indication for surgical involvement in the management of a patient with liver metastases is in the setting of colorectal cancer. This is principally

because colorectal cancers, among the common gastrointestinal and extra-abdominal malignancies, have the highest incidence of liver metastases as the sole site of regional failure. In addition, the problem is commonplace with over 50 000 patients per year developing colorectal liver metastases in the United States alone. For these reasons, the biology, clinical features, and indications for a variety of therapeutic interventions and their results are best understood for liver metastases from colorectal cancer.

An aggressive and invasive (surgical) approach to this problem is reasonable only if it can significantly alter the tumor's natural history and better the results from less invasive (systemic chemotherapy) therapy. The 5-year survival rates following resection of colorectal liver metastases average from 25 to 49 per cent in modern series. Natural history studies show that the median survival of untreated colorectal liver metastases is from 5 to 10 months. If this data is analyzed more closely, the 5-year survival rate for patients with a solitary unresected (but potentially resectable) liver metastasis was 2 per cent. The 3-year survival for patients with unresected multiple metastases was 1 per cent. Tumor response rates from chemotherapy based on 5-fluorouracil range from 22 to 25 per cent. The median survival in these series ranges from 9 to 12 months with few patients surviving longer than 3 years. These data show that an aggressive surgical approach can both alter the natural history and significantly improve survival when compared with supportive care or systemic therapy for patients with colorectal liver metastasis.

Patient evaluation

The majority of patients referred for evaluation and management of their liver metastases will be diagnosed by a rising serum level of carcinoembryonic antigen or routine surveillance imaging, principally abdominal CT. Patients presenting with symptoms related to their metastatic disease or who have metastatic disease suspected because of symptoms and signs have a significantly poorer survival than asymptomatic patients do. Patients should undergo an initial evaluation including thorough history, physical examination, and routine laboratory studies including a complete blood count, platelet count, serum electrolytes, and ‘liver function tests'. A current serum level of carcinoembryonic antigen should also be collected. The chest should be imaged by routine posterioanterior and lateral radiographs. If these are abnormal, they should be followed by a chest CT. The routine preoperative use of chest CT is costly and not currently recommended. The patient should also have undergone a routine colonoscopy within the previous 6 months. Following this initial evaluation, the liver should be imaged by a biphasic spiral CT scan enhanced with intravenous contrast. This scan uses a different contrast technique than routine CT scans enhanced with intravenous contrast that are performed for screening. These routine screening CTs are not adequate for definitive preoperative staging of the liver. Ill-defined liver abnormalities can be further characterized by ferumoxide-enhanced MRI. Angiographically assisted CT should be obtained if there remain any lingering questions about parenchymal abnormalities or in the rare circumstance that the anatomic detail provided by the biphasic CT/MRI is not adequate. Angiographically assisted CT can also be used to confirm the findings from biphasic CT during the initial institutional experience with this study. A celiac and superior mesenteric artery ‘angiogram' is obtained from vascular reconstruction of the CT's biphasic data set.

The definitive staging of intra-abdominal extrahepatic metastasis remains controversial. CT and MRI are marginally adequate studies, with sensitivities for detection of extrahepatic metastasis less than 60 per cent. ^{18}F-fluorodeoxyglucose PET imaging holds great promise, and should be considered if there is access to a linear accelerator. PET provides definitive evidence of metastatic disease if a ‘hot spot' located by PET corresponds to a defined anatomic abnormality on CT or MRI. Unfortunately, the cost associated with PET can be prohibitive and will limit its widespread application. Radionuclide-tagged monoclonal antibody scans have been used with variable success. The efficacy of tagged antibody scans needs to be shown in a larger patient population before recommendations for their use can be made.

Patient selection

The majority of patients with metastatic liver disease, referred for surgical opinion, will be accompanied by a screening abdominal CT scan. The next step in an initial evaluation is assessment of the adequacy of the screening CT. Cardiology and pulmonary consultations are obtained for older patients with significant cardiopulmonary disease. At this point, patients with severe comorbid illnesses and documented extrahepatic tumor are excluded from further consideration of liver-directed therapy. Depending on physician preference and institutional expertise, patients will then undergo a radiologic evaluation as detailed above. Following these definitive staging procedures, final treatment decisions are made. Patients with well-controlled comorbid illnesses, no definitive evidence of extrahepatic tumor, and four or fewer metastases arrayed within the liver so that they can be resected without compromise of hepatic function are scheduled for resection. Patients with well-controlled comorbid illnesses including mild hepatic dysfunction, no definitive evidence of extrahepatic tumor, and six or fewer metastases arrayed within the liver so that they could not be resected without compromise of hepatic function are scheduled for cryoablation/resection or radiofrequency ablation/resection. Patients with well-controlled comorbid illnesses, no definitive evidence of extrahepatic tumor, but more than six metastases within the liver are referred for systemic chemotherapy or are considered for hepatic artery infusion therapy.

Surgical resection

The principles and conduct of surgical resection for metastatic cancer are very similar to those described for treatment of hepatocellular carcinoma. Again, the underlying goal of resection is to obtain a greater than 1 cm resection margin of normal liver around the metastases. The type of resection is geared to that goal, and there is no clearly demonstrated superiority of larger (lobectomies) over more limited resections.

Surgical resection in the modern era can be performed safely with acceptable morbidity and mortality. In the centers with a large experience, the operative mortality is 2 to 4 per cent. The complication rates range from 15 to 50 per cent and include hemorrhage, liver dysfunction or failure, biliary leak or fistula, and cardiac and pulmonary complications. Median hospital stays following liver resection average approximately 2 weeks; however, the length of stay is decreasing, with stays of 1 week now common.

The 5-year survival rates from pooled large series range from 25 to 40 per cent (Table 5). The series with lower survival rates tend to be older with less accurate imaging technology, and less stringent patient selection criteria. Using current imaging technology and careful patient selection, 5-year survival rates as high as 49 per cent are seen. The 10-year survival rates drop to approximately 20 per cent, but these are older series and an improvement should be seen as the current 5-year data matures.

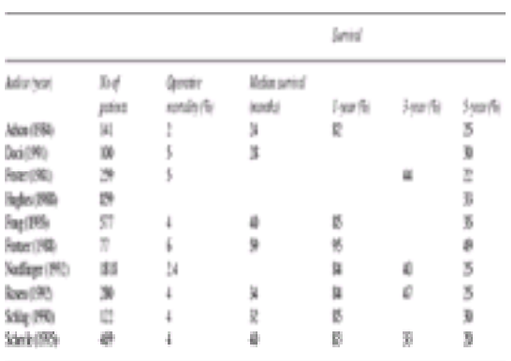


Table 5 Results of liver resection for metastatic colorectal carcinoma

Determinants of prognosis after resection

Data from large resection series has been subjected to statistical analysis so that associations with survival could be identified. As expected, the results are not uniform between series but comparisons show a number of similarities. Most, if not all, series show poor long-term outcomes for patients with the following tumor-related factors: extrahepatic metastases, particularly malignant portal adenopathy; intrahepatic ‘satellite' metastases; symptomatic presentation; tumors larger than 10 cm; more than 50 per cent of the liver involved with tumor; more than four metastases; and synchronous presentation of liver metastases and the primary tumor. Poor outcomes were also seen within the following intervention-related factors: positive margin or margin less than 1 cm following resection and perioperative blood transfusion.

Adjuvant systemic chemotherapy, following liver resection, has been used without the benefit of a prospective randomized trial. No benefit has been shown in retrospective reviews of chemotherapy based on 5-fluorouracil used in this setting. A small trial has shown that adjuvant hepatic intra-arterial chemotherapy was beneficial, but the numbers are too small to permit any definitive conclusion. The results of a larger American trial comparing observation with adjuvant intra-arterial fluorodeoxyuridine and systemic 5-fluorouracil following liver resection are currently pending. In the best series, 55 to 60 per cent of patients will develop recurrent colorectal cancer within 5 years. Examination of recurrences in these patients show that half will recur in the liver, with 85 per cent of these recurrences isolated to the liver. Other common sites of recurrence include the abdomen and pelvis (25 per cent) and lungs (25 per cent). Less than 5 per cent of recurrences develop in the bone

or brain. The majority of recurrences will occur within 2 years following resection. By considering the pattern and time course of recurrences, a postoperative surveillance program focusing on the abdomen, particularly the liver, and the lungs, which is most intense during the initial 2 postoperative years, is both logical and fruitful.

A measure of the relevance of any postoperative surveillance program is its ability to find recurrences that are treatable. Patients with isolated liver recurrences, detected at follow-up, are re-evaluated as described above and re-resected if they meet the criteria. Series of re-resections have shown that patients can be treated safely, with 5-year survival rates (25 to 30 per cent) rivaling those of the initial resection.

Hepatic artery infusion chemotherapy

Hepatic artery infusion chemotherapy exploits the preferential blood supply to neoplastic lesions in the liver. Thus, hepatic artery infusion theoretically increases the exposure of neoplastic cells to chemotherapy relative to normal hepatocytes. Regional therapy to the liver has several advantages over systemic therapy. Intra-arterial infusions can produce elevated hepatic drug concentrations relative to systemic levels for agents that have a high hepatic extraction. This characteristic results in high local exposure to the drug while limiting systemic toxicity. This situation is advantageous for drugs with steep dose–response rates, such as 5-fluorouracil and fluorodeoxyuridine, where higher drug concentrations result in increased measurable responses. Direct arterial perfusion of active agents such as fluorodeoxyuridine allows smaller infusion volumes to be used and facilitates treatment via implantable drug delivery systems. Regional hepatic chemotherapy for patients without evidence of extrahepatic disease may slow the development of systemic disease. One theory of colorectal carcinoma metastases postulates that there is a stepwise progression. Malignant cells initially spread via lymphatic and hematogenous routes to the liver, and cells from these metastases subsequently disseminate to the lung and the systemic circulation.

5-Fluorouracil and fluorodeoxyuridine have been the agents primarily used for intra-arterial therapy. Fluorodeoxyuridine appears to be the best agent for this use, since it has a high first-pass clearance rate with a hepatic extraction ratio of 0.7 to 0.9. The high first-pass clearance results in hepatic drug levels that are 400 times higher than systemic values. Pilot studies, using percutaneous catheters to deliver the drug continuously for 1 to 2 weeks, were able to show response rates as high as 30 to 50 per cent. With the development of an implantable infusion device, it became possible to deliver concentrated and prolonged infusions in the ambulatory setting.

Several randomized phase III trials have been performed comparing hepatic artery infusion with systemic chemotherapy (Table 6). In general, these reports demonstrate significantly higher complete and partial response rates in patients receiving intrahepatic infusions relative to systemic chemotherapy. None of the American studies has shown improvement in overall survival. The inability to show a survival advantage in these trials probably resulted from the enrollment of small numbers of patients and the practice of ‘crossing-over' patients to hepatic artery infusion if they failed to respond to systemic therapy. The two largest American studies allowed patient crossover to hepatic artery infusion therapy after failure of systemic therapy, thereby confounding the impact of hepatic artery infusion therapy on survival. Both studies did demonstrate a significant increase in survival for the crossover group compared with those who never received hepatic artery infusion. Certain trials also included patients with extrahepatic disease who were unlikely to derive any survival benefit from liver-directed therapy. In addition, many of these trials were started before the nuances of surgical pump placement and the significance of hepatobiliary toxicity due to fluorodeoxyuridine were appreciated. Treatment-limiting toxicities were seen in the hepatic artery infusion arms of many of the trials. For example, in the Mayo trial, nearly one-third of the patients randomized to hepatic artery infusion therapy never received any significant treatment. One recent meta-analysis of six randomized studies suggested small but significant improvements in 1- and 2-year survival rates for patients receiving hepatic artery infusion therapy. Two European studies demonstrated improved survival in patients treated with hepatic artery infusion; however, less than half of the patients in the control arms received systemic chemotherapy. Thus, definitive statements concerning the survival benefit of hepatic artery infusion over systemic therapy cannot yet be made. Some investigators, however, have clearly demonstrated improved symptom control and quality of life with hepatic artery infusion compared with systemic chemotherapy and supportive care. The toxicities of hepatic artery infusion noted in these trials include direct hepatic toxicity, biliary sclerosis, and gastric/duodenal irritation and ulceration. Systemic effects such as myelosuppression and mucositis are rarely seen after infusion therapy. Nausea, vomiting, and diarrhea may be noted following hepatic artery infusion but are attributable to local gastrointestinal inflammation rather than systemic effects.

Group	No. of patients	Response rate (%)		Survival (months)	
		HAI	Systemic	HAI	Systemic
MSKCC	162	52	20	18	12
NCOG	143	42	10	15.5	16
NCI	64	62	17	20	11
Mayo	69	48	21	12.5	10.5
Consortium	45	58	38	–	–
City of Hope	41	56	0	–	–
French	163	49	14	15	11
English	106	50	0	17	6.3

MSKCC, Memorial Sloan-Kettering Cancer Center; NCOG, Northern California Oncology Group; NCI, National Cancer Institute (US).

Table 6 Results of randomized trials of hepatic intra-arterial (HAI) compared with systemic chemotherapy for colorectal liver metastasis

One approach to decrease toxicity has been to lower the fluorodeoxyuridine dose. A phase II trial explored this option, decreasing the dose from 0.3 mg/kg.day to 0.1 mg/kg.day. Measured responses were equivalent to those seen with conventional doses with a 50 per cent complete response rate and median survival of 22.4 months. None of the patients treated with the lower dose of fluorodeoxyuridine required treatment termination as a result of hepatic artery infusion toxicity.

A different approach was attempted with the goal of reducing toxicity while maintaining the fluorodeoxyuridine dose. A randomized trial using fluorodeoxyuridine with or without dexamethasone was performed. Although no significant decrease in toxicity was documented, patients in the dexamethasome arm were able to receive greater overall drug doses, which resulted in improved response rates, well over 50 per cent in previously untreated patients.

New approaches designed to improve hepatic artery infusion responses have also included the addition of fluoropyrimidine modulators, such as leucovorin and interferon-a. Encouraging phase II data have been reported with this combination, but the addition of leucovorin to intra-arterial fluorodeoxyuridine therapy increased the rate of hepatic toxicity. Concurrent dexamethasone has helped to reduce biliary toxicity in this setting from 21 to 3 per cent while maintaining good response rates. In a recently published trial, this combination of fluorodeoxyuridine, leucovorin, and Decadron led to an objective response rate of 70 per cent with a median survival of nearly 25 months. Studies comparing hepatic artery infusion therapy with fluorodeoxyuridine, leucovorin, and dexamethasone against systemic 5-fluorouracil and leucovorin are ongoing.

The recognition of several important surgical considerations is fundamental during hepatic artery pump placement if one is to achieve the goals of bilobar hepatic perfusion while avoiding misperfusion of the stomach and duodenum. Preoperative angiography is essential to define variations in hepatic artery anatomy, which are estimated to occur in up to 50 per cent of patients. Conventional intra-arterial biplanar angiography is currently recommended, although it will probably be replaced by vascular reconstructions from biphasic CT in the near future. Specific methods have been described to achieve complete hepatic perfusion via either the gastroduodenal artery or anomalous branches of the hepatic artery. Maneuvers including ligation of non-dominant vessels and placement of dual-lumen pump devices have been described. The surgeons placing these pumps must recognize when these techniques are called for if they are to be assured of optimal delivery of drug to the liver. Evidence of extrahepatic disease is sought during the initial exploration of the abdomen and should include sampling of any suspicious peritoneal or omental nodules and frozen-section biopsies of the porta hepatis nodes or any suspicious nodes in the celiac axis. If any extrahepatic tumor is identified, especially in the portal nodes, the procedure is terminated. Misperfusion of the stomach and duodenum is avoided by performing a meticulous dissection and ligation of the right gastric artery and the small branches of the common hepatic artery along the superior border of the antrum and proximal duodenum. A cholecystectomy is then performed, but care is taken to avoid injuring the axial blood supply to the distal common bile duct. In most situations, the gastroduodenal artery is cannulated approximately 2 cm below its origin off the hepatic artery, and the tip of the catheter is secured just distal to the origin of the vessel to avoid disrupting arterial flow to the liver. Bilobar liver perfusion and absence of flow to the stomach and duodenum are verified by injection of 5 ml of fluorescein and examination with a Wood's lamp.

Cryoablation and radiofrequency ablation

Cryoablation refers to the *in situ* destruction of liver tumors by precisely and rapidly cooling the tumor and a zone of normal hepatic parenchyma to extreme subzero temperatures. By circulating a coolant, such as liquid nitrogen, through the core of the tumor, temperatures as low as -100°C can be achieved. Such profound tissue hypothermia results in both direct and indirect mechanisms of cell destruction. Details of the lethal effects of cryoablation have been well described, but essentially involve the formation of extracellular ice crystals leading to protein denaturation and the rupture of cell membranes. Indirectly, subzero temperatures result in microvascular thrombosis and tissue anoxia. Further damage occurs if the cells are allowed to thaw slowly and are then rapidly refrozen. Clinical studies have

suggested that two such freeze–thaw cycles may be necessary for optimal tissue destruction. Fundamental to the implementation of cryoablation as a viable treatment option for patients with unresectable liver metastases has been the development of intraoperative ultrasonography and refinement of the equipment used to deliver the coolant to the tumor.

Early reports demonstrated the feasibility and safety of treating liver metastases by cryoablation, while follow-up series have now helped to define the indications for this modality. The most significant clinical limitation of surgical resection is usually the number of lesions that can be safely removed while sparing sufficient hepatic parenchyma to avoid postoperative liver failure. Thus, the major indications for cryoablation are metastatic lesions that are not amenable to resection due to liver dysfunction from cirrhosis or because the number of lesions or the location of the tumors, usually bilobar, would risk insufficient postoperative hepatic function. For colorectal metastases, most centers with sufficient cryosurgical experience will limit this technique to patients with fewer than six metastases (generally less than 40 per cent of the liver volume), and tumors smaller than 6 to 8 cm in greatest diameter. Most surgeons will avoid large, central tumors near the hilum of the liver and major bile ducts, which are easily damaged by freezing.

Similarly to other forms of liver-directed therapy, cryoablation probably offers little advantage to patients with extrahepatic metastatic disease. The operative technique involves intraoperative ultrasound localization and monitoring of the cryoprobe placement into the tumor. In addition, intraoperative ultrasound allows detection of tumors not recognized on preoperative scans and is essential for monitoring the freeze margins of the ‘iceball'. The cryoprobe is a vacuum-insulated device that circulates supercooled liquid nitrogen through the probe's tip. A 3-mm blunt-tipped probe creates a freeze zone up to 4 cm, and the 8-mm trocar point probe creates a freeze zone of up to 6 cm. A single probe or combination of probes can be utilized to achieve complete freezing of the tumor and an additional 1 cm beyond the margin to ensure complete cryoablation. The entire process of probe introduction, freeze–thaw cycles, and probe extraction is monitored with real-time intraoperative ultrasound.

Major complication rates are reportedly between 10 and 40 per cent, and mortality rates are 2 per cent. Complications may include hemorrhage, biliary fistulas, hepatic or subphrenic abscesses, and more rarely, hepatic failure, coagulopathy, and cardiac arrest. Disease-free survival rates and patterns of failure have now been reported from several centers. With median follow-up times of 18 to 36 months, most series report 5-year actuarial disease-free survival rates of 15 to 28 per cent. These rates are comparable with those reported for liver resection, and in light of the reported mortality rates of less than 2 per cent, it becomes obvious why cryoablation has generated such interest and enthusiasm. In one series, postoperative carcinoembryonic antigen changes were extremely predictive of ultimate outcome. For the group in which carcinoembryonic antigen levels returned to the normal range, median survival exceeded 3 years.

Tumor destruction due to thermal damage, as opposed to freezing, is the goal of radiofrequency ablation. The radiofrequency technique uses high-frequency alternating current that passes from an electrode into the tumor and surrounding tissue. With application of an alternating current, ions within the tumor attempt to follow the change in direction of the current, resulting in frictional heating and coagulation necrosis within the tumor. Radiofrequency ablation uses monopolar or bipolar needle electrodes placed into the tumor under radiologic guidance. These electrodes can be placed in the operating room under ultrasound guidance by minimally invasive (laparoscopic) techniques or percutaneously by using CT or ultrasound-directed imaging. Preliminary results from treatment of hepatic tumors have been promising. These techniques can be safely applied, with mortality rates under 1 per cent and morbidity rates under 3 per cent. Radiofrequency ablation also appears to be effective, with local recurrence rates after treatment between 2 and 3 per cent in larger series. Careful attention must be paid to tumor size in order to prevent local recurrence. Studies have shown that as tumor size increases over 3 cm, the local recurrence rates rise sharply. The adequacy of treating larger liver tumors by multiple ablations is unproved and ongoing improvements in radiofrequency technology are directed towards treatment of these larger tumors. What is clear from these early studies is the substantially lower morbidity and mortality rates from radiofrequency ablation compared with cryoablation. These are principally due to the need for laparotomy associated with cryoablation. At present, though, cryoablation is superior for tumors ranging in size between 3 and 6 cm.

Systemic chemotherapy

An in-depth discussion of systemic chemotherapy for metastatic colon cancer is beyond the scope of this chapter and readers are referred elsewhere. Despite two decades of intense clinical investigation, cancer metastatic to the liver remains relatively resistant to systemic chemotherapy. Only the fluoropyrimidines (5-fluorouracil and fluorodeoxyuridine) and irinotecan, a topoisomerase I inhibitor, have been shown to have consistent therapeutic activity, and numerous studies have not demonstrated other agents to be superior to these drugs. The best and most consistent results with 5-fluorouracil have been obtained when it is combined with the folate modulator leucovorin. Response rates for this combination range from 10 to 60 per cent with the average partial response rate of 30 to 40 per cent and complete response rate of under 5 per cent. Median survival from this combination therapy is from 12 to 14 months. Irinotecan is useful as a salvage therapy by virtue of its 15 to 25 per cent response rate in patients with progression of tumor on 5-fluorouracil. Wider use of irinotecan has been limited by its substantial gastrointestinal and hematologic toxicity. Overall, chemotherapy is purely palliative and a poor substitute for the potentially curative therapies discussed previously.

Embolization

Liver metastases can be treated by embolization of the blood supply. A cannula is introduced into the femoral artery and advanced until the celiac axis has been cannulated. An angiogram is then performed and the vessels supplying the metastases identified. If possible, these vessels are then superselectively cannulated. Absolute alcohol, gelfoam, or steel coils may be introduced to occlude the blood supply. Embolization is most effective for patients with hepatic metastases from neuroendocrine tumors and sarcomas because of their rich blood supply and the resultant reduction of symptoms attributable to the secretion of a hormonally active substance from neuroendocrine tumors. Embolization has not been shown to improve survival of patients with metastases from colorectal carcinoma and other gastrointestinal malignancies.

Radiotherapy

The efficacy of radiotherapy in the treatment of metastases of the liver is limited by the radiosensitivity of normal hepatocytes and the relative insensitivity of metastatic carcinoma. The maximum dose that can be administered before inducing radiation hepatitis is 35 Gy; this dose is too low for curative irradiation of metastatic carcinoma. Radiotherapy is occasionally useful in the palliation of severe pain from hepatic metastases, but does not prolong survival.

Gene therapy

Using modern molecular techniques, many of the genetic abnormalities in cancer are now partially understood. In general, both genes that potentiate the malignant phenotype, called oncogenes, are functioning and/or genes that suppress the malignant phenotype, called tumor suppressor genes, are lost within tumor cells. Gene therapy strategies that target and inhibit oncogene function or replace lost tumor suppressor gene function are being developed. Liver metastases are potentially good targets for gene therapy, since the gene can be delivered directly into the tumor via the hepatic artery. A key field of investigation is focused on the tumor suppressor gene p53, which is mutated in up to 50 per cent of patients with colorectal cancer. When p53 is reintroduced into p53 mutant tumor cells, these cells undergo programmed cell death called apoptosis. Ongoing studies are examining the possible intra-arterial delivery of ‘normal' p53 into hepatic metastases with mutated p53. This and other gene therapy trials will be expanding rapidly as the clinical use of gene therapy moves more into mainstream medicine.

Neuroendocrine (including carcinoid) metastases

Both primary gastrointestinal neuroendocrine and carcinoid malignancies metastasize to the liver as the most common site of extranodal dissemination. Metastases from these malignancies pose unique features that make them excellent candidates for liver-directed therapy. These features include: a less aggressive behavior manifested as a more indolent clinical course and prolonged survival when compared with hepatic metastases from adenocarcinoma; and endocrinopathy, related to the secretion of active hormones from liver metastases directly into the systemic circulation. In general, the severity of endocrinopathy is directly related to the hepatic tumor burden.

The evaluation of patients with neuroendocrine metastases is similar to that described for patients with colorectal cancer. An exception to the similarity is the evaluation of tumor markers. Since endocrinopathy from ectopic hormone oversecretion is common, the selective sampling of serum hormone levels is indicated. The selection of hormones to be sampled should be based on suggestive clinical symptoms, signs, and routine laboratory studies. The use of imaging studies is also similar to that described above, with the exception of ^{18}F-fluorodeoxyglucose PET imaging, which remains experimental for neuroendocrine tumors. Somatostatin receptor scintigraphy plays the same role in the management of neuroendocrine metastases as does ^{18}F-fluorodeoxyglucose PET imaging in the management of colorectal metastases. Overall, the sensitivity for detection of intrahepatic and extrahepatic metastases from neuroendocrine tumors by somatostatin receptor scintigraphy has been shown to be from 85 to 92 per cent. The specificity for detection of intrahepatic metastases by somatostatin receptor scintigraphy has also been shown to be as high as 93 per cent. These data support the contention that somatostatin receptor scintigraphy has a potentially important role in the imaging of neuroendocrine malignancies, particularly for the detection of extrahepatic and occult intrahepatic metastases.

Patient selection is less well defined for patients with neuroendocrine liver metastases when compared with colorectal carcinoma. In general, patients with neuroendocrine liver metastases that are potentially resectable should be resected. The 5-year survival rates following resection for this tumor are excellent, with an average of 70 per cent. The survival rate is double when compared with patients who are not resected. Equally as important, resolution of the associated

endocrinopathy was seen in 90 to 95 per cent of these patients. Patients with small to moderate-sized low-volume metastases who are not candidates for resection, for anatomic reasons, should be considered for either cryoablation or radiofrequency ablation. For patients with disseminated liver metastases, which are not amenable to resection or ablation, hepatic transplantation is an alternative. In a recent review of 103 patients undergoing liver transplantation for neuroendocrine metastases there was a 5-year survival rate of 47 per cent. The disease-free survival in this group was unfortunately only 24 per cent. Those patients who did well were under the age of 50 with metastases isolated to the liver. These data support the conclusion that transplantation can provide good long-term palliation and occasional cure in selected patients with hepatic metastases from neuroendocrine tumors.

Non-surgical palliative options have also been successful for treatment of neuroendocrine liver metastases. Hepatic artery emboliza-tion without chemotherapy has been able to produce an objective tumor response in from 50 to 60 per cent of patients, with a median duration of response of approximately 12 months and a 5-year survival rate of 60 per cent. The addition of chemotherapy, usually doxorubicin, mixed with lipiodol and coinjected with the occlusive substance improves the tumor response rate to 60 to 100 per cent and improves symptoms (from tumor and endocrinopathy) in 70 to 100 per cent of patients. Side-effects following embolization are common and are as described for treatment of hepatoma. The incidence of liver failure (10 to 15 per cent) is much lower than for hepatoma and is a reflection of normal liver function in patients with neuroendocrine metastases.

Neuroendocrine tumors have been relatively responsive to chemotherapy, particularly streptozotocin. Response rates for single-agent streptozotocin range from 40 to 60 per cent. The addition of doxorubicin to streptozotocin improves the response rates up to 70 per cent and the median duration of survival to 18 months. These results are slightly better than the combination of streptozotocin and 5-fluorouracil, except for gastrinoma. The combination of streptozotocin, 5-fluorouracil, and doxorubicin may produce the highest response rates, but the data from several series show a wide range of variation. Major side-effects from streptozotocin include renal dysfunction (20 to 70 per cent) and hepatic dysfunction (30 per cent). Hormonal therapy with the somatostatin analog octreotide is effective in controlling symptomatic endocrinopathy in most patients. This drug has some therapeutic effect with tumor stabilization, seen in up to 50 per cent of patients.

Liver metastases from other primary tumors

Liver metastases from primary malignancies, others than colorectal and neuroendocrine, are usually a part of a systemic metastatic dissemination and rarely the focus of liver-directed therapy. On those rare occasions, surgical opinion will be sought about a patient with an excellent performance status who has low-volume metastases confined to the liver. The data on these patients are limited to a few small retrospective studies of very highly selected patients. In these studies, the overall 5-year survival rates range from 15 to 37 per cent. Analysis of prognostic criteria is difficult, although patients who underwent a curative resection, had a single metastasis, had no extrahepatic metastases, and presented with metachronous not synchronous metastasis had a better survival. In general, treatment of a single or low-volume metastasis may be considered in highly selected patients. Selection criteria include: the patient has been or is being treated with standard therapy for the particular metastasis; the metastasis has been stable in size or slowly growing over the period of chemotherapy or observation; there is no evidence of extrahepatic metastases following thorough clinical and radiologic evaluation; the patient has an excellent performance status; and the liver metastasis will determine patient survival far in advance of other potential metastatic disease. The relative roles of resection and ablation are not well defined, therefore resection should be pursued whenever possible.

Of the tumor types resected, patients with Wilm's tumors, renal cell carcinoma, and testicular tumors had the best results. Patients with breast carcinoma, sarcomas, adrenocortical carcinoma, and gynecologic cancers had less favorable results, but appeared to do better than those patients with gastrointestinal primaries, particularly those of the stomach and pancreas.

Patients with malignant gastrointestinal stromal tumors (gastro-intestinal sarcomas) commonly present with metastases confined to the liver. Although resection in these patients is palliative, it may extend their median survival up to 5 years and is indicated, given the variable natural history of this tumor and lack of effective alternative therapy.

Fifty per cent of patients with uveal melanoma may present with liver metastases as the sole or initial site of dissemination. Although the median survival of these patients is only 5 months, selected patients benefit from liver-directed therapy. The majority of patients are treated with chemoembolization and a 36 per cent response rate has been shown using cisplatin-based regimens. Palliative resection may be considered for those patients with large symptomatic metastases or who are not candidates for chemoembolization. Prolonged survival following resection has been seen in rare anecdotal reports.

Further reading

Adson MA, van Heerden JA, Adson MH, Wagner JS, Ilstrup DM. Resection of hepatic metastasis from colorectal cancer. *Archives of Surgery* 1984; **119**: 647–51. [Older series (*n* = 141) of hepatic resections with 25 per cent 5-year survival rate.]

Advanced Colorectal Cancer Meta-Analysis Project. Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: Evidence in terms of response rate. *Journal of Clinical Oncology* 1992; **10**: 896–903. [Meta-analysis of experience with first-line chemotherapy for advanced colorectal cancer.]

Baron RL. Detection of liver neoplasms: techniques and outcome. *Abdominal Imaging* 1994; **19**: 320–4. [Review of imaging for liver tumors.]

Berdah SV, Delpero JR, Garcia S, Harwigsen J, Le Treut YP. A Western surgical experience of peripheral cholangiocarcinoma. *British Journal of Surgery* 1996; **83**: 1517–21. [Review of European experience (*n* = 31) with resection for peripheral cholangiocarcinoma showing 32 per cent actuarial 5-year survival.]

Berney T, Mentha G, Roth AD, Morel P. Results of surgical resection of liver metastasis from non-colorectal primaries. *British Journal of Surgery* 1998; **85**: 1423–7. [Current morbidity and mortality justify use in highly selected patients.]

Bhattacharya S, Novell JR, Winslett MC, Hobbs KE. Iodized oil in the treatment of hepatocellular carcinoma. *British Journal of Surgery* 1994; **81**: 1563–71. [Review of lipiodol uses for diagnosis and therapy in hepatocellular carcinoma.].

Bismuth H, Chiche L. Surgical anatomy and anatomical surgery of the liver. In: Blumgart LH, ed., *Surgery of the liver and biliary tract*, pp. 3–9. Churchill Livingstone, London, 1994. [Excellent descriptions of functional hepatic anatomy.]

Bismuth H, Chiche L, Adam R, Castaing D, Diamond T, Dennison A. Liver resection versus transplant for hepatocellular carcinoma in cirrhotic patients. *Annals of Surgery* 1993; **218**: 145–51. [Improved survival for transplant in patients with small and uninodular or binodular tumors.]

Blumgart LH, Fong Y. Surgical options in the treatment of hepatic metastasis from colorectal cancer. *Current Problems in Surgery* 1995; **32**: 346–55. [Comprehensive review of management options and pertinent details.]

Bruix J. Treatment of hepatocellular carcinoma. *Hepatology* 1997; **25**: 259–62. [Current overview of management options.]

Busuttil RW, Farmer DG. The surgical treatment of primary hepatobiliary malignancy. *Liver Transplantation and Surgery* 1996; **2**: 114–30. [Comprehensive review of management for all primary hepatic malignancies.]

Cady B *et al.* Surgical margin in hepatic resection for colorectal metastasis: a critical and improvable determinant of outcome. *Annals of Surgery* 1998; **227**: 566–71. [Review stressing importance of 1-cm surgical margin.]

Carr BI, Flickinger JC, Lotze MT. Hepatobiliary cancers. In: DeVita VT, Hellman S, Rosenberg SA, eds, *Cancer: principles and practice of oncology*, pp. 1087–114. Lippincott-Raven, Philadelphia, 1997. [Comprehensive review of all aspects of primary hepatic malignancies.]

Casavilla FA *et al.* Hepatic resection and transplantation for peripheral cholangiocarcinoma. *Journal of the American College of Surgeons* 1997; **185**: 429–36. [Institutional experience (American) with 54 patients showing hepatic resection and orthotopic liver transplant are effective therapies when the tumor can be removed with adequate margins, the lesion is singular, and lymph nodes are not involved.]

Chang AE, Schneider PD, Sugarbaker PH, Simpson C, Culnane M, Steinberg SM. A prospective randomized trial of regional versus systemic continuous infusion 5-fluorodeoxyuridine chemotherapy in the treatment of colorectal liver metastasis. *Annals of Surgery* 1987; **206**: 685–93. [Trial showing statistically significant improvement in responses and survival (portal node negative patients) for hepatic artery infusion therapy.]

Chen H, Hardacre JM, Uzur A, Cameron JL, Choti MA. Isolated liver metastasis from neuroendocrine tumors: Does resection prolong survival? *Journal of the American College of Surgeons* 1998, **187**: 88–93. [Answer: Yes. Review showing excellent survival data with resection.]

Chen M-F, Jan Y-Y, Wang C-S, Jeng L-B, Hwang T-L. Clinical experience in 20 hepatic resections for peripheral cholangiocarcinoma. *Cancer* 1989; **64**: 2226–32. [Asian experience with resection of peripheral cholangiocarcinoma.]

Curley SA, Chase JL, Roh MS, Hohn DC. Technical considerations and complications associated with the placement of 180 implantable hepatic arterial perfusion devices. *Surgery* 1993; **114**: 928–34. [A how-to and how-not-to guide for hepatic artery infusion pump placement.]

Curley SA, Izzo F, Delrio P. Radiofrequency ablation of unresectable primary and metastatic hepatic malignancies: results in 123 patients. *Annals of Surgery* 1999; **230**: 1–8. [Large current series of

results for radiofrequency ablation in hepatocellular carcinoma and colorectal liver metastasis.]

Curley SA, Roh MS, Chase JL, Hohn DC. Adjuvant hepatic arterial infusion chemotherapy after curative resection of colorectal liver metastasis. *American Journal of Surgery* 1993; **166**: 743–6. [First series to show survival advantage of adjuvant hepatic artery infusion therapy after liver resection.]

Daneker GW. Liver. In: Wood WC, Skandalakis JE, eds, *Anatomic basis of tumor surgery*, pp. 504–45. Quality Medical Publishing, St. Louis, 1999. [Comprehensive review of hepatic anatomy and techniques for liver resection and hepatic artery infusion pump placement.]

Delbeke D *et al.* Staging recurrent metastatic colorectal carcinoma with PET. *Journal of Nuclear Medicine* 1997; **38**: 1196–201. [Prospective study showing greater than 90 per cent accuracy for detection of intrahepatic and extrahepatic colorectal metastasis.]

Doci R, Gennari L, Bignami P, Montalto F, Morabito A, Bozzetti F. One hundred patients with hepatic metastasis from colorectal cancer treated by resection: analysis and prognostic determinants. *British Journal of Surgery* 1991; **78**: 797–801. [Older resection series with 30 per cent actuarial 5-year survival and analysis of significant prognostic indicators.]

Elias D *et al.* Resection of liver metastasis from a non-colorectal primary: indications and results based on 147 monocentric patients: *Journal of the American College of Surgeons* 1998; **187**: 487–93. [Recommends resection of neuroendocrine, testicular, and renal metastasis, indications for resection of metastasis from other sites 'obscure'.]

Fischman AJ. Positron emission tomography in the clinical evaluation of metastatic cancer. *Journal of Clinical Oncology* 1996; **14**: 691–96. [An overview of PET technique and interpretation.]

Fong Y, Blumgart LH. Hepatic colorectal metastasis: current status of surgical therapy. *Oncology* 1988; **12**: 1489–503. [A concise update of their 1995 *Current Problems in Surgery* article.]

Fortner JG, Silva JS, Golbey RB, Cox EB, Mclean BJ. Multivariate analysis of a personal series of 247 consecutive patients with liver metastasis from colorectal cancer: I. Treatment by hepatic resection. *Annals of Surgery* 1984; **199**: 306–16. [Review of large personal series of resections.]

Fuhrman GM, Curley SA, Hohn DC, Roh MS. Improved survival after resection of colorectal liver metastasis. *Annals of Surgical Oncology* 1995; **2**: 537–41. [Modern era resection series with 44 per cent actuarial 5-year survival rate.]

Groupe d'Etude et de Traitment du Carcinome Hepatocellulaire. A comparison of lipiodol chemoembolization and conservative treatment for unresectable hepatocellular carcinoma. *New England Journal of Medicine* 1995; **332**: 1256–61. [In patients with hepatocellular carcinoma and cirrhosis, chemoembolization reduced tumor growth, caused acute liver failure, and did not improve survival over conservative management.]

Harrison LE *et al.* Hepatic resection for noncolorectal nonneuroendocrine metastasis: a 15-year experience with 96 patients. *Surgery* 1997; **121**: 625–32. [Primary tumor type, disease-free interval, and curative resection predict benefit.]

Hughes K *et al.* Resection of the liver for colorectal carcinoma metastasis: a multi-institutional study of patterns of recurrence. *Surgery* 1986; **100**: 278–84. [Registry of Hepatic Metastasis data on 607 patients concluding that: hepatic resection is effective, adjuvant therapy after resection should be directed at both the lung and liver, and positive pathologic margins or bilobar metastases indicate an increased risk for hepatic recurrence.]

Ichida T, Van Thiel DH, Hassanein T. The medical management of hepatocellular carcinoma (HCC) in Japan: a review with implications for hepatocellular carcinoma seen in the West. *Hepato-Gastroenterology* 1996; **43**: 1575–83. [Concise review of selection, results, and algorithm for medical management of hepatocellular carcinoma.]

Ikeda D *et al.* A multivariate analysis of risk factors for hepatocellular carcinogenesis: a prospective observation of 795 patients with viral and alcoholic cirrhosis. *Hepatology* 1993; **18**: 47–53.

Ishak KG, Malt RA. Sarcomas of the liver and spleen. In: Raaf JH, ed., *Management of soft tissue sarcomas*, pp. 165–80. Year Book Medical Publishers, Chicago, 1992.

Iwatsuki S *et al.* Hepatic resection versus transplantation for hepatocellular carcinoma. *Annals of Surgery* 1991; **214**: 221–9. [Improved survival rates for all stages of hepatocellular carcinoma with transplant over resection in patients with cirrhosis.]

Jaques DP, Coit DG, Casper ES, Brennan MF. Hepatic metastasis from soft tissue sarcoma. *Annals of Surgery* 1995; **221**: 392–7. [Following hepatic resection of 14 patients with high-grade leiomyosarcomas, all have had recurrences (most in the liver) with a median survival of 30 months.]

Johnson PJ. Why can't we cure primary liver cancer? *European Journal of Cancer* 1995; **31**: 1562–4. [Thoughtful editorial on current limitations of treatment for hepatocellular carcinoma.]

Kemeny N *et al.* Phase II study of hepatic arterial floxuridine, leucovorin, and dexamethasone for unresectable liver metastasis from colorectal carcinoma. *Journal of Clinical Oncology* 1994; **12**: 2288–95. [Report of currently most effective regimen for hepatic artery infusion with 78 per cent response rate in previously untreated and 52 per cent in previously treated patients.]

Lai EC, Fan ST, Lo CM, Chu KM, Liu CL, Wong J. Hepatic resection for hepatocellular carcinoma. An audit of 343 patients. *Annals of Surgery* 1995; **221**: 291–8. [A 22-year experience showing improved recent results due to changes in management strategies and technologic advances.]

Le Treut YP *et al.* Results of liver transplantation in the treatment of metastatic neuroendocrine tumors. A 31-case French multicentric report. *Annals of Surgery* 1997; **225**: 355–64. [Transplant can control symptomatic endocrinopathy and survival for patients with metastasis from carcinoids.]

Lewis KH, Chezmar JL. Hepatic metastasis. *MRI Clinics of North America* 1997; **5**: 319–30. [Review of MRI technique and contrast agents for metastasis.]

Liver Cancer Study Group of Japan. Primary liver cancer in Japan. Clinicopathologic features and results of surgical treatment. *Annals of Surgery* 1990; **211**: 277–87. [Review of 4765 cases of primary liver malignancy with careful review of demographics, clinical features, and surgical outcomes.]

Livraghi T, Lazzaroni S, Meloni F, Torzilli G, Vettori C. Intralesional ethanol for the treatment of unresectable liver cancer. *World Journal of Surgery* 1995; **19**: 801–6. [Series of 210 patients treated with percutaneous ethanol injection resulting in 41 per cent and 33 per cent 5-year survivals for patients with Child's class A disease with a single or multiple hepatocellular carcinoma, respectively (15 per cent local recurrence rate).]

Mazzaferro V *et al.* Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *New England Journal of Medicine* 1996; **334**: 693–9. [Liver transplantation is effective treatment for small, unresectable hepatocellular carcinoma in patients with cirrhosis.]

Meta-analysis Group in Cancer. Reappraisal of hepatic arterial infusion in the treatment of nonresectable liver metastasis from colorectal cancer. *Journal of the National Cancer Institute (USA)* 1996; **88**: 252–8. [Meta-analysis of randomized trials confirms that hepatic artery infusion can achieve much higher tumor response rates than systemic chemotherapy in patients with liver metastases from colorectal cancer.]

Moertel CG. Chemotherapy for colorectal cancer. *New England Journal of Medicine* 1994; **330**: 1136–41. [Review of 5-fluorouracil chemotherapy and biochemical modulation.]

Norton JA, Fraker DL. Tumor markers. In: Sabiston DC, ed., *Textbook of surgery*, pp. 491–507. WB Saunders, Philadelphia, 1991. [Comprehensive review of tumor markers and their utility in specific cancers.]

Olthoff KM *et al.* Adjuvant chemotherapy improves survival after liver transplantation for hepatocellular carcinoma. *Annals of Surgery* 1995; **221**: 734–43. [Adjuvant chemotherapy (5-fluorouracil, doxorubicin, platin) after transplantation can improve disease-free survival, even in stage III and IV disease.]

Pichlmayr R *et al.* Role of liver transplantation in the treatment of unresectable liver cancer. *World Journal of Surgery* 1995; **19**: 807–13. [Favorable indications for transplantation: stage II hepatocellular carcinoma, fibrolamellar hepatocellular carcinoma, epitheliod hemangioendothelialoma, hepatoblastoma, and neuroendocrine liver metastasis.]

Que FG, Nagorney DM, Batts KP, Linz LJ, Kvols LK. Hepatic resection of metastatic neuroendocrine carcinoma. *American Journal of Surgery* 1995; **169**: 36–42. [Four-year survival rate of 73 per cent and relief of symptoms in 90 per cent following resection in 74 patients.]

Ravikumar TS. Laparoscopic staging and intraoperative ultrasonography for liver tumor management. *Surgical Oncology Clinics of North America* 1996; **5**: 271–81.

Ringe B, Pichlmayr R, Wittekind C, Tusch G. Surgical treatment of hepatocellular carcinoma: experience with liver resection and transplantation in 198 patients. *World Journal of Surgery* 1991; **15**: 270–85. [Liver resection is the treatment of choice for primary liver cancer, while transplantation may be indicated, especially for non-resectable or recurrent lesions.]

Rosen CB, Donohue JH, Nagorney DM. Liver resection for metastatic colon and rectal carcinoma. In: Cohen AM, Winawer SJ, eds, *Carcinoma of the colon, rectum, and anus*, pp. 805–21. McGraw-Hill, New York, 1995.

Roth JA, Cristiano RJ. Gene therapy for cancer: What have we done and where are we going? *Journal of the National Cancer Institute (USA)* 1997; **89**: 21–39. [Comprehensive review of the current state of gene therapy for cancer.]

Rothenberg ML *et al.* Phase II trial of irinotecan in patients with progressive or rapidly recurrent colorectal cancer. *Journal of Clincal Oncology* 1996; **14**: 1128–35. [Initial phase II trial showing a 23 per cent response rate with a duration of 6 months for patients refractory to 5-fluorouracil.]

Scheele J, Stangl R, Altendorf-Hofmann A, Gall FP. Indications of prognosis after hepatic resection for colorectal secondaries. *Surgery* 1991; **110**: 13–29. [Review of 266 patients with identification of factors related to survival including: the presence and extent of mesenteric lymph node involvement, grade III/IV primary tumor, synchronous diagnosis of metastases, satellite metastases, limited resection margins, and non-anatomic.]

Selby R, Kadry Z, Carr B, Tzakis A, Madariaga JR, Iwatsuki S. Liver transplantation for hepatocellular carcinoma. *World Journal of Surgery* 1995; **19**: 53–8. [Postoperative 5- year actuarial survivals of 105 patients correlated with the pathologic stage of the tumor. The overall 5-year survival for all patients was 36 per cent. Nodal disease, bilobar tumor, and macroscopic venous invasion were

significant poor-prognosis features.]

Simonetti RG, Liberati A, Angiolini C, Pagliaro L. Treatment of hepatocellular carcinoma: a systematic review of randomized controlled trials. *Annals of Oncology* 1997; **8**: 117–36. [Systematic review indicated that non-surgical current treatments are ineffective (systemic chemotherapy) or minimally and uncertainly effective (embolization, tamoxifen, and interferon).]

Small WC *et al.* Preoperative determination of the resectability of hepatic tumors: efficacy of CT during arterial portography. *American Journal of Radiology* 1993; **161**: 319–22. [Study showing 91 per cent accuracy for staging of patients with colorectal liver metastasis by angiographically assisted planar CT.]

Steele G. Cryoablation in hepatic surgery. *Seminars in Liver Disease* 1994; **14**: 120–5. [Review of cryoablation for treatment of liver metastasis.]

Tsai SL, Liaw YF. Etiology and pathogenesis of hepatocellular carcinoma. *Digestive Surgery* 1995; **12**: 7–15. [Review of multiple etiologic factors for hepatocellular carcinoma.]

Vauthey JN *et al.* Factors effecting long term outcome after hepatic resection for hepatocellular carcinoma. *American Journal of Surgery* 1995; **169**: 28–35. [In a review of 106 patients, early cirrhosis (Child–Pugh A) did not adversely affect survival. Vascular invasion predicted long-term outcome.]

Wagner JS, Adson MA, van Heerden JA, Adson MH, Ilstrup DM. The natural history of hepatic metastasis from colorectal cancer: a comparison with resective treatment. *Annals of Surgery* 1984; **199**: 502–8. [Landmark early paper showing superiority of surgical resection (*n* = 141) when compared with historical controls treated non-surgically.]

Zeman RK *et al.* Helical (spiral) CT of the abdomen. *American Journal of Radiology* 1993; **160**: 719–25. [Review of the technology and clinical applications of spiral CT.]

Zhou X-D, Tang Z-Y. Management of hepatocellular carcinoma: long term outcome in 2639 cases. *Japanese Journal of Chemotherapy* 1997; **24**(Suppl. 1): 9–15. [Review of large series of surgically managed patients.]

30.5 The Budd–Chiari syndrome

J. Michael Henderson

Pathophysiology
Etiology
Chronic myeloproliferative disorders
Webs
Tumors
Pregnancy and oral contraceptives
Paroxysmal nocturnal hemoglobinuria
Hypercoagulable states
Clinical presentation
Acute fulminant Budd–Chiari syndrome
Subacute Budd–Chiari syndrome
Chronic Budd–Chiari syndrome
Diagnosis
Radiologic investigation
Liver biopsy
Hematologic evaluation
Management
Hematologic management
Outcomes
Summary
Further reading

Budd–Chiari syndrome encompasses a group of disorders caused by obstruction to hepatic venous outflow. The range of presentations of the clinical syndrome is from mild, non-specific, upper abdominal symptoms to a fulminant course with extensive hepatic necrosis. Budd originally described the pathologic features of hepatic vein thrombosis in 1846 while the clinical syndrome of hepatomegaly, ascites, and abdominal pain was described by Chiari in 1899. However, the syndrome now more broadly includes any thrombotic or non-thrombotic occlusion at the hepatic veins or the suprahepatic inferior vena cava leading to liver outflow obstruction. The venous outflow obstruction results in marked elevation of sinusoidal pressure, intense congestion of the liver, progressive necrosis of the hepatic parenchyma, and hepatocellular death.

The investigation and management of Budd–Chiari syndrome depends on a multidisciplinary approach with input from hepatology, hematology, radiology, and surgery. The rarity of this syndrome often leads to a delay in diagnosis, but suspicion of hepatic venous outflow obstruction mandates a careful work up with the priorities of defining the following.

1. Is there a mechanical outflow obstruction?
2. Is there ongoing acute or chronic liver damage?
3. What is the underlying etiology for the syndrome?
4. What are the treatment options?

This chapter will define a method for working through these steps based on current data.

Pathophysiology

Budd–Chiari syndrome is caused by obstruction to hepatic venous outflow. This obstruction means that the liver sinusoids cannot drain while blood continues to flow into the liver. This results in congestion of the sinusoids, increased sinusoidal pressure, hepatocyte necrosis, and progressive liver damage. If this passes to the chronic phase, centrilobular fibrosis results. Untreated, patients die from progressive liver damage over several months to years.

The normal physiology of hepatic venous drainage is through the right, middle, and left hepatic veins, with the caudate lobe having separate short veins draining into the inferior vena cava. In addition, there are usually several small hepatic veins passing directly from the right lobe into the intrahepatic inferior vena cava. In Budd–Chiari syndrome, if outflow obstruction only affects some of these veins, the segments of the liver that maintain venous drainage may be spared, will hypertrophy significantly, and there is therefore a more chronic rate of progression of the total syndrome. The extent of hepatic venous thrombosis correlates with the clinical manifestations of the syndrome and its prognosis. When there is total obstruction to all hepatic venous outflow, the severe congestion and swelling will stretch Glisson's capsule with severe pain and a fulminant course. On the other hand, the more common sequence is for one or two of the major hepatic veins to occlude and the symptoms may be less severe leading to delay in diagnosis until there is further obstruction of other major venous outflow.

Veno-occlusive disease of the liver presents with different pathophysiology but has similar outcomes. Most common in patients undergoing bone marrow transplantation, this syndrome results in a non-thrombotic occlusion of small sublobular branches of the hepatic veins causing a sinusoidal outflow obstruction. This in turn can lead to hepatocyte necrosis and a full-blown syndrome.

Etiology

Etiologies are listed in Table 1.

Myeloproliferative disorder	50
Tumors	10
Estrogen use/pregnancy	10
Hypercoaguable states	5
Paroxysmal nocturnal hemoglobinuria	5
Caval or hepatic vein webs	5
Idiopathic	15

Table 1 Etiologies of Budd–Chiari syndrome (%)

Chronic myeloproliferative disorders

These are the most common etiologic factor in Budd–Chiari syndrome. Polycythemia rubra vera is the most common of these although other myeloproliferative disorders such as myelofibrosis, essential thrombocythemia, and chronic lymphocytic leukemia may cause hepatic vein thrombosis. These disorders are characterized by a common origin of a malignant stem cell change followed by proliferation. The outcome of this process is an increased viscosity, a low-grade disseminated intravascular coagulation, and intravascular thrombosis. The liver is the main site of clearance of plasminogen activator, and a deficiency of hepatic antiplasmin may contribute to localization of thrombosis in the hepatic veins. Evaluation and detection of these hematologic disorders is more difficult in patients with Budd–Chiari syndrome than in their usual occurrence because of associated portal hypertension with increased plasma volume and splenomegaly which may mask the increased red blood cell mass and changes of thrombocythemia. Early detection of an underlying myeloproliferative disorder can be made using spontaneous endogenous erythroid colony formation *in vitro* and/or in bone marrow biopsies. Up to three-quarters of patients with ‘idiopathic' Budd–Chiari syndrome have been shown to fall into

this category.

Webs

Webs in the suprahepatic inferior vena cava or hepatic venous outflow are a common cause of Budd–Chiari syndrome in the Far East and South Africa. In the United States, webs should be considered in immigrants from that region but are not frequently seen in the native populations.

Tumors

Tumors as a cause of Budd–Chiari syndrome account for approximately 10 per cent of cases. Most of these tumors are intrahepatic hepatocellular carcinomas that directly invade the hepatic veins. However, other tumors such as renal and/or adrenal tumors may involve the vena cava and/or the hepatic veins themselves. It is the mechanical obstruction caused by these tumors with intrinsic or extrinsic compression of the hepatic veins that lead to Budd–Chiari syndrome.

Pregnancy and oral contraceptives

Pregnancy and oral contraceptives are cited in the etiology of Budd–Chiari syndrome and it is usually in the first 2 months after delivery that pregnancy is implicated as a risk factor. It has been postulated that increased levels of factors VII and VIII along with increased fibrinogen may be etiologic. While the reported risk in patients on the oral contraceptive pill is 2.4 over age-matched controls, other etiologic factors should be sought as some of these patients may have an underlying hematologic disorder.

Paroxysmal nocturnal hemoglobinuria

Paroxysmal nocturnal hemoglobinuria may accompany other severe hematologic disorders such as aplastic anemia and the acute leukemias. Characterized by complement-induced platelet activation, these patients are markedly hypercoagulable with extensive thrombosis of all outflow tracts and a devastating fulminant course.

Hypercoagulable states

Hypercoagulable states account for a relatively low percentage of patients with Budd–Chiari syndrome. However protein C, protein S, and antithrombin III deficiencies should all be sought in patients with hepatic venous thrombosis and no other defined etiology. More recently, the factor V Leiden mutation has been shown to be a frequent cause of hereditary thrombophilia. A recent analysis has documented a significantly increased prevalence of this mutation in Budd–Chiari syndrome, so it is a further factor which should be sought.

Clinical presentation

Budd–Chiari syndrome is relatively uncommon in Europe and the United States, being much more common in India, South Africa, and the Orient. It can occur at any age, but is most common in the third and fourth decades; it is slightly more common in women than in men. Budd–Chiari syndrome may present in acute, subacute, or chronic forms depending on the factors outlined in the pathophysiology above.

Acute fulminant Budd–Chiari syndrome

This is uncommon and only occurs when there is total obstruction of all venous outflow with severe congestion, hepatocyte necrosis, and liver failure. This usually implies a severe underlying hematologic disorder which must be fully defined in making management decisions.

Subacute Budd–Chiari syndrome

This presents with symptoms over several weeks. Right upper quadrant pain and ascites are the dominant symptoms, but the absence of other stigmas of liver disease such as jaundice, muscle wasting, and spider angiomas often result in failure to make the diagnosis. Hepatomegaly is present, but may be difficult to detect in the face of significant ascites. Splenomegaly is not a feature in early Budd–Chiari syndrome, only occurring as portal hypertension develops later in the course. The most frequent misdiagnoses are of a gynecologic or gastrointestinal malignancy or an intra-abdominal inflammatory process causing the ascites.

Chronic Budd–Chiari syndrome

This presents with portal hypertension and stigmas of chronic liver disease. Ascites, variceal bleeding, and muscle wasting may all be present, and it is usually because ‘something does not fit' in the work up that Budd–Chiari syndrome is considered and diagnosed as outlined below.

In all stages of the disease, non-specific symptoms such as general malaise, tiredness, anorexia, and nausea are unhelpful. In considering the clinical presentation of disorders in the differential diagnosis such as malignancies listed above, right-sided cardiac problems, or early diagnosis of other liver diseases, these non-specific symptoms will be similar. It cannot be overemphasized that the diagnosis of Budd–Chiari syndrome depends on a high index of suspicion.

Diagnosis

Laboratory studies, as with the clinical presentation, are overall unhelpful in making a diagnosis of Budd–Chiari syndrome. Liver tests usually show minimal elevation of aminotransferases and bilirubin, a marginal decrease in serum albumin, and a mild prolongation of prothrombin time. Considering the degree of hepatocellular damage that is often seen at biopsy, it is surprising how minimal these laboratory changes appear. The one exception is the unusual patient with fulminant Budd–Chiari syndrome who may show marked laboratory abnormalities. Ascites fluid analysis should be done but the findings are highly variable with total protein from 1.5 to 3 g. The serum to ascites albumin gradient of greater than1.1 g/dl is usually found, but is not diagnostic. The ascites cell count is variable with most patients having fewer than 100 white blood cells/mm^3. Spontaneous bacterial peritonitis does not occur in Budd–Chiari syndrome.

Diagnosis depends on a combination of appropriate radiologic imaging and liver biopsy. If the diagnosis is confirmed, then full hematologic evaluation becomes important. The sequence to be followed can therefore be divided into the phases of: (i) Is there a mechanical outflow obstruction? (ii) Is there ongoing or chronic hepatic damage? and (iii) Is there an underlying hematologic disorder?

Radiologic investigation

Radiologic investigation must follow an appropriate sequence. Clinical suspicion of Budd–Chiari syndrome should lead to hepatic ultrasound with Doppler flow studies of the main hepatic veins. In a normal subject, the three major hepatic veins can be readily visualized and phasic flow documented within them. Inability to demonstrate these veins or to only show small and stenotic vessels with no phasic flow raises a high index of suspicion of hepatic vein occlusion. Figure 1 shows ultrasound imaging of hepatic vein with appropriate phasic flow within it. If this can be demonstrated for all major hepatic veins, Budd–Chiari syndrome is excluded. Concurrent with hepatic vein imaging, the portal vein should be visualized to document patency and presence of prograde flow to the liver within it. This is an important further piece of information for management decisions.

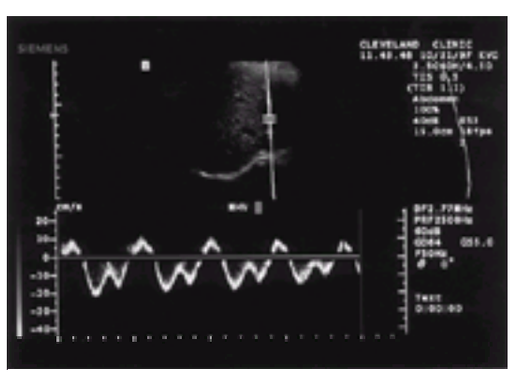


Fig. 1. Doppler ultrasound of hepatic veins. The patent hepatic vein can be seen and the tracing shows normal phasic flow pattern. If these two components are absent, the diagnosis of Budd–Chiari syndrome must be entertained.

Other radiologic imaging modalities such as hepatic scintigraphy, computed tomography, and magnetic resonance imaging have all been used both to assess liver morphology and for vascular imaging in Budd–Chiari syndrome. While each of these may provide some supportive evidence for Budd–Chiari syndrome, they are not first-line diagnostic modalities and by and large are unnecessary.

The gold standard for radiologic diagnosis of Budd–Chiari syndrome is hepatic venography and this should be performed if the Doppler ultrasound is equivocal or strongly supports the diagnosis. Hepatic venography may be done either through a transfemoral or a transjugular approach, and should attempt to cannulate each of the major hepatic veins directly. Pressure measurements are of little value in making the diagnosis as reliable pressures cannot be measured in the occluded or stenotic veins. Contrast material should be injected at each site and will give the typical spider's web appearance that is characteristic of Budd–Chiari syndrome (Fig. 2). The spider's web represents contrast in recannulized vessels coursing on the surface of the liver which are collaterals that are attempting to decompress the congested sinusoids. In involved segments, no normal named hepatic vein can be identified. As indicated above, the disease may only affect one or two of the three major hepatic veins, and it is for this reason that all the major veins must be studied. Venography also includes visual and pressure assessment of the inferior vena cava which may be compressed by the swollen liver (Fig. 3) or thrombosed. This information helps plan decompressing therapy.

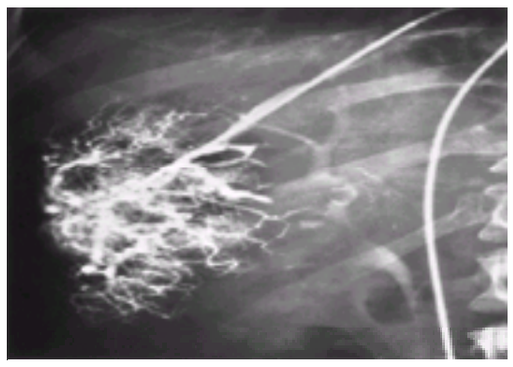


Fig. 2. The typical spider's web of Budd–Chiari syndrome seen by hepatic venography. The interconnecting small branches represent collaterals attempting to decompress the sinusoids. No major hepatic vein is visualized.



Fig. 3. Inferior vena caval compression. The swollen liver of Budd–Chiari syndrome has compressed the inferior vena cava with a resultant pressure gradient from the infrahepatic cava to the right atrium.

The role of celiac and/or superior mesenteric arteriography is more controversial in the work up of patients with Budd–Chiari syndrome. These studies should be used when Doppler ultrasound fails to demonstrate an open portal vein with flow. Arteriography, followed through to the venous phase, is important to document if there is thrombosis of the portal vein and its major tributaries as this may obviate some of the surgical options.

Liver biopsy

This must be performed if the radiologic imaging studies confirm or are highly suspicious for Budd–Chiari syndrome. This is the confirmatory test for a histologic diagnosis in addition to assessing the degree of ongoing hepatocyte necrosis and liver fibrosis. It is often said that the presence of massive ascites and a degree of coagulopathy may make biopsy difficult, but if that is the case, ascites should be drained and the coagulopathy corrected to allow this essential diagnostic step to be taken. Separate biopsies of the right and left lobe should be done to evaluate the degree of involvement of both sides of the liver. Transjugular liver biopsy is not a good option in these patients unless the biopsy needle can be advanced some distance into the occluded hepatic veins to assure adequate tissue samples. The essential features on liver biopsy to be evaluated by the pathologist are the degree of centrilobular congestion, hepatocellular necrosis, degree of lobular collapse, and finally extent of fibrosis and/or cirrhosis.

Laparoscopy has been suggested as a diagnostic modality in evaluating Budd–Chiari syndrome. Its use should be limited to cases where there remains a diagnostic dilemma although it does provide an alternative way for bilobar biopsies under direct vision. Laparoscopy is an adjunctive rather than a primary diagnostic tool.

Hematologic evaluation

Hematologic evaluation is the third major diagnostic step. This is summarized in Table 2 and requires the input of an experienced hematologist who understands the interrelation of the Budd–Chiari syndrome with the pathophysiologic changes of portal hypertension and the impact of this on the hematologic diagnostic tests. The newer modalities of spontaneous endogenous erythroid colony formation and measurement of factor V gene for the Leiden mutation should be performed where available.

• Routine hematology with blood smear
• Bone marrow to include reticulin stain and cytogenetics
• Erythroid colony formation
• Leukocyte alkaline phosphatase
• Platelet aggregometry
• Red blood cell mass
• Sucrose hemolysis
• Protein C, protein S, and antithrombin III levels
• Factor V Leiden mutation

Table 2 Hematologic evaluation of patients with Budd–Chiari syndrome

Management

The management of Budd–Chiari syndrome is divided into two phases, first the treatment of the liver and second the treatment of the underlying etiology. Treatment of the liver damage component is based on biopsy findings. Treatment of any underlying hematologic disorder is an ongoing process that includes management of the underlying hematologic disorder and may include anticoagulation.

The three options in the management of the liver injury are:

- non-surgical management
- surgical decompression
- liver transplantation.

Figure 4 illustrates the main categories of biopsy findings and the associated therapies. The decision on non-surgical treatment or surgical management is based on the biopsy.

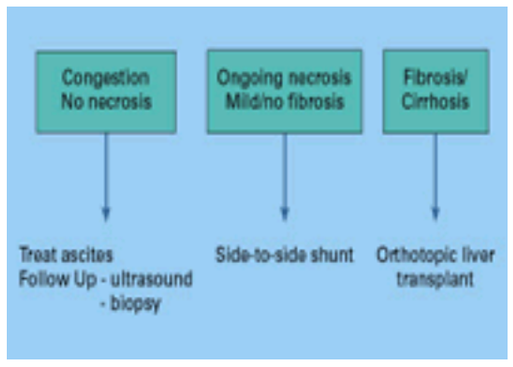


Fig. 4. Liver biopsy findings. The three broad categories of liver biopsy findings in this figure dictate the management courses.

Non-surgical management must be used with caution as the disease may result in progressive liver damage. Figure 5 shows a liver biopsy with mild Budd–Chiari syndrome with some sinusoidal dilata-tion but minimal evidence of ongoing necrosis. Management of ascites in patients such as this with salt restriction and diuretics is appropriate, but further follow-up with repeat ultrasound and repeat liver biopsies in 3 to 6 months is mandatory. In the acute setting, providing the biopsy does not show significant ongoing necrosis, thrombolytic therapy with streptokinase or urokinase can be undertaken, but again requires careful invasive follow-up. A peritoneal venous shunt does nothing to deal with the basic pathophysiology of Budd–Chiari syndrome and is a purely palliative measure that should not be used unless all other options have been eliminated. Web dilatation with angioplasty may play a role, but again requires reassessment of liver pathology. When non-surgical management is embarked upon, there must be a low index for moving forward to surgical decompression if the follow-up biopsies show progression to necrosis or there is refractory ascites.

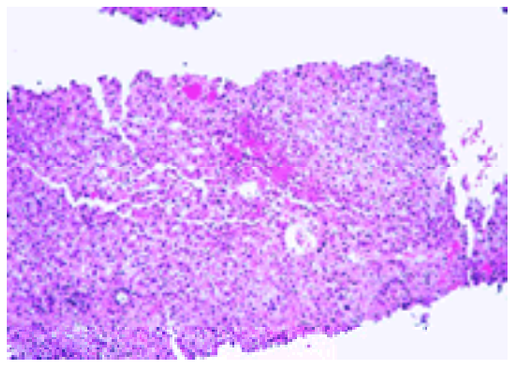


Fig. 5. Mild Budd–Chiari syndrome of this biopsy shows sinusoidal dilation with minimal evidence of necrosis.

Figure 6 shows an example of acute Budd–Chiari syndrome with ongoing hepatocyte necrosis under both high- and low-power views. This patient needs sinusoidal decompression by side-to-side portal systemic shunt. The principle in the management of this patient is that the sinusoids can be decompressed by using the portal vein as an outflow track from the obstructed sinusoids and any side-to-side type of shunt will suffice. The options are broad. A side-to-side portacaval shunt will be adequate if the inferior vena cava is not obstructed. Equally, a mesocaval shunt is an infrahepatic shunt that will decompress the sinusoids when there is no caval obstruction. However, if there is severe caval obstruction, as shown by a cava to atrial pressure gradient greater than 20 mmHg, a mesoatrial shunt may be required to decompress the sinusoids. These shunts are illustrated in Fig. 7 and Fig. 8. Other options are available to decompress the sinusoids. Transjugular intrahepatic portal systemic shunts (**TIPS**) have been used in Budd–Chiari syndrome, applying the same principle of the portal vein being the outflow track from the obstructed sinusoids. In the enlarged liver of Budd–Chiari syndrome, TIPS need to be long, often with two or three stents, and if there is an underlying thrombotic problem, careful monitoring is required to assure continued patency. Are mesoatrial shunts still indicated? The issue is important in that this is a complex procedure with a long synthetic graft. The availability of intravascular stents to stent open an obstructed inferior vena cava may obviate the need for mesoatrial shunts by combining an infrahepatic portacaval or mesocaval shunt with an intrahepatic inferior vena cava stent.

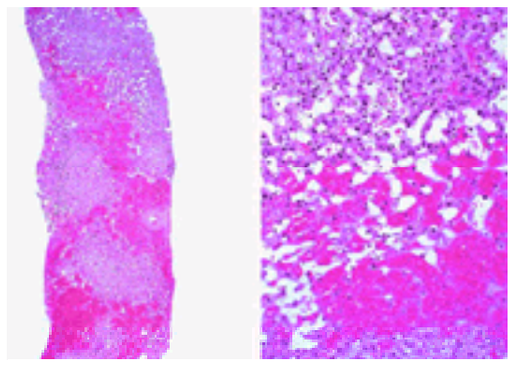


Fig. 6. Acute Budd–Chiari syndrome. The low- and high-power views show the marked sinusoidal congestion, ongoing hepatocyte necrosis, and evidence of severe liver injury.

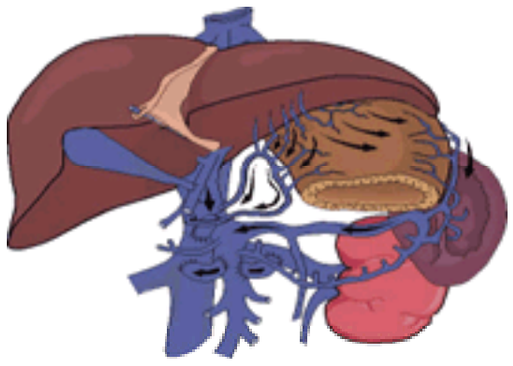


Fig. 7. Side-to-side infrahepatic shunts. Portacaval, mesocaval, and mesorenal shunts are illustrated that all serve to use the portal vein as a decompressive route from obstructed sinusoids in Budd–Chiari syndrome.

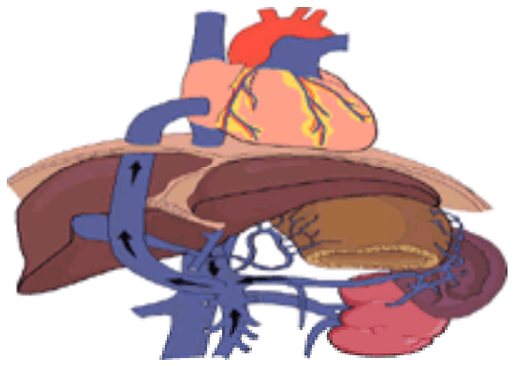


Fig. 8. Mesoatrial shunt. A long synthetic poly(tetrafluoroethene) (PTFE) graft is taken from the superior mesenteric vein in front of the liver, through the diaphragm, and anastomosed to the right atrium for decompression of the liver using the portal vein as an outflow tract from the sinusoids.

From the pathophysiologic point of view, any type of decompression of the sinusoids will halt the ongoing necrosis, stabilize the liver disease, and relieve the symptoms of Budd–Chiari syndrome in the acute setting.

Figure 9 illustrates the third type of biopsy that may be seen in Budd–Chiari syndrome with severe fibrosis and cirrhosis. There are only small islands of regenerative nodules of hepatocytes in this densely fibrotic liver. This patient requires a liver transplant. No attempt to decompress this liver will relieve the obstruction or alleviate the symptoms. Progression to this degree of scarring is usually the result of recurrent thrombotic episodes involving progressively more hepatic veins over several years. The liver has progressively attempted to compensate for loss of hepatic segments by hypertrophy of the remaining segments before they too become involved in the process. Marked distortion of gross liver anatomy is therefore usually seen on ultrasound, CAT scan, or direct visualization of the liver. Transplant is often difficult in these patients with the recipient hepatectomy being complicated by the many collaterals which have formed from the surface liver to the diaphragm in an attempt to decompress the sinusoids spontaneously. The transplant surgeon needs to be aware of this and take appropriate intraoperative precautions to move through his phase.

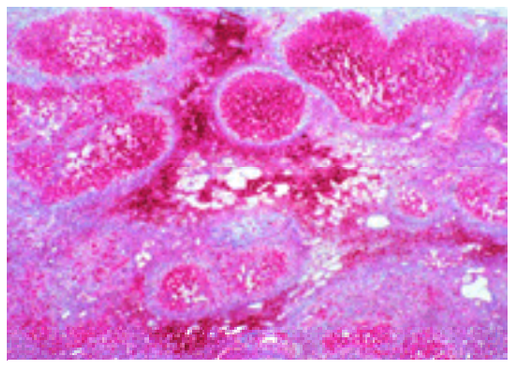


Fig. 9. Chronic Budd–Chiari syndrome. This liver biopsy shows the liver is almost entirely fibrotic with small islands of hepatocytes as regenerative nodules. This Budd–Chiari syndrome has reached endstage liver disease.

Hematologic management

This must proceed in parallel with management of the liver disease. It is important that any underlying hematologic disease is identified preoperatively, therapy initiated, and decisions made as to optimal timing for surgery considering the risk to the liver and the risk of further vascular thrombosis. Management must be continued long term, and, in addition to specific therapy aimed at any underlying blood disorder, may require long-term anticoagulation.

Outcomes

Budd–Chiari syndrome is rare, and there are no prospective, randomized, controlled trials looking at outcomes. Large series amount to approximately 50 patients. Reports have largely been based on local available expertise. However, in the last two decades there has been increasing recognition of the need to use different therapies and to tailor the management to the findings as outlined in this chapter. The reported experience with non-surgical methods is largely anecdotal with case reports of thrombolytic therapy, or web dilatations, and of conservative management of symptoms.

Experience with TIPS for Budd–Chiari syndrome is growing as exemplified by the Freiburg experience. In this series of 12 patients managed by TIPS, the stent could be placed successfully in all patients and alleviated symptoms in the 10 patients with subacute and chronic disease. The two patients with fulminant disease progressed to liver failure and died. Long-term follow-up with TIPS is still awaited in the series or other reports.

Table 3 summarizes reported experience of some of the larger series of surgical decompression by side-to-side shunts. These series include both infra- and suprahepatic shunts depending on the status of the inferior vena cava. However, some have also included infrahepatic shunts combined with caval stent. Overall, reasonable outcomes have been achieved with survival as indicated in this table, but it has been increasingly recognized that decompression with a surgical shunt may not be optimal therapy when there is significant fibrosis and cirrhosis.

Center	Publication year	No. of patients	Hepatocytic	Sporadic	Survival % (No. of years)
Hopkins	1990	26	11	15	85 (5)
Emory	1990	24	10	11	86 (5)
San Diego	1989	13	13	-	85 (4)
Mayo	1985	10	9	1	90 (5)
Baylor	1992	10	3	7	70 (5)
Paul Brousse	1990	21	20	1	96 (5)
Milan	1990	16	16	-	83 (5)
Hannover	1995	12	8	4	50 (5)
Toronto	1996	21	21	-	57 (5)

Table 3 Experience and outcome of side-to-side portal systemic shunt for acute Budd–Chiari syndrome

Experience with liver transplantation for Budd–Chiari syndrome is summarized from five centers in Table 4. While early experience had fairly significant hospital mortality, usually related to the preoperative treatments used and technically difficult procedures as outlined above, the overall results have improved. The major issue in facing a decision for transplantation is the severity of any underlying hematologic disorder and the potential impact of post-transplant immunosuppression on that disease. Equally, the need for anticoagulation has to be carefully considered in these patients as the correct timing in initiating this is essential to avoid early graft thrombosis. Some of the transplant series have reported significant retransplant rates related to vascular thrombosis following initial transplantation.

Center	Publication year	No. of patients	Hospital mortality (%)	Survival (% (no. of years)
Pittsburgh	1985	25	30	50 (5)
Cambridge	1988	17	12	71 (2)
Baylor	1992	14	14	76 (3)
Hannover	1995	43	28	69 (5)
Toronto	1996	7		67 (5)

Table 4 Experience and outcome of orthotopic liver transplantation for Budd–Chiari syndrome

Summary

The algorithm in Fig. 10 summarizes the overall approach to the diagnosis and management of Budd–Chiari syndrome. Clinical suspicion of Budd–Chiari syndrome should lead to a Doppler ultrasound of the main hepatic veins. Failure to identify these vessels or lack of phasic flow should lead to further evaluation with venography. Liver biopsy should be performed at the same time to evaluate both lobes of the liver. Confirmation of a diagnosis of Budd–Chiari syndrome should lead to definitive management based on the severity of the findings on the biopsy. Treatment options range from minimal intervention for patients with sinusoidal congestion but no ongoing hepatocyte necrosis, through surgical decompression for those with acute Budd–Chiari syndrome and ongoing liver damage, to liver transplantation for those with endstage liver disease.

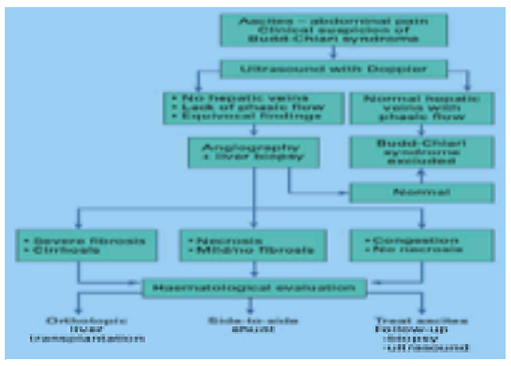


Fig. 10. Algorithm for the management of Budd–Chiari syndrome.

Further reading

Diagnostic evaluation

De Stefano V, Teofili L, Leone G, Michiels JJ. Spontaneous erythroid colony formation as the clue to an underlying myeloproliferative disorder in patients with Budd-Chiari syndrome or portal vein thrombosis. *Seminars in Thrombosis and Hemostasis* 1997; **23**:411–18. [An update on myeloproliferative disorders and Budd-Chiari syndrome.]

Henderson JM, *et al.* Surgical options, hematologic evaluation, and pathologic changes in Budd-Chiari syndrome. *American Journal of Surgery* 1990: **159**: 41–8; discussion 48–50. [The author's series of patients with Budd-Chiari syndrome.]

Mitchell MC, Boitnott JK, Kaufman S, Cameron JL, Maddrey WC. Budd-Chiari syndrome: etiology, diagnosis and management. *Medicine* 1982: **61**: 199–218. [Etiologic and diagnostic workup defined.]

Shill M, Henderson J, Tavill A. The Budd-Chiari syndrome revisited. *Gastroenterolgist* 1994; **2**: 27. [Outlines diagnostic workup for the Budd-Chiari syndrome.]

Valla D *et al.* Primary myeloproliferative disorder and hepatic vein thrombosis. A prospective study of erythroid colony formation in vitro in 20 patients with Budd-Chiari syndrome. *Annals of Internal Medicine* 1985; **103**: 329–34. [The ‘classic' paper on myeloproliferative disorders and Budd-Chiari syndrome.]

Therapy

Campbell DA Jr, *et al.* Hepatic transplantation with perioperative and lone term anticoagulation as treatment for Budd-Chiari syndrome. *Surgery, Gynecology and Obstetrics* 1988; **166**: 511–18.

Hemming AW, Langer B, Greig P, Taylor BR, Adams R, Heathcote EJ. Treatment of Budd-Chiari syndrome with portosystemic shunt or liver transplantation. *American Journal of Surgery* 1996; **171**: 176–80.

KleinAS, Cameron JL. Diagnosis and management of the Budd-Chiari syndrome. *American Journal of Surgery* 1990; **160**: 128–33.

Orloff MJ, Girard B. Long term results of treatment of Budd-Chiari syndrome by side to side portacaval shunt. *Surgery, Gynecology and Obstetrics* 1989; **168**:33–41.

Ringe B, *et al.* Which is the best surgery for Budd-Chiari syndrome: venous decompression or liver transplantation? A single-center experience with 50 patients. *Hepatology* 1995: **21**: 1337–44.

[These papers outline selective experience in centers using decompression and transplant to treat Budd-Chiari syndrome.]

31.1 Benign diseases of the biliary tract

Julian Britton, Kenneth I. Bickerstaff and Adrian Savage

Introduction
Anatomy
Physiology
Bile
Biliary pain
Pathology
Inflammation
Obstruction
Neoplasm
Investigation of the biliary tract
Plain radiography
Ultrasonography
Contrast radiology
Liver function tests
Biliary manometry
Disease of the biliary tract
Congenital abnormalities
Acquired disease
Management of gallbladder stones
Gallstone dissolution therapy
Cholecystectomy
The postcholecystectomy syndrome
Stones in the bile duct
Origin
Clinical presentation
Investigation
Management
Biliary peritonitis
Surgical damage to the bile ducts
Aetiology and prevention
Diagnosis
Treatment
Complications and results
Postinflammatory stricture
Ampullary stenosis
Sclerosing cholangitis
Aetiology
Diagnosis
Treatment
Biliary fistula
External biliary fistula
Internal biliary fistula
Recurrent pyogenic cholangitis
Pathology
Diagnosis
Treatment
Biliary hydatid disease
Benign tumours of the biliary tract
Acknowledgement
Further reading

Introduction

Benign diseases of the biliary tract are one of the most common surgical problems in the world. Gallstones in the West and oriental cholangitis in the East afflict millions of people. Surgery plays an important part in treatment and over half a million cholecystectomies are performed each year in the United States of America. Percutaneous, laparoscopic, and endoscopic procedures are replacing conventional open abdominal surgery but they all require a thorough understanding of the anatomy, physiology, and pathology of the biliary tract.

Anatomy

This is covered in Chapter 30.1. Figure 1 is reproduced here for convenience.

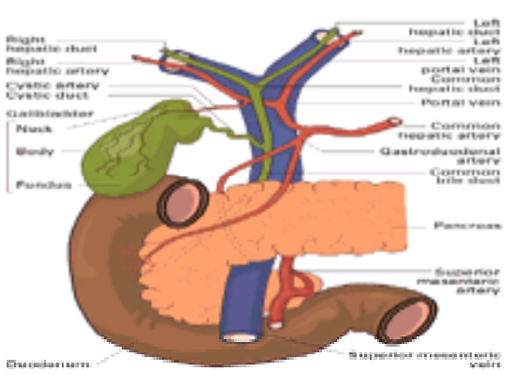


Fig. 1. Anatomy of the extrahepatic bile ducts and gallbladder.

Physiology

Bile

The prime function of the biliary tract is to convey bile from the liver, where it is formed, to the duodenum. Along the way bile is stored and concentrated in the gallbladder until it is required. Bile helps in the digestion of certain foodstuffs and it forms a major excretory pathway. Between 250 and 1000 ml of bile are secreted by human hepatocytes every day. The main constituent is water, along with bile acids, bile pigments, cholesterol, phospholipids, and all the inorganic ions found in plasma (Table 1). The bile acids are the most important by weight. The concentrations of all these molecules are subsequently modified by the epithelium of both the bile ducts and the gallbladder. The pH of bile-duct bile is generally greater than 7 and the inorganic ions are normally present in concentrations slightly higher than in plasma, with the exception of chloride, which is usually lower (Table 2). This is important in clinical practice because patients with a significant bile fistula lose electrolytes rapidly.

	g/l	%
Water		98.0
Bile salts	6.5–14	0.7
Inorganic salts		0.7
Bile pigments	0.12–1.35	0.2
Fatty acids	1.6–4.1	0.15
Lecithin		0.1
Cholesterol	0.8–1.8	0.06

Between 250–1000ml of hepatic bile is secreted in 24h

Table 1 Composition of adult human hepatic bile

	Range	Mean
pH	6.2–8.5	7.5
Sodium	131–164	149
Potassium	2.6–12.0	5.0
Chloride	89–118	101
Bicarbonate	17–55	30
Calcium	3.3–4.1	
Magnesium	1.4–3.0	

Table 2 Electrolyte content (mEq/l) of adult human hepatic bile

The bile acids, cholic and chenodeoxycholic acid, are synthesized in the liver from cholesterol and are the major pathway of cholesterol excretion. In the colon they are metabolized by bacteria to deoxycholic and lithocholic acid, respectively. All of them except lithocholic acid are absorbed back into the portal circulation and are then re-excreted into the bile (Fig. 2). This enterohepatic circulation of bile salts takes place several times each day. Because of this continuous circulation, there is always a pool of bile acids in the body. A small amount is lost from the pool each day in the stools and is made up by hepatic synthesis.

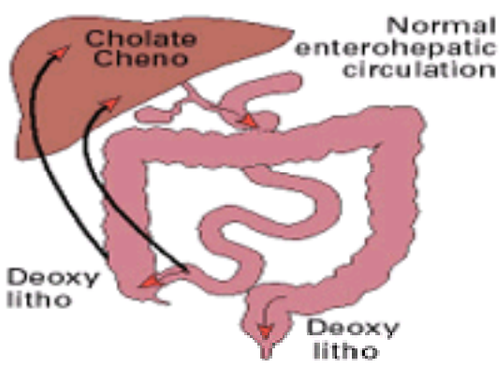


Fig. 2. Enterohepatic circulation of bile salts: because of reabsorption from the small and large bowel only a small proportion of the bile-salt pool is lost in the stool each day; this amount is replaced by synthesis of bile salts in the liver.

The main function of the bile acids in bile is to maintain cholesterol in solution, which they do by forming micelles. Hydrophobic cholesterol is encased in hydrophilic phospholipid and bile salts. Only when the concentration of cholesterol exceeds the capacity of the bile salts and phospholipid to maintain it in micellar solution does bile become lithogenic and liable to form stones. Clinically, increasing artificially the concentration of bile salts in bile makes it possible to render bile unsaturated in cholesterol and so enable dissolution of cholesterol gallstones (see Fig. 21). Ingestion of food increases bile flow, probably mediated by vagal and hormonal stimuli, although bile salts themselves have a powerful choleretic effect.

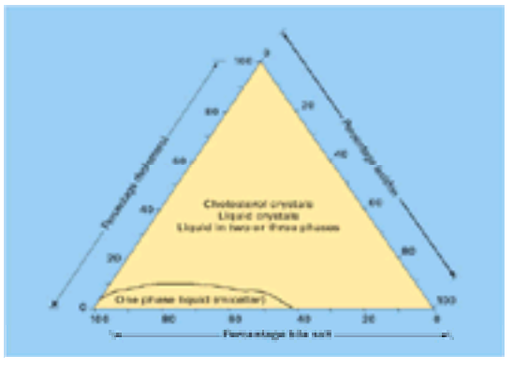


Fig. 21. Phase diagram of the solubility of cholesterol in bile, showing the physical state of all the possible combinations of bile salts in aqueous solutions. The area below the line in the bottom left represents the maximum amount of cholesterol solubilized by any mixture of lecithin and bile salt. Combinations above and to the right of this area means that cholesterol is present in bile in crystal or liquid crystal form.

Bilirubin, the major degradation product of haemoglobin, is conjugated with glucuronide by the hepatocyte and actively secreted into bile. Conjugated bilirubin is converted by the action of bacteria in the bowel into urobilinogen, which is reabsorbed into the portal circulation and re-excreted. Some urobilinogen reaches the systemic circulation and is then excreted in the urine, where it can be detected. The absence of urobilinogen from the urine implies complete obstruction of bile flow into the bowel.

The gallbladder acts as a store for bile between meals and concentrates the bile by active reabsorption of water and electrolytes. About half the hepatic output of bile is stored in the gallbladder; the other half trickles into the duodenum continuously. Food, and especially fat, in the duodenum releases cholecystokinin, which is the most important stimulus for gallbladder contraction. Contraction of the gallbladder causes the sphincter of Oddi and the second part of the duodenum to relax. The bile duct acts purely as a conduit, for it does not show peristaltic activity. The sphincter of Oddi certainly plays an important part in bile-duct function. The muscle of the sphincter is separate from the muscle of the duodenal wall both embryologically and functionally, although the two are usually co-ordinated. This co-ordination occasionally breaks down, but whether this is of clinical significance is not known. The role of the autonomic nerves to the gallbladder and the bile duct is also uncertain.

Biliary pain

At endoscopic retrograde cholangiopancreatography, distension of the bile duct with contrast causes pain and it is probable that a similar effect occurs from distension in normal life. While a rapid increase in intraductal pressure causes pain, slower increases lead to discomfort. Pain from stones in either the bile duct or the gallbladder probably results from acute obstruction to flow. Spasm of the sphincter of Oddi is probably painless, although it may also lead to a rapid rise in intraductal pressure. Pain nerve endings are widely distributed throughout the biliary tract. Most pain fibres return to the central nervous system via the splanchnic nerves but a significant minority run with the vagus, right phrenic, and intercostal nerves. This wide distribution probably explains some of the clinical variations in the perception of biliary pain.

Pathology

Inflammation

Acute and chronic inflammation of the biliary tract are both common. Inflammation can be caused by chemicals, bacteria or parasites; the cause of primary sclerosing cholangitis is unknown. Following obstruction of the gallbladder by a stone in the cystic duct, bile salts leak into the mucosa and cause inflammation. Bacterial infection soon follows. Bacteria are often present in normal bile, particularly if there are stones present or if there is an anastomosis between the biliary tree and the bowel. The most common infecting bacteria are *Escherichia coli, Klebsiella*, and *Streptococcus faecalis*. Anaerobic bacteria are present, but in much smaller numbers. Most significant infections are with multiple organisms. The frequent occurrence of bacteria in bile, even in the absence of acute inflammation, means that prophylaxis with an appropriate antibiotic before any surgical or radiological procedure on the biliary tract is wise.

The Chinese liver fluke *Clonorchis sinensis, Ascaris lumbricoides*, and occasionally a daughter cyst from a ruptured hydatid cyst of the liver are the common parasitic infestations of the biliary tract, usually in the Middle East and Asia. Bacterial cholangitis is a common accompaniment and malignant change in the biliary epithelium may follow chronic infection with *C. sinensis*.

Obstruction

Obstruction of the main bile duct leads to morphological changes in the liver with a marked inflammatory infiltrate in the portal tracts. Scarring and fibrosis around the bile ducts result in the eventual development, after weeks or months of obstruction, of secondary biliary cirrhosis. Hepatocyte function deteriorates progressively from the onset of obstruction and is slow to recover once the obstruction is relieved. Similar pathological changes can be seen in the gallbladder wall when chronic inflammation occurs secondary to gallstones. The gallbladder becomes shrunken, fibrotic, and non-functional (see Fig. 22).



Fig. 22. Gallbladder removed at operation showing severe chronic cholecystitis.

Neoplasm

Neoplasia of the glandular biliary epithelium is unusual and benign tumours are almost unknown. Occasionally, adenomatous polyps are seen in the gallbladder or the bile duct.

Investigation of the biliary tract

Plain radiography

A plain abdominal radiograph may show radio-opaque gallstones (Fig. 3) or air in the biliary tree (Fig. 4). This is a useful investigation in patients who present acutely and is an essential preliminary to any contrast study of the biliary tract as 10 per cent of gallstones are radio-opaque. Calcification in the right upper quadrant is not proof of biliary-tract disease and nor does the presence of gallstones mean that they are the cause of the patient's symptoms.



Fig. 3. Radio-opaque gallstones on a plain radiograph; the outline of the stones suggests that they are in the gallbladder not the kidney.

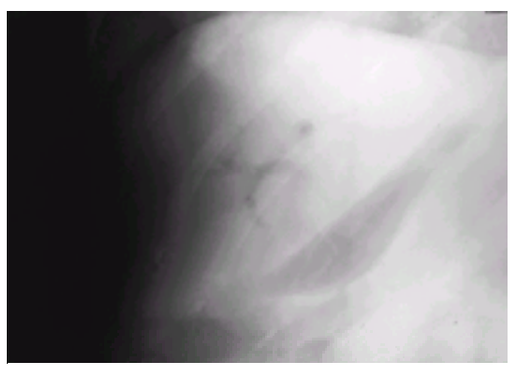


Fig. 4. Air in the biliary tree on a plain radiograph.

Two unusual appearances are diagnostic of biliary disease. Gas in the lumen or in the wall of the gallbladder is diagnostic of emphysematous cholecystitis, while the appearance of air in the biliary tree and distal obstruction of the small bowel suggests gallstone ileus. Other causes of biliary air are a biliary-enteric fistula either created surgically (Fig. 4) or as a result of a stone ulcerating from the gallbladder into the bowel, usually the duodenum and very occasionally the colon.

Ultrasonography

An ultrasound scan is the initial investigation of any patient suspected of biliary-tract disease. It is non-invasive, painless, and can be easily performed on patients who are unwell. Adjacent organs can be examined simultaneously. Ultrasonography may provide unsatisfactory results in patients with obesity, ascites and gaseous distension, and is dependent on the skill and experience of the operator.

Stones within the gallbladder can be detected with a sensitivity and specificity of over 90 per cent. They appear as reflective foci that cast an acoustic shadow (Fig. 5) and that move with changes in posture; they may form a layer in the gallbladder. The presence of local tenderness and the thickness of the gallbladder wall allow acute and chronic cholecystitis to be distinguished. A layer of oedema may also appear as a sonolucent zone between the gallbladder and the liver in patients with acute cholecystitis. Polyps within the lumen can be identified.

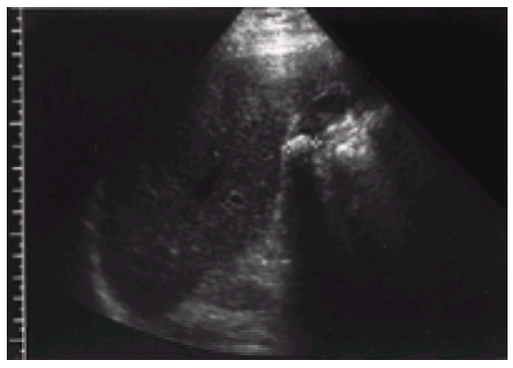


Fig. 5. Ultrasonogram showing gallstones in a contracted, thick-walled gallbladder.

Gallbladder function can be assessed by ultrasound. Volume is calculated from the ultrasonic dimensions of the gallbladder (Volume = length × width × height × p/6). A normal gallbladder will reduce in volume by two-thirds after a suitable fatty drink.

Ultrasonography is always the first imaging technique used in a patient with jaundice or cholangitis. Dilatation of the ducts establishes an extrahepatic obstructive cause for the jaundice and may also identify stones in the ducts (Fig. 6). An expert can identify the site of the obstruction in three-quarters of patients and the cause in one-third. Ultrasound can also be used during surgery to detect bile-duct stones.

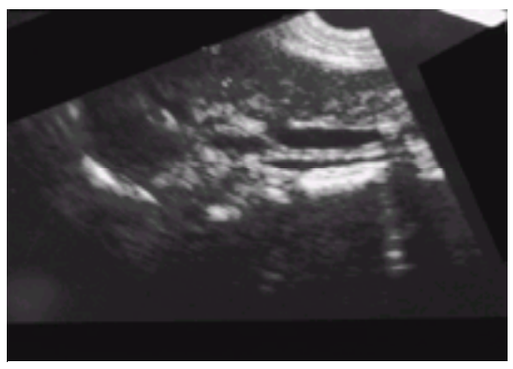


Fig. 6. Ultrasonogram showing a gallstone in a dilated common bile duct; there is a typical sonic shadow behind the stone.

Contrast radiology

Oral cholecystography

Although this was the standard method of investigating the gallbladder for many years, it has now been superseded by ultrasonography. A biliary contrast medium based on tri-iodobenzoic acid, which is primarily excreted by the liver, is taken by mouth. It is absorbed into the portal venous system, transported across liver cells into bile and concentrated in the gallbladder, where it becomes visible on a radiograph (Fig. 7). Patients are then given a fatty meal; a normal gallbladder contracts. Failure of the gallbladder to opacify is evidence of gallbladder disease, provided that the contrast material has been absorbed and the patient is not jaundiced.

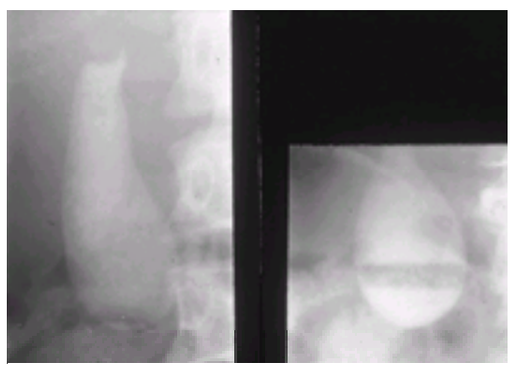


Fig. 7. Oral cholecystogram showing a layer of gallstones within the gallbladder; this is only seen when the patient stands erect (right-hand picture).

Percutaneous transhepatic cholangiography

A 22-gauge, flexible Chiba needle is advanced through the skin and well into the liver under local anaesthesia and antibiotic cover. Blood coagulation must be normal. Contrast medium is gently injected as the needle is slowly withdrawn. The injection is watched on an image intensifier: when contrast enters a bile duct, withdrawal is stopped. Contrast is then injected to fill the whole biliary tree (Fig. 8). In skilled hands this can always be achieved if the bile ducts are dilated and is also possible in about two-thirds of patients in whom the ducts are normal in size. It may be necessary to puncture both the right and the left hepatic systems to outline all the ducts. Bile leakage into the peritoneum, cholangitis, and haemorrhage are the major complications, occurring in about 4 per cent of patients. For this reason it is best to plan to relieve any obstruction, either surgically or by the percutaneous insertion of a drain or stent, during the initial examination. There is a mortality of about 0.5 per cent.

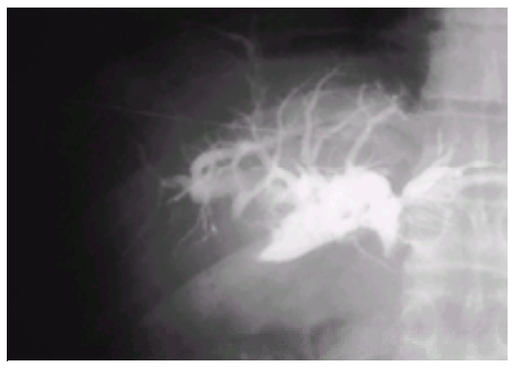


Fig. 8. Percutaneous transhepatic cholangiogram showing grossly dilated ducts. The stricture at the choledochojejunostomy developed 16 years after the original surgery, which was performed to repair a damaged bile duct during a cholecystectomy. The Chiba needle can be seen at the top left of the picture.

Endoscopic retrograde cholangiopancreatography (ERCP)

A full description of the technique of ERCP is given in Chapter 14.2. The ability to outline the biliary tree and the pancreatic duct as well as to inspect the ampulla of Vater has completely revolutionized the management of many benign biliary problems. Not only has diagnosis been improved (Fig. 9) but therapeutic techniques have been developed that have significantly improved patient care. ERCP requires technical skill and sophisticated imaging, and is not as widely available as the percutaneous approach. The endoscopic route is preferred initially because it does not transgress the peritoneum, but there are occasions when it is essential to outline the bile duct from above as well as from below.

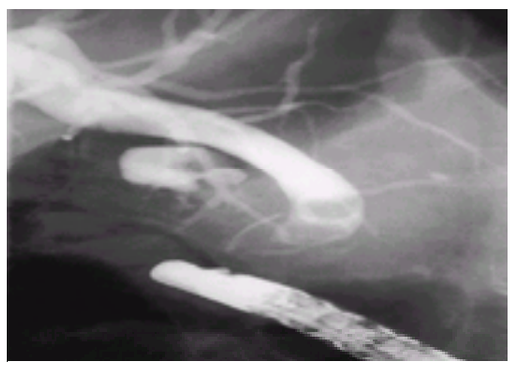


Fig. 9. Endoscopic retrograde cholangiogram showing stones in the common bile duct. This patient has earlier had a Polya partial gastrectomy so that the endoscope approaches the ampulla backwards from the duodenojejunal flexure along the third part of the duodenum.

Computed tomography (CT)

CT is useful in the diagnosis of biliary-tract disease, although the evidence is usually indirect. It is easier to identify disease in the liver, the pancreas, and the gallbladder rather than the bile duct itself, although it is simple to see the ducts when they are dilated (Fig. 10). CT is probably superior to ultrasonography in identifying the level and the cause of biliary obstruction but ultrasound is cheaper, simpler, and safer.

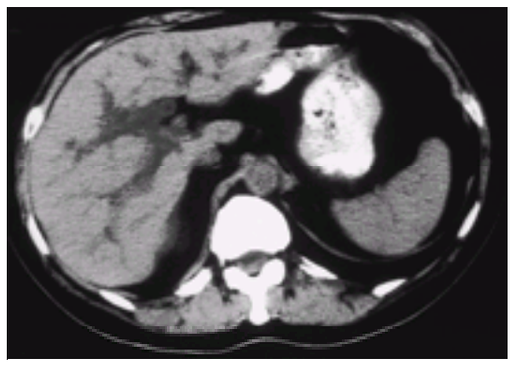


Fig. 10. CT scan of the liver showing grossly dilated bile ducts.

Malignant disease in or around the biliary tree can best be staged by CT; this may also be needed to allow a biopsy to be taken. It is not of great value in the management of the common inflammatory conditions, where ultrasound is of greater use. However, ultrasonography and CT should be regarded as complementary investigations and in difficult cases it would be quite usual to obtain both.

Other imaging techniques

There is a variety of other imaging techniques that are sometimes of value. Intravenous cholangiography, where the bile ducts are outlined by intravenously administered contrast excreted in the bile, has largely been superseded by ERCP, percutaneous transhepatic cholangiography, and magnetic resonance cholangiography.

Isotopic scanning of the gallbladder may be useful in the diagnosis of acute cholecystitis. A technetium-labelled derivative of iminodiacetic acid (**HIDA**, PIPIDA) is given intravenously and images are then recorded by gamma-camera (Fig. 11). In acute cholecystitis the cystic duct is occluded and so the gallbladder will not opacify. If it does opacify, acute cholecystitis can be confidently excluded. A HIDA scan is particularly useful in the diagnosis of acute acalculous cholecystitis. False-positive results may occur in alcoholic liver disease and also in patients maintained on total parenteral nutrition, in whom the gallbladder is atonic and so the isotope may fail to enter it despite a patent cystic duct. HIDA scanning is also used to assess the rate of excretion through a biliary–enteric anastomosis.

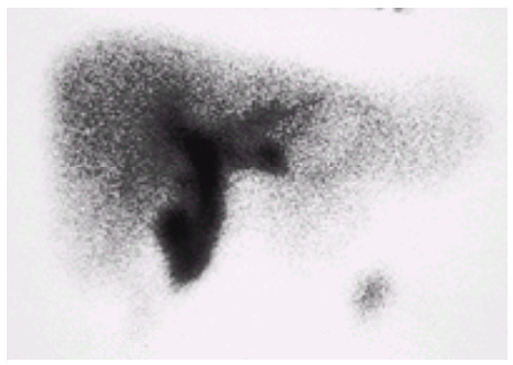


Fig. 11. HIDA scan of the liver and the bile ducts showing a normal excretion pattern; the gallbladder can just be seen as a small nubbin off the main bile duct.

Arteriography is essential before any major operation on the biliary system because of the wide variation in biliary arterial anatomy (Fig. 12).

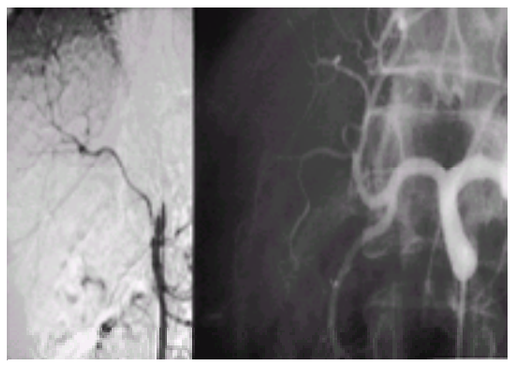


Fig. 12. Hepatic angiogram showing one of the common anomalies. The left hepatic artery arises from the coeliac plexus (the radiograph on the right) but the right hepatic artery comes from the superior mesenteric artery (the radiograph on the left). This arrangement is found in about 25 per cent of patients.

Magnetic resonance imaging (**MRI**) of the liver, gallbladder, and pancreas provides similar images to those obtained from CT. It is particularly useful in patients, such as children, where repeated exposure to X-rays is undesirable. Accurate anatomical images of the bile duct and the pancreatic duct can also be obtained with MRI using new techniques and new contrast materials. It seems likely that magnetic resonance cholangiopancreatography will supersede ERCP and percutaneous transhepatic cholangiography for the diagnosis of biliary-tract and pancreatic disease (Fig. 13).



Fig. 13. Image of the bile duct from a magnetic resonance scan (by courtesy of Dr D. Lomas).

Liver function tests

Biochemical measures of liver function abound. In a surgical context, interest has focused on bilirubin, alkaline phosphatase, which is excreted by liver cells, and transferase enzymes, which are predominantly located within liver cells. Changes in these three have traditionally been used to differentiate between intra- and extrahepatic causes of jaundice. A rise in alkaline phosphatase signifies an extrahepatic obstructive cause whilst changes in enzyme levels indicate disease within the liver cells themselves. These changes are never totally reliable and as indicators they have been superseded by ultrasonography and the demonstration of dilatation of the bile ducts. They are, however, useful indicators of disease severity, and their main value is to monitor the effects of treatment. Prothrombin time also measures s liver function, since it depends on the synthetic functions of the liver. Prolongation of the prothrombin time, which might lead to excessive haemorrhage, must be corrected by the administration of vitamin K or fresh frozen plasma before embarking on any surgical procedure.

Biliary manometry

Biliary manometry is used to assess the function of the sphincter of Oddi. Pressure traces can be obtained either by placing a special perfusion catheter across the sphincter from below at ERCP or from above during the course of exploring the common bile duct at operation. Placement from below at ERCP is used to detect stenosis and dyskinesia of the sphincter of Oddi in patients with persistent pain following a cholecystectomy. Sphincter stenosis is diagnosed by an elevated basal sphincter pressure and these patients are often cured of their symptoms by an endoscopic sphincterotomy. Other manometric abnormalities have been identified, such as rapid phasic contractions, excessive retrograde contractions and a paradoxical response to cholecystokinin, but their clinical relevance is not yet clear. Manometry in conjunction with peroperative cholangiography can detect small stones in the bile duct at operation but the technique is time-consuming and difficult to perform accurately. It is not much used.

Disease of the biliary tract

Congenital abnormalities

Biliary atresia (see also Chapter 42.6)

Biliary atresia has an incidence of 1 in 12 000 live births and presents in the first week of life as cholestatic jaundice. Untreated it pursues a relentless course, with progressive liver failure and death before the age of 3 years. The cause of the atresia is unknown but failure of vacuolation of the solid biliary bud in the early weeks of intrauterine life is one possible explanation. Inflammatory obliteration of previously patent ducts occurring about the time of birth is currently considered to be the cause. Viruses and metabolic disorders are two possible causes for the inflammation.

Classification

Three pathological types of atresia are recognized: type 1, atresia of the common bile duct; type 2, atresia of the common hepatic duct; and type 3, atresia of the right and left hepatic ducts. Types 1 and 2 are relatively easy to correct surgically. In type-3 disease there is often a conical area of fibrous tissue at the hilum of the liver that

contains within it many small biliary radicals. The number of these ducts decreases with time and this adds to the urgency of diagnosis.

Diagnosis

Biliary atresia is notoriously difficult to diagnose. Jaundice in the new born can be caused by hepatocellular disease, intrahepatic bile-duct hypoplasia or a choledochal cyst as well as biliary atresia. Conventional methods of investigation that are useful in adults are less so in infants. For example, there is rarely any intrahepatic duct dilatation so that ultrasonography is of limited help. Histological examination of a liver biopsy is the most reliable method of diagnosis but in clinical practice a range of tests is usually needed.

Treatment

The discovery on microscopy of small biliary radicals in the fibrous tissue at the hilum of the liver has led to the development of portoenterostomy for the treatment of type-3 biliary atresia. In this procedure all of the fibrous tissue and atretic ducts at the porta hepatis are resected *en bloc*. An open enterotomy in the antimesenteric border of a jejunal loop is then sutured to the edge of the fibrous tissue in the porta hepatis. Bile drains from the liver into the bowel through the tiny ducts at the liver hilum. This Kasai operation can lead to long-term survival in nearly half the patients (Fig. 14). If atresia involves only the extrahepatic bile ducts, it is possible to anastomose the residual dilated hepatic ducts to a Roux-en-Y loop of small bowel. Unfortunately, the results of hepaticojejunostomy are not much better than for portoenterostomy. Various modifications of the Kasai procedure have been suggested, including bringing up the Roux loop to the skin surface in order to allow easy access to the hepatic anastomosis.

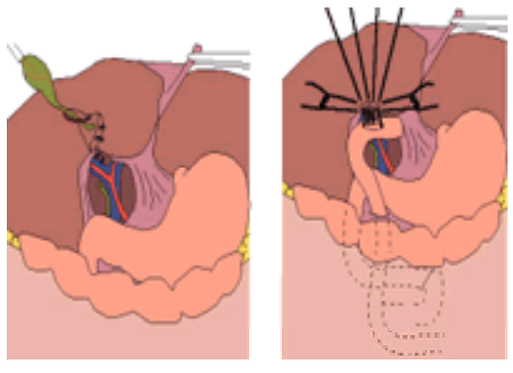


Fig. 14. The Kasai procedure: the abnormal tissue at the hilum of the liver is excised and a Roux-en-y loop of jejunum is sutured to the raw surface at the hilum of the liver.

Bacterial cholangitis and portal hypertension are the two common complications of the Kasai operation. Cholangitis requires treatment with systemic antibiotics and the exclusion of a surgically correctable obstruction such as kinking of the Roux loop. Portal hypertension consequent upon hepatic fibrosis is a long-term problem: injection sclerotherapy is the treatment of choice for children who develop oesophageal varices.

Even though the Kasai operation has proved a significant advance in the management of biliary atresia the results are not completely satisfactory. However, the results of liver transplantation in children have improved so much that it is now the treatment of choice for children with endstage liver failure (see Chapter 16.6). As there are few child liver donors it is often necessary to reduce the size of an adult liver for transplantation into an infant, but the surgery and the immunosuppression are tolerated without difficulty. Portoenterostomy remains the first treatment for biliary atresia; liver transplantation is available when portoenterostomy is impossible or liver function fails.

Choledochal cyst

Choledochal cyst is an aneurysmal dilatation of the bile duct. It is a rare condition, with an incidence between 1:100 000 and 1:150 000 live births in the West, although it is probably more common in the East. Females are affected two to four times more often than males. Most patients present in infancy, although a significant minority are diagnosed as adults.

Aetiology

The cause is unknown. Partial biliary obstruction and a weakness of the wall of the bile duct are the two basic defects required for the formation of a cyst: both these abnormalities may be congenital or acquired. Abnormal recanalization of the bile duct, which is complete by the fifth week of intrauterine development, could result in areas of stenosis and dilatation. This conjecture is supported by the occurrence of choledochal cysts in neonates and by the finding of multiple areas of stenosis and dilatation in some patients with a choledochal cyst. A cyst may also develop after trauma or from fibrosis and stenosis of the distal common bile duct from recurrent cholangitis associated with stones in the bile duct. Early changes of this type are sometimes seen at ERCP.

Classification

Cystic dilatation may affect any part of the biliary system; five patterns have been described (Fig. 15). Type 1, a cystic or fusiform dilatation of the common bile duct, is the most common (82 per cent). Type 2 (3 per cent) is a supraduodenal diverticulum of the common bile duct, and type 3 (5 per cent) is a diverticulum of the intraduodenal bile duct, or choledochocele. Type 4 (9 per cent) consists of multiple cysts: type-4A cysts affect both the intrahepatic and extrahepatic bile ducts, while type-4B cysts affect the extrahepatic duct only. Type 5 (1 per cent) describes cysts of the intrahepatic bile ducts. These may be solitary or multiple, and this type includes Caroli disease. They can vary in size from 2 cm in diameter to giant cysts and the wall is composed of fibrous tissue that may be up to 1 cm thick. The cyst is lined by cuboidal biliary epithelium that is often ulcerated in adults.

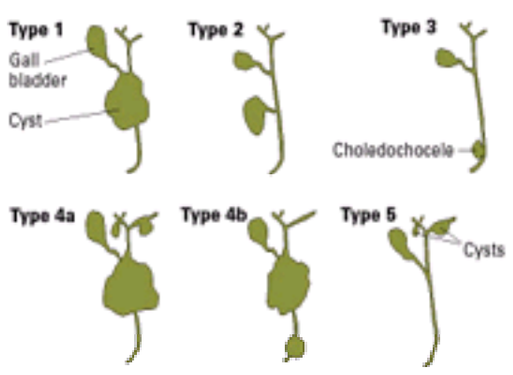


Fig. 15. Classification of choledochal cysts. Type 1 is the most common.

Diagnosis

Choledochal cysts present either as a mass or with biliary obstruction. The classic triad of abdominal pain, jaundice, and a mass occurs in less than one-half of patients. Most adults present with jaundice or cholangitis. Infants also present with jaundice but may vomit due to duodenal compression and may also have a palpable mass in the abdomen.

Ultrasonography will usually confirm the diagnosis. Cholangiography is absolutely essential to delineate the biliary anatomy accurately and thus determine the best

approach to surgical treatment. RCP is the easiest method of obtaining a cholangiogram (Fig. 16) but percutaneous and operative cholangiography may also be needed. An arteriogram to delineate the relation of the cyst to the hepatic artery and the portal vein may be very valuable before embarking on surgical treatment.



Fig. 16. Type I choledochal cyst, packed with stones, which presented in adult life. There is also a stricture of the common hepatic duct immediately at the junction with the cyst.

Treatment

Once diagnosed, the choledochal cyst is treated surgically. Although numerous operations have been described, a type-1 cyst should ideally be excised completely. The distal bile duct is then anastomosed to the small bowel, usually the jejunum, via a Roux-en-Y loop. In infants the structures are sufficiently mobile to allow use of the duodenum. Anastomoses between the cyst wall and the bowel are usually unsatisfactory in the long term. The cyst wall is often devoid of mucosa and is lined only by granulation tissue, so that a stricture develops. This leads to recurrent cholangitis and the development of stones within the cyst. The high risk of carcinoma developing in a choledochal cyst is also a reason for complete excision. This is usually possible but if difficult, it is acceptable simply to excise the lining of the cyst and thus protect the portal vein by a partial thickness of the cyst wall. Previous surgery, recurrent cholangitis, and portal hypertension all make treatment difficult.

Type-2 cysts can usually be excised, the defect in the common bile duct being repaired by primary suture over a T-tube brought out through a separate incision in the duct. Small choledochoceles (type 3) may be treated by endoscopic sphincterotomy. Larger ones require a surgical sphincterotomy using the transduodenal approach. Type-4 cysts are treated by a combination of techniques determined by the precise anatomy in each individual case. Segmental resection of the liver may be necessary, particularly if there are intrahepatic stones, strictures or abscesses as well as an extrahepatic cyst. Hepaticojejunostomy is then necessary to reconstitute biliary drainage.

Complications

Cysts may rupture spontaneously in infants, and cholangitis develops in both adults and children. Gallstones can develop within the cyst and are more common in adults (Fig. 16). Secondary biliary cirrhosis supervenes in 15 per cent of adults, with chronic obstruction of the bile ducts. Carcinoma develops within the cyst in 8 per cent of patients and is a fatal complication. The diagnosis is usually made at operation or at autopsy. Most patients are in their thirties and 75 per cent of tumours are adenocarcinomas, although squamous carcinoma and cholangiocarcinoma also occur. The mean survival after diagnosis is 8.5 months.

Acquired disease

Gallstones

The most common acquired abnormality of the biliary system is the presence of gallstones. In most patients these remain dormant, but they may cause biliary colic simply by mechanical obstruction. More frequently, the stones lead to acute or chronic inflammation within the biliary tree and cause symptoms. Inflammation without stones can occur in both the gallbladder and the bile duct, but is unusual, as are infarction of the gallbladder and benign neoplasms of the biliary tract. Damage to the bile duct is rare at the time of cholecystectomy. It may lead to the development of a benign biliary stricture, which is one of the most serious complications of biliary-tract surgery. Abnormalities of biliary function probably do occur but at the present time they are poorly understood.

Prevalence

Gallstones are very common, with a prevalence at autopsy of 11 to 36 per cent. There are at least 5 million people in the United Kingdom and 25 million in the United States of America with gallstones. Overall the prevalence has probably increased in Western societies over the last 50 years, and it certainly increases with age from 4 per cent of people in the third decade of life to 27 per cent in the seventh. This increase may be related to changes in the biochemistry of bile with age. Women are three times more likely than men to develop stones, and first-degree relatives of patients with gallstones have a two-fold greater prevalence. There are geographical variations. The prevalence is very high in certain American Indian communities (Pima and Chippewa tribes), in Mexico, Sweden, the former Czechoslovakia and Chile, and low in Greece, Japan, India, and China. Certain conditions predispose to the development of gallstones. Obesity is a risk factor for gallstones in women under the age of 50 years. Pregnancy, but not use of the oral contraceptive pill, does probably predispose to the development of gallstones (Fig. 17). Dietary factors that have been implicated include a high energy intake, increased consumption of unrefined carbohydrate, and diets low in fibre. Crohn's disease, terminal ileal resection and jejunoileal bypass for obesity are associated with a four-fold increase in prevalence. Biliary infection and parasitic infestation of the biliary tree are important factors in the development of pigment stones in Asia but not in the West. Other diseases associated with the development of gallstones include diabetes mellitus, type IV hyperlipoproteinaemia, cirrhosis of the liver, gastric surgery, and total parenteral nutrition. Patients with haemolytic anaemia due to hereditary spherocytosis, sickle-cell disease and thalassaemia also show an increased prevalence of pigment stones.

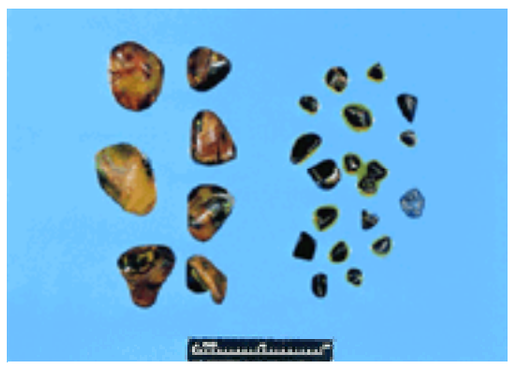


Fig. 17. Gallstones from a young woman with two children. The stones clearly divide into two groups of similar size and probably formed coincidentally with her two pregnancies.

Classification

Cholesterol and bile pigments are the two principal constituents of gallstones. In addition, calcium carbonate, phosphate, and palmitate are present in variable amounts. Pure cholesterol and pure pigment stones do occur, but most stones are composed of a mixture (Fig. 18). Predominantly cholesterol stones account for 75 per cent of all gallstones in the West. They are single or multiple, hard, and usually layered on cross-section (Fig. 19). Pigment stones are most common in Asia. They are usually black or brown in colour; brown stones crumble when squashed (Fig. 20). About 10 per cent of stones contain sufficient calcium to be radio-opaque (see Fig. 3).



Fig. 18. Mixed stones: they are clearly faceted and consist of layers of pigment and cholesterol.

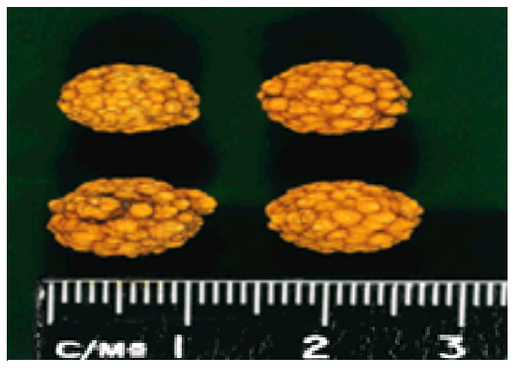


Fig. 19. Four pure cholesterol stones.

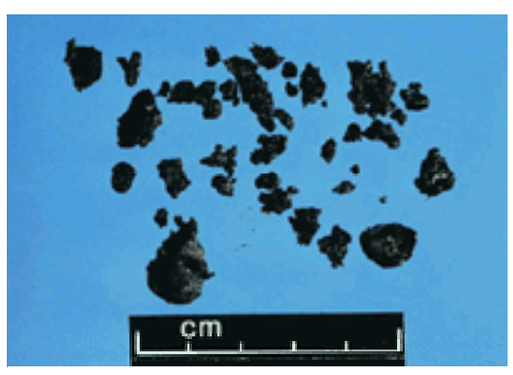


Fig. 20. Pigment gallstones: they are smaller and softer than mixed or cholesterol stones, and they are classically associated with haemolysis.

Formation

Cholesterol stones Cholesterol is insoluble in water, and is held in solution as micelles of cholesterol, phospholipids, and bile salts. Lecithin forms the major component of phospholipids; the bile salts are glycine or taurine conjugates. The physicochemical state of bile can be determined from a phase diagram (Fig. 21). The relative proportions of cholesterol, bile salts, and phospholipid are expressed as a percentage of the total lipid content and plotted on triangular coordinates. An increase in the cholesterol concentration or a decrease in the bile-salt concentration results in supersaturation of bile with cholesterol, with the formation of a liquid crystalline phase of cholesterol. The biliary lipid composition of normal bile and gallstone bile is virtually identical, and since at least half of the Western population have supersaturated bile there must be another factor responsible for the formation of stones. Cholesterol will only crystallize from a supersaturated solution if there is a nidus on which the crystals can form. This process is called nucleation, and the time taken for supersaturated bile to form crystals of cholesterol is known as the nucleation time. Normal bile takes 15 days to form crystals, compared with 3 days for bile from patients with cholesterol gallstones. Mucous glycoproteins from the gallbladder wall and bilirubinate have both been proposed as nucleating factors, while a bile protein has been proposed as an inhibitor. The nucleation phenomenon also depends on gallbladder function: the motility of the wall determines the degree of stasis and mixing of the bile within the lumen. One other significant finding is that the size of the bile-acid pool is reduced in many patients with gallstones, although the reason for this is unknown.

Pigment stones Pigment gallstones are formed of calcium bilirubinate and contain less than 25 per cent cholesterol (Fig. 20). They are usually small and multiple. As might be expected, they are more prevalent in patients with haemolytic disorders such as hereditary spherocytosis and sickle-cell disease, and in those with cirrhosis, who commonly have a mild degree of haemolysis. They are frequently found in oriental countries, where they are associated with parasitic infections. Bilirubin in bile is normally conjugated with glucuronide. The enzyme b-glucuronidase, which may be produced by bacteria such as *E. coli*, splits the molecule and the unconjugated bilirubin precipitates as the calcium salt. Hydrolytic enzymes from gallbladder mucosa might act in the same way.

Clinical presentation

Since the sixteenth century it has been realized that most people who have gallstones remain asymptomatic throughout their lives. In a few patients, stones are discovered by accident during the investigation of some other problem. In general these stones should be left alone. In a study of patients in whom gallbladder stones were discovered on screening, only 15 per cent developed biliary pain in the subsequent 15 years.

There may be a case for removing asymptomatic stones in diabetics because of their greater susceptibility to infection. For the same reason some units recommend a cholecystectomy prior to transplantation in patients who have gallstones. An incidental cholecystectomy for gallstones during a laparotomy for an unrelated condition can sometimes be appropriate because such patients are at greater risk of developing subsequent symptoms. There is no justification for removing the gallbladder with stones to prevent the later development of cancer.

Disease in the gallbladder affects only a small minority of people with gallstones. It presents with a variety of clinical syndromes, of which the most common are chronic cholecystitis, which appears slowly over time, biliary colic and acute cholecystitis, both of which develop rapidly. Patients occasionally present with the complications of one of these three, either locally in the gallbladder or from stones in the bile duct, and very occasionally from a stone that has ulcerated through into the bowel.

Chronic cholecystitis

Pathology

About two-thirds of patients present with chronic cholecystitis. The pathological changes, which often do not correlate well with symptoms, vary from those of an apparently normal gallbladder with minor chronic inflammation in the mucosa to a shrunken organ with gross transmural fibrosis and organized adhesions (Fig. 22). The mucosa is initially hypertrophied but later becomes atrophied; the epithelium protrudes into the muscle coat, leading to the formation of Aschoff –Rokitansky sinuses. The most severe form is represented by cholecystitis glandularis proliferans in which there are buried areas of hyperproliferative epithelium within the wall of the gallbladder (see Fig. 26). Rarely, dystrophic calcification may occur, resulting in the formation of a porcelain gallbladder (Fig. 23). Calcification is said to predispose

to malignant change.



Fig. 26. Severe cholecystitis glandularis proliferans; the very thick wall of the gallbladder is white and rather mucoid in appearance when it is cut open.



Fig. 23. A calcified (porcelain) gallbladder on a plain radiograph.

Diagnosis

The typical patient complains of recurrent attacks of right hypochondrial or epigastric pain, usually after meals, and particularly after consumption of fatty foods. The pain, which often occurs at night, varies from mild indigestion after eating to persistent, moderately severe pain in the right upper quadrant that may radiate round to the back and sometimes to the right shoulder or between the shoulder blades. The pain is often described as a tight band all the way round the upper abdomen; very occasionally the pain on the right-hand side of the abdomen is suppressed and the patient presents only with left-sided pain.

Nausea usually accompanies the pain, sometimes associated with vomiting. There may be additional minor symptoms such as flatulence and abdominal distension, but these are equally common in patients who do not have gallstones. There may be mild, right upper quadrant, abdominal tenderness, but examination is usually unremarkable.

It is often possible to make a confident clinical diagnosis of chronic cholecystitis in a patient with classical symptoms but the presence of gallstones should be confirmed by ultrasonography (see Fig. 5). When it is difficult to distinguish chronic cholecystitis from peptic ulceration, a hiatus hernia or diverticular disease, then further radiological and endoscopic examination may be required. Hiatus hernia and diverticular disease often occur together with gallstones, a condition known as Saint's triad.

It is easy to ascribe the patient's symptoms to the stones found on investigation when this is not the case: many patients with other conditions have gallstones. On the other hand, patients with symptoms that show enough features related to their proven gallstones are likely to improve with treatment.

Treatment

Once the diagnosis is established some form of active treatment is indicated, as the symptoms will almost always continue. Some patients can control their symptoms by taking care with their diet. Others are only occasionally troubled and simply require mild analgesia. More commonly, some form of surgical treatment is needed. The risks of surgery must be balanced against the potential benefits, and the views of the informed patient are just as important as the opinion of the doctor. Most surgeons prefer the patient to request an operation.

Biliary colic

Biliary colic is due to impaction of a stone in the neck of the gallbladder or the ampulla of Vater. The severe pain starts abruptly in the epigastrium, often at night after a heavy meal, and lasts for several hours. It is usually continuous and is associated with restlessness, vomiting, and sweating. The pain may radiate through to the back but does not radiate to the shoulder, as in acute cholecystitis. General examination may disclose a patient in obvious severe pain, with a mild tachycardia and normal temperature. Abdominal examination shows only mild tenderness in the epigastrium. In contrast to acute cholecystitis, tenderness over the gallbladder is absent. Occasionally, transient jaundice may occur with stones that impact at the the ampulla. Most patients need a strong analgesic given by injection, and after two attacks of severe biliary colic will want some form of definitive treatment. At operation the gallbladder is often normal on external appearance and shows only mild inflammatory changes on histological examination.

Acute cholecystitis

Pathology

About one-fifth of patients first present with acute cholecystitis; in about one-third there is clinical or pathological evidence of previous chronic cholecystitis. It is usually due to persistent impaction of a stone in the neck of the gallbladder. The result is initially a chemical inflammation of the gallbladder wall, perhaps due to the mucosal toxin, lysolecithin, produced by the action of phospholipase on biliary lecithin. This is soon followed by bacterial infection. Because the cystic duct is occluded the inflammatory process is particularly aggressive and the gallbladder becomes acutely distended with accompanying lymphatic and venous obstruction. The serosa may be covered by a fibrinous exudate and subserosal haemorrhage gives the appearance of patchy gangrene. The gallbladder wall itself is grossly thickened and oedematous, and the underlying mucosa may show hyperaemia or patchy necrosis (Fig. 24). Histologically, three grades of inflammation are recognized: acute cholecystitis, acute suppurative cholecystitis, and acute gangrenous cholecystitis. Rarely, an abscess or empyema develops within the gallbladder, while perforation of an ischaemic area leads to a pericholecystic abscess, bile peritonitis or a cholecystenteric fistula.

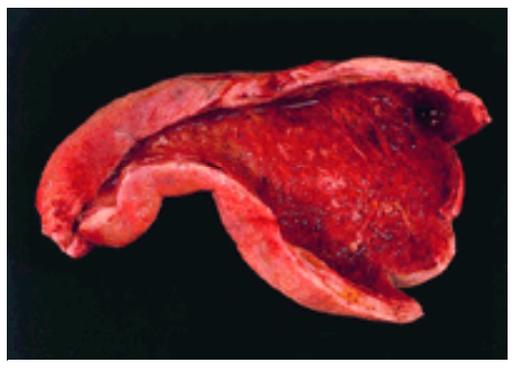


Fig. 24. A gallbladder removed at an emergency operation showing severe acute cholecystitis. Note the reddened mucosa and the thickening of the gallbladder wall.

Diagnosis

Patients present with acute upper abdominal pain but, in contrast to those with biliary colic, the pain has often been present for 2 or 3 days. Because the inflammation extends to the parietal peritoneum, the pain is well localized and it hurts the patient to move or breathe. Patients feel generally unwell, may have been febrile, and are anorexic. Physical signs vary with the severity of the inflammation, but there is usually some degree of fever and tachycardia. Mild jaundice is present in 10 to 15 per cent of patients. Right hypochondrial tenderness is invariable, and there may also be guarding, rigidity, and rebound tenderness as well. If the latter three physical signs are subdued it may be possible to feel the gallbladder itself. Murphy's sign (inspiratory arrest during subcostal palpation) is widely regarded as pathognomonic of cholecystitis. It is certainly present in patients with established acute cholecystitis, but it only reflects peritoneal irritation in the right upper quadrant, other causes of which include chronic cholecystitis, acute hepatitis, and a localized abscess around a perforated duodenal ulcer. There is usually a clear distinction between acute cholecystitis and biliary colic: this is important because the management is different.

In elderly patients, acute cholecystitis may present more insidiously and the frequent absence of typical physical signs results in a delay in diagnosis. In addition, the incidence of complications is higher and the prevalence of intercurrent illness combines to increase the mortality 10-fold. Acute cholecystitis is uncommon in children, when it is usually associated with gallstones that develop as a complication of haemolytic disease. Acalculous cholecystitis occurs in children, with severe sepsis.

Differential diagnosis

Clinically it can be difficult to distinguish acute cholecystitis from acute pancreatitis, acute appendicitis, acute pyelonephritis, perforation of a peptic ulcer, and, occasionally, biliary colic. A raised white-cell count and serum amylase may occur in several of these conditions. Patients with biliary colic rarely have a leucocytosis. Urine should always be tested and sent for culture if appropriate. One-quarter of patients have disturbed liver function, but not all of them will have stones in the bile duct. There are rarely any specific features of acute cholecystitis on plain radiographs, but ultrasound may localize the tenderness to the gallbladder and may demonstrate stones. Free air under the diaphragm on a chest radiograph implies perforation of a viscus, usually a peptic ulcer. A normal HIDA scan excludes acute cholecystitis (see Fig. 11).

Young women who present acutely with severe pain in the right upper quadrant and signs of peritonitis may have the Curtis–Fitz-Hugh syndrome. Clinically, these patients appear to have acute cholecystitis but ultrasonographic examination fails to show gallstones or any signs of acute cholecystitis. At laparoscopy or laparotomy the gallbladder is normal but there are string-like adhesions between the liver and the peritoneum. This perihepatitis is caused by infection with *Chlamydia trachomatis*. There may also be evidence of genital-tract infection with the same organism. The diagnosis can be confirmed by isolation of the organism from peritoneal fluid or by rising titres of chlamydial antibodies in the serum. Treatment is with oxytetracycline, 2 g daily for 10 days.

Acute viral hepatitis can sometimes present like acute cholecystitis. The acutely swollen liver is painful and tender but the systemic symptoms and the onset of jaundice soon make the true diagnosis clear.

Treatment

Acute cholecystitis resolves with conservative treatment in the majority of cases. If admission to hospital is necessary, patients require intravenous fluids, analgesia, and suspension of oral intake. Vomiting is unusual, but if present nasogastric aspiration is helpful. If the patient fails to respond intravenous antibiotics are prescribed. Our present choice is cefuroxime, 1.5 g three times a day.

Most patients should be offered a cholecystectomy, which should normally be undertaken on the next convenient operating list. There is no advantage in letting the acute illness subside and removing the gallbladder 6 weeks later, except in a patient who is unfit for surgery and whose condition could be improved by waiting.

Complications

An empyema of the gallbladder should be suspected clinically if the physical signs and symptoms fail to improve or worsen on conservative management. In particular, fever and tenderness in the right upper quadrant fail to abate, and there is a persistent or increasing leucocytosis. With time the gallbladder becomes necrotic and ruptures, resulting either in a localized abscess or in generalized peritonitis. An empyema is an abscess within the gallbladder and it must therefore be drained. The best method is to insert a pigtail catheter into the gallbladder under ultrasound control, as the gallbladder is usually adherent to the peritoneum of the abdominal wall (Fig. 25). If there is any doubt a transhepatic route for the catheter should be chosen. Percutaneous drainage is clearly less disturbing for the patient, who is usually quite ill and toxic. If it fails for any reason, then a conventional surgical approach must be adopted. Occasionally in these circumstances a safe cholecystectomy can be performed by an experienced surgeon. Otherwise a cholecystostomy is better. A cholecystostomy drain should remain in place until a contrast study shows that there is free drainage from the gallbladder into the bile duct and duodenum. This normally takes about 7 days.



Fig. 25. Radiograph showing a drain in the gallbladder in a patient with severe acute cholecystitis.

Acute emphysematous cholecystitis

This is a severe and fulminant form of acute cholecystitis that accounts for less than 1 per cent of cases. Stones are absent in 30 to 50 per cent. The patient is usually elderly and male, and 40 per cent have diabetes mellitus. It is caused by a mixture of bacteria that includes gas-forming organisms, and the pathognomonic diagnostic sign is gas within the wall or the lumen of the gallbladder seen on a plain radiograph. The onset of the disease is abrupt and the condition of the patient deteriorates rapidly. There is a high incidence of gangrene and perforation, and emergency cholecystectomy is needed.

Xanthogranulomatous cholecystitis

This is a rare but severe form of chronic cholecystitis in which the gallbladder is thickened and irregular with extensions of yellow xanthogranulomatous inflammation to adjacent organs. Foamy macrophages and giant cells are seen within connective tissue in the wall of the gallbladder. The condition is thought to be due to bile penetrating deeply into the gallbladder wall. The appearances both on investigation and at operation resemble carcinoma of the gallbladder; frozen-section histology at operation may be required because the two conditions are associated.

Acute acalculous cholecystitis

Acute cholecystitis can develop in the absence of stones in the gallbladder. It is most often seen in patients in the intensive care unit and is associated with severe illness such as multiple trauma, extensive burns, major surgery and sepsis, often in an elderly person. The cause is unknown but is thought to be related to gallbladder distension and bile stasis. The normal contraction of the gallbladder is inhibited in patients with sepsis and those on total parenteral nutrition, especially if opiate analgesics are given. This allows the development of biliary sludge, which may be demonstrated in the gallbladder of many patients with major illness.

Pathology

Pathological examination of the gallbladder reveals oedema of the serosa and muscular layers, with patchy thrombosis of arterioles and venules. Areas of necrosis may affect the underlying mucosa. One possibility is that activation of factor VII by trauma could lead to thrombosis of blood vessels in the seromuscular layer of the gallbladder.

Diagnosis

In a severely ill patient, the development of acute acalculous cholecystitis is usually insidious. The clinical features are similar to those of acute calculous cholecystitis but they are often masked by the underlying condition. Ultrasonography is the most useful investigation and may show biliary sludge in a tender, thickened gallbladder, but fails to demonstrate stones. All the indicators of liver function deteriorate, and a HIDA scan will fail to demonstrate the gallbladder.

Treatment

Acute acalculous cholecystitis requires urgent treatment. Percutaneous drainage of the gallbladder with ultrasound or CT guidance may be sufficient but an immediate cholecystectomy is sometimes necessary. The mortality varies with the nature of the underlying condition but is generally higher than for acute calculous cholecystitis.

Cholesterolosis

This is caused by the deposition of cholesterol in the mucosa and submucosa of the gallbladder wall and produces the classic macroscopic appearance of a ‘strawberry gallbladder'. Microscopy shows macrophages loaded with cholesterol. Ultrasound identifies the cholesterol in the wall as bright shiny spots and there may also be cholesterol stones within the lumen. Cholesterolosis may cause pancreatitis, perhaps as a result of small cholesterol crystals passing down the bile duct and briefly occluding the ampulla, so that symptomatic patients should be advised to undergo cholecystectomy.

Adenomyomatosis

Adenomyomatosis or cholecystitis glandularis proliferans is characterized on microscopy by hypertrophic smooth-muscle bundles and epithelial sinus formation. The gallbladder has a thickened wall that may be divided into two separate sections by a stricture or incomplete septum (Fig. 26). Granulomatous polyps develop in the lumen at the fundus. Inflammation develops later and gallstones are sometimes present. Symptomatic patients require a cholecystectomy. Others in whom the diagnosis is made but not treated require surveillance since adenomyomatosis can predispose to carcinoma.

Mucocele of the gallbladder

A mucocele of the gallbladder forms when a stone impacts in the cystic duct but bacterial infection does not occur. Bile is reabsorbed but the epithelium continues to secrete mucus and the gallbladder becomes distended (Fig. 27, Fig. 28). It is easily palpable and may even be visible but it is not tender. Such patients have somewhat subdued but nevertheless persistent symptoms, often including distressing nausea. Once diagnosed, patients should be advised to have a cholecystectomy. If infection occurs an empyema develops rapidly. This will require urgent percutaneous drainage or surgery. Rarely a mucocele of the gallbladder may perforate. Although pseudomyxoma peritonei has been reported to follow rupture of a mucocele, it probably only follows rupture of a cystadenoma or cystadenocarcinoma of the gallbladder.

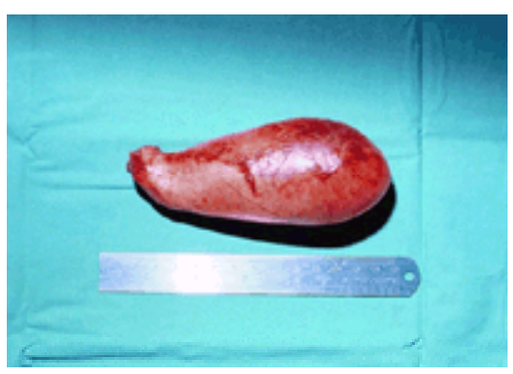


Fig. 27. A mucocele of the gallbladder, unopened.

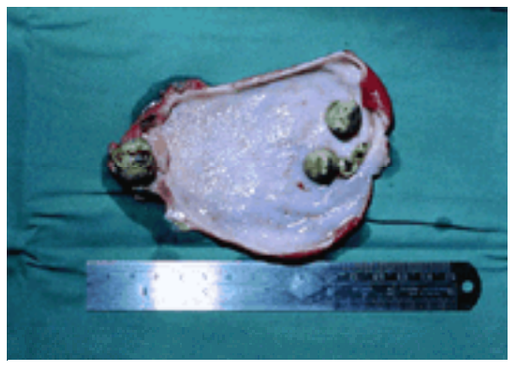


Fig. 28. The same specimen as Fig. 27 opened to show the white shiny mucosal surface and one stone stuck in Hartmann's pouch.

Torsion of the gallbladder

Infarction of the gallbladder due to torsion or volvulus is a rare event. Two anatomical anomalies permit torsion. Firstly the gallbladder may have no attachment to the liver, lying free in the peritoneal cavity suspended only by the cystic duct and artery. Secondly, and more commonly, the gallbladder is suspended from the liver by a narrow mesentery. Acute torsion causes right-sided abdominal pain and the tense, infarcted gallbladder may be palpable (Fig. 29). It is often misdiagnosed as acute

appendicitis. Intermittent torsion can occur and produces periodic bouts of pain.

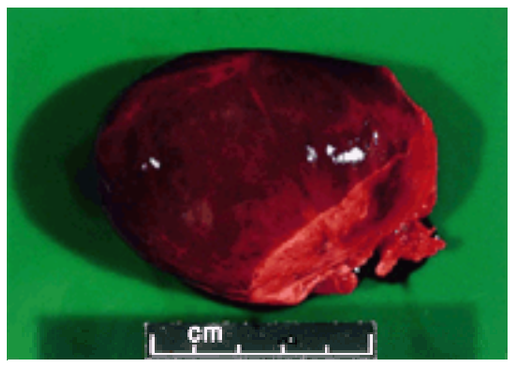


Fig. 29. Torsion of the gallbladder: the line across which the fundus of the gallbladder twisted can be clearly seen and the fundus is congested and gangrene is developing.

Biliary pain without stones

A small group of patients, usually young women, present with pain in the right hypochondrium that, in the opinion of everyone who sees them, is typical of biliary pain. However, all conventional biliary investigations, which may have been repeated on several occasions, are normal. Furthermore a minority benefit from cholecystectomy even though no pathological abnormality is discovered in the gallbladder at operation.

Now that cholesterolosis and adenomyomatosis can be excluded ultrasonographically before operation, interest has centred on the possibility that these patients have a functional disorder of the biliary tract. This idea has received support from the discovery that some of them will develop an identical pain after intravenous injection of cholecystokinin. The hope that this test would identify those patients who would benefit from a cholecystectomy has not been confirmed by clinical trials.

In practical terms, it is essential to exclude the irritable bowel syndrome, which can produce very similar symptoms to biliary pain. After explaining the position very carefully to the patient it is reasonable to proceed to a cholecystectomy. Unfortunately, the pain persists after operation in some patients; these form part of a group of patients with postcholecystectomy pain.

Management of gallbladder stones

The first successful cholecystectomy was performed by Langenbuch in 1882. Over the subsequent 100 years the open abdominal operation became the standard treatment for gallbladder stones. The mortality rate was reduced to 1 per cent and the morbidity rate to 10 per cent but it still took 6 to 8 weeks to recover and an alternative to open surgery had many attractions. Drugs that dissolved cholesterol gallstones were the first idea to be introduced. But they were only of use in a few patients. Then there were developments in engineering and other technologies that produced a bewildering array of methods for removing stones. All of them, with the possible exception of oral bile-salt therapy, were consigned to history with the introduction of laparoscopic cholecystectomy in 1987 by Phillippe Mouret in Lyons, France.

Gallstone dissolution therapy

Cholesterol gallstones can be dissolved by decreasing the cholesterol saturation of bile (see Fig. 21). This can be achieved by giving natural chenodeoxycholic acid (10–15 mg/kg per day) or synthetic ursodeoxycholic acid (8–12 mg/kg per day and up to 15 mg/kg per day in obese patients) by mouth. Both drugs are equally good at dissolving gallstones *in vivo* but the stones must be less than 15 mm in diameter, without calcification and within a functioning gallbladder. The drugs are least effective in obese patients and cannot be used in patients with acute cholecystitis or biliary colic. Liver function is carefully monitored during treatment and the stones are measured at 6 months. If there has been no reduction in size then there is no point in continuing with treatment. Small stones will dissolve in 6 months in 80 per cent of patients but larger stones will require up to 2 years' treatment. Even then only 14 per cent of patients were stone free in the largest reported study. After cessation of treatment 50 per cent of patients will develop recurrence of gallstones within 5 years. Overall, dissolution therapy provides a viable alternative to surgery in a very small proportion of carefully selected patients.

Cholecystectomy

Cholecystectomy is the most common major abdominal operation in the Western world. An elective operation can be planned for the convenience of the patient and the surgeon. Patients with acute complications usually respond to initial conservative treatment and should undergo operation on the next convenient list. As noted above, there is no surgical advantage in waiting for 6 weeks whilst the inflammation subsides although there may, on occasion, be medical advantages.

The essential features of the operation are common to any cholecystectomy. General anaesthesia and good exposure of the operative field are essential. The cystic duct and the cystic artery must be dissected, identified with certainty, occluded, and divided. The gallbladder is then removed from the liver bed. The place of routine peroperative cholangiography remains controversial. It helps to delineate the biliary anatomy and it will also identify stones in the bile duct. The differences between the various operations are in the number and the length of the incisions necessary and the time that it takes the patient to recover.

Laparoscopic cholecystectomy

Patient selection

Every patient who needs a cholecystectomy is potentially suitable for a laparoscopic procedure. There are no absolute contraindications. Peritoneal adhesions, pregnancy, and portal hypertension can make laparoscopy difficult. Obesity is not a contraindication. The view of the operative field is just as good in a fat patient as it is in a thin, and the absence of an abdominal incision reduces the incidence of complications.

The 1 in 20 chance that a laparoscopic operation will be converted to an open procedure must be taken into account when advising patients about an operation. The risk is higher when the disease to be treated is more severe or more acute. Patients must be warned about this beforehand and must consent to a laparotomy as well as a laparoscopic operation.

About 1 in 10 patients will also have stones in the bile duct. It is important to identify these patients in advance and to decide how these bile duct stones are to be removed. A history of jaundice, abnormal liver function tests, and dilatation of the bile ducts on ultrasonographic examination all suggest that there may be a stone in the duct. Patients with more than one positive feature must have a cholangiogram done. Some surgeons prefer to do this before the operation and obtain an intravenous or endoscopic (ERCP) cholangiogram. Others prefer to do peroperative cholangiography. Magnetic resonance cholangiography is likely to be used in the future (see Fig. 13). ERCP allows removal of the stones through an endoscopic sphincterotomy. When the results of the other investigations are available, a decision must be made whether to remove the bile-duct stones laparoscopically or to advise a conventional open cholecystectomy and exploration of the common bile duct. The decision will depend on the experience and practice of the individual surgeon.

Preoperative preparation

Most elective patients are seen in a preadmission clinic and admitted on the day of surgery. The results of a full blood count and liver function tests should be available. Blood should be grouped and serum saved for cross-matching. For inpatients, fluid depletion and electrolyte imbalance must also be corrected. Prophylaxis against deep venous thrombosis is necessary. Elective female patients should stop the contraceptive pill 1 month before admission. All patients should wear compression stockings and receive heparin 5000 units subcutaneously twice a day. Patients are asked to empty their bladder before coming to theatre and are not normally catheterized. Antibiotic prophylaxis with cefuroxime 1.5 g intravenously is given on induction of anaesthesia.

Operative technique

The patient lies flat on the operating table and the surgeon stands on the left side with the television monitors and the associated electronic equipment on either side at the head of the table (Fig. 30). Some surgeons prefer the patient in the Lloyd Davies position and stand between the legs to operate. A nasogastric tube is placed only if the stomach is distended with gas. It is removed at the end of the operation. The exposure of Calot's triangle is sometimes improved by elevating the head of the table and rotating the patient to the left.

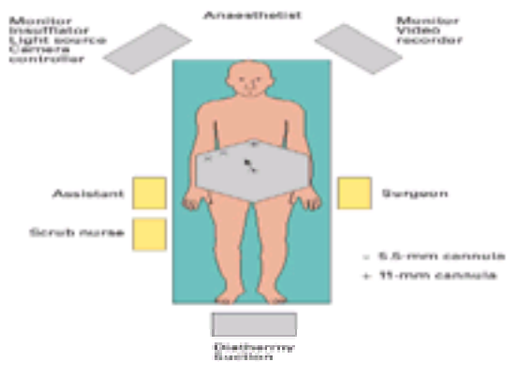


Fig. 30. The layout of the operating theatre during a laparoscopic cholecystectomy. The television monitors with the associated light source and camera controller are all located at the head of the operating table. The surgeon stands on the patient's left-hand side and watches the monitor at the patient's right shoulder. The diathermy machine is most conveniently located at the patient's feet along with the sucker.

Placing the cannulas The Veress needle and the first 11-mm cannula are placed through the umbilicus (Fig. 31). We do not hesitate to adopt an open technique if there is any difficulty and it is sometimes easier to do so if there is a paraumbilical hernia. The hernia can be repaired at the end of the operation. The abdomen is inspected as soon as the endoscope is inserted. It is important to check that the initial puncture has not caused damage to the bowel. The stomach and the liver are well seen and so is the gallbladder unless there are dense local adhesions. Three further ports are needed and they should be placed under direct vision and after careful assessment of the anatomy of the individual patient. An 11-mm port is placed in the epigastrium. The tip should come through the right leaf of the falciform ligament. Two 5.5-mm ports are placed immediately below the right costal margin (Fig. 32). The medial one should complement the epigastric port and lie in the correct position in relation to the operative field; the lateral one is placed over the fundus of the gallbladder. Occasionally, in fat patients, a fifth port to retract the duodenum and the colon is helpful and then a 5.5-mm cannula is inserted just to the left of the midline halfway between the umbilical and the epigastric ports. Cholangiography catheters and T-tube drains either come with their own special ports or are inserted through separate small incisions placed below the costal margin.

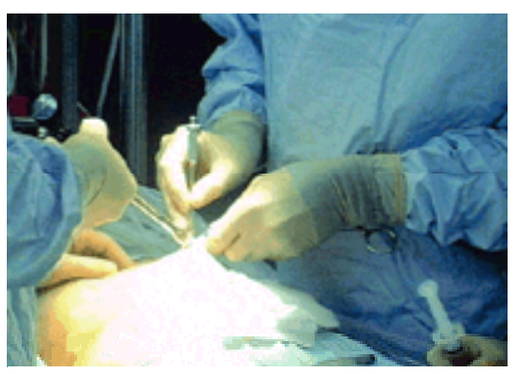


Fig. 31. The Veress needle is being inserted through the umbilicus; note that the abdominal wall is held up to provide countertraction.

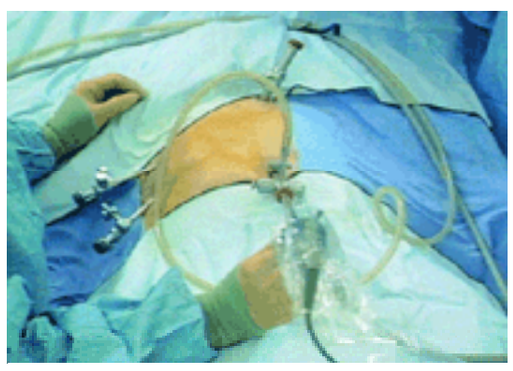


Fig. 32. Layout of the cannulas for a laparoscopic cholecystectomy. The picture is taken from the foot of the operating table looking towards the patient's head. The telescope with the attached television camera is placed through the umbilicus and an 11-mm port in the epigastrium and two 5.5-mm ports below the right costal margin. Gas insufflation is through the umbilical cannula.

Dissecting Calot's triangle A grasping forceps is placed through the right lateral port and is fixed to the fundus of the gallbladder. The gallbladder and the attached liver are then retracted upwards towards the patient's right shoulder. This opens up the subhepatic space (Fig. 33). This manoeuvre is sometimes easier if the gallbladder is emptied first. This is best done by plunging the lateral trocar and cannula directly into the fundus of the gallbladder and then aspirating down the cannula. Adhesions to the gallbladder are divided and a second grasping forceps is attached to the neck of the gallbladder and pulled laterally to expose Calot's triangle.

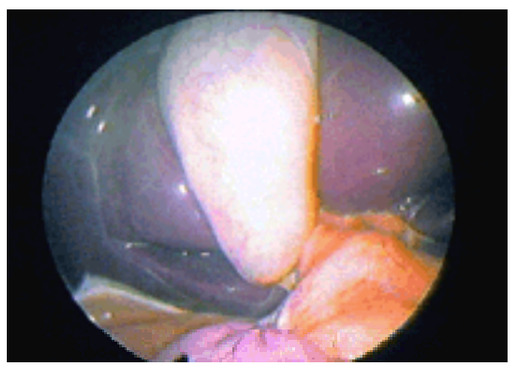


Fig. 33. View of Calot's triangle.

Dissection through the epigastric port with the dolphin-nosed or Petelin curved forceps (Fig. 34) starts at the junction of the gallbladder with the cystic duct. The peritoneum, fat, and loose areolar tissue around the gallbladder are grasped and gently torn down towards the bile duct. This continues until the inferior aspect of the cystic duct is clearly seen. Any troublesome bleeding should be stopped at once with diathermy but care must be taken not to burn the bile duct. The superior aspect of the cystic duct is cleaned in the same way and eventually the duct is dissected on all sides. Curved forceps can be useful for this (Fig. 35). The cystic duct is therefore clearly identified as a direct extension from the neck of the gallbladder or Hartmann's pouch. This is important because the junction of the cystic duct with the common bile duct is not always seen.

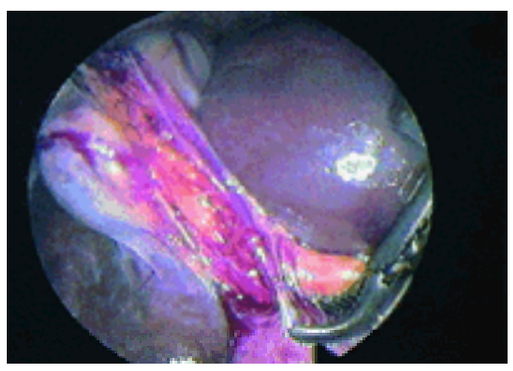


Fig. 34. Dissection within Calot's triangle with Petelin curved forceps.



Fig. 35. A curved Petelin's dissector alongside the dolphin-nosed dissecting forceps. Note the grooves on the outside of the blades of the dolphin-nosed instrument that make separation of the tissues easier.

Most of this dissection is done from the front, but, by varying the position and the placement of the forceps on the neck of the gallbladder, it is possible to expose the posterior aspect of Calot's triangle. This dissection can be helped by dividing the peritoneum attaching the gallbladder to the liver for a short distance.

The cystic artery is then exposed in the same way. Normally there is an obvious length of cystic artery running across Calot's triangle. Sometimes, however, the triangle is mostly occupied by the right hepatic artery and the cystic artery or arteries are very short. It is possible to misidentify these structures, and care must be taken not to clip and divide the right hepatic artery.

Once the cystic artery and the cystic duct are clearly identified, they are both clipped and divided (Fig. 36). Sometimes the cystic duct must be divided before the cystic artery can be safely dissected. The cystic duct should be occluded with a locking clip, and, if the duct is too large to fit inside, a clip it must be tied with an extracorporeal knot such as a Roeder loop (Fig. 37). Catgut, polyglactin, or polyglycolic acid are suitable materials.

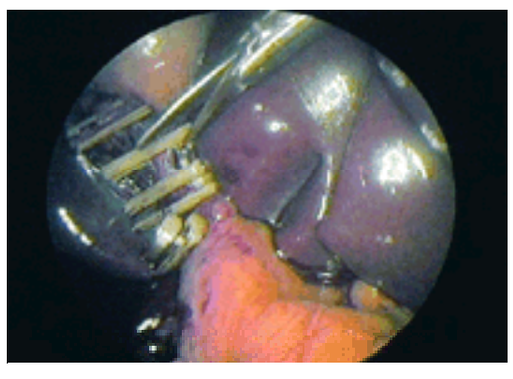


Fig. 36. The cystic duct has been clipped and divided. Three clips are already placed on the cystic artery. The scissors are about to divide beneath the top clip and so leave two clips on the arterial stump. The clips are made of synthetic material that will dissolve over about 90 days.

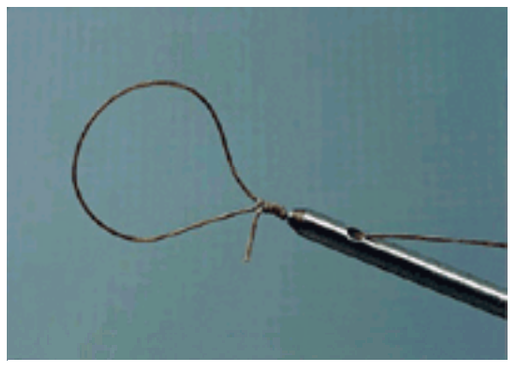


Fig. 37. A chromic catgut Roeder loop with a purpose-designed ligature pusher for endoscopic ligation.

The dissection is not always easy, particularly if there is a lot of fibrosis from chronic inflammation. The loss of tactile sensation is a disadvantage but there are some substitutes. Significant structures, such as an artery or a duct, contain elastic tissue and will twang if they are tweaked. Fibrous tissue has no such elasticity. Retrograde dissection from the fundus downwards is not feasible because upward retraction of the gallbladder is essential for exposure of Calot's triangle. As always, it is important not to rush a difficult dissection because time often reveals the anatomy. Furthermore, the variations in anatomy are no different from those in an open operation and the same rules therefore apply. No significant structure should be divided until the surgeon is completely happy that s/he understands the anatomy. If

there is any doubt then either a cholangiogram should be done or the procedure converted to an open operation.

Peroperative cholangiography The procedure is the same as at an open operation, although the technical details differ. The cystic duct is occluded close to the gallbladder with a clip and a small proximal incision is made in the cystic duct with fine scissors. A catheter is guided into the duct and held there with a special instrument or a loosely applied titanium clip. Air bubbles must be excluded and the cannulas, which are radio-opaque, should be held out of the way whilst the radiographs are taken. We usually use a 1.3-mm umbilical catheter inserted through an intravenous cannula (size 14), although specially designed cholangiography catheters with their own introducer are available and are easier to use (Fig. 38).



Fig. 38. A normal laparoscopic peroperative cholangiogram. The cystic duct is cannulated in the normal way and the cannula held in place either with a special instrument or with a removable metal clip. The laparoscopic ports are not radio-opaque and must be held out of the way whilst the radiographs are taken.

Dissection and removal of the gallbladder Once the cystic duct and the cystic artery are divided the free edge of the lesser omentum falls away. The midcostal grasper is repositioned just above the cystic duct clip and it is pulled upwards and away from the liver to expose the plane between the liver and the gallbladder. The gallbladder is then dissected off the liver bed with scissors or diathermy, starting at the bottom (Fig. 39). Bleeding points on the liver bed are coagulated as dissection proceeds. Dissection at the fundus is often difficult and simply requires patience and constant changes of position. Eventually the gallbladder is only just attached to the liver at the fundus. This is divided after a final inspection of the cystic duct and artery.

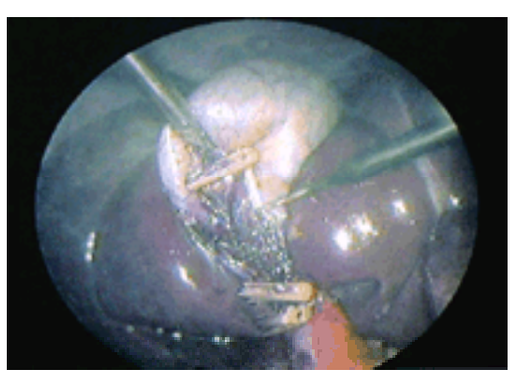


Fig. 39. Dissecting the gallbladder from the liver with a diathermy hook.

If the gallbladder is perforated, and this is preferable to entering the liver itself, bile and as many stones as possible should be removed by suction. Nevertheless, stones are often left behind in the peritoneal cavity where, somewhat surprisingly, they very rarely cause complications. It is certainly not necessary to try and retrieve every stone that is lost.

It is easiest to remove the gallbladder by using a retrieval bag (Fig. 40). This is rolled up and inserted through the epigastric cannula. The gallbladder and any loose stones are placed inside the bag, which is then withdrawn through the epigastric port. This may require enlargement of the incision in the abdominal wall if the gallbladder or the stones are large. Any bile or blood that have accumulated is sucked away and a drain can be left to the gallbladder bed if desired.

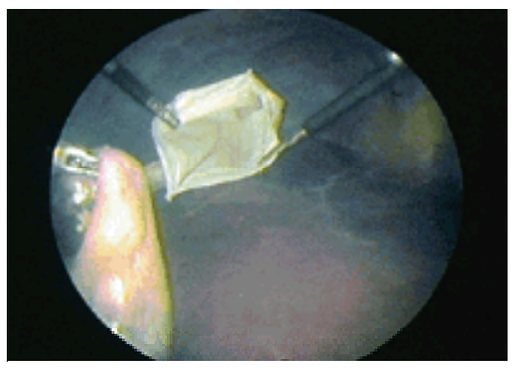


Fig. 40. The gallbladder has been dissected from the liver and is being placed inside a retrieval bag.

A final inspection of the abdomen is appropriate before all the carbon dioxide gas is allowed to escape and the cannulas are removed. The abdominal muscles are sutured if a wound has been enlarged and the skin is closed with adhesive tape rather than sutures.

Management of stones in the bile duct

The presence of stones in the bile duct should be established and the management of those stones decided before the patient is taken to theatre for a laparoscopic cholecystectomy. Preoperative cholangiography determines the best strategy.

Small stones in a small duct will probably pass spontaneously and are best left alone. Larger stones discovered on intravenous or magnetic resonance cholangiography can either be removed at an ERCP or at operation. If stones are identified at an ERCP, then the endoscopist should perform a sphincterotomy and remove the stones in appropriate patients. Laparoscopic cholecystectomy can follow at any convenient time thereafter.

A sphincterotomy in a patient under the age of 50 years is unwise unless there are special reasons such as obesity, severe jaundice or sepsis. Some stones are too large to remove endoscopically. Both these groups of patients require surgery. Laparoscopic extraction of stones from the bile duct through the cystic duct is perfectly possible. The cystic duct is first dilated with a balloon and then instruments, including a narrow cholangioscope, can be passed into the bile duct to retrieve the stones. Hepatic duct stones are difficult to extract, particularly if the cystic duct enters the common duct low down behind the pancreas. A fine tube can be left in the cystic duct at the end to act as a drain. It can also be used for cholangiography.

A conventional exploration of the bile duct through a choledochotomy is equally possible laparoscopically. It is particularly suitable for patients with a single stone in a

large dilated duct. A 30° forward oblique telescope gives the best view of the bile duct and a supraduodenal choledochotomy is made in the anterior surface of the duct with the diathermy hook and scissors. Stones can be washed out of the duct or removed with a balloon or a basket (Fig. 41). The ducts are then inspected with a choledochoscope and any remaining stones removed. At the end a T-tube is sutured into the bile duct.



Fig. 41. Laparoscopic exploration of the common bile duct. Two large stones are being extracted from the bile duct with a balloon. A clip on the cystic duct can be seen at the top of the picture and the grasping forceps is holding the choledochotomy open.

If stones are discovered in the bile ducts unexpectedly during the course of a laparoscopic cholecystectomy, their management depends upon their size and the experience of the operating surgeon. The choice lies between exploring the duct laparoscopically, converting to an open operation, or a postoperative ERCP. If stones are deliberately left behind a drain should be placed next to the cystic duct stump in case it leaks.

Difficulties, conversion, and complications

Inserting a laparoscopic port can cause spectacular damage to bowel or blood vessels. Invisible retroperitoneal aortic injury is one cause of acute collapse during a laparoscopic operation. Such injuries are fortunately very rare and must be resolved before the operation proceeds any further.

Sudden severe haemorrhage during a cholecystectomy is always a problem. It is best avoided in the first place. A torn cystic artery or a liver laceration are the common causes and both can be difficult to stop. The usual general principles apply. Blood and clot are sucked or washed away and the best possible exposure is obtained whilst the surgeon applies local pressure to the bleeding area. The scrub nurse prepares the diathermy, a titanium-clip applier or a suitable suture. When everything is ready the pressure is removed and the bleeding point is identified and secured. This is always more difficult to achieve in an endoscopic environment and there is sometimes no alternative to a laparotomy.

Conversion to an open operation demonstrates good surgical judgement and must not be regarded as a failure. The three main reasons why conversion is required are difficulty in identifying the anatomy, haemorrhage, and damage to another structure. Obviously the discovery of some other, unsuspected, condition may also require a laparotomy.

Every surgeon fears damaging a bile duct. In large published series the accident happens once in every 300 laparoscopic cholecystectomies. This is slightly more frequent than during an open operation. Great care should be taken to identify the anatomy correctly and to place any clips properly. A damaged bile duct should be repaired at once by an experienced surgeon, preferably at the time that it occurs.

Reactionary haemorrhage is rare and requires a laparotomy, although a repeat laparoscopy may be sufficient. Bile leaks from the stump of the cystic duct or the liver bed present with pain in the right upper quadrant, fever, and a subhepatic collection on ultrasonography. The bile is removed with a percutaneous drain. The leak is confirmed at an ERCP and placement of a stent in the bile duct usually resolves the leak. Surprisingly, a leak will not stop with a sphincterotomy alone. The drain remains until the drainage ceases and the stent is removed not less than 2 months later.

Benefits and results

Most patients are mobile and can eat and drink on the day of operation. Some patients are troubled by nausea and vomiting but only about half require opiate analgesia. Any drain can be removed after about 12 h. The average stay in hospital is about 24 h and many suitable patients can be treated as day cases. The majority are back at work within 2 weeks. The mortality rate after a laparoscopic cholecystectomy is about 0.1 per cent and the complication rate is about 4 per cent. Both figures are much better than for an open operation and the cosmetic benefits of four small incisions are important to many patients.

Open cholecystectomy

This has become an uncommon, but not an extinct, operation. It is most commonly performed after conversion of a laparoscopic operation. Occasionally it is an incidental addition to another procedure, but open cholecystectomy and exploration of the common bile duct are still required when bile duct stones cannot be removed in any other way. Because open cholecystectomy has become so uncommon it is vital that every opportunity be taken to teach the operation to surgeons in training.

Preoperative preparation

The preparation for an open operation is the same as for the laparoscopic procedure. Most patients can be admitted on the day of surgery and prophylaxis against infection and thromboembolism is essential.

Operative technique

Incision Four incisions can be used for cholecystectomy: midline, right paramedian, right subcostal, or a right transverse. A midline incision is useful when the diagnosis is not definite. A long subcostal incision gives the best exposure of the biliary tract when difficulties are expected. Access to the rest of the abdomen is poor, but fortunately improved preoperative diagnosis has reduced the need for a full diagnostic laparotomy. A transverse incision gives a better cosmetic result at the expense of exposure. A minicholecystectomy is performed through a very short subcostal incision. Choosing the most appropriate incision for any particular patient depends partly on the preference of the surgeon, partly on the build of the patient, and partly on the expected pathological condition.

Dissecting Calot's triangle The operative field should be exposed by retraction of the liver upwards, with traction anteriorly and to the right on the neck of the gallbladder with a suitable forceps, while a damp pack held by the assistant retracts the colon and the duodenum inferiorly. Occasionally it is helpful to bring the whole liver down into the wound with a pack placed over the dome of the liver. If the gallbladder is tense and difficult to grasp, the operation may be made easier by aspiration of the contents.

The peritoneum over the neck of the gallbladder is incised in front and behind and the contents of Calot's triangle displayed by a combination of blunt and sharp dissection (Fig. 42). Once the cystic duct and artery have been definitely identified, the cystic artery is ligated in continuity and divided between ligatures. If operative cholangiography is planned, then a sufficient length of the cystic duct is exposed for easy cannulation. Any stones in the cystic duct are milked back into the gallbladder and the cystic duct is ligated close to the gallbladder. If cholangiography is to be performed it should be at this stage. When satisfactory pictures have been obtained the cannula is removed and the cystic duct ligated or oversewn with an absorbable suture. The dissection of the gallbladder from the liver can begin either at the fundus or in the region of the cystic duct. Either way it is important to keep close to the gallbladder wall and diathermy will be needed to achieve haemostasis. A vacuum drain can be left in the gallbladder bed if appropriate.

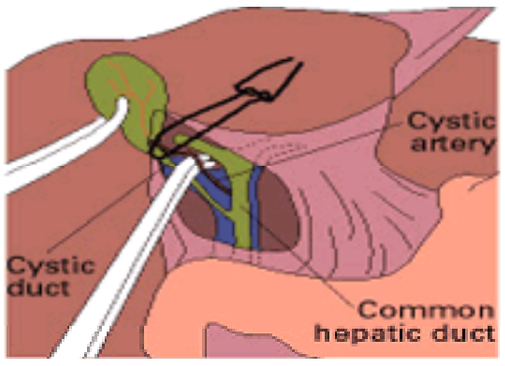


Fig. 42. Dissecting Calot's triangle as the first step in a cholecystectomy. The borders of Calot's triangle are the cystic duct laterally, the common hepatic medially, and the inferior margin of the liver superiorly.

Peroperative cholangiography An operative cholangiogram is most often performed to demonstrate the anatomy of the biliary tree. Sometimes it is to demonstrate the presence of stones in the ducts and it is invaluable when there is a leak of bile from an unidentified duct. From a practical viewpoint, good cholangiograms are only obtained with regular practice and attention to detail. A suitable cannula is tied into the cystic duct or a fine butterfly needle is introduced directly into the bile duct itself. After the careful elimination of any air bubbles from the syringe and cannula, dilute contrast material (25 per cent sodium diatrizoate) is then injected. Ideally this should be done during screening of the patient, but more commonly two radiographs are taken, the first after 3 to 5 ml and the second after a further 8 to 12 ml of contrast have been injected. On the first film the bile duct itself will be seen; the second film shows the intrahepatic biliary radicles and contrast in the duodenum (Fig. 43).



Fig. 43. An operative cholangiogram: note the cannula inserted into the cystic-duct stump (same patient as Fig. 22); the three large stones shown in the macroscopic specimen are clearly seen on the radiograph.

Filling defects in the contrast or the absence of flow into the duodenum are positive evidence of stones in the bile duct. Some false-positive results are obtained, and stones are actually removed from only two-thirds of ducts explored as a result of an abnormal peroperative cholangiogram.

Complications

Sudden arterial haemorrhage during a cholecystectomy usually arises from a torn cystic artery. The bleeding point should not be clipped; instead the wound should be packed tightly with a small swab. It is then essential to wait while haemostasis develops. During this time the surgeon can ensure that the exposure and the illumination are optimal. The wound is then kept dry with a strong sucker while the bleeding point is sutured with a fine 4/0 or 5/0 prolene stitch. Very occasionally the only way of controlling haemorrhage while the damage is being repaired is by occluding the hepatic artery with the fingers and thumb of the left hand placed across the entrance of the lesser sac (Pringle's manoeuvre).

Damage to the bile duct occurs perhaps once in every five hundred open cholecystectomies (Fig. 44). Damage is avoided by following every rule of the operation without fail on every occasion. As in a laparoscopic cholecystectomy the most important thing is to recognize the injury and to repair it there and then, either oneself or by asking a specialist colleague to help.



Fig. 44. Operative cholangiogram showing complete division of the bile duct with no contrast filling the intrahepatic ducts. The cannula is tied into the proximal common bile duct.

The mortality of open cholecystectomy is 1 per cent and the morbidity about 5 per cent. Pulmonary complications are the most common; wound infection, deep-vein thrombosis, and cardiovascular problems account for the rest. The overall figure for mortality conceals a considerable variation with age. Cardiac and respiratory disease are more frequent in elderly people, in whom it is more common to find complications from the stones themselves. The mortality rate in patients over the age of 70 years may reach 10 per cent.

Postoperative care

Opiates are required for pain relief after an open cholecystectomy. A bolus of the drug is best delivered intravenously from a syringe pump under the control of the patient (patient-controlled analgesia, PCA). Most patients will drink and move out of bed the day after the operation. A drain can be removed after 24 h unless there is bile present. If there is, then it should not be removed until the drainage has stopped. A prolonged ileus is uncommon and most patients will eat on the second postoperative day. Many patients will be ready to go home 3 or 4 days after operation. Most will need 6 to 8 weeks off work.

'Fundus first' cholecystectomy

Some surgeons feel that dissecting Calot's triangle first takes the surgeon too close to structures that should be avoided. They prefer to start the operation at the fundus and take the gallbladder off the liver first. By keeping immediately adjacent to the gallbladder wall the cystic artery and cystic duct are eventually exposed and can be tied well away from other important structures, which are often never seen. Operative cholangiography is performed towards the end of the operation if necessary. Bleeding from the gallbladder bed can sometimes be a nuisance and obscure the view of Calot's triangle but this approach may be easier with an acutely

inflamed gallbladder. It should certainly be adopted if the initial dissection of Calot's triangle in the conventional operation proves to be difficult.

Minicholecystectomy

Much of the morbidity from a conventional cholecystectomy arises from the incision of the abdominal wall needed to provide sufficient exposure. However, with modern imaging there is little need for either a laparotomy or an operative cholangiogram, which are the main reasons for a large wound. Minicholecystectomy is performed via a subcostal incision no more than 5 cm long placed right over the gallbladder, which is then dissected out fundus first. Metal clips are placed to occlude the cystic duct and the cystic artery. The incision is closed without drainage. There is little postoperative pain or systemic upset and patients can be discharged 2 or 3 days after their surgery. Controlled trials have shown that the results of minicholecystectomy are as good as those of the laparoscopic operation.

Acute cholecystectomy

This is usually an open operation from the outset, either because the diagnosis is only made at laparotomy or because it has been immediately obvious on laparoscopy that conversion to an open operation would be wise. It is a difficult operation that should only be undertaken by an experienced surgeon. The operation is always done fundus first and mostly with the fingers. It is easy in the presence of acute inflammation to enter the plane between the liver and the gallbladder, and with careful finger dissection to end up with the gallbladder on the pedicle of the cystic artery and the cystic duct. These two structures should then be tied separately with a thick ligature that does not cut through the friable tissues. Haemorrhage from the gallbladder bed, which can be rather brisk, is controlled with pressure from a pack until haemostasis supervenes. A drain should be left in the gallbladder bed. At any suggestion of difficulty or danger the operation should be converted to a cholecystostomy. This will always be the safer operation for the less experienced surgeon.

Cholecystostomy

Most patients with acute cholecystitis respond to conservative measures. A few do not and then go on to develop an empyema of the gallbladder that requires drainage. Ultrasound-guided percutaneous drainage with a pigtail catheter is the procedure of choice (see Fig. 25). Surgical drainage with a large tube drain after removing all the stones is rarely required. It is best to allow the inflammation to settle whichever method of drainage is employed and to remove the gallbladder sometime later when a laparoscopic operation is often possible.

Partial cholecystectomy

Sometimes during a cholecystectomy it becomes obvious that it is too dangerous to remove all the gallbladder. It is then wisest to excise as much of the gallbladder as possible and to remove any remaining stones from the lumen. The gallbladder lumen is closed with a suture and a drain is left in the wound. If necessary a further operation to remove the residual gallbladder tissue can be undertaken, usually with some difficulty, sometime later when the acute inflammation has subsided.

The postcholecystectomy syndrome

Severe biliary pain and nausea are effectively eliminated by cholecystectomy. Dyspepsia and symptoms associated with oesophageal reflux often persist. Up to half the patients who have had their gallbladder removed will admit to persistent symptoms 1 year later. The majority of these patients have only mild complaints, often do not seek medical advice, and find that the symptoms are associated with certain foodstuffs. Fatty foods are the most common culprit but there is a wide variation and patients soon learn to avoid those foods that upset them.

Similar but severe symptoms occur in 1 to 5 per cent of patients. They are more common in middle-aged patients with a long preoperative history, which may not have been typical of gallstone disease, and those who had a normal gallbladder removed. There are various possible explanations. In about one-third of patients a stone in the bile duct is found. In a further third another cause is found, such as pancreatic or liver disease, peptic ulceration or the irritable bowel syndrome.

The irritable bowel syndrome, in particular, is often associated with disease of the biliary tract and the symptoms can be hard to separate. Removing the gallbladder may remove some component of the complaints but leave those that are attributable to the bowel. If this is recognized before surgery it is wise to warn the patient.

Some patients appear to have disordered function of the bile duct and ampullary sphincter (biliary dyskinesia). This is often manifest at ERCP when the patient obviously experiences pain as the bile duct is filled with contrast. Physiological studies of the ampulla can identify these patients and then destruction of the ampullary sphincter is effective. In other patients the results of endoscopic sphincterotomy for postcholecystectomy symptoms are poor.

Some of these patients, particularly those in whom the gallbladder was normal, have often had other organs such as the appendix and the uterus removed. They may show symptoms of anxiety or depression and they tend to focus attention on relatively mild symptoms. It is important to establish that there is no objective organic disease in the biliary tract so far as possible and then to offer these patients counselling for the underlying problem.

Stones in the bile duct

Stones in the bile duct may lie dormant for many years and only come to light because of an episode of pain, jaundice, or cholangitis (see Fig. 9). They may also be discovered by ultrasonography at the same time as stones are seen in the gallbladder (see Fig. 6). About 10 per cent of patients with stones in the gallbladder also have stones in the bile ducts (choledocholithiasis). The incidence increases with age: one-quarter of patients over 60 years of age have stones in both sites. Most stones in the West are found in the common bile duct, whereas in the East hepatic duct stones are more usual.

Origin

Most common-duct stones originate in the gallbladder and migrate through the cystic duct into the bile duct. These secondary stones consist mostly of cholesterol. They often grow in size within the bile ducts.

Primary bile-duct stones form within the bile duct. They are usually bilirubinate stones of the soft brown type and they are associated with biliary stasis due to obstruction, infection, and the presence of foreign bodies such as food. Extra excretion of bilirubin as a result of haemolysis, from spherocytosis or thalassaemia for example, predisposes to this type of stone. In the Orient they are caused by infection sometimes associated with parasites within the biliary tract.

Clinical presentation

Although stones in the bile duct may be silent, the development of symptoms is potentially serious; obstructive jaundice, ascending cholangitis, and acute pancreatitis are all associated with major illness and death.

Less seriously, stones in the ducts can cause bouts of abdominal pain and dyspepsia indistinguishable from gallbladder symptoms, or of intermittent biliary colic with transient jaundice. Elderly patients with bile-duct stones sometimes present in apparently obscure ways with malaise, confusion, collapse, or septicaemia (Fig. 45). The cause is often only discovered when the routine liver function tests are found to be abnormal. Silent stones in the bile ducts are also discovered during investigation before a cholecystectomy. Ultrasonographic examination of the gallbladder is incomplete without an examination of the bile duct.

Fig. 45. Endoscopic cholangiogram of a patient admitted collapsed with septicaemia; *E. coli* was grown from a blood culture.

Obstructive jaundice

Occasionally, a small stone passes into the bile duct and impacts at the ampulla, causing pain and jaundice. The severity of the jaundice depends on the duration of the obstruction, but as the stone passes on the jaundice spontaneously resolves. A solitary stone may disappear from the biliary tree in this way, but normally some stones remain in a thick-walled gallbladder to support the diagnosis. Such patients need a cholecystectomy and a preoperative cholangiogram is essential to confirm that no stones remain in the duct.

More commonly there is a larger stone or stones within a dilated bile duct. Sometimes the number of stones in the duct leads to a significant impairment of bile flow. At other times a stone moves up and down within the duct and acts as a ball valve, causing pain and jaundice when it impacts but allowing the symptoms to resolve spontaneously when it moves away. The site of impaction is usually immediately above the ampulla, but it may be above a fibrotic narrowing in the bile duct caused by the stones themselves. Complete impaction of a stone causes severe, progressive jaundice.

Stones in the bile duct usually cause pain. However, from the pain alone it is not easy to distinguish obstructive jaundice due to stones from that due to malignant disease. Clinical examination normally reveals nothing except a jaundiced patient, and possibly some scratch marks from the intolerable itching. The gallbladder is not palpable since it is thick-walled and fibrotic, and it resists distension, although there is often mild tenderness in the right upper quadrant.

Many of these patients are elderly and require a prompt endoscopic sphincterotomy and extraction of their stones. If the stones cannot be removed, then a stent should be placed in the bile duct, which will relieve the jaundice. In some very elderly patients, particularly if the gallbladder has already been removed, a stent is effective and permanent treatment. Cholecystectomy can be performed later when the jaundice has resolved. In practice only 10 per cent of elderly patients have continuing symptoms and need surgery. Patients under the age of 50 years who are not profoundly jaundiced are best treated surgically.

Ascending cholangitis

Ascending cholangitis is still a fatal disease and it must be treated as a medical emergency. Fortunately, it is usually an easy diagnosis to make clinically, as most patients present with the classic symptoms of epigastric pain, rigors, and jaundice (Charcot's triad or Charcot's intermittent biliary fever). Elderly patients sometimes present simply with septicaemia or collapse, with little or no jaundice, and rarely the origin of a Gram-negative septicaemia is eventually traced back to the bile duct.

Pathology

Cholangitis is always associated with some degree of obstruction within the bile duct: stones in the ducts are the cause in 80 per cent of cases. Cholangitis is a rare presentation of malignant biliary obstruction, except in those with carcinoma of the ampulla. Patients with a benign biliary stricture commonly experience recurrent episodes of cholangitis and they always have bacteria in their bile, as do patients with an endoluminal prosthesis in place. Patients with stones nearly always have a positive bile culture whereas this is found in only 10 per cent of patients with malignant jaundice.

Bacteriology

Most of the bacteria cultured from the bile in patients with cholangitis are also found in the bowel. *Escherichia coli, S. faecalis* and *Klebsiella* spp. are the most common pathogens, but *Staphylococcus, Pseudomonas* and *Proteus* may occasionally be present. Anaerobic bacteria such as *Clostridium perfringens* and *Bacteroides fragilis*, although rarely cultured from gallbladder bile, are an important feature in cholangitis. Bacteria reach the liver in the portal vein and are normally cleared there by the reticuloendothelial system. There is also evidence of cholangiovenous reflux of organisms into the circulation when the systemic symptoms of cholangitis become apparent. More than one organism is present in over half of all patients, and there is some evidence of synergy between the aerobic and anaerobic organisms. Antibiotic treatment, which should always be vigorous, must take account of the polymicrobial nature of most infections.

Treatment

The obstructed bile duct must be drained adequately, by the most effective route, and as quickly as possible. However, the patient must first be resuscitated with intravenous fluids and antibiotics. Antibiotic treatment of septicaemia will produce improvement for a short period, but it will not cure unless the obstruction is relieved. Nowadays this can usually be achieved by an endoscopic sphincterotomy (Fig. 46). If this is not possible then percutaneous transhepatic drainage should be tried. Surgery to drain the bile duct is a last resort.

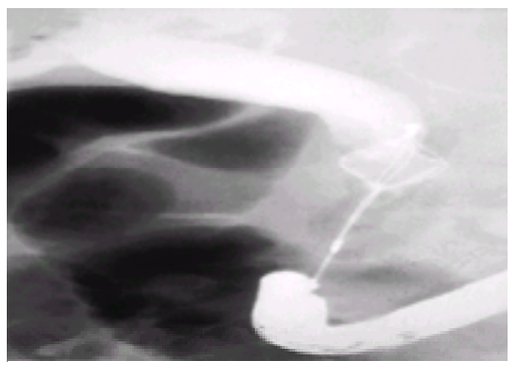


Fig. 46. Stone being extracted from the bile duct with a Dormier basket. This patient had also had a Polya gastrectomy some years previously. The endoscope approaches the ampulla along the afferent loop. This is always difficult and is only achieved in less than half the patients.

Complications

The septic process can develop in two different ways. Sometimes intrahepatic abscesses appear. They can be seen on a cholangiogram as cavities in association with intrahepatic ducts. Ultrasonographic or CT diagnosis is initially more difficult as there is little distinction between inflamed liver tissue and an abscess. Repeated examinations are required. Hepatic abscesses may rupture through the hepatic capsule and give rise to intraperitoneal perihepatic collections. The liver is often swollen and there is tension within the biliary tree. There is often a gush of purulent bile into the duodenum when the obstructing stone is released endoscopically.

Alternatively the sepsis may become systemic. Progressive renal and cardiac impairment ensues, and patients develop septic shock. Dialysis or haemofiltration may well be required. Occasionally the presenting feature of cholangitis is complete renal failure or cardiovascular collapse; such patients frequently die.

Acute pancreatitis

Acute pancreatitis is associated with gallstones (see Chapter 33.1). Impaction of a small stone at the ampulla and occlusion of the pancreatic duct is a cause of acute pancreatitis. An early ultrasonographic examination of the biliary tract is therefore essential in every patient who is admitted with acute pancreatitis, particularly if there is any change in the liver function tests. A few have evidence of a stone in the bile duct and an immediate endoscopic sphincterotomy and extraction of the stone is then well worthwhile as it may abort the episode of pancreatitis. There is no evidence that the pancreatitis is made worse by ERCP, although it is wise to avoid injecting contrast into the pancreatic duct.

Mirizzi syndrome

This is an unusual and specific cause of obstruction of the common hepatic duct by a stone impacted in Hartmann's pouch. The stone may simply press on the bile duct, but more commonly it ulcerates into the duct, creating a cholecystocholedochal fistula. Patients present with obstructive jaundice and on cholangiography there is a narrowing of the bile duct at the porta hepatis that can have the appearance of a cholangiocarcinoma (Fig. 47, Fig. 48). A variant of the syndrome occurs when a large stone in the cystic duct becomes impacted just above the junction with the bile duct and obstructs the common hepatic duct. This can also be confusing on cholangiography until it is appreciated that the obstructing stone lies in the cystic duct and not the bile duct. Surgery to remove the gallbladder and the gallstone is often extremely difficult because of severe inflammation and fibrosis. A gap in the bile-duct wall usually remains when the obstructing gallstone is removed. A small hole can be closed over a T-tube. A larger gap requires some form of reconstruction and part of the wall of the gallbladder or cystic duct is often available. For this reason it is wise to be circumspect about removing both these structures until the state of the bile duct is clear. If a T-tube is used for the repair, it must be brought out through a separate incision in the bile duct and not through the damaged area. Alternatively, a biliary enteric bypass may be the only safe way to drain the hepatic ducts.

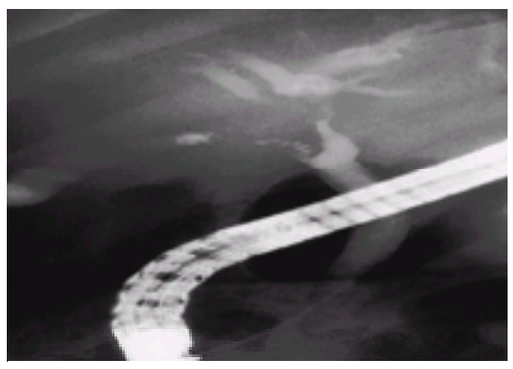


Fig. 47. See also Fig. 48. Two cholangiograms of the same patient taken 6 weeks apart. The first picture was diagnosed as showing a typical cholangiocarcinoma and an attempt was made to place a stent across the tumour. Six weeks later at a second ERCP a large stone is clearly seen in the common hepatic duct. It was easily removed at surgery. This is a classic example of Mirizzi syndrome and an initial erroneous diagnosis of malignant biliary obstruction is not uncommon.

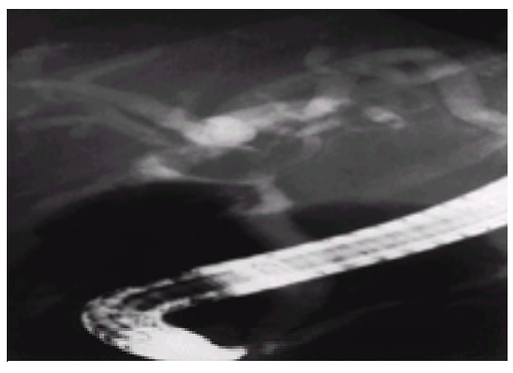


Fig. 48. See also Fig. 47. Two cholangiograms of the same patient taken 6 weeks apart. The first picture was diagnosed as showing a typical cholangiocarcinoma and an attempt was made to place a stent across the tumour. Six weeks later at a second ERCP a large stone is clearly seen in the common hepatic duct. It was easily removed at surgery. This is a classic example of Mirizzi syndrome and an initial erroneous diagnosis of malignant biliary obstruction is not uncommon.

Investigation

The most important investigation is ultrasonography of the liver, the bile duct, the gallbladder, and the pancreas. It should be undertaken on the least suspicion that there may be stones or another obstructive lesion in the bile duct. The ultrasonographer need only decide whether or not the bile ducts are dilated. The normal common bile duct should not be greater than 7 mm in diameter when measured on ultrasound in a starved patient (see Fig. 6). If the ducts are dilated, the patient has an extrahepatic obstructive cause for their symptoms. If the ducts are not dilated it is unlikely that there are stones in the bile duct, but there are important exceptions to this rule. If the examination is done very soon after a stone has entered the bile duct there may have been insufficient time for dilatation to have developed. The examination should be repeated 1 week later. In patients with cirrhosis of the liver the intrahepatic bile ducts are simply not able to dilate. If there is clinical uncertainty about the presence of a stone within the ducts, a cholangiogram is needed.

An experienced ultrasonographer can always detect dilatation of the ducts, but the site of the obstruction will only be identified in two-thirds of patients and the cause of the obstruction in one-third. Nevertheless, stones and strictures can sometimes be identified on ultrasonograms.

Before the introduction of ultrasonography, biochemical markers of liver function were important in differentiating surgical from medical jaundice. Their specificity and sensitivity were very poor and they are now only of historical interest. The main value of biochemistry nowadays is to quantify the severity and the duration of an obstruction and to monitor the effects of treatment.

CT has a limited place in the imaging of common-duct stones. The ultrasonographic examination may raise the possibility of a malignant obstruction, and a CT scan may be obtained before ERCP. CT detects dilatation of the ducts very reliably (see Fig. 10) and it is better than ultrasound at identifying the site and the cause of an obstruction. Magnetic resonance cholangiography (MRC) is likely to be useful in the future.

The prothrombin time is a marker of liver function as it is dependent on the synthesis of coagulation proteins in the liver. It should always be measured, even if the patient is not jaundiced. Patients with a prolonged prothrombin time should receive vitamin K and may also require fresh frozen plasma to correct a coagulation defect before embarking on an endoscopic sphincterotomy or surgery. Any patient who has a trace of jaundice and a fever must have blood cultures taken before treatment with antibiotics. This may be the only opportunity to identify the organism.

It can still be very difficult to differentiate surgical from medical causes of jaundice. Hepatitis and jaundice due to drugs occasionally develop in patients who also have stones in the ducts. It is important to take a drug history and as soon as hepatitis is suspected the immunological markers must be measured, and the laboratory must be warned of the risk of infection.

Management

When stones are suspected or discovered in the bile duct there should be little delay in confirming the diagnosis and treating them either surgically or at an ERCP. Percutaneous cholangiography is occasionally helpful. A detailed description of diagnostic and therapeutic ERCP is given in Chapter 14.2. The most appropriate treatment depends on the age of the patient, whether the patient has undergone previous cholecystectomy, and the skills available.

Stones in the duct and the gallbladder present

Surgical removal is the most appropriate treatment for young or middle-aged patients without serious coexisting disease and with uncomplicated ductal stones. The morbidity and mortality associated with surgical exploration of the bile duct and endoscopic sphincterotomy are very similar in this group of patients, and most of them need their gallbladder removed. Preoperative endoscopic clearance of the ducts may reduce the total time in hospital, but it has no other advantage and the risk of long-term complications from sphincterotomy is unknown.

Patients over the age of 60 years, poor-risk patients, and those with complicated ductal stones are best treated endoscopically. The morbidity and mortality of

endoscopic treatment is much less in this group of patients and only 1 in 10 elderly patients needs subsequent cholecystectomy. Patients between the ages of 50 and 60 years have to be treated on their individual merits.

Stones in the duct and the gallbladder absent

Patients with retained or recurrent stones following cholecystectomy should be treated endoscopically in the first instance, whatever their age. Retained stones are those detected soon after a choledochotomy. Recurrent stones come to light months or years later.

Retained stones, particularly if they have been left behind deliberately, require an ERCP as soon as convenient. They are occasionally difficult to remove, particularly if the ducts are small and the stones lie above a T-tube. Some stones can only be removed once the T-tube has itself been removed, and time must pass to allow the track to mature before removal of the T-tube is safe. Retained stones may pass spontaneously, and they can occasionally be flushed through the ampulla with saline, dissolved with mono-octonoin or methyltertbutylether (MTBE), or shattered by extracorporeal lithotripsy; none of these methods is totally reliable. Extracting retained stones along the mature T-tube track under radiological control (Burhenne technique) is very effective in skilled hands, but the patient has to keep the tube and its attendant bag for 6 weeks while the track matures. None of the methods is perfect, and a combination of a percutaneous and an endoscopic technique may be needed.

Recurrent stones are often multiple and they may be large. In most patients a generous endoscopic sphincterotomy will allow removal of the stones with a basket or balloon (see Fig. 46). Large stones can sometimes be fragmented with a crushing basket or lithotripsy and then removed in pieces. Elderly patients with stones that are too large to remove endoscopically can be treated by placing a stent in the bile duct. The stent aids drainage of bile, prevents impaction of the stones, and is an effective alternative treatment to open operation. The stents last for many months, even years, and rarely need to be changed. The overall success in clearing the bile duct endoscopically is about 80 per cent, stents may treat a further 5 to 10 per cent, and the remainder should be considered for an operation.

Stones in the duct and open cholecystectomy

The majority of these stones should be discovered before operation but some will be found on peroperative cholangiography. It is normally appropriate to remove all the stones from the duct at one operation, but there are some exceptions. Small stones in small ducts, particularly if they are impacted at the ampulla, may be best left alone. The small stones in the duct may pass spontaneously whilst the impacted ones should be extracted endoscopically after leaving a T-tube in the duct for safety. Similarly, in poor-risk patients it might be appropriate simply to close the abdomen with a T-tube in the bile duct and arrange a prompt endoscopic extraction. Stones can be removed surgically, either from above (the supraduodenal approach) or from below (the transduodenal approach). Occasionally both approaches are needed together. The transduodenal operation has been largely superseded by the endoscopic approach, which is safer and easier.

Supraduodenal exploration of the common bile duct

The common bile duct is exposed above the duodenum and it is helpful to mobilize the second part of the duodenum (Kocher's manoeuvre). The site of the choledochotomy should be as proximal as possible to allow choledochoduodenostomy if required and to leave the greatest length of bile duct above the choledochotomy should it be necessary to repair a postoperative bile-duct stricture.

Two stay sutures are placed on either side of the proposed choledochotomy and the duct opened longitudinally along the anterior wall (Fig. 49). A sample of bile is sent for culture. After the careful removal of any obvious stones, the ducts are gently flushed with an umbilical catheter. Choledochoscopy with either a rigid or a flexible choledochoscope is then performed, first upwards into the intrahepatic ducts (Fig. 50) and then downwards to the ampulla. The saline infusion not only provides a view (Fig. 51) but also washes stones and debris out of the ducts. It is easier to see the intrahepatic ducts and much more difficult to be sure that the retroduodenal portion of the bile duct is clear of stones. Desjardin's forceps, a Fogarty balloon catheter, and a Dormia basket can be helpful in extracting stones. Metal bougies should never be used in an attempt to dilate the ampulla. They rarely dilate the ampulla but usually create a fistula into the duodenum immediately above the ampullary opening.

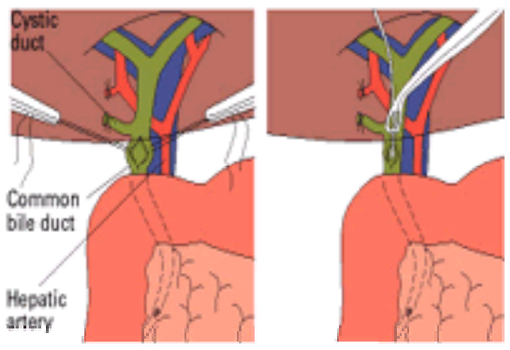


Fig. 49. The approach to supraduodenal exploration of the bile duct.

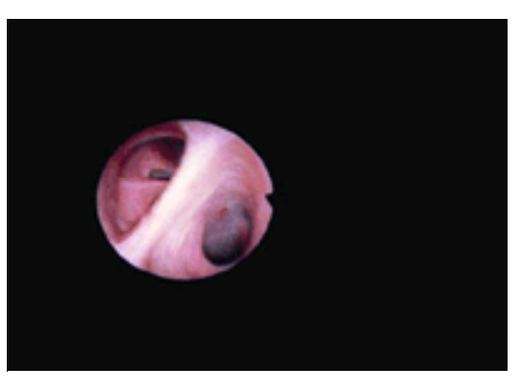


Fig. 50. The junction of the right and left hepatic ducts seen from within the bile duct at choledochoscopy. Second-order bile ducts are seen beyond. This picture and the next were taken through a flexible choledochoscope.

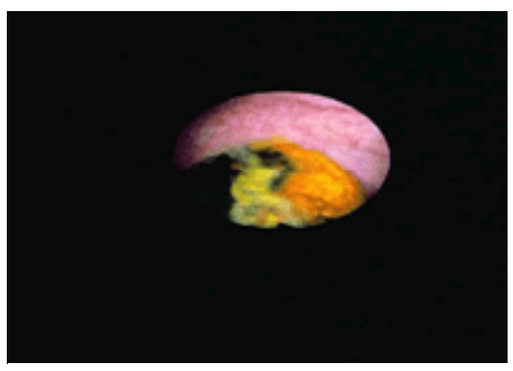


Fig. 51. A stone in the retroduodenal portion of the common bile duct seen at choledochoscopy.

Some surgeons like to repeat the cholangiogram to confirm that the duct has been cleared. However, it is difficult to obtain satisfactory pictures and the failure to

remove all the stones at an exploration does not matter. Indeed it is sometimes safer and wiser to leave a difficult stone behind for later retrieval than to cause damage to the bile duct by prolonged attempts at its removal.

Once all the stones and debris have been removed a 16-Fg guttered latex T-tube is placed in the common bile duct. The free end is brought out laterally through the abdominal wall. This position keeps the radiologist's hands away from the X-ray beam during cholangiography, and the size of tube will allow percutaneous extraction of a retained stone if this becomes necessary. The choledochotomy is then closed with a fine absorbable suture and a drain is placed in the subhepatic space before closing the wound.

Postoperatively the T-tube is allowed to drain freely into a sterile, closed drainage bag. A T-tube cholangiogram is obtained on the ninth or tenth day after surgery (Fig. 52) and, if the duct is clear, the T-tube is clamped. Provided the patient does not develop any pain or discomfort the tube can be removed 24 h later. Drainage from the T-tube site usually ceases within 48 h.

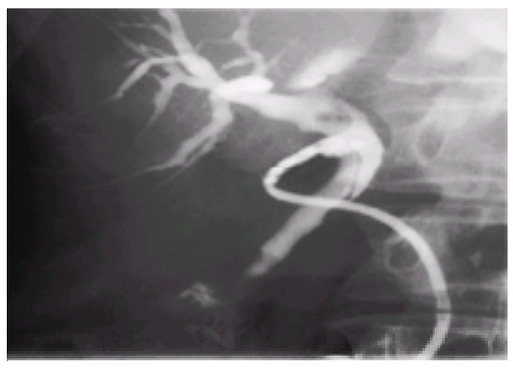


Fig. 52. T-tube cholangiogram showing a retained stone. The stone lies alongside the upper arm of the T-tube; it was removed at ERCP.

Transduodenal exploration of the bile duct

In this operation the bile duct is approached across the duodenum and through the ampulla. It is usually combined with a sphincteroplasty. The duodenum is fully mobilized and a longitudinal incision is made in the right lateral wall over the ampulla. A probe is then passed into the bile duct and the ampullary sphincter is divided with scissors. Fine absorbable sutures are placed to appose the mucosa of the bile duct to the duodenum and the stones are then extracted with Desjardin forceps. The choledochoscope can be used to ensure that all the stones have been removed and the duodenum is then closed. The advantage of this approach is that any missed stone will pass spontaneously. The disadvantage is the risk of pancreatitis from interference with the ampulla. Most patients who need a transampullary approach to their bile duct are better treated endoscopically.

Choledochoduodenostomy

Occasionally, an alternative to closure of the common bile duct over a T-tube after a supraduodenal exploration is a choledochoduodenostomy. Provided the bile duct is more than 15 mm in diameter the operation is quick and easy to perform, and there are no worries about retained stones. The vertical incision in the common bile duct is sutured to a longitudinal incision in the duodenum with a single layer of stitches. Results in elderly patients are satisfactory, but in patients who have had the anastomosis for a number of years recurrent cholangitis may develop. This is known as the ‘sump syndrome': infection arises from stones and vegetable matter that collect in the retroduodenal portion of the bile duct between the anastomosis and the ampulla. There may also be stenosis of the choledochoduodenostomy. Endoscopic sphincterotomy of the ampulla and balloon dilatation of the anastomosis may alleviate the symptoms but treatment is not very satisfactory.

Biliary peritonitis

Bile peritonitis is uncommon. Perforation of an acutely inflamed gallbladder, removal of a T-tube before a proper track has formed, and percutaneous transhepatic cholangiography are three possible causes. Clinically it is often difficult to diagnose because uninfected bile is not particularly irritant and the signs may be very subdued.

Surgery, either laparoscopic or open, is usually required to remove a perforated gallbladder. The T-tube will have to be replaced at an open operation if generalized peritonitis develops after its removal. Leakage of bile into the peritoneum after a percutaneous cholangiogram is uncommon because it is usual to leave a drain in the bile duct if it is not possible to relieve the obstruction. Any local peritonitis can usually be treated conservatively, although one can expect the leak to persist if there is persistent obstruction and no drain has been left in the bile duct. Local collections of bile that persist may be amenable to percutaneous drainage.

Surgical damage to the bile ducts

Damage to the bile duct is, as outlined above, a feared complication of cholecystectomy. If a stricture develops it can lead to recurrent cholangitis, chronic ill health and, unless corrected, ultimately to liver failure. Over the 100 years following the first cholecystectomy in 1882 the rules for the safe performance of the operation were well established. By the time the laparoscopic operation was introduced in 1987 the incidence of damage to the ducts had fallen to a very low level. Unfortunately, the limited view, the unusual manual skills that are needed, the lack of tactile sensation, and the difficulties of orientation and judgement of depth on a two-dimensional image have led to a small epidemic of damaged ducts as a consequence of a laparoscopic cholecystectomy. The lessons of a previous era have had to be relearnt and many damaged ducts have had to be repaired.

Aetiology and prevention

Poor surgical technique is the most common cause of a significant biliary injury and there is no difference, in this respect, between the laparoscopic or the open operation. The precise individual anatomy has not been correctly identified, although various anatomical and pathological factors may have made this difficult. The ducts are often narrow and the surgeon thinks they are too narrow to be the bile duct. The cystic duct may run alongside the bile duct for a distance, which leads the surgeon into the wrong plane. Anatomical variations of the main ducts also predispose to damage. The cystic duct may enter the right hepatic duct or sometimes there is no right duct and the right anterior hepatic duct runs very close to the cystic duct. During the operation, excessive fibrosis and inflammation in the porta hepatis and sudden inadvertent haemorrhage are both known to be dangerous and to put the bile ducts at risk.

There are three types of injury to the bile duct that occur during a laparoscopic cholecystectomy. The common mistake, as with the open operation, is to think the common bile duct is the cystic duct in an ‘easy' cholecystectomy. The ‘duct' is tied and divided, and a variable length of the common hepatic duct is then excised (see Fig. 44). Less common is a failure to identify correctly the structures around the junction of the cystic duct and bile duct because of severe inflammation and fibrosis. Least common are diathermy injuries. They usually involve the common hepatic duct and occur because the proximity of the duct to the diathermy probe, and the ability of heat to travel along metal clips, are not appreciated.

Inadequate exposure is the cause of most injuries during open cholecystectomy. An adequate and correctly placed initial incision is essential. Excessive traction is to be avoided and it is not necessary to trace the cystic duct right to the junction with the bile duct. Minicholecystectomy is deliberately performed through the smallest possible incision. Exposure is therefore minimal and dissection must stay immediately adjacent to the gallbladder wall until the cystic duct is reached. Correctly performed the operation is safe, but there is no margin for error.

If it proves difficult to identify the anatomical features during a laparoscopic operation it is safest to convert to an open procedure at once. A cholangiogram should be obtained as soon as possible, and this should also be the first move when difficulty is encountered during an open cholecystectomy. On the cholangiogram, it is important to check that the whole biliary system fills with contrast. When a section of the bile duct is removed it is usually the intrahepatic ducts that fail to fill (Fig. 44).

Diagnosis

In about one-quarter of patients the injury is recognized at the time of operation and in a further third it comes to light within the next month. Most of that one-third

present with jaundice, sometimes with cholangitis and sometimes with a biliary fistula. The remainder present months or years later with recurrent cholangitis. In the early postoperative period, ERCP is the most useful imaging technique for displaying the extent of the damage; this may provide an opportunity to place a stent if this is appropriate. Contrast medium injected along the track of a fistula may define the injury and the bile ducts adequately. In patients who present later, both ERCP and percutaneous transhepatic cholangiography may be needed to display the superior and the inferior aspects of the stricture. It may also be possible to relieve the obstruction by placing a self-expanding metal stent across the stricture. These patients with long-standing incomplete obstruction and infection have a significant risk of liver damage and portal hypertension. The presence and the severity of these complications require investigations such as a liver biopsy and oesophagoscopy for varices.

Treatment

Many surgeons will have realized with horror during a cholecystectomy that they have just clipped or tied the bile duct. The clip or the tie should be removed and nothing further needs to be done. Strictures do not develop afterwards.

Immediate repair of a damaged duct

A serious injury may take one of three forms. First, there may be a short incision into the bile duct, perhaps with a ragged edge if tearing was a feature of the injury. This is commonly caused by inferior traction on the cystic duct at the junction with the bile duct. Second, the bile duct may be divided clean across, perhaps obliquely, but there is no loss of length. Third, a portion of the duct may be excised.

P>Most surgeons are able to deal with the first two injuries. In the first, the incision can be closed directly with fine absorbable sutures, over a T-tube if necessary. In the second case the bile duct should be repaired with interrupted, fine absorbable sutures over a T-tube (Fig. 53). A T-tube should never be brought out through a damaged area of duct because the irritation, inflammation, and fibrosis are likely to increase the risk of a subsequent stricture. It must always be placed through a separate stab incision above or below the repair.

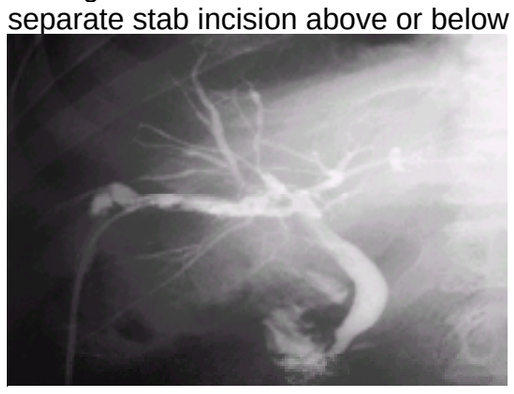


Fig. 53. Immediate end-to-end repair of a divided bile duct over a T-tube. The bile duct had been cut straight across and there was no loss of length. There is some leakage of contrast along the T-tube track but this was not clinically apparent. This is the same case as in Fig. 44.

The best results of treatment in the third type of injury are obtained if an experienced biliary surgeon is called to help. In most cases the common hepatic duct has been completely removed and only a tiny stump or even separate right and left hepatic ducts remain. This injury requires repair with an hepaticojejunostomy to a Roux loop of jejunum. The anastomosis should be supported with a tube or tubes that remain in place for at least 3 months (Fig. 54). If the damage to the duct is very proximal a choledochoduodenostomy may occasionally be safe, but there is always the risk of excessive tension on the anastomosis and a subsequent stricture. Partial excision of one wall of the duct is the most dangerous injury of all. It is tempting to mobilize the duct above and below and to suture the defect transversely over a T-tube. Experience is required to judge whether the degree of tension is excessive. If there is any doubt an hepaticojejunostomy is safer.



Fig. 54. Hepaticojejunostomy reconstruction of the bile duct following its complete division during a laparoscopic cholecystectomy. The anastomosis is stented with two U-tubes that go through the right and left hepatic ducts and out through the liver capsule.

Late repair of a damaged duct

Patients who present at any time after the original operation should be referred to a specialist unit. The first requirement is to control any sepsis by drainage of pus and appropriate antibiotics, and to ensure free drainage of bile so far as this is possible. It is then necessary to establish the precise nature and extent of the damage. The anatomy of all the bile ducts, both intra- and extrahepatic, must be defined, and the presence of stones or debris in obstructed ducts established. An arteriogram is essential before surgical reconstruction. It is not uncommon to find concomitant damage to the hepatic artery or the portal vein. Many of these patients are malnourished and need parenteral nutrition.

The objective of treatment is to achieve long-term, unobstructed drainage of bile to the bowel. Percutaneous and endoscopic techniques can provide temporary or permanent drainage with metal self-expanding or plastic stents. Metal stents are particularly useful for the definitive treatment of high-risk patients. Balloon dilatation of a stricture is usually inadequate and of short-term benefit only. Surgery, whether it be at a first or a subsequent operation, requires a direct mucosa-to-mucosa anastomosis of the bile duct to a Roux loop of jejunum. Inadequate treatment of a stricture is dangerous. The patient's symptoms may improve but there is nevertheless a slow and progressive deterioration in liver function. Such patients should be referred for assessment for liver transplantation at an early stage.

Complications and results

The best results are obtained when a bile-duct injury is discovered immediately and when a suitable tension-free repair is performed, which heals with minimal scarring. The worst results are in patients who have had many previous repairs and who have evidence of liver failure and portal hypertension. Injuries at the porta hepatis have a worse prognosis than more proximal damage, probably because they are technically more difficult to repair.

The operative mortality is at least 5 per cent, and uncontrollable haemorrhage and renal failure are the common causes of death, often associated with infection and an external biliary fistula. Many patients experience one or more major complications. Bile-duct repairs are notorious for the formation of a recurrent stricture. In the past about 1 in 3 patients could expect further trouble; recently this has fallen to 1 in 10. Despite these difficulties, they should be offered a further attempt at operation provided their liver function is still good: previous failure does not preclude a successful outcome.

Postinflammatory stricture

Narrowing of the bile duct is often seen at ERCP in association with chronic inflammation in or around the duct, usually from bile-duct stones and sometimes from chronic pancreatitis. In patients with stones the stricture tends to be in the retroduodenal portion of the duct and the endoscopist has to be sure that the stone will come

through the narrow area if it is engaged in a basket. Failure to realize this problem is the most common cause of a trapped basket during attempted endoscopic removal of a bile-duct stone (Fig. 55). Significant and short inflammatory strictures of the duct appear to respond well to balloon dilatation, although if dilatation fails surgery is needed. Chronic pancreatitis may cause narrowing of the proximal bile duct. Jaundice developing during an acute exacerbation usually fades spontaneously. When there is evidence of long-term obstruction from severe fibrosis, an end-to-side choledochoduodenostomy is indicated.

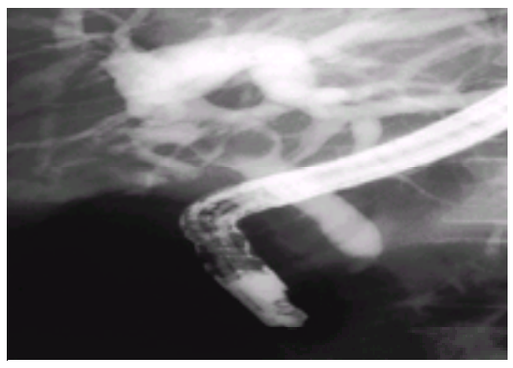


Fig. 55. The large stone in the common hepatic duct is wider than the proximal common bile duct. Unwisely the stone was trapped in a Dormier basket but it could not be retrieved through the narrow proximal common bile duct into the duodenum. The stone and the basket were both removed at an open operation performed the same day.

Ampullary stenosis

Ampullary stenosis presents with episodic pain with features that strongly suggest a biliary origin. Most of these patients will have undergone previous cholecystectomy. There are no absolute diagnostic criteria, but the combination of abnormal liver function tests, dilatation of the bile duct, delayed emptying of contrast, and difficulty in cannulating the ampulla at an ERCP are the most useful features. In specialist units, manometric studies of ampullary function and special provocation tests can help to identify such patients and indicate those who will benefit from a sphincterotomy. The precise histological changes are uncertain but in most cases there is excessive fibrosis and inflammation of both the mucosa and the smooth muscle of the ampulla. If the diagnosis is well established before operation, the results are good.

Sclerosing cholangitis

Sclerosing cholangitis is characterized by an obliterative inflammatory fibrosis of the biliary tract that leads to chronic liver disease. Sometimes the fibrosis is clearly secondary to stones in the bile duct or previous biliary surgery, but primary sclerosing cholangitis, in which these predisposing causes are absent, is a disease entity on its own. The appearance on cholangiography is diagnostic, although occasionally only time will exclude cholangiocarcinoma (Fig. 56). Primary sclerosing cholangitis was once regarded as a rare disease but the advent of ERCP has resulted in greater recognition of the condition.

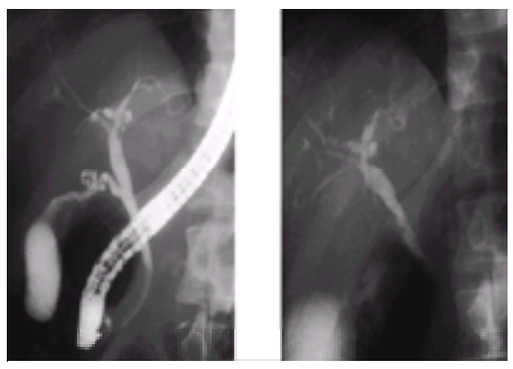


Fig. 56. Cholangiogram showing the typical appearances of primary sclerosing cholangitis with beading of both the intra- and extrahepatic ducts. The cholangiogram on the right was obtained 2 years after the one on the left and demonstrates the progression of the disease.

Aetiology

The cause of primary sclerosing cholangitis is unknown. The association with ulcerative colitis in two-thirds of patients suggests that chronic low-grade portal bacteraemia or the absorption of toxic bile acids from the diseased colon might be significant aetiological factors, but neither has much experimental support. Recently, phenotyping studies have shown a much higher frequency of HLA-B8, -DR3, -DQ2, and -DRw52A in patients with primary sclerosing cholangitis than in controls. These findings not only confirm the role of genetic factors but also suggest that the disease is immunologically mediated, as this phenotype is closely associated with a number of autoimmune diseases. Overall, current evidence suggests that primary sclerosing cholangitis is an immunologically mediated disease, perhaps triggered in genetically susceptible individuals by acquired toxic or infectious agents.

Diagnosis

Men are twice as commonly affected as are women and most patients present between the ages of 25 and 40 years. The usual symptoms are fatigue, intermittent jaundice, weight loss, upper abdominal pain, and pruritus. Attacks of acute cholangitis are surprisingly rare, unless there has been instrumental biliary intervention. Approximately half of all symptomatic patients have jaundice or hepatosplenomegaly. Many patients are discovered because of an abnormally high alkaline phosphatase on routine testing, usually during investigation for ulcerative colitis. Serum bilirubin and alkaline phosphatase are usually elevated, the latter more than the former; their concentrations also fluctuate during the course of the disease. The cholangiogram is diagnostic, with typical beading from irregular stricturing and dilatation of both the intra- and extrahepatic ducts (Fig. 56). Occasionally only the intrahepatic ducts are involved; very rarely the disease affects only the extrahepatic system. Liver histology is not often diagnostic. The early features are periductal fibrosis, portal oedema, and bile ductular proliferation. Later, fibrosis spreads into the liver parenchyma, leading ultimately to biliary cirrhosis.

Although primary sclerosing cholangitis and ulcerative colitis are closely linked, the course of each disease is apparently independent. The colitis usually extends throughout the colon but causes few symptoms. Colectomy makes no difference to the course of primary sclerosing cholangitis.

Treatment

There is no curative treatment. Trials of corticosteroids, immunosuppressants, ursodeoxycholic acid and antibiotics, either alone or in combination, have been universally disappointing. Management is directed towards minimizing symptoms and treating complications. Pruritus responds to cholestyramine; antibiotics are needed during episodes of cholangitis. Mechanical relief of a well-defined stricture is well worthwhile. In many patients the best approach is to place a stent across the obstruction, either percutaneously or endoscopically. Balloon dilatation of the strictures may also be effective.

Surgical treatment is controversial. Resection of extrahepatic strictures and reconstruction over Silastic stents produces good results in some series, but orthotopic liver transplantation is the only option available to young patients with primary sclerosing cholangitis and advanced liver disease. The 5 year survival after transplantation is approx. 85 per cent in most transplant centres. Primary sclerosing cholangitis recurs in about 20 per cent of such patients but it does not appear to affect the outcome adversely.

The average time between the onset of symptoms and death is about 12 years, and most patients die from hepatic failure. About one-quarter of patients with primary

sclerosing cholangitis eventually develop a bile-duct carcinoma, which frequently follows a very aggressive course. The results of liver transplantation in such patients is very poor.

Biliary fistula

Leakage of bile from the biliary tract can occur from the liver, the gallbladder, or the bile duct itself, and it may leak to the skin via the peritoneum or to the bowel. Some fistulas are created deliberately, such as a choledochoduodenostomy. Others develop from a pathological process, either from surgical complications, from ulceration of a stone, or from drainage of pus into an adjacent structure.

External biliary fistula

The most common external fistula develops after surgery. Even after a straightforward cholecystectomy there may be a little bile in a drain the following day. Occasionally, larger volumes of bile drain: this usually arises from the cystic duct and occurs because the clip or the tie has not been correctly placed or has come off; sometimes the leak comes from an accessory bile duct in the gallbladder bed. Provided there is no obstruction to bile flow the fistula will usually close spontaneously. If this does not happen it is necessary to place a stent in the bile duct.

A T-tube in the common bile duct is technically a fistula. Normally a T-tube cholangiogram will be performed before the T-tube is removed to confirm that there is free flow into the duodenum; the fistula closes rapidly once the T-tube is removed. Any delay in closure implies some degree of obstruction, such as a residual stone, and an ERCP is necessary.

The late development of a fistula after a cholecystectomy almost always signifies unrecognized damage to the bile duct and comes to light after the drainage of an abscess (Fig. 57). These patients are usually ill and septic. They need careful evaluation and investigation before any further surgical intervention.

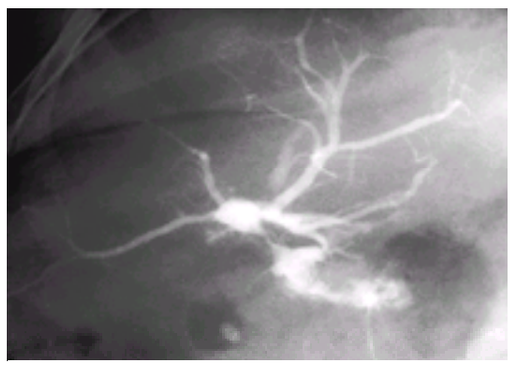


Fig. 57. External biliary fistula following a left hepatectomy for a chronic abscess in the left liver. The right posterior hepatic ductal system fills from a sinogram. The fistula was repaired with a loop of jejunum anastomosed to the right posterior hepatic duct.

Severe cholangitis occasionally leads to an intrahepatic abscess, which ruptures first into the perihepatic peritoneum. Biliary peritonitis rarely ensues because of surrounding adhesions, but when the abscess is drained externally then a fistula results. Such a fistula will only close when the proximal obstruction that caused the cholangitis is removed or relieved. This may not be possible with a malignant obstruction.

Any significant bile loss externally is accompanied by rapid fluid and electrolyte depletion, which must be vigorously replaced. If the patient will tolerate it, bile can be returned to the bowel via a nasogastric tube.

Internal biliary fistula

Spontaneous internal fistulas are uncommon and are usually discovered at cholecystectomy when a communication between the gallbladder and the duodenum becomes apparent as Hartmann's pouch is dissected away from the bowel. This usually results when a stone has ulcerated into the duodenum and disappeared in the faeces. There are no specific symptoms to suggest that this has happened, except when a large stone escapes and impacts in the terminal ileum, giving rise to gallstone ileus. Rarely, the stone ulcerates into the stomach or the colon. In the latter instance, patients have profuse diarrhoea as the bile is irritant to the colon.

The treatment is to remove the gallbladder and to close the hole in the bowel. It is quite possible to achieve this laparoscopically. It is very rarely necessary to resect bowel, but it is wise to leave a drain in the wound.

Recurrent pyogenic cholangitis

A specific type of cholangitis occurs in patients of Asiatic or Oriental origin, affecting predominantly the intrahepatic bile ducts, which contain soft stones and strictures (see also Chapter 47.6). The left hepatic system is more frequently affected than is the right, and liver abscesses are a common complication. Both sexes are affected equally, and the condition presents at a younger age than Western cholelithiasis. The gallbladder is only inflamed in about one-fifth of patients, and it rarely contains stones.

Pathology

Infection of small biliary radicles by bowel organisms, probably from an episode of gastroenteritis, is thought to be the cause. The disease is more common in malnourished people and in some populations there is an association with infection by *Clonorchis sinensis* and *Ascaris lumbricoides*. Bacterial enzymes split soluble conjugated bilirubin, forming bilirubinate sludge. Strictures of the ducts are also a constant feature, but it is uncertain whether the stones or the strictures appear first.

The primary pathology is in the bile ducts, and the liver is involved secondarily. In the acute stage the liver is oedematous and there is inflammation around the portal tracts and thrombophlebitis of the portal veins. After recurrent attacks the bile ducts become thickened and stenosed, surrounded by fibrous tissue and a chronic inflammatory infiltrate. Secondary changes develop in the liver.

Diagnosis

A clinical diagnosis is easy to make in the right context. There are typical symptoms of recurrent cholangitis in a young patient of Asian or Oriental origin and signs of chronic hepatic infection. Viral hepatitis is the principal differential diagnosis. Ultrasonography and ERCP are the main diagnostic investigations required, but blood culture and examination of the stools for parasites are also important.

Treatment

Treatment of the acute stage is directed at controlling the infection with intravenous fluids and antibiotics. Surgery is avoided unless the patient's condition deteriorates because of septicaemia from severe obstruction or generalized peritonitis from pancreatitis, rupture of an empyema of the gallbladder, or rupture of a distended hepatic duct on the surface of the liver. Acute operations are directed at draining the biliary tree with a T-tube through a choledochotomy, after clearing the duct of as many stones as possible.

Elective surgery is intended to remove the stones from the biliary tract and relieve any strictures that are present. This is difficult and tedious surgery, because the stones are very soft and do not wash away easily. On occasion some form of limited hepatic resection may be the best form of treatment; this is particularly useful if only the left hepatic system is diseased. Most surgeons also remove the gallbladder. As the name implies the disease tends to recur with time, although the ultimate prognosis is very unpredictable. When complications develop the outlook is poor.

Biliary hydatid disease

The liver is the most common site for a hydatid cyst in man (see also Chapter 47.2.4). Such cysts grow slowly in size and about two-thirds of the patients present with simple hepatomegaly. Of the remainder, one in eight are found by accident and a similar number present with biliary colic and transient jaundice due to rupture of the cyst into the biliary tree. Attention is mostly directed towards treatment of the primary cyst, which includes treatment with antihelminthics. Imaging of the biliary system is important. If hydatid elements are present in the ducts they must be removed through a choledochotomy at the same time as removal of the main cyst. Choledochoscopy before closure is useful to ensure that the duct is clear, and the choledochotomy is then closed over a T-tube. Sometimes it is wise to perform a transduodenal sphincteroplasty to ensure free drainage of any residual hydatid material into the bowel. Nowadays it might be easier to do this endoscopically soon after removal of the cyst. Occasionally, a biliary fistula persists after removal of hydatid cyst: ERCP and sphincterotomy with removal of any daughter cysts or hydatid debris from the bile ducts should allow the fistula to close.

Benign tumours of the biliary tract

Neoplasia of the biliary tract is uncommon. Carcinoma of the gallbladder and bile duct appear infrequently and benign tumours are decidedly rare. Polyps in the gallbladder are sometimes seen on an ultrasonogram or on a cholecystogram. The majority of these are cholesterol polyps or adenomyomas and not, therefore, true tumours. If gallstones are also present and causing symptoms, then the patient needs a cholecystectomy. If they are an isolated finding, most surgeons believe that they do sometimes cause symptoms; provided these are sufficiently severe, surgery is justified. True polyps or adenomas larger than 1 cm in diameter should be removed as there is a potential for malignant change. If operation is not appropriate and the polyps are less than 1 cm, then they should be kept under ultrasonographic surveillance and removal advised if they enlarge.

Papillomas of the bile duct occur most commonly at the ampulla and are usually small. They present with biliary-tract obstruction or recurrent pancreatitis. ERCP reveals the lesion unfolding into the duodenum as the sphincterotomy is made. These lesions are usually regarded as premalignant and they should be removed either surgically or endoscopically.

Adenoma of the bile duct is a rare condition that presents with jaundice. The ERCP appearances are typical (Fig. 58). Adenomas tend to recur and they can become malignant so they should be removed. Depending on their site in the biliary tract and their extent, this may require an hepatic resection.

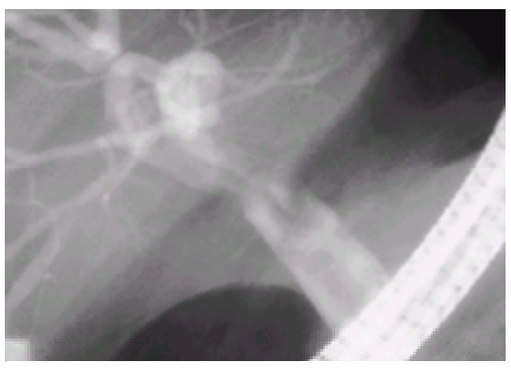


Fig. 58. Endoscopic retrograde cholangiogram of a patient with a villous adenoma of the left hepatic duct. The tumour can be seen extending out of the left hepatic duct and partially occluding the right hepatic duct. This was removed by a left hepatectomy.

Acknowledgement

The authors are grateful to Dr Roger Chapman for his help with the section on primary sclerosing cholangitis.

Further reading

Admirand WH, Small DM. The physicochemical basis of cholesterol gallstone formation in man. *Journal of Clinical Investigation* 1968; **47**: 1043–52. [The classic description of the biochemistry of cholesterol gallstone formation.]

Aldridge MC, Bismuth H. Gallbladder cancer: the polyp–cancer sequence. *British Journal of Surgery* 1990; **77**: 363–4. [Gallbladder polyps larger than 10 mm have malignant potential.]

Armstrong CP, Taylor TV, Jeacock J, Lucas S. The biliary tract in patients with acute gallstone pancreatitis. *British Journal of Surgery* 1985; **72**: 551–5. [A prospective study of 769 patients with gallstones of whom 59 developed acute pancreatitis.]

Atli AO, Coskun T, Ozenc A, Hersek E. Biliary enteric fistulas. *International Surgeon* 1997; **82**: 280–3. [A retrospective review of 31 patients seen over a 19-year period.]

Babb RR. Acute acalculous cholecystitis. A review. *Journal of Clinical Gastroenterology* 1992; **15**: 238–41. [A review with an emphasis on investigations.]

Barkun JS *et al.* Randomised controlled trial of laparoscopic versus mini cholecystectomy. *Lancet* 1992; **340**: 1116–19. [Laparoscopic cholecystectomy was more effective than mini-cholecystectomy in 70 randomly allocated patients.]

Baron RL *et al.* A prospective comparison of the evaluation of biliary obstruction using computed tomography and ultrasonography. *Radiology* 1982; **145**: 91–8. [Computed tomography identifies the site and the cause of biliary obstruction better than ultrasonography.]

Bergman JJGHM *et al.* Endoscopic balloon dilatation for common duct stones. *Lancet* 1997; **349**: 1124–9. [Balloon dilatation preserves the ampullary sphincter and is as safe and as effective as endoscopic sphincterotomy.]

Bilge A, Sozuer EM. Diagnosis and surgical treatment of hepatic hydatid disease. *HPB Surgery* 1994; **8**: 77–81. [A retrospective review of 226 patients seen in one department between 1978 and 1990.]

Blumgart LH, ed. *Surgery of the liver and biliary tract*, 3rd edn. Churchill Livingstone, Edinburgh, 1998. [The definitive, comprehensive, multiauthor textbook.]

Blumgart LH, Kelley CJ, Benjamin IS. Benign bile duct stricture following cholecystectomy: critical factors in management. *British Journal of Surgery* 1984; **71**: 836–43. [The management of iatrogenic biliary stricture following open cholecystectomy.]

Bouchier IAD. Gall stones. *British Medical Journal* 1990; **300**: 592–7. [A review of gallstone disease from a medical perspective with an emphasis on stone-dissolving drugs.]

Boyden EA. The anatomy of the choledochoduodenal junction in man. *Surgery Gynecology and Obstetrics* 1957; **104**: 641–52. [An unsurpassed description of the anatomy of the ampulla.]

Chen HH, Zhang WH, Wang SS. Twenty-two year experience with the diagnosis and treatment of intrahepatic calculi. *Surgery Gynecology and Obstetrics* 1984; **159**: 519–24. [A review of the management of 162 patients with intrahepatic bile-duct stones.]

Cooperberg PL, Burhenne HJ. Real-time ultrasonography. Diagnostic technique of choice in calculous gallbladder disease. *New England Journal of Medicine* 1980; **302**: 1277–9. [The sensitivity of ultrasound in the diagnosis of gallstones in 313 patients was 98 per cent and the specificity was between 93.5 and 97.7 per cent.]

Corazziari E, Funch Jensen P, Hogan WJ, Tanaka M, Toouli J. Functional disorders of the biliary tract. *Gastroenterology International* 1993; **6**:129–44. [A comprehensive review of gallbladder and sphincter of Oddi dysfunction.]

Corder AP, Scott SD, Johnson CD. Place of routine operative cholangiography at cholecystectomy. *British Journal of Surgery* 1992; **79**: 945–7. [Cholangiography should be performed selectively at cholecystectomy on the basis of abnormal liver function tests, a history of pancreatitis or jaundice, or dilatation of the bile ducts.]

Diamond T. Choledochal cyst. *Surgery* 1992; **10**: 75–7. [A review of the aetiology, diagnosis, and treatment of choledochal cyst.]

Fanning NF, Horgan PG, Keane FBV. Evolving management of common bile duct stones in the laparoscopic era. *Journal of the Royal College of Surgeons of Edinburgh* 1997; **42**: 389–94. [An audit

of 173 patients presenting with common bile-duct stones between 1990 and 1994.]

Fiore NF *et al.* An analysis of perioperative cholangiography in one thousand laparoscopic cholecystectomies. *Surgery* 1997; **122**: 817–23. [Laparoscopic cholangiography is most useful in patients with a high risk of common bile-duct stones.]

Folsch UR *et al.* Role of early ERCP in gall stone pancreatitis. *New England Journal of Medicine* 1997; **336**: 237–42. [Patients with cholangitis or progressive jaundice and acute pancreatitis benefit from an early ERCP.]

Freeman ML *et al.* Complications of endoscopic biliary sphincterotomy. *New England Journal of Medicine* 1996; **335**: 909–18. [A prospective multicentre study of 2347 patients.]

Ghumman E, Barry M, Grace P. Management of gallstones in pregnancy. *British Journal of Surgery* 1997; **84**: 1646–50. [Surgery should only be performed for complicated, non-resolving, biliary-tract disease in pregnancy.]

Gilchrist I, Ross B, Thomas WEG. Extracorporeal shockwave lithotripsy for common bile duct stones. *British Journal of Surgery* 1997; **84**: 29–32. [Extracorporeal shockwave lithotripsy following failed endoscopic sphincterotomy is a useful additional treatment.]

Glenn F. A 26 year experience in the surgical treatment of 5037 patients with non-malignant biliary tract disease. *Surgery Gynecology and Obstetrics* 1959; **109**: 591–606. [A unique description of the classical surgical treatment of benign biliary disease.]

Glenn F. Acute cholecystitis. *Surgery Gynecology and Obstetrics* 1976; **143**: 56–60. [A report of 2021 patients treated surgically for acute cholecystitis over a 42-year period before the advent of laparoscopic cholecystectomy.]

Glenn F, Reed C, Grafe WR. Biliary enteric fistula. *Surgery Gynecology and Obstetrics* 1981; **153**: 527–31. [105 patients with a biliary enteric fistula were seen over a 46-year period; three-quarters of them had a fistula between the gallbladder and duodenum.]

Gough MH. 'The cholecystogram is normal'—but. *British Medical Journal* 1977; **274**: 960–2. [Normal investigations do not exclude significant gallbladder disease.]

Gouma DJ, Obertop H. Acute acalculous cholecystitis. What is new in diagnosis and treatment. *HPB Surgery* 1992; **6**: 69–78. [A review of the cause and the management of acalculous cholecystitis.]

Gracie WA, Ransohoff DF. The natural history of silent gallstones. The innocent gallstone is not a myth. *New England Journal of Medicine* 1982; **307**: 798–800. [Routine cholecystectomy for silent gallstones is neither necessary nor advisable.]

Houston JP, Collins MC, Cameron I, Reed MWR, Parsons MA, Roberts KM. Xanthogranulomatous cholecystitis. *British Journal of Surgery* 1994; **81**: 1030–2. [A review of the clinical features, diagnosis, and treatment of 31 patients with this rare condition.]

Ishibashi T, Kasahara K, Yasuda Y, Nagai H, Makino S, Kanazawa K. Malignant change in the biliary tract after excision of choledochal cyst. *British Journal of Surgery* 1997; **84**: 1987–91. [A review of the risks of malignant change in 48 patients with choledochal cyst.]

Johnston DE, Kaplan MM. Pathogenesis and treatment of gallstones. *New England Journal of Medicine* 1993; **328**: 412–21. [Comprehensive review of the topic with 150 references.]

Kelly DA, Buick RG. Congenital abnormalities of the biliary tree. *Surgery* 1992; **10**: 149–52. [A review of the management of choledochal cyst.]

Khoo DE, Walsh CJ, Cox MR, Murphy CA, Motson RW. Laparoscopic common bile duct exploration: evolution of a new technique. *British Journal of Surgery* 1996; **83**: 341–6. [Laparoscopic exploration completely cleared the bile duct of stones in 38 of 51 patients.]

Lee YM, Kaplan MM. Primary sclerosing cholangitis. *New England Journal of Medicine* 1995; **332**: 924–33. [An up-to-date review of the pathogenesis, diagnosis, and treatment of primary sclerosing cholangitis.]

Le Quesne LP, Ranger I. Cholecystitis glandularis proliferans. *British Journal of Surgery* 1957; **44**: 447–58. [An old but comprehensive discussion of this unusual variant of chronic cholecystitis.]

Lillemoe KD *et al.* Major bile duct injuries during laparoscopic cholecystectomy. *Annals of Surgery* 1997; **225**: 459–71. [A prospective review from one institution of the radiological and surgical management of 89 patients with a major bile-duct injury.]

Lindsey I, Kitchen G, Leung D. Emphysematous cholecystitis. *Australia and New Zealand Journal of Surgery* 1996; **66**: 267–8. [A case report with a review of the literature.]

Losken A, Wilson BW, Sherman R. Torsion of the gallbladder: a case report and review of the literature. *American Surgeon* 1997; **63**: 975–8. [A single case report with a comprehensive reference list.]

McClement JW, Howard ER, Mowat AP. Results of surgical treatment for extrahepatic biliary atresia in the United Kingdom 1980–1982. *British Medical Journal* 1985; **290**: 345–7. [A survey of 114 infants treated for biliary atresia before the introduction of liver transplantation.]

Mitchell SA, Chapman RW. The management of primary sclerosing cholangitis. *Alimentary Pharmacology and Therapeutics* 1997; **11**: 33–43. [A review that concentrates on the treatment of primary sclerosing cholangitis.]

Motson RW, Wetter LA. Operative choledochoscopy: common bile duct exploration is incomplete without it. *British Journal of Surgery* 1990; **77**: 975–82. [A review of the collected results of rigid and flexible choledochoscopy in 4438 patients with stones in the bile ducts.]

Mueller PR, Harbin WP, Ferrucci JT, Wittenberg J, Van Sonnenberg E. Fine needle transhepatic cholangiography: reflections after 450 cases. *American Journal of Roentgenology* 1981; **136**: 85–90. [A review of the early experience with percutaneous transhepatic cholangiography.]

Musella M *et al.* Magnetic resonance imaging in evaluation of the common bile duct. *British Journal of Surgery* 1998; **85**: 16–19. [Magnetic resonance cholangiography could replace operative cholangiography and ERCP.]

Neoptolemos JP, Carr-Lock DL, Fossard DP. Prospective randomised study of preoperative endoscopic sphincterotomy versus surgery alone for common bile duct stones. *British Medical Journal* 1987; **294**: 470–4. [Surgery is better than preoperative endoscopic sphincterotomy.]

Neoptolemos JP, Carr–Locke DL, London N, Bailey I, Fossard DP. ERCP findings and the role of endoscopic sphincterotomy in acute gallstone pancreatitis. *British Journal of Surgery* 1988; **75**: 954–60. [Urgent ERCP and sphincterotomy is recommended for all patients with predicted severe gallstone pancreatitis.]

O'Connor MJ, Schwartz ML, McQuarrie DG, Sumner HW. Acute bacterial cholangitis. An analysis of clinical manifestation. *Archives of Surgery* 1982; **117**: 437–41. [Early drainage of an obstructed biliary tract is still the priority for treatment.]

Pemberton M, Wells AD. The Mirizzi syndrome. *Postgraduate Medical Journal* 1997; **73**: 487–90. [A review with case reports and a comprehensive reference list.]

Pitt HA, Postier RG, Cameron JL. Biliary bacteria. *Archives of Surgery* 1982; **117**: 445–9. [Positive bile cultures are associated with an increased incidence of wound infection and postoperative renal dysfunction.]

Ramirez P *et al.* Choledochoduodenostomy and sphincterotomy in the treatment of choledocholithisis. *British Journal of Surgery* 1994; **81**: 121–3. [A report of the long-term follow-up of exploration of the bile duct in 591 patients of whom 216 underwent supraduodenal choledochoduodenostomy.]

Richardson MC *et al.* Incidence and nature of bile duct injuries following laparoscopic cholecystectomy: an audit of 5913 cases. *British Journal of Surgery* 1996; **83**: 1356–60. [Thirty-seven bile-duct injuries occurred in this prospective audit over a 5-year period.]

Rodriguez–Sanjuan JC, Casado F, Fernandez MJ, Morales DJ, Naranjo A. Cholecystectomy and fistula closure versus enterolithotomy alone in gallstone ileus. *British Journal of Surgery* 1997; **84**: 634–7. [Simple removal of the gallstone from the small bowel is best in most patients.]

Rossi RL, Biliary tract injuries revisited. *Surgical Clinics of North America* 1994; **74**: 731–990. [An up-to-date review of the incidence, prevention, and treatment of bile-duct injuries.]

Russell RCG. Minimally invasive surgery. General surgery: biliary surgery. *British Medical Journal* 1993; **307**: 1266–9. [A review of laparoscopic biliary surgery including details of technique.]

Safrany L, Cotton PB. A preliminary report: urgent duodenoscopic sphincterotomy for acute gallstone pancreatitis. *Surgery* 1981; **89**: 424–8. [The original observation that urgent sphincterotomy allows immediate resolution of acute gallstone pancreatitis.]

Schirmer A, Rossi RL, Hughes KS, Munson JL, Braasch JW. Common operative problems in hepato-biliary surgery. *Surgical Clinics of North America* 1991; **71**: 1363–89. [Contains a useful discussion of intraoperative complications during cholecystectomy and the management of postoperative biliary fistula.]

Schwesinger WH, Diehl AK. Changing indications for laparoscopic cholecystectomy. *Surgical Clinics of North America* 1996; **76**: 493–504. [Pretransplant cholecystectomy, pre-emptive cholecystectomy before pregnancy and in young children are now all possible indications for surgery in patients with gallstones.]

Scragg RKR, McMichael AJ, Baghurst PA. Diet, alcohol, and relative weight in gall stone disease: a case-control study. *British Medical Journal* 1984; **288**: 1113–19. [Eating sugar and fat increases the risk of developing gallstones while obesity is a risk only in young women.]

Scragg RKR, McMichael AJ, Seamark RF. Oral contraceptives, pregnancy, and endogenous oestrogen in gall stone disease—a case-control study. *British Medical Journal* 1984; **288**: 1795–9. [Oral contraceptives and multiparity increase the risk of developing gallstones in young women.]

Shanahan D, Lord PH, Grogono J, Wastell C. Clinical acute cholecystitis and the Curtis Fitz-Hugh syndrome. *Annals of the Royal College of Surgeons of England* 1988; **70**: 44–6. [Female patients with right upper-quadrant pain and a normal ultrasound scan should be screened for the Curtis Fitz-Hugh syndrome.]

Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *New England Journal of Medicine* 1991; **324**: 1073–8. [The first large reported series of laparoscopic cholecystectomy.]

Stubbs RS. Wound infection after cholecystectomy: a case for routine prophylaxis. *Annals of the Royal College of Surgeons of England* 1983; **65**: 30–1. [Antibiotic prophylaxis reduced the incidence of wound infection from 15 to 3.6 per cent.]

Toscano RL, Taylor PH, Peters J, Edgin R. Mirizzi syndrome. *American Surgeon* 1994; **60**: 889–91. [A case report with a review of the literature.]

Van Sonnenberg E, D'Agostino HB, Goodacre BW, Sanchez RB, Casola G. Percutaneous drainage of the gall bladder. *Radiology* 1992; **183**: 167–70. [The results of percutaneous cholecystostomy in 127 patients.]

Vitetta L, Sali A, Little P, Nayman J, Elzarka A. Primary ‘brown pigment' bile duct stones. *HPB Surgery* 1991; **4**: 209–22. [Primary bile-duct stones are different from gallbladder stones and contain a low concentration of cholesterol.]

Weitemeyer R. The treatment of ampullary stenosis by endoscopic sphincterotomy (EST). *American Journal of Surgery* 1994; **167**: 493–6. [A review of 247 patients with ampullary stenosis. Symptoms were relieved by EST in 70 per cent.]

Wright KD, Wellwood JM. Bile duct injury during laparoscopic cholecystectomy without operative cholangiography. *British Journal of Surgery* 1998; **85**: 191–4. [A prospective audit of a consecutive series of 1200 laparoscopic cholecystectomies without operative cholangiography; three bile-duct injuries occurred.]

31.2 Malignant diseases of the biliary tract

Steven A. Ahrendt and Henry A. Pitt

Gallbladder cancer
Incidence
Etiology
Pathology and staging
Diagnosis
Resection
Palliative therapy
Adjuvant therapy
Survival
Cholangiocarcinoma
Incidence
Etiology and associated diseases
Staging and classification
Diagnosis
Surgical resection
Palliative therapy
Adjuvant therapy
Survival
Summary
Further reading

Malignancies of the biliary tract can be divided into two main groups, gallbladder cancer and cholangiocarcinoma. In most parts of the world, gallbladder cancers are two to four times more common than bile-duct malignancies. The vast majority of patients with gallbladder cancer have one or more gallstones, whereas cholelithiasis is no more common in patients with cholangiocarcinoma than in the general population. Malignancies of both the gallbladder and the bile ducts are almost always adenocarcinomas. Gallbladder cancers usually present with pain, whereas bile-duct tumors almost always present with jaundice. For both tumors, the radiologic work-up is similar, and resection with negative margins is the goal of therapy. Unfortunately, many patients present at an advanced stage and palliation is the only option. The role of chemotherapy and radiation as an adjuvant to surgical resection or for palliation remains controversial. When the tumor is detected at an early stage and completely resected, prolonged survival can be expected, but this occurs in only a small proportion of patients.

A full discussion of gallbladder cancer will be followed by details of cholangiocarcinoma.

Gallbladder cancer

Cancer of the gallbladder is an aggressive malignancy that occurs predominantly in elderly people and is strongly associated with gallstones. With the exception of early cases detected incidentally at cholecystectomy for gallstone disease, the prognosis for most patients is poor. Many of these tumors are unresectable at presentation and most can be managed nonoperatively. Recently, an aggressive surgical approach for patients with localized gallbladder cancer has produced encouraging results with acceptable morbidity.

Incidence

Gallbladder cancer is the fifth most common gastrointestinal malignancy. It is two to three times more common in females than males, in part due to the higher incidence of gallstones in women. Over 75 per cent of the patients are over the age of 65 years. Approximately 5000 new cases are diagnosed annually in the United States of America, and the overall incidence is 2.5 cases per 100 000 residents. Incidence varies considerably with both ethnic background and geographic location. In the United States, gallbladder cancer is more common in Native Americans, and the annual incidence in Native American females with gallstones approaches 75 cases per 100 000. Similarly, in Chile, where over 50 per cent of women over the age of 50 years have gallstones, adenocarcinoma of the gallbladder is the leading cause of cancer deaths among women.

Etiology

Several factors have been associated with an increased risk of developing gallbladder cancer. Among these factors, gallstones are the most common because of their high prevalence in the general population. The incidence of gallstones increases with age, and by the age of 75 years about 35 per cent of women and 20 per cent of men in the United States will have developed them. The association between gallstones and carcinoma of the gallbladder was observed by Mayo in 1903. Cholelithiasis is present in 75 to 90 per cent of cases of gallbladder cancer. Gallbladder cancer is approximately seven times more common in the presence of cholelithiasis and chronic cholecystitis than in people without gallstones. In addition, the risk of developing gallbladder cancer is higher in patients with symptomatic than asymptomatic gallstones. Approximately 1 per cent of all elective cholecystectomies for cholelithiasis will discover an occult gallbladder cancer.

More recently, an anomalous junction of the pancreatobiliary duct, exposure to carcinogens such as azotoluene and nitrosamines, a porcelain gallbladder, and other biliary disorders such as choledochal cysts and primary sclerosing cholangitis have been associated with gallbladder cancer.

Pathology and staging

Ninety per cent of cancers of the gallbladder are adenocarcinomas. Six per cent of gallbladder cancers demonstrate papillary features histopathologically; these tumors are commonly diagnosed while localized to the gallbladder and are also associated with a better overall survival. At diagnosis, 25 per cent of cancers are localized to the gallbladder wall, 35 per cent have associated metastases to regional lymph nodes or extension into adjacent organs, and 40 per cent have already metastasized to distant sites.

Lymphatic drainage from the gallbladder is predictable and correlates with the pattern of nodal metastases found in gallbladder cancer. Lymph from the gallbladder drains initially to the cystic duct node and then descends along the common bile duct to pericholedochal lymph nodes. It then proceeds to nodes posterior to the head of the pancreas and then to interaortocaval nodes. Secondary routes of lymphatic drainage include the retroportal and right celiac lymph nodes. The current **TNM** (tumor, node, metastasis) classification of the American Joint Committee on Cancer is shown in Table 1. The appropriate management and overall prognosis are strongly dependent on tumor stage.

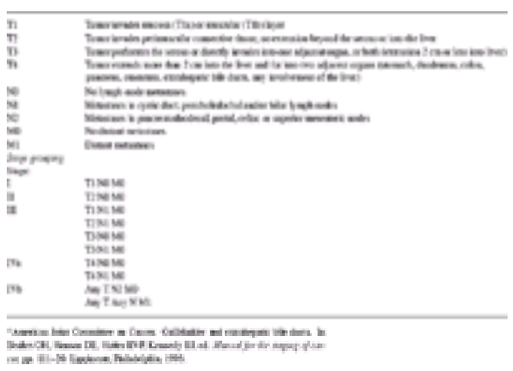


Table 1 TNM staging for gallbladder cancer*

Diagnosis

Clinical presentation

Gallbladder cancer most often presents with pain in the right upper quadrant of the abdomen mimicking acute or chronic cholecystitis. Weight loss, jaundice, and an abdominal mass are less common presenting symptoms. The preoperative diagnosis is difficult. In a recent series, only 8 per cent of 53 patients with carcinoma of the gall-bladder were diagnosed correctly preoperatively. The most common misdiagnoses included chronic cholecystitis (28 per cent), pancreatic cancer (13 per cent), acute cholecystitis (9 per cent), choledocholithiasis (8 per cent), and gallbladder hydrops (8 per cent).

Radiologic evaluation

Ultrasonography, computed tomography (**CT**), magnetic resonance imaging (**MRI**), cholangiography, and angiography may all be helpful in evaluating patients with gallbladder cancer. The sensitivity of ultrasound in the detection of gallbladder cancer ranges from 70 to 100 per cent. CT usually demonstrates a mass replacing the gallbladder or extending into adjacent organs (Fig. 1). Spiral CT also demonstrates invasion of the adjacent liver, which is present in as many as 70 to 80 per cent of the patients, as well as the adjacent vascular anatomy. With newer MRI techniques, gallbladder cancers may be differentiated from the adjacent liver and biliary obstruction and/or encasement of the portal vein may also be easily visualized (Fig. 2).

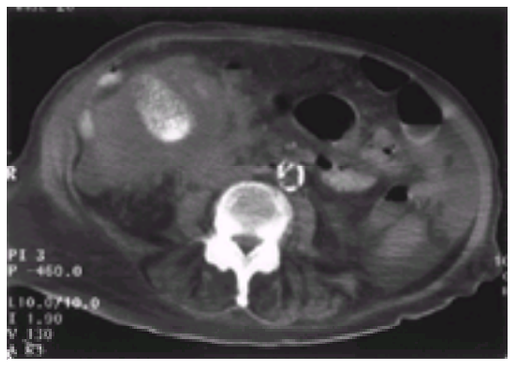


Fig. 1. CT scan demonstrating a large gallbladder cancer with extension into the duodenum; gallstones (calcifications) are present within the mass.

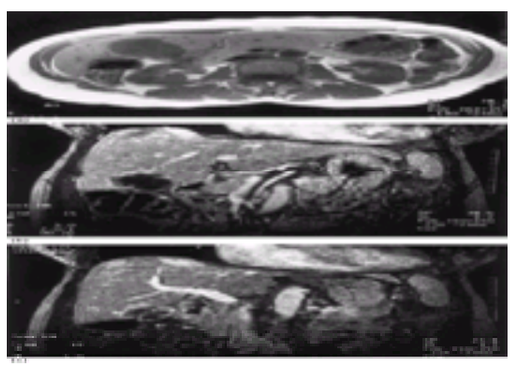


Fig. 2. (a) MRI scan demonstrating a mass arising within the gallbladder with extension into the hepatic parenchyma; (b) MR angiogram from the same patient demonstrating the relation between the hepatic artery and the mass; (c) MR angiogram from the same patient showing the relation between the main and right portal vein and the mass.

Cholangiography also may be helpful in diagnosing jaundiced patients with gallbladder cancer. The typical cholangiographic finding in gallbladder cancer is a long stricture of the common hepatic duct. Angiography may identify encasement of the portal vein or hepatic artery, but the newer spiral CT and MRI techniques usually provide the same information. If radiologic studies suggest that the tumor is unresectable (liver or peritoneal metastases, portal vein encasement, or extensive hepatic invasion), a biopsy of the tumor is warranted and can be performed under ultrasound or CT guidance.

Resection

The most effective treatment for localized gallbladder cancer is resection of the primary tumor along with the regional lymph nodes. After preoperative staging, patients with clinical stage I to IVa tumors warrant exploration if they are otherwise suitable operative candidates.

Simple cholecystectomy

The appropriate operative procedure for the patient with localized gallbladder cancer is determined by the pathologic stage. Patients with tumors confined to the gallbladder mucosa or submucosa (T1a) are usually identified following cholecystectomy for gallstone disease, have a negligible incidence of nodal metastases, and an overall 5-year survival approaching 100 per cent. More recently, with widespread use of the laparoscopic approach, recurrent cancer at port sites and peritoneal carcinomatosis have been reported after cholecystectomy, even in patients with *in situ* disease. The correct management of patients after laparoscopic cholecystectomy with or without bile spillage for an incidental gallbladder cancer remains unclear. Thus, those with suspected gallbladder cancer preoperatively should undergo open cholecystectomy to minimize the chance of bile spillage and tumor dissemination.

Extended cholecystectomy

Cancer of the gallbladder with invasion into (T1b) or beyond (stages II–IVa) the gallbladder muscularis is associated with an increasing incidence of metastases to regional lymph nodes and should be managed with an extended lymphadenectomy as part of the operative procedure. The extent of the lymphadenectomy is based on the operative findings and knowledge of the patterns of lymphatic spread in patients with gallbladder cancer. This dissection should include the cystic duct, pericholedochal, portal, right celiac, and posterior pancreatoduodenal lymph nodes. In patients without any grossly involved nodes, the common bile duct, hepatic artery, and portal vein are skeletonized from the porta hepatis to the pancreas and celiac axis, respectively. Following an extensive Kocher maneuver, the posterior pancreatoduodenal lymph nodes are also included with the specimen. In the presence of any grossly enlarged pericholedochal nodes, consideration should be given to resecting the common bile duct to avoid leaving a tumor-positive margin. Biliary–enteric continuity is restored with a Roux-en-Y hepaticojejunostomy.

Tumor extension into the hepatic parenchyma is common, and extended cholecystectomy should incorporate at least a 2-cm margin beyond the palpable or sonographic extent of the tumor to include any angiolymphatic extension from the main mass. For smaller tumors, this goal can be achieved with a wedge resection of the liver.

Extended resections

For stage IV tumors invading more than 2 cm into the liver, an anatomic liver resection may be required to achieve a histologically negative margin. Cancers originating in the gallbladder fundus can usually be treated adequately by resection of segments IVb and V (Fig. 3). Tumor extension into either the right hepatic artery or portal vein may only be resectable with a right hepatic lobectomy or trisegmentectomy.

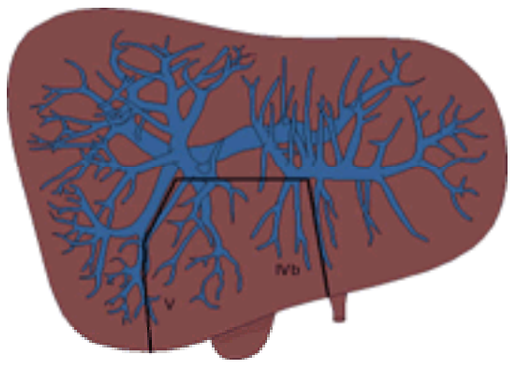


Fig. 3. Schema demonstrating extent of bisegmentectomy (segments IVb and V) in patients with T4 gallbladder cancer.

Local extension into the colon or duodenum requires an *en bloc* resection of these structures to achieve an adequate margin. In the absence of distant nodal, peritoneal, or hepatic metastases, these lesions may still be resectable for cure. Extension into the right colon should be managed with segmental colectomy, while duodenal invasion may require a pancreatoduodenectomy. A recent trend in Japan has been to perform a pancreatoduodenectomy as part of an extensive resection of lymph nodes even when the tumor does not extend directly into the duodenum or pancreas. However, in most series, the mortality of a combined liver resection and a pancreatoduodenectomy has been rather high, and this very extensive surgery should be undertaken with considerable caution.

Palliative therapy

Most therapy for gallbladder cancer is palliative. If a tissue diagnosis can be established in patients with an unresectable tumor, nonoperative palliation should be considered. Many of these patients have obstructive jaundice that can be managed with either an endoscopic or percutaneously placed biliary stent. Options include indwelling plastic or metallic endoprostheses and internal–external stents.

Pain is another problem that should be treated aggressively to improve quality of life. Narcotics should be used as necessary. Radiation therapy may also help with pain but takes a considerable time to administer and is rather costly. Percutaneous nerve block of the celiac ganglion may reduce the need for narcotics. Chemical splanchnicectomy can also be performed in patients who are found have unresectable tumors at the time of operative exploration. Obstruction of the gastric outlet is difficult to relieve nonoperatively and usually requires a gastrojejunostomy.

Adjuvant therapy

The results of chemotherapy in the treatment of patients with gallbladder cancer have been rather poor due to limited responsiveness. Partial response rates to 5-fluorouracil alone or in combination with methyl-CCNU [1(2-chloroethyl)-3-cyclohexyl)-1-nitrosourea] or streptozotocin range from 5 to 13 per cent, with a median survival of 10 to 21 weeks in treated patients. Recently, the combination of intravenous 5-fluorouracil, high-dose levofolinic acid, and oral hydroxyurea has achieved a partial response rate of 30 per cent with a median survival of 8 months in patients with unresectable gallbladder cancer.

External-beam and intraoperative radiation therapy have both been used in the management of patients with gallbladder cancer. No randomized studies have demonstrated improved survival with external-beam radiation alone as a postoperative adjuvant. In a retrospective review of 38 patients with gallbladder cancer managed with postoperative adjuvant therapy, overall survival was better in those who received greater than 4000 cGy than in those who received less. However, the response to the radiation therapy in this series was difficult to differentiate from the effect of resection alone. The use of intraoperative radiation therapy may provide a slight survival benefit in patients with locally advanced gallbladder cancer.

Survival

Extended cholecystectomy is well tolerated, with an operative mortality in several recent series of 0 to 1 per cent. Common postoperative complications include wound or intra-abdominal infection, hemorrhage, or delayed gastric emptying. Fatalities increase as the magnitude of the operative procedure increases. Extended cholecystectomy combined with a major hepatectomy or pancreatoduodenectomy has a mortality rate of 10 to 15 per cent, and adds the risks of biliary or pancreatic anastomotic leaks to the procedure.

As outlined above, survival in patients with gallbladder cancer is strongly influenced by the pathologic stage at presentation. Patients with cancer limited to the gallbladder mucosa and submucosa (T1a) have a uniformly excellent prognosis. Thirty-five patients with incidentally discovered, T1a gallbladder cancer recently reported by Shirai *et al.* had an overall 5-year survival of 100 per cent. Invasion into the muscular wall of the gallbladder increases the risk of recurrence after attempted curative resection. The reported 5-year survival for patients with T1b gallbladder cancer varies widely, ranging from 20 to 100 per cent (Table 2). In a survey of 172 major hospitals in Japan, the 5-year survival for patients with T1b tumors was 72 per cent, with many patients undergoing simple cholecystectomy, suggesting that this procedure is inadequate once muscular invasion is present.

Table 2 Actuarial survival in T1 gallbladder cancer

Invasion into either the muscularis or the subserosa increases the risk of metastasis to regional lymph nodes to 15 and 50 per cent, respectively. Tsukada *et al.* report a 5-year overall survival for patients with stage II disease of 80 per cent with extended cholecystectomy. Similar results have been reported with an aggressive surgical approach at the Memorial Sloan-Kettering Hospital. In patients with tumors confined to the gallbladder wall or with local extension into the liver or other adjacent organs (stages III and IV), long-term survival is possible after extended cholecystectomy. Several groups have recently reported 5-year overall survival for patients with stage III and stage IV gallbladder cancer of 40 to 63 per cent and 19 to 25 per cent, respectively (Fig. 4).

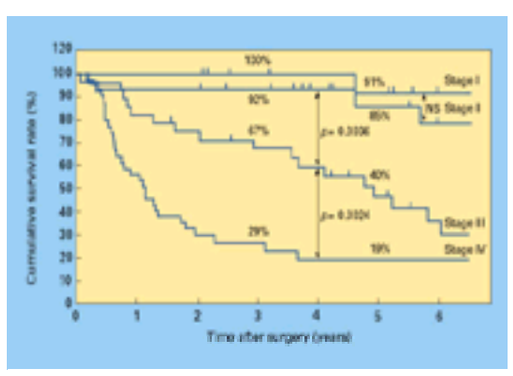


Fig. 4. Actuarial survival for patients with gallbladder cancer by pathologic stage (see Table 1). Survival in patients with stage I and stage II tumors was significantly greater than patients with stage III tumors. In addition, stage III patients survived significantly longer than patients with stage IV disease.

Tsukada *et al.* identified a curative, margin-negative resection as the most significant factor in predicting prognosis. Five-year survival (52 per cent) in patients undergoing resection who had microscopically negative margins was significantly greater than the 5-year survival (5 per cent) in patients with a microscopically positive margin (Fig. 5). The majority of patients with gallbladder cancer have, as already mentioned, advanced unresectable disease at the time of presentation. As a result, fewer than 15 per cent of all patients with gallbladder cancer are alive after 5 years. The median survival for stage IV patients at the time of presentation is only 1 to 3 months.

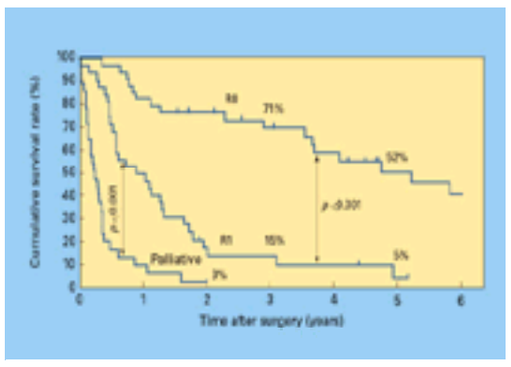


Fig. 5. Actuarial survival for patients with stage III and stage IV gallbladder cancer by resection-margin status. Results shown for patients managed with curative (negative margins) resection (R0), noncurative (microscopic positive margin) resections (R1), and palliative (grossly positive margin) resections (R2). Patients managed with R0 resection survived significantly longer than those receiving R1 resection. In addition, patients with extensive tumor resected with palliative intent survived a significantly shorter time than those with microscopic positive margins.

Cholangiocarcinoma

Cholangiocarcinoma is an uncommon tumor that may occur anywhere along the intrahepatic or extrahepatic biliary tree. The bifurcation of the hepatic duct is the most frequently involved site, and approximately 60 to 80 per cent of cholangiocarcinomas encountered at tertiary referral centers are found in the perihilar region. Distal tumors are the second most common, whereas purely intrahepatic cholangiocarcinomas occur with the lowest frequency. When possible, surgical resection does offer a chance for long-term, disease-free survival, but many patients will only be candidates for palliative bypass or operative or nonoperative intubation aimed at providing biliary drainage and preventing cholangitis and hepatic failure.

Incidence

Between 2500 and 3000 new cases of cholangiocarcinoma are diagnosed annually in the United States of America. The incidence increases with age, and these tumors occur with similar frequency in men and women. Overall, the incidence of cholangiocarcinoma in the United States is approximately 1.0/100 000 people per year.

Etiology and associated diseases

A number of diseases and environmental agents have been linked to cholangiocarcinoma, including primary sclerosing cholangitis, choledochal cysts, and hepatolithiasis. Features common to a number of these etiologic factors include bile-duct stones, biliary stasis, and infection. Cancers of the bile duct in patients with primary sclerosing cholangitis are most often extrahepatic, commonly occur near the bifurcation of the hepatic duct, and are difficult to differentiate from the multiple, benign strictures associated with this disease. The mean age at presentation in patients with cholangiocarcinoma and primary sclerosing cholangitis is in the fifth decade of life. Similarly, choledochal cysts are usually diagnosed in childhood or early adult life, and the risk of cholangiocarcinoma increases steadily with age. Hepatolithiasis is a definite risk factor for cholangiocarcinoma and will develop in between 5 and 10 per cent of the patients with intra-hepatic stones. Many other risk factors have been identified, including liver flukes, Thorotrast, dietary nitrosoamines, and exposure to dioxin.

Staging and classification

Cholangiocarcinoma is best classified into three broad groups: (1) intrahepatic, (2) perihilar, and (3) distal (Fig. 6). This classification correlates with the anatomic distribution and indicates the preferred treatment for each site. Intrahepatic tumors are treated like hepatocellular carcinoma, with hepatectomy when possible. Distal tumors are managed, in a similar fashion to other periampullary malignancies, by pancreatoduodenectomy. The perihilar tumors make up the largest group and are managed by resection of the bile duct with or without hepatic resection.

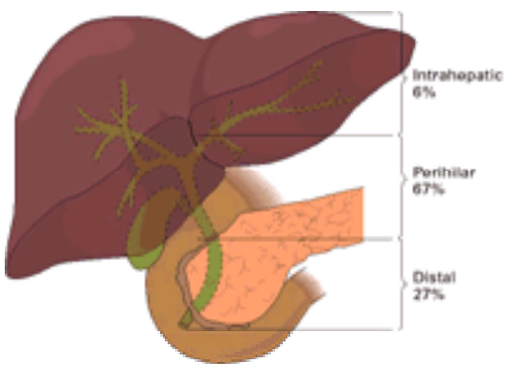


Fig. 6. Distribution of 294 cholangiocarcinomas into intrahepatic, perihilar, and distal subgroups.

Cancers of the bifurcation of the hepatic duct have also been classified by Bismuth according to their anatomic location. In this system, type I tumors are those confined to the common hepatic duct, and type II involve the bifurcation without involvement of secondary intrahepatic ducts. Type IIIa and IIIb tumors extend into either the right or left secondary intrahepatic ducts, respectively, and type IV tumors involve the secondary intrahepatic ducts on both sides.

Cholangiocarcinoma is also staged according to the TNM classification of the American Joint Commission on Cancer (Table 3). In this system, stage I tumors are limited to the mucosa or muscular layer of the bile duct whereas stage II tumors invade periductal tissues. Stage III tumors have metastases to regional lymph nodes whereas stage IV tumors either invade adjacent structures (IVa) or have distant metastases (IVb).

T1	Tumour invades mucosa or muscle layer
T2	Tumour invades perihepatic tissue
T3	Tumour invades adjacent structures (liver, pancreas, duodenum, gallbladder, colon, stomach)
N0	No regional lymph node metastasis
N1	Metastasis in cystic duct, pericholedochal and/or hilar lymph nodes
N2	Metastases in the peripancreatic (head only), periduodenal, periportal, retro- superior mesenteric, and/or posterior gastrocolic nodes and/or lymph nodes
M0	No distant metastasis
M1	Distant metastasis
Stage grouping	
Stage:	
I	T1 N0 M0
II	T2 N0 M0
III	T1 N1 M0 T2 N1 M0 T3 N0 M0 T3 N1 M0
IVA	T3 Any N M0
IVB	Any T Any N M1

*American Joint Committee on Cancer. Definition and morphology of bile ducts. In: Beahrs OH, Henson DE, Hutter RVP, Kennedy BJ, ed. *Manual for staging of cancer*, 2nd ed. pp. 111–20. Lippincott, Philadelphia, 1983.

Table 3 TNM staging for tumors of the extrahepatic bile ducts*

Diagnosis

Clinical presentation

More than 90 per cent of patients with perihilar or distal tumors present with jaundice; those with intrahepatic cholangiocarcinoma are rarely jaundiced until late in the course of the disease. Less common presenting clinical features include pruritus, fever, mild abdominal pain, fatigue, anorexia, and weight loss. Cholangitis is not a frequent presenting finding but most commonly develops after biliary manipulation by endoscopic or percutaneous techniques.

Radiologic evaluation

The goal of radiologic evaluation in patients with cholangiocarcinoma is the delineation of the overall extent of the tumor, including any involvement of the bile ducts, liver and portal vessels, and of distant metastases. An ordered sequence of tests will usually achieve these goals. The initial radiographic studies consist of either abdominal ultrasonography or CT scanning. Intrahepatic cholangio-carcinomas are easily visualized on CT scans, but perihilar tumors are often difficult to visualize on ultrasound and standard CT scans. A hilar cholangiocarcinoma will give a picture of a dilated intrahepatic biliary tree, a normal or collapsed gallbladder and extrahepatic biliary tree, and a normal pancreas. Distal tumors will lead to dilatation of the gallbladder and both the intra- and extrahepatic biliary tree.

After the finding of a dilatated bile duct, biliary anatomy has usually been defined cholangiographically through either the percutaneous transhepatic or the endoscopic retrograde routes. The most proximal extent of the tumor is the most important feature in determining resectability in patients with perihilar tumors, and the percutaneous route is favored in these patients because it defines the proximal extent of tumor involvement most reliably. Recently, magnetic resonance cholangiopancreatography has shown diagnostic accuracy comparable to percutaneous and endoscopic cholangiography. In a series of 14 patients with cholangiocarcinoma, diagnostic-quality magnetic resonance cholangiopancreatographic images were better at visualizing intrahepatic biliary anatomy than endoscopic retrograde cholangiopancreatography. Magnetic resonance cholangiopancreatography also has the advantage of obtaining two-dimensional images to define further an obstructing lesion. The sensitivity, specificity, and overall accuracy of magnetic resonance cholangiopancreatography at determining the level of biliary obstruction and the presence of a benign or malignant lesion are similar to those of endoscopic retrograde cholangiopancreatography.

Biopsy/cytology

In efforts to establish a tissue diagnosis, methods including percutaneous fine-needle aspiration biopsy, brush-and-scrape biopsy, and cytologic examination of bile have all been used. However, their sensitivity in detecting a malignancy is only approximately 30 to 50 per cent. Thus, prolonged efforts to obtain a preoperative tissue diagnosis are not indicated unless the patient is not a candidate for surgery.

Surgical resection

Curative treatment of patients with cholangiocarcinoma is only possible with complete resection. The operative approach depends on the site and extent of the tumor. In the presence of peritoneal or contralateral liver metastases, resection is contraindicated. For patients with anatomically resectable intrahepatic cholangiocarcinoma and without advanced cirrhosis, partial hepatectomy is the procedure of choice. Care must be taken to achieve a negative margin on the bile duct if the tumor approaches the hilum. Approximately 40 per cent of resectable intrahepatic cholangiocarcinomas will be either multiple and/or involve both hepatic lobes. Hepatic resection should be planned to remove completely all tumor with an adequate margin, which most commonly involves a formal lobectomy or trisegmentectomy.

Patients with perihilar tumors involving the bifurcation or proximal common hepatic duct (Bismuth types I or II) who have no vascular invasion are candidates for local excision. Biliary–enteric continuity is restored with bilateral hepaticojejunostomies. If a negative margin cannot be achieved on either the right or left, resection of the involved lobe is indicated. Similarly, if preoperative evaluation suggests a Bismuth type IIIa or IIIb tumor, right or left hepatic lobectomy, respectively, should be planned (Fig. 7). To achieve negative margins, resection of the adjacent caudate lobe should be part of this procedure.

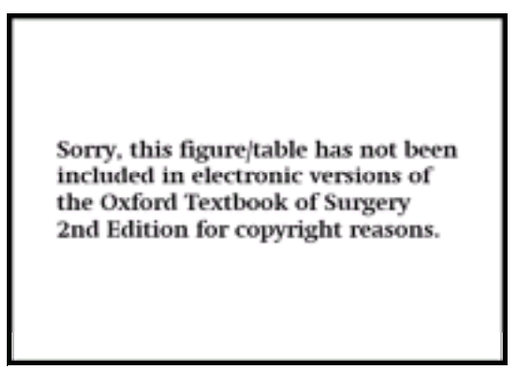


Fig. 7. (a) Diagram illustrating left hepatic and hilar resection of Bismuth type IIIb cholangiocarcinoma with preoperatively placed, transhepatic stents. (b) Diagram of resected left hepatic lobe and hilum with perihilar cholangiocarcinoma (top) and right hepatic lobe with divided right hepatic duct prior to reconstruction (bottom). (c) A Silastic transhepatic stent is placed through a right Roux-en-Y cholangiojejunostomy after left hepatic lobectomy.

For patients with resectable distal cholangiocarcinoma, pancreatoduodenectomy is the optimal procedure. Over 85 per cent of those with a tumor of the distal bile duct will be suitable for the pylorus-preserving modification. The pylorus-preserving procedure can be performed safely, is an adequate cancer operation, and provides a better quality of life than more radical operations.

Palliative therapy

Surgical exploration should be undertaken in ‘good risk' patients without evidence of metastatic or locally unresectable disease; however, intraoperatively more than half of these patients are found to have either peritoneal or hepatic metastases or, more likely, locally unresectable disease. In patients with extensive metastatic disease, preoperatively placed biliary stents may be left in place, but the gallbladder should be removed to prevent acute cholecystitis. Postoperatively, larger, but softer, transhepatic stents or metallic endoprosthesis may be placed over guidewires.

In patients with locally advanced, unresectable perihilar tumors, several operative approaches are available for palliation. These include a Roux-en-Y choledochojejunostomy with intraoperative placement of Silastic biliary catheters (Fig. 8) or a segment III or V cholangiojejunostomy. A tissue diagnosis of cholangiocarcinoma should be established for patients who are candidates for postoperative chemoradiation. The advantages of this approach over nonoperative palliation include (1) removal of the gallbladder, (2) placement of larger, softer Silastic stents with a lower risk of hemobilia and improved patient comfort, and (3)

positioning of the stent into a defunctionalized Roux-en-Y jejunal limb to reduce the incidence of subsequent cholangitis.

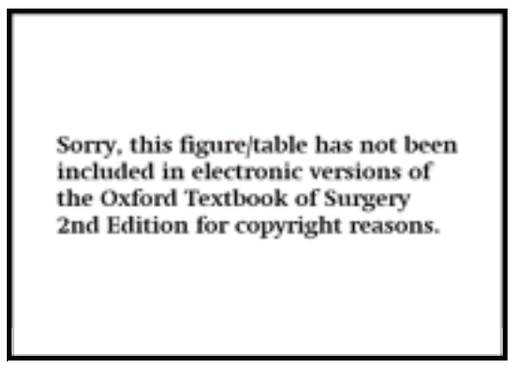


Fig. 8. (a) Diagram illustrating palliative intubation of Bismuth type IV, unresectable, perihilar cholangiocarcinoma with preoperatively placed transhepatic stents. (b) The common bile duct is divided distal to the tumor and a cholecystectomy is performed. (c) A Roux-en-Y choledochojejunostomy is constructed over Silastic transhepatic stents distal to the tumor.

Most tumors of the distal bile duct are resectable; but if resection is not possible due to vascular encasement, a cholecystectomy, a Roux-en-Y hepaticojejunostomy proximal to the tumor, and a gastro-jejunostomy to prevent obstruction of the gastric outlet should be performed. A chemical splanchnicectomy should also be included to treat or prevent pain.

Adjuvant therapy

Numerous reports have suggested that radiation therapy improves survival for patients with cholangiocarcinoma, especially when resection is impossible. External-beam radiotherapy has been delivered to a total dose of 45 to 60 Gy. Innovative techniques to increase the dose delivered to the tumor have included the use of high dose-rate equipment, charged particles, intraoperative radiotherapy, brachytherapy with iridium-192 or cobalt-60 via percutaneous or endoscopic stents, and radioimmunotherapy with iodine-131-labeled anticarcinoembryonic antigen. No prospective, randomized trials have been reported, and in many reports the patients chosen to receive radiation therapy have been healthier than those in the comparison groups and have had localized or resectable tumors. Moreover, a recent, well-controlled, but not randomized, trial from Johns Hopkins reported no benefit for postoperative adjuvant radiation.

Chemotherapy alone, using 5-fluorouracil, mitomycin C or cisplatin, has not been shown to improve survival in patients with either resected or unresected cholangiocarcinoma. Response rates to multidrug chemotherapeutic regimens also remain low. Given the potential radiosensitizing effect of 5-fluorouracil, the combination of radiation and chemotherapy may be more effective than either agent alone. Preliminary data in relatively small numbers of patients have been reported and are encouraging, but no data from randomized trials are available. However, giving chemoradiation to these patients is difficult, because of the high incidence of biliary sepsis.

Survival

Procedure-related morbidity and mortality, quality of survival, and long-term survival are highly dependent on the stage of disease at presentation and on whether the patient is treated by a palliative procedure or complete tumor resection. Intrahepatic cholangiocarcinomas often present in advanced stages, and multiple tumors and/or bilobar involvement are common. For resectable intrahepatic cholangiocarcinoma, overall 5-year survival ranges from 30 to 40 per cent. In comparison, overall 5-year survival for patients with resectable perihilar tumors has been only 10 to 20 per cent. In recent years, more hepatic resections have been performed for perihilar cholangiocarcinoma. Operative mortality for these more extensive procedures is 6 to 8 per cent. Five-year survivals of 33 to 46 per cent have been reported for patients undergoing these more extended resections who have negative margins microscopically (Fig. 9). Patients with cancer of the distal bile duct have the highest rate of resection; they have a median survival of 32 to 38 months and a 5-year survival rate of 28 to 45 per cent.

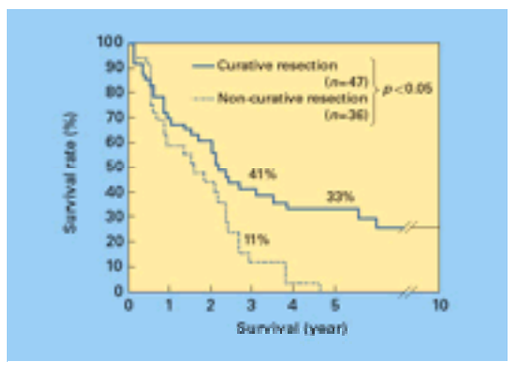


Fig. 9. Actuarial Kaplan–Meier survival for patients with perihilar cholangiocarcinoma undergoing curative or noncurative resection (microscopic tumor at surgical margin). Survival was improved ($p < 0.05$) in the curatively resected patients.

Hepatic transplantation has been employed in patients with unresectable intrahepatic and extensive perihilar tumors. Median survivals of up to 18 months have been reported in node-negative patients, but in some series the operative mortality has been as high as 25 per cent and, in general, node-positive patients have not survived longer than 2 years. Moreover, with the present shortage of organs, most groups are no longer transplanting livers for cholangiocarcinoma. Even with multimodality adjuvant therapy, median survival for unresectable intrahepatic tumors has only been 6 to 7 months. Similarly, median survival for patients with unresectable perihilar tumors varies between 5 and 8 months. One report suggests that survival for these patients may be slightly longer with surgical palliation, but no data from randomized studies are available. Good data on survival for unresectable tumors of the distal bile duct are not available because these rare cases tend to be included in larger series of unresectable pancreatic cancer.

Summary

Despite overall advances in the ability to diagnose and treat malignant tumors of the gallbladder and biliary tract, the prognosis for most patients with these malignancies remains poor. Improvements in diagnostic CT scanning and MRI have improved the ability to diagnose and stage these tumors noninvasively. Complete surgical resection remains the only curative treatment. The use of aggressive surgical approaches is likely to continue, and the challenge remains in being able to perform them safely in jaundiced and, sometimes, septic patients. Unfortunately, for many patients with malignant tumors of the biliary tract, curative resection will not be possible, and optimal palliation remains the goal of therapy. Finally, advances in adjuvant chemotherapy, radiotherapy, immunotherapy, or other innovative treatments will be required to improve further the overall prognosis of patients with either gallbladder cancer or cholangiocarcinoma.

Further reading

Ahrendt SA, Cameron JL, Pitt HA. Current management of patients with perihilar cholangiocarcinoma. In: Cameron JL, ed. *Advances in surgery*, Vol. 30, pp. 427–52. Mosby, St. Louis, 1996. [Review article outlining the management and prognosis of patients with perihilar cholangiocarcinoma.]

Ahrendt SA *et al.* Diagnosis and management of cholangiocarcinoma in primary sclerosing cholangitis. *Journal of Gastrointestinal Surgery* 1999; **3**: 537–68. [Single institution experience of 25 patients with primary sclerosing cholangitis and cholangiocarcinoma.]

American Joint Committee on Cancer. Gallbladder and extrahepatic bile ducts. In: Beahrs OH, Henson DE, Hutter RVP, Kennedy BJ, ed. *Manual for the staging of cancer*, pp. 111–20. Lippincott,

Philadelphia, 1993. [Staging manual, which includes the TNM staging system for gallbladder and extrahepatic bile duct cancer.]

Bartlett DL, Fong Y, Fortner JG, Brennan MF, Blumgart LH. Long-term results after resection for gallbladder cancer: implications for staging and management. *Annals of Surgery* 1996; **224**: 639–46. [Single institution review demonstrating improved survival after more extensive resections for gallbladder cancer.]

Bismuth H, Nakache R, Diamond T. Management strategies in resection for hilar cholangiocarcinoma. *Annals of Surgery* 1992; **215**: 31–8. [Series of 23 patients undergoing resection for hilar cholangiocarcinoma demonstrating long-term survival with negative resection margins.]

Cameron JC. *Atlas of surgery*, Vol. 1. Decker, Philadelphia, 1990. [Atlas illustrating surgical techniques for managing bile duct and gallbladder cancer at The Johns Hopkins Hospital.]

Cameron JL, Pitt HA, Zinner MJ, Kaufman SL, Coleman JA. Management of proximal cholangiocarcinoma by surgical resection and radiotherapy. *American Journal of Surgery* 1990; **159**: 91–8. [Retrospective review of single institution experience with perihilar cholangiocarcinoma.]

Carriaga MT, Henson DE. Liver, gallbladder, extrahepatic bile ducts, and pancreas. *Cancer* 1995; **75**: 171–90. [Epidemiological data on gallbladder and bile duct cancer from SEER database of the National Cancer Institute.]

Casavilla FA *et al.* Hepatic resection and transplantation for peripheral cholangiocarcinoma. *Journal of the American College of Surgeons* 1997; **195**: 429–36. [Single institution experience reporting comparable survival following resection or transplant for intrahepatic cholangiocarcinoma.]

Cubertafond P, Gainant A, Cucchiaro G. Surgical treatment of 724 carcinomas of the gallbladder: results of the French Surgical Association survey. *Annals of Surgery* 1994; **219**: 275–80. [Large multiinstitutional French experience with gallbladder cancer.]

de Aretxabala X *et al.* Gallbladder cancer in Chile: a report on 43 potentially resectable tumors. *Cancer* 1992; **69**: 60–5. [Review of surgical management of gallbladder cancer from a high incidence area.]

Ede RJ, Williams SJ, Hatfield ARW, McIntyre S, Mair G. Endoscopic management of inoperable cholangiocarcinoma using iridium-192. *British Journal of Surgery* 1989; **76**: 867–71. [Small experience with endoscopic brachytherapy in unresectable cholangiocarcinoma.]

Falkson G, MacIntyre JM, Moertel CG. Eastern Cooperative Oncology Group experience with chemotherapy for inoperable gallbladder and bile duct cancer. *Cancer* 1984; **54**: 965–9. [Early experience demonstrating poor response rates and survival with 5-FU alone in inoperable biliary tract cancer.]

Gall FP, Kockerling F, Scheele J, Schneider C, Hohenberger W. Radical operations for carcinoma of the gallbladder: present status in Germany. *World Journal of Surgery* 1991; **15**: 328–36. [Retrospective review of 113 patients with gallbladder cancer managed at a single German center.]

Gebbia V *et al.* Treatment of advanced adenocarcinomas of the exocrine pancreas and the gallbladder with 5-fluorouracil, high dose levofolinic acid and oral hydroxyurea on a weekly schedule. *Cancer* 1996; **78**: 1300–7. [Phase II trial reporting relatively high response rate of a 5-FU based chemotherapy regimen in gallbladder cancer.]

Klempnauer J *et al.* Resectional surgery of hilar cholangiocarcinoma: a multivariate analysis of prognostic factors. *Journal of Clinical Oncology* 1997; **15**: 947–54. [Large series of patients with perihilar cholangiocarcinoma illustrating significant clinical prognostic factors.]

Koyama K, Tanaka J, Kato S, Asanuma Y. New strategy for treatment of carcinoma of the hilar bile duct. *Surgery, Gynecology, Obstetrics* 1989; **168**: 523–9. [Initial report of a novel adjuvant chemoradiation regimen.]

Kraybill WG *et al.* Multidisciplinary treatment of biliary tract cancers. *Journal of Surgical Oncology* 1994; **55**: 239–45. [Retrospective analysis of 96 patients with biliary tract cancer demonstrating better survival in patients receiving greater than 4000 cGy radiation.]

Lai EC, Chu KM, Lo CY, Fan St, Lo CM, Wong J. Choice of palliation for malignant hilar obstruction. *American Journal of Surgery* 1992; **163**: 208–12. [Retrospective review of nonoperative stenting and surgical bypass for palliation of malignant jaundice.]

Langer JC, Langer B, Taylor BR, Zeldin R, Cummings B. Carcinoma of the extrahepatic bile ducts: results of an aggressive surgical approach. *Surgery* 1985; **98**: 752–9. [Early series demonstrating safety and prolonged survival following resection of bile duct cancer.]

Lillemoe KD, Pitt HA, Kaufmann SL, Cameron JL. Acute cholecystitis occurring as a complication of percutaneous transhepatic drainage. *Surgery, Gynecology, Obstetrics* 1989; **168**: 348–52. [Small series reporting frequent occurrence of acute cholecystitis in patients with an indwelling percutaneous transhepatic stent.]

Lipsett PA, Pitt HA, Colombani PM, Boitnott JK, Cameron JL. Choledochal cyst disease: a changing pattern of presentation. *Annals of Surgery* 1994; **220**: 644–52. [Large single center review of choledochal cysts primarily in an adult population.]

Lomanto D *et al.* Magnetic resonance cholangiopancreatography in the diagnosis of biliopancreatic disease. *American Journal of Surgery* 1997; **174**: 33–8. [Review of new technology for diagnosing biliary obstruction.]

Lowenfels AB, Lindstrom CG, Conway MJ, Hastings PR. Gallstones and risk of gallbladder cancer. *Journal of the National Cancer Institute* 1985; **75**: 77–80. [Case control study demonstrating association of race, sex, and gallstones with gallbladder cancer.]

Nakeeb A *et al.* Cholangiocarcinoma: a spectrum of intrahepatic, perihilar, and distal tumors. *Annals of Surgery* 1996; **224**: 463–75. [Large, single institution experience with cholangiocarcinoma introducing a new anatomic classification.]

Nervi F *et al.* Frequency of gallbladder cancer in Chile: a high-risk area. *International Journal of Cancer* 1988; **41**: 657–60. [Autopsy study demonstrating the importance of gallstones and female sex in gallbladder cancer from a high-incidence area.]

Nordback IH *et al.* Unresectable hilar cholangiocarcinoma: percutaneous versus operative palliation. *Surgery* 1994; **115**: 597–603. [Single institution, nonrandomized review demonstrating modest improvement in quality of life in patients undergoing operative pallation of unresectable cholangiocarcinoma.]

Ogura Y, Mizumoto R, Isaji S, Kusuda T, Matsuda S, Tabata M. Radical operations for carcinoma of the gallbladder: present status in Japan. *World Journal of Surgery* 1991; **15**: 337–43. [Multi-institutional review of 1686 patients with gallbladder cancer managed in Japan.]

Ouchi K, Suzuki M, Tominaga T, Saijo S, Matsuno S. Survival after surgery for cancer of the gallbladder. *British Journal of Surgery* 1994;**81**: 1655–7. [Analysis of prognostic factors in resectable gallbladder cancer.]

Piehler JM, Crichlow RW. Primary carcinoma of the gallbladder. *Surgery, Gynecology, Obstetrics* 1978; **147**: 929–39. [Review article highlighting clinical signs and symptoms in over 2000 patients with gallbladder cancer.]

Pitt HA. Curative treatment for pancreatic neoplasms: standard resection. *Surgical Clinics of North America* 1995; **75**: 891. [Controlled, single institution experience demonstrating no survival benefit of postoperative radiotherapy in cholangiocarcinoma.]

Pitt, HA, Dooley WC, Yeo CJ, Cameron JL. Malignancies of the biliary tree. *Current Problems in Surgery* 1995; **32**: 1–90. [Recent review of the management of biliary tract cancer.]

Pitt HA *et al.* Perihilar cholangiocarcinoma: postoperative radiotherapy does not improve survival. *Annals of Surgery* 1995; **221**: 788–98. [Review of pathological and functional data supporting the use of pylorus preserving pancreatoduodenectomy for patients with pancreatic cancer.]

Pitt HA, Grochow LB, Abrams RA. Cancer of the biliary tree. In: DeVita VT, Hellman S, Rosenberg SA, ed. *Cancer: principles and practice of oncology*, pp. 1114–28. Lippincott, Philadelphia, 1997. [Detailed review of etiology, diagnosis, and management of gallbladder and bile duct cancer.]

Ransohoff DF, Gracie WA. Treatment of gallstones. *Annals of Internal Medicine* 1993; **119**: 606–19. [Use of mathematical models to predict gallstone-related complications such as gallbladder cancer in asymptomatic and symptomatic patients.]

Robertson JM *et al.* Treatment of primary hepatobiliary cancer with conformal radiation therapy and regional chemotherapy. *Journal of Clinical Oncology* 1993; **11**: 1286–92. [Uncontrolled series of nine patients with cholangiocarcinoma receiving conformal radiotherapy with intra-arterial chemotherapy.]

Sheen-Chen SM, Chou FF, Eng HL. Intrahepatic cholangiocarcinoma in hepatolithiasis: a frequently overlooked disease. *Journal of Surgical Oncology* 1991; **47**: 131–5. [Study demonstrating the high incidence of cholangiocarcinoma in patients with intrahepatic stones.]

Shirai Y, Tsukada K, Ohtani T, Watanabe H, Hatakeyama K. Hepatic metastases from carcinoma of the gallbladder. *Cancer* 1995; **75**: 2063–8. [Pathological evaluation of the pattern of hepatic metastases from gallbladder cancer.]

Shirai Y, Yoshida K, Tsukada K, Muto T. Inapparent carcinoma of the gallbladder: an appraisal of a radical second operation after simple cholecystectomy. *Annals of Surgery* 1992; **215**: 326–31. [Study reporting excellent survival in patients with stage I gallbladder cancer.]

Shirai Y, Yoshida K, Tsukada K, Ohtani T, Muto T. Identification of the regional lymphatic system of the gallbladder by vital staining. *British Journal of Surgery* 1992; **79**: 659–62. [Mapping of the lymphatic drainage of the gallbladder in humans.]

Soyer P *et al.* Carcinoma of the gallbladder: imaging features with surgical correlation. *American Journal of Roentgenology* 1997; **169**: 781–5. [Pictorial review of imaging techniques used in

diagnosing gallbladder cancer.]

Sugiura Y *et al.* Extensive resection of the bile ducts combined with liver resection for cancer of the main hepatic duct junction: a cooperative study of the Keio Bile Duct Cancer Study Group. *Surgery* 1994; **115**: 445–51. [Multicenter study illustrating the importance of negative margins in combined bile duct and liver resections for perihilar cholangiocarcinoma.]

Todoroki T. Radiation therapy for primary gallbladder cancer. *Hepatogastroenterology* 1997; **44**: 1229–39. [Review article summarizing the rationale for radiation therapy in gallbladder cancer.]

Tsukada K *et al.* Outcome of radical surgery for carcinoma of the gallbladder according to the TNM stage. *Surgery* 1996; **120**: 816–22. [Single institution experience demonstrating importance of achieving negative surgical margins in resectional surgery for gallbladder cancer.]

White K, Kraybill WG, Lopez MJ. Primary carcinoma of the gallbladder: TNM staging and prognosis. *Journal of Surgical Oncology* 1988; **39**: 251–5. [Retrospective review demonstrating significance of TNM stage in prognosis.]

Whittington R, Neuberg D, Tester WJ, Benson AB, Haller DG. Protracted intravenous fluorouracil infusion with radiation therapy in the management of localized pancreaticobiliary carcinoma: a phase I Eastern Cooperative Oncology Group Trial. *Journal of Clinical Oncology* 1995; **13**: 227–34. [Phase I trial of chemoradiation in biliary tract malignancy.]

Wibbenmeyer LA, Wade TP, Chen RC, Meyer RC, Turgeon RP, Andrus CH. Laparoscopic cholecystectomy can disseminate *in situ* carcinoma of the gallbladder. *Journal of the American College of Surgeons* 1995; **181**: 504–10. [Review of patients with gallbladder cancer detected at laparoscopic cholecystectomy illustrating the hazards of spilling bile during the procedure.]

32.1 Aetiology, presentation, and investigation

George Hamilton

[Aetiology](#)
[Anatomical and physiological considerations](#)
[Haemodynamic considerations](#)
[The ‘backward' and ‘forward' theories of portal hypertension](#)
[Hyperdynamic circulation of portal hypertension](#)
[Plasma volume expansion](#)
[Epidemiology and natural history of portal hypertension](#)
[Oesophageal varices](#)
[Rupture of oesophageal varices](#)
[Presentation of portal hypertension](#)
[History and clinical findings](#)
[Investigation](#)
[Endoscopy](#)
[Imaging modalities in portal hypertension](#)
[Ultrasound and duplex scanning](#)
[Magnetic resonance imaging](#)
[Computed axial tomography (CT scan)](#)
[Diagnostic angiography](#)
[Further reading](#)

Aetiology

Many diseases of the liver or associated structures obstruct portal blood flow and result in portal hypertension. Such obstruction is accompanied by increased mesenteric blood flow, the so-called hyperaemia of portal hypertension. These phenomena are generally considered to be the two principal factors in the pathophysiology of portal hypertension.

Obstruction may occur in the hepatic veins, within the liver parenchyma, or in the portal vein itself. The most common cause of portal hypertension in Europe and North America is cirrhosis, which increases intrahepatic vascular resistance by fibrosis, thrombosis, and nodular regeneration. Portal vein obstruction by thrombosis or extrinsic compression is the second most common cause. Hepatic vein occlusion is the third most common but much less frequent cause.

A more precise classification of portal hypertension is by site of obstruction at the presinusoidal, sinusoidal, and post-sinusoidal levels (Table 1). Intrinsic liver disease is usually absent in patients with presinusoidal obstruction: the natural history and prognosis of portal hypertension is therefore relatively good. Post-sinusoidal obstruction can result in hepatocellular damage as a secondary congestive phenomenon. In the acute phase of the illness this may result in death from liver failure rather than from variceal haemorrhage. In patients with sinusoidal block such as occurs in cirrhosis, the degree of hepatocellular damage is the most important factor in the natural history of the disease. Death may result not only from variceal haemorrhage but also from liver failure, malnutrition, or infection.



Table 1 Classification of portal hypertension

Irrespective of cause, portal hypertension results in the development of collateral channels between the portal and systemic venous circulations, of which the most important clinically are those that develop in the oesophagus. These varices are the source of spectacular haemorrhage and thus from a clinical viewpoint, the most important complication of portal hypertension.

Anatomical and physiological considerations

The liver has a dual supply from both the hepatic artery and the portal vein. The superior mesenteric and splenic veins drain the splanchnic and splenic beds and join to form the portal vein. Entering into the origin of the portal vein is the coronary or left gastric vein which drains the lesser curve of the stomach and the lower oesophagus. The inferior mesenteric vein drains the left hemicolon and most of the rectum, and joins the splenic vein just before its confluence with the superior mesenteric vein (Fig. 1). After division into left and right branches, the hepatic artery parallels the portal system, and subsequently divides into branches corresponding to the segmental anatomy of the liver.

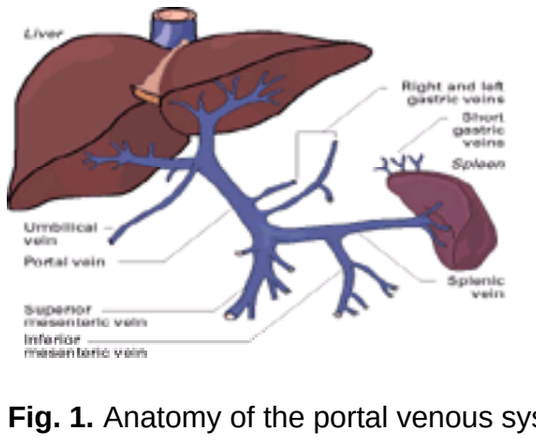


Fig. 1. Anatomy of the portal venous system.

The portal venous system is entirely devoid of valves. Numerous small tributaries connect the portal and systemic venous systems, and these can evolve into major collateral channels when portal hypertension supervenes. Formation of such collaterals is triggered when portal pressure rises above the normal level of 5 to 10 mmHg. This absence of valves is a feature of practical importance, in that it allows portal venous pressure to be measured during surgical procedures by cannulation of any small mesenteric or omental vein.

The most important of these portasystemic channels are the left gastric or coronary vein, which connects the oesophagocardiac venous plexus with the splenic or

portal vein; the short gastric and left gastroepiploic veins, which connect the oesophageal and gastric plexus with the splenic vein; numerous retroperitoneal portal radicles, which connect to the left renal vein via the left adrenal vein; the umbilical and periumbilical veins connecting to the left portal vein; and the inferior mesenteric vein connecting via the superior haemorrhoidal vein to the middle and inferior haemorrhoidal veins of the systemic circulation. The last are often responsible for the formation of large hypertensive haemorrhoids. Collaterals may also form across the diaphragm, within adhesions from previous abdominal surgery or from inflammatory bowel disease, or at mucocutaneous junctions, such as those that occur at ileostomies or colostomies (Fig. 2). In addition to these channels, intrahepatic shunts develop through which a significant proportion of portal venous flow can pass. These routes allow up to 80 per cent of the liver's blood supply to bypass the sinusoidal circulation.

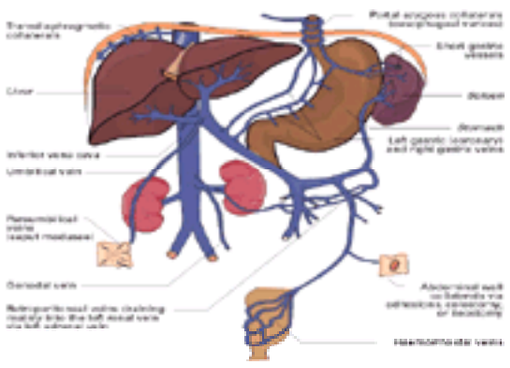


Fig. 2. Sites of portasystemic collateral circulation in portal hypertension.

The portal vein normally carries 75 per cent of the blood supply of the liver, with an average flow of 1200 ml/min. The hepatic artery supplies the remainder, at an average flow of 400 ml/min; it also provides 30 to 40 per cent of the liver's normal oxygen requirement since portal blood is already partly deoxygenated. When flow to the sinusoids is significantly reduced, a compensatory increase in hepatic arterial inflow takes place to provide the liver's oxygen requirements. However, intrahepatic shunting may cause up to 30 per cent of hepatic arterial flow to bypass the sinusoidal bed. Despite the compensatory increase in arterial flow, therefore, an overall significant reduction in hepatic tissue perfusion is highly likely to occur.

Haemodynamic considerations

An appreciation of the basic haemodynamic principles governing flow within blood vessels allows a better understanding of the pathophysiology of portal hypertension. As in all blood vessels, pressure within the portal system is determined by the interaction of flow and vascular resistance. As these parameters change, so does portal pressure. This relationship is expressed by Ohm's law,

$$D_p = Q \times R$$

where D_p is change in pressure, Q is flow, and R is resistance in a vessel.

Resistance is determined by several factors as expressed in Poiseuille's law,

$$R = 8nL/\pi r^4$$

where n is the coefficient of viscosity, L the vessel length, and r the vessel radius. Within the liver, viscosity and length of vessel are relatively constant and thus changes in vascular resistance are mainly due to changes in radius. Because the relationship is to the fourth power of the radius, small changes in portal vessel diameter will have profound effects on vascular resistance. By substitution of the equation for resistance into Ohm's equation, it can also be readily seen that pressure changes will also depend on changes in flow.

The normal liver has a very low intrahepatic resistance which can decrease with increased blood flow due to vascular dilatation. This inherent property of an extremely compliant hepatic circulation can maintain normal portal pressure within a wide range of portal flow. This capacity is lost in liver disease with dramatic increases in intrahepatic resistance. As well as structural disturbances associated with cirrhosis, inflammatory changes in the hepatic venous tree and deposition of fibrous tissue around the terminal hepatic venules and adjacent sinusoids have been described. In cirrhotic livers, intrahepatic resistance may also be affected by proliferation of myofibroblasts around the sinusoids and terminal hepatic venules, resulting in increased contractility which may raise resistance and contribute to portal hypertension. These myofibroblasts are known as hepatic stellate cells. These cells are star shaped with many branching arms and acquire contractile properties behaving in a similar manner to vascular pericytes. Thus in response to vasoactive substances such as endothelin, these cells by their action on the sinusoids at the microcirculatory level can increase resistance to portal blood flow. Sinusoids may be compressed by hepatocyte enlargement in regenerating parts of the liver. This can occur as a result of several toxic, infectious, or metabolic insults to the parenchyma and may explain the portal hypertension seen in non-cirrhotic conditions such as alcoholic hepatitis.

The ‘backward' and ‘forward' theories of portal hypertension

Despite the decompressive effects of collaterals and intrahepatic shunts, with up to 80 per cent of the portal blood flow bypassing the liver, portal pressure remains elevated. Two major hypotheses have been advanced to explain this phenomenon.

The ‘backward' theory postulates that portal hypertension is due to increased hepatic vascular resistance which develops as a specific response, so that in the presence of normal flow, pressure must increase. Hypertrophy of myofibroblasts within the sinusoidal bed is one possible mechanism by which this could be achieved. Ohm's law suggests that portal hypertension would develop only if normal liver blood flow was maintained, but logically, in the presence of extensive low-resistance collaterals, blood flow should be diverted away and thus portal pressure should decrease. In reality, however, the portal and splanchnic circulation is not only markedly increased but hyperdynamic, thus negating the likelihood that this theory can explain portal hypertension.

The ‘forward' theory was initially postulated by Banti in 1883. He suggested that splenomegaly, portal hypertension, and cirrhosis were the result of increased splanchnic arterial blood flow. The subsequent confirmation of increased splanchnic blood flow as a major feature of portal hypertension has led to further development of this theory. This now suggests that increased and hyperdynamic flow, together with decreased splanchnic precapillary resistance, maintain portal hypertension even in the face of extensive portasystemic shunting. Since Banti first postulated his theory, opinions on the relative contribution of increased splanchnic blood flow to the development of portal hypertension have varied. More recently Groszmann and other workers have increasingly implicated both theories as being involved in the pathophysiology of this condition.

It is currently believed that the principal and initial abnormality is increased vascular resistance to portal flow and that portal hypertension is then maintained by increased blood flow into the portal circulation, a phenomenon which has been confirmed conclusively both experimentally and clinically. Blood flow to the stomach, spleen, and intestines is increased by 50 per cent in portal hypertension: this is achieved largely by splanchnic vasodilatation and raised cardiac output.

Hyperdynamic circulation of portal hypertension

A hyperdynamic systemic circulation is frequently found in patients with portal hypertension. The intensity of this phenomenon was previously related to the degree of underlying hepatocellular dysfunction. However, this state is also seen in patients with extrahepatic causes of portal hypertension, in whom hepatic function is usually normal. It is portasystemic shunting and peripheral vasodilatation which are the major determinants in the pathogenesis of this clinical feature of portal hypertension. The hyperdynamic circulation is characterized by decreased systemic vascular resistance, decreased mean arterial pressure, increased splanchnic blood flow, increased cardiac index, and plasma volume expansion.

Several substances have been implicated as hormonal factors which act on both the splanchnic and systemic circulations to produce hyperaemia. Three main mechanisms are involved, namely increased levels of circulating vasodilators, increased endothelial production of local vasodilators, and decreased vascular responsiveness to endogenous vasoconstrictors. The most important of these are bile acids, serotonin, glucagon, and prostaglandins, in particular prostacyclin, and

nitric oxide.

Elevated levels of bile acids have been demonstrated in patients with liver disease and implicated as promoters of splanchnic hyperaemia. In the experimental situation however, lowering of serum bile acid levels to normal has no effect on the increased splanchnic blood flow, and their role remains uncertain. Serotonin has a powerful vasoconstrictor effect on all vascular smooth muscle and is normally released into the portal circulation by the enterochromaffin cells of the gastrointestinal tract. In experimental portal hypertension, selective blockade of serotonin receptors reduces portal pressure, thus supporting the role of serotonergic mechanisms in the pathogenesis of portal hypertension. Glucagon, a potent splanchnic vasodilator, is elevated in patients with portal hypertension, and probably accounts for 30 per cent of the increased splanchnic blood flow. The effect of somatostatin reducing portal pressure and blood flow may in part result from its inhibitory effect on the release of glucagon, in addition to its direct vasoactive properties.

The endothelium is of central importance in modulation of basic vascular tone and has been implicated as an important factor in the vasodilatation of portal hypertension. The two major substances implicated and produced by the endothelium are prostacyclin and nitric oxide.

Prostacyclin is a naturally occurring powerful vasodilator, production of which is elevated in the portal vein endothelium of rats with portal hypertension. Its production is directly related to the portal pressure. Prostacyclin production is also elevated in experimental arterial hypertension: this may represent a compensatory physiological response of endothelial cells to reduce hypertension. Patients with cirrhosis have extremely high levels of urinary 6-ketoprostaglandin F_{1a}, a stable metabolite of prostacyclin, and pharmacological inhibition of prostacyclin production reduces portal pressure in these patients.

Nitric oxide, or endothelial-derived relaxing factor, is released from l-arginine by the action of nitric oxide synthase and causes potent vasodilatation by its action on smooth muscle cells. Experimental evidence, mainly in the rat model of portal hypertension using partial portal vein ligation, suggests that nitric oxide is important in the development of portal hypertension. Increased levels of constitutive and inducible nitric oxide synthase, secondary to the raised levels of endotoxins and cytokines found in cirrhosis, are proposed as a mechanism resulting in production of nitric oxide and thus vasodilatation. However, this mechanism remains unproved in patients with cirrhosis, with contradictory results on the circulation found when inhibitors of nitric oxide synthase were administered. Much research continues into the role of nitric oxide which, although clearly important, is only one of several factors implicated.

There is experimental evidence for a decreased responsiveness to vasoconstrictor substances such as endothelin, noradrenaline, angiotensin II, and vasopressin. Thus hyporeactivity to these vasopressors may have a role in altering the balance between endogenous vasodilators and vasoconstrictors and thus contributing to vasodilatation and the production of a hyperdynamic circulation.

Thus several important vasoactive substances are delivered directly into the systemic circulation via portasystemic collaterals which carry 80 per cent of the portal blood flow. These substances will remain vasoactive since they have avoided deactivation by the hepatocytes.

Plasma volume expansion

Plasma volume expansion seems to follow directly from vasodilatation in the development of the hyperdynamic circulation. The proposed mechanism is one of response to the hypovolaemia induced by the dilatation of the systemic and splanchnic circulations. This results in a decreased central blood volume triggering the renin–angiotensin system, release of catecholamines, and antidiuretic hormone. The resultant sodium and water retention causes increased plasma volume which together with peripheral vasodilatation results in the hyperdynamic circulation of portal hypertension (Fig. 3).

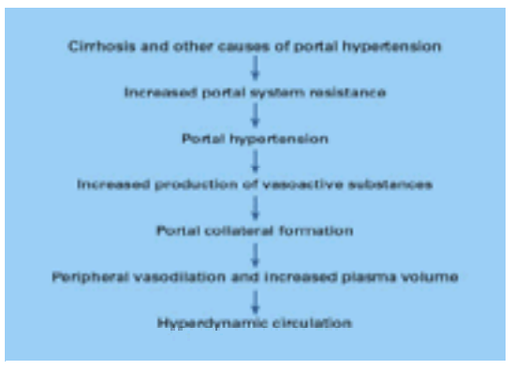


Fig. 3. Hyperdynamic circulation of portal hypertension.

Epidemiology and natural history of portal hypertension

Cirrhosis of the liver accounts for approximately 90 per cent of all cases of portal hypertension presenting in the West. In Eastern and tropical countries non-cirrhotic causes predominate. Distinct geographical distributions are found for non-cirrhotic portal fibrosis, schistosomiasis, and idiopathic portal hypertension.

In India, idiopathic portal fibrosis causes 20 to 30 per cent of all cases of portal hypertension; a further 20 to 30 per cent are due to portal vein thrombosis secondary to dehydration and portal infection in the neonatal period. In Japan approximately 10 per cent of cases of portal hypertension are idiopathic, as opposed to 1 to 6 per cent in the West. This condition remains most prevalent in the tropics.

Schistosomiasis affects more than 200 million people worldwide: *Schistosoma mansoni* is particularly prevalent in the Middle East, Africa, and South America, while *S. japonicum* is the causative agent in the Far East. Hepatic schistosomiasis is particularly common in Egypt, where oesophageal variceal bleeding secondary to this disease is the most common cause of upper gastrointestinal haemorrhage.

In children, extrahepatic portal vein obstruction is the main cause of upper gastrointestinal bleeding: this accounts for 40 to 50 per cent of all cases of portal hypertension in those under 17 years old. It is much more common in developing countries.

The natural history and prognosis of portal hypertension depends primarily on the underlying hepatic functional reserve. Conditions causing presinusoidal block but in which hepatocellular function remains good carry the best prognosis: the major lethal complication is haemorrhage from oesophageal varices. The mortality for each bleeding episode is about 5 per cent, and where facilities exist for prompt blood transfusion and injection sclerotherapy the long-term outlook is good. In post-sinusoidal block (Budd–Chiari syndrome or veno-occlusive disease) severe secondary hepatocellular damage in the acute phase can result in death from liver failure and massive ascites rather than from haemorrhage.

In patients with cirrhosis the overall mortality rate from oesophageal variceal bleeding is about 40 per cent. If the patient recovers from the acute bleed the risk of recurrent haemorrhage during the same hospital stay is 60 per cent; this increases to more than 80 per cent at 2 years. The long-term survival of patients with cirrhosis following variceal bleeding is poor, ranging from 6 to 35 per cent at 5 years. It is important to emphasize, however, that only about 30 per cent of patients with cirrhosis will ever experience such bleeding. The remainder die of liver failure, cachexia, and infection.

The outlook in cirrhosis is generally determined by the severity of associated hepatocellular disease. In 1964 Child introduced a classification of severity of liver disease based on clinical findings and liver function tests. Classification into grades A, B, or C (Table 2) allows long-term survival and operative risk to be predicted for each patient. Groups A (good liver function) and B (moderate impairment of liver function) have a 70 to 80 per cent survival rate at 1 year, while only 30 per cent of patients in group C (poor liver function) survive for 1 year. Child originally reported operative mortality rates of 5 per cent in groups A and B, and of 53 per cent in patients of group C undergoing portacaval shunting procedures. Using this classification, similar predictive patterns of mortality have been reported consistently throughout the literature. Patients with alcoholic liver disease or chronic active hepatitis also have particularly poor prognoses; patients with primary biliary cirrhosis generally have a better outlook.

Serum bilirubin (nmol/l)	< 34	35–51	> 68
(mg per cent)	< 2	2–3	> 3
Albumin (g/l)	35	28–35	< 28
Prothrombin time (s prolonged)	1–4	4–6	> 6
Ascites	Absent	Slight	Moderate
Encephalopathy	None	Minimal	Coma
Child's grade	A	B	C
points scored	5–6	7–9	10–15

Clinical and laboratory data allow long-term prognosis and interventional or operative risk to be assessed. This also provides the best method of comparison of different forms of treatment in patients with portal hypertension.

Table 2 Child's classification of patients with chronic liver disease

Acute hepatic decompensation may occur during a bleeding episode. This may result in deterioration such that a patient initially in a good prognostic group enters a lower Child's grade. This phenomenon is of crucial importance in timing of intervention and most particularly, in analysis by Child's grading of the efficacy of specific interventional procedures.

Oesophageal varices

Rupture of oesophageal varices (and, less commonly, of gastric varices) is the most dramatic complication of portal hypertension and that which is most likely to require surgical management. Elegant anatomical studies have revealed in detail the complex and peculiar portasystemic connection which occurs in the distal 2 to 5 cm of the oesophagus, precisely the zone where varices develop (Fig. 4). Four distinct layers of veins have been found in this zone, extending from the luminal surface to the adventitial layer. Intraepithelial veins or vascular epithelial channels drain into a superficial venous plexus, which lies just below the oesophageal epithelium. This plexus is in turn connected to a layer of deep intrinsic veins just outside the muscularis mucosae. The deep intrinsic veins connect with the outermost, external venous plexus coursing within the oesophageal adventitia via the perforating veins, which traverse the muscular layers of the oesophagus. Thus in this distal portion of the oesophagus, venous channels are concentrated largely in the mucosa and run in longitudinal palisades. These connect with the more deeply located submucous plexus in the cephalad oesophagus and stomach. This arrangement of veins probably underlies one of the intrinsic mechanisms of the physiological oesophageal sphincter, the so-called mucosal rosette, which is of importance in prevention of reflux.

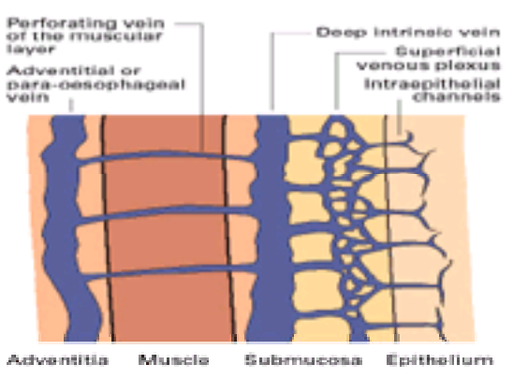


Fig. 4. Venous drainage of the distal oesophagus (after Kitano, 1986). Schematic, sectional representation of normal oesophageal wall displaying the venous anatomy.

In patients with portal hypertension this venous complex, situated at the watershed between the portal and systemic circulations, dilates dramatically and forms varices, classically running in three to five distinct trunks. The intraepithelial channels probably form the cherry red spots which can be seen endoscopically and are recognized as predictors of impending rupture. They are also found in histological specimens of oesophageal transection rings.

Studies using the oesophageal Doppler ultrasound probe have shown that blood flow and velocity within the palisade zone is complex and considerable. While flow is mainly cephalad, as would be expected, bidirectional flow has also been demonstrated, particularly where perforating veins from the deep adventitial plexus enter the varices. Turbulence of flow has also been demonstrated in these perforating veins. The palisade and perforating venous systems are therefore critical to the development of varices. Less severe variceal bleeds probably result from rupture of the intraepithelial channels, while torrential haemorrhage results from rupture of the deeper, high-flow intrinsic venous channels.

Gastric varices are a much less common source of haemorrhage, probably because of their deeper submucosal situation. These varices are mainly supplied by the short gastric veins and course into the deep intrinsic veins of the distal oesophagus. Congestive gastropathy, which is common in portal hypertension, is associated with mucosal congestion, resulting from increased submucosal arteriovenous communications which develop between the muscularis mucosae and dilated arterioles and venules. Such congested mucosa is particularly susceptible to gastritis and haemorrhage.

Rupture of oesophageal varices

Significant rupture occurs in only 30 per cent of patients, and the causes remain poorly understood. Two main theories for variceal rupture are commonly advanced. The erosive theory postulates that mucosal damage due to reflux oesophagitis causes erosion into the varix and thus haemorrhage. Reflux oesophagitis is not common, however, either at endoscopy or at operative inspection of the distal oesophagus. Inflammation is rarely found in specimens of oesophageal rings removed at transection, and controlled trials have found that cimetidine has no advantage over placebo in preventing recurrent variceal bleeding. Erosion of the oesophageal mucosa as a cause of rupture therefore seems unlikely.

The eruptive theory proposes that variceal rupture is related to pressure where the varix is in close proximity to the oesophageal lumen: thus it commonly occurs in the distal 2 to 5 cm of the oesophagus. Although increased portal pressure (greater than 12 mmHg) is required for variceal development, a direct correlation between portal pressure and risk of variceal bleeding has not been found. There is more circumspect evidence to support a relationship between the severity of portal hypertension and risk of variceal haemorrhage. Portal pressure in alcoholic patients following a variceal bleed was higher in those who did not survive. In addition, angiography and blood volume expansion, both of which may increase portal pressure, occasionally precipitate haemorrhage. Variceal haemorrhage has also been reported to occur after deep inspiration, coughing, and following the Valsalva manoeuvre, all of which cause an increased pressure differential between oesophageal varix and lumen. Local factors such as variceal wall thickness and that of the overlying mucosa are obviously also of importance.

Variceal size has been reported to be a risk factor for haemorrhage, the larger varices being more likely to bleed. The lack of relationship between pressure and variceal size reinforces the importance of local factors in varix development, however.

Groszmann has proposed that variceal wall tension may be the unifying predictor of rupture. Tension in a variceal wall can be derived from a modification of Laplace's law

$$T = TP \times r/w$$

where T is tension, TP is transluminal pressure, r is the vessel radius, and w is the wall thickness.

In oesophageal varices, transmural tension is the difference in pressure between the oesophageal lumen and that of the varix. Thus tension is directly related not only to this pressure difference, but also to the varix radius, and is inversely related to variceal wall thickness. These three variables correspond in the clinical situation to variceal size, thickness of epithelium overlying the varix, and degree of portal hypertension, all of which have been described as independent predictors of bleeding. When portal pressure is increased the more superficial protruding varix will have less supporting connective tissue and a larger radius than a deeper varix embedded in

supporting tissue which is under identical pressure. Wall tension will therefore be higher in these superficial varices, and rupture will result when this tension is no longer in equilibrium with the outwardly directed expansile force of portal pressure. When the elastic limit of the vessel has been reached, small changes in volume or radius will result in large changes in wall tension and imminent rupture (Fig. 5).

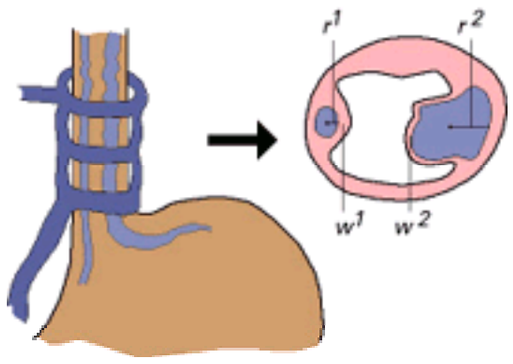


Fig. 5. Variceal wall tension (after Polio and Groszmann, 1986). Since $T = TP \times r/w$, variceal wall tension (T) is greater where radius (r) is larger and the wall (w) is thinner despite equal transmural pressure (TP); T_1 is greater than T_2.

The parameters considered in the above equation cannot be measured clinically. However, endoscopic classification of varices according to size and presence of cherry red spots and red wall markings, both of which are related to thinness of the overlying epithelium, has markedly increased the clinician's ability to predict impending variceal haemorrhage.

Reduction of variceal pressure by shunting, disconnection, or by pharmacological means decreases transluminal pressure, vessel radius, and wall tension. Sclerotherapy, by causing dense perivariceal scarring, will increase the surrounding supporting tissue, strengthening the variceal wall and thus decreasing tension.

Once rupture has occurred the severity of haemorrhage will be affected by haemodynamic factors and by the disordered haemostasis that frequently accompanies liver disease. The relevant haemodynamic factors have been expressed by Groszmann in the following equation.

severity of bleeding = area of rent/blood viscosity

It follows that large holes under greater pressure will bleed more severely than small holes under lower pressure. Blood viscosity is directly related to the haematocrit: reduction of haematocrit after haemorrhage, in anaemia, and after volume replacement with crystalloids may increase the severity of bleeding.

Patients with liver disease may show multiple disorders of haemostasis, including deficient production by the liver of the coagulation proteins, thrombocytopenia secondary to hypersplenism, impaired platelet function, and increased fibrinolytic activity. The relative importance of all these factors discussed in the clinical situation remains unclear. Haemorrhage generally responds most readily to transfusion of whole, preferably fresh, blood, and clotting factors such as fresh frozen plasma and platelets. Empirically, haemodynamic factors may be most important in the development of rupture, while coagulatory factors may be of greater importance once profuse haemorrhage is established.

Bacterial infection has recently been implicated in the pathogenesis of variceal bleeding. This hypothesis suggests that episodes of endotoxaemia or clinical sepsis, which frequently occur in patients with portal hypertension and cirrhosis, expose these patients to high levels of endotoxin. These induce production of endothelin and possibly of thromboxane, which cause hepatic stellate cell contraction and thus increased intrahepatic vascular resistance and increased portal pressure. Variceal haemorrhage would possibly be aggravated by the increased levels of platelet antiaggregatory agents such as prostacyclin and nitric oxide acting at the endothelial level at the site of rupture. This interesting hypothesis will be tested by a prospective randomized trial of antibiotic therapy.

Presentation of portal hypertension

Portal hypertension *per se* has little in the way of clinical features, and usually presents with complications such as variceal haemorrhage, ascites, or hypersplenism. Haemorrhage occurs in only 30 per cent of these patients: the remainder may present with a clinical history and findings indicative of liver disease, particularly cirrhosis. The clinical history is therefore of major importance when portal hypertension is suspected.

History and clinical findings

Careful questioning of the patient about alcoholism, past jaundice or hepatitis, exposure to hepatotoxins, and a past history of blood transfusion or drug abuse should be routine. A history of neonatal or intra-abdominal sepsis, or of a myeloproliferative disorder is suggestive of an extrahepatic portal vein block. A history of myeloproliferative disorder may indicate the Budd–Chiari syndrome. The use of oral contraceptives or other sex hormone-containing medications is also significant.

The most common clinical presentation of portal hypertension is profuse gastrointestinal haemorrhage: the severity, frequency, and dates of these episodes must be determined. Elucidation of symptoms of hepatic decompensation, in particular ascites and portosystemic encephalopathy as manifested by intellectual dulling, confusion, or coma, is important. Haematemesis is the most common presentation, but episodes of melaena without haematemesis may also occur. A history of dyspepsia or past peptic ulceration suggests bleeding from one of these sources and the results of any previous barium studies or endoscopies may be informative.

Careful physical examination may reveal the stigmata of liver disease, but in the absence of hepatocellular decompensation the single most important finding is an enlarged spleen. If the spleen cannot be felt, or if splenomegaly is not confirmed on imaging, portal hypertension is unlikely to be present. Although enlargement of the spleen is progressive, the degree of hypersplenism is not related to the severity of portal hypertension. Abdominal wall veins from portasystemic shunts may be visible, although this sign is uncommon. The caput medusae, in which several veins can be seen radiating out across the abdomen from the umbilicus, is the most striking sign. In patients with obstruction of the inferior vena cava in the Budd–Chiari syndrome or as a result of severe ascites, abnormal veins course upwards across the abdomen from the iliac fossas to cross the costal margin and drain into superior vena cava territory.

Rarely, a murmur or venous hum may be heard in the region of the xiphoid process, arising from turbulent flow in a large umbilical collateral. A systolic murmur overlying the liver may indicate alcoholic hepatitis or a primary hepatocarcinoma. The liver may be enlarged, particularly in alcoholic patients, or shrunken: the latter can be confirmed by careful percussion. A shrunken liver is particularly common in portal hypertension and is reported to be associated with higher portal pressures. The consistency of the liver should also be assessed: a firm nodular liver indicates cirrhosis, while a smooth soft liver is suggestive of extrahepatic portal obstruction.

Ascites usually develops in hepatocellular decompensation and is rarely due to portal hypertension alone. Peripheral oedema is also frequently present. Further signs of liver failure include palmar erythema, spider naevi, white nails, jaundice, and portasystemic encephalopathy. Rectal varices are common in portal hypertension but cannot be detected by digital examination alone. Bleeding from rectal varices is uncommon, rarely severe, and must obviously be distinguished from that associated with simple haemorrhoids.

Haemorrhage from oesophageal varices remains the most common presentation of portal hypertension: if these varices did not form, portal hypertension would be of virtually no clinical significance. Bleeding is typically dramatic and catastrophic, with massive haematemesis and circulatory collapse. Bleeding *per rectum* may also be profuse and fresh. A slow oozing of blood from varices may occasionally result in true melaena only. Variceal bleeding in patients with cirrhosis may cause marked deterioration of liver cell function. This, together with increased protein absorption from the blood-laden gut, may lead to portasystemic encephalopathy and coma.

Non-variceal gastrointestinal bleeding is also common in alcoholic patients who may suffer from peptic ulceration, gastric erosions, and the Mallory–Weiss syndrome.

Investigation

Clinical evaluation of the patient is of utmost importance, particularly when surgical intervention is being contemplated. Operative procedures in patients with compromised liver function carry significantly increased risk. Evaluation of the patient's hepatocellular reserve is therefore of crucial importance when selecting

treatment for bleeding varices. Such assessment may be best made in conjunction with a hepatologist or physician with a special interest in liver disease. Recommended laboratory investigations are listed in Table 3. Once the underlying liver disease has been diagnosed and liver function assessed, treatment should be selected on the basis of liver reserve, as classified by Child. The next stage in investigation is imaging of the varices and the portal circulation.

1.	Full blood count: anaemia, leucopenia, and thrombocytopenia in chronic liver disease and hypersplenism
2.	Liver function tests: indicators of primary cause of liver disease Serum albumin and prothrombin time are good measures of synthetic function γ-Glutamyl transpeptidase is raised in patients with excess alcohol intake
3.	Specific serum autoantibodies: markers of primary biliary cirrhosis and autoimmune chronic active hepatitis
4.	Serological markers of hepatitis B and C
5.	Tests for copper and iron metabolism (Wilson's disease and haemochromatosis)
6.	Serum α_1-antitrypsin levels, any patient with cirrhosis, particularly with a past medical history of neonatal jaundice
7.	Blood grouping

Table 3 General investigations of portal hypertension

Endoscopy

Endoscopy allows rapid and safe confirmation of the source of bleeding from the upper gastrointestinal tract. It is of vital importance in the exclusion of other causes of bleeding, such as peptic ulceration, gastritis, duodenitis, or an oesophageal tear. During active haemorrhage the source of bleeding may be obscured, particularly if it is from gastric varices or gastritis. Repeat endoscopy after 3 to 4 h will often find the stomach emptied of clot or blood, allowing visualization of the source. Once the diagnosis of variceal haemorrhage is confirmed, a skilled endoscopist can undertake immediate treatment by injection sclerotherapy.

Endoscopy also allows the size, distribution, and colour of varices to be determined. Varices are normally white and opaque: red coloration and the presence of cherry red spots are accurate predictors of variceal bleeding. Larger varices are also more likely to bleed. Endoscopic assessment of scarring or ulceration following injection sclerotherapy is of major importance, particularly when bleeding continues or recurs and procedures such as oesophageal transection are being considered. Surgical intervention should be avoided in patients with injection site ulceration: this is a frequent cause of haemorrhage which usually responds to medical treatment.

Imaging modalities in portal hypertension

Plain radiographs of the abdomen and chest are of limited value in the management of portal hypertension, but they may yield useful information. In an abdominal radiograph the size of the liver and spleen may be assessed and rarely, gas shadows in the portal circulation may be detected in patients with enterocolitis, intestinal infection, or disseminated intravascular coagulation syndromes (Fig. 6, Fig. 7).



Fig. 6. Abdominal radiograph demonstrating gross splenomegaly secondary to portal hypertension.



Fig. 7. Abdominal radiograph demonstrating calcified and thrombosed portal and splenic veins.

Widening of the left paravertebral shadow due to lateral displacement of the pleural reflection between aorta and vertebral column by a dilated hemiazygos vein may also be seen. Rarely, extensive para-oesophageal collaterals may appear as a posterior mediastinal mass (Fig. 8).



Fig. 8. Chest radiograph, lateral view: this shows a lobulated mass behind the heart due to para-oesophageal varices.

Where endoscopic facilities are not available, barium studies can be useful in demonstrating oesophageal and gastric varices (Fig. 9). The typical appearances are of a

dilated oesophagus containing multiple irregular filling defects, usually in the distal third but occasionally running throughout the entire oesophagus. Barium studies are also useful in diagnosing other causes of bleeding, such as peptic ulceration. For this reason a full upper gastrointestinal barium study is preferable to a barium swallow alone.

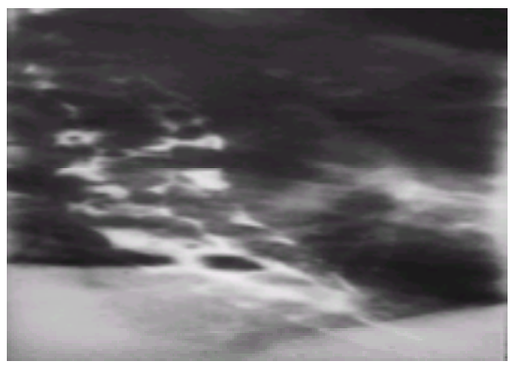


Fig. 9. Barium swallow. The characteristic sign of a string of beads in the oesophagus indicates oesophageal varices.

Once the diagnosis is made and operative intervention is being considered, the anatomical details of the portal circulation are needed. Classically this is obtained by angiography, but less invasive techniques such as duplex ultrasonography, contrast-enhancedCT scanning, and magnetic resonance imaging are being increasingly employed. Indeed in many centres angiography is now used rarely.

Ultrasound and duplex scanning

Real-time ultrasound scanning can provide much detailed and accurate information by totally non-invasive means. Because it is cheap, easy to perform, and well tolerated by the patient, duplex ultrasonography has become the primary investigation of choice in portal hypertension and indeed in all hepatobiliary disorders. It is of particular value for scanning the hepatic and splenic parenchyma for haemangioma, tumour, and cirrhosis. Thrombosis or occlusion of the portal, superior mesenteric, and splenic vein can be accurately detected. Enlargement of the portal vein (greater than 13 mm) and large collaterals can be visualized, helping to confirm the diagnosis of portal hypertension. Although the results are often compromised by duodenal or intestinal gas, its absolute safety is making ultrasound scanning the first line of investigation in portal hypertension. Recent advances in duplex ultrasound technology—in particular, colour-flow coding and power Doppler—allow faster and more accurate assessment of direction and flow in the portal circulation (Fig. 10). The accuracy is such that flow through portasystemic shunts can be measured, and duplex scanning is superseding angiography as the investigation of choice when graft occlusion is suspected. Duplex ultrasonography has become an important diagnostic and research tool with a promising future role in the investigation of portal hypertension.

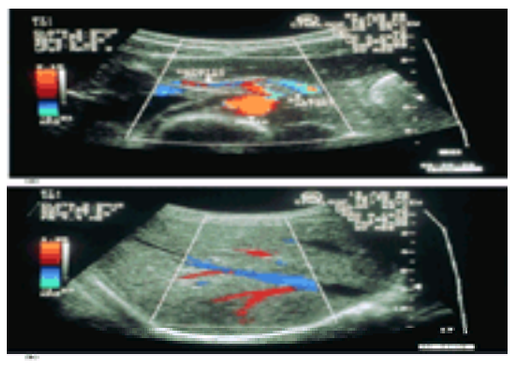


Fig. 10. (a) Duplex scans of the portal and hepatic vessels. The portal, splenic, and hepatic veins can be visualized by duplex scanning. Colour duplex, while not essential, facilitates imaging of deeper visceral vessels. Flow measurement and monitoring of blood flow direction can be routinely performed with this technique. Part (a) shows the structures at the coeliac axis. (b) Duplex scans of the portal and hepatic vessels. Part (b) elegant images of a normal hepatic vein and its branches.

Magnetic resonance imaging

This modality allows excellent parenchymal visualization in any plane, better blood vessel imaging than CT scanning, and is no longer limited by poor resolution and long imaging times. Recent advances in fast imaging (single breath hold), motion suppression techniques, and the development of MRI contrast agents have dramatically increased its value. The quality of imaging of the portal and hepatic vessels using gadolinium as a contrast agent is of such high quality that magnetic resonance angiography or MRA has superceded conventional angiography where MRI facilities are available (Fig. 11, Fig. 12 and Fig. 13).

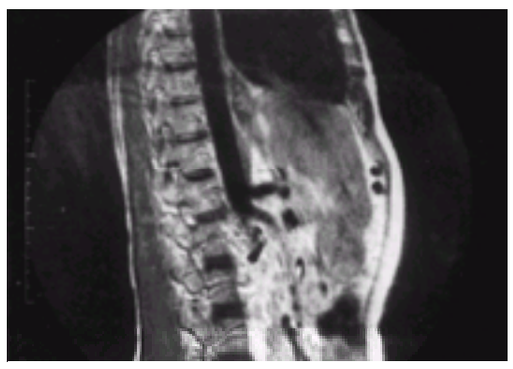


Fig. 11. Magnetic resonance imaging of the abdomen. The scan demonstrates the origins of the coeliac and superior mesenteric arteries. Note the large varices coursing along the anterior surface of the cirrhotic liver.

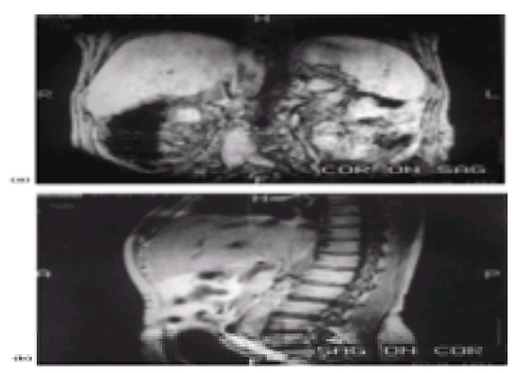


Fig. 12. (a and b) Magnetic resonance imaging, showing complete thrombosis of the hepatic and abdominal inferior vena cava of a patient with Budd–Chiari syndrome

secondary to myeloproliferative disease. Note the collaterals in the anterior abdominal wall and running along the aorta.

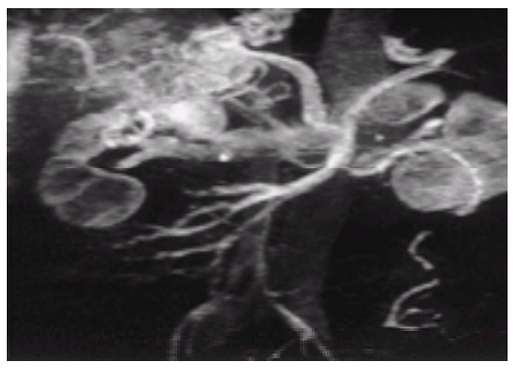


Fig. 13. Magnetic resonance angiogram using gadolinium contrast demonstrating the portal venous and systemic venous anatomy.

Computed axial tomography (CT scan)

The major advantage of CT scanning lies in the detection of focal liver disease and hepatosplenomegaly (Fig. 14). Used with contrast enhancement and dynamic spiral scanning, portal vein patency and para-oesophageal and retroperitoneal portasystemic collaterals can be demonstrated readily. Currently, this is rarely the diagnostic procedure of choice since ultrasound, and increasingly MRI, delineate these lesions more easily and accurately. CT assessment of liver parenchymal volume is used in some centres as part of the assessment and selection of candidates for liver transplantation.

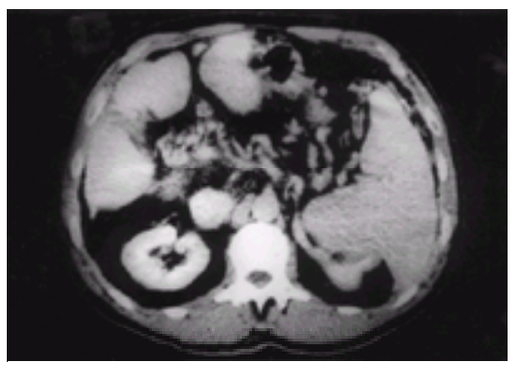


Fig. 14. Abdominal CT scan. Splenomegaly is clearly demonstrated with prominent retroperitoneal and hilar varices.

Diagnostic angiography

Angiography no longer plays a major role in the investigation of portal hypertension except where spiral CT or MRI facilities are not available. However, one major advantage of this catheter-based technique remains the capacity to measure hepatic blood flow, free and wedged hepatic pressures, and inferior vena cava pressures. Also, in many parts of the world, visualization of the portal system, particularly for identification of major portasystemic collaterals and provision of a map to allow planning of surgical intervention, continues to be by classical angiography. Therefore the various techniques of imaging of the portal circulation by angiography will be described.

Coeliac and superior mesenteric angiography

Coeliac and superior mesenteric angiography is the safest, most commonly used, but indirect means of visualizing the portal venous circulation (Fig. 15). Using the Seldinger technique under local anaesthesia, the coeliac trunk and superior mesenteric artery are selectively and individually cannulated using a small-bore arterial catheter (French 5) and 50 to 60 ml of contrast medium. The venous circulation is visualized awaiting the venous return of contrast after its arterial injection: this technique requires careful timing by the radiologist, and has been greatly simplified by the advent of digital subtraction angiography. Detailed imaging ofthe entire portal circulation with identification of major collaterals and shunts can be obtained routinely. Additional information, such as the presence of hepatofugal flow, may influence the choice of surgical procedure.



Fig. 15. Superior mesenteric angiogram. The venous phase shows a patent portal vein with a prominent left gastric vein and oesophageal varices.

Injection of contrast into the coeliac trunk allows the hepatic and splenic arterial circulation to be visualized. The intrahepatic circulation is frequently abnormal in cirrhosis, showing tortuosity or corkscrewing of the hepatic arterial branches. Haemangiomas, tumour circulation, hepatic artery-to-portal vein shunting, aneurysms, or anatomical variations of the major arteries can be readily and accurately demonstrated. Injection of contrast into the splenic artery allows evaluation of the splenic vein and coronary vein, and demonstration of oesophagogastric varices. Superior mesenteric angiography is used to delineate the superior mesenteric and portal veins and their collaterals.

Indirect visualization of the portal tree suffers from a degree of loss of detail, but this is rarely severe enough to warrant direct venography. Digital subtraction angiography is being used increasingly often with good results, but resolution is often poorer than that of conventional angiography.

Splenic venography

The best imaging of the splenic and portal veins, and collaterals is obtained using this technique, but at the cost of increased risk. Splenic venography is of particular value in the investigation of extrahepatic portal vein obstruction or when cavernous transformation of a thrombosed portal vein is suspected (Fig. 16). After infiltrationof local anaesthetic into tissues overlying the spleen, a fine-bore catheter is passed into the spleen and splenic pulp pressure is measured. Venography is then performed

by injection of contrast into the splenic pulp. Once imaging is completed, the catheter tract is sealed by injection of pledgets of gelatin foam. Providing any coagulopathy is corrected beforehand, this procedure is rarely complicated by significant haemorrhage, but extravasation of contrast around the spleen may cause pain, both local and referred to the left shoulder. The risk of complication remains higher than with angiography and splenic venography should therefore be reserved for patients in whom imaging by angiography is inadequate.



Fig. 16. Splenic venogram. The splenic injection gives the best definition of the portal venous circulation, here confirming the patency of a portacaval shunt.

Transhepatic venography

This technique is performed by percutaneous puncture of the liver, introduction of a fine-bore catheter into an intrahepatic radicle, and injection of contrast. Excellent imaging of the splenic and portal vein is obtained (Fig. 17) and varices can be embolized via the coronary or larger collateral veins. Initial success in stopping variceal haemorrhage is high (80 to 90 per cent), but there is a 25 to 30 per cent rebleeding rate within a few days. This, together with a high complication rate and the technically demanding nature of the procedure, has led to its virtual abandonment.

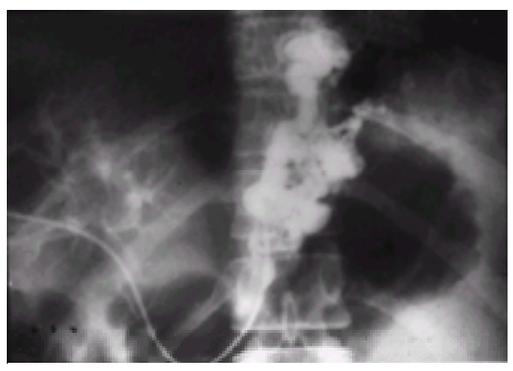


Fig. 17. Transhepatic portagram. Selective cannulation of the left gastric vein has been performed demonstrating oesophageal varices; embolization of bleeding varices is possible by this route but is now rarely employed because of technical difficulty and the high rebleeding rate.

Inferior vena cava and hepatic venography

In the Budd–Chiari syndrome the inferior vena cava is often severely stenosed or entirely occluded. The degree to which this vessel is compressed by the hypertrophied caudate lobe can be assessed both by visualization and measurement of the pressure gradient between the suprahepatic and infrahepatic inferior vena cavas (Fig. 18): a high gradient mitigates against treatment by portacaval shunt. Hepatic venography is of value in diagnosing hepatic vein thrombosis in patients with veno-occlusive disease associated with myeloproliferative disorders. Its most common use, however, is in measurement of free and wedged hepatic pressures in the assessment of cirrhosis and portal hypertension. These measurements are of particular value as a research tool in evaluation of the haemodynamic effects of drugs on the portal circulation.



Fig. 18. Inferior vena cavogram. These anteroposterior and lateral views of the inferior vena cava confirm the diagnosis of Budd–Chiari syndrome with compression of the hepatic vena cava by the massively hypertrophied caudate lobe.

Carbon dioxide angiography

Using digital subtraction with transjugular hepatic vein pump injection of carbon dioxide, excellent imaging of the portal circulation can be obtained (Fig. 19). This method is useful in diagnosis of portal vein thrombosis and has the advantage of avoiding contrast, which is particularly valuable in patients either at risk of or with established hepatorenal syndrome.

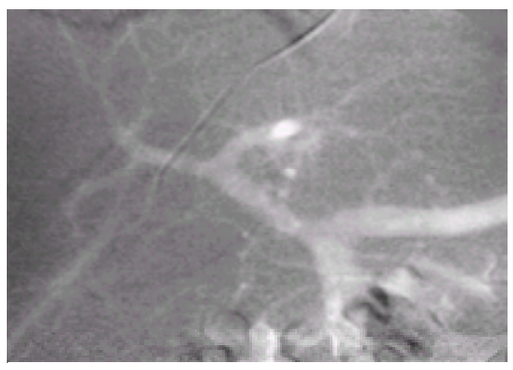


Fig. 19. Carbon dioxide portagram demonstrating the superior mesenteric, splenic, portal, and intrahepatic portal veins.

Further reading

Fernando B *et al.* N-acetylcysteine prevents development of the hyperdynamic circulation in the portal hypertensive rat. *Hepatology* 1998; **28**: 689–94. [Evidence that the hyperdynamic circulation can be prevented in experimental portal hypertension by blocking oxidant stress and that formation of reactive oxidant species may be important in the pathophysiology of portal hyprtension.]

Goulis J, Patch D, Burroughs AK. Hypothesis. Bacterial infection in the pathogenesis of variceal bleeding. *Lancet* 1999; **353**: 139–42. [A new hypothesis for variceal bleeding implicating endotoxin mediated increase in portal pressure and compromised haemostasis as two major factors which initiate bleeding.]

Gupta TK, Chen L, Groszmann RJ. Pathophysiology of portal hypertension. *Baillière's Clinical Gastroenterology* 1997; **11**: 203–19. [This is a good review of the pathophysiology of portal hypertension.]

Kitano S, Terblanche J, Kahn D, Bornman PC. Venous anatomy of the lower oesophagus in portal hypertension: practical implications. *British Journal of Surgery* 1986; **73**: 525–31. [This is an elegant anatomical study of the venous circulation of the lower oesophagus in portal hypertension.]

Sherlock S, Dooley J. The portal venous system and portal hypertension. In: *Diseases of the liver and biliary system*, 10th edn, pp. 135–75. Blackwell Science, Oxford, 1997. [An authoritative overview of the clinical entity of portal hypertension by one of the founding figures of hepatology.]

Sitzman JV, Li S-S, Adkinson F. Evidence for role of prostacyclin as a systemic hormone in portal hypertension. *Surgery* 1991; **109**: 149–53. [The importance of prostacyclin in the development of portal hypertension and the hyperdynamic circulation is presented.]

Spence RAJ, Terblanche J. Venous anatomy of the lower oesophagus: a new perspective on varices. *British Journal of Surgery* 1987; **74**: 659–60. [Further illumination of the lower oesophageal venous anatomy and its relevance to varices is presented.]

Tibballs JM, Watkinson AF. Imaging of the liver and biliary tract. In: Bircher J, Benhamou J-P, McIntyre N, Rizzetto M, Rodes J, eds, *Oxford textbook of clinical hepatology*, 2nd edn, pp. 549–78. Oxford University Press, 1999. [This is an excellent chapter describing imaging of the portal circulation.]

32.2 Emergency treatment of bleeding oesophageal varices

J. R. Galloway

Introduction
General principles of resuscitation
Pharmacotherapy
Endoscopic therapy
Balloon tamponade
Emergency decompression
Further reading

Introduction

The acute onset of bleeding from gastroesophageal varices continues to challenge both the internist and surgeon alike. Particularly frustrating is that cessation of bleeding, which can be achieved by any number of methods to be presented here, often does not correlate with ultimate survival of the patient. Indeed, patient survival after the bleeding has been stopped is primarily dependent upon the underlying function of the diseased liver or its hepatocellular reserve. This observation has led to a certain degree of medical mindlessness and lassitude, which causes many physicians to refer patients with variceal bleeding immediately to the interventional radiologist for an urgent transjugular intrahepatic portasystemic shunt (**TIPS**) procedure and just be done with it. This attitude, which may have some merit in its simplicity, ignores the many other well studied modalities available to the physician for obtaining control of variceal bleeding both in an emergency and in the long term.

These various treatment options which include pharmacotherapy, endoscopic therapy, balloon tamponade, total and selective shunts, devascularization procedures, and liver transplants should not be considered competitive but are in fact complementary (Table 1). Definitive therapy for bleeding varices should ultimately be tailored to individual patients in terms of their underlying liver function or hepatocellular reserve, as well as their need and candidacy for a liver transplant. Other factors, such as the type of liver disease, location and amount of variceal bleeding, comorbidities, and access to tertiary healthcare, should also be considered. Therefore, the immediate goal in the emergency treatment of patients with bleeding gastroesophageal varices is to stabilize them acutely, and bring them to elective evaluation. The purpose of the elective evaluation is to determine which available treatment option— medical, surgical, or radiographic—offers the patient the best survival advantage by controlling variceal bleeding while maintaining underlying hepatocellular function.

Pharmacotherapy
Octreotide
Pitressin + nitroglycerin
Endoscopic therapy
Variceal banding
Variceal sclerosis
Balloon tamponade
Decompression
Radiologic—TIPSS
Surgical—total and selective shunts
Devascularization procedures
Liver transplantation

Table 1 Therapies for variceal bleeding

General principles of resuscitation

The patient who presents with massive variceal bleeding is at significant risk of dying from hypovolemic shock and loss of oxygen carrying capacity. The well known general principles detailing the management of patients with hypovolemic shock govern the resuscitation of the patient with variceal bleeding. The patient should be admitted to an intensive care unit. Multiple, large-bore, intravenous accesses should be obtained immediately and a central venous line inserted, so that central venous pressure can be monitored. Routine pulmonary artery catheterization is unnecessary. In patients with cirrhosis and portal hypertension, it usually shows parameters consistent with a hyperdynamic state, such as increased cardiac output and low systemic vascular resistance. This hyperdynamic state is well described but its mechanism is not fully understood. The temptation to use vasopressors in this setting should be resisted, as it can lead to incomplete resuscitation of the still hypovolemic patient with variceal bleeding. A pulmonary artery catheter may, however, be of value in those patients with a history of significant atherosclerotic or valvular heart disease. Insertion of a Foley catheter to monitor urine output is required to ensure adequate volume resuscitation and to prevent renal failure. Careful observation of the patient's airway and oxygen administration to maintain arterial oxygen saturation at more than 94 per cent is also necessary. Early intubation of the patient with massive variceal bleeding is appropriate in maintaining adequate oxygenation and ventilation, as well as in the prevention of aspiration, particularly during any subsequent endoscopic treatment. Death from respiratory failure following significant aspiration during massive variceal bleeding occurs in many patients even though their bleeding is ultimately controlled. Clearly, this complication can be avoided if the principles of airway management are followed closely. Additionally, bacterial pneumonias are a common sequela following a major variceal bleeding episode. Serial chest radiographs, sputum cultures, and early institution of broad-spectrum intravenous antibiotics may be necessary to prevent death from pulmonary sepsis.

The judicious administration of crystalloid solutions should maintain blood pressure and urine output. The amount of salt given during this initial period must be carefully monitored, as too much sodium will only increase the amount of post-resuscitative ascites that may cause further morbidity should the patient survive. The appropriate use of fresh frozen plasma, platelets, and cryoprecipitate to correct any underlying coagulation defects is equally important. Finally, packed red blood cells are transfused to obtain adequate oxygen delivery by maintaining hemoglobin in the 9 to 10 g/dl range. While it is of paramount importance to maintain adequate hemodynamic status and oxygen delivery, ‘overresuscitation' of the patient should be avoided as it leads to a hypervolemic state that will not only worsen ascites, as noted above, but quite possibly can lead to increased portal pressures, and persistent or recurrent variceal bleeding. Finally, it is important to realize that during this early resuscitation phase, 50 per cent or more of patients will stop bleeding spontaneously without further treatment. For those who show evidence of continued bleeding, additional measures become necessary and these are outlined below.

Pharmacotherapy

The drugs available for control of acute variceal bleeding consist of vasopressin and the synthetic analog of somatostatin, octreotide. Both are potent, splanchnic vasoconstrictors that are administered intravenously to reduce portal hypertension and stop the acute bleeding episode. Non-cardioselective b-blockers, such as propranolol and nadolol, have been utilized in the prevention of variceal bleeding, but they have no role in acute variceal bleeding. Newer agents, such as ludothelin and other endothelial- produced factors, may prove to be of some value as further experience is gained.

Vasopressin was first used to treat variceal bleeding in 1959. Initial control of bleeding occurs in approximately 70 per cent of the patients who receive intravenous vasopressin, but recurrence of bleeding is high when the infusions are discontinued and survival is not necessarily improved. Because of its potent vasoconstrictor properties, vasopressin should be administered in the intensive care unit setting with appropriate hemodynamic monitoring including central venous pressure monitoring and an arterial line. Vasopressin is best administered through a central line because of the infrequent but significant complication of soft tissue necrosis when administered peripherally. Vasopressin should also be administered in conjunction with intravenous nitroglycerin. Nitroglycerin reduces the systemic vasoconstrictive side-effects of vasopressin, particularly coronary vasoconstriction, and has the additional effect of reducing portal pressure. A more common complication of vasopressin infusion is significant fluid retention, volume overload, and hyponatremia. Therefore, vasopressin should be delivered in as concentrated infusion as possible and maintenance intravenous fluid should be adjusted as necessary to prevent volume overload. Vasopressin is often delivered initially as a 20 unit bolus over 20 min in 200 ml of D5W, although this may not be necessary and may increase the risk of systemic side-effects. This is followed by a continuous infusion in a dose range from 0.1 to 0.4 unit/min. The concurrent infusion of nitroglycerin usually begins at 40 µg/min and is titrated to maintain mean arterial pressures in the range of 60 to 70 mmHg. In the past, a patient was weaned off vasopressin slowly in decreasing increments of 0.1 unit/min over a 24-h period once variceal bleeding had ceased. The reasoning for this is unclear and most likely unjustified. In current practice, vasopressin is usually terminated when cession of bleeding has been documented by endoscopic means. Because of the significant complications associated with vasopressin, terlipressin, an analog of vasopressin, has been widely utilized in Europe, although it is not available in the United States. It has the advantage of a longer half-life, and less diffuse systemic vasoconstriction while maintaining

effective splanchnic vasoconstriction. Controlled trials show improved bleeding control approaching 80 per cent and a survival benefit over placebo infusions.

Octreotide, the synthetic analog of somatostatin, has replaced to a large extent the use of vasopressin and nitroglycerin in the control of acute variceal bleeding. It too is a potent splanchnic bed vasoconstrictor when administered in a supraphysiologic dose. This change to octreotide as the drug of choice has resulted from it having been shown to be equally effective as the vasopressin/nitroglycerin combination in controlling variceal bleeding without the significant hemodynamic side-effects. Octreotide is administered continuously in a dose range between 50 and 100 μg/h after an initial 50 μg intravenous bolus.

Endoscopic therapy

Upper gastointestinal endoscopy plays an extremely important role in the management of acute variceal bleeding, both as a diagnostic modality and a therapeutic intervention. Early endoscopy in the patient with cirrhosis and suspected portal hypertensive bleeding not only documents the site of variceal bleeding, but also defines other potential causes of gastrointestinal bleeding such as a Mallory–Weiss tear, erosive gastritis, gastric and duodenal ulcer disease, and malignancy. The therapeutic role of upper endoscopy in acute variceal bleeding is also well established. Since the advent of flexible endoscopy during the 1970s, endoscopic variceal sclerotherapy has become the most widely used treatment in the management of acutely bleeding varices. Endoscopic sclerotherapy is the injection of a sclerosing agent (sodium morrhuate, sodium tetradecol, or ethanolamine) either into the lumen of the varix (intravariceal) or into the submucosal tissue surrounding the varix (paravariceal) in order to clot or tamponade the varix, respectively (Fig. 1). In episodes of acute variceal bleeding, endoscopic sclerotherapy will lead to cessation of bleeding in 90 to 95 per cent of cases. The technical aspects of variceal sclerotherapy, including the type of sclerosant used, intravariceal or paravariceal injection, and the number and frequency of sclerotherapy sessions, have all been extensively studied. No one technique or agent has been found to be significantly better than the others in controlling bleeding. The known risks of endoscopic sclerotherapy include esophageal perforation, post-sclerotherapy esophageal ulceration with bleeding, esophageal stricture formation, and rarely, sclerosant embolization to the pulmonary circulation with resultant acute respiratory distress.

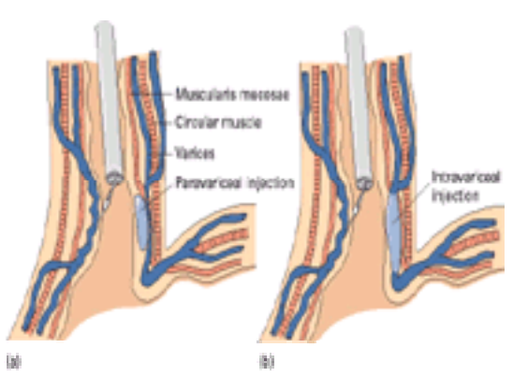


Fig. 1. Endoscopic variceal aclerotherapy. (a) Paravariceal injections tamponade the varix, while(b) intravariceal injections cause thrombosis of the varix.

In 1986 the technique of endoscopic variceal band ligation was introduced (Fig. 2). This technique is an adaptation of the successful experience in placing small rubber bands around hemorrhoids. This procedure involves the deployment of small rubber bands by a triggering device around the base of the varix that has been suctioned into a sheath at the end of an endoscope. The banded varix has its blood flow interrupted and usually sloughs off in the following 72 h leaving a small, shallow, non-bleeding ulcer that rapidly heals. Its initial acceptance was delayed until the advent of multiple band ligatures, which obviated the need to reinsert the endoscope repeatedly to band each varix. Multiple controlled trials comparing sclerotherapy with banding suggest that variceal band ligation has an equal control rate of acute variceal bleeding with lower complication rates of mortality, variceal rebleeding, esophageal perforation, and strictures. Its primary disadvantage is the difficulty in visualizing varices during massive bleeding due to a 30 per cent reduction in the field of view caused by the plastic channel used for the deployment of the bands. Nevertheless, endoscopic variceal band ligation has now replaced endoscopic sclerotherapy as the endoscopic therapeutic procedure of choice in most major portal hypertensive centers.

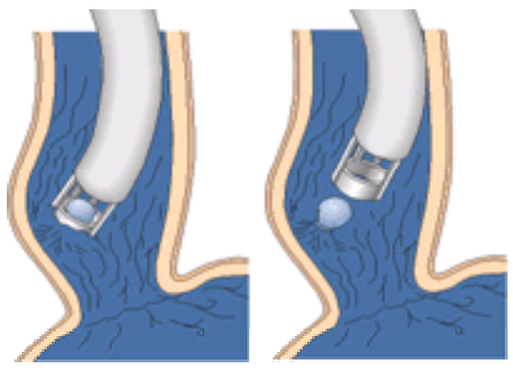


Fig. 2. Endoscopic variceal band ligation. The esophageal varix is suctioned up into the ligating device (a) and the base of the varix is ligated with an O-ring (b).

Although not part of this particular discussion on control of acute variceal bleeding, it should be noted that endoscopic variceal banding and sclerotherapy both play an important first-line role in the obliteration of varices over a more prolonged period of time in order to prevent recurrent variceal bleeding. Rebleeding rate in this more chronic form of control approaches 50 per cent for sclerotherapy and 35 per cent for band ligation, but many patients who experience rebleeding episodes can have their bleeding recontrolled with further sclerotherapy or banding.

Balloon tamponade

There are a few occasions when neither pharmacologic nor endoscopic measures will lead to control of acute variceal bleeding. In these cases, bleeding tends to be torrential and resuscitative efforts, though ongoing, continue to fall behind in terms of maintenance of hemodynamic status and oxygen delivery. It is during this difficult circumstance that balloon tamponade can be a life-saving procedure by stabilizing the patient and allow a more definitive decompression procedure to be performed (Fig. 3). There are two balloon tamponade devices available. The older Sengstaken–Blakemore tube, which is a three-lumen tube, and the newer four-lumen Minnesota tube. Both tubes have a gastric balloon, an esophageal balloon, and a distal gastric port for aspiration of gastric contents. The Minnesota tube's fourth lumen is a proximal aspiration port above the esophageal balloon whose sole purpose is the aspiration of salivary secretions from the cervical esophagus. Because it lacks this proximal port, the Sengstaken–Blakemore tube requires placement of a proximal orogastric or nasogastric tube.

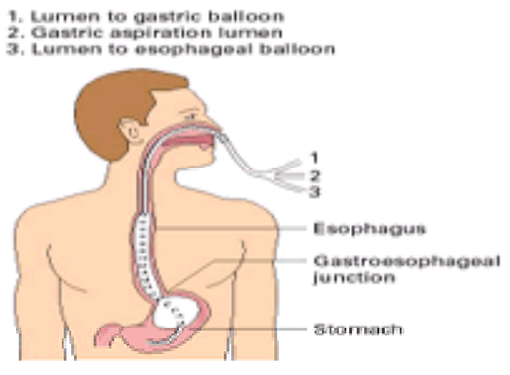


Fig. 3. The Sengstaken–Blakemore tube is used to tamponade acutely bleeding gastroesophageal varices. The Minnesota tube (not pictured) has a fourth lumen for

aspiration of salivary secretions from the cervical esophagus.

Although infrequently used, a brief discussion about the placement of these tubes is appropriate since actual placement is not as simple as the concept of tamponade. Usually by the time the decision is made to institute balloon tamponade, the patient is already intubated and ventilated due to continued bleeding and hemodynamic instability. If not, it is mandatory that the patient undergo urgent intubation and mechanical ventilation prior to insertion of either tube in order to prevent the risk of massive aspiration. Once the patient's airway has been controlled, the tube can be safely placed with less risk of aspiration. Although the package insert for both tubes shows placement through the nasopharynx, this is extremely difficult and placement of the well lubricated tube through the oral pharynx is recommended. Prior to placement of the tube, one should measure from the xiphoid process to the incisors the number of centimeters on the tube necessary to ensure complete placement of the gastric balloon within the stomach. This is usually more than 40 cm. These tubes are very flexible and when well lubricated are often difficult to manipulate. Soaking the tube in an ice-water bath prior to lubrication stiffens them and assists in their placement.

Once the tube has been inserted into the stomach beyond the number of centimeters approximated from the incisors, around 50 ml of air is injected into the gastric balloon. An upper abdominal radiograph is obtained immediately to document that the balloon is well within the stomach. Following this, approximately 250 to 350 ml of air is placed into the balloon and the balloon port clamped. The gastric balloon is then gently pulled superiorly against the gastroesophageal junction to effect tamponade on the varices. The old football helmets used to hold the tube in place are no longer utilized, and in most cases, the tube can be anchored to the mouthpiece holding the endotracheal tube in place, or taped into place at the nares using a small styrofoam device that accompanies most of these tubes. The balloon should not be pulled up too tightly against the gastroesophageal junction, as it is unnecessary in terms of stopping the bleeding, and it only increases the risk of mucosal ischemia and ulceration by the balloon. In nearly all cases, inflation of the gastric balloon alone will stop the variceal bleeding even if the main bleeding site is higher up in the esophagus since blood flow through the more superior esophageal varices is decreased by the interruption of blood flow by the tamponading balloon on the more inferior varices at the gastroesophageal junction. On those rare occasions when bleeding continues, it may be necessary to inflate the esophageal balloon. In most patients, 100 ml of air instilled into the esophageal balloon will obtain an appropriate amount of tamponade. The esophageal port should then be connected to a blood pressure manometer and monitored every 2 h. The pressure within the esophageal balloon should be kept in the 40 to 45 mmHg range. Occasionally air will need to be either removed or instilled in order to obtain this appropriate amount of pressure. If the Minnesota tube is being utilized, it is not necessary to place an oropharyngeal or nasogastric tube above the esophageal balloon once it is inflated.

Balloon tamponade is potentially dangerous. Complications include ulceration and esophageal rupture. Nevertheless, it is a very effective means of controlling massive variceal bleeding when other modalities have failed. It allows for stabilization of the patient, correction of coagulopathy, and treatment of any metabolic derangements that accompany a massive bleed. Balloon tamponade is a temporizing measure and not a definitive therapy. The balloon should not be left inflated for more than 48 h. Institution of balloon tamponade mandates decompression of the varices via TIPS or surgical shunting within that time period.

Emergency decompression

In those instances when less invasive procedures have failed to stop the bleeding, emergency decompression of the varices becomes necessary. Although surgical methods such as total and selective shunts have been utilized in the emergency setting in the past, the transjugular intrahepatic portasystemic shunt (TIPS) has now replaced surgical decompression as the procedure of choice in the control of acute variceal bleeding unresponsive to pharmacologic or endoscopic methods.

The TIPS essentially functions as a side-to-side portasystemic shunt, as it decompresses the gastroesophageal, splenic, superior mesenteric, and portal vein components of the portal circulation. It does so by diverting portal blood flow away from the hepatic sinusoids through a hepatic parenchymal tunnel expanded by metallic wire-mesh stent between the high-pressure portal venous system and the low-pressure hepatic veins. Doppler ultrasonography is necessary to ensure portal vein patency prior to performing the TIPS. The procedure is performed in an angiographic suite equipped with good-quality digital imaging. The first step is cannulation of the right internal jugular vein, which provides the most direct route to the hepatic veins. Through the introducing sheath within the internal jugular vein, a diagnostic angiographic catheter is advanced under fluoroscopic guidance into the hepatic veins, and hepatic venography is performed to assess the size and location of the hepatic veins (Fig. 4). Several coaxial catheters and needle systems are at present available to create the tract between the hepatic venous and the portal venous systems. A 16- or 21-gauge needle is then used to access the portal venous system, usually via the right portal vein (Fig. 5). Once the needle is inserted into the portal venous system, an angiographic guidewire is advanced into the portal vein followed by passage of the angiographic catheter (Fig. 6). A venogram is then performed through this catheter to define better the portal anatomy, the direction of flow within the portal vein and its branches, and the presence of portasystemic venous collaterals. The transhepatic tract is then dilated with an angioplasty balloon (Fig. 7) followed by the deployment of a balloon expandable stent (Fig. 8). Although a variety of metal stents have been used, the most popular is the expandable wire-mesh Wallstent. This stent has the advantage of being very flexible enabling it to negotiate any angle between the hepatic vein and portal vein. Once the Wallstent has been placed in the appropriate position, it is dilated with sequentially larger balloons until the portasystemic gradient has dropped to 12 mmHg or less. Usually a Wallstent of 10 mm maximal diameter is necessary to obtain this drop in portal pressure. A second stent may be necessary in parallel fashion to drop the portasystemic gradient sufficiently. A final portal venogram is obtained in order to assess stent position and patency, to ascertain flow within the portal vein branches, and to identify any persistent filling of gastroesophageal varices. Embolization of any persistent varices is sometimes added to the procedure although the clinical efficacy of this additional step is unproved.

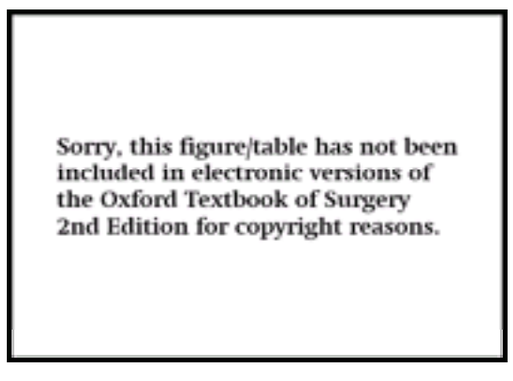


Fig. 4. TIPS: selective catheterization of the right hepatic vein.

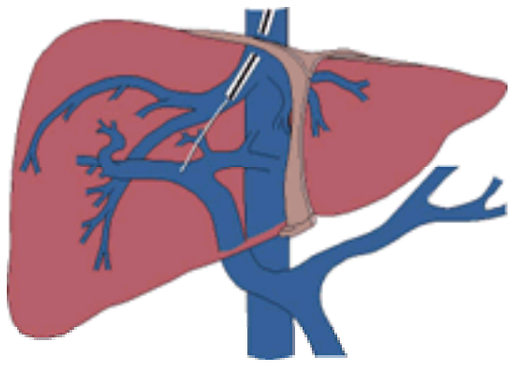


Fig. 5. TIPS: a 21-gauge needle punctures the hepatic parenchyma and is directed into the right portal vein.

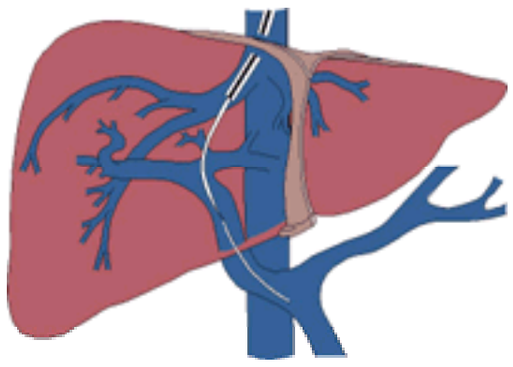


Fig. 6. TIPS: a guidewire is passed through the needle into the portal venous system via the right portal vein.

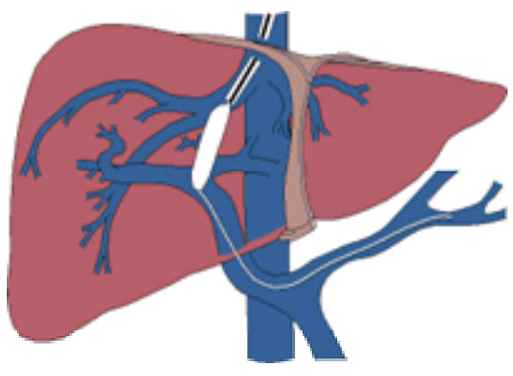


Fig. 7. TIPS: the hepatic parenchyma is dilated by the balloon to 8 mm.

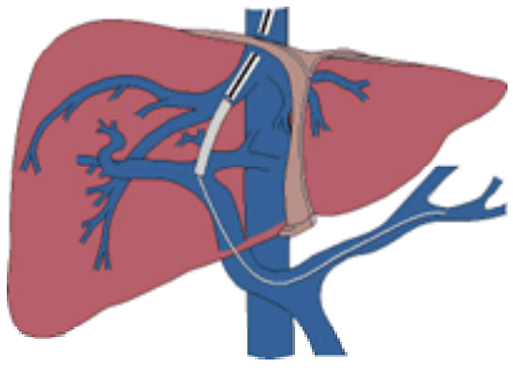


Fig. 8. TIPS: the metallic stent is deployed along the newly created parenchymal tunnel and further expanded to 10 mm.

Because of its relatively non-invasive nature when compared with surgical shunts, the TIPS has enjoyed widespread use since it became popularized in 1990. At present, the TIPS procedure can be performed with a technical success rate of approximately 90 per cent, and a procedural complication rate of approximately 10 per cent. Technical failures occur primarily due to massive ascites, which should be drained prior to a TIPS, very hard or atrophic livers, and occluded portal or hepatic veins. Procedural complications include subcapsular hematoma formation, hemorrhage from rupture of the liver capsule during needle puncture, and misplacement of the Wallstent too far up into the right atrium or too far down into the portal vein. Other technical complications include neck hematoma, puncture of the gallbladder or bile duct resulting in hemobilia, acute stent thrombosis, and acute thrombosis of the portal venous system. Acute systemic complications associated with the procedure include cor pulmonale, adult respiratory stress syndrome, sepsis, hepatic failure, myocardial infarction, bacterial endocarditis, and contrast-induced renal failure. The long-term stenosis rate appears to be approximately 50 per cent at 6 to 12 months, and there is a shunt thrombosis rate of 10 per cent at 6 months. Two mechanisms appear to contribute to stenosis and thrombosis: pseudointimal hyperplasia may narrow the lumen of the stent, and thickening and narrowing of the hepatic venous outflow tract, which tends to occur between 6 and 12 months after stent placement. These stenosis and thrombosis rates translate to a 20 per cent rebleeding rate during the first year. The TIPS also has a 30 per cent encephalopathy rate because it functions as a total shunt, and many patients requiring a TIPS have poor underlying hepatic function. A large number of these patients will, however, respond to traditional medical therapy for hepatic encephalopathy. Once the TIPS has been placed, a rigorous follow-up program to visualize the TIPS with Doppler ultrasonography every 4 months, and shunt catheterization with dilatation of any stenosed shunts, is necessary to maintain an adequate patency rate. Contraindications to a TIPS include right-sided heart failure, severe hepatic failure, and polycystic liver disease. Placement of a TIPS through a hepatic neoplasm should also be avoided because of the increased results of bleeding and tumor dissemination.

The success of the TIPS in the emergency setting has markedly reduced the number of surgical shunts performed for control of acute variceal bleeding. Surgical shunts are restricted primarily to those patients in whom a TIPS is technically not feasible, or are performed when an interventional radiologist trained in the TIPS is not available. Historically, the most common surgical shunt for emergency control of variceal bleeding has been the portacaval H-graft shunt because it can be rapidly performed resulting in quick and effective decompression of the portal venous system leading to control of variceal bleeding. Several variations of the portacaval shunt are also available. Of those, the mesocaval C-graft shunt is currently most recommended because it too can be performed rapidly and allows for total decompression of the portal venous system. Its primary advantage over the portacaval H-graft shunt is that it is performed below the transverse mesocolon away from the portahepatis, and, therefore, is less likely to interfere technically with any future possible liver transplant. The distal splenorenal shunt (**DSRS**) has been utilized in the urgent setting in patients who have had their variceal bleeding ‘slowed' pharmacologically or by balloon tamponade. Potential candidates for emergency DSRS should have reasonable hepatic function, appropriate anatomy, and absence of or medically controlled ascites. If emergency DSRS is considered, then early ligation of variceal inflow (the coronary veins) should be performed before the shunt. However, even in experienced hands, a DSRS can take twice as long as a portacaval or mesocaval shunt and, therefore, does not lend itself to the emergency setting. Surgical decompression of bleeding varices is discussed more thoroughly in Chapter 32.3.

A final algorithm which utilizes all of these modalities for control of variceal bleeding in a sequential non-competitive way is presented in Fig. 9.

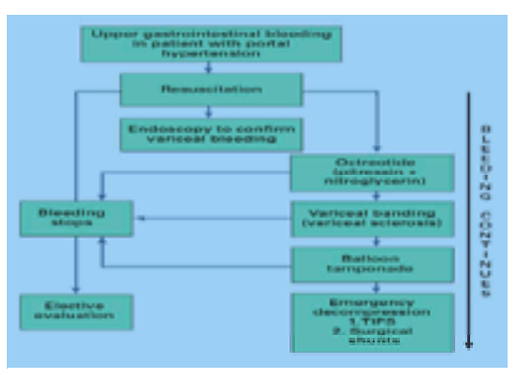


Fig. 9. An algorithm for the emergency treatment of bleeding esophageal varices.

Further reading

Brooks WS, Sclerotherapy: analysis and review. *Endoscopy Reviews* 1985; May/June: 11–53.

Burnett DA, Rikkers LF. Nonoperative emergency treatment of variceal hemorrhage. *Surgical Clinics of North America* 1990; **70**: 291–306.

Burroughs AK *et al.* Randomized, double-blind, placebo-controlled trial of somatostatin for variceal bleeding: emergency control and prevention of early variceal rebleeding. *Gastroenterology* 1990; **99**: 1388–95.

Caberra J *et al.* Transjugular intrahepatic portosystemic shunt versus sclerotherapy in elective treatment of variceal hemorrhage. *Gastroenterology* 1996; **110**: 832–9.

Cello JP *et al.* Endoscopic sclerotherapy versus percutaneous transjugular intrahaepatic portosystemic shunt (TIPS) after initial sclerotherapy in cirrhotic patients with acute variceal hemorrhage: a randomized controlled trial. *Annals of Internal Medicine* 1997: **126**: 950–865.

Conn HO. Transjugular intrahepatic portal-systemic shunts: the state of the art. *Hepatology* 1993; **17**: 148–58.

Fogel MR *et al.* Continuous intravenous vasopressin in active upper gastrointestinal bleeding: a placebo-controlled trial. *Annals of Internal Medicine* 1982; **96**: 565–9.

Ginson AES *et al.* Randomized trial of variceal band ligation versus injection sclerotherapy for bleeding esophageal varices. *Lancet* 1993; **342**: 391–4.

Grace ND. Variceal hemorrhage; pharmacologic approach. In: McDermott WV, Bothe A, eds, *Surgery of the liver*, pp. 33–314. Blackwell Scientific Publications, Boston, 1988.

Hwang SJ *et al.* A randomized controlled trial comparing octreotide and vasopressin in the control of acute esophageal variceal bleeding. *Journal of Hepatology* 1929; **16**: 320–5.

Infante-Rivard C, Esraola S, Villeneuve JP. Role of endoscopic variceal sclerotherapy in the longterm management of variceal bleeding: a meta-analysis. *Gastroenterology* 1989; **96**: 1087–92.

Laine L. Management of actively bleeding esophageal varices. *Gastrointestinal Endoscopy* 1997; **46**: 83–7.

Laine L *et al.* Endoscopic ligation compared with sclerotherapy for treatment of bleeding esophageal varices. *Annals of Internal Medicine* 1993; **119**: 1–7.

Panes J *et al.* Efficacy of balloon tamponade in the treatment of bleeding gastric and esophageal varices. Results in 151 consecutive episodes. *Digestive Diseases and Sciences* 1988; **33**: 454–8.

Paquet KJ, Feussner H. Endoscopic sclerosis and esophageal balloon tamponade in acute hemorrhage from esophagogastric varices: a prospective randomized controlled trial. *Hepatology* 1985; **5**: 580–3.

Ring EJ *et al.* Using transjugular intrahepatic portosystemic shunts to control variceal bleeding before liver transplantation. *Annals of Surgery* 1992; **116**: 304–9.

Rossle M *et al.* The transjugular intrahepatic portosystemic stent-shunt procedure for variceal bleeding. *New England Journal of Medicine* 1994; **330**: 165–71.

Sherlock S. Esophageal varices. *American Journal of Surgery* 1990; **160**: 9–13.

Stiegmann GV *et al.* Endoscopic sclerotherapy as compared with endoscopic ligation for bleeding varices. *New England Journal of Medicine* 1992; **326**: 1527–32.

Sung JYI *et al.* Octreotide infusion or emergency sclerotherapy for variceal hemorrhage. *Lancet* 1993; **342**: 637–41.

Terblanche J. Has sclerotherapy altered the management of patients with variceal bleeding? *American Journal of Surgery* 1990; **160**: 37–42.

Terblanche J, Burrouoghs AK, Hobbs KEF. Controversies in management of bleeding esophageal varices. *New England Journal of Medicine* 1989; **320**: 1393–8, 1469–74.

Teres J *et al.* Vasopressin/nitroglycerin infusion versus esophageal tamponade in the treatment of acute variceal bleeding: a randomized controlled trial. *Hepatology* 1990: **11**: 964–8.

Tsai YT *et al.* Controlled trial of vasopressin plus nitroglycerin versus vasopressin alone in treatment of bleeding esophageal varices. *Hepatology* 1986; **6**: 406–9.

Walker S *et al.* Terlipressin in bleeding esophageal varices; a placebo-controlled, double-blind study. *Hepatology* 1986; **6**: 112–15.

Westaby D, Williams R. Status of sclerotherapy for variceal bleeding in 1990. *American Journal of Surgery* 1990; **160**: 32–6.

32.3 Surgical shunt procedures

J. R. Galloway

Introduction
Total shunts
End-to-side portacaval shunt
Side-to-side portacaval shunt and its variants
Partial shunts
Selective shunts
Distal splenorenal shunt (DSRS)
Distal splenorenal shunt with splenopancreatic disconnection
Selective distal splenocaval shunt
The coronary–caval shunt
Definitive treatment for variceal bleeding
Further reading

Introduction

During the 40-year period from the early 1940s to the early 1980s, shunt procedures were the primary emergency as well as elective treatment for variceal bleeding (Fig. 1). Throughout this time, numerous studies were undertaken to determine the outcome of total shunts compared with medical therapy alone, total shunts compared with endoscopic therapy, total shunts compared with selective shunts, and selective shunts compared with endoscopic therapy. In 1983, orthotopic liver transplantation moved from the realm of experimental procedures and became an accepted therapy for the patient with endstage liver disease as a result of improved immunosuppression with cyclosporine. This achievement along with the wide acceptance of the transjugular intrahepatic portasystemic shunt (**TIPS**) procedure by the early 1990s appears to have reduced surgical shunts to a much more secondary role. Shunts are now rarely performed in the community, and even in major portal hypertensive centers, the number of shunts has decreased from well over a 100/year in the early 1980s to less than 25/year currently. Further, these shunts are being performed in a more highly selected group of patients than ever before.

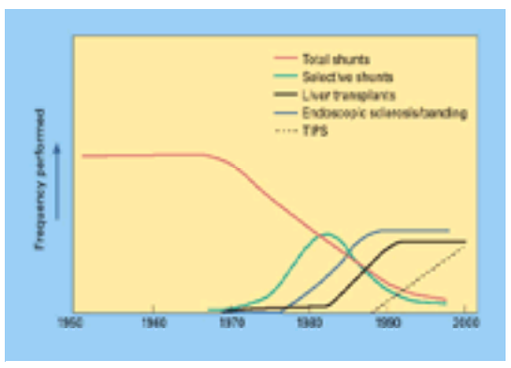


Fig. 1. Relative frequency of various treatment modalities for variceal bleeding during the past 50 years.

While liver transplantation and endoscopic methods of variceal bleeding control have appropriately siphoned away many of the patients who would have otherwise been shunted in earlier years, it is felt that the ubiquity of the TIPS has led to its overutilization, resulting in a decrease in the referral of patients for surgical shunt procedures. Despite these events, a working knowledge of the various shunt procedures available is necessary, including a thorough understanding of the indications for the various shunts, the technical pitfalls associated with each shunt, and the ultimate result that can be anticipated with each shunt. In most cases, however, the surgeon should perform the kind of shunt that he/she knows best how to do; especially if referral to a tertiary medical center where shunts are more routinely performed is not possible. Worrying about the debated pros and cons of each shunt is appropriate if the surgeon can perform all shunts equally well. However, the advantages of one particular shunt over another are potentially lost if the surgeon is considering a shunt in which he/she has little or no experience.

The goal of all shunts is to reduce variceal pressure, and thereby stop acute variceal bleeding and prevent recurrent variceal bleeding. All shunts, both total and selective, accomplish this goal more than 90 per cent of the time, although the ultimate long-term shunt patency rates may vary. The difference between total, partial, and selective shunts performed surgically is the degree in which portal perfusion is maintained to the cirrhotic liver, and to what effect any altered perfusion has ultimately on liver function and reserve. Total shunts are large-diameter shunts which decompress the entire portamesenteric venous system and divert all portal flow out of, or away from, the liver. Total shunting occurs because the patent shunt offers a pathway of lower resistance to the portal blood flow than the diseased liver. Partial shunts are those with an 8- or 10-mm diameter whose purpose is to reduce variceal pressures while maintaining some degree of portal flow. Selective shunts lower variceal pressures by decompressing only the gastric and splenic components of the portamesenteric venous circulation while maintaining portal perfusion of the cirrhotic liver.

Total shunts

A total portasystemic shunt can be accomplished in several ways and some are briefly presented here.

End-to-side portacaval shunt

The classic end-to-side portacaval shunt divides the portal vein just below its bifurcation and turns the mesenteric end of the portal vein down to the infrahepatic inferior vena cava where an end-to-side anastomosis is performed (Fig. 2). This end-to-side shunt completely diverts portal blood flow away from the liver. Unlike the other total shunts, it does not decompress the obstructed sinusoids of the liver because the portal vein is unable to act as an outflow tract. As such, the end-to-side portacaval shunt controls variceal bleeding extremely well, but does not necessarily alleviate ascites. It has an excellent long-term patency rate exceeding 90 per cent.

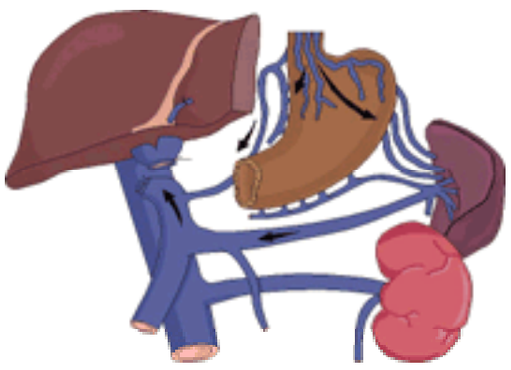


Fig. 2. End-to-side portacaval shunt.

There are several important technical details in performing the end-to-side portacaval shunt. Dissection should be limited to the relevant portion of the portal vein and the inferior vena cava whose exposure can be facilitated by reflecting the duodenum with the Kocher maneuver. The portal structures should be retracted medially and

superiorly away from the portal vein. Extreme caution is required during the dissection of the medial aspect of the portal vein along its entire length within the portahepatis. There are several potential venous tributaries entering the medial aspect of the portal vein, which are divided to ensure complete mobilization of the portal vein. A coronary vein arising superior to the junction of the splenic and superior mesenteric veins can be preserved unless division is necessary for further mobilization of the portal vein. Any replaced or accessory right hepatic artery running along the posterolateral aspect of the portal triad should be preserved. Division of a portion of the pancreas extending superiorly, or of a dense amount of lymphatic tissue lying posterior to the portal vein, is also required. The necessity of bringing down the end of the divided portal vein to the anterior medial aspect of the inferior vena cava without kinking or twisting should be observed. Finally, meticulous construction of the anastomosis, including the avoidance of suturing the anterior and posterior layers of the anastomosis together, is paramount.

Side-to-side portacaval shunt and its variants

All other types of total portasystemic shunts are variants of the side-to-side portacaval shunt. These shunts involve the joining of the portal vein, or one of its major tributaries, to the inferior vena cava or one of its major tributaries. They are usually more than 10 mm in diameter. The side-to-side portacaval shunt, the portacaval H-graft shunt, the mesocaval C-graft shunt, the mesorenal shunt, and the central splenorenal shunt are examples of what are functionally side-to-side shunts (Fig. 3). The closer the shunt is to the liver (i.e. the more central the shunt is), the greater the continuing patency rate because the veins involving the anastomosis are larger and have a higher flow. While prosthetic materials may increase the risk of thrombosis, their use may obviate the need for extensive mobilization of the veins used to form the shunt. This can be technically demanding and hazardous in a portal hypertensive field. Factors determining the choice of total shunt include the technical feasibility based on the preoperative radiographic evaluation of the relevant vessels, the surgeon's experience and preference, and the subsequent possibility of liver transplantation.

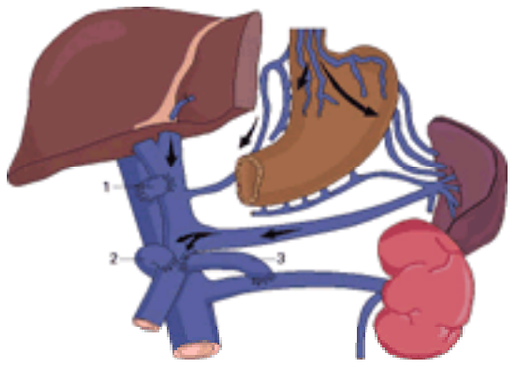


Fig. 3. Various interposition shunts. (1) Portacaval. (2) Mesocaval.(3) Mesorenal.

The side-to-side portacaval shunt is more difficult to perform than other total shunts because the portal vein and inferior vena cava may be too far apart to be anastomosed easily. Often, extensive dissection and mobilization of both veins, as described with the end-to-side portacaval shunt, are required in order for these veins to be joined easily. Additionally, their potential route of connection may be blocked by an enlarged caudate lobe, which may result from hepatic engorgement in Budd–Chiari syndrome or isolated hypertrophy in cirrhosis. Tension on the side-to-side anastomosis by pulling two distant veins together may result in narrowing of the shunt and predispose it to thrombosis. Because of these technical difficulties associated with side-to-side portacaval shunts, an interposition H-graft of a vascular prosthesis may be necessary. A short piece of ringed Gortex graft (polytetrafluoroethylene, or **PTFE**), 18 mm or more in size, has a reasonably low thrombosis rate and accomplishes the same physiologic goals as a direct side-to-side portacaval shunt. Dacron vascular grafts should be avoided in all portal hypertensive shunting procedures because of a higher thrombosis rate. Unfortunately, side-to-side portacaval shunts and the portacaval H-graft shunt both require extensive dissection within the portahepatis making any subsequent liver transplant more difficult technically by violating this area. Additionally, portacaval anastomoses are not feasible in patients with extrahepatic portal vein thrombosis regardless of the presence or absence of cirrhosis.

These technical concerns have resulted in the adoption of the interposition mesocaval C-shunt by many portal hypertensive surgeons as the procedure of choice for control of variceal bleeding in the emergency or partial emergency setting (Fig. 4). It is perhaps the easiest, quickest, and safest of all total decompressive shunt procedures. As a functional side-to-side total shunt, it is also extremely effective in the treatment of medically intractable ascites. Another advantage in this era of transplantation is that the shunt is performed below the transverse mesocolon well away from the liver. Therefore, it does not interfere with any subsequent portal dissection during liver transplantation and keeps portal hypertension to a minimum until implantation of the donor liver, after which the shunt is ligated. The mesocaval shunt has also been very useful in decompressing the congested liver secondary to hepatic vein thrombosis, also known as Budd–Chiari syndrome. A final advantage is that, despite a past relatively high shunt thrombosis rate of 20 per cent with previous graft materials, the use of autologous vein grafts, or ringed PTFE, along with the C-configuration of the shunt have yielded a very acceptable long-term patency rate approaching 90 per cent.

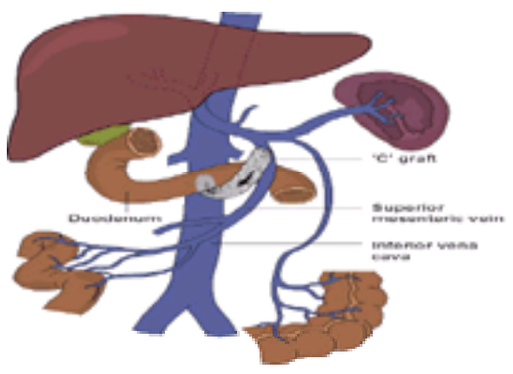


Fig. 4. Mesocaval C-shunt.

The mesocaval shunt can be performed through either a bilateral subcostal or long midline incision. The left neck is also draped for procurement of the left internal jugular vein if an autogenous vein graft is desired as its size makes for an ideal conduit. Alternatively, a ringed PTFE graft of at least 12 mm in diameter can be utilized. Upon entering the abdomen, the transverse mesocolon is elevated and the middle colic vein is followed as it courses down to the superior mesenteric vein. Palpation of the superior mesenteric artery at the root of the mesentery occasionally helps to identify the location of the superior mesenteric vein, which is to the right of this artery. However, these anatomic relationships are inconsistent and a wide transverse incision at the root of the mesentery best identifies the superior mesenteric vein. This vein is then freed from the surrounding connective tissue and lymphatics before it passes beneath the pancreas. Large lymphatic channels should be ligated to prevent postoperative chylous ascites. While small venous tributaries of the superior mesenteric vein are divided, it is generally not necessary to ligate any of its major branches. Occasionally, ligation of the middle colic vein facilitates a more proximal exposure of the superior mesenteric vein. At least 6 cm of vein should be exposed. The inferior vena cava is identified through the right colonic mesentery adjacent to the duodenum. The inferior vena cava need not be mobilized, but its anterior surface must be cleared for several centimeters proximally and distally below the renal veins. Mobilization of the right colon will help facilitate adequate exposure of the anterior surface of the inferior vena cava, especially in patients with a thickened or fatty mesentery, but this maneuver can result in significant bleeding in the presence of portal hypertension. The C-loop of the duodenum is partially kocherized and retracted superiorly for several centimeters for additional exposure of the inferior vena cava.

The internal jugular vein graft, or a 14- to 20-mm PTFE graft, is then trimmed to an appropriate length to avoid tension on the anastomoses; especially if the duodenum impinges on the graft. The inferior vena cava is then partially cross-clamped with a side-biting vascular clamp and a small button or ellipse approximating the size of the graft is cut from the anterior surface of the inferior vena cava. The vena caval anastomosis to the graft is then performed with a running 4–0 vascular suture, being careful to evert the anastomosis. Proximal and distal control of the exposed superior mesenteric vein is then obtained. Care must be taken not to injure the two or three relatively large tributary veins coming from the proximal jejunum and the right colon that enter the posterolateral and posteromedial aspects of the superior mesenteric vein. The superior mesenteric vein is then occluded proximally and distally with small vascular clamps while its larger venous tributaries are controlled with encompassing vessel loops. The graft is then anastomosed to the right anterolateral side of the superior mesenteric vein with a running 4–0 vascular suture. Care must be taken to prevent twisting or kinking of the graft. The final configuration of the graft should take on a C-shape as it takes off from the anterior surface of the inferior vena cava and circles around the second and third portions of the duodenum to enter into the right anterolateral side of the superior mesenteric vein. Erosion of the shunt into the duodenum is a theoretical possibility, but the author has never seen this complication. The ‘C' configuration of the shunt allows for a relatively easy and

large anastomosis to be performed between the graft and the main portion of the superior mesenteric vein. This technical advance contributes to improved patency rate in this shunt.

The Linton shunt, or central splenorenal shunt, is an end-to-side splenorenal anastomosis utilizing the proximal splenic vein as it joins the superior mesenteric vein to form the portal vein. As such, it functions as a total shunt, serving to decompress the entire portamesenteric venous system. It also incorporates a splenectomy and a limited gastric devascularization. For this reason, its use has been advocated in the setting of massive splenomegaly with splenic sequestration even though the associated pancytopenia, and in particular the thrombocytopenia, accompanying the portal hypertension, is usually of minimal clinical significance. This shunt is particularly difficult to perform, requiring extensive splenic vein dissection distally towards the tail of the pancreas in order to prevent kinking of the proximal splenic vein when it is brought down for anastomosis into the left renal vein. The long-term thrombosis rate of the Linton shunt approaches 30 per cent, although the rebleeding rate is somewhat lower due to the partial gastric devascularization and splenectomy. Restoration of portal perfusion, as a result of this high incidence of shunt thrombosis, may account for the lower encephalopathy rate attributed to this shunt.

Total portasystemic shunts are no longer widely used. Their position in the armamentarium against variceal bleeding has to a large extent been usurped by the TIPS procedure. Nevertheless, historically they have played an important role in surgical control of variceal bleeding with a combined success rate that exceeded 90 per cent over the long term in most patient populations. The reduction of ascites formation by total shunts has also been valuable. The primary disadvantage of total shunts has been the significant encephalopathy rate approaching 30 to 40 per cent as a result of hepatic deprivation of portal blood flow. Patient outcome has ultimately been determined by the severity of the underlying liver disease and its response to loss of portal perfusion. As a result, long-term survival rates are not improved by any of these shunt procedures as the mechanism of death changes from variceal bleeding and hypovolemic shock to progressive hepatic failure. Despite these difficulties, current recommendations for the use of total shunts include the following: (i) control of acute variceal bleeding in the emergency setting if a TIPS is unavailable or technically impossible; (ii) as a long-term bridge to hepatic transplantation in patients with recurrent variceal bleeding unresponsive to pharmacologic agents, endoscopic therapy, or a TIPS, and who are not candidates for a selective shunt; (iii) as a method of surgical decompression to relieve intractable ascites in patients where a TIPS cannot be performed; (iv) in patients who bleed from rectal, colonic, or stomal varices; and (v) for acute decompression of the congested liver with necrosis resulting from Budd–Chiari syndrome.

Partial shunts

Theoretically, partial portacaval or mesocaval H-graft shunts attempt to accomplish the same goal as selective shunts, which is the decompression of the varices while maintaining portal perfusion. The operation, as described by Sarfeh, is the placement of a small, 8- or 10-mm, interposition portacaval shunt of ringed PTFE graft combined with ligation of the coronary vein and other collateral vessels (Fig. 5). Postoperatively portal perfusion is preserved in 80 per cent of the patients with an 8-mm H-graft, as compared with only 50 per cent of those patients in whom a 10-mm graft is placed. Operative mortality is quite acceptable at 6 per cent. Control of variceal bleeding is equivalent to other surgical shunts at approximately 90 per cent. Shunt stenosis, or occlusion, approaches 20 per cent, but overall patency rate in general is excellent when those stenosed or occluded shunts are managed by radiographic interventional techniques. The encephalopathy rate is 10 to 15 per cent in reported series, once again suggesting the validity of maintaining some portal perfusion in order to prevent encephalopathy and hepatic failure. On the surface, it would appear that the TIPS and the small-bore portacaval H-graft shunt should be similar in their efficacy. A recently reported randomized trial comparing the TIPS with 8-mm H-graft portacaval shunts shows that portal pressures are reduced by both procedures, but the reduction is more significant in the surgical shunt. Major variceal rebleeding occurred in 11 per cent of the patients after TIPS placement, but did not occur in the surgically placed small-bore shunt group. Finally, they have also documented a much higher incidence of shunt occlusion in the TIPS groups when compared with the H-graft shunt group. It was concluded that the surgically placed partial shunt may have some advantages over the radiographically placed TIPS, but further studies are warranted.

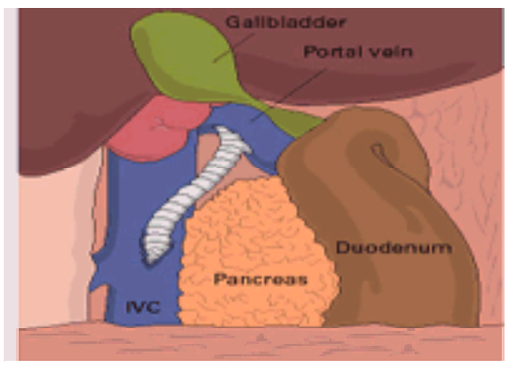


Fig. 5. Small-bore H-graft portacaval shunt (partial shunt).

Selective shunts

Selective shunts are those shunt procedures which are designed to decompress the gastroesophageal varices selectively while maintaining portal hypertension and prograde portal vein blood flow into the liver. The most common of these operations is the distal splenorenal shunt reported by Warren and colleagues in 1967. Variations of this shunt exist, including the distal splenorenal shunt with splenopancreatic disconnection and the distal splenocaval shunt. The Inokuchi shunt, also known as the coronary caval shunt, is another type of selective shunt. The concept of selective variceal decompression was brought about by the concern over the increased incidence of hepatic encephalopathy and progressive liver failure in patients undergoing total shunts. It was postulated that by maintaining portal perfusion to the cirrhotic liver, the risk of hepatic encephalopathy and liver failure would be ameliorated while at the same time excellent control of variceal bleeding would be obtained. Ultimately, it was hoped that the long-term survival rate after selective shunting would be improved in the cirrhotic population.

Distal splenorenal shunt (DSRS)

The distal splenorenal shunt has been used in clinical practice for some 30 years now, and is the most commonly performed selective shunt (Fig. 6). The DSRS consists of an anastomosis between the distal end of the splenic vein to the side of the left renal vein. This results in selective decompression of the gastroesophageal varices through the short gastric veins entering into the spleen, across the splenic parenchyma, out the splenic vein, and into the left renal vein. In addition, ligationof the coronary vein and the gastroepiploic vein along thegreater curvature of the stomach serves to disconnect the superior mesenteric and portal veins from the gastrosplenic components of the splanchnic venous circulation, leading to maintenance of portal hypertension and prograde portal blood flow into the liver. The DSRS is contraindicated when anatomically it cannot be performed due to splenic vein thrombosis or if the patient has medically intractable ascites.

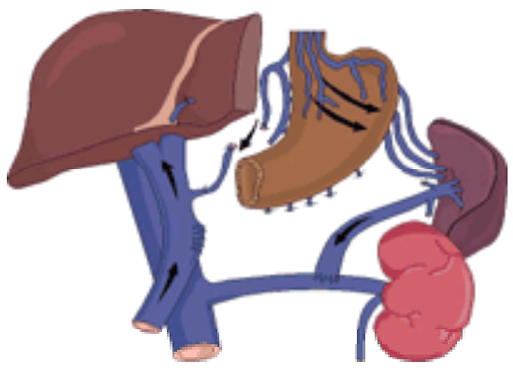


Fig. 6. Distal splenorenal shunt.

Suitable anatomy to perform the DSRS must be defined by preoperative angiography. A selective splenic arteriogram with venous phase visualizes the size and course of the splenic vein. Splenic vein thrombosis precludes the performance of the DSRS. Additionally, a small splenic vein of 5 to 6 mm, while it may not be a direct contraindication to the DSRS, does predict a higher thrombosis rate, as well as a higher early rebleed rate due to ineffective variceal decompression through the small

shunt. Selective cannulation of the left renal vein is important to help determine the juxtapositioning between the splenic vein and the left renal vein. Left renal vein venography also defines any anatomic variance, such as a retroaortic or circumaortic left renal vein which may lead to a more difficult DSRS. If a completely retroaortic vein is identified, the anterior branch of a circumaortic vein is small, or there is retrograde flow down the gonadal vein suggesting left renal vein outflow obstruction, then a distal splenocaval shunt to insure adequate trans-splenic decompression of the gastroesophageal varices may be necessary. Finally, a selective arteriogram of the superior mesenteric artery with venous phase defines the presence of portal vein or superior mesenteric vein thrombosis, and assists in identifying the size and number of coronary vein collaterals which it is necessary to ligate to insure a complete selectivity of the DSRS.

The DSRS is performed through a bilateral subcostal incision. The lesser sac is entered through the gastrocolic ligament, by ligating the gastric epiploic veins along the greater curvature of the stomach. Criss-crossing the gastroepiploic vessels and their tributaries directly on the stomach several times ensures an adequate porta-azygous disconnection and partial gastric devascularization at this level. The entire greater curve is taken down from the pylorus to the level of the short gastric veins, which are left intact as they are a critical component in the transplenic decompression of the esophageal and gastric varices. The splenocolic ligament is also carefully taken down to allow greater access to the proximal splenic vein and tail of the pancreas, particularly if a complete splenopancreatic disconnection is to be performed. The stomach is then retracted anteriorly and superiorly while the transverse colon is retracted inferiorly. Once the lesser sac has been entered, the groove between the pancreas and the fourth portion of the duodenum is identified, and the retroperitoneal tissues are incised longitudinally along the inferior border of the pancreas. Several small retroperitoneal veins and lymphatics should be ligated. As dissection continues, the pancreas is elevated anteriorly and rotated superiorly, and the splenic vein is identified along the posterior aspect of the pancreas. The junction of the splenic vein with the superior mesenteric vein is defined and the splenic vein is encircled here with a vessel loop (Fig. 7). The anterior surfaces of the superior mesenteric vein and portal vein are usually free of any venous tributaries, which allows for easy separation from the undersurface of the pancreas, as in the Whipple procedure. This allows visualization of any coronary veins that sometimes enter the superior crotch of the splenic vein and portal vein junction. Care must also be taken to ligate a small branch of the superior mesenteric artery that feeds into the neck of the pancreas just inferior to where the splenic vein joins the superior mesenteric vein. Dissection of the splenic vein from its groove in the pancreas starts at the junction of the splenic vein with the superior mesenteric vein and is directed toward the tail of the pancreas. This is the most difficult aspect of the shunt. It is important to free the posterior aspect of the splenic vein completely from the retroperitoneum before attempting this dissection. The splenic vein should be cleared of all perivenular tissues, so that dissection occurs directly on the vein. This allows for the exact identification and careful ligation of the multiple pancreatic venous tributaries that one will encounter as one removes the splenic vein from the pancreatic groove (Fig. 8). Unless the splenopancreatic disconnection is to be performed, dissection to the distal third of the pancreas usually yields sufficient length of vein for the creation of a tension-free anastomosis.

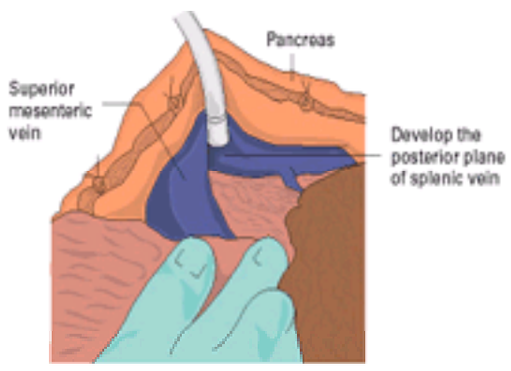


Fig. 7. DSRS: dissection of the junction of the splenic and superior mesenteric veins.

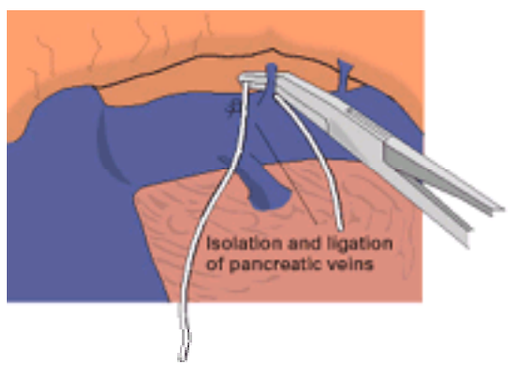


Fig. 8. DSRS: dissection of the splenic vein out of the pancreatic groove.

Attention is then turned to the retroperitoneum where the left renal vein is to be identified as it courses over the anterior surface of the aorta. Careful study of the preoperative angiograms enables one to find the left renal vein more readily. The retroperitoneal tissues between the pancreas and duodenum are incised over the aorta, and adjacent lymphatics are meticulously ligated. The left renal vein is identified and dissected free from surrounding lymphatics and fatty tissues. The left adrenal vein, entering the superior aspect of the left renal vein, is ligated in continuity and divided. This allows the left renal vein to be completely mobilized and elevated out of the retroperitoneum from the renal hilum to its junction with the inferior vena cava. The left adrenal vein is occasionally quite large and identified before the left renal vein. While it can be traced inferiorly to its junction with the left renal vein, the temptation to use it as part of the shunt should be resisted. Care must be taken not to injure the small descending lumbrical veins that exist off of the inferior aspect of the left renal vein. The gonadal vein is usually left intact to serve as an additional outflow tract for decompression unless it becomes necessary to ligate it for additional mobilization of the left renal vein.

The distal splenic vein is divided near its junction with the superior mesenteric vein, and the distal stump is oversewn with 5–0 vascular suture. The left renal vein is elevated out of the retroperitonum, clamped with a partially occluding vascular clamp, such as a Satinsky or Kay–Lambert clamp, and a longitudinal venotomy made. The splenic vein is then anastomosed to the anteriosuperior surface of the left renal vein taking care to avoid kinking or twisting of the splenic vein. It is usually necessary to trim a portion of the splenic vein in order to prevent redundancy of the vein. The posterior row of the anastomosis is carried out with a running 5–0 vascular suture (Fig. 9). The anterior row is completed with interrupted 5–0 vascular sutures to allow the shunt to balloon open when flow is restored. This anterior row of interrupted sutures also allows for early postoperative shunt angioplasty if postoperative angiography reveals any type of major shunt stenosis or gradient. Following completion of the actual shunt (Fig. 10), the coronary azygous disconnection is performed by identification and ligation of the coronary vein and several of its tributaries. This vein is sometimes ligated during the splenic vein dissection. Otherwise, it is found along the superior aspect of the pancreas coursing toward the lesser curvature of the stomach just anterior to the left gastric artery. In highly selected patients consisting primarily of those with Childs A and B cirrhosis, the overall operative mortality rate in nine published series of the DSRS is 6 per cent. The variceal rebleeding rate is 5.9 per cent, and the shunt thrombosis rate is 4.4 per cent. The overall incidence of hepatic encephalopathy at a median follow-up of 4 years is 15 per cent. The majority of these episodes are mild and can be managed medically with protein restriction, lactulose, and/or neomycin. While it has been well documented in the literature that overall long-term survival rate after a DSRS is not improved, there is a tendency toward improved survival in patients with good hepatic function. This is particularly true if they have non-alcoholic cirrhosis. Abstinence from alcohol in patients with compensated alcoholic cirrhosis may also yield a survival advantage after a DSRS. As with total shunts, the survival for patients undergoing selective shunting ultimately depends upon the underlying liver disease with 1-year survival rates ranging from 75 to 90 per cent, and 5-year survival rates ranging from 50 to 75 per cent. In the Emory experience of the DSRS, during the era of transplantation, a consecutive series of 147 DSRS in patients with Childs A and B cirrhosis achieved a 91 per cent 1-year and 77 per cent 3-year survival rate. The key in obtaining good results after a DSRS is to select patients with adequate liver function in whom there is a reasonable life expectancy from their liver disease.

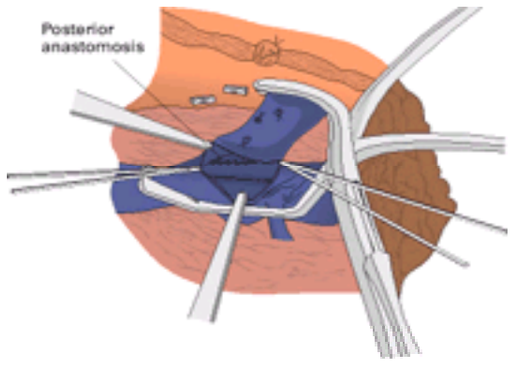


Fig. 9. DSRS: creating the end-to-side splenorenal anastomosis.

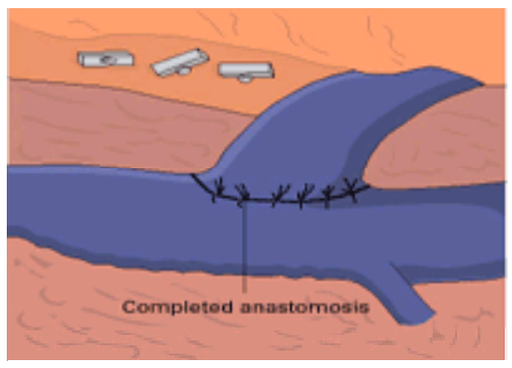


Fig. 10. DSRS: the final appearance of the splenorenal anastomosis with a gentle curve out of the pancreas.

Most postoperative complications following the DSRS are similar to those of the other shunts including intra-abdominal bleeding requiring re-exploration (approximately 1 per cent), and the usual infections and sepsis. Early postoperative liver failure and encephalopathy should be relatively low when compared with total shunts if the patient has been well chosen. There are, however, a few postoperative complications that are relatively unique to the DSRS. As mentioned earlier, any degree of ascites is aggravated by the DSRS due to maintenance of portal hypertension. Ascites is further exacerbated by the volume expansion that occurs during general anesthesia, as well as by the extensive porta-azygous disconnection as part of the DSRS, which leads to an additional increase in portal hypertension. The use of sodium-free fluids and a salt-restricted diet postoperatively along with the early institution of spironolactone and the judicious use of furosemide is almost always necessary. Failure to obtain diuresis of the ascites in association with an elevation of the white blood cell count, abdominal tenderness, and fever should alert one to the possibility of subacute bacterial peritonitis and/or chylous ascites. Chylous ascites, which occurs from the disruption of retroperitoneal lymphatics during left renal vein dissection, is particularly problematic. Treatment often requires multiple large-volume paracenteses, a very low fat diet (< 20 g), and parenteral nutrition before slowly resolving over a 4- to 6-week period. A temporary peritoneovenous shunt has been necessary on rare occasions. Another complication specific to the DSRS is portal vein or splenic vein thrombosis. Fortunately, portal vein thrombosis usually does not cause major problems and resolves spontaneously over several weeks, especially if it is non-occlusive. Splenic vein thrombosis, of course, leads to loss of the DSRS. If discovered early in the postoperative period, re-exploration in an attempt to declot the splenic vein, correct any technical problems, and preserve the shunt, is warranted. A splenectomy and gastric devascularization is necessary if shunt patency cannot be restored in order to prevent recurrent bleeding. Another even more rare complication is gastric paresis and transverse colon atony. The temporary loss of peristalsis in these organs is thought to represent a retraction injury that occurs in an effort to expose the pertinent structures of the lesser sac. Management is primarily medical until stomach and colon function return. Surprisingly, acute pancreatitis is an unlikely occurrence despite the significant manipulation of the pancreas during the DSRS.

Distal splenorenal shunt with splenopancreatic disconnection

In patients with non-cirrhotic portal hypertension, which includes extrahepatic portal vein thrombosis and primary hepatic fibrosis, selective variceal decompression by the DSRS represents optimal surgical therapy because postoperative portal perfusion is maintained for years and survival is improved. This also appears to be true for patients with well compensated non-alcoholic cirrhosis and schistosomiasis. In the case of alcoholic cirrhosis, portal perfusion is unfortunately lost in 60 per cent of the patients within 1 year after the DSRS, and long-term survival is no better than that achieved bytotal portasystemic shunts. In 1984, Warren proposed that this loss of portal perfusion after the DSRS occurred primarily via a system oftranspancreatic collaterals. This is a so-called ‘pancreatic siphon' developed between the high-pressure portal system and the low-pressure splenorenal anastomosis as the portal blood ‘sought' any available resistance pathway around the increased resistance bed of the diseased liver (Fig. 11). Other pathways, such as the transgastric and transcolonic collaterals, seemed to play a lesser role in this phenomenon. Loss of hepatic trophic factors such as insulin from the liver through this pancreatic syphon, and the resultant decrease in liver size and function, was hypothesized as the main reason why the DSRS did not prolong survival in those with alcoholic cirrhosis. It was projected that if the splenic vein was completely separated from the pancreas during the DSRS, also known as the splenopancreatic disconnection, then the pancreatic syphon would be disrupted and the postoperative portal perfusion and selectivity of the operation would be preserved (Fig. 12). Essentially, the DSRS with splenopancreatic disconnection evolved as an effort to achieve more effective selective shunting.

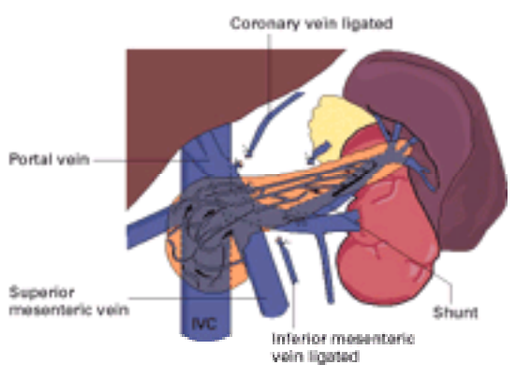


Fig. 11. The pancreatic siphon diverting portal blood flow from the high-pressure portal system to the low-pressure distal splenorenal shunt.

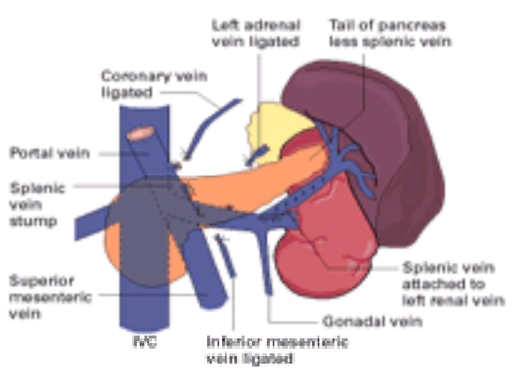


Fig. 12. The distal splenorenal shunt with splenopancreatic disconnection.

Shortly after Warren's observations were published, Inokuchi described similar collateralization after selective shunting which he termed portal malcirculation and corroborated its physiologic significance to the loss of portal perfusion. In 1989, a 4-year assessment of the DSRS with splenopancreatic disconnection in 78 patients was published. This study verified portal perfusion was maintained in the alcoholic patient, as well as the non-alcoholic patient, when a DSRS with splenopancreatic disconnection was performed. Hepatic function in the alcoholic group was maintained as well, along with continued excellent control of variceal bleeding. Survival data showed an operative mortality rate of 6.4 per cent despite the increased complexity of the procedure, and an overall 4-year survival rate of 70 per cent with no significant difference between those with alcoholic and non-alcoholic cirrhosis. It was concluded that the DSRS with splenopancreatic disconnection significantly improved maintenance of portal perfusion and survival in patients with alcoholic cirrhosis requiring selective shunt for variceal bleeding when compared with the standard DSRS. In patients with non-alcoholic cirrhosis, the data did not show the addition of splenopancreatic disconnection to offer any advantage over the standard DSRS.

The technical aspects of the DSRS with splenopancreatic disconnection document this procedure to be more difficult than the DSRS alone. The abdomen is entered through a bilateral subcostal incision. After dividing the gastrocolic ligament from the pylorus to the short gastric vessels, the splenocolic ligament is taken down. This exposes the hilum of the spleen, and provides access to the lesser sac, splenic vein, and tail of the pancreas while interrupting the major collateral pathway between the splenic vein and the mesentery of the splenic flexure. Splenopancreatic disconnection is the complete dissection of the splenic vein out of the pancreas from its junction with the superior mesenteric vein to its bifurcation at the splenic hilum. Dissection of the splenic vein should proceed directly on the vein with the posterior inferior surface of the vein isolated before the anterior superior surfaces are approached. The junction of the splenic vein and superior mesenteric vein should be controlled early in the dissection. Small splenopancreatic perforating veins are ligated and divided as they are encountered. Division of the splenic vein at its junction with the superior mesenteric vein will facilitate dissection of the splenic vein from the pancreas as does intermittent clamping of the splenic artery to decrease splenic blood flow. Superiorly rotating the pancreas with the surgeon's left hand behind the gland aids exposure as the splenic vein passes through the upper border of the pancreas, also known as the pancreatic groove. In most cases, the splenic vein courses above the pancreas for several centimeters before entering the hilum of the spleen. Controlling the vein above the pancreas requires special care, since injury to the vein at the hilum may cause irreversible harm and loss of the ability to complete the shunt. If dissection of the splenic vein from the pancreas is difficult, complete mobilization of the spleen from its bed, with rotation of the spleen and pancreas medially, will allow dissection of the splenic vein from a posterior approach. The splenorenal anastomosis is performed in the previously described manner along with interruption of the coronary veins.

This advancement in the DSRS obviously represents an increased technical challenge. The following observations are pertinent to the performance of the procedure. The more normal the pancreas, the easier the dissection. Unfortunately, the pancreas is often hard and indurated in alcoholic patients who benefit most from this procedure. The length of the pancreatic tail and its relationship to the splenic vein at the hilus is also variable. A short pancreas makes for an easier DSRS with splenopancreatic disconnection. The quality of the vein may also determine the ability to complete the dissection. A thin-wall vein with fine tributaries is more difficult to disconnect totally. The splenic vein in patients who have had repeated endoscopic variceal sclerotherapy frequently has a periphlebitis, which also makes it more difficult to dissect. Additionally, any injury to the splenic vein at the hilus is a more serious injury than an injury at the portal vein junction. The former may be irreparable and necessitate splenectomy with no shunt, while the latter can usually be dealt with by dissecting out more vein. Finally, the long-term benefit of maintaining portal perfusion of the liver and improved survival appears to justify the extra operative time and risk when performed by experienced hands.

Selective distal splenocaval shunt

Preoperative radiographic evaluation may identify certain anatomic and technical factors which make the distal splenorenal anastomosis difficult if not impossible. The left renal vein may not be suitable due to congenital anomalies, such as a completely retroaortic vein, or a circumaortic renal vein with a very small anterior component. Prior surgery, such as a left nephrectomy or left adrenalectomy, decreases the feasibility of a splenorenal anastomosis. In patients with massive splenomegaly, the splenorenal anastomosis is made technically challenging by a long torturous splenic vein that is juxtaposed directly in front of or even below the left renal vein. Finally, a left renal vein requiring lumbar or gonadal vessels as major outflow tracts secondary to renal vein stenosis can be anticipated to develop renal vein hypertension as the extra volume of splenic vein flow enters the left renal vein once the splenorenal anastomosis is formed. Temporary inadequate variceal decompression is the result. These problems can be overcome, and the principles of selective shunting maintained, by constructing a direct splenocaval anastomosis when the left renal vein is not suitable for a DSRS.

The operation is similar to the DSRS with several technical modifications. Usually it is necessary to dissect a greater amount of splenic vein out of the pancreas in order to span the extra distance to the inferior vena cava. If the splenic vein is not long enough, then an interposition graft may be required. The internal jugular vein is a suitable graft, and the left neck can be prepped for harvesting of this vein in anticipation of performing a distal splenocaval shunt. Exposure of the infrarenal inferior vena cava requires taking down the ligament of Treitz to free the fourth part of the duodenum caudally. Dissection then crosses in front of the aorta and behind the superior mesenteric artery to identify the inferior vena cava in the retroperitoneum. The inferior vena cava is mobilized from the right gonadal vein to the left renal vein for side-clamping. The anastomosis is fashioned like a conventional DSRS and must be done without tension. After removing a 1-cm button from the left anteriolateral surface of the inferior vena cava, the superior and inferior corners of the splenic vein are anchored. The back row is sewn with a 5–0 continuous vascular suture, and the anterior row is completed by using interrupted 5–0 vascular sutures. Mortality, shunt patency, maintenance of portal perfusion, and control of variceal bleeding are not different from that obtained with a conventional DSRS. Therefore, the principal physiologic goals of selective shunts, decompression of the gastrosplenic compartment to control variceal bleeding and maintenance of portal perfusion, are accomplished by the distal splenocaval shunt.

The coronary–caval shunt

Almost concurrently with the development of the distal splenorenal shunt, the concept of selective decompression was also being studied by Inokuchi in Japan. Because of the observed anatomic differences in his patients, primarily with post-hepatitis cirrhosis, Inokuchi was able to develop selective decompression of the gastroesophageal varices by creation of the coronary–caval shunt, also termed the left gastric venacaval shunt (Fig. 13). This operation can be performed frequently in Japan because the coronary vein tends to be a single large tributary feeding into the gastroesophageal junction and is more suitable for dissection and anastomosis. It is rarely performed in the Western setting because the coronary vein is rarely large enough in diameter or strong enough to hold sutures because of its thin-walled nature. Nevertheless, this shunt is mentioned because it supports the concept of selective variceal decompression in maintaining portal perfusion and liver function while successfully decompressing gastroesophageal varices and preventing bleeding. Operative mortality rate is 3 per cent and the variceal rebleeding rate is approximately 8 per cent. Survival rates at 1, 5, and 10 years are 90, 65, and 60 per cent, respectively.

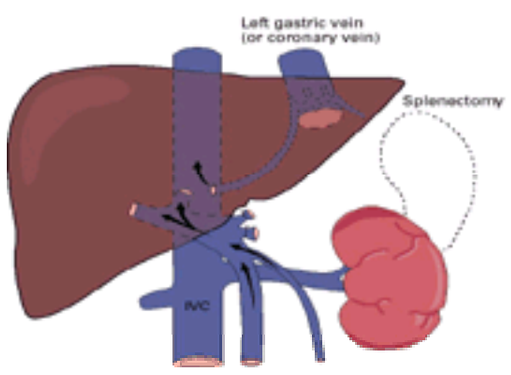


Fig. 13. The coronary–caval shunt. The gastroesophageal varices are selectively decompressed into the inferior vena cava. A splenectomy is also performed.

Definitive treatment for variceal bleeding

As the twenty-first century becomes a reality, the definitive treatment of variceal bleeding is now considered to be a multidisciplinary approach involving both non-operative and operative alternatives. Non-operative alternatives include: pharmacotherapy, balloon tamponade, endoscopic sclerotherapy or banding, and the transjugular intrahepatic portasystemic shunt (TIPS). Operative approaches include those that have been discussed here under the categories of total and selective shunts, as well as devascularization procedures. Ultimately, hepatic transplantation provides the most ideal means for portal decompression while restoring hepatic

function to normal. Unfortunately, with a limited donor pool, it is important to utilize alternative therapies to control variceal bleeding in patients with well compensated liver disease. The question arises as to what role each of these modalities play in the armamentarium against variceal bleeding. Through the years, clinical experience has shown that these various treatments are not necessarily competing therapies, despite what randomized trials in the literature might lead one to believe. Instead, each therapy has a distinct role. The selection of a specific therapy is based on numerous factors including: type of liver disease present, underlying hepatic function, location of the bleeding varices, candidacy for liver transplant, where the patient lives, access to tertiary health care, and finally, the associated operative risks and comorbid factors.

Ultimately, the definitive therapy chosen should balance the risk of rebleeding should that therapy fail compared with the risk of liver failure brought on by the institution of that therapy, and/or by the natural history of the underlying liver disease. Numerous management algorithms have been devised which have taken into consideration all of these factors. Reference is made to a tremendously detailed algorithm by Dr Layton Rikkers, a leader in the field of portal hypertensive surgery and hepatic transplantation, in *Current practice of surgery*, Vol. 3, 1995. A more simplified algorithm is offered in Fig. 14. These algorithms emphasize the necessity for a multidisciplinary team approach and, if followed, ultimate control of the patient's bleeding varices can be quite successful. Certain guiding primary principles have been utilized in the generation of these algorithms for the prevention of recurrent variceal bleeding. They are the following: (i) full evaluation after stabilization from the acute variceal bleeding episode; (ii) institution as first-line therapy both b-blockade and endoscopic banding or sclerosis, which should lower rebleeding rate from 80 per cent without treatment to approximately 40 per cent; (iii) radiographic decompression (TIPS) of those patients who fail first-line therapy if liver function is poor; (iv) surgical decompression, preferably with selective shunting, for first-line therapy failures if liver function is good; (v) liver transplantation for appropriate candidates with endstage liver disease utilizing a TIPS as a bridge to transplantation if first-line therapy fails to control recurrent variceal bleeding; and (vi) utilization of a TIPS as a salvage procedure in the patient with recurrent variceal bleeding and poor liver function who is not a candidate for liver transplantation.

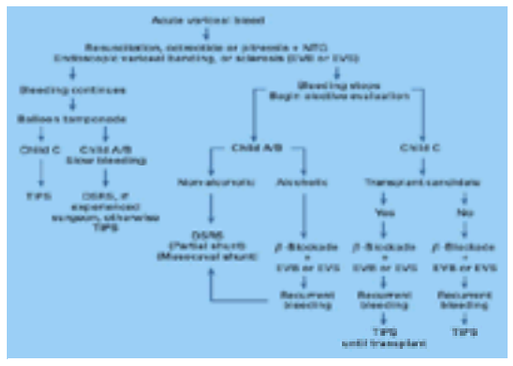


Fig. 14. Algorithm for definitive therapy of variceal bleeding.

It should be noted that the above fourth guiding principle reflects a firm belief in the importance of maintaining portal perfusion in order to prevent post-shunt encephalopathy and progressive liver failure. However, the data supporting the concept of selective shunting over total shunts is mixed at best. There are many reasons for this, but it is reasonably clear that all surgical shunts prevent recurrent variceal bleeding at least 90 per cent of the time, encephalopathy rates are somewhat less in selective shunts than total shunts (10 to 15 compared with 30 to 40 per cent), and survival rates are not significantly different between selective and total shunts. It is further apparent that patients with well compensated liver function can do well in the short term regardless of the type of shunt performed. Experienced portal hypertensive surgeons know this intuitively and, therefore, tend to perform the shunt with which they have the most experience. Perhaps one of the most important questions remaining is whether or not the TIPS in its avoidance of major surgical stress offers any short-term and/or long-term advantages over selective shunting in patients with good-function cirrhosis. Hopefully, an ongoing, multicenter, randomized controlled trial will yield a definitive answer.

Finally, what about those circumstances when patients with variceal bleeding do not have timely access to a well-defined, multidisciplinary approach for the definitive management of their variceal bleeding? In these, hopefully, rare instances, the original recommendation remains that the patient be treated with whatever modality the surgeon or physician has most readily available to him and is most comfortable in performing. Indeed, this appears to be the approach that is already occurring in many non-academic medical centers where pharmacologic and endoscopic methods are used for acute variceal bleeding, and the TIPS is utilized for more long-term control. It should, however, be remembered that the shunt operations, as described herein, offer excellent control of variceal bleeding with very acceptable mortality and morbidity rates when performed by experienced hands. Unfortunately, they can be technically demanding and the number of surgeons comfortable in performing them, particularly in the emergency setting, is continuing to decline.

Further reading

Atta HM *et al.* Selective splenocaval shunt. Report of 26 cases and review of the literature. *Archives of Surgery* 1991; **126**: 582–5.

D'Amico G, Pagliaro L, Bosch J. Treatment of portal hypertension: a meta-analysis review. *Hepatology* 1995; **22**: 332–6.

Galloway JR, Henderson JM. Management of variceal bleeding in patients with extra-hepatic portal vein thrombosis. *American Journal of Surgery* 1990; **160**: 122–7.

Grace ND. A hepatologist's view of variceal bleeding. *American Journal of Surgery* 1990; **160**: 26–31.

Grace N. The side to side portal caval shunt revisited. *New England Journal of Medicine* 1994; **330**: 208–12.

Henderson JM. Liver transplantation for portal hypertension. *Gastroenterology Clinics of North America* 1992; **21**: 197.

Henderson JM. Surgical measures in prevention of recurrent variceal bleeding. *Gastroenterology and Endoscopy Clinics of North America* 1992; **2**: 151–66.

Henderson JM *et al.* Distal splenorenal shunt with splenopancreatic disconnection. A four year assessment. *Annals of Surgery* 1989; **210**: 332–41.

Henderson JM *et al.* Endoscopic variceal sclerosis compared with distal splenorenal shunt to prevent recurrent variceal bleeding in cirrhosis. A prospective randomized trial. *Annals of Internal Medicine* 1990; **112**: 262–9.

Henderson JM *et al.* Selective shunt in the management of variceal bleeding in the era of liver transplantation. *Annals of Surgery* 1992; **216**: 248–55.

Henderson JM, Barnes DS, Geisinger MA. Portal hypertension. *Current Problems in Surgery* 1998; **35**: 379–452.

Inokuchi K. Present status of surgical treatment of oesophageal varices in Japan: a nation wide survey of 3588 patients. *World Journal of Surgery* 1985; **9**: 171–80.

Inokuchi K, Sugimachi K. The selective shunt for variceal bleeding: a personal perspective. *American Journal of Surgery* 1990; **160**: 48–53.

Langer B, ed. Treatment of portal hypertension in 1994—State of the art. *World Journal of Surgery* 1994; **18**: 169–258.

Maffei-Faccioli A *et al.* Selective variceal decompression and its role relative to other therapies. *American Journal of Surgery* 190; **160**: 60–6.

Mazzaferro V *et al.* Liver transplantation in patients with previous portasystemic shunt. *American Journal of Surgery* 1990; **160**: 111–16.

Mikkelson WP. End-to-side portacaval anastomosis. In: Malt RA, ed. *Surgical techniques illustrated*, pp. 436–41. WB Saunders, Philadelphia, 1985.

Millikan WJ *et al.* Surgical rescue for failure of cirrhotic sclerotherapy. *American Journal of Surgery* 1990; **160**: 117–21.

Orloff MJ, Bell RH. Long-term survival after emergency portacaval shunting for bleeding varices in patients with alcoholic cirrhosis. *American Journal of Surgery* 1986; **151**: 176–83.

Ottinger LW. The Linton splenorenal shunt in the management of bleeding complications of portal hypertension. *Annals of Surgery* 1982; **196**: 664–8.

Planas R *et al.* Portacaval shunt versus endoscopic sclerotherapy in the elective treatment of variceal hemorrhage. *Gastroenterology* 1991; **100**: 1078.

Rikkers LF. Is the distal splenorenal shunt better? *Hepatology* 1988; **8**: 1705–7.

Rikkers LF. Portal hypertension. In: Levine BA *et al.*, eds, *Current practice of surgery*, Vol 3. Churchill Livingstone, New York, 1995.

Rikkers LF, Jin G. Surgical management of acute variceal hemorrhage. *World Journal of Surgery* 1994; **18**: 193–9.

Rikkers LF *et al.* Shunt surgery versus endoscopic sclerotherapy for variceal hemorrhage. Late results of a randomized trial. *American Journal of Surgery* 1993; **165**: 27–33.

Sanfey H, Cameron JL. Mesocaval shunts. In: Rutherford RB, ed., *Vascular surgery*, 2nd edn, pp. 1029–51. WB Saunders, Philadelphia, 1984.

Sarfeh IJ, Rypins EB, Mason GR. A systematic appraisal of portacaval H-graft diameters: clinical and hemodynamic perspective. *Annals of Surgery* 1986; **204**: 356–63.

Sarr MG, Herlong HF, Cameron JL. Longterm patency of the mesocaval C-shunt. *American Journal of Surgery* 1986; **151**: 98–103.

Spina GP *et al.* Distal splenorenal shunt versus endoscopic sclerotherapy in the prevention of variceal bleeding. A meta-analysis of 4 randomized clinical trials. *Journal of Hepatology* 1992; **16**: 338–45.

Vargas HE, Gerber D, Abu-Elmagel K. Management of portal hypertension-related bleeding. *Surgical Clinics of North America* 1999; **79**: 1–22.

Warren WD, Millikan WJ. Selective transplenic decompression procedure: changes in technique after 300 cases. *Contemporary Surgery* 1981; **18**: 11–30.

Warren WD, Zeppa R, Forman JS. Selective transplenic decompression of gastro esophageal varices by distal splenorenal shunt. *Annals of Surgery* 1967; **16**: 437–46.

Warren WD *et al.* Splenopancreatic disconnection: improved selectivity of distal splenorenal shunt. *Annals of Surgery* 1986; **204**: 346–54.

32.4 Non-shunt procedures in management of variceal bleeding

George Hamilton

Introduction
Historical background
Indications
Preoperative preparation
Current non-shunt procedures
Oesophageal transection
Portoazygous disconnection
The Hassab procedure
The Sugiura procedure
Results of non-shunt procedures
Conclusion
Further reading

Introduction

Most episodes of variceal haemorrhage will be successfully treated by resuscitation and injection sclerotherapy, and indeed the efficacy of this approach has been confirmed by several well-controlled trials. Nevertheless, in a significant minority either acute bleeding will persist, or rebleeding will occur in the near or long term. In this situation there is little doubt that some form of portosystemic decompression is the most successful treatment for bleeding.

However, this objective is achieved at a cost, because disabling portosystemic encephalopathy and deterioration in liver function over subsequent years frequently develops in patients who have been treated by shunting. Also the operative mortality of emergency shunting procedures is high, even where hepatic reserve is good. Furthermore, in variceal bleeding from extrahepatic portal hypertension, 30 to 50 per cent of children and 40 to 50 per cent of adults will have extensive thrombosis of the splenic and mesenteric veins which makes decompressive procedures impossible. For all these reasons an alternative to portal decompression is required and clearly the repertoire of any surgeon involved in the treatment of variceal haemorrhage must include non-shunt procedures.

Historical background

Many different non-shunt approaches to the treatment of bleeding oesophageal varices have been developed ranging from the direct attack by oesophagogastrectomy to variations on the theme of portoazygous disconnection and these are listed in Table 1. Indeed, this problem has taxed the ingenuity of surgeons since the late nineteenth century when splenectomy and omentopexy were originally performed. The wide range of procedures devised since then reflects the generally unsatisfactory results achieved, particularly so in the presence of poor liver function. In recent years, the advent of stapling devices has reawakened interest in portoazygous disconnection as a suitable alternative to shunting. Currently, the importance of these procedures in comparison with injection sclerotherapy, transjugular intrahepatic portosystemic shunt (**TIPS**), and portocaval shunting is not known.

Splenectomy	Total gastrectomy
Omentopexy, splenopneumopexy, etc.	Hepatic and splenic artery ligation
Coronary vein ligation	Splenectomy and coronary vein ligation
Transthoracic ligation of varices (Boerema, Crile, and Milnes Walker)	Tanner's procedure (subcardiac portoazygos disconnection)
Boerema button	Hassab procedure
Stapled oesophageal transection	Sugiura procedure

Table 1 Non-shunt procedures for variceal haemorrhage

Indications

In most centres the first line of management of variceal haemorrhage is by resuscitation, medical therapy with a variety of vasoactive agents, balloon tamponade as a holding procedure, and injection sclerotherapy of the varices. Operative treatment is generally reserved for continued bleeding, or where recurrent variceal haemorrhage is a problem (Table 2). For example, it is now clearly established that if bleeding is not controlled after two sets of variceal injection, high rebleeding and mortality rates result from further attempts at injection; in this group emergency operative intervention is the best option. Increasingly over recent years TIPS has become the procedure of choice in this group of patients because of the possibly lower morbidity and mortality, and the preservation of a pristine operative field for subsequent liver transplantation. Indeed transplantation is the most effective treatment for liver failure complicated by bleeding varices. However, where liver transplantation is not indicated or available, TIPS may not be the best treatment option in view of the high incidence of portosystemic encephalopathy, and the significant rate of shunt stenosis.

1. Failed sclerotherapy: mortality becomes very high after the second unsuccessful injection and emergency oesophageal transection is highly effective
2. Absence of suitable veins for obtaining an extensive thrombosis of splenic and mesenteric veins causing extrahepatic portal hypertension
3. Mild liver dysfunction in the presence of splenomegaly and hypersplenism, e.g. mansonic schistosomiasis
4. High probability of portosystemic encephalopathy with a shunt, particularly where work demands a high level of skill or intellectual activity
5. Recurrent bleeding uncontrolled by injection in poor liver function, i.e. Child's grade C
6. Absence of expertise or facilities for emergency injection sclerotherapy: stapled oesophageal transection is an effective in controlling bleeding

Table 2 Indications for non-shunt procedures

Variceal bleeding in the presence of good hepatic function but a large spleen, for example in mansonic schistosomiasis, is a clinical scenario where a non-shunt procedure has many advantages over portal decompression. Usually massive splenomegaly and hypersplenism occur in a young patient with an excellent prospect of good lifelong hepatocellular function. Shunting not only carries an immediate high incidence of portosystemic encephalopathy but also, in non-selective decompressive procedures, an inexorable deterioration of liver function will take place over the years. This hepatocellular deterioration becomes highly relevant in patients with conditions such as schistosomiasis or extrahepatic portal hypertension where life expectancy and liver function are good. Oesophageal transection and oesophagogastric devascularization (Hassab and Sugiura procedures) offer low rates of rebleeding and advantages over shunting in terms of preserving liver function and avoiding portosystemic encephalopathy. Unfortunately the dearth of properly constructed trials comparing shunt and non-shunt techniques makes it impossible to be more objective about these two surgical approaches. Obviously, when the splenic and superior mesenteric veins are thrombosed there is no alternative to using a non-shunt technique. Portosystemic encephalopathy is a devastating complication adversely affecting intellectual function.

Patients with cirrhosis increasingly come to liver transplantation in the endstages of their disease. Previous surgery, particularly portocaval shunting, will complicate the procedure but will also increase its morbidity and mortality, and adhesions around the left lobe of the liver and throughout the left upper quadrant from non-shunt procedures will undoubtedly increase the technical problems of the recipient hepatectomy. Arguably, however, dealing with these adhesions might be less difficult than the dissection around the portal structures with the added problems of sudden massive portal hypertension which occurs during the disconnection of patent portocaval shunt in a transplant. At present, where transplantation is not yet indicated, and if TIPS is not available, a mesocaval shunt is the operation most favoured by transplanters for control of variceal haemorrhage because of the minimal dissection around the liver. Simple oesophageal transection, however, will reliably control haemorrhage where injection sclerotherapy has failed, with much less adhesion development, and in this emergency situation must be the procedure of choice.

As in shunting, the more major devascularization procedures have a higher operative mortality in the emergency situation, particularly in the setting of poor hepatic reserve. Once again, an expeditiously performed simple oesophageal transection, using the stapler but without major devascularization, has been shown to be extremely effective in treatment of haemorrhage uncontrolled by medical therapy or injection.

Preoperative preparation

Full investigation and assessment of liver disease and function is essential. Any clotting abnormality should be corrected, particularly in the presence of hypersplenism, and renal function assessed, monitored, and optimally maintained before, during, and after the procedure. Preoperative MR angiography may be of value if devascularization is to be performed, providing a ‘road map' of the collateral circulation around the oesophagogastric region. If splenectomy is to be performed, vaccination against pneumococcal infection is probably of value, as is maintenance low-dose penicillin cover for the perioperative period, and for at least 2 years afterwards.

Current non-shunt procedures

Techniques such as transthoracic ligation of oesophageal varices, Tanner's high gastric disconnection, and splenic artery and coronary vein ligation have little to offer over current medical and surgical treatment and have been largely abandoned. However, these methods find occasional use in the emergency situation, particularly where equipment and expertise may be lacking, and where transfer of a desperately ill patient across large distances to an experienced centre is not feasible. In the vast majority of centres, the currently favoured non-shunting procedures are simple stapled oesophageal transection, oesophagogastric devascularization, or a combination of the two.

Oesophageal transection

Interest in portoazygous disconnection was reawakened after Vanhemmel in 1974 first used a stapling device (the Russian SPTU stapling gun) in transection of the oesophagus for bleeding varices. With the advent of simpler more effective stapling guns (Autosuture and Ethicon) came a worldwide upsurge of interest in oesophageal transection, particularly in the emergency situation.

The procedure is performed via a left subcostal or upper midline incision. The left lobe of the liver is retracted and the oesophagogastric junction identified; retraction may be difficult in a rigid cirrhotic liver and in this case, mobilization of the left lobe may be necessary. The overlying peritoneal reflection is incised using diathermy to coagulate the multiple small vessels which develop in this layer as a result of portal hypertension. The oesophagus is mobilized with its surrounding para-oesophageal tissues, from the oesophagocardiac junction to the upper margins of the diaphragmatic crus, a length of approximately 5 cm. Difficulty may be experienced with this mobilization in patients who have previously undergone injection sclerotherapy, where thickening of the para-oesophageal tissues often develops. Despite this handicap it is always possible to mobilize the oesophagus without complication. Ideally the vagus nerves should be mobilized, although no postvagotomy complications were found in a recent large series from the Royal Free Hospital (London) where no specific attempt to mobilize the vagi was made. A strong linen thread is passed around the mobilized distal oesophagus and a short anterior gastrotomy is made. After assessing the lumen of the oesophagus with passage of calibrated obturators, the appropriate size of staple gun is passed into the oesophagus. The head is then separated from the anvil, and the linen thread tied tightly on to the shaft of the gun just above the oesophagogastric junction. After removal of any sling or tape around the oesophagus, the anvil and staple cartridge are approximated and the gun fired. In one manoeuvre the oesophagus is transected with removal of a 1-cm ring of oesophagus, and the ends reanastomosed with a secure double layer of steel clips (Fig. 1).

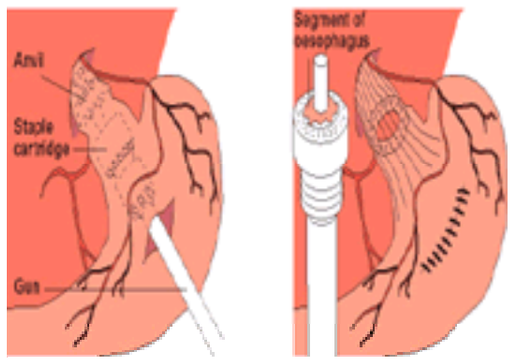


Fig. 1. Staple transection of the oesophagus. The head of the staple gun is introduced into the lower oesophagus via an anterior gastrotomy and a stout linen thread is tied firmly down on to the opened head of the gun about 1 cm above the oesophagocardiac junction. The head is closed on to the anvil and the gun fired, simultaneously transecting and removing a 1-cm ring of oesophagus and reanastomosing it with a double layer of staples. The ‘doughnut' of resected oesophagus is inspected to ensure completeness of the transection.

The gun is opened and the doughnut inspected, a complete ring indicating a technically satisfactory transection. A nasogastric tube is carefully guided through the anastomosis and its tip placed in the body of the stomach before closure of the gastrotomy. The abdomen is closed without drainage, the patient allowed to take small sips of water, and the nasogastric tube usually removed after 2 days before allowing return to normal oral intake.

This technically straightforward procedure can be performed within 1 to 2 h and usually with little blood loss. Varying degrees of devascularization can be added to the procedure with most authors advocating ligation of peri-oesophageal collaterals and the left gastric or coronary veins.

Portoazygous disconnection

The principle behind this surgical approach is the disconnection of the venous circulation of the distal oesophagus and cardia from the hypertensive portal circulation by division of all the feeding vessels. The most frequently employed operations are the Hassab devascularization and Sugiura devascularization with oesophageal transection.

The Hassab procedure

The Hassab procedure consists of devascularization of the upper half of the stomach and oesophagus (Fig. 2). The first step is usually splenic artery ligation followed by careful mobilization of the spleen. This mobilization, as in all dissections in portal hypertension, requires patient ligation and coagulation of multiple collaterals within the peritoneal reflections, and after individual ligation and division of the short gastric vessels, the spleen is removed. The whole proximal stomach is then devascularized from the terminal two branches of the left gastric artery at the incisura angularis upwards by ligation and division of the lesser and greater omentum, and of the posterior gastric adhesions. After division of the oesophagogastric reflection of peritoneum and mobilization of the vagi, the distal 7 to 8 cm of oesophagus is mobilized and all feeding vessels are ligated and divided. Exposure in this part of the procedure is much facilitated by the use of costal margin retractors. The distal 3 cm of oesophagus and proximal 5 cm of stomach may then be opened longitudinally, thus displaying the varices and allowing obliteration of each variceal column by undersewing from as high as possible within the oesophagus with an absorbable suture. After positioning of a nasogastric tube the oesophagogastrotomy is carefully closed by suturing or stapling. Some authors recommend closure by swinging a flap of stomach wall into the oesophageal defect, thus minimizing oesophageal stricturing.

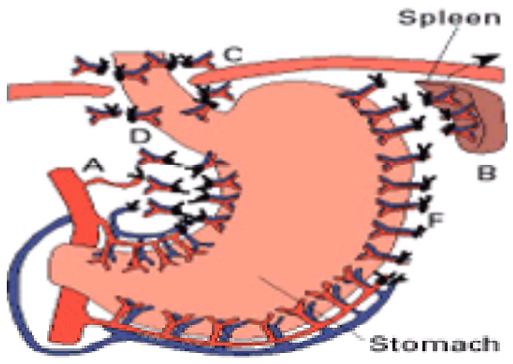


Fig. 2. The Hassab procedure. The distal oesophagus is mobilized by dissection via a midline or left transverse incision up to and including retraction or division of the diaphragmatic crura. (A) The left gastric or coronary vein is divided. (B) Splenectomy is performed. (C) The distal 8 to 10 cm of oesophagus is devascularized. (D and E) The lesser and greater curves are devascularized. The oesophagus can be opened and the variceal trunks underrun with absorbable suture; a pyloroplasty is sometimes also performed.

The Sugiura procedure

The Sugiura operation is a much more radical development of the above method, classically performed in two staged procedures. At the first operation, via a left thoracotomy, the distal intrathoracic oesophagus is devascularized and an oesophageal transection performed. Six weeks later, via an upper abdominal midline incision, the intra-abdominal oesophagus and proximal stomach are devascularized by lesser and greater curve division and splenectomy.

Vagotomy and pyloroplasty are then performed (Fig. 3). This massive procedure has been modified into a one-stage operation using a transabdominal approach facilitated by the use of costal margin and sternal retractors. After division of the crura of the diaphragm, 10 cm of oesophagus can be devascularized, a staple transection performed via a gastrotomy, and the rest of the abdominal part of the operation completed.

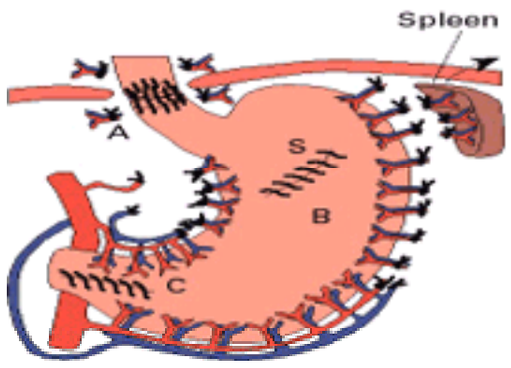


Fig. 3. The Sugiura procedure. This operation is performed either in two stages (thoracic then abdominal) or in one transabdominal procedure (modified Sugiura). A devascularization is performed as in the Hassab procedure, but with more extensive devascularization of the distal thoracic oesophagus; this can be achieved transhiatally without recourse to thoracotomy. A staple transection and usually pyloroplasty are then performed.

Results of non-shunt procedures

Because of high long-term rebleeding rates of up to 50 per cent, simple oesophageal transection without some form of devascularization is rarely performed as a definitive, elective procedure. However, in the emergency situation, transection alone has been shown to be highly effective at controlling variceal haemorrhage and in a randomized controlled trial from the Royal Free Hospital was found to have morbidity and mortality similar to injection sclerotherapy across all grades of liver disease. Rebleeding does not occur before 3 months, but when it does develop, is frequently not from varices at all but from ulceration at the transection line which is readily controlled by omeprazole. The incidence of portosystemic encephalopathy is very low and dysphagia, although frequent, is usually transient, but when persistent is cured by dilatation.

The major complications of devascularization procedures are those of the oesophageal transection, possibly compounded by devascularization ischaemia, namely dysphagia, ulceration, and stricture formation, but with an extremely low rate of rebleeding and recurrence of varices when compared with simple transection alone. All of these procedures are characterized by little or no increase in portosystemic encephalopathy.

The majority of series of devascularization procedures report a significantly high risk of rebleeding although there is considerable variability in reported results. The exception to this is Sugiura's results with his two-stage operation, wherein he reported an overall mortality of 4.9 per cent with a mortality of 13.3 per cent for emergency procedures. In a different series from Kuwait, Abouna reported a similar overall mortality of 7.7 per cent for his one-stage modification of the Sugiura technique. Both series had extremely low rebleeding rates of 1.5 and 3.4 per cent, respectively, with no increase in portosystemic encephalopathy. These results are the best reported for any treatment for oesophageal varices, but unfortunately no Western group has been able to reproduce them, largely it is claimed because of the higher preponderance of alcoholic cirrhosis and hepatitis in the West. Indeed the Western experience of the Sugiura technique is characterized by high morbidity and mortality. However, excellent results have been reported in the management of variceal bleeding in schistosomal portal hypertension where long-term liver function is good. Devascularization procedures, in particular the modified Sugiura, provide the treatment of choice for this condition.

Conclusion

Despite the success of injection sclerotherapy and TIPS, there remains a place for surgical treatment of bleeding oesophageal varices. In general, non-shunt procedures offer advantages over TIPS and portosystemic shunting mainly in terms of lower incidence of portosystemic encephalopathy and maintenance of portal perfusion of the liver. Simple oesophageal transection is a proved, effective means of controlling acute variceal haemorrhage, particularly where injection sclerotherapy has failed. In this emergency situation this simple approach has many advantages and is probably the treatment of choice. However, this method is less valuable in the elective situation because of its high rebleeding rate, and thus the modified Sugiura procedure has much to offer, although Western experience has been disappointing. Because there is so little objective clinical data and properly conducted comparison is not available, the choice of approach, shunt or non-shunt, is determined by the local disease patterns, local surgical expertise, and management bias.

Further reading

Abouna GM, Baissony H, Al-Nakib BM, Menkarios AT, Silva OSG. The place of Sugiura operation for portal hypertension and bleeding esophageal varices. *Surgery* 1987; **101**: 91–8. [Description and results of a modification of the Sugiura procedure in a large series of patients.]

Alam MK, el-Sayed GS, Abdulmajeed AM, al-Dohayan A. Effect of previous sclerotherapy on the outcome of gastro-oesophageal devascularisation and oesophageal transection in bleeding oesophageal varices. *British Journal of Surgery* 1996; **83**: 1702–5. [Sclerotherapy changes to the oesophago-gastric junction do not compromise staple transection or devascularization.]

Bornman PC, Terblanche J, Kahn D, Jonker MA, Kirsch RE. Limitations of multiple injection sclerotherapy sessions for acute variceal bleeding. *South African Medical Journal* 1986; **70**: 34–6. [In poor risk patients requiring more than two sessions of injection sclerotherapy for bleeding varices during the same admission, mortality is extremely high and other methods of control recommended.]

Burroughs AK, Hamilton G, Phillips A, Mezzanotte G, McIntyre N, Hobbs KEF. A comparison of sclerotherapy with staple transection of the esophagus for the emergency control of bleeding from oesophageal varices. *New England Journal of Medicine* 1989; **321**: 857–62. [Important randomized prospective comparison of sclerotherapy with transection.]

Chaudhary A, Aranya RC. Devascularisation following endoscopic sclerotherapy of oesophageal varices: dangers and difficulties. *British Journal of Surgery* 1991; **78**: 1249–51. [Sclerotherapy changes to the peri-oesophageal tissues found to be severe, resulting in cases of post-transection leakage and intraoperative perforation: severity of changes probably related to number of

sclerotherapy sessions.]

Gouge TH, Ranson JH. Oesophageal transection and paraoesophagogastric devascularization for bleeding oesophageal varices. *American Journal of Surgery* 1986; **151**: 47–54. [Poor results reported with the unmodified Sugiura procedure in an American series.]

Hassab MA. Gastro-oesophageal decongestion and splenectomy in the treatment of oesophageal varices in bilharzial cirrhosis: further studies with a report on 355 operations. *Surgery* 1967; **61**: 169–76. [Original report by Hassab of his procedure undertaken in a large series of patients.]

Hosking SW, Johnson AG. What happens to oesophageal varices after transection and devascularization. *Surgery* 1987; **101**: 531–4. [Description of the changes at the oesophago-gastric junction developing after injection sclerotherapy.]

Huizinga WK, Angorn PA, Baker LW. Esophageal transection versus injection sclerotherapy in the management of bleeding oesophageal varices in patients at high risk. *Surgery, Gynecology, Obstetrics* 1985; **160**: 539–46. [Prospective randomized study comparing sclerotherapy with transection, which found better outcome for surgery.]

Kaye GL, McCormick PA, Siringo S, Hobbs KEF, McIntyre N, Burroughs AK. Bleeding from staple line erosion after esophageal transection: effect of omeprazole. *Hepatology* 1992; **15**: 1031–5. [Rebleeding occurring after staple transection is due to staple line oesophageal erosions and can be controlled by omeprazole.]

Kollias J, Jeans PL, Padbury RT, Toouli J. Gastric devascularisation and splenectomy for bleeding gastric varices. *Australia and New Zealand Journal of Surgery* 1995; **11**: 804–7. [Gastric devascularization is effective in the treatment of bleeding oesophageal varices.]

Matory WE Jr, Sedgwick CE, Rossi RL. Non-shunting procedures in management of bleeding esophageal varices. *Surgical Clinics of North America* 1980; **60**: 281–95. [Early description of non-shunt procedures in preinjection era.]

McCormick PA *et al.* Esophageal staple transection as a salvage procedure after failure of acute injection sclerotherapy. *Hepatology* 1992; **15**: 403–6. [Emergency staple transection is an effective treatment for patients who rebleed despite injection sclerotherapy.]

Raia S, Mies S, Alfieri Jr F. Portal hypertension in mansonic schistosomiasis. *World Journal of Surgery* 1991; **15**: 176–87. [Brazilian experience of schistosomal portal hypertension confirming high incidence of portosystemic encephalopathy (PSE) after splenorenal shunt; no PSE found after devascularization procedures.]

Sugiura M, Futagawa S. Results of six hundred and thirty-six oesophageal transections with paraoesophagogastric devascularization in the treatment of oesophageal varices. *Journal of Vascular Surgery* 1984; **1**: 254–60. [Author's report of a huge series of oesophageal varices which were treated by the Sugiura technique.]

Tsimoyannis EC, Gossios KJ, Tsianos EV. Modified Sugiura's procedure may improve liver function in portal hypertension. *Hepatogastroenterology* 1997; **44**: 492–5. [Study on a small number of patients with bleeding oesophageal varices treated by the Sugiura procedure where postoperative improvement in hepatic function was found.]

32.5 Ascites and its complications

Robert M. Strauss and Jules L. Dienstag

Classification of ascites and differential diagnosis
Pathogenesis
Clinical features
Examination of the ascitic fluid
Management
Medical management
Surgical management
Spontaneous bacterial peritonitis
Pathogenesis
Predisposing factors
Bacteriology
Diagnosis
Treatment
Hepatorenal syndrome
Diagnosis
Treatment
Further reading

The term ascites is derived from the Greek word *askitos* (bladder, belly, bag) and denotes the presence of excessive fluid in the peritoneal cavity. It may present as an isolated clinical finding or be seen in the context of generalized edema.

Classification of ascites and differential diagnosis

Ascites may be caused by hepatic, renal, or cardiac decompensation; by hepatic venous outflow obstruction; and by malignant, inflammatory, or infectious processes in the peritoneum. Several different classification schemes have been proposed. Traditionally, one discriminating between transudative and exudative ascites has been adopted widely (Table 1). This classification implies different pathogenic mechanisms. It has been suggested by various groups that given the poor distinction between transudative and exudative ascites as defined by the ascitic fluid protein concentration, it may be more precise if ascitic fluid is defined as portal hypertensive and non-portal hypertensive. This distinction is based on the difference between the serum and ascites albumin concentrations and defines the pathogenesis of the ascites.



Table 1 Classification of ascites

Ascites is not the only cause of marked abdominal swelling. Gross obesity, gaseous distention, visceromegaly, or abnormal masses can all be confused with fluid excess. A massive ovarian cyst, hydatid cyst, and, exceptionally, pregnancy with or without polyhydramnios, may be confused with ascites.

Pathogenesis

Normally, there is just enough free fluid in the peritoneal cavity to lubricate the peritoneal surfaces. Ascites occurs when there is an imbalance of factors that favor the flow of fluid from the vascular space (usually representing a state of excess total body sodium) and/or when there is exudation of fluid through infection or malignant implantation on the peritoneum.

Chronologically, sodium retention is the earliest alteration in kidney function observed in patients with cirrhosis. The systemic circulation is also frequently disturbed in patients with cirrhosis. Plasma volume and cardiac output are increased while arterial pressure and systemic vascular resistance are reduced leading to a hyperdynamic circulatory syndrome. Cardiac output may increase two- to threefold and reach values similar to those found in patients with high-output cardiac failure or healthy adults during intense physical exercise. Arterial pressure and systemic vascular resistance may be low, similar to patients with septic shock. Normally, the higher pressure at the arteriolar end of the capillary allows passage of fluid without protein into the pericapillary space, while reabsorption takes place at the venous end of the capillary where hydrostatic pressure is lower than oncotic pressure. In cirrhosis, portal venous hypertension increases the filtration pressure at the capillary level with transudation of fluid, while low albumin levels decrease vascular oncotic pressure. Two potential mechanisms have been proposed to explain the pathogenesis of ascites. The ‘underfilling hypothesis' suggests that the primary abnormality is a sequestration of fluid within the splanchnic vascular bed secondary to portal hypertension and reduced oncotic pressure. These events cause a decrease in effective circulating blood volume, which leads to compensatory renal retention of sodium by increasing aldosterone production. Loss of fluid from the intravascular compartment results in reduced plasma volume which is sensed by cardiopulmonary and arterial receptors resulting in renal sodium and water retention to compensate for the reduced plasma volume. The retained fluid is unable to fill the intravascular compartment and suppress the sodium and water retaining signals to the kidney because fluid is constantly leaking into the peritoneal cavity thus creating a vicious cycle. Although this theory appears logical and accounts for several physiologic observations, it does not correspond well with systemic hemodynamic abnormalities typically associated with cirrhosis. For example, if this were entirely accurate, plasma volume and cardiac index would be reduced and systemic vascular resistance increased. In cirrhosis, the converse happens, plasma volume and cardiac index are markedly increased and systemic vascular resistance markedly reduced.

The ‘overflow hypothesis' suggests that renal dysfunction (perhaps mediated by a hepatorenal reflex) is the primary event leading to sodium and water accumulation in the absence of intravascular volume depletion. This hypothesis is supported by the fact that acute constriction of the portal vein in dogs is associated with renal vasoconstriction, renal sodium and water retention, and renin stimulation. These effects can be abolished in the homolateral, denervated kidney, but not in the contralateral innervated kidney. Not all data, however, substantiate the ‘overflow' hypothesis. For example, patients with decompensated cirrhosis show an activation of the sympathetic nervous and renin–angiotensin–aldosterone systems and non-osmotic release of vasopressin, which should be suppressed, not stimulated, by renal sodium and water retention with volume expansion.

A new hypothesis that integrates the underfill and overfill hypotheses into one is that of peripheral arterial vasodilation. The underlying assumption of this hypothesis is that patients have peripheral vasodilation as a consequence of arteriovenous shunts in the splanchnic, dermal, and pulmonary circulations. The peripheral vasodilation that accompanies arteriovenous shunts causes renal sodium retention even if the plasma volume is increased. An expanded plasma volume is insufficient to fill the disproportionately enlarged arterial tree resulting from arterial vasodilation. Because sodium retention may not be sufficient to fill the enlarged vascular compartment, renin, angiotensin, aldosterone, and vasopressin are not suppressed.

Finally, recent data suggest that nitric oxide is an early mediator that participates in the hemodynamic changes of cirrhosis, including hyporesponsiveness to endogenous vasoconstrictors. Increased production of nitric oxide, therefore, may play an important role in the pathogenesis of ascites.

Clinical features

Within the Western world, the presence of ascites is in most instances due to underlying cirrhosis of the liver; only about 20 per cent of patients may have less common, non-hepatic causes of ascites. Initially, patients may not be aware of the presence of fluid within the abdomen. As the amount increases, however, the patient may become aware of distension, a sensation of fullness, or of tight-fitting clothes. Tense ascites may be associated with respiratory distress, anorexia, nausea, early satiety, pyrosis, ventral hernia, or abdominal pain. Body weight generally increases, but if the ascites is associated with alcoholism, poor nutrition, neoplasia, or the wasting of endstage chronic liver disease, weight may be stable or decreased. Some patients with tense ascites complain of numbness or paraesthesias with sensory loss in the distribution of the lateral cutaneous nerve of the thigh, presumably as a result of pressure where the nerve emerges behind the inguinal ligament (meralgia paraesthetica).

Physical examination of patients with massive ascites reveals a tensely distended abdomen with tightly stretched skin, bulging flanks, and an everted umbilicus. Smaller amounts of fluid (in excess of 120 ml) may be detected by eliciting shifting abdominal dullness, a fluid wave, or periumbilical dullness with the patient on hands and knees; this last maneuver is known as a positive ‘puddle sign.' When massive ascites is present, not uncommonly found are presacral, scrotal, and pedal edema, as well as pleural effusions, especially on the right side. Ascites is usually accompanied by other signs of chronic liver disease, such as jaundice, spider angiomas, hepatic or splenic enlargement, or prominent abdominal wall veins (‘caput medusa'). In some instances, ascites has been semiquantified on physical examination (1+ to 4+), but this classification does not allow for interobserver variation and is of no use in most clinical situations.

Radiographic imaging examinations are helpful for confirming or extending impressions gained on physical examination. Ultrasonography and computed tomography can detect as little as 30 ml of ascitic fluid. Even routine plain radiographs of the abdomen may be helpful, demonstrating generalized haziness and loss of the psoas shadow (ground glass appearance) as well as centralization and separation of bowel loops by ascites. In morbidly obese patients or those with a thick anterior abdominal wall, clinical detection of ascites and shifting dullness may not be possible. In these instances, ultrasound examination may be especially helpful in demonstrating the presence of ascites.

Examination of the ascitic fluid

Diagnostic paracentesis is a simple, relatively safe procedure that should be done in every patient with new-onset ascites and can help determine the cause of ascites. Points of entry into the peritoneal cavity with the patient in the supine position include the left or right lower quadrant and the avascular linea alba below the umbilicus. Previous scar sites should be avoided. Regardless of the point of entry, patients should empty their bladders before paracentesis.

After a needle insertion site is selected, the area is disinfected with iodine solution, and skin and subcutaneous tissues can be infiltrated with a local anesthetic. The operator's gloved hand retracts the skin caudad in relation to the abdominal wall while a needle is inserted (needle size is dependent on thickness of the panniculus). The direction of the needle is changed three times, in ‘Z' fashion, to minimize the likelihood of post-tap ascites leak. The character of the fluid (clear, straw-colored, thick, odorous, cloudy, milky, and bloody) should be noted, and laboratory analysis should be performed. Traditionally, the concentration of protein in the ascitic fluid and the ascites:serum lactate dehydrogenase ratio has been used to determine whether the ascites is exudative (protein more than 2.5 g/100 ml, ascites:serum lactate dehydrogenase ratio greater than 0.6) or transudative (protein less than 2.5 g/100 ml, ascites:serum lactate dehydrogenase ratio less than 0.6). Unfortunately, the distinction between exudative and transudative ascites is not absolute. Some patients with infectious and malignant ascites have a transudate, while patients with cirrhotic ascites may have an exudate. The serum:ascites albumin concentration gradient (a difference rather than a ratio) is felt by most authorities to be more accurate than the exudate/transudate distinction in classifying ascites. This method is based on the concept of oncotic-hydrostatic balance; patients with portal hypertension have a large difference between their serum and ascitic fluid albumin concentrations (greater or equal to 1.1 g/dl), while patients with ascites secondary to other processes have serum:ascites albumin gradients of less than 1.1 g/dl. For this method of classification to be accurate, however, the specimens must be obtained simultaneously (at least on the same day), the patient should be hemodynamically stable, and the specimens should be tested with the same assay system. The assay technique should also be accurate at low albumin ranges. This simple classification allows the clinician to determine if the ascites is related to portal hypertension or to an alternative cause with a reliability of greater than 90 per cent.

Ascitic fluid cell counts (white blood and differential counts) should be determined. An absolute count lower than 250 polymorphonuclear cells/mm^3 usually denotes uninfected ascites, while counts above this suggest infection. Ascitic fluid albumin, total protein, and triglyceride (if chylous ascites is suspected) concentrations should be measured. Lactic dehydrogenase, glucose, and pH of ascites are rarely helpful. Cytology and bacteriologic analysis should also be performed. Enough fluid should be obtained to perform Gram stains, routine cultures, and staining for acid-fast bacilli. Mycotic and mycobacterial cultures should be performed if indicated, and amylase levels should also be determined.

An ascitic fluid polymorphonuclear count above 250/mm^3 suggests, but a count above 500/mm^3 establishes with more than 80 per cent certainty, the diagnosis of spontaneous bacterial peritonitis. Although ascites pH and lactate levels have been proposed as helpful tests, they are not sufficiently discriminating to be of diagnostic value. Ascitic pH values below 7.32 and lactate levels above 32 mg/dl are seen not only in spontaneous bacterial peritonitis, but also in malignant ascites, pancreatic ascites, and tuberculous peritonitis.

Chylous ascites refers to a milky or creamy appearance of peritoneal fluid resulting from the presence of thoracic or intestinal lymph. Chylous ascites contains Sudan-staining globules that can be detected microscopically and an increase in triglyceride level that can be documented by chemical examination. Although generally regarded as being secondary to malignancy or lymphoproliferative disorder, chylous ascites is found in up to 20 per cent of patients with cirrhosis, resulting, perhaps, from a torn or ruptured lymphatic duct. Chylous ascites may also be found in some patients who have undergone portacaval shunt surgery.

Clinicians are often concerned about the risk of hemoperitoneum and/or infection at the time of paracentesis. These complications are rare and, in general, the risks of not performing paracentesis and missing an important diagnosis (e.g. peritonitis) far outweigh those of performing this simple procedure (Table 2).

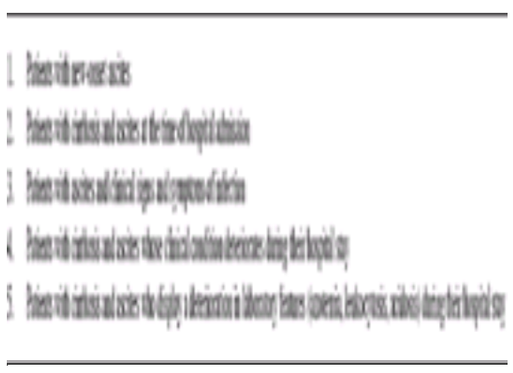


Table 2 Indications for diagnostic paracentesis

Management

Some patients may respond to simple measures, while others may require a series of therapeutic approaches that will be only palliative. Generally speaking, the more advanced the liver disease, the worse the prognosis, and, therefore, the less effective therapy will be.

Medical management

General measures

As in any disease process, attention should be directed to treating the underlying cause. Attempts to improve liver function are worthwhile; alcoholic patients should abstain from alcohol and all patients should receive optimal nutrition. Bed rest is usually recommended but is of little practical value, especially in patients whose urine sodium concentration is less than 10 mEq/1. The theoretical benefit of bed rest derives from the fact that an upright posture activates the renin–angiotensin–aldosterone and a-adrenergic systems; however, the effect of bed rest is negligible compared with that of conventional therapy. Sodium intake should be restricted to 1.5 to 2 g day; lower limits are rarely practical as patients will be non-compliant. If renal function is normal and hyponatremia is absent, fluid restriction is not necessary.

In patients with cirrhosis and ascites, prostaglandins are felt to be helpful in preserving renal function by maintaining glomerular filtration rate and free water clearance. Administration of drugs that inhibit prostaglandin synthetase, such as non-steroidal anti-inflammatory drugs, should be avoided in such patients.

Diuretics

Most patients require diuretic therapy. Dye dilution studies have demonstrated that, in the absence of edema, only 0.5 kg/day of ascitic fluid can be mobilized without compromising the intravascular compartment; however, many authorities advocate an approximate limit of 1 kg (1 litre) a day for medical diuresis. The presence of peripheral edema with ascites permits a more liberal mobilization of fluid with diuretics. Excessive diuresis may result in azotemia, hyponatremia, encephalopathy, and hepatorenal syndrome.

Aldosterone antagonists

Spironolactone is an aldosterone antagonist that promotes natriuresis and potassium retention. The initial dose ranges from 50 to 200 mg/day given in two or four doses. A useful indicator to gauge the efficacy of natriuresis is the concentration of urinary sodium relative to potassium. Ideally, urinary electrolytes should be measured in the first morning urine specimen after the patient has abstained from sodium-containing food for 8 to 12 h. If the sodium concentration of the urine remains lower than the potassium concentration, the diuresis is suboptimal and the dose of spironolactone should be increased by 50 to 100 mg every 2 to 3 days until urinary sodium exceeds urinary potassium. Theoretically, there is no absolute dose limit; however, hyperkalemia and azotemia usually prevent unlimited dose escalation. Most patients respond to a dose of less than 300 mg/day. Serum electrolytes and renal function should be monitored closely depending on the responses to and duration of diuretic therapy. Gynecomastia and galactorrhea are recognized complications of spironolactone therapy and can be dose-limiting. If spironolactone causes hyperkalemia, a potassium-wasting diuretic can be added.

Amiloride is another potassium-sparing diuretic which, although not an aldosterone antagonist, has been used successfully to treat ascites. It increases renal excretion of sodium and chloride without significantly affecting the glomerular filtration rate.

Loop diuretics

These agents act by increasing the delivery of sodium to the distal tubule where ability to reabsorb sodium is exceeded. The most common loop diuretic employed is furosemide; an initial dose of 20 mg/day may be increased sequentially, especially in patients with peripheral edema. Loop diuretics should be used as supplements to spironolactone and not as the primary diuretic. Hypokalemia may develop despite administration of potassium-sparing diuretics, and hyponatremia may also develop or worsen. Metabolic alkalosis and hepatic encephalopathy are other potential complications.

Thiazide diuretics

These are usually ineffective when given alone and they may increase renal production of ammonia. They can, however, be used in conjunction with spironolactone either to reduce potassium retention or to provide a synergistic effect.

Nitric oxide inhibitors

Data supporting a pathogenic role for nitric oxide in the formation of ascites suggest that nitric oxide inhibitors could be developed to treat cirrhotic ascites. This potential approach will be the subject of active investigation.

Large-volume paracentesis

During the 1950s, before diuretics were available, large-volume paracentesis was employed widely. This was, however, later implicated as a cause of hypotension, renal insufficiency, symptomatic hyponatremia, and encephalopathy and the procedure was virtually abandoned for the next two decades. The procedure was resurrected in the 1980s by Rodès and his colleagues in Barcelona, and by Kao *et al.* in the United States and was demonstrated to be safe and well-tolerated, without hemodynamic, renal, or hepatic consequences. Numerous groups have since performed randomized studies comparing large-volume paracentesis (with or without albumin infusion) with diuretic therapy and peritoneovenous shunting (see below). Most studies have concluded that large-volume paracentesis associated with intravenous albumin infusion is a fast, effective, and safe treatment for cirrhotic ascites. In general, patients undergoing paracentesis have a shorter hospital stay than those treated with diuretics. The Barcelona group has demonstrated the safety and clinical value of total-volume paracentesis, performed over 1 to 2 h, with no effect on renal, hepatic, or hormonal features, provided that intravenous albumin is administered. Some groups have questioned the necessity for expensive albumin infusion after paracentesis and have been treating patients quite successfully and safely without albumin infusions.

Large-volume paracentesis *per se* is quite simple. It should be performed in sterile conditions and under local anesthesia. The Barcelona group uses a needle (preferably a sharp metal needle with a 7-cm long, 17-gauge metal, blunt-edged cannula with side holes) that is inserted in the left lower abdominal quadrant. Once the needle has entered the peritoneal cavity, the inner part is removed and the fluid can be mobilized with the aid of a suction pump with large volume capacity. After the procedure, patients remain recumbent for at least 2 h on the side opposite to the paracentesis to avoid leakage of fluid. Typically, after several days, diuretic therapy is resumed. Paracentesis may be repeated when the clinical condition dictates—usually after several weeks, but not uncommonly after a few days. In Barcelona, therapeutic paracentesis is ‘first-line' therapy for patients with tense cirrhotic ascites; however, although quite widely used, large-volume paracentesis has not been embraced by all centers. The need for rapid paracentesis is often not compelling; medical therapy is quite effective in most patients, the need to return for frequent procedures is difficult for many patients, and complications do occur. In addition, levels of opsonins have been shown to be reduced after paracentesis. This could theoretically predispose these patients to spontaneous bacterial peritonitis, although this complication is rarely realized. Complications of paracentesis include infection, dissection of ascitic fluid along fascial planes to the scrotum or pleural space, hemorrhage, perforation of the bowel with generalized peritonitis or abdominal abscess, and retention of catheter fragments in the abdominal cavity. In experienced hands, however, these complications are rare.

Head out-of-water immersion

A few clinicians resort to this modality when conservative measures have failed. In the United States, this therapeutic mode is rarely used. Head out-of-water immersion decreases plasma renin activity, aldosterone, vasopressin, and norepinephrine, as well as increases the percentage of water load and urinary sodium excretion. Obviously, patients with encephalopathy cannot be subjected to this approach.

Surgical management

Peritoneovenous shunt

In 1974, LeVeen and colleagues developed a pressure-activated, one-way valve for use as a peritoneovenous shunt. One limb of the shunt lies free in the peritoneal cavity, and the venous opening of the efferent limb inserts into the superior vena cava near its entrance into the right atrium. Flow in the shunt is threaded subcutaneously and maintained if there is a 3 to 5 cmH_2O pressure gradient between the valve and its venous end. If the gradient falls below this level, the valve closes, preventing blood from flowing back into the shunt tubing. Two additional shunts are available, the Denver shunt and the Cordis–Hakim shunt, which incorporate a pumping mechanism at the abdominal end. Most patients, however, find ‘manual pumping' of these shunts difficult, and they represent little improvement over the LeVeen shunt. Even with a peritoneovenous shunt, most patients require diuretics at a reduced dose. These shunts are also associated with concomitant improvement in renal blood flow.

A Veterans Administration Cooperative Study of LeVeen shunts diminished initial enthusiasm for this procedure. Over 3000 alcoholic patients with cirrhotic ascites were enrolled in this study. Those refractory to diuretics were randomized to continued diuretic therapy plus therapeutic paracentesis or LeVeen shunt placement. Patients randomized to LeVeen shunt therapy did not show improved survival; the number of infections increased and the need for sodium restriction and diuretics continued, even in those with functioning shunts. Excessive shunt failure was also observed. A study published in 1989 compared medical treatment with peritoneovenous shunt therapy; the latter relieved the ascites promptly, reduced the length of hospital stay, and delayed recurrence of ascites. The duration of survival and number of complications in the medical and shunt groups was similar. This study suggested that peritoneovenous shunting is indicated in patients whose ascites is not treated satisfactorily by medical therapy, including those with recurrent ascites after multiple large-volume paracentesis procedures, and who have acceptable operative risk.

Early complications of peritoneovenous shunts are valve malfunction, intravascular volume overload, pulmonary edema, disseminated intravascular coagulation, hypokalemia, air embolism, peritonitis, sepsis, and ascites leakage. Late complications are shunt occlusion (in about 30 per cent of patients), hepatic vein thrombosis, variceal hemorrhage, and peritoneal fibrosis with bowel obstruction. At least half of the patients will experience one or more complications. Mortality rates for these

patients with endstage cirrhosis and refractory ascites are quite high; about half succumb within the first postoperative year. This procedure is contraindicated in patients with sepsis, heart failure, and active peritonitis. The best results are achieved in a limited subset of patients with diuretic-resistant ascites but relatively well-preserved hepatic function.

Between the anticipated complications and revision requirements of peritoneovenous shunts on the one hand and the availability of transjugular intrahepatic portasystemic shunts and liver transplantation (see below) on the other, LeVeen shunts have been all but abandoned.

Portasystemic shunts

Use of portasystemic shunts for treatment of cirrhotic ascites is warranted only in those patients in whom all other forms of therapy have failed. Many randomized trials were done several decades ago comparing medical therapy with different types of portasystemic shunts, but most such studies were designed to evaluate shunts as therapy for bleeding esophageal varices. The shunt operation that is most effective at relieving ascites has not been established; in fact, ascites is reduced after any type of portasystemic shunt as a consequence of decreased portal flow and decreased intrahepatic congestion. Among the most commonly performed shunts, splenorenal and portacaval shunts and their variants have been proven effective in relieving ascites. Ascites may actually increase transiently after distal splenorenal (Warren) shunt surgery, but, ultimately, both distal and proximal (Linton) shunts lower the tendency to accumulate ascites. Between the two types of portacaval anastomosis, the end-to-side portacaval shunt decompresses the entire splanchnic bed but not hepatic sinusoids; after this shunt, ascites may persist. The side-to-side portacaval shunt, which decompresses both the splanchnic and hepatic sinusoidal beds, is very effective in relieving ascites; it is the decompressive shunt of choice for the treatment of ascites in the Budd–Chiari syndrome (when hepatic function is well preserved). The choice of portasystemic shunt should be individualized, taking into account factors such as the experience and preference of local surgeons. In contemporary practice, surgical shunts are rarely, if ever, performed for the management of ascites.

Transjugular intrahepatic portasystemic shunt (TIPS)

Angiographically guided insertion of a stent between an intrahepatic vein and the portal vein has been used to reduce portal hypertension. The use of TIPS became very popular in the treatment of bleeding from esophageal varices. In 1993, an uncontrolled, small but encouraging study was published in which 14 patients with intractable ascites were treated with this technique. TIPS procedures were adopted enthusiastically for the treatment of variceal bleeding and ascites in most university and many private hospitals by the early 1990s. The proceedings of a National Institutes of Health Consensus Development Conference published in 1995 designated TIPS as accepted treatment for uncontrollable acute variceal bleeding as well as variceal bleeding refractory to medical therapy. At the time, data on TIPS as therapy for refractory ascites were limited, but TIPS was described at the consensus conference as promising for use in ascites. Since then, several prospective uncontrolled studies have been published demonstrating the value of TIPS for the treatment of ascites.

In approximately 5 per cent of patients with ascites, a pleural effusion may develop and is usually right sided. In the absence of cardiac or pulmonary disease, the presence of a pleural effusion in a patient with cirrhosis is known as hepatic hydrothorax. As opposed to ascitic fluid in the abdominal cavity, even small volumes of fluid within the pleura may be associated with respiratory distress, which may require thoracentesis for removal of the fluid. This procedure may be associated with a high frequency of complications in patients with cirrhosis. Once the diagnosis of hepatic hydrothorax is established, medical therapy consisting of salt restriction and diuretics is initiated. When these conservative measures are exhausted and the patient remains symptomatic, TIPS appears to be the most effective form of treatment, as supported by the publication of several study reports in recent years.

Liver transplantation

Although beyond the scope of this chapter, liver transplantation is the treatment of choice for patients with liver failure and otherwise intractable ascites, unless contraindications exist.

Spontaneous bacterial peritonitis

An important complication of ascites is spontaneous bacterial peritonitis, defined as an infection of ascitic fluid in the absence of any obvious intra-abdominal source.

Pathogenesis

Spontaneous bacterial peritonitis can occur in patients with ascites of any cause; the mechanisms underlying it remain unknown. Plausible contributing factors include transmural translocation of bacteria secondary to increased gut permeability to enteric organisms, derangement in abdominal lymphatic circulation, hematogenous seeding from a distant source, decreased function of the reticuloendothelial system in patients with chronic liver disease, qualitative neutrophil dysfunction, impaired leukocyte chemotaxis, and decreased antibacterial capacity of ascitic fluid (deficiency of opsonic activity).

Predisposing factors

Patients with cirrhosis are at high risk of developing spontaneous bacterial peritonitis, which usually occurs in patients with advanced liver disease. Patients with a low serum ascitic fluid protein level (less than or equal to 1 g/dl) have a 10-fold-higher frequency incidence of spontaneous bacterial peritonitis than those with a high ascitic fluid protein content. Also, patients with cirrhosis and gastrointestinal bleeding are at high risk of spontaneous bacterial peritonitis. Twenty per cent of patients with cirrhosis and gastrointestinal hemorrhage are infected at the time of hospital admission. Invasive procedures such as intravascular catheters and bladder catheters have been proposed as favoring bacteremia and spontaneous bacterial peritonitis in patients with cirrhosis, but data do not support this association. Nevertheless the use of these catheters should minimized.

Bacteriology

A single organism is the cause of over 90 per cent of cases. The most common organisms are *Escherichia coli* and *Streptococcus* spp.; other organisms may also be implicated (Table 3).

Micro-organism	*Percentage of cases*
Gram-negative bacilli	70
Escherichia coli	
Klebsiella spp.	
Others*	
Gram-positive cocci	10–30
Streptococcus pneumoniae	
Other Streptococcus	
Enterococci	
Staphylococcus spp.	
Anaerobes and microaerophilus	6–14
Unusual cases◇	

*Includes *Acinetobacter* spp., *Serratia* spp., *Proteus* spp., *Pseudomonas* spp., *Providencia* spp., *Salmonella* spp., *Citrobacter* spp., *Enterobacter* spp.
†Includes *Bacteroides* spp., *Clostridium* spp.
◇Includes diphtheroids, *Neisseria* spp., *Candida* spp., *Listeria monocytogenes*.

Table 3 Organisms responsible for spontaneous bacterial peritonitis

Diagnosis

Prompt recognition of spontaneous bacterial peritonitis requires not only an awareness of the presenting symptoms and signs, but also a high index of suspicion in any patient with ascites and clinical deterioration. Spontaneous bacterial peritonitis can vary in its presentation from being clinically dramatic to totally asymptomatic. Fever of about 38°C (100°F) is the most common presenting feature and occurs in 50 to 80 per cent of patients. Abdominal pain, usually diffuse, occurs in 25 to 72 per cent of patients; rebound tenderness is elicited in over 50 per cent of patients. None of these findings is specific for spontaneous bacterial peritonitis, however; abdominal complaints are common in patients with cirrhosis and ascites in the absence of this complication, and spontaneous bacterial peritonitis can occur in the absence of abdominal pain or fever.

Other subtle indicators suggestive of spontaneous bacterial peritonitis include diarrhea, worsening renal insufficiency, and refractoriness to diuretics, hypothermia, and unexplained encephalopathy. Peripheral blood leukocytosis is common but may be masked by hypersplenism. Despite negative ascitic fluid cultures (which may occur in up to 50 per cent of patients), the presence of more than 250 polymorphonuclear cells/mm^3 in ascites suffices to support the diagnosis and to mandate therapy.

Once the diagnosis is entertained, paracentesis is indicated. Ascitic fluid should be submitted promptly to the laboratory for aerobic and anaerobic cultures, cell count with differential count, and Gram stain of centrifuged fluid, as well as other routine laboratory assessments. To increase the likelihood of a positive ascites fluid culture, the operator should inoculate 10 ml of ascitic fluid directly into blood culture bottles at the bedside. ‘Bedside' inoculation increases the sensitivity of ascites culture and decreases the time between inoculation of the culture and appearance of bacterial growth.

The diagnosis of spontaneous bacterial peritonitis is suggested by a polymorphonuclear cell concentration of 250 cell/mm^3 or higher but felt to be established more conclusively by a concentration of 500 cells or more/mm^3. Ascitic pH below 7.3, a lactate level above 32 mg/100 ml, glucose, specific gravity, and protein concentrations are less helpful. Most cases of peritonitis in patients with cirrhosis result from spontaneous bacterial peritonitis; therefore, the threshold for surgery, which can be fatal, should be very high. Conversely, peritonitis secondary to an intra-abdominal process such as a perforated ulcer or diverticulitis can occur in patients with cirrhosis and ascites; failure to recognize such a source of infection and to intervene surgically and promptly is almost universally fatal. Hallmarks of ‘surgical' peritonitis include polymorphonuclear cell counts above 10 000/mm^3, polymicrobial peritonitis, low ascites glucose, and high levels of lactate dehydrogenase and protein.

Ascitic fluid from patients with HIV infection as well as from other immunosuppressed patients should be sent routinely for mycobacterial and mycological stains and cultures in addition to standard tests.

Treatment

Once a presumptive diagnosis of spontaneous bacterial peritonitis has been made, prompt and appropriate intravenous antibiotic therapy should be instituted long before culture results are known and even if cultures are negative. Empirically, a third-generation cephalosporin should be used, unless a Gram stain suggests that a narrower-spectrum antibiotic will suffice. Aminoglycosides should be avoided; they are associated with a high risk of nephrotoxicity and have unpredictable distribution in patients with cirrhosis. When antibiotic sensitivities of the organism are known, antibiotic coverage can be narrowed. Treatment should continue for 10 to 14 days, although shorter-duration therapy with third-generation cephalosporins is effective. Cefotaxime has become a commonly used antibiotic in this setting. Response to antibiotics can be judged clinically; repeat paracentesis in 48 h may be helpful but is not usually necessary. Once an episode is treated the frequency of recurrence is high, and the mortality is also high, a reflection of the underlying severity of the liver disease in these patients. Between 40 and 70 per cent of survivors of an episode of spontaneous bacterial peritonitis will experience another episode during the first year of follow-up. One-year probability of survival after the first episode of spontaneous bacterial peritonitis is approximately 30 to 40 per cent; therefore, spontaneous bacterial peritonitis seems to be a landmark in the natural history of cirrhosis and survivors should be considered for liver transplantation.

Prophylaxis with daily quinolone therapy (e.g. ciprofloxacin or norfloxacin) has been shown to reduce the frequency of not only recurrent but also spontaneous bacterial peritonitis. Patients admitted to hospital with cirrhosis and gastrointestinal bleeding or patients whose ascitic fluid has low protein constitute two groups at high risk of spontaneous bacterial peritonitis. Long-term primary or secondary antibiotic therapy reduces the incidence of spontaneous bacterial peritonitis, but patients should be monitored carefully for the development of infections caused by quinolone-resistant organisms. For patients taking long-term quinolone antibiotic coverage, gut antifungal prophylaxis (e.g. with mycostatin or chlotrimazole) is indicated.

Hepatorenal syndrome

Progressive ‘functional' renal failure occurs in patients with advanced liver disease, most of whom have tense ascites. This specific form of renal failure develops in the absence of clinical or laboratory evidence of common causes of renal failure that may occur in patients without liver disease. The kidneys are anatomically and histologically normal; if hepatic deterioration is reversed by transplantation, the renal failure is also reversible.

The prognosis for this syndrome is extremely poor; the majority of patients succumb within the first month of a diagnosis of hepatorenal syndrome.

Hepatorenal syndrome is characterized by impaired renal function, with marked abnormalities in renal arterial circulation and excess endogenous activity of vasoactive substances. In the kidneys there is marked arterial shunting from the outer renal cortex to the medulla, which results in a decreased glomerular filtration rate. In the extrarenal circulation, there is a predominance of arteriolar vasodilatation, which results in a reduction in total systemic vascular resistance and arterial hypotension as well as a high cardiac output.

The renin–angiotensin system participates in the renal vasoconstriction of the hepatorenal syndrome. Arginine vasopressin (anti-diuretic hormone) has vasoconstrictor activity as well, and plasma levels of this hormone are increased in patients cirrhosis and ascites. Endothelin may be another vasoconstrictive substance whose concentration in plasma is increased in cirrhosis. Whether endothelin causes renal vasoconstriction however, remains controversial.

Adenosine, some eicosanoids with vasoconstrictor effects (leukotriene E4), and endotoxin have been the focus of investigative attention and are being studied as potential factors leading to renal dysfunction in hepatorenal syndrome.

Vasodilators, such as prostaglandins, play little or no role in regulating renal perfusion under physiologic conditions in healthy patients, but the situation differs in patients with cirrhosis. Inhibition of prostaglandins by non-steroidal anti-inflammatory agents may impair renal perfusion in these patients by potentiating the effects of vasoconstrictor substances.

Nitric oxide may also play a role in patients with cirrhosis, especially by interacting with prostaglandins to maintain renal hemodynamics. When nitric oxide and prostaglandins are inhibited simultaneously in rats with cirrhosis and ascites, marked renal vaso-constriction develops. Members of the natriuretic peptide family may also be involved in the regulation of renal perfusion in patients with cirrhosis.

Diagnosis

Progressive azotemia, oliguria (less than 500 ml/day), and a concentrated urine with a ratio of urine:plasma osmolarity above 1.0, and a urinary sodium concentration below 10 mEq/1 in the presence of a normal urinalysis are the typical features (Table 4). The urine may contain small amounts of protein, hyaline, and a few granular casts. Oliguria may occur spontaneously but usually follows diuretic therapy, diarrhea, large-volume paracentesis (especially when performed without administration of plasma volume expanders, such as albumin), gastrointestinal hemorrhage, or sepsis. Before this diagnosis can be established, decreased intravascular volume must be excluded by expansion of the intravascular compartment with colloid and fluid as well as discontinuing diuretic therapy. In some circumstances, measurement of central venous pressure may be needed but is usually not necessary. If, after expansion of the intravascular compartment with colloid and fluid, sustained improvement in renal function (decrease in serum creatinine below 1.5 mg/dl or increase in 24-h creatinine clearance above 40 ml/min) does not occur, the diagnosis of hepatorenal syndrome can be made.

Biochemical characteristic	Prerenal azotemia	Hepatorenal syndrome	Acute tubular necrosis
Urine sodium concentration (mEq/l)	< 10	< 10	> 30
Urine to plasma creatinine ratio	< 30:1	> 30:1	< 20:1
Urine osmolality	At least 100 mosmol > plasma osmolality	At least 100 mosmol > plasma osmolality	Equal to plasma osmolality
Urine sediment	Normal	Unremarkable	Casts, cellular debris

Table 4 Differential diagnosis of acute azotemia in patients with advanced liver disease

Although serum creatinine is an accurate reflection of glomerular filtration rate in general, in patients with cirrhosis, creatinine measurements are insensitive indicators of this rate. In patients with cirrhosis, poor nutritional status and reduced muscle mass are associated with low endogenous creatinine production; therefore, a decreased glomerular filtration rate may be associated with only a slight increase in serum creatinine. Similarly, because other factors, such as gastrointestinal bleeding, may contribute to elevations of blood urea nitrogen in patients with cirrhosis, measuring these levels is also inaccurate in assessing renal function in these patients.

A diagnosis of hepatorenal syndrome in patients with cirrhosis is precluded when shock predates the development of renal impairment and rendered very difficult when patients have a history of recent use of nephrotoxic drugs, such as non-steroidal anti-inflammatory agents and certain antibiotics.

Treatment

Despite the identification of several predictive factors for the development of hepatorenal syndrome in patients with cirrhosis and ascites, little can be done to prevent the development of renal failure in those patients. Although no effective therapy for hepatorenal syndrome exists, precipitating factors should be avoided. Administration of plasma volume expanders, especially albumin, has been shown to reduce the occurrence of hepatorenal syndrome following large-volume paracentesis from nearly 10 per cent without volume expansion to 2 per cent in those who received treatment. Administration of antibiotics to prevent spontaneous bacterial peritonitis may also be helpful.

Numerous vasodilating agents such as phentolamine and papaverine have been administered, but have not been effective. Recently, several investigators have tried misoprostol with no beneficial effects. Vasoconstriction agents such as norepinephrine, metaraminol, angiotensin II, and ornipressin have also been used in an attempt to improve renal perfusion by increasing systemic vascular resistance and suppressing the activity of the systemic vasoconstrictors.

Despite anecdotal reports of hepatorenal syndrome reversal in patients treated with peritoneovenous shunting, this procedure is not recommended. Although peritoneovenous shunting may limit progression of renal failure, the procedure has no impact on overall survival and is accompanied by an unacceptably high frequency of mechanical, infectious, and hematologic complications. Side-to-side or end-to-side portasystemic shunts have been reported in rare instances to reverse hepatorenal syndrome, probably by reducing sinusoidal portal pressure and limiting the impact of endogenous vasoconstrictors. Practically, however, hepatic decompensation is so profound in patients with hepatorenal syndrome that surgical shunts carry too high a morbidity and mortality to be embraced. In contrast, TIPS yields the same level of portal decompression without the operative morbidity and mortality of surgical shunts. Although the application of TIPS to treatment of hepatorenal syndrome remains limited, preliminary reports of improvement in renal function and survival after TIPS have appeared. Therefore, before TIPS can be recommended for hepatorenal syndrome, further study is warranted. Hemodialysis has not been effective in reversing hepatorenal syndrome, but, occasionally, patients with epatorenal syndrome have been maintained briefly on hemodialysis while awaiting liver transplantation.

Liver transplantation remains the best approach for patients with endstage liver disease and hepatorenal syndrome. Unfortunately, patients with hepatorenal syndrome fall into the most decompensated category of transplantation candidates, and the outcome of transplantation is correspondingly disappointing. Patients with hepatorenal syndrome prior to transplantation experience more complications, spend more days in intensive care, and have higher in-hospital mortality. Still, despite this increased morbidity, long-term survival of patients with hepatorenal syndrome undergoing liver transplantation is excellent; their 3-year probability of survival is 60 per cent. Ideally, liver transplantation should be performed before the onset of hepatorenal syndrome.

Further reading

Arroyo V, Ginés P, Rimola A, Gaya J. Renal function abnormalities, prostaglandins and effects of nonsteroidal anti-inflammatory drugs in cirrhosis with ascites. An overview with emphasis on pathogenesis. *American Journal of Medicine* 1986; **81**: 104–22.[Very comprehensive review; physiology and pathophysiology.]

Bataller R, Ginés P, Guevarra M, Arroyo V. Hepatorenal syndrome. *Seminars in Liver Disease* 1997; **17**: 233–47. [Comprehensive review and key physiologic concepts regarding hepatorenal syndrome.]

Brensing KA, Textor J, Strunk H, Klehr HU, Schild H, Sauerbruch T. Transjugular intrahepatic portosystemic stent-shunt for hepatorenal syndrome. *Lancet* 1997; **349**: 697– 8. [Comprehensive review of TPS.]

Conn HO. Transjugular intrahepatic-portal systemic shunts: the state of the art. *Hepatology* 1993; **17**: 143–51. [The first comprehensive review of experience with TPS.]

Epstein M. Deranged sodium homeostasis in cirrhosis. *Gastroenterology* 1979; **76**: 622–35. [Important article describing the ‘earliest' derangement that occurs in cirrhosis.]

Ginés P *et al.* Comparison of paracentesis and diuretics in the treatment of cirrhotics with tense ascites. Results of a randomized study. *Gastroenterology* 1987; **93**: 234–41. [Important for the reader to compare one mode of therapy with the other.]

Ginés P, Fernández-Esparracha G, Arroyo V, Rodes J. Pathogenesis of ascites and cirrhosis. *Seminars in Liver Disease* 1997; **17**: 175–89. [Excellent review.]

Ginés P *et al.* Randomized comparative study of therapeutic paracentesis with and without intravenous albumin in cirrhosis. *Gastroenterology* 1998; **94**: 1493–502. [Insights into a frequently debated topic.]

Guarner C, Soriano G. Spontaneous bacterial peritonitis. *Seminars in Liver Disease* 1997; **17**: 203–17. [Excellent review.]

Lebrec D *et al.* Transjugular intrahepatic portosystemic shunt: comparison with paracentesis in patients with cirrhosis and refractory ascites: a randomized trial. *Journal of Hepatology* 1996; **25**: 135–44. [Well designed study comparing two treatment modalities.]

Martin P-Y, Gines P, Schrier RW. Nitric oxide as a mediator of hemodynamic abnormalities and sodium water retention in cirrhosis. *New England Journal of Medicine* 1998; **339**: 533–41. [A good explanation of the role of nitric oxide.]

Moskovitz M. The peritoneovenous shunt: expectations and reality. *American Journal of Gastroenterology* 1990; **85**: 917–29. [Good review of more established therapy for ascites.]

Ochs A *et al.* The transjugular intrahepatic portosystemic shunt-stent procedure for refractory ascites. *New England Journal of Medicine* 1995; **332**: 1192–7. [Excellent review.]

Runyon BA. Care of patients with ascites. *New England Journal of Medicine* 1994; **330**: 337–42. [Good review.]

Runyon BA. Treatment of patients with cirrhosis and ascites. *Seminars in Liver Disease* 1997; **17**: 249–60. [Another good review.]

Runyon BA *et al.* Bedside inoculation of blood culture bottles with ascites fluid is superior to delayed inoculation in the detection of spontaneous bacterial peritonitis. *Journal of Clinical Microbiology* 1990; **28**: 2811–12.

Runyon BA *et al.* The serum–ascites albumin gradient is superior to the exudate–transudate concept in the differential diagnosis of ascites. *Annals of Internal Medicine* 1992; **117**: 215–20. [This article changed the way we classify different types of ascitic fluid.]

Stanley MM *et al.* Peritoneovenous shunting as compared with medical treatment in patients with alcoholic cirrhosis and massive ascites. *New England Journal of Medicine* 1989; **321**: 1632–8. [Important article which sheds light on physiology and therapeutic aspects.]

Starzl TE, Demetris AJ, Van Thiel D. Liver transplantation. Parts 1, 2. *New England Journal of Medicine* 1989; **321**: 1014–22, 1092–9. [Excellent review.]

Strauss RM, Boyer TD. Diagnosis and management of cirrhotic ascites. In: Zakim D, Boyer TD, eds, *Hepatology, a textbook of liver disease*,. 3rd edn, pp. 764–88. WB Saunders Co., Philadelphia, 1996. [Very comprehensive review.]

Strauss RM, Martin LG, Kaufman SL, Boyer TD. Transjugular intrahepatic portal systemic shunt for the management of symptomatic cirrhotic hydrothorax. *American Journal of Gastroenterology* 1994; **89**: 1520. [Describes patients with intractable hepatic hydrothorax undergoing TPS with adequate resolution of their problem. Although a small study, it is the first publication on this clinical problem and the use of this therapeutic modality.]

33.1 Acute pancreatitis

Oscar Joe Hines and Howard A. Reber

Introduction
Pathogenesis of acute pancreatitis
Obstruction–secretion
Common-channel theory
Duodenal reflux
Increased pancreatic-duct permeability
Enzyme autoactivation
Etiology
Pathology
Diagnosis of acute pancreatitis
Diagnostic studies
Predicting severity
Prognostic signs
Pathophysiology of systemic manifestations
Cytokines
Treatment of acute pancreatitis
Endoscopic intervention
Surgical treatment
Surgery in severe acute pancreatitis
Risk factors for pancreatic infections
Diagnosis and bacteriologic findings
Timing of surgery
Infected pancreatic necrosis
Sterile pancreatic necrosis
Surgical techniques
Postoperative course
Prognosis
Long-term complications
Pancreatic abscess
Pancreatic pseudocyst
Pancreatic fistula
Further reading

Introduction

Acute pancreatitis is a nonbacterial inflammatory disease that results from intrapancreatic activation, release, and digestion of the organ by its own enzymes. Although the majority of attacks are mild and patients recover within a week, about 10 per cent of cases are severe, associated with a protracted course, and may be complicated by infection and/or multisystem organ failure and death. Acute pancreatitis is characterized clinically by acute abdominal pain, elevated concentrations of pancreatic enzymes in blood, and an increase in the amount of those enzymes excreted in urine. There may be just a single episode of pancreatitis or multiple episodes may recur. In the United States of America, the most common cause of acute pancreatitis is cholelithiasis.

Acute pancreatitis should be distinguished from chronic pancreatitis, which causes fibrosis and scarring of the pancreatic parenchyma that can produce pancreatic exocrine insufficiency and diabetes. In acute pancreatitis, the pancreas usually returns to normal after the acute inflammation subsides and pancreatic insufficiency does not result. Unlike patients with acute pancreatitis, those with chronic pancreatitis usually experience chronic abdominal pain for months or years that is not necessarily associated with acute inflammation. But they also may experience acute episodes of pancreatic inflammation whose symptoms and treatment are the same as in a patient with acute pancreatitis. The distinction between the two diseases is important because their causes and treatment are often different.

Pathogenesis of acute pancreatitis

Acute pancreatitis is thought to result from enzymatic digestion of the gland by its own enzymes, but the mechanism(s) by which enzyme activation occurs in humans remains obscure. The pathogenesis of pancreatitis will be discussed in the context of various theories that attempt to explain its clinical characteristics.

Obstruction–secretion

Complete obstruction of the pancreatic duct in animals produces pancreatic edema that may resemble a mild acute edematous pancreatitis in humans. However, the process generally resolves within a week and over several months the acinar tissue is replaced by fibrous scar. The pancreas takes on an appearance which resembles that of chronic pancreatitis. Partial duct obstruction, when combined with stimulation of pancreatic secretion, produces a more severe inflammation. This experimental condition may mimic the events in biliary pancreatitis, when a gallstone obstructs the pancreatic duct during pancreatic secretion. Acute alcoholic pancreatitis may also have a similar mechanism. Ethanol ingestion stimulates the secretion of gastric acid, which releases duodenal secretin, promoting pancreatic secretion. Alcohol also causes spasm of the sphincter of Oddi, which obstructs the flow of pancreatic secretions. Ethanol also may acutely decrease pancreatic blood flow, which could injure the pancreatic acini. This theory does not explain how the pancreatic enzymes are activated to begin the autodigestive process, however.

Common-channel theory

In 1902 Eugene Opie, a pathologist at Johns Hopkins Hospital, performed an autopsy on two patients who had died from acute pancreatitis. He found that a small gallstone that had migrated from the gallbladder into the distal common bile duct had lodged near the ampulla of Vater. He reasoned that the obstructing stone caused bile to reflux into the pancreatic duct, which entered the commonbile duct just proximal to the stone, to create a short common channel. Pancreatitis followed. Although it is unknown whether bile reflux is actually responsible for pancreatitis, it is well established that gallstones in the distal common bile duct may incite pancreatitis. The critical event is probably obstruction of the pancreatic duct. In animals, infected bile, which contains deconjugated bile salts, and bile incubated with pancreatic juice, which contains lysolecithin, can injure the lining of the pancreatic ducts. Their permeability is increased so that pancreatic enzymes can leak from them into the surrounding pancreatic parenchyma, which could initiate pancreatitis. On the other hand, fresh bile is apparently innocuous to the ducts.

Duodenal reflux

Reflux of duodenal contents into the pancreatic duct causes pancreatitis in experimental animals. The rare case of pancreatitis that occurs after a Billroth II gastrectomy with obstruction of the afferent loop probably has a similar cause. This theory explains pancreatic enzyme activation, since duodenal enterokinase would gain access to the pancreatic duct. Nevertheless, normal mechanisms in the duodenal wall and sphincter of Oddi efficiently prevent reflux and there is no direct evidence that duodenal reflux into the pancreatic ducts occurs in humans.

Increased pancreatic-duct permeability

In animal experiments, a variety of conditions increase the permeability of the pancreatic ducts so that pancreatic enzymes, normally restricted to the duct lumen, can leak into the surrounding tissue. These conditions include the acute ingestion of ethanol, direct exposure of the duct to deconjugated bile salts, pancreatic secretion against an obstruction, and acute hypercalcemia. If those enzymes have been activated, they cause acute pancreatitis. Nevertheless, there is no direct evidence that this occurs in humans.

Enzyme autoactivation

Before they are secreted, the pancreatic proteolytic enzymes are stored in an inactive state as zymogen granules in the acinar cells. Activation, which requires the conversion of the proenzyme trypsinogen to active trypsin, normally does not occur until the enzymes reach the lumen of the duodenum, where trypsinogen is exposed to duodenal enterokinase. Active trypsin then catalyzes the conversion of the remaining proenzymes into their active form and digestion of the chyme proceeds. Under certain conditions in experimental animals, acute pancreatitis occurs when proteolytic enzymes are activated inside the acinar cells themselves before secretion takes place. The zymogen granules appear to fuse with intracellular lysosomes that contain the enzyme, cathepsin B, that converts trypsinogen to trypsin. Presumably, the

presence of active trypsin inside the acinar cells injures them and initiates a cascade of events that lead to pancreatitis. These observations in animals may be important because they offer a novel explanation for enzyme activation that does not require the fact of exposure to duodenal enterokinase. However, there is no evidence that enzyme autoactivation occurs in human pancreatitis.

It should be evident that we have only an incomplete understanding of the pathogenesis of pancreatitis. The diverse causes of the disease suggest that any of a number of inciting factors may result in injury to the gland and that the pancreas responds in a limited fashion. Biliary pancreatitis is associated with acute inflammation and almost certainly follows transient obstruction of the pancreatic duct by a gallstone. It is unknown whether bile reflux, elevated ductal pressures, or other factors are important. Alcoholic pancreatitis is associated with chronic inflammation, although acute inflammatory episodes do occur. Whether acute alcohol ingestion produces acute pancreatitis by elevating ductal pressures, producing transient pancreatic ischemia, or by increasing the permeability of the pancreatic duct, is also unresolved.

Etiology

Gallstone pancreatitis accounts for about 90 per cent of cases of acute pancreatitis in the United States. Eradication of the gallstone disease prevents further attacks of pancreatitis. The etiologic mechanism probably involves the transient obstruction of the pancreatic duct by a gallstone in the common bile duct at the ampulla of Vater (see above). Gallstones are recoverable in the feces of over 90 per cent of patients within 10 days of an attack, suggesting that the obstruction is brief and that most such stones pass into the duodenum. The usual age at onset is in the mid- to late 40s, with women more commonly affected than men.

Alcoholic pancreatitis accounts for about 75 per cent of cases of chronic pancreatitis in the United States and many of these patients suffer acute episodes of pancreatic inflammation that are indistinguishable from acute pancreatitis of other causes. Sometimes, an episode of ‘binge drinking' precipitates an attack typical of acute pancreatitis: this usually occurs in a patient who is a chronic alcoholic in whom the pancreas has probably already been permanently damaged by alcohol even though other evidence of pancreatic disease is absent (e.g. pancreatic exocrine insufficiency, diabetes). However, occasionally there is no prior history of ethanol intake and the pancreas may be normal before the attack. The mechanism(s) of injury are unclear. Ethanol increases the permeability of the pancreatic ducts (see above), which might allow pancreatic enzymes normally contained within the ducts to leak into the surrounding tissue. If the enzymes are active, the pancreas is damaged. There is recent evidence that acute ethanol ingestion markedly but transiently depresses pancreatic blood flow. Thus, cellular damage also could be provoked by ischemia.

Hypercalcemic states, most commonly hyperparathyroidism, may cause acute and chronic pancreatitis. The mechanism is unclear, but hypercalcemia may favor the intraductal precipitation of calcium stones, which are seen in up to half of these patients. Calcium also influences the activation of certain pancreatic enzymes and a variety of synthetic and secretory events in the acinar cells. Acute hypercalcemia increases the permeability of pancreatic ducts, perhaps allowing enzymes to leak from them and damage the tissue.

Hyperlipidemia, especially that associated with elevated chylomicrons and very low-density lipoproteins, is a cause of acute pancreatitis. The mechanism may involve the activation of pancreatic lipase and the subsequent release of large amounts of toxic fatty acids into the pancreatic capillary circulation. Vascular endothelial damage might cause sludging of red blood cells, stasis, and pancreatic ischemic injury and inflammation. Hyperlipidemia can also occur transiently during an attack of pancreatitis, usually in alcoholics. In such patients, it is not thought to have a significant role.

Hyperlipidemia as a cause of pancreatitis should be recognized for two reasons. First, because the elevated serum lipids interfere with the chemical determination of amylase; serum amylase in these patients with acute pancreatitis is often reported as ‘normal'. The appearance of lactescent serum in a patient with acute abdominal pain, however, should alert the clinician to the correct diagnosis. Because the lipid is not excreted in the urine, the finding of a high urinary amylase excretion may still be diagnostic in this setting. Second, dietary or pharmacologic control of the hyperlipidemia minimizes the chances of recurrent episodes of pancreatitis.

Hereditary pancreatitis is a genetic condition transmitted as a Mendelian dominant trait. Genetic analysis has localized the abnormalities to chromosome 7q35 in several of the families studied; this results in an abnormal form of cationic trypsin that may be more resistant to enzymatic degradation. If this abnormal trypsin were released into the pancreatic parenchyma, and the protective mechanisms that would normally destroy it were ineffective, this could theoretically cause pancreatitis. The condition is rare, though over 250 patients have been reported. In most of the patients, symptoms typical of acute pancreatitis appear between the ages of 12 and 14 years. Relentless progression of the process is usual, with recurrent attacks of acute inflammation. Eventually many patients develop the typi-cal manifestations of chronic pancreatitis, including calcifications (32 per cent), diabetes (19 per cent), and steatorrhea from exocrine insufficiency (15 per cent). An increased incidence of pancreatic carcinoma has been noted in these patients. In most of the reported cases where endoscopic retrograde cholangiopancreatography has been performed, the pancreatic ducts have been dilated.

Postoperative (iatrogenic) pancreatitis occurs after a variety of procedures in proximity to the pancreas and the causes are usually obvious. They include direct injury to the gland (e.g. pancreatic biopsy, pancreatic resection) or obstruction of the pancreatic duct (e.g. the placement of a long-arm T-tube through the sphincter of Oddi into the duodenum; forceful dilation of the sphincter of Oddi during exploration of the common duct). For this reason, these latter two procedures are no longer performed. Pancreatitis also occasionally follows endoscopic sphincterotomy or surgical sphincteroplasty.

Sometimes the operation is distant from the pancreas and there is no obvious explanation. For example, pancreatitis complicates up to 5 per cent of cases of heart surgery that require cardiopulmonary bypass. It has been suggested that impaired pancreatic perfusion may damage the pancreas in this setting.

Pancreatitis can follow Billroth II gastrectomy if the afferent limb of the jejunum becomes obstructed. In this case, pancreatic juice and bile accumulate in the limb and may reflux into the pancreatic duct under pressure.

Drugs that cause acute pancreatitis are an uncommon but important cause of the disease. Those most commonly incriminated include steroids, azathioprine, 6-mercaptopurine, thiazide diuretics, furosemide (frusemide), sulfonamides, tetracycline, and estrogens (which produce pancreatitis by inducing hypertriglyceridemia).

Brief ductal obstruction may cause acute pancreatitis, but when it is persistent over months or years it causes chronic pancreatitis instead. Obstruction occurs in a variety of clinical settings. Chronic pancreatitis has been reported in patients with severe strictures at the point where fusion of the ventral and dorsal pancreatic ducts occurs during fetal life. Strictures occur also in other areas of the main duct and are probably due to scarring after trauma or inflammation from any cause. Pancreas divisum, a variant of normal pancreatic ductal anatomy, may cause pancreatitis because of duct obstruction. During the seventh week of fetal development, the ventral and dorsal pancreatic buds fuse. In 90 per cent of people, the ducts of the ventral and dorsal buds join to form the main pancreatic duct (Wirsung). This empties into the duodenum at the papilla of Vater with the common bile duct. The proximal portion of the dorsal bud duct becomes the minor pancreatic duct (Santorini), which empties into the duodenum through the minor papilla about 2 cm proximal to the papilla of Vater. In the remaining 10 per cent the ventral and dorsal ducts do not fuse, and the only the ventral duct that drains the small uncinate process of the pancreas empties through the major papilla. The dorsal pancreas (the rest of the head, the body, and tail) empties instead through the minor papilla. This anatomic arrangement is not usually associated with any disease. However, in a small percentage of people, repeated episodes of acute pancreatitis do occur. The cause is thought to be related to increased pressure in the pancreatic duct, which occurs when the relatively large volume of juice from most of the pancreas must drain through the small opening in the minor papilla. In those patients who are symptomatic, it has been suggested that inflammation and scarring of the minor papilla (e.g. from duo-denitis) may have further narrowed the already small opening. Opening the minor papilla either by endoscopic or surgical means may eliminate the pancreatitis if there have not been too many prior attacks. Otherwise, chronic obstructive pancreatitis may have developed, which is not improved by opening the sphincter. The condition affects mainly women and is often first diagnosed in the second and third decades of life, an earlier age than pancreatitis caused by ethanol, which affects people in their 40s, or gallstones, which affect people in their 50s and 60s.

Miscellaneous additional causes include the acute pancreatitis that follows scorpion venom poisoning (*Tityus trinitatis*), various infectious agents (mumps, group B coxsackie viruses, herpes simplex, mononucleosis), and exposure to anticholinesterase insecticides.

Idiopathic pancreatitis still accounts for 10 to 15 per cent of the total, suggesting that a number of other causes are still obscure. About one-third of patients with no obvious cause for their episodes of acute pancreatitis will eventually be found to have cholelithiasis. Thus, to avoid overlooking a treatable condition, an extensive search for the cause must always be made before the designation ‘idiopathic' is applied.

Pathology

When the acute pancreatitis is mild, pancreatic edema is the prominent feature; this is often referred to as edematous pancreatitis. Grossly, the pancreas appears boggy and indurated. Microscopically, the pancreatic lobules are separated by edema fluid without evidence of hemorrhage. There is a variable amount of interlobular infiltration by inflammatory cells. The acinar cells are preserved, and blood flow appears to be maintained in the small capillaries and venules. In more severe cases, the edema and accompanying inflammation may extend throughout the retroperitoneum and to surrounding organs. Acute fluid collections often develop. Fat necrosis, secondary to the interaction of activated pancreatic enzymes with the fat, is common; this appears as raised, yellow–gray areas in the retroperitoneum, omentum, or mesentery. During such an episode of acute inflammation, both the exocrine and endocrine functions of the gland are impaired for some weeks or even months. However, if the cause (e.g. cholelithiasis) and any complications (e.g. a pseudocyst) of the pancreatitis are eliminated, the pancreas usually returns to normal. Of

course, some scarring may persist after a severe attack, but both exocrine and endocrine function are normal and there is little if any histologic abnormality evident. In most cases, even multiple attacks of acute pancreatitis do not lead to chronic pancreatitis.

The most severe form of acute pancreatitis is characterized by pancreatic necrosis. Grossly, the pancreas first appears hemorrhagic and then becomes necrotic. Its surface may appear devitalized, although it can look normal if the dead tissue is located deep within the gland. Necrotic areas can be found adjacent to healthy lobules, along with an intense inflammatory reaction and cellular infiltration. This has been termed necrotizing or hemorrhagic pancreatitis.

Diagnosis of acute pancreatitis

The typical attack of acute pancreatitis begins with severe and persistent epigastric or upper abdominal pain that often radiates through to the back. Frequently, it follows the ingestion of a large meal and is associated with nausea, and persistent vomiting and retching. The findings are the same regardless of the cause, even if the event represents an episode of acute pancreatic inflammation in a patient with chronic pancreatitis.

The pain is of variable intensity and may be less severe with edematous pancreatitis than with the necrotizing form of the disease. Examination of the abdomen reveals tenderness, most marked in the epigastrium but sometimes present throughout. The bowel sounds are decreased or absent. Usually there are no masses palpable; when one is present, it most often represents a swollen pancreas, pseudocyst, or abscess. In necrotizing pancreatitis, the abdomen may be distended with intraperitoneal fluid. The temperature is only mildly elevated (100–101° F) in uncomplicated cases. There may be evidence of a pleural effusion, especially on the left side.

When the disease is more severe, the patient may also exhibit signs of profound fluid losses from sequestration of edema fluid and/or blood in the peripancreatic retroperitoneal spaces or in the peritoneal cavity (ascites). Severe dehydration, tachycardia, and hypotension may be present. In about 1 per cent of cases, a bluish color is evident around the umbilicus (Cullen's sign), or in the flanks (Grey Turner's sign), which represents blood that has dissected to those areas from the retroperitoneum near the pancreas in patients with hemorrhage into the gland.

Although the clinical presentation often suggests the correct diagnosis and the laboratory findings usually confirm it, it is important to stress that acute pancreatitis is a diagnosis of exclusion. Other acute upper abdominal conditions (e.g. perforated peptic ulcer, acute cholecystitis, gangrenous obstruction of the small bowel) must be considered in every case. If doubt remains, a laparotomy may be indicated for diagnosis (fewer than 2 per cent of cases), since the diseases with which acute pancreatitis is most likely to be confused are often lethal if not treated surgically.

Diagnostic studies

The hematocrit may be elevated because of dehydration, or low as a result of pancreatic or retroperitoneal blood loss. In the absence of suppurative complications, a white blood-cell count over 12 000/µl is unusual. Liver function tests are usually normal, but mild elevations of the serum bilirubin (over 4 mg/dl) may be seen, probably due to partial obstruction of the intrapancreatic portion of the common bile duct by the swollen head of the pancreas. Ultrasonographic examination may reveal evidence of pancreatic and peripancreatic edema or fluid collections. Computed tomographic (**CT**) scans may show similar changes. Nevertheless, these two studies are not commonly needed to diagnose pancreatitis; their main value is in the diagnosis and management of complications of the disease.

The tests most widely used to diagnose acute pancreatitis depend on the release of various pancreatic enzymes from the inflamed gland. The serum amylase concentration rises to more than 2.5 times normal within 6 h of the onset of an acute episode and usually remains elevated for several days. Values above 1000 IU/dl are characteristic (but not diagnostic) of biliary pancreatitis; lower values are typical of acute alcoholic pancreatitis. The sensitivity of amylase in diagnosing pancreatitis ranges from 55 to 80 per cent. Of patients with abdominal pain and hyperamylasemia, 70 per cent are found to have pancreatitis. The serum amylase concentration does not correlate with the severity of the episode of pancreatitis, however.

A patient with acute pancreatitis may have a normal serum amylase for several reasons. (1) The urinary clearance of amylase increases shortly after the onset of pancreatic inflammation. Thus, the serum concentration may fall to normal several days after the onset of the attack because more is excreted in the urine. For this reason, the urinary excretion of amylase may be measured as well; the excretion of more than 5000 IU/24 h is abnormal. (2) In patients with hyperlipidemia and pancreatitis, the chemical determination of amylase is interfered with by the lipid. Serum amylase values are usually falsely normal in this setting (see above). Urinary amylase measurement is still accurate and is usually elevated, however. (3) In patients with chronic pancreatitis who have suffered significant destruction of pancreatic parenchyma, there may be insufficient amylase released to raise the serum concentration above normal, even with an episode of acute inflammation. Pancreatic fibrosis may also limit the permeability of the gland and its capillaries so that absorption of amylase into the blood is restricted.

It is important to recognize that the serum and urinary amylase excretion both may be increased in patients without pancreatitis. Examples are patients with perforated duodenal ulcer, gangrenous cholecystitis, obstruction of the small bowel, and other acute intra-abdominal inflammatory conditions. This increase probably occurs because the pancreatic amylase leaks directly into the peritoneal cavity (e.g. perforated viscus), or translocates through the inflamed, permeable bowel wall and is absorbed into the blood.

The usual laboratory assay measures amylase isoenzymes from both pancreatic (**p-type** amylase) and salivary or other (**s-type** amylase) sources. Thus, elevations in the serum amylase in some patients without pancreatic disease may be due to excessive s-type amylase alone. Examples are salivary or ovarian tumors, chronic sialadenitis, or liver disease. Macroamylasemia is another condition in which the serum amylase is elevated but the pancreas is normal. In this abnormality, the amylase molecules (usually s-type) exist in the serum as large macromolecular aggregates whose effective size and molecular mass are much greater than those of the normal amylase molecule (about 55 000 Da). These macromolecules are too large for glomerular filtration and excretion by the kidney so their serum concentration increases. Urinary excretion of amylase is low. The identification of both macroamylasemia and s-type amylase can be made by electrophoretic analysis of the serum. This should be done when the serum amylase concentration is high but the clinical picture is inconsistent with the diagnosis of pancreatitis.

Other pancreatic enzymes have also been measured in an effort to improve the diagnostic accuracy of serum amylase determinations. The greatest experience has been with serum lipase, which appears to be more reliable than amylase in diagnosing acute pancreatitis, probably because the pancreas is the main source of lipase in the blood. Lipase has a greater sensitivity (85–100 per cent) than amylase and stays elevated longer than amylase, making it a better test for patients who present later in the course of the disease. Like amylase, a variety of other diseases may lead to a rise in lipase, including cholecystitis, perforated viscus, and even bone fractures. If a cut-off of three times the normal lipase concentration is used for diagnosis of acute pancreatitis, the sensitivity and specificity are close to 100 per cent. Radioimmunoassay kits are now available for the rapid determination of serum trypsin and elastase concentrations, and these also appear to be sensitive markers of pancreatic inflammation. To date, they have been used mostly in research protocols.

Radiologic studies are not commonly used to diagnose acute pancreatitis, but certain nonspecific abnormalities are often present. The most frequent finding on a plain abdominal film is the dilatation of an isolated loop of intestine (duodenum, jejunum, or transverse colon) adjacent to the pancreas (sentinel loop). A gas-filled, ascending and right transverse colon that stops abruptly in the mid or left transverse colon (colon cut-off sign) is due to colonic spasm near the inflamed pancreas. An upper gastrointestinal series may show a widened duodenal loop or a swollen ampulla of Vater because of edema in the head of the pancreas. Calcification in the pancreas itself signifies the presence of chronic pancreatitis, but is not diagnostic of acute inflammation. Chest films may reveal pleural effusions, most commonly on the left side.

In patients with biliary pancreatitis a plain abdominal film may show radiopaque gallstones in the region of the gallbladder (10–15 per cent of the time). However, the best way to confirm the presence of gallstones when biliary pancreatitis is suspected is with ultrasound, which may also reveal an edematous, swollen pancreas or acute peripancreatic collections of fluid. Serial ultrasound examinations may be useful in following the clinical course of patients with protracted or complicated pancreatitis or fluid collections. In about 20 per cent of cases of acute pancreatitis, ultrasonography of the pancreas is technically unsatisfactory because of the presence of overlying bowel gas.

A CT scan of the pancreas is not usually required for the diagnosis of acute pancreatitis, but should be obtained in all patients whose illness has not begun to resolve within 2 to 3 days, or whenever complications are suspected. Possible findings include pancreatic edema, a mass of inflamed peripancreatic tissue and pancreas, pancreatic and/or peripancreatic fluid collections (acute pseudocysts), or abscess. Extension of the inflammation from the area of the pancreas to other tissue planes (retrocolic, perinephric) may be evident and implies a more serious prognosis. The adequacy of pancreatic vascular perfusion can also be estimated if the CT examination is done first without, and then with, intravenous contrast material, which must be injected rapidly as a bolus. A viable pancreas ‘enhances' as the contrast material flows through it; lack of enhancement suggests necrosis of that part of the pancreas (Fig. 1). This technique represents a major advance in the diagnosis and management of complicated acute pancreatitis, since patients with significant pancreatic necrosis (more than 50 per cent of the pancreas) are more likely to develop pancreatic infection and to require surgical intervention. If infection is suspected, a percutaneous narrow-gauge needle can be introduced into collections of fluid, especially near poorly perfused pancreatic tissue, using CT or ultrasound as a guide. The aspirated material should be sent for Gram stain, and bacterial and fungal culture. Proof of infection is an indication for laparotomy and surgical drainage.

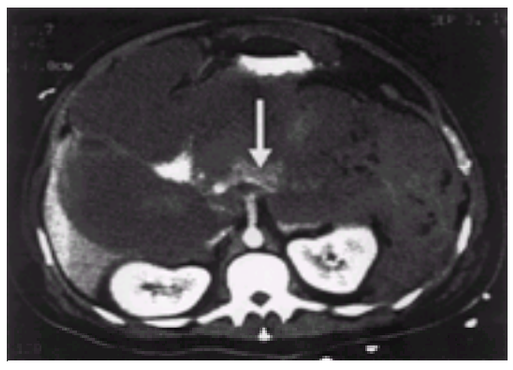


Fig. 1. Helical CT scan of the abdomen revealing very little viable pancreas.

Predicting severity

Prognostic signs

Early identification of patients who are at increased risk for the development of serious complications of pancreatitis allows them to be managed more aggressively, which may decrease the mortality rate. Three different approaches have been used as follows. (1) Specific biochemical markers have been sought that discriminate between edematous and necrotizing pancreatitis, since the latter is known to have a higher rate of complications. In necrotizing pancreatitis, the serum concentration of a_2-macroglobulin, which binds trypsin, falls, and those of a_1-protease inhibitor and complement (C3 and C4) increase. So far, the best discrimination between edematous and necrotizing pancreatitis has been achieved with serum C-reactive protein, a nonspecific acute-phase reactant that is elevated in necrotizing pancreatitis. C-reactive protein is still not widely used outside of the research setting, however. Interleukin (**IL**) 6 has been studied as an early predictor of disease severity. In patients with severe and complicated acute pancreatitis, median serum concentrations of IL-6 are significantly higher than in those with only a mild attack. In other studies, serum concentrations of IL-8 and neutrophil elastase in patients with severe, complicated pancreatitis were higher than in patients with uncomplicated disease. (2) Multiple biochemical and/or physiologic measurements have been made during the first several days after hospital admission, which, taken together, have prognostic value. Ranson developed a series of 11 different measures to estimate the severity of acute pancreatitis (Table 1). Other modifications of these measures exist, but Ranson's original criteria are still used widely. The APACHE II scoring system also has been applied to patients with acute pancreatitis. Unlike the Ranson system, it appears to be a valid estimate of prognosis when used at any time during the disease. (3) Contrast-enhanced CT scanning provides the most objective information about the extent of the inflammatory process, the degree of pancreatic necrosis, and the presence of infection, all of which affect the treatment and the prognosis. However, because it is an invasive and expensive test, it should not be done on all patients. It should be done on those who appear to be at higher risk based on any of the preceding prognostic criteria (e.g. over 3 Ranson's signs), those whose illness is not resolving within 2 to 3 days of admission, or for any other reason if a complication is suspected.

At admission or diagnosis
Age > 55 years
White blood count > 16 000/mm^3
Blood glucose > 200 mg/dl
Serum lactic dehydrogenase > 350 iu./l
Serum glutamic oxaloacetic transaminase > 250 u/dl
During the initial 48 h
Hematocrit fall > 10 per cent
Blood urea nitrogen rise > 5 mg/dl
Serum calcium below 8 mg/dl
Arterial Po_2 < 60 mmHg
Base deficit > 4 mEq/l
Estimated fluid sequestration > 6 l

Table 1 Ranson criteria

Pathophysiology of systemic manifestations

Patients with acute pancreatitis may develop hypotension secondary to hypovolemia caused by fluid sequestration. In response to inflammation, two or more liters of fluid may be lost into the abdominal cavity and retroperitoneum. The elevated hematocrit frequently seen is indirect evidence of the relative hypovolemia that exists.

Oliguria, decreased cardiac output, and hypotension are common during severe acute pancreatitis; these may be the result of hypovolemia from the fluid sequestration. A systemic capillary leak may also occur secondary to the release of inflammatory mediators such as histamine or bradykinin. A so-called myocardial depressant factor has been isolated from the serum of experimental animals with acute pancreatitis and with hypovolemic shock, but this has not yet been characterized.

Oliguria is common in patients with pancreatitis and fluid sequestration, but the urine output usually increases when intravascular volume is repleted. In the majority of patients with uncomplicated disease, the blood urea nitrogen can rise to as high as 40 mg/dl and the creatinine to about 6 mg/dl. These values return to normal with adequate fluid replacement. In more severe cases, or when fluid replacement is inadequate, acute renal failure may ensue. When renal failure complicates acute pancreatitis, the mortality rate may be as high as 70 per cent.

Arterial hypoxemia is present in 50 to 60 per cent of patients with acute pancreatitis, but most do not require endotracheal intubation and mechanical ventilation. When respiratory support is necessary, it indicates a more severe impairment of pulmonary function, with an expected increase in associated mortality. Arterial blood gas measurements and serial chest radiographs should be part of the early assessment in all patients with acute pancreatitis. The majority improve rapidly as the disease resolves, but some deteriorate. Significant progression of respiratory failure within the first 48 h is associated with an increase in fatality. Because hypoxemia usually precedes any observable changes in the chest radiograph, there may be a significant time lag between its appearance and radiographic evidence of pulmonary disease. Eventually, abnormalities on a chest radiograph can be seen in over half of all patients with acute pancreatitis.

Many of the physiologic abnormalities observed in the respiratory failure associated with acute pancreatitis are similar to those seen in adult respiratory distress syndrome occurring from other causes. The mechanisms of pulmonary injury in acute pancreatitis are unknown. Sepsis may contribute to the respiratory failure in patients with infection, but many who develop pulmonary dysfunction have sterile disease. High serum concentrations of pancreatic enzymes may contribute to respiratory failure. Cytokines released from the pancreas or liver have also been implicated (see later).

Significant abnormalities occur in the metabolism of glucose and calcium. Elevated blood glucose concentrations (130–200 mg/dl) are seen in 50 to 75 per cent of patients with acute pancreatitis. Although blood sugars greater than 200 mg/dl are associated with increased mortality, this is only a reflection of the severity of the pancreatitis. Death rarely occurs as a result of hyperglycemia or of diabetic complications alone. A number of factors contribute to the hyperglycemia. In patients with acute pancreatitis the serum glucagon concentrations are much higher than in patients with stress from other causes. Elevated serum growth hormone and cortisol concentrations, both of which can produce hyperglycemia, are also found in patients with acute pancreatitis. The elevated concentrations of glucagon and glucose usually return to normal within 18 to 21 days after resolution of the pancreatitis.

Hyperglycemia in patients with acute pancreatitis should not be treated unless it is severe. Insulin is not very effective in lowering the serum glucose of patients with hyperglucagonemia, so high doses may be necessary. Because the hyperglucagonemia of acute pancreatitis tends to resolve suddenly and unpredictably, giving large amounts of insulin to these patients to correct an elevated blood glucose may produce dangerous hypoglycemia.

Mild hypocalcemia is frequent in patients with pancreatitis and, while the serum calcium is usually only minimally depressed, profound hypocalcemia can occur; this implies a poor prognosis. A number of factors may contribute to the hypocalcemia, as follows. (1) Hypoalbuminemia is common in acute pancreatitis because vascular

permeability is increased and albumin is lost from the intravascular space. Since some of the serum calcium is normally bound to albumin, some of it is lost in this way. In this situation, the serum ionized calcium is often normal despite a low total serum calcium. Therefore, in the laboratory examination of a patient with acute pancreatitis, calcium should be measured in conjunction with albumin to determine accurately the amount of ionized calcium. Otherwise, if only total calcium is measured, a falsely low value for ionized calcium may be obtained. (2) Fat necrosis, with saponification of calcium into soap, also may account for the loss of as much as 1 to 2 g of calcium. However, this is not sufficient to account for the level of hypocalcemia sometimes seen in acute pancreatitis.

Parathyroid hormone production, release, and peripheral activity in response to hypocalcemia are all normal in acute pancreatitis. Hypomagnesemia may also be present in patients with acute pancreatitis, which is important since hypocalcemia may be impossible to correct without also treating the hypomagnesemia. In the majority of patients with acute pancreatitis, hypocalcemia is transient and mild. Calcium replacement is not necessary unless the ionized calcium is also decreased or the patient is symptomatic. Hypokalemia, if present, should be corrected at the same time, because rapid correction of hypocalcemia in this clinical setting may precipitate cardiac arrhythmias.

Cytokines

Although the release of activated pancreatic enzymes into the circulation undoubtedly accounts for some of the manifestations of acute pancreatitis, this cannot explain all the systemic effects of severe disease. Recent studies suggest that the production and release of a variety of inflammatory mediators by the inflamed pancreas contributes as well. For example, acute pancreatitis is associated with the rapid release of IL-1, IL-6, and tumor necrosis factor (**TNF**)-a into the serum, and the serum concentrations of these cytokines correlate well with the degree of pancreatic inflammation. One study found that serum from patients admitted with a diagnosis of acute pancreatitis had elevations of IL-6 that correlated with disease severity. IL-6 concentrations correlate with those of two other markers, C-reactive protein and phospholipase-A activity, suggesting a causal role for IL-6 in the acute-phase response. IL-1, IL-6, and TNF-a are all elevated in the serum of animals with acute pancreatitis. One study also demonstrated that pretreatment with anti-TNF antibodies attenuated the expected rise of serum TNF, glucose, and amylase.

Platelet-activating factor (**PAF**), another inflammatory mediator, also seems to play an important part in the pathogenesis of the acute disease. PAF is a low molecular-weight phospholipid that stimulates platelet, macrophage, and neutrophil aggregation, and promotes hypotension and capillary leakage. Treatment with PAF antagonists appears to ameliorate the severity of acute pancreatitis.

The mechanisms of cytokine production are only beginning to be defined. In the pancreas, they appear to be released primarily from tissue macrophages, although recruitment of leucocytes and other inflammatory cells may contribute to the process as well. The release of histamine and other secretory proteins into the interstitial space from the inflamed pancreas may cause the initial activation of macrophages and endothelial cells, with release of such cytokines as IL-8, PAF, and with neutrophil recruitment. Cytokines derived from the pancreas may travel to the liver via the portal venous circulation; there they could cause liver injury and additional release of cytokines derived from the hepatic Kupffer cells. These cytokines could enter the systemic circulation via the hepatic venous drainage and be responsible for the widespread effects on the other organ systems.

Treatment of acute pancreatitis

The treatment of uncomplicated acute pancreatitis is medical, and is directed primarily toward the restoration of fluid and electrolyte balance and the avoidance of secretory stimulation of the pancreas. Nasogastric suction should be instituted in many patients, except those with mild disease not associated with significant vomiting or pain. The placement of a nasogastric tube does not appear to affect the course of the disease, but rather serves to improve the symptoms associated with gastric distension.All oral intake should be withheld until the ileus has resolved and pain is absent. The enzyme abnormalities have usually returned to normal by that time. For unclear reasons, the serum enzymes sometimes remain elevated even though the patient is asymptomatic. The resumption of oral intake is usually tolerated in these circumstances. Occasionally, symptoms will recur when oral intake is resumed. When this happens, another period of fasting is indicated. Fluid requirements may be considerable since large volumes can be sequestered in the retroperitoneum adjacent to the pancreas. The amount of crystalloid and colloid (albumin, blood) given should be sufficient to maintain an adequate hematocrit, circulating blood volume, and urine output. It should be stressed that the most important aspect of early resuscitation is adequate and aggressive fluid replacement. Renal failure from inadequate replacement is a frequent finding in patients who die from this disease. Moreover, the pancreatitis itself may progress from the impaired pancreatic perfusion that can accompany the shock state.

Electrolyte derangements may be present if vomiting has been significant and dehydration is severe. A hypokalemic, hypochloremic metabolic alkalosis is most common. Hypocalcemia may occur in patients with severe pancreatitis. It must be treated urgently with parenteral calcium because it predisposes to cardiac arrhythmias. The prognosis appears to be related to the degree of hypocalcemia and the ease with which it is corrected. Hypomagnesemia is also common, particularly in alcoholic patients, and the magnesium should be replaced. Patients with acute pancreatitis who develop arterial hypoxemia (*P* AO_2 less than 70 mmHg) require supplemental O_2. Because the onset of this is often insidious and may precede any changes on the chest film, it may be prudent in most patients to determine arterial blood gases every 12 h for the first several days after admission to the hospital, and for longer in those who appear to have more serious disease. The occasional patient who develops adult respiratory distress syndrome will require endotracheal intubation and mechanical ventilation.

A variety of drugs have been shown to exert no beneficial effect in patients with acute pancreatitis. These include anticholinergics, gastric acid antisecretory agents, glucagon, somatostatin, and inhibitors of proteolytic enzymes (e.g. aprotinin). None of these medications has been found to alter significantly the course of pancreatitis. Recent data from studies in patients with severe acute pancreatitis suggest that treatment with a PAF antagonist (Lexipafant®) was of value. Eighty-three patients were randomized either to Lexipafant® treatment or to placebo for 3 days. The active drug reduced serum IL-8 on day 1 and prevented the rise in IL-6. After 72 h there was a reduction in the severity of organ failure. Seven of 12 patients with severe disease receiving Lexipafant® recovered from organ failure, in contrast to only two of 11 patients receiving saline. Further studies are under way in Europe and the United States. In the future, IL-1 receptor antagonists and Il-10 administration may prove beneficial to the patient with severe disease.

Parenteral nutrition has no specific beneficial effect on the course of pancreatitis. It should be used as in any other patient from whom oral intake must be withheld for prolonged periods and enteral nutrition is not possible. A prospective randomized study evaluating parenteral nutrition in patients with pancreatitis found no difference in complications or hospital stay. Elemental diets introduced via tube into the small intestine do not avoid pancreatic stimulation.

Peritoneal lavage has been used in patients with severe acute pancreatitis and intraperitoneal fluid to remove toxins and various metabolites from the peritoneal cavity, and minimize their systemic absorption. Although controlled trials failed to show that this decreased mortality (Table 2), there was some evidence that the incidence of cardiopulmonary complications was lessened. For this reason, and because there are anecdotal reports of patients improving rapidly during lavage therapy, it is still used occasionally.

Study*	Treatment	n	Complications (%)	Deaths (%)
Trapnell et al.	Control	12	50	17
	Lavage—7 days	12	75	36
Ellis et al.	Control	20	30	5
	Lavage	19	42	21
Mayer et al.	Control	45	38	35
	Lavage—3 days	45	71	38
Ranson et al.	Lavage—2 days	15	40	20
	Lavage—7 days	14	22	0

*See Further Reading.

Table 2 Utility of peritoneal lavage in acute pancreatitis

Endoscopic intervention

Biliary pancreatitis is caused by a gallstone that obstructs the pancreatic and bile ducts. In the majority of cases, the gallstone dislodges spontaneously and usually passes into the intestine. When this happens, the patient's clinical course begins to improve and the episode is generally short lived. If the stone remains impacted, the attack of pancreatitis may become severe and refractory to the usual treatment. Impaction should be suspected when a patient with pancreatitis is known to have gallstones, when the serum bilirubin concentrationis elevated above 4 mg/dl, when the alkaline phosphatase is elevated, and when the clinical course does not improve within 24 to 36 h. If a skilled endoscopist is available, endoscopic retrograde cholangiopancreatography with endoscopic sphincterotomy to remove the stone may abort

the attack. Four randomized studies have evaluated the utility of urgent sphinterotomy and stone extraction (Table 3); they appear to show benefit in patients who have severe disease and are not responding to resuscitative therapy. If a skilled endoscopist is not available, surgical removal of the impacted stone is indicated.

Study*	Treatment	Complications (%)	Deaths (%)
Neoptolemos *et al.*	ERCP/ES	19	0
	Conservative	63	13
Nowak *et al.*	ERCP/ES	14	1
	Conservative	34	11
Fan *et al.*	ERCP/ES	20	3
	Conservative	76	18
Fölsch *et al.*	ERCP/ES	17	5
	Conservative	14	3

*See Further Reading.

Table 3 Early endoscopic retrograde cholangiopancreatography and endoscopic sphincterotomy (ERCP and ES) in acute and severe gallstone pancreatitis

Surgical treatment

Surgery is contraindicated in uncomplicated acute pancreatitis. However, surgery is occasionally performed in patients with abdominal pain where the diagnosis is unclear and they are found to have acute pancreatitis. If the pancreatitis is mild to moderate in severity and gallstones are present, a cholecystectomy and operative cholangiogram should be obtained. Stones in the common duct should be removed. The pancreas should not be disturbed. For severe pancreatitis, the biliary tree should be drained if stones are present (cholecystostomy), but cholecystectomy probably should not be done. The common duct should be decompressed with a T-tube if stones are present; if a stone is impacted at the ampulla of Vater, it should be removed. It is unclear whether drains should be placed in the vicinity of the pancreas; many surgeons would not do this, thinking that the foreign body might predispose to infection. If necrotic pancreas or peripancreatic tissue is found, it should be debrided and drained widely.

In most patients with biliary pancreatitis, the diagnosis is obvious without the need for surgery and the pancreatitis resolves within 2 or 3 days. In these patients, cholecystectomy should be performed during the same stay, usually 4 to 6 days after the original admission. The older practice of discharging the patient and delaying cholecystectomy for 6 weeks was associated with a 60 to 80 per cent incidence of recurrence of pancreatitis. On the other hand, patients with severe episodes of biliary pancreatitis should be discharged once they have recovered, and cholecystectomy should be deferred for at least 6 weeks. By this time the inflammation has subsided and the operation can be performed without increased risk.

Surgery in severe acute pancreatitis

The rationale for surgery in severe pancreatitis has evolved over the last 50 years. In the early part of the twentieth century, a total pancreatectomy was often recommended, which included resection of living pancreas as well as that portion which had undergone necrosis. The reasoning was that removal of all of the gland would abort the inflammatory process and recovery would then ensue. Although some patients also had pancreatic and/or peripancreatic infection, the presence of infection was not a critical determinant in the decision to operate. The approach was soon abandoned because of unacceptably high mortality rates in these critically ill patients.

Today it is clear that those patients with infection of the necrotic pancreas or peripancreatic debris benefit from surgical debridement and drainage of the infected and devitalized tissue. Viable pancreas is not removed. The systemic organ failure that ultimately is responsible for the demise of such patients is believed to be caused by the infection. Surgical drainage and debridement of this material provide the best chance of recovery.

There remains a group of critically ill patients with severe acute pancreatitis, usually with significant pancreatic and/or peripancreatic necrosis, in whom infection is not present. The role of surgery in these patients with so-called sterile necrosis continues to be controversial. In general, evidence is accumulating that this group is managed effectively by nonsurgical means in the majority of cases. When surgery is performed, it is done in an attempt to drain and debride the necrotic tissue which is the putative source of various noxious substances believed to be responsible for the ongoing illness. Or it may even be that infection was present, but that its diagnosis was overlooked. The validity of this approach is as yet unproved.

Risk factors for pancreatic infections

Secondary infection of necrotic pancreatic and peripancreatic tissue is a common source of morbidity in acute pancreatitis. The origin of the bacteria is generally felt to be the gastrointestinal tract, although hematogenous seeding from infected intravenous lines or other sources could occur. Infections are more likely in patients with extensive pancreatic necrosis and severe pancreatic inflammation. Pancreatic infection occurs in up to 10 per cent of all cases of acute pancreatitis. Acute necrotizing pancreatitis develops in about 10 per cent of all patients suffering from acute pancreatitis, and in them about 30 per cent become infected.

Another risk factor for pancreatic infection is the duration of the disease. Beger found that among patients who underwent surgical exploration within the first week after the onset of pancreatitis, 24 per cent had infected pancreatic necrosis. From the second to the third week there was an increase of from 36 cent to 71 per cent, and then a decline to an infection rate of 32.5 per cent when operation was done after the fourth week. Similar results were reported for CT-guided fine-needle aspiration of pancreatic necrosis.

Diagnosis and bacteriologic findings

The diagnosis of pancreatic infection is made most reliably by CT- or ultrasound-guided, fine-needle aspiration, with Gram staining and culture of the aspirate. The technique is safe, accurate, and can be performed rapidly. Although it is tempting only to perform fine-needle aspiration in patients who exhibit a septic clinical picture (e.g. high fever and white-cell count), some patients with infection have a low-grade fever and a white-cell count below 15 000. Thus, in most patients, fine-needle aspiration should be performed in those who have evidence of necrosis and fluid collection on the CT scan.

Pancreatic infections usually are caused by Gram-negative enteric bacteria. The most frequently isolated pathogens are *Escherichia coli* (approximately 35 per cent), *Klebsiella pneumonia* (approximately24 per cent) and *Enterococcus* spp. (approximately 24 per cent). Other less frequently isolated organisms are staphylococci (approximately 14 per cent), *Pseudomonas* ssp. (approximately 11 per cent), *Proteus* (approximately 8 per cent), aerobic streptococci (approximately 7 per cent), *Enterobacter* (approximately 7 per cent), and *Bacteroides* (approximately 6 per cent). Anaerobic pathogens occur in about 6 per cent of cases, and fungal infections are seen frequently, especially in patients who have been on multiple antibiotics for long periods. It may be that monomicrobial infections (58 per cent of cases) are more likely to contain Gram-positive organisms and that polymicrobial infections (42 per cent of cases) contain Gram-negative pathogens.

All patients with necrotizing pancreatitis are treated prophylactically with a broad-spectrum antibiotic (e.g., imipenem) until therapy can be further refined by sensitivity testing of material obtained by fine-needle aspiration or at operation. However, in patients with proved infection, antibiotic therapy is adjunctive only; definitive management requires surgery.

Timing of surgery

Infected pancreatic necrosis

In most cases, infection is proved by Gram stain and culture of pancreatic or peripancreatic fluid obtained by fine-needle aspiration. In a minority of patients, gas bubbles are evident on CT in the area of the pancreas (Fig. 2): if so, fine-needle aspiration is unnecessary. Gas should be assumed to be the product of bacterial fermentation from infection. In either case, the presence of infected pancreatic necrosis is an undisputed indication for surgical intervention. The patient's condition should be optimized and surgery undertaken within 24 to 48 h.

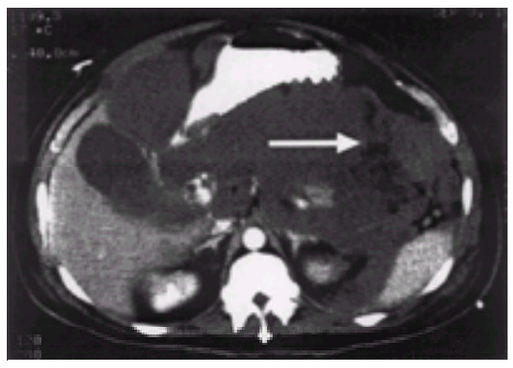


Fig. 2. Helical CT scan revealing gas bubbles in the pancreatic phlegmon.

Except in the unusual condition of fulminating acute pancreatitis with organ failure and a rapidly progressive downhill course soon after admission to the hospital, most of these patients should not undergo operation during the first week of their illness. When clinical deterioration is rapid and surgery is undertaken during the first week, most die. The outcome is better when surgery is postponed at least until the second week or later, when the margins of the pancreatic necrosis have become better defined and the acute inflammation has subsided somewhat. Fortunately, in most cases the disease has been progressing for a week or more by the time the diagnosis of infected pancreatic necrosis has been established and the need for surgery is evident.

Sterile pancreatic necrosis

Almost all patients with sterile pancreatic necrosis should receive conservative, non-surgical management in the intensive care unit. Most will eventually recover. Nevertheless, some fail to improve after weeks of treatment, or they may even begin to deteriorate. In these patients, consideration for surgery still seems reasonable for the reasons presented earlier. The timing of operation in this small group cannot be given precisely. In general, it seems best to continue to treat them conservatively for at least 3 to 4 weeks before concluding that surgery should be performed.

Surgical techniques

The goals of surgery are to remove infected and devitalized pancreatic and peripancreatic tissue, to drain pus and other fluid collections, and to leave drains behind that can be used for continuous postoperative lavage of the affected areas. Wide exposure is required for a thorough exploration of the peripancreatic and retroperitoneal tissues. The CT scan provides a ‘road map' to those areas that require drainage, so that uninvolved tissue planes do not need to be opened and unnecessarily contaminated. Viable pancreatic parenchyma should be preserved. Hemostasis may require suture ligation of bleeding vessels, but significant bleeding usually suggests that the surgeon should limit further dissection in that area. Indeed, the inflammatory process has usually thrombosed many of the vessels that otherwise would have bled extensively. Once most of the necrotic material has been removed and all of the fluid collections drained, several large sump drains are placed for postoperative lavage in the areas most involved. Normal saline at a rate of 1 l/h is used for at least the first 24 h and then gradually the rate is decreased over a number of days, according to the patient's clinical course and the character of the drainage. The fascia of the abdominal wound is closed and the skin left open.

Patients with extended areas of infected necrotic tissue, or large multiloculated or multiple abscesses, may be treated by open drainage and packing. The operative procedure is the same, but the debrided areas are packed and the incision is left open. This approach requires reoperation under general anesthesia every 2 to 3 days until it is evident that there is no more infection and clean granulation tissue begins to cover the debrided surfaces. Then the packing is removed and the wound is allowed to heal by secondary intention.

Postoperative course

Patients who are treated by open packing are committed to at least several reoperations before they recover completely. However, about 20 per cent of patients treated with the closed technique also require at least one reoperation to drain recurrent or persistent areas of infection. An abdominal CT scan obtained at weekly intervals will help guide further therapy and document improvement during the postoperative course.

The most common local postoperative complications are hemorrhage and intestinal fistulas. Postoperative hemorrhage is associated with a high mortality rate, as bleeding often occurs from major peripancreatic blood vessels (e.g. splenic, superior mesenteric or portal veins; gastroduodenal or pancreatic arteries). Arteriography with embolization of bleeding vessels may be effective, but some patients require urgent operation to stop the bleeding. Enterocutaneous fistulas (pancreatic, duodenal, small bowel, or colon) occur in up to 30 per cent of patients. They can be caused directly by the necrotizing infection, by iatrogenic trauma at the time of the debridement, or by erosion into the bowel of an adjacent surgical drain. Most of the bowel fistulas eventually close spontaneously without operative intervention.

Prognosis

The mortality rate for all patients with acute pancreatitis is about 10 per cent. Necrotizing pancreatitis associated with infection has a mortality rate of about 20 per cent, although there is some evidence that earlier diagnosis of infection and aggressive surgical intervention may lower this figure.

Long-term complications

Pancreatic abscess

A pancreatic abscess is a collection of purulent material within a defined cavity and with little if any associated necrosis (Fig. 3). It is different from infected pancreatic necrosis and an infected pancreatic pseudocyst. Pancreatic abscesses usually require 3 to 4 weeks to become apparent. Diagnosis should be suspected when the patient begins to run a septic course, and it can be confirmed with CT and percutaneous fine-needle aspiration of the pus. An external drain should be left in place. Unlike infected pancreatic necrosis, nonsurgical tube drainage may be effective. However, unless the patient improves rapidly within 24 to 48 h of external drainage, surgical drainage of the abscess is required. Antibiotics are adjunctive only.

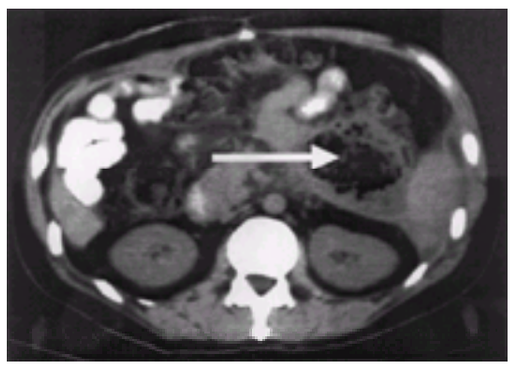


Fig. 3. CT scan demonstrating a pancreatic abscess.

Pancreatic pseudocyst

A pancreatic pseudocyst is a collection of fluid, usually in the vicinity of the pancreas, which develops in association with a leak of pancreatic juices from the inflamed parenchyma or from a disrupted duct (Fig. 4). The wall of the pseudocyst is comprised of fibrous, non-epithelialized tissue. Occasionally a pseudocyst may present at

great distance from the pancreas (e.g. thorax, groin) when the fluid dissects through tissue planes. As many as 30 per cent of patients with acute pancreatitis form acute fluid collections around the time of theacute attack, but these must be distinguished from chronic pseudocysts. The majority of these acute ‘pseudocysts' resolve without intervention. Only about 5 per cent of these patients develop chronic pseudocysts, which are characterized by their ovoid or spherical shape and well-formed wall. Because of this natural history, acute ‘pseudocysts' (fluid collections) should be managed expectantly. If they develop into chronic pseudocysts, they may require treatment.



Fig. 4. Endoscopic retrograde cholangiopancreatogram demonstrating a pancreatic pseudocyst communicating with the main pancreatic duct.

The management of pseudocysts varies according to their size and the presence of associated symptoms. Asymptomatic pseudocysts up to 5 to 6 cm in diameter may be safely observed, and are usually followed with either serial ultrasound or CT examinations. Larger cysts or pseudocysts of any size that are symptomatic require treatment.

Symptoms are most often from gastrointestinal obstruction when the cyst distorts the stomach or duodenum, or abdominal pain. Serious complications also can occur, although they are uncommon (fewer than 5 per cent of cases); these include hemorrhage into the cyst, perforation of the cyst, and infection of the cyst. Hemorrhage is usually caused by erosion of the splenic or gastroduodenal artery or other major vessel within the wall of the cyst, and the bleeding is usually confined to the cyst lumen. The diagnosis should be suspected if there are clinical signs of hypovolemia and a falling hematocrit. There may be abdominal pain and a mass may be palpable. An abdominal CT scan shows the cyst with the contained blood clot. Angiography confirms the diagnosis and the radiologist should attempt to embolize the bleeding vessel. If not, emergency surgery with ligation of the vessel or excision of the cyst is required. Perforation of a pseudocyst is a surgical emergency that is characterized by the sudden onset of intense abdominal pain with peritonitis. Patients require urgent surgery, with irrigation of the peritoneal cavity and usually external cyst drainage. Infection of a pseudocyst should be suspected if signs of sepsis develop. Diagnosis by CT scan, and treatment by percutaneous cyst aspiration and drainage, are usually effective.

In the absence of a life-threatening complication, elective surgery of pseudocysts is usually delayed until the cyst has developed a mature wall that will hold sutures at the time of repair. For those cysts that develop following an episode of acute pancreatitis, this requires 4 to 6 weeks. In most cases, the patient can eat and be discharged from the hospital during the interval. Pseudocysts that resolve spontaneously will usually do so during this time.

Pseudocysts may be treated surgically, or by endoscopic or radiologic drainage. Endoscopic methods require the placement of a plastic stent through the stomach or duodenal wall into the adjacent cyst. The stent is eventually removed and in about 80 per cent of cases the cyst is permanently eradicated. These endoscopic techniques require expertise that still is not widely available. Radiologic approaches usually consist of percutaneous external drainage of the cyst, with eventual removal of the drainage catheter many weeks later. Many of these pseudocysts recur. Surgical treatment usually consists of drainage of the cyst internally to either the stomach (cystgastrostomy) or to a Roux-en-Y limb of jejunum (cystjejunostomy). Both are safe and effective, with recurrence rates under 10 per cent. If the pseudocyst is in the tail of the pancreas, a distal pancreatectomy with excision of the cyst may be best. The recurrence rate is under 1 per cent.

Pancreatic fistula

In the setting of acute pancreatitis, a pancreatic fistula is usually the result of a ductal disruption associated with pancreatic necrosis and it usually becomes apparent as the patient recovers from operative debridement. In a large series of 556 patients, pancreatic fistulas occurred in 9 per cent of those with necrotizing pancreatitis. The diagnosis is made by finding a high amylase (usually many thousands of units per liter) in the drain effluent, and the condition is treated by leaving the drain in place. Many such fistulas will close spontaneously, provided that ductal continuity can be re-established as healing occurs, infection is eradicated, and nutrition is adequate. Parenteral nutrition is usually not required and most patients are able to eat a regular diet. There is no evidence that oral intake delays resolution of fistula. The use of somatostatin does not appear to hasten closure of a fistula, although if it is a high-output fistula (i.e., more than 200 ml/day), the secretory inhibitor may simplify management of the patient. Fistulas that persist for as long as 1 year, or those whose anatomic characteristics preclude spontaneous closure (e.g. duct obstruction between fistula and duodenal lumen; duct discontinuity), will require operative repair. This is best done by creating an anastomosis between the pancreatic duct at the point of the leak and a Roux-en-Y limb of jejunum. The success rate of operative repair is greater than 90 per cent.

Further reading

Agarwal N, Pitchumoni CS, Sivaprasad AV. Evaluating tests for acute pancreatitis. *American Journal of Gastroenterology* 1990; **85**: 356–66. [The use of serum pancreatic enzymes, ultrasonography (US), and computed tomography (CT) are evaluated in making the diagnosis of pancreatitis.]

Austin J, Reber HA. The pathologic features of acute pancreatitis. In: Howard J, Jordan G, Reber HA, ed. *Surgical diseases of the pancreas*. Lea and Febiger, Philadelphia, 1986.

Beger HG, Bittner R, Block S, Buchler M. Bacterial contamination of pancreatic necrosis: a prospective clinical study. *Gastroenterology* 1986; **91**: 433–8. [This study documents the temporal rates of bacterial infection of pancreatic necrosis.]

Buchler M, Malfertheiner P, Schoetensack C, Uhl W, Beger HG. Sensitivity of antiproteases, complement factors and C-reactive protein in detecting pancreatic necrosis. Results of a prospective clinical study. *International Journal of Pancreatology* 1986; **1**: 227–35. [This study found that C-reactive protein and a_2-macroglobulin have a 95 and 85 per cent overall detection rate for pancreatic necrosis, respectively.]

Clavien PA *et al.* Acute pancreatitis and normoamylasemia: not an uncommon combination. *Annals of Surgery* 1989; **210**: 614–20. [A cohort of 352 patients with acute pancreatitis were studied for causes of normal amylase levels during the attack. Normal amylase does not predict severity in these patients and the authors caution against the use of amylase for this purpose.]

Corfield AP, Cooper MJ, Williamson RCN. Acute pancreatitis: a lethal disease of increasing incidence. *Gut* 1985; **26**: 724–9. [This study characterizes a large cohort of patients in England with acute pancreatitis and documents the etiology and outcome in these patients.]

Demmy T, Burch JM, Feliciano DV, Mattox KL, Jordan GL JR. Comparison of multiple-parameter prognostic systems in acute pancreatitis. *American Journal of Surgery* 1988; **156**: 492–5. [This study examines several classification systems for pancreatitis and finds each fairly equal with regard to the ability to predict outcome.]

Dominguez-Munoz JE *et al.* Monitoring of serum proteinase–antiproteinase balance and systemic inflammatory response in prognostic evaluation of acute pancreatitis. Results of a prospective multicenter study. *Digestive Diseases and Sciences* 1993; **38**: 507–13. [This study examined a_2-macroglobulin, C3, C4, a_1-protease inhibitor, C-reactive protein, and polymorphonuclear elastase in differentiating between severe and mild pancreatitis. The study concludes that elastase is a useful marker.]

Evans FC. Pancreatic abscess. *American Journal of Surgery* 1969; **117**: 537–40. [This article reviews the presentation and treatment of pancreatic abscesses.]

Fan ST, Lai EC, Mok FP, Lo CM, Zheng SS, Wong J. Early treatment of acute biliary pancreatitis by endoscopic papillotomy. *New England Journal of Medicine* 1993; **328**: 228–32. [This large prospective study examines the use of early ERCP and papillotomy for the treatment of acute pancreatitis. The group concludes that it is very useful and may lower mortality from this disease.]

Folsch UR, Nitsche R, Ludtke R, Hilgers RA, Creutzfeldt W. Early ERCP and papillotomy compared with conservative treatment for acute biliary pancreatitis. The German Study Group on Acute Biliary Pancreatitis. *New England Journal of Medicine* 1997; **336**: 237–42. [This study finds that early ERCP is useful in patients with acute pancreatitis and jaundice.]

Gerzof SG *et al.* Early diagnosis of pancreatic infection by computed tomography-guided aspiration. *Gastroenterology* 1987; **93**: 1315–20. [This study determines that percutaneous fine needle aspiration of pancreatic necrosis is a safe and accurate way to determine infection.]

Grewal HP, Mohey el Din A, Gaber L, Kotb M, Gaber AO. Amelioration of the physiologic and biochemical changes of acute pancreatitis using an anti-TNF-a polyclonal antibody. *American Journal of Surgery* 1994; **167**: 214–18. [Pretreatment with anti-TNF antibody in rats improves the parameters of pancreatitis after bile acid duct injection-induce pancreatitis.]

Gumaste V, Roditis N, Mehta D, Dave PB. Serum lipase levels in nonpancreatic abdominal pain versus acute pancreatitis. *American Journal of Gastroenterology* 1993; **88**: 2051–5. [The authors evaluate the use of serum amylase and lipase in differentiating pancreatitis from other abdominal pathology.They conclude that a lipase level three times higher than normal is most useful in differentiating these two.]

Heath DI, Cruickshank A, Gudgeon M, Jehanli A, Shenkin A, Imrie CW. Role of interleukin-6 in mediating the acute phase protein response and potential as an early means of severity assessment in acute pancreatitis. *Gut* 1993; **34**: 41–5. [Utilizing a peak interleukin-6 concentration of above 130 u/ml, this study was able to distinguish between severe and mild attacks of acute pancreatitis with a sensitivity of 100 per cent and specificity of 71 per cent.]

Hurley JE, Vargish T. Early diagnosis and outcome of pancreatic abscess in pancreatitis. *American Surgeon* 1987; **53**: 29–33. [This study finds that patients with high levels of lactate dehydrogenase, leukocytosis, metabolic acidosis, hypoxemia, hypocalcemia, and fluid sequestration have a higher chance of formation of pancreatic abscesses and a worse outcome.]

Ihse I, Evander A, Gustafson I, Holmberg JT. Influence of peritoneal lavage on objective prognostic signs in acute pancreatitis. *Annals of Surgery* 1986; **204**: 122–7. [Although serum amylase levels fell quicker in patients with pancreatitis who underwent peritoneal lavage, no significant improvement in complications, mortality, or hospital stay was observed.]

Isenmann R, Büchler MW. Infection and acute pancreatitis. *British Journal of Surgery* 1994; **81**: 1707–8. [This review describes the etiology and treatment of infections manifesting as a result of acute pancreatitis.]

Kingsnorth A, Galloway S, Formela L. Randomized, double blind phase II trial of lexipafant, a platelet-activating factor antagonist, in human acute pancreatitis. *British Journal of Surgery* 1995; **82**: 1414–20. [This prospective European study documents the use of Lexipafant in acute pancreatitis and finds that it reduced IL-6 and IL-8 levels and improved mortality in patients with severe disease.]

Konturek SJ *et al.* Role of platelet activating factor in pathogenesis of acute pancreatitis in rats. *Gut* 1992; **33**: 1268–74. [This study finds that platelet activating factor diminishes blood flow, increases pancreatic weight, and impacts pancreatic pathology in a similar way to caerulein-induced pancreatitis in rats.]

Lefer AM. Vascular mediators in ischemia and shock. In: Cowley RA, Trump BF, ed. *Pathophysiology of shock; anoxia, and ischemia*. Williams and Wilkins, Baltimore, 1982. [This paper describes the mediators of shock and specifically discusses a myocardial depressant factor.]

Leser HG *et al.* Elevation of serum interleukin-6 concentration precedes acute-phase response and reflects severity in acute pancreatitis. *Gastroenterology* 1991; **101**: 782–5. [This study found that the levels of interleukin-6 correlate with the severity of pancreatitis and death from pancreatitis.]

Levant JA, Secrist DM, Resin H, Sturdevant RA, Guth PH. Nasogastric suction in the treatment of alcoholic pancreatitis. A controlled study. *Journal of the American Medical Association* 1974; **229**: 51–2. [This randomized controlled clinical trial determines that nasogastric suctioning does not effect outcome in patients with acute pancreatitis.]

Mayer AD *et al.* controlled clinical trial of peritoneal lavage for the treatment of severe acute pancreatitis. *New England Journal of Medicine* 1985; **312**: 399–404. [This study examines 91 patients with severe acute pancreatitis and determines that routine peritoneal lavage offers no advantage.]

Naeije R, Salingret E, Clumeck N, Detroyer A, Devis G. Is nasogastric suction necessary in acute pancreatitis? *British Medical Journal* 1978; **ii**: 659–60. [This study concludes that, in patients with mild to moderate acute pancreatitis, the use of nasogastric suctioning does not effect outcome.]

Nath BJ, Warshaw, AL. Pulmonary insufficiency. In: Bradley EL, ed. *Complications of pancreatitis*. Saunders, Philadelphia, 1982. [This paper specifically addresses pulmonary complications during acute pancreatitis.]

Neoptolemos JP, Carr-Locke DL, London NJ, Bailey IA, James D, Fossard DP. Controlled trial of urgent endoscopic retrograde cholangiopancreatography and endoscopic sphincterotomy versus conservative treatment for acute pancreatitis due to gallstones. *Lancet* 1988; **ii**: 979–83. [In this study, patients with severe gallstone pancreatitis who underwent early ERCP fared better than those who did not with regard to complications, death rate, and length of hospital stay.]

Nowak A, Nowakowska-Dulawa E, Marek T A, Rybicka J. Final results of the prospective, randomized, controlled study on endoscopic sphincterotomy versus conventional management in acute biliary pancreatitis. *Gastroenterology* 1995; **108**: A380. [The authors conclude that early ERCP diminishes complications and mortality in patients with gallstone pancreatitis.]

Opie EL. The relation of choledocholithiasis in disease of the pancreas and to fat necrosis. *Bulletin of the Johns Hopkins Hospital* 1901; **12**: 19–20. [Describes the autopsy findings in two patients who died of pancreatitis.]

Pederzoli P, Bassi C, Vesentini S. A randomized multicenter clinical trial of antibiotic prophylaxis of septic complications in acute necrotizing pancreatitis with imipenem. *Surgery, Gynecology, Obstetrics* 1993; **176**: 480–3. [The authors conclude that in patients with pancreatic necrosis imipenern decreases the incidence of subsequent infection of this necrotic material.]

Pezzilli R *et al.* Serum interleukin-6, interleukin-8, and b_2-microglobulin in early assessment of severity of acute pancreatitis. Comparison with serum C-reactive protein. *Digestive Diseases and Sciences* 1995; **40**: 2341–8. [The results of this study suggests that, when assayed during the first 24 h of disease onset, interleukin-6 and interleukin-8 are better markers than b_2-microglobulin or C-reactive protein for evaluating the severity of acute pancreatitis.]

Ranson JHC. Etiologic and prognostic factors in human acute pancreatitis: a review. *American Journal of Gastroenterology* 1982; **77**: 633–8. [The authors conclude that lavage lasting at least 7 days decreases sepsis and death from sepsis in patients with acute pancreatitis.]

Ranson JH, Berman RS. Long peritoneal lavage decreases pancreatic sepsis in acute pancreatitis. *Annals of Surgery* 1990: **211**: 708–16.

Ranson JH, Rifkind KM, Roses DF, Fink DS, Eng K, Spencer FC. Prognostic signs and the role of operative management in acute pancreatitis. *Surgery, Gynecology, Obstetrics* 1974: **139**: 69–81. [This study identifies factors that help to determine the identification of patients with severe pancreatitis. The authors also conclude that the avoidance of early laparotomy whenever practical, the prolongation of nasogastric suction until all evidence of pancreatic inflammation has resolved, careful monitoring of respiratory function, and early treatment of pulmonary complications, and peritoneal dialysis in patients with severe disease may improve outcome.]

Sarr MG, Sanfey H, Cameron JL. Prospective randomized trial of nasogastric suction in patients with acute pancreatitis. *Surgery* 1986; **100**: 500–4. [This study concludes that the routine use of nasogastric suctioning results in no improvement in patients with mild to moderate pancreatitis.]

Sax HC *et al.* Early total parenteral nutrition in acute pancreatitis: lack of beneficial effects. *American Journal of Surgery* 1987; **153**: 117–24. [The routine use of TPN in patients with acute pancreatitis does not improve outcome and the incidence of catheter-related sepsis is higher in these patients.]

Teerenhovi O, Nordback I, Eskola J. High volume lesser sac lavage in acute necrotizing pancreatitis. *British Journal of Surgery* 1989; **76**: 370–3. [Lesser sac lavage does not improve outcome in patients with necrotizing pancreatitis according to this study.]

Warshaw AL. Pancreatic abscess. *New England Journal of Medicine* 1972; **287**: 1234–6. [Describes the treatment of pancreatic abscesses.]

Weaver DW, Bouwman DL, Walt, AJ, Clink D, Resto A, Stephany J. A correlation between clinical pancreatitis and isoenzyme patterns of amylase. *Surgery* 1982; **92**: 576–80. [The isoamylase levels were determined in patients with pancreatitis and the authors conclude that this test is inexpensive and useful in determining the etiology of hyperamylasemia.]

Whitcomb DC *et al.* Hereditary pancreatitis is caused by a mutation in the cationic trypsinogen gene. *Nature Genetics* 1996; **14**: 141–5. [A mutation of the trypsin molecule is described in patients with hereditary pancreatitis.]

Widdison AL, Karanjia ND, Alvarez C, Reber HA. Sources of pancreatic pathogens in acute necrotizing pancreatitis. *Gastroenterology* 1991; **100**: A304. [The authors conclude that infection of the pancreas may originate from many different sources including the colon, gallbladder, and renal pelvis.]

33.2 Chronic pancreatitis

C. W. Imrie and S. J. Govil

Introduction
Incidence and prevalence
Nomenclature
Definition
Epidemiology, pathology, subtypes
Relation between chronic pancreatitis and pancreatic cancer
Pathogenesis of alcohol-induced chronic pancreatitis
Chronic obstructive pancreatitis
Pancreas divisum
Clinical features and complications
Pain
Obstruction of adjacent organs
Splenic venous thrombosis
Exocrine insufficiency
Endocrine insufficiency
Ascites or pleural effusion
Pancreatic pseudocyst
Haemorrhage
Miscellaneous features
Associated disease
Diagnosis
Endoscopic ultrasonography
ERCP
Tests of pancreatic function
Treatment
Medical management
Extracorporeal shockwave lithotripsy (ESWL)
Endoscopic approaches
Surgical approaches
Pseudocysts
Further reading

Introduction

For surgeons, three major aspects of chronic pancreatitis are of importance. The first is the management of pain, which is the dominant symptom in the majority of patients. The second is the more recent knowledge that chronic pancreatitis is a precursor to carcinoma in a proportion of patients, and the third that the differential diagnosis between chronic pancreatitis and carcinoma may be difficult.

Many specialist groups of clinicians have acquired considerable knowledge from the management of large numbers of patients and, for this reason, specialist management is to be commended. A major drawback is the lack of randomized, controlled studies of different management approaches. Claims have been made that medical therapy with antioxidant agents represents a major advance, while endoscopically placed stents, with or without additional extracorporeal shockwave lithotripsy and minimally invasive thoracoscopic splanchnicectomy, represent alternatives to direct surgical approaches. The need for critical assessment of management algorithms and well-conducted clinical trials of differing management approaches is great indeed. There are special problems in conducting clinical trials because alcohol abuse is the major cause of chronic pancreatitis (70 per cent of patients), making the unreliable behaviour patterns of the patients an additional consideration.

Incidence and prevalence

While chronic pancreatitis is said to be uncommon, with a prevalence at autopsy of 0.03 to 0.4 per cent, it is now being diagnosed much more frequently in life on the evidence of endoscopic retrograde (**ERCP**) and magnetic resonance cholangiopancreatography (**MRCP**).

Annual incidence figures of 8.2 cases per 100 000 and a prevalence of 27.4 per 100 000 were found in the prospective Copenhagen Pancreatitis Study. Data collection is, however, unsatisfactory and very few incidence studies have been performed.

Nomenclature

In 1983 and 1984, conferences in Cambridge, England and Marseilles, respectively, resulted in a simplification of the earlier classification of chronic pancreatitis to, in the 1983 meeting, a initial single category; in the Marseilles meeting a subgroup of chronic obstructive pancreatitis was defined.

Definition

Chronic pancreatitis is a continuing inflammatory disease of the pancreas that typically presents with upper abdominal and back pain. In 10 to 15 per cent of patients, pain may be absent. Only 15 to 20 per cent have exocrine insufficiency and a similar proportion are diabetic. The only sign of an inflammatory process may be fibrosis, indicative of earlier inflammation.

Epidemiology, pathology, subtypes

Typically this disease occurs in younger men with a very heavy alcohol intake over a period of approx. 8 to 12 years. Alcohol abuse (a daily intake of at least 80 g) is a feature of approx. 70 per cent of our patients, with most of the other 30 per cent coming under the unsatisfactory label of idiopathic chronic pancreatitis. However, within the 30 per cent is an increasing number of patients with hereditary pancreatitis associated with one of two genetic defects in the trypsinogen molecule that results in its intracellular activation rather than activation in the duodenum. International studies in both the United States of America and Europe are collecting data that are beginning to indicate that the proportion of patients with a hereditary pancreatitis is slowly increasing. Such patients often begin their clinical course with recurrent attacks of acute pancreatitis, progressing to a chronic form with features that are not different from those of other patients. One of the trypsinogen abnormalities does result in a younger age group being affected and both are associated with a higher incidence of pancreatic carcinoma.

Relation between chronic pancreatitis and pancreatic cancer

An important collaborative international study of 2015 patients published in 1993 identified 56 pancreatic carcinomas over a mean follow-up period of 7.4 years. Lowenfels and colleagues recruited patients from several countries and the expected number of carcinomas calculated from country-specific incidence data and adjusted for age and sex was 2.13; this yielded a standardized incidence ratio (ratio of observed:expected cases) of 26.3. The conclusion was that the cumulative risk of patients with chronic pancreatitis developing carcinoma increased steadily with time; at 10 years it was 1.8 per cent and at 20 years 4 per cent. This observation was found to be independent of nationality, sex, or the form of chronic pancreatitis. While 70 per cent of patients with chronic pancreatitis have alcohol as the aetiological factor and many of these people are also cigarette smokers, the data from the Lowenfels' study would suggest a direct link with the disease process, while not totally excluding alcohol abuse as being an independent factor of some importance.

Pathogenesis of alcohol-induced chronic pancreatitis

Most patients who develop alcohol-induced acute pancreatitis have been drinking heavily for a considerable number of months (usually 3 to 7 years), although occasional exceptions have been reported. Those who continue to drink then usually progress to chronic pancreatitis, with progressively increasing fibrosis of the acinar tissue and the relative sparing of the islets until a late stage. The viscosity of pancreatic secretion increases, with protein precipitates developing; in addition, the

fibrotic process tends to form strictures of the main and major ducts.

Stone formation is consistently of the same material, with stones in India, Europe, and the United States having a nucleus of protein in which nickel and iron can be found; around this, layers of calcium carbonate develop, which may form spicules. The calcium carbonate usually forms at least 95 per cent of the stone material; a typical example of the macroscopic appearance is shown in Fig. 1.

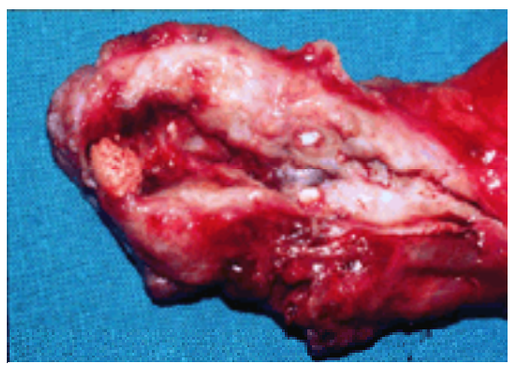


Fig. 1. A distal pancreatectomy resection specimen from a 23-year-old woman containing pancreatic-duct stones.

The exact mechanism of alcohol-induced pancreatitis is still poorly understood, but there is probably a genetic predisposition to determine whether the pancreas, liver, heart or brain are primarily affected in those who drink heavily for long periods of time. This is borne out by the finding of normal liver biopsies in patients coming to surgical therapy for chronic pancreatitis; the lead time before other organs become involved by the adverse affects of alcohol to a clinically perceptible extent may be as long as 15 to 20 years. The role of antioxidants in the protective mechanisms is probably very important and cigarette smoking has an additional adverse affect in those who abuse alcohol. In addition, those who smoke 20 cigarettes per day, irrespective of alcohol intake, have a threefold increase in the risk for developing pancreatic cancer.

The chronic pancreatitis that develops in children and young adults in India and other areas nearer the equator has been labelled tropical chronic pancreatitis; various postulates that single dietary factors are responsible for its development have failed to stand the test of careful audit and scrutiny in controlled studies. Claims about relative deficiencies of antioxidants in the diet of these patients are reasonably strong.

Chronic obstructive pancreatitis

This subgroup of patients with chronic pancreatitis is typified by stricture of a major or the main duct, with the obstruction causing progressive chronic pancreatitis on the proximal side of the gland. This change may be restricted to the tail and result in local pseudocyst formation or a fistula with ascites, as depicted in Fig. 2. A stricture occurring nearer the ampulla can result in two-thirds of the gland being affected, as shown in Fig. 3. The inception of this form of chronic pancreatitis may be a single attack of acute pancreatitis in which the main duct is seriously damaged; a pancreatic pseudocyst may have been documented during the postacute phase of that attack. Blunt abdominal trauma damaging the main duct by crushing the pancreas against the vertebral column and the presence of a slow-growing tumour can produce a similar picture.

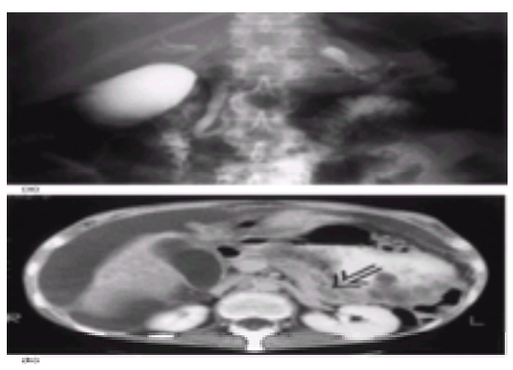


Fig. 2. (a) An ERCPgram showing the leakage point of a pancreatic fistula from the distal disrupted pancreatic-duct system (arrow); (b) a computed tomogram of the same patient shows the rupture area in the distal pancreatic duct (arrow) and ascitic fluid, particularly around the liver.

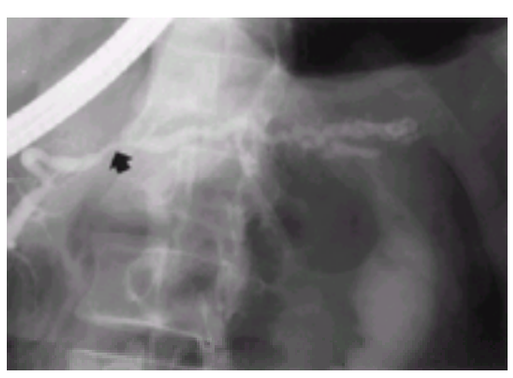


Fig. 3. An example of a focal area of compression of the main pancreatic duct with signs of distal obstruction—chronic obstructive pancreatitis.

Pancreas divisum

In this condition, which has been variously reported to occur in between 1 to 7 per cent of the population, with the lowest incidence found in Japan and the highest in Belgium, there is non- or malunion of the main duct (Wirsung) and the accessory duct of Santorini. The condition occurs in so many people that it alone does not cause chronic pancreatitis, but stricture formation seems more common at the orifice of the accessory duct in these patients (Fig. 4). A segmental dorsal chronic pancreatitis may therefore develop; treatments ranging from sphincteroplasty of the accessory duct through endoscopic stenting to resection of the dorsal pancreas have all been advocated. Controlled studies are not available.

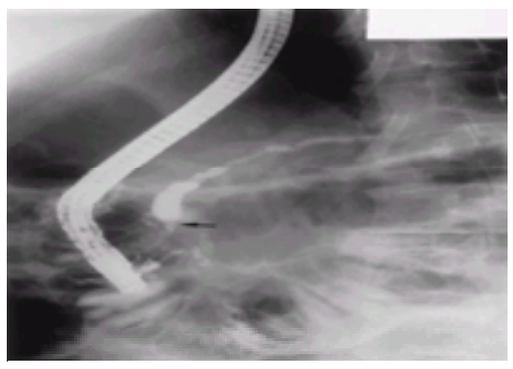


Fig. 4. An example of a stricture close to the duodenal orifice of the accessory duct of Santorini in a patient with pancreas divisum.

Clinical features and complications

Pain

Although 5 to15 per cent of patients with proven chronic pancreatitis suffer no pain or only occasional discomfort; the remainder have their life dominated by upper abdominal and back pain, which is frequently exacerbated by stimulation of pancreatic secretion by either food or alcohol. There is often a characteristic 30- to 40-min delay after such stimulation before the pain reaches a peak. The severity of pain is such that the patient is compelled to adopt certain physical positions that minimize the problem. Leaning forward in a sitting position or bending in a fetal position are particularly helpful, and ever-increasing amounts of analgesics are prescribed, which have their own side-effects and potentially exacerbate the problem of pain by causing constipation. While the opiate side-effect of slowing gut transit time can be very helpful in the 15 to 20 per cent with steatorrhoea, it may be a major undesirable side-effect in the 80 per cent not affected in this way. For those who have a heavy alcohol intake already, the temptation to drink more heavily in order to lessen the pain results in a vicious circle of recurrent episodes of pain, analgesic intake, and superadded alcohol. Episodes of pain may last from one to many hours; measurements of blood amylase and lipase indicate only modest or moderate elevations. Quality of life is adversely affected to a great degree.

Measurements show raised pressure in the pancreatic tissue and ducts; where the duct system is grossly dilated, decompression by endoscopic stenting or surgical drainage with stone or stricture removal are very effective. However, the relation between the radiological picture and the pain problem is far from clear and difficult to understand in many patients.

The histological and histochemical features of the nerves serving the pancreas change, with swelling of the nerves and disruption of the Schwann cells. It is postulated that this leaves the nerves themselves exposed to various transmitters released from inflammatory cells and there is evidence of an increase of substance P, a neurotransmitter associated with pain transmission, in chronic pancreatitis. No studies have yet compared patients who have chronic pancreatitis and pain and those without pain because there is no clinical indication for pancreatic biopsy in the second group and no safe methods for obtaining such biopsies yet exist.

Obstruction of adjacent organs

The lower common bile duct running through the head of pancreas is particularly vulnerable to compression (Fig. 5), with the resultant increase in back pressure in the biliary tree being a possible cause of pain. Most patients with this relatively common feature have a considerably elevated alkaline phosphatase while the bilirubin may be normal or marginally raised. Discomfort and pain associated with this are not common. Temporary stenting of the lower common bile duct with an endoscopically introduced plastic stent is good therapy, but the stents need to be removed or changed at 3 to 6 months.

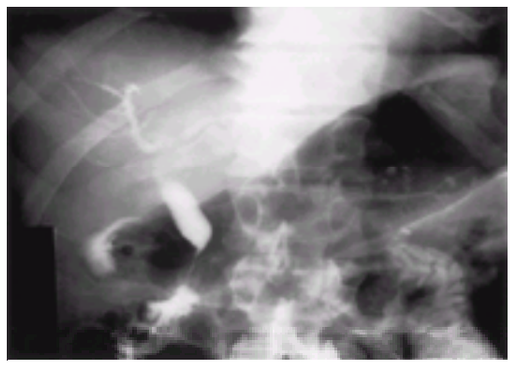


Fig. 5. An example of compression of the lower common bile duct in a patient with chronic pancreatitis (note also the distal pancreatic-duct system with ectatic side branches viewed through the gas-filled stomach).

Compression of the second part of duodenum may be so great that a clinical picture indistinguishable from pyloric stenosis develops. Gastrojejunostomy may be the only way to relieve this problem, which is less frequent than biliary compression.

Colonic obstruction can occur in association with peripancreatic inflammation; this may be very difficult to differentiate radiologically from the picture of a carcinoma in the transverse colon and the stricture may be too tight to allow full colonoscopy. This is a much less common complication and one that we have only witnessed twice, once in the right and once in the left transverse colon.

Splenic venous thrombosis

This is an important complication of chronic pancreatitis that results in segmental portal hypertension with splenomegaly, and with occasional varices developing in the lower oesophagus and proximal stomach. When a severe problem exists, splenectomy with distal pancreatectomy is usually successful but can be a time-consuming and difficult procedure, owing to the peripancreatic inflammatory process.

Exocrine insufficiency

The pancreatic reserve is such that steatorrhoea only develops in a small proportion of patients. Stools tend to be bulky, with a particularly offensive smell, pale in colour, and difficult to flush in a standard toilet. Increased fat in the diet may exacerbate this problem greatly. Pancreatic supplementation with the new microsphere pancreatin capsules is usually effective. It has been shown that it is best to encourage these patients to control the number of capsules taken themselves rather than have a rigid limitation to around six capsules per day. Proton-pump inhibitors or H_2-receptor antagonists protect the pancreatin capsules from acid destruction.

Endocrine insufficiency

The islets of Langerhans tend to be spared until the very late stages of the disease, but diabetes is occasionally an early part of the presentation. The endoscopists and surgeons who carry out life-changing procedures often have to face an ensuing weight gain in their patients, and it is not unusual for diabetes to present 6 or 12 months after successful interventional surgery because the total islet output of insulin was sufficient for a 50-kg patient but soon becomes insufficient with quality weight gain. It is important to warn patients of this potential development.

Ascites or pleural effusion

Disruption of the main pancreatic duct can occur in chronic pancreatitis and may present with either ascites (Fig. 2) or pleural effusion. Quite frequently the diagnosis is missed, and even when samples of fluid have been aspirated an analysis of amylase or lipase content may be overlooked in preference to bacteriological and cytological studies. Typically the amylase content is very high and the albumin content is also higher than one normally finds in ascitic fluid associated with hepatic disease. Endoscopic stenting of the disrupted duct or surgery may be necessary.

Pancreatic pseudocyst

This may be associated with disruption of a main duct and is a variant of the architectural disruption found in pancreatic ascites; or it may be impossible to demonstrate any link with the main duct or a major duct. Special care must be taken in patients with chronic pancreatitis to verify that a cystic tumour is not present as misdiagnosis is possible.

Haemorrhage

Bleeding within a pancreatic pseudocyst or simply from erosion of a major vessel by the disease process may present with haemorrhage along the pancreatic duct, either acutely or chronically (haemosuccus pancreaticus). Even more rarely, bleeding may occur into the retro- or intraperitoneal spaces. Therapeutic angiographic control is to be preferred to surgery. Bleeding from oesophagitis, or from varices associated with splenic venous thrombosis, gastritis, peptic ulceration, or even duodenitis must also be considered in these patients.

Miscellaneous features

Nausea, vomiting, and upper abdominal discomfort are common. Weight loss (partly from a fear of eating because of pain induction) and diarrhoea is a frequent phenomenon, combined with malnutrition in which vitamin deficiencies are to be expected.

Associated disease

Occasionally liver cirrhosis, cardiomyopathy, or brain damage from alcohol abuse may accompany the chronic pancreatitis but this is unusual in the early years of the disease. The link with pancreatic cancer is discussed above.

Diagnosis

In the advanced form of the disease the diagnosis can be strongly suggested by the history and confirmed by the presence of calcification on straight abdominal radiographs, often accompanied by malabsorption and diabetes. Earlier forms of the disease can present particular difficulties but the recent emergence of MRCP presents a major advance in diagnosis of this condition. The quality of imaging is usually not quite as precise as that of ERCP, but a negative MRCP is certainly very valuable and positive identification avoids both the inconvenience and potential danger of ERCP (Fig. 6 and Fig. 7). To most clinicians involved in the diagnosis and management of chronic pancreatitis, new algorithms have been devised that do, of course, depend on the resources available in a particular institution but increasingly would follow a line from straight radiography, to standard ultrasonography to MRCP.

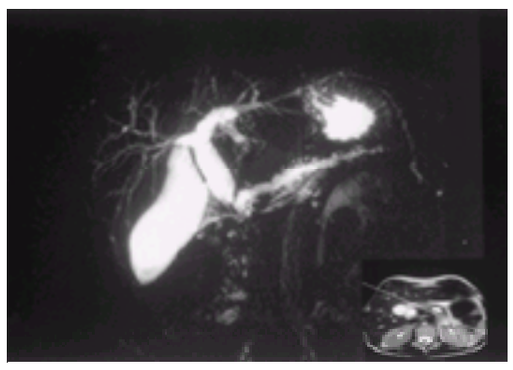


Fig. 6. An example of a modern MRCP scan (with no oral or intravenous contrast) clearly demonstrating chronic pancreatitis with compression of the lower common bile duct and a dilated gallbladder.

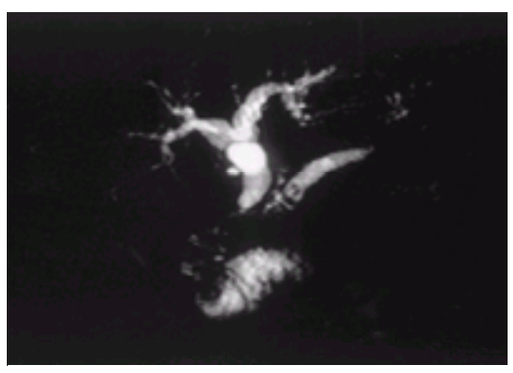


Fig. 7. A similar MRCP scan showing stones in a dilated pancreatic duct with compression of the lower common bile duct and dilatation of the biliary tree.

Computed tomography can be particularly valuable for early detection of calcification. It is also useful for delineating duct disruption and ascites (Fig. 2(b)); when combined with angiographic techniques it can outline the effects of splenic venous thrombosis or occlusion.

Endoscopic ultrasonography

This is slowly becoming increasingly available and, while operator dependent, has the potential to provide very useful information on cystic abnormalities, possible tumours, disruption of major ducts, and varices associated with segmental portal hypertension.

ERCP

This has been the mainstay of the diagnosis of ductal damage as well as early intraduct calcification. It has the major advantage of simultaneous brush biopsy to enable cytological assessment of any stricture, and the therapeutic potential of stent placement is also most important, even though its value is still debated. Decompression by biliary stenting is generally accepted to be very effective but there is controversy about the usefulness of draining the pancreatic duct with plastic stents, as these tend to block completely by 6 to 12 months and may retard the flow of pancreatic juice considerably sooner.

Even more contentious is the advocacy by a few doctors of the value of the endoscopic placement of expanding metal stents in strictures of the main pancreatic duct.

Tests of pancreatic function

These have been increasingly disregarded by clinicians over the last 10 to 15 years, partly due to their relative inaccuracy and also to the inconvenience to patients. Structural investigations such as those outlined above have dominated the diagnostic approach. Outside major centres there is an increasingly limited utilization of both

the indirect and direct tests of pancreatic function.

Indirect tests

These less interventional investigations, which involve no enteral intubation, are preferred by patients. In the Bentiromide test the patient ingests a synthetic tripeptide (*N*-benzoyl-L-tyrosyl-*p*-amino-benzoic acid); this is cleaved by chymotrypsin to release *p*-aminobenzoic acid, which is absorbed from the small intestine, partially conjugated by the liver, and then excreted in the urine. The same principle is used in the pancreolauryl test in which the esther fluorescein dilaurate is hydrolysed to release fluorescein, which is absorbed and then excreted in the urine to be detected by spectrophotometry.

More recently, measurements of both faecal elastase and chymotrypsin to test for each enzyme's secretion following ingestion of their respective oral substrates have gained ground and claims have been advanced that these are the most practical tests in this difficult area.

Secretin–cholecystokinin test

This remains the ‘gold standard' for pancreatic function but requires the presence of a double-lumen tube to aspirate both the stomach and the duodenal contents to analyse pancreatic secretions after stimulation by exogenous hormones. The marker [^{14}C]polyethylene glycol is used to ensure that at least 85 per cent of the duodenal juice is recovered. The outputs of bicarbonate, trypsin, and amylase as well as lipase can be measured. The test has a sensitivity of up to 90 per cent but specificity is in the region of 80 per cent. As with most tests it is most accurate in those with the most gross pancreatitis. Indeed one of the major deterrents to the use of tests of pancreatic function is their inability to detect minor degrees of exocrine insufficiency.

Treatment

The major indications for treatment are:

1. intractable pain;
2. fear of carcinoma;
3. the development of structural complications

In the 15 per cent of patients who are pain free, the need to consider therapeutic intervention is much less pressing. Initial drug treatment for pain, moving on to endoscopic and allied measures before surgery, is the normal algorithm. However, for chronic pancreatitis predominantly in the head of the gland there is evidence that surgical decompression will not only help the pain but also slow down or stop the deterioration in both exocrine and endocrine function.

Medical management

There are strong claims that antioxidant therapy minimizes or even removes pain and that pancreatic extracts can also behave in a similar fashion. Convincing randomized, controlled clinical trials are totally absent. In the present state of knowledge these approaches can certainly be tried as a first line of management, with a modest expectation of success. One problem is that the most used product, containing selenium, methionine, and vitamins A, C, and E (Antox), is registered as a food additive and is difficult to prescribe and quite expensive to purchase. The recommended dose is four to six capsules per day, more convenient than the 14 tablets per day required before Antox became available.

Analgesics, moving initially through non-steroidal anti-inflammatory drugs to opiates, are the usual recommendation. The danger of opiates causing major gut problems and exacerbating the pain cycle has already been noted. The pain seems to be very severe in most patients and it is understandable that doctors end up prescribing major opiates in considerable doses. Apart from the usual prescription of oral dihydrocodeine, pethidine, or morphine preparations, the use of skin patches of fentanyl and of transcutaneous electrical nerve stimulation (TENS) machines should be considered.

The treatment of exocrine insufficiency may require the prescription of many more than the usual six capsules per day of pancreatin to reverse steatorrhoea, but dietary restriction of fat and the provision of drugs that minimize the acid-destructive effect on enzyme supplements such as Creon by the stomach are also very important. In practice it can prove very difficult to eliminate steatorrhoea and a considerable proportion of patients with this complication of chronic pancreatitis only partially respond to replacement with pancreatic extract containing either 10 000 or 25 000 units of lipase per capsule. The major advantage to the patient of the high-dose lipase capsules is the smaller number that normally have to be taken per day; this has proved particularly beneficial in children with cystic fibrosis, and to a lesser extent in those with chronic pancreatitis.

In our clinical experience only a small proportion of patients have a beneficial effect on pain from the prescription of pancreatic extract and similarly the use of antioxidant preparations containing selenium, methionine and the vitamins A, C and E provides only a limited reduction in pain. However, the patients who do benefit are very consistent in their claims and this justifies a trial of such measures in all patients as we have no means of identifying those likely to benefit.

Extracorporeal shockwave lithotripsy (ESWL)

In patients with stones in the pancreatic ducts the use of a two-plane ESWL has proved very successful in expert hands. Both the Brussels group (Dumonceau *et al.*) and certain Japanese workers have shown that stones can be readily fragmented by a short series of exposures. While the European team reckon that endoscopic removal of the fragments is the norm, the Japanese have questioned whether even this is necessary, as 18 of 22 patients passed these stone granules spontaneously. However, at present only a few centres in the world have availed themselves of either the machines or the expertise to assess this approach.

Endoscopic approaches

Cremer, Deviere and their colleagues are particularly strong in their advocacy of an endoscopic approach to many of the problems of chronic pancreatitis. They claim good pain relief by first using endoscopic sphincterotomy of the main pancreatic duct, followed by stone extraction after ESWL and the placement of a plastic stent, where appropriate, through areas of stricture formation. Where complete clearance of the main pancreatic duct was achieved the results were particularly encouraging, but they do emphasize that it is advisable to use 10 FG stents and to be aware that they tend to block at between 6 and 12 months. Their experience is that 50 to 70 per cent of patients with chronic pancreatitis require prolonged stenting, but they do not have a controlled study and a long follow-up of results is not yet available.

A comparison of endoscopic with surgical therapy by independent assessors is essential before claims of the preferred treatment can be substantiated. The risks of endoscopic therapy are less than those of surgery, but the recurrent procedures and the total risk of complications, as well as the total expense, require assessment. While possible morbidity is acknowledged in the endoscopic approach, fatality is rare. An important factor in any attempt to compare outcomes is the relatively small number of patients who have yet been subjected to endoscopic therapy. Ultimately a combination of treatments may prove best in the management of any single group of patients. The major existing studies are outlined in Table 1, Table 2, Table 3, Table 4 and Table 5.



Table 1 Role of biliary stents in chronic pancreatitis

Study[a]	No. of patients	No. in whom stone complete ...	Mean ... (months)	Follow-up	Pain relief	Complications
Sherman et al. (1991)	32	23			67% complete	12.5%
Kozarek et al. (1992)	12	11	17		9/11 complete	
Dumas et al. (1994)	50	39	15		39/50 complete	Pancreatitis in 4 (8%), ... in 2 (5%)
Hennessy et al. (1988)	15/16 with pain	6	14		7/16 complete; 5/16 decreased pain	19%

[a] Sources: Sherman S et al. Gastrointestinal Endoscopy 1991; 37(5): 533–7; Kozarek RA et al. American Journal of Gastroenterology 1992; 87(8): 864–9; Dumas R et al. Gastrointestinal Endoscopy 1994; 40(6): 708–14; Hennessy WB et al. Gastrointestinal Endoscopy 1988; 34(6): 543–8.

Table 2 Efficacy of endoscopic treatment of pancreatic stones in patients with chronic pancreatitis

Study[a]	No. of patients	No. in whom stone fragmentation achieved	No. in whom stone clearance achieved	Follow-up (months)	Pain relief
... Thune et al. (1993)	8	6	5 total	17	5 complete
Sauerbruch et al. (1992)	14	11	10 total 1 partial	24	12 complete 1 improved
Delhaye et al. (1992)	123	122	72 total		48/68 complete 18/68 improved
... Chan et al. (1992); Schreiber et al. (1990); ... Cremaniga et al. (1992)	41	35	34 total 5 partial	30 12	31 complete 3/5 complete

[a] Sources: Ann Thune E et al. American Journal of Gastroenterology 1993; 88(8): 1203–6; Sauerbruch T et al. Gut 1992; 33(7): 968–72; Delhaye M et al. Gastroenterology 1992; 103(1): 189–96; Yamada M, Ohashi K et al. Endoscopy 1990; 22(1): 3–8; Chan R et al. American Journal of Gastroenterology 1990; 85(7); 1288–94; Schreiber F et al. Journal of Gastroenterology and Hepatology 1990; 5(5): 567–81; Cremaniga G et al. Gastrointestinal Endoscopy 1992; 38(2): 159–64.

Table 3 Efficacy of ESWL in fragmentation and endoscopic clearance of pancreatic duct stones

Study[a]	No. of patients	Early pain relief	No. in whom stone cleared	Pancreatitis relieved or improved	Follow-up after treatment	Complications
Ramesh et al. (1995)	15	14 (74%)	6 (40%)	9/15 (60%)	18 mo	10/15 pancreatitis ... Abscess, bleeding, ... exploratory laparotomy
Sauer et al. (1993)	19	4 (21%)	2 (10%)	12 (63%) 3 (16%) improved	20.5 months	Early complications in 7 patients Later complications in 2 patients after median follow-up of 20 months
Rösch et al. (1991)	21	17 (74%)	13 (62%)	13 (62%)	1 mo	Severe ... dropped ... after 1st) after 2 months and ... remained free ... for maximum of 4 months.
Lennon et al. (1991)	16	7 (70%)			21 months	75% minor complications

[a] Sources: Ramesh J et al. Endoscopy 1995; 27(9): 638–44; Sauer BG et al. Gastrointestinal Endoscopy 1993; 39(5): 665–9; Rösch T et al. Gastrointestinal Radiology 1991; 176(1): 27–31; Lennon JM et al. Endoscopy 1991; 23(1): 13–16.

Table 4 Efficacy of pancreatic stents for patients with chronic pancreatitis and a dominant stricture of the main pancreatic duct

Study[a]	No. of patients	Patients with duct changes on stent removal	Nature of changes	Resolution of changes
Smith et al. (1994)	63	49 (80%)	Duct irregularity 49%, ductal narrowing 35.5%, side-branch changes 15.5%	Complete 64%
				Partial 23%
				None 7%
Sherman et al. (1994)	16 (22.2%)	4 (56%)		1 of 4 patients assessed at 18 months: pancreatitis
	75 (22.5%)	17 (68%)		
Sullivan et al. (1992)	10 transient ductectasia	10 (100%)	Normal ducts: narrowing in 3 patients	Partial in 3 patients
	19 ductectasia	5 (26%)	Abnormal ducts: improvement in 7 patients; deterioration in 4 patients	
Kozarek (1990)	9 dilated ducts	9 dilated ducts	Pseudocysts, duct dilatation, stenosis, side-branch ectasia	
	29 pancreatitis	4 pancreatitis		
Dumas et al. (1994)	50	39 (78%)		

[a] Sources: Smith MT et al. Gastrointestinal Endoscopy 1994; 40(3): 328–33; Sherman S et al. Gastrointestinal Endoscopy 1994; 40(6): 276–82; Sullivan CB et al. American Journal of Roentgenology 1992; 159(4): 753–5; Kozarek RA. Gastrointestinal Endoscopy 1990; 36(5): 99–105; Dumas R et al. Gastrointestinal Endoscopy 1994; 40(6): 708–14.

Table 5 Alterations in pancreatic-duct morphology induced by stents

Surgical approaches

Interruption of nerve pathways

While transcutaneous coeliac blockade has proved a considerable benefit in the therapy of pancreatic cancer the short-term advantage has resulted in few advocates of its use in the management of chronic pancreatitis. Direct interruption of the nerves at open operation is very difficult: the senior author of this chapter has striven to perform this on eight occasions with only four successes because removal of the coeliac ganglia is fraught with problems of identification. Stripping of the adventitia from the coeliac axis and superior mesenteric artery is usually required, which is not easy in patients who have often had multiple prior procedures.

A newer and much less hazardous approach is that of minimally invasive thoracoscopic splanchnicectomy, which we have found can be performed in a 24-h admission and has consistently good initial results, with eradication of the need for opiates in most patients and a sustained response at 1 year in over 60 per cent.

Andren-Sandberg and colleagues in Sweden have very similar results and the procedure is remarkably straightforward, being carried out with the patient in a prone position. At first a double-lumen endotracheal tube was used, but we have found this to be unnecessary and instead use small chest drains (18 FG), which can be removed before the patient leaves the recovery rooms. The identification of the splanchnic nerves is usually not a problem unless major lung disease or previous surgery or extreme kyphosis cause problems. As with most other approaches, this has not been subjected to a randomized study but our initial experience in over 20 patients with chronic pancreatitis is certainly encouraging, with over 60 per cent having sustained benefit at a year.

Bypass and drainage

In patients who have major duodenal compression and/or obstruction of the lower common bile duct, surgical bypass, laparoscopic or open, can deal with this problem readily by gastrojejunostomy and cholecysto- or choledochojejunostomy. The much less common problem of colonic obstruction is best dealt with by resection and end-to-end anastomosis.

Pain in the presence of a tight stricture and distal dilated duct is best dealt with by drainage of the obstructed duct system by reroofing the duct (preferably using a harmonic/diathermy scalpel) and then anastomosing the Roux loop longitudinally along the anterior surface of the pancreas. Where the obstruction is in the neck, or more distally, and the head of the gland is not involved, then such measures in themselves usually prove beneficial (Fig. 8). However, in the majority of patients the head of the gland is also involved and often enlarged. In such patients, simple longitudinal pancreatojejunostomy is much less likely to be successful.



Fig. 8. A late check of the integrity of a pancreatojeunostomy more than 10 years after the operation; the ERCP shows good outlining of the Roux loop.

Resections

Duodenum-preserving, head-of-pancreas resections

Increasingly it is recognized that a ‘compartment syndrome' with increased tissue pressure and possible chronic ischaemia, usually in the head of pancreas, contributes to the pain and malfunction of the gland. Based on this reasoning, Beger from Ulm devised an operation with subtotal resection of the head of pancreas, duodenal preservation, and drainage of the body and tail into a Roux loop. The anterior 80 per cent of the head of the gland is resected and sometimes the lower common bile duct is drained into the Roux loop, as well as the more distal pancreatic duct (Fig. 9). A randomized study showed this to be superior to a pylorus-preserving pancreatoduodenectomy (Table 6) and the pioneering Ulm team have claimed few complications and major benefit in over 300 patients. The preferred suture for anastomosis of the Roux loop to the pancreas is 3/0 Prolene or a similar material, with either interrupted or continuous technique. Klempa also compared the Beger operation with the classical Whipple procedure in a randomized study and the results favoured the duodenum-preserving operation.

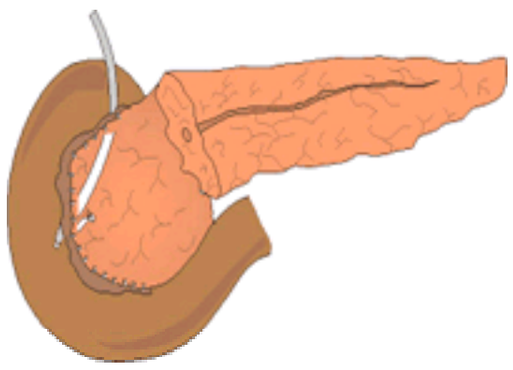


Fig. 9. Diagram of the area of resection of most of the head of the pancreas with duodenal conservation in the Beger operation.

	Beger (n = 20)	Whipple (n = 20)
Mortality	0	0
Morbidity	15%	20%
Mean hospital stay	13 days	14 days
Mean weight gain at 6 months*	4.1 kg	1.9 kg
Proportion pain free at 6 months*	75%	40%
Proportion employed at 6 months	80%	67%
Proportion requiring readmission	13%	27%
Glucose tolerance and insulin secretion*	Unchanged	Reduced
Pancreaticobiliary l test	Unchanged	Unchanged

*Statistically significant, $p < 0.05$.
Source: Buchler *et al.* (1995).

Table 6 Results of prospective, randomized trial between Beger and pylorus-preserving Whipple procedures

Following Beger's ideas, Frey of Sacramento, California, described a much less radical partial resection of the head of the gland in which he removed between 5 to 6 g of tissue to decompress and enhance drainage of the major ducts (Fig. 10). This lesser operation in terms of tissue resection has been compared by Izbicki and colleagues with the Beger procedure; few differences were found in outcome in this randomized study (Table 7). Izbicki has gone on to a further comparison of the Frey operation with pylorus-preserving pancreatoduodenectomy and again found the lesser procedure to have the more favourable results, though pain relief was good with both (Table 8). These randomized studies do represent the way ahead for our understanding of the optimal therapeutic approaches. Unless there is gross and extensive fibrosis in the head of the gland, the advantage of the Frey operation in being less radical and yet having a good prospect of improvement for the patient makes it an attractive first step in therapy unless there is greater local experience of the more radical Beger procedure. Such an approach may mean that the minority of patients require subsequent consideration for a greater amount of resection of pancreatic tissue at a further procedure. Neither the Beger nor the Frey procedure are particularly appropriate where the duct system is of normal calibre and up until recently no drainage procedure has been considered applicable to this type of patient. However, Izbicki has shown that there is potential in creating a V-shaped channel down to the pancreatic duct in such patients and then applying a Roux loop over the whole of the anterior surface of the gland.

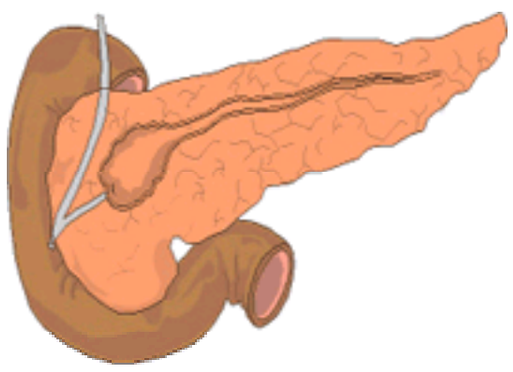


Fig. 10. Diagram of the limited resection (usually only 5–6 g of tissue from the head of the gland) in the Frey operation.

	Beger (n_1=λ_{10}=λ_{11}=λ_{12}=λ_{13}=20)	Frey (n_1=λ_{10}=λ_{11}=λ_{12}=λ_{13}=22)
Mortality	0	0
Morbidity*	4 (20%)	2 (9%)
Global quality of life score:		
Preoperative	28.6	28.6
Postoperative	85.7	85.7
Pain score:		
Preoperative	62.25	68.5
Postoperative	5	4
Abnormal faecal chymotrypsin:		
Preoperative	40%	48%
Postoperative	50%	50%
Pathological serum insulin:		
Preoperative	55%	45%
Postoperative	65%	50%

*Statistically significant, $p < 0.05$.
Source: Izbicki *et al.* (1995).

Table 7 Results of prospective, randomized trial between Beger and Frey procedures (mean follow-up 1.5 years)

	Frey ($n = 31$)	PPPD ($n = 30$)
Mortality	1 (3.2%)	0
Morbidity	19.4%	53.3%
Global quality of life score:		
Preoperative	28.6	28.6
Postoperative	85.7	71.4
Pain score:		
Preoperative	75	75
Postoperative	0	0

Source: Izbicki *et al.* (1998).

Table 8 Results of prospective, randomized trial comparing Frey operation and pylorus-preserving pancreatoduodenectomy (PPPD) (mean follow-up 2 years)

Pylorus-preserving pancreaticoduodenectomy can be a useful operation in patients who have the disease predominantly confined to the head of the gland, but it has been shown in the studies referred to above to be inferior to both the Beger and Frey approach (Table 6 and Table 8). Nevertheless it has some applicability in management where surgeons do not have sufficient experience of duodenum-preserving, head-of-pancreas resections.

Distal pancreatectomy is applicable to the subgroup of patients with chronic obstructive pancreatitis and is particularly useful in their management. A smaller number of patients who have the more traditional type of chronic pancreatitis may be treated with this approach, which classically involves a splenectomy. Laparoscopic distal pancreatectomy involving the use of stapling devices across the neck of the gland and at the splenic hilum, with removal of intervening tissue and dependence of the spleen on the vasa brevia, has been described; the number of patients so far treated with this approach is small.

We and others have advocated total preservation of splenic artery and vein by dissecting the diseased pancreas away from the vessels at open operation. This dissection can be very time consuming, but it is advantageous to retain the spleen and thereby minimize the late risk of overwhelming postsplenectomy infection. Splenectomy is an extra hazard to those who are diabetic.

Total pancreatectomy

This is the final surgical resort and can result in a life free from pain, but up to a third of patients may still have pain similar to that before the operation; when this is considered (in addition to the inevitable diabetes and steatorrhoea) there are few surgeons who will embark on this approach. Total pancreatectomy often follows a series of failed earlier efforts but the brittle type of diabetes and the relative hyperglycaemia that must be maintained in the absence of glucagon do, of necessity, mean that the patient has a moderate quality of life and is at the risk of fatal hypoglycaemia; this has been found by most surgeons who have performed a reasonable number of such procedures.

The extraction of islet tissue and its reimplantation into the liver by injection along a mesenteric vein has been performed experimentally in small groups of patients subject to total pancreatectomy. The long-term results of such autotransplantation have not fully been assessed and enthusiasm for such an approach must be tempered until careful audit is available.

Pseudocysts

If the pseudocyst is distorting the stomach then either endoscopic or surgical drainage into the stomach is a good approach after a trial of subcutaneous octreotide (300–600 µg/day), which may result in spontaneous resolution. Pseudocysts not adjacent to the stomach are best drained internally into a Roux loop, but when located in the tail of the gland a resection with preservation of splenic vessels is the optimum. It bears repeating that all cystic lesions in and around the pancreas must be verified to be pseudocysts and not cystic tumours.

Most reports of endoscopic, percutaneous, or surgical therapy of pancreatic pseudocyst do not clearly differentiate between whether it is a complication of acute or of chronic pancreatitis. Many of the claims for superiority of technique and results cannot be correctly assessed. In a personal series of 200 consecutive pseudocysts, 150 complicated acute pancreatitis and only 50 were identified as a complication of chronic disease. We found a mortality of 8.5 per cent in the former group and none of the 50 with chronic pancreatitis died. Internal drainage techniques were employed in most patients.

Further reading

Ammann RW, Akovbiantz A, Largiader F, Schueler G. Course and outcome of chronic pancreatitis. Longitudinal study of a mixed medical-surgical series of 245 patients. *Gastroenterology* 1984; **86** : 820–8.

Axon AT, Classen M, Cotton PB, Cremer M, Freeny PC, Lees WR. Pancreatography in chronic pancreatitis: international definitions. *Gut* 1984; **25** : 1107–12.

Beger HG, Buchler M, Bittner R, Oettinger W, Roscher R. Duodenum-preserving resection of the head of the pancreas in severe chronic pancreatitis. Early and late results. *Annals of Surgery* 1989; **209** : 273–8.

Beger HG, Krautzberger W, Bittner R, Buchler M, Limmer J. Duodenum-preserving resection of the head of the pancreas in patients with severe chronic pancreatitis. *Surgery* 1985; **98** : 467–72.

Bockman DE, Buchler M, Malfertheiner P, Beger HG. Analysis of nerves in chronic pancreatitis. *Gastroenterology* 1988; **94** : 1459–69.

Buchler M, Friess H, Muller MM, Beger HG. Randomised trial of duodenum-preserving pancreatic head resection versus pylorus-preserving Whipple in chronic pancreatitis. *American Journal of Surgery* 1995; **169** : 65–70.

Copenhagen Pancreatitis Study. An interim report from a prospective epidemiological multicenter study. *Scandinavian Journal of Gastroenterology* 1981; **166** : 305–12.

Dumonceau JM *et al.* Endoscopic pancreatic drainage in chronic pancreatitis associated with ductal stones: long term results. *Gastrointestinal Endoscopy* 1996; **43** (6): 547–55.

Greenlee HB, Prinz RA, Aranha GV. Long-term results of side-to-side pancreatico-jejunostomy. *World Journal of Surgery* 1990; **14** : 80–4.

Izbicki JR, Bloechle C, Broeing DC, Knoefel WT, Kuechler T, Broelch CE. Extended drainage versus resection in surgery for chronic pancreatitis. A prospective randomized trial comparing the longitudinal pancreaticojejunostomy combined with local pancreatic head excision with the pylorus-preserving pancreatoduodenectomy. *Annals of Surgery* 1998; **228** (6), 771–9.

Izbicki JR, Bloechle C, Knoefel WT, Kuechler T, Binmoeller KF, Broelsch CE. Duodenum-preserving resection of the head of the pancreas in chronic pancreatitis: a prospective, randomized trial. *Annals of Surgery* 1995; **221** (4): 350–8.

Klempa I *et al.* Pancreatic function and quality of life after resection of the head of the pancreas in chronic pancreatitis. A prospective, randomized comparative study after duodenum-preserving resection of the head of the pancreas versus Whipple's operation. *Chirurgie* 1995; **66** : 350–9.

Lowenfels AB *et al.* Pancreatitis and the risk of pancreatic cancer. *New England Journal of Medicine* 1993; **328** : 1433–7.

Niederau C, Grendell JH. Diagnosis of chronic pancreatitis. *Gastroenterology* 1985; **88** : 1973–95.

Singer MV *et al.* Classification of chronic pancreatitis. In: Beger HG, Warshaw AL, Buchler M *et al.*, eds. *The pancreas*, Vol. 1, Ch. 66. Blackwell Scientific, Oxford, 1998.

Worning H *et al.* Alcoholic chronic pancreatitis. In: Beger HG, Warshaw AL, Buchler M *et al.*, eds. *The pancreas*, Vol 1, Ch. 67. Blackwell Scientific, Oxford, 1998.

33.3 Trauma to the pancreas

Alan R. Berry

Trauma to the pancreas
Indicators of pancreatic injury
Operative examination of the pancreas
Management
Patients treated non-operatively
Patients undergoing surgery
Further reading

Trauma to the pancreas

Injury to the pancreas was first formally recognized and reported in 1827. Although the problem has since become more familiar, compared with the incidence of injury to other viscera it is still uncommon and therefore is frequently overlooked. If the injury is not diagnosed initially it subsequently becomes apparent when complications occur, up to 6 weeks after the event. When the diagnosis is delayed, treatment is much more difficult, morbidity is high, and the mortality rate is about 20 per cent. A small proportion of patients, in whom the diagnosis is missed or the injury successfully treated conservatively, will develop chronic pancreatitis, secondary to stricturing of the pancreatic duct, as long as several years later (Fig. 1).



Fig. 1. Endoscopic retrograde cholangiopancreatogram showing pancreatic-duct stricture and distal chronic pancreatitis 4 years after a severe crush injury.

The trauma may be either blunt or penetrating. The penetrating type is about three times more common and is associated with a higher mortality rate. Up to 90 per cent of patients also have other serious injuries, which contribute to the high mortality. Other viscera that are particularly liable to be damaged in patients with pancreatic trauma are the liver, spleen, and small intestine.

Pancreatic trauma occurs in between 1 and 3 per cent of patients who have suffered blunt or penetrating abdominal injuries. This figure is low, but is increasing as high-speed motor vehicle accidents and civil violence become more common. It is most frequently seen in young adults, and in more men than women. The experience of any surgeon is likely to be small and successful management will therefore be based on the accrued published experience of others. The possibility of pancreatic injury should always be considered in patients who have suffered abdominal trauma. If there is sufficient concern the pancreas should be examined thoroughly, either through a laparotomy, which may be being performed for other injuries, or by careful investigation. Ductal injury in particular must be identified early if complications are to be avoided.

Indicators of pancreatic injury

Serum amylase is elevated in 50 to 90 per cent of patients with pancreatic trauma, but false-positive results are also common. Trauma is a common cause of acute pancreatitis in children, accounting for one-third of all such cases in the United States. The pancreas of a child is vulnerable and easily damaged during sledging or bicycle accidents, or by physical abuse. Children often present late, the accident having been regarded as trivial; in these patients, hyperamylasaemia is always present.

Ultrasonography is a simple method by which patients suspected of having sustained pancreatic trauma can be assessed. It may demonstrate pseudocysts but it is unreliable in detecting ductal injury. Computed tomography (**CT**), if available, is more reliable and will successfully detect pancreatic fractures and contusions, as well as post-traumatic pseudocysts. In the acute setting, CT scans performed with intravenous and oral contrast are reportedly 85 per cent sensitive for diagnosing a pancreatic injury in children. However, ductal integrity can best be assessed by endoscopic retrograde pancreatography, which is the investigation of choice in a stable patient when pancreatic injury is suspected. The classification of pancreatic injuries is shown in Table 1.

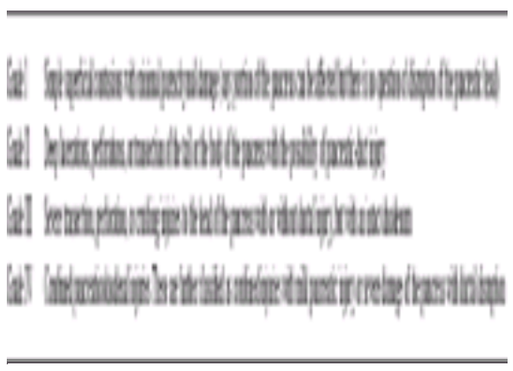


Table 1 Classification of pancreatic injuries (Lucas 1977)

The majority of surgeons first encounter a pancreatic injury while performing a laparotomy for some more obvious coexisting injury. This is the best time to detect pancreatic injury, and correct treatment can avoid future complications and possible death. Examination of the pancreas in this situation is difficult, however, particularly when other injuries seem more pressing. A planned approach is essential. Patients at risk are those who have received blunt trauma to the upper abdomen during a road-traffic accident or who have an obvious penetrating injury to the upper abdomen. After other, immediately life-threatening injuries have been dealt with, the pancreas should be examined. Clues to injury are the presence of blood in the lesser sac or, more commonly, a retroperitoneal haematoma in the upper abdomen that extends into the transverse mesocolon. There may be evidence of haematoma lateral to the second part of the duodenum. These findings demand thorough exploration and examination of the gland.

Operative examination of the pancreas

Access to the head of the gland is gained by dividing the peritoneum lateral to the second part of the duodenum and lifting the head away from the posterior abdominal wall and to the left (Kocher's manoeuvre) (Fig. 2; 1). The body and tail of the gland are best examined from within the lesser sac. Access is gained by freeing the greater omentum from its avascular attachment to the transverse colon (Fig. 2; 2). This allows access deep to the omentum and superior to the transverse mesocolon.

The tail of the gland can be approached by deflecting the spleen to the right (Fig. 2; 3). This is neither particularly easy nor quick, and in emergency situations the tail can be examined adequately without the need for this manoeuvre. Division of the ligament of Trietz and mobilization of the duodenojejunal flexure allows examination of the third and fourth parts of the duodenum (Fig. 2; 4). This may be made easier if the hepatic flexure of the colon is mobilized to the left.

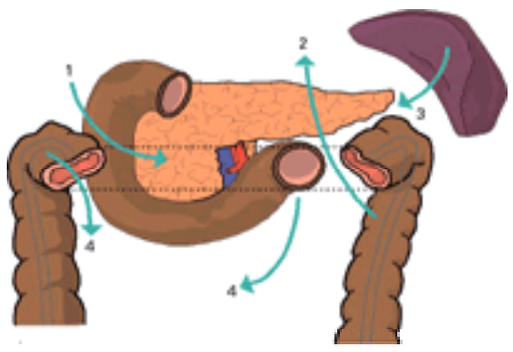


Fig. 2. Operative examination of the pancreas.

Injury to the gland may be identified by the presence of a capsular tear, by obvious haemorrhage from the gland, or, rarely, by its complete disruption. Ductal injury, however, may be very difficult to confirm, with nothing more obvious than some bruising in the gland. If facilities allow, a pancreatic ductogram should be obtained at the initial laparotomy: this can be done by intraoperative endoscopic pancreatography or by performing a transduodenal cannulation of the pancreatic duct. In practice, however, such refined endoscopic techniques are only available in a few specialized centres, and even satisfactory transduodenal pancreatograms are extremely difficult to obtain and interpret when performed in an emergency situation.

Management

Patients treated non-operatively

A recent, large, retrospective review of pancreatic trauma in children has shown that 79 per cent of grade I and II injuries can be successfully managed non-operatively. Careful observation, in the absence of clinical deterioration in children, rather than the conventional surgical exploration, was advocated.

Patients undergoing surgery

Grade-I injuries can be managed safely by drainage of the pancreatic bed: this is best achieved by bringing large tube drains out posterolaterally through the loins. Unless there is definite evidence of ductal injury, this is the treatment of choice.

If there is an obvious injury to the duct within the tail or body of the gland, a distal pancreatectomy should be performed, with ligation of the stump of the duct and drainage of the pancreatic bed. This approach in grade-II injuries carries fewer complications than drainage alone.

Grade-III and -IV injuries, in which the head of the gland is damaged, with or without trauma to the duodenum, may require a Whipple's pancreaticoduodenectomy to be performed. Fortunately this procedure is indicated in only 2 to 3 per cent of patients with pancreatic injuries. The operation is technically difficult, mainly because the very narrow calibre of the normal pancreatic and bile ducts makes reconstructions very difficult. In this situation the pancreatic anastomosis can be made by invaginating the stump of the gland into a loop of jejunum, and the biliary anastomosis may be facilitated by ligating the distal common bile duct and anastomosing the gallbladder to an accessible loop of jejunum. Many have advocated drainage alone as the treatment for these severe injuries because of the technical difficulties of pancreaticoduodenectomy and the poor results of surgery. In critically ill patients with multiple trauma, it would seem prudent to institute drainage of the pancreas in the first instance. Appropriate investigations and secondary surgery can be performed at a later date in those patients who survive their other injuries. The results of surgical treatment of pancreatic trauma are disappointing: the overall mortality rate is about 20 per cent. The mortality associated with pancreaticoduodenectomy in these patients approaches 50 per cent, but the results of less radical treatment of these major injuries (grades III and IV) are little better.

The prognosis of major pancreatic injuries (grades III and IV) is unlikely to improve. However, with prompt diagnosis and early treatment, patients who have grade-I and -II injuries should make a satisfactory recovery.

Further reading

Jones RC. Management of pancreatic trauma. *American Journal of Surgery*, 1985; **150**: 698–704. [Review of a large North American single centre experience of 500 cases of pancreatic trauma (72 per cent penetrating). The authors advocate conservative treatment but emphasize the importance of identifying ductal injury.]

Keller MS, Stafford PW, Vane DW. Conserative management of pancreatic trauma in children. *Journal of Trauma* 1997; **42** (6): 1097–100. [Review of 154 children, identified from the National Paediatric Trauma Registry, admitted to paediatric trauma centres in the United States with pancreatic injury. The authors conclude that early surgical intervention is unwarranted in the absence of deterioration or major ductal injury.]

Lucas CE. Diagnosis and treatment of pancreatic and duodenal injury. *Surgical Clinics of North America*, 1977; **57**: 49–65. [Describes the most commonly used classification of pancreatic injuries.]

33.4 Pancreatic cancer

Charles J. Yeo and John L. Cameron

Epidemiology and etiology
Pathology
Ductal adenocarcinoma
Adenosquamous carcinoma
Acinar cell carcinoma
Giant cell carcinoma
Pancreatoblastoma
Serous and mucinous neoplasms
Intraductal papillary-mucinous neoplasms
Other rare tumors
Growth factors
Genetics
Clinicopathologic staging
Diagnosis and clinical staging
Palliative intervention
Non-operative
Operative
Resectional therapy
Pancreaticoduodenectomy for tumors of the head, neck, or uncinate process
Distal pancreatectomy for tumors of the body and tail of the pancreas
Radiation and chemotherapy
Hormonal therapy
Gene therapy
Further reading

Epidemiology and etiology

In the United States the incidence rates for pancreatic cancer increased nearly threefold from 1920 to 1978. Since then, the annual pancreatic cancer incidence rate has remained constantat nine new cases per 100 000 population, ranking it eleventh in incidence among all cancers. In nearly all European countries, the incidence rates of pancreatic cancer have continued to rise. Throughout the world, over 185 000 new cases of pancreatic cancer occur yearly, with a death-to-incidence ratio of approximately 0.99. Not only is pancreatic cancer a major worldwide public health problem, but it remains a significant management challenge.

The demographics of pancreatic cancer have been widely investigated (Table 1). The incidence rates for pancreatic cancer increase steadily with age, with over 80 per cent of cases occurring between the ages of 60 and 80. Gender differences in pancreatic cancer have been equalizing over recent years, although the tumor still occurs more frequently in men.

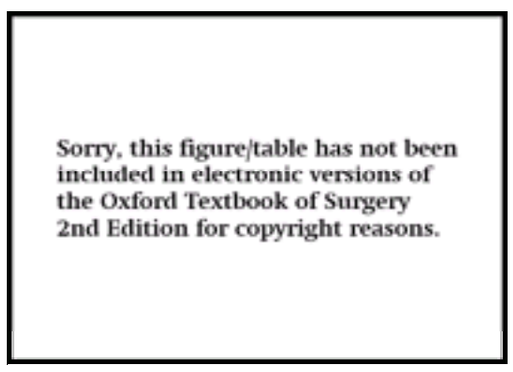


Table 1 No caption available.

Worldwide geographic differences in the incidence of pancreatic cancer occur. It is generally true that the incidence rates of pan-creatic cancer are highest in Western and industrialized countries, and lowest in underdeveloped nations.

Six genetic syndromes are associated with an increased risk of developing pancreatic cancer. These syndromes include hereditary non-polyposis colorectal cancer, familial breast cancer associated with the BRCA2 mutation, Peutz–Jeghers syndrome, ataxia-telangiectasia syndrome, familial atypical multiple mole melanoma syndrome, and hereditary pancreatitis.

The association of diabetes to pancreatic cancer has been studied by many investigators. Although the data are somewhat variable, most reports indicate no consistent association of diabetes with cancer of the pancreas, except when cases are included where the diabetes was diagnosed immediately prior to the cancer diagnosis. This relationship suggests that diabetes is more commonly an early symptom of pancreatic cancer rather than a causative influence.

Chronic pancreatitis appears to be linked to pancreatic cancer. Patients with chronic pancreatitis followed for several years have up to a 16-fold increased risk of pancreatic cancer. The cumulative 25-year risk of pancreatic cancer in patients with any form of chronic pancreatitis approaches 4 per cent. Other disorders with apparent increased risks of pancreatic cancer include thyroid and other endocrine tumors, cystic fibrosis, elevations of sex hormones, and pernicious anemia.

Of all the environmental factors that have been studied, there can be no doubt that cigarette smoking is a risk factor for pancreatic cancer. Case–control studies have found that smoking increases the risk of pancreatic cancer, with an odds ratio between 1.3 and 5.5 for current smokers. There appears to be a dose–response relationship with both the number of cigarettes smoked and the duration of smoking.

The relationship between nutrition and diet in pancreatic cancer has been extensively studied. From a number of case–control studies, there appears to be an association between pancreatic cancer and increasing intakes of carbohydrate, cholesterol, meat, salt, dehydrated food, fried food, refined sugar, soybeans, and nitrosamines. The risks are unproven for the excess ingestion of fat, b-carotene, and coffee. A protective influence has been reported for entities such as dietary fiber, vitamin C, fruits, vegetables, no preservatives, raw foods, pressure cooking, and microwave cooking.

Numerous studies have noted an increased risk of pancreatic cancer with certain occupational factors such as exposures to leather tanning, textiles, and various specific chemicals such as ethylene chlorhydrin, halogenated hydrocarbons, chlorinated water, and DDT.

Pathology

Cancer of the pancreas includes a number of pathologically distinct neoplasms, with varying morphologies, modes of presentation, and overall outcome (Table 2). Cancer of the pancreas can be broadly classified as primary, metastatic, or systemic. Primary cancers arise in the pancreas, and they can show either endocrine or non-endocrine differentiation. The focus of the discussion in this chapter will be upon the most common primary non-endocrine cancer: adenocarcinoma. Metastatic cancers originate elsewhere and spread hematogenously or via lymphatics to the pancreas. Systemic malignancies derive from the blood or lymph nodes and by definition simultaneously involve multiple sites, one of which may be the pancreas (Table 2).

Primary solid non-endocrine epithelial tumors
Ductal adenocarcinoma
Adenosquamous carcinoma
Acinar cell carcinoma
Giant cell carcinoma
Pancreatoblastoma
Primary cystic non-endocrine epithelial tumors
Serous cystic neoplasms
Mucinous cystic neoplasms
Intraductal papillary-mucinous neoplasms
Solid and cystic papillary neoplasms
Acinar cell cystadenocarcinoma
Endocrine tumors
Other tumors

Table 2 Pathology of pancreatic cancer

Ductal adenocarcinoma

This is the most common primary malignancy of the pancreas, accounting for over 75 per cent of all primary non-endocrine cancers. Two-thirds of ductal adenocarcinomas arise in the head, neck, or uncinate process of the pancreas, 15 per cent originate in the body and tail of the gland, and 20 per cent diffusely involve the gland. Grossly, ductal adenocarcinomas are poorly-defined firm masses with a whitish-yellow appearance that often obstruct and dilate the adjacent distal common bile duct or main pancreatic duct. Microscopically, ductal adenocarcinomas contain infiltrative glands of various shapes and sizes surrounded by dense, reactive fibrous tissue (Fig. 1). Many ductal adenocarcinomas infiltrate into vascular spaces, perineural spaces, and lymphatic spaces. At the time of presentation, most ductal adenocarcinomas have metastasized to peripancreatic lymph nodes.

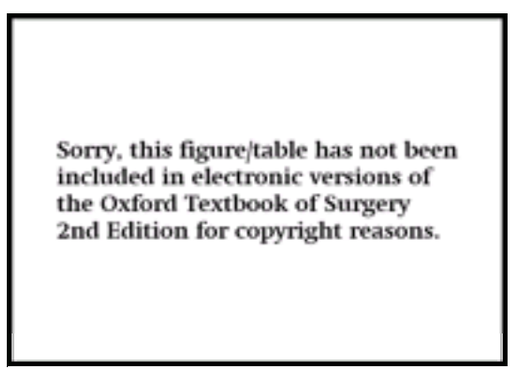


Fig. 1. A moderately differentiated ductal adenocarcinoma containing infiltrative glands of various shapes and sizes surrounded by dense, reactive fibrous tissue. The duct-forming cells are markedly atypical.

Histologically identifiable precursor lesions appear to progress to infiltrating ductal adenocarcinoma of the pancreas (Fig. 2). These precursor lesions originate within pancreatic ducts and ductules, and are frequently found adjacent to pancreatic cancer. Multiple linesof evidence suggest that pancreatic cancer may progress from flat duct lesions, to papillary duct lesions without atypia, to papillary ductlesions with atypia, and finally, to infiltrating adenocarcinoma.

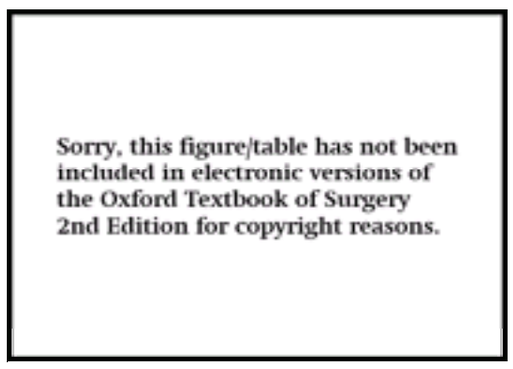


Fig. 2. Progression (from left to right) from normal ductal epithelium, to low-grade pancreatic intraepithelial neoplasia, to high-grade pancreatic intraepithelial neoplasia, to infiltrating cancer.

Adenosquamous carcinoma

This is a rare variant of ductal adenocarcinoma that demonstrates both glandular and squamous differentiation. The biologic behavior of this tumor appears to be identical to that of ductal adenocarcinoma.

Acinar cell carcinoma

This is an uncommon tumor with a distinct histologic appearance and occasionally an unusual clinical presentation. While most patients present with biliary or gastrointestinal obstruction, up to 20 per cent of patients also develop subcutaneous fat necrosis, an erythema nodosum-like rash, peripheral eosinophilia, and polyarthralgia. It is believed that these clinical signs and symptoms result from the systemic release of lipase by the tumor. Histologically these tumors form acini and their cells display eosinophilic, granular cytoplasm. Patients with these tumors have a better prognosis than do those with ductal adenocarcinoma.

Giant cell carcinoma

This is a rare pancreatic neoplasm marked microscopically by large, uninucleated or multinucleated tumor cells, with striking pleomorphism. The hyperchromatic nuclei are sharply angulated and contain prominent nucleoli and numerous, sometimes bizarre, mitotic figures. A minority of giant cell carcinomas manifest osteoclast-like giant cells. In this variant, the nuclei are round and uniform, and the tumors contain abundant osteoid and poorly formed glandular structures.

Pancreatoblastoma

This is also termed pancreatic carcinoma of infancy and occurs primarily in children and adolescents. These tumors grossly appear firm and lobulated, and microscopically they demonstrate small cells with a syncytial growth pattern, eosinophilic cytoplasm, and round benign-appearing nuclei. The prognosis is better than for ductal adenocarcinoma.

Serous and mucinous neoplasms

Of the primary cystic non-endocrine epithelial tumors, serous and mucinous neoplasms are the most common. Serous cystic neoplasms are more common in women, have a gross appearance of a well circumscribed multiloculated or uniloculated cyst, and microscopically are lined by a layer of simple cuboidal cells. Nearly all these

neoplasms are benign.

Mucinous cystic neoplasms include three types of neoplasms along a spectrum: the mucinous cystadenoma, the intermediate tumor, and the mucinous cystadenocarcinoma. Mucinous cystadenomas contain a single layer of epithelium without atypia, occasionally with the presence of a dense stroma (similar to that seen in the ovary) beneath the epithelial cells. In intermediate tumors, the epithelium may form papillas, the cells may show atypia, and the architecture may be more complex. Mucinous cystadenocarcinomas demonstrate invasion of the neoplastic epithelial cells into the surrounding stroma. Even in this latter subgroup, patients with resected tumors typically live much longer than patients with ductal adenocarcinoma, with over 50 per cent of patients surviving 5 years.

Intraductal papillary-mucinous neoplasms

These occur with equal frequency in both genders and are often diagnosed at endoscopy by the finding of mucin oozing from the ampulla of Vater. These tumors are typically soft and marked by mucus-filled dilated pancreatic ducts. Most patients with these tumors are curable by pancreatic resection, although metastases do occur.

Other rare tumors

Rare tumors of the pancreas include schwannoma, leiomyosarcoma, liposarcoma, primitive neuroectodermal tumor, and malignant fibrous histiocytoma. Tumors can metastasize to the pancreas, with the most common primary tumor sites being breast, lung, colorectum, melanoma, and renal cell carcinoma. Most pancreatic lymphomas are non-Hodgkin lymphomas. The pancreas may also be involved with common leukemias such as chronic myelogenous leukemia, acute lymphoblastic leukemia, and acute lymphocytic leukemia.

Growth factors

Various polypeptide growth factors and their receptors appear to play a role in the regulation of pancreatic cancer cell growth. These growth factors are produced by many cells, act at or near their sites of expression through autocrine and paracrine mechanisms, and may also exert their effects before release from the cell via a so-called ‘juxtacrine' mechanism.

The epidermal growth factor receptor binds several ligands such as epidermal growth factor, transforming growth factor-a, epiregulin, and others. Overexpression of epidermal growth factor receptor and its ligands has been correlated with enhanced pancreatic cancer metastatic potential and increased tumor invasiveness.

The transforming growth factor-b polypeptide family has also been studied in pancreatic cancer. Transforming growth factor-b appears to be overexpressed in certain pancreatic cancers, and such overexpression appears to promote tumor angiogenesis and enhance adhesiveness.

The fibroblast growth factor family of polypeptides has been found to be overexpressed in pancreatic cancer. Fibroblast growth factor appears to participate in the enhancement of pancreatic cell growth and may contribute to abnormal epithelial–mesenchymal interactions within the growing neoplasm.

Genetics

Cancer of the pancreas has been the focus of much recent molecular study. Tumor suppressor genes, oncogenes, and DNA mismatch repair genes have all been found to be genetically altered in pancreatic ductal adenocarcinoma. The tumor suppressor genes *p53, p16*, and *DPC4* are inactivated frequently in sporadic adenocarcinoma of the pancreas (Table 3).

Sorry, this figure/table has not been included in electronic versions of the Oxford Textbook of Surgery 2nd Edition for copyright reasons.

Table 3 No caption available.

p53 is a well-characterized tumor suppressor gene which lies on chromosome 17p. *p53* function appears to be inactivated in up to 75 per cent of all pancreatic cancers. The inactivation of *p53* gene function appears to lead to the loss of two important controls of cell growth: the regulation of cellular proliferation, and the induction of cell death (apoptosis).

The *p16* gene resides on chromosome 9p, and it appears to be altered in up to 95 per cent of all pancreatic cancers. Alteration of the *p16* gene product appears to inhibit the phosphorylation of a number of growth and regulatory proteins, leading to a failure of cellular growth and relatively unchecked cellular proliferation.

A newly identified tumor suppressor gene *DPC4* resides on chromosome 18q, and appears to be inactivated in half of all pancreatic cancers. The *DPC4* gene product has striking amino acid homology to a family of proteins called MAD proteins, which may play a role in signal transduction via the transforming growth factor-b family of cell surface receptors.

The most important oncogene studied to date in pancreatic cancer is the *k-ras* oncogene. Point mutations in codons 12, 13, or 61 of the *k-ras* oncogene impair the intrinsic GTPase activity of its gene product, resulting in a protein that is constitutively active in signal transduction. *k-ras* mutations have been found in between 80 and 100 per cent of pancreatic cancers, most of these being mutations in codon 12. *k-ras* mutation appears to occur early in the development of pancreatic cancer, making it a target for molecular-based tests for early detection.

Approximately 4 per cent of pancreatic cancers have been found to contain genetic alterations in DNA mismatch repair genes. The mismatch repair enzymes coded by the genes function as heterodimers to repair single base pair changes and small insertions/deletions that occur during DNA replication. Tumors found to contain mutations in DNA mismatch repair genes are histologically notable for expanding borders, poor differentiation, and a prominent syncytial growth pattern. This subgroup of tumors may have a more favorable prognosis.

To date, six clinical syndromes have been identified as being associated with pancreatic cancer (Table 4). Hereditary pancreatitis is an autosomal dominant disorder with 80 per cent penetrance and variable expressivity that is characterized by recurrent bouts of severe epigastric pain and hyperamylasemia, with onset usually before the teenage years. The hereditary pancreatitis mutation eliminates a trypsin-sensitive hydrolysis site in the chain linking the two globular domains of the trypsin molecule, rendering the patient susceptible to recurrent episodes of pancreatic autodigestion and acute pancreatitis.

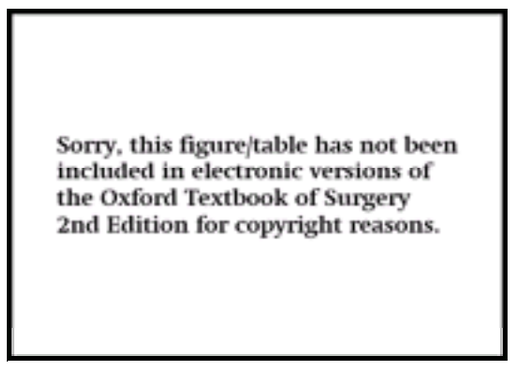


Table 4 No caption available.

Hereditary non-polyposis colorectal cancer is associated with an increased risk of cancers of the colorectum, endometrium, stomach, ovary, and pancreas. This syndrome is caused by germline mutations in one of the several DNA mismatch repair genes.

Familial breast cancer associated with the *BRCA2* tumor suppressor mutation is associated with an increased risk of not only breast cancer but also cancer of the ovaries, prostate, colon, and pancreas.

The familial atypical multiple mole melanoma syndrome is characterized by multiple nevi, atypical nevi, and melanomas. The gene responsible for this syndrome is the *p16* tumor suppressor gene.

The ataxia-telangiectasia syndrome is associated with disabling cerebellar ataxia, telangiectasias, and humoral and cellular immune deficiencies. The gene responsible for this syndrome is termed the *ATM* gene. Carriers of this gene mutation are at increased risk for the development of breast, ovarian, biliary tract, stomach, and pancreatic cancer.

The Peutz–Jeghers syndrome is characterized by hamartomatous polyps of the gastrointestinal tract, and mucocutaneous melanin deposition. The responsible ene has been recently identified on chromosome 19, and it encodes abnormal forms of a serine threonine kinase known as *STK11*.

Clinicopathologic staging

The most commonly used staging system for pancreatic cancer is the Union Internationale Contra Le Cancer (UICC) system, which is based upon TNM factors (Table 5). This is a relatively easy to use system that allows stage classification via tumor, lymph node, and metastatic parameters. A more complex system has been proposed by the Japan Pancreas Society (Table 6). This system adds other factors to the classification such as serosal invasion, retroperitoneal invasion, and invasion of the portal venous system. This system is considerably more difficult to apply, and has gained only limited use.

Stage grouping	*T*	*N*	*M*	*5-year survival (%)*
Stage I	T_1 or T_2	N_0	M_0	20–40
Stage II	T_3	N_0	M_0	10–25
Stage III	Any T	N_1	M_0	10–15
Stage IV	Any T	Any N	M_1	0–8

Tumor (T): T_1 = limited to pancreas; T_2 = extension directly to duodenum, bile duct, or peripancreatic tissues; T_3 = extension directly to stomach, spleen, colon, or adjacent large vessels.
Lymph nodes (N): N_0 = no lymph node metastases; N_1 = lymph node metastases.
Distant metastasis (M): M_0 = no distant metastases; M_1 = distant metastases.
Adapted from UICC. *TNM classification of malignant tumors*. 4th edn. Springer Verlag, New York, 1987.

Table 5 Union Internationale Contra Le Cancer (UICC) staging of pancreatic cancer

Stage grouping	*T*	*S*	*RP*	*PV*	*N*	*M*	*5-year survival (%)*
Stage I	T_1	S_0	RP_0	PV_0	N_0	M_0	35–45
Stage II	T_2	S_1	RP_1	PV_1	N_1	M_0	15–25
Stage III	T_3	S_2	RP_2	PV_2	N_2	M_0	5–15
Stage IV	T_4	S_3	RP_3	PV_3	N_3	M_1	0–10

Tumor (T): T_1 = 0 to 2cm; T_2 = 2.1 to 4cm; T_3 = 4.1 to 6cm; T_4 = ¿ 6.1cm.
Serosal invasion (S); Retroperitoneal invasion (RP); Portal venous invasion (PV): 0 = absence of invasion; 1 = suspected invasion; 2 = definite invasion; 3 = severe invasion.
Lymph nodes (N): N_0 = no metastasis; N_1 = primary lymph node group metastasis; N_2 = secondary lymph node group metastasis; N_3 = tertiary lymph node group metastasis.
Distant metastasis (M): M_0 = no distant metastasis; M_1 = distant metastasis.

Table 6 Japan Pancreas Society stage classification

Diagnosis and clinical staging

Most patients with pancreatic cancer come to medical attention after the development of jaundice. Jaundice occurs as a result of the neoplasm, arising in the head of the pancreas, obstructing the intrapancreatic portion of the common bile duct. Accompanying symptoms may include weight loss, abdominal pain, pruritus, weakness, alteration of bowel habits, and anorexia. With careful questioning, early symptoms of vague, somewhat diffuse upper abdominal or back discomfort may have been present for many months prior to the development of jaundice, and may have been overlooked by the patient or attributed to other causes.

Pancreatic cancer may present in an atypical fashion. In 10 per cent of patients new-onset diabetes may be the first clinical feature. Occasionally, acute pancreatitis may be the first manifestation of a pancreatic neoplasm, related to partial obstruction of the pancreatic duct.

Other symptoms found in some patients include nausea and/or vomiting, related to luminal narrowing of the upper gastrointestinal tract. In patients with tumors of the pancreatic head, the second portion of the duodenum is the usual site of obstruction. In patients with tumors of the pancreatic body, the obstruction occurs at the duodenojejunal junction.

The most common physical finding at initial presentation is jaundice. A palpable gallbladder may also be found. In cases of advanced disease, there may be evidence of cachexia, muscle wasting, and a nodular texture of the liver consistent with metastatic disease. Physical findings that serve as markers of disseminated tumor include left supraclavicular adenopathy (Virchow's node), periumbilical lymphadenopathy (Sister Mary Joseph's node), and the finding of metastases in the pelvis encircling the perirectal region (Blumer's shelf).

Laboratory studies typically reveal elevated serum bilirubin, alkaline phosphatase, and g-glutamyl transpeptidase, with mild to modest elevations of the hepatic aminotransferases. Normochromic anemia and hypoalbuminemia may reflect the chronic nature of the neoplastic process and its nutritional sequelae. In patients with

deep jaundice, the coagulation parameters should be checked because prolonged exclusion of the bile from the gastrointestinal tract leads to malabsorption of the fat-soluble vitamins and decreased hepatic production of vitamin K-dependent clotting factors, leading to a prolongation of the prothrombin time.

Many tumor markers have been studied in an attempt to facilitate an early diagnosis of pancreatic cancer. At present, there are no accurate and reliable serum markers that can be used for the diagnosis of pancreatic cancer. The best of the available markers is the carbohydrate antigen 19–9 (CA19–9), a Lewis blood group-related mucin. Using an upper limit of normal of 37 units/ml, CA19–9 only approaches an 80 per cent accuracy in identifying patients with pancreatic cancer. Increasing the cut-off value to 200 units/ml further increases the accuracy to almost 95 per cent. The combined use of CA19–9 with imaging modalities such as ultrasound, CT, or endoscopic retrogade cholangiopancreatography (**ERCP**) improves the diagnostic accuracy of each individual test, such that the accuracy can approach, but not achieve, 100 per cent. Results with other markers such as SPAN-1, CA-50, and DUPAN-2 have all been disappointing. One hope for the future lies in advances in molecular genetic technology, making early detection of small curable pancreatic cancer a reality.

In the appropriate clinical setting, high-quality spiral (also termed helical) CT is the preferred non-invasive imaging modality used for the diagnosis and staging of pancreatic cancer. The tumor typically appears as a hypodense focal lesion associated with pancreatic enlargement (Fig. 3). Unless there are obvious liver lesions consistent with hepatic metastases, the CT scan is not truly diagnostic of pancreatic cancer, but is highly suggestive.



Fig. 3. A spiral CT image revealing a large hypodense mass in the head of the pancreas in a patient with obstructive jaundice. Associated findings include dilated intrahepatic ducts and a distended gallbladder.

The new technique of ultrafast spin-echo MRI has been reported to be more sensitive than classic (non-spiral) CT scanning, for the diagnosis and staging of pancreatic cancer. While CT is usually more cost effective and less time consuming, pancreatic MRI is rapidly evolving. Advances in MRI techniques may lead it to play a more prominent role in pancreatic imaging in the future. While conventional MR scanning parameters require lengthy scanning times, fast spin-echo or turbo spin-echo techniques have been useful in reducing imaging times. On T_1-weighted images, pancreatic adenocarcinoma appears as a hypointense pancreatic mass relative to the more hyperintense normal pancreas (Fig. 4).

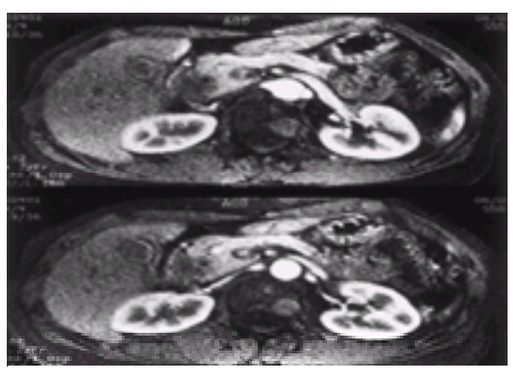


Fig. 4. Two T_1-weighted MR images using contrast enhancement with gadolinium. A mass in the head of the pancreas appears as a hypointense area.

ERCP allows direct imaging of the pancreatic duct, the site of origin of most pancreatic cancers. The sensitivity of ERCP for the diagnosis of pancreatic cancer is quite high, with the finding of a long irregular stricture in an otherwise normal pancreatic duct being virtually pathognomonic of pancreatic cancer in the appropriate clinical setting (Fig. 5). With the current sophistication of spiral CT scanning, the routine practice of diagnostic ERCP is unsupported. Diagnostic ERCP is probably best reserved for the evaluation of patients with presumed pancreatic cancer and obstructive jaundice in whom no mass is evident on CT, the symptomatic but non-jaundiced patient without an obvious pancreatic mass, or the occasional patient with chronic pancreatitis in whom the development of a pancreatic neoplasm is suspected based upon clinical deterioration. Therapeutic ERCP does play a role in some cases, as it allows biliary decompression via placement of an endoprosthesis (see Palliative intervention, below).

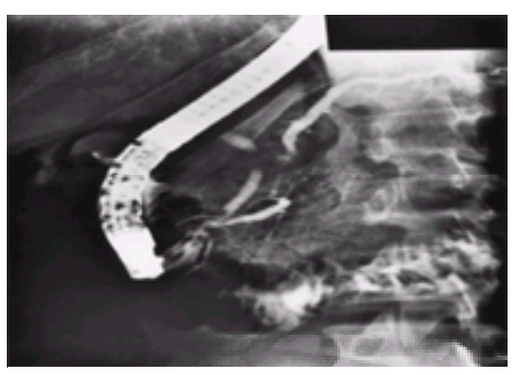


Fig. 5. An ERCP in a patient with obstructive jaundice, revealing a classic ‘double-duct' sign. There is evidence of tumor at the genu (or knee) of the common bile duct and pancreatic duct.

Positron emission tomography (**PET scanning**) uses the increased metabolism of glucose by pancreatic cancer cells as the basis of imaging. At present the application of PET scanning is limited because of the expensive and highly sophisticated technology needed in the absence of a clear-cut benefit over state-of-the-art CT scanning. Investigations continue in this area.

The use of percutaneous pancreatic biopsy in the diagnostic work-up of a patient with a suspected pancreatic cancer remains controversial. While percutaneous biopsy can be safe, serious complications have been reported. Tumor dissemination along the subcutaneous needle tract has occurred, and concerns regarding tumor dissemination from capsular puncture have arisen. It is our philosophy that percutaneous biopsy has no role in the evaluation of a good-risk operative candidate with a clinically resectable pancreatic mass. A negative percutaneous biopsy would not preclude operative exploration and resection. However, there is a role for percutaneous pancreatic biopsy in a poor-risk patient in whom a major pancreatic resection is not possible, because they may be a candidate for palliative chemoradiation therapy. Also, some form of tissue diagnosis is mandatory in patients who are undergoing consideration for preoperative chemoradiation therapy (neoadjuvant protocols). Percutaneous biopsy should further be considered in patients whose clinical presentation and imaging studies are not suggestive of pancreatic

adenocarcinoma, but rather of a more uncommon entity such as pancreatic lymphoma.

While the diagnosis of pancreatic cancer is a separate concept from the clinical staging of pancreatic cancer, in practice there is substantial overlap between diagnostic and staging modalities. Spiral CT serves as the major staging tool for patients with pancreatic cancer. Spiral CT has the ability to demonstrate liver metastases and can determine the proximity of the primary neoplasm to major peripancreatic vascular structures such as the celiac axis, superior mesenteric artery, and the mesenteric venous structures. For tumors of the head, neck, or uncinate process of the pancreas, occlusion of the superior mesenteric vein or portal vein with the presence of periportal collateral vessels is a reliable sign of unresectability. In contrast, for tumors of the body and tail of the pancreas, occlusion of the splenic vein with perigastric collaterals does not preclude resection, and should not be used as a sign of unresectability.

Preoperative angiography was used more commonly in the past to predict resectability rate and to give information regarding vascular anatomy. While the angiographic finding of venous encasement often represents tumor ingrowth into the vessel (Fig. 6), at times such a venous abnormality will not be found to be associated with tumor invasion, and rather to be a result of the desmoplastic reaction surrounding the tumor. Some studies have shown that angiography adds to the reliability of CT, while other studies have shown that high-quality spiral CT can predict unresectability better than angiography alone. At present, angiography is not considered a routine step in the staging of patients with pancreatic cancer. We believe angiography should be applied selectively in cases where vascular involvement by CT scan is questionable, or in patients who have undergone previous operative palliation or chemoradiation therapy (because in such cases intraoperative vascular assessment may be difficult because of previous operation or radiation-induced scar formation).

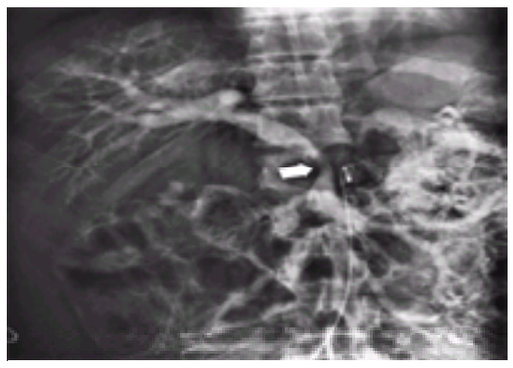


Fig. 6. Angiography: The venous phase of a superior mesenteric artery injection revealing encasement (narrowing) at the superior mesenteric vein–portal vein confluence (arrow).

The relatively new technique of endoscopic ultrasound (**EUS**) produces high-frequency images of the pancreas using the wall of the stomach and duodenum as an acoustic window. Although EUS itself does not provide a tissue diagnosis, the development of EUS-guided fine-needle aspiration makes the acquisition of tissue samples possible. EUS can be a sensitive method for the detection of pancreatic tumors, and it can also be used to stage the local extent of pancreatic cancer (Fig. 7). Staging via EUS has been reported to be equivalent or possibly superior to the combination of CT plus ERCP for cancers less than 3 cm in size, although further improvements in spiral CT will likely increase the accuracy of CT for such small pancreatic tumors. While EUS can be particularly useful in providing staging information regarding the extent of the primary tumor (T stage), EUS cannot reliably assess metastatic involvement of the liver, nor can it clearly establish lymph node involvement. Importantly, EUS is quite operator-dependent, and a high level of expertise is necessary for optimal results.



Fig. 7. Endoscopic ultrasonography (EUS) image using linear array echoendoscope, revealing a mass in the head of the pancreas with no vascular invasion of the superior mesenteric artery (SMA), superior mesenteric vein (SMV), or portal vein (portal).

A continuing controversy in the staging of patients with pancreatic cancer regards the use of laparoscopy. The rational for the application of laparoscopy comes from data which indicate that between 20 and 25 per cent of patients staged by all other modalities will be determined to have unanticipated peritoneal or liver metastases discovered at laparoscopy. However, the inherent bias favoring the utilization of laparoscopy involves a presumed equivalence of non-operative palliation with operative palliation in patients with pancreatic cancer. Laparoscopic staging only makes sense if patients who are deemed to be unresectable and spared laparotomy can be effectively palliated non-operatively. Staging laparoscopy can be performed with minimal morbidity and mortality on an outpatient basis evaluating the entire peritoneal surface, the pericolic gutters, the hemidiaphragms, and the pelvis as well as the surfaces of the liver. There are varying degrees of expertise in the application of laparoscopy, with some highly experienced groups performing a more extensive laparoscopic evaluation. Early reports of laparoscopy for the staging of pancreatic cancer reported findings of positive peritoneal disease precluding resection in over 70 per cent of patients. However, these studies were reported prior to the widespread availability of spiral CT. More recent studies have shown that the rate of positive peritoneal findings approaches 20 to 25 per cent for patients with pancreatic cancer staged via spiral CT, with there being a difference in the results based upon the site of origin of the pancreatic primary. For example, patients presenting with obstructive jaundice secondary to tumors of the head of the pancreas have only a 15 to 20 per cent incidence of unexpected intraperitoneal metastases after routine staging studies. In contrast, up to 50 per cent of patients with cancer of the body and tail of the pancreas have unexpected peritoneal metastases. Based upon these data, staging laparoscopy is best supported for patients with cancer of the body and tail of the pancreas. For good-risk patients with tumors in the head of the pancreas who present with obstructive jaundice and no evidence of liver metastases, many surgeons believe the optimal management involves surgical palliation, to include biliary–enteric bypass, gastrojejunostomy, and alcohol celiac nerve block. Staging laparoscopy serves no purpose in such a setting.

An algorithm for the management of patients with pancreatic cancer is shown in Fig. 8. When spiral CT shows a resectable, localized mass in the head, neck, or uncinate process of the pancreas which is causing biliary obstruction, the patient undergoes operative exploration without further investigation. Patients with resectable tumors undergo pancreaticoduodenectomy while those found to have unresectable tumor undergo palliative procedures. For patients with tumors of the body and tail of the pancreas which appear resectable by spiral CT, staging laparoscopy is appropriate prior to laparotomy. In patients with unresectable pancreatic cancer as determined by spiral CT, the algorithm is largely dependent upon the presence or absence of biliary obstruction. In those patients with obstructive jaundice and a limited life expectancy, a non-operative route of biliary drainage is preferred. In patients with good performance status and symptoms of gastric outlet obstruction related to tumor involvement, operative palliation to include biliary–enteric bypass, gastrojejunostomy, and alcohol celiac nerve block appears appropriate. Non-jaundiced patients without symptoms of gastroduodenal obstruction are managed non-operatively.

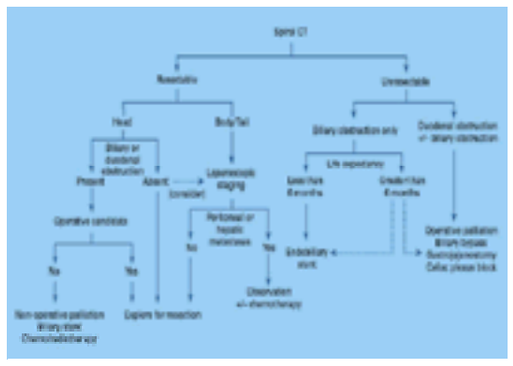


Fig. 8. Algorithm for use in the clinical setting of presumptive pancreatic cancer.

Palliative intervention

Non-operative

Non-operative management is appropriate in patients who are determined by staging studies to have distant metastases, unresectable local disease, or disseminated intra-abdominal tumor. It is also appropriate in patients with acute or chronic debilitating diseases that make anesthesia and surgery prohibitive.

Non-operative biliary decompression can now be achieved either by endoscopic or percutaneous transhepatic catheters in nearly all patients. The technique of endoscopic biliary stent (endoprosthesis) insertion is now standardized, with a technical success rate exceeding 90 per cent (Fig. 9). Complications of endoprosthesis placement include the early risks of cholangitis, pancreatitis, or duodenal perforation, and the late risks of cholecystitis, stent migration, and stent occlusion. Some improvement in stent patency has been reported with the use of metallic expandable endoprostheses.

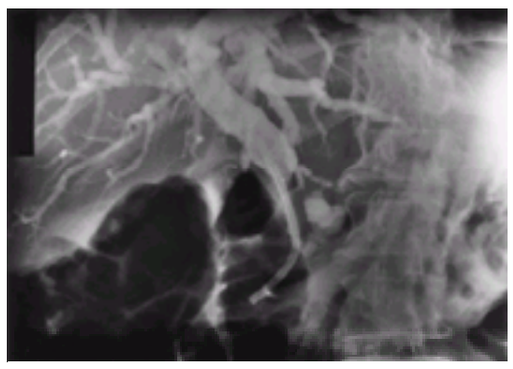


Fig. 9. An image obtained immediately following a successful ERCP. A 10 French plastic endoprosthesis has been placed across a malignant distal bile duct obstruction. A markedly dilated intrahepatic biliary tree is visualized.

Percutaneous transhepatic biliary drainage offers an alternative route for the relief of biliary obstruction. Diagnostic cholangiography first defines the site of biliary duct obstruction and serves as a road map for the biliary drainage (Fig. 10). Biliary drainage is performed with an internal–external catheter which can be exchanged for a larger diameter, softer tube at subsequent manipulations. Experience has accumulated with the placement of percutaneous metallic endoprostheses as well. Complications of percutaneous transhepatic catheter drainage include hemobilia related to the transhepatic route, bile peritonitis, bile pleural effusion, cholangitis, pancreatitis, and acute cholecystitis.

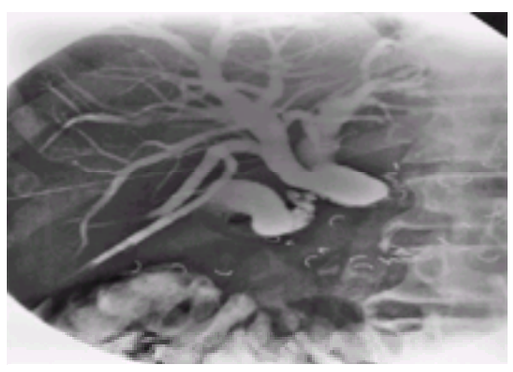


Fig. 10. A percutaneous transhepatic cholangiogram in a patient with obstructive jaundice. There is a malignant obstruction of the biliary tree at the level of the junction of the common hepatic duct and cystic duct. Staples are visible from a remote laparotomy for peptic ulcer disease.

The endoscopic approach is the method of choice for the non-operative palliation of jaundice in patients with pancreatic cancer. The endoscopic approach has a comparable success rate with the percutaneous approach, and the endoscopic approach is generally better tolerated and associated with a lower degree of procedure-related morbidity.

Tumor-associated pain can be a disturbing and incapacitating symptom of pancreatic cancer. For many patients, this symptom is poorly managed. Analgesic therapy is guided by the Three Step Analgesic Ladder of the World Health Organization. Tumor-associated pain is best treated with long-acting oral analgesics in appropriate doses, with the most commonly used drug being long-acting morphine sulfate. In patients who cannot take oral medications, topical analgesics (fentanyl) worn as continuous-release cutaneous patches can be highly effective. Poorly controlled pain is usually the result of inadequate analgesic dosing. Several non-operative treatment modalities exist to manage intractable pain that does not respond to oral or topical pain medication. These modalities include percutaneous celiac nerve block performed with either fluoroscopic or CT guidance, and external beam radiation therapy. Recently, thorascopic splanchnicectomy and endoscopic chemical splanchnicectomy performed with the aid of ultrasound guidance have been reported as techniques to control cancer-associated pain, however the techniques are too new to allow assessment.

The final symptom that may require palliation is duodenal obstruction. Until recently, there was little option besides surgical bypass (gastrojejunostomy). Recently, biliary-type metallic endoprostheses have been placed to palliate duodenal obstruction. The endoprosthesis is placed within the native duodenum, at the site of tumor infiltration. Careful assessment of this promising approach is required.

Operative

Palliative surgery for pancreatic cancer is designed to relieve biliary obstruction, avoid or treat duodenal obstruction, palliate tumor-associated pain, and to improve quality of life.

The surgical options for palliation of obstructive jaundice all include some form of an internal biliary bypass. The three most common techniques include

choledochoduodenostomy, cholecystojejunostomy, or hepatico- (choledocho-) jejunostomy. In general, choledochoduodenostomy is avoided because of concerns regarding the proximity of the anastomosis to the tumor, with the possibility of recurrent jaundice. The use of the gallbladder as a conduit (via cholecystojejunostomy) should also be avoided, again because of a relatively high incidence of recurrent jaundice. The most favored technique is a biliary–entebic bypacs to the jejunum (Fig. 11), performed as either an end-to-side hepaticojejunostomy or an end-to-side choledochojejunostomy, with the gallbladder being removed prior to dissection and mobilization of the extrahepatic biliary tree.

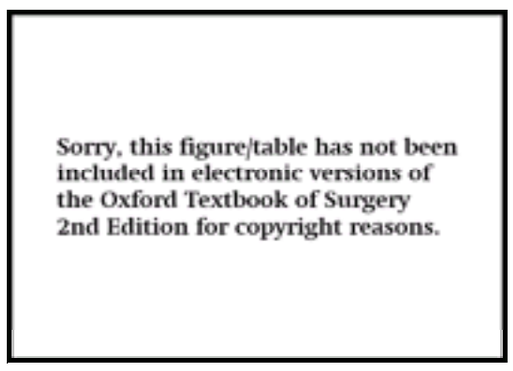


Fig. 11. An illustration depicting the anatomy following one method of palliative intervention. The biliary–enteric anastomosis is shown as a retrocolic end-to-side hepaticojejunostomy using a jejunal loop. A jejunojejunostomy is performed below the transverse mesocolon, to divert the enteric stream away from the biliary–enteric anastomosis. Also shown is a retrocolic gastrojejunostomy.

At the time of diagnosis of pancreatic cancer, approximately one-third of patients will have symptoms of nausea and/or vomiting. As the malignant disease progresses, duodenal obstruction manifests in an increasing number of patients. Obstruction can occur at either the duodenal C loop (by a cancer in the head of the pancreas), or at the ligament of Treitz (by a cancer in the body of the pancreas). While there is some controversy over the use of prophylactic gastrojejunostomy in patients who do not have mechanical obstruction at the time of laparotomy, it is our policy to perform gastrojejunostomy in patients undergoing operative palliation of pancreatic cancer. We create a retrocolic, isoperistaltic loop gastrojejunostomy, using the jejunum 20 to 30 cm beyond the ligament of Treitz. The anastomosis is created somewhat posterior, in the most dependent portion of the gastric greater curvature. Using this technique the incidence of early delayed gastric emptying is low, and hospital discharge is not delayed. Vagotomy is not performed in this setting, as it may contribute to delayed gastric emptying. Instead, routine acid secretory inhibition agents (such as histamine H_2- receptor antagonists or substituted benzamidazoles) are used to prevent marginal ulceration at the gastro-jejunal anastomosis.

The abdominal and back pain associated with unresectable pancreatic cancer can be a major disabling symptom for the patient and is addressed by intraoperative chemical splanchnicectomy directed at the celiac axis. This is performed by the injection of 20 ml of 50 per cent alcohol on each side of the aorta at the level of the celiac axis.

Results for the surgical palliation of unresectable pancreatic cancer at The Johns Hopkins Hospital are shown in Table 7. These data indicate that surgical palliation can be accomplished with a low in-hospital mortality rate, with a mean postoperative survival of nearly 8 months.

In-hospital mortality	2.5%
No postoperative complications	63%
Postoperative complications	37%
Wound infection	10%
Cholangitis	8%
Delayed gastric emptying	8%
Postoperative length of stay*	14 days
1-year survival	19%
2-year survival	4%
Mean survival	7.7 months
Longest survivor	27 months
Late complications	
Gastric outlet obstruction	4%
Recurrent jaundice	2%

*Note: These length of stay data reflect our practice from over 5 years ago. Current postoperative length of stay approaches 10 days.

Table 7 The Johns Hopkins experience with surgical palliation of periampullary cancer (*n* = 118 patients)

Resectional therapy

Pancreaticoduodenectomy for tumors of the head, neck, or uncinate process

Pancreaticoduodenectomy is currently the resectional procedure of choice for patients with pancreatic cancer localized to the head, neck, or uncinate process of the pancreas. In patients undergoing exploration, the initial portion of the operative procedure is dedicated to the assessment of resectability via a thorough intraoperative evaluation. An extensive Kocher maneuver is performed, elevating the duodenum and head of the pancreas out of the retroperitoneum. The porta hepatis is carefully assessed and the gallbladder is removed. Early division of the extrahepatic biliary tree allows caudal retraction of the distal common bile duct and opens the plane to visualize the anterior portion of the portal vein. The superior mesenteric vein is most easily identified during the performance of an extensive Kocher maneuver. In most cases, the proximal gastrointestinal tract is divided 2 cm distal to the pylorus, to yield a pylorus-preserving pancreaticoduodenectomy (Fig. 12). Multiple options exist for the reconstruction of the pancreas, bile duct, and gastrointestinal tract. Most commonly, the reconstructive technique anastomoses the pancreas first, followed by the bile duct and the duodenum. An alternative for pancreatic–enteric reconstruction involves the use of a pancreaticogastrostomy (Fig. 13), which has been touted as a technique to reduce the incidence of pancreatic fistula.

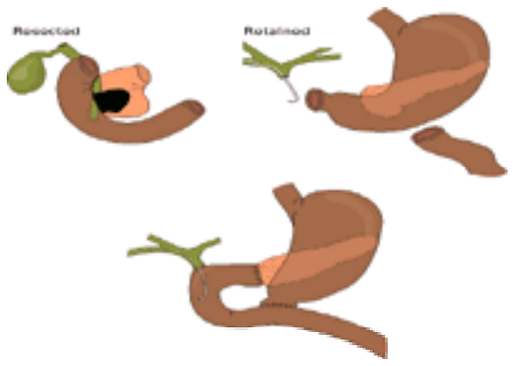


Fig. 12. Pylorus-preserving pancreaticoduodenectomy. Top left: The structures resected include the duodenum (except for the initial 1 to 2 cm beyond the pylorus); head, neck and uncinate process of the pancreas, with tumor (black); gallbladder; and distal extrahepatic biliary tree. Top right: The structures retained include the entire stomach, pylorus, proximal 1 to 2 cm of duodenum, body and tail of the pancreas, proximal biliary tree, and jejunum distal to the ligament of Treitz. Bottom: The reconstruction is shown as a proximal end-to-end pancreaticojejunostomy, hepaticojejunostomy decompressed via a percutaneous transhepatic catheter, and a distal duodenojejunostomy.

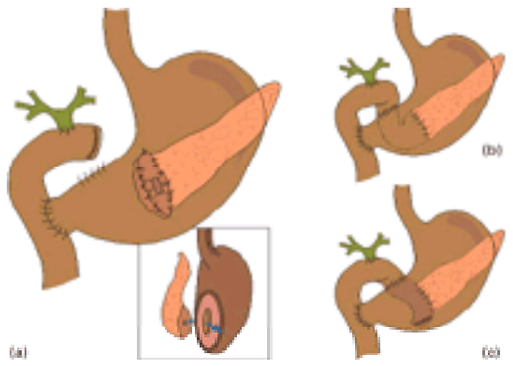


Fig. 13. Schematic illustration of (a) pancreaticogastrostomy, (b) end-to-end pancreaticojejunostomy, and (c) end-to-side pancreaticojejunostomy. The inset details the posterior gastrotomy.

The operative mortality rate for pancreaticoduodenectomy is currently less than 3 per cent in surgical centers with a large experience of the procedure. Leading causes of postoperative death include postoperative sepsis, hemorrhage, and cardiovascular events. In contrast to the low mortality rate, the incidence of postoperative complications can approach 40 to 50 per cent (Table 8).

Sorry, this figure/table has not been included in electronic versions of the Oxford Textbook of Surgery 2nd Edition for copyright reasons.

Table 8 No caption available.

The prognosis for patients with resected adenocarcinoma of the head, neck, or uncinate process of the pancreas is determined by several factors: clinical pathologic staging, tumor biology, molecular genetics, perioperative factors, and the use of postoperative adjuvant therapy. Overall, patients undergoing pancreaticoduodenectomy for pancreatic adenocarcinoma have a median survival rate ranging from 12 to 20 months, with a 5-year survival rate of 15 to 20 per cent. The more advanced the clinicopathologic staging, the worse the prognosis. Tumor DNA content correlates with prognosis, as patients undergoing resection with diploid tumors fare significantly better than those with aneuploid tumors. Further, as the number of tumor suppressor gene mutations increases, there appears to be an increasing risk of death from cancer. A further short-term outcome factor following pancreaticoduodenectomy is that patients operated on at high-volume institutions experience lower perioperative mortality and a shorter hospital stay than patients operated on at institutions with little or no experience with the resection.

Several controversies remain pertaining to the technique of pancreaticoduodenectomy. While pylorus preservation is favored for most patients who undergo this resection, some surgeons routinely perform a classic pancreaticoduodenectomy, which adds a distal gastrectomy. The issue of the safest technique of pancreatic–enteric reconstruction remains unsettled, with there being no consensus regarding the most secure method. Pancreaticojejunostomy is most typically performed, although others perform pancreaticogastrostomy. A further controversy involves the extent of lymph node dissection that should be added to pancreaticoduodenectomy. Extended (radical) pancreaticoduodenectomy was popularized several decades ago and has been reported to improve the resectability and survival rates in patients with pancreatic adenocarcinoma. The underlying strategy is to remove the pancreatic tumor with a wide soft-tissue margin encompassing a retroperitoneal lymph node dissection, with or without resection of the superior mesenteric vein–portal vein confluence. To date there are no published prospective, randomized data that demonstrate a consistent survival advantage for patients undergoing extended (radical) resection.

Distal pancreatectomy for tumors of the body and tail of the pancreas

Patients with tumors in this location frequently are diagnosed after some delay because these tumors do not obstruct the bile duct or the gastrointestinal tract in an early stage. As a result, the likelihood that curative resection will be possible is less than 10 per cent for these tumors. Nevertheless, if the diagnosis is made when the tumor is not metastatic, not encasing the celiac axis or mesenteric vessels, and not extending into the retroperitoneum, then resection remains an option. In these cases, many groups perform staging laparoscopy to search for metastatic disease not visualized by standard staging studies such as CT scan or endoscopic ultrasound. In the absence of findings of unresectability, mobilization of the inferior surface of the body of the pancreas from the retroperitoneum may be helpful to assess for retroperitoneal involvement. Localized tumors without extensive vascular or retroperitoneal involvement are appropriate for resection. Involvement of the splenic vein does not indicate unresectability. In general, splenic preservation is not indicated for malignant neoplasms, therefore the spleen is mobilized out of the retroperitoneum and resected *en bloc* with the distal pancreas (Fig. 14). Once the tumor has been elevated out of the retroperitoneum, the splenic vein can be controlled at a variable distance from its junction with the portal vein. Closure of the pancreatic resection line can be safely performed with mattress sutures or with a linear stapler. Overall, patients undergoing distal pancreatectomy for pancreatic adenocarcinoma have median survival rates ranging from 7 to 13 months, with a 5-year survival rate of approximately 10 per cent or less. Tumor size, lymph node involvement, and surgical margins all affect long-term outcome.

Sorry, this figure/table has not been included in electronic versions of the Oxford Textbook of Surgery 2nd Edition for copyright reasons.

Fig. 14. An illustration near the completion of a distal pancreatectomy and splenectomy for a large tumor in the body of the pancreas. The spleen and tail of the pancreas have been mobilized out of the retroperitoneum. The pancreatic parenchyma is being divided using the electrocautery.

Radiation and chemotherapy

A major obstacle to the cure of adenocarcinoma of the pancreas following any form of surgical resection is the burden of remaining regional subclinical disease. The classic approach to the treatment of this disease burden involves the combination of radiation and chemotherapy. Early observations regarding such combined modality therapy combined 5-fluorouracil and external beam radiation in patients with advanced disease. Several subsequent experiences have supported the use of 5-fluorouracil and external beam radiation therapy in patients who have undergone resection for pancreatic cancer. Unfortunately, the number of patients enrolled in some of these studies was small, the therapy protocols have varied, and there have been slow patient accrual rates. Thus, there remains some skepticism as to

whether postoperative adjuvant chemoradiation therapy is beneficial in patients following pancreatic resection.

A recent study from The Johns Hopkins Hospital prospectively evaluated 174 patients with resected adenocarcinoma of the head, neck, or uncinate process of the pancreas treated between 1991 and 1995 (Fig. 15). A Cox proportional hazards multivariate model indicated that the use of adjuvant chemoradiation therapy was associated with improved outcome. Based upon these data and the earlier results from the Gastrointestinal Tumor Study Group, we continue to recommend the combination of 5-fluorouracil based chemotherapy and radiation therapy in patients who have undergone pancreatic resection.

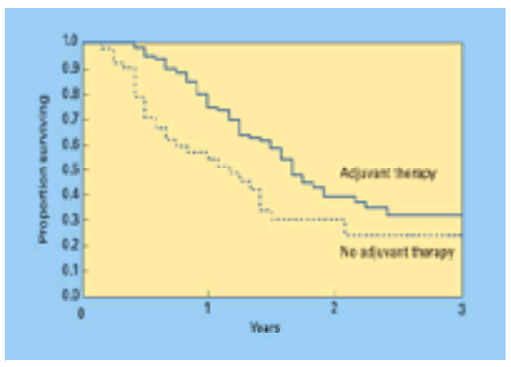


Fig. 15. The actuarial survival curves comparing patients receiving postoperative adjuvant chemoradiation therapy (*n* = 120) with those not receiving therapy (*n* = 53; *p* = 0.003). The median survivals are 19.5 months with adjuvant therapy and 13.5 months without therapy. (Not randomized.)

A newer concept for the treatment of patients with pancreatic cancer involves using chemotherapy and radiation therapy prior to surgical exploration and attempted resection, with this approach being termed neoadjuvant therapy. The theoretical advantages of this approach include a decreased potential burden of tumor at operation, sterilization of tumor cells prior to intraoperative manipulation, and the ability to administer systemic therapy earlier, prior to the postoperative period. Newer neoadjuvant chemoradiation protocols have yielded improvements over older protocols, with the newer techniques involving rapid fractionation radiation therapy over a 2-week period with concurrent 5-fluorouracil. The current controversy between the proponents of (i) surgical resection followed by postoperative chemoradiation and (ii) the neoadjuvant chemoradiation approach followed by surgical resection remains unresolved.

There have been many studies of antitumor therapies for patients with unresectable advanced pancreatic cancer. Many trials evaluating the use of chemotherapy and radiation therapy both alone and in combination have shown a marginal improvement in survival, often with relatively high toxicity rates and some negative impact on quality of life assessment. Recently, a somewhat novel chemotherapeutic agent, gemcitabine, has been released by the United States Food and Drug Administration for patients with advanced pancreatic carcinoma. Gemcitabine is an advance over 5-fluorouracil based chemotherapy in patients with advanced pancreatic cancer, with marginal improvement in median survival, improved pain control, and improved performance status. Further, gemcitabine also appears to be a potent radiation sensitizer, with applicability to patients when combined with radiation therapy. Multiple other agents are currently being studied for a role in the treatment of patients with pancreatic cancer.

Hormonal therapy

In vivo and *in vitro* experiments indicate that estrogens and androgens may affect pancreatic cancer growth. In clinical trials, hormone manipulation therapy has not resulted in measurable treatment benefit.

Further, two gastrointestinal hormones, cholecystokinin and gastrin, have been studied largely in animal models for efficacy against pancreatic cancer. Human trials are awaited.

Gene therapy

Although the conceptual basis for gene therapy initially was founded on the correction of single gene (monogenetic) inherited disorders, recombinant DNA technology has been applied to patients with cancer where multiple genetic defects are observed. Currently, an extensive number of anticancer phase I protocols, incorporating an array of gene transfer systems, are under investigation. Human pancreatic cancer has been a difficult target for extensive experimentation. However, early trials of a cytokine-secreting pancreatic adenocarcinoma vaccine have been undertaken. The rationale for these trials comes from preclinical studies with murine tumor models, which have demonstrated that tumor cell vaccines engineered to secrete certain cytokines can elicit systemic immune responses capable of eliminating established tumors. While the impact of this vaccine strategy on patients with pancreatic cancer remains uncertain, it is hoped that such therapy, combined with surgical resection and improvements in chemoradiation, may be associated with improvements in outcome.

Further reading

Brat DJ, Lillemoe KD, Yeo CJ, Warfield PB, Hruban RH. Progression of pancreatic intraductal neoplasms to infiltrating adenocarcinoma of the pancreas. *American Journal of Surgical Pathology* 1998; **22**: 163–9. [Data supporting a precursor progression for pancreatic cancer, from ductal lesions to carcinoma.]

Brennan MF, Moccia RD, Klimstra D. Management of adenocarcinoma of the body and tail of the pancreas. *Annals of Surgery* 1996; **223**: 506–12. [A discussion of the management of patients with cancers in the body and tail of the pancreas.]

DiGiuseppe JA, Yeo CJ, Hruban RH. Molecular biology and the diagnosis and treatment of adenocarcinoma of the pancreas. *Advances in Anatomic Pathology* 1996;**3**: 139–55. [A review of the molecular biology of pancreatic cancer.]

Fernandez-del Castillo C, Warshaw AL. Laparoscopic staging and peritoneal cytology. *Surgical Oncology Clinics of North America* 1998; **7**: 135–42. [A review of laparoscopic staging in pancreatic cancer.]

Goggins M *et al.* Pancreatic adenocarcinomas with DNA replication errors (RER+) are associated with wild-type *k-ras* and a characteristic histopathology: poor differentiation, a syncytial growth pattern, and pushing borders suggest RER+. *American Journal of Pathology* 1998; **152**: 1501–7. [The first report of pancreatic cancers with DNA mismatch repair gene abnormalities.]

Gold EB, Goldin SB. Epidemiology of and risk factors for pancreatic cancer. *Surgical Oncology Clinics of North America* 1998; **7**: 67–91. [A thorough review of the epidemiology of pancreatic cancer.]

Hahn SA *et al. DPC4*, a candidate tumor-suppressor gene at human chromosome 18q21.1. *Science* 1996; **271**: 350–3. [The initial report of the DPC4 tumor suppressor gene.]

Hruban RH, Petersen GM, Ha PK, Kern SE. Genetics of pancreatic cancer: from genes to families. *Surgical Oncology Clinics of North America* 1998; **7**: 1–23. [A review of the familial aspects of pancreatic cancer inheritance, including data from the National Familial Pancreas Tumor Registry.]

Jaffee EM *et al.* Development and characterization of cytokine-secreting pancreatic adenocarcinoma vaccine from primary tumors for use in clinical trials. *The Cancer Journal from Scientific American* 1998: **4**: 194–203. [A description of the development for and application of a tumor vaccine strategy to treat patients with pancreatic cancer following surgery.]

Kaufman SL. Percutaneous palliation of unresectable pancreatic cancer. *Surgical Clinics of North America* 1995; **75**: 989–99. [A review of percutaneous palliation.]

Korc M. Role of growth factors in pancreatic cancer. *Surgical Oncology Clinics of North America* 1998; **7**: 25–41. [A review of growth factors in pancreatic cancer.]

Lichtenstein DR, Carr-Locke DL. Endoscopic palliation for unresectable pancreatic cancer. *Surgical Clinics of North America* 1995; **75**: 969–88. [A review of endoscopic palliation.]

Lillemoe KD, Cameron JL, Kaufman HS, Yeo CJ, Pitt HA, Sauter PK. Chemical splanchnicectomy in patients with unresectable pancreatic cancer: a prospective randomized trial. *Annals of Surgery* 1993; **217**: 447–57. [A prospective randomized trial demonstrating the benefit of alcohol splanchnicectomy for palliating pain associated with pancreatic cancer.]

Miller AR *et al.* Neoadjuvant chemoradiation for adenocarcinoma of the pancreas. *Surgical Oncology Clinics of North America* 1998; **7**: 183–97. [A recent review of neoadjuvant chemoradiation therapy for pancreatic cancer.]

Rozenblum E *et al.* Tumor-suppressive pathways in pancreatic carcinoma. *Cancer Research* 1997; **57**: 1731–4. [A synthesis of the tumor suppressor story in pancreatic cancer, focusing on *p53, p16*, and *DPC4*.]

Wilentz RE, Hruban RH. Pathology of cancer of the pancreas. *Surgical Oncology Clinics of North America* 1998; **7**: 43–65. [A comprehensive review of the pathology of pancreatic cancer, with many photomicrographs.]

Yeo CJ. Pylorus-preserving pancreaticoduodenectomy. *Surgical Oncology Clinics of North America* 1998; **7**: 143–56. [A discussion of the most commonly used resectional procedure for pancreatic cancer.]

Yeo CJ, Cameron JL. Pancreatic cancer. *Current Problems in Surgery* 1999; **36**: 57–152. [An expanded discussion of the topic of pancreatic carcinoma.]

Yeo CJ *et al.* Pancreaticoduodenectomy for cancer of the head of the pancreas: 201 patients. *Annals of Surgery* 1995; **221**: 721–33. [The clinical experience with a large group of patients from a single institution, detailing prognostic factors predictive of outcome.]

Yeo CJ *et al.* Pancreaticoduodenectomy for pancreatic adenocarcinoma: postoperative adjuvant chemoradiation improves survival: a prospective, single institution experience. *Annals of Surgery* 1997; **225**: 621–36. [Data to support the use of chemoradiation therapy following pancreaticoduodenectomy.]

Yeo CJ, *et al.* Periampullary adenocarcinoma: analysis of 5-year survivors. *Annals of Surgery* 1998; **227**: 821–31. [An evaluation of patients with pancreatic cancer who survived more than 5 years following pancreaticoduodenectomy.]

Yeo CJ *et al.* Symposium on pancreatic cancer: 1998 update. *Journal of the American College of Surgeons* 1998: **187**: 429–42. [A review of molecular biology, staging, surgery, and non-operative treatments of pancreatic cancer.]

33.5 Insulinomas and other tumours

Jeffrey A Norton and Hop Le

Introduction
Insulinomas
Physiology of insulin production
Clinical presentation
Diagnosis
Pathology
Radiologic localization
Treatment
Gastrinoma
Symptoms
Diagnosis
Pathology
Radiographic localization
Treatment
Other functioning islet-cell tumors
Glucagonoma
Vasoactive intestinal peptide tumor (VIPoma)
Somatostatinoma
Nonfunctioning islet-cell tumors
Rare islet-cell tumors
Medical management
Tumor localization
Surgical therapy
Further reading

Introduction

Neuroendocrine tumors of the pancreas are uncommon, with an estimated incidence of between 6 and 10 per million population per year. Insulinomas and gastrinomas are the two most common functional islet-cell tumors. Some tumors, such as those that produce pancreatic peptide, may appear to be nonfunctional as this hormone has no symptoms in man. The clinical syndromes produced by excessive hormone secretion are their distinguishing feature. Although characteristic constellations of symptoms suggest a diagnosis, the specific diagnosis depends on measuring elevated serum concentrations of a circulating hormone plus the appropriate abnormality in a biologic substance. For example, in order to diagnose insulinoma one must find the serum insulin is elevated and the serum glucose reduced. With the exception of insulinomas, which are usually benign, islet-cell tumors are generally malignant but slow growing. Patients may live for years with distant metastases (Table 1).

Type	Malignancy (%)	Site	MEN-1
Insulinoma	< 10	Pancreas	5–10 %
VIPoma	50	Pancreas + duodenum	Occasional
Somatostatinoma	100	Pancreas + duodenum	Occasional
PPoma	100	Pancreas	High
Non-functioning	60	Pancreas	High
Glucagonoma	100	Pancreas	Occasional
Gastrinoma	60	Pancreas Duodenum	15–24 %

Table 1 Spectrum of malignancy, site, and prevalence of multiple endocrine neoplasia type 1 (MEN-1) in islet-cell tumours

The pancreatic islets of Langerhans contain at least five different cell types: b-cells, which produce insulin; a-cells, which produce glucagon; g-cells, which produce somatostatin; F-cells, which produce pancreatic polypeptide; and enterochromaffin cells, which produce serotonin. Islet cells are part of the APUD (*a*mine *p*recursor *u*ptake and *d*ecarboxylation) system. They stain positively for neurone-specific enolase and chromogranin A. Pancreatic islets may produce hormones that are not normally present in the pancreas, such as gastrin, adrenocorticotropin (**ACTH**), vasoactive intestinal polypeptide, and growth hormone-releasing hormone. Neuroendocrine cells are also found in the proximal portion of the duodenum and the antrum of the stomach, where they produce gastrin and somatostatin.

Although many pancreatic islet-cell tumors are multihormonal, one peptide generally predominates and is responsible for the clinical syndrome. Approximately 10 to 20 per cent of islet-cell tumors arise in association with multiple endocrine neoplasia type 1 (**MEN-1**) (Table 1). MEN-1 is inherited as an autosomal-dominant trait; it is characterized by tumors of multiple endocrine organs including pituitary, pancreas, and parathyroid. The gene for MEN-1 is called menin and has been localized to the long arm of chromosome 11. It is a tumor-suppressor gene, but its exact function is unknown. In patients with MEN-1, islet-cell tumors are always multifocal, may be multihormonal, and can occur in either the pancreas or the duodenum, or both; gastrinomas are the most common functional tumor in these patients, but any type of functional islet-cell tumor can occur. Non-functioning islet-cell tumors also occur and measurement of the serum pancreatic polypeptide has been used to screen for an islet-cell tumor.

Insulinomas

Insulinomas are the most common pancreatic islet-cell tumor, with an incidence of approximately 3 to 6 per million population per year. They are generally small (less than 2 cm), solitary, benign tumors located within the pancreas. They may be multiple in the setting of MEN-1. Fewer than 10 per cent are malignant. The mean age of patients with insulinoma is 45 years, but the tumor can occur at any age and generally occurs at a younger age in MEN-1. Patients with MEN-1 who have insulinoma usually present with the tumor in the third decade of life. Presentation is with neuroglycopenic symptoms that may be subtle and the diagnosis is often missed; it depends on an understanding of normal insulin secretion and glucose homeostasis, and of the changes that occur during a prolonged fast.

Physiology of insulin production

Because the measurement of insulin and glucose is essential to the diagnosis of insulinoma, it is important to understand the mechanisms of insulin synthesis, secretion, release, and control. Insulin is synthesized in the rough endoplasmic reticulum in a large form called preproinsulin. A 24-amino acid peptide is then cleaved to yield proinsulin, which is transported to the Golgi apparatus where it is packaged into secretory granules. These granules contain proteases that cleave the C-peptide sequence of the molecule to yield biologically active insulin. Therefore, when insulin is secreted, equimolar amounts of C-peptide and insulin can be detected in the serum. Although some proinsulin is normally released from b-cells, it is usually less than 25 per cent of the serum concentration of insulin. Measurement of serum C-peptide and proinsulin is useful in the differential diagnosis of hypoglycemia as the finding of elevated concentrations rules out factitious hypoglycemia.

Clinical presentation

Patients with insulinoma have symptoms caused by hypoglycemia. Most commonly they have neuroglycopenic symptoms that include seizures, difficulty awakening, visual disturbances, confusion, lethargy, weakness, or transient motor deficits. Erroneous psychiatric or neurologic diagnoses are common and insulinomas may go undiagnosed for years. Hypoglycemia also causes catecholamine release with subsequent sympathetic discharge leading to sweating, anxiety, and palpitations. Symptoms from insulinomas tend to occur early in the morning after an overnight fast or during periods of vigorous exercise. Patients learn that food relieves or avoids symptoms, and most with insulinoma gain weight. A family history of hypoglycemia or other endocrine abnormalities should be sought to exclude the presence of

MEN-1.

Diagnosis

In 1935, before the development of radioimmunoassay and the ability to measure serum insulin, Whipple proposed the diagnostic triad of hypoglycemic symptoms induced by fasting, blood glucose of less than 45 mg/dl, and relief of symptoms following glucose administration. Now, the ability to measure serum insulin in relation to glucose has simplified the diagnosis. The presence of severe hypoglycemia and an inappropriately elevated insulin is essential for the diagnosis. The diagnostic test of choice is a 72-h fast during which serum glucose and insulin are measured every 6 h, which requires admission to hospital. The fast is terminated at 72 h or when the patient develops neuroglycopenic symptoms. At that point, serum glucose, insulin, C-peptide, and proinsulin are measured, and intravenous glucose administered. Most patients with insulinoma develop neuroglycopenic symptoms within 12 to 24 h; some will mange to fast for 48 h, but none will fast for 72 h, so completion of the diagnostic fast excludes insulinoma as a cause for the hypoglycemia. During the fast, careful observation is necessary to avoid any complications of hypoglycemia and to exclude the use of either exogenous insulin or oral hypoglycemic drugs, which cause factitious hypoglycemia, an important condition in the differential diagnosis. Patients with factitious hypoglycemia are generally women and work as nurses or other health-care professionals. Urinary sulfonyl urea is also measured to exclude the use of oral hypoglycemic drugs.

The diagnosis of insulinoma is made based primarily on the results of the fast. It is important to note that patients must develop neuroglycopenic symptoms (such as confusion, seizures, sleepiness, drowsiness), at which time the blood samples are drawn and the fast terminated. Insulinoma is diagnosed by a serum glucose of less than 45 mg/dl and a concomitant insulin of more than 5 µunits/ml. Patients with an insulinoma have an immunoreactive insulin (µunits/ml):glucose (mg/dl) ratio greater than 0.3. They also have a higher serum proinsulin than normal individuals or those with factitious hypoglycemia; a ratio of proinsulin to insulin of more than 25 per cent is consistent with the diagnosis of insulinoma. C-peptide concentrations are also elevated. If factitious hypoglycemia is caused by the administration of exogenous insulin, appreciable amounts of proinsulin will not be detected. If it is caused by bovine or porcine insulin, it may be possible to detect antibodies to insulin. Thus the typical patient with an insulinoma will have neuroglycopenic symptoms during the fast, a serum glucose of less than 45 mg/dl, a serum insulin greater than 5 µunits/ml, a glucose:insulin ratio in excess of 0.3, a proinsulin:insulin ratio of more than 25 per cent, elevated C-peptide, undetectable urinary sulfonyl urea, and no insulin antibodies.

Pathology

Approximately 80 to 90 per cent of insulinomas are small (under 2 cm) and solitary, distributed equally within the head, body, or tail of the pancreas. Approximately 5 to 10 per cent may occur as multiple islet-cell tumors, usually associated with MEN-1, and a similar proportion are malignant, as best determined by the presence of metastases to lymph nodes or liver. A diffuse microadenomatosis, in which multiple, small, nonencapsulated tumors or nodules are distributed throughout the pancreas (nesidioblastosis), may occur in neonates as a cause of hypoglycemia, but this entity does not occur in adults. Insulinomas appear reddish brown because of increased vascularity. Microscopic examination cannot distinguish between benign and malignant insulinomas, although malignant lesions tend to be larger (average size 6 cm). Malignant insulinomas are diagnosed at the time of surgery by the identification of metastases in either lymph nodes or liver. In MEN-1 with multiple insulinomas the one responsible for the syndrome is usually a large, easily identifiable tumor, the removal of which will reduce the hypoglycemia. Pathologic examination of pancreatic specimens from patients with MEN-1 demonstrates that many other tumors are present; these generally do not secrete insulin.

Radiologic localization

Because insulinomas are small, their detection by conventional, noninvasive radiographic imaging studies is poor. Computed tomography (**CT**), magnetic resonance imaging (**MRI**), and ultrasonography identify only a minority of insulinomas (less than 40 per cent). However, either MRI or CT scans should be obtained in order to exclude liver metastases or a large malignant primary. Spiral CT is better than conventional CT; it should be performed with intravenous contrast and special thin sequences through the pancreas. On CT, insulinomas appear bright or as a blush because of increased numbers of blood vessels compared to the adjacent pancreas. On MRI, insulinoma and other islet-cell tumors will appear bright on sequences such as T_2 and inversion recovery. With ultrasound, insulinoma appears sonolucent compared to the more echo-dense pancreas. Each of these studies will only image approximately 20 to 40 per cent of insulinomas (Table 2).

Study	Insulinoma (sensitivity %)	Gastrinoma (sensitivity %)
Ultrasound	36	20
CT	17	62
MRI	25	26
Endoscopic ultrasound	80	40–85
SRS	0–62	91
Angiography	30	68
PVS	77	73
Angiographic stimulation	85–100	54–84

PVS, portal venous sampling; SRS, somatostatin-receptor scintigraphy.

Table 2 Localization studies for insulinoma and gastrinoma

Somatostatin-receptor scintigraphy, which images tumors based on the density of these receptors, has become the noninvasive imaging study of choice for islet-cell tumors other than insulinoma, which have a low density of receptors and thus are not imaged by this technique. Therefore, noninvasive imaging studies are of limited use in localizing insulinoma and only one study, such as CT, is indicated to image a large tumor or liver metastases.

Because of the inability accurately to localize insulinoma on CT, MRI, and somatostatin-receptor scintigraphy, some have recommended surgical exploration with intraoperative ultrasound and the avoidance of other costly, invasive, preoperative imaging studies. However, others have stated that invasive localization studies should be performed. Endoscopic ultrasonography was introduced in the early 1990s and has become one of the studies of choice to identify insulinoma. Its principle is to use the gastrointestinal tract to navigate the ultrasound probe closer to the pancreas for high-resolution images. It is observer dependent and is not available in all centers, but there are reports that it can correctly identify approximately 80 to 90 per cent of insulinomas; its use is recommended for certain institutions with appropriate expertise, where it may be the only preoperative imaging study that is necessary (Fig. 1).

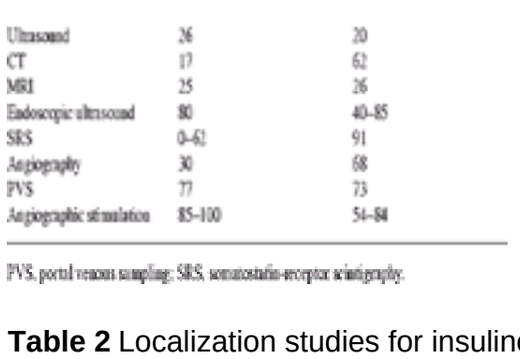


Fig. 1. Endoscopic ultrasonography of insulinoma. The endoscope is passed into the stomach and the balloon inflated with saline. A 2-cm insulinoma is identified in the tail of the pancreas (three arrows). Notice how the tumor is sonolucent compared to the more echo-dense pancreas. The tumor is near the splenic vein (v, single arrow), which courses through the body and tail of the pancreas.

As insulinomas are evenly distributed among the head, body, and tail of the pancreas, localization of the region of the gland bearing the tumor may be helpful. Sampling of insulin concentrations in the portal vein and its tributaries is able correctly to indicate the region of the pancreas containing the tumor in approximately 75 per cent of patients, because the concentration of insulin is highest in the vein from that area. This information is potentially very useful to the surgeon, who is then able

to remove the section of the pancreas containing the tumor, even if the tumor cannot precisely be identified intraoperatively. However, portal venous sampling has disadvantages. It is expensive, invasive, and tedious; it may result in significant complications, including pain and hemorrhage. Another study that appears to have replaced portal venous sampling is selective intra-arterial calcium angiography, which combines an angiogram with hepatic venous sampling, based on the finding that intravenous calcium stimulates insulinomas to secrete insulin. It has the potential to image the tumor (Fig. 2) and it can also determine the region of the pancreas containing it. The study is performed by sequentially catheterizing selective arteries that perfuse different regions of the pancreas, then selectively injecting calcium into the artery and measuring the insulin concentrations in the hepatic vein. There is a marked, rapid increase in the insulin concentration in the hepatic vein following injection of calcium into the artery that perfuses the insulinoma (Fig. 3). The calcium angiogram has correctly imaged approximately 50 per cent of insulinomas and correctly identified the region of the pancreas bearing the tumor in 85 per cent. Therefore, because it can image the tumor and provide regional localization, it is the preoperative study of choice for patients with insulinoma.

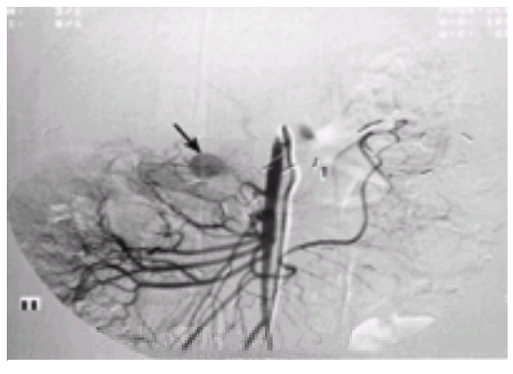


Fig. 2. Selective angiogram of the superior mesenteric artery showing a vascular blush in the head of the pancreas consistent with an insulinoma (arrow).

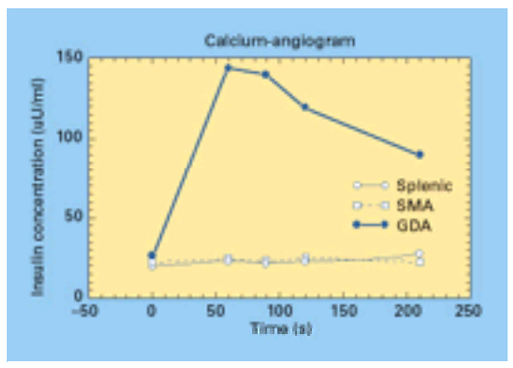


Fig. 3. Calcium angiogram of a patient with insulinoma. Calcium was injected sequentially into the splenic artery, superior mesenteric artery (SMA), and gastroduodenal artery (GDA). Insulin concentrations were measured at short intervals later in the hepatic vein. When calcium was injected into the GDA (the artery that perfused the tumor), insulin concentrations rapidly increased in the hepatic vein, localizing the tumor to the head of the pancreas.

Treatment

Surgery is the only curative treatment for patients with insulinoma, but the operation is not urgent and the patient is first managed medically. Symptoms of hypoglycemia can often be controlled by the use of frequent small meals, setting an alarm clock to awaken from sleep for feeding during the night, and the use of cornstarch to prolong the absorption of nutrients from the gut. Diazoxide has been used to control effectively hypoglycemia in approximately 60 per cent of patients with insulinoma. It may cause prolonged hypotension during anesthesia and should be discontinued at least 1 week before surgery. Calcium-channel blockers may also inhibit insulin secretion and help improve the hypoglycemia in some patients with insulinoma. Octreotide, the potent, long-acting analog of somatostatin, also decreases insulin secretion, but is generally ineffective at controlling the hypoglycemia of insulinoma. The presence of MEN-I should be excluded by testing for other components of the syndrome, which include primary hyperparathyroidism, other endocrine pancreatic islet-cell tumors, pituitary tumors, lipomas, and carcinoid tumors of the gut. Other manifestations of MEN-1 are excluded by a careful history asking about other family members, examining for lipomas, and measuring the serum calcium, prolactin, pancreatic polypeptide, and gastrin.

As insulinomas are usually benign, small, and uniformly distributed throughout the pancreas, the goal of surgery is precise identification and excision of tumor with preservation of normal pancreas and spleen. The pancreas is fully exposed to allow complete palpation and inspection. After dividing the gastrocolic ligament, a wide extended Kocher maneuver is performed, which includes mobilization of the right colon. In some instances the lateral peritoneal attachments of the spleen are incised to elevate the spleen and the tail of the pancreas to facilitate identification of a small tumor within the tail. The peritoneum along the inferior border of the gland is incised to allow palpation of the gland between the thumb and forefinger. The essential means of identifying the insulinoma is intraoperative ultrasound, which is performed with a high-resolution, near-field, real-time transducer (7.5–10 MHz). Doppler flow allows more accurate discrimination of the tumor and adjacent ducts, arteries, and veins. Insulinomas, as mentioned above, appear sonolucent compared to the more echo-dense normal pancreas. Masses/tumors are always imaged in two directions to record them in three dimensions. Intraoperative ultrasound can also be used to identify precisely the relations of the tumor to other vital structures such as the pancreatic duct, bile duct, arteries and veins, which allows safe excision. It has been used to facilitate the enucleation of nonpalpable insulinomas within the pancreatic head. Recent case reports have suggested that resection of insulinoma can be accomplished laparoscopically. The surgeon is positioned between the legs of the patient as for a Nissen fundolication. The pancreas is exposed using the harmonic scalpel to divide the gastrocolic ligament. Ultrasound is then used to identify the tumor, and it is enucleated using ultrasound for guidance. The procedure is identical to the open operation and laparotomy is done if the tumor is not identified.

Large, localized malignant tumors, which are uncommon (less than 5 per cent), or tumors that cannot be removed by enucleation because of their relation to the pancreatic duct or other vital structures, are resected by either distal pancreatectomy or Whipple pancreaticoduodenectomy. Blind distal pancreatectomy, a procedure that was formerly recommended, is no longer indicated because recent studies show that occult tumors are usually within the head of the pancreas, an area that is difficult to palpate. Many large series from different institutions have shown that more than 90 per cent of patients can have successful surgery and complete correction of the hypoglycemia with this approach. Further, because as much normal pancreas as possible is preserved, there is a low incidence of diabetes mellitus postoperatively. Patients with MEN-1 and insulinoma are treated in the same manner, except that a dominant, large islet-cell tumor is usually identified and resected. Complications of the excision of insulinoma are primarily those associated with uncontrolled pancreatic drainage, which include abscess, fistula, pseudocyst, or wound infection. Proper drainage of the pancreatic enzymes by closed suction usually prevents these complications (Table 3).

Tumor	*n*	*Tumor found; number and (%)*	*Cured; number and (%)*
Insulinoma	25	24 (96)	24 (96)
Gastrinoma	91	75 (82)	59 (64)

Data from: Doherty *et al. Surgery* 1991; *110*: 989–97; Norton *et al. Annals of Surgery* 1992; *215*: 8–18.

Table 3 Results of surgery for insulinoma and gastrinoma

Malignant insulinoma

Surgery in patients with malignant insulinoma may still be curative, but this can only be accomplished if all the tumor is completely removed. The primary tumor is excised by either Whipple pancreaticoduodenectomy or subtotal pancreatectomy–splenectomy, depending on the location. Secondary liver metastases are removed by either multiple wedge resections or lobectomy. Resection of both pancreatic and hepatic disease is indicated if the operative procedure planned can remove all gross disease with a prediction of acceptable mortality and morbidity. Resection of all identifiable tumor appears to improve survival and may even cure a subset of patients. It is important to remember that patients with liver metastases still have a long-term survival of over 20 per cent if left untreated; with complete resection, survival increases to approximately 80 per cent. Another indication for debulking surgery is amelioration of the signs and symptoms of hypoglycemia, especially in patients who do not respond adequately to medical management. This indication applies to patients with metastatic insulinoma who do not respond to octreotide, verapamil, diazoxide, or frequent feeding. Surgery may also be indicated in patients whose islet-cell tumor causes life-threatening symptoms such as gastrointestinal hemorrhage, or bile or intestinal obstruction. In these instances, the tumor may either be resected or bypassed to relieve the symptoms.

Chemotherapy for metastatic islet-cell tumors has had a 30 per cent partial response rate with no complete responses published. The active drugs include doxorubicin, 5-fluorouracil, and streptozotocin. These drugs have been used in combination and individually, with combination providing more responses. In Sweden, interferon-a has had reasonable response rates in patients who have failed chemotherapy. Recent studies have used long-acting analogs of somatostatin as antitumor agents to slow tumor growth and metastasis. These hormonal therapies have been effective at controlling medical signs and symptoms of hormonal excess and are indicated for that purpose, but the antitumor effects are less clear.

Chemoembolization, a combination of simultaneous occlusion of the hepatic artery and chemotherapy, has had some dramatic antitumor responses in individuals with large hepatic tumors. Symptoms have improved in nearly every patient and hormonal concentrations have significantly decreased in approximately 80 per cent of individuals, suggesting that this may have a real benefit as disease burden progresses. However, side-effects and complications may occur and some patients have developed liver abscesses. Alcohol injection, cryotherapy, and radiofrequency ablation have been used in a few patients with some benefit (Fig. 4).



Fig. 4. Flow diagram for the diagnosis and management of insulinoma.

Gastrinoma

In the initial description of the Zollinger–Ellison syndrome in 1955, Zollinger and Ellison included a triad of clinical findings: peptic ulceration of the jejunum, gastric acid hypersecretion, and an islet-cell tumor of the pancreas. The precise incidence of gastrinoma is unknown, although it is estimated that between 1 and 3 persons per million develop gastrinoma each year. The incidence is approximately 0.1 to 1 per cent of patients with peptic ulcer disease. Gastrinomas are the most common functional tumor of pancreatic islet cells in patients with MEN-1, and approximately 20 per cent of patients with Zollinger–Ellison syndrome will have MEN-1.

Symptoms

Sporadic Zollinger–Ellison syndrome occurs most commonly in the fifth decade of life. Patients with Zollinger–Ellison syndrome and MEN-1 are younger and usually present in the third decade. Men develop a gastrinoma twice as often as women. Most patients with Zollinger–Ellison syndrome complain of epigastric pain and indigestion caused by gastric acid hypersecretion. The typical patient has duodenal ulcers. Some have multiple ulcers, recurrent ulcers, and ulcers in unusual locations such as the jejunum. Perforated peptic ulcer occurs in approximately 10 per cent; conversely, 10 per cent of patients with Zollinger–Ellison syndrome never have any signs or symptoms of peptic ulcer disease. Secretory diarrhea occurs in 40 per cent of patients with the syndrome and is the presenting complaint in 20 per cent. The diarrhea is associated with increased intestinal transit time and malabsorption. Esophageal symptoms are also common. Most patients experience heartburn and dysphagia. Endoscopy demonstrates changes consistent with esophagitis, including erosion, inflammation, stricture, ulceration, and perforation. Each of these symptoms is related to acid hypersecretion, control of which completely removes them all.

Diagnosis

Measurement of the fasting serum gastrin is an essential study for the diagnosis of Zollinger–Ellison syndrome; patients with the syndrome will have an elevated serum gastrin (over 100 pg/ml) (Table 4). Gastric acid antisecretory medications must be discontinued for 1 week before testing; the most common cause of an elevated serum gastrin is iatrogenic achlorhydria secondary to medicines that inhibit acid output. Another essential study for the diagnosis is measurement of basal acid output. Patients with Zollinger–Ellison syndrome have an elevated basal acid output (more than 15 mEq/h in patients without previous surgery to reduce gastric acid output, or more than 5 mEq/h in those with prior acid-reducing operations. Measuring the output of gastric acid may exclude the diagnosis of achlorhydria, which is another cause of elevated serum gastrin. A less accurate but acceptable method for assessing acid output is to measure gastric pH; over pH 3 is inconsistent with the diagnosis of Zollinger–Ellison syndrome.

Cause of hypergastrinaemia	BAO	Serum (gastrin) response to secretin test	Test result (Hypergastrinaemia)
Gastrinoma	Elevated	> 200pg/ml Δ	↑
Hypochlorhydric states	-	↑	↑
G-cell hyperplasia	+	↑	↑↑
Retained antrum	+	↑	+
Short gut syndrome	+	↑	+
Gastric outlet obstruction	+	↑	+

BAO, basal acid output; ↑, no appreciable change.

Table 4 The differential diagnosis of hypergastrinemia

There are some rare clinical conditions that mimic Zollinger–Ellison syndrome: hypergastrinemia accompanied by gastric acid hypersecretion may occur in patients with massive small-bowel resection, gastric-outlet obstruction, antral G-cell hyperplasia, and retained gastric antrum. Each of the above may be differentiated from Zollinger–Ellison syndrome by a secretin stimulation test (see below). A fasting serum gastrin of more than 1000 pg/ml and a basal acid output of more than 15 mEq/h is diagnostic of Zollinger–Ellison syndrome, but, in 70 per cent of patients the fasting serum gastrin is between 100 and 1000 pg/ml, and the gastric pH is less than 3; in these patients a secretin provocative test is necessary to establish the diagnosis. After an overnight fast, 2 units/kg of secretin is given intravenously, and serum gastrin measured at 0, 2, 5, 10, and 15 min. Patients with Zollinger–Ellison syndrome have an increase in serum gastrin of more than 200 pg/ml over the basal concentration. The secretin test has also been used to detect recurrence following surgical excision of gastrinoma. Patients with gastrinoma will have an abnormal secretin test before they develop an elevated basal gastrin or disease that can be imaged. Radiographic imaging studies in these patients are not necessary unless the secretin test

becomes abnormal.

In summary, the diagnosis of Zollinger–Ellison syndrome is made biochemically by the finding of elevated fasting serum gastrin and basal acid output. The secretin stimulation test is used to confirm the diagnosis.

Pathology

Unlike insulinomas, which are commonly benign and located within the pancreas, gastrinomas are malignant and may be extrapancreatic; 80 per cent occur within the ‘gastrinoma triangle', which includes both the head of the pancreas and the duodenum. Primary gastrinomas are found within the submucosa of the wall of the duodenum (Fig. 5). They also occur at other sites including the pancreas, liver, stomach, ovary, and heart. Despite the fact that gastrinomas may metastasize to lymph nodes, primary gastrinomas may also occur in lymph nodes, as shown by long-term cure following excision of a lymph-node gastrinoma only. Primary gastrinomas to the left of the superior mesenteric artery may be more malignant than those within the gastrinoma triangle. Duodenal gastrinomas have a 60 per cent incidence of lymph-node metastases, which is similar to that of malignant pancreatic tumors, but pancreatic tumors have a higher probability of liver metastases, present in approximately one-third of patients at the time of diagnosis. Liver metastases are the feature limiting survival: patients without liver metastases do not die from the tumor; those with them have a 10-year survival of 30 per cent. Gastrinomas may disseminate to lungs, bone, heart, and adrenal glands. Malignant metastatic gastrinomas may secrete other hormones beside gastrin; those that secrete ACTH have an especially poor prognosis.

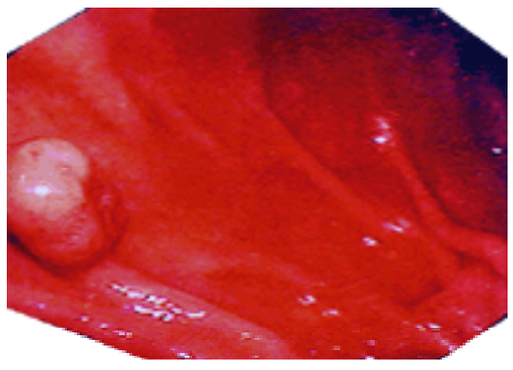


Fig. 5. Endoscopy of a gastrinoma in the second portion of the duodenum: a submucosal tumor is identified (left side of photograph); it measures 1 cm and is protruding into the lumen; it was later excised and found to be a gastrinoma.

Multiple endocrine tumors may occur within the pancreas and the duodenum in patients with MEN-1; immunohistochemical studies suggest that the gastrinomas commonly occur within the duodenum. The malignant potential of islet-cell tumors in patients with MEN-1 is similar to that of sporadic tumors. The tumors occur at a younger age in patients with MEN-1 and metastasize with a similar frequency.

Radiographic localization

For localization of gastrinoma, ultrasound has a low sensitivity of 20 to 30 per cent but its specificity approaches 92 per cent. CT detects approximately 50 per cent of primary and metastatic gastrinomas, but lesions of less than 1 cm are seldom identified. Approximately 30 per cent of gastrinomas of between 1 and 3 cm are seen on CT, while nearly all larger tumors are imaged. CT detects approximately 80 per cent of pancreatic gastrinomas, but only 35 per cent of extrapancreatic. Higher sensitivity is possible with newer techniques such as dynamic CT scanning, but small tumors may still be missed. MRI can demonstrate gastrinomas from an increased signal intensity during certain sequences. It is especially useful in identifying liver metastases and in distinguishing tumors from benign hemangiomas (Fig. 6), but, despite special sequences, MRI finds only 25 per cent of primary gastrinomas.



Fig. 6. MRI of gastrinoma: inversion recovery sequence of a scan in a patient with Zollinger–Ellison syndrome showing a primary tumor within the posterior head of the pancreas and a metastasis in the right lobe of the liver. Gastrinomas appear bright on this sequence; MRI is especially useful for showing liver tumors.

Somatostatin-receptor scintigraphy is the noninvasive imaging study of choice to localize both primary and metastatic gastrinoma, and is the most useful non-invasive imaging study in Zollinger–Ellison syndrome; as a very high proportion of gastrinomas have somatostatin receptors, approximately 80 to 90 per cent of tumors will be imaged (Fig. 7). It is as sensitive as all other imaging studies combined and alters the chosen management in 50 per cent of patients; it has 100 per cent positive predictive value so a positive study is also very specific for gastrinoma; it is able to image both primary and metastatic gastrinomas with equal efficiency, but it may miss small duodenal gastrinomas, so a negative scintigram does not exclude these.

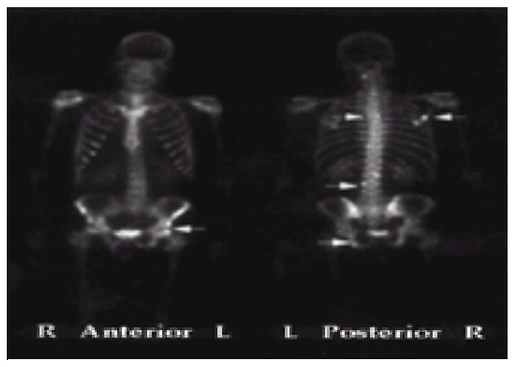


Fig. 7. Somatostatin-receptor scintigraphy in a patient who developed locally recurrent gastrinoma in lymph nodes; the study was done to exclude distant disease. The scintigram demonstrated numerous bony metastases (arrows), which precluded repeat surgery; the patient was treated with chemotherapy.

Endoscopic ultrasonography is a new method used to image gastrinomas presurgically. It is one of the best methods for identifying pancreatic gastrinomas, but it has difficulty with duodenal tumors. The ability to identify tumor is partially related to the background, which in the pancreas is homogeneous but in the duodenum inhomogeneous; therefore, small gastrinomas within the duodenal wall may be missed. Endoscopic ultrasonography has a sensitivity of 50 per cent, 75 per cent, and 63 per cent for duodenal, pancreatic, and lymph-node gastrinomas, respectively. It may be difficult to distinguish normal lymph nodes from those containing gastrinoma

because both appear sonolucent, and thus false-positive results may occur. Despite these potential problems, endoscopic ultrasonography has a role in the preoperative imaging of gastrinomas. As experience increases, so will an improvement in results.

Selective angiography and regional localization studies have been used in gastrinoma. Gastrinomas are visible on angiography as a blush within the pancreatic or liver circulation. Angiography is able to identify approximately 40 to 60 per cent of gastrinomas, both primary and metastatic. Regional localization by portal venous sampling for gastrin provides accurate localization in approximately 75 per cent of patients. Secretin, a secretagogue that is known to stimulate the secretion of gastrin from gastrinoma, has been combined with selective angiography in an attempt to obviate the need for portal venous sampling; the principle is the same as for the calcium angiogram in insulinoma (see above). The secretin angiogram is done by selectively injecting secretin sequentially into arteries that perfuse different regions of the pancreas and liver; shortly after the serum gastrin is measured in the hepatic vein. When the artery that perfuses the tumor is the one injected with secretin, the concentration of gastrin in the hepatic vein increases substantially and the gastri-noma is localized. This study combines the potential to image the tumor (as a vascular blush) with regional localization. Unlike insulinomas, which are uniformly distributed throughout the entire pancreas, most occult gastrinomas are found in the gastrinoma triangle, the area around the head of the pancreas and duodenum. Therefore, regional localization in patients with gastrinoma has usually suggested that the tumor is in the region of the head of the pancreas, which includes the duodenum. In these patients, invasive localization studies are not indicated.

Treatment

Because the symptoms of Zollinger–Ellison syndrome can now be controlled medically in all patients, total gastrectomy is no longer indicated to manage the gastric acid hypersecretion. Proper medical management requires precise measurement of gastric acid output and a dose of medication that will effectively suppress it. Patients with Zollinger–Ellison syndrome need larger doses of medication than those with peptic ulcer disease. Relief of symptoms is an unreliable indicator of the effectiveness of medical therapy. Gastric acid output must be maintained at 10 mEq/h before the next dose of medication, and at less than 5 mEq/h if there has been previous surgery to reduce acid secretion. Patients with esophagitis from acid reflux require acid outputs of less than 1 mEq/h.

Omeprazole and lansoprazole act by binding the H^+-K^+ ATPase on the luminal aspect of the parietal cell and dramatically inhibit gastric acid secretion. These drugs also have a longer duration of action than histamine-receptor antagonists. Patients are started on omeprazole 20 mg every 12 h. The dose is increased until the acid secretion is controlled; the mean total dose for acid control is 80 mg/day. Experimental studies have suggested that the long-term use of omeprazole is associated with the development of gastric carcinoid tumors, most likely due to chronic achlorhydria and hypergastrinemia as carcinoids have occurred with both ranitidine and omeprazole. This effect has been described in patients with Zollinger–Ellison syndrome in the context of MEN-1 who are maintained for long periods on omeprazole. Therefore, in these patients it is mandatory to perform periodic gastric endoscopy to exclude the development of carcinoid tumors.

The goal of surgery is complete resection of gastrinoma for possible cure of Zollinger–Ellison syndrome; this may be achievable in patients with both localized and metastatic tumor. Long-term follow-up demonstrates that the malignant tumoral process is the most important determinant of prognosis and survival. Therefore, all patients with sporadic gastrinoma and no contraindications to surgery should undergo laparotomy with the goal of potential cure and/or control of that process. Recent evidence suggests that resection of primary gastrinoma decreases the probability of liver metastases, which are the factor limiting survival, since patients with liver gastrinoma may subsequently die from Zollinger–Ellison syndrome. Resection of tumor has not as yet been shown to improve survival, but it is anticipated that, because surgery reduces the chance of spread to the liver, it will do so.

An upper abdominal incision is made. Adequate exposure of the entire abdomen is indicated because extrapancreatic, extraintestinal primary gastrinomas have been identified within the ovary, mesentery, liver, and stomach. After general exploration, the operative focus is primarily the liver, pancreas, and duodenum. Operative ultrasonography is performed on the liver and pancreas; gastrinomas appear sonolucent. Lymph nodes are removed as completely as possible from the pancreatic head, common bile duct, and celiac axis, even if they appear normal, as approximately 60 per cent of gastrinomas within the pancreas and duodenum have nodal metastases. The surgeon also concentrates on the duodenum with transillumination and duodenotomy. An endoscope is passed into the duodenum and the duodenal wall transilluminated; small duodenal gastrinomas are lucent. On palpation at the time of duodenotomy, gastrinomas feel like a firm nodule within the submucosa. Duodenotomy detects more gastrinomas than transillumination because tumors on the medial wall of the duodenum will be detected only by duodenotomy. Care must be taken to avoid injury to the pancreatic or bile duct. Catheters can be passed through the cystic duct to aid in identification of the common bile duct. Duodenal gastrinomas should be excised with a full-thickness wall of duodenum around the tumor. Tumors in the pancreatic head are enucleated and tumors within the body or tail are removed by distal pancreatectomy–splenectomy. Liver metastases can be wedge resected during the same procedure if the predicted morbidity and mortality are acceptable.

Surgery is indicated unless radiographic studies demonstrate unresectable metastatic disease (bilobar diffuse liver metastases or bony metastases; Fig. 7). Even when preoperative imaging studies fail to identify tumor, surgery is still indicated because gastrinoma will be identified during exploration in more than 90 per cent of these patients. With improved intraoperative methods and increased awareness that occult tumors are within the duodenum, gastrinomas are now found in nearly every patient. Occasionally, gastrinoma is found within a lymph node(s) only. Some suggest that a small primary tumor within the duodenum (microgastrinoma) has been missed, but, as outlined above, long-term follow-up indicates that some (40 per cent) of these patients remain disease-free, consistent with the notion that primary gastrinomas of lymph nodes really do exist. Overall, surgery produces disease-free survival in approximately 60 per cent of patients (range 9 to 100 per cent) with sporadic Zollinger–Ellison syndrome, and actuarial survival is excellent. Approximately one-half of patients who are initially disease free will experience a recurrence, with a long-term cure rate of 30 per cent. The expertise of the surgeon appears to be the single, most important factor in achieving a good surgical outcome, as small tumors, especially in the duodenum, may be missed. When patients are not cured by surgery and have localized gastrinoma on imaging studies, repeat operations are indicated; surgery will again render approximately 30 per cent disease free in the long term. All patients should be carefully followed for recurrent disease following excision of gastrinoma. The most sensitive method is the secretin test. If an abnormal test develops, somatostatin-receptor scintigraphy and CT should be performed. Table 3 summarizes the results of surgery for patients with Zollinger–Ellison syndrome. When all tumor is resected or no tumor is identified at careful laparotomy, the 10-year survival is greater than 85 per cent. However, for incompletely resected, unresectable gastrinoma or recurrent tumor, the 5-year survival ranges from 14 to 95 per cent (mean 43 per cent), and the 10-year survival is 25 per cent (Table 3).

Metastatic gastrinoma

Metastases occur in 25 to 90 per cent of patients with Zollinger–Ellison syndrome, and are the most common cause of mortality and morbidity. Although gastrinoma is indolent and some patients can do well for long periods with distant metastases, the overall 5-year survival for those with metastases is less than 40 per cent. Recent studies suggest that, on careful follow-up and repetitive imaging studies, 26 per cent of patients will have liver metastases that do not increase in size and do not need any antitumor treatment. However, 42 per cent of patients with liver metastases will experience rapid tumor growth and will eventually die from tumor. Clinical and laboratory features that suggest a rapid growth rate include a raised serum gastrin, the presence of diffuse, bilobar liver metastases, and bony metastases. The most effective chemotherapeutic regimen is a combination of doxorubicin, 5-fluorouracil, and streptozotocin. This regimen has a 40 per cent response rate, but no complete responses have been described and survival is not prolonged. Hepatic arterial chemoembolization with an emulsion of iodized oil, 5-fluorouracil, and streptozotocin has been used with some efficacy, but the effect on survival is unknown and the associated complications may be significant. Octreotide has been used as an antitumor agent, with minimal results. Interferon-a is associated with flu-like symptoms and chronic fatigue; its antitumor effects have been minimal.

Paradoxically, aggressive surgery to debulk and remove distant metastases of gastrinoma is a very effective treatment, shown to prolong survival. Resected patients have a 5-year survival of 60 per cent compared to 38 per cent in patients with inoperable disease. Further, some of these patients can still be rendered completely disease-free. However, long-term follow-up studies suggest that at 10 years most will have developed biochemical evidence of recurrent disease. Aggressive resection of localized metastases is indicated in the patients in whom most or all tumor can be removed (Fig. 8).

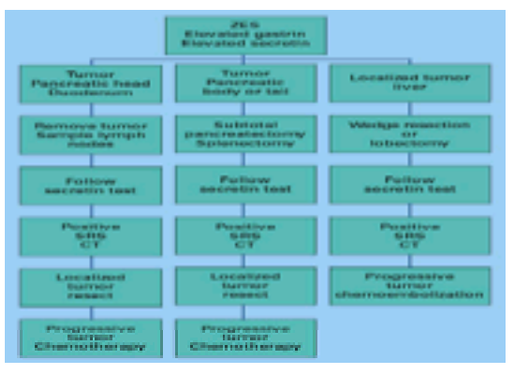


Fig. 8. Flow diagram for the diagnosis and management of gastrinoma.

Other functioning islet-cell tumors

Glucagonoma

Glucagonoma is a malignant tumor that usually presents in the fifth or sixth decade of life. Patients have a characteristic raised, red, itchy rash called necrolytic migratory erythema. The rash is typically seen on the lower extremities, and in intertriginous and perioral areas. It is initially erythematous and scaly, but can progress to bullous lesions that slough. The patients also have hypoaminoacidemia, weight loss, type 2 diabetes mellitus, severe muscle wasting or cachexia, and a high probability of deep venous thrombosis and pulmonary embolism.

Decreased plasma concentrations of amino acids and elevated concentrations of glucagon (over 500 pg/ml) are found. Because the cachexia is so severe, most patients are initially managed by total parenteral nutrition with supplemental zinc, trace metals, and insulin. This regimen improves the hypoaminoacidemia, malnutrition, serum zinc, rash, and functional status for surgery. Long-term management of patients with metastatic glucagonoma has relied on the use of octreotide, which can reduce the circulating plasma glucagon and improve the rash and malnutrition. Glucagonomas are usually located within the pancreas, and are metastatic and unresectable at the time of diagnosis.

Vasoactive intestinal peptide tumor (VIPoma)

The VIPoma syndrome is also called the pancreatic cholera syndrome, the Verner–Morrison syndrome, or the WDHA (*w*atery *d*iarrhea, *h*ypokalemia, and *a*chlorhydria) syndrome. VIPomas produce severe secretory diarrhea that causes hypokalemia, hypochlorhydria, hypovolemia, and dehydration. Patients with this tumor commonly have 5 to 10 l of stool per day. The diarrhea will persist despite depriving them of oral intake. These patients can complain of abdominal cramping, weakness, and flushing; the weakness is caused by hypokalemia and dehydration. They may also have hypercalcemia. Most VIPomas arise within the pancreas but extrapancreatic tumors have been described. Elevated serum pancreatic polypeptide has been associated with pancreatic but not extrapancreatic VIPomas. The diagnosis of VIPoma is made when the fasting plasma VIP is greater than 500 pg/ml in the presence of secretory diarrhea. Because the dehydration and electrolyte abnormalities are severe, it is necessary to correct them before surgery. Before octreotide was available, patients required large amounts of fluid, potassium, and chloride. Octreotide dramatically reduces the secretory diarrhea, making presurgical correction of the electrolyte abnormalities and dehydration safe and easy (see Management below).

Somatostatinoma

Somatostatinomas are very rare malignant tumors that arise in either the duodenum or pancreas. The somatostatinoma syndrome includes steatorrhea, cholelithiasis, type 2 diabetes mellitus, and hypochlorhydria. Most patients have advanced metastatic tumors and thus also have weight loss. They may present with abdominal pain secondary to the cholelithiasis. Because the symptoms are subtle and may be associated with other conditions, patients with somatostatinoma have locally advanced or distant metastatic disease at the time of diagnosis. Some somatostatinomas have been diagnosed incidentally at the time of cholecystectomy. Somatostatin-like activity can be measured by immunologic assays and is essential to the diagnosis.

Nonfunctioning islet-cell tumors

Nonfunctioning islet-cell tumors or pancreatic polypeptide-producing tumors (PPoma) do not have a clinical syndrome related to excessive hormone secretion. They are usually malignant and large at the time of diagnosis, and produce symptoms secondary to their mass effects. Some patients with MEN-1 or Von Hippel–Lindau syndrome may undergo screening for these tumors, diagnosed by an elevated serum pancreatic polypeptide; this finding necessitates a CT scan of the pancreas, which identifies a large tumor. The large tumors may cause symptoms of extrahepatic bile-duct obstruction, intestinal bleeding secondary to invasion of a major vessel within the gut, or intestinal obstruction secondary to direct invasion. Patients may also have hepatic metastases.

Rare islet-cell tumors

Islet-cell tumors can produce a variety of unusual hormones including growth hormone-releasing factor, ACTH, parathyroid hormone-related peptide, neurotensin, and serotonin. Carcinoid tumors that secrete serotonin are rarely localized to the pancreas and usually occur within the small intestine or appendix. Islet-cell tumors have been found to cause severe hypercalcemia secondary to parathyroid hormone-related peptide, Cushing syndrome secondary to ectopic ACTH, and acromegaly secondary to growth hormone-releasing factor. Rarely, such tumors produce neurotensin, which may cause diarrhea, hypotension, and flushing.

Medical management

Medical therapy can be used to control the signs and symptoms of excessive hormonal secretion; it may also be used to try to control the tumoral process but it is not curative. As outlined, patients with glucagonoma often have severe malnutrition and need intense nutritional management before surgery. The malnutrition can be corrected by complete (glucose, fat and amino acids) intravenous feeding with added trace elements (zinc) and insulin (for the type 2 diabetes). Interestingly, we have demonstrated that correction of the hypoaminoacidemia will ameliorate their rash. The common deep venous thrombosis and pulmonary embolism can be prevented with either a Greenfield filter or heparin.

Patients with somatostatinoma have gallstones, so cholecystectomy is recommended at the time of surgery. In patients with VIPoma, as outlined, the severe dehydration and electrolyte abnormalities usually require intravenous fluids for presurgical correction. The long-acting somatostatin analog octreotide decreases serum VIP, stops secretory diarrhea, and dramatically simplifies management. Severe hypercalcemia secondary to a parathyroid hormone-related peptide-producing tumor must be controlled by the use of intravenous normal saline, lasix, mithramycin, and diphosphonates. Cushing syndrome from ACTH-producing islet-cell tumors may be controlled by drugs such as ketoconazole, aminoglutethimide, and RU 486. However, medical control of the hypercortisolism is most often inadequate and these patients require bilateral adrenalectomy for control of Cushing syndrome if complete resection of the ACTH-producing tumor is impossible. Serotonin-producing islet-cell tumors (carcinoid syndrome) require octreotide as a premedication at the time of surgical resection in order to prevent the occurrence of the carcinoid crisis.

Tumor localization

Since these tumors are generally malignant, they are not occult and will be visible on CT or MRI. Some studies have demonstrated that MRI is more sensitive for detection of liver metastases than CT. CT may be preferred for imaging the primary tumor, because CT images have clearer anatomic resolution than MRI. Further, CT is less expensive. Somatostatin-receptor scintigraphy should also be used to detect primary tumors and metastases, and to determine the utility of octreotide therapy. Approximately 80 to 90 per cent of these rare islet-cell tumors will image on somatostatin-receptor scintigrams.

Surgical therapy

Cure or control of the tumoral process can only be accomplished if all tumor is completely removed. These tumors are generally more malignant and larger than the common types, such that they are resected by either pancreaticoduodenectomy or subtotal pancreatectomy–splenectomy, depending on their primary location. Secondary liver metastases are removed at either the same time or subsequently by wedge resection or lobectomy, depending on their location and size. Resection of all identifiable tumor appears to improve survival and may cure approximately 20 to 30 per cent of patients. These operations must be done safely as the long-term survival without treatment is excellent. Patients with untreated liver metastases have a 5-year term survival of 20 per cent.

Another indication for surgery is to reduce the signs and symptoms of excessive hormone secretion, especially in patients who do not have a good response to medical management. This indication may apply to patients with VIPoma and glucagonoma who do not respond to octreotide. Reduction of tumor mass will usually reduce circulating concentrations of hormones and symptoms. Surgery may improve the symptoms or it may reduce the medication required for symptom control. In some instances, such as ACTH-secreting islet-cell tumor, surgical resection of the target organs (both adrenals) may substantially improve the quality and duration of life. Surgery may also be indicated in nonfunctional islet-cell tumors that develop life-threatening symptoms from tumor mass such as gastrointestinal hemorrhage and biliary or intestinal obstruction. In these conditions, the tumor may either be removed or bypassed to relieve the symptoms.

Chemotherapy, and chemoembolization by a combination of simultaneous hepatic arterial occlusion and doxorubicin chemotherapy, has had dramatic antitumor responses in some patients with large hepatic tumors (see under Malignant insulinoma above). Symptoms have improved and hormonal concentrations significantly decreased, but side-effects and complications may occur and some patients have developed liver abscesses. Alcohol injection and cryotherapy have been used in a few patients without clear benefit. Recently, radiofrequency ablation of islet-cell tumors within the liver has been successful at improving hormone concentrations and reducing tumor volume. Durable long-term responses will be necessary before this method can be recommended unequivocally.

Further reading

Alexander HR *et al.* Prospective study of somatostatin receptor scintigraphy and its effect on operative outcome in patients with Zollinger–Ellison syndrome. *Annals of Surgery* 1998; **228**(2): 228–37. [This prospective study of 35 consecutive patients with Zollinger-Ellison syndrome concluded that somatostatin receptor scintigraphy was the most sensitive preoperative imaging study for extrahepatic gastrinomas.]

Amikura K *et al.* Role of surgery in management of adrenocorticotropic hormone-producing islet cell tumours of the pancreas. *Surgery* 1995; **118**(6): 1125–30. [This retrospective chart review of 12 patients with ectopic adrenocorticotropic hormone-producing islet cell tumors of the pancreas revealed that this aggressive tumor may benefit from palliative bilateral adrenalectomy because biochemical cures have been reported after aggressive surgical resection.]

Arcenas AG, Ajani JA, Carrasco H, Levin B, Wallace S. Vascular occlusive therapy of pancreatic endocrine tumours metastatic to the liver. *Frontiers in Gastrointestinal Research* 1995; **23**: 439–50. [Textbook reviewing vascular occlusive techniques as one method of treating metastic pancreatic endocrine tumor to the liver.]

Arnold WS, Fraker DL, Alexander HR, Weber HC, Norton JA, Jensen RT. Apparent lymph node primary gastrinoma. *Surgery* 1994; **116**(6): 1123–9; discussion, 1129–30. [Resection of apparent lymph node primary gastrinoma lead to a 43 per cent disease free state with a median 5.3 year follow-up period.]

Cohen MS, Picus D, Lairmore TC, Strasburg SM, Doherty GM, Norton JA. Prospective study of provocative angiograms to localize functional islet cell tumours of the pancreas. *Surgery* 1997; **122**(6): 1091–100. [Provocative angiography for insulinoma (*n* = 4) and Zollinger–Ellison syndrome (Zollinger–Ellison syndrome *n* = 8) is the preoperative localization study of choice because there are few false-positive and false-negative results.]

Doherty GM, Shawker TH, Eastman RC, Gorden P, Norton JA. Results of a prospective strategy to diagnose, localize and resect insulinoma. *Surgery* 1991; **110**: 989–97. [The diagnostic workup of insulinoma includes a supervised fast, with portal venous sampling being the most sensitive preoperative localization technique, and US being an essential tool for intraoperative detection.]

Fraker DL, Norton JA, Alexander HR, Venzon DJ, Jensen RT. Surgery in Zollinger–Ellison syndrome alters the natural history of gastrinoma. *Annals of Surgery* 1994; **220**(3): 320–8; discussion, 328–30. [Surgical exploration with attempted curative gastrinoma resection resulted in a decrease in the incidence of hepatic metastases (3 per cent (3/98)) compared with medical management (23 per cent (6/26)).

Gibril F *et al.* Somatostatin receptor scintigraphy: its sensitivity compared with that of other imaging methods in detecting primary and metastatic gastrinomas. A prospective study. *Annals of Internal Medicine* 1996; **125**(1): 26–34. [See comments.] [Somatostatin receptor scintigraphy is the single most sensitive (92 per cent) method for imaging primary or metastatic liver lesions in patients with Zollinger–Ellison syndrome.]

Grant CS, van Heerden J, Charboneau JW, James EM, Reading CC. Insulinoma. The value of intraoperative ultrasonography. *Archives of Surgery* 1988; **123**(7): 843–8. [Review of current localization techniques in 36 patients who were surgically treated for insulinoma revealed that intraoperative ultrasound can enhance intraoperative identification of the tumor.]

Jaskowiak NT, Fraker DL, Alexander HR, Norton JA, Doppman JL, Jensen RT. Is reoperation for gastrinoma excision indicated in Zollinger–Ellison syndrome? *Surgery* 1996; **120**(6): 1055–62; discussion, 1062–3. [Reoperation for recurrent or persistent gastrinoma resulted in a disease free state in 30 per cent of the patients and therefore should be considered for patients with imageable disease.]

Krenning EP *et al.* Somatostatin-receptor scintigraphy in gastroenteropancreatic tumours. *Annals of the New York Academy of Sciences* 1994; **733**: 416–24. [Review of the role of somatostatin-receptor scintigraphy in the localization of pancreatic endocrine tumors.]

Norton JA, Doppman JL, Jensen RT. Curative resection in Zollinger–Ellison syndrome: results of a 10-year prospective study. *Annals of Surgery* 1992; **215**: 8–18. [This prospective study of 73 patients with Zollinger–Ellison syndrome suggests that patients with the localized sporadic syndrome can have the gastric hypersecretion managed medically. Seventy-eight per cent of them are able to have complete gastrinoma resection.]

Rosch T *et al.* Localisation of pancreatic endocrine tumours by endoscopic ultrasonography. *New England Journal of Medicine* 1992; **326**(26): 1721–6. [See comments.] [Endoscopic ultrasound can be highly sensitive (82 per cent) and specific (95 per cent) in the localization of pancreatic endocrine tumours.]

Sugg SL *et al.* A prospective study of intraoperative methods to diagnose and resect duodenal gastrinomas. *Annals of Surgery* 1993; **218**(2): 138–44. [Prospective comparison of the ability of palpation, IOUS, IOE, and duodenotomy to detect gastrinomas in 35 patients with Zollinger–Ellison syndrome found that duodenum is the most common location for gastrinoma and that duodenotomy should be routinely performed in all explorations of patients with the syndrome.]

Weber HC *et al.* Determinants of metastatic rate and survival in patients with Zollinger–Ellison syndrome: a prospective long-term study. *Gastroenterology* 1995; **108**(6): 1637–49. [Prospective follow-up of 185 Zollinger–Ellison syndrome patients found that the presences of liver metastases is the primary determinant of survival.]

34.1 The acute abdomen

Julian Britton

[Introduction](#)
[Anatomy](#)
[Physiology and pathology](#)
[Clinical diagnosis](#)
- [History](#)
- [Examination](#)
- [Investigations](#)

[Making a diagnosis](#)
- [The classic case](#)
- [A question of elimination](#)
- [Anatomy and pathology](#)
- [Age and sex](#)
- [Statistics](#)
- [The pattern of disease](#)
- [Listing the possibilities](#)
- [A repeat visit](#)
- [The return visit](#)
- [Computer assistance](#)
- [The pragmatic approach](#)
- [The use of analgesia](#)
- [Spot diagnosis](#)

[Learning to make a diagnosis](#)
- [Recording the facts](#)
- [Always make a diagnosis](#)
- [Reporting to a senior](#)
- [Review the analysis](#)

[The differential diagnosis](#)
- [Inflammation](#)
- [Obstruction](#)
- [Colic](#)
- [A major catastrophe](#)
- [A mass](#)
- [Specific diagnostic problems](#)
- [Unusual problems](#)

[Conclusion](#)
[Further reading](#)

Introduction

Acute disease within the abdomen is common and many patients with abdominal symptoms present every day to doctors working in the community. Within a Western population of half-a-million people, between 5 and 10 patients are admitted to a surgical ward each day with acute abdominal pain. One or two more will complain of acute abdominal symptoms after an accident. By definition the illness starts suddenly and most patients present to a hospital within 7 or 10 days of the onset of symptoms. In the majority of patients, symptoms arise from disease within the abdominal cavity itself, but occasionally they originate elsewhere in the body.

The range of disease extends from the relatively trivial to the immediately life threatening, and attempts to reach a diagnosis must sometimes be curtailed in the interests of immediate treatment. More commonly there is time to take a history, to examine the patient, and to organize the investigations that will be helpful in establishing a diagnosis and planning treatment. Accurate recording of the relevant facts is vital, and a clear understanding of the anatomy and pathophysiology of intra-abdominal disease is necessary for both diagnosis and treatment. These patients are therefore ideal for training junior members of a surgical team.

Some patients require early surgery. This itself varies from a simple, straightforward procedure to a highly complex operation that stretches the ability and skill of even the most experienced surgeon. The immediate feedback that an emergency operation provides on the accuracy and the adequacy of the preoperative assessment and preparation is another reason why the patient with an acute abdomen is an important part of surgical training.

Anatomy

A good knowledge of normal and abnormal abdominal anatomy, and particularly surface anatomy, is essential. Variations within and between individuals are obvious, but normal anatomy also changes with age, posture, respiration, disease, and previous surgery. Nevertheless, with experience most surgeons carry a remarkably accurate mental picture of the expected internal position of any particular organ in any particular patient.

The embryological development of the abdomen is relevant in two respects. The intestine and all its associated organs such as the liver and the pancreas develop initially as midline structures. Thus visceral pain is usually felt along the midline of the abdomen. The gut also has a segmental origin so that the division into foregut, midgut, and hindgut exactly correlates with the vascular supply, and, correspondingly, pain is felt in the epigastrium, the umbilical area, and the hypogastrium. Certain congenital abnormalities can predispose to acute abdominal complications.

In contrast to the visceral peritoneum, the parietal peritoneum is innervated by somatic nerves. Pain is therefore accurately localized to the site of irritation of the abdominal wall and is accompanied by a reflex contraction of its muscles. This applies both to the anterior and the posterior abdominal walls. Psoas spasm from acute appendicitis and a scoliosis concave to the side of intra-abdominal inflammation are two good examples. Inflammation confined to the pelvis may not, however, be accompanied by spasm of anterior abdominal muscles and this may cause clinical confusion. This is because the somatic nerves that supply the organs in the pelvis do not supply the muscles of the anterior abdominal wall.

When describing the findings of abdominal examination the surface is best divided into six areas by a transverse line going through the umbilicus and longitudinal lines running through the tip of the ninth rib on each side (Fig. 1). Thus there are epigastric and hypogastric areas in the middle, and an iliac fossa and hypochondrium laterally. It is often also useful to describe a periumbilical area. How-ever, it is important to realize that none of these divisions has a true anatomical basis.

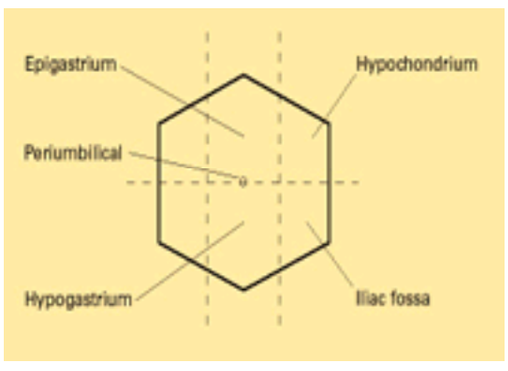


Fig. 1. The descriptive areas of the abdominal wall.

Physiology and pathology

Normal physiology is rapidly disrupted by the onset of acute intra-abdominal disease. Many patients vomit, and gastrointestinal secretion, absorption, and motility all change in the presence of obstruction, luminal infection, or peritonitis. Urine is reduced in volume and altered in content, usually secondary to redistribution of fluid in the body compartments but sometimes because of a direct toxic effect on the kidneys.

The mediation of abdominal pain is not well understood. It is perfectly possible to handle the intra-abdominal organs and even divide the bowel of a conscious patient without causing any pain. However, distension or stretching of the bowel wall is accompanied by reflex contraction of the smooth muscle in the wall, which is immediately painful. This may be due to transient ischaemia of the muscle. The pain fibres run with the splanchnic sympathetic nerves to the spinal column, where they are distributed segmentally. The pain is localized to the abdominal cavity but not to the precise segment of bowel that is being stretched. Other pathways within the spinal column are also stimulated, and vomiting, which is a common accompaniment of severe pain, can also be centrally mediated.

The gastrointestinal tract is a significant source of a wide variety of hormones. These change in response to acute disturbances of function but whether this is a primary or a secondary effect is not yet clear. Inflammation is the most common cause of acute pathology within the abdomen, followed by obstruction, haemorrhage, trauma, and ischaemia.

Bacteria, viruses, fungi, parasites, and chemicals can all cause inflammation: bacteria from the bowel, such as *Escherichia coli, Streptococcus faecalis*, and various anaerobes, are by far the most important. Other bacteria that cause acute abdominal pain are *Salmonella* and *Shigella* spp., *Yersinia*, and *Campylobacter*. Acute inflammation normally develops into clinical significance over hours rather than minutes or days, and progression either to suppuration or resolution also takes time. Perforation and ischaemia develop in minutes and cause very acute symptoms. Resolution, whatever the underlying pathology, always takes longer than development. Neoplasia, neurogenic, and metabolic disorders occur less commonly but they are all well-recognized causes of acute abdominal pain.

Some of these pathological processes are closely interlinked. There are a number of causes of intestinal obstruction, of which neoplasia is one. Peritonitis from perforation of the bowel into the potential peritoneal space usually arises from local ischaemia, but this may in turn be caused by inflammation or obstruction that has progressed to strangulation. The clinical presentation and the physiological consequences of obstruction or peritonitis may be similar whatever the cause, but a careful history and examination should enable the underlying pathology to be discerned.

Clinical diagnosis

Most patients with an acute abdomen can be managed using simple clinical skills (Table 1). An accurate history and a thorough examination are often sufficient to make a diagnosis and recommend treatment; modern investigations can help and may reassure the anaesthetist that the patient is fit for an operation. The primary objective when the patient and the doctor first meet is, therefore, to elicit the symptoms and the signs necessary to make a rapid and accurate diagnosis. It is sometimes obvious that the patient is in severe pain or seriously ill. The necessary immediate treatment must then take precedence over making a diagnosis.

	Percentage of cases
Non-specific abdominal pain	34
Acute appendicitis	28
Acute cholecystitis	10
Small-bowel obstruction	4
Acute gynaecological disease	4
Acute pancreatitis	3
Renal colic	3
Perforated peptic ulcer	2
Cancer	2
Diverticular disease	1
Miscellaneous	9

Table 1 Causes of acute abdominal pain seen in hospitals in the developed world (after de Dombal, 1991)

Unfortunately, even the most experienced clinicians only make a correct clinical diagnosis of acute abdominal pain on four occasions out of five; younger doctors and those who practise in the community are only right half the time. Many attempts have been made to improve on these results and one method that has attracted much attention is computer-assisted diagnosis. By a curious coincidence this has simply taught us once again that taking an accurate history and examining the patient carefully are still the most important factors in making a correct diagnosis.

History

Many patients will make their own diagnosis as one listens to their story: the art of taking a history is to induce every patient to do so. Doctor and patient have not usually met before, and the style and the approach of the doctor really do matter. A relaxed, confident manner and a smile always help, and you must make it absolutely plain to the patient that they have your complete attention and that you have plenty of time to listen, even if this is not so. You should discourage interruptions by other members of staff or requests to answer the telephone.

Patients like to be treated as individuals. Go and sit by their bedside knowing their name and introduce yourself clearly with your own name. Some patients will immediately start to describe their symptoms and must be left to continue. Others look for a cue from the doctor. Simple, non-specific questions such as ‘what has happened?' or ‘why have you come to hospital?' are best. Some will then give their history spontaneously; others reply in only a few words and then need prompting again. It is occasionally better initially to engage the patient in conversation about something entirely unrelated, such as their job or their family, and then when they are relaxed lead the discussion back to the acute problem. This is particularly useful with very anxious patients. The most difficult patient is the one who is garrulous about everything but the reason they have come for help. Often there is nothing for it but to stop the flow of words deliberately and redirect the patient to the current problem. It is difficult to do this without appearing rude or disinterested: beware of the temptation to assume that there is little wrong with these patients. They are sometimes simply frightened.

Most patients come to the end of their story spontaneously, and sometimes they have told you everything you need to know in perfect order. Never intervene to clarify a point of detail but do stop the `atient ghen thepinformation they offer becomes irrelevant: it is important not to overload the brain with too many facts. When the patient has finished there will usually be some points that need amplifying or some further information that is essential. This is best obtained by asking direct, but not leading, questions. It is very easy indeed to suggest the answer you want either by the words you use, your facial expression, or the manner in which you speak or behave. If you do this the answers will be unreliable. Asking questions is also an art that requires tact and skill. Short, specific questions are best, and they must be phrased clearly without using jargon and in language the patient understands. Some patients, like most politicians, do not answer the question they are asked. You should insist, politely, on a specific answer if one is possible. No two doctors ever obtain exactly identical histories: a young surgeon may be amazed to hear a patient give a totally contradictory reply to an apparently identical question from a senior colleague. It is also surprising how often it is the very last thing the patient says that clinches the diagnosis.

Not everyone can give a history themselves. Most children are shy or frightened, although others, even the very young, sometimes tell a perfect story. The confused and the mentally handicapped are often unreliable as regards facts, while the memory of an elderly patient who is ill is often faulty. A relative or a friend must then relate the history, but the clinician should remember that his or her personality then intervenes. This is a particular problem if the patient is foreign and the history has to be taken through an interpreter.

Complete attention to the patient and absolute concentration on everything he or she says and how it is said are essential. Observation of the patient is slightly different from inspection during the examination. It encompasses demeanour as well as an assessment of personality, mood, and reaction to the illness. Movement, particularly expressive movement of the hands, is always useful. Patients with peritonitis lie quite still and look ill, patients with colic really do roll around, and patients with cholelithiasis often describe the pain radiating round into the flanks with their hands, for example. Obvious and significant physical signs such as gross abdominal distension with audible borborygmi, jaundice, or the smell of melaena should not be ignored: they all point to a specific pathology that may be confirmed by specific questions.

Allowing the patient to talk freely does not prevent recording the facts in a systematic fashion. In most hospitals this has to be done freehand but there are advantages in specially designed forms. The information is recorded systematically, and omissions are obvious and can be corrected at once. Such forms also require the clinician to be specific about the features of certain symptoms.

Pain

Most patients admitted with an acute abdominal problem complain of abdominal pain. Cope, in his classic book, observed that acute pain lasting for more than 6 h in a previously fit patient usually has a surgical cause. It is also a most important symptom: detailed enquiry about the nature of the pain will often indicate the correct diagnosis.

Site

The first thing to establish is the precise site of the pain that the patient has now. Some patients are extraordinarily obtuse about this, partly because they have difficulty in answering and partly because they often do not understand why you want to know. It is best to ask the patient to point with one finger to where the pain is worst and to record this site in the notes. Those who wave a hand vaguely everywhere probably do not have too much wrong with them. Pain often moves during the course of an illness and it is then worthwhile asking where the pain was situated at the beginning.

Radiation

Radiation of the pain to other parts of the body is often diagnostic. Radiation of the pain to the testicle in ureteric colic, to the shoulders in acute cholecystitis, and to the knee with an obstructed obturator hernia are specific and typical examples. Sometimes patients volunteer that a pain radiates elsewhere but more commonly it is necessary to ask directly.

Onset

Some patients can say exactly when the pain started. They may be able to give a time or say what they were doing. This always suggests a significant cause and an acute pathological process, such as perforation or strangulation. Pain that wakes the patient up at night is also significant, although it is not often possible to describe the acuteness of onset. Sometimes pain is not the first symptom the patient noticed and this may suggest a medical cause, as with the vomiting from gastroenteritis or the marked anorexia of hepatitis. The duration of the illness gives some idea how far any pathology may have progressed and this can be correlated with the findings on clinical examination.

Some patients relate the onset of their pain to an injury. Apparently mild trauma is occasionally followed by serious intra-abdominal injury; on the other hand it is more common for patients, after the onset of the symptoms, to try and relate them to an injury. This can be dangerously misleading, as with acute testicular torsion for example.

Frequency

There are two aspects to frequency. Alterations in the pain since this episode started are useful pointers to the immediate diagnosis, whereas pains that have come and gone in a similar way in the previous weeks or months suggest a longer term and more chronic disease process. Variations in intensity in the short term can be classified into two types. Either the pain is constant or it comes and goes. If it comes and goes with some degree of regularity, it is colic. Constant pain is associated with inflammatory conditions and colicky pain with distension of smooth muscle, as described below.

Aggravation and alleviation

Any movement makes the pain of peritonitis worse, while lying still makes it better. Acute exacerbation of the pain on walking, breathing, coughing, or going over a bump in the road on the ride in to the hospital is equivalent to rebound tenderness on examination. Pain in the shoulder on lying down comes from diaphragmatic stimulation by an irritant fluid. The fluid is often blood from an intra-abdominal injury or an ectopic pregnancy. Analgesics usually make the pain better; this can be deceiving. Sometimes vomiting temporarily relieves the pain of obstruction.

Severity and type

Pain is a very subjective symptom and people's reaction to it varies widely. Accompanying signs such as sweating and tachycardia give the observer some idea of severity, but this only establishes that there is something wrong with the patient, which is often perfectly obvious anyway. Most patients find it very difficult to describe the nature of their pain and require prompting. No particular diagnoses are suggested by such descriptions as boring, dragging, sharp or dull, and they are best avoided.

Nausea and vomiting

These are two quite separate symptoms and both are useful in diagnosis. Nausea may precede vomiting but it need not do so and neither does vomiting always follow nausea. Nausea by itself is a less specific symptom, although it is a common accompaniment of gallstone disease. Anorexia is a separate and somewhat non-specific symptom since most people, and particularly children, lose their appetite when they are unwell. Pain normally precedes vomiting in surgical disease of the abdomen whereas the reverse is often the case in medical conditions.

Vomiting is a classic symptom of intestinal obstruction and it usually accompanies colic. Vomiting often occurs after a bout of pain in obstruction and the shorter the interval between the two the higher the obstruction. The vomit itself is initially green in colour but turns yellow and then frankly faecal as the obstruction persists. Retching without vomiting suggests acute torsion of an intra-abdominal structure.

Vomiting does not often accompany perforation of a peptic ulcer or intra-abdominal haemorrhage, and it is a late event in distal obstruction of the large bowel if it occurs at all. Nausea and anorexia are more common than vomiting in appendicitis.

Bowel function

Diarrhoea and constipation are two potentially confusing symptoms because they mean different things to different people. It is important first to establish the patient's normal bowel habit and the normal consistency of the stool, and then to decide if there have been any recent changes. Diarrhoea to some people simply means frequent defaecation of normal faecal material, whereas repeated loose watery stools are of greater interest to the surgeon. When true diarrhoea is present it is important to establish whether other members of the household are afflicted. The presence of blood, slime, or the black, tarry stools of melaena are all of obvious diagnostic value. If intestinal obstruction is suspected, then failure to pass wind as well as stool is important.

Gynaecological symptoms

Symptoms arising from the uterus, fallopian tubes, and ovaries are a common reason for admission to hospital with acute abdominal pain. Furthermore, the ‘negative laparotomy' rate is highest in young women. Questions about normal and abnormal menstrual function, vaginal discharge, and the risk of pregnancy are therefore essential. Tact and sensitivity are required, but the answers really do matter: a ruptured ectopic pregnancy is a potentially lethal condition.

Urinary symptoms

Alterations in the pattern of micturition suggest disease of the urinary tract. Frequency is linked with inflammation, while anuria is most commonly caused by acute retention in elderly men. Pain on passing urine must be separated into two classes. Abdominal pain exacerbated by micturition suggests irritation of the peritoneal surface of the bladder, while stinging pain in the urethra on urination is characteristic of infection. Patients should also be asked about the colour of the urine and the presence of blood or pus. Dysuria is a symptom that means different things to different doctors, and the term should not be used without specifying what is meant.

Past history

Any previous medical problem may be relevant to the cause of an acute admission for abdominal pain and it will certainly be relevant to the management. Chronic indigestion can be a useful pointer to a possible cause of peritonitis. A past history of abdominal surgery is important because adhesions have now overtaken hernias as the most common cause of intestinal obstruction. Patients often report previous episodes of abdominal pain and it is useful to establish whether this episode is identical. If it is, then chronic surgical diseases that flare up intermittently must be considered. Recurrent acute pancreatitis would be a good example.

Drugs

Many people take therapeutic drugs. Most patients, when asked, think only of those prescribed by the doctor but in many countries in the world, including the United Kingdom, it is possible to buy drugs without a doctor's prescription and these may be relevant too. Diuretics and sympathomimetic drugs may be implicated in the onset of acute retention, digoxin overdose classically causes vomiting followed by abdominal pain, and many drugs cause cholestatic jaundice.

Not all patients know what drugs they are taking and pills may be transferred from bottle to bottle so that the labels are unreliable. Ultimately, a direct enquiry to the doctor or the pharmacist who wrote or supplied the prescription may be necessary.

Examination

No experienced doctor completely separates examination from taking the history. Observation begins the moment the doctor meets the patient and does not end until they part company. Most clinicians rapidly assimilate, almost unconsciously, many features of a new patient, and not all of them can easily be described in words. Attitude, alertness, mood, agitation, sweating, respiration, movement, the eyes, the colour, the facial expression, the pulse, the handshake, and many other factors are all put together to give an instant impression of the severity of the illness and sometimes the diagnosis. The restlessness of a patient with colic is in marked contrast to the immobility of peritonitis. The gaunt patient with sunken eyes, a weak, thready pulse, and little respiratory or abdominal movement looks the same today as did patients two-and-a-half thousand years ago when Hippocrates first described the facies of severe peritonitis.

First impressions can, of course, be false and they are not a substitute for a systematic examination. Some would say that examination does not add much to a well taken history but more evidence to help unravel a diagnosis is usually welcome. As with the history, examination of the whole patient is relevant in the overall management, although here we are concerned with the signs that are important in the diagnosis of the acute abdomen.

Vital signs

Pulse rate, respiratory rate, temperature, and blood pressure are all essential observations. The initial values on admission may be misleading because of the hustle and bustle of the journey to hospital, but subsequent measurements are important in any patient whose condition is observed following admission. The charts may give a general clue as to the diagnosis. An increase in respiratory rate suggests pulmonary pathology rather than abdominal disease. An isolated rise in temperature certainly indicates disease but it does not specify where, nor does it necessarily signify infection. The height and the course of a fever in an adult may point to a diagnosis; in children, fever is an unreliable guide as it is notoriously labile.

Consistent changes in these four vital signs over time are useful indicators of progressive pathology. A persistent rise in the pulse with an accompanying fall in the blood pressure is sure evidence that a peptic ulcer is still bleeding; increasing fever means that an empyema of the gallbladder needs draining. Changes in pulse and blood pressure following abdominal trauma are useful, although they usually indicate the need for active treatment rather than specifying the underlying diagnosis.

General features

There are many signs found elsewhere in the examination that indicate disease within the abdomen. General features of the patient, such as anaemia, jaundice, and facial flushing, all have a direct relevance to abdominal diagnosis. The pallor of fear must not be confused with the pallor of anaemia, and cyanosis often accompanies an acute intra-abdominal catastrophe. In children, acute inflammation of the upper respiratory tract can present with abdominal pain and examination is not complete until the tonsils and the ear-drums have been inspected. Here, however, we are primarily concerned with the abdominal signs.

Examination of the abdomen

Physical examination of the abdomen follows the time-honoured sequence of inspection, palpation, percussion, and auscultation. Many signs can be seen and few patients, even young children, object to simple observation. Palpation can be painful and it is certainly unusual. Explaining what you are doing helps a patient to relax and so does distraction with conversation. Sometimes palpation with a stethoscope is useful. Percussion and auscultation are less useful in the abdomen than in the chest.

Different doctors obtain different histories and variations in the interpretation of physical signs are even more marked. Natural variation is compounded by the lack of universal agreement on the definition of some physical signs. Despite this the basic findings should be recorded in the notes. Eponymous signs are best avoided. In practice they are rarely absolutely pathognomonic of one condition.

Inspection

Inspection of the abdomen is a subtle art. First and foremost both the patient and the examiner should be comfortable. The patient must lie as flat and as straight as possible with the head on a single pillow. Daylight and warmth are desirable, and adequate exposure of the abdomen essential, although it is kind to keep the genitalia covered until they are actually examined.

Time should then be spent simply looking but looking in an intelligent and thoughtful way. Most important physical signs can often be seen (Fig. 2). The history will have given some clues as to possible diagnoses, and there will be specific signs to look for while remembering that negative findings are equally important. Previous abdominal operations will have been noted in the history and the scars can be examined. Their only importance now is that there may be an incisional hernia or underlying intraperitoneal adhesions. Obvious discoloration is always important. Bruising from a seat-belt injury or the blue-grey discoloration in the flanks or around the umbilicus from haemorrhagic pancreatitis are both good examples.

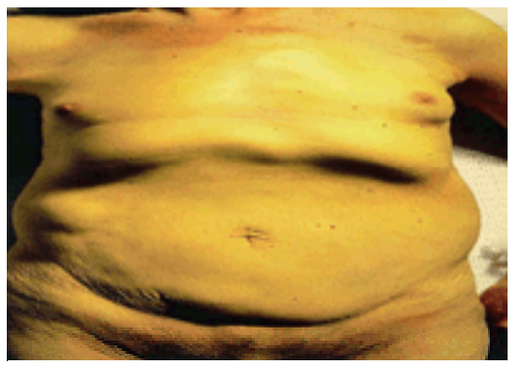


Fig. 2. All the necessary physical signs can be seen in this patient confidently to diagnose cancer of the pancreas. The patient is jaundiced, the skin is hanging loosely indicating severe weight loss, and the distended gallbladder can be seen in the right hypochondrium.

Shape

The first thing to decide is whether the shape, symmetry, and contour of the abdominal wall are normal. Generalized distension is usually obvious except in obese patients, when it can be very difficult to decide if the abdomen is simply fat. The most common cause of generalized distension is a fetus. Excess fluid and air in the gut

and ascites are the common pathological causes of distension; this is usually symmetrical. Asymmetrical distension is best judged from the end of the bed and is caused either by a mass within the abdominal cavity or a lump in the abdominal wall. The two can be differentiated because the latter must always move with the abdominal wall whereas intraperitoneal lumps do not necessarily do so.

Movement

The abdominal wall normally moves with respiration. With the patient lying on his or her back the abdominal wall rises up on inspiration as the diaphragm descends and falls back on expiration. If this respiratory movement hurts, then the patient will try to reduce or eliminate any movement by keeping the abdominal wall over the painful area still. This can often be seen and the effect can be enhanced by asking the patient to take a deep breath. Another common technique, but one that is less useful in the author's experience, is to ask the patient to blow their tummy out and to suck it in. Patients with peritonitis find this painful, as they do when asked to cough. Sometimes, in thin patients, it is possible to see the muscles of the abdominal wall contract spontaneously in response to the painful stimulus. This is visible guarding.

Sometimes movement within the abdominal cavity can be seen on the surface. Aortic pulsation and fetal movements are both normal, and so, occasionally, in elderly individuals or those with gastroenteritis, is visible peristalsis. It is, however, a classic sign of intestinal obstruction. Distended loops of bowel can be seen through the abdominal wall and peristaltic contractions can often also be seen. These contractions are sometimes accompanied by borborygmi audible with or without a stethoscope. Patience is needed, and sometimes peristalsis can be stimulated by palpation of the abdomen.

Palpation

Palpation of the abdomen requires warm hands, short fingernails, and care. By convention the doctor sits on the patient's right with the right hand flat on the abdomen in a comfortable position. Students, however, should learn to be ambidextrous because sometimes only the left side of the patient is accessible and some organs, such as the gallbladder, are occasionally easier to feel from the left.

Superficial palpation should consist of gentle movements of the whole hand. Deep palpation is achieved by gentle pressure and by flexion of the metacarpophalangeal joints whilst keeping the fingers extended. It is best to begin by asking the patient where the abdomen hurts and then to start palpating in the opposite corner. Work towards the painful area but do take care. Once hurt, few patients will relax. The signs then become difficult to interpret and are sometimes actually misleading.

The abnormalities of importance in the acute abdomen separate into three groups. There are the signs associated with peritonitis, those that accompany a mass or enlargement of one of the solid organs, and finally those that differentiate the causes of abdominal distension.

Signs of peritonitis The four signs of peritonitis are tenderness, guarding, rigidity, and rebound tenderness. Eliciting these signs is painful and it is better to see than to hear the pain. A flicker of the eyelids or a facial grimace are quite sufficient to establish the presence of pain, although guarding and rigidity are usually felt.

Tenderness This is present when any palpation of the abdominal wall causes pain. It is either present or absent, although it is also possible to establish the extent of the tenderness over the abdominal wall. It is not easy to assess severity because patients vary so much in their reaction to pain. It is useful to establish where in an individual patient the pain is worst. Pain arising from the parietal peritoneum is accurately localized and patients can often point to the site of most intense pain. The examiner can also ask the patient to compare the intensity of pain by direct pressure in the four quadrants of the abdomen.

Guarding There are different opinions about the physical signs of guarding and rigidity, so the examiner must be specific about what s/he actually means. In the author's opinion, guarding is present when there is reflex contraction of the muscles of the abdominal wall when the examining hand palpates it and thus causes slight pain. This may be seen but is more commonly felt.

Rigidity Again there is no generally accepted definition of this sign, but the most useful description is of an involuntary increase in the resting tone of the muscles of the abdominal wall. It may be localized or generalized. It is felt as an increased resistance of the abdominal wall to palpation. The intensity varies from minor increases in tension right up to the typical generalized, board-like rigidity classically associated with perforation of a peptic ulcer.

Rebound tenderness This is the most important physical sign of the four. It can be a difficult sign to elicit but when present it establishes the presence of peritonitis. It occurs when inflamed visceral peritoneum moves across and irritates the parietal peritoneum, and is best detected by percussion. This produces small movements of the underlying tissues, causes least pain, and can even localize the sign to specific areas within the abdomen. The classical method of detecting rebound tenderness by gross depression of the abdominal wall with the hand and then sudden release (hence the term ‘release tenderness') is both crude and unkind, and while sometimes useful should generally be abandoned. Rebound tenderness is also a symptom. Movement such as walking or the jolting of a vehicle may exacerbate the abdominal pain, and it always worth enquiring about this whilst taking the history.

Abdominal swellings It is essential to establish the size of all the solid intra-abdominal organs during palpation and equally important to identify any abnormal masses. When the liver, spleen, and kidneys are enlarged there are certain specific signs that must be sought. When an abnormal mass is felt either within or separately from the solid organs, then all the usual rules relating to the examination of lumps apply, although it may be impossible to assess swellings that lie deep within the abdominal cavity. Particular attention should be paid to the anatomical origin of the lump. Here mobility, and movement with respiration and pulsation, are useful. It is always helpful to establish that a swelling is cystic. Sometimes tenderness and the other signs of peritonitis coexist with an abdominal swelling.

Abdominal distension Abnormal abdominal distension may be caused by an abdominal swelling but flatus, fluid, and faeces are more common. Pregnancy is generally obvious and faeces are easily discovered on rectal examination. Excess gas or fluid within the abdominal cavity is easy to demonstrate, but establishing the presence of free intraperitoneal air or ascites can be difficult.

Groins and genitalia No abdominal examination is complete without examination of the groins and the genitalia, particularly in men. Hernias are common but not always obvious. A small femoral hernia in a large woman is easily missed. If the hernia is the cause of an obstruction it will also be tense, tender, and irreducible, but it may not be very large.

Scrotal abnormalities such as testicular torsion and epididymo-orchitis can present with abdominal pain, but there are always abnormal scrotal signs on examination.

Rectal and vaginal examination No patient likes a rectal or a vaginal examination but they are essential. Again, the examination needs to be conducted with thought. Consider all the anatomical structures in the pelvis, including the prostate and the cervix, and look at the glove for blood or pus when the examination is finished. Rectal tenderness on the patient's right side may be the only sign of pelvic appendicitis. A swelling in a fallopian tube on vaginal examination may be the only sign of an ectopic pregnancy.

Percussion

Percussion of the abdomen has three specific uses. First, it is the best method of eliciting rebound tenderness. Second, it is the most sensitive method for detecting enlargement of the bladder. Third, shifting dullness determines the presence or absence of ascites. It has a subsidiary role in confirming the size of the liver and spleen, and may sometimes be useful in outlining an intra-abdominal mass.

Auscultation

Auscultation of the abdomen is not very helpful but the presence or absence of bowel sounds is a useful physical sign. Qualitative observations are less reliable. Nevertheless, an increase in the magnitude and the frequency of bowel sounds accompanies mechanical intestinal obstruction whilst a succussion splash, which can sometimes be heard without a stethoscope, is a sure sign of obstruction. Bowel sounds that definitely disappear during observation of a patient with abdominal pain and tenderness indicate the onset of peritonitis and the need for a laparotomy.

Investigations

Although investigations are more or less routinely requested in most patients with acute abdominal pain, very few of the tests are actually valuable in making a diagnosis. In a few patients no investigations are necessary because the diagnosis is clinically obvious. In the majority the cause of the pain is initially uncertain and it is hoped that tests will help. Older and more experienced surgeons maintain that it is preferable to wait and see in these circumstances. They argue that significant disease is usually progressive and when the patient is re-examined after an interval the physical signs are more marked and the diagnosis easier. Younger surgeons think that the delay gives time for complications to develop, with a consequent increase in postoperative complications that diagnostic investigations might avoid.

However, their enthusiasm for investigation can also delay a necessary operation if the tests take too long to perform. In a few patients an accurate diagnosis that is essential for correct treatment can only be made with the help of special investigations.

We are most concerned here with tests that will help in the diagnosis and the management of the patient within the first 24 h of admission. After that the number of tests that can sometimes be useful is vast and they are considered in the individual subject chapters. Analysis of venous blood and various radiographs are the most popular immediate investigations, with the addition of ultrasonographic examination and computed tomography. They can be divided into two groups. Those tests that help in diagnosis and those that help in management.

Tests useful in diagnosis

Testing the urine

Simple clinical inspection of the urine should still be regarded as an essential part of examination of the abdomen. Urine containing tiny amounts of blood looks smoky. Sugar and ketones can both be smelt. Infected urine may be cloudy or blood-stained, smells unpleasant, and contains nitrites and white blood cells. Confirmation of all these findings using biochemical sticks is convenient and easy.

When an infection of the urinary tract is suspected, then a carefully collected urine specimen should also be sent immediately to the laboratory for analysis. It is not easy for any patient to provide a true mid-stream urine specimen, and they must be both helped and supervised. Even then, contamination can be a problem and there are occasions, particularly in women and children, when a catheter specimen should be collected. Urethral catheterization is usually appropriate, but suprapubic puncture of the bladder provides the least contaminated specimen and carries the least risk of introducing an infection. Even though the culture result will not be available for a few days the sample must be sent immediately otherwise the opportunity to identify the organism responsible may be lost, as most patients with a urinary-tract infection presenting with acute abdominal pain will need immediate treatment with antibiotics.

Blood tests

White blood-cell count Many significant causes of acute abdominal pain are associated with some degree of inflammation. As a consequence, an increase both in the absolute numbers of white cells and in the proportion of neutrophils might be expected. The reverse observation is also true: an increase in the white-cell count indicates the existence of inflammation. It is always necessary to interpret the result in the clinical context, for the inflammation may not necessarily be within the abdomen. A value within the normal range does not exclude intra-abdominal inflammation.

This very simple way of looking at the white-cell count is not the most useful. It is more helpful to interpret the result in a statistical sense. In other words, the probability of a patient with a normal white-cell count having acute appendicitis, for example, is low, whilst the chances with a raised count are higher (Table 2).

White cell count	*Percentage chance of appendicitis*
Up to 8000	5
8001–10000	7
10001–12000	18
12001–15000	36
Over 15000	67

Table 2 Relation between total white blood-cell count and chance of appendicitis (de Dombal, 1991)

The same observations may be made about an excess of neutrophils in the differential white-cell count. Indeed the results of all such tests used to establish a diagnosis should ideally be analysed in this way. In practice a normal white-cell count is often used to reassure the surgeon who wants to wait and see, while an increased count supports a decision to operate. The surgeon should realize, however, that the test is then being used to help in a management decision and not to make a diagnosis.

Serum amylase Acute pancreatitis usually presents with the symptoms and signs of peritonitis, and normally patients with peritonitis warrant an immediate laparotomy. Surgery is, however, best avoided in patients with acute pancreatitis. The rise in serum amylase that usually accompanies pancreatitis allows the correct diagnosis to be made and a laparotomy is thus averted.

Because the result is so important for both diagnosis and treatment it is essential to appreciate the limitations of the test. Other intra-abdominal catastrophes, such as a perforated peptic ulcer, a ruptured aortic aneurysm, or dead gut, can cause a modest rise in the serum amylase, while if the blood sample is taken too long after the onset of the pancreatitis the enzyme's concentration may have reverted to normal and so give a false-negative result. Again, a statistical approach can be adopted. A low serum amylase carries a low chance that the patient has acute pancreatitis whilst a high concentration implies a high chance (but not a certainty) that pancreatitis is indeed the diagnosis.

Radiological investigations

Plain abdominal radiographs Controversy surrounds the use of plain abdominal radiology. Sometimes the films confirm the clinical diagnosis, add further detail, and modify the management of an individual patient. At other times the films are simply misleading, although occasionally they suggest a diagnosis that the clinician has not considered. One thing is certain. Not every patient with acute abdominal pain needs an abdominal radiograph. When one is requested the doctor should be clear what information s/he hopes to gain and s/he must have the skills to interpret the films if no radiologist is available.

Traditionally two films are taken, one with the patient lying supine and the other with the patient sitting or standing erect. Modern protagonists of a single supine film point out that little additional information is derived from the erect film and add that not every patient with acute abdominal pain can safely or comfortably sit or stand. Some radiologists prefer, as an alternative to the erect film, to lie the patient on their right side and then take a lateral radiograph (the lateral decubitus view). In the author's opinion an erect view does, on occasion, add useful information whereas the lateral decubitus view usually does not. It provides only a limited view of the abdominal cavity and free intraperitoneal gas is better seen on a chest radiograph.

Abdominal films are more use in some circumstances than in others. None of the radiological signs of acute appendicitis is truly helpful, but radiological examination should be performed in patients with suspected intestinal obstruction and those who have suffered abdominal trauma (Fig. 3). Stones in the kidney, the ureter, or the gallbladder are sometimes confirmed on a plain film, and calcification of the wall of an abdominal aortic aneurysm may be the only clue to its presence. Radiology of the abdomen is more useful in older patients, who tend to have more significant pathology and thus more abnormalities on such films. It is important to remember that the presence of abnormalities on any abdominal radiograph is valuable but their absence is meaningless.



Fig. 3. A supine abdominal radiograph demonstrating grossly distended loops of small bowel typical of intestinal obstruction (by courtesy of Dr D. Lindsell).

Chest radiography A good-quality, erect chest radiograph is the best film with which to confirm the presence of free intraperitoneal air (Fig. 4). This can be seen as a black crescent, sometimes with an air/fluid level, underneath one or both diaphragms. Proximal perforations of the bowel tend to lead to larger amounts of free air than distal ones; if the perforation has occurred some time before presentation, as can happen in patients with diverticular disease, the margin of the pneumoperitoneum on the radiograph is often rather hazy and irregular. There may also be a small pleural reaction above the diaphragm.



Fig. 4. An erect chest radiograph demonstrating free gas under the diaphragm (by courtesy of Dr B. Shepstone).

In very old and very young patients, pneumonia and pleurisy present with referred abdominal pain. Fractures of lower ribs may indicate a ruptured spleen or lacerated liver.

Intravenous urography Renal colic is usually an easy clinical diagnosis to make because of the characteristic distribution of the pain. An emergency ultrasonographic examination may show a hydronephrosis on the affected side and will occasionally identify a calculus. An urgent intravenous urogram adds additional information and will indicate the site of an obstruction, and outline the degree of dilatation of the urinary tract and the size of an offending stone.

Occasionally, emergency urography is useful. Delayed excretion of contrast on the side where the patient complains of pain confirms the diagnosis. A normal urogram effectively excludes the diagnosis, provided the examination is done within a short time of the last episode of pain. Other causes of the abdominal pain can then be considered. Intravenous urography is also useful in trauma to the urinary tract. Most such patients have haematuria. The degree and the site of any damage may be displayed and the presence of a normally functioning kidney on the unaffected side can be confirmed.

Ultrasound examination

Ultrasound is widely used in the diagnosis of acute abdominal pain. Its place in elective diagnosis of conditions affecting the upper abdomen, the pelvis, and the retroperitoneum is already established; it is also useful in the emergency patient. Gallstones and an aortic aneurysm are easy to see (Fig. 5), as are the oedematous gallbladder wall and a tear in the aneurysm. Transabdominal and vaginal ultrasound are both useful in the pelvis to identify swellings of the uterus, ovaries or fallopian tubes. The ultrasound probe can also be used, like the examining hand, to identify the specific structure that hurts.

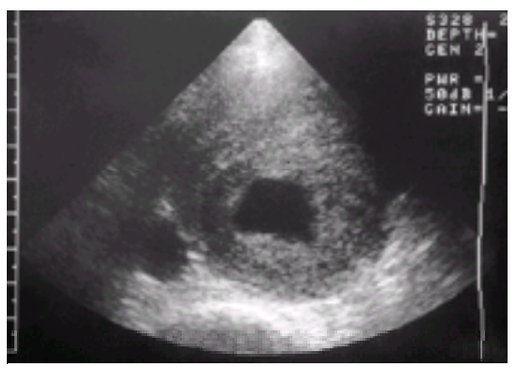


Fig. 5. An ultrasonogram of the abdomen demonstrating a large aortic aneurysm filled with concentric rings of thrombus surrounding a residual lumen (by courtesy of Dr D. Lindsell).

Ultrasound is less useful in examining the bowel because of the presence of gas. However, the inflamed appendix often lies behind the caecum and contains little air. Certainly the ultrasound probe can localize the tenderness to this specific area and sometimes it can also demonstrate an oedematous tubular structure at the site where a retrocaecal appendix should lie.

Doppler ultrasound, which demonstrates flow in vessels, can help decide the cause of acute testicular pain. The hyperaemia of epididymo-orchitis is in marked contrast to the ischaemia of torsion.

Following trauma, ultrasound can demonstrate the presence of free intraperitoneal fluid and look for damage to the liver, spleen, kidneys and pancreas. It cannot identify blood clot very well and it is of no practical use in looking for injury to the gut. Ultrasound is more readily available and can be brought to the patient's bedside but computed tomography is more accurate.

Computed tomography (CT)

CT has become an essential tool in the diagnosis and the treatment of the acute abdomen. Swelling of the pancreas and peripancreatic oedema will confirm a diagnosis of acute pancreatitis. Lack of perfusion of parts of the pancreas on a contrast study implies potential pancreatic necrosis and identifies a patient with a high risk of complications. CT with contrast is the best method of confirming a diagnosis of acute diverticulitis. It can also identify a pericolic abscess, which may best be treated by CT-guided drainage. CT is of occasional use in the diagnosis of an abdominal aortic aneurysm.

Many patients with abdominal trauma require an immediate laparotomy. Other patients are less acutely injured and CT is helpful in identifying the nature and the extent of any intra-abdominal injury.

Tests useful in management

Many of the tests that are useful in diagnosis also have a role in management. A progressive reduction in the white-cell count or an improvement in the radiological signs of obstruction after treatment both indicate resolution of the pathology. A large number of other tests also help in the treatment of a patient, many of which are undertaken soon after the patient is admitted to hospital. Some of them also play a part in diagnosis as well.

Blood tests

Haemoglobin concentration and packed cell volume The clinical diagnosis of anaemia is not always reliable, and in any patient who may possibly have an anaesthetic it is clearly important to know the oxygen-carrying capacity of the blood. The initial haemoglobin value does not indicate the volume of blood lost in patients with overt evidence of acute haemorrhage, but sequential measurements can give a rough guide, provided any blood transfused is taken into account. Occasionally, the discovery of an unexpectedly low haemoglobin can help in diagnosis: carcinoma of the caecum as a cause for intestinal obstruction with anaemia is a classic example. Packed cell volume accurately reflects the severity of fluid loss in a dehydrated patient and it is a good guide to the adequacy of rehydration.

Creatinine and electrolytes Most patients with significant intra-abdominal pathology should have their creatinine and electrolyte concentrations measured on admission. The initial values must be interpreted in the clinical context, particularly if the patient is dehydrated; in most circumstances it is the serum potassium that is the most important because of its role in cardiac function. Serial values are vital for proper postoperative fluid management.

Liver function tests Most patients with an acute abdomen due to liver and biliary disease are jaundiced. The depth of the jaundice reflects the severity of the pathology, and it is rare to need to measure the liver function acutely. It is, however, essential to obtain a blood sample on admission for later analysis because subsequent biochemical deterioration, which may not be clinically obvious, will demand further action. This particularly applies to elderly patients in whom the signs and symptoms of biliary disease are often obscure. The diagnosis is sometimes not even considered until abnormal results of liver function tests are discovered.

Calcium concentration This is only of immediate value in patients with acute pancreatitis. Depleted values are an indirect guide to the diagnosis and are used in some severity-scoring systems. When low calcium threatens to induce tetany, intravenous calcium supplements will be needed. Calcium is always measured in patients with renal colic, in whom evidence of hyperparathyroidism is sought, but hypercalcaemia is rarely found.

Blood gas analysis An arterial blood sample should be analysed in a patient who is severely ill with an acute abdomen from whatever cause. Many such patients are covertly hypoxic, and the result of blood gas analysis may indicate the need for immediate ventilatory support. More commonly, patients will need ventilation after an emergency operation; preoperative values are then a useful measure of the patient's progress.

Blood gas analysis is also a component of many scoring systems to assess the severity of acute pancreatitis.

Radiology

If a chest radiograph is not necessary for diagnosis it is unlikely that it will be needed in the management of the patient. Nevertheless, there are times when, although a clinical diagnosis can be made, a chest radiograph should be obtained simply to provide a baseline. It is often useful to know that postoperative changes in a number of investigations, particularly the chest radiograph, were not present before surgery.

Contrast radiology Conventional contrast radiology is rarely needed as an emergency, although an urgent barium or air enema is important in a child with suspected intussusception. In adults with obstruction of the large bowel a limited barium enema examination is sometimes useful to establish the presence of a mechanical rather than pseudo-obstruction. In patients with obstruction of the small bowel where the cause is obscure or resolution is not occurring as fast as expected, then a small-bowel enema is always helpful.

Electrocardiography

Anyone over the age of 40 years who presents with acute abdominal pain, particularly if the diagnosis is not straightforward, should have an electrocardiograph. Very occasionally a myocardial infarct will present with abdominal pain and recovery is unlikely to be helped by an unnecessary laparotomy.

Endoscopy and arteriography

Emergency gastroscopy and colonoscopy, occasionally performed on the operating table, are helpful in patients who present with acute gastrointestinal haemorrhage. Precise localization of the bleeding point is essential for effective treatment. Mesenteric angiography may be needed as well. In both instances, treatment as well as diagnosis may be possible.

Endoscopy has no part to play in the diagnosis of a perforated peptic ulcer or diverticulitis. It may make matters worse by blowing air into the peritoneal cavity through the perforation.

Peritoneal lavage

This is a useful investigation in patients with abdominal trauma, particularly if they are unconscious or are otherwise unable to cooperate in an abdominal examination. The presence of significant amounts of blood or intestinal contents in the wash-out fluid is a clear indication for an urgent laparotomy.

Excess neutrophils in fluid aspirated from, or washed out of, the peritoneum is a reliable indicator of the presence of peritonitis. The test is rarely used because it gives no indication of the site or the cause of the inflammation.

Laparoscopy

Laparoscopy is invaluable in the diagnosis and the treatment of patients with acute abdominal pain, although it does involve a general anaesthetic. All the organs responsible for the common causes of acute abdominal pain, such as the appendix, the gallbladder, the fallopian tubes, the ovaries, most of the small bowel, the stomach and the sigmoid colon, are easily seen through the laparoscope. In many cases it will be appropriate to treat the problem as well. The technique of laparoscopic appendicectomy is now well established. In any event, it is usually possible to use a better-placed and often smaller incision once the diagnosis is established even if a laparotomy is needed.

Laparoscopy has a limited role in the victim of abdominal trauma. If blood is found, then the endoscopist must look for the source and decide if the severity is sufficient to justify a laparotomy. If intestinal contents are found, a laparotomy is obligatory.

Making a diagnosis

No one really understands how a doctor makes a diagnosis, although the process has been analysed many times. In theory it is simply a matter of collecting all the relevant facts and analysing them correctly. The contrast between the junior clinician who takes time and trouble over the patient and yet makes the wrong diagnosis half the time and the senior colleague who asks a few questions, performs a limited examination, and is right eight times out of ten shows that this is not the whole story. Very few patients present with all the symptoms and signs of their disease and only experience can teach the clinician which few questions to ask, how to ask them, and how to interpret the answers correctly in the context of the individual patient. The last skill is particularly important when some of the facts conflict. Experience and constant practice are certainly essential for maximum accuracy.

In actual clinical practice, several methods are used to make a diagnosis. Some involve purely practical considerations whilst others look at the same data in different ways. Most clinicians use all the methods at one time or another, often together, and usually without giving the matter a second thought.

The classic case

In this unusual circumstance the patient gives the typical history of a classical cause of abdominal pain with every symptom in its correct place. Examination reveals all the expected signs and the diagnosis is obvious, even to the least experienced clinician. There is really no place for a differential diagnosis nor are investigations necessary. All that is left to be done is to organize the correct treatment.

A question of elimination

More commonly a variety of diagnoses are suggested by the patient's initial complaint. Most doctors then ask further questions to support or exclude each individual diagnosis. The method reaches its peak with the most experienced clinician who makes a diagnosis on the basis of half-a-dozen questions and a very limited examination of the abdomen. This is the culmination of a natural development in most doctors. To begin with they are taught to obtain information in a systematic way and to record it all in a standard format. Later they learn that certain symptoms and signs are commonly associated with certain conditions and so they construct in their minds algorithms for each individual patient. With experience the method works well but it is dangerous for the beginner because of the risk of following the wrong pathway early in the ‘decision tree'.

Anatomy and pathology

Another approach is to consider which anatomical structure could be the cause of the symptoms and signs. Abdominal anatomy is broadly the same in everyone and most patients can localize their pain to one area of the abdomen. Each organ within that area is then considered in turn as the source of the symptoms. The most difficult area is the left upper quadrant, where there are no structures that commonly cause acute abdominal pain. Anatomy can also be used to construct a list of all the other structures that could cause abdominal pain at any particular site.

A particular pathological process is suggested by certain symptoms and signs. This may, in turn, suggest a diagnosis. Acute intra-abdominal inflammation is usually accompanied by a fever and abdominal tenderness, whilst mechanical obstruction is characterized by colic and vomiting. Sometimes obstruction and inflammation coexist but this in itself shortens the diagnostic list to only a few possibilities.

Age and sex

The importance of sex in relation to the possible causes of abdominal pain is obvious but age is an equally valuable discriminator. Cancer is more common in the old than the young, for example, whereas non-specific abdominal pain is a disease of the young.

Statistics

Acute appendicitis and non-specific abdominal pain account for approximately two-thirds of all the patients admitted to hospital with abdominal pain in Europe and North America (see Table 1). This means that for two-thirds of the time the only important distinction the surgeon has to make is between these two conditions. Looked at another way this also means that if it was possible to separate these two diagnoses accurately from all the other possibilities, then the clinician would make a correct diagnosis in two-thirds of the patients.

The pattern of disease

Common things occur commonly. The most common surgical cause of acute abdominal pain in a patient admitted to a hospital anywhere in the world is acute appendicitis. The next most common cause in Africa is small-bowel obstruction; in the West it is acute cholecystitis. Knowledge of local epidemiology is therefore useful in the diagnosis of abdominal pain, although the patterns of disease will change over time.

Listing the possibilities

This is essential for the young clinician but a luxury for the surgeon in charge who has to decide on treatment. The latter keeps a list in his head but the former will be wise to write the possibilities down, preferably in order of probability. There is truth in the saying that you will not make a diagnosis unless you think of it: making a list may jog the memory, as the rarities are easily forgotten. However, a list of possible diagnoses only provides a framework on which to base further enquiry. It is of no direct help in planning treatment.

A repeat visit

Most patients with abdominal pain will have some investigations done on admission, particularly if the diagnosis is not obvious. The results usually support or refute a diagnosis that has already been considered, but they sometimes suggest another condition. This applies particularly to the abdominal radiographs. It is then always worth repeating the clinical examination at once to test whether this possible diagnosis would fit with the findings. It is true that this is the reverse way to interpret information but it is, nevertheless, often a good way to make a correct diagnosis.

The return visit

Sometimes the initial symptoms and signs are insufficient to make a diagnosis. In these circumstances, time is an invaluable ally. If spontaneous improvement occurs over the ensuing few hours, then the need to make a prompt diagnosis lessens; if the disease process progresses, the symptoms, and particularly the signs, will worsen. In practice the clinician returns to examine the patient again 3 or 4 h after admission. If the symptoms and signs are worse, then the diagnosis will usually be apparent. If there is improvement, then no surgical treatment is needed, but it is also probable that no specific diagnosis will ever be made. If there has been no change, then a further examination a few hours later again will be needed. Some people call this approach ‘active observation'.

There are two risks attached to a ‘wait and see' policy. The first is that the passage of time may allow a complication to develop, with a consequent increase in morbidity. The second is a lack of intellectual rigour: the clinician waits until the patient has recovered or it is perfectly obvious that a laparotomy is essential and never bothers to make a clinical diagnosis.

Computer assistance

Only a clinician can obtain the information that is needed from a patient with acute abdominal pain and decide on the best form of management. This information has first to be analysed in order to make a diagnosis before treatment can begin. It is here that computers can help, although this help does not derive entirely from sophisticated data analysis. In the Western world the diagnostic accuracy of the first doctor who sees a patient with acute abdominal pain in hospital is about 45 per cent. Improving the quality of the information obtained by the rigid definition of symptoms and signs and the amount of information collected with the use of special forms will increase this initial accuracy to 60 per cent. This figure can be improved to about 70 per cent by giving feedback about their accuracy to individual doctors and computer analysis of the data from individual patients by comparison with a large database of information from patients with abdominal pain of known cause. When the results of investigations are available, this further improves to 75 per cent. Performance is improved at every level of surgical experience, and the very best clinicians can make an accurate diagnosis of the cause of acute abdominal pain 80 per cent of the time. This improvement in diagnostic accuracy is also reflected in an improved outcome for the patient, with a substantial reduction in the number of normal or perforated appendices removed at operation and fewer serious surgical errors.

The pragmatic approach

The pragmatists claim that the diagnosis lies between acute appendicitis and non-specific abdominal pain in two-thirds of the patients; those in the former group need an operation whereas those in the latter do not. They also claim that many of the remaining one-third of patients with another diagnosis will benefit from a laparotomy when the signs indicate that one is needed. So the clinician faced with a patient with acute abdominal pain has only one decision to take. Do the symptoms and signs justify a laparotomy? If the answer is ‘yes' than the diagnosis will become obvious at operation. If the answer is ‘no' then a period of active observation is all that is needed. The main disadvantages are a high rate of ‘negative laparotomy' and the complications that ensue from inappropriate operations.

Many younger surgeons would substitute a laparoscopy for a laparotomy in this management plan. They would also recommend laparoscopy earlier in the course of

the illness and with less justification than for a laparotomy. Whether this leads to better and more accurate treatment with less complications is not yet clear.

There are occasions when a practical approach has to be adopted. It may be obvious that a laparotomy is essential but a diagnosis cannot be made because the patient is unconscious or is too ill to co-operate. Alternatively, there may be no time to make a complete diagnosis. This will certainly be the case when the available surgical services are overwhelmed by multiple casualties. Absent bowel sounds as an indication of peritonitis would often be sufficient justification for a laparotomy in these circumstances.

The use of analgesia

It is a good rule that patients with acute abdominal pain are not given analgesia until a diagnosis has been made and the treatment determined. However, common humanity often demands that pain is relieved before a surgical opinion is obtained. Provided that the time of administration and the dose of the drug are recorded, an experienced clinician should not be deceived. Analgesia does not eliminate all the physical signs but it does subdue them and due allowance must be made. Once a diagnosis has been made, every patient must be given adequate analgesia at once, whatever other treatment is planned.

Sometimes a patient cannot be properly examined because of the severity of their pain. Adequate analgesia, perhaps given intravenously, may then make an examination, and so a more accurate diagnosis, possible.

Spot diagnosis

This is included only to be condemned. Immediately visible symptoms and signs in the abdomen can often be interpreted in several ways. Furthermore, the doctor who jumps to conclusions tends to make subsequent observations fit the chosen diagnosis rather than to analyse them dispassionately. There are no circumstances where an instant diagnosis of an acute abdomen should be made. It is simply not safe.

Learning to make a diagnosis

How does the young clinician obtain the skills to match his or her seniors in the diagnosis of acute abdominal pain? A determination to make the correct diagnosis on every occasion and plenty of experience are the basic answers, but there are a few techniques that will help.

Recording the facts

Computer analysis has shown the importance of obtaining all, but no more than all, the necessary information to make a diagnosis. Whilst it is usual for the house surgeon or intern to record a complete case history, residents or registrars should confine themselves to recording in their own writing the facts relevant to the diagnosis. This forces the clinician to decide which facts are important and starts the process of excluding more information than can be processed. This is easier if a special data-collection form is used.

Always make a diagnosis

The surgical trainee must always decide upon, and write down, one diagnosis, even if a list of other alternatives is appended. After all, a single diagnosis will become essential as soon as the trainee takes on the responsibility for deciding on treatment. To begin with the diagnosis will be wrong at least half the time, but as confidence and experience are gained this figure will improve.

Reporting to a senior

Asking a senior colleague for advice about the diagnosis and management of a patient is always good training. The discipline of putting the right facts together in a logical order and presenting them correctly is excellent practice at analysing the information.

Review the analysis

Finally, it is always worthwhile to compare the findings on admission with the final diagnosis. It is often possible to identify where a sign or symptom was misinterpreted and also to appreciate that certain groups of findings correlate with certain diagnoses.

The differential diagnosis

It is not the intention here to discuss in detail all the possible causes of acute abdominal pain. Rather it is to point out certain features that will help in the differentiation of the more common causes of pain and also to discuss a number of unusual problems that are not dealt with elsewhere.

There are various ways in which to classify the causes of acute abdominal pain. Bailey and Love compared the differential diagnosis of acute appendicitis to a house with two stories corresponding to the upper and lower abdomen, the pelvis (‘the basement'), the thorax (‘the attic'), and the retroperitoneal structures (‘the backyard') (Table 3). The underlying anatomical analysis is still useful, but in practice many experienced clinicians will first decide whether the features of the case suggest inflammation, obstruction, colic, a major catastrophe, or simply the presence of a mass. In many instances, more than one condition is involved. Inflammation can cause obstruction and unresolved obstruction will eventually lead to inflammation. Nevertheless the distinction is useful in practice because within each category there are then a number of possible causes for the findings.

The attic
Pneumonia
Pleurisy
The upper storey
Perforated peptic ulcer
Acute cholecystitis
The ground floor
Gastroenteritis
Mesenteric adenitis
Intestinal obstruction
Crohn's disease
Meckel's diverticulitis
The basement
Salpingitis
Ectopic gestation
Ruptured luteal cyst
Twisted ovarian cyst
The backyard
Ureteric colic
Acute pyelonephritis
Ruptured abdominal aortic aneurysm
The electrical installation
Herpes zoster

Table 3 Differential diagnosis of acute appendicitis (after Bailey and Love, 1962)

Inflammation

Inflammation within the peritoneal cavity usually arises from one of the intra-abdominal organs, although primary peritonitis does rarely occur in young girls. In most instances the inflammation is secondary to infection or ischaemia. Initially the inflammation is confined to the organ of origin and this makes the diagnosis easier as the physical signs will also be localized. Untreated and progressive inflammation will eventually lead to gangrene and necrosis, with consequent perforation of the viscus and the development of generalized peritonitis. Both the location of the initial signs and the speed of onset of generalized peritonitis will give a clue to the original site of the inflammation. The prognosis worsens with the progress of the condition and so the clinician is always anxious to treat localized peritonitis before perforation occurs.

Localized peritonitis

Acute appendicitis

Many clinicians consider this to be a complete diagnosis in itself but, in older textbooks, acute catarrhal appendicitis, acute obstructive appendicitis, acute perforated appendicitis, and an appendix mass are all described as separate clinical entities. The last two certainly present a different clinical picture from the first two and sometimes it is even possible to identify the colic and the vomiting that marks the start of an obstructive appendicitis. But with the exception of an appendix mass, where conservative management is usually preferred, the other three varieties all require an early operation so that differentiating between them is hardly worthwhile. Even so, it is always wise to insist on adequate resuscitation of the patient with perforated appendicitis before surgery.

The other common error is to assume that all the patients with acute appendicitis present to a surgeon and, furthermore, that they all end up with an appendicectomy. Acute inflammation sometimes resolves spontaneously, and acute appendicitis is no exception. A proportion of the patients diagnosed as having non-specific abdominal pain have mild acute appendicitis, and a few such patients are readmitted a few days or weeks later with clear-cut appendicitis or an appendix mass.

Half the patients manifest all the classical symptoms and signs of appendicitis but the other half presents in a variety of ways that range from the misleading to the bizarre. The diagnosis is often difficult at the extremes of life and during pregnancy. The young may not be keen to reveal their symptoms and in the old there are often few physical signs. Stoical and muscular young men may have convincing symptoms but trivial tenderness. Pregnancy distorts the anatomy and the uterus can itself be the source of the pain. Diarrhoea and vomiting with abdominal pain will suggest gastroenteritis to most doctors, but they can be caused by appendicitis; appendicitis may also accompany gastroenteritis. The variable anatomical position of the appendix in normal people can also mislead the clinician. There may be few abdominal signs with pelvic appendicitis and the doctor who omits the rectal examination will miss the diagnosis. Retrocaecal appendicitis often causes microscopic haematuria from inflammation of the adjacent ureter and this can lead to an incorrect diagnosis of urinary-tract infection or ureteric colic.

Bizarre presentations include a lumbar abscess from missed retrocaecal appendicitis, a perineal sinus from pelvic appendicitis, and appendicitis in the presence of an appendicectomy scar.

The list of differential diagnoses is very long: many conditions can mimic appendicitis, from right-sided pneumonia to salpingitis, and including a ruptured aortic aneurysm (Table 3). At least one in five appendices removed at operation is normal histologically. This seldom matters, for another surgical cause for the pain is found at operation in about half these patients, although the complications that can ensue from an unnecessary operation should not be overlooked. A greater danger is to delay an operation until perforation has occurred, when the patient is worse, surgery is more difficult, and the complication rate is higher. The surgeon who never removes a normal appendix is undoubtedly exposing other patients to the risk of perforation.

Non-specific abdominal pain

Most patients with abdominal pain never discover the cause. Only about one patient in every 15 with pain is admitted to a hospital, and of those who are admitted four out of every 10 leave without a diagnosis. Some patients find these statistics reassuring, but others find the doctor's inability to find a cause for their symptoms distressing. It is particularly important to reassure the latter group that a diagnosis of non-specific abdominal pain does not imply that there is no pain nor that there is no cause. The alternative title of non-surgical abdominal pain is more accurate in the sense that these patients do not need an operation, but it is inaccurate since there may still be an underlying surgical problem.

The predominant symptom is always pain, but there is usually abdominal tenderness as well, often in the right iliac fossa. This normally implies peritonitis but the signs are never sufficient to justify the diagnosis. Some people call this peritonism, implying thereby irritation of the peritoneum without inflammation; this term is best avoided as it has no true pathological explanation.

Such patients are usually actively observed, and the symptoms and the signs subside as mysteriously as they came. Up to a point, therefore, it is a diagnosis of exclusion and one that is only made in retrospect, but there are a few pointers towards a positive diagnosis. Most of the patients are young and two-thirds of them are women. The pain is rarely made worse by movement, it rarely moves its position in the abdomen, and about one-third of patients will keep their eyes closed during the abdominal examination (the ‘closed eyes' sign).

Some of these patients do, in fact, have minor versions of recognized clinical conditions such as mesenteric adenitis, threadworm infestation, gynaecological pain from ovulation, or torsion of a colonic appendix epiploicae. Incomplete intestinal obstruction may not be clinically obvious and it may resolve before the diagnosis is made if the loop of bowel releases itself or the adhesion tears. Obscure abdominal pain in the elderly person, which is uncommon, is often associated with cancer, particularly cancer of the colon. Social and psychological factors play a very important part in some patients. This is usually because of anxiety about the minor abdominal pains that afflict everyone at some time or another rather than being a primary cause in themselves.

Acute cholecystitis

This is usually an easy diagnosis to make and even easier to confirm with the use of ultrasound. In the developed world it is the most common cause of acute abdominal pain in elderly people and the third most common cause in the younger age groups.

There are three important circumstances in which the diagnosis is difficult. In the elderly person the symptoms are sometimes obscure and the signs minimal. There is often no fever and the white-cell count is normal. Acute acalculous cholecystitis can develop insidiously in patients who are very ill for another reason. They are often being fed intravenously in intensive care. Ultrasonography can confirm the diagnosis, but only if it is considered. The third difficulty arises when a patient gives a classical history of acute gallstone disease but has no stones on ultrasonographic examination, even though the ultrasound probe may identify the gallbladder as the source of the pain. Sometimes the gallstones are too small to be seen, and occasionally the patient has passed their only stone. The diagnosis is then only made in retrospect when the stones have either reformed or grown in size. Alternative explanations for acute abdominal pain in the right upper quadrant in the absence of gallstones are exacerbation of a peptic ulcer, renal colic, and chlamydial perihepatitis (the Curtis–Fitz-Hugh syndrome).

Acute pancreatitis

Sometimes acute pancreatitis can be confidently diagnosed on the basis of the patient's symptoms and signs. Constant epigastric pain radiating to the back and bruising in the flanks or around the umbilicus, which can be very difficult to see, are the key features. A high serum amylase merely confirms the clinical opinion. At other times it is difficult to distinguish pancreatitis from other causes of upper abdominal peritonitis such as a perforated ulcer, ischaemic bowel, or biliary colic, particularly if there is only a modest rise in the serum amylase. CT is always helpful in this circumstance. It will confirm or refute a diagnosis of acute pancreatitis and may, in the latter case, identify an alternative cause for the patient's pain.

Most patients with an acute flare up of chronic pancreatitis make the diagnosis for themselves. The signs are often subdued and there is usually little or no rise in the serum amylase. Acute pancreatitis is a rare complication of surgery. It has a poor prognosis. It is particularly difficult to diagnose, especially if the patient has had an abdominal operation. Frequently the diagnosis is only made when the serum amylase is found to be raised as part of a biochemical screen. This reinforces the general point that serum amylase should be measured in any patient with signs of peritonitis. High serum concentrations generally support the diagnosis; modest elevations demand re-evaluation of the patient and consideration of other possibilities.

Acute diverticular disease

This may present in a variety of ways. The most common is acute localized inflammation of a diverticulum. Less common presentations are generalized peritonitis from perforation of a diverticulum, a pericolic abscess, intestinal obstruction from adhesions, a faecal fistula, and acute rectal haemorrhage.

Acute pain and tenderness in the lower abdomen and focused on the left iliac fossa are the hallmarks of acute diverticulitis. Initially the symptoms and signs are fairly widely spread, even into the upper abdomen, but as the inflammation resolves the signs tend to localize to the left iliac fossa. There is usually an alteration in bowel habit and often frequency of micturition. A rectal examination is essential, as tenderness or even a mass may be palpable; the most important differential diagnosis is carcinoma of the large bowel. CT with oral, rectal, and intravenous contrast will confirm the diagnosis. A barium enema or a colonoscopy is best deferred until the acute episode has subsided because of the risk of perforating the inflamed diverticulum.

Acute ileitis

Some patients operated upon for appendicitis have instead acute inflammation of the terminal ileum. Crohn's disease can certainly present in this way, although there is usually a history of diarrhoea and weight loss as well as the recent pain. Infection with *Yersinia enterocolitica* can also cause a self-limiting ileitis. The latter diagnosis is made by sequential serological studies.

Acute caecal diverticulitis

This is a rare condition that presents in a similar fashion to acute appendicitis. At operation the appendix is normal and a large mass is felt in the caecum or ascending colon. Most surgeons mistake this mass for carcinoma and so perform a right hemicolectomy. If the right diagnosis is made, resection is unnecessary.

Acute gynaecological problems

Rupture of an ovarian follicle, ectopic pregnancy, salpingitis, ovarian cysts, and fibroids are the common gynaecological causes of acute abdominal pain. Most of them present with other abdominal symptoms and signs that lead to a mistaken diagnosis of acute appendicitis, but all of them can be diagnosed clinically. A vaginal examination is essential. Reliable early pregnancy tests and pelvic ultrasonography are also valuable aids to diagnosis, as is laparoscopy.

Salpingitis is the most common gynaecological cause of lower abdominal pain in women. It is often bilateral, is often accompanied by a fever and a vaginal discharge, may be associated with an intrauterine contraceptive device, and tends to be persistent and recurrent. The normal rupture of an ovarian follicle in the middle of the menstrual cycle (*mittelschmerz*) is sometimes accompanied by sufficient bleeding to cause significant lower abdominal pain. The diagnosis can often be made on ultrasonography by the presence of a small amount of fluid in the pouch of Douglas. The bleeding is rarely of any magnitude and no treatment apart from rest and analgesia is usually needed. Haemorrhage from an ectopic pregnancy is often frightening and can be lethal. The difficulty is that many patients do not even realize that they are pregnant. Tactful but specific questions about the possibility of pregnancy are essential: never forget that previous sterilization does not mean that further pregnancy is impossible. Clips come adrift, ties come undone, and tubes recanalize. The residual damage to the fallopian tube means that an ectopic pregnancy is more likely than usual.

Benign ovarian cysts twist; malignant cysts can rupture, bleed, infiltrate surrounding structures, and cause obstruction of the small bowel. Twisted cysts are easy to diagnose as there is a tender, central, lower abdominal mass arising out of the pelvis, although they are not always easy to distinguish from a degenerating fibroid.

Urinary-tract infection

This can mean anything from severe acute pyelonephritis to mild cystitis. The symptoms and the signs vary accordingly. Mild cystitis rarely presents to a hospital, but pyelonephritis certainly does and right-sided infection is easy to confuse with acute appendicitis. Normally the urine is obviously infected on testing but occasionally acute inflammation of the renal parenchyma can precede urinary symptoms and the appearance of pus in the urine by a day or two.

A perirenal abscess and a pyonephrosis both give rise to abdominal pain, but there are usually many other physical signs to suggest the true diagnosis.

Testicular pain

Occasionally patients with testicular torsion present with lower abdominal discomfort as the predominant symptom rather than severe scrotal pain. Confusion only arises when examination of the genitalia is omitted as part of the abdominal examination. The symptoms of acute epididymitis and torsion of a testicular appendage usually focus on the scrotum rather than any abdominal pain.

Meckel's diverticulitis

It is an exceptional diagnostician who can separate Meckel's diverticulitis from acute appendicitis. It can be done because the abdominal tenderness lies closer to the midline than with classical appendicitis, but there are no other distinguishing features. The difference is of no practical importance since the true diagnosis is soon apparent at operation.

Generalized peritonitis

Generalized peritonitis is easy to diagnose, but the causes are legion. Up to a point, identifying the precise cause is irrelevant, because a laparotomy is mandatory if the patient is to survive. The diagnosis is then immediately apparent. However, any surgeon knows the difficulty of oversewing a perforated duodenal ulcer in a fat patient from a lower left paramedian incision, and one incision is better then two for the patient, leaving aside the benefits to the surgeon's pride. Any abdominal organ can rupture as a result of inflammation, ischaemia, or trauma and flood the peritoneum with blood, bile, urine, or intestinal contents. It is always worth trying to decide which fluid is present and from whence it arises.

Perforation of the gut with leakage of intestinal contents is the most common problem. Bowel content is the most irritant substance and gives rise to all the classical symptoms and signs of peritonitis. The three common causes are a perforated peptic ulcer, perforated diverticular disease, and perforated appendicitis. Stercoral perforation of the colon due to severe constipation is a rare variant of perforated diverticulitis. Distinguishing between the various causes depends on taking a very careful history to try and decide the symptoms at the onset of the illness and an equally careful evaluation of the abdominal signs. Some patients are simply too ill to co-operate. Patients with a perforated ulcer may give a history of indigestion or the consumption of ulcerogenic drugs. Sometimes it is possible to decide that the signs are worst in one particular area of the abdomen.

Blood, bile, and urine are all rather less irritant to the peritoneum and give rise to less marked physical signs. Blood, which usually comes from a ruptured spleen, an ectopic pregnancy, or a ruptured aortic aneurysm characteristically gives rise to tenderness and rebound tenderness but very little guarding or rigidity. Uninfected urine and bile can both be present in the peritoneum in large amounts with very few signs at all, although once infection is present then the signs are usually very marked.

Obstruction

Intestinal obstruction

Abdominal pain, vomiting, abdominal distension, and constipation are a quartet of very obvious symptoms and signs that make the diagnosis of intestinal obstruction easy. If the patient's presentation to the doctor is delayed, the hyperactivity of the bowel that is so obvious at the onset of the obstruction may have been replaced by paralysis, with disappearance of the colicky pain and any visible peristalsis.

Not every patient with obstruction requires immediate surgery; it is important to try and decide on the cause. Uncomplicated obstruction due to adhesions, which is now the most common problem, is best treated conservatively to begin with, as many episodes will subside spontaneously. An obstructed hernia, which means occlusion of the lumen of the bowel within the hernia, requires a prompt operation. Strangulation implies impairment of the blood supply to the bowel and also demands an urgent operation. It is indicated by a shocked patient with peritonitis as well as the signs of obstruction, although these will not be present if only omentum or extraperitoneal fat are trapped in the hernia. An inguinal or incisional hernia should be obvious, but it is extraordinarily easy to overlook a small, tense, femoral hernia in a fat patient. The various rare intra-abdominal hernias are rarely diagnosed before surgery.

Both adhesions and hernias usually obstruct the small bowel. Obstruction of the large bowel is most commonly due to cancer, usually of the colon but sometimes of the ovary. Volvulus of the bowel, either large or small, initially causes obstruction, but strangulation rapidly supervenes if the bowel is not untwisted.

If there is doubt about the reality or the severity of an obstruction, the use of contrast radiology is always helpful. A small-bowel enema is a useful way to decide whether an obstruction of the small bowel needs an operation. A conventional barium enema will help to plan an operation for an obstruction of the large bowel and will identify patients with pseudo-obstruction.

Rare causes of obstruction, such as gallstone ileus, bolus obstruction from worms or foodstuffs, an obturator hernia, and malrotation, are usually diagnosed at the laparotomy, although there is enormous satisfaction in making such a diagnosis in advance. It is sometimes possible if the diagnosis is considered and the relevant symptoms and signs sought.

Acute retention of urine

Acute retention of urine also presents with abdominal distension and pain but the patient will volunteer that they have not passed urine for some time. Most patients are male and elderly. The tense, distended bladder is easy to see and to feel, and there is instant relief when a catheter is passed. The soft, flaccid bladder associated with chronic retention, which sometimes presents acutely, is much more difficult to feel or to percuss, and ultrasonography is often of help. Sympathomimetic drugs,

sometimes in small doses in proprietary medicines, and constipation, are two factors that can precipitate acute retention in susceptible people.

Patients with pelvic peritonitis are sometimes sent into hospital with a diagnosis of acute retention. The inflammation has led to dehydration and severe oliguria along with lower abdominal pain and distension. The true diagnosis is often not made until a catheter is passed and only a small amount of urine is retrieved.

Colic

Abnormal contraction of smooth muscle causes the regular and intermittent pain that is called colic. Within the abdomen only the gut, the renal tract, the uterus, and the biliary tract cause such a pain. The site and the distribution of the pain are different in each case and so they are easy to tell apart except, sometimes, for right-sided renal colic and biliary pain. The most severe intestinal colic accompanies gastroenteritis, although it is a classic sign of small-bowel obstruction. Stones are the common cause of renal and biliary colic, although blood clot and pus can cause ureteric colic quite as severe as that due to a stone. Most, but not all, women know when they are pregnant and are about to deliver a baby, but uterine colic due to a miscarriage, or even severe dysmenorrhoea, sometimes present as acute abdominal pain.

A major catastrophe

A ruptured abdominal aortic aneurysm, acute haemorrhagic pancreatitis, and acutely ischaemic bowel are the three common explanations causing a patient to present at hospital severely shocked immediately after the onset of acute abdominal pain. Shock is also seen in any patient with significant intra-abdominal pathology who presents some time after the onset of their symptoms, but the history will usually identify the delay and lead to the correct diagnosis.

A ruptured aneurysm is never easy to feel in a hypotensive patient and severe acute pancreatitis is sometimes only diagnosed at laparotomy. Patients with dead bowel inside the abdomen look and sometimes smell as though they are dying, which in one sense is correct. Those with strangulated gut usually show signs of obstruction, while a mesenteric embolus is easy to diagnose if the patient is fibrillating but very difficult if the thrombus lies on the endocardium after a silent myocardial infarct.

A mass

Occasional patients with abdominal pain who present acutely have a mass in the abdomen on palpation. The first rule is to try and define the organ from which the lump arises and then to describe all the physical characteristics of the lump, remembering particularly to test for pulsation. An ovarian cyst and an obstructed loop of bowel are common causes of a lump; apart from these, lumps are rare. They can arise from any intra-abdominal organ. Most obscure abdominal masses are subjected to intense preoperative investigation; in reality they all require exploration at laparotomy, when their nature and their origin will be revealed. Even then, all the preoperative predictions are sometimes found to be wrong.

Specific diagnostic problems

Abdominal trauma

The injured patient with a rapidly distending abdomen and signs of exsanguination obviously requires an immediate laparotomy. However, it is often difficult to decide whether a trauma victim has significant intra-abdominal injury. Most such patients have other injuries, a few are unconscious, and alcohol is a factor in some. Complaints of abdominal pain should always be taken seriously, even though few patients can give a good history and abdominal examination is limited because the normal responses are impaired. Furthermore, some abdominal injuries, particularly intestinal injuries, do not become apparent immediately: repeated reassessment is therefore essential.

It is always worthwhile to establish the mechanism of injury as this can give some indication of the likely damage. Direct, blunt trauma in a car crash or from the handlebars of a pedal cycle, kicks, punches, and falls from a height can all rupture the kidney, the spleen, the liver, the pancreas or the bowel. A lap and diagonal seat-belt will tear the small-bowel mesentery if the lap belt lies too high across the abdomen. A fractured pelvis can perforate the bladder, and is often associated with rupture of the urethra.

The feasibility of a complete examination will vary, but the abdominal surgeon should certainly pay attention to the general condition of the patient, the other injuries sustained, and, in particular, to the chest. Unexplained and persistent tachycardia and hypotension despite active resuscitation implies continuing haemorrhage. If the site of bleeding is not visible the blood is either in the chest or the abdomen. Fractured ribs can damage abdominal organs as well as the lungs.

Peritoneal lavage is a useful indicator of the need for a laparotomy. Frank blood, intestinal contents, or heavily bloodstained lavage fluid clearly indicate the need for a laparotomy. The kidney is the most frequently injured organ in blunt abdominal trauma, and intravenous urography is essential in the management of traumatic haematuria. Too many surgeons have removed an injured but solitary kidney. CT of the abdomen can identify retroperitoneal injuries that are not apparent on clinical examination and will also demonstrate the extent of an injury within a solid organ.

In civilian life, penetrating abdominal wounds are less common than blunt injury, but most will require surgical exploration. It is usually impossible to tell how far a wound has penetrated and no simple investigation, except perhaps laparoscopy, is helpful.

Common sense must be applied: in the trauma victim it is almost always better to ‘look and see' rather than to ‘wait and see' if there is a continuing possibility of a significant intra-abdominal injury.

Acute abdominal pain in children

Nine out of 10 children with abdominal pain either have acute appendicitis or non-specific abdominal pain (Table 4). The other child is likely to have either a urinary-tract infection or intussusception. Urinary infection presents with the classical symptoms of fever, frequency, and stinging pain on micturition. The diagnosis is only confirmed when pus cells are found on microscopy of a correctly collected urine sample and a significant number of pathogenic bacteria grow on culture. In contrast, intussusception rarely presents with all the classical signs of colicky abdominal pain, a palpable abdominal mass, and redcurrant-jelly stools. Any infant—most patients are under 5 years of age—who develops severe, intermittent, acute abdominal pain manifested as acute episodes of screaming should be suspected of harbouring an intussusception. An ultrasonographic examination will confirm the diagnosis and an enema either with barium, which can also make the diagnosis, or with air may also reduce the intussusception.

	Percentage of cases
Non-specific abdominal pain	62
Acute appendicitis	32
Urinary-tract infection	2
Intussusception	1
Miscellaneous	3

Table 4 Causes of acute abdominal pain in children seen in hospitals in the developed world (after de Dombal, 1991)

Two conditions that may cause non-specific abdominal pain in children deserve particular comment. Constipation is often said to be the most common cause of abdominal pain, although few such patients reach hospital. In young children, in whom a rectal examination is to be avoided, it is perhaps permissible to make the diagnosis on a plain abdominal radiograph, although even this is unnecessary if faecal masses can be felt in the abdomen. Secondly, examination of the ears and the throat are as essential as palpation of the abdomen in young children: both tonsillitis and otitis media can present with abdominal pain.

The acute abdomen in the tropics

Acute appendicitis and non-specific abdominal pain are the most common causes of abdominal pain all over the world. Disease patterns vary, however, and few patients with non-specific pain are admitted to a hospital in the Third World. Of the specifically tropical diseases only two are of real concern. Firstly, infestation with worms is a significant cause of abdominal colic, and sometimes of intestinal obstruction. Secondly, pain and tenderness in the right upper quadrant are more likely to be caused by amoebic hepatitis with a liver abscess rather than acute cholecystitis. Amoebic colitis can present with all the signs of peritonitis but the typical shallow shaggy ulcers seen on sigmoidoscopy should make the correct diagnosis clear. When the inflammation is confined to the caecum (typhlitis), the signs mimic acute appendicitis.

Unusual problems

The abdominal wall

Rectus-sheath haematoma

This rare condition can mimic intra-abdominal pathology. The haematoma develops from rupture of the inferior epigastric artery in the lower half of the abdomen. Pregnant women are particularly affected; there is sometimes a history of injury, and the right side is involved twice as commonly as the left.

The onset of the pain is acute and it is often accompanied by nausea and vomiting. There is marked tenderness in the iliac fossa so that it is easy to misdiagnose appendicitis. Two physical signs will reveal the true diagnosis. It may be possible to show that the tenderness, and the swelling if there is one, are confined to the abdominal wall. Secondly, bruising of the skin may be visible. Sometimes this only appears a few days later and it is often at a distance from the site of maximum tenderness. Ultrasonography and CT will both make the diagnosis for certain.

Once diagnosed the haematoma will slowly resolve with rest; in patients who undergo surgery the diagnosis becomes apparent as the abdominal incision is made. It is then worth tying off the bleeding vessel; this is also needed in the rare patient in whom the haemorrhage does not stop.

Similar bleeding sometimes arises from spontaneous rupture of an intercostal artery. The dramatic bruising of the abdominal wall spreading round from the lower thorax, often in a segmental distribution, is unmistakeable. The haematoma will normally resolve with rest.

Abdominal-wall hernia

Small epigastric, periumbilical, and tiny incisional hernias, which usually contain only extraperitoneal fat, can be very hard to find and are surprisingly painful, even without strangulation of the contents. Epigastric hernias often cause peculiar digestive symptoms, which can lead to diagnostic confusion.

Medical causes of abdominal pain

Many patients who need an operation for acute abdominal pain also have other medical problems, and some strictly medical conditions are treated by surgeons simply because they usually present as a surgical problem. Here we consider those patients who present with an acute abdomen but the underlying cause requires the care of a physician.

Pulmonary problems

Inflammation of the pleura at the base of the lungs can present as abdominal pain because of irritation of the lower intercostal nerves that supply the abdominal wall. Pneumonia in the very young and the very old, and pulmonary embolism at any age, can present in this way. Usually there are some symptoms and signs referrable to the chest, although in infants pneumonia may only be diagnosed on a chest radiograph.

Occasionally the chest abnormality is itself secondary to pathology below the diaphragm. Rarely, perforated diverticular disease produces minimal symptoms from the peritonitis but leads to some subphrenic infection. Pneumonia supervenes because of poor diaphragmatic movement and by the time the patient comes to the hospital the abdominal symptoms are long since forgotten. Conservative management of the abdominal problem is usually appropriate anyway, so the diagnostic error is rarely of serious consequence. Cholecystitis can sometimes present in the same way.

Cardiac causes

Occasionally, pain from an acutely distended liver is the presenting feature of right-sided heart failure, although other physical signs are usually also present. Myocardial infarction is said, on occasion, to present with continuous epigastric pain. The main distinguishing feature is the complete lack of any epigastric tenderness.

Drugs

Warfarin and digoxin are the two common drugs that cause abdominal pain, although neither is as widely used as in the past. Digoxin toxicity presents with abdominal pain and vomiting. Warfarin anticoagulation can cause a spontaneous retroperitoneal haemorrhage with an associated paralytic ileus and severe abdominal pain radiating to the back. Haematuria is a clue to the diagnosis and the prothrombin time is usually excessively prolonged.

Diabetes

Very rarely, diabetes mellitus first manifests itself with acute abdominal pain. The diagnosis is rarely difficult because the patient, who is usually an adolescent, is obviously severely ill with impairment of consciousness and marked dehydration. Sometimes the sweet, ketotic smell typical of hyperglycaemic ketoacidosis pervades the whole examination room.

Hepatitis

Inflammation of the liver can sometimes cause intrahepatic cholestasis. If the acutely distended liver is also painful it is easy to think that the patient has extrahepatic obstructive jaundice. Both drugs and viral infections can produce this clinical picture, although the matter is resolved as soon as the bile ducts are of normal size on an ultrasonographic examination.

Blood

About half the children with Henoch–Schönlein purpura complain of abdominal pain. The pain is due to bleeding into the wall of the bowel and this can lead to an intussusception. The diagnosis will only be missed if the characteristic skin rash is absent. Spontaneous and painful intra-abdominal haemorrhage is also a feature of haemophilia.

Spurious abdominal pain

There are a few strange people in every society who enjoy being admitted to a hospital. Some of them are looking for a bed, some want drugs, some like attention, and a few need help. Most are well known to their own local medical services and some even achieve a wider notoriety.

Such patients usually present themselves in the accident and emergency department and give a dramatic history of some acute abdominal event. Renal colic, a potential rupture of the spleen, and a fall astride a bar are all popular. They can always simulate the necessary physical signs. They often appear in very severe pain. A tiny cut on the lip, a finger, or the genitalia is quite sufficient to produce haematuria. They will also claim to be allergic to intravenous contrast media so that it is impossible to confirm a diagnosis of renal colic. Many are admitted for observation and it may be some days before the true diagnosis is revealed.

Many such patients are simply too good to be true. They fail to realize that it is exceptionally rare for any patient to have every recorded symptom of their chosen diagnosis. Most patients give an address that is a long way away and some describe themselves as lorry drivers. Other clues are early and repeated demands for

opiate analgesia and many previous admissions to hospital. In the classical case of Munchausen syndrome there are also multiple abdominal scars.

Even when the diagnosis is suspected it is difficult to confront the patient, since doctors are naturally loath to accuse patients of lying. It is better to investigate the background by speaking on the telephone to doctors who have treated the patient in the past, either in a hospital or in the community, and making it clear to the patient that this is taking place. Demands for pain relief should be met by the offer of non-opiate analgesics. Most patients simply leave the hospital, often without telling the staff, once this process of enquiry starts or when they fail to obtain the service that they were seeking.

Conclusion

The patient with acute abdominal pain always presents the surgeon with the challenge of making the correct diagnosis. Without that diagnosis the right treatment cannot be offered.

Making the correct diagnosis is never easy. It demands attention to detail in taking the history and examining the patient, and clarity of thought in analysing the information that is obtained. Investigations may help, but in many places in the world there are no facilities for further investigation. There the management of every patient depends entirely on the clinical skills of the doctor. Finally, every surgeon should remember that for every five patients admitted to a hospital with abdominal pain at least one has his or her diagnosis changed during the course of their admission and two leave the hospital without being given a cause for their pain. This does not mean that there was no cause. It does mean that our skill in making a diagnosis needs to be improved.

Further reading

Adams ID *et al.* Computer aided diagnosis of acute abdominal pain: a multicentre study. *British Medical Journal* 1986; **293**: 800–4. [A study of 16 737 patients confirms that computer aided diagnosis improves diagnostic accuracy, decision-making and patient outcome.]

Bailey H, Love M. Vermiform appendix. In: *A short practice of surgery*, 12th edn, pp. 978–1002. Lewis, London, 1962. [A classical account of the diagnosis and treatment of appendicitis.]

Banarjee A. Factitious disorders presenting as acute emergencies. *Postgraduate Medical Journal* 1994; **70**: 68–73. [A review of medical and surgical factitious illness.]

Campbell JPM, Gunn AA. Plain abdominal radiographs and acute abdominal pain. *British Journal of Surgery* 1988; **75**: 554–6. [A review of 5080 patients with acute abdominal pain, with particular reference to the 45 per cent who had a plain abdominal radiograph taken.]

Cope Z. *The early diagnosis of the acute abdomen*, 9th edn. Oxford University Press, 1948. [Any early edition of this classical text is as relevant today as it was when it was first written in 1921.]

de Dombal FT. *Diagnosis of acute abdominal pain*, 2nd edn. Churchill Livingstone, Edinburgh, 1991. [A compilation of scientific studies of abdominal pain completed over 20 years.]

de Dombal FT. Acute abdominal pain in the elderly. *Journal of Clinical Gastroenterology* 1994; **19**: 331–5. [A comparative review of acute abdominal pain in patients under and over the age of 50 years.]

Editorial. Analgesia and the acute abdomen. *British Medical Journal* 1979; **279**: 1093. [Urgent relief of pain is good treatment and is, nowadays, unlikely to do harm by delaying diagnosis.]

Gallegos N, Hobsley M. Abdominal pain: parietal or visceral? *Journal of the Royal Society of Medicine* 1992; **85**: 379. [How to distinguish pain arising from the abdominal wall from pain arising within the abdomen.]

Gray DWR, Collin J. Non-specific abdominal pain as a cause of acute admission to hospital. *British Journal of Surgery* 1987; **74**: 239–42. [A review of the causes of non–specific abdominal pain.]

Gray DWR, Dixon JM, Collin J. The closed eyes sign: an aid to diagnosing non-specific abdominal pain. *British Medical Journal* 1988; **297**: 837. [The first description of a new physical sign.]

Hoffman J, Rasmussen OO. Aids in the diagnosis of acute appendicitis. *British Journal of Surgery* 1989; **76**: 774–9. [A review of all the tests for diagnosing acute appendicitis.]

Jones DJ. Appendicitis. *British Medical Journal* 1992; **305**: 44–7. [How to diagnose and treat appendicitis.]

Jones PF. Active observation in management of acute abdominal pain in childhood. *British Medical Journal* 1976; **274**: 551–3. [Active observation of children with abdominal pain helps to make a diagnosis and is safe.]

Jones PF. *Emergency abdominal surgery*, 2nd edn. Blackwell Scientific Publications, Oxford, 1987. [A comprehensive text with additional information about the management of children with abdominal pain.]

McAdam WAF, Brock BM, Armitage T, Davenport P, Chan M, de Dombal FT. Twelve years' experience of computer-aided diagnosis in a district general hospital. *Annals of the Royal College of Surgeons of England* 1990; **72**: 140–6. [Computer aided diagnosis maintained improvements in diagnosis and treatment for 10 years.]

Macartney FJ. Diagnostic logic. *British Medical Journal* 1987; **295**: 1325–31. [A detailed review of how doctors make a diagnosis.]

Martin RF, Rossi RL, ed. Abdominal emergencies. *Surgical Clinics of North America* 1997; **77**: 1226–470. [An up-to-date series of articles on the diagnosis and management of the acute abdomen with an American perspective.]

Masters K, Levine BA, Gaskill HV, Sirinek KR. Diagnosing appendicitis during pregnancy. *American Journal of Surgery* 1984; **148**: 768–71. [A review of 36 patients undergoing laparotomy for suspected appendicitis in pregnancy.]

Mohammed R, Goy JA, Walpole BG, Brentnall EW, Miller T. Munchausen's syndrome. A study of the casualty ‘Black books' of Melbourne. *Medical Journal of Australia* 1985; **143**: 561–3. [A review of 21 cases of Munchausen syndrome with a further 20 probable cases.]

O'Shea B, McGennis A, Cahill M, Falvey J. Munchausen's syndrome. *British Journal of Hospital Medicine* 1984; **1**: 269–74. [A review of Munchausen syndrome with guidance on management.]

Paterson–Brown S. Strategies for reducing inappropriate laparotomy rate in the acute abdomen. *British Medical Journal* 1991; **303**: 1115–18. [A review of the invasive and non-invasive techniques that help in the diagnosis of the acute abdomen.]

Paterson–Brown S, Vipond MN. Modern aids to clinical decision making in the acute abdomen. *British Journal of Surgery* 1990; **77**: 13–18. [An up-to-date review of all the aids to making a diagnosis in patients with acute abdominal pain.]

Puylaert JBCM *et al.* A prospective study of ultrasonography in the diagnosis of appendicitis. *New England Journal of Medicine* 1987; **317**: 666–9. [Ultrasound diagnosis of acute appendicitis was correct in 39 of 52 patients. It was 100 per cent accurate in excluding appendicitis in 31 patients.]

Sherck JP, McCort JJ, Oakes DD. Computed tomography in thoraco-abdominal trauma. *Journal of Trauma* 1984; **24**: 1015–21. [Thoracoabdominal CT scanning of stable trauma patients accurately identifies injury to the liver, spleen, kidney, and the retroperitoneum.]

Siddiqui MN, Abid Q, Qaseem T, Hameed S, Ahmed M. ‘Spontaneous' rectus sheath haematoma: a rare cause of abdominal pain. *Journal of the Royal Society of Medicine* 1992; **85**: 420–1. [CT is useful in the diagnosis of rectus-sheath haematoma.]

Stewart J, Finan PJ, Courtney DF, Brennan TG. Does a water soluble contrast enema assist in the management of acute large bowel obstruction: a prospective study of 117 cases. *British Journal of Surgery* 1984; **71**: 799–801. [A water-soluble contrast enema confirms the diagnosis of mechanical obstruction and prevents unnecessary surgery in patients with pseudo-obstruction.]

Thomson HJ, Jones PF. Active observation in acute abdominal pain. *American Journal of Surgery* 1986; **152**: 522–5. [A prospective study of 220 patients showed that a period of active observation reduced the ‘negative laparotomy' rate.]

Williamson RCN, Cooper MJ, ed. *Emergency abdominal surgery*. Vol.17, *Clinical surgery international*. Churchill Livingstone, Edinburgh, 1990. [Multiple chapters that focus on the surgical treatment of the acute abdomen.]

34.2 The peritoneum and peritonitis

D. McWhinnie

Anatomy of the peritoneal cavity
Peritonitis
- **Localized or generalized peritonitis**
- **Signs and symptoms**
- **Acute peritonitis**
- **Chronic (sclerosing) peritonitis**

Further reading

Anatomy of the peritoneal cavity

The peritoneal cavity is lined by the peritoneal membrane that consists of a single layer of mesothelial cells supported by loose connective tissue. The membrane comprises two components in continuity: the visceral peritoneum covers the abdominal organs and the parietal peritoneum lines the abdominal wall (Fig. 1). The cavity is thus a closed sac except for the fimbriated ends of the fallopian tubes. The mesothelial cells secrete serous fluid which provides a lubricant to allow gliding movements of the viscera. In the normal healthy adult the peritoneal cavity contains about 100 ml of serous fluid, secreted from a total peritoneal surface area of approximately 2 m^2. This large surface area of thin semipermeable peritoneal membrane can be utilized for peritoneal dialysis.

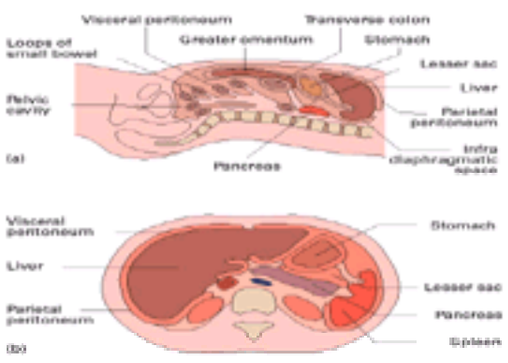


Fig. 1. Peritoneal cavity. (a) Coronal section; (b) transverse section showing the dependency of pelvic cavity and infradiaphragmatic space in the recumbent patient. The visceral and parietal peritoneum are shown lining the abdominal organs and abdominal wall.

The technique of peritoneal dialysis has been used in the treatment of renal failure since the late 1940s but was of limited intermittent application until the development of the implanted Tenckhoff silicon rubber catheter in 1968 (see Chapter 17.16). This allowed peritoneal dialysis to be performed continuously and safely for longer periods, commonly of several years' duration. Dialysis fluid is stored in sterile PVC bags and during an ‘exchange' the bag is connected to the Tenckhoff catheter by PVC tubing. The dialysate is encouraged to enter the abdominal cavity by gravity, through elevation of the bag. Between three and five exchanges of 1 to 2 litres of peritoneal dialysate are performed daily. Clearance of particular electrolytes is achieved by reducing their concentration in the dialysate. Fluid balance is monitored by the measurement of daily weights and measured by altering the proportion of ‘hypertonic' (3.8 per cent glucose solution) and ‘isotonic' (2.36 per cent glucose solution) exchanges. While the dialysis fluid is in the peritoneal cavity, the uraemic toxins diffuse from the blood supply of the mesentery and peritoneum. At the end of each dialysis cycle the dialysate is drained by gravity and returned to the bag. This simple technique of continuous ambulatory peritoneal dialysis (**CAPD**) has been adopted in many dialysis centres as initial therapy for renal failure in patients with diabetes, in young patients with prospects of an early transplant, and in those patients with poor vascular access.

The anatomy of the abdominal cavity determines the pattern of fluid movement within the abdomen. Functionally, four compartments exist within the peritoneal cavity: the pelvis, the right and left paracolic gutters, and the infradiaphragmatic space (Fig. 1). As the paracolic gutters slope into the infradiaphragmatic space superiorly and over the pelvic brim inferiorly, fluid in the recumbent patient collects under the diaphragm and in the pelvis, forming pelvic or subphrenic abscesses (see Chapter 34.2).

The pattern of innervation of the visceral and parietal peritoneum determines the pattern of abdominal pain that occurs when the peritoneum is inflamed. As the visceral peritoneum receives only afferent innervation from the autonomic nervous system, stimuli from the visceral peritoneum are localized only to the foregut, midgut, or hindgut in the midline. The visceral nerves do not mediate pain or temperature but respond to distension, traction, or pressure. In contrast, the parietal peritoneum is innervated by both somatic and visceral afferent fibres which allow the acute localization of pain when the parietal peritoneum is inflamed. The differing effects of visceral and parietal irritation can be seen in the clinical example of acute appendicitis (see Chapter 27.1). Initially the pain is periumbilical (visceral midgut due to distension of the appendix) before moving to the right iliac fossa as the adjacent parietal peritoneum is irritated.

Peritonitis

Localized or generalized peritonitis

Peritonitis is defined as inflammation of the peritoneal cavity, where the peritoneal fluid increases in volume with the passage of a transudate rich in leucocyte polymorphs and fibrin. Initially, peritoneal inflammation is often localized and the affected area contained by a wrapping of greater omentum, adjacent bowel, and fibrinous adhesions. If the inflammatory focus is part of an ongoing process, or if host defences are lowered, localized peritonitis may progress to life-threatening generalized peritonitis. There is massive exudation of inflammatory fluid into the peritoneal cavity causing hypovolaemia, often compounded by toxaemia from absorbed products and septicaemia if infection is present. Diffuse peritoneal irritation causes peristaltic paralysis with the cessation of bowel motility.

Signs and symptoms

The clinical features of peritonitis are dependent on both the aetiology and the progression of the inflammation. The early manifestations of peritonitis following disease of an abdominal viscus are characterized by the primary disease process itself and a detailed discussion can be found in the appropriate chapter. An overview of the signs and symptoms of the acute abdomen are detailed n Chapter 34.1. In summary, the signs and symptoms of peritonitis secondary to a diseased abdominal viscus are as follows. Initial inflammation of the visceral peritoneum overlying the damaged organ results in poorly localized pain, through stimulation of the visceral autonomic nerves. Irritation of the nearby parietal peritoneum results in localization of the pain. Associated symptoms include malaise, nausea and vomiting, and a low-grade fever. The four cardinal signs of peritonitis, consisting of tenderness, guarding, rigidity, and rebound tenderness may also be elicited. When the peritonitis is generalized the patient is clearly unwell, with marked fever, dehydration, and absent bowel sounds. Pain is diffuse throughout the abdomen and is exacerbated by even the slightest movement. Shoulder-tip pain is diagnostic of diaphragmatic inflammation.

Acute peritonitis

Acute peritonitis may be classified as primary (spontaneous), where an infection has arisen *de novo* within the peritoneum, or secondary, where the inflammatory process involving the peritoneum is the result of an identifiable primary process (Table 1).

Acute peritonitis
Primary (spontaneous)
Secondary
Acute suppurative
Granulomatous
Chemical (aseptic)
Interventional
Traumatic
Drug-induced
Chronic (sclerosing) peritonitis
Infectious
Drug-induced
Chemical
Foreign-body
Carcinomatosis

Table 1 Aetiology of peritonitis

Primary (spontaneous) peritonitis

Idiopathic peritonitis is uncommon, constituting about 1 per cent of all cases of peritonitis and occurs when no obvious source for the peritoneal infection can be demonstrated. It is a diagnosis by exclusion and is confirmed in retrospect when the results of blood cultures or peritoneal swabs become available. Formerly, pneumococci were implicated, but in recent years haemolytic streptococci, *Escherichia coli*, and *Klebsiella* spp. are now more frequently cultured.

It was classically described in young girls where the port of entry was presumed to be through the fallopian tubes. Adult primary peritonitis arises via haematogenous spread or translocation of bacteria through the bowel wall, especially in the presence of exogenous (e.g. steroid therapy) or endogenous (intercurrent disease) immunosuppresssion. Pneumococcal infection is well recognized in children after a splenectomy and in those who have nephrotic syndrome. Adults with cirrhosis are prone to a wide variety of spontaneous bacterial infections because of the proteinaceous nature of the ascitic fluid which provides an excellent culture medium for bloodborne bacteria to proliferate.

Successful treatment of primary peritonitis, without resorting to laparotomy or laparoscopy, is rare. Only when the peritoneal tap and culture of peritoneal fluid reveals a non-enteric organism can conservative antibiotic therapy be instituted with caution.

Secondary peritonitis

Acute suppurative peritonitis

This is the most common form of peritonitis encountered by the surgeon and results from the perforation of a viscus (e.g. appendix, peptic ulcer, colonic diverticulum, or gallbladder), ischaemia of an intra-abdominal organ (e.g. strangulated hernia, volvulus, mesenteric artery occlusion), or extension of an existing infection of an abdominal organ (e.g. appendix abscess, liver abscess, pyosalpinx). Surgical intervention is the treatment of choice and in most cases is mandatory, being aimed at the primary disease process. Supportive therapy before laparotomy includes opiate analgesia, antibiotic therapy (against both aerobes and anaerobes), correction of hypovolaemia and electrolyte imbalance by intravenous fluid replacement, and nasogastric aspiration to reduce abdominal distension and prevent aspiration pneumonia through inhalation of vomit.

Granulomatous peritonitis

The presence of chronic granulomas within the abdominal cavity in association with peritonitis is the result of infection, most commonly with tuberculosis but occasionally with fungi such as *Candida albicans*. Tuberculous peritonitis is usually associated with a tuberculous focus within the abdomen (e.g. intestine, mesenteric lymph node, fallopian tubes), but occasionally can develop by haematogenous spread from a focus outside the abdominal cavity, such as in the lung. The causative agent is *Mycobacterium tuberculosis*, nowadays usually of the human rather than the bovine strain, following widespread pasteurization of milk. Although the incidence of the disease is low in Western countries, occurring in the chronically ill or malnourished patient, it is still prevalent in the Indian subcontinent and in the Far East. Immigrants from these locations remain susceptible to the condition. Females are twice as commonly affected as males but all age groups are involved.

Pathologically, tuberculous peritonitis may present with ascites or with intra-abdominal adhesions, but there is often considerable overlap between the two forms of the disease. Clinical presentation is insidious with vague generalized abdominal pain, low-grade fever, anorexia, and weight loss. The patient with ascites may also present with gross abdominal distension, while those with intra-abdominal adhesions may present with pain and distension or bowel obstruction. Preoperative diagnosis is uncommon. Tuberculous ascites is difficult to differentiate from other forms of ascites. The leucocyte count may demonstrate a lymphocytosis and lymphocytes also predominate in the ascitic fluid. Centrifugation and direct microscopy of the ascitic fluid rarely demonstrate the tubercle bacillus. Culture of the ascitic fluid is positive for tuberculosis in less than one-half of all cases. At laparotomy, the entire peritoneal cavity is studded with tuberculous nodules and the initial differential diagnoses include carcinomatosis, starch peritonitis, and widespread fat necrosis. In the patient who presents with subacute or acute bowel obstruction, widespread fibrinous adhesions with matting intestines and omentum predominate. Here the differential diagnosis includes other forms of sclerosing peritonitis (see below). Confirmation of tuberculous peritonitis is made by histopathology and treatment consists of combination chemotherapy (rifampicin, isoniazid, and ethambutol) for up to 12 months.

Chemical (aseptic) peritonitis

Aseptic peritonitis refers to the peritoneal inflammation from substances other than bacteria, but bacterial contamination and overgrowth soon follow. A perforated peptic ulcer provides the most severe and common form of chemical peritonitis with gastric juice and bile contaminating the peritoneal cavity. Biliary peritonitis alone may follow gangrene and perforation of the gallbladder, or, after open or laparoscopic cholecystectomy, may be the result of an unrecognized division of an accessory hepatic duct, an insecure ligature or clip on the cystic duct remnant, delayed necrosis of part of the common bile duct wall following injudicious use of cautery, or displacement of a T-tube following exploration of the common bile duct. Blood in the peritoneum is also a cause of peritoneal irritation after slow bleeding (e.g. a ruptured graafian follicle or following splenic injury) rather than from a catastrophic haemorrhagic event such as ruptured aneurysm where the primary pathology itself overshadows the peritoneal irritation. Meconium and urine may also precipitate chemical peritonitis.

Interventional peritonitis

Endoscopy of the gastrointestinal tract may precipitate acute peritonitis through colonoscope perforation of a diverticulum or inadvertent perforation of the oesophagogastric junction during oesophageal dilatation. Similarly, the urinary bladder may be perforated during diagnostic or therapeutic cystoscopy leading to a urinary leak within the abdomen. The recent expansion of interventional radiological procedures has precipitated a multitude of assaults on the abdominal cavity, such as CT-guided biopsy and drainage of abscesses, mesenteric angiography and therapeutic embolization, and percutaneous transhepatic cholangiography and stenting, all providing further potential for peritonitis.

Peritonitis may follow abdominal surgery where bowel and gastric contents, blood, and urine escape into the abdominal cavity following anastomotic dehiscence. In patients with renal failure treated by continuous ambulatory peritoneal dialysis, a permanent indwelling catheter in the abdominal cavity provides a portal of entry for exogenous bacteria despite the use of stringent aseptic techniques during dialysis exchanges (see Chapter 17.16).

Traumatic peritonitis

Abdominal trauma may produce acute peritonitis in several ways. Penetrating wounds of the abdomen without visceral injury may provide a route for exogenous bacterial contamination.

Penetration of a visceral organ may precipitate the spillage of visceral contents into the peritoneal cavity. Several blunt trauma may disrupt intra-abdominal organs directly or indirectly through disruption of their vascular supply.

Drug-induced peritonitis

Warfarin anticoagulation can cause peritoneal irritation and peritonitis through leakage from a spontaneous retroperitoneal haematoma. The symptoms of acute

peritonitis have also been described during treatment with the antituberculous agent, isoniazid.

Chronic (sclerosing) peritonitis

Chronic or sclerosing peritonitis is the end result of a heterogeneous group of conditions (Table 1). Classically, sclerosing peritonitis is characterized by dense adhesions, especially between loops of small bowel, and in the most extreme cases the entire small bowel and even the large intestine and liver are cocooned in a dense adhesive membrane of fibrous tissue (Fig. 2). The classic description of sclerosing peritonitis relates to the b-adrenoceptor blocker, practolol. This drug was first introduced in 1970, but the first reports of thickening of serosal membranes appeared 4 years later and the drug was withdrawn from general use in 1976. During this 6 year period approximately 200 patients worldwide were diagnosed as having sclerosing peritonitis. Patients presented with subacute small bowel obstruction or acute-on-chronic small bowel obstruction. Treatment consisted of surgical stripping of the cocoon of fibrous tissue from the underlying intestine and was a long and tedious procedure. Intra-abdominal fistula formation and death was not uncommon. In the early 1980s sclerosing peritonitis made a reappearance in patients undergoing continuous ambulatory peritoneal dialysis. Patients presented with impaired ultrafiltration capacity due to the deposition of fibrous membrane on the peritoneum, and the clinical features of bowel obstruction, either acute or chronic. Epidemiological studies demonstrated that the sclerosis was due to the chlorhexidine and alcohol solution used to sterilize the connectors of the CAPD catheters at the time of peritoneal dialysis ‘exchanges'. Approximately 1 ml of antiseptic solution was able to enter the peritoneal cavity at the time of each exchange causing irritation to the peritoneal membrane. Other mechanisms which may precipitate sclerosing peritonitis in patients treated by CAPD include the presence of acetate in the dialysate and recurrent CAPD peritonitis with fibrin deposition over the peritoneal membrane.

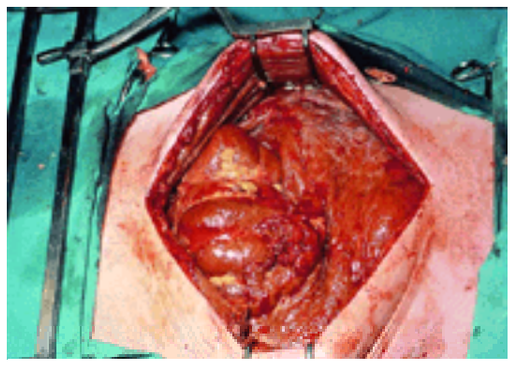


Fig. 2. Sclerosing peritonitis following CAPD. The small bowel, large bowel, and liver are cocooned in a fibrous membrane. (By courtesy of Dr B. Junor.)

Foreign-body granulomatous peritonitis was not uncommon in patients who had undergone laparotomy in the days before the introduction of ‘talc-free' surgical gloves. The main component of surgical talc is starch and starch peritonitis was first described in 1956 and is due to starch sensitivity. Some 2 to 6 weeks after an initial laparotomy from which the patient usually makes an uneventful recovery, the patient presents with either ascites or signs and symptoms of bowel obstruction. At laparotomy, multiple nodules are seen over the parietal and visceral peritoneum, mimicking malignant deposits with dense adhesions between loops of bowel. Treatment consists of lysis of adhesions and the diagnosis is confirmed by histology where birefringent granules are observed. Administration of steroids is ineffective in preventing intestinal adhesions.

Other forms of chronic peritonitis which may form dense adhesions include the adhesive form of tuberculous peritonitis (see above) and carcinomatosis.

Further reading

Brown P, Baddeley H, Read AE, Davies JD, McGarry J. Sclerosing peritonitis, an unusual reaction to a b-adrenergic-blocking drug. *Lancet* 1974; **ii**: 1477–81. [The first report of sclerosing peritonitis.]

Ellis H. Pathological changes produced by surgical dusting powders. *Annals of the Royal College of Surgeons of England* 1994; **76**: 5–8. [The pathology of starch granulomas and other irritants.]

Johnson CC, Baldessarre J, Levison ME. Peritonitis: update on pathophysiology, clinical manifestations, and management. *Clinical Infectious Diseases* 1997; **6**: 1035–45. [Extensive overview of peritonitis.]

Kittur DS, Korpe SW, Raytch RE, Smith GW. Surgical aspects of sclerosing encapsulating peritonitis. *Archives of Surgery* 1990; **125**: 1626–8. [Surgical outcome of sclerosing peritonitis.]

Maddaus MA, Ahrenholz D, Simmons RL. The biology of peritonitis and implication for treatment. *Surgical Clinics of North America* 1988; **68**: 431–43. [Pathophysiology of peritonitis in review.]

Rigby RJ, Hawley CM. Sclerosing peritonitis: the experience in Australia. *Nephrology, Dialysis, Transplantation* 1998; **13**: 154–9. [Extensive review of all aspects of sclerosing peritonitis.]

Saklayen MG. CAPD peritonitis. Incidence, pathogens, diagnosis and management. *Medical Clinics of North America* 1990; **74**: 997–1010. [Epidemiology of CAPD peritonitis.]

Willcox CM, Dismukes WE. Spontaneous bacterial peritonitis. A review of pathogenesis, diagnosis and treatment. *Medicine (Baltimore)* 1987; **66**: 447–56. [Review of primary peritonitis.]

34.3 Trauma: injuries to the abdominal wall and mesentery

Ralph L. Warren and Charles J. McCabe

Injuries of the abdominal wall
Injuries of the mesentery
Summary
Further reading

Injuries of the abdominal wall

Injuries of the abdominal wall are important not only in their own right, but also because they serve as markers of underlying visceral injuries. Traumatic forces to the abdomen first impact on the abdominal wall. Penetrating trauma, by definition, always causes injury to the structures of that wall, ranging from small puncture wounds to large, irregular lacerations (Fig. 1). Blunt trauma can be more deceptive: it can injure just the abdominal wall, both that wall and underlying viscera, or, when the structures of the wall are elastic (as in children), just the abdominal viscera, leaving the wall itself relatively unscathed.



Fig. 1. Significant trauma to the abdominal wall and chest from a boating accident.

Penetrating trauma, especially stab wounds, not uncommonly produces isolated injuries of the abdominal wall, but these are usually not life-threatening. Stab wounds to the abdomen may penetrate the full thickness of the abdominal wall without injuring intra-abdominal viscera. In the past, stab wounds were locally explored in the emergency department to determine whether or not the deep fascia had been transgressed; if so, then exploratory laparotomy was performed. Even with fascial penetration, however, at least 50 per cent of these patients were found to have no significant visceral injury upon exploration. This fact has led to the increasing use of non-invasive [ultrasonographic, computed tomographic (**CT**), and/or laparoscopic] evaluation and even non-operative observational management of abdominal stab wounds in an effort to avoid unnecessary laparotomy. Hemoperitoneum is no longer an absolute indication for laparotomy. Absolute indications for operation are evisceration (Fig. 2), hemodynamic instability, and peritonitis. Stab wounds of the abdominal wall require routine wound care (hemostasis, irrigation, debridement), fascial repair if there is any significant defect of the deep fascia, and loose skin closure. If laparotomy is performed, the deep fascia is often best repaired from the inside. A short course of oral antibiotics, for example dicloxacillin or cefazolin for 2 to 3 days, is prudent for lacerations of the abdominal wall not requiring laparotomy, but drains are not necessary unless persistent oozing or an accumulation of fluid is expected.



Fig. 2. Evisceration as a result of an abdominal stab wound.

Gunshot wounds have a much higher (over 90 per cent) rate of injury to the intra-abdominal organs, and high-energy projectiles can cause much more extensive disruption. Basic surgical principles of wound care (hemostasis, irrigation, and debridement) are keys to treatment. For large, contaminated wounds, intravenous antibiotics and delayed primary closure have been successfully used since the Vietnam conflict. In general, gunshot wounds of the abdomen have, until recently, required exploratory laparotomy. However, some low-energy gunshot (small-caliber handgun) wounds that appear on physical examination to be tangential without any manifestations of visceral injury (hemorrhage, peritonitis) can be, and have been, managed non-operatively. Diagnostic peritoneal lavage can be used in such cases to rule out intraperitoneal injury: a red-cell count in lavage fluid of less than 10 000 cells/mm^3 is adequately reliable in excluding visceral injury.

Blunt trauma to the abdomen can produce isolated injuries to the wall or to viscera, or combined injuries to both. Isolated injuries to the wall include abrasions, contusions (Fig. 3), and lacerations, which are managed in exactly the same way as such soft-tissue wounds elsewhere on the body. ‘Degloving' injuries of the abdominal wall, in which the subcutaneous fat is sheared apart, avulsing the skin from the fascial layers and thereby depriving it of its blood supply, may result, for example, from the tyres of a motor vehicle passing over the victim's abdomen. These injuries are also managed in the same way as degloving injuries elsewhere.



Fig. 3. Flank abrasion: hematoma should alert the physician to potential renal and colonic injury.

The ‘seatbelt syndrome', often associated with the ‘seatbelt sign' (Fig. 4), results from compression from an improperly positioned (too high, over the abdomen instead

of the bony pelvis) lap seatbelt in a motor-vehicle crash. The force can disrupt the body of the rectus abdominis muscle, as well as the external and internal oblique muscles, or cause avulsion of the abdominal musculature from its attachments at the costophrenic arch or pelvic bones. Uniquely associated with the seatbelt syndrome is a fracture of the posterior process and/or the body of the lumbar spine, first described by Chance in 1948 (the fracture now bears his name), resulting from hyperflexion of the spine about a fixed axis anterior to the vertebral column. Also included in the seatbelt syndrome are fractures of the neck of the pancreas and avulsions of the small-bowel mesentery (see below). Many trauma surgeons consider a true seatbelt sign as an indication for operation, as it is associated with such a high incidence of visceral injury.



Fig. 4. The seatbelt sign should create a high index of suspicion for intra-abdominal injury.

Two isolated injuries of the abdominal wall, traumatic hernia and hematoma of the rectus sheath, are unique to the abdominal region and may or may not be due to seatbelt injuries. Traumatic hernia of the abdominal wall is associated with a specific diagnostic triad: the immediate appearance of a hernia with intact skin after blunt abdominal trauma, signs of injury at the time of the initial medical evaluation, and no identifiable hernia sac at exploration. The hernia may present as an easily palpable, anterior abdominal mass, or become apparent only during exploration of the abdomen. Abdominal hernias should be repaired at the time of presentation, usually using non-absorbable suture (polypropylene, nylon). Rarely, especially when the hernias have been neglected because of a delay in diagnosis, the defect may be large and require a prosthetic graft to effect closure. Injuries that result in tissue necrosis and/or significant contamination can create enormous management problems and are fortunately rare. Contamination is, of course, a relative contraindication to the use of prosthetic material, but if unavoidable, polytetrafluoroethylene (Gor-tex®) has been successfully used and is relatively resistant to infection.

A hematoma of the rectus sheath is a result of disruption of the large superior or inferior epigastric vessels, which course within and along the posterior aspect of the rectus muscles on each side. It presents as a tender abdominal mass, which, unlike true intra-abdominal masses, usually increases in prominence when the patient tenses the abdominal muscles (Bouchacourt's sign). The hematoma can often be visualized with ultrasound, and an abdominal CT scan delineates the lesion very nicely (Fig. 5). Most rectus hematomas can be managed non-operatively. A few patients, however, especially those with coagulopathy or on anticoagulation therapy, will have evidence of expansion (increasing pain and signs of continuing blood loss and transfusion requirement) and may need operative ligation or angiographic embolization of the bleeding vessels.

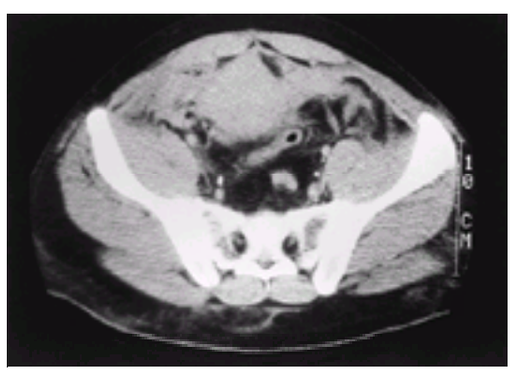


Fig. 5. Rectus sheath hematoma visualized by CT.

Injuries of the mesentery

Injuries of the bowel mesentery are common, both as isolated injuries and in combination with injuries of other intraperitoneal structures. Penetrating trauma, in general, causes lacerations. Less commonly, either because of injury to a large mesenteric-root vessel or extensive combined bowel and mesenteric disruption (most commonly secondary to gunshot wounds), the bowel is rendered ischemic. The specific diagnosis of mesenteric injury is rarely made preoperatively. Rather, the patient presents with accompanying intra-abdominal hemorrhage, hypotension, a fluid requirement, falling hematocrit, abdominal tenderness and/or distension (note that several units of blood can easily be accommodated in the peritoneal cavity without being detectable as distension), or hemoperitoneum detected by diagnostic peritoneal lavage, ultrasound, or CT scan.

Blunt trauma to the abdomen causes mesenteric injury in approx. 5 per cent of all patients, and is due to compression and deceleration forces. The injury can result in mesenteric hematomas (Fig. 6), less frequently free intraperitoneal hemorrhage, or devascularization of the bowel causing ischemia (Fig. 7). Compression tears the intima of the mesenteric vessels, leading to secondary thrombosis. Mesenteric hematomas and free hemorrhage manifest clinically as hypotension and fluid requirement, falling hematocrit, and shock, or are detected by diagnostic peritoneal lavage, ultrasound, or CT; mesenteric hematomas can also be directly visualized by CT. The preoperative diagnosis of mesenteric avulsion with bowel ischemia requires a high index of suspicion, as no imaging studies are able to detect it. It occurs most commonly after a sudden, severe, abdominal compressive force, as in the ‘seatbelt syndrome', or when a motor vehicle's tyres run over a pedestrian. Acceleration/deceleration forces are also a common mechanism.

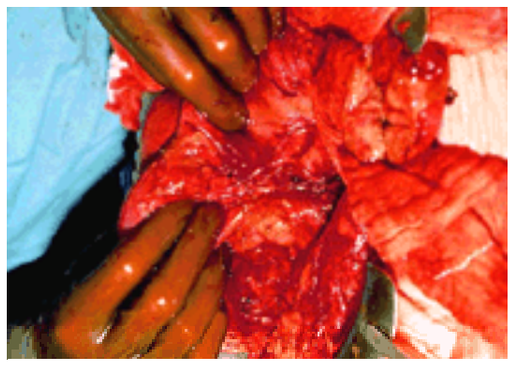


Fig. 6. A mesenteric hematoma resulting from blunt abdominal trauma; there was no intestinal ischemia and the hematoma was not expanding.



Fig. 7. Avulsion of the mesentery can result in ischemic bowel.

All patients undergoing abdominal exploration for trauma require careful inspection of the entire length of the intestine as an integral part of operative management. A search should be made for mesenteric hematomas, mesenteric and bowel tears (Fig. 8), and ischemic bowel. Non-expanding hematomas and contusions of the mesentery, found by CT or at operation, do not require exploration if the bowel is definitely viable. If there is evidence of continuing blood loss or the hematoma is seen at operation to be expanding, or if the bowel is not viable, proximal vascular control should be obtained, the hematoma should be explored, and the bleeding vessels ligated. Bowel resection should be accomplished if necessary for ischemic intestines. Where injury to the superior mesenteric artery or vein results in extensive ischemia, the vessel should be repaired using autogenous saphenous vein, either as a patch or end-to-end interposition graft. Non-viable intestine should be resected and primary anastomosis carried out, closing any mesenteric defects to prevent future internal hernia. Intraoperative Doppler-flow examination or intravenous injection of fluorescein and inspection of the bowel under ultraviolet light can help determine the adequacy of vascular supply. If bowel viability or graft patency are tenuous, the proximal bowel can be brought out as an enterostomy and the distal end either brought out as a mucous fistula or closed as a Hartmann pouch, or a planned 'second-look' operation should be done within 24 h.



Fig. 8. Disruption of the antimesenteric small intestine as a result of compressive forces.

Summary

Injuries of the abdominal wall must be evaluated both in and of themselves, and, more importantly, as clues to possible underlying visceral injuries. Associated injuries to the head, thorax, and extremities often occupy the physician's attention, and diagnostic efforts are usually, and initially should be, focused on detecting underlying intra-abdominal visceral injury. The diagnosis of injuries to the abdominal wall and mesentery relies heavily on physical examination and a high index of suspicion, with the physician consciously looking for signs of injury, followed by appropriate imaging if needed. Management of these injuries is usually straightforward, following basic surgical principles. Complicated defects may require innovative therapy.

Further reading

Appleby JP, Nagy AG. Abdominal injuries associated with the use of seatbelts. *American Journal of Surgery* 1989; **157**: 457–8. [Gastrointestinal and lumbar spine injuries are commonly associated with seatbelt use.]

Asbun HJ, Irani H, Roe EJ, Bloch JH. Intraabdominal seatbelt injury. *Journal of Trauma* 1990; **30**: 189–93. [The seatbelt sign should create a high level of suspicion for intra-abdominal visceral injury.]

Blaisdell FW, Trunkey DD. *Abdominal trauma*, 2nd edn. Thieme Medical, New York, 1993. [The best general reference on abdominal trauma.]

Carillo EH, Bergamini TM, Miller FB, Richardson JD. Abdominal vascular injuries. *Journal of Trauma* 1997; **43**: 164–71. [Compressive forces with elevation of an intimal flap can lead to occlusion of mesenteric vessels resulting in visceral ischaemia.]

Chance CQ. Note on a type of flexion fracture of the spine. *British Journal of Radiology* 1948; **21:**, 452–3. [The first description of the Chance fracture.]

Garrett JN, Braunstein PW. The seatbelt syndrome. *Journal of Trauma* 1962; **2**: 220–38. [The first description of the 'seatbelt syndrome'.]

Malangoni MA, Condon RF. Traumatic abdominal wall hernia. *Journal of Trauma* 1983; **23**: 356–7. [An excellent review of traumatic hernias.]

34.4 The omentum

D. McWhinnie

Torsion of the omentum
Aetiology
Pathology
Clinical features
Treatment
Tumours of the omentum
Omental cysts
Solid tumours of the omentum
Pathology
Clinical features and treatment
Further reading

Torsion of the omentum

Torsion or volvulus of the greater omentum is defined as a rotation of the organ in its longitudinal access around a narrow pedicle. It may be classified as primary or secondary.

Aetiology

In primary or idiopathic torsion, a redundant, mobile segment of omentum rotates around a proximal fixed point in the absence of any associated intra-abdominal pathology. Although first described by Eitel in 1899, it remains a relatively rare condition with approximately 250 cases reported in the world literature. While the precise cause is unknown, both predisposing and precipitating factors in the pathogenesis of the condition can be identified.

Predisposing factors include an anatomical abnormality of the omentum itself, such as a bifid omentum, accessory omentum, and a narrowed omental pedicle. Accumulations of omental fat in the obese, especially in children, is another well recognized factor. The normal anatomical arrangements for the omental vessels may also give rise to torsion. The omental veins are larger, longer, and more tortuous than the arteries and this redundancy allows venous kinking and obstruction, which provides a fixed point around which a self-perpetuating torsion may occur. A higher incidence of torsion on the right is related to the greater size and mobility of the right side of the omentum.

Precipitating factors for torsion are those which cause displacement of the omentum. At risk are manual workers who experience heavy exertion, especially if associated with repetitive blunt abdominal trauma or exposure to vibrating tools. Other precipitating factors include coughing and straining, sudden change in body position, and hyperperistalsis from overeating.

Secondary torsion is more common than the primary type and is associated with pre-existing abdominal pathology. The omentum twists between two fixed points and its distal edge is attached directly or by adhesions to cysts, tumours (primary or secondary), foci of intra-abdominal inflammation, post-surgical wounds or scarring, internal hernia, or external hernial sacs. The majority of cases occur in patients with inguinal hernias. Precipitating factors of primary torsion also initiate secondary torsion.

Pathology

The omentum rotates a variable number of times around a pivotal point, usually in a clockwise direction. The venous return is compromised and the distal omentum becomes congested and oedematous. Resultant haemorrhagic extravasation stimulates aseptic peritonitis with a characteristic accumulation of serosanguineous fluid in the peritoneal cavity. As the torsion proceeds, arterial occlusion leads to acute haemorrhagic infarction and eventual necrosis of the omental segment (Fig. 1). If the mass is not excised it becomes atrophic and fibrotic, and on rare occasions the pedicle may even auto-amputate.

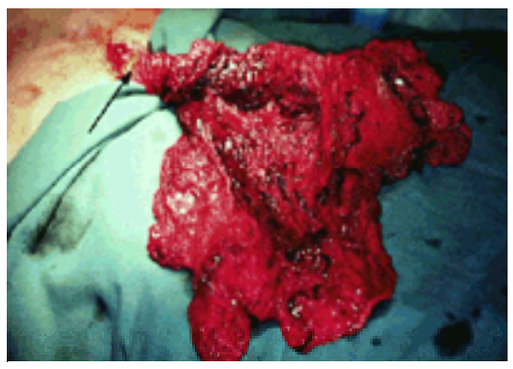


Fig. 1. Torsion of the greater omentum around a proximal pivotal point (arrow) (by courtesy of Dr V. Myllyla).

Clinical features

Omental torsion occurs either in childhood or in adulthood around the fourth decade of life. Men are affected twice as often as women and the majority of patients are overweight. The signs and symptoms reflect the underlying localized peritonitis. As the affected segment is usually on the right, the sudden onset of pain, with rebound tenderness and guarding is also right sided, and is commonly mistaken for acute appendicitis. Other possible diagnoses include acute cholecystitis, twisted ovarian cyst, or torsion of an epiploric appendage. The non-specific features of peritonitis are not usually sufficient to allow an acute preoperative diagnosis. If CT scanning or MRI is performed, the diagnosis can be made on the characteristic fat spiral pattern of the rotated omentum (Fig. 2). The finding of free serosanguineous fluid in association with a normal appendix, gallbladder, pelvic organs, and bowel should alert the surgeon to the possibility of omental torsion.

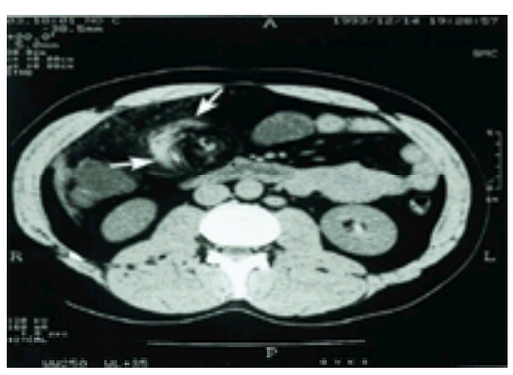


Fig. 2. CT scan of omental torsion (arrows) resembling a spiral pattern. The fatty spiral mass represents necrotic omentum.

Treatment

The condition only becomes clinically apparent once vascular thrombosis of the omental vessels has occurred, and is irreversible even if the omentum is derotated. Treatment consists of resection of the affected portion of omentum. Any diseased process associated with secondary torsion, such as inguinal hernia, also requires correction. Although omental torsion is not life threatening, segmental omentectomy reduces morbidity and removes inflammatory tissue which may later serve as a focus of intra-abdominal adhesions. Postoperative recovery is usually rapid and morbidity is minimal.

Tumours of the omentum

Omental cysts

Pathology

Most cysts of the omentum are of lymphatic mesothelioma origin. All are rare.

Cystic lymphangioma

In childhood, omental cysts are usually caused by developmental abnormalities of the lymphoid tissue, such as obstruction of the lymphatic channels or by growth of congenitally misplaced lymphatic tissue. They are variously called chylous cysts, cystic hygromas, or cystic lymphangiomas, and are benign. They vary greatly in size, the smallest being only a few centimetres in diameters, and can be unilocular or multilocular. Histologically each cyst has an endothelial lining similar to cystic hygroma of the neck, and contains many foamy macrophages, giving the fluid a milky appearance.

Cystic mesothelioma

Omental cysts of mesothelial origin occur almost exclusively in adult life, usually in women under the age of 50 years. Although they are benign, local recurrence often occurs after surgical excision. They appear as large multicystic masses similar to cystic lymphangiomas, but histologically they are lined by flattened or cuboidal mesothelial cells and the cyst fluid is clear, containing mucopolysaccharides. Although the aetiology is unknown there is no association with asbestos exposure.

Dermoid cysts

As with dermoid cysts elsewhere in the body, omental dermoid cysts are lined with squamous epithelium and may contain epithelial structures such as hair and teeth.

Pseudocysts

Omental pseudocysts are caused by fat necrosis or abdominal trauma with haematoma formation. They are lined with fibrous tissue and contain bloodstained fluid.

Clinical features and treatment

Many omental cysts are small and asymptomatic and may only be discovered incidentally at laparotomy or autopsy. Large cysts may present with diffuse abdominal distension or as asmooth, mobile, palpable mass in the lower midline. Characteristically they are non-tender unless complicated by torsion of the omentum or intestinal obstruction. Plain radiographs of the abdomen may demonstrate a soft tissue shadow and barium studies may show displacement of bowel. The differential diagnosis includes mesenteric, peritoneal, or retroperitoneal cysts and tumours but the diagnosis is usually made at laparotomy. Treatment consists of surgical excision of the cyst.

Solid tumours of the omentum

Pathology

Secondary tumour

The vast majority of omental neoplasms are metastatic carcinomas arising from ovary, bowel, or pancreas and these are often associated with abdominal ascites. The rare, diffuse, malignant mesothelioma of the peritoneum which is associated with the exposure to fibrous minerals such as asbestos also consistently involves the omentum.

Primary tumour

Primary solid tumours of the omentum are exceptionally rare and may be benign or malignant. The majority are of smooth muscle origin. Malignant potential is difficult to predict from the histology but approximately one-third are frankly malignant. Before making a diagnosis of primary smooth muscle tumour of the omentum, leiomyosarcoma of the uterus or gastrointestinal tract giving rise to omental metastases must be carefully excluded. Other rare primary omental tumours include fibroma, fibrosarcoma, lipoma, and liposarcoma. Infantile myxoid hamartomas, found in infants under 1 year old, consist of multiple nodular lesions which show histological resemblance to myxoid liposarcoma; the clinical course is invariably benign.

Clinical features and treatment

Benign primary tumours of the omentum, when sufficiently large, present with a palpable abdominal mass or diffuse distension and require surgical excision. Malignant primary omental tumours are highly invasive and often present late with involvement of adjacent organs. Radical surgical excision of both the omentum and the involved organs may be required, but often palliative surgery is the only treatment option.

Further reading

Kimber CP, Westmore P, Hutson JM, Kelly JH. Primary omental torsion in children. *Journal of Paediatrics and Child Health* 1996; **32**: 22–24. [Aetiology and clinical features of juvenile omental torsion.]

Mainzer RA, Simoes A. Primary idiopathic torsion of the omentum. *Archives of Surgery* 1964; **88**: 974–983. [Overview of adult omental torsion.]

Morson BC, Dawson IMP, Day DW, Jass JR, Price AB, Williams GT. *Morson's and Dawson's gastrointestinal pathology*, 3rd edn. Blackwell, London, 1990. [Classification of omental tumours and cysts.]

Myllyla V, Leskinen M, Paakko E, Kiviniemi H, Kairaluoma MV. CT appearance of omental torsions: fat spiral pattern. *Fortschritte der Röntgenologie* 1994; **161**: 559–560. [Gross pathology and imaging of omental torsion.]

Stout AP, Hendry J, Purdie FJ. Primary solid tumours of the great omentum. *Cancer* 1963; **16**: 231–243. [Definitive classification of tumours of the omentum.]

34.5 Retroperitoneal fibrosis

David Cranston

Introduction
Benign
Malignant
Clinical picture
Diagnosis
Management
Further reading

Introduction

Retroperitoneal fibrosis was first clearly described by the French urologist Albarran, at the beginning of the twentieth century, as ‘stenosing periureteritis'. In 1902 he practised his operation of *liberation externe*, designed to disengage the ureters from the scar tissue and fibrous masses that surrounded them. In 1948, Ormond described two patients with diffuse fibrosis of the retroperitoneal tissues who presented with backache, anaemia, and anuria, and he established the clinical and pathological entity of idiopathic retroperitoneal fibrosis more clearly. In the 1960s the disease became known as retroperitoneal fibrosis, having previously be known under various names including chronic periureteritis and fibrosing retroperitonitis, as it described more accurately the nature of the disease. An increasing number of causes of retroperitoneal fibrosis are now recognized, and can be divided into benign and malignant.

Benign

Idiopathic retroperitoneal fibrosis, in which no aetiological factor is found, comprises two-thirds of the benign cases, and is the best known of the group of fibrosing syndromes that include mediastinal fibrosis and sclerosing cholangitis. A dense plaque of fibrous tissue forms in the lower abdomen and extends laterally and downwards from the renal arteries over the promontory of the sacrum, encasing, but not usually infiltrating, the hollow tubes, aorta, inferior vena cava, and ureters (Fig. 1). The central portion of the plaque consists of dense scar tissue, while the growing margins have the histological appearance of chronic inflammation, with a mixturepof mononuclear sells inderspersed with occasional giant cells and eosinophils. Idiopathic retroperitoneal fibrosis may occur in areas where the walls of the aorta or other arteries have severe atherosclerosis with damage to the media. When this occurs, insoluble lipid (ceroid) leaks into the periaortic tissues and induces an IgG-mediated immune response. Chronic periaortitis has been suggested as a more appropriate term for this condition (see section 7.2). Idiopathic retroperitoneal fibrosis normally affects men in the fifth or sixth decade of life, although it has been reported in children as young as 7 years of age, and in adults up to the age of 85. One uncommon but well-recognized cause of this condition is the drug methysergide maleate that has been taken for migraine. Occasionally, long-standing infections of the urinary tract, especially with extravasation, can lead to retroperitoneal fibrosis, although extravasation in the absence of infection seldom leads to fibrosis. Haemorrhage into the retroperitoneal space from a leaking abdominal aortic aneurysm is said to be a cause of this condition, but this seems unlikely, and it is more probable that the association of abdominal aortic aneurysm with idiopathic fibrosis is due to an immune response to insoluble lipid. More recently, retroperitoneal fibrosis has been described in a patient with Parkinson's disease treated with pergolide, and as a complication of intra-arterial stents and angioplasty.

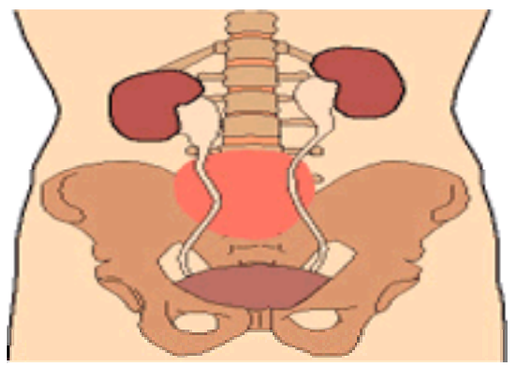


Fig. 1. The classical site of retroperitoneal fibrosis.

Malignant

The most common malignant tumour presenting as retroperitoneal fibrosis is a lymphoma, and the diagnosis may be missed at laparotomy if a deep biopsy is not taken. Carcinoma of the breast, stomach, pancreas, colon, renal tract and prostate, and carcinoid tumours, may be associated with retroperitoneal fibrosis due to metastasis. It may be difficult to identify the underlying metastatic tumour.

Radiotherapy and the treatment of cancer can also cause retroperitoneal fibrosis, although this is much less common today with more precise localization of the field to be irradiated. Chemotherapy, especially in the treatment of testicular tumours, may leave fibrous retroperitoneal masses that can involve the ureter. These may or may not contain residual tumour, and may require surgical removal.

Clinical picture

Men are affected twice as commonly as women. In the early stages the underlying disease is responsible for the presentation, which is relatively non-specific, including loss of appetite and weight, fever, sweating, and malaise. Later in the disease the clinical features reflect the major complication of the disease, ureteric obstruction, which, if bilateral, can lead to anuria and renal failure. The clinical picture may also be associated with an underlying cause such as an abdominal aortic aneurysm. Hypertension may be present in up to 60 per cent of the cases, and pyuria is a common finding. A girdle distribution of pain in the abdomen is found in up to 90 per cent of patients, described as insidious, dull, and not colicky.

Diagnosis

The diagnosis of retroperitoneal fibrosis is made conventionally from the history, examination, and investigations. Usually the erythrocyte sedimentation rate is high. The typical appearance on the intravenous urogram or antegrade pyelogram is that of medial displacement of the ureters with dilatation of the ureter and pelvis above (Fig. 2). Up to one-third of patients will have a non-functioning kidney at the time of presentation due to long-standing obstruction. Ultrasonographic examination may demonstrate the fibrous plaque as an echo-free mass with smooth borders, thickest at the sacral promontory. Both computed tomographic scanning and magnetic resonance imaging can define the area of fibrosis precisely (Fig. 3). Typically this extends from the renal vessels down to the level of the bifurcation of the great vessels, and out laterally to the psoas muscles. Fine-needle biopsy of the mass may be helpful in confirming the presence of malignant disease, but a negative result does not exclude malignancy.

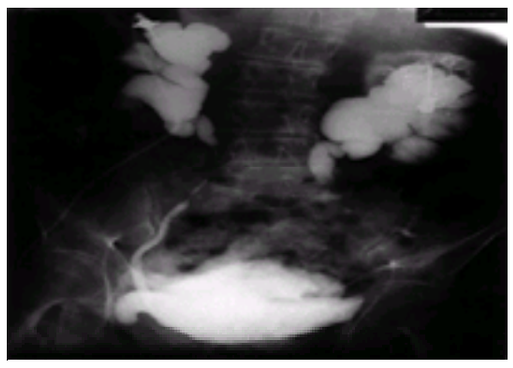


Fig. 2. Antegrade pyelogram showing medial deviation of the ureter with obstruction.

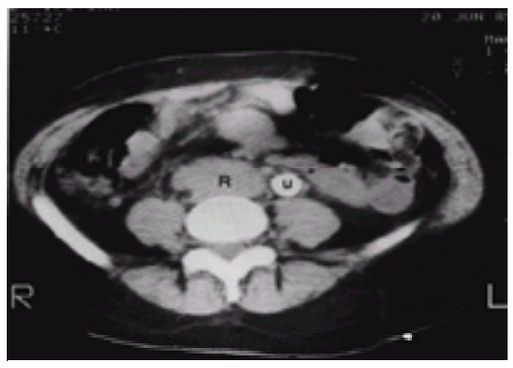


Fig. 3. Computed tomograph, showing retroperitoneal fibrosis (R) surrounding the aorta and inferior vena cava with ureteric obstruction (U).

Management

Surgical treatment is often necessary for retroperitoneal fibrosis, but a number of medical therapies have been useful.

The role of steroids remains a subject of debate. They may decrease the oedema often associated with retroperitoneal fibrosis and in this way help in resolution of the obstruction. If steroids are used, it is probably wise to stop them when the erythrocyte sedimentation rate returns to normal. Spontaneous resolution of retroperitoneal fibrosis in the absence of treatment is known to occur. More recently the anti-oestrogen tamoxifen has been used successfully in some patients, and cyclophosphamide is another drug that reportedly has some benefit.

If the patient is unwell due to renal failure the obstruction can be relieved by nephrostomies or JJ stents, passed in either an antegrade or retrograde fashion. Here it is important to watch for and correct the fluid and electrolyte losses due to diuresis after relief of the obstruction. Once the emerency has been resolved, there is time to find the underlying cause.

Laparotomy is often necessary to free the ureters from the encasing fibrous tissue. At the same time, biopsies should be taken to exclude an underlying malignant process. It is essential that large, deep biopsies be taken for this purpose as it is very difficult to exclude malignancy with small, superficial ones.

In the presence of renal dysfunction, ureterolysis is performed. With care, a plane can usually be found between the ureters and fibrous tissue. It is often easier to begin just above the bladder, as the ureter is usually free at this point and, as the plaque is entered, a plane can be found between it and the ureter. It is often helpful to place a sling around the ureter, insert a right-angled forceps between the plaque and the ureter, and then cut down on to the forceps with the scalpel. The insertion of a stent or catheter up the ureter helps to identify it. Great care must be taken on the right-hand side to avoid damage to the inferior vena cava, which is often very difficult to identify in the inflammatory fibrotic mass.

After freeing the ureters from the fibrous sheath that surrounds them, omentum can be mobilized and used to wrap them throughout their length, keeping them free of fibrous tissue (Fig. 4). In mobilizing the omentum, it must be remembered that the blood supply runs vertically down the gastroepiploic arch, and that the majority of the blood to the arch comes from the right side. While wrapping the ureters in omentum does not guarantee freedom from recurrent obstruction, it seems to be better than other techniques.

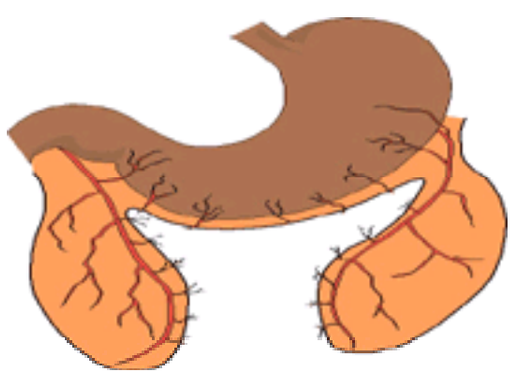


Fig. 4. Mobilization of the omentum before wrapping around the ureters.

In patients who are unwell, or who have one non-functioning kidney, it may be better to operate on only one side rather than both. However, wherever possible the procedure should be done bilaterally even in the presence of unilateral disease, as involvement of the other side is almost certain with the passage of time.

Rarely, dense fibrosis is restricted to the lower ureters, and it is possible to reimplant a normal ureter into the bladder with a psoas hitch or a Boari flap. Occasionally, long strictures form after the initial ureterolysis, and these may have to be treated with more specialized techniques, such as renal autotransplanation.

Close follow-up is important to check on progressive or recurrent disease, although bilateral ureterolysis usually has a satisfactory outcome, especially if the renal impairment was not too severe in the first place.

Further reading

Albarran J. Retention renale par periureterite, liberation externe de l'uretere. *Association France Urologie* 1905; **9**: 511–17. [First description of the disease.]

Baker LRI *et al.* Idiopathic retroperitoneal fibrosis. A retrospective analysis of 60 cases. *British Journal of Urology* 1988; **60**: 497–503. [A review of 60 cases.]

Charlton CAC. The use of steroids in a form of retroperitoneal fibrosis. *Proceedings of the Royal Society of Medicine* 1968; **61**: 875–6. [A discussion of the value of steroids.]

Degressys GE, Dunnick NR, Sivleman PM, Cohan RH, Illescas FF, Castagno A. Retroperitoneal fibrosis: use of CT in distinguishing among possible causes. *American Journal of Roentgenology*

1986; **146**: 57–60. [Discussion of imaging techniques.]

Mikkelsen D, Lepor H. Innovative surgical managment of idiopathic retroperitoneal fibrosis. *Journal of Urology* 1989; **141**: 1192–6. [Surgical options for the difficult cases.]

Mitchinson MJ. Retroperitoneal fibrosis revisited. *Archives of Pathology and Laboratory Medicine* 1986; **110**: 784–6. [Discussion of the pathology of the disease.]

O'Flynn D. Retroperitoneal fibrosis In: McDougal WS, ed. *Rob and Smith's operative surgery (Urology)*, 4th edn, pp. 243–54. Butterworths, London, 1986. [A good surgical review.]

Ormond JK. Bilateral ureteral obstruction due to envelopment and compression by an inflammataory retroperitoneal process. *Journal of Urology* 1948; **59**: 1072–9. [A classic historical paper.]

Resnick MI, Kursh ED. Retroperitoneal fibrosis. In: Walsh PC, Retik AB, Vaughan ED, Wein AJ, ed. *Campbell's urology*, 7th edn, pp. 403–9. WB Saunders: Philadelphia, 1998. [Excellent review article.]

Suby HI, Kerr WS, Graham JR, Fraley E. Retroperitoneal fibrosis: a missing link in the chain of pathogenesis *Journal of Urology* 1965; **93**: 144–52.[Discussion of pathology.]

Tresidder GC, Blandy JP, Singh M. Omental sleeve to prevent recurrent retroperitoneal fibrosis around the ureter. *Urology International* 1972; **27**: 144–8. [Surgical techniques in the disease.]

Yuh WTC, Barloon TJ, Sickels WJ, Kramolowsky EV, Williams RD. Magnetic resonance imaging in the diagnosis and follow up of idiopathic retroperitoneal fibrosis. *Journal of Urology* 1989; **141**: 602–5. [Discussion of MRI in the diagnosis.]

34.6 Retroperitoneal neoplasms

Herbert C. Hoover, Jr.

Introduction
Classification of retroperitoneal neoplasms
Histology of retroperitoneal sarcomas
Clinical presentation
Diagnostic evaluation
Operative considerations
Results of therapy
Resection rate
Extent of resection
Operative mortality
Operative morbidity
Survival
The role of chemotherapy and radiation therapy
Chemotherapy
Radiation therapy
Summary
Further reading

Introduction

The retroperitoneum is a large potential space bounded anteriorly by the posterior peritoneum, posteriorly by the spine and back muscles, superiorly by the diaphragm, inferiorly by the levators, and laterally by the flank muscles at the level of the anterior superior spine of the iliac crest to the tip of the twelfth rib. In this vast potential space, retroperitoneal masses tend to become very large before producing signs or symptoms, thus accounting for the poor prognosis for most malignant tumors arising there. Although the pancreas, kidneys, ureters, and adrenals are retroperitoneal structures, neoplasms of these organs are not generally included in the analysis of retroperitoneal neoplasms and will not be considered in this section. With rare exceptions, retroperitoneal neoplasms are sarcomas, lymphomas, or benign lesions. Lymphomas and benign tumors will be discussed only in terms of their differentiation from sarcomas, which will be the focus of our discussion. Retroperitoneal sarcomas are rare, representing only about 0.1 to 0.2 per cent of all malignancies overall and only about 10 to 15 per cent of all soft-tissue sarcomas. They form approximately 40 per cent of all retroperitoneal masses.

Classification of retroperitoneal neoplasms

Most retroperitoneal tumors are of mesodermal origin, and both benign and malignant tumors can arise from many different tissues, such as: adipose, lipoma and liposarcomas; smooth muscle, leiomyoma and leiomyosarcomas; connective tissue, fibroma and fibrosarcoma; striated muscle, rhabdomyoma and rhabdomyosarcoma; lymph vessels, lymphangioma and lymphangiosarcoma; lymph nodes, nodal hyperplasia and lymphoma; blood vessels, hemangiomas and hemangiosarcoma, or benign and malignant hemangiopericytoma. Benign and malignant tumors may also originate from the nervous system, including nerve-sheath tumors such as non-encapsulated fibroma, encapsulated neurilemmoma, or malignant schwannoma. Tumors arising from the sympathetic system include ganglioneuroma, sympathicoblastoma, and neuroblastoma. Rarely, tumors may arise from heterotopic adrenocortical and chromaffin tissue, such as carcinomas arising from adrenocortical tissue, malignant non-chromaffin paraganglioma, paraganglioma, and pheochromocytoma. Urogenital-ridge tumors are rare, as are tumors arising from embryonic remnants, such as benign and malignant teratomas and chordomas. At least 16 cases of retroperitoneal synovial sarcoma have been reported. Primary tumors in virtually any organ can metastasize to the retroperitoneum, metastatic testicular carcinoma being the most common. Benign cysts also occur in the retroperitoneum, as they do in all other body cavities.

Histology of retroperitoneal sarcomas

The histologic types of retroperitoneal sarcoma seen most commonly in several major series are listed in Table 1. Malignant fibrous histiocytoma is being diagnosed with increasing frequency, probably because of a better understanding of the histopathology of the tumor and the reclassification of many tumors previously diagnosed as pleomorphic variants of liposarcoma, fibrosarcoma, and rhabdomyosarcoma. Histologic grade is of more prognostic significance than the cell of origin. Mesodermally derived tumors of differing embryonic origin tend to behave in a similar biologic fashion, but manifest increasingly aggressive behavior clinically as they become less differentiated or assume a higher grade. The United States National Cancer Institute grading uses a composite of histopathologic features that includes necrosis, cellularity, pleomorphism, and mitosis. Each histologic type has a spectrum of biologic behavior. Some tumors exhibit a narrow spectrum such as prolonged benign course (e.g. myxoid liposarcoma), but others, such as synovial sarcoma, are consistently aggressive and are always assigned a grade 3 (highest grade). The degree of necrosis has the strongest predictive value for overall survival: prognosis is worse as the percentage of necrosis increases. All of these grading criteria are probably more applicable to soft-tissue sarcomas in sites other than the retroperitoneum, where even low-grade lesions respond poorly to treatment, presumably because of their large size at presentation and the difficulty in achieving wide surgical margins on excision.

Histological type	*No. seen*	*% of total*
Liposarcoma	92	31
Leiomyosarcoma	70	24
Fibrosarcoma	33	11
Rhabdomyosarcoma	23	8
Neurofibrosarcoma	22	7
Miscellaneous	21	7
Malignant fibrous histiocytoma	15	5
Spindle-cell carcinoma	7	2
Hemangiopericytoma	6	2
Synovial sarcoma	5	2

Table 1 Histological types of retroperitoneal sarcoma: composite of four major series (294 cases)

Clinical presentation

The mean age in most series is around 55 years. There is a slight male predominance. Unless found incidentally during a laparotomy or on a computed tomographic (**CT**) scan, retroperitoneal sarcomas are usually large at presentation. Physical signs are few, other than a palpable abdominal or pelvic mass, which is seen in 50 to 75 per cent of patients. Up to 60 per cent of patients present with pain, which is usually very mild.

Diagnostic evaluation

CT scanning has revolutionized the investigation of patients with suspected retroperitoneal neoplasms and, if done with radiographic contrast media, may be the only study other than a chest radiograph that is needed preoperatively. The size of the tumor and any anatomic changes secondary to its growth are easily visualized, and tumor invasion of adjacent organs can be demonstrated or suggested. CT scanning is, however, not useful in predicting the histologic type or grade of soft-tissue sarcomas. Most tumors in the retroperitoneum appear as soft-tissue masses containing focal areas of necrosis (Fig. 1). If fatty elements are obvious, a diagnosis of liposarcoma is possible, but differentiation from a benign lipoma is uncertain. CT scanning cannot reliably distinguish between retroperitoneal lymphomas and sarcomas: although lymphomas tend to be homogeneous on CT scanning and often envelope the inferior vena cava and aorta, while sarcomas are usually heterogeneous, these findings are unreliable.

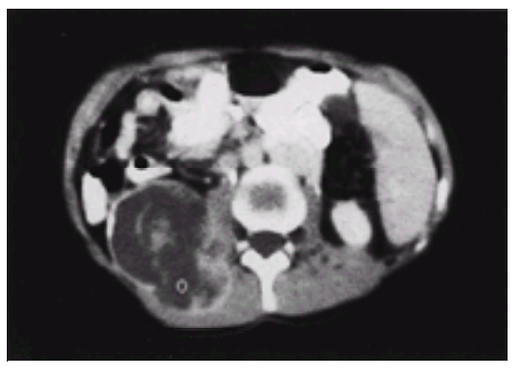


Fig. 1. CT scan of left retroperitoneal sarcoma with invasion into paraspinal muscles.

Although not used in most reported series, magnetic resonance imaging is felt by some radiologists to offer advantages over CT scanning for assessing retroperitoneal tumors. They argue that spacial assessment is better, due to the possibility of using differently oriented planes (axial, sagittal, coronal, and oblique). A clear view of the vascular structures is given without the need for contrast material.

The functional state of at least one kidney must be demonstrated by either an intravenous-contrast CT scan or an excretory urogram because the *en bloc* resection of one kidney is often required for curative resections. Although upper gastrointestinal series and a barium enema study often demonstrate displacement or even invasion, these findings are easily demonstrated at laparotomy, and thus these imaging studies are usually unnecessary. It is not unreasonable to obtain a CT scan of the lungs, although pulmonary metastases are unusual at the time of presentation of retroperitoneal sarcoma. Arteriography is helpful in showing the extent of the lesion and the tumor's blood supply; it often reveals displacement of major vessels. Although arteriography should be considered in patients with extensive lesions, it rarely adds significantly to a good-quality contrast CT. If performed, a ‘flush' aortogram is usually obtained first, followed by selective arterial injections as indicated. Findings suggestive of neoplasia include neovascularity, venous lakes, tumor blush, and vessel encasement. Because malignant fibrous histiocytoma tends to occur in the renal area, the demonstration of an extrarenal arterial supply is helpful in deciding to save the kidney. A dominant lumbar or intercostal arterial supply adds to the likelihood that the tumor has a retroperitoneal origin. If the arterial supply is celiac, or inferior or superior mesenteric, an intraperitoneal origin is likely, but retroperitoneal tumors are often supplied partially by the superior or inferior mesenteric arteries. Unfortunately, vascularity correlates poorly with the histologic type and cannot reliably differentiate between benign and malignant tumors. Invasion and obstruction of arteries are rare, although veins, especially the vena cava, are more vulnerable. An inferior vena cavagram can be helpful in assessing any invasion of the vena cava and/or renal veins. In the absence of widespread metastatic disease, the operability and resectability of these tumors should rarely be determined on the basis of radiologic findings, which are often misleading.

A histologic diagnosis should be secured before proceeding with a heroic, multiorgan resection for large tumors. Since histologic type and grade are important considerations, fine-needle aspirates are seldom adequate. In the past, even core biopsies were discouraged, as the limited amount of material for study often leads to great difficulty in classification. With modern pathologic techniques, CT-guided core biopsies can commonly provide enough tissue for an accurate diagnosis. For small masses that can predictably be resected *en bloc*, a preoperative diagnosis is less important.

Operative considerations

All patients should undergo a full bowel preparation because a limited resection of the colon or rectum is commonly required.

A limited midline incision is usually best for the initial exploration; this should be placed directly over the center of the mass so that it can be extended readily in either direction. If the tumor is in the upper retroperitoneum towards or invading the diaphragm, a thoracoabdominal approach may be indicated. The retroperitoneal or flank approach is less satisfactory than is an abdominal incision in allowing the surgeon to perform an *en bloc* resection of involved organs or to control the major arteries and veins supplying the tumor. The abdominal portion of the incision is opened first for the exploration to determine resectability, and a careful search is made for hepatic or peritoneal metastases. If biopsy of the tumor has not previously been performed or was inadequate, the surgeon has to decide whether a histologic diagnosis is needed. Incisional wedge biopsies should be obtained only from patients who have obviously inoperable disease or where lymphoma is suspected. A retroperitoneal biopsy carries the risk of contaminating the entire peritoneal cavity and retroperitoneal space, potentially negating any chance for cure. Great care must be taken to isolate the area of biopsy and to obtain absolute hemostasis.

Localized tumors, even those that are huge, should be removed completely to provide both biopsy material and ensure the most effective primary therapeutic approach. This should include an *en bloc* resection of involved organs (most commonly the kidney, tail of the pancreas, or colon), the border of resection being well beyond the visible limits of the tumor whenever possible. This should be 1 to 2 cm into normal, uninvolved tissue whenever possible. The apparent encapsulation is usually a pseudocapsule containing normal as well as neoplastic cells. The surgeon must resist the temptation of removing the tumor from its pseudocapsule if a curative resection is the intent; although such dissection is often rather easy, recurrence is almost certain. Retroperitoneal sarcomas are often fixed as they invade muscles of the posterior parietal wall, which are themselves immobile. Fixation is not a sign of unresectability unless there is extensive involvement of irreplaceable or unremovable structures. The majority of these tumors become totally resectable with persistent dissection, though wide margins are usually not possible.

For large tumors, the tumor mass often makes it impossible to get control of the aorta and other great vessels early in the dissection as one would prefer. To insist on doing so will often lead to frustration and an inappropriate conclusion of unresectability. The approach should be to work around the periphery of the tumor and do any ‘easy' dissection first. When the dissection becomes difficult, one should go somewhere else. The only operative plan to follow is to maintain flexibility and the willingness to readjust continually your approach. It is often surprising to the less experienced how many ‘unresectable' tumors can be safely and completely resected by an experienced surgeon who refuses to give up just because the ‘going gets tough'. It is not surprising that the resectability rate varies between 35 and 100 per cent (Table 2) as so much depends upon the attitude and aggressiveness of the surgeon.

Series	*Total no. of cases/no. resected*	*Total 5-year survival (%)*
UCLA (1981)	54/33 (61%)	
Memorial (1981)	158/77 (49%)	40
MCV (1984)	47/18 (38%)	50
Roswell Park (1985)	68/27 (40%)	64
NCI (1985)	37/37 (100%)	38
Fort Sam Houston (1986)	20/7 (35%)	43
Sweden (1988)	32/16 (50%)	46
Minnesota (1989)	50/31 (62%)	48
Princess Margaret (1994)	104/45 (43%)	36
Roswell Park (1996)	87/83 (95%)	63
Memorial (1997)	151/134 (89%)	36
MEAN	(60%)	(46%)

Table 2 Resectability and survival in retroperitoneal sarcoma

Lymphomas tend to be more diffuse, in which case a limited biopsy specimen should be taken, and the margins of the tumor should be marked by clips. Rarely, localized lymphomas can be totally resected. It is more acceptable to resect totally what proves to be a lymphoma than to biopsy a potential sarcoma with the attendant risk of seeding. Surgical judgement is critical here. Unfortunately, the rarity of these tumors means that few surgeons have much experience in this evaluation.

Results of therapy

Resection rate

The resection rate varies considerably depending upon the outlook of the surgeon, but is usually about 60 per cent (Table 2). Liposarcomas and neurogenic sarcomas have the highest rate of complete resection, and leiomyosarcomas are consistently the least resectable. There is increasing evidence that subtotal or palliative

resections are of benefit to many patients, especially those with extensive but low-grade tumors.

Extent of resection

Resection of adjacent organs is commonly necessary for complete removal of retroperitoneal sarcomas and is required in up to 73 per cent of patients. Kidney and adrenal are the most commonly resected organs, followed by colon, pancreas, and small bowel.

Operative mortality

Several recent series report no deaths in the resection of retroperitoneal sarcomas. The mortality rate should be less than 5 per cent with modern anesthesia and support.

Operative morbidity

Morbidity is obviously dependent upon the extent of resection and the adjacent organs resected. Common problems are related to small-bowel and colonic ileus, small-bowel perforation, fistulas, and intra-abdominal abscess.

Survival

Interpretation of survival in the reported series (Table 2) is difficult. The data given on histological grade and stage are usually scant, the numbers of patients are generally small, and the follow-up period often short. However, some conclusions are obvious. Complete resection offers the only hope for long-term survival. Figure 2 shows the dramatic difference in survival between those patients at Roswell Park Cancer Institute between 1977 and 1994 having wide excision (margin of 2 cm or more)and those with local excision (margin of a few millimeters).

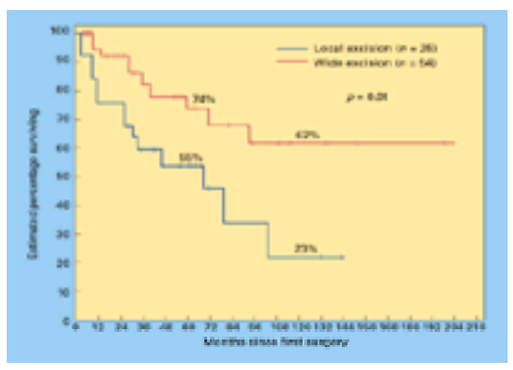


Fig. 2. Overall survival for patients having wide excision (*n*=54) and those with local excision (*n*=25) of their retroperitoneal sarcoma (*p*=0.01).

The tumor grade is the other major determinant of survival in most series. Figure 3 shows the survival curves according to the histologic grade in the Roswell Park series. As in other series of retroperitoneal sarcomas, the survival of patients with grade II sarcoma parallels that of those with grade III sarcoma. This suggests that a two-grade system may be simpler and more expedient for these unusual tumors.

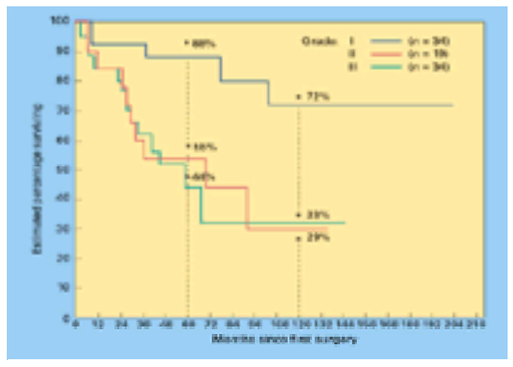


Fig. 3. Survival of retroperitoneal sarcomas (primary or locally recurrent) according to the histologic grade of the tumor, recognizing three grades (*p*=0.006).

Unfortunately, survival intervals even in patients whose tumors are totally resected continue to decline beyond 15 years, showing that retroperitoneal sarcomas are rarely totally cured. The most recent series from Memorial reported a recurrence rate of approx. 5 per cent per year from the time of the initial operation. Of the patients who were disease-free for 5 years or more from the initial resection, 40 per cent had recurrences by 10 years. Figure 4 shows the time to recurrence after resection in the Memorial series. Recurrences are eventually seen in 70 to 80 per cent of patients: these are often in the original tumor bed and often resectable. In a series from the University of California, Los Angeles, as in most, 85 per cent of the recurrences were local, and metastases occurred in only 15 per cent of patients, primarily to liver and/or lungs. Some series have reported up to seven re-excisions of the same tumor, producing long-term survival in spite of a very low eventual cure rate. These facts support the need for careful, long-term follow-up and aggressive surgical resection of recurrences when prospects of complete resection are possible.

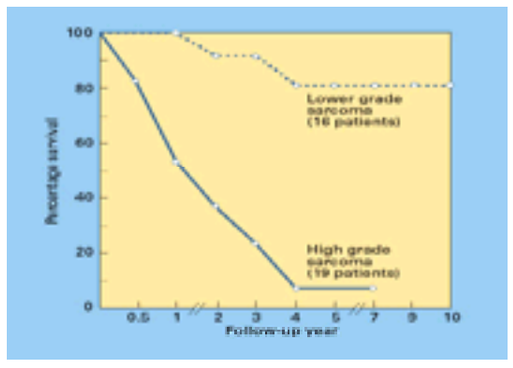


Fig. 4. Survival according to histologic grade of 35 patients with retroperitoneal sarcoma (1971–77) following complete excision.

Tumors often become less differentiated and more aggressive with each recurrence, with shorter tumor-free intervals.

The role of chemotherapy and radiation therapy

Chemotherapy

There is no evidence to suggest that chemotherapy used as an adjuvant to total surgical resection is of any benefit. It cannot be recommended outside of a clinical trial. Multiple-drug chemotherapy (usually including Adriamycin) is usually recommended, but is of questionable benefit.

Radiation therapy

The most important factor in obtaining long-term control and disease-free survival is the complete resection of the tumor, although this produces a low likelihood of long-term tumor control. Essentially all series reporting the best local control have employed high-dose radiation after total resection. In the Massachusetts General Hospital series, a dose of at least 6000 cGy was required for control. Unfortunately, review of the literature reveals that the role of radiation therapy remains unproved. The National Cancer Institute randomized patients to either intraoperative radiation therapy with low-dose external-beam therapy or high-dose external-beam therapy. There were significantly fewer local recurrences in the arm given intraoperative radiation therapy but there was no survival advantage in either treatment arm. There were significant radiation complications in both: primarily radiation enteritis in the patients receiving high-dose external-beam therapy postoperatively and peripheral neuropathy in the intraoperatively treated patients. Because these large tumors often displace most of the bowel out of the treatment field, the use of preoperative external-beam radiation therapy followed by a postresection intraoperative boost has considerable theoretical merit. This approach has been piloted and early data are encouraging, but it has not been tested in prospectively randomized trials.

Essentially all of these patients are at high risk of recurrence postoperatively, especially if they have a high-grade tumor. Clinical trials with innovative approaches to adjuvant radiation therapy are imperative as it is unlikely that more aggressive resections will have a substantial impact.

Summary

Retroperitoneal neoplasms are rare but are most commonly sarcomas, lymphomas, or benign lesions. They usually reach a large size before presenting with a palpable abdominal or pelvic mass, or with abdominal or back pain, often with weight loss and vague gastrointestinal symptoms. Of the sarcomas, liposarcomas are reported most often in the retroperitoneum, but malignant fibrous histiocytoma has been diagnosed with increasing frequency since its description. An abdominal and/or pelvic CT scan is the most important evaluation in planning resection, which is the only potentially curative therapy for retroperitoneal sarcomas. An arteriogram and inferior vena cavagram may also be helpful. An abdominal approach is usually advised for large retroperitoneal tumors to allow for the *en bloc* resection of involved adjacent organs and achieve optimum control of the main arterial and venous supply of the tumor. Potentially curable lesions should be radically resected and not removed from their pseudocapsule. Unfortunately, only about 50 per cent of tumors are resectable, and nearly 75 per cent require resection of adjacent organs. Operative mortality should be less than 5 per cent and morbidity rates should be low. Survival is dependent primarily upon the grade of malignancy and the stage, with a 50 per cent 5-year survival overall. Unfortunately, nearly 80 per cent of patients eventually suffer a recurrence, but they may benefit from repeated resections. Adjuvant chemotherapy has no role outside of a controlled clinical trial. Adjuvant irradiation is probably of benefit but requires more definitive trials.

Further reading

Alvarenga JC, Ball ABS, Fisher C, Fryatt I, Jones L, Meirion Thomas J. Limitations of surgery in the treatment of retroperitoneal sarcoma. *British Journal of Surgery* 1991; **78**: 912–16. [This large series of 120 patients focuses on the importance of multiorgan involvement and its effect on the ability to resect completely these tumors.]

Heslin MJ *et al.* Prognostic factors associated with long-term survival for retroperitoneal sarcoma: implications for management. *Journal of Clinical Oncology* 1997; **15**: 2832–9. [Important data on late recurrence are presented with long-term follow-up showing approximately 5 per cent recurrence per year from the time of initial operation.]

Karakousis CP, Velez AF, Gerstenbluth R, Driscoll DL. Resectability and survival in retroperitoneal sarcomas. *Annals of Surgical Oncology* 1996; **3**:150–8. [This series reports a remarkable 100 per cent resectability rate for 55 primary tumors and 87 per cent for 32 locally recurrent tumors, with no operative mortality.]

Shiloni E, Szold A, White DE, Freund HR. High-grade retroperitoneal sarcomas: role of an aggressive palliative approach. *Journal of Surgical Oncology* 1993; **53**: 197–203. [A strong case is made for aggressive subtotal excision in tumors that cannot be totally resected.]

Sindelar WF *et al.* Intraoperative radiotherapy in retroperitoneal sarcomas. Final results of a prospective, randomized, clinical trial. *Archives of Surgery* 1993; **128**: 402–10. [This article presents the final results of the NCI retroperitoneal sarcoma trial, with a median follow-up of 8 years and with all patients followed-up for a minimum of 5 years.]

Willett CG *et al.* Intraoperative electron beam radiation therapy for retroperitoneal soft tissue sarcoma. *Cancer* 1991; **68**: 278–83. [This report makes a strong case for preoperative radiation therapy followed by postresection intraoperative radiation therapy, which should be evaluated in a clinical trial.]

35.1 Inguinal and femoral hernias

Clare Cheek and Andrew Kingsnorth

Hernias in general
Inguinal hernias
Introduction
Aetiology
Anatomy
Symptoms and signs
Examination
Diagnosis
Direct and indirect hernias
Further investigations
Treatment
Outcomes of surgery
Complications of groin hernia repair
Femoral hernia
Anatomy
Presentation and diagnosis
Differential diagnosis
Treatment
Further reading

Hernias in general

A hernia is a weakness in the abdominal wall through which contents can protrude. A hernia can be present at birth. Hernias usually present as a bulge or lump that disappears when the patient lies down. They can be primary or recurrent; they may be reducible or irreducible. Irreducible hernias may be either incarcerated or they may be strangulated (Table 1, Fig. 1).

Primary	a hernia in which no operation has been performed
Recurrent	a hernia occurring after a previous hernia repair. Two thirds of the operations are for recurrence after operation on the previously intact groin (Nyhus's definition)
Reducible	the contents of the hernia can return to the abdominal cavity
Irreducible	the contents of the hernia are retained outside the abdomen
Strangulated	an irreducible hernia with ischaemia of enclosed contents
Richter's hernia	hernia that contains part of the circumference of the bowel wall, which can cause strangulation without obstruction (Fig. 1)
Littre's hernia	hernia that contains a Meckel's diverticulum
Maydl's hernia	hernia that contains two adjacent loops of bowel, the segment between the loops may be strangulated
Amyand's hernia	hernia containing a Meckel's diverticulum
Obturator hernia	protrusion of the hernial sac and its contents into the obturator canal, and then into the pelvic cavity, without herniation of the contents of the sac

Table 1 Definitions relating to hernias



Fig. 1. Richter's hernia (partial enterocele). The antimesenteric circumference of the bowel is first held by the rigid neck of the hernial sac, usually a femoral or obturator hernia. The situation is progressive, from (a) partial involvement of the bowel circumference without obstruction, to (b) subacute obstruction, to (c) complete obstruction and strangulation of the incarcerated bowel.

All hernias have a sac, contents, and a neck. Hernias may contain omentum, small bowel, large bowel, appendix, bladder, or ovary.

Inguinal hernias

Introduction

In various countries of the world, inguinal hernias are between 8 and 20 times more common than femoral hernias. They are 10 times more common in males than females, and 55 per cent occur on the right side. Seventy per cent of inguinal hernias are indirect and 30 per cent direct. Bilateral hernias are four times more often direct than indirect. Approximately 85 000 inguinal hernias are repaired each year in the National Health Service in England, and 750 000 in the United States.

Inguinal hernias are present at birth in 3 to 5 per cent of full-term babies and up to 30 per cent of preterm babies. Of infants presenting with a congenital inguinal hernia, 10 to 20 per cent will have a contralateral hernia. Approximately 80 inguinal hernia repairs per 10 000 live births are performed in infants aged under 1 year. Rates for hernia surgery have a second peak in the 55- to 85-year age group, in which 75 hernia repairs per 10 000 population are performed yearly.

Aetiology

Indirect inguinal herniation arises from incomplete obliteration of the patent processus vaginalis. Not all individuals with a patency will develop a hernia; at autopsy 15 to 30 per cent of adult males without clinically apparent hernia have a patent processus vaginalis. Other factors that raise intra-abdominal pressure, such as continuous ambulatory peritoneal dialysis, ascites, respiratory and urinary-outflow obstruction, may exacerbate the development of groin hernias.

Patients with connective-tissue disorders such as Ehlers–Danlos syndrome have a high incidence of hernias, and smokers with groin hernias have been shown to have abnormalities of collagen metabolism resulting in a generalized weakness of connective tissue. Genetic factors are important in the aetiology of inguinal hernia: a study of congenital indirect inguinal hernias in China on 280 families indicated that the mode of transmission was autosomal dominant, with incomplete penetrance and paternal transmission. Affected males may inherit a gene associated with indirect inguinal hernia from their fathers. Many patients recall an episode of heavy manual work such as lifting in the development of their inguinal hernia. However, the underlying weakness and predisposing cause (e.g. indirect sac, smoking) is probably a

more important contributory factor.

Anatomy (Fig. 2)

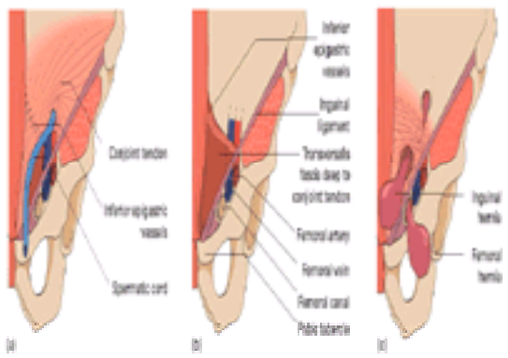


Fig. 2. The anatomy of the inguinal region.

The inguinal canal

The inguinal canal passes obliquely downwards and medially from the internal to the external inguinal ring. It is 4 cm long in the adult, but in the neonate the deep ring lies posterior to the external ring. In the male the inguinal canal contains the spermatic cord and the ilioinguinal nerve; in the female it contains the round ligament and the ilioinguinal nerve. The anterior wall of the inguinal canal is formed by the aponeurosis of the external oblique; it is reinforced in the lateral third by the fibres of the internal oblique muscle. The posterior wall of the inguinal canal is formed by the transversalis fascia; it is reinforced in the medial third by the conjoint tendon, which is the fused tendon of the internal oblique and transversus abdominis muscles.

The inferior wall of the inguinal canal is formed by the in-rolled edge of the inguinal ligament. The superior wall is formed by the arching fibres of the internal oblique and transversus abdominis.

Internal inguinal ring

This is where the round ligament or spermatic cord emerge through the transversalis fascia. Medial to the inguinal ring are the inferior epigastric vessels. The surface markings of the internal ring are 1.25 cm above a point midway between the anterior superior iliac spine and the pubic tubercle.

External inguinal ring

The external ring is a V-shaped opening immediately superior to the pubic tubercle.

Symptoms and signs (Fig. 3)



Fig. 3. Inguinal hernia.

A groin hernia may present as a lump in the groin that first appears in some patients after an episode of heavy lifting. Approximately 7 per cent of patients report a sudden event that could have caused the hernia. Often the hernia is noticed during standing or straining; typically the patient is aware that the lump disappears on lying down and is uncomfortable after exertion. Symptoms caused by complications of the hernia, such as abdominal distension and colicky abdominal pain, may be noticed by the patient before they notice the groin lump.

It is important to take a full history because hernias can develop as a result of raised intra-abdominal pressure such as occurs with respiratory, urinary or bowel problems, or during peritoneal dialysis.

Examination

A patient with a suspected groin hernia should be examined both lying and standing. They should be asked to cough so that a cough impulse and change in size of the hernia may be detected.

On inspection, a lump or bulge in the groin is noted, and this may be smaller or disappear when the patient is supine. It may be just in the groin or may extend down into the scrotum. It may be reducible or irreducible. If the hernia extends into the scrotum, it must be distinguished from a hydrocele or testicular swelling. An inguinal hernia is reduced through the superficial inguinal ring, which lies above and medial to the pubic tubercle. This distinguishes it from a femoral hernia, in which the neck of the hernia is below and lateral to the pubic tubercle. An indirect hernia can usually be controlled by pressure over the deep ring; a direct one, however, cannot, as it emerges forwards through the posterior wall of the inguinal canal medial to the deep ring.

Percussion and auscultation may be performed to determine whether the hernia contains bowel.

Diagnosis

An inguinal hernia is usually diagnosed on history and examination; further investigations are not generally required. However, in patients who have a typical history yet a hernia cannot be palpated, further investigations may be required. The differential diagnosis of an inguinal hernia is femoral hernia, hydrocele, undescended testicle, lymph nodes, lipoma, aneurysm of the femoral artery, and saphena varix.

Direct and indirect hernias

Inguinal hernias may be either indirect, the neck of the hernia emerging lateral to the inferior epigastric vessels and following the course of the inguinal canal, or direct, in which the sac emerges medial to the inferior epigastric artery.

Indirect hernias occur at all ages; direct hernias are more common in the over-40 age group. It was once thought to be an important clinical skill to be able to demonstrate the difference between the two types; however, this is notoriously inaccurate and expert surgeons have an accuracy of only 50 per cent in distinguishing between indirect and direct inguinal hernias.

Direct hernias have a smaller risk of strangulation than indirect hernias, owing to their wider necks. Indeed the repair of an easily reducible, confidently diagnosed, direct inguinal hernia, usually presenting as a prominent bulge in the groin of an elderly man, is not considered mandatory.

Further investigations

Herniography has been used for many years in the diagnosis of suspected, but impalpable, groin hernias. It involves puncture of the abdominal wall and injection of non-ionic contrast medium into the peritoneal cavity. It is highly accurate in experienced hands, but is an invasive procedure. It is also useful in distinguishing between direct and indirect hernias.

Computed tomographic scanning is a useful, non-invasive method of imaging groin hernias. It is possible to distinguish between direct and indirect hernias, but it is thought to be less sensitive when compared with herniography, and is expensive.

The role of magnetic resonance imaging in the diagnosis of impalpable hernias of the groin is yet to be fully evaluated. A recent study using dynamic real-time imaging was accurate in diagnosing impalpable hernias and was able to demonstrate the hernial sacs moving in and out of the defect.

Treatment

Surgery is the usual treatment for groin hernias, and with the use of local anaesthesia few patients should be refused surgery. Selected patients who do not wish to have surgery or who wish to work whilst awaiting surgery can be offered a truss.

Non-operative treatment is recommended only for asymptomatic, reducible, direct hernias in elderly individuals where an improvement in quality of life is unlikely. Some hernias enlarge and become symptomatic, particularly the indirect, sliding type; in such cases, and particularly in younger patients, surgery should be advised.

Trusses

Trusses have been used in the treatment of groin hernias for many years; approx. 40 000 are prescribed per annum in the United Kingdom. Trusses are appliances designed to be worn throughout the day to keep the hernia reduced. There are two main types: elastic trusses, which need replacing every 6 months, and spring trusses, which need to be renewed every 2 years. It is essential that the truss be correctly measured for and fitted if it is to benefit the patient. Patients must be fully instructed on how to fit and wear their truss. There is no evidence to suggest that truss wearing causes increased risk of recurrence after hernia surgery.

Surgical principles

The main principles of hernia surgery are herniotomy: opening and dealing with the hernia sac, and herniorrhaphy: repair of the posterior wall of the inguinal canal.

Herniotomy

A transverse or oblique skin incision is made 1 to 2 cm above the inguinal ligament and the external oblique aponeurosis is opened along the line of its fibres to the external inguinal ring. The hernial sac is dissected out from the cord structures and, if indirect, is carefully dissected up to the deep inguinal ring. The sac is opened, the contents are inspected, and reduced into the abdominal cavity. The sac is then transfixed and excised. If the sac is large and extends down into the scrotum, it is divided proximally; it is not necessary to dissect out the whole sac from the scrotum, which may actually be harmful and compromise the blood supply to the testicle. If the hernia is a sliding type, that is, the wall of the sac is formed by bladder or colon, the sac is closed with a purse-string suture and reduced into the abdomen. Direct sacs do not usually need to be opened; they are usually inverted. In children a herniotomy is all that is required; repair of the posterior wall is not necessary.

Herniorrhaphy The posterior wall is then repaired. It is strengthened by using sutures or non-absorbable mesh; there are several different types of repair, which are outlined below.

Strangulation of inguinal hernias

The risk of strangulation varies between 0.3 and 2.9 per cent per annum depending on type, and is higher in the first 3 months of the hernia developing. It occurs more commonly with indirect hernias. The patient may present with severe pain in the groin or with symptoms and signs of obstruction.

It is important that the patient be adequately resuscitated before surgery: preoperative blood tests are necessary to determine the degree of dehydration and electrolyte imbalance; abdominal radiographs will indicate if the patient is obstructed; a chest radiograph and electrocardiogram will be required before anaesthesia. Intravenous fluids and a urinary catheter to monitor urine output are essential; a central venous line may also be helpful in ensuring optimum rehydration. Several hours may be required to achieve this. Broad-spectrum antibiotics are necessary to prevent sepsis, and a nasogastric tube to decompress the stomach and small bowel is helpful in decreasing the risk of aspiration.

The principles of surgery are the same as for elective repair except that the contents of the hernial sac need to be inspected and their viability assessed. If bowel resection is required, this can be done safely through the groin. There has been concern about the use of mesh in emergency hernia repairs, but when antibiotic prophylaxis is used there is no increase in infective complications.

Anaesthesia

Groin hernias can be repaired under local, regional, or general anaesthesia. Local anaesthesia is very popular in the United States, where it is used in 70 per cent of hernia repairs, but is less commonly used in the United Kingdom (6 per cent of cases), which is due to surgical conservatism. Local anaesthesia is usually given by a skilled practitioner (surgeon or anaesthetist), either as a regional nerve block or ‘infiltrate as you operate' using lignocaine (lidocaine) or bupivacaine. A typical dose regimen is 2 mg/kg body wt of bupivacaine plain or 4 mg/kg with adrenaline (epinephrine), or 3 mg/kg body wt lignocaine plain or 7 mg/kg with adrenaline.

The recommended dose of local anaesthetic should not in general be exceeded. However, the safety margin in the recommended safe dose is wide, as illustrated by serial postoperative plasma concentrations following doses that approach the maximum recommended for lignocaine or bupivacaine. For instance, on administering lignocaine with adrenaline to the maximum dose of 7 mg/kg, peak plasma concentrations ranged from 0.23 to 0.9 mg/l; the threshold for toxicity is 5 mg/l. The administration of 20 ml 0f 0.5 per cent plain bupivacaine resulted in venous plasma concentrations of 0.07 to 1.14 mg/l; cardiovascular toxicity occurs at plasma concentrations greater than 4 mg/l. Thus many surgeons widely exceed the recommended doses without any adverse outcomes. Nevertheless, it is important that patients be closely monitored during surgery as complications due to local anaesthetic agents can occur. An anaesthetist or nurse anaesthetist should be available in the operating room during the procedure. Surgery can be technically more demanding under local anaesthesia and adequate training must be ensured.

The advantages of local anaesthesia are that it has fewer adverse effects on respiratory function than both general and regional anaesthesia, and patients return to normal activities more quickly.

General anaesthesia has improved and short-acting agents such as propofol enable prompt recovery. In the emergency procedures, general anaesthesia is usually required as the repair may be more demanding and bowel resection may be required.

Antibiotics

In emergency hernia repair, antibiotics are mandatory (see above) as there is a greater risk of infective complications postoperatively. In the elective case there has been an increase in the use of prophylactic antibiotics, especially when prosthetic materials are used in the repair, but there is little evidence to support their routine use. The use of antibiotic in the wound or by local infiltration has been found to be of benefit in reducing postoperative wound infections.

Day case

The majority of groin hernias can be repaired as a day-case procedure, which is beneficial in decreased rates of wound infection, earlier return to normal activities, and cost savings. Specialist units in the United States and Europe have shown that the majority of hernias can be repaired on an ambulatory basis. However, the British National Health Service statistics for 1993 show that only approximately 20 per cent of groin hernias were repaired as a day-case procedure.

Patients undergoing hernia repair as a day case must be generally fit, ASA (American Association of Anesthesiologists) grade 1 or 2, and must have adequate social and domestic back-up.

Patients must be assessed before admission for suitability for day-case surgery and will require adequate written instructions on their postoperative care.

Techniques for hernia repair

Sutured

Darns using a non-absorbable suture were once popular in England, but the popularization of the simple and effective Lichtenstein method has eclipsed this obsolete and inferior technique. A darn repair includes plication of the transversalis fascia supplemented by a loose weave of interlocking suture material between the conjoint tendon and the inguinal ligament. Ischaemia induced by the plication and gaps in the weave are probably responsible for the high recurrence rate.

Shouldice repair

In randomized, controlled trials the Shouldice repair has been found to give lower recurrence rates than other suture techniques. Four layers of sutures are required. Traditionally, sutures were made of stainless-steel wire, but polypropylene sutures, which are easier to handle, can be used. This method involves division and resection of the cremaster muscle. After the hernial sac has been dealt with the transversalis fascia just medial to the internal ring is elevated to separate it from the underlying fat. The transversalis fascia is the divided for the whole length of the inguinal canal down to the pubic tubercle. The flaps of the transversalis fascia are lifted up and, using a double-breasting technique, the transversalis fascia is reconstituted. The lower lateral flap is sutured to the deep surface of the upper medial flap, commencing at the medial end of the canal; then the free margin of the upper flap is sutured to the lower flap. A further two layers of sutures are inserted from the deep surface of the conjoint tendon to the upturned edge of the inguinal ligament (Fig. 4).

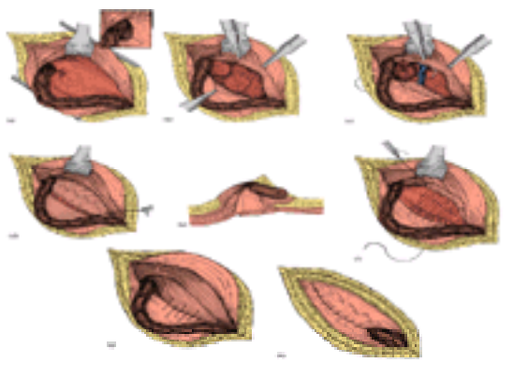


Fig. 4. Shouldice repair. (a) After the neck of the sac has been divided at the deep inguinal ring the fascia transversalis of the deep opening is identified and assessed. If the ring is normal sized the stump of the sac is reduced and no more need be done. If the ring is dilated the transversalis fascia should be carefully dissected and divided from internal ring to pubis. (c) If the subjacent extraperitoneal fat and peritoneum is bulging, a ‘trick of the trade' is to pack it down with a gauze swab; this must be removed before the sutures are snugged tight. (c) Suturing the lower lateral flap of fascia transversalis to the undersurface of the upper medial flap along the ‘white line' or ‘arch', and (d) completing the overlap of the fascia transversalis repair: the margin of the upper medial flap is sutured to the anterior surface of the lower lateral flap; a neat closure up to the cord makes a new deep ring (e). (f) The aponeurotic, white part of the internal oblique tendon and the conjoint tendon are used to reinforce the repair. (g) The anterior aponeurotic surface of the internal oblique aponeurosis is loosely sutured to the aponeurosis of the external oblique medially. (h) The external oblique aponeurosis is closed, anterior to the cord; thus the inguinal canal is reconstituted with the cord obliquely traversing it.

Mesh

Prosthetic mesh has recently become popular for repair of inguinal hernia. The Lichtenstein technique uses a patch of polypropylene mesh approximately 8 × 16 cm, tailored to the individual patient's anatomy. The mesh is trimmed and a slit is made at its lateral edge, creating two tails, an upper and a lower; the upper tail is wider (two-thirds) than the lower (one-third). The mesh is sutured to the inguinal ligament below, and overlaps and is sutured to the conjoint tendon above, using polypropylene. The tails of the mesh are both secured to the inguinal ligament. The mesh has been found to shrink by as much as 20 per cent as the fibrous tissue contracts, so it is important to use a mesh of more than adequate size. This technique has good results and is easy to learn (Fig. 5).

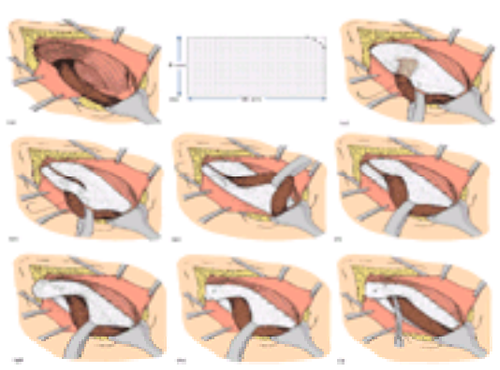


Fig. 5. Lichtenstein repair. (a) Wide dissection of the posterior wall of the canal. (b) Shaping the mesh. (c) Initial half of continuous suture to allow the mesh to overlap the pubic tubercle and appose to the inguinal ligament. (d) The mesh is slit (one-third below, two-thirds above), up to the medial margin of the internal ring. (e) The lower ‘tail' of the mesh is flipped behind the cord, followed by the continuous suture with its needle, and the cord is retracted upwards. (f) The continuous suture line along the inguinal ligament is now continued to the lateral border of the internal ring. (g) Three or four sutures tack the mesh cranially. (h) ‘Tails' are overlapped and crossed, and a single suture placed to create a new internal ring. (i) Sutures may now be placed lateral to the ring to prevent shifting and curling. An artery clip is run down between the mesh and the new internal ring to ensure an adequate aperture.

Laparoscopic repair

The first laparoscopic hernia repair was performed in 1979, since then this method has been adopted by some enthusiasts. Initial repairs closed off the internal ring with sutures, but due to the high recurrence rates, laparoscopic repairs now use mesh to repair the defect.

There are two main techniques: transabdominal preperitoneal (TAPP) and totally extraperitoneal (TEP). General anaesthesia is required for laparoscopic hernia repair. The transabdominal preperitoneal technique involves insertion of the laparoscope into the peritoneal cavity and dissection of the hernia by elevating the peritoneum behind Hesselbach's triangle. Polypropylene mesh is inserted to repair the hernia; it may be secured with staples or sutures after the repair, and the peritoneum is closed over the operation site to ensure that abdominal contents cannot lie directly in contact with the mesh.

The totally extraperitoneal approach does not involve insertion of the laparoscope into the peritoneal cavity; the dissection in this approach is in the extraperitoneal plane. The plane is developed and carbon dioxide gas is insufflated; several ports are inserted to aid in dissection. Mesh is used to repair the hernia once the sac has been reduced; there is no need to secure the mesh.

This laparoscopic approach has the advantage of fewer wound complications and an earlier return to normal activities. The operation is more expensive, as it initially involves a longer operating time and also often uses disposable equipment. It is technically demanding and has a ‘steep learning curve', and for this reason it is mainly undertaken by surgeons with a special interest in laparoscopic surgery.

Outcomes of surgery

Patients should be able to return to normal as soon as resolving pain permits; traditionally, they had been advised to have 6 weeks or even 3 months off work. A randomized, controlled trial of navy recruits showed no difference in recurrence rates when patients were randomized to return to full activities at 3 weeks or at 3 months. The major factors affecting return to activity are motivation and financial incentives. An American study compared two groups of patients and found that in those receiving workers' compensation the duration of postoperative pain was 27 days with return to work at 36.5 days. In those with commercial insurance (self-employed) the duration of postoperative pain was 7.5 days and return to work at 8.5 days.

Patients should not begin driving until 3 to 4 days after surgery as the foot reaction time does not return to normal until then.

Complications of groin hernia repair

Wound complications

Haematoma can be prevented by meticulous haemostasis. Wound infection results in an increased risk of recurrence, and it should be treated swiftly with antibiotics and if necessary opening of the wound.

Seroma occurs more commonly after mesh repair. Large, unresolving, symptomatic seromas can be treated by aspiration under aseptic conditions.

Testicular complications

The incidence of ischaemic orchitis and testicular atrophy is less than 1 per cent after primary repair of inguinal hernia. If previous scrotal or groin surgery has been performed, the risk of testicular damage is greater than 1 per cent. During dissection of the hernial sac from the spermatic cord there is a risk of damage to the testicular artery; however, there is a blood supply to the testicle from the cremasteric vessels and from branches of the internal pudendal artery. It is important that the testicle is not delivered from the scrotum during hernia repair as this will increase the risk of atrophy by disrupting the blood supply to the testicle from arterial anastomoses in the scrotum. Ischaemic orchitis usually presents 4 days after surgery with swelling and pain in the testicle; sometimes it completely resolves but in 25 per cent of cases the testicle gradually atrophies.

Nerve injury

The iliohypogastric, ilioinguinal, and genitofemoral nerves are vulnerable to injury during surgery for inguinal hernia. Division of one of these nerves during surgery, however, will cause little sensory deficit and will prevent this problem. The problem occurs if the nerve is damaged by diathermy or traction. The patient may then develop persistent pain. A local anaesthetic injection can be useful to confirm the neuralgia that has developed, i.e., the sensory nerve that has been affected. If the local anaesthetic was effective, then the nerve can be permanently ablated using phenol, or surgery can be performed to alleviate the symptoms permanently.

Port-site hernias

Small-bowel obstruction occurs in 0.3 per cent of patients after laparoscopic hernia repair, more commonly in the transabdominal approach. Occasionally this is due to a port-site hernia, a complication that can be avoided by adequate closure of the fascial defects created by the ports.

General complications

Visceral injury can occur in open hernia repair as well as during laparoscopic repair. It is important that the contents of the hernial sac be inspected before their reduction into the abdomen. The bladder can be at risk during repair of femoral and direct inguinal hernias; large bowel can be injured in repair of sliding hernias. In laparoscopic hernia repair the viscera can be injured during insertion of the trocars; there have also been several case reports of small-bowel obstruction after transabdominal preperitoneal repair due to adhesion of bowel to the exposed mesh.

Mortality

The mortality after elective inguinal hernia repair is low, but there is still a significant mortality after repair of a strangulated groin hernia when bowel resection is required. The report of the National Confidential Enquiry into Postoperative Deaths in 1991/2 reported the death rate as 7 per cent.

Recurrence

Recurrence after hernia surgery has been defined by Marsden as a weakness of the operation area necessitating a further operation or the provision of a truss (Marsden's definition). Recurrence rates of approximately 10 per cent for primary inguinal hernia repair are reported from United States audit figures. Specialist centres report figures of 1 to 5 per cent. Examination of patients after hernia surgery has found that 50 per cent are unaware of a recurrence. Although most hernias that recur do so within 5 years, recurrences can appear at any time up to 25 years after surgery: 25 per cent of recurrences have arisen at 2 years, 60 per cent by 5 years, and 75 per cent by 10 years.

Femoral hernia

Approximately 5000 femoral hernias are repaired in England in the National Health Service each year; 4 per cent of these are for recurrent hernias. In the United States, 25 000 femoral hernias are repaired annually, which represents 4 per cent of all operations for groin hernia. Femoral hernias are less common than inguinal by a factor of approximately 1 to 25 and are four times more common in women. They are found more often in elderly and multiparous individuals, but can present in the nulliparous woman or rarely in childhood (Fig. 6).



Fig. 6. Femoral hernia

Anatomy (Fig. 7)

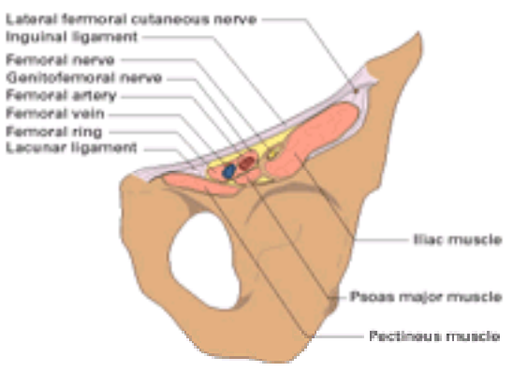


Fig. 7. Anatomical boundaries of the femoral canal.

Femoral hernias emerge through the groin from the femoral canal, which has rigid boundaries consisting of medially the lacunar ligament, anteriorly the inguinal ligament, posteriorly the pectineal ligament, and laterally the femoral vein. The femoral canal normally contains lymph nodes and loose connective tissue, which forms the characteristic ‘onion skin' over the fundus of the hernia sac.

Presentation and diagnosis

A femoral hernia may be an incidental finding or it may present as a symptomatic groin swelling. A relatively large proportion (30–80 per cent) present with strangulation and/or obstruction due to the narrow neck of the hernial sac and the sharp edge of the lacunar ligament.

Differential diagnosis

Lymph nodes, lipoma, inguinal hernia, aneurysm of the femoral artery, saphena varix, and varicocele (occasionally seen in preg-nancy).

Treatment

Surgery is mandatory once the diagnosis of femoral hernia has been confirmed. Ideally, the operation should be within a month of diagnosis and patients must be warned to present to hospital immediately if they have any signs of obstruction or strangulation. If the diagnosis of femoral hernia is in doubt, surgical exploration should be performed.

Emergency treatment

Strangulated hernias are treated surgically, but resuscitation is very important and the patients must be adequately rehydrated and receive prophylactic antibiotics before surgery. Although urgent surgery is required, several hours may be necessary to ensure the patient is optimally resuscitated.

Surgery

The principle is the same in all of the three main approaches to femoral hernia repair: dissection of the sac, inspection of the contents, ligation of the sac, and hernia repair, usually with sutures to approximate the inguinal and pectineal ligaments.

The low approach (Lockwood)

This method has been advocated for elective repair when the contents of the sac are viable; it is simple and easy.

After preparing the patient for surgery a transverse or skin-crease incision is made over the lump and the sac is dissected out from the several layers of connective tissue that may surround it. The sac is opened and the contents inspected. If viable the contents are reduced; if non-viable omentum is present it is excised. If the sac contains non-viable bowel, it is best to proceed to a lower midline incision in order to perform a bowel resection. It can often be difficult to reduce the contents of the sac and it may be necessary to make an incision in the lacunar ligament to release its neck. (An abnormal obturator artery is present in 20 per cent of cases: usually it lies on the lateral side of the femoral canal; rarely it lies medially close to the lacunar ligament; and occasionally this can lead to troublesome bleeding when the lacunar ligament is divided.) The sac can then be ligated and excised.

The repair is then made with a non-absorbable suture to approximate the inguinal ligament to the pectineal ligament, taking care to protect the femoral vein, which lies laterally (Fig. 8).

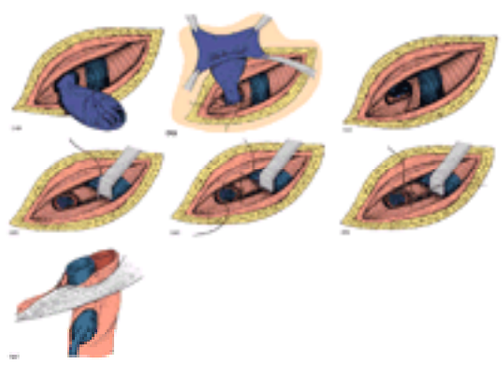


Fig. 8. Femoral hernia repair, the low approach. (a) The sac mobilized. (b) Closure of the sac. (c) The canal after closure of the sac. (d) Retraction of the femoral vein laterally enables visualization of the pectineal (Cooper's) ligament; the first suture is introduced but not approximated. (e) The next suture picks up the inguinal ligament and the subjacent iliopubic tract of the fascia transversalis. Care must be taken to avoid the cord structures in the wall. (f) The sutures are placed to form the base of an isosceles triangle with the apex at the pubic tubercle. (g) The knot is tied deeply at the medial side away from the femoral vein.

Transinguinal (Lotheissen)

This technique involves a standard oblique groin incision above and parallel to the inguinal ligament, followed by incision of the external oblique and opening of the inguinal canal. The femoral hernia is approached by opening the transversalis fascia at the back wall of the canal. The hernial sac is then reduced and the repair done as described above. This technique is less popular as it is thought to weaken the inguinal canal.

High approach (McEvedy)

This approach is generally considered the optimum method for strangulated femoral hernias. Classically an oblique or paramedian skin incision is made. However, an

acceptable modification is the use of a unilateral Pfannenstiel incision, which can be extended across the midline if a laparotomy becomes necessary. The rectus sheath is then opened longitudinally and the rectus muscle retracted medially. The transversalis fascia can then be pushed away from the inguinal ligament to expose the femoral canal. If the sac is empty it can be withdrawn into the abdomen and the hernia repaired from above by approximation of the inguinal and pectineal ligament using non-absorbable sutures. If the hernia sac cannot be reduced, the peritoneum is opened and by gentle traction from above the contents are withdrawn from the sac. If there are still difficulties, the use of pressure from below and division of the lacunar ligament should enable this manoeuvre to be completed. If the contents are not viable a resection can be performed through the same incision. The hernia can then be repaired and the peritoneum closed, and then the rectus sheath repaired.

A similar type of repair can be performed in a preperitoneal approach using a lower midline incision.

Laparoscopic repair

Femoral hernias can be repaired laparoscopically either by a transabdominal or extraperitoneal approach. However, the laparoscopic operation requires extensive dissection and has to be done under general anaesthesia, the equipment required is more expensive, and at present there is no evidence to suggest that it is more efficacious than the open repair.

Plug repair

Plugs of mesh can be inserted into the femoral canal to repair the hernia. This has been found to give low rates of recurrence, but migration of the mesh and infections, which are thought to occur more commonly when mesh is rolled up, have been described.

Which repair?

The traditional teaching has been that the low approach should be used for elective repair and the McEvedy should be performed for emergency hernia repair. Systematic review of the literature has shown little evidence to support this and it is quite acceptable to use any of the approaches. However, the transinguinal approach may result in a higher recurrence rate.

Further reading

Cheek CM, Black NM, Devlin HB, Kingsnorth AN, Taylor RS, Watkins D. Systematic review on groin hernia surgery. *Annals of the Royal College of Surgeons of England* 1998; **80** (Suppl. 1). [A meta-analysis of 4977 citations from the hernia literature, providing evidence-based best practice.]

Devlin HB, Kingsnorth AN. *Management of abdominal hernias*, 2nd edn. Chapman and Hall, London, 1998.[The definitive English language text.]

Royal College of Surgeons of England. *Guidelines on the management of groin hernia in adults*. RCS, London, 1992. [Practical care pathways from diagnosis to operation and postoperative care.]

35.2 Umbilical, epigastric, and rarer abdominal wall hernias

C. Daniel Smith

Umbilical hernia
Congenital hernia
Infantile hernia
Adult paraumbilical hernia
Epigastric hernia
Anatomy
Diagnosis
Treatment
Spigelian hernia
Anatomy
Diagnosis
Treatment
Obturator hernia
Anatomy
Diagnosis
Treatment
Lumbar hernias
Anatomy
Diagnosis
Treatment
Sciatic hernias
Anatomy
Diagnosis
Treatment
Perineal hernias
Anatomy
Treatment
Further reading

Umbilical hernia

There are three distinct types of hernia that occur around the umbilicus: (i) congenital (omphalocele or exomphalos), (ii) infantile umbilical hernia, and (iii) adult paraumbilical hernia.

Congenital hernia

This condition occurs in 1 in 5000 births and is associated with other serious congenital abnormalities in 60 per cent of cases.

Anatomy

During intrauterine development the amniotic sac contains the embryologic midgut. At 10 weeks of gestation the gut normally returns to the abdominal cavity. When this doesn't occur, the umbilical canal fails to close and at birth a broad funnel-shaped defect is present at the umbilicus. Viscera covered only by peritoneum protrude through this abdominal wall defect.

Diagnosis

The diagnosis is immediately evident at the time of birth, with an obvious protrusion of abdominal viscera through the umbilicus. Its location in the midline and the presence of peritoneum covering the herniated viscera distinguishes this congenital abdominal wall defect from gastroschisis, where the abdominal wall defect is off midline, and the herniated viscera are uncovered by either peritoneum or skin.

Treatment

Urgent surgical repair should be performed before rupture of the sac occurs or infection supervenes. With small defects (< 5 cm in diameter), the viscera can usually be returned to the abdominal cavity with closure of the abdominal wall and skin defect. In larger defects, the abdominal cavity is usually too small to allow return of the viscera and primary closure. Initial treatment usually aims to cover the peritoneal sac with an artificial material such as a Silastic bag, which is sutured to the edges of the defect, followed by staged reduction of the contents with eventual skin closure. Inevitably an incisional hernia results which can be repaired in later childhood.

Infantile hernia

This hernia is present in 10 per cent of Caucasian infants (male:female ratio is 2:1) and 90 per cent of children of African decent. In babies of low birth weight, the incidence is as high as 75 per cent. The occurrence is also increased in Down syndrome, Beckwith–Wiedemann syndrome, and in the presence of ascites.

Anatomy

At birth, following division of the umbilical cord, the stump heals by granulation and scarring to fuse with the umbilical ring of the abdominal wall. Failure of fusion at the abdominal wall allows a peritoneal sac to protrude, usually at the superior margin of the ring. The infantile hernia, as opposed to the congenital type, is always covered with skin, and reaches its maximal size 1 month after birth.

Diagnosis

This hernia is usually symptomless and presents as an easily reducible lump which becomes more prominent during crying and coughing. Incarceration or strangulation of this hernia is extremely rare, and congenital umbilical hernias rarely enlarge over time. The hernia contents usually remain virtually unchanged in size until just before final closure.

Treatment

This hernia will spontaneously disappear in 93 per cent of children by the age of 2 years. Consequently, operative treatment in the newborn baby is deferred to allow time for spontaneous closure. Tapes, binders, and trusses to reduce the hernia contents are not recommended as they may lead to skin infection or necrosis. Surgical repair is indicated for any symptoms associated with the umbilical hernia, or if the hernia persists beyond 2 years of age. Hernias greater than 2 cm in size are less likely to close spontaneously. The operation is performed under general anesthesia on an outpatient basis. A semicircular infraumbilical skin incision allows elevation of the umbilicus and exposure of the edges of the fascial defect. Herniated content is usually reduced without entering the peritoneum. The fascial defect is then closed transversely with absorbable or non-absorbable suture. The base of the umbilicus is sutured to the fascia to invert the umbilicus and restore its normal contour.

Adult paraumbilical hernia

Most adult umbilical hernias are acquired. About 10 per cent of adult umbilical hernias are congenital hernias carried into adulthood.

Anatomy

The superior rim of the umbilicus is the site of attachment of the round ligament and the remnants of the urachus and umbilical arteries, thus creating a weak area in the abdominal fascia. Additionally, the lowest tendonous insertion of the rectus abdominis muscle into the linea alba is at the level of the superior umbilical rim. Stretching of the abdominal wall due to multiple pregnancies, ascites, or obesity predisposes to the development of an umbilical hernia. The condition usually occurs after the age of 35 years and, due to its association with pregnancy, is five times more common in females.

Diagnosis

This hernia tends to enlarge progressively over time and may be asymptomatic depending on size and body habitus. It may produce localized pain as the fascial defect enlarges or herniated content stretches overlying subcutaneous tissue and skin. Gastrointestinal symptoms commonly occur owing to traction between the hernia contents and the stomach or transverse colon. When the hernia sac contains bowel, colic due to intermittent intestinal obstruction is possible. Thinning, discoloration, and necrosis of the skin may occur in patients with larger hernias. The paraumbilical hernia usually has a small neck and incarceration and strangulation are common. Diagnosis can be particularly challenging in the obese patient, where an abdominal wall defect can be impossible to feel, and a large hernia sac and content can be hidden within an abdominal pannus. CT scanning or ultrasound may be the best diagnostic tests in obese patients in whom an umbilical hernia is suspected.

Treatment

Once diagnosed, an umbilical hernia should be repaired. In select cases where operative risk is excessive or in bedridden patients, umbilical hernias with a large defect and a low likelihood of incarceration need not be repaired. However, eventually these large hernias lead to disability and pain, requiring repair at a time when loss of abdominal domain makes repair extremely problematic. The elderly and those with ascites are at especially high risk of significant complications from an untreated umbilical hernia. In the presence of ascites, rupture of an umbilical hernia or necrosis and infection of overlying skin has a mortality of up to 30 per cent. Therefore, in patients with ascites, elective repair should be offered before these complications develop. Medical control of ascites perioperatively is critical to minimize wound complications and recurrence. Tapes, binders, and trusses often lead to skin infection or necrosis and are not recommended. Preoperative weight loss in the obese is extremely important to increase the likelihood of a successful operative repair.

Small hernias can be repaired under local anesthesia. Larger hernias require general anesthesia. Most umbilical hernias can be repaired on an outpatient basis. Elderly patients and those with ascites or medical comorbidities will require admission to hospital. A curvilinear infraumbilical incision provides the best access for encircling the hernia sac at its base and exposing the fascial edges of the defect. The raphe of the umbilicus often requires release and dissection off of the hernia sac. Entering the peritoneum is not necessary unless abdominal content is adherent to the sac or incarcerated and requiring investigation for viability. Small hernias (< 3 cm in diameter) can usually be closed primarily in a transverse fashion using interrupted non-absorbable sutures. A fascial edge of 1 cm should be exposed to ensure adequate fascial bites are taken. Larger hernias may require repair with a non-absorbable mesh sutured securely to the fascial edges, thereby allowing a ‘tension-free' repair. Securing the mesh to the undersurface of the fascia creates a ‘buttressed' repair which may have mechanical advantages over an onlay repair. To maximize this buttressed effect, when repairing recurrent or especially large primary umbilical hernias, laparoscopy has been used to place a large piece of mesh intraperitoneally, and thereby maximize the mesh overlying normal abdominal wall. The base of the mobilized umbilicus is sutured to the fascia using an absorbable suture to invert the umbilicus and restore its normal contour.

Avoidance of heavy lifting is recommended for 4 to 6 weeks after operative repair of an umbilical hernia. Following a posterior buttressed repair, return to heavy lifting and usual activity may be allowed earlier. Patients with ascites require careful medical control of their ascites postoperatively. In those with ascites, paracentesis or placement of a peritoneovenous shunt at the time of herniorrhaphy may help minimize postoperative wound breakdown and recurrence.

Epigastric hernia

All primary hernias occurring in the midline of the abdomen, with the exception of those of the umbilicus, are hernias of the linea alba. These hernias are far more common above the umbilicus and are termed epigastric hernias. An epigastric hernia is present in 5 per cent of individuals at autopsy, and 25 per cent of individuals have multiple hernias. This hernia may present at any age, but is most common between 20 and 50 years of age, and is three times more common in men than in women.

Anatomy

Epigastric hernias are protrusions of preperitoneal fat through a fascial defect in the decussating fibers of the supraumbilical portion of the linea alba. The defect usually occurs where the linea alba is pierced by a blood vessel. A peritoneal sac may accompany fat through the defect, containing omentum or rarely bowel.

Diagnosis

The majority of epigastric hernias are asymptomatic. Vague upper abdominal pain and nausea, associated with epigastric tenderness may be present. These symptoms tend to be more severe when the patient is supine, owing to traction on the hernia contents. A lump, which may be tender, is usually palpable in non-obese subjects. Gangrene of the contents of the hernia occasionally occurs, producing severe epigastric tenderness and localized muscular rigidity. These features may mimic those of an intra-abdominal catastrophe.

The presence of a non-tender epigastric hernia should never be considered to be an adequate explanation for dyspepsia or epigastric pain except following extensive investigation of the upper gastrointestinal tract. Conversely, in obese patients with chronic upper abdominal symptoms, an epigastric hernia may remain undiagnosed for years because it is often not palpable. CT scanning or ultrasound evaluation may be necessary to diagnose an epigastric hernia in the obese patient.

Treatment

Once diagnosed, these hernias should be repaired. As with umbilical hernias, select patients with large defects and a low risk of incarceration may not need repair. This circumstance is very unusual. Only in the non-operative candidate should any form of truss or external compression be utilized. An upper midline incision affords the best exposure. The hernia sac is dissected free from the surrounding subcutaneous tissues and at least 1 cm of the fascial edges of the defect exposed. Since up to 25 per cent of epigastric hernias are multiple, a thorough search for additional hernias should be carried out. If other midline defects are found, the linea alba is incised incorporating the multiple hernias into one defect. The hernia is then closed longitudinally with interrupted or running non-absorbable suture. Wound drains are unnecessary and, depending on the size of the repair, patients are instructed to avoid heavy lifting for 4 to 6 weeks postoperatively.

Spigelian hernia

Spigelian hernias are uncommon. They are most common in women over 50 years of age.

Anatomy

The transversus abdominis muscle becomes aponeurotic at the semilunar line which stretches from the ninth rib to the pubic tubercle. The part of the aponeurosis between the semilunar line and the lateral edge of the rectus muscle is called the spigelian aponeurosis. This area of transition from muscle to aponeurosis is an area of potential abdominal wall defects. It is through such a defect that a spigelian hernia emerges, passing between the fibers of overlying internal oblique muscle and spreading out deep to the external oblique muscle. Most spigelian hernias appear below the umbilicus, level with the termination of the posterior rectus sheath (linea semicircularis), close to a point one-third of the way along a line between the umbilicus and the anterior superior iliac spine. The hernia usually lies lateral to the rectus sheath and extends out towards the iliac fossa (Fig. 1); occasionally it lies within the sheath alongside the rectus muscle (Fig. 2). Only rarely does it penetrate the external oblique muscle to lie subcutaneously.

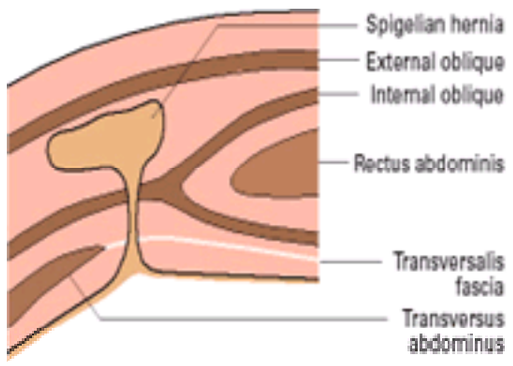


Fig. 1. A spigelian hernia emerging lateral to the rectus sheath.

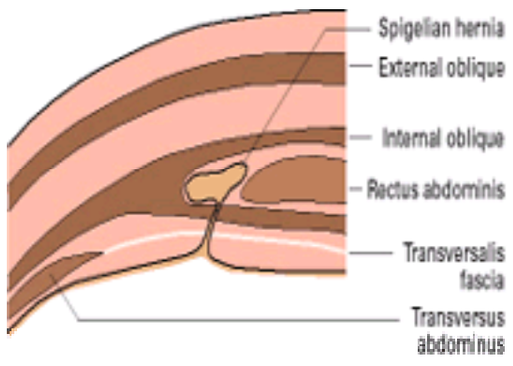


Fig. 2. A spigelian hernia emerging within the rectus sheath.

Preperitoneal fat most commonly herniates, but a hernia sac of peritoneum may contain small bowel, colon, or omentum. The fibrous bands of the spigelian fascia give these hernias a rigid neck making incarceration and strangulation common.

Diagnosis

Because the hernia sac or content is usually deep to the external oblique muscle, these hernias are difficult to palpate and the diagnosis is often obscure, especially in the obese patient. Frequently, the only symptoms are obscure abdominal pain or small bowel obstruction. In a thin standing patient, a mass with a cough impulse may be palpable in the iliac fossa; this disappears on lying down. Twenty per cent of these hernias have strangulated at presentation resulting in a tender mass in the abdominal wall which may be difficult to differentiate from an abdominal wall hematoma, a muscular tear, or an intra-abdominal inflammatory mass. Ultrasound and CT may be useful in confirming the diagnosis, especially in the obese patient. Recently, laparoscopy has proved useful in both the diagnosis and repair of spigelian hernias.

Treatment

Once diagnosed or suspected, these hernias should be operatively repaired. A transverse incision is made over the site of localized tenderness or palpable mass, and the external oblique fibers split to expose the hernia sac. Herniated content is reduced, the sac excised, and the defect closed by direct suture. If the diagnosis is in doubt, or the location of the hernia unclear, laparoscopy through an infraumbilical site will allow identification of the abdominal wall defect for subsequent laparoscopic or open repair. Alternatively, a midline laparotomy with reduction of the hernia and repair from the inside can be performed.

Obturator hernia

These hernias are rare, with fewer than 600 cases having been reported in the world literature. Most surgeons will never see an obturator hernia.

Anatomy

The hernia follows the obturator canal which courses between the superior pubic ramus and the obturator membrane and normally transmits only the obturator nerve and vessels. The sac of an obturator hernia exits the obturator canal and spreads out deep to the adductor muscles in the groin, where it is difficult to detect clinically. The sac usually contains small bowel, often as a Richter's type hernia. Less commonly, omentum, colon, fallopian tube, ovary, or bladder herniate. Strangulation is common because of the size and rigidity of the obturator canal.

Diagnosis

The majority of these hernias occur in elderly women who have recently lost weight. The higher incidence in women is probably due to the wider pelvis and flatter angle of the obturator canal. Most patients present with acute groin pain, or abdominal symptoms ranging from small bowel obstruction to mild and obscure discomfort. Approximately one-third of patients report having had similar episodes in the past. Occasionally, straining due to constipation or other factors appears to precipitate herniation.

The diagnosis is rarely made preoperatively; it is usually made during laparotomy for small bowel obstruction. Even if the correct diagnosis is entertained preoperatively, the pathognomonic signs (Table 1) are present in fewer than 50 per cent of cases.

1. Howship–Romberg sign—pain radiating to medial thigh with hip extension, internal rotation, and adduction
2. Hannington-Kiff sign—loss of adductor reflex elicited 5cm above the knee (compression on obturator nerve)
3. Bruising below the medial edge of the inguinal ligament due to strangulation and blood in the hernia sac
4. A palpable tender mass on vaginal examination (per rectal examination)

Table 1 Pathognomonic signs of obturator hernia

Treatment

Whether known preoperatively or found intraoperatively, an obturator hernia should always be repaired. If the diagnosis is made preoperatively, an extraperitoneal

approach provides ideal exposure for hernia reduction and repair. However, since the diagnosis is rarely made preoperatively, most obturator hernias are repaired during urgent laparotomy for small bowel obstruction. Regardless of approach, after placing the patient in the Trendelenburg position, the hernia sac can be seen disappearing into the obturator canal. If the hernia has reduced spontaneously, a defect can be felt just below the ischiopubic tract at the obturator membrane. An incarcerated hernia should be reduced by gentle traction. If compromised bowel is present in the hernia, the bowel should be clamped before reduction to prevent spillage. Rarely, it may be necessary to enlarge the obturator canal by incising posteromedial to the neck of the hernia, thereby avoiding damage to the obturator nerve. After reduction, bowel is carefully inspected and non-viable bowel resected.

Without infection or spillage of intestinal content, the defect is repaired with an overlay of non-absorbable mesh. Primary suture closure or overlay of urinary bladder may be necessary when spillage has occurred. The mortality in these patients remains high (13 to 20 per cent) primarily due to their advanced ge and the small number of correct preoperative diagnoses.

Lumbar hernias

Lumbar hernias are uncommon, with fewer than 300 cases having been reported in the literature. This hernia is most commonly the consequence of trauma (penetrating or operative). More rarely it is congenital or acquired.

Anatomy (Fig. 3)

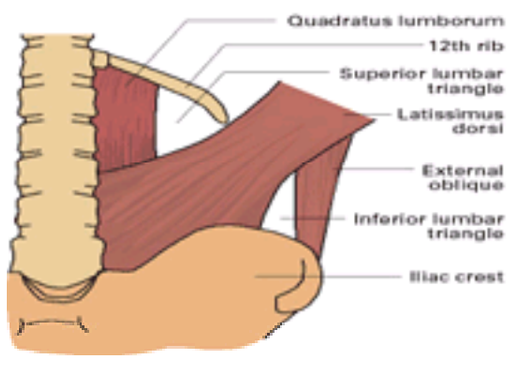


Fig. 3. The anatomy of the superior and inferior lumbar triangles.

In the area between the twelfth rib and the iliac crest a number of back and abdominal muscles come together and create an area of potential weakness. This area can be divided into the superior and inferior lumbar triangles. The inferior triangle (of Petit) is the usual site of congenital lumbar herniation, which accounts for 25 per cent of cases. The boundaries of this area are the iliac crest inferiorly, the latissimus dorsi muscle superomedially, and the posterior boundary of the external oblique muscle superolaterally. Acquired hernias are rare in this area, except when both internal and external tables of iliac crest have been removed for bone grafting.

The quadratus lumborum, twelfth rib, and internal oblique muscle form the superior lumbar triangle. This is the usual site of acquired hernias secondary to trauma (surgery, penetrating injury), infection (such as wound infection following nephrectomy), or of those occurring spontaneously.

These hernias most commonly contain small or large bowel and are at little risk of strangulation owing to their wide neck.

Diagnosis

Lumbar hernias usually occur in middle-aged men. They present as a lumbar bulge that appears with standing and disappears on lying down. The only symptom is usually a dull ache. Large hernias may contain a significant segment of colon causing abdominal bloating or constipation. CT scan may be helpful in diagnosis, especially in the obese patient.

Treatment

Lumbar hernias should be repaired operatively if symptomatic, or if there are signs of strangulation or incarceration. The hernia can be repaired through a flank or posterior approach, with the patient in the lateral position, or through an anterior retroperitoneal approach. The sac is emptied and closed off. The muscle defect is best repaired using non-absorbable suture and mesh. If the hernia is small it may be possible to coapt adjacent muscles with non-absorbable sutures, but in most cases simple suture repair has a high failure rate. When mesh is used, it is ideally placed deep to the defect in a buttressed fashion ensuring adequate overlay and secure fixation to the edges of the defect. Alternatively, a flap of fascia lata and gluteus maximus can be mobilized and rotated up to close the defect. Giant lumbar hernias are unlikely to strangulate and have a high recurrence rate. These may be best treated non-operatively.

Sciatic hernias

These are extremely rare hernias.

Anatomy

The hernia sac exits the pelvis through either the greater sciatic foramen, above or below the piriformis muscle, or, most commonly, through the lesser sciatic foramen. The sac then lies deep to the gluteus maximus, where it is well hidden unless it is large enough to protrude below the buttock crease.

Diagnosis

Diagnosis is usually made at the time of laparotomy for small bowel obstruction caused by the hernia. The bowel is seen disappearing out through a posterior pelvic defect, behind the broad ligament in women.

Treatment

As with obturator hernia, the patient is placed in the Trendelenburg position and the contents of the sac reduced. The neck of the sac is ligated, and the defect is covered by fascia mobilized from the piriformis muscle.

Perineal hernias

Perineal hernias are uncommon. They are typically the result of prior pelvic surgery.

Anatomy

They occur through defects in the muscular pelvic floor and usually follow pelvic exenteration or abdominoperineal resection of the rectum. The hernia contents are usually small bowel or bladder. There are rarely any associated symptoms.

Treatment

The usual approach is transabdominal. Once small bowel and bladder have been mobilized out of the pelvis, the defect is closed with non-absorbable mesh. If there is potential for infection, synthetic implants are best avoided and gracilis muscle can be transposed from the thigh to form a sling across the defect via a perineal

approach.

Further reading

Artioukh DY, Walker SJ. Spigelian herniae: presentation, diagnosis and treatment. *Journal of the Royal College of Surgeons of Edinburgh* 1996; **41**: 241–3. [A series of 19 patients with spigelian hernia reviewing presentation, diagnosis, and treatment.]

Felix EL, Michas C. Laparoscopic repair of spigelian hernias. *Surgical Laparoscopy and Endoscopy* 1994; **4**: 308–10. [Details technique of laparoscopic diagnosis and repair in four patients with spigelian hernia.]

Gianetta E *et al.* Hernia repair in elderly patients. *British Journal of Surgery* 1997; **84**: 983–5. [Detailed experience with types of hernias in the elderly, and their repair.]

Losanoff J, Kjossev K. Sciatic hernia. *Acta Chirurgica Belgica* 1995; **95**: 269–70. [Case report of a young patient with a sciatic hernia, including a review of techniques of repair.]

Naude G, Bongard F. Obturator hernia is an unexpected diagnosis. *American Journal of Surgery* 1997; **174**: 72–5. [Review of experience with 6 patients with obturator hernia, emphasizing the difficulty in diagnosis, and frequently fatal outcome.]

Nyhus LM, Condon RE, eds. *Hernia*, 4th edn. J.B. Lippincott Co., Philadelphia, 1995. [Contemporary chapter by one of the world's most recognized herniologists.]

Skinner MA, Grosfeld JL. Inguinal and umbilical hernia repair in infants and children. *Surgical Clinics of North America* 1993; **73**: 439–50. [Review of the unique aspects of pediatric hernias including their etiology, diagnosis, and repair.]

So JB, Palmer MT, Shellito PC. Postoperative perineal hernia. *Diseases of the Colon and Rectum* 1997; **40**: 954–7.

35.3 Wound dehiscence, incisional hernia, and parastomal hernia

Adrian Savage and Peter M. Lamont

Preoperative factors
Operative factors
Type of incision
Technique of closure
Suture material
Postoperative factors
Clinical features of wound disruption
Management of wound dehiscence
Incisional hernia
Clinical features
Management and repair
Parastomal hernia
Management and repair
Further reading

Wounds may fail in one of two ways. Wound dehiscence describes partial or complete disruption of an abdominal wound closure with protrusion or evisceration of the abdominal contents. Incisional hernia is defined as an abnormal protrusion of a viscus through the musculoaponeurotic layers of a surgical scar. Since the time at which a scar may be said to have formed is open to debate, the full healing of the skin incision is used to make a convenient distinction between wound dehiscence and incisional hernia. Dehiscence of the wound occurs before cutaneous healing, while incisional hernias lie under a well-healed skin incision. A parastomal hernia is the result of an incisional hernia occurring adjacent to a surgically created stoma. The factors contributing to these complications are similar and wound dehiscence may be associated with the subsequent development of an incisional hernia.

Disruption of an abdominal wound occurs in less than 1 per cent of patients undergoing major abdominal surgery. Incisional hernia is more common, occurring in approximately 10 per cent of patients. The cause of wound failure is related to the preoperative condition of the patient, the technique of wound closure, and postoperative complications.

Preoperative factors

The preoperative factors implicated in wound dehiscence and incisional hernia are advanced age, male sex, previous irradiation, jaundice, uraemia, anaemia, diabetes, malnutrition, malignant disease, vitamin C depletion, obesity, and the administration of steroids or cytotoxic drugs. Clinical studies of these variables in the background to wound failure are sparse since they rarely occur in isolation. Experimental studies have, however, found delayed healing of laparotomy wounds in uraemic and anaemic animals, supporting a role for these factors in the cause of wound dehiscence.

Operative factors

Norris wrote that ‘the elimination of postoperative wound dehiscence is entirely within the jurisdiction of the operating surgeon'. The type of incision, the choice of suture material, and the method of wound closure are all of importance in subsequent wound dehiscence but less important in the development of incisional hernia.

Type of incision

Midline, standard paramedian and transverse incisions all have a similar incidence of incisional hernia. Upper abdominal incisions may be more prone to dehiscence than lower abdominal incisions. Herniation is considerably more common through incisions more than 18 cm in length and when a stoma is created through the wound. There is a very low incidence of incisional hernia (0.37 per cent) and no wound dehiscence associated with the use of a lateral paramedian incision, but this approach is time-consuming and not suitable for emergencies. Appendicectomy wounds and Pfannensteil incisions have a particularly low incidence of wound dehiscence and incisional herniation.

Incisional hernias through port sites used for laparoscopic surgery are now being reported, with an incidence of 0.021 to 0.77 per cent. Except in infants, such hernias are confined to the sites of 10-mm or larger trocars. Such incisions are difficult to close at the fascial layer without extending the skin incision, which would negate to some extent the cosmetic advantage of the laparoscopic approach. These hernias are important because of the propensity of small bowel to become trapped as an incarcerated Richter's hernia.

Technique of closure

The theoretical advantages of a mass closure of abdominal incisions have found support in many clinical studies on the prevention of wound dehiscence. For example, the wound dehiscence rate with layered closure with catgut was 10.3 per cent compared with a rate of 0.93 per cent after mass closure with interrupted steel sutures. The use of deep tension sutures to support a layered catgut closure reportedly prevents wound dehiscence. A reduction in the rate of wound dehiscence from 3.8 per cent to 0.8 per cent was found on changing the technique of closure from layered catgut to mass closure with nylon or Dexon®. However, mass closure and layered closure of abdominal incisions are associated with similar rates of incisional hernia.

The technique of mass closure depends on the placement of sutures through all layers of the anterior abdominal wall, except the skin, at a distance of more than 1 cm from the edge of the musculoaponeurotic layer. The earliest description of this technique advocated the use of interrupted sutures of stainless-steel wire placed as a figure-of-8, ‘far and near' suture. Subsequent studies have shown that this is unnecessary and a continuous suture is no less effective. Such sutures should not be placed more than 1 cm apart, and the total length of suture used in a continuous mass closure should be more than four times the length of the wound. The sutures should not be placed under tension, which may result in ischaemia of the tissue enclosed within the mass closure and the development of incisional hernia. Multiple ‘buttonhole' hernias may appear along the edges of an abdominal wall closure where the sutures have torn through the tissues as the wound has swollen in the postoperative period.

Suture material

The use of catgut in the closure of all but appendicectomy wounds is now uncommon because of the high rate of wound dehiscence. Stainless-steel wire has long since given way to nylon for ease of use. The use of nylon sutures and a mass closure technique is associated with an incidence of wound dehiscence below 1 per cent. Jenkins reported only one dehiscence in 1505 abdominal wound closures. However, persistent wound sepsis and discharging sinuses have been associated with the use of non-absorbable sutures, both braided and monofilament. Modern synthetic absorbable sutures that retain their strength far longer than catgut have certain theoretical advantages. One is that the persistence of suture material over many years may result in the formation of ‘buttonhole' hernias where the suture material itself has eroded defects through the muscle layers. Polyglycolic acid (Dexon®), polygalactin (Vicryl®), and polydioxanone (PDS®) have been used successfully for the closure of laparotomy wounds, though polyglycolic acid may not absorb efficiently in the presence of sepsis. Polydioxanone is a monofilament suture similar in appearance and ease of use to nylon, and with the same rate of wound dehiscence and incisional hernia.

Postoperative factors

Increased intra-abdominal pressure due to inadequate postoperative analgesia, vomiting, the development of a postoperative chest infection resulting in coughing, and gross distension from paralytic ileus are important in both wound dehiscence and incisional hernia. Improvements in anaesthetic technique, postoperative care for the prevention of chest infection, and good postoperative analgesia have contributed to a reduction in the incidence of wound disruption in recent years. Wound sepsis also predisposes to the development of wound dehiscence and incisional hernia.

Clinical features of wound disruption

Wound disruption may be occult or overt, and partial or complete. Overt wound dehiscence follows removal of the skin sutures. The skin incision may partly or completely open to allow frank evisceration or show the presence of a herniation of bowel through a partial or complete defect in the musculoaponeurotic closure.

Occult wound dehiscence occurs with disruption of the musculoaponeurotic layers beneath intact skin sutures.

Wound dehiscence is at least twice as common in men as in women and is more common in patients over the age of 60 years. Wound disruption occurs on the sixth to ninth postoperative day in over 55 per cent of patients. Twenty-one per cent of the dehiscences occur on removal of the sutures. The patient may show signs of gross dehydration, a rise in temperature and pulse rate, and a peripheral leucocytosis, especially if an occult dehiscence has been overlooked for more than 24 h. Signs of bronchopneumonia and meteorism may also be present. A copious serosanguinous discharge presages dehiscence in about one-third of patients. Alternatively, a boggy swelling or frank sepsis of the wound may be the only signs of dehiscence of the fascial layer.

Management of wound dehiscence

The first priority in the treatment of wound dehiscence is the correction of fluid depletion and electrolyte disturbance. Patients are often dehydrated, and cardiovascular collapse secondary to sepsis and shock from evisceration of the bowel must be treated before giving a general anaesthetic for repair of the dehiscence. If occult dehiscence is suspected, removal of skin sutures allows direct inspection of the fascial closure. While resuscitation is in progress, the dehiscence is covered by the liberal application of gauze swabs soaked in normal saline; a Velcro® corset or many-tailed bandage may prevent further evisceration.

Confirmed dehiscence of an abdominal wound is best treated by resuturing. At operation the skin sutures and the remnants of the previous fascial closure are removed. The edges of the wound are debrided of necrotic tissue and the wound resutured with a careful mass closure using No. 1 nylon or polydioxanone. Gross abdominal distension may be due to an ileus or intestinal obstruction secondary to the early formation of adhesions. Decompression of the bowel by retrograde milking of the intestinal contents into the stomach and nasogastric aspiration may considerably ease the process of closure of the wound, and the division of any obstructing adhesions is important.

Very rarely, the patient is unfit for surgery or the disrupted wound is too grossly contaminated to allow immediate surgery. Packing of the wound to return the bowel to the abdominal cavity followed by strapping may allow the patient's condition to improve over a few days so that the wound may be closed as a secondary procedure. Wound dehiscence treated conservatively is inevitably followed by the development of incisional hernia. Even after resuturing, incisional hernia develops in almost 50 per cent of patients.

The prognosis of wound dehiscence is grave, and becomes worse with advancing age and with gross suppuration of the wound. Mortality rates of between 15 and 24 per cent have been reported. The outcome is better if the wound disruption is recognized and treated early, in patients whose wounds are clean, and if prolapse of the intestine does not occur. Death is most commonly due to multisystem failure.

Incisional hernia

Clinical features

Incisional hernias can develop at any age. Although clinical evidence of incisional hernia may be delayed for more than 10 years after laparotomy and less than half are apparent at 1 year, the use of radiopaque markers has shown separation of the musculoaponeurotic layers as early as 1 month postoperatively in patients who subsequently develop incisional hernias.

The presentation of incisional hernia depends on the site of the original wound, the size of the neck of the hernia, the size of the hernia, and the presence of complications. Small defects in the scar may result in large hernias, and this may predispose to incarceration and strangulation. Large hernias are unsightly and may give rise to abdominal discomfort (Fig. 1). Pressure necrosis and ulceration may occur in the skin overlying a large hernia.

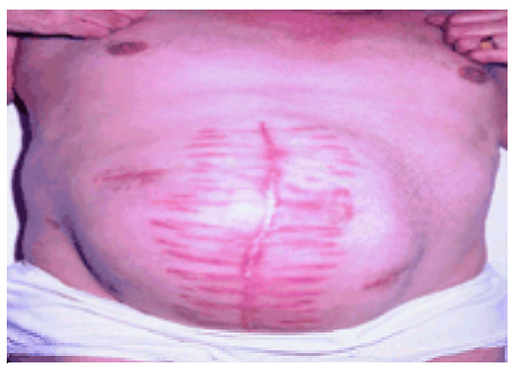


Fig. 1. Incisional hernia following laparotomy for peritonitis.

Management and repair

The majority of incisional hernias are asymptomatic, and the majority of symptomatic hernias may be managed conservatively. Obese patients benefit from weight reduction, since this reduces intra-abdominal pressure and may render a symptomatic hernia asymptomatic. Some patients find benefit from support by an elastic corset. In the fit patient, the indications for elective repair of incisional hernia are discomfort, enlargement, unacceptable appearance, or significant risk of strangulation. The presence of incarceration or strangulation is an indication for urgent or emergency surgery.

Unfortunately, factors that predispose to the development of incisional hernias are relative contraindications to repair. Weight reduction improves the chance of successful repair in obese patients. Stopping tobacco smoking and optimizing respiratory function reduce the chances of postoperative cough. In patients with large incisional hernias, the contents of the hernia may have lost the right of domain in the abdomen: return of the contents to the abdominal cavity may result in increased intra-abdominal pressure, splinting of the diaphragm, and a significant reduction in pulmonary reserve in patients with chronic respiratory disease. Recurrence of malignant disease, cachexia, ascites, renal failure, and hepatic failure are also contraindications to repair. The preoperative assessment of patients undergoing repair of incisional hernias is therefore important. Prophylactic administration of antibiotics is indicated to reduce the incidence of wound sepsis and the chance of failure of the repair, particularly if a non-absorbable mesh is implanted. In patients with massive incisional hernias, the use of a progressive pneumoperitoneum by nitrous oxide insufflation via a Veress needle or fine-bore indwelling cannula placed under local anaesthesia, for a period of 4 to 8 days before surgery, may expand the abdominal cavity and facilitate the repair.

Many different surgical approaches have been described for the repair of incisional hernias: no single technique is satisfactory for all hernias and surgical treatment has a high failure rate. In general, tension-free repairs using either prosthetic mesh or fascia have a much lower failure rate at around 10 per cent than the failure rate of 40 per cent for sutured repairs where the defect is brought together under tension. The types of repair may be divided into five basic categories.

1. The defect may be repaired in the same way as a laparotomy wound, with a mass closure with No. 1 nylon or polydioxanone.
2. The rectus sheath may be overlapped as in the ‘Mayo', double-breasted, ‘vest-over-pants' repair.
3. The defect may be repaired with a darn of nylon or fascia lata.
4. Lower midline incisions may be amenable to closure by swinging muscle over to close the defect.
5. Large defects may be repaired by implanting a non-absorbable mesh of tantalum, Marlex®, Mersilene®, or polytetrafluoroethylene.

The initial steps in the repair of an incisional hernia are the same, irrespective of the technique used. Generally, the original incision is reopened, and it is often necessary to excise an ellipse of redundant skin. The skin and subcutaneous tissues are dissected from the hernia sac back to the defect in the musculoaponeurotic layers and 3 to 4 cm of the fascia around the hernia are exposed.

Mass closure of an incisional hernia may be appropriate if the defect can be approximated without undue tension. The peritoneum is opened, adhesions to the undersurface of the scar lysed to clear 3 to 4 cm on the peritoneal surface and the hernia sac, and its covering of weak scar tissue is excised. The hernia is then closed with a mass closure of interrupted or continuous non-absorbable suture such as No. 1 nylon. The sutures are placed according to the principles of mass closure of a

primary wound. The repair may be reinforced by the addition of interrupted, far-and-near sutures or with an onlay graft of Marlex® mesh.

Opening of the peritoneum may result in postoperative ileus, and abdominal distension may compromise the security of the repair. In addition, dissection of the peritoneal and fibrous sac of the hernia may prove difficult and result in inadvertent enterotomy of underlying bowel. The ‘keel operation' is an extraperitoneal repair for midline incisional hernias that avoids opening the peritoneum, minimizes postoperative ileus, and allows early mobilization. The hernial sac and the neck of the hernia are cleaned of fibrofatty tissue and the hernia is inverted by the placement of interrupted mattress nylon sutures to close the defect in the musculoaponeurotic layer. The use of relaxing incisions in the anterior rectus sheath should be avoided because the repair may fail through them (Fig. 2).

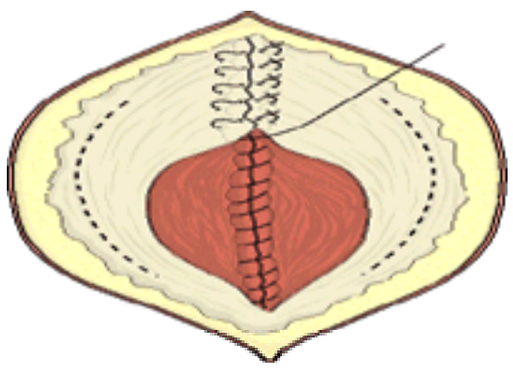


Fig. 2. The ‘keel operation' for extraperitoneal closure of an incisional hernia is performed by cleaning the edges of the hernia sac, which is then inverted by two layers of interrupted mattress sutures. Relaxing incisions may be made as shown, but herniation through these incisions may occur.

The Mayo ‘double-breasted' repair described for the repair of paraumbilical and epigastric hernias may be used to close small incisional hernias, especially if the direction of the original incision was transverse (for example, a hernia through an incision previously used for a transverse colostomy). After excising the hernial sac, the anterior rectus sheath is overlapped by a double layer of interrupted nylon mattress sutures so that the upper layer of the rectus sheath overlies the lower layer (Fig. 3).

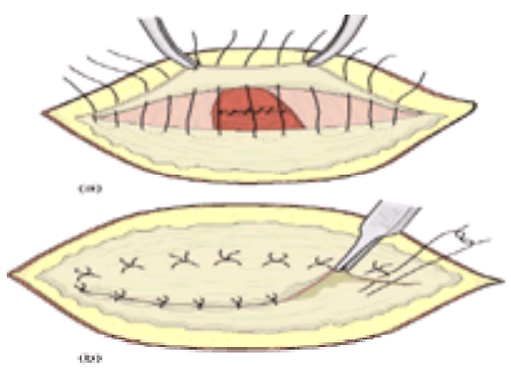


Fig. 3. The Mayo ‘double-breasted' repair closes a defect by overlapping the upper part of the defect over the lower edge.

Repairs with fascia lata and nylon darning have largely been replaced by the use of non-absorbable implants. However, midline incisional hernias may be repaired by incising the anterior rectus sheath 1 cm from the edge of the hernial defect on each side and suturing the medial edges of the anterior rectus sheath to invert the hernial sac. The repair is completed by placing a nylon darn between the lateral edges of the incision in the anterior rectus sheath.

Nuttall has described a classic repair of lower abdominal, midline incisional hernias by overlapping the rectus abdominis muscle on each side. The anterior rectus sheath is incised on each side of the defect, and the rectus abdominis detached from its insertion as close as possible to the pubic symphysis. Each rectus abdominis muscle is then reattached to the opposite side of the pubic tubercle with nylon sutures, and the overlap of muscle loosely sutured together. The operation is completed by the closure of the anterior rectus sheath with non-absorbable sutures.

Large defects should be repaired by implanting a mesh of non-absorbable material. Tantalum gauze has now been replaced with materials such as knitted monofilament polypropylene (Marlex®) and polytetrafluoroethylene. The patch is sutured to the edges of the defect in the musculoaponeurotic layers with a non-absorbable nylon suture. Both intra- and extraperitoneal placement have been described. Marlex® is best placed outside the peritoneal cavity as the mesh may cause adhesions and there are reports of erosion into the intestine and enterocutaneous fistula formation with intraperitoneal placement. Polytetrafluoroethylene excites little tissue reaction, which may be a disadvantage unless the patch is placed within the peritoneal cavity (Fig. 4). A closed suction drain is placed deep to the skin closure and left until drainage has ceased. Postoperative care is as for any patient undergoing abdominal surgery, except that the patient is taught to sit up in ways that do not place tension on the repair.

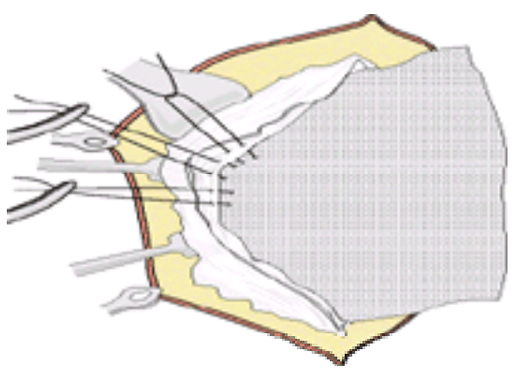


Fig. 4. Marlex® or Mersilene® mesh may be used to close a large defect. It is sewn in place with interrupted, non-absorbable sutures. Usually, the mesh is placed outside the peritoneum.

The rate of recurrence of incisional hernias following suture repair is between 25 and 44 per cent, and second repairs are equally unsuccessful. Many studies have compared the recurrence rate of hernias repaired by suture and by placement of mesh, and the results show on average a three-fold reduction in recurrence after mesh repair to under 10 per cent.

Parastomal hernia

Hernias develop alongside 5 to 10 per cent of colostomies and 3 to 10 per cent of ileostomies. Parastomal hernias are more common when the stoma is sited lateral to the rectus muscle than when the stoma is brought out through the rectus abdominis. The siting of a stoma in the incision used for laparotomy is associated with a very high incidence of incisional hernia. The extraperitoneal approach for the formation of a colostomy does not prevent the development of a parastomal hernia. A chronic cough, obesity, malnutrition, postoperative sepsis, and abdominal distension may predispose to the formation of a parastomal hernia.

Management and repair

The majority of parastomal hernias are asymptomatic, and only 10 to 20 per cent of patients require repair. Small hernias may be controlled by the use of a well-designed colostomy belt. The indications for surgical intervention are mainly related to difficulties in maintaining the stoma appliance. If the hernia is large, it may be difficult to apply a bag. The patient may be unable to see the stoma because of a large hernia (Fig. 5). Reduction of the hernia on lying down and its prolapse on standing may dislodge the appliance. The development of strangulation of a parastomal hernia is an absolute indication for repair. Some patients find large parastomal hernias cosmetically unacceptable. The contraindications to repair are the same as for repair of an incisional hernia.



Fig. 5. A large parastomal hernia causing difficulty in the application of a colostomy bag.

Two surgical options are availabe for the repair of a parastomal hernia. The stoma may be resited elsewhere and the original stoma site closed, repairing the hernia at the same time: this is appropriate for patients in whom the original stoma site is unsatisfactory. In some patients who have had multiple abdominal incisions, however, the choice of sites for a stoma may be limited. Under these circumstances, local repair of the hernia is indicated.

Local repair may be made via a peristomal incision. The defect in the musculature of the abdominal wall is then repaired with interrupted nylon sutures. An alternative is to approach the stoma via an incision at least 10 cm away, using the original laparotomy incision. The stoma is approached by subcutaneous dissection and repaired by the placement of a collar of Marlex® or other non-absorbable mesh. The technique has the advantage of avoiding a fresh surgical incision in the vicinity of the stoma, which may cause problems with the fitting of an appliance.

Further reading

Azurin DJ, Go LS, Arroyo LR, Kirkland ML. Trocar site herniation following laparoscopic cholecystectomy and the significance of preexisting umbilical hernia. *American Surgeon* 1995; **61:** 718–20.

Caldironi MW *et al.* Progressive pneumoperitoneum in the management of giant incisional hernias: a study of 41 patients. *British Journal of Surgery* 1990; **77**: 306–8.

Donaldson DR, Hegarty JH, Brennan TG, Buillou PJ, Finan PJ, Hall TJ. The lateral paramedian incision—experience with 850 cases. *British Journal of Surgery* 1982; **69**: 630–2.

Dudley HAF. Layered and mass closure of the abdominal wall: a theoretical and experimental analysis. *British Journal of Surgery* 1970; **57**: 664–7. [The science to support mass closure.]

Ellis H. Abdominal wall incisions and their healing. In Ellis H, Lamont P, eds. *Problems in general surgery. Vol.* 6 (2), *Wound healing*, pp. 249–61. Lippincott, Philadelphia, 1989. [An excellent overview of the causes of burst abdomen and incisional hernia based on a large personal series of 1129 patients.]

Gillon JF, Begin GF, Marecos C, Fourtanier G. Expanded polytetrafluoroethylene patches used in the intraperitoneal or extraperitoneal position for repair of incisional hernias of the anterolateral abdominal wall. *American Journal of Surgery* 1997; **174**: 16–19.

Goligher JC, Irvin TT, Johnston D, de Dombal FT, Hill GL, Horrocks JC. A controlled clinical trial of three methods of closure of laparotomy wounds. *British Journal of Surgery* 1975; **62**: 823–9.

Greenall MJ, Evans M, Pollock AV. Midline or transverse laparotomy? A random controlled clinical trial. Part 1: Influence on healing. *British Journal of Surgery* 1980; **67**: 188–90. [Direction of laparotomy incision is in the main irrelevant.]

Jenkins TPN. The burst abdominal wound: a mechanical approach. *British Journal of Surgery* 1976; **63**: 873–6. [Classic paper that proved the theory in H. A. F. Dudley's paper.]

Jones TE, Newell ET, Brubaker RE. The use of alloy steel wire in the closure of abdominal wounds. *Surgery Gynecology and Obstetrics* 1941; **72**: 1056–9. [Good description of the early history of mass closure.]

Maingot R. Umbilical and incisional hernia. In Maingot R, ed. *Abdominal operations*, p. 1618. Appleton-Century-Crofts, New York, 1980.

Norris JD. A review of wound healing and the mechanics of dehiscence. *Surgery* 1939; **5**: 775–86. [One of the first to recognize that wound dehiscence is preventable.]

Nuttall HCW. Rectus transplantation for midline incisional herniae. *British Journal of Surgery* 1937; **25**: 344–50. [Classic description of his repair.]

Pollock AV, Evans M. Early prediction of late incisional hernias. *British Journal of Surgery* 1989; **76**: 953–4. [The use of radioopaque markers to study the development of incisional hernia.]

Sjödahl R, Anderberg B, Bolin T. Parastomal hernia in relation to the site of the abdominal stoma. *British Journal of Surgery* 1988; **75**: 339–41.

Zrukowski ZH, Matheson NA. 'Button hole' incisional hernia: a late complication of abdominal wound closure with continuous non-absorbable sutures. *British Journal of Surgery* 1987; **74**: 824–5. [Description of 'buttonhole' incisional hernia.]

36.1 Acute vaginal bleeding

I. Z. Mackenzie

Introduction
Prepubertal bleeding
Bleeding during the menstrual years
Examination
Investigations
Management
Postmenopausal bleeding
Conclusion
Further reading

Introduction

Bleeding from the genital tract is of variable significance and its management depends upon the age of the patient, and the volume of blood that is being lost. Apart from bleeding as a complication of pregnancy, it is only rarely so heavy and life-threatening that urgent first-aid measures are required. Thus, for those women aged between 15 and 50 years, the possibility of pregnancy must always be considered before suggesting certain investigations and treatments.

Although many of the causes for bleeding will apply in all age groups—prepubertal, during the menstrual years, and postmenopausally—some are very uncommon at certain ages, are unlikely to result in excessive blood loss, and need not be considered in detail. However, management of individual causes is generally the same, irrespective of age, but precise treatments vary and need to be discussed separately.

Prepubertal bleeding

Serious vaginal bleeding in a young girl is most likely to be due to trauma—and usually accidental, such as after falling astride furniture. Sexual abuse, if very traumatic, may rupture the intact hymenal ring, and this should always be considered as a possibility, with appropriate investigation and follow-up. The source of bleeding will generally be obvious on inspection of the external genitalia, revealing a laceration of the labia or hymen, or a vulval haematoma. If blood is issuing through the introitus and showing no signs of ceasing, examination under anaesthetic is necessary to explore the vagina for possible penetrating injuries; adequate examination without anaesthetic will not be possible in most young girls, and should probably not be attempted. Examination under anaesthetic is best performed using a hysteroscope or cystoscope passed into the vagina in the very young girl, and in older children a paediatric laryngoscope may be used to view the cervix and upper vagina. Consideration should be given to ensuring that the bladder and rectum have not been damaged; if the anterior or posterior wall is lacerated, bladder or rectal examinations should also be performed. Packing the vagina with 1-inch (2.5 cm) ribbon gauze dampened with proflavine cream will control bleeding from superficial lacerations if it is not too heavy and the trauma not penetrating; an indwelling urinary catheter may be necessary until the pack is removed 24 h later. Suturing of any actively bleeding lacerations, using 2/0 catgut or other absorbable material, may be required. If a vulval haematoma has developed, incision and drainage with ligation of the bleeding point should be performed; occasionally, oozing continues and one or two deep 2/0 catgut sutures will usually control any persisting loss.

Rarely, the bleeding might come from a tumour of the lower genital tract. Adenosis vaginae, clear-cell adenocarcinoma of the vagina, and sarcoma botryoides of the vagina or cervix are possibilities. Biopsy specimens should be taken to obtain a histological diagnosis, and packing of the vagina should control active bleeding. If blood is coming through the cervix, this might be due to precocious puberty or the result of an oestrogen-secreting tumour of the ovaries. A bimanual pelvic examination with one finger introduced into the rectum should be performed to feel the size of the uterus and feel for the ovaries; the ovaries should not be palpable unless they are enlarged by tumour. If tumour is suspected, further investigation is necessary, with plasma gonadotropin and steroid assays and pelvic ultrasonography, followed by laparoscopy or laparotomy if the investigations suggest the suspicion to be correct. If heavy bleeding persists, curettage under anaesthesia, using a 3-mm aspiration curette and not requiring any cervical dilatation, is atraumatic; it must be remembered that the uterocervical canal in young girls is usually only 5 cm in length and care must be taken to avoid the use of excessive force, which might lead to uterine perforation. Oral progestogens, such as norethisterone (5 mg twice a day), should control any continuing bleeding within 24 h. Subsequent referral to a specialist is advised.

In a young child, a foreign body ‘lost' in the vagina should be considered, since this may result in infection with some bloodstained discharge. The vagina should be examined under anaesthetic as described above to retrieve the foreign body. Table 1 lists causes of vaginal bleeding found in 52 girls aged 10 years or younger referred to a specialist unit for investigation.

Unknown aetiology	25%
Precocious puberty	21%
Genital-tract tumours	21%
Vulval lesions	10%
Urethral prolapse	10%
Trauma	8%
Vulvovaginitis	6%

Table 1 Aetiology of vaginal bleeding in 52 girls aged 10 years or younger referred to a specialist unit (Hill et al. 1989)

Bleeding during the menstrual years

Acute vaginal bleeding may be either provoked or spontaneous, and may occur in non-pregnant women or in association with pregnancy. In the non-pregnant, unprovoked bleeding may occur at the time of expected menstruation, following a delay in menstruation, or between menstrual periods (intermenstrual bleeding). Provoked bleeding may follow coitus (postcoital bleeding), be associated with the insertion or removal of an intrauterine contraceptive device, or result from some ‘surgical' procedure involving the uterus or cervix, including criminal attacks upon a pregnancy to procure abortion. The relation of the bleeding to the menstrual cycle and the onset of the last menstrual period are thus of major importance in deciding which are the most probable differential diagnoses and the most appropriate management strategy. Table 2 lists possible causes and likely amounts of bleeding for the different types of bleeding that might be encountered in the non-pregnant woman.

	Prepubertal bleeding	Intermenstrual bleeding	Postcoital	Menorrhagia	Postmenopausal
Vaginitis	–	+	+	–	+
Vaginal foreign body	+	+	+	–	+
Vaginal malignancy	++++	+++	+++	–	++
Vulval trauma	++++	+++	–	–	++
Dermatitis vulvae	–	+	–	–	+
Cervical carcinoma	–	+	+++	–	–
Cervicitis	–	+	++	–	+
Cervical polyp	–	++++	+++	+++	+++
Cervical malignancy	++++	+++	+++	++++	++++
Endometrial polyp	–	++	–	++++	++++
Endometrial malignancy	–	++	–	++++	++++
Intrauterine contraceptive device	–	+	–	++++	–
Accidental or criminal	–	+++	–	++++	++++
Tubal malignancy	–	+	–	–	+
Functional ovarian tumour	++++	–	–	++	++++
Ovarian dysfunction	–	+	–	++++	+

–, No bleeding; +, a slight show; ++, light bleeding; +++, moderate bleeding; ++++, heavy bleeding requiring emergency attention.

Table 2 Causes and maximum amounts of abnormal bleeding from the genital tract likely to be experienced in the non-pregnant woman

Examination

General examination should be performed, with particular attention paid to possible pre-existing anaemia and cardiac decompensation. Lymphadenopathy, particularly in the supraclavicular fossae (Troisier's sign), might herald possible malignant disease; chest examination is important for the same reason. If pregnancy is a possibility, the breasts may be full and tender, with areolar pigmentation, prominent Montgomery's tubercles, and skin marbling from subdermal venous dilatation. Abdominal examination is necessary to detect any organomegaly, presence of pelvic tumours, and ascites.

Inspection of the vulva will give some indication of the rate of blood loss and the nature of the bleeding, whether fresh or old. Pooling of blood in the vagina with clot formation may conceal the rate of loss. Speculum examination will identify vaginal and cervical causes; examination during the acute phase will allow confirmation of the site of the bleeding, whether from a lesion in the lower genital tract or from the uterus. Bimanual examination should be performed to assess the consistency and surface texture of the cervix, and to note whether the internal os is closed. Cervical excitation should be determined, which will be positive in cases of ectopic tubal pregnancy. The position, size, shape, and consistency of the uterus should be assessed, and a note made of its mobility and tenderness. Other pelvic masses should be sought, most notably ovarian and tubal, and whether there is any associated tenderness.

Investigations

Following resuscitation that might be required to restore the circulating intravascular volume, some urgent investigations are appropriate. In all cases, haemoglobin, blood group, and rhesus type should be checked; in certain ethnic groups, a haemoglobin electrophoresis should be performed. Pregnancy should always be considered and, if it is a possibility from the history and clinical findings, it can be further confirmed by checking a urine or serum immunological test. The result will be positive 28 days after conception; with the specific tests using monoclonal antibody to b-human chorionic gonadotropin it will be positive within 14 days of conception; a negative result can be misleading.

Pelvic ultrasonography, ideally using a vaginal probe, may be particularly helpful in the very early weeks of pregnancy and in particularly obese women, although it may prove unhelpful or give confusing results unless the ultrasonographer is experienced in the examination of early pregnancy.

Management

Management depends initially upon the volume of blood lost, the clinical condition, and whether the patient is pregnant or not.

Non-pregnant patients

If a bleeding cervical polyp on a narrow stalk is seen, this can be avulsed with polyp forceps, grasping the polyp between the blades and twisting the forceps until the polyp comes away: the base will not bleed as a rule, but if it does, touching with a silver nitrate stick will be sufficient to control the loss. In patients in whom vaginal bleeding is heavy and a local lesion is not obvious or not amenable to ‘first-aid' treatment, examination under anaesthetic, combined with uterine curettage, should be performed.

Cervical dilatation and uterine curettage

If bleeding is not excessive, this procedure can be done with a paracervical nerve block, but if the loss is heavy, a general or regional anaesthetic is preferable. After asepsis and draping, the vulva is inspected and then a bimanual pelvic examination is performed to determine the size, position, consistency, shape, and mobility of the uterus. Palpation in the fornices is necessary to detect any adnexal masses. A Sim's or Auvard's speculum is passed to inspect the vaginal walls and the cervix. The anterior lip of the cervix is then stabilized with a volsellum forceps and the uterocervical canal sounded: in the normal non-pregnant woman, it is usually approx. 7.5 cm long. The cervix is then dilated with graduated dilators to a maximum of 7 mm, taking care not to insert the dilators with excessive force, so perforating the uterine fundus. If a broad-based cervical polyp is present, it should be avulsed by excision at the base and a haemostatic No. 1 catgut, or similar absorbable suture, on a trocar pointed needle should be inserted to close the defect.

The uterine cavity should then be explored. If the equipment is available, hysteroscopy using normal saline as a distending and flushing medium under pressure is probably the best way of examining the endometrial cavity. If there is heavy bleeding, insufflation with carbon dioxide is usually inadequate for inspection of the cavity. Hysteroscopy will allow the detection of pathology including endometrial polyps, submucous or intraluminal fibroids, and endometrial malignancies. If a polyp or pedunculated intraluminal fibroid is seen, this can usually be removed with polyp forceps. These forceps should be passed in the closed position to the uterine fundus and then opened, rotated through 180°, closed, rotated back through 180°, and withdrawn. This should be repeated until the polyp or fibroid has been seen within the jaws of the forceps and thus removed. Hysteroscopic examination should then be repeated to confirm that all the tissue has been removed. Systematic endometrial curettage should then be performed using a sharp curette passed to the fundus and withdrawn, applying pressure against the uterine wall. This procedure should be repeated until the anterior, posterior, and both lateral surfaces have been curetted. Finally, the fundus should be curetted with two or three strokes. If hysteroscopic examination is not available, the procedure for exploring the cavity with polyp forceps and sharp curette is followed, although small polyps may escape detection. All the tissue removed should be collected and sent in 10 per cent formalin solution for histological examination.

If the bleeding is coming from the cervix or vagina, and there are some inflammatory changes or ulcerated areas, these should be biopsied; a wedge of tissue from the suspect area is excised and one or two No. 1 chromic catgut sutures on a trocar-pointed needle will be required to provide haemostasis. The tissue specimen should be sent for urgent histological examination. If an intrauterine contraceptive device is in position and it is thought that it may be responsible for the bleeding, it should be removed with a pair of polyp forceps, used as described above for intrauterine examination: it is imperative to advise the patient that she is no longer protected against conception.

If a laceration of the vagina is encountered, which will probably have been anticipated from the history, it should be carefully examined to determine its position, depth, and extent. Trauma to the anterior wall could involve the bladder or urethra, and a careful exploration of both by cystoscopy may be appropriate; trauma to the posterior wall might involve the rectum, and a rectal examination should be performed to determine whether the mucosa has been damaged. In the event of either bladder or rectal mucosa being breached, a careful formal repair of the damage will be required, ideally by an expert to reduce the chances of fistula formation.

In most instances, acute bleeding will subside after the surgical procedure. In some cases of severe menorrhagia in which no pathology is found, the bleeding may be controlled by the oral administration of high doses of a progestagen such as norethisterone 10 mg 6 hourly until the loss ceases, and then continued at 10 mg 12 hourly for a total treatment time of 3 weeks. On discontinuing, a menstrual loss, which should not be heavy, will occur. Alternative treatment with other medications such as danazol or ethamsylate may be used. More recently, uterine tamponade using a distensible intrauterine balloon has been described as an emergency procedure; the balloon can be introduced without anaesthetic and left in place for 24 h and then removed, without the need for anaesthesia.

Rarely the bleeding is due to an ulcerated, prolapsed, pedunculated, intraluminal fibroid presenting at the cervix, the os being widely dilated to accommodate the tumour. It may be appropriate to remove the fibroid transvaginally by applying traction to it with tissue forceps to bring the pedicle into view. This can then be transfixed with No. 2 chromic catgut, or similar material, on a trocar-pointed needle, and the pedicle divided distal to the transfixion. If this is not possible, a hysterectomy may then be necessary.

Pregnant patients

Initial management depends upon the volume of bleeding, the size of the uterus, and the state of the cervix. Table 3 lists the likely nature of the bleeding that may be encountered in different situations. During the first trimester, an initial pelvic examination is essential to determine whether the cervix is closed and abortion threatened, or open and abortion inevitable. For the closed cervix, treatment involves bedrest with some light sedation, if necessary until the bleeding abates, when slow mobilization can begin. If the patient is rhesus (D) negative, anti-D immunoglobulin (250 iu) should be given by intramuscular injection. At 8 weeks' gestation or later, an ultrasound examination is helpful to demonstrate a live fetus and continuing pregnancy. Subsequent pregnancy management is essentially unchanged from the usual routine. Bleeding at later gestations is dealt with in Chapter 36.12.

Spontaneous abortion:	
Threatened	++
Inevitable	++++
Ectopic gestation	+
Attempted criminal abortion	++++
Molar pregnancy	++++
Lower genital-tract lesions (polyps etc.)	++

+, A slight show; ++, light bleeding; ++++, heavy bleeding requiring attention.

Table 3 Bleeding during early pregnancy

An inevitable abortion is diagnosed when vaginal bleeding is associated with menstrual-like cramps, which may be severe, and the internal cervical os is dilated sufficiently for the introduction of a finger. If products of conception are felt in the os, they should be removed with sterile sponge-holding forceps, especially if the patient is in marked pain or is hypotensive with a bradycardia due to vagal stimulation; this is referred to ‘cervical shock'. If bleeding is excessive, ergometrine (0.5 mg intramuscularly) should be given to make the uterus contract and to reduce bleeding. All rhesus (D)-negative women should be given 250 iu anti-D immunoglobulin within 60 h of the start of bleeding. Since many spontaneous abortions between 8 and 13 weeks' gestation are ‘incomplete', with pieces of placental tissue retained within the uterine cavity, surgical evacuation is often advised. Ultrasound examination of the uterus may be used to assess the completeness of the abortion, but the results are often misleading.

Surgical evacuation of retained products of conception

Adequate anaesthesia, such as a paracervical block, regional block, or induction anaesthesia, and strict aseptic precautions are essential. After preparing the skin and vaginal, and draping, urinary voiding having been previously advised, the size and position of the uterus are determined on bimanual examination. Ergometrine, 0.5 mg given intramuscularly before the start of the surgical procedure, should ensure a contracted, firm-walled uterus, reducing bleeding and the chance of uterine perforation. A uterine sound should not be passed at the start of the procedure for fear of perforating the soft pregnant uterine wall. Cervical dilatation should not be necessary and, in some instances, the index finger can be passed through the dilated cervix into the uterine cavity and the retained products felt, dislodged, and expelled through the open cervix. Subsequent retrieval of any remaining pieces of placental tissue is achieved by stabilizing the cervix and uterus with a sponge-holding forceps, grasping the anterior cervical lip, and introducing a second sponge-holding forceps through the cervix to the uterine fundus, where it should be opened, rotated through 180°, closed, and withdrawn. Systematic curettage of the entire cavity with a blunt or flushing curette will ensure that all tissue is removed. If the tissue removed from the uterus has the typical appearance of ‘grape-like vesicles', it should be sent for histological examination to identify any trophoblastic tumour.

On completion, there should be minimal bleeding and, once recovery from anaesthesia is complete and observations are satisfactory, the patient can be discharged.

If a criminal abortion attempt is suspected, broad-spectrum antibiotic cover should be given after bacteriological specimens have been collected. A check for consumptive coagulopathy should be made in cases of Gram-negative septicaemia. Evacuation of the uterus should be performed, as previously described, as soon as possible. Trauma to the vaginal wall and uterus should be searched for carefully, and, if either exists, close observations must be instituted to watch for the development of peritonitis or evidence of damage to bowel or bladder. Postoperative recovery observations must also include the monitoring of hepatic and renal function for evidence of impending failure during the first 48 h after admission.

In the event of hydatiform molar change being identified on histological examination of the evacuated products, the patient should be recalled 2 weeks after the initial evacuation and, if heavy bleeding is persisting, a further curettage should be performed, and any tissue recovered sent for histological examination. At the same time, a urine or serum sample should be collected and sent for assay of b-human chorionic gonadotropin. Cases should be notified to the appropriate trophoblastic tumour registry. If the concentrations of gonadotropin are maintained or rising when retested 2 weeks later, referral to a specialist centre for further investigation and possible chemotherapy should be arranged. If they are falling, repeated checks should be instituted at 1- to 2-monthly intervals for the subsequent 24 months, to ensure that they remain at non-pregnant values; further pregnancy should be prevented for this period, but the combined oral contraceptive pill is best avoided since recurrent disease rates may be increased. If the concentration of b-human chorionic gonadotropin starts to rise, a new pregnancy needs to be excluded, and referral to the specialist centre will be required.

Patients with ectopic pregnancies frequently faint and may notice a small volume of vaginal blood loss or ‘prune juice' discharge, but only rarely experience heavy vaginal bleeding. Ectopic pregnancy may occur in any women in whom pregnancy is a possibility, and the diagnosis must be considered in the presence of vaginal bleeding and lower abdominal pain, even without a history of amenorrhoea.

Postmenopausal bleeding

This is defined as bleeding from the genital tract 12 months or more after the last menstrual period in a woman aged 45 years or older. Its causes are essentially the same as those described for intermenstrual bleeding and menorrhagia, with a greater emphasis on cervical and endometrial malignancy (see Table 2). Oestrogen-secreting ovarian tumours, such as granulosa-cell and theca-cell tumours, most frequently present as heavy postmenopausal bleeding. Additionally, the possibility that endogenous steroids have been given to relieve menopausal symptoms should be considered and specific enquiry made.

All cases of postmenopausal bleeding require investigation irrespective of the volume of blood loss. Vaginal ultrasonography, outpatient and inpatient hysteroscopy, and uterine curettage are used for this purpose. Although only urgent if bleeding is excessive, the need to investigate is important since endometrial pathology is present in 20 per cent of cases and endometrial carcinoma in 10 per cent.

Conclusion

Vaginal bleeding can be very heavy and life-threatening. It is important in the first instance to decide whether the patient could be pregnant. The management required to control the bleeding may then be urgent, especially if the patient is pregnant. For the pregnant woman, management depends upon the prognosis for the pregnancy, whether a threatened abortion or an inevitable loss. For the non-pregnant, the identification of the cause for the bleeding is necessary to enable appropriate treatment to be provided.

Further reading

Bagshawe KD, Lawler SD. Hydatidiform mole and choriocarcinoma. In: Turnbull AC, Chamberlain GV, ed. *Obstetrics*, pp. 453–68. Churchill Livingstone, Edinburgh, 1989.

Dewhurst J. *Practical paediatric and adolescent gynaecology*. Marcel Dekker, New York, 1980.

Hill NCW, Oppenheimer LW, Morton KE. The aetiology of vaginal bleeding in children. A 20 year review. *British Journal of Obstetrics and Gynaecology* 1989; **96**: 467–70.

MacKenzie IZ, Bibby JG. Critical assessment of dilatation and curettage in 1029 women. *Lancet* 1978; **ii**: 566–8.

Neuwirth RS. Some new applications for hysteroscopy. *Contemporary Obstetrics and Gynecology* 1987; **21** (Spec. Iss.): 1–9.

36.2 The management of acute pelvic pain

Michael D. G. Gillmer and Bruce C. Dunphy

Introduction
History
Examination
Investigation
Laparoscopy
Treatment
Further reading

Introduction

The gynaecological causes of acute pelvic pain include haemoperitoneum, infection, abnormal pregnancy, and vascular complications.

Haemoperitoneum may be due to retrograde menstruation, rupture of functional or neoplastic ovarian cysts, or an ectopic pregnancy. It may also follow uterine perforation during the insertion of an intrauterine device or an operation on the uterus, such as dilatation and curettage, or hysteroscopy.

Salpingitis due to sexually transmitted disease is the most common infective gynaecological cause of pelvic pain. Other causes include septic abortion or rupture of a pelvic abscess. In postmenopausal women, pyometra, usually associated with endometrial malignancy, may rarely cause acute pain. Important non-gynaecological infective causes of acute pelvic pain include appendicitis and diverticulitis. Occasionally in women with cystitis, pyelonephritis, or ureteric colic, acute pelvic pain may be the primary complaint; this last diagnosis is especially likely in women with a pelvic kidney. Other less common non-gynaecological causes of acute pelvic pain include regional ileitis and lower-bowel obstruction, due to malignancy, mesenteric occlusion, or pelvic vein thrombophlebitis. Orthopaedic causes include sacroiliac strain, referred pain from a prolapsed intervertebral disc, or degenerative changes in the lumbar spine.

In addition, several complications of early pregnancy may present with acute pelvic pain. These include inevitable abortion, ectopic pregnancy, and rupture of a corpus luteum cyst. In later pregnancy, red degeneration of a fibroid or placental abruption may also cause acute pelvic pain.

Torsion of an ovarian cyst or, more rarely, a pedunculated subserous fibroid is also a cause of acute pelvic pain.

The most important differential diagnoses are acute salpingitis, an ovarian cyst accident, or a tubal ectopic pregnancy, and the most prominent question in the mind of any clinician managing a woman in the reproductive age-group should be ‘does this woman have an ectopic pregnancy?' Ectopic pregnancy remains an important cause of maternal death and, although this follows sudden collapse in approximately a third of cases, the remaining deaths are potentially avoidable.

History

The classic presentation of an ectopic pregnancy depends on its site of implantation in the fallopian tube. There is usually amenorrhoea of 6 to 8 weeks' duration, and longer if the pregnancy is situated in the ampullary end of the tube. Pregnancy symptoms are variable and may, or may not, be elicited. The pain, which is due to intraperitoneal bleeding, is usually unilateral and present in the lower abdomen and pelvis. If there is significant intraperitoneal bleeding, the blood may track up under the diaphragm and cause referred shoulder-tip pain. Vaginal bleeding tends to be scanty and dark brown in colour, and usually begins a few hours after the onset of the pain. Pregnancy symptoms, such as nausea and breast tenderness, may have been present but frequently cease shortly before the onset of pain and bleeding. Factors that predispose to tubal ectopic pregnancy include intrauterine contraceptive devices, the progestogen-only pill, tubal damage following salpingitis or appendicitis, and embryo transfer following *in vitro* fertilization. As with acute appendicitis, the presentation is varied and, when suspected, cannot be excluded on the basis of history alone. An ectopic pregnancy can be associated with almost any pattern of bleeding, and the pain may not be unilateral.

Pelvic inflammatory disease, due to salpingitis, is usually but not always bilateral, and may be associated with fever, rigors, and a purulent vaginal discharge. There may also be a history of abnormal vaginal bleeding. The pain tends to be an increasingly severe constant ache, but may become colicky if a pyosalpinx develops causing significant tubal distension. Vomiting tends, paradoxically, to be more common than in women with tubal pregnancies. Predisposing factors include a history of previous pelvic infection, an intrauterine contraceptive device, or a recent change of sexual partner.

Pain associated with torsion of an ovarian tumour, normal adnexum, or a pedunculated fibroid, is usually of sudden onset, confined to one side, and colicky in nature. Vomiting is also common.

Examination

Lower abdominal guarding and rebound tenderness indicate significant pelvic pathology, but physical findings in women with acute pelvic pain may be misleading. Unilateral signs usually indicate an ectopic pregnancy, ovarian accident, or extragenital pathology. Bilateral signs commonly indicate pelvic inflammatory disease but do not exclude any of the above diagnoses. The uterus is usually slightly enlarged and softened with an ectopic pregnancy, and this may cause confusion with an early intrauterine pregnancy. The cervix, however, dilates during abortion and the vaginal blood loss is generally heavier. lthough pain on movement of the cervix, so-called ‘cervical excitation' pain, is typical of a ‘leaking' tubal ectopic pregnancy, it will also be observed whenever there is pus or blood in the pelvis, causing peritonism.

The classic acute presentation of a ruptured ectopic pregnancy involves shock, signs of a significant haemoperitoneum, and profound anaemia. In this situation a pelvic examination is contraindicated as it may cause further, ‘life-threatening' intra-abdominal haemorrhage. On the other hand, a patient with a chronic ‘leaking' ectopic pregnancy may occasionally present with severe anaemia, a stable pulse and blood pressure, acute pain, and a chronic pelvic haematocele. Tubal ectopic pregnancies are seldom palpable unless they are of advanced gestation, situated in the ampullary portion of the tube, or are associated with a pelvic haematocele.

The presence of a pelvic mass separate from the uterus usually indicates an ovarian tumour, a pyosalpinx, a tubo-ovarian mass, or occasionally a subserous fibroid. Although fibroids rarely cause acute pelvic pain, pedunculated fibroids may undergo torsion and present a clinical picture of acute colicky pain and vomiting, indistinguishable from that due to torsion of an ovarian tumour. Red degeneration, due to ischaemic necrosis of a fibroid, causes a constant severe ache. It is usually associated with pregnancy but it may also occur in perimenopausal women.

Copious purulent vaginal discharge, visible on speculum examination, suggests a sexually transmitted infection, or possibly a septic inevitable or incomplete abortion. Vaginitis and cervicitis may, however, also cause a profuse infected vaginal discharge and this sign cannot therefore be considered to be indicative of pelvic inflammatory disease.

A fever in excess of 38°C in a woman with acute pelvic pain is usually due to salpingitis, septic abortion, appendicitis, a urinary tract infection, or some other extragenital infection. Lesser degrees of fever may, however, occur with any of the pelvic pathologies.

Investigation

Urinary b-human chorionic gonadotrophin (**b-hCG**) tests are inexpensive and should be the first line of investigation in sexually active women with acute pelvic pain. Modern tests have a detection limit of 25 to 50 iu/l with false-negative rates of less than 2 per cent. A positive result with this very sensitive technique indicates an absolute need to determine whether the pregnancy is intrauterine or extrauterine; a negative result does not, however, exclude an ectopic pregnancy.

Urine must also be examined microscopically and cultured. It should, however, be remembered that white cells may appear in urine in association with an ovarian accident or appendicitis. In addition, the presence of pyuria does not exclude an ectopic pregnancy. Cervical and high vaginal swabs should be cultured, looking particularly for *Chlamydia* spp. and *Neisseria gonorrhoeae*. If *Chlamydia* culture is not available, then tests for *Chlamydia* antigen should be used. If the patient has a pyrexia of more than 38°C, blood cultures should be obtained.

Although a full blood count and erythrocyte sedimentation rate should be performed, these are of limited value as the haemoglobin concentration may not be reduced in women with an ectopic pregnancy, except where there is massive intraperitoneal bleeding or a pelvic haematocele. In addition, the finding of an elevated white cell

count or erythrocyte sedimentation rate is non-specific, as these may be elevated in any of the conditions that cause acute pelvic pain. The erythrocyte sedimentation rate may, however, be useful in monitoring the response of pelvic inflammatory disease to antibiotic therapy. Elevated serum concentrations of antibodies to genital tract pathogens such as *Chlamydia* do not necessarily indicate acute infection and are of minimal diagnostic value in this clinical situation.

Abdominal and vaginal ultrasound scanning is of undoubted value in diagnosing pelvic pathology. Intrauterine pregnancy can be diagnosed before 7 weeks using a vaginal probe, and thereafter with an abdominal transducer. Ultrasound techniques are, however, of limited value for the diagnosis of tubal ectopic pregnancy and, although reports of ‘cystic structures' in the adnexal regions or evidence of free fluid in the pouch of Douglas may arouse suspicion, a diagnosis of a tubal ectopic pregnancy cannot be made with certainty unless a fetal heart can be seen beating outside the uterine cavity. In addition, the absence of ultrasound evidence of an intrauterine pregnancy in a woman with a positive pregnancy test does not necessarily indicate that the pregnancy is ectopic, as very small amounts of placental tissue in women with a missed or incomplete abortion can produce a positive urinary b-hCG pregnancy test. Furthermore, it must be emphasized that although identification of an intrauterine pregnancy greatly reduces the likelihood of an ectopic pregnancy, it does not exclude this diagnosis. Ultrasound scanning of the pelvis and lower abdomen, by experts using high-resolution equipment, is, however, extremely valuable for the diagnosis of ovarian cysts, fibroids, and missed abortion, and for identifying viable intrauterine pregnancies.

Whenever there is any doubt, a laparoscopy must be performed in order to make a definitive diagnosis.

Laparoscopy

The decision to perform a laparoscopy is dictated by the clinical situation and the speed with which the above investigations can be performed. When there is an obvious haemoperitoneum or the patient is shocked, an urgent laparotomy is indicated, as laparoscopy could be hazardous for a patient in this condition and could delay the arrest of bleeding. On the other hand, if the woman's condition is stable and technical difficulties are anticipated because of obesity, or a history of gross pelvic or abdominal adhesions, a conservative course may be preferable. If the diagnosis is in doubt and a laparoscopy is either contraindicated or laparoscopic equipment is not available, then a laparotomy, or possibly a minilaparotomy using a Cusco speculum, should be performed. It must be emphasized that an examination under anaesthesia will not exclude an ectopic pregnancy and has no part in the modern management of this condition. In addition, provided that appropriate operative facilities are available, there can be no justification for performing a culdocentesis.

A careful and methodical approach should be used when assessing the pelvis and abdominal cavity laparoscopically. The anterior and posterior surfaces of the uterus should be inspected and both fallopian tubes visualized throughout their length. Minor degrees of inflammation of the fallopian tubes may prove difficult to distinguish from the hyperaemia induced by the carbon dioxide used to insufflate the abdominal cavity. The ovaries should be lifted up to enable the whole surface of each to be inspected and the pouch of Douglas should be examined carefully to exclude endometriosis or adhesions. In addition, the presence of any free fluid or blood should be noted. If there is evidence of infection, fluid should be aspirated and swabs taken from the fallopian tubes for bacteriological examination. The parietal peritoneum should be examined for adhesions and, whenever possible, the appendix should be identified and examined throughout its length. The gallbladder should also be inspected to exclude cholecystitis and the perihepatic area should be examined to exclude adhesions due to the Fitz-Hugh and Curtis syndrome (chlamydial perihepatitis). It is vital that the findings are documented accurately for future reference.

When a laparoscopy is performed in women with acute pelvic pain, approximately 35 per cent will have acute or chronic pelvic inflammatory disease; 20 per cent will have an ectopic pregnancy; 15 per cent, an ovarian cyst; 5 per cent, adhesions not associated with pelvic inflammatory disease; and 2 per cent, appendicitis. In the remaining cases, no cause will be identified.

Treatment

When a tubal ectopic pregnancy is diagnosed and active bleeding is suspected, the primary concern is to achieve haemostasis. A laparotomy should therefore be performed, without delay, through a low transverse suprapubic incision, and haemostasis achieved as rapidly as possible by applying an artery clamp on either side of the ectopic pregnancy. The affected tube should then either be removed completely or partially, depending on the site of the ectopic pregnancy and the extent of the tubal damage. There is, however, no correct way to proceed once the ectopic pregnancy has been removed. Ideally, a minimum amount of tube should be removed and the cut ends ligated so as to allow a tubal reanastomosis at a later date. It is best to resist the temptation to perform an adhesolysis or further tubal surgery at the time of the ectopic pregnancy, as conditions at the time of the acute episode mitigate against a successful outcome. Although traditional treatment of a tubal ectopic pregnancy involves a laparotomy and either a partial or total salpingectomy, or occasionally a salpingo-oophorectomy, more conservative forms of treatment which aim to conserve tubal function are nowadays considered to be preferable if the ectopic pregnancy is still intact. These include ‘milking' the ectopic pregnancy out of the ampullary end of the tube when possible or, if the ectopic is in the midportion of the tube, performing a ‘linear salpingostomy', in which the tubal pregnancy is removed through a vertical incision on the antimesenteric border of the fallopian tube overlying the site of the ectopic. Where the necessary equipment and expertise are available, this latter procedure can also be performed under laparoscopic control, thus avoiding the need for a laparotomy and resulting in a more rapid postoperative recovery. Whether the procedure has been performed endoscopically or by laparotomy, it is advisable, at the conclusion, to wash out the peritoneal cavity with warm isotonic saline, thus minimizing the likelihood of postoperative ileus and adhesion formation. Prophylactic antibiotics should be considered postoperatively.

Medical treatment of unruptured ectopic pregnancy using methotrexate injected intramuscularly or directly into the ectopic has been advocated in recent years. This requires serial monitoring of serum b-hCG concentrations following injection to ensure complete resolution of the ectopic gestation. Additional oral administration of the antiprogesterone mefipristone has been shown to enhance the efficacy of methotrexate therapy.

Acute or chronic pelvic infection should be treated initially with intravenous antibiotics, even if there is no evidence of a bacteraemia or septicaemia. A combination of metronidazole, a third-generation cephalosporin, and doxycycline should be used initially, pending the bacteriological results obtained from the cervical, vaginal, and fallopian tube swabs, peritoneal fluid, and blood cultures, regardless of the suspected aetiology. Subsequent treatment should be based on the antibiotic sensitivities, but must include an antichlamydial agent, such as tetracycline or erythromycin. Ideally, antibiotic therapy should be continued for at least a month, rotating through two or three different combinations of antibiotics. If a tubo-ovarian mass or abscess is observed during laparoscopy, this is not an indication for surgery. A laparotomy should only be performed if the patient has septic shock or if the mass fails to reduce in size with effective antibiotic therapy. If a pelvic abscess is situated in the pouch of Douglas, it may be more appropriate to drain this by making an incision through the posterior vaginal fornix rather than performing a laparotomy. The patient's sexual partner should be advised to attend a genito-urinary medicine clinic for screening and appropriate treatment.

Further reading

Rocker I. Gynecological pain. In: Rocker I, ed., *Pelvic pain in women—diagnosis and management*, pp. 103–31. Springer-Verlag, Berlin, 1990. [A comprehensive analysis of the aetiology and clinical features of gynaecological pelvic pain.]

Stabile I, Grudzinskas JG. Ectopic pregnancy: what's new? *Progress in Obstetrics and Gynaecology* 1994; **11**: 281–309. [A review of the modern management of ectopic pregnancy, including medical and endoscopic surgical treatments.]

36.3 Acute pelvic sepsis and tubo-ovarian abscess

Robert G. Forman

Acute pelvic sepsis
Introduction
Clinical features
Management of pelvic sepsis
Tubo-ovarian abscess
Treatment
Medical therapy
Drainage of tubo-ovarian abscesses
Conservative or radical surgery for tubo-ovarian abscesses
Rupture of a tubo-ovarian abscess
Conclusion
Further reading

Acute pelvic sepsis

Introduction

Acute pelvic sepsis, sometimes referred to as acute pelvic inflammatory disease, is one of the most frequent causes of morbidity in previously healthy young women. Up to 30 per cent of gynaecological admissions in some developing countries are attributable to this disease. The infection is located in the upper genital tract and typically involves the endometrium and both fallopian tubes. In the milder forms of infection oophoritis is absent. Isolated vaginitis, commonly seen in association with local vaginal pathogens such as *Candida albicans* and *Trichomonas vaginalis*, is a distinct clinical entity from acute pelvic sepsis.

Ninety-nine per cent of all cases of acute pelvic sepsis are due to ascending infection. The remaining 1 per cent is caused by local spread from other pelvic structures, most frequently the appendix. The ascending infection arises in association with sexually transmitted disease. Until the late 1970s *Neisseria gonorrhoeae* was the organism classically associated with acute pelvic sepsis, and this may still be the case in some developing countries. At the beginning of the 1980s it became apparent, initially in Europe and later in the United States, that non-gonococcal infection was increasing in incidence. Currently it is believed that up to 50 per cent of all cases of acute pelvic sepsis are caused by *Chlamydia trachomatis*, whereas gonoccocal infection only accounts for approximately 10 per cent.

Several factors predispose to an increased risk of pelvic sepsis. These include age of first sexual activity, number of sexual partners, and the presence of an intrauterine contraceptive device. Although the last is controversial, there is a well-documented seven- to ninefold increased risk of pelvic sepsis in nulligravid women who use this form of contraception. Therapeutic pregnancy termination is associated with a 0.5 per cent incidence of acute pelvic sepsis within 3 weeks of surgery. The incidence in criminally induced abortions is much higher. There is also evidence linking the presence of bacterial vaginosis and the use of vaginal douching with pelvic inflammatory disease.

The importance of pelvic sepsis lies not only with the morbidity associated with acute stages of the disease but also with the serious nature of the chronic sequelae which may result. Damage to the endosalpinx and pelvic adhesion formation leads to chronic pelvic pain, deep dyspareunia, and an increased likelihood of reproductive disorders. The incidence of infertility increases from 10 per cent after a single episode of pelvic sepsis to 50 per cent after three documented infections. Ectopic pregnancy occurs in 1 in 16 women in their first pregnancy following pelvic sepsis, which is a 10-fold increase compared with a control population.

Clinical features

Typically, the patient with acute pelvic sepsis presents with fever and systemic signs of toxic illness. There is marked bilateral tenderness in both iliac fossas, usually associated with signs of peritonism. Speculum examination of the vagina discloses a purulent discharge from the external cervical os. Cervical excitation pain may be elicited. The uterus and both adnexas are tender. In the presence of a pyosalpinx or tubo-ovarian abscess, a mass may be palpable, although the degree of tenderness may prevent accurate assessment via the vaginal fornices.

Not all cases of acute pelvic sepsis present in such a florid way, and the disease is subject to both under- and overdiagnosis. In its milder forms it may be mistaken for appendicitis, gastrointestinal disturbance, dysmenorrhoea, or urinary tract infection, and may be inadequately treated. Conversely, a low-grade pelvic sepsis is not uncommonly an inappropriate diagnosis of exclusion. It is argued that the diagnosis can only be made at laparoscopy, at which time bacteriological swabs can be obtained from the endosalpinx and peritoneal fluid. While this is a counsel of perfection, diagnostic laparoscopy has not yet been widely accepted for this indication.

Management of pelvic sepsis

On admission to hospital, a full blood count should be obtained as a baseline. Leucocytosis is common and can be useful in monitoring the response to therapy. Additional supportive evidence of inflammation can be obtained by an elevation of erythrocyte sedimentation rate or C-reactive protein. Urea and electrolytes should be assessed if the patient is toxic or dehydrated. Blood cultures are essential in the presence of pyrexia. Bacteriological swabs are taken from the urethra and cervix. If possible, these should be Gram-stained immediately. If facilities are available, swabs from the cervical endothelium should be assessed for *Chlamydia trachomatis*.

Any constitutional disturbance should be corrected by rehydration, and antibiotic therapy must be commenced immediately. The antibiotics chosen should be active against chlamydia and gonococcus as well as against coliforms and anaerobic species. The last are probably not causative agents but are opportunistic organisms. It is difficult to treat acute pelvic sepsis adequately using single-agent therapy. Treatment is always commenced as soon as bacteriological specimens have been obtained and prior to the results of culture being available. Chlamydia and gonococcus are not easy to isolate, so effective therapy should not be reduced even when culture results are known, although further specific therapy can be added if indicated. There are currently no data comparing parenteral with oral therapy although most clinicians empirically favour the parenteral route until signs of clinical improvement are observed. Examples of parenteral regimens suggested by the Centers for Disease Control and Prevention in the United States include clindamycin and gentamicin or cefoxitin and doxycyline (can be given orally). Oral therapies include ofloxacin and metronidazole for 14 days.

As there is such a high incidence of sexual transmission, it is vital that male partners are seen and treated with an effective antibiotic regimen such as outlined above. This is often overlooked, particularly as male partners may well be asymptomatic.

Tubo-ovarian abscess

The other important sequela of acute pelvic sepsis, and perhaps the most likely to present to a general surgeon, is tubo-ovarian abscess. Many so-called tubo-ovarian abscesses are not abscesses at all, as they do not involve the ovarian stroma but are a conglutination of tube and ovary to surrounding structures. These are more correctly called tubo-ovarian complexes. In practice, tubo-ovarian abscesses, tubo-ovarian complexes, pyosalpinx, and ovarian abscesses are difficult to distinguish clinically and the management is the same.

Tubo-ovarian abscesses arise secondary to damage to the endosalpinx provoked by the inflammatory process. Purulent material exudes from the tubal ostium into the pouch of Douglas. The ovary can become involved if the infected material gains access via a corpus luteum. In other cases, a tubo-ovarian complex forms due to contiguous spread from the tube to the bowel, bladder, and the contralateral adnexa.

The presentation of tubo-ovarian abscess is similar to that of an episode of acute pelvic sepsis. Ninety per cent of patients complain of abdominal pain. Fever and leucocytosis are usually present, although in 20 per cent these features are absent. Up to two-thirds of patients give no prior history of pelvic sepsis. There is evidence, at least in adolescents, that patients with tubo-ovarian abscess have a less severe constitutional disturbance than those with uncomplicated acute pelvic inflammatory disease. The differential diagnosis from uncomplicated acute pelvic sepsis is made on the basis of an adnexal mass. However, transvaginal ultrasound indicates that the mass may be clinically undetectable in approximately 40 per cent of patients.

Diagnosis is confirmed by ultrasound, using either transabdominal or, preferably, transvaginal routes. Both radionucleotide scanning and computed axial tomography

have also been proposed, although their use is not widespread. Recent studies suggest magnetic resonance imaging has great potential in establishing the diagnosis of tubo-ovarian abscess.

Treatment

Until recently, the treatment of tubo-ovarian abscesses was aggressive, prompted by fears of rupture. The mortality from a ruptured tubo-ovarian abscess in the early literature exceeded 50 per cent. The principle behind modern management of tubo-ovarian abscesses should still be to avoid rupture, although more thought should now be given to preservation of ovarian and reproductive function than has previously been the case. Therapeutic options include medical treatment, surgical drainage, and conservative and radical surgical excision.

Medical therapy

It is now believed that medical therapy should be the first line of treatment in acute pelvic sepsis, including that complicated by tubo-ovarian abscess. Anaerobes are isolated from tubo-ovarian abscesses in over two-thirds of cases, and treatment should be directed towards them. The poor results of medical treatment in the earlier literature reflected the lack of suitable antimicrobial therapy active against anaerobic organisms. Treatment includes intravenous antibiotics and fluids, nasogastric suction, and blood transfusion, when necessary. The response to therapy is measured by symptomatic improvement, reduction in fever and leucocytosis, and a decrease in the size of the mass. In a study of 160 cases treated in this way, 69 per cent of patients had a good initial response to medication and the remainder required early surgery. Nearly one-third of those who responded initially to medical therapy subsequently underwent a surgical procedure. Patients responding to medical therapy tended to be younger nulliparous women with unilateral adnexal involvement. In another review of 856 patients with medically treated tubo-ovarian abscesses, 67 per cent responded well initially, and the pregnancy rate in the 295 patients who wished to conceive and were followed up was 10 per cent.

It seems reasonable to conclude from these results that medical therapy is appropriate for an unruptured tubo-ovarian abscess (with no associated surgical disorders), but if no response is seen after 48 h, surgery is indicated.

Drainage of tubo-ovarian abscesses

Tubo-ovarian abscesses can be drained in several ways, and the literature abounds with suggestions. A frequently adopted approach is a posterior colpotomy, which will be discussed in more detail below. Transabdominal and extraperitoneal routes have also been used. Laparoscopic drainage has been advocated and, in skilled hands, no difference in patient outcome is observed compared with drainage at laparotomy. Considerable operator expertise is required in elective operative laparoscopy prior to attempting this procedure in the context of suppurative lesions of the pelvis.

Pelvic abscesses, including tubo-ovarian abscesses, have been drained using a transgluteal approach through the greater sciatic foramen. Catheter placement was guided by computed axial tomography. Surgery was avoided in over 80 per cent of patients, but it is unlikely that this method will be adopted except in specialist centres. Another technique, which is potentially of more widespread interest, involves vaginal drainage under transvaginal ultrasound control. In one study, 10 patients with tubo-ovarian abscess, demonstrated by ultrasound, underwent transvaginal aspiration under light sedation. Between 10 and 120 ml of pus were aspirated, and accurate bacteriological diagnosis was obtained in six cases. Another randomized study demonstrated that patients treated by antibiotics combined with transvaginal ultrasound aspiration had a more favourable outcome than those treated by antibiotics alone.

The traditional method of draining tubo-ovarian abscesses has been posterior colpotomy. The technique involves exposing the posterior vaginal fornix by upwards traction on the anterior lip of the cervix. The junction of the cervix and pouch of Douglas is then transversely incised. The incision is enlarged by opening a clamp in the wound and pus is drained. All loculations should be broken down digitally. Bacteriological specimens are obtained and a drain is left *in situ.* Corrugated drains are preferable to tubal drains as the latter may block. A review of 384 colpotomy drainages described 23 patients (6 per cent) who developed peritonitis, with six deaths. Subsequent surgery was often necessary in the other patients.

Posterior colpotomy is performed less often than previously. It is now recognized that it is only safe when the abscess is midline, fluctuant, and dissecting the rectovaginal septum. However, in situations where more advanced surgical facilities are not readily available, or where there are patient contraindications such as obesity, this can be a useful technique.

Conservative or radical surgery for tubo-ovarian abscesses

On the assumption that all pelvic organs harbour infective microfoci, hysterectomy and bilateral salpingo-oophorectomy was, until recently, the standard surgical procedure for tubo-ovarian abscesses. Undoubtedly, this radical approach offers the best chance of a definitive cure, and in older patients who have completed their families it may still be the most appropriate. The consequence of surgical castration can be overcome using hormone replacement therapy, and as there is no longer any risk of endometrial malignancy, unopposed oestrogens can be administered. Unfortunately, patients with acute pelvic sepsis or tubo-ovarian abscess are usually younger women of low parity. The principle behind treatment in these patients should be to remove the infected tissue but to attempt to retain reproductive potential.

Until recently it might have been argued that women with gross disruption of tubal anatomy and function are very unlikely to conceive, and attempts to conserve these organs are misguided, but the advent of *in vitro* fertilization has changed this situation. Dense pelvic adhesions are no barrier to oocyte retrieval using a transvaginal approach and pregnancy rates in leading centres average 25 to 30 per cent per *in vitro* fertilization cycle. At least one ovary and the uterus need to be conserved if *in vitro* fertilization is to be possible at a later date.

Unilateral adnexectomy has been reported to be the most frequently performed surgical procedure for the management of tubo-ovarian abscess. In a study of 50 patients for whom follow-up details were available, 14 per cent required further surgery and a further 14 per cent subsequently had a successful intrauterine pregnancy. A complication rate of nearly 40 per cent, with wound sepsis in six patients and damage to either bowel or bladder in the remainder, has been reported for a group of 28 patients.

The Pfannenstiel incision, the gynaecologist's traditional approach to the pelvis, is not recommended in surgery for the management of tubo-ovarian abscess because of the high risk of wound or subfascial infection when the recti are separated from the fascia. Either a muscle-splitting transverse or a midline or paramedian incision are more suitable, although the latter scar tends to be weaker.

Dissection inside the pelvis can be difficult, as the infected tissues are friable and tissue planes obscured. Care should be taken to free the uterus and adnexas from bowel and bladder, as these structures are the most frequently damaged. An attempt should also be made to identify the ureters, especially if the tubo-ovarian abscess involves the side walls of the pelvis. The pelvis should be irrigated with normal saline solution prior to abdominal closure. If conservative procedures are performed, corrugated drains should be inserted and brought out either through a posterior colpotomy or through the abdomen. If hysterectomy has been necessary, the vaginal vault can be left open and the vaginal wall edges oversewn. It is wise to bring out a large drain through the vagina.

Closure of the fascial layers is best achieved using non-absorbable monofilament and, if the skin is closed directly, interrupted sutures are to be preferred to a continuous suture. However, delayed primary closure of skin and subcutaneous tissues may be considered.

Rupture of a tubo-ovarian abscess

Rupture of a tubo-ovarian abscess is rare now, due to earlier initiation of effective antibiotic treatment of pelvic sepsis. However, if access to medical facilities has been delayed, rupture is potentially a most serious complication, which is associated with high mortality. Abundant free pus is present in the abdomen, with loculations between bowel and mesentery. Pus is sometimes present in the subphrenic spaces. The traditional treatment for a ruptured tubo-ovarian abscess involves hysterectomy and bilateral salpingo-oophorectomy. Surgery in these very ill patients should be as short as possible and more conservative measures, including peritoneal lavage and unilateral or bilateral adnexectomy, have been advocated. Peritoneal lavage involves freeing the bowel from adhesions along its length then irrigating the cavity with normal saline. Drains are sited above the liver and spleen and in the pouch of Douglas, and are brought out abdominally. Postoperative lavage is performed daily, using 2 litres of dialysis solution containing antibiotics, and is continued for 3 to 4 days until the lavage fluid returns clear.

Operative mortality of a group of patients treated in this way was 7.1 per cent, which was similar to that in other series with high hysterectomy rates. Mortality was greatest in the presence of subphrenic pus and when the bowel was damaged during surgery.

Conclusion

Acute pelvic sepsis is nearly always a complication of sexually transmitted disease and its incidence is increasing. Broad-spectrum antibiotic therapy active against

chlamydia, gonococcus, and anaerobes should be administered as soon as the patient presents to a medical practitioner. Tubo-ovarian abscess is a consequence of untreated or ineffectively treated pelvic sepsis, and the first-line treatment should be medical. Surgery is reserved for cases failing to respond to medical therapy or where rupture is thought likely. If surgery is appropriate, conservative measures will maintain the patient's reproductive potential while removing the primary focus of infection. In older patients who have completed their families, radical surgery may be appropriate.

Further reading

Boardman LA, Peipert JF, Brody JM, Cooper AS, Sung J. Endovaginal sonography for the diagnosis of upper genital tract infection. *Obstetrics and Gynecology* 1997; **90**: 54–7.

Bulas DI, Ahlstrom PA, Sivit CJ, Blask AR, O'Donnell RM. Pelvic inflammatory disease in the adolescent: comparison of transabdominal and transvaginal sonographic evaluation. *Radiology* 1992; **183**: 435–9.

Butch RJ *et al.* Drainage of pelvic abscesses through the greater sciatic foramen. *Radiology* 1986; **158**: 487–91.

Caspi B, Zalel Y, Or Y, Bar Dayan Y, Appelman Z, Katz Z. Sonographically guided aspiration: an alternative therapy for tubo-ovarian abscess. *Ultrasound in Obstetrics and Gynecology* 1996; **7**: 439–42.

Centers for Disease Control and Prevention. 1998 Guidelines for Treatment of Sexually Transmitted Disease. *Morbidity and Mortality Weekly Report* 1998; **47**: 1–7.

Ginsburg DS, Stern JL, Hamod KA, Genadry R, Spence MR. Tubo-ovarian abscess: a retrospective review. *American Journal of Obstetrics and Gynecology* 1980; **138**: 1055–8.

Ha HK *et al.* MR imaging of tubo-ovarian abscess. *Acta Radiologica* 1995; **36**: 510–14.

Henry-Suchet J, Soler A, Loffredo V. Laparoscopic treatment of tubo-ovarian abscesses. *Journal of Reproductive Medicine* 1984; **29**: 579–82.

Landers DV, Sweet RL. Current trends in the diagnosis and management of tubo-ovarian abscess. *American Journal of Obstetrics and Gynecology* 1985; **151**: 1098–110.

Lipscomb GH, Ling FW. Tubo-ovarian abscess in postmenopausal patients. *Southern Medical Journal* 1992; **85**: 696–9.

Method MW, Urnes PD, Neahring R, Sciarra JJ, Keith L. Economic considerations in the use of laparoscopy for diagnosing pelvic inflammatory disease. *Journal of Reproductive Medicine* 1987; **32**: 759–64.

Perez-Medina T, Huertas MA, Bajo JM. Early ultrasound-guided transvaginal drainage of tubo-ovarian abscesses: a randomized study. *Ultrasound in Obstetrics and Gynecology* 1996; **7**: 435–8.

Quan M. Pelvic inflammatory disease: diagnosis and management. *Journal of the American Board of Family Practice* 1994; **7**: 110–23.

Rivlin ME. Conservative surgery for adnexal abscess. *Journal of Reproductive Medicine* 1985; **30**: 726–30.

Rivlin ME, Golan A, Darling MR. Diffuse peritoneal sepsis associated with colpotomy drainage of pelvic abscess. *Journal of Reproductive Medicine* 1982; **27**: 406–10.

Slap GB *et al.* Recognition of tubo-ovarian abscess in adolescents with pelvic inflammatory disease. *Journal of Adolescent Health* 1996; **18**: 397–403.

Teisala K, Heinonen PK, Punnonen R. Transvaginal ultrasound in the diagnosis and treatment of tubo-ovarian abscess. *British Journal of Obstetrics and Gynaecology* 1990; **97**: 178–80.

Vermeeren J, TeLind RW. Intra-abdominal rupture of pelvic abscess. *American Journal of Obstetrics and Gynecology* 1954; **68**: 402–9.

Weström L. Incidence, prevalence and trends of acute pelvic inflammatory disease and its consequences in industrialized countries. *American Journal of Obstetrics and Gynecology* 1980; **138**: 880–92.

36.4 Ovarian accidents

Michael D. G. Gillmer

Introduction
Symptoms
Signs
Investigation
Operative management
Further reading

Introduction

Ovarian accidents that may be encountered at laparotomy include haemorrhage, rupture, torsion, infection, and degeneration. These complications usually involve ovaries that are enlarged, but haemorrhage and, rarely, torsion may also occur with ovaries of normal size, and are therefore especially likely to be misdiagnosed preoperatively. All of these ovarian complications are associated with pain and should ideally be distinguished from other causes of acute pelvic pain before embarking on a laparotomy. These include ectopic pregnancy, spontaneous abortion, salpingitis, appendicitis, diverticulitis, ileitis, ureteric colic, intestinal obstruction, torsion of other pelvic organs, and retroplacental haemorrhage or red degeneration of a fibroid during pregnancy.

Symptoms

The nature and timing of the pain will frequently be helpful in establishing the correct diagnosis. Pain caused by ovulatory bleeding will usually be at mid-cycle and is commonly unilateral (so-called *mittelschmerz*). It may also follow ovarian hyperstimulation, especially in women undergoing gonadotrophin treatment. Pain due to the rupture of a corpus luteum cyst will usually occur during early pregnancy, but may also occur late in the menstrual cycle, while intraperitoneal bleeding due to rupture of other functional ovarian cysts may occur at any time and cause confusion. Although most functional cysts rupture spontaneously, this may also occur during intercourse, labour, or pelvic examination. The pain is of acute onset and may be associated with a feeling that something has ‘given way'. It is usually followed by signs of peritonism, such as vomiting and diarrhoea. Shoulder-tip pain due to fluid under the diaphragm may also occur. Rupture of a dermoid cyst, although rare, releases extremely irritant sebaceous material into the peritoneal cavity, causing an acute abdomen. Bleeding into the pelvic cavity is generally of sudden onset and causes a constant ache of variable severity. Large haemorrhages may cause widespread abdominal pain and referred diaphragmatic pain in one or both shoulders, together with shock, which may be profound.

If the ovary is enlarged, torsion or bleeding into an ovarian cyst are more likely than haemorrhage, but intraperitoneal bleeding may also occur with the rupture of an endometriotic or neoplastic cyst. Torsion typically causes severe pain of sudden onset, to one or other side of the midline. There may also be a history of intermittent episodes of similar pain. Vomiting and irritability of the adjacent bowel or bladder may occur. Shock is rarely present and, although there may be a tachycardia, the woman is usually afebrile.

Signs

Small ovarian tumours lying in the pelvis can only be palpated on vaginal or rectal examination, and may be impossible to feel if there is guarding due to peritoneal irritation. Most lie posterior to the uterus, except for dermoid cysts (benign teratomas), which are frequently found in the uterovesical pouch.

Ovarian tumours that are large enough to be palpable through the abdominal wall tend to displace the bowel above and laterally and to lie centrally above the uterus. As a result, an ovarian tumour can easily be distinguished from ascites by the absence of flank dullness. The position of the tumour is, however, of little value in distinguishing it from an enlarged uterus or distended bladder. It is also difficult to be certain from which side of the pelvis the tumour arises, as even when an ovarian tumour is lying to one side, it may be found to arise from the opposite ovary.

Investigation

Before embarking on a laparotomy for a suspected ovarian accident, it is important to exclude other pelvic and abdominal tumours for which surgery is not required, such as a gravid uterus, distended bladder, or pelvic kidney. Rarely, an enlarged liver or spleen, or simply obesity, may also cause confusion. If necessary, empty the bladder with a catheter. Always consider the possibility of pregnancy and perform a pregnancy test on urine, if necessary, using a sensitive monoclonal b-human chorionic gonadotrophin kit. A placental souffle, identified with a portable Doppler ultrasound transducer, is also suggestive of pregnancy, while detection of fetal heart sounds is diagnostic.

Careful pelvic examination will indicate whether the mass is attached to, or a part of, the uterus. In addition, always bear in mind the possibility of a pelvic kidney.

Abdominal or pelvic ultrasound should be performed whenever possible, and will usually give clear evidence of an ovarian cyst, but cannot, even in experienced hands, be relied upon to distinguish between solid ovarian and uterine tumours, or to identify very early intrauterine or ectopic pregnancies. Radiographs of the abdomen and pelvis are seldom helpful but may, rarely, reveal calcification due to teeth in a dermoid cyst. If the tumour is not large and the diagnosis remains in doubt, then a diagnostic laparoscopy should always be performed.

Operative management

The surgical procedure adopted depends on the nature of the ovarian accident, the age of the patient, and, if a tumour is present, whether or not this appears to be malignant.

Haemorrhage into a functional ovarian cyst of less than 5 to 6 cm in diameter, diagnosed laparoscopically, is self-limiting and should be managed conservatively. If there is active intraperitoneal bleeding from a ruptured functional cyst or corpus luteum, haemostasis may be achieved laparoscopically using bipolar diathermy or, if available, the ‘cold coagulator' or contact Nd-YAG laser. If laparoscopic techniques fail, or are not feasible because of heavy bleeding, then a laparotomy should be performed to control the haemorrhage by oversewing the bleeding points, or by removing the cyst or, if necessary, the whole ovary.

Rupture of benign ovarian cysts usually occurs at the site of previous ischaemic degeneration, but may also occur if the cyst is endometriotic, papilliferous in nature, or malignant. If the cyst is of benign appearance and a plane of cleavage can be identified between the cyst and normal ovarian tissue, then an ovarian cystectomy should be performed and the ovary reformed. If normal ovarian tissue cannot be identified and the tumour is unilateral, a salpingo-oophorectomy is preferable. In women of childbearing years, ovarian tissue should be conserved whenever possible. However, in those rare circumstances where bilateral benign ovarian tumours are found and a bilateral salpingo-oophorectomy is necessary, because no normal ovarian tissue can be identified in either ovary, the uterus should always be conserved so that the woman may subsequently consider *in vitro* fertilization using ‘ovum donation'. If the cyst is endometriotic in nature, additional pelvic endometriotic tissue may be present, especially around the uterosacral ligaments. An effort should be made to excise or cauterize this tissue, taking great care to avoid ureteric damage, especially when using diathermy cautery.

Features suggesting malignancy, which may be noted at the time of laparotomy, include solidity or solid areas in a cystic tumour, areas of haemorrhage in the tumour, large surface blood vessels, tissue fungating through the surface of the ovary, free peritoneal fluid, a bilateral distribution, and metastases on the peritoneum or in the omentum or liver. When the tumour is clearly malignant, bilateral salpingo-oophorectomy is usually necessary, especially in peri- or postmenopausal women. Ideally, the tumour should, if possible, be removed intact to prevent contamination of the peritoneal cavity with the cyst contents. If there is free peritoneal fluid, this should always be sampled and sent for cytological examination. It is usual to perform a total abdominal hysterectomy at the same time in women who have completed their family, but if there is doubt about the diagnosis, and the tumour is confined to one ovary, then the uterus should be conserved. All large tumour nodules of more than 1 cm in diameter should be excised if possible, to ‘debulk' the tumour and make it more amenable to subsequent chemo- and/or radiotherapy. The greater omentum should also be excised to reduce the risk of subsequent large-bowel obstruction. If there is any doubt about the diagnosis, especially in younger women who wish to conserve their fertility, a frozen-section biopsy examination should be arranged whenever possible.

Thorough peritoneal lavage using warm isotonic saline solution should be performed at the end of the procedure, to remove blood and potentially irritant cyst fluid from the peritoneal cavity. This practice will reduce the incidence of postoperative ileus and adhesion formation, and is especially important if the tumour is a mucinous

cystadenoma as it may lessen the risk of subsequent pseudomyxoma peritonei.

If tubo-ovarian torsion is suspected or diagnosed laparoscopically, then an immediate laparotomy is indicated. If the torsion is recent or incomplete, the ovary and tube may be viable and can be conserved, using a suture to stabilize them and prevent a recurrence. If the ovary contains a cyst, an ovarian cystectomy should be performed. More commonly, the ovary and tube have suffered irreversible ischaemic damage and are gangrenous, necessitating salpingo-oophorectomy.

Tubo-ovarian infection (see also Chapter 36.3) may occasionally be identified during a laparotomy performed for suspected appendicitis or inflammatory bowel disease. If the infection is unilateral, it may be appropriate to perform a salpingo-oophorectomy. Pelvic inflammatory disease is, however, more commonly bilateral and usually occurs in young women of childbearing age. The temptation to perform a bilateral salpingo-oophorectomy must therefore be resisted. Obvious pus should be drained and conservative management with intravenous antibiotics should be instituted in an attempt to conserve fertility.

Degeneration of ovarian tumours may be identified as a coincidental finding at laparotomy. Most large, solid, ovarian tumours display necrosis or haemorrhage in their centres. These tumours should be treated by salpingo-oophorectomy.

Further reading

Peel KR. Benign and malignant tumours of the ovary. In: Whitfield CR, ed., *Dewhurst's textbook of obstetrics and gynaecology for postgraduates*, pp. 759–79. Blackwell Scientific, Oxford, 1995. [A detailed review of the pathology, clinical features, and management of benign and malignant ovarian tumours.]

Tumours of the ovary. In: Tindall VR. *Jeffcoate's principles of gynaecology*, 5th edn, pp. 450–82. Butterworths, London, 1987. [A comprehensive overview of the pathology and clinical management of ovarian tumours.]

36.5 Ovarian tumours

Ira R. Horowitz

Introduction
Functional cysts
Corpus luteum
Theca luteal cysts
Benign ovarian tumors and neoplasms
Endometriosis
Mature cystic teratomas
Ovarian malignancies in girls, adolescents, and young women
Germ cell tumors
Dysgerminoma
Endodermal sinus tumors
Embryonal carcinoma
Polyembryoma of the ovary
Gestational trophoblastic neoplasia (choriocarcinomas)
Immature teratomas
Sex cord stromal tumors of childhood and adolescence
Sertoli–Leydig cell tumors
Epithelial ovarian tumors
Risk factors
Screening
Surgical staging
Second-look laparotomy
Secondary cytoreduction
Quality of life
Follow-up and subsequent care
Further reading

Introduction

During the first three decades of life ovarian malignancies are very rare. Between 5 and 10 per cent of premenopausal ovarian neoplasms are malignant compared with 30 per cent of those in postmenopausal patients. Ovarian tumors in girls and adolescent females account for approximately 3 to 5 per cent of all ovarian neoplasms. In this age group approximately 65 per cent of ovarian neoplasms have been found to be germ cell tumors.

Functional cysts

These are the most common cysts present in the ovary. These can present as follicular, corpus luteal, or theca luteal cysts. They are benign and usually resolve spontaneously. These functional cysts can have diameters of up to 8 cm. Frequently, they are the result of dysfunction in the hypothalamus–pituitary–ovary axis and are an incidental finding on examination. Occasionally, symptoms such as pressure or peritonitis may be present from a mass effect or rupture. The peritoneal signs are usually resolved within 48 h after cyst rupture. This is in contradistinction to a ruptured tubo-ovarian abscess or ectopic pregnancy which will not resolve spontaneously and will need medical and/or surgical intervention.

Corpus luteum

Corpus luteal cysts result from persistently functioning corpora lutea. These cysts are also self-limiting and may express prolonged progesterone secretion. The classic triad of a persistent corpus luteum as described by Haldan is normal menses followed by spotting, a slightly enlarged adnexal mass, and unilateral pelvic discomfort.

Treatment

Although these cysts are usually self-limiting, the presence of a hemorrhagic corpus luteal cyst may necessitate laparoscopic intervention to control the bleeding and resultant hemoperitoneum. It is imperative that a serum pregnancy test be obtained to rule out the presence of a ruptured ectopic pregnancy rather than a hemorrhagic corpus luteal cyst. The majority of cysts will resolve spontaneously in approximately 4 to 10 weeks. Oral contraceptives used to suppress ovulation in this population resolve the problem in approximately 80 per cent of cases. Medical therapies other than oral contraception include oral, intramuscular, or subcutaneous progesterone delivery or gonadotropin-releasing hormone (**GnRH**) analogs. Surgical treatment as previously described is frequently through laparoscopy for treatment of ovarian torsion, persistent pain, or a ruptured hemorrhagic corpus luteum.

Theca luteal cysts

Theca luteal cysts are frequently bilateral and secondary to hyperstimulation from an intrauterine pregnancy or gestational trophoblastic disease. These cysts can be present in up to 50 per cent of hydatidiform moles and 10 per cent of choriocarcinomas. Bilateral luteomas of pregnancy can also exist during a normal intrauterine pregnancy. These masses have been reported with diameters greater than 20 cm. If left alone, they too will regress spontaneously within 6 weeks of delivery or treatment of the gestational trophoblastic neoplasm. Additional causes of theca luteal cysts include increased exogenous stimulation of the ovaries such as clomiphene citrate, GnRH analogs, human menopausal gonadotropins, and human chorionic gonadotropins (**hCG**).

Treatment

The majority of theca luteal cysts and luteomas of pregnancy regress spontaneously. As previously mentioned, if torsion occurs this would result in a surgical evaluation such as laparoscopy or laparotomy. Recent data for early torsion or intermittent torsion suggest that these patients can be treated conservatively without oophorectomy.

Benign ovarian tumors and neoplasms

Endometriosis

In addition to the functional cysts previously described, several benign neoplasms can also present as ovarian tumors. Endometriosis frequently presents itself in the third and fourth decades of life. Although classically thought of as endometrial implants in the peritoneal and celomic cavities secondary to retrograde menstruation, hematogenous and celomic metaplasia are accepted as alternative mechanisms. Ovarian endometriomas can present as small surface implants or large endometriomas measuring over 4 cm. These lesions are not malignant and do not have malignant predisposition. However, severe endometriosis can result in severe dysmenorrhea, anovulation, and/or infertility.

Treatment

Treatment of ovarian endometriomas is frequently surgical. GnRH analogs, oral contraceptives, and progestational agents are not successful in treating large endometriomas. Medical therapy is occasionally efficacious in treating small capsule implants. This diagnosis is usually made on laparoscopy. Ovarian endometriomas do not usually present as isolated lesions and coexist with celomic lesions, colon implants, and possible obliteration of the cul de sac. Surgical treatment consists of resecting endometrial implants and adhesions as well as the endometrioma of the ovary. Frequently this can be done by the skilled laparoscopist or via a laparotomy. If the patient desires future fertility, all attempts should be made to manage these lesions conservatively. Surgical excision, fulguration, and laser ablation are equally efficacious in managing these patients.

Mature cystic teratomas

Approximately 10 to 12 per cent of all ovarian neoplasms are benign cystic teratomas. These lesions develop from totipotential cells and consist of well differentiated ectodermal and mesodermal elements. Bilaterality occurs in approximately 10 to 15 per cent of patients. Cystic teratomas (dermoid cysts) frequently have a smooth capsule and can grow up to 30 cm in diameter. Entering the cyst will reveal the presence of bone, hair, sebaceous material, nerve tissue, and/or teeth, in addition to other tissues from the three germ layers. Functional teratomas, such as those consisting of thyroid tissue (struma ovarii), may be present in approximately 5 to 20 per cent of benign teratomas. Struma ovarii is present when the majority of the solid component is thyroid tissue. Thyroid toxicosis can present secondary to struma ovarii. Ovarian carcinoids can also be present as primary lesions in benign cystic teratomas. Patients presenting with the classic symptoms of carcinoid syndrome with unilateral adnexal masses should be evaluated for carcinoids in a mature cystic teratoma. Metastatic carcinoids to the ovary are frequently bilateral.

Treatment

Metastatic transformation occurs in less than 3 per cent of benign teratomas. Squamous cell carcinoma arising from Rokitansky's nodule accounts for approximately 80 per cent of these malignant lesions. The majority of mature cystic teratomas are diagnosed after ultrasound evaluation reveals fat or calcification in the ovarian tumor. There are also incidental findings when patients undergo abdominal radiographic evaluation for an unrelated pathology. On physical examination these lesions are anterior to the uterus as if floating above the uterus. This is secondary to the fat content in the benign teratoma. All of these lesions are treated surgically. Operative laparoscopy or operative endoscopy or laparotomy is used to perform an ovarian cystectomy with preservation of the remaining ovary. In large lesions oophorectomy may be required via laparoscopy or laparotomy. The contralateral ovary should be evaluated, and if an abnormal area is present, a cystectomy performed. All attempts should be made to close the capsule to prevent adhesion and infertility.

Ovarian malignancies in girls, adolescents, and young women

The preponderance of ovarian malignancies in females under the age of 20 years are germ cell tumors. Twenty-three per cent of the germ cell ovarian tumors presenting in this age group are malignant. Epithelial tumors, although possible in this age group, usually occur in women of 40 years and older, but can occur in their third and fourth decade. Girls and adolescents present with symptoms of increasing abdominal discomfort, abdominal girth, urinary frequency, satiated appetites, and (at times) weight loss. Germ cell tumors that produce hormones may also result in signs of precocious puberty, including early menarche and advanced Tanner stage. It is frequently difficult to obtain a satisfactory bimanual pelvic examination in the younger patient. The clinician must therefore rely on transvaginal or abdominal ultrasound to evaluate the pelvis. More expensive studies such as computed axial tomography (CT scans) and magnetic resonance imaging (MRI) are frequently used to evaluate metastatic disease to abdominal viscera, brain, and lungs. Tumor markers such as a-fetoprotein, hCG, and carcinoembryonic antigen (**CEA**) may be present in pure or mixed germ cell tumors. CA125, a tumor marker frequently used in epithelial ovarian cancers, is not useful in this group of patients. CA125 has a low sensitivity and specificity and may be elevated from conditions such as inflammatory bowel disease, pelvic inflammatory disease, endometriosis, pregnancy, leiomyoma of the uteri, and various benign conditions involving the celomic cavity. In addition to radiologic evaluation and assays for tumor marker levels (a-fetoprotein, b-hCG, and CEA) an intravenous pyelogram can assist in delineating the course of the ureter. External compression may obstruct the ureter and may require intervention such as a percutaneous nephrostomy tube prior to surgery to protect the affected kidney.

Germ cell tumors

Germ cell tumors arise from the primitive germ cells and can develop into dysgerminomas or embryonal carcinomas. The latter is further subdivided into extraembryonic and embryonic tumors.

Dysgerminoma

Dysgerminoma is the most common malignant germ cell tumor accounting for approximately 1 to 2 per cent of primary ovarian neoplasms and 2 to 5 per cent of ovarian malignancies. More than 90 per cent of these lesions are present in patients younger than 30 years of age, with survival of 75 to 90 per cent. As with benign cystic teratomas, bilateral involvement is present in 10 to 15 per cent of patients presenting with this tumor.

Treatment

Treatment consists of surgical debulking. If residual disease is present, attempts should be made to resect all of the disease without subjecting the patient to extensive resection of abdominal viscera. Although traditionally radiation therapy has been a mainstay of treatment for these lesions, studies have shown chemotherapy to be efficacious. If no residual disease is present, four to six courses of therapy would be appropriate. Radiographic evaluation and physical examination should be used postoperatively for follow-up. All attempts should be made to preserve the contralateral ovary and uterus. Pregnancy should be delayed for at least 1 year following chemotherapy. Pure dysgerminomas do not have any tumor markers present. Mixed germ cell tumors can present with a dysgerminoma and embryonal cell line. Tumor markers are then useful in following patients during treatment. The most commonly used chemotherapeutic regimens are **BEP** (bleomycin, etoposide, cisplatin), **VBP** (vinblastine, bleomycin, cisplatin), and **VAC** (vincristine, actinomycin, and cyclphosphamide).

Endodermal sinus tumors

Endodermal sinus tumors were originally described as extra-embryonal germ cells resembling the endodermal sinus of Duval in the yolk sac of the rat placenta. Endodermal sinus tumors including mixed forms are the second most common germ cell tumor. Median age of presentation is approximately 19 to 20 years. These lesions are usually bilateral, with 5 per cent present in a benign cystic teratoma. Ninety per cent of patients with these lesions present with a primary symptom such as abdominal pelvic pain. In addition to a-fetoprotein, these lesions can produce low levels of a_1 -antitrypsin. As previously mentioned, the a-fetoprotein can be used as a measure of tumor response to therapy. Aggressive surgical treatment of these patients should be made. Preservation of the adnexa and uterus are important in those patients desiring future fertility. Postoperative adjuvant chemotherapy with protocols utilizing VAC, VPB, and BEP is a mainstay for the treatment of these patients. The a-fetoprotein level should be used as a marker to determine the number of courses required to treat this lesion. If no a-fetoprotein is present or it returns to within normal range in less than four courses, four courses should be administered. If the level continues to be elevated, additional therapy should be administered to the patient.

Embryonal carcinoma

Embryonal lesions usually present at 14 to 15 years of age. Embryo-nal carcinomas can differentiate into a multitude of tumors including embryonic tumors, common teratomas, polyembryomas, endodermal sinus tumors, and choriocarcinomas. These totipotential cells present with elevated levels of a-fetoprotein and hCG.

Treatment

Treatment consists of surgical exploration, tumor debulking, unilateral adnexectomy, and adjuvant chemotherapy. Regimens used include those discussed for endodermal sinus tumors—VAC, VBP, and BEP. Initial treatment with BEP for embryonal carcinomas appears to be appropriate after unilateral oophorectomy. Since these lesions appear to be unilateral, unilateral salpingo-oophorectomy is all that is required.

Polyembryoma of the ovary

Polyembryoma of the ovary is rare and composed of ‘embryoid bodies'. These lesions occur in premenarchal patients, frequently with precocious puberty and signs of pseudopuberty. a-Fetoprotein and hCG titers may also be elevated.

Treatment

Treatment consists of surgical resection of the affected ovary, evaluation of the celomic cavity, lymph nodes, and contralateral ovary, and treatment with adjuvant chemotherapy, as previously described.

Gestational trophoblastic neoplasia (choriocarcinomas)

Non-gestational choriocarcinomas are rare. Lesions of the ovaries usually occur in patients less than 25 years of age with up to 50 per cent exhibiting signs of

isosexual puberty. Pure choriocarcinomas can be treated with methotrexate, actinomycin, and cyclophosphamide, or alternative regimens as previously described including BEP or VBP. Other regimens for advanced trophoblastic disease utilized multiple agents. b-hCG titers are elevated in this group of patients and are followed to determine tumor response. hCG titers can be plotted on log paper and a line drawn to determine how many additional courses patients require upon reverting to a normal level.

Immature teratomas

Immature teratomas can occur within the first three decades of life. Although unilateral, these lesions may coexist with a benign cystic teratoma in the contralateral ovary. If a pure immature teratoma is present, these lesions do not produce steroids or tumor markers.

Treatment

Treatment is surgical with the removal of the affected ovary and a unilateral salpingo-oophorectomy. If no metastatic disease is present, the treatment is complete. If the patient does not desire further fertility, a total abdominal hysterectomy and bilateral salpingo-oophorectomy should be performed. Metastatic lesions should be treated with the previously mentioned adjuvant chemotherapies including VAC, PVB, BEP, and etoposide and platinum.

Sex cord stromal tumors of childhood and adolescence

Sex cord stromal tumors of the ovary include granulosa, theca, Sertoli–Leydig, collagen-producing stromal cells, or their embryonic precursors. Granulosa cell tumors account for less than 7 per cent of all ovarian tumors in the first two decades of life with the majority of juvenile type. These lesions can be bilateral in approximately 2 per cent of patients and therefore require only a unilateral salpingo-oophorectomy in patients desiring future fertility. Attempts should be made to leave the contralateral ovary in place, as well as the uterus, to permit future fertility. These tumors are relatively slow growing. Advanced lesions can be treated with adjuvant chemotherapy, as previously described, including BEP. The adult granulosa tumorscan also be treated as previously described with adjuvant chemotherapy including BEP, VAC, and PAC (cisplatin, Adriamycin, cyclophosphamide).

Sertoli–Leydig cell tumors

Sertoli–Leydig (arrhenoblastoma) tumors account for less than 2 per cent of ovarian neoplasms. These tumors produce androgens and thereby result in virilization of the female patient. Clinical signs include acne, hirsutism, clitoromegaly, male-pattern baldness, breast atrophy, and voice changes. Androstanedione and testosterone levels are elevated, with dehydroepiandrosterone sulfate being normal or elevated.

Treatment

Treatment again is conservative surgery with unilateral salpingo-oophorectomy and evaluation of the contralateral ovary and peritoneal cavity. If the patient does not desire future fertility, a total abdominal hysterectomy and bilateral salpingo-oophorectomy should be performed. These lesions are not very neoplastic and therefore do not respond well to adjuvant therapy. For patients who require adjuvant therapy, chemotherapy with the previously noted germ cell regimens as well as radiation therapy have been reported to be efficacious in certain cases.

Epithelial ovarian tumors

These tumors arise from the epithelium, celomic epithelium, or mesothelium. Seventy-five per cent of epithelial tumors are serous type, 20 per cent are mucinous, 2 per cent are endometrioid, and the remainder are clear cell carcinoma and undifferentiated.

Risk factors

Risk factors include advancing age, family history (including BRACA1 and BRACA2), nulliparity, and possibly fertility drugs. However, the risks are decreased with prolonged breast feeding, oral contraceptives, and in multiparous patients. Unfortunately, epithelial tumors of the ovary are usually not identified early in their course owing to their non-specific symptoms. Early symptoms are usually vague including urinary frequency, constipation, and abdominal discomfort. As the disease progresses, abdominal distension, bloating, nausea, early satiety, anorexia, and vaginal bleeding may occur. The median age for patients presenting with epithelial tumors is about 64 years old, with less than 1 per cent of tumors occurring at less than 30 years.

Screening

Screening programs have not been successful in ovarian cancer. Physical examination, tumor markers including CA125, and ultrasound (including duplex Doppler) have been utilized to assist in identifying those patients with early ovarian cancer. As previously mentioned in this chapter, CA125 has a very low sensitivity and specificity in the premenopausal patient. Unfortunately, that also holds for the menopausal patient in this type of cancer. Ultrasound is being used with increased frequency. It is imperative that not only cyst formation be evaluated but the morphology of the cyst, thickness of the septum, and blood flow (pulsatile index) of the ovarian vessels. A pulsatile index of less than 1 increases the likelihood that a malignancy may be present. Complex masses on ultrasound in a menopausal patient require surgical evaluation. The oophorectomy may be done via laparoscopy or laparotomy.

Surgical staging

Ovarian cancer spreads by three routes: trans-celomic, lymphatic, and hematogenous. As such, staging for ovarian carcinoma is surgical (Table 1). Most patients who are presumed to have early disease clinically are upstaged after surgery. Approximately 28 per cent of stage I and 43 per cent of clinical stage II carcinomas are upstaged when appropriate surgical staging is performed. As previously mentioned, the identification of early tumors is rare in this disease. However, these tumors may present at an early stage and in a premenopausal patient who desires fertility treatments. Prior to surgical staging the patient evaluation should include a Pap smear, mammogram, chemistry profile, complete blood count with differential platelets, and a chest radiograph. An intravenous pyelogram, CT, MRI or barium swallow may be used in situations where advanced disease is anticipated.

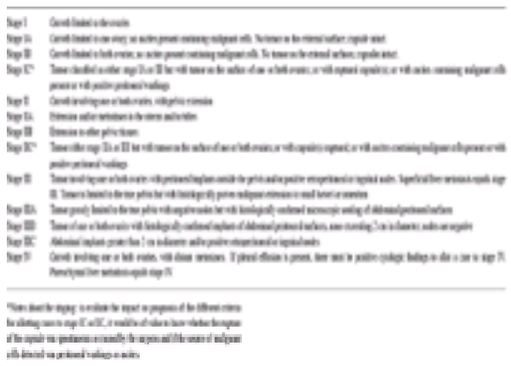


Table 1 FIGO staging for carcinoma of the ovary.

A thorough staging procedure begins with a complete evaluation of the celomic cavity upon entering the abdomen. Washings of the abdomen and pelvis are then obtained as well as diaphragm. The pelvis is evaluated and if the lesion appears to be unilateral and confined to the ovary with no distant disease, a unilateral salpingo-oophorectomy should be performed in those patients desiring future fertility. If the patient does not desire future fertility, a total abdominal hysterectomy and bilateral salpingo-oophorectomy should be performed. When no gross disease is present an omentectomy should be performed, pelvic and periaortic lymph nodes should be sampled to the level of the renal vein, and random peritoneal biopsies should be taken. In those patients with advanced disease, all attemptsshould be made to remove the disease present. Optimal cytoreduction of less than 1 cm markedly improves survival after adjuvant chemotherapy. In patients with less than 2 cm of disease outside of the pelvis, pelvic and periaortic lymph nodes should also be obtained to rule out a more advanced stage. In patients with bulky disease of 2 cm or more outside the pelvis or positive pleural effusions, a lymph node dissection is not needed since these patients have a stage IIIC or IV cancer, respectively, which

would not change by performing lymph node dissection.

Stages Iag1 and Ib2 can be treated conservatively with unilateral adnexectomy or total abdominal hysterectomy and bilateral salpingo-oophorectomy in those not desiring future fertility. Adjuvant chemotherapy is not required in this group of patients. Stage Iag2, Ibg2, and g3 as well as stage IC should be treated with three or four courses of a platinum-based chemotherapy. In the United States, cisplatin taxol or carboplatin taxol are used. In Europe, platinum cyclophosphamide or platinum analog taxol are used. Whole abdominal irradiation has also been reported to be efficacious in patients with early disease, but the side-effects of this modality outweigh that of cytotoxic therapy. Studies have also been done with intraperitoneal potassium-32 and have been shown to be efficacious. Patients with stage II, III, or IV disease are treated with platinum analog taxol or platinum cyclophosphamide for six courses of chemotherapy. In those with bulky disease after the initial surgery, many gynecologic oncologists have recommended neoadjuvant chemotherapy with three courses of chemotherapy, followed by interval debulking and an additional three to six courses of therapy.

Second-look laparotomy

Second-look laparotomies continue to remain controversial. Ovarian cancer in 30 to 50 per cent of patients with a negative second-look laparotomy recurs within 3 to 5 years. Patients with stage I or II carcinoma would have an 85 to 95 per cent probability of a negative second-look laparotomy, whereas those with stage III or IV disease would only have a 30 to 45 per cent chance of a negative result. Therefore, it is not recommended that patients with stage I or II disease have a second-look laparotomy. In patients with stage III or IV disease, negative second-look laparotomies may benefit from consolidation chemotherapy. There are several trials currently evaluating the role of consolidation therapy in negative second-look laparotomies. One in particular is the use of intraperitoneal potassium-32 by the Gynecologic Oncology Group in the United States. It is imperative that any patient who desires to undergo a second-look laparotomy should be willing to accept additional chemotherapy if positive. If the patient does not desire future therapy, they should be managed expectantly with CA125 and physical examination. If however they desire salvage therapy, the following drugs have been shown to be efficacious: topotecan, etoposide, gemcitabine, hexamethylmine, and doxil. Other drugs that can be used in third-line therapy include: tamoxifen, 5fu and leucovorin, ifosfamide, and mesna.

Secondary cytoreduction

Unlike a second laparotomy which is to evaluate microscopic disease or disease not identified radiographically, secondary cytoreduction would be extensive tumor cytoreduction during a second-look laparotomy. This procedure is even more controversial than second-look laparotomy. Survival is not significantly altered in these patients.

Quality of life

It is imperative that the surgeon bears in mind the quality of life in this group of patients. With the recent data supporting neoadjuvant chemotherapy, extensive cytoreduction at the expense of quality of life is no longer warranted. Patients may benefit from an additional laparotomy after three courses of chemotherapy. Instrumentation such as ultrasonic surgical aspiration has also enhanced our ability for optimal cytoreduction while reducing the number of bowel resections and diaphragmatic debulkings for these patients.

Follow-up and subsequent care

As it is difficult to identify these patients at an early stage, it is difficult to determine the best mode of follow-up. At the present time, the patient should be followed with a physical examination every 3 months for the first 2 years, followed by evaluations every 6 months for the subsequent 3 years. If CA125 levels were elevated upon presentation, these would be efficacious in following the patient for recurrent disease. However, difficulties will arise if the CA125 is slightly elevated and no disease can be identified radiographically or upon physical examination. There are no studies to support treating these patients at this stage as opposed to waiting for them to become symptomatic or disease to be palpated. Despite this, patients continue to fixate on their CA125 levels and frequently will not permit the clinician to manage them expectantly.

These are exciting times with respect to the treatment of ovarian carcinoma. Our patients are much more knowledgeable of the disease process as a result of the internet and their ability to research ovarian cancer. Numerous agents are becoming available each year to treat this devastating disease. Unfortuantely, the cost of diagnosing and treating ovarian cancer continues to increase. Only through collaborative research studies will we be ble to define the most successful treatment protocols.

Further reading

Berek JS. Epithelial ovarian cancer. In: Berek JS, Hacker NF, eds, *Practical gynecologic oncology*, 2nd edn, pp. 327–76. William & Wilkins, Baltimore, 1994. [This chapter provides a thorough review of the contemporary management of ovarian carcinoma.]

Berek JS, Griffith CT, Leventhal JM. Laparoscopy for second-look evaluation in ovarian cancer. *Obstetrics and Gynecology* 1981; **58**: 192. [One of the original publications assessing second-look evaluation in ovarian carcinoma.]

Berek JS, Hacker NF, Lagasse LD, Nieberg RK, Elashoff RM. Survival of atients following secondary cytoreductive surgery in ovarian cancer. *Obstetrics and Gynecology* 1983; **61**: 189–93. [This article provides the basis for many of the published recommendations pertaining to secondary cytoreductive surgery.]

Burke JW, Morris M. Secondary cytoreductive surgery for ovarian cancer. *Obstetrics and Gynecology Clinics of North America* 1994; **21**: 167–76. [The role of secondary cytoreduction remains controversial; however this publication provides a contemporary update.]

Deppe G, Malviya VK, Boike G, Malone JM, Jr. Use of cavitron surgical aspirator for debulking of diaphragmatic metastases in patients with advanced carcinoma of the ovaries. *Surgery, Gynecology, Obstetrics* 1989; **168**: 458. [Diaphragmatic metastases are frequently difficult to excise without increased morbidity via surgical resection. The surgical aspirator permits debulking of these lesions with minimal morbidity.]

Hacker NF, Berek JS, Lagasse LD. Gastrointestinal operations in gynecologic oncology. In: Knapp RE, Berkowitz RS, eds, *Gynecologic oncology*, 2nd edn, pp. 361–75. McGraw-Hill, New York, 1993. [This chapter provides a review of the extensive nature of surgical debulking and gastrointestinal resection required to obtain optimal cytoreduction.]

Horowitz IR, De La Cuesta RS. Benign and malignant tumors of the ovary. In Carpenter SEK, Rock JA, eds, *Pediatric adolescent gynecology*, pp. 397–416. Raven Press, New York, 1992. [This chapter provides a thorough review of benign and malignant tumors of the ovary in the pediatric adolescent population.]

Hoskins WJ *et al.* Influence of secondary cytoreduction at the time of second-look laparotomy on the survival of patients with epithelial ovarian carcinoma. *Gynecologic Oncology* 1989; **34**: 365–71. [This article provides a review of secondary cytoreduction at the time of second-look, including those that appear to be clinically negative.

Hoskins WJ, Bundy BN, Thigpen TJ, Omura GA. The influence of cytoreductive surgery on recurrence-free interval and survival in small volume Stage III epithelial ovarian cancer: a Gynecology Oncology Study Group. *Gynecologic Oncology* 1992; **47**: 159. [A landmark study reviewing the influence of cytoreductive surgery in small volume stage III carcinomas.]

Hunter RW, Alexander NDE, Soutter WP. Meta-analysis of surgery in advanced ovarian carcinoma: is maximum cytoreductive surgery an independent determinant of prognosis. *American Journal of Obstetrics and Gynecology* 1992; **166**: 504. [This article provides a thorough review and re-evaluation of the role of maximum cytoreduction in the treatment of patients with ovarian carcinoma.]

Lorigan PC, Crosby T, Coleman RE. Current drug treatment guidelines for epithelial ovarian cancer. *Drugs* 1996; **51**: 571–84. [Provides a review of the current treatment guidelines for epithelial ovarian carcinoma.]

Markman M *et al.* Ninety-six-hour infusional paclitaxel as salvage therapy of ovarian cancer patients previously failing treatment with 3-hour or 24-hour paclitaxel infusion regimens. *Journal of Clinical Oncology* 1998; **16**: 1849–51. [Reviews the role of 96-h infusion as salvage therapy in ovarian cancer patients.]

McGuire WP *et al.* Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. *New England Journal of Medicine* 1996; **334**: 1–6. [Landmark study evaluating the role of taxol versus cyclophosphamide in the treatment of ovarian cancer.]

Nguyen HN, Averette HE, Janice KM. Ovarian carcinoma: a review of the significance of familial risk factors and the role of prophylactic oophorectomy in cancer prevention. *Cancer* 1994; **74**: 545–55. [Reviews the role of prophylactic oophorectomy in patients with a family history of ovarian carcinoma.

NIH Consensus Conference Ovarian Cancer: Screening, Treatment and Follow-Up. *Journal of the American Medical Association* 1995; **273**: 491–7. [The NIH Consensus report reviews the recommendations for screening, treatment and follow-up.]

Omura GA *et al.* Long-term follow-up and prognostic factor analysis in advanced ovarian carcinoma: The Gynecologic Oncology Group experience. *Journal of Clinical Oncology* 1991; **9**: 1138. [This article reviews prognostic factors in patients with advanced ovarian carcinoma.]

Rubin SC *et al.* Recurrence after negative second-look laparotomy for ovarian cancer: analysis of risk factors. *American Journal of Obstetrics and Gynecology* 1988; **159**: 1094. [This article reviews the

role of second-look and the delayed recurrence found in patients with a negative second-look.]

van der Burg MEL *et al.* The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. *New England Journal of Medicine* 1995; **332**: 629–34. [This article reviews the role of neoadjuvant chemotherapy prior to surgical debulking of advanced ovarian carcinoma.]

Williams L. The role of secondary cytoreductive surgery in epithelial ovarian malignancies. *Oncology* 1992; **6**: 25–32. [Another article discussing the role of secondary cytoreduction in ovarian carcinoma.]

Williams S, Blessing JA, Liao S-Y, Ball H, Hanjani P. Adjuvant therapy of ovarian germ cell tumors with cisplatin, etoposide, and bleomycin: a trial of the Gynecologic Oncology Group. *Journal of Clinical Oncology* 1994; **12**: 701–6. [This article reviews the role of cisplatin, etoposide, and bleomycin in germ cell tumors.]

Young RC *et al.* Adjuvant therapy in stage I and stage II epithelial ovarian cancer: results of two prospective randomized trials. *New England Journal of Medicine* 1990; **322**: 1021–7. [This study reviews the role of adjuvant therapy in treating early stage I and stage II epithelial carcinomas of the ovary.]

Zambetti M, Esrobedo A, Pilotti S, DePaolo G. Cisplatinum/vinblastine/bleomycin combination chemotherapy in advanced or recurrent granulosa cell tumors of the ovary. *Gynecologic Oncology* 1990; **36**: 317–20. [This article reviews the role of cisplatinum, vinblastine, and bleomycin in treatment of granulosa cell tumors.]

36.6 Uterine fibroids

William D. Boyd and F. Mark Charnock

[Introduction](#)
[Pathology](#)
[Clinical features](#)
[Treatment](#)
- [Myomectomy](#)
- [Hysterectomy](#)
- [Surgical points to consider](#)
- [Recent advances](#)
- [Fibroids and pregnancy](#)

[Further reading](#)

Introduction

The correct term for a fibroid is a leiomyoma. It is estimated to be the most common tumour in the human body, being present in approximately 20 per cent of women at autopsy. Hence general surgeons are likely to encounter a uterine fibroid more than once in their careers in more than one presentation.

The cause of fibroids is unknown but the main contender is the stimulation provided by unopposed oestrogens. They are more common in women who have not borne children and in women of black African origin. The latter tend to develop fibroids when young even despite having borne children. In women of European origin, fibroids tend to cause symptoms from around the age of 30 years onwards. After the menopause they atrophy in most cases, due both to decreased oestrogen levels and blood supply. In 10 to 15 per cent of cases, fibroids are found together with endometriosis, which can make surgery difficult.

Pathology

Macroscopically the appearance is of a firm, round tumour in the uterine wall, which itself may be thickened and hypertrophied. The tumours usually arise in the body of the uterus or, less commonly, in the cervix; infrequently, they arise from the round ligaments. Fibroids within the wall of the uterus are described as being intramural; if projecting from the peritoneal surface of the uterus, as subserous; if between the layers of the broad ligament, as intraligamentary; and if projecting into the uterine cavity, as submucous (Table 1). Subserous fibroids may become pedunculated (Fig. 1) and submucous ones polypoid, even protruding through the external cervical os. Pedunculated fibroids may become adherent to other structures, particularly the omentum, gain a second blood supply, and lose their uterine attachment; they are then termed parasitic. Fibroids arising in the cervix or intraligamentary fibroids may displace the ureters sometimes quite a distance from their normal position, placing them at risk during surgical removal.

Intramural:
Uterine body
Cervix
Subserous:
Pedunculated
Parasitic
Intraligamentary
Submucous:
Pedunculated

Table 1 The different sites of fibroids

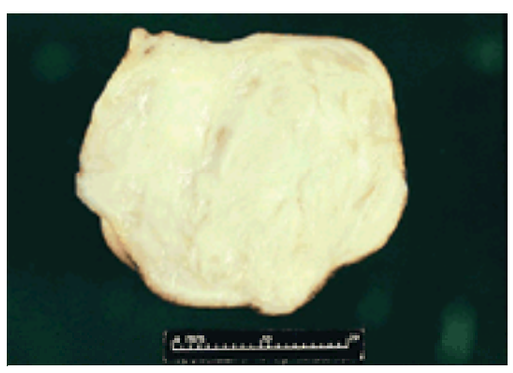


Fig. 1. The typical white whorled appearance of the cut surface of a pedunculated fibroid.

Fibroids may be single or multiple. Their size varies from microscopic to large enough to fill the whole abdomen. Characteristically they are firm, but as a result of degeneration may be soft and cystic, or rock hard when calcified. They have a false capsule of compressed uterine muscle. The blood supply enters from the periphery and thus fibroids are relatively avascular.

Various degenerative changes, encouraged by poor blood supply, occur in fibroids. These include hyaline degeneration, cystic degeneration, calcification, necrosis, red degeneration, sarcomatous degeneration, and infection. Other rare complications, such as benign metastases in the lung, and an association with polycythaemia, have been described.

The more important changes from the surgeon's point of view are red degeneration, calcification, and sarcomatous degeneration. Calcification is most often seen in pedunculated fibroids or in women well beyond the menopause. The fibroids becomes rock hard and will show up on radiography as a calcified mass. Red degeneration (Fig. 2) most often occurs in pregnancy or near the menopause. It is similar to the process of infarction and gives rise to severe pain. Sarcomatous degeneration is a very infrequent occurrence (1:1000 up to 8:1000), although about two-thirds of uterine sarcomas do arise in fibroids. The median age of presentation is in the late 40s. A better prognosis is reported for premenopausal women. A rapid, painful increase in size of a fibroid can be a suspicious sign of sarcomatous change. The treatment is total hysterectomy and bilateral salpingo-oophorectomy. Five-year survival rates vary from 20 to 63 per cent.



Fig. 2. Haemorrhage into an intramural fibroid (red degeneration).

Clinical features

The clinical presentation is variable (Table 2), with many fibroids being symptomless and being found unexpectedly on a routine pelvic examination, for example while performing a cervical smear, or at laparotomy. While most symptoms lead to presentation to a gynaecologist, some will first present to a general surgeon and other specialties. Many patients complain of an abdominal swelling. In large fibroids, pressure effects on the pelvic veins may result in, or exacerbate, varicose veins and haemorrhoids. A large tumour filling the abdomen may cause dyspnoea. Pressure effects on the bladder or bladder neck may cause frequency of micturition and stress incontinence, and retention of urine, respectively. Irregular menstruation and infertility are also associated with fibroids. Pain, which may be acute and severe, may result from torsion of a pedunculated fibroid or that associated with red degeneration. Whether the pain is due to the fibroid or some other less obvious cause is often a diagnostic problem.

Symptom or sign	Specialist referral
Subfertility, irregular menses	Gynaecologist
Abdominal swelling	Surgeon, gynaecologist
Varicose veins, haemorrhoids	Surgeon
Dyspnoea	General physician
Frequency of micturition, stress incontinence, retention of urine	Urologist, gynaecologist

Table 2 Referral patterns at fibroid presentation

The main differential diagnosis is from ovarian tumours when as a rule, on vaginal examination, the uterus can be detected as separate from the swelling. However, if the ovarian tumour is adherent to the uterus, difficulty arises. A pelvic ultrasound examination in most cases can delineate an ovarian from a uterine mass, although it is not unusual for the correct diagnosis to be reached only at laparotomy.

Treatment

The incidental finding of fibroids (leiomyomas) at laparotomy should be noted and rarely is surgery indicated as an unplanned procedure in this situation. The only time when surgery would be considered is where a fibroid was pedunculated and had undergone torsion causing symptoms. This can be considered to be a true emergency procedure.

Myomectomy

The main reason for performing a myomectomy (a local enucleation of the fibroid) is to improve fertility prospects for the management of menorrhagia exacerbated by fibroids in women who wish to retain their fertility potential, or to relieve pressure symptoms. The main complication during surgery is uncontrollable haemorrhage, which can require hysterectomy and blood transfusion. However, it is rare that a hysterectomy has to be performed, although it is essential to counsel patients preoperatively. A current approach in these patients is to pretreat with gonadotrophin-releasing hormones, which reduces blood loss at surgery and the need for a hysterectomy. Hence, there is no place for emergency myomectomy by a general surgeon or a gynaecologist except in the instance mentioned above.

Gonadotrophin-releasing hormone analogues are effective in shrinking fibroids and in reducing blood loss at myomectomy with large fibroids. This is now accepted practice before myomectomy for large fibroids. These analogues suppress ovarian function and can result in troublesome menopausal symptoms. Some of these can be controlled with norethisterone, 5 mg daily.

Transcervical endoscopic resection or laser vaporization of submucous fibroids, after a preoperative course of suppressant analogues, is an already developed technique. The scarring of the resected area may hinder fertility.

Rupture of a myomectomy scar during pregnancy is rare. However, some obstetricians will opt for elective caesarean section at 38 weeks' gestation in patients who have had a previous myomectomy.

Hysterectomy

The standard treatment in a woman who has completed her family and who is usually aged over 35 years is a hysterectomy. In a woman under 45 years of age the problems involved in removal of both ovaries and subsequent hormone replacement therapy must be discussed before surgery. The rationale behind this is as follows: there is a 4 to 5 per cent risk of ovarian disease developing after hysterectomy and a 0.2 per cent (0.1–0.5 per cent) risk of developing ovarian cancer in the conserved ovaries; the function of the ovaries has declined significantly by this age, the average age of the menopause being 51 years in Europe.

Another application of suppressant analogues is to shrink a large fibroid sufficiently to allow a vaginal hysterectomy rather than an abdominal operation. Care must be taken in selecting these patients as there is a definite association of fibroids with endometriosis, which is a relative contraindication to vaginal hysterectomy.

Surgical points to consider

Myomectomy

In performing a myomectomy the surgeon must be careful to avoid damaging the fallopian tubes in their course through the myometrium. Incisions on the posterior uterine wall are to be avoided as far as possible because of the increased risk of adhesions involving the intestines, the ovaries, and the fallopian tubes. Adhesions involving the ovaries and fallopian tubes may hinder ovum pick-up. The ‘hood' operation (as devised by Victor Bonney), in which the uterine incision is placed anteriorly over the fundus, is a useful method for removing posterior intramural fibroids so as to avoid this complication.

Long-term problems associated with performing a myomectomy are recurrence of fibroids necessitating subsequent surgery and uterine rupture during pregnancy.

Meticulous haemostasis during surgery is essential and to aid this a clamp (Bonney's clamp) can be placed at the level of the upper cervix temporarily to clamp the

uterine vessels. A modification of this is to tie a Foley-type catheter around the upper cervix, having first fashioned a hole in the broad ligament on both sides. In conjunction with this method, the uterine blood supply coming from the ovarian artery can be interrupted medial to the ovary by using a rubber-shod, Bainbridge-type, vascular clamp.

Hysterectomy

The ureter is at risk of trauma during surgery, particularly with either a cervical or a broad-ligament fibroid where vision can be limited. The course of the ureter can be very distorted or associated with endometryosis. Endometriosis canplead to dense adhesions involving the bowel, greater omentum, peritoneum lining the pouch of Douglas, ovaries, and fallopian tubes.

Recent advances

With the tremendous advances in endoscopic equipment, many centres are now performing both hysterectomies and myomectomies laparoscopically, with obvious benefits to the patient. However, as these skills are exacting and the operations time-consuming, there is little if any role for the general surgeon in this area. Embolization of the uterine arteries or of the vessels feeding the fibroids via a catheter introduced through the femoral artery relieves symptoms in 85 per cent of patients. As the effects on subsequent pregnancy are unclear, it is probably best reserved for women who have completed their family.

Fibroids and pregnancy

The incidence of fibroids in pregnancy varies between 0.5 and 5 per cent. Red degeneration of the fibroid can occur and give rise to acute abdominal pain, vomiting, low-grade pyrexia, exquisite localized peritoneal tenderness over the surface of the fibroid, and a marked leucocytosis.

The differential diagnosis includes many non-obstetrical causes of abdominal pain in the pregnant woman, of which the general surgeon should be aware. In attempting to arrive at a diagnosis, Alder's sign may be of help in differentiating between pain and tenderness of genital and extragenital origin. The source of pain is located by palpation with the patient lying flat. She is then turned on her left side with the examining hand still in position. If the pain shifts to the left it is likely to originate in the uterus or adnexae but if there is no change the source is probably extragenital (e.g. appendicitis).

Unless there is serious doubt about the diagnosis the treatment is conservative, with bed rest, sedation, and analgesia; resolution usually occurs in 4 to 7 days. Rarely, laparotomy may be required if the symptoms worsen and the possibility of an acute surgical condition such as appendicitis cannot be ruled out. A pedunculated fibroid may have undergone torsion and secondary degeneration, which provides the sole indication for myomectomy during pregnancy. Fibroids should not be removed at caesarean section.

Further reading

Alder N. A sign of differentiating uterine from extra uterine complications of pregnancy and puerperium. *British Medical Journal* 1951; **2**: 1194–5.

Amias AG. Abdominal pain in pregnancy In: Turnbull A, Chamberlain G, ed. *Obstetrics*, pp. 608–9. Churchill Livingstone, London, 1989.

Candiani GB, Fedele L, Parazzini F, Villa L. Risk of recurrence after myomectomy. *British Journal of Obstetrics and Gynaecology* 1991; **98**: 385–9.

Corson SL, Brooks PG. Resectoscopic myomectomy. *Fertilization and Sterilization* 1991; **55**: 1041–4.

DeLancey JO. A modified technique for hemostasis during myomectomy. *Surgical Gynecology* 1992; **174**: 153–4.

Donnez J, Gillerot S, Bourgonjon D, Clerckx F, Nisolle M. Neodymium:YAG laser hysteroscopy in large submucous fibroids. *Fertilization and Sterilization* 1990: **54**: 999–1003.

Dubuisson JB *et al.* Gonadotrophin releasing hormone agonist and laparoscopic myomectomy. *Clinical Therapy* 1992; **14** (Suppl. A): 51–6.

Friedmann AJ, Barbieri RL, Doubilet PM, Fine C, Schiff I. A randomized double blind trial of a gonadotrophin releasing hormone agonist (leuprolide) with or without medroxy progesterone acetate in the treatment of leiomyomata uteri. *Fertilization and Sterilization* 1988; **49**: 404–9.

Friedman AJ, Daly M, Juneau-Norcross M, Fine C. Recurrence of myomas after myomectomy in women pretreated with leuprolide acetate depot or placebo. *Fertilization and Sterilization* 1992; **58**: 205–8.

Glavind K, Palvio DH, Lauritsen JG. Uterine myoma in pregnancy. *Acta Obstetrica et Gynaecologica Scandinavica* 1990; **69**: 617–19.

Moghissi KS. Hormonal therapy before surgical treatment for uterine leiomyomas. *Surgery, Gynecology, Obstetrics* 1991; **172**: 497–502.

Monaghan JM, ed. *Bonney's gynaecological surgery*, 9th edn, pp. 87–9. Baillière Tindall, London, 1986.

Monaghan JM, ed. *Bonney's gynaecological surgery*, 9th edn, pp. 91–2. Baillière Tindall, London, 1986.

Oram D, Jacobs I. Improving the prognosis in ovarian cancer. In: Studd J, ed. *Progress in obstetrics and gynaecology*, Vol. 6, pp. 412–13. Churchill Livingstone, London, 1987.

Persuad V, Arjoon PD. Uterine leiomyoma: incidence of degenerative change and a correlation of associated symptoms. *Obstetrics and Gynecology* 1970; **35**: 432.

Persuad V, Knight LP. Malignant mesenchymal tumours of the corpus uteri. *West Indian Medical Journal* 1968; **17**: 96.

Shaw RW, Soutter WP, Stanton SL. *Gynaecology*, 2nd edn, pp. 441–56. Churchill Livingstone, London, 1997.

Sutton C. Operative laparoscopy. *Current Opinions in Obstetrics and Gynecology* 1992; **4**: 430–8.

Verkauf BS. Myomectomy for fertility enhancement and preservation. *Fertilization and Sterilization* 1992; **58**: 1–15.

Warthington-Kirsch RL, Popky GL, Hutchings FL, Jr. Uterine arterial embolisation for mix of leiomyomas: quality of life assessment and clinical response to radiology. 1998; **208**: 625–9.

36.7 Endometriosis

David H. Barlow

Introduction
The nature of endometriosis
The forms of advanced disease
Symptoms of endometriosis
Clinical signs of endometriosis
Suspecting pelvic endometriosis: findings at laparotomy
Ovarian cysts
Adhesive disease
Intestinal lesions
Disease limited to the pelvis
Postoperative hormone therapy
Further reading

Introduction

Endometriosis is a relatively common problem in women of reproductive age. Since it cannot be diagnosed accurately without seeing inside the abdomen it is well nigh impossible to discover the true prevalence of endometriosis but estimates suggest that it affects at least one per cent, but less than five per cent, of women of reproductive age. In selected populations which are more likely to have diagnostic laparoscopy, such as those with chronic pelvic pain or infertility, the prevalence of endometriosis is reported to be much higher.

The management of endometriosis is normally within the jurisdiction of the gynaecologist and it is sensible to involve a gynaecologist in the case at an early stage. In gynaecological practice today endometriosis should rarely have to involve laparotomy since the surgical management of most endometriosis problems is by laparoscopic surgery. If this approach is taken then the surgeon is only likely to find him/herself confronted by significant endometriosis if it is encountered unexpectedly during surgery for an acute abdomen or as a coincidental finding during elective surgery for another problem. There are undoubtedly pitfalls for the surgeon which could lead to inappropriate and possibly excessive surgery. This may be disastrous (i) for the woman's reproductive potential and may be carried out in a situation where an adequate reproductive history (past experience and future intentions) has not been taken; or (ii) for the woman's future oestrogen production if the ovaries are removed. In gynaecological practice this latter action now demands specific informed consent and discussion if bilateral oophorectomy is a possibility. In general the gynaecological management of endometriosis is surgically conservative. Radical surgery involving hysterectomy and possibly bilateral oophorectomy is generally regarded as a late stage procedure for chronic intractable pelvic pain. This chapter will not attempt a comprehensive account of the gynaecological management of endometriosis, but will provide a limited perspective on that management and mainly provide information relevant to a surgeon who might need to be able to suspect endometriosis in advance of surgery or unexpectedly encounters it at laparotomy where it presents the surgeon with a dilemma since open surgery is not usually the treatment of choice.

The nature of endometriosis

Endometriosis is the presence, in sites outside the uterus, of glands and stroma which histologically resemble endometrium. The origin of this tissue remains uncertain but the two most likely explanations are either (i) metaplastic transformation or (ii) seeding and implantation of menstrual endometrium which has reached the peritoneal cavity by retrograde menstruation, which is a common event. The consequences of endometriosis depend on the site: the most common sites are the ovaries, the pelvic peritoneal surfaces, and the posterior aspect of the uterus. Other sites such as bladder and bowel are occasionally affected and some distant sites are rarely involved but often quoted because the cyclical bleeding at sites such as the umbilicus or lung are noteworthy.

An important unresolved current controversy surrounds the question of whether minimal and mild endometriosis usually reflects normal female physiology or is an early stage of disease usually leading on to the more significant forms of advanced disease. These more advanced forms are (i) ovarian endometriotic ‘cysts', (ii) extensive peritoneal involvement with adhesion formation, and (iii) deep uterovaginal septum disease. It is argued that localized implants on peritoneal surfaces may be transient and sporadic physiological events, with the retrograde endometrial tissue surviving for a time after retrograde menstruation but if left alone it would eventually clear. In this hypothesis the advanced forms are considered to be true pathological processes but probably distinct from minimal endometriosis.

Endometriotic implants are now recognized to have many appearances in addition to the traditional teaching which represented endometriosis as either ‘chocolate cysts' or bluish-black peritoneal nodules or ‘powder burn' lesions which may be associated with a local inflammatory reaction, fibrosis, and adhesions. These are now known to represent only part of the spectrum of peritoneal endometriosis. Biopsy studies indicate that much peritoneal endometriosis is not bluish-black in appearance: small haemorrhagic and non-haemorrhagic lesions represent what may be stages in the progression of the disease. These less classical appearances have been correlated with the rate of biopsy confirmation of the diagnosis. There is a close correlation for white opacified peritoneum (81 per cent), red flame-like lesions (81 per cent), and glandular lesions resembling endometrium (67 per cent), and reasonable correlations for subovarian adhesions (50 per cent), yellow-brown patches (47 per cent), and circular peritoneal defects (45 per cent). The range of peritoneal appearances is illustrated in Fig. 1.

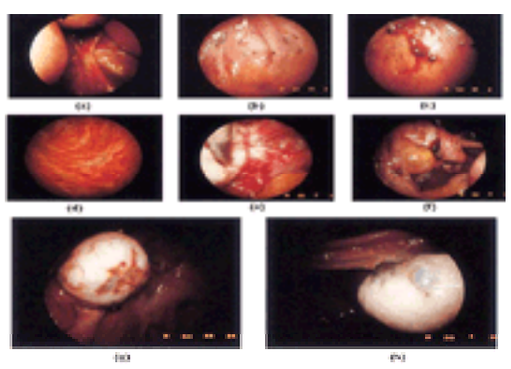


Fig. 1. The range of peritoneal and ovarian appearances of endometriosis. (a) Classical ‘powder burn' lesions on uterosacral ligaments. (b) White peritoneum with deep implants.(c) Surface haemorrhagic blebs with neovascularization. (d) Implant and extensive neovascularization. (e) Peritoneal vesicles, a peritoneal defect, and subovarian adhesions.(f) Severe pelvic adhesive disease, extensive adhesions obliterating pouch of Douglas, ovary, and fallopian tube bound down. (g) Ovarian surface implants. (h) Small endometriotic cysts. (Fig. 1(a),Fig. 1(b),Fig. 1(c),Fig. 1(d),Fig. 1(e),Fig. 1(f),Fig. 1(g) and Fig. 1(h) by courtesy of Dr Roseanne Jelley and Dr Patrick Magill.)

The issue of diagnosis has been further compounded by the evidence that biopsy samples from visually normal peritoneum often contain microscopically detectable endometrial tissue. This observation has not made an impact on clinical practice despite having been made several years ago and is of little relevance to the situations which might confront a surgeon.

The inflammation associated with endometriotic implants is probably due to the production of prostanoids and angiogenic cytokines. The peritoneal reaction can be severe, causing the formation of extensive adhesions, particularly after surgery. In such women a second laparotomy may reveal densely adherent organs, matted loops of bowel, and pelvic organs making it difficult to distinguish individual structures (Fig. 2).

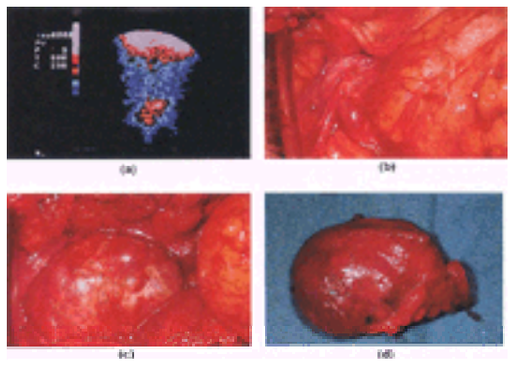


Fig. 2. A 45-year-old woman with previous hysterectomy and left salpingo-oophorectomy for endometriosis. Normal serum Ca-125. (a) Ca-125 immunoscintigraphy posterior image showing focus of high Ca-125 expression in right pelvis. (b) Laparotomy reveals dense adhesions between loops of bowel and other structures. (c) Dissection of adhesions reveals ovarian mass. (d) Removed ovary containing a 6-cm diameter ‘chocolate cyst'.

The forms of advanced disease

Extensive peritoneal involvement with adhesion formation

Some women have extensive involvement of the peritoneal surfaces with an excessive peritoneal reaction. This may take the form of clear blebbing and formation of large peritoneal vesicles which may become detached from the underlying structures. Whole surfaces may be affected and biopsy reveals the pathology to be endometriosis even though there may be no red, blue, or black lesions to be seen. Alternatively the extensive peritoneal involvement may be in the form of widespread reddish implants. The peritoneal involvement may be associated with a tendency to adhesion formation which can lead to obliteration of the pouch of Douglas and in some cases can be so extensive that the pelvic organs can be unrecognizable.

Ovarian ‘cysts'

This form of endometriosis can be confused with haemorrhagic luteal cysts which also produce ovarian cysts that contain inspissated blood which has a ‘chocolate' appearance. These luteal cysts can be several centimetres in diameter but would not involve adhesion to adjacent structures. Endometriotic cysts frequently involve adhesion to the underlying ovarian fossa and it is thought that many result from an initial adhesion of the ovary to a peritoneal lesion in the ovarian fossa which then bleeds into the adherent portion of ovary. This displaces the ovary so that an apparent ‘cyst' pushes into the ovary but the cyst contents are actually outside the ovarian cortex. In such cases any surgical attempt to mobilize the ovary will inevitably rupture the ‘cyst' (Fig. 1). Some other endometriotic cysts are thought to arise by invagination of the ovarian surface at the site of endometriotic implants. Such ‘cysts' would again be extraovarian, being outside the ovarian cortex.

Deep uterovaginal septum disease

This is also referred to as deep invasive endometriosis. This form of endometriosis has attracted much attention in recent years having been relatively ignored previously. It is the form of disease which provides the greatest surgical challenge and carries the highest risk of significant complications such as bowel or ureteric damage. These lesions are deposits of endometriotic tissue deep in the uterovaginal septum which may be invisible at laparoscopy so that a routine diagnostic laparoscopy might be reported as showing no endometriosis. Sadly these lesions can be particularly painful and tender, especially at menstruation. The origin of these deep lesions is another controversial issue. Suggestions include direct invasion from the peritoneal cavity, invagination of implants on the peritoneum in the lowest aspect of the pouch of Douglas, or that they may arise directly in the septum. It is noteworthy that these deep lesions often involve not only glands and stroma but also muscle cells. For this reason it is suggested that this is actually a form of adenomyosis.

Endometriosis detected at a general surgical laparotomy tends to be one of the more severe forms of pelvic disease or intestinal endometriosis, since the more subtle appearances are usually restricted to the pelvis and seen only on close inspection with the laparoscope.

Symptoms of endometriosis

The characteristic symptom of endometriosis is menstrual or immediately premenstrual pain. This pain may reflect stretching of tissues by a local ‘menstruation', a local effect of prostaglandins, or might relate to tissue damage and adhesion formation. There is wide variation in the extent of distress, and this correlates poorly with the extent of observed disease. Many women with endometriosis are asymptomatic: some of these have severe disease discovered either after the incidental detection of an ovarian cyst or during investigation of infertility.

Deep dyspareunia can occur, probably resulting from pressure on affected organs such as a fixed retroverted uterus, affected ovaries, uterosacral ligaments, and the rectovaginal septum, but these areas can be affected without such symptoms. In one series of more than 700 women only one in four of those who were sexually active and had disease affecting the pouch of Douglas (cul-de-sac) reported dyspareunia.

Endometriosis is not a common cause of an acute abdomen but if it occurs it could be through bowel involvement causing obstruction or by rupture of an ovarian ‘chocolate cyst'. This is an uncommon complication but is more likely than torsion of an endometriotic cyst, since such cysts tend to be locally adherent.

Clinical signs of endometriosis

Clinical examination may increase suspicion of the presence of pelvic endometriosis. An ovarian mass or masses, thickening or tenderness in the uterosacral or pouch of Douglas areas, and fixed uterine retroversion are all features which should warn the surgeon of an increased risk of endometriosis, particularly if accompanied by characteristic symptoms. Non-invasive investigative methods such as serum markers, ultrasonic imaging, or CT scanning have not proved sufficiently specific to have found a place in routine diagnosis. The most promising non-invasive investigation is magnetic resonance imaging. This is now recognized to identify lesions of 1 cm or larger, but lesions of less than 1 cm must still be regarded as equivocal using MRI. On T_2-weighted images the endometriotic implants give a bright signal. If there is an ovarian cyst, this gives a bright signal on T_2 weighting if the contents of the cyst are fatty (dermoid cyst) or contain haemosiderin (endometriotic cyst). If fat suppression is then applied the fat is no longer bright but the endometrioma remains bright, thus facilitating the characterization of lesions not possible by other non-invasive means. Despite MRI providing the most useful diagnostic information, its widespread use in decision making is limited by cost.

Suspecting pelvic endometriosis: findings at laparotomy

The different findings at laparotomy present the surgeon with different considerations. Factors to be weighed in decision making include whether the woman intends pregnancy in the future, whether she is experiencing subfertility, whether she is symptomatic, and what therapy would have been chosen for that set of circumstances if they had been detected in advance by laparoscopy.

The likely findings at laparotomy can be classified into broad categories as ovarian cysts, adhesive disease, intestinal lesions, and disease limited to the pelvis. Localized deep rectovaginal septum disease is not likely to be a prominent problem at laparotomy although it may be encountered at colorectal surgery.

Ovarian cysts

A laparotomy may be performed following rupture of an ovarian cyst. Endometriotic ‘chocolate cysts' can be several centimetres in diameter, and their rupture releases viscous inspissated blood into the peritoneal cavity. Resection of such cysts is appropriate since medical treatment tends to have little impact. The gynaecological management of such cysts would preferably be via laparoscopic surgery. An unruptured endometriotic cyst would be drained and the lining visualized and any endometriotic foci ablated by laser or diathermy. If the cyst is large then some authorities favour drainage by laparoscopy followed by hormonal suppressive therapy (gonadotrophin-releasing hormone (**GnRH**) agonist) and ablation of endometriotic foci as an interval laparoscopic procedure. Generally attempts to remove the lining prove fruitless since there is no distinct cyst wall, the cyst lining being the invaginated ovarian cortex. Sometimes the cyst lining can be stripped and this may be more likely if the ‘chocolate cyst' is actually a haemorrhagic corpus luteum cyst and not endometriosis. If there is already a ruptured cyst and it is discovered by the surgeon

at laparotomy, it is still preferable that the management is approximated to the scheme described although if the abdomen is open then even a large cyst can be dealt with by ablating endometriotic foci without the need for a two stage process. The key principle is that the approach should be as conservative as possible.

If there is certainty that the woman does not intend to have children in the future, it can be appropriate to remove the whole ovary, having ensured that the other ovary is healthy. If bilateral cysts are found it is important to preserve some ovarian tissue. The woman's specific consent to the possibility of oophorectomy should have been obtained before the operation.

Until relatively recently it was considered inappropriate to remove both ovaries yet leave the uterus intact since the retained uterus would complicate the necessary subsequent hormone replacement therapy. This argument still applies, with consent, if there is no wish for a future pregnancy. If pregnancy may be desired, the retention of the uterus without the ovaries can be considered if the woman is prepared to undergo a pregnancy by ovum donation, where she is not the genetic mother.

Adhesive disease

Pelvic adhesions are common in association with endometriosis (Fig. 1(e) and Fig. 1(f)), and may cause extensive matting of tissue including pelvic organs and/or bowel (Fig. 1(f) and Fig. 2). These adhesions should be separated, taking care to minimize tissue handling, drying of tissues, and haemorrhage. The extent to which adhesions will reform is uncertain, and there is interest in adhesion barrier products which are under evaluation. In current gynaecological practice adhesolysis is electively performed via the laparoscope when possible since laparoscopic techniques are thought to cause fewer subsequent adhesions and the lens of the scope takes the eye of the surgeon close to the operation site with substantial magnification facilitating fine dissection. The carbon dioxide laser is particularly useful for adhesolysis because the energy generated by the laser beam is dissipated within 0.1 mm; thus with experience it can be used to separate bowel from other structures.

Intestinal lesions

Surgeons are occasionally confronted at laparotomy by endometriosis affecting either the superficial serosa or the muscular layer of the bowel. Most series reporting this have been small. One series of 127 cases, which excluded superficial lesions, reported sigmoid and rectal involvement in 52 per cent, and appendix, ileal, and caecal disease in 22, 17, and 5 per cent, respectively. One-quarter of patients with disease of the sigmoid and rectum had rectovaginal septum involvement, and ileal disease always involved the last 50 cm of the ileum. Other sites are rarely affected.

Analyses of women seen by surgeons and gynaecologists reveal different viewpoints and patient populations between the two specialties (Table 1).

	Surgeons	Gynaecologists
Number of patients	36	30
Age (years)	43 (30–66)	35 (27–45)
Known genital endometriosis (%)	0	70
Bowel symptoms (%)	60	50
Barium enema (% +ve)	60	50
Preoperative hormone therapy (%)	0	30
No bowel excision (%)	14	47
Castration (%)	28	0
No genital exploration (%)	19	0
Selective excision of genital endometriosis (%)	0	30
Postoperative hormone therapy (%)	0	47

Table 1 The different clinical practice in the management of sigmoid colon and rectal endometriosis between surgeons and gynaecologists participating in a study of bowel endometriosis (adapted from Lansac et al. 1987)

Surgeons are more likely to resect bowel and to perform castration but are less consistent in checking the pelvic organs. They are also less likely to perform selective excision of genital lesions and to prescribe hormonal therapy.

A perimenstrual barium enema with double contrast can be valuable if there is clinical suspicion of intestinal endometriosis. This will reveal features such as filling defects, stricture, and external compression. Sigmoidoscopy can also be useful, revealing the endometriotic lesions and permitting biopsy. Again, MRI may provide valuable information.

At the time of laparotomy, excision of bowel lesions thought to be endometriotic is justified. There is a fairly high recurrence rate for bowel symptoms after medical treatment, as well as a risk of subsequent obstruction or malignant transformation of intestinal endometriosis. If this is thought to be an outside possibility, appropriate consent should have been obtained and bowel preparation given.

Disease limited to the pelvis

The management of endometriosis limited to the reproductive organs and pelvic peritoneum, and without significant ovarian cysts, should take the form of assessment of the extent of disease and action should be guided by a knowledge of the woman's symptoms and fertility requirements.

The first step should be to score the endometriosis as indicated by the revised American Fertility Society classification (Fig. 3 and Fig. 4). This scoring system grades the severity of disease and facilitates comparability of assessment between patients. If the surgeon does not have the scoring system available then it is valuable to record the extent of endometriosis in detail in the operation note. The nature of the pelvic organs render them particularly suitable for a freehand drawing which would include the uterus, tubes, and ovaries. Localized implants can be represented schematically as dots and extensive peritoneal involvement can be represented by shading. Adhesion between structures is often represented as some form of hatching. The important point is that, since endometriosis tends to be a chronic relapsing problem, an accurate record of this type will be valuable to subsequent clinicians who may have to manage the case. Once the assessment has been made, treatment options can be considered.

PERITONEUM	ENDOMETRIOSIS	<1 cm	1–3 cm	>3 cm
	Superficial	1	2	4
	Deep	2	4	6
OVARY	R. Superficial	1	2	4
	Deep	4	16	20
	L. Superficial	1	2	4
	Deep	4	16	20
	POSTERIOR CULDESAC OBLITERATION	Partial		Complete
		4		40
CUL-DE-SAC	ADHESIONS	<1/3 Enclosure	1/3 - 2/3 Enclosure	>2/3 Enclosure
	R. Filmy	1	2	4
	Dense	4	8	16
	L. Filmy	1	2	4
	Dense	4	8	16
TUBE	R. Filmy	1	2	4
	Dense	4*	8*	16
	L. Filmy	1	2	4
	Dense	4*	8*	16

Fig. 3. The revised American Fertility Society grading system for endometriosis.

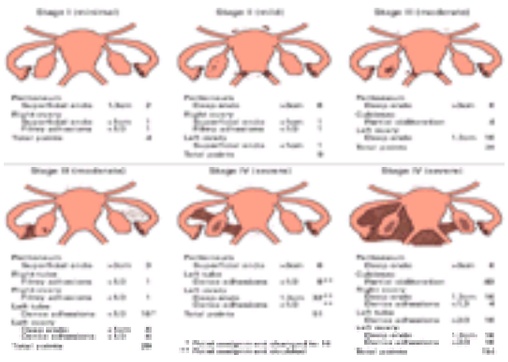


Fig. 4. Examples of endometriosis scores based on the revised American Fertility Society grading system.

If the woman is thought to have endometriosis-related pain, the lesions should be ablated using laser, endocoagulation, or electrodiathermy. Any surgeon attempting such ablation must be aware that some peritoneal lesions overlie sensitive structures such as the ureter. The most likely sites for ablation will be the surface of the uterus, the ovaries, and the pelvic peritoneum.

If the woman is thought not to have pain relating to endometriosis then, if the endometriosis is moderate or severe, it is still appropriate to ablate lesions. When the disease is minimal or mild the case for intervention is less clear, particularly in asymptomatic women who are being investigated for infertility. There is no good evidence that medical treatment improves on pregnancy rates achieved in untreated women with minimal or mild disease. There is now evidence from a large, randomized, controlled trial to suggest that ablation of implants by laser or diathermy at diagnostic laparoscopy can improve the subsequent pregnancy rate over the next 9 months in women with minimal or mild endometriosis, but there are conflicting results from other studies. At laparotomy, if lesions are seen, this evidence would appear to support the ablation of implants where it is judged that it can be done without damaging structures such as ureter or bowel.

The extent of surgical intervention must be influenced by desired fertility. Ovaries affected only by surface lesions should not be removed. Lesions can be ablated or treated medically, but unless significant ovarian cysts are present the ovary should not be damaged further since it is of major importance for continued ovarian activity and future pregnancy, whether achieved naturally or by assisted conception.

Postoperative hormone therapy

Since it is difficult to ablate all lesions surgically there is a logical case for a period of hormone treatment of perhaps 3 to 6 months, but there is very limited evidence at present on the combined application of surgical and medical management. Medical therapy should be omitted if there is a desire to conceive quickly.

If medical therapy is required, the best evidence of efficacy is for the use of danazol, gestrinone, or the GnRH-agonist drugs (buserelin, nafarelin, goserelin, or leupolide). These alternatives are equally effective in the treatment of endometriosis but have different side-effects and in general women report a greater acceptance of the hypoestrogenic side-effects of GnRH agonists than of the relatively androgenic side-effects of danazol or gestrinone. There is a growing body of evidence on the use of ‘addback' therapy to remove the hypoestrogenic consequences of GnRH-agonist therapy without losing the efficacy of the agonist. Progestogens, such as medroxyprogesterone acetate or norethisterone, are thought to be effective, but there is a less substantial body of efficacy data; however, some women will find the progestogenic side-effects more acceptable. Even less efficacy information is available for the option of prescribing a combined oral contraceptive pill, preferably continuously since this will generally abolish mentrual events and provide relief from dysmenorrhoea. Since the medical management issues can be complex, it is preferable to involve a gynaecologist in the management.

Women who undergo bilateral oophorectomy should receive oestrogen replacement therapy to counter their, otherwise, increased risk of postmenopausal osteoporosis, myocardial infarction, and symptoms of the menopause.

Anxiety that oestrogen replacement might stimulate recurrence of endometriosis has some foundation, but the few published studies suggest that the risk of recurrence is low. The consequences of an untreated premature menopause are sufficiently severe for the risk of oestrogen replacement to be worth taking. The decision may be more difficult in patients with intestinal endometriosis since the one follow-up study in which all patients had bowel endometriosis reported an 18 per cent recurrence rate. It should be remembered, however, that endometriosis may recur, whatever therapy is used. Following open conservative surgery the recurrence rate is reported at 3, 6, and 9 years as 10.5, 26.6, and 47.2 per cent, respectively.

For many women the management of their endometriosis, if it is severe, can become a test of endurance. They suffer chronic pain and dyspareunia and can have recurrent episodes of disease which may eventually so exhaust medical and surgical options (including pelvic clearance) that there is little to offer except the services of a pain clinic.

Further reading

Bellina JH, Hemmings R, Voros IJ, Ross LF. Carbon dioxide laser and electrosurgical wound study with an animal model. A comparison of tissue damage and healing patterns in peritoneal tissue. *American Journal of Obstetrics and Gynecology* 1984; **148**: 327–34. [Description of the effects of different laparoscopic treatment modalities.]

Brough RJ, O'Flynn K. Recurrent pelvic endometriosis and bilateral ureteric obstruction associated with hormone replacement therapy. *British Medical Journal* 1996; **312**: 1221–2. [Description of a serious complication of endometriosis.]

Delicata RJ, Clarke GW, Roy MK, Shaw RW, Carey PD. Presentation of endometriosis to general surgeons: a 10 year experience. *British Journal of Surgery* 1996; **83**: 711.

Donnez J, Nisolle M, Gillet N, Smets M, Bassil S, Casanas Roux F. Large ovarian endometriomas. *Human Reproduction* 1996; **11**: 641–6. [Discussion of two-stage management of ovarian endometriosis.]

Donnez J, Nisolle M, Gillerot S, Smets M, Bassil S, Casanas Roux F. Rectovaginal septum adenomyotic nodules: a series of 500 cases. *British Journal of Obstetrics and Gynaecology* 1997; **104**: 1014–18. [Large series of rectovaginal disease, discusses morphological nature of this form of endometriosis.]

Evers JL. Endometriosis does not exist; all women have endometriosis. *Human Reproduction* 1994; **9**: 2206–9. [Discussion of possibility that mild endometriosis may be physiological and not pathology.]

Feichtinger W, Kemeter P. Pregnancy after total ovariectomy achieved by ovum donation. *Lancet* 1985; **ii**: 722–3. [Justification for conservation of uterus at bilateral oophorectomy.[

Gordon R, Evers K, Kressel, Laufer I, Herlinger H. Double contrast enema in pelvic endometriosis. *American Journal of Roentgenology* 1982; **138**: 549–52.

Goulbourne JA, MacLeod DA. Endometriosis presenting as an acute abdomen. *British Journal of Clinical Practice* 1986; **40**: 124–6.

Gray LA. Endometriosis of the bowel: role of bowel resection, superficial excision, and oophorectomy in treatment. *Annals of Surgery* 1973; **177**: 580–7.

Grodstein F *et al.* Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *New England Journal of Medicine* 1996; **335**: 453–61. [Review of the relevance of hormone replacement therapy to cardiovascular disease risk.]

Groff T. Classifications. In: Schenken RS, ed., *Endometriosis: contemporary concepts in clinical management*, pp. 145–67. Lippincott, Philadelphia, 1989. [Discussions of the issues involved in the classification system of endometriosis.]

Hornstein MD, Hemmings R, Yuzpe AA, Heinrichs WL. Use of nafarelin versus placebo after reductive laparoscopic surgery for endometriosis. *Fertility and Sterility* 1997; **68**: 860–4. [Evidence on the use of postoperative medical therapy.]

Jansen RPS, Russell P. Nonpigmented endometriosis: clinical, laparoscopic, and pathological definition. *American Journal of Obstetrics and Gynecology* 1986; **155**: 1154–9. [Description of the variety of non-pigmented lesions in endometriosis and histological correlations.]

Karthaus M, Prahst A, Geissler RG, Hertenstein B, Degenhardt F, Ganser A. Acute abdomen due to endometriosis as a diagnostic and therapeutic challenge in the treatment of acute myelocytic leukemia. *Annals of Hematology* 1997; **74**: 29–31.

Kennedy SH, Williams IA, Brodribb J, Barlow DH, Shaw RS. A comparison of nafarelin acetate and danazol in the treatment of endometriosis. *Fertility and Sterility* 1990; 53:924–8. [Description of efficacy and side-effects of medical treatments.]

Koninckx PR, Martin DC. Deep endometriosis: a consequence of infiltration or retraction or possibly adenomyosis externa. *Fertility and Sterility* 1992; **58**: 924–8. [Discussion of the morphology of deep infiltrative endometriosis.]

Koninckx PR, Muyldermans M, Moerman P, Meuleman C, Deprest J, Cornillie F. CA 125 concentrations in ovarian ‘chocolate' cyst fluid can differentiate an endometriotic cyst from a cystic corpus luteum. *Human Reproduction* 1992; **7**: 1314–17.

Lansac J, Pierre F, Letessier E. Digestive endometriosis: reults of a multicentre investigation. *Contributions to Gynecology and Obstetrics* 1987; **16**: 192–204.

Macafee C, Greer H. Intestinal endometriosis. A report of 29 cases and a survey of the literature. *Journal of Obstetrics and Gynaecology of the British Empire* 1960; **67**: 549–55.

Murphy AA, Green WR, Bobbie D, dela Cruz, ZC, Rock JA. Unsuspected endometriosis documented by scanning electron microscopy in visually normal peritoneum. *Fertility and Sterility* 1986; **46**: 522–4. [Evidence for the presence of microscopic endometriosis in normal-looking peritoneum.]

Nisolle N, Donnez J. Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. *Fertility and Sterility* 1997; **68**: 585–96. [Discussion of the nature of different forms of endometriosis.]

O'Connor DT. *Endometriosis*, p. 71. Churchill Livingstone, Edinburgh, 1987. [Monograph on endometriosis which details a large Australian case series.]

Redwine DB. Age-related evolution in color appearance of endometriosis. *Fertility and Sterility* 1987; **48**: 1062–3. [Correlation of different peritoneal lesions with age suggesting a possible pattern of progression appearances of lesions.]

Redwine DB. Conservative laparoscopic excision of endometriosis by sharp dissection: life table analysis of reoperation and persistent or recurrent disease. *Fertility and Sterility* 1991; **56**: 628–34. [A description of disease recurrence risks.]

Revised American Fertility Society classification of endometriosis. *Fertility and Sterility* 1985; **43**: 351–2.

Wheeler JM, Malinak LR. Recurrent endometriosis. *Contributions to Gynecology and Obstetrics* 1987; **16**: 13–21. [A description of disease recurrence.]

36.8.1 Uterine malignancy

C. O. Granai, Walter H. Gajewski and Bruno Arena

Introduction
Diagnosis
Preoperative assessment
Pathology
Patterns of spread
Staging
Risk stratification
Management
Historic aspects
Current procedure
Postoperative treatments
Special cases
Further reading

Introduction

Cancer of the endometrium is the most common gynecologic pelvic malignancy in developed nations. Within the United States there are an estimated 36 000 new cases per year. The mean age of diagnosis is the late fifth decade in menopause. Since most of these cancers present with vaginal bleeding, any postmenopausal bleeding should alert the clinician to the possibility of endometrial cancer. Epidemiologic risk factors include obesity, nulliparity, diet, advancing age, and use of unopposed estrogen. Oral contraception and, surprisingly, smoking, correlate negatively with endometrial cancer.

Diagnosis

The physical examination is not directly helpful in establishing the diagnosis. Because of the occult location of endometrial cancer, the pelvic examination and Pap smear are usually normal. Today, diagnosis is generally made by endometrial biopsy, typically indicated to evaluate abnormal bleeding or any postmenopausal bleeding. The least complicated method of endometrial biopsy is recommended. The results of simple, well-tolerated, in-office sampling techniques correlate well with those obtained by the formerly standard dilatation and curettage under anesthesia, which still has a back-up role when the diagnosis remains unclear after more cost-effective methods of biopsy have been tried. Multiple studies have sought to correlate ultrasonographic evaluation of the thickness of the endometrial strip with the pathology in postmenopausal women. Although the risk of cancer is rare when the endometrial strip is 5 mm or less, we believe the standard for diagnosis remains the endometrial biopsy.

Preoperative assessment

The endometrial biopsy establishes the diagnosis of malignancy. Endocervical curettage can identify occult endocervical extension of the cancer. Histology, grade, and endocervical involvement, as determined by endometrial biopsy, can affect choice of treatment. Few other preoperative tests in clinically early endometrial cancer are fruitful, but the chest radiograph and often a computed tomographic scan are frequently done before surgery, looking for distant disease. CA125, the serum tumor marker for ovarian cancer, is occasionally elevated in endometrial cancer; changes in any such increases can be used in monitoring treatment response. Without relevant symptoms, cystoscopy, sigmoidoscopy, and pelvic ultrasonography are invariably negative and hence can be avoided. In contrast, magnetic resonance imaging (**MRI**) can determine the extent and depth of uterine involvement as well as the pelvic distribution of the disease. Nevertheless, MRI need not be done routinely, as it is costly and rarely alters the choice of treatment. A notable exception would be where the preservation of fertility is concerned. Premenopausal women can develop endometrial cancer, typically as a result of polycystic ovarian disease or other conditions of estrogen excess. Occasionally such patients are candidates for hormonal therapy (i.e., high-dose pro-gesterone) in an attempt to conserve fertility. This experimental and risky approach is only rational if MRI demonstrates no myometrial invasion. Proper counselling of the patient is imperative in anycase.

Pathology

Endometrioid adenocarcinoma of the endometrium is by far the most prevalent malignant histological type in the uterus; together with the other adenocarcinomas, adenosquamous, papillary serous, and clear-cell cancers, 90 per cent of uterine cancers are accounted for. Differentiating between these histological subtypes is important as each has its own biologic behavior and outcome. For example, the 5-year survival for the typical early-stage endometrioid cancer can be 90 per cent. This excellent survival contrasts sharply with that of around 40 and 45 per cent reported with papillary serous and clear-cell carcinomas, respectively. Additionally, and unlike the endometrioid type, papillary serous carcinomas of the endometrium behave biologically more like epithelial ovarian cancer (i.e., spreading and recurring diffusely on peritoneal surfaces).

Patterns of spread

The biologic pattern of spread of the typical advancing endometrioid carcinoma is, first, expansion within the endometrial cavity, then invasion of the myometrium, followed by invasion of the vascular/lymphatic space, which allows metastatic spread to the pelvic and, in turn, the para-aortic lymph nodes. The para-aortic nodes were thought to be the initial/preferential site of nodal metastases for endometrial cancer, but this clinical axiom has been disproved by extensive surgical–pathologic studies conducted by the Gynecologic Oncology Group.

Another important site for lymphatic metastasis is the vagina. Although tumor implantation and direct extension to the vagina can also occur, most endometrial metastases to the vagina are lymphatic. This observation has major therapeutic implications, since vaginal metastases/recurrence usually represent only the ‘tip of the iceberg' rather than an isolated recurrence. As a result, the present International Federation of Gynaecology and Obstetrics staging system has allotted vaginal metastases a high stage (IIIb). In at least one-third of cases, vaginal disease is a ‘marker' for simultaneous, albeit occult, systemic disease. This reality accounts for the high failure of salvage treatment if it is directed only at obvious local or regional endometrial recurrences.

Further progression of the tumor occurs directly within the pelvis and exfoliated malignant cells can implant throughout the peritoneal cavity. Vascular metastases to the lung, liver, bone, and brain do occur, but late in the disease process. ‘Involvement' of the endometrium and ovaries with simultaneous ‘endometrioid' cancers is not uncommon. An incidence of simultaneous endometrial and ovarian cancer of endometrioid type as high as a 10 per cent has been reported. These findings may represent either synchronous/separate primaries or metastatic disease, from one site to the other, which is an important distinction to make with great implications for prognosis and treatment.

Staging

By convention (Federation of Gynaecology and Obstetrics), the staging of endometrial cancer is now done by surgery (Table 1). The change from the former clinical staging system to a surgical approach brings a more comprehensive understanding of the true disease distribution. The extent of disease was frequently misjudged when only preoperative clinical findings were used to determine the stage. For example, the Gynecologic Oncology Group found that clinical staging underestimated the true (surgical–pathological) stage in at least 10 per cent of cases. The greater accuracy of surgical staging realistically identifies and classifies patients by existing risk factors and allows a better statistical comparison of treatments. Federation of Gynaecology and Obstetrics staging of endometrial cancer considers not only disease distribution, but further subcategorizes the disease based on the histologic grade/degree of differentiation of the tumor. Since abnormal bleeding is an early warning sign of endometrial cancer, most women are diagnosed with early/low-stage cancers. Indeed, the majority will have no evidence of disease beyond the uterus, 75per cent being stage I and 10 per cent stage II. The remaining 15 per cent have advanced-stage disease.

Stage IA	G1G3	Tumor limited to endometrium
Stage IB	G1G3	Invasion to less than one-half the myometrium
Stage IC	G1G3	Invasion to more than one-half the myometrium
Stage IIA	G1G3	Endocervical glandular involvement only
Stage IIB	G1G3	Cervical stromal invasion
Stage IIIA	G1G3	Tumor invades serosa and/or adnexa, and/or positive peritoneal cytology
Stage IIIB	G1G3	Vaginal metastases
Stage IIIC	G1G3	Metastases to pelvic and/or para-aortic lymph nodes
Stage IVA	G1G3	Tumor invasion of bladder and/or bowel mucosa
Stage IVB		Distant metastases including intra-abdominal and/or inguinal lymph nodes

G, grade

Table 1 Federation of Gynaecology and Obstetrics staging for carcinoma of the corpus uteri

Risk stratification

By using the information learned from surgical staging and other prognostic variables, patients can be rationally divided into prognostic groups. Data from the Gynecologic Oncology Group's landmark prospective study of surgical–pathologic staging for endometrial cancer have simplified and validated the list of prognostic factors. Prognosis in stage I endometrial cancer is predicted by tumor grade, depth of myometrial invasion, invasion of the vascular/lymphatic space, age, hormonal-receptor status, and tumor histology (clear-cell and papillary serous carcinomas have a poorer prognosis).

In the Gynecologic Oncology Group study, patients with well-differentiated adenocarcinomas had only a 3 per cent risk of pelvic lymph-node involvement, compared with 18 per cent for those with poorly differentiated lesions. If no myometrial invasion was present, the risk of nodal metastases was only 1 per cent. This risk increased to 25 per cent if the carcinoma invaded the myometrium to the outer third. Patients with no invasion of the vascular/lymphatic space had a 7 per cent risk of pelvic lymph-node metastases, whereas those who had such invasion had a 27 per cent risk. Moreover, invasion of the vascular/lymphatic space has been associated with an increase in the rate of recurrence from 10 to 44 per cent, and may be a more important risk factor than myometrial invasion.

Information from DNA-ploidy and S-phase fraction analysis may be useful in planning treatment, but the practical, clinical importance of this remains undefined.

The prognostic value of malignant cytology (washings) from the abdomen and pelvis has been controversial. Although uncommon as the sole risk factor present in a case of endometrial cancer, data suggest that positive pelvic washings are an important independent predictor of risk. Indeed, in the Gynecologic Oncology Group series, malignant cytology alone was associated with a 5-year survival of 56 per cent. In recognition of the poorer outcome and risk for diffuse recurrence implied by the finding of positive washings, the Federation of Gynaecology and Obstetrics has raised the stage of patients with malignant cytology from I to IIIa (Fig. 1).



Fig. 1. Management plan for stage I endometrial cancer as practiced at the Women and Infants Hospital. TAH/BSO, total abdominal hysterectomy/bilateral salpingo-oophorectomy.

Another observation relevant to the surgical management of endometrial cancer is that, contrary to former belief, the para-aortic lymph nodes (stage IIIc) are not the prime site of nodal metastases. Rather, the pelvic lymph nodes are more commonly involved. Overall, 9 per cent of patients with stage I endometrial cancer had metastases in pelvic lymph nodes, compared to 5 per cent with para-aortic disease. The risk of positive lymph nodes depends on the presence of other prognostic factors (e.g. grade, vascular/lymphatic invasion). Up to 61 per cent of stage I patients had positive pelvic lymph nodes and up to 30 per cent had involvement of para-aortic lymph nodes. However, the para-aortic nodes themselves are rarely the exclusive sight of nodal metastases, occurring in only 2 per cent of cases overall. Thus, in the absence of other major risk factors identified before or during surgery (e.g. high tumor grade plus deep invasion; intraperitoneal disease) examining clinically normal para-aortic nodes by biopsy carries an unacceptably high risk of morbidity.

As suggested, the maximal clinical value of prognostic factors comes when they are combined and used proactively during all phases of the intervention. For example, the tumor grade and endocervical involvement can be ascertained before laparotomy, even though they may be altered by findings at surgery. When ‘high-risk' factors are identified preoperatively, such as a poorly differentiated cancer or cervical extension, a need for more complex surgery and staging can be anticipated and a referral to a gynecologic oncologist made.

The proactive use of prognostic factors is even more profound at laparotomy, where new, meaningful information is uncovered that shapes the progressing operation. For example, despite clinical appearances to the contrary, unanticipated gross extrauterine disease can still be encountered at surgery. Extrauterine tumor is a major poor-risk factor associated with a high probability of nodal metastases (pelvic 33 per cent; para-aortic 8 per cent). Finding extrauterine tumor then justifies a more extensive staging operation than usual, possibly including the sampling of lymph nodes from both the pelvic and para-aortic regions. Other major risk factors can also be determined intraoperatively. Specifically, the depth of myometrial invasion and endocervical involvement can be quickly identified by frozen section. Surgery is then tailored accordingly, using the combined knowledge of both the pre- and intraoperative risk factors. After surgery, more detailed pathologic analysis becomes the basis for the final risk stratification, which determines the need for adjuvant therapy.

Management

Historic aspects

Traditionally, endometrial cancer has been managed with radiation followed a few weeks later by hysterectomy. This sequence was credited with a 70 to 90 per cent 5-year survival for patients with early disease. It is understandable, given the good survival record relative to that for most other cancers, that the time-honored approach (radiation followed by surgery) went unchallenged.

However, it eventually became obvious that the same duration of survival could be achieved when surgery, not radiation, was the primary/initial management step. Moreover, by not giving radiation as the primary treatment, the true pathologic findings were not obscured by radiation damage. Equally effective and as well tolerated by patients after surgery as it was preoperatively, radiation is now employed selectively, based on the actual disease distribution.

Current procedure

The first step in the management of most endometrial cancers is surgery. Exploratory laparotomy, cytological washings, extrafascial total hysterectomy, bilateral salpingo-oophorectomy, intraoperative assessment of myometrial invasion and disease distribution, and sampling of lymph nodes, if indicated, are usually done. The exact intervention to be used is adjusted according to patient's age, tumor histologic cell type and grade, and intraoperative findings. For example, finding gross extrauterine disease in the upper abdomen may diminish the effectiveness of surgery. In that case, a limited palliative operation, including hysterectomy, should be considered, preventing the otherwise inevitable extensive bleeding that results from untreated endometrial cancer. Recently, a role for debulking surgery in the setting

of advanced disease has been suggested by small studies in showing improved survival among patients where optimal debulking (<1 cm) was achieved.

For those patients who are deemed inoperable for medical reasons, primary radiation, including external beam and intracavitary implants, is an alternative to surgery. The cure rate, however, even with early disease, is significantly less than that achieved by hysterectomy alone. Fortunately, despite the first (stereotypic) impression to the contrary, most patients, including the very elderly, tolerate surgery well.

In the program in women's oncology, Women and Infants Hospital/Brown University, all patients with gynecologic cancer are presented prospectively to the multidisciplinary tumor board. The typical treatment protocols recommended by the board for the primary and secondary management of endometrial cancer are shown in Fig. 2, Fig. 3, Fig. 4 and Fig. 5.

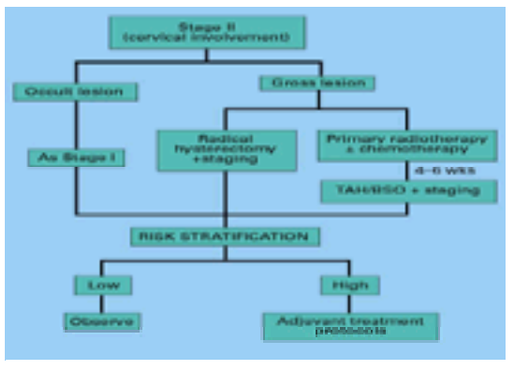


Fig. 2. Management plan for stage II endometrial cancer as practised at the Women and Infants Hospital. TAH/BSO, total abdominal hysterectomy/bilateral salpingo-oophorectomy.

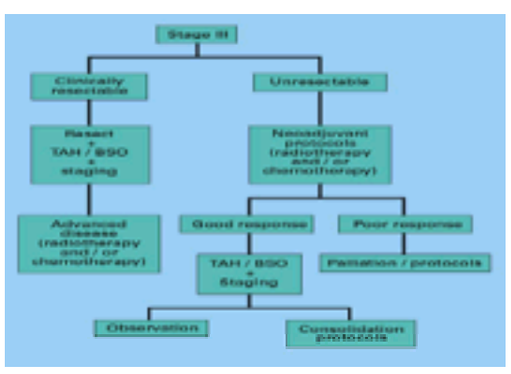


Fig. 3. Management plan for stage III endometrial cancer as practised at the Women and Infants Hospital. TAH/BSO, total abdominal hysterectomy/bilateral salpingo-oophorectomy.

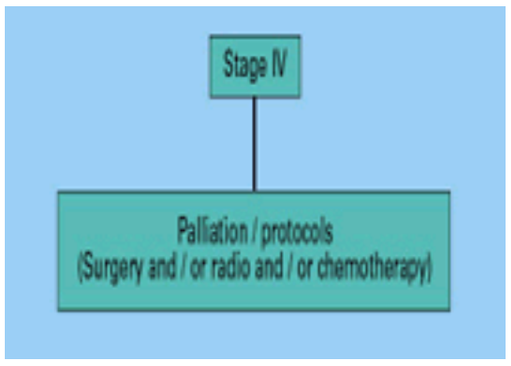


Fig. 4. Management plan for stage IV endometrial cancer as practised at the Women and Infants Hospital.

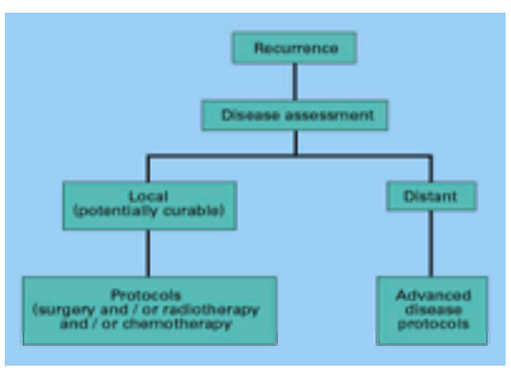


Fig. 5. Management plan for recurrent disease in endometrial cancer as practised at the Women and Infants Hospital.

Postoperative treatments

Adjuvant

Patients with a poor prognosis can be accurately identified by surgical staging (i.e. after surgery). Pelvic radiation is the most common adjuvant treatment given to those at a high risk for local recurrence. Currently, at the Women and Infants Hospital/Brown University, standard treatment protocols recommend, with exceptions, postoperative radiation therapy for patients with ‘stage I' disease and deep myometrial invasion or vascular/lymphatic invasion or extension to the endocervix, assuming that in these cases the lymph nodes are disease free. Alternatively, when the lymph nodes are affected, extrauterine disease is identified, or the washings contain malignant cells, the strategic value of radiation is less clear and entry of the patient into experimental protocols for the treatment of advanced disease is recommended. Systemic treatment seems called for, possibly in conjunction with pelvic (and occasionally extended-field) radiation.

Although radiation has been used for a long time as a major treatment for endometrial cancer, surprisingly few data suggest that pelvic radiation improves survival beyond that achieved by surgery alone. Anecdotal and smaller retrospective studies have suggested improved regional control and survival is sometimes achieved in subsets of patients given adjuvant radiation. However, larger studies have not confirmed the survival advantage and prospective studies are few and of limited scope (i.e. exclude high risk patients).

Local recurrences, particularly in the vagina, can be reduced by radiation given either as an external beam to the whole pelvis or as a vaginal cylinder, albeit at a possible cost of radiation-induced morbidity including major and minor gastrointestinal and genitourinary complications, and diminished sexual function. A randomized, prospective study conducted by the Gynecologic Oncology Group to determine whether radiation therapy improves outcome in intermediate-risk patients. Though still needing longer follow-up, pelvic radiation decreased local regional recurrence but did not as yet show beneficial effect on survival. At least the patients with a good prognosis can be spared the potential problems of radiation by using primary surgery and stratifying for risk postoperatively.

Advanced disease

Effective treatment for advanced endometrial cancer is elusive. Pelvic radiation therapy is often tried with high-risk or high-stage disease, but with generally discouraging results. This problem is understandable, since treating a diffuse malignant process using a local/regional therapy is bound to fail. Research studies of protocols to improve management are required to develop appropriate systemic treatments.

Although logical, and once highly recommended, systemic hormonal therapy with high-dose progesterone has proved, in carefully controlled studies, to be largely ineffective. Occasionally, however, responses may occur, particularly with grade 1 lesions (most likely to have estrogen and progesterone receptors) at sites outside the radiated field. Similarly, there are reports of responses to chemotherapy but larger studies have yet to identify the role, if any, for the routine use of currently available chemotherapy. At present, in spite of the lack of convincing evidence, standard systemic management for advanced endometrial cancer combines variations of Adriamycin, cisplatinum, cytoxan, ifosfamide, VP-16, and Taxol for their short-term benefit which can be quite dramatic. Combining local/regional radiation plus chemotherapy is a strategy also being explored. In general, entry into experimental protocol should be considered as the primary recommendation.

Recurrent disease

Recurrent endometrial cancer has a surprisingly poor prognosis. The true scope of the recurrence is often belied by the seemingly limited extent of the clinical disease. A common example would be finding an almost innocuous nodule of recurrent endometrial carcinoma in the upper vagina. Given that the mode of metastasis to the vagina is usually lymphatic, there is often occult, systemic/diffuse disease being disseminated simultaneously by the same means but as yet unidentified clinically and radiographically,. Thus, despite their common application in this situation, local/regional treatments directed at the obvious site of recurrence alone have been unsuccessful. There have been reports, nonetheless, of occasional salvage cures from using radiation (if not previously given) and still rarer successes with salvage surgery (e.g. pelvic exenteration) if the recurrence is truly a local phenomenon. Unfortunately, most patients have widespread disease and thus are candidates for palliative treatment only, often using systemic chemotherapeutic protocols. The most active chemotherapeutic regimen presently in use is cisplatin and Adriamycin. Newer agents such as Taxol and Doxil (liposomal adriamycin) are the subjects of continuing clinical investigation.

Surgically restaging patients assumed to have early local recurrence can play an important part in the decision on optimal treatment. Further, where possible, complete surgical resection of the recurrent tumor enhances the salvage therapy.

In general, for advanced/recurrent endometrial cancer, surgery and radiation are of limited value for local control and, although desperately needed, systemic therapies are merely briefly palliative.

Special cases

As with all malignancies, a variety of less common uterine neoplasms exist. The oncologist/surgeon should be aware of these ‘special cases' and their need to be managed accordingly.

Papillary serous and clear-cell cancers

Both are a relatively uncommon malignancies in the endometrium. However, papillary serous carcinoma as an endometrial cancer is noteworthy because of its biologic behavior. The patterns of spread and prognostic factors characteristic of the typical endometrioid adenocarcinoma do not apply to these subtypes: on the contrary, both behave biologically more like epithelial ovarian cancer. In addition to their histologic similarities to their ovarian counterparts, these endometrial cancers have a propensity to spread widely and diffusely into the abdominal cavity on peritoneal/serosal surfaces. Virulent behavior can be expected from both subtypes, even in the absence of myometrial invasion or of other poor prognostic factors useful in predicting risk in the stereotypic endometrial adenocarcinoma. Moreover, aggressive behavior can occur with only a small, seemingly insignificant admixture of the histological subtype, often overlooked against a dominant background of simultaneous endometrioid carcinoma.

After surgery, endometrial papillary serous carcinoma is difficult to manage, even with combination chemotherapy using cisplatin and Taxol, which is effective in papillary serous cancer of the ovary. Despite the limited success to date, systemic therapy seems necessary and best, given the biologic behavior of the endometrial variant of papillary serous cancer.

Other uterine malignancies

About 5 per cent of uterine tumours are sarcomas, emanating from either the myometrium or the endometrial stroma. Although not generally realized, the mixed mullerian mesodermal tumor (formerly known as carcinosarcoma) and not the leiomyosarcoma is the most common connective-tissue malignancy in the uterus. Next in frequency are leiomyosarcoma (very rare in the general population compared with their benign counterparts, the uterine leiomyomas/fibroids) followed by endometrial stromal sarcomas. Each of these malignancies has unique clinical and pathologic features.

Mixed mullerian mesodermal tumours of the uterus have been subcategorized by pathologists in many ways; from a practical, clinical viewpoint there is little difference among the subtypes. In general, the primary treatment for mixed mullerian mesodermal tumours and leiomyosarcomas is surgery (total abdominal hysterectomy or bilateral salpingo-oophorectomy) and staging, followed by adjuvant treatment. Both lesions usually recur locally/regionally and distantly. Thus, radiation to the pelvis as adjuvant treatment may reduce local recurrence, but has little, if any, impact on overall survival. Experimental protocols using systemic treatments for these sarcomas are being investigated. To date, adjuvant chemotherapy, which seems intuitively important, has not been shown to reduce the risk of recurrence or improve survival. In advanced or recurrent disease, ifosphamide or cisplatin show the best, although limited, response rates (20–25 per cent)

Many other tumors that can also occur in the uterine connective tissues (e.g. endometrial stromal tumours) are part of a spectrum of neoplasia with benign to malignant variants. A detailed analysis of any connective-tissue tumor to define its position in the continuum of connective-tissue neoplasia, and hence what biologic behavior is to be anticipated, is essential.

The presence of rare uterine tumors of this type is often unsuspected prior to, or even during, laparotomy. Fortunately, surgery alone is often curative, particularly for low-grade, early-stage lesions. If the tumor recurs, another local aggressive resection or even resection of distant disease can be appropriate, depending on the histological features. More often, however, advanced or recurrent connective neoplasia is treated systemically. Some of these neoplasms respond dramatically to hormonal ablation or to high-dose progesterone therapy. Others occasionally respond to combination chemotherapy. Unfortunately, the majority of high-grade sarcomas that cannot be cured surgically only respond to systemic treatment for a short time.

Preserving fertility

Occasionally, a young woman with a very well-differentiated and limited adenocarcinoma of the endometrium will want to preserve her ability to conceive. Typically, such patients have a medical history and physical findings consistent with chronic anovulation. After appraisal of the potentially severe risks involved, these special patients may elect for an unconventional management.

Barring clinical evidence of advanced disease or a suggestion of myometrial invasion on MRI, patients can be offered intense primary hormonal management using high-dose progesterone instead of hysterectomy. The anticipated response to medical management is amenorrhea. If amenorrhea occurs, treatment is continued for 3 to 4 months. A thorough dilatation and curettage with hysteroscopy must follow the limited trial of medical management (or sooner if the clinical behavior does not conform to expectations).

If the endometrial cancer is ‘resolved' at the completion of hormonal therapy, assistance in achieving pregnancy with fertility drugs can be given before the need for hysterectomy develops.

Cancer of the fallopian tubes

Although the fallopian tubes have a direct physical relation to the uterus, the rare tubal cancer is biologically and behaviorally more analogous to epithelial ovarian cancers. Malignancy of the fallopian tubes is therefore staged and managed like that of the ovary rather than of the uterus.

The diagnosis of a tubal malignancy is usually fortuitous or comes about retrospectively after surgery is performed for what is thought to be an ovarian mass. Preoperative symptoms are sometimes present, the classic one being a profuse, intermittent, watery vaginal discharge. Pain and an abnormal Pap smear, suggesting adenocarcinoma but not otherwise explained by routine testing (colposcopy, endocervical curettage, and dilatation and curettage), can also lead to the diagnosis of cancer of the fallopian tubes. In general, although such cancers behave clinically and pathologically like ovarian epithelial cancers, their overall prognosis seems worse.

Further reading

Aalders J, Abeler V, Kolstad P, Onsrud M. Postoperative external irradiation and prognostic parameters in stage I endometrial carcinoma: clinical and histopathologic study of 540 patients. *Obstetrics and Gynecology* 1980; **56**: 419.

Ali S, Wells M. Mixed mullerian tumors of uterine corpus: a review. *International Journal of Gynecologic Cancer* 1993; **3**: 1.

American College of Obstetricians and Gynecologists. *Carcinoma of the endometrium* (ACOG Technical Bulletin no. 162).

Ball HG *et al.* A phase II trial of Taxol in advanced and recurrent adenocarcinoma of the endometrium: a Gynecologic Oncology Group study. *Gynecologic Oncology* 1995; **56**: 120.

Burke TW, Morris M. Surgery for malignant tumors of the uterine corpus. In: Gerschenson DM, DeCherney AH, Curry SL, ed. *Operative gynecology*, pp. 371–3. Saunders, Philadelphia, 1993.

Burke TW *et al.* Prospective treatment of advanced or recurrent endometrial carcinoma with cisplatin, doxorubicin, and cyclophosphamide. *Gynecologic Oncology* 1991; **40**: 264.

Carcangiu ML, Chambers JT. Uterine papillary serous carcinoma: a study on 108 cases with emphasis on the prognostic significance of associated endometrioid carcinoma, absence of invasion, and concomitant ovarian carcinoma. *Gynecologic Oncology* 1992; **47**: 298.

Chambers JT *et al.* Uterine papillary serous carcinoma. *Obstetrics and Gynecology* 1987; **69**: 109.

Creasman WT. Endometrial cancer: incidence, prognostic factors, diagnosis, and treatment. *Seminars in Oncology* 1997; **24**(1) (Suppl. 1): S1-140–50.

Creasman WT *et al.* Surgical pathologic spread patterns of endometrial cancer. *Cancer Gynecology and Obstetrics* 1987; **21**: 35.

Davies JL *et al.* A review of the risk factors for endometrial carcinoma. *Obstetrics and Gynecology Survey* 1981; **36**: 107.

Granai CO, ed. Cancers unique to women. *Surgical Oncology Clinics of North America* 1998; **7**(2).

Granai CO, Steinhoff M, Gajewski W, Arena B. DNA ploidy and S-phase fraction in stage I grade I endometrial adenocarcinomas. *Gynaecopathology* 1998; **3**: 3–5.

Grimes DA. Diagnostic dilation and curettage: a reappraisal. *American Journal of Obstetrics and Gynecology* 1982; **142**: 1.

Gurney H, Murphy D, Crowther D. The management of primary fallopian tube carcinoma. *British Journal of Obstetrics and Gynaecology* 1990; **97**: 822.

Hacker NF. Uterine cancer. In: Berek JS, Hacker NF, ed. *Practical gynecological oncology*, pp. 285–326. Williams and Wilkins, Baltimore, 1989.

International Federation of Gynecology and Obstetrics. Annual report on the results of treatment in gynecologic cancer. *International Journal of Gynecology and Obstetrics* 1989; **28**: 189–90.

Karlsson B, Granberg S, Wikland M, Ylostalo P, Torvid K, Marsal K, Valentin L. Transvaginal ultrasonography of the endometrium in women with postmenopausal bleeding—a Nordic multicenter study. *American Journal of Obstetrics and Gynecology* 1995; **172**(5): 1488–94.

Larson B, Silfversward C, Nilsson B, Petterson F. Prognostic factors in uterine leiomyosarcoma. *Acta Oncologica* 1990; **29**: 185.

Levi F *et al.* Chemotherapy of advanced ovarian cancer with 4'–*O*-tetrahydropyranyl doxorubicin and cisplatin: a randomized phase II trial with an evaluation of circadian timing and dose-intensity. *Journal of Clinical Oncology* 1990; **8**: 705.

Meerwaldt JH *et al.* Endometrial adenocarcinoma, adjuvant radiotherapy tailored to prognostic factors. *International Journal of Radiation, Oncology, Biology, Physics* 1989; **18**: 299.

Mittal KR, Schwartz PE, Barwick KW. Architectural (Federation of Gynaecology and Obstetrics) grading, nuclear grading, and other prognostic indicators in stage I endometrial adenocarcinoma with identification of high-risk and low-risk groups. *Cancer* 1988; **61**: 538.

Morris PJ, Malt RA, ed. *Oxford textbook of surgery*, Vol. 12, pp. 1438–43. Oxford Medical, 1994.

Morrow CP *et al.* Relationship between surgical–pathological risk factors and outcome in clinical stage I and II carcinoma of the endometrium: a GOG study. *Gynecologic Oncology* 1991; **40**: 55.

Muntz HG *et al.* Primary adenocarcinoma of the fallopian tube. *European Journal of Gynaecological Oncology* 1989; **10**: 239.

Onsrud M, Kolstad P, Normann T. Postoperative external pelvic irradiation in carcinoma of the corpus stage I: a controlled clinical trial. *Gynecologic Oncology* 1976; **4**: 22.

Parazzini F, La Vecchia C, Bocciolone L, Franceschi S. The epidemiology of endometrial cancer. *Gynecologic Oncology* 1991; **41**: 1.

Phillips GL, Prem KA, Adcock LL, Twiggs LB. Vaginal recurrence of adenocarcinoma of the endometrium. *Gynecologic Oncology* 1982; **13**: 323.

Piver MS. Progesterone therapy for malignant peritoneal cytology: surgical stage I endometrial adenocarcinoma. *Seminars in Oncology* 1988; **15**: 50–2.

Piver MS, Recio FO, Baker TR, Hempling RE. A prospective trial of progesterone therapy for malignant peritoneal cytology in patients with endometrial carcinoma. *Gynecologic Oncology* 1992; **47**: 373–6.

Randall ME, Spirtos NM, Dvoretsky P. Whole abdominal radiotherapy versus combination chemotherapy with doxorubicin and cisplatin in advanced endometrial carcinoma (phase III): Gynecologic Oncology Group study no. 122. *Journal of the National Cancer Institute Monographs* 1995; **19**: 13–5

Rubin SC, ed. *Chemotherapy of gynecologic cancers* (Society of Gynecologic Oncologists Handbook). Lippincott-Raven, New York, 1996.

SGO Medical Practice and Ethics Committee. *Clinical practice guidelines: management of gynecologic cancers*. October 1996. [Suggestions about the clinical concerns and management of endometrial cancers.]

Shimm DS *et al.* Management of high-grade stage I adenocarcinoma of the endometrium: hysterectomy following low dose external beam pelvic irradiation. *Gynecologic Oncology* 1986; **23**: 183.

Silverberg SG *et al.* Carcinosarcoma (malignant mixed mesodermal tumor) of the uterus. A Gynecologic Oncology Group pathologic study of 203 cases. *International Journal of Gynecologic Pathology* 1990; **9**: 1.

Sivridis E, Buckley CH, Fox H. The prognostic significance of lymphatic vascular space invasion in endometrial adenocarcinoma. *British Journal of Obstetrics and Gynaecology* 1987; **94**: 991.

Society of Gynecologic Oncologists. Practice guidelines: uterine corpus—sarcomas. *Oncology* 1998; **12**(2).

Thigpen JT *et al.* A randomized comparison of doxorubicin alone versus doxorubicin plus cyclophosphamide in the management of advanced or recurrent endometrial carcinoma: a Gynecologic Oncology Group study. *Journal of Clinical Oncology* 1994; **12**: 1408.

Wolfson AH *et al.* The prognostic significance of surgical staging for carcinoma of the endometrium. *Gynecologic Oncology* 1992; **45**: 142.

Young RH, Clement PB. Pathology of the uterine corpus. *Human Pathology* 1991; 776.

Zuna RE, Behrens A. Peritoneal washing cytology in gynecologic cancers: long-term follow-up of 355 patients. *Journal of the National Cancer Institute* 1996; **88**(14): 980–7.

36.8.2 Cervical malignancy

C. O. Granai, Stephen S. Falkenberry, Hector M. Tarraza, Walter H. Gajewski, Bruno Arena, Margarett Ellison and Michael A. Steller

Introduction
Pathology
The spectrum of cervical squamous neoplasia
Preinvasive disease
Microinvasion
Invasive disease
Clinical features
Diagnosis
Staging
Treatment
Early clinical disease
Advanced clinical disease
Survival
Recurrent cervical cancer
Summary and trends
Further reading

Introduction

Cervical cancer is the second most common malignancy among women worldwide, although in the United States of America it ranks only eighth. Approximately 450 000 new cases are diagnosed internationally each year, with nearly 200 000 deaths attributable to the disease. Furthermore, in developing countries, cervical cancer remains the leading cause of cancer-related death in women. In the United States, cytologic screening has resulted in a 70 per cent decline in deaths from cervical cancer in the past 50 years. The American Cancer Society estimates that, during 2000, in the United States there would be 12 800 new cases of invasive cervical cancer and 4600 deaths from the disease. Despite the ability to detect and cure most cases of local disease, 35 per cent of patients still develop persistent, or present with widely metastatic, disease, for which there is no consistently effective therapy. Taken together, these observations underscore the need to promote screening worldwide, and to develop more effective approaches for the treatment, or better, the prevention of cervical cancer.

A large body of evidence has accumulated to indicate that human papillomavirus (**HPV**) is the venereally transmitted agent serving as an important cofactor in the etiology of cervical carcinoma. Extensive epidemiologic data now strongly associate HPV with a spectrum of anogenital neoplasia including condyloma, cervical dysplasia, and cervical carcinoma. The risk of cervical cancer has long correlated with sexual intercourse at an early age, a history of multiple sexual partners, smoking, a large number of pregnancies, and a history of sexually transmitted diseases. Indeed, over time, many different sexually transmitted diseases, including gonorrhea and herpesvirus, have been thought to be the key etiologic agents responsible for cervical cancer. In retrospect, each might better have been regarded a ‘marker' for high-risk behavior through which the actual etiologic agents/cofactors were encountered, such that, in concert with other cofactors, cervical neoplasia was initiated.

In contrast to the other sexually transmitted diseases once held suspect, HPV seems to deserve the dubious distinction of causing some cases of cervical cancer. HPV (a heterogeneous group of DNA viruses consisting of 9-kb, closed circular and double-stranded DNA) can infect numerous sites in the body, such as the skin, mouth, esophagus, larynx, and anogenital tract. More than 70 different HPV genotypes have thus far been identified, of which approx. 20 show a predilection for tissues of the anogenital tract.

HPV DNA is detected in more than 90 per cent of all tumors of the uterine cervix. Mucosotropic HPVs are grouped into low-risk or high-risk categories on the basis of each genotype's association with a benign or malignant disease process. For example, HPV 6 and 11 are commonly detected in condyloma accuminata but are virtually never in cervical carcinoma; they are therefore considered ‘low-risk.' In contrast, high-risk types 16 and 18 can be detected in nearly 70 per cent of squamous-cell carcinomas of the cervix. Other high-risk types include HPV 31, 33, and 45. Molecular biology has shown that the *E6* and *E7* genes of HPV 16 and 18 encode for oncoproteins that can immortalize human keratinocytes. This potential appears to be limited to high-risk types, since E6 and E7 proteins from HPV 6 and 11 are nontransforming. The high-risk HPV E6 and E7 oncoproteins alter the regulation of cell growth by inactivating the products of tumor-suppressor genes *p53* and *pRB* (retinoblastoma), respectively. It has been postulated that the relatively low binding capacities of the low-risk HPV E6 and E7 proteins to p53 and pRB, respectively, in part account for their less aggressive functional properties.

In addition to HPV, cervical cancer is associated with another virus. A strong correlation exists between the human immunodeficiency virus (**HIV**) and intravenous drug use. By current definition, the development of cervical cancer in an ‘HIV-positive' woman categorizes her as having fulminant acquired immune deficiency syndrome. The long recognized, clear association between squamous-cell cancer of the cervix and sexual/social behavior permits the malignancy to be categorized as a sexually/socially transmitted disease, thus theoretically preventable.

Pathology

Over 90 per cent of invasive cervical cancers are histologically defined as squamous cell in origin; the other 10 per cent are various adenocarcinomas and less common subtypes. Although a less important distinction, most squamous cell cancers are of the large-cell, nonkeratinizing subtype and arise from the cervical transformation zone where columnar epithelium is constantly being replaced by squamous epithelium through metaplasia. In contrast, adenocarcinoma arises primarily from the columnar epithelium of the endocervical glands.

Most studies show that squamous and adenocarcinomas have similar prognoses and clinical behavior when prognostic factors are controlled for, but adenosquamous cancers, particularly the glassy-cell variant, carry a significantly poorer prognosis, usually requiring multi-modal therapy even for early-stage disease.

There are also many, less common histologic types of cervical cancer, each with unique biologic behaviors and prognoses. Adenoid cystic carcinomas generally carry a poor prognosis, whereas adenoid basal and verrucous carcinomas of the cervix have good prognoses. The verrucous carcinoma, characterized by pushing borders, is only locally aggressive and treatable by local resection. The infrequent small-cell cancer of neuroendocrine origin is an extremely virulent cervical neoplasm that can present with paraendocrine symptoms and early dissemination. Attempts to treat small-cell cancers of the cervix have been modeled upon therapies used against the synonymous lesion of the lung. But even when surgery, radiation, and chemotherapy are combined, the prognosis for this cervical malignancy is poor. Rarely, lymphomas, embryomas, embryonal rhabdomyosarcomas, leiomyosarcomas, malignant mixed muellerian tumors, melanomas, carcinoids, and primary sarcomas arise from the uterine cervix.

Finally, cancers found on the cervix are not always primary. Malignancies from the corpus of the uterus, breast, colon, and kidney are known to metastasize to the cervix.

(Given the dominating prevalence of squamous-cell cancer, the remainder of the chapter will address primarily the clinical/surgical management of that type.)

The spectrum of cervical squamous neoplasia

Preinvasive disease

Invasive squamous-cell cancer is frequently preceded by, or coexistent with, squamous intraepithelial lesions [dysplasias; cervical intraepithelial neoplasias (**CIN**)]. This finding reinforces the concept that cervical neoplasia is a ‘continuum' of disease ranging from mild dysplasia (CIN 1) to invasive cancer. Typically, slow (over 2 to 20 years), orderly biologic progression occurs across the continuum from preinvasive neoplasia to frankly invasive squamous cancer. Peterson followed 127 patients with untreated carcinoma *in situ* of the uterine cervix; 30 per cent had developed invasive cervical cancer by the tenth year of observation. While as many as 60 per cent of cases with low-grade epithelial neoplasia resolve spontaneously, rapid, unpredictable, malignant transformation can occur, possibly in conjunction with high-risk features such as HPV subtypes 16 and 18, HIV, and smoking. This ‘rapidly progressive' variant of cervical neoplasia challenges the ‘cost-effective' wisdom of recommendations for screening which increasingly suggest that longer intervals between Papanicolaou (**Pap**) smears are acceptable.

Microinvasion

The next level of neoplasia after carcinoma *in situ* (CIN 3, high-grade squamous intraepithelial lesions) is microinvasion. Microinvasion is a concept suggesting that an invasive cancer can exist yet be so minimal that it is still a local process (easily resectable for cure). Fundamental to the concept is a total absence of risk for regionally advanced disease or metastasis.

Expert opinions differ in the practical definition of microinvasion. The Society of Gynecologic Oncologists (**SGO**) defines the lesion pragmatically as a malignancy with a depth of invasion of up to 3 mm (from the base of the epithelium), without involvement of lymph vascular spaces, and without multiple foci of invasion (confluency). In contrast, the staging of cervical cancer by the International Federation of Obstetrics and Gynecology (**FIGO**) is more complex and subdivides ‘microinvasion' into two strata (Table 1). Indeed FIGO's higher subdefinition of stage Ia (also known as Ia^2, and also inferred to be ‘microinvasion') allows for a depth of invasion of the cervical stroma of between 3 and 5 mm and the presence of involvement of vascular spaces.



Table 1 International Federation of Obstetrics and Gynecology (FIGO) staging classification of cervical carcinoma

The deeper penetration (within FIGO stage Ia^2) has been associated with a greater risk of metastasis to pelvic lymph nodes. In one study the risk of pelvic nodal spread was 6.8 per cent for patients at FIGO stage Ia^2 and 0.21 per cent for patients where the SGO definition of less than 3-mm invasion was used. Similarly, other studies found that with stromal invasion of less than 3 mm the risk of metastasis to lymph nodes was very low, 0 to 1 per cent, but when invasion deepened to between 3.1 and 5 mm the risk increased to 4 to 7.3 per cent. Moreover, in a disturbing study, 85 patients at FIGO stage Ia^2 were followed for 5 years after conservative treatment (nonradical surgery) and three died from recurrent disease. Also troubling is that FIGO's definition of microinvasion remains silent on the importance of the involvement of lymph vascular spaces as an independent risk factor for local recurrence and metastasis.

Considering that ‘unambiguous curability' is the keystone of the definition of microinvasion as a clinical entity, the above results are incongruous and conceptually unacceptable. A safer clinical axiom is that ‘microinvasion' exists but must be very carefully defined. In that context, the FIGO designation of stage Ia^2 should be considered only a technicality and not a guideline inclusive of the concept of microinvasion or the more minimalist treatment it implies. Though the practical value of this higher degree of microinvasion is still being defined, it will likely prove to identify patients at intermediate risk needing an intermediate surgical intervention (e.g. modified radical hysterectomy).

Currently, the concept of microinvasion has yet to find universal acceptance in relation to adenocarcinomas. Since these lesions arise in the glandular epithelium of the cervix, measuring the actual depth of invasion is difficult and interobserver variability is high. As a practical matter, most centers continue to treat early invasive adenocarcinomas of the cervix in a manner consistent with stage Ib squamous lesions.

Invasive disease

Following microinvasion in the neoplastic spectrum is frankly invasive squamous-cell cancer. Inherent in this ultimate form of neoplasia are the properties of angiogenesis, uncontrolled growth, and metastatic potential.

Malignant progression

Beginning on and expanding in the cervix, the squamous lesion, over time, usually extends laterally into the paracervical and parametrial soft tissues, ultimately involving the pelvic side wall and obstructing the ureters. Simultaneously, the lesion often spreads inferiorly into the vagina and superiorly into the lower uterine segment. Direct, early extension into the bladder or rectum can occur but is uncommon.

Once the cancer has invaded the underlying cervical stroma it may enter the paracervical and parametrial lymphatic and vascular spaces. As a consequence, the parametrial obturator, internal iliac, and external iliac lymph nodes become sites of metastasis. As the malignancy ascends the lymphatic chain, the common iliac, para-aortic, thoracic, and the supraclavicular groups become involved. The risk of para-aortic nodal disease is under 1 per cent if the pelvic nodes are uninvolved. In collected series, patients with stage I (Table 1) cervical cancer have a 15 per cent incidence of metastasis to pelvic lymph nodes, stage II 27 per cent, and stage III 46 per cent; by comparison, the para-aortic nodes have an incidence of metastases of 2, 15, and 30 per cent, respectively. With rare exception, nodal metastasis is orderly without ‘skip lesions' to more superior locations.

The most common sites of distant metastases include the lungs, mediastinum, bone, and liver. Hematogenous dissemination occurs in fewer than 10 per cent of patients.

Clinical features

The majority of women with early occult cervical cancer are asymptomatic. The first clinical manifestation of invasive cervical cancer is usually abnormal vaginal bleeding. Among symptomatic patients there are complaints of intermenstrual bleeding, postcoital (the ‘classic' but less common clinical history), and postmenopausal bleeding occurring in friable, frequently ulcerated, cervical epithelium. When cervical cancer is necrotic, ‘vaginal discharge' is a common feature. Pain (pelvic, lumbosacral, or gluteal) is a symptom that often heralds advanced disease. For example, abdominal and low back pain may be caused by metastases to para-aortic lymph nodes. Occasionally, at presentation, there may be hematuria or rectal bleeding, suggesting involvement of the bladder or rectum.

Diagnosis

The diagnosis of cervical cancer is established by biopsy and histologic evaluation. All clinical lesions on the cervix should be evaluated by biopsy. Even when the cervix is clinically normal (i.e., the cancer is occult), cytologic examination by Pap smear usually reveals atypical, dysplastic, or frankly malignant squamous or glandular cells. Despite a strong correlation between cervical cytologic and histologic/biopsy features, cytology itself is insufficient for diagnosing any degree of neoplasia, certainly to include invasive cancer. Rather, abnormal Pap smears are ‘red flags' warning of possible neoplasia. The appropriate evaluation of the suspicious Pap smear consists of colposcopic examination with biopsy of the visible lesions and endocervical curettage to evaluate the endocervix above the colposcopic field of vision. Upon histologic confirmation of invasive cancer, the depth of cervical penetration/invasion must be determined. When the cervical biopsy reveals more than 3 mm of invasion or the presence of involvement of lymph vascular spaces, no further biopsy is required, and staging and treatment planning should proceed. The diagnostic (and defining) cervical cone biopsy should be performed when ‘microinvasion' is suspected from a Pap smear or cervical biopsy, the depth of invasion is uncertain and possibly quite limited, colposcopic examination fails to visualize the invasive lesion or explain the suspicious cytologic features, or endocervical curettage reveals dysplastic or malignant cells in the absence of ectocervical invasive cancer.

Staging

In contrast to ovarian and endometrial malignancies, which are staged surgically, the FIGO staging of cervical cancer is clinical and based on its usual malignant progression (Table 1). As such, the single most important staging procedure is the clinical pelvic examination. Chest radiography and radiographic evaluation of the

ureters are also performed routinely. Though intravenous pyelography is still officially cited as the standard means of evaluating the genitourinary system in patients with cervical cancer, computed tomography (**CT**) provides more comprehensive information on lymph-node status and extent of pelvic–abdominal disease, so wherever possible the CT scan has largely replaced the intravenous pyelogram. Cystoscopy and proctoscopy are not cost-effective, and are rarely useful for early (e.g. stage Ia[1] -Ib[1]) disease but should be performed more often in the staging of more advanced disease. The value of radiographic assessment of lymph-node status is limited, but it may be useful in treatment planning when obviously involved nodes are found.

Heller studied 320 patients with stage IIb, III, and IV disease by lymphangiogram, CT, and ultrasonography; when evaluating pelvic lymph nodes, the false-negative rates were 14, 25, and 30 per cent, respectively. In other studies evaluating the para-aortic nodes, CT scans had a false-negative rate of 10 to 15 per cent and a false-positive rate of 20 to 25 per cent. In selected cases, magnetic resonance imaging may be helpful in depicting tumor involvement of the lower uterine segment and extracervical extension. Though not part of the current FIGO staging, pretreatment staging by imaging (e.g., CT, MRI, PET scans) as well as by molecular and cytologic/histologic means seems the future in cervical cancer as in many other areas of oncology.

Treatment

Early clinical disease

Microinvasive squamous carcinoma as defined by SGO or FIGO stage Ia[1] can be treated by either cone biopsy or extrafascial total hysterectomy (Rutledge type I; Table 2), depending on the patient's desire to preserve fertility. Currently the term ‘microinvasion' should not be applied to adenocarcinomas in treatment planning (see above); past attempts to do so resulted in unacceptably high rates of recurrence.

Type	Description
I	Extrafascial (simple) hysterectomy; parametrial ligament is incised, allowing lateral deflection of ureter
II	Modified radical hysterectomy; removal of medial half of cardinal and uterosacral ligaments, upper third of vagina removed
III	Removal of entire cardinal and uterosacral ligaments, upper third of vagina removed
IV	Removal of all parametrial tissue, superior vesical artery, and three-quarters of vagina
V	Removal of portions of rectal cervix and bladder

Table 2 Rutledge classification of hysterectomy

For early cervical cancer slightly beyond the SGO criteria for microinvasion (e.g. FIGO Ia[2]), less-than-radical surgery should be used only with the understanding that there exists a relatively increased risk of lymph-node metastasis and local recurrence. Such patients may be treated with either modified radical hysterectomy (Rutledge type II) with pelvic lymphadenectomy or primary radiation therapy, depending on their age and medical condition.

Stage Ib[1] and IIa cervical cancers are treated with either radical hysterectomy (Rutledge type III) with pelvic lymphadenectomy (Meigs' procedure) or radical radiation therapy using both external pelvic and brachytherapy. Although similar 5-year survival rates are reported for surgery and radiation therapy, surgery as primary treatment has advantages, such as conservation of the ovaries in premenopausal women, and radiation is associated with greater long-term morbidity. Radiation causes vaginal fibrosis, which often results in disturbances in sexual function. Chronic bladder and bowel problems resulting from radiation require major surgical intervention in 8 per cent of patients. Another major advantage of primary surgery over radiation for cervical cancer is that it can be used to determine the extent of disease more precisely by performing lymphadenectomy. The better understanding of disease distribution that results from surgical staging permits a more exacting separation of patients into groups with good and poor prognosis, thus determining those in need of postoperative adjuvant treatment.

The treatment of stage Ib[2] (bulky Ib) cervical cancer is discussed in the section ‘Special cases of early cervical cancer.'

Radical surgery

This was first described by Clark in 1896. In 1912, Wertheim introduced the radical hysterectomy as a primary treatment for cervical cancer; during his career he carried out 500 radical operations with a substantial 30 per cent operative mortality and a survival of 40 per cent. In 1944, Meigs from the Massachusetts General Hospital expanded on the Wertheim approach by including pelvic lymphadenectomy and a more extensive parametrial extirpation. Despite the greater complexity of the operation, Meigs reported a mortality of only 1 per cent, and the 5-year survival was better (90 per cent for node-negative patients and 42 per cent for node-positive). Acute complications of radical hysterectomy and pelvic lymphadenectomy then, as now, included hemorrhage, infection, and deep venous thrombophlebitis. Delayed complications are bladder atony in 3 per cent, genitourinary fistula in 1 to 2 per cent, and lymphoceles.

Modern surgical technique, anesthesia, and postoperative care have lessened operative morbidity since Meigs' report, which explains why surgery has emerged as the primary treatment of choice for early cervical cancer, even for patients once dogmatically considered too elderly. If, however, at laparotomy unexpected intraoperative findings are encountered (e.g. unresectable disease or nodal metastases), departure from the intended radical hysterectomy may be appropriate in deference to the pursuit of an alternative (e.g. radiation with or without chemotherapy) more consistent with the new-found extent of disease.

Prognostic factors/risk stratification/adjuvant therapy

The Gynecologic Oncology Group (**GOG**) and other investigators have analyzed pathologic findings predictive of survival (Table 3). Survival is correlated with tumor size, vascular invasion, parametrial extension, depth of stromal invasion, and lymph-node status. Other factors, such as microvessel counts and indicators of angiogenesis, also correlate with outcome. Some factors, such as tumor size, depth of stromal invasion, and possible involvement of lymph vascular spaces, may be surrogates for lymph-node metastasis rather than independently prognostic. The presence of positive lymph nodes carries a range of prognoses that seemingly depend on the location, size, and number of nodes involved (Table 4).

Factor	*Positive pelvic lymph nodes (%)*	*p*
Histologic grade:		
1	9.7	
2	18.9	0.01
3	28.8	
Maximum clinical tumor diameter (cm)		
0.1–1.0	12.2	
1.1–2.0	15.8	0.2
2.1–3.0	16.3	
3.1+	23.0	
Depth of invasion (mm):		
<5	3.4	
6–10	15.1	
11–15	23.2	0.0001
16–20	30.8	
21+	33.6	
Inner third	4.5	
Middle third	15.3	0.0001
Outer third	26.4	
Parametrial extension:		
Negative	13.5	
Positive	43.2	0.0001
Capillary/lymphatic spaces:		
Negative	8.2	
Positive	25.4	0.0001
Weight of uterus:		
Negative	18.2	
Positive	25.0	0.4

Table 3 Pathologic evaluation of 940 patients with stage Ib cancer of the cervix

Status	5-year survival (%)
Lymph nodes negative	90
Lymph nodes positive	50
One positive lymph node	65
More than one positive lymph node	35
Unilateral positive nodes	60
Bilateral positive nodes	30

Table 4 Correlation between lymph-node status and survival for stage Ib cervical cancer

Based on the information obtained at operation, patients at high risk for recurrence can benefit from, or be eligible for, protocols using adjuvant radiotherapy and/or chemotherapy. While patients with multiple positive pelvic lymph nodes and/or involved surgical margins have had fewer local recurrences after postoperative radiation, the survival benefit has been less clear. Selection bias confounds the old data; definitive proof that postoperative radiation improves survival among high-risk patients has remained elusive. Furthermore, while other prognostic factors such as involvement of the lymph vascular spaces, tumor diameter, and depth of invasion are associated with an increased risk of local recurrence, even here adjuvant radiation therapy has inconsistently reduced that risk. Recently, however, the GOG reported preliminary data from a randomized, prospective trial demonstrating that adjuvant radiation decreased the risk of recurrence for high-risk patients. Severe radiation toxicity to the gastrointestinal and genitoturinary systems was 2.9 per cent but most toxicity was transient. The effect on overall survival has not yet been calculated. Nevertheless, based on a long-held inference and by analogy with another, recent GOG 123 study of bulky, stage Ib cervical cancer, postoperative adjuvant radiation, with sensitizing chemotherapy, seems on its way to becoming the standard for high risk patients. Furthering this sense is the recently reported trial of the Southwest Oncology Group (SWOG 8797) showing a 51 per cent reduction in the risk of death among high risk cervical cancer patients following radical hysterectomy if treated with chemotherapy and radiation compared with those who received postoperative radiation alone. Thus, adjuvant postoperative radiation therapy, now likely with chemotherapy, should be offered to patients with positive surgical margins and/or involvement of multiple pelvic lymph nodes. When available, further data from the GOG trial on adjuvant treatment for involvement of lymph vascular spaces, deep stromal invasion, and large tumor diameter should be useful in making firmer treatment recommendations.

A role for adjuvant chemotherapy here, as in breast and ovarian cancer, would seem logical, although there are few data. Small studies have suggested that it can improve outcome for some patients with high-risk disease after surgery. Again, more trials are needed, probably, as suggested by recent GOG trials, in conjunction with concomitant radiation.

Special cases of early cervical cancer

The bulky cervix

Some cases of ostensible ‘early' cervical cancers merit special management. The so-called barrel cervix (stage Ib^2) is an example. When carcinoma arising within the endocervical canal expands drastically the entire cervix (e.g. to more than 6 cm in diameter), the now bulky cervix is less suitable geometrically for receiving uniform, high-dose brachyradiation for cure. The barrel cervix is so large and distorted that curative tumoricidal isodose curves encompassing the entire ‘barrel' are not always possible, leaving instead the peripheral and hypo-oxygenated areas of the lesion undertreated. This accounts for a reported central failure rate of 15 to 25 per cent after treatment of the bulky cervix with radiation alone (the treatment many experts believe is currently the recommendable standard).

The best clinical management for the bulky cervix is controversial. Many centers still believe that modified radical radiation (approxi-mately 4000 cGy to the whole pelvis, and a single tandem and colpostat brachytherapy treatment) followed by extrafascial hysterectomy to enhance local control is proper in the management of the barrel-shaped cervix. The rationale for this approach has been to remove surgically the now radiation-shrunken/radiation-resistant, persistent central disease within the cervix. Using this sequence of treatments, Rutledge *et al.* and later Gallion *et al.* reported that central failures were reduced from 17 to 19 per cent to 2 per cent. Unfortunately, statistically significant improvement in survival for this more involved, possibly more morbid, approach was not demonstrated by this or other series. The GOG has now reported improved survival and fewer relapses for patients with stage Ib bulky cervical cancer on using weekly cisplatinum and radiation followed by hysterectomy as compared to radiation alone followed by hysterectomy. The recurrence rate for the former group was 20.8 per cent compared to 35.7 per cent for the latter. This equaled a 46 per cent reduction in risk of death for patients receiving weekly cisplatin along with radiation. The implications of this randomized trial are large but possibly still premature. Thus, for the moment, despite the various experts, no absolute recommendations can be made for treating the bulky cervix. In time there may be a proven advantage to radiation with sensitizing chemotherapy, possibly followed by surgery (or not). In the meanwhile, careful pretreatment staging, planning, and discussions are paramount.

Cervical cancer in pregnancy

Cervical cancer in pregnancy is particularly complex and emotionally distressing. Management must be based on the stage of the disease, the gestational age at diagnosis, and the patient's desires about the pregnancy. Microinvasive disease can usually be managed with cone biopsy to achieve a negative surgical margin. More invasive, stage I tumors should be treated individually. When diagnosed in early pregnancy, frankly invasive cases (Ia^2–IIa) are usually, and certainly have been in the past, treated with radical hysterectomy as in the nonpregnant woman. The evolution of superb neonatal care of infants, even for those weighing fewer than 500 g, has brought new considerations and new dilemmas for patients' decisions upon interventional strategies. In order to preserve the pregnancy to viability (26 weeks' gestation in some neonatal intensive-care nurseries), various approaches have been tried, including cases in which platinum-based chemotherapy has been used as a neoadjuvant during the pregnancy and definitive therapy delayed until after the birth. Obviously then, when cervical cancer is diagnosed in the third trimester, definitive therapy can usually be delayed until fetal maturity and delivery are accomplished. In cases of advanced cervical cancer in pregnancy, treatment must be individualized. All patients with cervical cancer in pregnancy are best managed at tertiary centers by an informed team including high-risk obstetricians, neonatal pediatricians, and gynoncologists, all of whom serve to advise the patient in her most difficult choice.

Old and new operations

In an effort to treat effectively early but invasive cervical cancer and preserve the patient's (desired) potential fertility, trachelectomy has been/is very occasionally employed at some centers. Removal of the entire cervix with preservation of the uterine fundus has been technically feasible for years. Whether it is an effective and compassionate strategy remains to be determined, but it does seem to have some rationale, as does the combination of radical vaginal hysterectomy and laparoscopic lymphadenectomy in other settings. At the moment, however, neither approach is the standard of care in the United States.

Advanced clinical disease

Radiation plus sensitizing chemotherapy

Stage IIb to IVa cervical cancers, by default, had traditionally been treated with primary radical radiation therapy. Five-year survivals vary with the stage and the reporting institution but are generally between 20 and 50 per cent. Many failures of radiation therapy are the result of treating large tumor volumes, which are beyond the curative capacities of tolerable radiation doses. Also, at least 30 per cent of patients with stage IIb to IVa cancer have metastatic disease in the para-aortic lymph nodes. Extended-field radiation, covering the pelvis and incorporating a para-aortic ‘chimney', has had minimal impact on widespread nodal disease of this type and no impact on the probably concomitant, but at present occult, more distant metastases. As for the para-aortic adenopathy, the problem is that substantial, tumoricidal doses of radiation cannot be given because of their toxicity. Thus constrained, radiation has but a limited ability to sterilize the malignant lymph nodes, especially those that are large or widely dispersed. Hacker *et al.* have reported improved survival in women who have undergone surgical removal of enlarged para-aortic lymph nodes prior to extended-field radiation.

Recently, however, the standard of care for the treatment of advanced cervical cancer has been changed. The role of radiosensitizing chemotherapy using agents such as cisplatinum and 5-fluorouracil is still evolving but the concept has demonstrated its valuein several important trials (GOG 85, GOG 120, GOG 123, RTOG 90-01, SWOG 8797). In each, cisplatinum (with or without 5-fluorouracil) and concurrent radiation resulted in a significant reduction in risk of recurrence and death from cervical cancer stages II-IVA. These dramatic findings led to an announcement by the National Cancer Institute informing physicians of this new standard of care (i.e.,

radiation with sensitizing chemotherapy).

Stage IVb cervical cancer is incurable and, accordingly, treatment is individualized and palliative. For some patients, local, high-dose external radiation therapy can be used as an unconventional but creative means to control vaginal bleeding or localized pain due to bone involvement. The advantage of this approach over conventional, 1.8-Gy daily fractions to a total of 45 Gy is less intrusion into the remaining life of the patient. Other palliative measures such as opioid, steroid, nonsteroidal, and neuroabsessive analgesia should be used aggressively to minimize suffering from incurable disease. In patients with incurable disease in whom reasonable therapeutic options have been exhausted, urinary diversion or stenting for obstruction is not appropriate. The natural history of cervical cancer in this setting is progression to fatal uremia, a relatively painless, humanely lethal process.

Surgery

Occasionally a patient with stage IVa squamous carcinoma is better treated with primary surgery (anterior or total exenteration) when the tumor is deemed surgically resectable and it otherwise presents an unfavorable geometry for local control with radiation. Enthusiasm for surgery must be tempered, however, since 40 per cent of these patients will have involved para-aortic lymph nodes (i.e., distant disease). If found at surgery the enlarged lymph nodes should be considered for removal, as doing so may enhance control and survival. Nevertheless, involved para-aortic nodes or distant metastases are general contraindications to exenterative surgery.

Reports from Japan have suggested that primary radical hysterectomy is effective for select, locally advanced cervical cancers (stage IIb). Surgery here offers the same advantages as discussed previously. Following surgery, adjuvant therapy is carefully considered.

Chemotherapy

There are currently three settings in which chemotherapy is employed in the treatment of cervical cancer:

1. as a radiosensitizer during primary radiation therapy (see previous sections);
2. in the neoadjuvant treatment of locally advanced disease (Ib^2–IVa);
3. in the treatment of metastatic disease (IVb).

In all of the above, cisplatinum is the agent that has shown the greatest activity in the treatment of squamous carcinoma of the cervix. Response rates of 23 per cent, including 10 per cent complete responses, have been reported in GOG trials using single-agent cisplatinum, 50 mg/m^2 every 3 weeks. Other agents, such as ifosfamide, doxorubicin, 5-fluorouracil, paclitaxel, and bleomycin, have lower (less than 15 per cent) response rates. Because of the increased toxicity relative to response rate in multiagent regimens, single-agent cisplatinum remains the agent of choice in most cases where the palliation of advanced and recurrent disease is the realistic intent.

The value of neoadjuvant chemotherapy in the treatment of locally advanced cervical carcinoma remains debated. While studies have shown objective responses, some resulting in ‘downstaging' to a stage of operability, only Sardi *et al.* have consistently demonstrated long-term survival advantage. Studies on this important issue are under way.

Women and Infants' Hospital/Brown University treatment

All patients at the Women and Infants' Hospital/Brown University with gynecologic malignancies are presented to a multidisciplinary, prospective Tumor Board for recommendations on primary treatment, and again as new information or events warrant. Primary-treatment protocols typically suggested by the Board for cervical cancer are as follows.

Stage Ia

Early stage Ia^1 (less than 1 mm cervical stromal invasion without involvement of the lymph vascular spaces) is managed either minimally with cone biopsy or, if preserved fertility is not desired, by extrafascial hysterectomy with the removal of suspicious lymph nodes. For from 1 to 3 mm of stromal invasion, typical management is modified radical hysterectomy and select pelvic lymphadenectomy, unless the patient has an overwhelming desire to conserve her fertility and is willing to assume the risk of less proven alternatives. Para-aortic lymphadenectomy is performed if pelvic nodes are involved.

Stage Ib–IIa

The majority of patients with stage Ib (defined by SGO criteria) and early IIa cancers undergo radical hysterectomy, pelvic lymphadenectomy, and para-aortic lymphadenectomy only as indicated. At laparotomy, the actual surgery may be modified, redirected, or stopped if unexpected circumstances (such as parametrial extension or the involvement of para-aortic lymph nodes) are encountered. After surgery, patients are stratified, based on the surgical–pathologic findings, and those deemed at high risk are offered adjuvant treatment on various research protocols (or with radiation therapy and sensitizing chemotherapy). If positive pelvic nodes are encountered, an alternative to radical surgery, in specific cases, is primary radical radiation with chemotherapy.

Stage Ib2 (bulky tumors larger than 6 cm) are usually managed with primary radiation, typically with sensitizing platinum chemotherapy, followed by strong consideration of extrafascial hysterectomy, in part depending on response to radiation therapy.

Stage IIb and beyond

Patients with advanced disease are encouraged to consider participating in various experimental protocols as an alternative to what is otherwise typically used—radiation therapy with sensitizing chemotherapy. Off protocol, patients with stage IIb to IVa disease and without evidence of involvement of para-aortic lymph nodes or distant metastases on CT scan are considered for pretreatment, extraperitoneal sampling of para-aortic nodes. If these are positive, treatment is usually extended-field radiation with sensitizing cisplatinum; if negative, standard pelvic radiation is combined with sensitizing chemotherapy. Occasional patients with stage IVa disease are recommended for primary anterior exenteration, though most are treated by primary radiation and sensitizing chemotherapy. Women with stage IVb disease not opting for protocols are offered platinum-based chemotherapy or other palliation.

Survival

The overall 5-year survival of patients with cervical cancer is only 60 per cent. Stage of disease is the single, most important predictor of survival. Five-year survival for patients with stage Ia disease is 98 per cent, for stage Ib to IIa 75 to 85 per cent, stage IIb 55 per cent, stage III 10 to 50 per cent, and for stage IVb essentially none. Discouragingly, despite improved radiotherapeutic techniques and better delineation of the extent of the disease, survival rates for cervical cancer had not improved much in the last 30 years. Recent advances which combine modalities, such as chemoradiation, will likely improve upon past outcomes, however.

Recurrent cervical cancer

The overall recurrence rate of cervical cancer is 35 per cent. In contrast, only 10 to 20 per cent of patients with stage Ib disease experience recurrence; this is increased to 45 per cent if affected nodes were found originally. Recurrence is generally within the first 2 to 3 years following primary treatment, and most patients who have recurrence die from the disease.

Recurrent cervical cancer generally has a poor prognosis and must be treated within the context or constraint of the site of the recurrence and the type of therapy previously given (surgery and/or radiation). Sites of recurrence are categorized as local/central (potentially resectable and curable), regional (in the pelvis, but extending to the side wall, or lymph nodes, unresectable and rarely curable), and distant.

A true local (central) recurrence after surgery can be managed with radiation therapy, with reported survivals as high as 25 to 50 per cent. In previously radiated patients with central recurrences, the selective use of pelvic exenteration offers a 30 to 60 per cent 5-year survival if total resection is possible. Because of the high rate of complications (Table 5) and the major long-term physical and psychologic effects of this radical procedure, exenterative surgery is only indicated if cure is possible. Extensive counseling on the details of surgery, operative morbidity, mortality, and the long-term impact of colostomy and urinary stoma on body image and sexual function is imperative.

Operative mortality	7
Serious infection	5
Small-bowel obstruction	5–27
Gastrointestinal fistula	8
Genitourinary fistula	5

Table 5 Complications associated with pelvic exenteration (percentages)

Candidates for surgery must be carefully selected, restaged, and should have no evidence of disease beyond the central pelvis. The triad of unilateral edema of the lower extremities, sciatic pain, and ureteral obstruction suggests almost certain unresectability, and thus is a relative contraindication to exenteration. Unfortunately, even in an asymptomatic patient seemingly with clinically resectable disease, exenteration is often not possible due to intraoperative findings. When this occurs, treatment options include systemic chemotherapy and palliation.

Summary and trends

A less predictable, more virulent, cervical neoplasia associated with high-risk cofactors such as HIV, subtypes of human papillomavirus, and cigarette smoking may be emerging. Unfortunately, women with risk factors for cervical cancer, such as HIV infection and high-risk behavior, are less likely as a group to seek gynecologic screening and care. The future screening and management of cervical neoplasia will need to be adjusted to fit this evolving biologic pattern and social behavior. Traditional treatment of cervical cancer yields good results for early-stage disease, but with past, standard approaches (i.e. radiation alone) the prognosis for high risk, low-stage, and advanced disease remained poor and grounded. Finally, however, chemoradiation strategies, possibly combined with surgery, together allied with better staging, and risk stratification, hold promise for improving upon outcomes. Strategies for early detection such as full compliance with screening guidelines, prevention strategies stressing cessation of smoking, condom use, and the avoidance of high-risk behavior have the greatest potential for reducing morbidity and mortality from cervical cancer. Contact with the at-risk population through education and outreach is essential. A viricidal agent that eradicates HPV or an HPV vaccine could potentially eliminate this disease altogether. Promising areas of research in the treatment of cervical cancer include neoadjuvant chemotherapy, HPV-directed immunotherapy, and antiangiogenesis therapy.

Further reading

Abayomi O *et al.* The value of ‘routine tests' in the staging evaluation of gynecologic malignancies: a cost effectiveness analysis. *International Journal of Radiation Oncology, Biology, Physics* 1982; **8**: 241.

American College of Obstetricians and Gynecologists. *Diagnosis and management of invasive cervical carcinomas*. Technical Bulletin, No. 138, 1989.

Averette HE, Lichtinger M, Sevin BU, Girtanner RE. Pelvic exenteration: a 15-year experience in a general hospital. *American Journal of Obstetrics and Gynecology* 1984; **150**: 179.

Averette H, Nelson JH, Jr, Ng AB, Hoskins W, Boyce JG, Ford JH. Diagnosis and management of microinvasive (stage la) carcinoma of the uterine cervix. *Cancer* 1976; **38**: 414.

Benedetti-Panici P *et al.* Locally advanced cervical adenocarcinoma: is there a place for chemo-surgical treatment ? *Gynecologic Oncology* 1996; **61**(1): 44–9.

Berek JS *et al.* Adenocarcinoma of the uterine cervix: histologic variables associated with lymph node metastasis and survival. *Obstetrics and Gynecology* 1985; **65**: 46.

Braly PS. The NIH consensus conference on cervical cancer: implications for practice. *Primary Care Update for Obstetricians/Gynecologists* 1997; **4**(5): 179–83.

Burghardt E *et al.* Microinvasive carcinoma of the uterine cervix (International Federation of Gynecology and Obstetrics stage la). *Cancer* 1991; **67**: 1037.

Burghardt E *et al.* The new FIGO definition of cervical cancer stage 1a: a critique. [Editorial.] *Gynecologic Oncology* 1997; **65**(1): 1–5.

Cancer Statistics 1998. *Journal of the American Society*, Ca Cancer J Clin, 1998; **48**(1): 25.

Creasman WT. New gynecologic cancer staging. [Editorial.] *Gynecologic Oncology* 1995; **58**: 157–8.

Creasman WT *et al.* Management of stage la carcinoma of the cervix. *American Journal of Obstetrics and Gynecology* 1985; **153**: 164.

Delgado G. Stage Ib squamous cancer of the cervix: the choice of treatment. *Obstetrics and Gynecology Surveys* 1978; **33**: 174

Delgado G *et al.* A prospective surgical pathological study of stage I squamous carcinoma of the cervix: a Gynecologic Oncology Group Study. *Gynecologic Oncology* 1989; **35**: 314.

Duguid HLD, Duncan ID. In: Singer A, ed. *Cancer of the cervix: diagnosis and treatment*, p. 77. Saunders, Philadelphia, 1985.

Elkas J, Farias-Eisner R. Cancer of the uterine cervix. *Current Opinion in Obstetrics and Gynecology* 1998; **10**(1): 47–50.

Fanning J *et al.* Cisplatin, 5-fluorouracil, and ifosfamide in the treatment of recurrent or advanced cervical cancer. *Gynecologic Oncology* 1995; **56**(2): 235–8.

Fields AL *et al.* Mature results of a phase II trial of concomitant cisplatin/pelvic radiotherapy for locally advanced squamous cell carcinoma of the cervix. *Gynecologic Oncology* 1996; **61**(3): 416–22.

Finan MA, Kline RC. Controversies in the management of stage Ib cervical cancer. *Contemporary Obstetrics/Gynecology* 1998: 77,78, 85–100.

Fletcher GH, Wharton JT. Principles of irradiation therapy for gynecologic malignancy. *Current Problems in Obstetrics and Gynecology* 1978; **2**: 2.

Graham JB, Sotto LSJ, Paloucek FP. *Carcinoma of the cervix*. Saunders, Philadelphia, 1962.

Greenlee RT, Murray T, Bolden S, Wingo PA. Cancer Statistics, 2000. *CA-A Cancer Journal For Clinicians* 2000; **50**: 32–3.

Heller PD, Malfetano JH, Bundy BN. Clinical pathologic study of stages IIb, III, and IVa carcinoma to the cervix: extended diagnostic evaluation for paraaortic node metastasis (a GOG study). *Gynecologic Oncology* 1990; **38**: 425.

Hopkins MP, Lavin JP. Cervical cancer in pregnancy. [Editorial.] *Gynecologic Oncology* 1996; **63**(3): 293.

Hoskins PJ *et al.* Small cell carcinoma of the cervix treated with concurrent radiotherapy, cisplatin, and etoposide. *Gynecologic Oncology* 1995; **56**(2): 218–25.

International Federation of Gynecology and Obstetrics. Annual report on the results of treatment in gynecologic cancer. *International Journal of Gynecology and Obstetrics* 1989; **28**: 189.

Iwasaka T *et al.* Adjuvant chemotherapy after radical hysterectomy for cervical carcinoma: a comparison with effects of adjuvant radiotherapy. *Obstetrics and Gynecology* 1998; **91**(6): 977–81.

John M *et al.* Final results of a phase II chemoradiation protocol for locally advanced cervical cancer: RTOG 85–15. *Gynecologic Oncology* 1996; **61**(2): 221–6.

Keys HM, Bundy BN, Stehman FB, *et al.* Cisplatin, radiation and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. *New England Journal of Medicine* 1999; **340**: 1154–61.

Lawrence WD. Advances in the pathology of the uterine cervix. *Human Pathology* 1991; **22**: 792.

Look KY *et al.* An analysis of cell type in patients with surgically staged stage Ib carcinoma of the cervix: a Gynecologic Oncology Group Study. *Gynecologic Oncology* 1996; **63**(3): 304–11.

Marcial VA, Marcial LV. Radiation therapy of cervical cancer: new developments. *Cancer* 1993; **71**: 1438.

Meigs JV. Carcinoma of the cervix—the Wertheim operation. *Surgery, Gynecology and Obstetrics* 1944; **78**: 195.

Miller B *et al.* Intestinal fistulae formation following pelvic exenteration: a review of the University of Texas M.D. Anderson Cancer Center experience, 1957–1990. *Gynecologic Oncology* 1995; **56**(2): 207–10.

Morris M, Eifel PJ, Lu J, *et al.* Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *New England Journal of Medicine* 1999; **340**: 1137–43.

Peters WA, Liu PY, Barrett R, *et al.* Cisplatin, 5-fluorouracil plus radiation therapy are superior to radiation therapy as adjunctive therapy in high-risk, early-stage carcinoma of the cervix after radical hysterectomy and pelvic lymphadenectomy: report of a phase III intergroup. *Gynecology and Obstetrics* 1999; **72**: 443.

Peterson O. Spontaneous course of cervical pre-cancerous conditions. *American Journal of Obstetrics and Gynecology* 1956; **72**: 1063.

Piver MS, Rutledge FN, Smith PJ. Five classes of extended hysterectomy of women with cervical cancer. *Obstetrics and Gynecology* 1974; **44**: 265.

Raju KS, Bates TD, Taylor RW. Primary adenocarcinoma of the cervix: treatment and results. *British Journal of Obstetrics and Gynaecology* 1987; **94**: 1212.

Reeves WC, Brinton LA, Garcia M. Human papilloma virus infection and cervical cancer in Latin America. *New England Journal of Medicine* 1989; **320**: 1437.

Richart RM. Causes and management of cervical intraepithelial neoplasia. *Cancer* 1987; **60**: 1951.

Roberts WS. Major morbidity after pelvic exenteration: a 7 year experience. *Obstetrics and Gynecology* 1987; **69**: 617.

Rose PG, Bundy BN, Watkins EB, *et al.* Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. *New England Journal of Medicine* 1999; **340**: 1144–53.

Rubin SC, ed. *Chemotherapy of gynecologic cancers* (Society of Gynecologic Oncologists handbook). Lippincott-Raven, Philadelphia, 1996.

Serur E *et al.* Neoadjuvant chemotherapy in stage Ib2 squamous cell carcinoma of the cervix. *Gynecologic Oncology* 1997; **65**(2): 348–56.

SGO Medical Practice and Ethics Committee. *Clinical practice guidelines, management of gynecologic cancers*, October 1996.

Shingleton HM, Gusberg SB. Radical hysterectomy. In: Gusberg SB, Shingleton HM, Deppe G, ed. *Female genital cancer*, p. 534. Churchill Livingstone, New York, 1988.

Ueki M, Okamura S, Maeda T. Individualization of patients for adjuvant chemotherapy after surgical treatment of cervical cancer. *British Journal of Obstetrics and Gynaecology* 1987; **94**: 985.

van Nagell JR, Depriest PD, Higgins PV, Powell DE. Surgical therapy for cervical cancer. In: Gershenson DM, DeCherney AH, Curry SL, ed. *Operative gynecology*, p. 271. Saunders, Philadelphia, 1993.

Vermorken JB. The role of chemotherapy in squamous cell carcinoma of the uterine cervix: a review. *International Journal of Gynecologic Cancer* 1993; **3**: 129.

Wertheim E. The extended abdominal operation for carcinoma uteri. *American Journal of Obstetrics and Gynecology* 1912; **66**: 169.

Whitney CW, Sause W, Bundy B, *et al.* Randomized comparison of fluorouracil plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: a Gynecologic Oncology Group study. *Journal of Clinical Oncology* 1999; **17**: 1339–48.

36.8.3 Radical hysterectomy

C. O. Granai and Hector M. Tarraza

Rationale
Technique
Examination under anesthesia and opening the abdomen
Exploration of the abdomen
Exploration of the pelvis
The ovaries
Pelvic lymphadenectomy
Dissection of the superior pelvic ureter
Mobilization of the uterus
Dissection of the inferior pelvic ureter
Transection of the parametrium and paracervical tissues
Vaginectomy
Suprapubic catheter and drains
Morbidity
Further reading

Cervical cancer is the second most common malignancy in women worldwide, even though cervical cytologic screening (the Pap smear) has substantially reduced the incidence of this carcinoma in many regions. The treatment of cervical cancer is based on establishing its anatomic distribution (stage) within the patient and then managing that disease distribution in accordance with the specific biologic behavior of the neoplasm itself (pathways of contiguous and lymphatic spread). Generally, for early tumors (stages Ib and IIa), radical hysterectomy is now the better curative option; whereas for the more advanced stages of disease, radiation therapy, chemotherapy, and occasionally surgery, alone or increasingly in combination, become the treatment/s of choice.

Although abdominal hysterectomy was first described in 1895, not until Wertheim published his experience with 500 cases in 1912 did the operation become popular. Wertheim's high surgical mortality rate of 20 per cent was understandably problematic, even for that era. Unfortunately, the less morbid surgical alternative, radical vaginal hysterectomy, was, for technical reasons, applicable only to one-fifth of patients.

Ironically, the introduction of radium therapy, while improving the effectiveness of treatment for cervical cancer, was a major setback in the evolution of radical hysterectomy. Although several centers in Europe and the United States of America continued to perform the operation, most clinics at that time abandoned the approach in deference to the newer, safer, radium therapy. However, in 1944, Dr JV Meigs, in Boston, Massachusetts, published details of 47 patients treated by combining radical hysterectomy with bilateral pelvic lymphadenectomy using relatively modern surgical and medical techniques. He described curative results without the high operative mortality rates of previous series. Meigs's report reinstituted radical hysterectomy, but done in combination with bilateral pelvic lymphadenectomy, as an effective treatment option for early-stage (Ib, IIa) cervical carcinoma, as well as for other selected pelvic malignancies (i.e., stage II endometrial carcinoma and stage I vaginal carcinoma confined to the upper vagina).

Rationale

Radical hysterectomy is designed to be curative, based on the Halstedian principle of radical *en bloc* dissection. All the tissues adjacent to the cervix, including the paracervical and parametrial tissues, and upper vagina, are removed together; indeed, all tissues medial to the iliac vessels are incorporated in the dissection. A bilateral pelvic lymphadenectomy is conducted along the iliac vessels and in the obturator space. In clinical circumstances (e.g. stage Ib, IIa) where survival rates after treatment are equivalent, radical hysterectomy has decided advantages over radiotherapy: preserving ovarian function; affording superior vaginal and sexual function; allowing comprehensive surgical staging.

Furthermore, surgery avoids the nuisances and greater complications of extensive (i.e. radical) radiotherapy, such as long-term morbidity to the gastrointestinal and genitourinary tracts, and radiation-induced secondary malignancies. For its part, radical hysterectomy carries possible morbidity (some severe) similar to that encountered with other major pelvic operations, including ureteral injury. This acknowledged, in experienced hands, the operation is usually safe and effective, and can yield 5-year survivals in excess of 90 per cent with stage Ib carcinoma of the cervix.

Technique

Examination under anesthesia and opening the abdomen

Under general anesthesia, with a three-way urinary catheter in place, the pelvis is examined in search of clinical findings that would contraindicate surgery, such as obvious paravaginal or parametrial neoplasm. Assuming that no contraindication exists, the abdomen is opened through an infraumbilical midline incision (Fig. 1) or at times via a Pfannenstiel incision. Upon entering the peritoneum, normal saline is instilled into the abdomen and pelvis and is then recovered for histocytologic evaluation.

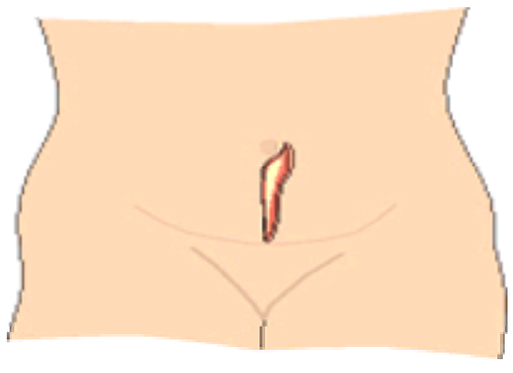


Fig. 1. Infraumbilical/abdominal approach to the pelvis.

Exploration of the abdomen

A systematic exploration of the abdomen is begun, starting in the upper part, where the visceral and peritoneal surfaces are inspected and palpated in search of metastatic disease. Given the biologic behavior of cervical cancer, special attention is given to the para-aortic lymph nodes as a potential site of metastasis. Any suspicious findings should be removed for examination. If results are positive, although some controversy exists, the operation is generally contraindicated, since under these circumstances (i.e., advanced surgical stage) surgery alone is not curative. But, evolving evidence of the effectiveness of radiation therapy given in conjunction with sensitizing chemotherapy after radical hysterectomy leaves open the additive value of removing surgically the central/primary disease. The final answer to this intraoperative dilemma (i.e., to complete the radical hysterectomy or not) awaits randomized trials.

Exploration of the pelvis

If no disease is detected in the upper abdomen, a self-retaining retractor is placed in the incision and the bowel is packed away. The uterus is grasped by Kelly clamps across the insertion of the utero-ovarian ligament, fallopian tube, and round ligament. The round ligament is also clamped laterally, close to the inguinal ring, and is then divided in the center, which in effect transects the broad ligament and enters the retroperitoneum. The anterior leaf of the broad ligament is further incised inferiorly and across the uterovesical fold, while the posterior leaf is opened superiorly and lateral to the infundibulopelvic ligament; if need be, this incision can be

extended still further lateral to the colon (Fig. 2). These peritoneal incisions allow the surgeon to dissect the bladder off the cervix and upper vagina, and then to expose fully and to dissect the pelvic retroperitoneum inferiorly and posteriorly into the avascular paravesical and pararectal spaces. In so doing, a thorough exploration for advanced pelvic malignancy can be made. If any is found, further surgery may be contraindicated (although the same provisos pertain as discussed above for para-aortic adenopathy). The condensation of tissue separating the paravesical and pararectal spaces and connecting the lateral cervix to the pelvic side-wall is commonly referred to as the ‘web'; in proper anatomic terms it is the cardinal ligament.

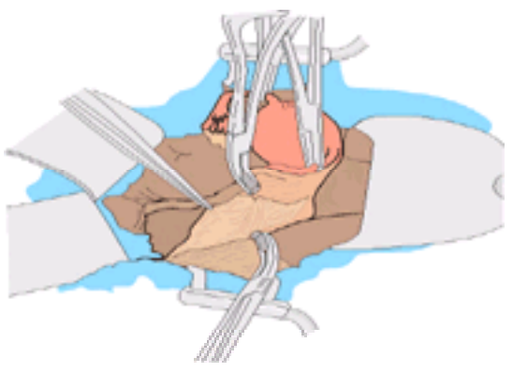


Fig. 2. Transection of the round ligament to expose the retroperitoneal pelvic spaces.

The ovaries

If the patient wants to retain her ovarian function, the adnexa should be divided at the utero-ovarian ligament and, with the ureter directly visualized, the ovary with its vascular pedicle is mobilized by dissecting superiorly. This accomplished, the ovary can be pexed superiorly, high above the standard pelvic radiation field, a placement that often permits continued ovarian function, even if adjuvant radiation therapy to the pelvis becomes necessary. Occasionally, however, ovaries relocated to this unnatural retroperitoneal site can become problematic as they form cysts, physiologic and otherwise, which can be symptomatic and create diagnostic dilemmas. If the patient does not want to retain ovarian function, the vascular pedicle can be clamped, cut, and ligated (Fig. 3).

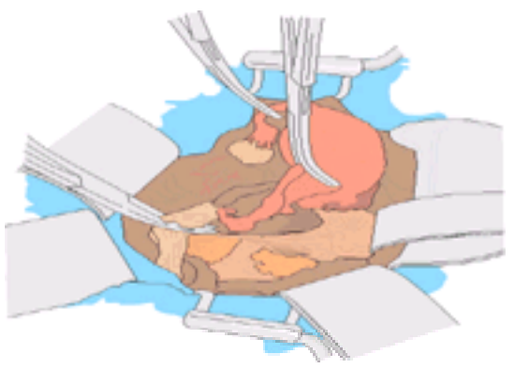


Fig. 3. Ligation of the infundibular ovarian vascular pedicle.

Pelvic lymphadenectomy

A pelvic lymphadenectomy, facilitated by hemostasis and marking the area of involvement with titanium surgical clips and electrocautery, is performed along the iliac vessels, starting from the mid-common iliac artery and moving inferiorly. The dissection is taken into the obturator space, where the obturator nerve is identified beneath the external iliac vein. All lymph-bearing tissue between the obturator nerve and the obliterated hypogastric artery is removed (Fig. 4). As part of the lymphadenectomy, the uterine artery is identified at its origin from the internal iliac, where it is doubly clamped, cut, ligated or clipped, then dissected medially (Fig. 5). Using this systematic approach, a total bilateral pelvic lymphadenectomy is performed for prognostic, and possibly therapeutic, advantages. Some surgeons elect to perform aspects of the lymphadenectomy after the radical hysterectomy has been completed. Doing so creates a more open surgical field, making the lymphadenectomy easier. In that scenario the the lymphadenectomy is initially carried out to the extentof identifying the vessels and landmarks as required for the radical hysterectomy to be properly performed. The remainder of the lymphadenectomy is then carried out after the uterus has been removed. Any suspicious lymph nodes (e.g. enlarged) would be sampled early on, however.

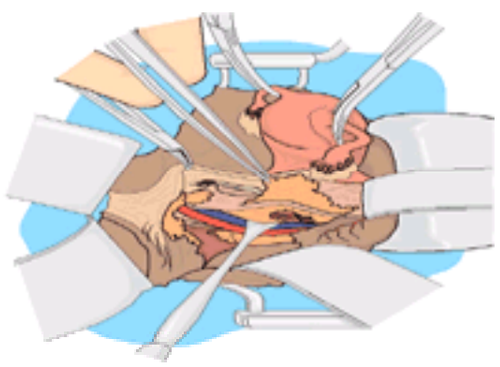


Fig. 4. Pelvic lymphadenectomy removing all node-bearing tissues along all major vessels.

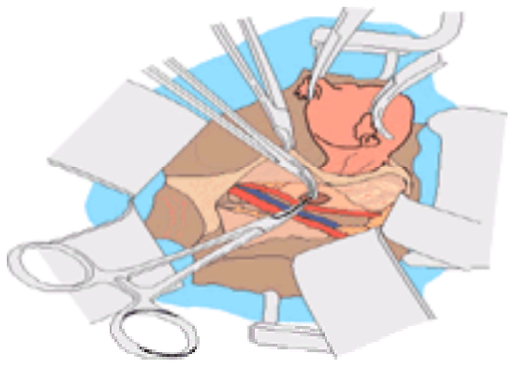


Fig. 5. Ligation of the uterine artery at its take-off from the internal iliac artery.

Dissection of the superior pelvic ureter

As part of the retroperitoneal dissection, the superior portion of the pelvic ureter is mobilized free of its peritoneal attachments, starting below the pelvic brim, moving inferiorly to the level at which it intersects with the uterine artery, which has been transected and moved medially, and is then dissected over the ureter (‘water under the bridge') (Fig. 6). The blood supply to the ureter is conserved.



Fig. 6. Dissection of the superior pelvic ureter while on Penrose traction.

Mobilization of the uterus

In the Okabayashi modification of the radical hysterectomy, before the ureter is completely free of its inferior attachments, the rectovaginal septum is opened by incising the peritoneum to the cervix and dissecting the rectum from the upper two-thirds of the vagina. This separation allows complete exposure of the uterosacral ligaments; these are doubly clamped, cut, and suture ligated at their insertion lateral to the rectum (Fig. 7). This maneuver allows some mobilization of the uterus, releasing it anteriorly and permitting better visualization of the course of the ureter, as well as better surgical access to the specimen for the remainder of the case.

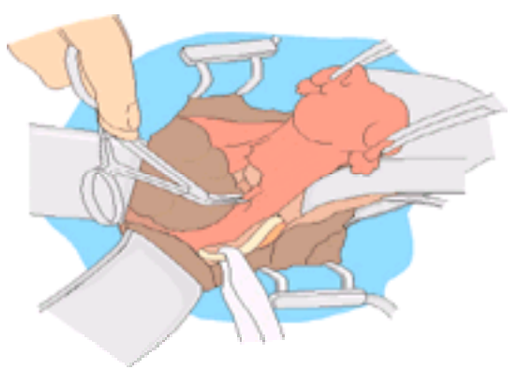


Fig. 7. Ligation of the uterosacral ligaments.

Dissection of the inferior pelvic ureter

With the uterus mobilized and retracted anteriorly the inferior ureter can be more readily dissected free of its attachments, beyond its intersection with the uterine artery, through its ‘tunnel', to its final insertion into the bladder (Fig. 8). This delicate dissection is facilitated by the use of right-angle clamps and fine ligatures. At completion the ureter is completely mobilized, but not denuded, from the upper pelvis to its bladder entry, and lies in the lateral pelvis.

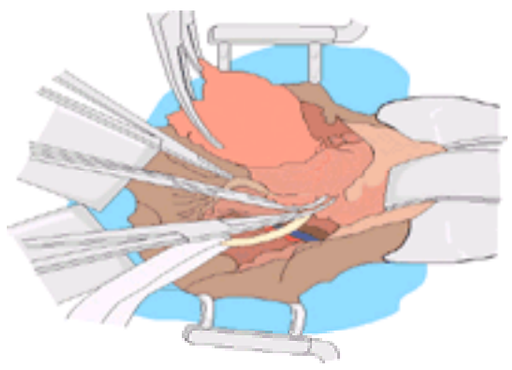


Fig. 8. Dissection of the inferior pelvic ureter.

Transection of the parametrium and paracervical tissues

The ureter having been dissected free from the ultimate surgical specimen, the cardinal ligaments can now, in a stepwise fashion, be double clamped as far lateral as the vessels on the pelvic side-wall, cut, and suture ligated (Fig. 9). This transection is carried inferiomedially along the base of the cardinal ligament to the appropriate level at which the vaginectomy is to be done.

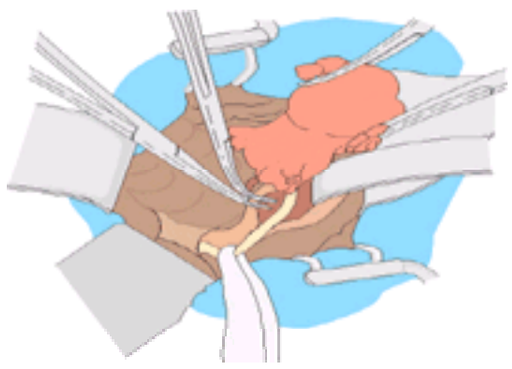


Fig. 9. Transection of all paracervical tissue beneath the inferior pelvic ureter.

Vaginectomy

At this point, the entire specimen has been surgically freed, with the exception of its attachment to the vagina. The vagina is transected at least 3 or 4 cm below the clinically evident disease so that negative surgical margins can be obtained (Fig. 10). Surgical stapling, using absorbable staples, is an expeditious alternative to traditional clamping and suturing as a technique actually to transect and to close the vagina. The vaginectomy being complete, the specimen, consisting of the uterus, cervix, upper vagina, parametrium, and paracervical tissues, is passed off the field.

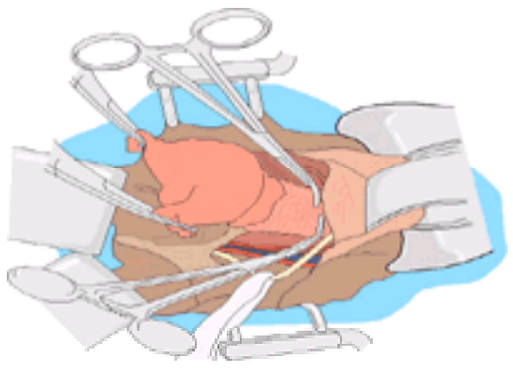


Fig. 10. Upper vaginectomy to remove the radical hysterectomy specimen.

Suprapubic catheter and drains

Because prolonged bladder dysfunction is inherent in the extensive pelvic dissection fundamental to radical hysterectomy, some surgeons place a suprapubic catheter prior to closing the abdomen. Suprapubic catheters are generally more comfortable for patients than is prolonged transurethral catheterization. After radical hysterectomy, 2 to 6 weeks of bladder decompression can be needed, awaiting the return of normal voiding. Typically, bilateral, percutaneous, closed suction drains are placed into the lateral retroperitoneal areas inferiorly beneath the vaginal cuff to evacuate the sometimes large, but evanescent, volumes of lymphatic fluids that pool after lymphadenectomy. The peritoneum is left open and a continuous mass closure is used for the abdominal incision (Fig. 11). Figure 12 demonstrates the gross appearance of a radical hysterectomy specimen.

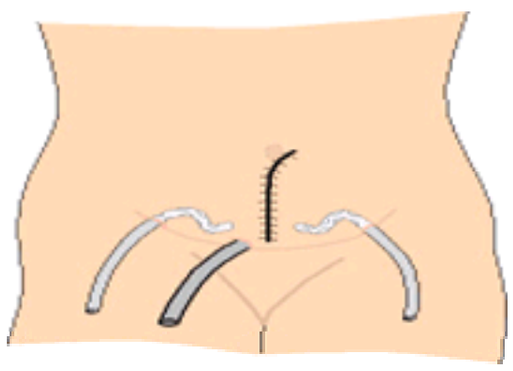


Fig. 11. Suprapubic catheter in the urinary bladder and two closed suction drains in the retroperitoneal spaces.

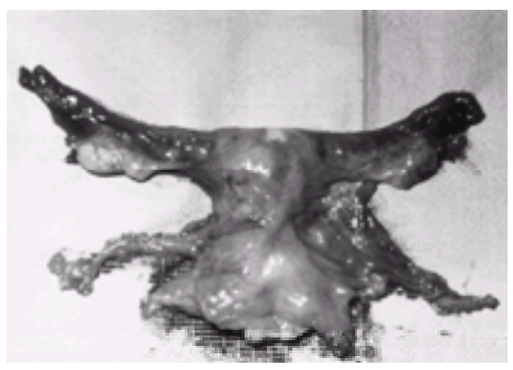


Fig. 12. Gross specimen of a radical hysterectomy; note the paracervical and parametrial tissue removed. The upper third of the vagina is included.

Morbidity

Although some major and minor morbidity exists for radical surgery of all types, there are specific problems/complications associated with radical hysterectomy. As mentioned, devascularization of the ureter can lead to fibrosis, obstruction, and fistula formation. To help prevent this classic complication, the ureter should not be denuded of its longitudinal microvascular sheath during its dissection from the peritoneum. Another notable but not unique risk is major hemorrhage. The pelvic veins are notorious for being clinically obscure, but when they are cut, bleed profusely. Prolonged tamponade is often the only means of controlling the inaccessible and life-threatening hemorrhage that can result. The mean operative blood loss, even with properly performed radical hysterectomies, frequently exceeds 1000 ml. As part of the extensive dissections, injury to the obturator nerve can occur; this affects the patient's ability to abduct her leg. Care in retraction and during lymphadenectomy is needed to reduce the chance of this and other neuropathy. Substantial lymphocysts can develop as a consequence of the lymphadenectomy, but are less common if percutaneous closed-suction catheters are left *in situ* until the drainage is minimal and the retroperitoneum is left open.

Further reading

Abitol MM, Davenport JH. Sexual dysfunction after therapy for cervical carcinoma. *American Journal of Obstetrics and Gynecology* 1974; **119**: 181.

American Cancer Society. Cancer statistics 1990. *Ca—a cancer journal for clinicians* 1990; **40**.

American College of Obstetricians and Gynecologists. *Clinical review*, Vol. 3(3), May–June 1998.

***Annual report on the results of treatment in gynecological cancer*, Vol.** 20. International Federation of Gynecology and Obstetrics, Stockholm, 1988.

Gauthier P *et al.* Identification of histopathologic risk groups in stage IB squamous cell carcinoma of the cervix. *Obstetrics and Gynecology* 1985; **66**: 569.

Granai CO, Gajewski WH. Gynecologic cancer. In: Skee RT, Lachant NA, eds. *Handbook of cancer chemotherapy*, pp. 288–317. Little, Brown Co., 1999.

Gusberg SB, Shingleton HM. Diagnosis and principles of treatment of cancer of the cervix. In: Gusberg SB, Shingleton HM, Deppe G, ed. *Female genital cancer*, p. 275. Churchill Livingstone, New

York, 1988.

Hoffman M, Roberts WS, Cavanagh DC. Are radiation-induced cancers real? *Contributions to Obstetrics and Gynecology* 1988; **31**: 45.

Meigs JV. Carcinoma of the cervix—the Wertheim operation. *Surgery, Gynecology, Obstetrics* 1944; **78**: 192.

Morley GW, Seski JC. Radical pelvic surgery versus radiation therapy for stage I carcinoma of the cervix (exclusive of microinvasion). *American Journal of Obstetrics and Gynecology* 1976; **126**: 785.

Morrow CP, Townsend DE, ed. Cancer of the vagina. In: *Synopsis of gynecologic oncology*, 2nd edn. Wiley, New York, 1981.

Newton M. Radical hysterectomy or radiotherapy for stage I cervical cancer: a prospective comparison with 5 and 10 year follow-up. *American Journal of Obstetrics and Gynecology* 1975; **123**: 535.

***NIH consensus statement*, Vol.** 14(1), *Cervical cancer*, April 1996.

Okabayashi H. Radical abdominal hysterectomy for cancer of the cervix uteri. *Surgery, Gynecology, Obstetrics* 1921; **33**: 335.

Orr JW, Jr. Cervical cancer. *Surgical Oncology Clinics of North America* 1998; **7**(2) (*Cancers unique to women*): 299–316.

Orr JW, Jr, *et al.* Correlation of peri-operative morbidity and conization–radical hysterectomy interval. *Obstetrics and Gynecology* 1982; **59**: 726.

Rutledge FN. The role of radical hysterectomy in adenocarcinoma of the endometrium. *Gynecologic Oncology* 1974; **2**: 331.

SGO Medical Practice and Ethics Committee. *Clinical practice guidelines, management of gynecologic cancers*, October 1996.

Shingleton HM, Kim RY. Treatment of cancer of the cervix. In: Gusberg SB, Shingleton HM, Deppe G, ed. *Female genital cancer*, p. 297. Churchill Livingstone, New York, 1988.

Shingleton HM, Orr JW, Jr. *Cancer of the cervix: its diagnosis and treatment*, 2nd edn. Churchill Livingstone, Edinburgh, 1987.

Strockbine MF, Hancock JG, Fletcher GH. Complications in 831 patients with squamous cell carcinoma of the intact cervix treated with 3000 rads or more whole pelvic irradiation. *American Journal of Roentgenology* 1970; **108**: 293.

Tarraza HM, Ellerkmann RM. Pelvic radical surgery. *Surgical Oncology Clinics of North America* 1998; **7** (*Cancers unique to women*): 399.

Taussig FJ. Iliac lymphadenectomy for group II cancer of the cervix. *American Journal of Obstetrics and Gynecology* 1943; **45**: 733.

Wertheim E. The extent of the abdominal operation for carcinoma. *American Journal of Obstetrics* 1912, **66**: 169.

36.9 Perineal abscesses

William D. Boyd and F. Mark Charnock

Introduction
Surgery of Bartholin's gland
Aspiration and antibiotic treatment
Incision and drainage
Marsupialization
Excision
Further reading

Introduction

A number of conditions can present as perineal abscesses, the most common being a Bartholin's abscess due to blockage of the duct of the gland with the subsequent infection of the duct and/or cyst. Table 1 lists some of the less common conditions that can present as a perineal abscess.

Systemic conditions:
Behçet's syndrome
Crohn's disease
Local conditions:
Hidradenitis suppurativa (Fox–Fordyce disease)
Infected perineal laceration, episiotomy
Epidermoid and implantation dermoid cysts
Tuberculosis
Herpetiform lesions (infected)
Bowel fistulas

Table 1 Differential diagnoses of a perineal abscess

Surgery of Bartholin's gland

A Bartholin's cyst is a cyst of the duct and not of the gland. The gland is usually compressed around the periphery of the cyst. Surgical excision of the cyst can be associated with a considerable blood loss due to the vascularity of the area. The remaining gland, which will still be functional, may reaccumulate and present as a symptomatic cyst, and need re-excision. As in other areas of the body when a cyst has become infected, incision and drainage rather than attempted excision is the wiser of the available treatment options, which are:

1. aspiration and antibiotic treatment;
2. incision (cruciate or linear) and drainage (wick, Word catheter, Foley catheter);
3. marsupialization;
4. excision.

All of these procedures can be performed under local anaesthetic, except perhaps excision. However, as the abscess and the area around it are usually very tender, most women will opt for general anaesthesia. These procedures are all performed with the patient in the lithotomy position.

Aspiration and antibiotic treatment

In the very early stages of infection of a Bartholin's gland a course of antibiotic therapy may be sufficient treatment. However, once a fluctuant abscess has formed, drainage is needed. Aspiration can be performed with a 16- or 18-gauge needle under local anaesthetic. The patient should be started on a course of metronidazole and either a penicillin or a cephalosporin antibiotic. The contents of the cyst should be examined by microbiological culture. If the results of the culture suggest a particular organism not covered by the above combination, then the appropriate changes should be made. Up to 80 per cent return of function of the gland has been reported with aspiration and antibiotic therapy.

Incision and drainage

Incision and drainage is the standard treatment for any abscess. The incision should be made in the vestibular area close to the hymen through an area of fluctuance (Fig. 1(a)). A drain (corrugated, red rubber drain) or a wick (1.5-cm ribbon gauze) is then inserted for 24 h. Because the skin has a tendency to reseal, recurrence is common. Therefore, removing an ellipse of skin (Fig. 1(b)) may help keep the edges apart and prevent this recurrence. Likewise, making a cruciate incision over the area of maximum fluctuance and either suturing back or excising the skin edges may also prevent the aperture resealing.

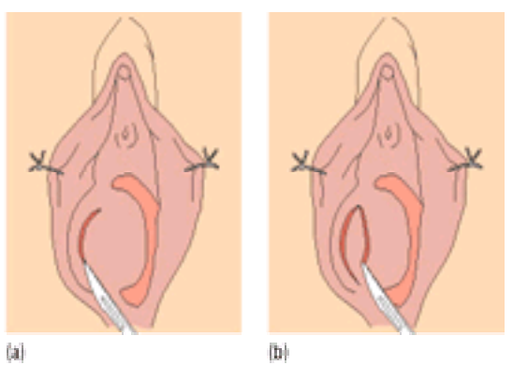


Fig. 1. Treatment of Bartholin's abscess: (a) incision; (b) drainage.

A Word or a Foley catheter can be inserted through a stab incision into the abscess cavity and the self-retaining bulb inflated with saline. The free end of the Word catheter is tucked away into the vagina (Fig. 2). In the case of the Foley catheter a ligature is placed around the catheter approx. 7.5 cm from the bulb, occluding all lumens. The catheter is then cut close to the ligature. This end is placed in the vagina. The catheter is left in place for 3 to 4 weeks until the tract of the wound becomes covered with new epithelium and so forms a new duct and maintains gland function.

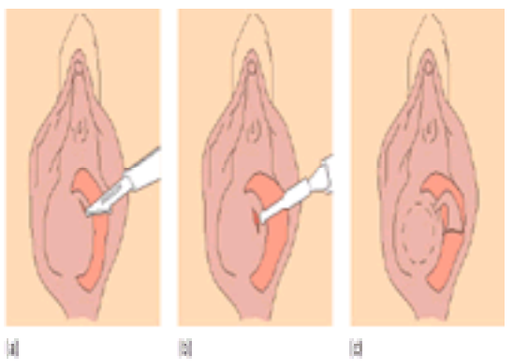


Fig. 2. Insertion of Word or Foley catheter in treatment of Bartholin's abscess.

Marsupialization

Marsupialization of a Bartholin's abscess also allows some gland function to be retained. An incision is made as described above, either linear or cruciate (Fig. 1(b)). The cyst wall is then sutured to the skin edges with interrupted plain 2/0 catgut. An effort should be made to leave a large aperture, as this will shrink as the abscess cavity shrinks.

Excision

Excision of a Bartholin's abscess can be difficult as the abscess wall is very adherent to the surrounding tissues and the usual tissue planes are not easily defined. A brisk blood loss can ensue and a drain may be needed to prevent haematoma formation and blood transfusion. A linear incision is fashioned as described above. The abscess is then enucleated using a combination of sharp and blunt dissection with curved dissecting scissors. Haemostasis must be meticulous, using diathermy and fine ties. Once the abscess has been removed the dead space left behind is obliterated using interrupted absorbable sutures and interrupted sutures to the skin as well (Fig. 1, Fig. 3). If, while dissecting the abscess, the cavity is inadvertently opened, marsupialization can be performed.

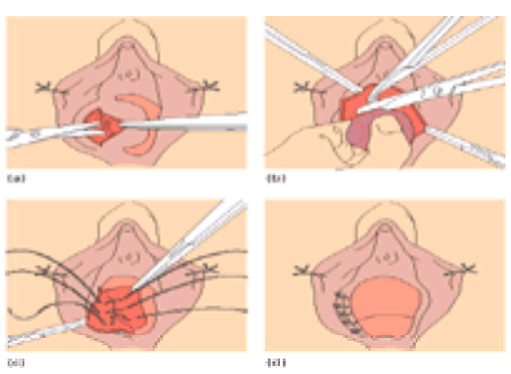


Fig. 3. Excision of Bartholin's abscess.

The patient should be encouraged to use the bath or bidet frequently but thereafter to keep the vulval area clean and dry. If there is a solid area within the abscess cavity a biopsy specimen should be taken and sent for histological examination to exclude malignancy. In any woman over the age of 40 years, malignancy should be seriously considered.

Further reading

Axe S *et al.* Adenomas in minor vestibular glands. *Obstetrics and Gynaecology* 1986; **68**: 16.

Cheetham DR. Bartholin's cyst: marsupialisation or aspiration? *American Journal of Obstetrics and Gynecology* 1985; **152**: 569–70.

Monaghan JM. Operations on the vulva. In: Monaghan JM, ed. *Bonney's gynaecological surgery*, pp. 119–21. Baillière Tindall, London, 1986.

Nichols DH, Randall CL. Minor and ambulatory surgery. In: Nichols DH, Randall CL, ed. *Vaginal surgery*, 3rd edn, pp. 163–5. Williams and Wilkins, New York, 1989.

36.10 Late spontaneous abortion

I. Z. Mackenzie

Introduction
Clinical presentation
Investigations and management
Threatened abortion
Inevitable abortion
Missed abortion
Surgical evacuation of the placenta or piece of placenta
Trophoblastic tumours
Factors associated with spontaneous late abortion
Incidence and recurrence
Further reading

Introduction

Spontaneous abortions occurring after the first 13 weeks of gestation will be considered, as these sometimes represent problems for the general surgeon. The strict definition of abortion is determined by law and thus may vary between countries: most commonly it is up to 28 weeks, although in the United Kingdom it was reduced to 24 weeks in 1991. However, it is modern obstetric and gynaecological practice to view bleeding up to 24 weeks as threatened abortion and thereafter antepartum haemorrhage, with the fetus having a chance of survival under intensive care. Therefore management is, to a certain extent, dictated by the circumstances at presentation and, in particular, the gestation period.

Clinical presentation

In most cases, pregnancy will have been suspected by the patient and subsequently confirmed. The primary presenting feature is vaginal bleeding, often with some associated lower abdominal pain, usually described as similar to menstrual cramps. The degree of discomfort is often a pointer to the stage that the process has reached, although this can be misleading. Similarly, the volume of blood lost through the vagina, as reported by the patient and observed by the physician, is often indicative of an inevitable abortion, but, on occasion, may be misleading. Unless the pregnancy had already ceased growing and the fetus has died (a missed abortion), the usual systemic symptoms and breast signs of pregnancy are likely to be positive at the time of presentation. Fetal movements may have been felt if the gestation has reached 18 to 22 weeks in those pregnant for the first time, and 16 to 20 weeks in those who have been pregnant before.

Differential diagnoses for the vaginal bleeding to be considered from the history include an intrauterine molar pregnancy (trophoblastic tumour); an extrauterine pregnancy, which is likely to have implanted into the abdominal cavity once the second trimester has been reached; and sources of bleeding from the lower genital tract such as cervical polyps, cervicitis, and cervical carcinoma.

The clinical examination must include palpation of the lower abdomen to exclude signs suggestive of intraperitoneal bleeding (as would be found in an abdominal pregnancy or from a traumatized uterus resulting from a criminal assault upon pregnancy). The size of the uterus must be assessed, to compare with that expected from the calculation of duration of pregnancy using the first day of the last menstrual loss. Bimanual pelvic examination is necessary to confirm pregnancy, to determine uterine size as accurately as possible, and to assess the degree of cervical dilatation: if the cervix is closed, the abortion is threatened and if open, the abortion is inevitable. Lesions of the lower genital tract involving the vagina and cervix will be palpable or visible with the aid of a vaginal speculum.

Investigations and management

In all cases of abortion, blood should be checked for haemoglobin, group, and rhesus type. Patients who are rhesus (D) negative should receive intramuscular immunoglobulin prophylaxis within 60 h of the onset of bleeding; 250 iu should be given up to 20 weeks of gestation and 500 iu thereafter. A Kleihauer count on a sample of maternal blood taken at this time gives an approximate guide to the volume of any fetomaternal haemorrhage that may have occurred, and further doses of immunoglobulin can be given; this in general is only advised from 20 weeks' gestation.

Threatened abortion

In those cases where abortion is threatened an ultrasonographic examination is helpful, particularly if the uterus is smaller or larger than expected from the menstrual dates. Hopefully, this will demonstrate an apparently normal viable fetus of appropriate size for gestation. It may, however, demonstrate fetal death, with absent fetal cardiac pulsations and movements (missed abortion), or a molar pregnancy, with the classical ‘snowstorm' sonographic picture. If real-time ultrasound is not available, the small Doppler instruments can be used to demonstrate the presence of the fetal circulation. Circulation should be detected from 14 weeks' gestation and occasionally as early as 12 weeks, although difficulty may be encountered in obese women. A urine or serum immunological pregnancy test can be performed but is not particularly helpful, especially since the placenta may continue to secrete human chorionic gonadotropin after fetal death has occurred. Very high levels of gonadotropin, however, are suggestive of trophoblastic disease, and, once the diagnosis has been confirmed by ultrasound, the appropriate management should be instituted.

Bed rest with mild sedation forms the initial approach to management. The volume of vaginal blood loss should be noted. In most instances, heavy vaginal bleeding is associated with progression to an inevitable abortion. In a few cases, however, marked bleeding may occur and persist over days or weeks; this is usually associated with placental implantation over the cervix. In this instance, conservative measures are maintained and include transfusion when necessary. If blood loss persists and threatens the patient's health or life, termination of the pregnancy may be advised. This will either be as an abortion or delivery of the baby alive by inducing labour. Delivery by caesarean section might be considered if the gestation is 24 weeks or more and facilities for intensive neonatal care are available.

In the majority of cases, bleeding settles within a matter of days and mobilization can occur, with normal daily activities resuming shortly thereafter. It is common practice to caution against coitus until 2 weeks have elapsed after the fresh bleeding has stopped, although there is no good scientific evidence to support this advice. The use of oral tocolysis, such as salbutamol (8 mg three times a day), has been advocated in the later gestational period, and this needs to be maintained until well into the third trimester (36 weeks' gestation); again, there is no evidence that the prognosis is improved as a consequence. There is now considerable evidence of improved chances of neonatal survival and fewer complications with the prophylactic administration of steroids from 24 weeks' gestation if preterm delivery becomes inevitable or necessary. Dexamethasone 12 mg for two doses 24 h apart or betamethasone 8 mg for two doses 24 h apart and repeated weekly are usually prescribed and continued at that dose until about 34 weeks' gestation.

The abortion of one twin and the retention of the other, to be delivered more than 7 weeks later, has been described. Strategies used to enhance the prospects for the remaining twin have included cervical cerclage and prolonged bed rest; there is no objective evidence in support of these.

Inevitable abortion

Once the cervix is dilated, abortion is inevitable. Some gynaecologists have suggested inserting a cervical circumsuture. The value of this is doubtful and abortion frequently follows, although there are occasional reports of success. There remains the risk of cervical laceration if effacement and dilatation continue and the suture is not urgently removed. Almost invariably, uterine contractions will begin and will result in expulsion of the fetus and placenta. If contractions are slow to establish and the process becomes prolonged, or if the fetal membranes have ruptured and contractions do not start, the cervix being partially dilated, uterine activity can be augmented with intravenous oxytocics, such as oxytocin 100 mu/min, to speed up the process. Local administration of prostaglandin pessaries, such as prostaglandin E_2 (25 mg on one occasion) or the prostaglandin E_1 analogue gemeprost (1 mg at 3-hourly intervals), may be used. As an alternative, misoprostol 800 µg vaginally or 400 µg orally at 3-hourly intervals is a cheaper option, although the drug is not licensed for this purpose at present.

When the fetus is expelled, ergometrine (0.5 mg intramuscularly) should be given to try to ensure a firmly contracted uterus and help speed placental separation, reducing blood loss. In the absence of excessive bleeding it is appropriate to wait a few hours for the placenta to be expelled. If it is not forthcoming, or if it has been delivered incomplete, surgical evacuation will be necessary. Following evacuation, whether spontaneous or with the aid of surgery, the uterus should be well contracted. There should then be minimal bleeding, and provided the vital signs are stable, the patient can be discharged. Bromocriptine (2.5 mg daily for 14 days) will prevent lactation in patients who were more than 20 weeks' pregnant at the time of abortion, although this is often not necessary. Permission should be sought for

pathological examination of the fetus and placenta. The possibility of infective causes for the abortion (see Table 1) as well as fetal karyotypic abnormality should be considered and appropriate specimens collected for investigation.

Maternal systemic disease	Uterine causes	Fetal causes
Infection, Rubella, Listeria, cytomegalovirus, Toxoplasma, parvovirus, Campylobacter, Chlamydia, Mycoplasma, Vibrio, Treponema	Cervical incompetence, Congenital malformation, Fibroids, Intrauterine contraceptive device in situ	Polyhydramnios, Multiple pregnancies, Congenital anomalies, Karyotypic abnormalities
Trauma, accidental, iatrogenic		
Pre-existing disease, diabetes, renal, thyroid, hypertension		

Table 1 Aetiological factors associated with spontaneous second-trimester abortion

Follow-up should be offered to discuss the possible aetiological factors that may have been detected or suspected from the pathological examination of the fetus and placenta and other investigations, and to advise upon future pregnancy management.

Missed abortion

This should be suspected if the uterus is at least 4 weeks smaller than anticipated from the menstrual dates, pregnancy symptoms having disappeared, and fetal movements having ceased if previously present. Immunological pregnancy tests may not become negative for 3 to 4 weeks, but the diagnosis can be confirmed by ultrasonography. If the uterus is not larger than that equivalent to a 15-week gestation, evacuation should be managed by cervical dilatation and uterine aspiration (using a paracervical block, regional block, or general anaesthetic); with larger gestational sizes, evacuation should be achieved by induction of a miscarriage, using oxytocics. Some gynaecologists prefer the surgical approach up to 20 weeks' gestation and even later.

Uterine aspiration of a missed abortion

The administration of a prostaglandin such as prostaglandin E_2 10 mg, gemeprost 1 mg, or misoprostol 800 µg vaginally 2 to 3 h before surgery can assist the operation by softening and dilating the cervix and aids uterine contraction after the abortion, reducing blood loss. Alternative strategies include the insertion of cervical tents of laminaria, Dilapan®, or Lamicel® approximately 12 h before surgery, which are removed at the operation. After the induction of general or regional anaesthesia, vulval and vaginal preparation with aqueous chlorhexidine or a similar solution is followed by draping. If a paracervical block is chosen, 10 ml lignocaine (lidocaine) with 1:200 000 solution of adrenaline (epinephrine), or a similar anaesthetic, is injected into the cervical stroma on each side, avoiding the cervical branches of the uterine arteries. Bimanual examination is then performed to confirm the size and position of the uterus. For those patients not treated with prostaglandins preoperatively, intravenous ergometrine 0.5 mg or oxytocin 5 u are given to contract the uterus and reduce operative blood loss. The cervix is then stabilized with a sponge-holding forceps and dilated with graduated dilators to a maximum of 10 mm. The uterus is then aspirated with a curette of appropriate size (to a maximum of 10 mm). The procedure is completed when no further products of conception are seen escaping along the suction tubing and the cavity feels empty with the characteristic ‘grating' feeling experienced with the curette against the uterine wall. For the more advanced pregnancies managed surgically, cervical dilatation needs to be greater, often up to 14 mm. The fetus is removed piecemeal with ovum forceps, the cavity finally being fully evacuated with the aspiration equipment. Subsequent care is as described for spontaneous inevitable abortion. Histological examination of the tissue removed is generally only able to confirm apparently normal placental tissue and little else.

Prostaglandin induction of a late missed abortion

Probably the most appropriate and effective method of evacuating a missed abortion when the uterus is larger than a 15-week pregnancy involves the administration of prostaglandins intramuscularly or into the vagina. A number of different protocols have been described, and the most efficient are 15(s)-15-methylprostaglandin F_{2a} intramuscularly 6-hourly; prostaglandin E_2 25 mg vaginally plus intravenous oxytocin after 18 h; or the prostaglandin E_1 analogues gemeprost 1 mg or misoprostol 800 µg vaginally 3-hourly augmented with intravenous oxytocin after 18 h.

Treatment is maintained until expulsion of the pregnancy has occurred. Adequate analgesia, such as diamorphine (10 mg at 4-hourly intervals), should be provided as necessary: the miscarriage is managed as described for an inevitable abortion. If it is incomplete, a surgical evacuation to remove any retained products will be necessary.

Surgical evacuation of the placenta or piece of placenta

If expulsion of the pregnancy is incomplete, either following spontaneous abortion or after prostaglandin induction in cases of ruptured membranes or missed abortion, surgical evacuation under anaesthesia of any retained pieces of placenta will be necessary. A paracervical block is usually insufficient, especially to remove a placenta retained in the uterine fundus. Epidural, spinal, or general anaesthesia is preferred and should be followed by appropriate asepsis and draping; the bladder should be empty or catheterized if necessary. Removal of placental tissue is generally best accomplished by digital exploration of the uterine cavity, which reduces the chance of traumatic damage to the uterine wall: sharp instruments, such as a uterine sound, should be avoided for this reason. Following the digital removal of all placental tissue, gentle blunt curettage covering all the walls and the fundus of the uterus should be performed to confirm that the cavity is empty and to remove some of the decidua. Ergometrine 0.5 mg or oxytocin 5 u intravenously can be given at this stage to ensure a well-contracted uterus that is firm against the curette, and will help to reduce further unnecessary bleeding. On completion, having ensured that the uterus is well contracted and that bleeding is minimal, the volume of blood lost is estimated. If there is doubt about whether the placenta is normal, it should be sent for histological section, and, if sepsis is suspected, a sample should be sent for bacteriological culture. Prophylactic broad-spectrum antibiotics effective against aerobic and anaerobic organisms should be considered.

After recovery from the anaesthetic and once vital signs are stable, provided the uterus is well contracted and bleeding is minimal, the patient may be discharged from medical care.

Trophoblastic tumours

Trophoblastic disease, including hydatidiform mole and choriocarcinoma, may be suspected with a history of vaginal bleeding and excessive symptoms of pregnancy. The diagnosis is made by pelvic ultrasonography, and, once confirmed, a chest radiograph should be taken, followed by evacuation of the uterus. An intravenous line should be in place and a ready supply of cross-matched blood available, since heavy bleeding may occur at the time of abortion. The evacuation is managed as described for a missed abortion; in all cases, if prostaglandins have been used to provoke expulsion of the products of conception at the later gestational sizes, a post-abortion uterine curettage should be performed to ensure that all pieces of placental tissue have been removed. All products of conception recovered should be examined histologically to confirm the preoperative diagnosis and to determine the precise degree of differentiation of the tumour.

Follow-up should be arranged to institute repeated assays of b-human chorionic gonadotropin in maternal blood in order to monitor any persisting or recurrent disease. All cases should be registered with the appropriate tumour centre for advice on further management.

Factors associated with spontaneous late abortion

Table 1 lists some of the factors that have been associated with spontaneous abortion in the second trimester. In most instances, investigation of possible causes will only be considered once abortion has occurred. Specific post-abortion investigations to be considered should include the following.

1. Enquiry about possible maternal infections.
2. Detection and quantitation of maternal antibodies are worth considering, but require the testing of two blood samples at 3-week intervals to detect any change in titre, while some of the viral infections, such as parvovirus, are best diagnosed by culturing samples of placental tissue.
3. A glucose tolerance test is worth performing if there are other signs/symptoms of diabetes.

4. A hysterogram will demonstrate any developmental anomaly of the Müllerian ductal system, such as a septate or subseptate uterus, or more severe abnormalities, such as a bicornuate (Fig. 1) or a didelphic uterus: some of these anomalies are amenable to plastic operations. Intramural and submucosal fibroids distorting the endometrial cavity and possibly resulting in abortion will also show up on a hysterogram and could be removed by abdominal or vaginal myomectomy. Cervical incompetence is extremely difficult to diagnose, although it may be suspected if the radiopaque dye has disappeared from cavity of the uterus on a delayed film.



Fig. 1. Hysterographic illustration of a bicornuate uterus in a patient with recurrent late abortion.

5. Fetal conditions predisposing to abortion, if not evident at the time, should be detected at post-mortem examination, and karyotyping using fetal blood or skin will detect any chromosomal anomaly. When appropriate, parental karyotyping should be performed to check for the presence of chromosomal translocations.
6. Red-cell isoimmunization, if severe, can result in fetal death and abortion from about 20 weeks' gestation. Investigation for maternal red-cell antibodies should be made.

Specific reasons for the abortion will be detected in few cases and, for the majority of couples, reassurance can be given for future pregnancies.

Incidence and recurrence

Spontaneous abortions are often quoted as occurring in 10 per cent of recognized pregnancies; once 13 weeks' gestation has been reached the chance of a spontaneous abortion is 0.7 to 1.0 per cent. This rate appears to decline as pregnancy advances towards 28 weeks' gestation. Although spontaneous losses are more common in older women, the increase for second-trimester losses does not seem to be significant. Overall, the risk of the next pregnancy being expelled spontaneously as a late abortion or preterm labour is approximately 15 per cent, and increases to 30 per cent after two consecutive losses.

Further reading

Charles D, Hurry DJ. Cervical incompetence. In: Fuchs F, Stubblefield PG, ed. *Preterm birth. Causes, prevention and management*, pp. 98–111. Macmillan, New York, 1984. [Discusses the diagnosis and options for management of the incompetent cervix to preserve a pregnancy.]

Elder MG, Hendricks CH, ed. *Preterm labour*. Butterworth International Medical Reviews, London, 1981. [Considers the physiological mechanism, consequences, and management of late threatened abortion.]

Mackenzie IZ, Davies AJ, Embrey MP. Fetal death *in utero* managed with vaginal prostaglandin E_2 gel. *British Medical Journal* 1979; **i**: 1764. [Describes various protocols for the administration of prostaglandins vaginally for inducing abortion in cases of fetal death.]

Main DM, Main EK. Management of preterm labor and delivery. In: Gabbe SG, Niebyl JR, Simpson JL, ed. *Obstetrics—normal and abnormal pregnancy*, pp. 689–738. Churchill Livingstone, New York, 1986. [Describes the management of women presenting with spontaneous and threatened abortion and preterm labour.]

Stray-Pederson B, Stray-Pederson S. Etiologic factors and subsequent reproductive performance in 195 couples with a history of habitual abortion. *American Journal of Obstetrics and Gynecology* 1984; **148**: 140–6. [Considers the prognosis for women with a history of recurrent spontaneous abortion]

Tabor A, Philip J, Marson M, Bang J, Obel FB, Norgard-Pederson B. Randomised controlled trial of genetic amniocentesis in 4606 low-risk women. *Lancet* 1986; **i**: 1287–93. [The most objective study evaluating the effect of diagnostic amniocentesis upon subsequent abortion.]

Wallenburg HCS, Keirse MJNC, Fiere HMP, Blacquiere JF. Intramuscular administration of 15 (*s*)-15-methyl prostaglandin $F_{2\,a}$ for induction of labour in patients with fetal death. *British Journal of Obstetrics and Gynaecology* 1980; **87**: 203–9. [Describes the use of the intramuscular protocol of a prostaglandin analogue for inducing abortion in cases of fetal death.]

Woolfson J, Fay T, Bates A. Twins with 54 days between deliveries. Case report. *British Journal of Obstetrics and Gynaecology* 1983; **90**: 685–7. [A report of the possible value of inserting a cervical suture to close the cervix after the spontaneous expulsion of the first of twins, the second being delivered some weeks later with improved chances of survival.]

36.11 Caesarean section

Gordon M. Stirrat

Introduction
Trends in incidence
Contribution to maternal mortality and morbidity
Types of procedure
Clinical indications
Operative procedures for emergency caesarean section
Preoperative preparation
Anaesthesia for caesarean section
Operative techniques
Intraoperative problems
Blood transfusion and caesarean section
Postoperative care
Postoperative complications
Situations requiring special consideration
Maternal
Fetal
Mode of delivery in subsequent pregnancies
Further reading

Introduction

Caesarean section is the delivery of a child through incisions in the anterior abdominal wall and uterus. The origin of the word ‘caesarean' is uncertain, but probably derives from the Latin verb *caedere*, meaning ‘to cut'. The oldest authenticated record of a live child being born by this method was in Sicily in 508 BC. Up to the sixteenth century the operation was mainly carried out after death to save the life of the child, and this is still necessary today on rare occasions.

The first verifiable caesarean section on a living woman was performed in Wittenberg in Germany by Trautmann in 1610; the patient died 25 days later. The first such operation in Great Britain was in Edinburgh in 1737: the baby was stillborn and the mother died shortly afterwards. Mary Donelly, an Irish midwife, is credited with performing the first caesarean section in the British Isles in which the mother survived, in 1738. A 14-year-old woman carried out the first recorded procedure in the United States on herself in 1822. The first of the twins she was carrying was born normally but the second was born after she opened her own abdomen with a razor. The fate of the illegitimate twins is unknown but the woman survived after two doctors closed the abdominal wound.

It was not until the late nineteenth century that closure of the uterine incision became routine, with a consequent improvement in maternal mortality. The first satisfactory method for suturing the upper-segment incision was described by Sanger in 1822. Vertical incision of the uterus through the lower, rather than the upper, segment was first described by Osiander from Gottingen in 1805. Kehere, also in 1882, described the precursor of the transverse lower-segment operation; this incision was subsequently championed by Munro Kerr, the Scottish obstetrician, who first used it in 1911. However, it was almost 40 years before the reduced maternal mortality associated with the approach was generally accepted in Great Britain.

An elective caesarean section is one where a decision to effect delivery by this route has been made before the onset of labour. All the rest are, rather unsatisfactorily, classified as ‘planned' or ‘unplanned' emergency caesarean sections. The term ‘planned emergency' means that appropriate assessment and preparation had occurred, whereas in an ‘unplanned emergency' procedure, full preparation was not possible due to (to quote the United Kingdom Department of Health) ‘over-riding urgency in the management of the patient'. Emergency caesarean section is the main subject of this chapter.

Trends in incidence

Fig. 1 shows the marked variation in rates of caesarean section for the United States of America, England and Wales, and The Netherlands between 1965 and 1995, and for Chile from 1986, compiled from a variety of sources. Between 1968 and 1988 the incidence of caesarean section increased by factors of 4.6, 3.0, and 3.2 in the United States, England and Wales, and The Netherlands, respectively. By 1984, it was the most common surgical procedure in the United States. High rates of caesarean section are a feature of maternity care in parts of Latin America. Brazil had the highest recorded rate in the world, of 31 per cent, in 1980. In Chile, the rate in 1986 was already 28 per cent and had increased by one-third to 37 per cent in 1994! Several reasons have been suggested for the variation in rates within and between nations. Differences in obstetric characteristics of, and risk markers in, population served may be important in between-hospital comparisons but do not explain major national differences. Referral bias of ‘high risk' cases to ‘high-tech' centres will result in more interventions; but the uncritical application of technology tends to increase the rate with no demonstrable benefit. The rate tends to increase with maternal age, which cannot be explained by socioeconomic or age-related obstetric or medical risk factors. ‘Risk' may be perceived differently in older women. The direct correlation between socioeconomic status and rates for caesarean independent of maternal age, parity, birth weight, ethnicity, or pregnancy complications suggests that the rates are not an indicator of medical need.

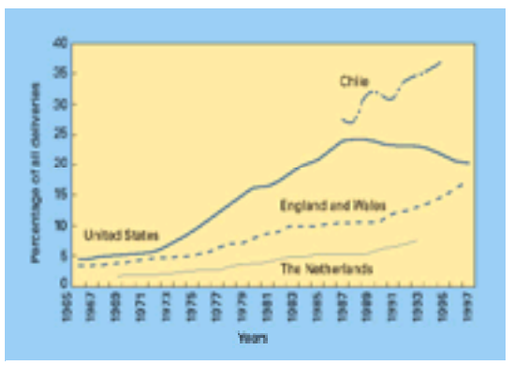


Fig. 1. Caesarean section rates for England and Wales, the United States of America, and The Netherlands 1965–1996, and for Chile from 1986.

Policies and attitudes of the obstetrician are major factors in determining mode of delivery. Fear of litigation is the most common reason cited for the increase in caesarean section rates. Previous caesarean section is the most common indication, despite evidence that it is frequently unjustified. As the rate of caesarean section increases for such obstetric problems as breech presentation, previous caesarean section, or malposition of the fetal head, the less able will succeeding generations of obstetricians be to use more conservative means. The rise in the rate of caesarean section in Chile correlates with the increasing number of births carried out under private insurance. A downward trend can be seen in the rate in the United States from 1988 but this decline has stalled from 1997. Rates have continued to increase in England and Wales and, possibly, The Netherlands. Contrary to expectations, this has been more due to a greater rate of increase in emergency than elective caesarean sections. In England, for example, the rates of emergency and elective caesarean sections, respectively, rose from about 6 and 5 per cent in 1989/90 to about 9 and 6 per cent in 1994/95. Lowering the rate to 15 per cent has become a national public health goal in the United States but this policy has been challenged. Among the strategies for doing so in the United States are:

- the presence of a nurse–midwife throughout labour
- 24-h cover by an experienced obstetrician
- agreed, evidence-based guidelines for the management of labour, and use of epidural anaesthesia
- greater (but appropriate) recourse to trial of labour after previous caesarean section
- audit of caesarean section rates for each obstetrician and comparisons with regional/national standards
- eliminating financial incentives.

To these may be added attempting external cephalic version at term for women with an otherwise uncomplicated breech presentation.

Contribution to maternal mortality and morbidity

Between 1991 and 1996, 104 (40 per cent) of the 262 maternal deaths in the United Kingdom directly attributable to childbirth were associated with caesarean section. Regrettably the total number of caesarean sections carried out is not known with accuracy. It is, therefore, not possible to state the fatality rate per procedure. It can be seen from Table 1 that the maternal mortality rate in the United Kingdom from 1985 to 1996 was, on average, four times greater for emergency caesarean section than for the elective procedure. Between 1990 and 1993 the rate was 6.4 times greater!The main causes in rough rank order remain pulmonary embolus, hypertensive disease, haemorrhage, and sepsis. Table 2 illustrates the extent to which ‘substandard care' contributed to the outcome.

	1985–1987	1988–1990	1991–1993	1994–1996	1985/1996
Elective	6 (13)	19 (33)	8 (14)	9 (20)	42 (30)
'Planned' emergency	25 (54)	32 (56)	38 (64)	26 (58)	121 (58)
'Unplanned' emergency	15 (33)	6 (11)	13 (22)	10 (22)	44 (21)
Total emergency	40 (87)	38 (66)	51 (86)	36 (80)	165 (80)
TOTAL	46	57	59	45	207

Percentages in parentheses, aggregate totals in bold type.
;'–ASQ-3 There is no complete description in the association file that matches this paragraph. Each style was converted individually.–;Source: 'Reports on Confidential Enquiries into Maternal Deaths in the United Kingdom 1985–97' (see Further Reading).

Table 1 Number of direct maternal deaths associated with emergency versus elective caesarean sections in the United Kingdom, 1985–1996

Factor	Concern
General anesthesia/intubation	Main cause. Difficult intubation or aspiration of gastric contents. No deaths associated with epidural anesthesia
Absence of intensive care facilities	The urgent operation is performed in maternity hospitals
Inadequate control of hypertension	Main therapeutic neglect factor
Failure responsibility to senior staff	Senior obstetrician and anaesthetic staff need to be informed and involved in time
Failure of junior staff to consult and/or to seek consultation with the colleagues	Senior obstetrician and anaesthetic staff need to be informed and involved in time
Operation done too late	Should be initiated against risk of long labour too early
Poor management of haemorrhage	Protocols for management of major obstetric haemorrhage must be established
Other aspects of poor management	For example, antenatal preparation, fluid balance, thromboembolism
Miscellaneous suggestions	No recommendations

Table 2 Major factors contributing to substandard care

Types of procedure

There are two types of procedure, the ‘classical' caesarean section, in which the upper uterine segment is incised longitudinally, and the lower-segment caesarean section. In the latter the lower segment is usually incised transversely, although a longitudinal incision starting in the lower segment is sometimes necessary.

Clinical indications

Caesarean section must not be carried out without good reason. It is justified when risk to the life and health of the baby and/or mother is greater if the pregnancy continues and if vaginal delivery cannot be achieved safely.

This situation can arise as a result of problems with uterine contractions, with the dimensions of the bony pelvis, or with the fetus. One of the common reasons given for carrying out an emergency caesarean section is ‘fetal distress'. This is difficult to define and often no more than an uncorroborated statement about concern for fetal welfare. Another reason given is ‘failure to progress in labour'. This often includes other indications, such as failed induction of labour, ‘abnormal uterine action', cephalopelvic disproportion, prolonged labour, and ‘maternal distress'. There is a strong iatrogenic element in both causation and interpretation in this category. Among other reasons are ante-partum haemorrhage, including both placenta praevia and placental abruption; failed ‘trial of labour' with a breech presentation; other malpresentations, such as transverse lie, face, or brow presentation; cord presentation or prolapse; and severe pre-eclampsia or eclampsia. It is rarely indicated in the context of actual or impending death of the mother. Fetal brain damage will begin about 8 to 10 min after failure of maternal circulation. If life support can be continued, or if a post-mortem caesarean section can be performed within that time, an attempt to save the fetus is justifiable.

Classical caesarean section is rarely necessary in modern practice. This option may, however, be used when a structural abnormality or space-occupying lesion (e.g. a fibroid) makes the lower-segment approach impossible. It is also applicable when the fetus is trapped in a transverse lie after rupture of the membranes and all the liquor has drained away; in some cases of anterior placenta praevia when the lower segment is excessively vascular; if hysterectomy is scheduled at the same time (very rare); or when the woman is already dead or moribund and rapid delivery of he baby is vital.

Operative procedures for emergency caesarean section

Preoperative preparation

Administrative and social

Several important conditions should, if at all possible, be fulfilled before proceeding. The woman, her partner, the surgeon, the anaesthetist, the paediatrician, the midwife, and all the nurses involved must know why the operation is thought to be necessary, and the paediatrician must be present in theatre for the delivery. If the operation is being done by an obstetrician in training, the case should be discussed with the senior obstetrician in charge of the woman's care or with the senior obstetrician on call. Written, informed consent should be obtained from the woman. If her consciousness is impaired, consent should be obtained from the next of kin. A fully conscious and otherwise competent woman has the right under common law to refuse permission for surgery even though, in the words of a decision by the United Kingdom Court of Appeal in 1997, ‘the consequence may be the death or serious handicap of the child she bears, or her own death'.

The woman's partner may ask to be present in theatre during the procedure in order to support her. This is usually acceptable if regional anaesthesia is being used and the woman is therefore awake, and if someone is available to look after the partner. Partners must understand that their presence is a privilege and not a right, and must agree to leave the operating theatre if asked to do so. Presence during surgery under general anaesthesia is less acceptable because the primary reason for attendance (i.e., support of the woman) cannot be fulfilled and it may be traumatic to see her unconscious and paralysed on a surgical table. Theatre staff are also more fully occupied and may not be able to look after the partner.

Pre-anaesthetic

Antacid therapy must be given to all women being (or at high risk of being) delivered by caesarean section. Magnesium trisilicate is ineffective. Suggested guidelines are given in Table 3. No other premedication is routinely necessary. Nasogastric tubes are not effective in emptying the stomach and, if not withdrawn before induction of anaesthesia, may increase the risk of aspiration.

Table 3 Guidelines for antacid therapy before caesarean section

Presurgical

Routine cross-matching of blood is not necessary provided that the woman's blood group and a recent haemoglobin concentration are known and a least 2 units of group O RhD-negative, Kell-negative blood are immediately available within the maternity unit. A blood sample should be sent to the blood transfusion laboratory in case cross-matching needs to be undertaken later. Guidelines for prior cross-matching of blood are suggested in Table 4.

Table 4 Guidelines for cross-matching of blood for caesarean section

Aseptic catheterization of the bladder should be performed in theatre (before induction of general anaesthesia) but an indwelling catheter is not routinely necessary. If a fetal scalp electrode is in place it should be removed. The operating table must be adjustable to allow head-down tilt. The mother should be placed on it with at least a 10° left lateral tilt to reduce the risk of maternal supine hypotension and fetal compromise. The management of women considered to be an inoculation risk is considered below.

The abdominal wall should be painted with a suitable antiseptic preparation from xiphisternum to midthigh and sterile towels draped around the area of the incision. Towel clips should not be fixed to the skin.

Anaesthesia for caesarean section

Factors that affect the conduct of anaesthesia during pregnancy include the physiological adaptations to pregnancy, such as an increased tendency to coagulation, smooth-muscle relaxation, which increases the risk of oesophageal reflux, and more rapid development of hypoxia during apnoea. There may be aorta–caval compression caused by the gravid uterus, and delayed stomach emptying in the patient who has been in labour. The need to maintain uterine tone in order to avoid catastrophic haemorrhage means that extreme caution is necessary if volatile anaesthetic agents are being considered. The risk of depressing fetal function and, therefore, neonatal state also limits the choice of drugs.

The choice of anaesthetic procedure for caesarean section is determined by the skill of the anaesthetist and surgeon, the degree of urgency, and informed patient preference. For an emergency, epidural anaesthesia started *de novo* is often too slow. The options are either to insert a spinal block or ‘top up' an epidural used during labour. Both of these will be rapid enough for some circumstances but only general anaesthesia will provide the surgeon with the ability to proceed within a few minutes of induction.

General anaesthesia

This has the advantages of rapidity of onset, a low failure rate, and operator control over respiratory, cardiovascular, and central nervous systems. However, its disadvantages include risks of failed intubation, gastric aspiration, maternal awareness, and depression of neonatal activity. General anaesthesia is indicated if the patient is hypovolaemic, semiconscious or convulsing, or if urgent surgery is required. It is contraindicated in the presence of maternal disease such as myasthenia or malignant hyperpyrexia.

For general anaesthesia to be used with the maximum degree of safety an intravenous line must be established; induction should take place in a fully equipped and staffed operating theatre; the patient should be preoxygenated for 3 to 4 min; cricoid pressure must be applied properly during induction and must be maintained until the endotracheal tube cuff is inflated to avoid aspiration; and a drill for failed intubation must be established.

Regional anaesthesia

The risks of general anaesthesia are avoided (unless total or high spinal block occurs), blood loss is decreased, and mother–infant bonding and breast feeding are facilitated. However, the block may fail, be inadequate, or may be too extensive. Regional anaesthesia is preferred if intubation has previously been difficult, if the patient has experienced an allergic response to previous general anaesthetic, or if a functioning block has been established earlier.

In order to make the procedure as safe as possible, both anaesthetist and assistant must be trained in the techniques to be used; full facilities must be prepared for induction of general anaesthesia and resuscitation; the patient must be prepared as if for general anaesthesia; patients should be given an intravenous ‘preload' of 1 litre of crystalloid; and the sensory block should cover from S4 to T4. The use of combined subarachnoid and epidural anaesthesia is increasing, but the decision to use one or other (or both) must be determined by the available facilities and skills, and the needs of the individual patients.

Operative techniques

There are many safe ways of performing a caesarean section: the techniques are best learnt from an experienced obstetrician.

Lower-segment caesarean section

The skin incision can be transverse or vertical (Fig. 2). The transverse incision may bleed more, but it is less painful, heals more quickly, is less liable to herniate, and the scar is usually hidden beneath the hairline. It is, therefore, generally preferred. A suprapubic midline incision may be useful for the delivery of three of more infants and is necessary for classical section.

If a transverse skin incision (Fig. 2(a)) is used, a straight incision (10–12 cm) is made approximately two finger breadths above the symphysis pubis. The subcutaneous tissue and rectus sheath are divided transversely and haemostasis secured (Fig. 3(a)). Subaponeurotic dissection of the sheath is then carried out and perforating vessels are ligated or diathermied. The rectus muscles are separated digitally in the midline to expose the parietal peritoneum (Fig. 3(b)).

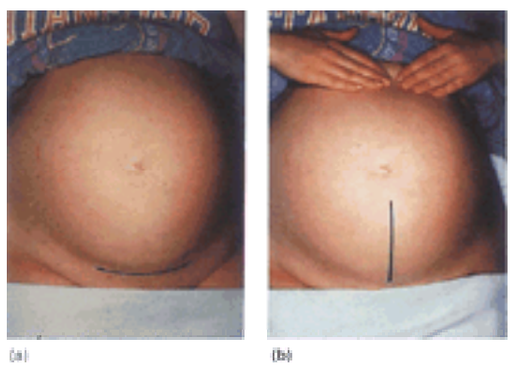


Fig. 2. (a) Line of transverse suprapubic skin incision; (b) line of midline subumbilical skin incision.

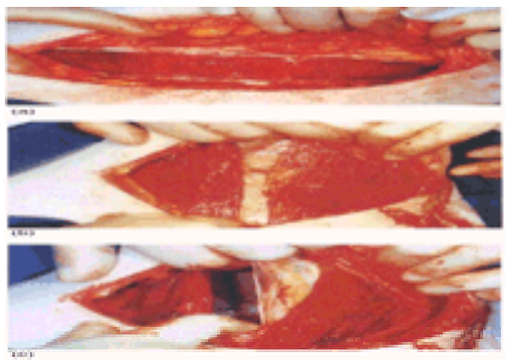


Fig. 3. (a) Rectus sheath incised bilaterally; (b) separation of rectus muscles to reveal parietal peritoneum; (c) division of parietal peritoneum longitudinally.

A midline incision (Fig. 2(b)) is performed as for other lower abdominal surgery. The parietal peritoneum is divided vertically (Fig. 3(c)), taking care not to damage bowel or bladder. Abdominal packs are unnecessary. A Doyen retractor is inserted suprapubically to widen access to the visceral peritoneum and retract the bladder. The lower uterine segment can now be seen (check for dextrorotation). The loose uterovesical peritoneum is divided transversely and the upper border of the bladder freed digitally and displaced caudally. The Doyen retractor is slipped into the retrovesical space. Using a scalpel, an 8- to 10-cm superficial incision (with a slight upward curve) is made in the lower segment (Fig. 4(a)). This is this is then deepened in the midline until the cavity is entered and fully extended by stretching with both index fingers (Fig. 4(b)). The Doyen retractor is then removed.

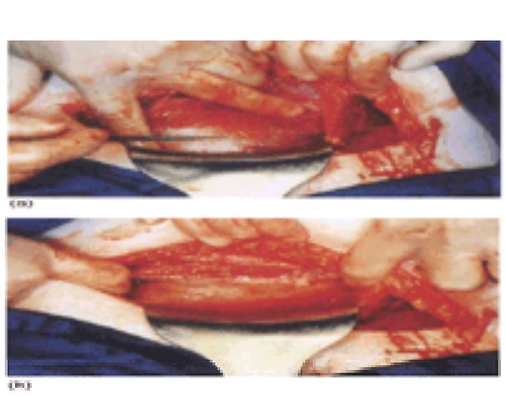


Fig. 4. (a) Incision of lower uterine segment; (b) completion of incision in lower segment to reveal presenting part.

Delivery of the baby

Undue haste makes extension of the uterine incision and heavy bleeding more likely. If the head presents, a hand is passed inside the uterus below and behind it to lift and guide it through the incision (Fig. 5). Suction of the mouth can now be performed. Fundal pressure may be exerted by the assistant once the head is free. Short obstetric forceps may be used for controlled delivery. If the head is moulded deeply in the pelvis it may be pushed up from below by an assistant.



Fig. 5. Delivery of fetal head.

All presentations other than cephalic are best delivered by the breech. The feet are grasped and delivered, and the trunk is delivered by traction and rotated slightly to the patient's left. This brings the left arm into view, which is delivered by hooking a finger into the elbow. The right arm is delivered similarly after rotating the trunk to the right. The face will now appear in the wound and oral suction can be performed. It is important to control delivery of the head using short obstetric forceps.

The baby is placed on the mother's abdomen (to prevent blood draining towards the placenta), where the umbilical cord is clamped and cut. The baby is then passed to the gowned paediatrician.

Delivery of the placenta

Syntocinon 5 units should be given intravenously when the head is delivered and the placenta should then be delivered by cord traction when the uterus contracts. Manual removal can be performed if necessary. The uterine cavity should be explored using a gauze swab wrapped around a finger to confirm complete removal of placenta and membranes.

Uterine closure

Closure can be facilitated by eventrating the uterus on to the anterior abdominal wall but this may be painful if epidural anaesthesia is being used. Otherwise, reinsertion of the Doyen retractor gives access.

The upper and lower borders and the lateral extremities (angles) of the incision are identified and held by Green-Armytage forceps. Care is needed: the upper edge will now be much thicker than the lower because of uterine retraction and the posterior wall is sometimes confused with the anterior edge of the incision. The uterus should be closed in two layers using No. 1 sutures on round-bodied needles (Fig. 6). Synthetic polyglycolic acid sutures (e.g. Vicryl®) are now preferred to chromic catgut. The first suture is placed to secure the right angle: it is cut long and held in the artery forceps. The left angle is then sutured and, with the short end held, the suture is continued across the inner two-thirds of the myometrium and tied into the suture on the right angle. The second suture starts at the left angle. An invaginating (Lembert type) suture can be used. The eventrated uterus is returned to the abdominal cavity and the incision is checked for persistent bleeding points, which are sutured to secure haemostasis. The visceral peritoneum is closed with a continuous suture as above. The fallopian tubes and ovaries are inspected, and swabs and instruments are checked.

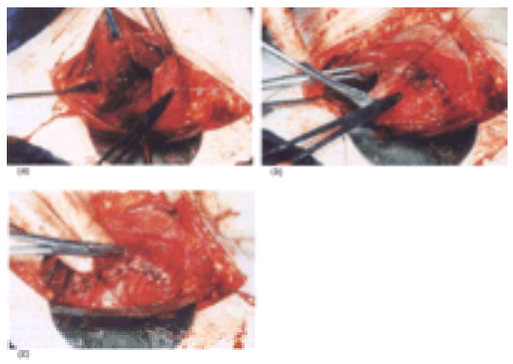


Fig. 6. (a) Suture in left angle of uterine incision; (b) first line of sutures in lower uterine segment; (c) second line of sutures in lower uterine segment.

Abdominal closure

The peritoneal cavity should be swabbed clear of blood and amniotic fluid, paying particular attention to the paracolic gutters. The abdomen is then closed in the conventional manner. The suture material used for closure is a matter for personal preference. Careful technique and scrupulous attention to haemostasis are far more important. Some operators advise routine subcutaneous and/or subfascial drainage of transverse incisions. A subcuticular stitch to the skin using prolene or one of the newer absorbable materials gives a good cosmetic result.

Classical caesarean section

The uterus is approached through a midline incision with the rectus sheath and peritoneum incised vertically. The upper segment of the uterus is incised in its lower half and in the less vascular midline. It is, therefore, necessary to check the position of the fundus and for dextrorotation before making the incision. Delivery of the baby is easier by the breech method whatever the presentation. Closure of the incision is as for the lower segment, except that three layers of suture are necessary, and conventional procedures for abdominal closure are followed.

Longitudinal incision in lower segment

A vertical incision starting in the lower segment may be necessary if the lower segment is undeveloped, for example when very preterm delivery by caesarean section is required.

A midline skin incision may be best. The lower segment is approached routinely and the visceral peritoneum incised transversely. The upper flap is reflected upwards and the bladder downwards. A small, vertical incision is made in the lower segment down to the membranes and the incision carefully enlarged caudally and cranially by scissors. Care must be taken to deliver the very premature baby as atraumatically as possible. Closure is as for the conventional lower-segment procedure.

Intraoperative problems

Transverse lie

If a conventional lower-segment caesarean section has been embarked upon in the presence of a transverse lie with ruptured membranes, it may not be possible to deliver the baby through the transverse uterine incision because the shoulder is likely to be impacted. On no account must the arm be delivered into the wound: this renders safe delivery impossible. An inverted-T incision into the upper segment should be made, although the scar heals less well than in either the transverse or vertical uterine incisions. Transverse lie with ruptured membranes is, therefore, a possible indication for a classical caesarean operation.

Placenta praevia

This seldom causes major problems for delivery, even if it is anterior. It is best to avoid incising the anterior placenta. The placental site in the lower segment tends to bleed heavily because it does not contract so well, but hot packs, pressure, and patience are usually sufficient to control the bleeding. Rarely, the placenta is morbidly adherent because of excessive invasion of the trophoblast (placenta accreta). This can give rise to major haemorrhage and should be dealt with as described below.

Haemorrhage from uterine vessels

The average blood loss at caesarean section is about 1000 ml. Haemorrhage is best prevented by precise and unhurried surgery and by careful delivery of the fetal head. Every centre must have an effective protocol to deal with major obstetric haemorrhage, such as that suggested in the Department of Health's 1988–90 ‘Report on Confidential Enquiries into Maternal Deaths in the United Kingdom'. Excessive bleeding is often due to trauma to the uterine vessels: deep and wide sutures usually control the bleeding but the ureters must be identified to prevent damage (see below). Hot packs and pressure should be applied for sufficient time to allow haemostasis and for preparations to be made to deal with major haemorrhage. If not already present, a senior obstetrician and anaesthetist should be called for and, if the situation is not under control, they must attend. The key to successful haemostasis is calm and purposeful action. Aortic pressure is useful as a temporary measure and can be life saving.

Ligation of the internal iliac (hypogastric) arteries and/or hysterectomy should be considered in patients who require intraoperative blood transfusion in the presence of significant continuing haemorrhage. These procedures must not be performed too late, but neither should they be undertaken too early.

A coagulation defect may be the primary cause of the haemorrhage (as in severe placental abruption) or may develop secondarily: this must be considered and dealt with. Other management is as for any major haemorrhage associated with surgery.

Damage to the lower segment

A thin or fibrous lower segment may be torn during delivery of the baby, particularly in patients who have undergone a previous caesarean section. The tear is usually in the midline towards the cervix and below the bladder, which must be dissected free and inspected for damage before repair of the tear.

Damage to the bladder or ureters

The bladder may be accidentally damaged during opening of the parietal peritoneum (particularly in patients who have had previous surgery), or after previous caesarean section. It may also be damaged if the lower segment tears. The damage must be recognized and repaired if fistula formation is to be avoided. Continuous bladder drainage is necessary for up to 10 days.

Damage to the ureters is uncommon but usually occurs during attempts to stop haemorrhage from uterine vessels or during repair of a damaged bladder. Recognition and immediate repair are vital. This is best achieved by a surgeon experienced in the technique.

Blood transfusion and caesarean section

Haemorrhage associated with caesarean section is a major world-wide cause of maternal morbidity and mortality. It was the second most common cause of caesarean section-related deaths in the United Kingdom from 1990 to 1993, accounting for 17 per cent of the total. Ekeroma *et al.* have reviewed the suggested guidelines for the use of blood products in surgery produced, among others, by the United States ‘Report of Health and Human Services' and Rosen *et al.* (see Further Reading). Table 5 lists those that apply directly to caesarean section.

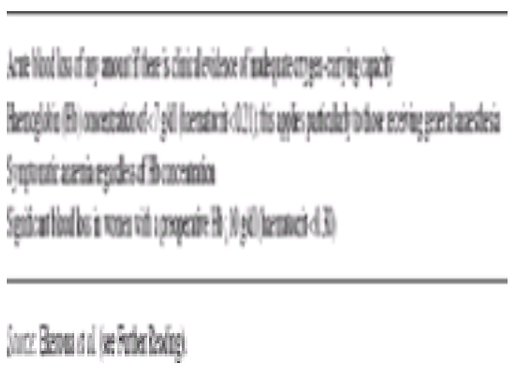


Table 5 Guidelines for blood transfusion in caesarean section

Postoperative care

The same principles apply as for general surgery. Aspiration and hypoxia must be prevented. Adequate pain relief is important, and early mobilization should be the aim, since this reduces the risk of thrombo-embolism. In the absence of postoperative complications, and if the baby is well, the woman can expect to go home about 5 days after delivery. Before discharge she should be counselled about contraception and management of any future deliveries.

Postoperative complications

Haemorrhage

Poor postoperative observation with failure to appreciate the degree of intra-abdominal and/or vaginal haemorrhage, failure to control it, and inadequate replacement are among the most common causes of serious sequelae of caesarean section. The anaesthetist and obstetrician should consult about the need to transfuse (see above) and, where possible, a haematologist should be involved in the management of massive haemorrhage. It is worth repeating here that it is imperative that clinical guidelines be established, made known, and frequently revisited for the management of massive obstetric haemorrhage from whatever cause including, of course, caesarean section. These should include management of those who have a conscientious objection to, or otherwise refuse, blood transfusion.

In the absence of any other reason for staying in hospital, a woman may be allowed home with a haemoglobin of as low as 7 g/dl (haematocrit 0.21) as long as she is not symptomatic, and is both receiving and taking adequate haematinic therapy.

Infection

Sepsis, which contributes significantly to morbidity, can be minimized by proper attention to sterility and surgical technique. Prophylactic administration of antibiotics reduces infective morbidity: broad-spectrum penicillins [e.g. amoxycillin (amoxicillin) with clavulanic acid] are as effective as cephalosporins. A higher incidence of side-effects and poorer compliance with longer courses have to be balanced against the lesser efficacy of short courses of drugs.

Thromboembolism

Women are at increased risk of thromboembolism after caesarean section, particularly if there is a past history of thrombosis, severe pre-eclampsia, or if the patient is obese. Suggested risk categories are shown in Table 6. Early mobilization is important for prevention, and high-risk patients should wear leg stockings and receive low-dose subcutaneous heparin until the fifth postoperative day. Management of established, deep venous thrombosis or pulmonary embolism is described elsewhere. Iodine-125 should be avoided for diagnosis because iodine is excreted in breast milk. Warfarin is not excreted in breast milk to any significant extent. Breast feeding is not, therefore, contraindicated during its use.

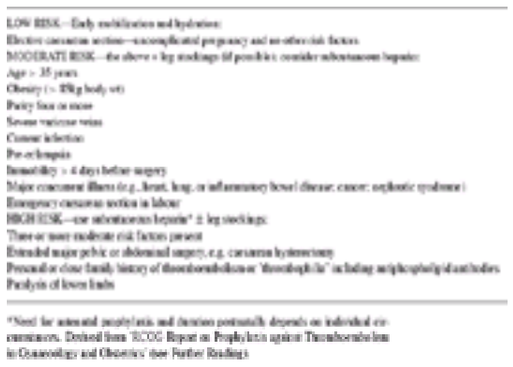


Table 6 Risk categories for thromboembolism

Wound dehiscence

This is generally preventable by good surgical technique. It is usually due to inadequate haemostasis and failure to drain ‘oozing' wounds.

Paralytic ileus

Postoperative distension of the abdomen is usually associated with caesarean section performed after very prolonged labour. Paralytic ileus is managed by gastric suction and administration of parental fluids until resolution is complete. Failure to treat may result in caecal perforation.

Psychological stress

Many women find emergency caesarean section to be particularly stressful. This should be borne in mind in their subsequent postnatal care.

Situations requiring special consideration

Maternal

Inoculation-risk women

Table 7 gives an indication of those women who have an increased likelihood of carrying hepatitis B and/or C virus, human immunodeficiency virus (**HIV**), or other agents spread by inoculation. Splashing of maternal blood on the faces of surgeons and assistants is common, and glove puncture with skin contamination occurs in about 1 in 3 caesarean sections. The incidence of percutaneous (‘needlestick' or ‘sharps') injuries is not known but is under-reported. Suture needles cause most of the injuries. The average risk of HIV infection after percutaneous exposure to HIV-infected blood is 0.3 per cent. Without vaccination, the risk of acquiring hepatitis B from an hepatitis B early antigen-positive patient after percutaneous injury may be as high as 30 per cent.

Known or suspected AIDS or ARC
HIV antibody-positive
HBsAg-positive/anti-HBe-negative
History of acute hepatitis in past 6 weeks[1]
Chronic active hepatitis or cirrhosis[1]
Resident in ‘high-risk area'—mainly in developing world
Childhood spent in developing world[2]
Having visited high-risk area in past 5 years[2]
Treated using blood products[2]
Sexual partners of HIV-risk men[2]
Intravenous drug abusers

AIDS, acquired immune deficiency syndrome; ARC, AIDS-related complex; HBe, s, hepatitis B early, surface antigen (Ag); HIV, human immunodeficiency virus.
[1]Can be excluded if found to be HBsAg-negative but HBeAg- and HBe antibody-positive
[2]Can be excluded if HIV-negative and more than 1 year since last exposure.

Table 7 Inoculation-risk women

Therefore, guidelines should be set down about the care of inoculation-risk patients. These must include immunization against hepatitis B virus and strategies for minimizing percutaneous injury. It is advised that those carrying out operative deliveries on inoculation-risk women should wear adequate eye protection, proper protective clothing, and double gloves. Body fluids should be handled with care and soiled garments disposed of as recommended in local inoculation-risk guidelines. Mouth-operated suction devices for the infant should be avoided.

Severe pre-eclampsia

The combination of emergency caesarean section and severe pre-eclampsia is a significant contributor to maternal mortality: the altered haemodynamics in pre-eclampsia are often not appreciated and proper precautions are not taken. Since it is a high-output state with reduced plasma volume, infusion of excessive fluid causes pulmonary oedema, which can be fatal. All patients with severe pre-eclampsia who are undergoing caesarean section must be carefully monitored postoperatively in a high-dependency area using at least a central venous pressure line and possibly a Swann–Ganz catheter. Intravenous infusion of crystalloid must be controlled very carefully. Oliguria is to be expected initially: diuretics are contraindicated since they merely reduce the circulating volume further.

Bleeding disorders or anticoagulants

Advice must be taken from a haematologist about women with an acquired coagulation defect or bleeding diathesis. Warfarin therapy is best avoided antenatally because it cannot be readily reversed (and is associated with fetal anomalies). Subcutaneous heparin does not increase the risk of bleeding and the action of intravenous heparin is reversible on administration of protamine sulphate. It can occasionally cause thrombocytopenia or osteopenia when use is prolonged.

Congenital or acquired cardiac disease

The risk of surgery to the mother must not be underestimated. Meticulous anaesthetic and surgical technique and access to intensive-care facilities are vital.

Other drugs

Special anaesthetic precautions need to be taken for women taking corticosteroids, b-agonists and antagonists, and antihypertensive agents.

Uterine anomalies

These are rarely severe enough to cause surgical problems. Classical caesarean section or a longitudinal incision in the narrow segment may be necessary if the uterus is not properly formed or is too narrow for safe transverse incision.

Fetal

Multiple pregnancy

This can usually be dealt with by lower-segment caesarean section, but a midline skin incision is advisable for higher-order multiples (triplets and above). The risks of maternal supine hypotension and postpartum haemorrhage are increased. Conjoined twins are very rare, but the possibility should be checked by ultrasound when twins are diagnosed. Classical caesarean section is necessary: if found unexpectedly the ‘inverted T' incision (see above) will be necessary.

Other fetal anomalies

A major hydrocephalus may make decompression of the head necessary before delivery. The diagnosis should have been made by preoperative ultrasonographic examination: a large but normal head may be mistaken for hydrocephalus and radiography may give a false-positive diagnosis.

Concurrent operations

Tubal ligation is a simple surgical addition to caesarean section but must be used sparingly because there is an increased risk of subsequent request for reversal. Prior counselling is mandatory and written permission from the woman absolutely necessary. The failure rate may be slightly greater even when tubal ligation is carried out

properly at caesarean section.

Caesarean hysterectomy is usually performed for control of major haemorrhage: the difficult decision must be made before it is too late for the patient to survive (see above). It should be avoided for other indications if at all possible and is not justifiable for sterilization. The temptation to remove subserous fibroids (myomectomy) at caesarean section must be resisted because bleeding from the incised pregnant uterus is difficult to control.

An accident to an ovarian cyst (such as rupture or torsion of a teratoma) is a rare cause of an acute abdomen during pregnancy. If it occurs after about 28 weeks, caesarean section is likely to be necessary to gain access. During such an operation for obstetric reasons, both ovaries should be checked visually. Neoplastic cysts should be removed but the surgery should be the least possible until a full histological diagnosis is made when, if necessary, definitive surgery can be planned. Histological examination of frozen sections is prone to overdiagnose malignancy in these circumstances.

Mode of delivery in subsequent pregnancies

This will depend on the cause of the previous caesarean section. If a ‘trial of labour' is to be allowed, it must take place in a hospital equipped to deal with emergencies (such as ruptured uterus and hypovolaemic shock) and able to proceed quickly to repeat caesarean section if necessary. Elective caesarean section is advised if the cause is clearly recurrent (e.g. severe cephalopelvic disproportion); there has previously been uterine rupture or damage to the bowel or bladder; or with a malpresentation. Labour must be managed with great care. A safe outcome for mother and baby is directly related to the quality of supervision of the ‘trial of labour'. If vaginal delivery is attempted, oxytocin must be used only after very careful consideration and with the strongest of indications. Epidural anaesthesia can be used; the pain of ruptured uterus will break through the epidural block.

Even in the late 1990s, one-third of caesarean sections in the United States were elective operations on women who had previously had caesarean sections. However, successful and safe vaginal delivery has been shown to be possible in between 50 and 75 per cent of women delivered by caesarean section in a previous pregnancy. Neither elective repeat caesarean section nor trial of labour are risk free. Length of hospital stay and the need for transfusion are greater, and the incidence of post-partum pyrexia is greater after repeat caesarean section. Trial of labour is associated with an increased risk of uterine rupture of less than 1 per cent in women with one previous section. The risk is three times greater in women labouring after two or more caesarean sections. Even in sub-Saharan Africa, where severe cephalopelvic disproportion is a greater problem than in developed countries, vaginal delivery can be achieved in between 50 and 70 per cent of women offered trial of labour. For trial of labour to be safe in this context, in addition to the facilities noted above, antibiotics, anticonvulsants, plasma expanders and, if at all possible, safe blood transfusion should be available. Transport is often a practical problem in the developing countries.

Further reading

Boulvain M, Fraser WD, Brisson-Carroll G, Faron G, Wollast E. Trial of labour after caesarean section in sub-Saharan Africa: a meta-analysis. *British Journal of Obstetrics and Gynaecology* 1997; **104**: 1385–90. [This is not a true meta-analysis, since it reviews 17 non-randomized and retrospective studies, but it provides useful clinical information.]

Department of Health. *NHS maternity statistics, England: 1989–90 to 1994–95* (Statistical Bulletin 1997/28).London: Department of Health, 1997. [A summary of NHS maternities in the period described, including comparisons with similar data from 1985 and earlier years.]

Department of Health. *Triennial reports on confidential enquiries into maternal deaths in the United Kingdom 1985–1993*. HMSO, London. [An invaluable compilation and review of all maternal deaths in the United Kingdom.]

Ekeroma AJ, Ansari A, Stirrat GM. Blood transfusion in obstetrics and gynaecology. *British Journal of Obstetrics and Gynaecology* 1997; **104**: 278–84. [A critical review of the use of blood products internationally in obstetrics and gynaecology with recommendations.]

Flamm BL, Goings JR, Liu Y, Wolde-Tsadik G. Elective repeat cesarean delivery versus trial of labour: a prospective multicenter study. *Obstetrics and Gynaecology* 1994; **83**: 927–32. [A non-randomized, prospective, multicentre comparison of outcomes in over 7000 women in Southern California who had previously been delivered by caesarean section.]

Murray SF, Pradenas FS. Cesarean births in Chile, 1986 to 1994. *Birth* 1997; **24**: 258–63. [This analysis is based on insurance fund data and hospital reporting systems for both public- and private-sector maternity care.]

Notzon FC *et al.* Cesarean section delivery in the 1980s: international comparison by indication *American Journal of Obstetrics and Gynecology* 1994; **170**: 495–504. [A comparison of current levels and trends in Norway, Scotland, Sweden and the United States.]

***RCOG Report of the Working Party on Prophylaxis against Thromboembolism in Gynaecology and Obstetrics*.** Royal College of Obstetricians and Gynaecologists, London, 1995. [A comprehensive review of the available evidence.]

Rosen NR, Bates LH, Herod G. Transfusion therapy: improved patient care and resource utilisation. *Transfusion* 1993; **33:** 341–7. [Reports patient health benefits and cost savings accruing from the implementation of transfusion guidelines based on national standards.]

Sachs BP, Castro MA, Frigoletto F. The risks of lowering the cesarean-delivery rate. *New England Journal of Medicine* 1999; **340**: 54–7. [This ‘sounding board' paper challenges the wisdom of the aim to reduce the caesarean section rate to a target figure. They are concerned about the safety of such an ‘authoritarian' policy.]

***US Report of Health and Human Services 1989: Indications for the use of red blood cells, platelets and fresh frozen plasma*.** Washington DC. [This sets national standards for the effective and efficient use of blood roducts.]

36.12 Third-stage complications

Michael D. G. Gillmer

Introduction
Primary postpartum haemorrhage and retained placenta
Primary postpartum haemorrhage
Retained placenta
Genital tract injuries
Perineal and vulval trauma
Cervical and vaginal trauma
Uterine rupture
Acute uterine inversion
Further reading

Introduction

A number of complications of the third stage of labour may involve a general surgeon, especially where experienced obstetric help is unavailable.

Primary postpartum haemorrhage and retained placenta

The placenta may be retained either as a result of a constriction ring in the lower uterus or, more rarely, because of morbid adherence at the site of a previous uterine scar, so-called placenta accreta. Separation may be complete or partial, and the latter is frequently associated with primary postpartum haemorrhage.

Primary postpartum haemorrhage

This is defined as a vaginal blood loss exceeding 500 ml within 24 h of delivery, but most haemorrhages occur during or immediately after the third stage of labour. Common causes include partial separation of the placenta, uterine atony, and genital tract trauma.

Excessive bleeding after vaginal delivery requires immediate action. The general condition of the patient must be assessed, and 0.5 mg of ergometrine, should be given intravenously, regardless of whether or not the placenta has been expelled. The uterine fundus should then be massaged until it contracts and, if the placenta is retained, a vaginal examination should be performed to assess its site. If placental delivery does not appear to be imminent or if bleeding persists, summon assistance, set up an intravenous infusion, and request cross-matched blood. Oxytocin, 10 units intravenously followed by an intravenous infusion of 50 units in 500 ml of normal saline, should also be given to keep the uterus firmly contracted. If the placenta has been delivered and the uterus remains relaxed despite these measures, then it should be compressed firmly between a fist introduced into the anterior fornix of the vagina and a hand placed on the lower abdomen, so-called bimanual compression. Carboprost, 250 µg, should then be given by deep intramuscular injection. This may be repeated at intervals of 15 min for a maximum of eight doses.

Retained placenta

If the placenta has not been delivered and remains within the uterine cavity, arrange for manual removal. This should be performed under general anaesthesia unless there is already an effective epidural in place. Place the patient in the lithotomy position and with the fingers held together insert a hand into the vagina. The other hand is placed over the uterine fundus and used to push the uterus down towards the pelvis. The edge of the placenta should then be identified and the vaginal hand gently introduced into the uterus along the plane of cleavage between the placenta and uterine wall. Ideally the whole placenta should be separated and removed intact, after which a further 0.25 mg of ergometrine should be administered intravenously. There is a serious risk of traumatic uterine rupture during this procedure. This usually occurs if the operator uses excessive force when insinuating his hand through the constriction ring, or fails to ‘guard' the uterine fundus by pushing it down on to the vaginal hand. An inexperienced surgeon may also mistake the lower segment for the uterine cavity and attempt to deliver the whole upper segment, instead of the placenta. Rarely the placenta is morbidly adherent and no true cleavage plane will be identified. As this situation creates a serious risk of uterine perforation and severe haemorrhage, hysterectomy is the safest course of action in a woman who has completed her family. If bleeding is not excessive and only a part of the placenta is retained, then a conservative approach may be adopted, and the placenta left *in situ.* This, however, creates a significant risk of infection and secondary postpartum haemorrhage, and should only be used when the patient is adamant that she wishes to have further children. Prophylactic antibiotics should be administered after manual removal of the placenta to minimize the risk of infection.

If the uterus is well contracted and bleeding persists after the placenta has been delivered, genital tract trauma must be excluded.

Genital tract injuries

Factors predisposing to genital tract trauma during delivery are inexpert operative delivery, previous surgery, and congenital abnormalities, such as a double cervix or vagina.

Perineal and vulval trauma

These are common during the delivery of the fetal head, but serious injury can usually be avoided by an appropriately timed episiotomy. Labial and clitoral lacerations usually heal without treatment, but should be sutured, especially if they are deep or bleed persistently. A vulval haematoma may form beneath intact skin but, more commonly, develops insidiously deep to an inadequately repaired laceration and may extend to fill the ischiorectal fossa with more than 500 ml of blood. The haematoma should be drained under general anaesthesia and, when possible, all bleeding points should be ligated. Prophylactic antibiotics are advisable and blood transfusion may be necessary.

Cervical and vaginal trauma

These are uncommon but should be suspected if bleeding persists despite vulval haemostasis and a well-contracted uterus. Predisposing factors include previous cervical surgery, misapplication of forceps or the vacuum extractor, and traumatic manipulation of a malpresentation. Examination should be performed with the woman under general anaesthesia in the lithotomy position and requires good illumination, large retractors, and skilled assistance. Large vaginal lacerations should be oversewn, taking care to obliterate dead space and avoid damage to the underlying rectum. The cervix should be grasped with large sponge-holding forceps and drawn down to facilitate inspection. The apex of all tears must be identified and wide, deep sutures inserted to achieve haemostasis. When sutures are inserted into the lateral vaginal fornices great care must be taken to avoid the ureters.

If the apex of a tear cannot be visualized, or if a broad ligament haematoma has formed, then a laparotomy is required to repair the trauma. If the maternal condition deteriorates due to oligaemic shock, a pelvic haematoma or uterine rupture must be excluded.

Uterine rupture

Uterine rupture occurs in less than 1 in 5000 labours and is usually due to oxytocic hyperstimulation, operative manipulation, or obstructed labour, especially in women with a pre-existing uterine scar. It may also occur before labour, either as a ‘silent' event with minimal lower abdominal pain and tenderness, or as a result of seat belt trauma in a motor accident. Uterine rupture in labour is most common in multigravidas, and usually involves the lower segment.

Signs of impending rupture include fetal distress, local pain, and tenderness. These are, however, common in labour and are usually missed. Vaginal bleeding may occur, but concealed intraperitoneal haemorrhage is usually associated with fetal death and rapidly leads to severe oligaemic shock. Severe placental abruption may display similar clinical features and should always be considered, especially if the bleeding is largely concealed.

Whenever possible, experienced anaesthetic and surgical staff should be called to assist as the maternal mortality in this condition is high. As soon as resuscitation has been established, a laparotomy should be performed using a vertical midline incision. The fetus should be removed, if it has not already been delivered, and the site of bleeding identified and arrested. The damage can then be assessed and a decision made whether to repair the rupture or perform a hysterectomy. Hysterectomy is the

procedure of choice, but the uterus should be conserved, whenever possible, if the patient is primigravid or desires more children. Optimal haemostasis can be achieved with total hysterectomy, but a subtotal hysterectomy, leaving the cervix, may be necessary if the patient's condition is unstable. Extension of the rupture into the upper vagina or bladder may result in ureteric damage and difficulties with haemostasis. If necessary, urological assistance should be sought, and when an open bladder injury has occurred, continuous drainage should be maintained for 10 to 14 days. Uncontrollable haemorrhage may necessitate ligation of the anterior division of one or both internal iliac arteries. It is rarely necessary to pack the pelvis.

Postoperative care includes careful blood and fluid replacement with the aid of central venous pressure measurements, an indwelling Foley catheter, and regular checks of coagulation status and blood electrolyte concentrations. Prophylactic antibiotics should also be administered.

Acute uterine inversion

Uterine inversion may occur spontaneously, but is usually the result of mismanagement of the third stage of labour. Spontaneous inversion is extremely rare, but may be caused by a sustained rise in intra-abdominal pressure, such as vomiting, when the placenta is adherent at the fundus of an atonic uterus. Fundal pressure or sustained traction on the umbilical cord with an adherent fundal placenta and atonic uterus is, however, a much more common cause of uterine inversion.

Initially a dimple develops in the fundus of the uterus and may be palpable by abdominal examination. The fundus subsequently reaches the cervix, and passes through it to present at the introitus. In extreme cases, the uterus, cervix, and vagina invert completely. If the placenta has been delivered, then the rough, plum-coloured surface of the uterus will be visible. With mild degrees of inversion the woman complains of lower abdominal pain, but in more severe forms this is followed rapidly by profound shock. Haemorrhage may be minimal, especially if the placenta remains adherent.

Immediate resuscitation and repositioning of the uterus are vital. Delay results in increasing oedema and this makes replacement much more difficult. The patient should be placed in the lithotomy position under general anaesthesia in an operating theatre, and the placenta, if still attached, should be separated. The fundus should then be grasped and gently but firmly returned through the cervix. If this proves difficult because of a cervical contraction ring, then constant, steady pressure should be maintained, if necessary with the aid of a uterine relaxant, until the uterus slides into place. Alternatively, if the equipment is available, O'Sullivan's ‘hydrostatic' method can be employed. Copious warm sterile saline is run into the vagina from a height of 1 m above the patient while the introitus is occluded with both hands. This usually corrects the inversion quite dramatically. Rarely, it may be necessary to perform a laparotomy and pull the fundus back through the cervix with large, tissue forceps.

After the inversion has been corrected, oxytocics should be given and an intravenous infusion of oxytocin administered to keep the uterus well contracted. Care must be taken to exclude other trauma, and a urinary catheter should be passed to ensure that the urine is clear. Prophylactic antibiotics should also be given routinely to reduce the risk of infection.

Further reading

Hibbard BM. Abnormalities of the third stage of labour and obstetric trauma. In: *Principles of obstetrics*, pp. 676–88. Butterworths, London, 1988. [A description of all complications of the third stage of labour, including those caused by birth trauma.]

Loeffler F. Postpartum haemorrhage and abnormalities of the third stage of labour. In: Chamberlain G, ed., *Obstetrics*, pp. 729–34. Churchill Livingstone, Edinburgh, 1995. [An overview of the causes and management of postpartum haemorrhage and other third stage complications.]

36.13 Surgery for infertility

Arlene J. Morales and John A. Rock

Introduction
Patient selection
Intraoperative principles for the reproductive surgeon
Operative instrumentation
Magnification instruments
Laparotomy instruments: microsurgery
Laparoscopy
Energy source
Lasers (*L*ight *A*mplification *S*timulated *E*mission of *R*adiation)
Electrosurgical energy
Mechanical energy
Adhesion prevention techniques
Diagnostic laparoscopy/hysteroscopy
Microlaparoscopy under conscious sedation
Congenital abnormalities of the reproductive tract
Uterine leiomyomas
Intrauterine adhesions
Fallopian tubal surgery
Endometriosis/pelvic adhesions
Ovarian surgery
Conclusions
Further reading

Introduction

Many important strides have occurred in the past 15 years in the surgical management of the infertile woman. The advent of diagnostic laparoscopy that became popular in the 1980s followed by the subsequent development of operative laparoscopic procedures spurred on the development of better instrumentation which allowed the use of microsurgery at laparotomy to become less frequent. Hysteroscopy, once performed as simply a diagnostic tool, has advanced to replace frequently the need for laparotomy for correction of many congenital uterine anomalies. With the advent of these refined laparoscopic and hysteroscopic tools, the reproductive surgeon has had to add a new dimension to their armamentarium while maintaining microsurgical techniques which are critical in any surgery of the female reproductive tract.

The next decade will continue to challenge the reproductive surgeon as the issue of outcome assessment will be expanding especially with the increasing success and efficiency of assisted reproductive technology as an alternative therapy for some women.

Patient selection

Patient selection for surgery can be a complex procedure requiring education about surgical and non-surgical options. Although a surgical course may be recommended, a competent patient's decision regarding therapy must be respected. Contemporary decision making involves the shared responsibility of the surgeon and patient. The probability of good surgical success is dependent on preoperative decisions regarding their options. The presentation of options must take into account the following variables: limited data from randomized surgical trials on outcomes, incomplete knowledge of published data, the surgeon's experiences and biases, patient's personal preference or their religious moral and ethical issues. An appreciation of the limitations of current surgical knowledge in the infertile woman is paramount for the reproductive surgeon. Other variables unique to infertility surgery such as the age of the woman, duration of infertility, as well as natural history of different pelvic pathophysiologies in the woman must all be taken into account when counseling the potential patient.

In addition to the preoperative evaluation of the pelvis, there are a host of other factors that should be considered in the infertile woman. Most importantly is the age of the woman, as decreased monthly fecundity is well-recognized. A better prognostic indicator of oocyte quality is a cycle day 3 follicle-stimulating hormone level, especially in women over the age of 35. The normal ranges of the follicle-stimulating hormone vary depending on which assay is used. Generally, the higher the follicle-stimulating hormone level is over a certain threshold, the worse the prognosis of achieving pregnancy. Elevated estradiol levels on cycle day 3 also convey a poor prognosis. Evaluation of sperm quality in the patient's partner is also important, as the patient may decide not to proceed with surgery if her partner is azoospermic. Other factors to consider are length of infertility, previous pregnancy history, and ovulatory status.

The success of proposed surgery must be delineated with the clinical outcome well defined. Published ‘success' rates appear in the literature and these usually reflect results in the hands of an experienced surgeon. The results will likely be altered by the skill and experience of the operating surgeon. Before undergoing any surgery, the role of a ‘second-look' laparoscopy in reducing ‘postoperative' adhesions may be considered as an adjunct to the initial surgery. There are no randomized clinical trials demonstrating efficacy in women, but many reproductive surgeons believe it is beneficial.

Intraoperative principles for the reproductive surgeon

Gynecologic microsurgery is a highly developed skill that demands a significant degree of neuromuscular hand–eye co-ordination. This co-ordination can only be obtained after repeated practice and close observation.

The positioning of the patient cannot be emphasized strongly enough. A poorly-positioned patient cannot only make the surgery more difficult and limit access but can cause injury (nerve compression) to the patient. At laparoscopy and hysteroscopy, the patient should be placed in a modified lithotomy position, with both arms at her side if possible. This will allow both the surgeon and assistant full access to instruments. If a surgery is expected to last more than 2 h, sequential air compression hose for deep vein thrombosis prophylaxis in the patient should be considered.

The selection of the surgical approach: laparotomy versus laparoscopy has changed considerably over the last several years. The preferred operative treatment for most infertility surgery is endoscopic except for tubal anastomosis, multiple myomectomies, and a ‘frozen' pelvis, where laparotomy is a viable option. The advantages of laparoscopic surgery include improved visualization (magnification 6× to 8×) with ability to diagnose and treat simultaneously, less tissue trauma, less *de novo* adhesion formation, quicker and easier recovery for the patient, and less complication rates. The disadvantage is the lack of tactile sense, two-dimensional view, instrument limitations, and a need for more co-ordinated movement with assistance.

Complication rates for operative laparoscopy are relatively low as they are with laparotomy. There are no direct comparative data on complications regarding these two surgical approaches. The complications of laparoscopy are injuries to bowel, vessels (hemorrhage) and anesthetic complications. Relatively unique complications of laparoscopy include respiratory embarrassment because of increased intra-abdominal carbon dioxide gas pressure and deep Trendelenburg, subcutaneous emphysema, cervical laceration, and uterine perforation. The unfortunate need for occasional unipolar instrumentation in laparoscopy also increases the risk of electrosurgical injury.

The 1993 membership survey of the American Association of Gynecologic Laparoscopists reported on over 80 000 laparoscopic procedures performed in 1993, of which 49 697 were operative laparoscopy cases. The rate/1000 laparoscopies for hemorrhage and bowel/urinary injury increased from 6.8 to 7.9 and 2.8 to 5.5, respectively. The increased rate of hemorrhage and bowel/urinary tract injuries since 1991, may be associated with the increased performance of laparoscopic advanced operative procedures.

Operative instrumentation

Magnification instruments

The laparoscope and hysteroscope can both magnify and demagnify tissue. At visualization, 1 cm from tissue, magnification is approximately six times, at 2 cm four times, and at 3 cm two times with no magnification occurring at 4 cm. At distances greater than 4 cm, demagnification occurs and structures appear smaller than reality. At laparotomy, operative loupes can provide up to 6× magnification, while operating microscopes can magnify up to 30 times reality.

Laparotomy instruments: microsurgery

Microsurgical instruments includes jeweller's forceps, dissecting rods, iris and spring-handed scissors, microneedle carriers, and a microelectrode and/or bipolar cautery apparatus. An assortment of glass, plastic, titanium, and Teflon-coated rods are useful in elevating adhesions during lysis and excision. The surgeon must be aware, however, that the glass, plastic, and Teflon rods may be damaged by a carbon dioxide laser beam. The sharp, pointed iris scissors perform precision cuts needed on the fallopian tube, and dull-pointed scissors (Struely scissors) are used during ovarian cystectomies. Serrated microscissors have the advantage of preventing tissue slippage when the blades are closed. A rounded-handle microneedle carrier is chosen for 8-O and smaller sutures. A small plastic needle carrier suffices for 4-O to 7-O sutures. Suture material for microsurgery should have minimal reactivity and good suture ‘memory'. Sutures of 7-O to 9-O are suggested for tubal surgery and 4-O to 6-O sutures for ovarian reconstruction. A 4-mm tapered needle with a cutting tip easily passes through tissue, with minimal resistance, and is therefore recommended for most microsurgical procedures.

Laparoscopy

For laparoscopies, ancillary instruments includes hydrodissection devices, grasping and biopsy forceps, scissors, probes, endosutures, and stapling devices.

The devices for hydrodissection are extremely useful for almost all operative laparoscopies. Large volumes of fluids and gas can be evacuated quickly, just as importantly, the cannula end may be used as a blunt probe for tissue dissection. There are a multitude of shape and sizes of forceps, scissors, and probes. Most of these instruments are unipolar, although there are increasingly newer bipolar instruments. In addition to sutures for laparoscopic application, there are ‘looped' sutures such as the Endoloop (Ethicon), that are frequently used when ligating large pedicles. Laparoscopic stapling devices (except for single clip applicators) have also eased the ability to do exenterative laparoscopic surgery and are rarely used in infertility surgery except for resection of a hydrosalpinx.

Energy source

The choice of energy choice is primarily dependent on the surgeon. We will briefly review the advantages and disadvantages of the most frequent sources: lasers, electrosurgical, and mechanical (Table 1).

Type	CO_2	Argon	Nd:YAG	KTP
Medium	Gas	Gas	Crystal	Crystal
Wavelength	10.6μ	$0.488/0.514\mu$	1.06μ	0.532μ
Preferentially absorbed by:	Water	Hemoglobin/melanin	Tissue proteins	Hemoglobin/melanin
Cutting ability	High	Low	Fair	Low
Hemostatic ability	Low	Fair	High	Fair
Depth of penetration	<1 mm	1–3 mm	5–7 mm	1–3 mm

Table 1 Comparisons of surgical lasers used in reproductive surgery

Lasers (*L*ight *A*mplification *S*timulated *E*mission of *R*adiation)

Lasers have been used in microsurgery since 1974. The laser has become an integral tool in reproductive surgery. Each type of laser has unique advantages and disadvantages. The four basic lasers in reproductive surgery include the carbon dioxide, argon, Neodymium:yttrium aluminum garnet (Nd:YAG) and the Nd:YAG passed through a potassium titanyl phosphate crystal. The characteristics of each laser depends on the wavelength of the specific laser. The tissue effect of a laser is primarily dependent on the power density. The direct relationship of power density to power output and its inverse square relationship to spot size remains of paramount importance in laser use. The exposure time is the most important determinant of thermal target tissue effect.

Carbon dioxide

The carbon dioxide laser was the first laser to be used extensively in gynecology and remains the most widely used laser among reproductive surgeons. The carbon dioxide laser is preferentially absorbed by tissue with a high water content, with 90 per cent of the energy absorbed and transformed into heat within the initial 30 μm of tissue—resulting in vaporization. Because epithelial cells have a high water content, these cells can be vaporized with minimal damage to surrounding structures. Although precise cutting with the carbon dioxide laser can be accomplished, its hemostatic properties are poor comparatively. Increasing the spot size (defocusing the beam) will decrease power density while increasing surface area treated and allow for some hemostasis. Although the carbon dioxide laser is the most commonly used laser in reproductive surgery, it has several remaining disadvantages. Because the carbon dioxide laser is in the infrared spectrum (invisible to the eye), an accompanying visible helium–neon laser is used in parallel. These laser beams must be well aligned and regular maintenance of the instrument is required. Additionally, as the carbon dioxide laser beam cannot be delivered through fibers, a rigid waveguide through articulating arms containing mirrors and lenses must be used. Vaporization results in plume that must be evacuated. Any ancillary instruments used must be roughened to diffuse reflection of the carbon dioxide beam. Blackened objects are not sufficient, as they will still reflect infrared radiation. Keeping the peritoneal surfaces moistened will also enhance the safety margin. The carbon dioxide laser has been employed successfully for lysis of adhesions, vaporization of endometriosis, excision of ectopic pregnancies, ablation of uterosacral ligament, oophorectomy, and performance of cuff salpingostomy.

Argon and potassium titanyl phosphate

The energy from these lasers is preferentially absorbed by pigmented molecules, such as hemoglobin, and heat is distributed much more widely in the tissue than with the carbon dioxide laser but not the Nd:YAG. These lasers, therefore, cut better than the Nd:YAG and less than carbon dioxide lasers and have better coagulation than the carbon dioxide laser and less than the Nd:YAG. Both the potassium titanyl phosphate and argon lasers, like the Nd:YAG laser, can be passed through a flexible fiber, and therefore it can be used through the laparoscope and hysteroscope. Because the argon laser beam passes through clear peritoneal tissues, it becomes an excellent tool for vaporizing pigmented endometriotic lesions. Because of its shorter wavelength, the argon laser can be transmitted through a flexible quartz fiber. However, an aiming beam is necessary for accurate direction. Use of the argon laser in reproductive surgery was first reported in five women in 1983. The argon laser is able to ablate ovarian endometriomas, lyse pelvic adhesions, create neosalpingostomies for distal tubal occlusion, and remove ectopic gestations.

Nd: YAG

The wavelength of this laser is absorbed by tissue proteins with significant scattering and markedly increased thermal effect of this laser beam deep in the tissue. If the deeper tissue heats up and explodes, a ‘popcorn' effect occurs, which should be avoided. The use of a sapphire tip not only limits the laser back-scatter but enhances the laser's excellent cutting properties through a touch technique. However, caution should be used when using the sapphire tip because of its high temperatures. Carbon dioxide emboli have been reported when this gas is used to cool the tip during hysteroscopic surgery. The Nd:YAG laser has been used to perform vaporization of endometriosis, adhesiolysis, linear salpingostomy, uterosacral ligament transsection, hysteroscopic submucous myoma resection, endometrial ablation, and ovarian wedge resection for polycystic ovarian disease.

Electrosurgical energy

Most important to understand in the use of this energy source, is the difference in unipolar versus bipolar surgery and cutting current and coagulation current. With unipolar surgery, an electrode that is used by the surgeon is the ‘active' electrode and a ‘ground' electrode is placed remotely, usually the leg of the patient to carry the electron flow. Because of the different sizes of these two electrodes, the power density current is effective at the small electrode and diluted by the larger ground electrode without tissue effect. The pathway of this current is the path of ‘least resistance', usually to the ground electrode. In bipolar surgery, the flow of current occurs between the two electrodes of the surgical instrument and results in a more localized thermal effect. The cutting current heats the tissue so rapidly, that the cells are vaporized. Fulguration (superficial coagulation/hemostasis) is achieved with intermittent short bursts of high voltage. Desiccation (deeper coagulation) occurs when the

electrical current passing through the tissue causes heating and drying.

Mechanical energy

The oldest technique is the use of scissors. The largest advantage is the availability of tissue for surgical pathology with no ‘laser' damage effect compromising the tissue. The use of disposable scissors largely has replace the frequent problem of ‘dulled' scissors.

Adhesion prevention techniques

Postoperative adhesion formation resulting in suboptimal tubo-ovarian function remains a dilemma for the reproductive surgeon. No ancillary modality can circumvent the importance of good microsurgical technique in the prevention of adhesions. Microsurgical technique encompasses the philosophy of gentle manipulation of tissue to avoid tissue trauma, continual irrigation (maintenance of moist tissue), and meticulous hemostasis with minimal tissue handling.

During laparotomy, microsurgical instruments are used to prevent injury by crushing tissue. Precise hemostasis is necessary to prevent contact of free blood with tissue. Small blood vessels to be coagulated are seen more easily under magnification provided by the operating microscope or by loupes. Tissue abrasion can be avoided by use of atraumatic instruments and by the surgeon's awareness of serosal injury when placing moist or dry abdominal packs near the bowel. Continuous intermittent irrigation with fluids, is necessary to avoid tissue drying, to prevent blood accumulation in the peritoneum, and to dilute foreign-body contaminants. Foreign-body exposure may be reduced by rinsing the surgeon's gloves to remove exposed talcum powder prior to surgery. In addition, the use of reactive sutures, such as chromic or silk, should be minimized or avoided. Tissue trauma is further limited by using small, atraumatic needles for suture placement.

It is generally believed that operative laparoscopy produces less *de novo* adhesion formation than laparotomy. Additional factors thought to be protective are the absence of the drying air of the operating room to tissues, cooling of tissue by room air, foreign-body exposure (talc), abrading of tissue by packing of bowel, and reduction in oozing from small vessels immediately after surgical injury due to increased intra-abdominal pressure by pneumoperitoneum.

Methods used to minimize postoperative adhesion formation include the use of steroids, heparin, dextran, non-steroidal anti-inflammatory agents, and placement of large volumes of fluid in the pelvis, although none of these agents are of proven efficacy. Adhesions prevention barriers such as oxidized regenerated cellulose, e.g. Interceed (Ethicon, Inc., Somerville, NJ, USA) and a surgical membrane, e.g. Gore-Tex provide alternative options.

Reproductive surgeons have advocated a second-look laparoscopy from 1 to 12 weeks postoperatively to lyse postoperative adhesions while they are still forming. Often this adhesiolysis can be performed adequately through the laparoscope, possible avoiding a major procedure at a later date. Long-term follow-up studies on the reformation of the filmy adhesions is currently in progress.

Diagnostic laparoscopy/hysteroscopy

Laparoscopy plays a major part in infertility management, but because of rapidly developing non-surgical diagnostic and therapeutic modalities such as vaginal ultrasound and *in vitro* fertilization and cost-containment concerns, there is no established standard for its use. Its use varies from inclusion in all primary infertility evaluations to occasional use after several reasonable therapeutic efforts have failed. With the re-emergence of microlaparoscopy under conscious sedation, the use of laparoscopy as vital in all infertility evaluations has been proposed.

It is critical that the anatomical survey done during diagnostic laparoscopy or hysteroscopy be performed meticulously and systematically. A general panoramic inspection is first performed and then a more detailed investigation pursues.

Hysteroscopic surgery has greatly advanced over the last decade. Advances in instrumentation have provided the reproductive surgeon with therapeutic options for the treatment of septate uterus, uterine adhesions, and submucous fibroids. Hysteroscopes are available in both a rigid and flexible sheath. In addition to carbon dioxide as a distension media for hysteroscopy, dextrose in water, regular saline, 1.5 per cent glycine, sorbitol and dextran in 10 per cent dextrose can all be used. Use of a low-flow hysteroscopic insufflator of carbon dioxide is an excellent media for diagnostic hysteroscopes but blood, debris can obscure visualization. A laparoscopic insufflator can never be used as the flow rate is excessive and carbon dioxide embolus can occur resulting in death of a patient. Dextran is a viscous non-conductive media which is immiscible with blood. Its disadvantage is the arduous use in continuous flow cases and possibility of inducement of anaphylactic shock. If electrosurgical instruments are to be used, only non-conductive media (sorbitol, 1.5 per cent glycine) can be used. Diagnostic hysteroscopies and hysterosalpingograms are complementary and of particular value in the infertile woman with an abnormal uterine cavity on hysterosalpingogram.

Microlaparoscopy under conscious sedation

The use of laparoscopy under conscious sedation is not new. It has been used for over a decade in the performance of tubal ligation. The use of microlaparoscopy has seen a resurgence in the United States over the last 2 years secondary to improved instrumentation, less patient invasiveness and an increased effort to minimize surgical costs. The microlaparoscope range from 2 to 5 mm in diameter. For the practised laparoscopist, a diagnostic microlaparoscopy takes about 10 min. The three areas that are significantly different than routine diagnostic laparoscopy are: (a) smaller field of vision on camera; (b) less of a pneumoperitoneum (less abdominal discomfort); and (c) less Trendelenberg positioning of the patients (increased patient comfort). With practice, the technique of microlaparoscopy can be mastered relatively quickly. As instruments improve, the limited visualization should improve. Despite this, the differences are significant enough, that formal training should exist prior to attempting this surgical approach.

Congenital abnormalities of the reproductive tract

Uterine anomalies occur in 0.1 per cent to 2 per cent of all women, in 4 per cent of women with infertility, and in 15 per cent of women with recurrent abortion. The müllerian (paramesonephric) ducts develop lateral to the mesonephric ducts as a result of an infolding of the coelomic epithelium late in the fifth or early in the sixth week following fertilization. These ducts grow caudally and fuse in the eighth week to form a single structure consisting of the uterus, cervix, and upper vagina. The lower aspect of the vagina is derived from the posterior wall of the urogenital sinus between the two orifices of the Wolffian (mesonephric) ducts; canalization of the entire vagina is normally completed by the twenty-second week of gestation. Anomalies of the lower third of the vagina and external genetalia generally occur as the results of inappropriate intrauterine androgen stimulation between the sixth and thirtieth week of gestation; whereas anomalies of the remainder of the reproductive tract occur as a result of a defect in lateral or vertical fusion between the eighth and seventeenth weeks of gestation (Table 2).

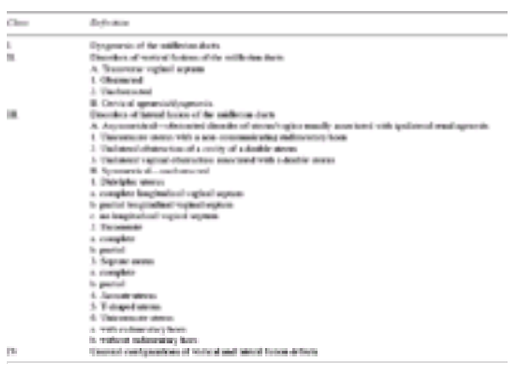


Table 2 Classification of uterovaginal anomalies

A recent four tier classification system provides a simplified and practical approach to these anomalies for the clinician.

Class I anomalies encompass dysgenesis of the müllerian ducts. Although complete absence of the entire müllerian duct system is an uncommon event, agenesis of portions of the müllerian ducts is not uncommon. Mayer–Rokitansky–Kuster–Hauser syndrome (dysgenesis of the uterus with agenesis of the upper two-thirds of the vagina) is reported to occur in 1 of 4000 to 5000 females and is the second most common cause of amenorrhea. Accompanying renal (40 per cent) and skeletal (12

per cent) anomalies are often noted.

Vaginal agenesis can either be treated with the McIndoe procedure, in which a split-thickness skin graft over a vaginal mould is sewn in place within a newly dissected vaginal space or by the procedure of choice, in which use of sequential sized vaginal (Frank or Ingram) dilators are used to create a neovagina. Up to 70 per cent success (creation of functioning neovagina) rates have been reported for the dilator technique (Rock, TeLinde). The modified McIndoe procedure is best used in those women with previously failed surgical procedure. The success of this procedure is directly related to satisfactory split-thickness skin graft (usually obtained by the Padgett electro-dermatome) and the patient motivation. Up to an 83 per cent success rate by an experienced surgeon has been reported (Rock, TeLinde).

Class II anomalies encompass disorders of vertical fusion of the müllerian ducts such as transverse vaginal septa and cervical agenesis/dysgenesis. The transverse vaginal septum can be in the upper, middle, or lower third of the vagina and can be complete or partial. Management of these septum is resection, most effectively accomplished by a vaginal approach. Occasionally, a laparotomy, with passage of a sound through the uterus/cervix to delineate better the superior aspect of the septum is useful. The septum in the upper third of the vagina can be particularly difficult to resect. A novel technique recommended includes use of ultrasound-directed needle aspiration to decompress the hematocolpos, continuous endometrial suppression and vaginal dilation prior to resection of the septum to shorten the distance between the septum. Surgical repair of cervical agenesis/dysgenesis has been uniformly disappointing.

Class III anomalies encompass symmetrical and asymmetrical disorders of lateral fusion of the müllerian ducts. Symmetrical defects are grouped into six anomalies: didelphic, bicornuate, septate, arcuate, unicornuate, and T-shaped uterus. Most lateral fusion disorders are diagnosed by hysterosalpingogram during evaluation for repeated pregnancy loss or infertility. Surgical management of the woman with a didelphic uterus is uncommon today, except for occasional excision of longitudinal vaginal septum where dyspaurenia is resultant. A unification procedure is technically difficult with disappointing results and not usually indicated. Women with bicornuate uterus are thought to experience an increased incidence of pregnancy wastage, but many of these women go on to uneventful pregnancies. If a surgical corrective technique is recommended, there are three abdominal unification procedures: the Strassmann metroplasty, the modified Jones metroplasty, and the Tompkins metroplasty. The Strassmann metroplasty is the most commonly used technique and should be considered the procedure of choice. The Strassmann metroplasty, begins with visualization of the pelvic cavity. Not infrequently, a broad peritoneal band will be identified from the posterior aspect of the bladder to the anterior wall of the rectum, passing between the two uterine bodies (Fig. 1). This band, if present, should be excised to allow better visualization. An incision through the uterine serosa from cornu to cornu extending through the myometrium should be made until both cavities are entered. The myometrium and mucosa are then closed in three layers using interrupted 2-O polyglycolic suture, starting in the lower uterine segment anteriorly and posteriorly and ending at the uterine fundus, i.e. the uterus is opened horizontally and closed vertically. The first layer should be directed so that the suture knots are tied within the uterine cavity. Finally, the serosal edges are then imbricated with a running baseball stitch, using a similar 4-O or 5-O suture.

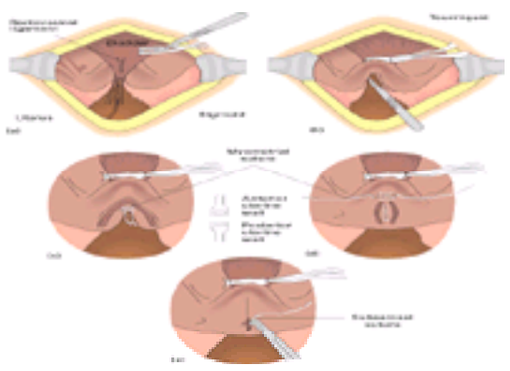


Fig. 1. The Strassmann metroplasty with modification. (a) If a rectovesical ligament is found, it should be removed. (b) An incision is made on the medial side of each hemicorpus and carried deep enough to enter the uterine cavity. The edges of the myometrium will evert to face the opposite side. (c,d) The myometrium is approximated by use of interrupted vertical figure-of-eight no. 3-O polyglycolic acid sutures. One should avoid placing sutures too close to the interstitial portion of the fallopian tubes. (e) A continuous no. 5-O polyglycolic acid subserosal suture is used as a final layer. Tourniquets are removed, and defects in the broad ligament are closed.

The septate uterus occurs more often than a didelphic or bicornuate uterus. The procedure of choice for correction of a septate uterus is transcervical metroplasty. In the past, surgical management of the septate uterus consisted of the modified Jones metroplasty or the Tompkins metroplasty. The hysteroscopic approach is more effective and results in fewer complications than the approach by laparotomy. When a hysteroscopic resection of a vaginal septum is done; it is usually necessary for a second surgeon simultaneously to visualize the outside of the uterus by laparoscopy to minimize the risk of uterine perforation. The septum may be cut using scissors, a resectoscope with electric current or laser, and then the tissue within the septum retracts, leaving a normal-appearing cavity. The use of adjuvants such as intrauterine devices, antibiotics, or hormonal therapy to prevent intrauterine scarring following septum resection is often performed, but has not been studied in randomized clinical trials. If a preoperative recurrent abortion evaluation is negative, patients properly selected for surgical correction should anticipate an improvement of between 75 and 80 per cent in conceiving and delivering a child. The arcuate uterus is a mild form of an asymptomatic septate uterus representing minimal distortion from normal uterine architecture that requires no surgical intervention.

The unicornuate uterus results in a reduced uterine volume and maybe associated with premature deliveries and fetal malpresentations. If a non-communicating rudimentary horn is present, the patient is at risk for menstrual regurgitation and pregnancy, or both leading to rupture of the rudimentary horn. In this case, the rudimentary horn should be excised. The fallopian tube from the rudimentary horn may be conserved and reimplanted into the unicornuate uterus. No surgical therapy of the unicornuate horn is needed.

The T-shaped uterus is but one of the cervical/uterine/tubal abnormalities seen in diethylstilbestrol-exposed females. These women are at increased risk of reproductive disorders. No surgical techniques have proven of benefit in altering the uterine cavitary deformities.

Class IV anomalies encompass unusual configurations of vertical and lateral fusion defects. Individualized surgical approaches can be used in women with these unusual configurations. An understanding of the embryologic origins of the tissue is of paramount importance in surgical planning.

Uterine leiomyomas

Uterine leiomyomas are the most common tumors of the female reproductive tract. Women have a 40 per cent chance of developing a detectable leiomyomas during their reproductive years. These tumors are made up of smooth muscle in an interdigitating pattern connected by fibrous connective tissue. Leiomyomas are thought to be monoclonal in origin. Sarcomatous changes occur in 0.1 per cent of enucleated tumors and are most often found in rapidly growing tumors during the adolescent and postmenopausal periods. Because uterine myomas are estrogen sensitive, they are found most commonly in women of reproductive age.

The majority of leiomyomas are asymptomatic; because of their benign nature, they require no surgical therapy. The diagnosis of leiomyomas is usually made on ultrasound, although other radiologic techniques such as pelvic magnetic resonance imaging can be used for surgical planning. Surgical therapy may be indicated because of hypermenorrhea (and ensuing anemia), dysmenorrhea, pelvic pain and extrinsic pressure on pelvic viscera, hydronephrosis, constipation, urinary frequency, recurrent fetal wastage, and rapid growth of the myomas. Any patient presenting with complaints attributable to these tumors should undergo a thorough evaluation to rule out other causes of these symptoms before treatment of the leiomyomas is undertaken.

There is an absence of evidence defining the conditions under which uterine leiomyomas can contribute to infertility. Certainly, the woman with multiple submucous leiomyomas would seem to be at risk for abnormal and poor implantation. Prior to undergoing assisted reproductive technology and after conservative attempts at pregnancy, it seems logical to recommend a hysteroscopic resection of the submucous fibroids as a therapeutic option. When recommending surgical removal of fibroids for infertility, a complete evaluation for other causes of infertility is of paramount importance. Additionally, the fibroids must be of sufficient size and/or location to warrant surgical intervention, as in the example noted above.

Although removal of leiomyomas through a laparoscopic approach have been described, there is consensus that removal of very small or pedunculated fibroids in the absence of symptoms is not a worthwhile endeavor. Resection of large leiomyomas through the laparoscope can be accomplished but can take a prolonged time period secondary to the need for morcellation of the specimen for either removal through the laparoscopic ports, a minilaparotomy or vaginal culdoplasty. Currently, the large majority of reproductive surgeons will perform most abdominal myomectomies through a laparotomy approach. The principles of a myomectomy for infertility is to use as few incisions as possible, preserve the integrity of the cornual aspects of the fallopian tubes, control hemostasis, and minimize contributors of postoperative adhesion formation. The most critical time of the surgery, is the initial inspection of the uterus, when the surgical plan is decided. Ideally, all the uterine myomas should

be removed through one anterior serosal incision to minimize the risk of postoperative adhesion formation. The uterine incision should be made with a cold knife to minimize serosal damage and carried through to the ‘pseudocapsule' of the myoma. The ‘pseudocapsule', a condensation of fibroconnective tissue, should be visible. The pseudocapsule must be incised to enucleate the myoma (Fig. 2). The myoma is grasped with a tenaculum and placed on traction. The myoma can be dissected from its surrounding normal myometrium using a periosteal elevator to assist in blunt dissection and Metzenbaum or Struely scissors to lyse adhesive bands. Usually a single artery supplying the tumor can be identified, then clamped and ligated. The location of the endometrium, especially in a uterus that is significantly distorted, must always be identified. The myometrial defect should be closed using a 3-O polyglycolic suture in a concentric, spiraling purse-string fashion from the base of the defect continuing towards the serosa. The key to closing the myometrial defect is in obliterating all potential dead space in order to decrease the risk of postoperative bleeding and infection. The serosa should be approximated with a 5-O non-reactive suture using subserosal imbricating stitch that does not create tissue necrosis.

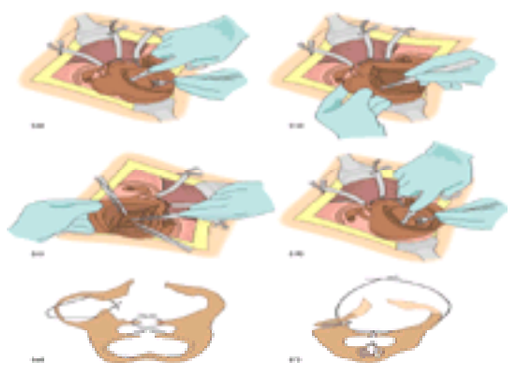


Fig. 2. The sequence of steps in a multiple myomectomy is shown in these illustrations. (a) Through a transverse Maylard incision, tourniquets are placed to occlude the uterine and ovarian artery flow. Through a single incision in the anterior myometrium, a large anterior myoma is removed first. All other myomas are removed through this incision. (b) A smaller intramural myoma is removed through the same incision. (c) To avoid making a separate incision in the posterior uterine wall, a large posterior myoma is removed through the uterine cavity. After an incision has been made through the anterior endometrium, an incision is made in the posterior endometrium directly over the posterior myoma. (d) The myoma in the posterior uterine wall is dissected from its bed and removed through the uterine cavity. An incision in the posterior uterine serosa is thus avoided. (e) Multiple sutures (no. 2-O delayed absorbable) are used to close the defect in the posterior uterine wall first. Incisions in the uterine cavity are closed. Then the defects in the anterior uterine wall are closed. (f) Trimming excess myometrium from the anterior wall allows a better approximation of the myometrium. The edges of the serosa are closed with a continuous ‘baseball' stitch with no. 4-O delayed-absorbable sutures.

Hysteroscopic resection of infertility-related fibroids can be done on an outpatient basis very successfully. The myomas can be resected using a resectoscope, potassium titanyl phosphate or Nd:YAG laser, or scissors with electrosurgery. Usually, tissue fragments float away but a continuous outflow–inflow media system will help alleviate this problem. If hemostasis is a problem, a 5-ml intrauterine balloon can be placed intraoperatively for uterine tamponade and removed 4 to 6 h after the procedure. If a large area of endometrium has been denuded, estrogen therapy (ex. Premarin 1.25 mg daily)for 6 weeks should be considered with a progestin withdrawal.

Approximately 60 per cent of women who have received a myomectomy for infertility are able to conceive, although the strength of evidence supporting this statement is weak. Women in whom the uterine cavity is entered at the time of transabdominal myomectomy should undergo a caesarean section for their subsequent deliveries to lessen the chance of uterine rupture during labor. Although uterine myomas can be excised successfully, the ability to prevent the formation of new myomas does not exist. There is a 20 per cent recurrence rate of symptomatic myoma within 5 years after myomectomy.

Intrauterine adhesions

Although the principle cause of uterine synchiae is thought to be trauma to the endometrium, infection can also result in adhesions. The consequences of Asherman's syndrome may be quite varied ranging from amenorrhea to hypomenorrhea. The surgical management of intrauterine adhesions has changed dramatically over the years, from blinded attempts at adhesion resection to hysteroscopically directed lysis of adhesions with electrosurgical current, laser, or scissors. The use of postoperative estrogen therapy has been recommended.

Fallopian tubal surgery

Today, women with a tubal disorder leading to infertility have the option of undergoing major surgery or *in-vitro* fertilization (IVF) which currently has a 25 to 35 per cent success rate per cycle. The continued improvement in pregnancy rates from IVF has left only certain tubal surgical procedures efficacious. There is no good quality evidence upon which to make the recommendation of tubal surgery versus IVF. Certainly, the skill of the surgeon, preference and age of the patient, financial impediments for either therapy are all part of the decision-making process done by the patient and physician. Patients with suspected tubal disease should therefore undergo a thorough evaluation, so that realistic estimates on the prognosis of surgical repair can be conveyed and can be compared with current success rates of IVF. Procedures commonly performed on the fallopian tube to enhance fertility include adhesiolysis, repair of distal and proximal occlusion, and tubal anastomosis. Most of these surgeries except tubal anastomosis are usually approached laparoscopically.

Of note, recent data suggests that the presence of a hydrosalpinx is detrimental to IVF implantation rates because of suspected embryotoxicity of the hydrosalpingeal fluid. Over the last year, a change in practice has started to occur in the United States. There is increasing supportive data on resection of hydrosalpinx prior to IVF.

For adhesiolysis, fimbriolysis is performed when the fimbria appear generally healthy but may have become attached or phimotic. The individual fimbria are gently dissected apart in an avascular plane. Care should be taken to remove the adhesions at both ends of attachment, rather than just break the adhesions. A common scenario in which this surgery is performed is adhesion formation from previous appendectomy where the suspicion of intratubal disease is low.

Fimbrioplasty is the surgical technique employed for more severe tubal disease. Fimbrioplasty refers to surgery performed for repair of parietal occlusion of the fimbriated end of the fallopian tube. The term, neosalpingostomy, refers to the repair of a completely occluded fallopian tube. Partial occlusion can be improved by using microsurgical techniques to perform the fimbrioplasty. When repairing distal occlusion, often a dimple can be found at the distal fimbria, serving as a landmark for the tubal ostium. The tube can be entered at this point with a small curved clamp, which is then spread; tissue planes are bluntly dissected as the clamp is gently withdrawn. To evert the endosalpinx, a small skin hook is placed approximately 1 cm into the lumen, and the tubal wall is pulled upward. The fimbria are secured with several interrupted 6-O or 7-O polyglycolic sutures placed through the tubal serosa, the everted ends of the tube near the edge, and back through the tubal serosa, thereby burying the knot between serosal surfaces (Fig. 3). Success is related to the state of the fimbria as well as to whether the ovaries are free of adhesions. If pregnancy does not ensue 6 months to 1 year following this surgery, a diagnostic laparoscopy should be performed to evaluate the anatomic status and patency of the tubes. Adhesions may be lysed at this time, but if occlusion and dilation were to recur, IVF would be the next appropriate step as repeat surgical repairs have an abysmal success rate.

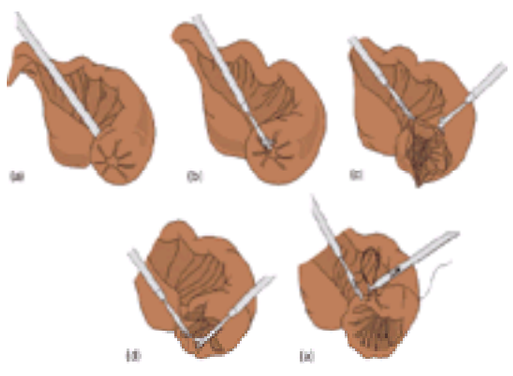


Fig. 3. Salpingostomy. (a) The occluded distal end of the tube usually has a centrally placed avascular area from which avascular scarred lines extend in a cartwheel manner. (b) The first incision is made along an avascular line toward the ovary. (c) Avascular lines are incised by viewing from within the tube along the circumference of the initial opening. (d) Cutting along the avascular lines is continued until a satisfactory stoma is fashioned. (e) The flaps can be averted by placing two or three no.

6-O absorbable synthetic sutures.

The diagnosis of proximal tubal occlusion is made only after multiple attempts to demonstrate patency by hysterosalpingography or chromopertubation during laparoscopy. Either test alone is associated with a significant false-positive occlusion rate, probably secondary to tubal spasm. Traditionally, proximal tubal occlusion has been treated microsurgically by excising the occluded proximal segment and reimplanting the distal tube in the posterior uterine wall in which the healthy intramural portion of the fallopian tube is identified. Proximal tubal occlusion can be recanalized with hysteroscopic or fluoroscopic guidance, using either a balloon catheter or a Teflon-coated guidewire. The success rates for achieving tubal patency using these less invasive techniques (approximately 80 to 90 per cent) have compared favorably with laparotomy, with the obvious advantage of significantly reducing the morbidity. IVF though is a reasonable option if fluoroscopic dilatation of proximal tubal occlusion is unsuccessful.

It is estimated that 1 to 2 per cent of women seek a reversal of a previous sterilization procedure. The prognosis of a tubal anastomosis depends on: the type of sterilization that was done (e.g. unipolar electrocautery is worst; a clip is best); the combined tubal length following the procedure; the segment of tube that is to be approximated and the age of the patient. All of these factors will aid the patient and physician to conclude as to the appropriate course of action.

Approximation of tubal segments of the same diameter offers a better outcome than that of segments requiring surgical modification in order to achieve proper diameter width. Patency rates of at least 70 per cent are expected with an isthmic to isthmic anastomosis in which the combined length is greater than 6 cm, with few pregnancies achieved when the combined length is less than 3 cm. Technically, the anastomosis procedure should always be carried out under magnification. The scarred segments of the tube are resected and the muscularis is reapproximated with 7-O or 8-O polyglycolic interrupted sutures at 12 o'clock, 6 o'clock, 9 o'clock, and 3 o'clock. It is advisable to bring the portions of the tube together over a 2-O nylon stent. If the anastomoses is being done using the ampullary portion of the tube, only a one-layer closure may be possible. If the anastomoses is being done using the isthmic portion of the tube, usually a two-layer closure is feasible with the first layer encompassing the tubal muscularis (excluding the endosalpinx) The second layer approximates the tubal serosa and outer portion of the fallopian tube muscle with a 6-O polyglycolic suture in either an interrupted or running fashion. Patency of the tube can be evaluated by injecting indigo carmine dye into the uterine fundus (Fig. 4). The site of anastomosis need not be watertight to permit subsequent intrauterine pregnancies. A 91 per cent tubal patency rate and a 83 per cent live birth rate was reported by a group of experienced surgeons in women with a isthmic-to-isthmic anastomosis.

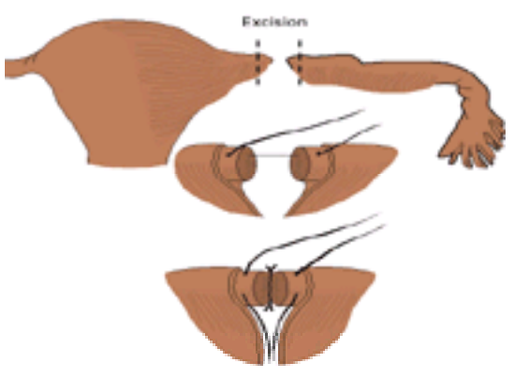


Fig. 4. Isthmic-isthmic anastomosis. Patency of obstructed lumina is established by excision of obstructed segment. The lumina may be approximated over a no. 2-O nylon splint in two layers after the mesosalpinx is approximated. Three sutures are placed in the muscularis, and the serosa is then approximated.

Endometriosis/pelvic adhesions

Although endometriosis was first described by Rokitansky in 1860, the pathophysiology of endometriosis remains an enigma. The treatment of endometriosis is based on a combination of scant scientific evidence, clinical experience, and personal opinion. Our lack of knowledge regarding the natural history of endometriosis blunts our ability to prognosticate outcomes. Current therapeutic options of endometriosis are based on symptoms, either pain or infertility. The evidence that endometriosis in its early stages is a causative factor in female infertility is controversial and subsequently its treatment is also not fully validated. Based on three prospective studies in infertile women, the prevalence of endometriosis was between 21 per cent and 38.5 per cent of women with primary infertility, 13 per cent of women with secondary infertility, and 5.2 per cent of fertile women. There are increasing supportive data that minimal and mild cases of endometriosis are related to infertility, although the data on causality is weak at best. Medical therapy is rarely recommended for therapy in the infertile woman with minimal to mild endometriosis in the absence of pelvic pain. The aim of conservative (non-curative) surgical therapy for endometriosis is the removal of disease and restoration of normal anatomy.

Whether surgery is done by laparotomy or laparoscopy, similar modalities of hemostasis and destruction of endometriotic implants can be used. No one technique has been shown to be superior to the other techniques. The two most common techniques are electrosurgery and the carbon dioxide laser. Conservative surgery may include lysis of adhesions, excision of endometrial implants and endometriomas, plication of uterosacral ligaments, anterior uterine suspension, and presacral neurectomy.

Conservative surgery is associated with a 40 to 75 per cent conception rate, depending on the severity of the disease. The only randomized clinical surgical trial in endometriosis revealed that there was an increased pregnancy rate in those women whom had undergone operative laparoscopy of minimal–mild disease compared with women whom had undergone a diagnostic laparoscopy over the subsequent 3-year period. It is estimated that eight women with mild–minimal disease would need an operative laparoscopy to result in one additional pregnancy based on the results of this trial. Certainly, these data provide the first good quality evidence regarding a surgical role in the treatment of infertile women with minimal–mild endometriosis.

Ovarian surgery

Any ovarian surgery undertaken in a reproductive age women, should use microsurgical techniques if possible. Loss of ovarian cortex may increase a woman's risk of impaired fertility.

When performing ovarian reconstruction, the reproductive surgeon should attempt to preserve two important structures. The first is the specialized peritoneal layer of epithelium covering the condensed stroma of the ovarian cortex, which contains the primordial follicles. Removal of as little of the ovarian cortex as is possible is paramount, especially in view of the availability of oocyte retrieval and IVF. Secondly the fimbria-ovarica (which extends from the tubal fimbria to the ovary) should not be disrupted because it contains skeletal muscle that aids the fimbria in the capture of oocytes.

The two most common procedures involving the ovaries are ovarian cystectomy and ovariolysis. When simple cysts, endometrioma, or benign cystic teratomas are to be removed, the visceral peritoneum is incised to the level of the cystic capsule. The cyst is dissected free from the ovarian stroma using both sharp and blunt dissection. Where the cyst is densely adherent to the ovarian cortex, as is often the case with ovarian endometrioma, the involvedcortex can be removed with the intact cyst wall. Compression of the ovarian base, either by hands or laparoscopic instruments, will stabilize the ovary and allow for hemostatic control. Once the cyst is removed, the bleeding is controlled by 3-O polyglactin sutures placed at the base of the ovarian defect. Once hemostasis is achieved the surgical ovarian defect can be left open to close by secondary intention, in the hope of reducing adhesion formation (Fig. 5).

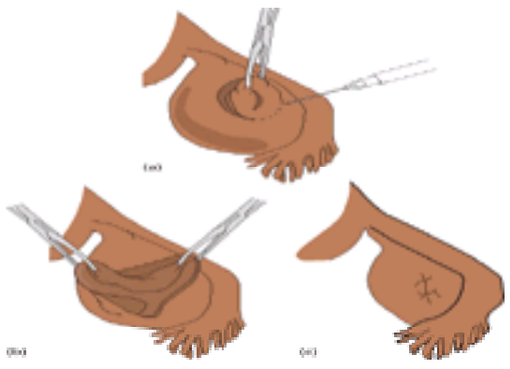


Fig. 5. Laparoscopic ovarian cystectomy after fenestration of the cyst.(a) The cut edges of the ovarian cortex and cyst wall are held and teased apart. (b) The cyst wall can be stripped off by twisting it around the grasping forceps. Hydrodissection may be helpful. (c) Large defects can be closed with laparoscopic suturing. Smaller incisions are left to heal by second intention.

Lyses of adhesions from the ovarian surface may be accomplished by sharp scissors, laser energy, or the cutting current of electrosurgery. Electrosurgery will vaporize the adhesive tissue and at the same time maintain hemostasis. Adjuvants for adhesion prevention such as placement of surgical membranes should be considered.

Conclusions

The increasing efficiency and accessibility of IVF to more women, the more difficult becomes the decision of surgical approaches to restorative anatomic position for improvement in infertility versus bypassing the anatomic defects and proceeding directly on to IVF. With perhaps decreasing opportunities for the trainee to learn reproductive surgery, it is incumbent to continue teaching these specialized surgical techniques at every opportunity. With the addition of better instrumentation, infertility surgery has become much less invasive, as the majority of surgeries can now be accomplished through the laparoscope. The acknowledgement of the need for outcome data in reproductive surgery cannot be overly emphasized. It is only through outcomes data, that the most appropriate therapeutic route for our patients will be realized.

Further reading

DeCherney AH, Russel JB, Graebe RA *et al.* Resectoscopic management of müllerian fusion defects. *Fertility and Sterility* 1986; **45**: 726. [Widely read description of technique of hysteroscopic metroplasty.]

Hulka JF, Halme J. Sterilization reversal: results of 101 attempts. *American Journal of Obstetrics and Gynceology* 1988; **159**: 767. [Results of microscopic isthmic-isthmic anastomoses.]

Hurst BS, Rock JA. Preoperative dilatiation to facilitate repair of the high tranverse vaginal septum. *Fertility and Sterility* 1992; **57**: 1351. [Novel technique to ease the resection of high transverse vaginal septums.]

Karabacak O, Tiras MB, Taner MZ *et al.* Small diameter versus conventional laparoscopy: a prospective, self-controlled study. *Human Reproduction* 1997; **12**: 2399. [Microlaparoscopy was less invasive but less sensitive in terms of diagnosis of endometriosis, adhesions.]

Levy BS, Hulka JF, Peterson HB, Phillips JM. Operative laparoscopy: American Association of Gynecologic Laparoscopists, 1993 membership survey. *Journal of the American Association of Gynecologic Laparoscopists* 1994; **1**: 301. [A comparative database of laparoscopic complications between 1993 and 1991 in the United States.]

Mahmood TA, Templeton A. The relationship between endometriosis and semen analysis: a review of 490 consecutive laparoscopies. *Human Reproduction* 1989: **4**: 782.

Mahmood TA, Templeton A. Prevalence and genesis of endometriosis. *Human Reproduction* 1991; **6**: 544.

Mais V, Ajossa S, Marongiu D *et al.* Reduction of adhesion reformation after laparoscopic endometeriosis surgery: a randomized trial with an oxidized regenerated cellulose absorbable barrier. *Obstetrics and Gynecology* 1995; **86**: 512. [A randomized surgical trial of Interceede reporting significant prevention of adhesions at second-look laparoscopy 12 to 14 weeks after initial therapy.]

Marcoux S, Maheyx R, Berube S. Laparoscopic surgery in infertile women with minimal or mild endometriosis. Canadian Collaborative Group on Endometriosis. *New England Journal of Medicine* 1997; **337**: 217. [The only randomized surgical trial of intervention in infertile women with minimal or mild endometriosis.]

McIndoe A. Treatment of congenital absence and obliterative conditions of the vagina. *British Journal of Plastic Surgery* 1950; **2**: 254. [Initial description of classic McIndoe surgical procedure.]

Menken J, Trussel J, Larsen U. Age and infertility. *Science* 1986; **233**: 1389–94. [Delineates the detrimental effect of age and decreased fertility.]

Nackley AC, Muasher SJ. The significance of hydrosalpinx in IVF. *Fertility and Sterility* 1998: **69**: 373. [Recent review of the evidence for and against resection of hydrosalpinx for improved IVF success.]

Operative Laparoscopy Study Group. Post-operative adhesion development following operative laparoscopy: evaluation at early second-look procedures. *Fertility and Sterility* 1991; **55**: 700. [A sentinel article about the advantage and use of second-look laparoscopy in the United States.]

Robboy SJ, Mehta K, Norris HJ. Malignant potential and pathology of leiomyomatoisis tumors of the uterus. *Clinical Obstetrics and Gynecology* 1990; **2**: 2–9. [Reviews the incidence, prevalence of benign and malignant leiomyomas.]

Rock JA, Keenan DL. Surgical correction of uterovaginal anomalies. In: Sciarra JJ, ed., *Gynceology and obstetrics*, Vol 1, pp. 1–30. Harper and Row, New York, 1992. [A simplified practical classification of uterovaginal abnormalities is presented.]

Strandell A, Waldenstrom U, Nilsson L *et al.* Hydrosalpinz reduces IVF/ET preganancy rates. *Human Reproduction* 1994; **9**: 861. [The first retrospective study noting the detrimental effect of hydrosalpinx on IVF.]

Verkauf BS. The incidence, symptoms and signs of endometriosis in fertile and infertile women. *Journal of the Florida Medical Association* 1987; **74**: 671.

36.14 Rupture of the pregnant uterus

J. P. Anthony and F. Mark Charnock

Aetiology of ruptured uterus
Scarred uterus
Unscarred uterus
Clinical features
Making the diagnosis
Antepartum
Intrapartum
Postpartum
Complications
Management
Prevention
Management of established rupture
The principles of management
Surgical procedure
Closure of the defect
Hysterectomy
Subtotal hysterectomy
Perioperative care
Postoperative care
Histological examination
Prognosis
Future pregnancies
Summary
Further reading

Uterine rupture is an important cause of both maternal and fetal death worldwide, although advances in obstetric management have reduced adverse outcomes to a minimum in developed countries. The incidence varies widely, but most authors suggest figures of 1:1200–1500 deliveries. The maternal mortality can be high, with haemorrhage and its complications being the most important contributory factors. Nevertheless, only six maternal deaths occurred from uterine rupture in the United Kingdom between 1985 and 1987 in relation to 2.3 million deliveries. In no mother was death due to rupture of a caesarean section scar, although one patient had had a previous caesarean section.

Aetiology of ruptured uterus

Often a distinction is made between rupture of a scarred uterus and ‘spontaneous' rupture where there is no previous scar. However, the term ‘spontaneous' is misleading in that such events are usually associated with obstetric trauma during labour (Table 1).

Uterine scar
Caesarean section
Other:
Myomectomy
Hysterectomy
Termination
Curettage
Evacuation
Obstructed labour
Oxytocin infusion, especially with obstructed labour
Trauma
Obstetric:
Difficult instrumental delivery
Breech extraction
Internal podalic version
Manual removal of morbidly adherent placenta
Non-obstetric:
Road traffic accident
Violence
Gynaecological
Cornual pregnancy
Unknown

Table 1 Causes of uterine rupture

Scarred uterus

Previous caesarean section is the most common factor, and while rupture is reported in only 0.3 to 0.7 per cent of labours in the presence of a previous lower-segment scar, it occurs in up to 9 per cent of labours following a classical uterine scar.

The number of previous lower-segment scars appears not to increase the risk of subsequent rupture. A 3 per cent incidence of occult or subclinical ruptures or dehiscences has been reported when an elective repeat caesarean section was performed. In a similar group, 0.5 per cent ruptures and 1.9 per cent dehiscences were reported. Provided the uterine scar does not extend into the upper segment, it is insignificant with regard to rupture whether the lower-segment incision is transverse or vertical.

A lower-segment scar typically ruptures only during labour, particularly in the late first or second stage when uterine activity reaches its peak. Upper-segment scars, in contrast, may rupture before labour, when the presentation mimics closely that of abruptio placentae.

Other uterine scars (see Table 1) have been associated with rupture in pregnancy or in labour. Patients with such scars require skilled obstetric management.

Unscarred uterus

Less commonly, ruptures may be associated with obstetric trauma from instrumental deliveries or intrauterine manipulations; with abdominal trauma; and with uterine anomalies, such as a pregnancy within a uterine cornu. The last may mimic ectopic pregnancy in presentation.

A combination of factors could be present, with rupture occurring in a scarred uterus, which may or may not involve the previous scar.

Clinical features

The presentation may vary from dramatic, with collapse of the mother, to vague symptoms and signs with the diagnosis being delayed by many hours. Furthermore, routine exploration of previous scar sites, either digitally via the cervix after vaginal delivery or directly at caesarean section, may reveal hitherto unsuspected weaknesses or ruptures in the scar. Such occult defects may be divided into ‘rupture' requiring surgical repair or intervention, or ‘dehiscence' not requiring intervention. While the classical features are listed in Table 2, some or all may be absent.

Maternal:
Vaginal bleeding in labour
Abdominal pain
Cessation of progress in labour
Shock
Signs of acute abdomen
Postpartum haemorrhage
Fetal:
Fetal distress
Fetal death
Dehiscence:
No symptoms
Vaginal examination—defect palpable

Table 2 Clinical features of uterine rupture

Signs such as scar tenderness are unreliable, and suprapubic tenderness is common in many labours. Fears that epidural analgesia would mask symptoms of scar rupture are probably unfounded as features other than tenderness often predominate.

Making the diagnosis

Antepartum

Antenatal presentation of uterine rupture can often be confused with that of abruptio placentae. The presentation is not necessarily dramatic initially, and the diagnosis should be considered when mothers with a scarred uterus present with vaginal bleeding or abdominal pain. Maternal shock and fetal compromise or death are common, with fetal survival being rare.

Intrapartum

To detect uterine rupture as early as possible, meticulous observation is essential, particularly in mothers with previous uterine surgery. Regular assessment of the strength of uterine contractions and observations of maternal pulse and blood pressure, fetal heart rate pattern, and the progress of labour are all facilitated by the routine use of a partogram. Vaginal bleeding should be noted promptly. Continuous electronic fetal monitoring is advisable where facilities exist to detect early signs of fetal compromise. Intrauterine pressure monitoring is of value in confirming delayed progress in the late first stage despite good uterine activity. Unfortunately, no consistent pattern of uterine pressure presages uterine rupture.

Postpartum

In many cases, the diagnosis is delayed until several hours after delivery. These cases are often associated with a more serious outcome in association with severe haemorrhage. Continued close observation in the postdelivery period is essential, therefore, for mothers at risk of rupture.

Complications

The major maternal complications of uterine rupture are outlined in Table 3.

Haemorrhage
Amniotic fluid embolism
Disseminated intravascular coagulation
Associated with obstructed labour:
Sepsis
Dehydration
Acidosis
Associated with operative intervention

Table 3 Complications of uterine rupture

Management

Prevention

In some areas, symphysiotomy has reduced the risk of uterine rupture by avoiding a caesarean scar. In those mothers with a previous caesarean section, it is probably wise not to allow a subsequent delivery should an upper-segment incision have been made, a more common event nowadays with caesarean sections being performed for very preterm deliveries. In multipara, oxytocin should be used only after careful clinical consideration. Early recourse to caesarean section rather than attempting complicated operative vaginal deliveries has become ever more established in obstetric practice in developed countries.

Phelan *et al.* allowed a trial of labour in patients with more than one previous caesarean section and achieved successful vaginal delivery in 82 per cent of mothers with one previous caesarean section (1348/1637 cases), 72 per cent after two caesarean sections (107/149 cases), and 90 per cent after three (9/10 cases). Those least likely to succeed in a vaginal delivery were those in whom the previous caesa-rean section had been for failure to progress in labour due to cephalopelvic disproportion.

Management of established rupture

The factors influencing management are indicated in Table 4; it will be influenced by whether the rupture occurs antenatally, during labour, or after delivery. The severity of injury depends on whether the rupture involves the placental site, cervix, bladder, or ureters. The risk of bladder involvement is strongly related to the number of previous caesarean sections as the tissues are so often densely adherent. Ruptures not involving a previous uterine scar are usually more dramatic, requiring urgent measures to avoid serious sequelae from hypotension and haemorrhage.

Time of rupture
Site of rupture
Severity of rupture
Damage to other structures
Pre-existing maternal condition

Table 4 Factors influencing management of uterine rupture

The principles of management

Resuscitation and stabilization of the patient's general condition are vital (Table 5), with blood transfusion being invariably required. Undue delay before surgery can worsen the condition of an already compromised patient, particularly if disseminated intravascular coagulation occurs.

Resuscitation:
Blood
Fluid
Acid–base
General anaesthesia
Control of haemorrhage
Surgical procedures
Drainage
Antibiotics
Catheter
Postoperative surveillance
Histology

Table 5 Management of uterine rupture; principles

If the rupture follows prolonged obstructed labour, the mother may be dehydrated, ketotic, and acidotic. Marked biochemical disturbances must be tackled in these patients. The health of the mother must always take priority over the well-being of the fetus, although, ideally, one aims to deliver a live fetus from a well mother.

Skilled general anaesthesia is mandatory and will permit greater control of the patient's overall haemodynamic condition. Experienced help must be sought, as the operative procedures during laparotomy can be technically very difficult.

Control of haemorrhage

Control of haemorrhage is a priority and manual compression of the aorta or ligation of one or both internal iliac arteries can be useful. Because of the enormous collateral circulation, both iliac arteries may need ligation. Care should be taken to identify the ureters properly. In the face of brisk haemorrhage, closure or clamping of the edges of the defect temporarily, before any definitive procedure, will reduce the blood loss.

Surgical procedure

Closure of the defect

Closure of the defect with sterilization may be the procedure of choice if the patient is severely shocked. The patient's wishes must be considered when choosing an appropriate procedure, as several series have reported pregnancies following conservative surgery (repair of the defect only). The patient must be counselled subsequently on the risk of repeat rupture, and she may then elect for sterilization.

Hysterectomy

Hysterectomy is often the procedure of choice in view of the risk involved with subsequent pregnancies. At hysterectomy, the tissues must be handled gently as they are very vascular and oedematous and the pelvic ligaments are lax. The greatly enlarged vessels should be carefully clamped, and it is essential to double-tie all pedicles. The risk of tearing the bladder wall during dissection should not be underestimated and can be reduced by sharp dissection, keeping to the midline. This will also reduce the risk of trauma to the venous plexuses, which lie laterally behind the bladder base.

Subtotal hysterectomy

Subtotal hysterectomy is appropriate when safe identification of the ureters is impossible. As identification of the lower edge of the dilated cervix can be difficult, part of the cervix may be left behind. The subtotal operation may be the procedure of choice if the surgeon is inexperienced.

Perioperative care

Both the pelvis and the rectus sheath should be drained, especially in the presence of infection or where extensive dissection of adherent tissues was necessary. Wide-bore suction drains are helpful with the pelvis. Large subrectus haematomas develop easily if disseminated intravascular coagulation supervenes. An indwelling urinary catheter is mandatory, to assist both in monitoring fluid balance and in detecting haematuria, which may signify unrecognized bladder injury. Before the end of the abdominal procedure, examining the patient in the lithotomy position can assist in diagnosing whether the entire cervix has been removed, as well as excluding associated vaginal injuries or lacerations. Antibiotic prophylaxis with broad-spectrum agents is advised. The risk of sepsis is particularly high when membranes have been ruptured for a long time.

Postoperative care

This should include observation for signs of infection, careful monitoring of fluid balance, renal function and coagulation status, and adequate blood replacement. Central venous pressure monitoring can be invaluable.

Histological examination

The hysterectomy specimen will permit identification of abnormal trophoblast invasion at the site of the rupture and the full extent of the trauma. It also serves as a check that all the cervix has been removed.

Prognosis

Mortality and morbidity result largely from haemorrhage, impaired renal perfusion due to hypovolaemia, and infection. In one large series over half of the ruptures occurred in the presence of previous uterine scarring, the majority of which had mild clinical features with little bleeding from the avascular scar site, and were repaired conservatively. When more extensive lacerations have developed, or in cases where rupture was traumatic or spontaneous, 85 per cent of such patients required immediate hysterectomy. The overall maternal mortality was 3 per cent, with all deaths occurring in patients who ruptured a previously intact uterus. Fetal mortality was 27.8 per cent, with death being three times more common in rupture of an unscarred uterus. Furthermore, infant morbidity was relatively high amongst those surviving a traumatic uterine rupture; central or peripheral nerve injuries or respiratory distress symptoms were the most common sequelae.

Future pregnancies

In one study a good outcome was found, with only a single repeat rupture occurring in 22 subsequent pregnancies following conservative surgery. Some patients even achieved vaginal deliveries. Only two out of 28 patients had involuntary infertility after the procedure. As fear of subsequent pregnancies can be an important factor,

conservative surgery should only be performed when the couple actively desire further children. Subsequent pregnancy after conservative surgery should usually be managed by elective caesarean section in the later third trimester, with elective sterilization when the couple's family is complete.

Summary

Uterine rupture remains a serious and life-threatening event. The increasing use of caesarean section for delivery in the Western world will lead to an increased number of mothers at risk of scar rupture in subsequent pregnancies. Maternal and fetal outcome is worse following rupture of an unscarred uterus, often in association with the use of oxytocics. If the morbidity and mortality are to be kept to a minimum, the management of uterine rupture should involve careful collaboration between obstetricians, anaesthetists, and the intensive care-unit team.

Further reading

Aguero D, Kizer S. Obstetric prognosis of the repair of uterine rupture. *Surgery, Gynecology, Obstetrics* 1968; **124**: 528–30.

Arulkumaran S, Gibb DMF, Ingemarsson I, Kitchener HC, Ratnam SS. Uterine activity during spontaneous labour after previous lower-segment caesarean section. *British Journal of Obstetrics and Gynaecology* 1989; **96**: 933–8.

Arulkumaran S, Ingemarsson I, Ratnam SS. Oxytocin augmentation in dysfunctional labour after previous caesarean section. *British Journal of Obstetrics and Gynaecology* 1989; **96**: 939–41.

Beckley S, Gee H, Newton JR. Scar rupture in labour after previous lower uterine segment caesarean section; the role of uterine activity measurement. *British Journal of Obstetrics and Gynaecology* 1991; **98**: 265–9.

Department of Health. *Report of confidential enquiries into maternal death in the United Kingdom 1985–87*, p. 68. HMSO, London, 1991.

Dewhurst CJ. The ruptured caesarean section scar. *Journal of Obstetrics and Gynaecology of the British Commonwealth* 1957; **64**: 113–18.

Garnet JD. Uterine rupture during pregnancy: an analysis of 133 patients. *Obstetrics and Gynaecology* 1964; **23**: 898–905.

Merrill BS, Gibbs CE. Planned vaginal delivery following caesarean section. *Obstetrics and Gynaecology* 1978; **52**: 50–2.

O'Connor RA, Gaughan B. Pregnancy following simple repair of the ruptured gravid uterus. *British Journal of Obstetrics and Gynaecology* 1989; **96**: 942–4.

Phelan JP, Clark SL, Diaz F, Paul RH. Vaginal birth after caesarean. *American Journal of Obstetrics and Gynecology* 1987; **157**: 1510–15.

Reyes-Ceja L, Cabrera R, Insfran E, Herrera-Cassa F. Pregnancy following previous uterine rupture. *Obstetrics and Gynaecology* 1969; **34**: 387–9.

Sheth SS. Suturing the tear as treatment in uterine rupture. *American Journal of Obstetrics and Gynecology* 1969; **105**: 440–3.

Shrinsky DC, Benson RC. Rupture of the pregnant uterus: a review. *Obstetrics and Gynaecology Survey* 1978; **33**: 217–32.

37 Soft tissue tumours

William C. Wood

[Introduction](#)
[Incidence](#)
[Epidemiology](#)
- [Molecular epidemiology](#)
- [Chemical carcinogenesis](#)
- [Trauma or foreign bodies](#)
- [Chronic lymphedema](#)
- [Radiation](#)

[Benign tumors of soft tissue](#)
- [Non-aggressive benign tumors](#)
- [Aggressive benign tumors](#)

[Soft tissue sarcomas](#)
- [Clinical appearance](#)
- [Imaging](#)
- [Diagnosis](#)
- [Staging](#)
- [Natural history](#)
- [Treatment](#)
- [Retroperitoneal sarcoma](#)
- [Isolated recurrences and metastases](#)

[Further reading](#)

Introduction

The term ‘soft tissue' refers to the extraskeletal connective tissue that accounts for more than 50 per cent of body weight. It includes muscle, tendon, fat, fascia, and synovium. These all arise embryologically from the mesenchyme. Although soft tissue tumors are all thought to arise from primitive mesodermal tissue, they vary widely in their histologic appearance and in the mature tissue type that they resemble. By convention, tumors arising in the peripheral nerves are also grouped with soft tissue tumors, as their presentation and treatment are similar.

Benign soft tissue tumors such as lipoma or benign fatty tumor closely resemble specific normal adult tissues. They exhibit little tendency to local invasion, rarely recur after local excision, and do not metastasize. Malignant soft tissue tumors are termed sarcomas (from the Greek ‘sarkoma', meaning fleshy growth), while malignant tumors arising from epithelial tissues are termed carcinomas. Exceptions to this generalization are tumors arising from the endothelium of blood or lymphatic vessels, or the mesothelial lining of body cavities, which are grouped with soft tissue tumors. A classification has been developed by Enzinger using the principle of similarity of histologic appearance to normal tissues (Table 1). Unlike most other human neoplasms, the soft tissue tumors are not grouped simply into benign or malignant. Occupying a middle ground between sarcomas and benign neoplasms is a group of locally aggressive tumors such as desmoid tumors, aggressive fibromatoses, atypical lipomas, and aggressive neurofibromas. These do not metastasize and are not called sarcomas, but require more comprehensive treatment than do other benign soft tissue tumors because their characteristics locally resemble malignant behavior.

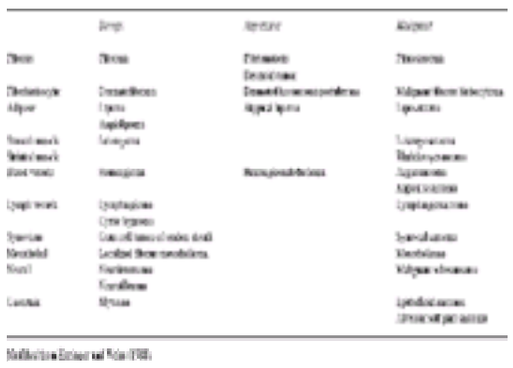


Table 1 Histologic classification of soft tissue tumors

Incidence

Benign tumors of soft tissue are far more common than sarcomas, outnumbering them by about 100:1 in biopsy series. As many lipomas and hemangiomas are never submitted for biopsy examination, the actual ratio is probably considerably higher. Lipomas are the most common human neoplasm. Sarcomas are rare when the volume of tissue from which they may arise is considered, with an annual age-adjusted incidence of about 2:100 000. Just over 5000 soft tissue sarcomas are diagnosed annually in the United States, for example, accounting for 0.7 per cent of newly diagnosed invasive cancers (excluding squamous and basal cell carcinoma of the skin). In children under 15 years of age they account for over 6 per cent of all cancers.

Epidemiology

Molecular epidemiology

The pathogenesis of most soft tissue tumors is poorly understood. Genetic factors are associated with certain benign tumors: about 5 per cent of lipomas and angiolipomas are multiple and familial. Oncogenes were first described in an avian sarcoma model by Peyton Rous in 1911. Soft tissue sarcomas have no established genetic predisposition, but occur with increased frequency in patients with genetically transmitted conditions such as intestinal polyposis, Gardner's syndrome, and basal cell nevus syndrome. Patients with von Recklinghausen's disease have a 15 per cent chance of developing a soft tissue sarcoma. This is associated with NF1 gene expression. As the germ cell mutations associated with these conditions are defined, the association with sarcomas of soft tissue is being studied and appears to reflect the loss of tumor suppressor activity of certain genes. The mutations of RB genes responsible for retinoblastoma are also associated with an increased risk of later sarcoma. Mutations of the RB genes are seen in sporadic sarcomas as well. *p*53, another tumor suppressor gene, is found to exhibit germline mutations in the Li-Fraumeni families who are at increased risk for sarcoma. This gene is also found to be mutated in sporadic sarcomas. This is a very active field of investigation. It is hoped that certain genetic markers will be predictive of response to certain therapies, for instance sensitivity to radiation or to certain cytotoxic agents, and allow therapy to be specifically tailored to a given tumor. Some experimental models suggest the feasibility of this approach. *p*53 nuclear overexpression and loss of pRB expression are both associated with increased tumor grade, but such work remains at an investigational level without clinical integration. The investigation into the molecular mechanism of the development of Kaposi's sarcoma in association with AIDS has been a model for future research in this field. Kaposi's sarcoma has been found to be linked to the presence of human herpes virus 8 (**HHV-8**). It is found in all types of Kaposi's sarcoma, not only those that are AIDS associated. The role of HHV-8 has proved quite complex with the initial tumor arising as an angiohyperplastic–inflammatory lesion mediated by inflammatory cytokines and angiogenic factors triggered by the infection. Over time this reactive lesion evolves into a sarcoma, demonstrated by monoclonal growth in the later nodular lesions.

Chemical carcinogenesis

Chemical carcinogens have been implicated in anecdotal reports but no link has been established. Exposure to dioxin, phenoxyacetic acid, or to agent orange (containing dioxin) has not been shown convincingly to have an association with development of sarcomas in case–control studies, yet conflicting results have been published.

Trauma or foreign bodies

Although trauma is often incidental to the discovery of a soft tissue tumor by a patient, no evidence exists of any link between the two. Anecdotal reports of sarcoma occurring in association with foreign bodies such as bullets and shrapnel pieces exist, but prospective studies of large numbers of patients with metallic or Silastic implants have not shown any associated tumor development.

Chronic lymphedema

There is an increased incidence of lymphangiosarcoma in lymphedematous extremities, first described in women's arms following radical mastectomy (Steward–Treves' syndrome). It has also been seen in lymphedema resulting from filariasis confirming the association with the lymphedema rather than breast cancer or chest wall irradiation.

Radiation

Radiation exposure is associated with a later incidence of sarcoma estimated at between 2 and 4/1000 patients after 10 years. Although any type of soft tissue sarcoma can arise, malignant fibrous histiocytoma, angiosarcoma, lymphangiosarcoma, or extraskeletal osteosarcoma are most common. Most soft tissue tumors arise in individuals in whom no predisposing factor can be identified.

Benign tumors of soft tissue

Non-aggressive benign tumors

Lipomas are among the most common benign tumors. They are usually subcutaneous, not tethered or fixed to the underlying fascia, multilobulated, and vary greatly in size. Angiolipomas contain many small vessels and are usually tender or painful. Their major significance is their confusion with liposarcomas. A biopsy should be taken or they should be removed when evidently growing, when larger than 5 cm, or if symptomatic. Biopsies should be taken of large lipomas, or those tethered to fascia or deep to the fascia, as recommended for liposarcomas.

Neurofibromas and neurilemmomas both arise from Schwann cells. Simple excision is curative, but when they arise in nerves of clinical significance, every effort should be made to preserve the affected nerve. Multiple neurofibromas are seen in patients with von Recklinghausen's disease and are too numerous to be removed: 10 to 15 per cent of these patients will develop neurogenic sarcomas. Neurofibromas should be removed when they undergo rapid enlargement or if they become painful, as well as for cosmetic reasons.

Dermatofibromas, also called sclerosing hemangiomas, occur in the skin and have a very characteristic appearance. They are usually about 1 cm in diameter and do not tend to grow. Excision is curative when desired. Myxomas and mesenchymomas present as soft tissue tumors; the diagnosis is made histologically. They are almost always cured by simple excision. Mesotheliomas arise in the pleura, pericardium, or peritoneum. They project but do not infiltrate and are usually cured by excision.

Hemangiomas are vascular neoplasms that are usually present at birth in the head and neck area, but they can be anywhere. Most congenital hemangiomas regress spontaneously by the age of 5 years, even if they grow rapidly during the first months of life. It is extremely rare for such lesions to compromise vital structures, but low-dose radiation to a very focused field may be used to treat this situation. Hemangiomas of extremities may not regress, and can result in limb enlargement. Such patients are often best treated by angiographic embolization, or sometimes by surgical dearterialization. The hemangioma may involve bone, and surgery too near the tumor may be accompanied by severe or even fatal hemorrhage. Lymphangiomas are similar tumors which can be surgically excised. Conservative management of both of these tumor types allows preservation of function and achieves a favorable cosmetic result. Ongoing definition of mediators of growth of these lesions is leading to anecdotal reports of dramatic response to therapy with targeted biologic modulators and specific cytokines. This is a rapidly evolving clinical aspect of research in angiogenesis.

Ganglions are fibrous walled cysts, most commonly seen on the dorsal surface of the wrist. They result from an excrescence of synovial tissue and are filled with synovial fluid. Simple excision is usually curative. They are easily confused with giant cell tumors of the tendon sheath, which are also most commonly seen in the wrist or hand and are also treated by simple excision.

Leiomyomas most commonly arise from the smooth muscle of the uterus and gastrointestinal tract. Leiomyoblastomas are similar to leiomyomas, but tend to grow large and to develop areas of hemorrhage and cystic degeneration. Both are treated by simple excision.

Nodular fasciitis has often been mistaken for a fibrosarcoma. These tumors usually occur in the superficial or deep fascia and grow rapidly when they first appear. In weeks to months they reach 3 to 5 cm and then usually stop growing. They are generally eliminated by simple excision.

Aggressive benign tumors

Desmoid tumors or aggressive fibromatosis can arise from any fascial tissue. These lesions are remarkable for their infiltrative nature and fall into two groups. Abdominal desmoid tumors arise from the muscular wall of the abdomen, often in association with abdominal incisions, and are especially common in postpartum women, in whom they may be related to hormonal changes and surgical trauma. If resected to clear margins they seldom recur. Extra-abdominal desmoid tumors behave more aggressively. They are common about the shoulder or within an extremity, where they infiltrate diffusely. Although these tumors do not metastasize, they cause local destruction of function. Surgical resection short of amputation is often impossible, but wide resection is the preferred treatment when it can be accomplished with minimal morbidity. Resection to close margins or to positive margins is generally followed by recurrence. Irradiation is effective in controlling desmoid tumors if 60 Gy can be delivered. The hormonal milieu may affect desmoid tumors, and treatment with progesterone, tamoxifen, indomethacin, and ascorbic acid have all been reported to induce remissions. These therapies are associated with minimal toxicity and deserve further study, but they usually fail to produce a response.

Atypical lipomas and atypical fibrous histiocytomas are generally located deep in the soft tissue or muscle; they appear atypical histologically and although they do not metastasize they tend to recur locally after resection, particularly if wide margins are not achieved.

Dermatofibrosarcoma protuberans arises insidiously in the skin or subcutaneous tissue. These tumors do not metastasize, even when multiply recurrent, but they behave locally as though malignant. It is important to achieve control at the time of first resection, with generous (at least 2 cm) margins to minimize the risk of further recurrence. Half of all patients treated with simple excision and more limited margins will suffer recurrence; amputation may be required and death may result. Radiation therapy as a supplement to conservative resection appears beneficial but has not been subject to controlled trial. Radiation has been used in place of very radical resection in some patients with a favorable result.

Soft tissue sarcomas

Clinical appearance

Soft tissue sarcomas can arise in any part of the body and usually present as a painless lump. When they arise in the retroperitoneum or deep in the thigh or buttock they may be quite large before they are first noticed; on the foot or wrist very small masses may be apparent. The larger the soft tissue mass, the more likely it is to be a sarcoma, and this diagnosis should be considered for any mass that is at or below the level of the deep fascia. Progressive growth increases the probability that the tumor will prove malignant. Careful physical examination will give the dimensions of the palpable tumor and provide information about the infiltrative nature of its borders, fixation to normal structures, and involvement of skin, muscle, bone, major blood vessels, or nerves. The insidious feature of soft tissue sarcomas is their ‘benign' presentation that too frequently induces a false sense of security in both patient and physician.

Imaging

MRI provides superb visualization of the extent of soft tissue tumors. In comparative studies it is generally superior to CT scanning, but not to such a degree that it must be used if CT is more available. The detail afforded by these techniques has eliminated the need for angiography in treatment planning.

Diagnosis

A biopsy specimen of any putative sarcoma should be obtained: the technique used is of great importance. Fine-needle aspiration cytology will often yield a diagnosis of malignancy, but the classification and grading of a sarcoma cannot reliably be performed from cytology alone. A core-needle biopsy allows examination of a thin core of tissue and permits recurrent sarcoma to be recognized. Sample size is limited, however, and may hinder the accuracy of diagnosis and grading of soft tissue sarcomas. Good communication with an expert soft tissue pathologist is essential in this stage of the evaluation. Incisional rather than excisional biopsy is preferred for all but the smallest tumors. The excision of any tumor greater than 3 cm in diameter contaminates the environs with nests of tumor cells that may be carried for some distance along tissue planes by dissecting hematomas, and creates a need for more radical therapy than would otherwise have been needed.

The correct placement of a biopsy incision is also of far greater importance in the management of soft tissue sarcomas than is the case for most other tumors. The biopsy should be performed by a surgeon who is prepared to treat the tumor if it proves to be a sarcoma, and who can place the biopsy incision where it does not compromise the planned treatment. Biopsy incisions should be oriented longitudinally in the extremities, and over the direction of underlying muscle, placed for inclusion in possible later resection of skin. General anesthesia is preferable to local anesthesia since these tumors are usually at the level of or beneath the deep fascia. Attempts to obtain biopsy samples at this depth under local anesthesia are often limited by the patient's discomfort, and therefore produce inadequate samples of tissue. Cautery should not be applied until the specimen has been removed, to avoid heat artifacts. Meticulous hemostasis and closure in multiple layers to avoid hematoma formation is beneficial. Inappropriate biopsy techniques contributed to diminished quality of care in a high proportion of patients with sarcoma studied by the Musculoskeletal Tumor Society. Classification and grading of soft tissue tumors should be reviewed by subspecialty pathologists when treatment decisions will relate to the histologic type or grade: proper management of soft tissue sarcomas depends upon accuracy in classification and grading, which in turn depend upon representative well-stained slides interpreted by one with considerable interest and experience in sarcomas.

Staging

The first aspect of TNM staging is the evaluation of the tumor itself (T). In addition to physical examination, radiologic studies are essential. CT and MRI have revolutionized the local staging of sarcomas. The size of sarcomas, measured as the longest axis diameter, carries independent prognostic information for risk of distant metastasie (Table 2).

	Grade 1		*Grades 2 and 3*	
Size (mm)	*No.*	*%*	*No.*	*%*
< 25	5	0	17	6
26–50	11	0	48	23
51–100	12	0	55	38
100–200	8	0	33	52
> 200	2	0	6	83

Table 2 Distant metastases by size of primary tumor

The evaluation of regional lymph nodes (N) is confined to physical examination. Much is made of the rarity of regional nodal metastases in patients with soft tissue sarcoma, averaging 3.9 per cent in 12 series comprising more than 2500 patients. A review of 323 patients with clinical M0 sarcomas treated at the Massachusetts General Hospital identified 19 (5.9 per cent) with regional lymph node involvement. The incidence varied with stage, being 0, 2, and 12 per cent in grades I, II, and III, respectively. Four of five patients with epithelioid sarcomas had regional node disease: any palpable regional lymph nodes in a patient with grade III sarcoma should be viewed with suspicion.

The search for metastases (M) in soft tissue sarcomas is focused on the lungs. Although sarcomas of the retroperitoneum commonly metastasize to the liver, the lung is easily the most common site of distant metastases from other sites. CT scans of the chest are consequently an essential part of initial staging of patients with sarcoma.

Unlike most other solid tumors, where classification by stage is based upon TNM alone, grade (G) is the dominant factor in assigning the stage of soft tissue sarcomas and is most closely related to prognosis (Fig. 1). Grade I sarcomas (well-differentiated) are less likely to recur locally or to exhibit distant metastases than are grade III sarcomas (poorly differentiated). The grade of the tumor is based on the presence of necrosis, the number of mitotic figures per 10 high-power fields, cellularity, and the extent of stroma. The grade of the tumor is more predictive of clinical course than the histologic classification and also subject to less interobserver variation. The pathologic grade is combined with the TNM classification to assign a stage to soft tissue sarcomas (Table 3).

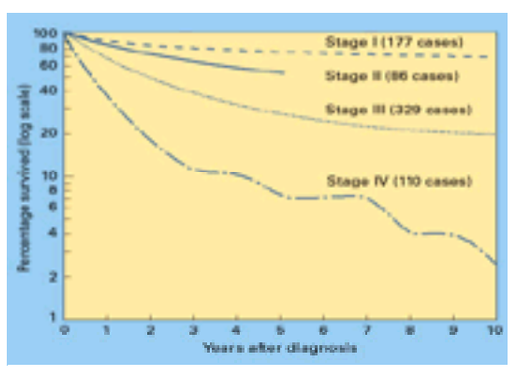


Fig. 1. Survival as a function of grade of soft tissue sarcomas. These survival curves reflect the staging system based on tumor grade as the primary determinant of stage.

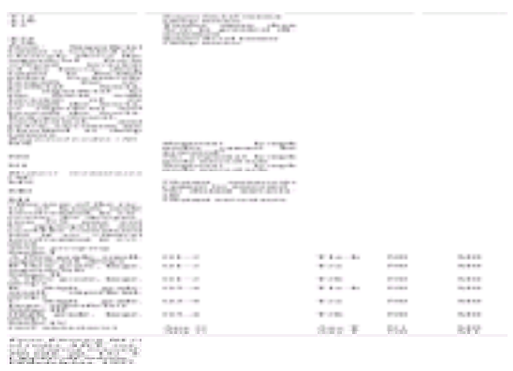


Table 3 Staging soft tissue sarcomas

Natural history

Although the stage incorporating tumor grade is more predictive of clinical course than is the histologic classification of soft tissue sarcomas, some unique features exist based on the tissue subtype. However, even among pathologists with particular experience and expertise in soft tissue sarcomas, there is often discordance of histologic classification. While striated muscle comprises more than one-third of human body mass, rhabdomyosarcomas are rare, although they are the most common soft tissue sarcomas in children under 15 years of age. These are managed best by a combination of treatments, chosen according to the location of the tumor. When all gross tumor can be resected with minimal morbidity, resection followed by adjuvant chemotherapy affords excellent control, radiation therapy being reserved for close margins. Over one-third of childhood rhabdomyosarcomas originate in the head and neck, where removal of biopsy specimens is followed by primary irradiation and adjuvant chemotherapy. Such combined therapy has improved the 3-year survival rate of children from about 20 per cent in the early 1960s to more than 70 per cent now. Survival ranges from 88 per cent for patients with localized disease to 32 per cent for children who have metastatic disease present at the time of diagnosis.

Liposarcoma is a malignant fatty tumor and may be multicentric. It is not known whether this is true multicentricity or whether these other foci represent metastases to other fatty tissue sites. Liposarcomas are frequently large, arising in the proximal extremity, trunk, or retroperitoneum. Many tumors exceeding 23 kg in weight have been reported, including one weighing 125 kg. They may regress quite dramatically following radiotherapy, but well-differentiated liposarcomas may be aggressive locally and recur many times before presenting with metastases. Pleomorphic and lipoblastic liposarcomas tend to be of higher grade and to exhibit distant metastases.

Malignant fibrous histiocytoma is the most frequent classification applied to soft tissue sarcomas, and encompasses many tumors that would have been classified as fibrosarcomas or pleomorphic rhabdomyosarcomas in the past. It implies a high-grade sarcoma that lacks specific features of differentiation. The disease is currently thought to be derived from fibroblasts rather than the histiocyte that gave the tumor its name. Although classified as a soft tissue sarcoma, malignant fibrous histiocytomas may arise from bone and a variety of other tissues. Fibrosarcoma was once the most common soft tissue sarcoma: when studying older series one should remember that most fibrosarcomas would now be classified as malignant fibrous histiocytomas or other types.

Leiomyosarcomas arise from or resemble smooth muscle cells; they can arise from the walls of blood vessels at any site in the body. When they arise in the wall of the gut, their behavior correlates with the number of mitoses in 50 high-power fields. For other soft tissue sarcomas the number of mitoses in 10 fields is used in determining the grade of the tumor. Synovial sarcomas arise most commonly in the lower extremities, in a muscle away from a joint. Histologic glandularity is an important prognostic feature: tumors with over 50 per cent glandularity and less than 15 mitoses in 10 high-power fields are unlikely to prove fatal if properly treated.

Neurofibrosarcomas (also called malignant schwannomas and malignant neurilemmomas) originate from or resemble neural sheath tissue. They frequently arise in patients with von Recklinghausen's disease, about 15 per cent of whom will develop sarcomas. Epithelioid sarcomas usually arise in the extremity, where they often spread to areas of skin, fatty tissue, and bone. Whether this represents multicentricity or a pattern of metastasis is not known. They metastasize to regional lymph nodes twice as often as to the lungs. Some surgeons advocate elective dissection of regional draining lymph node groups, and a biopsy should be taken of any palpably enlarged node. Surgical margins should be more generous for epithelioid sarcomas than for most other soft tissue sarcomas. Alveolar soft part sarcomas are rare, even for sarcomas, and have no benign counterpart. In adults the tumor usually affects an extremity, while in children it is usually a head and neck tumor. Protracted but ultimately fatal progression is a common clinical course.

Angiosarcomas account for only 1 to 2 per cent of soft tissue sarcomas, but these are usually high grade and behave aggressively. Lymphangiosarcoma arising in an edematous extremity is often diffuse and radical amputation is typically required for its control. Many of these tumors arise in the skin and subcutaneous tissues, in contrast to the deep-tissue location of most soft tissue sarcomas. Half of all angiosarcomas arise in the head and neck: these are particularly aggressive, often progress, and are unresponsive to treatment. Any attempt to resect them should be preceded by examination of numerous biopsy specimens from the anticipated resection margin, permanent-section histology being used to plan the resection. Evidence of microscopic tumor is often present 6 to 8 cm from the apparent tumor border on permanent sections, even though it was not apparent on frozen sections.

For many years Kaposi's sarcoma was a rare sarcoma seen in elderly Jewish men in Europe or the United States of America. It was clinically characterized as multiple pigmented sarcoma nodules, usually beginning on the legs as red nodules and slowly spreading. The condition was rather indolent and could be excised or treated by irradiation. Later, it was recognized as a common neoplasm in certain areas of Africa, where it behaved more aggressively but was sensitive to combination cytotoxic chemotherapy. A third presentation of Kaposi's sarcoma was reported in transplant recipients who were receiving immunotherapy. Although it affects well under 1 per cent of transplant recipients, such patients have a 200-fold increase in risk.

Since 1981 Kaposi's sarcoma has presented in epidemic form in patients with AIDS. Its clinical presentation and behavior is similar to that seen in immunosuppressed transplant recipients. Kaposi's sarcoma is not seen in transfusion-related AIDS and is uncommon in AIDS associated with drug abuse, but it occurs in nearly half of male homosexuals with AIDS. The epidemic form of Kaposi's sarcoma presents with smaller lesions, usually on the chest, arms, head, and neck. It frequently involves the mucosa of mouth and anus and half of affected patients have visceral lesions by the time of diagnosis. These are usually asymptomatic, although gastrointestinal tract lesions may bleed.

Epidemic Kaposi's sarcoma has its own staging system that is based on extent of disease and modified by A and B subtypes, which refer to systemic symptoms. Minimal tumor without B symptoms, which include weight loss, fever, or ongoing infection, is either observed or treated with a-interferon; if B symptoms are present, single-agent chemotherapy is generally used, typically with the vinca alkaloids. More extensive tumor is generally treated with interferon or chemotherapy in the absence of B symptoms; in patients with B symptoms or rapid growth of tumor, more intensive chemotherapy with doxorubicin and/or etoposide is suggested. Irradiation is directed toward symptomatic or disfiguring tumor nodules. A response to interferon therapy is largely seen in those without B symptoms. Although patients often respond to chemotherapy, these are commonly partial responses and of limited duration.

Treatment (Fig. 2)

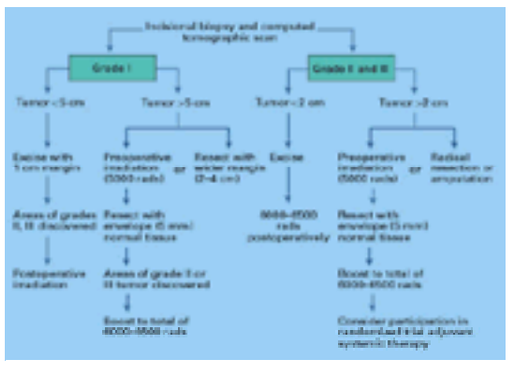


Fig. 2. Management of soft tissue sarcomas.

Surgery

Although treatment is determined by the specific diagnosis and the stage of the soft tissue sarcoma, the goal is to eliminate both the primary tumor and any metastatic disease. Basic local control may be accomplished by surgical resection. In the first half of the twentieth century this was most commonly attempted by resection of all apparent tumor, often to margins reported to be free by pathologic evaluation. Subsequent review of patients so treated reported local failure rates ranging from 50 to over 90 per cent (Table 4), and inapparent tumor obviously extended well beyond visible disease. Wider excisions with more generous margins improved the local failure rate to just less than 50 per cent. In the 1950s and 1960s the widespread acceptance of radical surgery, particularly radical amputations for extremity sarcoma, lowered the rate of local failure to between 12 and 20 per cent. In patients for whom surgery is the only available treatment, the lessons of this era are still appropriate. Local control of soft tissue sarcoma requires resection with a generous margin of seemingly normal tissue. For low-grade sarcomas (other than epithelioid) this may be as little as 1 cm of clean margin; for high-grade sarcomas, 4 cm in the closest dimension is required. For tumors in the center of a muscle compartment, this goal can be accomplished by removal of the entire compartment from origin to insertion, which clearly causes appreciable functional and cosmetic morbidity. Many soft tissue sarcomas arise in less convenient locations, near a joint or between two muscle compartments. One recent analysis found that only 30 per cent of soft tissue sarcomas of the extremities occurred in deep intracompartmental locations, and only one-half of these could be treated by compartment resection or primary myectomy. Treatment of such lesions of the extremity by surgery alone requires amputation. For high-grade sarcomas of the foot, a below-knee amputation would be required: an above-knee amputation is needed to achieve local control of such tumors affecting the ankle or lower leg. For tumors of the mid- to distal thigh, hip disarticulation

suffices, but proximal sarcomas of high grade near the hip require hindquarter amputation or hemipelvectomy. The latter may be modified according to the location of the primary tumor. If the iliac wing can be spared, rehabilitation to ambulation with a prosthesis is aided. Similar procedures exist for high-grade sarcomas of the upper extremity, up through interscapulothoracic or forequarter amputation for sarcomas of the proximal arm and distal shoulder. This involves removal of both clavicle and scapula and can be modified to encompass adjacent or attached ribs for more proximal tumors of lower grade. Resection of sarcomas should not be limited to standard procedures, but should be tailored to the biology and location of the individual tumor.

Local excision	Rosenwell Park, 1975	95
Wide excision	Rosenwell Park, 1975	60
Radical resection	Memorial Sloan-Kettering, 1975	20
Amputation	Memorial Sloan-Kettering, 1975	7
Radiation and surgery	Massachusetts General, 1980	6
Chemotherapy, radiation, and surgery	UCLA, 1984	

Table 4 Comparison of local control

Although such radical surgical procedures greatly affect the patient, they are often well tolerated: many patients demonstrate great resilience and resourcefulness in dealing with the loss of a major body part and the attendant disability and cosmetic defect. Physical therapy, rehabilitation, the availability of sophisticated prostheses, and psychosocial support are all essential parts of caring for the patient faced with such ablative surgery.

Radiotherapy

Radiation therapy has been used in attempts to achieve local control of soft tissue sarcomas. Although effective control was sometimes achieved, shrinkage of the tumor was followed all too often by progression. These events led to the concept that soft tissue sarcomas were ‘radioresistant'. The use of high-dose radiation increased the rate of tumor control, but the attendant morbidity also increased. Local control rates, even following aggressive irradiation, are generally inferior to those achieved by surgery alone. Consequently, radiation alone should be reserved for patients whose tumors cannot be resected because of size or location, for those who are medically unfit for surgery, or those who refuse surgical management. As mentioned above, desmoid tumors, although they are not true sarcomas, are well managed by radiation therapy alone.

Combination therapy

Combined surgery and irradiation was pioneered by Suit and colleagues. Experimental data suggested that radiation at a level of 60 Gy would control microscopic deposits of sarcoma cells, although larger solid sarcomas would have hypoxic areas that might not be eliminated at this dose. However, surgery alone needs to be as radical to achieve local control since microscopic deposits of sarcoma cells extend beyond the primary tumor mass. Radiation, therefore, could be used to eliminate subclinical extensions of disease into the surrounding, normal appearing tissue with surgical resection of gross tumor, sparing adjacent major nerves and vessels.

A clinical example contrasting combined treatment with surgery alone might be a grade III, 6-cm malignant fibrous histiocytoma arising in the gastrocnemius muscle of the left leg of a 26-year-old woman (G3 T2 N0 M0). Surgical treatment alone (Fig. 3) would require at least resection of the posterior compartment of the calf. If the posterior compartment were small, or if the tumor was tethered or too close to the deep structures of the leg, amputation above the knee would be required to achieve local control of the sarcoma. Alternatively, preoperative irradiation using a four-field technique that spares the anterior tissues at 2 Gy/day for a total of 45 Gy could be followed by a resection of the tumor with a small envelope of normal tissue 3 to 5 mm thick to ensure that all gross tumor was removed. After wound healing is complete a 15- to 20-Gy boost dose of irradiation to the region at highest risk (brachytherapy or particle therapy using electron or proton beams) would be administered for a total dose of 60 to 65 Gy. Such treatment is at least as effective as radical resection or amputation alone in achieving local control of soft tissue sarcomas, and produces minimal cosmetic or functional impairment (Table 4).

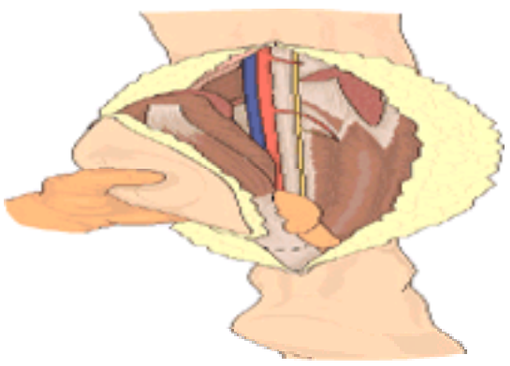


Fig. 3. Radical surgery alone. Resection of the posterior compartment of the calf for a small soft tissue sarcoma arising in the center of that compartment. For a low-grade sarcoma located well within the compartment, surgery can be used without radiation to achieve local control. There is considerable functional and cosmetic loss with removal of the gastrocnemius and soleus muscles.

Radiation therapy can be added to limited surgery for control of soft tissue sarcomas either pre- or postoperatively. A strong theoretical case can be made for delivering the radiation prior to the tumor resection (Fig. 4). First, the volume of tumor subjected to irradiation may be limited to those tissues at risk by virtue of the biology of the sarcoma and its anatomic location, while postresection radiation must include all tissue planes that may have been contaminated by the surgical procedure, drain tracts, or hematoma-borne clumps of viable tumor cells. This option generally requires a larger radiation field. Second, radiation therapy can begin immediately, while postoperative radiotherapy is delayed for a week or even longer, depending on the rate of wound healing. It may be delayed for several weeks by persisting seromas. Third, after irradiation the tumor commonly regresses in size and develops a plane of peritumor edema that marks the borders of infiltrative sarcomas quite well. Regression allows defined resections to be performed, sparing the majority of normal tissue in the region. The well-oxygenated tumor at the periphery of the sarcoma is most likely to be sterilized, increasing the distance between the dissection and the location of any viable tumor, which tends to be near the center of the tumor mass in relatively hypoxic regions of the sarcoma. This plan further diminishes the likelihood of tumor implantation at the time of conservative resection.

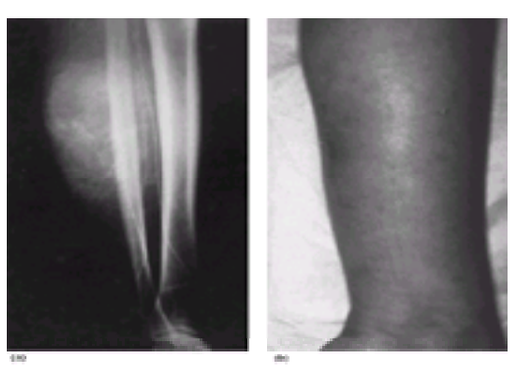


Fig. 4. Preoperative irradiation and surgery. An 8-cm soft tissue sarcoma of the posterior compartment, too large to resect with surgery alone, treated with preoperative

irradiation and surgery. (a) Arteriogram of tumor. (b) Minimal cosmetic and functional deficit.

Pre- and postoperative irradiation have not been compared in a prospective, randomized trial. In a single institution's experience, preoperative irradiation was administered (Table 5). Patients with smaller tumors of lower grade were more likely to have undergone resection prior to referral, often with the expectation that the tumor was benign. Many of the tumors were not resectable at the time of referral. Although there was no apparent difference in outcome for the smaller tumors, larger tumors showed better local control following preoperative radiation therapy.

Size (cm)	Postoperative No.	Local control (%)	Preoperative No.	Local control (%)
< 25	14	82	5	100
26–50	45	85	9	89
51–100	29	84	32	91
101–150	8	83	17	100
151–200	5	0	9	73
> 200	1	0	6	100
Total	102	80	78	92

Table 5 Local control as a function of pre- or postoperative irradiation (Massachusetts General Hospital, 1988)

When irradiation is added to resection, the surgical procedure is directed toward the removal of all gross tumor with a sufficient margin of normal tissue to allow confirmation of complete excision—gross margin of 5 to 10 mm. The resection should not be performed with a cautery knife because evaluation of surgical margins cannot be accurately performed. The surface of the specimen should be painted with India ink by the surgeon or pathologist to facilitate evaluation of the resection margins. Attention is then focused on elimination of dead space and prevention of hematomas or seromas. When a large potential dead space exists at the conclusion of the resection, it is best filled with a transposed muscle flap. If the tissues are sufficiently mobile to fall together without the need for transposed tissue, suction drains alone may suffice. The area should be immobilized, particularly if it is in an extremity, and suction drains should be left in place until the threat of seroma formation is past. Early mobilization and drain removal accounts for many wound complications in this setting. Poor healing is most common where irradiated tissues are very thin, such as about the knee, or in areas of abundant fat, such as the buttock or proximal thigh. In the latter setting, fat necrosis following radiation and surgery can lead to late wound breakdown requiring revision. When breakdown occurs it is often best managed by coverage with a myocutaneous flap.

Chemotherapy, surgery, and irradiation for local control

Eilber pioneered the use of intra-arterial doxorubicin followed by limited surgical resection and irradiation for the treatment of extremity sarcomas. This technique allowed excellent local control while sparing the extremity. No direct comparison has been performed between the use of irradiation and limited surgery and this treatment with the addition of intra-arterial chemotherapy. When Eilber compared intravenous (systemic) and intra-arterial routes for doxorubicin administration, the results were equal.

Adjuvant chemotherapy

The higher the grade of soft tissue sarcomas, and the greater the size, the higher the risk of metastases. Local control is usually achieved with either radical surgery or radiation and surgery; survival reflects the distant dissemination of sarcoma. Adjuvant chemotherapy to increase survival rates seemed possible.

Cytotoxic chemotherapy offers temporary remission to many patients with advanced, unresectable sarcomas or widespread metastatic disease. A host of combinations of active agents have been evaluated, in both randomized and non-randomized trials. The most active agents are doxorubicin and ifosfamide, the latter typically given with mensa for bladder protection. A combination favored in many centers is MAID (mensa, adriamycin [doxorubicin], ifosfamide, and dacarbazine) and another combines epirubicin and ifosfamide. With aggressive dose-intense programs, response rates have varied from less than 20 to over 65 per cent in selected series. The proportion of patients achieving a complete clinical remission varies from none to over 20 per cent, but is most commonly 10 to 15 per cent. Some of those achieving complete remission enjoy long-term survival freeof disease.

These same agents and combinations have been used as adjuvants in an attempt to destroy subclinical metastases in patients with locally controlled high-grade soft tissue sarcomas. Several non-randomized, comparative trials have suggested that this approach confers a survival benefit, but the majority of randomized trials have failed to demonstrate any advantage. One trial that suggested a benefit was performed at the National Cancer Institute (**NCI**), United States. As seen in Table 6, survival of the treated patients paralleled that of other studies, but the control group fared less well. The chemotherapy program used in the NCI trial was quite toxic, and a high proportion of patients developed some degree of cardiac impairment, demonstrable by exercise testing or at rest. It was initially suggested that the benefit seen in this trial might reflect the use of chemotherapeutic agents at the maximum tolerable dose. A subsequent trial performed by the same investigators using lower doses of drugs did not show any survival advantage with the use of higher drug doses. No beneficial adjuvant chemotherapy program for high-grade soft tissue sarcomas has been established. This may reflect a modest benefit in an uncommon tumor leading to clinical trials of insufficient size to demonstrate benefit. An overview of published randomized trials suggests a significant benefit to adjuvant chemotherapy in high-grade soft tissue sarcoma, but it has been faulted as not including all trials in existence. This remains an area of active investigation.

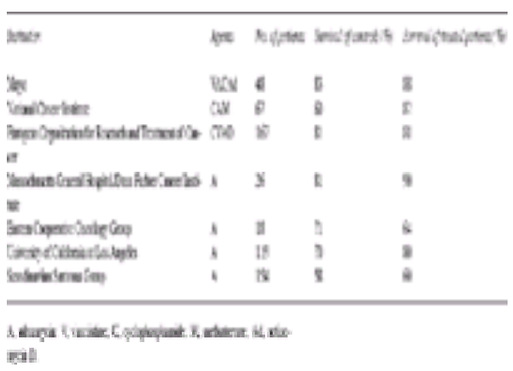


Table 6 Adjuvant chemotherapy for soft tissue sarcoma of the extremity

Retroperitoneal sarcoma

Sarcomas arising in the retroperitoneum are a particular problem because their adjacency to essential organs does not allow a wide resection. Without being able to assure that a margin of normal tissue including inapparent microscopic extentions of disease has been removed, local failure is very likely. Radiation has been combined with surgery in an attempt to sterilize these micrometastases likely to be residual after resection. Clips placed by the surgeon at the extent of tumor resection aid the radiation oncologist in treatment planning. When a diagnosis is made prior to resection, which should usually be the case with contemporary imaging techniques, radiation may be given prior to resection. This allows the tumor to displace bowel and other radiation-sensitive tissues, protecting them from high radiation doses. The same radiation given postresection would find the bowel loops adherent to the tumor bed and exposed to the full dose of radiation. Because retroperitoneal sarcomas are relatively uncommon (e.g. fewer than 500 annually in the United States), insufficient numbers exist to assess the value of combined therapy in this situation on a prospective basis. Just as soft tissue sarcomas of most of the body commonly metastasize to the lungs as the capillary field in which they ‘catch',

retroperitoneal sarcomas tend to metastasize to the liver.

Isolated recurrences and metastases

Patients who have an isolated recurrence or a limited number of pulmonary metastases may be treated by aggressive local or regional therapy; more than one-third of such patients survive for 5 years. Giuliano and associates reported achieving local control in 92 per cent of patients with isolated local recurrence of extremity or truncal soft tissue sarcoma. The actuarial 5-year survival in surgically treated patients was 87 per cent.

Patients with pulmonary metastases from soft tissue sarcoma can be successfully managed by resection. A more favorable prognosis correlates with fewer metastases, a longer interval between treatment of the primary tumor and diagnosis of the pulmonary metastases, and a longer doubling time, measured from interval chest radiographs. An analysis by Roth *et al.* confirmed the value of these prognostic factors, but found that no single limit is sufficiently accurate to exclude patients from one attempt at control by resection of pulmonary metastases. Patients who present with recurrent pulmonary metastases following thoracotomy and resection of metastases may be successfully treated by a second procedure. Of 29 patients so treated in the NCI series, there was a 22 per cent 3-year actuarial survival. Because surgical resection of pulmonary metastases has a reasonable expectation of success, patients with high-grade soft tissue sarcomas should be monitored at 6-month intervals for at least 3 years after primary therapy. The use of chest CT scans has been advocated for this follow-up, but there are no data to suggest their superiority to chest radiographs.

Further reading

Antman K *et al.* Pooled results from three randomized adjuvant studies of doxorubicin versus observation in soft tissue sarcoma: 10 year results. *Proceedings of ASCO* 1990; 9: 1204.

Benjamin RS *et al.* Should patients with advanced sarcomas be treated with chemotherapy? *European Journal of Cancer* 1998; **34**: 958–65.

Chevreau C *et al.* Phase I–II trial of intensification of the MAID regimen with support of lenograstim (rHuG-CSF) in patients with advanced soft-tissue sarcoma (STS). *American Journal of Clinical Oncology* 1999; **22**: 267–72.

Demetri GD. Major developments in the understanding and treatment of soft-tissue sarcomas in adults. *Current Opinion in Oncology* 1998; **10**: 343–7.

Demetri GD *et al.* NCCN sarcoma practice guidelines. National Comprehensive Cancer Network. *Oncology* 1998; **12**: 183–218.

Eilber FR *et al.* High-grade soft tissue sarcomas of the extremity: UCLA experience with limb salvage. *Progress in Clinical and Biological Research* 1985; **201**: 59–74.

Ensoli B, Sirianni MC. Kaposi's sarcoma pathogenesis: a link between immunology and tumor biology. *Critical Reviews in Oncogenesis* 1998; **9**: 107–24.

Ensoli B, Sturzl M. Kaposi's sarcoma: a result of the interplay among inflammatory cytokines, angiogenic factors and viral agents. *Cytokine and Growth Factor Reviews* 1998; **9**: 63–83.

Enzinger FM, Weiss SW. *Soft tissue tumors*, 3rd edn. CV Mosby, St Louis, 1995.

Giuliano AE, Eilber FR, Morton DL. The management of locally recurrent soft-tissue sarcoma. *Annals of Surgery* 1982; **196**: 87–91.

Ham SJ *et al.* Soft tissue sarcoma of the extremities. A multimodality diagnostic and therapeutic approach. *Cancer Treatment Reviews* 1998; **24**: 373–91.

Kiel KD, Suit HD. Radiation therapy in the treatment of aggressive fibromatosis (desmoid tumors). *Cancer* 1984; **54**: 2051–5.

Mankin HJ, Lange TA, Spanier SS. The hazards of biopsy in patients with malignant primary bone and soft-tissue tumors. *Journal of Bone and Joint Surgery* 1982; **8**: 1121–7.

Mazeron JJ, Suit HD. Lymph nodes as sites of metastases from sarcomas of soft tissues. *Cancer* 1987; **60**: 1800–8.

Neipel F, Fleckenstein B. The role of HHV-8 in Kaposi's sarcoma. *Seminars in Cancer Biology* 1999; **9**: 151–64.

O'Byrne K, Steward WP. The role of chemotherapy in the treatment of adult soft tissue sarcomas. *Oncology* 1999; **56**: 13–23.

Orlow I *et al.* Alterations of 1NK4A and INK4B genes in adult soft tissue sarcomas: effect on survival. *Journal of the National Cancer Institute* 1999; **91**: 73–9.

Pisters PW. Combined modality treatment of extremity soft tissue sarcomas. *Annals of Surgical Oncology* 1998; **5**: 464–72.

Roth JA *et al.* Differing determinants of prognosis following resection of pulmonary metastases from osteogenic and soft tissue sarcoma patients. *Cancer* 1985; **55**: 1361–6.

Seynaeve C, Verweij J. High-dose chemotherapy in adult sarcomas: no standard yet. *Seminars in Oncology* 1999; **26**: 119–33.

Wood WC *et al.* Radiation and conservative surgery in the treatment of soft tissue sarcoma. *American Journal of Surgery* 1984; **147**: 537–41.

Zalupski MM *et al.* Defining the role of adjuvant chemotherapy for patients with soft tissue sarcoma of the extremities. In: Salmon SE, ed., *Adjuvant therapy of cancer VII*, pp. 386–92. JB Lippincott, 1993.

38.1 Hamartomas, vascular malformations, and congenital defects

Calvin O. McCall and Thomas G. Cropley

Hamartomas
Hamartomas of epidermal structures: epidermal nevi
Hamartomas of dermal structures: connective tissue nevi
Hemangiomas and vascular malformations
Lymphangiomas
Further reading

Hamartomas

Hamartomas of epidermal structures: epidermal nevi

The term nevus refers to a hamartomatous malformation of the skin. Nevi may be principally composed of dermal elements (connective tissue nevi), melanocytes (nevocellular nevi), vascular elements, or epidermal elements (epidermal nevi). Epidermal nevi are verrucous or wart-like and are frequently present at birth, although they may develop later in childhood or (rarely) in adult life. They may be small, resembling true viral warts, but more typically they form large, linear plaques or linear arrays of warty papules. Rarely, large areas of the body surface may be covered by linear or swirling lesions. Epidermal nevi are typically asymptomatic, although large, linear nevi may become spontaneously inflamed, with pruritus and erythema of the lesions and surrounding skin.

Epidermal nevi are generally of little consequence to the health of the patient. Very rarely, they may be associated with congenital malformations of underlying osseous and soft-tissue structures, particularly when they occur on the head and neck (epidermal nevus syndrome). Cytologic atypia is not seen and malignant transformation within epidermal nevi is rare, except in nevus sebaceous (see below). Epidermal nevi are most commonly removed for cosmetic reasons. Such treatment may involve full-thickness excision with primary closure, partial-thickness excision, electrodesiccation, CO_2 laser ablation, dermabrasion, or cryosurgery. Regardless of the treatment method employed, significant scarring is a risk.

The nevus sebaceous is a type of epidermal nevus that occurs predominantly on the scalp and face (Fig. 1). Histologically, the same type of epidermal proliferation as is seen in other epidermal nevi characterizes nevus sebaceous, but large, mature sebaceous glands and apocrine glands are also present. Before puberty, the nevus sebaceous is typically a smooth-surfaced, flat plaque with a slightly yellowish or yellow–orange color. When it occurs on the scalp, which is the most common site, the absence of hair within the lesion may bring it to attention. At puberty, the sebaceous elements within the nevus begin to hypertrophy and the surface becomes more verrucous. Nevus sebaceous may undergo malignant transformation, with basal-cell carcinoma forming within the lesion. Since such transformation occurs in only about 5 per cent of patients with nevus sebaceous, the common practice of excising these lesions prophylactically is controversial. Yet, there is support for removing these lesions before puberty. Benign tumors such as syringocystadenoma papilliferum may also occur within these nevi.

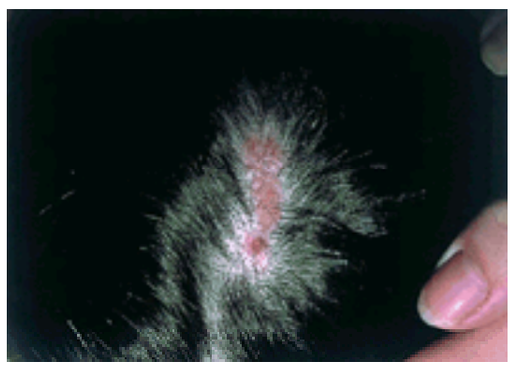


Fig. 1. Nevus sebaceous.

Hamartomas of dermal structures: connective tissue nevi

Connective tissue nevi are generally present at birth. They consist of firm, flesh-colored papules and plaques, and are usually solitary. The orifices of hair follicles may be prominent within the lesions, giving them the appearance of pig skin. Connective tissue nevi are usually idiopathic and are not associated with other disease, except the shagreen patch, which occurs in tuberous sclerosis. Histologically, connective tissue nevi are composed of dense aggregations of collagen fibers, elastic fibers, or both. Connective tissue nevi are harmless, although they may be excised for cosmetic or functional reasons.

Hemangiomas and vascular malformations

Infantile hemangiomas (Fig. 2), the most common neoplasms of infancy, result from a benign proliferation of endothelial cells within the superficial dermis, the deep dermis and subcutis, or in both locations. Previously, infantile hemangiomas were classified as ‘capillary', ‘cavernous', or ‘mixed' hemangiomas. Histopathologically, these terms are incorrect and classification is more correctly based on their location within the dermis: ‘superficial' or ‘deep'. A given lesion may exhibit components of both. This classification has clinical, prognostic, and therapeutic utility. Superficial hemangiomas tend to be bright red and exophytic; deep hemangiomas appear as subcutaneous nodules, some with a blue tint and some with a superficial component including telangiectases.



Fig. 2. Infantile hemangioma.

Infantile hemangiomas occur in approximately 3 per cent of all newborns and can be found in up to 10 per cent of 1 year olds. They may be present at birth but are more commonly noted after the second week of life. Hemangiomas are most common on the head and neck, but may occur anywhere on the body including the mucous membranes. Eighty per cent are solitary. Individual lesions typically begin as pink macules that thicken to form red–purple papules or plaques with a spongy consistency. They may grow to exceed 10 cm in diameter. The surface of a superficial lesion may be studded with small capillary tufts, which produce the resemblance to a strawberry, thus, the name ‘strawberry hemangioma'. Hemangiomas lacking a superficial component will not have this characteristic.

Infantile hemangiomas seldom require treatment; approximately 70 per cent involute spontaneously by the age of 6 or 7 years, with an excellent cosmetic result.

Treatment may be required, however, for lesions causing functional impairment, hemorrhage, or extensive tissue destruction. Administration of oral or intralesional corticosteroids may stimulate gradual involution and is often used initially for rapidly growing hemangiomas. Surgical excision and CO_2 laser ablation are often successful, but may produce unacceptable scarring. Other treatments include cryotherapy, sclerosing, embolization with compression, and radiotherapy. The pulsed-dye laser is effective for some superficial hemangiomas and is the treatment of choice for residual telangiectases in resolving hemangiomas. It is not effective for deep lesions.

Cutaneous hemangiomas, when multiple, may be associated with visceral hemangiomas and significant mortality. All children with multiple lesions should be considered for imaging procedures to rule out visceral involvement. In the case of large hemangiomas, the Kasabach–Merritt phenomenon (thrombocytopenia and disseminated intravascular coagulation) must be considered. Affected children may exhibit evidence of thrombocytopenia and anemia, as well as rapid changes in the size or character of their hemangioma. Such children require immediate hematologic evaluation, as the mortality rate is greater than 20 per cent. Recent studies suggest this phenomenon is not associated with common hemangiomas but with a more aggressive vascular neoplasm known as kaposiform hemangioendothelioma, which may be congenital or acquired during early life. Admission to hospital, blood products, surgical resection, and systemic steroids have typically been used to treat these life-threatening neoplasms. Recently, our increased understanding of angiogenesis has resulted in new therapies and potential therapies. Interferon-a_{2a} and -a_{2b} are now indicated for the treatment of life-threatening hemangiomas. These interferons decrease the production of the angiogenic protein, fibroblast growth factor. They are administered subcutaneously over a period of 6 to 12 months. Spastic diplegia, fever, neutropenia, and mild elevations of liver enzymes may complicate administration. Other inhibitors of angiogenesis are in clinical trials and may prove useful in the future.

Cherry hemangiomas are generally smaller than strawberry hemangiomas, and tend to be round and bright red. They begin to appear during childhood and continue to develop throughout adult life. Cherry hemangiomas seldom require treatment except for cosmetic reasons: they are easily treated with shave excision, punch excision, or electrodesiccation. Occasionally, they thrombose, changing in color from bright red to purplish-black. This color change may alarm patient and physician alike, since the apparent sudden appearance of a black papule raises the spectre of nodular malignant melanoma. Complete excision and microscopic examination are indicated in such cases.

Port-wine stains (nevus flammeus), unlike hemangiomas, do not represent a proliferation of excessive numbers of vascular elements, but are caused by ectasia of normal numbers of vessels within the skin. The superficial dermal vessels that comprise port-wine stains are capillary size or slightly larger. Port-wine stains are commonly present at birth; they occur in approximately 0.3 per cent of neonates. Early in life, port-wine stains are pink or purplish-pink in color, and macular. With time, the vascular channels become more ectatic, imparting a lumpy appearance to the skin. Mature port-wine stains are frequently dark purple. Although most common on the head and neck, port-wine stains may also be seen on the trunk (particularly in the pectoral area) and on the extremities (usually the arms). Until recently, their treatment was unsatisfactory and produced unacceptable scarring. The development of tuneable dye lasers has revolutionized their treatment, making it possible to remove these lesions partially or completely without excessive scarring.

Occasionally, port-wine stains are associated with other problems. Unilateral glaucoma may be associated with port-wine stains of the orbital region. The Sturge–Weber syndrome is characterized by the presence of a port-wine stain in the distribution of the first branch of the trigeminal nerve associated with a similar vascular malformation of the underlying cerebral cortex and meninges (Fig. 3). Seizures, mental retardation, and motor cortical deficits are seen in many patients. Similarly, the Klippel–Trenaunay–Weber syndrome is the association of a port-wine stain of an extremity or the trunk with arteriovenous malformation and hemihypertrophy.



Fig. 3. Port-wine stain of Sturge–Weber.

Pyogenic granulomas are common vascular lesions that present as rapidly growing papules or nodules, which are friable (Fig. 4). They may bleed profusely after minor trauma. ‘Pyogenic granuloma' is, however, a misnomer. This lesion is an eruptive superficial hemangioma unrelated to infection. Treatment with electrosurgery, CO_2 laser ablation, or simple excision is usually curative, although lesions may recur. Since the differential diagnosis of rapidly growing and bleeding nodules includes nodular malignant melanoma, these lesions should be biopsied.



Fig. 4. Pyogenic granuloma.

Lymphangiomas

Lymphangiomas are composed of capillary-sized lymphatic vessels or larger, cavernous lymphatics. Lymphangioma circumscriptum is a common type of lymphangioma resulting from a superficial proliferation of capillary-sized lymphatic vessels. They appear as multiple vesicles on the cutaneous surface, and have a deeper component consisting of large, fluid-filled spaces (cisternae) within adjacent muscle tissue. The appearance of the superficial portion of lymphangioma circumscriptum has been likened to frog eggs. Small precipitates of red blood cells can often be seen at the bottom of the lymphatic vesicles. Treatment of lymphangioma circumscriptum may be attempted for cosmetic reasons. Both excision and CO_2 laser ablation have been successfully employed, but recurrences are common because of the deep cisternae.

Cavernous lymphangiomas and cystic hygromas are proliferations of lymphatic vascular tissue consisting of larger lymphatic spaces than those seen in lymphangioma circumscriptum. Cystic hygromas are the more common of the two; they most frequently occur on the neck, although they may be seen on the head or inguinal regions. Cystic hygromas behave like benign tumors and grow slowly. Large lesions may cause functional impairment, particularly on the neck. Surgical excision is difficult, but highly successful for cystic hygromas. Excision of cavernous lymphangiomas is less likely to be curative and recurrence is common.

Further reading

Beaubien ER, Ball NJ, Storwick GS. Kaposiform hemangioendothelioma: a locally aggressive vascular tumor. *Journal of the American Academy of Dermatology* 1998; **38**: 799–802. [This report

details the clinical and histologic features of kaposiform hemangioendothelioma.]

Dellon AL, Luethke R, Wong L, Barnett N. Epidermal nevus: surgical treatment by partial-thickness skin excision. *Annals of Plastic Surgery* 1992; **28**: 292–6. [This case report provides evidence that partial thickness excision is the treatment of choice for this nevus, but is not without cosmetic risk. See also the 'Invited Comment' by Ernest N. Kaplan.]

Fox BJ, Lapins NA. Comparison of treatment modalities for epidermal nevus: a case report and review. *Dermatologic Surgery and Oncology* 1983; **9**: 879–85. [The authors review the effectiveness of multiple therapies for the treatment of epidermal nevi.]

Jones EW, Heyl T. Naevus sebaceus. *British Journal of Dermatology* 1970; **82**: 99–117. [The authors present the clinical and histologic characteristics of the nevus sebaceous.]

Peachey RD, Lim C-G, Whimster IW. Lymphangioma of the skin. Review of 65 cases. *British Journal of Dermatology* 1970; **83**: 519–27. [This is an in-depth review of cutaneous lymphangiomas and includes a summary of surgical outcomes.]

Requena L, Sangueza OP. Cutaneous vascular proliferations. Part II. Hyperplasias and benign neoplasms. *Journal of the American Academy of Dermatology* 1997; **37**: 887–920. [This article is one in a series of articles extensively reviewing cutaneous vascular lesions.]

Sarkar M *et al.* Thrombocytopenic coagulopathy (Kasabach–Merrit phenomenon) is associated with kaposiform hemangioendothelioma and not with common infantile haemangioma. *Plastic and Reconstructive Surgery* 1997; **100**: 1377–86. [The authors suggest that common hemangiomas are not associated with the Kasabach-Merrit phenomenon.]

Solomon LM, Fretzin DF, Dewald RL. The epidermal nevus syndrome. *Archives of Dermatology* 1968; **97**: 273–85. [Developmental abnormalities occasionally associated with epidermal nevi are presented in this article.]

Tamayo L *et al.* Therapeutic efficacy of interferon alfa-2b in infants with life-threatening giant haemangiomas. *Archives of Dermatology* 1997; **133**: 1567–71. [Subcutaneous interferon a-2b is presented as an option for treating life-threatening hemangiomas.]

38.2 Benign tumours and cysts of the skin and its appendages

Calvin O. McCall and Thomas G. Cropley

Epidermal tumors
Epidermal (epidermoid) cysts
Dermal tumors
Benign tumors of epidermal appendages
Further reading

Epidermal tumors

Seborrheic keratoses are common, benign lesions that appear in middle age (Fig. 1). They can occur anywhere on the body except the mucous membranes. They are most common on the trunk and face, generally sparing the palms and soles. Individual lesions range in size from a few millimeters to several centimeters, are papular, and commonly have a ‘stuck on' appearance due to their superficial nature. Their surface may be verrucous or smooth and waxy. Seborrheic keratoses vary from flesh colored to black, but they can be pink, especially if irritated. They have no premalignant potential and can usually be distinguished from malignant melanoma, pigmented basal cell carcinoma, and nevi on clinical grounds. However, the resemblance of darkly pigmented seborrheic keratoses to malignant melanoma can be striking, and irritated lesions can resemble malignancies as well. When the diagnosis is in doubt, an excisional biopsy is recommended. Superficial or shave biopsies should be avoided whenever a diagnosis of malignant melanoma is considered. Removal of typical seborrheic keratoses is seldom indicated. For cosmetic purposes, removal can be accomplished by shave excision, curettage, or cryosurgery. Full-thickness excision and primary closure are not necessary. The purported association between eruptive seborrheic keratoses and malignancy, ‘the sign of Leser-Trelat', continues to be debated.

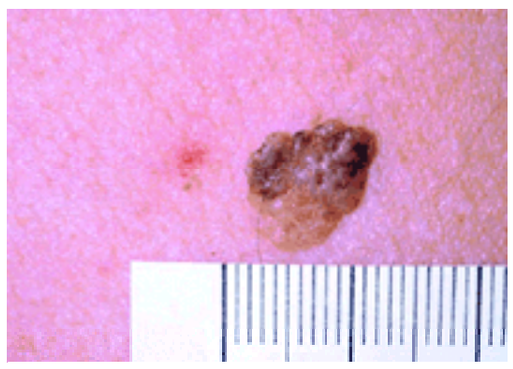


Fig. 1. Seborrheic keratosis.

Dermatosis papulosa nigra refers to the presence of numerous waxy black papules, variants of seborrheic keratoses, which occur in the periorbital region. This entity is seen in dark-skinned individuals, and is of cosmetic significance only. Lesions can be treated by shave excision, electrosurgery, or cryosurgery. Any treatment should be performed cautiously, since post-treatment pigmentary changes may be equally cosmetically distressing to the patient.

Clear cell acanthomas are uncommon, benign tumors of the skin occurring primarily on the lower extremities in adults. They typically appear as shiny, red to brown papules or plaques measuring up to 2 cm in diameter. A peripheral scale may be present. The differential diagnosis of clear cell acanthoma includes squamous cell carcinoma, amelanotic malignant melanoma, and pyogenic granuloma. Excisional biopsy is usually indicated to confirm the diagnosis, and is curative. When multiple lesions are to be treated, cryotherapy can be utilized to diminish scarring.

Epidermal (epidermoid) cysts

Epidermal cysts, often incorrectly referred to as sebaceous cysts, are the most commonly encountered cysts of the skin (Fig. 2). They occur in childhood, but are more common in early adulthood; the most common sites are the scalp, face, and back. Lesions are solitary or multiple, and present as 0.5 to 5.0 cm, spherical, slightly compressible, intradermal, or subcutaneous masses. A dilated follicular orifice may connect the cyst to the cutaneous surface. The wall of the cyst is composed of fully differentiated squamous epithelium and keratinocytes shed from the cyst wall form a white cheesy material within the cyst. Surgical treatment is required only if a cyst is symptomatic, and should be delayed if the cyst is severely inflamed or infected (Fig. 3). If not previously inflamed or manipulated surgically, cysts can be bluntly dissected free of the surrounding tissue following a simple equatorial incision of the overlying skin. Cysts on the face should be removed utilizing the smallest incision possible to minimize cosmetic deformity. Inflammation and secondary infection of epidermal cysts lead to fibrosis of the surrounding dermis, making removal more difficult. Such cysts are difficult to dissect free and *en bloc* excision becomes the treatment of choice. Squamous cell carcinoma and basal cell carcinoma can rarely form in the epidermal cyst wall. Pathologic examination of excised cysts is prudent.

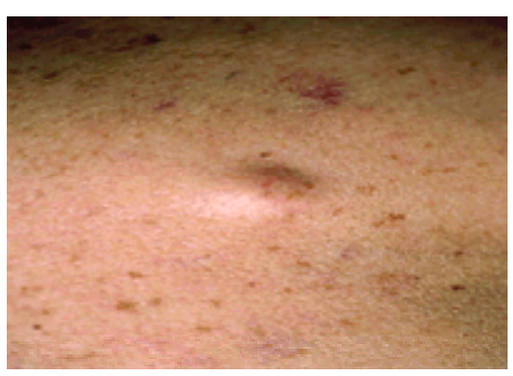


Fig. 2. Epidermal cyst.

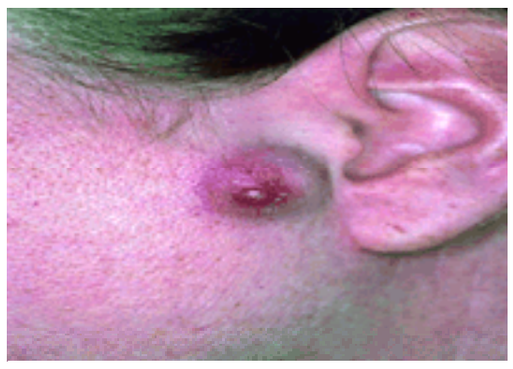


Fig. 3. Epidermal cyst, inflamed.

Multiple epidermal cysts may be associated with familial polyposis and osteomas in patients with Gardner's syndrome. Milia, epidermal cysts measuring 1 to 2 mm, may also be found in large numbers, but have no known association with heritable syndromes. Milia occur most frequently on the face, are a cosmetic nuisance, and can be removed with a needle or a No. 11 scalpel blade.

Pilar cysts, also known as trichilemmal cysts, are less common than epidermal cysts, and occur almost exclusively on the scalp. They are indistinguishable from epidermal cysts on clinical grounds, although small pilar cysts on the scalp often feel quite firm, unlike epidermal cysts. The wall of a pilar cyst is composed of epithelium resembling the outer root sheath of hair follicles. The cyst cavity is filled with dense glassy keratin resembling hair biochemically. Pilar cysts seldom become infected and are usually removed by equatorial incision, blunt dissection, and lateral compression.

Steatocystoma multiplex is inherited as an autosomal dominant trait. Multiple, thin-walled cysts develop during adolescence on the upper chest, neck, and back. Aspiration of the cyst contents may reveal a drop of oily fluid. Surgical treatment is usually not indicated; conservative management is frequently the best course.

Digital mucinous cysts present as 4 to 5 mm diameter, bluish, cystic-appearing nodules overlying the distal phalanx proximal to the nail fold. A longitudinal groove may form in the adjacent nail plate as the result of downward pressure by the cyst on the nail matrix. Mucinous cysts are not true cysts, since they lack an epithelial lining. Rather, they consist of a focus of mucinous degeneration of the dermis. Treatment is difficult: excision usually requires an advancement procedure for closure, and may result in distortion of the nail. Cryosurgery is frequently efficacious with less risk of nail distortion.

Dermal tumors

Neurofibromas, as a manifestation of neurofibromatosis, tend to be multiple with more than 100 lesions often present. Most commonly, however, these benign tumors of nerve sheath origin occur as solitary, flesh-colored tumors that protrude from the skin, without the associated findings of neurofibromatosis. They range in size from pinpoint to many centimeters in diameter. Smaller lesions characteristically invert partially when pressed with a finger (‘button-holing'). Patients with a neurofibroma should be examined carefully for the presence of other neurofibromas, café-au-lait spots, and axillary freckling—all markers for neurofibromatosis. Small, solitary neurofibromas can be excised simply, but large lesions often have an extensive subcutaneous component making excision more complex. All eroded, ulcerated, or otherwise atypical neurofibromas should be excised and submitted for microscopic examination to rule out malignant degeneration (sarcoma).

Dermatofibromas are common cutaneous lesions resulting from the proliferation of dermal fibroblasts. They can occur anywhere on the body. When compressed laterally these firm intradermal nodules often ‘dimple'. The surface is frequently shiny or corrugated and hyperpigmented. The differential diagnosis includes dermatofibrosarcoma protuberans, the malignant counterpart of dermatofibroma, metastatic carcinoma, Kaposi's sarcoma, and desmoplastic malignant melanoma. The correct diagnosis is usually apparent clinically, but incisional or excisional biopsy may be indicated in difficult cases. Treatment of dermatofibromas is often unsatisfactory; lesions frequently recur following excision and postsurgical scarring may be a cosmetic problem. Unless dermatofibromas are particularly large, become irritated, or present a diagnostic problem, it is often best not to treat them.

Acquired digital fibrokeratomas are benign lesions that develop on the lateral part of the fingers or toes in adults. Lesions are solitary, firm, and flesh-colored; they may be dome-shaped or conical, and are surrounded by a small collarette of skin (Fig. 4). Simple excision with primary closure is curative.



Fig. 4. Digital fibrokeratoma.

Fibrous papules of the nose present as 2 to 3 mm, superficial, flesh-colored papules on the nose or perinasal skin. These benign, fibrovascular proliferations may resemble small basal cell carcinomas: the diagnosis of fibrous papule of the nose is frequently made postoperatively by the pathologist. No treatment is required.

Leiomyomas of the skin are uncommon. Their presentation varies but most often they are multiple, tender, red to brown, dermal nodules. Malignant degeneration is not likely, but removal may be desirable for relief of pain or cosmetic concerns. Simple excision is the preferred treatment. If not amenable to surgery, calcium channel blockers may reduce pain by preventing contraction of smooth muscle cells.

Lipomas are benign tumors of the subcutaneous fat. They can appear at any time but are most common in middle age. These soft, dermal nodules may reach sizes greater than 10 cm. Although usually asymptomatic, pain may result from compression of nerves or adjacent structures, and angiolipomas, highly vascular lipomas, are frequently tender. When painful or interfering with functioning, treatment may be required. Simple incision, blunt dissection, and lateral compression easily express small lesions. Percutaneous liposuction is a relatively new treatment option for lipomas. Large lesions may envelop nerves and vessels, thus complicating surgical therapy. Additionally, the frontalis-associated subfascial lipoma is problematic. Typically it presents as a slightly protruding subcutaneous mass on the lateral forehead in older men. Dissection is extremely difficult because of the highly vascular muscle that invests it.

Benign tumors of epidermal appendages

A summary of tumors of the epidermal appendages is presented in Table 1.

Name	Site(s) of predilection	Appearance	Cell type	Comment
Trichofolliculoma	Face	Flesh-colored papule with tuft of white hair	Hair follicle	
Trichoepithelioma	Central face	Flesh-colored papule	Hair follicle	Multiple lesions are inherited; solitary ones are not inherited
Pilomatricoma	Face, arms	Large firm nodules, purple-red color	Hair matrix	40% occur in children
Pilar tumor of the scalp	Scalp	Large nodular, may ulcerate	Follicle outer root sheath	Elderly women 80%
Trichilemmoma	Face, neck, ears	Flesh-pink papule	Follicle outer root sheath	Multiple lesions occur in Cowden's syndrome
Sebaceous adenoma	Face, trunk	Yellow-orange papule	Sebaceous gland	Multiple lesions occur in Muir–Torre syndrome
Sebaceous epithelioma	Face	Yellow papule	Sebaceous gland	Muir–Torre syndrome
Apocrine/eccrine hidrocystoma	Eyelids	Blue cyst	Apocrine	
Syringocystadenoma papilliferum	Scalp, face	Crusted, wart-like nodule	Apocrine	Some arise in nevus sebaceus
Cylindroma	Scalp	Pink, firm nodules, "Turban tumors"	? Apocrine	Associated with multiple trichoepithelioma
Eccrine hidrocystoma	Periorbital	Tiny, blue cyst	Eccrine gland	
Syringoma	Periorbital	Flesh-colored papule	Eccrine duct	
Eccrine poroma	Hands, feet	Red nodule, may ulcerate	Eccrine duct	
Eccrine spiradenoma	Legs/others	Firm dermal nodule	Eccrine duct	Often tender

Table 1 Benign tumors of epidermal appendages

Further reading

Betti R, Bruscagin C, Inselvini E, Palvarini M, Crosti C. Successful cryotherapic treatment and overview of multiple clear cell acanthomas. *Dermatologic Surgery* 1995; **21**: 342–4. [Successful

treatment of multiple clear cell acanthomas is presented in this case report.]

Rampen FHJ, Schwengle LEM. The sign of Leser-Trelat: Does it exist? *Journal of the American Academy of Dermatology* 1989; **21**: 50–5. [The authors discuss the possible relationship between eruptive seborrheic keratoses and internal malignancy.]

Salasche SJ, McCollough ML, Angellon VL, Grabski WJ. Frontalis-associated lipoma of the forehead. *Journal of the American Academy of Dermatology* 1989; **20**: 462–8. [This deep fatty tumor is described and its surgical treatment is illustrated with photographs and diagrams.]

Schwartz RA. Sign of Leser-Trelat. *Journal of the American Academy of Dermatology* 1996; **35**: 88–95. [The author presents criteria for defining the sign and supports its paraneoplastic significance.]

38.3 Benign acquired pigmented lesions of the skin

Arthur J. Sober and Richard G. B. Langley

Distinguishing benign pigmented lesions from melanoma
Benign acquired nevi as markers of increased melanoma risk
Spitz nevi (spindle and epithelioid cell nevi)
Neurocutaneous syndromes
Further reading

Benign acquired pigmented lesions are important for three reasons. First, these lesions must be distinguished from cutaneous melanoma, which occasionally may be a diagnostic problem (Table 1). Second, some benign acquired pigmented lesions are potential precursors (e.g. dysplastic nevi) for melanoma and are markers of increased risk for cutaneous melanoma (Fig. 1). People with greatly increased numbers of benign acquired nevi (even non-dysplastic) are also at increased risk of melanoma developing. Third, the presence of lentigines and café-au-lait macules may indicate the presence of certain neurocutaneous syndromes, some of which have surgical implications (see Table 3).

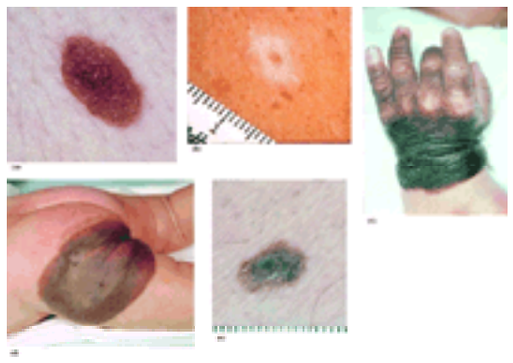


Fig. 1. (a) Compound nevus. (b) Halo nevus.(c) Congenital melanocytic nevus (which may be a precursor to melanoma). (d) Giant congenital melanocytic nevus (which may be a precursor to melanoma). (e) Dysplastic nevus (which may be a precursor to or marker for melanoma).

Distinguishing benign pigmented lesions from melanoma

Most benign acquired pigmented lesions are not difficult to distinguish from cutaneous melanoma since they lack the irregularity of border, surface, and pigment pattern found in radial growth phase melanoma. Lesions that may present difficulty include some dysplastic nevi, irritated seborrheic keratoses, traumatized hemangiomas, pyogenic granulomas, pigmented basal cell carcinomas, and blue nevi. The last, with its melanin pigment located in the mid-dermis, may resemble the nodular form of cutaneous melanoma. Table 1 shows the most characteristic features of common benign pigmented lesions.

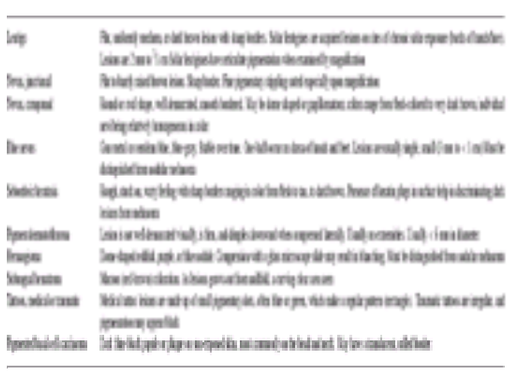


Table 1 Benign acquired pigmented lesions

Benign acquired nevi as markers of increased melanoma risk

Melanocytic nevi can be either precursors or markers of increased risk for melanoma. As precursors, it is recognized that approximately one-third to one-half of all melanomas arise within previously benign melanocytic nevi (usually dysplastic nevi). The development of a change in color or size in a previously stable pigmented skin lesion may be one of the earliest changes in the development of melanoma. Signs such as ulceration and bleeding are less common in early melanoma and more frequently occur in advanced disease.

Quantification of nevi has also been determined to correlate with the risk of developing melanoma. In separate studies by Swerdlow and by Holman, the presence of increased numbers of benign acquired nevi greatly increased the risk for cutaneous melanoma. Holman noted about an 11-fold increase in patients with the presence of 10 or more palpable nevi on the arms. Swerdlow reported a risk of more than 60-fold when 50 or more nevi of diameter 2 mm or more are present.

Clinically atypical nevi or dysplastic nevi (also called Clark's nevi, B-K mole, atypical nevus, FAMMM mole) are recognized markers for cutaneous melanoma, and are potential precursor lesions. The dysplastic nevus was first described in patients with familial melanoma by Clark and named B-K moles after the index families studied, and by Lynch who termed this the familial atypical multiple mole melanoma (FAMMM) syndrome. In melanoma-prone families, the risk for cutaneous melanoma appears to be transmitted as a dominant trait. Dysplastic nevi (one to many) are present in about half of the family members, and in nearly all of those who develop melanoma. Similar findings have been reported in Caucasian families from Europe and Australia. There is also an increased frequency of multiple primary melanomas in these families. In some patients the melanoma clinically appears to arise in association with a dysplastic nevus (and may also be seen in histopathologic contiguity). In others the presence of the dysplastic nevus appears to serve as a marker of increased risk. The lifetime risk of melanoma in an individual with dysplastic nevi from a family with two affected members may exceed 50 per cent.

Both the number and type of nevi have been identified as elevating the risk of melanoma in a recent case–control study by Tucker *et al.* In this study 716 patients with recently diagnosed melanoma were compared with 1014 controls and examined for presence, type, and size of melanocytic nevi. The presence of dysplastic nevi conferred an elevated risk of melanoma and the degree of risk was associated with the number of such nevi; a solitary lesion doubled the risk of melanoma whereas 10 or more dysplastic nevi resulted in a 12-fold elevation of risk. Elevated risk was also seen with the number of small and large non-dysplastic nevi. Patients with 50 to 99 small nevi or more than 10 large non-dysplastic nevi had twice the risk of developing melanoma compared with controls.

The short arm of chromosome 1p has been linked in some American families to the familial syndrome. Other researchers have failed to reproduce linkage to chromosome 1p in subsequent studies. More recently, genetic linkage studies have localized a familial melanoma gene on chromosome 9p21. Positional cloning studies by Nabori and Kamb later identified a tumor suppressor gene, p16 $^{INK4\ (CDKN2)}$, on chromosome 9p21 and identified this as the possible candidate for melanoma suceptibility.

Table 2 lists the features that distinguish dysplastic nevi from ordinary benign acquired nevi. No one feature is unique. Some highly irregular dysplastic nevi may

resemble cutaneous melanoma, in which case histopathologic examination may be necessary to distinguish the two.

Clinical feature	*Dysplastic nevi*	*Benign acquired nevi*
Color	Variable mixture of tan brown, black, or red/pink within a single nevus; nevi may look very different from each other	Uniformly tan or brown
Shape	Irregular borders; pigment may fade off into surrounding skin; macular portion at the edge of the nevus	Round; sharp, clear-cut borders between the nevus and the surrounding skin; may be flat or elevated
Size	Usually more than 5 mm, may be more than 10 mm; occasionally smaller than 5 mm	Usually less than 6mm in diameter
Number	Often very many (more than 100), but occasionally may be only one	In a typical adult: 10 to 40 nevi scattered over the body; perhaps 15 per cent of patients have no nevi
Location	Sun-exposed areas; the back is the most common site, but dysplastic nevi may also be seen on the scalp, the breasts, and buttocks	Generally on the sun-exposed surfaces on the skin above the waist; the scalp, breasts, and buttocks are rarely involved

Table 2 Clinical features distinguishing dysplastic nevi from benign acquired nevi

Sporadic forms of dysplastic nevi affect about 5 to 10 per cent of the Caucasian population of the United States of America. Whether all these individuals have an increased risk for melanoma is not known. Kraemer has estimated a 6 per cent lifetime risk in individuals who have one or more dysplastic nevi and lack a family history of either dysplastic nevi or cutaneous melanoma. In a study of 452 patients (mixed familial and sporadic) with dysplastic nevi, followed for a mean of 27 months by Rigel *et al.*, 18 melanomas were diagnosed (12 *in situ*, 6 invasive). Since this figure is so much higher than the estimated 1 per cent lifetime risk, monitoring of patients with dysplastic nevi seems warranted.

Spitz nevi (spindle and epithelioid cell nevi)

The Spitz nevus, also termed benign juvenile melanoma, described by Sophie Spitz in 1948, deserves special comment since the diagnosis and appropriate management of these lesions may be difficult. Also termed spindle and epithelioid cell nevi, from the histopathologic appearance, the Spitz nevus does not only occur in children and adolescents, but also in adults. The typical history is the sudden appearance of a flesh-colored to pink or red dome-shaped papule or nodule (although in individuals with a dark complexion the lesion may be deeply pigmented). Mitoses and cells with an atypical appearance may be present, and the pathologist may be tempted to regard these as melanoma.

However, since melanoma is rare in children under the age of 10 years (< $1/10^6$ per year), all lesions diagnosed as melanoma not arising in congenital nevi in this age group should be reviewed by the pathologist with the possibility of Spitz nevus in mind. A second opinion from a melanoma reference pathologist may be of value if there is any concern about diagnosis. Diagnostic biopsy should ideally be excisional and should extend to the full depth of the lesion, since the feature of maturation with depth is helpful in the evaluation. Failure to mature with depth may indicate melanoma. Any lesions that the pathologist regards as equivocal, including such adjectives as ‘atypical', should probably be treated with complete excision to prevent local recurrence. These patients should undergo regular cutaneous examination to exclude the unlikely possibility that the lesion was actually a melanoma masquerading as a Spitz nevus. Until more definitive techniques are available to separate Spitz nevi from melanoma, both overdiagnosis and underdiagnosis are likely to occur. Local recurrence has been reported to be infrequent in incompletely removed Spitz nevi that have ‘classical' histopathologic features.

Neurocutaneous syndromes

The presence of benign lentigines and café-au-lait macules are helpful in the recognition of certain neurocutaneous syndromes (Table 3). Several of these syndromes may present with signs and symptoms requiring surgical therapy for alleviation. An example would be the development of acoustic neuroma in patients with neurofibromatosis or gastrointestinal polyps in Peutz–Jegher's syndrome. Table 4 presents the clinical features of neurofibromatosis.

Syndrome	*Features*	*Inheritance**	*Cutaneous findings*	*Associated features*
Multiple lentigines syndrome	LEOPARD syndrome	AD	Multiple, hyperpigmented macules, especially on the face Mucosal surfaces spared	1—lentigines, 2—ECG abnormalities 3—ocular hypertelorism 4—pulmonary stenosis 5—abnormal genitalia 6—retardation of growth 7—deafness (sensorineural)
LAMB and NAME syndromes	Syndromes are same	?	Multiple lentigines–ephelides, blue nevi, mucocutaneous myxomas	1—lentigines, 2—atrial myxoma 3—mucocutaneous myxomas 4—blue nevi 5—nevi 6—atrial myxoma 7—mucocutaneous myxomas 8—ephelides
Peutz–Jegher's syndrome	Perioral mucocutaneous pigmentation	AD (?)50	Brown-colored macules Characteristically around the mouth, on lips and buccal mucosa. May also occur on the hands and feet	Gastrointestinal polyps, especially in jejunum Increased risk of gastrointestinal and some genitourinary malignancies
Cronkhite–Canada syndrome	-	?	Brown macules, commonly on face and extremities. Alopecia and dystrophic nail changes.	Multiple gastrointestinal polyps

* AD, autosomal dominant; ER, spontaneous remission.

Table 3 Syndrome characterized by lentiginosis

Pigmentary
Axillary freckling
Café-au-lait spots
General hyperpigmentation
Hypopigmented macules (rare)
Large pigmented macules overlying plexiform neurofibroma
Lisch nodules (iris)
Non-pigmentary
Angiomas
Atrophic hypoplastic macules (rare)
Blue-red macules
Juvenile xanthogranulomas
Neurofibroma
Plexiform neurofibroma
Pruritus
Segmental hypertrophy

Table 4 Cutaneous manifestations of neurofibromatosis

Further reading

Clark WH Jr, Reimer RR, Greene M, Ainsworth AM, Mastrangelo MJ. Origin of familial malignant melanomas from heritable melanocytic lesions. ‘The B-K mole syndrome'. *Archives of Dermatology* 1978; **114**: 732–8.

Holman CDJ, Armstrong BK. Pigmentary traits, ethnic origin, benign nevi and family history as risk factors for cutaneous malignant melanoma. *Journal of the National Cancer Instute* 1984; **72**: 257–66.

Kamb A *et al.* A cell cycle regulator potentially involved in genesis of many tumor types. *Science* 1994; **264**: 436–40.

Kaye VN, Dehner LP. Spindle and epithelioid cell nevus (Spitz nevus). Natural history following biopsy. *Archives of Dermatology* 1990; **126**: 1581–3.

Kraemer KH *et al.* Dysplastic naevi and cutaneous melanoma risk. *Lancet* 1983; **ii**: 1076.

Langley RGB, Fitzpatrick TBF, Sober AJ. Clinical characteristics. In: Balch CM, Houghton AN, Sober AJ, Soong SJ, eds, *Cutaneous melanoma*, 3rd edn, pp. 81–101. Quality Medical Publishing, St Louis, 1998.

Langley RGB, Barnhill RL, Sober AJ, Mihm MC Jr, Fitzpatrick TBF. Cutaneous melanoma. In: Freedberg IM, Eisen AZ, Wolff K, Austen FK, Goldsmith LA, Katz SI, Fitzpatrick TB eds. *Fitzpatrick's dermatology in general medicine*, 5th edn. McGraw-Hill, New York, 1999: 1080–116.

Lynch HT, Fusaro RM, Kimberling WJ, Lynch JF, Danes FS. Familial atypical multiple mole melanoma (FAMMM) syndrome: segregation analysis. *Journal of Medical Genetics* 1983; **20**: 342–4.

Nobori T, Miura K, Wu DJ, Lois A, Takabayashi K, Carson DA. Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature* 1994; **368**: 753–6.

Rigel DS *et al.* Dysplastic nevi, markers for increased risk for melanoma. *Cancer* 1989; **63**: 386–9.

Sober AJ, Koh HK, Tran N-L, Washington CV Jr. Melanoma and other skin cancers. In: Fauci AS *et al.*, eds, *Harrison's principles of internal medicine*, 14th edn, pp. 543–9. McGraw-Hill, New York, 1998.

Sober AJ, Langley RGB. Current treatment of melanoma. In: Rakel RE, ed., *Conn's current therapy, 50th anniversary edition*, pp. 809–12. WB Saunders, Philadelphia, 1998.

Spitz S. Melanomas of childhood. *American Journal of Pathology* 1948; **3**: 591–609.

Swerdlow AJ *et al.* Benign melanocyte nevi as a risk factor for malignant melanomas. *British Medical Journal* 1986; **292**: 1555–9.

Tucker MA *et al.* Clinically recognized dysplastic nevi. A central risk factor for cutaneous melanoma. *Journal of the American Medical Association* 1997; **277**: 1439–44.

38.4 Premalignant and malignant tumours of the skin

Calvin O. McCall and Thomas G. Cropley

Epidermal tumors
Actinic keratoses
Cutaneous horns
Basal cell carcinoma
Squamous cell carcinoma
Keratoacanthoma
Dermal tumors
Dermatofibrosarcoma protuberans
Malignant fibrous histiocytoma
Atypical fibroxanthoma
Juvenile xanthogranuloma
Cutaneous metastases
Further reading

Epidermal tumors

Actinic keratoses

Actinic keratoses, also known as solar keratoses, are common premalignant tumors of the epidermis, which occur in sun-exposed skin (Fig. 1). Lesions are typically multiple, and may be so numerous as to alter most of the sun-exposed portion of the cutaneous surface. Most commonly, actinic keratoses are poorly defined papules with a surface scale, which may be more easily felt than seen. Their color ranges from pink to dark brown. Hypertrophic actinic keratoses are less common, and are fleshier papules or nodules surmounted by a thick scale or horn-like excrescence. Actinic keratoses are most common in fair-skinned people and are unquestionably the result of chronic ultraviolet exposure. Similar lesions may develop as a result of exposure to other tumor inducers, particularly ionizing radiation and arsenic.



Fig. 1. Actinic keratoses.

Histologically, actinic keratoses represent the earliest type of neoplastic transformation of the epidermis. Anaplasia and disorderly proliferation of keratinocytes are seen in the lower epidermis, but, unlike squamous cell carcinoma *in situ*, full-thickness atypia is not seen. An identical process involving the vermilion border of the lower lip is termed actinic cheilitis.

The reported malignant transformation rates vary greatly for actinic keratoses, ranging from less than 1 per cent to over 10 per cent. How-ever, recent studies support the association of these lesions with squamous cell carcinoma. Studies also support the value of early therapy. Because they are very superficial, a variety of modalities can be used to treat actinic keratoses. With the possible exception of large hypertrophic lesions, full-thickness excision is not indicated. In the United States, discrete lesions present in low numbers are most commonly treated with liquid nitrogen cryosurgery (see Section 32.7). Small numbers of lesions may also be treated with curettage. When more diffuse actinic damage is present, topical chemotherapy with 5-fluorouracil cream for 2 to 4 weeks is effective. Patients should be warned that widespread inflammation and crusting will develop. The use of topical tretinoin for the treatment of actinic damage has been studied, but its role in the treatment of actinic keratosis is poorly defined. Regardless of the treatment chosen, the patient's skin should be protected from further ultraviolet damage by chemical or opaque sunscreens and physical protection such as hats and long-sleeved shirts. Patient education regarding sun protection is essential.

Cutaneous horns

Cutaneous horns are elongated, keratinous projections from the skin, ranging in size from a few millimeters to many centimeters (Fig. 2). Because they are suggestive of underlying anaplasia, biopsy examination is warranted. Most cutaneous horns arise from actinic keratoses but they may also result from seborrheic keratoses, warts, keratoacanthomas, squamous cell carcinomas, and basal cell carcinomas.



Fig. 2. Cutaneous horn. The base of the horn must be checked for the presence of squamous cell carcinoma.

Basal cell carcinoma

Basal cell carcinoma is the most common form of skin cancer in Caucasian patients. It is rare in Asian and black populations, except in those with predisposing syndromes such as xeroderma pigmentosa or the nevoid basal cell carcinoma syndrome. Basal cell carcinoma is a disease of adults; affected children and infants have only rarely been described. In the vast majority of cases, excess exposure to short-wave ultraviolet light (UVB range, 290 to 320 nm) precipitates basal cell carcinoma. As a result, these tumors arise in sun-exposed sites, and have a particular predilection for the head and neck. Other risk factors for the development of basal cell carcinoma include ionizing radiation, burn scars, and chronic exposure to arsenic, all of which have also been implicated in the pathogenesis of squamous cell carcinoma. An autosomal dominantly inherited syndrome known as the nevoid basal cell carcinoma syndrome was described in the 1960s and the responsible gene has been identified. Affected patients develop innumerable basal cell carcinomas beginning in early childhood in association with odontogenic cysts of the mandible, other bony abnormalities, and medulloblastoma. A characteristic facies with prominent frontal bossing is usually seen. Basal cell carcinomas are also common in

patients with xeroderma pigmentosa, in association with multiple squamous cell carcinomas, multiple malignant melanomas, and other cutaneous tumors.

The origin of basal cell carcinoma is controversial. For many years the basal cell of the epidermis was presumed to be the originating cell type, but more recently it has been proposed that a pluripotential epithelial cell, perhaps originating in the pilosebaceous epithelium, is the true progenitor cell type. The cells of basal cell carcinoma are small and usually not cytologically atypical. In sections stained with hematoxylin and eosin these cells appear deep blue and are similar to normal epidermal basal cells. In addition to the proliferation of these basaloid tumor cells, stromal changes are invariably found within and peripheral to the tumor mass. In the nodular and superficial types of basal cell carcinoma the stromal change consists of an abundant deposition of mucopolysaccharides. Sclerosing basal cell carcinomas exhibit intense dermal fibrosis. The origin of the stromal changes is not known. Dermal fibrosis is presumably the principal reason that metastases from basal cell carcinoma are rare. In general, invasive growth of tumor nests follows the path of least resistance. Rather than invading bone or cartilage, the tumor tends to glide along the perichondrium or periosteum as it grows, sometimes to a distance many centimeters from the origin. Invasion along neurovascular bundles is uncommon but may be seen in the sclerosing type and other ‘aggressive' types of basal cell carcinoma. The three principal types of basal cell carcinoma, and several special subtypes, are discussed below.

Nodular basal cell carcinoma

Nodular basal cell carcinoma is the most common type of basal cell carcinoma (Fig. 3). The tumor consists of a single expansile nodule of tumor cells and their dermal stroma. The lesions are grossly dome-shaped or spheroidal, and to the naked eye appear shiny, with a translucent or gelatinous optical quality. Numerous small blood vessels are usually seen within the tumor, as a result of tumor-induced angiogenesis. They are typically flesh colored, although small, grayish-blue foci of pigmentation are often present deep within the substance of the tumor. In extreme cases, this pigmentation may be diffuse within the tumor. Such lesions are referred to as pigmented basal cell carcinomas (Fig. 4) and may be confused with nodular malignant melanoma.



Fig. 3. Nodular basal cell carcinoma.

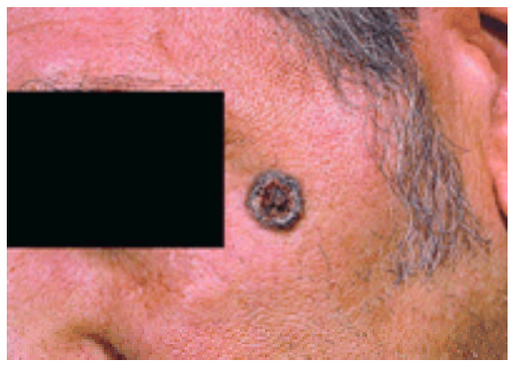


Fig. 4. Pigmented basal cell carcinoma.

Nodular basal cell carcinomas begin as small papules 1 to 3 mm in size. As the papules enlarge, their radial growth tends to be greater than their vertical growth, resulting in flattened hemispherical plaques that can grow to many centimeters in diameter. Tumors greater than 1 cm in diameter usually outgrow their blood supply, resulting in central ulceration.

Diagnosis is usually made by shave biopsy, but pigmented lesions are more appropriately completely excised if a diagnosis of malignant melanoma is considered. Tumors measuring less than 1 cm in diameter are rarely deeply invasive, and are frequently treated with electrodesiccation and curettage, cryosurgery, or excision. Regardless of the therapeutic modality used, treatment must extend beyond the margins of clinically apparent disease. For well-defined basal cell carcinomas less than 2 cm in diameter, evidence suggests a 4 mm margin of normal skin removed during treatment will be adequate in more than 95 per cent of cases. Excision is the treatment of choice for lesions greater than 1 cm in diameter, although those arising in ‘high-risk' areas and recurrent tumors are best managed with Mohs micrographic surgery. Certain areas are at higher risk for recurrence because invasive tumor nests may track along cartilage or bone for great distances, and because tumor may be sequestered at connective tissue interfaces. These sites include the nasal tip, the alar groove, the medial canthus (in which case orbital invasion may occur), the preauricular and postauricular sulci, and the external auditory canal. Radiation therapy is an acceptable alternative for treatment of ‘high risk' and recurrent lesions, but it is usually reserved for individuals unable to tolerate surgery. New treatment modalities are being investigated, but photodynamic therapy, intralesional interferon, and intralesional 5-fluorouracil have not yet been approved as standard therapy.

Superficial basal cell carcinoma

Superficial basal cell carcinoma is the second most common type of basal cell carcinoma. It is unusual on the head and neck, appearing primarily on the trunk and proximal extremities. Clinically, superficial basal cell carcinoma resembles squamous cell carcinoma *in situ*, presenting as a thin, scaly pink plaque with undulating or irregular margins. Frequently, a narrow, thread-like translucent rim, similar in appearance to nodular basal cell carcinoma, is seen at the periphery of the expanding lesion. This tumor primarily spreads by radial extension of small pseudopodia of tumor cells and their stroma. Ulceration and deep dermal invasion usually occur only after the plaque has reached several centimeters in diameter. Excision is often not the treatment of choice for superficial basal cell carcinoma: these lesions may be large, and the resultant surgical defect may be difficult to repair primarily. In addition, the microscopic extensions of tumor cells into the adjacent skin are clinically inapparent, so wide margins may be necessary to ensure removal. Cryosurgery, electrodesiccation and curettage, and topical chemotherapy with 5-fluorouracil offer cure rates similar to those achieved following excision (about 90 per cent), frequently with less morbidity. Recurrences are treated with Mohs micrographic surgery or excision with conventional histologic frozen-section control.

Sclerosing basal cell carcinoma

Sclerosing (morphea-like, fibrosing) basal cell carcinoma is the third most common type of basal cell carcinoma. Its distribution is the same as that of nodular basal cell carcinoma; most lesions are located on the head and neck. Histologically, sclerosing basal cell carcinoma comprises narrow cords of tumor cells encased in dense scar-like fibrosis. This resemblance to scar tissue is also apparent at the clinical level and the tumor usually lacks the gelatinous appearance of other types of basal cell carcinoma. Lesions are firm, indurated, and may be bound down to the adjacent tissues. The surface of the tumor is usually smooth and slightly atrophic, and it may be lighter in color than the surrounding skin. Tiny translucent papules typical of basal cell carcinoma may be seen at the edge of the scar-like plaque.

Sclerosing basal cell carcinomas behave more aggressively than nodular or superficial basal cell carcinomas. Deep invasion along neurovascular bundles, tissue planes, and into adjacent soft tissues is usual. Surgical extirpation of this tumor is fraught with difficulty, and recurrence is likely unless microscopic margin control is undertaken. The Mohs' technique is superior to conventional frozen sections, which do not allow complete examination of the surgical margins.

Although sclerosing basal cell carcinoma is the most commonly encountered ‘aggressive' form of this disease, a histologically distinct form of basal cell carcinoma with a similar propensity for deep, infiltrative spread has been identified in recent years. This basosquamous, or metatypical carcinoma, consists of cells typical of basal cell

carcinoma, as well as cells that exhibit squamous metaplasia. Like sclerosing basal cell carcinoma, these tumors are locally aggressive and are difficult to remove without microscopic margin control.

Squamous cell carcinoma

Squamous cell carcinoma is a malignancy of epidermal keratinocytes. After basal cell carcinoma, it is the most frequently encountered malignant neoplasm of the skin. In the Caucasian population, squamous cell carcinoma most commonly results from chronic actinic damage and the incidence increases with age. However, squamous cell carcinoma also develops at mucocutaneous interfaces (Fig. 5) as a result of tobacco use or human papillomavirus infection, at sites treated with ionizing radiation, in burn scars, and in chronic ulcers of the skin. Factors other than solar damage account for the majority of squamous cell carcinomas in darkly pigmented individuals. Although actinically induced squamous cell carcinoma rarely metastasizes, tumors induced by other factors, particularly those at mucocutaneous interfaces, are aggressive and metastasize frequently. The incidence of squamous cell carcinoma and the rate of metastasis are greater in immunosuppressed patients, especially following organ transplantation.

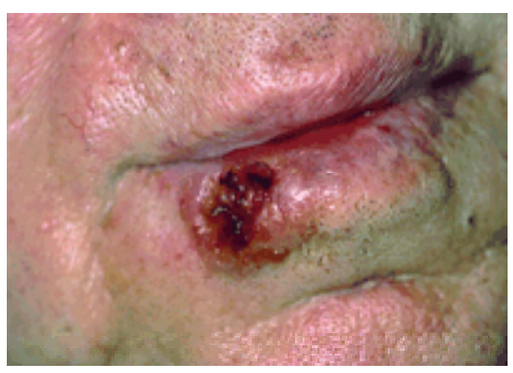


Fig. 5. Squamous cell carcinoma in a common site on the lower lip.

The clinical presentation of cutaneous squamous cell carcinoma is extraordinarily variable. As a rule it presents as an exophytic nodule or a plaque, which may ulcerate. These lesions are typically pink or flesh-colored in Caucasian patients; they may be hypopigmented or hyperpigmented in those with dark complexions. There is often evidence of aberrant keratinization, resulting in the formation of a scale or a cutaneous horn (see above). All of the modalities used for the treatment of basal cell carcinoma can be utilized for the treatment of squamous cell carcinoma. Complete excision, including Mohs micrographic surgery, is the treatment of choice for squamous cell carcinoma because of its metastatic potential.

Squamous cell carcinoma *in situ* is common. It may occur following chronic actinic damage, or in skin that is not exposed to the sun. In the latter case, the process is frequently referred to as Bowen's disease. Squamous cell carcinoma *in situ* arising from the genital skin is referred to as erythroplasia of Queyrat. Both actinically induced squamous cell carcinoma *in situ* and Bowen's disease (Fig. 6) present as scaly, well-defined plaques with irregular borders. In Caucasian patients, lesions are typically pink. Histologically, squamous cell carcinoma *in situ* does not invade the dermis, although invasive squamous cell carcinoma may develop within lesions. In the past, squamous cell carcinoma *in situ*, particularly in sun-protected skin, has been associated with internal malignancy of various types. However, a recent population-based study has cast doubt on this concept. Squamous cell carcinoma *in situ* may be treated by excision, although other modalities such as cryosurgery, electrodesiccation with curettage, and radiation are also effective. Topical chemotherapy with 5 per cent 5-fluorouracil is also occasionally used.

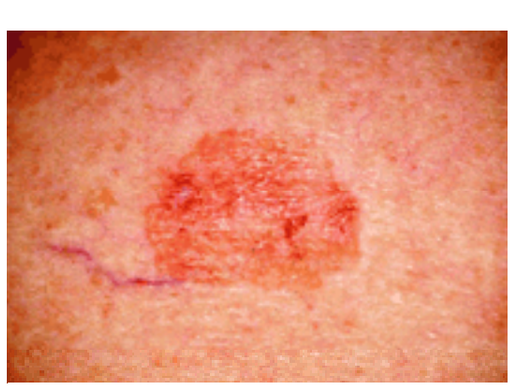


Fig. 6. Squamous cell carcinoma *in situ* (Bowen's disease).

Keratoacanthoma

Keratoacanthoma bears many similarities to squamous cell carcinoma, and many clinicians consider it to be a low-grade carcinoma, despite its typically self-limited nature. Classically, the tumor is domed or conical, and has a central crater filled with keratinaceous material (Fig. 7). The lesion is pink and lacks the gelatinous translucence and telangiectasia of basal cell carcinoma.Keratoacanthomas appear suddenly and often grow rapidly (several centimeters in a month). They range in size from several millimeters to several centimeters and are generally solitary. Multiple keratoacanthomas are rare and most frequently occur in patients with a family history of the condition.

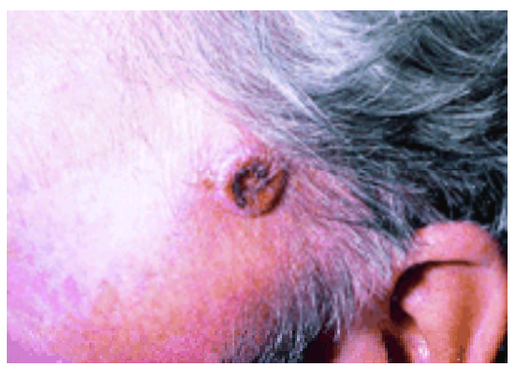


Fig. 7. Keratoacanthoma.

The pathology of keratoacanthoma resembles squamous cell carcinoma, and differentiation between the two entities may be difficult. Fully developed keratoacanthomas are expansile tumors composed of large, cytologically bland keratinocytes with ‘ground glass' cytoplasm. Unlike squamous cell carcinoma, the base of the tumor is flat and does not exhibit invasion of the deeper dermis. It is important to include the deep margin of the tumor in the biopsy specimen, as this may be the principal means by which the pathologist can make the correct diagnosis. The biopsy should also include the peripheral margin of the tumor.

Keratoacanthomas typically regress spontaneously within 1 year. Postinvolutional scarring is common; for that reason the lesions are frequently excised despite their seemingly benign nature. Large keratoacanthomas and those developing in immunosuppressed patients tend to be most aggressive. Their behavior is more like that of true squamous cell carcinoma, with locally destructive growth, the potential for invasion along neurovascular structures, and metastasis. Surgical excision is the

treatment of choice, but may not always be feasible, particularly when lesions are very large or multiple. Intralesional 5-fluorouracil or bleomycin is often used in such cases. Superficial radiation therapy (600 to 1000 cGy) is effective and safe for the treatment of most lesions. Recently, synthetic retinoids have been used in the management of patients with multiple keratoacanthomas.

Dermal tumors

Dermatofibrosarcoma protuberans

Dermatofibrosarcoma protuberans is a low-grade sarcoma of fibroblast origin, and may be thought of as the malignant equivalent of dermatofibroma. Ninety per cent of cases occur in adults, most often in the third and fourth decades of life. The lesions are firm intradermal plaques, which later become multinodular. Slow extension into the surrounding dermis and subjacent connective tissue is typical. Metastasis to regional lymph nodes and to visceral organs can occur after many years.

Dermatofibrosarcoma protuberans is difficult to remove surgically. Recurrence occurs in up to 50 per cent of cases, although recent series have suggested that an excision margin of 3 cm may reduce the recurrence rate to 20 per cent. Mohs' surgery may reduce the recurrence rate further, but considerable experience is necessary to identify the infiltrating strands of sarcoma cells on frozen sections. Regardless of the technique used, wide surgical excision with at least 3 cm margins is the treatment of choice.

Malignant fibrous histiocytoma

Malignant fibrous histiocytoma is an uncommon sarcoma, usually arising on the extremities. Although the tumor may arise within subcutaneous tissue or fascia, primary cutaneous malignant fibrous histiocytoma is occasionally encountered. Its presentation is similar to that of dermatofibrosarcoma protuberans (see above). The principal prognostic factor for malignant fibrous histiocytoma arising in the skin is whether or not the tumor has penetrated the underlying fascia. The incidence of metastases is approximately 10 per cent in cases without fascial involvement, and 30 per cent or more if fascia is invaded. The surgeon should be prepared to excise the cutaneous tumor, underlying fat, fascia, and a block of subjacent muscle.

Atypical fibroxanthoma

Atypical fibroxanthoma is a dermal tumor consisting of spindle and epitheloid cells. Debate exists over the nature of this tumor. Some feel it is benign, while others consider it a low-grade malignancy similar to malignant fibrous histiocytoma. It is generally found in sun-exposed regions of the skin in elderly patients, but it can occur in sun-protected areas in younger individuals. Special techniques should be used to establish a firm diagnosis, and complete excision is recommended because of a potential for metastasis.

Juvenile xanthogranuloma

Juvenile xanthogranuloma is a common cutaneous lesion encountered primarily in young children. Most lesions are solitary, but patients with multiple lesions have been described. Clinically, the tumor is a yellowish or yellowish-brown papule or nodule that arises rapidly. Histologically, it consists of a proliferation of histiocytoid cells with foamy cytoplasm. This benign lesion tends to regress spontaneously. Generally no treatment is needed, although excision is usually curative.

Cutaneous metastases

Cutaneous metastases occur in approximately 5 per cent of patients with internal malignancy and are the presenting sign in about 1 per cent. The incidence of cutaneous metastases tends to parallel the incidence of primary tumors. Carcinoma of the breast accounts for the vast majority of cutaneous metastases in women, and primary lung carcinoma is the most common source of cutaneous metastases in men. Table 1 lists the clinical characteristics of cutaneous metastatic carcinoma.

Breast	Anterior chest wall nodules; large indurated plaques in ipsilateral chest, scalp occasionally	Inflammatory carcinoma mimics cellulitis
Lung	Scalp; randomly scattered	30% adenocarcinoma 30% squamous cell 4% undifferentiated Key for presenting sign
Breast	Scalp; randomly scattered	
Colon	Abdominal, periumbilical skin	
Gastric/upper gastrointestinal tract	Umbilicus (Sister Joseph's nodule); abdominal wall; randomly scattered	
Rectal and rectal	Local skin	
Ovary	Periumbilical	
Melanoma	Randomly scattered; often in multiple crops	Blue-black, brown, or red nodules

Table 1 Clinical features of cutaneous tumor metastases.

Cutaneous metastases most commonly form solitary or multiple intradermal nodules, which are typically hard and indurated. On the scalp, which is a common site, an expansile mass of tumor cells may destroy hair follicles, resulting in alopecia. Pigmented metastases are most frequently seen in malignant melanoma. The so-called Sister Joseph's nodule is a solitary, firm umbilical nodule seen as a feature of metastatic carcinoma of gastrointestinal or pancreatic origin (Fig. 8). Several hematologic malignancies, including Hodgkin's disease, non-Hodgkin's lymphoma, and myelogenous leukemias may involve the skin. T-cell lymphomas and adult T-cell leukemia are particularly prone to cutaneous infiltration due to the epidermotropic nature of many malignant T-cell clones. Hematologic malignancies most commonly produce dermal plaques or nodules, which vary greatly in color, and may be solitary or multiple. Secondary changes such as purpura and ulceration are common. The differential diagnosis of infiltrated plaques or nodules in patients with leukemia or lymphoma often includes opportunistic infections due to bacteria, mycobacteria, or fungi. Diagnostic biopsy of representative lesions for histopathologic examination and culture is appropriate.

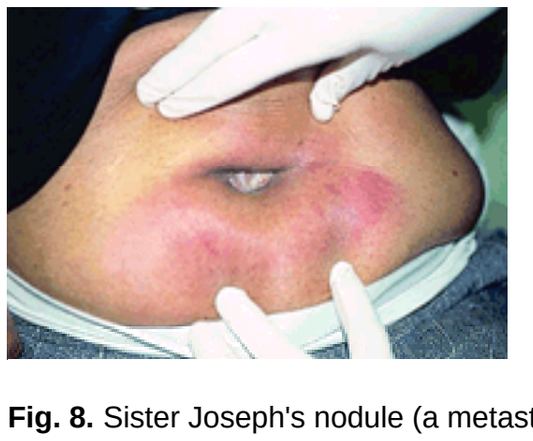


Fig. 8. Sister Joseph's nodule (a metastasis from a gastrointestinal carcinoma).

Further reading

Arbesman H, Ransohoff DF. Is Bowen's disease a predictor for the development of internal malignancy? *Journal of the American Medical Association* 1987; **257**: 516–18. [This critical review failed to demonstrate that Bowen's disease is a predictor for internal malignancy.]

Asada M *et al.* Solid basal cell epithelioma possibly originates from the outer root sheath of the hair follicle. *Acta Dermato-Venereologica* 1993; **73**: 286–92. [The authors present evidence that basal

cell carcinomas originate from cells of the follicular unit.]

Fritsch C, Goerz G, Ruzicka T. Photodynamic therapy in dermatology. *Archives of Dermatology* 1998; **134**: 207–14. [Recent developments in cutaneous photodynamic therapy are reviewed in this article.]

Goette DK. Topical chemotherapy with 5-fluorouracil. A review. *Journal of the American Academy of Dermatology* 1981; **4**: 633–49. [This is a detailed review of 5-fluorouracil and its uses in dermatology.]

Gorlin RJ. Nevoid basal cell carcinoma syndrome. *Medicine* 1987; **66**: 98–113. [The author presents an in-depth review of the syndrome and its various manifestations.]

Greenway HT, Cornell RC, Tanner DJ, Peets E, Bordin GM, Nagi C. *Journal of the American Academy of Dermatology* 1986; **15**: 437–43. [This report summarizes a pilot study showing interferon to be effective in treating basal cell carcinomas.]

Haydey RP, Reed ML, Dzubow LM, Shupack JL. Treatment of keratoacanthomas with oral 13-cis-retinoid acid. *New England Journal of Medicine* 1980; **303**: 560–2. [The authors discuss the treatment of multiple keratoacanthomas with surgery and an oral retinoid.]

Lookingbill DP, Spangler N, Sexton FM. Skin involvement as the presenting sign of internal carcinoma. A retrospective study of 7316 cancer patients. *Journal of the American Academy of Dermatology* 1990; **22**: 19–26. [This is a detailed review article examining cutaneous metastasis from internal malignancies.]

Miller BH *et al.* Nonsurgical treatment of basal cell carcinomas with intralesional 5-fluorouracil/epinephrine injectable gel. *Journal of the American Academy of Dermatology* 1997; **36**: 72–7. [Intralesional %-fluoroeracil was effective in treating basal cell carcinomas in this clinical trial.]

Mittelbronn MA, Mullins DL, Ramos-Caro FA, Flowers FP. Frequency of pre-existing actinic keratosis in cutaneous squamous cell carcinoma. *International Journal of Dermatology* 1998; **37**: 677–81. [This study supports the correlation between actinic keratoses and squamous cell carcinomas.]

Ratner D *et al.* Mohs micrographic surgery for the treatment of dermatofibrosarcoma protuberans. *Journal of the American Academy of Dermatology* 1997; **37**: 600–13. [Mohs micrographic surgery is presented as an effective treatment for this cutaneous malignancy.]

Schwartz RA. Keratoacanthoma. *Journal of the American Academy of Dermatology* 1994; **30**: 1–19. [The author reviews this tumor's features, behavior, and response to treatment.]

Weiss JS *et al.* Topical tretinoin improves photoaged skin. A double-blind vehicle-controlled study. *Journal of the American Medical Association* 1988; **259**: 527–32. [Cutaneous solar damage may be diminished by a topical retinoid according to this study.]

Wolf DJ, Zitelli JA. Surgical margins for basal cell carcinoma. *Archives of Dermatology* 1987; **123**: 340–4. [This study examined the margins required for reliable routine surgical treatment of basal cell carcinomas.]

38.5.1 Recognition, staging, and prognosis of cutaneous melanoma

Arthur J. Sober and Richard G. B. Langley

Introduction
Epidemiology

Recognition

Prognostic factors

Follow-up
Screening
Further reading

Risk profiles

Biopsy
Staging

Management

Introduction

Cutaneous melanoma is an increasingly frequent and potentially lethal malignancy that results from the malignant transformation of melanocytes in the epidermis, or occasionally the dermis. Early recognition followed by prompt surgical excision offers the best chance for cure for cutaneous melanoma. Primary tumor thickness remains the most critical prognostic indicator in this malignancy. While early diagnosis and surgical excision of *in situ*, or early invasive melanomas, is curative in most patients, the efficacy of treatment of advanced melanoma remains limited, and the prognosis of metastatic disease remains guarded. In the 1940s melanoma was invariably detected in its late stages and patients died rapidly from metastatic disease. As the early signs of the disease have been widely published, and educational campaigns have been promoted to the public, an increased awareness of this malignancy has resulted. Improvements in the overall 5-year survival have been documented as the rate in the 1940s was in the order of 40 per cent and current rates are 88 per cent. It is believed that such improvements in the overall 5-year melanoma survival rate are the result of improved detection.

Epidemiology

The American Cancer Society has estimated that in 1999 44 200 Americans will develop invasive melanoma and 7300 will die of this disease. The frequency in children is less than one case per million per year, but after the age of 15 it is about 100 times more frequent. The incidence in black and east Asian people, as well as in dark-skinned Caucasians, is one-tenth to one-twentieth that of light-skinned Caucasians.

The incidence of melanoma has increased over the past two decades more rapidly than that of any other tumor. This increased frequency has been seen worldwide and is approximately 3 to 8 per cent per year. Although the cause of this increase is unknown, the short period over which this change has occurred suggests that some environmental or behavioural change is responsible. The increases are especially noteworthy in adolescents and young adults. In men aged 15 to 35, melanoma is the second most common cause of death from cancer in the United States; in New Zealand it accounts for 23 per cent of all cancers diagnosed in individuals of this age group. The most frequent site of cutaneous melanoma is the upper back in men and the lower leg from knee to ankle, as well as the upper back, in women. In black, east Asian, and dark-complexioned Caucasian subjects, the most common sites are palms, soles, nailbeds, and mucous membranes. Several lines of evidence indirectly link solar exposure to the increasing incidence of melanoma (Table 1). The incidence of melanoma varies inversely with latitude within countries, and the disease is more common in individuals of British extraction in Australia than in the United Kingdom. In addition, melanoma tends to occur most frequently on sun-exposed sites, particularly those exposed to intermittent solar radiation. The higher rates seen in light-complexioned populations also support the role of solar radiation.



Table 1 Evidence for and against a role for sunlight in the etiology of cutaneous malignant melanoma

Risk profiles

Table 2 ranks the risk profiles associated with the development of melanoma: the highest risk is associated with a persistently changing mole. Increase in size and change in color are the two most common characteristics of early cutaneous melanoma. The increased risk associated with increasing numbers of nevi and with sporadic and familial dysplastic nevi is reviewed elsewhere (see Chapter 38.3). Factors associated with only a small increase in relative risk may be important if they affect a substantial proportion of the population. MacKie has proposed an algorithm that can be used in screening to distinguish between high- and low-risk people. Those who freckle, have three or more dysplastic nevi, three or more severe sunburns, or 20 or more nevi have an increased risk of about 600-fold (men) or 200-fold (women) over people with none of these factors.

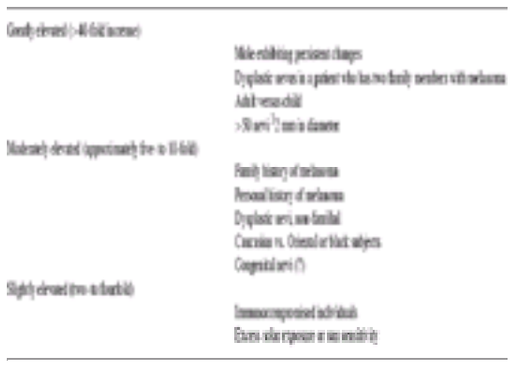


Table 2 Risk factors for cutaneous melanoma

Recognition

Most melanomas can be distinguished from non-malignant cutaneous lesions (Fig. 1). Table 3 lists those factors in the history and physical examination which are of diagnostic help, while Table 4 lists features distinguishing the four types of cutaneous melanoma. Approximately 5 per cent of melanomas may be clinically amelanotic and in these cases the diagnosis can only be made by the pathologist. Even with the best clinical skills, not all melanomas can be diagnosed clinically. A biopsy

specimen should be obtained when melanoma is a diagnostic possibility.

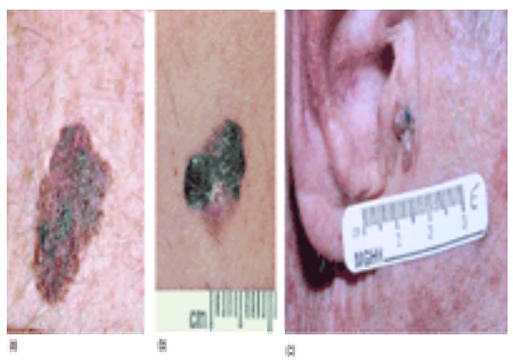


Fig. 1. (a) Early superficial spreading melanoma, level II. (b) Superficial spreading melanoma. Vertical and radial growth phases are present. (c) Superficial spreading melanoma (note amelanotic areas).

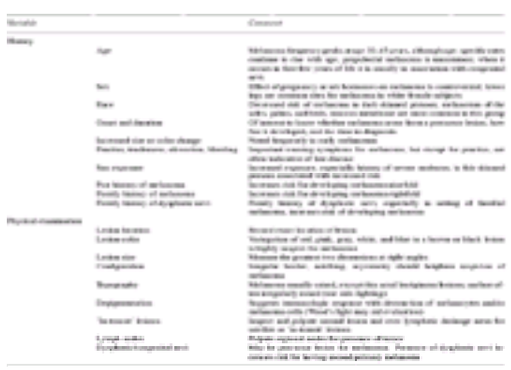


Table 3 Guide to clinical evaluation of a suspected melanoma

Type	Site	Average age of diagnosis (years)	Duration of disease (months to years)	Color
Lentigo maligna melanoma	Sun-exposed surfaces, particularly malar region of cheek and temple	70	5–20 or larger	In the portion, shades of brown and tan predominate, but variegated gray occasionally present; in nodules, shades of reddish brown, bluish gray, bluish black
Superficial spreading melanoma	Any site (most common site: upper back and in women, lower legs	40–50	1–7	Shades of brown mixed with black; red, white, blue, and/or shades of reddish brown and white may exist, and the border of lesion is at least in part visibly irregularly pigmented
Nodular melanoma	Any site	40–50	Months to 5 years	Reddish blue (papule or bluish black; otherwise same as color as normal with thickness of that
Acral lentiginous melanoma	Palms, soles, and/or subungual, mucous membrane	60	1–10	In the precious, shade brown predominantly, as mixed lesions (plaques) brown-black or blue-black predominantly

[a] Activity from the precursor stage (benign melanosis, inactive) contributes to the epidermis.

Table 4 Clinical features of malignant melanoma

Biopsy

Total excisional biopsy with narrow margins is the procedure of choice for the diagnosis of cutaneous melanoma: this provides the entire specimen for histopathologic examination and also provides definitive therapy for benign lesions. Removal of a partial biopsy specimen appears to be safe and is a reasonable course of action when total excisional biopsy is not easily accomplished. Either punch (trephine) or incisional biopsy can be used. Shave biopsies are not appropriate if a melanoma is suspected, as this procedure disrupts the tumor and precludes accurate microstaging. Final determination of melanoma type and thickness (important for prognosis) must await complete excision, however. When partial biopsies are performed, biopsy of the most raised portion will usually be sufficient for diagnosis. In flat or plaque-type lesions biopsy of the darkest portion most frequently produces a representative specimen.

Staging

It is important to accurately stage melanoma patients for appropriate management, and to determine prognosis. The traditional three-stage classification system has been supplanted by the American Joint Commission on Cancer four-stage system in most academic centers and in clinical trial investigation. This system is listed in Table 5.



Table 5 The American Joint Cancer Commission classification for staging melanoma of the skin

Prognostic factors

The thickness of the primary tumor in millimetres (Breslow scale) is the single most important prognostic factor for melanoma patients with primary cutaneous melanoma. There is a reasonably direct correlation between the primary tumor thickness (measured vertically from the granular cell layer down to the deepest tumor cell) and survival at 5, 8, or 10 years. There is also an inverse relationship between thickness of the primary tumor and time of death (Table 6). A small number of patients with very thin tumors develop metastatic disease; this may occur 10 years or more after removal of the primary tumor. As melanoma may have a prolonged natural history, comparison of results at 10 years after diagnosis, is more consistent than comparisons made after only 5 years.

Risk category	Melanoma thickness (mm)	5-year survival[a] (%)
Minimum	<0.76	96–99
Low	0.76–1.50	87–94
Intermediate	1.51–4.0	66–77
High	>4.0	<50

[a] Crude survival rates.

Table 6 Prognosis of cutaneous melanoma based on primary tumor thickness

Clark's levels of invasion, which historically preceded thickness measurement, also correlate well with outcome. In the Clark system, level II tumors show penetration of the papillary dermis, in level III the papillary dermis is involved, level IV enters the reticular dermis, and level V enters the subcutaneous fat. Using this system it is sometimes difficult to distinguish level III from level IV. In addition, the designation of polypoid tumors of substantial thickness as level III underestimates the risk associated with these lesions.

Several centers have evaluated other factors that add to the ability to predict outcome (Table 7). Almost all studies have shown that women with localized disease survive longer than men: the location of tumors on favorable sites such as the lower extremity may explain this difference, although some studies suggest that location alone is not sufficient to explain this difference in survival. Increased numbers of mitoses, the histologic presence of an ulcer, and the presence of microscopic satellites, defined as discrete nests of tumor cells at least 0.05 mm in diameter separated from the main body of the tumor and located in the reticular dermis or subcutaneous fat, are all associated with a worse prognosis. Tumors in certain sites such as scalp, hands and feet, and mucous membranes also appear to have worse prognoses. Recently, Clark *et al.* have proposed a six-factor model to separate low-, intermediate-, and high-risk patients that have downward penetration of the tumor (vertical growth phase). The six factors are tumor thickness of 1.70 mm or more, sex, axial plus acral locations versus non-acral extremities, regression, tumor infiltration by lymphocytes, and mitotic rate. By objectively evaluating prognostic factors, appropriate surgical programs can be established and the efficacy of experimental adjuvant therapies can be assessed.



Table 7 Prognostic variables for clinical stage I/II melanoma

Management

Surgical margins of the primary tumor

In the past, considerable controversy has surrounded the choice of resection margins in the management of primary cutaneous melanoma. Traditionally, a wide local excision involving a margin of 5 cm was considered the standard of care. This recommendation of a 5 cm margin has been attributed to W. Sampson Handley who in the 1907 Hunterian lectures based this recommendation on a single autopsy examination of a metastatic deposit in a 34-year-old patient. With the institution of the Breslow thickness scale, low-risk patients could be identified for whom narrower surgical margins appeared to be safe. In 1977, Breslow and Macht noted that the width of the surgical excision margin did not appear to be an important factor in outcome for melanomas less than 0.75 mm thick. Recent prospective, randomized, controlled, surgical trials have provided evidence to support the safety and efficacy of narrower surgical margins for thin and intermediate thickness melanomas. The World Health Organization study which randomized 612 patients, controlled for major prognostic criteria, with melanomas less than 2 mm in thickness to undergo excision with 1 or 3 cm margins. No significant difference in disease-free survival and overall survival rates between the groups in the initial publication at a mean follow-up of 55 months was reported; nor was a significant difference reported in a published update of this trial with a mean follow-up of 90 months. The actuarial survival of the patients undergoing a narrow excision margin was 89.6 per cent and was not statistically different from the wide excision group of 90.3 per cent (p = 0.64). The local recurrence rate was 3 per cent occurring between 5 months and 7 years after the primary excision, and all were in the narrow excision group and in patients with a primary melanoma between 1.1-mm and 2.0-mm thick. Based on these findings the WHO study comparing 1-cm versus 3-cm margins has, for melanomas less than 1.0-mm thick, confirmed the efficacy and safety of 1.0-cm margins. Because of the trend to local recurrences in the 1.1 to 2.0-mm group it was recommended by these authors that longer follow-up is needed in this group before issuing a final recommendation.

The second important trial to support narrower surgical margins in the treatment of primary cutaneous melanoma was the Intergroup Melanoma Surgery Trial which prospectively enrolled 486 patients to undergo 2-cm versus 4-cm margins for melanomas between 1-mm and 4-mm thick. No statistically significant difference was noted in local recurrence or survival between the study groups in the initial report with a median follow-up of 72 months, or in the recent update with a median follow-up of 91 months. The 5-year disease-free survival rate was 79 per cent in the narrow excision group, and 81 per cent for the wider excision group. In addition, the local recurrence rate was not significantly different (p = 0.72) between the 2-cm and 4-cm groups constituting 2.1 per cent and 2.6 per cent of the patients, respectively. This trial confirmed the effectiveness and safety of a 2-cm margin for intermediate thickness melanomas.

In summary, there is now strong evidence that 1-cm margins are safe and efficacious for melanomas less than 1 mm in thickness. For melanomas greater than 1 mm in thickness there remains controversy regarding optimal excision margins. Based on the trend to increased local recurrences in the 1.1-mm to 2.0-mm group from the WHO study, and the lack of benefit to any wider excision as determined in the Intergroup study, some centers recommend 2.0-cm margins for intermediate thickness melanomas.

The optimal excision margin for *in situ* or thick melanomas(>4 mm) has not been rigorously studied to date in a controlled, randomized surgical trial. The most recent National Institutes of Health consensus conference in 1992 recommended 0.5-cm margins for melanomas *in situ.* For thick primary melanomas (>4 mm), many surgeons prefer to take margins greater than 2.0 cm, although there is a lack of evidence to support or refute this practice (Table 8).

Tumor thickness (mm)	Margins of surgical excision (cm)	Closure	Sentinel node studies
<1.0	1	Primary	No
1.0–4.0	1–2	Primary, graft, or flap	+
>4.0	2–3	Primary, graft, or flap	±

[a] Based on current practices at Massachusetts General Hospital. The management of each patient is individualized. It may not be possible to achieve the recommended margins of excision in certain anatomic sites, e.g., on the face or distal extremities. Also, skin grafts or flaps may be necessary for surgical defects in the latter locations.

Table 8 Guidelines for surgical management of patients with clinical stage I/II melanoma

While well designed and conducted randomized, controlled surgical trials have provided important information to guide clinical practice, strict guidelines should not be routinely recommended in the surgical management of cutaneous primary melanoma, and each case should be evaluated individually. Melanomas near a vital structure may require a reduction in the margin, whereas poor histologic prognostic factors may suggest a biologically more aggressive melanoma and support a wider margin. Melanoma excision at special sites, such as the fingers, toes, sole of the foot, and ear also require separate surgical and functional considerations. For further discussion of surgical management of melanoma at special sites see the monograph by Balch *et al.*

Elective lymph node dissection

The role of elective lymph node dissection (ELND) in patients with American Joint Commission on Cancer stage I/II melanoma has been one of the most controversial aspects in the surgical management of melanoma. The debate has focused primarily on the role of ELND in intermediate thickness melanomas, as thin melanomas rarely have nodal metastases making the value of ELND in this group low. Similarly, most patients with melanomas thicker than 4 mm fail from systemic metastases and ELND does not appear to improve their prognoses. The controversy is being clarified in light of recent randomized studies and the development of minimally invasive staging procedures. One of these studies, the Intergroup Melanoma study, involved a prospective randomized surgical trial of 740 American Joint Commission on Cancer stage I and II patients with intermediate thickness melanomas (1 to 4 mm), of the trunk or extremity. Patients were randomly selected to receive excision of the primary melanoma (with 2- or 4-cm margin) and to receive ELND or observation. With a median follow-up of 7.4 years, this trial failed to show a statistically significant overall 5-year survival benefit from ELND (86 per cent) compared with the observation group (82 per cent). A subset analysis did identify a survival benefit in patients that were less than 60 years of age that had ELND, and in those subgroups having ELND that had non-ulcerated lesions, and in melanomas between 1 and 2-mm thick. This trial has been criticized on methodologic grounds for conducting subgroup analyses, and these possible survival benefits of ELND in this trial have not been widely accepted.

Sentinel lymph node biopsy has been recently developed and is a minimally invasive staging procedure which may obviate the need for performing an ELND. This technique involves using lymphoscintigraphy to map precisely the nodal basin draining the primary melanoma, and thus identify the initial lymph node that drains from the primary tumor to determine if melanoma is present in that basin. One per cent Lymphazurin Blue dye is injected around the lesion to permit intraoperative localization of the first node draining the lymphatic basin, the sentinel node, that the primary melanoma drains to. This allows subsequent histopathologic examination to determine the presence or absence of tumor in the blue sentinel node. The sentinel lymph node biopsy facilitates identification of those patients who have subclinical micrometastases in a minimally invasive way and target these patients for node dissection. Selective lymph node biopsies therefore can identify only those with nodal involvement without subjecting the vast majority who lack metastases to the morbidity of ELND. Morton *et al.* in the original studies were able to identify a sentinel node in 81 per cent of 243 patients, and metastases were detected in 21 per cent of lymphadenectomy specimens. Complete lymphadenectomy was performed in all cases, and false-negative results were seen in only 1 per cent (two of 243) of patients with metastases were detected in non-sentinel nodes exclusively. More recently a somewhat higher false-negative rate has been reported.

To improve the intraoperative localization of the sentinel node a radioactive tracer, technetium-99, has been utilized with intraoperative localization with a hand-held gamma probe. Using this technique, 98 per cent (118 of 121) of sentinel nodes were identified in one study by Krag *et al.* with less tissue dissection required than with the conventional blue dye method. Other recent refinements of the sentinel node procedure have involved attempts by pathologists to improve the detection of micrometastases in the sentinel node after removal. Conventional routine histopathology and immunohistochemistry may yield false-negative results on pathologic examination of the sentinel node, prompting the use of reverse transcriptase assays for messenger RNA for the tyrosinase gene. Wang *et al.* were able to improve the detection of melanoma metastases in sentinel nodes using this technique; 66 per cent per cent (19 of 29) of sentinel nodes were found to have melanoma compared with 38 per cent (11 of 29) of patients undergoing sentinel lymph node biopsy with conventional histopathologic examination. The significance of ‘metastases' detected solely by polymerase chain reaction remains to be determined.

Adjuvant therapy

Until recently, there has been no standard therapy for patients with resected melanoma who are at high risk for recurrence (adjuvant therapy). The Eastern Co-operative Oncology Group (ECOG) 1689 trial has recently demonstrated improvement in median relapse-free survival and 5-year survival from interferon-a2b (20 MU/m^2/day, given intravenously, 5 days/week for 4 weeks; this was followed by maintenance therapy of 10 MU/m^2/day, given subcutaneously, three times a week for 48 weeks) in treated node-positive resected patients compared with controls. Ongoing trials with interferon in combination or comparison with other adjuvant agents (i.e. melanoma vaccines) and other adjuvant therapies should provide further information in this regard. The high expense and substantial morbidity associated with interferon-a use has not led to its universal use in all patients who had post-resection of nodal metastases.

Limb perfusion

In specialized centers perfusion of the affected extremity with phenylalanine mustard at elevated temperature has produced good control of in-transit lesions and satellite lesions. Although such perfusion is reasonably well accepted as a palliative measure, its benefit as an adjuvant in the treatment of primary extremity melanoma has not yet been established in prospective randomized studies. In a recent comparative investigation, melanoma patients matched for current prognostic factors showed a similar outcome after treatment with surgery alone or with perfusion. Perfusions with combinations of therapy including tumor necrosis factor has yielded a higher percentage of initial response but also a higher relapse rate by 6 months.

Distant disease

Therapeutic lymph node dissection appears to be warranted in patients with no significant disseminated disease; in the presence of disseminated disease, a more systemic approach is usually considered. The response to chemotherapy is greatest for disease in the skin, lymph nodes, and lung; tumors in bone, brain, and liver rarely respond. Long-term remission can occasionally be achieved by aggressive surgical removal of an isolated metastasis in the brain or other organ, but multiple metastases often appear within months. Table 9 shows possible approaches to the treatment of melanoma in various sites.

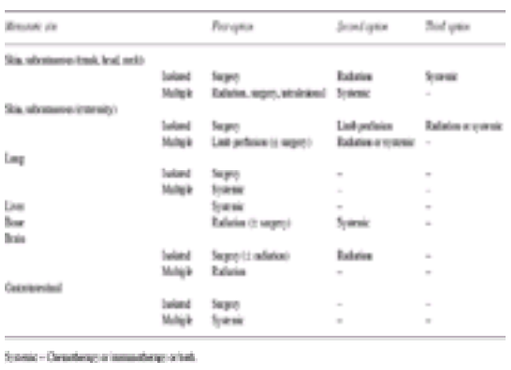


Table 9 General guidelines for treatment sequences in metastatic melanoma

Follow-up

Follow-up of melanoma patients is critical to enable detection of early recurrence and for the diagnosis of second or subsequent tumors at an early, curable, stage. Table 10 depicts the follow-up schedule currently in use in the Melanoma Center at the Massachusetts General Hospital.

Thickness (mm)	*Frequency of follow-up (months)*	*Duration of follow-up (years)*[a]
In situ	3	×1 visit then every 12 months
<1.5 mm	3	×2
	then every 6 months	×3 years,
	then every 12 months[b]	
1.5–4 mm	3	×2 years,
	then every 6 months	×3 years
	then every 12 months[b]	
>4 mm	every 3 months	×2 years
	then every 6 months	×3 years
	then every 12 months[b]	

[a]Ten-year follow-up for all melanoma patient. [b]If a patient has clinically atypical moles, the interval may continue every 6 months.

Table 10 Melanoma follow-up schedule at Massachusetts General Hospital Melanoma Center

Screening

The best hope for cure of cutaneous melanoma lies in early diagnosis and appropriate surgical removal. Public and professional education campaigns have resulted in a progressive fall in median melanoma thickness at diagnosis and an increase in 5-year survival. Awareness of risk factors and of the characteristics of early melanoma should result in continued improvement in survival.

Further reading

Balch CM, Urist MM, Karakousis CP, Barnhill R, *et al.* Efficacy of 2-cm surgical margins for intermediate-thickness melanomas (1 to 4 mm). Results of a multi-institutional randomized surgical trial. *Annals of Surgery* 1993; **218**: 262–7.

Balch CM *et al.*, eds. *Cutaneous melanoma*, 3rd edn. Quality Medical Publishers, St. Louis, 1998.

Beahrs OH, Henson DE, Hutter RVP, Myers MH, eds. *Manual for staging of cancer*, 3rd edn, pp. 139–44. American Joint Commission on Cancer. JB Lippincott, Philadelphia, 1988.

Breslow A. Tumor thickness, level of invasion, and node dissection in stage I cutaneous melanoma. *Annals of Surgery* 1975; **182**: 572–5.

Clark WH Jr, From L. Bernardino EA, Mihm MC, Jr. The histogenesis and biologic behavior of primary human malignant melanoma of the skin. *Cancer Research* 1969; **29**: 705–26.

Clark WH Jr, *et al.* Model predicting survival in stage I melanoma based upon tumor progression. *Journal of the National Cancer Institute* 1989; **81**: 1893–904.

Friedman R, Rigel D, Kopf A, Harris M, Baker D. *Cancer of the skin*. WB Saunders, Philadelphia, 1991.

Johnson TM, Smith JW 2nd, Nelson BR, Chang A. Current therapy for cutaneous melanoma. *Journal of the American Academy of Dermatology* 1995; **32**: 689–707.

Kirkwood JM, Strawderman MH, Ernstoff MS, Smith TJ, Borden EC, Blum RH. Interferon alfa-2b adjuvant therapy of high-risk resected cutaneous melanoma: the Eastern Cooperative Oncology Group Trial EST 1684. *Journal of Clinical Oncology* 1996; **14**: 7–17.

Koh HK, Kligler BE, Lew RA. Sunlight and cutaneous malignant melanoma: evidence for and against causation. *Photochemistry and Photobiology* 1990; **51**: 765–79.

Krag DN, Meijer SJ, Weaver DL, *et al.* Minimal access surgery for staging of malignant melanoma. *Archives of Surgery* 1995; **130**: 659–60.

Lienard D, Eggermont AM, Kroon BB, Schraffordt Koops H, Lejeune FJ. Isolated limb perfusion in primary and recurrent melanoma: indications and results. *Seminars in Surgical Oncology* 1998; **14**: 202–9.

Mackie RM, Freudenberger R, Aitchison TC. Personal risk factor chart for cutaneous melanoma. *Lancet* 1989; **ii**: 487–90.

Rhodes AR *et al.* Risk factors for cutaneous melanoma: a practical method of recognizing predisposed individuals. *Journal of the American Medical Association* 1987; **258**: 3146–54.

Veronesi U *et al.* Delayed regional lymph node dissection in stage I melanoma of the skin of the lower extremities. *Cancer* 1982; **49**: 2420–30.

Veronesi U *et al.* Thin stage I primary cutaneous malignant melanoma: comparison of excision with margins of 1 or 3 cm. *New England Journal of Medicine* 1988; **318**: 1159–62.

38.5.2 Surgical management of malignant melanoma

Grant W. Carlson

Epidemiology
Diagnosis
Pathology
Growth phases of melanoma
Histologic types of melanoma
Biopsy
Staging
Microstaging
TNM staging system
Other prognostic factors
Local treatment
Regional treatment
Elective lymph node dissection
Sentinel lymph node biopsy
Therapeutic lymph node dissection
Axillary lymph node dissection
Groin dissection
Head and neck dissection
Isolated limb perfusion
Surgical therapy of metastatic melanoma
Adjuvant therapy
Further reading

Epidemiology

The incidence of malignant melanoma is increasing faster than any other cancer and has tripled in the past four decades. Malignant melanoma now accounts for 3 per cent of all cancers. It is estimated that in the year 2000, 1 in 75 of the American population will develop the disease. It is estimated that 38 300 new cases of melanoma were diagnosed in the United States in 1996, resulting in 7300 deaths. Melanoma has the highest annual incidence rate of any cancer in white individuals between the ages of 25 and 29 and in white men of 35 to 39 years old.

Risk factors for the development of melanoma are summarized in Table 1. The risk of developing an additional primary melanoma in a patient who has had a melanoma is approximately 5 per cent. People with blonde or red hair, fair complexion, blue eyes, and a propensity to sunburn are associated with an increased risk of melanoma. Melanoma is primarily a disease of white people with individuals of Northern or Central European ancestry being at highest risk. Recreational exposure to sunlight, especially a history of blistering sunburn, has been associated with an increasing risk of melanoma. Radiation in the ultraviolet B range is the most critical element. The presence of a large number of benign moles has been associated with an increasing risk of melanoma. A family history of melanoma or skin cancer is associated with an increased risk of melanoma. Familial atypical multiple mole and melanoma syndrome (FAMMM) is used to describe families with two or more kindreds with melanoma and the presence of atypical moles.

Personal history of melanoma
Fair complexion
Race
Sunlight exposure
Benign nevi
Family history of melanoma

Table 1 Risk factors for developing malignant melanoma

Diagnosis

Malignant melanoma arises from a pre-existing nevus in approximately 60 per cent of cases. The clinical accuracy of diagnosis is approximately 50 per cent for non-dermatologists so the index of suspicion must be high. An ABCD rule has been developed that helps clinicians identify moles which warrant biopsy. A stands for lesion asymmetry, B for border irregularity, C for variegated color, and D for diameter over 6 mm. Other signs of melanoma include itching, bleeding, ulceration, changes in a pre-existing benign mole, or development of a new pigmented lesion.

Pathology

Growth phases of melanoma

The growth of melanoma has been characterized histologically as occurring in two distinct phases: a radial and vertical phase. A radial growth phase is characterized by melanoma tumor cells in the epidermis and papillary dermis, with development of a raised irregular surface on the skin. The risk of nodal metastasis is low in a purely radial growth phase.

Vertical growth into the deeper layers of the skin is associated with increasing nodularity of the lesion and increased propensity for metastasis. As discussed below, the depth of invasion correlates directly with prognosis.

Histologic types of melanoma

Melanomas are classified into four major categories based on clinical and histologic features: superficial spreading, nodular, lentigo maligna, and acral lentiginous melanomas. When matched for tumor thickness, no prognostic significance is found between the various types of melanoma.

Superficial spreading melanomas

These are the most common variety of melanoma, seen in about 70 per cent of cases. The lesion usually arises within a pre-existing nevus. There is often a history of a slowly evolving change in the precursor lesion over several years. The initial growth is above the basement membrane.

Nodular melanomas

These represent about 10 to 15 per cent of cutaneous melanomas. They are generally dark blue-black and are more symmetric and uniform than other melanomas. The ABCD rule does not apply to nodular melanomas. Nodular melanomas typically begin in uninvolved skin. They characteristically lack a radial growth phase and have a shorter clinical onset than superficial spreading melanomas.

Lentigo maligna melanomas

These account for 10 to 15 per cent of cutaneous melanomas. They typically occur on the sun-exposed areas of the head and neck and hands. Clinically, they are large, flat, tan lesions with areas of dark brown or black coloration that look like a stain on the skin. They are characterized by a prolonged radial growth phase. They do not have the same propensity to metastasize as melanomas of other growth patterns.

Acral lentiginous melanomas

These characteristically occur on the palms or soles, or beneath the nail beds. They make up 2 to 8 per cent of melanomas in Caucasian people as opposed to 40 to 60 per cent of melanomas in people of black, Hispanic, and Asian descent.

Other types or melanoma

These include neurotropic melanoma, which tend to extend along peripheral nerves and resemble cutaneous fibrous tumors. They commonly occur in the head and neck region and are prone to multiple recurrences and metastases. Desmoplastic melanoma is a rare type of spindle cell melanoma that often arises from pigmented lesions. Amelanotic melanoma has no identifiable pigment by light microscopy but stains with HMB45 and S100 immunostains.

Biopsy

Total excisional biopsy is preferred over incisional or shave biopsy. This increases the likelihood of accurate staging which is vital for determining prognosis and definitive treatment planning. The biopsy should include a 1- to 2-mm margin of normal skin and subcutaneous fat. Frozen section analysis is inaccurate and has no role in the management of melanoma. Incisional or punch biopsies are reserved for large lesions or those that involve critical areas: digits, palm of the hand, sole of the feet. The area of clinically greatest tumor thickness should be included in the biopsy specimen.

Staging

Microstaging

The most important prognostic factor in primary melanoma is Breslow's tumor thickness. An oculomicrometer is used to measure the thickness from the top of the granular layer of the epidermis to the deepest contiguous tumor cell at the base of the lesion. Many investigators have documented an inverse correlation between Breslow's tumor thickness and survival. Thin melanomas are defined as being less than 1 mm in thickness, intermediate melanomas 1 to 4 mm., and thick melanomas measure more than 4 mm.

Clark devised a system to classify melanomas according to the level of invasion relative to histologically defined landmarks in the skin. Clark's five levels are defined in Table 2. Clark's levels correlate with tumor thickness and prognosis. The system is subjective and is hampered by the variability in skin thickness throughout the body.

Level	Characteristics
I	All tumor cells confined to the epidermis without invasion through the basement membrane. Equivalent to melanoma in situ
II	Tumor cells penetrate through the basement membrane into the papillary dermis but do not reach or extend into the reticular dermis
III	Tumor cells completely fill the papillary dermis and abut the reticular dermis but do not extend into it
IV	Tumor cells extend into the reticular dermis
V	Tumor cells extend into the subcutaneous fat

Table 2 Clark's levels of melanoma invasion

Several studies have demonstrated that tumor thickness conveys more prognostic information than does Clark's level of invasion. Gradations of Breslow's thickness within Clark's levels III, IV, and V have been associated with different survival rates. Patients with tumors of a given thickness have essentially the same survival regardless of level of invasion.

TNM staging system

The American Joint Committee on Cancer (**AJCC**) developed afour-stage TNM system that divides clinically localized melanomas into subgroups based on Breslow's tumor thickness (Table 3). Stage I includes T1N0M0 (£ 0.75 mm.) and T2N0M0 (0.76 to 1.5 mm.); stage II includes T3N0M0 (1.51 to 4.00 mm.) and T4N0M0 (> 4 mm.). AJCC stage III designates regional involvement and stage IV metastatic disease. Survival as a function of AJCC stage in shown in Table 4.



Table 3 American Joint Committee on Cancer Classification System

AJCC stage	TNM categories	Breslow's thickness (mm)	5-year survival rate (%)
I	T1, N0, M0	≤ 1.75	90–98
	T2, N0, M0	0.76–1.50	70–95
II	T3, N0, M0	1.51–4.0	60–80
	T4, N0, M0	>4	40–50
III	Any T, N1 or 2, M0		30–40
IV	Any T, any N, M1a or M1b		<10

Table 4 American Joint Committee on Cancer (AJCC) staging for melanoma and 5-year survival

Other prognostic factors

Ulceration

The presence of ulceration is second only to tumor thickness as a prognostic factor in localized melanoma. A study from the University of Alabama at Birmingham and the Sydney Melanoma Unit found the 10-year survival for stage I and II melanoma to be 79 per cent in the absence of ulceration and 50 per cent when ulceration was present.

Sex

Female patients have a better survival than do men with melanoma. This is partly explained by the predominance of extremity lesions and the melanomas tend to be less frequently ulcerated.

Location

Patients with extremity lesions have a better survival than those with trunk or head and neck melanomas. Subset analysis reveals that in the head and neck region, scalp melanomas have a worse prognosis than lesions of the face or neck. Patients with melanomas located on the hands and feet had a significantly worse prognosis than those with lesions on the arm or leg.

Age

Advancing age correlates significantly with a shortened survival.

Histological subtypes

The histological subtypes of melanoma do not have prognostic significance when the tumor thickness at the time of presentation is controlled.

Local treatment

There is an important relationship between tumor thickness and local recurrence of melanoma. Handley in 1907 recommended 5-cm excision margins based on one autopsy study of a patient with metastatic melanoma in the groin. This recommendation influenced the surgical treatment of malignant melanoma for over half a century. There have been two prospective randomized trials comparing various excision margins. The WHO trial compared 1-cm with 3-cm excision margins for melanomas of less than 2 mm in thickness. The overall survival rates were the same for both groups of patients. For melanomas of 1 to 2 mm in thickness, the local recurrence was 6 per cent for 1-cm margins and 1 per cent for 3-cm margins. The WHO trial concluded that patients with melanomas up to 2 mm could be safely treated with 1-cm margins.

A subsequent prospective randomized trial that compared 2-cm with 4-cm margins in intermediate-thickness melanoma (1 to 4 mm) was conducted by the Intergroup Melanoma Committee. Local recurrences increased as tumor thickness increased but there were no significant differences between the treatment groups. No prospective randomized trials have been conducted to provide recommendations for melanomas thicker than 4 mm. Retrospective data suggest that 2-cm margins would be appropriate.

The surgical margin is tailored to the thickness of the lesion. Depth of excision has not been clearly defined but the current belief is that fascial excision does not influence recurrence or in-transit metastases. For lesions less than 1 mm in thickness, a 1-cm margin is sufficient; for 1- to 4-mm lesions, a 2-cm margin is adequate; and for lesions greater than 4 mm, 2-cm or greater margins are obtained (Table 5). The type of closure depends on the margin of excision required and the anatomic location of the lesion. Most excisions can be performed with primary closure. Lesions of the palm of sole usually require a graft or flap closure. Subungual melanomas usually require digital amputation at the distal interphalangeal joint.

Stage I <1 mm thick	◊	Wide local excision with 1cm margins	◊	No lymphatic mapping
Stage II 1–4 mm thick	◊	Wide local excision with 2cm margins or If 1–2mm thick, 1–2cm margins acceptable in anatomically difficult areas or If > 4mm thick, then ? 3cm margins	◊	Lymphatic mapping
Stage III T any, N1, N2	◊	Excision of primary + complete node dissection Inguinofemoral nodes: deep groin dissection if pelvic CT positive or if Cloquet's node is positive. Consider deep groin dissection if clinically positive nodes, or three superficial nodes positive		

Table 5 Surgical treatment of primary malignant melanoma

Regional treatment

The management of a draining lymph node basin that is clinically negative is controversial. The clinical examination is inaccurate, with 20 to 30 per cent of intermediate-thickness melanomas (1 to 4 mm) harboring micrometastatic disease.

Elective lymph node dissection

In order for elective lymph node dissection to be effective, one must be able to identify patients at high risk for subclinical disease. Lesions less than 1 mm have an extremely low incidence of nodal metastases making elective lymph node dissection unnecessary. Lesions greater than 4 mm in thickness have a high incidence of distant metastases making elective lymph node dissection less likely to affect the ultimate outcome. Retrospective data have shown that lymph node dissections in patients with micrometastatic disease have improved survival over those presenting with macroscopic disease. Piepkorn and colleagues recently reviewed several series comparing survival for patients undergoing positive elective lymph node dissection with therapeutic dissections. They noted a 20 per cent improvement in the median 5-year survival for those with clinically negative but histologically positive nodes as opposed to those who present with grossly involved lymph nodes.

There are three large, non-randomized, prospective trials of patients with intermediate-thickness lesions who underwent elective lymph node dissection. Improved survival was noted in all three series 8 to 10 years after elective lymph node dissection.

There have been four prospective, randomized trials evaluating elective lymph node dissection with observation (Table 6). One WHO trial studied distal extremity melanomas of all thicknesses. No benefit from elective lymph node dissection was seen. The majority of cases were thin melanomas and 85 per cent were females who tend to have a survival advantage. The study has also been criticized because distal extremity melanomas tend to do better than ones at other anatomic sites and patients were not stratified for ulceration.

Trial	Anatomic site(s)	Tumor thickness	Results
WHO	Trunk, extremities	All	No benefit
Mayo	Trunk, extremities	All	No benefit
WHO	Trunk	>1.5 mm	Benefit for males with 1.5-4.0 mm
Intergroup	All	1-4 mm	Benefit for 1-2 mm

Table 6 Prospective randomized trials of elective lymph node dissection

The Mayo Clinic performed a randomized trial including all thicknesses of melanoma, but two-thirds of the lesions were less than 1.5 mm. There were three randomization arms: wide local excision; wide local excision and elective lymph node dissection; and wide local excision with elective lymph node dissection performed 30 to 60 days after treatment of the primary lesion. No survival benefit for elective lymph node dissection was found, but only 7 per cent of patients undergoing elective lymph node dissection had positive nodes documented.

A second trial was conducted by the WHO. Lesions had to be at least 1.5 mm and confined to the trunk. Interim analysis has not yet been published, but preliminary reports demonstrate that females fare better than males and that a benefit with elective lymph node dissection has been observed for patients in the 1.5- to 4-mm subset. However, the entire cohort did not benefit from elective lymph node dissection. Results of the Melanoma Intergroup randomized trial have recently been published. Eight hundred patients with melanomas 1 to 4 mm in thickness were randomized to elective lymph node dissection or observation. Patients were stratified by tumor thickness, ulceration, and anatomic subsite. Overall there was no benefit of elective lymph node dissection for the entire group but subset analysis showed a benefit for lesions 1 to 2 mm in thickness.

Sentinel lymph node biopsy

Sentinel lymph node biopsy is an effective method of detecting metastatic disease in a nodal basin. Advantages include: (i) the technique can be performed under local anesthesia; (ii) only patients with proven nodal metastases will undergo formal lymphadenectomy; (iii) those without nodal metastases will be spared the morbidity of unnecessary lymphadenectomy; and (iv) it allows for more careful histologic examination of lymph nodes by serial sectioning, immunohistochemistry, and DNA analysis. The technique was originally described by Morton and colleagues as a method of identifying patients with primary melanoma who may be harboring clinically occult nodal metastases. The technique assumes that lymphatic drainage from the site of the primary tumor follows an orderly progression through afferent lymphatics into a sentinel lymph node(s) before spreading into non-sentinel lymph nodes found in the regional nodal basin. A vital blue dye (Lymphazurin or blue patent V) is injected into the dermis around the primary tumor. Five minutes after injection, an incision is made over the draining nodal basin and dissection performed to identify a blue-stained lymphatic channel leading to a blue-stained lymph node(s). Identification of the sentinel lymph node is successful in over 85 per cent of cases using this method (Table 7). The false-negative rate, defined as the percentage of nodal basins that harbor metastases in nodes other than the sentinel lymph node when the sentinel lymph node is negative, has been reported to be less than 4 per cent in several studies. Recurrences in a nodal basin after a negative sentinel lymph node biopsy have been reported in less than 1 per cent of cases.

Author	Patients/numbers	Site	Success rate (%) Blue dye	Gamma probe	Combined
Albertini	106	All	69.5	83.5	95
Alex	15	Head and neck	75	98	95
Borstch	117	Head and neck	92	-	95
Gadelkarim	39	Trunk/extremities	100	-	-
Kapteijn	100	All	84	-	99.5
Krag	171	All	-	98	-
Liu-yen	35	Extremities	100	-	-
Morton	233	All	82	-	-
Pijpers	135	All	83	100	100
Wells	58	Head and neck	67	95	95
Thompson	118	All	88	-	-
Overall	1075		84	90.5	97

Table 7 Sentinel lymph identification

Krag and colleagues described using technetium-labeled sulfur colloid instead of vital blue dye for lymphatic mapping. A hand-held g-probe is used intraoperatively to localize the sentinel lymph node pericutaneously. This allows for smaller incisions, less dissection, and a greater sensitivity in sentinel lymph node identification. Preoperative dynamic lymphoscintigraphy may help in distinguishing second-echelon nodes which are not the sentinel lymph node (Fig. 1). The success rate of sentinel lymph node identification using this method is 97 per cent. Use of the blue dye and radioactive colloid are complementary, especially when the primary lesion is near the nodal basin. The background radioactivity of the primary lesion makes finding the sentinel lymph node difficult when the g-probe is used alone.

Fig. 1. Lymphoscintigram after injection of technetium-labeled sulfur colloid around a right forehead melanoma. The tracer localized a sentinel lymph node in the preauricular area.

Sentinel lymph nodes can be serially sectioned and immunohistochemical stains for HMB45 and S100 used to improve the detection of microscopic disease. A recent study reported a 26 per cent increase in detection (serial sectioning 8 per cent and immunohistochemistry 18 per cent) of micrometastatic disease using this combination of methods.

Therapeutic lymph node dissection

Patients who present with stage III melanoma have a poor prognosis. Calabro and colleagues reviewed 1001 patients with malignant melanoma and regional nodal disease. The 5- and 10-year overall survival rates were 33 and 28 per cent, respectively. The overall regional recurrence rate was 16 per cent. The number of nodal metastases and the presence or absence of extranodal disease were independent predictors of regional metastases.

Axillary lymph node dissection

A complete dissection, including level III lymph nodes medial to the pectoralis minor muscle, is performed. The axillary vein is skeltonized and the insertion of the pectoralis minor muscle is frequently divided to aid in removal of level III nodes.

Karakousis and colleagues reported their experience with 133 patients treated with axillary dissection for metastatic melanoma. The supra-axillary fat pad overlying the brachial plexus was frequently removed and found to contain metastatic nodes in 6 of 44 patients. The regional recurrence rate after complete axillary dissection was 19 per cent.

Groin dissection

The extent of groin dissection for metastatic melanoma has been the subject of much debate. Opponents to deep groin dissection argue that the poor survival associated with iliac and obturator node involvement do not justify the potential for increased morbidity seen with this procedure. Karakousis and Driscoll reviewed their experience with 48 patients with metastatic deep groin nodes. The 5- and 10-year survival rates were 34 and 25 per cent, respectively.

Patients with more than three clinically involved inguinal nodes or Cloquet's node involvement have an approximate 25 per cent incidence of iliac node involvement. Preoperative pelvic CT scan can be helpful in assessing the status of the pelvic lymph nodes. Metastatic involvement of Cloquet's node or multiple grossly involved inguinal nodes are used as an indication to perform a deep groin dissection by many surgeons.

Head and neck dissection

Radical neck dissection has been the traditional treatment for metastatic melanoma to the cervical lymph nodes. The Sydney Melanoma Unit reviewed their experience of 397 patients with melanoma treated with cervical node dissections. The incidence of recurrence in the neck was 24 per cent overall: 28 per cent when nodes were histologically positive and 13 per cent when nodes were histologically negative at the time of surgery.

The role of modified neck dissection in the therapeutic setting is uncertain. Byers reported a recurrence rate in the neck of 50 per cent for patients treated with clinically and histologically positive nodes compared with only 17 per cent in those treated for subclinical disease. Adjuvant radiation therapy administered in a high dose per fraction is currently being studied to reduce the regional recurrence rate.

Isolated limb perfusion

In-transit metastases arising in the skin or subcutaneous tissue between the primary site and the regional lymph node basin are an uncommon site of recurrence. Isolated limb perfusion has been developed to treat this type of recurrence. It allows for prolonged perfusion with cytotoxic or biologic agents. Melphalan is the most common agent used. It is heated to 41°C and perfused for up to 90 min. Hyperthermic isolated perfusion with melphalan alone or combined with tumor necrosis factor-a results in regression of more than 90 per cent of cutaneous in-transit metastases. Studies have not found isolated limb perfusion effective in the adjuvant setting.

Surgical therapy of metastatic melanoma

Spread of melanoma beyond the regional lymph nodes has a grave prognosis (median survival 6 months) and is usually treated by systemic agents. Patients with isolated metastases to the lung or gastrointestinal tract that can be completely resected will have a prolonged disease-free survival. Palliative resection of brain metastases is occasionally indicated.

Adjuvant therapy

The Eastern Cooperative Oncology Group has reported the results of a large, randomized, multicenter study in patients with high-risk melanoma. Administration of adjuvant interferon-a2b resulted in improvement in relapse-free and overall survival compared with observation alone. In this trial, 287 patients with thick primary (T4N0) or node-positive (N1 to 2) melanoma were randomized either to adjuvant interferon-a2b therapy at 20 MIU/m^2 every day (intravenously), 5 days/week for 4 weeks, followed by 10 MIU/m^2 three times weekly (subcutaneously) for 48 weeks, or observation.

Interferon-a2b therapy significantly increased median relapse-free survival by 9 months and produced a 42 per cent improvement in the 5-year relapse-free survival rate. In addition, interferon-a2b therapy significantly increased median overall survival by 1 year and produced a 24 per cent improvement in the 5-year overall survival rate.

Further reading

Albertini JJ *et al.* Intraoperative radio-lympho-scintigraphy improves sentinel lymph node identification for patients with melanoma. *Annals of Surgery* 1996; **223**: 217–24.

Alex JC *et al.* Localization of regional lymph nodes in melanomas of the head and neck. *Archives of Otolaryngology, Head and Neck Surgery* 1998; **124**: 135–40.

Balch CM, Soong SJ, Milton GW, Shaw HM, McGovern VJ, Murad TM. A comparison of prognostic factors and surgical results in 1786 patients with localized (stage I) melanoma treated in Alabama, USA, and New South Wales, Australia. *Annals of Surgery* 1982; **196**: 677–84.

Balch CM, Soong SJ, Shaw HM, Milton GW. An analysis of prognostic factors in 4000 patients with cutaneous melanoma. In: Balch CM and Milton GW, eds, *Cutaneous melanoma*, pp. 321–52. JB Lippincott Co., New York, 1985.

Balch CM *et al.* Efficacy of 2-cm surgical margins for intermediate-thickness melanomas (1 to 4 mm). *Annals of Surgery* 1993; **218**: 262–9. [Randomized trial of 2- versus 4-cm margins in the treatment of intermediate-thickness melanomas.]

Balch CM, Soong SJ, Bartolucci AA. Efficacy of an elective regional lymph node dissection of 1 to 4 mm thick melanomas for patients 60 years of age or younger. *Annals of Surgery* 1996; **224**: 255–66. [Melanoma Intergroup randomized trial of elective lymph node dissection in the treatment of intermediate-thickness melanoma.]

Bostick P *et al.* Intraoperative lymphatic mapping for early-stage melanoma of the head and neck. *American Journal of Surgery* 1997; **174**: 536–9.

Breslow A. Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. *Annals of Surgery* 1970; **172**: 902–8.

Byers RM. The role of modified neck dissection in the treatment of cutaneous melanoma of the head and neck. *Archives of Surgery* 1986; **121**: 1338.

Calabro A, Singletary SE, Balch CM. Patterns of relapse in 1001 consecutive patients with melanoma nodal metastases. *Archives of Surgery* 1989; **124**: 1051–5. [Large retrospective review of patterns of relapse after lymph node metastases in melanoma.]

Clark WHJ. The histogenesis and biologic behavior of primary human malignant melanomas of the skin. *Cancer Research* 1969; **29**: 705–27.

Geara FB, Ang KK. Radiation therapy for malignant melanoma. *Surgical Clinics of North America* 1996; **76**: 1383–98.

Gershenwald J, Thompson W, Mansfield P. Patterns of of recurrence following a negative sentinel lymph node biopsy in 243 patients with stage I or II melanoma. *Journal of Clinical Oncology* 1998; **16**: 2253–60.

Godellas CV *et al.* The identification and mapping of melanoma regional nodal metastases: Minimally invasive surgery for the diagnosis of nodal metastases. *American Surgeon* 1995; **61**: 97–101.

Handley WS. The pathology of melanotic growths in relation to their operative treatment. *Lancet* 1907; **i**: 927. [Landmark paper that influences surgical treatment of melanoma for over half a century.]

Kapteijn BA *et al.* Localizing the sentinel node in cutaneous melanoma: Gamma probe detection versus blue dye. *Annals of Surgical Oncology* 1997; **4**: 156–60.

Karakousis CP, Driscoll DL. Positive deep nodes in the groin and survival in malignant melanoma. *American Journal of Surgery* 1996; **171**: 421–2.

Karakousis CP, Hena MA, Emrich LJ, Driscoll DL. Axillary node dissection in malignant melanoma: Results and complications. *Surgery* 1990; **108**: 10–17.

Kirkwood JM, Strawderman MH, Ernstoff MS. Interferon a-2b adjuvant therapy of high-risk resected cutaneous melanoma: The Eastern Cooperative Oncology Group trial EST 1684. *Journal of Clinical Oncology* 1996; **14**: 7–17. [The first randomized trial that showed a survival benefit for adjuvant therapy after primary treatment of melanoma.]

Krag DN *et al.* Minimal-access surgery for staging of malignant melanoma. *Archives of Surgery* 1995; **130**: 654–8. [Landmark paper describing the use of radioactive tracers and intraoperative probe to localize sentinel lymph nodes.]

Lingam MK, Mackie RM, McKay AJ. Intraoperative identification of sentinel lymph node in patients with malignant melanoma. *British Journal of Cancer* 1997; **75**: 1505–8.

Messina JL. Pathologic examination of the sentinel lymph node in malignant melanoma. *American Journal of Pathology*, 1999; **23**: 686–90.

Milton GW, Shaw HM, McCarthy WH, Pearson L, Balch CM, Soong SJ. Prophylactic lymph node dissection in clinical stage I cutaneous malignant melanoma: results of surgical treatment in 1319 patients. *British Journal of Surgery* 1982; **69**: 108 *American Journal of Surgery* 11.

Morton DL *et al.* Technical details of intraoperative lymphatic mapping for early stage melanoma. *Archives of Surgery* 1992; **127**: 392–9. [Landmark paper describing sentinel lymph node mapping using vital blue dye.]

O'Brien CJ, Gianoutsos MP, Morgan MJ. Neck dissection for cutaneous maligant melanoma. *World Journal of Surgery* 1992; **16**: 222–6. [Large series of patients with head and neck melanoma treated by neck dissection.]

Piepkorn M, Weinstock MA, Barnhill RL. Theoretical and empirical arguments in relation to elective lymph node dissection for melanoma. *Archives of Dermatology* 1997; **133**: 995–1002.

Pijpers R, Borgstein PJ, Meijer S, Hoekstra OS, van Hattum LH, Teule GJ. Sentinel node biopsy in melanoma patients: dynamic lymphoscintigraphy followed by intraoperative gamma probe and vital dye guidance. *World Journal of Surgery* 1997; **21**: 788–92.

Reintgen DS, Cox EB, McCarty KS, Vollmer RT, Seigler HF. Efficacy of elective lymph node dissection in patients with intermediate thickness primary melanoma. *Annals of Surgery* 1983; **198**: 379–85.

Reintgen D, Cruse CW, Wells K, Berman C. An orderly progression of melanoma nodal metastases. *Annals of Surgery* 1994; **220**: 759–67.

Ross M, Reintgen DS, Balch C. Selective lymphadenecomy: Emerging role of lymphatic mapping and sentinel node biopsy in the management of early stage melanoma. *Seminars in Surgical Oncology* 1993; **9**: 219–23.

Sim FH, Taylor WF, Pritchard DJ, Soule EH. Lymphadenectomy in the management of stage I malignant melanoma: A prospective randomized study. *Mayo Clinic Proceedings* 1986; **61**: 697–705. [Randomized trial performed at the Mayo Clinic evaluating elective lymph node dissection in the treatment of stage I melanoma.]

Thompson JF *et al.* Sentinel lymph node status as an indicator of the presence of metastatic melanoma in regional lymph nodes. *Melanoma Research* 1995; **5**: 255–60.

Veronesi U. Delayed regional lymph node dissection in stage I melanoma of the skin of the lower extremities. *Cancer* 1982; **49**: 2420–30. [WHO randomized trial of elective lymph node dissection in the treatment of stage I melanoma.]

Veronesi U, Cascinelli N. Narrow excision (1-cm margin). A safe procedure for thin cutaneous melanoma. *Archives of Surgery* 1991; **126**: 438–41. [Randomized trial of 1- versus 3-cm margins in the treatment of early melanoma.]

Wells KE *et al.* Sentinel lymph node biopsy in melanoma of the head and neck. *Plastic and Reconstructive Surgery* 1997; **100**: 591–4.

38.6 Cutaneous problems occurring in the surgical patient

David A. Wrone, Richard G. B. Langely and Arthur J. Sober

Drug reactions
Therapy
Contact dermatitis
Contact dermatitis in the surgeon
Infections of the skin
Common bacterial infections
Skin infection in the immunocompromised host
Cutaneous problems in transplant patients
Toxic shock syndrome
Xerosis
Miliaria
Decubitus ulcers (pressure ulcers, bedsores)
Hidradenitis suppurativa (suppurative hidradenitis, apocrinitis)
Cellulitis (erysipelas)
Folliculitis
Abscess (furuncle, boil)
Carbuncle
Herpes simplex virus
Further reading

Numerous cutaneous problems can occur in patients cared for by surgeons. Some, such as drug reactions, xerosis, decubiti, or miliaria, are not specific to the surgical patient, unlike others, such as contact dermatitis induced by dressing materials (tape, tincture of benzoin) or problems with stoma maintenance. A discussion of the more common cutaneous problems follows: stomal care will be considered elsewhere. Dermatologic consultation may be of value if diagnosis or management appear problematic.

Drug reactions

Surgical patients are exposed to multiple medications in the preoperative, operative, and postoperative periods. If an adverse reaction occurs, every effort should be made to identify the causative agent as specifically as possible. Time course, probability, appearance of the reaction (type), and the patient's history of drug allergy are helpful in determining the most likely agent(s). Table 1 lists the drugs associated with specific patterns of cutaneous reaction. The generalized, morbilliform eruptions (erythematous macules and papules) are not usually associated with life-threatening reactions. If medication producing this type of eruption is essential, it can frequently be continued; the eruption may or may not disappear. Close monitoring for progression of the reaction is advised.

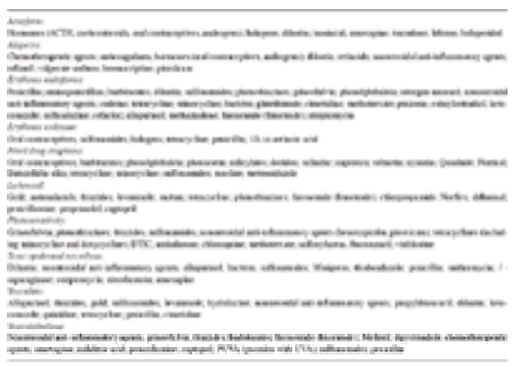


Table 1 Selected drugs associated with cutaneous eruptions with distinctive appearances

The risk of life-threatening anaphylaxis is increased in patients who develop urticarial eruptions and every effort should be made to identify the causative agent. Administration should be stopped if possible. Patients should be monitored for signs and symptoms of respiratory or circulatory difficulty. Treatment with drugs causing bullous erythema multiforme or toxic epidermal necrolysis should be stopped. Patients developing toxic epidermal necrolysis have a mortality rate of over 50 per cent; management in a burn centre is recommended, as death usually results from fluid and electrolyte imbalance or sepsis from a cutaneous source.

The course is helpful in determining the most likely cause of the reaction: charting the onset and cessation of medications by date helps in the evaluation. Most cutaneous reactions will occur within 1 week of starting treatment, but a longer interval is seen in 50 per cent of patients who develop reactions to ampicillin and semisynthetic penicillins. Eruptions to medications that have been used for some time can occur, although they are less common than those developing in response to a newly introduced drug. Reactions can begin as long as 1 week after a medication has been stopped and can continue for 7 to 14 days: persistence of the rash for several days after discontinuation of a suspected cause does not mean that the incorrect drug has been withdrawn. The administration of high doses of corticosteroids, as is common in neurosurgical units, can mask a drug reaction, which may become apparent only when the corticosteroid doses are tapered.

Medications that frequently cause rashes include sulfonamides, penicillins, and blood products; others, including digitalis preparations, antacids, vitamins, minerals, and stool softeners, rarely cause cutaneous reactions. Table 2 lists drugs frequently associated with reactions, while Table 3 shows those that rarely cause cutaneous problems.

Drug	%
Amoxicillin	6.3
Trimethoprim–sulfamethoxazole	5.9
Ampicillin	5.2
Semisynthetic penicillins	3.6
Blood, whole human	3.5
Corticotropin	2.8
Platelets	2.8
Ipodate	2.8
Erythromycin	2.3
Sulfisoxazole	1.7
Penicillin	1.6
Gentamicin sulphate	1.6
Practolol	1.6
Cephalosporins	1.3
Quinidine	1.2
Plasma protein fraction	1.2
Digoxine	1.1

Table 2 Drugs producing cutaneous reactions in at least 1 per cent of exposed individuals



Table 3 Drugs received by at least 500 patients with no skin reactions

A carefully taken history of drug allergy is essential. Many patients have had a previous reaction to a similar medication. They may also have inadvertently been exposed to a cross-reacting drug.

Identification of the offending agent may not be possible in some patients who are receiving many drugs simultaneously. In this situation, elimination of all non-essential medication is recommended, with continuation of essential drugs and close monitoring.

Therapy

Most drug reactions do not require specific treatment, other than for symptomatic relief of itch. Emollients (moisturizers) may be helpful for mildly symptomatic patients. Oral antihistamines (hydroxyzine, 10 to 25 mg, or chlorpheniramine maleate, 4 mg every 4 to 6 h as required) may also be of help in patients for whom the drowsiness is not intolerable and other medical contraindications are absent. Oral steroids are usually not warranted.

Contact dermatitis

Products employed by surgeons may produce allergic contact dermatitis in susceptible patients. The presence of vesicles or bullae, geographic location, sharp borders, and itch are features that point toward contact dermatitis rather than superficial infection (cellulitis, bullous impetigo), which it sometimes resembles. In an unsensitized individual the eruption usually starts seven or more days after contact with the inducer and may last up to 3 weeks. In a previously sensitized individual, onset after 24 to 72 h is more typical; the duration may be 10 days to 2 weeks. A careful history of prior sensitizations should be elicited. along with a record of current exposures. Common sources of contact dermatitis in surgical patients include topical preparations containing neomycin or diphenhydramine, tincture of benzoin, metal appliances, and tape (Table 4). Neomycin sensitivity is common, affecting approximately 1 per cent of the population. Sensitivity is much more common (15–30 per cent) after neomycin is applied to chronic erythematous skin, as in venous stasis dermatitis. Once sensitized, cross-reactions with other topical aminoglycosides can occur. Concomitant sensitization with bacitracin is also common. Allergy to topical sulfonamides can result in cross-reactions to *p*-amino compounds such as *p*-aminobenzoic acid-containing sun blocks, benzocaine, and hair dye. Sensitization to systemic antihistamines is rare, yet topical applications commonly produce allergic contact dermatitis: topical diphenhydramine is ineffective and best avoided.

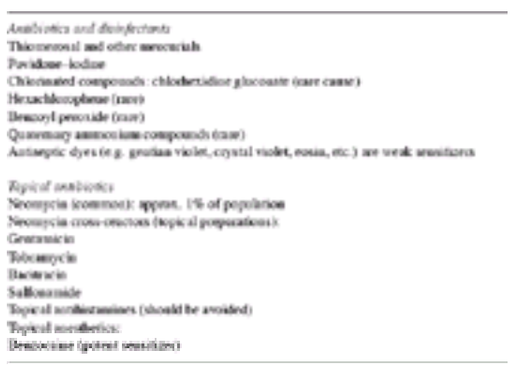


Table 4 Contactants in the hospitalized patient

Among the topical anesthetics, benzocaine is a potent sensitizer, and sensitive patients will also react to procaine, tetracaine, and other agents. Safe topical anesthetics for use in benzocaine-sensitive persons include lidocaine (lignocaine), pramoxine (pramocaine), mepivacaine, pyrrocaine, and prilocaine hydrochloride.

Treatment of contact dermatitis requires cessation of the product and avoidance of agents that may cross-react. In mild cases these efforts are sufficient. If vesicle formation is substantial or if serous drainage is prominent, wet–dry saline solution-soaked gauze dressings applied for 30 min three or four times daily help dry the lesions and remove the accumulated drainage. Topical steroids are not effective.

If the eruption is very extensive, with substantial symptoms or involvement of major structures (swollen shut eyes, edema of genitalia), oral cortisone may help, if this is not medically contraindicated. Typical doses would be prednisone, 60 mg/day, with milk or antacid for 3 days, reducing the dose by 5 mg/day over a period of 2 weeks.

Contact dermatitis in the surgeon

Surgeons have an increased risk of certain types of contact dermatitis because of their occupational exposure. Table 5 lists agents that may cause contact dermatitis in surgeons and surgical personnel.

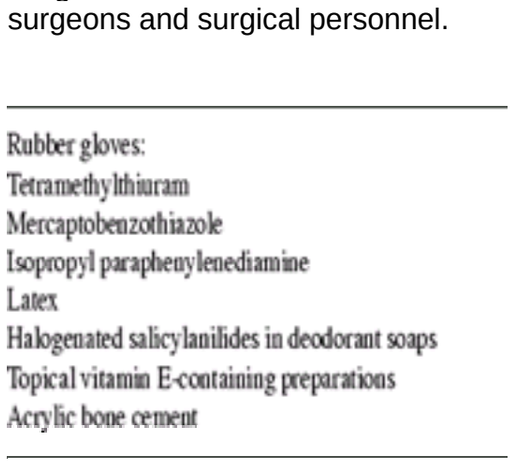


Table 5 Agents producing allergic contact dermatitis in surgical personnel

Contact dermatitis can be either irritant or allergic in nature. Hexachlorophene, a preoperative scrubbing agent, is a common cause of primary irritant dermatitis, but actual allergic contact dermatitis to this agent is rare. Allergy to components of rubber gloves should be suspected in eruptions that stop abruptly above the wrist. Antioxidants and accelerators are the major cause of allergic contact dermatitis to rubber gloves. Such sensitizers include tetramethylthiuram, mercaptobenzothiazole,

and isopropyl *p*-phenylenediamine. Surgeons who develop such sensitivity may benefit from the use of hypoallergenic gloves. Orthopedic surgeons are exposed to acrylic bone cement, which can penetrate rubber gloves and produce sensitization. Excessive dryness and fissuring of the skin in the affected areas is accompanied by paresthesias of the fingertips. Contact with cement must be avoided by the sensitized individual. The double-glove technique may be successful in preventing contact dermatitis unless the sensitized individual is allergic to methylmethacrylate monomer. Starch glove powder is another source of irritation; allergic contact dermatitis can be avoided by alternative use of talc. Evaluation of potential offending agents can be performed by patch testing with standardized agents.

Allergic reactions to latex have been increasingly seen in recent years and may affect up to 10 per cent of operating-room person-nel. The hypersensitivity is either type I (IgE mediated) or type IV (delayed type). Reactions typically include contact dermatitis, urticaria, angioedema, rhinoconjunctivtis, asthma, and anaphylaxis. The surgeon should be alert for patients with latex allergy: those at particular risk include atopic patients, patients who have undergone multiple surgical procedures, and patients with spina bifida who have genitourinary abnormalities. Patients or health-care workers who are suspected to have latex allergy should be referred for diagnosis (skin testing, radioallergosorbent test, enzyme-linked immunosorbent assay). Surgeons who have patients with latex allergy need to operate in a latex-free environment, i.e., latex-free gloves, intravenous equipment, catheters, tourniquets, airway equipment. Emergency carts should also be available with non-latex products in hospital wards.

Infections of the skin

Common bacterial infections

Impetigo

This superficial infection of the outer layers of the skin is characterized by the presence of honey-colored crusts. Previously, most cases were due to infection by streptococci; occasionally, mixed streptococcal–staphylococcal infection was seen. Now the majority of cases are caused by staphylococci alone; a few mixed infections are still seen. This condition is especially common in small children. The differential diagnosis includes contact dermatitis and infected atopic dermatitis.

Treatment

Traditionally, treatment required a 10-day course of oral antibiotics, either penicillin VK, 250 mg every 6 h (or the equivalent dose in small children), or erythromycin, 250 mg every 6 h. Topical treatment with mupirocin ointment (Bactroban) three times daily has recently been used successfully, with the addition of oral antibiotics if topical therapy fails. For extensively crusted areas, wet–dry saline soaks help debride and dry the area. Complications of impetigo are unusual, but rare cases of post-streptococcal glomerulonephritis have been reported.

Moniliasis (candidiasis)

Moniliasis is a common problem in hospital patients, especially those who are bedridden, febrile, and diaphoretic. Other predisposing factors include obesity (deep body folds), exposure to broad-spectrum antibiotics, diabetes, and oral contraceptives. An erythematous, macerated, scaly eruption is noted in intertriginous areas. Satellite pustules are present. Microscopic examination of KOH-immersed scrapings of the pustular contents demonstrates pseudohyphae. Extensive eruptions can involve the entire back and buttocks.

Therapy is aimed at decreasing the factors producing a warm, moist environment. Ambulation, where possible, is usually helpful. Applications of nystatin cream twice daily, or one of the imidazole creams such as miconazole or clotrimazole twice daily, usually produce improvement. Adjunctive measures include drying the area with a fan several times daily and the use of 1 per cent hydrocortisone cream lightly applied twice daily for a few days to give symptomatic relief.

Skin infection in the immunocompromised host

Skin infections in the compromised host fall into four groups: infection originating in skin and similar in appearance to that in patients with an intact immune system; extensive infections, which are usually more limited in the immunocompetent; infection from cutaneous opportunistic pathogens; and spread of infection from a noncutaneous site to the skin.

The appearance of infectious processes may be altered in the immunocompromised patient and diagnosis may only be possible through histopathologic examination of tissue, using both routine stains and special stains for yeast, fungi, and acid-fast organisms, or by culture of biopsy tissue for routine and unusual organisms, or by culture of biopsy tissue for routine and unusual organisms. Table 6 describes the gross appearances of cutaneous lesions and some of the associated organisms found in these patients.



Table 6 Cutaneous and subcutaneous lesions caused by opportunistic organisms that infect compromised patients

Cutaneous problems in transplant patients

Skin problems frequently seen in transplant patients fall into three categories: accelerated development of malignancies; increased frequency, extent, and recalcitrance of cutaneous infections; and medication-related problems.

The most common cutaneous malignancies are of the non-melanoma types, including squamous-cell and basal-cell carcinoma, and keratoacanthoma (benign). These tumors appear earlier in life than anticipated, but do arise in sites commonly exposed to the sun. There is an increased incidence of squamous-cell carcinoma in transplant patients than in the population at large. Management is similar to that in non-transplant patients. Some keratoacanthomas behave in a more aggressive fashion in transplant patients; they may best be treated as squamous-cell carcinomas.

Table 7 lists the cutaneous infections that are seen more often in transplant patients. Opportunistic systemic infections may present on the skin. Histopathologic examination of specially stained biopsy specimens and culture should be performed on undiagnosed cutaneous lesions in these patients.

Viral:
Molluscum contagiosum
Warts
Dermatophytoses
Tinea versicolor
Moniliasis

Table 7 Common infections in transplant patients

Cutaneous side-effects of medication include those related to oral steroids, such as thinning of skin, increased bruising, striae, acne, moon facies, and hirsutism. Cyclosporin, one of the immunosuppressants, also produces hirsutism in some patients.

Toxic shock syndrome

Toxic shock syndrome, while initially associated with the use of superabsorbent tampons in young women, has increasingly been seen in the postoperative period from surgical wound infections. It is caused by enterotoxins released by a strain of *Staphylococcus aureus*. The syndrome is characterized by the presence of fever, hypotension, confusion, and often a scarlatiniform eruption. It can cause multiorgan failure, including gastrointestinal (vomiting, diarrhea), muscular (myalgias), renal insufficiency, hepatic (raised aspartate and alanine aminotransferases, bilirubin), and central nervous system (disorient-ation).

It is critical to recognize toxic shock syndrome, as the case fatality rate is reportedly 4.8 per cent in menstrual and 12.5 per cent in non-menstrual cases.

Treatment consists of immediate use of parenteral antistaphylococcal antibiotics and supportive measures.

Xerosis

Surgical patients in hospital often suffer from dry, scaly, and itchy skin (xerosis). The hospital environment of frequent bathing, use of antibacterial soaps, and low humidity often exacerbates this pre-existing problem. Xerosis can be at least partially be overcome by the use of oil or superfatted soaps, decreased frequency of bathing, and an increased environmental humidity (30–40 per cent). Frequent application of emollients, including compounds containing a-hydroxy acids, is a cornerstone of therapy and prevention.

Miliaria

Miliaria or ‘prickly heat' is a common problem in febrile, diaphoretic patients in overheated hospital rooms. It results from blockage of the sweat orifices, resulting in a sweat-containing vesicle appearing on the skin (miliaria crystallina); the duct may rupture and become inflamed, forming erythematous papules (miliaria rubra); influx of polymorphonuclear leucocytes produces the third stage, miliaria pustulosa. The differential diagnosis includes moniliasis, folliculitis, and drug-induced eruptions. Reduction of temperature and humidity, and the encouragement of walking, are usually sufficient to resolve the condition. Symptomatic improvement may be hastened by application of agents, such as 2 to 3 per cent salicylic acid in 70 per cent isopropanol, that remove the top layers of the stratum corneum and reopen the ducts.

Decubitus ulcers (pressure ulcers, bedsores)

No one measure or combination of measures is effective in preventing or healing all decubitus ulcers. The primary pathophysiologic event is chronic pressure on skin over a bony prominence. Tissue ischemia follows, producing areas of necrosis that are susceptible to infection. Friction and prolonged contact with urine or feces are additional predisposing factors. Resection of an underlying bony prominence may be necessary to allow healing to occur.

The most common sites of decubitus ulcers are the sacrum, tibial crest, ischial tuberosities, greater trochanters, iliac spine, spinous processes, heels, malleoli, knees, and elbows. Decubiti may also occur over the occiput and on the ears. Elderly, debilitated, paralysed, and unconscious patients are at greatest risk.

Prevention is the best approach to decubitus care. Frequent repositioning of the patient (every 1–2 h) usually prevents the development of these ulcers. Ambulation, when possible, also diminishes the risk. Good nutrition is essential to the maintenance of cutaneous integrity and also to facilitate healing once decubiti have developed. Correction of anemia is an additional adjunctive measure. The ulcer itself is managed in the same way as any cutaneous ulcer.

Skin in areas at risk should be examined frequently for the presence of localized erythema, which indicates the presence of increased pressure and is a forerunner to ulceration. Alternating-pressure mattresses, water-filled mattresses, and ‘egg-crate' pads (thick, papillated foam) are helpful in the prevention and treatment of ulcers.

Incision and drainage of any infected space and removal of necrotic tissues is an important first step in treatment. Various topical powders, creams, and semipermeable dressings have been used with varying degrees of success. Compounds in the research category, including growth factors, may prove to be helpful adjuncts to standard therapy. When medical approaches to ulcer healing have failed, several surgical approaches are possible. Complete excision of the ulcer with split-thickness skin-graft coverage may heal ulcers that have poor granulation tissue, necrotic edges, and lack of epithelial ingrowth. Full-thickness grafts or rotational flaps and pedicles are helpful when bone and periosteum have been damaged or when little subcutaneous support is present. The use of myocutaneous flaps is gaining in popularity.

Hidradenitis suppurativa (suppurative hidradenitis, apocrinitis)

This inflammatory, suppurative condition of unknown cause involves the apocrine glands and is frequently seen as a tetrad of hidradenitis suppurativa, severe cystic acne (acne conglobata), pilonidal sinus, and perifolliculitis capitis (dissecting folliculitis). While medical management can frequently be helpful early in the disease process, patients with extensive and persistent disease can only be managed effectively by surgery. Since apocrine glands are not fully developed until puberty, prepubertal patients are uncommon. Once developed, hidradenitis may persist into older adult life. The condition affects areas in which the apocrine glands are found: axillae, groin, buttocks, scrotum, perineum, submammary, areolar, and periumbilical skin. Lesions initially present as tender, erythematous to flesh-colored, subcutaneous nodules. Treatment at this stage includes incision and drainage, and oral antibiotics, but severe scarring may result. Progression results in the development of multiple, large, interconnected, chronic, deep-seated abscesses with sinus tracts between lesions and to the overlying skin. Treatment is surgical, and may include the unroofing of sinus tracts to allow healing by granulation, or total excision of the apocrine-containing tissues (this is more successful in the axilla than in the groin).

Early in the course of the disease, histopathologic examination of the skin shows a perifolliculitis with infiltrating cells comprising lymphocytes, polymorphonuclear leucocytes, and histiocytes. The primary abnormality may be apocrine follicular occlusion. The pilosebaceous structures are destroyed by resulting abscesses; sinus tracts lined with epidermis and extensive scar formation are late findings. Differential diagnosis includes the fistulas of regional ileitis, furunculosis, infected cysts, developmental fistulas, cystic acne, and lymphogranuloma venereum.

Treatment

Non-specific measures, including weight reduction for obese individuals, improved hygiene, and incision and drainage or intralesional steroids are helpful early in the infection. Short courses of antibiotics may also be successful (oral erythromycin or oral tetracycline in divided doses, a total of 1 g daily for 1–2 weeks). Chronic, low-dose antibiotic therapy may help reduce flares of the condition in patients with multiple, recurrent lesions. Oral 13-*cis* retinoic acid has produced temporary improvement in some severely affected patients, but side-effects include malformed fetuses in women receiving the drug during pregnancy, calcium deposits in tendon and bone, increased blood lipids, night blindness, and severe dryness.

Surgical treatments include unroofing of sinus tracts with healing by secondary intention and wide excision of apocrine gland-containing areas that are chronically affected. Good results can be achieved in the axilla following wide excision and primary closure (Pollock procedure). Squamous-cell carcinoma occasionally arises in chronically scarred areas.

Cellulitis (erysipelas)

This non-suppurative infection of the connective tissue (dermal) and subcutaneous layers of the skin may be caused by several types of bacteria, of which hemolytic streptococci are the most common. These are usually group A organisms, but b-hemolytic streptococci of groups B, C, or G may also be responsible.

Clinical appearance

Erythema, swelling, and tenderness spread rapidly, with sharp or vaguely defined borders. A defined portal of entry into the skin may or may not be present (puncture site, cutaneous ulcer, surgical wound, fissures in skin secondary to tinea pedis). Moderate to high fever may be present, with or without local adenopathy. Predisposing conditions include saphenous venectomy, chronic lymphedema, and phlebitis. The inflammation may be relatively superficial with well-demarcated borders. Patients with swollen arms or legs after dissections of proximal regional lymph nodes should be given a prescription for penicillin V or another antibiotic in anticipation of need, so that they can treat themselves at the first sign of erysipelas.

Differential diagnoses include contact dermatitis, superficial thrombophlebitis, and postinjection chemical inflammation.

Therapy

Oral antibiotics are the most important, single component in treatment; admission to hospital may be necessary for non-compliant, diabetic, immunosuppressed, or septicemic patients. Nafcillin, 2 g every 4 h, is one recommended program for the inpatient. Dicloxacillin, 500 mg every 4 h, or cefazolin, 1 g every 6 h, are reasonable alternatives for outpatient therapy. Treatment should be continued for 10 days. Adjunctive measures include application of heat to the area by warm, moist compresses, elevation of the area, and rest. Failure to respond in 48 to 72 h may indicate the presence of an abscess, which may require incision and drainage.

Folliculitis

This is a superficial pustular eruption localized to the opening of the hair follicle, commonly caused by *S. aureus* infection. Predisposing factors include obesity, diabetes, moist humid climate, and tight or occlusive clothing. A small pustule with a hair emerging is the typical clinical finding. Healing usually occurs without scarring. Contact with mineral oils or tar products may produce a sterile folliculitis; a similar sterile folliculitis may occur beneath adhesive plasters or adhesive dressings. Differential diagnoses includes miliaria pustulosa (see above), inflammatory tinea, and moniliasis.

Therapy

Folliculitis is often self-limiting. Improvement in hygiene and the use of antibacterial scrubs such as 3 per cent hexachlorophene or chlorhexidine gluconate may clear the infection; wearing of loose-fitting cotton clothing often helps. If the infection is persistent or extensive, treatment with dicloxacillin or erythromycin (each 250 mg four times daily for 7 days) is usually sufficient. Recurrence is common.

Abscess (furuncle, boil)

An abscess is a localized collection of pus within the skin associated with erythema, tenderness, and showing marked infiltration by polymorphonuclear leucocytes. An abscess that involves the hair follicle is termed a furuncle. *Staphylococcus aureus* is the most common causative organism; streptococci and Gram-negative organisms may also be present.

Epidemiology

Furuncles are uncommon in children, except in those with atopic dermatitis; the incidence increases at puberty, becoming common in adolescents and young adults. Men are more frequently affected than women. Most patients have intact immune systems. Local mechanical factors participate in the pattern of distribution of lesions.

Clinical appearance

The earliest lesion is a small, tender, inflammatory nodule in a follicular location. Pustulation occurs, followed by necrosis and discharge. Healing results in scar formation. Furuncles may be single or multiple. Common sites are the buttocks, anogenital area, face, neck, arms, wrists, and fingers. Cystic acne, hidradenitis suppurativa, and inflamed epidermal inclusion cysts are the lesions that most commonly need to be distinguished from furuncles.

Therapy

Since the course of furuncles is often self-limiting, therapy may not be necessary. Large lesions should be treated by incision and drainage. Antibiotic therapy, such as dicloxacillin 250 to 500 mg every 4 h for 7 days, is also warranted in this situation. Patients who show persistent recurrences may be nasal or perineal carriers of the causative organism.

Carbuncle

A carbuncle is a suppurative extension of several contiguous furuncles into the subcutaneous fat. The maintenance of fascial attachments to the skin results in the production of multilocular compartments, particularly in the nape of the neck, upper back, hips, and thighs. Men in middle to late life are most commonly affected; predisposing factors include malnutrition, diabetes, cardiac failure, prolonged steroid therapy, and generalized dermatoses. *Staphylococcus aureus* is the most common cause.

Clinical features

The initial lesion is a painful, dome-shaped nodule, which may increase in size over several days to 10 cm or more. After about 1 week, pus-draining follicular orifices are noted, followed by necrosis of skin and deep ulceration. Fever and malaise may precede or accompany the development of the lesion.

Treatment

A carbuncle should be treated with a penicillinase-resistant systemic antibiotic such as dicloxacillin (250–500 mg every 6 h), particularly if cellulitis is suspected. If the lesion is large, incision and drainage is indicated.

Herpes simplex virus

Herpes zoster and herpes simplex are both common in hospital patients. Herpes simplex appears clinically as grouped vesicles on an erythematous base and may be provoked by febrile episodes. Herpes zoster may appear as clusters of vesicles located unilaterally in the distribution of a single dermatome. The differential diagnosis includes contact dermatitis and cellulitis. The diagnosis can be confirmed by viral culture or, more rapidly and less expensively, by staining a smear of the vesicle base with Giemsa dye and examining the microscopic preparation for the presence of multinucleated giant cells.

Therapy for herpes simplex is aimed at reducing pain, promoting healing, and decreasing the potential for secondary bacterial infection. Local measures, such as wet–dry saline dressings three or four times daily, may be all that is required. Herpes zoster may initially need large doses of medication for pain relief. Oral acyclovir, 800 mg five times daily (alternately, famciclovir 500 mg or valacyclovir 1000 mg every 8 h), may be administered for 7 to 10 days. Secondary bacterial infection, if it develops, can be treated with appropriate antibiotics, selected on the basis of the results of Gram-staining or culture. Persistent localized herpes simplex in patients with the acquired immune deficiency syndrome may require treatment with oral acyclovir in doses of 200 to 400 mg five times daily until the lesions are completely healed. Relapses are common.

Disseminated herpes zoster is rare in healthy people but more common in patients with lymphoma. This condition will usually respond to intravenous acyclovir in doses of 10 mg/kg every 8 h (if renal function is normal).

Topical antivirals are largely ineffective in altering the clinical outcome of herpes simplex virus infections in individuals with normal immunologic capabilities.

Further reading

Ariyan S, Krizek TJ. Hidradenitis suppurativa of the groin treated by excision and spontaneous healing. *Plastic and Reconstructive Surgery* 1976; **58**: 45–7. [A description of the authors' surgical technique.]

Bell BA, Ellis H. Hidradenitis suppurativa. *Journal of the Royal Society of Medicine* 1978; **71**: 511–15. [A discussion of hidradenitis suppurativa.]

Bigby M, Jick S, Arndt K. Drug-induced cutaneous reactions. *Journal of the American Medical Association* 1986; **256**: 3358–63. [A classic article that discusses the results of the Boston Collaborative Drug Surveillance Program providing quantitative data on cutaneous drug reactions.]

Blacker KL, Stern RS, Wintroub BU. Cutaneous reactions to drugs. In: Fitzpatrick TB *et al.*, ed. *Dermatology in general medicine*, 4th edn, pp. 1783–95. McGraw-Hill, New York, 1993. [A review with excellent tables describing cutaneous reaction to drugs.]

Britton JW, Fajardo JE, Krafte-Jacobs B. Comparison of mupirocin and erythromycin in the treatment of impetigo. *Journal of Pediatrics* 1990; **117**: 827–9. [A randomized trial showing that topical mupirocin can be as effective as oral erythromycin for impetigo.]

Champion RH, Burton JL, Ebling FJG, Rook A, Wilkinson DS. *Textbook of dermatology*, 5th edn. Blackwell Scientific, Oxford, 1992. [A comprehensive dermatology textbook that provides further reading on most dermatology subjects.]

Fisher AA. Contact dermatitis in surgeons. *Journal of Dermatologic Surgical Oncology* 1975; **1**: 63–7. [A succinct but slightly dated review of contact dermatitis in surgical personnel.]

Kan K, Schulzer M, Chow A. Clinical spectrum of nonmenstrual toxic shock syndrome: comparison with menstrual TSS by multivariate discriminate analysis. *Clinical Infectious Disease* 1993; **16**: 100–6. [A discussion of toxic shock syndrome.]

Kanja LF, Wilking SVB, Phillips TJ. Pressure ulcers. *Journal of the American Academy of Dermatology* 1998; **38**: 517–36. [A comprehensive review that includes pathophysiology.]

Layden L, Wilson SE. Bacterial diseases of the skin. In: Rakel RE, ed. *Conn's current therapy—1987*, p. 660. Saunders, Philadelphia, 1987. [A review concentrating on therapy choices.]

Lever WF, Elder DE. *Lever's histopathology of the skin*, 8th edn, p. 461. Lippincott-Raven, Philadelphia, 1997. [A dermatopathology text with a relevant discussion of the pathophysiology of hidradenitis suppurativa.]

Mustoe TA *et al.* A phase II study to evaluate recombinant platelet-derived growth factor-BB in the treatment of stage 3 and 4 pressure ulcers. *Archives of Surgery* 1994; **129**: 213–19. [An article describing the results of a randomized trial suggesting that a platelet derived growth factor may be helpful in healing some ulcers.]

Rietschel RL, Fowler JF, Fisher AA. *Fischer's contact dermatitis*, 4th edn. Williams and Wilkins, Baltimore, 1995. [A comprehensive textbook with an excellent description of each antigen that may be problematic for surgeons.]

Rubin RH, Young LS. *Clinical approach to infection in the compromised host*, 3rd edn. Plenum, New York, 1994. [A text especially useful for its review of etiologic agents in skin disease in immunocompromised patients.]

Sanford JP *et al. The Sanford guide to antimicrobial therapy*. Antimicrobial Therapy Inc., Vienna VA, 1997. [A well referenced and updated guide for selecting appropriate antibiotic and antiviral therapies for skin infections.]

Sugarman B. Infection and pressure sores. *Archives of Physical Medical Rehabilitation* 1985; **66**: 177–9. [An article showing that pressure ulcers can be infected with bacteria, but that the infection may not be clinically obvious.]

Sussman GL, Beezhold D. Allergy to latex rubber. *Annals Internal Medicine* 1995; **122**: 415–18. [A review article on latex allergy.]

Tasche C, Angela J, Jayaram B. Surgical treatment of hidradenitis suppurativa of the axilla. *Plastic and Reconstructive Surgery* 1975; **55**: 559–62. [A review article with a description of the authors' surgical technique.]

38.7 Surgical therapies used in dermatology

Calvin O. McCall and Thomas G. Cropley

Excisional surgery
Electrosurgery
Cryosurgery
Mohs micrographic surgery
Laser therapy
Further reading

Excisional surgery

Although many therapeutic modalities are available to cutaneous surgeons, they most commonly utilize excisional surgery under local anesthesia. Biopsy or complete removal of elevated, superficial lesions can be accomplished rapidly by ‘shave excision.' However, this superficial procedure must not be used to excise nodular lesions or lesions with atypical pigmentation where malignant melanoma is in the differential diagnosis. In such cases the shave technique makes microscopic interpretation impossible. For most excisions an elliptical or fusiform procedure is preferable. To improve the cosmetic outcome, the ellipse excision is performed with the long axis oriented in the direction of the relaxed skin-tension lines. A border of normal skin should be taken around the lesion. The amount of normal skin excised depends on the goal of the procedure (biopsy or cure) and the nature of the lesion. When possible, the excision should extend into the subcutaneous fat for ease of closure and microscopic interpretation. Alternatives are an incisional biopsy or a ‘punch' biopsy utilizing a commercially available, tubular blade. With a diagnosis, the appropriate margin for excision can be planned.

Electrosurgery

Many surgeons use electrosection (cutting) and electrocoagulation in conjunction with scalpel surgery. Electrofulguration and electrodesiccation are techniques more commonly used by dermato-logists and dermatologic surgeons. In electrofulguration, the tip of the electrode needle is held a short distance above the skin. A spark jumps the gap from the electrode tip to the lesion's surface. This method produces very superficial tissue necrosis and is thus suitable only for superficial lesions such as seborrheic keratoses. Electrodesiccation requires the tip of the electrode to contact the cutaneous surface; this produces a hemispherical zone of necrosis deeper than that achieved with electrofulguration. Both techniques require removal of charred tissue with a curette or scalpel as treatment proceeds. Electrodesiccation with curettage is frequently employed in the treatment of warts, seborrheic keratoses, and other small, benign skin lesions. Malignant lesions, particularly basal-cell carcinomas, can also be treated with this technique. The inability to assess surgical margins for residual malignancy after this procedure makes it less desirable than excision in many cases; it is generally considered inadequate for the treatment of recurrent or sclerotic malignancies.

All electrosurgical procedures require adequate local anesthesia. Electrosurgery should be used with great caution in patients with cardiac pacemakers: although seldom a problem, the radiofrequency output of the electrosurgical units may be sensed by some older pacemakers, causing them to malfunction. Electrosurgical equipment should not be used in proximity to alcoholic solutions or to dry gauze because of enhanced flammability.

Cryosurgery

Cryosurgery is a destructive modality in which a cryorefrigerant, usually liquid nitrogen, is delivered to the cutaneous surface. It may be done with a cotton-tipped applicator, a spray, or a recirculating, closed probe. The extreme cold causes cell death as a result of intracellular ice crystal formation and electrolyte concentration.

Regardless of the technique used, cryosurgery may be employed in the treatment of benign and malignant lesions. Benign lesions such as warts or actinic keratoses are treated for 10 to 20 s. Basal-cell and squamous-cell carcinomas require longer treatment times, typically two or more freeze cycles of greater than 30 s each. Only the spray or cryoprobe units should be used for treatment of skin cancers, since the depth of ice formation with the cotton-tipped applicator is insufficient to treat invasive tumors. Treatment is followed by edema, erythema, and blister formation. Healing takes place by secondary intention; in the case of large skin cancers this may require several weeks. Treatment of superficial lesions leaves little or no scarring, but unsightly pigmentary changes may result, especially in patients with dark skin. Pigmentary changes and scars invariably result from treatment of skin cancers. For this reason, cryosurgery is not an appropriate treatment for cutaneous malignancies when the cosmetic outcome is important.

The principal advantage of cryosurgery over other treatments for benign lesions is the speed with which multiple lesions can be treated. Disadvantages in the treatment of malignancies exist: surgical margins cannot be evaluated for residual malignancy and this technique may not be adequate for ‘high risk' lesions (see below). Anesthesia is not generally required for treatment of small, superficial lesions, but larger tumors should be treated under local anesthesia.

Mohs micrographic surgery

This is a surgical technique whereby the radial and vertical extent of tumor growth can be mapped intraoperatively, allowing very precise extirpation of even very small foci of carcinoma. Unlike conventional vertical sections, the histologic sections used in the Mohs technique are cut parallel to the cutaneous surface and tangentially through the epidermis; this allows simultaneous visualization of the entire deep and lateral margins of the excision specimen. Excision of the tumor proceeds in stages: after initial debulking of the visible tumor, additional horizontal sections may be excised sequentially. Each section is examined for the presence of residual tumor at the deep or lateral margins and subsequent sections are taken of any residual foci. The procedure is complete when no residual tumor can be found at any margin.

Originally, the Mohs technique employed *in situ* chemical tissue fixation. In 1974, Tromovitch and Stegman described the freshtissue technique, which employs frozen histologic sections rather than *in situ* chemical fixation. This technique is now commonly employed for the treatment of ‘high risk' cutaneous malignancies: recurrent malignancies, aggressive basal-cell carcinomas, and carcinomas arising in areas at high risk of recurrence.

After successful removal of the tumor, conventional reconstructive surgical procedures are usually employed to repair the resultant defect. Large or complex defects are frequently treated by the combined efforts of the dermatologic surgeon, plastic surgeon, otolaryngologist, or neurosurgeon as appropriate.

Laser therapy

A number of lasers are used in dermatology. In general terms, a laser consists of an optical resonator, a power source, and an optical mechanism to direct the beam toward the target. Lasers are commonly named according to the lasing material within the optical resonator, which when energized produces light. The beam produced by a laser is a unique type of electromagnetic radiation that is monochromatic, coherent (i.e., all light waves are in phase), and highly collimated. The utility of a given type of laser depends primarily on the wavelength of light produced and the mode of delivering the light. The emitted light is characterized as continuous, pulsed, or Q-switched. Table 1 gives the optical characteristics and uses for some of the lasers currently available.

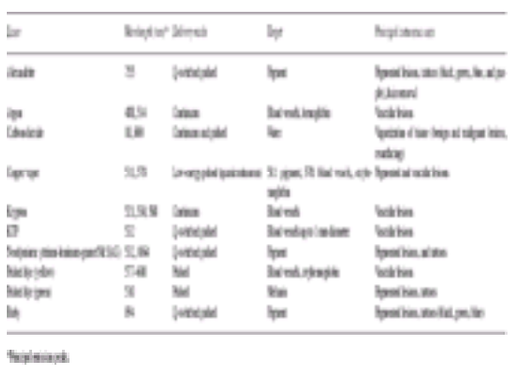


Table 1 Characteristics of lasers commonly used in dermatology

Carbon dioxide lasers are commonly employed in dermatologic surgery as well as in gynecologic and general surgery. The infrared light produced by these devices is absorbed by water, and the resulting tissue vaporization is non-selective. This type of laser is particularly useful for removal of warts, syringomas, epidermal nevi, and other superficial lesions. Healing takes place by secondary intention. The CO_2 laser may be used in a defocused mode to remove excess tissue with a sculpting or 'airbrush' technique, as is employed in the treatment of rhinophyma. The CO_2 laser invariably causes some degree of scarring, and, for that reason, its use in the treatment of port-wine stains and tattoos has largely become obsolete. A robotic scanning device that 'chops' the beam into extremely brief pulses allows precise removal of very thin layers of epidermis and dermis. Laser 'resurfacing' using this type of CO_2 laser is employed for cosmetic treatment of facial wrinkling and other manifestations of chronic sun damage.

Tuneable dye lasers are also commonly used in dermatology. They employ an organic dye as the lasing medium. The dye is stimulated to emit photons by 'pumping' it with an argon laser beam or a flashlamp. The output wavelength of dye lasers can be adjusted, or tuned, to match the absorption spectrum of target chromophores. The short pulse duration of the dye laser, which prevents thermal damage at distances greater than a few microns from the target, greatly minimizes the risk of scarring. Thus, it is now possible to treat port-wine stains and other superficial vascular malformations of the skin with minimal cosmetic risk.

Ruby lasers have also recently gained utility in dermatology. They were developed in the 1950s and until recently have had few legitimate medical applications. By coupling a ruby laser with an electronic mechanism know as a Q-switch, which compresses the output of the laser into brief pulses of extremely high energy, it is possible to destroy pigment granules in tattoos. In early trials, the Q-switched ruby laser produced clearing of up to 80 per cent of amateur tattoos and approx. 30 per cent of professionally applied tattoos. Tattoos composed of black carbon pigment (such as post-traumatic tattoos) respond particularly well; red areas in colored tattoos do not respond to this treatment. Most tattoos can be removed partially or completely without appreciable scarring. Hypopigmentation occurs in most patients, but usually resolves spontaneously after 6 to 24 months. Other Q-switched lasers have been developed and are used for treatment of colored tattoos (Q-switched neodymium: yttrium- aluminum-garnet laser).

Research in cutaneous laser therapy is currently very active. Lasers are being studied as light sources for photodynamic therapy.Lasers and other high-output light sources are also being studied in attempts to achieve permanent hair removal.

Further reading

Alster TS, Thiers BH, ed. Lasers in dermatology. *Dermatologic Clinics*, 1997; **15**: 355–534. [The authors present a review of laser biophysics and the characteristics of the most commonly used lasers.]

Boughton RS, Spencer SK. Electrosurgical fundamentals. *Journal of the American Academy of Dermatology* 1987; **16**: 862–7. [Applicable electrophysical fundamentals and safety considerations are presented in this review.]

Fitzpatrick RE. Laser resurfacing. *Advances in Dermatology*, 1997; **13**: 463–501. [This article is a general review of laser cutaneous resurfacing.]

Fitzpatrick RE, Goldman MP, Satur NM, Tope WD. Pulsed carbon dioxide laser resurfacing of photoaged facial skin. *Archives of Dermatology* 1996; **132**: 395–402. [This study demonstrates the efficacy of the pulsed $C0_2$ laser in resurfacing photodamaged facial skin.]

Mohs FE. Chemosurgery. A microscopically controlled method of cancer excision. *Archives of Surgery* 1941; **42**: 279–95. [Dr Mohs' original chemosurgical technique is presented.]

Shriner DL, McCoy DK, Goldberg DJ, Wagner RF, Jr. Mohs micrographic surgery. *Journal of the American Academy of Dermatology* 1998; **39**: 79–97. [This publication reviews Dr Mohs' technique as it is currently performed for the treatment of skin tumors.]

Taylor CR *et al.* Treatment of tattoos by Q-switched ruby laser. *Archives of Dermatology* 1990; **126**: 893–9. [The use of the Q-switched ruby laser for tattoo removal is discussed in this well illustrated article.]

Zacarian SA. Complications, indications, and contraindications in cryosurgery. In: Zacarian SA, ed. *Cryosurgery for skin cancer and cutaneous disorders*, p. 283. Mosby, St Louis, 1985. [The author presents the fundamentals of cutaneous cryosurgery for a variety of skin disorders.]

39.1 Genitourinary anatomy

Francis J. McGovern and John D. Seigne

The retroperitoneum
The adrenal glands
The kidney
The ureter
The bladder
The prostate gland
The urethra
The penis
The testicles and scrotum
Further reading

The retroperitoneum

The retroperitoneum is the area that lies between the posterior abdominal wall and the peritoneal cavity. Its superior boundary is the diaphragm; the inferior boundary is the pelvic floor. Anteriorly is the peritoneum, posteriorly the vertebral column and transversus abdominis muscles. The retroperitoneum contains the kidneys, adrenals, ureters, bladder, and prostate gland. The aorta and the vena cava with their branches lie on the vertebrae, as do the lymphatic channels, nodes, and the cisterna chyli (Fig. 1). The lumbar nerve plexus lies within the psoas muscle and the sacral plexus lies on the pyriformis muscle. The sympathetic nerves arise from the celiac plexus and sympathetic chain on the bodies of the lumbar vertebrae; the parasympathetic nerves arise from the vagus nerve and the sacral outflow.

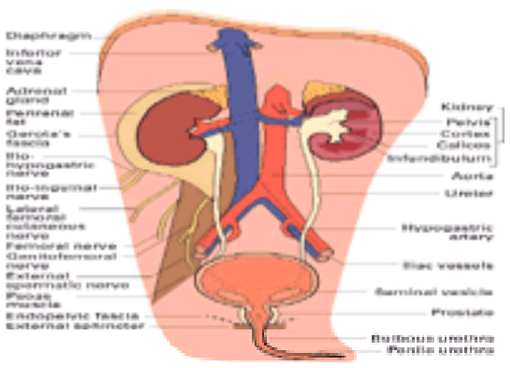


Fig. 1. The retroperitoneum: a cut-away view of the left kidney illustrates its cross-sectional anatomy.

The adrenal glands

The adrenals are paired endocrine glands found at the superomedial aspect of both kidneys. The right gland is triangular and the left is crescentic. The anatomical relations of the glands are the same as for the upper poles of their respective kidneys. Each adrenal consists of a cortex, which secretes glucosteroids, mineralosteroids, and sex steroids, and a medulla, which secretes adrenaline and noradrenaline. The adrenals receive an arterial supply from the aorta, and from the renal and inferior phrenic arteries. The left adrenal vein drains to the ipsilateral renal vein. The right adrenal has very short veins that go directly to the vena cava. Sympathetic fibres reach the adrenal medulla from the celiac plexus.

The kidney

The kidneys lie on the psoas and quadratus lumborum muscles. The left kidney (hilum opposite L1) is slightly higher than the right (hilum opposite L1–2), which is displaced inferiorly by the liver. Each kidney, and its adrenal gland, is surrounded by a layer of fat encircled by Gerota's fascia. Anterior to the right kidney are the bare area of the liver, the duodenum, and the ascending colon. Anterior to the left kidney lie the spleen, the posterior surface of the stomach, the pancreas, the descending colon, and the jejunum.

The renal artery is the first major lateral branch of the abdominal aorta. On the right this artery is approx. 3 cm long; it passes behind the inferior vena cava before branching in the renal hilum. The left renal artery passes directly from the aorta to the hilum. Thirty per cent of individuals have more than one renal artery, the most common variant being an aberrant upper-pole vessel. The renal veins lie anterior to the arteries. The left renal vein receives the adrenal vein, the gonadal vein, and one lumbar vein before crossing the aorta and draining into the inferior vena cava. The right renal vein is shorter and usually has no tributaries. The lymphatics drain to aortocaval nodes. Sympathetic and parasympathetic nerves pass from the celiac plexus to both kidneys and probably have a role in regulating renal blood flow.

The kidney consists of a cortex (site of the glomeruli) and a medulla (site of the collecting ducts). The collecting ducts drain via papillae into minor calices, which combine in groups of two or three to form one of six to seven major calices, which together make up the renal pelvis, the most posterior structure of the hilum (Fig. 1). Common congenital abnormalities of the kidney include unilateral agenesis, ectopia, horseshoe kidney, and a fused pelvic kidney (Fig. 2).

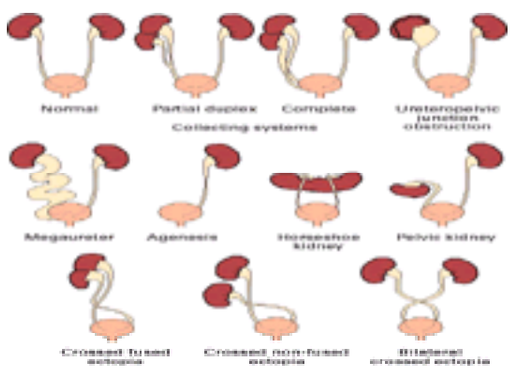


Fig. 2. A diagrammatic representation of the common renal and ureteric anomalies.

The ureter

The ureter is a narrow muscular tube that transmits urine from the renal pelvis to the bladder, a distance of about 25 cm. It runs through the retroperitoneum along the psoas muscle, at the tips of the lumbar transverse processes (an important radiological landmark). It crosses the pelvic brim at the bifurcation of the common iliac artery and lies on the levator ani muscle before swinging medially into the base of the bladder at the level of the ischial spine. Anteriorly, the ureter is adherent to the peritoneum for much of its length, and it can easily be injured during pelvic and abdominal surgery.

The ureter receives blood from the aorta, and from the renal, gonadal, common iliac, vesical, and uterine arteries. The lymphatics drain with each of these vessels. The somatic sensory nerves innervate the ureter in a segmental fashion, which explains why ureteric obstruction can give rise to referred ipsilateral testicular pain.

There are three anatomical points of narrowing within the ureter at the ureteropelvic junction, the site where it crosses the iliac vessels, and at the ureterovesical junction. Renal stones commonly impact at these sites.

A common congenital abnormality is duplication of the ureter, which may be complete or partial. When duplication is partial, the ureter from the upper-pole moiety joins the lower-pole ureter at any point from the renal pelvis to the bladder. When duplication is complete, the ureters cross, with the lower-pole ureter entering the bladder more laterally and having a tendency to reflux urine up to the kidney (Fig. 2).

The bladder

In men, the bladder sits on top of the prostate gland, in front of the rectum, behind the lower abdominal wall and pubic symphysis. In women the bladder lies on the perineal membrane and anterior to the uterus. The bladder is covered superiorly by loosely adherent peritoneum that allows it to expand upwards on filling. The ureters enter the bladder base posterolaterally and, together with the urethra, form a triangular area, the trigone. The bladder wall contains the smooth muscle known as the detrusor (Fig. 3).

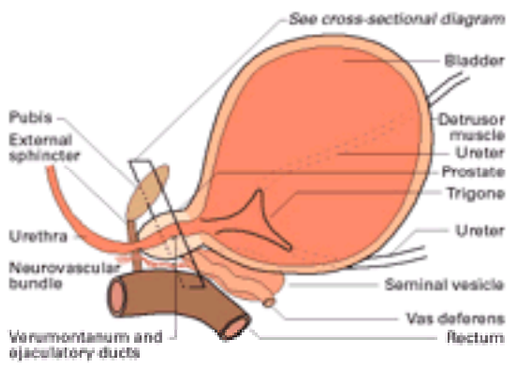


Fig. 3. A sagittal section through the pelvis at the level of the pubic symphysis; the level and plane of the cross-sectional diagram (Fig. 4) is shown.

The blood supply of the bladder is derived from the internal iliac artery via its superior and inferior vesical branches. The venous drainage is through the prostatovesical plexus to the internal iliac veins. The lymphatic drainage is to the obturator, internal, and external iliac nodes. The sympathetic nervous supply is from the hypogastric plexus (T5–L4), and is largely inhibitory to the detrusor and stimulatory to the bladder neck: this allows the bladder to expand and store urine. The parasympathetic supply originates from the sacral outflow (S1–5) and has the opposite effect on the bladder muscle, initiating and maintaining voiding.

The prostate gland

The prostate gland is a chestnut-sized organ, present only in the male, that lies below the bladder, nestled underneath the pubic symphysis. Immediately posterior to it lie the seminal vesicles and the ampullae of the ductus deferens, the combined excretory duct of which (the ejaculatory duct) passes through the prostate into the urethra. Separating these structures from the rectum is a fibrous sheath, derived from a fused extension of the peritoneum, known as Denonvilliers' fascia. The urethra passes through the centre of the prostate and is joined in its distal two-thirds by the ejaculatory ducts, which enter the urethra posterolaterally. These landmarks serve to divide the prostate into four anatomic zones. The peripheral zone, which is about 1 cm thick and extends around the posterior surface of the prostate, contains the prostatic capsule. This is the area in which most prostatic malignancies arise. Lying between the peripheral zone posteriorly and the urethra and ejaculatory ducts anteriorly is the central zone. The transitional zone lies in the superior part of the gland, around the urethra above the level of the ejaculatory ducts. This is the area that enlarges in benign prostatic hyperplasia. The area anterior to the urethra, below the level of the ejaculatory ducts, is known as the fibromuscular stroma (Fig. 4).

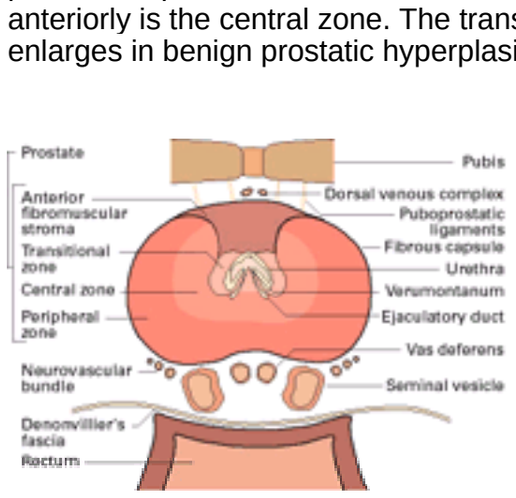


Fig. 4. A cross-section through the prostate to show its relation to the surrounding structures and its internal zonal anatomy. The level and plane of this diagram are shown in Fig. 3.

The arterial supply is from the inferior vesical artery. The prostate is surrounded by a plexus of veins that drain to the internal iliac system. The lymphatics drain to the obturator, external, and internal iliac nodes. The nervi ergenti (parasympathetic), which run to the penis and control erectile function, travel at the base of the prostate, with vessels, at the five and seven o'clock positions. These can be spared in a radical prostatectomy if one is trying to preserve potency (Fig. 4).

The urethra

The urethra conducts urine from the bladder to the outside. The female urethra is about 3 cm long and lies in the anterior vaginal wall. It is surrounded at its midpoint by the somatic external sphincter (supplied by the pudendal nerve S234), which, in combination with the internal sphincter at the level of the bladder neck, maintains continence.

The male urethra is divided into four segments. The prostatic urethra, extending from the bladder neck to the apex of the prostate, is the widest part of the urethra and is notable for a small protuberance on its posterior surface near its distal end, known as the verumontanum. This serves as an important endoscopic landmark indicating the most proximal aspect of the somatic external sphincter. The ejaculatory ducts and the prostatic glands drain into the floor of the urethra at the level of the verumontanum.

The membranous urethra, running between the apex of the prostate and the external perineal membrane, is 2 cm long and is surrounded by the somatic external sphincter (innervated by the pudendal nerve, S234). The prostatomembranous junction is the most common site of urethral disruption associated with pelvic trauma.

The bulbar urethra extends from the perineal membrane to the beginning of the pendulous urethra. It is surrounded by the bulbospongiosus muscle and is the area most often injured when a person falls astride an object.

The pendulous urethra is contained within the penis. Just proximal to the meatus (the narrowest part of the urethra) there is a slight dilatation: this area is known as the navicular fossa.

The penis

The penis consists of three expansile fibrovascular tubes: two corpora cavernosa and the corpus spongiosum. The two corpora cavernosa lie side by side along the

dorsal aspect of the penis. The corpus spongiosum runs along the ventral surface and contains the urethra. At the head of the penis the corpus spongiosum enlarges to form the glans. Both corpora cavernosa are surrounded by a tough fibrous sheet, the tunica albuginea, within which the fibrovascular core can expanded to produce an erection. The corpora are attached posteriorly to the ischial rami, where they are surrounded by the ischiocavernosa muscle that helps support and firm an erection. The two corpora cavernosa unite at the pubic symphysis and together with the corpus spongiosum form the penis (Fig. 5).

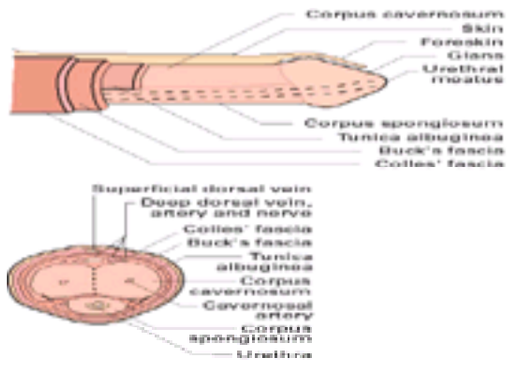


Fig. 5. A lateral cut-away view and cross-section of the penis to show the arrangement of the internal structures and their relation to the surrounding fascial layers.

The penis receives its blood supply from the pudendal artery, which is the terminal branch of the internal iliac artery. The superficial and deep dorsal veins of the penis pass through the suspensory ligament to the periprostatic plexus. The lymphatics pass to the inguinal nodes. The nerve supply is from the pudendal and pelvic plexus.

The testicles and scrotum

The testicles are paired gonads that arise on the posterior abdominal wall at the level of the kidneys during the 7th week of intrauterine life. They descend during gestation, preceded by a fibrous structure, the gubernaculum; the testes should be located in the scrotum at birth. As they pass through the anterior abdominal wall they take a covering from each of its three muscle layers, the most prominent being the internal oblique, which becomes the cremesteric muscle. A sleeve of peritoneum, the processus vaginalis, accompanies the testis on its journey through the inguinal canal. This sleeve is normally obliterated shortly after birth; if it persists an indirect inguinal hernia may form in its tract. The remnant of peritoneum covering the anterior surface of the testicle is called the tunica vaginalis; this may fill with fluid to form a hydrocele (Fig. 6).

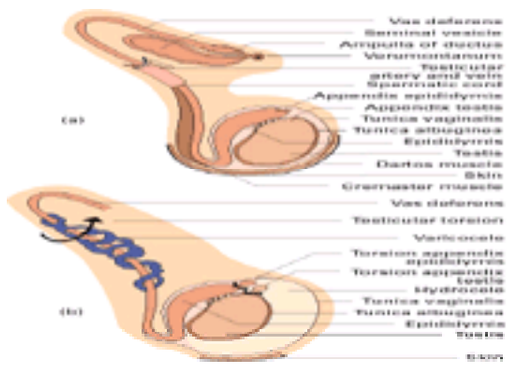


Fig. 6. (a) Cross-sectional view of a normal testicle and spermatic cord;(b) an illustration of some of the common scrotal abnormalities.

The testicle is surrounded by a thick fibrous sheath, the tunica albuginea. Inside each testicle there are approximately 400 seminiferous tubules, which produce sperm. These tubules drain to the epididymis, vas deferens, and, ultimately, the seminal vesicle and ejaculatory ducts.

The testicular artery arises from the aorta at the level of the second lumbar vertebra and travels through the retroperitoneum and the entire length of the cord to the testicle. The veins form a plexus in the scrotum (the pampiniform plexus), which usually unites at the level of the internal inguinal ring and drains into the vena cava on the right and the renal vein on the left. Dilation of the testicular veins of the pampiniform plexus is known as a varicocele (Fig. 6). The lymphatics drain to the para-aortic nodes, aortocaval nodes, and paracaval nodes at the level of the renal hilum.

The scrotum is a sack of thickened skin with muscle fibres in Colles' fascia, forming the rugose dartos muscle. It maintains the testicles at 1 to 2° below body temperature, allowing optimal spermatogenesis.

Further reading

Abrahams PH, Hutchings RT, Marks SC. *McMinn's color atlas of human anatomy*, 4th edn. Mosby, London, 1998.

Brooks J. Anatomy of the lower urinary tract and male genitalia. In: Walsh PC, Retik AB, Darractt-Vaughan E, Wein AJ, ed. *Campbell's urology*, 7th edn, pp. 89–128. Saunders, Philadelphia, 1997.

Kabalin J. Surgical anatomy of the retroperitoneum, kidneys and ureters. In: Walsh PC, Retik AB, Darractt-Vaughan E, Wein AJ, ed. *Campbell's urology*, 7th edn, pp. 49–88. Saunders, Philadelphia, 1997.

Williams P *et al.*, ed. *Gray's anatomy*, 38th edn. Churchill Livingstone, London, 1995.

39.2 Voiding function and dysfunction

Michael H. Safir, Angelo E. Gousse and Shlomo Raz

Introduction
Classification of voiding dysfunction
Functional anatomy
- Bladder
- Urethra

The 'voiding cycle'—normal filling and emptying
- Bladder filling
- Bladder emptying

Clinical evaluation of voiding dysfunction
- History
- Physical examination

Neurogenic urinary dysfunction
- Neurotransmitters
- Overview
- Spinal cord injury
- Sacral root disease
- Diabetes mellitus
- Brain injury
- Following radical pelvic surgery

Non-neurogenic voiding dysfunction
- Failure to store (bladder)
- Failure to store (outlet)
- Failure to empty (bladder)
- Failure to empty (outlet)
- Bladder neck dysfunction

Functional bladder testing
Conclusion
Further reading

Introduction

The continual cycle of bladder filling and emptying is a process which requires efficient and precise function of the lower urinary tract, which we define as the bladder and the urethra, as well as the neuromuscular and fascial support structures for these organs. Interruption of the voiding cycle by disease may have both systemic consequences (e.g. uremia or urosepsis) and important social implications, since urinary disability is generally perceived as a frailty of aging or chronic infirmity. Simply uncovering the disease is often difficult and may be veiled by both patient embarrassment and social stigma. Social perception has been cited a major factor relating delayed reports of voiding symptoms to the health care provider.

The economic significance of voiding dysfunction may be measured by analyzing total health care cost per disease entity. As an example, one decade ago in the United States, the total economic costs for incontinent patients was estimated to be $10.6 billion, the major costs of which were divided among purchasing caregivers' time and supplies, such as diapers. The personal and social costs are more difficult to estimate; however, the impact of freedom from diapers or from 'fear of accidents' may be dramatic and may greatly exceed the expectations of the physician.

The non-urologic surgeon will benefit from an improved understanding of voiding function and dysfunction for a variety of reasons, among them the following. (1) Operations such as abdominopelvic resection for colon cancer may result in significant voiding impairment, either as a consequence of neurologic or direct urinary organ injury. (2) Early postoperative voiding dysfunction is quite common and may be complicated by preoperative urologic disease. (3) Neuromuscular disease of the lower urinary tract, due to common neural pathways, may herald or progress in tandem with neurogenic disease of other pelvic viscera, such as colon or small bowel. (4) Trauma, either violent or intentional (e.g. therapeutic radiation), may result in global pelvic injury and dysfunction (e.g. radiation cystitis and radiation proctitis). (5) Pelvic surgeons should be familiar with operative techniques of urinary reconstruction, such that other regional surgery does not jeopardize or dismantle previous surgery.

The focus of this chapter is to provide an understanding of normal and pathologic lower urinary tract function and to present operative and non-operative treatment alternatives. We have separated neurogenic from non-neurogenic voiding dysfunction to clarify the topic of voiding dysfunction, particularly since neurogenic voiding dysfunction has historically been mired with confusing language and complex descriptions. We hope to provide a description of voiding function, both normal and pathologic, and present both operative and non-operative approaches to therapy. To illustrate topics, we have chosen to provide representative examples of diseases that may be encountered by surgeons, rather than providing exhaustive lists. As some common disease processes, such as benign prostatic hyperplasia, have been covered in detail elsewhere in this text, we would refer the reader to these sections when more detail is needed.

Classification of voiding dysfunction

We define voiding dysfunction as a general term which applies to any disorder of bladder filling or emptying and which may result from any category of disease: neoplastic, infectious, metabolic, traumatic, inflammatory, autoimmune, or psychogenic. Because of the breadth of this scope, classifying voiding dysfunction has been a daunting task, and has led to confusion both for the urologist and non-urologist. Some systems, such as the Krane–Siroky Urodynamic Classification, have distinguished disease based upon findings of functional filling and voiding studies (urodynamics). We believe that the easiest and most useful classification system simply distinguishes all forms of voiding dysfunction as either a failure to store urine or as a failure to empty urine, as proposed by Wein. Either type of failure may be due to dysfunction of the bladder or of the outlet, resulting in four major categories of voiding dysfunction. Most disorders may be individually classified from among these forms, although some disease processes may result from both classification entities.

Functional anatomy

Bladder

The bladder is the central organ of the lower urinary tract. Structurally it may be divided into four regions: dome, body, trigone, and neck. From a physiologic standpoint, it may more simply be divided into the bladder body and the bladder neck. Both the bladder body and neck are composed of three layers: urothelium (transitional cell epithelium), a smooth muscle layer, and a collagen-rich adventitial layer. Physiologically, anatomic differences between bladder body and outlet allow the bladder to serve both as a pump and as a reservoir. These essential and opposing functions may be understood by considering the differences in smooth muscle organization between the two regions. The smooth muscle of the bladder body (detrusor muscle) is not histologically partitioned into distinct layers. Instead, the myocytes possess multidirectional polarity, a property that is believed to allow the bladder to contract with synchronous reduction in all dimensions of the bladder lumen.

As a response to bladder filling, muscle fibers in the bladder body undergo elongation and flattening, together with lengthening of both elastin and collagen. In combination, these changes enable the healthy bladder to fill without a significant alteration in intravesical pressure, a property that we refer to as bladder compliance. Disease of the lower urinary tract, such as bladder outlet obstruction from benign prostatic hyperplasia, may initially result in muscular hypertrophy, coinciding with a relative reduction in collagen. While this adaptive response may initially be beneficial to allow bladder emptying, long-standing or severe bladder obstruction ultimately results in bladder wall fibrosis (increased collagen and decreased muscle), leading to (i) viscoelastic deterioration of the bladder wall and (ii) the inability of the bladder to store increasing urine volumes under low pressure (reduced bladder compliance). With continued obstruction, chronic detrusor fatigue impairs bladder emptying and may contribute to chronic urinary retention.

At the bladder neck, specialized smooth muscle orientation supports continence at rest and facilitates low-pressure voiding during micturition. Different from the heterogeneous muscular orientation in the bladder body, muscular orientation at the bladder neck assumes an organized latticework that has been likened to a camera shutter. Contraction of the bladder neck during filling promotes continence, while relaxation during micturition facilitates voiding. Although the normative bladder neck is closed at rest, competence of the bladder neck at rest is not a prerequisite for continence, since up to 20 per cent of normal continent females have an open, seemingly

patulous bladder neck during bladder filling.

Urethra

The male urethra is anatomically divided, proximally to distally, into five regions: prostatic, membranous, bulbar, penile, and fossa navicularis. The prostatic urethra, approximately 3 to 4 cm long, traverses through the prostate analogous to a drinking straw piercing through an orange. Voiding impairment attributable to benign prostatic hyper-plasia may, in part, be due to impingement on the urethral lumen by constrictive forces, both passive (bulky prostatic hyperplasia) and active (heightened smooth muscle tone of the prostatic capsule). The membranous urethra is surrounded by the bulk of the external urethral sphincter, which comprises predominately slow-twitch fibers. These fibers support ‘unaware' continence by providing relatively indefatigable contractions over long periods of time. On the other hand, less abundant fast-twitch fibers provide for immediate volitional interruption of the urinary stream. The regions of the urethra distal to the membranous urethra are responsible for passive conduction of urine and do not actively participate in continence.

The female urethra is short (approximately 4 cm) and consists of infolded urothelium enveloped in a rich estrogen-dependent vascular sponge which in turn is surrounded by a coat of smooth muscle and fibroelastic tissue (Fig. 1). Since many healthy women have an incompetent bladder neck at rest, continence may additionally depend upon: (i) adequate function of the external urethral sphincter, (ii) the ‘washer effect' (water-tight seal) contributed to by a well-estrogenized and healthy spongy portion of the urethra, and (iii) integrity of vesicourethral structures of support. The distal urethra, as in males, functions only as a conduit for urine.

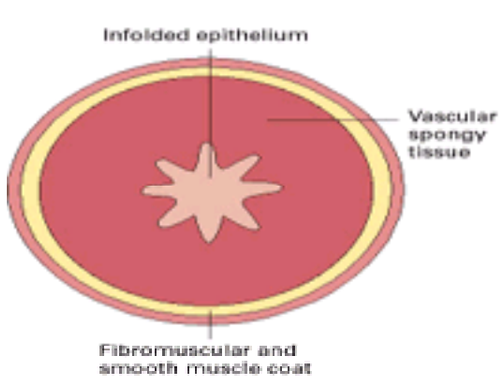


Fig. 1. The female urethra is deeply invested within estrogen-responsive vascular spongy tissue. The enriched sponge allows mucosal coaptation, which is believed to support continence.

The ‘voiding cycle'—normal filling and emptying

Bladder filling

A healthy adult bladder is capable of low-pressure urinary storage and must allow storage of at least a few hours of urine, during times of both normal and diuretic urinary production. Storage must initially be imperceptible to the patient and must gradually produce the urge to void without the precipitant need to empty the bladder. This sensation of fullness is transmitted though sacral nerve roots (principally S_3) to the spinal cord. Without necessary neural inhibition, a motor reflex arc is initiated, resulting in forceful contraction of the bladder. Normal childhood neural development allows inhibition of this reflex to avoid inappropriate and involuntary loss of urine. Since these inhibitory impulses are believed to originate in the cerebral cortex, loss of inhibition as a result of adult neurologic disease (e.g. stroke) may result in ‘re-emergence' of uninhibited reflex voiding, leading to urinary urgency or urge incontinence. The ability of the bladder to accommodate urine is also mechanically influenced by bladder compliance, defined simply as the ability of the bladder to accommodate increases in bladder volume without an increase in intraluminal pressure.

Bladder emptying

Bladder emptying ordinarily is a voluntary function that results from an organized cascade of neuromuscular changes in both the bladder and the bladder outlet. The first change observed during micturition is not bladder contraction, but voluntary relaxation of the pelvic floor musculature and synchronous relaxation of the external urethral sphincter. Following relaxation of the external sphincter, the trigonal musculature contracts, exerting opening traction on the bladder neck. Following this so-called ‘funneling' of the bladder neck, contraction of the detrusor finally occurs, resulting in an isometric bladder contraction and, ultimately, micturition, when the intravesical pressure begins to exceed bladder outlet pressure. In healthy patients, contraction simply occurs until the bladder is empty, after which the bladder outlet contracts and the detrusor relaxes in preparation for urine storage. In patients who have a poorly contractile bladder or in those with bladder outlet obstruction, voiding occurs only during the interval in which intraluminal bladder pressure exceeds outlet pressure. As an example, inappropriate contraction of the external sphincter during voiding, often seen among patients with spinal cord injuries and referred to as detrusor external sphincter dyssynergia (**DESD**), results in abrupt interruption of the urinary stream at the moment when outlet and detrusor pressures are isobaric. It is easy, in this sense, to understand why such patients are prone to the sequela of residual urinary stasis (urinary tract infection, bladder stones, etc.) and why potentially dangerous detrusor pressure may be transmitted to the kidneys.

Clinical evaluation of voiding dysfunction

History

At a minimum, evaluation of the patient with voiding dysfunction should consist of the patient's account of voiding dysfunction, focusing on duration of dysfunction, time course of onset, and factors that ameliorate or aggravate their symptoms. Certain descriptive elements may help to organize the patient's complaints: diuria (daily urinary frequency), nocturia (night-time frequency), urgency (immediate need to void), hesitancy (ability to void immediately at will), abdominal straining during micturition, change in force or caliber of stream, dysuria (painful voiding), intermittency of stream, and post-void residual sensation. To objectify symptoms, many urologists rely on symptom scores such as the American Urological Association's symptom index. The index was specifically designed to assess lower urinary tract symptoms in older males, although the resolution of this measure to distinguish among specific diagnostic entities has been challenged.

Individually, urinary symptoms do not target specific urologic diagnoses and, indeed, may not accurately identify urologic disease. Symptoms such as nocturia may be physiologic and therapeutic for the patient with congestive heart failure and lower extremity edema. Urinary urgency is common among widely disparate diseases, including urinary tract infection, detrusor hyperreflexia (uninhibited detrusor contractions resulting from neurologic disease), and detrusor instability (uninhibited detrusor contractions not resulting from neurologic disease).

As a symptom, urinary incontinence may be characterized according to the event that precipitates urine loss: sudden increase in intra-abdominal pressure (stress incontinence), leakage preceded by urinary urgency (urge incontinence), or leakage of both sorts (mixed incontinence). These forms of incontinence should be distinguished from overflow incontinence in which urine involuntarily spills from a full bladder, resulting from urinary retention associated with either bladder outlet obstruction (e.g. benign prostatic hyperplasia) or a poorly contractile bladder. When evaluating a patient with incontinence, it is commonplace to characterize pad usage (per day) and to clarify the relative severity of stress and urge leakage in patients demonstrating mixed incontinence.

Physical examination

A directed physical examination to uncover the etiology of voiding dysfunction should, at a minimum, include a thorough examination of the abdomen, a pelvic examination (rectal examination or vaginal/rectal), and a focused neurologic examination. Palpation and percussion of the suprapubic abdomen may reveal a distended bladder, which may typically be identified when bladder volumes exceed 150 ml. Examination of the genitalia may identify gender-related sources of voiding difficulty (e.g. urethral meatal stenosis in the male or urethral hypermobility in the female).

Valid initial impressions of neurologic function may be obtained even before initiating the medical interview. Simple observation may provide important clues to neurologic disease. Signs of dementia, tremor, and unilateral paresis may be observed and may provide evidence of existing or undiagnosed disease (e.g. stroke, multiple sclerosis, or Parkinson's disease) which individually may affect voiding function. Examination of the back may unusually reveal occult spinal dysrrhaphism, sometimes manifested cutaneously as a midline dimple overlying the spine (spina bifida occulta) or short gluteal cleft (sacral agenesis). Both of these segmentation

abnormalities of the spine, as well as other forms of spinal dysrrhaphism, may cause neurogenic voiding dysfunction, predominately due to impaired sacral root function.

When performing a directed physical examination, spinal cord or nerve root disease may also be suggested if dermatomal sensory loss is discovered. Since thoracolumbar (sympathetic) and sacral (parasympathetic) nerve root integrity promotes functional voiding, important dermatomal landmarks to examine include T10 (umbilicus), L1 (scrotum or labia), L5 (great toe), and S3 to S5 (perianal skin). Motor function of the lower extremities should also be examined for signs of spinal cord or spinal root disease. Gross motor function may be easily appraised by isolating lumbrosacrally innervated muscle groups, most importantly in the lower extremities: L4 to S1 (tibialis anterior), L5 to S2 (gastrocnemius), and L4 to S1 (toe extensors). The bulbocavernosus reflex, an external anal sphincter contraction elicited by squeezing on the glans penis (or clitoris), is a sacral reflex arc. Absence of the reflex may suggest sacral nerve injury, but may be absent in up to 20 per cent of normal females.

Neurogenic urinary dysfunction

Neurotransmitters

The voluntary process of bladder emptying and filling is governed by an actively debated and complex neurophysiologic model. From a functional standpoint, it is convenient to focus initially on neurotransmitters, which may be divided into cholinergic, adrenergic, and non-adrenergic non-cholinergic peripheral transmitters. Cholinergic (muscarinic-type) receptor sites in the lower urinary tract predominate throughout the bladder body and base. Stimulation of the preganglionic parasympathetic fibers of the pelvic nerve (S2 to S4) results in release of acetylcholine from postganglionic parasympathetic fibers and initiates a sustained bladder contraction, irrespective of whether the contraction was associated with volitional voiding or if it resulted from an uninhibited and involuntary reflex contraction (detrusor hyperreflexia).

Adrenergic receptors are sparser within the body and base of the bladder, and their role in the micturition cycle is more controversial. They are clustered about the bladder neck (more significantly in males) and are present within the prostatic capsule and prostatic smooth muscle. The suggested role of adrenergic neurotransmitters may be summarized by three clinical postulates: (i) a-adrenergic receptor activation results in increasing bladder outlet resistance via smooth muscle contraction at the bladder neck and within the prostate, (2) b-adrenergic receptor activation facilitates increasing accommodation of the bladder, and (3) adrenergic activation may impair bladder contractility via an inhibitory effect on parasympathetic ganglionic transmission. Non-adrenergic non-cholinergic peripheral neurotransmitters, such as enkephalin, neuropeptide Y, and somatostatin have all been demonstrated in cell bodies within the sacral parasympathetic nucleus and have also been implicated in mediation of the voiding cycle.

Overview

The topic of neurogenic urinary dysfunction includes a wide spectrum of diseases, ranging from the elderly patient who has had a stroke and complains of urinary frequency and urgency, to the patient who is quadriplegic and dependent on assisted voiding. Although many factors may impact on the type and severity of voiding dysfunction following neurologic injury or disease, location of the injury is most predictive. This insight has been clinically helpful in anticipating the type of urinary dysfunction based upon the site of disease. Tumors above the brain stem, for example, are likely to cause detrusor hyperreflexia, a pattern of dysfunction also seen with stroke at the same location. Clinically, patients with both diseases may complain of urinary frequency and urgency, as well as urge incontinence. Since the abundance of disparate neurologic diseases precludes a comprehensive review, we have chosen to consider common and representative diseases, the lessons of which may be applied to diseases at similar anatomic locations

Although treatment of each patient must be individualized, it is helpful to consider four principal goals of treatment for virtually any patient with neurogenic urinary dysfunction: (i) maintenance of urinary continence, (ii) preservation of kidney function via low-pressure storage and emptying of the lower urinary tract, (iii) avoidance of urinary tract infection, and most importantly (iv) initiation of a treatment plan that is attentive to the patient's goals and abilities. Despite the intent and thoughtfulness of a treatment plan, the success or failure of urologic intervention may be measured against satisfaction of these goals.

Spinal cord injury

Spinal cord injury is a devastating event that may have lifelong urologic consequences for the patient. In North America, injury to the cord occurs most commonly as a consequence of motor vehicle accident (40 per cent), falls (20 per cent), and gunshot wounds (13.6 per cent). Prior to the advent of improved assisted voiding techniques (e.g. clean intermittent catheterization of the bladder), renal failure was the leading cause of death following spinal cord injury, usually developing as a consequence of functional bladder outlet obstruction and recurrent urinary tract infection.

When acutely evaluating the patient with spinal cord injury, often the only necessary urologic intervention is provision of reliable bladder drainage, assuming the absence of direct urologic organ injury (e.g. bladder rupture, renal laceration). Lower urinary tract dysfunction in the acute setting of spinal cord injury takes the form of spinal shock, a period of impaired excitability of the cord at and below the level of injury. Because of disruption of both afferent and efferent neural pathways, the bladder remains acontractile and areflexic, leaving patients in urinary retention and dependent on assisted voiding techniques.

Following resolution of spinal shock, usually after approximately 6 weeks, there is a recovery phase that is heralded by the return of reflex sacral activity. In a simple sense, the urologic sequelae of spinal cord injury develop as a direct result of isolation of neurons below the site of injury from the brain and higher neurologic centers, making the lower urinary tract dependent on primitive, often dysfunctional, local reflexes to direct bladder filling and emptying. Although there may be maintenance of volume-dependent bladder contraction, bladder emptying may be incomplete due to inappropriate contraction of the external sphincter synchronous with detrusor contraction. This phenomenon, detrusor external sphincter dyssynergia (DESD), is the dominant voiding pattern following suprasacral spinal cord injury. Because a specialized portion of the midbrain (pontine micturition center) is responsible for orchestrating the finely tuned synergy of sphincter relaxation and detrusor contraction during healthy voiding, disruption of the spinal pathways that connect the center to the sacral roots leads to dysfunctional, uncoordinated voiding. Despite the simplicity of this model, other patterns of voiding dysfunction may occur following spinal cord injury, including ‘normal' voiding, detrusor hyperreflexia without DESD, and detrusor areflexia (absence of bladder contractions) (Table 1).

Spinal level	Normal (%)	Detrusor hypocontractility with DESD (%)	Detrusor hyperreflexia without DESD (%)	Detrusor areflexia (%)
Cervical	0	55	30	15
Thoracic	0	90	10	0
Lumbar	0	30	30	40
Sacral	12	12	12	64

DESD, detrusor external sphincter dyssynergia.
Urodynamic patterns are clustered by level of injury, although disparate findings among injured patients preclude accurately predicting an individual's voiding dynamics based solely upon level of injury.

Table 1 Spinal cord injury

It is difficult to predict long-term voiding ability among patients with spinal cord injuries. This remains as a frustrating realization in light of the important implications that voiding impairment has with regard to social and vocational reintegration. Some attempts have been made to refine this prognosis. In one large study, if perianal pinprick sensation was intact on presentation (S4 to S5), implying intact spinothalamic fibers (incomplete injury), 65 per cent of patients voided without assistance at 1 year, whereas no patient with absent pinprick sensation voided without assistance during this period. Once neurologic recovery has reached a plateau, patients require rigorous surveillance to assure continued low-pressure bladder filling (favorable compliance). Deterioration of bladder compliance may lead to high-pressure bladder filling, which has been established as an independent risk factor for renal impairment among patients with spinal cord injury.

For the majority of patients with spinal cord injury, clean intermittent catheterization is preferable to an indwelling catheter because of a lower rate of urinary tract infection with the former and beneficial bladder ‘rehabilitation' consequent to restoration of the storage and emptying cycle. Patients without the manual dexterity to perform self-catheterization are still suitable candidates if assisting nursing care is readily available.

Sacral root disease

Central herniation of lumbar disks is a useful model to understand sacral lesions such as cauda equina syndrome. Typically, lesions occur at the L4 to L5 vertebral level and may result in an acontractile bladder and partial or complete urinary retention due to compression of S2 to S4 nerve roots, most importantly the S3 root. Disease usually involves injury to both afferent and efferent limbs of the sacral reflex arc, such that the patient has both impaired sensation and poor bladder emptying. Alternatively, disk disease may also result in nerve irritation and resultant detrusor hyperreflexia. Other disorders of sacral nerve roots that may cause voiding dysfunction include cauda equina tumor, multiple system atrophy, tabes dorsalis, and viral sacral myeloradiculopathy (e.g. herpes simplex virus, cytomegalovirus, Epstein–Barr virus, and HIV).

Diabetes mellitus

Lower urinary tract voiding signs and symptoms attributable to diabetes mellitus have been collectively referred to as diabetic cystopathy. Diabetic cystopathy results from both a peripheral and an autonomic neuropathy and has historically been associated with long-standing or poorly controlled disease. Urodynamic studies on selected patients with diabetes have identified abnormal bladder function in approximately one-third of patients; however, only approximately 10 per cent are thought to be symptomatic. When symptomatic, the earliest alteration may simply be a gradual reduction in voiding frequency, due to neuropathy of stretch-sensitive afferent neurons. Since the changes are gradual, the manifestations of diabetic cystopathy may be clinically inapparent unless screened for by provocative history. The ‘endstage' of diabetic cystopathy has been clinically and urodynamically defined as including impaired bladder sensation, increased cystometric capacity, poor detrusor contractility, impaired flow, and large post-void residual. Other series of unselected symptomatic patients have challenged this generalization, citing a more significant frequency of detrusor hyperreflexia and frequent outflow obstruction among male diabetics. Whatever the cause, poor bladder emptying may put the patient at risk of recurrent urinary tract infection and obstructive renal insufficiency.

Therapy for the patient with diabetes is focused to enhance bladder emptying and includes double voiding (two closely timed trips to the washroom) and timed voiding (scheduled hours at which to void). Patients who do not thrive with these conservative measures, including those patients with frequent urinary tract infections or those with evidence of upper urinary tract deterioration from hydronephrosis, should be managed with clean intermittent catheterization. Pharmacotherapy, principally including cholinergic agents such as bethanechol chloride (acetylcholine altered to resist hydrolysis) has been disappointing.

Brain injury

Injury to the brain, as a consequence of stroke, tumor, or direct trauma, has the potential for significant deterioration of voiding function. Suprapontine injury is most common and will be focused on, since isolated midbrain (pontine) injury is uncommon and results in less predictable voiding impairment. Acute injury represents disruption of the corticopontine pathway, the integrity of which is believed to allow inhibition of detrusor contractions. The net effect may be an ‘uninhibited bladder', which clinically manifests as symptoms of urinary frequency and urge incontinence. Incontinence may be temporary and up to 41 per cent of patients may return to continence within 2 weeks. If incontinence persists, patients may benefit from a combination of both behavioral modification including fluid restriction and frequent timed voiding as well as anticholinergic medication, such as oxybutinin.

Following radical pelvic surgery

Marshall corroborated the anecdotal observation of postoperative urinary retention following abdominopelvic resection in 1946. A retrospective review of 600 patients undergoing colorectal surgery identified voiding dysfunction in 60 per cent of patients undergoing abdominopelvic resection and only 8 per cent of those who only underwent colostomy. More recent retrospective studies have demonstrated voiding dysfunction in approximately 50 per cent of patients who have undergone abdominopelvic resection. Injury may occur either during the abdominal or perineal portion of the operation. Injury from above can only occur below the rectal levator hiatus, since the main neural trunks to the bladder and urethra lie below the levator fascia. Perineal dissection may cause injury to the pudendal nerves as they lie on the undersurface of the ischium and pubic bone. Management includes temporary use of an indwelling(catheter with a voiding trial and decatheterization when thm patient is mobile. As i guideline, if a patient is unable to void 4 weeks following surgery, he should bm referred for urologic evaluation. Many of these patients ultimately will rely on intermittent self-catheterization for bladder emptying.

Non-neurogenic voiding dysfunction

Failure to store (bladder)

Despite a well-functioning urinary outlet, continence may be sabotaged by the inability of the bladder to function as a capacious, low-pressure reservoir. A healthy adult bladder is expected to fill physiologically to approximately 400 to 500 ml of urine without an appreciable increase in intravesical pressure and without the precipitous need to void. The ability of the bladder to store urine may be challenged in two major ways: involuntary bladder contractions and deterioration of bladder compliance.

Involuntary bladder contractions

An involuntary bladder contraction is defined by the International Continence Society as a sudden increase in detrusor pressure that is not volitional, regardless of whether the contraction or its sensory impetus is perceived by the patient. When symptomatic, patients may complain of the abrupt need to void and, often, the inability to reach the restroom before urinary leakage occurs (urge incontinence). When involuntary bladder contractions result from neurogenic urinary dysfunction, the condition is deemed detrusor hyperreflexia, while conditions unrelated to specific neurologic disease are referred to as detrusor instability. When detrusor instability occurs in men, it is usually a consequence of benign prostatic hyperplasia or other forms of bladder outlet obstruction (e.g. urethral stricture). In a consecutive series of patients with benign prostatic hyperplasia, Rosier and colleagues prospectively identified objective detrusor instability in 20 per cent of 185 patients. When encountered in female patients, it may coexist with stress urinary incontinence or may develop from obstruction caused by anti-incontinence surgery.

Deterioration of bladder compliance

Disease processes that result in bladder fibrosis and contracture may also challenge functional bladder capacity. Genitourinary tuberculosis initially produces irritative voiding symptoms, but in late stages may result in bladder fibrosis and contracture unless adequately treated with antituberculosis medications. Radiation cystitis is a common sequela of pelvic irradiation, affecting as many as 20 per cent of patients receiving external beam therapy for pelvic malignancy (e.g. bladder and rectal carcinoma). Most cases occur acutely and subside within a year of treatment, while the ‘late sequela', consisting of transmural bladder fibrosis due to endarteritis, may not present until 8 to 10 years following therapy. Although preventative agents such as sodium pentosanpolysulfate and therapeutic maneuvers such as hyperbaric oxygen have shown promise clinically, the mainstay of treatment for patients demonstrating characteristic frequency, urgency, and dysuria remains anticholinergic medication (e.g. oxybutinin) and behavioral modifications (fluid restriction, timed voiding, etc.).

Interstitial cystitis is an inflammatory disease of the bladder that is marked by subacute development of often disabling symptoms including urinary frequency, urgency, and painful bladder filling. The specific etiology of the disease, which usually occurs is women of 30 to 70 years old, has not been established; however, the disease is generally perceived to stem from disruption of the protective layer of glycosaminoglycans and type IV collagen underlying the mucosa of the bladder, allowing leakage of urine into the bladder wall. Alternative models of disease have implicated as-yet-unidentified infectious agents or autoimmune phenomena. Treatment is usually palliative and most commonly consists of hydrodistension of the bladder and intravesical administration of various topical agents (e.g. DMSO and heparin). Patients with severe bladder contraction may require augmentation cystoplasty or substitution cystoplasty.

Acute and chronic bladder inflammation (cystitis) may result in a broad variety of voiding dysfunction. Acute inflammation, as seen during acute bacterial cystitis, usually manifests as so-called irritative voiding symptoms, including urinary frequency, urgency, and dysuria. Chronic inflammation may also result in irritative symptoms, but may also result in reduction of functional bladder volume due to bladder fibrosis, such as is observed in diseases such as genitourinary tuberculosis and, more rarely, sporadic diseases such as eosiniphilic cystitis.

Failure to store (outlet)

The International Continence Society divides the definition of stress urinary incontinence into three categories: a symptom, a sign, and a condition. The symptom indicates the patient's statement of involuntary loss of urine during physical exertion. The sign denotes the observation of loss of urine from the urethra synchronous with physical exertion (e.g. coughing, laughing, or straining). Spontaneous urinary incontinence may also occur when the patient is seated or standing. Gravitational urinary loss during postural change may be observed in the most severe cases. The condition is the involuntary loss of urine occurring when, in the absence of a detrusor contraction, the intravesical pressure exceeds the maximum urethral pressure. Accurate diagnosis often requires urodynamic investigation in addition to

careful history and physical examination.

Stress urinary incontinence may occur in both male and female patients. In the male patient, stress urinary incontinence is most commonly iatrogenic and may be associated with prostate and bladder neck endoscopic or open surgery. In the female patient, stress urinary incontinence has a multifactorial etiology we will describe below.

Female stress urinary incontinence

Major risk factors for stress urinary incontinence in the female include parturition, laxity of the pelvic floor, neuropathic changes related to aging, and estrogen deficiency secondary to menopause. Other possible contributing factors include pelvic trauma, neurologic impairment, and prior failed anti-incontinence procedures. An estimated 30 to 65 per cent of women experience stress urinary incontinence with urgency and urge incontinence. Stress urinary incontinence accompanied with urge incontinence is commonly termed mixed urinary incontinence. Urethral hypermobility associated with pelvic floor laxity and intrinsic sphincter competency should be evaluated in all female patients with stress urinary incontinence.

Traditionally, the distinction between stress urinary incontinence existing as either urethral hypermobility or intrinsic sphincteric deficiency was important because it contributed to selection of the surgical procedure. Since many authorities currently believe that nearly all female patients with stress urinary incontinence have a component of intrinsic sphincteric deficiency, this clinical distinction is playing less of a role in dictating surgical treatment.

Urethral hypermobility, present in the majority of women with stress urinary incontinence, is rotational descent of the proximal urethra and bladder neck into the vagina (Fig. 2). Urinary leakage contributed to by hypermobility is associated with sudden increases in abdominal pressure attendant to activities such as coughing and straining. Urethral hypermobility can physically be observed by wide excursion of urethrovesical angle change determined by a Q-tip test or, radiographically, by a lateral standing cystogram with relaxing and straining views.

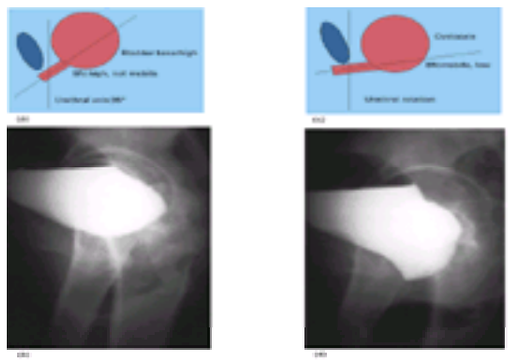


Fig. 2. Normal female anatomy places the bladder base and bladder neck in a ‘high' position behind the pubic symphysis (a, b). Loss of vesicourethral support, as is seen clinically among patients complaining of stress urinary incontinence, is associated with deterioration of bladder support (cystocele) as well as development of urethral hypermobility (c, d).

Intrinsic sphincter deficiency is also a significant component of incontinence in women. Intrinsic sphincter deficiency is defined as a deficiency in urethral sphincter function that is unrelated to the urethral support. The most common causes are prior incontinence surgery, prior pelvic surgery (radical hysterectomy, abdominoperineal resection of rectum), neurologic disorders, urethral mucosal atrophy related to estrogen deficiency, and radiation injury. Intrinsic sphincter deficiency results in poor urethral mucosal coaptation and stress urinary incontinence with minimal increase in abdominal pressure or gravitation.

Abdominal leak point pressure (**ALPP**) is a urodynamic diagnosis that is defined as the minimum intravesical pressure that produces leakage during a Valsalva-type maneuver. Performed in the standard fashion, as originally reported by McGuire, the ALPP reliably and reproducibly classifies patients based on the presence and magnitude of urethral sphincteric dysfunction. An ALPP less than 60 cmH_2O is consistent with intrinsic sphincter deficiency, 61 to 90 cmH_2O is in the ‘gray' zone, and an ALPP greater than 90 cmH_2O is not indicative of significant deficiency. ALPP developed and has now become popular as a result of the inability of the urethral pressure profile to identify sphincteric dysfunction reliably.

Surgical treatments for female stress urinary incontinence may be grouped into four major types of procedures: (i) retropubic suspensions (Burch, Marshall–Marchetti–Krantz, and Lapides laparoscopic bladder neck suspensions), (ii) transvaginal suspensions (Pereyra, Raz, Stamey, and Gittes bladder neck suspensions, (iii) anterior colporrhaphy, and (iv) sling procedures (pubovaginal sling using autologous fascia, cadaveric allograft fascia, anterior vaginal wall tissue *in situ*, and synthetic material as a graft). Intraurethral and/or periurethral bulking agents such as collagen, autologous fat, PTFE paste, and silicon have been used to treat stress urinary incontinence associated with intrinsic sphincter deficiency. The surgical details and illustrations of various surgical procedures used to treat female stress urinary incontinence are illustrated in Fig. 3.

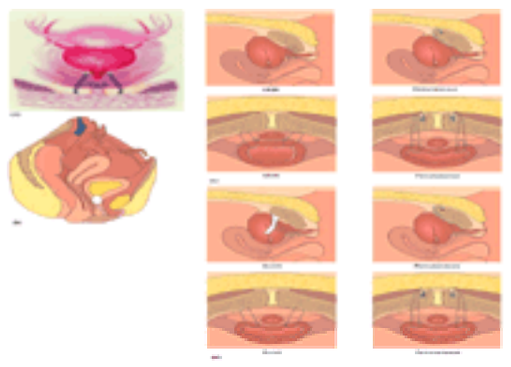


Fig. 3. Common types of operations to correct incontinence are shown. (a) Pubovaginal sling. (b) Transurethral injection of gluteraldehyde cross-linked collagen (Contigen™). (c) Marshall–Marchetti–Kranz procedure (MMK). (d) Burch colposuspension.

Recently, after an extensive meta-analysis of 282 screened articles using available data, Leach and colleagues summarized a report on the surgical management and outcome of female stress urinary incontinence. The panel found sufficient acceptable data (48 months and longer follow-up) to conclude that surgical treatment of stress urinary incontinence is effective, offering a long-term cure in a significant percentage of women. The evidence supports surgery as an initial therapy and as a secondary form of therapy after failure of non-operative treatments for stress urinary incontinence. Retropubic suspensions and slings are the most efficacious procedures for long-term success (in terms of cure/dry rates). Anterior colporrhaphy procedures are the least likely of the four major categories to be efficacious in the long term. It does appear that retropubic suspensions and sling procedures are associated with slightly higher complication rates, including postoperative voiding dysfunction such as urinary retention and associated irritative symptoms.

Non-operative management of female stress urinary incontinence includes biofeedback, behavioral modification, Kegel exercises, and pharmacotherapy with a-agonists or anticholinergic agents. Overall, the outcome results of these types of monotherapy are poorer than that of surgical treatments or bulking agents.

Male stress urinary incontinence

Non-iatrogenic stress urinary incontinence is uncommon in the male. Prostate surgery is the most common iatrogenic cause of this condition. While the incidence of stress urinary incontinence is uncommon after transurethral prostatectomy (less than 4 per cent), the incidence varies from 5 to 87 per cent after post-radical prostatectomy. Predisposing factors include neurologic disorders, radiation injury, pelvic trauma, and radical pelvic surgery. The male functional sphincter unit is divided

into: the proximal urethral sphincter, which includes the bladder neck to the level of the verumontanum; and the distal urethral sphincter, which includes the urethral muscularis, periurethral striated muscle, and the rhabdosphincter at the level of the external sphincter. Unlike the female, male stress urinary incontinence is not based on urethral hypermobility or pelvic floor relaxation but rather on direct injury to the sphincteric unit. Otherwise stated, all male stress urinary incontinence is secondary to intrinsic sphincter deficiency.

Transurethral or open prostatectomy removes the proximal ureteral sphincter mechanism. Therefore, postoperative continence relies on an intact distal urethral sphincter. During a transurethral resection of the prostate, resection distal to the verumontanum may damage the distal urethral sphincter, resulting in post-prostatectomy stress urinary incontinence. Maintaining continence after radical prostatectomy is related to surgical technique and other variables that are yet to be identified. Although bladder dysfunction may play a role in post-prostatectomy stress urinary incontinence, intrinsic sphincter deficiency is by far the more important contributing factor. Biofeedback, fluid restriction, anticholinergic agents, and a-agonists may alleviate stress urinary incontinence with a major bladder component and minimal intrinsic sphincter deficiency, but will not cure patients with extensive sphincteric incompetence. The current surgical options in the management of male intrinsic sphincter deficiency include: (i) transurethral or periurethral bulking agents (cure rate less than 50 per cent), (ii) insertion of an artificial urinary sphincter around the bulbar urethra (cure rate 70 to 80 per cent), and (iii) less commonly, placement of a male bulbar urethral sling or compression device (cure rate 50 to 60 per cent).

Failure to empty (bladder)

Isolated detrusor dysfunction leading to failure to empty, in the absence of neurologic disease, is an uncommon disease in young patients. When present, it is usually associated with hereditary muscle disease, chronic bladder infection, or mural fibrosis. Conversely, failure to empty is quite common in the elderly patient. Elbadawi and colleagues have described an association between poor clinical bladder emptying and axonal and muscular degeneration, superimposed on ‘normal' age-related histologic changes. The syndrome of detrusor hyperactivity with impaired contractility is believed to represent these changes in association with clinical instability due to alterations in intracellular junctions.

Failure to empty (outlet)

Despite an adequate sustained detrusor contraction, functional or anatomic obstruction occurring at any point from the bladder neck to the meatus may prevent bladder emptying. Failure to empty due to the urinary outlet is most commonly a disease of males, because of the length and complexity of the urethra. The most common cause of obstructive uropathy in the older male is benign prostatic hyperplasia and is described in detail elsewhere in this textbook.

Stricture disease

Urethral stricture disease develops as a maladaptive inflammatory response to urethral injury, most commonly from transurethral instrumentation, infection (gonococcal and non-gonococcal urethritis), and trauma. Clinically, the disease presents with obstructive voiding symptoms and urinary tract infection, most commonly seen when bladder emptying is impaired. The diagnosis is made using cystourethroscopy (Denmark pictures) and radiographic techniques such as retrograde urethrography and voiding cystourethrography (Fig. 4).



Fig. 4. Retrograde urethrography demonstrates discrete urethral stricture within the bulbar urethra. Fine annular strictures may be amenable to endoscopic incision, although open urethroplasty most reliably offers long-term cure.

Surgical treatment options for urethral stricture include urethral dilation with sounds, optical internal urethrotomy, placement of an intraurethral stent, or open urethroplasty. Urethral dilatation is performed with either the use of rigid, metal urethral sounds or specialized instruments called filiforms and followers. Optical internal urethrotomy has classically been performed under direct vision utilizing a ‘cold knife', but may be performed, with equal efficacy, utilizing the argon laser. Both optical internal urethrotomy and urethral dilation represent excellent means of achieving short-term ‘cure' of stricture and may be safely and easily performed.

Following the advent of endoscopic transurethral stricture incision (direct vision urethrotomy), the surgical pendulum swung toward this and other ‘minimally invasive' modes of treatment, also including urethral dilatation with sounds. However, concern has focused on the durability of these options. A recent, prospective, randomized study evaluating optical internal urethrotomy compared with urethral dilation identified equal efficacy for both types of treatment, with a 1-year recurrence of 40 per cent for short strictures (less than 2 cm) and 80 per cent for long strictures (greater than 4 cm). These modest results have returned the pendulum back toward open urethral reconstruction (urethroplasty) (Fig. 5) and new technologies, such as transurethral stent placement.

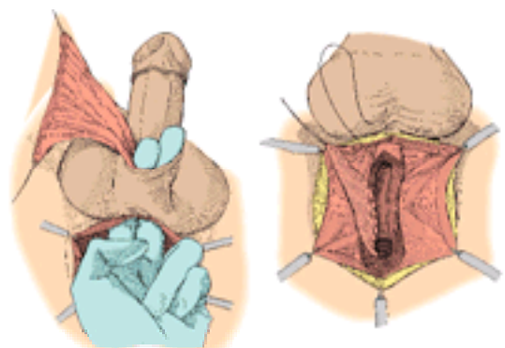


Fig. 5. A fasciocutaneous flap derived from penile skin may be employed to provide urethral width during urethroplasty. The flap may either be rolled into an interposition tube following resection of a segment of urethra or, more commonly, may simply be used as an onlay flap after incising the fibrosed urethra (shown).

Bladder neck dysfunction

The inability to time opening of the bladder neck properly is a well-described disorder that occurs most commonly in young and middle-aged men, but has also been described in women. Patients usually report both obstructive and irritative symptoms and have been labeled as possessing psychogenic dysfunction due to a perceived association of anxiety, normal prostatic examination, adequate emptying, and often young age. Video-urodynamic testing, a form of functional imaging of the lower urinary tract (see below), is the procedure of choice to evaluate this problem. During video-urodynamic testing, the diagnosis is corroborated by demonstrating elevation of voiding pressure, low urinary flow, and fluoroscopic level of obstruction at the bladder neck due to poor funneling of the bladder neck at the outset of voiding. Treatment is endoscopic incision of the bladder neck.

Functional bladder testing

The urodynamic study may be as simple as bedside ‘eyeball' urodynamics or as complex as software-driven multichannel studies. As a general rule, the ‘right' test is

the simplest test which may reliably reproduce the patient's complaint. However complex or simple the study, the urodynamic examination may be classified as one, some, or all of five individual tests, the combination of which is determined by the indication of the study. These tests include: filling cystometry, uroflow with or without pressure-monitoring, sphincter electromyography, urethral pressure profiles, and real-time fluoroscopy (video-urodynamics) (Fig. 6).

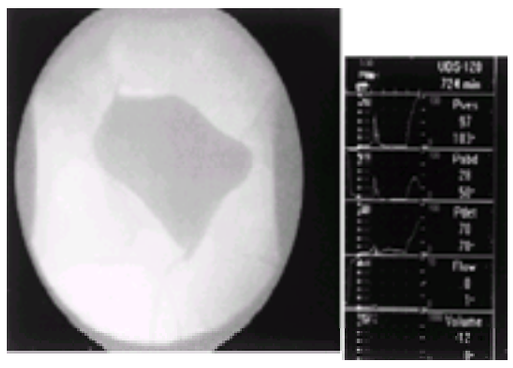


Fig. 6. Video-urodynamic evaluation offers real-time fluoroscopy of the lower urinary tract synchronous with transduction of bladder and urethral pressures. This image demonstrates normal funneling of the bladder neck just prior to the onset of micturition. The detrusor pressure (P_{det}) at this time of isometric contraction approaches 75 cmH_2O.

Filling cystometry focuses on the storage capability of the lower urinary tract and is accomplished by infusion of fluid (usually water, carbon dioxide gas, or radiographic contrast) into the bladder via a narrow transurethral catheter that is capable of simultaneous infusion of fluid and transduction of intravesical pressure (P_{ves}). P_{ves} itself represents the sum of intrinsic detrusor forces (P_{det}) and extravesical abdominal forces (P_{abd}), such as resulting from coughing or a Valsalva-type maneuver. Any involuntary detrusor contraction during a filling cystometrogram is abnormal and may represent either detrusor hyperreflexia or detrusor instability. Poor bladder compliance, a tonic increase in P_{det} during filling such as may be seen with neurogenic urinary dysfunction, may lead to upper urinary tract dysfunction unless overflow urinary leakage occurs at ‘safe' pressures (less than 40 cmH_2O). Uroflow testing provides information about urinary flow (Q) by indirectly measuring voiding volume while monitoring the duration of micturition. The most useful derivative of this study is Q_{max}, defined as the maximum urinary flow rate. The shortcoming of uroflow is that poor flow may represent either urinary outlet obstruction or poor detrusor function. Although solitary uroflow may be performed without a transduction catheter, uroflow with simultaneous transduction of intravesical pressure (utilizing a urodynamic catheter) may help to identify a high pressure–low flow pattern characteristic for outlet obstruction. The addition of real-time contrast fluorography (video-urodynamics) may help to enhance diagnostic utility by revealing functional disease of the urinary tract (e.g. urethral hypermobility).

Electromyography, in concert with other urodynamic tests, provides information regarding activity of the muscles of the pelvic floor, via either perineal needle electrodes or surface electrodes. The most useful application of electromyography is to aid in clarifying the diagnosis of detrusor external sphincter dyssynergy, which is characterized by inappropriate electromyography activity during detrusor contraction.

Clearly, not all patients require such testing, but it should be strongly considered in patients who: (i) have suspected or clinically progressing neurogenic urinary dysfunction, (ii) have specific complaints (e.g. stress incontinence) that cannot be reproduced in the office, (iiii) have undergone previous anti-incontinence surgery, and (iv) may not have a solid diagnosis based upon less complex studies.

Urodynamics have evolved into elegant physiologic testing of voiding dysfunction based, in part, on refinement of equipment, but more so based upon the insights of physician-scientists who have revolutionized our understanding of voiding pathophysiology. Although adjunct testing is a powerful tool, it is properly used to corroborate or dispute a clinical diagnosis, not as a substitute for a diligent routine urologic assessment.

Conclusion

The focus of surgical care may often draw focus away from significant, but often overlooked, patient health issues. However, it is relatively simple to identify bladder health issues by developing in your repertoire a routine of simple questions and elementary examination skills. Most diseases of voiding are readily treatable after consideration of a few diagnostic possibilities. Due to the prevalence of voiding dysfunction and the implications of aging, it is important for the surgeon to be aware of common urologic diseases and general paradigms of diagnosis and treatment.

Further reading

Beck R, Fowler CJ, Mathias CJ. Genitourinary dysfunction in disorders of the autonomic nervous system. In: Rushton DN, ed., *Handbook of neuro-urology*, pp. 281–301. Marcel Dekker, New York, 1994.

Becker HC, Miller J, Noske HD, Klask JP, Weidner W. Transurethral laser urethrotomy with argon laser: experience with 900 urethrotomies in 450 patients from 1978 to 1993. *Urologia Internationalis* 1995; **55**: 150–3.

Benoit G, Pelhuche A, Motyka G, Charpentier B. Urodynamic study of patients awaiting a simultaneous pancreas and kidney transplant. *Transplant Proceedings* 1994; **26**: 289–95.

Blaivas JG, Zayed AAH, Labib KB. The bulbocavernosus reflex in urology: a prospective study of 299 patients. *Journal of Urology* 1981; **126**: 197–9.

Brocklehurst JC, Andrews K, Richards B, Laycock PJ. Incidence and correlates of incontinence in stroke patients. *Journal of the American Geriatric Society* 1985; **33**: 540–2.

Buck AC, Reed PI, Siddiq YK, Chisholm GD, Russel-Fraser T. Bladder dysfunction and neuropathy in diabetes. *Diabetologia* 1976; **12**: 251–8.

Burney RE, Maio RF, Karunas R. Incidence, characteristics, and outcome of spinal cord injury at trauma centers in North America. *Archives of Surgery* 1993; **128**: 96–9.

Chancellor MB, Rivas DA. The American Urological Association symptom index for women with voiding symptoms: lack of index specificity for benign prostatic hyperplasia. *Journal of Urology* 1993; **150**: 1706–9.

Dean R J, Lyton B. Urologic complications of pelvic irradiation. *Journal of Urology* 1978; **119**: 64–70.

Dionko AC, Koppenhoefer R. Bethanechol chloride in neurogenic bladder dysfunction. *Urology* 1976; **8**: 455–8.

Dixon J, Gosling J. Structure and innervation in the human. In: Torrens M, Morrison T, eds, *The physiology of the lower urinary tract*, pp. 3–22. Springer Verlag, London, 1987.

Elbadawi A, Yalla SV, Resnick NM. Structural basis of geriatric voiding dysfunction. II. Aging detrusor: normal versus impaired contractility. *Journal of Urology* 1993; **150**: 1657–67.

Gow JG, Barbosa S. Genitourinary tuberculosis: a study of 1117 cases over a period of 34 years. *British Journal of Urology* 1984; **56**: 449–53.

Griffiths DJ, Harrison G, Moore K, McCracken C. Variability of post-void residual urine volume in the elderly. *Urological Research* 1996; **24**: 23–6.

Hampson SJ, Woodhouse CRJ. Sodium pentosanpolysulfate in the management of hemorrhagic cystitis: experience with 14 patients. *European Urology* 1994; **25**: 40–8

Hoyle CHV, Lincoln J, Burnstock G. Neural control of pelvic organs. In: Rushton DE, ed., *Handbook of neuro-urology*, pp. 1–54. Marcel Decker, New York, 1994.

Jeter KF. The social impact of urinary incontinence. In: Raz S, ed., *Female urology*, pp. 80–6. WB Saunders, Philadelphia, 1996.

Kaplan SA, Te AE, Blaivas JG. Urodynamic findings in patients with diabetic cystopathy. *Journal of Urology* 1995; **153**: 342–4.

Krane RJ, Siroky MB. Classification of neurogenic disorders. In: Krane RJ, Siroky MB, eds, *Clinical neurourology*, Ch. 15. Little Brown, Boston, 1979:308–30.

Marshall VF. Incidence of voiding impairment following abdominopelvic resection. *Journal of Urology* 1946; **55**: 324–8.

McGuire EJ, Woodside JR, Borden TA, Weiss RM. The prognostic value of urodynamic testing in myelodysplastic patients. *Journal of Urology* 1981; **126**: 205–9.

Parsons CL, Housley T, Schmidt JD, Lebow D. Treatment of interstitial cystitis with intravesical heparin. *British Journal of Urology* 1994; **73**: 504–7.

Perez-Marrero R, Emerson CE, Maharajah DO, Juma S. Prolongation of response to DMSO by heparin maintenance. *Urology* 1993; **41**: 64–6.

Peterson NE. Eosinophilic cystitis. *Urology* 1985; **26**: 167–71.

Plunkett JM, Braren V. Clean intermittent catheterization in children. *Journal of Urology* 1979; **121**: 469.

Rohner TJ Jr, Rohner JF. Urinary incontinence in America: the social significance. In: O'Donnell PD, ed., *Urinary incontinence*, pp. 4–6. Mosby, St. Louis, 1997.

Rosier PF, de la Rosette JJ, Wijkstra H, Van Kerrebroeck PE, Debruyne FM. Is detrusor instability in elderly males related to grade of obstruction? *Neurourology and Urodynamics* 1995; **14**: 625–33.

Shenot PJ, Rivas DA, Watanabe T, Chancellor MB. Early predictors of bladder recovery and urodynamics after spinal cord injury. *Neurourology and Urodynamics* 1998; **17**: 25–9.

Sirls LT, Choe JM. The incontinence history and physical examination. In: O'Donnell PD, ed., *Urinary incontinence*. Mosby, St. Louis, 1997:105–10.

Stormont TJ, Suman VJ, Oesterling JE. Newly diagnosed bulbar urethral strictures: etiology and outcomes of various treatments. *Journal of Urology* 1993; **150**: 1725–30.

Susset JG, Regnier CL. Viscoelastic properties of the bladder and urethra. In: Yalla SV, McGuire EJ, Elbadawi AE, Blaivas JG, eds, *Neurourology and urodynamics*. Macmillan Publishing Company, New York, 1988: 580–5.

Turner-Warwick R. A urodynamic review of bladder outlet obstruction in the male and its clinical implications. In: Turner-Warwick R, Whiteside G, eds, *Symposium on Clinical Urodynamics: the Urologic Clinics of North America*, vol.6, p. 171. Saunders, Philadelphia, 1979.

Warren JW. Interstitial cystitis as an infectious disease. *Urologic Clinics of North America* 1994; **21**: 31–9.

Wein AJ. Neuromuscular dysfunction of the lower urinary tract. In: Walsh PC, Retik AB, Stamey TA, Vaughn ED, eds, *Campbell's urology*, p. 576. WB Saunders, Philadelphia, 1992.

Weiss JP, Boyland FP, Mori H. Treatment of radiation-induced cystitis with hyperbaric oxygen. *Journal of Urology* 1985; **134**: 352–60.

Wilson CB, Leopard J, Nakamura RM, Cheresh DA, Stein PC, Parsons CL. Selective type IV collagen defects in the urothelial basement membrane in interstitial cystitis. *Journal of Urology* 1995; **154**: 1222–6.

39.3 Trauma to the genitourinary tract

David M. Rudnick and Alex F. Althausen

Urethra
Anterior urethral trauma
Posterior urethral trauma
Bladder
Symptoms and signs
Diagnosis
Treatment
Ureter
Diagnosis
Treatment
Kidney
Penetrating injury
Non-penetrating injury
Radiographic diagnosis
Treatment
Injuries to the external genitalia
Blunt trauma
Penetrating injuries
Burn injuries
Further reading

Genitourinary trauma is rarely an isolated event, and as such must be diagnosed and managed in the context of the injuries overall. Basic principles of trauma management must be followed, including securing a patent airway, maintaining respiration and perfusion, and evaluating neurologic injury. In the severely injured patient with multisystem trauma, the urologic organs in their hidden retroperitoneal location can be overlooked initially, but if not attended to may be a source of significant short- and long-term morbidity, and even death. The trauma surgeon must therefore always keep in mind the possibility of urologic injury, and look for signs, symptoms, and associated injuries, which will lead to timely diagnosis and management.

As a rule the genitourinary tract should be evaluated from the urethral meatus to the kidney; that is, in a retrograde manner. We will consider each part of the urinary tract in succession, discussing mechanism of injury, signs and symptoms, and evaluation and management of the possible urologic injuries.

Urethra

Anterior urethral trauma

Iatrogenic manipulation, straddle injuries of the perineum, and wounds from gunshots or antipersonnel devices are the most common causes of trauma to the anterior urethra. Trauma to the bulbous urethra usually results from a fall astride an object, causing the perineum and the underlying urethra to be crushed against the inferior edge of the pubic symphysis. The most frequent lesion is contusion and hemorrhage within Buck's fascia. If this fascia is torn, there is subcutaneous extravasation of blood and urine, which is enclosed within Colles' fascia (Fig. 1). This extravasation cannot extend out to the thighs because of the firm attachment of the fascia to the ischiopubic rami. The extravasated fluids travel downward into the scrotum and upward around the root of the penis. They then extend further upward beneath Scarpa's fascia, the superficial fascia of the abdomen (Fig. 2).

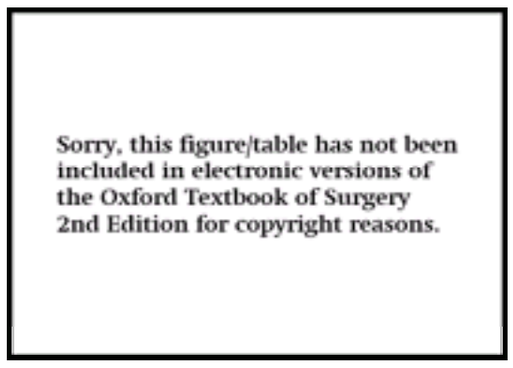


Fig. 1. Bulbar urethral tear resulting from straddle injury.

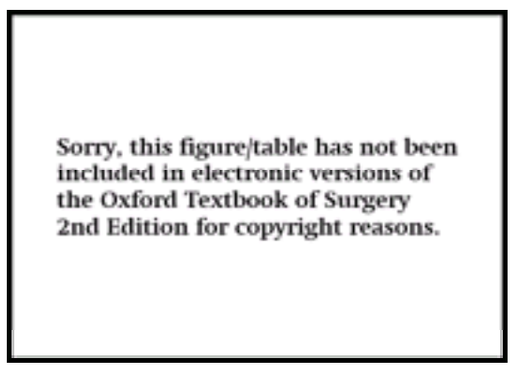


Fig. 2. Extravasation plane of bulbar urethral tear.

Diagnosis

Pain and swelling at the site of injury, with ecchymosis extending along the previously described fascial planes, are the rule. Blood at the urethral meatus and difficulty in voiding are also common findings. When anterior urethral injuries are severe, voiding may produce marked pain and increase the swelling in the perineum and scrotum as urine and blood collect in the periurethral tissues.

A retrograde urethrogram with a water-soluble contrast medium should be performed prior to further manipulation; this can be accomplished under gentle pressure using a Brodney penile clamp or a 12F Foley catheter inserted just far enough into the urethra to allow the catheter balloon to be inflated with 3 ml of fluid. The catheter will occlude the urethra at the fossa navicularis. The degree of injury is usually commensurate with the extent of extravasation of contrast material. However, if a urethral catheter has been previously passed without prior radiographic determination of the urethral injury, a pericatheter urethrogram can be performed (Fig. 3).

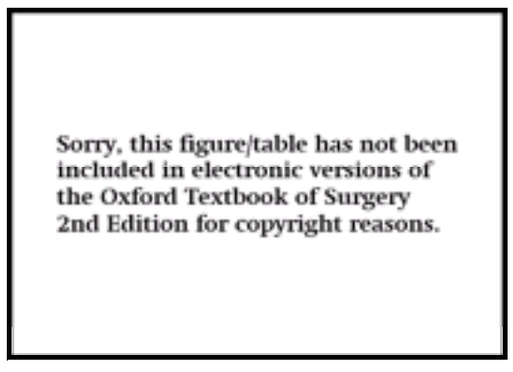


Fig. 3. Pericatheter urethrogram.

Treatment

Once the diagnosis of anterior urethral injury has been made, the initial management is based on the prevention of infection, extravasation of blood and urine, and loss or displacement of the urethral lining. If not attended to, any of these, alone or in combination, may cause subsequent stricture.

If urethrography demonstrates only contusion, observation is all that is required. The perineal, scrotal, and penile hematoma and ecchymoses are usually resorbed without the need for incision and drainage. The dysuria and urethral bleeding dissipate in several days. However, delayed strictures do occur in patients in whom the urethral lining was demonstrably intact at the time of initial injury. The patient should therefore be re-evaluated after 6 to 12 months, even if there are no symptoms of outlet obstruction.

Anterior urethral injuries with evidence of partial separation and minimal xtravasation should be treated initially with drainage by urethral catheter and broad-spectrum antibiotics. The catheter is left in place for 3 weeks to allow the laceration to line itself with new epithelium. It should be coated with an inert material such as Silastic, and should be no larger than a 14F for an adult. A larger catheter may not allow adequate drainage of the extravasated fluids around it. If the periurethral tissues become infected or if the swelling increases, drainage and debridement are necessary. Anterior urethral injuries with separation of the urethra or significant blast injury require perineal incision and drainage. The transected urethra is then debrided, mobilized, and obliquely spatulated. The urethra can be mobilized proximally and distally, allowing defects of 1.5 to 2 cm to be approximated without tension. Primary urethral anastomosis is carried out over a 24F catheter using fine, absorbable sutures (Fig. 4). The catheter is then removed and a 14F Foley catheter is replaced for drainage. Insertion of a suprapubic tube further aids urinary drainage and removes the need for a catheter change if the Foley fails to function. The urethral catheter is removed in 10 to 14 days and a voiding cystourethrogram is done through the suprapubic tube to check for fistulas. If healing is incomplete the cystotomy tube should remain in place for another week without replacing the urethral catheter. When primary reapproximation is impossible because of the length of injured urethra, tube grafts, scrotal inlays, or other urethroplastic procedures can be used. One-stage operations are ideal, but multistaged procedures are occasionally required. If the patient is critically ill from associated injuries the initial treatment is placement of a suprapubic cystotomy tube, with perineal drainage if the extravasation appears significant. Elective surgery may be carried out 3 to 4 months later, after the tissue reaction has bated.

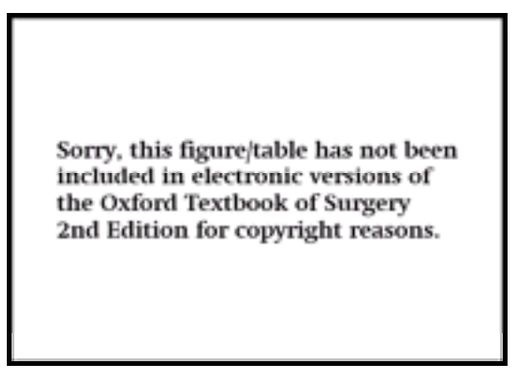


Fig. 4. Repair of bulbar urethral tear.

Posterior urethral trauma

Most injuries to the posterior urethra are the result of pelvic fractures sustained during motor vehicle accidents. Gunshot and stab wounds account for less than 10 per cent of these injuries in the United States of America. Five per cent of patients with pelvic fractures have urethral injury; this figure rises to 50 per cent when there is disruption of the pelvic ring. Fractures of the pubic rami and symphysis cause rupture of the puboprostatic ligaments. The tremendous force needed to cause the fracture of the pelvic ring also causes shearing of the pelvic organs, the most fragile of which is the posterior urethra. The shearing of the bladder and prostate against the fixed genitourinary diaphragm causes a partial or complete tear of the membranous urethra (Fig. 5).

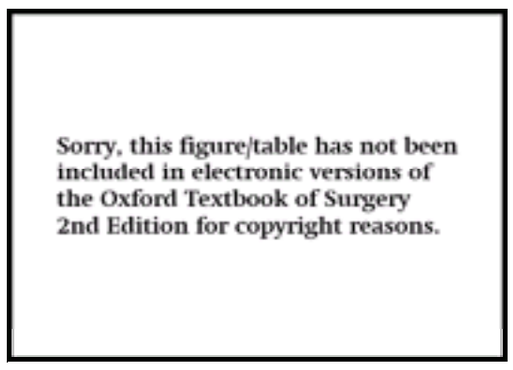


Fig. 5. Posterior urethral injury.

Diagnosis

The injury should be suspected in any patient with a pelvic fracture. He may be unable to void, and may present with blood at the urethral meatus and a palpable bladder. Digital rectal examination should be performed to rule out an associated rectal injury. The prostate often is not palpable within the pelvic hematoma. A retrograde urethrogram with water-soluble contrast will show periprostatic extravasation and no contrast within the bladder if the urethral tear is complete. An incomplete rupture will show some contrast medium within the bladder as well as periprostatic or periurethral extravasation. A cystogram per suprapubic cystotomy is necessary to allow full evaluation of the bladder. In patients with complete posterior urethral disruption there is a typical teardrop deformity of a high-riding bladder caused by a perivesical accumulation of blood. Urine will also extravasate, if the bladder neck is not in spasm. Since this mechanism of injury also frequently leads to injury of the body or neck of the bladder, it is essential to complete the evaluation of the lower urinary tract even when a catheter cannot be passed retrograde.

Treatment

The treatment of posterior urethral injuries remains controversial. It is our opinion that a urethral catheter should not be passed when a rupture is suspected or established. The insertion of a Foley catheter can convert an incomplete tear into a complete rupture as well as introduce infection. It can also increase the gap between the apex of the prostate and the tear, thereby extending the urethral stricture that almost always results. Initial treatment should consist only of the placement of a large, suprapubic cystostomy tube. Drainage of the retropubic space is not recommended. Some centers advocate placement of an aligning Silastic catheter, with antegrade and retrograde endoscopy as needed during the open placement of a suprapubic cystostomy. This catheter is not placed on traction or even intended to be a drain, but merely an alignment guide for the prostate to descend on to as the pelvic hematoma is reabsorbed.

If the rupture is incomplete the suprapubic tube may be clamped in 2 weeks to initiate voiding, if an antegrade cystourethrogram through the suprapubic tube shows no extravasation. The tube is removed after 3 weeks, when urothelialization of the injury is usually complete. This kind of injury may heal without stricture formation. When the rupture is complete, various methods of urethral anastomosis, with or without partial pubectomy, or scrotal inlay procedures (described in standard urologic texts) can be carried out 6 months after the initial injury. This time lag allows the extent of the stricture to declare itself and periurethral induration to subside. Primary repair of the urethra at the time of injury produces an unacceptable rate of postoperative stricture and erectile impotence, and often is not practical in a multiply injured patient with a pelvic hematoma. The impotence rate following primary repair is 30 per cent; in patients undergoing delayed repair it approaches 10 percent. The stricture rate after primary urethral realignment is as high as 75 per cent in most reports and urinary stress incontinence affects 25 per centof patients; following a delayed procedure these figures drop to 10 and 15 per cent, respectively.

Bladder

Blunt trauma causes 75 per cent of bladder injuries; penetrating missiles cause the remainder. A direct blow over a full bladder may cause it to rupture as a result of a sudden increase in intravesical pressure. If there is a tear in the anterior bladder wall an extraperitoneal rupture results in the extravasation of urine into the paravesical tissues (space of Retzius) (Fig. 6(A)). When the tear is in the posterior wall, retrovesical extravasation occurs (Fig. 6(B)). If the rent in the bladder is at the dome, the urine passes into the peritoneal cavity (Fig. 6(C)). Anterior perforation of the bladder by bony spicules may occur in patients with severe pelvic fractures (Fig. 6(D)).

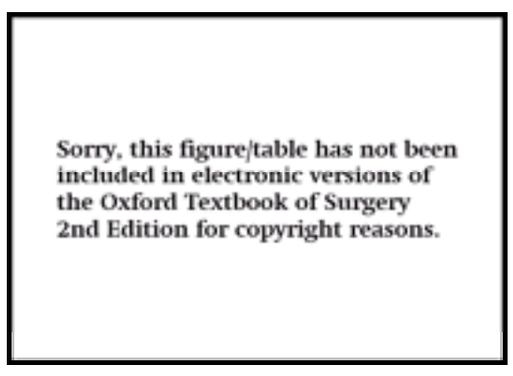


Fig. 6. Blunt trauma to the bladder with rupture.

Fractured pelvis is the injury most commonly associated with a ruptured bladder (at least 80 per cent). There is no apparent correlation between the type of bladder injury and the type of pelvic fracture. However, the greater the compression or disruption of the pelvic ring the more likely are bladder and urethral injuries. Only 10 per cent of patients with pelvic fractures sustain significant bladder injuries and 5 per cent have posterior urethral damage. It is not unusual for intra-abdominal injury to be associated with bladder rupture. Penetrating missiles involve the colon, rectum, or small bowel in at least 80 per cent of patients. About 40 per cent of patients with injury to the bladder have also sustained injuries to the liver, spleen, and colon.

Complications arise mainly from the associated injuries rather than from isolated bladder rupture. Factors associated with mortality are: age greater than 60 years; type I fractures (comminuted, involving three or more parts of the pelvic ring); four to five associated organ injuries; presentation in shock; and pedestrian injuries. The mortality rate in patients with bladder rupture is 10 to 20 per cent; this is not due to the bladder injury itself but to associated injuries. Uncontrolled bleeding from torn pelvic vessels is best managed by placement of an external fixator, or when necessary, by selective angio-graphy and embolization.

Symptoms and signs

Patients with extraperitoneal extravasation caused by rupture of the anterior bladder wall develop suprapubic pain, tenderness, and swelling. Ecchymosis of the skin overlying the bladder is also common. Intraperitoneal rupture produces the desire to void but the inability to do so. Lower abdominal pain is another symptom, which, in the later phase, develops into the findings of peritonitis. Uremia may result when urine comes into contact with the peritoneum: this acts as a semipermeable membrane, dialyzing the high urea nitrogen in the urine and causing a rapid rise in blood urea nitrogen. Fever and shock result if a long time has elapsed since the bladder rupture, and if the urine is infected.

Diagnosis

Gross hematuria occurs in most patients who are able to void: bloody urine is found at the time of urethral catheterization, or the bladder may be empty if there is a large intraperitoneal tear. Urethral injury should be excluded prior to the insertion of a Foley catheter. If this is not possible, a retrograde urethrogram should be performed first. When the catheter has been inserted a static cystogram can be obtained, taking oblique as well as anteroposterior views. Intraperitoneal rupture is indicated by the presence of contrast material within the abdominal cavity. A teardrop deformity of the bladder occurs classically with an extravesical collection of blood and urine in the space of Retzius. A postdrainage film is mandatory as it may disclose extravasation of contrast material in the retrovesical space that had been hidden by the contrast-filled bladder during the full phase of the cystogram. In patients with clinical evidence of a ruptured bladder, a static cystogram may be normal if the small tear has ‘sealed'. Distension of the bladder may therefore be needed: this is achieved with a ‘stress' cystogram, in which an extra 60 ml of contrast is put into the bladder after the pressure has equalized on the standard static cystogram. A plain radiograph of kidney, ureter, and bladder prior to the cystogram is essential in determining the presence of pelvic fractures or foreign bodies. An intravenous pyelogram allows the upper urinary tracts to be assessed for associated ureteral or renal injuries, but it diagnoses bladder ruptures in no more than 20 per cent of known bladder injuries. In medical centers using computed tomography (**CT**) to evaluate trauma, a CT cystogram performed with retrograde gravity filling of the bladder provides sufficient visualization of the bladder to diagnose rupture. A bladder filled by contrast excreted by the kidneys is not satisfactory: false positives may occur as the result of extravasation of contrast from other sources such as the bowel, the upper urinary tract following intravenous administration, or the posterior urethra following retrograde urethrography. In questionable cases, or if laceration of the bladder neck is suspected, the bladder may be filled with contrast under fluoroscopy. Fig. 7 presents an algorithm that will aid the physician in assessing the patient with pelvic and suspected urologic injuries.

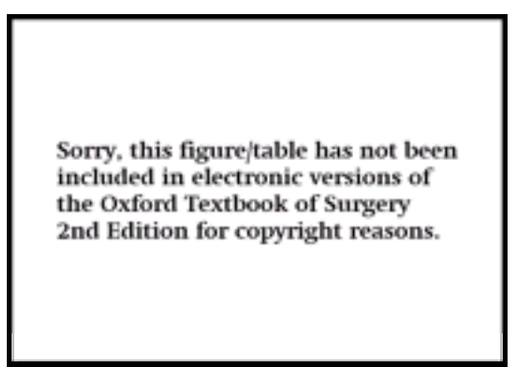


Fig. 7. Algorithm for evaluation and management of lower urinary-tract injuries associated with pelvic fractures.

Treatment

Bleeding from a contused bladder usually resolves with drainage by urethral catheter. Most extraperitoneal ruptures of the bladder may be managed nonoperatively by urethral catheter. The ideal candidate for this form of management is a patient who has mild hematuria, no tearing of the bladder neck, and is not going to the operating room for associated injuries. In these patients, large-catheter (24F), low-pressure urethral drainage will most often lead to spontaneous sealing of the bladder injury. The catheter should be left in place for 2 weeks and removed only after a cystogram demonstrates no extravasation. Heavy hematuria that might cause catheter-occluding clots, massive extravasation, and, most importantly, an injury to the bladder neck all require operative repair. Patients undergoing laparotomy for associated injuries should have their bladder injuries repaired at the same time. We favor an intraperitoneal approach through the dome of the bladder, with debridement if necessary and two-layer closure with absorbable suture from the inside of the bladder. With this method the retroperitoneum does not need to be entered or drained. A suprapubic catheter and a urethral catheter (unless associated with a urethral injury) should be left for drainage. The Foley is removed when the urine clears and the suprapubic catheter is removed after a negative cystogram. Broad-spectrum antibiotics must be administered postoperatively, since this is a fertile area for infection.

Intraperitoneal rupture of the bladder must be managed surgically as the peritoneal cavity will provide a low-pressure reservoir for continued urinary leakage and intraperitoneal contents may herniate into the bladder through the dome, preventing closure of the rent. The preferred method is a transperitoneal approach to the dome of the bladder, with vertical cystotomy and inspection of the rest of the bladder for associated extraperitoneal injuries, then two-layer closure with absorbable suture. Suprapubic cystotomy and urethral catheters should be placed and managed as described above for surgical treatment of extraperitoneal ruptures.

Ureter

Because of their size, location, and mobility, the ureters are uncommonly injured by trauma, but are not infrequently damaged as a complication of pelvic or endoscopic surgery. Excluding iatrogenic causes, gunshot wounds account for over 90 per cent of ureteral injuries, with the upper, middle, and lower parts being injured with equal frequency. Laceration of the ureter by a penetrating missile is associated with injuries to the small bowel (80 per cent), the colon (60 per cent) and the vena cava (20 per cent). Blunt ureteral trauma is more common in children and occurs as a result of either a congenitally dilated ureter or an extreme flexion injury causing a tear at the ureteropelvic junction.

Diagnosis

A high index of suspicion based on the mechanism of injury and the location of the penetrating trauma is necessary to make the diagnosis upon initial presentation. Many patients are severely injured and have had few diagnostic studies before being taken to the operating room. Up to one-third of patients will have no gross or microscopic hematuria, although over half will present with gross hematuria. Intravenous urography is frequently nondiagnostic, with findings of a delayed nephrogram or ureteral dilatation on the side of the injury: this may be because in many patients there is not time for a full study because of the severity of the injury; full studies have a greater accuracy of close to 80 per cent. The sensitivity of CT is unknown but is presumably as good or better than that of intravenous urography because of its greater ability to demonstrate small amounts of contrast extravasation. The retrograde pyeloureterogram is the most sensitive radiographic study and the most useful in planning surgical repair, but is often impractical in the acute setting. Intraoperative diagnosis can be made by the intravenous or intraureteral administration of indigo carmine dye. Missed ureteral injuries may present with abdominal pain, prolonged ileus, tenderness, fever, and a rising creatinine as urine accumulates in the retroperitoneum. Although gross or microscopic hematuria may be present, the drainage of urine through a surgical wound can be the first indication of ureteral injury. Ureteral injuries may be very subtle, and long delays in diagnosis and treatment may result in the loss of a kidney in up to one-third of cases.

Treatment

Primary surgical repair should be undertaken as soon as the diagnosis is made, if it is to be successful. The operative technique varies with the site and extent of the ureteral injury. High-velocity missiles create more tissue injury than might be apparent on simple inspection. Adequate debridement of the injured tissues, an infection-free area of surgery, watertight mucosa-to-mucosa closure with fine, absorbable, monofilament sutures, tension-free spatulated anastomosis, and good urinary drainage with ureteral stents and/or nephrostomy tubes are essential. If the above principles are not adhered to, ischemia, necrosis, and disruption or stricture of the ureteral anastomosis may result. Nephrectomy may then become necessary.

Lower ureteral injury

In the adult a quick and acceptable technique is to debride the lower ureter back to healthy tissue with a viable blood supply and implant it in a spatulated, freely refluxing fashion into the dome of the bladder. In a child, nonrefluxing reimplantation of the ureter in the standard Leadbetter–Politano fashion is the preferred repair (Fig. 8). If the length of the ureter after debridement and mobilization does not lend itself to reimplantation, transureteroureterostomy or a psoas hitch with refluxing reimplantation may be performed (Fig. 9).

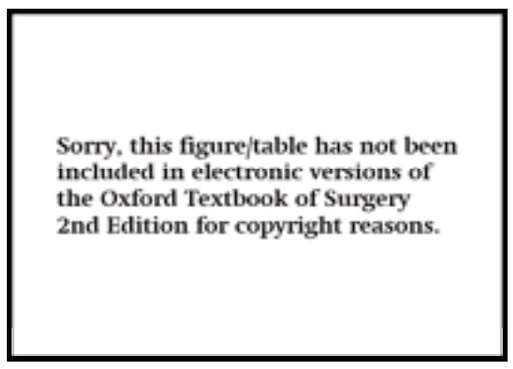


Fig. 8. Nonrefluxing reimplantation of a lower ureteral injury.

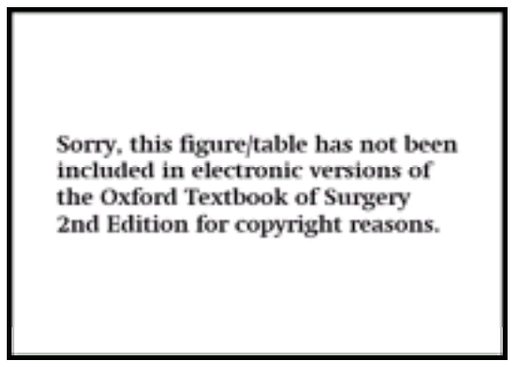


Fig. 9. Transureteroureterostomy.

Midureteral injury

The injured midureter is best repaired by primary anastomosis (Fig. 10). This segment of ureter is the most mobile and therefore more readily reapproximated.

Downward mobilization of the kidney may give the surgeon several extra centimeters of ‘ureteral' length.

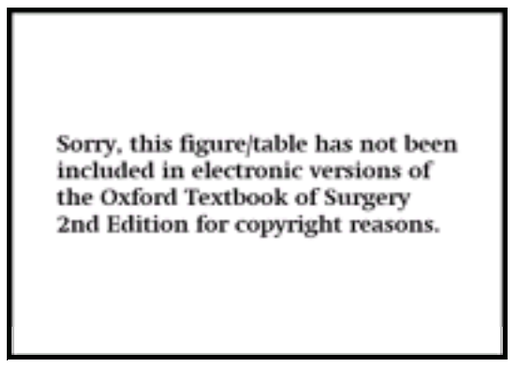


Fig. 10. Repair of midureteral injury.

Proximal ureteral injury

The preferred technique of repairing the upper ureter is primary anastomosis. Mobilization of the kidney and the use of the various methods of pyeloplasty aid in diminishing the gap between the healthy ends of the ureter.

The formation of an ileal ureter or kidney autotransplantation should be considered when the damage is so extensive that the other repairs are technically impossible. A discussion of iatrogenic ureteral injuries is not within the scope of this chapter.

Kidney

Renal injuries caused by blunt trauma or penetrating missiles may involve the parenchyma, collecting system, or vascular pedicle. Successful management of renal trauma depends on accurate and timely staging. Fig. 11 presents an algorithm for the diagnosis and management of renal trauma. Signs and symptoms of renal injury include gross or microscopic hematuria, which is present in greater than 90 per cent of renal injuries but may be absent in up to one-third of pedicle injuries, flank ecchymosis or mass, and shock.

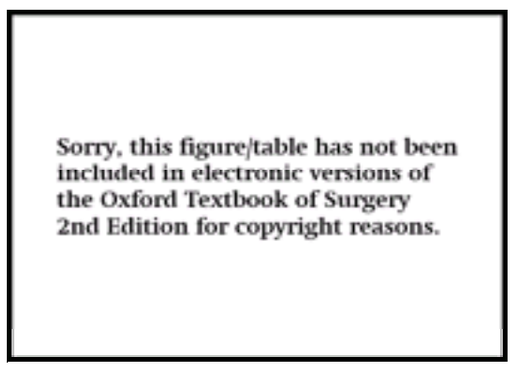


Fig. 11. Algorithm for evaluation and management of suspected renal trauma.

Penetrating injury

Penetrating wounds of the kidney usually result from gunshots or stabbing; they are often complicated by perforation of other structures such as abdominal organs, chest, spine, or major vessels, depending on the character and direction of the missile. Approximately 90 per cent of gunshot wounds and 60 per cent of stab wounds to the kidney are associated with intra-abdominal injuries. The liver and colon may be injured in up to 60 per cent of patients with injuries to the right kidney. Injuries to the left kidney are associated with injury to the spleen, stomach, or pancreas in 60 per cent of patients. Chest injuries occur in 20 per cent of all patients with injury to the kidneys. Any penetrating injury from the nipple line downwards should be considered to be intra-abdominal as well as intrathoracic.

The wounds in the renal and surrounding tissues may be perforating, lacerating, or explosive (Fig. 12). The degree of injury depends on the speed and shape of the missile as well as on the internal structure of the organ. Damage created by a bullet is related to the muzzle velocity. As the missile enters the body, gaseous vaporization produces more tissue damage. The bullet also tumbles in its trajectory and occasionally fragments, causing still more damage and creating an exit wound that is larger than the entry wound.

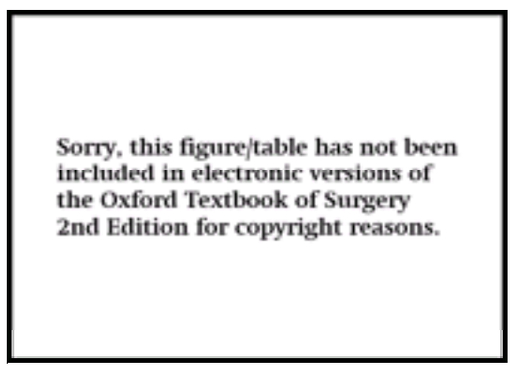


Fig. 12. Penetrating renal injury.

Deep parenchymal lacerations, tears in the collecting system, and rupture of the renal vascular pedicle are classified as major renal injuries (Fig. 13). Penetrating injuries to the kidney result in major injuries in 40 to 67 per cent of cases.

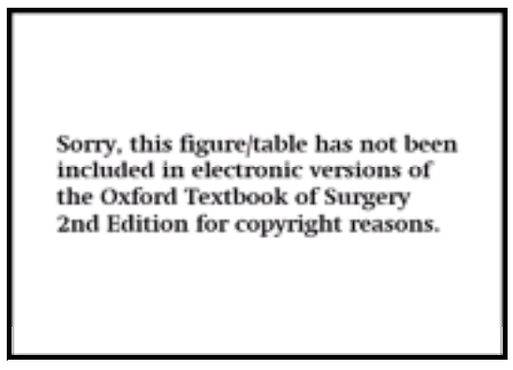


Fig. 13. Major renal injuries.

Non-penetrating injury

Blunt trauma causes 65 per cent of all renal injuries: this usually results from a direct blow to the abdomen that compresses the kidneys between the lower ribs and the spine, causing contusion and laceration (Fig. 14).

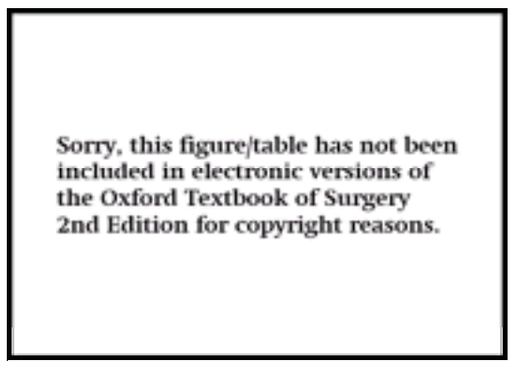


Fig. 14. Blunt renal trauma.

Indirect injury is much less common: this acceleration–deceleration type of injury is caused by falling and landing on the buttocks or feet, or when the patient is abruptly stopped as in a motor vehicle accident (Fig. 15(a)). The kidney remains in motion with respect to the rest of the body and aorta. The media and adventia of the renal artery stretch readily, whereas the relatively inelastic intima tears (Fig. 15(b)). The intimal tear initiates clotting, resulting in distal propagation of thrombus and eventual total thrombosis of the renal artery (Fig. 15(c)). Injuries to the renal pedicle are caused by blunt trauma in 85 per cent of cases; 90 per cent of these occur in patients under 25 years of age. The artery is injured in 70 per cent of patients, the vein in 20 per cent, and both vessels in the remaining 10 per cent. The mortality rate approaches 50 per cent, due mainly to the severity of associated injuries.

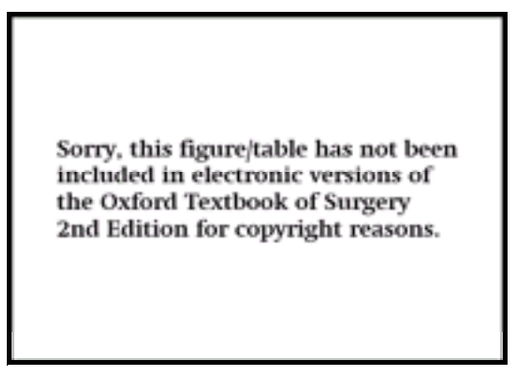


Fig. 15. Mechanism of injury of intimal tear of the renal artery following blunt trauma.

Fortunately, renal injury caused by blunt trauma is minor in the majority of patients. Contusion of the kidney or small, subcapsular hematomas are the rule (Fig. 16).

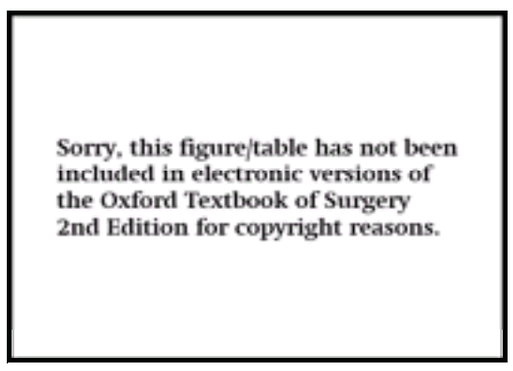


Fig. 16. Minor renal injuries.

Radiographic diagnosis

The need for radiographic staging is determined by the clinical staging. In adults who sustain blunt trauma with no episodes of hypotension, have only microscopic hematuria, and have no associated injuries or mechanism of injury suspicious for a major renal injury, radiographic staging is not necessary. Radiographic evaluation is essential in all other patients in whom there is a question of renal trauma. The purpose of radiographic staging is to determine the site and extent of renal injury. In centers equipped with CT, the abdominal scan with oral and intravenous contrast is the single best study for the evaluation of the kidneys in the hemodynamically stable patient. It is faster and more accurate than the intravenous urogram, and has the added ability to diagnose many other nonurologic injuries. Where the CT scan is not available, the intravenous urogram with nephrotomograms is still the mainstay for evaluating renal trauma in the emergency setting. Where no imaging is possible before taking the patient to the operating room for the correction of hemodynamic instability, the renal injury must be staged visually at laparotomy after a one-shot intravenous pyelogram with 2 cc/kg of iodinated contrast, with a flat plate of the abdomen taken at 5 min postinjection. This study most importantly demonstrates function and the presence of a contralateral kidney. Arteriography is rarely indicated but may be useful if angiography for other vascular injuries is being

undertaken, if intravenous pyelography is inconclusive and CT is not available, or if therapeutic embolization is contemplated.

Treatment

The ability to treat renal trauma nonoperatively follows from sound clinical and radiographic staging (Table 1); the inability to stage adequately a renal injury, usually due to an urgent need for the operating room, demands surgical staging. The proper treatment of the injured patient depends on a systematic approach combining clinical, radiographic, and surgical modalities (see Fig. 11).

Grade I	Hematuria or renal contusion, normal radiographic study, or nonvisualization without renal laceration
Grade II	Hematuria or renal laceration < 1 cm deep
Grade III	Laceration > 1 cm deep into the parenchyma without urinary extravasation
Grade IV	Laceration > 1 cm deep with urinary extravasation, or segmental renal thrombosis
Grade V	Multiple major lacerations, or pedicle injury

Table 1 Grading of renal trauma by the American Association of Surgery of Trauma

Minor (grade I and II) injuries are properly treated by observation. The only absolute indication for operative management is hemodynamic instability due to renal injury, or injury to the pedicle in a solitary kidney. Major lacerations of the collecting system with medial extravasation and large devascularized segments are strong relative indications. Relative indications that may be either observed or operated upon include major parenchymal lacerations with lateral extravasation, vascular injuries with a functional contralateral unit diagnosed after 4 to 6 h, and any major injury in a patient already undergoing laparotomy for associated injuries who can tolerate the extra time in the operating room needed to repair the kidney.

Complications of nonoperative management include delayed bleeding with a transfusion requirement, urinoma, or abscess. Late complications of nonoperative treatment may include atrophy of the kidney following unrecognized injury to the renal pedicle or hypertension. Delayed operation on the injured kidney often leads to nephrectomy. Complications of initial operative management include nephrectomy, hemorrhage, and urine leak.

Operative technique in renal injuries

Associated visceral injuries are common in patients with major blunt and penetrating renal trauma: the transabdominal approach therefore allows the best assessment of the patient as well as the best vascular control. Once the peritoneal contents have been explored and other injured organs have been taken care of, attention should be given to the injured kidney, which is best approached from the midline.

The posterior peritoneum is incised over the aorta at the level of the renal pedicle (Fig. 17(a)). A vascular clamp is applied to the renal artery (Fig. 17(b)), remembering that the right renal artery is posterior to the left renal vein. This maneuver allows the renal bed to be explored without excessive bleeding (Fig. 17(c)).

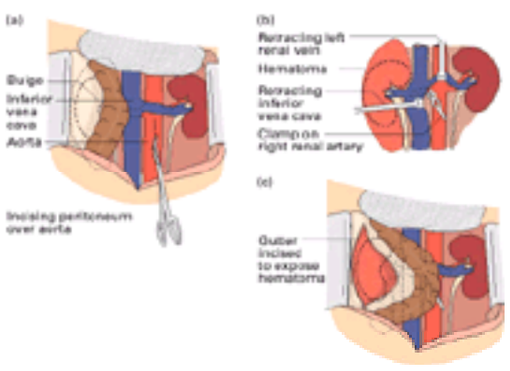


Fig. 17. Control of the renal pedicle.

If the renal pedicle is not controlled prior to opening Gerota's fascia, the tamponade effect is lost, and the need for nephrectomy increases as the clot is removed and the kidney explored. There is a 10 to 30 per cent need for nephrectomy even under the most favorable operative conditions.

When renal repair is possible the operative guidelines are: accurate assessment of the degree of injury; adequate debridement of injured tissue; meticulous hemostasis; watertight repair of the collecting system; and adequate drainage of the renal bed (Fig. 18).

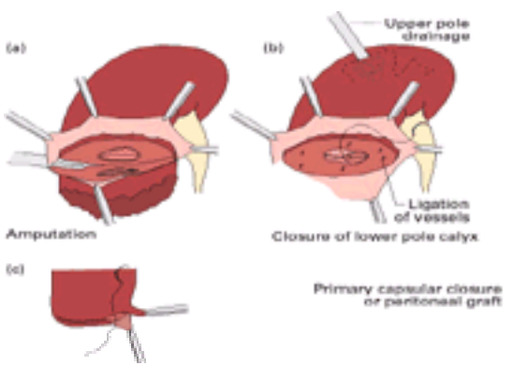


Fig. 18. Technique of lower-pole nephrectomy.

Injuries to the external genitalia

Wounds of the penis, urethra, and scrotum caused by blunt trauma or penetrating missiles may be associated with injuries to the thighs, perineum, bony pelvis, bladder, urethra or rectum. Thermal, electrical, and chemical burns may also involve the genitalia. In civilian life, injuries usually result from industrial accidents; fragmentation devices cause the majority of genital injuries in combat. Trauma can be either major or minor, with contusion, laceration, or puncture. Major injury can result from burns. Partial or complete amputation of the penis is not uncommon in acts of self-mutilation.

Blunt trauma

Blunt trauma, with crushing of the genitals, may cause severe edema and hemorrhage. Because of the elasticity of the tissues, scrotal and penile swelling is rapid. There is pain and occasionally difficulty in voiding. Traumatic conditions such as hematocele, hematoma of scrotal layers, rupture of the tunica, and intratesticular hematoma, with or without actual rupture, can now be diagnosed by ultrasonography, which greatly aids the urologist in determining those patients who require surgical intervention. Those who can be treated nonsurgically require the application of icepacks, pressure, and scrotal elevation. Difficulty in voiding can be overcome by a suprapubic cystocath. Once the edema begins to subside, in 4 to 5 days, heat, instead of ice, will accelerate the absorption of tissue fluid.

Testicular rupture is an unusual consequence of blunt trauma because the tunica albuginea covering the testicles is so tough. The patient experiences pain, with nausea and vomiting; there may also be a fever. The expanding scrotal mass requires incision and drainage, with adequate debridement of the extruded testicular tissues. The tunica is then closed with fine chromic sutures and a pressure dressing applied. Scrotal elevation and bed rest for several days diminish the pain. A testicular (nuclear) scan 1 month after the repair allows assessment of the post-traumatic viability of the testicle. Spermatogenesis is frequently impaired following severe injury, but endocrine function may be maintained.

Rupture of the penile corporal bodies (penile fracture) is treated similarly to rupture of the testicle. Presentation is usually with an audible snap during intercourse, followed by detumescence, hematoma, and pain. Urethral injury must be ruled out. The tunica albuginea of the affected corpus cavernosum is then exposed, debrided, and closed. A urethral catheter is placed. Penile rupture may produce impotence, due to damage of the neurovascular bundle, or a chordee due to corporal scarring. These problems can be overcome by using vasoactive drugs to achieve erection or by corrective surgery.

Degloving injuries of the penile and scrotal skin may occur with industrial accidents. The skin is usually very mobile and may therefore protect the deeper tissues from injury. Immediate repair, after debridement, with split-thickness skin grafts, or primary closure, decrease the chance of infection or sepsis. If the repair of the scrotum does not allow for the concomitant replacement of the testicles, they can be placed subcutaneously in the medial thigh until secondary repair becomes possible.

Penetrating injuries

Most penetrating wounds of the genitals should be treated as potentially infected. The basic principles in management are therefore debridement, lavage, repair, drainage, and antibiotics. When primary closure is possible, a tension-free anastomosis with absorbable sutures is performed.

The severed penis requires microvascular repair. The dorsal arteries, and the deep and superficial veins, need to be anastomosed to allow adequate tissue survival and function. The amputated penis is washed and prepared with betadine, wrapped in sterile gauze, and placed in a plastic bag submerged in cold saline until the patient can be made ready for the operating room. Repair of this injury, as well as others of the external genitalia, requires a good knowledge of plastic surgical techniques.

Burn injuries

Thermal injuries to the external genitalia are treated similarly to burns on other areas of the body. Antiseptic dressings, and conservative debridement and skin grafting when necessary, are the cornerstones of management. Chemical burns to the external genitalia are usually superficial and can be managed by thorough irrigation and dressings, and when necessary debridement and grafting. Electrical burns often cause extensive soft-tissue damage despite minimal surface damage. Electrical injuries should be observed for the initial 24 h to allow demarcation of affected tissue and only then undertake debridement and grafting

Further reading

Koraitim M. Pelvic fracture urethral injuries: the unresolved controversy. *Journal of Urology* 1999; **161**: 1433–44. [Review of the literature on pelvic fractures and urethral injuries. An algorithm for evaluation and management of pelvic fractures and lower urinary tract injuries is presented.]

Cline KJ, Mata JA, Venable DD, Eastham JA. Penetrating trauma to the male external genitalia. *Journal of Trauma* 1998; **44**: 492–4. [Modern review of the management of penetrating trauma to the penis and scrotum.]

Palmer LS, Rosenbaum RR, Gershbaum MD, Kreutzer ER. Penetrating ureteral trauma at an urban trauma center: ten year experience. *Urology* 1999; **54**: 34–6. [Series of patients with ureteral injuries. This study demonstrates the pitfalls of intravenous urography and urinalysis in diagnosing these injuries.]

McAninch JW, Carrol PR. Renal exploration after trauma: indications and reconstructive techniques. *Urologic Clinics of North America* 1989; **16**: 203–12. [A review of the indications and surgical techniques of renal reconstruction in trauma.]

Miller K, McAninch JW. Radiographic assessment of renal trauma: our 15 year experience. *Journal ofUrology* l995; **154**: 352–5. [This report establishes that significant renal injuries are vanishingly rare in patients suffering blunt abdominal trauma without either gross hematuria or shock.]

39.4 Urinary complications of spinal abnormalities and injuries

Marcus Drake and Jeremy Noble

Introduction
Neuroanatomy and neurophysiology
Innervation
Reflex control of the lower urinary tract
Lower urinary reflexes
Classification of spinal abnormalities
Clinical features
Investigation
Urological management of spinal cord injury
General considerations
Spinal shock
Facilitation of storage
Facilitation of emptying
Urinary diversion
Urinary tract infection
Stones
Vesicoureteric reflux
Other spinal conditions
Conclusion
Further reading

Introduction

The lower urinary tract stores urine at low intravesical pressures during bladder filling and permits efficient and unobstructed expulsion of urine during voiding. The process is controlled by complex neuronal pathways, so pathological processes affecting the spinal cord often have profound effects on lower urinary tract function. This chapter outlines the relevant neuroanatomy and reviews urological consequences of spinal cord pathology and their management. The emphasis is on traumatic spinal cord injury, but the general principles can be extrapolated to other spinal cord pathologies.

Neuroanatomy and neurophysiology

The lower urinary tract is regulated by neural circuits located in the brain, lumbosacral spinal cord, and peripheral ganglia which integrate somatic and autonomic mechanisms. Voluntary control of micturition is learned during early childhood. Neuronal control is closely related in the male to the control of sexual function.

Innervation

The lower urinary tract receives somatic, sympathetic, and parasympathetic nerves containing both afferent and efferent axons.

Efferents

Parasympathetic nerves provide the majority of excitatory input to the detrusor muscle. Preganglionic axons travel from the intermediolateral horn in the sacral spinal cord (S2–S4) to ganglia in the pelvic plexus. Sympathetic preganglionic fibres originate from the thoracolumbar spinal cord (T11–L2) and pass via the sympathetic chain to the inferior mesenteric and pelvic plexuses. Postganglionic sympathetic fibres travel in the hypogastric nerves to target organs. Somatic efferents arise from Onuf's nucleus in the ventral grey matter of the sacral spinal cord (S2–S4). They run peripherally in the pudendal nerve and supply the external urethral musculature (S2–3), pelvic floor (S3), anal sphincter, ischiocavernosus, and bulbocavernosus. Higher central nervous system control of efferents is carried in the medial, lateral, and ventral reticulospinal tracts. These control and co-ordinate detrusor activity with sphincter function.

Afferents

Sensory inputs arise from vesical tension/pain and urethral mucosa receptors. Afferents enter the posterior horn of the sacral spinal cord and decussate on to interneurones and projection neurones. Sacrobulbar tracts carry information on bladder volume and spinothalamic tracts carry pain and temperature inputs.

Reflex control of the lower urinary tract

Functionally the lower urinary tract consists of two units, the detrusor smooth muscle of the bladder and the bladder outlet, the latter comprising the bladder neck, urethra, and the striated muscle external urethral sphincter. They are controlled in a reciprocal manner by central pathways in an on–off switching circuit. Normally reflex activity is integrated in the pontine micturition centre, which is subject to modification by higher centres.

Lower urinary reflexes

Storage

Intravesical pressure stays low during filling due to the compliant nature of the bladder wall and inactivity of parasympathetic efferents. Tonic sympathetic activity maintains tone in the smooth muscle of the bladder outlet. Sympathetic nerves also inhibit bladder activity in animals, but this role has not been established in humans. Electromyographic activity within the urethral sphincter increases during bladder filling, reflecting increased efferent output in somatic nerves.

Voiding

At full bladder capacity, afferent discharges from tension receptors lead to a reversal of the pattern of efferent outflow. Parasympathetic nerves become active, with inhibition of sympathetic and somatic pathways. Consequent sphincter relaxation and detrusor contraction raises intravesical pressure and causes urine to flow. Stimulation of urethral receptors by urine perpetuates the reflex until the bladder is completely empty.

Abnormal reflexes

Loss of central connections does not preclude detrusor activity, since reflex circuits in the distal spinal cord can also mediate bladder contraction. In normal adults these circuits are usually not manifest past infancy, but in the presence of spinal cord pathology they may become prominent. It is uncertain whether their presence represents emergence of primitive reflexes, or whether they are a result of synaptic plasticity in the central nervous system.

The afferent limb of the reflex arc may change, with unmyelinated C-fibres becoming significant. Reflex detrusor contraction may therefore occur in response to unusual stimuli such as instillation of cold water into the bladder. Alteration of the neurotransmitters has been identified in certain pathological conditions and non-adrenergic non-cholinergic neurotransmission may emerge in the presence of spinal cord pathology.

Classification of spinal abnormalities

See Table 1.

Congenital
Spina bifida
Degenerative
Multiple sclerosis
Motor neurone disease
Non-metastatic manifestation of malignancy
Syringomyelia
Subacute combined degeneration of the cord
Friedrich's ataxia
Trauma
Traumatic spinal cord injury
Iatrogenic
Cord compression
Central disc prolapse
Tumour (metastasis, meningioma, neuroectodermal tumours)
Arteriovenous malformation
Epidural abscess
Meningeal haemorrhage
Aneurysmal bone cyst
Infarction
Cardiac arrest
Systemic hypotension
Anterior spinal artery occlusion
Inflammatory/infective
Meningitis
Tuberculosis
Transverse myelitis
Neurosyphilis

Table 1 Classification of spinal abnormalities

Clinical features

Most spinal cord pathologies are insidious in onset and many follow an intermittently progressive course. Spinal reflex activity is typically preserved or enhanced in these circumstances. In contrast, immediately following spinal cord injury overt reflex activity ceases distal to the injury. This period of ‘spinal shock' lasts a variable length of time (days to years). Neurological examination is undertaken to ascertain level and completeness of injury.

On recovery of reflex activity, voiding is usual for suprasacral injuries and is termed hyperreflexia. Such voiding can not be initiated or inhibited voluntarily by the patient, but it may be possible to induce micturition by stimulating specific trigger areas such as the inner thigh or the perineum. Co-ordination of detrusor contraction with sphincter relaxation may be lost (‘detrusor sphincter dyssynergia'), resulting in poor flow and high pressure voiding. Inability to void, which is termed areflexia, is usual for sacral or more distal injuries, but a proportion of higher lesions cause areflexia. Most injured patients will be unable to feel any stimulus below the injury level. Some patients retain urethral sensation, especially in incomplete injuries. Absence of pain sensation means that stones and other urinary tract problems are often asymptomatic, or present with urinary tract infections or catheter blockage, rather than ureteric colic.

Patients with injury above T6 may suffer autonomic dysreflexia, a potentially life-threatening syndrome of disordered autonomic function characterized by excessive sweating, flushing, severe headache, and hypertension. It can be precipitated by bladder distension, urological intervention, constipation, pressure sores, or any stimulus that would be painful to an uninjured person. Attacks resolve on removal of the cause. Sublingual nifedipine can be used to control hypertension.

Renal failure used to be the main cause of mortality in these patients, as a result of high pressure voiding, vesico-ureteric reflux, stones, urinary tract infection, and amyloidosis. Improved management has raised life expectancy after spinal cord injury to near normal, but regular surveillance of renal function is vital as deterioration can be rapid and asymptomatic.

Investigation

On return of reflex activity, urodynamics is undertaken to identify patients with hypocompliance, detrusor sphincter dyssynergia, or high pressure hyperreflexia. Such features necessitate a more interventional approach and frequent follow-up. Some centres perform renography at this stage as a baseline for future comparison in assessing management. Upper tract imaging, usually ultrasound, detects any hydronephrosis or stone formation. Later investigation of chronic spinal cord injury is tailored according to risk factors or change in clinical status (Fig. 1).

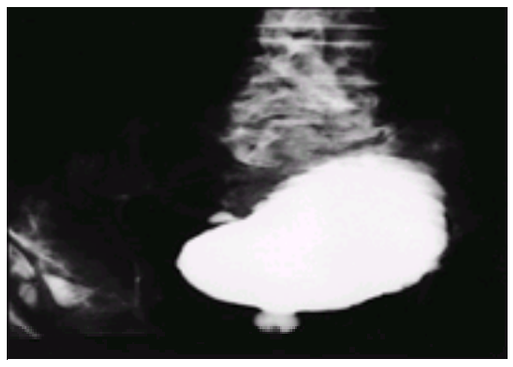


Fig. 1. Cystogram of a neuropathic bladder showing trabeculation, a small diverticulum, and an open bladder neck.

Urological management of spinal cord injury

General considerations

Management decisions must consider the capabilities of the patient, especially hand function, limb spasms, ability to transfer, motivation, sexual activity, bowel function, social support, transport, finance, and pending litigation. Many patients have preferred options, acquired from discussion with other injured patients and their own knowledge. An increasing number are reluctant to undertake irreversible procedures in the hope that their spinal cord injury may come to be curable in the future. The ideal of management is to achieve low pressure voiding with control of timing and full continence between voids. This is rarely achieved, so in practice it is more common to facilitate the less effective component, whether storage or voiding. One of the mainstays of management is clean intermittent catheterization, either by the carer, or the patient if hand function is good.

Regular follow-up is required, as deterioration of renal and bladder function may be rapid and asymptomatic. Routine serum urea and electrolytes may be combined with upper tract imaging, renography, and urodynamics. Reasons to change management include deterioration in upper tract function, worsening urodynamic parameters, urinary tract infections, drug side-effects, and skin problems.

Spinal shock

Management varies between centres. A catheter may be placed until the return of reflex activity, a suprapubic catheter being preferable in females to prevent urethral damage. Alternatives include early institution of clean intermittent catheterization or bladder training, in the hope of encouraging early recovery of reflex voiding. These require careful supervision to prevent bladder overdistension. At this stage of injury it is impossible to predict subsequent behaviour of the urinary tract.

Facilitation of storage

Treatment aims to convert a small capacity, low compliant, high pressure bladder into a high capacity, compliant, low pressure bladder, even if this abolishes spontaneous voiding and makes clean intermittent catheterization necessary. Storage may be improved either by reducing detrusor hyperreflexia or by increasing outlet resistance.

Reduction of detrusor hyperreflexia

Drugs used to facilitate storage work by reducing hyperreflexic contractions and sensation. Anticholinergic agents (e.g. propantheline, oxybutinin, tolteridine, and amitriptyline) are effective, but they cause side-effects including dry mouth, tachycardia, blurred vision, and constipation. Baclofen, a polysynaptic inhibitor used to treat skeletal muscle spasm, is capable of reducing detrusor hyperreflexia. Capsaicin, an irritant derivative of hot red peppers, blocks conduction in unmyelinated C-fibres

and release of neuropeptides from nerve endings. Intravesical capsaicin increases functional bladder capacity and reduces irritative symptoms.

Neuromodulation is the modification of the sensory and/or motor function of the bladder by electrical or magnetic stimulation of the sacral nerve roots. Stimulation parameters can be adjusted to enable facilitation of storage or voiding or penile erection. Implantable electrodes enable chronic neuromodulation, but they often have to be removed because of infection or migration. Neuromodulation is not effective in all patients and preservation of the sacral spinal cord is a requirement.

Augmentation enterocystoplasty improves symptoms in up to 90 per cent of cases, but carries the risk of metabolic and malignant effects. Detrusor myomectomy is an alternative in which the detrusor is dissected from the underlying mucosa, effectively creating a bladder autoaugmentation. Detrusor myomectomy avoids the adverse effects of enterocystoplasty, and it does not preclude subsequent cystoplasty or deafferentation, but variable success has been reported. For both techniques careful patient selection is essential. Because residual urine volume is increased, clean intermittent catheterization may be necessary following either procedure. Therapeutic overdistension, in which the bladder is exposed to high pressure stretching, is now rarely performed as the beneficial effects are limited in hyperreflexia and there may be long-term problems.

Increase in outlet resistance

The pelvic floor can be strengthened by chronic electrical stimulation using intravaginal or anal stimulators. a-Adrenergic agonists, for example ephedrine, work by stimulation of the bladder neck. Their effectiveness is variable and they should be avoided in patients prone to autonomic dysreflexia. Surgical options include periurethral injection of collagen or Teflon which serves to bulk the tissues, thereby reducing lumen size. Suspension of the vesicourethral junction, using a vaginal, retropubic, or laparoscopic approach, can also be effective. Recent improvements in artificial urinary sphincters have led to their increasing use. They are successful in 80 per cent of cases, but there is a risk of spontaneous failure, infection, or erosion into the urethra. Artificial sphincters typically have a working life of about 10 years. Where these procedures are unsuccessful the urethra or bladder outlet may be closed and the bladder drained through a suprapubic catheter or urinary diversion.

Facilitation of emptying

Voiding can be achieved in some individuals by idiosyncratic manoeuvres such as suprapubic tapping or stroking a trigger zone to set off a hyperreflexic contraction. Other patients empty by expressing the bladder, raising the intra-abdominal pressure to force urine past the weak external sphincter. Expression risks raising intravesical pressure to dangerous levels and is no longer recommended, but patients already established on the technique may be reluctant to change their accustomed method.

Patients with bladder outflow obstruction may respond to a-adrenoceptor antagonists. Cholinergic agonists such as bethane-chol, given to stimulate detrusor contraction, do not elicit useful bladder emptying in patients with neurological problems.

Severe detrusor sphincter dyssynergia or bladder outlet obstruction may necessitate transurethral resection, bladder neck incision, or sphincterotomy. Urethral stenting to hold the bladder outlet patent fell out of favour because of major problems with encrustation, migration, and difficulty in removal. Newer stents, such as the Memokath, are achieving better results. Detrusor sphincter dyssynergia has been treated with some success by the injection of botulinum toxin directly into the external urethral sphincter.

Sacral anterior root stimulation is the treatment that comes closest to the ideal of low pressure voiding with control of timing and full continence between voids. An implantable device with stimulatory electrodes in the sacral foramina is used to generate voiding. Simultaneous posterior rhizotomy eliminates reflex incontinence between stimulations by abolishing the micturition reflex. Sacral anterior root stimulation can only be applied where distal innervation is intact and where the bladder is capable of contraction. The equipment is expensive.

Urinary diversion

In certain cases it is most appropriate simply to drain the bladder, either by means of a catheter (urethral or suprapubic) or surgery (formation of a conduit or continent reservoir). Catheters are associated with risk of blockage, stones, and stricture formation and it is uncertain whether the long-term outcome in relation to protection of the upper tracts is worse.

Urinary tract infection

Frequent urinary tract infections may result from indwelling foreign bodies, incomplete bladder emptying, or vesicoureteric reflux, but often no cause is found. Late diagnosis means that pyonephrosis and renal abscesses are common. Patients are also at risk of urethral and psoas abscesses. Symptomatic recurrent urinary tract infections require correction of the cause, high fluid intake, and scrupulous hygiene. Antibiotic therapy should be reserved for patients with systemic complaints and its use as prophylaxis is best avoided.

Stones

Renal stones are common in patients with spinal cord injuries and they may be associated with deteriorating renal function. Risk factors include urinary stasis, alkaline urine, infection by urea-splitting organisms, and skeletal demineralization caused by reduced mobility. Bladder stones are due to incomplete emptying, foreign bodies, or urinary tract infections. Stones may be asymptomatic, or they may present with urinary tract infections, cloudy urine, or autonomic dysreflexia. Management follows the same principles as stones in uninjured people.

Vesicoureteric reflux

Vesicoureteric reflux can be managed by submucosal injection of Teflon or collagen paste to narrow the ureteric orifice into a slit. Ureteric reimplantation is the definitive antireflux procedure, but it can be extremely difficult with a thick-walled, trabeculated, neuropathic bladder. It is sometimes necessary, therefore, to divert the urinary stream in order to protect the upper tracts.

Other spinal conditions

As for traumatic spinal cord injury, motivation and commitment are important deciding factors in the success of treatment. The spinal pathology may influence life expectancy and in some conditions intellectual function will be impaired. Otherwise the general principles and treatments described above can be applied.

Conclusion

The complex central control of the lower urinary tract means it is almost inevitably affected by spinal cord pathology. Careful and thorough investigation with regular follow-up is necessary to select appropriate treatment and identify those at risk of upper tract deterioration. By this approach, life expectancy for patients with traumatic spinal cord injury is near normal and quality of life need not be influenced detrimentally by the lower urinary tract.

Further reading

Blaivas JG. The neurophysiology of micturition: a clinical study of 550 patients. *Journal of Urology* 1982; **127**: 958–63. [A seminal work on the central control of micturition.]

Dmochowski RR, Ganabathi K, Leach GE. Non-operative management of the urinary tract in spinal cord injury. *Neurourology and Urodynamics* 1995; **14**: 47–55. [Long-term follow-up of patients managed with drugs and/or clean intermittent catheterization.]

Rijkhoff NJM, Wijkstra H, van Kerrebroeck PEV, Debruyne FMJ. Urinary bladder control by electrical stimulation: review of electrical stimulation techniques in spinal cord injury. *Neurourology and Urodynamics* 1997; **16**: 39–53. [A clear and comprehensive review.]

Thomas DG, O'Flynn KF. Spinal cord injury. In: Mundy AR, Stephenson TP, Wein AJ, eds, *Urodynamics: principles, practice and application*, 2nd edn, pp. 345–58. Churchill Livingstone, Edinburgh, 1994. [An excellent description of practical urodynamics.]

Trop CS, Bennett CJ. Autonomic dysreflexia and its urological implications: a review. *Journal of Urology* 1991; **146**: 1461–9. [A full explanation of this important problem.]

39.5 Urinary stone disease

David M. Rudnick and Stephen P. Dretler

Introduction
Incidence
Composition and etiology
Calcium oxalate
Struvite
Uric acid
Cystine
Other
Treatment
Techniques of retrograde and antegrade endoscopy
Endoscopic intracorporeal lithotripsy devices
Extracorporeal shock-wave lithotripsy
Open surgery
Bladder and urethral stones
Urethral calculi
Bladder calculi
Ureteral stones
Symptoms
Diagnosis
Treatment
Renal stones
Symptoms
Diagnosis
Treatment
Summary
Further reading

Introduction

Urinary calculus disease is one of the oldest known to man. Egyptian mummies from 3500 BC have been found to have renal and bladder calculi. Practical and ingenious methods of stone management have evolved over the millennia. Anticipating modern lithotripsy, the Egyptians used gum to attach a diamond to the tip of a hollow reed, inserted it into the bladder through the urethra, and the patient then walked, allowing the diamond to fragment the more fragile bladder stone. In ancient Greece, the seriousness and ubiquity of stone disease stimulated the development of a specialty of practitioners, non-physicians, who treated the problem, leading Hippocrates to include in his oath the promise of physicians not to ‘cut for stone but to leave that to the practitioners of the art'. The Arab physicians, renowned pharmacologists during the European Dark Ages, were reported to have developed solvents to dissolve calculi; unfortunately, their formulas have been lost to history. During the Middle Ages, itinerant barber–surgeons, many of whom were clergymen, developed a technique for the perineal removal of bladder stones that relieved many sufferers and killed many others. The technique required the patient to be held by strong men in the exaggerated lithotomy position (Fig. 1). A perineal incision was made and, in under 1 min, a long knife was used to incise the plane between the rectum and prostate, creating a perineal vesicostomy. Using abdominal pressure, the bladder stone was pushed towards the perineum and grasped by two fingers inserted in the perineum. The most famous of these stone surgeons was Frère Jacques of nursery-rhyme fame. Over the centuries, more sophisticated methods of stone crushing were developed. With the advent of anesthesia in 1846, surgical removal of calculi evolved as the most common treatment for stones that could not be passed spontaneously. Surgical treatment of upper urinary calculi and stone-crushing forceps placed in the bladder through the urethra remained the technique of choice until the last decade, when less invasive methods of stone management were developed.



Fig. 1. An engraving from 1535 that demonstrates the ancient barbaric method of perineal lithotomy for the removal of a bladder stone.

Fragmentation by extracorporeal shock-wave lithotripsy has revolutionized the management of stone disease. Percutaneous nephrostolithotomy has now been substituted for surgery in the majority of patients who have a stone too large for fragmentation. For the removal of ureteral calculi that will not pass, surgery has been almost completely replaced by endoscopic techniques, including ureteroscopy with basket extraction or *in situ* fragmentation by electrohydraulic, ultrasonic, or laser techniques. In fully equipped major centres, surgery is now performed on fewer than 2 per cent of patients with calculi who require intervention. However, becauseof their expense, it is unlikely that these new techniques will be available for more than 30 per cent of the world population.

The diagnosis and treatment of urinary stone disease has three goals: symptomatic relief for the patient, preservation of renal function, and elimination of a source for infection or bleeding. Management thus depends on the anatomy and symptoms of the patient, the degree of obstruction and function in the involved renal unit, the presence of infection or gross hematuria, the size and composition of the stone, the experience of the treating physician, and the availability of the various treatments.

Incidence

The highest incidence of calculi occurs between the ages of 30 and 50 years; the 3:1 male:female ratio is unexplained. Three per cent of the population of the United States will form a stone some time during their lives. There is a recurrence rate of approximately 7 to 8 per cent a year; within 5 years, the first-time stone former has a 40 per cent chance of developing another. Although the disease is more common in family members, with the exception of cystine stones (autosomal recessive) there is no genetic pattern in the usual forms of calculus disease.

Composition and etiology

Calcium oxalate

Calcium oxalate is the most common constituent of urinary lithiasis in the industrialized world. Crystallization occurs when the solubility product of calcium and oxalate in urine is exceeded. Hypercalciuria (more than 300 mg calcium in a 24-h urine collection in males; more than 250 mg in females) may be secondary to hypercalcemia (in hyper-parathyroidism, paraneoplastic syndromes, sarcoidosis etc.), a consequence of increased intestinal absorption of calcium, skeletal demineralization or decreased renal calcium reabsorption, or it may be idiopathic.

Oxalate is chelated in the small intestine by dibasic anions such as calcium. Oxalate that remains free is then absorbed in the colon and excreted in the urine. Urinary oxalate is therefore increased in patients who ingest large amounts of oxalate, or those with anatomic or physiologic short-bowel syndromes (resection or Crohn's disease) that cause fat malabsorption, calcium soap formation, and increased delivery of free oxalate to the colon. Calcium oxalate calculi (‘jackstones') are the most

radiodense. They are brown and black, and may have diagnostic spicules (Fig. 2).

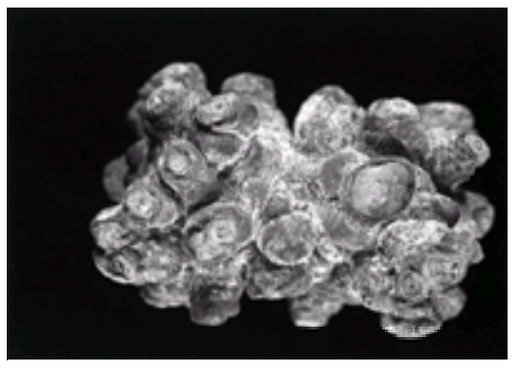


Fig. 2. A calcium oxalate monohydrate stone: their nodularity may be exaggerated to spicule formation, thus gaining the name ‘jackstone'.

Struvite

Magnesium ammonium phosphate stones occur as a consequence of the combined presence of a foreign body or obstruction and infection by a urea-splitting organism, usually *Proteus*. Such stones are less dense than oxalate and are composed of very fragile, yellow and white, chalky substances; they often fill the entire calyces and may take the form of ‘staghorn' calculi (Fig. 3). Bacteria live in the stone itself and therefore clearance of the entire stone is essential to eradicate infection.

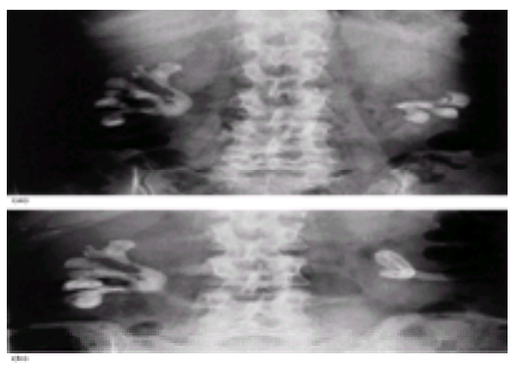


Fig. 3. (a) A plain abdominal radiograph of bilateral staghorn calculi, the right being larger than the left. (b) The same patient following percutaneous removal of the left staghorn calculus: the nephrostomy tube is in place; the large right calculus remains.

Uric acid

Uric acid stones are associated with dehydration, a fixed low urine pH, and hyperuricosuria (greater than 750–800 mg/day), which may occur without associated hyperuricosemia. Uric acid stones are radiolucent unless they fill the calyceal system, when their large mass may cause them to be minimally radiodense.

Cystine

Cystine stones are less radiodense than calcium oxalate calculi. They have a yellow–tan appearance of maple sugar. Cystine stones are the principal clinical manifestation of the autosomal-recessive disease of cystinuria, which is a disorder of the absorption and excretion of the dibasic amino acids. The 24-h urine collection will reveal an excretion of more than 400 mg cystine.

Other

Urinary calculi, most commonly calcium phosphate stones, can occur with prolonged immobilization leading to bone resorption or due to distal renal tubular acidosis.

Treatment

The treatment of urinary stone disease includes open surgery, percutaneous techniques, and the use of extracorporeal or endoscopic methods to fragment calculi into spontaneously passable concretions.

Techniques of retrograde and antegrade endoscopy

The endoscopic approach to the ureter is a natural extension of cystoscopy. The ureteroscope may vary from 7F to 12F in diameter. The smaller-diameter ureteroscopes can be introduced into the ureter without the need for dilation. Ureteroscopes have a fiberoptic (rigid or flexible) light channel and a working port through which catheters, baskets, lasers, ultrasonic lithotrodes, or electrohydraulic probes may be inserted and through which irrigation fluid is infused. Ureteroscopy is usually done under regional or general anesthesia. The ureteroscope is inserted under direct vision and advanced to the level of the calculus (see Fig. 8(a)). The calculus is then manipulated or fragmented (described below) (see Fig. 8(b), Fig. 8(c)). Early ureteroscopes were all rigid, encased in metal, and had inflexible light bundles. The newest generation are flexible, and, when necessary, may be advanced to the kidney for diagnosis or treatment.

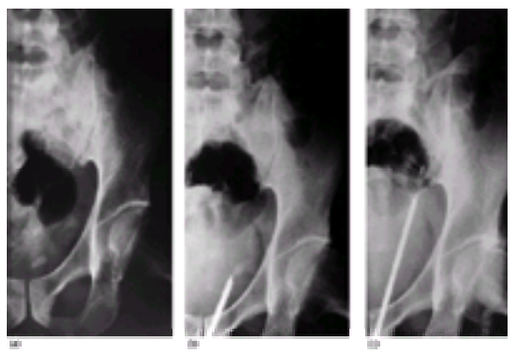


Fig. 8. (a) A distal left ureteral stone; (b) a ureteroscope through the bladder into the ureteral orifice and up to the level of the stone with beginning fragmentation; (c) no further stone remains; the ureteroscope is within the ureter.

The percutaneous method was the first non-surgical technique developed for the removal of renal calculi. A trocar needle is inserted into the kidney and a guidewire is passed through the trocar. After dilation over the guidewire, a 28F to 30F sheath is placed through the parenchyma to the collecting system. Rigid and flexible

nephroscopes can then be used to grasp and extract, or fragment and aspirate intrarenal or upper ureteral calculi. This technique is now used for large calculi or those that do not respond to extracorporeal shock-wave lithotripsy. Complications of percutaneous nephrostolithotomy include parenchymal bleeding, hydrothorax or pneumothorax, injury to adjacent organs, and incomplete stone removal.

Endoscopic intracorporeal lithotripsy devices

Electrohydraulic lithoptripter

The electrohydraulic lithotriptor was conceived in the 1930s in St Petersburg. It depends on the creation of an electrical spark, lasting 2 to 5 µs, in a fluid medium. The spark vaporizes water, creating a tiny ‘cavitation bubble'; this expands and then implodes, creating a pressure wave by both expansion and contraction. The pressure wave overcomes the tensile strength of the calculus's crystal bonds and induces fragmentation. Its effectiveness depends on the relative fragility of the calculus: struvite, calcium oxalate dihydrate (yellow), and uric acid stones are the most fragile, while calcium oxalate monohydrate (black), cystine, and calcium phosphate stones are harder and more difficult to fragment. The spark must be kept 2 mm from the urothelial wall to avoid mucosal injury or perforation. The electrohydraulic lithotriptor must be used under direct vision; extreme care and an experienced operator are required to prevent tissue injury. It is used through the operating port of cystoscopes, ureteroscopes, and percutaneous equipment.

Ultrasonic lithotripsy

The ultrasonic lithotrode is a hollow tube placed under direct vision into the bladder by a cystoscope, into the ureter through a ureteroscope, and into the kidney percutaneously (Fig. 4). The ultrasonic probe is hollow and vibrates in four directions, causing a jack-hammer effect on the stone. The stone is fragmented and the pieces are removed by suction through the hollow center of the lithotrode. The vibrations may cause tissue injury. The sonotrode is rigid and cannot be placed in the urinary tract via flexible instruments. It is an efficient form of lithotripsy and is especially effective for the larger, harder renal calculi that do not respond to less invasive techniques.

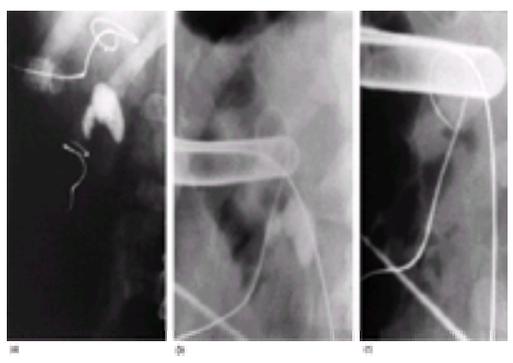


Fig. 4. Percutaneous ultrasonic nephrostolithotomy: (a) a large stone in the right kidney; (b) a percutaneous guidewire placed down the ureter and a large percutaneous tube placed into the collecting system of the right kidney; (c) abdominal film of the kidney after ultrasonic desiccation and aspiration of the stone fragments performed through a percutaneous tract; there is no stone remaining.

Pneumatic lithotripsy

The pneumatic lithotrode may be placed through a ureteroscope and may be as small as 2.4F. It is semirigid and has a flexure of 20°. A pneumatic pulse drives the solid tip at a rate of 5 to 10 Hz, causing very effective stone fragmentation.

Laser lithotripsy

The pulsed dye laser operates at 504-nm wavelength for 1 ms through a 320-nm core diameter, silica-coated, quartz fiber. It delivers 140 mJ of energy with each pulse. The low energy per pulse and the minimal absorption of 504-nm light by tissue make this a safe and effective method of stone fragmentation. The laser fiber is passed through the operating port of a ureteroscope. The small fiber size, accuracy, and safety allow successful fragmentation of ureteral stones with minimal invasiveness. More recently the holmium:YAG laser has been used for stone fragmentation. This laser has a 2100-nm wavelength and operates in a 350-ms pulsed mode at 5 to 30 Hz and energy settings of 0.2 to 2 J. The holmium causes cavitation due to vaporization of water at the stone–fluid interface. Unlike the pulsed dye laser, the holmium has the capacity to destroy tissue, and therefore it must not come into contact with the urothelium. Despite this risk, the holmium effectively fragments hard stones such as calcium oxalate monohydrate and cystine, produces smaller residual fragments, and causes less migration of the stone with each shock.

Endoscopic manipulation devices

In addition to the intracorporeal lithotripsy devices available for use through endoscopes, there are a number of stone-grasping devices that may be used to extract fragments from the urinary tract. These include various configurations of wire baskets, grasping forceps, and stone-crushing forceps. Many of these instruments are available in small and flexible enough configurations to be used through even the smallest flexible ureteroscopes.

Extracorporeal shock-wave lithotripsy

Perhaps no treatment in urology has changed the basic approach to a condition as swiftly and as effectively as this technique. It was developed in Munich in the late 1970s and, where available, is the most widely accepted treatment for the majority of renal calculi.

The first-generation lithotriptor used a spark-plug-like electrode placed underwater at the base of a semi-ellipsoid. This apparatus was submerged in the bottom of a tub of water. The spark created from the electrode vaporized water, creating shock waves that were reflected from the side walls of the semi-ellipsoid and focused at a point (*F2*) 13 to 15 cm from the origination (*F1*) (Fig. 5, Fig. 6). Biplanar fluoroscopy is used to site the patient's stone at the point *F2*. Using a power of 18 to 24 kV, shock waves are propagated, and the stones are slowly and completely crumbled to spontaneously passable fragments (Fig. 6). Each shock was triggered so that it occurred during the refractory period of the QRS complex, preventing cardiac arrhythmia, and treatments were limited to 2000 shocks to prevent renal injury. Since the introduction of the original unit, modifications and variations have been developed. The shock waves may now be created by electronically stimulating a ceramic dish (piezoelectric) or by electromagnetically stimulating a membrane. The shock waves need not be given only during the QRS refractory period, and 3000 shocks may be administered at each treatment. Newer models are portable and may be used without a water bath, using only a water-filled membrane to allow propagation of the shock wave. The force and configuration of the waves required the patient to be anesthetized when using the early models; however, decreasing the size of the *F2* focal point, decreasing the power at that point, and enlarging the site of shock-wave entry have all contributed to the use of extracorporeal shock-wave lithotripsy without anesthesia. The newer machines are less powerful, and are often less efficient; some patients with hard calculi may require multiple treatments. Ideally, repeat treatments are given 1 week apart to allow edema to resolve. The current debate is whether a patient should undergo multiple treatments with shock-wave lithotripsy and perhaps have residual small fragments, or undergo a more invasive, percutaneous stone removal. Second- and third-generation lithotriptors use lower pressures and therefore a higher number of shocks per treatment is allowed. Extracorporeal shock-wave lithotripsy is used to treat stones in the kidney and ureter, and has been used to fragment bladder calculi. Because of difficulty in visualizing some calculi over the bony pelvis, and because of the increased distance of the stone from the skin in the distal ureter, shock-wave lithotripsy is more efficient in the kidney and upper ureter than in the presacral and lower ureter. Its major complications are perinephric bleeding secondary to shock-wave trauma and ureteral obstruction by fragments. The bleeding is rarely severe enough to require intervention.

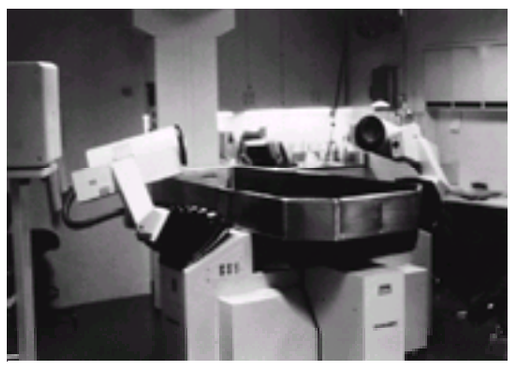


Fig. 5. First-generation Dornier HM3 Lithotripter; Massachusetts General Hospital, Boston, MA, USA.

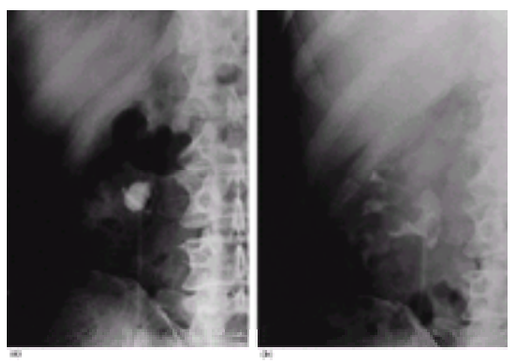


Fig. 6. (a) An abdominal film that shows a stone in the right renal pelvis; there is a stent in place that passes from the renal pelvis to the bladder.(b) The same stone after extracorporeal shock-wave lithotripsy. The small, sandy fragments are spread throughout the kidney. The stent is in place to prevent ureteral obstruction and to allow the fragments to migrate to the bladder without causing obstruction.

Open surgery

In major stone-treatment centers in the United States, open surgery for urinary calculi is used in between 1 and 5 per cent of cases. Advances in minimally invasive techniques have supplanted this time-honored form of surgery, but there are situations that still demand its use. The most common reasons to proceed to open surgery for urolithiasis are failure of less invasive methods, anatomic abnormalities involving the renal collecting system, and the presence of comorbid medical conditions in the patient. The most common open procedure is a ureterolithotomy for a larger than 1.0 cm, hard, difficult to manage, upper ureteral stone.

Bladder and urethral stones

Urethral calculi

A calculus that obstructs the urethra is infrequent, since the calculi that spontaneously migrate down the ureter are usually small enough to be passed through the urethra.

Symptoms

The symptoms of urethral obstruction secondary to a calculus are classic: the patient complains of the sudden onset of a midstream obstruction to the flow of urine. The outflow obstruction may be complete, or a decreased or split stream may be described.

Diagnosis

Palpation of the penile, scrotal, or perineal urethra will usually reveal the calculus. An abdominal radiograph that includes the genitalia will usually show the stone. Before endoscopic examination, the diagnosis may be confirmed by a retrograde urethrogram.

Treatment

Urgent relief of obstruction may be attempted by passage of a urethral catheter. The urethra should be anesthetized with 2 per cent lidocaine (lignocaine) jelly and the well-lubricated catheter passed, attempting to relieve the obstruction and manipulate the calculus to the bladder, where it may be crushed at a later date. If the catheter fails to dislodge the calculus and the patient is unable to void, relief of obstruction must be obtained by means of suprapubic drainage. The patient's urethral stone may then be treated by endoscopic methods. A cystoscope with biopsy forceps or stone-crushing forceps is placed in the urethra under regional or general anesthesia. The stone can usually be dislodged to the bladder, where it is grasped and extracted or crushed. A calculus that cannot be dislodged, manipulated, crushed or fragmented is rare. However, if this occurs, the urethra is incised longitudinally at the site of the palpable calculus, which then can be popped out. The urethra is closed with 4–0 absorbable sutures, and a stenting catheter is left in place for 7 to 10 days and removed when a urethrogram shows no extravasation. Appropriate broad-spectrum antibiotic coverage is given.

Bladder calculi

Symptoms

The symptoms of bladder calculi may include hematuria, dysuria, and urinary frequency. Except in areas of the world where protein deficiency leads to endemic metabolic bladder-stone disease, bladder calculi occur as a result of bladder-outlet obstruction, chronic bladder infection, or the presence of a foreign body (urinary catheter). When obstruction is the cause, the calculi may be composed of uric acid, calcium oxalate, or struvite (magnesium ammonium phosphate).

Diagnosis

The urinalysis will show red and white blood cells. Urine culture may be positive but is not diagnostic of a bladder calculus. The urine cytological report may show cells with inflammatory changes characterized as ‘atypical'. An abdominal radiograph will show bladder calculi only if they are radiopaque. An ultrasonographic examination of the bladder will show the shadowing of a lucent calculus that moves freely with change in patient position. An intravenous urogram may disclose the filling defect of a radiolucent calculus, but this finding may be confused with the lucent defect of a blood clot or bladder tumor. A computed tomographic (**CT**) scan will show a dense, white defect within the bladder that will migrate to the dome when the patient is placed prone.

Treatment

Bladder calculi are usually treated in conjunction with measures to relieve urinary-outlet obstruction, the most common cause of which is benign prostatic hyperplasia. Calculi of less than 3 cm in diameter can be fragmented during transurethral endoscopic procedures by any of the available devices for intracorporeal lithotripsy. Endoscopic stone-crushing forceps can then be used to fragment the stone to pieces that may be irrigated from the bladder. Calculi larger than 3 cm are more difficult to fragment and extract transurethrally, and may be best approached by suprapubic cystotomy; however, there are advocates for the use of the holmium laser, which will fragment the stone to small particles. If the suprapubic route is used, outlet obstruction may be treated by simultaneous removal of the prostate; alternatively, a suprapubic tube may be left for 7 days and a transurethral resection of the prostate done as a secondary procedure. Broad-spectrum antibiotics should be

administered, regardless of the route of treatment. Zealous treatment of a large bladder stone and simultaneous transurethral prostatectomy can lead to severe sepsis and bleeding.

Complications include perforation of the bladder from the use of the endoscopic lithotripsy device, mucosal bleeding, and systemic infection.

Ureteral stones

Ureteral calculi account for two-thirds of all calculi brought to the attention of a physician. One-half of all ureteral calculi pass spontaneously; the other half require intervention. Fig. 7 shows the likelihood of spontaneous stone passage over a period of time. Note that the width of the calculus, not its length, is the factor that correlates best with spontaneous passage. Calculi of less than 8 mm in diameter may be allowed to pass spontaneously if they are not causing pain or obstruction or are associated with an infection. However, the patient must be monitored because calculi may cause silent obstruction. As shown in Fig. 7, calculi over 5 mm in diameter may take many months to pass. Irreversible renal injury may occur after only 1 to 2 weeks of complete obstruction. One month of total obstruction can cause total loss of the kidney; therefore, following a stone's progress with abdominal radiography is not sufficient and ultrasonography must be added to rule out obstruction. As a rule of thumb, an asymptomatic, non-obstructing stone more than 5 mm wide that stays in one position for more than 3 months has a very small chance of spontaneous passage. Similarly, it must be appreciated that a calculus which becomes symptomatic while in the upper ureter has less likelihood of passing spontaneously than a similarly sized calculus discovered in the distal ureter.

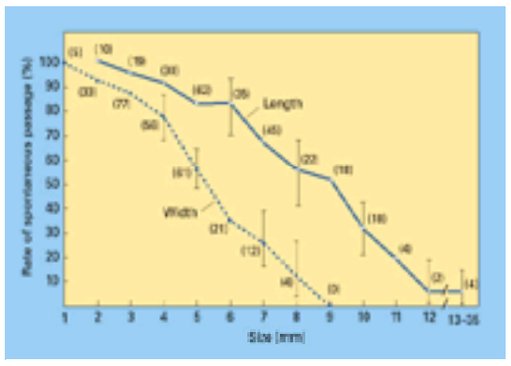


Fig. 7. Rates of spontaneous passage of ureteral calculi within 1 year in intervals of 1 mm for length and width. Numbers in parentheses indicate number of calculi passed spontaneously; vertical bars indicate 90 per cent confidence intervals.

Symptoms

Perhaps no other medical condition has as characteristic a clinical presentation as that of the colic caused by the movement of a ureteral calculus. Typically, the patient presents with the sudden onset of excruciating flank or abdominal pain, which may radiate to the scrotum or labia and/or into the ipsilateral costovertebral angle. Pain in the testes or scrotum usually indicates a stone at about L3 but may indicate a stone as high as the ureteropelvic junction, since the T12, L1 ilioinguinal nerves share a common root with the innervation of the upper ureter. Unlike the patient with a perforated viscus, who will be silent and ashen-faced, the patient passing a ureteral calculus will be writhing in agony, unable to lie quietly, and may alternatively lie, sit, or stand, and may be inconsolable. Physical examination will rarely elicit any signs of peritoneal irritation. However, calculi in the right lower ureter may be confused with inflammation of the appendix. If the ureter is partially or totally obstructed, tenderness may be present in the ipsilateral costovertebral angle. Although it is rare to have point tenderness over the course of the ureter, deep palpation may elicit pain over the area of the impacted stone.

Diagnosis

Except in the rare condition of total urinary obstruction, the urinalysis will show red blood cells. The diagnosis of renal colic in the absence of red cells in the urine should be carefully questioned. White blood cells may not be seen unless there is an associated infection.

An abdominal radiograph will show the presence of the stone along the course of the ureter in 60 to 70 per cent of patients. The absence of a calculus on the abdominal film does not exclude the diagnosis of a ureteral calculus: the calculus may be radiolucent or poorly radiopaque, may be obscured by the bony pelvis or overlying bowel contents, or be confused with pelvic phleboliths. The distinction between a phlebolith and a calculus is made on the basis that calculi are irregular and solid while phleboliths are round and often have a lucent centre. An intravenous urogram is the classic non-invasive study that is used to confirm the diagnosis. There is often delayed function on the side of the calculus, 1 to 2 h or more being needed for the contrast to concentrate on that side. Follow-up films should be taken hourly until the contrast columnates to the site of the obstruction. Calculi over the lumbar transverse processes or the sacrum may be obscured until contrast medium reaches that level. If contrast moves into the retrovesical ureter, the overlying contrast-filled bladder may obscure the distal ureter. The patient must be instructed to void and an oblique film of the pelvis obtained. If the stone can be identified on the initial abdominal film, if it is less than 5 mm in diameter (high likelihood of passage), and if the patient's symptoms abate, the intravenous urogram may be postponed and obtained later if symptoms recur or if there is no spontaneous passage over an appropriate period of observation.

At many major medical centers, the initial radiographic study of choice for suspected urolithiasis is the non-contrast, 5-mm cut, spiral CT scan. The spiral CT rapidly acquires data and therefore is not prone to missing small areas of the torso due to respiratory motion. All stones, whether radiolucent or opaque on plain film, are easily visualized on CT scan. The finding of a high-density body in the expected course of the urinary tract, along with secondary signs of an enlarged kidney, hydronephrosis, or perirenal or periureteral stranding, confirms the diagnosis. If there is doubt about whether the visualized object is in the urinary tract, contrast can be administered and the patient rescanned. Using this protocol an accuracy of over 95 per cent has been achieved. Furthermore, other causes of abdominal or flank pain, such as aortic aneurysm, gallbladder disease, and diverticulitis, can be diagnosed or suspected from a CT scan. Recent research suggests that the composition of the stone may be deduced from the Hounsfield units it registers, with uric acid having readings in the 200 to 400 range, calcium oxalate in the 900 to 1300 range, and cystine in the 600 to 700 range.

Treatment

If the patient has a high-grade obstruction from a calculus that is unlikely to pass, remains symptomatic, or is infected, intervention is required. The method of intervention depends on the equipment available, the experience of the operator, and the position of the stone. The American Urological Association (**AUA**) recently convened a panel of experts to make recommendations for the management of ureteral calculi based on a thorough evaluation of the peer-reviewed literature on the subject and their personal expertise. They evaluated extracorporeal lithotripsy, endoscopy, and open surgery for their stone-free rates, complications, and need for secondary interventions, and made generalized treatment recommendations based on stone size and location.

Upper ureteral calculi

These are defined as calculi lodged between the ureteropelvic junction and the narrowing of the ureter as it courses over the iliac vessels. The AUA panel recommends extracorporeal shock-wave lithotripsy, if available, as the first line of treatment in most patients with stones under 1 cm. The reported stone-free rate with this treatment is 80 to 85 per cent. Unless there is sepsis or high-grade obstruction, these calculi may be fragmented *in situ* without first being bypassed by the endoscopic placement of a ureteral stent. While stenting may help to manage the symptoms of the passage of fragments, there is no evidence that it improves their clearance. Ureteroscopic management of upper ureteral calculi by electrohydraulic lithotripsy or laser may be undertaken by experienced endoscopists when shock-wave lithotripsy is not available or has failed; however, freedom from stone is obtained in only 50 to 60 per cent of cases reported in the literature. Either the stone cannot be reached or it becomes only partially fragmented and is dislodged and migrates to the kidney. If extracorporeal shock-wave lithotripsy and an experienced endoscopist are not available, a ureterolithotomy through a subcostal or dorsal lumbotomy approach is the most definitive and most successful treatment for an impacted upper ureteral stone. In the presence of fever or urosepsis, percutaneous nephrostomy is the safest and most definitive way to drain the obstructed infection, with treatment of the stone after an adequate course of antibiotics.

Upper ureteral calculi larger than 1.0 cm may be treated either by shock-wave lithotripsy or percutaneous nephrostolithotomy, with equivalent success rates in the range of 70 per cent. Retrograde ureteroscopy has a success rate in this population of less than 50 per cent as residual calculi frequently migrate to the kidney.

Success with these larger stones is usually dependent on their fragility.

Lower ureteral calculi

The AUA panel recognizes both shock-wave lithotripsy (in the prone position to allow placement of the lower ureteral stone at the focal point of the shock waves) and ureteroscopy with intracorporeal fragmentation (Fig. 8)or basketing as acceptable treatment options. The success rate with extracorporeal shock-wave lithotripsy is only 80 per cent, with most failures because of poor radiographic visualization. Experienced endoscopists are able to perform ureteroscopy with stone fragmentation or removal in 90 to 99 per cent of these patients. The high success rate and low cost have resulted in ureteroscopy being the most commonly used method of intervention for an impacted lower ureteral stone. Newer generations of extracorporeal shock-wave lithotriptors may shift the weight to non-invasive therapy for all but the large and most impacted calculi. Rarely, when these techniques fail or are unavailable, surgical removal through a Gibson or transvesical incision is performed. Again, in the presence of infection, the obstruction should be relieved by indwelling stent of percutaneous nephrostomy and the infection treated before manipulation or treatment of the stone.

‘Steinstrasse'

This is a new medical term and literally means ‘stone street.' It is a condition that follows the use of extracorporeal shock-wave lithotripsy for renal or ureteral calculi. Small pieces of fragmented calculi collect and obstruct in the distal ureter, like sand occluding a straw (Fig. 9). The endoscopic approach to this condition is difficult: ureteroscopic resolution may take many hours and result in ureteral injury. The most prudent management of pain, sepsis, or obstruction caused by ‘steinstrasse' is to place a percutaneous nephrostomy tube and allow the fragments to pass spontaneously. If the condition is not resolved in 4 to 6 weeks, endoscopic removal is performed.



Fig. 9. Steinstrasse': the right lower ureter is filled with fragments following extracorporeal shock-wave lithotripsy; these are stone fragments that fill the ureter like sand in a straw and may cause ureteral obstruction.

Renal stones

The management of renal calculi depends on the patient's symptoms, the size and suspected composition of the calculus, and its position within the kidney.

Symptoms

The patient with a renal calculus may present with flank or abdominal pain, fever, and/or gross or microscopic hematuria. If infection is present, pyuria will be seen. Although the patient may be in considerable pain, unless the stone is impacted below the ureteropelvic junction the degree of colic does not approach that seen with a ureteral stone. Physical examination may reveal flank tenderness over the costovertebral angle or anteriorly over the lower pole of the kidney. A tense, swollen kidney is tender to deep palpation. Severe sepsis and shock may occur as a result of a stone obstructing the outflow of infected urine.

Diagnosis

An ultrasonographic examination will almost always show the presence of a renal calculus and is useful in determining the degree of obstruction; however, the size of a renal stone is difficult to ascertain by ultrasound and an abdominal radiograph is required. An intravenous urogram may then be performed to determine the position of the stone within the kidney and the presence of any associated anatomical abnormalities, such as obstruction at the ureteropelvic junction, ureteral obstruction, or stenotic infundibulum. As with acute colic caused by ureteral stones, we have begun to use the spiral CT scan to diagnose renal stones and to plan treatment. As previously described, information about stone size, renal anatomy, and perhaps stone composition can be obtained swiftly and accurately with this method. Its chief disadvantage in the diagnosis of renal stones is that it does not give as detailed a map of the collecting system as is desirable when planning therapy for large or calyceal stones.

Intervention is indicated for significantly symptomatic, infected or obstructing calculi, or those that show progressive growth on serial evaluations. Renal calculi of 6-mm diameter have a 60 per cent chance of causing a symptomatic episode within 5 years.

Treatment

Extracorporeal shock-wave lithotripsy is the most successful treatment for most renal calculi of less than 2.5 cm in their greatest diameter in patients who do not have obstruction to the egress of fragments. The exceptions are cystine calculi, very radiodense pure calcium oxalate monohydrate calculi, and calcium phosphate (brushite) calculi of more than 1.5 cm in diameter. In these exceptions, or for renal calculi that are proximal to a narrowed infundibulum or ureteral obstruction, and for calculi larger than 2.5 cm, percutaneous nephrostolithotomy is the preferred method of treatment. The success of shock-wave lithotripsy depends on both the size and fragility of the calculus.

Stones made of uric acid may be dissolved *in situ* by alkalinization of the urine with sodium bicarbonate or other available alkalinizing agents such as potassium citrate. The patient should keep the urine pH greater than 6.5 to achieve sufficient alkalinization for dissolution. A 1- to 2-cm calculus may take 6 to 8 weeks to dissolve. The usual dose of potassium citrate is 20 mEq four times a day. The addition of 300 mg of allopurinol will decrease uric acid saturation.

Struvite stones must be treated and eliminated in order to prevent recurrent infection and regrowth of stone. An AUA guidelines panel recommends percutaneous nephrostolithotomy followed by shock-wave lithotripsy or repeat percutaneous endoscopy for residual fragments as the standard approach to struvite staghorn calculi, and either percutaneous nephrostolithotomy or shock-wave lithotripsy for small struvite stones in normal kidneys.

In a patient with a poorly functioning or non-functional kidney with a symptomatic stone, nephrectomy is a reasonable alternative.

If required, the open surgical approach depends on the size and position of the calculus. Calculi in the renal pelvis are approached via an 11th or 12th rib incision and removed via a pyelotomy. If a branch of the calculus extends to an infundibulum and calyx, the pyelotomy incision can be extended into the corresponding infundibulum. If the stone cannot be completely removed by this method, a nephrotomy may be used to expose it. Staghorn calculi with narrow infundibulums and multiple calyces filled with stone may require surgical treatment by anatrophic nephrolithotomy (Fig. 10).

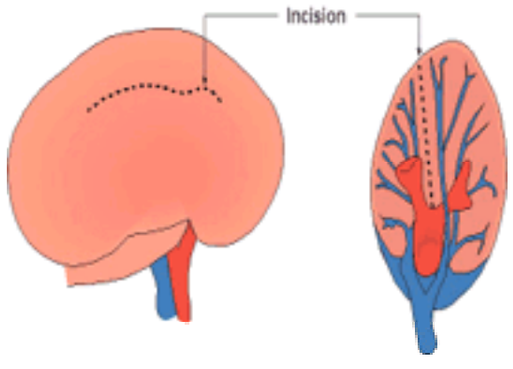


Fig. 10. Anatrophic nephrolithotomy: lines of incision along the posterior segmental blood supply.

Summary

The role of the surgeon in the management of stone disease is to relieve obstruction and eradicate the offending calculus. Technological advances such as extracorporeal shock-wave lithotripsy, miniaturization of endoscopes, and better sources of intracorporeal lithotripsy have made this job safer and more effective.

Further reading

Chaussy C, Schmiedt E, Jocham D, Brendel W, Forssmann B, Walther V. First clinical experiences with extracorporeally induced destruction of kidney stones by shock waves. *Journal of Urology* 1982; **127**: 417–20.

Coe FL, Keck J, Norton ER. The natural history of calcium urolithiasis. *Journal of the American Medical Association* 1977; **238**: 1519.

Furlow WI, Bucchiere JJ. The surgical fate of ureteral calculi: review of Mayo Clinic experience. *Journal of Urology* 1976; **116**: 559.

Herring LC. Observations of 10 000 urinary calculi. *Journal of Urology* 1962; **88**: 545.

Prien EL. Composition and structure of urinary stone. *American Journal of Medicine* 1968; **45**: 654.

Segura JW, Patterson DE, LeRoy AJ, May GR, Smith LH. Percutaneous lithotripsy. *Journal of Urology* 1983; **130**: 1051–4.

Segura JW *et al.* Nephrolithiasis Clinical Guidlines Panel summary report on the management of staghorn calculi. *Journal of Urology* 1994; **151**: 1648.

Segura JW *et al.* Ureteral Stones Clinical Guidelines Panel summary report on the management of ureteral calculi. *Journal of Urology* 1997; **158**: 1915.

Ueno A, Kawamura T, Ogawa A, Takayasu H. Relation of spontaneous passage of ureteral calculi to size. *Urology* 1977; **10**: 544–6.

39.6 Male sexual dysfunction and infertility

Lewis Kriteman and Jerry Yuan

Physiology of erection
Incidence and epidemiology
Pathogenesis and etiology
- Psychogenic disorders
- Neurogenic disorders
- Endocrinologic disorders
- Arteriogenic disorders
- Cavernosal disorders
- Other causes

Evaluation
Management—non-surgical
- Oral agents
- Intracavernous injections
- Intraurethral agent (MUSE)
- Vacuum device

Management—surgical
- Penile prostheses
- Vascular surgery

Infertility
- History
- Physical examination
- Semen analysis
- Endocrine evaluation
- Secondary andrology testing
- Radiographic studies for male infertility
- Testicular biopsy
- Genetics of male infertility
- Assisted reproductive technique in male infertility
- Treatment of male infertility

Further reading

Male sexual dysfunction is defined as the inability to achieve satisfactory sexual relationships, and encompasses problems with erection, emission, ejaculation, or orgasm. This section of the chapter will focus on the first of these problems, erectile dysfunction, which strictly implies the inability to achieve or maintain an erect penis with sufficient rigidity for intercourse. With increasing understanding of the physiology of penile erections, there have been great advances made over the past 20 years in the diagnosis and treatment of erectile dysfunction.

Physiology of erection

The processes culminating in a penile erection involve the complex interactions between psychologic, neurologic, hormonal, and vascular factors. A fundamental knowledge of penile physiology is essential to the diagnosis and treatment of erectile dysfunction. The penis is innervated by the autonomic and somatic nervous systems. The parasympathetic nerves, originating in the intermediolateral nucleus of the spinal cord at levels S_2 to S_4, travel down the pelvic nerve via the pelvic plexus innervating the penis through the cavernous nerves. Penile sensation is carried back to the spinal cord via the pudendal nerve. The major blood supply to the penis is from the paired internal pudendal arteries which are terminal branches of the hypogastric arteries. The flaccid corpora cavernosum exhibit contracted arteries, arterioles, and sinusoids. Smooth muscles in the arteriolar tree are the key to erections. With stimulation of the parasympathetic nerves to the penis, there is relaxation of the sinusoidal smooth muscles and arteriolar vasodilatation resulting in increased arterial blood flow into the corpora. As the sinusoidal system fills with blood and expands, its walls push up against the tunica albuginea trapping and compressing the subtunical venular plexus. With further stretching of the tunica, the emissary veins are collapsed thereby effectively reducing the venous outflow from the penis. Recently the neurotransmitter responsible for sinusoidal smooth muscle relaxation has been identified as endothelium-derived relaxing factor; this endothelial-dependent relaxation is mediated by nitric oxide and an accumulation of cyclic guanosine monophosphate (**cGMP**) (Fig. 1).

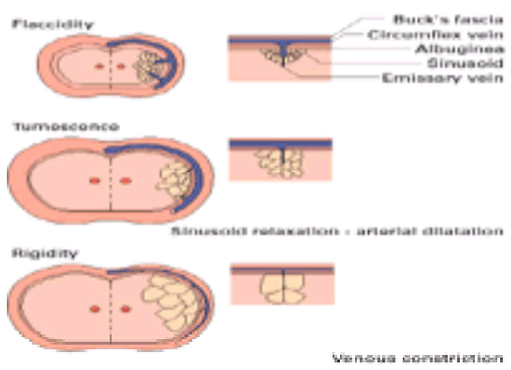


Fig. 1. Physiology of penile erection. During tumescence, the smooth muscles are relaxed, producing arterial dilation and filling of the sinusoidal spaces. These spaces compress the subalbugineal venous plexus, reducing outflow through the emissary veins and increasing intracavernous pressure. At rigidity the whole deep venous system outside the corpora is compressed between the tunica albuginea and Buck's fascia.

Incidence and epidemiology

The incidence of impotence increases with age and was first noted by Kinsey in 1948: at age 40 only 1 in 50 men were impotent, whereas at age 65 the frequency increased to 1 in 4 men. It is estimated that impotence affects 10 to 20 million men in the United States alone. Some investigators report that 35 per cent of married men over the age of 60 suffer from impotence. Another large study classified impotence as minimal, moderate, and complete, and found a combined prevalence of 52 per cent for men aged between 40 and 70 years who were not in institutions. In this latter review, subject age was the most important variable followed by heart disease, hypertension, diabetes, medications, and level of anger and depression. Feldman and colleagues found that subjects with hypertension and heart disease who also smoked cigarettes had a much greater incidence of complete impotence.

Pathogenesis and etiology

There have been many different systems proposed to classify erectile dysfunction. We offer a system which combines the various causes of impotence with current understanding of physiology and anatomy.

Psychogenic disorders

In the 1950s more than 90 per cent of impotence was attributed to psychogenic causes. Today most physicians would consider these disorders to be responsible for 10 to 30 per cent of cases. Sexual behavior is controlled by the cerebral cortex, limbic system, and hypothalamus, which all play a role in sending both stimulatory and inhibitory messages to the erection center in the spinal cord. These higher centers can have a direct inhibitory effect on the erection center, and can cause release of excess catecholamines which increase penile cavernosal smooth muscle tone.

Neurogenic disorders

This group of causes can be simply categorized as lesions of the brain, the spinal cord, the peripheral nerves, and neuropathies. Parkinson's disease, Alzheimer's disease, cerebrovascular accidents, tumors, and epilepsy all have been implicated to some degree in both loss of libido and impotence. Spinal cord lesions encompass trauma, multiple sclerosis, spina bifida, syringomyelia, tabes dorsalis, amyotrophic lateral sclerosis, and cord compression from a herniated disk or tumor. Reflexogenic erections have been observed in 95 per cent of patients with complete upper motor neuron lesions. However, potency is preserved in lower motor neuron lesions in only 25 per cent of patients. Diabetic peripheral neuropathy with resultant microangiopathy and impotence is a common etiologic factor. Some authors believe that diabetes may affect the cavernous nerve terminals by depleting their stores of neurotransmitters, especially nitric oxide. Alcoholism and vitamin deficiency have also been found to decrease neurotransmitter release in the cavernous nerves. The incidence of peripheral nerve injury as a cause of impotence is quite high. Surgical procedures including radical prostatectomy, abdominal perineal resection, and external sphincterotomy have all been implicated in iatrogenic impotence. Finally, trauma to the pelvis with resultant cavernous nerve or artery damage can also cause erectile dysfunction.

Endocrinologic disorders

Hormonal deficiencies and imbalances can certainly suppress sexual interest and impair erections. Hypogonad states of any kind can be associated with decreased libido and the frequency of nocturnal erections. Additionally, hyperprolactinemia, hyperthyroidism, hypothyroidism, Cushing's syndrome, and Addison's disease have likewise been implicated. Patients with low testosterone levels do derive some benefit from exogenous replacement. However, those patients with borderline levels demonstrated no improvement of potency with testosterone therapy.

Arteriogenic disorders

Any damage to the hypogastric or cavernosal arteries, either from trauma or long-standing atherosclerotic disease, decreases the arterial perfusion pressure and flow to the penile erectile tissue. The incidence and age of onset of coronary artery disease and erectile dysfunction are parallel. Most often arteriogenic impotence is a component of systemic arterial disease, and as such shares common risk factors including hypertension, hyperlipidemia, cigarette smoking, and diabetes mellitus.

Cavernosal disorders

An impaired venous occlusive mechanism has been proposed as one of the most common vasculogenic causes of erectile dysfunction. The development of large venous channels and degenerative changes in the tunica albuginea resulting in inadequate compression of the emissary veins have both been implicated as causative factors. Additionally, structural alterations in the cavernosal smooth muscle and insufficient trabecular smooth muscle relaxation have been observed in these patients. Cigarette smoking may induce local vasoconstriction and penile venous leakage thus contributing to dysfunction.

Other causes

Various classes of medications can cause impotence. Almost all antihypertensives, especially the centrally acting drugs (methyldopa, clonidine, reserpine, and guanethidine), have been implicated. The a- and b-adrenergic blocking agents, thiazide diuretics, and spironolactone all can cause impotence. Antidepressants, tranquilizers, and antipsychotics are also included. Chronic medical conditions such as renal insufficiency, severe pulmonary and heart disease, and cirrhosis have been associated with an increased incidence of impotence.

Evaluation

A thorough history and physical examination are the mainstay in evaluating a patient with erectile dysfunction. Specific questions should be asked concerning the patient's perception of his impotence. Moreover, reduction in the level of libido may point to an endocrinologic etiology. If possible, an interview with the partner is often helpful in confirming the patient's history. The duration and nature of onset of sexual dysfunction, the presence or absence of early morning erections, and intermittency often are helpful in differentiating organic from psychogenic etiologies. Sudden onset of impotence, especially associated with any psychologic stress, often points towards a psychogenic cause. Many physicians incorporate into their initial evaluation the concept of a goal-directed therapeutic treatment plan as described by Lue and associates. The various methods of treatment for the erectile dysfunction are presented with particular attention shown to the couple's understanding of the therapeutic options, their reasonable expectations of each, and any associated side-effects involved.

Erectile dysfunction may be multifactorial. A thorough medical history and a complete list of medications often helps the physician determine appropriate further work-up and treatment recommendations. A history of coronary or peripheral vascular disease, diabetes, chronic renal failure, tobacco and alcohol use, psychologic, neurologic, or chronic debilitating disease all can contribute to, or be the primary cause of, impotence. Changes in a patient's antihypertensive medication may be all that is necessary to improve his potency. Pertinent surgical history such as an abdominoperineal resection should be noted.

Physical examination should focus on the genitals with regard to penile plaques and chordee, and the testicles. A thorough rectal examination should always be performed to assess for prostate cancer, prostatitis, and to check the rectal tone. Findings of gynecomastia and atrophic testes should prompt an endocrine evaluation for hypogonadism or hyperprolactinemia. A careful neurologic examination may show evidence of peripheral neuropathy, and a deficiency in perineal sensation or lack of a bulbocavernosus reflex may also imply a neurologic disorder. Davis-Joseph and colleagues found that history and physical examination alone yielded a 95 per cent sensitivity with regards to organic erectile dysfunction but only a 50 per cent specificity. Abnormalities detected during this directed examination may guide the physician to the best approach in further evaluation and treatment.

Laboratory testing should be included in the initial evaluation to identify any potentially contributing conditions. Laboratory evaluation should include a complete blood count, serum chemistries including renal and hepatic function, and urinalysis. The level of prostatic specific antigen should be measured in all men over the age of 50 years. Initial hormonal evaluation generally includes measurement of free and total testosterone. A low testosterone level should be verified with an early morning repeat prior to instituting replacement therapy. Abnormal testosterone values should prompt further evaluation with prolactin, luteinizing hormone (**LH**), and follicle-stimulating hormone (**FSH**) levels to rule out hyperprolactinemia and to differentiate primary from secondary hypogonadism.

Following the initial work-up, the patient and physician can then decide whether to pursue further evaluation for an organic disorder, or a referral for psychologic counseling. Some practitioners, however, elect to perform nocturnal penile tumescence testing or formal sleep studies to assess nocturnal penile erections during rapid eye movement sleep. Presumably, a normal nocturnal penile tumescence test suggests that the patient has the capacity for erotically induced erections, and may identify those men with psychogenic impotence. However, several investigators have discounted the usefulness of nocturnal penile tumescence testing, and report that sexual performance correlates poorly with results of this.

Evaluating patients with an intracavernous injection of a vasodilating agent may also be helpful in assessing the vascular status of the penis. A normal erection results from the complex process of increased arterial inflow, normal sinusoidal relaxation, and appropriately diminished venous outflow in response to the pharmacologic agent. An adequate erection following low-dose injection supports the diagnosis of either psychogenic impotence or mild organic impotence with intact penile vasculature, whereas no response to injection with a high dose of a single or multiagent mixture strongly suggests a significant vascular etiology which may require penile prosthesis placement. Additional work-up may include duplex ultrasonography to assess the arterial inflow, and cavernosometry to diagnose veno-occlusive erectile dysfunction. Finally, arteriography is indicated in the young patient whose impotence results from pelvic trauma with arterial disruption, and is used preoperatively to obtain an accurate view of the anatomy prior to revascularization (Fig. 2).



Fig. 2. Detailed algorithm for erectile dysfunction.

Management—non-surgical

The etiology of erectile dysfunction may be multifactorial. Often it is difficult to differentiate organic from psychogenic components. Once the reversible causes of impotence have been addressed and have been treated appropriately, and those patients with psychogenic impotence have been referred for counseling, the question of further management must be addressed. Over the last two decades the emergence of new medications has changed the way we manage impotence. There now is a much larger role for non-surgical treatments, which include the following,

Oral agents

Adrenergic, dopaminergic, and seratonergic receptors exist in the brain centers involved with erection, libido, and ejaculation. Most patients prefer to try the oral medications as their initial treatment. Yohimbine, an a-adrenergic antagonist, is a widely prescribed drug which has been shown to be no more effective than placebo in improving potency. Nevertheless, this medication continues to be used in patients with mild to moderate impotence. Trazadone is an antidepressant which has a positive effect on nocturnal penile tumescence and sexually stimulated erections. Some efficacy has been shown with the combination of yohimbine and trazadone. Recently a new medication, sildenafil (Viagra), which acts by inhibiting the breakdown of cGMP and causes penile vascular smooth muscle relaxation, has been approved in the United States. Preliminary studies have demonstrated a 75 per cent efficacy rate.

Intracavernous injections

The use of intracavernous pharmacologic agents for the diagnosis and treatment of impotence has become very popular. The most serious adverse effects of injection therapy are priapism and the development of penile plaques. The incidence of priapism ranges from 5 to 10 per cent depending on the dose and agent(s) injected. The agents most commonly used are papavarine, phentolamine (Regitine), and prostaglandin E_1 (Alprostadil), singly or in combination. Papavarine inhibits phosphodiesterase which leads to increased cyclic adenosine monophosphate resulting in vasodilatation. It is inexpensive and stable at room temperature, but there is a high incidence of priapism and fibrosis formation. Phentolamine is often injected with papavarine, and works as a competitive a-adrenergic receptor inhibitor. Prostaglandin E_1 causes smooth muscle relaxation and vasodilatation. This medication has been shown to have a higher response rate with less priapism and fibrosis. A combination of all three medications has been formulated which may be effective in patients unresponsive to single-agent injections.

Intraurethral agent (MUSE)

Recently a new delivery system has enabled prostaglandin E_1 to be administered in the form of a urethral pellet. The alprostadil pellet is placed in the urethra and dissolves across the urethral membrane into the surrounding vascular tissue. MUSE may be effective with either mild to moderate organic impotence or psychogenic impotence. Initial reports found this treatment to be successful in two-thirds of patients with a very low incidence of side-effects.

Vacuum device

This device consists of a plastic cylinder connected to either a hand or electric pump, and a rubber constricting ring, and is useful in almost all cases of erectile dysfunction. Negative pressure is used to achieve venous engorgement of the penis at which point the ring is placed at the base of the penis to prevent venous outflow. A combination of intracavernous injection with the vacuum device and constriction ring may enhance the erection in the more difficult patients. The majority of men using these devices report adequate erections for coitus with satisfaction rates ranging from 68 to 83 per cent.

Management—surgical

Non-surgical options for treatment of erectile dysfunction should be considered and completely explored with the patient prior to any surgical procedure. The two major categories of surgical intervention include placement of a penile prosthesis and vascular surgery.

Penile prostheses

Penile prostheses fall into two broad categories: the malleable or semirigid prosthesis, and the inflatable prosthesis (Fig. 3). Placement of any device, however, precludes future physiologic erections. This point should be made clearly to the patient and his partner. Additionally, no prosthesis will restore the full length or girth of erections compared with the natural erection. The two major complications from prosthesis placement are mechanical failure and infection. With the newer devices now available, reoperation secondary to device failure has decreased substantially, and can be expected in 5 per cent of cases when the device has been in place for 5 to 10 years. Semirigid devices usually are less prone to mechanical failure. Certain patients are at a greater risk for developing prosthesis-related infection from their surgery. These include patients with spinal cord injuries, insulin-dependent diabetes, and those undergoing reoperation. Patient and partner satisfaction with these devices ranges from 60 to 80 per cent.

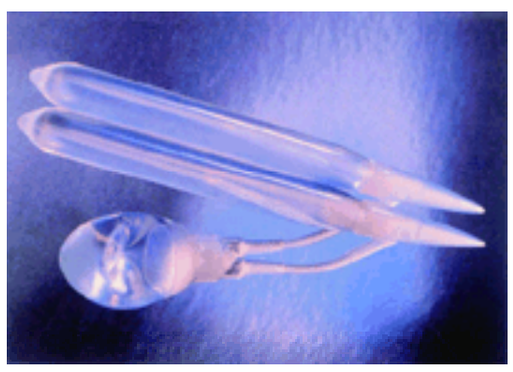


Fig. 3. Inflatable penile prosthesis (by courtesy of the Mentor Corporation, Goleta, California).

Vascular surgery

Vascular surgery for erectile dysfunction can be divided into two major areas: penile arterial revascularization and surgery for veno-occlusive disorders. The selection criteria for vascular surgical procedures should be rigorous. Patients who are best suited for revascularization procedures are younger men who have sustained pelvic trauma with a discrete arterial lesion noted on arteriography. Most patients with arteriogenic impotence have diffuse atherosclerotic changes and are not candidates for a revascularization procedure. The success rate of this type of procedure varies from 50 to 70 per cent even in the most experienced hands. An additional 10 to 20per cent of patients will achieve potency after surgery by the help of intracavernous injections. Most penile arterialization procedures utilize one of the inferior epigastric arteries which is anastomosed to the dorsal penile artery, the deep dorsal vein, or a combination of both vessels.

Strict selection criteria should also be stressed prior to recommending surgery for a veno-occlusive disorder. These patients tend to have erections that cannot be maintained with lack of response to intracavernous injections despite normal arterial duplex studies. The faulty veno-occlusive mechanism can then be assessed with cavernosometry and cavernosography. Success rates range from 40 50 per cent.

Infertility

Infertility, as defined by the inability to conceive after 1 year of unprotected intercourse, affects 15 per cent of all couples. Male factor infertility is being increasingly recognized, and may be contributory in half of all infertility cases. History, physical examination, and semen analysis are the initial steps in the evaluation of male infertility; other tests such as sperm functional assays, antisperm antibody screening, genetic testing, and testis biopsy are performed as indicated. The causes for male infertility are diverse and can be divided into three major categories: (i) hypogonadotropic, or pretesticular, infertility due to hypothalamic or pituitary deficiency which results in testicular understimulation; (ii) testicular deficiency problems with reduced or absent spermatogenesis due to intrinsic spermatogenic defects, gonadotoxins, or varicocele; and (iii) obstructive, or post-testicular, infertility with adequate spermatogenesis but complete or partial blockage along the male reproductive track.

Evaluation of male infertility is straightforward in most men and can be expediently accomplished. A detailed history and physical examination should be the first step, and a dedicated infertility questionnaire is an invaluable tool used in the patient's initial visit (Fig. 4 and Fig. 5).

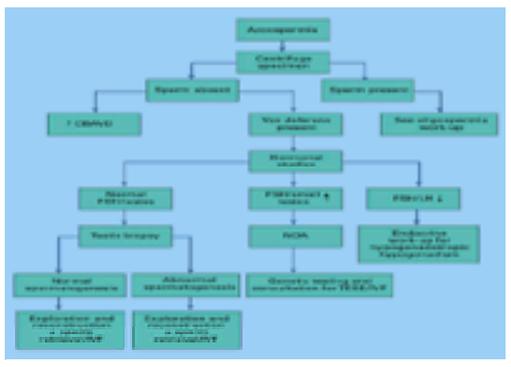


Fig. 4. Detailed algorithm for azoospermia.

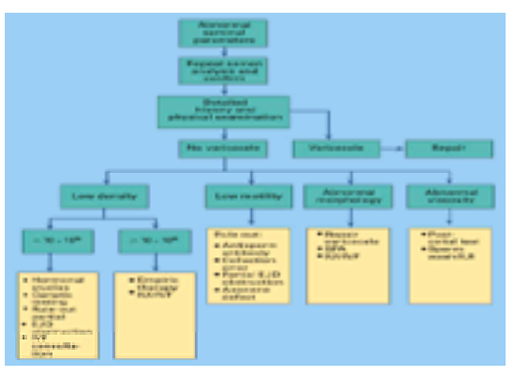


Fig. 5. Detailed algorithm for oligospermia.

History

Certain factors in a patient's history should be elicited during his initial visit. Particular attention should be shown to the following.

1. Previous fertility by either partner would suggest a causative factor in the other; in men who have fathered children in the past and are unable to do so now, three conditions tend to be more common: varicocele because of its progressively deleterious effect on spermatogenesis, acquired obstruction, or female factors.
2. A sexual history should include the frequency and timing of intercourse, and the use of lubricants which may be toxic to sperm.
3. A medical history should include all past and present illnesses and the associated treatments, all of which may adversely affect fertility. Illicit drug use and testicular toxin exposure, including radiation, may result in severe and possibly irreversible testicular injury. Medications such as cimetidine, sulfasalazine, and calcium channel blockers have also been implicated in male-factor infertility.
4. A surgical history should include any procedures which may result in vasal obstructions such as hernia or hydrocele repairs, and those procedures which may result in abnormal ejaculation such as retroperitoneal lymph node dissection for testicular cancer or bladder neck surgery for voiding dysfunction.
5. The female partner's history is also considered; significant female factors may influence the extent of the male factor evaluation and treatment options. Additionally, advancing female age with tubal obstruction and concurrent obstructive azoospermia in the male is better served with *in vitro* fertilization (**IVF**) rather than surgical reconstruction in both.

Physical examination

The male reproductive system is readily examined, and a detailed physical examination is critical. Often the physical findings demonstrate the diagnosis, and eliminate the need for further tests. The key aspects of the male infertility examination include the following.

1. The overall body habitus: is the patient well developed with normal secondary male characteristics?
2. Are there any scars in the abdomen, groin, or scrotal area?
3. Does the phallus appear normal? Urethral strictures may interfere with ejaculation, and hypospadias with chordee may prevent proper vaginal penetration and deposition of semen.
4. The testes: one should note the number, location, size, consistency, and the presence of any mass, asymmetry, and tenderness. Azoospermia with normal testes would suggest adequate spermatogenesis and obstruction, whereas atrophic testes would suggest the opposite.
5. The epididymis and vas deferens: vasal atresia, either unilateral or bilateral, may be seen in carriers of cystic fibrosis or in those with a mesonephric anomaly with coexisting renal malformation. In men with normal-volume azoospermia and palpably normal testes, the presence of epididymal distention strongly suggests an obstructive etiology whereas a collapsed or normal-feeling epididymis would suggest a defect in spermatogenesis.
6. The spermatic cords: varicocele, or venous dilation of the pampiniform plexus of the spermatic cord, may be found in 40 to 70 per cent of subfertile men. Most are on the left side with infrequent bilaterality. The diagnosis of a varicocele is established by palpation of the spermatic cord with the patient supine and standing. Examining the patient after prolonged standing or during Valsava with a pencil Doppler device may be helpful in patients with a difficult examination due to retractile testes or thickened cords.
7. Rectal examination: prostatic induration or mass may be present in men with chronic prostatitis or cystic seminal vesicles; either may result in ejaculatory duct obstruction.

Semen analysis

Semen analysis is an integral part of the male infertility evaluation and the specimen should be collected by masturbation following 2 to 3 days of abstinence. Minimal standards of adequate seminal parameters have been established (Table 1). Human spermatogenesis spans 3 months and is easily disturbed; semen analysis is better deferred until 3 or more months following any known illness. Two or more semen analyses with consistent bulk parameters within 20 per cent of each other should be obtained prior to drawing any conclusion. Azoospermic samples should be centrifuged and the pellet examined to rule out extreme oligospermia. The key components of a complete semen analysis are as follows.

Ejaculate volume	1.5–5.0ml
Sperm concentration	> 20 million/ml
Viability	60%
Motility (forward progression)	> 60%
Morphology	> 60% normal forms
Presence of:	Fructose
Absence of:	Agglutination
	Pyospermia

Table 1 Normal values for semen analysis

Volume

Low ejaculate volume, 1.5 ml or less, may be due to a collection error including an insufficient abstinence period, a hypoandrogenic state, or ejaculatory disorders. Retrograde ejaculation is diagnosed by examining the postejaculate urine for sperm. Absence of seminal fructose also suggests an obstructive process at the ejaculatory duct level or seminal vesicle atresia.

Sperm density

Adequate sperm density, or sperm count, has been adjusted downward to 20 million/ml and is based on statistical analysis of natural pregnancy rates in a large study population. Normal sperm density is significantly higher at 70 to 80 million/ml and patients should be advised of this important difference. Density multiplied by motility percentage, or total motile sperm count, is the most accurate predictor of pregnancy rate.

Motility

Sixty per cent or more sperm should demonstrate active forward progression (grade 3 to 4 on a 0 to 4 scale). Motility should also be sustained when observed over time. Reduced motility, or asthenospermia, may result from a variety of causes including antisperm antibody, genital tract infection, and immotile cilia syndrome. One needs to ensure the specimen was properly collected, examined in a timely manner, and was not exposed to extreme heat or cold. Viability staining is required to differentiate non-viable sperm, or necrospermia, from asthenospermia.

Morphology

The significance of morphology is less well understood. Pathologically defined morphologic defects such as acrosome-deficient round-headed sperm or tail stump syndrome are established causes for male infertility and are rare. Abnormal morphology, or teratospermia, may be reflective of disturbed spermatogenesis and is often observed in conjunction with other seminal abnormalities.

Pyospermia

Most seminal round cells are immature germ cells; distinguishing these from white blood cells (pyospermia) requires special staining. Pyospermia may be seen with genital tract infections; however, in the absence of infection, the significance of pyospermia is uncertain although the release of potentially toxic reactive oxygen species is one possible mechanism by which pyospermia may adversely affect sperm function.

Endocrine evaluation

Male infertility due to isolated endocrine abnormality is infrequent. Hormonal evaluation for FSH and testosterone is indicated in men with severe oligospermia (< 10 million/ml), or in those with a history and physical finding suggestive of a central neurologic process including headache, anosmia, or erectile dysfunction with loss of libido (e.g. hyperprolactinemia). Elevated FSH suggests primary testicular failure, especially if it is more than twice the normal value. Low FSH/LH/testosterone is seen with hypothalamic or pituitary deficiency with testicular hypostimulation. Pituitary MRI for tumor and endocrinology consultation should also be considered in these latter patients.

Secondary andrology testing

A large number of tests are currently available to assess various aspects of sperm function; however, most subfertile men will not require these tests and the associated expense.

Sperm penetration assay

A sperm penetration assay examines the physiologic ability of sperm to transform and interact with eggs. Zona pellucida-free hamster eggs are used as targets. Poor assay results may be due to defects in various steps required for sperm–egg interaction and may be predictive of poor intrauterine insemination or conventional IVF outcome.

Hemizonal assay

A hemizonal assay is similar to a sperm penetration assay while substituting a human zona pellucida for the hamster eggs. The zona is divided in half; each half is then incubated with sperm from either the patient or a fertile donor and compared for sperm–zona binding.

Antisperm antibody

Sperm-bound (direct) IgA and IgG antibodies may result in reduced motility and possibly abnormal sperm–egg interaction, fertilization, and implantation. The immunobead test is widely accepted as the assay of choice for direct antisperm antibody testing. Serum antisperm antibody is non-specific and is not routinely obtained.

Postcoital test

Examination of postcoital mucus for the presence of motile sperm is indicated in couples suspected of abnormal sperm–cervical mucus interaction. A postcoital test requires precisely timed intercourse and proper collection. An abnormal result may be due to either male or female factors, and further testing using donor mucus and sperm (Penetrak) may be required to determine the causative factor involved.

Other tests

Other studies to assess acrosome reaction, sperm capacitation with computer-assisted semen analysis, and sperm tail ultrastructure with electron microscopy are warranted in selected cases.

Radiographic studies for male infertility

Radiographic studies, including ultrasound, are employed in selected cases of male infertility. Transrectal ultrasound is indicated in men with reduced seminal volume to assess the prostate and seminal vesicles for normal development and evidence of ejaculatory duct obstruction. Direct seminal vesicle puncturing with seminal fluid

analysis and contrast instillation may also be helpful, and may obviate the need for a formal vasogram. Color Doppler ultrasound is very sensitive in the diagnosis of varicoceles; we use this technique in patients whose physical examination is uncertain, but not as part of routine evaluation since surgical repair of subclinical, or radiographically diagnosed only, varicoceles does not appear to be helpful in most subfertile men. Renal ultrasound or an intravenous pyelogram is indicated in men with vasal atresia and negative cystic fibrosis screening because of the associated incidence of renal anomalies. Azoospermic men with a normal testicular biopsy may require vasal fluid analysis and a formal vasogram to assess distal vasal patency. An internal spermatic venogram for varicoceles is likewise reserved for the equivocal cases or for those opting for radiographic embolization as the definitive treatment for their varicoceles.

Testicular biopsy

A testicular biopsy is mostly performed in azoospermic men to differentiate testicular failure, or non-obstructive azoospermia, from normal spermatogenesis and obstruction, or obstructive azoospermia. Non-obstructive azoospermia can be predicted in men with elevated FSH and atrophic testes; diagnostic testicular biopsy in these men is not indicated unless IVF is being considered. Histologic patterns in non-obstructive azoospermia include: (i) germ cell aplasia, or Sertoli cell only, in which the seminiferous tubules are devoid of any germ cell element; (ii) maturation arrest in which the progressive germ cell development is arrested prematurely; and (iii) hypospermatogenesis in which spermatogenesis is complete but the overall germ cell population is reduced. Many of these patients may exhibit a combination of these findings, and patchy foci of full spermatogenesis may be present in testes with predominantly Sertoli-only or maturation arrest. An intraoperative ‘touch prep' biopsy may be performed in men with presumed obstructive azoospermia undergoing surgical exploration; a touch prep is highly accurate in predicting adequate spermatogenesis, and allows for reconstruction to proceed immediately without waiting for the final histology report. Testicular sperm extraction may also be performed in conjunction with IVF by intracytoplasmic sperm injection. Testicular sperm extraction is successful in extracting sperm in up to 50 per cent of men with non-obstructive azoospermia and a trial testicular sperm extraction may be performed prior to committing to IVF by intracytoplasmic sperm injection. Testicular sperm extraction reliably yields sperm in men with obstructive azoospermia.

Genetics of male infertility

A number of genetic conditions may result in male infertility either as part of a well described entity such as Prader–Willi syndrome or as isolated male infertility. Genetic testing is increasingly important since IVF with intracytoplasmic sperm injection is now capable of overcoming the most severe and previously hopeless forms of male infertility associated with these conditions. Genetic testing is indicated in the following cases, especially if IVF with intracytoplasmic sperm injection is being considered.

Vasal atresia

The majority of men with vasal atresia are cystic fibrosis carriers (autosomal recessive) with more than 600 mutation variants described to date. A second group of men with vasal atresia are negative for cystic fibrosis mutations but have renal anomalies due to a common mesonephric developmental defect. If the cystic fibrosis carrier status is confirmed, the wife should also be screened prior to IVF as the incidence of cystic fibrosis mutations in the general population is 1 in 25. If both partners are positive for cystic fibrosis mutations, consideration should be given for either adoption, inseminations with donor sperm, or sperm retrieval with IVF and intracytoplasmic sperm injection in conjunction with preimplantation genetic diagnosis with selective embryo transfers.

Severe idiopathic oligospermia and non-obstructive azoospermia

Structural and numerical chromosome abnormalities such as balanced translocation and Kleinfelter syndrome may be detected via karyotyping. In men with normal karyotype, Y-chromosome microdeletion may be present in 8 to 13 per cent of cases of non-obstructive azoospermia and to a lesser extent in severe oligospermia (< 10 million/ml). The diagnosis of Y-chromosome microdeletion is established by analysis of DNA obtained from peripheral lymphocytes with polymerase chain reaction. Microdeleted sperm is capable of oocyte fertilization and term pregnancy by intracytoplasmic sperm injection, although male offspring will probably exhibit similar microdeletions when tested.

Assisted reproductive technique in male infertility

Intrauterine insemination has long been used and is relatively inexpensive; it requires a minimum of 1 to 3 million motile sperm following processing. More recently, IVF with intracytoplasmic sperm injection has revolutionized the treatment for those with severe male-factor infertility and obstructive lesions not amenable to reconstruction. IVF with intracytoplasmic sperm injection can be accomplished with a very small number of sperm by directly injecting a single sperm into each egg. The source of sperm may be from any portion of the male reproductive tract including the testes. Ovarian hyperstimulation and egg retrievals are carried out under the care of the gynecologist. The overall live birth rate per egg retrieval attempt is about 30 per cent. It is an expensive endeavor with significant patient drop-out rate due to the side-effects from ovarian hyperstimulation. IVF is associated with a 30 per cent likelihood of multiple gestations which often result in premature delivery and costly perinatal care. IVF with intracytoplasmic sperm injection should not replace cause-specific treatments for male infertility such as varicocelectomy or vasectomy reversal, and should be reserved for those in whom more cost-effective treatment options do not exist.

Treatment of male infertility

Hypogonadotropic state

Hypothalamic or pituitary deficiency is manifest by low LH/FSH/testosterone levels and is treatable with appropriate hormonal replacement.

Varicoceles

Varicoceles exert a wide range of impact on seminal parameters ranging from none to absolute azoospermia with testicular atrophy. Their effect may be progressive as evidenced by the higher incidence of varicoceles (70 per cent) in men with secondary infertility. Correction of clinical varicoceles in the subfertile male unequivocally improves pregnancy rates and should be offered as the treatment of choice. Inguinal ligation with optical magnification is the most common approach; other approaches include high ligation via a retroperitoneal approach, laparoscopy, and venographic embolization. Seminal improvement may be evident as early as 3 months, although most patients will require up to 1 year or more to plateau. The likelihood of seminal improvement after varicocelectomy is 70 per cent with the overall pregnancy rate at 30 to 40 per cent.

Congenital vasoepididymal obstruction

Normal testicular biopsy with normal-volume azoospermia and intact vas deferens identify those with congenital obstructions which may be either intratesticular or at the vasoepididymal level. Cystic fibrosis testing is advised in these men. Microsurgical reconstruction with epididymovasostomy may be feasible. In men who have failed or are not amenable to surgical reconstruction, sperm extraction with IVF by intracytoplasmic sperm injection is an alternative.

Vasectomy reversals

Three to six per cent of men with vasectomies request reversals. Vasectomy reversal is highly successful and avoids the high cost and potential complications of IVF with intracytoplasmic sperm injection. Most men will require vasovasostomy, and some may require epididymovasostomy, mostly in those whose vasectomies were performed more than 10 years ago. Overall pregnancy rate ranges from 25 per cent (epididymovasostomy) to 60 per cent or higher (vasovasostomy) depending on the obstructive intervals. Concurrent sperm retrieval with cryopreservation should be considered in men undergoing epididymovasostomy because of the 40 to 50 per cent persistent azoospermic rate; in men with successful epididymovasostomy and vasovasostomy, postoperative cryopreservation of ejaculated sperm is also advised because of the 10 per cent chance of secondary occlusion.

Vasal atresia

In men with bilateral vasal atresia, sperm retrieval with IVF by intracytoplasmic sperm injection is the only option for biologic parenthood. In men with unilateral vasal atresia and azoospermia, contralateral ejaculatory duct obstruction secondary to a seminal vesicle cyst may coexist and may be corrected with transurethral unroofing. Cystic fibrosis testing is mandatory in these men.

Ejaculatory duct obstructions

A number of conditions may lead to ejaculatory duct obstructions. These include: müllerian, seminal vesicle, or ejaculatory duct cysts; seminal vesicle atresia or

hypoplasia; and prostatic scarring with central calcifications. Partial ejaculatory duct obstruction is also possible and is suspected in men with severe oligospermia and/or asthenospermia and transrectal ultrasound findings suggestive of ejaculatory duct obstruction. Transurethral resection is the treatment of choice; some patients may require sperm retrieval with IVF by intracytoplasmic sperm injection.

Ejaculatory dysfunction

Normal ejaculation requires the deposit of seminal fluid in the prostatic urethra as mediated by an intact lumbar sympathetic chain in conjunction with adequate bladder neck closure at the time of ejaculation. Disruption of any of the above may result in anejaculation or retrograde ejaculation. Common causes include diabetic neuropathy, spinal cord injury, retroperitoneal lymph node dissection, psychogenic anejaculation, and previous bladder neck surgery. Anejaculation is managed with penile vibratory stimulation or electroejaculation in conjunction with assisted reproductive technique. In men with retrograde ejaculation, sympathomimetics such as imipramine or pseudoephedrine may be effective in achieving antegrade ejaculation or sperm may be retrieved from postejaculate urine for assisted reproductive technique.

Antisperm antibody

Immunosuppression, sperm washing with intrauterine insemination, or IVF with intracytoplasmic sperm injection may be required.

Idiopathic oligospermia

In many subfertile men, no explanation is found. This is a reflection of our incomplete understanding of human reproductive biology. Treatment outcome with a variety of medications such as clomiphene citrate has been inconsistent and many may require assisted reproductive technique.

Non-obstructive azoospermia

Centrifuged pellet examination should be performed to rule out extreme oligospermia. Genetic testing may influence the overall decision-making process and is highly recommended. Testicular sperm extraction with IVF by intracytoplasmic sperm injection is possible in some cases of non-obstructive azoospermia; IVF by intracytoplasmic sperm injection for non-obstructive azoospermia appears to have a lower live birth rate than IVF performed for obstructive azoospermia. Adoption and donor sperm inseminations are other options.

Further reading

Belker AM *et al.* Results of 1469 microsurgical vasectomy reversals by the Vasovasostomy Study Group. *Journal of Urology* 1991; **145**: 505–11.

Bors E, Comarr AE. Neurologic disturbances of sexual function with special reference to 529 patients with spinal cord injury. *Urology Survey* 1960; **10**: 191.

Condra M, Morales A, Surridge DH. The unreliability of nocturnal penile tumescence recording as an outcome measurement in the treatment of organic impotence. *Journal of Urology* 1986; **135**: 280.

Cookson MS, Nadig PW. Long-term results with vacuum constriction device. *Journal of Urology* 1993; **149**: 290.

Davis-Joseph B, Tiefer L, Melman A. Accuracy of the initial history and physical examination to establish the etiology of erectile dysfunction. *Urology* 1995; **45**: 498.

Diokno AC, Brown MB, Herzog AR. Sexual function in the elderly. *Archives of Internal Medicine* 1990; **150**: 197–200.

Feldman HA, Goldstein I, Hatzichristou DG. Impotence and its medical and psychosocial correlates: Results of the Massachusetts Male Aging Study. *Journal of Urology* 1994; **151**: 54–61.

Graham CW, Regan JB. Blinded clinical trial of testosterone enanthate in impotent men with low or low-normal serum testosterone levels. I>International Journal of Impotence Research 1992; **4**: 144.

Jarow JP. Seminal vesicle aspiration in the management of patients with ejaculatory duct obstruction. *Journal of Urology* 1994; **152**: 899–901.

Jow WW, Steckel J, Schlegel PN. Motile sperm in human testis biopsy specimens. *Journal of Andrology* 1993; **14**: 194–8.

Junemann K-P, Lue TF, Luo JA. The effect of cigarette smoking on penile erection. *Journal of Urology* 1987; **138**: 438–41.

Kim N, Azadzio KM, Goldstein I. A nitric oxide-like factor mediates nonadrenergic, noncholinergic neurogenic relaxation of penile corpus cavernosum smooth muscle. *Journal of Clinical Investigation* 1991; **88**: 112.

Kolettis PN, Thomas AJ. Vasoepididymostomy for vasectomy reversal: a critical assessment in the era of intracytoplasmic sperm injection. *Journal of Urology* 1997; **158**: 467–70.

Lewis RW. Surgery for erectile dysfunction. In: Walsh PC, Retik AP, Stamey TA, Vaughan ED, eds. *Campbell's urology*, 7th edn, pp. 1215–35. WB Saunders, Philadelphia, 1998.

Lewis RW, Barrett DM. *Modern management of male erectile dysfunction*, vol. 14, pp. 162–67. *AUA Update Series*, 1995. American Urological Association, Houston.

Lue TF. Impotence—a patient's goal-directed approach to treatment. *World Journal of Urology* 1990; **8**: 67–74.

Lue TF. Physiology of penile erection and pathophysiology of erectile dysfunction and priapism. In: Walsh PC, Retik AP, Stamey TA, Vaughan ED, eds, *Campbell's urology*, 7th edn, pp. 1157–1179. WB Saunders, Philadelphia, 1998.

Madgar I *et al.* Controlled trial of high spermatic vein ligation for varicocele in infertile men. *Fertility and Sterility* 1995; **63**: 120–4.

Mak V, Jarvi KA. The genetics of male infertility. *Journal of Urology* 1996; **156**: 1245–57.

Marmar JL, DeBenedictis TJ, Praiss DE. The use of vacuum constrictor device to augment a partial erection following an intracavernous injection. *Journal of Urology* 1988; **140**: 975.

Maters WH, Johnson V. *Human sexual response*. Little Brown, Boston, 1970.

Michal V, Kovac J, Belan A. Arterial lesions in impotence: phalloarteriography. *International Angiology* 1984; **3**: 247.

Morales A, Surridge DH, Marshall PG, Fenemore J. Nonhormonal pharmacological treatment of organic impotence. *Journal of Urology* 1982; **128**: 45.

Motorsi F, Strambi LF, Guazzoni G. Effect of yohimbine-trazadone on psychogenic impotence: a randomized, double-blind placebo-controlled study. *Urology* 1994; **44**: 732.

Saenz de Tejada I, Goldstein I, Azadzio K. Impaired neurogenic and endothelium-mediated relaxation of penile smooth muscle from diabetic men with impotence. *New England Journal of Medicine* 1989; **320**: 1025–30.

Saenz de Tejada I, Moroukian P, Tessier J. Trabecular smooth muscle modulates the capacitor function of the penis. Studies on a rabbit model. *American Journal of Physiology* 1991; **260**: H1590–5.

Schlegel PN. Fertility option after vasectomy. *Fertility and Sterility* 1997; **67**: 133–41.

Schlegel PN. Is assisted reproduction the optimal treatment for varicocele-associated male infertility? A cost-effective analysis. *Urology* 1997; **49**: 83–90.

Schlegel PN, Shin D, Goldstein M. Urogenital anomalies in men with congenital absence of the vas deferens. *Journal of Urology* 1996; **155**: 1644–8.

Shabsigh R, Fishman IJ, Scott FB. Evaluation of erectile impotence. *Urology* 1988; **32**: 83–90.

Sigman M, Jarow JP. Endocrine evaluation of infertile men. *Urology* 1997; **50**: 659–64.

Tournaye H, Verheyen G, Ubaldi F. Are there any predictive factors for successful testicular sperm recovery on azoospermic patients? *Human Reproduction* 1997; **12**: 80–6.

39.7 Benign prostate disease

David Cranston and Damian Hanbury

Prostatic hyperplasia
Introduction
Function of the prostate
Pathophysiology
Clinical features
Investigations
Treatment
Conclusion
Prostatitis
Introduction
Clinical features and treatment
Further reading

Prostatic hyperplasia

Introduction

Hyperplasia is the most common benign condition to affect the prostate; its treatment forms a large part of the urosurgical workload in Europe and the United States of America. Despite uncertainty about its causes and natural history, it is evident that the incidence of benign prostatic hyperplasia increases with age. The doubling time of the hyperplastic gland is slow and it is rare to see patients presenting before the age of 55 years. Berry and his colleagues reviewed the results of five autopsy studies and looked at the age-related prevalence of histologically diagnosed, benign prostatic hyperplasia: the prevalence begins rising in the fourth decade and approaches 90 per cent by the ninth. American men now have a 29 per cent chance of undergoing a transurethral prostatectomy during their lifetime. The economic cost to the National Health Service in the United Kingdom from prostatic surgery lies between £58m and £75m each year. There is a much lower incidence of benign prostatic hyperplasia in Asia, and in China and Japan it is an uncommon condition.

Function of the prostate

The function of the prostate remains unclear. It contributes about 10 per cent of the seminal fluid, has a beneficial effect on sperm motility, and produces numerous substances, such as citric acid, acid phosphatase, prostaglandins, and zinc, the roles of most of which are unknown.

Pathophysiology

There is considerable evidence for steroid hormonal control of prostatic growth. Normal prostatic development is only seen in the presence of functioning testes. John Hunter recognized that in mammals with a mating season, there was a regular swelling and shrinking of the prostate along with the testes. Benign prostatic hyperplasia does not occur in castrated males, or pseudohermaphrodites who cannot convert their testosterone to the active metabolite, dihydrotestosterone. The role of oestrogens is controversial but they do stimulate benign prostatic hyperplasia in the experimental dog model; ageing in men is associated with a gradual decline in testosterone and a concomitant increase in oestrogens. Recently, prostatic growth factors have been isolated that stimulate the growth *in vitro* of fibroblasts and epithelial cells. Benign prostatic hyperplasia originates in the periurethral or transitional zone of the gland. The periurethral glands and connective tissue stroma undergo hyperplasia and the smooth muscle hypertrophies, which has the effect of enlarging the gland, causing narrowing of the prostatic urethra and obstruction (Fig. 1).

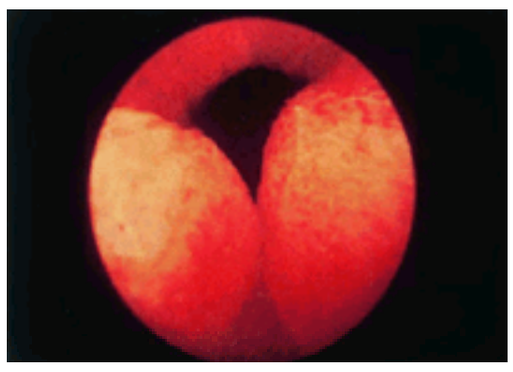


Fig. 1. Cystoscopic view of an occlusive prostate.

Clinical features

The clinical features of benign prostatic hyperplasia are caused by prostatic outflow obstruction, although there is no clear relation between the size of the gland and the symptoms. Increasing outflow resistance commonly causes a mixture of symptoms from the lower urinary tract, including those that were classically known as ‘obstructive': hesitancy, poor stream, a sensation of incomplete voiding due to residual urine in the bladder after micturition, and ultimately acute or chronic urinary retention. These symptoms may be mixed with others such as urgency, urge incontinence, frequency, and nocturia. Patients with appreciable urinary symptoms due to benign prostatic hyperplasia may experience impairment of their health in many other ways, including disturbance of sleep, lack of energy, and interference with employment, sexual function, social life, and holidays.

Symptoms secondary to urinary infection and bladder stones can result from incomplete bladder emptying (Fig. 2). Occasionally the bladder decompensates, leading to chronic urinary retention with overflow. These patients often present with enuresis and little else in the way of obstructive symptoms. The most severe complication from benign prostatic hyperplasia is that of a high-pressure chronic retention with renal failure, in which symptoms may be minimal and a poorly emptying bladder with hypertension are often the only clinical features (Fig. 3 and Fig. 4).



Fig. 2. Late intravenous urogram showing ‘fish-hooking' of right ureter, a small right-sided bladder diverticulum, multiple loosened bladder stones, and a huge basal prostatic impression.



Fig. 3. Late intravenous urogram showing a chronic retention and ureteric dilatation; the patient presented with enuresis and weight loss.



Fig. 4. A patient with chronic urinary retention.

Investigations

A detailed history and full examination, including digital rectal, are the prerequisites to diagnosis. The urine should be examined by dipstick, and by microscopy and culture. Serum creatinine and electrolytes should be checked if there is any suspicion of chronic retention. If polydipsia is suspected, a frequency/volume chart will establish this diagnosis.

Symptoms in the lower urinary tract are notoriously unreliable and unreproducible, so it is helpful to have further objective anatomical and physiological information. The most useful is the urinary flow rate, in which the patient is asked to void into a machine from a ‘comfortably full' bladder; the flow is then converted into a graph tracing (Fig. 5). A maximum flow rate of less than 10 ml/s from a voided volume of more than 200 ml strongly suggests obstruction.

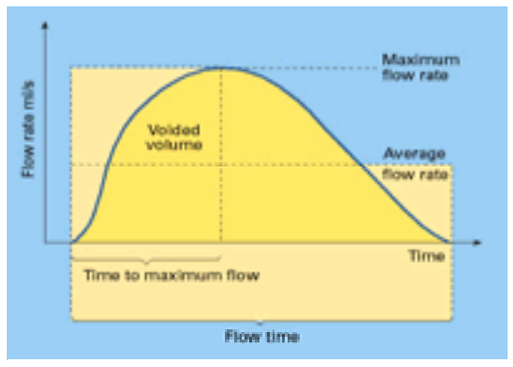


Fig. 5. Characteristics of a free-flow study: flow rate is plotted against flow time and from this the maximum and the average flow rates can be calculated, together with the voided volume; the figure shows a normal profile in a patient without obstruction.

Ultrasonographic imaging of the urinary tract gives information on the size of the prostate, the residual volume after micturition, bladder stones, and any dilatation of the upper urinary tract. Ultrasonography has replaced intravenous urography as the investigation of choice for demonstrating the relevant anatomy in this condition. A good history and examination together with a flow-rate study and ultrasonography should be sufficient to establish the diagnosis in the majority of cases. Formal urodynamic assessment is indicated in younger patients or where irritative symptoms predominate.

Treatment

The aims of treating benign prostatic hyperplasia are to improve the quality of life by removing outflow obstruction and restoring easy micturition. In the absence of serious complications such as renal failure, symptomatic patients should be offered the choice of immediate treatment or watchful waiting; the patient's preferences should be the key factor in the recommendation of any treatment, as all therapies have their side-effects. The options range from watchful waiting, through drug therapy to all types of prostatic ablation; they are discussed in more detail next.

Drug therapies

a-Blockers act by relaxing the smooth muscle of the prostate and bladder neck, which receives innervation via a_1-adrenergic receptors. They do have two major side-effects, retrograde ejaculation and postural hypotension, although the latter is minimized by taking the first dose at night. Younger patients or those with small prostates seem to benefit most from a-blockade.

The evidence that functioning testes may be necessary in the development of benign prostatic hyperplasia has led a number of investigators to assess the role of hormonal treatments. Androgen withdrawal causes regression of benign prostatic hyperplasia, but the side-effects of loss of libido and potency make it an unacceptable treatment. One major study of finasteride, a 5a-reductase inhibitor that blocks the conversion of testosterone to dihydrotestosterone, has shown that treatment over a 4-year period in 3000 men reduced the obstructive symptoms, the prostate volume, and also, to a small degree, the probability of surgery and acute retention. However, it took 4 months before any benefit was noted.

Surgical treatment

Transurethral prostatectomy

There are a variety of surgical options in the treatment of benign prostatic hyperplasia, although transurethral prostatectomy at present remains the ‘gold standard' by which other methods should be evaluated. Transurethral prostatectomy has increased in popularity over the last three decades with the development of the

resectoscope, and this operation has led to a marked reduction in postoperative morbidity and length of hospital stay compared with open prostatectomy. In 1989/90 some 93 per cent of all prostatectomies in National Health Service hospitals in the United Kingdom were performed transurethrally and this figure rose to 98 per cent when prostate surgery performed by specialist urological surgeons was evaluated.

The widespread introduction of transurethral prostatectomy without its rigorous assessment in randomized, controlled trials has led to debate about its effectiveness and the risks relative to those of the previously conventional open procedure. Some studies suggested a higher mortality in patients undergoing transurethral prostatectomy over those undergoing open surgery. However, more recent work has suggested that this difference may be due to difficulties in adjusting adequately for comorbidity.

It has been found that 78 per cent of patients experience improvement of symptoms after a prostatectomy and 74 per cent are happy with the result. It would certainly appear that outcome is related to the severity of the symptoms, with those patients who experienced the most severe symptoms preoperatively having the greatest improvement postoperatively.

The incidence of complications following transurethral prostatectomy varies between different series but larger studies have reported an immediate postoperative morbidity of 18 per cent, with haemorrhage, clot retention, failure to void after catheter removal, and infection of the urinary tract being the common complications. Total incontinence is very rare, although minor degrees of incontinence are much more common and the incidence of urinary-tract infection in the 3 months after surgery is reportedly as high as 25 per cent. The incidence of retrograde ejaculation is variable, but all patients should be warned of this before surgery. Transurethral syndrome is an unusual complication following transurethral resection and is characterized by raised blood pressure, bradycardia, mental confusion, and nausea. It is thought to be due to intoxication with the irrigating solution, usually glycine, and is related to the length of time of the procedure. The operative mortality of transurethral prostatectomy ranges from 0.2 to 1 per cent, the most common cause of postoperative death being cardiovascular complications.

Transurethral prostatectomy is not without its complications and these, together with a trend towards less invasive surgery, have led to increasing pressure on urological surgeons to search for alternative ways of treating benign prostatic hyperplasia, some of which are considered next.

Bladder neck incision (transurethral incision of the prostate)

This technique, initially described by Orandi in 1973, involves an incision in the capsule of the prostate in the seven o'clock and five o'clock position, opening up the prostatic urethra and forming a groove or tunnel through which voiding can take place. It does not involve resection of prostatic tissue and for this reason is said to carry less morbidity and a shorter operating time. Long-term follow-up shows symptomatic relief in 80 per cent of patients, but it is not recommended in patients with a prostate weighing over 35 g.

Laser prostatectomy

The term ‘laser prostatectomy' embraces two distinct techniques, contact and non-contact. The two lasers most commonly used in urology are the neodymium:yttrium aluminium garnet (Nd:YAG) and the holmium. The mechanism of non-contact laser prostatectomy is similar in many respects to microwave hyperthermia and cryosurgery, which rely primarily on thermal coagulation and delayed sloughing of the tissue with coagulative necrosis, and usually requires catheterization for 1 to 2 weeks. This technique enjoyed initial popularity but the long-term results were less encouraging and interest in it is declining.

In the contact technique the laser probe directly touches and immediately vaporizes prostatic tissue, the net result being the immediate removal of obstructing tissue in a manner equivalent to transurethral prostatectomy. Improvement in the urinary stream is usually prompt. The potential advantages of contact laser prostatectomy over conventional transurethral prostatectomy are in peri- and postoperative morbidity. A reduction in blood loss can lead to a reduced requirement for transfusion and the possibility of shorter times for catheterization, which in turn could allow selected patients to undergo prostatectomy in a day-surgery unit, with the potential for a major reduction in treatment costs. The most comprehensive study of contact laser prostatectomy has been the Oxford laser prostate trial, which was set up as a prospective, double-blind, randomized, controlled trial of transurethral prostatectomy and contact laser prostatectomy, with economic evaluation of both procedures. The American Urology Association symptom score was used as the primary outcome measure. Peak urinary flow rate, treatment-related complications, and reoperation and health service costs were secondary outcome measures. The results showed that perioperative blood loss and transfusion requirements were statistically significantly reduced by laser prostatectomy. There was no clinically significant difference at 3 years between transurethral prostatectomy and contact laser prostatectomy in terms of mean change in symptom scores and flow rates. There are distinct perioperative advantages in favour of contact laser treatment, but some disadvantages in terms of recatheterization and reoperation rates.

Hyperthermia

Heat can be delivered selectively to the prostate by microwave probes placed in the rectum or in the prostatic urethra. The depth of penetration of the microwave is dependent on its frequency and the heating of the tissue results from polarization of the small molecules in the oscillating electromagnetic field. Conventionally, hyperthermia consists of raising the body temperature to between 42 and 45°C. More recently, the concept of thermotherapy has been introduced, in which the prostate is heated above 45°C in order to produce more definite tissue changes. Hyperthermia treatments require several sessions lasting 30 to 60 min; thermotherapy requires a single treatment lasting 1 h. These treatments can be performed on outpatients with minimal morbidity, but, although good symptomatic responses have been reported, there is very little improvement in the minimal flow rate. Complications include haematuria, urinary-tract infection, and rectal pain; 20 per cent of patients may need catheterization for up to 7 days.

Prostatic stents

The first stent used for prostatic obstruction was described in 1980 by Fabian and since then a variety of temporary and permanent prostatic stents have been developed. Many have the advantage that they can be introduced under local anaesthetic, although the potential for dislodgement, calcification and infection, and for interference in sexual function, remains.

Cryotherapy

This technique appears to be effective in certain groups of patients but requires a prolonged period of catheterization, making it less attractive than are other options.

Transurethral balloon dilatation of the prostate

Dilatation of the prostatic urethra has been practised for many centuries using metal sounds; in the mid-nineteenth century, metal dilators were designed with the specific purpose of rupturing the prostate gland and the bladder neck. While successful in relieving obstruction they were associated with significant morbidity and their use was abandoned. It sees a revival of interest once or twice a century but as yet without any greater success.

Conclusion

Benign prostatic hyperplasia continues to account for a major part of the workload of urological surgeons. This is likely to increase with the increasing age of the population and will continue to present a challenge for urological surgeons to develop new and effective ways of managing the disease. No doubt new drug therapies will be developed for patients in whom watchful waiting is not an option, the role of laser treatment is now becoming clearer, and other treatment techniques are on the horizon that will need careful evaluation in randomized, controlled trials.

Prostatitis

Introduction

Prostatitis is one of the more common inflammatory conditions encountered in urological practice. Accurate data on its incidence and prevalence are not available, but recent statistics from the United States indicate that there were more physician visits for prostatitis than for benign prostatic hyperplasia or prostate cancer. Prostatitis may be due to acute or chronic bacterial infection; it can also occur in the absence of bacterial growth, and it has been suggested that ‘chronic pelvic pain syndrome' would be a more appropriate term in patients without demonstrable infection. Demonstration of bacteria in the postprostatic massage of urine or in expressed prostatic secretions, when the midstream urine shows no growth, is highly diagnostic of bacterial prostatitis. *Escherichia coli* or *Klebsiella* and *Proteus* spp. often cause bacterial prostatitis; pseudomonads and enterococci are less common. Prostatitis is seen in men of all ages, and typically is associated with pain in the perineum and suprapubic areas associated with frequency and urgency of micturition. Abdominal examination is normally unremarkable but the characteristic feature is a very tender prostate on

rectal examination. Occasionally, a fluctuant abscess may be palpable within the gland.

Clinical features and treatment

Acute bacterial prostatitis often produces generalized malaise and fever associated with symptoms localized to the prostate and acute cystitis. It responds dramatically to antimicrobial drugs. Sometimes it is necessary to treat the patient in hospital and prolonged oral antibiotic therapy of at least 30 days is recommended to prevent the development of chronic bacterial prostatitis.

Chronic bacterial prostatitis is one of the most common causes of relapsing infection of the urinary tract in men and in the past has been difficult to treat, although the newer fluoroquinolone antibiotics have been shown to be safe and effective in this condition.

Non-bacterial prostatitis is more common than bacterial prostatitis. The cause is unknown and the treatment often empirical and of variable effectiveness; *Chlamydia trichomatis* remains a possible aetiological agent, although the evidence is far from definite. A trial of tetracycline or one of its derivatives, together with a-adrenergic-blocking agents, is sometimes of help, especially in patients with perineal pain from the prostate, as the prostate is rich in a-adrenergic receptors and symptoms may be relieved by these drugs.

A significant number of patients with prostatitis have an underlying anxiety about carcinoma, and reassurance, where appropriate, plays an important part in management.

Further reading

Barry MJ, Mulley AG, Fowler FJ, Wennenberg J. Watchful waiting versus immediate transurethral resection for symptomatic prostatism. *Journal of the American Medical Association* 1988; **259**: 3010–17. [An examination of the advantages and disadvantages of non-surgical versus surgical treatment.]

Berry SJ *et al.* The development of human benign prostatic hyperplasia with age. *Journal of Urology* 1984; **132**: 474–9.

Concato J *et al.* Problems of comorbidity in mortality after prostatectomy. *Journal of the American Medical Association* 1992; **267**: 1077–82.

Doll HA, Black NA, Flood AB, McPherson CK. Patient-perceived health status before and up to 12 months after transurethral resection of the prostate for benign prostatic hypertrophy. *British Journal of Urology* 1993; **71**: 297–305. [Classic paper from the patient's viewpoint.]

Doll HA, Black NA, McPherson K, Flood AB, Williams GB, Smith JC. Mortality, morbidity and complications following transurethral resection of the prostate for benign prostatic hypertrophy. *Journal of Urology* 1992; **147**: 1566–73.

Edwards LE *et al.* Transurethral resection of the prostate and bladder neck incision: a review of 700 cases. *British Journal of Urology* 1985; **57**: 168–71.

Keoghane SR *et al.* A double blind randomised controlled trial and economic evaluation of transurethral resection versus contact laser vaporisation for benign prostatic enlargement—three year follow up. *British Journal of Urology* 2000; in press.

Khoury S. Future directions in the management of benign prostatic hyperplasia. *British Journal of Urology* 1992; **70** (Suppl.): 27–32. [A look at potential advances in the treatment of benign prostatic hyperplasia.]

Linder A *et al.* Complications in hyperthermia treatment of benign prostatic hyperplasia. *Journal of Urology* 1990; **144**: 1390–2.

Luzzi G. The prostatitis syndromes *International Journal of STD and AIDS* 1996; **7**: 471–8. [Excellent review papers on this difficult subject.]

Luzzi G. Chronic pelvic pain syndrome. [Editorial]. *British Medical Journal* 1999; **318**: 1227–8.

McConnell *et al.* The effect of finasteride on the risk of acute urinary retention and the need for surgical treatment among men with benign prostatic hyperplasia. *New England Journal of Medicine* 1998; **338**: 557–63. [Excellent long-term study on finasteride.]

Nickel JC. Clinical guidelines for prostatitis. *European Urology* 1998; **34**: 455–6.

Peeling WB. Diagnostic assessment of benign prostatic hyperplasia. *Prostate Supplement* 1989; **2**: 51–68. [Good overview of the assessment of benign prostatoc hyperplasia.]

Reddy PK. Balloon dilatation of the prostate: principles and techniques. *Journal of Endourology* 1991; **5**: 93–8.

Roos NO *et al.* Mortality and re-operation after open and transurethral resection of the prostate for benign prostatic hyperplasia. *New England Journal of Medicine* 1989; **320**: 1120–4. [Classic review paper which caused much debate as to the safety of transurethral resection versus open prostatic surgery.]

Watson GM. Minimally invasive therapies of the prostate. *Minimally Invasive Therapy* 1992; **1**: 231–40.

39.8 Cancer of the prostate

John Naitoh, Jean B. deKernion, and Daniel Shoskes

Introduction
Epidemiology and genetics
Grading systems
Modes of spread
Staging systems
Screening
Diagnosis
Indications for evaluation
PSA blood test
Management of localized disease
Pretreatment evaluation
Treatment options for localized prostate cancer
Treatment of metastatic disease
Hormonal therapy
Treatment of the complications of metastatic prostate cancer
Recommended treatment by stage
Clinical stage T1a and T1c
Clinically organ-confined (stage T2)
Locally advanced: stage T3
Hormone-sensitive, metastatic disease (N^{+}, M^{+})
Hormone-refractory disease
Further reading

Introduction

Prostate cancer is the most common cancer found in American men, and is the second leading cause of cancer death. In 1997, approximately 209 900 new cases were diagnosed in the United States alone, with over 41 800 deaths attributed to this malignancy. However, despite these large numbers, there continues to be controversy about the use of screening tests that can detect cancer in asymptomatic individuals, and about the optimal treatment if a localized (organ-confined) prostate cancer is diagnosed.

Here we will first review the ‘non-controversial' aspects of pathology, epidemiology, and the clinical presentation of prostate cancer. Then we will discuss the controversies that surround its diagnosis and treatment, including the risks and benefits of ‘watchful waiting', radiotherapy, and surgery.

Epidemiology and genetics

Prostate cancer is a common finding at autopsy, but it is a cause of death in only a relatively small proportion of men. Although 60 to 80 per cent of 80-year-old men have histological evidence of prostate cancer, only 3 per cent of men will actually die from it. The present problem is the clinical differentiation between an incidental, non-lethal and a clinically significant, potentially lethal tumor.

Despite the fact that the majority of men will develop only incidental tumors, prostate cancer is yet a major cause of disease and death throughout the world. In the United States it ranks second to lung cancer as a cause of cancer deaths in men, with the lifetime risk for developing a clinically apparent prostate cancer ranging between 5 and 8 per cent. While the median age for clinical diagnosis is currently 72 years, this average will probably decrease over time because increasing numbers of younger men are having their cancer detected in prostate-specific antigen (**PSA**)-based screening programmes.

Family history is one of the most important risk factors for the development of prostate cancer. First-degree relatives of those with this cancer have an eightfold increased risk, especially if the cancer becomes clinically manifest at a younger age (below 60 years). Recently, a highly penetrant but rare familial prostate cancer gene was localized to the chromosome region 1q24–25. While this gene is believed to be involved in only about one-third of all familial prostate cancers, over 80 per cent of men who have this susceptibility gene will develop a clinically significant prostate tumor as compared to the 5 per cent incidence found in non-carriers of this gene.

Epidemiological studies show marked geographical and racial differences in the incidence of prostate cancer. North Americans (especially African Americans) and Western Europeans have the highest rates of clinically significant prostate cancer, while Asians living in the East have the lowest. Interestingly, while the mortality rates show significant geographic variance, the autopsy prevalence (i.e., the incidence of clinically insignificant prostate cancer) is fairly constant worldwide.

As immigration from a low-incidence to a high-incidence country results in the eventual elevation of the individual's cancer risk to that of the adopted country, it has been hypothesized that environmental and dietary factors are more important than ethnic factors for the development of a clinically significant prostate cancer. Low dietary vitamin D, low selenium, and high dietary fat have all been correlated with the development of prostate cancer, but dietary fat intake has proved of particular interest. Because dietary fats can be metabolized into steroid hormones, it has been proposed that high-fat diets result in altered metabolism of testosterone, which may then be able to promote the growth and development of prostate cancer. As a corollary, it has also been shown that low-fat diets or diets enriched in w-3 fatty acids or soy proteins can decrease testosterone production and create an environment that is adverse for the development of prostate cancer.

Grading systems

There are several existing systems for grading adenocarcinomas by their risk for progression, but the most popular is the Gleason, which a3signs t8e tumor`cells to a category ranging from 1 to 5 based on their degree of differentiation, gland formation, and their relation to the stroma. Given that multiple histological patterns are usually found within a given tumor, the Gleason score that is the sum of its two most common patterns is commonly used. Gleason sum scores can therefore range from 2 (most well differentiated) to 10 (most poorly differentiated).

In general the Gleason score can be correlated with the stage of the tumor and the overall prognosis. However, while the Gleason system is simple and reproducible, it has some limitations. First of all, determining the Gleason score of a prostate tumor from a few needle biopsies has proved difficult and unreliable. Up to 40 per cent of core biopsies will underestimate the actual grade of the cancer found once the entire prostate has been removed and examined. Secondly, while the Gleason score provides very useful prognostic information if it is less than 3 or greater than 7, the information obtained from intermediate Gleason scores is less useful.

Thus, owing to the limited prognostic information obtained from the Gleason system on the biological potential of a given tumor, many other systems and tumor markers have been devised with the hope of gaining more accurate information. Besides the creation of other grading systems that consider cellular morphology as well as glandular structure, it has also been proposed that flow cytometry (tumor DNA ploidy), p53, p27, and other molecular markers may all be able to provide a more accurate prognosis than the Gleason score and clinical staging.

Modes of spread

In its early stages, prostate cancer is confined to the gland by the thin surrounding capsule. In its more aggressive form, it can escape the confines of the gland, either at the apex (where there is no gland capsule) or by lateral penetration through the capsule. The most common point of direct spread is where the erectile nerves perforate the prostate capsule along the posterolateral aspect of the gland, although the cancer can also invade the capsule anteriorly, or it can spread into the seminal vesicles or the bladder neck.

The primary lymphatic drainage of the prostate is via the internal iliac, perivesical, external iliac, obturator, and presacral nodes. The secondary lymphatic drainage includes the inguinal, common iliac and para-aortic nodes. Despite this well-described pattern of lymphatic drainage, up to 15 per cent of patients who have lymph-node metastases will have ‘skip lesions' that bypassed the primary drainage sites and went directly to the common iliac or para-aortic chains.

Hematogenous dissemination can also occur, based on the finding that the majority of patients who have advanced prostate cancer will also have detectable prostate cells circulating in the peripheral blood and bone marrow. A rich venous plexus surrounds the prostate and connects to the venous drainage of the spine: this collection

of veins (Batson's plexus) is potentially one of the reasons why spinal metastases are common in advanced prostate cancer.

In addition, prostate cancer cells themselves have a direct tropism to bone, with bone metastases occurring in 85 per cent of patients who die of prostate cancer. However, unlike other neoplasms, prostate cancer induces the formation of blastic not lytic bone lesions. Thus, whenever an older man presents with bone pain and the plain radiograph shows an osteoblastic lesion, the primary differential diagnosis is metastatic prostate cancer.

Staging systems

By convention, the International Union Against Cancer TNM system is currently the standard staging system used in the medical literature. Updated in 1997, the TNM system is largely based on the findings of the digital rectal examination (Table 1).

Table 1 The 1997 American Joint Committee on Cancer clinical staging ystem for prostate cancer (TNM)

Screening

For a screening programme to be effective, the cancer must be detectable with minimal discomfort, low risk, and acceptable financial cost. Early diagnosis must also be able to improve the quantity and quality of the patient's life; there must be an effective treatment available for the cancer; the treatment must be acceptable to the patient; and active treatment of the asymptomatic patient must be able to provide a better outcome than observation alone.

There is at present no evidence that screening for prostate cancer based on the PSA blood test fulfils all of the requirements of an effective screening programme. If the detection of prostate cancer is based on the presence of voiding symptoms or on the findings by digital rectal examination, then only a minority of patients will be found to have curable, localized disease at the time of diagnosis. On the other hand, if screening with PSA is able to detect a significant number of tumors that will not actually shorten the patient's life, then the morbidity and mortality of that screening and subsequent treatment could be worse than the disease. Unfortunately, it is as yet not known whether screening programmes are resulting in the over- or undertreatment of prostate cancer.

Despite these uncertainties, there is some evidence that screening for prostate cancer is beneficial. First, there is evidence that PSA-based screening programmes are able to detect only clinically significant tumors (i.e., larger and more aggressive tumors that will cause significant symptoms or decrease the patient's lifespan if left untreated). Secondly, given the lack of effective treatments for advanced prostate cancer, early detection and treatment is currently the only approach that can decrease its death rate. And finally, there is some evidence that the aggressive diagnosis and treatment of prostate cancer is having an effect: decreases in its incidence and mortality rate were seen for the first time in 1997. While it is too early to attribute these changes to the implementation of the screening programmes that began in the mid-1980s, PSA-based screening programmes have demonstrably resulted in a decrease in the number of patients who present with metastatic tumors or a markedly elevated PSA.

Given the lack of concrete evidence to support or refute the benefits of a screening programme for prostate cancer, the decision to offer such screening must be made on an individual basis, depending on the patient's age, health status, family history, personal risk for prostate cancer, and his personal beliefs. However, it must also be remembered that there are some who should undergo intensive evaluation because they have an increased risk of dying from prostate cancer. These individuals at risk include all African American men, and, as explained earlier, all men who have a first-degree relative with prostate cancer. In these men there may be a genetic predisposition to aggressive prostate cancer, and the American Urologic Association recommends their evaluation with annual PSA tests and rectal examinations beginning at the age of 40 years. For all other men, once they understand the risks and benefits of screening, as outlined above, testing for prostate cancer should begin at the age of 50 if they desire it. Following an initial screening, annual examinations can then be considered, with any changes on rectal examination or a PSA rise of greater than 0.7 ng/ml per year serving as an indication for biopsy. Once the patient is over the age of 70, or once he develops significant underlying illnesses or another incurable malignancy that will decrease his life expectancy to less than 10 years, then screening for prostate cancer can be halted.

Diagnosis

For many years, prostate cancer could only be detected from an abnormal finding on digital rectal examination, an elevated prostatic acid phosphatase, and the results of transperineal biopsy. Over the last 10 years, new methods have proliferated (transrectal ultrasound-guided prostate biopsies, serum PSA) that have made the diagnosis of prostate cancer easier (Fig. 1). This has resulted in a large increase in the numbers of patients diagnosed.

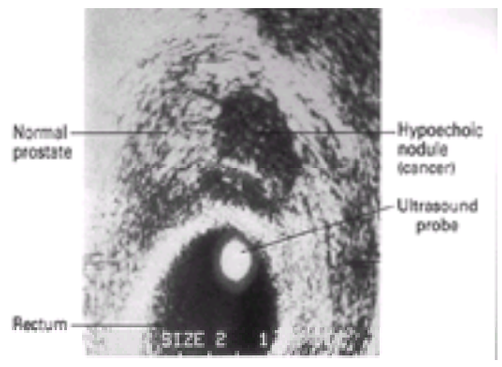


Fig. 1. Transrectal ultrasound image of prostate. Note hypoechoic nodule (arrow) found to be adenocarcinoma on biopsy.

Indications for evaluation

For any male with voiding difficulties, or where there are signs and symptoms of metastatic cancer, a search for prostate cancer is reasonable. Such patients may have symptoms due to prostate cancer. In these circumstances, either curative treatments (such as radical prostatectomy or radiation therapy) or palliative treatments (such as hormone therapy) can be used to decrease the symptoms and improve quality of life.

Patients who have an abnormal finding on digital rectal examination should also be tested for prostate cancer. In general, any area of asymmetry, nodularity, or induration in the prostate should be evaluated (by either a finger- or ultrasound-guided needle biopsy), since up to 50 per cent of these changes will be caused by cancer. However, while digital rectal examination is cheap, rapid, and well tolerated, the problem with using it alone as a screening tool is its lack of specificity and its inability to detect cancers before they have spread. Over 50 per cent of patients whose cancer was detected by digital rectal examination had already developed locally advanced or metastatic disease at the time of diagnosis.

PSA blood test

The discovery of PSA has revolutionized the diagnosis and treatment of prostate cancer. The traditional method of direct palpation of the prostate was only able to detect cancers after they had become large and therefore had a greater chance of being advanced. In contrast, PSA allows the detection of tumors before they become palpable, and therefore can detect cancer at an earlier stage.

PSA is a 34-kDa protein produced by the glandular tissue of the prostate. While it is largely prostate-specific, small amounts can also be found in the periurethral and salivary glands, although these traces are usually serologically undetectable and are therefore clinically insignificant.

The detection of PSA in serum is related to the extent of its leakage from the prostate. Overall, the leakage is relatively lower from benign than cancerous tissue: serum PSA is increased by only 0.3 ng/ml per gram of hyperplastic tissue, but by approximately 3.5 ng/ml per gram of tumor. Thus, while there is a potential overlap in serum PSA between men who have cancer and those who simply have very large prostates, cancer overall causes a greater rise in PSA per gram of tissue than benign prostatic hypertrophy, with the generally accepted normal range for PSA being between 0.0 to 4.0 ng/ml.

When this point is used as a cut-off value, the sensitivity of PSA in detecting cancer is between 73 and 87 per cent. However, the performance characteristics of PSA are highly dependent on the population tested and the relative risk of its members having a malignancy. For those who have a prostate nodule, approximately 50 per cent will be found to have prostate cancer independent of the value for serum PSA, and all of these men should be biopsied however low their serum PSA (interestingly, in this group of men, 20 per cent who are diagnosed with a prostate cancer after an abnormal finding on digital rectal examination will have a ‘normal' PSA of less than 4.0 ng/ml). In contrast, for men who have a normal prostate on digital rectal examination, the incidence of cancer is between 4 to 7 per cent for those who have a PSA of less than 4.0 ng/ml, and between 31 to 42 per cent for those who have a PSA of greater than 10 ng/ml.

Despite the enthusiasm for PSA as a screening tool, there are definite limitations to its use. One concern about the widespread use of PSA-based screening is the detection of indolent, clinically insignificant tumors. As was discussed above, histological evidence of prostate cancer is almost an inevitable consequence of aging, with cancers that progress to cause illness and death affecting only 8 per cent of all men. Theoretically, since PSA-based screening programmes can detect small tumors before they are palpable (clinical stage T1c), it is possible that some of these will be so small and slow growing that they will never represent a threat to the patient's health or life. However, recent data, as mentioned above, suggest that PSA-based screening programmes are picking up only clinically significant, potentially lethal tumors.

It must also be remembered that PSA is just prostate specific, and not cancer specific, with the serum PSA only reflecting the amount leaking into the bloodstream from the prostate. Thus, any other medical condition that increases the volume of benign prostate tissue (such as benign prostatic hypertrophy), or any condition that can increase the leakage of PSA into the bloodstream (such as perineal trauma, urinary-tract infection, prostatitis, urethral catheterization, prostate biopsy, or recent ejaculation) can cause the PSA to rise into the abnormal range. Due to these confounding factors, the specificity of PSA is limited, with only 25 per cent of men who have a PSA of more than 4.0 ng/ml actually having prostate cancer.

Beacuse of the low specificity of the PSA test, whereby four patients will have to undergo evaluation with prostate ultrasonography and biopsy for every cancer that is detected, many perturbations of PSA have been devised to improve its performance. PSA velocity, free PSA, and age/racial adjustments have all been proposed as ways of improving the specificity of PSA as a screening test.

PSA velocity uses the rate of change in PSA over time to predict the presence or absence of prostate cancer. In the normal prostate, hyperplasia and hypertrophy occur with aging, resulting in a slow but steady increase in serum PSA. Given that cancer cells proliferate more quickly than benign prostate cells, the serum PSA will rise more rapidly in the man who has cancer than in one who has a benign condition such as hypertrophy. While a change in PSA of greater than 0.75 ng/ml per year can indicate the presence of cancer (sensitivity 72 per cent, specificity 90 per cent), the utility of PSA velocity has been limited by the 15 per cent error of the PSA test itself, and by the other benign conditions (such as infection, inflammation) that can spuriously increase the PSA.

The recent discovery of multiple forms of PSA in the bloodstream has resulted in the development of new screening tests for prostate cancer. PSA is normally found conjugated to a_1-antichymotrypsin in the bloodstream, with cancer tissue making more a_1-antichymotryp-sin than normal tissue. Thus it was expected that the free (unconju-gated) PSA would be lower in those with prostate cancer than in healthy individuals.

Early studies found that free PSA ratios were lower in men with prostate cancer than in those with benign prostatic hypertrophy, with later reports showing that the average free:total PSA ratio was 0.18 in those with prostate cancer and 0.28 in benign prostatic hypertrophy. Subsequent studies showed that free:total PSA ratios appear to have the greatest utility when the PSA is between 3 to 10 ng/ml, where the diagnosis of cancer is less certain. In this group, the specificity of the free:total PSA test was better than that of the PSA alone, while the sensitivity of the two tests was nearly identical. By using the free:total PSA test, up to 40 per cent of the negative biopsies could have been eliminated if a cut-off value of 0.20 had been applied prospectively to the study population.

Later studies disputed the idea that there is a single free:total PSA value that can accurately discriminate between cancer and benign prostatic hypertrophy. Cut-off values adjusted to different prostate sizes or different PSA values have been proposed as ways to improve further the utility of the free:total PSA test. By using these adjustments, the 90 per cent sensitivity for cancer detection was preserved, while eliminating 33 per cent of the unnecessary biopsies.

Thus, free:total PSA has the potential to improve the ability of PSA to discriminate between cancer and benign prostatic hypertrophy. This appears to be especially true in men whose PSA is between 4 and 10, since the risk of cancer is more definitive at the extreme ranges of PSA. Interestingly, free:total PSA might be particularly useful in younger men whose absolute PSA is low but whose PSA is abnormal based on age-adjusted ranges. Up to 15 per cent of these men will be diagnosed with cancer during serial evaluation. The application of a free:total PSA ratio cut-off of 0.23 identified 93 per cent of the men in this category who had prostate cancer and reduced the number of negative biopsies by 30 per cent. The free:total PSA might also be useful in the serial evaluation of a man who has an elevated PSA but a history of multiple negative prostate biopsies; in our practice, the free:total ratio is used to decide if he can safely be observed, or if more extensive biopsies are needed.

Out of all of the different adjustments made to improve the performance of PSA as a screening test, the best appears to involve the use of age-adjusted ranges. With the recognition that (i) the prostate will increase in size with age, and (ii) PSA increases with age, it was logical to conclude that the normal range of PSA will also increase with age. From population-based studies of men who have no evidence of cancer, the mean and 95 per cent confidence intervals for PSA have been established, and these values have been indexed to patient age (see Table 2).

Age	Caucasian	African American	Asian
<50	0-2.5	0-2	0-2
50–60	0-3.5	0-4	0-3
60–70	0-4.5	0-4.5	0-4
>70	0-6.5	0-5.5	0-5

Table 2 PSA reference ranges normalized for patient age and ethnicity

By using these reference ranges, it was argued that these normal values would increase the detection of cancer in younger men who will have a greater likelihood of benefiting from screening and treatment, while decreasing the numbers of older men who undergo prostate biopsy where the benefits of cancer detection are less certain. It was predicted that these reference ranges would result in a 15 per cent increase of cancer detected in men under 60 years of age, while it would decrease the number of biopsies performed in men over 70 years old by 44 per cent.

Subsequent to the development of age-adjusted ranges, the next logical step was an adjustment for ethnicity given the different rates of prostate cancer seen in

Caucasians, Asians and African Americans. Asians appear to have low rates of prostate cancer, and have correspondingly low PSA. In contrast, African Americans have high incidences of high-grade, aggressive prostate cancer, and therefore also need a lower limit for PSA in order to maximize the detection of their tumors while they are still curable (see Table 2).

In summary, the main limitation of PSA as a tumor marker is the fact that it is prostate specific not prostate cancer specific. The optimal definition of ‘normal' for PSA is still undefined, with age adjustments, ethnic adjustments, or determination of the amount of PSA bound to plasma proteins all ways that potentially could improve its specificity. Unfortunately, none of these methods is ideal, with improved specificity often at the cost of lowered sensitivity. Thus the search for new tumor markers has continued. Prostate-specific membrane antigen, human kallekrein, prostate stem-cell antigen, and DD3 are some of the molecular candidates that may replace PSA as the next-generation marker for the early detection of prostate cancer, with the hope that they will not just be prostate specific but also cancer specific.

Management of localized disease

Pretreatment evaluation

The standard pretreatment evaluation of the patient who has localized prostate cancer involves tumor staging, an evaluation of comorbidity, and as a baseline assessment of urinary and sexual function. Unfortunately, at this time there is a paucity of published information on how these factors can be accurately and easily measured in clinical practice.

For standard presurgical staging, most practitioners use a combination of the digital rectal examination, PSA, bone scan (in selected patients), and CT scan (in selected patients). Digital rectal examination has been the mainstay of clinical staging because it is cheap, rapid, and relatively non-invasive. However, it is affected by significant interobserver variability, by limited diagnostic accuracy, and by poor sensitivity. While the presence of T3 or T4 disease on digital rectal examination is very specific, there is also a tendency for it to understage the cancer that has spread microscopically (see Table 3). With these difficulties, digital rectal examination is relatively unreliable for determining the presence or absence of organ-confined disease.

Clinical stage based on DRE	Percentage understaged (pathological T3 or T4)
Stage T1A (found on TURP)	0
T1C (normal examination)	40
T2A (nodular induration < ½ of one lobe)	38
T2A (nodular induration > ½ of one lobe)	64–76

DRE: digital rectal examination; TURP: transurethral resection of prostate.
Constructed from data reported in: Catalona WJ. Patient selection for, results of, and impact on tumor resection of primary staging radical prostatectomy. Urologic Clinics of North America 1990; 17: 819–34.

Table 3 The risk that clinical staging underestimates the actual volume of prostate tumors

With improved understanding of the biology of PSA, this serum marker has also been used to stage patients before surgery. In studies on surgically treated patients, those who had a preoperative PSA of less than 4 ng/ml were almost uniformly cured by radical prostatectomy. In addition, patients whose PSA is less than 10 ng/ml demonstrably have an extremely low risk of detectable skeletal metastasis, and therefore do not need a staging bone scan before treatment. Finally, in those who have small, low-grade prostate tumors with a low PSA, the staging pelvic lymphadenectomy can also be deferred, owing to their low incidence of metastasis.

In contrast, once the PSA is greater than 10 to 20 ng/ml, then the chance that organ-confined cancer will be found at operation is markedly decreased, and the risk that cancer will recur after treatment is markedly increased. Similar data are recorded in the radiotherapy literature, where a pretreatment PSA of greater than 20 to 30 ng/ml is regarded as a probable reflection of occult metastatic disease.

However, the utility of PSA in staging is limited in the majority of patients who have clinically localized prostate cancer. PSA has not performed well when used alone as a staging tool in some case series, owing to the fact that most patients who undergo surgery have a PSA that is only mildly elevated (between 4 and 10 ng/ml), and given that there are cases where the PSA is elevated by conditions other than prostate cancer. Thus, the absolute PSA may not reflect the patient's actual tumor burden or stage.

The radionuclide bone scan is also a standard part of staging for the patient with localized prostate cancer. Given its propensity to spread to the axial skeleton, bone scans can identify patients who have metastatic disease and who therefore will not benefit from surgical treatment. Unfortunately, bone scans lack both sensitivity and specificity, with other medical diseases such as healing fractures, degenerative joint disease or Paget's disease yielding false-positive results: MRI or plain radiographs are often needed to confirm the diagnosis. Furthermore, given the low incidence of skeletal metastasis in patients who have low-grade, low-stage tumors, the bone scan can usually be deferred if the PSA is less than 10 ng/ml. In contrast, the risk of skeletal metastasis is markedly increased once the pretreatment PSA is greater than 20 to 30 ng/ml: in these patients, a bone scan is required to rule out the presence of skeletal metastasis.

The use of other imaging studies for staging prostate cancer is more controversial. Historically, CT scanning had only limited utility for cancer staging because of its low level of resolution. However, recent technical advances have improved its ability to detect smaller structures, and it has been suggested that modern (spiral or thin-cut) CT scanning might have the highest sensitivity of all current imaging methods for detecting the presence of nodal metastasis. The advantages of CT scanning are that it can identify small but pathologically enlarged lymph nodes (more than 1 cm across), and provide a route for biopsy of these suspicious lesions without the need for surgery. However, the main limitations of CT scanning are its lack of specificity (since inflammatory, non-cancerous nodes are also be detected), its inability to provide information about the local extent of the primary tumor, and its inability to detect micrometastatic disease.

Others have proposed that MRI can provide additional information on staging for clinically localized prostate cancer, especially where clinical staging and PSA provide ambiguous information. Limited studies have shown that MRI (especially endorectal-coil MRI) can provide useful information on capsular penetration and invasion of the seminal vesicles where the PSA and digital rectal examination are indeterminate. Some studies have also shown that MRI findings can independently predict the occurrence of PSA relapse after radical prostatectomy. MRI cannot, however, be considered a standard method for staging as these results have not been replicated at many other centres, and the equipment required for endorectal-coil MRI is available in only a few places.

In summary, the standard staging for the patient with prostate cancer involves the pretreatment PSA and digital rectal examination. In those who have small prostate nodules and a PSA of less than 10 ng/ml, no further evaluation is needed. In contrast, those who have a PSA greater than 10 ng/ml should have a bone scan, while CT scanning or MRI can also be used in selected circumstances, although the data supporting their use are more limited.

Treatment options for localized prostate cancer

While it is clear that many men will benefit from aggressive treatment with surgery or radiotherapy, it is also apparent that some who have a clinically localized prostate tumor will be overtreated, owing to the indolent nature of their tumors, while others will have a recurrence after therapy despite a thorough staging evaluation. However, as there is no biomarker that can reliably separate those who have indolent cancers from those who have biologically aggressive lesions, and as there is no reliable method for detecting the presence of micrometastatic disease, all men who have a localized prostate cancer and a reasonably long life expectancy should consider undergoing treatment rather than ‘watchful waiting'.

Treatment for the man who has a clinically localized prostate cancer should be tailored to his individual characteristics, and to the grade and stage of his tumor. To date no randomized trial has compared equivalent populations of patients and definitively shown a survival benefit for either radical prostatectomy or radiation therapy. Interpretation of treatment outcomes is further complicated by the understanding that many will survive equally long without any therapeutic intervention other than endocrine therapy and symptomatic treatment of metastases. The variable behavior of prostate cancer and the confounding presence of other illnesses in the age group affected make objective interpretation of the available data impossible. Until a test is developed that can accurately define the potential lethality of a specific tumor in a specific patient who has other specific illnesses, other criteria must be used by the individual in his choice of treatment, more often based on the risks of side-effects than the chance of cure.

Side-effects of therapy must be considered in the decision-making. Radical prostatectomy can result in wound infection, hemorrhage, and deep-vein thrombosis, and has a perioperative mortality of up to 1 per cent. In the long term, erectile dysfunction, incontinence, and urethral stricture can also compromise quality of life after radical prostatectomy. For the man whose sexual potency is important or whose risk for complications is great, non-surgical therapies or watchful waiting may be preferred. Newer surgical techniques can minimize periurethral dissection and preserve the erectile nerves, but the impact of radical prostatectomy on sexual function can still be significant. In contrast, the effects of radiotherapy can be much more mild, while watchful waiting has no immediate effects on sexual function or continence. However, radiotherapy and watchful waiting result in a trade-off, where the decreased risk of immediate side-effects from treatment are exchanged for the increased risk that local symptoms will occur later if the cancer progresses.

Watchful waiting

In some cases, the risk of cancer progression in the patient's lifetime is so low that no intervention is warranted. Most physicians would agree that this applies to men over 80 years old with localized disease and to men who are over the age of 70 with stage T1a disease. Some European centres also advocate observation for younger men who have well- to moderately differentiated stage T2a and stage T2b disease.

Studies on watchful waiting show that T1–2 disease will have a 10-year local progression of 72 per cent, metastasis rates of 23 per cent, and cause-specific eath rates of 8 per cent. (but it must be remembered that most of these men received hormonal therapy at the time of tumor progression). While these rates of progression appear to be similar to those seen with other forms of therapy, these men cannot be directly compared to those who undergo radiotherapy or radical prostatectomy. In general, those assigned to watchful waiting are older, have a lower volume of disease, and lower-grade tumors (some tumors were diagnosed by cytology alone). Furthermore, neither PSA nor prostatic acid phosphatase were measured over time in these studies, meaning that the actual progression rates of the tumors were probably much higher. Thus, watchful waiting should be applied with caution to men with life expectancies of more than 10 years, with a full disclosure of the risks and benefits of such an approach.

External-beam radiotherapy

Even more difficult than the decision between treatment and watchful waiting is the choice between surgical treatment and radiotherapy. In most European centres, external-beam radiotherapy is the primary treatment for localized prostate cancer in all men who are less than 80 years old. In North America, it is usually the first choice for treatment only in men who have moderate comorbidity, or who are between 70 and 80 years old. Standard therapy is 5000 cGy over 5 weeks to the pelvis with a further boost to the prostate of 2000 cGy over 2 weeks, although the use of brachytherapy in the prostate is also becoming more commonplace.

Radiotherapy can be an effective treatment for many men who have organ-confined (T1/T2) and locally advanced (T3) prostate cancers. Local control for stage T1 and T2 tumors can be achieved in 80 to 95 per cent of patients overall, with 5-, 10-, and 15-year survival rates of 80, 60, and 37 per cent, respectively. While the primary concern in using this treatment for men with a life expectancy above 15 years is the high rate of PSA elevation after treatment, there is no conclusive evidence that surgery offers a significant overall survival advantage over radiotherapy.

Radiotherapy does have certain advantages over radical prostatectomy. There is no need for admission to hospital, and the risks of an operation and prolonged postoperative recuperation are avoided. The risks of disabling bladder dysfunction and incontinence are lower with radiotherapy, although impotence can still be a problem over time. However, the side-effects of treatment, the protracted time course of radiotherapy (daily treatments over several weeks), and questions about its efficacy relative to surgical treatment are factors that make it a far from ideal option for all patients.

The complications of radiotherapy have been well defined, although variations in dosage, route (external beam versus brachytherapy), and dosage fractionation will affect the rates of complications (Table 4). Proctitis has occurred in up to 30 per cent of patients while undergoing treatment, with 1 to 2 per cent having long-term rectal problems including stenosis or chronic proctitis. Cystitis and gross hematuria are further side-effects that can occur during radiotherapy to the pelvis. Fortunately, most of these irritative voiding symptoms will resolve after treatment, with only 2 per cent of men having severe, long-term problems following external-beam radiotherapy. Erectile dysfunction can also be associated with radiotherapy, with approximately 50 to 66 per cent of potent men becoming impotent after treatment. However, unlike the erectile dysfunction associated with surgery, the impotence that follows radiotherapy tends to be progressive, and can occur many months to years after the therapy.

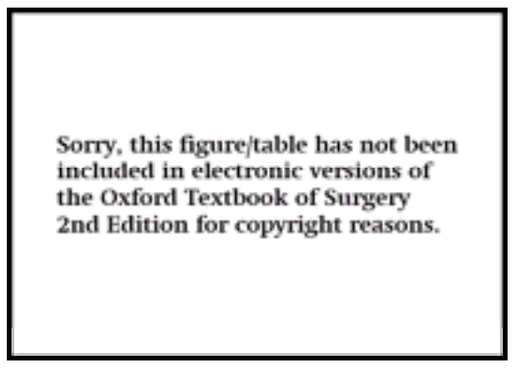


Table 4 No caption available.

Surgery

In the United States, for men who are under the age of 70 years, who have no significant medical comorbidity, and who have a clinically localized prostate cancer, radical prostatectomy is the primary treatment. Due to (i) the aging of the population, (ii) the advent of PSA-based screening, (iii) the development of transrectal ultrasonography, which makes prostate biopsies easier to perform, and (iv) the major improvements in surgical technique, the use of radical prostatectomy has increased six-fold between 1984 to 1990. However, despite improvements in the ability to detect cancer at earlier stages, and in surgical technique and patient selection, biochemical (PSA-based) recurrence rates after surgery are still between 20 and 30 per cent, with the side-effects (impotence, incontinence) potentially worse than the symptoms that could be caused by the cancer itself. The main issue is the preoperative selection of patients who will best benefit from the surgery, both in terms of cancer control and overall quality of life.

Radical prostatectomy is the complete surgical removal of the prostate and seminal vesicles. As originally performed, either by a retropubic or a perineal approach, there was a high incidence of complications, particularly impotence and incontinence. In the early 1980s, better understanding of the neuroanatomy of the pelvis allowed for the preservation of the periprostatic neurovascular bundles responsible for erection (Fig. 2 and Fig. 3). Improved attention to anatomical detail and further modifications in surgical technique have resulted in further improvements in intraoperative blood loss, and in postoperative continence and sexual function.

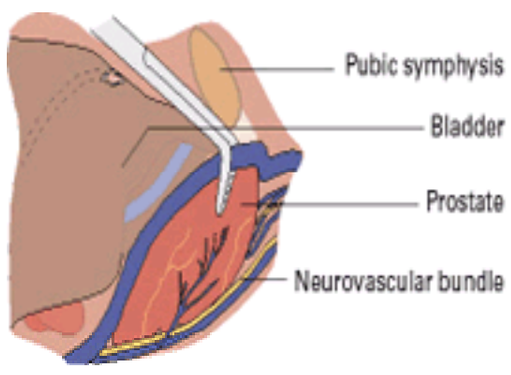


Fig. 2. Surgical anatomy of the prostate gland with demonstration of the neurovascular bundles.

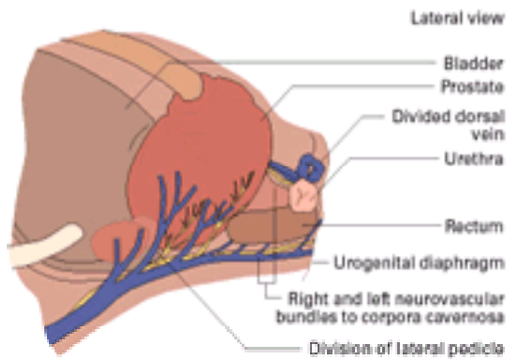


Fig. 3. Radical prostatectomy, division of lateral pedicles. Note preservation of neurovascular bundles.

Most series report 5-year disease-free survival rates of between 60 to 80 per cent following radical prostatectomy, with the probability of cancer control related to tumor stage. Between 80 and 90 per cent of patients who have organ-confined cancer, 75 per cent who have T3a/T3b tumors, and 10 to 20 per cent who have pathological T3c lesions, will remain free of recurrence 5 years after surgery.

The risks of surgery must also be considered as they represent a much greater challenge to the patient than with radiotherapy. As outlined earlier, the short-term complications of radical prostatectomy include stricture, rectal injury, wound infection, and deep-vein thrombosis. Early studies reported varying incidences of these complications but the more recent series clearly show that the incidence of significant, short-term complications is low, owing to improvements in surgical technique (Table 5).

Complication (out of 1342 cases)	*No. (%)*
Deep-vein thrombosis/embolism	34 (2.6)
Myocardial infarction/arrhythmia	19 (1.4)
Wound infection/hematoma	17 (1.3)
Lymph leak/prolonged drainage	8 (0.6)
Foley catheter malfunction	5 (0.4)
Rectal injury	3 (0.3)
Death within 30 days of surgery	3 (0.2)

Table 5 Early complications of radical prostatectomy

Over the long term, postoperative impotence can be a significant problem, and it can occur even if both of the erectile nerves were successfully preserved during surgery. Besides the ability to save the nerves, the preservation of potency is mainly dependent on the age of the patient. Postoperative potency rates of up to 80 per cent have been seen in patients who were 50 to 60 years old, while the rate drops to 50 to 60 per cent in men who were 60 to 70 years old. In contrast, men who were over the age of 70 at the time of surgery were rarely potent following radical prostatectomy.

Another potential long-term complication is urinary incontinence. Most modern surgical series report that approximately 20 per cent of patients who undergo radical prostatectomy will develop stress urinary incontinence, with less than 1 per cent having severe leakage.

Comparison of radiotherapy with radical prostatectomy

The side-effects and long-term complications of radiotherapy and radical prostatectomy have been well characterized, but it has been impossible to compare directly the efficacy of these treatments, as the populations who receive them are usually not comparable. Historically, patients who underwent radiotherapy tended to be older and had more locally advanced tumors. Surgery and radiotherapy have also been difficult to compare because studies on the outcome of radiotherapy have always been based on clinical stage, while those on surgery have been based on pathological stage. However, the tendency of the surgeon to underestimate the clinical disease stage is somewhat counterbalanced by a similar tendency for the radiotherapist to overestimate the clinical extent of the tumor.

Clinical trials to compare the efficacy of radiotherapy and surgery have been started, but as yet they have not provided definitive answers to this controversy. In comparable groups of men who underwent either radical prostatectomy or radiotherapy, biochemical recurrence rates (as defined by a detectable PSA in the radical prostatectomy group or a rising PSA after radiotherapy) appear to be similar for the first 5 years after treatment. While the evidence so far suggests that the short-term results of radiotherapy and surgery are equivalent, the longer-term results are not yet available.

In contrast, some studies with longer-term follow-up suggest that cancer control might not be as good with radiotherapy as it is with surgery. The only completed randomized trial to date that compared the two treatments (from the United States Veterans Administration Co-operative Research Group) favoured radical prostatectomy, although it had methodological flaws that limited acceptance of its results. A subsequent study also showed a long-term, disease-free survival advantage for surgical patients over those who received radiotherapy, noting that 22 per cent of the former and 39 per cent of the latter had PSA progression. A structured review of the relevant literature also suggested that long-term PSA recurrence rates are higher in patients treated with radiotherapy than in those treated with radical prostatectomy. In patients who had had more than 15 years of follow-up, the biochemical (PSA-based) relapse-free rates ranged between 19 and 46 per cent in the radiotherapy group, and between 40 to 75 per cent in the radical prostatectomy group.

Thus, while it is unknown whether or not radiotherapy can control prostate cancer as effectively as radical prostatectomy, the long-term data seem to slightly favour surgery. It is hoped that studies from the Prostate Outcome Research Team (PORT) will clarify this controversy in the near future, as they compare the outcomes of different treatment approaches while trying to control for tumor stage, tumor grade, patient age, and comorbidity. Until these studies have been completed, patients must be given all three options for therapy, with the decision largely based on their preference.

Summary

In general, the approach we use is as follows. For men who have an expected lifespan of less than 5 years (i.e., those more than 80 years old), watchful waiting is recommended. For those who have a 5- to 10-year expected lifespan, external-beam radiotherapy is recommended, as it can give similar results to surgery with fewer complications. For those who are under 70 years old with an expected lifespan of more than 15 years, surgery is likely to give superior results, although the overall survival benefit will probably be small. In those who have clinical stage T3 disease and a very high risk of metastatic cancer, radiotherapy in combination with hormone therapy might be the best option (see below).

Treatment of metastatic disease

Hormonal therapy

The favourable response of prostate cancer to androgen ablation was first described by Huggins in 1941, and it has remained the treatment of choice for metastatic disease. Over 90 per cent of men with metastatic prostate cancer will have a measurable response to endocrine therapy. While androgen-deprivation therapy is not

curative, the mean duration of response is approximately 3 years, with some having a much longer response.

The testicles produce about 95 per cent of the circulating androgens under the central control of the pituitary gland and hypothalamus (Fig. 4). The adrenals supply the remaining 5 per cent, although it is uncertain whether adrenal androgens alone can support the growth of prostate cancer cells.

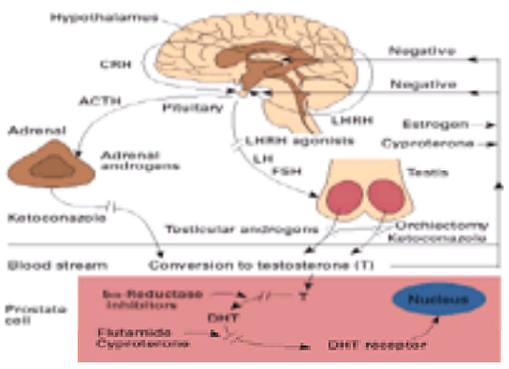


Fig. 4. Endocrine control of prostate growth via androgen production, and pathways blocked by hormone treatments to control prostate cancer. LH, luteinizing hormone; LHRH, luteinizing hormone-releasing hormone; FSH, follicular-stimulating hormone; DHT, dihydrotestosterone; ACTH, adrenocorticotrophic hormone; CRH, corticotropin-releasing hormone.

There are many forms of endocrine manipulation that can be used to treat metastatic prostate cancer. Clinically, androgen ablation can be achieved by interfering with any step of androgen production or action, from blocking the action of the hypothalamic–pituitary axis to stimulate testosterone production to blocking the interaction between testosterone and its receptor (Fig. 4). While there used to be controversy about the optimal use of endocrine therapy, recent studies have provided information that can help both patients and clinicians. The choice of early rather than delayed therapy, the use of total androgen ablation or ablation of testicular androgens alone, and the choice of the therapeutic agent have all been the focus of a research effort, with answers now becoming apparent.

The traditional choices for the patient who needs to undergo androgen-deprivation therapy include bilateral orchiectomy, estrogens, luteinizing hormone-releasing hormone (**LHRH**) agonists, and steroid receptor-blocking agents. Given that each approach has advantages and limitations, the individual's choice has to be based on his ability to comply with the regimen, its simplicity, the safety and side-effects of the treatment, his ability to accept the treatment, its efficacy, and its cost.

Orchiectomy

Since the original report by Huggins, orchiectomy has been the ‘gold standard' for androgen-deprivation therapy. With orchiectomy, one can achieve rapid and effective reductions in serum testosterone, leaving only the small contribution made by the adrenal glands. Castrate levels of testosterone can be found within 3 h after surgery.

The main advantages of bilateral orchiectomy are that it is safe, effective, inexpensive, and that it can rapidly decrease the serum testosterone. For the most part, bilateral orchiectomy can be done under local anesthesia, with an overall cost for the entire procedure of approximately $3000 (in the United States). As it is a one-time treatment, no further costs are incurred by the patient. Orchiectomy also has advantages where compliance is a problem, or where access to a health-care provider is limited.

The main side-effects of orchiectomy are related to the loss of testosterone, including gynecomastia, impotence, and hot flushes. Other problems include surgery-related complications (hematoma, wound infection), loss of self-image, and irreversibility. However, despite these considerations, cost-based analyses have clearly shown the superiority of orchiectomy for the majority of men who have metastatic prostate cancer.

Based on British data, assuming an equivalent efficacy between orchiectomy and LHRH agonists, for patients who have a response that is less than 2 years' duration, the LHRH agonist was more cost-effective, but for those who have more than 2 years' response, bilateral orchiectomy was more cost-effective. While this study could be faulted because of the high cost and 5-day hospital stay for orchiectomy, and the relatively cheapness of the chosen LHRH agonist, it is clear that orchiectomy was the most cost-effective approach for the majority of those with metastatic prostate cancer.

Estrogens

Exogenous estrogen is believed to block androgen metabolism and the growth of prostate cancer at several levels: feedback inhibition of the hypothalamic–pituitary axis, direct suppression of testicular androgens, increased concentrations of sex steroid-binding globulins and prolactin, and direct suppression of tumor growth. To date, the most widely used agent has been diethylstilbestrol at doses of 3 to 5 mg/day. While 1 mg/day is ineffective, the majority of men will have a castrate concentration of testosterone at 3 mg/day, with some requiring 5 mg/day. The main advantages of diethylstilbestrol are its low price (approximately $100 per year in the United States), its effectiveness in achieving the castrate state and as an anticancer agent, and its simple dosing schedule.

Unfortunately, the incidence of life-threatening side-effects increases with increased dose. In addition to the standard side-effects of androgen-deprivation therapy, diethylstilbestrol has also been associated with an increased risk of gynecomastia, thromboembolic phenomena, cardiovascular complications, and sudden death. In fact, once the costs of breast irradiation and those incurred for the treatment of diethylstilbestrol-associated cardiovascular complications are included, the cost advantage for diethylstilbestrol is lost. Furthermore, at a dose of 5 mg, the survival benefit that is due to the hormone treatment of the cancer is lost because of the increased cardiovascular morbidity that is secondary to it. Thus, given the side-effect profile and the existence of less toxic and equally effective hormonal agents, diethylstilbestrol has become a treatment of historical import only (this is especially true because it is no longer available in most countries, owing to the lack of manufacturers willing to produce this inexpensive but potentially toxic agent).

LHRH analogues

LHRH analogues result in a flare of testosterone after about 3 days, with it then paradoxically dropping to castrate concentrations within 2 weeks through downregulation of pituitary receptors and depletion of luteinizing-hormone stores. The clinical effect is equivalent to orchiectomy or diethylstilbestrol, although somewhat delayed.

LHRH agonists (such as leuprolide or goserelin acetate) have become the mainstay of androgen-deprivation treatment in the United States. These produce a lower incidence of gynecomastia and fluid retention but a higher incidence of hot flushes than the estrogens. In addition, unless the early ‘flare' period is covered by diethylstilbestrol or an antiandrogen, there is a danger that the tumor will be stimulated in the first 2 weeks of treatment, leading to an exacerbation of symptoms including urinary retention, spinal-cord compression, and pain from skeletal metastasis.

LHRH analogues are available as daily subcutaneous injection, as a nasal spray, as a monthly depot injection, or as an implant that lasts 3 months. The main advantages of LHRH agonists are their safety, their reversibility (if intermittent androgen-deprivation therapy is to be considered), and their effectiveness in achieving castrate concentrations of testosterone.

However, LHRH agonists are also associated with significant inconvenience and cost: regular visits to the doctor are required for the injections, compliance with the regimen can be problematical where access to the health-care provider is limited, and finally, the cost of the medication itself can be prohibitive. In US dollars, the LHRH agonist will cost each patient approximately $5000 per year, or $12 500 for the entire course of therapy (assuming a median survival time of 30 months for the patient with metastatic prostate cancer). Based on the entire cost of an orchiectomy, the break-even point for cost in the United States between orchiectomy and LHRH agonist is at 6 months, after which LHRH agonists become more expensive.

Total androgen blockade

The goal of total androgen blockade is to block the effect of both testicular and adrenal androgens. Typical regimens have combined either orchiectomy or LHRH

analogues with a pure antiandrogen (such as flutamide or bicalutamide).

The early results were promising. The American National Cancer Institute study that compared an LHRH analog with or without flutamide showed a statistically significant increase in time to progression and a reduction in mortality rate if total androgen blockade was used, although the overall improvement in survival was only 3 months. Similar results were found in a Canadian trial that showed an overall improved survival of 5 months. Subset analysis in both of these studies suggested that patients with a low burden of disease and good performance status benefited most from total androgen blockade.

A cost analysis also suggested that the use of total androgen blockade is reasonable. Given that antiandrogens cost approximately $10 000 per year (in the United States), it was estimated that total androgen blockade costs between $13 000 to $48 000 per life-year saved, depending on how much of a survival benefit was expected from it. Since dialysis costs approximately $40 000 per life-year saved, the argument was made that the cost of total androgen blockade is similar to that of other, well-accepted medical interventions.

However, despite these initially promising results, more recent studies suggest that the utility of total androgen blockade is much more limited. In a large, randomized, prospective study of orchiectomy compared with orchiectomy plus antiandrogen, no differences in overall survival were seen. Subset analysis failed to identify any subgroup that benefited from total androgen blockade. These results, in combination with the high cost of antiandrogen therapy, have basically put to rest the concept of total androgen deprivation.

Other hormonal agents

Ketoconazole

Ketoconazole is an imidazole derivative that inhibits cholesterol metabolism, resulting in ablation of both testicular and adrenal androgen production. It is the most rapidly acting medical treatment for androgen ablation, with castrate concentrations of testosterone achieved within 8 h. Primary side-effects are nausea, gynecomastia, impotence, pruritus, liver dysfunction, and adrenal insufficiency; therefore, ketoconazole is often given together with hydrocortisone.

This medication is extremely useful in the patient who needs emergency therapy (such as in impending spinal-cord compression), given its rapid onset of action. A second condition in which ketoconazole appears to be useful is for the patient who is failing in first-line hormone therapy: there are now anecdotal reports of durable PSA responses in men who have prostate cancer that is progressing despite previous orchiectomy or treatment with an LHRH agonist.

Glucocorticoids

Prednisone at a dose of 15 mg/day can decrease PSA in castrate men, with a reported 20 per cent objective improvement in symptoms. Whether this effect is due to effects on testosterone or on inflammation around the tumor is unclear.

Treatment of the complications of metastatic prostate cancer

Spinal-cord compression

This devastating complication must always be considered in patients with prostate cancer who have back pain or decreased sensation in lower limbs. The finding of cord compression should be considered an emergency: patients who have pain or paraesthesia can often recover function if they are treated promptly, while those who present with paralysis rarely recover function after treatment. Given that patients with metastases can live for many months, preservation of function is of critical importance for the preservation of qualityof life.

Spinal-cord compression should be suspected in any patient who is having progressively worsening back pain, sciatica, weakness in the lower extremities, or urinary/fecal incontinence. Once suspected, the diagnosis can be confirmed by MRI: because of the high incidence of diffuse spinal involvement, it is important to image the entire spine and not just the area that is symptomatic. Treatment usually involves emergency neurosurgical consultation, high-dose steroids, and either external-beam radiation or decompressive laminectomy.

Urinary obstruction

Local extension of prostate cancer is sometimes found at the time of diagnosis, or may occur after external-beam radiotherapy. Obstruction of bladder outflow can present with hesitancy, dribbling, straining, slowing of the urinary stream, or urinary retention. The ureters can also become obstructed either due to the local extent of the primary tumor along the trigone or to extrinsic compression from involved retroperitoneal lymph nodes.

Management of urinary-tract obstruction will depend on its degree, the severity of symptoms, and the prior treatment that the patient has received. For the man who has mild voiding symptoms from prostate cancer, or who has unilateral ureteral obstruction but otherwise normal renal function, either observation or endocrine therapy can be the primary approaches.

In contrast, the man who has significant obstructive voiding symptoms, or recurrent gross hematuria, should undergo a limited transurethral resection of the prostate to alleviate his discomfort. Likewise, the man who has a reasonable life expectancy and bilateral ureteral obstruction should also be treated, using either antegrade or retrograde ureteral stenting to stabilize renal function. The stents can usually be removed in 2 to 3 months, once endocrine therapy has adequately reduced the size of the tumor.

In contrast, for the man who has developed ureteric obstruction while on androgen-deprivation therapy, the prognosis is grim and caution should be used before any intervention. Comparative studies have shown that procedures to unobstruct the kidneys in the presence of a hormone-refractory prostate cancer do not improve quality of life or increase overall life expectancy. Thus those who have ureteral obstruction and hormone-refractory prostate cancer need to be counselled for their poor prognosis, with consideration given to hospice care without any further invasive surgery. Likewise, for the man who has progressive, hormone-refractory prostate cancer and urinary retention, prostatic stents, urethral catheterization or a suprapubic tube can be used to drain the bladder until he succumbs from his disease.

Bone pain

For most patients who have isolated, symptomatic bone metastasis, external-beam radiotherapy can be used to palliate symptoms. For those who have metastatic disease in the spine, radiotherapy can also be used to stabilize the axial skeleton and prevent the later occurrence of spinal-cord compression.

For those who have more diffuse symptomatic skeletal disease, systemic radioactive agents such as strontium-89 can be used, resulting in a significant decrease in pain scores and the need for narcotics. Unfortunately, this therapy is associated with bone-marrow suppression and has only minimal long-term efficacy.

Aggressive pain management should also be instituted, as the patient who has advanced prostate cancer can live for quite some time. An approach that is similar to the World Health Organization Pyramid should be used. Acetaminophen (paracetamol), non-steroidal anti-inflammatory agents, and oral narcotics are usually the initial steps for outpatient pain management, and should provide adequate relief for the majority of patients. For those who have more severe symptoms, consultation from a cancer pain-clinic team should be obtained for therapy that is based on neuroleptics, narcotic patches, epidural implants, and nerve blocks.

Recommended treatment by stage

The following are our own recommendations and represent an approach that, in 1998, would be considered somewhat conservative in North America and aggressive in Europe. Each treatment choice is recommended to the patient as our own preference: but all treatment options with their risks, benefits, and the strength of the data supporting them are given as alternatives. Some of the information given here repeats and reinforces the account given above.

Clinical stage T1a and T1c

Treatment decisions for the man who has T1a prostate cancer must be based on confirmation that he has a small, low-grade tumor. Our current policy is to perform repeat transrectal ultrasound biopsies in those who have stage T1a tumors, unless they have been poorly resected, in which case a repeat transurethral resection plus a needle biopsy is appropriate. If the resampling is negative, then patients can be followed serially with digital rectal examination and PSA testing at 6-month intervals, with any significant change in either signalling the need for further biopsies. In contrast, if, on re-evaluation, the patient has a more significant tumor, then treatment

decisions need to be based on his age and general medical condition.

Similarly, the management of those who have clinical T1c tumors must be influenced by their age and the tumor grade. As almost all stage T1c tumors are clinically significant (either high-grade or of significant volume), and as many are in fact pathologically much more advanced than anticipated from the digital rectal examination (with involvement of periprostatic fat in 44 per cent and extension into the seminal vesicles in 8 per cent), we generally recommend definitive treatment for almost all younger men who have a T1c tumor, even if they only have a single focus of well-differentiated tumor found on a single biopsy core. In contrast, for those who are older, or where there is a question about their 10-year life expectancy, we tend to be more conservative provided that a second set of extensive biopsies, including deep transition-zone biopsies, shows no significant tumor volume and no tumor with a grade greater than Gleason 3.

Clinically organ-confined (stage T2)

With the diagnosis of a stage T2 prostate cancer, the treatment decision can be difficult. The standard choices include radiotherapy, radical prostatectomy, and watchful waiting, but, as discussed earlier, the final choice cannot be based on any firm scientific evidence that one form of treatment is better than another. Therefore, decisions have to be based more on the side-effects, long-term risks, and financial and emotional costs of the different therapies. In general, young healthy patients have historically been steered towards adical prostatectomy in the United States, while older patients are steered towards observation or radiotherapy. This is due to the different risks and benefits that have been associated with radical prostatectomy, radiotherapy, and observation.

For men who are in good medical condition and are under the age of 70, radical prostatectomy is commonly recommended, although the evidence supporting this approach over radiotherapy is extremely limited. With improved patient selection, anesthetic technique, and hospital care, the surgical complications are minimal, with the average hospital stay following radical prostatectomy decreased to 3 days. In most cases, the patient is able to eat a regular diet within 2 days of surgery, with full activity resumed within a month.

As there is no apparent difference in cure rates between surgery and radiotherapy, the choice for surgery has to be based on other criteria. The main differences between radiotherapy and surgery for treating prostate cancer are the more accurate staging information and the higher risks of long-term, treatment-related complications from surgery.

With radiotherapy, tumor stage is established on the PSA and digital rectal examination, both of which are poor predictors of actual tumor volume. Furthermore, up to 40 per cent of patients are found to have been undergraded when the results from the needle biopsy are compared to the Gleason score of the final specimen. Thus, the actual tumor grade and stage remain unknown for the patient who undergoes radiotherapy, with that information only inferred from the response to treatment. In contrast, following radical prostatectomy, more accurate prognostic information can be given, and high-risk patients who would benefit from adjuvant therapy can be identified.

However, this improved prognostic information comes at the cost of a higher risk for the patient. While radiotherapy is not associated with the involuntary loss of urine, most current series report that approximately 20 per cent patients who undergo radical prostatectomy will have some degree of stress urinary incontinence after surgery, with approximately 1 per cent having severe leakage. Postoperative impotence can also be a more common problem after radical prostatectomy. While the technique of peeling the erectile nerves off the lateral edge of the prostate has markedly improved the outlook for adequate sexual function, the return of normal erections also depends on the preoperative sexual function and the patient's age. In general, younger men have a better chance of preserved potency if they are treated by radical prostatectomy with nerve sparing, while older men have a better chance of preserved sexual function after radiotherapy, as nerve sparing rarely works once the patient is over the age of 70.

Radiotherapy has other advantages over surgery. It is well tolerated, not associated with hemorrhagic or anesthetic risks, and does not require admission to hospital or a significant recovery period. Normal activity can usually be maintained during radiotherapy, although daily visits to the radiotherapist are required for 4 to 6 weeks, and many patients do report feeling fatigued at the end of the treatment programme. However, radiotherapy also has disadvantages compared to surgery. Aside from some uncertainty about the success of the treatment, radiotherapy can have significant side-effects. While incontinence is not associated with radiotherapy, severe bladder irritation (urgency, pain and frequency) can occur in 5 per cent of patients. Furthermore, rectal irritation (diarrhea, rectal urgency, tenesmus, and rectal bleeding) can occur in 3 to 10 per cent of patients, with impotence occurring in 40 to 50 per cent. And finally, over the long term (greater than 10 years), radiotherapy may not provide as good a chance of cure as surgery. After radical prostatectomy, the majority of recurrences are within the first 4 years, while for radiotherapy, recurrences can arise over a much longer period of time since the prostate may not have been completely ablated. Thus, while cure rates for radiotherapy and surgery appear to be equivalent at 5 years, these results may not hold up over longer follow-up.

In summary, there is no single, best treatment for organ-confined prostate cancer. Until definitive data are collected that prove in a randomized, prospective manner that one approach is superior to the other, the treatment choice has to individualized to the patient's goals for health maintenance, and to his age and the presence of other illness. Until there are believable data on efficacy, watchful waiting, radical prostatectomy, and radiotherapy (including brachytherapy) can all be considered acceptable treatment options, with the differences in side-effects determining when they are to be used.

Locally advanced: stage T3

There is no consensus on the optimal treatment for men who have locally advanced prostate cancer. Decisions about the management of the patient with T3 prostate cancer must be based on the known prognostic factors that predict his overall life expectancy and determine the risk of tumor progression. As up to 50 per cent of patients who have a clinical T3 lesion will have lymphatic metastases at the time of surgery, as therapy of any kind is rarely curative, and as surgery overall is not provenly superior to radiotherapy in the management of these tumors, the treatment for the man who has locally advanced prostate cancer can involve androgen-deprivation therapy alone, surgery, radiotherapy, or a combination of these.

For T3 disease, surgery or radiation alone are rarely curative. Recurrence rates after treatment can be as high as 80 per cent, with approximately 50 per cent of patients having positive nodes, and 2 per cent invasion of the tumor into adjacent organs. Therefore, therapies that combine systemic hormonal manipulation with local treatment of the primary tumor have been devised.

Neoadjuvant hormone therapy has not improved the chance of long-term cure after surgery, but might have some benefit for those treated with radiotherapy. Radiation therapy alone was associated with an overall, clinical disease-free, and PSA disease-free, survival of 71, 43, and 26 per cent, respectively, while the comparable proportions for the group that received neoadjuvant hormones were 84, 61, and 46 per cent. Whether or not this beneficial response will translate into improvements in overall survival is unknown. Alternative strategies for T3 prostate cancer include cryotherapy (freezing the prostate with liquid nitrogen), a combination of external-beam radiation with brachytherapy (therefore allowing the administration of higher radiation doses to the prostate), or combinations of standard therapies with gene treatments that try to enhance a tumor's sensitivity to the immune system.

Hormone-sensitive, metastatic disease (N^+, M^+)

Prostate cancer classically spreads to the bones and lymph nodes: in such patients, androgen-deprivation therapy is the mainstay. However, with the definition of an entire population of men who have metastatic cancer based on the serum PSA and yet have no symptomatic or radiological evidence of malignancy, the timing of, and indications for, androgen-deprivation therapy have become controversial.

Preliminary data had suggested a survival benefit for immediate rather than delayed androgen-deprivation therapy (where treatment is offered only when symptoms appear), but flaws in study design limited the strength of this conclusion. For example, a recent study from Great Britain that showed a survival benefit for early as compared to deferred androgen-deprivation therapy basically compared patients who had had immediate treatment with a group that had received no treatment at all.

There are other randomized, prospective studies that have failed to show a survival benefit for early androgen-deprivation therapy. Given the lack of conclusive data, patients who have asymptomatic, metastatic disease should be offered the choice between early or delayed endocrine therapy. While there are potential (and unproven) survival benefits for those who undergo early hormonal therapy, the patient must also understand that the treatment is associated with significant side-effects, including osteoporosis, loss of libido, impotence, hot flushes, malaise, and gynecomastia. In contrast to the asymptomatic, symptomatic patients should receive immediate endocrine therapy since they can then see a rapid improvement in their symptoms. In emergencies (such as impending spinal-cord compression), bilateral orchiectomy or ketoconazole therapy should be instituted, as these can result in the most rapid decline in serum testosterone.

For other patients, the choices include total androgen deprivation, bilateral orchiectomy alone, or a depot LHRH analog alone. Bilateral orchiectomy is the simplest and least expensive of the options, and has the advantage that monthly visits to the physician for injections are not required (especially important where access to the provider is limited, or where the patient has a history of non-compliance). On the other hand, acceptance of orchiectomy has been problematical in some populations for cultural reasons, with the use of LHRH agonists representing the primary mode of androgen-deprivation therapy in the United States.

LHRH agonists and orchiectomy are equivalent in their ability to reduce the serum testosterone. The main differences between the two are related to (i) reversibility, and (ii) cost. Intermittent hormone ablation has been used to treat advanced prostate cancer in some patients. In this approach, LHRH agonists are used cyclically, with times where the patient is off of the medication. This approach can be advantageous to the patient who wants to minimize side-effects, as sexual function and energy can return to normal during those times. For those treated with LHRH agonists, intermittent therapy is an option. For those treated by orchiectomy, the loss of testicular androgens is permanent. However, it must also be remembered that the long-term efficacy of intermittent androgen deprivation has yet to be proved, although animal studies suggest that this approach will not decrease the amount of time that a patient will respond to androgen-deprivation therapy.

The other consideration is financial. Given the need for cost containment in modern medicine, and given the identical efficacy of orchiectomy and LHRH agonists, in most circumstances surgical castration should be the preferred method for androgen deprivation because it is cheaper.

As 95 per cent of testosterone production comes from the testicles and 5 per cent from the adrenal glands, there was concern that the relatively small amount of adrenal androgens could promote growth of the cancer. Because of this fear, the concept of total androgen blockade was developed, where orchiectomy or an LHRH analog are used to ablate testicular androgen production, while an androgen-receptor blocker (flutamide or bicalutamide) is used to eliminate the effect of the adrenal androgens. While preliminary evidence showed improved survival for patients treated by total androgen blockade, later prospective, randomized trials failed to show any survival benefit. In addition, the cost of antiandrogens can be prohibitive, running up to US$15 000 per year.

Hormone-refractory disease

There has been a recent reconsideration of the term ‘hormone refractory,' given that many patients who fail a first-line hormonal manipulation will often respond to a second or even a third. With this understanding came a modification in nomenclature, as (i) it has been argued (see above) that the low concentrations of testosterone coming from the adrenal glands may also be contributing to the growth of the cancer, and (ii) that the antiandrogen medications in common use (flutamide, bicalutamide) may actually develop agonist properties that can stimulate cancer growth as the tumor progresses.

For patients who have androgen-resistant but hormone-sensitive tumors, this implies that the ablation of testicular androgen production alone is no longer effective. In such men, the addition of an antiandrogen to block the testosterone receptor (e.g. flutamide or bicalutamide), or the discontinuation of the antiandrogen (if the patient is already on one) may be able to induce a transient response. While there are few data on the initiation of an antiandrogen at the time of tumor progression, the antiandrogen withdrawal phenomenon has been well described, in which stopping the antiandrogen results in tumor regression that can last for 3 to 6 months in approximately 33 per cent of patients.

The other option for the hormone-sensitive but androgen-insensitive patient involves the use of steroids in combination with ketoconazole (ketoconazole is a potent inhibitor of adrenal steroid production). Therapy combining ketoconazole and hydrocortisone can produce significant decreases in serum PSA and cancer symptoms in a small subset of patients. While one of the mechanisms of ketoconazole involves the suppression of adrenal androgen production, a direct cytotoxic effect is probably important as well.

Unfortunately, most of these secondary interventions for metastatic prostate cancer are only effective for 2 to 3 months, after which the tumor begins to grow again. At this point, the patient should be considered to have hormone-refractory disease (i.e., both androgen- and hormone-resistant). For the atient who is young and has a good performance status, enrolment in a chemotherapy trial should be considered if one is easily available. For other patients, hospice care should be considered.

Until recently, prostate cancer has been considered to be a chemoresistant tumor. However, with the recent development of new chemotherapy agents and with a better understanding of the molecular biology of prostate cancer, there are new pharmacological manoeuvres that can induce short-term responses in patients who have hormone-refractory prostate cancer. PSA-based response rates ranging from 50 to 60 per cent have been seen for chemotherapeutic agents such as paclitaxel, docetaxel, and vinorelbine. While these PSA responses have not yet resulted in any improvement in overall survival, these treatments should still be considered valuable, as improved functional status and subjective improvement in cancer pain can be seen after chemotherapy.

Where the patient is too debilitated to undergo chemotherapy, palliation of symptoms alone should be the goal of treatment. As most patients with hormone-refractory prostate cancer can live for many months, the preservation of function is critical for the maintenance of their quality of life. In cases of urinary retention, a limited transurethral resection is reasonable; otherwise consideration can be given to inserting a chronic indwelling catheter. For patients with ureteral obstruction secondary to lymphadenopathy, nephrostomy tubes or internal stents can be placed for drainage, although most of these men will have a very short life expectancy and will not benefit from aggressive attempts to preserve renal function.

For the patient with prostate cancer metastatic to the skeleton, adequate pain control must be ensured through the use of anti-inflammatory medications and narcotics (see earlier). Systemic strontium or spot radiotherapy treatments can also be effective in some cases to improve pain control. In addition, vigilance must be maintained for the symptoms of spinal-cord compression (such as sudden changes in gait, or a sudden change in bowel or bladder habits), as many men who have advanced prostate cancer have a significant lifespan to come despite their hormone-refractory state: the detection and active treatment of unstable spinal segments will allow them to remain ambulatory and functional for as long as possible.

Further reading

Albertsen PC. Screening for prostate cancer is neither appropriate nor cost effective. *Urologic Clinics of North America* 1996; **23**: 521–30. [This review provides a concise summary of the arguments for and against prostate cancer screening.]

Arcangeli CG, Ornstein DK, Keetch DW, Andriole GL. Prostate-specific antigen as a screening test for prostate cancer. *Urologic Clinics of North America* 1997; **24**: 299–306. [This review nicely summarizes all of the ways that PSA can be manipulated to improve its specificity as a screening test, including PSA density, velocity, and age-adjusted ranges.]

Catalona WJ *et al.* Evaluation of the percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. *Journal of the American Medical Association* 1995; **274**: 1214–20. [The first large series that showed a potential benefit for the use of free:total PSA.]

deKernion JB, Belldegrun A, Naitoh J. Radical prostatectomy for localized prostate cancer: indications, techniques and results. In Belldegrun A, Kirby RS, Oliver T, eds. *New perspectives in prostate cancer*, pp.185–204. Oxford, Isis Medical Media,1998. [This chapter provides a concise review of the surgical treatment of prostate cancer, including the risks and benefits of surgery as compared to radiotherapy and watchful waiting.]

Epstein JI, Partin AW, Sauvageot J, Walsh PC. Prediction of progression following radical prostatectomy: a multivariate analysis of 721 men with long term follow-up. *American Journal of Surgical Pathology* 1996; **20**: 286–92. [An update of Partin's earlier work that looked at the relative importance of different clinical factors in predicting tumour recurrence and pathological stage.]

Dowd GJ, Veltri RW, Orozco M, Miller CM, Oesterling JE. Update on the appropriateness of staging evaluation for newly diagnosed prostate cancer. *Journal of Urology* 1997; **158**: 687–98. [The best summary of the current state of the art for the staging of prostate cancer.]

Gronberg H *et al.* Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (*HPC1*) locus. *Journal of the American Medical Association* 1997; **278**: 1251–5. [This landmark paper was the first to identify the locus for the human prostate cancer susceptibility gene.]

Huggins C, Hodges CV. Studies on prostatic cancer. 1. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. *Cancer Research* 1941; **1**: 293–7. [The classic paper that described the hormone-dependent growth of prostate cancer.]

Johansson JE, Adami HO, Andersson SO, Berstroum R, Holmberg L, Krusemo UB. High 10-year survival rate in patients with early, untreated prostatic cancer. *Journal of the American Medical Association* 1992; **267**: 2191–6. [The definitive European ‘watchful waiting' series.]

Naitoh J, Belldegrun A. The Fourth Annual CaPCure Retreat: presenting the cutting edge in prostate cancer research. *Prostate Cancer and Prostate Diseases* 1998; **1**: 48–53. [This article summarizes all of the recent advances in the molecular biology and immunology of prostate cancer, along with a discussion of the novel treatment approaches that are undergoing study.]

Naitoh J, Belldegrun A. Prostate cancer gene therapy. *Prostate Cancer and Prostate Diseases* 1999 (in press). [This review article summarizes all of the recent developments in the gene therapy of prostate cancer.]

Naitoh J, Zeiner RL, deKernion JB. The diagnosis and treatment of prostate cancer. *American Family Physician* 1998; **57**: 1531–39. [A brief review of the screening controversy for the general practitioner, as well as the risks and benefits of prostate cancer screening.]

Oesterling JE *et al.* Serum prostate-specific antigen in a community based population of healthy men: establishment of age specific reference ranges. *Journal of the American Medical Association* 1993; **270:**, 860. [This paper provided definitive evidence for the utility of age-adjusted ranges for PSA-based screening programmes.]

Stamey TA, Kabalin JN, Ferrari M. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of prostate III. Radiation treated patients. *Journal of Urology* 1989; **141**: 1084–7. [One of the first papers to describe the utility of PSA in the follow-up of patients who were treated for prostate cancer.]

Stamey TA *et al.* Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of prostate II. Radical prostatectomy treated patients. *Journal of Urology* 1989; **141**: 1076.

Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. *New England Journal of Medicine* 1987; **317**: 909–16. [The seminal paper that reported the utility of PSA as a screening test for prostate cancer.]

Stein A, DeKernion J, Smith RB, Dorey F, Patel H. Prostate specific antigen levels after radical prostatectomy in patients with organ confined and locally extensive prostate cancer. *Journal of Urology* 1992; **147**: 942–6. [One of the first case series showing PSA recurrence rates following radical prostatectomy.]

Stephenson RA, Smart CR, Mineau GP, James BC, Janerich DT, Dibble RL. The fall in incidence of prostatic carcinoma: on the down side of a PSA induced peak in incidence. Data from the Utah Cancer Registry. *Cancer* 1996; **77**: 1342–8. [One of the first papers to suggest that there has been an impact of aggressive screening and treatment on the incidence of prostate cancer.]

Terris MK, McNeal JE, Stamey TA. Detection of clinically significant prostate cancer by transrectal ultrasound-guided systematic biopsies. *Journal of Urology* 1992; **148**: 829–32. [One of the first papers to describe the use of prostate ultrasound to detect prostate cancer.]

Walsh PC, Donker PJ. Impotence following radical prostatectomy: insight into etiology and prevention. *Journal of Urology* 1982; **128**: 492–7. [The classic work that revolutionized the surgical treatment of prostate cancer.]

Yang R, Naitoh J, Filipinos J, deKernion JB, Reiter RE, Loda M. Low levels of p27 protein expression predict poor disease free survival in patients with pathological T2a–T3b adenocarcinoma of the prostate. *Journal of Urology* 1998; **159**: 941–4. [This paper discusses a new prognostic marker for prostate cancer: as the molecular biology of prostate cancer becomes clearer, such papers will be more common.]

39.9.1 Tumours of the kidney

W. Turner and D. Cranston

Introduction
Renal cell carcinoma
- Incidence
- Aetiology
- Pathogenesis
- Histopathology
- Clinical features
- Investigation
- Treatment
- Follow-up and recurrence
- Prognosis

Oncocytoma and angiomyolipoma
Sarcoma
Nephroblastoma
- Incidence
- Aetiology
- Pathogenesis
- Histopathology
- Presentation
- Investigation
- Treatment
- Chemotherapy and radiotherapy
- Prognosis

Further reading

Introduction

Tumours of the kidney include those of the parenchyma and urothelial tumours in the renal pelvis (see Chapter 39.9.2). Parenchymal tumours include renal cell carcinoma, nephroblastoma (Wilms' tumour), oncocytoma, angiomyolipoma, lymphoma, and sarcoma.

Renal cell carcinoma

Incidence

Renal tumours comprise 2 per cent of all malignant tumours. There were 3776 registrations of malignant tumours of the kidney (International Classification of Diseases, 9th revision, 189) in adults in England and Wales in 1988 and these were the tenth most common tumours in men. Renal cell carcinoma is unusual below the age of 40 years; it presents at a median age of 65 to 74 and is twice as common in men as in women (Fig. 1 and Fig. 2). The incidence is increasing in both sexes, particularly over the age of 65 years (Fig. 3). Scandinavia has a high incidence, Western Europe's and North America's incidence is lower, and Japan's incidence is low.

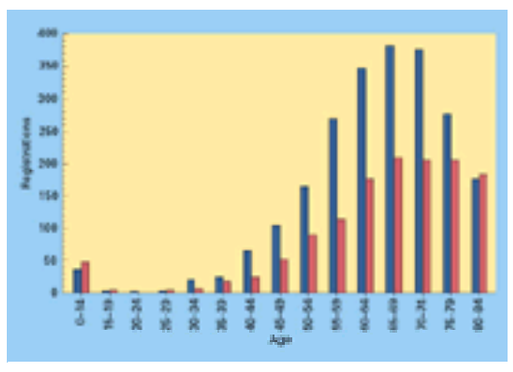


Fig. 1. Registrations of malignant tumours of the kidney in 1988 in England and Wales. The blue column indicates males, the red columns indicate females. Data are from the Office of Population Censuses and Surveys, 1994.

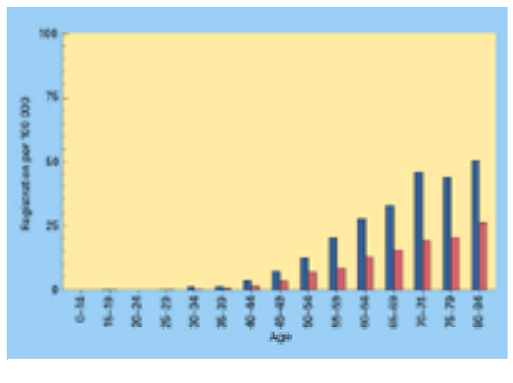


Fig. 2. Registrations of malignant tumours of the kidney per 100 000 of population in 1988 in England and Wales. The blue column indicates males, the red columns indicate females. Data are from the Office of Population Censuses and Surveys, 1994.

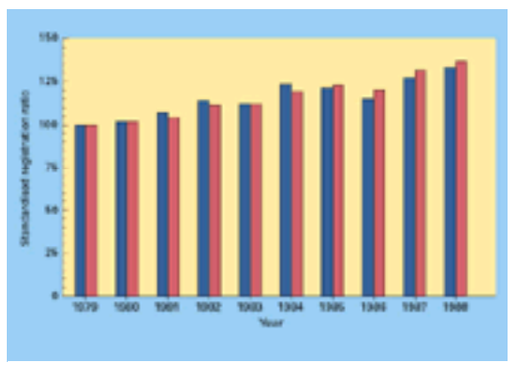


Fig. 3. Standardized registration ratios (SRR) of malignant tumours of the kidney from 1979 to 1988 in England and Wales. SRR is an index which enables comparison of incidences in populations with different age structures. The blue columns indicates males, the red columns indicate females. Data are from the Office of Population

Censuses and Surveys, 1994.

Aetiology

The aetiology of renal cell carcinoma remains uncertain, although there are a number of known associations, the most clear of which are genetic. Some patients have a familial tendency to renal cell carcinoma: this can be classified by the histology of the resulting tumour (Table 1), and has been reviewed recently by Maher.

Clear cell renal cell carcinoma	Papillary renal cell carcinoma
von Hippel-Lindau disease	Familial papillary renal cell carcinoma
Chromosome three translocations	Tuberous sclerosis
Familial clear cell renal cell carcinoma	

Table 1 Types of familial renal cell carcinoma

von Hippel-Lindau disease is well reviewed by Maher and colleagues, and Linehan and colleagues give an elegant account of the identification of the von Hippel-Lindau gene. This gene downregulates transcriptional elongation and suppresses expression of growth factors such as vascular endothelial growth factor: this may explain the occurrence of vascular central nervous system tumours in von Hippel-Lindau disease. Patients with this disease have an early age of onset of renal cell carcinoma compared with sporadic cases and tend to develop bilateral and multicentric tumours. The median survival in von Hippel-Lindau disease is around 50 years, with a 70 per cent risk of developing renal cell carcinoma by 60 years.

An increased risk of renal cell carcinoma is associated with smoking and obesity. Associations with renal cell carcinoma have been found in coke-oven workers and those with occupational exposure to cadmium or petrochemicals. In patients receiving dialysis for endstage renal failure, acquired cystic kidney disease is common, and this is a risk factor for developing renal cell carcinoma. Regular renal imaging to detect renal cell carcinoma has been suggested for patients on dialysis.

Pathogenesis

Renal cell carcinoma may present because of direct, lymphatic, and haematogenous spread or with paraneoplastic syndromes. Direct spread occurs through perirenal fat and Gerota's fascia into surrounding structures. The incidence of lymphatic spread is around 25 per cent, with involvement of the renal hilar, para-aortic, paracaval, and mediastinal nodes. Autopsy evidence suggests that lymph node metastases virtually never occur without distant metastases. Venous involvement occurs in up to 30 per cent of cases. Distant metastases occur in 20 to 30 per cent of patients at presentation, most often in the lung or bone, but also in the liver, brain, skin, or numerous other occasionally reported sites.

Histopathology

Macroscopic

Renal cell carcinoma is usually a solitary, well-demarcated mass containing yellow or brown tumour with haemorrhage and necrosis (Fig. 4). Cysts are common.

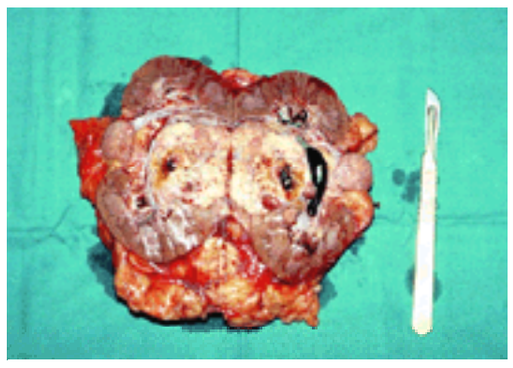


Fig. 4. Macroscopic appearance of a renal cell carcinoma.

Microscopic

The histopathological classification of renal cell carcinoma, including growth pattern, cell of origin, cytogenetics, and molecular genetics, is clearly reviewed by Motzer and colleagues. The common types are clear cell carcinoma and papillary carcinoma: loss of the von Hippel-Lindau gene is almost invariable in the former, but does not occur in the latter.

Clinical features

Presentation

The manner of presentation of symptomatic renal cell carcinoma depends on the pathological spread of the disease, and on numerous possible so-called paraneoplastic syndromes. The clinical features of patients with renal cell carcinoma are shown in Table 2.

Haematuria	55%
Pain	37%
Mass	19%
Triad	5%
Fever	12%
Weight loss	31%
Hypertension	16%
Anaemia	21%
Polycythaemia	5%
Raised ESR	33%

Table 2 Clinical features in renal cell carcinoma

The classic triad of haematuria, loin pain, and a palpable mass is rare, although 80 per cent of patients have one or more of these features. Occasionally an ectopic (e.g. pelvic) kidney may develop a tumour (Fig. 5). The primary site may also cause microscopic haematuria, and this is an important means of detection of this tumour, although less than 1 per cent of all patients with any form of haematuria will have renal cell carcinoma. Spread into the renal vein is common (Table 3), and inferior vena cava extension occurs in 9 per cent (Fig. 6, Fig. 7). Left renal vein involvement in men may cause a rapid-onset varicocele, but this is very rare. Around a quarter of patients have distant metastases at presentation, the frequency of occurrence at different sites is shown in Table 4. Metastases may sometimes produce local effects such as pneumonia, haemoptysis, bone pain, pathological fracture, or symptoms of an intracerebral mass, and they may occur at odd sites such as on the scalp or tongue (Fig. 8, Fig. 9). Systemic effects such as hypertension, weight loss, fever, and night sweats are common.

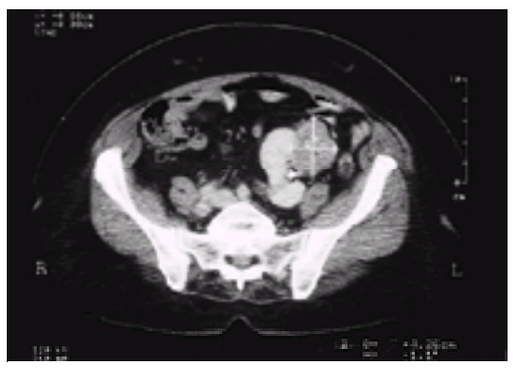


Fig. 5. Renal cell carcinoma in a pelvic kidney.

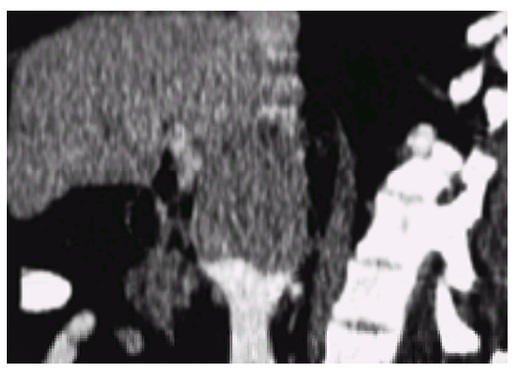


Fig. 6. Renal cell carcinoma with extension into the inferior vena cava on MRI scan (saggital section).

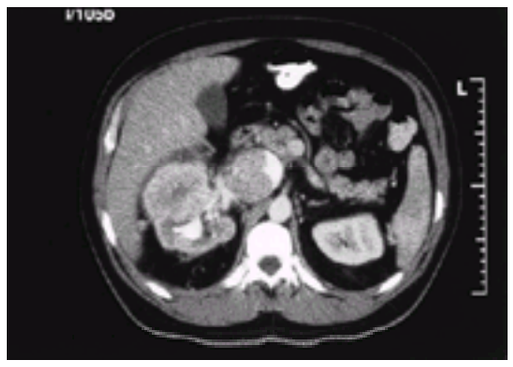


Fig. 7. Renal cell carcinoma with extension into the inferior vena cava on CT scan.



Fig. 8. Metastatic renal cell carcinoma on scalp.



Fig. 9. Metastatic renal cell carcinoma on tongue.

Renal vein spread	23%
Inferior vena caval spread	9%
Right atrial spread	0.43%

Table 3 Venous spread in renal cell carcinoma

Bone	38%
Lung	30%
Liver	6%
Brain	7%
Soft tissue	4%

Table 4 Metastatic site in renal cell carcinoma

Paraneoplastic syndromes

One-third of patients present with a non-urological feature, so-called paraneoplastic syndromes. These are early manifestations of tumours and most resolve after nephrectomy. Non-specific syndromes and specific endocrine syndromes are recognized. They are well reviewed by Sufrin and colleagues. These syndromes may cause diagnostic difficulty. The more frequent ones are shown in Table 2.

Anaemia can occur without haematuria. Increased or decreased platelet production and a variety of coagulopathies can occur. Thirty per cent of patients have abnormal liver function tests and often diffuse hepatomegaly, but without hepatic metastases: this is Stauffer's syndrome. Its importance is that the changes may be wrongly ascribed to metastases, and a potentially curative nephrectomy witheld. The hepatic abnormalities reverse with nephrectomy. Neuropathies, myopathies, vasculitis, amyloidosis (with nephrotic syndrome), raised a-fetoprotein, and disordered iron metabolism may also occur.

Renal cell carcinomas may produce renin and this can cause hypertension. Sixty per cent of patients with renal cell carcinoma have elevated erythropoietin levels, although only 5 per cent have erythrocytosis. Renal cell carcinomas may produce ectopic hormones such as parathormone, adenocorticotrophin, prolactin, human chorionic gonadotrophin, and glucagon.

Investigation

This has three purposes: first, to demonstrate a suspicious mass in the renal substance; second, to distinguish it from other lesions not requiring radical surgery (see Table 5); and third, to stage the disease and assess its effects. It should be borne in mind that an increasing number of tumours are being detected by chance during ultrasound, computed tomography, or magnetic resonance examinations performed for other reasons, so the first goal of imaging may be reached serendipitously.

Complex renal cyst
Haematoma
Abscess
Tuberculosis
Infarct
Localized inflammatory pseudotumour
Angiomyolipoma
Oncocytoma
Lymphoma
Metastasis

Table 5 The radiological differential diagnosis of a suspected renal cell carcinoma

Demonstration of a solid renal mass

There has been a clear move away from the traditional techniques of intravenous urography and arteriography, in favour of ultrasound, computed tomography, and magnetic resonance imaging. The evaluation and management of renal masses has recently been reassessed in detail by Wolf.

An intravenous urogram will demonstrate a renal mass and contralateral renal function. Ultrasound will best show if the lesion is cystic or solid (Table 6). This distinction is important because cysts occur in up to 25 per cent of the population. Wall irregularities occur in 17 per cent of cysts, and 14 per cent have septa when injected with contrast. Septa thicker than 1 mm, or with attached solid elements suggest a renal cell carcinoma. Calcification occurs in up to 3 per cent of cysts and 20 per cent of calcified cystic lesions will be renal cell carcinoma; on plain radiography, 13 per cent of renal cell carcinomas contain calcification. A good discussion of so-called complex renal cysts is given by Kawashima and colleagues.

Trans-sonic
Anechoic at high and low gain
Well defined, smooth margins
Posterior wall enhancement

Table 6 The sonographic criteria for diagnosis of a simple renal cyst

If the lesion is solid or if there is doubt about whether it is a simple cyst, computed tomography with and without contrast is indicated. Enhancement with contrast suggests a renal cell carcinoma until proved otherwise, although it may be seen with oncocytomas, angiomyolipomas, columns of Bertini, or metastases. An experienced radiologist should be able to distinguish a simple cyst from solid lesions with ultrasound and computed tomography in at least 95 per cent of cases. Magnetic resonance imaging has high resolution, and may distinguish better than computed tomography between cystic and solid lesions, it may better detect haemorrhage into cysts, and it can be used with gadolinium contrast in those with allergy to iodinated contrast who cannot have contrast computed tomography. The relative merits of computed tomography and magnetic resonance imaging in the diagnosis and staging of renal cell carcinoma are well discussed by McClennan and Deyoe.

Biopsies of indeterminate renal lesions can be made under radiological guidance, but a negative result does not rule out malignancy, so most lesions are removed surgically rather than sampled by biopsy, because of the high likelihood of renal cell carcinoma.

Staging

Renal cell carcinoma is staged for three reasons: first, to plan treatment; second, to assess prognosis; and third, to compare patient groups in research protocols. The tumour–node–metastasis (TNM) system (Table 7) or, particularly in North America, Robson's modification of Flocks and Kadesky's system (Table 8) is used.

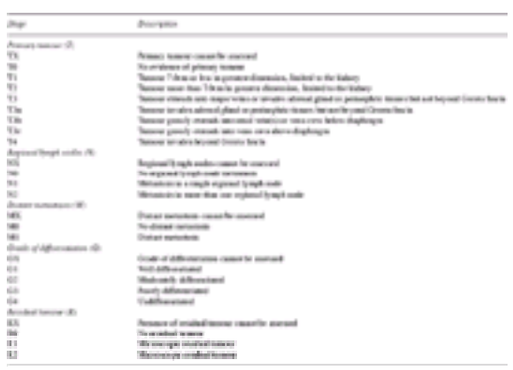


Table 7 Tumour–node–metastasis (TNM) classification for renal cell carcinoma

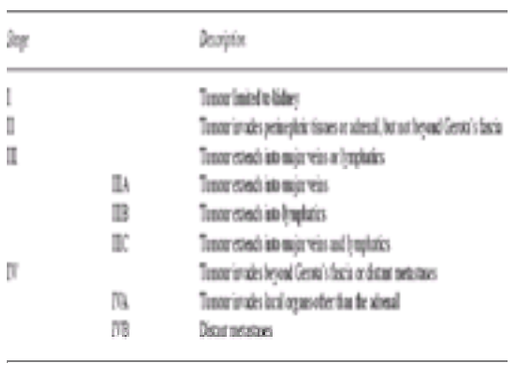


Table 8 Robson and colleagues' modification of Flocks and Kadesky' classification for renal cell carcinoma

The best current imaging modality to stage a renal cell carcinoma is computed tomography without and with contrast. The accuracy should be around 95 per cent. The major source of staging error is between T2 and T3a disease, namely the presence or absence of spread into the perinephric fat. A technically satisfactory colour Doppler ultrasound scan by an experienced radiologist which shows no venous spread is highly reliable, but where the scan is equivocal, or suggests venous thrombus, magnetic resonance imaging is now the study of choice to define best the cranial extent of the thrombus. The importance of this is that it can critically affect the surgical approach (see below). The liver is examined radiologically during the computed tomography study of the primary. Plain chest radiography is required if computed tomography or magnetic resonance imaging of the chest have not been done when examining the primary tumour. Bone scintigraphy is done in the presence of a raised alkaline phosphatase, with very large tumours, or if planned resection includes removal of extensive caval or atrial thrombus.

Treatment

Nephrectomy

Radical nephrectomy is now the standard operation for renal cell carcinoma. It entails early ligation of the renal artery then vein, and removal *en bloc* of Gerota's fascia (with its contents which include the kidney, the adrenal gland, and the perinephric fat), the upper ureter, and the renal vessels. It superseded simple nephrectomy in the 1960s when Robson began staging renal cell carcinomas and found that spread into the perinephric fat worsened prognosis. No evidence has subsequently shown that radical nephrectomy confers a survival advantage compared with simple nephrectomy. Radical nephrectomy can be done transperitoneally, via flank or thoracoabdominal incisions, or through a lumbotomy. The choice largely depends on the surgeon's familiarity with a particular incision, unless caval thrombus is present.

Ipsilateral adrenalectomy is controversial. The incidence of synchronous adrenal involvement is around 5 per cent. It occurs most often with upper pole tumours larger than 5 cm, so in other circumstances, many surgeons now would not resect the adrenal gland.

Survival rates after radical nephrectomy in relation to tumour stage are shown in Table 9.

	5 years (%)	10 years (%)
T1	80	50
T2	71	60
T3a	64	42
T3b	51	27
T4	23	No data
Positive nodes	17	No data
Synchronous metastases	3	No data
Metachronous metastases	23	No data

Table 9 Survival after radical nephrectomy for renal cell carcinoma

Lymphadenectomy

Radical nephrectomy with abdominal lymphadenectomy has been advocated because of a belief that a small number of patients without identified distant metastases have lymph node metastases. If these are of low volume, lymphadenectomy might be curative. However, there is no evidence to confirm a survival advantage. No benefit from the improved staging provided by lymphadenectomy currently exists because there is no effective therapy for metastatic disease. Survival is poor in patients with positive nodes (Table 9).

Removal of venous tumour thrombus

Around 23 per cent of patients have renal vein thrombus, and around 9 per cent have vena caval tumour thrombus (Table 3), although 50 per cent of the latter have nodal or distant metastases. Renal vein thrombus can be removed after sideclamping the junction of the renal vein and the inferior vena cava.

In the absence of metastases, removal of caval thrombus can be followed by prolonged survival, although the operative mortality is 5 to 10 per cent. Infradiaphragmatic thrombus can be removed after proximal caval control. Cardiopulmonary bypass and hypothermic circulatory arrest has been used to facilitate removal of supradiaphragmatic thrombus. The median 5-year survival after radical nephrectomy and removal of caval thrombus from a highly selected group of 230 patients from 11 series was 62 per cent. Clearly this requires meticulous preoperative assessment, to exclude metastases, and to define the level of the thrombus to allow the optimal surgical approach. This assessment now would certainly include isotopic bone scanning and spiral CT of the thorax, to exclude the two most frequent sites of metastases.

Conservative renal surgery

Conservative surgery is an issue for two groups of patients (see Table 10): those for whom radical surgery would make dialysis inevitable (mandatory indications), and those for whom conservative surgery may be desirable (elective indications). The technical feasibility of conservative surgery is presupposed, and for most surgeons this implies a small lesion (less than, say, 6 cm) away from the midpoint or hilum of the kidney. The results of conservative surgery for mandatory indications have been sufficiently good that the indications have been extended to produce the elective indications. Novick, one of the pioneers of conservative renal surgery, has recently reviewed its role.

Mandatory	Elective
Bilateral tumours	Incidentally detected tumour
Tumour in a single kidney	Familial RCC or von Hippel-Lindau disease
Patients with moderate/severe renal impairment	Normal contralateral kidney, conservative surgery preferred by surgeon or patient

Table 10 Indications for conservative renal surgery

Conservative surgery includes enucleation, local excision, and partial nephrectomy. Although renal cell carcinomas often seem to have a pseudocapsule and thus to be amenable to enucleation, at least a rim of normal-looking kidney should be removed because microscopic infiltration is very common. So-called bench surgery, whereby the kidney is excised, the tumour is removed from the kidney *ex vivo* and the kidney is then autotransplanted back into the patient, has not proved to be useful. This is largely because renal function in the operated kidney is transiently impaired after surgery.

Local recurrence is a major potential problem following conservative surgery, because 10 to 20 per cent of kidneys with renal cell carcinomas have satellite lesions and this can occur with tumours smaller than 5 cm in maximum diameter. In 14 series including a total of 897 patients, the median local recurrence rate was 4.5 per cent (although this may be greater in von Hippel-Lindau disease), and this is very similar to the 3 per cent median local recurrence rate from 6 series which included 1130 patients who had a radical nephrectomy. Survival following partial nephrectomy seems at least as good as following radical nephrectomy for tumours of similar pT stage.

When conservative renal surgery is not technically possible in a patient with a mandatory indication, the possibility of transplantation after a period of dialysis must be considered. A period of 2 years of dialysis was felt to be desirable, so as to avoid transplantation in those who would develop early metastases. However, improvements in immunosuppression regimens now make a year of dialysis reasonable, because the risk of tumour development on dialysis may now exceed the risk of developing metastases on immunosuppression.

Laparoscopic nephrectomy

This technique has been used for patients with renal tumours following demonstration of its technical feasibility in benign renal disease. Whilst it may have a place in highly skilled hands, it must be regarded as experimental at present.

Embolization

The renal artery can be occluded with a variety of materials introduced via an arterial catheter. The most appropriate role for embolization is in advanced disease with a symptomatic primary site. The symptoms may be relieved with less morbidity than by a nephrectomy. Embolization can also reduce the vascularity of very large tumours before surgery and embolization followed by surgery in patients with multiple metastases may produce regression of metastases.

Radiotherapy

Neither pre- nor postoperative radiotherapy improves survival. Radiotherapy relieves pain from skeletal metastases in up to 70 per cent of patients and can be used in conjunction with internal fixation, which can stabilize long-bone or spinal metastases.

Surgery for metastases

Renal cell carcinoma is unusual in that solitary metastases (particularly pulmonary or cerebral) can be resected with long-term survival on occasions, although this is more frequent with metachronous metastases (Table 9). In the presence of metastases, nephrectomy may be performed occasionally to control local symptoms such as pain or haematuria.

Hormone therapy and cytotoxic agents

Progestins, androgens, anti-oestrogens, steroids, and over 35 cytotoxic agents, either singly or in combination, have been used to treat disseminated disease, but none has consistently produced a response rate of over 10 per cent: the frequent overexpression of the multidrug resistance gene *mdr* in renal cell carcinoma cells seems important here.

Immunotherapy

Occasional instances of spontaneous regression suggest that the immune system is involved in the balance between host and disease in renal cell carcinoma, and so it was thought that immunotherapy might be effective against metastatic disease. The explosion of molecular biological technology over the past two decades has greatly increased our knowledge of many aspects of the biology of renal cell carcinoma and has suggested new treatment options for metastatic disease. Unfortunately, no form of systemic treatment substantially improves survival.

Interferons are proteins made by cells in response to a variety of antigenic signals: they have a direct antiproliferative action on tumour cells *in vitro*, they stimulate host lymphoid cells and macrophages, and they enhance antigen presentation. Interleukins are lymphokines produced by T lymphocytes: interleukin 2 (**IL-2**) is a growth and activation factor for lymphocytes, including natural killer cells. Recombinant DNA technology has enabled interferons and IL-2 to be produced in large amounts.

In addition to immunotherapy with cytokines, cellular immunotherapy has been used to treat advanced renal cell carcinoma. Lymphokine-activated killer (**LAK**) cells are produced by exposing harvested peripheral lymphocytes to IL-2 *in vitro*, and tumour-infiltrating lymphocytes (**TILs**) are obtained by *in vitro* expansion (with IL-2) of a single cell suspension from a radical nephrectomy specimen.

Over 1600 patients have been treated with interferons: the range of overall response rate to interferon-a is 12 to 18 per cent, with a typical duration of 6 to 10 months and usually mild or moderate toxicity. The response rate may be greater with pulmonary metastases. IL-2 monotherapy for metastatic renal cell carcinoma has been used widely and shows a response rate of about 18 per cent, lasting for up to 34 months. The toxicity of high dose IL-2 is considerable, with renal failure and cardiovascular failure the most serious problems. There is an associated mortality of 1 to 2 per cent. Lower dose subcutaneous IL-2 produces similar response rates, with less morbidity and mortality, although the duration of the responses is uncertain. Combining interferon-a and IL-2 has shown no clear benefit in randomized comparisons.

Adoptive immunotherapy, using LAK cells or TILs, has shown that TILs have greater antitumour activity, but there are no clear answers yet on the role of either treatment modality. Potential future therapies in renal cell carcinoma include radiolabelled monoclonal antibodies, the use of ‘suicide' genes (delivering virus-directed prodrug enzyme such as thymidine kinase), protection of drug-sensitive tissue by *mdr-1* gene insertion, and blockade of tumour *mdr* gene action.

Follow-up and recurrence

Renal cell carcinoma may recur in isolation many years after nephrectomy, but most recurrences occur within 3 years. The most common sites of relapse are the lung, bone, and liver, so follow-up is directed at early detection of metastases at these sites. Chest radiography, liver function tests, and serum alkaline phosphatase seem sensible investigations: the role of either ultrasound or computed tomography of the renal bed or contralateral kidney is unclear. It may be realistic to tailor follow-up to the pathological stage of the tumour. If relapse occurs, and the patient is a candidate for further surgery, full restaging to exclude multifocal recurrence is mandatory before attempting excision of either local recurrence, metachronous primary tumour, or isolated metastases.

Prognosis

The prognosis of renal cell carcinoma is difficult to predict because its natural history is variable. Very large primary tumours with venous extensions sometimes have no associated metastases, but metastases have been reported from renal cell carcinomas smaller than 2 cm. The vast majority of patients with localized disease who are not treated die of their disease and patients who are apparently tumour free may develop and die from metastases over 20 years after nephrectomy. However, the incidence of spontaneous regression of renal cell carcinoma is second only to that of melanoma and there are 5-year survivors amongst patients with untreated advanced disease.

Prognostic factors

Robson identified five prognostic factors which are the basis of current classifications: distant metastases, regional lymph node involvement, venous extension, local spread, and tumour grade. Despite extensive basic scientific research, no other factors have improved upon these. Tumour stage is generally accepted as the most important prognostic variable, with the presence or absence of metastases foremost amongst the parameters of stage. Lymph node status has independent significance, but venous involvement *per se* does not. Current and future prognostic factors have been reviewed recently on behalf of the International Union against Cancer.

Oncocytoma and angiomyolipoma

Between 5 and 10 per cent of solid renal masses are either oncocytomas or angiomyolipomas. The affected kidney is often removed for presumed renal cell carcinoma and this is probably overtreatment for oncocytoma or angiomyolipoma. These tumours may often be asymptomatic and will be detected more commonly with the increasing use of abdominal ultrasound, computed tomography, and magnetic resonance imaging. This may lead to nephrectomies which would not otherwise have been done and may be unnecessary.

Oncocytoma is a benign neoplasm of distal tubular origin. It is twice as common in men as in women and is diagnosed at a median age of 60 years. It is usually well encapsulated and rarely penetrates the renal capsule. Oncocytomas are often spherical and are a homogeneous brown colour, although larger lesions may have a characteristic central stellate scar macroscopically and on cross- sectional imaging. Distinction from oncocytic renal cell carcinoma may be difficult pathologically.

Angiomyolipoma is a hamartoma, which occurs in women, usually around the age of 60 years, in over 80 per cent of cases. They are associated with tuberous sclerosis (mental retardation, epilepsy, and sebaceous adenomas), which is caused by dominantly inherited genes, but most cases now are sporadic. They contain muscle, vascular tissue, and fat and are well circumscribed, but not encapsulated. They may vary in colour, with composition, from yellowish to pinkish-grey.

Both oncocytomas and angiomyolipomas can be bilateral and multicentric, and both can coexist with renal cell carcinoma. Oncocytomas are seldom symptomatic, whereas angiomyolipomas may present with symptoms, particularly if over 4 cm in maximum diameter. Flank pain, a mass, microscopic haematuria, and rupture with acute retroperitoneal haemorrhage may occur in angiomyolipoma.

Both lesions may be diagnosed correctly with radiology either after chance finding or after symptoms. This may allow alternatives to routine nephrectomy for solid renal masses which were previously all thought to be renal cell carcinoma.

Patients with small peripheral lesions, particularly if asymptomatic, may be left under radiological review. Those with larger, symptomatic, bilateral, multicentric, or indeterminate lesions may be treated by partial nephrectomy, or enucleation for angiomyolipoma. Deaths apparently from oncocytoma may actually be caused by oncocytic renal cell carcinoma.

Sarcoma

About 1 per cent of renal tumours are sarcomas and their incidence increases with age. Distinction from renal cell carcinoma is usually only possible with histopathology. All forms of sarcoma can occur in the kidney, but leiomyosarcoma accounts for 60 per cent. Leiomyosarcoma of the kidney tends to compress the

kidney rather than invade it and it metastasizes early and widely. Treatment is by radical nephrectomy. Survival has been increased after chemotherapy in some cases.

Nephroblastoma

Incidence

Nephroblastoma (Wilms' tumour) is the most common intra-abdominal tumour of children and accounts for 10 per cent of childhood malignancy. The condition has recently been reviewed in detail by Weiner and colleagues, and Petruzzi and Green.

The incidence is 1 per 10 000 live births or 7 per million children under the age of 15 years. There were 83 registrations of kidney and unspecified urinary tract tumours (International Classification of Diseases, 9th revision, 189) in children in England and Wales in 1988. The peak incidence is between 2 and 4 years and few cases occur after the age of 7.

Aetiology

About 1 per cent of children with nephroblastoma have associated congenital anomalies. Genetic influence in the aetiology of nephroblastoma has been borne out by detailed study of its molecular genetics. Nephroblastoma was the model for Knudson's two-hit hypothesis which proposes that two separate factors act to cause the disease. This was because cases with an associated genetic predisposition occur at earlier ages and are more commonly bilateral than sporadic cases. It was proposed that a tumour-suppressor gene which was expressed in normal individuals acted to regulate cellular proliferation in some way. Loss of this gene, Knudson's first hit, then allows tumour development in the presence of any other predisposing factor, the second hit.

Genetic defects were localized to chromosome 11p13, where a tumour-suppressor gene (*WT1*) has been identified that encodes a transcription factor and is normally expressed in the early fetal kidney. A second locus (*WT2*) has been recognized at 11p15, but the lack of linkage with either *WT1* or *WT2* in many cases of familial Wilms' tumour suggests the presence of at least one other genetic locus. Loss of heterozygosity studies have implicated sites on chromosomes 1p and 16q, although these may influence tumour progression, rather initiation. Thus the tumour seems not to conform to the two-hit hypothesis involving a single gene, as initially proposed.

Pathogenesis

Like renal cell carcinoma, nephroblastoma may present because of direct, lymphatic, and haematogenous spread. Direct spread occurs through perirenal fat and Gerota's fascia into surrounding structures. Renal hilar, para-aortic, paracaval, and mediastinal nodes may become involved. Venous involvement occurs in about 5 per cent of cases, with intracardiac extension in less than 1 per cent. Distant metastases occur most often in the lung, but can occur in the liver or other sites.

The United Kingdom Medical Research Council, the American National Wilms' Tumour Study Group (**NWTSG**) and the International Society for Paediatric Oncology (**SIOP**) have undertaken clinical trials in nephroblastoma and have provided a great deal of vital information about the disease and its treatment. The NWTSG has developed a staging system (Table 11).

Stage	Description
I	Tumour limited to kidney, completely resected without rupture
II	Tumour extends beyond kidney by direct invasion of capsule, extrarenal vascular extension, or tumour spillage; completely resected (includes biopsied tumour)
III	Incomplete surgical resection with residual tumour remaining within the abdomen (includes nodal metastases, tumour spillage, or peritoneal seeding)
IV	Any local stage with haematogenous metastasis
V	Bilateral renal involvement at time of diagnosis

Table 11 National Wilms' Tumour Study Group classification system

Histopathology

Macroscopic

Most primary tumours are more than 10 cm across and weigh over 500 g. The tumour is usually soft and friable with areas of necrosis, haemorrhage, and cyst formation. There is usually a well-defined pseudocapsule.

Microscopic

The tumour is of renal origin and contains elements of blastema, tubules, and stroma. There may also be heterologous tissue elements such as muscle and bone.

One of the most critical prognostic features is the presence or absence of anaplastic histology, defined as extreme nuclear atypia, with multipolar mitotic features and markedly enlarged nuclei. Anaplasia is rare in patients under 2 years, but occurs in 13 per cent of those over 4 years old. Focal or diffuse anaplasia can occur, the latter predicts responsiveness to aggressive chemotherapy.

Other histological abnormalities within both affected and normal kidneys are regarded as precursor lesions. These include nephrogenic rests (foci of abnormally persistent nephrogenic blastema) and nephroblastomatosis (multiple or diffuse nephrogenic rests in one or both kidneys). Nephrogenic rests occur in less than 1 per cent of infant kidneys at autopsy, but in 30 to 40 per cent of kidneys removed for Wilms' tumour.

Presentation

Nephroblastoma usually presents as a mass in the flank, with abdominal pain and haematuria in about 25 per cent, and fever, anorexia, and vomiting in about 5 per cent. In contrast to neuroblastoma, which may also present with an abdominal mass, the child is apparently well, rather than ill. Anomalies associated with nephroblastoma include aniridia, hemihypertrophy, Beckwith–Wiedemann syndrome, and external genitalia deformities.

Investigation

Ultrasound is usually the initial investigation and it will confirm that a solid mass arises from the kidney. The other kidney is examined, the liver is checked for metastases, and the renal vein and inferior vena cava are imaged to exclude venous extensions. As with renal cell carcinoma in adults, the intravenous urogram has been replaced by CT, which gives more anatomical detail and an indication of renal function. The same reservations regarding precise definition of spread into the perinephric fat, the liver, local lymph nodes, and the venous system apply to the use of CT in Wilms' as to its use in renal cell carcinoma. A chest radiograph, blood count, and biochemical profile are performed. Definitive staging, however, requires information from surgery and histology.

Treatment

In contrast to renal cell carcinoma, radiotherapy and chemotherapy have much to offer after nephrectomy in nephroblastoma. Survival from nephroblastoma has improved dramatically in recent decades with improved surgical care and refinement of radiotherapy and chemotherapy in the light of clinical trials.

Nephrectomy

Radical nephrectomy through an anterior transperitoneal approach is performed. The objective is complete tumour removal without spillage, which doubles the relapse rate. The opposite kidney is mobilized, inspected, and palpated to exclude a bilateral tumour. This occurs in about 5 per cent of cases and one-third of them are not detected preoperatively. The usual approach to bilateral tumours is a radical nephrectomy on the side of the larger tumour and removal of as much of the tumour as possible on the other side. This may necessitate a contralateral nephrectomy and dialysis followed by transplantation.

There is debate about the renal hyperperfusion syndrome which follows nephrectomy and contralateral partial nephrectomy, and can lead to chronic deterioration in renal function. It is stimulating a renal-sparing approach to surgery, particularly in the light of effective chemotherapy.

Lymphadenectomy

Extensive lymphadenectomy is not advocated, but lymph nodes are sampled to improve staging.

Removal of venous tumour thrombus

Preoperative chemotherapy is given when venous tumour thrombus is found at diagnosis and the thrombus may well recede. Removal of intracardiac tumour may require cardiopulmonary bypass.

Chemotherapy and radiotherapy

In contrast to renal cell carcinoma, nephroblastoma is sensitive to both radiotherapy and cytotoxic drugs. Numerous studies have investigated how these treatments can maximize benefit and minimize side-effects. Before the NWTSG and SIOP studies began, postoperative radiotherapy was the standard of care in children with nephroblastoma: however, its complications in this setting include the development of second tumours, scoliosis, and radiation enteritis.

The SIOP studies investigated the use of preoperative radiotherapy and chemotherapy to try to downstage tumours and reduce the need for postoperative radiotherapy. In the NWTSG studies, however, preoperative treatment is avoided on the grounds that more accurate staging can be done in the untreated patient, and that up to 8 per cent of children prove, after nephrectomy, to have some other condition.

The SIOP studies showed that preoperative vincristine and actinomycin-D were as successful as preoperative radiotherapy in downstaging and reducing surgical tumour spillage from rupture. Later studies have shown that postoperative chemotherapy reduces relapse and that lung metastases can be treated effectively by chemotherapy and salvage lung resections.

The NWTSG studies showed that anaplasia predicts poor outcome and the need for intensive treatment. Combination chemotherapy was shown to be more effective that single agents. Postoperative radiotherapy in addition to chemotherapy was found to be unnecessary in children under 2 years, those with stage I tumours, and those with stage II tumours and favourable histology. The addition of doxorubicin to vincristine and actinomycin-D enabled the dose of postoperative radiotherapy to be halved in children with stage III and favourable histology. Pulsed and shorter duration chemotherapy regimes were found to be as effective as longer ones. The latest NWTSG study is assessing the role of prognostic factors in predicting the likelihood of relapse after treatment, and this may then generate subsequent studies using these factors to stratify treatment groups.

Recurrence after treatment usually occurs in the first 2 years after nephrectomy, is difficult to manage, and a multidisciplinary approach is most successful. Ifosfamide, cisplatin, and etoposide all have activity against nephroblastoma and are used to treat recurrence.

Prognosis

The most important prognostic factors are anaplastic histology and stage. Four-year overall survival rates in the NWTSG 3 study for children with favourable histology were: stage I, 96 per cent; stage II, 91 per cent; stage III, 91 per cent; and stage IV, 81 per cent; and for children with unfavourable histology, the 4-year survival rates were: stage I, 89 per cent and stages II to IV, 54 per cent.

Further reading

Bono AV, Lovisolo JAJ. Renal cell carcinoma—diagnosis and treatment: state of the art. *European Urology* 1997; **31**(suppl 1): 47–55.[General review of renal cell carcinoma from a European perspective.]

Gelb AB. Renal cell carcinoma. Current prognostic factors. *Cancer* 1997; **80**: 981–6. [One of a series of papers from an International Union against Cancer (UICC) Workshop.]

Hrouda D, Muir GH, Dalgleish AG. The role of immunotherapy for urological tumours. *British Journal of Urology* 1997; **79**: 307–16. [General review of immunotherapy in urology.]

Johnsen JA, Hellsten S. Lymphatogenous spread of renal cell carcinoma: an autopsy study. *Journal of Urology* 1997; **157**: 450–3. [An extensive Scandinavian autopsy series reviewed with respect to renal cell carcinoma.]

Kawashima A, Goldman SM, Sandler CM. The indeterminate renal mass. *Radiologic Clinics of North America* 1996; **34**: 997–1015. [Review of the role of imaging in renal masses.]

Levy DA, Slaton JW, Swanson DA, Dinney CPN. Stage specific guidelines for surveillance after radical nephrectomy for local renal cell carcinoma. *Journal of Urology* 1998; **159**: 1163–7. [Follow-up guidelines from the MD Anderson Cancer Center, based on their experience.]

Linehan WM, Lerman MI, Zbar B. Identification of the von Hippel-Lindau (VHL) gene. Its role in renal cancer. *Journal of the American Medical Association* 1995; **273**: 564–70. [A review from the pioneering National Institutes of Health group that identified the VHL gene.]

McClennan BL, Deyoe LA. The imaging evaluation of renal cell carcinoma: diagnosis and staging. *Radiologic Clinics of North America* 1994; **32**: 55–69. [A review of the role of imaging in renal cell carcinoma.]

McDougall EM, Clayman RV, Elashry OM. Laparoscopic radical nephrectomy for renal tumour: the Washington University experience. *Journal of Urology* 1996; **155**: 1180–5. Case series from one of the pioneers of laparoscopic nephrectomy.]

Maher EH. Inherited renal cell carcinoma. *British Journal of Urology* 1996; **78**: 542–5. [A review of the genetics of the inherited renal cancer syndromes.]

Maher ER *et al.* Clinical features and natural history of von Hippel-Lindau disease. *Quarterly Journal of Medicine* 1990; **77**: 1151–63. [An extensive review of von Hippel-Lindau disease.]

Marple JT, MacDougall M. Development of malignancy in the end-stage renal disease patient. *Seminars in Nephrology* 1993; **13**: 306–14. [Broad review of malignancy in advanced renal disease.]

Motzer RJ, Bander NH, Nanus DM. Renal-cell carcinoma. *New England Journal of Medicine* 1996; **335**: 865–75. [General review of renal cell carcinoma from a North American perspective.]

Novick AC. Nephron-sparing surgery for renal cell carcinoma. *British Journal of Urology* 1998; **82**: 321–4. [An update on nephron-sparing surgery from the best known exponent.]

Ono Y, Katoh N, Kinukawa T, Matsuura O, Ohshima S. Laparoscopic radical nephrectomy: the Nagoya experience. *Journal of Urology* 1997; **158**: 719–23. [Case series of laparoscopic nephrectomy from Japan.]

Petruzzi MJ, Green DM. Wilms' tumour. *Pediatric Clinics of North America* 1997; **44**: 939–52. [North American perspective on Wilms' tumour.]

Reinberg Y, Matas A, Manivel C, Gonzalez R, Gillingham KJ, Pryor JL. Outcome of renal transplantation or dialysis in patients with a history of renal cancer. *Cancer* 1992; **70**: 1564–7. [Follow-up of endstage renal disease treatment in patients previously treated for renal cancer.]

Srigley JR *et al.* Current prognostic factors—renal cell carcinoma. *Cancer* 1997; **80**: 994–6. [A review of prognostic factors from the UICC workshop.]

Sufrin G, Chasan S, Golio A, Murphy GP. Paraneoplastic and serologic syndromes of renal adenocarcinoma. *Seminars in Urology* 1989; **7**: 158–71. [An extensive review of the paraneoplastic syndromes.]

Swanson DA *et al.* Future prognostic factors for renal cell carcinoma. *Cancer* 1997; **80**: 997–8. [A look ahead from the UICC workshop.]

Turner WH, Zingg EJ. Renal cell carcinoma: diagnosis and surgical treatment. In: Whitfield H, Hendry WF, Kirby R, Duckett JW, eds. *Textbook of genitourinary surgery*. Blackwell Science, Oxford,1998: 1117–41. [A review, with extensive collation of data from individual case series, about the presentation, diagnosis, and treatment of renal cell carcinoma.]

Weiner JS, Coppes MJ, Ritchey ML. Current concepts in the biology and management of Wilms' tumour. *Journal of Urology* 1998; **159**: 1316–25. [A paediatric urology perspective on Wilms' tumour.]

Wolf JS Jr. Evaluation and management of solid and cystic renal masses. *Journal of Urology* 1998; **159**: 1120–33. [an extensive literature review on the assessment of renal masses.]

39.9.2 Urothelial tumours

T. S. O'Brien

Epidemiology
Aetiology
Pathology
Clinical features
Investigations
Diagnostic investigations
Staging investigations
Treatment
Carcinoma *in situ*
Superficial bladder cancer
pTa tumours
pT1 tumours
Invasive bladder cancer
Bladder preservation
Transurethral resection
Radiotherapy
Bladder reconstruction
Urinary diversion
Construction of a neobladder
Chemotherapy
Disseminated disease
Urothelial tumours of the upper urinary tract
Epidemiology
Aetiology
Pathology
Clinical features
Investigations
Treatment
Adjuvant therapy
Further reading

Urothelial tumours arise from the transitional cell epithelium of the urinary tract. Bladder cancer accounts for over 97 per cent and can present a formidable challenge to the urologist. It may also cause a patient more distress than any other tumour.

Epidemiology

There are approximately 45 000 new cases of bladder cancer diagnosed each year in the United States and 10 000 patients a year die of the disease. Two-thirds of the patients are men, making it the fourth commonest cause of cancer death in men after lung, colorectal, and prostate cancers. In England and Wales in 1991, 11 470 cases were diagnosed and 4768 people died of the disease. The majority of the patients diagnosed with bladder cancer are elderly, with 80 per cent aged over 60 years and 45 per cent aged over 70 years. In 1994 in England and Wales 668 people under the age of 64 years died of the disease. In all countries the incidence has been increasing at a rate of about 1 per cent per annum since the 1950s.

Aetiology

Bladder cancer was first reported to be associated with certain occupations in 1895 when Rehn in Germany noted three cases in dye workers. Nowadays bladder cancer is regarded as a prescribed industrial disease in occupations in which chemicals of the arylhydrocarbon group are used. These include benzidine, a- and b-napththylamines, aminobipenyl, and orthotolidine. These substances formerly found use in the paints, dyestuffs, rubber, cable, and chemical industries, but none have been used since the mid-1960s. As the latent period between exposure and development of the disease averages 25 years, sporadic cases are still seen. Nowadays in the majority of patients the exact causative agent will be unclear, but an average consumption of 15 cigarettes a day doubles the risk of developing the disease and about 40 per cent of patients with bladder cancer are smokers. Other recognized causes include chronic overingestion of phenacetin-containing analgesics, treatment with the cytotoxic agent cyclophosphamide, and pelvic irrradiation. Worldwide, bilharzia caused by infestation of the bladder with *Schistosoma haematobium* remains an important aetiological factor. In Egypt 80 per cent of bladder cancers are associated with bilharzia and in these instances the tumour will be squamous cell in origin.

Pathology

Benign urothelial tumours are extremely rare. The inverted papilloma is occasionally seen in patients with bladder outflow obstruction. It is a benign proliferative tumour most commonly located at the trigone or around the bladder neck. Malignant tumours are of three main histological types. Transitional cell tumours account for over 97 per cent of urothelial tumours in the Western world. Squamous tumours account for only 1 to 2 per cent of urothelial tumours in the United Kingdom, but in Egypt squamous cancers account for 80 per cent of bladder cancers. These arise in patches of the bladder which have undergone squamous metaplasia as a result of infestation with *S. haematobium*. These bilharzial tumours are nodular exophytic lesions that are usually well differentiated with a low incidence of lymph node and distant metastases. In the Western world squamous cancers are most often caused by irritation of the mucosa from calculi or indwelling catheters. Consequently around 5 per cent of patients with paraplegia will develop a squamous carcinoma. Adenocarcinomas are extremely rare, accounting for less than 0.5 per cent of bladder cancers in the United Kingdom. These are associated with persistent urachal tissue or ectopia vesicae.

Transitional cell carcinomas are morphologically one of three types. Papillary tumours (Fig. 1) account for 70 per cent of bladder cancers. These tend to be pedunculated and grow into the lumen of the bladder. They are usually superficial, well differentiated lesions. Twenty per cent of bladder cancers are solid lesions (Fig. 2): these may appear sessile or solid with ulceration, necrosis, and calcification. They tend to be invasive lesions and are less well differentiated than papillary lesions.

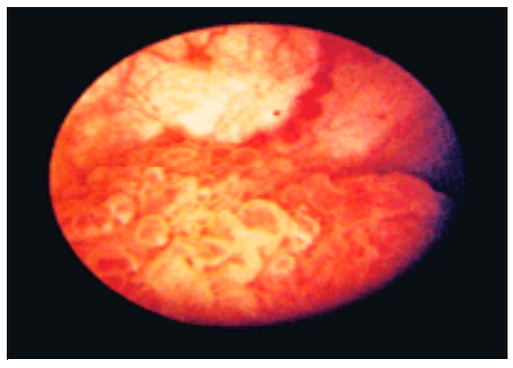


Fig. 1. Papillary bladder tumour identified at cystoscopy.

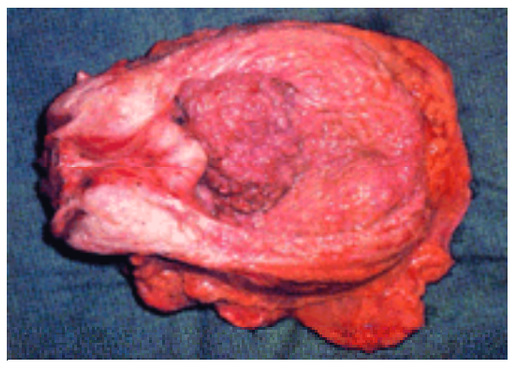


Fig. 2. Cystoprostatectomy specimen showing an invasive bladder tumour.

Carcinoma *in situ* appears as flat red velvety patches in the bladder and may be difficult to distinguish from the changes of chronic infection or inflammation. Histologically it is poorly differentiated transitional cell carcinoma confined to the mucosa. It is categorized as primary if it is the sole lesion, and secondary if the carcinoma *in situ* is found in addition to the presence of another discrete tumour. Cytological examination of the urine will reveal malignant cells in 90 per cent of patients. The behaviour of carcinoma *in situ* is difficult to predict but an important prognostic feature is the presence of irritative voiding symptoms. Only 10 per cent of patients without symptoms will progress whereas 65 per cent of symptomatic patients will progress to develop invasive disease.

Sixty per cent of cancers arise on the lateral walls of the bladder and around 30 per cent around the base or trigone.

Tumours are graded as well differentiated (G1), moderately differentiated (G2), or poorly differentiated (G3) according to the degree of cellular anaplasia. There is a strong correlation between tumour stage and grade. Most well differentiated tumours are superficial and most poorly differentiated tumours are invasive. Bladder cancers are staged according to either the new 1997 AJCC/UICC TNM classification or the Jewett–Strong–Marshall system (Table 1). Tumour stages that have been determined pathologically are prefixed ‘p' (e.g. T1 becomes pT1) (Fig. 3).

TNM		Jewett
Tis (5%)	Carcinoma *in situ*	0
Ta (15%)	The tumour does not invade muscle and does not cross the lamina propria of the bladder mucosa	0
T1 (30%)	The tumour extends beyond the lamina propria of the mucosa into subepithelial connective tissue but does not infiltrate detrusor muscle	A
T2 (15%)	Tumour invades detrusor muscle	
	T2a: tumour invades inner half of detrusor muscle	B1
	T2b: tumour invades outer half	B2
T3 (30%)	Tumour invades perivesical tissue	C
	T3a: microscopically	
	T3b: macroscopically (extravesical mass)	
T4 (5%)	Tumour invades any of the following	D1
	T4a: tumour infiltrates prostate, uterus, or vagina	
	T4b: tumour fixed to pelvic side wall or abdominal wall	
N0	No regional lymph node involvement	
N1	Involvement of a single node less than 2 cm in size	D1
N2	Involvement of a single lymph node greater than 2 cm in size or multiple small nodes	D2
N3	Involvement of a lymph node mass greater than 5 cm in diameter	
Metastases		
M0	No metastases evident	
M1	Metastases evident	D2

Figures in parentheses indicate the approximate percentage of tumours in each stage at the time of presentation.

Table 1 Staging systems for bladder cancers (Jewett–Strong–Marshall 1952 and AJCC/UICC 1997).

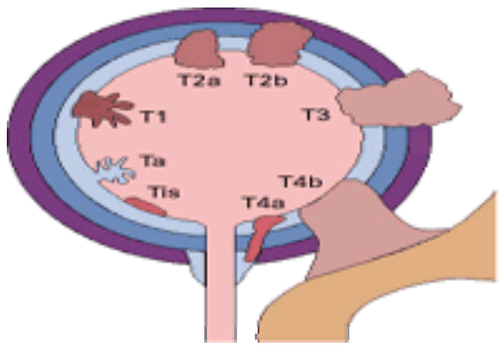


Fig. 3. Staging system for bladder cancer (see Table 1).

The genetic changes underlying the development of bladder cancers are being unravelled and deletions on chromosome 9 are the most frequent finding. Over 50 per cent of T1 tumours show deletions on the long arm of chromosome 9. In muscle, invasive alterations in the expression of the tumour suppressor gene p53 are common and may be responsible for the progression to a muscle invasive phenotype.

Tumour stage and grade, and the presence of associated carcinoma *in situ* are the best established prognostic parameters in bladder cancer. In fact over 90 per cent of patients who die of bladder cancer have G3 disease. Nevertheless, current prognostic parameters are inexact, and some superficial tumours remain localized whereas others that appear identical recur, or progress. Assigning a prognosis is a particular problem in G2 disease. Extensive multivariate analysis of large data sets has produced clinical risk profiles that are useful in day to day clinical use. Tumour grade, invasion of the lamina propria (pT1), number of tumours at presentation, tumour size greater than 3 cm, completeness of the initial resection, and whether or not a tumour is seen at the 3-month check cystoscopy are identified as the main clinical factors.

There has been an explosion of interest in identifying the molecular characteristics that determine a tumour's behaviour, both to assign more accurately a prognosis to an individual tumour and to identify new therapeutic targets. There is a strong correlation between the content of DNA and the stage and grade of tumours, so that progression and recurrence appear to correlate with aneuploidy and a high proliferative rate. The p53 tumour suppressor gene appears to play an important role in regulating bladder cancer. T1 tumours that express wild-type (normal) p53 are reported to have a progression rate of around 2.5 per cent per year compared with 20 per cent in tumours expressing mutant p53. The validity of making treatment decisions based on p53 status has not been tested however. Other key molecules appear to be the retinoblastoma tumour suppressor gene; the oncogene, epidermal growth factor receptor; and the angiogenic factor, vascular endothelial cell growth factor.

Clinical features

Painless haematuria is the predominant symptom in over 80 per cent of patients with up to 10 per cent of patients presenting with clot retention. The symptom complex of bladder irritability with dysuria, frequency, and urgency may be seen in around 20 per cent of patients. In these instances the patients will often have invasive lesions or carcinoma *in situ.* Other patients may present with recurrent urinary tract infections, loin pain secondary to an obstructed ureter, sterile pyuria, or as an incidental finding at cystoscopy. Increasingly, patients are identified as a result of routine ‘dipstick' testing of urine for haematuria: 4 per cent of these patients will have a bladder cancer.

Clinical examination will usually be normal. Occasionally in advanced disease an abdominal or pelvic mass may be palpable.

Investigations

Diagnostic investigations

The clinical problem is usually that of unexplained macroscopic haematuria. Twenty per cent of these patients will have a urothelial tumour. All patients should undergo urine culture, examination of the urine for malignant cells, an intravenous urogram, and a cystoscopy. An ultrasound examination of the kidneys is also recommended as it is a more sensitive technique than intravenous urography in identifying small renal tumours (less than 3 cm). The cystoscopy is conveniently performed using the flexible cystoscope under local anaesthesia. The intravenous urogram may show a filling defect within the bladder (Fig. 4), but is used principally to exclude a source of bleeding from the upper urinary tract. Urinary cytology is particularly helpful in suspected cases of carcinoma *in situ*, when 90 per cent of specimens will be positive. By contrast, in up to a half of patients with well differentiated tumours, cytological examination of the urine will be negative. Trans-abdominal ultrasound of the bladder is a sensitive technique for identifying papillary tumours larger than 1 cm in diameter (Fig. 5), but is a poor technique for identifying flat bladder cancers.

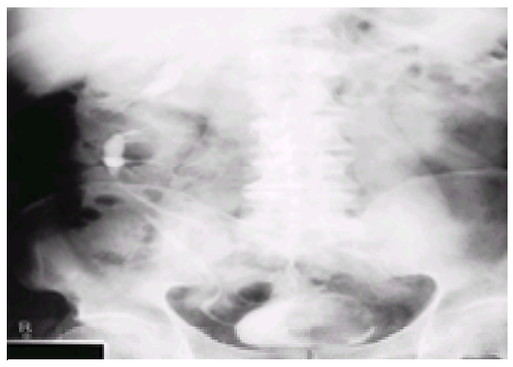


Fig. 4. Intravenous urogram showing a large filling defect in the left side of the bladder. Note also the crossed fused ectopia of the kidneys.



Fig. 5. Ultrasound appearance of a large bladder tumour projecting into the lumen of the bladder.

Urine-based diagnostic tests are of great interest because of their potential value in identifying bladder tumours non-invasively. Several are now available which aim to detect tumour-specific molecules in the urine. Tests under evaluation include those for human complement factor H-related protein, which is a component of the basement membrane (the BARD BTA test); telomerase, an enzyme which prevents chromosomal shortening in proliferating neoplastic cells; nuclear matrix protein 22, which is a nuclear transcription regulator; and a cell surface molecule detected by monoclonal antibody M344 (IMMUNOCYT). These tests have been reported to have a sensitivity of up to 80 to 90 per cent for detecting bladder tumours but cannot at present be considered as an alternative to diagnostic cystoscopy.

Staging investigations

Patients with tumours that are shown to be superficial to the detrusor muscle on histological examination need no further staging investigations. In patients with muscle invasive tumours it is necessary to differentiate between those with localized extension of the tumour and those with advanced pelvic or disseminated disease. This is best achieved by computed tomography (**CT**) (Fig. 6). It is an excellent technique for judging extravesical extension and offers a definite advantage over clinical staging by bimanual examination, which may understage a patient in up to a half of cases. CT scanning is less effective at differentiating between T2a and T2b lesions. No further advantage is offered by magnetic resonance imaging.

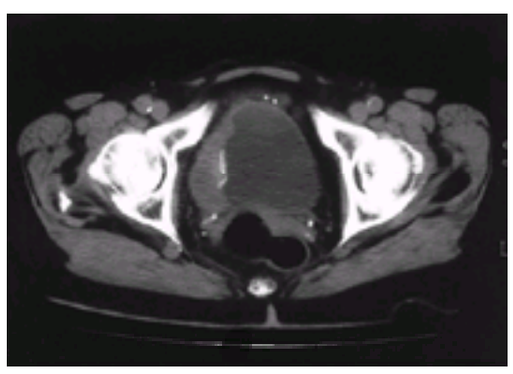


Fig. 6. Computed tomogram showing an invasive tumour in the right side of the bladder.

Pelvic lymphadenectomy provides the ‘gold standard' for pelvic nodal staging and is the only technique for distinguishing pathological from reactive change. However, lymphadenectomy is rarely performed other than at the time of cystectomy.

Treatment

All patients with bladder tumours should initially undergo transurethral resection of the tumour. This can be performed under general or regional anaesthesia. A bimanual examination of the bladder should be performed both before and after the resection in order to stage the tumour clinically. Samples of the tumour and its muscular base should be sent separately to the laboratory to aid the detection of muscular invasion. Random biopsies of apparently normal bladder distant from the tumour can be taken to look for the presence of associated carcinoma *in situ.* Further management is then largely based on the stage and grade of the tumour.

Carcinoma *in situ*

This accounts for around 10 per cent of symptomatic bladder cancers. If untreated, 50 to 60 per cent of tumours will progress to invasive disease. The initial treatment is cystoscopic resection of all macrosopic disease. Localized carcinoma *in situ* may need no further treatment, the patients merely having regular examination of their urine for malignant cells and cystoscopic surveillance.

Widespread or recurrent localized carcinoma *in situ* can be treated by intravesical instillations of **BCG** (bacille Calmette–Guérin). These are given weekly for 6 weeks with a further 3-week course after 3 months. The treatment is highly effective and has superseded treatment with other intravesical agents such as Thiotepa, Epodyl,

and mitomycin C. Side-effects of treatment with BCG are common with about half the patients experiencing an increase in frequency and dysuria. The dysuria and frequency will usually improve spontaneously but may be helped by oxybutynin, 5 mg three times daily by mouth. Two per cent of patients will develop signs of systemic BCG infection requiring antituberculous medication. After BCG around 70 per cent of patients will have resolution of symptoms and complete regression of the tumour beyond 2 years. Half of patients will be disease free at 5 years. Treatment with BCG appears also to reduce the overall incidence of progression of carcinoma *in situ* to invasive disease. Maintenance BCG regimens with BCG administered weekly for 3 weeks every 6 months for 3 years can further reduce the risk of recurrence and progression of carcinoma *in situ*, but at the cost of increased side-effects. The mechanism of action of BCG is thought to be immunological rather than a direct cytotoxic effect and as such represents a landmark in immunotherapy for cancer. Patients who fail to respond to BCG or who relapse on treatment should be offered cystectomy. Systemic chemotherapy and external beam radiotherapy are ineffective in treating carcinoma *in situ.*

Superficial bladder cancer

Seventy per cent of bladder cancers are superficial to the muscle layers of the bladder at presentation. The majority are well differentiated. Around 60 per cent of patients will develop a recurrence and 20 per cent will progress to develop invasive disease, so cystoscopic surveillance is essential. Check cystoscopies should be performed 3-monthly in the first year, 6-monthly in the second year, and yearly thereafter. A single dose of intravesical mitomycin or epirubicin given within 24 h of the initial tumour resection has been demonstrated to reduce the recurrence rate by approximately 40 per cent and so should be administered in all patients.

The findings at the initial 3-month check cystoscopy give a useful guide to prognosis. If no recurrence is identified at 3 months, 80 per cent of patients will have no further recurrence. By contrast, of those patients noted to have a recurrence at 3 months, only 10 per cent will have no further recurrence. Biopsies should always be taken of recurrences as 20 per cent of superficial tumours progress to become invasive.

pTa tumours

These comprise 35 per cent of all bladder cancers and are usually well differentiated (G1). Less than 5 per cent of pTa tumours progress to invasive disease and the 10-year survival is around 96 per cent.

pT1 tumours

These comprise 30 per cent of bladder cancers. The outlook is less favourable than with pTa tumours, with a 10-year survival around 65 per cent. More of these tumours are graded G2 or G3.

The pT1 G3 tumour is particularly sinister with up to 50 per cent progressing to invasive disease. In patients with pT1 G3 tumours a more aggressive approach to treatment is probably appropriate using adjuvant intravesical BCG (see above) or cystectomy.

Patients with multiple superficial tumours, tumours that have recurred at each check cystoscopy, or tumours with carcinoma *in situ* adjacent to the primary tumour require a course of adjuvant intravesical therapy after transurethral resection. Current evidence suggests that BCG is the agent of choice. The treatment regimen for BCG in superficial bladder tumours is the same as that used for carcinoma *in situ.* The recurrence rate at 2 years can be reduced from around 60 per cent with no treatment to around 20 per cent after treatment with BCG.

Patients who fail to respond to BCG or who develop recurrences after treatment can be offered cystectomy. Superficial bladder tumours do not respond to external beam radiotherapy.

Invasive bladder cancer

Invasive bladder cancer carries a poor prognosis with conventional treatment offering a 5-year survival of only 50 per cent or less. At centres in the United Kingdom attempts at cure have traditionally centred largely on bladder preservation that generally utilizes external beam radiotherapy, whereas in the United States cystectomy and bladder reconstruction has found favour. The goal in bladder preservation is eradication of the cancer with maintenance of satisfactory native bladder function. Only if the cancer is not cured is excisional surgery considered. The reconstructive approach involves cystectomy with the urine flow being either diverted or a neobladder constructed.

Bladder preservation

Transurethral resection

Transurethral resection alone is inadequate treatment for most muscle invasive bladder cancers, although some patients with T2 tumours can undoubtedly be cured. Overall, in patients with T2 tumours treated by transurethral resection alone, 5-year survival is around 40 per cent. These poor results in apparent T2 tumours could be caused by inadequate clearance, understaging of tumours, or early blood-borne spread of tumour. However, if patients are very carefully selected, 10-year survival in around 75 per cent has been reported. Tumours should be ideally solitary, papillary, less than 3 cm in size, and a biopsy of the bladder wall deep to the tumour must be clear if this approach is adopted. In view of the uncertainty of this approach and the disappointing results in most series, urologists tend to treat T2 tumours more aggressively.

Radiotherapy

Radiotherapy has been the mainstay of the treatment of invasive bladder cancer in the United Kingdom, but the situation is probably changing with increasing numbers of urologists favouring cystectomy. Overall, however, only 35 per cent of patients can expect to be cured by radiotherapy. Clinical features that are associated with a poor response include multifocal tumours, widespread carcinoma *in situ*, an obstructed ureter, or high tumour stage. Survival in patients with T2 tumours averages 40 per cent and in those with with T3 tumours 25 to 30 per cent. The prognosis in patients with T4 lesions treated by radiotherapy is dismal, with only 5 per cent surviving 5 years. Radical external beam radiotherapy comprises approximately 65 gray given in 2-gray fractions over 7 weeks. Side-effects are common, with around 75 per cent of patients experiencing worsening frequency and dysuria. Persistent disabling radiation cystitis occurs in about 5 per cent of patients. Following a course of radiotherapy patients require regular cystoscopic surveillance. If recurrent tumour is identified, a salvage cystectomy can be considered. In many patients a salvage cystectomy will not be feasible because of debility or the advanced nature of the tumour. In those proceding to cystectomy the overall 5-year survival is around 40 per cent. In those patients in whom no residual tumour is found in the resected specimen, 5-year survival improves to 70 per cent. Five-year survival is 50 per cent if superficial tumour is found in the specimen and only 20 per cent if invasive tumour is found.

Bladder reconstruction

In view of the disappointing results of radiotherapy in the treatment of invasive bladder cancer surgeons, particularly in the United States, have increasingly turned to cystectomy as the primary treatment modality. In men the standard operation involves radical cystoprostatectomy including pelvic lymphadenectomy. Urethrectomy is mandatory if there is tumour anywhere in the urethra but otherwise is not essential, and if the urethra can be preserved, orthotopic reconstruction of the bladder becomes feasible. Recently, nerve-sparing cystoprostatectomy has been described with the advantage that sexual function is likely to be preserved. It remains to be seen if the long-term results of this operation match those of the standard radical cystoprostatectomy.

In women the bladder is removed *en bloc* with the uterus, fallopian tubes, ovaries, and anterior wall of the vagina. A urethrectomy has traditionally been performed routinely as over a third of women with invasive bladder cancer have urethral involvement, but some authors now report preservation of the urethra allowing orthotopic bladder reconstruction in women. In any event the urethra should not be preserved in women if there is urethral tumour.

There are essentially two approaches to dealing with the urinary stream.

Urinary diversion

Traditionally this utilizes an isolated segment of ileum as a urinary conduit to an abdominal wall stoma. Urine drains continuously and is collected in a bag attached to the skin around the stoma. More recently, continent urinary diversions have been described. In these procedures an intra-abdominal pouch is fashioned from ileum (the Koch pouch) or the ileocaecal segment (the Indiana pouch). The pouch can then be catheterized intermittently by the patient through the abdominal stoma.

Construction of a neobladder

A neobladder is fashioned by forming a pouch of ileum or sigmoid. This pouch can then be sutured to the internal urethral meatus. The patient is continent because the external urinary sphincter is preserved. Patients either void spontaneously by use of the Valsalva manoeuvre or catheterize the pouch intermittently via the urethra. A major advantage is the avoidance of an abdominal stoma.

The best 5-year-survival rates that might be expected following radical cystectomy for invasive bladder cancer are 70 per cent for T2 tumours, 40 per cent for T3 tumours, and only 5 to 10 per cent for T4 tumours. Superficially these results appear superior to those that can be achieved with radiotherapy, particularly with less advanced tumours. However, it is very difficult to make direct comparisons between surgically staged cases and clinically staged cases. Clinical staging undoubtedly understages many T2 tumours.

Chemotherapy

The generally poor results of both radiotherapy and surgery in the treatment of invasive bladder cancer are explained by the presence of unrecognized distant disease at the time of the initial treatment. Accordingly, urologists are turning increasingly towards chemotherapy in the treatment of invasive bladder cancer.

Single agent treatment has so far proved disappointing, but the introduction of multiagent regimens has brought fresh hope. In particular, **M-VAC** (methotrexate, vinblastine, adriamycin, cisplatin) shows great promise. Chemotherapy can be administered after definitive treatment (adjuvant chemotherapy) or prior to definitive treatment (neoadjuvant chemotherapy). Clinical trials have provided conflicting evidence concerning the benefits or otherwise of chemotherapy, and the results of definitive studies are awaited.

Disseminated disease

The treatment of disseminated bladder cancer is essentially palliative as patients have a median survival of only 6 months. Multiagent chemotherapy using M-VAC can double the median survival with about 15 per cent of patients surviving 2 years, but it carries a significant morbidity in patients who may already be very weak.

Urothelial tumours of the upper urinary tract

Epidemiology

Urothelial tumours of the the upper urinary tract are rare and account for approximately 5 per cent of all urothelial tumours. In England and Wales around 600 cases are diagnosed each year. Urothelial tumours of the renal pelvis are four times more common than those of the ureter and account for 5 per cent of all renal tumours. Men are twice as susceptible as women and the peak incidence is in the seventh decade.

Aetiology

Aetiological factors are similar to those in bladder cancer so that cigarette smoking accounts for around half of cases. Seventy per cent of patients who develop an analgesic-induced nephropathy will develop a transitional cell carcinoma of the upper urinary tract. In the Balkans one-half of all renal tumours are urothelial and are associated with the presence of Balkan nephropathy.

Pathology

Ninety per cent of upper tract urothelial tumours are transitional cell carcinomas. In 2 to 5 per cent of cases the tumours are bilateral.

Upper tract urothelial tumours are associated with carcinoma of the bladder: 3 per cent of patients with a bladder cancer will develop a tumour of the upper urinary tract at some stage, and 50 per cent of patients with an upper tract urothelial tumour will develop a bladder cancer. Interestingly, bladder tumours that develop in the presence of an upper tract transitional tumour and the upper tract transitional tumour itself have been demonstrated to be monoclonal, suggesting that the bladder tumour arises by implantation of tumour cells shed from the upper tract tumour.

Transitional cell carcinoma of the upper urinary tract may spread by direct invasion, or by invasion of lymphatics, blood vessels, or adjacent mucosa. Mucosal invasion probably accounts for the high incidence (30 per cent) of stump recurrences following conservative surgery.

The system of staging is similar to that used for bladder cancer (Table 2). Stage and grade are closely correlated. Median survival is 6 years in those with low-grade tumours and 1 year in those with high-grade tumours. Median survival is 8 years in those with low-stage tumours (Ta or T1), but only 1 year in those with advanced stages (T2 to 4).

Histological finding	*Grabstald-Cummings*	*TNM*
Carcinoma *in situ*	I	Tis
Non-invasive papillary	I	Ta
Invades subepithelial connective tissue	II	T1
Invades muscle	III	T2
Invades renal parenchyma or peripelvic tissue or periureteral tissue	III	T3
Invades adjacent organs	IV	T4
Lymph node metastases	IV	N1–3
Distant metastases	IV	M1

Table 2 Staging systems for upper tract urothelial tumours (Grabstald–Cummings 1975 and AJCC/UICC 1997)

Clinical features

Seventy-five per cent of patients will present with haematuria. The bleeding will usually be throughout the stream. The patient may occasionally pass a clot cast of the ureter. A third of patients will have loin pain due to obstruction of a ureter, or may have ureteric colic which can mimic that due to the much more commonly seen ureteric stone.

Investigations

The clinical problem is usually that of unexplained haematuria, so patients will undergo the same investigations as patients with suspected bladder cancer. The intravenous urogram will usually show a filling defect in the renal pelvis or ureter. The differential diagnosis of the filling defect includes overlying bowel gas, radiolucent stone, blood clot, or a sloughed papilla. If doubt exists after the urogram as to the nature of the filling defect, a retrograde ureterogram can be performed.

Optional investigations which may be useful include analysis of ureteral urine samples for malignant cells, and brush biopsy of a suspected tumour.

Ureteroscopy and biopsy is used increasingly in the diagnosis of upper urinary tract tumours. By providing tissue for histological examination, ureteroscopic biopsy may allow detection of patients with low-stage/grade lesions suitable for conservative surgery. The flexible steerable ureteroscope has revolutionized the diagnosis of these tumours because it makes instrumentation and examination of the ureter so atraumatic. Computed tomography provides the best non-invasive means of staging the

tumour preoperatively.

Treatment

The traditional treatment for upper tract urothelial tumours is nephroureterectomy with excision of a cuff of bladder around the ureteric orifice. The justification for this radical treatment has been the reported high incidence (30 to 75 per cent) of ureteral recurrences in patients treated by more conservative surgery. In patients treated by nephroureterectomy, 5-year survival is around 90 per cent for stage Ta, 45 per cent for stage T1, 23 per cent for stages T2/3, and 0 per cent for stage T4.

More conservative surgery is increasingly finding favour for patients with low-grade/stage tumours in whom the results of local excision are equivalent to those achieved with nephroureterectomy. Conservative surgery can be performed using an antegrade approach with resection of the tumour through a nephroscopic port. This is best for renal pelvic lesions. Brachytherapy to the access tract probably helps reduce the incidence of track site recurrence. Retrograde ureteroscopic approaches can be used for the treatment of ureteric or renal pelvic lesions. The flexible ureteroscope facilitates this, and ablation of the tumour with laser is particularly convenient because of the small-calibre fibre through which the laser energy can be transmitted. However, conservative endourological surgery should still be considered an experimental approach if the patient has a normal contralateral kidney, but is particularly appropriate if the patient has a single kidney or has bilateral tumours.

Adjuvant therapy

The radiotherapy and chemotherapy regimens for the treatment of urothelial tumours of the upper urinary tract are similar to those used in bladder cancer. Each may have some benefit but a definitive role has yet to be established for either.

Further reading

Bloom HJ, Hendry WF, Wallace DM, Skeet RG. Treatment of T3 bladder cancer: a controlled trial of preoperative radiotherapy and radical cystectomy versus radical radiotherapy. *British. Journal of Urology* 1982; **54**: 136–51. [A classic study comparing a bladder preservation strategy with cystectomy.]

Esrig D *et al.* Accumulation of nuclear p53 and tumour progression in bladder cancer. *New England Journal of Medicine* 1994; **331**: 1259–64. [A comprehensive analysis of the results of cystectomy for invasive bladder cancer and assessment of the effect of the p53 tumour suppressor gene.]

Herr H *et al.* Intravesical bacillus Calmette–Guérin therapy prevents tumour progression and death from superficial bladder cancer: ten year follow-up of a randomised trial. *Journal of Clinical Oncology* 1995; **13**: 1404–8. [The best evidence that BCG makes a difference in preventing progression of high-risk superficial bladder cancer.]

Kurth KH, Denis LJ, Bouffioux C. Factors affecting recurrence and progression in superficial bladder cancer. *European Journal of Cancer* 1995; **31**: 1840–6. [Thorough analysis of EORTC trials for clinical risk factors for progression and recurrence in superficial bladder cancer.]

Mameghan H, Fisher R, Mameghan J, Brook S. Analysis of failure following definitive radiotherapy for invasive transitional cell carcinoma of the bladder. *International Journal of Radiation. Oncology, Biology, Physics* 1995; **31**: 247–54. [A useful practical guide to successes and failures following radiotherapy for bladder cancer.]

Raghavan D, Shipley W, Garnick M, Russell P, Richie J. Biology and management of bladder cancer. *New England Journal of Medicine* 1990; **322**: 1129–38. [A readable excellent overview of all aspects of bladder cancer with extensive reference list.]

Solsona E, Iborra I, Ricos JV, Monros JL, Casanova J, Calabing C. Feasibility of transurethral resection for muscle infiltrating carcinoma of the bladder—long term follow up of a prospective study. *Journal of Urology* 1998; **159**: 95–9. [The best study of transurethral resection alone for invasive bladder cancer.]

Stadler WM, Kuzel TM, Raghavan D. Metastatic bladder cancer: advances in treatment. *European Journal of Cancer* 1997; **33**: S23–26. [An update on chemotherapy in advanced bladder cancer.]

Sternberg C. Neoadjuvant and adjuvant chemotherapy in locally advanced bladder cancer. *Seminars in Oncology* 1996; **23**: 621–32. [An excellent review of the role of chemotherapy in bladder cancer.]

Tolley D, Parmar M, Grigor K, Lallemand G. The effect of intravesical mitomycin C on recurrence of newly diagnosed superficial bladder cancer: a further report with 7 years of followup. *Journal of Urology* 1996; **155**: 1233–8. [A single dose of chemotherapy given within 24 h of resection dramatically reduces future recurrence rates of superficial bladder cancer.]

39.9.3 Tumours of the testicle

Sam D. Graham, Jr and D. Cranston

Introduction
Pathology
Etiology
Staging
Clinical features
Investigations
Treatment
Other tumors of the testicle
Conclusion
Further reading

Introduction

Testicular tumors are one of the most common solid tumors of men between the ages of 15 and 35 years, with an estimated incidence of 2/100& 000 men at risk. In the United States, this accounts for an annual rate of 7600 cases in 1998 and an annualized death rate of 400 males. Carcinoma of the testicle is seen predominantly in white males, with the highest incidence reported in Scandinavia and New Zealand and the lowest incidence in Africa and Asia. Before the advent of *cis*-platinum-based chemotherapy and the multidisciplinary approach, the mortality rates from this cancer exceeded 50 per cent, and as can be seen from the above statistics, is currently below 10 per cent.

Pathology

Most testicular cancers (90–95 per cent) arise from the germ cells (spermatogenic cells). The usual classification of testicular tumors is based upon the histologic patterns of seminomatous and non-seminomatous elements (Table 1). Most tumors, however, contain multiple histologic elements and are designated as mixed germ-cell tumors. In addition to the histologic elements characterizing the testicular cancers, there are two types of primitive cells, trophoblastic and yolk-sac cells, that may also be present. Trophoblastic elements may be present in all histologic types of testicular cancers and are responsible for the production of b-human chorionic gonadotrophin, which can be detected in the patient's serum. Yolk-sac tumors produce a-fetoprotein, which can also be detected in serum, and both of these provide good tumor markers, the concentrations of which depend on the amount of tumor in the body. Serial determinations of a-fetoprotein and b-human chorionic gonadotrophin after surgery closely reflect the effectiveness of treatment and allow a reliable means of determining distant recurrence.

GERM-CELL TUMORS
Seminomas
Classic
Spermatocytic
Non-seminomatous tumors
Teratoma
Embryonal-cell
Yolk-sac carcinoma
Choriocarcinoma
NON-GERM-CELL TUMORS
Stromal tumors
Leydig-cell
Sertoli-cell
Secondary tumors
Lymphoma
Metastatic

Table 1 Histologic characterization of testicular carcinomas

Classically, germ-cell tumors have been divided into seminomatous and non-seminomatous. Seminomas appear in older men, with a peak incidence at 34 to 40 years of age, while the non-seminomatous germ-cell tumors appear in younger men aged 25 to 35 years. However, it is now clear that there is often a range of different malignant tissues in each tumor, which are often a mix of seminomatous and non-seminomatous elements. The earlier classification of testicular tumors is now largely superseded by systems that attempt to quantify how much of each tissue is present. The presence of yolk-sac or trophoblastic tissue tends to be associated with a worse prognosis than the presence of differentiated tissue. The British Testicular Tumour Panel formalized the British version of testicular tumor nomenclature and this can be compared (Table 1) with that of the World Health Organization, which modified the earlier North American classification of Mostofi and Price.

Etiology

Evidence supports the importance of congenital factors, with a higher risk in men with a family history of testicular tumor. The most common etiological factor is cryptorchidism: up to 14 per cent per cent of patients with testicular tumors have a history of maldescent; their incidence of testicular cancer is 3 and 14 times that in otherwise healthy males. The highest risk of testicular carcinoma is in those patients with testicles positioned within the abdomen rather than the scrotum. Usually, in patients with cryptorchidism, the tumor forms within the cryptorchid testicle, but up to 25 per cent form in the contralateral testicle. Other factors associated with testicular carcinoma include infection with the human immunodeficiency virus, history of previous testicular carcinoma, and Caucasian race.

Staging

A variety of clinical staging systems have been devised for testicular tumors, but the most commonly used is the AJCC/UCC (TNM) system (Table 2) The testicle itself is classified T1 to T4 depending on the extent of spread through it. Lymph-node involvement is either N1, involving the para-aortic nodes, or N2, involving nodes in the pelvis or mediastinum. M signifies bloodborne metastases in lung, liver, or brain. Computed tomographic (**CT**) scanning is one of the most effective ways of demonstrating spread. Testicular tumors spread locally into the epididymis and spermatic cord, and by lymphatic invasion along the lymphatics to retroperitoneal, para-aortic lymph nodes. Patients with invasion of the cord structures or tunica are at the highest risk for nodal metastases. Choriocarcinoma can also metastasize hematogenously, bypassing the retroperitoneal nodes.

TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma *in situ*
T1	Tumor has not spread beyond the testicle and epididymis
T2	Tumor has spread to blood vessels, lymphatics, or tunica vaginalis
T3	Tumor invades spermatic cord
T4	Tumor invades scrotum
NX	Regional nodes cannot be assessed
N0	No nodal metastasis
N1	Metastasis to single lymph node ≤ 2cm in greatest dimension
N2	Metastasis to single lymph node ¿ 2cm, but ≤ 5cm, or multiple nodes ≤ 5cm
N3	At least one node ¿ 5cm
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis is present
S1	LDH < 1.5 × normal, *β*-hCG < 5000mIU/ml, *α*-FP < 1000ng/ml
S2	LDH 1.5–10 × normal, *β*-hCG 5000–50000mIU/ml, *α*-FP 1000–10000ng/ml
S3	LDH ¿ 10 × normal, *β*-hCG ¿ 50000mIU/ml, *α*-FP 10000ng/ml

LDH, lactate dehydrogenase; *β*-hCG, *β*-subunit, human chorionic gonadotropin; *α*-FP, *α*-fetoprotein.

Table 2 AJCC/UCC staging of testicular carcinoma

Clinical features

The majority of patients with testicular tumors present with a painless lump in the testicle (Fig. 1), although up to 30 per cent are first diagnosed after presentation with metastatic disease. Examination discloses a lump in the testis and not separate from it. A testicular carcinoma is a tumor that is often detected by self-examination. Other presentations include a sudden scrotal swelling such as a hydrocele, or systemic symptoms such as gynecomastia. Any recent testicular swelling not associated with voiding symptoms or an active sediment should be suspect, and, in the absence of any laboratory test that unequivocally points to the diagnosis, inguinal exploration should be considered.



Fig. 1. A seminoma in a 45-year-old man presenting as a painless lump in the testicle.

Investigations

Ultrasonographic scanning, while not essential, will often confirm the clinical diagnosis. Blood should be sent for tumor markers (a-fetoprotein and b-human chorionic gonadotrophin). This should be done both pre- and postoperatively.

Treatment

Orchidectomy should be performed through an inguinal approach with a clamp placed on the spermatic cord, which theoretically prevents dislodgement of tumor during manipulation of the testicle (Fig. 2, Fig. 3). A frozen-section biopsy examination can be done if there is any diagnostic doubt, but this is not normally necessary. The cord is then transfixed and ligated at the internal inguinal ring, and the testicle removed. CT scans allow clinical staging of the disease.

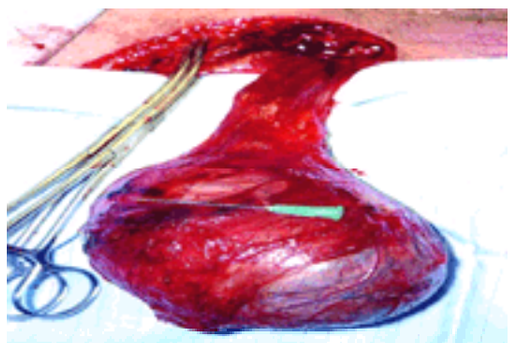


Fig. 2. Orchidectomy for seminoma.



Fig. 3. Postorchidectomy specimen showing large seminoma.

Subsequent treatment depends on the histologic type and staging of the tumor. A seminoma with negative serum markers after orchidectomy and without any detectable metastasis can be treated by radiotherapy of 3000& Gy to the retroperitoneal lymph nodes; survival is almost 100 per cent. Recently, it has been suggested that an equally good effect can be obtained with a short course of carboplatinum.

In other non-seminomatous germ-cell tumors with no evidence of spread and normal tumor markers, the options are either to keep the patient under surveillance with monthly measurement of markers and CT scans, or to perform a radical dissection of the retroperitoneal lymph nodes, which is indicated in patients with tumors involving the lymphatics, vessels, tunica, or cord structures. In the absence of involvement of these structures, the option of surveillance can be considered.

Regardless of the original histologic type, if there is evidence of retroperitoneal nodal disease, combination chemotherapy is given and the patient is followed by regular measurement of tumor markers and CT scans. When the tumor volume is small, the masses will usually disappear completely, with a survival rate of over 96 per cent. When the tumor volume is large, a mass that may contain active tumor can remain after chemotherapy. In these cases a retroperitoneal lymph-node dissection is performed so that the tissue can be examined, and if tumor is found further chemotherapy may be given. Cisplatin is one of the most active agents in the treatment of testicular cancer and various drug regimens including this agent have been advocated. One of the more commonly used regimens for advanced, non-seminomatous germ-cell tumors involves four cycles of bleomycin, etoposide, and cisplatin.

Other tumors of the testicle

Other tumors are rare, constituting between 5 and 10 per cent of all testicular tumors, and include Leydig-cell and Sertoli-cell tumors, both of which have a good prognosis.

Lymphoma of the testis accounts for 5 per cent of testicular tumors and is the most frequent testicular tumor in men over 50 years of age. The testis is diffusely enlarged to 5& cm in diameter or more. Survival is poor in patients with bilateral disease or with lymphoma at other sites, but is better if the disease is confined to the

testis.

Metastatic spread of tumors to the testis is very rare.

Conclusion

Testicular cancer has become one of the most curable of solid neoplasms and serves as an excellent demonstration of the success of a multidisciplinary approach to the treatment of malignant disease.

Further reading

Doherty AP, Bower M, Christmas TJ. The role of tumor markers in the diagnosis and treatment of testicular germ cell cancers. *British Journal of Urology* 1997; **79**(2): 247–52.

Donohue JP, Foster RS. Retroperitoneal lymphadenectomy in staging and treatment. The development of nerve-sparing techniques. *Urologic Clinics of North America* 1998; **25**(3): 461–8.

Farrer JH, Walker AH, Rajfer J. Management of the post-pubertal cryptorchid tests: a statistical review. *Journal of Urology* 1985; **134**: 1071.

Herr HW, Sheinfeld J. Is biopsy of the contralateral testis necessary in patients with germ cell tumors? *Journal of Urology* 1997; **158**(4): 1331–4.

Mostofi FK, Price B. Tumors of the male genital system. In: *Atlas of tumor pathology*, 2nd series, Fascicle 8. Armed Forces Institute of Pathology, Washington DC, 1973.

Mostofi FK, Sobin LH. *International histological classification of tumors of testes*. World Health Organization, Geneva, 1977.

Pugh RCB, Cameron K. Teratoma. In: Pugh RCB, ed. *Pathology of the testis*, pp. 199–244. Blackwell Scientific, Oxford, 1976.

Richie JP. Neoplasms of the testis. In: Walsh PC, Retik AB, Stamey TA, Vaughan ED, ed. *Campbell's urology*, 6th edn, pp. 1222–63. Saunders, Philadelphia, 1992.

Sharir S, Foster FS, Donohue JP, Jewett MA. What is the appropriate follow-up after treatment? *Seminars in Urologic Oncology* 1996; **14**(1): 45–53.

Steele GS. Richie JP. Current role of retroperitoneal lymph node dissection in testicular cancer. *Oncology* 1997; **11**(5): 717–29; discussion, 730–7.

39.9.4 Carcinoma of the penis

D. Cranston

Introduction
Pathology and staging
Clinical features
Management
Balanitis xerotica obliterans
Further reading

Introduction

Squamous cell carcinoma of the penis is the most common tumour of the penis. It is seen in uncircumcised men and is extremely rare in those who have been circumcised. Neonatal circumcision offers almost 100 per cent protection, but circumcision later in life does not seem to be as protective. Accounting for less than 1 per cent of all malignancies, squamous cell carcinoma occurs most frequently in the 50- to 60-year-old age group and has an incidence in the United States of 1 to 2 cases per 100 000 per annum. There is increasing evidence that squamous cell carcinoma of the penis is associated with the human papilloma virus and is associated with poor hygiene, being unusual in men who are able to retract and clean the foreskin.

In situ carcinoma of the penis was first described in 1911 and is known as erythroplasia of Queyrat if it involves the preputial sac, or Bowen's disease if (rarely) it involves the skin of the shaft of the penis. It is a precancerous condition that responds to local irradiation or 5-flurouracil cream. Erythroplasia of Queyrat presents as a red velvet-like plaque on the glans or prepuce that can be solitary or multifocal, and about one-third of patients will go on to develop invasive carcinoma.

Squamous cell carcinoma is usually seen on the glans or inner surface of the prepuce, and may have a superficial or vertical growth pattern, the latter being much more aggressive.

Pathology and staging

Squamous cell carcinoma of the penis may be seen with varying degrees of mitotic activity. It may be very well differentiated and present as a verrucous carcinoma, most are well to moderately differentiated, and rarely it is undifferentiated. It arises on the glans and spreads locally invading the corpora spongiosum and then the venous system, although initially it tends to spread through the lymphatics to the superficial and deep inguinal and the pelvic lymph nodes, especially around the obturator fossa. It is unusual for the pelvic nodes to be involved in the absence of inguinal node involvement. Although half of all patients will present with enlarged nodes at the time of diagnosis, less than a third will actually have nodal metastases. This is due to the fact that there is often associated infection that causes enlargement of the localized inguinal lymph nodes and it is important not to assume that they are infiltrated with tumour as they often resolve following treatment of the primary tumour.

The most common classification is that of Jackson (Table 1).

Stage I (A)	Tumours confined to glans, prepuce, or both
Stage II (B)	Tumours extending on to shaft or corpora
Stage III (C)	Tumours with inguinal metastases that are operable
Stage IV (D)	Tumours involving adjacent structure, inoperable inguinal metastasis, or distant metastasis

Table 1 Jackson's classification for carcinoma of the penis

Clinical features

Classically, carcinoma of the penis (Fig. 1) appears in elderly men with poor hygiene of the uncircumcised penis. The first sign of penile carcinoma is usually an ulcerated area over the glans, which slowly progresses. If the patient has a phimosis, then a lump may be felt under the foreskin, or a discharge may be noted from the penis. Rarely the first sign may be an enlarged inguinal node. The diagnosis is made by taking a biopsy of the lesion. Computed tomography, magnetic resonance imaging, and urethroscopy may all be helpful in certain situations to establish evidence of spread.



Fig. 1. Carcinoma of the penis.

Management

Treatment of squamous cell carcinoma of the penis is controversial. Occasionally, stage I disease may be cured by circumcision if the tumour is *in situ* or very small and limited to the foreskin. Tumours on the glans of the penis can be treated locally with radiotherapy, often using radioactive iridium-192 wires over the penis, or interstitially. Local excision can be performed either by laser surgery using carbon dioxide or neodymium YAG, or partial amputation as a guillotine resection with 2 cm tumour clearance, leaving the urethra 1.5 cm longer to form a good urethrostomy. In stage II disease, radiotherapy may be tried although local amputation is often necessary. Stage III disease requires radical amputation of the penis with reconstruction of the urethra with a scrotal flap. In stage IV disease, preliminary radiation may be attempted, but this often fails and block dissection of the inguinal nodes may be worthwhile, although once the inguinal nodes are involved the outlook becomes much worse.

Squamous cell carcinoma of the penis is partially responsive to chemotherapy, especially bleomycin, methotrexate, and cisplatin, and may be worth consideration.

Careful follow-up is necessary. Relapses occur in 10 to 50 per cent of patients receiving radiotherapy or laser treatment. Most local recurrences following conservative therapy occur within 3 years. The prognosis in squamous cell carcinoma of the penis is mainly related to the presence and extent of lymph node involvement, and varies from 85 per cent or greater 5-year survival with negative nodes, to 50 to 60 per cent survival with unilateral nodes, and 15 per cent for Jackson stage II not treated with primary lymphadenectomy.

Balanitis xerotica obliterans

This most frequently presents as a white patch on the prepuce or glans occurring in both circumcised and uncircumcised men. Histologically it is similar to lichen sclerosis et atrophicus. Meatal stenosis is treated by dilatation, meatotomy, or topical steroids. Malignant change occurs rarely.

Further reading

Jackson SM. The treatment of carcinoma of the penis. *British Journal of Surgery* 1966; **53**: 33. [Classic staging of the disease.]

Lynch DF, Schellhammer P. Tumors of the penis. In: Walsh PC, Retik AB, Vaughan ED, Wein AJ, eds, *Campbell's urology*, 7th edn, pp. 2453–85. WB Saunders, Philadelphia, 1998. [Good overview.]

Pizzocaro G, Piva L, Bandieramonte G, Tana S. Up to date management of carcinoma of the penis. *European Urology* 1997; **32**: 5–15. [Good review paper.]

39.10 Torsion of the testicle

D. Cranston

Introduction
Etiology
Diagnosis

Introduction

Torsion of the testis or strictly speaking torsion of the spermatic cord has been a recognized surgical emergency since Delasiauve first described a case affecting an undescended testicle in 1839 at Hotel Dieu in Paris. It remains one of the few true surgical emergencies, and in young men still causes an unacceptably high incidence of testicular loss, which now carries with it an increasing threat of litigation. Urgent referral and urgent exploration remain the key to preserving testicular viability. Torsion may occur at any age but is most common around puberty; it is unusual after the age of 25 years.

Etiology

The usual predisposing cause of testicular torsion is a high investment of the tunica vaginialis, which allows the testicle to hang like a bell clapper inside and the testicle and spermatic cord to rotate within the tunica. This is a bilateral abnormality and it is important to fix both sides at the time of surgery. Intermittent testicular torsion is a well recognized entity.

Diagnosis

The diagnosis is made primarily on the history and examination. A high level of suspicion is mandatory in young men who present with acute pain and swelling, where it has been shown that testicular torsion accounts for nearly 90 per cent of acutely presenting scrotal symptoms in the 13- to 21-year age group. Vomiting is often a feature. The scrotum must always be examined, as in the early stages pain may be referred to the groin or iliac fossa. The testis often has a high, horizontal lie in the scrotum. Oedema and erythema of the scrotum are usual features in torsion and do not support a diagnosis of epididymo-orchitis, which is very unusual in this age group. Torsion of a testicular appendage is more common in prepubertal boys, as is orchitis and idiopathic scrotal oedema. Rarely haemorrhage into a testicular tumour can present with acute scrotal pain.

Investigative techniques are usually unnecessary and delay exploration. Colour Doppler ultrasound may be helpful in the diagnosis but can be misleading, especially in cases of intermittent torsion where a hyperaemia can occur after spontaneous untwisting.

Surgery within 4 h usually allows testicular preservation, some atrophy occurs between 4 and 8 h, and after 10 h ischaemic necrosis is virtually inevitable (Fig. 1). Fixation of the testicle should be bipolar with non-absorpable sutures, as there are case reports of recurrent episodes of torsion after fixation with catgut sutures. The contralateral side should be fixed at the same time unless there is severe infection secondary to ischaemic necrosis. Orchidectomy is the best course of action if the testis is non-viable. Semen quality is reduced in men following unilateral torsion, and while the mechanism remains unclear, there is some evidence to suggest that restoring the blood supply to an ischaemic testis stimulates the production of antitestis and antisperm antibodies.

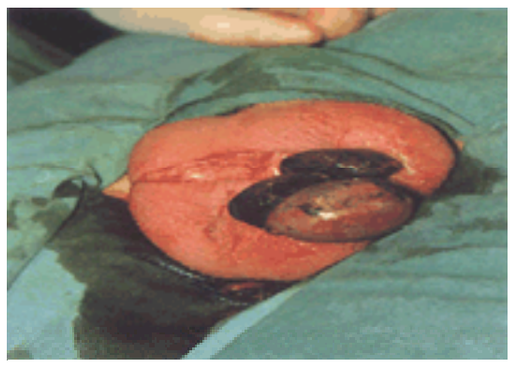


Fig. 1. Torsion of the testicle.

39.11 Vasectomy

D. Cranston

Introduction
The operation
Follow-up
Complications
Medicolegal aspects
Reversal of vasectomy
Immunology
Conclusion
Further reading

Introduction

Vasectomy is the most reliable method of birth control, used by over 43 million people worldwide. By convention, in the United Kingdom, two specimens of semen are examined after vasectomy to confirm the success of the operation, before allowing the couple to abandon other forms of contraception. Previously vasectomy was thought by most who had it to be irreversible and without failure; neither supposition is correct.

The operation

Vasectomy is usually done under local anaesthesia, although a few surgeons prefer a general anaesthetic. At the Elliot-Smith Clinic in Oxford, 30 000 vasectomies have been performed over the last 30 years. In the surgical technique usually employed there, bilateral, 1-cm incisions are made in the scrotum after locating the vas and infiltrating with 2 to 3 ml of 1 per cent lignocaine (lidocaine) with adrenaline (epinephrine). Once the skin has been cut, the vas is held with forceps and pulled into the wound to form an inverted ‘U'. The adventitia is stripped away, 1 to 2 cm removed, and the lumen cauterized. The ends of the vas are separated into different tissue planes and dropped back under the skin. No sutures are used, and the vas is not routinely sent for histological examination. The author now uses the ‘Li' method: this is described as the ‘no scalpel' method, and uses very pointed forceps, which can breach the skin, although a scalpel is often easier for this initial incision.

Follow-up

Two semen samples are usually taken at 16 and 18 weeks after vasectomy; 90 per cent of men will be negative for sperm at this stage. In a recent survey of 5000 men at the Elliot-Smith Clinic, 4 per cent had a positive second sample at 18 weeks, following a negative first sample. A few men (2.5 per cent) do not become negative and have a few persistent sperm in their semen (as opposed to millions of sperm when the vas has not been divided). In this situation some clinics give a ‘special clearance' when at least 7 months have elapsed since the vasectomy and two consecutive semen samples have shown sperm counts of less than 10 000/ml.

In a recent prospective study it has been found that transitory reappearance of sperm after successful vasectomy occurs in 0.6 per cent of men. This incidence is 18 times greater than the reported pregnancy rate following successful vasectomy.

Complications

Although unusual, complications do arise following vasectomy, and it is important to make this clear to those men who are requesting it. Philp reviewed 16 000 vasectomies and found that 7.7 per cent sought advice for pain, 3.6 per cent for bleeding, and a scrotal haematoma developed in 0.9 per cent; 80 per cent of men returned to work in 3 days and 96 per cent within a week. Sperm granulomas develop in 10 to 15 per cent of patients, and long-term chronic pain is seen in 0.1 per cent. This may be relieved by excision of the sperm granuloma. Earlier suggestions of a link between vasectomy and testicular or prostate cancer have not been substantiated; neither is there any proof of an association between vasectomy and coronary arterial disease.

Medicolegal aspects

All vasectomy patients and their partners should be counselled about the small possibility of late failure, and warning of failure should be recorded. It is not essential in the United Kingdom for the wife or partner to sign the consent form, although many surgeons prefer it if they do. The risk of pregnancy after successful vasectomy (when there have been two negative sperm counts) is 1 in 2–3000. Six cases have recently been reported in the literature where fatherhood after vasectomy was proved by DNA analysis but was associated with persistently negative semen analysis; in one case, eight samples were all negative.

Reversal of vasectomy

It is estimated that 1 to 3 per cent of men who have had a vasectomy will seek a reversal, most of them when entering a new relationship following marriage breakdown. Various techniques of vasovasostomy have been described: some surgeons advocate a two-layer anastomosis using an operating microscope, while others use 7-0 prolene sutures with magnifying spectacles. The published pregnancy rate following reversal varies between 40 and 80 per cent in different series.

Immunology

After experimental ligation of the vas deferens in animals, sperm can be found in the regional lymph nodes; circulating antisperm antibodies can be detected in the serum of 60 to 80 per cent of men after vasectomy, but their effects; if any, are unknown.

Conclusion

Vasectomy remains the most reliable method of birth control, but it is not free of complications. All vasectomy patients and their partners must be warned of the small possibility of late failure. Reversal of vasectomy may be possible, but success cannot be guaranteed.

Further reading

Davies AH, Sharp RJ, Cranston D, Mitchell RG. The long term outcome following ‘special clearance' after vasectomy. *British Journal of Urology* 1990; **66**: 211–12.

Hendry WF. Vasectomy and vasectomy reversal. *British Journal of Urology* 1994; **73**: 337–44.

Moller H, Knudsen LB, Lynge E. Risk of testicular cancer after vasectomy: cohort study of over 73 000 men. *British Medical Journal* 1994; **309**: 295–9.

O'Brien TS, Cranston D, Ashwin P, Turner E, MacKenzie IZ, Guillebaud J. Temporary reappearance of sperm 12 months after vasectomy clearance. *British Journal of Urology* 1995; **76**: 371–2.

Philp T, Guillebaud J, Budd D. Complications of vasectomy: review of 16 000 patients. *British Journal of Urology* 1984; **56**: 745–8.

Smith JC, Cranston D, O'Brien T, Guillebaud J, Hindmarsh J, Turner AG. Fatherhood without apparent spermatozoa after vasectomy *Lancet* 1994; **344**: 30.

40.1 Development of surgery of the heart and great vessels

Stephen Westaby

Further reading

Early students of the circulation were preoccupied with structure rather than function. The field attained the status of an art in the collaboration of the anatomist, Andreas Vesalius (1540–1564) and the artist Jan Van Calcar, a student of Titian. Such studies did, however, stimulate studies of function. Leonardo De inci (1452–1519) wrote that the contraction of the heart lasted about half as long as the resting period, and he built glass models of the heart valves to illustrate their function (Fig. 1). In De Vinci's life-time Jean Fernel wrote the first book devoted solely to bodily function, and originated the disciplinary name, ‘physiology'. Harvey estimated the stroke volume of the ventricles by measuring their capacity and showed that the heart must pump out in a few minutes ‘a volume of blood greater than that contained in the whole body'.

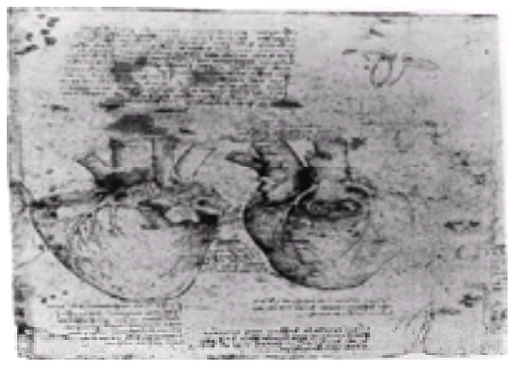


Fig. 1. De Vinci's illustrations of the heart.

Harvey described the circulation of blood in 1628, but it was more than 250 years later before meaningful progress occurred in the surgical treatment of cardiovascular disease. The concept of circulatory support began in 1812 when Le Gallois wrote ‘If one could substitute for the heart a kind of injection of arterial blood, either naturally or artificially made, one would succeed easily in maintaining alive indefinitely any part of the body whatsoever'. In the 1850s Brown Sequard performed morbid but enlightening experiments to this end: ‘after decapitation of a dog, cannulae were inserted into common carotid and vertebral arteries on both sides and oxygenated blood injected with a syringe. During 15 minutes that the perfusion lasted there were movement of the eyes and jaws'.

In practical terms, as in most fields of surgery, cardiac surgery began with the treatment of trauma. Successful suture of experimental cardiac wounds was achieved by Bloch in the 1880s. The first clinical success is attributed to Rehn in 1897. In 1902 Hill, in the United States, performed successful cardiac suture on the kitchen table by the light of oil lamps. At the time Billroth made the much quoted comment that ‘any surgeon who would attempt an operation on the heart should lose the respect of his colleagues'. This concept was upheld by the profession for many years until it gradually lost credibility in the decades after the end of the Second World War.

In 1899, Provost and Batelli devised an effective method for defibrillating the heart of an animal by opening the chest and applying directly to it 240 V of alternating current. It was 1947 before this was first applied successfully to a patient. There are few reports of cardiovascular operations before 1900. In 1888, Matas opened an aneurysmal sac and performed intravascular suture of the arterial openings. Weil and Delorme advocated removal of scar tissue surrounding the heart for treatment of constrictive pericarditis in the 1890s, though this relatively simple procedure was not undertaken clinically until 1920.

The practical development of cardiovascular surgery has, therefore, occurred almost entirely within the twentieth century. At a time when the investigation of cardiac disease relied simply on physical examination and ausculation, rheumatic valve disease was the principle interest. Brunton (1902) suggested that mitral stenosis should be amenable to surgical therapy; shortly afterwards Munro urged ligation of the patent ductus arteriosus. In 1912, Tuffier described a successful transaortic aortic valvotomy and in 1923 Cutler and Levine reported the first transventricular aortic valvotomy. Both operations were extremely adventurous for the time and received more criticism than acclaim (Fig. 2). In 1925, Sir Henry Souttar described a method for relieving rheumatic mitral stenosis by introducing a finger through the left atrial appendage (Fig. 3). The patient, a 19-year-old girl, survived with good symptomatic relief. At the time Souttar wrote ‘The problem is to a large extent mechanical and as such should already be within the scope of surgery, were it not for the extraordinary nature of the conditions under which the problem must be attacked. In view of the extreme danger to the brain from even the shortest check to its blood supply, any manipulations which are carried out must therefore be executed in the full flow of the bloodstream and they must not perceptively interfere with the contractions of the heart.' Although Souttar and others clearly foresaw the possibility of surgical solutions for cardiac disease, heart surgery became a moral issue, in much the same way as did cardiac transplantation some 50 years later. Souttar readdressed the need for artificial circulatory support, but cardiac surgery was to remain limited by these constraints for a further 30 years. Perhaps this was just as well for there were no antibiotics, no blood transfusion service, and anaesthesia for open chest surgery had not been developed.

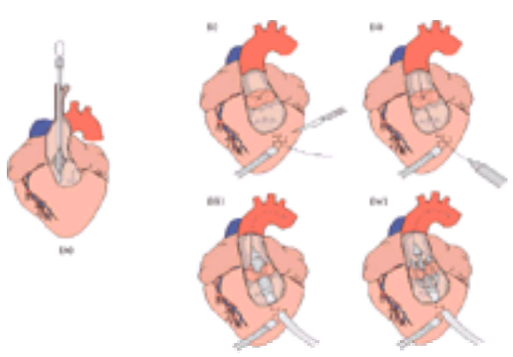


Fig. 2. Methods for transaortic (a) and transventricular (b) aortic valvotomy.

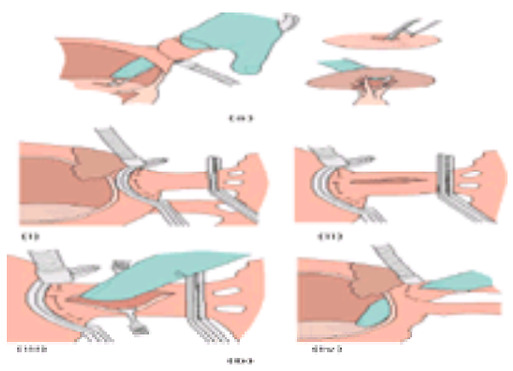


Fig. 3. Mitral valvotomy via (a) the left atrial appendage (b) left superior pulmonary vein.

In 1945, Strieder closed a patent ductus arteriosus, but the patient died of infection. Fifteen months later Gross performed the first successful ligation of a patent ductus and subsequently described the procedure for division and suture of the dividend ends.

The demonstration that some types of heart disease could be treated surgically provided an added incentive for accurate diagnosis. An endoscopic approach to internal cardiac structures was first attempted by Jacobaeus and Liljestrand in 1914 (Fig. 4). In 1929, a technique for right heart catheterization was described by Forssman, who practised transvenous insertion of a ureteral catheter in cadavers and then performed the procedure on himself when his colleagues refused to assist him. Little attention was given to this method until 1941 when Cournand, Richards, and others began to use it more extensively. Subsequently, they and Forssman received a Nobel Prize for their work.

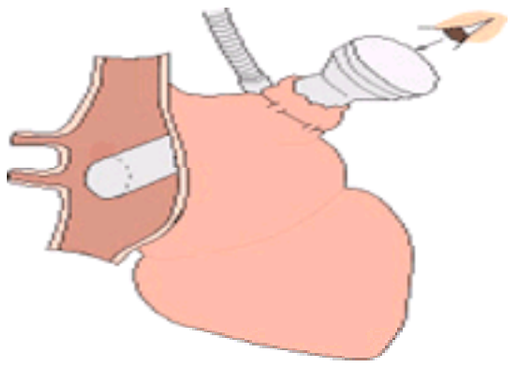


Fig. 4. Bolton examining cardioscope. This followed earlier attempts by Jacobeaus and Liljestrand.

The Second World War provided an additional stimulus to progress in cardiac surgery. In 1944, Harken was in charge of the first American Surgical Chest Center located with the 160th General Hospital of the United States Army in Cirencester. He became fascinated by the problem of cardiac wounds and foreign bodies in the heart. In all, Harken carried out 134 operations for the removal of cardiovascular foreign bodies (Fig. 5), 78 of which were within or related to the great vessels and 56 within or closely applied to the heart. Not one of his patients died. He subsequently returned to Boston to develop surgery of the mitral and aortic valves. According to Harken the credit for suggesting surgery for congenital pulmonary stenosis should go to O'Shaunessy. A description of the technique appeared in the text of a Hunterian lecture he prepared before his death in 1940. The unpublished text contains a drawing of the stenotic valve and another depicting its incision with a valvulotome inserted through the right ventricle.

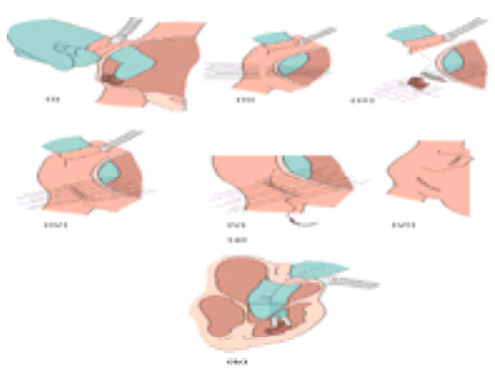


Fig. 5. Harker's method for extraction of an (a) intra-atrial and (b) intraventricular foreign body.

Surgery for coronary artery disease originated in 1943 when Beck attempted to create a new cardiac blood supply by removal of part of the epicardium and onlay of pectoral muscle against the surface of the heart. Nineteen years later, Vineberg implanted the internal mammary artery into the wall of the left venticle as a treatment for angina. The first experimental procedure for coarctation of the aorta was described by Park and Blalock in 1944, using subclavian artery to bypass obstruction in dogs. Claggett of the Mayo Clinic used this method clinically (Fig. 6) and shortly afterwards Crafoord described resection of coarctation with end-to-end anastomosis. This procedure required proximal and distal cross-clamping of the aorta, a procedure that was only conceived after Crafoord had torn a patent ductus during ligation and was forced to clamp the aorta for 28 min while repair was undertaken. When no neurological damage resulted, Crafoord reasoned that one should be able to occlude the aorta of a patient with coarctation for sufficient time to allow resection and reanastomosis.

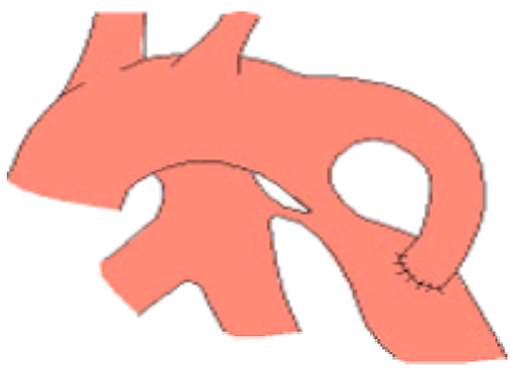


Fig. 6. Claggett's method for subclavian bypass of coarctation of the aorta.

After Blalock's experiments with the subclavian artery, Taussig suggested that this might be used to divert blood to the lungs of patients with congenital pulmonary stenosis, a concept which would revolutionize the treatment of cyanotic congenital heart disease (Fig. 7). At the time, there were grave doubts as to whether a severely cyanosed patient with heart disease could withstand anaesthesia, let alone the temporary occlusion of a pulmonary artery. Nevertheless, the first operation in 1944 was a success and in 1946 Potts described a procedure in which a side-to-side anastomosis between the aorta and pulmonary artery could be performed to the same effect, with the use of partial occluding clamps. Even more ingenious was the creation, by Blalock and Hanlon, in patients with transposition of a large interatrial defect, by closed means.

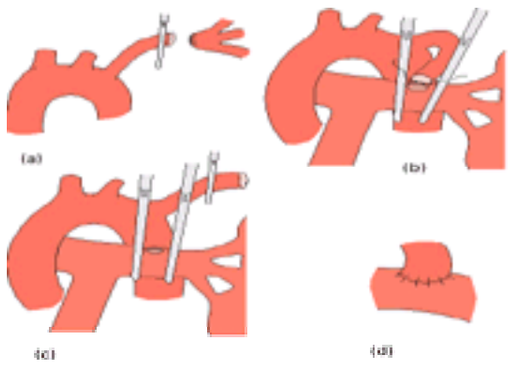


Fig. 7. Blalock Taussig systemic pulmonary arterial shunt.

The enormous clinical benefit derived from relatively simple, mechanical surgical procedures greatly increased the interest in the scope of cardiovascular surgery for both congenital and acquired defects. The prospect of performing surgical procedures on intracardiac structures fascinated many investigators who sought more and more ingenious methods for achieving their goals. Notable amongst these were Bailey, Harken, Brock, Smithy (who himself died from aortic stenosis), and Tubbs, who in addition to performing the first successful closure of a patent ductus with endocarditis, developed mitral commisurotomy. Mitral valve surgery produced many problems including torn atria, atrial thrombus and systemic embolism, dysrhythmias, and serious mitral regurgitation. Bailey's first successful mitral commisurotomy was performed in June of 1948, after four previous failures. He had already been refused operating time in three of the five hospitals in Philadelphia in which he had privileges. Further failures would close these hospitals to him. He therefore scheduled his fourth and fifth operations on the same day in two hospitals. The fourth patient died, but Bailey had a taxi waiting and completed a successful operation on the fifth patient before the hospital could cancel. A week later this patient was flown to Chicago and was presented triumphantly at a cardiological meeting there.

The simple concept of relief of obstruction in cardiovascular surgery continued in 1948 when, within a few weeks, Holmes Sellors then Brock slit the stenotic valve in patients with valvular pulmonary stenosis using a transventricular approach (Fig. 8). In 1950, Brock reported the use of blind right ventricular infundibular resection for the treatment of tetralogy of Fallot using a punch type instrument. Subsequently, a variety of blind procedures were used in an attempt to close atrial and ventricular septal defects (Fig. 9). These included the blind suture method of Gordon Murray in 1948, the circumclusion operation of Sondergaard in 1950, atrioseptopexy by Bailey in 1952, and the atrial well technique of Robert Gross in 1953 (Fig. 10). From 1952, a number of methods for catheterization of the left side of the heart were described, including the transbronchial, transthoracic, and transeptal approaches, and it became increasingly apparent that satisfactory reproducible and safe repair of more complex abnormalities required direct vision of intracardiac structures. The aortic valve in particular was never satisfactorily dealt with by closed techniques. Blind approaches either from the left ventricle with an expanding dilator or from the aorta with an operating tunnel stitched to an aortotomy incision were occasionally successful but there were many spectacular failures. Surgery of the aortic valve therefore had to await the advent of open heart surgery and a reliable valve prosthesis.

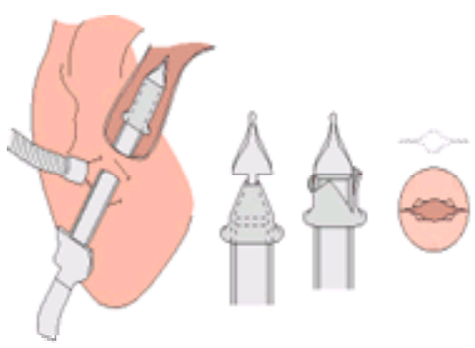


Fig. 8. Transventricular approach to the stenotic pulmonary valve. The Nichols double edged valvutome is illustrated.

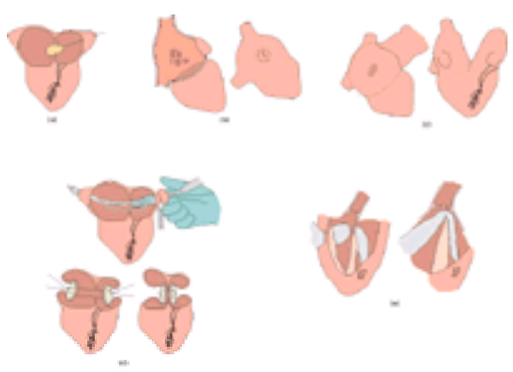


Fig. 9. Early ‘blind' methods for closing atrial septal defects (a–d). (e) ‘Pericardial tube' for closure of ventricular septal defect.

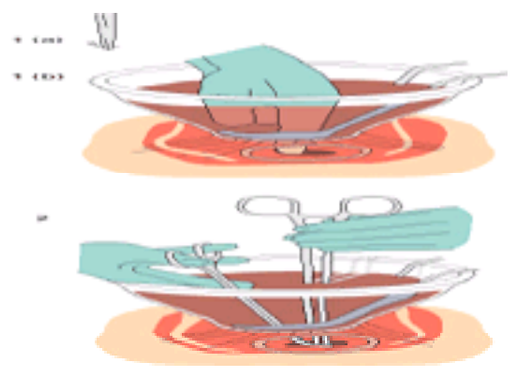


Fig. 10. The Gross atrial ‘well' technique for repair of atrial septal defects by direct suturing or by application of a patch of polythene sheeting.

Initially, there were three separate approaches to the goal of intracardiac surgery on a heart which was not required to support the circulation. The principles of hypothermia and temporary circulatory arrest were established by Bigelow, whose experimental findings were reported in 1950. Moderate systemic hypothermia (32°C) achieved by topical cooling in association with venous inflow occlusion by temporary ligation of the superior and inferior vena cava gave the surgeon 8 to 10 min to close an atrial or ventricular septal defect before irreversible cerebral damage ensued. The first clinical success using this method was reported by Lewis in 1952, using cold rubber blankets, and was later repeated by Swan in 1953, using an iced water bath technique which was eventually adopted by others.

Swan's hypothermic bath technique was cumbersome, messy, and required a prolonged rewarming process. Brock therefore developed a technique for venovenous cooling with a cooling coil and then returned blood to the heart. The procedure was quicker and more reliable, but the patient had to be heparinized. It was never widely

used.

The first successful open resection of infundibular stenosis and closure of ventricular septal defect was performed by Scott using hypothermia. Drew and Anderson subsequently used profound hypothermia with longer periods of arrest for intracardiac surgery in adults.

The second method, that of donor cross-circulation was originally considered by Blum and Megibow in 1950. Kerr (1952) and Southworth and Pierce (1952) used cross-circulation for partial or complete cardiac bypass in experimental animals, but experienced many technical difficulties by failing to use a pump to control the interchange of blood between donor and recipient. Andreasen and Watson (1952) presented a detailed physiological analysis of their experiments with cross-circulation and emphasized the value of pump control and the reduced flow concept based on their earlier azygos flow studies. Lillehei and Varco (1955) successfully applied donor cross-circulation to the clinical correction of ventricular septal defects, initially reporting results in eight patients with no mortality or significant morbidity. Lillehei's spectacular success obtained by linking the femoral vessels of an adult donor to the vessels of a child whose heart would be taken out of circuit for an intracardiac repair is one of the most exciting landmarks in the history of surgery (Fig. 11). Usually, the adult was a parent of the child. For the first operation the donor was an unrelated volunteer taxi driver. With further application of this technique it became apparent that the risks to the donor were unacceptable. An otherwise effective method of circulatory support was therefore abandoned in favour of cardiopulmonary bypass.



Fig. 11. Aerial view of the operating room during Lillehei's first cross circulation operation.

In 1935, Gibbon developed an apparatus which supported the circulation of a cat for 35 min while its pulmonary artery was occluded with a clamp. Over the next decade, the inability to oxygenate the blood efficiently hampered further developments until 1950, when it was recognized that turbulence in the oxygenator circuit could produce much higher rates of oxygenation. This discovery, known as the boundary phenomenon, increased the contact between red cells and oxygen. At the same time Tom Watson, Chairman of the board of IBM, provided financial and engineering support to build a heart–lung machine on a more sophisticated scale. Gibbon then began extracorporeal circulatory experiments on large dogs and performed successful intracardiac operations. Over a 3-year period mortality in animals fell from 80 per cent to 12 per cent, and by 1953 he was prepared to take on clinical cases.

The occasion presented itself when he repaired atrial septal defect with a large left to right shunt in an 18-year-girl. He used a pump oxygenator system: total bypass time was 45 min and for 26 min the circulation was taken over completely by the machine. The extracorporeal circuit consisted of a roller pump and an oxygenator made from stainless steel vertical screens over which the blood flowed producing turbulence with good oxygen exchange. The screens were enclosed in a plastic shell with oxygen bubbled directly into the closed environment. Coronary suction was used to recover the coronary sinus blood and a left ventricular vent was inserted to eliminate the risks of air embolus, since the heart continued to beat. The girl made an uneventful recovery. This achievement was the culmination of 22 years of investigation and generated an explosium of technological progress in cardiac surgery. Interestingly, at the time, it was not recognized as an important breakthrough, largely because the techniques of hypothermia and cross-circulation were more readily available to clinical surgery. Gibbons' following three patients died and he himself became disillusioned. It was a further 8 years before Sharp performed the first open pulmonary embolectomy, thus fulfilling Gibbons' original notion. Gibbons' stationary vertical screen oxygenator was adopted by Kirklin and later by Holmes Sellors. In Stockholm, Bjork devised an apparatus consisting of stainless steel discs rotating in a bath of blood. Melrose, improved the efficiency of Bjork's machine by rotating an inclined drum through which blood flowed slowly by gravity across series of shallow trough-like discs. This machine was used for supportive perfusion in sick patients with aortic stenosis: aortic valvotomy was still performed as a closed procedure but it was hoped that extracorporeal circulation would increase the likelihood of survival. In December 1953, a woman with severe aortic and mitral valve disease survived a double closed valvotomy performed using the Melrose machine. Six further patients were operated on, but only the first survived and clinical work was halted. In April 1957, the Hammersmith Hospital team operated on a 30-year-old woman with a secundum atrial septal defect and pulmonary hypertension. The postoperative course was stormy but the patient survived and is still alive and well. However, because atrial septal defects could be closed with hypothermia alone there was continued scepticism regarding the heart–lung machine. In an attempt to dispel this, the paediatric cardiologist, Bonham Carter referred 50 children with ventricular septal defects to Hammersmith Hospital, with no questions asked until all 50 operations had been completed. The operations were a resounding success, with an overall mortality rate of 20 per cent. This fell to only 4 per cent in those with uncomplicated defect without pulmonary hypertension or anatomical complications. These early operations used potassium citrate to arrest the heart. The Melrose technique was to inject 2 ml of a 25 per cent solution of potassium citrate in 50 ml of autologous heparinized blood into the aortic root (with the ascending aorta clamped). This produced ideal operating conditions on a flaccid bloodless heart, with rapid return to normal beating on release of the aortic cross-clamp. The technique, a forerunner of cold potassium cardioplegia, was used to treat over 100 patients without any apparent ill-effects. However, reports from North America suggested that the potassium solution caused myocardial damage and it was reluctantly abandoned in favour of electrical fibrillation with hypothermia, or coronary artery cannulation and perfusion.

Cardiopulmonary bypass using a pump oxygenator system deve-loped rapidly and increased greatly the scope of cardiac surgery. Nevertheless, side-effects, particularly severe pulmonary dysfunction, bleeding tendencies, and renal failure, occurred frequently and were attributed to damage to blood components caused by the pump oxygenator system. The so-called ‘postperfusion syndrome' was attributed to the adverse effects from contact of blood with foreign surfaces. Drew addressed this problem by keeping the patient's lungs as the oxygenator and using a two-pump system to replace both right and left sides of the heart. To this he added a cooling coil and explored the use of profound hypothermia with body temperatures as low as 8 to 10°C. The first successful application of this technique was in January 1959 for closure of atrial and ventricular septal defects in a 4-year-old child. Apart from Ross, who has recently used the two-pump system for coronary artery surgery, others did not adopt this technique. Deep hypothermia is nevertheless used routinely with conventional cardiopulmonary bypass, and total circulatory arrest for surgery of congenital heart defects in infants and young children, and for extensive surgery on the thoracic aorta in adults. Barrett-Boyes pioneered the use of surface cooling in infants: he employed ice bags followed by core-cooling to 10 to 20°C with the pump oxygenator.

Modifications and refinement of the heart–lung machine continued during the 1960s and 1970s. The next major step was the development of the membrane oxygenator, a device which reduced trauma to blood cells by separating blood from gas. With the elimination of time constraints, attention switched to the treatment of more complex congenital and acquired heart defects. Closed mitral valvotomy was well established for mitral stenosis, but there were no satisfactory procedures for aortic valve disease or mitral regurgitation. Various reparative procedures were attempted (annuloplasty by Wooller, pericardial inserts by Ross and Cleland) but there was clearly a demand for a valve prosthesis. Hufnagel designed and used a caged ball valve prosthesis for insertion in the descending thoracic aorta of patients with severe aortic regurgitation (Fig. 12). Further developments along similar lines by Starr, Harken, and McGovern led to the development of a satisfactory caged ball prosthesis for insertion in the subcoronary aortic position (Fig. 13). Initial attempts to use the prosthesis in the mitral area failed because of protrusion of the cage into the outflow tract of the left ventricle. Despite this, in 1962 Starr and Edwards reported a series of 16 patients who had undergone mitral valve replacement with a ball valve prosthesis: 10 survived. In 1964, Melrose reported both experimental and clinical use of a tilting disc valve made of polypropylene. The valves were manufactured and used clinically by Wooller and Logan. This prosthesis was haemodynamically sound and initially highly successful. With time more and more patients returned with extensive thrombus formation around the prosthesis, and its use was abandoned. Soon afterwards, Bjork designed a similar tilting disc valve which was manufactured by the Shiley Company and used clinically in competition with the Starr valve from Edwards Laboratories. Some surgeons, including Brock, had serious reservations about implanting foreign material within the heart and great vessels. Brock was convinced that natural tissues from cadavers should be used and began by using an aortic tube homograft for repair of coarctation. Ross carried out the first aortic valve replacement using an aortic homograft as an emergency in London in 1962, the valve homograft having been prepared by Gunning in Oxford. Both he and Barrett-Boyes worked enthusiastically with homograft replacement, refining the methods of collection, preparation, and preservation. Later Ross varied the technique by transferring the native pulmonary valve to the aortic area and replacing the patient's pulmonary valve with an aortic valve homograft. Senning replaced the aortic valve with autogenous facia lata, initially with excellent early results. The procedure was later abandoned when valve failure followed shrinkage and calcification of the native material. Nevertheless, many surgeons found homograft valve replacement tedious and difficult to carry out and in order to overcome these drawbacks attempts were made to mount pericardium, dura mater, or homograft valves on a metal frame to which a sewing ring was attached. A little later a pig valve heterograft, again prepared by Gunning in Oxford, was implanted in a patient by Binet in France, and soon after Allison and Gunning in Oxford also used these pig valves. When Carpentier showed that glutaraldehyde preparation removed most of the antigenic properties of pig or calf valves, heterograft valves were stent mounted and have subsequently proven very successful.

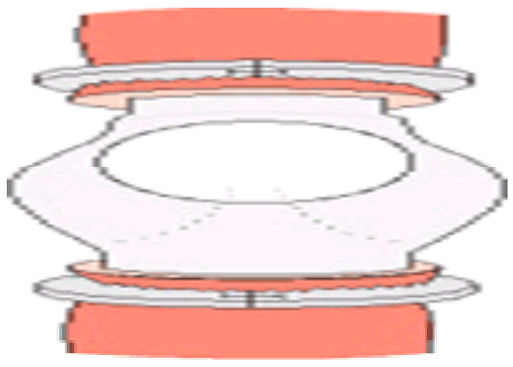


Fig. 12. The Hufnagel prosthetic ball valve sewn into the descending aorta.

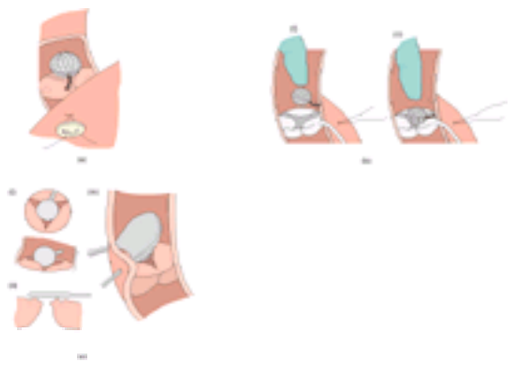


Fig. 13. Early attempts at an aortic valve prosthesis. (a) Fixation of tail of ball by passage through eyes of plastic button and tying. (b) (i) Prosthetic valve during systole. (ii) Prosthetic valve during diastole. (c) Disadvantages of some of these valves. (i) Ball of fixed spherical shape does not tamponade a triangular opening well. (ii) Flat disc does not tamponade at opening in an irregular surface well. (iii) Two-tailed disc may be held above valve orifice in such a way as to be incapable of function.

Revascularization of the ischaemic myocardium remained a considerable challenge. As early as 1937, O'Shaunnessy had suggested the technique of omentopexy, whereby omentum was transferred in continuity from the abdomen to the surface of the heart to encourage collateral circulation. Pneumopexy, partial coronary sinus ligation, and pericardial abrasion were all tried, with dubious outcome. Vineberg's implantation of the internal mammary artery into myocardium provided a more promising line of attack. He showed that an open-ended artery buried in the myocardial sinusoids could establish vascular connections and supply additional blood to an ischaemic left ventricle. Effler and Favaloro adopted this technique enthusiastically and were well placed when Mason Soanes developed the technique of coronary arteriography. This revolutionized surgery for coronary artery disease by demonstrating the site of coronary occlusion and providing the concept of coronary bypass. Favaloro used autogenous saphenous vein, reporting the technique in 1968. Although coronary artery bypass surgery would eventually account for more than 80 per cent of cardiac operations, the procedure was initially regarded with scepticism and relatively few operations were carried out for the next 5 years. Although it was clear that successful coronary artery bypass relieved angina, there were doubts about the ability to prolong life and its superiority over increasingly effective medical treatment of angina.

From 1973 to 1978, several British centres took part in a European multicentre, randomized trial to compare long-term results of coronary bypass with those of medical management. The results together with those of American trials, and careful analysis of subsets with different degrees of coronary disease, established the superiority of surgery, particularly in patients with two and three vessel coronary disease, left main stem obstruction, and high-grade proximal left anterior descending obstruction. Impaired left ventricular function from previous myocardial infarction reinforces the decision for surgery and the superiority of surgical management.

In December 1967, Barnard astonished the world by performing the first clinical human cardiac transplant, based on the pioneering research work of Shumway at Stanford. The patient died from acute rejection 18 days later, but a second patient lived for 18 months and produced a wave of enthusiasm throughout the surgical world. In Britain, Ross performed three transplants but the longest survivor lived for only 43 days and publicity surrounding excision of the beating donor heart brought the programme to an end. This led to a clear definition of brain death designed to convince the public and medical profession that an individual would be accepted as a donor only when independent life would never be possible. Fortunately, Shumway continued with his transplant programme when most others had given up in disillusionment. His remarkable achievements led English and Yacoub separately to resume the British programme in 1979.

Further reading

Barnard CN. A human cardiac transplant: an interim report of a successful operation performed at Groote-Schuur Hospital, Capetown. *South African Medical Journal* 1967; **41**: 1271–4.

Blalock A. Cardiovascular surgery, past and present. *Journal of Thoracic Cardiovascular Surgery* 1966; **51**: 153–64.

Blalock RC, Cambell M. Discussion on surgery of the heart and great vessels. *Proceedings of the Royal Society of Medicine* 1951; **44**: 995–1005.

Cleland WP. The evolution of cardiac surgery in the United Kingdom. *Thorax* 1983; **38**: 887–96.

Drew CE, Anderson IM. Profound hypothermia in cardiac surgery. *Lancet* 1952; **i**: 748–50.

Gibbon JH. Application of a mechanical heart and lung apparatus to cardiac surgery. *Minnesota Medicine* 1954; **37**: 171–80.

Gross RE. Surgical correction for coarctation of the aorta. *Surgery* 1945; **18**: 673–78.

Harken DE. Foreign bodies in and in relation to the thoracic blood vessels and heart. I. Techniques for approaching and removing foreign bodies from chambers of the heart. *Surgery, Gynecology and Obstetrics* 1946; **83**: 117–25.

Kirklin JW. Open heart surgery at the Mayo Clinic: the 25th anniversary. *Mayo Clinic Proceedings* 1980; **55**: 339–41.

Kirklin JW, Pacifico AD. Surgery for acquired valvular heart disease. *New England Journal of Medicine* 1973; **288**: 133–8.

Lillehei CW, Cohen M, Warden HE, Zeigler NR, Varco RL. The results of direct vision closure of ventricular septal defects in eight patients by means of controlled cross circulation. *Surgery, Gynecology and Obstetrics* 1955; **10**: 447–66.

Souttar HS. The surgical treatment of mitral stenosis. *British Medical Journal* 1925; **2**: 603–6.

Starr A, Edwards ML. Mitral replacement: clinical experience with a ball valve prosthesis. *Annals of Surgery* 1961; 154–726.

Vineberg A. Treatment of coronary artery insufficiency by implantation of the internal mammary artery into the left ventricular myocardium. *Journal of Thoracic Surgery* 1952; **23**: 42–54.

40.2 Cardiopulmonary bypass and myocardial protection

Stephen Westaby

Cardiopulmonary bypass
The bypass circuit
Conduct of cardiopulmonary bypass
Optimal variables for cardiopulmonary bypass
Profound hypothermia and total circulatory arrest
Interactions between blood and foreign surfaces
The damaging effects of cardiopulmonary bypass
Protection of the heart during open heart surgery
Development of techniques for myocardial protection
Cardioplegic solution
Blood cardioplegia
Cardioplegia in clinical practice
Further reading

Special techniques are required for circulatory support and protection of the myocardium and vital organs during detailed repair of the non-beating, empty heart. During 35 years of open-heart surgery with cardiopulmonary bypass, methods have evolved that have reduced the mortality and morbidity of certain straightforward procedures almost to zero. Widespread application of modern cardiovascular technology now allows safer, more routine, surgical correction of cardiac defects at the extremes of age, from the preterm neonate with complex congenital heart disease to the late octogenarian with calcific aortic stenosis. Even hearts with extensive myocardial infarction and rupture can be repaired successfully without further damage using cold cardioplegic arrest. Nevertheless, exposure of blood to foreign surfaces still has widespread damaging effects (the whole-body inflammatory response) and myocardial protection could be further improved. Short cardiopulmonary bypass and aortic cross-clamp times greatly reduce the potential for morbidity and mortality associated with cardiac surgery. There is scope for future research and development, and no room for complacency.

Cardiopulmonary bypass

Cardiopulmonary bypass is a method of whole-body perfusion in which the pumping action of the heart and the oxygenation of blood by the lungs are replaced by an extracorporeal circuit. Elimination of blood flow through the heart and lungs provides a controlled environment for surgery of the heart and great vessels, and less frequently of the major airways. The technique enables hundreds of thousands of cardiac operations to be performed annually, with a mortality in some categories of patients of less than 1 per cent.

The bypass circuit

The main components of the bypass machine are an oxygenator and an arterial pump, which substitute for the lungs and heart, respectively, for the duration of the surgical procedure. Other components include provision for blood defoaming, filtration, temperature regulation, monitoring and safety devices. A simple circuit is shown in Fig. 1.

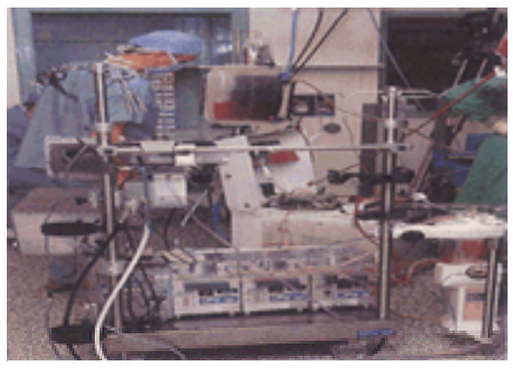


Fig. 1. Cardiopulmonary bypass circuit.

Venous blood is diverted from the heart using a large-bore cannula inserted in the right atrial appendage (Fig. 2). If the right side of the heart has to be opened, separate cannulas are inserted into the superior and inferior venae cavae, usually introduced through the right atrium. Purse-string sutures are snared around the incisions to produce a blood- and airtight seal. The venous cannula is connected to a reservoir sited below the operating table so that blood drains by siphon. Once established, this acts as a low-pressure suction system controlled by the difference in height between the operating table and the venous reservoir (Fig. 3).

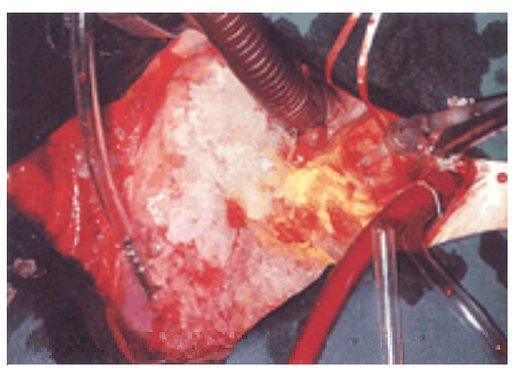


Fig. 2. Arterial and venous cannulas in the aorta and right atrial appendage.

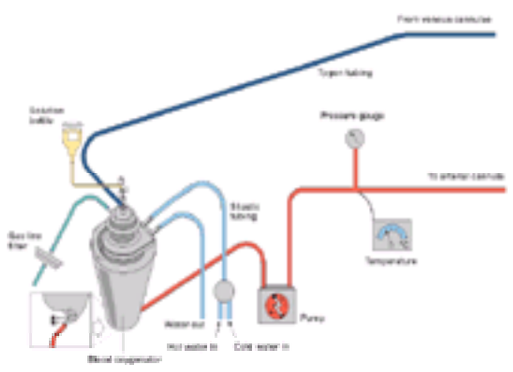


Fig. 3. Cardiopulmonary bypass circuit: venous drainage occurs by siphon effect due to the difference in height between the operating table and venous reservoir.

The oxygenator

There are two types of commercial disposable oxygenator, both of which contain an integral heat exchanger to regulate the blood outflow temperature.

Bubble oxygenator

These have been used since the early days of cardiopulmonary bypass. A rapid jet of oxygen is forced through a volume of blood to cause the bubbles, which produce a large blood–gas interface across which gas transfer occurs by partial-pressure gradients. The smaller the bubbles the more efficient the oxygenation because of the relative surface area:volume ratio. However, smaller bubbles are more difficult to remove effectively before the blood is returned to the patient. Carbon dioxide also rapidly enters the gas phase and larger bubbles are more effective for its removal. In practice, the oxygen flow in bubble oxygenators creates enough bubbles in a range of sizes required to fulfil both functions. Clearance of bubbles before delivery of the arterialized blood to the circulation is achieved by exposure to surface-action silicone rubber, filtration, and settling (Fig. 4). High gas flows in the oxygenator increase both the emission and dispersion time of microbubbles, and high rates of blood flow decrease the time available for settling. Both are undesirable.

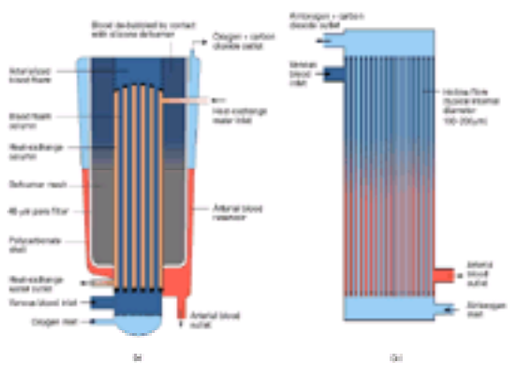


Fig. 4. Bubble (a) and membrane (b) mechanical oxygenators.

Although bubble oxygenators are relatively simple in construction, cheap, and easy to incorporate within the bypass circuit, they have inherent disadvantages. Inefficient removal of bubbles may lead to widespread microembolization. A direct blood–gas interface damages the cellular elements of blood, particularly platelets, and may produce fibrinous microemboli. Blood gas control is imprecise. Blood concentrations of carbon dioxide can be maintained by varying the flow rate of oxygen, but this is at the expense of precise control of arterial oxygen tension. From a safety standpoint there is an inherent capacity for air or oxygen to be drawn through the oxygenator should the venous reservoir empty; this can result in massive arterial gas embolism. Equally, if the gas supply is interrupted, blood entering the bubble generator could induce catastrophic failure. Nevertheless, modern bubble oxygenators are reliable and are currently used in at least two-thirds of bypass operations worldwide.

Membrane oxygenator

Blood is passed over a membrane permeable to oxygen and carbon dioxide but which separates the two over a wide area. Venous blood drawn from a reservoir after filtration and de-airing is pumped through the oxygenator, where any remaining bubbles tend to pass through the inert microporous membrane into the gas phase. Gaseous nitrogen is highly insoluble and can be confined to the gas phase, so an oxygen/air mixture is used to modulate the arterial oxygen tension. Carbon dioxide tension is regulated by varying the gas flow or by altering the carbon dioxide tension in the gas mixture. Blood gas and hydrogen ion concentrations can therefore be maintained with considerable accuracy. Far fewer microbubbles enter the circulation and massive gas embolism is almost impossible. Should venous obstruction lead to an empty reservoir, air pumped into the oxygenator passes across the membrane into the gas phase, rendering passage to the patient unlikely.

The membrane itself may take the form of flat sheets or hollow fibres. Flat-sheet membranes are usually arranged as a ‘fan fold' with the blood on one side and the gas on the other, providing an exchange surface area of 2 to 3 m^2. Baffles placed between the layers of the stack allow free flow of fluids, though the internal resistance of this type of oxygenator is inherently high. The oxygenator is therefore placed between the pump head and the patient. Hollow-fibre oxygenators are compact devices containing membrane capillaries with an internal diameter of between 100 and 200 μm. Gas flows through the fibres, which are arranged in bundles. In both flat-plate and hollow-fibre oxygenators the blood flows in eddies, which bring more red cells to the gas-exchange surface during transit. The intricacy of the blood path increases the time taken in setting up the apparatus because any air contained within the system must be purged with soluble carbon dioxide to prevent the formation of bubbles. Although the membrane presents a large surface of foreign material to the blood, the propensity for cellular activation or damage is less than that associated with a gas–liquid interface. The relative merits of bubble and membrane oxygenators are summarized in Table 1. Membrane oxygenators have distinct advantages and are now more competitively priced.

	Bubble oxygenator	Membrane oxygenator
Advantages	Relatively cheap	Separate blood and gas phases
	Relatively simple in construction	Creation of far fewer bubbles
	Easy to set up quickly	Greater accuracy in blood gas control
		Massive gas embolism is almost impossible
Disadvantages	Introduction of arterial bubbles,	More expensive
	which can form microemboli	More difficult to set up
	Large blood/gas interface with potential for cell damage	
	Imprecise blood gas control	
	Potential for massive arterial gas embolism	

Table 1 Comparison of bubble and membrane oxygenators

The arterial pump

This is a simple, mechanical, double-arm roller pump with speed control that allows the production of variable flow rates. For any given setting, the rollers rotate at a constant rate to provide an output that can equal or exceed the normal range for cardiac output. The rollers act on a smooth, flexible, and strong tube of silicone rubber, which is less prone to spalliation (the splitting off of small particles) than is polyvinyl chloride, from which the remainder of the circuit tubing is made. The rollers are non-occlusive in order to avoid direct damage to the cellular elements of blood. The fundamental difference in mechanical pumping as opposed to physiological cardiac pumping is the absence of a pulsatile wave form, though pulsatile systems are now available.

Besides blood drained directly from the right atrium, the heart–lung machine has two further pump heads to return blood from the pericardium or that vented from the left side of the heart in order to prevent cardiac distension or maintain a dry field. Because blood scavenged from these sites is mixed with air, fat globules, and debris from the operating site, it is returned to the cardiotomy reservoir for defoaming and filtration before re-entering the oxygenator. The cardiotomy reservoir may be integral with the oxygenator or separate. Oxygenated, pumped blood is returned to the body on the arterial side of the circulation by a cannula situated in the ascending aorta or femoral artery (Fig. 2). Most surgeons employ a 40-μm pore diameter filter in the arterial line to remove bubbles or particles that evade upstream filtration,

particularly when a bubble oxygenator is used (Fig. 5). The need for a filter on the arterial line is less certain in membrane oxygenator circuits, since the filter itself may damage blood or generate gaseous or fibrinous emboli.

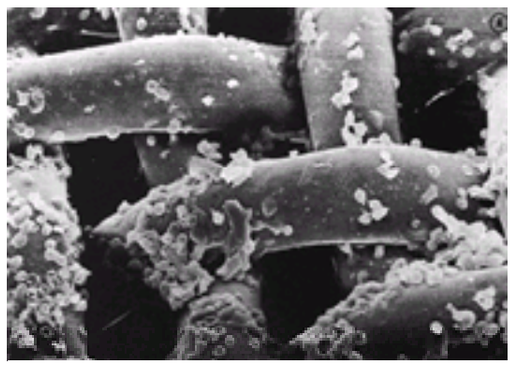


Fig. 5. Arterial-line filter with deposited elements of blood.

The priming fluid

The extracorporeal circuit must be primed with fluid to exclude air before use. In the early days the machine was filled with whole blood, but large transfusions of autologous blood generated problems and were associated with considerable haemolysis within the circuit. There was a high incidence of coagulopathy and organ damage related to capillary sludging and complement activation at low flow rates. This led to the ‘clear-fluid prime'; most centres now use total haemodilution for adult surgery unless the patient had a low preoperative haematocrit. Most adult circuits have a priming volume of between 2 and 2.5 l. We have reduced this to 1.0 to 1.5 l. At the onset of perfusion there is a rapid infusion of clear fluid, producing haemodilution and a haematocrit around 25 per cent. Babies and infants have a relatively small total blood volume in comparison to the circuit volume so that blood must be added proportionally. Introduction of the fluid prime has greatly reduced the morbidity and mortality of cardiopulmonary bypass.

Before bypass, the arterial and venous pipes are connected to the oxygenator and filled with a balanced electrolyte solution such as Ringer's lactate. The priming fluid is pumped around the blind-loop circuit in order to dislodge and disperse adherent bubbles. Passage through the heat exchanger ensures that the prime is at a uniform temperature of between 35 and 37°C before bypass commences. A pre-bypass filter of 0.2-μm diameter may be used temporarily within the close circuit to remove bacteria and other particles. The tubes connecting the patient, pump and oxygenator are made of transparent polyvinyl chloride and are usually supplied as customized sets assembled under clean conditions and sterilized by g-irradiation. The circuit is assembled using ‘no touch' techniques.

For patients in advanced cardiac failure or with compromised renal or pulmonary function, maintenance of plasma colloid osmotic pressure is important. A haemofilter unit in the arterial line allows fluid to be removed during the bypass procedure; this reduces the final degree of haemodilution and eliminates extra volume in the form of cardioplegic solution. Ultrafiltrate is removed from the circulation by creating a pressure gradient across the semipermeable membrane of the haemofilter. Colloid osmotic pressure may be increased by the use of human plasma protein or cryoprecipitate-poor plasma. Artificial plasma expanders such as gelatine solution or hydroxyethyl starch can also be added to the prime, though hydroxyethyl starch may be associated with more severe postoperative bleeding due to platelet dysfunction.

Conduct of cardiopulmonary bypass

Anaesthesia is induced by a neuroleptic analgesic technique combining high-dose analgesia with a neuroleptic or an anxiolytic agent. Anaesthetic maintenance during bypass usually relies on ‘top-up' doses during rewarming. This, together with hypothermia, provides adequate but sometimes uneven anaesthesia. Continuous intravenous anaesthesia with volatile agents, or more recently with propofol, has proved satisfactory despite cardiodepressant tendencies. Nitrous oxide should never be used during the peri-bypass period since it rapidly augments the size of intravascular bubbles. Deleterious effects have been reported following use of nitrous oxide even 20 min after the end of bypass.

Heparin (300–400 U/kg) is given 5 to 10 min before arterial cannulation. The activated clotting time should be at least 400 s. Blood gas analysis should be undertaken before commencing bypass to confirm that carbon dioxide tension lies within normal limits (4.5–6.0 kPa), since hypocarbia caused by hyperventilation interferes with the autoregulation of cerebral blood flow.

The closed bypass circuit loop is then divided between clamps, removing the pre-bypass filter. The surgeon inserts and connects the arterial cannula, taking care to exclude air bubbles. The perfusionist confirms arterial pressure in the line and then venous drainage is achieved by inserting a two-stage venous cannula into the right atrial appendage or by cannulating the superior and inferior venae cavae separately. The cannulas are connected to the venous line to complete the circuit.

The surgeon begins cardiopulmonary bypass by telling the perfusionist to ‘go on'; s/he opens the arterial line and starts the roller pump slowly. When the perfusionist is confident that an adequate volume of blood can be perfused without an excessive rise in resistance or pressure, the venous line is opened, allowing blood to drain into the venous reservoir. This begins a siphon and as levels in the reservoir rise the speed of the arterial pump is increased, drawing fluid from the reservoir into the oxygenator and then the aorta. When the level in the reservoir is constant, all the blood entering the machine is being returned to the patient. Full flow has been achieved and ventilation of the lungs is suspended. The arterial-line pressure is checked, since excessive pressure may indicate dangerous malpositioning of the aortic cannula.

For most bypasses the temperature of the perfusate is reduced by the heat exchanger, which in turn causes systemic hypothermia in the patient. Moderate hypothermia (28–32°C) gives a measure of protection for the brain in case of episodes of poor perfusion or hypotension (Table 2). Cerebral oxygen consumption is less than 50 per cent of normal at this temperature and circulatory arrest may be tolerated for 5 to 8 min (long enough for a skilled team to cope with equipment failure). Full, unobstructed venous drainage empties the heart so that it does not eject. Oxygenated blood is returned from the heart–lung machine at a rate of 2.4 l/m^2 per minute. Pressure is governed by the peripheral vascular resistance, which is adjusted pharmacologically by the anaesthetist and perfusionist to obtain means of between 50 and 70 mmHg (Fig. 6). Most oxygenators are run at 1:1 gas:blood ratio and there is seldom a problem achieving full saturation. The *P*o$_2$ often reaches 300 mmHg or more. While the desired temperature is reached the ascending aorta is clamped and cardioplegic arrest is induced by infusion of 1 l of hyperkalaemic electrolyte solution at 4°C into the coronary arteries. If aortic regurgitation is present, the aorta is opened and the solution delivered directly into the coronary ostia. Cardioplegia stops the heart in diastole for 30 min or more; this period can be increased by using topical hypothermia with iced slush in the pericardium. Further doses of cardioplegic solution are administered if electrical activity returns on the electrocardiogram. This provides ideal conditions for surgery to be performed on the cold, flaccid heart that is emptied of blood and easily manipulated.

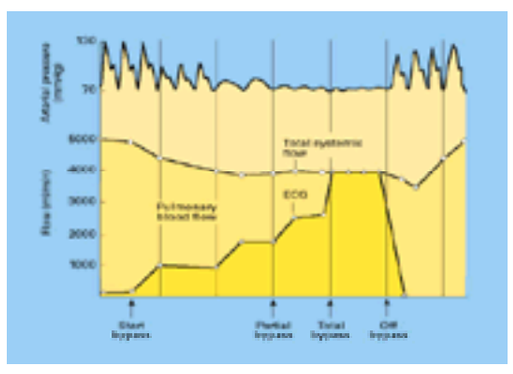


Fig. 6. Blood flow and perfusion pressure during a standard adult bypass.



Table 2 Recommended flow rates during cardiopulmonary bypass

Removal of the aortic cross-clamp reperfuses the myocardium and electrical activity resumes. Fifty per cent of patients show reversion to spontaneous rhythm; the remainder develop ventricular fibrillation requiring defibrillation. When the heart begins to beat, the aortic valve is open and blood is ejected against the resistance of the perfusion pressure. Great care must be taken to evacuate air from the cardiac chambers to prevent cerebral or coronary air embolism. This procedure is an integral and routine part of every open operation and air embolism by macrobubbles is rare.

Some 15 to 20 min before completion of the cardiac repair the surgeon instructs the perfusionist to rewarm the patient. This is achieved by passing warm water through the heat exchanger coils, though temperature gradients greater than 10°C between water and blood must be avoided. The duration of rewarming depends on the depth of cooling and the weight and body surface area of the patient. A heating blanket on the surface of the operating table assists in rewarming; the aim is to regain a temperature of 37°C by the time that the cardiac repair is complete.

After surgical correction the heart should be more efficient through relief of obstruction, elimination of intracardiac shunt, or increase in myocardial blood flow. This may be balanced by short-term injury from ischaemia, reperfusion, and surgical manipulation. Interstitial oedema reduces cardiac compliance and increased atrial filling pressures may be needed. Five minutes before weaning from cardiopulmonary bypass, ventilation is restarted. A few forceful inflations of the lungs by hand help to expel residual air from the left atrium or pulmonary veins. Cardiopulmonary bypass is discontinued by gradually reducing venous drainage so that less blood is available for the heart–lung machine and more for the heart. As venous pressure rises the heart begins to eject; the venous line is then clamped and the atrial filling pressures gradually adjusted by controlled transfusions from the reservoir. The circulating blood volume is slowly augmented until arterial pressure is satisfactory. If systemic pressure is low, coronary perfusion pressure may be inadequate. Catecholamine support is occasionally necessary in patients with myocardial dysfunction, though calcium is an excellent inotrope at this stage. Problems of rate or rhythm may be addressed by temporary epicardial pacing. Severe myocardial depression with left ventricular failure may require intra-aortic balloon counterpulsation.

Acute hypertension due to residual hypothermia and vasoconstriction is common, particularly after coronary surgery, and jeopardizes vascular anastomoses. Increased afterload also adversely affects the myocardium during the recovery period from aortic cross-clamping and cardioplegia. This should be treated pharmacologically with intravenous glyceryl trinitrate or sodium nitroprusside.

Optimal variables for cardiopulmonary bypass

Pressure and flow

Conventional perfusion techniques are associated with a reduction in the mean arterial blood pressure to around 50 to 70 mmHg; this is attributable to a reduction in blood viscosity produced by haemodilution techniques. When considering optimal pressures and flow, the brain is the critical organ. Early studies introduced the concept of an established formula for calculating adequate mean pump flow during perfusion. Flow rates of 2.0 to 2.4 l/m^2 per minute provide satisfactory maintenance of acid–base balance and oxygen consumption, and preserve acceptably normal patterns of cerebral function as indicated by the electroencephalogram. Flows greater than 2.4 l/m^2 per minute at normothermia produce no further increase in oxygen consumption, and increase blood trauma and the risk of gaseous emboli. Recommended rates of flow for certain groups are shown in Table 2.

It is usual to employ moderately hypothermic bypass (25–32°C), which reduces overall oxygen demand and permits the use of lower flows and pressures (see Table 2). With lower flow rates the non-coronary collateral flow is reduced, thereby slowing cardiac rewarming during aortic cross-clamping. A flow rate of 1.2 l/m^2 per minute maintains global oxygenation adequately at 25°C. There is a close correlation between oxygen consumption and flow rates modified by hypothermia. Blood lactate concentration is an effective indicator of microcirculatory perfusion and should be less than 5 mmol/l. Mixed venous oxygen saturation can be used as an indicator of very poor perfusion by shunting.

A fall in mean pressure below 50 mmHg is associated with adverse electroencephalographic changes. Pressor agents such as noradrenaline (norepinephrine) are used to increase the cerebral blood flow at the expense of mesenteric, renal, and extremity flow. Mean pressures above 100 mmHg should be avoided by pharmacological manipulation of peripheral vascular resistance. When reperfusing the ischaemic heart after cardioplegia or the fibrillating heart during intermittent aortic cross-clamping, a pressure of 70 to 80 mmHg is optimal, though greater pressures are necessary in patients with hypertrophied left ventricles. High systemic pressures should be avoided when discontinuing bypass since afterload reduction reduces cardiac work and oxygen consumption.

The standard roller pump produces non-pulsatile flow. Modifications of the roller-pump systems have been used to impart pulsatility with synchronization that allows a systolic pressure pulse to be delivered during the patient's diastole. Early investigations suggested that increased peripheral resistance and lactic acidosis during cardiopulmonary bypass could be diminished with pulsatile flow. Measurements of change in vascular resistance, oxygen consumption, and lactate production suggested a generally beneficial effect of pulsatility. The increased energy imparted by pulsatile flow at a given mean arterial pressure is thought to improve lymph flow and provide more effective opening of capillary beds. In particular, urine output is improved by pulsatility. However, many perfusion-related problems that prompted the development of pulsatility have disappeared with recent improvements in oxygenators, priming solutions, and the general conduct of cardiopulmonary bypass. Its use is therefore not widespread.

Acid–base balance and the regulation of cerebral blood flow

Cerebral blood flow influences cerebral function following cardiopulmonary bypass, through adequacy of oxygen delivery for cerebral metabolism and by flow-dependent delivery of microemboli to the cerebral vasculature. The influence of carbon dioxide ($P\mathrm{aco}_2$) on cerebral autoregulation during cardiopulmonary bypass is important.

The majority of cardiac surgical procedures employ hypothermic cardiopulmonary bypass at average temperatures of 26 to 32°C, and hypothermia increases the solubility of carbon dioxide in blood. Until recently the influence of acid–base management techniques and changes in $P\mathrm{aco}_2$ on cerebral blood-flow regulation and flow/metabolism coupling have been poorly understood. At 37°C the intracellular ratio of H^+/OH^- concentrations is in equilibrium with that of the blood; this is approx. 1:16, maintained by protein buffering and modulated by variation of the $P\mathrm{aco}_2$, which establishes an intracellular:extracellular gradient for hydrogen ions of about 4:1. Disturbance of the blood hydrogen:hydroxyl ion ratio may be expected to cause intracellular biochemical inefficiency or damage. During hypothermic perfusion, it has been common practice to maintain the blood gas values normally found at 37°C at all temperatures; this is known as ‘pH stat' management. However, cerebral blood flow increases during cardiopulmonary bypass with pH stat, implying impairment of autoregulation; this is now thought to be due to artificial maintenance of blood pH at the physiological level of 7.42 during hypothermia by addition of CO_2 to the oxygenator, an action that changes the acid–base status towards a state of respiratory acidosis, which is known to have a vasodilating effect on cerebral blood vessels.

Rahn and Reeves advocated the concept of so-called alpha-stat regulation, whereby preservation of a constant relative alkalinity of blood is favoured at different temperatures by allowing a decrease in $P\mathrm{aco}_2$ and a simultaneous increase in pH during the development of hypothermia. In practice, this is achieved by maintaining $P\mathrm{aco}_2$ constant by analysis of blood gases at 37°C without temperature correction throughout the procedure. Serial measurements of cerebral blood flow before, during, and after cardiopulmonary bypass with alpha-stat CO_2 regulation show that cerebral autoregulation remains intact when CO_2 is not added to the oxygenator. Cerebral hyperaemia and impaired autoregulation during hypothermic cardiopulmonary bypass can therefore be explained by the relative hypercarbic state resulting from addition of CO_2. When a non-temperature-corrected $P\mathrm{aco}_2$ of approx. 40 mmHg is maintained, cerebral blood flow is lower and cerebral autoregulation is better preserved than when the pH at once concept is used. With alpha-stat control, cerebral blood flow is independent of pressure changes and dependent upon cerebral

metabolism and oxygen consumption. This is important, since global increases in cerebral blood flow due to elevation of $Paco_2$ may occur at the expense of potentially ischaemic areas, and CO_2-induced cerebral dilation can critically reduce perfusion pressure in the circle of Willis. This may jeopardize areas of brain that are dependent on flow through stenosed vessels. In addition, cerebral hyperaemia may increase the flow-dependent delivery of microemboli.

Profound hypothermia and total circulatory arrest

The technique of deep hypothermia and circulatory arrest was developed in response to the need for safe correction of complex intracardiac congenital heart defects in small infants; this, together with cardioplegic arrest of the heart, provides an operative field that is completely bloodless and without cannulas to interfere with the intracardiac procedure. The technique is used for repair of all complex intra-atrial defects in infants weighing less than 7 kg.

Repair of other congenital defects in infants weighing less than 4 kg also uses circulatory arrest: cannulation of the inferior and superior venae cavae may be technically difficult in small infants with tetralogy of Fallot or ventricular septal defects. Although intra-atrial surgery may not be required, the heart may be small and the technical aspects of operating during continuous cardiopulmonary bypass may be formidable. Circulatory arrest is also used in adult patients undergoing extensive surgery on the ascending, transverse, and descending thoracic or thoracoabdominal aorta.

Hypothermia causes a significant reduction in oxygen consumption, which is directly related to the magnitude of the decrease in temperature. The central nervous system is the least tolerant tissue to hypoxia at normothermia, followed by the kidneys, liver, and heart. The range of ischaemic tolerance spreads from 5 to 6 min for brain cells to many hours for fat and skin. Hypothermia roughly doubles this tolerance time for each 5°C of cooling: the brain is protected from ischaemic damage for 6 to 9 min at 32°C, for 40 min at 20°C, and for 100 min at 12°C, though temperatures of less than 18°C are associated with other problems. Hypothermia shifts the oxyhaemoglobin dissociation curve to the left and, as blood is cooled, there is a slight increase in the solubility of oxygen in the plasma. These effects counterbalance the reduction in systemic oxygen delivery. Acid–base status is generally maintained when cooling proceeds in a gradual fashion. Serum glucose is elevated and is associated with a decrease in plasma insulin. There is probably decreased systemic utilization of glucose, as well as inhibition of insulin release secondary to hypothermia. There are no major alterations in serum electrolytes. Circulating concentrations of catecholamines increase and free fatty acids are released, probably because of central nervous activation of the sympathetic nervous system. The increase in glucose and free fatty acids during hypothermia may provide additional metabolic substrates for the heart and brain during the arrest interval.

Experimental evidence suggests that renal blood flow is substantially reduced during surface or core cooling. Reversible derangement of renal function occurs in patients undergoing hypothermia and circulatory arrest, though it is possible that decreases in cardiac output and increases in systemic vascular resistance during cooling may be chiefly responsible for renal dysfunction.

In 1971, Barrett-Boyes described the combination of surface and core cooling with a pump oxygenator during surgical correction of congenital cardiac defects. Surface cooling to 32°C followed by core cooling on cardiopulmonary bypass to 18°C was used routinely for many years. There has been a recent trend away from prolonged surface cooling. Measurements of acid–base status, oxygen consumption, serum lactate concentration and electrolytes, together with various serum enzymes including creatinine phosphokinase, transaminases, and dehydrogenases, have shown no significant difference between patients cooled to 20°C by cardiopulmonary bypass alone and those undergoing a combination of surface and core cooling. The surgical approach at most centres currently uses minimal surface cooling by exposure during anaesthesia in conjunction with core cooling on cardiopulmonary bypass.

During the induction of anaesthesia and the insertion of monitoring lines the infant is left undercovered, allowing the temperature to drift down. Rectal and nasopharyngeal temperatures are monitored with the knowledge that brain temperature always lags behind during cooling. The child is placed on the operating table on a temperature mattress, which is adjusted for surface cooling. The head is packed in ice bags. When the temperature reaches 34–32°C, median sternotomy is performed, the ascending aorta cannulated, and a single venous cannula inserted into the right atrial appendage. The bypass circuit is primed with packed red cells and a balanced electrolyte solution to create a haematocrit of 23 to 25 per cent during bypass. This degree of haemodilution improves peripheral perfusion during hypothermia while maintaining adequate oxygen delivery. During core cooling, frusemide (furosemide), 1 mg/kg, is administered if urine flow falls below 1 ml/kg per 30 minutes. When cardiopulmonary bypass is commenced the perfusate temperature is gradually lowered, maintaining a 10°C temperature gradient between the water bath and the blood; this 10°C gradient should never be exceeded. Flow rates are adjusted to obtain 120 ml/kg per minute in a newborn infant or 2.4 l/m^2 per minute in older infants. It is usual to add steroids to the pump prime before circulatory arrest since they have a protective effect against cellular oedema. When the rectal temperature reaches 18°C the aortic cross-clamp is applied, cardioplegic solution is infused into the aortic root, and cardiopulmonary bypass is discontinued. The venous blood is drained from the right atrium into the oxygenator, after which the venous cannula can be removed, facilitating intracardiac manipulation in a completely bloodless field. When the surgical procedure is complete the venous cannula is reinserted and the pump slowly restarted. When venous return commences, pump flow can be increased. Rewarming is carried out at approx. 2.4 l/m^2 per minute with a maximal 10°C gradient between rectal temperature, blood temperature, and the heat exchanger. The heart is de-aired and the aortic cross-clamp released to reperfuse the myocardium. Spontaneous rhythm is rarely restored until the temperature reaches 30 to 32°C. At completion of rewarming the patient is weaned from cardiopulmonary bypass. Calcium chloride (100 mg/kg) is administered before termination of bypass for its inotropic effects and to increase the blood calcium. After reversal of the action of heparin by protamine, platelet transfusions are given to all infants, since profound hypothermia results in thrombocytopenia.

The safe circulatory arrest time during profound hypothermia (18–20°C) is considered to be between 50 and 70 min. Safety is defined as the absence of any structural or functional damage as a result of the procedure. In practice, 30 min is regarded as completely safe, whereas 40 min has only a 90 per cent probability of safety. Sensitive tests of damage are difficult to apply in humans: experimental studies of cerebral morphology in animals have been used to extrapolate the limits of safety. In children who have recovered from a successful operation with an uncomplicated postoperative course, it is difficult to detect neurological deficit or specific abnormality of cerebral function when circulatory arrest has not exceeded 1 h at a rectal temperature of 20°C. Psychological assessment and measurements of the intelligence quotient do not provide a sensitive index of injury sustained from this technique. Isolated incidents of cerebral damage usually occur in patients who have had a stormy postoperative course or who have suffered cardiorespiratory arrest at some time.

In adults undergoing total circulatory arrest, barbiturate infusion (thiopentone 30–40 mg/kg) is the only technique proved to reduce cerebral metabolic oxygen requirements. Profound hypothermia together with barbiturates can reduce cerebral oxygen consumption to approximately 11 per cent of the normothermic, provided that the critical ischaemic time remains at 30 to 45 min. After this, progressive cellular damage occurs, probably as a result of free-radical generation. Superoxide is produced at reperfusion, with subsequent uncontrolled intracellular release of calcium ions. It is hoped that administration of the calcium-channel blocker nimodapine before ischaemia may improve outcome, but the search for clinically effective free-radical scavengers continues.

Interactions between blood and foreign surfaces

Improved methods for protection of the ischaemic myocardium and advances in surgical technique have greatly reduced surgical mortality. Morbidity associated with many types of operation stems predominantly from the damaging effects of cardiopulmonary bypass. Biocompatibility within this system consequently assumes great significance. Circulation of blood outside its natural endothelialized channels cannot be regarded as a biological process and damaging effects always occur. The artificial environment created by plastics, glass, and metal causes many alterations in the structure and function of blood, which flows through the circuit for periods of a few minutes to several hours (Fig. 7) (Table 3). Collectively, the adverse clinical manifestations of extracorporeal circulation were known as the ‘postperfusion syndrome' and more recently as the ‘whole-body inflammatory response'. They may complicate postoperative convalescence and may contribute to death from multisystem organ failure, particularly in the very sick or those at extremes of age.



Fig. 7. Foreign materials in a bubble oxygenator include polyethylene, polyurethane, polyvinyl chloride, and nylon.

Blood component	Induced changes
Proteins	
Albumin, γ-globulin, fibrinogen etc.	Adsorption, conformational changes, increased enzymatic degradation
Clotting factors	Activation of intrinsic coagulation causes the activation of clotting factors, enabling the ultimate conversion of fibrinogen into fibrin
Lipids	Adsorption, conformational changes
Formed elements	
Platelets	Adhesion, release, aggregation, platelet plug formation, deterioration of function, removal from the circulation, interference with intrinsic coagulation
Erythrocytes	Adhesion, release, haemolysis, removal from circulation, deterioration of function, interference with coagulation mechanism
Leucocytes	Adhesion, deterioration of function, removal from circulation

Table 3 Alterations in blood structure caused by circulation through an artificial environment

The interactions of the ‘contact system' of plasma and of the activated humoral cascades with the cellular elements of blood and tissues of the body are complex (Fig. 8, Fig. 9, Fig. 10 and Fig. 11). Many synthetic materials are exposed to blood within the various types of oxygenator. In bubble oxygenators the allegedly biocompatible materials include nylon, polyurethane, carbon-coated polyurethane, and polyethylene, all within a polycarbonate shell. The coils of the heat exchanger consist of carbon-coated aluminium. Membrane oxygenators may be of silicone rubber or microporous material, such as Teflon, polyethylene, polypropylene, or polytetrafluoroethylene (Gore-Tex). In bubble oxygenators the effects of direct gas–blood interaction include protein denaturation, complement activation, increased fragility of red cells, susceptibility to haemolysis, and denaturation of platelet membrane materials causing platelet aggregation, clumping, and removal from circulation. These effects of gas-interface oxygenators may be caused more by repeated renewal of the interface than by the nature of the interface itself. Membrane oxygenators have the advantage of eliminating effects from a blood–gas interface, though complement activation occurs by direct interaction of blood with foreign surfaces. Nevertheless, changes in plasma proteins, lipoproteins, and red cells that occur with gas-interface oxygenators are less evident with membrane oxygenators.

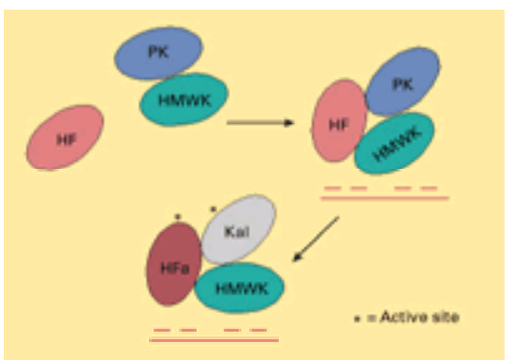


Fig. 8. Contact activation of Hageman factor by reaction with prekallikrein (PK) and high molecular-weight kininogen (HMWK).

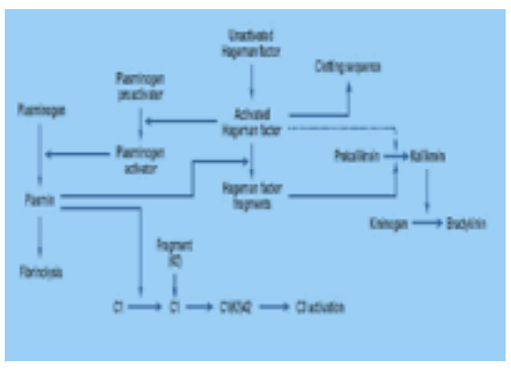


Fig. 9. Interactions of the ‘contact system' of plasma.

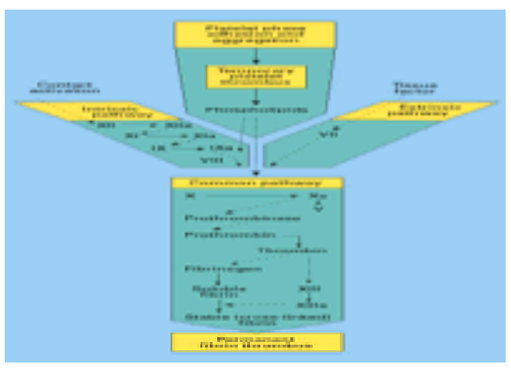


Fig. 10. Activation of the clotting cascade.

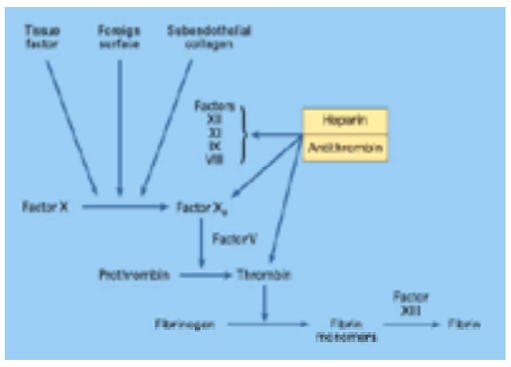


Fig. 11. Interaction of heparin with the clotting sequence.

In clinical circuits with membrane oxygenators a gas–blood interface still occurs in the cardiotomy suction apparatus that retrieves spilt blood from the operative field. This system consists of metal or plastic suckers and vents that aspirate and accelerate blood under negative pressure generated by a separate roller pump. An efficient

and safe cardiotomy suction system should have a large storage capacity, minimal suction to avoid blood trauma, minimal admixture of air, and reliable defoaming capacity. A wide variety of microemboli of gas, fat, fibrin, tissue debris, cellular clumps and foreign materials such as bone, wax, calcium from heart valves, glove powder, dust, and swab fragments are aspirated, and an efficient filter in the system is advantageous. Cardiotomy reservoirs have been designed to incorporate filters and defoaming chambers in an attempt to remove these microemboli.

The bubbles created by suction are qualitatively different from those generated in a bubble oxygenator, since they consist mainly of nitrogen, a relatively insoluble gas that may persist in microbubble form after passage through both the defoamer in the cardiotomy reservoir and the oxygenator. A large proportion of the red-cell destruction that occurs during cardiopulmonary bypass results from mechanical trauma and turbulence induced by cardiotomy suction. Aspiration of blood from the pericardium and intracardiac chambers causes microaggregation of formed blood elements, and should be minimized.

The damaging effects of cardiopulmonary bypass

An understanding of the nature of bioincompatibility in extracorporeal circuits requires an insight into the pathogenesis and effects of perfusion-related damage. When pronounced, this consists of a generalized increase in pulmonary capillary permeability (non-cardiogenic haemorrhagic pulmonary oedema), renal impairment, bleeding diathesis, neurological changes, and fever of non-infective origin. The ‘postperfusion syndrome' causes generalized organ dysfunction that is transient and inconsequential after most straightforward adult operations, but may necessitate respiratory or renal support, blood transfusion, or re-entry for diffuse abnormal bleeding in small infants or after prolonged bypasses for complex repair.

‘Pump lung' ranges from barely noticeable interstitial oedema to rare but fatal non-cardiogenic pulmonary oedema. Most patients have increased alveolar/capillary oxygen difference and increased fluid in the tracheobronchial tree. There is shunting due to both venous mixture and alveolar collapse. Total airflow resistance increases and there are measurable changes in pulmonary compliance. The changes in mechanical properties have been ascribed to non-cardiogenic increase in extravascular lung water (Fig. 12).



Fig. 12. Plain chest radiograph in a child with non-cardiogenic pulmonary oedema following cardiopulmonary bypass.

Both light and electron microscopy show consistent pathological changes in lung biopsies taken from patients up to 4 h after bypass (Fig. 13). There is engorgement of the pulmonary vascular bed, microatelectasis, and both interstitial and alveolar haemorrhage. There is damage and swelling of the capillary endothelial cells and, intracellularly, of the matrix compartment of the mitochondria and of the endoplasmic reticulum. The pericapillary space shows dispersion, and there are accumulations of leucocytes and their degranulation products in apposition to the capillary membranes, with local endothelial damage. Oedema and rarefaction of the cytoplasm can be seen within type I pneumocytes, and interstitial mast cells show loss in density and degranulation. These ultrastructural changes can be related to the duration of perfusion.

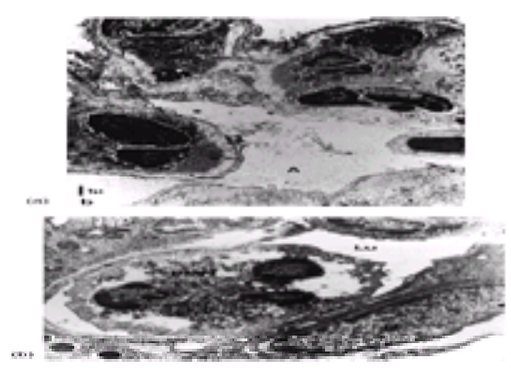


Fig. 13. Electron micrographs from lung after cardiopulmonary bypass:(a) granulocytes trapped in the pulmonary capillaries; (b) granulocyte in apposition with the pulmonary capillary membrane in a state of degranulation.

On total bypass the lung is hypoperfused and not ventilated, but is left partially inflated with 100 per cent oxygen. The lack of blood supply, with subsequent damage to type II pneumocytes, produces a decrease or total lack of surfactant production, which may contribute to the development of alveolar collapse. The lung is also the primary filter for intravascular emboli during partial bypass with the lungs in circuit. Nevertheless, the likely mechanism for pulmonary dysfunction is that of the ‘whole-body inflammatory response', with complement activation as the stimulus for organ damage (Fig. 14, Fig. 15). White cells activated by complement anaphylotoxins are known to aggregate in the pulmonary circulation at the end of cardiopulmonary bypass and 50 per cent of the total circulating neutrophils are found transiently in the pulmonary capillaries (Fig. 16(a),Fig. 16(b) and Fig. 16(c)). Release of protease enzymes (such as elastase) and oxygen free radicals can directly damage the pulmonary capillary membranes. Denaturation of a-protease inhibitor by free radicals may well facilitate proteolytic damage.

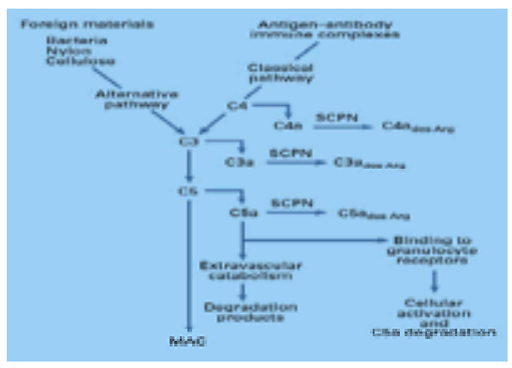


Fig. 14. Complement activation by classical and alternative pathways with generation of the anaphylatoxins C3a and C5a.

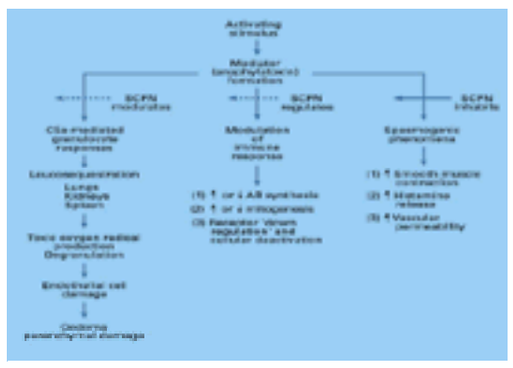


Fig. 15. Consequences of complement activation and anaphylatoxin release.

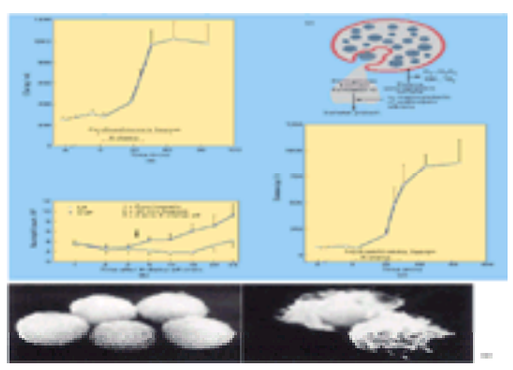


Fig. 16. (a) Complement anaphylatoxin C3a during and after cardiopulmonary bypass. (b) Sequestration of neutrophils in the pulmonary circulation after aortic cross-clamp release, resumption of cardiac activity, and pulmonary blood flow. (c) Release of lysosomal enzymes and generation of oxygen free radicals by ‘activated' polymorphonuclear leucocytes. (d) Elastase before, during, and after cardiopulmonary bypass. (e) Morphology (scanning electron microscopy) of unstimulated (left) and stimulated (right) granulocytes.

In support of the whole-body inflammatory response hypothesis, concentrations of the complement anaphylotoxin C3a have been shown to relate directly to postoperative organ system dysfunction (Fig. 17(a),Fig. 17(b) and Fig. 17(c)). Complement activation occurs by the alternative pathway. Plasma concentrations of elastase rise by a factor of 10 to 30 times during the course of cardiopulmonary bypass. The end-point of the inflammatory response is a generalized increase in microvascular permeability. In normal lungs, interstitial oedema does not create serious gas-exchange problems, but after major surgery reduced compliance and increased work of breathing may lead to a need for prolonged intubation, increased risk of infection, and death. Such complications explain the need for routine postoperative ventilation of cardiac surgical patients in most units. However, ventilation is unnecessary for the great majority of adult cardiac operations with bypass times of less than 1 h. In Oxford, 85 per cent of adult patients are extubated in a recovery area within 2 h of leaving the operating room.

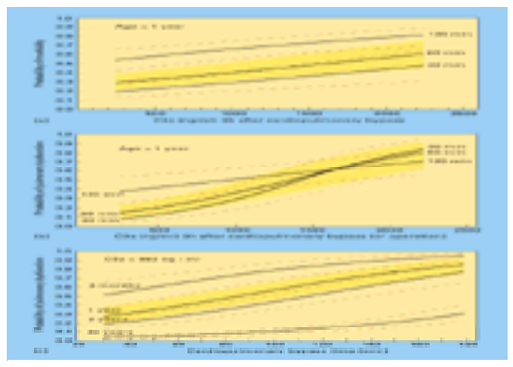


Fig. 17. (a, b) Probability of morbidity after 30, 60, or 120 min of cardiopulmonary bypass related to C3a concentrations 3 h after cardiopulmonary bypass. (c) Probability of morbidity related to duration of cardiopulmonary bypass in patients of different age. The young are particularly susceptible to the damaging effects.

Treatment of a severe episode of non-cardiogenic pulmonary oedema may include infusion of adrenaline (epinephrine), which has antiallergic, bronchial smooth-muscle relaxing, and haemodynamic properties, and pharmacological doses of steroids (equivalent to methylprednisolone at 30 mg/kg) to improve pulmonary blood flow, stabilize the endothelial wall, and stop the leakage of protease enzymes into the interstitial space. Positive end-expiratory pressure is used to improve gas exchange by increasing functional residual capacity, decreasing intrapulmonary shunting, and increasing interstitial pressure. Aggressive tracheobronchial toilet is essential for clearance of pulmonary oedema fluid.

Severe, acute renal failure requiring haemodialysis develops in about 2 per cent of patients after cardiopulmonary bypass, usually in association with multisystem organ failure. These patients are usually critically ill with poor left ventricular function and have a predicted mortality in excess of 70 per cent. The highest mortality rate is associated with the presence of respiratory failure, stroke, hypotension, and infection. Frank neurological deficit and major stroke occur in less than 2 per cent of patients undergoing cardiac surgery. Many of these problems are related to pre-existing carotid disease or macroscopic embolism from within a cardiac chamber. However, detailed psychometric testing discloses cerebral changes in the great majority of patients following cardiopulmonary bypass. Historically these have been related to microembolism, and recently microemboli have been demonstrated in the retinal vessels following cardiopulmonary bypass. The relation between subjectively imperceptible and transient psychological changes and well-being remains unclear. It is not unusual to find patients reading a newspaper 24 h after cardiac surgery, though the attention span is limited. Residual late psychometric impairment is usually mild, not apparent to the patient, and usually limited to mild cognitive impairment that has no significance in functional terms. Factors predisposing to major, long-term intellectual dysfunction include cardiac failure before surgery and global impairment of left ventricular function. Patients and their relatives should be warned that some impairment of memory, concentration, and reaction time may be noticed for a number of months after surgery, but that this usually recovers or improves to a level that will not interfere with normal activity. Patients in whom intraoperative control of $Paco_2$ is poor have a significantly higher incidence of postoperative neurological deficit on psychometric testing than those in whom $Paco_2$ is assiduously monitored and corrected. Alpha-stat (as opposed to pH-stat) regulation of blood gases preserves cerebral autoregulation and is the method of choice if cerebral effects are to be minimized.

Coagulation disorders seen after cardiopulmonary bypass are multiple and complex, and detailed description is beyond the scope of this chapter. Postoperative bleeding is minimized by careful surgical haemostasis, though abnormal bleeding may result from heparin–protamine imbalance, thrombocytopenia combined with abnormal platelet function, defects of the clotting cascade, and abnormal fibrinolysis. Bleeding tendencies are rare in patients undergoing short, moderately hypothermic bypass using membrane oxygenation and controlled cardiotomy suction to minimize platelet damage. When abnormal bleeding occurs in the presence of normal clotting factors, platelet dysfunction may be caused by binding of fibrinogen degradation products to surface membrane receptors. The protease inhibitor aprotinin protects against this and can be used in selected patients at substantial risk from postoperative haemorrhage (some early reoperations and patients with bacterial endocarditis). We do not advocate routine use of this drug and we do not use it for patients undergoing coronary bypass grafts. We have not found it to be useful in patients undergoing profound hypothermia and circulatory arrest, in whom platelet dysfunction is particularly common.

Treatment of abnormal bleeding depends on the precise cause, but includes protamine administration according to activated clotting time, infusion of platelets, fresh frozen plasma, or cryoprecipitate, and, rarely, the administration of an antifibrinolytic agent such as ee-aminocaproic acid. Cardiac surgeons know that warm fresh blood transfused directly from a donor is the best treatment for intractable, diffuse abnormal bleeding, though haematologists are at odds with this practice. Topical administration of cryoprecipitate or aprotinin to bleeding surfaces within the pericardium has proved effective when the chest cannot be closed safely. Reinfusion of

shed mediastinal blood by an autotransfusion system is used routinely in many centres.

For patients with profuse haemorrhage in the recovery area there should be a low threshold for re-entry, even if surgical bleeding is deemed unlikely. A surgical bleeding point is often identified; if not, packing of the open chest may be necessary. Multiple transfusions of donor blood and even continuous reinfusion of poor-quality, shed mediastinal blood greatly increase overall morbidity from pulmonary and renal dysfunction.

Protection of the heart during open heart surgery

Development of techniques for myocardial protection

Operations on the heart are facilitated by having the ascending aorta cross-clamped and the heart bloodless and quiet through cessation of coronary blood flow. Consequently, the heart is subject to global myocardial ischaemia and, even under conditions of moderate or deep hypothermia, the anoxic safe period without irreversible and fatal damage does not exceed 30 min. As early as 1955, Melrose and colleagues at the Hammersmith Hospital (London) used chemical cardioplegia with potassium citrate to arrest the heart and reduce myocardial oxygen demand. Although this technique was initially successful, it was abandoned when reports from the United States of America described a high incidence of ventricular fibrillation postoperatively and myofibrillar necrosis histologically following 30 min of ischaemia. Cardiac standstill was then induced by anoxic aortic cross-clamping or by continuous electrical fibrillation. Myocardial metabolic requirements were reduced by systemic hypothermia plus topical cooling in the pericardium. Intermittent coronary perfusion was used during the procedure to supply oxygen.

Hypothermia reduces cardiac rate and myocardial oxygen consumption, but at temperatures below 28°C, coronary resistance increases and subendocardial blood flow decreases. Topical hypothermia with iced saline slush in the pericardium can selectively decrease myocardial temperature below systemic temperature and increases myocardial protection. However, the heart fibrillates and continuous electrical fibrillation itself produces a 50 per cent rise in oxygen consumption and rapid fall in ATP, creatinine phosphate, and glycogen stores. During coronary arterial perfusion in a fibrillating heart, blood flow is diverted away from the subendocardial surface and any rise in left ventricular end-diastolic pressure further decreases subendocardial perfusion.

Intermittent coronary perfusion with moderate hypothermia and cardiac standstill induced by ventricular fibrillation has been used successfully for myocardial protection for many years. Nevertheless, this technique provides inadequate myocardial oxygenation, metabolite substrate delivery, and wash-out of catabolic products. Intermittent perfusion for 5 min followed by ischaemia for 15 min decreases myocardial contractility, compliance, and increases myocardial oedema. If the cyclical periods of aortic clamping are prolonged beyond 90 min there is a progressive fall in myocardial adenine nucleotides. Cellular swelling and myocardial oedema follow depletion of energy stores and failure of active membrane transport of small molecules, particularly sodium, potassium and chloride. Severe myocardial oedema reduces ventricular diasystolic function, stroke volume, and cardiac output.

In spite of surgical disadvantages, operations have been performed on normothermic, perfused, empty, beating hearts, a method advocated as optimal as late as 1975. Valve surgery can be carried out on a perfused, empty, beating heart by the technique of individual coronary arterial perfusion, using small individual cannulas placed in the ostia of the right and left coronary arteries and perfused with oxygenated blood by way of separate pumps and lines from the heart–lung machine. Continuous coronary perfusion has been combined with ventricular fibrillation, and similar techniques are still successfully applied by some to coronary arterial surgery. Nevertheless, the need for prolonged periods of cardiac arrest and improved myocardial preservation led to renewed interest in chemical cardioplegia. The hypothesis underlying the use of cold cardioplegic myocardial preservation is that it reduces myocardial oxygen demand during the ischaemic period of aortic cross-clamping to such low levels that the myocardial energy stores are sufficient to maintain cell structure and the energy-dependent cell membrane pumps that preserve transcellular gradients of sodium, potassium, calcium, and magnesium. Thus myocardial cell viability and function are preserved. Histological studies have shown a higher incidence of endomyocardial damage with continuous coronary perfusion and intermittent aortic cross-clamping than with cold cardioplegic arrest.

Experimental studies undertaken in the 1970s illustrated that the colder the heart during the ischaemic period, the better the myocardial protection. They also showed that abrupt electromechanical dissociation of the heart at the onset of ischaemia combined with rapid myocardial cooling provided better myocardial protection than either method alone. When the heart is electromechanically quiescent the energy demands of the myocardium are determined primarily by temperature, and myocardial energy is derived primarily from anaerobic metabolism of glycogen and glucose. The low energy output of anaerobic metabolism is sufficient to maintain myocardial viability during prolonged ischaemia if the energy demands are substantially lowered by cooling.

Experimental research and clinical experience indicate that the cardioplegic solution should be at 4°C and the myocardial temperature should be lowered to between 12 and 15°C at the start of the procedure and then maintained at between 15 and 20°C throughout the ischaemic period.

Cardioplegic solution

The ideal cardioplegic solution must be cold when infused (4°C), should induce cardioplegia rapidly, should produce no direct damage to cell membranes, and should minimize intracellular ionic changes. The chemical principles for inducing cardiac arrest are shown in Table 4. Potassium rapidly infused into the cross-clamped aorta depolarizes myocardial cells, producing sustained diastole. Asystole decreases the amount of ATP consumption and reduces oxygen demand. A potassium dose of 20 to 24 mEq in a crystalloid delivery solution produces a prompt arrest. Doses above 30 mEq/l increase coronary vascular resistance because of a calcium-activated rise in left ventricular wall tension. When blood is used in the cardioplegic solution to increase oxygen delivery the solution must contain 30 mEq/l KCl, since some potassium is taken up by the red cells. Magnesium is used at a concentration of 50 mEq/l to depress both the inherent rhythmicity of pacemaker cells and myocardial contractility. Pacemaker cells are unstable due to a slow inward flux of sodium, which eventually reduces action potentials to the firing levels. Magnesium blocks the inward flux of sodium and interferes with repolarization. Calcium and magnesium compete at cell-membrane receptor sites to activate or slow neuromuscular transmission. Magnesium inhibits the release of calcium by the sarcoplasmic reticulum, thereby inhibiting calcium influx and stabilizing potassium channels. Magnesium also suppresses myocardial ATPase activity and activates enzymes that transfer phosphate from ATP to ADP.

1. Myocardial depletion of calcium
2. Myocardial depletion of sodium
3. Elevation of extracellular potassium
4. Elevation of extracellular magnesium
5. Infusion of local anaesthetic agents
6. Infusion of calcium and antagonistic agents

Table 4 Chemical principles for inducing cardiac arrest

Because of the excellent operating conditions and effectiveness of myocardial preservation, cold cardioplegic arrest has become the most commonly used method of operative myocardial preservation. Hypothermic solutions in clinical use vary considerably in their components. Table 5 shows the constituents of the St Thomas' cardioplegic solution used at the Oxford Heart Centre; this employs a temperature of around 4°C. Hypothermia of the perfusate appears to be the most important factor in lowering of cellular energy requirement during ischaemia. In the 1960s, both cold blood and crystalloid solutions without any additional arresting agent were used experimentally and shown to provide satisfactory myocardial protection in the dog heart for ischaemic periods of 60 min. Arrest using potassium or procaine at normothermia proves only slightly better than unmodified normothermic ischaemic arrest. Nevertheless, the effects of an arresting agent plus hypothermia are additive and result in better postischaemic ventricular function than when either is used alone. Utilization of cellular energy stores is lowered by progressive levels of myocardial hypothermia and the lower the myocardial temperature of the arrested heart, the better the postischaemic myocardial performance.

Substance	Ringer's solution (mmol)	Cardioplegic solution (mmol)	Final concentration (mmol/l)*
Calcium chloride	2.25	0	2.20
Potassium chloride	4.02	15.06	10.59
Sodium chloride	147.53	0	144.15
Magnesium chloride	0	15.59	15.9682
Procaine hydrochloride	0	1.00	0.98
Distilled water	1000ml	20ml	1020ml

*Concentration corrected for final volume of 1020ml.
Osmolarity = approximately 300 mmol/kg H2O.

Table 5 St Thomas' cardioplegic solution no. 1

Arresting agent

The majority of cardioplegic solutions contain either potassium or procaine as the arresting agent. Potassium produces arrest by membrane depolarization, procaine by preventing the propagation of an electrical stimulus. Studies comparing the two agents have found no difference in postischaemic myocardial performance; both appear to decrease cellular energy requirements equally, in spite of their different mechanisms of action. Magnesium has also been used as an arresting agent and is present in high concentration in the Kirsh solution. Additives such as nifedipine and verapamil have been used effectively in animal experiments.

Calcium

Cardioplegic solutions appear to work well both with and without calcium in the medium. Proponents of calcium-containing solutions have supported its use by citing the ‘calcium paradox'. This syndrome was described in isolated rat heart experiments following normothermic aerobic perfusion with calcium free solution followed by perfusion with a calcium-containing medium. This perfusion sequence resulted in an irreversible loss of electrical and mechanical activity of the heart. The syndrome appears to be modified by profound hypothermia, which may account for the clinical results achieved with calcium-free solutions. Nevertheless, the presence of calcium has not been shown to be harmful and may well prove beneficial.

pH

Cellular acidosis is markedly deleterious for anaerobic metabolism, and this supports the use of a solution with pH 7.4 or above. Animal experiments have shown poor postischaemic recovery following the use of acidic cardioplegic solutions. Solutions used in clinical practice therefore contain bicarbonate or tris(hydroxymethyl)aminomethane (THAM) as their buffer.

Glucose

Glucose is used to provide substrate during anaerobic metabolism in some types of cardioplegic solution, though the benefit has not been conclusively proved.

Osmolarity

Wide variations in osmolarity of the perfusate are clearly harmful and some of the deleterious effects of the original Melrose solution (450 mmol) have been attributed to hyperosmolarity. Postischaemic ventricular performance following use of hyperosmolar solutions is poor when compared to the effect of solutions with an osmolarity between 300 and 320 mmol. Hypo-osmolar solutions (272 mmol) cause myocardial oedema. It therefore appears the solution should be at least iso-osmolar, and if hyperosmolar should not exceed 380 mmol.

Blood cardioplegia

Cardioplegia undertaken with cold oxygenated blood plus an arresting agent has the theoretical advantage of intermittently replenishing cellular energy levels by aerobic metabolism during period of reinfusion. However, it has been suggested that the lower limit of blood temperature should be 22°C since at lower temperatures the rheology of blood may be altered, with sludging of red cells and deleterious effects on the microcirculation. Blood cardioplegia therefore has the potential disadvantage of allowing a higher rate of energy consumption during ischaemia because the myocardial cooling is not as efficient as that with crystalloid solutions. Whereas myocardial temperatures below 15°C can easily be achieved with crystalloid solutions at 4°C, this degree of cooling cannot be achieved with blood at 22°C. Nevertheless, experimental findings and practical clinical experience with blood cardioplegia have shown little difference between the two methods.

Blood cardioplegia demands more elaborate systems for administration, and a higher concentration of potassium chloride (28–30 ml/l) is required to produce asystole. This need for a higher potassium chloride concentration is probably due to potassium influx into the red blood cell. Since the hyperkalaemic blood used for cardioplegia is usually returned to the cardiopulmonary bypass circuit, a significant rise in serum potassium is more likely then with crystalloid cardioplegia.

Since the chemically arrested heart at 15°C requires 0.3 to 0.5 ml oxygen/min per 100 g, the need to supply oxygen to the heart by blood cardioplegia is questionable. Oxygen demand of this magnitude could well be met be oxygenating asanguinous solutions, and may in any case be met by non-coronary collateral flow (estimated at about 2 ml/min per 100 g). Clinical trials comparing blood with crystalloid cardioplegia have shown that crystalloid cardioplegia provides more sustained and consistent myocardial protection. With blood cardioplegia, multiple doses seem necessary for comparable protection.

Evidence has accumulated to suggest that normothermic, potassium-containing, oxygenated blood cardioplegia (the so-called hot shot) is beneficial at the end of hypothermic cardiac arrest. Replenishment of oxygen and substrate by this technique appears to reduce ischaemia-induced ventricular failure.

Cardioplegia in clinical practice

Cold crystalloid cardioplegia produces rapid induction and sustained maintenance of cardiac arrest with profound myocardial hypothermia. Operative techniques should avoid injury to the coronary arteries or myocardium at the time of infusion and must subsequently avoid damage from reperfusion with blood during release of the aortic cross-clamp. In practice, variable rates and levels of myocardial cooling may occur, particularly in the presence of severe coronary arterial disease and through rewarming by non-coronary collateral flow. Retrograde delivery of cardioplegia via the coronary sinus has been used to circumvent the problem of coronary occlusion. For this method a balloon catheter is inserted through a purse-string suture in the right atrial wall and manipulated into the coronary sinus. The balloon is inflated and cardioplegia instilled under low pressure.

The efficacy of cold crystalloid cardioplegia is directly related to the level of myocardial cooling. Myocardial temperature monitoring can be performed using a needle thermistor in the septum in order to sustain a myocardial temperature below 20°C and preferably as close to 10°C as possible. The thermistor is inserted 1.5 cm to the right of the left anterior descending coronary artery and angled superiorly into the septum. When delivering cardioplegia, attention should be focused on those areas of the myocardium known from the coronary angiogram to have the poorest blood supply.

For coronary arterial bypass, repair of most congenital conditions, and mitral valve surgery, infusion of cardioplegia occurs directly through a needle in the aortic root after the cross-clamp application. This route can also be used in the treatment of aortic stenosis following vent insertion to ensure against left ventricular distension. In patients with aortic regurgitation, the aorta is opened and cardioplegia delivered directly into the coronary ostia through hand-held cannulas. Reinfusion of cardioplegia is often necessary for maintenance of myocardial arrest and adequate hypothermia. Further infusions are usually given at intervals of 25 to 30 min whilst the cross-clamp is in place. Should electrical activity resume earlier, because of collateral flow, reinfusion is undertaken sooner. Cardioplegic reinfusion washes out metabolic end-products and, if glucose-containing solutions are used, may replenish glucose for anaerobic metabolism. When myocardial temperature is measured directly, cardioplegia is repeated when the septal temperature rises above 20°C. The rate of myocardial temperature rise can be slowed by systemic hypothermia and topical hypothermia with ice slush in the pericardial sac. This reduces the tendency to rewarm through non-coronary collaterals that connect the left ventricle to the bronchial circulation.

During reperfusion, after cross-clamp removal and recovery of the heart before weaning from cardiopulmonary bypass, the heart is particularly susceptible to high

arterial pressure. At the time of cross-clamp release, perfusion pressure is therefore deliberately reduced to around 50 mmHg until the heart obtains tone and begins to beat or fibrillate. There is a high initial rate of spontaneous resumption of sinus rhythm, although ventricular fibrillation is common during the early reperfusion period. This probably results from electrical instability during resolving myocardial hypothermia and the transition of the cellular membrane from an arrested state to normal function. When the blood temperature is 35 to 37°C at aortic unclamping, rewarming of the hypothermic myocardium usually takes 5 to 10 min. Attempts to discontinue cardiopulmonary bypass before full recovery with stable rhythm may well prove unsuccessful.

When properly administered, cold potassium cardioplegia appears to maintain the entire myocardium in a protected and completely reversible state during periods of ischaemia lasting more than 2 h. Clinical reports have described patients with ischaemic periods of up to 3.5 h with normal postoperative myocardial function. Cardioplegia is superior to either intermittent coronary perfusion or prolonged ischaemia modified by profound topical hypothermia. Factors that decrease the effectiveness of coronary perfusion or topical hypothermia, including left ventricular hypertrophy, severe coronary disease, and prolonged aortic cross-clamping, do not appear to decrease the effectiveness of cold potassium cardioplegia. Further refinement in the composition of cardioplegic solutions and methods of administration will probably lead to even better clinical results.

Further reading

Akins K. Non cardioplegic myocardial preservation for coronary revascularisation. *Journal of Thoracic and Cardiovascular Surgery* 1984; **88**: 174–7.

Bretschneider HJ. Myocardial protection. *Thoracic and Cardiovascular Surgery* 1908; **28**: 295–8.

Brody WR, Reitz BA. Topical hypothermic protection of the myocardium. *Annals of Thoracic Surgery* 1975; **20**: 66–70.

Buckberg GD. A proposed ‘solution' to the cardioplegic controversy. *Journal of Thoracic and Cardiovascular Surgery*, 1979; **77**: 803–6.

Fox LS, Blackstone EH, Kirklin JW, Bishop SP, Bradley EL. Relationship of brain blood flow and oxygen consumption to perfusion flow rate during hypothermic cardiopulmonary bypass. *Journal of Thoracic and Cardiovascular Surgery* 1984; **87**: 658–64.

Fox LS, Blackstone EH, Kirklin JW, Stewart RW, Samuelson P. Relationship of whole body oxygen consumption to perfusion flow rate during hypothermic cardiopulmonary bypass. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 239–43.

Hearse DJ. The protection of the ischaemic myocardium: surgical success vs clinical failure. *Progress in Cardiovascular Diseases* 1988; **30**: 381–402.

Kirklin JW, Conti VR, Blackstone EH. Prevention of myocardial damage during cardiac operations. *New England Journal of Medicine* 1979; **301**: 315–21.

Kirklin JK, Westaby S, Blackstone EH, Kirklin JW, Crenoweth DE, Pacifico AD. Complement and the damaging effects of cardiopulmonary bypass. *Journal of Thoracic and Cardiovascular Surgery* 1983; **86**: 845–51.

Kirklin JW, Barratt Boyes BG, ed. *Cardiac surgery*, p. 83. Wiley, New York, 1986.

Mavroudis C. To pulse or not to pulse. *Annals of Thoracic Surgery* 1978; **25**: 259–61.

Murkin JM, Farrar JK, Tweed WA, McKenzie FN, Guiraudon G. Cerebral autoregulation and flow/metabolism coupling during cardiopulmonary bypass: the influence of $P\text{aco}_2$. *Anesthesia and Analgesia* 1987; **66**: 825–32.

Rahn H, Reeves RB, Howell BJ. Hydrogen ion regulation, temperature and evolution. *Annual Review of Respiratory Diseases* 1975; **112**: 165–9.

Sade RH, Bartles DM, Dearing JP, Campbell LJ, Loadholt CB. A prospective randomised study of membrane vs bubble oxygenators in children. *Annals of Thoracic Surgery* 1980; **29**: 502–7.

Swain JA. Cardiac surgery and the brain. *New England Journal of Medicine* 1993; **329**: 1119–20.

Takamoto S *et al.* Comparison of single dose and multiple dose crystalloid and blood potassium cardioplegic during prolonged hypothermic aortic occlusion. *Journal of Thoracic and Cardiovascular Surgery* 1980; **79**: 19–24.

Utley JR, ed. *Pathophysiology and Techniques of Cardiopulmonary Bypass*, Vol. 1. Williams and Wilkins, Baltimore, 1982.

Utley, JR, ed. *Pathophysiology and Techniques of Cardiopulmonary Bypass*, Vol. 2. Williams and Wilkins, Baltimore, 1983.

Van Oeveren W *et al.* Effects of aprotinin on haemostatic mechanisms during cardiopulmonary bypass. *Annals of Thoracic Surgery* 1987; **44**: 640–5.

Westaby S. Aspects of biocompatibility in cardiopulmonary bypass. *Critical Reviews of Biocompatibility* 1987; **3**: 193–234.

Westaby S. Organ dysfunction after cardiopulmonary bypass: a systemic inflammatory reaction initiated by the extracorporeal circuit. *Intensive Care Medicine* 1987; **13**: 89–94.

40.3 Circulatory support for the failing heart

Francisco A. Arabía, Jack G. Copeland and Richard G. Smith

Patient selection
Devices
Intra-aortic balloon pump
Axial-flow pump
Centrifugal pumps
Pulsatile paracorporeal pumps
Internal ventricular assist devices (pulsatile intracorporeal pumps)
Total artificial hearts
Anticoagulation
Complications
Bleeding
Thromboembolism
Infection
Hemolysis
Multiorgan failure
Technical tips
Future directions
Further reading

Orthotopic heart transplantation is recognized as the best therapy for endstage congestive heart failure. However, under the current procurement system, approximately 15 to 30 per cent of potential recipients die while waiting for a donor heart. This discrepancy generates a realistic need for circulatory assist(ance) devices for cardiac support and replacement.

Mechanical circulatory assistance was initially used for postcardiotomy cardiogenic shock, but more recently it has been used for temporary cardiac support, ‘bridge to recovery' (see below), or until a donor heart becomes available for transplantation, ‘bridge to transplantation'(see below). The use of all these devices has continued to increase in recent years. This chapter discusses the range of devices currently used for cardiovascular assistance and the surgical technique for implantation. It is possible to state with some certainty that these devices and their descendants will one day become an alternative to transplantation.

Patient selection

This is probably the most important factor in determining success with mechanical circulatory assistance. Early experience with the use of postcardiotomy devices for cardiac recovery showed that patients younger than 60 years had a survival rate of 21 to 31 per cent, those older than 60 years a survival rate of 12 per cent, and those over 70 years a survival rate of 6 per cent. Other factors contributing to successful outcomes include careful selection of assist device, and the skill, experience, and judgment of the implanting team. Table 1 lists basic criteria used as guidelines for placement of ventricular assistance. These guidelines comprise essentially a definition of cardiogenic shock. When the criteria are met, prompt intervention is important as prolonged hypotension (longer than 12 h) is associated with multisystem failure and poor recovery. Measurements of cardiac output and atrial pressure are necessary to assess the patient's need for a device. These measures are also very helpful in the immediate postoperative period, especially after the implantation of an univentricular device.

Cardiac index < 2.0l/m^2 per minute
Systemic vascular resistance > 2100dyn/sec per cm^2
Atrial pressure > 20mmHg
Systemic hypotension (systolic pressure < 80mmHg)
Urine output < 20 ml/h with:
Metabolic acidosis
Failure to separate from cardiopulmonary bypass

Table 1 Criteria for insertion of cardiovascular assist devices

The necessity for biventricular rather than univentricular support must be addressed very early in the evaluation. Patients who need a bridge to transplantation may have uni- or biventricular support, depending on their preoperative hemodynamics. In general, those who have low right atrial pressure, normal or slightly elevated pulmonary vascular resistance, and no ventricular arrhythmias may benefit from left ventricular support only. Approximately 20 to 25 per cent of patients who receive left ventricular support alone may require right ventricular assistance.Table 2 summarizes current criteria used to differentiate between the need for uni- or biventricular support.

Right ventricular failure: RVAD	
Left atrial pressure < 15mmHg	
Right atrial pressure > 20mmHg	
Few or no arrhythmias	
Near normal left ventricular function by echocardiogram, ventriculogram, or multigated acquisition scan	
Left ventricular failure: LVAD	
Left atrial pressure > 20mmHg	
Right atrial pressure < 15mmHg	
Few or no arrhythmias	
Near normal pulmonary vascular resistance	
Normal right ventricular function by echocardiogram, ventriculogram, or multigated acquisition scan	
Biventricular failure: BIVAD	
Right and left atrial pressures > 20mmHg	
Ventricular tachycardia or fibrillation	
Severely impaired right and left ventricular function by echocardiogram, ventriculogram or multigated acquisition scan	

RVAD, right ventricular assist device; LVAD, left ventricular assist device; BIVADS, biventricular assist devices.

Table 2 Guidelines for selection of cardiovascular assist devices

An additional factor to be considered when choosing a ventricular support device is whether a continuous-flow or a pulsatile device should be used. The use of an intra-aortic balloon pump with a continuous-flow device (centrifugal pump) appears to improve organ function. If a pulsatile flow is chosen, a choice must also be made between synchronous or asynchronous contractions of the device with the native heart. In the past, synchronous counterpulsatile flow was generally used when reversible heart damage was suspected and augmentation of blood flow to the endocardium was desired to facilitate myocardial recovery. Asynchronous pulsatile flow was used when myocardial damage was irreversible and maximal blood flow was required to maintain end-organ function. There are now case reports of patients being weaned off asynchronous devices who experienced adequate recovery of their myocardium; these cases are now called ‘bridge to recovery'.

Transesophageal echocardiography has been found extremely useful during the implantation of devices because it can immediately provide information on the proper orientation of the device so that it can function under optimal conditions. It is also helpful in evaluating right ventricular function when only a left ventricular assist device is utilized. It also provides helpful assistance in de-airing the heart or atria after implantation and before discontinuing cardiopulmonary bypass.

Devices

A large number of devices have been invented in the last few decades. Many have not left the drawing board while others have made it to laboratory experimentation (see Table 3). Those presented have made it into the clinical arena or offer a great deal of hope.

Device	Power supply	Flow range (l/min)	Indications	Percutaneously
Intra-aortic balloon pump	Electrical	0.5–0.8	Left ventricular failure	No
Axial-flow pump	Electrical	<3.5	Left ventricular failure	No
Centrifugal	Electrical	2.6–5.0	Right ventricular, left ventricular or biventricular	No
Paracorporeal				
Thoratec	Pneumatic	<4.5	Right ventricular, left ventricular or biventricular	Yes
Abiomed BVS	Pneumatic	<5.0	Right ventricular, left ventricular or biventricular	No
Intracorporeal ventricular				
Novacor LVAS	Electrical	<8.0	Left ventricular failure	Yes
Heartmate	Pneumatic or electrical	<8.0	Left ventricular failure	Yes
Total artificial heart				
Cardiowest	Pneumatic	<15.0	Biventricular failure	Yes
Abiomed*	Electromechanical	<10.0	Biventricular failure	Yes
Cleveland Clinic–Nimbus*	Hydraulic	<8.0	Biventricular failure	Yes
Baylor*	Electrical	<9.0	Biventricular failure	Yes
Penn State*	Electrical	<9.0	Biventricular failure	Yes

*Experimental devices.

Table 3 Circulatory assist devices

Intra-aortic balloon pump

This is the simplest and most frequently used method of circulatory assistance today (Fig. 1). Its operation is timed with the electrocardiographic or aortic pressure waveform. The device inflates during diastole (counterpulsation), propelling blood into the coronary arteries and the peripheral circulation. The effects of this pump on the circulation include augmentation of the diastolic pressure, a decrease in the afterload, and a decrease in the myocardial oxygen consumption. Use of an intra-aortic balloon pump improves cardiac function, augmenting cardiac output by approximately 10 per cent or 500 to 800 ml/min. The device is usually used in combination with inotropic agents.

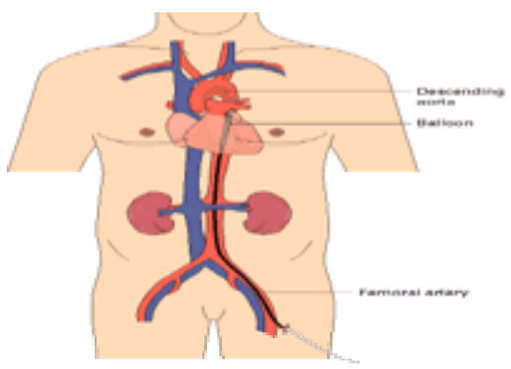


Fig. 1. Schematic diagram of an intra-aortic balloon pump inserted via the left femoral artery with the balloon properly distal to the left subclavian artery.

The intra-aortic balloon pump has several therapeutic uses in modern medicine. It is widely used in left ventricular failure following cardiopulmonary bypass. It is also used for patients with cardiogenic shock who are on maximal inotropic support while waiting for cardiac transplantation, or for patients with cardiogenic shock following myocardial infarction or intractable angina. In cases of right ventricular failure, the pump has been placed in the pulmonary artery.

Contraindications for the use of an intra-aortic balloon pump include aortic valvular insufficiency, aortic aneurysm, and severe aortoiliac or femoral disease precluding the insertion of the balloon. When aortic valvular insufficiency exists, augmentation of the diastolic pressure will cause an increase in the regurgitant flow across the valve and result in left ventricular distension. In patients with aortic aneurysm, the aortic wall may be perforated during the insertion of a balloon pump or as a result of pressure augmentation by the pump.

The intra-aortic balloon pump is usually inserted via one of the femoral arteries. It can be placed in the operating room, in a cardiac catheterization laboratory, or by the bedside utilizing local anesthesia and intravenous sedation. Under special circumstances, this device could be placed via the axillary arteries or directly into the ascending aorta through a median sternotomy/thoracotomy. It is always advanced in the aorta to a location distal to the take-off of the left subclavian artery. Its position is recorded by fluoroscopy, chest radiography, or echocardiography.

The pump is available in three different sizes. The 34-cc (ml) balloon is recommended for patients who are less than or equal to 61 inches (approximately 155 cm) tall, the 40-cc (ml) balloon for patients who are between 62 (approximately 158 cm) and 71 (approximately 180 cm) inches tall, and the 50-cc (ml) balloon for patients who are taller than 71 inches.

Axial-flow pump

P>The Hemo-pump or Nimbus pump (Johnson and Johnson, Rancho Cordova, CA) is a relatively simple device that provides nonpulsatile flow rates of up to 3.5 l/min. It is essentially a miniature axial-flow pump at the end of a catheter. The pump is inserted through a 12-mm woven graft that is sutured to a femoral or iliac artery. Under fluoroscopic guidance, the cannula is advanced into the aorta, across the aortic valve, and its tip positioned at the apex of the left ventricle (Fig. 2). The axial pump resides in the descending aorta in a 17 × 17 mm cylindrical housing at the end of a 20-cm flexible inflow cannula. The pump aspirates blood from the left ventricle and ejects it into the descending aorta. The device generates 3 l/min when rotating at 25 000 rev/min and discharging into an output load of 100 mmHg pressure. The Hemo-pump is a disposable, ‘one-time use' device that is currently under clinical investigation in the United States. Stationary blades provide directional flow of the blood as it exits the axial pump.

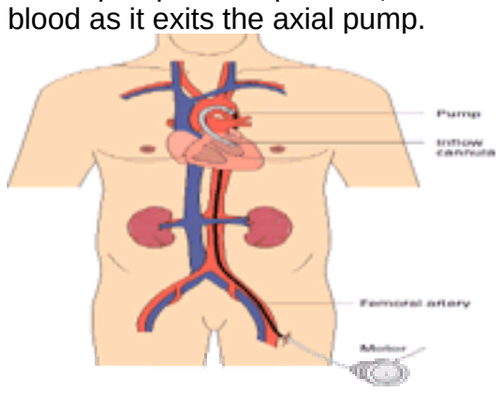


Fig. 2. Diagram of an axial-flow pump inserted via the left femoral artery and positioned at the apex of the left ventricle.

Indications for placement of an axial-flow pump are similar to those for the intra-aortic balloon pump. The Hemo-pump has been used in patients with refractory cardiogenic shock after coronary arterial bypass grafting, acute myocardial infarction, ischemic cardiomyopathy, postpartum cardiomyopathy, and rejection of a

transplanted heart. A contraindication to its use is severe aortoiliac disease that will prevent introduction of the inflow cannula and pump into the heart.

Centrifugal pumps

These devices utilize centrifugal force to generate a nonpulsatile flow. Flow rate is dependent on the rotational speed of the pump. Blood enters the pump at the apex and is accelerated outward. They are used to assist the left, right (see Fig. 3(a), Fig. 3(b)), or both ventricles. The inflow cannula to the pump originates from the left or right atrium and the outflow goes to the aorta or pulmonary artery. The centrifugal pump is extracorporeal, with the inflow and outflow cannulas passing through the chest wall. The pump is electromagnetically coupled to a motor that connects to the control console to seal completely the console from the blood.

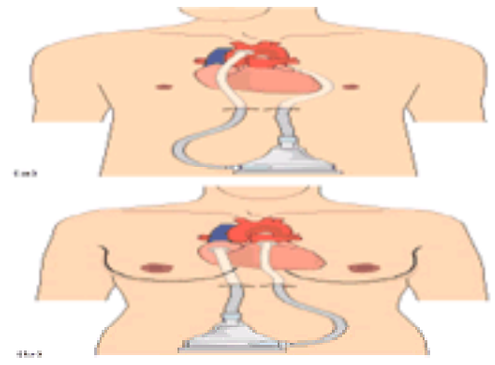


Fig. 3. (a) Diagram of a centrifugal pump positioned as a left ventricular assist device with the inflow cannula in the left atrial appendage and the outflow conduit in the ascending aorta; (b) diagram of a centrifugal pump positioned as a right ventricular assist device with the inflow cannula in the right atrium and the outflow conduit in the pulmonary artery.

Flow rates from 2 to 5 l/min are obtained with these devices. When centrifugal pumps are used, it is recommended that they not be operated at outputs of less than 0.5 l/min. When patients are weaned from these devices and flow rates below 2 l/min are used, anticoagulation is increased to lower the potential for thrombus formation.

Centrifugal pumps have been used in cardiogenic shock, as a bridge for transplantation, to support a failing heart after transplantation, and for patients who cannot be weaned from cardiopulmonary bypass.

A centrifugal pump can also be connected to an oxygenator (Fig. 4). In this setting, it can be utilized emergently to provide partial cardiopulmonary bypass for a short period of time if the patient is expected to recover or be bridged to a more sophisticated device. This configuration is utilized in some institutions for routine intraoperative cardiopulmonary bypass. In emergency the system is usually placed via the femoral artery and vein. It is good practice to place a small cannula off the arterial line to perfuse the distal leg when the main cannula is placed in the femoral artery. The main cannula tends to compromise flow to the lower leg; by placing a second, smaller cannula in the femoral artery, some flow is maintained distally. This technique is beneficial in small patients with small arteries. A flow meter can be attached to this second cannula.

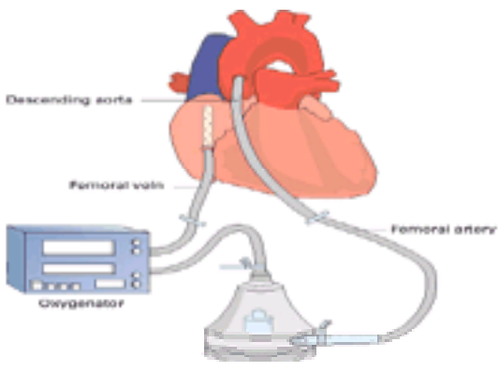


Fig. 4. Centrifugal pump with oxygenator used for partial cardiopulmonary bypass via the femoral vessels.

Pulsatile paracorporeal pumps

Paracorporeal pumps provide uni- or biventricular pulsatile flow. A system of two or four cannulas provides in- and outflow to the blood pumps. The most commonly used devices have been Thoratec (Berkeley, CA) and Abiomed (Danvers, MA). These two devices are commercially available.

The Thoratec is a pneumatically driven, paracorporeal prosthetic ventricle. This system has been used successfully as a bridge to recovery or transplantation. The inflow cannula is placed in the right or left atrium. Unidirectional mechanical valves are located in the inflow and outflow ports of the device. The outflow conduits are anastomosed to the ascending aorta or pulmonary artery via a Dacron graft. However, better filling of the left device is obtained if the its inflow cannula originates from the patient's left ventricular apex. A polyurethane sac divides the blood chamber from the air chamber within the prosthetic ventricle. The stroke volume from each paracorporeal ventricle is about 65 ml. The device is capable of delivering flow of up to 6.5 l/min. The pneumatic drives connect to a console that is a few feet from the patient and the system can be triggered in synchronous or asynchronous mode.

The Thoratec system can be installed in different configurations, including as a single ventricular (Fig. 5) or a biventricular assist device (Fig. 6(a), Fig. 6(c)). The left ventricular pump can be used as a bridge to recovery by placing a left atrial cannula in the left atrial appendage or in the interatrial groove. The outflow cannula is anastomosed to the ascending aorta. The suture more commonly used to anastomose the outflow conduit of any device to the aorta or pulmonary artery is of polypropylene, reinforced with pledgets at the superior and inferior poles of the anastomosis; this can sometime be accomplished without cardiopulmonary bypass. When the device is intended as a bridge to transplantation, it is recommended that the inflow cannula to the device be anastomosed to the left ventricular apex; a special cannula has been designed for this purpose. It is in this configuration that the maximal output is obtained for this particular device. Cardiopulmonary bypass is required when the left ventricular apical cannula is used. Once on bypass the apex of the heart is lifted. Six to eight interrupted, pledgeted sutures are placed around the circumference of the apex. The apex is the cored out; the diameter of the core is approximately 1 cm. When a right-sided device is also used, the inflow cannula is placed in the right atrium. The outflow cannula is anastomosed to the pulmonary artery. All cannulas are placed in the upper abdomen and tunneled through the skin and fascia into the pericardial space. The ventricular assist devices are filled with saline and then connected to the respective inflow and outflow cannulas. A 5-mm incision is made in the outflow conduits and a catheter is advanced through the outflow valve into the ventricular assist device to remove all remaining air; this is done in each device, one at a time. The incision in the outflow conduits is closed with 4–0 or 5–0 polypropylene. The patient is then placed in the Trendelenburg position. If cardiopulmonary bypass is utilized, the single ventricular or biventricular assist devices are allowed to take over the circulation by increasing the rate and decreasing the cardiopulmonary bypass flow rate. If cardiopulmonary bypass is not utilized, then the single or the biventricular devices are allowed to take over the circulation. Protamine is administered; catheters are placed to drain the mediastinum. The sternum is closed in the usual fashion. The inflow and outflow cannulas need to be secured to the skin to prevent motion and possible dislodgment.

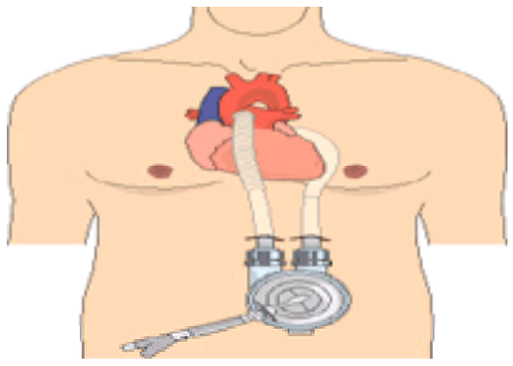


Fig. 5. Diagram of the Thoratec left ventricular assist device with the inflow cannula in the atrial appendage and the outflow conduit in the aorta.

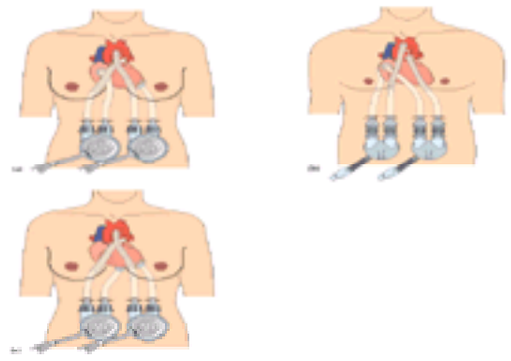


Fig. 6. (a) Diagram of the Thoratec system in the biventricular configuration: the left ventricular assist device (VAD) has an inflow cannula connected to the left ventricular apex and the outflow conduit is to the aorta; the right VAD inflow is from a cannula in the right atrium and the outflow in the pulmonary artery. (b) Diagram of the Thoratec system in the biventricular configuration using a left atrial cannula in the interatrial groove as the inflow conduit. (c) Diagram of Thoratec with biventricular cannulation.

The Abiomed BVS 5000 system is also a paracorporeal, pulsatile, pneumatic device that can provide short-term uni- or biventricular support (Fig. 7). Each blood pump consists of two polyurethane chambers that are separated by trileaflet polyurethane valves. Blood flows continuously through both chambers during systole and during diastole by gravity from the patient's atrium into the first chamber of the blood pump. From here, it flows passively across the inlet valve into the active, ventricle-like chamber where it is pneumatically propelled into the patient's great vessels. The pumps are operated by a console that determines the pulsatile rate and systole:diastole ratio based on the compressed air flow into and out of the chambers. The system maintains a stroke volume of 82 ml and was designed as a support system for hearts that have sustained reversible damage, particularly in failure to wean from cardiopulmonary bypass. It has also been used as a bridge to transplantation. The Abiomed can provide flows of up to 5 l/min.

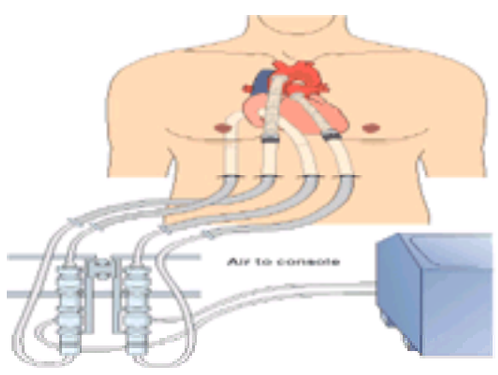


Fig. 7. Diagram of the Abiomed BVS 5000 system used for biventricular support.

The Abiomed BVS 5000 is placed via a median sternotomy; cardiopulmonary bypass may be required. The atrial cannula has a lighthouse tip. The left atrial cannula can be placed via the atrial appendage, below the interatrial groove, or through the dome of the left atrium. The right cannula can be placed in the mid-free wall or right atrial appendage. The arterial cannula utilizes a 14-mm woven Dacron graft instead of a lighthouse tip. These grafts are trimmed to attain optimal cannula position and attached to the ascending and pulmonary artery. The cannulas are externalized below the costal margin.

Internal ventricular assist devices (pulsatile intracorporeal pumps)

The Novacor (Division of Baxter Health Care Co., Oakland, CA) and the HeartMate (Thermo Cardiosystem, Inc., Wobern, MA) represent prototypes of implantable, univentricular assist devices that provide pulsatile assistance to the failing left ventricle. The Novacor is an electrically driven pump which energizes a solenoid that moves a dual pusher plate to propel blood into the aorta. Originally described in 1983 for providing long-term cardiac assistance, the Novacor may well become the first totally implantable, permanent ventricular assist device. Its main application has been as a bridge to cardiac transplantation, but it has also been used to support patients in profound cardiogenic shock after cardiac surgery. The Novacor is implanted in the left upper quadrant of the abdomen, anterior to the fascia of the rectus abdominis muscle (Fig. 8). It weighs approximately 3.3 kg and occupies a volume of about 400 ml. The inflow to the device comes via a low-porosity Dacron conduit, which is anastomosed to the apex of the left ventricle (Fig. 9). The outflow is via a similar Dacron conduit, which is usually anastomosed to the ascending aorta or, occasionally, the abdominal aorta. Bioprosthetic (porcine) inflow and outflow valve conduits maintain unidirectional flow. The power line and vent are placed subcutaneously and exit the patient in the right lower quadrant of the abdomen. A series of sensors are attached to the pump mechanism and exit the body to connect to a console where information about filling, stroke volumes, pump rate, and energy usage are displayed for analysis. This device can provide flows up to 8 l/min. The currently available Novacor left ventricular assist device has the inflow and outflow porcine valves incorporated into the conduits themselves. This new model has the letters ‘pc' attached to it, which refer to the valve conduits. The change in valve configuration should decrease the areas of stasis around the valves and the potential for thrombus formation. Currently a protocol exists to send patients home with a Novacor while waiting for heart transplantation.

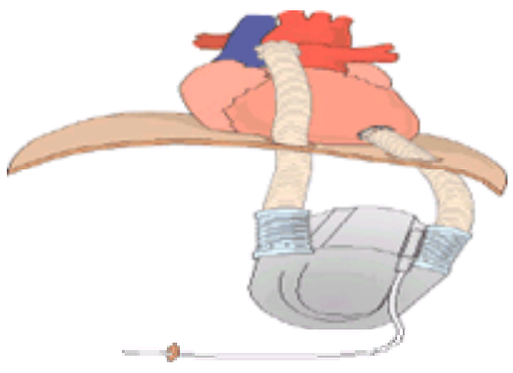


Fig. 8. Diagram of the Novacor left ventricular assist device with valve conduits, with inflow in the left ventricular apex and the outflow to the aorta; the device rests in

the preperitoneal space.

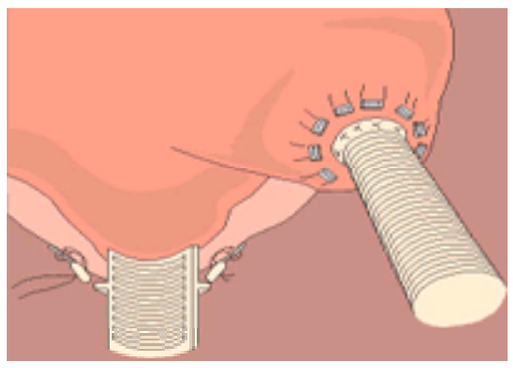


Fig. 9. Novacor inflow cannula attached to the left ventricular apex with pledgeted sutures.

The surgical implantation of the Novacor requires cardiopulmonary bypass. A median sternotomy is made, with extension of the skin incision in the midline to below the umbilicus. The sternum is divided in the usual fashion. The subcutaneous tissues in the abdomen are divided to the fascia of the rectus muscle but not through it. The midline in the fascia should be preserved intact to avoid incisional hernias at a later date. The fascia is incised to the left of the midline over the rectus abdominis muscle. It is than elevated from the muscle all the way to the subcostal margin superiorly, laterally to the anterior axillary line, and inferiorly to the superior iliac crest. The fascia is also incised to the right of the midline and elevated from the right rectus abdominis, superiorly to the subcostal margin, laterally to the midclavicular line, and inferiorly to the same level as the left incision. Tunnels are made through the diaphragm, keeping in mind the position of the left ventricular apex so that the inflow cannula to the Novacor has a gentle curvature and no tension. This tunnel is to the left of the midline. The outflow conduit from the Novacor is passed through a tunnel to the right of the midline, keeping in mind the need for a gentle curvature to go from the Novacor outlet to the ascending aorta. Great care should be used when creating the tunnels in order not to enter the abdominal cavity. Any perforation of the peritoneum should be repaired at this time. A single or double purse-string is placed in the ascending aorta and a single purse-string in the right atrial appendage. The purse-string in the aorta should be placed keeping in mind the end-to-side anastomosis of the outflow conduit. Blood is withdrawn from the right atrium to preclot the inflow and outflow conduits of the Novacor. Heparin is then administered. The new aortic conduit is albumin-impregnated Dacron and does not require preclotting. The ascending aorta is cannulated and the right atrium is cannulated with a two-stage venous cannula. The outflow conduit is passed through its tunnel. A partial-occluding clamp is place in the ascending aorta. An aortotomy is made for a length of about 1.5 cm. The outflow conduit is stretched and a tunnel is made from the subcutaneous pocket through the anterior aspect of the diaphragm into the pericardial space. It is then anastomosed to the ascending aorta with a running 4–0 polypropylene suture. Pledgets are used to anchor the stitches at the toe and heel of the Dacron conduit. This anastomosis can usually be done without cardiopulmonary bypass. The conduit is then totally occluded, the partial-occluding clamp on the aorta is removed, and the anastomosis is inspected for hemostasis. Placement of the inflow cannula requires cardiopulmonary bypass as a core has to be removed from the left ventricular apex; this does not require arresting the heart or fibrillating it. The apex of the heart is lifted with the surgeon's left hand or by placing surgical sponges underthe heart. Twelve interrupted, pledgeted sutures are placed around the circumference of the apex (Fig. 9). The apex is cored out; the core has a diameter of about 1.5 cm. The inflow cannula is sutured with the 12 pledgeted sutures. The conduit is tunneled from the pericardium through the upper aspect of the diaphragm to the left side of the subcutaneous pocket. Additional interrupted stitches may be required between the apex and the conduit to prevent bleeding. The inflow and outflow bioprosthetic valves are placed in the device.The conduits are then connected to the device after filling it with saline solution in order to de-air the Novacor. The drive-line to the Novacor is placed in the pocket and tunneled out through the skin in the right lower quadrant of the abdomen. The patient is placed in Trendelenburg position. A 1-cm transverse incision is placed in the outflow conduit while occluding its distal aspect. The Novacor is then allowed to pump by providing single contractions in order fully to de-air the system. Once all the air has been removed a 4–0 polypropylene stitch is used to close the incision in the conduit. The Novacor is allowed to take over the circulation by increasing the rate and cardiopulmonary bypass is discontinued. Protamine is then given once the patient is stable. Two mediastinal, soft suction catheters are placed in the pericardial space. Two large-bore suction catheters (Jackson–Pratt drains; Baxter Health Care Corp., Irving, CA) are also placed around the Novacor to prevent hematoma formation. The incision in the abdominal wall is closed in a single layer with a running 0 Dexon in the fascial layer followed by vertical skin staples. If necessary, interrupted 2–0 polypropylene mattress stitches could be used to provide additional strength to the wound. The sternotomy is closed in a regular fashion. The skin may then be reapproximated with staples.

The HeartMate assist device is a totally implantable, pneumatically or electrically driven, dual-chamber pump in a rigid titanium housing (Fig. 10). A flexible polyurethane diaphragm divides the air and blood chambers. The metallic side of the blood chamber is made of sintered titanium microspheres. This surface promotes lining of the device with the patient's own blood products to minimize clot formation. This ventricular assist device is positioned either intra-abdominally or preperitoneally without entering the abdomen. The inlet Dacron conduit is anastomosed to the apex of the left ventricle and the outlet conduit goes to the aorta. Porcine valves (25 mm), located at the inlet and outlet of the device, provide unidirectional flow. The pneumatic drive line is positioned subcutaneously and exits the skin at a distance from the device to connect to the external console. The device can generate up to 8 l/min of blood flow. It was the first device approved by the Food and Drug Administration as a bridge to heart transplantation.

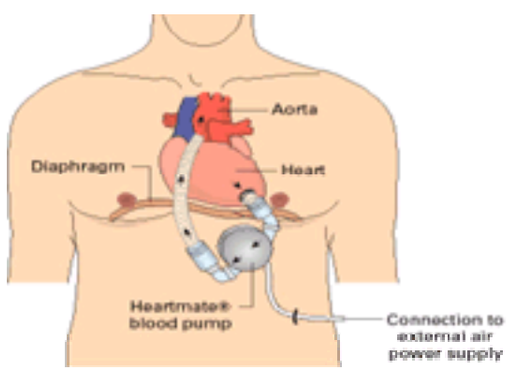


Fig. 10. Diagram of the HeartMate assist device.

The HeartMate is implanted through a median sternotomy that extends to the umbilicus. Cardiopulmonary bypass is instituted from the right atrium to the ascending aorta. The heart could be given cardioplegia after aortic cross-clamping or fibrillated without the need to cross-clamp. The pump is placed intra- or preperitoneally in the left upper quadrant. Its inflow and outflow conduits are attached to the device with ratable joints, so that they can be adjusted to fit the patient. The inflow cannula of the HeartMate is shorter than that of the Novacor.

The left ventricular apex is cored with a special circular knife. A Teflon-covered, reinforced Silastic sewing ring is then attached to the opening with interrupted, pledgeted sutures. The inflow and outflow conduits are connected to the blood pump. The inlet conduit is tunneled through an incision in the diaphragm, aligned with the mitral valve, then secured to the sewing ring. The outflow conduit, a polyester graft, is anastomosed end-to-side to the ascending aorta with a running 4–0 polypropylene suture. To ensure proper seating of the device, the outflow graft must be of an appropriate length; therefore, percutaneous lines are exteriorized before cutting the graft. A tunneled incision in the left lateral abdominal wall above the iliac crest allows passage of the pneumatic drive line. In the case of the vented electrical device, two stab incisions are made, one for the electric leads and the other for the vent tube. When the pump is properly seated, the graft can be cut so that it lies over the diaphragm, just to the right of the median sternotomy. Proper placement of the graft helps to prevent complications at the time of resternotomy for transplantation.

The omentum is wrapped around the intraperitoneal portion of the drive line to prevent bowel adhesion, and the pump is secured to the abdominal wall. Before pumping with a left ventricular assist device is initiated, the patient is placed in the Trendelenburg position. The outflow graft is cross-clamped near the anastomosis, and air is vented from the heart and the graft with a 19G needle. Pump flow is started and gradually increased, while cardiopulmonary bypass flow is gradually decreased. At this point, transesophageal echo-cardiography helps to confirm proper placement of the inflow graft. The effect of heparin is reversed, and before closure of the sternotomy and laparotomy incisions, thoracic drainage tubes are placed.

Preperitoneal implantation of the HeartMate requires the preparation of a small preperitoneal pocket behind the right rectus muscle. The device sits in a space between the posterior rectus sheath and the internal oblique fascia. The drive line is tunneled subcutaneously and exits the body in the left lower quadrant. Two large-bore suction drains (Jackson–Pratt drains; Baxter Healthcare) are placed in the most dependent aspect of the pocket.

Total artificial hearts

This provides complete support of the circulation, but requires removal of the native heart. Although the device has been used to provide permanent support, its most important role is to serve as a bridge to cardiac transplantation. Total artificial hearts are now undergoing experimental and clinical evaluation; these devices are the CardioWest 70 (Fig. 11), the Abiomed, the Cleveland Clinic–Nimbus (Fig. 12), the Baylor, and the Penn State.



Fig. 11. The CardioWest 70 total artificial heart.



Fig. 12. The Cleveland Clinic–Nimbus total artificial heart.

The first clinical use of an artificial heart was in 1969 by Cooley, when he implanted the Liotta Heart to support a patient for 64 h until a donor heart became available for transplantation. Use of the total artificial heart as a permanent device was initiated by DeVries in 1982, when he implanted a Jarvik 7 in a patient who survived 112 days. The first successful use of the Jarvik total artificial heart as a bridge for cardiac transplantation was by Copeland in 1985.

The CardioWest 70 (formerly known as Jarvik 70) is a pneumatically driven total artificial heart that is under investigation by the Food and Drug Administration to be utilized as a bridge to heart transplantation. The prosthetic ventricles are made of polyurethane and Medtronic-Hall mechanical valves provide unidirectional flow. Blood and air are separated by a four-layer, segmented polyurethane diaphragm that retracts during diastole and is displaced forward by compressed air during systole to propel blood out of the prosthetic ventricle. The device can provide flows of up to 15 l/min, but is usually used to provide flows at 6 to 8 l/min.

The controller on the device adjusts the heart rate, systolic duration, and drive-line pressures for each of the ventricles. The total artificial heart is operated so that there is incomplete filling but complete emptying with each stroke. The atrial pressure on each side determines ventricular filling; as atrial pressure increases, a higher stroke volume and cardiac output is obtained. The CardioWest total artificial heart has been used extensively in Europe and is undergoing clinical trials in selected centers in the United States as a bridge to transplantation.

Preparation of the artificial heart for implantation

Three steps are taken in preparation for implantation before giving heparin: the arterial grafts are prepared, the atrial cuffs are trimmed to appropriate size, and the drive lines are tunneled through the skin.

The grafts are first preclotted three times with the patient's blood before the heparin is given. After exposure to the blood, approximately 30 cc (ml) for each graft each time, they are stretched and left to dry for about 5 min and then preclotted again. Finally, after preclotting three times and the 5-min drying periods, the grafts are coated on the outside with biologic glue (cryoprecipitate with calcium and topical thrombin), placed once more in a stretched position, and allowed to dry. This is done early in the operation so there is plenty of time to obtain excellent preclotting of the grafts. If the patient has been heparinized before the decision to implant the total artificial heart is made, the arterial conduits are preclotted with a combination of heparinized blood, protamine, and thrombin. The ‘quick connects' of the atrial cuffs are trimmed. The edge of the ‘quick connect' for anastomosis to the atrium is cut to a radius of 5 to 7 mm extending out from the connect (Fig. 13). It is cut in a completely circular fashion. We then stretch and invert them.

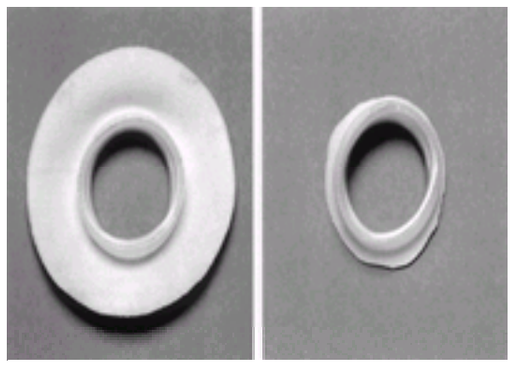


Fig. 13. (a, b) Atrial cuffs before and after being cut, leaving a 5- to 7-mm rim

The drive lines for the ventricles are passed through their subcutaneous pathways before the patient is heparinized. The left-sided ventricular drive line is positioned in the epigastrium at about the level of the midclavicular line and approximately 2 inches (5 cm) below the costal margin. A semicircular incision with a skin flap is made on the left upper quadrant. A long clamp is placed through the subcutaneous tissue, through the rectus fascias, rectus muscle, and into the chest. The pathway is enlarged somewhat by opening the clamp and then a 1 inch (2.5 cm) Penrose drain is passed through it. The end of the drive line is then placed in the Penrose drain and advanced approximately 4 to 5 inches (up to approximately 13 cm). The Penrose drains are pulled through the pathway, which delivers the drive line through the skin. The right drive line is placed at least 1½ to 2 inches (up to approximately 5 cm) away from the left so that no necrosis between the two exit sites will result. The

incision and passage of the line are made in a similar fashion. The ventricles are then positioned lateral to the wound and covered with a towel while the rest of the procedure takes place. This provides ample opportunity for small bleeders in the drive-line pathway to clot.

Removal of the recipient's heart

The aorta, superior and inferior vena cava are cannulated in a standard fashion. Umbilical-tape ‘chokers' are used on the cavae. Dissection around the aorta and pulmonary artery is limited to the proximal portion of the aorta in anticipation of later transplantation, thus leaving some untouched areas that will not be very fibrotic. Cardiopulmonary bypass is instituted and the heart is fibrillated. Total bypass is instituted by pulling on the choker tapes. The heart is fibrillated and its excision begun. The excision is significantly different for that used for transplantation. It seeks to preserve the annulus of both the tricuspid and mitral valves. Thus, an incision is made on the ventricular side of the atrioventricular groove of the right ventricle (Fig. 14); this can be done with a knife and extended with either a knife or scissors. It is extended anteriorly across the right ventricular outflow tract and just proximal to the pulmonary valve. Posteriorly it is extended to the interventricular septum and then across the septum, again staying on the left ventricular side of the atrioventricular groove and preserving the entirety of the mitral annulus. The anterior and posterior lines of the incision are dissected apart from each other out to the level of the pulmonary bifurcation. Finally, the excess muscle, on the right and left sides, is trimmed down to near the atrioventricular valves. Chordae are trimmed away and a 2-mm edge of valve tissue is left intact, along with the annulus. The atrial cuff generally extends 1 cm beyond the atrioventricular valve and consists of residual ventricular muscle and fat in the atrioventricular groove. All chordae are then trimmed away. The portion of the cuff in the left ventricular outflow tract consists of the residual anterior leaflet of the mitral valve and some aortic tissue. Most of the aortic tissue is trimmed away, but some is left intact since it is felt to present strong tissue for the sewing of the ‘quick connect' cuff (Fig. 15). The great vessels are separated from the remaining ventricular myocardium just above the valvular level, and then separated from each other.



Fig. 14. A ventriculectomy 1 cm below the level of the atrioventricular groove is necessary for adequate placement of the CardioWest total artificial heart.

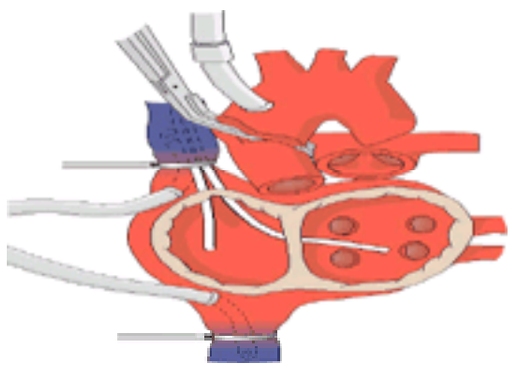


Fig. 15. Diagram of the ventricular rim with the atrioventricular valves removed; the aorta and pulmonary artery have been prepared for placement of the outflow conduits.

Next, the entrance of the coronary sinus into the right atrium is oversewn; this prevents backflow of blood through the coronary sinus and out to the cut vessels on the atrioventricular groove.

We then encircle the outer walls of the entire right and left atrial cuff complex with Teflon felt buttresses (Fig. 16), placed in such a way that they can be used for strengthening the anastomosis to the ‘quick connect', and also to tamponade and control all possible bleeding from the portion of the cuff at the atrioventricular groove. The buttresses are cut to approximately 10 to 12 mm in width and are generally 10 cm in length. It most often takes at least two of them to extend around the entire cuff. They are placed on the outer edge of the cuff and sewn in place with a running 3–0 polypropylene suture. A long (MH) needle is used to accomplish this and, on completion, the left and right atrial cuffs are surrounded.

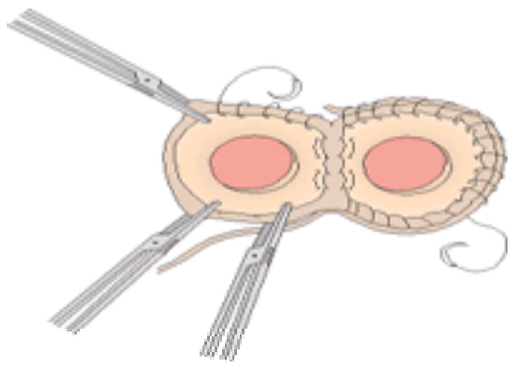


Fig. 16. Felt strips sutured around ventricular cuffs; atrial ‘quick connects' are being sutured in place.

Attachment of the grafts and atrial ‘quick connects', and testing for hemostasis

The atrial ‘quick connect' is sewn first, which is inverted, and placed inside the left atrial cuff on the lateral wall. A 3–0 polypropylene running stitch on an MH needle is used, taking care to tailor the atrial cuff and the ‘quick connect' into a single, hemostatic suture line. The line includes both the free wall of the atrium, which is buttressed with a Teflon felt, and the atrial septum, which has no buttressing material. After completing this entire suture line, a similar procedure is done with the right ‘quick connect'. Again, the connector is inverted and placed within the atrium, the suture line is run, and after completing both suture lines, i.e., for the left and right atrial cuffs, we then evert the ‘quick connects' so that they will be in the normal position.

Checking for hemostasis is now done with the plastic tester made to fit within the ‘quick connect'. A syringe [60–100 cc (ml) in volume] is used to inject into a three-way stop connected with the tester to test the left atrial suture line; the surgeon places a hand posterior to the left atrium and compresses the right and left pulmonary veins while the surgical assistant injects saline mixed with a small amount of blood into the left atrium (Fig. 17). A search for leaks is then made; if there are any, sutures are placed immediately. On the right side the cavae are already obstructed by the caval tapes, thus fluid is simply injected into the right atrium under pressure. Again, any leaks are immediately closed with a 3–0 MH polypropylene suture.

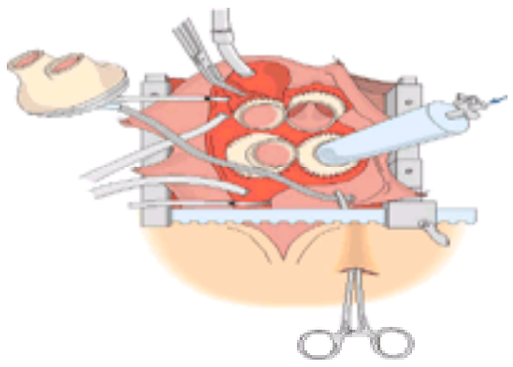


Fig. 17. Left atrial cuff being tested for anastomotic leaks: arterial conduits are already in place; it is easier to check the atrial conduits before anastomosing the arterial conduits.

Next the anastomosis of the great vessels is made, with that for the pulmonary artery first. The length of pulmonary artery present is that of the entire main vessel from just above the commissures of the valve. The length of the conduits for the great vessels is determined by placing the prosthetic ventricles in their position within the pericardial cavity and interposing the conduit between the aortic or pulmonic valve and its respective great vessel. The conduit is then cut to the appropriate length. An anastomosis is made end-to-end with running 4–0 polypropylene, beginning with the lateral wall and running to the back wall of the anastomosis from the inside. A similar anastomosis is made with the aortic suture line (Fig. 18). Following this, we use the aortic tester, which is inserted into the aortic ‘quick connect'. Saline is injected under pressure, leaks sought, and any found closed with 4–0 polypropylene. The pulmonary artery needs to be cross-clamped in order to test the integrity of its anastomosis to the conduit. The pulmonary arterial and aortic tester is the same but smaller than the one used for the atrial ‘quick connect'.

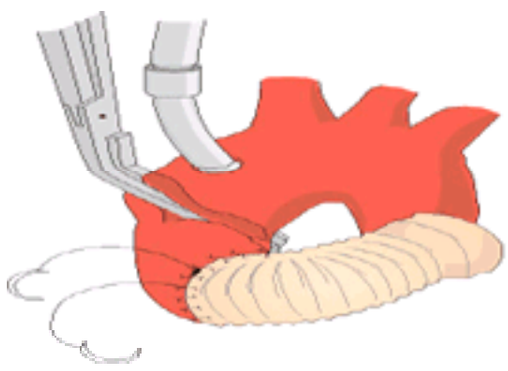


Fig. 18. Aortic conduit is anastomosed to the aorta with a running suture of 4–0 polypropylene.

Implantation of the ventricles and de-airing

Once adequate hemostasis of all suture lines has been established, placement of the ventricles is begun. First, the left ventricle is connected. The left atrial ‘quick connector' is grasped with two large Mayo clamps placed side by side, taking a good bite of the connector (Fig. 19). The opposite side of the plastic fitting for the connector of the device is placed within the connector and the operator pulls with the Mayo clamps and pushes the device into the ‘quick connect'; this takes some practice and should be done a number times in mock settings before it is actually attempted in the operating room. The position in which the heart sits for the duration of the support of the patient is determined by the orientation of the ventricle as it is placed into the atrial ‘quick connect'. Therefore, where exactly the aortic graft will connect to the left ventricle and the anticipated position of the left ventricle should be carefully assessed before the atrial ‘quick connect' is completed. It is then an easy matter to snap on the aortic connector, taking care not to twist the graft or aorta. While this is being done, the ventricle should be filled with saline through the aortic valve as well as the graft. Once the connection has been made, the patient is placed in a steep Trendelenburg position and large vent sites are placed in the highest point of the aortic graft and the aorta for the removal of air. The right ventricle is then connected. The atrial connection is made first, again taking care with the orientation of the ventricle so that the direction of flow from the outlet valve is appropriate for the anatomy of the patient. After the atrial connection has been made, the pulmonary graft connection is made again, taking care not to twist it. Before connecting the pulmonary graft the chokers on the superior and inferior vena cava should be removed; this allows a flow of blood into the right atrium and the right ventricle, and flushes air out as the connection to the pulmonic graft is made (Fig. 20).

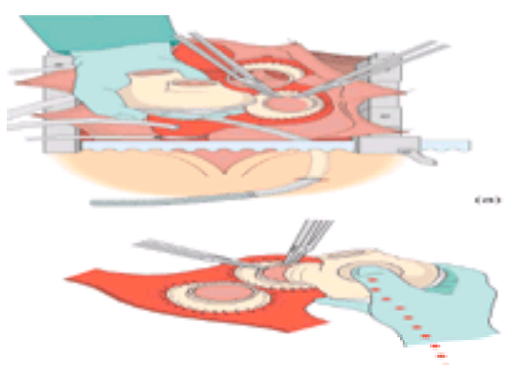


Fig. 19. (a) Two Mayo clamps are used to grasp the side of the atrial ‘quick connect' as the inflow aspect of the prosthetic ventricle is brought near. (b) The inflow connector is pushed into the atrial ‘quick connect' with steady pressure. The same maneuver is performed for the right prosthetic ventricle.

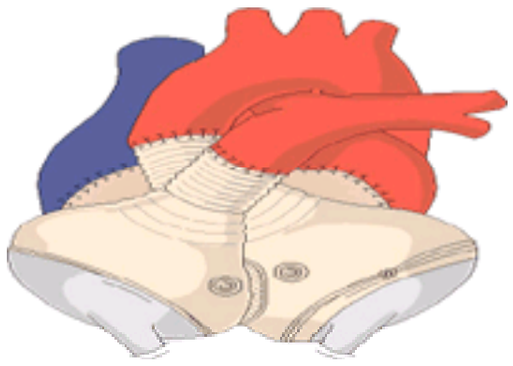


Fig. 20. The CardioWest total artificial heart connected to the respective atria and great vessels.

Then, with the patient in an extremely steep Trendelenburg position and with the lungs being slowly ventilated, pumping is begun at a very slow rate. The ventricles, as well as the atria, are agitated at this time. If available, transesophageal echocardiographic monitoring for air bubbles in the atria and aorta can help greatly in deciding when the device has been completely de-aired. As air is slowly removed from the device, the pumping rate and pressure are increased. Generally this process takes

about 10 min and should be done with patience and attention to complete removal of air before the artificial heart takes over from the heart–lung machine. During this time, decreasing the flow on the heart–lung machine temporarily may help move air through the lungs and into the device. Once the operator is satisfied that all air is out of the device, the vent sites may be closed and full pumping with the device begun as the patient is weaned off the heart–lung machine. The patient should be kept in steep Trendelenburg for an additional 15 to 20 min.

Following this, as the bed is flattened out, the initial attempts at positioning the ventricles within the mediastinum can be made. Optimally, the pleura on both sides should not be opened and the pericardium should be left intact for primary closure if possible or for closure with an interposed Gor-Tex membrane if not. In smaller patients there may be a need to force the right ventricle under the left edge of the sternum. When this is done, care should be taken to examine the left pulmonary veins and the inferior vena cava for evidence of compression; this check is facilitated by transesophageal echo-cardiography.

Great attention to hemostasis is necessary at this time. Optimally, these patients should be treated with aprotinin, 60 000 units per kilogram in total, given as follows: one-third as a bolus in the pump, one-third as a bolus by anesthesia, and one-third over a 4-h period by continuous infusion. A test dose of aprotinin should be given before these boluses are delivered. Aprotinin has helped greatly in obtaining hemostasis in these procedures. It is an absolute requirement that complete hemostasis be obtained before closing the chest. Bleeding cannot be permitted, since transfusion results in significant hemolysis and sensitization of these patients, who may, therefore, be disqualified for transplantation. Once absolute hemostasis has been obtained, chest tubes are placed in the mediastinum. We usually use a right-angle tube on the diaphragmatic surface and a straight tube that lies next to the right ventricle. The pericardium is closed or a Gor-Tex graft is used; this is not tight and it is placed to facilitate opening of the sternum at time of transplantation. Irrigation with copious amounts of vancomycin solution 1 g/l and Betadine solution, one ampoule per liter, is undertaken before closure. The sternum and remaining incision are closed routinely, taking care to observe cardiac output from the device as well as central venous pressure, and device filling when the chest is closed, since closure may alter the anatomy specifically causing pressure on the left-sided pulmonary veins, inferior vena cava, and occasionally the right-sided pulmonary veins. If decreased flow is noted, then the chest must be reopened and changes made in the position of the device. The most helpful change has often been to mobilize the diaphragmatic attachment of the pericardium, allowing the device to sit more leftward in the chest; this requires the left pleura to be opened, and in this case the lung itself should be protected from becoming adherent to the device by using pericardium or Gor-Tex membrane, since such adhesions may be extremely dense and difficult to deal with at the time of transplantation, and may, if they involve the lung, cause significant lung tears and bleeding.

Comment

Four total artificial hearts have currently been developed in the United States as permanent devices (see also above). The Abiomed total artificial heart is an electrohydraulically driven system capable of generating flows in excess of 10 l/min. Two blood pumps replace the failing ventricles. Each ventricle has an inlet and outlet trileaflet valves manufactured from polyurethane. This device has undergone chronic animal studies in excess of 60 days.

The Cleveland Clinic–Nimbus total artificial heart is another type of hydraulic drive system that replaces the ventricles. It can provide flows of more than 8 l/min at an afterload of 110 mmHg. It has undergone chronic animal studies up to 120 days, maintaining normal organ and cell function.

The Baylor total artificial heart is a completely implantable electromechanical device that can replace both ventricles. An electric motor provides rotational motion that is then converted to rectilinear motion in order to propel pusher plates. It can provide flows of more than 9 l/min. The Penn State Heart is an electrically driven total artificial heart replacement for permanent use.

Anticoagulation

Protocols for anticoagulation differ among institutions using mechanical assist devices, mainly because no single regimen seems to be more effective than the others. Many agents have been used at different times after implantation and in varying combinations. These agents include low molecular-weight dextran, heparin, warfarin, aspirin, and dipyridamole.

A typical regimen at our institution includes the use of dextran at 25 to 50 ml/h and dipyridamole at 100 mg every 6 h on the first postoperative day. Heparin by continuous intravenous infusion to maintain the partial thromboplastin time at 50 to 60 s is started when postoperative bleeding has been controlled, usually on the second or third postoperative day. At the time heparin is started, dextran is discontinued. The patient is maintained on heparin and dipyridamole unless a relatively lengthy implant time is anticipated. In this case, warfarin is favored over heparin, with maintenance of the prothrombin time at approximately 18 to 22 s (INR 2.5 to 3.5.) If there is clinical evidence of thromboembolism, aspirin is added at a dose of 81 to 325 mg/day.

More recently, for patients implanted with total artificial hearts, we have modified our regimen to include the use of other agents such as ticlopidine, 250 mg every other day, and pentoxifylline, 1200 mg/day, in combination with heparin, warfarin, aspirin, and dipyridamole to stabilize platelet function and to balance the different coagulation pathways. In order to monitor the effectiveness of anticoagulation therapy, in addition to the usual prothrombin and partial thromboplastin times, we examine factor X, fibrinogen, fibrin degradation products, platelet count, bleeding time, and *in vitro* platelet agregatability in response to four different stimulators [adenosine, epinephrine (adrenaline), collagen, and arachidonic acid). Antithrombin III in plasma and serum is used to calculate the antithrombotic potential index relating the amount of antithrombin III available to interact against thrombin. The use of thromboelastography, probably the best available indicator of overall coagulation mechanisms, and the calculation of the thrombodynamic potential index, give a measurement of the coagulability of blood by examining the rapid phase of clot formation.

Complications

Common complications encountered with the use of circulatory assist devices include bleeding, thromboembolism, infection, hemolysis, and multiorgan failure.

Bleeding

Bleeding can be significant at the time of placement of a circulatory assist device, particularly if cardiopulmonary bypass has been used for a prolonged period of time. Activation of platelets and the coagulation system by exposure of the blood to artificial surfaces and turbulent flow patterns may lead to a severe coagulopathy. Control of bleeding from suture lines, conduits, and raw surfaces as a result of deficiencies in the coagulation system can be a significant problem. The bleeding time becomes greater than 30 min after 2 h of cardiopulmonary bypass, owing to platelet dysfunction. Depending upon the extent of platelet damage, platelet function may be normalized within 30 min after stopping the cardiopulmonary bypass. A prolonged bleeding time with a platelet count above 100 000 indicates platelet dysfunction that may require exogenous platelet administration.

Desmopressin, a synthetic analog of the hormone vasopressin, has been useful in augmenting platelet function. Its mode of action is mediated by the factor VIII complex and, therefore, adequate amounts of factor VIII are required. A decrease in bleeding time may be measured at 30 to 90 min following administration.

Bleeding may also become a significant problem at the time of device explantation. Dense adhesions between the device and the intrathoracic organs may cause diffuse hemorrhage. The recent introduction of aprotinin into the pharmacologic armamentarium during implantation and/or explantation of devices might decrease bleeding significantly and increase survival.

Thromboembolism

The formation of macroscopic thrombus in circulatory assist devices is commonly seen at explantation even when the patient has had ‘adequate anticoagulation'. Many of the devices have internal sites that may provide a nidus for the formation of thrombus. Other factors that appear to be involved in the development of thrombus within assist devices are a low or reduced blood flow and infection. The low-flow state and the contact between blood and an artificial surface accelerates thrombus formation. Infection can lead to a hyper-coagulable state, with activation of inflammatory cells and mediators leading to the formation of thrombus.

Embolic complications can occur at any time. The brain is by far the most sensitive and most reported organ affected by thromboembolism. However, irreversible neurologic damage is uncommon; transient ischemic attacks are more frequent. Embolism to the kidneys, ophthalmic artery, lungs, and heart via the coronary arteries have also been documented.

Infection

Patients who need circulatory assistance are at risk for developing infections not only because of the presence of a large foreign body, but also because they are often debilitated and malnourished. Prophylactic antibiotics are routinely used during the perioperative period. The longer the device remains implanted, the higher the incidence of infection. Thrombosis, thromboembolism, and infection are more common beyond 30 days of support and these complications may be interrelated.

Pneumonia and mediastinitis are the most prevalent infections encountered. Other types include infection in and around the device involving not only the device itself and its inflow and outflow conduits, but also the drive lines. These infections require systemic antibiotics and surgical debridement.

Hemolysis

Hemolysis is often seen in patients who have been on cardiopulmonary bypass for an extended period of time. Chronic hemolysis can be associated with anemia requiring transfusions, which may sensitize a potential transplant patient to tissue antigens, making them a less compatible potential recipient. It may also be associated with chronic renal failure. If used as recommended, all currently available devices do not produce excessive hemolysis.

Multiorgan failure

Many patients are in some degree of cardiogenic shock preoperatively; this may persist postoperatively if the assist device fails to produce the desired increase in cardiac output. Renal failure, associated with azotemia and abnormal free-water clearance, has a high mortality. Managing renal failure with hemodialysis, ultrafiltration, peritoneal dialysis, or a combination of these techniques usually does not have a significant impact on the outcome. Pulmonary failure is usually manifested by hypoxemia, increased intrapulmonary shunting, and decreased compliance. Jaundice and elevations in the liver enzymes are the most common evidence of liver failure, although an elevated bilirubin may represent blood transfusion-related hemolysis. Gastrointestinal failure can be manifested by bleeding from the gastrointestinal tract and/or prolonged ileus. The outcome for patients with multiorgan failure is directly related to the number of organs involved; the more organs involved in failure, the higher the mortality.

Technical tips

- Use aprotinin when excessive bleeding is suspected/predicted
- Close the pericardium whenever possible
- Drain the pericardial space
- Use polytetrafluoroethylene membrane to wrap inflow and outflow cannulas
- Use pledgets for outflow-conduit anastomosis

Future directions

The circulatory assist devices represents significant medical and engineering achievements. In the near future, pneumatically driven assist devices may be replaced by electromechanically or electromagnetically driven units, flows may be either pulsatile or nonpulsatile, moving diaphragms may be replaced by high-speed turbines; and power supplies may be internal, external, or transfer energy across the skin. The advent of this new technology may one day change the management of heart disease.

Further reading

Arabia FA, Smith RG, Rose DS, Arzouman DA, Sethi GK, Copeland JG. Success rates of long term circulatory assist devices currently used for bridge to heart transplant. *American Society for Artificial Internal Organs Journal* 1996; **42**(5): M542–6.

Arabia FA *et al.* Cost analysis of the Novacor LVAS as an outpatient bridge to heart transplant. *American Society for Artificial Internal Organs Journal* 1996; **42**(5): M546–9.

Bolman RM *et al.* Circulatory support with a centrifugal pump as a bridge to cardiac transplantation. *Annals of Thoracic Surgery* 1989; **47**: 408–12.

Burnett CM *et al.* Improved survival after Hemopump insertion in patients experiencing postcardiotomy cardiogenic shock during cardiopulmonary bypass. *American Society for Artificial Internal Organs Transactions* 1990; **36**: M626–9.

Burton NA *et al.* A reliable bridge to cardiac transplantation: the TCI left ventricular assist device. *Annals of Thoracic Surgery* 1993; **55**: 1425–31.

Butler KC, Moise JC, Wampler RK. The Hemopump: a new cardiac prosthesis device. *IEEE Transactions on Biomedical Engineering* 1990; **37**: 193–6.

Champsaur G *et al.* Use of the Aviomed BVS 5000 as a bridge to cardiac transplantation. *Journal of Thoracic and Cardiovascular Surgery* 1990; **100**: 122–8.

Copeland JG *et al.* Bleeding and anticoagulation. *Annals of Thoracic Surgery* 1989; **47**: 88–95.

Copeland JG *et al.* Orthotopic total artificial heart bridge to transplantation: preliminary results. *Journal of Heart Transplantation* 1989; **8**: 124–38.

Copeland JG *et al.* Bridge to transplantation. Indications for Symbion TAH, Symbion AVAD. In: Akutsu T, Koyanagi H, ed. (Proceedings) *The 3rd international symposium on artificial heart and assist devices*, pp. 303–8. Springer, Tokyo, 1991.

Deeb GM *et al.* Clinical experience with the nimbus pump. *American Society for Artificial Internal Organs Transactions* 1990; 36: M632–6.

Drinkwater DC, Laks H. Clinical experience with centrifugal pump ventricular support at UCLA Medical Center. *American Society for Artificial Internal Organs Transactions* 1988; **34**: 505–8.

Farrar DJ *et al.* Heterotopic prosthetic ventricles as a bridge to cardiac transplantation. A multicenter study in 29 patients. *New England Journal of Medicine* 1988; **318**: 333–40.

Farrar DJ *et al.* Thoratec VAD system as a bridge to heart transplantation. *Journal of Heart Transplantation* 1990; **9**: 415–23.

Griffith BP. Interim use of the Jarvik-7 Artificial Heart: lessons learned at Presbyterian University Hospital of Pittsburgh. *Annals of Thoracic Surgery* 1989; **47**: 158–66.

Guyton RA *et al.* Postcardiotomy shock: clinical evaluation of the BVS 5000 biventricular support system. *Annals of Thoracic Surgery* 1993; **56**: 346–56.

Harasaki H *et al.* Progress in Cleveland Clinic–Nimbus total artificial heart development. *American Society for Artificial Internal Organs Journal* 1994; **40**: M494–8.

Hill JD. Bridging to cardiac transplantation. *Annals of Thoracic Surgery* 1989; **47**: 167–71.

Hooper TL *et al.* Successful use of the left ventricular assist device for primary graft failure after heart transplantation. *Journal of Heart Transplantation* 1988; **7**: 385–7.

Joyce LD *et al.* Summary of the world experience with clinical use of total artificial heart as heart support devices. *Journal of Heart Transplantation* 1986; **5**: 229–35.

Kaan GL *et al.* Management of postcardiotomy cardiogenic shock with a new pulsatile ventricular assist device. *American Society for Artificial Internal Organs Transactions* 1991; **37**: 559–63.

Kaye MP. The registry of the International Society for Heart and Lung Transplantation: tenth official report—1993. *Journal of Heart and Lung Transplantation* 1993; **4**: 541–8.

Kormos RL *et al.* Experience with univentricular support in mortally ill cardiac transplant candidates. *Annals of Thoracic Surgery* 1990; **49**: 261–72.

Lick S *et al.* Use of the Symbion biventricular assist device in bridging to transplantation. *Annals of Thoracic Surgery* 1993; **55**: 283–7.

McCarthy PM, Wang N, Vargo, R. Preperitoneal insertion of the HeartMate 10001P implantable left ventricular assist device. *Annals of Thoracic Surgery* 1994; **57**: 634–8.

Muneretto C *et al.* Total artificial heart: survival and complications. *Annals of Thoracic Surgery* 1989; **47**: 151–7.

Orime Y *et al.* The Baylor total artificial heart. *American Society for Artificial Internal Organs Journal* 1994; **40**: M499–505.

Ott RA *et al.* Clinical choices for circulatory assist devices. *American Society for Artificial Internal Organs Transactions* 1990; **36**: 792–8.

Ott RA, Joyce L, Emery RW. Current status of mechanical circulatory assistance. *Bulletin of the Minneapolis Heart Institute Foundation* 1988; **6**: 13–18.

Oz MC, Goldstein DJ, Rose EA. Preperitoneal placement of ventricular assist devices: an illustrated stepwise approach. *Journal of Cardiac Surgery* 1995; **10**: 288–94.

Pae WE *et al.* Long term results of ventricular assist pumping in postcardiotomy cardiogenic shock. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 434–41.

Parnis SM *et al.* Chronic *in vivo* evaluation of an electrohydraulic total artificial heart. *American Society for Artificial Internal Organs Journal* 1994; **40**: M489–93.

Pennington DG *et al.* Use of the Pierce–Donachy ventricular assist device in patients with cardiogenic shock after cardiac operations. *Annals of Thoracic Surgery* 1989; **47**: 130–5.

Pennington DG *et al.* Circulatory support 1988: patient selection. *Annals of Thoracic Surgery* 1989; **47**: 77–81.

Pennington DG *et al.* Univentricular versus biventricular assist device support. *Journal of Heart and Lung Transplantation* 1991; **10**: 228–63.

Phillips SJ *et al.* Hemopump support for the failing heart. *American Society for Artificial Internal Organs Transactions* 1990; **36**: M629–32.

Portner PM *et al.* Implantable electrical left ventricular assist system: bridge to transplantation and the future. *Annals of Thoracic Surgery* 1989; **47**: 142–50.

Rosado LJ, Copeland JG. Orthotopic heart transplantation: recent advances. *Primary Cardiology* 1990; **16**: 33–47.

Rose DM *et al.* Technique and results with a roller pump left and right heart assist device. *Annals of Thoracic Surgery* 1989; **47**: 130–5.

Szefner J, Cabrol C. Control and treatment of hemostasis in patients with a total artificial heart: the experience of La Pitie. In: Pifarre P, ed. *Anticoagulation, hemostasis, and blood preservation in cardiovascular surgery*, pp. 237–64. Hanley I. Belfus, Philadelphia, 1993..

40.4 Vascular and cardiac biological and non-biological prostheses

Graeme L. Hammond and Michael A. Coady

[Vascular prostheses](#)
[Biologic grafts](#)
[Autogenous grafts](#)
[Allografts (homografts)](#)
[Heterografts (xenografts)](#)
[Synthetic grafts](#)
[Textile grafts](#)
[Non-textile grafts](#)
[New developments with prosthetic grafts](#)
[Bioresorbable grafts](#)
[Gene therapy](#)
[Antithrombogenic coatings](#)
[Endothelial seeding](#)
[Antibiotic bonding](#)
[Vascular stents](#)
[Self-expanding stents](#)
[Clinical applications](#)
[Cardiac prostheses](#)
[Bioprosthetic valves](#)
[Mechanical valves](#)
[Further reading](#)

The development of vascular prostheses in the early 1950s and of cardiac prostheses in the early 1960s extended surgical treatment to diseases that had been previously beyond this field, and set the stage for the development of artificial assist devices that are now at the forefront of surgical research. Each year in the United States, there are approximately 1.4 million surgical procedures performed that require arterial prostheses. Although room for improvements in materials and design remains, the clinical use of vascular and valvular prostheses has reduced the mortality associated with ruptured aortic aneurysm from 100 per cent to 30 per cent, and that of symptomatic aortic stenosis from 100 per cent to 5 per cent.

Vascular prostheses

A prosthesis may be defined as an apparatus meant to replace an injured or absent body part in order to restore necessary function. In cardiovascular surgery, a vascular prosthetic graft functions to replace the arterial segment compromised by atherosclerotic occlusive disease, trauma, or aneurysmal degenerative disease. These devices may consist of a biologic or synthetic graft, or vascular stent, allowing the affected segment to be bypassed, excised, reopened, or repaired. Table 1 displays a classification of vascular grafts and stents according to their method of construction.

Biologic		*Synthetic*		*Stents*
Autografts		Textile		Balloon-expandable
	Arterial		Woven Dacron	
	Venous		Knitted Dacron	
Allografts		Non-textile		Self-expanding
	Arterial		Teflon (ePTFE)	
	Venous		Polyurethane	
	Umbilical vein		resorbable	
Xenografts				Stent-grafts
	Bovine carotid			
	Canine carotid			

Table 1 Classification of prosthetic grafts and stents

There are several considerations when selecting a prosthetic graft. The ideal prosthesis should be biocompatible with the host, physically durable, resistant to infection, non-reactive, non-thrombogenic, and easily implantable. In addition, it should be available in a variety of sizes and lengths. Currently no prosthetic material completely fulfills all of these expectations, however, a number of acceptable alternatives, particularly for larger vessels, are available.

Biologic grafts

Autogenous grafts

Arterial autografts

Arterial autografts were first introduced in 1964 for vascular replacement.These grafts serve as excellent arterial substitutes for several reasons. They have an intrinsic blood supply, do not degenerate with time, and have the ability to heal in an infected field. In children, they also demonstrate proportional arterial growth.

Although an arterial autograft is an ideal arterial substitute, size and availability preclude its use in most circumstances. The most common arterial autograft used is the internal mammary during coronary artery bypass surgery. It has a patency rate superior to that of saphenous vein grafts, with an overall improvement in overall patient survival. Autografts from other locations, however, may be procured without the need for replacement. Fibromuscular dysplastic lesions of the renal arteries may be treated with a hypogastric artery autograft, as the vessel is of a similar diameter. In addition, carotid, popliteal, and femoral aneurysms are ideal sites for arterial autografts because the segments are short and the diameter approximates that of the external iliac artery.

Arterial autografts also may be constructed from an arterial segment previously occluded by atherosclerosis. The thrombosed superficial femoral artery has been used for this purpose following endarterectomy. This graft can be used as a patch, bypass, or may be used after removal of an infected synthetic graft. Furthermore, arterial autografts may be used in patients with anastomotic false aneurysms as interposition grafts to repair the defect after excision of the aneurysm.

Venous autografts

The greater saphenous vein serves as the conduit of choice for revascularization of small-calibre vessels. It is the longest vein in the body, readily accessible for use in coronary bypass and peripheral vascular surgery. Alternate autogenous venous grafts include the lesser saphenous, cephalic and basilic, the internal jugular, and the superficial femoral vein.

Various factors influence the long-term patency of vein grafts. Vein graft failure may result from both stenotic and aneurysmal disease. Stenosis results from intimal hyperplasia, operative trauma, valve fibrosis, and atherosclerosis. Early graft occlusion (<30 days) is primarily related to technical errors or thrombotic complications. Intermediate graft occlusion (<2 years) occurs secondary to neointimal hyperplasia or vein stenosis. Intimal hyperplasia is an adaptive response of the vein to the high pressure of arterial flow, and is normally a self-limiting process, which reaches a steady-state within 2 years; however in focal areas it can lead to significant stenosis. Late stenosis (<2 years) occurs from progression of atheroma and aneurysmal deterioration of the graft.

Recently identified risk factors for infrainguinal vein graft stenosis include a significant smoking history, and high systemic levels of plasma fibrinogen, Lipoprotein (a), and 5-hydroxytryptamine. Accelerated atherosclerosis occurs in veins that have become arterial conduits. Graft arteriosclerosis occurring within 3 to 5 years after

surgery has been reported in 15 per cent of femoropopliteal vein grafts and in 7 per cent of aortocoronary vein grafts.

Allografts (homografts)

Arterial allografts.

Fresh or preserved arterial allografts were introduced as potential vascular conduits in the 1940s because of the lack of acceptable synthetic vascular conduits. Problems that occur with arterial allografts include immediate rejection, viral transmission, and occlusion (fresh allografts), as well as severe calcification, early rupture, late occlusion, and significant aneurysmal dilatation (preserved allografts). These allografts may be preserved both by irradiation and freeze drying.

Venous allografts

Venous allografts were studied in the early 1960s and 1970s when it became recognized that autogenous saphenous vein was often absent or inadequate. Techniques have been developed with cryopreservation with dimethylsulfoxide (DSMO) and liquid nitrogen to reduce host immunologic responses; however more recent results have been discouraging.

Experience with allografts is still experimental at this stage, as immunologic responses to the foreign tissue and vessel wall degenerative changes pose continued problems. Long-term assessment is needed before before expanding the range of application.

Umbilical vein grafts

Dardik *et al.* introduced the glutaraldehyde-tanned human umbilical vein graft in the mid-1970s. The Biograft (Biovascular, Inc.) is rendered non-antigenic secondary to cross-linking of the amino groups of the polypeptide collagen chains. While some centers have shown that umbilical vein grafts have equivalent patency rates for above-knee femoral–popliteal bypassess compared to saphenous, most studies have shown that saphenous vein grafts are superior below the knee. Although there has been some debate, no significant difference in patency rates between ePTFE and umbilical vein has been demonstrated in the literature. The major problem with umbilical vein grafts appears to be to be aneurysm formation, reported as high as 65 per cent after 5 years. Recent improvements in the processing of umbilical vein grafts have resulted in less aneurysm formation (17 per cent over a 6-year period); however, this remains a continued disadvantage.

Heterografts (xenografts)

Bovine and canine carotid xenografts

Grafts constructed from other animal species may demonstrate considerable limitations secondary to immune rejection, despite glutaraldehyde tanning. Additional limitations include the risk of viral transmission, aneurysm formation, and infection. Although minimal experience with the bovine heterograft is available, experimental results of canine carotid xenografts have been limited to the laboratory.

Synthetic grafts

The most commonly used synthetic grafts are made primarily from one of two materials: Dacron (polyester) or ePTFE. Both textile and non-textile grafts are used.

Textile grafts

Dacron is a multi-filament yarn consisting of small filaments that are given a particular texture to make the graft soft and elastic. The yarn can be made into a prosthetic graft by weaving, knitting, or braiding. In woven grafts the fabric is interlaced in an over-and-under pattern. They have little or no stretch in any direction, are tightly constructed, of low porosity, and relatively strong and unyielding. They may be more difficult to sew because of the small interstices and relative stiffness, but tend to minimize or eliminate bleeding through the graft.

Woven grafts are designed specifically for use in patients who require systemic heparinization and are thus ideally suited for replacement of the thoracic aorta. Graft stiffness is usually not a problem when reconstructing the aorta after resection of fusiform or saccular aneurysms, as the resection margins tend to be thick, fibrous, and easily incorporate sutures. A serious dilemma arises, however, with thoracic aortic dissections. In this case, the aortic tissues are extremely thin, tenuous, and friable. Attempts to sew a stiff, non-compliant graft into these tissues are often fraught with further tearing, sometimes to the point of avulsion of a suture line. On the other hand, low porosity is a requirement as patients undergoing surgery for type A dissections are fully heparinized, on cardiopulmonary bypass and often develop severe preoperative or intraoperative coagulopathies. A satisfactory compromise has been developed by Meadox in the low porosity Veri-Soft woven graft. In this case, the Dacron yarn is designed to be soft enough to allow good flexibility of the graft, but yet the weave is tight enough to prevent hemorrhage through the interstices. We have been very satisfied with the Veri-Soft graft as a replacement for the ascending aorta in type A thoracic dissections. At our institution, however, we also toughen the aorta with glutaraldehyde to help ensure the stability of the aortic tissues before performing the anastomosis.

Knitted Dacron grafts have primarily a longitudinal (warp) or circumferential (weft) orientation. The wrap configuration is most commonly preferred because of its greater stability. Knitted grafts have a greater porosity, with enhanced healing, excellent handling, and are less likely to fray. These grafts need to be preclotted to seal the interstices of the porous fabric and to decrease bleeding. In addition, the graft is inherently weaker, which may lead to gradual dilatation. Knitted grafts are designed for abdominal and peripheral vascular procedures. In general, they are easily sewn to blood vessels because of their soft, compliant nature. This is particularly helpful when the native vessel is heavily calcified resulting in irregular vascular margins, needle blunting, or deflection of the suture path. The compliance of the graft allows a snug fit despite the appearance of the suture line and minimizes bleeding at anastomotic sites. However, after removal of the clamps, bleeding, that can sometimes be considerable, will occur through the graft until clot forms in the interstices. Use of these grafts is contraindicated in the heparinized patient due to the likelihood of uncontrollable hemorrhage. A modification of the standard knitted graft is the double velour graft (Meadox) that incorporates a velour inner and outer pile to enhance the incorporation of clots in the graft interstices. Dacron grafts of less than 5 mm in diameter should not be used due to their high incidence of thrombosis. This observation precludes the use of Dacron grafts for obstructive disease below the knee. For grafts in the suprainguinal region, however, Dacron remains the material of choice.

It was originally thought that a high porosity (knitted) graft was necessary to allow endothelial cells and their subendothelial matrix to invade and adhere to the inner graft surface. This mechanism, it was reasoned, was necessary to reduce thrombosis and therefore increase patency. Although prosthetic vascular grafts may develop a neo-endothelium in canine and non-human primate models, the experience with humans has shown that re-endothelialization occurs only sporadically or not at all. In fact, once blood courses through a synthetic prosthetic graft, the luminal surface is coated with a layer of protein. This layer, consisting primarily of fibrinogen also attracts and activates platelets and various coagulation proteins. Thus, the graft becomes lined with a neointima consisting of a thin layer of proteinaceous material and fibrin. Therefore, for practical purposes, re-endothelialization does not occur. To test the effect of woven versus knitted grafts on patency, an imaginative study was carried out by Robicszek *et al.* who designed bifurcation grafts in which one limb of the graft was knitted and one limb was woven. The grafts were implanted in 143 consecutive patients with aortoiliac atherosclerosis or aneurysms. Analysis failed to reveal any difference in patency rates between the woven and knitted limbs during the observation period that ranged from 1 month to 2 years.

Graft impregnation with collagen, albumin, or gelatin is being increasingly used in the production of Dacron grafts. These modifications are designed to reduce the potential for graft leak. After the graft is placed, the collagen undergoes biodegradation, thus allowing normal healing. These grafts are extremely useful in the heparinized patient. However, in the non-heparinized patient, there has been no evidence of reduced total blood loss or improved patency. All impregnated grafts are currently significantly more expensive than standard synthetic grafts.

Non-textile grafts

Expanded polytetrafluoroethylene grafts

Although Dacron grafts performed well in the aortoiliac region, the poor long-term performance of these grafts in the infra-inguinal region motivated a search for new prosthetic materials.

The advantage of ePTFE grafts include the fact that they do not need to be pre-clotted, are biocompatible, resistant to dilatation, and do not leak through the substance of the graft. The smoothness of ePTFE may have several advantages: they may be more resistant to infection due to resistance of the graft wall to bacterial adherence, and they may be easier to thrombectomize than Dacron conduits. The smooth surface may inhibit the adherence of platelets and fibrinous material to the conduit

surface, thereby reducing thrombogenicity and allowing the use of smaller diameter grafts for grafting peripheral vessels. ePTFE is impervious to blood but is not resistant to fibroblast infiltration at anastomotic sites.

The disadvantage of ePTFE grafts is that they are non-compliant, and bleeding through suture holes and gaps between sutures is common. Because of this deficiency, the use of ePTFE material, in the systemically heparinized patient or in those with active or potential coagulopathies, is not recommended. Most ePTFE grafts are not pleated, as are Dacron grafts, and therefore tend to kink more easily.

The superiority of ePTFE versus Dacron regarding long-term patency has received mixed reviews. There is no difference in the development of initial hyperplasia at anastomotic sites or in the patency of aortic bifurcation grafts made from the two materials. However, there appears to be a somewhat higher patency rate for ePTFE grafts when used in the axillofemoral position or in the femoral–popliteal–tibial position. Neither ePTFE or Dacron is comparable in patency with reversed saphenous vein.

Polyurethane grafts

Polyurethane grafts initially promised durability, low thrombogenicity, compliance, and high patency because of its smooth lining and elastic, thin outer wall. Unfortunately, these grafts have exhibited a disappointing high rate of degeneration and aneurysm formation.

New developments with prosthetic grafts

Bioresorbable grafts

Tissue engineering is a new and rapidly expanding field, in which techniques are being developed for culturing a variety of tissues both *in vitro* and *in vivo* using polymer ‘scaffolds' to support tissue growth. Bioresorbable grafts exhibit cell-adhesion properties which enable cell attachment and tissue ingrowth. The native vessel regenerates over a scaffold of absorbable material. Both smooth muscle cells and endothelial cells have been successfully cultured over the biodegradable material. Studies involving absorbable woven polyglactin acid (Vicryl)prosthesis have demonstrated replacement by endothelialized vessels containing myofibroblasts and fibroplasia. While autologous venous or arterial vessels are generally used, not all patients possess adequate conduit for revascularization.

Preliminary studies have achieved good cell densities of both smooth muscle cells and endothelial cells on biodegradable polymer scaffolds. Biodegradable grafts remain a tantalizing option for future bypass strategies; however, at present they are still very much in the early experimental stages.

Gene therapy

Recent developments in gene transfer techniques have emerged as possible new therapeutic options in the treatment of vascular diseases. Studies have established the feasibility of direct *in vivo* gene transfer into the vasculature by reporter genes such as B-galactosidase or luciferase. After endothelial denudation, von der Leyen *et al.* recently restored endothelial cell nitric oxide synthase expression in the vessel wall by using a Sendai virus/liposome *in vivo* gene transfer technique. This restored nitric oxide production and inhibited neointima formation by 70 per cent. This suggests that endothelial cell nitric oxide synthase transfection may be a potential therapeutic approach in treating neointimal hyperplasia. In addition, Mann *et al.* demonstrated that an intraoperative gene therapy approach using anti-sense oligonucleotide blockage of medial smooth muscle proliferation can prevent the accelerated atherosclerosis that is responsible for autologous vein graft failure. Gene therapy may play a significant part in the treatment of vascular diseases in the future.

Antithrombogenic coatings

Recent studies have shown that carbon-lined ePTFE grafts decrease platelet accumulation and thus decreases potential thrombogenicity. Preliminary studies with carbon-impregnated grafts tend to have better short-term patency than the standard PTFE graft.

Ritter *et al.* in 1998 demonstrated, that if air was removed from small diameter ePTFE grafts, the acute thrombogenicity was decreased. Heparin bonding further decreased the thrombogenicity in this study. Other studies have shown that heparin bonding of ePTFE grafts decreases surface thrombogenicity. Coupling heparin through a surface bound substrate to porcine pericardial bioprosthesis has also shown prevention of calcific degeneration in preliminary studies.

Endothelial seeding

As mentioned above, humans lack the ability to endothelialize the intraluminal surface of vascular conduits. Indeed, the improved patency reported for venous conduits compared with synthetic grafts may be attributable to an intact endothelial surface. Endothelial cells have been cultivated from vessels and have been grown in culture. These cells may then be incubated with a synthetic or biologic graft (human umbilical vein) to line the graft with an endothelium.

This technology, however, has not been as successful as initially hoped. Differences in seeding methods may contribute to the failure. Recently, however, there have been encouraging results with improved seeding technique. These techniques, although presently quite tedious, hold promise, especially if gene therapy is employed in combination with endothelial seeding.

Antibiotic bonding

Infection-resistant ePTFE grafts are of interest because of their potential clinical use in sites of frank bacterial contamination, and in the replacement of infected grafts. A number of antibiotics have been bonded to ePTFE by a variety of methods. Initial results have been encouraging.

Vascular stents

Vascular stents were initially designed to complement balloon angioplasty by preventing acute vessel closure from elastic recoil, distal embolization, and plaque dissection.

Several different stents have been developed, including self-expanding and balloon-expandable varieties. These stents currently enable a percutaneous deployment and offer a low metal mass to reduce thrombogenicity and intimal hyperplasia (Fig. 1, Fig. 2).

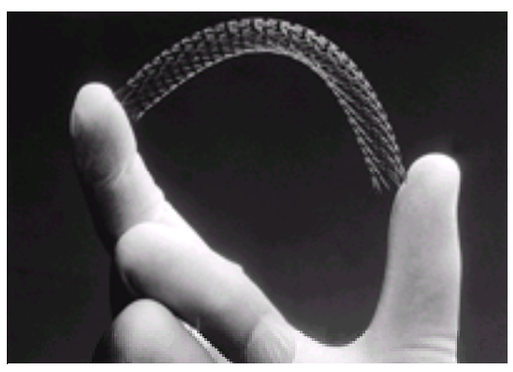


Fig. 1. Memotherm stent. Nitinol shape memory permits the stent to ‘remember' its original shape once deployed in tortuous anatomic situations and complex lesions (by courtesy of C.R. Bard, Inc.)

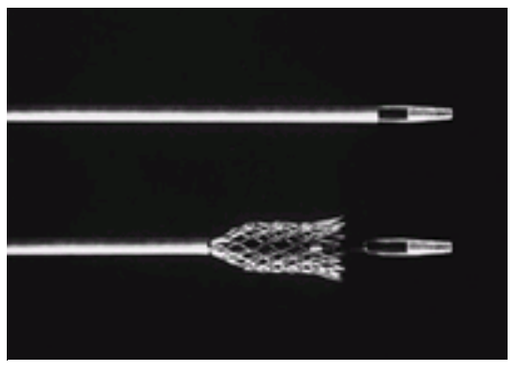


Fig. 2. Memotherm stent and stent applicator (By courtesy of Angiomed).

Self-expanding stents

The Wallstent (Shneider) is made of stainless steel filaments woven in a criss-cross tubular pattern. This stent is highly flexible in the longitudinal direction. The Wallstent is used clinically in the coronary and peripheral arteries. This stent is compressed and contained on a delivery catheter. During deployment, the outer membrane of the Wallstent is retracted, allowing the stent to expand as it is released. The expanded size is 20 to 40 per cent shorter than its compressed state. The overall radial force is somewhat weak, and additional balloon dilatation may accelerate the rate of expansion, allowing the stent to reach its maximum diameter.

The Memotherm stent (Fig. 1) is a self-expanding stent made of Nitinol, a nickel–titanium alloy which gives it flexibility as well as radial strength. It has a diamond-shaped design which minimizes shortening of the stent upon deployment unlike other expandable stents which may shorten by as much as 60 per cent once deployed. It is available in a variety of lengths (30 to 110 mm) and diameters (7 to 12 mm).

Clinical applications

Aortic stenting

Endovascular stent grafting for the treatment of descending thoracic aortic aneurysms is an alternative technique that is less invasive, potentially less expensive, and associated with less morbidity than open repair. Although these techniques are new, they hold promise as future methods for treatment of thoracic aortic pathology in very high-risk patients. Patients likely to benefit from endovascular stenting include elderly individuals, those patients with cardiac, pulmonary, or renal failure, and those who have undergone prior surgery on the thoracic aorta.

Stanford University is leading the way in this new technology. For endovascular stent-grafting of aortic aneurysms, their recent series reports an early mortality rate of 9 ± 3 per cent, with paraplegia occurring in 4 ± 2 per cent and stroke in 5 ± 3 per cent. Actuarial survival estimates were 87 ± 4 per cent at 1 year, 81 ± 6 per cent at 2 years, and 81 ± 6 per cent at 4 years. Since these patients represent a high-risk cohort, the survival estimates are satisfactory, and are comparable to those of patients who underwent open procedures.

Recent investigations have also evaluated stent-grafts for treatment of aortic dissection. Dake *et al.* recently studied the placement of endovascular stent-grafts across the primary entry tear for the management for acute aortic dissection originating in the descending thoracic aorta. The endovascular stent-grafts were custom-designed for each patient according to measurements obtained from imaging studies. The stainless-steel endoskeleton framework was composed of Z-shaped stent elements and covered with woven polyester graft materials, or polytetrafluoroethylene graft material. Placement was successful in all 19 patients in this series, with complete thrombosis of the thoracic aortic false lumen in 15 of 19 patients (79 per cent); partial thrombosis was achieved in 4 of 19 patients (21 per cent). Revascularization of ischemic branch vessels, with subsequent relief of corresponding symptoms, occurred in 76 per cent of the obstructed branches. Three of the 19 patients died within 30 days (16 per cent); there were no deaths and no instances of aneurysm or aortic rupture during the average follow-up of 13 months. Although further evaluation is required, these initial results suggest that stent-graft coverage of the primary entry tear may be a promising new treatment for selected patients with aortic dissection.

Brewster *et al.* and the vascular group at the Massachusetts General Hospital reported their initial experience with 28 patients that underwent endovascular abdominal aortic aneurysm repair with stented bifurcation grafts and compared them with an equal number of patients undergoing open repair. They demonstrated significant advantages in terms of reduced blood loss, postoperative systemic complications, intensive care unit stay and length of hospital stay. However, they were somewhat offset by a higher incidence of vascular complications with six of 28 patients showing an endoleak at discharge and four patients requiring reintervention at 6 weeks for limb ischemia. These results are quite consistent with a learning curve and will undoubtedly improve with experience.

Iliac stenting

Kauffmann *et al.* in 1991 reported the results of a randomized study comparing iliac stenting with percutaneous transluminal angioplasty. The study indicated statistically significant differences in the complication rate (two of 62 after stenting and five of 69 after angioplasty) and patency (95 per cent after stenting and 72 per cent after angioplasty). Clinical improvement after 2 years was 89 per cent after stenting and 70 per cent after angioplasty. This is probably due to the greater luminal diameter that is achieved with stenting (Fig. 3). At the present time, iliac stenting is indicated as an adjunct for percutaneous transluminal angioplasty; longer follow-up is required before definitive recommendations can be made regarding stenting of all ileal lesions.

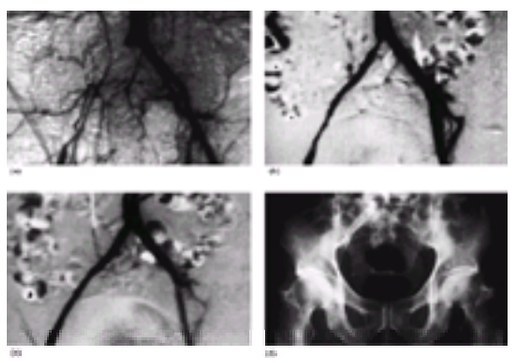


Fig. 3. (a) Intravenous digital subtraction angiography (DSA) before angioplasty. Complete iliac communis and externa occlusion (right). (b) DSA after lysis and dilatation. (c) DSA angiogram after stent implantation. (d) Plain radiograph after stent implantation (by courtesy of Angiomed).

Femoropopliteal stenting

Studies have shown a significant rate of early re-stenosis with stents in the femoropopliteal region. This may reflect the longer length, the smaller vessel diameter, and the longer stented surface area. Strecker *et al.* recently reported the long-term results of femoropopliteal artery stents in 80 patients after technical failure of balloon angioplasty. The 2-year primary patency rate was 51 per cent and the 3-year rate was 48 per cent. The 2-year primary patency rate was significantly better for shorter stents (59 per cent) versus that for longer stents (30 per cent) and for treated stenoses (73 per cent) versus treated occlusions (33 per cent). Because of the rate of restenosis secondary to intimal hyperplasia, femoropopliteal stents are not presently recommended for femoropopliteal lesions. Surgical revascularization is the preferred treatment, especially for long-segment disease. Stenting may be useful for treatment of percutaneous transluminal angioplasty related dissections and in patients with limb-threatening ischemia and no surgical alternative.

Renal artery stenting

Rees *et al.* reported results using the Palmaz stent in ostial renal lesions in 28 hypertensive patients. Stents were placed to treat elastic recoil immediately after conventional angioplasty in 20 patients and re-stenosis after percutaneous transluminal angioplasty in eight patients. At follow-up, hypertension was cured in three patients and improved in 15 patients, with a cumulative cure or improvement of 64 per cent at 6 months. Follow-up angiography was performed in 18 patients; re-stenosis was present in seven patients and was accompanied by a relapse of hypertension in only three patients. Although results are preliminary, renal stents may be beneficial in patients with poor results from conventional angioplasty for ostial atheroma.

Palmaz and Wallstents have also been used in the management of renal ischemia associated with acute aortic dissections. Results with these endovascular techniques appear promising. Long-term studies regarding durability and effectiveness, however, await further investigation.

Cardiac prostheses

The decision to use any particular heart valve is often based on personal clinical experience, an experience complicated by the evolutionary nature of prosthetic heart valve design and the myriad of devices now available. Although mechanical valves offer satisfactory hemodynamic performance, their well recognized thromboembolic risks help promote the development of biologic valves which, by partially decreasing the non-biologic–biologic interface, had the advantage of being less thrombogenic. This advantage was clearly confirmed in the mid 1970s by several clinical series showing that porcine valves were less thrombogenic than mechanical valves. As is often the case with newly introduced prosthetic valves, however, the initial experience showed great promise only to be followed in the early 1980s by the realization that the biologic valves themselves were also associated with their own set of problems, the most ominous of which was tissue failure. Therefore, against the risk of failure in the biologic valves must be weighed the risk of malfunction, thromboembolism, and anticoagulation related hemorrhage in the mechanical valves. The valves to be considered in this chapter are outlined in Table 2. There are two types of bioprosthetic valves (heterografts and homografts, Fig. 4) which may be composed of either bovine or porcine valvular or pericardial tissue. There are three general categories of available mechanical valves (ball and cage, tilting disk, and bi-leaflet, Fig. 5).

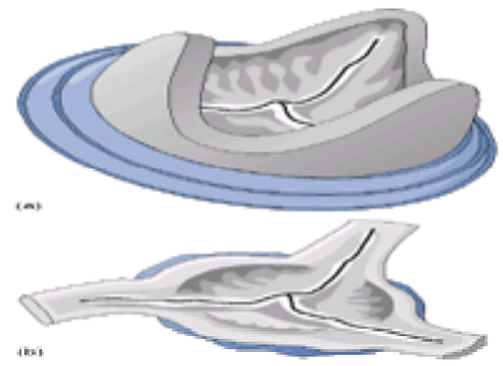


Fig. 4. (a) Carpentier–Edwards Porcine Heterograft. (b) Cryo-Life Homograft.

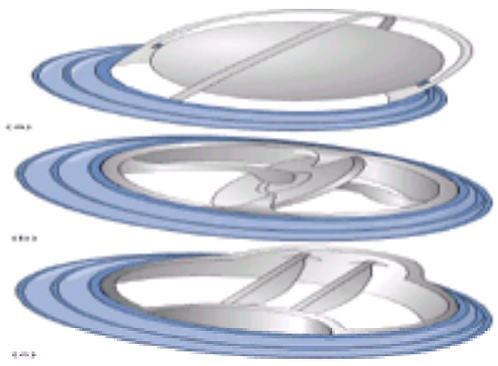


Fig. 5. (a) Starr–Edwards ball and cage mechanical valve. (b) Medtronic Hall tilting disk valve. (c) St. Jude Medical bi-leaflet valve.

Bioprosthetic		*Mechanical*	
Heterograft		Ball and cage	
	Hancock Carpentier-Edwards Ionescu-Shiley		Starr-Edwards
Homograft		Tilting disk	
	CryoLife		Bjork-Shiley Medtronic-Hall Omniscience
		Bileaflet	
			St Jude Medical Carbomedics Edwards-Duromedics

Table 2 Classification of prosthetic heart valves

Bioprosthetic valves

Porcine valves are obtained from freshly slaughtered pigs and treated so that the final product is a non-living valve, mounted on a Dacron stent. Upon removal from the pig, the endothelial cells quickly die and are easily removed by washing. The subendothelial elastin layer is immunogenic and is removed by treating the valve with the enzyme ficin. The sole remaining layer is collagen, which has sufficient species homogeneity to prevent rejection. However, native collagen lacks the strength to withstand arterial pressures over millions of cycles and is therefore strengthened by cross-linking with glutaraldehyde. The valve is then sewn to the stent and stored in 0.065 per cent glutaraldehyde solution.

Pericardial valves were initially introduced in the 1980s. These valves were made of bovine pericardium and designed to fabricate a trileaflet valve. Early studies demonstrated durability problems and leaflet tearing. This led to structural changes and improvements in durability. Intermediate results are encouraging.

Homograft valves were initially not well accepted due to several inherent problems in obtaining fresh specimens, sterilization, storage, and operative technique. In addition, early methods of preservation led to structural deterioration. Renewed interest in homografts has resulted from improvements in preservation. Aortic, pulmonary, and mitral valve allografts are currently cryopreserved at ultra-low temperatures. Since 1984, CryoLife has processed over 25 000 donated hearts and documented the implantation of more than 25 000 cryopreserved allograft valves at 500 institutions in the United States, Canada, and Europe. In a recent clinical outcome analysis of 1698 CryoValve allograft implants, there were no occurrences of thromboembolic or anticoagulant-related complications. Further, 10-year freedom from structural deterioration for aortic and pulmonary valve allografts were 82 per cent and 88 per cent, respectively. Patients receiving cryopreserved heart valve allografts had a low incidence of endocarditis, with 29 (1.7 per cent) reported occurrences. No reported case of endocarditis was attributed to the allograft valve.

Mechanical valves

The Starr–Edwards mechanical valve has undergone many modifications since its introduction in 1961. Its current configuration, consists of a silastic ball, stainless

steel struts, and a Dacron sewing ring. The Starr–Edwards valve has the longest history of continuous usage of any valve and remained a mainstay of cardiac surgery 30 years after its initial introduction. However, an inherent problem with ball and cage design valves is the high profile of the cage and tertiary orifice obstruction by the ball. Because of this, the Starr–Edwards valve is now usually reserved for cases, such as subacute bacterial endocarditis, where one wants a large, fluffy, redundant sewing ring that can fill gaps in the aortic or mitral annulus.

All ball in cage valves have three orifices. As shown in Fig. 6, the primary orifice is the orifice through which the blood must go when leaving the left ventricle. The secondary orifice is the orifice subtended by the angle made from the ball in the open position and the primary orifice. The height of the cage determines the area of the secondary orifice. In proper valve design, of course, the primary and secondary orifices will be equal. The tertiary orifice is the orifice between the perimeter of the ball in the open position and the walls of the aorta. It is this orifice that cannot be controlled by valve design and can cause obstruction to blood flow. The ball valve, therefore, creates a paradoxical problem. If one wishes to insert a valve with a large primary orifice, one will then create a small tertiary orifice due to the size of the ball. Or conversely, a large tertiary orifice will result in a small primary orifice. Because of the dual problems of tertiary orifice obstruction and high cage profile, the Starr–Edwards valve should not be used in the aortic position of patients with narrow aortic roots or in the mitral position in patients with small left ventricular chambers. The incidence of thromboembolism is also higher with the Starr–Edwards valve compared with the bi-leaflet valves.

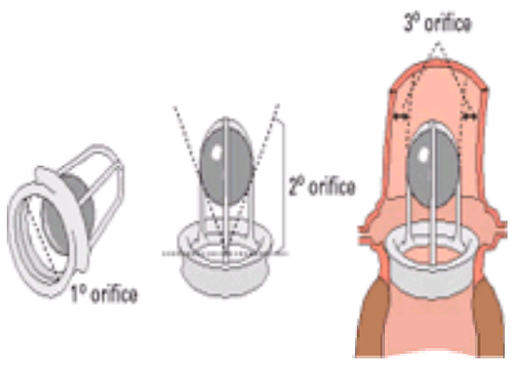


Fig. 6. Three orifices produced by ball and cage valves.

To obviate the problems associated with ball and cage design valves, the so-called low profile valves were developed. The Bjork–Shiley convexoconcave tilting-disk valve resulted in a high incidence of structural failure, and has been withdrawn. Since 1994, the Medtronic-Hall and Omniscience tilting disk valves have been available in the United States. These valves are low profile and may be used in patients with small ventricles. The effective orifice size is larger than with the ball and cage valves in the smaller sizes. The most significant problem associated with these valves has been thrombosis, which has been reported despite anticoagulation. Surgical technique must be perfect when implanting these valves, as retained chordae or leaflets may cause valvular interference.

The most commonly used low profile valves are the St. Jude and Carbomedics bi-leaflet valves. The St. Jude valve is made from pyrolight carbon with a Dacron sewing ring. There has been no postoperative hemodynamic or clinical performance difference between the bi-leaflet and tilting disc valves.

To evaluate the relative merits and shortcomings of mechanical and biologic prostheses and their long-term outcome, we conducted a longitudinal analysis of biologic and mechanical valves inserted at the Yale-New Haven Hospital over an 11-year period. We were unable to demonstrate that there was a clear advantage to the generalized use of either of these two types of valves. However, there is important information that supports the use of specific valves for specific indications.

Mechanical valves rarely develop mechanical failure; however, if clinical problems develop a year or so after implanting a mechanical valve, one should keep in mind that the most common types of malfunction are tissue ingrowth with primary orifice obstruction or tissue ingrowth with leaflet or ball jamming. In our series of 510 mechanical valves followed for 10 years, we observed 10 incidents of malfunction by these modes. On the other hand, of 606 biologic valves that were inserted, 57 valves developed torn leaflets with insufficiency and 11 developed calcified leaflets with stenosis. Therefore, the possibility of valve failure must be an important consideration when deciding upon the use of a tissue valve. Knowing the natural history of tissue valves is very helpful in deciding upon who should receive a porcine valve. Tissue valves fail much more rapidly in younger age group patients. In pediatric patients, more than 90 per cent of valves will fail over a 5-year period. In the adult age group (18 years of age) the failure rate decreases from 7 per cent per year of patients age 18 to 30 to 2 per cent per year of patients age 70 to 80. In the actuarial analysis presented in Fig. 7, patients receiving biologic valves were arbitrarily divided into patients above and below 50 years of age to demonstrate the effective age on valve failure slopes. However, when we divided the biologic group into those above and below 60 years age, the actuarial slopes were again significantly different, this time at the $p < 0.01$ level. Accordingly, freedom from failure decreases in a linear fashion with increasing age, but there may be no clear-cut age above which freedom from failure becomes negligible. Nevertheless, it is an important attribute of biologic valves that they may provide relatively safe management in a group of patients in whom anticoagulation may be difficult or hazardous and who frequently have other medical disorders that may require surgical intervention.

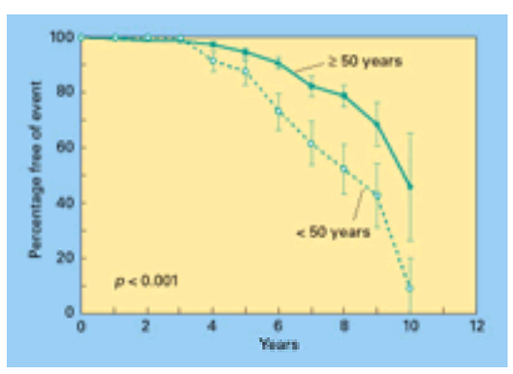


Fig. 7. Actuarial analysis of valve failure in patients above and below 50 years of age who underwent valve replacement with biologic valves.

Although valve failure is not an important consideration for mechanical valves, thromboembolism is. Fig. 8 shows the actuarial analysis of all valves in which thromboembolism, including thrombosed valves, was documented. Although the incidence of thromboembolism was significantly lower with biologic valves, the sequela of thromboembolism seemed to be just as serious in the biologic group as the mechanical group. For example, there were four fatal cerebral vascular accidents in 13 patients with biologic valves and six in 37 patients with mechanical valves. All thrombosed valves (one biologic and seven mechanical), resulted in deaths.

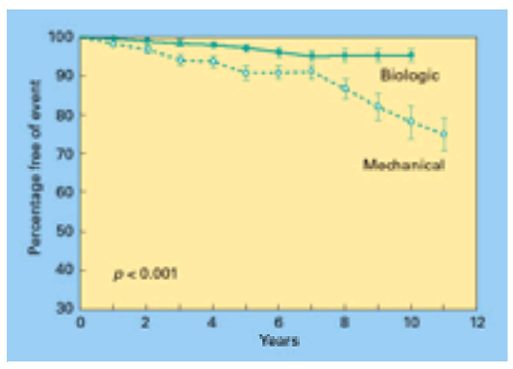


Fig. 8. Actuarial analysis of all valves in which thromboembolism, including thrombosed valve, was documented.

Interestingly, there is no difference between biologic and mechanical valves when followed over a 10-year period. As portrayed in the actuarial analysis in Fig. 9, when total morbidity and valve-related mortality were combined and plotted for biologic and mechanical valves, there was significantly less trouble ($p < 0.01$) with the biologic valves over the first 5 years of the study and significantly less trouble ($p < 0.001$) with mechanical valves over the last 5 to 6 years of the study. However, over the entire study, there was no significant difference between the groups.

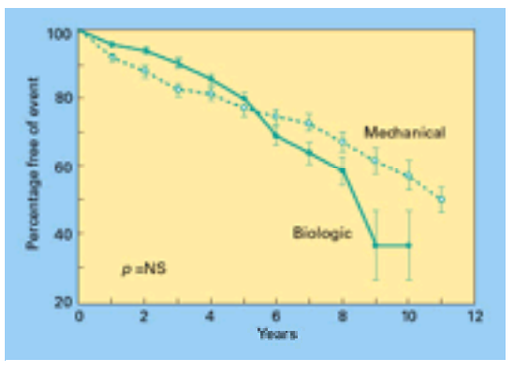


Fig. 9. Actuarial analysis of total morbidity and valve related mortality for biologic and mechanical valves: years 1 to 5 $p < 0.01$; years 5 to 10 $p < 0.001$; years 1 to 10, not significant.

Improvements in mechanical valves will obviously require the development and use of less thrombogenic materials. The tissue failure problem with biologic valves may be eased somewhat by the introduction of low pressure-fixation valves. Fixation under low pressure has the advantage of helping to preserve the tertiary conformation of collagen molecules. Thus closure energy is partially dissipated through disruption of non-covalent bonding interactions, which are able to reform during opening rather than through disruption of covalent bonds that can never reform because of the non-living nature of the tissue.

Further reading

Aupart M, Neville P, Dreyfus X *et al.* The Carpentier-Edwards pericardial aortic valve: Intermediate results in 420 patients. *European Journal of Cardiothoracic Surgery* 1994; **8**(5): 277–80.

Babatasi G, Bara L, Massetti M, Galateau F, Bloch MF, Boulakia-Cohen F, Khayat A, Samama MM. Value of carbon in the thrombogenicity of cardiovascular prostheses. *Annales de Cardiologie et d'Angeiologie* 1997; **46**(1): 7–14.

Bacourt F. Prospective randomized study of carbon-impregnated polytetrafluorethylene grafts for below-knee popliteal and distal bypass: results at 2 years. *Annals of Vascular Surgery* 1977; **11**(6): 596–603.

Bergamini TM, Bandyk DF, Govostis D *et al.* Infection of vascular prostheses caused by bacterial biofilms. *Journal of Vascular Surgery* 1988; **7**(1): 21–30.

Boylan MJ, Lytle BJ, Loop FD, *et al.* Surgical treatment of isolated left anterior descending coronary stenosis. Comparison of left internal mammary artery and venous autograft at 19 to 20 years of follow-up. *Journal of Thoracic and Cardiovascular Surgery* 1994; **107**: 657–62.

Brewster DC, Geller SC, Kaufman JA, Cambria RP, Gertler JP, LaMuraglia GM, Atamian S, Abbott WM. Initial experience with endovascular aneurysm repair: comparison of early results with conventional open repair. *Journal of Vascular Surgery* 1998; **27**(6): 992-1005.

Callow AD. Arterial homografts. *European Journal of Vascular and Endovascular Surgery* 1996; **12**(3): 272–81.

Cantelmo NL, Quist WC, Lo Gerfo FW. Quantitative analysis of anastomotic intimal hyperplasia in paired Dacron and PTFE grafts. *Journal of Cardiovascular Surgery* 1989; **30**: 910–15.

Capps SB, Fronk DM, Haug KD, Rose BK. *CryoValve heart valve allograft clinical experience*. Cryolife, Inc., Kennesaw, GA, 1998.

Carpenter JP, Tomaszewski JE. Human saphenous vein allograft bypass grafts: immune response. *Journal of Vascular Surgery* 1998; **27**(3): 492–9.

Carpentier A, Dubost C, Lane E *et al.* Continuing improvements in valvular bioprostheses. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 27–42.

Chakfe N, Kretx JG, Petit H *et al.* Albumin-impregnated polyester vascular prosthesis for abdominal aortic surgery: an improvement? *European Journal of Vascular Surgery* 1996; **12**: 346–53.

Chanda J, Kuribayashi R, Abe T. Heparin in calcification prevention of porcine pericardial bioprostheses. *Biomaterials* 1997; **18**(16): 1109–13.

Cheshire NJ, Wolfe JH, Barradas MA, Chambler AW, Mikhailidis DP. Smoking and plasma fibrinogen, lipoprotein (a) and serotonin are markers for postoperative infrainguinal graft stenosis. *European Journal of Vascularand Endovascular Surgery* 1996; **11** (4): 479–86.

Dake MD, Miller DC, Semba CP, *et al.* Transluminal placement of endovascular stent-grafts for the treatment of descending thoracic aortic aneurysms. *New England Journal of Medicine* 1994; **331**(26):1729–34.

Dake MD, Kato NK, Mitchell RS, *et al.* Endovascular stent –graft placement for the treatment of acute aortic dissection. *New England Journal of Medicine* **340** (20) 1546–1552.

Dardik H, Ibrahim IM, Dardik I. Successful arterial substitution with modified human umbilical vein. *Annals of Surgery* 1976; **183**(3): 252–8.

Fann JI, Mitchell RS, Dake MD, *et al.* The results of endovascular grafting of descending thoracic aortic aneurysms. In Yao JST, ed. *Vascular surgery: Twenty years of progress*.. Appleton and Lange, Stamford, 1996, pp. 241–54.

Fann JI, Slonin SM, Miller DC. Endovascular treatment of thoracic aortic aneurysms and dissection. In Calligaro, Dougherty, Hollier, eds. *Diagnosis and treatment of aortic and peripheral arterial aneurysms*. W.B. Saunders Co., Philadelphia, 1999, pp. 97–111.

Fiore AC, Naunheim KS, D'Orazio S *et al.* Mitral valve replacement: randomized trial of St. Jude and Medtronic-Hall prostheses. *Annals of Thoracic Surgery* 1992; **54**: 68–73.

Gossetti B, Irace L, Intrieri F *et al.* The salvage of the lower limbs: choice of the prosthetic material in femoro-distal and axillo-femoral bypass. *Italian Journal of Surgical Science* 1988; **18**: 69–73.

Hammond GL, Geha AS, Kopf GS, Hashim SW. Biological versus mechanical valves. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 182–98.

Johnson WC, Squires JW. Axillo-femoral (PTFE) and infraimguinal revascularixation (PTFE and umbilical vein). *Journal of Cardiovascular Surgery* 1991; **32**: 344.

Karner J, Schemper M, Teleky B, Kretschmer G, Piza F, Polterauer P. Aorto-Y-bifurcation graft: Dacron versus PTFE. Preliminary results of a randomized prospective study. *International Surgery* 1988; 73: 218–20.

Kato, Matsuda T, Kaneko M, Ueda T, Kuratani T, Yoshioka Y, Ohnishi K. Experimental assessment of newly devised transcatheter stent-graft for aortic dissection. *Annals of Thoracic Surgery* 1995; **59**: 908–15.

Kauffman GW, Richter GM, Noldge G, Roeren T, Landwehr P, Lackner K. Four-years' experience with a balloon-expandable endoprosthesis. Experimental and clinical application. *Radiologe* 1991; **31**(4): 202–9.

Kopf GS, Geha AS, Hellenbrand WE, Kleinman CS. Fate of left-sided cardiac bioprosthetic valves in children. *Archives of Surgery* 1986; **121**: 488–90.

Leseche G, Ohan J, Bouttier S, Palombi T, Bertrand P, Andreassian B. Above-knee femoropopliteal bypass grafting using endothelial cell seeded PTFE grafts: five-year clinical experience. *Annals of Vascular Surgery* 1995; **9** (Suppl.): S15–23.

von der Leyen HE, Gibbons GH, Morishita R, Lewis NP, Zhang L, Nakajima M, Kaneda Y, Cooke JP, Dzau VJ. Gene therapy inhibiting neointimal vascular lesion: *in vivo* transfer of endothelial cell nitric oxide synthase gene. *Proceedings of the National Academy of Sciences, USA* 1995; **92**(4): 1137–41.

Magilligan DJ, Lewis JW, Heinzerling RH, Smith D. Fate of a second porcine bioprosthetic valve. *Journal of Thoracic and Cardiovascular Surgery* 1983; **85**: 362–70.

Mann MJ, Gibbons GH, Kernoff RS, *et al.* Genetic engineering of vein grafts resistant to atherosclerosis. *Proceedings of the National Academy of Sciences, USA* 1995; **92**(10): 4502–6.

Marrin CAS, Birkmeyer JD, O'Connor G. The Bjork-Shiley dilemma. *Annals of Thoracic Surgery* 1993; **55**: 1361–4.

Miller DC, Oyer PE, Stinson EB *et al.* Ten-to-fifteen year reassessment of the performance characteristics of the Starr–Edwards Model 6120 mitral valve prosthesis. *Journal of Thoracic and Cardiovascular Surgery* 1983; **85(1)**: 1–20.

Muhl E, Gatermann S, Iven H, Dendorfer A, Bruch HP. Local application of vancomycin for prophylaxis of graft infection: release of vancomycin from antibiotic-bonded Dacron grafts, toxicity in endothelial cell culture, and efficacy against graft infection in an animal model. *Annals of Vascular Surgery* 1996; **10**(3): 244–53.

Niklason LE, Langer RS. Advances in tissue engineering of blood vessels and other tissues. *Transplant Immunology* 1997; **5**(4): 303–6.

Nunn DB, Freeman MN, Hudgkins PC. Post-operative alterations in size of Dacron grafts: an ultrasonic evaluation. *Annals of Surgery* 1979; **189**(6): 741–5.

Okahara K, Kambayashi J, Shibuya T, Kawasaki T, Sakon M, Dohi Y, Oka Y, Ito S, Miyake S. An infection-resistant PTFE vascular graft; spiral coiling of the graft with ofloxacin-bonded PTFE thread. *European Journal of Vascular and Endovascular Surgery* 1995; **9**(4): 408–14.

Pelletier LC, Carrier M, Leclerc Y *et al.* Porcine versus pericardial bioprostheses: A comparison of late results in 1,593 patients. *Annals of Thoracic Surgery* 1989; **47**: 352–61.

Qvarfordt PG, Stoney RJ. Arterial autografts. *Acta Chirurgica Scandinavica (Supplementum)* 1985; **529**: 37–45.

Qvarfordt PG, Reilly LM, Ehrenfield WK, *et al.* Surgical management of vascular graft infections – local treatment, graft excision, and methods of revascularization. In Bernhard VM, Towne JB, eds. *Complications in vascular surgery*. New York, Grune and Stratton, 1985, pp. 499–511.

Rees CR, Palmaz JC, Becker GJ, *et al.* Palmaz stent in atherosclerotic stenoses involving the ostia of the renal arteries: preliminary report of a multicenter study. *Radiology* 1991; **181**(2): 507–14.

Richter GM, Palmaz JC, Allenberg JR, Kauffmann GW. Transluminal stent prosthesis in aneurysm of the abdominal aorta. Initial experiences with a new procedure. *Radiologe* 1994; **34**(9): 511–18.

Ritter EF, Kim YB, Reischl HP, Serafin D, Rudner AM, Klitzman B. Heparin coating of vascular prostheses reduces thromboemboli. *Surgery* 1997; **122**(5): 888–92.

Ritter EF, Fata MM, Rudner AM, Klitzman B. Heparin bonding increases patency of long microvascular prostheses. *Plastic and Reconstructive Surgery* 1998; **191**(1): 142–6.

Robicsek F, Duncan GD, Anderson CE *et al.* Indium 111-labeled platelet deposition in woven and knitted Dacron bifurcated aortic grafts with the same patient as a clinical model. *Journal of Vascular Surgery* 1987; **5**: 833–7.

Rutherford RB, Jones DN, Bergentz S-E *et al.* Factors affecting the patency of infrainguinal bypass. *Journal of Vascular Surgery* 1988; **8(3)**: 236–46.

Slonim SM, Nyman U, Semba CP, Miller DC, Mitchell RS, Dake M. Aortic dissection: Percutaneous management of ischemic complications with endovascular stents and balloon fenestration. *Journal of Vascular Surgery* 1996; **23**: 241–53.

Sottiurai VS, Yao JST, Flinn WR, Batson RC. Intimal hyperplasia and neointima: An ultrastructural analysis of thrombosed grafts in humans. *Surgery* 1983; **93**: 809–17.

Stinson EB, Griepp RB, Bieber CP, Shumway NE. Aortic valve allografts for mitral valve replacement. *Surgery* 1975; **77**: 861–7.

Stinson EB, Griepp RB, Oyer PE, Shumway NE. Long-term experience with porcine aortic valve xenografts. *Journal of Thoracic and Cardiovascular Surgery* 1997; **73**: 54–63.

Stoney RJ, DeLuccia N, Ehrenfeld WK, *et al.*, Aortorenal arterial autografts. *Archives of Surgery* 1981; **116** (11): 1416-22.

Strecker EP, Boos IB, Gottmann D. Femoropopliteal artery stent placement: evaluation of long-term success. *Radiology* 1997; **205**(2): 375–83.

Strobel R, Boontje AH, Van Den Dungen JJ. Aneurysm formation in modified human umbilical vein grafts. *European Journal of Cardiothoracic Surgery* 1996; **11**(4): 417–20.

Vasseur B, Hammond GL. New technique for repair of ascending thoracic aortic dissections. *Annals of Thoracic Surgery* 1989; **47(2)**: 318–19.

Veith FJ, Gupta S, Daly V. Management of early and late thrombosis of expanded polytetra-fluorethylene (PTFE) femoropopliteal bypass grafts: favorable prognosis with appropriate reoperation. *Surgery* 1980; **87(5)**; 581–7.

Wylie EJ. Vascular replacement with arterial autografts. *Surgery* 1965; **57**: 14.

Yeager A, Callow AD. New graft materials and current approaches to an acceptable small-diameter vascular graft. *Transactions of the American Society for Artificial Organs* 1988; **34**(2): 88–94.

Zacharias RK, Kirkman TR, Clowes AW. Mechanisms of healing in synthetic grafts. *Journal of Vascular Surgery* 1987; **6**: 429–36.

40.5 Principles of Blood conservation, cardiac anaesthesia, and cerebral protection

Catherine R. Grebenik and Michael E. Sinclair

Blood conservation
Risks of transfusion
Conservation techniques
Cardiac anaesthesia
Narcotics
Other anaesthetic agents
Intraoperative ischaemia and preoperative medications
Cerebral protection
Hypothermia
Carbon dioxide management
Anaesthetic agents
Calcium antagonists
Hyperglycaemia
Further reading

Blood conservation

Risks of transfusion

The transfusion of homologous blood products carries multiple risks that are cumulative as more blood is transfused. Blood transfusion is immunosuppressive and dose-dependently associated with an increased risk of postoperative infection; large-volume transfusions of donor blood are associated with increased pulmonary and renal complications. The need for blood products for patients undergoing cardiac surgery imposes significant costs and demands on hospital blood banks; it is estimated that around 10 per cent of all blood transfusions in the United States are associated with coronary arterial surgery. Recent fears about the transmission of the human immunodeficiency virus (**HIV**) have heightened earlier concerns over risks such as transfusion reactions, transmission of other viral and bacterial diseases, coagulopathy, acute respiratory distress, isoimmunization, and febrile reactions.

Although current blood-banking techniques make bacterial contamination of blood rare, the threat of transmission of viruses remains. Testing of blood donors for HIV antibody did not begin until 1985; in 1984 an estimated 7200 patients contracted transfusion-associated acquired immune deficiency syndrome. Of these, 27 per cent had received the infected unit during cardiac surgery. Although the risks have decreased appreciably since 1985, tests for HIV antibody do not detect all infected donors: HIV antibody may not be present until 6 to 14 weeks after infection, as determined by lymphocyte culture or experimental HIV antigen and antibody assay. Although voluntary self-disqualification by donors at risk of HIV infection reduces the risk of transmission, some high-risk individuals continue to donate blood. Current estimates of the risks of HIV transmission from screened blood vary from 1:40 000 to 1:1 000 000.

Other clinically significant illnesses can be transmitted by transfusion, hepatitis viruses being the most commonly transmitted agents. Cytomegalovirus and Epstein–Barr virus are transmitted to previously unexposed recipients, but infection is usually asymptomatic, except in immunocompromised hosts. Malaria, syphilis, and babesiosis have also been transmitted. Human T-cell lymphotropic virus is present in donor populations in the United States; this virus is associated with adult T-cell leukaemia and a chronic progressive myelopathy. The possibility of transmission of new-variant Creutzfeldt–Jakob disease cannot be discounted. Although no instance has yet been documented as a result of transfusion, the long latency period of these illnesses makes any such association difficult to prove.

Conservation techniques

There has been a gradual and progressive fall in blood use in cardiac surgery. Between 1967 and 1970 an average of 12 units of whole blood was transfused to each patient. In 1973, the average patient received 8 units. Over the last 20 years, new techniques have reduced the average amount transfused to 1 to 3 units per patient, although transfusion practices vary widely between institutions. Total mean blood loss during first-time cardiac surgery has been estimated at 1400 ml. In one series, the use of a comprehensive blood-conservation programme allowed an astonishing 97 per cent of 2326 patients undergoing coronary arterial bypass grafting to avoid any banked blood products.

The first principle in reducing the need for blood products in cardiac surgery is careful surgical technique with meticulous attention to potential bleeding points. Surgery should be rapid: the coagulopathy associated with cardiopulmonary bypass is directly related to its duration, and variations in surgical expertise account for large differences in postoperative bleeding. When bleeding occurs, early surgical re-exploration will reduce the total volume of transfusion as well as decreasing local fibrinolysis, which may contribute to a coagulopathy. Maintenance of normothermia is highly desirable: hypothermia causes reversible platelet dysfunction that may exacerbate a bleeding tendency.

Blood conservation is based on two principles, haemodilution and autologous transfusion. The use of a crystalloid prime with haemodilution is now the standard procedure in adult patients; low priming-volume oxygenators allow even small patients to be managed without excessive haemodilution. The use of normovolaemic haemodilution has been extended into the postoperative period, with more stringent criteria for the transfusion of red blood cells applied to most patients. A haematocrit of 22 to 25 per cent in completely revascularized patients is considered acceptable in many institutions. However, the lowest acceptable haematocrit will vary from patient to patient, depending on their physiological condition and cardiovascular reserve.

Some type of autologous transfusion is used in most cardiac procedures. Forms of autologous transfusion currently in use include preoperative blood donation, intraoperative withdrawal of one or more units of blood before cardiopulmonary bypass for reinfusion after bypass, intraoperative autotransfusion of blood suctioned from the operative field, and the postoperative infusion of mediastinal blood collected in chest-tube drainage reservoirs.

Preoperative donation of autologous blood for intraoperative use in patients undergoing cardiovascular procedures was reported as early as 1967. Interest remained low in the 1970s, owing to concerns over administrative difficulties and cost. There is, unfortunately, no cost saving from the use of predonated blood; in fact, administrative costs may be considerably greater for autologous blood than for routine stock. The logistics of organizing a predonation programme are far from simple, but current concerns surrounding the risk of transmission of HIV virus, and realization about the significant morbidity and mortality of transfusion-related hepatitis have renewed interest in this practice. The use of recombinant human erythropoietin increases the ability of patients to donate autologous blood and may alleviate problems with anaemia, albeit with increased cost. Although autologous blood is commonly obtained for use in patients undergoing orthopaedic or other major surgical procedures, acceptance has been slow in the field of cardiac surgery because of concerns over the possible effects of blood donations by patients with coronary arterial and valvular heart disease, although several reports have documented the safety of preoperative autologous donation by these patients. Birkmeyer *et al.* undertook a prospective, multi-institution study of autologous predonation in cardiac surgery and concluded that it was not cost-effective and that, even for individual patients, the risks of donating autologous blood might outweigh the benefits.

The withdrawal of one or more units of autologous blood before the start of cardiopulmonary bypass for reinfusion following bypass is a common method of blood conservation. Reports on the efficacy of this procedure in decreasing requirements for homologous blood are conflicting. Several studies have reported significant decreases in both banked-blood requirements and blood loss after bypass but their control groups were not always well matched. Others found no difference in blood requirements when compared with well-matched controls, or in the quantity of postoperative bleeding. One study actually reported an increase in banked-blood requirements in patients receiving autologous rather than homologous blood.

The intraoperative use of a cell-scavenging device during cardiac surgery significantly reduces the requirements for banked blood. Although such devices have been available for many years, early models were labour-intensive, and their use was associated with complications such as haemolysis, coagulopathies, emboli, renal failure, and pulmonary abnormalities. Current models are completely automated and consist of a regionally heparinized collection system for the retransfusion of washed red blood cells, with a haematocrit of 50 to 60 per cent. The initial capital outlay is high and disposable parts are expensive. Salvaged red blood cells have a normal survival time and higher concentrations of 2, 3-diphosphoglycerate than banked blood. A remaining problem is that dilutional coagulopathy can occur after several blood volumes have been processed by cell salvage because coagulation factors are discarded during the washing process.

Reinfusion of shed mediastinal blood has been shown to be safe. Shed blood is defibrinated by the action of the mediastinum and anticoagulant is not required. Mediastinal drainage must be collected in a special container: the Sorenson Receptaseal autotransfusion system consists of a rigid, outer reusable canister with a disposable, sterile liner made of two plastic bags, each with 170-μm filters. Administration through a 40-μm filter is recommended. Most protocols require the reinfusion

of shed blood every 4 h if drainage exceeds 250 ml; quantities smaller than this are discarded. Evaluation of shed blood shows haematocrits of 18 to 25 per cent, platelets of 60 000 to 70 000, decreased fibrinogen, elevated fibrin-split products, and elevated plasma haemoglobin. Platelets are non-functional but the survival time of red blood cells is normal. Controversy remains over the value of reinfusing shed mediastinal blood. In small volumes its use does not adversely affect recipients' coagulation profiles or increase the need for platelets or fresh frozen plasma. However, larger volumes may be associated with delayed bleeding and haemostatic derangement, and therefore be counterproductive. Thus, the reinfusion of shed mediastinal blood may not save substantial amounts of autologous red cells and its use may be deleterious. Concerns also remain over the risk of introducing infection. Although cultures obtained from mediastinal blood were positive after 1 week of incubation in 5 to 25 per cent of samples, cultures obtained from patients following transfusion of this blood were negative and none of the patients who received culture-positive blood had signs of clinical infection. The incidence of postoperative systemic and wound infections does not seem to be greater in patients receiving mediastinal rather than homologous blood. Four per cent of cultures from blood-bank units have been reported as positive.

Finally, any discussion of blood conservation must mention the drugs that modulate blood activation. These include tranexamic acid, desmopressin (dDAVP), and aprotinin, a naturally occurring protease inhibitor of bovine origin. Aprotinin provenly reduces blood loss and transfusion requirements in ‘redo' and high-risk primary coronary arterial surgery. It appears to act by preserving platelet adhesion and reducing fibrinolysis, and it also inhibits the activation of neutrophils and of the kallikrein system. Whilst it is undeniable that prophylactic antifibrinolytic drugs are effective in reducing bleeding, and there has been great enthusiasm in particular for aprotinin, concerns remain about their routine use both for economic reasons and for fear of potential thromboembolic complications. The prothrombotic effects of aprotinin have been implicated in an increase in perioperative myocardial infarction in patients undergoing coronary arterial surgery. Until this concern is resolved, it may be prudent to restrict the use of this drug to those at highest risk of excessive bleeding, for example patients with acute aortic dissection and systemic infection.

Cardiac anaesthesia

The choice of anaesthetic technique for patients undergoing cardiac surgery depends on numerous factors, including the experience and preference of the anaesthetist: no data support the superiority of one technique over any other. The goal of the anaesthetist is to provide unconsciousness, analgesia, amnesia, and suppression of the neuroendocrine response to surgery. Ideally the anaesthetic technique should incorporate the possibility of an early return to spontaneous breathing.

Narcotics

In the early days of cardiac surgery, the only anaesthetics available were the inhalational agents, principally halothane. In 1969, Lowenstein published a report on the use of morphine in doses of 0.5 to 3 mg/kg as the principal anaesthetic for patients undergoing aortic-valve replacement. Patients with poor ventricular function had more stable haemodynamics than those treated by earlier techniques using inhalational agents, although problems such as hypotension due to histamine release and arterial and venous dilation, increased fluid requirements, and a high incidence of intraoperative recall were encountered. With the introduction of fentanyl, a highly lipid-soluble, rapidly acting synthetic narcotic, morphine was largely abandoned. The first report of the use of high-dose fentanyl for patients undergoing cardiac surgery was published in 1978.

Fentanyl is approximately 80 times as potent as morphine, and is used in doses of 10 to 100 μg/kg. Another synthetic opioid, sufentanil, has also become popular as an anaesthetic agent. It is approximately 5 to 10 times as potent as fentanyl, has a slightly more rapid onset, and has a shorter duration of action due to greater lipid solubility. Unlike morphine, fentanyl and sufentanil do not cause histamine release and, in clinical doses, neither is a myocardial depressant. Their effects on the peripheral vasculature are the subject of controversy, but fentanyl reportedly has either no effect on or decreases peripheral vascular resistance in humans. Remifentanil is a new, synthetic, piperidine-derivative opioid that is metabolized by plasma esterases and has an extremely short plasma half-life of 3 to 5 min. The short half-life necessitates administration by continuous infusion and permits high doses to be given, but with rapid emergence from anaesthesia that is independent of the duration or total dose of drug administered. The swift return to spontaneous respiration after stopping remifentanil infusion suggests that its use should be ideal for so-called fast tracking of cardiac surgical patients. Unfortunately, the rapid dissipation of opiate effect is rather difficult to manage in clinical practice.

The synthetic opioids cause a dose-dependent bradycardia, mediated by the vagus nerve. Severe bradycardia or asystole has been reported with the muscle relaxant vecuronium in combination with narcotics, particularly sufentanil. This problem can be minimized by using the muscle relaxant pancuronium, which is vagolytic. High-dose, and occasionally low-dose, narcotics cause rigidity of the chest wall, which can be severe enough to make ventilation impossible. This rigidity may be due to effects on the central nervous system, although the precise mechanism is unknown. The incidence and severity of this problem may be decreased by the prior administration of a small dose of a non-depolarizing muscle relaxant. All opioid agonists produce dose-dependent respiratory depression through a direct effect on the brainstem respiratory centres; at the doses commonly used in cardiac surgery, this generally results in a need for postoperative mechanical ventilation.

Fentanyl is extensively metabolized to inactive metabolites by the liver, with a high hepatic extraction ratio, indicating that metabolism is perfusion-dependent: anything that decreases hepatic blood flow will therefore decrease the rate of metabolism. Metabolites are excreted by the kidney. The elimination half-life is 185 to 219 min, but this is significantly prolonged in elderly patients (945 min). Sufentanil also undergoes hepatic metabolism, with a high hepatic extraction ratio, but some of its metabolites are active. The elimination half-life is 148 to 164 min. Metabolites are excreted in the urine and faeces. Prolonged respiratory depression has been reported in patients with renal failure. Although large doses of narcotics cause progressive slowing of the electroencephalogram and unconsciousness, intraoperative recall has been reported. The use of amnesiac agents such as scopolamine or benzodiazepines as premedication, and intraoperative supplementation, are advocated.

The choice of opiate largely depends on personal preference. Although early reports suggested that sufentanil provided greater haemodynamic stability than fentanyl, there are no convincing data to support the superiority of one opioid over another. Proponents of sufentanil argue that it allows earlier extubation. This effect is dose-dependent and patients who receive larger doses may require the same duration of postoperative ventilation as those given fentanyl. One significant difference between the various opiates is cost: sufentanil and remifentanil are significantly more expensive than fentanyl.

The advantages of high-dose narcotics include ease of administration, haemodynamic stability, even in patients with depressed ventricular function, the provision of a basal level of anaesthesia at all times during the procedure, and extension of analgesia into the postoperative period. Disadvantages include the requirement for postoperative mechanical ventilation, an inconsistent ability to block the hypertensive response to periods of surgical stimulation, and the suggestion that there is a ceiling effect beyond which additional drug is of no benefit. In patients with good ventricular function it is frequently necessary to supplement the narcotic with low doses of inhalational agents or vasodilators in order to control hypertension and provide stable haemodynamics. High doses of narcotics may induce an acute tolerance. Higher doses have been advocated (above 100 μg/kg of fentanyl or 20–30 μg/kg of sufentanil) in order to block hypertensive responses intraoperatively. However, there is no dose that is able to block haemodynamic and endocrine responses in all patients. In other words, there may be no achievable plasma concentration that consistently provides complete anaesthesia in all patients. Extremely high doses of narcotics, therefore, have little advantage.

Other anaesthetic agents

The advantages of using inhalational agents such as halothane, enflurane, or isoflurane include the ability to increase or decrease the dose at will, a decrease in myocardial oxygen consumption as a result of drug-induced myocardial depression, and the lack of prolonged effect, which allows extubation whenever it is deemed medically appropriate. Disadvantages include myocardial depression, which may not be tolerated in patients with poor ventricular function, a greater incidence of hypotension and haemodynamic variability than occurs with opioids, a risk of potentiating dysrhythmias, the possibility of coronary steal with isoflurane, and a lack of postoperative analgesia.

When used alone, narcotics can be associated with hypertension and tachycardia during surgical stimulation in patients with normal left-ventricular function. It has therefore become common practice to use a moderate to large dose of opiate as a background anaesthetic, with the addition of low to moderate doses of inhalational or intravenous anaesthetic agents as a supplement to ensure unconsciousness and to assist in controlling haemodynamic variables. The agents used include all of the volatile anaesthetic drugs (halothane, enflurane, isoflurane, sevoflurane, and desflurane) and the intravenous anaesthetics propofol and midazolam.

Isoflurane controls blood pressure principally as a result of direct vasodilatation, while enflurane acts as a negative inotropic agent. Isoflurane therefore became a popular choice as the supplement to narcotics for patients undergoing revascularization. However, in 1983 Reiz *et al.* published a study of 21 patients undergoing coronary arterial bypass grafting who were given 1 per cent end-tidal isoflurane. Electrocardiographic evidence of myocardial ischaemia was seen in 10 out of 21 patients, and it was concluded that isoflurane was a direct coronary vasodilator that had caused ischaemia as a result of a coronary steal.

Since that time, enormous controversy has been generated over the use of isoflurane in patients with coronary arterial disease undergoing both cardiac and non-cardiac operations. Numerous studies in humans and animals have produced varying results. Moffitt *et al.* measured coronary-sinus flow and oxygen content in humans and concluded that isoflurane dilated coronary vessels to a significantly greater degree than equipotent doses of halothane and enflurane. In contrast, Tarnow *et al.* found that, compared with other agents, isoflurane actually protected patients against pacing-induced ischaemia. Buffington *et al.*, using a microsphere technique to measure coronary blood flow in the dog, demonstrated coronary steal with isoflurane but not with halothane: although equipotent anaesthetic doses of the agents were used, the resulting haemodynamics were significantly different between groups; halothane reduced the mean perfusion pressure from the control of 55 mmHg to 51 mmHg, while isoflurane decreased it to 41 mmHg. These differing pressures might have obscured the effects of the two agents, but, although that is a significant flaw, this is the only study to demonstrate coronary steal directly. However, it is not analogous to the clinical situation where inhalational agents are titrated against

haemodynamic effects. In *in vitro* preparations of isolated segments of coronary artery, neither halothane nor isoflurane are direct vasodilators.

Two large clinical studies evaluated outcome in coronary arterial bypass surgery using various anaesthetic agents. Slogoff and Keats prospectively randomized 1012 patients to receive enflurane, halothane, isoflurane, or sufentanil. There was no difference between groups in the incidence of pre-bypass ischaemia (28–33.5 per cent), postoperative myocardial infarction (3.6–4.7 per cent), or death (1.2–2.4 per cent). Hypotension occurred twice as often in patients receiving an inhalational agent rather than sufentanil, and hypertension was twice as common in patients given sufentanil. The incidence of tachycardia (heart rate greater than 110 beats/min) was no different with different agents. Tachycardia was the only haemodynamic abnormality associated with a significantly increased incidence of ischaemia: 48 per cent of patients with tachycardia became ischaemic compared to 29.5 per cent of those with no haemodynamic abnormalities ($p<0.05$). Tuman *et al.* prospectively studied 1094 patients undergoing coronary revascularization who received fentanyl, sufentanil, diazepam plus ketamine, or halothane. During the intravenous anaesthesia, halothane, enflurane, or isoflurane were used as a supplemental anaesthetic. There was no difference between groups in the incidence of postoperative myocardial infarction, low cardiac-output state, or death. Thus, no clinical studies have demonstrated any difference in outcome in patients with coronary arterial disease receiving isoflurane rather than other inhalational or intravenous agents.

Although there is some experimental evidence to support the concept of isoflurane-induced coronary steal, the data are often contradictory and no study has shown it to be a clinically relevant phenomenon. Because isoflurane maintains cardiac output, unlike the other inhalational agents, which are potent myocardial depressants, it continues to be commonly used as an anaesthetic agent in patients with coronary arterial disease.

The use of intravenous anaesthetic agents such as propofol and midazolam has been extensively investigated in cardiac anaesthesia. Propofol has an especially desirable pharmacokinetic profile for use as a component of a ‘fast track' anaesthetic protocol but has a number of undesirable haemodynamic effects, in particular hypotension. Propofol is undoubtedly a systemic vasodilator, with effects on both arteriolar tone and venous capacitance; there remains controversy over whether it is also a direct myocardial depressant. Most of the studies undertaken have examined the effects of propofol in patients with ischaemic heart disease but normal left-ventricular function. Sherry *et al.* evaluated the effects of a propofol/fentanyl combination in patients with low cardiac-output states and concluded that the use of propofol was acceptable. Phillips *et al.* studied patients undergoing aortic-valve surgery and found that propofol, although causing some myocardial depression, was not contraindicated.

Benzodiazepines such as midazolam are excellent hypnotic and amnesiac agents, but in combination with opiates at induction may produce severe hypotension and haemodynamic instability.

The major drawback to the use of total intravenous anaesthesia is the need for a suitable drug-delivery system comparable to that of the volatile anaesthetic vaporizer. The infusion rate of an intravenous drug does not give much information about its resulting plasma concentrations, the pharmacokinetics of which are profoundly affected by age, cardiopulmonary bypass, haemodilution, and hypothermia, amongst other factors. Computer-controlled, pharmacokinetic-based, drug-infusion systems can rapidly achieve and maintain physician-specified, blood concentrations of anaesthetic, which the anaesthetist can titrate against the clinical response. Unfortunately, the optimal target concentrations for intravenous anaesthetic agents in specific clinical situations have yet to be defined.

In summary, it is increasingly uncommon to rely on a single drug for the induction and maintenance of anaesthesia. The most popular anaesthetic techniques for patients undergoing cardiac surgery are those in which synthetic opiates are given in moderate to high dose, supplemented by other hypnotic agents. The choice of technique should be based on the patient population and local practices of intensive-care management. If early extubation is desired, the combination of a small to moderate dose of narcotic with inhalational agents or total intravenous anaesthesia is preferable. In unstable patients or those with poor ventricular function, higher doses of narcotics may be employed to avoid myocardial depression and other undesirable haemodynamic effects. In these patients the requirement for postoperative mechanical ventilation is likely to be controlled by factors other than the duration of opiate effect.

Intraoperative ischaemia and preoperative medications

Although studies have shown that intraoperative ischaemia is common in patients undergoing coronary revascularization, the prognostic significance of these episodes is controversial. In 1518 patients undergoing coronary arterial bypass grafting, the occurrence of pre-bypass electrocardiographic evidence of ischaemia correlated with the incidence of postoperative myocardial infarction. Another study found no relation between either preoperative or pre-bypass electrocardiographic ischaemia and postoperative outcome in 50 patients undergoing coronary arterial bypass grafting. Post-bypass electrocardiographic changes suggestive of ischaemia were also common (40 per cent of patients). Leung *et al.* examined the prognostic significance of intraoperative ischaemia as defined by new abnormalities of regional wall motion detected by transoesophageal echocardiography. Although neither pre-bypass abnormalities of regional wall motion nor electrocardiographic ischaemia correlated with poor outcome, post-bypass ischaemia was predictive of adverse events. A new abnormality of regional wall motion detected by echocardiography is a more sensitive and specific indicator of ischaemia than are electrocardiographic changes.

Patients with coronary arterial disease are managed medically with b-adrenergic-blocking drugs, calcium entry-blocking drugs, and nitrates, often in combination. Several studies have examined the effects of preoperative medications on the incidence of intraoperative myocardial ischaemia. In all studies, preoperative b-blockade was associated with a significantly lower incidence of intraoperative ischaemia compared to therapy with calcium entry-blocking agents. Intraoperative electrocardiographic evidence of ischaemia was seen in 34 per cent of 444 patients receiving b-blocking agents and in 53 per cent of those not receiving these drugs. Tachycardia (heart rate above 109) was more common in patients not receiving b-blockers and doubled the incidence of perioperative ischaemia. Other studies have confirmed that tachycardia is the haemodynamic change most commonly associated with intraoperative ischaemia; the majority of intraoperative ischaemic events are not associated with any change in haemodynamics.

All preoperative anti-ischaemic agents should be continued in patients undergoing coronary revascularization. b-Adrenergic-blocking drugs are associated with a significantly lower incidence of perioperative ischaemia, although the prognostic significance of pre- and post-bypass ischaemia remains uncertain.

Cerebral protection

Neurological dysfunction is one of the major causes of morbidity after heart surgery. The incidence of postoperative stroke is around 1 to 2 per cent and that of early neuropsychological dysfunction between 40 and 80 per cent, the exact figure depending somewhat on the subtlety of the examining technique. These percentages have remained relatively static despite advances in surgical skills, anaesthetic techniques, and bypass technology. Risk factors for postoperative cerebral dysfunction include a history of previous stroke, transient ischaemic attacks or cerebrovascular disease, hypertension, diabetes, atheroma of the ascending aorta, prolonged bypass time, periods of profound hypotension, and, above all, advanced age. The incidence of stroke after heart surgery in the over-75-year-old is about 9 per cent and the increasing age of the population undergoing surgery may be one reason for lack of success in reducing morbidity from cerebral damage. As the population of patients undergoing surgery continues to age, so the incidence of age-related adverse effects will increase unless better prophylactic or therapeutic measures are used. The cost benefit of reducing cerebral damage is substantial: fatality, length of intensive-care and hospital stay, and hospital costs are all greatly increased by stroke.

Most cerebral damage is thought to be due to micro- and macroembolization from atheroma, thrombus and valve debris, gas bubbles, and particulate matter from the bypass circuit producing focal ischaemia. Open-chamber surgery is associated with a two to three times greater risk of stroke than closed-chamber bypass surgery, although the incidence of subtle cognitive changes is similar. Focal ischaemia may be exacerbated by the presence of intra- and extracranial cerebrovascular disease with compromised autoregulation and poor collateral flow. Global ischaemia may occur during periods of acute hypotension or total circulatory arrest. The various models of cerebral ischaemia can be defined as shown in Table 1.

Type	Description
Focal	Vascular occlusion: stroke, embolus
	Space-occupying mass: oedema, tumour, haemorrhage
Global, incomplete	Shock; hypotension; cardiopulmonary bypass
Global, complete	Cardiac arrest; deep hypothermic circulatory arrest

Table 1 Types of cerebral ischaemia

The subject of cerebral protection is complex and controversial; measures currently employed in efforts to prevent cerebral damage include hypothermia, blood-sugar control, nimodipine, and various anaesthetic agents including barbiturates, isoflurane, propofol, and etomidate. The use of membrane oxygenators and arterial-line filters may reduce cerebral embolization. Optimal blood gas and acid-base management, and alterations in surgical technique, such as identifying areas of aortic plaque and avoiding cannulation through them, may also help to reduce cerebral damage. Transoesophageal echocardiography has been advocated for the detection of atheromatous disease in the aorta to enable the surgeon to avoid dislodging plaques, which may cause intraoperative stroke.

Currently, the major areas of research interest are in the modulation of pathological processes initiated by ischaemia and reperfusion. Massive release of excitatory neurotransmitters (predominantly glutamate) into the extracellular space during reperfusion following an ischaemic episode triggers a cascade of events leading to a build up in intracellular calcium and sodium. This in turn causes degradation of cellular membranes, activating a wide range of receptors and second-messenger systems. Mitochondrial calcium accumulation promotes the generation of oxygen free radicals, which further stimulates the inflammatory process and can lead to additional injury and neuronal death during reperfusion. Drugs that can inhibit the release of excitatory transmitters or otherwise modulate this cascade of events may potentially be effective in reducing secondary brain injury. A plethora of drugs is under investigation in the search for agents that block the actions of glutamate and other excitotoxic amino acids. Free-radical scavengers may affect another link in the chain of events in reperfusion injury, and a number of other drugs have been reported to have previously unsuspected neuroprotective properties, including aprotinin and acadesine, an adenosine-regulating agent. The challenge remains to find a neuroprotective agent with proven efficacy in improving neurological outcome after cardiac surgery.

Hypothermia

Hypothermia remains the mainstay of cerebral protection for anticipated and predictable episodes of cerebral ischaemia. Profound hypothermia to temperatures of 15 to 18°C, with circulatory arrest or low-flow bypass, is used in the repair of complex congenital lesions in children and in adults undergoing aortic-arch surgery. Cooling decreases the rate of all brain activity, both electrical and metabolic. For every 10°C decrease in temperature, the cerebral metabolic rate decreases by a factor of 2.2; at 28°C, it is approximately 45 per cent of normal. At 18°C the electroencephalogram is flat and the cerebral metabolic rate is 20 per cent of normal. Cooling protects the brain by diminishing the depletion of high-energy phosphate and inhibiting the release of excitatory neurotransmitters. Profound hypothermia allows periods of complete brain ischaemia of around 30 min to be tolerated without overt brain damage, but longer periods of circulatory arrest may have serious neurological consequences. The maximum duration of circulatory arrest that does not cause permanent impairment in the central nervous system is uncertain. In an attempt to reduce the risks from, and prolong the safe duration of, deep hypothermic circulatory arrest, techniques such as continuous retrograde cerebral perfusion and asanguinous hypothermic cerebral perfusion (so-called cerebroplegia) have been used, but remain essentially unproved. That hypothermia does not protect totally is undisputed. Bellinger *et al.* made a prospective, randomized study of infants undergoing the arterial-switch procedure that convincingly demonstrated the superiority of hypothermic low-flow bypass over total circulatory arrest in long-term neurodevelopmental outcome.

Conventional wisdom has extended the undoubted value of hypothermia in total circulatory arrest to routine use during cardiopulmonary bypass; some authorities suggest that mild to moderate hypothermia reduces cerebral dysfunction after cardiopulmonary bypass. If this is true, then it would seem likely that the use of normothermic bypass would be associated with an increase in postoperative cerebral dysfunction. Although one randomized study purported to show a three-fold increase in the incidence of stroke after normothermic as compared to hypothermic perfusion, other large experiences of normothermic bypass have not demonstrated excessive rates of stroke.

The use of hypothermia is not without consequence: it is associated with dysrhythmias, increased systemic vascular resistance, a left shift of the oxygen–haemoglobin dissociation curve, impaired coagulation, and a considerable metabolic demand for oxygen with rewarming. In patients undergoing vascular surgery, unintentional postoperative hypothermia is associated with an increased incidence of myocardial ischaemia during the early postoperative period. Deep hypothermia requires additional time on cardiopulmonary bypass for cooling and rewarming, and is associated with the consumption of clotting factors and with renal dysfunction.

Carbon dioxide management

Blood cooling increases carbon dioxide solubility and, if all else is constant, causes a fall in $PaCO_2$ and a relative alkalosis. The $PaCO_2$ has profound effects on cerebrovascular resistance, with hypercarbia causing increases and hypocarbia decreases in cerebral blood flow. Acid–base management during hypothermic cardiopulmonary bypass thus has significant effects on cerebral blood flow. With the ‘alpha-stat' method of pH regulation the pH is maintained at 7.4 and $PaCO_2$ at 5.3 kPa as measured at the temperature in the blood-gas analyser (37°). In the ‘pH-stat method' a correction factor for the patient's body temperature is applied to the measured $PaCO_2$, and the corrected $PaCO_2$ is maintained at about 5.3 kPa with a pH of 7.4. The pH-stat method requires the addition of exogenous carbon dioxide to the oxygenator gas mixture as the temperature falls, and when measured at the temperature of the blood-gas analyser the $PaCO_2$ is around 6.3 to 6.5 kPa. The resultant respiratory acidosis counteracts the alkaline shift with hypothermia. Under mildly hypothermic conditions, pH-stat management results in pressure-dependent cerebral blood flow. In contrast, alpha-stat management maintains cerebral autoregulation and metabolic coupling. The use of alpha-stat management has been associated with a reduced incidence of postoperative neuropsychiatric dysfunction in adults. This may be due to a reduced embolic load secondary to lower cerebral blood flow. In contrast, however, in infants undergoing deep hypothermic circulatory arrest there is some evidence to suggest that the luxury cerebral perfusion induced by the pH-stat strategy may induce more even brain cooling and improve long-term outcome. More work is needed to define the optimal acid–base management in all situations.

Anaesthetic agents

Most general anaesthetics decrease the cerebral metabolic rate by decreasing the electrical activity of the brain. If cerebral blood flow falls in parallel with metabolic rate, then the potential exists to reduce the embolic load delivered to the cerebral circulation. Barbiturates cause cerebral vasoconstriction with decreased cerebral blood flow, decreased intracranial pressure, and a fall in cerebral metabolic rate by as much as 50 or 60 per cent due to diminished cerebral electrical activity. Barbiturates are effective in incomplete global or focal ischaemia: animal studies show that barbiturates, given before or during an ischaemic event, can prolong the brain's tolerance to the insult, resulting in a smaller infarct. There are fewer data available in humans, but Nussmeier *et al.* reported an improvement in neuropsychiatric function in patients who received an average dose of 40 mg/kg thiopentone to produce electroencephalographic suppression on bypass during cardiac procedures requiring an open ventricle. Patients receiving thiopentone required more vasopressor support at the termination of bypass, and extubation was delayed. However, others have been unable to demonstrate a protective effect of barbiturates in patients undergoing coronary arterial surgery. Intraoperative barbiturate therapy is currently considered to be useful where focal ischaemia may occur, such as in carotid endarterectomy, cerebral bypass procedures, or when clipping intracerebral aneurysms. Dose regimens vary among studies, but electroencephalographic monitoring allows the titration of thiopentone to the degree of electroencephalographic suppression. Propofol and etomidate both have similar effects to thiopentone.

Inhalational agents produce dose-dependent reductions in cerebral metabolic rate but increase blood flow, which can be detrimental in the presence of increased intracerebral pressure. Although not all studies are in agreement, there is good evidence that isoflurane provides cerebral protection during focal and incomplete ischaemia as a result of metabolic suppression, and it may be the anaesthetic of choice for patients in whom hypotension needs to be induced.

Calcium antagonists

Ischaemia disrupts intracellular calcium homeostasis. There is accumulating evidence that the calcium entry-blocking drug nimodipine provides cerebral protection in a variety of circumstances, although the precise mechanism of protection is unknown. Most importantly, nimodipine may be effective even when given after an ischaemic insult; it has been shown to reduce the extent of cerebral infarction when given up to 48 h after the acute event. While nimodipine appears to provide no benefit after head injury, it has found a place in the management of patients with cerebral vasospasm after subarachnoid haemorrhage. The only trial so far in patients undergoing valve surgery was terminated prematurely because of increased bleeding complications and fatalities in the treated group, with no evident beneficial effect of nimodipine. Nevertheless, the use of calcium blockers deserves further investigation.

Hyperglycaemia

Clinical and laboratory data suggest that glucose may play a part in the damaging effects of cerebral ischaemia. In the absence of oxygen, glucose is metabolized to lactic acid, which may increase tissue damage. Studies in animals have shown that elevated glucose during cerebral ischaemia is associated with a worse neurological outcome. Patients with hyperglycaemia following ischaemic stroke have a poorer neurological outcome than those with a normal serum glucose. Current recommendations are to avoid the intraoperative use of glucose-containing solutions, and to monitor serum glucose closely in patients at risk for cerebral ischaemia.

Further reading

Belisle S, Hardy JF. Hemorrhage and the use of blood products after adult cardiac operations; myths and realities. *Annals of Thoracic Surgery* 1996; **62**: 1908–17. [An excellent and up-to-date review

of the literature relating to bleeding and transfusion of blood products in cardiac surgery.]

Bellinger DC *et al.* Developmental and neurologic status of children after heart surgery with hypothermic circulatory arrest or low-flow cardiopulmonary bypass in infant heart surgery. *New England Journal of Medicine* 1995; **332**: 549–55. [A randomized study of infants undergoing the arterial-switch operation, which demonstrates that heart surgery performed with deep hypothermic circulatory arrest as the predominant support strategy is associated with an increased risk of neurological abnormalities when compared with surgery performed using hypothermic low-flow bypass.]

Birkmeyer JD *et al.* Cost-effectiveness of preoperative autologous donation in coronary artery bypass grafting. *Annals of Thoracic Surgery* 1994; **57**: 161–9. [A prospective, multi-institution study, which showed that preoperative autologous donation was not cost-effective, and suggested that the risks of donating autologous blood might outweigh the benefits.]

Hardy JF, Belisle S, Janvier G, Samoma M. Reduction in requirements for allogeneic blood products: nonpharmacologic methods. *Annals of Thoracic Surgery* 1996; **62**: 1935–43. [A review of the evidence for non-pharmacological methods of blood conservation in cardiac surgery.]

Janssens M, Hartstein G, David JL. Reduction in requirements for allogeneic blood products: pharmacologic methods. *Annals of Thoracic Surgery* 1996; **62**: 1944–50. [A literature review of the use of drugs in the prophylaxis and treatment of bleeding during and after cardiac surgery.]

P>**Leung JM, O'Kelly B, Browner WS, Tubau J, Hollenberg M, Mangano DT.** Prognostic importance of postbypass regional wall-motion abnormalities in patients undergoing coronary artery bypass surgery. *Anesthesiology* 1989; **71**: 16–25. [Segmental wall-motion abnormalities seen on transoesophageal echo following cardiopulmonary bypass, but not those in the pre-bypass period, appeared to predict adverse cardiac outcome.]

Lowenstein E, Hallowell P, Levine F, Daggett W, Austen G, Laver M. Cardiovascular response to large doses of intravenous morphine in man. *New England Journal of Medicine* 1969; **281**: 1389–93. [A pivotal paper demonstrating that large doses of morphine sulphate could be used with safety in patients with minimal circulatory reserve.]

Mora CT, Dudek C, Torjman MC, White PF. The effects of anaesthetic technique on the hemodynamic response and recovery profile in coronary revascularization patients. *Anesthesia and Analgesia* 1995; **81**: 900–10. [After induction with fentanyl, 90 patients were randomized to receive anaesthetic maintenance with one of four techniques (enflurane, fentanyl, propofol, or thiopentone). Short-acting drugs were, not surprisingly, associated with earlier recovery times.]

Murkin JM, ed. Protection of the brain during cardiopulmonary bypass: diagnosis, etiology and therapeutics. *Journal of Cardiothoracic and Vascular Anesthesia* 1996; **10**; 2–139. [An in-depth, single-topic review.]

Nussmeier NA, Slogoff S. Neuropsychiatric complications after cardiopulmonary bypass: cerebral protection by a barbiturate. *Anesthesiology* 1986; **64**: 164–70. [High-dose thiopentone (mean dose 39.5 mg/kg) given to patients undergoing cardiac surgery reduced postoperative cerebral dysfunction but at the cost of an increased need for inotropes and a prolonged recovery time.]

Ovrum E, Holen EA, Tangen G. Consistent nonpharmacological blood conservation in primary and reoperative coronary artery bypass grafting. *European Journal of Cardiothoracic Surgery* 1995; **9**: 30–5. [A study of simple, non-pharmacological methods of blood conservation showed that it is possible to greatly reduce blood utilization. An amazing 97 per cent of 2326 patients were managed without autologous predonation or homologous transfusion.]

Philbin DM, Rosow CE, Schneider RC, Koski G, D'Ambra MN. Fentanyl and sufentanil anesthesia revisited: how much is enough? *Anesthesiology* 1990; **73**: 5–11. [A study that demonstrated that suppression of haemodynamic and hormonal responses during coronary arterial bypass grafting surgery is not related to opioid dose.]

Reiz S, Balfors E, Sorenson MB, Ariala S, Friedman A, Truedsson H. Isoflurane—a powerful coronary dilator in patients with coronary artery disease. *Anesthesiology* 1983; **59**: 91–7. [This much-quoted study involved the administration of 1 per cent end-tidal isoflurane, which decreased mean arterial pressure by 40 per cent. Myocardial ischaemia was noted in 10 out of 21 patients and the authors suggest that isoflurane might contribute to regional myocardial ischaemia by virtue of coronary vasodilatation not related to normal autoregulation.]

Shaw PJ *et al.* Early neurological complications of coronary artery bypass surgery. *British Medical Journal* 1985; **291**: 1384–7. [A prospective study of 312 patients undergoing elective surgery demonstrated a high incidence of lesser neurological abnormalities that were only detectable by detailed neurological examination.]

Slogoff S, Keats AS. Randomised trial of primary anesthetic agents on outcome of coronary artery bypass operations. *Anesthesiology* 1989; **70**: 179–88. [An important prospective study of over 1000 patients randomized to receive either enflurane, halothane, isoflurane or sufentanil as the primary anaesthetic for coronary arterial surgery. New ST-segment depression was noted in 30 per cent of patients, but the incidence of ischaemia, infarction, or death was not different between groups.]

Tuman KJ, McCarthy RJ, Spiess BD, DaValle M, Dabir R, Ivankovich AD. Does choice of anesthetic agent significantly affect outcome after coronary artery surgery. *Anesthesiology* 1989; **70**: 189–98. [A prospective study of 1094 patients receiving one of five primary anaesthetic techniques (high-dose fentanyl, moderate-dose fentanyl, sufentanil, ketamine with diazepam, or halothane). Multivariate analysis of data suggests that many other factors are significantly more important as determinants of outcome than is anaesthetic technique.]

Tuman KJ, McCarthy RJ, Najafi H, Ivankovich AD. Differential effects of advanced age on neurologic and cardiac risks of coronary artery operations. *Journal of Thoracic and Cardiovascular Surgery* 1992; 1510–17. [A prospective study of 2000 patients found the incidence of postoperative neurological events in patients aged over 75 years to be twice that of patients aged 65 to 74 and nine times greater than that in patients aged below 65 years.]

40.6 Cardiovascular monitoring and postoperative care

Catherine R. Grebenik and Michael E. Sinclair

Cardiovascular monitoring
Clinical monitoring
Direct monitors of the cardiovascular system
Indirect monitors of cardiovascular function
Postoperative cardiac intensive care
Environment
Maintenance of tissue perfusion and oxygenation
Fluid replacement
Ventilation
Postoperative myocardial ischaemia and early extubation
Analgesia
Prevention and treatment of complications
Hypertension
Bleeding
Cardiac tamponade
Dysrhythmias
Low cardiac output
Further reading

Cardiovascular monitoring

The aims of perioperative cardiovascular monitoring are to detect and quantify potentially dangerous haemodynamic changes, and to direct treatment and assess its effects.

Routine cardiovascular monitoring during and after cardiac surgery includes clinical monitoring and the continuous measurement and display of the electrocardiogram, arterial pressure, central venous pressure, and pulse oximetry. Intermittent measurements of urine output, blood gases, and the core–peripheral temperature gradient also provide useful indirect information about the state of the cardiovascular system. Other variables that may be monitored in high-risk patients include left atrial pressure, pulmonary arterial pressures, and cardiac output. Transoesophageal echocardiography has become a widely used tool in the last decade and complements routine haemodynamic monitoring by providing information about cardiac structure and function that may aid diagnosis as well as treatment.

Clinical monitoring

Basic clinical monitoring consists of the direct observation of the patient by anaesthetist, surgeon, or nurse. It includes the assessment of skin colour and capillary refill time, palpation of peripheral pulses, skin temperature and sweating, and observation of respiratory rate and rhythm, chest movement and the noises of breathing. This monitoring requires no mechanical or electrical instrumentation and its importance must not be underestimated. Information obtained from mechanical monitors is subject to many potential errors, and should be interpreted in relation to the clinical picture.

Direct monitors of the cardiovascular system

Electrocardiogram

The electrocardiogram provides continuous information about heart rate and rhythm, and analysis of changes in the ST segments may help to detect myocardial ischaemia. Measured at 60 ms from the J point, horizontal or downward depression of more than 1 mm or elevation of the ST segment of greater than 1.5 mm is an indicator of ischaemia. New monitors incorporate automated ST-segment monitoring. A multiple-lead electrode system consisting of four limb leads and a fifth at the V5 position allows selection of seven out of the 12 standard leads, V5 and II being the most helpful for detecting anterolateral and inferior ischaemia, respectively. Electrocardiographic monitoring is subject to many types of interference in the operating theatre, and artefacts are common.

Arterial blood pressure

Direct monitoring of arterial blood pressure is essential for all but the simplest of cardiac surgical procedures so that rapid changes in blood pressure can be followed accurately. In addition, an arterial line permits repeated blood gas estimations. Indirect pressure measurements (using a blood-pressure cuff) may be inaccurate, particularly in obese, hypertensive, hypothermic, or shocked patients.

Direct arterial-pressure monitoring consists of an indwelling arterial cannula connected to a pressure transducer by a length of rigid, fluid-filled manometer tubing. Amplification of the signal from the transducer is used to create a pressure trace on an oscilloscope. A digital read-out of systolic, diastolic, and mean pressure is produced by averaging several beats. The arterial line is kept open either by intermittent flushing or by continuous infusion of a small volume of heparinized saline.

Sites of arterial pressure monitoring

The radial artery is most commonly used for pressure monitoring. It is accessible, easy to cannulate, and short-term cannulation has few complications. Allen's test can be used to assess the adequacy of collateral circulation to the hand, but the complication rate from radial cannulation is so low that this test is of little practical relevance. Other possible sites of arterial cannulation include the brachial, axillary, femoral, posterior tibial, and dorsalis pedis arteries. The femoral artery is avoided by those who believe it to be a potential source of infection with organisms from the perineum. Prolonged femoral arterial cannulation has, however, been reported without major complications.

Inaccuracy in monitoring

Transduced blood pressure is subject to inaccuracies due to transducer position, calibration, and the mechanical characteristics of the recording system. Before use the system is zeroed against a reference point at the level of the right atrium.

The pressure measured from a peripheral part of the arterial system does not always accurately reflect the intra-aortic pressure. Progressive alteration of the arterial pressure wave occurs as it is transmitted through the vascular system; measured systolic and diastolic pressures may be either higher or lower than aortic pressures, depending on the degree of vasomotor tone and the rigidity of the vascular tree. Aortic to radial pressure gradients are particularly common immediately after cardiopulmonary bypass.

Central venous pressure

Central venous pressure is measured as an indication of right ventricular end-diastolic volume, using a catheter in a central (intrathoracic) vein. Venous pressure is affected by a number of interrelated factors: circulating blood volume, right and left heart function, peripheral and pulmonary vascular resistance, and intrathoracic pressure. In patients with normal ventricular and pulmonary function, changes in central venous pressure occur roughly in parallel with left heart filling pressures. This relation does not persist in pathological conditions, however.

Central venous pressure is measured with reference to a fixed zero point, at the level of the right atrium. The position of the zero point will affect the absolute value of the central venous pressure; in a supine patient the pressure measured with reference to the mid-axillary line will be higher than that measured with reference to the sternal angle. Any change in central venous pressure will be of the same magnitude, irrespective of the zero point.

Single measurements of central venous pressure are rarely helpful unless the measurement lies well outside the normal range (1–10 mmHg in a patient breathing spontaneously). Serial measurements and the response of central venous and arterial pressures to a fluid challenge (a rapid infusion of 100–200 ml) are, however,

useful indicators of circulatory status and right ventricular function.

The units of central venous pressure depend on the method of measurement: if a simple saline manometer is used the measurement will be in centimetres of water (cmH_2O) but with an electronic pressure transducer the measurement is in mmHg (1 cmH_2O = 1.36 mmHg).

Routes of central venous cannulation

Many techniques of central venous cannulation have been described. The most reliable is via the right internal jugular vein, since its short, straight course almost assures that the catheter tip will pass into the superior vena cava or right atrium. For short-term perioperative use the internal jugular veins are preferred to the subclavian veins, since there is a lower incidence of major complications such as pneumothorax and arterial puncture. Central venous cannulation via peripheral arm veins is likely to provide correct intrathoracic placement of the catheter tip in only about 40 per cent of patients.

Left atrial pressure

In many patients with left-sided heart disease or pulmonary disorders the central venous pressure is not a reliable indicator of left heart filling pressures. In these patients, left atrial pressure can be measured directly via a catheter inserted at the time of surgery. The use of left atrial catheters carries a number of unique risks, including the possibility of air or clot emboli to the systemic circulation and the potential for bleeding at the time of removal.

Pulmonary arterial pressures and cardiac output

Pulmonary arterial pressures can be measured with a flow-directed pulmonary arterial catheter, also known as a Swan–Ganz catheter. The balloon-tipped catheter is inserted into the central venous system and ‘floated' through the right heart into the pulmonary artery using the changes in pressure waveform as the catheter passes through the chambers of the heart to guide insertion (Fig. 1). The catheter can be used to measure central venous pressure, pulmonary arterial systolic, diastolic, and mean pressures, and pulmonary capillary-wedge pressure (also known as pulmonary artery occlusion pressure), which is the pressure measured when the catheter balloon is inflated so that the catheter is impacted into a peripheral pulmonary artery (Fig. 2). The catheter then registers the pressure transmitted retrogradely from the left atrium, providing an estimate of left heart filling pressure.

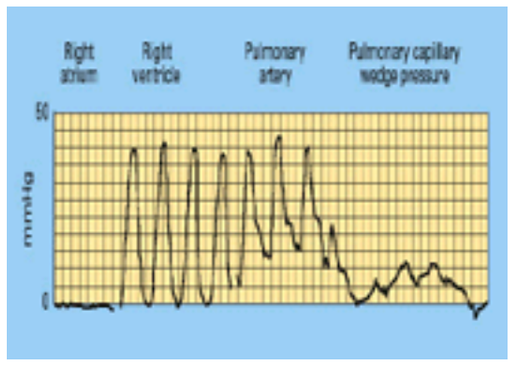


Fig. 1. Pressure changes seen during the passage of a pulmonary arterial catheter. The waveform becomes markedly pulsatile as the catheter passes through the tricuspid valve into the right ventricle. Passage into the pulmonary artery is shown by a rise in the diastolic pressure. Advancing the catheter further causes it to impact into a peripheral pulmonary arteriole and the waveform changes once again to show pulmonary capillary-wedge pressure. On deflation of the balloon the pulmonary arterial pressure trace should return.

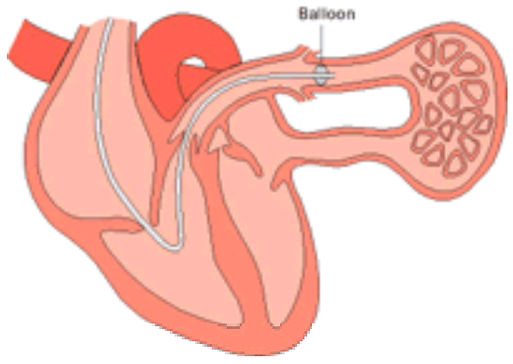


Fig. 2. The pulmonary arterial catheter with balloon inflated is wedged into a pulmonary arteriole. In this position the tip of the catheter is isolated from its upstream arterial pressure and records the pressure transmitted from the left atrium through the pulmonary capillary bed.

The pulmonary arterial catheter, when fitted with a thermistor, can also be used to measure cardiac output by thermodilution. Once the cardiac output is known, stroke volume, systemic and pulmonary vascular resistance, and left and right ventricular stroke-work indices can all be calculated. Mixed venous blood can be sampled from the catheter, or an oximetric pulmonary arterial catheter can be inserted to permit continuous monitoring of mixed venous oxygen saturation. Advantages of placing a pulmonary arterial catheter include the ability to assess pharmacological and fluid therapy by serial measurement of filling pressures and cardiac output. In patients with abnormal right-ventricular function or pulmonary hypertension, a pulmonary arterial catheter has significant advantages over a central venous catheter, which is not a sensitive monitor of left heart filling pressure in these situations. The principal disadvantages of the pulmonary arterial catheter are cost and the incidence of complications, albeit low, such as pulmonary arterial rupture and pulmonary infarction.

Controversy exists over the need to place a pulmonary arterial catheter in all patients undergoing cardiac surgery. A large study involving more than 1000 patients undergoing coronary arterial surgery concluded that the use of a pulmonary arterial catheter did not favourably alter outcome. However, serious flaws included a lack of patient randomization, differences in anaesthetic techniques, greater use of pressors in patients electively receiving a pulmonary arterial catheter, older age and higher risk categories in patients receiving catheters, and a higher incidence of reoperation in these patients. The effects of the monitoring alone cannot, therefore, be determined. In the field of critical care, a recent large observational study by Connors *et al.* revealed increased fatality in patients who had a pulmonary arterial catheter inserted when compared with matched patients who did not receive a catheter. However, this study has also been much criticized. The Society of Critical Care Medicine has issued a consensus statement that reviews the indications for, and evidence supporting the use of, pulmonary arterial catheters.

Despite recent controversy over its use, the pulmonary arterial catheter has contributed greatly to knowledge and understanding of cardiovascular pathophysiology and pharmacology, and it is undoubtedly useful in identifying haemodynamic abnormalities that are not obvious from clinical examination. In view of their expense and potential complications, it seems reasonable to restrict the use of pulmonary arterial catheters in cardiac surgery to patients ith poor ventricular function, unstable angina, recent infarction, valvular abnormalities, and those undergoing emergency operations. For stable patients with normal left-ventricular function and good distal vessels who are expected to have short cardiopulmonary bypass times, a central venous catheter is adequate.

Pulse oximetry

Oxygen saturation can be monitored non-invasively and continuously by pulse oximetry. The oximeter probe measures the pulsatile absorption of light from two light-emitting diodes and calculates the ratio of oxygenated to deoxygenated haemoglobin. The result is a waveform whose magnitude depends on the pulse pressure and the arterial oxygen saturation. Because of the sigmoid shape of the oxygen dissociation curve, a large change in PaO_2 above 10 kPa produces only a small change in saturation, but if the PaO_2 is less than 10 kPa a small change in PaO_2 produces a larger change in oxygen saturation. Pulse oximetry is a valuable and reassuring monitor of oxygen saturation during weaning from mechanical ventilation and after extubation. Unfortunately the accuracy, and therefore the usefulness, of the

instrument is greatly reduced in conditions in which skin blood flow is reduced, such as hypothermia or low cardiac output.

Transoesophageal echocardiography

Transoesophageal echocardiography is a form of relatively non-invasive monitoring that provides continuous, two-dimensional echocardiographic images from a transducer positioned in the oesophagus. The apparatus consists of a crystal mounted on the tip of a gastroscope. This is inserted in the oesophagus under direct vision following anaesthetic induction and intubation (Fig. 3). Transoesophageal echocardiography allows continuous monitoring of a two-dimensional image of the cardiac structures during surgery. It complements routine haemodynamic monitoring and gives information that cannot be obtained by other means. Left ventricular function, volume, and wall-motion abnormalities are assessed from the mid-papillary muscle, short-axis view of the left ventricle (Fig. 4). This view is considered optimal because all three major coronary arteries supply a portion of the ventricle at this level. It is a highly sensitive method of detecting segmental wall-motion abnormalities and has been increasingly used to monitor intraoperative myocardial ischaemia.



Fig. 3. An example of an Advanced Technology Laboratories (ATL) transoesophageal echocardiography probe. Two control levers at the base of the probe allow anterior, posterior, and lateral manipulation of the probe tip in the oesophagus.

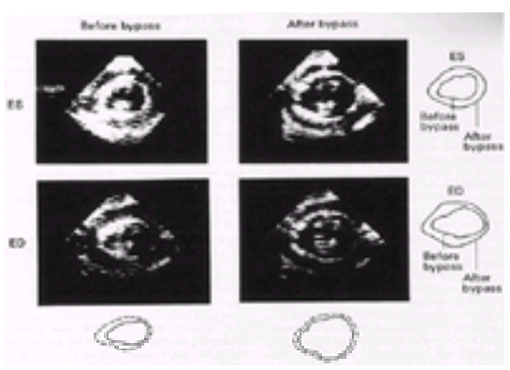


Fig. 4. Mid-papillary, short-axis view of the left ventricle. Measurement of end-systolic (ES) and end-diastolic (ED) areas allows calculation of ejection fraction areas pre- and post-bypass. The tracings to the right of the echo images reveal an increase in both ES and ED areas post-bypass.

Advances in probe technology have fuelled an explosive increase in clinical applications. Pulse-wave and colour-flow Doppler capabilities allow evaluation of cardiac output, valvular function, and intracardiac shunts. Transoesophageal echocardiography is a highly sensitive tool considered essential in many centres to detect the presence of residual mitral regurgitation or paravalvular leaks following valve repairs and replacements. Newer models of the system contain biplane probes, which allow imaging in two planes. This modification allows the evaluation of thoracic aortic aneurysms and improves the assessment of regurgitant lesions. Transoesophageal echocardiographic systems with continuous-wave Doppler capabilities can measure blood-flow velocity with great accuracy and are particularly useful for the evaluation of patients with congenital heart disease. Paediatric-sized probes are available.

Transoesophageal echocardiography is considered a safe procedure, although the potential for oesophageal injury does exist. Oesophageal pathology such as varices, prior oesophageal surgery, or the presence of a stricture is a contraindication to placement of the probe.

Limitations to the usefulness of transoesophageal echocardiography include the inability to obtain images before and during anaesthetic induction, the ability to image in only a single plane in most systems, the significant expense of purchasing the equipment, and the requirement for additional training of the anaesthetist in echocardiography. Interpretation of wall-motion abnormalities is highly subjective and the operator needs to possess considerable technical and interpretative skills to avoid the possibility of providing misleading data that may produce inappropriate decisions and actions.

The practical benefits of the use of transoesophageal echocardiography remain unproved except in a few specific areas such as in assessment of the thoracic aorta and in mitral-valve surgery. Like the pulmonary arterial catheter, transoesophageal echocardiography has its advocates and devotees who use it for all cases, but its routine use for the monitoring of myocardial ischaemia has not so far been shown to improve outcome in a controlled trial.

Indirect monitors of cardiovascular function

Useful information about cardiovascular function can be gained by monitoring its effects on other organ systems.

Urine output

Urine output is critically dependent on adequate renal perfusion pressure and intravascular volume. In an adult, maintenance of a urine output of more than 0.5 ml/kg per hour without the use of diuretics is a reassuring sign of satisfactory cardiac output and renal blood flow.

Blood gas analysis

Inadequate cardiac output and consequent tissue hypoxia produces metabolic acidosis and a decrease in mixed venous oxygen saturation. Appropriate treatment will reverse these changes.

Temperature

The core–peripheral temperature gradient has been used as a guide to cardiac output, since peripheral vasoconstriction is one of the compensatory mechanisms for a fall in output. It is, however, an unreliable indicator of cardiovascular function, since it is affected by a number of variables including ambient temperature and the presence or absence of peripheral vascular disease.

Postoperative cardiac intensive care

The major differences between a cardiac surgical patient and a patient who has undergone non-cardiac major surgery are the profound physiological effects of cardiopulmonary bypass, combined with temporary myocardial dysfunction, haemodilution, cooling, and potential coagulopathy. The aims of postoperative care are the optimization of tissue perfusion and oxygenation, and the prevention and, if necessary, prompt treatment of complications.

Environment

Patients are nursed in a high-dependency area with the same monitoring facilities as in the operating theatre and with access to blood gas and electrolyte analysis. A small minority of patients who have multisystem disorders will require admission to an intensive-care unit, but the majority can be safely managed in a recovery area. This area should be close to the operating theatre, and there should be adequate space for equipment and the performance of emergency surgical procedures. A one-to-one nurse-to-patient ratio is essential in the immediate postoperative period, until the patient is haemodynamically stable and extubated. The majority of patients undergoing routine elective surgery will be fit to leave the critical-care area within 24 h of surgery.

Maintenance of tissue perfusion and oxygenation

Circulatory management during the postoperative period is an extension of the management begun during weaning from cardiopulmonary bypass. The aim is to achieve and maintain an adequate cardiac output to ensure oxygenation at the cellular level. This goal requires optimization of the determinants of cardiac output and of oxygen-carrying capacity. Some degree of decreased myocardial function is almost invariable in the immediate postoperative period as the heart recovers from the insult of surgery.

Fluid replacement

In common with all surgical patients, the postoperative cardiac patient requires fluids both for maintenance of normal homeostasis and to replace abnormal losses. Hypovolaemia is frequently seen in the early postoperative period due to excessive bleeding or diuresis, increased capillary permeability, and the effects of vasodilators or rapid rewarming. Blood is given to maintain a haematocrit of 25 to 30 per cent; autotransfusion of red cells salvaged from mediastinal drainage is used in some centres in an attempt to decrease the need for homologous blood transfusion.

Controversy exists over the employment of colloid as opposed to crystalloid solutions to maintain the circulation. In the United Kingdom the usual practice is to restrict crystalloid administration on the first postoperative day to 0.5 to 1.0 ml/kg per hour of 5 per cent glucose because of the obligatory retention of sodium and water after major surgery, and to use colloid solutions (plasma or plasma expanders) to increase the circulating volume. The gelatine derivatives Haemaccel and Gelofusine are the cheapest colloids and are just as effective as volume expanders as the more expensive solutions. Potassium is added to intravenous fluids to maintain extracellular potassium at or above 4.5 mmol/l and avoid hypokalaemia, which may be associated with dysrhythmias.

Ventilation

The use of controlled ventilation after cardiac surgery became widespread in the 1960s, with the belief that it allowed a smoother and more stable recovery from surgery. As surgical and anaesthetic techniques have improved, a number of studies have shown that earlier return to spontaneous respiration and early extubation are reasonable options in selected patients. The increase in demand for cardiac surgery and the need for cost containment have driven the trend towards earlier extubation. Prolonged intensive-care stay not only costs more but potentially limits the number of operations that can be performed. Early extubation improves patient comfort, decreases the need for sedative drugs, reduces nursing dependency, and may decrease morbidity and length of hospital stay.

There are few prospective randomized trials of early extubation after cardiac surgery. Quasha *et al.*, in 1980, compared early extubation (within 8 h of surgery) with late extubation (after 18 h) in patients undergoing coronary arterial surgery, and found that earlier extubation did not impair either haemodynamic or respiratory performance and might be beneficial. More recently, Cheng *et al.* documented a significant improvement in intrapulmonary shunting in the group randomized to early extubation. There are no studies that have been able to demonstrate improvements in respiratory function related to a period of artificial ventilation, and in fact the reverse may be true. A retrospective analysis of 47 patients following cardiac surgery noted a significantly higher incidence of atelectasis and a longer intensive-care stay in those patients in whom extubation had been delayed, and in a study of ventilation–perfusion distribution a substantial shunt was demonstrated during mechanical ventilation, with no difference in the level of shunt or ventilation–perfusion inequality between ventilation and spontaneous respiration. Although respiratory complications are relatively common after cardiac surgery, evidence that a period of positive-pressure ventilation reduces either the severity or the duration of pulmonary dysfunction is lacking. There are, however, a number of well-documented adverse effects from mechanical ventilation. These include reduction in venous return and cardiac output, which is particularly marked with the use of positive end-expiratory pressure, and decreases in renal and splanchnic perfusion. There is also considerable morbidity associated with endotracheal intubation, notably impaired coughing, cessation of mucus transport, cilial loss, and respiratory epithelial damage. These factors predispose to atelectasis and bacterial colonization. Furthermore, when extubation is delayed, the need for sedation and analgesia in order to tolerate the presence of a tracheal tube is considerably increased. Quasha *et al.* found that patients in whom extubation was delayed consumed 50 per cent more morphine and 400 per cent more benzodiazepine. This in itself is likely to reduce cardiac output, since all sedative drugs have some myocardial depressant effect. Finally, there is always the possibility, however rare, of technical mishaps such as ventilator disconnection or mechanical failure. Thus, neither endotracheal intubation nor mechanical ventilation can be considered to be benign or harmless.

Factors that should be considered before ventilatory support is withdrawn include haemodynamic stability, resolution of the effects of narcotics and muscle relaxants, adequate arterial blood gases with inspired oxygen concentrations of 50 per cent or less, and normal core temperature.

Extubation can be done when, in addition to the above, the patient is awake and alert, and able to maintain adequate gas exchange without distress. By these criteria many patients can be extubated within a few hours of surgery. Early extubation is part of a overall strategy that relies on appropriate anaesthetic and surgical techniques and postoperative management. However, each patient must be assessed individually and some will undoubtedly benefit from a longer period of controlled ventilation after surgery. Patients with severe postoperative cardiac dysfunction are best managed by continuing controlled ventilation to ensure adequate oxygenation and reduce the workload and oxygen cost of breathing.

Postoperative myocardial ischaemia and early extubation

In the early postoperative period, patients undergoing coronary arterial bypass surgery may be especially vulnerable to myocardial ischaemia, despite successful revascularization. Catecholamines remain markedly elevated for some time after surgery, suggesting that the stress of bypass does not end with the operation. These postoperative ischaemic episodes appear to be more severe than at other perioperative times, and there may be a relation between post-bypass ischaemia and adverse outcome. In high-risk, non-cardiac surgical patients, postoperative myocardial ischaemia certainly appears to be the most important predictor of adverse outcome. This led Mangano to the investigate the effect of a heavy sedative–analgesic regimen in the early postoperative period with the aim of blunting activation of the sympathetic nervous system and thus reducing myocardial ischaemia. In Mangano's study the use of a continuous infusion of sufentanil, in a dose not dissimilar to an anaesthetic regimen, reduced the incidence and severity of postoperative ischaemic episodes when compared with intermittent bolus doses of intravenous morphine. However, when intensive analgesia was stopped, the incidence of late postoperative ischaemia was similar in both groups of patients, and overall, there was no significant difference in adverse cardiac outcome. This work has major implications for postoperative care of patients after coronary arterial surgery, since the use of such analgesic regimens precludes early extubation. Preliminary data suggest that early extubation certainly does not increase the risk of ischaemia or infarction when compared to late extubation, but much more work is needed to determine the exact relation between perioperative stress, extubation time, and myocardial ischaemia. Until the results of large outcome studies are reported the use of heavy postoperative sedation remains of unproved benefit.

Analgesia

Effective analgesia helps to relieve anxiety and tachycardia and hypertension, and is a vital component of so-called fast-track management. In the early postoperative period, intravenous opiates, given by continuous infusion, are the mainstay of treatment. Supplementation with non-steroidal anti-inflammatory drugs improves the quality of analgesia and has a morphine-sparing effect. Most patients can be transferred to oral analgesics by the second postoperative day. Regular and effective analgesia will improve mobilization and well being, and speed recovery from surgery.

Prevention and treatment of complications

Hypertension

Hypertension is common after coronary arterial surgery. High blood pressure increases cardiac work and oxygen consumption, and may lead to myocardial ischaemia and increased bleeding. Recovery of consciousness with inadequate analgesia and sedation is a common but easily treated cause. Vasodilators such as glyceryl trinitrate or sodium nitroprusside are effective treatment for hypertension. Both drugs have a short duration of action, and the dose can be titrated against response. Reflex tachycardia can be treated by a b-blocking drug, although these should be used with caution in patients with left ventricular dysfunction.

Bleeding

Excessive postoperative blood loss is due to inadequate haemostasis, a deficiency in coagulation, or both. Between 3 and 5 per cent of patients undergoing cardiac surgery require more than 10 units of blood over the operative period. Postoperative hypertension will increase bleeding and should be avoided, especially in those in whom the surgical repair is fragile. Occasionally, sudden massive haemorrhage will require immediate surgical re-entry. This is almost always due to disruption of stitches and may occur after a brief episode of marked hypertension. Continued heavy bleeding (more than 4 ml/kg per hour) in a patient with normal or near-normal clotting is an indication for early surgical re-exploration: when coagulation is abnormal the decision to re-explore is more difficult. Coagulopathy after cardiopulmonary bypass is frequently multifactorial. Platelet function is affected by preoperative treatment with aspirin, haemodilution, and mechanical destruction during the bypass. The most likely cause of a bleeding tendency after cardiopulmonary bypass is a qualitative or quantitative decrease in platelets, and initial management should be directed to bring the platelet count to greater than 100 000/μl. In addition, concentrations of circulating clotting factors may be greatly reduced, and excessive fibrinolysis can occur. Residual, unneutralized heparin may also contribute to bleeding. Treatment of bleeding should include aggressive replacement of circulating volume and red cells, together with platelets and clotting factors. The use of positive end-expiratory pressure is popular with surgeons as a method of limiting bleeding after cardiac surgery. However, a randomized, controlled trial failed to show any benefit from this manoeuvre, which may be harmful in the hypovolaemic patient.

Cardiac tamponade

Tamponade occurs if blood accumulates in the mediastinum, producing mechanical obstruction to cardiac filling. This causes hypotension due to a fall in stroke volume, and an increase in filling pressures, together with tachycardia, oliguria, and peripheral cooling. Hypotension from tamponade may improve initially with administration of fluids and inotropic support, but if the accumulation of blood continues, circulatory collapse is inevitable. The diagnosis rests on a high index of suspicion, although chest radiographs may show an enlarging heart shadow, and echocardiography may be helpful in demonstrating pericardial clot and a compressed ventricle. Pulsus paradoxus (an exaggeration of the normal fall in arterial pressure with inspiration) is often cited as a sign of cardiac tamponade, but is rarely seen. The treatment of tamponade is mediastinal exploration with removal of clot. During preparation for resternotomy, imminent cardiovascular collapse may be delayed by further increasing filling pressures and heart rate, whilst blood pressure is maintained by peripheral vasoconstriction (keep the patient ‘full, fast and squeezed tight'). Opening the chest usually produces a marked haemodynamic improvement.

Dysrhythmias

Rhythm disturbances are common after cardiac surgery. Many factors may contribute: hypokalaemia, acid–base abnormalities, hypothermia, myocardial ischaemia, circulating catecholamines, and surgical damage. Treatment includes correction of hypokalaemia and acidosis, administration of antidysrhythmic drugs, and occasionally overdrive pacing or DC cardioversion. About 30 per cent of patients undergoing coronary arterial surgery will develop atrial fibrillation by the third postoperative day. Prophylactic administration of amiodarone may be effective in preventing this.

Low cardiac output

Postoperative myocardial dysfunction can be caused by failure of myocardial protection during aortic cross-clamping, technically inadequate surgery, perioperative infarction, dysrhythmias, air embolism, or coronary spasm; pre-existing ventricular dysfunction is a major risk factor. Acute cardiac failure is characterized by hypotension and hypoperfusion associated with vasoconstriction, oliguria, and acidosis. Successful management depends on optimization of cardiac output and reduction in cardiac workload by physiological, pharmacological, and mechanical means. General supportive measures include the treatment of acidosis and hypoxaemia. A pulmonary arterial catheter may be required to direct treatment.

Determinants of cardiac output

Heart rate and rhythm

Cardiac output is the product of heart rate and stroke volume. During bradycardia, increasing the rate will therefore increase the cardiac output (if stroke volume were unchanged). At heart rates above 110 to 120 beats/min, however, shortening of the duration of diastole starts to impair ventricular filling and markedly reduces stroke volume. Myocardial ischaemia may occur with severe tachycardia, and the function of prosthetic valves may be significantly impaired. Hence, excessive increases in heart rate are undesirable.

Dysrhythmias will adversely affect cardiac output and should be promptly diagnosed and treated. The co-ordinated atrial contraction that occurs in sinus rhythm contributes up to 30 per cent of ventricular filling and is vital for adequate cardiac performance in patients with poor ventricular function. Sequential dual-chamber pacing (pacing both atrium and ventricle with an appropriate atrioventricular delay) mimics the effect of sinus rhythm, and is useful when heart block complicates surgery.

Preload

Measurements of both right and left heart-filling pressures are required to optimize preload. Although increased filling pressures are needed to augment cardiac output, ventricular overdistension and pulmonary congestion must be avoided.

Contractility

Myocardial contractility is increased by the use of positively inotropic drugs. Catecholamines such as dopamine, dobutamine, adrenaline (epinephrine), and isoprenaline will improve contractility, albeit at the cost of an increase in myocardial oxygen demand. The overall clinical effects depend on the relative degrees of stimulation of a- and b-adrenergic receptors. The newer phosphodiesterase inhibitors such as amrinone, milrinone, and enoximone combine positive inotropy with vasodilatation; these drugs may find a place in the treatment of low cardiac-output syndrome.

Afterload

The use of vasodilator drugs to reduce afterload decreases myocardial wall tension and oxygen demand, and allows the ventricles to empty more fully. Successful treatment is marked by improving cardiac output in the face of unchanged or only slightly decreased arterial pressure. A combination of vasodilators and inotropes is often required to maximize cardiac output.

Mechanical support of the failing heart

When pharmacological support is insufficient to reverse the low cardiac-output syndrome, mechanical assistance should be considered. The intra-aortic balloon pump is a volume-displacement device in which inflation and deflation of an elongated balloon in the descending thoracic aorta is timed to coincide with the cardiac cycle. Balloon inflation increases diastolic blood pressure, thus improving coronary flow: active deflation just before systole unloads the aorta and reduces resistance to ventricular ejection. The balloon pump does not generate blood flow but augments existing cardiac output and increases coronary perfusion. In contrast, ventricular-assist devices are capable of maintaining cardiac output without ventricular function. At present, experience with these devices is limited, but it seems likely that their use will increase as knowledge and technology advance.

Further reading

Cheng DCH. Fast-track cardiac surgery: economic implications in postoperative care. *Journal of Cardiothoracic and Vascular Anesthesia* 1998; **12**: 72–9. [Davey Cheng discusses the topical issues of cost containment and efficient resource utilization in cardiac surgery.]

Cheng DCH *et al.* Morbidity outcome in early versus conventional tracheal extubation after coronary artery bypass grafting: a prospective randomised controlled trial. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 755–64. [This study examined the effects of early (1–6 h) versus late (12–22 h) extubation in 120 patients and found that early extubation significantly decreased intensive-care unit and hospital length of stay.]

Chong JL, Pillai R, Fisher A, Grebenik C, Sinclair M, Westaby S. Cardiac surgery: moving away from intensive care. *British Heart Journal* 1992; **68**: 430–3. [This paper describes the use of a cardiac surgical recovery area and demonstrates that fewer than 7 per cent of patients require conventional intensive-care management after cardiac surgery.]

Connors AF *et al.* The effectiveness of right heart catheterization in the initial care of critically ill patients. *Journal of the American Medical Association* 1996; **276**: 889–97. [An observational study of the effect of right heart catheterization during the first 24 h of intensive-care unit stay in 5735 critically ill patients on subsequent survival, length of stay and cost of care. The use of a pulmonary arterial

catheter was associated with greater fatality and increased resource utilization.]

Kaplan JA, ed. *Cardiac anesthesia*, 3rd edn. Saunders, Philadelphia, 1993. [A seminal textbook for the cardiac anaesthetist.]

Lefemine AA, Harken DE. Postoperative care following open-heart operations: Routine use of controlled ventilation. *Journal of Thoracic and Cardiovascular Surgery* 1966, **52**: 207–16. [An important study of two non-randomized groups each of 18 patients who were either ventilated or allowed to breathe spontaneously after a variety of (mostly valvular) heart surgery. Advantages of ventilation included avoidance of tracheostomy and reduction in hypoxia, acidosis, hypercarbia, hypotension and atelectasis.]

Mangano DT *et al.* Postoperative myocardial ischaemia. Therapeutic trials using intensive analgesia following surgery. *Anesthesiology* 1992; **76**: 342–53. [A randomized, controlled study of 106 patients undergoing coronary arterial bypass graft surgery who received either low-dose intermittent intravenous morphine or continuous sufentanil infusion postoperatively in the intensive care unit. The incidence of electrocardiographic ischaemia was significantly less in patients receiving sufentanil but, on stopping sedation, the incidence and severity of ischaemic episodes was similar in both groups.]

Pulmonary Artery Catheter Consensus Conference. Consensus statement. *Critical Care Medicine* 1997; **25**: 910–25. [In patients undergoing cardiac surgery the recommendation of the Consensus Conference was that the routine use of a pulmonary arterial catheter was unnecessary in low-risk patients but may be useful in high-risk patients, especially those with clinically important left ventricular dysfunction.]

Quasha *et al.* Postoperative respiratory care: a controlled trial of early and late extubation following coronary artery bypass grafting. *Anesthesiology* 1980; **52**:135–41. [The first prospective trial of early extubation demonstrated that early extubation (at a mean of 2 h) was achievable, and that compared with late extubation (mean of 18 h) was associated with a lowered incidence of postoperative respiratory complications.]

Swan HJC *et al.* Catheterization of the heart in man with the use of a flow-directed balloon-tipped catheter. *New England Journal of Medicine* 1970; **283**: 447–51. [The original report of bedside pulmonary arterial catheterization without fluoroscopy by using a balloon-tipped catheter—the paper that moved pulmonary arterial catheters from the catheter laboratory to the intensive care unit.]

Tuman KJ *et al.* Effect of pulmonary artery catheterization on outcome in patients undergoing coronary artery surgery. *Anesthesiology* 1989; **70**: 199–206. [A non-randomized study of 1094 patients undergoing coronary arterial surgery, which concluded that a pulmonary arterial catheter did not favourably alter outcome.]

40.7.1 Management of congenital heart disease

Nicholas Archer and Ingegerd Östman-Smith

Introduction
Terminology
Physiological descriptions
Presentation and clinical features of congenital heart defects
Fetal
Neonatal and infant
Older children and adults
Investigation of suspected congenital heart disease
Chest radiograph
Electrocardiography
Hyperoxia test (nitrogen washout test)
Cardiac ultrasound scanning
Cardiac catheterization
Magnetic resonance imaging (MRI)
Choice of investigations
Management of symptomatic congenital heart disease
Medical management
Postoperative follow-up
Palliative procedures in congenital heart disease
Conclusions
Further reading

Introduction

Congenital structural abnormalities of the heart occur in approximately 8 per 1000 live births. Some of these malformations are haemodynamically trivial, but others cause mortality and major morbidity in early childhood. About 30 per cent cause symptoms in the early months after birth. As palliative and corrective surgery become increasingly effective, more individuals will reach adult life with significant congenital heart lesions; many require further surgery. In order to understand the types and effects of cardiac abnormalities it is necessary to consider modern descriptive terminology before considering physiological consequences.

Terminology

Sequential chamber analysis

Sequential chamber analysis is a way of describing the morphological arrangement of any heart in a logical manner without making reference to known or conjectured embryological processes which have given rise to the situation. If systemic veins drain normally into a morphological right atrium, pulmonary veins drain normally to a morphological left atrium, each atrium drains through an appropriate valve into a morphologically concordant ventricle, and each ventricle gives rise to an appropriate great artery, the heart is said to have normal connections. Having normal connections is not inconsistent with the heart being in an abnormal position, and the majority of congenital heart defects are found in a normally connected heart. Abnormality of connection at any level (atrial, ventricular, arterial) should be clearly and precisely described, referring to right or left in a morphological rather than a positional sense. Atrial situs solitus means usual atrial arrangement with a morphologically right atrium being to the right of the morphological left atrium, whereas in atrial situs inversus there is a mirror image reversal of this pattern. Atria and ventricles have certain characteristics which allow their morphological recognition clinically as well as surgically. Recognition of the morphology of particular cardiac chambers in life is most readily achieved by ultrasound imaging but electrocardiography, radiography, and angiocardiography may all provide helpful information. Once morphological connections have been described any abnormalities in the relative position of structures can be given, for example aorta to left of pulmonary artery. Table 1 gives a simplified outline of normal and abnormal connections based on the work of Anderson and others. Van Praagh's system of segmental analysis has many similarities to sequential chamber analysis and is widely used in North American literature; it is rather more complex and relates to embryological processes. When the basic arrangement of a heart has been described, further lesions can be elucidated, for example stenosed valves or septal defects. Another, more traditional and complementary way of describing and classifying heart defects is by basic physiological disturbances.



Table 1 Outline of sequential chamber analysis for describing congenital cardiac abnormalities

Physiological descriptions

Congenital heart abnormalities can be grouped together by their physiological characteristics, thereby allowing general statements to be made about the adverse effects of the lesion. Some abnormalities fit into several physiological groups.

Six physiological divisions will be considered: obstructive and regurgitant lesions, left to right shunts, right to left shunts, common mixing situations, transposition haemodynamics, and functional as opposed to structural heart disease (Table 2).

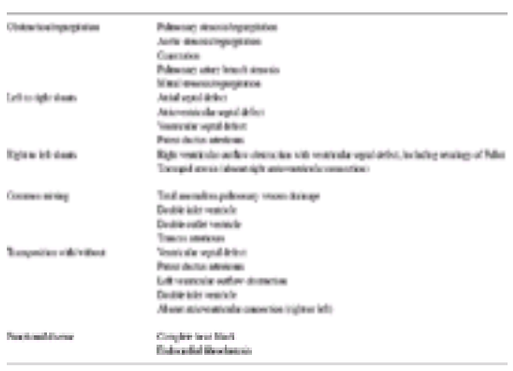


Table 2 Physiological classification of congenital heart defects with examples

Obstructive and regurgitant lesions

A simple obstruction, such as arterial valve stenosis or coarctation, will result in hypertrophy of the cardiac chamber proximal to it, usually without gross dilatation of the chamber initially, particularly in pulmonary or aortic stenosis, but not in mitral stenosis (a rare congenital lesion) which does produce left atrial enlargement. Regurgitation of a valve is far more likely to produce dilatation of the proximal chamber and consequent radiographic evidence of enlargement of the heart. Both pressure and volume overload of a chamber may eventually result in adverse effects further back in the circulation than the chamber immediately proximal to the lesion. Isolated pulmonary outflow obstructive lesions will not alter the amount of blood passing through the lungs: there needs to be an abnormal gap (atrial or ventricular septal defect) or persistence of a normal gap (patent foramen ovale in critical pulmonary stenosis in the newborn) for this to occur.

Left to right shunts

The common lesions thus classified are atrial and ventricular septal defects, atrioventricular septal defects, and patent ductus arteriosus. In patients with uncomplicated disease, blood flows from left to right heart at a rate determined by the size of the hole and resistance to flow downstream from it. Thus, very little blood will cross even a large ventricular septal defect until the normal postnatal fall in pulmonary vascular resistance occurs. Hence, such a defect is frequently not suspected in a newborn, who may present with a murmur or heart failure some weeks later. High pulmonary blood flow can cause heart failure as the left ventricle becomes volume overloaded by the increased pulmonary venous return in those with ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defects. Volume overloading from atrial septal defects occurs chiefly to the right ventricle, and heart failure in childhood is rare. So-called ostium primum atrial septal defects are correctly classified as partial atrioventricular septal defects and as such may be associated with heart failure in infancy. High pulmonary blood flow or high perfusion pressure in the pulmonary vascular tree may result in permanent constriction of the pulmonary arterioles with consequent elevation of pulmonary vascular resistance. In this event, right-sided pressures may exceed left heart pressures and right to left shunting occurs producing cyanosis (Eisenmenger complex). The likelihood of Eisenmenger haemodynamics and its timing varies from lesion to lesion, being early and almost universal in complete atrioventricular septal defects and only exceptionally encountered in simple ostium secundum atrial septal defects. Much rarer left to right shunting lesions include aortopulmonary window and coronary arteriovenous fistula.

Right to left shunting lesions

Because right heart pressures are normally lower than left after the first few days of life right to left shunting occurs if there is both obstruction to blood flow through the right heart and a route for the blood to cross to the left such as an atrial or ventricular septal defect. In young infants, a patent foramen ovale allows such shunting to occur. The net result of this arrangement is that desaturated blood bypasses the lungs and reaches the systemic circulation to produce systemic arterial desaturation and probably recognizable cyanosis, depending on the degree. Examples of right to left shunting lesions are given in Table 2.

Common mixing

Mixing of desaturated and fully oxygenated blood results in a degree of systemic arterial desaturation, the severity of which will depend chiefly on whether there is increased or decreased lung blood flow. Mixing can occur at any level from before venous blood drains into the heart (for example total anomalous pulmonary venous drainage to the superior vena cava) until when blood leaves the heart (as in truncus arteriosus). Tricuspid atresia (absent right atrioventricular connection) exemplifies a lesion in which two of these physiological mechanisms coexist, with right to left shunt and common mixing.

Transposition

Transposition exists when the aorta arises from the morphological right ventricle and the pulmonary artery arises from the left ventricle (ventriculoarterial discordance). Clearly, there has in transposition to be some mixing of systemic and pulmonary blood to sustain life: this can be at atrial, ventricular, or great arterial levels. Other lesions such as obstruction (pulmonary stenosis or coarctation) or obligatory common mixing (tricuspid atresia) may also be present.

Functional lesions

Functional congenital heart diseases include bradyarrhythmias and tachyarrhythmias in the absence of macroscopic structural abnormalities and heart muscle disease already present at birth.

Presentation and clinical features of congenital heart defects

Fetal

Transabdominal fetal ultrasound examination is increasingly being used to detect cardiac abnormalities from about 18 weeks gestation; transvaginal scanning is possible a few weeks earlier but is not widely practised for cardiac assessment. Routine screening of all pregnancies by identifying four normally proportioned heart chambers at a general anomaly scan is followed by a more detailed (and time consuming) cardiac scan for those in whom an abnormality is suspected in the screening scan. Examination in detail of other groups is also offered to those with previously affected children (recurrence risk is usually about 3 per cent), those who themselves have congenital heart disease (recurrence risk varies), those mothers with diabetes (risk 5 per cent) or other conditions or treatments associated with an increased risk of cardiac problems in their offspring, those with a fetus thought to have other (non-cardiac) abnormalities, and those parents with an inheritable condition with a possible cardiac component. Professional and parental views on termination of pregnancy vary and are beyond the scope of this contribution but management of what can be a very stressful experience for couples requires sensitivity to the broader issues. Increasing numbers of babies are being born with cardiac abnormalities diagnosed antenatally; this may affect the choice of the place of delivery and should facilitate the immediate neonatal management, and, in particular, help prevent babies from becoming severely symptomatic before serious heart lesions are suspected.

Fetal arrhythmias which may produce heart failure (hydrops fetalis) occur in the presence of structurally normal or abnormal hearts. Fetal supraventricular tachycardia can be treated by drug administration to the mother or directly to the fetus so as to restore sinus rhythm prior to delivery in many cases.

Neonatal and infant

Congenital heart disease may present in the early weeks of life in a number of ways. A cardiac abnormality may be revealed by investigation (ECG or ultrasound) before it is clinically apparent, the investigation having been undertaken for some other reason, such as Down's syndrome or oesophageal atresia. Heart failure in a young baby causes sweatiness, breathlessness, and poor feeding. Physical signs include tachycardia, tachypnoea, intercostal indrawing, and hepatomegaly. Peripheral oedema is a very late feature of neonatal and infant heart failure, and the jugular venous waveform and pressure cannot be reliably evaluated. Cyanosis is only present in gross pulmonary oedema, although some conditions causing heart failure may also cause cyanosis. Congenital cardiac abnormalities causing heart failure in the newborn and in infancy are given in Table 3. Acquired heart disease, such as myocarditis, may also cause heart failure in this age group as in any other.

Hypoplastic left heart
Coarctation with patent ductus arteriosus, and with or without ventricular septal defect
Interrupted aortic arch
Critical aortic stenosis
Transposition of the great arteries with patent ductus arteriosus, and /or ventricular septal defect
Left to right shunt lesions
Endocardial fibroelastosis
Univentricular atrioventricular connection with increased lung blood flow
Unobstructed total anomalous pulmonary venous drainage

Table 3 Congenital heart lesions which cause heart failure in neonates and infants

Central cyanosis (Fig. 1) can be difficult to recognize in newborn infants, particularly if they are plethoric, even by newborn standards, if they have facial petechiae after birth (traumatic cyanosis), or if they are of non-white race. Cyanosis of hands and feet is very common in healthy babies in the first few days of life. Cyanosis may be caused by respiratory disease, in which case respiratory distress is usually very marked. Some improvement is often seen when high concentrations of oxygen are breathed; blood gases analysis reveals hypercapnia and chest radiographs usually allow a respiratory diagnosis to be made. If doubt still exists, an ECG and a formal hyperoxia test may help but cardiac ultrasound scanning is likely to be needed. Cyanosis may also be a feature of hypoglycaemia, hypocalcaemia, and various abnormal neurological states, the mechanisms for these involving respiratory depression or elevated pulmonary vascular resistance. Methaemoglobinaemia causes a baby to look a slate blue/grey colour: this is easily confused with central cyanosis but arterial oxygenation is rarely abnormal, even though the blood itself is black. Cardiac causes for cyanosis can be classified into three main categories: common mixing, right to left shunting, and transposition (Table 2). Some congenital lesions fall into more than one category. Generally speaking, babies with common mixing conditions without associated obstruction to blood flow into the lungs are likely to be less desaturated and do not present as deeply cyanosed newborns; those with transposition without a large associated ventricular septal defect or patent ductus arteriosus present very early in life. Infants with conditions associated with high lung blood flow develop heart failure.



Fig. 1. Central cyanosis at 3 to 4 weeks of age.

Murmurs in newborn babies and infants may be normal innocent noises; serious congenital heart disease can also exist in the absence of murmurs. Nevertheless, much store is placed on listening for murmurs at routine examinations. History and other features of the physical examination are important in helping to determine whether further investigation of a murmur is needed. Arterial valve stenosis causes a murmur which may be heard from birth; left to right shunting lesions usually do not as the shunt will be small until pulmonary vascular resistance falls. An innocent murmur arising from the pulmonary arteries may be heard over the upper chest and laterally at birth and for the first few months of life. Tricuspid regurgitation secondary to perinatal asphyxial stress may cause a transient pansystolic murmur at the lower left sternal edge; a small ventricular septal defect may cause an identical murmur in the newborn period. This has a high likelihood of disappearing, with the defect closing in the first year of life.

Newborn babies and older infants may collapse due to sepsis and many other causes, but arrhythmias (with or without structural heart disease) and congenital heart lesions must be borne in mind, particularly if collapse occurs in the first week or two of life and is associated with cardiomegaly and hepatomegaly. Infants with lesions in which the systemic circulation is dependent upon patency of the ductus arteriosus often present with collapse; there may have been earlier concern about developing heart failure or poorly felt femoral pulses. Lesions associated with duct dependent systemic circulation are listed in Table 4. If pulmonary blood flow is critically dependent on the ductus arteriosus, cyanosis will appear or worsen when the duct closes. Collapse with respiratory distress will follow if severe hypoxaemia results in metabolic acidosis before appropriate intervention occurs (Table 5).

Hypoplastic left heart
Interrupted aortic arch
Coarctation—infant type
Critical aortic stenosis
Aortic atresia

Table 4 Lesions with duct dependent systemic circulation

Pulmonary atresia with intact ventricular septum
Tetralogy of Fallot with pulmonary atresia
Tricuspid atresia with concordant ventriculo-arterial connection and small ventricular septal defect or pulmonary stenosis
Transposition of the great arteries with severe left ventricular outflow obstruction
Double outlet right ventricle with pulmonary atresia or severe stenosis

Table 5 Lesions with duct-dependent pulmonary circulation

Abnormality of the femoral pulses, such as impalpability or diminished pulse pressure compared with upper limb pulses, should always be sought when assessing the cardiovascular system of young babies, difficulty in feeling femoral pulses suggests coarctation of the aorta.

Older children and adults

Heart failure and cyanosis (Fig. 2) are unusual presentations of congenital heart disease beyond infancy, but they may be features of the progressive natural history, particularly as modified by surgery. Occasionally, a patient with Eisenmenger haemodynamics from an unrecognised ventricular septal defect or patent ductus arteriosus may present with cyanosis in later childhood or adult life; heart failure in middle age is a manifestation of some congenital lesions which may have escaped earlier detection, such as atrial septal defects or coarctation of the aorta. Sinus of Valsalva aneurysm does not cause symptoms until rupture in adult life causes heart failure. Hypertension and cerebrovascular accidents in young adults is one mode of presentation of coarctation of the aorta. Syncope, especially on exertion, angina, or even sudden death may be the first indication of congenital left ventricular outflow obstruction. ‘Funny turns' due to arrhythmias associated with congenital heart lesions may result in previously unrecognized conditions, such as atrial septal defect or congenitally corrected transposition, being detected. Infective endocarditis may be the presenting feature of a congenitally bicuspid aortic valve. Asymptomatic murmurs in childhood are frequently innocent but can be due to a large number of structural heart lesions. Many of these are indications for prophylaxis against endocarditis. Some, including aortic stenosis, coarctation, patent ductus arteriosus, and atrial septal defect, can be of major long-term haemodynamic importance. The most common innocent, normal murmurs in children over the age of 1 year are a venous hum and the vibratory midsystolic murmur described by Still. Features of these murmurs are given in Table 6; and in both of these conditions cardiovascular examination is

otherwise normal.



Fig. 2. Central cyanosis at 6 years.

Venous hum	Base of heart under clavicle
	Louder in diastole
	Louder in inspiration
	Abolished lying flat/head down, by moving or pressing on neck
Still's innocent systolic murmur	Between lower left sternal edge and apex
	Vibratory
	Louder lying
	Louder with bell of stethoscope
	Abolished by Valsalva manoeuvre
	No precordial thrill

Table 6 Features of common innocent murmurs after infancy

Investigation of suspected congenital heart disease

Chest radiograph

A simple posteroanterior (or anteroposterior in neonates and infants) film usually suffices, although lateral views occasionally add further information. Information to be gained relevant to a complete cardiovascular diagnosis is listed in Table 7 (Fig. 3, Fig. 4).

Heart
Site
Size
Contour
Side of apex
Spine and ribs
Lung pathology—collapse/consolidation/overinflation
Mediastinal displacement
Bronchial and abdominal situs
Aortic arch side
Pulmonary vascular markings

Table 7 Features of relevance to cardiovascular system on plain chest radiograph



Fig. 3. Chest radiograph, neonate with cardiomegaly and pulmonary plethora.



Fig. 4. Chest radiograph, preterm newborn infant with tetralogy of Fallot showing oligaemic lung fields.

Electrocardiography

This yields extremely valuable information on rhythm, chamber size and dominance, and myocardial state. Echocardiography is more sensitive at determining left ventricular hypertrophy. There are certain ECG patterns which are almost diagnostic in particular clinical settings: left axis deviation in a cyanosed neonate with oligaemic lung fields makes tricuspid atresia virtually certain, while a superior, ‘north west' QRS axis in Down's syndrome points to a complete atrioventricular septal defect.

Hyperoxia test (nitrogen washout test)

As originally described, this test helps distinguish cyanotic congenital heart disease in the newborn from respiratory disease by the fact that a rise in arterial oxygen tension above 20 kPa in an infant breathing 85 per cent oxygen is not usually seen with cyanotic heart disease.

The availability of ultrasound imaging has reduced the need for the hyperoxia test, but it is still useful where ultrasound is not readily available. It is also useful in alerting the physician to the presence of more complex cardiac disease in an infant who is not clearly cyanosed but who may still not respond normally to breathing a high concentration of oxygen. Persistent pulmonary hypertension of a newborn with a structurally normal heart may cause a failed hyperoxia test, although the response may be labile. Hyperoxia testing can be performed using transcutaneous oxygen tension electrodes if no other indication for arterial sampling exists and a conclusive result (pass or fail) is obtained. Pulse oximetry will not provide such discriminatory information because of the characteristics of the haemoglobin oxygen dissociation curve.

Cardiac ultrasound scanning

Echocardiography has revolutionized the investigation and management of congenital heart disease. M-mode echocardiography, ultrasound imaging, and Doppler in all its modalities (continuous wave, pulsed wave, and colour flow) allow precise anatomical and functional information to be obtained by experienced operators (Fig. 5 and Fig. 6). An increasing number of patients are referred to surgery without the need for invasive investigation. Intraoperative ultrasound scanning shows promise in reducing the incidence of unexpected complications or unsatisfactory repairs by allowing their detection in the operating theatre. Transoesophageal, as opposed to precordial, scanning is useful in studying atrial anatomy but is not practicable in young infants.



Fig. 5. High left parasternal ultrasound scan showing patent ductus arteriosus. Main pulma, main pulmonary artery; LPA, left pulmonary artery; desc ao, descending aorta.



Fig. 6. Apical four chamber ultrasound scan showing perimembranous ventricular septal defect with tricuspid valve tissue partly occluding it by forming a so-called aneurysm. LA, left atrium; RA, right atrium; LV, left ventricle; RV, right ventricle; VSA, ventricular septal aneurysm.

Cardiac catheterization

This is often used to evaluate haemodynamics and eludicate anatomical details, but these roles are being modified rapidly by echocardiography. Although mortality and morbidity rates are low, they must be weighed against the benefit to be gained in a particular case. Interventional cardiac catheterization is finding an increased role in congenital heart disease (Fig. 7) and is making some operations, such as pulmonary valvotomy, a rarity (Table 8).

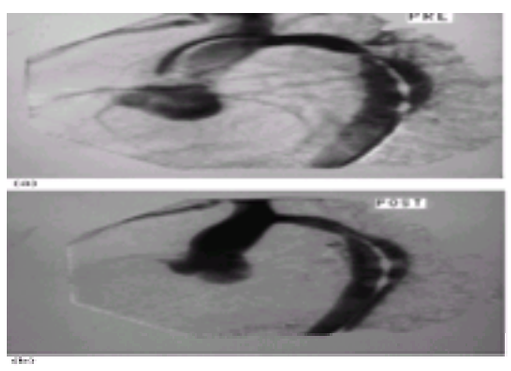


Fig. 7. (a) Digital subtraction aortogram showing recoarctation following previous neonatal subclavian flap repair of coarctation. (b) Similar left anterior oblique aortogram after balloon dilatation relief of recoarctation.

Balloon atrial septostomy	Transposition of the great arteries[a]
	Obstruction to flow out of atrium[a]
Blade atrial septostomy	Failed balloon septostomy[a]
Balloon valvuloplasty	Pulmonary stenosis[a]
	Aortic stenosis[a]
	Pulmonary atresia, heart assisted procedure[a]
Balloon angioplasty	Coarctation[a]
	Reconstruction[a]
Stent closure	Patent ductus arteriosus[a]
	Atrial septal defect[a]
Embolization	Aortopulmonary collaterals[a]
	Cerebral arteriovenous fistula[a]
Foreign body retrieval[a]	
Accessory pathway ablation[a]	
Stenting of stenosed vessels (pulmonary arteries)[a]	

[a]Standard treatment.
[b]Of value under certain circumstances.
[c]Being evaluated.

Table 8 Interventional cardiac catheterization in congenital heart disease

Magnetic resonance imaging (MRI)

Anatomical details are clearly displayed by MRI, although to date clear clinical advantage over ultrasound or angiography is only recognized in a few conditions, such as locating pulmonary arteries in patients with truncus arteriosus or pulmonary atresia with ventricular septal defect. Information on cardiac output and pulmonary vascular resistance can be obtained from magnetic resonance studies but whether this will supersede assessment with ultrasound techniques remains to be seen.

Choice of investigations

Deciding on the correct investigative approach at the right time needs careful judgement taking into account risks and benefits as well as the age of the patient and the degree of co-operation required and likely to be obtained. A knowledge of the possible natural histories of conditions subject to corrective or palliative surgery is not always available and careful repeated follow-up by clinical and investigational means is required. Investigations with minimum hazard, inconvenience, and expense should be used when-ever possible.

Management of symptomatic congenital heart disease

The key to optimal management of both asymptomatic and symtomatic congenital heart disease is a complete anatomical diagnosis and it is important not only to identify the major structural cause of symptoms but also to document normality of the rest of the cardiac anatomy including the systemic and pulmonary venous connections and the presence or absence of a patent ductus arteriosus, as the presence both of a patent ductus or bilateral superior venae cavae with the left draining to the coronary sinus have major implications for the management of cardiopulmonary bypass. Furthermore, if the surgeon encounters unexpected additional lesions at surgery the correct identification of the anatomy uses up precious bypass-time. To achieve a complete anatomical diagnosis it may be necessary to perform more than one echocardiographic examination, with the second one carried out under sedation to facilitate obtaining optimal subcostal and suprasternal views, and haemodynamic evaluation with cardiac catheterization and angiography may be needed as outlined above.

Medical management

The sick neonate

The sick neonate generally presents because either pulmonary or systemic blood flow is duct-dependent, or because there is inadequate contribution from pulmonary venous return to the blood circulating in the arterial circulation as in transposition, a situation called ‘inadequate mixing'. The focus of medical management should be to rapidly bring the patient into optimal condition for emergency surgery. The initial management consists of restoring duct-patency by infusion of prostaglandins, either prostaglandin E_2 in a dose of 0.005 to 0.010 µg/kg/min, or a dose of Prostaglandin E_1 of 0.05 to 0.1 µg/kg/min. The survival of infants with duct-dependent cardiac lesions has been dramatically improved by the introduction of these drugs. At the same time, it is of vital importance to correct accompanying acidosis rapidly, using sodium bicarbonate if spontaneous ventilation is adequate to blow off CO_2 generated and if hypernatraemia is not present, but using THAM (tris-hydroxymethyl-aminomethane) if there is respiratory distress or an accompanying respiratory component to the acidosis. If there is respiratory distress or profound metabolic acidosis it is often desirable to electively intubate and ventilate the infant to reduce the systemic oxygen requirements associated with increased respiratory work. Neonates, particularly those with cyanotic heart disease, are prone to hypoglycaemia and blood glucose should be closely monitored and rapidly corrected if subnormal. A substantial portion of sick neonates may develop hypocalcaemia if maintained on intravenous glucose infusions only, and this occurs even in infants without partial or complete DiGeorge syndrome, and as hypocalcaemia adversely affects both blood pressure and myocardial contractility close attention to ionized serum calcium levels is mandatory. It is also important to avoid hyponatraemia and hypokalaemia which increase the risk of perioperative complications, and thus it is never appropriate to keep these sick neonates on an infusion of dextrose only without appropriate electrolyte supplements.

22q11-deletions causing partial or complete DiGeorge syndrome (or Shprintzen's syndrome) may be associated with any conotruncal cardiac malformation but particular high risk lesions are interrupted aortic arch, truncus arteriosus, and pulmonary atresia with VSD or tetralogy of Fallot associated with a right-sided aortic arch. Apart from the tendency to develop hypocalcaemia the other major practical importance is the possibility of associated immunodeficiency with an absent or reduced number of T-cell lymphocytes. Such infants are at risk of graft-versus-host reaction if transfused with lymphocyte-containing blood products, particularly with the large volume of blood required for cardiopulmonary bypass. Thus, infants with high-risk cardiac lesions should have their chest radiograph carefully scrutinized for absence or presence of a thymic shadow, and blood sent for 22q11-deletion detection by FISH-test and quantification of T and B-cell lymphocyte subclasses prior to cardiac surgery. Often the result of these investigations will not be available at the time of surgery and infants with high risk lesions should only be transfused with irradiated blood products until such time as it has been established that an adequate T-cell lymphocyte count is present. If there is complete thymus aplasia found at sternotomy, there will most probably be insufficient circulating lymphocytes for chromosome culture from the peripheral blood to be successful and a skin biopsy should be sent for chromosome analysis from fibroblast culture.

It is probably wise not to give nasogastric feeding to infants that may be at increased risk of necrotizing enterocolitis because their clinical presentation included a period of gut ischaemia (e.g. coarcation of the aorta or interrupted aortic arch), but in other infants nasogastric feeding of expressed breast milk or an appropriate infant formula should be considered as the best way of maintaining electrolytes and of reversing a catabolic state and trying to build up some tissue glycogen stores prior to emergency surgery. In neonates that have been anuric for a period it is vitally important to re-establish renal perfusion and urine flow rapidly, and they should immediately be given 1 mg/kg of frusemide intravenously, and if there is no response within 1 h further interventions such as a single mannitol dose of 0.5 g/kg and/or 2 mg/kg of aminophylline infused over 20 min and followed by 2 mg/kg of frusemide should be considered. A dopamine infusion at 5 ng/kg/min may also be helpful, particularly if overall systemic blood pressure is remaining low. Anuria in duct-dependent systemic circulation should never be treated by volume alone. If there is poor myocardial function, as is often the case with critical aortic stenosis or severe coarctation, it is helpful to load the child with digoxin prior to surgery (as long as the serum potassium is > 3.5 mmol/l). A dose of 5 µg /kg slowly intravenously, or 7 µg /kg given gastrically, can be safely given twice within a few hours even in the anuric child in order to rapidly achieve therapeutic serum levels, but size and timing of further maintenance doses would depend on subsequent urine output.

With modern management of cardiopulmonary bypass and postoperative care it is increasingly common that the first surgical procedure is an attempt at correction of the anomaly, but for patients that have lesions that are unsuitable for neonatal correction, or lesions that can only be palliated, strategies for surgical palliation are discussed below.

The sick older infant

The older infant is likely to present with features of chronic heart failure, often secondary to large left to right shunts with excessive pulmonary blood flow at high pressures. The first task is to optimize control of heart failure with the use of digoxin, diuretics, and, where appropriate, after-load reduction with captopril or enalapril. Use of either loop-diuretics or thiazides lead to secondary hyperaldosteronism and either diuretic will work best if combined with spironolactone. Potassium conservation is important as breast fed babies and babies not yet taking solids have a low potassium intake, and will become hypokalaemic unless loop diuretics are combined with a potassium-sparing diuretic. Where the left to right shunt is large, it is appropriate to introduce, for example, captopril early on, as it is better tolerated in infants who are not yet on massive doses of diuretics and bordering on a hypovolaemic state. It is also wise not to give big doses of captopril at the same time as powerful loop-diuretics when you are introducing the drug, and generally it is better to give captopril a few hours before the diuretics rather than vice versa. After-load reducing agents should not be used where there is systemic outflow obstruction. When the heart failure is under acceptable control, the focus should shift to optimizing nutrition, as these infants often present with failure to thrive. To achieve this, calorie-supplementation of breast milk or formula feeds, early introduction of solid feeds, and overnight continuous gastric drip feeding via a nasogastric tube are all strategies that may be needed. When a patient has been on large doses of diuretics

preoperatively it is wise to check not only serum potassium, sodium, and calcium but also serum magnesium among the preoperative blood tests, as magnesium depletion is not uncommon and will lower the threshold for perioperative arrhythmias.

The older child

The older child requiring elective surgery is less likely to require medical antifailure treatment, but other types of medication may be appropriate prior to surgery. For instance children with coarctation of the aorta benefit by being established on a b-blocker prior to surgery, as it makes the postoperative hypertension invariably following coarctation-resection in the older child easier to control. The b-blocker chosen should be a lipophilic one in order for it to exert central as well as peripheral effects (i.e. do not use atenolol for this indication). The postoperative hypertension is due to reflex-activation of sympathetic nerves with massive release of catecholamines and in the intensive care unit it should not be treated with vasodilators alone, but ideally with a combined a- and b-blocker such as labetalol given as an intravenous infusion. After 24 to 48 h this can be exchanged for oral b-blocker therapy. Antihypertensive medication aiming for complete normalization of blood pressure should be maintained for about 3 months postoperatively in order to try to reset central baroreceptor control to normal levels, and then tailed off gradually rather than discontinued suddenly. Other patients that may benefit from preoperative b-receptor blockade are those patients with gross left ventricular hypertrophy secondary to severe aortic stenosis at risk for the development of ‘stone heart' on bypass; there is experimental evidence that propranolol reduces the risk of this complication. Some patients with tetralogy of Fallot may have been treated with b-blockers to control hypercyanotic spells, in that situation it is crucial not to discontinue the b-blocker before the anaesthetic, as with up-regulated b-receptors such a patient is at increased risk of a hypercyanotic spell on the induction of anaesthetic. It is much safer to continue the usual b-blocker dose on the morning of surgery, and simply consider the degree and type of b-receptor blockade present when choosing postoperative inotrope support if required. b-receptor blockade is competitive and can always be overcome by a dose increase of agonist, but it is important to use an agonist of the same type as the blocker used—do not use a non-selective agonist such as isoprenaline if the patient has been treated with a selective b-blocker preoperatively.

Parents and children should be counselled about risk factors for both acute and subacute bacterial endocarditis. In the young child acute bacterial endocarditis, often introduced via skin sepsis, is a more common problem than subacute bacterial endocarditis, and eczema is a very important risk factor. Thus parents must be induced to treat eczema aggressively and oral antibiotics cidal to *Staphylococcus aureus* should be given if eczema becomes infected, or with other purulent skin lesions. Good dental hygiene and regular dental checks should be encouraged from an early age and consideration to early fluoride supplementation should be given where the water supply has a low fluoride content. Dental procedures causing gum trauma should be covered by antibiotic prophylaxis, that is polishing and scaling as well as dental extractions. It is helpful to issue the child with a card giving instructions about recommended antibiotic prophylaxis for both surgical and dental procedures. Finally, both child and parents should be persuaded that body-piercing jewellery should not be worn.

At all ages of the child, appropriate counselling and psychological support to the parents is important, but the older the child the greater the need to focus on the psychological preparation of the child for the type of operation that is planned, and for the stay in hospital and on the intensive care unit.

Postoperative follow-up

Short-term follow-up

Postoperatively the infant or child needs to be monitored for any sign of haemodynamic problems with the cardiac repair, and before discharge home the result of the cardiac operation should be assessed by echocardiography and Doppler. Particular note should be taken of the presence of any pericardial fluid, and in any infant that is unexpectedly tachypnoeic diaphragmatic movements should be assessed by ultrasound to rule out phrenic nerve palsy. In order to reduce risk of infections of the skin incision, parents should be advised not to bathe infants until the skin over the incision is intact and scabs have fallen off. Parents and general practitioners need to be advised about symptoms that may indicate the onset of postpericardotomy syndrome, which generally appear 1 to 8 weeks postoperatively. It is desirable that patients that have had bypass-surgery should have at least one further postoperative cardiac ultrasound within that period looking for increasing pericardial effusion. When present, postpericardotomy syndrome will often respond to anti-inflammatory doses of aspirin or indomethacin, or in severe cases steroids, but in rare cases the onset may be rapid causing tamponade needing pericardial drainage and on occasions surgery with creation of a pericardial window or even pericardectomy may be required.

Post-operative diuretic therapy is generally best tailed off gradually rather than stopped abruptly as the secondary hyperaldosteronism caused by diuretic therapy may lead to acute fluid retention if big doses of diuretics are stopped suddenly.

Long-term follow-up

Patients that have had initial palliative surgery need regular monitoring in order to determine optimal timing of future surgical interventions depending on factors such as pulmonary artery growth and body size, and whether the pulmonary vascular bed is adequately protected from excessive flow/and or pressure. However, even patients that have had ‘corrective surgery' often require long-term follow-up. For instance patients that have had correction of coarctation of the aorta require life-long surveillance of blood pressure and monitoring of the growth of the site of repair. Patients who have had cardiac surgery involving closure of perimembranous ventricular septal defects are at a low, but definite, risk of late development of conduction tissue fibrosis and increasing atrioventricular block, so even patients with a successful VSD repair merit infrequent long-term follow-up. Patients that have had right ventricular outflow incisions, or repairs involving intraventricular baffles, are at significant risk of late development of ventricular arrhythmias and should have occasional 24-h ECG monitoring as part of routine follow-up. The same applies to patients who have had Mustard or Senning corrections or other types of intra-atrial baffles, as they are at considerable risk of late development of sinus node dysfunction and atrial flutter or fibrillation, as well as at risk for late baffle obstruction and dysfunction of the right ventricle that is supporting the systemic circulation.

Palliative procedures in congenital heart disease

Palliative operations and interventional cardiac catheterization have several applications in patients with congenital heart disease. They are used to permit survival of critically ill neonates, to prepare the circulation for later more curative surgery, and to permit an acceptable quality of life in patients with complex malformations not amenable to corrective surgery. A low-risk palliative operation, which results in a good haemodynamic state, can sometimes be a preferable option to a high-risk ‘corrective' operation. The later development of heart surgery for complex congenital heart disease was only made possible by two pioneering palliative procedures: the Blalock–Taussig shunt to increase pulmonary blood flow, and the Rashkind balloon atrial septostomy to improve mixing at the atrial level.

In order to have a satisfactory quality of life, certain haemodynamic conditions must be met: systemic venous and pulmonary venous return must be unobstructed; systemic arterial outflow from the heart must be unobstructed; pulmonary blood flow must be sufficient to give acceptable arterial oxygen saturations, but at a relatively normal pressure, and not so high as to cause congestive heart failure; and if the great vessels are transposed there must be adequate mixing of systemic and pulmonary venous return.

Some patients with ‘univentricular hearts' and pulmonary stenosis fulfil these criteria and may remain completely asymptomatic into adulthood; however, in many patients with complex lesions palliative procedures are used to create these conditions.

Increasing mixing of systemic and pulmonary venous return

Transposition of the great vessels is typical of a condition in which severe arterial desaturation results from inadequate mixing of systemic and pulmonary venous return. There are, however, many other conditions where poor mixing or unfavourable streaming of systemic and pulmonary venous return result in severe hypoxia in the systemic circulation. The initial palliative treatment should be an infusion of prostaglandin E_1 or E_2. This usually improves arterial oxygenation significantly by allowing mixing at ductal level, and by increasing pulmonary blood flow. However, this solution is only temporary and further management depends on the type of eventual surgical correction that is planned. If an arterial switch procedure is planned in the next few days, prostaglandin treatment alone may be sufficient; if an atrial redirection procedure or a later arterial switch is planned it is usually necessary to improve the arterial oxygenation by an atrial septostomy. In the first days or weeks of life, this can be achieved by the use of a Rashkind septostomy catheter, introduced either via the saphenofemoral venous junction or via the umbilical vein. This procedure used to be undertaken only in the cardiac catheterization laboratory, but can be performed in the paediatric intensive care unit under the guidance of ultrasound scanning (Fig. 8).

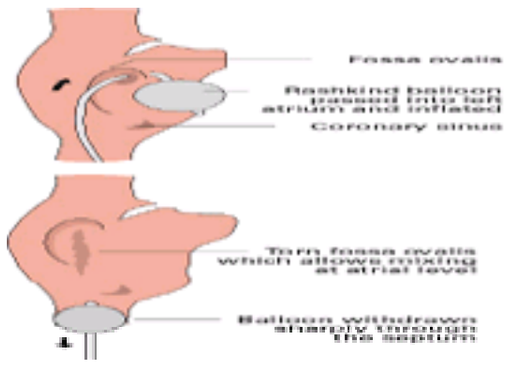


Fig. 8. Rashkind's septostomy is used to improve mixing of systemic and pulmonary venous return. A specially designed catheter with a specially strong inflatable balloon at the tip is introduced into the venous side of the circulation, either via the umbilical vein, cut down to the saphenofemoral junction, or through a percutaneously introduced sheath into the femoral vein. It is advanced into the left atrium under either radiological or echocardiographic screening control. The balloon is then withdrawn forcibly to tear the fossa ovalis and increase venous mixing at atrial level.

In later life, or if a larger opening in the atrial septum is required, a surgical atrial septostomy is generally the procedure of choice. This septostomy can be performed as a closed procedure (Blalock–Hanlon septostomy) or, more commonly, as an open atrial septostomy using cardiopulmonary bypass. Rarely, where there are relative contraindications for surgery, a cardiac catheter blade septostomy, using a percutaneously introduced catheter with a knife blade, can be used to enlarge existing atrial septal defects. However, the opening achieved is not as large as a surgically created defect, and neighbouring structures may accidentally be damaged.

Increasing pulmonary blood flow

A group of cardiac malformations presents with cyanosis secondary to reduced pulmonary blood flow. The most common is tetralogy of Fallot; others include pulmonary atresia with or without a ventricular septal defect, tricuspid atresia with aorta arising from the main chamber and the pulmonary artery from the outlet chamber, and a double outlet right ventricle with valvar or infundibular pulmonary stenosis. If hypoxia is severe in the neonate, prostaglandin E_1 or E_2 infusion temporarily reopens a closing ductus arteriosus. Oral prostaglandin E_2 administration may be used to maintain an open ductus for a period of weeks or even months, but the need for frequent administration makes it a difficult home-based treatment.

In many patients it is necessary to consider creating a surgical shunt to increase pulmonary blood flow. In the neonate the procedure of choice is the modified Blalock–Taussig shunt (Fig. 9). This involves a prosthetic tube (usually Gore-Tex or Impra) being anastomosed end-to-side with either the right or the left subclavian artery and connected end-to-side with the ipsilateral pulmonary artery. If the subclavian artery is extremely small the innominate artery is occasionally used. This shunt has a good patency rate, particularly if a tube of 5 mm, rather than 4 mm, diameter can be used. As long as the shunt is unilateral and connected to confluent central pulmonary arteries it rarely causes excessive pulmonary blood flow, and it is comparatively easy to take down at the time of further surgery. Many surgeons feel that a 2 to 3-month period of treatment with low-dose aspirin, with or without dipyridamole, reduces the risk of thrombotic occlusion.

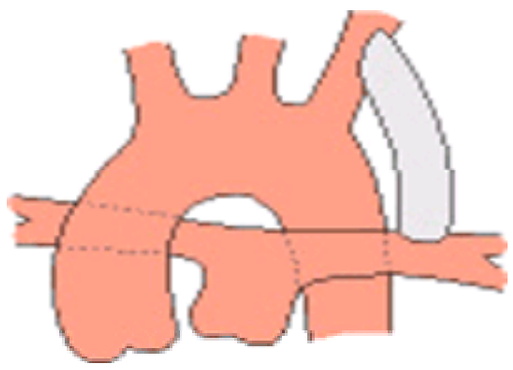


Fig. 9. The modified Blalock systemic/pulmonary shunt is performed through a left or right thoracotomy (fourth interspace). A 5-mm Impra graft is anastomosed to both the subclavian and pulmonary arteries using side or cross clamps.

In older children, particularly those with a non-correctable cardiac malformation who require permanent palliation, the classical Blalock–Taussig shunt (Fig. 10) can produce an excellent long-term result. This shunt is created by anastomosing the right or the left subclavian artery end-to-side to the ipsilateral pulmonary artery and ligating the distal subclavian artery. The arm still receives adequate blood supply from collaterals around the shoulder. The particular advantage with this shunt is the fact that no prosthetic material is used and that the subclavian artery will grow with the patient. It is also useful in the setting of non-confluent small pulmonary arteries where a prosthetic tube shunt can produce excessive flow if connected to one lung only. Some of Helen Taussig's original patients remained alive and reasonably well for over 25 years on their initial shunt.

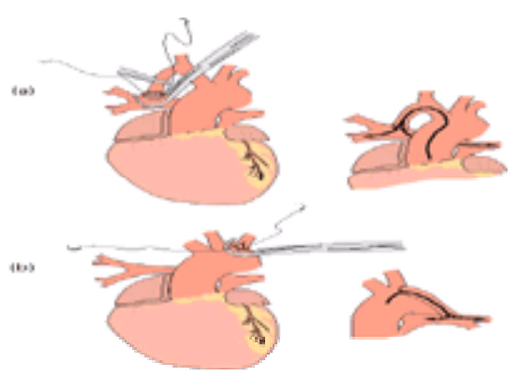


Fig. 10. (a) In the classical procedure the subclavian artery itself is divided and detached at the apex of the chest and brought down to the pulmonary artery in the hilum. (b) For the classical Blalock shunt on the left a longitudinal incision can be made through the root of the subclavian and closed in a transverse manner to prevent kinking as the vessel is brought down to the pulmonary artery.

The Waterston shunt (a fenestration between ascending aorta and right pulmonary artery) and the Potts shunt (a fenestration between descending aorta and left pulmonary artery) are rarely used nowadays. The former has a relatively high incidence of pulmonary vascular disease and of severe distortion of the right pulmonary artery, and the latter is extremely difficult to take down at the time of later surgery. Some surgeons still use the Waterston shunt when the pulmonary arteries are extremely small, but a central shunt using a tube prosthesis from the aortic arch to the main pulmonary artery is preferable (Fig. 11). The central shunt is also useful in patients with pulmonary atresia, ventricular septal defect, and multifocal pulmonary blood supply, in whom it helps to enlarge a hypoplastic central pulmonary artery confluence when used in conjunction with attempts at unifocalization of pulmonary blood supply.

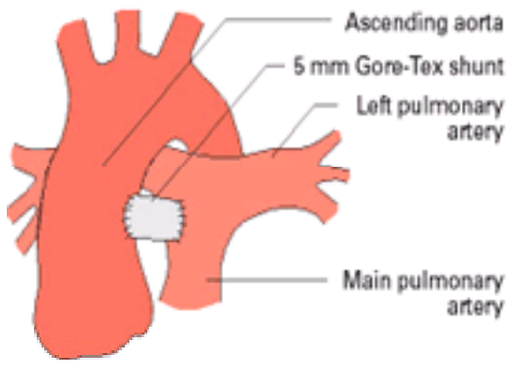


Fig. 11. A central shunt is performed through a median sternotomy incision using side clamps. A short length of Gore-Tex or Impra is interposed between the two vessels and can be easily closed at a second procedure using clips.

The Glenn shunt is a venous shunt which, in its classical form, is constructed by anastomosing the right superior vena cava end-to-end with the divided right pulmonary artery (Fig. 12(a)). For a period it went out of favour because some patients developed late pulmonary arteriovenous fistulae with increasing cyanosis. This complication was thought to have been a reaction of the pulmonary vasculature to the absence of pulsatile flow but may be a response to the absence of a trophic factor present in hepatic venous blood. There has been a great resurgence in the use of this shunt in recent years, as in many ways it is ideal for preparing the circulation for a later Fontan procedure, since it provides good pulmonary blood flow at very low pressure, thus avoiding the risk of precipitating a rise in pulmonary vascular resistance. Furthermore, it does not cause a volume overload of the systemic ventricle, which may lead to dysfunction, particularly diastolic, of this chamber. Nowadays this shunt is usually constructed as a bidirectional Glenn shunt by anastomosing the right superior vena cava end-to-side to the right pulmonary artery (Fig. 12(b)). This has the advantage of distributing pulmonary blood flow equally to the two lungs and preserving some pulsatility in the pulmonary vascular bed if an additional source of pulmonary blood flow is preserved. Used in this way it is sometimes referred to as a ‘hemi-Fontan'. In patients with laevoisomerism (‘bilateral left-sidedness') systemic venous drainage may be via a left, or bilateral venae cavae, and under these circumstances a left-sided, or bilateral Glenn shunt may be constructed. The only drawback of this shunt is that it cannot be used in neonates as their pulmonary vascular resistance is too high. It is mostly used in children over 6 months of age, although occasionally it has been performed in 3-month-old infants. If used as the only source of pulmonary blood flow, pulmonary artery growth may be compromised long-term as even in an infant a Glenn-shunt will provide a pulmonary blood flow that is maximally about 60 per cent of systemic blood flow.

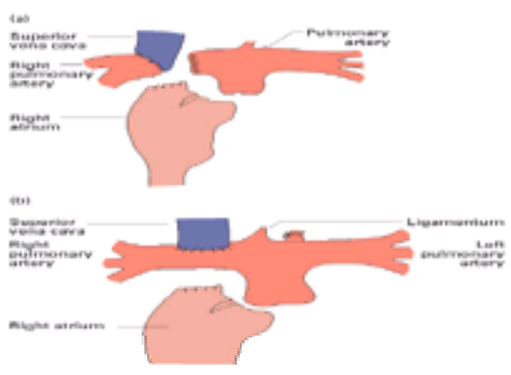


Fig. 12. (a) The classical Glenn shunt is performed through a right thoracotomy in the fourth interspace without the use of cardiopulmonary bypass. The transected superior vena cava is anastomosed directly to the right pulmonary artery. (b) In the bidirectional Glenn shunt either side clamping or cardiopulmonary bypass can be used. Superior venal caval blood is then distributed to both pulmonary arteries. Total cavopulmonary connection can be achieved using cardiopulmonary bypass at a later date if necessary.

Where reduced pulmonary blood flow is mainly due to valvular obstruction an obvious approach is to enlarge the pulmonary valve orifice; however, in all patients with a large coexisting ventricular septal defect there is the risk of flooding the pulmonary circulation at too high a pressure. The most controlled method is probably via catheter balloon dilatation with a carefully sized balloon, but a surgical outflow-tract enlargement is occasionally used as a first preparatory stage in the treatment of patients with tetralogy of Fallot and a severely hypoplastic main pulmonary artery.

In pulmonary atresia with intact ventricular septum, the right ventricle is usually severely hypoplastic and unable to support the pulmonary circulation, even if continuity between right ventricle and pulmonary artery can be re-established with a valvotomy. However, if the management is by a systemic-to-pulmonary shunt only, the right ventricle will not grow. Optimal management therefore consists of securing pulmonary blood flow, usually with a modified Blalock shunt but occasionally with long-term prostaglandin treatment, followed by a valvotomy or right ventricular outflow reconstruction to establish forward flow through the right ventricle. It is important that the outflow obstruction is relieved as completely as possible: there is usually substantial tricuspid valve incompetence in this condition, and this severely compromises forward flow if there is a large outflow gradient remaining. However, before decompressing the right ventricle it is important to establish that there are no important coronary artery sinusoids communicating with the right ventricular cavity and creating a partially right ventricular-dependent coronary artery circulation. If there is only a patent foramen ovale, or if the atrial septal defect is small, it is important also to secure unobstructed inflow of systemic venous blood across the atrial septum into the left heart chambers by carrying out a septostomy, usually a Rashkind septostomy.

Reducing excessive pulmonary blood flow

Whereas at birth the pulmonary vascular resistance is approximately equal to systemic vascular resistance, postnatally it should normally drop rapidly, so that at 3 months of age it is normally a small fraction of that of the systemic circulation (resistance ratio about 0.02–0.1). Thus, if there are large defects in the interventricular septum, if both arteries arise from the same ventricle, or if a functionally common ventricle exists, the pulmonary circulation will be flooded by high flow at systemic pressure. This often precipitates congestive heart failure and carries the risk of causing irreversible pulmonary vascular disease. If it is caused by a single ventricular defect alone immediate surgical closure of the defect is usually appropriate. On the other hand, if there are multiple ventricular septal defects, particularly if they are of the ‘Swiss cheese' variety, or if there is a common atrioventricular septal defect with unbalanced ventricles or other complications of atrioventricular valve morphology, or if there are associated lesions such as coarctation of the aorta, the pulmonary circulation can be protected by pulmonary artery banding (Fig. 13). In this operation a silk tie is applied around the main pulmonary artery and tightened until a desired ratio between distal pulmonary artery pressure and systemic artery pressure is achieved. Pulmonary artery banding is also used to protect the pulmonary circulation in patients with complex heart disease, but the actual tightness of the band may vary depending on the timing of future planned procedures. Since the band does not grow with the patient, the average pulmonary artery band will initially still admit moderately excessive flow. It usually produces an ideal balance between pulmonary and systemic flow at around 12 months of age, and then gradually restricts pulmonary blood flow so that some right to left shunting starts occurring, usually between 2 and 4 years of age. In some patients with complex heart disease it may be appropriate to detach the pulmonary arteries from the heart and to feed the pulmonary circulation through a shunt; this applies, for example, in the Norwood stage 1 procedure for hypoplastic left heart syndrome, where the pulmonary artery is used to reconstruct an aortic arch.

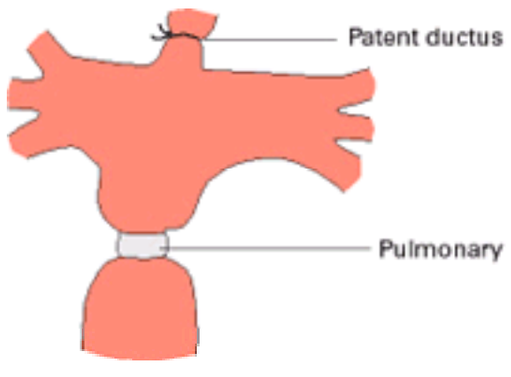


Fig. 13. Pulmonary artery banding is performed to reduce blood flow to the lungs in patients with large left to right intracardiac shunts. The main pulmonary artery is approached through a left thoracotomy. Associated defects, such as coarctation and patent, ductus can also be corrected. The band is tightened progressively whilst observing both pulmonary and systemic blood pressure and oxygen saturation. There is a balance between the band being sufficiently tight to prevent pulmonary vascular disease and deterioration in haemodynamics and oxygen saturation.

Removing inflow obstruction

The most common form of inflow obstruction in congenital heart disease consists of atresia of one of the atrioventricular valves. In patients with atresia of the right atrioventricular valve (usually morphologically a tricuspid valve) the systemic venous return has to cross a patent foramen ovale or an atrial septal defect to reach the ventricles. If this opening is restrictive in a neonate a Rashkind balloon atrial septostomy is performed; in the older infant or child an open atrial septostomy is preferable. Atresia of the left atrioventricular valve (usually morphologically a mitral valve) is associated with pulmonary venous congestion unless a good size atrial septal defect is present to allow unrestricted flow into the right atrium, and a Rashkind atrial septostomy is often required in the neonate. In later life the interatrial communication may often need to be enlarged by an open atrial septectomy. Unless it is anticipated that atrial septal structures will be used in subsequent surgery, the excision of the atrial septum in this situation should be wide, rather than including only the fossa ovalis membrane, so that functionally a single atrium is achieved. Where there is congenital severe stenosis, as opposed to atresia, of an atrioventricular valve balloon dilatation of the valve, at least of the right atrioventricular valve, may find a role in management but this still remains to be evaluated.

Relief of systemic outflow obstruction

Correctable causes of outflow obstruction, such as valvular stenosis or coarctation, should be corrected in patients with complex congenital heart disease, since if they coexist with an unprotected pulmonary circulation the effect of pulmonary artery banding will be severely diminished by the systemic outflow gradient, and irreversible pulmonary vascular disease will occur early. An example of a difficult situation is the child with either tricuspid atresia or double-inlet left ventricle where the great arteries are ‘transposed' so that the pulmonary artery comes off the main chamber (the left ventricle), and the outflow chamber is filled via a restrictive ventricular septal defect and gives rise to the aorta, often with muscular subaortic stenosis and/or coarctation of the aorta. In the past these children were treated with pulmonary artery banding with or without coarctation repair, but the mortality rate was high and the few survivors developed pulmonary vascular disease. In recent years, more radical palliative procedures are therefore becoming adopted. One approach is a variant of the Damus–Kaye–Stansel procedure, in which the pulmonary artery is transected near the bifurcation and the proximal end is anastomosed end-to-side to the aorta. The pulmonary circulation is then fed via a shunt. The mortality rate associated with this procedure is substantial, often related to the difficulty in the immediate postbypass period in establishing sufficient pulmonary blood flow through a shunt into a pulmonary vascular bed that is adapted to flow at high pressures. There have also been some long-term problems with incompetence of the previously pulmonary, now systemic, valve, but generally the haemodynamic state of the survivors is good. Lincoln has advocated using surgical enlargement of the ventricular septal defect (Fig. 14) (with or without a subaortic gusset) in these patients, and the surgical mortality of this procedure appears lower. However, it is difficult to achieve a sufficiently large ventricular septal defect in a small neonate, and restriction of flow between main chamber and outflow chamber may recur. This procedure needs to be combined with pulmonary artery banding if carried out as the first palliation. Again, the symptomatic state of survivors is good. One advantage with this operation is that it can be used to improve the haemodynamics even in patients with pulmonary vascular disease.

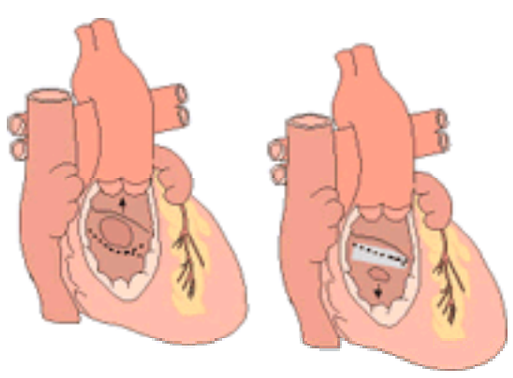


Fig. 14. Safe areas for enlargement of the ventricular septal defect in patients in whom the aorta arises from the infundibular chamber. When the defect is in normal position (left) it is enlarged towards the aortic valve to avoid the conduction bundle. When the defect is muscular, the location of the bundle varies but is often superior so defect enlargement is usually toward the trabecular chamber.

The most complete form of systemic outflow obstruction is aortic atresia, usually combined with left ventricular hypoplasia to create the hypoplastic left heart syndrome, which is the most common cause of death from congenital heart disease. As mentioned above, Norwood has described an initial palliative procedure, which by using the pulmonary artery to reconstruct the aortic arch, transforms the heart to a ‘univentricular heart of right ventricular morphology', with the pulmonary circulation fed from a systemic–pulmonary shunt into a reconstructed pulmonary confluence. If pulmonary vascular resistance remains low in later life a ‘correction' is attempted, using a Fontan or total cavopulmonary connection procedure. The combined surgical mortality rate of this series of operations remains substantial in most centres, and there are doubts about the long-term ability of the right ventricle and half a right atrium to support the systemic circulation. There are reasons to suspect that many patients with Norwood palliation will eventually require a heart transplant in later life. The other possible active treatment in this condition is a neonatal heart transplant, and the Norwood stage 1 operation may serve as a holding procedure, with the aim of later transplantation. However, the supply of potential human neonatal heart donors will never match the need if all children with the hypoplastic left heart syndrome receive active treatment, and it is important to discuss fully with the parents whether they really wish for active treatment to be carried out.

Reducing right-to-left shunting

There are many conditions in which right-to-left shunting across an atrial septal defect or a patent foramen ovale aggravates arterial hypoxia in obstructive right-sided heart lesions. This situation is usually best dealt with by correction of the underlying anomaly. One special case in which a ‘palliative' closure of the atrial septal defect may be helpful is Ebstein's anomaly of the tricuspid valve. Because the tricuspid valve is usually severely incompetent, and sometimes mildly stenotic, in severe cases of the anomaly, right-to-left shunting at the atrial level is often substantial if there is an atrial septal defect present. If hypoxia is the major symptom, and the right ventricle is judged to be able to accommodate pulmonary blood flow, the mortality associated with closure of the atrial septal defect alone is much lower than if the procedure is combined with valve replacement or tricuspid reconstructive surgery.

General comments

It is important that the choice of palliative procedure is based on the fullest diagnostic information possible. This can often be obtained non-invasively by cardiac ultrasound and Doppler studies. Considerable thought has to be given to the likely final correction or palliation required. For instance, in a child with transposition of the great arteries, severe pulmonary stenosis, and a ventricular septal defect with a straddling atrioventricular valve, a biventricular repair is not possible. Initially, a Rashkind balloon atrial septostomy should be considered, even if the child is not critically hypoxic, because long-term mixing at atrial level will become important. Later the child is likely to need a procedure to increase pulmonary blood flow. As the final palliation is likely to be a Glenn shunt combined with an open atrial septectomy, or a total cavopulmonary connection, it is important that if any systemic-to-pulmonary shunt has to be performed before the child is large enough for a Glenn shunt, it should be a left-sided shunt, usually a modified Blalock, as a right-sided Blalock shunt would make a later Glenn shunt very difficult. Thus, if palliative procedures are chosen correctly, and carried out at the optimal time, even a child with uncorrectable congenital heart disease can have a good future.

Conclusions

The optimal management of congenital heart disease should aim at creating favourable conditions for attempting complete surgical repairs as described in the following chapters wherever the anatomy allows it. However, good surgical results both from complete repairs and from palliative procedures require detailed understanding of the anatomy, regular monitoring of the child, individualized medical supporting therapy, and close liaison between the paediatric cardiologist and the cardiac surgeon.

Further reading

Anderson RH, Becker AE, Lucchese FA, Meier MA, Rigby ML, Soto B. *Morphology of congenital heart disease. Angiographic, echocardiographic and surgical correlates*. Castle House, Tunbridge Wells, Kent, 1983.

Anderson RH, Tynan M. Sequential chamber localisation. *Recent Advances in Cardiology* 1981; **8**: 265–85.

Archer N, Burch M. *Paediatric cardiology: an introduction.* Edward Arnold, London, 1998.

Bank ER, Hernandez RJ. CT and MR of congenital heart disease. *Radiological Clinics of North America* 1988; **26**: 241–62.

Ebert PA. *Atlas of congenital cardiac surgery*. Churchill-Livingstone, New York, 1989.

Gussenhoven EJ, Vanherwerden LA, Roselandt J, Ligtvoet KM, Bos E, Witsenburg M. Intraoperative two-dimensional echocardiography in congenital heart disease. *Journal of the American College of Cardiology* 1987; **9**: 565–72.

Kirklin JW, Barratt-Boyes BG. *Cardiac surgery*, Vols 1 and 2, 2nd edn. Churchill-Livingstone, New York, 1993.

Nihill MR. Catheterisation and angiography. In: Garson A, Bricker JT, Fisher DJ, Neish SR (eds). *The science and practice of paediatric cardiology*, 2nd edn. Williams and Wilkins, Baltimore, 1997, pp. 995–1019.

Park MK. *Paediatric cardiology for practitioners*, 3rd edn. Mosby, St Louis, 1996.

Park MK, Guntheroth WG. *How to read pediatric ECGs*, 3rd edn. Mosby, St Louis, 1992.

Perloff JK, Child JS. *Congenital heart disease in adults*. WB Saunders, Philadelphia, 1991.

Salmon AP, Keeton BR, Sethia B. Developments in interventional catheterisation and progress in surgery for congenital heart disease: achieving a balance. *British Heart Journal* 1993; **69**: 479–80.

Seward J, *et al.* Transesophageal echocardiography: technique, anatomic correlations, implementation and clinical applications. *Mayo Clinic Proceedings* 1988; **63**: 649–80.

Sreeram N, *et al.* Changing role of noninvasive investigation in the preoperative assessment of congenital heart disease: a nine year experience. *British Heart Journal* 1990; **63**: 345–9.

Stumpflen I, Stumpflen A, Wimmer M, Bernaschek G. Effect of detailed fetal echocardiography as part of routine pre natal ultrasonagraphic screening on detection of congenital heart disease. *Lancet* 1996; **348**: 854–7.

Van Praagh R. Terminology of congenital heart disease. *Circulation* 1977; **56**: 139–43.

Vargo T. Cardiac catheterisation: hemodynamic measurements. In: Garson A, Bricker JT, Fisher DJ, Neish SR (eds). *The science and practice of pediatric cardiology*, 2nd edn. Williams and Wilkins, Baltimore, 1997, pp. 961–93.

Wylie J, Wren C, Hunter S. Screening for fetal cardiac malformations. *British Heart Journal* 1994; **71** (suppl.): 20–27.

40.7.2 Patent ductus arteriosus and aortic coarctation

Emile A. Bacha and Gus J. Vlahakes

Patent ductus arteriosus
Anatomy and physiology
Diagnosis
Complications
Treatment
Operative technique
Results
Coarctation of the thoracic aorta
Embryology and pathology
Diagnosis
Natural course and indications for operation
Operative technique
Results and complications
Further reading

Patent ductus arteriosus

Closure of the patent ductus arteriosus by Gross in 1938 marked the beginning of modern cardiac surgery. Although the patent ductus is frequently associated with other cardiac anomalies, this chapter will concentrate on the infant in whom this is an isolated problem.

Anatomy and physiology

In a full-term infant the ductus arteriosus connects the main pulmonary trunk with the descending aorta about 5 to 10 mm distal to the origin of the subclavian artery. It is 5 to 10 mm in length and has an average diameter of 10 mm. The aortic orifice is wider than the pulmonary, although large variations in size and shape are possible. The recurrent branch of the left vagus nerve loops around the lateral and inferior aspect of the ductus and ascends to the larynx (recurrent laryngeal nerve).

The ductus arteriosus is crucial for the fetal circulation, permitting flow from the right ventricle (and hence the placental circulation) to the systemic bed, bypassing the high-resistance prenatal pulmonary circulation. The ductus thus carries 55 to 60 per cent of the combined ventricular output: it is therefore nearly equal in size to the descending aorta and angiographically appears as a smooth extension of the main pulmonary artery into the descending aorta.

Postnatal closure of the ductus occurs in two stages. Some 10 to 15 h after birth (in full-term infants), contraction of the medial smooth muscle leads to shortening of the duct and protrusion of intimal cushions into the lumen, resulting in a functional closure. Some 2 to 3 weeks after birth, infolding of the endothelium with necrosis and proliferation of subintimal layers leads to fibrosis and permanent sealing of the lumen. Closure begins at the pulmonary end of the duct; it may remain incomplete at the aortic end.

Ductal closure is mediated by the interaction of oxygen tension and endogenous prostaglandins. Prostaglandins, especially prostaglandin E_2, keep the ductus open, while high oxygen tension after birth leads to duct closure. Response to prostaglandin and oxygen varies with duct maturity: in the full-term infant the duct is sensitive to oxygen, while in the premature infant, prostaglandins have a dominant effect. Consequently, in premature infants the failure of ductal closure results from incomplete ductal development, whereas in full-term infants a patent ductus arteriosus results from structural abnormalities of ductal tissue. Under physiologic conditions, other factors such as vasoactive substances (such as acetylcholine, bradykinin, and endogenous catecholamines, or variations in pH) may contribute to closure of the ductus.

The ductus arteriosus is completely closed by 8 weeks of age in 88 per cent of infants with an otherwise normal cardiovascular system. The incidence of isolated patent ductus arteriosus in term infants is approx. 1 in 200 live births; it accounts for 5 to 10 per cent of all types of congenital heart disease. The incidence increases markedly with increasing prematurity.

Diagnosis

Clinical symptoms are related to the shunt flow between the aorta and pulmonary artery, which is determined by the size of the ductus, systemic and pulmonary vascular resistances, and systemic blood pressure.

In neonates with a large patent ductus arteriosus, pressures in the aorta and pulmonary artery are practically equal. As neonatal pulmonary vascular resistance decreases after birth, left-to-right shunting rapidly increases, and severe left ventricular failure with pulmonary edema may occur early in infancy.

Physical examination shows a hyperdynamic precordium, occasionally with a palpable systolic thrill and cardiac enlargement. The pulse is bounding and the pulse pressure is wide. The first and second heart sounds are accentuated; a murmur is present, which is more systolic than ‘machinery' in its characteristics. Occasionally a mid-diastolic mitral flow rumble can be heard at the apex.

Many patients with moderate patent ductus arteriosus remain asymptomatic until the second decade of life, when easy fatiguability and exertional dyspnea occur. Findings on physical examination are less marked than those in patients with a large patent ductus arteriosus. The classical continuous murmur, which masks the heart sounds and is described as ‘machinery' in its quality, leads to the diagnosis.

Most patients with a small patent ductus arteriosus are asymptomatic, a systolic murmur in the second left interspace being the only detectable abnormality.

Clinical symptoms improve in infants with a significant, uncorrected left-to-right shunt as pulmonary vascular resistance increases. Once pulmonary vascular resistance exceeds the systemic vascular resistance, cyanosis develops and the condition may then be irreversible. Generally, these changes begin by the age of 2 years; they may take several years to develop.

Electrocardiogram

Left ventricular hypertrophy and left ventricular enlargement with left atrial hypertrophy are present in patients with a significant shunt, but may be absent if the shunt is small. Right ventricular hypertrophy occurs with the appearance of increased pulmonary hypertension.

Radiography

The chest radiograph may be normal in patients with a small patent ductus arteriosus. Enlargement of the heart with pulmonary overperfusion, with or without pulmonary edema, and a prominent ascending aorta are typical findings in patients with a moderate or large patent ductus arteriosus.

Echocardiogram, cardiac catheterization, and angiography

In infants and young children with typical, uncomplicated patent ductus arteriosus, cardiac catheterization is not necessary for diagnosis, decision-making, and surgical ligation. Two-dimensional echo-cardiography with Doppler flow imaging is usually definitive and allows direct visualization of the ductus, confirmation of ductal flow, and assessment of the magnitude of the shunt.

If catheterization is performed, an increase of pulmonary arterial blood oxygen content of more than 0.5 ml/dl or a saturation increase of more than 4 to 5 per cent over that of right ventricular blood indicates a significant left-to-right shunt. If the anatomic condition is unclear, angiocardiography or magnetic resonance imaging can

provide additional information (Fig. 1).

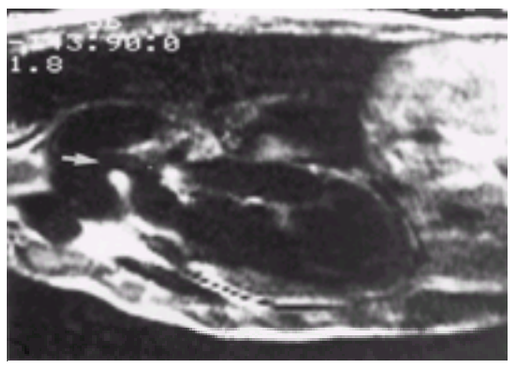


Fig. 1. Magnetic resonance imaging in a sagittal–coronal view showing a patent ductus arteriosus (arrow) in an adult.

Complications

Congestive heart failure, pulmonary hypertension with vascular damage, infective endarteritis, and aneurysmal dilatation of the ductus complicate the clinical course of patients with untreated patent ductus arteriosus.

Treatment

Adult and full-term infant

Once the diagnosis has been made, surgical closure of the duct is indicated. Elective operation should be performed in asymptomatic children before the age of 2 years. Surgical closure within the first 6 months of birth is indicated only in symptomatic patients, since spontaneous closure of the duct may still occur.

Medical management in this group focuses on the treatment of cardiac failure; indomethacin (see below) is not effective for producing duct closure beyond the newborn period. Occlusion of the patent ductus arteriosus can be achieved by transfemoral catheterization (Gianturco coils or Rashkind double-umbrella device) in selected cases in larger infants and children. However, short patent ductus and other anatomic variations are associated with reopening and residual patency. Videothoracoscopic techniques are also increasingly being used for closure of patent ductus arteriosus by clip application in older infants and children.

Pulmonary hypertension is not a contraindication to surgery, as long as the shunt flow is predominantly left to right and is not reversed because of severe pulmonary vascular disease.

Preterm infant

All neonates with a birth weight of 1000 g or less have a patent ductus arteriosus; this is hemodynamically significant in 40 per cent of these infants. A symptomatic or significant patent ductus arteriosus is strongly correlated with respiratory distress, necrotizing enterocolitis, and cardiac failure. Increased systemic oxygenation, diuretic therapy, and indomethacin (0.2 mg/kg intravenously) can induce ductal closure in most neonates. Up to three doses of indomethacin can be given safely at an interval of every 12 to 24 h, provided renal function is monitored closely. Contraindications to indomethacin therapy include bleeding disorders, impaired renal function, intracerebral hemorrhage, and necrotizing enterocolitis.

If medical treatment is ineffective or contraindicated, surgical closure is warranted. To avoid a potentially dangerous transport to the operating room, surgical closure can be safely performed in the neonatal intensive care unit with the full surgical team in attendance. Local analgesia can be added to systemic sedation to provide enough anesthesia.

Operative technique

Open

The chest is entered through a posterolateral thoracotomy in the fourth interspace and the mediastinal pleura is opened over the upper descending aorta (Fig. 2). The ductus is freed from its attachments to the mediastinal tissue and a heavy ligature is passed around it to facilitate further dissection. The recurrent laryngeal nerve should not be dissected free. The ductus can be interrupted by division or ligation; both produce excellent results.

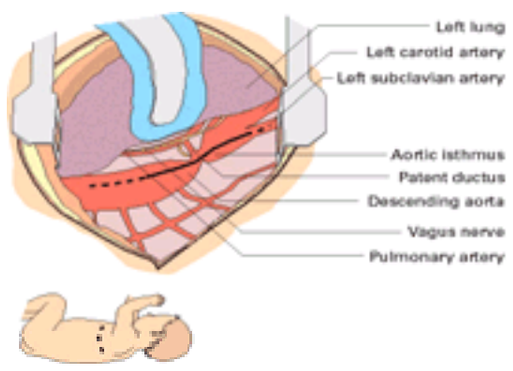


Fig. 2. Surgical approach to patent ductus and aortic coarctation. The chest is entered by a posterolateral incision through the left fourth intercostal space. The lungs are retracted anteriorly and the pleura incised. The incision in patients with isolated patent ductus arteriosus (solid line) is not as extended as in those with aortic coarctation (dashed line).

In preterm infants, interruption of the ductus with a medium-sized, stainless-steel or titanium, hemostatic clip (Hemoclip®; Weck and Co.) is the preferred technique. In infants and children under 1 year of age the ductus is usually ligated with one strand of No. 1 braided polyester ligature (Tevdek®; Deknatel Inc.) on the aortic side and a hemostatic clip on the pulmonary arterial side (Fig. 3). In older children, or in those with a large ductus, the ductus should be divided, the aortic and pulmonary ends being oversewn with continuous 5–0 polypropylene (Prolene®; Ethicon Inc.) suture (Fig. 4). Closure of patent ductus arteriosus in adults, especially when the aortic end is calcified, presents special problems. Here it is best to avoid dissection or clamping of the calcified aortic side of the ductus. Vascular clamps are placed on the noncalcified, pulmonary arterial side, and the rim of pulmonary artery surrounding the insertion of the ductus can be divided between the two clamps and oversewn. Alternatively, cardiopulmonary bypass and closure through the open pulmonary artery may be required.

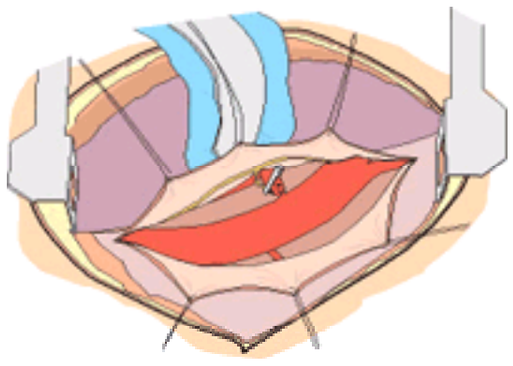


Fig. 3. Interruption of patent ductus arteriosus in infants and children under 1 year of age: after identification of the recurrent laryngeal nerve and dissection of the ductus, the aortic side of the ductus is ligated; a metal clip is placed on the pulmonary artery side.

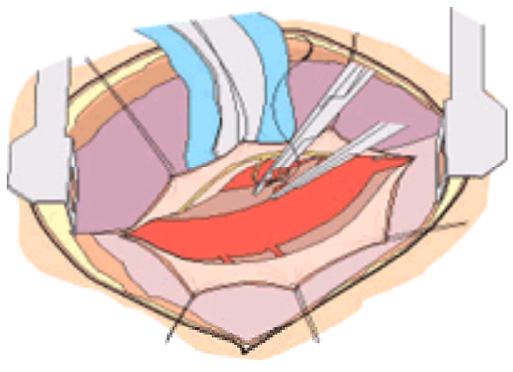


Fig. 4. Division of patent ductus arteriosus in older children or in patients with a large ductus: vascular clamps are placed at the aortic and pulmonary arterial ends of the ductus; after division, the aortic and pulmonary arterial ends are closed with a continuous suture and the clamps removed.

Postoperative care of patients with repair of an uncomplicated patent ductus arteriosus should be uneventful. The chest tube is removed in the operating room or a few hours after surgery, and the child is usually discharged 24 to 48 h after surgery.

Video-assisted thoracoscopic approach

Under general anesthesia and in the right lateral decubitus position, four 3-mm incisions are made in the posterolateral chest, along the line of a standard posterolateral thoracotomy. Ports are placed through these incisions without cutting muscle or spreading the ribs. Endoscopic instruments are passed through the ports. An expandable retractor holds the lung aside. The ductus is freed from the surrounding structures and a standard vascular clip is placed with an endoscopic applicator. Transesophageal echocardiography is used to demonstrate the elimination of ductal flow during clip application.

Results

Potential complications, especially in preterm infants, include residual patent ductus arteriosus, injury to the recurrent laryngeal or vagus nerves, trauma to the lungs (with or without pneumothorax), chylothorax, pneumonia, and infection of the wound. However, the operation is considered to be one of the safest in cardiac surgery. Hospital mortality is almost zero in patients with an uncomplicated patent ductus arteriosus. Cardiac failure, pulmonary hypertension, and advanced age increase the operative risk; closure of an uncomplicated patent ductus arteriosus in infancy or childhood provides a normal life expectancy. These results support closure of a patent ductus even in asymptomatic patients, when compared to the risks and complications of leaving the condition.

Coarctation of the thoracic aorta

Coarctation of the thoracic aorta affects 5 to 8 per cent of infants with congenital heart disease, with a male:female ratio of 2:1 for isolated coarctation. The lesion varies considerably in severity of constriction, as well as in its length and location. It may be associated with other left-sided obstructive cardiac lesions.

Embryology and pathology

Two major types of coarctation can be identified, depending on the blood supply of the lower body.

The ‘adult' type

A localized, thin membrane of infolded aortic media and intimal hyperplasia is usually found in a juxta- or postductal position, and is considered to be derived from ductal tissue. Blood flow to the lower body is supplied by the left ventricle through the ascending aorta and most patients survive to adult life. Other cardiac anomalies are rare, but the most common is a bicuspid aortic valve.

The ‘infantile' or ‘preductal' type

A local membrane is associated with a diffuse narrowing of the aortic isthmus, called tubular hypoplasia. Blood supply to the lower body may be provided by the right heart through a patent ductus arteriosus; patients generally become symptomatic during infancy. This lesion is often associated with other cardiac abnormalities such as patent ductus arteriosus, atrial and ventricular septal defects, and other left-sided obstructive lesions, such as aortic and mitral stenosis.

The exact cause of coarctation remains controversial: ectopic ductal tissue or changes in the fetal flow pattern in the ductus and ascending aorta have been suggested.

Collateral arterial circulation from branches of the subclavian, internal mammary, intercostal, and scapular arteries is variable. It is well developed in adults but unusual in infants.

Diagnosis

Infants with coarctation generally present with congestive heart failure from volume overload if a ventricular septal defect is present, and from increased afterload in the case of severe coarctation with an intact ventricular septum. Signs of systemic hypoperfusion, possibly including metabolic acidosis, can be present. Physical examination shows a hyperdynamic precordial impulse and a nonspecific pansystolic murmur along the left upper sternal border. Murmur characteristics may change with different associated cardiac anomalies. In patients with normal cardiac output, hypertension can usually be detected in the upper limbs. A systolic pressure difference between upper and lower extremities of more than 20 mmHg is usually present, but may be absent if cardiac failure is severe, when reduced aortic blood flow yields a smaller gradient across the coarctation.

Older children and adults are generally asymptomatic. The first finding is either a murmur or systemic hypertension. Physical examination shows decreased, delayed, or absent femoral pulses. In less severe cases or when collateral vessels provide enough distal blood flow, peripheral pulses may be relatively equal. Measurements of blood pressure in the upper and lower extremities are necessary to detect the gradient, which will often increase with exercise of the lower extremities.

A pressure difference between both arms suggests that the origin of one of the subclavian arteries, in most cases the left, is below the coarctation.

Electrocardiogram

Infants with symptomatic coarctation may show deviation of the right axis and signs of right ventricular hypertrophy. In older children and adults, left ventricular hypertrophy is the dominant sign.

Radiography

The symptomatic infant generally shows cardiomegaly. In older patients the chest radiograph reveals a prominent left ventricle with the cardiac silhouette at the upper limit of normal. After 8 to 10 years of age, characteristic rib notching may occur, caused by markedly enlarged intercostal arteries.

Echocardiography, cardiac catheterization, and angiography

Echocardiography can establish the diagnosis and assess associated cardiac anomalies, especially in the critically ill infant. In neonates and young infants, cardiac catheterization is usually not done and echocardiographic data are sufficient. In older patients, especially if associated cardiac lesions are present or suspected, cardiac catheterization and angiography, or magnetic resonance imaging (Fig. 5), are used to delineate the precise anatomy of the lesion and the collateral circulation, and in adults, to confirm the status of the coronary arteries.

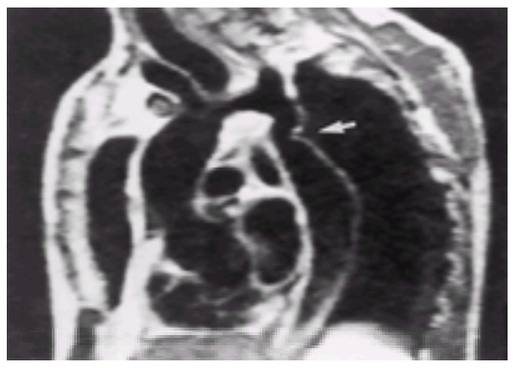


Fig. 5. Sagittal–coronal magnetic resonance image of an aortic coarctation (arrow) in an adult; the ascending aorta is dilated and the isthmus is hypoplastic.

Natural course and indications for operation

Life expectancy in patients with isolated coarctation is about half of that of the normal population: death is due mainly to heart failure, rupture of the aorta, and cerebral hemorrhage.

Treatment of coarctation is surgical. Elective repair is now recommended at 3 to 6 months of age, or at the time of diagnosis if the coarctation is discovered after this age; delay increases the risk of persistent postoperative hypertension.

The use of prostaglandin E_1 is routine in neonates who are *in extremis* secondary to coarctation; in the majority of cases, this is to restore or maintain ductal patency and thereby support lower body circulation by the right ventricle, as well as to avoid left heart failure. The initial dose is 0.1 µg/kg per minute; this can be reduced in small increments to a minimum dose of 0.05 µg/kg per minute as long as ductal patency is maintained. After stabilization of the medical condition of the infant, surgery should be performed as soon as possible. Although uncommon, lack of an immediate response to prostaglandin treatment is an indication for emergency surgical intervention.

Operative technique

Before placing the patient in the right lateral decubitus position, a monitoring line is placed in the right radial artery. Another arterial line in a lower extremity (or the umbilical artery in a neonate) is ideal for assessing the gradient after repair. The chest is entered via a left posterolateral thoracotomy in the fourth intercostal space or through the bed of the fifth rib. The mediastinal pleura is incised over the aorta posterior to the vagus nerve, starting about 4 cm below the coarctation and up to the entire proximal portion of the left subclavian artery (Fig. 2). Staying just superficial to the adventitia, the proximal left subclavian artery with the adjacent parts of the aortic arch and the ligamentum arteriosus are dissected and isolated with tapes. Dissection of the aorta is continued distal to the coarctation, avoiding injury to the friable intercostal arteries. The ligamentum arteriosus or patent ductus is then ligated and divided.

Although several techniques have been described over the years, three have now emerged as the techniques of choice for routine repair of coarctation: the subclavian flap aortoplasty, resection and end-to-end anastomosis, and the synthetic patch aortoplasty. The subclavian flap is preferred only in young infants with a very long coarctation or hypoplastic isthmus (Fig. 6). The aortic cross-clamp is placed proximally between the left common carotid and subclavian arteries, and a second clamp is placed well distal to the coarctation but preferably proximal to the intercostal arteries. The subclavian artery is ligated, divided, and incised longitudinally. The incision is continued onto the aorta and is carried beyond the coarctation. The local membrane shelf is excised carefully, taking care not to penetrate the posterior aortic wall. The subclavian flap is turned down and sutured continuously along the edges of the incision using 5–0 or 6–0 monofilament absorbable sutures (polydioxanone, PDS®: Ethicon Inc.). Alternatively 6–0 or 7–0 interrupted nonabsorbable sutures can be used. It is crucial for the success of this technique that the aortotomy extends well beyond the coarctation.

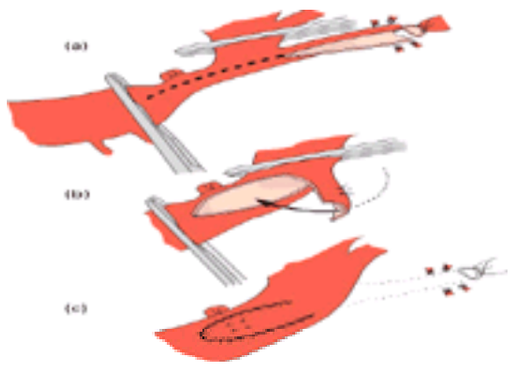


Fig. 6. Subclavian flap aortoplasty for repair of coarctation in young patients with a hypoplastic isthmus. (a) Following pleural incision, the coarctation, the adjacent aorta, and the entire proximal subclavian artery are dissected. A patent ductus or ligamentum arteriosum is then ligated and divided. Clamps are placed and the subclavian artery is ligated and longitudinally incised. The incision is carried well beyond the coarctation (dashed line), and the membrane shelf is carefully excised, avoiding penetration through the posterior aortic wall. (b) The subclavian artery is turned down. (c) The arterial flap is sewn into place.

In the vast majority of cases, however, resection of the coarctation with direct end-to-end anastomosis (Fig. 7) of the aorta is preferred. After sufficient mobilization of the proximal and distal aorta, the proximal aortic clamp is placed at the origin of the left subclavian artery and the distal clamp as described above. In newborns and infants the aorta can be mobilized substantially over a long distance. The aorta is transected proximal and distal to the coarctation, and the ends are approximated. Anastomosis is then performed with continuous 6–0 or 5–0 absorbable monofilament suture in infants and children to allow the anastomosis to grow, and with 3–0 polypropylene in adults because it need not grow. Alternatively, an interrupted technique can be used.

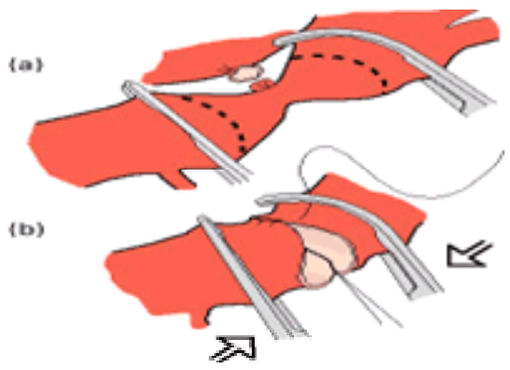


Fig. 7. Resection and end-to-end anastomosis: (a) after sufficient dissection and mobilization of the aorta and division of the patent ductus or ligamentum arteriosum, clamps are placed, and the area of coarctation is resected; (b) the aortic ends are approximated by gentle traction on the clamps, and the anastomosis is performed.

If, despite exhaustive mobilization of the aortic arch and distal aorta, approximation of the divided ends is impossible without tension on the suture line, a woven Dacron tube graft is inserted in adults. A minimum graft diameter of 18 mm is required to avoid a residual pressure gradient. Prosthetic patch aortoplasty is currently controversial because of extensively documented aneurysm formation in the aortic wall opposite the prosthetic patch. The aneurysm probably forms because of excessive removal of the coarctation membrane.

In technically difficult situations, especially in reoperations, a Dacron bypass graft can be used in adults to bridge the coarctation, usually without complete cross-clamping of the aorta.

After completion of the anastomoses, the distal clamp is removed, followed by the slow opening of the proximal clamp to prevent a rapid drop in blood pressure. The mediastinal pleura is closed and a single chest tube inserted for pleural drainage.

Results and complications

Hypertension can be observed in many patients for 2 or 3 days after surgery, sometimes combined with mild abdominal discomfort. If hypertension remains untreated, abdominal tenderness, fever, and leucocytosis can occur; intestinal necrosis due to mesenteric arteritis is occasionally seen. Preoperative treatment with b-blocker drugs and postoperative treatment with b-blockers, labetalol, nitroprusside, and angiotensin-converting enzyme inhibitors have made such complications rare.

Another rare but serious complication is paraplegia, which occurs in about 0.4 per cent of patients after operation. Insufficient collateralization, resulting in reduced perfusion of the anterior spinal artery during the time of aortic clamping, appears to be the cause.

Recurrence of coarctation, defined as a resting peak-pressure gradient greater than 20 mmHg across the repair area, is a late complication; this may be associated with operations performed in early infancy and the use of end-to-end anastomosis. The subclavian flap angioplasty has improved the outcome in young infants, but controversy remains over which is the best technique. There is a low but present incidence of arm dysfunction following standard subclavian flap aortoplasty. Percutaneous balloon angioplasty can be used with good results in selected patients with recurrent coarctation. It is not recommended as a primary technique in patients who have not undergone previous surgical treatment.

The mortality rate during the hospital stay in patients treated for isolated coarctation, with or without associated patent ductus arteriosus, is approx. 1 per cent, and continues to decrease. Almost all late deaths occur within 1 year of operation. Approximately 90 per cent of patients who undergo a coarctation repair after infancy and leave hospital are still alive after 25 years.

Further reading

Bhatt V, Nahata MC. Pharmacologic management of patent ductus arteriosus. *Clinical Pharmacology* 1989; **8**:17–33. [General review of the medical management of patent ductus.]

Burke RP, Wernovsky G, van der Velde M, Hansen D, Castaneda AR. Video-assisted thoracoscopic surgery for congenital heart disease. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 499–507. [A series of 46 patients aged 2 hours to 14 years undergoing video thoracoscopic surgery for congenital heart disease, including 26 patients with patent ductus. Includes follow-up with respect to residual patency.]

Cassady G *et al.* A randomized, controlled trial of very early prophylactic ligation of the ductus arteriosus in babies who weighed 1000 g or less at birth. *New England Journal of Medicine* 1989; **320**: 1511–16. [A series of 84 neonates weighing less than 1000 g with patent ductus, randomized to either ongoing medical therapy or surgical ligation. Primary difference between treatment groups was a decreased incidence of necrotizing enterocolitis in surgically treated neonates.]

Cohen M, Fuster V, Steele PM, Driscoll D, McGoon DC. Coarctation of the aorta. Long-term follow-up and prediction of outcome after surgical correction. *Circulation* 1989; **80**: 840–5. [Twenty-year follow-up in a series of 646 patients undergoing surgical repair of aortic coarctation. Anatomic and physiologic factors identified that correlate with best long-term survival.]

Coster DD, Gorton ME, Grooters RK, Thieman KC, Schneider RF, Soltanzadah H. Surgical closure of the patent ductus arteriosus in the neonatal intensive care unit. *Annals of Thoracic Surgery* 1989; **48**: 386–9. [A series of 115 neonates with patent ductus who underwent surgical ligation in the intensive care unit. There was no surgical mortality and no complications associated with conduct of the procedure in the intensive care unit.]

Daniels CJ, Cassidy SC, Teske DW, Wheller JJ, Allen HD. Reopening after successful coil occlusion for patent ductus arteriosus. *Journal of the American College of Cardiology* 1998; **31**: 444–50. [A small series (20 patients) who underwent catheter-based closure of patient ductus using coil embolization. Patients with residual patency were identified, as well as risk factors for this complication.]

Gersony WM. Coarctation of the aorta. In: Adams FH, Emmanouilides GC, ed. *Heart disease in infants, children, and adolescents*, 3rd edn, pp. 188–215. Williams and Wilkins, Baltimore, 1983. [An excellent chapter detailing all aspects of aortic coarctation. Includes an extensive reference list.]

Gold JP, Cohn LH. Operative management of the calcified ductus arteriosus. *Annals of Thoracic Surgery* 1986; **41**: 567–8. [A surgical technique for management of the calcified patent ductus in older patients.]

Hallman GL, Cooley DA, Gutgesell HP. *Surgical treatment of congenital heart disease*. Lea and Febiger, Philadelphia, 1987. [A general discussion of the traditional surgical aspects of treatment of patent ductus and aortic coarctation.]

Hellenbrand WE, Allen HD, Golinko RJ, Hagler DJ, Lutin W, Kan J. Balloon angioplasty for aortic recoarctation: results of valvuloplasty and angioplasty of congenital anomalies registry. *American Journal of Cardiology* 1990; **65**: 793–7. [A series of 200 patients who underwent balloon angioplasty of aortic recoarctation (age range 1 month to 26 years since primary operation. Results and morbidity/mortality are described.]

Heymann MA. Patent ductus arteriosus. In: Adams FH, Emmanouilides GC, ed. *Heart disease in infants, children and adolescents*, 3rd edn, pp. 158–71. Williams and Wilkins, Baltimore, 1983. [An excellent chapter dealing with detailed aspects of patent ductus arteriosus. Includes an extensive list of references.]

Kirklin JW, Barratt-Boyes BG. *Cardiac surgery*. Churchill Livingstone, New York, 1988. [Excellent chapters on aortic coarctation and patent ductus detailing surgical anatomy, pathophysiology, surgical treatment, and outcome. Extensively referenced with literature of interest to surgeons.]

Kron IL *et al.* Incidence and risk of reintervention after coarctation repair. *Annals of Thoracic Surgery* 1990; **49**: 920–6. [A series of 197 patients from 1967 to 1989 who underwent surgery for aortic coarctation. Risk factors for recoarctation are examined and analyzed.]

Laborde F, Folliguet TA, Etienne PY, Carbognani D, Batisse A, Petrie J. Video-thoracoscopic surgical interruption of patent ductus arteriosus. Routine experience in 332 pediatric cases. *European Journal of Cardio-Thoracic Surgery* 1997; **11**: 1052–5. [A large series of patients who underwent videothoracoscopic ligation of patent ductus. Only short-term follow-up is given, but excellent results are reported.]

Liberthson RR. Congenital heart disease in the child, adolescent, and adult. In: Eagle KA, Haber E, DeSanctis RW, Austen WG, ed. *The practice of cardiology*, 2nd edn, pp. 1157–65. Little, Brown, Boston, 1989. [General discussion of aortic coarctation and patent ductus in older children and adults.]

Metzdorff MT, Cobanoglu A, Grunkemeier GL, Sunderland CO, Starr A. Influence of age at operation on late results with subclavian flap angioplasty. *Journal of Thoracic and Cardiovascular Surgery* 1985; **89**: 235–41. [Authors examined late follow-up following subclavian flat coarctation repair. Risk of recoarctation was greatest when the technique was applied to neonates.]

Mitchell SC, Korones SB, Berendes HW. Congenital heart disease in 56,109 births. Incidence and natural history. *Circulation* 1971; **43**: 323–33. [Extensive natural history study detailing incidence of

congenital heart disease in a large cohort of births.]

Rudolph AM, Heymann MA, Spitznas U. Hemodynamic considerations in the development of narrowing of the aorta. *American Journal of Cardiology* 1972; **30**: 514–25. [Discussion of the influence of hemodynamics on the embryologic development and structural growth of the aorta. The principle can be generalized to other types of congenital heart defects.]

Siblini G, Rao PS, Nouri S, Ferdman B, Jureidini SB, Wilson AD. Long-term follow-up results of balloon angioplasty of postoperative aortic recoarctation. *American Journal of Cardiology*. 1998; **81**: 61–7. [A series of 33 patients who underwent balloon dilation of aortic recoarctation, with 1 to 10 year follow-up available for 32 patients. Excellent long-term results.]

Singh TP, Morrow WR, Walters HL, Vitale NA, Hakimi M. Coil occlusion versus conventional surgical closure of patent ductus arteriosus. *American Journal of Cardiology* 1997; **79**: 1283–5. [A study that examines the cost effectiveness of catheter-delivered coil occlusion versus surgical ligation of patent ductus.]

Sweeney MS, Walker WE, Duncan JM, Hallman GL, Livesay JL, Cooley DA. Reoperation for aortic coarctation: techniques, results, and indications for various approaches. *Annals of Thoracic Surgery* 1985; **40**: 46–9. [A series of 53 surgical patients who underwent surgery for recurrent aortic coarctation using a variety of surgical techniques.]

Waldhausen JA, Nahrwold DL. Repair of coarctation of the aorta with a subclavian flap. *Journal of Thoracic and Cardiovascular Surgery* 1966; **51**: 532–3. [Original description of the subclavian flap technique for repairing aortic coarctation.]

40.7.3 Atrial, ventricular, and atrioventricular septal defects

Ravi Pillai and Kenny Sin

Atrial septal defects
Clinical presentation
Investigations
Morphology
Surgical management
Ventricular septal defects
Clinical presentation
Investigations
Morphology
Surgical management
Early and late outcome
Atrioventricular septal defects
Clinical presentation
Investigations
Morphology
Surgical management
Early and late outcome
Further reading

Atrial septal defects

Atrial septal defects are the most common intracardiac defects encountered in surgical practice. They may occur in isolation or in association with other more complex cardiac anomalies. These defects were not clinically recognized until the 1940s and have been successfully closed since late in that decade. A variety of closed techniques used in the era before bypass have now been superseded by open operations on cardiopulmonary bypass. Surgical closure using hypothermia with inflow occlusion is still practised in developing countries with comparable results. The limitation of non-bypass techniques lies in the difficulties in adapting to unforeseen lesions such as partial anomalous pulmonary venous drainage or very large atrial septal defects. The three main types are secundum atrial septal defects (defects of the fossa ovalis), sinus venosus defects, and coronary sinus defects.

Clinical presentation

Twenty-five per cent of the population have a persistent foramen ovale. These are usually asymptomatic unless the interatrial communication is large. The clinical picture reflects the left to right shunt leading to increased pulmonary blood flow. The increased volume overload increases the right ventricular end-diastolic volume, the long-term consequence being right heart failure. The increased pulmonary blood flow eventually leads to pulmonary hypertension with shunt reversal: right to left shunting leads to cyanosis and overt cardiac failure. In children with significant interatrial communication, recurrent respiratory infections and dyspnoea on exertion may be presenting features. Right-sided cardiac failure is more likely to be seen in older patients, some of whom are in the seventh decade of life. Such symptoms are accompanied by signs of right ventricular enlargement and a pulmonary systolic murmur due to the increased blood flow. A systolic murmur in the tricuspid area is suggestive of gross right ventricular enlargement and dilatation of the ring of the right atrioventricular valve. The first heart sound has a widely split, loud second component. Mild cyanosis in patients with an isolated atrial septal defect is suggestive of a very high or low defect or a coronary sinus-type interatrial septal defect, the right to left shunting in this setting being due to ‘streaming' of the blood flow.

Investigations

Radiology confirms an enlarged right atrium and ventricle. The main pulmonary artery is also enlarged, with evidence of increased blood flow to both lung fields. Right bundle-branch block is an invariable finding on electrocardiography. Echocardiography allows direct visualization of the defect. Increased right ventricular volume with parodoxical septal movement reflects the left to right shunt. In patients with uncomplicated defects, this is the investigation of choice and obviates the need for cardiac catheterization. Cardiac catheterization continues to be of value in infants who may have additional abnormalities, as well as in older patients in whom assessment of pulmonary vascular resistance is important in planning management. A step-up in oxygen saturation at the mid-atrial level is diagnostic of a secundum atrial septal defect. Such step-ups at the level of the vena caval orifices indicate high or low defects. Selective injection into the pulmonary veins will define the drainage pattern.

Morphology

Ostium secundum defect

This is the most common form of atrial septal defect and represents a true deficiency of the floor of the fossa ovalis (Fig. 1). The defect is bounded by the superior, anterior, posterior, and inferior limbus. The limbus itself may be deficient to a variable extent. Absence of the inferior part of the limbus gives rise to a low atrial septal defect, bounded inferiorly by the opening of the inferior vena cava and the eustachian valve. Deficiency of the posterior limbus results in a posterior atrial septal defect.

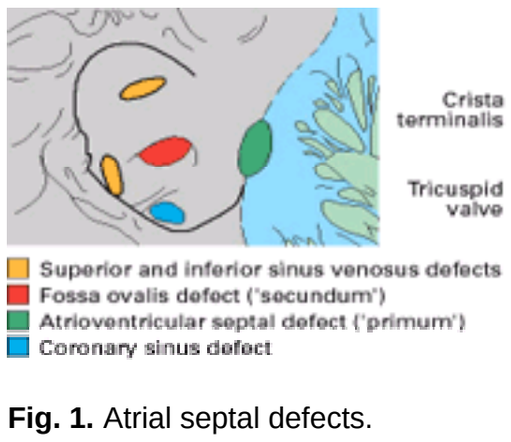


Fig. 1. Atrial septal defects.

Sinus of Valsalva defect

This defect is unrelated to the fossa ovalis and superior to the limbus. It lies at the mouth of the superior vena cava and is almost always associated with abnormal drainage of the right pulmonary veins (commonly the right upper and middle veins).

Coronary sinus defect

These defects are associated with an unroofed coronary sinus into which a left-sided superior vena cava commonly drains.

Surgical management

A significant proportion of secundum defects close during the first year of life. Haemodynamically important defects that persist beyond this period are closed electively at a socially convenient time. A persistent, significant left to right shunt would eventually produce the consequences described above. A pulmonary to systemic (Q_p:Q_s) flow ratio greater than 1.5 is an indication for elective surgery. On this basis, following diagnosis in infancy, planned surgery is normally undertaken between 1 and 2

years of age.

Surgical technique

Closure of the defect is performed on cardiopulmonary bypass established with an aortic cannula and two vena caval cannulas. Under moderate hypothermia (32–34°C) the heart is arrested with cold cardioplegic solution following aortic cross-clamping. Arrest of the heart allows careful inspection of the right atrium and its morphology. The defect itself may be repaired by primary direct suture, or using a pericardial or Dacron patch. The direct suture is satisfactory in patients with adequate fossa ovalis tissue. Larger defects are closed with a patch, avoiding distortion of the atrial chambers. Autologous pericardial patches are used in preference in view of their reduced thrombogenicity. In older patients, associated regurgitation of the tricuspid and mitral valves can be repaired concurrently.

Results

Contemporary techniques have reduced the early mortality after surgical repair to zero. Following repair there is complete relief of symptoms. Arrhythmias associated with the volume overload to the right atrium also improve.

Ventricular septal defects

The first reported repair of a ventricular septal defect was in 1954, using the technique of cross-circulation. Their closure today is performed on cardiopulmonary bypass, with or without deep hypothermia and circulatory arrest. Ventricular septal defects occur in isolation or associated with other more complex anomalies, including tetralogy of Fallot, transposition of the great arteries, tricuspid atresia, and truncus arteriosus. The most common associated lesion remains a persistent arterial duct (6 per cent).

Clinical presentation

Symptomatic presentation reflects the size of the ventricular septal defect and the resulting shunt. Infants may present with a precordial pansystolic murmur only; a large defect may cause gross cardiac failure, with tachypnoea, hepatomegaly, and failure to thrive. A high pulmonary blood flow gives rise to a mitral diastolic flow murmur and a loud pulmonary component of the second heart sound. While 80 per cent of ventricular septal defects presenting at birth may close in infancy, only 25 per cent of those persisting at 1 year do so. The long-term outcome in a persisting, haemodynamically significant, ventricular septal defect is pulmonary hypertension. The rise in pulmonary vascular resistance eventually leads to shunt reversal and cyanosis.

Investigations

Radiography shows cardiomegaly with pulmonary plethora, reflecting the increased pulmonary blood flow and biventricular hypertrophy in patients with a significant shunt. The development of irreversible pulmonary hypertension gives rise to large proximal pulmonary arteries with decreased peripheral lung markings. Associated compression of the bronchioles gives rise to alveolar hyperinflation. Two-dimensional Doppler echocardiography demonstrates clearly the position of the defect and the gradient across it. In addition, the size of the ventricles and any associated abnormalities may be visualized. Angiocardiography demonstrates all four chambers and enables the size of shunt and pulmonary vascular resistance to be calculated. Most importantly, a left ventricular injection reveals the profile of the interventricular septum and demonstrates the presence of additional ventricular septal defects.

Morphology

Ventricular septal defects have been classified in a variety of ways. More recent analysis of the morphology of the ventricular septum allows the defects to be divided into two types, perimembranous and muscular (Fig. 2). In addition, the muscular septum is divided into three components: inlet septum, trabecula septum, and outlet septum. The vast majority of defects (80 per cent) are perimembranous in type, abutting on to the membranous part of the ventricular septum. The relation of these defects to the adjacent muscular septum allows them to be described as being inlet, trabecular, or outlet perimembranous defects.

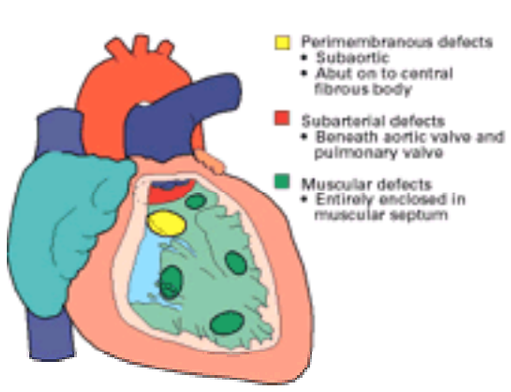


Fig. 2. Ventricular septal defects.

The conduction axis (bundle of His) passes along the left ventricular aspect of the posteroinferior rim of perimembranous defects. The septal leaflet of the tricuspid valve overlies inlet-type defects with varying chordal attachments to the rim. Outlet perimembranous defects are bounded superiorly by a varying amount of the outlet septum. Almost total absence of the outlet septum produces a subarterial defect, with only a fibrous rim separating the aortic and pulmonary valves. Excessive override of one or other of the great vessels leads to a double-outlet left or right ventricle. Such a defect may also be associated with aortic regurgitation secondary to prolapse of the right (most commonly) coronary cusp. In patients with long-standing defects the sinus itself prolapses, giving rise to an aneurysm of the sinus of Valsalva.

Muscular defects vary in size and are usually located centrally or anteriorly; they may be single or multiple. Anteriorly and centrally placed, multiple ventricular septal defects give rise to a ‘Swiss cheese' appearance that is difficult to manage surgically (Fig. 2).

Surgical management

Primary closure of the defect at presentation is the general rule in symptomatic patients. In neonates and in patients with the Swiss cheese defect, banding of the pulmonary artery will reduce pulmonary blood flow and allow growth; repair is undertaken at about 5 years of age in the latter group. Closure of a single ventricular septal defect is undertaken within 6 months of banding. Persistent defects in asymptomatic patients more than 12 months of age are electively closed within 2 years to prevent the development of pulmonary hypertension in the long term. Successful closure at this age may be considered curative. Closure is no longer an option once significant pulmonary hypertension develops with or without shunt reversal: such patients require heart–lung transplantation, though closure with a bilateral lung transplant may become possible.

Surgical technique

Closure of the defects is undertaken on cardiopulmonary bypass, with or without deep hypothermia and circulatory arrest. The heart is arrested by instilling cardioplegic solution into the aortic root following aortic cross-clamping. The two most common approaches to the defects are through the right atrium and via the tricuspid valve or through a transverse or longitudinal right ventriculotomy. The ventriculotomy approach is particularly useful in outlet-type defects. A left ventricular or transpulmonary approach is used less often. The left ventricular approach provides access to anteriorly placed, muscular ventricular septal defects while the transpulmonary provides good exposure for subarterial defects.

Perimembranous defects are all closed using a Dacron patch, placed in position using either interrupted or continuous sutures. Posteroinferiorly the placement of sutures through the base of the septal leaflet of the tricuspid valve avoids damage to the underlying bundle of His. Circular muscular defects also require patch closure, while oval or slit-like defects may be closed directly by interrupted sutures reinforced with Teflon or pericardial pledgets. Multiple defects are closed separately or with a large patch, as with the Swiss cheese defects.

Early and late outcome

Early mortality for repair of isolated, uncomplicated ventricular septal defects approaches zero in centres familiar with the operative and perioperative care required by patients with such defects. Incremental risk factors include young age (under 3 months), multiple defects, particularly the Swiss cheese type, and established pulmonary hypertension. Immediate postoperative complications are related to transient atrioventricular dissociation, complete heart block, low-output cardiac failure, or residual shunts. A degree of shunting across Dacron patches is to be expected for up to 48 hours. Significant shunts are investigated early and may require reoperation.

Long-term outcome is excellent in patients who undergo repair before any irreversible pulmonary vascular changes. Complete heart block is a rare occurrence; however, right bundle-branch block occurs in 30 to 40 per cent of patients, irrespective of the approach used for the repair.

Atrioventricular septal defects

Atrioventricular septal defects have in the past been called partial or complete atrioventricular canal defect, endocardial cushion defect, or ostium primum defects. The realization that there is no deficiency of the interatrial septum has now led to the universal acceptance of the term ‘atrioventricular septal defect'. The other terms are more easily understood if described as atrioventricular septal defects with separate or common atrioventricular orifice.

The first successful repair of the separate-orifice defect was reported in 1955. Early repairs carried out on cardiopulmonary bypass were associated with considerable morbidity and mortality. More recently, the improved understanding of the morphology and function of the atrioventricular valve and the disposition of the conduction tissue has decreased the risk incurred during surgery.

Clinical presentation

Clinical presentation reflects the spectrum of the defect. In the presence of an atrial communication only with a competent left atrioventricular valve, the signs and symptoms mimic those of a secundum atrial septal defect, and patients present in the first decade of life or even later. At the other end of the spectrum; patients with a common atrioventricular valve orifice and shunting at the atrial and ventricular levels develop signs and symptoms soon after birth, when there is gross cardiac failure associated with tachypnoea, hepatomegaly, poor peripheral perfusion, and failure to thrive. These children develop significant pulmonary hypertension within 24 months and rarely survive beyond this age. The cardiac impulse is prominent and an apical systolic murmur signifies regurgitation of the left atrioventricular valve. Left to right shunting and atrial overload produces a mid-diastolic flow murmur. The second heart sound is loud and widely split, with an accentuated pulmonary component. Mild cyanosis may be present when a ‘primum' type defect is associated with a large secundum defect, giving rise to a common atrium. More significant cyanosis, present in 20 per cent of patients, is indicative of more complex associated abnormalities such as tetralogy of Fallot, double-outlet right ventricle, transposition of the great arteries, or a completely unroofed coronary sinus. A persistent arterial duct is an associated anomaly in about 10 per cent of patients with atrioventricular septal defects.

Investigations

Chest radiographs invariably demonstrate cardiomegaly with pulmonary plethora. Due to the gross left to right shunting the main pulmonary artery and the vessels at the hila are prominent. Loss of peripheral lung markings indicates established pulmonary vascular disease and hypertension. The electrocardiogram shows left-axis deviation and biventricular hypertrophy. Atrial fibrillation occurs late and may point to poor long-term prognosis. Two-dimensional and Doppler echocardiography adequately define the anatomy associated with these defects. When a clear diagnosis of a defect associated with a double orifice and no interventricular communication is established, repair is undertaken on this evidence. However, in the presence of a common atrioventricular valve with an interventricular communication and atrioventricular valve regurgitation, cardiac catheterization is performed. Angiocardiography demonstrates the extent of the shunts and allows calculation of pulmonary vascular resistance. Other abnormalities may also be excluded. Typically, a left ventricular injection of contrast shows regurgitation of the left atrioventricular valve directly into the right atrium: this is significant in 40 to 60 per cent of patients. Anterior displacement of the aorta and left ventricular outflow tract together with the discrepancy in length between the inlet and outlet portions of the ventricular septum gives rise to a ‘goose-neck' deformity on ventriculography.

Morphology

Atrioventricular septal defects are characterized by absence of the membranous and muscular part of the atrioventricular septum (Fig. 3). The defect is associated with a resultant anomaly of the atrioventricular valve, together with displacement of the aorta and the course of the conduction axis.

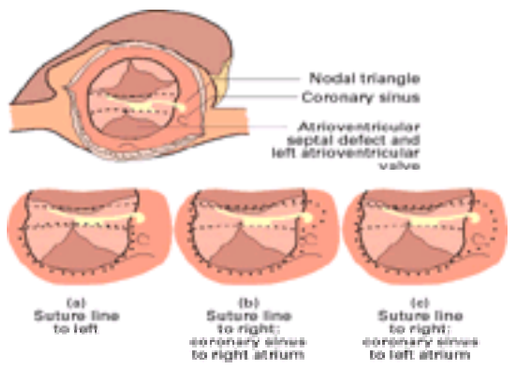


Fig. 3. Conduction tissue disposition and choice of placement of the interatrial baffle in atrioventricular septal defects (by courtesy of Drs S.Y. Ho and R.H. Anderson).

The aortic root normally lies wedged between the mitral and tricuspid valves. In patients with atrioventricular septal defect the aorta is displaced anteriorly, producing an elongated and narrowed left ventricular outflow tract. The atrioventricular valve is a five-leaflet structure. The crest of the interventricular septum is straddled by an anterior and posterior bridging leaflet: the so-called cleft in the left atrioventricular valve is merely the space between the left components of the anterior and posterior bridging leaflets. In defects associated with separate orifices, the bridging leaflets are joined by a connecting tongue of tissue, giving rise to two orifices. Further attachments by valve tissue to the crest of the interventricular septum prevent any interventricular communication, producing the ‘primum' defect. The extent of bridging of the anterior leaflet is variable, and formed the basis of the Rastelli classification of the common-orifice defect: type A defects are those with minimum bridging, and type C those where extensive bridging to the anterior leaflet occurs, extending across into the right side and being attached to the medial papillary muscle together with the anterosuperior leaflet of the right atrioventricular valve.

Competence of the atrioventricular valve is maintained in many patients. The acceptance of the left atrioventricular valve as a trileaflet structure has modified the surgical approach to repair of the defect.

The atrioventricular node normally lies in the triangle of Koch, which is bounded by the tendon of Todaro, the coronary sinus, and the tricuspid valve. In patients with atrioventricular septal defect, this node is displaced posteriorly and inferiorly. It lies in a new nodal triangle, bounded by the annulus of the right component of the bridging leaflet, the coronary sinus, and the lower edge of the interatrial septum. The common trunk passes deep to the posterior bridging leaflet and along the crest of the interventricular septum on the left side (Fig. 3). The interventricular septum, though complete, has a ‘scooped out' appearance (Fig. 4). In addition there is significant discrepancy between the inlet and outlet portions, the former being shorter than the latter. This discrepancy on the surface results in a shorter diaphragmatic surface compared to the left cardiac silhouette. The interventricular septum may be deviated into one or other ventricle, thereby reducing ventricular volume, and this has a significant impact on surgical outcome.

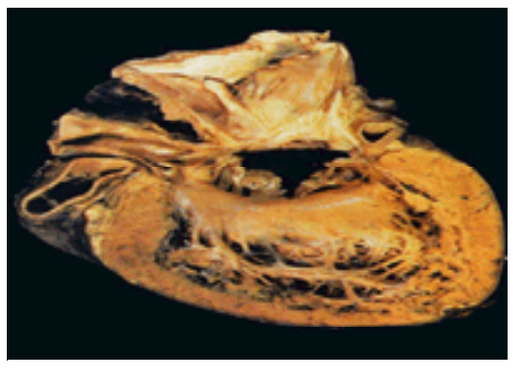


Fig. 4. Appearance of the interventricular septum in atrioventricular septal defects (by courtesy of Drs S.Y. Ho and R.H. Anderson).

Surgical management

Patients with an interatrial communication only and a competent left atrioventricular valve may present in the first decade of life or later. Closure is undertaken electively to avoid the long-term complications of increased pulmonary blood flow and right-sided overload. Patients with a common atrioventricular valve (single orifice) present soon after birth or within 12 months. In very small neonates, palliative procedures such as pulmonary arterial banding may be undertaken to reduce pulmonary blood flow and allow growth. With modern techniques, total repair is undertaken at presentation in more patients. This approach avoids the not insignificant morbidity and mortality associated with banding.

Surgical technique

Repair is undertaken on cardiopulmonary bypass with moderate hypothermia or with deep hypothermia and circulatory arrest. The heart is arrested with cold cardioplegia instilled into the aortic root following cross-clamping. In both types of atrioventricular septal defect the repair of the interatrial and ventricular communication, the preservation or restoration of atrioventricular valve function, and the avoidance of damage to the conduction tissue are essential to success.

Early repair of the single orifice (common atrioventricular valve) was previously undertaken using two separate patches to close the ventricular and atrial defects. While early results showed significant mortality, repair using a one-patch technique improved early outcome, although the technique requires careful resuspension of the left atrioventricular valve to avoid stenosis. With the two-patch technique the interventricular communication is closed with a Dacron patch and the atrial defect with a pericardial patch. A single Dacron patch is used for the single-patch method. It is tempting to close the ‘cleft' in the atrioventricular valve completely in an attempt to create a normal appearance of the mitral valve, but the morphology of the left atrioventricular valve is such that this will cause narrowing of the orifice and postoperative regurgitation due to distortion of the geometry of the valve apparatus. In a large proportion of patients the valve is competent and therefore no attempt should be made to ‘reconstruct' it. Competence is assessed by filling the left ventricular cavity with cold saline and studying the ‘lie' of the leaflets, as in diastole. This may localize small areas of regurgitation that require repair. In a significant number of patients there is no regurgitation and nothing further need be done.

An understanding of the disposition of the atrioventricular conduction axis is crucial in determining the site of attachment of the inferior edge of the interatrial patch. The safest technique is to suture the lower edge of the baffle along the right component of the inferior bridging leaflet, thereby keeping to the right of the non-branching bundle. Subsequently, the suture line passes lateral to the coronary sinus before running along the lower edge of the atrial septum. This leaves the coronary sinus in the left atrium, but the resulting shunt is of no haemodynamic significance unless there is a left superior vena cava draining into the coronary sinus. However, the associated orifice of the sinus is much larger and sutures may be safely placed within it. A third option is to place the inferior suture line into the left ventricular component of the bridging leaflet along the septal crest. Care is taken to place sutures through leaflet tissue only. It is possible to place all sutures to the left and avoid injury to the conduction axis (Fig. 3).

Any associated anomalies are corrected at the same time. The improved understanding of anatomy has all but eliminated the incidence of complete heart block. Measurement of left atrial pressure with the heart beating allows assessment of the repair and function of the left atrioventricular valve.

Early and late outcome

Reports on the early mortality in recent years vary from 0 to 30 per cent. Absence of an interventricular communication greatly improves outcome, with the patients tending to present later. Risk factors include small age and size, anatomy of the left atrioventricular valve and extent of regurgitation, presence of other associated anomalies, and the preoperative symptomatic status.

Following successful repair, the procedure may be regarded as ‘curative'. A small number of patients may require replacement of the left atrioventricular valve. Residual ventricular septal defects present early and invariably require reoperation and closure.

Further reading

Barrett-Boyes BG, Neilze JM, Clarkson PM, Shadey GC, Brandt PWT. Repair of ventricular septal defect in the first two years of life using profound hypothermia—circulatory arrest techniques. *Annals of Surgery* 1976; **184**: 376. [Review of 57 infants who underwent ventricular septal defect repair, with 2/8 and 2/49 deaths in infants with and without an associated coarctation, respectively. Psychometric testing in 19 children at 3 to 4 years of age showed no evidence of cerebral damage.]

Becker AE, Anderson RH. *Cardiac pathology.* Churchill Livingstone, London, 1982.

Becker AE, Anderson RH. Atrioventricular septal defects: What's in a name? *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 461. [An anatomical review.]

Bedford DE, Sellors TH, Sommerville W, Belcher JR, Besterman EMM. Atrial septal defect and its surgical treatment. *Lancet* 1957; **i**: 1255. [A comprehensive review including a series of 40 patients with 2.5 per cent mortality.]

Collins G, Calder L, Rose V, Kidd L, Keith J. Ventricular septal defect: clinical and haemodynamic changes in the first five years of life. *American Heart Journal* 1972; **84**: 695. [Prospective study of 200 consecutive infants with isolated ventricular septal defect showing that when flow ratios are greater than 2, mortality was 7 per cent, palliative surgery was required in 9 per cent, spontaneous closure occurred in 6 per cent, and 10 per cent developed Eisenmenger reaction.]

Frater RW. Persistent common atrioventricular canal: anatomy and function in relation to surgical repair. *Circulation* 1965; **32**: 120. [A review of the anatomy of the mitral valve in endocardial cushion defect including a report of a series of nine patients who underwent repair, with one death, and good echocardiographic result in the remainder at 1-year follow up.]

Kirklin JW, Barratt-Boyes BG. *Cardiac surgery*. Wiley, Chichester, 1986.

Kirklin JW, Blackstone EH, Bargeron LM, Pacifico AD, Kirklin JK. The repair of atrioventricular septal defects in infancy. *International Journal of Cardiology* 1986; **13**: 333. [A retrospective analysis of surgical results comparing the early (1967–1982) with the current (1984–1985) experience. The results and trend suggests that primary repair in the first year of life is advisable for most patients with atrioventricular septal defects. Predicted 10-year survival for patients operated upon was 99 per cent and 95 per cent for partial and complete defects, respectively.]

Lillehei CW, Cohen M, Warden HE, Ziegler NR, Varce RL. The results of direct viscoid closure of ventricular septal defects in eight patients by means of controlled cross circulation. *Surgery, Gynecology, and Obstetrics* 1955; **101**: 446. [A landmark article on intracardiac repair in children utilizing controlled cross-circulation.]

Lillehei CW, Cohen M, Warden HE. Direct vision intracardiac correction of congenital anomalies by controlled cross circulation: results in 32 patients with ventricular septal defects, tetralogy of Fallot and atrioventricular commissures defects. *Surgery* 1955; **38**: 11. [An update of the earlier series reporting 12/32 deaths with no donor mortality.]

Lewis FJ, Tanfic M. Closure of atrial septal defects with the aid of hypothermia: experimental accomplishments and the report of the one successful case. *Surgery* 1953; **33**: 52. [A report on the development of intracardiac repair using hypothermia and inflow occlusion in the canine model culminating in the successful repair of an atrial septal defect in a child.]

Murray G. Closure of defects in cardiac septa. *Annals of Surgery* 1948; **128**: 843. [A review of the anatomy and report of four patients who underwent external suture of cardiac septal defects.]

Pillai R, Ho SY, Anderson RH, Shinebourne EA, Lincoln JCR. Malalignment of the interventricular septum with atrio-ventricular septal defect—its implications concerning conduction tissue disposition. *Thoracic and Cardiovascular Surgery* 1984; **32**: 1–3. [An anatomical review with a report of an anomalous siting of the atrioventricular node associated with a malaligned interventricular septum in atrio-ventricular septal defect.]

Pillai R, Ho SY, Anderson RH, Lincoln C. Ostium primum atrioventricular septal defect: an anatomical and surgical review. *Annals of Thoracic Surgery* 1986; **41**: 458. [A review of anatomy and 84 patients who underwent surgical correction with an early and late mortality of 2.4 per cent.]

Rastelli GC, Weidman WH, Kirklin JW. Surgical repair of the partial form of persistent common atrioventricular canal with special reference to the problem of mitral valve incompetence. *Circulation* 1965; **31–32** (Suppl. 1): 31. [A retrospective review of 131 patients who underwent repair of a partial atrioventricular canal, with a hospital mortality of 3.8 per cent. No heart block occurred in the last 93 patients and results have been good in general, although there was evidence of residual mitral regurgitation in many patients.]

Rizzoli G *et al.* Progress in the surgical treatment of ventricular septal defect: an analysis of a twelve years experience. *Thoracic and Cardiovascular Surgery* 1983; **31**: 382. [A review of 267 patients who had ventricular septal defect (VSD) repair. Low weight, early operative date, multiplicity of defects, and presence of major associated lesions had independent significant incremental effect on hospital mortality. The risk of operation for infants with multiple VSDs approaches that of a single VSD in patients over 12 kg of weight if concomitant lesions do not complicate surgical treatment.]

40.7.4 Malalignment anomalies of the interventricular septum

Andrew Parry, David Hamilton and Pankaj Mankad

Anatomic considerations
Congenital cardiac defects associated with malalignment of the interventricular septum
Malalignment of the atrioventricular septum
Malalignment of the outlet septum
Malalignment of the muscular septum
Tetralogy of Fallot
Clinical features
Complications and natural history
Physical examination
Investigations
Indications and timing of operation
Principles and techniques of operations
Repair of pulmonary atresia with ventricular septal defect
Repair of tetralogy of Fallot with anomalous origin of the left anterior descending coronary artery from right coronary artery
Postoperative care and complications
Results
Univentricular heart
Clinical features and diagnosis
Natural history
Surgical management
Results
Further reading

Anatomic considerations

The interventricular septum contains membranous and muscular components. The membranous portion is a relatively small fibrous area which extends from the inferior limit of the interatrial septum to the superior extent of the muscular interventricular septum. The mitral and the tricuspid valves are attached to this part of the septum. The plane of mitral valve attachment is at a higher level than the plane of the tricuspid valve (Fig. 1). Thus, a small portion of the membranous septum lies below the muscular portion of the interatrial septum to the left of the tricuspid annulus, and this separates the right atrium from the left ventricle. Absence of this portion of the septum results in an intracardiac shunt from the left ventricle to the right atrium (Gerbode defect). The muscular part of the interventricular septum is composed of the inlet (behind the septal leaflet of the tricuspid valve), and the trabecular and outlet portions (Fig. 2). The outlet septum, also called the infundibular septum, represents septation of the bulbus cordis into right and left ventricular outflow tracts. As a result of the difference in configuration of the two ventricles, the various components of the interventricular septum normally subtain considerable angles to each other. The inlet septum is nearly at right angles to the outlet septum and is at about 45° to the trabecular septum.

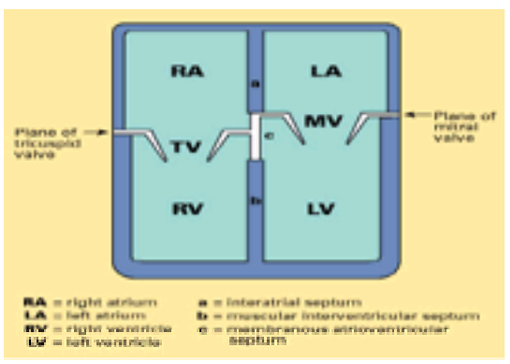


Fig. 1. Schematic diagram of septation of the heart. Absence of the inferior portion of ‘c' results in an interventricular septal defect (perimembranous ventricular septal defect). Absence of the superior portion of ‘c' results in a left ventricular–right atrial defect (Gerbode). This occurs because the left atrioventricular valve is on a superior plane to the right atrioventricular valve. Absence of the whole of ‘c' is associated with a total atrioventricular canal defect.

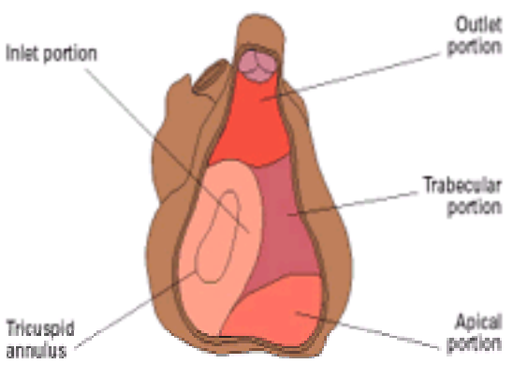


Fig. 2. Diagrammatic representation of the components of muscular interventricular septum viewed from the right ventricular side.

Congenital cardiac defects associated with malalignment of the interventricular septum

The normal alignment of the interventricular septum is often disturbed by the developmental anomalies which result in various congenital cardiac malformations. Malalignment anomalies can be divided according to the region of septum involved and there is a spectrum in the degree of malalignment seen for each.

Malalignment of the atrioventricular septum

This results in a perimembranous inlet ventricular septal defect, sometimes associated with straddling of the tricuspid valve (i.e. the tension apparatus of the valve is attached to both sides of the septum or to the septal crest) and/or overriding where the valve annulus is related to both ventricular chambers.

Malalignment of the outlet septum

Anterosuperior deviation of the outlet septum in relation to the rest of the septum results in a typical subarterial ventricular septal defect. The outlet septum may be anteriorly or posteriorly malaligned. In either case a subarterial ventricular septal defect occurs. Such a defect is commonly encountered in tetralogy of Fallot, double-outlet right ventricle, and transposition of the great arteries. With anterior malalignment the outlet septum impinges on the right ventricular outflow tract producing some degree of pulmonary or infundibular stenosis or, in its most severe form, pulmonary atresia. If the septum is posteriorly malaligned the ventricular septal defect is associated with left ventricular outflow tract obstruction and aortic stenosis. This is commonly seen with interrupted aortic arch. Tetralogy of Fallot, the

most common cyanotic heart defect, is described in detail below.

Malalignment of the muscular septum

The spectrum of malalignment of the muscular septum seen is very wide. At one extreme it may give rise to univentricular atrioventricular connection. Of these, double-inlet left ventricle is the most common example: the main ventricular chamber is of left ventricular morphology and the rudimentary ventricular chamber has right ventricular morphology. This is discussed briefly below.

Tetralogy of Fallot

Tetralogy of Fallot is caused by a single embryologic maldevelopment: anterosuperior deviation of the outlet septum (Fig. 2). This of necessity gives rise to the three secondary abnormalities: a large malalignment ventricular septal defect (usually of similar size to the aortic valve annulus), right ventricular outflow tract obstruction (exacerbated by the presence of abnormal hypertrophied muscle bundles which act like calipers varying the caliber of the outflow tract, see Fig. 6), and a dextraposed and overriding aorta. The tertiary abnormality, right ventricular hypertrophy, develops due to outflow obstruction. Although the malalignment is primarily responsible for the right ventricular outflow tract obstruction, other mechanisms such as anomalous right ventricular muscle bundles and accessory tissue derived from either the membranous ventricular septum or from the septal leaflet of the tricuspid valve may also contribute. Developmentally, tetralogy of Fallot results from the maldevelopment and unequal partitioning of the distal bulbus cordis. Lack of normal rotation of the distal bulbus creates a situation where the aortic segment is not completely transferred towards the left ventricle and comes to overlie the right ventricular outflow tract (Fig. 3). The condition can therefore be considered as part of a spectrum of congenital cardiac malformations characterized by transfer of the aorta from a position above the left ventricle, overriding a ventricular septal defect, towards a position directly above the anatomic right ventricle.

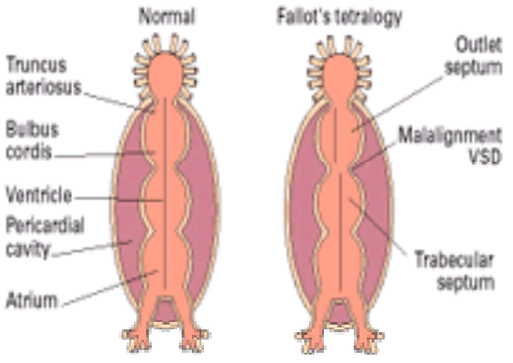


Fig. 3. Schematic representation of normal and abnormal septation of the endocardial heart tube as encountered in tetralogy of Fallot. Note the malalignment of the outlet septum in relation to the trabecular septum resulting in ventricular septal defect and overriding aorta.

A variety of patterns of infundibular stenosis can be recognized ranging from a well-formed infundibular chamber with a low-lying infundibular stenosis to absence of the infundibular septum resulting in a subpulmonary ventricular septal defect. Pulmonary flow may also be obstructed by pulmonary valve stenosis (60 to 70 per cent of patients) or atresia, supravalvular narrowing, stenosis of the pulmonary artery bifurcation, localized stenosis or atresia of either pulmonaryartery (usually the left), or incomplete distribution or hypoplasia of branch pulmonary arteries within the lung. When associated with pulmonary valvar atresia the normal impetus for pulmonary artery development is absent and in these cases the main and branch pulmonary arteries may be completely absent. When this occurs blood flow to the lungs may be provided completely by the persistent ductus arteriosus or by developmentally primitive segmental vessels arising directly from the aorta or other systemic arteries. The infinitely variable spectrum of right ventricular outflow obstruction must be assessed thoroughly preoperatively. Corrective surgery should relieve all possible sites of obstruction.

Clinical features

The classic clinical presentation of tetralogy of Fallot is with cyanosis due to right to left intracardiac shunting of blood across the large ventricular septal defect into the overriding aorta. The severity and timing of onset of cyanosis varies with the degree of obstruction to the pulmonary blood flow. Infants with severe infundibular and valvar stenosis/atresia, or with valvar stenosis and diffuse right ventricular outflow tract hypoplasia are often deeply cyanosed from birth. They rarely suffer from ‘cyanotic spells' but the cyanosis increases gradually. By contrast, infants with dominant infundibular stenosis have a delayed onset of cyanosis: ‘cyanotic spells' due to infundibular spasm may not occur until the second year of life. Such spells become less frequent with age, presumably because the infundibular stenosis becomes ‘fixed' due to the secondary development of endocardial fibrosis. As the child begins to walk, cyanosis is almost always accompanied by dyspnea and squatting is often adopted which may be nature's attempt to improve pulmonary flow by increasing systemic arterial resistance. A few children and adults with tetralogy are acyanotic at rest and only mildly cyanotic on exercise because the pulmonary valve is normal and infundibular stenosis mild.

Unusual clinical presentations are occasionally encountered. Patients in whom the intracardiac shunt is predominantly from left to right may remain acyanotic for many years, presenting within the first and second decades of life as the infundibular stenosis slowly increases in severity. At the other extreme, infants born with pulmonary valve atresia are ductus arteriosus dependent and are cyanosed from the first day of life. In a small subset of patients who may require surgical relief in infancy, presentation is initially that of a large ventricular septal defect, resulting in pulmonary plethora and congestive cardiac failure at 2 to 3 months of age. Cyanosis at rest develops gradually at about 6 months of age.

Most infants with tetralogy of Fallot have a patent foramen ovale. Associated cardiac anomalies such as atrial septal defect, persistence of a left superior vena cava, aberrant origin of right subclavian artery, persistence of the ductus arteriosus, complete atrioventricular canal (often associated with Down's syndrome), and multiple ventricular septal defects, though rare, may also occur.

Complications and natural history

Cerebrovascular accident as a result of cerebral thrombosis may occur at any age in patients with severe cyanosis and polycythemia. This problem is compounded by relative anemia due to iron deficiency because of increased blood viscosity. Dehydration may be a precipitating factor. Hemiplegia may also follow paradoxical embolism or a cerebral abscess. In older patients, rupture of bronchial collateral vessels may lead to massive hemoptysis. Without surgical intervention, cyanosis and polycythemia increase. This is due in part to the ‘fixed' orifice pulmonary valve which becomes relatively more stenotic as the child grows, but also because of a tendency to develop thrombosis of the pulmonary arteries with progressive reduction in pulmonary blood flow. The few patients who survive into the fourth or fifth decades of life die from chronic congestive heart failure due to secondary cardiomyopathy resulting from right ventricular pressure overload and chronic hypoxia in association with polycythemia.

Physical examination

Cyanosis and clubbing are often obvious in the older child with tetralogy of Fallot. The jugular venous pressure is normal and the radial pulse may be of low volume. A moderate intensity, mid-systolic pulmonary ejection murmur is audible, and maximal in the second and third intercostal spaces. This becomes less prominent when stenosis is very severe and may disappear completely during a cyanotic spell as a result of infundibular shut-down. A continuous murmur from a previously created systemic to pulmonary artery shunt may be audible over the precordium.

Investigations

Laboratory tests

Older infants and children usually have an elevated red blood cell count and hemoglobin level, the degree of which correlates with the degree of arterial desaturation and thus with the severity of outflow tract stenosis. The hemoglobin level may be falsely low if iron deficiency is present, in which case the hematocrit is a more accurate reflection of the severity than the hemoglobin.

Radiographic findings

Chest radiographs in affected children may show a typical ‘boot-shaped heart' with oligemic lung fields and demonstrate the position of the aortic arch, which is right sided in approximately 25 per cent of cases. Patients in the second or third decades of life may show progressive kyphoscoliosis.

Electrocardiogram

The electrocardiogram reveals right ventricular hypertrophy and absence of Q-waves with low-voltage R-waves in the left precordial leads. In the right chest leads the R-wave is dominant in infants, usually with an associated upright T-wave.

Echocardiogram

The echocardiogram clearly demonstrates the ventricular septal defect, aortic override, and narrowing of the right ventricular outflow tract. It also delineates any associated intracardiac defects, the course of the coronary arteries, and is usually adequate alone in cases of pulmonary stenosis. When pulmonary atresia is present however, definitive preoperative evaluation is dependent upon cardiac catheterization and high-quality angiocardiography to delineate the presence and anatomy of collateral vessels.

Cardiac catheterization and angiocardiography

This study historically has been necessary to demonstrate all morphologic features of the malformation. With improvements in echocardiography the role of angiocardiography in standard tetralogy of Fallot is less significant, but when pulmonary atresia is present it is essential to characterize any major aortopulmonary collateral arteries. Fig. 4 shows some of the characteristic features of tetralogy of Fallot as seen on cineangiogram. During the procedure it is important to measure distal pulmonary artery pressures, especially in collateral arteries, to ensure the absence of pulmonary vascular disease within them as these vessels are often unprotected.

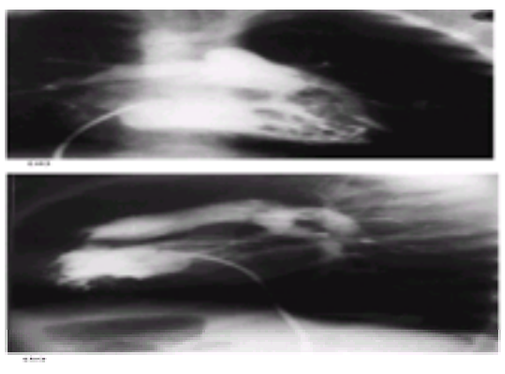


Fig. 4. (a) Cineangiogram of right ventricle in tetralogy of Fallot showing outflow tract obstruction due to hypertrophied muscle bundles (anteroposterior view). (b) Cineangiogram of right ventricle in tetralogy of Fallot showing outflow tract obstruction due to hypertrophied muscle bundles (lateral view).

Indications and timing of operation

The diagnosis of tetralogy of Fallot is an indication for operation in view of its poor prognosis without surgery and the very good early and late results following total correction. However, the morphologic and clinical variations are so numerous that it is important to assess each patient individually with regard to the timing and the type of surgery to be undertaken. Patients with uncomplicated anatomy with pulmonary valve stenosis and with adequately developed intrapulmonary arteries should undergo total intracardiac correction as a primary procedure or following an initial palliative systemic–pulmonary shunt. Although full repair should ideally be performed before the age of 3 to 4 years, older age is not a contraindication to operation. Some centers advocate primary total correction in infancy with excellent results. Part of the rationale for early intervention is based on evidence that development of the distal vascular bed of the lung is dependent on adequate pulmonary blood flow. In addition there is increasing evidence that correcting the right ventricular outflow tract obstruction allows normal development of the pulmonary valve and decreases the requirement for transannular patching. However, the ultimate decision on whether to perform one-stage or two-stage correction depends upon the characteristics of the right ventricular outflow tract obstruction, the experience of the surgeon and his team, and the available postoperative care. Regular audit and analysis of surgical results is essential in making these decisions.

Neonates who are duct dependent with pulmonary atresia require either total correction or a systemic–pulmonary shunt in the early neonatal period and are maintained on a prostaglandin infusion until this can be performed. Children with pulmonary atresia and major aortopulmonary collateral arteries should require no neonatal intervention, but early surgery to maximize the effective recruitment of pulmonary vascular segments is recommended. If complete one-stage unifocalization is selected as the appropriate surgical approach, it is optimal for the child to be 3 to 6 months of age.

Infants and children with severe cyanotic spells may be treated with propranolol (2 to 4 mg/kg.day) while awaiting surgery. This minimizes or abolishes infundibular muscular spasm in many cases.

Principles and techniques of operations

Palliative operations

The objectives of palliative surgery are to increase pulmonary blood flow and to allow hypoplastic pulmonary arteries to grow in caliber.

Systemic to pulmonary artery shunts

Shunts may be created either centrally between the ascending aorta and the main pulmonary artery (Davidson or Waterston anastomosis, Fig. 5(a)) or peripherally between the subclavian artery and right or left pulmonary artery by direct anastomosis (classic Blalock–Taussig shunt, Fig. 5(b)) or using a tube of polytetrafluoroethylene (**PTFE**) (modified Blalock–Taussig shunt, Fig. 5(c)).

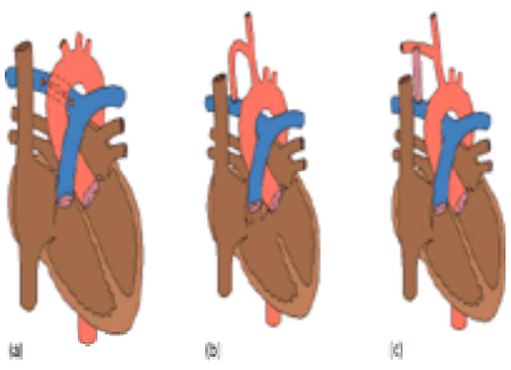


Fig. 5. Diagram showing various types of systemic–pulmonary artery shunts. (a) Waterston shunt: direct anastomosis between the ascending aorta and right pulmonary artery. (b) Classic Blalock–Taussig shunt, that is, anastomosis of the divided subclavian artery to the right pulmonary artery. (c) Modified Blalock–Taussig shunt, that is, interposition of PTFE between the subclavian and right pulmonary arteries.

The Blalock–Taussig or modified Blalock–Taussig shunt continues to be the shunt of choice, even in infancy, as it inherently controls the amount of pulmonary blood flow to some extent. The classic form should be performed on the side opposite to the aortic arch so that the subclavian artery originating from the innominate artery is used. This minimizes the risk of kinking of the subclavian artery at its origin. The Waterston anastomosis provides satisfactory symptom relief in small infants, but has lost favor because of the relatively high incidence of distortion of the right pulmonary artery and the risk of creating too large a shunt, resulting in cardiac failure, and later, pulmonary vascular hypertension. Pott's shunt (between the descending aorta and the left pulmonary artery) has been almost completely abandoned for the same reasons and because of difficulties encountered closing the anastomosis at the time of total repair. Since the Blalock–Taussig shunt is the most common type of shunt performed in infancy, it will be described in detail.

Blalock–Taussig shunt For a classic Blalock–Taussig shunt a thoracotomy is performed on the appropriate side and the lung retracted posteriorly. The main pulmonary artery is identified and an adequate length dissected taking care not to injure the phrenic nerve. Two snares of heavy silk are passed round the upper and lower lobe pulmonary arteries which, when tightened, prevent retrograde bleeding from the arteriotomy. The mediastinal pleura is incised from the pulmonary artery up to and along the subclavian artery. On the right, a groove is then created by blunt dissection between the trachea and the superior vena cava, safeguarding the vagus nerve and its recurrent laryngeal branch. The dissection is continued to the subclavian artery. The most accessible portion of the vessel is dissected and a tape passed around it. The subclavian artery is then tied distally close to the first rib and divided just proximal to the ligature while controlling the proximal end in a vascular clamp. On the right side, the artery is then drawn through the loop of the recurrent laryngeal nerve using a right-angled forceps passed beneath the vagus and the recurrent nerve. The subclavian artery is turned down and its distal end trimmed and cleaned. Extra length can be obtained by adequate mobilization of the innominate and carotid arteries and, if necessary, by dividing the inferior pulmonary ligament.

Proximal and distal control of the pulmonary artery is achieved and the artery opened transversely on its superior border. End-to-side anastomosis of the subclavian artery to the pulmonary artery is performed using 6/0 or 7/0 Prolene. In small infants, heparin (1 mg/kg) is often administered before the subclavian artery is divided. Once the anastomosis is completed, the snares are released and swabs are used to apply gentle pressure to the suture line. In small infants it is important to maintain a systolic blood pressure of over 70 to 80 mmHg, if necessary by using inotropic support. Hypotension in the early postoperative period predisposes to clotting of the anastomosis. In case of excessive oozing, heparin is reversed with protamine sulfate.

A modified Blalock–Taussig shunt may be performed through either a thoracotomy or a median sternotomy. The advantage of the latter approach is that a large segment of the pulmonary artery can be safely dissected out and controlled, allowing the anastomosis to be performed well away from the bifurcation into upper and lower branches. Stricture formation at this bifurcation is a well recognized complication of palliative shunts and may be irretrievable. A suitable length of appropriate diameter PTFE graft (anything from 3 to 6 mm depending on the child's size and the surgeon's technique) is trimmed at both ends and anastomosed to the inferior aspect of the subclavian artery and the superior surface of the pulmonary artery using 6/0 or 7/0 Prolene or Gortex suture. Once hemostasis is secured the chest is closed, leaving a chest drain *in situ.* When using prosthetic materials it is advisable to administer aspirin (10 mg/kg) rectally as soon as possible following formation of the anastomoses to minimize the risk of thrombus formation.

Partial correction

If the pulmonary arteries are small yet confluent and there is concern that this may produce right ventricular hypertension if complete repair is performed, a partial correction may be undertaken. The aim of this procedure is to relieve as much right ventricular outflow tract obstruction as possible and improve flow into the pulmonary arteries in order to promote their growth without risking right ventricular failure. Techniques for minimizing the obstruction are tailored to the individual patient and include resection of obstructing muscle bundles, pulmonary valvotomy, transannular patching, and pulmonary artery reconstruction. On occasion, insertion of a conduit from the right ventricular outflow tract to the pulmonary artery may be performed. Closure of the ventricular septal defect is undertaken at a later date when the pulmonary vasculature has developed sufficiently to allow acceptable right ventricular pressures (RV/LV pressure ration less than 0.5) following closure of the defect.

Total correction

The principles of total correction are patch closure of the ventricular septal shunt and relief of right ventricular outflow tract obstruction. This includes resection of ynfundibelar muscular stenosis, relief of pulmonary valvar obstruction, repair of main or branch pulmonary artery stenosis (which may be present either as a manifestation of the anomaly or as a result of a previously constructed shunt), and enlargement of the right ventricular outflow tract and hypoplastic pulmonary valve annulus. This may be achieved by the use of a monocusp (either native or aortic homograft) or by a free transannular patch of Dacron cloth or pericardium. In addition, previously constructed systemic to pulmonary artery shunts are ligated. It is therefore imperative that the surgeon enters the operating theater with a clear concept of the cardiac morphology as displayed by the cineangiocardiogram and echocardiogram.

The chest is opened through a median sternotomy. Preparations are made to establish cardiopulmonary bypass. If systemic–pulmonary artery shunts are present, they are dissected and ligatures are passed around them. Intravenous heparin (3 mg/kg) is given. Appropriate cannulas are inserted in the aorta and in the right atrium for bypass. Cardiopulmonary bypass is established and the systemic–pulmonary artery shunts are occluded without delay. Similarly, if present, the persistent ductus arteriosus must be ligated and divided. Division is particularly important if pulmonary valve insufficiency results at the end of the repair as the pulmonary artery dilation and elongation which accompanies this may produce stenosis at the ductal insertion site due to tethering of the pulmonary artery at this point. The patient is cooled to 25 to 28°C: infants weighing less than 10 kg may be placed under deep hypothermia (18°C) and total circulatory arrest. The aorta is cross-clamped and the myocardium is protected by infusing an initial dose of 20 ml/kg of cold crystalloid or blood cardioplegia. Repeat doses of half the volume of the initial dose of cardioplegic solution may be given every 30 min.

If preoperative assessment suggests the necessity of a transannular patch, a vertical right ventriculotomy is performed; this is controlled by two stay sutures on the outflow tract and later extended across the annulus onto the pulmonary artery. If there is associated valvular stenosis but the pulmonary artery annulus is of adequate size, either an atrial or a ventricular approach may be used. If a ventricular approach is adopted, a transverse ventriculotomy and a separate pulmonary arteriotomy can be used. The direction and extent of the right ventriculotomy is dictated by the distribution of the coronary arteries.

After opening the heart the ventricular septal defect should first be closed to allow firm placement of the sutures. If muscle resection is undertaken initially, there is no inherent strength in the raw cut muscle edges. The ventricular septal defect is closed using an appropriate sized circular patch of two-way stretch-woven Dacron or tanned pericardium with pledgetted Ethibond or Prolene sutures. Care is taken to avoid placing stitches in the region of the conduction bundle, the distribution of which is shown in Fig. 7. On completion of the ventricular septal defect closure the obstructing muscle bundles are divided. Identification of muscle bundles through the right ventriculotomy is depicted in Fig. 6. Initially, the parietal extension of the infundibular septum is dissected away from the right ventricular free wall and from the ventriculo-infundibular fold and divided transversely, safeguarding the right coronary cusp of the aortic valve. This increases the diameter of the infundibulum at its rightward margin. Any obstructing muscle bundles to the left of the outflow tract (septoparietal bundles) are excised completely. Low-lying obstruction is relieved by dividing the obstructing trabeculations above the moderator band which is usually preserved intact. Throughout the procedure, great care must be taken to protect the tension apparatus of the tricuspid valve.

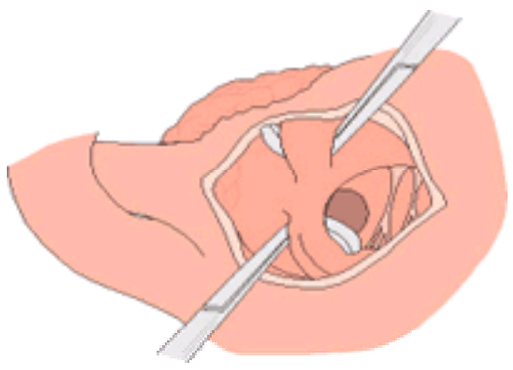


Fig. 6. The heart opened through right ventriculotomy. Obstructing muscle bundles prior to resection from the outflow tract.

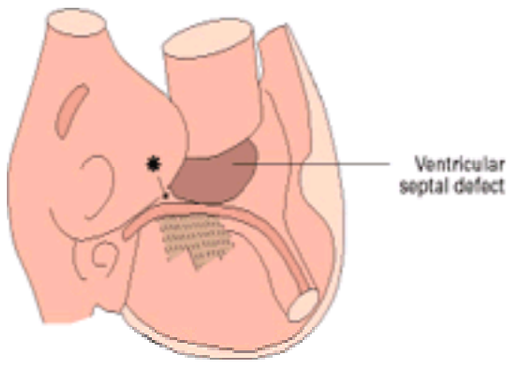


Fig. 7. Cardiac specialized conduction tissue in tetralogy of Fallot. Penetrating bundle is marked by asterisk, left bundle branch in dotted lines. Note proximity of bundle of His to the lower margin of the ventricular septal defect.

If pulmonary valvotomy is necessary, a pulmonary arteriotomy is performed taking care not to injure the pulmonary valve and commissures. Commissural fusion is divided after mobilizing cusp tissue from the vessel wall. In order to relieve the stenosis adequately it is often necessary to excise the thickened cusp edges, but this will normally result in some pulmonary valve incompetence.

The decision of whether to enlarge the pulmonary valve annulus is based on a critical assessment of its size. Preoperative measurements of the combined cineangiographic diameter of the right and left branch pulmonary arteries are compared with the diameter of the descending thoracic aorta at the diaphragm. At the time of operation, the absolute size of the main pulmonary artery and its relationship to the size of the aortic root is assessed. The true diameter of the pulmonary annulus can be estimated using Hegar dilators. The observed values are compared with standard values for the age and weight of the child and, using a predefined formula, the post-repair ratio of peak systolic pressure in the right ventricle to that in the left ventricle ($P_{RV/LV}$) is estimated. A ratio of less than 0.8 is acceptable, but 0.5 to 0.7 is desirable.

If the above observations suggest that the valve annulus requires enlargement, the transannular patch can take the form of a monocusp (native or aortic homograft) or a piece of pericardium, Dacron, or PTFE. The valve annulus is incised at the location of one of the commissures to maximize the amount of functional pulmonary valve tissue which is preserved. A patch of appropriate size is prepared and sewn in with continuous Prolene sutures. Since the purpose of the monocusp is to maintain reasonable pulmonary valve function, it is important that the bases of the native pulmonary valve cusps are in exact alignment with the base of the monocusp homograft (Fig. 8, Fig. 9). Alternatively, an entire homograft valve may be inserted (Fig. 10(a),Fig. 10(b),Fig. 10(c) and Fig. 10(d)).

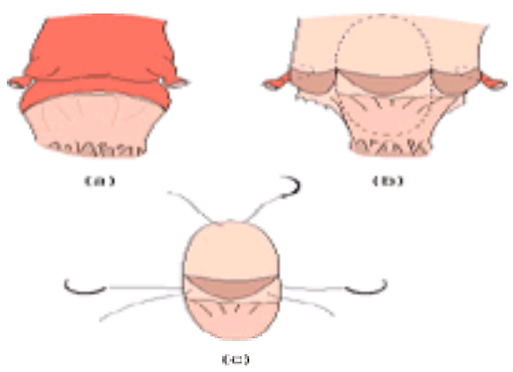


Fig. 8. Preparation of the monocusp homograft: the coronary artery stumps are sutured (a), the homograft tube is laid open (b), and a single cusp with aortic wall and the mitral leaflet is carefully dissected (c). Note the stay sutures at the base of the cusp for proper alignment to the base of the native pulmonary valve cusps.

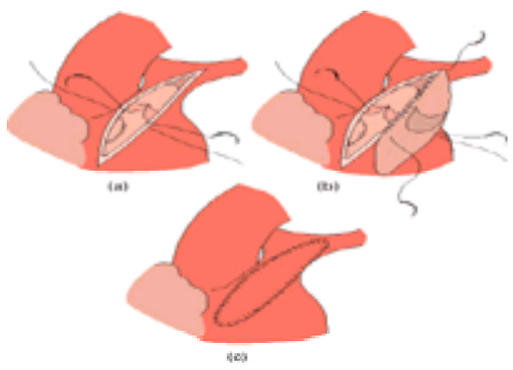


Fig. 9. Steps in the reconstruction of right ventricular outflow using monocusp homograft. (a) The right ventriculotomy is extended across the annulus onto the main pulmonary artery and stay sutures applied to the base of the valve cusps. (b) The monocusp homograft is properly aligned and sutured to the margins of the ventriculotomy and arteriotomy using continuous suture technique. (c) The completed suture line.

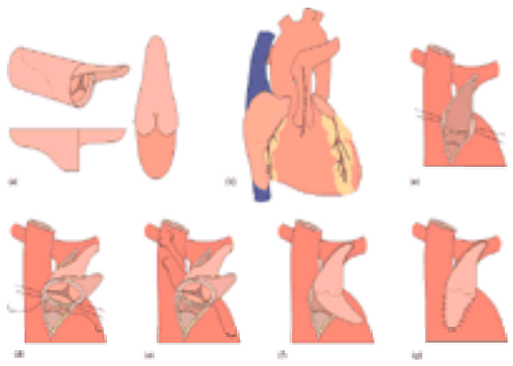


Fig. 10. Preparation and implantation of the complete homograft in pulmonary atresia: the preparation of the tube (a); right ventriculotomy is extended across the annulus onto the main (b), and if necessary, the left pulmonary artery (c); the margins of the incision are laid open (d); the suturing of the homograft is commenced at the level of pulmonary valve annulus (e, f, and g) and continued systematically along the pulmonary arteriotomy and right ventriculotomy.

If it is only necessary to widen the right ventricular outflow tract proximal to the valve, the right ventriculotomy is closed by the inclusion of a patch of Dacron, PTFE, or pericardium. The pulmonary arteriotomy is closed by direct suture or with an autologous or xenograft pericardium patch. Depending on the age of the patient and the anticipated degree of postoperative right ventricular injury, it may occasionally be advisable to leave a patent foramen ovale to allow unloading of the right ventricle. This is particularly true in patients in whom hypoplastic pulmonary arteries are being corrected and in neonates. Leaving a patent foramen ovale maintains systemic output and prevents right ventricular failure at the expense of some cyanosis. If a true atrial septal defect (as opposed to a patent foramen ovale) is present, it is closed via a separate right atriotomy or through the tricuspid valve. Closure of the ventriculotomy and the pulmonary arteriotomy can be performed following the release of the

aortic cross-clamp, thereby reducing the duration of myocardial ischemia. The patient is rewarmed and, following evacuation of air especially from the left side of the heart, cardiopulmonary bypass is discontinued. In small neonates it is advisable to leave both left and right atrial pressure monitoring lines to ensure adequate filling of the left side of the heart and reasonable assessment of right ventricular performance. Heparin is reversed with protamine sulfate. Following confirmation of hemostasis, the chest is closed leaving pericardial and mediastinal drains *in situ.* It is advisable to insert ventricular pacing wires in case these are required during the postoperative period. If sequential pacing is contemplated, two further electrodes are attached to the right atrium close to the sinoatrial node. Especially in neonates, junctional ectopic tachycardiac is common and adequate pacing ability is essential.

Repair of pulmonary atresia with ventricular septal defect

Pulmonary atresia implies pulmonary valve atresia with or without absence of the main pulmonary artery; there may thus be a gap from the right ventricle to the confluence of the branch pulmonary arteries. At birth, pulmonary blood flow is dependent on persistence of the ductus arteriosus. Surgically, either a shunt or complete repair (by closure of the ventricular septal defect and insertion of a suitable conduit in the form of an antibiotic-sterilized or cryopreserved aortic or pulmonary valved homograft, or a valve-bearing Dacron tube graft) is performed. It is essential in these patients to determine the complete segmental blood supply of each lung preoperatively as on occasion the native pulmonary arteries are hypoplastic or absent. In these circumstances pulmonary blood flow is supplied by developmentally primitive major aortopulmonary collateral arteries which usually originate from the descending thoracic aorta but may also come from the head and neck vessels or coronary arteries. Management of these latter individuals is problematic. Many surgeons undertake a staged repair with recruitment of the collateral vessels into a central confluence and insertion of a shunt into this confluence. These confluences may be connected up at a later date with establishment of an anatomically normal circulation via the use of a right ventricular–pulmonary artery conduit and closure of the ventricular septal defect. Alternatively, a more radical approach has been pioneered by Hanley and associates. In this approach termed one-stage unifocalization through a median sternotomy, all the collateral vessels are anastomosed together, a right ventricular–pulmonary artery conduit (usually a homograft) is inserted, and the ventricular septal defect is closed. This approach, essentially championed by a single group, has produced excellent results with operative mortality of 11 per cent and medium-term survival of 78 per cent.

Repair of tetralogy of Fallot with anomalous origin of the left anterior descending coronary artery from right coronary artery

It is important that this anomaly is recognized preoperatively as the left anterior descending coronary artery may be buried in the myocardium or fat, especially if pericardial adhesions from a previous sternotomy have obliterated the pericardial space. The passage of this artery across the right ventricular outflow tract poses a special danger if a vertical incision into the right ventricular outflow tract is required. The available options are: (i) transverse ventriculotomy and repair using a conduit from the right ventricle to the pulmonary artery thus bypassing the critical area, or (ii) dissection of the left anterior descending coronary artery from its bed in the right ventricular wall throughout most of its length from its origin to within a few millimeters of the interventricular groove. The ventriculotomy is then made beneath the artery and across the annulus if necessary. If this option is selected, great care must be taken not to damage any coronary artery or branches.

Postoperative care and complications

Following shunt operations, babies are nursed in a special unit whose nursing staff are familiar with the management of neonates. A rare, but often lethal, complication following a shunt is hemorrhagic pulmonary edema which can produce hypoxia and rapid clinical deterioration. The chest radiograph is diagnostic showing severe vascular congestion of the ipsilateral lung. This is due to the direct effect of the sudden increase in pulmonary blood flow and may be an indication for immediate reoperation. The anastomosis is narrowed appropriately. Low-dose heparin may be continued in the early postoperative period and is replaced later by low-dose aspirin to maintain the shunt patency. Less marked overperfusion of the lungs is usually controlled with diuretic therapy which may be required for several weeks. Following reconstructional surgery, patients may require cardiorespiratory support including sedation and ventilation and infusion of an inotropic agent such as dopamine. Complete heart block, which is rare and usually temporary, is managed by ventricular pacing.

Patients with tetralogy of Fallot show a particular tendency to accumulate interstitial, pleural, and peritoneal fluid early in the postoperative period. This may be due to chronic cyanosis which affects the permeability of capillary membranes. Fluid retention is particularly common in the early period and diuretic therapy is often necessary, keeping the central venous pressure as low as possible while maintaining cardiac output.

Results

The creation of classic and modified Blalock–Taussig shunts in children with uncomplicated tetralogy of Fallot carries a low operative mortality rate (0.1 to 2.5 per cent). However, when performed in very young infants with pulmonary atresia the mortality rate can be as high as 10 to 15 per cent. The Waterston shunt has a higher complication rate than the Blalock–Taussig shunt. While palliative transannular patching with cardiopulmonary bypass has a low hospital mortality in children and young adults, it can be in the region of 10 to 20 per cent in infants.

Since tetralogy of Fallot is morphologically and clinically a highly variable entity, it is difficult to make generalizations about surgical results. However, it is certain that the in-hospital mortality after repair of uncomplicated tetralogy of Fallot has steadily declined over the last three decades. Under proper conditions, the mortality rate in this group of patients, whether the repair is carried out primarily or as a two-stage procedure, is approximately 4 to 8 per cent. The risk is higher in patients with coexisting pulmonary arterial problems, associated cardiac anomalies, or in those with high hematocrit or a small left ventricle. The long-term results following total repair, both in terms of late survival and quality of life, are excellent. Late complications are not infrequent however. Restenosis of the right ventricular outflow tract may occur, particularly if a conduit is used in a young child. Also, the pulmonary valve regurgitation which is often produced may lead to significant right ventricular dilatation with attendant dysrhythmias and deterioration in right ventricular function. These complications require reoperation with the insertion of a valved conduit. The operation cannot therefore be considered curative and the survival curve beyond 10 years is slightly but significantly less than that of a matched general population.

Univentricular heart

The terminology used to describe a univentricular heart has varied in the literature. In the past, reference has been made to a single ventricle, common ventricle, double-inlet left ventricle, primitive ventricle, and corbiatriatum triloculare. These are synonymous with univentricular atrioventricular connection. Although this congenital abnormality has been known to cardiac pathologists for a number of years, it is only in the past 15 years that significant advances in its surgical management have been reported.

Univentricular hearts can be subclassified according to the morphology of the main ventricular chamber. This may exhibit left ventricular morphology (the most common variety), right ventricular morphology, or an indeterminate trabecular pattern. From the surgical standpoint it is important to distinguish between patients with and without an outlet chamber. This gives origin to one or both of the great arteries. The ventricular septal defect (bulboventricular foramen) connecting the main chamber to the outlet chamber may be restrictive in size. In the univentricular heart of left ventricular morphology, the atria connect with the posterior ventricle, and the outlet chamber (minor ventricle) is always positioned anterosuperiorly either to the left or to the right of the main ventricular chamber, or very occasionally centrally. The reverse is true for the univentricular heart of right ventricular morphology. Different morphologic types of univentricular heart are shown in Fig. 11. Other features of surgical importance are the type of atrioventricular connection (double inlet, or absent right or left atrioventricular connection), mode of atrioventricular connection (two valves, common valve, or a single valve), type of ventriculoarterial connection (concordant, discordant, double outlet, single outlet), presence or absence of subpulmonary or pulmonary stenosis, and any associated malformations, which occur in approximately one-third of patients.

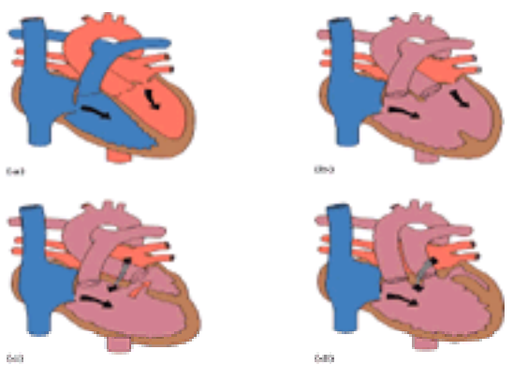


Fig. 11. Normal and univentricular hearts. (a) Normal heart with intact interventricular septum and two atrioventricular valves leading into the respective ventricular chamber (black arrows). (b) Common ventricle: a large ventricular septal defect with two atrioventricular valves opening into the respective ventricular chambers (black arrows). (c) Univentricular heart of left ventricular morphology. (d) Univentricular heart of right ventricular morphology. Black arrows depict the double inlet to the main ventricular chamber, while the red arrow shows the position of the bulboventricular foramen leading into the rudimentary ventricular chamber. The bulboventricular foramen is restrictive in (d).

Clinical features and diagnosis

The clinical features vary with the morphology of the pulmonary valve and subpulmonary region. Patients without pulmonary stenosis (approximately one-third) have a presentation similar to that of a large ventricular septal defect except that, since the main chamber acts as a common mixing chamber, mild cyanosis may be present. A mild to moderate pulmonary stenosis may cause no symptoms in the early years of life although cyanosis and a systolic murmur usually ensues in childhood. Severe pulmonary stenosis or pulmonary atresia usually presents with severe cyanosis in the neonatal period; this may be mistaken for tetralogy of Fallot.

The electrocardiogram and chest radiograph raise the suspicion of a univentricular heart which is confirmed by echocardiography. Cineangiography may be carried out if surgical treatment is under consideration.

Natural history

The natural history of the condition is almost as varied as its morphology. It is usually a lethal cardiac abnormality and results in a greatly reduced life expectancy. The most common causes of death are dysrhythmias, congestive cardiac failure, pneumonia, and sudden death.

Surgical management

In view of the poor natural history of the condition, complete repair of a univentricular heart is desirable if the surgical risks are not too great. The type of surgery and the approximate timing are highlighted in Fig. 12. Repair may be by one of two methods: septation or the Fontan procedure. In the former two parallel circulations are created, whereas in the latter two circulations in series are produced. This latter circulation is obviously hemodynamically disadvantageous and the single ventricle is required to perform more work. For this reason great care must be taken to preserve the ventricle as well as is possible and great thought must go into the decision-making process.

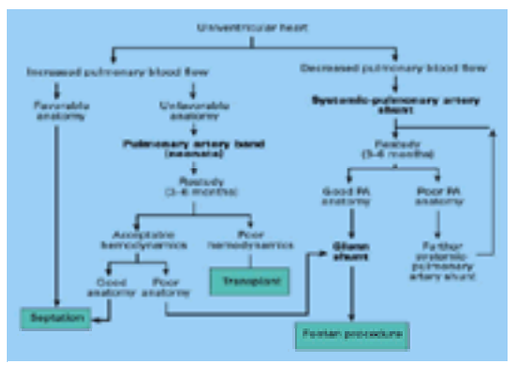


Fig. 12. Flow chart of the suggested course of management in the univentricular heart.

The principles involved in maximizing ventricular performance are: minimizing afterload, volume overload, and ischemia. Therefore, all outflow obstruction should be relieved be it coarctation, ventricular outflow tract obstruction, or restriction at the bulboventricular foramen. If found to be restrictive it is possible to excise some muscle around the bulboventricular foramen to enlarge it, but great care must be exercised so as not to injure the conduction system. Careful thought must also go into the assessment of the heart prior to the first step of palliation; if the heart is inevitably to be a single ventricle repair and there is any question regarding the adequacy of the bulboventricular foramen, a Damus–Kaye–Stansel aortopulmonary anastomosis is likely to give better long-term results. However, the increased morbidity associated with a hypothermic bypass run must be balanced against the likelihood of the bulboventricular foramen becoming restricitve. Volume overload should also be avoided; atrioventricular and ventriculoarterial valvar regurgitation should be aggressively dealt with to prevent fibrotic degeneration of the myocardium. owever, it should be remembered that the Blalock–Taussig shunt also presents a significant volume overload. In many institutions this is leading to surgeons performing Glenn shunts (superior vena cava to pulmonary artery) at younger ages, even as young as 3 to 5 months. The Glenn shunt, a stepping stone to a Fontan, reduces the volume overload of the ventricle to a single cardiac output.

As well as protecting the ventricle, the pulmonary circulation must also be protected. For this reason, in patients with excessive pulmonary blood flow, pulmonary artery banding should be performed in the neonatal period. This will decrease the risk of pulmonary vascular changes developing.

The principles of initial management are that patients with high pulmonary blood flow and cardiomegaly resulting in cardiac failure receive pulmonary artery banding. Those presenting with severe cyanosis are offered a systemic–pulmonary artery shunt, which is often combined with atrial septectomy if the atrial septal defect is restrictive or if one of the atrioventricular valves is stenotic or atretic.

Pulmonary artery banding

This is performed through an anterolateral thoracotomy on the same side as the pulmonary artery or a median sternotomy. The pericardium is opened and the dilated pulmonary artery is separated from the aorta with great care. Minimal dissection is performed to diminish the risk of band migration. A pressure-monitoring catheter is placed in the distal pulmonary artery. A 3- to 4-mm wide Teflon tape is marked at the formulated length (Trussler's formula: 22 mm + 1 mm/kg) and then passed around the main pulmonary artery. With the tape in place, the two ends are stitched together at a slightly looser fit than ultimately anticipated. Metal clips can then be applied progressively more proximally to ‘fine tune' the band tightness. The excess band can ultimately be trimmed away. Heart rate, oxygen saturation, and systemic and pulmonary arterial pressures are noted throughout the procedure. If bradycardia develops due to a significant reduction in oxygen saturation, the band is loosened a little; if the band is too slack as judged by the pressure difference across the band (distal pulmonary artery pressure should be between 30 and 50 per cent of the systemic pressure), the band is tightened. Arterial saturations for completely mixing lesions should be approximately 85 per cent at the end of the procedure. One or two stitches are placed between the pulmonary artery adventitia and the band to prevent it migrating towards the bifurcation of the pulmonary artery as this may cause severe obstruction at the origin of the right main pulmonary artery. The pericardium and the chest are closed leaving a drain *in situ.* A postoperative cinefilm of a child with pulmonary artery band in place is shown in Fig. 13.

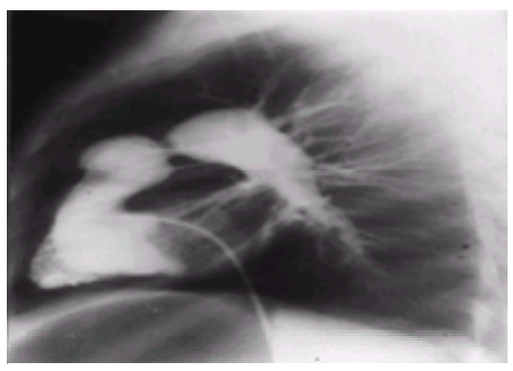


Fig. 13. Lateral cineangiogram of right ventricle and pulmonary arteries showing banded main pulmonary artery.

The ‘repair': septation compared with Fontan-type procedure

Following an appropriate palliative procedure, indications for repair by septation or by modified Fontan procedure include significant or increasing symptoms, progressive cardiac enlargement, and a rising hematocrit and hemoglobin level. In view of the increased operative risks incurred when a valved extracardiac conduit is added to septation to bypass severe subpulmonary stenosis, or when concomitant atrioventricular valve replacement is necessary, septation is usually not offered to these categories of patients. The modified Fontan procedure is preferred for patients with low pulmonary arterial pressure and resistance, since this operation creates a wide direct anastomosis between the right atrium and the pulmonary arteries. The optimal age for performing a Fontan procedure is unclear. Indeed, some authorities argue that a Glenn shunt with a small additional source of blood flow (such as a systemic–pulmonary artery shunt) is a better final palliation. Further follow-up is required to answer this question. Another lively argument in the Fontan literature pertains to the type of Fontan to be performed (Fig. 14, Fig. 15, Fig. 16 and Fig. 17). Although the classic Fontan and its variants are rarely performed today, total cavopulmonary connection, lateral tunnel Fontan, and extracardiac conduit Fontan all have their proponents. The first two have the advantage that they can grow with the patient and as they are fashioned out of natural tissue the possibility of thrombosis is low. However, they both have the risk of atrial enlargement (sometimes to enormous size) with the possibility of associated atrial dysrhythmias and pulmonary venous obstruction. The extracardiac Fontan cannot grow with the patient and therefore should not be performed until a graft of similar diameter as the adult inferior vena cava can be inserted (approximately 15 kg). In addition, as prosthetic material is used there is the risk of conduit thrombosis. However, there is little chance of the atria distending and becoming problematic.

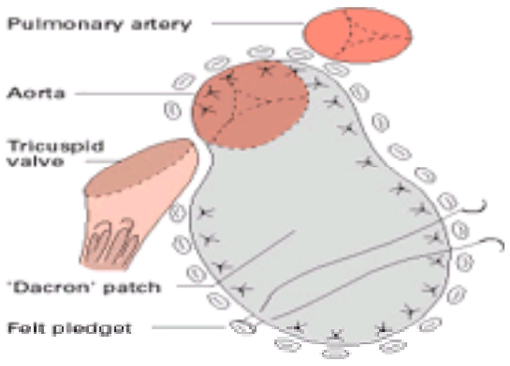


Fig. 14. Septation patch in a univentricular heart.

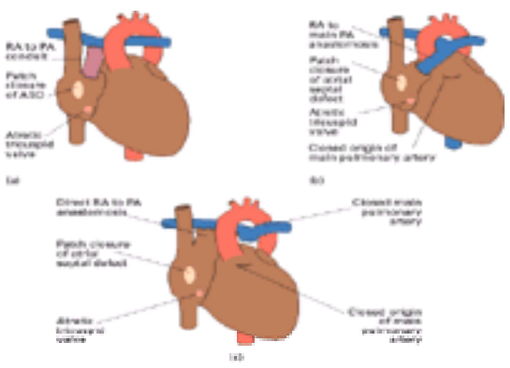


Fig. 15. Diagrams illustrating the types of atriopulmonary connections (Fontan procedures). (a) Right atrium to pulmonary artery (RA to PA) conduit. (b) Kreutzer's modification in which the patient's own pulmonary valve is connected, along with the pulmonary trunk, to the right atrium. (c) Direct anastomosis of the right atrial appendage to the undersurface of the right pulmonary artery, often using an onlay patch of pericardium.

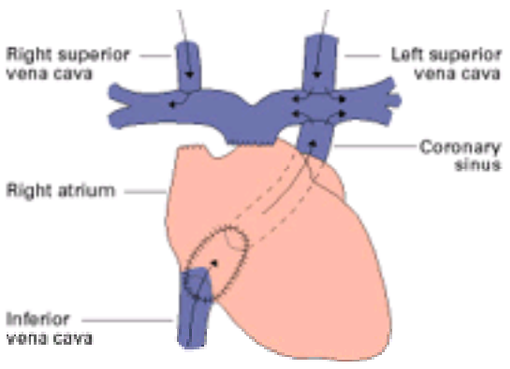


Fig. 16. Total cavopulmonary anastomosis. In this patient the presence of a left superior vena cava drained into a much dilated coronary sinus. This was used as the extension of the inferior vena cava to the left main pulmonary artery.

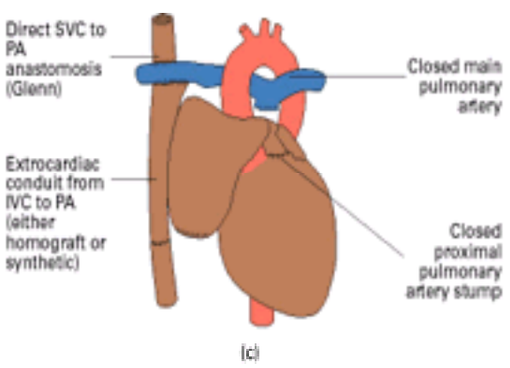


Fig. 17. Extracardiac Fontan. The right atrium is completely bypassed except for the coronary sinus blood.

In view of the potential for cardiac transplantation in the future, all operative procedures are planned to avoid development of pulmonary vascular disease and distortion of the pulmonary arteries. The technical aspects of septation or modifications of the Fontan procedure are beyond the scope of this chapter. However, Fig. 14, Fig. 15, Fig. 16 and Fig. 17 give an overview of different procedures. Some spectacular surgical successes have been reported from a few centers when septation has been performed in hearts of left ventricular morphology with a left-sided outlet chamber, giving rise to an unobstructed aorta.

Results

Hospital mortality rates following septation are still high. However, since the technical details and indications for septation continue to evolve, this current high mortality may not continue in the future. The mortality rate associated with modified Fontan procedures varies between 10 and 20 per cent and is largely dependent upon careful patient selection. The medium-term results of both septation and modified Fontan procedures are considered encouraging in many patients.

Further reading

Anderson RH, Allwork SP, Ho SY, Lennox CC, Zuberbuhler JR. Surgical anatomy of tetralogy of Fallot. *Journal of Thoracic and Cardiovascular Surgery* 1981; **81**: 887. [An anatomic study of postmortem specimens from patients with tetralogy of Fallot with particular reference to features of surgical significance.]

Hamilton DI, Di Eusanio G, Piccoli GP, Dickinson DF. Eight years experience with intracardiac repair of tetralogy of Fallot. *British Heart Journal* 1981; **46**: 144. [An overall review of one team's experience over an 8-year period dealing with tetralogy of Fallot identifying lessons learned and how these altered clinical practice.]

Kirklin JW, Karp RB. *The tetralogy of Fallot from a surgical viewpoint*. WB Saunders, Philadelphia, 1970. [A comprehensive textbook dealing with all aspects of tetralogy of Fallot from a surgical viewpoint.]

Reddy VM, Liddicoat JR, Hanley FL. Midline one-stage complete unifocalisation and repair of pulmonary atresia with ventricular septal defect and major aortopulmonary collaterals. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 832–44. [The experience with 10 patients over a 20-month period of an aggressive new approach to tetralogy of Fallot with pulmonary atresia and multiple aortopulmonary collateral arteries with particular reference to early and mid-term results.]

Stark J, de Leval M. *Surgery for congenital heart defects*. Grune & Stratton, London, 1983. [A general textbook giving an overview of surgical procedures for all forms of congenital heart disease.]

40.7.5 Abnormalities of the atrioventricular valves

Ravi Pillai and Kenny Sin

Introduction
Abnormalities of the right atrioventricular valve
Tricuspid atresia
Ebstein's anomaly
Abnormalities of the left atrioventricular valve
Further reading

Introduction

Simple morphological abnormalities of the right and left atrioventricular valves are uncommon congenital lesions. They include stenosis or absence, straddling valves, or abnormalities of the morphologically tricuspid right or the bicuspid (mitral) left atrioventricular valve. Additionally, a variety of anatomical abnormalities may exist, not only in the setting of situs solitus (usual) but in association with situs inversus or situs ambiguus (right or left atrial isomerism). Such lesions may be detected clinically at the time of birth, in infancy, in adolescence, or in late adulthood. Furthermore, they may be associated with abnormalities of atrioventricular or ventricular arterial connections.

In this chapter the common abnormalities of the right and left atrioventricular valves primarily in situs solitus are discussed.

Abnormalities of the right atrioventricular valve

The most common abnormality of the right atrioventricular valve is absence or atresia. In others the tricuspid valve may be partially dysplastic and displaced downwards into the right ventricular cavity (Ebstein's anomaly).

In hearts with ventricular septal defects there may be a variance in chordal attachment, leading to straddling of the interventricular septum to a greater or lesser degree.

Tricuspid atresia

Complete absence of the right atrioventricular valve occurs in 1 to 3 per cent of all congenital cardiac abnormalities. The vast majority of patients present at birth and the condition is one cause for cyanotic congenital heart disease (so-called blue babies).

Morphology

In the majority of patients there is complete absence of the right atrioventricular connection (Fig. 1) with ventriculoarterial concordance (i.e., left ventricle connected to the aorta). In about 30 per cent the ventriculoarterial connection is discordant (transposition); in these patients the presence of an adequate ventricular septal defect and unrestricted pulmonary blood flow prevents cyanosis in the neonatal period. Such patients may present later in life. In the majority, however, the absent right atrioventricular valve gives rise to a muscular right atrium, with a large interatrial communication through a secundum defect. There is total absence of the inlet portion of the right ventricle, with the main and outlet chamber of that ventricle varying in size. The size of the associated ventricular septal defect may well have a bearing on the capacity of the residual chamber. The septal defect is rarely restrictive and there may be associated pulmonary atresia with small main and branch pulmonary arteries.

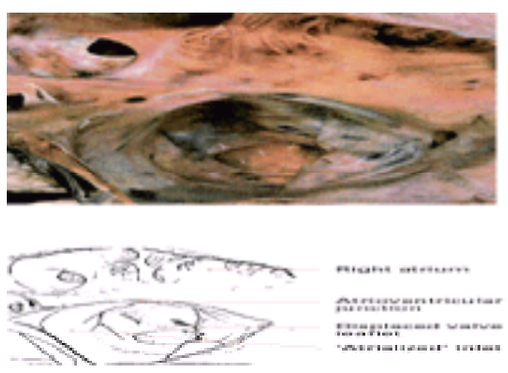


Fig. 1. Tricuspid atresia demonstrating blind-ending right atrium.

Occasionally there is considerable valve tissue that on examination is perforated, or, more commonly, the valve tissue is represented by a nub of membranous tissue related to the fibrous body. The atrioventricular node, usually in the triangle of Koch, is thus displaced posteriorly and inferiorly towards the membranous nub.

Clinical features

Most of these children present at the time of birth or a few weeks afterwards. Most of them are desaturated and are therefore clinically cyanotic. In a small proportion with a large ventricular septal defect or in the presence of ventriculoarterial discordance (transposition), progressive cyanosis may be associated with a closing ventricular septaldefect. Alternatively, in this small group with transposition of the great vessels, there may be excessive pulmonary blood flow giving rise to tachypnoea, failure to feed, and poor weight gain.

On examination there is a systolic murmur from the apex to the left sternal edge, with a palpable left ventricular heave. Children with associated pulmonary atresia invariably have a persistent ductus arteriosus, giving rise to a continuous murmur. Left untreated, the majority of children born with cyanosis become progressively more cyanotic and dyspnoeic, and may have syncopal attacks. Most of these children will die within the first 9 months of life if not treated.

Investigations

Chest radiograph

In patients with tricuspid atresia and situs solitus there are reduced vascular lung markings, reflecting reduced pulmonary blood flow. The heart shadow itself is within normal limits, with reduced prominence of the main, left, and right pulmonary arteries. These features are reversed in the small group of patients with tricuspid atresia and ventriculoarterial discordance.

Electrocardiography

The electrocardiogram shows left-axis deviation with signs of left ventricular hypertrophy. The left ventricle, being the recipient of both systemic and pulmonary venous return, tends to be large, dilated, and hypertrophic.

Echocardiography

This investigation, combined with the present-day availability of colour flow Doppler echocardiography, is the mainstay of investigation and diagnosis of tricuspid

atresia. Most features of this condition, as well as any related abnormalities, may be clearly seen.

Angiocardiography

In a small number of patients, cardiac catheterization is needed to delineate the size of the main and branch pulmonary arteries, as well as to confirm the findings of any other associated abnormality. In those patients who undergo initial palliation, catheterization, apart from confirming morphological features, allows the measurement of pulmonary vascular resistance and thereby facilitates corrective surgery.

Clinical management

Following the diagnosis of tricuspid atresia in a cyanosed child at birth, attempts to increase pulmonary blood flow by keeping the arterial duct patent are made. This is achieved in the neonatal period by an infusion of prostaglandin E_1, following which a palliative shunt procedure is performed.

In the smaller group of patients with ventriculoarterial discordance, who invariably present much later after birth, the haemodynamic disturbance is due to excessive blood flow to the lungs. Palliative surgery is performed in the neonatal period or in the first few weeks or months of life. Following such palliation, corrective surgery to separate the systemic and pulmonary circulations permanently is performed once the child is 30 months or older.

Surgical correction

Palliative procedures Palliative procedures fall into two main categories. The first and most common is for reduced pulmonary blood flow and cyanosis. Several systemic-to-pulmonary artery shunts have been described over the last 40 years. One of the early shunts described by Blalock and Taussig was in the context of treatment of tetralogy of Fallot; it consists of an anastomosis between the right subclavian artery and the right pulmonary artery. An end-to-side anastomosis between the right pulmonary artery and the superior vena cava was described by Glenn in 1958, specifically for the palliation of tricuspid atresia. In the modern era, the construction of a classical Glenn shunt is felt to compromise the planning of further ‘corrective' surgery.

Our preference is for the concept of the Blalock–Taussig shunt, in which a connection is created between the subclavian artery and the pulmonary artery, using prosthetic materials, such as Dacron, Gore-Tex, or Impra (polytetrafluoroethylene), with the exception of where the shunt in itself is to be considered a permanent treatment. The initial shunt is placed on the same side as the aortic arch, bearing in mind the possible need for a second shunt as the child grows.

Modified Blalock–Taussig shunt With general anaesthesia and continuous monitoring of arterial and systemic venous pressure, the child is placed in a left lateral position. The chest is entered through a left lateral thoracotomy. After the left subclavian artery has been identified the left lung is gently retracted inferiorly and posteriorly, exposing the left pulmonary hilum. The pleura over the left hilum is incised and the left main pulmonary artery and its branches dissected out.

It is not unusual for a large number of thin-walled collateral vessels to be present. Careful haemostasis at this stage avoids excessive postoperative bleeding. Distal branches of the pulmonary artery are controlled with circular sutures, while the totally free, main pulmonary artery allows a clamp to be placed proximally towards the pulmonary bifurcation. The left subclavian artery is dissected free, after which heparin 1 mg/kg body weight is given. Following this the Gore-Tex, Impra, or Dacron conduit, varying in size between 3 and 5 mm in diameter, is anastomosed to the site of the origin of the left subclavian, having occluded the subclavian completely with a side-biting clamp. The distal end of the conduit is anastomosed to the left main pulmonary artery, having removed a small button from the superior surface of the arterial wall.

On completion, first the distal and then the proximal clamps are released. On securing haemostasis a single intercostal drain is placed in position and the chest closed in layers.

Corrective procedures In 1968, Fontan described the first procedure to separate permanently the systemic and pulmonary circulations in tricuspid atresia.

The original procedure consisted of a Glenn shunt (right pulmonary artery to superior vena cava anastomosis) and further anastomosis between the right atrial appendage and the main pulmonary artery, following division and oversewing of the main pulmonary artery.

In the early experience described, a homograft valve was inserted between the inferior vena cava and the right atrium. In the last 20 years there have been several modifications of the procedure by Fontan himself, as well as others. Today, the phrase ‘Fontan' or ‘modified Fontan' procedure has been attached to a large variety of procedures that seek to divide the systemic and pulmonary circulations.

Ebstein's anomaly

Ebstein described the features of this syndrome in 1866, but it was not until the 1950s that Hunter and Lillehei described methods of surgical correction of the morphological abnormalities.

Patients may present soon after birth, in childhood, or in late adulthood. This anomaly may coexist with other associated abnormalities, which may have significant impact on prognosis.

Morphology

The characteristic features of Ebstein's disease are abnormality of the tricuspid valve both in structure and position, and an abnormal right ventricle with resulting gross enlargement of both the right ventricular and atrial chambers. The septal and posterior leaflets of the tricuspid valve may be dysplastic, with abnormal attachment of the chordal apparatus to the ventricular wall. There may be a varying degree of tethering of the chordae as well as attachment of the leaflets themselves to the ventricular wall. In addition, the commissure between the septal and posterior leaflets is displaced inferiorly into the body of the ventricular cavity to a greater or lesser extent. The junction between the septal and anterior leaflet, however, remains in a normal position, as does the anatomy of the conduction tissue with a normally sited atrioventricular node. The anterior tricuspid leaflet is usually normal in appearance but appears enlarged, tending to billow out into the atrium during systole.

The right ventricular wall proximal to the septal and posterior leaflet tends to become ‘atrialized', with a smooth appearance, and may be thinned out (Fig. 2). The right ventricle is grossly enlarged, while the only functional parts of the cavity, mainly the apex and the infundibulum, are variably developed.

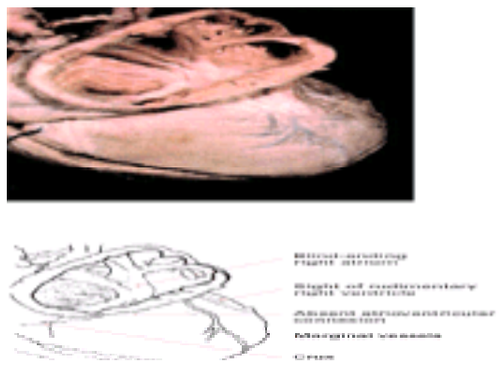


Fig. 2. Ebstein's anomaly demonstrating displaced position of the septal and posterior tricuspid leaflets and the atrialized portion of the right ventricle.

Associated abnormalities

About 60 per cent of patients have associated secundum atrial septal defect. Other common associated lesions include pulmonary stenosis or pulmonary atresia, persistent arterial duct, and ventricular septal defect. In addition, many patients have conduction abnormalities such as Wolff–Parkinson–White syndrome, and

paroxysmal and re-entrant supraventricular tachyarrhythmias.

Clinical presentation

Neonates may present soon after birth with cyanosis and congestive cardiac failure. Often these babies have associated abnormalities such as pulmonary stenosis or pulmonary atresia and a right-to-left atrial shunt. In the absence of such extra anomalies and depending on the degree of tricuspid regurgitation, presentation may be later and indeed they remain relatively asymptomatic late into adulthood. The potential for supraventricular tachyarrhythmias and other conduction defects can occasionally cause sudden death.

Chest radiography shows a large, rounded cardiac shadow with unusually clear lung fields. In extreme cases associated with poor forward blood flow the lungs themselves may be hypoplastic.

Echocardiography, both M-mode and two-dimensional, is used extensively as the diagnostic tool of choice. The size of the right ventricle and its function are easily determined, as is the extent of tricuspid regurgitation and the presence of any associated anomalies.

Cardiac catheterization and angiography is reserved for patients in whom specific haemodynamic data are required before planning surgery. The extent of tricuspid regurgitation may be significantly underestimated because of the grossly enlarged right ventricular and atrial cavities.

In patients with reasonable tricuspid valve function the shunt at atrial level may in fact be from left to right. In this event the lung fields may appear to be hyperaemic, with clinical signs of high pulmonary blood flow.

Treatment

Given the spectrum of functional disability the surgical procedures performed vary from complex repairs to simple closure of the secundum-type atrial septal defect. In the very sick neonate who is in cardiac failure and is cyanotic following initial resuscitation, the Starnes operation is the procedure of choice.

On cardiopulmonary bypass using deep hypothermic circulatory arrest, the orifice of the tricuspid valve is closed with a patch. The atrial septal defect is enlarged and a central aortopulmonary shunt constructed. The procedure essentially converts the anatomy to one of a univentricular heart. Following this initial procedure, a Fontan type repair is carried out at a later stage.

In the majority of patients, however, surgery is undertaken late. In over 60 per cent of patients presenting in infancy, adolescence, or late adulthood, the tricuspid valve is repaired, with plication of the atrialized portion of the right ventricle in the majority of cases. The coexisting atrial defect is closed with a patch.

There are two commonly described methods of repair. The Danielson repair consists of a transverse plication of the atrialized portion with a posteroinferior suture plication of the valve annulus, thereby converting the tricuspid valve to a monocusp valve, with functional competence being provided by the large anterior leaflet.

The Carpentier repair consists of a longitudinal plication of the atrialized portion with a annuloplasty using a ring. In a small proportion of patients with good function of the tricuspid valve and good residual function of the right ventricle, closure of the atrial septum alone may bring about symptomatic relief.

Surgical repair in adolescence, in the presence of good ventricular function, produces good long-term results. Sudden death from arrhythmias secondary to aberrant conduction pathways still remains a possibility. Such pathways may be ablated either at the time of surgery or before surgery, percutaneously in the catheter laboratory.

Abnormalities of the left atrioventricular valve

Abnormalities of the left atrioventricular valve or mitral valve at the time of birth may occur in a variety of settings. The valve may be imperforate, absent, atretic, or stenosed because of fusion of the subvalvular apparatus. Fusion of the papillary muscles produces the so-called parachute mitral valve.

Mitral valve atresia or absence is often associated with the hypoplastic left-heart syndrome in which the left ventricle and the aorta are small and hypoplastic. Occasionally the presence of a large ventricular septal defect gives rise to an adequate left ventricular cavity. In such cases, ventriculoarterial discordance (transposition) is the usual anatomical feature.

Echocardiography is the mainstay of diagnosis, and many of these children are kept alive by reversed flow through a persistent arterial duct.

Surgical repair in many of these conditions is complex and usually involves early palliation followed by subsequent conversion of the haemodynamics to one of univentricular heart that is right ventricular in type. In a proportion of patients the only morphological abnormality is a cleft in the anterior mitral leaflet (Fig. 3). Such a cleft lies in the aortic outflow and may give rise to varying degrees of mitral regurgitation. These patients may present early in life or may indeed remain asymptomatic well into adult life.

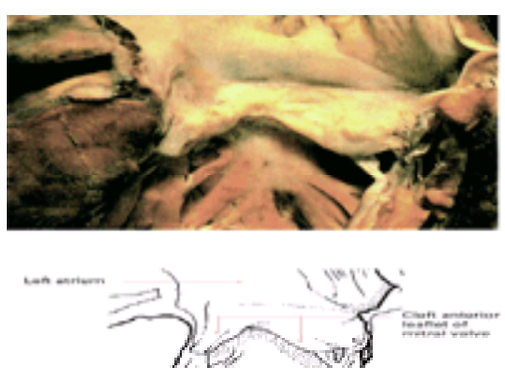


Fig. 3. True cleft of the anterior mitral leaflet.

When a direct surgical approach to the mitral valve is contemplated, preservation of the valve is of paramount importance, particularly in infants and children. Even in adults the majority of cleft mitral valves may be repaired successfully. A proportion of patients with connective tissue disorders such as Marfan syndrome may present early in life with mitral regurgitation due to stretching of the chordal apparatus with prolapse of segments of the cusps into the left atrium. In the early part of life, some of these valves may be amenable to repair.

If valve replacement is required, mechanical valves are used predominantly in the younger patient in order to avoid the accelerated calcification and early valve dysfunction associated with tissue valves. The disadvantage of mechanical valves is the need for permanent oral anticoagulation.

Further reading

Becker AE, Anderson RH. *Cardiac pathology*. Churchill Livingstone, Edinburgh, 1982.

Carpentier A *et al.* A new reconstructive operation for Ebstein's anomaly of the tricuspid valve. *Journal of Thoracic and Cardiovascular Surgery* 1988; **96**: 92–101. [A retrospective review of 14 patients treated surgically for Ebstein's anomaly of which 13 underwent repair with two deaths. The surviving patients had a marked improvement in functional class; echocardiography showed a normal-shaped right ventricle and good tricuspid valve function in all patients but one.]

Danielson GK, Driscoll DJ, Mair DD, Warnes CA, Oliver WC. Operative treatment of Ebstein's anomaly. *Journal of Thoracic and Cardiovascular Surgery* 1992; **104**: 1195–202. [A retrospective review of 185 patients of which 58.2 per cent had tricuspid valve reconstruction, 36.5 per cent had valve replacement, and 5.3 per cent a modified Fontan or other procedure. Early mortality was 6.3 per

cent, with 92.9 per cent of patients on follow-up in New York Heart Association class I or II.]

Del Torso S, Kelly MJ, Kalff V, Venables AW. Radionuclide assessment of ventricular contraction at rest and during exercise following the Fontan procedure for either tricuspid atresia or single ventricle. *American Journal of Cardiology* 1985; **55**: 1127–32. [A radionuclide assessment of ventricular function in 18 asymptomatic patients after a Fontan operation which showed that only two patients had both normal ejection fraction at rest and normal exercise response.]

Fontan F, Baudet E. Surgical repair of tricuspid atresia. *Thorax* 1971; **26**: 240–8. [A landmark article on the Fontan operation for tricuspid atresia.]

40.7.6 Stenosis and atresia of the arterial valves

Marshall L. Jacobs

Pulmonary stenosis
Critical pulmonary stenosis of the neonate
Pulmonary stenosis in infants and children
Pulmonary atresia with intact ventricular septum
Clinical features
Surgical management
Congenital aortic stenosis
Clinical features
Surgical management
Aortic atresia
Clinical features
Surgical management
Further reading

Pulmonary stenosis

In this section, pulmonary stenosis is taken to mean primarily valvular pulmonary stenosis in hearts with intact ventricular septum. Valvular pulmonary stenosis may coexist with infundibular pulmonary stenosis, but obstruction of the right ventricular outflow tract in association with a ventricular septal defect is considered elsewhere.

Critical pulmonary stenosis of the neonate

Neonates with critical valvular pulmonary stenosis are dependent upon patency of the ductus arteriosus for pulmonary blood flow. The pulmonary valve is most often a uniform cone of fibrous tissue with a central or eccentric orifice considerably smaller than the pulmonary annulus itself. Two or three fibrous ridges radiate from the central orifice to the annulus, dividing the valve tissue into two or three leaflets that correspond to the two or three, usually well developed sinuses of Valsalva. The leaflet tissue is in general thicker than that of a normal, non-stenotic pulmonary valve. In hearts with critical pulmonary stenosis, the right and left pulmonary arteries are usually well developed, though moderate to severe branch pulmonary artery hypoplasia is occasionally seen. Similarly, it is rare for the cavity of the right ventricle to be severely reduced in size, though a mild to moderate reduction of right ventricular cavity size relative to normal is most common. Abnormal connections between the epicardial coronary arteries and the right ventricular cavity (sinusoidal or coronary–cameral fistulas) are very rare in hearts with critical pulmonary valvar stenosis, in contrast to their relatively greater prevalence in hearts with pulmonary atresia and intact ventricular.

Clinical features

Neonates with critical pulmonary stenosis generally present within the first few days of life with irritability, tachypnea, and hypoxemia which is due to right to left shunting at the atrial level but will become extreme and be associated with metabolic acidosis if the ductus arteriosus begins to close. Radiographic examination of the chest reveals clear to hyperlucent lung fields and normal to large cardiac silhouette. Two-dimensional echocardiography is diagnostic, revealing the thickened, immobile pulmonary valve as well as defining the size of the right ventricular cavity and the degree of tricuspid valve regurgitation. Initial medical therapy consists of ensuring the patency of the ductus arteriosus, for example by infusion of prostaglandin E_1.

Surgical management

Relief of pulmonary valve stenosis is the goal of surgery, and this can be accomplished using a variety of adjunctive techniques. Closed pulmonary valvotomy, open valvotomy with temporary venous inflow occlusion, and open valvotomy with cardiopulmonary bypass with or without cardioplegia all give good results. A small increment in the effective cross-sectional area of the valve orifice results in significant augmentation of antegrade flow from the right ventricle to the pulmonary arteries. It is uncommon that a systemic to pulmonary artery shunt is required to augment pulmonary blood flow. In hearts where the right ventricular–pulmonary trunk junction (‘annulus') is very small, pulmonary valvotomy alone may be insufficient and a transannular patch may be required. In this uncommon situation, it is advisable to add a systemic to pulmonary artery shunt, as transannular patching alone may present a risk of severe hypoxia postoperatively as a result of an acute right ventricular dysfunction and resultant right to left shunting across a patent foramen ovale.

Percutaneous balloon valvotomy is an important non-surgical method of treatment of critical pulmonary stenosis in the neonate. This is accomplished in the cardiac catheterization laboratory and is with increasing frequency supplanting surgical valvotomy as the initial method of management of critical pulmonary stenosis. Since the vast majority of neonates with critical pulmonary stenosis can be stabilized by ensuring patency of the ductus arteriosus prior to palliative or definitive therapy, the hospital mortality for this group has in recent years been reduced to between 5 and 10 per cent.

Pulmonary stenosis in infants and children

The spectrum of pulmonary valve stenosis presenting after the neonatal period ranges from mild to moderate valvular obstruction that may remain relatively stable throughout life, to severe pulmonary stenosis associated with suprasystemic right ventricular pressure. The pulmonary valve is generally made up of two or three leaflets which are relatively well developed in contrast to those of the neonate with critical pulmonary stenosis. There may be complete or partial commissural fusion. The leaflet tissue is generally thicker than in normal valves, and the pulmonary artery wall may be pulled inward or tethered at the site of commissural cusp attachment. Some patients, particularly among those with Noonan's syndrome, have pulmonary valve dysplasia wherein the valve tissue itself is very thickened and the leaflets are quite rigid. Immobility of the leaflets due to thickening and rigidity, rather than commissural fusion, results in the significant pressure gradient between the right ventricle and the pulmonary arteries. Pulmonary valvular stenosis of any type may be accompanied by secondary infundibular hypertrophy resulting in infundibular stenosis.

Clinical features

About one-third of patients with valvular pulmonary stenosis and intact ventricular septum are asymptomatic when first evaluated. The diagnosis is suspected after auscultation of an ejection type of systolic murmur. Among those with symptoms, dyspnea on effort is the most common complaint. Cyanosis at rest or with exercise results from right to left shunting at the atrial level, which is the consequence of diminished compliance of the right ventricle relative to that of the left ventricle. Untreated valvular pulmonary stenosis may eventuate in the development of right ventricular failure. More importantly, infants in whom right ventricular pressure is elevated to near systemic levels at rest are at risk for sudden death. The electrocardiogram generally shows prominent P-waves associated with right atrial enlargement, and elevation of the height of the R-wave in the right precordial leads, together with incomplete right bundle branch block. Radiographic examination of the chest is non-specific, but often shows an enlarged cardiac silhouette due primarily to prominence of the right atrium. As in neonates with critical pulmonary stenosis, two-dimensional echocardiography is diagnostic. Cardiac catheterization may be undertaken for complete hemodynamic evaluation, but most importantly to provide access for balloon pulmonary valvotomy which is efficacious except in cases where the valve is severely dysplastic.

Surgical management

If relief of valvular pulmonic stenosis is not achieved at cardiac catheterization, it can be reliably accomplished by surgical means. Cardiopulmonary bypass with mild to moderate hypothermia is the most commonly used form of circulatory support. The pulmonary valve is generally exposed by means of a vertical arteriotomy in the proximal main pulmonary artery. Incision of fused commissures back to the level of the annulus often provides complete or near complete relief of valvular obstruction. When there is tethering of the wall of the main pulmonary artery to the commissural attachments, these should be incised. If the leaflets remain significantly immobile despite commissurotomy, as in instances of severe pulmonary valvular dysplasia, then partial or complete pulmonary valvectomy may be necessary. The resulting pulmonary insufficiency is well tolerated, as long as right ventricular function is not significantly compromised, there is a competent tricuspid valve, and the pulmonary vascular resistance is not significantly elevated. For patients with simple valvular pulmonary stenosis who present after the neonatal period, hospital mortality for relief of pulmonary stenosis by surgical means or by percutaneous balloon valvotomy approaches zero. The results are somewhat less uniform in patients with Noonan's syndrome, and in those with severe degrees of hypoplasia of the pulmonary annulus or of the right ventricular cavity.

Pulmonary atresia with intact ventricular septum

Pulmonary atresia with intact ventricular septum is a congenital heart malformation in which there is no luminal continuity between the right ventricular chamber and the pulmonary arteries. Hearts with pulmonary atresia and intact ventricular septum include a spectrum of morphology with regard to both the nature of the junction of the right ventricle and the pulmonary trunk, and the degree of development of the right ventricle and the tricuspid valve. Generally, the right and left pulmonary arteries and the main pulmonary artery are normal or near normal in size. Almost invariably, the ductus arteriosus is the exclusive source of pulmonary blood flow. Only rarely are major aortopulmonary collateral vessels present.

Most often, right ventricular cavity size is reduced and the wall is significantly hypertrophied. The right ventricle may have an infundibulum which ends blindly in a well developed but imperforate pulmonary valve plate. Alternatively, the atresia may be ‘muscular'. In such cases, there is no identifiable valve or valve-like tissue between the most distal portion of the right ventricle and the main pulmonary trunk. There is a high degree of correlation between the diameter of the tricuspid valve, and right ventricular cavity size. While right ventricular cavity size is less than normal in over 90 per cent of patients, those with normal or near normal size of the tricuspid annulus (and a concomitant moderate to significant degree of tricuspid incompetence) can be expected to have the most developed right ventricular chambers. This construction makes sense, as chamber development is undoubtedly related to the volume of blood flowing into and out of the right ventricle, despite atresia of the outflow tract. Another important aspect of morphology is the presence of coronary artery–right ventricular fistulas which are present in between one-third and one-half of patients with pulmonary atresia with intact ventricular septum. In the most extreme cases, there are intrinsic abnormalities such as stenoses or occlusion of the principal coronary arteries, resulting in a coronary circulation that is partially or completely right ventricular dependent.

Clinical features

Nearly all infants with pulmonary atresia and intact ventricular septum will present during the first few days of life with cyanosis. Progressive closure of the ductus arteriosus is associated with increasing hypoxemia and metabolic acidosis. Radiographic examination of the chest typically discloses clear lung fields with normal or diminished pulmonary vascular markings. The cardiac silhouette may be normal or enlarged. The electrocardiogram is not particularly helpful, as there may be electrocardiographic evidence of right ventricular hypertrophy despite diminished right ventricular cavity size. Two-dimensional echocardiography is diagnostic. This study should provide information regarding the dimensions of the tricuspid valve, the presence and degree of tricuspid insufficiency, right ventricular cavity size, and the nature of the right ventricular outflow obstruction. Frequently, right ventricular–coronary artery fistulas can be appreciated by two-dimensional echocardiography and colour flow mapping. Cardiac catheterization, however, is necessary to define with confidence the nature of the coronary circulation and the potential for its dependence upon the right ventricle and ventricular–coronary fistulas. Without treatment, patients with pulmonary atresia with intact ventricular septum succumb as a result of severe hypoxia and metabolic acidosis. Virtually all patients can be stabilized by maintenance of patency of the ductus arteriosus, for example with infusion of prostaglandin E_1

Surgical management

Choice of initial surgical procedure for patients with pulmonary atresia with intact ventricular septum depends upon the degree of development of the tricuspid valve and right ventricular cavity, and on the nature of the coronary circulation (including the presence or absence of right ventricular–coronary fistulas and the potential for a coronary circulation that is right ventricular dependent). Though not invariably true, it is generally the case that those patients with the larger tricuspid valves and right ventricular chamber sizes are less likely to have significant fistulas between the right ventricle and the coronary circulation. If the tricuspid valve and right ventricle are well developed, and decompression of the right ventricle does not pose a risk of coronary ischemia as a consequence of a steal from the coronary circulation via ventricular–coronary fistulas, then a procedure to establish antegrade flow from the right ventricle to the pulmonary artery trunk is undertaken. This may be incision or excision of the pulmonary valve plate if present, or establishment of continuity between the right ventricle and pulmonary artery trunk by means of a transannular patch or even a ventriculoarterial conduit. As right ventricular compliance remains markedly diminished for a period of time even after decompression, it is virtually always the case that patients having pulmonary atresia with intact ventricular septum will require an additional source of pulmonary blood flow at the time of initial surgery. This may be accomplished by construction of a systemic to pulmonary artery shunt, or by assuring prolonged patency of the ductus arteriosus, as for example by formalin infiltration of the wall of the ductus. This initial procedure may be accomplished with or without extracorporeal circulatory support. It is possible to place a transannular right ventricular outflow tract patch without the use of cardiopulmonary bypass, and to maintain patency of the ductus arteriosus by formalinization of its wall (Fig. 1). This formalinized ductus generally remains patent over a period ranging from several weeks to a few months. As the compliance of the right ventricle improves over time, the degree of right to left shunting at the atrial level through the foramen ovale gradually diminishes. With this operation, some patients achieve a ‘normal' circulation utilizing four cardiac chambers without any cardiac level shunts. Other patients require an additional procedure later to achieve additional relief of right ventricular outflow tract obstruction and to eliminate potential for right to left shunting at the atrial level. While we have been pleased with the results of the methodology described above (formalinization of the ductus arteriosus), many surgeons prefer to create an additional source of pulmonary blood flow at the time of pulmonary valvotomy by interposition of a 3.5 to 4 mm PTFE graft between a systemic artery and a pulmonary artery. In such cases, it is of course necessary to close the shunt at a later date. This can be accomplished at the time of a subsequent surgical procedure, or by occlusion of the shunt with coils or other devices in the cardiac catheterization laboratory.

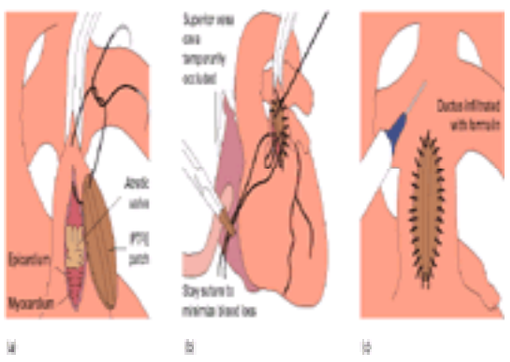


Fig. 1. Surgical management of pulmonary atresia with intact ventricular septum in a neonate. (a) The distal main pulmonary artery is occluded with a vascular clamp. Pulmonary blood flow is ensured by patency of the ductus arteriosus. A longitudinal incision is made over the proximal main pulmonary artery and extended onto the infundibulum of the right ventricle, at which level the incision is carried down only through the epicardium. Augmentation of the right ventricular outflow tract with a Teflon patch is begun. (b) After most of the circumference of the Teflon patch has been sutured in place, temporary vena cava occlusion is accomplished with a vascular clamp. Beneath the unsecured portion of the transannular patch, the scalpel is used to extend the incision through the myocardium of the right ventricular infundibulum and the pulmonary valve plate. Caval occlusion is discontinued and the remainder of the patch is secured with continuous suture. (c) Following completion of the right ventricular outflow tract reconstruction with transannular patch, formalin is infiltrated beneath the adventitia of the ductus arteriosus to ensure its continued patency.

Those patients with the smallest tricuspid valves and least well developed right ventricular chambers, and those in whom connections between the right ventricular chamber and coronary arteries together with intrinsic abnormalities of the coronary arteries result in a coronary circulation that is right ventricle dependent, are ultimately best served by a modification of Fontan's procedure. As such, their initial surgical management does not include decompression of the right ventricle, but rather they are helped by construction of a systemic to pulmonary artery shunt. After a period of maturation, when the pulmonary vascular resistance has fallen, they undergo a modification of Fontan's operation in one or two stages.

For patients with pulmonary atresia and intact septum, palliative or definitive surgery is required during the neonatal period. Without surgical treatment, about 50 per cent of babies die within the first month of life and nearly all die before 12 months of age. Construction of a systemic to pulmonary artery shunt alone is not associated with growth of the tricuspid valve and right ventricular chamber, and thus should be reserved for those patients who will ultimately be managed by means of a modification of Fontan's procedure. For those with a right ventricle that is judged to have the potential for eventually managing all of the systemic venous return, a procedure to optimize antegrade flow from the right ventricle to the main pulmonary artery trunk (such as a transannular outflow tract patch) should be combined with a procedure to provide an additional, but temporary, source of pulmonary blood flow, such as formalinization of the ductus or construction of a Blalock–Taussig shunt. In a multi-institution study conducted by the Congenital Heart Surgeons' Society, survival was 81 per cent at 1 month after the first intervention and 64 per cent at 4 years. This group of patients represented the entire spectrum of right ventricular morphology, and they were managed with a variety of treatment plans.

Congenital aortic stenosis

Congenital valvular aortic stenosis is a form of left ventricular outflow tract obstruction caused by abnormal development of the aortic valve cusps and the aortic annulus. Critical aortic valve stenosis of the neonate is a condition in which the severity of obstruction at the level of the aortic valve results in a systemic circulation that

is dependent upon flow from the right ventricle, via the ductus arteriosus. More commonly, valvular aortic stenosis becomes evident after the newborn period. The majority of patients with stenosis severe enough to warrant operative intervention in infancy and childhood have a bicuspid aortic valve. The larger of the two cusps frequently contains a central thickened ridge or raphe representing the rudimentary commissure between two fused cusps. There is usually fusion peripherally at one or both commissures. In about 30 per cent of cases, the valve is tricuspid, with thickening of the leaflets and variable degrees of fusion at the commissures. Rarely, the valve has a unicuspid configuration with only one commissure. With severe valvular aortic stenosis, there is virtually always concentric hypertrophy of the left ventricle. Some neonates present with a very extreme degree of hypertrophy and with fibrosis of the endocardium and subendocardium. Endocardial fibroelastosis may be present as a result of ischemia of the developing myocardium. In some instances, severe aortic valve stenosis in the neonate may be associated with varying degrees of stenosis or hypoplasia of the mitral valve, and a smaller than usual left ventricular cavity. These cases should be considered part of the spectrum of hypoplastic left heart syndrome.

Clinical features

Neonates with critical aortic stenosis present with poor systemic perfusion, pallor, tachypnea, tachycardia, and a mild degree of cyanosis. The murmur of aortic stenosis may be unimpressive because of diminished flow from the left ventricle. Spontaneous narrowing of the ductus arteriosus is associated with a profound hypoperfusion state and metabolic acidosis, and progresses rapidly to death unless supportive therapy to re-establish patency of the ductus arteriosus is undertaken (i.e. infusion of prostaglandin E_1).

Clinical signs of aortic stenosis presenting in older infants and children include an ejection systolic murmur at the base of the heart radiating to the carotid vessels and accompanied by a systolic ejection click. There may be a third or fourth heart sound present. The electro-cardiogram usually shows severe left ventricular hypertrophy and there may be a left ventricular strain pattern as well. In neonates and infants with severe aortic stenosis, radiographic examination of the chest usually reveals increase in heart size. This may not be so in older children, though left ventricular prominence and prominence of the ascending aorta may be present.

Two-dimensional echocardiography is diagnostic. In neonates with critical aortic stenosis and in infants with severe congestive heart failure, the Doppler derived estimate of the left ventricular outflow tract gradient is of little importance, as flow may be significantly diminished. In neonates and infants with severe valvular aortic stenosis, cardiac catheterization adds little diagnostic information. However, percutaneous balloon valvotomy is increasingly being utilized as the initial treatment of many patients in this category. Access is achieved by means of either the femoral, umbilical, or right carotid artery.

Surgical management

Surgical valvotomy can be accomplished in neonates and infants using a variety of techniques. As with critical valvular pulmonary stenosis, a relatively small increment in the effective valve orifice can result in significant diminution in the degree of obstruction. It is of course of paramount importance in valvular aortic stenosis to achieve relief of obstruction while creating a minimum amount of incompetence of the valve. Valvotomy can be effectively accomplished under direct vision via an incision in the ascending aorta during cardiopulmonary bypass with or without cardioplegic arrest. Alternatively, valvotomy can be accomplished via an incision in the ascending aorta during a brief period of inflow stasis, or ‘blindly' by passing dilators through the apex of the left ventricle and advancing them antegrade through the aortic annulus. Using these various techniques, mortality rates as low as 10 per cent and as high as 65 per cent have been reported for surgical palliation of neonates.

Aortic valvotomy in older children is generally accomplished using cardiopulmonary bypass with a moderate degree of hypothermia and aortic cross-clamping with cardioplegic myocardial protection. The goal of the operation is to minimize the degree of obstruction at the level of the aortic valve without creating a significant degree of insufficiency. This goal is usually accomplished by incision of fused commissures out to the level of the annulus, and often by removal of nodular and fibrous excrescenses from the valve leaflets themselves in order to enhance their mobility. One cannot overemphasize the importance of avoiding incision of the raphe in the congenitally bicuspid aortic valve, as this is certain to result in a significant degree of aortic insufficiency. Reintervention is likely to be necessary in the majority of patients who undergo aortic valvotomy early in childhood. In a long-term study performed at the National Institutes of Health (excluding those who underwent initial valvotomy as neonates and infants), about 90 per cent of children and young adults were free from reintervention for at least 10 years after the initial operation. The actuarial estimate of freedom from reintervention by 20 years after the initial operation was 60 per cent, and by 40 years was only 10 per cent. Reintervention appears to be required earlier and more frequently when the initial valvotomy has been performed during the neonatal or infant period.

Aortic atresia

Aortic atresia is the developmental anomaly in which the aortic valve is imperforate. This is virtually always associated with a hypoplastic ascending aorta. The systemic circulation is dependent upon flow from the right ventricle through the ductus arteriosus into the aorta. Flow in the ascending aorta is retrograde and includes only the flow to the coronary arteries. In the vast majority of cases, aortic atresia is associated with atresia or hypoplasia of the mitral valve, and absence or severe hypoplasia of the left ventricle. This is known as hypoplastic left heart syndrome, and is the fourth most common congenital cardiac anomaly presenting in the first year of life. It is the most common cause of death due to cardiac disease in infancy.

In rare instances, aortic atresia is associated with an unrestrictive ventricular septal defect. In these cases, normal left ventricular development can occur. Thus, a primary myocardial abnormality is unlikely to be the cause of hypoplastic left heart syndrome.

Clinical features

Physical examination typically shows a mildly cyanotic infant with tachypnea and tachycardia. Depending on the degree of patency of the ductus arteriosus at the time of evaluation, peripheral pulses may be normal, diminished, or absent. Progressive closure of the ductus arteriosus is associated with a profound hypoperfusion state leading rapidly to death. Re-establishment of patency of the ductus arteriosus by means of an infusion of prostaglandin E_1 is life saving. The electrocardiogram generally shows right atrial enlargement and right ventricular hypertrophy. Radiographic examination of the chest is non-specific. Cardiomegaly is common and pulmonary vascular markings may be prominent. Two-dimensional echocardiography is diagnostic in this lesion. Cardiac catheterization may be destabilizing to the critically ill neonate and adds little if any important physiological information.

Surgical management

Being the most common cardiac anomaly associated with a single dominant ventricle, aortic atresia or hypoplastic left heart syndrome lends itself ultimately to reconstruction by means of a modification of Fontan's operation. This, however, cannot be accomplished in the neonate because of the prohibitively high pulmonary vascular resistance. Initial surgical palliation is embodied by three basic principles. (1) The aorta must be associated directly with the right ventricle in a fashion guaranteeing unobstructed flow from right ventricle to the systemic circulation, with growth potential obviating further aortic surgery. (2) Pulmonary blood flow must be regulated for proper growth and development and maturation of the pulmonary vasculature, in order to avoid the development of pulmonary vascular obstructive disease and to minimize the volume load on the right ventricle. (3) A large interatrial communication is necessary to avoid pulmonary venous hypertension and its consequences.

The initial palliative procedure, as described by Norwood, is accomplished utilizing deep hypothermia and circulatory arrest. Extracorporeal circulation is established with a bypass loop between the right atrium and the main pulmonary artery. The right and left branch pulmonary arteries are occluded with tourniquets and patency of the ductus arteriosus is maintained thus assuring adequate systemic perfusion. The branch vessels of the aortic arch are occluded with tourniquets in preparation for circulatory arrest. After removal of the right atrial cannula, septum primum is excised, ensuring a large and unrestrictive interatrial communication. The main pulmonary artery is transected immediately proximal to the origin of the right pulmonary artery. The distal main pulmonary artery is oversewn with a patch to ensure continuity between the right and left pulmonary artery branches. The ductus arteriosus is ligated and transected at its entrance into the thoracic aorta. The incision is extended distally onto the thoracic aorta and proximally into the aortic arch and ascending aorta. The entire aortic arch is augmented with a gusset of cryopreserved pulmonary artery homograft. Proximally, the transected main pulmonary artery is anastomosed to the ascending aorta and the homograft gusset is then sewn to the remainder of the proximal main pulmonary artery. Pulmonary blood flow is provided by a systemic to pulmonary artery shunt. This may be a central shunt placed between the reconstructed aorta and the confluence of the pulmonary arteries, or a modified Blalock–Taussig shunt interposed between the innominate artery and right pulmonary artery (Fig. 2). Postoperatively, the critical balance between systemic and pulmonary blood flow is maintained through judicious control of ventilatory parameters, as the pulmonary vascular resistance is exquisitely sensitive to alterations in carbon dioxide tension and pH.

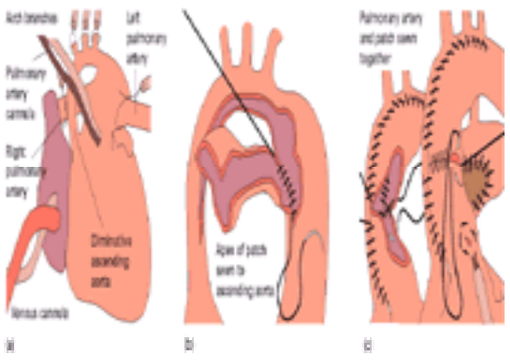


Fig. 2. Palliative reconstructive surgery for hypoplastic left heart syndrome. (a) Extracorporeal circulation is established with a bypass loop between the right atrium and main pulmonary artery. The right and left branch pulmonary arteries are occluded with tourniquets to ensure adequate systemic perfusion via the ductus arteriosus. During hypothermic circulatory arrest, the branch vessels of the aortic arch are occluded with tourniquets. (b) After the ductus arteriosus has been transected at its entrance into the thoracic aorta, the incision is extended distally on to the thoracic aorta and proximally into the aortic arch and diminutive ascending aorta. The entire aortic arch is augmented with a gusset of cryopreserved pulmonary artery homograft. (c) The transected proximal main pulmonary artery is anastomosed to the ascending aorta. The arch reconstruction is then completed by anastomosis of the homograft gusset to the remainder of the proximal main pulmonary artery. In this case, a central shunt is interposed between the augmented aorta and the confluence of the pulmonary arteries.

Currently, hospital survival following initial palliative surgery for hypoplastic left heart syndrome is at a level of 70 to 80 per cent at selected centers. The initial palliative procedure is characterized by pulmonary and systemic circulations in parallel, and thus an obligatory volume load on the functional single ventricle. As it is desirable to reduce the volume load on the single ventricle as early in life as possible, the conversion to the Fontan-type connection has been divided into two stages. In the middle of the first year of life, the hemi-Fontan procedure is accomplished. The superior vena cava is associated with the branch pulmonary arteries, and the systemic to pulmonary shunt is closed. Thus the volume work of the single ventricle is reduced to that of the systemic circulation only. Later, usually at about 2 to 3 years of age, the Fontan procedure is completed. Return from the inferior vena cava is directed to the pulmonary arteries either by means of a lateral atrial tunnel or with use of an extra cardiac conduit.

While staged reconstructive surgery for aortic atresia has gained increasingly widespread acceptance in the last decade, some centers still choose neonatal heart transplantation as initial and definitive treatment for this challenging lesion. The Congenital Heart Surgeons' Society in North America reported the intermediate term results of a multi-institutional study enrolling 323 neonates with aortic atresia. Results varied considerably between participating institutions. The best results were obtained at four of the high volume centers. At these four institutions, survival at 1, 3, 12, 24, and 36 after entry was 77, 70, 64, 62, and 61 respectively. Survival among the four institutions was similar, though two institutions used a staged reconstructive surgery protocol and two of the institutions used a heart transplantation protocol.

Further reading

Bailey LL *et al.* Orthotopic transplantation during early infancy as therapy for incurable congenital heart disease. *Annals of Surgery* 1988; **208**: 279. [This seminal publication describes the methodology and results of neonatal heart transplantation at one of the institutions which pioneered this form of therapy as definitive management for aortic atresia and other forms of hypoplastic left heart syndrome.]

DeBoer DA, Robbins RC, Maron BJ, McIntosh CL, Clark RE. Late results of aortic valvotomy for congenital valvular aortic stenosis. *Annals of Thoracic Surgery* 1990; **50**: 69. [This study, undertaken at the National Institutes of Health, describes the long-term follow-up of a cohort of patients who underwent surgical aortic valvotomy early in childhood.]

Hanley FL, Sade RM, Blackstone EH, Kirklin JW, Freedom RM, Nanda NC. Outcomes in neonatal pulmonary atresia with intact ventricular septum. *Journal of Thoracic and Cardiovascular Surgery* 1993; **105**: 406–27. [Results of the Congenital Heart Surgeons' Society multi-institutional study of neonates with pulmonary atresia and intact ventricular septum. The prognostic importance of morphologic features such as size of the tricuspid valve annulus and presence of coronary–cameral fistulas is included.]

Hanley FL, Sade RM, Freedom RK, Blackstone EH, Kirklin JW, and the Congenital Heart Surgeons' Society: Outcomes in critically ill neonates with pulmonary stenosis and intact ventricular septum: a multi-institutional study. *Journal of the American College of Cardiology* 1993; **22**: 183–92. [This publication describes the outcomes of a multi-institutional study of neonates with pulmonary stenosis.]

Jacobs ML, Norwood WI, eds. Hypoplastic left heart syndrome. In: *Pediatric cardiac surgery: current issues*, pp. 182–92. Butterworth-Heinemann, Boston, 1992. [This monograph provides an in-depth discussion of the surgical management of hypoplastic left heart syndrome.]

Jacobs ML, Blackstone EH, Bailey LB. Intermediate survival in neonates with aortic atresia. *Journal of Thoracic and Cardiovascular Surgery* 1998; in press. [The initial report of outcomes in the Congenital Heart Surgeons' Society multi-institutional study of 323 neonates with aortic atresia.]

Norwood WI, Lang P, Castaneda AR, Campbell DN. Experience with operations for hypoplastic left heart syndrome. *Journal of Thoracic and Cardiovascular Surgery* 1981; **82**: 511–19. [An early report from one of the principle pioneers in surgical management of aortic atresia, outlining the evolution of the approach to initial surgical palliation.]

Norwood WI, Jacobs ML, Murphy JD. Fontan's procedure for hypoplastic left heart syndrome. *Annals of Thoracic Surgery* 1992; **54**: 1025–30. [Results of a large institutional experience with patients with aortic atresia and other forms of hypoplastic left heart syndrome managed by a modification of Fontan's procedure.]

Turley K, Bove EL, Iannettoni M, Yeh J, Cotroneo JV, Galdieri RJ. Neonatal aortic stenosis. *Journal of Thoracic and Cardiovascular Surgery* 1990; **99**: 679. [A pooled experience with surgical intervention in neonates with valvar aortic stenosis. This report provides a basis for comparison with results of balloon dilation of the aortic valve in neonates.]

Zeevi B, Keane JF, Fellow KE, Lock JE. Balloon dilation of critical pulmonary stenosis in the first week of life. *Journal of the American College of Cardiology* 1988; **11**: 821. [Report of a large institutional experience with percutaneous balloon dilation of valvar stenosis in neonates.]

40.7.7 Transposition of the great arteries

Lester C. Permut and Hillel Laks

Morphology
Pathophysiology
Simple transposition
Transposition of the great arteries with ventricular septal defect
Transposition of the great arteries with ventricular septal defect and left ventricular outflow tract obstruction
Natural history
Diagnosis
History and physical examination
Electrocardiogram
Chest radiograph
Cardiac catheterization
Management
Simple transposition of the great arteries
Transposition of the great arteries with ventricular septal defect
Transposition of the great arteries, ventricular septal defect, and left ventricular outflow tract obstruction
Summary
Severe right ventricular failure following atrial switch procedures
Congenitally corrected transposition of the great arteries
Morphology
Pathophysiology
Diagnosis
Management
Further reading

Transposition of the great arteries is a congenital heart defect in which the aorta arises from the right ventricle and the pulmonary trunk arises from the left ventricle. This results from abnormal septation of the truncus arteriosus during the third to fourth weeks of fetal development. Transposition occurs in 19.3 to 33.8 per 100 000 live births, and is the most common cardiac cause of neonatal cyanosis. It is more common in males than in females.

Morphology

Transposition of the great arteries may be classified according to the presence or absence of associated cardiac abnormalities. Simple transposition refers to cases in which no other hemodynamically significant defects are present. It can also be described as transposition with intact ventricular septum, and is found in 60 per cent of patients.

Complex transposition of the great arteries indicates the coexistence of associated hemodynamically significant congenital heart defects. The two most common are ventricular septal defect, which occurs in 28 per cent of patients with transposition of the great arteries, and ventricular septal defect with left ventricular (subpulmonic) outflow tract obstruction, occurring in 12 per cent of patients with transposition of the great arteries.

The position of the great vessels in transposition of the great arteries may vary considerably, but most often the aorta arises anteriorly from the morphologically normal right ventricle, slightly to the right of the main pulmonary trunk. The pulmonary artery usually arises posteriorly and to the left of the aorta from the morphologically normal left ventricle. The relationship of the aorta to the pulmonary trunk is noted by the designation dextro- or d-transposition of the great arteries. The aorta may rarely arise to the left of the pulmonary trunk (l-transposition of the great arteries) or directly posterior to it. Thus, transposition refers only to the discordance between the great vessels and the ventricles from which they arise and not to the anatomic and spatial relationships between the great vessels themselves.

The coronary arteries in transposition of the great arteries arise from the aorta, but variability in their take-off and course is common. This becomes important in the planning of surgical repair. The most frequent patterns are illustrated in Fig. 1.

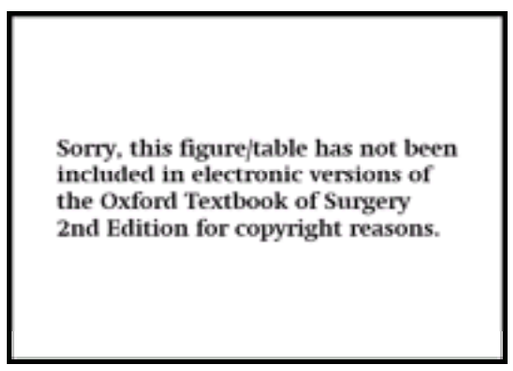


Fig. 1. Coronary artery patterns in patients with transposition of the great arteries.

Pathophysiology

Simple transposition

The basic pathophysiologic abnormality in simple transposition of the great arteries is that the pulmonary and systemic circulation are aligned in parallel rather than in series, as outlined schematically in Fig. 2. Desaturated systemic venous blood returns to the right atrium and right ventricle, but is then directed back to the systemic circulation through the aorta. Similarly, oxygen-rich pulmonary venous blood enters the left atrium and left ventricle where it is pumped, via the pulmonary trunk, to the lungs. The failure of oxygenated blood to reach the systemic circulation results in cyanosis, severe acidosis, and myocardial ischemia. Mixing of blood through a patent foramen ovale and patent ductus arteriosus may attenuate these changes and allow survival beyond the perinatal period. They also result in increased pulmonary blood flow, which may produce pulmonary congestion and the subsequent development of pulmonary vascular disease.

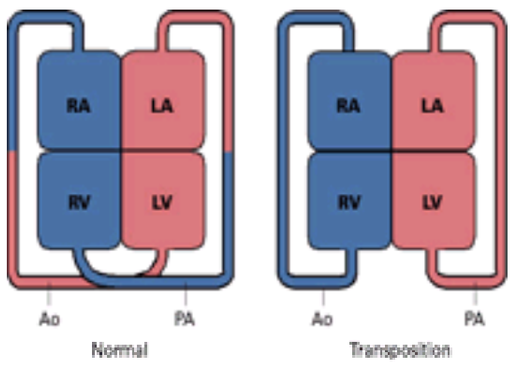


Fig. 2. Diagram of the circulation in a normal heart and in transposition of the great arteries. RA, right atrium; RV, right ventricle; LA, left atrium; LV, left ventricle; PA,

pulmonary artery; Ao, aorta.

Transposition of the great arteries with ventricular septal defect

The presence of a ventricular septal defect will alter the pathophysiologic effects outlined above. As pulmonary vascular resistance decreases normally after birth, systemic venous blood will be shunted from the right ventricle to the left ventricle across the septal defect. Thus, some unoxygenated blood will pass through the pulmonary circulation and return to the left atrium. Increased left atrial blood flow results in dilatation of the left atrium and stretching of the atrial septum, thereby enlarging the foramen ovale. The net result is an increased mixing of systemic and pulmonary venous blood with consequent lesser degrees of cyanosis and acidosis. The increased pulmonary blood flow may, however, cause congestive cardiac failure and induce the subsequent development of pulmonary vascular disease.

Transposition of the great arteries with ventricular septal defect and left ventricular outflow tract obstruction

This combination of defects, depending on the severity of the obstruction, may result in diminished pulmonary artery blood flow and shunting of blood across the ventricular septal defect from the left ventricle to the right ventricle. The degree of obstruction may progress rapidly in the first few hours after birth, and a marked diminution in pulmonary blood flow may occur. Simultaneously, as right ventricular volume and pressures increase, right atrial enlargement develops, leading to right-to-left shunting through the foramen ovale. The net result is marked desaturation of left atrial and left ventricular blood with consequent severe cyanosis and acidosis.

Natural history

Survival of patients with transposition of the great arteries depends on mixing of blood between the systemic and pulmonary circulations. In simple transposition, this must occur through a patent foramen ovale and/or a patent ductus arteriosus. Failure of adequate mixing after birth results in death in the first 24 h of life. Physiologic closure of the foramen ovale and ductus results in a 6-month mortality of 90 per cent in untreated infants. Patients with transposition of the great arteries and ventricular septal defect may survive into early childhood, but without treatment, pulmonary vascular disease and pulmonary hypertension develop. If left untreated, 34 per cent of patients with simple transposition of the great arteries and 78 per cent of patients with transposition of the great arteries and ventricular septal defect develop irreversible pulmonary vascular disease which renders them inoperable by 12 months of age. The onset of pulmonary hypertension is associated with rapid deterioration and subsequent death. Patients with transposition of the great arteries, ventricular septal defect, and left ventricular outflow tract obstruction have the worst prognosis. Rapidly worsening cyanosis occurs early in the postnatal period, resulting in poor survival without correction.

Diagnosis

History and physical examination

Cyanosis at or shortly after birth is commonly the initial finding in patients with transposition of the great arteries. The differential diagnosis thus includes other causes of cyanosis at birth including truncus arteriosus, tetralogy of Fallot, tricuspid atresia, and total anomalous pulmonary venous return. Transposition of the great arteries with ventricular septal defect is usually associated with less cyanosis, whereas transposition of the great arteries with ventricular septal defect and left ventricular outflow tract obstruction is associated with more severe cyanosis. In simple transposition, the only cardiac finding on physical examination may be a single accentuated second heart sound resulting from aortic valve closure anteriorly near the chest wall. Patients with associated ventricular septal defect, with or without subpulmonic stenosis, also have a holosystolic murmur. If the flow across the ventricular septal defect is large, a thrill may also be present.

Electrocardiogram

Electrocardiographic findings are generally limited to right atrial and right ventricular hypertrophy. Patients with transposition of the great arteries and ventricular septal defect, with or without left ventricular outflow tract obstruction, may have both left and right ventri-cular hypertrophy, but this is uncommon early in infancy.

Chest radiograph

Three principal radiologic changes are found in patients with transposition of the great arteries. Cardiomegaly, often with a characteristic oval-shaped heart, results from enlargement of the right ventricle which supports the systemic circulation. Narrowing of the superior mediastinum is seen as a result of the abnormal anteroposterior relationship of the aorta and pulmonary trunk. Pulmonary vascular markings are increased due to the increase in pulmonary blood flow (Fig. 3). Similar findings occur in transposition of the great arteries with ventricular septal defect. Cardiomegaly and mediastinal narrowing are accompanied by diminished pulmonary vascular markings in patients with transposition, ventricular septal defect, and left ventricular outflow tract obstruction.



Fig. 3. Chest radiograph of a patient with simple transposition of the great arteries demonstrating ‘egg-shaped' heart, narrow superior mediastinum, and increased pulmonary vascular markings.

Cardiac catheterization

Cardiac catheterization and ventriculography have been the definitive means of diagnosing transposition of the great arteries. Origin and position of the great arteries, the presence or absence of an associated ventricular septal defect, and the presence or absence of subpulmonic stenosis can be demonstrated with relative ease (Fig. 4, Fig. 5). Additionally, aortic root injection or selective coronary arteriography can delineate details of the coronary artery anatomy. In neonates, the craniocaudal projection gives a clear image of the coronary artery origins and distribution.

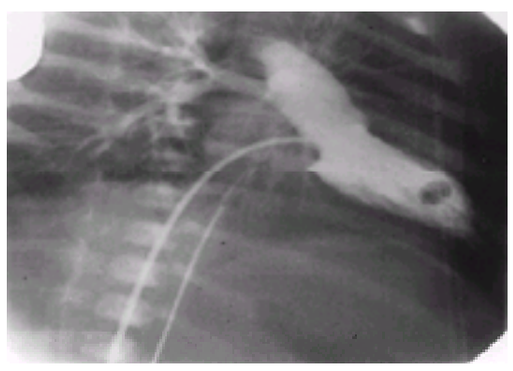


Fig. 4. Right ventriculogram in transposition of the great arteries demonstrating the aorta arising from the morphologically right ventricle.

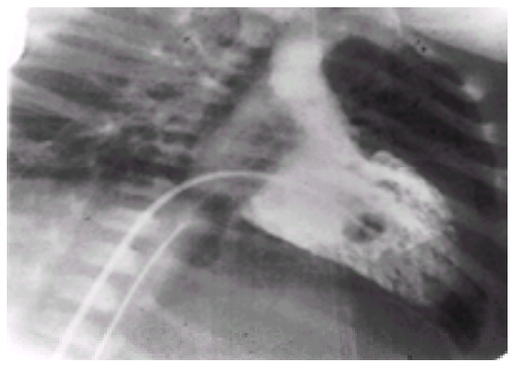


Fig. 5. Left ventriculogram in transposition of the great arteries demonstrating the pulmonary artery arising from the morphologically left ventricle.

Echocardiography

Two-dimensional Doppler echocardiography has replaced cardiac catheterization as the means of definitive diagnosis in patients with transposition of the great arteries in many institutions. This non-invasive technique can be performed easily at the bedside and obviates the need for vascular access and the use of nephrotoxic contrast materials. Echocardiography readily demonstrates the abnormally related great arteries, and the presence of ventricular septal defects, as well as the presence and degree of subpulmonic stenosis. Views through the aortic root can also demonstrate the anatomy of the proximal coronary arteries (Fig. 6). Echocardiography may be used as the sole method of preoperative assessment in most patients with simple transposition of the great arteries and transposition of the great arteries with ventricular septal defect. Catheterization is reserved for patients with more complex anatomy, or those in whom echocardiographic assessment is deemed technically inadequate.

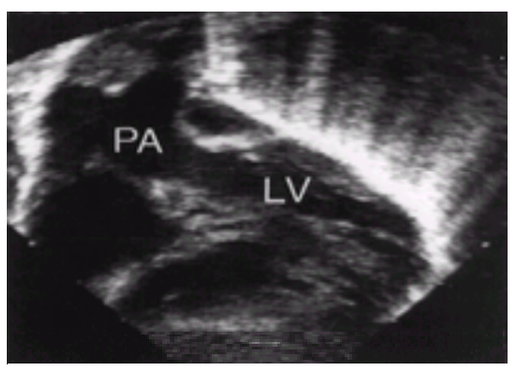


Fig. 6. Echocardiogram in transposition of the great arteries demonstrating the pulmonary artery arising from the morphologically left ventricle. LV, left ventricle; PA, pulmonary artery.

Management

Appropriate palliative and definitive treatment of transposition of the great arteries varies with the specific pathophysiologic derangements present.

Simple transposition of the great arteries

Palliation

The immediate therapeutic concern, once the diagnosis of transposition of the great arteries has been made, is to allow mixing of blood from the systemic and pulmonary circulations, thereby improving systemic oxygen saturation. The first method employed was described by Blalock and Hanlon in 1950 and involved surgical excision of a portion of the atrial septum without the use of cardiopulmonary bypass. In 1966, Miller and Rashkind revolutionized the management of these patients by introducing the non-surgical catheter technique of balloon atrial septostomy. A balloon catheter is passed across the atrial septum through the patent foramen ovale from right to left. The balloon is inflated and the catheter is forcibly withdrawn back into the right atrium, tearing the atrial septum. The procedure can be performed at the time of cardiac catheterization under fluoroscopic guidance, or at the bedside under echocardiographic guidance. The resultant large atrial septal defect almost always allows adequate mixing, and balloon septostomy has become the mainstay of initial management in transposition of the great arteries. Blalock–Hanlon atrial septectomy may rarely be required in patients who do not respond to balloon septostomy. Prostaglandin E_1 infusion is used to maintain patency of the ductus arteriosus during initial evaluation and therapy, but should not be considered more than a temporary measure. Severe metabolic acidosis may require treatment with sodium bicarbonate, although correction of systemic oxygen desaturation is the primary therapy for acidosis. Supportive therapy may include the use of digitalis and diuretics to treat ventricular failure.

Definitive treatment

The surgical management of simple transposition of the great arteries has undergone a rapid evolution in the past decade. The basic principle of definitive treatment is the redirection of blood flow such that the systemic and pulmonary circulations are placed in series rather than in parallel. This is accomplished by ‘switching' blood flow either within the atria or at the level of the great arteries.

Senning, in 1959, described the first technique for definitive repair of transposition of the great arteries by ‘switching' blood flow at the atrial level. The wall of the right atrium and the translocated atrial septum are used to create an intra-atrial baffle. The procedure is illustrated in Fig. 7. The right atrium is opened parallel to the interatrial groove, just anterior to the crista terminalis. The incision is extended superiorly to the base of the right atrial appendage and inferiorly toward the valve of the inferior vena cava. The left atrium is opened by dissection and incision of the interatrial groove, with a counterincision in the right superior pulmonary vein. The atrial septum is exposed through the right atriotomy, and a septal flap is created with its base positioned laterally at the interatrial groove. The flap is then rotated posteriorly and sutured to the wall of the left atrium posterior to the left atrial appendage and mitral valve and anterior to the left pulmonary veins so as to cover the orifices of the pulmonary veins. The posterior edge of the right atriotomy is next sutured to the right atrial wall anterior to the superior vena cava, and then to the remnant of the atrial septum. Inferiorly, the edge of the atrial baffle is sutured to the valve of the inferior vena cava and the right atrial wall just posterior to the coronary sinus. This baffle is thus used to divert blood from the orifices of the inferior and superior vena cavae to the mitral valve. The anterior edge of the right atriotomy is then anastamosed to the posterior edge of the left atriotomy, completing the repair.

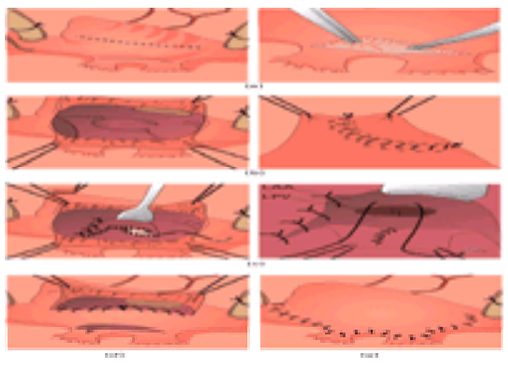


Fig. 7. Senning procedure. (a) Right atrial incision. Inset demonstrates dissection of interatrial groove. (b) Creation of atrial septal flap and augmentation with pericardium. (c) Suture of atrial flap over pulmonary vein orifices. (d) Creation of baffle from the vena cavae to the mitral valve and incision of left atrium. (e) Suture of the anterior right atrial wall to pulmonary veins, baffling blood from the pulmonary veins to the tricuspid valve. LAA, left atrial appendage; LPV, left pulmonary veins.

In 1964, Mustard described an alternative method of atrial switch which was technically simpler than that developed by Senning (Fig. 8). Following right atriotomy and complete atrial septectomy, a baffle of pericardium (or synthetic graft material such as Dacron or Goretex) is sutured to the atrial wall to divert flow from the caval orifices to the mitral valve. Pulmonary venous return flows lateral to the baffle into the tricuspid orifice, and then to the right ventricle and aorta.

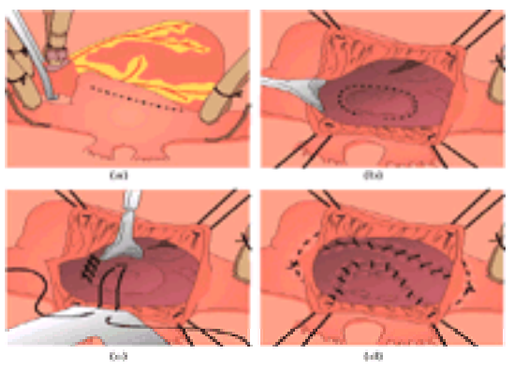


Fig. 8. Mustard procedure. (a) Right atrial incision. (b) Incision line for excision of atrial septum. (c) Initiation of baffle suture line. (d) Completed baffle directing blood flow from the vena cavae to the mitral valve.

Until about 1985, atrial switch using either the Senning or Mustard procedures was the technique of choice for defintive repair of transposition of the great arteries. These operations were performed in patients 6 weeks of age or older, and palliation with balloon atrial septostomy and occasionally atrial septectomy was required to allow survival from birth to operation. The major decision facing the surgeon at that time concerned choosing between Senning's and Mustard's technique. Although Senning was the first to report successful repair in 1959, other centers failed to duplicate these reults, largely owing to the technical difficulty of this procedure at a time when cardioplegia was not used and when open heart procedures generally had a high mortality rate. This prompted the development of Mustard's simplified operation in 1964. The Mustard procedure, which has been performed with a 1 per cent early mortality rate and a 5 per cent late mortality rate, was widely adopted for the treatment of transposition of the great arteries. However, it soon became recognized that there was a significant late morbidity related to the procedure. Superior vena cava obstruction occurred in up to 10 per cent of patients and pulmonary vein obstruction, usually left sided, occurred in 5 per cent. Many centers, therefore, returned to the Senning procedure as the technique of choice. The Senning procedure uses only autologous atrial tissue, which will theoretically grow with the patient, preventing subsequent obstruction of the pulmonary venous return. Recent reports indicate comparably low early and late mortality after either procedure.

Despite these excellent survival data, a number of major concerns remained about switching blood flow at the atrial level, regardless of the specific technique employed. First, atrial arrhythmias have been noted in up to 40 per cent of patients after Senning or Mustard procedures, and 10 per cent require pacemaker insertion by 10 years. Loss of the atrial kick may seriously impair cardiac output, and treatment regimens including antiarrhythmic agents and pacemakers add significantly to patient morbidity. A second concern regards the ability of an anatomic right ventricle to support the systemic circulation after an atrial switch. The bellows-like morphology of the right ventricle is ideally suited to pump high volumes at low pressures. Although short-term results have been good, doubts regarding the long-term capabilities of a systemic right ventricle are now being confirmed. Right ventricular dysfunction and failure is reported in 3 to 10 per cent of long-term survivors, and tricuspid regurgitation (an indicator of right ventricular pressure and volume overload) is present in 4 per cent. Exercise tolerance is also diminished in these patients. Finally, the necessary interval between birth and the atrial switch procedure is associated with a mortality of 5 to 10 per cent, which must be added to the early operative mortality rate in comparing this with other procedures.

Switching blood flow at the arterial level was initially attempted in the 1950s by Mustard. Success of this ‘anatomic' repair for transposition of the great arteries was dependent on two principal factors: the ability to transpose the coronary arteries to the ‘neo'-aorta and the ability of the left ventricle, previously connected to the low-pressure pulmonary circulation, to maintain the systemic circulation. After his initial attempts, Mustard abandoned this operation primarily because of the technical difficulties imposed by the coronary anastomoses. While certain patterns of coronary anatomy make arterial switch difficult or impossible, improvements in microvascular surgical techniques as well as the concerns regarding long-term results of the atrial switch prompted renewed interest in this procedure. Jatene reported the first successful operation in a 42-day-old child with transposition of the great arteries and ventricular septal defect in 1975. The volume load from the ventricular septal defect ‘prepared' the left ventricle for support of the systemic circulation. Yacoub reported use of the arterial switch for transposition of the great arteries with intact septum after preliminary preparation with a pulmonary artery band and systemic-to-pulmonary shunt. Eber reported the first small series of neonatal arterial switch procedures and Castaneda reported the first successful large series of arterial switches in the neonate. The technique is illustrated in Fig. 9. The aorta and pulmonary trunk are divided after careful assessment of the coronary anatomy. The coronary ostia are excised with a surrounding button of aortic wall and are reimplanted into the adjacent neo-aorta, care being taken to avoid angulation, kinking, or tension. The neo-pulmonary artery, from which the coronary buttons have been removed, is now repaired using a W-shaped pericardial patch. Early techniques included a graft of pericardium, dura mater, or synthetic material to bridge the proximal and distal pulmonary artery. Currently, mobilization of the branch pulmonary arteries and displacement of the pulmonary trunk bifurcation anterior to the aorta, as described by Lecompte, allows tension-free direct anastomosis of the proximal and distal pulmonary artery. Closure of the atrial septal defect is accomplished during a brief period of circulatory arrest.

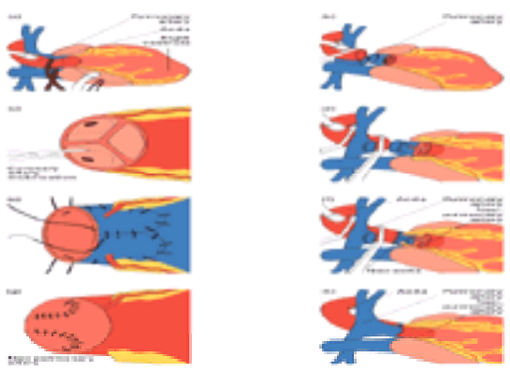


Fig. 9. Arterial switch procedure. (a) Site of aortic transection. (b) Transection of aorta and pulmonary trunk. (c) Excision of coronary arteries with button of aortic wall. (d) Pulmonary artery transposed anterior to the aorta. Incisions in ‘neo-aorta' at sites of coronary transfer. (e) Anastomosis of coronary artery to neo-aorta. (f) Anastomosis of neo-aorta to aortic arch. (g) Pericardial patches repairing defects in neo-pulmonary artery. (h) Anastomosis of neo-pulmonary artery to distal pulmonary trunk. Ao, aorta; PA, pulmonary artery.

Proper timing of the operation is essential for a successful outcome after the arterial switch, largely due to changes in the pulmonary vascular resistance. Before birth, the pulmonary vascular resistance is high and right and left ventricular size and wall thickness are similar. The normal postnatal decrease in pulmonary vascular resistance results in continued growth and hypertrophy of the systemic ventricle, with a relative decrease in the wall thickness of the pulmonary ventricle. Patients with simple transposition of the great arteries must, therefore, undergo an arterial switch in the first few weeks of life, preferably by 2 weeks of age, before the left ventricular mass becomes inadequate to maintain the systemic circulation. In rare instances (e.g. those with sepsis or delayed diagnosis), patients who are otherwise good candidates may be unable to undergo the arterial switch in the neonatal period. Pulmonary artery banding in early infancy induces a rapid increase in left ventricular wall mass, thereby ‘preparing' the left ventricle for a delayed arterial switch within a few weeks. Construction of a systemic-to-pulmonary artery shunt concomitant with pulmonary artery banding may be necessary to provide an adequate systemic oxygen saturation.

The other critical factor determining the outcome after the arterial switch procedure is the coronary artery anatomy. As illustrated in Fig. 1, the most common pattern of coronary distribution in transposition of the great arteries is ‘normal', and is easily amenable to arterial switch. However, variations in coronary anatomy are frequently present, and require careful planning to avoid kinking or tension upon transfer to the neo-aorta. The presence of certain patterns, including single coronary artery, intramural coronary artery, and origin near the aortic valve commisure was previously considered a contraindication to arterial switch. Although associated with a somewhat higher risk of postoperative myocardial ischemia and death, techniques have been developed which allow coronary transfer in such situations, so that arterial switch may be theoretically possible for all patterns of coronary anatomy. Preoperative recognition of such high risk patterns is essential for proper planning of these higher risk arterial switch procedures.

Although a learning curve exists, centers with experience report an early mortality rate between 2 and 5 per cent for neonatal arterial switch. Retropulmonary course of a left coronary artery arising from the right posterior sinus was the only morphologic predictor of increased mortality in one large long-term study. Reports comparing the atrial and the arterial switch demonstrate a significantly better systemic ventricular ejection fraction and a decreased incidence of atrial arrhythmias in the arterial switch group, without a difference in overall mortality. Morbidity after the arterial switch procedure is primarily related to pulmonary artery stenosis and occlusion of the left anterior descending coronary artery. Stenosis of the pulmonary artery requiring reoperation is reported in 5 to 10 per cent of patients. The use of absorbable suture material and a generous pericardial patch reconstruction at the sites of coronary button excision allows growth of the pulmonary artery anastamosis. The incidence of late pulmonary artery stenosis has declined with its use. Occlusion of the left anterior descending coronary artery occurs in up to 14 per cent of patients following the arterial switch and is thought to occur from kinking or tension on the vessel following reimplantation. Increased experience should reduce the incidence of this complication. It should be noted that most patients with coronary occlusion diagnosed late after surgery demonstrate normal left ventricular function, presumably due to the excellent coronary collaterals which develop in this age group. Long-term monitoring is necessary to demonstrate the continued adequacy of such collaterals. Intermediate and long-term reports demonstrate normal left ventricular size and function by echocardiography, without evidence of deterioration over time. Subsegmental myocardial perfusion abnormalities, detected by technetium-99m sestamibi imaging, are common in children 4 to 8 years following arterial switch, but are not associated with echocardiographic, electrocardiographic, or clinical evidence of ischemia. Formal exercise tolerance testing appears to be normal late after arterial switch.

The neonatal arterial switch is currently the procedure of choice for simple transposition of the great arteries. In patients who cannot undergo definitive repair in the first few weeks of life, pulmonary artery banding and systemic-to-pulmonary shunt followed by delayed arterial switch is indicated. In these neonates, arterial switch can usually be performed between 4 and 7 days after banding. Atrial switch is reserved for the increasingly rare situations in which the coronary anatomy is deemed unamenable to correction by arterial switch.

Transposition of the great arteries with ventricular septal defect

Palliation

Initial management considerations are identical to those outlined above for patients with simple transposition of the great arteries. Excessive pulmonary blood flow resulting from shunting across the ventricular septal defect causes congestive cardiac failure and was formerly palliated by pulmonary artery banding. This may rarely still be required, but is reserved for patients too ill to undergo arterial switch and for those with complex anatomy.

Definitive treatment

The general strategy in treatment of transposition of the great arteries with ventricular septal defect is redirection of blood flow to establish the pulmonary and systemic circulations in series, and closure of the defect. The atrial switch procedures with ventricular septal defect closure were found to have poor results in this group of patients. The operative mortality was high and late right ventricular dysfunction with tricuspid regurgitation was more frequent than in patients with transposition of the great arteries and intact septum. This prompted the early adoption of the arterial switch with closure of the ventricular septal defect for these patients. As with simple transposition, atrial switch or arterial switch may be utilized. However, some differences exist with regard to timing of repair in transposition of the great arteries with ventricular septal defect. Systemic-to-pulmonary shunting across the septal defect produces volume overloading of the left ventricle with consequent maintenance of left ventricular wall thickness despite the postnatal decrease in pulmonary vascular resistance. These patients can therefore tolerate the arterial switch procedure after the first few weeks of life. In fact, the initial successful application of the arterial switch was limited to older infants with transposition of the great arteries and ventricular septal defect. Delayed operation allows cardiac and coronary artery growth, thereby reducing the difficulty of coronary transfer. Improved early and late results were obtained as well as increased experience in anastomosis of smaller coronary vessels. This prompted application of the arterial switch to patients with simple transposition of the great arteries in the neonatal period. Although patients with transposition of the great arteries and ventricular septal defect can undergo later repair, the success of neonatal repair for simple transposition and the risk of morbidity from excessive pulmonary blood flow have resulted in the current recommendation for neonatal arterial switch and ventricular septal defect closure. The technique is as described above, with Dacron patch closure of the defect through the right atrium during a brief period of circulatory arrest. Atrial switch is reserved for patients with coronary anatomy not amenable to arterial switch.

Transposition of the great arteries, ventricular septal defect, and left ventricular outflow tract obstruction

Palliation

Patients with transposition of the great arteries, ventricular septal defect, and left ventricular outflow tract obstruction have diminished pulmonary blood flow. As with other types of transposition, initial therapy includes atrial septostomy and prostaglandin E_1 to maintain patency of the ductus arteriosus. These patients require construction of a systemic-to-pulmonary artery shunt to augment pulmonary blood flow and allow survival until definitive repair can be performed.

Definitive treatment

The anatomy in these patients is not amenable to repair by either atrial or arterial switch procedures. Direct excision of the subpulmonic left ventricular outflow tract is only occasionally possible, in which case it is performed in conjunction with arterial switch and septal defect closure. In most cases, direct excision risks injury to the mitral apparatus and conduction system, and is difficult to perform without left ventriculotomy which may impair ventricular performance. Instead, right ventriculotomy is performed and patch closure of the ventricular septal defect is accomplished with baffling of the left ventricular outflow through the defect into the aorta. The pulmonic valve is oversewn, and a valved conduit is placed between the right ventricle and the pulmonary trunk. The procedure, first described by Rastelli, is illustrated in Fig. 10.

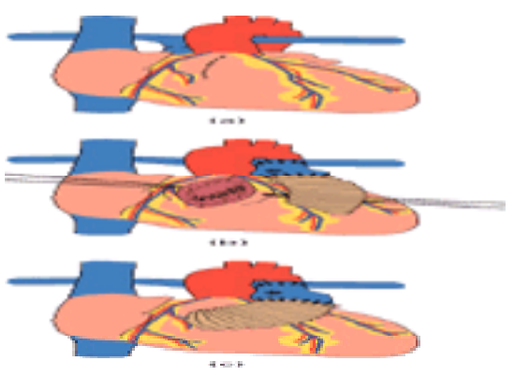


Fig. 10. Rastelli procedure. (a) Incisions in right ventricular outflow tract and pulmonary trunk. (b) Ventricular septal defect patch visualized through the right ventriculotomy. Composite Dacron–homograft conduit anastomosed to the pulmonary arteriotomy on the left side of the aorta.(c) Completed repair with anastomosis of

the Dacron portion of the conduit to the right ventriculotomy.

Operation is deferred until 3 to 5 years of age so that placement of an adult-sized conduit is possible. Although operative mortality rates were high in early series, recent results indicate an acceptable rate of less than 5 per cent. Morbidity is principally secondary to late conduit obstruction, usually related to sternal compression and development of a fibrointimal peel. The use of cryopreserved homograft conduits and placement of the conduit well to the left of the sternum have significantly reduced the incidence of this complication. If obstruction or growth necessitates conduit replacement, the procedure can be performed safely with mortality rates under 1 per cent.

More recently, in an effort to avoid the morbidity of conduit replacement, direct anastomosis of the pulmonary trunk to the right ventriculotomy, as described by Lecompte, has been reported for treatment of transposition of the great arteries, ventricular septal defect, and left ventricular outflow tract obstruction. Short-term results in a small number of patients appear promising. Intermediate- and long-term follow-up is necessary in this patient group to demonstrate the absence of right ventricular dysfunction resulting from the obligatory pulmonary insufficiency associated with this procedure.

Summary

Transposition of the great arteries is the most common form of cyanotic congenital heart disease in the neonate, and is associated with a grave prognosis if left untreated. Diagnosis may be suspected by physical examination and chest radiographs, and confirmed by echocardiography in most situations. Cardiac catheterization with coronary angiography may occassionally be required. Initial treatment is accomplished by balloon atrial septostomy and prostaglandin E_1 infusion. Neonatal arterial switch has become the accepted definitive procedure of choice for patients with transposition of the great arteries, with or without ventricular septal defect. Delayed arterial switch with prior preparation of the left ventricle by pulmonary artery banding should be performed in patients with simple transposition in whom delayed diagnosis or medical contraindications prohibit neonatal arterial switch. Atrial switch is reserved for the rare situations in which coronary anatomy is unamenable to arterial switch. The Rastelli procedure is employed for transposition of the great arteries with ventricular septal defect and left ventricular outflow tract obstruction. Arterial switch following prolonged preparation with pulmonary artery banding may obviate or delay cardiac transplantation in patients with severe right ventricular failure occurring late after atrial switch procedures. Excellent early, midterm, and late results continue to improve as more experience is obtained.

Severe right ventricular failure following atrial switch procedures

As described above, severe right ventricular failure, tricuspid regurgitation, and atrial arrhythmias are commonly seen late after atrial switch procedures. Previously, cardiac transplantation has been the only surgical option. Recently, however, protocols for pulmonary artery banding and delayed arterial switch have been reported. These older patients, in contrast to infants, require a longer period of banding to allow adequate ‘preparation' of the left ventricle, typically 6 months or more. Banding is not well tolerated and in most cases the band must be serially tightened. Adequacy of preparation is determined by echocardiographic assessment of left ventricular wall mass and hemodynamic measurement of near- systemic left ventricular pressure. Patients with successful preparation undergo arterial switch procedure as described above. Takedown of the Mustard or Senning baffle and reconstruction of the atrial septum is performed as well.

Late arterial switch for failed atrial switch operations has only been reported in a small number of cases, but mortality (both during preparation and perioperatively) is around 25 to 30 per cent. Increasing experience may allow development of improved preparation protocols and better ability to select between delayed arterial switch and transplantation.

Congenitally corrected transposition of the great arteries

In some patients, transposition of the great arteries may be associated with discordance between the atria and the ventricles. Systemic venous blood returns to the right atrium, fills the morphologically left ventricle, and exits via the pulmonary artery to the lungs. Pulmonary venous return to the left atrium is directed to the morphologically right ventricle and aorta. Thus, the congenitally abnormal relationships between the atria and ventricles correct the circulatory derangements which would otherwise be caused by the transposed great arteries (i.e. a physiologic correction occurs). Congenitally corrected transposition of the great arteries is a rare congenital defect, occurring in 0.008 per cent of live births. The terms left and right ventricle will refer to morphology rather than spatial position for the remainder of this chapter.

Morphology

In 1 per cent of cases, congenitally corrected transposition of the great arteries is present without associated abnormalities. The atria are normal in position and morphology. The right atrium connects to the left ventricle through the mitral valve. Fibrous continuity exists between the mitral and pulmonic valves. Similarly, the left atrium connects to the right ventricle through the tricuspid valve, which is separated by a muscular ridge from the aortic valve annulus. The left ventricle is positioned posteriorly and to the right of the right ventricle. The aorta arises anterior and to the left of the pulmonary trunk.

The coronary arterial distribution is usually a mirror image of normal, but the relationship between the coronary vessels and the ventricles is unchanged. Thus, the left main coronary artery arises from a right aortic sinus of Valsalva and bifurcates into anterior descending and circumflex branches. The anterior descending artery courses in the anterior interventricular groove (located more rightward than in normal hearts) to the apex. The circumflex coronary artery passes rightward to the right atrioventricular groove with obtuse marginal branches to the left ventricular free wall. The right coronary artery arises from a left aortic sinus and enters the left atrioventricular groove, giving rise to a posterior descending artery in most cases.

The abnormal connection between atria and ventricles results in significant abnormalities of the conduction system, specifically involving the atrioventricular node and penetrating bundle. A normally located ‘posterior' atrioventricular node is found in its usual location near the coronary sinus, but does not connect with the penetrating bundle. A second anteriorly located atrioventricular node exists in these hearts and connects to the bundle branches, thereby allowing atrioventricular conduction.

Pathophysiology

Patients with congenitally corrected transposition of the great arteries in whom no associated defects coexist may be completely asymptomatic. Right ventricular failure and tricuspid regurgitation develops in 21 per cent of these patients as the right ventricle is required to support the systemic circulation. Tricuspid regurgitation may be caused by annular dilatation with normal leaflet morphology, or by abnormal leaflets with apical displacement similar to that found in Ebstein's malformation. The abnormal anatomy of the atrioventricular node and penetrating bundle results in an 8 to 20 per cent incidence of spontaneous atrioventricular block.

An associated ventricular septal defect is present in 60 to 70 per cent of patients with congenitally corrected transposition of the great arteries. Left-to-right shunting causes excessive pulmonary blood flow and volume overload of the morphologic right ventricle. Atrioventricular conduction disturbances are more frequent when a ventricular septal defect is present. The coexistence of left ventricular outflow tract obstruction and ventricular septal defect produces right-to-left shunting with resultant decreased pulmonary blood flow and cyanosis. The pathophysiologic picture is identical to that of tetralogy of Fallot. Left ventricular outflow tract obstruction is most commonly annular or subvalvar in nature, but may occur at the pulmonic valve as well.

Diagnosis

Most patients with congenitally corrected transposition of the great arteries present because of the effects of the associated cardiac defects or because of heart block. Physical examination of patients with ventricular septal defect reveals a holosystolic murmur; those who also have left ventricular outflow tract obstruction present with cyanosis and an ejection murmur heard best at the right sternal border. Chest radiography is unremarkable in uncomplicated congenitally corrected transposition of the great arteries. Cardiomegaly with pulmonary plethora will be seen in patients with ventricular septal defect, and cardiomegaly with diminished pulmonary vascular markings is present in those who also have left ventricular outflow tract obstruction. Cardiomegaly and pulmonary edema may be seen in patients with tricuspid regurgitation. The physical examination and radiographic findings will not distinguish congenitally corrected transposition of the great arteries from morphologically normal hearts with similar defects, but electrocardiography may. Abnormal septal depolarization results in right precordial q waves and absent q waves in the left precordial leads, a pattern directly opposite to that found in normally related ventricles. Atrioventricular block, although not pathognomonic, is commonly present. A definitive diagnosis of congenitally corrected transposition of the great arteries and its associated defects can be made by echocardiography. The abnormally related ventricles and great vessels are easily identified by two-dimensional echocardiography, as are ventricular septal defects and left ventricular outflow tract obstruction. Doppler flow studies allow estimation of the left ventricular outflow tract gradient and the degree of tricuspid regurgitation. Cardiac catheterization with angiocardiography remains the ‘gold standard' of diagnosis, but as with transposition of the great arteries, is being supplanted by echocardiography.

Management

Classic repair

Congenitally corrected transposition of the great arteries does not, in and of itself, require treatment. The management is based on the presence of coexisting lesions. The ‘classic' approach has involved direct treatment of the coexisting lesions, with no attempt to correct the atrioventricular and ventriculoarterial discordance.

Congenitally corrected transposition of the great arteries with ventricular septal defect

The ventricular septal defect is managed in a manner similar to that in infants with morphologically normal hearts. Infants with severe congestive heart failure unresponsive to medical therapy including digoxin and diuretics should undergo patch closure of the defect. Patients with minimal or no heart failure can undergo elective repair at about 1 year of age. The defect is usually accessible through a right atriotomy with retraction of the mitral valve leaflets for exposure. Sutures are placed on the right ventricular side of the defect, as the conduction tissues lie superficially in the anterior left ventricle.

Left ventricular outflow tract obstruction

This is a difficult problem in patients with congenitally corrected transposition of the great arteries. Since the obstruction is usually at the annular and subvalvar levels, simple valvotomy and resection is not effective. Annular enlargement and resection of subpulmonic muscle bundles has been reported, but with high rates of recurrence. Placement of a transannular patch, as in tetralogy of Fallot, is not possible since the circumflex coronary artery crosses immediately anterior to the left ventricular outflow tract. Thus, a left ventricle to pulmonary artery conduit is required. Conduits should be placed from the left ventricular apex around the right heart border to the pulmonary artery, thereby avoiding sternal compression. There is often little space for such a conduit in neonates and infants; these patients would also quickly outgrow the small conduits used. Therefore, in infants and small children with severe cyanosis, a systemic-to-pulmonary artery shunt (usually a modified Blalock–Taussig shunt) is constructed as a palliative procedure. Definitive repair with conduit placement is performed around 5 or 6 years of age.

Tricuspid regurgitation

Tricuspid valve replacement is indicated in patients with significant symptomatic tricuspid regurgitation and minimally symptomatic patients with severe tricuspid regurgitation and marked right ventricular dilatation. Repair for the Ebstein's type of valve in congenitally corrected transposition of the great arteries is rarely successful, so valve replacement is usually preferred. The tricuspid valve is approached through the left atrium similar to mitral valve operations in the morphologically normal heart.

Atrioventricular block

The presence of atrioventricular block, either primarily or as a result of surgical repair of other lesions, requires placement of a permanent pacing system. Maintenance of atrioventricular synchrony is preferable since the right ventricle may be more susceptible to failure and may benefit from the ‘atrial kick.' In adults and larger children, VDD or DDD pacing systems may be implanted transvenously or epicardially at the time of operation. Epicardial placement is necessary in smaller children.

Results

Operation for congenitally corrected transposition of the great arteries with ventricular septal defect is associated with a 10 per cent mortality rate. Repair of left ventricular outflow tract obstruction or tricuspid valve replacement have a 10 to 20 per cent mortality rate. The incidence of postoperative atrioventricular block is 10 per cent. As in patients undergoing atrial switch procedures for simple transposition of the great arteries, late right ventricular failure occurs frequently with resultant poor mid- and long-term results.

Anatomic repair

Given the poor outcome following ‘classic' repair of congenitally corrected transposition of the great vessels, several centers have recently reported results of anatomic repair. This consists of atrial switch procedure combined with arterial switch and ventricular septal defect closure in patients without left ventricular outflow tract obstruction (Fig. 11), and atrial switch combined with the Rastelli procedure for patients with left ventricular outflow tract obstruction. The Senning technique is preferred for the atrial switch, although the Mustard procedure may occasionally be indicated in patients with abnormal atrial site or morphology.

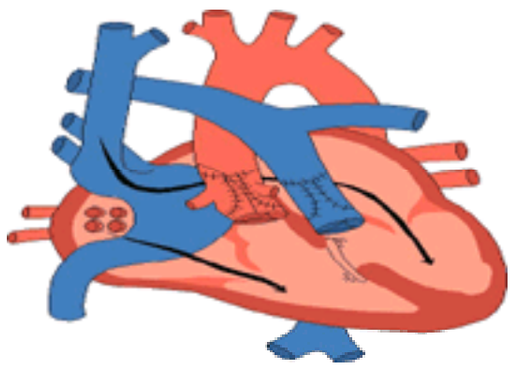


Fig. 11. Senning and arterial switch procedure for congenitally corrected transposition of the great arteries and ventricular septal defect.

Early mortality ranges from 7 to 15 per cent. Although most patients undergoing these procedures have been around 3 to 4 years old, successful repair in infants has been reported. Intermediate and long-term results are not yet available, but are expected to be superior to those of ‘classic' repair. If early mortality for these technically complex and long procedures continues to improve, anatomic correction will become the treatment of choice for congenitally corrected transposition of the great vessels.

Further reading

Anderson RH, Becker AE, Arnold R, Wilkinson JL. The conduction tissues in congenitally corrected transposition. *Circulation* 1974; **50**: 911–23.

Anderson RH, Becker AE, Gerlis LM. The pulmonary outflow tract in classically corrected transposition. *Journal of Thoracic and Cardiovascular Surgery* 1975; **69**: 747–57.

Arensman FW *et al.* Assessment of coronary and aortic anastamoses after anatomic correction of transposition of the great arteries. *Journal of Thoracic and Cardiovascular Surgery* 1985; **90**: 597–604.

Borow K, Arensman FW, Webb C, Radley-Smith R, Yacoub M. Assessment of left ventricular contractile state after anatomic correction of transposition of the great arteries. *Circulation* 1984; **69**: 106–12.

Bove EL. Senning's procedure for transposition of the great arteries. *Annals of Thoracic Surgery* 1987; **43**: 678–80.

Castaneda AR, Norwood WI, Jonas RA, Colon SD, Sanders SP, Lang P. Transposition of the great arteries and intact ventricular septum: anatomical repair in the neonate. *Annals of Thoracic Surgery* 1984; **38**: 438–42.

Colan SD, Boutin C, Castaneda AR, Wernovsky G. Status of the left ventricle after arterial switch operation for transposition of the great arteries: hemodynamic and echocardiographic evaluation. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 311–21.

Corno A, George B, Pearl J, Laks H. Surgical options for complex transposition of the great arteries. *Journal of the American College of Cardiology* 1989; **14**: 742–9.

Day RW, Laks H, Drinkwater DC. The influence of coronary anatomy on the arterial switch operation in neonates. *Journal of Thoracic and Cardiovascular Surgery* 1992; **104**: 706–12.

Egloff L *et al.* Congenitally corrected transposition of the great arteries: a clinical and surgical study. *Thoracic and Cardiovascular Surgeon* 1980; **28**: 228–32.

Friedberg DZ, Nadas AS. Clinical profile of patients with congenital corrected transposition of the great arteries. A study of 60 cases. *New England Journal of Medicine* 1970; **282**: 1053–9.

Hayes JH, Gersony WM. Arrhythmias after the Mustard operation for transposition of the great arteries: a long-term study. *Journal of the American College of Cardiology* 1986; **7**: 133–7.

Ilbawi MN *et al.* Preparation of the left ventricle for anatomical correction in patients with simple transposition of the great arteries. *Journal of Thoracic and Cardiovascular Surgery* 1987; **94**: 87–94.

Karl TR, Weintraub RG, Brizard CP, Cochrane AD, Mee, RBB. Senning plus arterial switch operation for discordant (congenitally corrected) transposition. *Annals of Thoracic Surgery* 1997; **64**: 495–502.

Kato H, Nakano S, Matsuda H, Hirose H, Shimazaki Y, Kawashima Y. Right ventricular function after atrial switch operation for transposition of the great arteries. *American Journal of Cardiology* 1989; **63**: 226–30.

Kirklin JW, Barratt-Boyes BG. *Cardiac surgery*, 2nd edn. Churchill Livingstone, New York, 1993.

Laks H. The arterial switch procedure for the neonate: coming of age. *Annals of Thoracic Surgery* 1989; **48**: 3–4.

Lecompte Y *et al.* Reconstruction of the pulmonary outflow tract without prosthetic conduit. *Journal of Thoracic and Cardiovascular Surgery* 1982; **84**: 727–33.

Marx GR, Hougen T, Norwood WI, Fyler DC, Castaneda AR, Nadas AS. Transposition of the great arteries with intact ventricular septum: results of Mustard and Senning operations in 123 consecutive patients. *Journal of the American College of Cardiology* 1983; **1**: 476–83.

McGrath LB, Kirklin JW, Blackstone EH, Pacifico AD, Kirklin JK, Bargeron LM Jr. Death and other events after cardiac repair in discordant atrioventricular connection. *Journal of Thoracic and Cardiovascular Surgery* 1985; **90**: 711–28.

Milanesi O, Siew YH, Thiene G, Frescura C, Anderson RH. The ventricular septal defect in complete transposition of the great arteries: pathologic anatomy in 57 cases with emphasis on subaortic, subpulmonary, and aortic arch obstruction. *Human Pathology* 1987; **18**: 392–6.

Musewe NN *et al.* Cardiopulmonary adaptation at rest and during exercise 10 years after Mustard repair for transposition of the great arteries. *Circulation* 1988; **77**: 1055–61.

Mustard WT. Successful two-stage correction of transposition of the great vessels. *Surgery* 1964; **55**: 469–72.

Newfeld EA, Paul MH, Muster AJ, Idriss FS. Pulmonary vascular disease in transposition of the great vessels and intact ventricular septum. *Circulation* 1979; **59**: 525–30.

Senning A. Surgical correction of transposition of the great arteries. *Surgery* 1959; **45**: 966–79.

Trusler GA. The Mustard procedure: a commentary. *Annals of Thoracic Surgery* 1987; **43**: 570–1.

Trusler GA, Castaneda AR, Rosenthal A, Blackstone EH, Kirklin JW, the Congenital Heart Surgeons Society. Current results of managemant in transposition of the great arteries, with special emphasis on patients with associated ventricular septal defect. *Journal of the American College of Cardiology* 1987; **10**: 1061–71.

Turina M, Siebenmann R, Nussbaumer P, Senning A. Long-term outlook after atrial correction of transposition of the great arteries. *Journal of Thoracic and Cardiovascular Surgery* 1988; **95**: 828–35.

Weindling SN *et al.* Myocardial perfusion, function, and xercise tolerance after the arterial switch operation. *Journal of the American College of Cardiology* 1994; **23**: 424–33.

Wernovsky G *et al.* Factors influencing early and late outcome of arterial switch operation for transposition of the great arteries. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 289–302.

Williams WG *et al.* Early and late results of a protocol for simple transposition leading to an atrial switch (Mustard) repair. *Journal of Thoracic and Cardiovascular Surgery* 1988; **95**: 717–26.

Yacoub MH, Radley-Smith R, Hilton CJ. Anatomical correction of complete transposition of the great arteries and ventricular septal defect in infancy. *British Medical Journal* 1976; **i**: 1112–14.

Yagihara T *et al.* Double switch operation in cardiac anomalies with atrioventricular and ventriculoarterial discordance. *Journal of Thoracic and Cardiovascular Surgery* 1994; **107**: 351–8.

40.7.8 Congenital anomalies of the thoracic aorta and main pulmonary arteries

Stephen Westaby

[Interrupted aortic arch](#)
[Persistent truncus arteriosus](#)
[Indications of surgery](#)
[Anomalies of the aortic arch](#)
[Double aortic arch](#)
[Malformations associated with right aortic arch](#)
[Malformations associated with left arch](#)
[Anomalies of the brachiocephalic vessels](#)
[Pulmonary artery sling](#)
[Aneurysm of the sinus of Valsalva](#)
[Aortopulmonary window](#)
[Aortoventricular tunnel](#)
[Origin of the right or left pulmonary artery from the ascending aorta](#)
[Further reading](#)

Patent ductus arteriosus and coarctation of the aorta are among the most common congenital cardiovascular defects. However, other anomalies of the great vessels, such as aortopulmonary window or pulmonary artery sling, are very rare (Table 1). Most defects of the aorta or main pulmonary arteries present, or are detectable, in infancy when successful surgical correction is feasible. Delayed intervention risks the development of irreversible or inoperable systemic or pulmonary hypertension (Table 2).

Disease	*Percentage*
Ventricular septal defect	30.5
Atrial septal defect	9.8
Patent ductus arteriosus	9.7
Pulmonic stenosis	6.9
Coarctation of the aorta	6.8
Aortic stenosis	6.1
Tetralogy of Fallot	5.8
Complete transposition of the great arteries	4.2
Persistent truncus arteriosus	2.2
Tricuspid atresia	1.3
All others	16.5

Table 1 Frequency of occurrence of cardiac malformations at birth

Aortic shunts
Patent ductus arteriosus
Aorticopulmonary septal defect
Truncus arteriosus
Pulmonic atresia, ventricular septal defect, and large "bronchial" collateral vessels
Ventricular shunts
Ventricular septal defect
Single ventricle
Transposition of the great arteries with ventricular septal defect
Double-outlet right ventricle*
Tricuspid atresia and ventricular septal defect without pulmonic stenosis
Mitral atresia and ventricular septal defect or single ventricle
Atrioventricular canal
Atrial shunts
Atrial septal defect: secundum, primum, sinus venosus
Common atrium
Total anomalous pulmonary venous return
Partial anomalous pulmonary venous return
Transposition with atrial septal defect

Table 2 Congenital cardiac lesions that can be complicated by the Eisenmenger reaction

Interrupted aortic arch

This rare defect is characterized by a total lack of continuity between the aortic arch and the descending aorta (Fig. 1). Sometimes there may be an atretic fibrous connection without a lumen. The most common site of the interruption is between the left common carotid and subclavian arteries (Type B, 55 per cent of cases). Obstruction occurs distal to the left subclavian (Type A) in 45 per cent and between the innominate artery and left carotid (Type C) in 5 per cent. In these patients a large patent ductus arteriosus supplies the lower half of the body with desaturated blood from the right ventricle, whereas the arch vessels are supplied with oxygenated blood from the left ventricle. Interrupted aortic arch is commonly associated with other defects, including large ventricular septal defect (in 95 per cent of patients), single ventricle, truncus arteriosus, aortopulmonary window, and transposition of the great arteries.

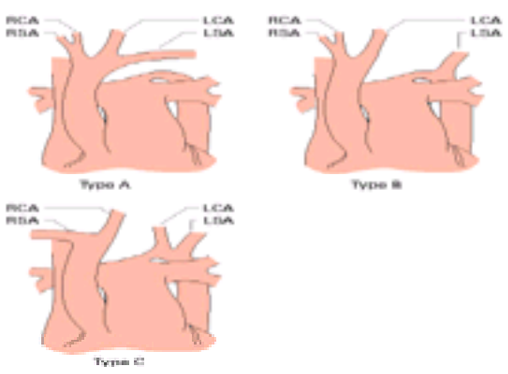


Fig. 1. Classification of interrupted aortic arch (RSA, LSA, right and left subclavian arteries; RCA, LCA, right and left carotid arteries).

Presentation is with severe heart failure in the neonatal period (Table 3) and survival beyond infancy is uncommon (Table 4). Unoperated, the median age at death is 10 days, with an 80 per cent mortality rate within the first 4 weeks of life. Rarely, when the ductus remains widely patent and intracardiac lesions are not complex, presentation is delayed in a manner similar to that of coarctation. When the ductus arteriosus begins to close and pulmonary vascular resistance falls, severe congestive cardiac failure develops, with pulmonary plethora and hypoperfusion of the lower half of the body. The femoral pulses are decreased or absent, and metabolic acidosis, anuria, and renal failure develop rapidly. Echocardiography and cardiac catheterization are used to define the anatomy of the aortic arch and determine the nature and extent of associated intracardiac lesions. When the diagnosis is established administration of prostaglandin E_1 maintains patency of the ductus arteriosus until surgical treatment can be undertaken.

Obstructive lesions	Volume overload
Hypoplastic left heart	Persistent truncus arteriosus
Coarctation of the aorta	Double outlet right ventricle (without pulmonary stenosis)
Severe aortic stenosis	Complete atrioventricular canal defect
Cor triatriatum	Large arteriovenous connections—and those with multiple major cardiac defects

N.B. Patients with lesions associated with reduced blood supply to the lungs rarely de-velop heart failure.

Table 3 Conditions presenting with heart failure in the neonatal period

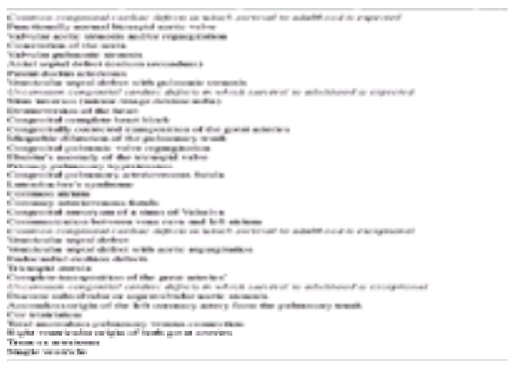


Table 4 Classification of congenital heart disease according to incidence and survival to adulthood

The type of surgical procedure performed depends on the site of interruption and the presence of associated cardiac anomalies, particularly left ventricular outflow tract obstruction. For patients with interruption distal to the left subclavian artery (type A), the situation is similar to that in preductal coarctation. The descending aortaand aortic isthmus are mobilized, and ductus ligated and divided, and direct end-to-end anastomosis is performed between the transverse and descending segments of aorta. When the distance between the two segments does not permit anastomosis, continuity can be established by disconnecting the left subclavian artery and turning this down on to the descending aorta.

Interruption between the left carotid and left subclavian arteries (type B), or rarely between the innominate and left carotid arter-ies (type C), is repaired using sternotomy rather than thoracotomy, and cardiopulmonary bypass with profound hypothermia and total circulatory arrest. With the circulation arrested the proximal descending aorta is mobilized using the ductus arteriosus for traction, until a direct anastomosis between the ascending and descending aorta can be made. Friable ductal tissue must be excised. With this technique associated cardiac defects, particularly ventricular septal defect, are repaired simultaneously (Fig. 2).

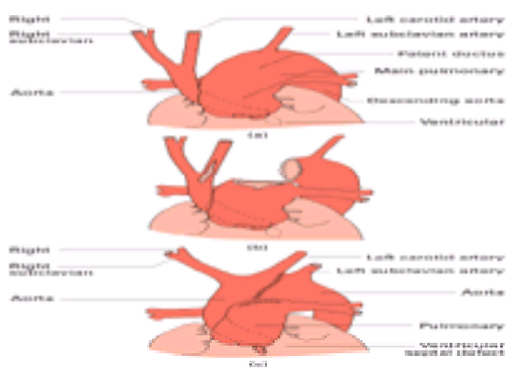


Fig. 2. Repair of interrupted aortic arch (Type B). (a) Type B interrupted aortic arch with ventricular septal defect and patent ductus arteriosus. Note discrepancy between aortic and pulmonary artery diameters and the great distance between ascending and descending aorta. (b) All identifiable ductal tissue has been excised. The incision in the distal ascending aorta has been extended into the carotid artery to allow better approximation and the wideST possible anastomosis. If necessary, the incision in the descending aorta can be extended into the subclavian artery. (c) The completed repair. The ventricular septal defect has been patched via a ventriculotomy and the foramen ovale closed. An end-to-side anastomosis with a single running suture has been created, and the pulmonary artery has been repaired.

Correction of the interrupted aortic arch carries a high risk of mortality. Whether associated cardiac defects are repaired at the same operation or secondarily, hospital mortality may be as high as 80 to 100 per cent. The best series report early mortality of about 30 per cent, though late complications such as subaortic stenosis, aortic anastomotic narrowing, or tube graft gradients cause additional mortality in childhood.

Persistent truncus arteriosus

Persistent truncus arteriosus is an uncommon anomaly, comprising only 2 to 3 per cent of congenital cardiac defects. The defect consists of a single common outflow trunk arising from the base of both ventricles and giving rise to the systemic, pulmonary, and coronary arterial circulations (Fig. 3(a), Fig. 3(b)). The truncus straddles an obligatory malalignment-type ventricular septal defect, which is usually large. The truncal annulus and valve are larger than the normal aortic annulus. About 65 per cent of patients have a tricuspid truncal valve, 25 per cent have four cusps, and the remainder have either bicuspid or pentacuspid valves (Fig. 4). Different morphological classifications have been proposed according to the anatomical configuration of the pulmonary arteries. The widely accepted classification is that of Collett and Edwards (Fig. 5). In Type 1 a single pulmonary trunk arises near the base of the truncus and divides into right and left branches; in Type 2 the right and left pulmonary arteries arise separately but close together usually from the posterior aspect of the truncus; in Type 3 one or both pulmonary arteries arise independently and more laterally from either side of the truncus; and in Type 4 there is absence of true pulmonary arteries with the pulmonary supply via large bronchial arteries. Currently, patients with Type 4 truncus in which the lungs are supplied by systemic arteries are grouped with patients having pulmonary atresia and ventricular septal defect. This defect is commonly referred to as ‘pseudotruncus'. Abnormal location of the coronary ostia, often with one or other main coronary artery arising from the non-coronary sinus of Valsalva, is found in approximately 50 per cent of patients. A degree of truncal valve incompetence is present in 50 per cent of patients. Truncal valve stenosis occurs infrequently, but may be severe. Arch anomalies such as coarctation, hypoplasia, or interruption may be found in association, and approximately 30 per cent of patients have a right-sided aortic arch. In a small proportion of patients, the right pulmonary artery arises from the truncus and the left pulmonary artery from a persistent patent ductus arteriosus outside the pericardium (Fig. 6).

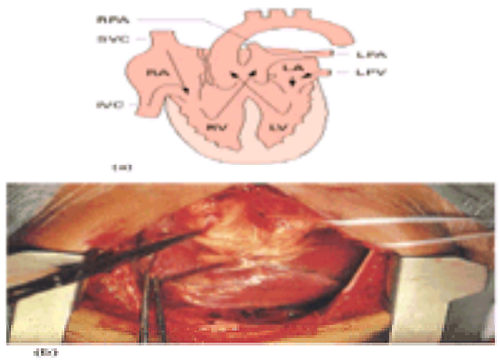


Fig. 3. (a) Schematic diagram of truncus arteriosus. The ventricular septal defect underlying the truncal root permits both ventricles to eject into the common vessel which supplies systemic, pulmonary and coronary circulations. SVC = superior vena cava; IVC = inferior vena cava; RA = right atrium; LA = left atrium; RV = right ventricle; LV = left ventricle; RPA = right pulmonary artery; LPA = left pulmonary artery; LPV = left pulmonary vein. (b) Appearance of the heart at operation. A single large trunk leaves the heart giving origin to the aorta and pulmonary arteries.

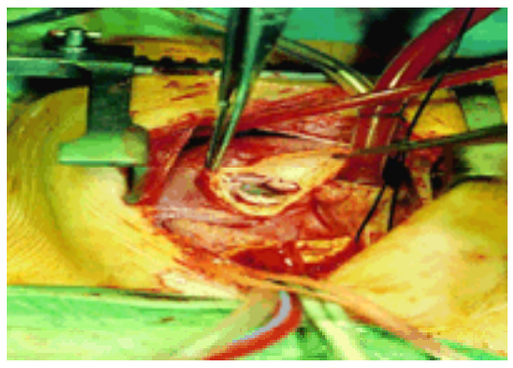


Fig. 4. Truncal valve with four cusps.

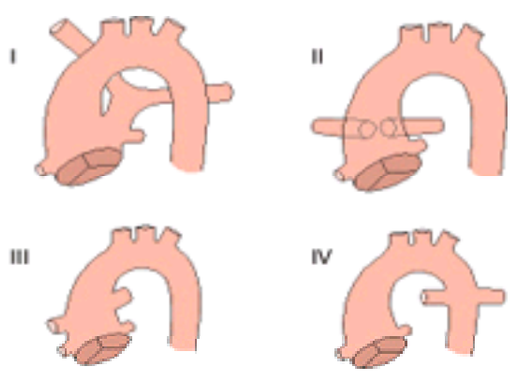


Fig. 5. Classification of truncus arteriosus.

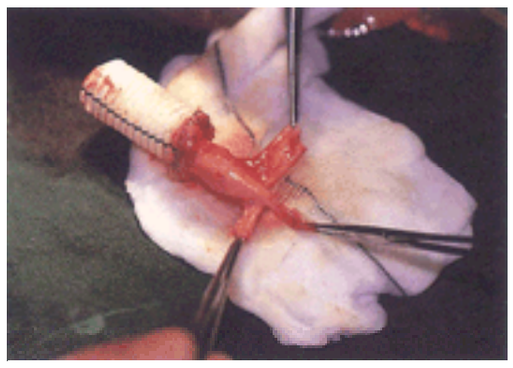


Fig. 6. Composite conduit of fresh pulmonary valve and arteries sewn to collagen impregnated Dacron (Hemashield) used to repair Type I truncus with discontinuous pulmonary arteries.

The haemodynamics in truncus arteriosus are determined by the differential resistance to flow in the systemic and pulmonary vascular beds. Equal systolic pressures exist in both ventricles and the truncus and blood from both ventricles is ejected directly into the common trunk where mixing occurs. Pulmonary blood flow and the development of pulmonary hypertension are influenced by the size of the pulmonary arteries, obstruction in these vessels, and by pulmonary arterial resistance. Systemic arterial saturation is always reduced and is influenced by the volume of pulmonary blood flow. In the neonate, pulmonary blood flow is initially restricted by high pulmonary vascular resistance. When this falls cardiac failure may occur because of large pulmonary blood flow and left ventricular volume overload. High pulmonary arterial pressure and flow eventually results in obstructive pulmonary vascular disease, which may appear within the first 6 months of life.

Unoperated, 90 per cent of patients die of heart failure during early infancy. Approximately 5 per cent of patients with moderate stenosis at the origin of both pulmonary arteries survive for many years with near normal function (natural pulmonary artery banding). In general only 10 per cent of patients survive past the age of 1 year. After this age mortality is determined by the severity of pulmonary vascular disease, which may allow survival to adolescent or adult life.

The clinical presentation is determined primarily by the status of the pulmonary vascular bed. Infants with pulmonary stenosis are often cyanotic in the neonatal period and present with symptoms of hypoxia. Patients with large pulmonary arteries develop congestive heart failure as early as the first week of life, though more often this is delayed until the second or third week. These patients are minimally cyanotic. In more than two-thirds of patients there is the harsh systolic murmur of ventricular septal defect, together with a systolic thrill at the second or third left interspace (Table 5). Stenotic pulmonary arteries may result in a continuous murmur extending over both sides of the upper chest. Excessive pulmonary blood flow may result in a diastolic rumble due to augmented flow across the mitral valve. In most cases the chest radiograph shows cardiomegaly with increased pulmonary vascular markings. The cardiac pedicle is particularly narrow. Two-dimensional echocardiography with Doppler studies confirm the diagnosis and identify associated lesions such as additional ventricular septal defects, coarctation, or interrupted aortic arch. Although the diagnosis is made by non-invasive means, accurate haemodynamic and detailed anatomical evaluation is required before surgical treatment.



Table 5 Principal causes of heart murmurs

Cardiac catheterization is performed with right and left heart studies and angiography. Usually the catheter readily enters the truncus from the right ventricle. Injection of contrast into the prox-imal portion of the truncus defines the abnormal origin of the pulmonary vessels and on the basis of these findings surgical correction can be planned. The presence and magnitude of truncal valve regurgitation is noted. Truncal and pulmonary artery saturations are equal and high in infants presenting with heart failure secondary to high pulmonary blood flow. Truncal saturation less than 80 per cent usually suggests the presence of severe pulmonary vascular disease.

Indications of surgery

Infants presenting with severe heart failure uncontrolled by medical therapy require urgent surgical correction. Prior to satisfactory results from complete correction, pulmonary artery banding was generally undertaken, but this has now been abandoned since mortality exceeded 50 per cent and often failed to protect against pulmonary vascular disease. Nevertheless, some survivors of banding are suitable for correction in adolescent life if distortion of the pulmonary vessels has not rendered them inoperable. For those infants initially stabilized by medical treatment, corrective surgery is usually recommended at the age of 5 to 6 months of age. Delay beyond 6 months increasingly risks development of intractable pulmonary vascular disease. Patients with a pulmonary vascular resistance greater than 10 units/m^2 are generally regarded as inoperable.

The corrective procedure involves detachment of the pulmonary arteries from the truncus followed by closure of the defect in the aorta. The ventricular septal defect is then repaired through a right ventriculotomy which, in turn, will be used as the new right ventricular outflow. Continuity between the pulmonary arteries and the right ventricle is then established by a valved conduit. In Oxford, the operation is undertaken using profound hypothermia and total circulatory arrest. Carefully selected paediatric aorta homografts are used to create the new main pulmonary artery. Excess homograft aorta is used to close the defect in the truncus (neo-aorta) after mobilization of the pulmonary arteries. A Dacron or Gore-tex patch is used to close the ventricular septal defect. For patients with a single pulmonary artery arising from the truncus, continuity of the extra pericardial left pulmonary artery has been established using a bifurcated pulmonary arterial homograft (Fig. 6). Proprietary conduits of Dacron containing a porcine valve prosthesis are also used for correction but have a high late incidence of dysfunction secondary to endothelial overgrowth and valve calcification.

Anomalies of the aortic arch

Embryologically the aortic arch and its major intrathoracic branches, together with the main pulmonary arteries and ductus arteriosus, arise from six pairs of vascular arches which connect paired right and left dorsal aorti with the unpaired ventral vessel (Fig. 7). The normal left-sided aortic arch results from regression of the eighth segment of the right dorsal aorta. The normal aortic arch passes to the left of the trachea, giving rise to the innominate, left common carotid, and left subclavian arteries. The ductus arteriosus connects the arch just distal to the subclavian artery with the pulmonary artery. However, abnormalities in development present the theoretical potential for bilateral aortic arches, and alterations in the normal sequence of development give rise to numerous variations in the anatomy of the aortic arch and brachiocephalic branches.

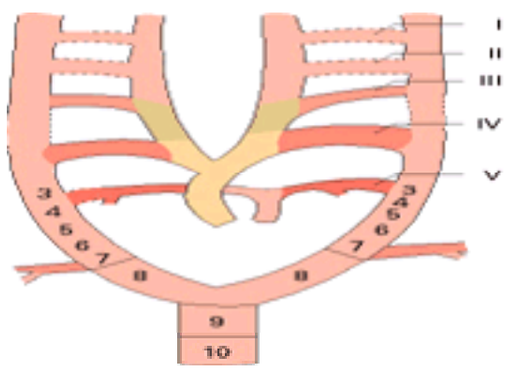


Fig. 7. Fetal aortic arch system. Roman numerals indicate the paired aortic arches and Arabic numerals the segments of the dorsal aortae. Only the seventh dorsal intersegmental artery is shown. Structures delimited by dashed lines ultimately disappear.

In practice, most vascular rings result from lack of regression of the eighth segment of the right dorsal aortic arch or an abnormal regression of other segments. The abnormal vessels form a ring which may completely encircle the mediastinal structures. Symptoms arise due to compression of the trachea, oesophagus, or both. Tracheal compression produces inspiratory stridor or intermittent lower respiratory tract infections, whereas oesophageal compression may cause dysphagia. Double aortic arch is the most common symptomatic anomaly, though vascular rings also occur with malformations of both left and persistent right aortic arch systems (Fig. 8).

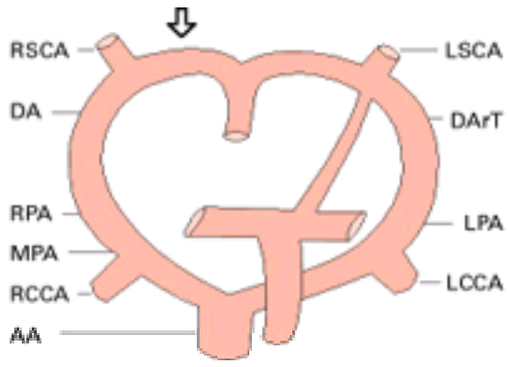


Fig. 8. Schematic representation of double aortic arch. The open arrow in this and subsequent figures indicates sites of abnormal persistence of a vascular structure, in this case the 8th segment of the right dorsal aorta. RSCA and LSCA indicate right and left subclavian arteries respectively; RCCA and LCCA, right and left common carotid arteries respectively. MPA, RPA, and LPA indicate main, right, and left pulmonary arteries respectively. AA, ascending aorta. DA, descending aorta. DArT refers to ductus arteriosus.

Double aortic arch

This results from failure of regression of the eighth segment of the right dorsal aorta with persistence of both right and left dorsal aortic arches. The ascending aorta

bifurcates in front of the trachea to give two separate arches which surround the trachea and oesophagus (Fig. 8). In 75 per cent of patients, the right arch is larger and crosses behind the mediastinal structures, giving rise to a left-sided descending aorta. The carotid and subclavian vessels arise separately, two from each arch. The ductus arteriosus is most commonly left-sided and may contribute to the vascular ring by drawing the pulmonary artery backwards on to the anterior aspect of the trachea.

Most infants with double aortic arch show signs of severe respiratory distress within the first 6 months of life. A high index of suspicion is required to make the diagnosis. Dysphagia or feeding difficulties may accompany respiratory distress. The plain chest radiograph may show double aortic arches, and there may be visible evidence of tracheal compression. Barium instilled into the oesophagus may confirm the impression of vascular ring and allow the major (larger) arch to be identified. If barium swallow is not conclusive, aortography may be required.

Surgical division of the smaller arch with mobilization of the mediastinal structures is required for babies with significant symptoms. When the left arch is smaller, the mediastinum is approached through a left thoracotomy and the arch is divided just distal to the left subclavian artery. Since chronic compression of the trachea produces atretic cartilages, airways obstruction may require weeks or months to resolve.

Malformations associated with right aortic arch

A right-sided aortic arch occurs almost exclusively in association with vascular ring or a separate congenital heart defect (Fig. 9). A right aortic arch with a left innominate artery followed by right carotid and right subclavian artery, is associated with 98 per cent of patients with cyanotic congenital heart disease (often tetralogy of Fallot or persistent truncus arteriosus). In these patients the ductus arteriosus connects the innominate artery to the left pulmonary artery and there is no vascular ring. In contrast, in right-sided aortic arch with aberrant left subclavian artery, the left subclavian arises as the most distal brachiocephalic vessel, and passes posterior to the oesophagus to reach the left arm (Fig. 10). The ductus arteriosus then connects the left subclavian artery with the left pulmonary artery, completing the vascular ring. Concomitant cardiac defects occur in only 10 to 15 per cent of patients and symptoms of compression arise in approximately 20 per cent: these are less severe than in those with double aortic arch. Clinical presentation occurs between the ages of 6 months and 2 years. The proximal portion of the left subclavian artery develops embryologically from the eighth segment of the left dorsal aorta and is consequently greatly dilated. This produces a large posterior oesophageal indentation. When symptoms are significant, surgical correction is by division of the ductus arteriosus or ligamentum with mobilization of the ring away from the trachea and oesophagus.

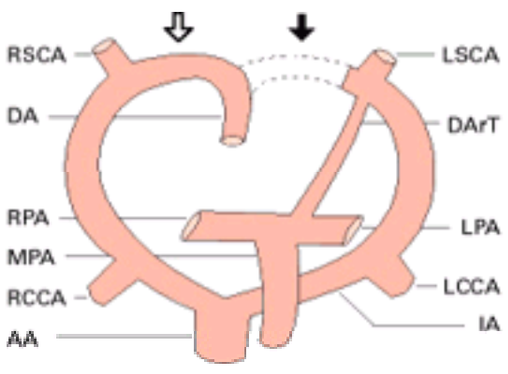


Fig. 9. Schematic representation of right aortic arch with mirror image brachiocephalic branching. Regression of the eighth segment of the left dorsal aorta and persistence of the analogous segment on the right are indicated by the arrows. IA, innominate artery. RSCA and LSCA indicate right and left subclavian arteries respectively. AA, ascending aorta. DA, descending aorta. DArT refers to ductus arteriosus.

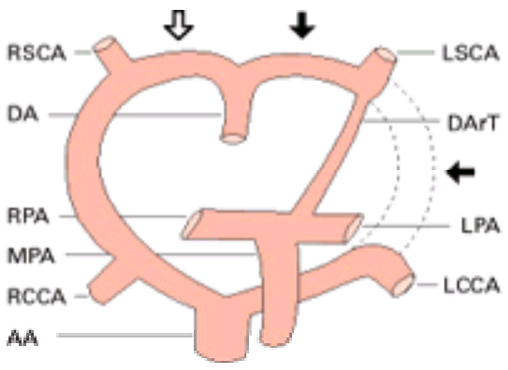


Fig. 10. Schematic representation of right aortic arch with aberrant left subclavian artery. Arrows indicate abnormal regression of the left 4th aortic arch with persistency of the 8th segment of the right dorsal aorta. RSCA and LSCA indicate right and left subclavian arteries respectively; RCCA and LCCA, right and left common carotid arteries respectively. MPA, RPA, and LPA indicate main, right, and left pulmonary arteries respectively. AA, ascending aorta. DA, descending aorta. DArT refers to ductus arteriosus.

In a rare condition, known as cervical aortic arch, a right-sided aortic arch is elongated to extend into the neck above the clavicle. Apart from the abnormal location of the arch, the anomaly is identical to right arch with aberrant left subclavian artery. The ductus or ligamentum on the left completes the ring. Presentation is often with a pulsatile mass in the supraclavicular fossa, though symptoms may relate to mediastinal compression. Barium swallow shows findings similar to the previous lesions except for more cranial location. Because of the rarity of this lesion, angiography is usually employed for precise identification.

Malformations associated with left arch

Vascular rings are seldom found with a left aortic arch, although mirror image equivalents of right arch anomalies are occasionally encountered. The most common anomaly is aberrant right subclavian artery, resulting from regression of the right fourth arch, and persistence of the right eighth dorsal segment. The right subclavian arises distal to the left and passes behind the oesophagus to the right arm. Because the ductus is usually normally sited, no ring is formed and dysphagia rarely, if ever, occurs. Presentation is almost exclusively in adult life when barium swallow may demonstrate an oblique posterior indentation sited between T2 and T4. Rarely, when significant dysphagia is proven, division of the aberrant vessel and mobilization away from the posterior aspect of the aorta will provide symptomatic relief. The surgical approach is by left thoracotomy and care must be taken to oversew the divided right subclavian artery which retracts away behind the oesophagus into the left hemithorax. In one Oxford patient the condition was associated with congenital tracheo-oesophageal fistula.

Anomalies of the brachiocephalic vessels

Variations in origin of the innominate or left carotid arteries are common and usually asymptomatic. Origin of the left carotid artery from the innominate artery may cause anterior tracheal compression and origin of the innominate artery unusually distal on the arch may produce upper airways obstruction as the vessel passes from left to right. Symptoms are usually investigated by bronchoscopy when exaggerated pulsation is evident on the anterior tracheal wall, or the brachial pulse is eliminated by internal pressure from the bronchoscope. Airways obstruction is seldom severe enough to warrant surgery.

Midarch coarctation of the aorta with narrowing of the arch vessels, and congenital septation of the aortic arch with carotid stenosis have been described in Oxford (Fig. 11 and Fig. 12). In both cases aortic arch replacement with reimplantation of the arch vessels was undertaken during circulatory arrest at 18°C.

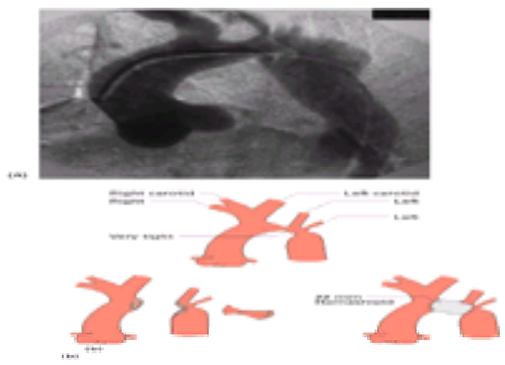


Fig. 11. (a) Mid arch coarctation of the aorta with subclavian stenosis and poststenotic dilatation. (b) Method of repair by aortic arch replacement using cardiopulmonary bypass and circulatory arrest.

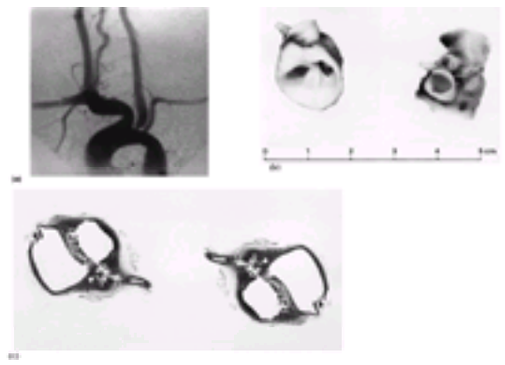


Fig. 12. Congenital septation of the aortic arch with carotid stenosis. (a) Arch aortogram. (b) Resected specimen. (c) Histology of the aortic septation suggested an origin from ductal tissue.

Pulmonary artery sling

This rare but clinically important anomaly occurs when the left pulmonary artery arises from the right pulmonary artery, extrapericardially, or from an elongated main pulmonary artery. The vessel then traverses to the right, crossing the superior aspect of the right main stem bronchus and lower trachea before turning acutely to the left and passing between the trachea and oesophagus to reach the hilum of the left lung (Fig. 13). Pulmonary artery sling is associated with developmental abnormalities of the tracheal cartilages (complete rings) or tracheobronchial compression with tracheomalacia, resulting in respiratory obstruction. The anomaly is thought to originate from failure of the proximal left sixth aortic arch to connect with the plexus of vessels developing in the primitive lung. It is more common in males and usually produces symptoms in the first weeks of life. Expiratory wheeze and hyperexpansion of the right lung result from compression of the tracheobronchial angle. Secondary infection of the affected lung occurs frequently, and death in misdiagnosed cases often occurs in the first 6 months. Wheezing may be confused with asthma, and clinical diagnosis is rarely made prior to investigation of an abnormal chest radiograph showing hyperinflation of the right lung and left shift of the mediastinum. Barium swallow shows anterior indentation of the oesophagus at the level of the carina, and is pathognomonic. Pulmonary angiography is usually carried out for clarification since the defect is rare and requires surgical correction.

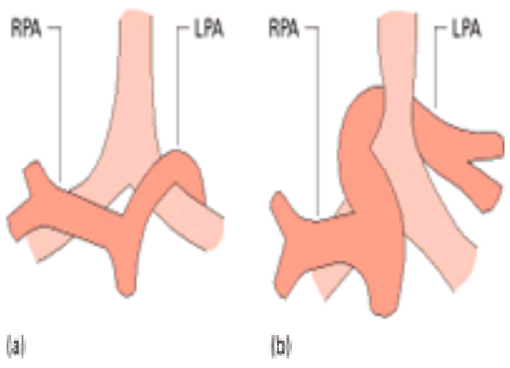


Fig. 13. Pulmonary artery sling. (a) Anterior view of the pulmonary arteries and trachea showing normal relationship between the right (RPA) and left (LPA) pulmonary arteries to their respective bronchi. (b) Anterior view showing the abnormally rightward origin of the left pulmonary artery with its subsequent course around the trachea.

Left thoracotomy is usually used for division and reimplantation of the left pulmonary artery, as described by Potts in 1954. The ligamentum arteriosum or persistent ductus arteriosus is divided and the left pulmonary artery mobilized between the trachea and oesophagus. After heparinization the left pulmonary artery is divided as close as possible to the right pulmonary artery, and the proximal end oversewn. The main pulmonary artery is then approached through a pericardial window in front of the left phrenic nerve. A partial occlusion clamp is used to isolate a portion of the main pulmonary artery into which the left pulmonary artery is implanted in an end-to-side manner. Often the surgical procedure is well tolerated, though tracheobronchial obstruction persists, and may require long-term intubation or tracheoplasty. Some centres prefer a median sternotomy approach and perform the procedure with the patient on cardiopulmonary bypass. The hospital mortality rate, predominantly due to airway complications, is 40 to 50 per cent. Diffuse anatomical tracheal stenosis with complete cartilaginous rings and severe stridor has a particularly poor outlook. Late thrombosis of the reimplanted pulmonary artery is common.

Aneurysm of the sinus of Valsalva

Aneurysm of the sinus of Valsalva is a thin-walled windsock-type out pouching, usually found in the right coronary sinus or adjacent part of the non-coronary sinus (Fig. 14(a), Fig. 14(b)). The tubular out-pouching from the wall of the sinus takes an intracardiac course, protruding into the right atrium or right ventricle (Fig. 15(a), Fig. 15(b). Symptoms occur when acute rupture of the sinus causes an aortocardiac fistula, though this rarely occurs in infancy or childhood.

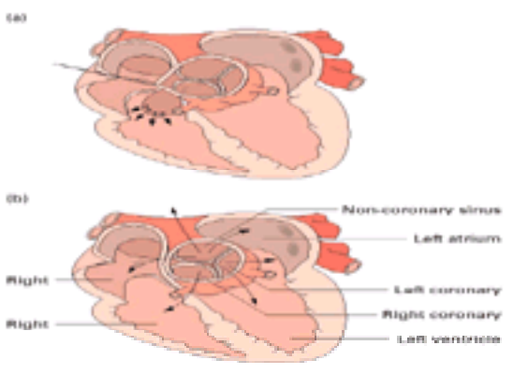


Fig. 14. (a) Aneurysm of the sinus of Valsalva. (b) Direction of rupture of aneurysm.

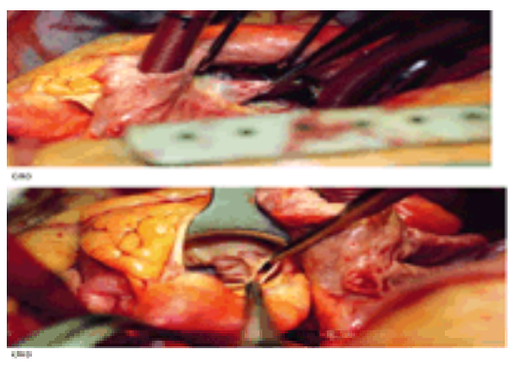


Fig. 15. Ruptured ‘windsock' aneurysm (a) protruding into the right atrium, (b) withdrawn into the aortic root.

The aneurysm occurs at a site of thinning of the wall of the aortic sinus just above the annulus, and is associated with absence of normal elastic and muscular tissue in this area. The sinus of origin determines the direction of protrusion and site and chamber into which rupture occurs. Very rarely aneurysms arise from the left coronary sinus or the posterior portion of the non-coronary sinus and rupture into the adjacent pericardium or left atrium. Other congenital cardiac defects are present in the majority of patients with aneurysm of the sinus of Valsalva. The most common are ventricular septal defects, aortic valve anomalies, and pulmonary stenosis. Coarctation of the aorta, patent ductus arteriosus, atrial septal defects, subaortic stenosis, and tetralogy of Fallot also occur.

Rupture of the aneurysm causes acute symptoms of breathlessness and precordial pain simulating myocardial infarction in 35 per cent of patients. In about 45 per cent of patients the only symptom is effort dyspnoea, and in 20 per cent rupture occurs without symptoms. Rupture may be precipitated by physical exertion or rarely by automobile accidents, bacterial endocarditis, or cardiac catheterization. The characteristic physical findings at this stage are those of aortic regurgitation and often tricuspid regurgitation, an otherwise unusual combination. The characteristic murmur is loud, harsh, superficial, and continuous, with either systolic or diastolic accentuation (Table 6). An elevated jugular venous pressure with prominent V waves occurs through right heart failure or direct entrance of the fistula into the right atrium. Liver enlargement with pulsation and often epigastric pain is progressive.

Table 6 Rupture of aneurysm of the sinus of Valsalva—differential diagnosis

The diagnosis is usually made on clinical grounds in patients with acute severe symptoms. Two-dimensional echocardiography and cardiac catheterization with angiocardiography are used to confirm the diagnosis and identify the sites of origin and rupture. Associated cardiac lesions, including ventricular septal defect, valvular aortic regurgitation, and pulmonary stenosis are also identified. The magnitude of left-to-right shunt through the fistulous communication can be measured, together with the pulmonary vascular resistance. The natural history of aneurysm of the sinus Valsalva is difficult to determine since many aneurysms fail to rupture during the patient's lifespan. Eighty per cent of patients with aneurysm of the sinus of Valsalva are male and in patients under going surgery rupture occurs predominantly in the third or fourth decade of life. Mean survival time after rupture is approximately 4 years, though in patients with a large shunt death may occur in less than 4 weeks from congestive heart failure. In unoperated patients, endocarditis complicates the syndrome in about 8 per cent.

Surgical correction is recommended for all patients. This is undertaken on cardiopulmonary bypass with cold cardioplegic arrest. The approach to the sinus is from both the aortic root as for aortic valve replacement, and through the right atrium, where the windsock and site of rupture can be identified within the right atrium or right ventricle. The dual approach allows accurate repair and avoids damage to the aortic valve. The most common associated anomaly, ventricular septal defect, can also be closed through the right atriotomy. Even if unsuspected, a ventricular septal defect should be sought: in one Oxford patient this remained undiagnosed during investigation through prolapse of an aortic valve cusp. The windsock is opened widely and excised, leaving a margin with which closure can be secured by direct suture or by Dacron patch. Operative mortality is extremely low and the late results excellent.

Aortopulmonary window (Fig. 16(a),Fig. 16(b),Fig. 16(c) and Fig. 16(d))

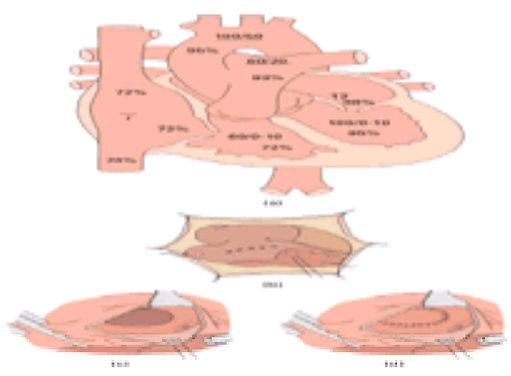


Fig. 16. Aortopulmonary window (a) and its surgical repair (b–d).

Aortopulmonary window is a rare malformation caused by incomplete formation of the septum in the developing truncus arteriosis. The result is an abnormal communication between the aorta and the pulmonary artery distal to their semilunar valves. Symptoms and signs are of congestive heart failure in the newborn. Pulses are bounding and pulse pressure may be wide. The heart is enlarged clinically and on chest radiography the lung fields are plethoric. The differential diagnosis of aortopulmonary window is persistent ductus arteriosis, persistent truncus arteriosis, and ventricular septal defect. Two-dimensional echocardiography will provide the diagnosis by demonstrating two separate semilunar valves related to the two great arteries.

Infants in severe heart failure are operated upon as soon as the diagnosis is established. Those whose heart failure is controlled medically should be operated upon before pulmonary vascular obstructive disease develops. The heart is approached through a midline sternotomy. Using cardiopulmonary bypass, often with deep

hypothermia and circulatory arrest, the aorta is opened and the aortopulmonary window closed with a Dacron patch. Care is taken to visualize the left coronary artery to avoid inclusion in the suture line. Associated lesions such as patent ductus arteriosus or ventricular septal defect are closed simultaneously. Long-term outlook is good when surgery is performed before onset of pulmonary vascular disease.

Aortoventricular tunnel

Aortic left ventricular tunnel is a rare defect which presents at birth with severe congestive heart failure and signs of torrential aortic regurgitation. The aortic orifice is usually sited anteriorly and separated from the right sinus of Valsalva by a prominent transverse supravalvar ridge. The tunnel passes downwards to communicate with the left ventricle and is usually visible externally as an extracardiac bulge. The morphology is distinct from aneurysm of the sinus of Valsalva. Aortic right ventricular tunnel is rare. In this case the tunnel communicates with the infundibulum of the right ventricle rather than with the left ventricle.

Infants presenting with aortoventricular tunnel are investigated by two-dimensional and pulsed Doppler echocardiography to elucidate signs of suspected massive aortic regurgitation. The principal differential diagnosis in this age group is congenital coronary artery fistula. Cardiac catheterization and cineangiography are used to confirm the diagnosis. Urgent surgical treatment, consisting of patch closure of the aortic defect, is then required. There have been multiple reports of successful closure of aortic left ventricular tunnel, though late development of aortic regurgitation has marred the long-term result, possibly due to aortic cusp retraction following direct suture of the defect. Patch closure should avoid this complication. The first successful closure of an aortic right ventricular tunnel was by Westaby in 1990 (Fig. 17).

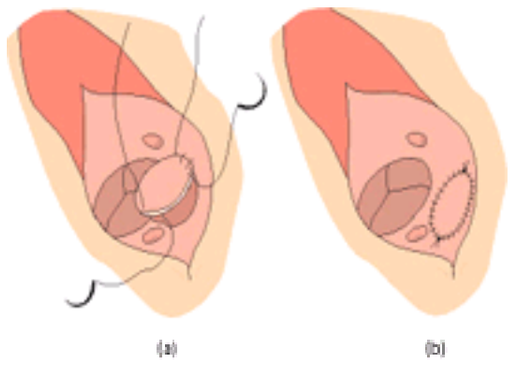


Fig. 17. Patch repair of aorta–right ventricular tunnel.

Origin of the right or left pulmonary artery from the ascending aorta

In this rare anomaly, usually the right but occasionally the left pulmonary artery arises from the ascending aorta in the presence of separate aortic and pulmonary valves and without interposition of the ductus arteriosus. Rarely, both right and left pulmonary arteries arise from the ascending aorta, but in the presence of separate aortic and pulmonary valves: the condition does not constitute truncus arteriosus.

Although less than 100 patients with this anomaly have been described, 70 per cent of untreated patients die before the age of 6 months, and 80 per cent are dead by 1 year. Patients who survive beyond the first months of life develop pulmonary vascular disease in both lungs. The development of pulmonary hypertension in the normally connected left lung is consistent with the finding that experimental ligation of a single pulmonary artery in calves within 24 h of birth results in severe pulmonary hypertension in the opposite lung within 8 weeks. In 80 per cent of patients, aortic origin of the right pulmonary artery occurs in conjunction with another anomaly: patent ductus arteriosus is present in about 50 per cent of patients. Less common associations are with tetralogy of Fallot, ventricular septal defect, coarctation of the aorta, interrupted aortic arch, aortopulmonary window, and atrial septal defect. Clinical presentation occurs in early infancy, with respiratory distress and congestive heart failure. Cyanosis may occur because of venous admixture in the lungs or due to reversed shunting through a patent foramen ovale or ductus arteriosus. Bounding peripheral pulses are obtained through rapid run-off from the aorta into the right lung and left-to-right shunting.

The plain chest radiograph shows cardiomegaly with a globular shaped heart and pulmonary plethora often equal on both sides. Diagnosis is established by two-dimensional echocardiography and confirmed by cardiac catheterization, when pulmonary vascular resistance is calculated as a guide to operability.

Surgical correction is undertaken with profound hypothermia with total circulatory arrest. The anomalous right pulmonary artery is mobilized beneath the superior vena cava and disconnected from the ascending aorta. The resulting defect may be closed directly or with a patch. The aorta is then lifted and pulled leftward to give access to the right side of the main pulmonary artery, into which the mobilized right pulmonary artery is implanted. Surgical results in the absence of associated anomalies are good when reimplantation is performed prior to the development of severe pulmonary vascular disease. Normal late pulmonary artery pressures are obtained, presumably with normal life expectancy.

Further reading

DiDonato RM, Fyfe DA, Puga FJ, Danielson GK, Ritter DG, Edwards WD, McGoon DC. Fifteen-year experience with surgical repair of truncus arteriosus. *Journal of Thoracic and Cardiovascular Surgery* 1985; **89**: 414–20.

Doty DB, Richardson JV, Falkovsky GE, Gordonova MI, Burakovsky VI. Aorto-pulmonary septal defect: haemodynamies angiography and operation. *Annals of Thoracic Surgery* 1981; **32**: 244–9.

Ebert PA, Turley K, Stanger P, Hoffman JIE, Heyman MA, Rudolf AM. Surgical treatment of truncus arteriosus in the first 6 months of life. *Annals of Surgery* 1984; **200**: 451–6.

Kirklin JW, Barratt-Boyes BG, eds. *Cardiac surgery*. John Wiley and Sons, New York, 1980: pp.1035–80.

Mitchell SC, Korones SB, Berendes HW. Congenital heart disease in 56 109 births. Incidence and natural history. *Circulation* 1971; **43**: 323–8.

Roesler M, De Leval M, Chrispin A, Stark J. Surgical management of vascular ring. *Annals of Surgery* 1983; **117**: 139–43.

Sade RM, Rosenthal A, Fellows K, Castaneda AR. Pulmonary artery sling. *Journal of Thoracic and Cardiovascular Surgery* 1975; **69**: 333–8.

Van Mierop LHS, Kutsche LM. Interruption of the aortic arch and coarctation of the aorta. Pathogenetic relations. *American Journal of Cardiology* 1984; **54**: 829–34.

40.7.9 Anomalies of pulmonary and systemic venous return

Emile A. Bacha and Gus J. Vlahakes

Abnormalities of pulmonary venous return
Definition
Embryology
Total anomalous pulmonary venous connection
Partial anomalous pulmonary venous connection
Cor triatriatum
Pulmonary venous stenosis
Anomalies of the systemic venous system
Persistent left superior vena cava
Unroofed coronary sinus
Abnormalities of the inferior vena cava
Further reading

Abnormalities of pulmonary venous return

Congenital abnormalities of the pulmonary venous system are rare, although they probably occur more frequently than is indicated by clinical and autopsy studies. Their variety is diverse, although certain patterns predominate.

Definition

The venous drainage of the lungs may be abnormal in three ways.

Anomalous connection

Instead of draining to the left atrium, the pulmonary veins are connected to the right atrium or one of its tributaries. The condition is termed total anomalous pulmonary venous connection when all the veins connect anomalously, and partial anomalous pulmonary venous connection when one or more (but not all) of the veins connect anomalously.

Obstruction without anomalous connection

Although the pulmonary veins drain as normal to the left atrium, there is some obstruction to blood flow. When obstruction occurs at a single point in the left atrium, the condition is referred to as cor triatriatum. The obstruction may also selectively involve one or more pulmonary veins, causing pulmonary venous stenosis.

Abnormal numbers of pulmonary veins

The veins connect without obstruction to the left atrium. Physiology is normal, and the condition is chiefly of interest to the thoracic surgeon, as it may modify the technique necessary for a pulmonary resection.

Embryology (Fig. 1)

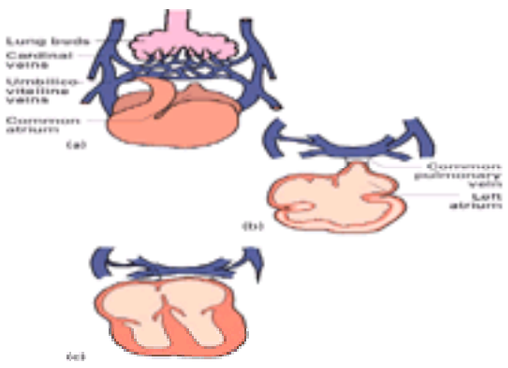


Fig. 1. Embryonic development of pulmonary venous drainage: (a) derivatives of the cardinal and umbilicovitelline veins join developing lung buds to form the primitive pulmonary venous drainage; (b) the left atrium grows to join the developing confluence of pulmonary veins; (c) failure of fusion of the left atrium into the common pulmonary vein can result in total anomalous pulmonary venous connection as shown.

The lungs develop as an outpouching from the foregut. The venous drainage of the primitive lung buds is into a common pulmonary vein, which drains to the cardinal veins. The cardinal veins form the future superior vena cava, left innominate vein, and coronary sinus. The connections with the cardinal veins become atretic when the common pulmonary vein fuses with an outgrowth of the left atrium: any abnormality of this developmental pathway leads to an anomalous connection. For example, if the common pulmonary vein fails to fuse with the left atrium and one or more of the connections to the cardinal veins persists, total anomalous pulmonary venous connection results.

Total anomalous pulmonary venous connection

This condition accounts for about 2 per cent of all cases of congenital heart disease.

Morphology

The common pulmonary vein is retained and connects anomalously to one or more of three sites: supracardiac, to the left innominate vein or superior vena cava; cardiac, to the right atrium or coronary sinus; infracardiac, to the inferior vena cava or portal vein. Blood flow in the anomalous connection may be obstructed, especially in patients with infracardiac connection; in 30 per cent of patients there may be other associated congenital heart defects.

Physiology

Since all pulmonary venous blood flows eventually to the right atrium, systemic and pulmonary venous blood are mixed.

Distribution of pulmonary venous blood depends on the presence and size of the atrial septal defect that is invariably present, any associated pulmonary arterial hypertension, and its effects on flow across the atrial septal defect. As the pulmonary venous blood is highly saturated with oxygen, the right-to-left shunt through this defect does not usually lead to deep cyanosis.

Any additional obstruction in the pulmonary venous drainage leads to pulmonary venous congestion, elevated pulmonary vascular resistance, and decreased pulmonary blood flow. This produces a clinical picture of pulmonary congestion with deepening cyanosis.

Presentation

Tachypnea occurs in the first few weeks of life; noncardiac causes must be excluded. Deep cyanosis is unusual; a pulmonary systolic flow murmur is usually present, but the clinical signs are usually unimpressive.

The chest radiograph is usually normal, although an abnormal shadow may indicate an anomalous vessel: the figure-of-eight or ‘snowman' sign reveals total anomalous pulmonary venous connection to the left innominate vein. Pulmonary venous obstruction may manifest as a diffuse haziness and congestion. The electrocardiogram shows right ventricular hypertrophy resulting from pulmonary hypertension.

Although echocardiography is usually diagnostic, showing the common pulmonary vein lying posterior to the left atrium, cardiac catheterization and angiography are still indicated in some institutions for accurate anatomic and physiologic assessment.

Treatment

Unique to this anomaly is the absence of definitive means of medical palliation; therefore, a severely obstructed, total anomalous pulmonary venous connection represents almost the only true surgical emergency across the entire spectrum of congenital heart surgery. Treatment is surgical in all cases, and should be undertaken urgently following diagnosis: the only exception is in the rare case of the adult patient with established pulmonary vascular disease, in whom surgical repair is contraindicated.

Surgery is performed on cardiopulmonary bypass, usually with deep hypothermia and circulatory arrest. It is always necessary to redirect pulmonary venous blood to the left atrium: this is usually achieved by direct anastomosis between the common pulmonary vein and the back of the left atrium (Fig. 2): abnormal connections are ligated, and the atrial septal defect closed. When total anomalous pulmonary venous connection is to the right atrium, a pericardial baffle can be used to direct blood through the atrial septal defect into the left atrium.

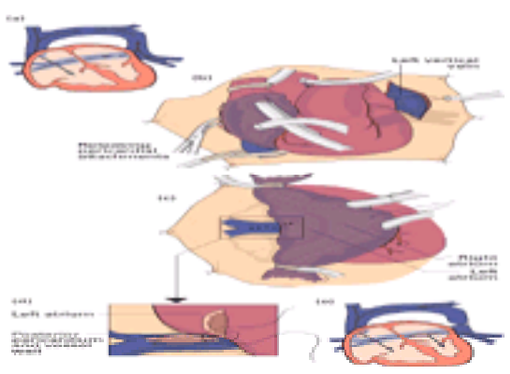


Fig. 2. Operative repair of supracardiac total anomalous pulmonary venous connection: (a) the anomalous connection with the common pulmonary vein draining via a vertical vein to the innominate vein, superior vena cava, and right atrium; (b) first stage of operative repair showing cannulation for cardiopulmonary bypass and surgical steps to mobilize the back of the heart and the left vertical vein; (c) the heart may be rotated toward the left (or alternatively the apex may be elevated) and transverse incisions are made across the left atrium and across the common pulmonary vein [*Note*: the left atrium may be small in these circumstances, and to ensure a wide open anastomosis, the incision in the left atrium must be made from the interatrial septum, extending towards the left and into the left atrial appendage if needed]; (d) anastomosis is then performed between the common pulmonary vein and the left atrium, usually using absorbable monofilament suture material; (e) the left vertical vein is ligated, resulting in the circulatory pathway as shown with the common pulmonary vein draining via the anastomosis into the left atrium.

Postoperative severe pulmonary hypertension and right ventricular failure may be a significant problem, particularly in patients with preoperative pulmonary venous obstruction. This is managed by monitoring pulmonary arterial pressure, deep sedation and respiratory support, control of acidosis, and sometimes by administration of pulmonary vasodilators such as inhaled nitric oxide, prostaglandins or low-dose isoproterenol infusion; in extreme cases, support by an extracorporeal membrane oxygenator may be required. Such treatment may be required for several days before successful ventilatory weaning can be achieved.

Results and prognosis

Untreated, 80 per cent of patients die within 1 year, and most die in the first 3 months of life. Those who survive develop fixed pulmonary vascular disease, right ventricular failure, and deepening cyanosis (Eisenmenger syndrome) by their second decade.

The results of surgery have improved since the first repair, performed in 1956 by Lewis and Varco. Current hospital mortality is below 10 per cent, and most of these die from pulmonary hypertension. Late complications are uncommon, but the principal problem is pulmonary venous obstruction, which usually manifests itself within 6 months of the original surgery. If this obstruction is due to an anastomotic stricture, the anastomosis should be revised; however, it is usually due to diffuse fibrosis in individual pulmonary veins and is difficult to treat surgically; some patients may be suitable candidates for percutaneous balloon dilation. Long-term prognosis is excellent, and many treated patients with no pre-existing pulmonary vascular disease have a normal lifespan.

Partial anomalous pulmonary venous connection

This is present in 0.7 per cent of all individuals autopsied. Nine per cent of patients with an atrial septal defect have associated partial anomalous pulmonary venous connection.

Morphology

The most common varieties are connection of the left superior pulmonary vein to the left innominate vein, and connection of the right superior and/or inferior pulmonary veins to the superior vena cava. When there is an associated defect high in the atrial septum, this is termed the sinus venosus syndrome. Other common connections are from the right superior and/or inferior pulmonary veins to the right atrium, or to the inferior vena cava; in the latter, hypoplasia of the right lung is common, as are other congenital defects involving the heart and diaphragm.

The majority of patients with partial anomalous pulmonary venous connection have an associated atrial septal defect.

Physiology

High pulmonary blood flow results from a left-to-right shunt, as in uncomplicated atrial septal defect. A single anomalous vein alone causes an increase in pulmonary flow of about 20 per cent and is therefore of minimal hemodynamic significance unless a large atrial septal defect is present. Several anomalously connected veins produce a more variable pattern, which depends on the presence, size, and location of any atrial septal defect, and the relative resistances of pulmonary and systemic vascular beds.

Presentation

This depends on the number of anomalously connected veins. One anomalous vein is usually undetectable clinically and most patients have a normal lifespan. When most of the veins are involved, the clinical picture is similar to that of total anomalous pulmonary venous connection.

The presentation in the majority is similar to that of uncomplicated atrial septal defect. There are no symptoms in childhood, but if pulmonary blood flow is very high, patients may develop pulmonary hypertension and right ventricular failure in later years. Most are diagnosed by the presence of a pulmonary flow murmur and wide, fixed splitting of the second heart sound, as in atrial septal defect. When there is hypoplasia of the right lung, as in those with partial anomalous connection to the

inferior vena cava, respiratory infections are common.

The chest radiograph may show the abnormally draining vessel. When the connection is to the inferior vena cava, also known as the ‘scimitar syndrome', a crescentic shadow is usually visible in the right lower lung field. The electrocardiogram shows peaked P waves and right ventricular hypertrophy.

Echocardiography may show abnormal entry sites of anomalous veins and may suggest the magnitude of the left-to-right shunt. As the main differential diagnosis is uncomplicated atrial septal defect, cardiac catheterization is often omitted and surgical correction based on echocardiographic findings. A search for a possible associated partial anomalous pulmonary venous communication should be made during repair of any atrial septal defect; thus it may be diagnosed perioperatively, provided the surgeon has a high index of suspicion.

Indications for surgery

Surgical correction of a partial anomalous communication should be undertaken in any patient with a ratio of pulmonary to systemic blood flow greater than 2:1 and without severe pulmonary hypertension that would preclude surgery. Isolated partial anomalous venous connection of the whole of one lung is also an indication for repair, since the opposite lung is the only one able to return oxygenated blood to the systemic circulation; should this lung become diseased later in life, fatal anoxia is a potential hazard.

Operative treatment

Cardiopulmonary bypass with or without circulatory arrest is always required. The anomalous flow of blood into the left atrium is usually corrected by the use of an intracardiac baffle, which is positioned to ‘tunnel' blood from the anomalous entry site via the atrial septum to the left atrium (Fig. 3). When the anomalous connection is to the left innominate vein, the anomalous vein can be transferred to the left atrial appendage.

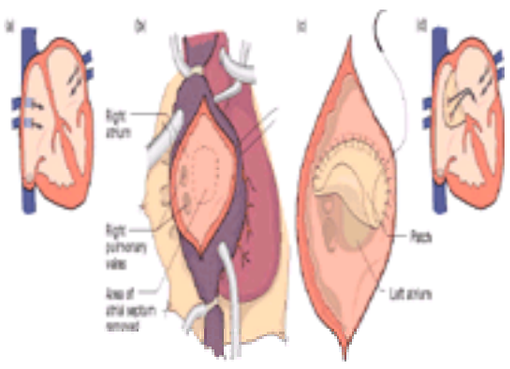


Fig. 3. Operative repair of partial anomalous pulmonary venous onnection: (a) anomalous drainage of the right lung into the right atrium; (b) first stage of operative repair showing cannulation for cardiopulmonary bypass—in this example, part of the atrial septum must be removed to create an atrial septal defect, which forms part of the pathway for the correction; (c) utilizing a patch of pericardium, the right pulmonary veins draining into the right atrium are baffled across the surgically created atrial septal defect; (d) resulting pulmonary venous drainage pathway following correction.

Results and prognosis

This is similar to that of uncomplicated atrial septal defect, hospital mortality being less than 1 per cent. Long-term survival is excellent when surgery is undertaken before the development of pulmonary hypertension, when the prognosis is poorer.

Cor triatriatum

This constitutes 0.4 per cent of congenital heart disease.

Morphology

The pulmonary veins drain into a single chamber, the so-called third atrium, which lies posterior to the left atrium. The two chambers share a common wall composed of a diaphragm of tissue with a central aperture. This aperture is only 3 mm to 1 cm in diameter, resulting in pulmonary venous obstruction.

Manifestations

The condition presents in the first few years of life with the symptoms and signs of pulmonary congestion: the narrower the aperture the earlier the onset and the more severe the symptoms. Echocardiography shows the abnormal chamber, the diaphragm, and the size of the aperture. Cardiac catheterization is performed if other abnormalities are suspected.

Treatment and prognosis

The prognosis of untreated patients depends on the size of the aperture, but since 75 per cent die in infancy, surgery is usually performed urgently. Cardiopulmonary bypass is established and the diaphragm is removed via an incision through the right atrium and atrial septum. Operative mortality depends on the degree of preoperative pulmonary venous congestion and pulmonary hypertension. Long-term results are excellent; there is a risk of pulmonary venous stenosis.

Pulmonary venous stenosis

In this rare condition one or more pulmonary veins are locally or diffusely narrowed, often near their junction with the left atrium. Associated cardiac abnormalities are the rule. Pulmonary venous congestion results, leading to right heart failure. Despite medical treatment, most patients die by the age of 4 years. Surgical therapy is unrewarding, particularly when stenosis is diffuse, and the condition often progresses postoperatively despite an initially good result. Percutaneous dilatation of localized pulmonary venous stenoses is now being undertaken; short-term results are encouraging, but the long-term benefits are unknown.

Anomalies of the systemic venous system

Abnormalities of the systemic veins are common, and most have no physiologic consequences. They may, however, complicate the placement of central venous cannulas and pulmonary arterial catheters; they may also modify the technique of venous cannulation for cardiopulmonary bypass, as well as some thoracic operations. Right-to-left shunting results when a systemic vein connects to the left atrium.

Persistent left superior vena cava

This vessel connects the left innominate vein to either the coronary sinus or the left atrium; the latter connection causes right-to-left shunting, which may be sufficient to lead to cyanosis. Operative treatment is advised to prevent systemic embolism. If a right-sided superior vena cava is also present, the abnormal left superior vena cava may simply be ligated; otherwise it must be transposed to the right atrium.

Unroofed coronary sinus

In this condition the coronary sinus communicates normally with the right atrium, but there is a defect in its common wall with the left atrium. The result is a left-to-right shunt similar to an atrial septal defect. Closure of the opening of the coronary sinus into the right atrium allows coronary sinus blood to drain into the left atrium. The

small right-to-left shunt that results is hemodynamically insignificant.

Abnormalities of the inferior vena cava

The inferior vena cava has complex embryologic origins and it may drain anomalously into either the azygous vein or left atrium. Only in the latter case is surgical correction indicated, for the same reasons as in a connection of the left superior vena cava to the left atrium.

Further reading

Buckley MJ, Behrendt DM, Goldblatt A, Laver MB, Austen WG. Correction of total anomalous pulmonary venous drainage in the first month of life. *Journal of Thoracic and Cardiovascular Surgery* 1972; **63**: 269–74. [Report of the first successful correction of infradiaphragmatic anomalous pulmonary venous return.]

Castaneda AR, Jonas RA, Mayer JE, Jr, Hanley FL. *Cardiac surgery of the neonate and infant*. Saunders, Philadelphia, 1994. [A good description of the contemporary approach to anomalous venous return, including diagnostic, surgical, and postoperative management aspects.]

Kirklin JW, Barratt-Boyes BG. *Cardiac surgery*, pp. 499–540. Wiley, New York, 1986. [An excellent chapter detailing the management of these lesions. Anatomy and pathophysiology is discussed. Follow-up data are presented, along with an analysis of natural history and surgical outcome data.]

Lamb RK, Qureshi SA, Wilkinson JL, Arnold R, West CR, Hamilton DI. Total anomalous pulmonary venous drainage. *Journal of Thoracic and Cardiovascular Surgery* 1988; **96**: 368–75. [Summary of a 17-year clinical experience with anomalous pulmonary venous return in 80 patients during an early period in this field (1968–1985).]

Oelert H, Schafers HJ, Stegmann T, Kallfelz HC, Borst HG. Complete correction of total anomalous pulmonary venous drainage: experience with 53 patients. *Annals of Thoracic Surgery* 1986; **41**: 392–4. [An early report of total correction of anomalous pulmonary venous return in 53 patients.]

Richardson JV, Doty, DB, Siewers RD, Zuberbuhler JR. Cor triatriatum (subdivided left atrium). *Journal of Thoracic and Cardiovascular Surgery* 1981; **81**: 232–8. [A report of 21 cases of cor triatriatum sinister encompassing a range of anatomic complexities. The surgical approach used depends on the age of the patient.]

Sano S, Brawn WJ, Mee RBB. Total anomalous pulmonary venous drainage. *Journal of Thoracic and Cardiovascular Surgery* 1989; **97**: 886–92. [A contemporary series of 44 patients (1979–1987) operated on early in life. Late complications are identified at long-term follow-up.]

Tucker BL, Lindesmith GG, Stiles QR, Meyer BW. The superior approach for correction of the supracardiac type of total anomalous pulmonary venous return. *Annals of Thoracic Surgery* 1976; **22**: 374–7. [A report detailing a surgical approach to supracardiac anomalous pulmonary venous return involving direct anastomosis between the left atrium and common pulmonary vein.]

40.7.10 Miscellaneous congenital cardiac defects

Arvind K. Agnihotri and Gus J. Vlahakes

Coronary arterial anomalies
Introduction
Coronary ostial anomalies
Anomalous origin of the left circumflex artery from the right aortic sinus or right coronary artery
Anomalous origin of the left main coronary artery from the right aortic sinus
Anomalous origin of the right coronary artery from left aortic sinus
Anomalous origin of the left anterior descending artery from the right aortic sinus
Anomalous origin of the left main coronary artery from the pulmonary artery
Origin of the right coronary artery from the pulmonary artery
Single coronary artery
Coronary anomalies associated with abnormalities of the cardiac chambers
Congenital coronary arterial fistula
Congenital aneurysms of the sinus of Valsalva
Ectopia cordis
Cardiac diverticula
Further reading

Coronary arterial anomalies

Introduction

At 3 weeks' gestation, the cardiovascular system appears as a pair of cardiogenic cords. These cords migrate in a cephalad direction and form a single heart tube. Early metabolic exchange takes place via sinusoids, which represent a prolongation of trabeculae characteristic of developing myocardium. Late in the fourth week of gestation, the coronary vasculature arises as isolated blood islands. These islands multiply over the remaining ventricle, and then coalesce to form a primitive vascular network.

At 5 weeks, coronary buds arise from the truncus arteriosus (which will later undergo septation and division to form the aorta and pulmonary trunk). The appearance of the coronary buds may be triggered by the approach of the vascular network, which develops first at the apex and progresses toward the truncus arteriosus. The luminal continuity between the proximal coronary buds and the distal epicardial vessels forms during the seventh week of gestation, resulting in the coronary circulation in its final configuration.

The incidence of important coronary arterial anomalies varies between 0.2 and 1.2 per cent, depending on whether the series is based on pathologic or on angiographic criteria. Anomalies of the coronary arteries may involve their origin, course, branches, or termination. With the increased use of cardiac angiography over the last 25 years, the recognition of variations in the coronary circulation has increased. The definition of what is a ‘normal' is subject to debate, particularly for a clinically insignificant variation found routinely. The following discussion focuses on major abnormalities with known clinical significance.

Coronary ostial anomalies

Two coronary ostia are normally present, one located at the right and the second at the left sinus of Valsalva; however, three or four coronary ostia are considered to be normal variants. Most commonly a conal branch arises from a third coronary ostium, rather than from its usual origin from the right coronary artery. This small, accessory conal artery is present in 30 to 50 per cent of normal hearts. The second most common cause for extra ostia results from the absence of a left main trunk; this occurs in about 1 per cent of normal hearts and usually gives rise to independent ostia for the left anterior and circumflex arteries.

Coronary ostia normally arise 1 cm or less above the aortic valve annulus. A high take-off refers to an ostium arising more than 1 cm above the annulus; a low take-off originates from below the annulus. These variations may complicate cannulation of the coronary arteries during cardiac catheterization or delivery of cardioplegia during cardiac surgery. Ostia located near an aortic valve commissure or in the posterior (noncoronary) sinus are abnormal.

Normally, the aortic opening of a coronary artery is equal to or greater than the diameter of the artery. Coronary ostial stenosis or atresia can occur if the ostium and possibly a few millimeters of the coronary vessel fail to canalize properly. This rare condition is potentially lethal: the severity of symptoms depends on the degree of stenosis and presence of collateral flow. If an affected infant requires surgery, bypass grafting should be performed. The subclavian artery is most commonly used, due to the limited success of saphenous vein bypass grafts in this age group.

Anomalous origin of the left circumflex artery from the right aortic sinus or right coronary artery

An origin of the left circumflex artery from the right coronary artery or from the right coronary ostium is the most frequent anomaly occurring in otherwise normal hearts. In one series, it was found in 26 of 45 patients (58 per cent) with angiographically confirmed coronary arterial anomalies. The distal course of the left circumflex artery follows its normal path posterior to the aorta and into the left atrioventricular sulcus. The left main coronary artery continues only as the left anterior descending artery with its normal distribution.

The anomaly is diagnosed by coronary angiography. An avascular region in the posterior lateral left ventricle during left coronary injection suggests an anomalous origin of the left circumflex artery, assuming that there is no proximal obstruction in the left circumflex artery. Page *et al.* have described the ‘aortic root sign', in which a profile view of the anomalous circumflex artery, coursing posteriorly behind the right sinus of Valsalva, is demonstrated during left ventriculography in a right anterior oblique projection.

This anomaly must be identified in patients undergoing valve surgery in order to prevent injury to the circumflex artery during valve replacement.

Anomalous origin of the left main coronary artery from the right aortic sinus

Anomalous origin of the left main coronary artery from the right sinus of Valsalva is classified on the basis of the course of the artery as it passes from its origin on to the left heart. It may run anterior to the pulmonary trunk, posterior to the aorta, between the aorta and pulmonary trunk (Fig. 1(a)), or within the interventricular septum beneath the right ventricular infundibulum. The right coronary artery arises normally from the right aortic sinus. This anomaly is noted in 8.9 per cent of patients with angiographically confirmed coronary arterial anomalies.

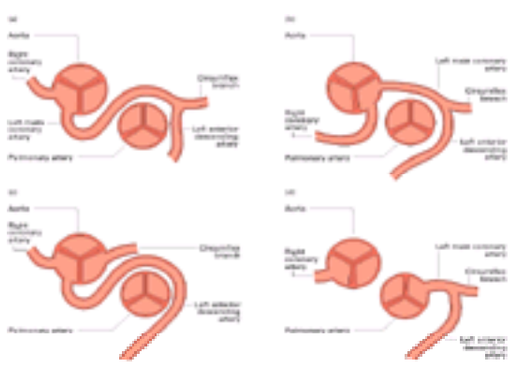


Fig. 1. Anomalies of the coronary arteries: (a) anomalous left main coronary artery arising from the right sinus of Valsalva and coursing between the aorta and pulmonary artery; (b) anomalous right coronary artery arising from the left sinus of Valsalva and coursing between the aorta and pulmonary artery; (c) anomalous left anterior descending artery arising from the right sinus of Valsalva and coursing between the aorta and pulmonary artery; (d) anomalous left main coronary artery

arising from the pulmonary artery.

When the anomalous left artery passes between the aorta and the pulmonary artery, myocardial ischemia or sudden death may occur. The mechanism is not clearly defined, but most sudden deaths occur after physical exertion: in one study, 13 of 16 patients (81.3 per cent) who died suddenly were engaged in, or had just completed, physical exercise. One suggestion is that as the left coronary artery arises from the right sinus of Valsalva it forms an acute angle at its origin, resulting in compromise of the arterial lumen as it courses left along the contour of the aorta. As cardiac output rises with exercise, the aorta expands and stretches the left main coronary artery, producing an occlusion of the slit-like orifice. Others have proposed that during exercise the left main coronary artery becomes compressed between the dilated aortic root and the pulmonary trunk, which is firmly anchored to the infundibular septum. The significance of hemodynamic compression between the great arteries is supported by the observation of a patient in whom the left main coronary artery arose from the proximal right coronary artery without an oblique take-off. However, there are doubts over whether a pulmonary artery under normal pressure can constrict an anomalous coronary artery perfused with systemic pressure. A symptomatic 14-year-old male with an anomalous left main coronary artery arising from the right sinus of Valsalva and traversing between the aorta and pulmonary artery underwent surgical enlargement of his left main orifice. He remained asymptomatic 9 years later, despite working in heavy construction.

This anomaly is more common in men than women, and almost all victims of sudden death are young males. Patients may present with exertional syncope, dizziness, and angina: symptoms which can mimic other cardiac abnormalities such as aortic stenosis. Physical findings such as murmurs and abnormal peripheral pulses can be helpful in the differential diagnosis. The resting electrocardiogram is normal in almost all young patients. A stress electrocardiogram should be carried to maximal effort to increase its reliability as a screening test for symptomatic patients. If evidence of myocardial ischemia is obtained, a coronary arteriogram should be performed. If a normal maximal stress electrocardiogram is obtained in a young male with definite exercise-induced symptoms, a coronary arteriogram should still be considered to rule out this coronary anomaly.

Young patients with an anomalous left main coronary artery originating from the right sinus of Valsalva and coursing between the aorta and pulmonary artery should undergo surgery to prevent exercise-induced myocardial ischemia and to reduce the high incidence of sudden death. Coronary artery bypass grafting to the left anterior descending artery and to the left circumflex artery relieves symptoms and objective evidence of myocardial ischemia. Surgical enlargement of a slit-like left coronary ostium may also relieve symptoms.

Anomalous origin of the right coronary artery from left aortic sinus

Both right and left coronary arteries arise from the left sinus of Valsalva in 27 per cent of patients with angiographically confirmed coronary arterial anomalies. The right coronary artery courses between the aorta and pulmonary artery with a normal distal distribution (Fig. 1(b)). The left coronary artery is normal, both in its origin from the left sinus of Valsalva and in its final distribution. Presenting complaints of dyspnea, dizziness, and exertional syncope are similar to those of patients with an anomalous left coronary artery arising from the right sinus of Valsalva and coursing between the aorta and pulmonary artery. The cause of ischemia appears to be either compression of the anomalous artery between the aorta and pulmonary artery during exertion, or occlusion of the slit-like orifice of the anomalous artery, which is stretched by the aorta during exercise. Although patients with this anomaly have a history of myocardial ischemia or myocardial infarction, sudden death was not seen in one autopsy series of 18 patients: surgical intervention in asymptomatic patients is not indicated. However, if the patient develops signs or symptoms of myocardial ischemia, an operation such as aortocoronary bypass is warranted.

Anomalous origin of the left anterior descending artery from the right aortic sinus

An anomalous left anterior descending artery can arise from the right sinus of Valsalva or the right coronary artery. It courses either between the aorta and the pulmonary artery (Fig. 1(c)) or anterior to the right ventricular infundibulum before finally reaching the anterior interventricular sulcus. The left coronary artery, arising from the left sinus of Valsalva, continues only as the left circumflex artery. In one series, this anomaly was found in 4.4 per cent of 45 patients with coronary arterial anomalies confirmed by angiography. Sudden death can occur when the left anterior descending artery courses between the aorta and pulmonary artery; consequently, aortocoronary bypass is recommended. Among patients with tetralogy of Fallot, the anomalous left anterior descending artery frequently courses anterior to the right ventricular infundibulum. It is important to recognize this anomaly before surgery, since the surgical approach may need to be altered.

Anomalous origin of the left main coronary artery from the pulmonary artery

This anomaly was found in 10 of 3800 patients catheterized for cardiac congenital disease. In this condition, the left coronary artery arises from the main pulmonary artery but its distal distribution remains normal (Fig. 1(d)). The right coronary artery is normal, both in its origin from the right sinus of Valsalva and in its distal distribution. In 1933, Bland *et al.* described a 3-month-old boy with ‘recurring attacks of dyspnea, pallor, and profuse sweating' in addition to an ‘acute discomfort precipitated by the exertion of nursing'. At autopsy the left main coronary artery was found to arise from the pulmonary artery; the right coronary artery was normal. This anomaly in infants has been referred to as the Bland–White–Garland syndrome.

The anomaly may present in infancy or in adulthood; the determining factor in the outcome is the degree of collateral circulation. In the infantile type, collateral circulation is inadequate, causing significant myocardial ischemia with resultant congestive heart failure. During fetal and neonatal life, the pressures within the pulmonary artery and aorta are approximately equal; consequently, the pulmonary artery perfuses the left main coronary artery. Early in life, the pulmonary arterial pressure begins to fall; at 12 months it is approximately one-quarter that of the aortic pressure. Survival depends on collateral channels forming between the right coronary artery and the anomalous left coronary artery, providing retrograde flow in the left main coronary artery, with arterial blood being supplied by the right coronary artery. If collateral circulation between the right and left coronary arteries is well established, the patient may survive to adulthood, despite the presence of a left-to-right shunt through the coronary circulation. The transition from antegrade to retrograde flow within the left main coronary artery is frequently fatal.

The patient is relatively asymptomatic in the first few weeks of life, but the progressive fall in the pulmonary arterial pressure produces symptoms such as angina-like episodes associated with feeding, paroxysmal crying, dyspnea, tachypnea, and restlessness at 2 to 3 months of age. Poor feeding results in minimal weight gain. Most patients deteriorate because of congestive heart failure and recurrent pulmonary infections. Mitral regurgitation can also occur due to ischemia or infarction of the papillary muscles.

Physical examination often reveals signs characteristic of heart failure such as cardiomegaly, hepatomegaly, and râles. Auscultation may reveal a non-specific systolic murmur at the base or an apical systolic murmur due to mitral regurgitation. The electrocardiogram frequently demonstrates Q waves and elevation of the ST segment in the lateral precordial leads, indicative of anterolateral infarction. The chest radiograph shows cardiomegaly and interstitial pulmonary edema.

Adult patients are frequently asymptomatic, and they may experience dyspnea or angina only during strenuous physical exertion. On auscultation, a continuous precordial murmur may be detected, as well as the apical pansystolic murmur of mitral regurgitation. The resting electrocardiogram is almost always abnormal, with changes in the ST segment. The chest radiograph may be normal, or may show cardiomegaly. In adults there is a high incidence of sudden death: this was the initial clinical manifestation in 8 of 10 patients who died from anomalous origin of the left main coronary artery from the pulmonary artery.

In both forms of this anomaly, cardiac catheterization with angiography is diagnostic. Injection into the aortic root shows a large right coronary artery with retrograde filling of the left main coronary artery through collateral vessels. Absence of retrograde flow through the left main coronary artery and into the pulmonary artery indicates poor collateral flow. In the infant, apparent absence of retrograde flow may be due to the persistence of elevated pulmonary arterial pressure and antegrade flow in the left coronary artery, particularly if there is associated mitral regurgitation.

The mortality rate in untreated patients is high. In a review of 58 patients in whom the left main coronary artery arose from the pulmonary artery, 46 (79 per cent) had died by 13 months of age, usually from congestive heart failure. The remaining 12 patients lived longer than 15 years, but eight of those died suddenly.

Simple ligation of the left main coronary artery produces a single coronary arterial system, with variable outcome. Current surgical treatment aims to reconstruct a two-coronary artery circulation. Most commonly, the left main coronary artery is directly reimplanted on the aorta after excising the ostium with a rim of pulmonary artery. Alternatively, a bypass can be performed distal to a ligation. Because of the poor patency of saphenous vein, most procedures in infants use subclavian artery as a bypass conduit. In a series of 39 surgical repairs that resulted in a two-coronary artery circulation, actuarial survival was 84 per cent at 10 years, but left ventricular dilatation persisted in 73 per cent of patients, and 39 per cent had persistent regional abnormalities of wall motion. Rarely, the left main ostium is arterialized by creating an aortopulmonary window and baffling it within the main pulmonary artery to the left main ostium.

Origin of the right coronary artery from the pulmonary artery

This anomaly is less common than anomalous origin of the left main coronary artery from the pulmonary artery. The right coronary artery arises from the right

pulmonary sinus, but its distal distribution is normal. The left coronary artery is normal in its origin and distribution.

Patients usually do not complain of angina; physical examination may only disclose a continuous murmur with diastolic accentuation. Definitive diagnosis is made by angiography: injection of the left coronary artery demonstrates retrograde filling of the anomalous right coronary artery through collaterals. Since cardiac arrest has been reported in patients with this anomaly, surgery is recommended. The procedure of choice is excision of the right coronary ostium with a rim of pulmonary artery and its reimplantation directly into the aorta. In older patients, proximal ligation and aortocoronary bypass may be performed.

Single coronary artery

A single coronary artery can follow the course of the normal right or left coronary artery, can divide into two branches with each branch following either the right or left coronary artery, or it may not correspond with any normal coronary arterial pattern. This anomaly was noted in two of nine patients with coronary arterial anomalies confirmed by angiography. Some patients are asymptomatic; others may present with congestive heart failure or sudden death. Diagnosis is confirmed by cardiac catheterization and angiography. In these patients, proximal coronary atherosclerosis is especially serious, since a second coronary artery is not present to contribute collaterals. Aortocoronary bypass is performed in symptomatic patients with confirmed coronary atherosclerosis.

Coronary anomalies associated with abnormalities of the cardiac chambers

Specific structural heart defects are associated with abnormalities of the coronary circulation, and these anomalies are discussed elsewhere (see Fig. 1, Chapter 40.7.7). Within the context of the present discussion a few principles are worth noting. (1) In general, the morphology of the developing ventricle dictates the coronary pattern. Thus, in ventricular inversion the coronary arteries are usually inverted and are typically appropriate to the chamber they supply. (2) Although ostial morphology is quite variable in congenital heart disease, the origin of the ostia tends to be from the aortic cusps, which are situated next to the aortopulmonary septum. Thus even in aortic atresia the coronaries tend to arise from the hypoplastic aorta (which may have reversed flow). (3) The left anterior descending artery is influenced by the development of the pulmonary conus. As such, hypoplasia of the pulmonary infundibulum in tetralogy of Fallot is associated with an increased chance of the left coronary artery arising as a branch of the right coronary artery.

In pulmonary atresia and in hypoplastic left ventricle, abnormal communications between the ventricular cavities and the coronary circulation are frequently seen. These ‘ventriculocoronary connections' usually result in reversal of the involved coronary flow during systole, thereby functioning as ‘venting' channels for the involved chamber.

Congenital coronary arterial fistula

This is a communication between a coronary artery and an atrium, ventricle, pulmonary artery, superior vena cava, or coronary sinus. During embryonic development, intratrabecular spaces of the myocardium communicate with the coronary arteries. Coronary arterial fistulas may form if a large intratrabecular space persists; this results in a connection between the coronary artery and a cardiac chamber. Fistulas involving the right heart structures are more common than those involving the left: fistulas usually arise from the right coronary artery and terminate somewhere on the venous side of the heart. The most common drainage site is the right ventricle, followed by the right atrium and pulmonary artery.

Many infants and children are asymptomatic; their fistulas are discovered only after a murmur is heard on a routine physical examination. Older patients (over 20 years of age) may present with fatigue, dyspnea, angina, and congestive heart failure. The symptoms vary, depending on the size of the left-to-right shunt and the degree to which myocardial flow is compromised. Patients may also present with other complications, including endocarditis, aneurysm formation, and rupture. Diagnosis is often suspected by echocardiography. Angiography confirms the lesion and defines the coronary circulation.

Successful surgical repair was first performed by Crafoord and Biorck in 1947. The goal of operative therapy is closure of the fistula without myocardial damage. Meticulous epicardial dissection and anatomic identification are essential. Recently, transcatheter embolization (coil embolization) has been used with increasing frequency in selected patients. Although early results suggest that this method is both safe and effective, the open technique is still preferred in patients needing concomitant operations, those with distal fistulas, when an adjacent coronary vessel is at risk, and in the very young.

Congenital aneurysms of the sinus of Valsalva

These are thought to arise from defective development of aortic elastic tissue above the annulus of the aortic valve. Aortic pressure gradually causes an aneurysm to form at the thinned aortic wall, most commonly at the right aortic sinus. The aneurysm is often described as having a ‘wind-sock' appearance. The aneurysm may rupture and form a fistula, usually into a low-pressure chamber such as the right ventricle or right atrium. The majority of patients have associated cardiac abnormalities, such as ventricular septal defects and aortic valvular anomalies.

Unruptured aneurysms of the sinus of Valsalva are rarely discovered before angiography, which is usually performed for associated cardiac lesions such as ventricular septal defects. Acute rupture results in sudden chest pain and dyspnea, as well as the appearance of a loud, continuous murmur. A widened pulse pressure may suggest aortic incompetence. Echocardiography can help confirm the diagnosis, while cardiac catheterization will define the fistula and any associated cardiac anomalies, as well the size of the shunt.

A combined surgical approach is recommended: the aneurysm is resected through the involved cardiac chamber and the fistula is closed through the aorta. Additional procedures, such as closure of a ventricular septal defect or reconstruction or replacement of an aortic valve, may be necessary.

Ectopia cordis

This is a rare malformation, occurring in about six per million live births, in which the heart is partially or completely outside of the thoracic cavity. There are four basic types of ectopia cordis classified according to the location of the heart: cervical, thoracic, thoracoabdominal, and abdominal. Cervical ectopia is found only in malformed fetuses, and only one case of true abdominal ectopia has been reported. In one review of ectopia cordis, 62.5 per cent of all cases were of the thoracic type. Thoracoabdominal ectopic cordis is associated with as Cantrell's pentad (or Cantrell's syndrome), which includes a defective abdominal wall, a bifid sternum, deficiency of diaphragm, deficiency of diaphragmatic pericardium, and intracardiac defects. This syndrome is thought to result from defective mesodermal development at 14 to 18 days of embryonic life.

The objective of surgical treatment is to replace the heart into the mediastinum. This operation may be complicated by the presence of other associated defects such as omphalocele, and diaphragmatic and sternal defects, which are present in thoracoabdominal ectopia cordis. Since the thoracic cavity is often too small to accept the heart, reconstruction of the chest wall may be necessary. In a recent review of the experience at Boston's Children's Hospital, five of 10 children who underwent attempted repair of thoracic or thoracoabdominal ectopia cordis survived beyond infancy, with four of the five ultimately having successful intracardiac repair.

Cardiac diverticula

Congenital diverticula can occur in any one of the four cardiac chambers and are defined as protrusion of saccular deformities of one or both ventricles. They usually arise in the apex, and consist of thinned myocardium, which may contain fibrous septations. In contrast to cardiac aneurysms, diverticula have all layers of the ventricular wall intact, and are capable of contraction. Atrial diverticula can cause arrhythmias, and if on the left side of the heart, they can be a source of cerebral emboli. Muscular ventricular diverticula can rupture or lead to congestive heart failure. Ventricular diverticula are frequently associated with other congenital abnormalities, such as midline thoracoabdominal defects. Surgical resection, usually with cardiopulmonary bypass, is the recommended therapeutic approach.

Further reading

Angelini P. Normal and anomalous coronary arteries: definitions and classification. *American Heart Journal* 1989; **117**: 418–34. [Comprehensive review of normal and congenitally abnormal coronary artery anatomy with 314 references.]

Cantrell JR, Haller JA, Ravitch MM. A syndrome of congenital defects involving the abdominal wall, sternum, diaphragm, pericardium, and heart. *Surgery, Gynecology, Obstetrics* 1958; **107**: 602–14. [Original report of multiple congenital cardiac defects that has come to be known as Cantrell's pentad or Cantrell's syndrome, including ectopia cordis.]

Carrel T, Tkebuchava T, Jenni R, Arbenz U, Turina M. Congenital coronary fistulas in children and adults: diagnosis, surgical technique and results. *Cardiology* 1996; **87**: 325–30. [A series of 11 patients with congenital coronary fistulae (age range 4–64 years), nine of whom were managed by surgery. Authors conclude that this lesion is managed best by surgery.]

Cheitlin MD, DeCastro CM, McCallister HA. Sudden death as a complication of anomalous left coronary origin from the anterior sinus of Valsalva. A not so minor congenital anomaly. *Circulation*

1974; **50**: 780–7. [Description of sudden death in patients where there is anomalous origin of the left coronary artery from the anterior sinus ofValsalva, emphasizing the potential importance of this coronary anomaly.]

Conte G, Pellegrini A. On the development of the coronary arteries in human embryos, stages 14–19. *Anatomy and Embryology* 1984; **169**: 209–15. [Excellent discussion of normal coronary artery development in human embryos.]

Davia JE, Green DC, Cheitlin MD, DeCastro C, Brott WH. Anomalous left coronary artery origin from the right coronary sinus. *American Heart Journal* 1984; **108**: 165–6. [A report of two cases of anomalous left coronary arteries arising from the right coronary sinus, one with surgical therapy and one with medical therapy.]

Fontana RS, Edwards JE. *Congenital cardiac disease: a review of 357 cases studied pathologically*. Saunders, Philadelphia, 1962. [A general discussion of anatomic pathology in congenital heart disease based on extensive necropsy study.]

George JM, Knowlan DM. Anomalous origin of the left coronary artery from the pulmonary artery in an adult. *New England Journal of Medicine* 1959; **261**: 993–8. [A case report of a left coronary artery arising from the pulmonary artery in an adult.]

Hornberger LK, Colan SD, Lock JE, Wessel DL, Mayer JE. Outcome of patients with ectopia cordis and significant intracardiac defects. *Circulation* 1996; **94** (Suppl. II): II-32–7. [A description of a series of 13 patients with ectopia cordis, 10 of which were managed surgically. The series emphasizes factors that may potentially affect outcome.]

Hutchins GM, Kessler-Hanna A, Moore GW. Development of the coronary arteries in the embryonic human heart. *Circulation* 1988; **77**: 1250–7. [A detailed study of coronary artery development derived from studies of specimens in the Carnegie Embryological Collection.]

Jokl E, McClellan JT, Ross GD. Congenital anomaly of the left coronary artery in young athletes. *Journal of the American Medical Association* 1962; **182**:572–74. [Early report of a 14-year-old athlete who died suddenly and was found to have a coronary artery anomaly.]

Kimbiris D, Iskandrian AS, Segal BL, Bemis CE. Anomalous aortic origin of coronary arteries. *Circulation* 1978; **58**: 606–15. [A review of 7000 coronary angiograms with the finding of 45 patients (0.64 per cent) with coronary artery anomalies. The relevant clinical course is reviewed.]

Kirklin JW, Barratt-Boyes B. *Cardiac surgery*. Wiley, New York, 1986. [Excellent discussions of lesions described in this chapter detailing surgical anatomy, pathophysiology, surgical treatment, and outcome. Extensively referenced with literature of interest to surgeons.]

Neufeld HN, Schneeweiss A. *Coronary artery disease in infants and children*. Lea and Febiger, Philadelphia, 1988. [General discussion of coronary artery disease in young patients that also includes congenital coronary anomalies.]

Page HL, Engel HJ, Campbell WB, Thomas CS, Jr. Anomalous origin of the left circumflex coronary artery. Recognition, angiographic demonstration and clinical significance. *Circulation* 1974; **50**: 768–73. [A review of 2996 coronary angiograms from which 20 patients were identified with anomalous origin of the circumflex circulation. Series emphasizes angiographic signs useful for detecting this anomaly.]

Roberts WC. Major anomalies of coronary arterial origin seen in adulthood. *American Heart Journal* 1986; **111**: 941–63.7 [Excellent and extremely well referenced review of the subject of coronary artery anomalies.]

40.8.1 Coronary artery disease: risk factors, diagnosis, and medical treatment

J. C. Forfar and B. Gribbin

Introduction
Risk factors for coronary artery disease
Blood lipids
Hypertension
Smoking
Other risk factors
The influence of risk factors on the development of the atheromatous plaque
Diagnosis of coronary disease—methods of investigation
Electrocardiography
Echocardiography and Doppler ultrasound
Radionuclide imaging and angiography
Further reading

Introduction

Coronary artery disease is the major cause of death in Western industrialized societies. In Britain it accounts for one in three deaths in men and one in four deaths in women. In the United States over 500 000 deaths annually are attributed to coronary disease. One half of all deaths from acute myocardial infarction occur within an hour of symptom onset, often before treatment can influence outcome. Morbidity statistics are less reliable, but the British Regional Heart Study, using a questionnaire and electrocardiography (ECG) in middle-aged men drawn from 24 British towns, showed that 8 per cent of individuals had angina and 15 per cent had definite or possible ECG evidence of ischaemic heart disease. The burden placed on the community and attendant economic and health care costs is enormous: in Britain coronary heart disease is estimated to be responsible for 12 per cent of all working days lost and the cost of treatment is over £700 million per year.

There is a marked difference in death rates due to coronary disease between countries; for example, a 10-fold greater age-standardized death rate for men aged 35 to 74 years in Scotland compared with Japan. Within Europe, threefold differences in death rates and disease incidence are seen with Finland and the United Kingdom higher than Italy, France, and Spain. Both environmental and genetic factors are relevant but individuals who migrate tend to acquire the risk characteristics of their new locality over a generation emphasizing lifestyle and exposure to risk factors over genetic susceptibility.

There are also marked contrasts in coronary disease mortality trends between developed and developing countries. In the United States, western Europe, and Australia, mortality has been falling between 15 and 50 per cent for at least 20 years. By contrast, rates continue to rise in eastern Europe, including Poland, Hungary, Bulgaria, and the Czech Republic. The reduction could be due to a fall in disease incidence or case fatality rates, or both. The management of acute myocardial infarction in particular has improved over this time, with case fatality rates halved, but there has also been an increased awareness of risk factor avoidance.

Risk factors for coronary artery disease

The three major modifiable risk factors are elevated serum cholesterol, high blood pressure, and cigarette smoking. Anyone doubles the risk of coronary heart disease but the presence of all three increases the risk by more than eight times. The magnitude of the risk and the impact of relative and absolute risk is age-dependent. For example, cigarette smoking in young males confers up to a fivefold increase in relative risk of myocardial infarction compared with threefold for the general population. Absolute risk, however, is lower in younger compared with older males. Other factors are presented in Table 1.

Elevated cholesterol (low-density lipoprotein)	
Hypertension	
Cigarette smoking	
Family history	
Other risk factors:	
	Age
	Male sex
	Diabetes
	Elevated triglycerides
	Obesity
	Hypothyroidism
	Renal disease
	Physical inactivity
	Psychosocial

Table 1 Risk factors for coronary artery disease

Blood lipids

Cholesterol is a necessary constituent of cell membranes, bile salts, and steroids. Triglycerides are used mainly as an energy supply. Both are transported in the plasma as complexes (lipoproteins), which can be classified according to their density. Low density lipoprotein (LDL) is the major cholesterol-carrying component of plasma and the main determinant of coronary risk. Very low density lipoprotein (VLDL) carries triglycerides and has been associated with coronary risk in most studies, although the magnitude of risk independent of other lipoprotein fractions is uncertain. High density lipoprotein (HDL) plays a part in the removal of cholesterol from cells and its transport to the liver and has an inverse relationship to coronary risk; it may be protective, at least in subjects with elevated LDL levels. Lipoprotein (a) has a similar protein and lipid composition to LDL but may confer specific risk for coronary risk in familial cases and may have thrombogenic as well as atherogenic effects.

In familial hypercholesterolaemia, there is a deficiency or absence of LDL receptors in liver cells. The heterozygous form affects about 1 in 500 of the population: these individuals lack 50 per cent of receptors and have serum cholesterol levels of 8 to 13 mmol/l. In the rare homozygous form, cholesterol levels may reach over 20 mmol/l; affected individuals suffer from accelerated atherosclerosis in the aortic root and coronary tree, which may lead to surgery in the first two decades of life and even heart and liver transplantation. However, the most common cause of elevated LDL cholesterol is not a single gene defect but a polygenic disorder in which LDL synthesis is increased and catabolism decreased. Renal disease, hypothyroidism, liver disease, alcoholism, dysproteinaemia, and steroid treatment may be secondary causes of hyperlipidaemia.

Epidemiological studies

The close relationship between serum cholesterol levels and the risk of developing coronary disease has been shown in over twenty prospective studies in different countries. The association in the Multiple Risk Factor Intervention Study is shown in Fig. 1. There is no lower limit for cholesterol as far as coronary risk is concerned with the same association in countries with low cholesterol and hence low absolute risk, such as China, as in those with higher levels and high absolute risk, such as the United Kingdom. Cholesterol is an independent risk factor, but the risk is magnified by the presence of hypertension or smoking.

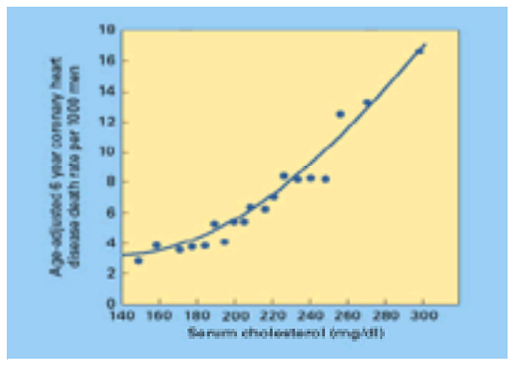


Fig. 1. Relationship between serum cholesterol and coronary heart disease in 361 662 men aged 35 to 57 years during an average follow-up of 6 years. Each point represents the median value for 5 per cent of the population.

Clinical trials of cholesterol lowering

Although trials involving dietary change alone have been inconsistent with regard to coronary event reduction, studies with powerful lipid-lowering drugs (3-hydroxy-3-methylglutaryl coenzyme A or HMG CoA reductase inhibitors; ‘statins') have consistently shown that mortality, the incidence of coronary events, and the need for coronary interventions is reduced by at least 25 per cent in patients with established disease. The Scandinavian Simvastatin Survival Study (4S) involved 4444 patients with coronary disease and showed a 37 per cent reduction in coronary events and 42 per cent reduction in coronary deaths. The Cholesterol And Recurrent Events trial (CARE) using pravastatin showed similar results in patients at somewhat lower absolute risk with lower LDL levels. Total mortality in this study was not reduced but stroke was reduced by 31 per cent. Data from the Long-term Intervention with Pravastatin in Ischaemic Disease trial (LIPID) confirms these findings. These important studies have finally proved the lipid hypothesis with regard to secondary prevention and have substantially lowered the threshold for prescription of such drugs in high-risk patients. These studies have overshadowed some 17 earlier studies using coronary angiography as a surrogate endpoint for cholesterol lowering therapy. Coronary stenoses in patients on lipid lowering therapy progress less rapidly and rarely regress compared with controls but the principal protective action is plaque stabilization with reduction in risk of plaque rupture and formation of local thrombus.

Although a significant fall in overall mortality in primary prevention studies had not been clearly demonstrated prior to 1995, analysis of pooled data from primary prevention trials suggested benefit in terms of coronary event reduction. The West of Scotland Coronary Prevention Study in 6595 men at relatively high risk from coronary disease (mean cholesterol 7 mmol/l) showed that over a 4.9-year follow up pravastatin reduced all cause mortality by 22 per cent, cardiovascular mortality by 32 per cent, and myocardial infarction (fatal and non-fatal) by 31 per cent.

A concern of both primary and secondary prevention studies with lipid-lowering drugs in the early 1990s was a possible increase in cancers and violent death in the treated groups. All four of the recent major intervention studies have not shown such an association, making statistical chance the most likely explanation for the earlier findings.

It may be estimated that treating 1000 patients with documented coronary disease and cholesterol above 6 mmol/l for 5 years with a statin will save 150 cardiovascular events, including 37 fatal and nonfatal infarcts and 62 coronary interventions (bypass and angioplasty).

Hypertension

Hypertension is an independent risk factor for stroke, cardiac failure, and coronary artery events. Meta-analysis of intervention trials has shown that for a treatment-related fall in diastolic blood pressure of 5 mmHg, the incidence of stroke was reduced by 38 per cent and coronary heart disease by 16 per cent in patients with mild hypertension. While the risk of stroke attributable to hypertension is totally reversed, that for coronary heart disease is partially so. The reasons for this difference are still controversial and include detrimental effects of thiazide diuretics on blood lipids, glucose tolerance and potassium, effects of b-blockers on lipid metabolism, and possible increases in coronary risk in patients with established coronary disease whose diastolic pressure is lowered below 80 mmHg. It is likely that the delay between blood pressure reduction and alteration in the natural history of coronary atheromatous disease will reduce the ability of trials lasting only a few years to identify changes in clinical events.

The elderly also benefit from blood pressure reduction. They may present with systolic hypertension because of decreased arterial compliance with ageing and impaired baroreflex sensitivity and treatment may be difficult because of increased sensitivity to drug side-effects particularly postural hypotension and compliance difficulties. Thiazide diuretics appear to be the drugs of first choice.

Smoking

Cigarette smoking is estimated to be responsible for one in five deaths of which approximately half are from cardiovascular disease. The Multiple Risk Factor Intervention Study showed that for a given level of cholesterol, cigarette smoking increased coronary death rates. It has a number of potentially hazardous effects, including activation of the sympathetic nervous system with enhanced platelet reactivity, impaired endothelium-mediated vasodilatation, and a carbon monoxide-induced fall in the oxygen carrying capacity of the blood. Women over the age of 35 years who take the contraceptive pill and smoke are at an increased risk of having a myocardial infarction. However, the effect of cigarette smoking is largely one of increasing the risk in subjects made susceptible by other risk factors, especially elevated cholesterol.

Cessation of smoking reduces the overall mortality rate progressively over ten years towards that of the age-matched non-smoker. Most of the excess risk has abated by two years. Following myocardial infarction, there is a 25 to 50 per cent reduction in mortality and half the risk of recurrent infarction in smokers who stop, compared with those who continue.

Other risk factors

These are partly associated with the major risk factors already described. Clustering of risk factors in an individual or a group is common. Obesity, particularly centripetal obesity, is a clinical manifestation of the coronary-prone individual; however, its reputation as a risk factor is largely due to its association with hypertension, diabetes, and elevated cholesterol levels. Obesity has not been clearly defined as a major independent risk factor.

Non-insulin-dependent and insulin-dependent diabetics, and those with impaired glucose tolerance are at increased risk of coronary artery disease. Hyperinsulinaemia is associated with increased coronary risk but partly through its association with hypertension, hypertriglyceridaemia, low HDL cholesterol, glucose intolerance, and obesity (insulin resistance).

Haemostatic factors associated with coronary risk include fibrinogen, Factor VII, and Factor VIII. Fibrinogen is strongly related to tobacco consumption and dietary fat to Factor VII levels.

A diet rich in saturated fat correlates with coronary event rates both within and between countries. In the Seven Countries Study, dietary fat was the most important factor explaining the geographical variation in coronary disease. Low intakes of polyunsaturated fat increased the risk of coronary disease. Intake of monounsaturated fat, such as a diet rich in olive oil, was negatively correlated with coronary disease. A family history of premature coronary disease may be partly related to environmental factors such as a diet rich in saturated fats and smoking. Having a first-degree relative with clinical coronary disease below 55 years confers a significant risk. Females are at lower risk than males, at least up to the menopause. This may be related to a higher level of HDL cholesterol in women, but the difference has become less marked, possibly due to an increase in the prevalence of smoking among younger women.

The role of triglycerides as a separate risk factor is controversial, but an elevated level may be associated with other lipid abnormalities and other risk factors. Moderate habitual exercise leads to a fall in blood pressure and a rise in HDL cholesterol; this has been shown to reduce the risk of cardiac events in American college graduates who retained an active life-style compared with their more sedentary contemporaries. Further support for the protective effect of exercise comes from a large British study in middle-aged civil servants followed for a mean period of 9 years. This showed that to be effective as a means of reducing the risk of fatal and non-fatal myocardial infarction exercise had to be moderately intense and current; a past history of vigorous exercise was not protective. Older men (aged 55 to 64 years) also

benefited, and required less vigorous exercise.

Limited alcohol use (about one drink daily) has been shown to exert a preventative benefit.

The influence of risk factors on the development of the atheromatous plaque

LDL cholesterol is taken up by cells via LDL receptors. Cholesterol is removed from the interstitial tissues by the lymphatics and from cells by HDL cholesterol. The arterial intima absorbs LDL particles through the endothelial layer, and has no lymphatic drainage; the concentration of LDL cholesterol is therefore greater in the intima than in most other tissues. LDL particles not taken up by intimal cells are modified and oxidized, forming aggregates which are engulfed by macrophages and may be transformed into foam cells. Accumulations of foam cells can be seen macroscopically even in the first decade of life, and are known as fatty streaks.

The further development of the atheromatous plaque is determined by a series of complex interactions between endothelial cells, lipoproteins, platelets, smooth muscle cells, and collagen. The combination of a high level of circulating LDL cholesterol, low HDL cholesterol, and endothelial injury, possibly triggered by factors such as hypertension, smoking, or simply as a result of a distorted pattern of blood flow, may increase LDL entry into the intima, overwhelming the foam cells. These die, releasing their cholesterol core. Modified LDL particles accumulate in the deeper intimal layer as lipid droplets which eventually coalesce into the lipid pool of the mature atheromatous plaque. Platelets play an important part as they aggregate on areas of endothelial damage, leading to release of vasoconstrictive and growth factors which promote smooth muscle cell proliferation and collagen deposition. Some plaques are predominantly fibrous, and others lipid laden. The end result is often a plaque containing a lipid pool core bordered by smooth muscle cells and foam cells and covered by a fibrous tissue cap (Fig. 2).

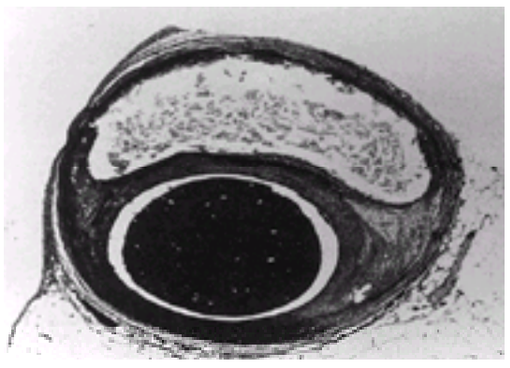


Fig. 2. Photomicrograph of a histological transverse section of an eccentric stenosis. The lumen contains gelatine/barium and the lipid pool is an apparently open space (upper portion) within the intima. The pool is separated from the lumen by the fibrous cap. (Reproduced by courtesy of Professor M. Davies.)

The characteristics of the plaque may be important in determining the nature of the patients' symptoms and the risk of an acute ischaemic event. Plaques vulnerable to local thrombosis include those with a large lipid core and thin fibrous cap. Fissuring of the cap allows blood to enter the lipid core forming intraplaque thrombus and possible (but not inevitable) luminal thrombus. Plaques that are largely fibrous are at lower risk f triggering local thrombus. Endothelial cells may also be removed from the surface of an artery allowing formation of a platelet monolayer and the risk of thrombus propagation. Plaques may be eccentric or concentric, and vary in their lipid/smooth muscle cell/fibrous tissue composition. In arteries with eccentric plaques, part of the wall circumference is healthy: this may be prone to dynamic change in cross-sectional area causing intermittent narrowing of the lumen and angina which is not consistently related to effort. The concentric plaque is likely to be rigid with fixed luminal narrowing; this gives a more consistent pattern of symptoms related to an increase in myocardial oxygen demand.

The majority of cases of unstable angina and myocardial infarction are triggered by disruption of the fibrous cap. This is most likely to occur at the edge of the cap at its junction with the relatively healthy wall and in plaques which have a large unsupported lipid core. If the split is superficial, thrombosis may follow platelet aggregation and release of vasoconstrictive factors, creating a subtotal occlusion of the artery and the clinical features of unstable angina. However, if there is deep exposure of collagen and the lipid-rich pool, thrombus formation within the intima extends into the lumen and is more likely to cause total occlusion and myocardial infarction.

Coronary angioscopy in patients with unstable angina has confirmed autopsy findings: plaque rupture with superimposed non-occlusive thrombus at different stages of maturity. Each deposit may be related to periods of pain and with microinfarcts in the distal coronary bed, suggesting embolization. Knowledge of these arterial changes has alerted cardiologists to the appearances of irregular undercut lesions and filling defects due to thrombus commonly seen during coronary angiography in unstable angina patients.

Angiographic studies show that the stenosis present before plaque rupture occurs need not be severe; this explains why asymptomatic and active subjects can quite suddenly experience unstable angina, myocardial infarction, or sudden arrhythmic death. Other studies in patients with coronary disease who have died of non-cardiac causes show that plaque fissuring may not be uncommon and in only a minority is associated with thrombotic occlusion. The fibrous cap can heal with or without spontaneous lysis of thrombus; incorporation of intimal thrombus into the new plaque structure determines whether the resulting stenosis is unchanged or progresses.

Diagnosis of coronary disease—methods of investigation

Electrocardiography

The resting ECG may be normal in patients with widespread coronary disease. Minor abnormalities must be interpreted in the light of the patient's history: ST segment depression and T-wave flattening or inversion are consistent with ischaemia but also occur with hypertensive heart disease, after digoxin treatment and during hyperventilation. Pathological Q-waves on the ECG is firm evidence of coronary disease as long as a cardiomyopathy has been excluded.

ST segment depression recorded fortuitously during pain confirms the diagnosis of coronary disease; ST segment elevation implies severe proximal stenosis of a coronary artery, rarely isolated coronary spasm, or an evolving myocardial infarction.

During infarction, the ECG usually shows the classical changes of ST segment elevation in the distribution of the infarct, deep T-wave inversion, and evolving loss of R-wave height progressing to Q-waves. Cardiac arrhythmias may be life threatening, such as ventricular fibrillation and ventricular tachycardia, potentially serious, such as atrioventricular heart block or atrial fibrillation or usually benign, such as sinus bradycardia, ventricular ectopics, and idioventricular rhythm. The initial ECG is occasionally normal, but serial records will show change, and extending the lead system to V7 to V9 may show infarct changes in the posterolateral territory. Patients with inferior wall infarcts should have right-ventricular leads recorded to look for evidence of right-ventricular infarction. Patients who do not develop Q-waves tend to have smaller infarcts.

Ambulatory ECG monitoring for 24 h or longer has an important role in detecting cardiac arrhythmias but may also be used to detect ST segment shifts in patients thought to have vasospastic angina or episodes of silent ischaemia. In exceptional circumstances, such as unexplained syncope with a major suspicion of a brady- or tachy-arrhythmia, implantable ECG monitoring devices may be justified. Up to 50 per cent of such cases prove *not* to need specific anti-arrhythmic therapy such as a permanent pacemaker.

Exercise testing

This allows the likelihood of obstructive coronary disease to be assessed, functional capacity to be determined and prognosis to be estimated. It is of fundamental importance in the management of patients presenting with anginal chest pain and after myocardial infarction.

Several ECG changes are indicative of exercise-induced ischaemia. The most common is depression of the ST segment by at least 1 mm (0.1 mV) 80 ms after the J point (this is the junction of the QRS complex and ST segment). Horizontal or down-sloping ST depression is more specific for coronary disease than is up-sloping (Fig. 3). ST segment elevation usually occurs over an area of myocardial damage associated with akinesia or aneurysm formation, but represents induced ischaemia in the absence of Q-waves.

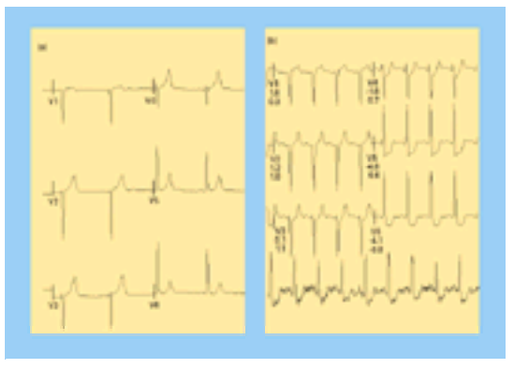


Fig. 3. ECG leads V1 to V6 showing marked ST segment depression during exercise: (a) before exercise and (b) at peak exercise.

Other findings related to myocardial ischaemia also need consideration. These include the level of exercise achieved, the systolic blood pressure response, the heart rate and workload at the onset of ischaemic changes, and symptoms of chest pain, breathlessness, and faintness. Exercise-induced ST depression can occur in hypertensive patients without angiographic evidence of coronary disease and in those with mitral valve prolapse. Digoxin may exaggerate ST segment changes and anti-anginal drugs such as b-blockers and calcium antagonists may restrain myocardial oxygen demand sufficiently to mask ischaemic changes which otherwise would be evident at low workload.

Exercise testing as a *diagnostic* test

The prevalence of coronary disease increases substantially as the history of chest pain becomes more typical of angina and in men more than women. There is, therefore, little diagnostic advantage in performing exercise tests in men with classical angina. The same is true of women with non-specific chest pain, in whom the likelihood of coronary disease is very low and the predictive value of a positive exercise test is only 10 per cent. Table 2 shows the average predictive value of positive and negative exercise tests in patients with different clinical presentations. For diagnostic purposes exercise testing is most useful in women with typical angina and patients with atypical angina. Women in the latter category with a positive result may, however, need to undergo additional tests such as an exercise thallium study (*vide infra*) in order to assess more accurately the post-test risk of coronary disease. Patients with marked ECG abnormalities early in the exercise period are more likely to have coronary disease.

	Men *Predictive value (per cent) of test*		*Women* *Predictive value (per cent) of test*	
	Positive	*Negative*	*Positive*	*Negative*
Typical angina	95	40	80	70
Atypical angina	85	75	55	90
Non-specific pain	40	90	10	95
Asymptomatic	30	95	5	‹99

Figures modified from those of Chaitman (1986).
Predictive value of positive test is TP/TP+FP × 100 and of a negative test TN/TN+FN × 100 where TP=true positive (i.e. positive exercise ECG and angiogram); FP=false positive (i.e. positive exercise ECG and negative angiogram); TN=true negative (i.e. negative exercise ECG and angiogram); FN=false negative (i.e. negative exercise ECG and positive angiogram).

Table 2 The predictive value of positive and negative exercise tests

The exercise test as an indicator of *prognosis*

The prognosis of patients with coronary disease is determined largely by left-ventricular function and the severity of coronary disease. Patients who develop marked ST segment depression of 2 mm or more at a low level of exercise (unable to complete stage 2 of the Bruce protocol) or at a low heart rate have almost a 90 per cent chance of having multivessel or left mainstem coronary disease, and hence a greater risk of *early* death. Loss of the normal blood pressure rise with exercise also places the patient in a high-risk group, especially if accompanied by ECG and clinical signs of severe ischaemia. Follow-up studies of patients with three-vessel coronary disease and normal or mildly impaired left-ventricular function show that survival is related to the level of exercise achieved. The 4-year mortality rate is 47 per cent in those unable to complete level 1 of the Bruce protocol (Fig. 4).

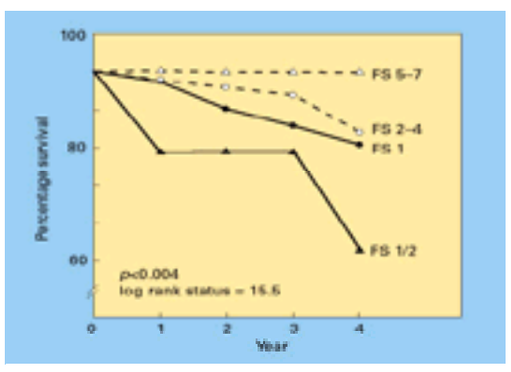


Fig. 4. Cumulative survival rates according to the final exercise stage (FS) achieved by 572 patients with three-vessel coronary artery disease and normal or mildly impaired left-ventricular function.

In any individual patient most prognostic information can be derived from knowledge of left-ventricular function, obtained by clinical examination or investigation, the peak exercise level attained, and the response of the ST segment.

Echocardiography and Doppler ultrasound

Echocardiography is a non-invasive investigation that can be applied serially at the bedside, and has many uses in patients with presumed coronary artery disease: (a) the dimensions and function of heart chambers can be studied; (b) regional wall motion abnormalities due to ischaemia or infarction can be detected; and (c) mechanical complications of myocardial infarction such as rupture of the interventricular septum and acute mitral regurgitation can be identified. It also helps in the recognition of transient myocardial ischaemia when carried out before, during, and immediately after stress testing using exercise or intravenous inotropic and chronotropic agents. Regional wall motion abnormalities may develop *de novo* or improve and again deteriorate according to the balance between myocardial oxygen supply and demand.

Radionuclide imaging and angiography

To complement standard exercise testing, radionuclide imaging techniques provide information about ventricular function, the extent and the functional significance of coronary stenoses; and aid the identification of viable but non-functioning myocardium.

At present, the main indications for perfusion imaging are (a) a low likelihood of coronary disease on clinical evaluation, but a positive exercise test; (b) a high pretest likelihood of coronary disease, but a negative exercise test; (c) a resting ECG, which is unsuitable as a marker of ischaemia (for example, digoxin effects, left bundle branch block, and pre-excitation); and (d) an inability to exercise. Thallium-201 or SestaMIBI are radioactive tracers avidly taken up by myocardial cells: their distribution is related to myocardial perfusion and cell viability. Injection intravenously at peak exercise allows a two-dimensional or preferably tomographic picture to be obtained using a gamma camera. Further imaging 3 to 4 h later allows comparison after redistribution of the isotope. Defects that resolve over time are considered to represent ischaemic but viable myocardium, while persistent defects indicate irreversible damage, or at least severe myocardial ‘stunning' (Fig. 5). Late redistribution of isotope (up to 24 h) seen by reinjection may further help to separate infarcted from viable myocardium.

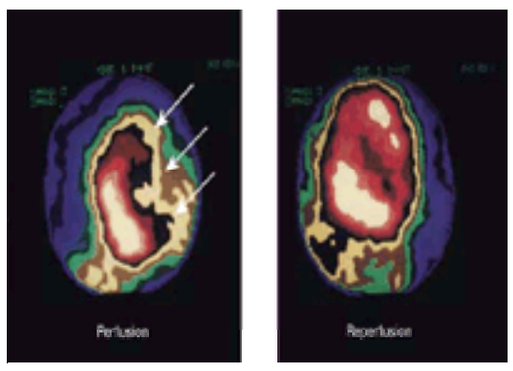


Fig. 5. Thallium perfusion scintigraphy on exercise (left; perfusion) and recovery (right; reperfusion) showing reversible myocardial ischaemia. A large area of low isotope uptake on exercise (arrowed) fills in on recovery showing ischaemia of the anterior left-ventricular wall. Right oblique projection.

Similar representation of myocardial perfusion can be achieved without exercise by injecting suitable tracers shortly after the intravenous injection of dipyridamole, which causes dilation of coronary arterioles. Redistribution of thallium uptake then defines areas of comparatively reduced perfusion. This is useful in patients who are unable to exercise. Infusion of adenosine has also been used as a coronary vasodilator, its use to some extent limited by minor side-effects, particularly chest discomfort, headache, and flushing.

Thallium-201 or SestaMIBI imaging is more sensitive and specific than standard exercise testing for the diagnosis of coronary disease but is considerably more expensive. However, when applied to selected patient groups it extends the evidence for or against the presence of coronary disease and determines the need for coronary arteriography. Detection of multiple or large perfusion defects indicates a high risk of cardiac morbidity and mortality.

Another radionuclide technique involves imaging of left-ventricular global and regional wall motion at rest and during supine exercise. Intravenous technetium-99m is used to label the blood pool and exercise-induced ischaemia detected by its effect on left-ventricular ejection fraction and regional wall motion. Patients with coronary disease whose ejection fraction falls with exercise have a higher incidence of cardiac events. Dobutamine can be used as a stress agent in those patients unable to exercise. In practice, radionuclide angiography is often used to gauge the functional relevance of known coronary disease and to justify an intervention on prognostic grounds.

Coronary arteriography

Coronary arteriography is usually combined with pressure measurement in the left ventricle, and single or biplane ventriculography. Performed percutaneously under local anaesthesia from the femoral artery or through a brachial arteriotomy, it defines the distribution and severity of coronary disease, collateral flow patterns, and the patency of saphenous vein, internal mammary, and other arterial bypass grafts.

Procedural risk is related to a number of variables, particularly the clinical state of the patient: complication risk is increased in those with angina at rest or pulmonary oedema and in those with severe left mainstem disease. In patients studied electively at one large British centre emergency revascularization was necessary in 0.24 per cent of patients undergoing coronary arteriography. Overall mortality should not exceed 0.1 to 0.2 per cent. Although accepted as the final arbiter of coronary disease, coronary angiography tends to underestimate disease severity compared with autopsy data, especially with lesions which appear to show a diameter stenosis of under 60 per cent.

Coronary arteriography is rarely performed solely for diagnostic purposes. Exceptions are those whose occupation may depend on the result, such as airline pilots, and where the specificity of less invasive diagnostic techniques is reduced. The main indications are: (a) to assess patients with limiting or unstable angina for revascularization; (b) to evaluate more accurately the prognostic relevance of positive non-invasive tests; and (c) to identify further jeopardized myocardium after myocardial infarction. Guidelines for coronary arteriography have been published with rates varying between 1000 and 4500 per million population per annum in the Western world.

Diagnosis and treatment of stable angina

Stable angina is a sensation of discomfort, often aching, gripping, or crushing in character, usually felt retrosternally but sited anywhere from the lower jaw to the epigastrium, including the arms and upper back. It is related to effort and emotional upset and eased by rest within seconds or minutes. In some cases the limiting factor is less clear and the patient may have difficulty in describing it other than a feeling of vague unease. Some patients may show the ‘warm-up' phenomenon with typical discomfort disappearing as the patient continues exercise and in others dyspnoea may be an anginal equivalent. The probability of an individual patient having coronary disease increases as the history becomes more typical of angina and with the unmasking of risk factors. Significant differences exist between males and females (Fig. 6).

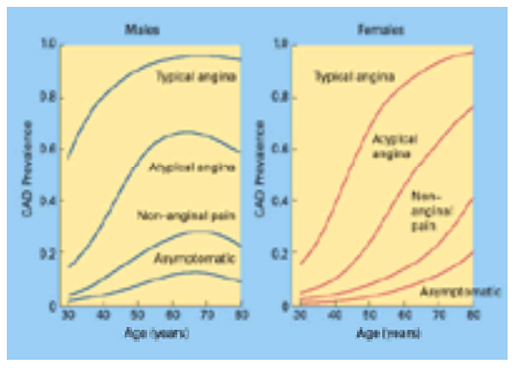


Fig. 6. Prevalence of coronary artery disease for males and females according to clinical history.

Examination may be normal, or provide supportive evidence of coronary disease, including lipid deposition such as tendon xanthomas, peripheral vascular disease, and hypertension. Auscultation of the heart may reveal a prominent fourth heart sound and patients with previous myocardial infarction may have features of cardiac failure.

The medical management of patients with stable angina aims: (a) to relieve symptoms and improve effort tolerance by means of drug treatment; (b) to identify high-risk subjects so that treatment designed to improve survival can be given; and (c) to advise on and treat risk factors in the expectation that disease progression will be slowed.

b-Blocking drugs are the mainstay of treatment. They act by reducing heart rate, cardiac contractility, and blood pressure, and thus myocardial oxygen consumption. They therefore redress the imbalance between demand and supply which causes ischaemia. They extend the duration of angina-free exercise and reduce the frequency and severity of asymptomatic ischaemia detected by ambulatory ECG monitoring. Side-effects consist mainly of loss of vigour and cold sensitivity. Contra-indications include sinus node disease, atrioventricular heart block, and asthma. Heart failure is often a contraindication but cautious use of b-blocking agents is gaining increasing acceptance in patients with reduced left-ventricular function, angina, and ventricular arrhythmias.

Calcium antagonist drugs include those with a predominant myocardial action such as verapamil and dihydropyridines such as nifedipine which have a largely peripheral effect. Verapamil and diltiazem are effective anti-anginal drugs when used singly but nifedipine should be prescribed with a b-blocker. Calcium antagonist drugs can exacerbate cardiac failure and cause flushing, ankle oedema, and headache.

A nitrate drug should be tried in most patients with angina and, if tolerated, continued in one of its many forms. The increased use of long-acting preparations has uncovered the phenomenon of nitrate tolerance. Tolerance can develop after 24 h of chronic exposure: preparations should therefore be prescribed to cover the period of maximum need but also to give adequate drug-free intervals. A twice daily regimen for isosorbide mononitrate avoids tolerance. Aspirin reduces the risk of myocardial infarction in patients with stable angina.

In chronic stable angina the severity of discomfort bears little relationship to the severity of coronary disease. As risk can be stratified by exercise testing most patients with this diagnosis should undergo a standard exercise test or radionuclide study. If this is impractical, those with a recent history of angina, those with a reduced exercise capacity compared with others of their age, and those with a worsening pattern of angina should be referred for consideration of coronary angiography. Some of the latter patients may have unstable angina, where exercise testing is inappropriate. Those with symptoms not adequately controlled by medication or at high risk of cardiac events as determined by exercise testing should undergo coronary arteriography and assessment for coronary intervention.

Diagnosis and treatment of unstable angina

This forms a clinical spectrum between stable angina and myocardial infarction, often called the acute coronary syndrome. There is a 10 to 15 per cent risk of progressing to myocardial infarction or sudden death. It can be defined as angina occurring with minimal effort and at rest but without a significant rise in cardiac enzymes. These patients may have pathological evidence of plaque fissuring and non-occlusive thrombus, and specialist advice and hospital care are essential.

Clinical management consists of antithrombotic treatment, administration of anti-anginal drugs, and early coronary arteriography if symptoms recur or if non-invasive testing after initial symptom control predicts high cardiac risk. Angioplasty or surgery may be required in some of these patients, such as those with recurring episodes of pain at rest with ECG changes despite adequate medical treatment; those with haemodynamic deterioration (e.g. left-ventricular failure or hypotension) during bouts of pain, and those with evidence of persistent ischaemic ECG changes in the absence of pain. Patients without these features should initially receive medical treatment followed by coronary arteriography if angina occurs during ward activity or if a predischarge exercise test proves to be positive. Routine coronary angiography in all patients with unstable angina is not indicated. Medical treatment of unstable angina reduces the risk of infarction and improves survival. Aspirin treatment reduces non-fatal myocardial infarction and death by approximately 30 per cent (Fig. 7). Intravenous unfractionated or subcutaneous low molecular weight heparin confers some additional benefit. b-Blockers reduce the risk of recurrent chest pain and early infarction but nifedipine does not, and may be detrimental unless given in conjunction with a b-blocker.

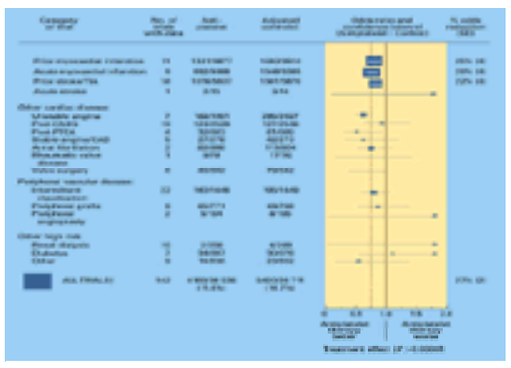


Fig. 7. Proportional effects of antiplatelet therapy on vascular events in 142 trials in some 83 000 patients subdivided by trial type. CABG = coronary artery bypass grafting; PTCA = percutaneous transluminal coronary angioplasty; CAD = coronary artery disease.

Patients with unstable angina should be admitted to a coronary care unit, given aspirin and, in the absence of contraindication a b-blocking drug such as metoprolol or atenolol with oral nitrates. Nifedipine can be added for patients already receiving b-blockers. Recurrent or continued pain should be treated with heparin and consideration of intravenous nitrates. This regimen relieves symptoms in the majority of patients with unstable angina. If it does not, coronary arteriography is appropriate. Heparin should be administered for at least 48 h after the last bout of chest pain and aspirin treatment should be long term. Patients who deteriorate haemodynamically may be stabilized by insertion of an intra-aortic balloon pump. The place of newer agents such as platelet glycoprotein 11b/111a receptor antagonists in unstable angina and coronary angioplasty is as yet uncertain but their role in coronary angioplasty and coronary stenting holds increasing promise.

In those patients with limiting chronic stable or persistent unstable angina, coronary angioplasty with or without intracoronary stent insertion is a highly effective treatment option. It has largely replaced coronary artery bypass grafting for single and double vessel coronary artery disease, provided lesions are technically suitable. More than 500 000 such procedures were performed in the United States of America in 1997. Threatened acute closure of angioplastied vessels, usually related to coronary dissection, has been largely solved by the development of coronary stenting. Restenosis of native coronary arteries (20 to 50 per cent) and stented coronary arteries (10 to 30 per cent) after angioplasty is related to vessel recoil and intimal hyperplasia. It remains a major problem but techniques including debulking procedures, stent coating, and local radiation therapy offer some promise for resolution of this important limitation.

Diagnosis and treatment of acute myocardial infarction

The patient presenting early with myocardial infarction usually has severe chest pain and appears anxious and frightened. Sweating and vomiting are common. Other physical signs are determined largely by vagal activity and by the size of the evolving infarct, but a wide area of impulse activity over the precordium is common in patients with anterior wall infarction and raised jugular venous pressure with right-ventricular infarction. There may be a third heart sound and a fourth is commonly present. Within 24 h, a pericardial friction rub may be audible.

The ECG usually confirms the diagnosis; a chest radiograph may show pulmonary congestion and echocardiography is very useful in confirming a wall motion abnormality early in the course of the illness, especially if the ECG is not diagnostic.

Elevation of cardiac enzyme levels is essential for the diagnosis, with increase in the cardiac isoenzyme of creatine kinase, total creatine kinase, and the less specific aspartate transaminase with one or other of the lactic dehydrogenase isoenzymes so that a profile of enzyme release, with early creatine kinase isoenzyme (24 h), intermediate aspartate transaminase (24 to 48 h) and late lactic dehydrogenase (48 to 72 h) peak can be built up. Recently, the measurement of the Troponin I or T isoenzymes has been successfully used to predict infarction or high cardiovascular risk *within* 6 h of acute chest pain. Infarct size can be approximated from total cardiac enzyme release.

Several drugs relieve myocardial ischaemia and improve survival of patients with myocardial infarction. When given intravenously, b-blockers reduce myocardial oxygen demand especially in patients with tachycardia and elevated blood pressure, and their role in improving prognosis has been established in several trials.

The greatest impact in terms of reducing mortality from infarction has followed the introduction of aspirin and thrombolytic drugs. About 80 per cent of patients with Q-wave infarcts studied by arteriography within 4 h show thrombotic occlusion of the infarct-related coronary artery. Large-scale trials have shown that intravenous thrombolytic drugs partially or completely reperfuse about 60 to 70 per cent of occluded arteries, with improved long-term survival. Combined with aspirin, 30-day mortality is reduced from 13 to 7 to 10 per cent. Benefit is closely related to the delay between symptom onset and the start of therapy (Fig. 8) but clearly extends to 12 h after infarction and up to 24 h in selected patients.

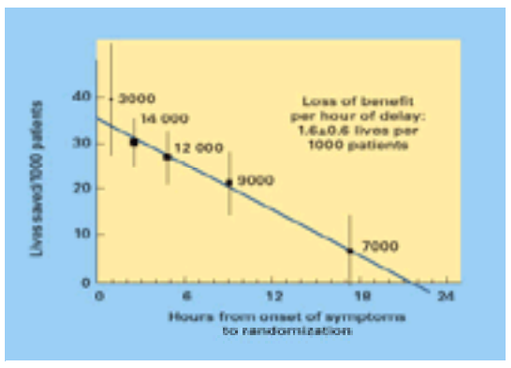


Fig. 8. Absolute mortality benefit with thrombolytic therapy for acute myocardial infarction from nine major randomized trials in 45 000 patients. Patient numbers are shown for each data point.

The two thrombolytic preparations used most commonly are streptokinase, and tissue plasminogen activator. Tissue plasminogen activator is more expensive and provides a more intense thrombus-targeted thrombolytic effect giving slightly greater patency of the infarct-related coronary artery at 90 min. It is, however, associated with a marginally greater risk of stroke from cerebral haemorrhage (3 per 1000 patients treated). Younger patients with extensive anterior wall infarction presenting early (who often have no collateral protection of the infarcting area) may benefit most from tissue plasminogen activator. Streptokinase has a longer lasting thrombolytic action with potentially useful antiplatelet effects from fibrin-degradation products. Hypotension can be seen with streptokinase but this may not confer extra risk. The thrombolytic effect of streptokinase may be inconsistent especially in those previously exposed to the drug where antibodies can limit the pharmacological effects for up to 1 year. Aspirin provides an additional 20 to 25 per cent reduction in 30-day mortality and is in routine use. This benefit emphasizes the importance of platelet aggregation in the evolution and propagation of thrombus in the infarct-related coronary artery. Future trials will combine thrombolytic agents with glycoprotein IIb-IIIa platelet receptor blockers

Heparin has been extensively studied in clinical trials of acute myocardial infarction but has not shown overwhelming advantage in routine clinical use. It may have a stronger case after tissue plasminogen activator as the thrombolytic agent and is useful in the prevention of secondary venous and arterial thromboembolism in selected high-risk patients such as those with heart failure, atrial fibrillation, and left-ventricular thrombus.

Thrombolytic drugs have been a major advance in acute management of myocardial infarction but difficulties remain. Failure of reperfusion as judged by the ECG is associated with high mortality but management is uncertain; options include repeat thrombolysis with the same or a different agent, emergency coronary angiography with or without intra-aortic balloon counterpulsation, and conservative strategies. Some patients have an absolute contraindication to thrombolytic agents, such as recent major surgery, previous cerebral bleeding, active peptic ulceration, or extensive cardiopulmonary resuscitation. Again, such patients have a high mortality. For these reasons, attention is turning increasingly to primary coronary angioplasty as an alternative. Abandoned for routine use *after* thrombolysis in the 1980s, this strategy is now used in up to 40 per cent of all acute myocardial infarcts in the United States of America but much less commonly in Europe, partly for economic and practical reasons; 80 to 90 per cent of acute infarction in the United Kingdom presents to a hospital without on-site angioplasty facilities. Primary coronary angioplasty for myocardial infarction is consistent with the open-artery hypothesis so clearly supported by the thrombolytic trials. Provided the patient can be transferred to an appropriate catheterization laboratory within 30 to 60 min, primary coronary angioplasty, with or without coronary stent implantation, offers high procedural success, greater utility, and improved epicardial flow in the infarct-related coronary artery compared with thrombolysis. Safe early patient discharge in low risk individuals may be possible. In experienced centres, 30-day mortality seems to be lower (2 to 6 per cent) but this crucial question remains controversial, especially in less experienced institutions. Clearly, primary coronary angioplasty for myocardial infarction is impractical for most district general hospitals in the United Kingdom. More data on this promising strategy is needed before it will become the acute treatment of choice for myocardial infarction.

Early management of myocardial infarction

Patients seen within 24 h of onset of possible myocardial infarction should be admitted to hospital without delay and assessed immediately. Opiate analgesia, aspirin, and oxygen should be given and in the absence of contraindications thrombolytic treatment started, provided the ECG shows 1 mm or more ST segment elevation in two sequential leads, or bundle branch block. If there is likely to be an unavoidable delay in getting the patient with a confident diagnosis to hospital then pre-hospital thrombolysis may be effective. The decision to initiate thrombolytic treatment more than 12 h after the onset of pain should be influenced by factors such as site of infarction (anterior wall infarcts are at higher risk than those in the inferior wall), persistent pain, ST segment elevation, and the presence of residual R-waves on the ECG. Streptokinase remains the most commonly prescribed thrombolytic. Primary coronary angioplasty should be considered as an alternative strategy to thrombolysis within 6 h of symptoms in large anterior infarcts especially in younger patients, with major haemodynamic deterioration and when thrombolytic therapy is contraindicated.

Angiotensin-converting enzyme inhibitors have an established role in most Q-wave myocardial infarcts. There is a modest acute mortality benefit (possibly via reduced cardiac rupture) that increases over the subsequent weeks reflecting the influence of angiotensin-converting enzyme inhibition on myocardial remodelling and hypertrophy after infarction. Reduction in the subsequent risk of heart failure appears to be the principal protective mechanism but a secondary anti-arrhythmic action may contribute. This protective action is likely to be a drug class effect but captopril, enalapril, lisinopril, and ramipril are of proven value in the post-thrombolytic era. b-Blockers should be routinely given after acute infarction, unless contraindicated. Nitrate therapy is useful in alleviating post-infarction angina but mortality benefit, and hence routine use remains unproven.

Further management of uncomplicated myocardial infarction

The patient can leave the coronary care unit after 24 to 36 h, be mobile at about 48 h, and be discharged at 6 to 7 days. Timing depends on the extent of myocardial damage. A pre-discharge exercise test should be performed in patients who have had thrombolytic treatment and those with a small enzyme rise or non-Q-wave infarct. In the former, there is the possibility of successful reperfusion and a severe residual coronary stenosis with an incomplete infarct. In both groups, multivessel coronary artery disease may exist. Patients with symptoms or ECG evidence of ischaemia or those with poor exercise capacity during a submaximal exercise test should undergo coronary arteriography.

For those with a satisfactory pre-discharge exercise test there is little to be gained from a further symptom limited test at 3 weeks, although the latter is an alternative means of assessment for uncomplicated infarcts. Those who complete stage 3 of the standard Bruce protocol without ischaemic changes have a predicted 1-year mortality of less than 2 per cent.

After acute myocardial infarction changes occur in the myocardium with early thinning of the infarcted area and over months dilatation of the left ventricle that may lead to clinical heart failure. Angiotensin-converting enzyme inhibitors reduce this risk. Aspirin should be continued long-term if tolerated. Secondary prevention is crucially important after myocardial infarction with cessation of smoking, reduction of cholesterol (most patients should receive ‘statin' therapy), blood pressure control, and lifestyle adjustment the major considerations.

Management of complications

Anti-arrhythmic treatment may be necessary early in the course of infarction. Ventricular fibrillation or fast ventricular tachycardia usually responds to 200 or 300 J defibrillation; if it occurs within 24 h of infarction prophylactic anti-arrhythmic drug treatment is usually not required. Intravenous lignocaine (lidocaine) is given for recurrent ventricular tachycardia, or for frequent symptomatic ventricular ectopy, especially after ventricular fibrillation or tachycardia. Intravenous magnesium infusion is safe and has a modest anti-arrhythmic effect in this context. Temporary transvenous cardiac pacing may be required for symptomatic heart block, asystole, and bradycardia not responding to atropine. If thrombolytic treatment has been given, venous access for pacing should be via a compressible site.

In the majority of patients with large antero-apical infarcts, echocardiography shows left-ventricular thrombus; this makes them liable to systemic embolism and stroke. Prophylactic anticoagulation reduces this risk: heparin should be administered initially and warfarin treatment continued for 6 months.

Doppler echocardiography is reliable for the diagnosis of ventricular septal rupture, rupture of the free wall of the left ventricle, and papillary muscle dysfunction or rupture leading to acute mitral regurgitation. All should be considered for early surgical repair.

Once in hospital, the most common cause of death is pump failure. Cardiogenic shock carries an extremely high mortality (40 to 90 per cent). The initial approach

involves optimizing left-ventricular filling pressures using direct measurement of pulmonary artery wedge pressure and administration of dopamine and other inotropic drugs such as dobutamine. Occasionally, underfilling of the left ventricle (from diuresis, fluid restriction, or right-ventricular infarction) mandates volume expansion to maintain cardiac output. Fluid replacement and inotropic and vasopressor drugs aim to maintain left-ventricular filling pressure at about 18 mmHg. However, in the absence of obvious hypovolaemia this approach should be subordinate to the urgent need to reperfuse ischaemic myocardium in the catheterization laboratory. Culprit lesion coronary occlusions can be identified and treated with angioplasty and/or intracoronary stenting. The survival rate of about 10 to 20 per cent for patients in cardiogenic shock may be improved up to threefold by emergency coronary angioplasty. Benefit is almost certainly related to the speed of re-establishment of satisfactory flow in the infarct-related coronary artery.

Further reading

American College of Cardiology/American Heart Association. Guidelines for the early management of patients with acute myocardial infarction. *Circulation* 1990; **82**: 664–707.

American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Subcommittee on Coronary Angiography). Guidelines for coronary angiography. *Journal of the American College of Cardiologists* 1987; **10**: 935–50.

American Heart Association and the National Heart, Lung, and Blood Institute. The cholesterol facts. A summary of the evidence relating dietary fats, serum cholesterol, and coronary heart disease. *Circulation* 1990; **81**: 1721–33.

Antiplatelet Trialists Collaboration; collaborative overview of randomised trials of antiplatelet therapy—I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *British Medical Journal* 1994; **308**: 81–106.

Bypass Angioplasty Revascularisation Investigation (BARI) Investigators. Comparison of coronary surgery with angioplasty in patients with multivessel disease. *New England Journal of Medicine* 1996; **335**: 217–25.

Cholesterol and Recurrent Events Trial Investigators. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *New England Journal of Medicine* 1996; **335**: 1001–9.

Chaitman BR. The changing role of the exercise electrocardiogram as a iagnostic and prognostic test for chronic ischaemic heart disease. *Journal of the American College of Cardiologists* 1986; **8**: 1195–210.

Cohen M *et al.* A comparison of low molecular-weight heparin with unfractionated heparin for unstable coronary artery disease. *New England Journal of Medicine* 1997; **337**: 447–52.

Collins R, Peto R, Baigent C, Sleight P. Aspirin, heparin and fibrinolytic therapy in suspected acute myocardial infarction. *New England Journal of Medicine* 1997; **336**: 847–60.

Davies MJ, Thomas AC. Plaque fissuring—the cause of acute myocardial infarction, sudden ischaemic death, and crescendo angina. *British Heart Journal* 1985; **53**: 363–73.

DeWood MA *et al.* Prevalence of total coronary occlusion during the early hours of transmural myocardial infarction. *New England Journal of Medicine* 1980; **303**: 897–902.

Falk E. Unstable angina with fatal outcome: Dynamic coronary thrombosis leading to infarction and/or sudden death. Autopsy evidence of recurrent mural thrombosis with peripheral embolization culminating in total vascular occlusion. *Circulation* 1985; **71**: 699–708.

Fishman DL *et al.* A randomised comparison of coronary stent placement and balloon angioplasty in the treatment of coronary artery disease. *New England Journal of Medicine* 1994; **331**: 496–501.

Gruppo Italiano per lo Studio della Streptochinasi nell' Infarto Miocardico (GISSI). Effectiveness of thrombolytic treatment in acute myocardial infarction. *Lancet* 1986; **i**: 397–402.

GUSTO Investigators. An International randomised trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 1993; **329**: 673–82.

ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction. *Lancet* 1988; **ii**: 349–60.

ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. A randomized comparison of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41 299 cases of suspected myocardial infarction. *Lancet* 1992; **i**: 735–70.

ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. A randomised factorial trial assessing oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995; **345**: 669–85.

Klein W *et al.* Comparison of low-molecular-weight heparin with unfractionated heparin acutely and with placebo for 6 weeks in the management of unstable coronary artery disease; Fragmin in unstable coronary artery disease study (FRIC). *Circulation* 1997; **96**: 61–8.

Lieu TA, Gurley RJ, Lundstrom RJ, Parmley WW. Primary angioplasty and thrombolysis for acute myocardial infarction; an evidence summary. *Journal of the American College of Cardiologists* 1996; **27**: 737–50.

Oler A, Whooley MA, Oler J, Grady D. Adding heparin to aspirin reduces the incidence of myocardial infarction in patients with unstable angina; a meta-analysis. *Journal of the American Medical Association* 1996; **276**: 811–15.

Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial: Risk factor changes and mortality results. *Journal of the American Medical Association* 1982; **248**: 1465–77.

Ross R. The pathogenesis of atherosclerosis—An update. *New England Journal of Medicine* 1986; **314**: 488–500.

Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease; the Scandinavian Simvastatin Survival Study (4S). *Lancet* 1994; **344**: 1383–9.

Shepherd J *et al.* Prevention of coronary heart disease with pravastatin in men with hypercholesterolaemia. *New England Journal of Medicine* 1995; **333**: 1301–7.

SOLVD Investigators. Effects of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left-ventricular ejection fractions. *New England Journal of Medicine* 1992; **327**: 685–91.

TIMI Study Group. Comparison of invasive and conservative strategies after treatment with intravenous tissue plasminogen activator in acute myocardial infarction (TIMI) phase II trial. *New England Journal of Medicine* 1989; **320**: 618–27.

Weiner DA *et al.* Prognostic importance of a clinical profile and exercise test in medically treated patients with coronary disease. *Journal of the American College of Cardiologists* 1984; **3**: 772–9.

40.8.2 Interventional cardiology

Douglas C. Morris

Percutaneous transluminal coronary angioplasty
Introduction
Angioplasty for chronic stable angina
Angioplasty for unstable angina
Angioplasty in acute myocardial infarction
Impact of lesion morphology on outcome
Angioplasty in specific subsets of the population
Alternative or complementary strategies to balloon angioplasty
Percutaneous balloon valvuloplasty
Percutaneous mitral balloon valvuloplasty
Percutaneous balloon aortic valvuloplasty
Percutaneous pulmonary valvuloplasty
Further reading

Percutaneous transluminal coronary angioplasty

Introduction

Over the past 20 or so years, coronary intervention has evolved from an investigational technique using a dilatation catheter composed of a minimally compliant cylindrical balloon to the widespread use of balloons of varying compliance and length, and of non-balloon debulking devices and stents (debulking refers to decreasing plaque mass). The role of intervention in coronary arterial disease has advanced from a therapy applicable only to discrete, proximal, non-calcified stenoses to a refined technique suitable for distal stenoses, geometrically complex stenoses, calcified stenoses, total occlusions, and saphenous vein grafts. Physician experience and expertise with the procedure has grown to the extent that angioplasty is appropriately utilized in challenging clinical situations such as unstable angina, acute myocardial infarction, multivessel disease, and in patients with poor left ventricular function. The exponential growth of coronary angioplasty since its initial clinical application in September 1977 by Andreas Gruentzig is reflected in Fig. 1.

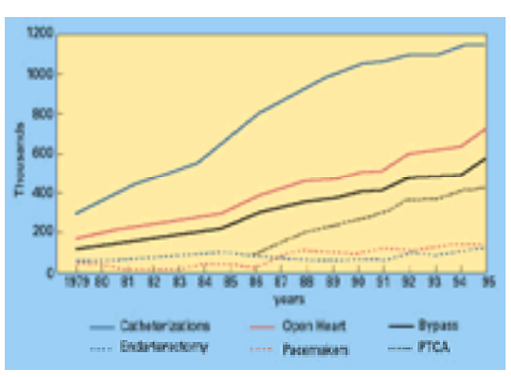


Fig. 1. Trends in cardiovascular operations and procedures (Source: National Center for Health Statistics and the American Heart Association).

Indications for coronary intervention

In selecting percutaneous coronary intervention as the most appropriate therapeutic approach in a given patient, the interventionist must balance the risk and consequences of failure with the likelihood and duration of success. The risks, durability, and completeness of revascularization should be acceptable in comparison with coronary bypass surgery. The potential for more complete relief of symptoms should outweigh the cost and short-term risk advantages of medical therapy.

Angioplasty for chronic stable angina

Single-vessel disease

Coronary angioplasty is generally considered the treatment of choice for patients with moderate or severe myocardial ischemia and anatomically suitable, single-vessel disease. Acceptance of this role for angioplasty has been primarily based upon the experience at multiple, high-volume, angioplasty centers and on observational studies. A large, prospective, observational study of treatment strategies in single-vessel disease found superior survival over the initial 5-year follow-up for angioplasty as compared to coronary bypass surgery for all sites except proximal, severe stenosis in the left anterior descending artery; the two therapies had an equivalent effect on survival in disease of the proximal part of that artery. The survival trend for all single-vessel disease favored angioplasty over medical therapy but not convincingly. A randomized trial of angioplasty and grafting of the internal mammary artery in 134 patients with isolated disease of the left anterior descending coronary artery showed no difference in mortality or myocardial infarction over a 2-year follow-up, but angioplasty led to more repeat revascularizations (25 per cent as compared to 3 per cent, $p < 0.01$). In comparison to medical therapy, angioplasty in single-vessel disease reduces the frequency of angina and improves exercise performance for at least 3 years according to data from a Veterans Affairs study.

Multivessel disease

The majority of patients presenting with symptomatic coronary arterial disease have multivessel involvement. Once angioplasty was established as a safe, effective therapeutic approach to single-vessel disease, the selection criteria were liberalized to include this larger cohort of patients. Its use in patients with multivessel disease has resulted in a high initial success rates and a low incidence of complications.

Five major trials have compared outcomes in patients with multivessel coronary disease randomized to angioplasty or surgery. The major difference between the outcome of the two therapies in all these trials is in the need for repeat revascularization (3 to 1 per cent for coronary arterial bypass grafting against 33 to 49 per cent for percutaneous transluminal coronary angioplasty). There is no difference in overall mortality between the two groups, and only in the German Angioplasty Bypass Investigation (GABI) trial was there a difference in the frequency of acute myocardial infarction (higher in the surgical group) (Table 1).

	ERACI	GABI	CABRI	EAST	BARI
No. of patients	127	359	1054	392	1829
Follow-up (months)	36	12	12	36	
Patients with three-vessel disease (%)	45	18	40	48	43
Patients receiving internal mammary artery (%)	77	37	81	98	82
Chronic occlusions treated with angioplasty*	Yes	No	Yes	Yes	Yes
Patients requiring reintervention (%)					
PTCA	37	44	16	49	
CaBG	3	4	3	11	
Mortality (PTCA vs CABG)	ND	ND	ND	ND	
Infarction (PTCA vs CABG)	ND	CABG+	ND	ND	
Angina (PTCA vs CABG)	ND	ND	PTCA+	PTCA+	

BARI, Bypass Angioplasty Revascularization Investigation; CABRI, Coronary Artery Bypass Revascularization Investigation; EAST, Emory Angioplasty Versus Surgery Trial; ERACI, Argentine Randomized Trial of Angioplasty versus Surgery; GABI, German Angioplasty Bypass Investigation; ND, no difference

Table 1 Randomized trials comparing percutaneous transluminal coronary angioplasty (PTCA) with coronary artery bypass graft (CABG) in patients with multivessel

disease

The differences in the rates of reintervention between angioplasty and bypass surgery are due to the significant incidence of restenosis in the early months after angioplasty and the higher frequency of incomplete revascularization with angioplasty. While the incidence of restenosis has been very favorably affected by the use of coronary stents, the threat of restenosis remains. Complete revascularization, which has been shown in the surgical experience to produce superior long-term results, is less frequently achieved with angioplasty than with surgery. The reasons include the presence of chronic total occlusions or lesions either unsuitable or technically difficult to dilate. In the Emory Angioplasty Versus Surgery Trial (EAST) only 71 per cent of the index segments were successfully dilated with angioplasty while in the Coronary Artery Bypass Revascularization Investigation (CABRI) only 54 per cent of the diseased vessels were successfully dilated. Most often, however, incomplete revascularization is a predetermined strategy to focus upon those stenoses likely causing the symptoms (culprit lesion) and to avoid complex lesions of little clinical significance in order to shorten the procedure and enhance safety. While ‘culprit lesion' angioplasty seems an appropriate therapeutic approach to relieve symptoms in those patients who are suboptimal candidates for bypass surgery, the ‘culprit lesion' may not be always clearly identifiable.

Angioplasty for unstable angina

Patients with unstable angina account for a majority of coronary angioplasties and are more likely than those with stable angina to experience major complications including death, myocardial infarction, and emergency coronary bypass. Initially, the favored approach in the patient with unstable angina was to defer angioplasty for a few days while stabilizing them on aggressive antianginal therapy, including aspirin and heparin. With the clinical availability and proven efficacy in this subset of patients of the platelet glycoprotein IIb/IIIa inhibitor, abciximab, the preferred approach is now early angioplasty combined with an intravenous bolus followed by a 12-h infusion of this monoclonal antibody. Heparin is administered in combination with this agent only during the procedure and in a reduced dose in order to reduce the risk of bleeding seen with the combination of 7E3 (abciximab) and full-dose heparin.

Angioplasty in acute myocardial infarction

In the treatment of acute myocardial infarction, coronary angioplasty was initially reserved primarily for those patients who were considered ineligible for thrombolysis (Table 2). The role of primary angioplasty, however, continue to expand in those patients who meet eligibility criteria for thrombolysis. Primary angioplasty offers the advantages of immediate confirmation of the diagnosis, thereby eliminating conditions masquerading as an infarction, a wider applicability than thrombolytics due to fewer contraindications, and freedom from the threat of intracranial hemorrhage. Most importantly, the initial studies on direct angioplasty of acute myocardial infarction established a higher patency rate (greater than 90 per cent) than with thrombolytics. This initial experience with primary angioplasty was followed by several randomized trials that confirmed the higher patency rate and demonstrated an improved clinical outcome with angioplasty compared to thrombolytics (Table 3). The restoration of grade 3 flow on the **TIMI** (thrombolysis of myocardial infarction) scale, which is significantly more likely with angioplasty, is the strongest determinant of short-term survival with any method of reperfusion. The excellent results reported with primary angioplasty are dependent on a short door-to-perfusion time (optimally less than 60 min) that is not achievable in all hospitals, and subject to a ‘learning curve' even for experienced operators. The data argue strongly in favor of primary angioplasty in the following circumstances.

Active or recent (4 weeks) internal bleeding
Intracranial neoplasm or recent head trauma
Prolonged, traumatic cardiopulmonary resuscitation
Suspected aortic dissection
History of hemorrhagic cerebrovascular accident
Recent (3 months) nonhemorrhagic cerebrovascular accident
Sustained systemic blood pressure over 200/120 mmHg in spite of medication
Trauma or surgery that is a potential bleeding source in the past 4 weeks

Table 2 Absolute contraindications to thrombolysis

Study	Approach	No. of patients	Success rate (%)	Reocclusion (%)	ischemia	Death (%)		
Grines, et al.	PTCA	195	97	10.3			2.6	
	TPA	210		30.0		$p < 0.001$	6.3	$p = 0.06$
Gibbons et al	PTCA	47	93	15.0			2.1	
	TPA	56		36.2			0.0	
Zijlstra, et al.	PTCA	70	98	9.0			0.0	
	SK	72		38.0		$p < 0.001$	4.0	$p = 0.13$
Ribeiro et al.	PTCA	50	80	8.0			4.0	
	SK	50		10.0		$p = NS$	2.0	$p = NS$
Grines GEB	PTCA	503	73/88*	9.9			5.7	
	TPA	573		15.5**		$p = NS$	7.0	$p = 0.37$

NS, not significant; PTCA, percutaneous transluminal coronary angioplasty; SK, streptokinase; TPA, tissue plasminogen activator.
*Lower per cent is core lab results.
**Combined for reinfarction and ischemia.

Table 3 Randomized studies of primary angioplasty

1. A patient who is thrombolytic-ineligible:
   a. nondiagnostic electrocardiogram;
   b. absolute contraindication for thrombolytics.
2. A patient within 12 h of the onset of an anterior myocardial infarction in whom clinical, electrocardiographic, and biochemical measures indicate a failure of thrombolytics to reperfuse.
3. A patient presenting within 12 h of an acute anterior myocardial infarction to a hospital with the facilities and an experienced team to deliver primary angioplasty.
4. A patient with an acute myocardial infarction accompanied by cardiogenic shock. (congestive heart failure, systemic hypotension, and tachycardia)
5. A patient over 65 years of age.

An unresolved issue about angioplasty in acute myocardial infarction is whether there is a benefit in administering either a bolus of a thrombolytic agent or a glycoprotein IIb/IIIa platelet inhibitor before angioplasty, particularly if there is to be a short delay before intervention.

Impact of lesion morphology on outcome

Recognition that lesion morphology played an important part in the success of angioplasty led to the American College of Cardiology/American Heart Association Percutaneous Transluminal Coronary Angioplasty Guidelines. In these guidelines, lesions were divided into type A (expected success greater than 85 per cent), type B (expected success 60 to 85 per cent), and type C (less than 60 per cent success) (Table 4). The success rates ascribed to each lesion type are not typical of results that are currently achievable with contemporary balloon technology. Furthermore, multiple type B characteristics will reduce the success rate and certain type B characteristics have a higher predictive value in determining success than others. In spite of an increase in the complexity of lesions being attempted, the advent of new devices, particularly stents, and antithrombotic strategies have weakened the prognostic value of the A, B, C lesion-grading system. The procedural success was 100 per cent for type A, 97.3 per cent for type B1, 97.1 per cent for type B2, and 87.4 per cent for type C lesions undergoing angioplasty in 1085 patients in the era of the new devices. Lesion length of greater than 20 mm, TIMI-I flow, calcification, angle greater than 90 per cent, lesions in old vein grafts, and chronic total occlusion were the predictors of procedural failure.

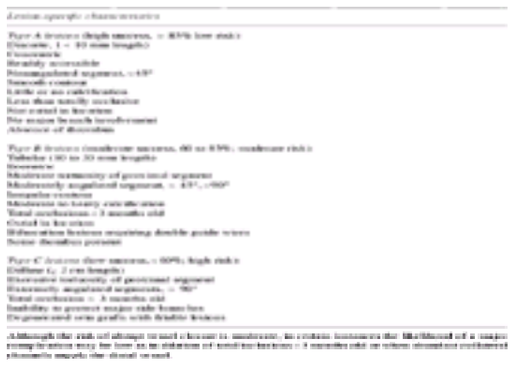


Table 4 Characteristics of type A, B and C lesions

Lesion characteristics that are associated with increased restenosis include proximal location in the left anterior descending artery, longer lesion length, chronic (over 3 months) total occlusions, angiographically visible thrombus, ostial location, and vein graft-body lesions.

Angioplasty in specific subsets of the population

Women

The risk of coronary angioplasty is higher in women than in men. The reasons for this higher risk in women is multifactorial. First, women generally have smaller coronary arteries, which can result in oversizing of balloons and stents. Smaller coronary arteries restrict the effective use of stenting as a bail-out technique and likely leads to an underestimation of the severity of the atherosclerosis. Secondly, women with coronary disease are typically older and have more comorbid factors than men with coronary disease. Finally, the Emory data suggest that volume depletion is more likely to occur in women, based upon their more frequent hypotensive response and an increased risk of death associated with a smaller body surface area.

Women, as compared to men, are also at increased risk for early and late morbidity and mortality following an acute myocardial infarction. Women presenting with myocardial infarctions, however, are older than their male counterparts, have a higher baseline prevalence of diabetes, hypertension and congestive heart failure, and present later in the course of their infarction. While thrombolytic therapy has demonstrated a 25 to 30 per cent reduction in early mortality in women and equivalent rates of recanalization of the infarct artery to those in men, in-hospital survival has remained lower for women than men with this therapy. Primary angioplasty when compared to thrombolytic therapy in acute myocardial infarction may improve survival in women and negate the sex-associated survival difference.

Advanced age

A steadily increasing number of invasive cardiovascular procedures are performed on the elderly patient. At the Emory Heart Center, the proportion of patients over 75 years of age has increased from 6 per cent in 1985–87 to 14 per cent in 1994–96. The explanations for this rapid emergence of coronary angioplasty as a therapeutic option in this segment of our population is multifactorial. First, there is the disproportionate growth in the elderly population. Secondly, this population has the highest prevalence of coronary atherosclerotic heart disease. Thirdly, coupled with this high prevalence of disease is a steadily increasing life expectancy. Finally, and most importantly, has been the reduction in morbidity and mortality with angioplasty. The most recent studies on angioplasty in elderly people have demonstrated initial success rates similar to those reported for younger age groups. The Emory data confirm this equivalency in angiographic and clinical success rates until age 80 years, when the success rates trend downward. In the Emory data, the major associated risks of cardiac death and Q-wave myocardial infarction did not increase significantly until after the age of 80 years. While most studies also noted this significant rise in complications occurring after the age of 75 or 80 years, the Washington Hospital Center results noted that the increase began after the age of 65. Their results also noted an increase in complications of vascular access beginning at the same age.

Since the reclosure rate after angioplasty in elderly people is similar to that in younger age groups, the older patient must not tolerate acute occlusion after angioplasty as well as his or her younger cohorts. Analysis of the Emory data suggests a confounding factor, namely, the elderly patient is more likely to have his of her reclosure treated by medical means, perhaps accounting for the increase in the postangioplasty myocardial infarction rate and possibly even the death rate. In other words, complications are more likely to be accepted in the very elderly patient than to proceed to surgery and its inherent risks.

Long-term results with coronary angioplasty show a higher cardiac death rate and more frequent recurrence of angina in the patient over the age of 75. As would be expected, the overall death rate is also higher. If complete revascularization is achieved in the elderly patient, then the differences in the cardiac death rate and recurrence of angina between the elderly and younger age groups disappear. Analysis of the Emory data reveals that while coronary disease is more extensive in the elderly patients the frequency of multisite angioplasty is the same as in younger patients, implying that myocardial revascularization is more often incomplete in the elderly patient.

Diabetes mellitus

The extent to which diabetes negatively affects the short-term outcome of coronary angioplasty remains controversial. Diabetic patients in the 1985–86 Percutaneous Transluminal Angioplasty Registry experienced more in-hospital death or myocardial infarction (10 compared to 4.5 per cent) even after multivariate adjustment for known predictors of adverse outcome. Conversely, analysis of the Emory data revealed no increased risk of death, myocardial infarction, or in-hospital coronary surgery among diabetic patients.

Long-term outcome after coronary angioplasty, however, is consistently adversely affected by diabetes mellitus, with an increased mortality and greater likelihood of myocardial infarction and repeat revascularization. Analysis of the Bypass Angioplasty Revascularization Investigation (**BARI**) data raised concern that survival among diabetic patients undergoing angioplasty for multivessel disease was lower than in those surgically revascularized with an internal mammary graft as one of the conduits. No other randomized trial or registry to date has confirmed the BARI findings and the present data do not support the conclusion that diabetic status alone should decide the revascularization strategy.

Alternative or complementary strategies to balloon angioplasty

Over the past two decades, other approaches to percutaneous revascularization of the coronary arteries have appeared. With the exception of stenting, these other devices have fallen from the scene or have slipped into the role of debulking devices to be used in concert with balloon angioplasty. At Emory University Hospital in 1997, balloon angioplasty was the sole technique used in 41.9 per cent of 1323 patients, 0.7 per cent were treated with directional atherectomy, 6.9 per cent were treated with rotational atherectomy, and none was treated with excimer laser; 51 per cent of the patients had stents implanted, which was a decided increase since 1995 when only 12.2 per cent received stents.

Rotational atherectomy is an appropriate interventional approach in the presence of calcium, in aorto-ostial and branch ostial lesions, in nondilatable lesions, and perhaps in long, ulcerated, and complex lesions. Highly angulated or thrombotic lesions are not appropriately treated with rotational atherectomy. Most interventionists use rotational atherectomy as an adjunct to balloon angioplasty but others do use the technique as a ‘stand-alone' approach. Neither approach has yet demonstrated a reduction in restenosis as compared to balloon angioplasty.

Two complications do seem to occur more frequently with rotational atherectomy than with balloon angioplasty, namely, the no-flow phenomenon and non-Q-wave infarction. The mechanism of the no-reflow phenomenon has not been established but is unlikely due entirely to distal obstruction by particulate matter generated by the device. Other possible explanations include platelet aggregation, endothelial dysfunction, and vasospasm. The prevalence of this occurrence is in the range of 7 to 20 per cent of patients undergoing rotational atherectomy. Non-Q-wave infarctions have been reported in between 0 and 10 per cent, with 4.9 per cent noted in the New Approaches to Coronary Intervention Registry.

The use of directional coronary atherectomy has decreased over the last decade. Many of the lesions initially thought to be most suitable for directional coronary atherectomy, such as aorto-ostial stenoses, eccentric stenoses, and bifurcation stenoses, are now most often treated with a combination of balloon angioplasty and stenting. Two large, multicenter trials comparing balloon angioplasty with directional coronary atherectomy, the Coronary Angioplasty versus Atherectomy Trial (CAVEAT) and the Canadian Coronary Atherectomy Trial (CCAT), both showed that directional coronary atherectomy was associated with a higher degree of angiographic success but in the CAVEAT trial this success was achieved with a higher periprocedural complication rate. Furthermore, there was no significant

difference in the restenosis rates between balloon angioplasty and directional coronary atherectomy in the two trials. Aggressive atherectomy techniques and adjunctive balloon angioplasty were found in the Balloon versus Optimal Atherectomy Trial (BOAT) to affect favorably the rate of restenosis but again at the cost of a higher frequency of myocardial infarctions.

The transluminal extraction catheter is used primarily in saphenous vein grafts containing thrombus or in large native vessels containing a large amount of angiographically visible thrombus. Its limited utility and, as a consequence, the lack of operator experience make it an impractical technique in many laboratories.

The excimer laser is effective in the treatment of aorto-ostial stenoses, total occlusions, and long lesions, but its superiority to simpler and less costly strategies has not been established.

Coronary stents were developed to address two specific complications of coronary angioplasty, namely, abrupt closure of the dilated artery and restenosis. The favorable impact of stenting on restenosis was established by the Benestent study and the Stent Restenosis study. The initial problems with coronary stents were related to subacute thrombosis and the vascular complications associated with the intense anticoagulation used to combat it. Recognition that incomplete attachment of the stent to the vessel wall and not just the thrombogenic material itself accounted for much of the subacute thrombosis, and the excellent results of the Benestent II pilot study (clinical success 99 per cent, subacute thrombosis 0, restenosis 13 per cent) without the use of anticoagulation, led to the discontinuation of anticoagulation after stenting. Presently, stents are frequently selected for the primary treatment of aorto-ostial lesions, proximal stenoses of the left anterior descending artery, saphenous vein stenoses, total occlusions, shelf-like lesions, and early recurrences. At other sites, stents are used for ‘bail-out' situations and for suboptimal results (greater than 20 per cent residual) in vessels of 3 mm or more.

Percutaneous balloon valvuloplasty

Percutaneous mitral balloon valvuloplasty

Any patient with symptomatic mitral stenosis should be considered for percutaneous mitral commissurotomy. Patient selection is based upon transthoracic, two-dimensional echocardiography. Valve rigidity, thickening, calcification, and subvalvular thickening are graded 0 to 4, with the sum of the gradings providing a score. Patients with a score of 8 or less are ideal candidates and those with a score over 12 should only be considered if surgery is not a viable option.

There are two transvenous techniques for performing percutaneous balloon mitral valvotomy; the wire-guided technique using either a single balloon or two balloons and the flow-directed method utilizing the Inoue balloon. The double-balloon method provides better hemodynamic results than the single-balloon. The flow-directed method with the Inoue balloon has become the most popular, as it gives hemodynamics similar to those of the double-balloon technique but is much simpler and quicker to perform.

Analysis of registry data on the Inoue balloon revealed that mitral valve replacement was necessary in 1 per cent of the patients, with a hospital mortality of 1.4 per cent. The mitral valve could not be crossed in 1.7 per cent and in 3.1 per cent an atrial septal defect greater than 1.5:1 was noted. Most of these shunts closed spontaneously within 3 months.

While numerous studies have shown excellent hemodynamic results of percutaneous mitral commissurotomy, data on the long-term results are somewhat scanty. The United States multicenter experience with the Inoue balloon showed that greater than 80 per cent of patients remain symptomatically improved during a 2-year follow-up (Fig. 2). Series from outside the United States of America based on a younger population of patients have shown even better long-term results.

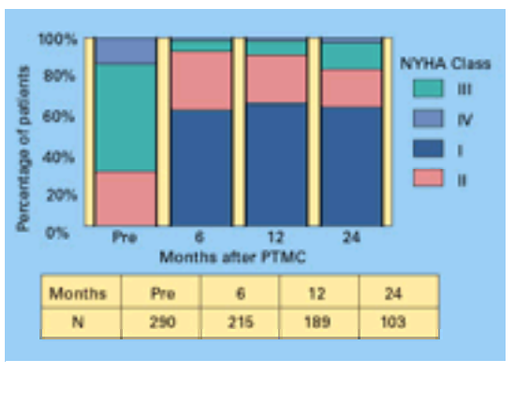


Fig. 2. The United States multicenter experience with the Inoue balloon. Almost all patients were significantly symptomatic prior to balloon dilatation. During the 2-year follow-up period, more than 80 per cent remained in New York Heart Association Class I or II.

Percutaneous balloon aortic valvuloplasty

While percutaneous balloon aortic valvuloplasty is an effective treatment option for children, it is only a palliative procedure in adults. Its use should be restricted to those patients who are not candidates for surgery or in those with poor left-ventricular function and a relatively low gradient across the aortic valve. In this latter group, valvuloplasty serve as a therapeutic trial to determine if their symptoms improve with intervention. Aortic valvuloplasty is safe in experienced hands.

Percutaneous pulmonary valvuloplasty

Balloon valvotomy has replaced surgery as the treatment of choice for pulmonic stenosis. The procedure can be performed safely and is much less invasive than a surgical approach. The procedure is indicated in any patient with isolated pulmonic stenosis and a resting gradient across the pulmonic valve of 50 mmHg or more. In the majority of patients a single balloon is used but double balloons should be used if the pulmonic annulus exceeds 19 mm. The balloon diameter should be 25 to 40 per cent larger than the pulmonary valve annulus to maximize the opening.

Further reading

American Heart Association. *Heart: stroke statistical update, 1997*, p. 27. AHA, Dallas, Texas, 1998.

Baim DS *et al.* Evaluating new devices. Acute (in-hospital) results from the new approaches to coronary intervention registry. *Circulation* 1994; **89**: 471–81.

Cowley MJ *et al.* Is traditionally defined complete revascularization needed for patients with multivessel disease treated by elective coronary angioplasty? *Journal of the American College of Cardiology* 1993; **22**: 1289–97. [Reviews question of incomplete revascularization during angioplasty.]

Feldman T, Hermann HC, Inoue K. Technique of percutaneous transvenous metral commissurotomy using the Inoue balloon. *Catheterization and Cardiovascular Diagnosis* 1994; Suppl 2: 26–34 1994. [Excellent supplemental issue on mitral valvuloplasty.]

Gibbons RJ *et al.* Immediate angioplasty compared with the administration of a thrombolytic agent followed by conservative treatment for myocardial infarction. *New England Journal of Medicine* 1993; **328**: 685–91.

Goy JJ *et al.* Coronary angioplasty versus left internal mammary artery grafting for isolated proximal left anterior descending artery stenosis. *Lancet* 1994; **343**: 1449–53. [Initial randomized study of angioplasty vs surgery in single vessel disease.]

Grines CL *et al.* A comparison of immediate angioplasty with thrombolytic therapy for acute myocardial infarction. *New England Journal of Medicine* 1993; **328**: 673–9. [Early randomized study of angioplasty vs thrombolytics in acute infarction.]

Holmes DR, Vlietstra RE. Balloon angioplasty in acute and chronic coronary artery disease. *Journal of the American Medical Association* 1989; **261**: 2109–15. [Early report of large angioplasty experience.]

Holmes DR *et al.* Percutaneous transluminal coronary angioplasty, alone or in combination with streptokinase therapy, during acute myocardial infarction. *Mayo Clinic Proceedings* 1985; **60**: 449–56. [Early experience with angioplasty in acute infarction.]

Lindsay J, Reddy VM, Pinnow EE, Little T, Picard AD. Morbidity and mortality rates in elderly patients undergoing percutaneous coronary transluminal angioplasty. *American Heart Journal* 1994; **128**: 697–702.

Mark DB *et al.* Continuing evolution of therapy for coronary artery disease: initial results from the era of coronary angioplasty. *Circulation* 1994; **89**: 2015–25. [Excellent observational study of treatment strategies in coronary disease.]

O'Keefe *et al.* Multivessel coronary angioplasty from 1980 to 1989: procedural results and long-term outcome. *Journal of the American College of Cardiology* 1990; **16**: 1097–102. [Initial large review of multivessel angioplasty.]

Parisi AF, Folland ED, Hartigan P. A comparison of angioplasty with medical therapy in the treatment of single-vessel coronary artery disease. *New England Journal of Medicine* 1992; **326**: 10–16. [Initial study randomizing patients to angioplasty vs medical therapy.]

Ribbeiro EE *et al.* A randomized trial of direct PTC versus intravenous streptolinase in acute myocardial infarction (abstr.). *Journal of the American College of Cardiology* 1991; **17**: 152A.

Rickards AF, Davies SW. Coronary angioplasty versus coronary surgery in the management of angina. *Current Opinion in Cardiology* 1995; **10**: 399–403. [Comparative review of two treatment strategies for angina.]

Ryan TJ *et al.* Guidelines for percutaneous transluminal coronary angioplasty: a report of the American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (subcommittee of Percutaneous Transluminal Coronary Angioplasty). *Journal of the American College of Cardiology* 1988; **12**: 519–40.

Stein B *et al.* Influence of diabetes mellitus on early and late outcome after percutaneous transluminal coronary angioplasty. *Circulation* 1995; **91**: 979–89.

Stone GW *et al.* Comparison of in-hospital outcome in men versus women treated by either thrombolytic therapy or primary coronary angioplasty for acute myocardial infarction. *American Journal of Cardiology* 1995; **75**: 987–92.

Tan K, Sulke N, Taub N, Sowton E. Clinical and lesion morphologic determinants of coronary angioplasty success and complications: current experience. *Journal of the American College of Cardiology* 1995; **25**: 855–65.

ten Berg JM *et al.* Long-term results after successful percutaneous transluminal coronary angioplasty in patients over 75 years of age. *American Heart Journal* 1996; **77**: 690–5.

The Epic Investigators. Use of monoclonal antibody directed against the platelet glycoprotein IIb/IIIa receptor in high-risk coronary angioplasty. *New England Journal of Medicine* 1994; **330**: 956–61. [Large multicenter trial establishing efficacy of glycoprotein IIb/IIIa..]

The Global Use of Strategies to Open Occluded Coronary Arteries in Acute Coronary Syndromes (GUSTO IIb) Angioplasty Substudy Investigators. A clinical trial comparing primary coronary angioplasty with tissue plasminogen activator for acute myocardial infarction. *New England Journal of Medicine* 1997; **336**: 1621–8. [Multicenter, randomized study of angioplasty vs thrombolytics in acute infarction.]

Weintraub WS, King SB, Douglas JS, Kosinski AS. Percutaneous transluminal coronary angioplasty as a first revascularization procedure in single-, double- and triple-vessel coronary artery disease. *Journal of the American College of Cardiology* 1995; **26**: 142. [Report of outcomes at a center with vast experience in angioplasty.]

Zijlstra F *et al.* A comparison of immediate coronary angioplasty with intravenous streptokinase in acute myocardial infarction. *New England Journal of Medicine* 1993; **328**: 680–4.

40.8.3 Coronary artery bypass

Robert A. Guyton

The origins of coronary artery bypass
Indications for coronary artery bypass
Comparison with medical therapy
Comparison of coronary artery bypass with percutaneous techniques
Summary of indications for coronary artery bypass
Technical considerations
Myocardial protection
Increased use of arterial conduits
Reoperations
Perioperative management
Modification of preoperative risk factors
Intraoperative management
Postoperative management
Results of coronary artery bypass grafting
Early outcomes
Late results of coronary artery bypass
Cost-effectiveness of coronary artery bypass
Evolving technology
Prevention of vein graft atherosclerosis
Transmyocardial revascularization
Less invasive coronary artery bypass
Further reading

The origins of coronary artery bypass

In early 1900s the apparent causal relationship between inadequate myocardial blood flow and angina pectoris was well appreciated. Alexis Carrel proposed a model of aortocoronary bypass and attempted such a bypass in a dog using the carotid artery as a conduit. Indirect methods of cardiac revascularization were tried in the 1930s, attempting to attach the pectoral muscle, the lung, or the omentum to the heart, allowing the development of collaterals. Claude Beck developed a procedure which encouraged the revascularization of the heart from the pericardial blood supply, the Beck I procedure. The epicardium and the inner surface of the pericardium were abraided, dilute acid and powdered asbestos were placed on the heart, and the coronary sinus was narrowed to a diameter of 3 mm. Vineberg, in Montreal, attempted to use the internal mammary artery as a source for collateral revascularization of the heart, pulling the internal mammary artery through a bluntly created tunnel in the myocardium. The patency of the Vineberg implant was high (80 per cent) but flow into the myocardium was limited.

Direct arterial revascularization was attempted in the 1950s and early 1960s. In 1953 Mustard in Toronto unsuccessfully attempted a carotid to coronary bypass. Bailey, Longmire, and Effler utilized coronary endarterectomy in the mid-1950s. In one of these operations, an internal mammary artery was used to repair a damaged right coronary artery in 1958 by William Longmire. In 1964 a similar situation was salvaged by an aortocoronary saphenous vein graft by Garrett and DeBakey. The first planned aortocoronary saphenous vein graft operation was probably conducted by Sabiston in 1962. Unfortunately this procedure was complicated by a fatal postoperative stroke. Selective coronary angiography was developed at the Cleveland Clinic in the 1960s by Sones. This allowed the development of direct aortocoronary revascularization by Rene Favaloro beginning with vein patch angioplasty after endarterectomy and followed by direct aortocoronary saphenous vein grafting (Fig. 1 and Fig. 2).

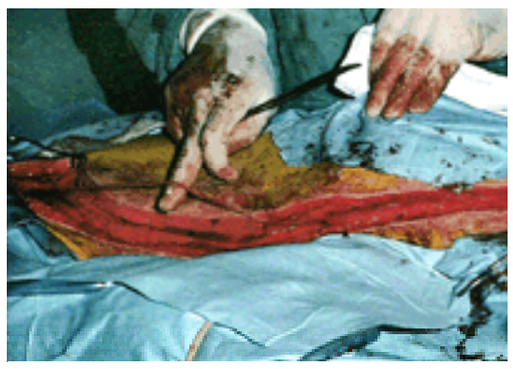


Fig. 1. Harvest of the long saphenous vein from groin to mid-calf—sufficient for four single grafts.

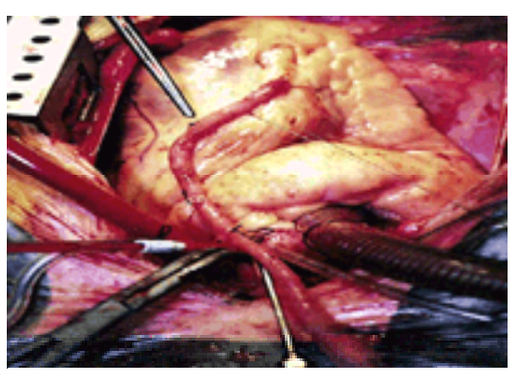


Fig. 2. Sequential long saphenous vein graft from left anterior descending coronary to its principal diagonal branch.

The internal mammary artery was utilized as a bypass conduit in 1954 by Murray in experimental studies. It was intentionally utilized in one patient by Goetz in 1961. A large clinical series was begun by Kolessov in Leningrad beginning in 1964. Utilizing an operating microscope George Green brought this technique to successful clinical application beginning in 1968. Convincing data for the superiority of the internal mammary artery as a conduit for coronary revascularization was brought forth in the early 1980s from Grondin and colleagues in Montreal and from Loop, Lytle, and colleagues at the Cleveland Clinic. Saphenous veins were demonstrated to have approximately a 50 to 60 per cent 10-year patency compared with approximately a 90 per cent 10-year patency for the internal mammary artery graft. Importantly, this also translated to a significantly decreased need for subsequent revascularization and improved patient survival.

Indications for coronary artery bypass

Comparison with medical therapy

The direct aortocoronary bypass operation developed by Favaloro was remarkably successful. With appropriate patient selection Favaloro reported mortality rates of less than 5 per cent even in early series. Within a decade advances in technology and continued careful selection led to series reporting mortality rates of less than 1 per cent in patients undergoing elective coronary bypass. There was an explosive growth of the utilization of coronary artery bypass as it was particularly effective in the relief of angina. Disabling angina remains the primary indication for coronary artery bypass and is a sufficient indication, if angiography shows bypassable distal

coronary arteries, with or without survival benefit. Freedom from first return of angina in a heterogeneous group of patients reported by Dr Kirklin in the 1991 American College of Cardiology/American Heart Association (ACC/AHA) Guidelines for Coronary Bypass Graft Surgery was 83 per cent at 5 years. The application of coronary bypass grafting in patients with disabling angina is not controversial. It is an effective and reliable therapy. Controversy has existed, however, in the application of coronary artery bypass in patients with mild angina in whom the risk and cost of operation must be balanced against a survival benefit without a marked improvement in the quality of life.

Coronary bypass surgery is one of the first operations to be prospectively evaluated with randomized, multi-institutional studies. Three major studies were accomplished in the 1970s: the Veteran's Administration Coronary Bypass Trial, the European Coronary Bypass Trial, and the Coronary Artery Surgery Study (**CASS**). Compared with current surgical practice, these early randomized trials focused upon a very select group of patients and utilized techniques both intraoperatively and postoperatively that would be considered substandard by current thinking. Only patients under the age of 66 were included (the typical mean age for many current institutions performing coronary artery bypass is 65), only one trial (CASS) included women, and most patients were very mildly symptomatic with class I or II angina (the European trial did include some patients with moderate angina). In general only those patients for whom the referring cardiologist felt that either medical therapy or coronary bypass would be equivalent therapy were included in the studies. In the CASS study 25 000 patients underwent coronary angiography and only 780 patients were eventually randomized to medical or surgical therapy (a randomization rate of only 3.1 per cent!). Clearly, generalization of these results to the bulk of patients currently undergoing evaluation for coronary bypass surgery is quite inappropriate.

Other factors now considered to be the standard of care were not utilized in the early randomized trials: (i) the Veteran's Administration study and the European study did not utilize internal mammary arteries and this artery was only utilized in 14 per cent of the patients enrolled and randomized in CASS; (ii) effective techniques of myocardial protection, refined in the 1980s, were not utilized; (iii) aspirin was not utilized in either the early or late postoperative period to maintain graft patency; and (iv) appropriate medical therapy was not utilized after operation—b-blockers were utilized in only half of the patients, ACE inhibitors were not utilized in those patients with poor ventricular function, and lipid lowering therapy was not utilized. Despite these factors which mitigated against benefit from coronary bypass grafting, several subgroups were identified with significantly improved survival in the surgically assigned group. The Veteran's Administration identified quickly that patients with left main coronary artery disease (greater than 50 per cent stenosis) should undergo coronary artery bypass surgery and subsequent trials omitted these patients. The CASS trial showed significant benefit in patients with triple vessel disease and impaired ejection fraction (between 0.34 and 0.50); 84 per cent of patients had an improved 7-year survival following surgery compared with 70 per cent of those who received medical treatment. Patients with more severe left ventricular function were excluded from the early randomized trials. The European study demonstrated that coronary artery bypass grafts improved overall survival in patients with three vessel disease (5-year survival 94 per cent compared with 82 per cent); improved survival if an important stenosis of the left anterior descending artery was present along with stenosis of at least one other vessel (5-year survival 93 per cent compared with 82 per cent), and significant improvement in patients with left ventricular dysfunction and a positive exercise test (5-year survival of 92 per cent compared with 79 per cent).

A comprehensive reanalysis of the three randomized studies, along with four smaller randomized trials conducted between 1972 and 1984, was conducted by Yusef together with the principal investigators from each of the randomized trials. Importantly, original data from each of the randomized trials were reanalyzed and, wherever possible, definitions were standardized. This led to the creation of a data set of some 2649 patients with chronic stable angina randomly allocated to coronary bypass or medical therapy. Abnormal ejection fraction was defined as less than or equal to 50 per cent diameter reduction and significant coronary stenosis was defined as greater than 50 per cent reduction. The cumulative mortality is depicted in Fig. 3. The cumulative mortality is tabulated in Table 1 along with risk reductions at 5, 7, and 10 years. Risk reductions with coronary bypass were significant at all three time intervals. It is notable that 40 per cent of patients initially assigned to medical treatment underwent coronary artery bypass by 10 years. These crossovers tend to occur in the highest-risk medical patients (those most likely to suffer an unfortunate event if they remained in the medical group) and therefore there is an underestimation of the relative benefits of coronary bypass. Because of perioperative mortality, the relative advantage of coronary bypass was not seen for 2 to 3 years, but the advantage in favor of initial coronary bypass increased up to 5 to 7 years. This relative advantage is narrowed again by 10 to 12 years. This narrowing of relative advantage is probably a consequence of three factors. First, long-term follow-up of any two modes of therapy will eventually lead to convergence of survival curves: all patients eventually die. Second, the high rate of surgery among high-risk patients initially assigned to medical therapy tends to ‘level' the results with time. Third, the development of vein graft atherosclerosis in the surgical group (approximately 50 per cent of the vein grafts are closed at 10 years) tends to diminish the relative advantage of revascularization in the surgical group, particularly since arterial conduits were rarely utilized in these patients.

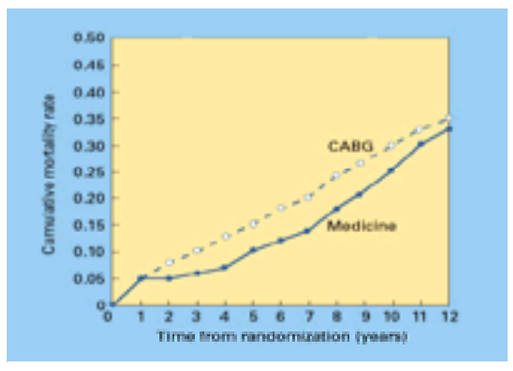


Fig. 3. Cumulative total mortality from meta-analysis of seven prospective randomized trials comparing coronary artery bypass grafting (CABG) with medical therapy.

Mortality (year)	CABG group (%)	Medical group (%)	Relative risk of CABG group	p value
5	10.2	15.8	0.61	*0.00001
7	15.8	21.7	0.68	*0.0004
10	26.4	30.5	0.83	0.005

*From Yusef et al. Lancet 1994; 344: 563–70.
CABG, coronary artery bypass graft.

Table 1 Meta-analysis of seven randomized trials—overall mortality*

Subgroup analysis was also highly revealing (Table 2). Notable is the fact that the relative benefit of coronary artery bypass, compared with medical therapy, at 5 years was approximately 60 per cent with regard to survival in many subgroups. A survival benefit was not demonstrated in those subgroups with minimal and early coronary disease (patients with one or two vessel disease without left anterior descending involvement and patients with normal exercise tests). In other subgroups the relative benefit was evident, but the absolute benefit was related to the risk of medical management. That is, the higher the risk of medical management for 5 years, the greater the absolute benefit of coronary artery bypass during the 5-year interval. A good example of this is the subgroup analysis related to left ventricular function (Fig. 4). Although the relative benefit of coronary bypass is essentially the same with normal or abnormal left ventricular function, the absolute benefit over 5 years was 5.2 per cent in patients with normal left ventricular function and 11.3 per cent in patients with abnormal left ventricular function. This is directly proportional to the relative risk of 5 years of medical treatment in patients with either normal or abnormal left ventricular function.

Five-year mortality	*CABG group (%)*	*Medical group (%)*	*Relative risk of CABG group*	*χ_1^2 (χ_2) P_1 value*
Left main disease	11.7	36.5	0.32	0.004
Triple vessel disease	10.1	17.6	0.58	< 0.001
One or two vessel disease				
No LAD disease	8.7	8.3	1.05	0.88
LAD disease present	8.6	14.8	0.58	0.05
Left ventricular function				
Normal	8.1	13.3	0.61	< 0.001
Abnormal	14.9	25.2	0.59	0.02
Exercise test				
Normal	9.0	11.6	0.78	0.38
Abnormal	8.7	14.8	0.52	< 0.001
Angina				
Class I, II	10.4	12.5	0.83	0.005
Class III, IV	12.8	22.4	0.57	0.001

*Yusuf et al. Lancet 1994; 344: 563–70.
CABG, coronary artery bypass graft; LAD, left anterior descending.

Table 2 Meta-analysis of seven randomized trials—subgroup analysis*

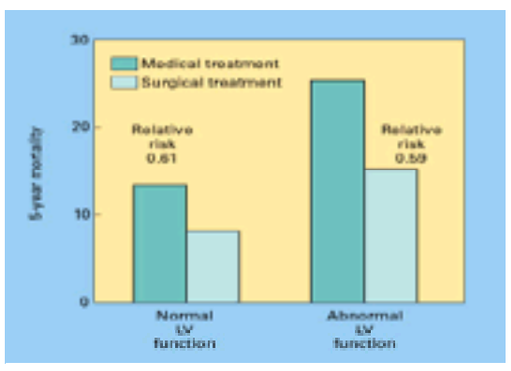


Fig. 4. The relative benefit of surgical therapy over medical therapy is the same for patients with normal left ventricular function compared with patients with abnormal left ventricular function. However, the absolute benefit over 5 years is greater for those patients with abnormal left ventricular function because the absolute risk of medical therapy is higher in patients with abnormal left ventricular function.

Another conclusion which appears to be supported by the subgroup analysis of the seven randomized trials is the fact that the relative benefit of coronary bypass is greater in patients in whom there is evidence of active ischemia: those patients with abnormal exercise tests compared with those with normal ones (relative risk 0.52 compared with 0.78) and those patients with class III or IV angina compared with those with class I or II (relative risk 0.57 compared with 0.83).

Analysis of survival in the randomized trials at 10 years or longer is most appropriately done by examining extension of overall survival during the first 10 years rather than examining cumulative survival at 10 years (Fig. 5). Examination of cumulative survival is problematic for two reasons. First, all survival curves will converge eventually at zero survival and if cumulative survival were the only yardstick, no difference in any mode of therapy would ever be demonstrable in the very long term. What is important is the integration of the area between the curves during the follow-up interval which allows a prediction of the average extension of life for the various treatment modalities. A second important reason that extension of survival is an appropriate means of analysis is the aforementioned fact that 40 per cent of the patients in the ‘medically treated' subgroup underwent coronary artery bypass during the 10-year observation interval. The ‘medically treated' group might therefore be best considered a ‘delayed coronary artery bypass' group in which coronary artery bypass was not offered initially for mild symptoms but was reserved for increasing symptoms as the disease process progressed. This is an important point as the clinician considers the appropriate therapy for patients with mild (non-disabling) symptoms in the current era. The meta-analysis strongly suggests that in patients with advanced coronary artery disease (patients with proximal left anterior descending disease, patients with triple vessel disease, or patients with impaired left ventricular function), early coronary artery bypass is appropriate despite the absence of disabling symptoms. This result was suggested, although less strongly, by the individual randomized trials. The extension of this conclusion to patients with proximal left anterior descending disease is possible only by the power available in the meta-analysis.

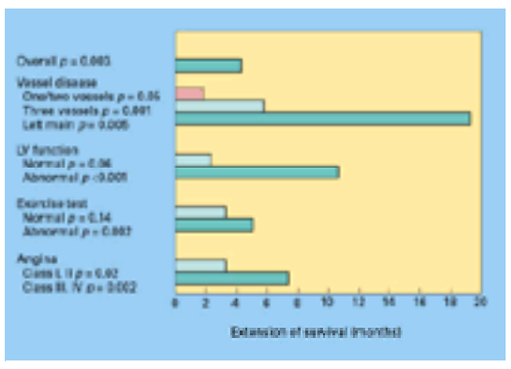


Fig. 5. Extension of survival by coronary artery bypass compared with medical therapy after 10 years of follow-up in various subgroups from meta-analysis of seven randomized trials.

Analysis of 10-year results depicted by extension of survival after 10 years of follow-up is depicted graphically in Fig. 5. Notable is the fact that the overall difference of 4.26 months between the ‘early coronary artery bypass' and ‘delayed coronary artery bypass' group is highly significant (p = 0.003) despite the fact that the cumulative mortality curves have begun to converge at 10 years (Fig. 3). Even if the curves converge completely as time progresses, the extension of survival benefit will persist (unless for some reason there should be a late survival disadvantage of the early coronary artery bypass group). Also notable in the 10-year analysis is the fact that patients with early atherosclerotic disease (normal left ventricular function, one or two vessel disease, normal exercise test) have a relatively small and non-significant benefit from early rather than delayed coronary artery bypass.

In addition to relief of disabling symptoms and prolongation of life, it has been proposed that another appropriate indication for coronary artery bypass is the prevention of adverse sequelae of coronary disease such as myocardial infarction. Registry studies have suggested that, in patients with advanced atherosclerotic disease, coronary artery bypass may decrease late myocardial infarction. In the meta-analysis mentioned above, no overall impact of coronary bypass on myocardial infarction could be demonstrated, related in part to perioperative infarction as an additive effect in the early coronary artery bypass group.

In summary, compared with medical therapy, coronary artery bypass provides superior relief of disabling symptoms in patients with bypassable vessels. Moreover, in patients with mild symptoms, coronary artery bypass provides a significant extension of survival in patients with advanced atherosclerotic disease, namely patients with proximal left anterior descending disease, triple vessel disease, left ventricular dysfunction related to prior ischemic damage, and patients with positive exercise tests. The introduction of angioplasty in the 1980s offered an alternative mode of revascularization which was quickly applied to patients with single vessel disease and has been extended to patients with multivessel disease, particularly as new percutaneous techniques have been developed with progressively decreasing restenosis rates.

Comparison of coronary artery bypass with percutaneous techniques

When percutaneous angioplasty was introduced in the early 1980s, it was proposed as a treatment for the early stages of coronary disease and generally applied to patients with proximal single vessel stenoses. Although it was widely considered to be an alternative to coronary artery bypass, in fact, in the early years, angioplasty was utilized as an alternative to medical therapy more commonly than it was utilized as an alternative to coronary bypass. Rates of coronary artery bypass did not decrease with the advent of angioplasty in the United States. More recently, however, angioplasty has been applied to patients with more advanced coronary disease and it has been suggested as an alternative to coronary artery bypass in patients with multivessel disease. This counterposition of angioplasty and coronary artery bypass led to several prospective randomized trials of the two therapies in patients with multivessel coronary disease. Unfortunately most of these trials have been

flawed by making inappropriate comparisons, by being underpowered, by inclusion of a large number of low-risk patients, and by utilizing endpoints which are inappropriate. In order to demonstrate with reasonable certainty that angioplasty and coronary bypass are equivalent therapies (i.e. to exclude a relative risk difference of 20 per cent with 90 per cent power) trials with approximately 4000 patients per treatment arm would be necessary. If one lowers one's goals and seeks to detect a relative risk difference of 30 per cent or greater, then trials with 2000 patients in each treatment arm would be appropriate, presuming that the patients are in moderate- to high-risk categories in which coronary artery bypass has been shown to be preferable to medical therapy. Moreover, one would require a high rate of adherence to the original treatment plan to prevent the confounding effect of crossover. Of the prospective trials that have been conducted, the Bypass Angioplasty Revascularization Investigation (**BARI**) was the largest trial. Unfortunately it was seriously underpowered with only approximately 900 patients in each treatment arm. Because this was the largest trial, however, it will be discussed in detail.

In the BARI trial 1829 patients with multivessel disease at 18 institutions were randomized to angioplasty or coronary artery bypass as initial therapy. The mean age of patients was 61 years. Forty-one per cent of the patients had triple vessel disease and 59 per cent had double vessel disease. In the group assigned to angioplasty, this was attempted for an average of 2.4 stenoses. Multivessel angioplasty was attempted in 70 per cent of the patients. At least one stenosis was successfully dilated in 88 per cent of the patients and all stenoses were successfully dilated in 57 per cent of the patients. Among the patients undergoing coronary bypass grafting as initial therapy, an average of 3.1 coronary arteries were bypassed per patient and all intended vessels were bypassed in 91 per cent of the patients. At least one internal mammary artery was utilized in 82 per cent of the patients. The median hospital stay after angioplasty was 3 days and the median stay after coronary bypass was 7 days.

The mortality and morbidity during the hospital stay was approximately equivalent in the two groups. These complications are presented in Table 3. Mortality was essentially identical with 1.3 per cent mortality in the coronary artery bypass group and 1.1 per cent in the angioplasty group. Most notably, the data in Table 3 indicate that the in-hospital complications of angioplasty are as high or higher than the major in-hospital complications of coronary bypass. If one totals the number of secondary procedures performed in the operating room, angioplasty suite, or renal dialysis suite, 70 such secondary procedures were required in the coronary bypass group or 0.08 procedures per patient. One hundred and forty such secondary procedures were required in the angioplasty group, or 0.16 secondary procedures per patient (Fig. 6). This is related primarily to a 10.2 per cent coronary bypass rate after angioplasty (but during the same hospital stay). With regard to neurologic sequelae, 11 patients in the coronary bypass group had stroke, coma, or dementia compared with 10 patients in the angioplasty group (1.2 per cent compared with 1.1 per cent) (Fig. 7). The only category of in-hospital complication in which coronary bypass was noted to be higher than angioplasty was in the category of Q wave myocardial infarctions. These infarctions were determined strictly by electrocardiographic analysis. If all complications listed in Table 3 are totaled (including myocardial infarction in hospital, repeat procedures, neurologic complications, congestive failure or pulmonary edema, cardiogenic shock, and respiratory failure), 184 complications occurred in the 892 patients undergoing initial coronary artery bypass grafting (0.21 complications per patient) and 216 complications occurred among the 904 undergoing multivessel angioplasty (0.24 complications per patient) (Fig. 8). This analysis is especially problematic as patients and cardiologists have the perception that the complication rate after angioplasty for multivessel coronary disease is lower with angioplasty than it is with surgery. This is simply not the case. It is true that after angioplasty patients return to work sooner and have a lower rate of post-procedure morbidity than those recovering from cardiopulmonary bypass and surgical incisions, which delay a full return to a normal lifestyle. However, for multivessel disease, the complication rate is similar for the two procedures.

Complication	CABG $n = 892$ No. (%)	PTCA $n = 904$ No. (%)	CABG worse than PTCA	PTCA worse than CABG
Death	12 (1.3)	10 (1.1)		
Q wave infarct	40 (4.5)	19 (2.1)	$p < 0.01$	
Emergency CABG	3 (0.3)	51 (5.6)		$p < 0.001$
Emergency PTCA	0	19 (2.1)		$p < 0.001$
Non-emergency CABG	0	28 (3.1)		$p < 0.001$
Non-emergency PTCA	0	11 (1.2)		$p < 0.001$
Peripheral embolectomy	3 (0.3)	11 (1.2)		$p < 0.05$
Reoperation for bleeding	70 (7.8)	4 (0.4)	$p < 0.001$	
Wound dehiscence or infection	27 (3.0)	4 (0.4)	$p < 0.001$	
Renal failure requiring dialysis	1 (0.1)	2 (0.2)		
Stroke	7 (0.8)	2 (0.2)		
Coma	0	4 (0.4)		
Dementia	4 (0.4)	4 (0.4)		
Congestive heart failure or pulmonary edema	75 (8.4)	74 (8.2)		
Cardiogenic shock	7 (0.8)	11 (1.2)		
Respiratory failure	28 (3.1)	71 (7.9)	$p < 0.05$	

Table 3 In-hospital complications of coronary artery bypass grafting (CABG) and percutaneous coronary angioplasty (PTCA) from the BARI trial

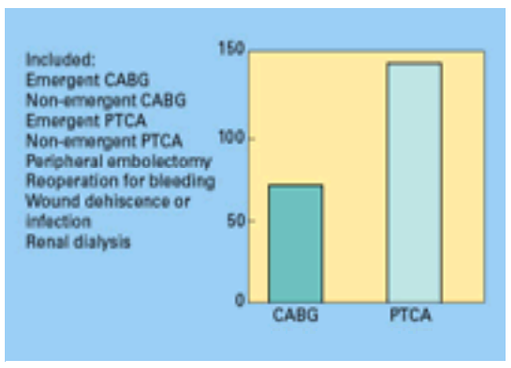


Fig. 6. In hospital, procedures in the BARI trial were significantly greater after angioplasty than after coronary bypass.

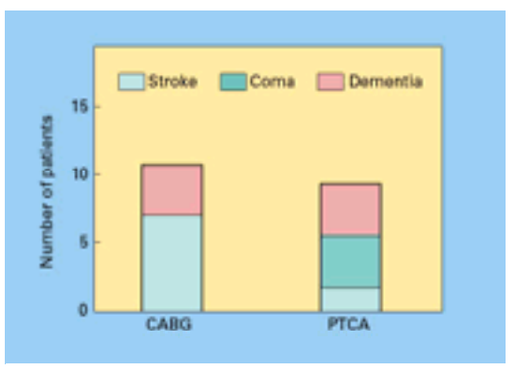


Fig. 7. Although seven patients suffered a stroke after coronary bypass in the BARI trial compared with only two patients after angioplasty, if one totals both type I and type II neurologic defects (stroke, coma, and dementia), 11 patients in the coronary bypass group had a major neurologic complication compared with 10 patients in the angioplasty group.

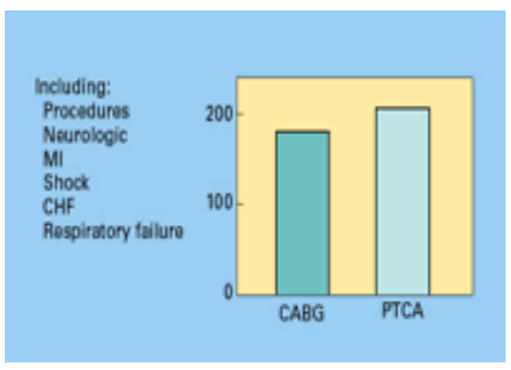


Fig. 8. The total complications in the BARI trial were slightly higher in the angioplasty group. This is contrary to the commonly held perception of the cardiology community and the general public.

The primary endpoint of the BARI trial was ‘all-cause death at 5 years'. The trial was designed to demonstrate equivalence between the two treatment methods and a sample size of 1200 patients per treatment arm was projected as necessary with an estimated 5-year mortality of about 5 per cent. The goal of the sample size was that an observed difference of 2.5 per cent would be excluded (‘the maximum acceptable absolute difference between CABG and PTCA event rates'). In fact only about two-thirds of the number of patients intended were actually randomized (1829 rather than 2400). The 5-year mortality among patients assigned to coronary bypass was 10.7 per cent and the mortality among those assigned to angioplasty was 13.7 per cent (absolute difference 3 per cent, $p = 0.19$. This nearly 22 per cent relative risk reduction in favor of coronary bypass did not reach statistical significance. However, because the study enrolled fewer patients than intended, the study had less than 40 per cent power to detect such a difference reliably. Therefore, the BARI study results are consistent with a greater than 2.5 per cent absolute difference in mortality in favor of coronary bypass surgery (the difference stated in the protocol). BARI did not demonstrate that coronary bypass and angioplasty are equivalent with regard to revascularization of multivessel disease. Indeed, an analysis of cardiac deaths revealed a 5-year mortality of 8 per cent in those assigned to angioplasty compared with 4.9 per cent in those assigned to coronary bypass (relative risk 1.55, significant with a p value of 0.02). Subgroup analysis of the BARI trial revealed that the primary difference between angioplasty and coronary artery bypass appeared to be in patients who were being treated for diabetes. Five-year mortality among those receiving treatment for diabetes assigned to coronary bypass was 19.4 per cent compared with 34.5 per cent among those assigned to angioplasty. This is probably related to the fact that patients receiving treatment for diabetes have more extensive coronary disease and, in particular, have more extensive distal coronary disease which is less easily treated by percutaneous therapy. Further analysis of this subgroup revealed that the benefit occurred primarily among patients treated with internal mammary artery grafts.

Another important consideration from the BARI trial is the frequency of repeat revascularization procedures during the follow-up interval. Over the 5 years of follow-up, 31 per cent of the patients randomized to angioplasty had undergone coronary artery bypass. Thirty-four per cent had had a repeat angioplasty. Eleven per cent of these patients had undergone both angioplasty and coronary artery bypass. In total only 45.5 per cent of patients initially randomized to angioplasty had survived 5 years without a repeat revascularization procedure. Among the patients randomized to coronary artery bypass grafting, only 1 per cent of patients underwent repeat coronary artery bypass grafting and 7.3 per cent underwent angioplasty. Ninety-two per cent of patients randomized to coronary artery bypass grafting had no repeat revascularization procedure during the 5 years of follow-up (Fig. 9).

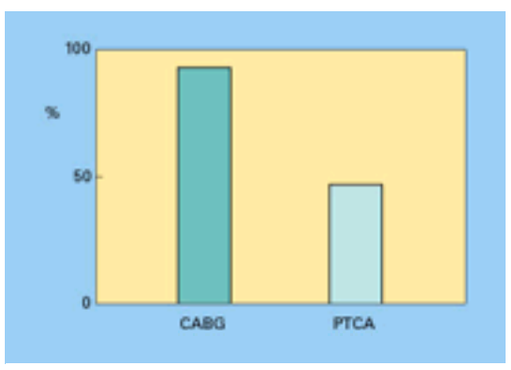


Fig. 9. Freedom from any repeat revascularization procedures was dramatically higher in the coronary artery bypass subgroup compared with the angioplasty subgroup in the BARI trial.

The BARI trial also conducted a prospectively designed analysis of comparative cost of the two procedures from a subgroup of participating centers, comprising a total of 934 of the1829 patients. The mean initial hospital cost of angioplasty was 65 per cent that of surgery, but after 5 years the accumulative cost for patients undergoing initial surgical therapy was only $2700 more than the cumulative cost of patients assigned to initial angioplasty (5 per cent difference). Since the surgical cohort had a 3 per cent higher overall 5-year survival, the cost of this survival benefit could be calculated. From the BARI trial the cost per year of survival benefit for surgical therapy was found to be approximately $26 000 per year.

Eight other randomized trials of coronary artery bypass compared with angioplasty have been published. Many of these trials have severe limitations. A number included large numbers of patients with single vessel disease. The German Angioplasty Bypass Surgery Investigation was limited by a restricted use of the internal mammary artery (37 per cent) and by a protracted delay between the time of randomization and surgical therapy (averaging 53 days). During this delay there was a 2.5 per cent mortality in the group of patients randomized to surgery. These four patient deaths in the pretreatment interval in the coronary bypass group represented 44 per cent of the total deaths in the coronary bypass group during the entire study period! Despite these major concerns about a number of the studies, a meta-analysis of the eight studies (excluding BARI) does provide some useful information. This meta-analysis included 3371 patients with a mean follow-up of 2.7 years (a number of trials had only 1 year follow-up). Among patients randomized to angioplasty 18 per cent required coronary bypass within a year and the subsequent need for coronary bypass was approximately 2 per cent per year. The requirement for any repeat revascularization procedure (angioplasty and/or coronary bypass graft) in the first year of follow-up was 34 per cent in the angioplasty group and only 3.3 per cent in the patients randomized to coronary bypass. Importantly, the prevalence of angina in the angioplasty group was 1.56 times as high after 1 year as in the coronary bypass group.

The Emory Angioplasty Surgery Trial (**EAST**) was one of the early, single institution trials comparing patients with multivessel coronary artery disease randomized to either angioplasty or surgery. Sixty per cent of the patients had double vessel and 40 per cent had single vessel disease. Follow-up included a non-invasive study for ischemia (a thallium scan) at 1 year and catheterization at 1 and 3 years. This led to repeat procedures in both the angioplasty and the surgery group at 1 year related to discovery of new stenoses. Follow-up is now available up to 8 years in this trial. The survival benefit in the EAST trial is fairly similar to the survival benefit seen in the BARI trial (Fig. 10). In the EAST trial because of even smaller numbers this difference is not significant (Fig. 6). At 8 years only 34 per cent of patients initially assigned to angioplasty are free from subsequent coronary bypass or angioplasty and 73 per cent initially assigned to coronary bypass are free of subsequent coronary bypass or angioplasty (Fig. 11). Notably, the need for subsequent coronary bypass is markedly different among the two groups, with only 3 per cent of patients undergoing coronary artery bypass grafting requiring a repeat procedure within 8 years and 29 per cent of patients undergoing angioplasty requiring coronary bypass within 8 years. The EAST trial did not reveal any difference in patients with diabetes assigned to either angioplasty or surgery. The power of the EAST trial would be inadequate to reveal such a difference. The philosophy of the EAST investigators was somewhat different from that of the BARI investigators as both angioplasty operators and surgeons were required to agree that angioplasty or surgery was appropriate for revascularization of all index segments prior to enrollment of patients in the study. The potential for complete revascularization was not a requirement for enrollment in the BARI trial. This may explain the difference in the two trials relative to patients with diabetes.

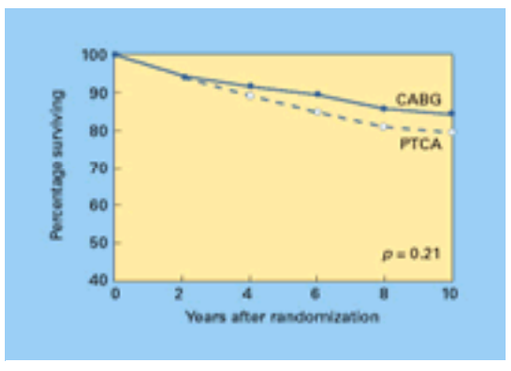


Fig. 10. Survival after randomization in the EAST trial is similar to the survival benefit seen in the BARI investigation. Of great interest is the fact that this survival benefit seems to extend to 10 years.

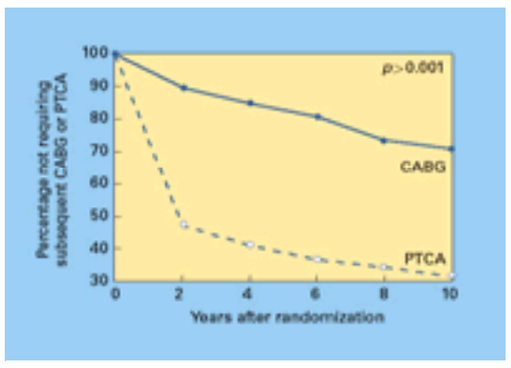


Fig. 11. The need for repeat revascularization in the EAST trial again demonstrated a dramatic difference between patients assigned to coronary bypass compared with those assigned to angioplasty.

Several other observations can be made from the EAST trial. The freedom from angina and freedom from any antianginal medications was significantly higher in the coronary bypass group compared with the angioplasty group. This difference diminished as follow-up was continued to 3 and 5 years as repeat revascularization was utilized in symptomatic patients (including considerable crossover in the angioplasty group to surgical therapy). Analysis of complications from the EAST trial supports the conclusion reached above, that the complication rate of angioplasty (if unplanned revascularization during the initial hospital stay is counted as a complication) is equally high as it is for coronary bypass during the initial hospital stay. The cumulative 3-year cost of angioplasty and coronary bypass was also almost identical during 3 years of follow-up in the EAST trial although the initial cost of angioplasty was approximately two-thirds that of the initial cost of surgery.

Summary of indications for coronary artery bypass

The coronary bypass operation is indicated for relief of symptoms and for prolongation of life. In patients with disabling angina, one can expect relief of angina in approximately 90 per cent of those with reasonable distal vessels. This effectiveness can be improved by nuclear scans demonstrating quantifiable ischemia at the time that the patient experiences angina. The results of the randomized trials of coronary bypass compared with percutaneous therapy have confirmed and extended this conclusion. Not only did coronary bypass effectively relieve angina in the symptomatic patients enrolled in the randomized trials, but freedom from angina and antianginal medications was superior in the coronary bypass cohorts compared with the angioplasty cohorts.

The benefit realized from the use of coronary bypass to relieve disabling symptoms must be balanced against the risk of the operation and adjusted to the potential activity level of the individual patients. The risk of operation may be very low in selected patients. In a series of approximately 1300 patients with single and double vessel disease in the early 1980s at Emory University with age less than 66, ejection fraction greater than 0.35, and no congestive heart failure, only one patient died during the hospital stay (hospital mortality 0.07 per cent). These young healthy patients had a very low risk and, in addition, had a potential for renewal of an active lifestyle for a prolonged period of time. In these patients coronary bypass for relief of disabling symptoms is strongly indicated. The patient at the other end of the spectrum would be 75 years old with severe arthritis and class II angina. The benefit of coronary bypass in this patient would be minimal as the patient's other illnesses would prevent a return to an active lifestyle and the risk of operation would be comparatively greater. In this instance, continued medical therapy or perhaps angioplasty would be appropriate therapy. In the treatment of angina with coronary bypass one must also be aware that as coronary disease progresses angina often returns. The hazard function for the return of angina is low for the first 5 years after operation and then begins to rise, related perhaps to closure of bypass conduits.

The second important indication for coronary bypass is prolongation of life. The aforementioned meta-analysis of the randomized trials of coronary bypass compared with medical therapy has defined patient subsets whose survival is enhanced by the coronary bypass operation. In particular, patients with advanced coronary artery disease benefit with regard to prolongation of survival from coronary artery bypass surgery. This includes patients with left main disease, triple vessel disease, and double or single vessel disease which includes proximal left anterior descending stenosis. Patients with double or single vessel disease not including proximal left anterior descending stenosis did not appear to benefit from the application of coronary bypass for prolongation of life. The recommendation for surgical therapy is more urgent if the patient has left ventricular dysfunction (related to prior myocardial damage or to hibernating myocardium) or if the patient has ongoing ischemia (unstable angina, markedly positive exercise test). In patients with severe left ventricular dysfunction, and chronic congestive heart failure, however, coronary artery bypass is unlikely to relieve symptoms unless nuclear studies can demonstrate a significant area of hibernating myocardium which might improve with revascularization.

Technical considerations

Progressive improvements in the techniques of coronary bypass have allowed surgeons to apply the procedure to increasingly difficult cases in the last quarter of a century. Most notable among these improvements have been improved techniques of myocardial protection, increased use of arterial conduits, and improved perioperative or adjunctive care (preoperative manipulation, anesthetic management, and postoperative care).

Myocardial protection

Coronary artery bypass was initially performed upon a beating heart with or without perfusion of the distal coronary segment. Early methods of hypothermic cardiac arrest which had been moderately successful in valve surgery were applied to coronary bypass with limited success. Cardioplegia techniques, the infusion of high concentration potassium solutions into the heart to stop contraction, greatly accelerated the application of coronary artery bypass as near complete cardiac recovery could be accomplished with simple crystalloid solutions. Refinement of cardioplegia techniques occurred in the late 1970s as blood cardioplegia was introduced and the concept of resuscitative cardioplegia emerged. With appropriately designed blood cardioplegic solutions, metabolically damaged myocardium or even infarcting myocardium can be resuscitated during the aortic cross-clamp interval such that the heart is stronger after the cross-clamp interval than it was prior to it. This has led to dramatic improvements in coronary bypass surgery for patients with very poor left ventricular function who have large areas of hibernating myocardium, and similarly dramatic improvements in patients in cardiogenic shock or with early, evolving myocardial infarction. For patients who have no impairment of myocardial metabolism prior to operation, crystalloid cardioplegia techniques offer superior visibility and near complete recovery of function. However, for patients with impaired myocardial function, especially those with metabolic compromise at the beginning of the operation, blood cardioplegia techniques are clearly preferable.

Increased use of arterial conduits

Although saphenous vein grafts have been the mainstay of coronary artery bypass surgery since its inception, there is increasing recognition that long-term patency of these conduits is a major concern. Only 50 to 60 per cent of saphenous vein grafts are patent at 10 years and a decreasing survival benefit and increasing return of angina appears to be associated with closure of these conduits. The Cleveland Clinic group and careful angiographic follow-up from Montreal have demonstrated conclusively that the internal mammary artery graft to the left anterior descending vessel offers benefit with regard to freedom from repeat revascularization and survival benefit in patients undergoing coronary artery bypass surgery (Fig. 12 and Fig. 13). The benefit of the use of a second internal mammary graft to lesser vessels (the right coronary artery or the circumflex artery) is less well established. There does appear to be some emerging evidence, again from Cleveland Clinic data, that this second arterial conduit offers some survival benefit.

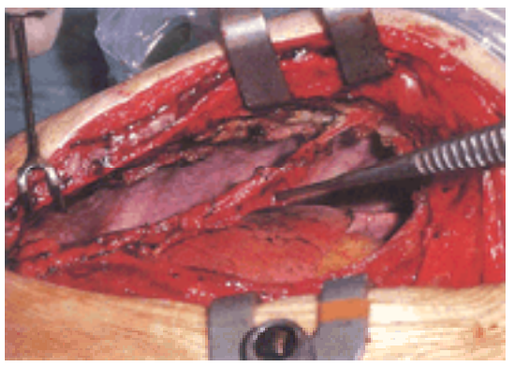


Fig. 12. The left internal mammary artery dissected from its bed in the parasternal position.

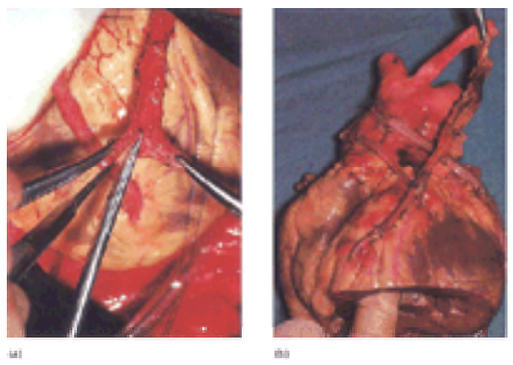


Fig. 13. (a) Left internal mammary pedicle prepared for anastomosis. (b) Postmortem specimen showing the left internal mammary artery originating from the left subclavian and grafted onto the left anterior descending coronary. Vein grafts supply an intermediate and circumflex vessel.

As repeat operations are becoming increasingly common, adequate saphenous vein grafts or internal mammary grafts for utilization as conduits have become limited and consideration of other conduits is necessary. Other arterial conduits which have been successfully utilized are the gastroepiploic artery (harvested from the greater curvature of the stomach and used either as an *in situ* or as a free graft), the inferior epigastric artery (harvested from behind the rectus sheath and utilized as a free graft), and the radial artery (utilized as a free graft). The radial artery in particular has an emerging popularity as a backup conduit in case saphenous veins and internal mammary arteries are not available and also as a first-line arterial conduit in an attempt to increase the number of arterial grafts performed. The radial artery was used some 20 years ago in France on a number of patients, but problems with spasm and early graft patency led to discontinuation of its use. Long-term follow-up of a number of these patients has revealed excellent long-term graft patency. Refined techniques of radial artery harvest have been developed such that early patency is now superior or equal to vein graft patency and intermediate-term patency appears to be considerably better than vein grafts although perhaps not quite as good as the left internal mammary. The left internal mammary graft with a 90 to 95 per cent 10-year patency remains the benchmark for long-term patency. Radial grafts, particularly since they are usually applied to secondary targets (smaller coronary arteries with less flow), are unlikely to match this benchmark, but appear to be quite promising in early series.

Reoperations

A very difficult technical problem has been coronary artery reoperations. One can expect an 85 to 90 per cent relief of angina with the initial coronary bypass operation, but only approximately a 70 per cent relief of angina with reoperation. Mortality is two to three times as high with coronary reopertion compared with the primary operation. Despite this, it is clear that most patients with disabling angina and reasonable distal targets should be offered coronary reoperation. There are several technical considerations that are important with reoperation. An oscillating saw is used to reopen the sternum. The sternal wires may be opened and not removed when the oscillating saw is used, preventing the saw from cutting into the soft tissue behind the posterior table of the sternum. The wires are then removed and the operation proceeds. Internal mammary artery retractors should be used to dissect the left side of the mediastinum if a patent internal mammary artery graft is present. Saphenous vein grafts often contain atheromatous debris which may be dislodged during dissection. For this reason manipulation of these vessels is avoided. This may require that dissection of the heart be accomplished after the patient is on cardiopulmonary bypass and after the heart is arrested. Severely diseased saphenous vein grafts from previous operations are replaced or at least ligated. Mildly diseased saphenous vein grafts may be left intact. It has been found to be particularly dangerous to replace a previous vein graft to the left anterior descending vessel with an internal mammary graft. The flow reserve in the internal mammary graft appears to be considerably less than that in the saphenous vein graft and anterior myocardial ischemia is an uncommon, but sometimes fatal, sequelae of substituting a new internal mammary graft for an old saphenous vein graft to the anterior wall. In this instance, the prudent technique is augmentation of the new internal mammary graft with a new saphenous vein graft to the same myocardial region.

Perioperative management

Analysis of the results of coronary bypass surgery, to be discussed in the next section, allow preoperative risk stratification. This is a particularly important topic as one applies the operation to more severely ill patients. Almost as important is the concept of risk neutralization by modification of preoperative factors and by perioperative management strategies.

Modification of preoperative risk factors

Many preoperative risk factors cannot be altered. These include age, severity of disease, prior myocardial damage, and concomitant peripheral vascular disease. Several preoperative risk factors, however, may be modified. Prominent among these are acuity of operation, evolving cerebral vascular damage, and right ventricular failure secondary to acute ischemia.

Acute myocardial ischemia often forces an urgent or emergency operation. Appropriate pharmacologic therapy, insertion of an intra-aortic balloon pump, or on occasion, percutaneous dilation of ‘culprit' stenoses may convert an emergency situation into an elective procedure. The benefit of temporizing therapy must be weighed against the risk of delay. The benefit is particularly appropriate if the patient has signs and symptoms of congestive heart failure which can be reversed, allowing the lungs and kidneys to recover prior to proceeding with coronary bypass.

A particularly dangerous situation which should be recognized prior to operation is right ventricular failure secondary to ischemic right ventricular injury (either myocardial necrosis or ischemia). Any patient who has suffered an acute right coronary occlusion (including those who have had subsequent reopening with thrombolytic therapy) should be evaluated for signs or symptoms of acute right heart failure. In an equivocal situation, echocardiography may be of benefit in determining right ventricular function prior to operations. If the patient has right ventricular dysfunction, delay of the operation for 4 to 6 weeks will allow recovery of the right ventricle and permit safe performance of coronary bypass. Acute right ventricular dysfunction is very difficult to reverse with coronary artery bypass and often leads to a fatal result if it is not recognized and avoided.

A recent cerebral vascular accident is another situation in which evaluation and occasionally delay will lead to an improvement in operative outcome. Evidence of a hemorrhagic component on a CT scan is especially ominous as this is associated with an 80 per cent chance of extension of neurologic damage during cardiopulmonary bypass. The operation should be delayed if at all possible if this is the case. A delay of 6 weeks or more is usually appropriate.

Preoperative evaluation of extracranial vascular disease is an important consideration as one attempts to reduce the risk of perioperative stroke. The incidence of stroke after coronary operation is related to extracranial cerebral vascular disease. The perioperative stroke rate is approximately 3 per cent for carotid stenoses less than 50 per cent, 8 per cent for stenoses between 50 and 80 per cent, and 11 per cent with stenoses greater than 80 per cent. Significant carotid disease has been discovered with increasing frequency in coronary bypass populations as older patients are referred for operation. Patient characteristics associated with carotid stenoses include age greater than 65, female gender, previous central nervous system symptoms suggestive of ischemia, cigarette smoking, peripheral vascular disease, and left main stenosis. The presence or absence of bruits is unreliable for detecting carotid stenoses. Carotid duplex examination, however, is an effective screening tool. Detection of carotid stenosis of 70 to 80 per cent should prompt consideration for prophylactic carotid endarterectomy. With skilled peripheral vascular surgeons this procedure can be performed under local anesthesia with a very low cardiac risk and a 1 or 2 per cent risk of stroke.

Intraoperative management

Anesthetic techniques

Great advances have been made by our anesthesia colleagues. In early series of coronary artery bypass, myocardial CKMB (creatine kinase MB form) levels were elevated in 10 to 15 per cent of patients prior to institution of cardiopulmonary bypass! Modern anesthetic techniques have provided gentle induction of anesthesia. Intraoperative monitoring techniques including use of the Swan–Ganz catheter and intraoperative transesophageal echocardiography have allowed safe performance of anesthesia even in very difficult situations.

Prophylactic use of the intra-aortic balloon pump may be indicated in patients with severe left ventricular dysfunction. A retrospective evaluation of 163 patients with a

left ventricular ejection fraction of 0.25 or less revealed a 2.7 per cent mortality for patients who received a prophylactic balloon pump placed on the day of operation compared with an 11.9 per cent hospital mortality for patients not receiving the balloon pump prophylactically. Placement of the intra-aortic balloon pump immediately prior to operation seems to afford a benefit similar to that achieved by placing the balloon pump prior to induction of anesthesia.

Management of aortic atherosclerotic disease

A leading cause of cerebrovascular damage after coronary artery bypass appears to be perioperative atheroembolism from the ascending aorta or arch. This risk dramatically increases with age. Embolization appears to arise intraoperatively from manipulation of the ascending aorta during cannulation, cross-clamping for cardioplegic arrest, and clamping for proximal anastomoses. Transesophageal echocardiography can suggest disease in the arch, but adequate visualization of the ascending aorta is difficult. Epivascular echocardiography is especially useful in this situation and can detect threatening disease of the aorta. If the aorta is more than 4 mm thick, one should use a ‘no clamp' technique for coronary artery bypass grafting. Often an appropriate site for cannulation can be found and a second site for placement of a proximal anastomosis can be located by epivascular echocardiography. The patient is placed on cardiopulmonary bypass and cooled to 20 to 25°C. Perfusion pressure is maintained at 70 to 80 mmHg and a vent is placed in the left ventricle to prevent left ventricular distension. Distal anastomoses are constructed with the heart fibrillating (or beating slowly). The vent flow is adjusted to maintain a positive pressure within the cardiac chambers. At 20 to 25°C the pump can be turned off for 3 to 5 min at a time while one or two proximal anastomoses are constructed in the selected areas of the ascending aorta without placement of any clamps on the aorta. This brief period of circulatory arrest is extremely well tolerated at these temperatures. This technique does not allow for resuscitation of ischemic myocardium or metabolically impaired myocardium as mentioned earlier, but it appears to provide adequate myocardial protection in most situations without the risk of atheroembolism associated with clamping the ascending aorta. If disease of the ascending aorta is particularly severe, it may be necessary to replace the ascending aorta with a Dacron graft under circulatory arrest and to place the proximal anastomoses on the newly inserted ascending aortic graft.

Minimizing the adverse systemic effects of cardiopulmonary bypass

Cardiopulmonary bypass causes a profound disruption of homeostasis. Allergic reactions are possible. A whole-body systemic inflam-matory response of variable degree occurs. Hemodilution and hypo-thermia cause derangements in cellular metabolism and capillary leakage. It appears that mild hypothermia causes less systemic derangement than profound hypothermia. Normothermic cardiopulmonary bypass is associated with an increased risk of cerebrovascular damage. Mild hypothermia appears to be protective when the brain is exposed to short periods of ischemia with micro-and macroembolization. In patients with cerebral vascular disease, perfusion pressures above 70 mmHg appear to be appropriate in an attempt to prevent cerebrovascular damage. Corticosteroids, administered preoperatively, are inexpensive and are likely to attenuate allergic reactions if they should occur. These steroids are ideally given 1 to 2 h prior to initiation of cardiopulmonary bypass. Aprotinin, in addition to its hemostatic role, is an effective anti-inflammatory agent which appears to reduce the length of hospital stay and cost in high-risk patients, particularly in patients receiving a coronary bypass reoperation. Perioperative leukocyte depletion (by filtration) also appears to reduce the whole-body inflammatory response to cardiopulmonary bypass and improves outcomes.

Reduction of perioperative infection

Meticulous operative technique is the first step in the reduction of perioperative infection. Focused use of electrocautery, minimal use of foreign agents (e.g. bone wax), and careful hemostasis are important. Preoperative antibiotic administration reduces the risk of mediastinal infection, but efficacy is dependent upon administration in a timely matter such that tissue levels of drug are adequate at the time of microbial exposure. These drugs should be given an hour or so preoperatively and administration must be completed prior to incision. Cephalosporins are the current agents of choice for prophylaxis of infection for coronary operations. There does not appear to be any data to suggest that more than 48 h of prophylaxis is beneficial. In patients who are allergic to penicillin, vancomycin is an appropriate alternative to cephalosporins.

Diabetes is an independent risk factor for wound infection after coronary artery bypass. A prospective study has demonstrated that aggressive glucose control using continuous insulin infusion led to a significantly decreased rate of deep sternal wound infection (a risk reduction of 50 per cent!). These data have led many institutions to attempt to maintain intraoperative glucose levels below 150 mg per cent, considerably lower than the 180 or 200 mg per cent utilized in previous years.

Management of perioperative bleeding

Most patients who are under treatment for coronary artery disease are on aspirin therapy. Aspirin decreases platelet aggregation and increases postoperative blood loss. If appropriate, withdrawal of aspirin therapy for 10 days preoperatively is useful. Aprotinin reduces intraoperative and postoperative blood loss and is particularly appropriate in reoperation. Aprotinin does not appear to decrease early graft patency. It can cause a deceptive prolongation of the activated clotting time and heparin administration must be carefully managed (either by using an abnormally prolonged activated clotting time as a target or by using intraoperative heparin assays to adjust heparin dosage). a-Aminocaproic acid (Amicar) also appears to be an effective and less expensive alternative to aprotinin in the reduction of perioperative bleeding.

Postoperative management

Prevention of dysrhythmias

Postoperative atrial fibrillation has been a major source of morbidity after coronary artery bypass, prolonging hospital stay and leading to a threefold increase in the risk of postoperative stroke. If b-blockers have been utilized preoperatively, they should be continued postoperatively. Postoperative withdrawal of b-blockers doubles the incidence of postoperative atrial fibrillation after coronary artery bypass. Almost every study of b-blockers administrated for the purpose of reducing postoperative atrial fibrillation has shown benefit. There may be an even greater benefit if b-blockers are begun preoperatively. Low-dose propanolol seems to be as effective as other b-blockers in prophylaxis against atrial fibrillation and is inexpensive.

Amiodarone is currently undergoing trials for use as a prophylaxis against postoperative atrial fibrillation. One early trial has suggested a risk reduction of 50 per cent. Digoxin and calcium channel blockers do not appear to have consistent benefit in prospective trials in the prevention of postoperative superventricular arrhythmias.

Graft patency

Aspirin appears to improve saphenous vein graft patency significantly throughout the first postoperative year. Administration should be begun within 12 h of completion of the bypass.

Preoperative administration appears to have no benefit that is additive to early postoperative administration. Prospective controlled trials have demonstrated that administration begun at 1, 7, or 24 h after operation is effective, but the benefit of therapy is lost when aspirin is begun more than 48 h after operation.

Ticlopidine is effective but appears to offer little advantage over aspirin except as an alternative in the patient allergic to aspirin. Ticlopidine is associated with neutropenia as a rare side-effect. Dipyridamole appears to have no additive benefit to aspirin for saphenous vein graft patency. Warfarin also has no role in maintaining saphenous vein graft patency.

Pharmacologic lipid manipulation

The Cholesterol Lowering Atherosclerosis Study was conducted in 162 men after coronary artery bypass. Patients were treated prospectively with placebo or combined cholesterol and niacin therapy. There was angiographic demonstration of a reduction in progression of atherosclerosis in the treated group. More effective lipid-lowering agents, HMG-CoA reductase inhibitors (‘the statins'), are now available. Lovastatin has been demonstrated to be efficacious in decreasing the angiographic progression of atherosclerotic vein graft disease in the Post Coronary Artery Bypass Graft Trial.

Smoking cessation

The Coronary Artery Bypass Study dramatically demonstrated the negative sequelae of continued smoking after coronary artery bypass. Among patients randomized to operation, survival was 84 per cent at 10 years for those who were able to stop smoking compared with 68 per cent for persistent smokers. Cessation of smoking after coronary bypass appears to lead to a survival expectancy which is equivalent to post-bypass patients who never smoked. There is also angiographic evidence of a difference: 52 per cent of non-smokers' saphenous vein grafts are free of disease at 5 years compared with only 39 per cent of persistent smokers' saphenous vein grafts.

It is clear that intense efforts at cessation should be made in every smoker after coronary artery bypass grafting. Smoking is an addictive disorder and treatment

generally should include drug therapy. The nicotine transdermal patch is a very cost-effective therapy with an average estimated cost per year of life saved ranging from $1000 to 2400. Nicotine patch therapy in conjunction with a behavior modification program led to a cessation rate of 20 per cent compared with a 9 per cent cessation rate for behavior modification alone. Pharmacologic antidepression therapy has also been shown to be effective for smoking cessation. A prospective trial of bupropion (similar to selective serotonin release inhibitors) led to a 23 per cent cessation rate at 1 year, twice that of the placebo group.

Hormonal manipulation

A number of observational studies have demonstrated reduced coronary mortality in postmenopausal women on estrogen replacement therapy. One long-term observational study focused upon estrogen replacement therapy in postmenopausal women after coronary bypass and demonstrated a reduction in all-cause mortality. Benefit was present even if estrogen replacement therapy was begun after coronary bypass.

Rapid recovery after operation

Rapid early extubation, ambulation, and chest tube removal on the first postoperative day, aggressive patient education, and targeted discharge on the third day have become standard goals of postoperative care. This ‘fast track' recovery is generally accompanied by telephone calls by professionals to the patient several times in the first 2 weeks after discharge. It appears that the shortest postoperative stays in hospital are followed by the fewest readmissions to hospital if effective protocols are established.

Results of coronary artery bypass grafting

Early outcomes

The immediate question posed by the patient is ‘What is my chance of living through this operation?'. In order to counsel the patient and his/her family appropriately, one must be able to predict the hospital mortality and major complications of the procedure, in particular cerebrovascular accidents and major wound infections.

Hospital mortality

In the last two decades a number of cardiac surgical databases have collected sufficient information on coronary artery surgery to predict with great reliability the mortality for patients undergoing coronary bypass. Unfortunately, differences in definitions of variables, variation in the variables collected, and institutional differences in operative techniques have made comparisons difficult from one data set to another. Jones and colleagues reviewed seven large data sets representing over 172 000 patients operated upon between 1986 and 1994 in an attempt to determine the predictive power of preoperative variables. Seven core variables were found to be predictive of mortality in all data sets: age, gender, prior heart surgery, left ventricular ejection fraction, per cent stenosis of the left main coronary artery, number of coronary arteries with greater than 70 per cent stenosis, and priority of operation. The variables related to acuteness of operation, age, and prior operation had the greatest predictive value. In addition, other variables were found which had a small additive impact on predictive accuracy: angioplasty during the admission for operation, recent (less than 1 week) infarction, history of angina, ventricular arrhythmia, congestive failure or mitral regurgitation, comorbidities including diabetes, cerebrovascular disease, and peripheral vascular disease, chronic obstructive pulmonary disease, and creatinine level (Table 4).

Age per year	1.05
Prior operation	1.0
> 1 prior operation	1.5
Priority of operation	
Elective	1.0
Urgent	1.2
Emergency	3.0
Salvage	8.7
Female gender	1.5
Vessel disease	
Left main disease > 50%	1.5
One vessel disease	1.0
Two vessel disease	1.0
Three vessel disease	1.2
Ejection fraction	
> 40%	1.0
30–39%	1.63
20–29%	2.21
< 20%	4.06

*Data derived from the Society of Thoracic Surgeons Database (333061 patients, Edwards *et al.*, *Annals of Thoracic Surgery* 1997, 63: 903–8) with exception of ejection fraction data which was derived from the New York State database (57187 patients, Hannan *et al.*, *Annals of Thoracic Surgery* 1994, 58: 1852–7).

Table 4 Relative risk of hospital death*

In all data sets age has predicted hospital mortality following coronary bypass. If age under 65 years has a relative risk of 1.0, data from the Canadian Network Data Base indicated that relative risk was 2.07 for ages 65 to 74 and 3.84 for those older than 75. It is interesting that when particularly young patients (under age 50) are compared with patients over age 70, the older patients had a higher hospital mortality, but long-term survival in the older patients more closely parallels the general population.

Female gender also predicts increased mortality following coronary bypass with a relative risk of 1.5 to 2. Smaller body size, associated with smaller coronary artery diameter, appears to account for the bulk of this difference. Despite increased hospital risk, longer-term results are not different from males.

Urgency of operation is a matter of particular importance. The definition of urgent or emergency operations is somewhat problematic as some data sets have allowed the definition of a ‘salvage' operation, implying a ‘last ditch' urgency. In general, an emergency operation has approximately three times the risk of an elective one. In this regard one might consider attempts to convert emergency operations to elective situations as discussed earlier in this chapter.

It is important to note that the generalization of risk stratification models to a particular population is problematic. The model must be calibrated to the overall mortality rates of the population (the institution or region to which it is applied) and the model must be updated as progress in therapy occurs.

Postoperative central nervous system abnormalities

Postoperative stroke is a complication dreaded perhaps as much as death after coronary artery bypass. Postoperative neurologic deficits are generally divided into two types: type I, those deficits associated with focal neurologic deficits and type II, deficits associated with deterioration in intellectual function or memory. In a prospective observational study attempting to identify the incidence of type I and type II deficits, 2108 patients at 24 institutions were observed. Of these patients, 3.1 per cent had a type I deficit and 3.0 per cent had a type II deficit. The hospital mortality was 21 per cent for patients with type I deficit and 10 per cent for those with a type II deficit. These neurologic complications doubled the length of hospital stay and increased by a factor of six the chance of nursing home placement after the hospital stay. A multivariate analysis revealed advanced age and history of significant hypertension to be independent predictors of both type I and type II deficits. Predictors of type I deficits include: proximal ascending aortic atherosclerosis identified at the time of operation (relative risk 4.52), history of prior neurologic symptoms (relative risk 3.2), use of intra-aortic balloon pump (relative risk 2.6), diabetes (relative risk 2.6), history of hypertension (relative risk 2.3), history of unstable angina (relative risk 1.8), and age (relative risk 1.75 per decade). Factors predictive of type II deficits included: postoperative dysrhythmia (including atrial fibrillation), a history of alcohol consumption, hypertension, prior coronary artery bypass, peripheral vascular disease, and congestive heart failure. It is evident from these various risk factors for type I and type II deficits that macroembolization appears to be the primary reason for type I deficits while hypoperfusion appears to be involved in many of the type II deficits.

Postoperative mediastinal infection

Mediastinal infection is a second major complication of coronary artery bypass which is unfortunately relatively common (occurring in 1 to 4 per cent in cases, approximately the same as the overall incidence of death or of stroke). It is associated with a mortality of approximately 20 per cent. Obesity is a factor recurrently identified as contributing to mediastinal infection. The relative risk appears to be approximately 2 for patients who are obese. Patients with diabetes, particularly patients requiring insulin, have an increased risk of mediastinitis. In a prospective trial as indicated earlier, aggressive attempts to maintain normal blood glucose levels with continuous intravenous infusion of insulin appear to impact significantly on the development of mediastinitis. Reoperation is associated with an increased risk of mediastinitis related to longer operations, additional dissection and tissue damage, and a higher likelihood of transfusion. The use of bilateral internal mammary arteries and resultant devascularization of the sternum appears to contribute to mediastinitis when combined with other risk factors, specifically diabetes and obesity. In patients who are not obese or diabetic the use of bilateral internal mammary grafts appears to carry little additional risk.

Late results of coronary artery bypass

Coronary artery bypass has been remarkably effective in both relief of symptoms and in prolongation of life (particularly in certain anatomic subsets as discussed earlier in this chapter). The 1991 American College of Cardiology/American Heart Association Guidelines for Coronary Artery Bypass provided a general understanding of expectation after bypass surgery. In a heterogeneous group of patients, survival at 5 years was 92 per cent and at 10 years 81 per cent. Freedom from first return of angina was 83 per cent at 5 years and 63 per cent at 10 years. The various factors that predict long-term results after coronary artery bypass have been analyzed in several studies. In an analysis of 24 000 patients from 1977 to 1994 at Emory, age, ejection fraction, diabetes, congestive heart failure, hypertension, procedure status, number of vessels diseased, and use or non-use of the internal mammary artery were identified as significant multivariate predictors of survival. Angina class, history of myocardial infarction, renal dysfunction, hypertension, and incomplete revascularization were important factors identified by univariate analysis (Table 5).

	Univariate p value	*p value*	*Multivariate Hazard ratio*
Age (per year)	* 0.0001	* 0.0001	1.056
Ejection fraction	* 0.0001	* 0.0001	0.979
Diabetes	* 0.0001	* 0.0001	1.330
Congestive failure	* 0.0001	* 0.0001	1.394
Hypertension	* 0.0001	* 0.0001	1.352
Procedure status*	0.0016	0.0010	1.140
Vessels diseased**	* 0.0001	0.0178	1.068
Angina class***	* 0.0001	0.0175	1.038
Incomplete revascularization	* 0.0001	0.0189	1.118
No internal mammary artery	* 0.0001	0.0098	1.089

*Elective, urgent, or emergency operation.
**One vessel disease, two vessel disease, three vessel disease, or left main disease.
***Angina class 0, 1, 2, 3, or 4.

Table 5 Correlates of long-term mortality from the Emory Cardiac Data Bank

Cost-effectiveness of coronary artery bypass

The most reasonable system of analysis of cost-effectiveness of coronary bypass appears to be an estimation of dollars spent per quality-adjusted life-year (**QUALY**) gained. A widely quoted analysis using this technique was compiled by Weinstein and Stason in 1982 utilizing data gathered from the randomized trials comparing medical therapy with coronary artery bypass. The cost of coronary artery bypass was relatively constant whether it was conducted for left main disease or for single vessel disease. Cost-effectiveness is therefore excellent when the procedure is applied to patient subgroups whose benefit in terms of survival or relief of symptoms compared with medical therapy is great (for example a patient with severe angina and left main disease). The cost-effectiveness of coronary bypass becomes poor, however, when preoperative symptoms are minimal and benefit in terms of survival is small (a patient with mild symptoms, no left anterior descending stenosis, and double vessel disease). A review by Kupersmith in 1995 converted these data to 1993 dollars as depicted in Fig. 14 and Table 6. Because coronary artery surgery is particularly effective in relieving angina, the cost-effectiveness of this operation, even in patients with single vessel disease, is not prohibitive if that patient has severe angina. In general, a cost-effectiveness of $20 000 to 40 000 per QUALY is consistent with other medical programs funded by society such as hemodialysis and treatment of hypertension. A cost per QUALY of less than $20 000 would be considered very cost-effective while a cost greater than $60 000 per QUALY would be considered expensive. As shown in Table 6, the cost-effectiveness of coronary bypass compares favorably with the cost of other generally accepted medical therapies.

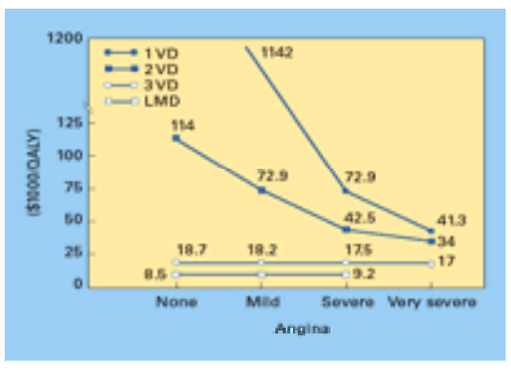


Fig. 14. A depiction of the cost-effectiveness of coronary artery bypass for one vessel disease (VD), 2VD, 3VD, and left main disease (LMD). Data are in 1993 dollars.

	US
CABG for left main stenosis, with/without angina	9000
CABG for three vessel disease, with/without angina	18000
CABG for two vessel disease, with severe angina, with left anterior descending stenosis	22000
CABG for two vessel disease, with severe angina, with no left anterior descending disease	61000
CABG for two vessel disease, with no angina, with left anterior descending stenosis	27000
CABG for two vessel disease, with no angina, with no left anterior descending disease	680000
CABG for one vessel disease, with severe angina	73000
PTCA for one vessel disease, with severe angina	9000
PTCA for left anterior descending stenosis, with mild angina	93000

*Adjusted to 1993 dollars from Kupersmith *et al.*, *Progress in Cardiovascular Diseases* 1995; 37: 307–46.

Table 6 Coronary artery bypass grafting (CABG) or percutaneous coronary angioplasty (PTCA) cost/quality-adjusted life-year compared with medical therapy*

Unfortunately much of the data concerning cost-effectiveness of coronary bypass is distinctly dated. The hospital cost of coronary artery bypass has been reduced by approximately two-thirds over the last decade. Long-term results are altered as arterial conduits are utilized, as aspirin is used in all patients after coronary bypass, as cholesterol-lowering manipulations are accomplished, and as angioplasty is utilized as adjunctive therapy after coronary bypass. Continuing reanalysis of cost-effectiveness data is a priority for thoracic surgeons.

Evolving technology

There are three areas of evolving technology in coronary revascularization which deserve mention in this chapter: gene therapy for prevention of vein graft atherosclerosis, transmyocardial laser revascularization, and less invasive coronary artery bypass.

Prevention of vein graft atherosclerosis

Endothelial damage occurs at the time of vein graft harvest and the endothelium is subsequently repopulated with endothelial cells. Signals are sent to the media of the vein graft which leads to the migration of fibroblasts to intimal levels with subsequent production of neointimal hyperplasia and later atherosclerosis. It appears that genetic confusion of this signal for transmigration of cells from the media to the intima can be accomplished by transfection of the vein graft with genetic material at the time of harvest. Since this has been accomplished in animal models, deterioration of vein grafts in mid-term studies has been greatly attenuated. This technique offers considerable promise for attenuation of vein graft atherosclerosis and perhaps extension of the duration of benefit of coronary bypass.

Transmyocardial revascularization

The reptilian heart provides nutrition to the myocardium with multiple communicating channels between the ventricular cavities and the myocardium. Indeed this network of communicating channels exists in the human fetal heart. Several attempts have been made to redevelop these connections from the ventricular cavity into the myocardium utilizing acupuncture and laser channels. The carbon dioxide laser has been utilized to create tiny channels from the epicardial to endocardial surface of the heart with subsequent sealing over of the epicardial end of the channels. Long-term channel patency has been reported in some animal experiments and in sporadic clinical case reports. Transmyocardial laser revascularization has been attempted in patients with severe angina refractory to medical therapy who are unsuitable for surgical revascularization, angioplasty, or heart transplantation. These patients generally have diffuse small-vessel disease and have generally undergone angioplasty and/or coronary bypass in the past.

A multicenter trial with transmyocardial laser revascularization as the sole therapy for 200 patients with refractory endstage coronary artery disease has recently been reported. The operative mortality rate was 9 per cent. The Canadian angina classification was significantly reduced from the preoperative status with follow-up extending to 1 year. Hospital admissions for angina decreased from 2.5 admissions per year in the year prior to treatment to 0.4 admissions in the year after treatment. Another multicenter, randomized, prospective study in 160 patients found that 72 per cent of patients with transmyocardial laser revascularization improved by at least two angina classes, while 69 per cent of the patients treated only with medical therapy had no change of angina class and 31 per cent became worse. Six months survival free of death, unstable angina, or class IV angina was 12 per cent in the medical management group and 73 per cent in the laser revascularization group. Despite these results which indicate subjective improvement in the patients, the mechanism of clinical benefit is very poorly understood. Animal experiments have suggested that some cardiac afferent nerve fibers are destroyed during laser revascularization raising concern that afferent denervation is the source of pain relief rather than relief of ischemia. It may be that the channels created by the laser lead to an inflammatory response and subsequent generation of neovascularity within the myocardium. Several investigators have suggested that endothelial growth factor might be additive to laser revascularization in benefit. Transmyocardial laser revascularization, despite the unknowns related to its mechanism, is a possible therapeutic option in a very difficult subgroup of patients.

Less invasive coronary artery bypass

Recently, attempts have been made to confer the benefit of coronary artery bypass on patients with coronary disease and, simultaneously, to diminish the morbidity of the operation either by performing the operation through a less invasive incision or by eliminating the use of cardiopulmonary bypass. Less invasive coronary bypass operations can be considered in three modalities. First, off bypass coronary bypass (**OBCAB**) is performed with a standard full median sternotomy (usually with a smaller skin incision). Mechanical stabilization is utilized to decrease the motion of the coronary arteries while distal anastomoses are performed. Proximal occlusion of the artery is usually accomplished with a vascular snare. Distal occlusion is not utilized for fear of causing a stenosis distal to the anastomoses. Second, minimally invasive direct coronary artery bypass (**MIDCAB**) is performed via a left anterior thoracotomy with or without removal of a small segment of rib or costal cartilage. This operation is performed on a beating heart with mechanical stabilization. Third, port access coronary artery bypass (**PA-CAB**) is performed with femoral cardiopulmonary bypass, cardioplegic arrest of the heart, and limited access incisions.

Potential benefits of OBCAB are the avoidance of cardiopulmonary bypass and limited retraction of the skin and sternum. This technique was first reported in 1967 by Kolessov in the Soviet Union. The technique was largely abandoned in the United States after effective cardiopulmonary bypass was developed. Large experiences with coronary bypass on a beating heart were accumulated in several other countries utilizing a variety of pharmacologic agents. Newly developed mechanical devices have greatly facilitated this technique. Various methods for elevation of the heart have allowed some access to the vessels on the lateral side and inferior surface of the heart.

The term MIDCAB should be reserved for performance of coronary artery bypass grafts without median sternotomy and without the use of coronary artery bypass. Generally, it is performed through a left anterior thoracotomy allowing access to the left anterior descending and diagonal vessels (and occasionally anterior marginal vessel). A similar technique may be utilized through a right anterior thoracotomy for bypass grafting of the right coronary artery. Generally, the internal mammary arteries are utilized. A recent modification of this technique involves a radial artery attached between the undissected internal mammary artery and the anterior vessels of the heart.

The use of OBCAB has led to a reduction in hospital stay from an average of 8 days in conventional coronary bypass to approximately 4 days. Patients generally return to work and social activities after 3 weeks rather than 6. Unfortunately, graft patency by angiography in an early series has been approximately 93 per cent (compared withan expected internal mammary graft patency of 98 per cent with conventional coronary artery bypass). Several surgeons have adopted a technique of intraoperative angiography to identify problems with off bypass anastomoses so that they can be corrected immediately, hoping to obtain angiographic patency rates equivalent to conventional procedures. As techniques are refined it is likely that off bypass techniques will prove to be quite acceptable and the sought after reduction in morbidity can be accomplished with preservation of the benefit of conventional procedures.

The third type of less invasive procedure is the closed chest, port access, video-assisted coronary artery bypass operation (PA-CAB). Aortic occlusion and cardioplegic arrest are central features of this technology. A triple lumen catheter with an aortic occlusion balloon at its distal end is used to accomplish endovascular aortic occlusion, cardioplegia delivery, and left ventricular decompression. With the heart arrested, coronary bypass can be performed on a decompressed heart through small ports with the aid of video cameras. This technique appears to allow for more complete revascularization and perhaps for more accurate anastomoses. Clinical trials of this technique are in progress.

The safety and efficacy of less invasive coronary bypass procedures remains to be determined. Although these techniques may prove to be inappropriate for a typical patient with multivessel disease, their utility in selected subgroups of patients is very likely to be demonstrable. For example, a patient requiring a single bypass to a large left anterior descending vessel appears to be an ideal candidate for less invasive procedures. Similarly, a patient requiring a bypass to a distal right coronary artery after two previous coronary artery bypass procedures is also likely to be easily approachable with off bypass techniques. The lateral surface of the heart is approachable via a left thoracotomy for reoperative patients requiring bypass grafts to the lateral wall. The combination of less invasive techniques with arterial conduits offers the possibility of a minimal morbidity with coronary bypass and arterial conduits to important vessels in the heart. A suggestion has been made that percutaneous techniques could then be used in a supplemental fashion to complete the revascularization procedure, a complimentary utilization of these two techniques which are so often seen as competitive.

Further reading

Chaitman BR *et al.* Myocardial infarction and cardiac mortality in the Bypass Angioplasty Revascularization Investigation (BARI) randomized trial. *Circulation* 1997; **96**: 2162–70. [This is an analysis of myocardial infarction and cardiac mortality in the largest prospective randomized trial of angioplasty versus coronary artery bypass.]

Charanjit SR, Gersh BJ, Yusuf S. Chronic coronary artery disease: coronary artery bypass surgery vs percutaneous transluminal coronary angioplasty vs medical therapy. In: *Textbook of evidence based cardiology*, pp. 368–92. BMJ Books, London, 1998. [An up-to-date discussion of randomized comparisons of medical therapy, coronary artery bypass and percutaneous techniques is presented.]

Comparison of coronary bypass surgery with angioplasty in patients with multivessel disease. The Bypass Angioplasty Revascularization Investigation (BARI) investigators. *New England Journal of Medicine* 1996; **335**: 217–25. [This is the original presentation of the largest randomized prospective trial of angioplasty versus coronary artery bypass surgery in multivessel disease.]

Edwards FH, Grover FL, Shroyer AL, Schwartz M, Bero J. The Society of Thoracic Surgeons National Cardiac Surgery Database: current risk assessment. *Annals of Thoracic Surgery* 1997; **63**: 903–8. [The Society of Thoracic Surgeons database presents, on an annual basis, a description of in-hospital morbidity and mortality for coronary artery bypass.]

King SB *et. al*. A randomized trial comparing coronary angioplasty with coronary bypass surgery. Emory Angioplasty versus Surgery Trial (EAST). *New England Journal of Medicine* 1994; **331**: 1044–50. [The EAST trial is a single institution comparison of angioplasty versus coronary artery bypass for patients with multivessel disease.]

Kupersmith J, Holmes-Rovner M, Hogan A, Rovner D, Gardiner J. Cost-effectiveness analysis in heart disease. Part III: Ischemia, congestive heart failure, and arrhythmias. *Progress in Cardiovascular Diseases* 1995; **37**: 307–46. [Cost effectiveness estimates are presented for coronary artery bypass and angioplasty.]

Pocock SJ *et. al*. Meta-analysis of randomised trials comparing coronary angioplasty with bypass surgery. *Lancet* 1995; **346**: 1184–9. [This meta-analysis includes several of the early randomized trials of angioplasty versus coronary artery bypass.]

Roach GW *et. al*. Adverse cerebral outcomes after coronary bypass surgery. Multicenter Study of Perioperative Ischemia Research Group and the Ischemia Research and Education Foundation Investigators. *New England Journal of Medicine* 1996; **335**: 1857–63. [This is an excellent prospective study of adverse cerebral outcomes after coronary bypass surgery.]

The Parisien Mediastinitis Study Group. Risk factors for deep sternal wound infection after sternotomy: a prospective, multicenter study. *Journal of Thoracic and Cardiovascular Surgery* 1996; **111**: 1200–7. [This prospective study identifies risk factors that can be used preoperatively to predict the risk of deep sternal wound infection.]

Thourani VH *et. al*. Influence of diabetes mellitus on early and late outcome after coronary artery bypass grafting. *Annals of Thoracic Surgery* 1999; **67**: 1045–52. [This recent manuscript presents the impact of diabetes and other risk factors on late outcome after coronary bypass grafting.]

Weintraub W *et. al.* Improving cost and outcomes of coronary surgery. *Circulation* 1998; (**Suppl.II**): II 23–8. [The dramatic improvement in cost and outcomes in coronary artery bypass surgery in the late 80s and early 90s is described.]

Yusuf S *et. al*. Effect of coronary artery bypass graft surgery on survival: overview of 10-year results from randomised trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration. *Lancet* 1994; **344**: 563–70. This is an outstanding reanalysis of the original data from the coronary artery bypass graft surgery trials.]

40.8.4 Surgery for complications of ischaemic heart disease

Matthew J. Wall, Jr and John C. Baldwin

Introduction
Acute coronary thrombosis
Ischemic mitral regurgitation
Postinfarction ventricular septal rupture
Left ventricular rupture, pseudoaneurysm, and aneurysm
Cardiac replacement for ischemic disease
Further reading

Introduction

Surgical procedures have been well developed for the treatment of ischemic heart disease. A small group of patients, particularly in the prethrombolytic era, developed significant complications of their ischemic heart disease which often required surgical therapy. While uncommon, these entities have a considerable morbidity and mortality and often require acute surgical therapy. These significant complications secondary to ischemic heart disease include acute coronary thrombosis, mitral regurgitation of ischemic etiology, postinfarction ventricular septal rupture, left ventricular rupture, ventricular aneurysms and pseudoaneurysms, and endstage dilated cardiomyopathy.

Acute coronary thrombosis

Appropriately timed revascularization of the myocardium following acute coronary occlusion significantly improves survival. Studies showing that alterations caused by brief periods of ischemia are completely reversible following reperfusion have resulted in vigorous searches for prompt and effective methods of restoring perfusion to the ischemic myocardium. The key to the success of this approach requires the presence of myocardium which can potentially be salvaged with these revascularization techniques.

The reperfusion phenomenon, despite intensive study, is not completely understood. Potentially deleterious effects from restoration of flow to ischemic tissue have been observed both clinically and experimentally. It is likely that at least some of this effect is mediated by the generation of superoxide radicals, which cause cellular edema, disrupt the integrity of cellular membranes, and increase vascular permeability. The net result appears to be myocardial injury and interstitial hemorrhage, manifested by both systolic and diastolic cardiac dysfunction. These effects can potentially be mitigated by so-called scavenger agents which can be introduced in the reperfusate solution. The optimal composition of the reperfusion milieu, however, has not yet been determined.

As the survival of acutely ischemic myocardial cells is time dependent, significant effort has been invested in developing techniques to restore perfusion rapidly. It is commonly held that reperfusion must be established within 4 to 6 h following ischemia to avoid irreversible myocardial injury. The aggressive and almost routine use of thrombolytic therapy for acute myocardia ischemia has had a significant effect on the preservation of ventricular function. Thus, the need for the high-risk emergency coronary artery bypass has been reduced to nil. In the acute setting, coronary artery bypass has been relegated to the protection of an area of myocardium at risk following successful reperfusion. Thus, thrombolytic therapy preserves myocardium and allows for the diagnosis of myocardium that will benefit from coronary artery bypass. With the preservation of ventricular mass and prevention of ischemic damage, surgery for complications of ischemic myocardium has decreased in frequency. While infrequent, unsuccessful thrombolysis, refractory symptoms consistent with ischemia, and cardiogenic shock are occasional indications for surgery. However, it is preferable to stabilize the patient's hemodynamic status by aggressive pharmacologic means to allow surgery to be performed on an urgent rather than emergency basis.

The success of thrombolytic therapy in averting the loss of myocardium following coronary thrombosis has improved the survival of patients with coronary artery disease. Pharmacologic thrombolysis affects the conversion of the glycoprotein, plasminogen, into plasmin. Agents such as streptokinase and the more potent, tissue plasminogen activator are effective in lysing established thrombus in nearly 75 per cent of patients. Indications for immediate surgical revascularization include failure to obtain adequate thrombolysis, or ongoing ischemia despite dissolution of the offending thrombus. The latter situation is usually caused by increased myocardial demands in the presence of a pre-existing, fixed stenosis. Patients with residual lesions which supply a significant area at risk for infarction may also require urgent surgery if the lesion is not suitable for angioplasty. In any event, thrombolytic therapy maintains the patency of the distal vessel allowing visualization on coronary angiography and thus facilitates identification of bypassable regions.

When it is clear that surgery is required, the thrombolytic therapy is discontinued. Despite the short half-lives of these agents, bleeding can be a significant issue during surgery and the blood bank should be alerted to have appropriate clotting factors available. The priority is to have effective prompt revascularization of the vulnerable area.

A common indication for emergency coronary bypass is failure of percutaneous transluminal coronary angioplasty with occlusion or dissection. In many groups a full surgical team and operating theater are maintained available during the performance of coronary angioplasty. However, many centers no longer maintain a dedicated operating theater, treating these cases as for other emergencies from the hospital. The indications for surgical intervention depend on the condition of the patient and not on the anatomic and geographic pattern. The clinical deterioration of the patient with persistent angina, life-threatening dysrhythmia, or hemodynamic instability despite aggressive pharmacotherapy continue to be indications for emergency intervention. The overall mortality after surgery following a failed angioplasty is in the range of 5 to 6 per cent. Evidence of myocardial infarction is demonstrated in up to 50 per cent of patients postoperatively in this scenario.

The goal of surgical revascularization, whether undertaken after failure of thrombolysis, rethrombosis, or complications of coronary angioplasty, is to identify major target vessels and bypass offending lesions with saphenous vein grafts. In more stable patients, in whom reperfusion has been successful and surgery is performed to protect an area of myocardium at risk, arterial conduits whenever feasible are desirable. Many patients may benefit from insertion of an intra-aortic balloon for counterpulsation: this may assist in achieving a survivable physiology and allow for a less urgent operation. Some patients gradually improve with counterpulsation alone over a period of days. Those who do not benefit from counterpulsation usually have a poor outcome regardless of the method of revascularization. Of patients who present in cardiogenic shock following acute coronary thrombosis, the 2-year survival does not exceed 40 per cent, regardless of whether surgical or non-surgical therapy is used. Factors which can affect survival following emergency coronary bypass are the period of time that has elapsed from onset of ischemia to reperfusion, the patient's pattern of coronary atherosclerosis, the prompt institution of intra-aortic counterpulsation, the patient's age, and the pre-existing level of ventricular function.

Ischemic mitral regurgitation

Mitral insufficiency secondary to coronary ischemia is a common entity occurring in approximately 20 per cent of patients with ischemic heart disease. However, less than 5 per cent of patients with mitral regurgitation after myocardial infarction have valvular regurgitation that is hemodynamically significant. The management and prognosis of the patient depends on the stage of their disease and the physiologic/anatomic derangement of the mitral valve as well as the degree of compensation.

Acute mitral regurgitation can occur from severe ischemia/infarction resulting in lack of function of the papillary muscle mechanism. In addition, dyskinetic segments of the ventricular wall may result in distortion of the mitral valve and the subvalvular apparatus. This prevents coaptation of the mitral valve leaflets, which is a key element for competency. This can be exacerbated in the ischemic ventricle that requires elevated left ventricular end diastolic pressures. For patients with chronic mitral regurgitation, the onset of symptoms and the occurrence of a new murmur can often be correlated with the onset of an acute myocardial infarction in the past. In addition, the presentation can be confused as some patients have episodic insufficiency as a result of intermittent myocardial ischemia. In the chronic phase, mitral regurgitation may be secondary to mitral annular dilatation, ventricular dilatation with displacement of the papillary muscles/subvalvular apparatus toward the apex of the left ventricle, or dysfunction/discoordination of the subvalvular apparatus. All these factors contribute to the key element, which is coaptation of the mitral leaflets in a uniform plane. Thus, surgical repair is based on the anatomic/physiologic derangement of the leaflets, annulus, and subvalvular apparatus.

The primary issue in these patients is achieving successful revascularization. A significant number of patients show substantial improve-ment following revascularization alone. Surgery aimed specifically toward the mitral valve should be reserved for those patients who have moderate to severe mitral regurgitation that results in hemodynamic compromise. Clinical judgement is essential in determining whether revascularization is likely to result in appreciable restoration of mitral valvular function or whether direct surgical repair/replacement of the mitral valve is required.

One of the most dramatic complications of ischemic heart disease leading to acute mitral regurgitation is the rupture of a papillary muscle. This occurs in approximately 2 per cent of patients following myocardial infarction, usually 1 week after the acute infarction. The posteromedial papillary muscle, which is supplied by the posterior

descending coronary artery, is commonly involved. The anterolateral papillary muscle is less affected as it has a dual blood supply from branches of the anterior descending and the circumflex coronary arteries. Symptoms arise almost immediately following rupture, although some patients may not manifest symptoms for 2 weeks. Patients usually develop pulmonary edema that can progress to cardiogenic shock. Aldhough echocardiography is usually sufficient to establish the diagnosis, papillary muscle rupture must occasionally be differentiated from ventricular septal rupture. Septal rupture can typically be excluded by the absence of a step-up in oxygen saturation in the right ventricle during placement of a pulmonary artery catheter. Moreover, pulmonary capillary wedge pressure tracings usually manifest striking V waves in the patient with papillary rupture. Coronary angiography will identify vessels for bypass grafting preoperatively but may not be tolerated in patients suffering profound hemodynamic deterioration.

The treatment of acute ischemic mitral regurgitation requires replacement of the valve, unless an isolated anatomic lesion involving the posterior leaflet can easily be repaired. Some clinicians have argued that repair with simultaneous revascularization may produce the lowest mortality rate in these patients. It must be remembered that the feasibility of repair is decreased by the presence of ischemic or necrotic tissue at the base of the papillary muscle or in the myocardium when reimplantation of chordae is contemplated (Fig. 1). This friable tissue holds sutures very poorly, making repair a tenuous procedure. Ischemic regurgitation associated with annular dilatation may be adequately treated by insertion of a mitral annular ring. However, it is not clear that repair is superior to replacement in combination with myocardial revascularization. The inherent uncertainty of restoration of function in valve repair and the possibility of subsequent replacement tend to support mitral valve replacement as a more prudent primary option in these very ill patients. Mitral valve repair is probably best reserved for the stable patient with established, chronic mitral valve dysfunction.

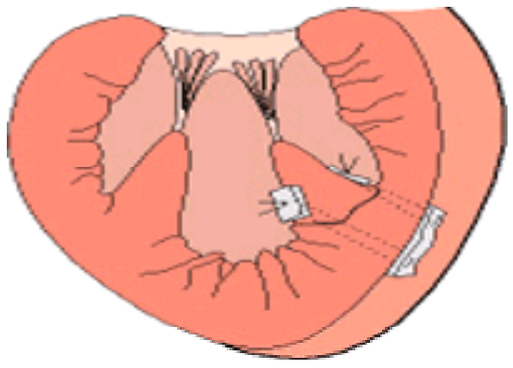


Fig. 1. Reimplantation of a ruptured papillary muscle. Teflon pledgets on both sides of the ventricular cavity are highly desirable.

With the current availability of a wide range of reliable and easily implanted mitral prostheses the choice of prosthesis is relatively unimportant and often regional. Prostheses with struts that extend into the left ventricle have to be carefully placed to avoid strut perforation of the posterior ventricular wall. This can be an issue particularly in cases with infarction in that region. In these situations a low-profile device may be warranted. There has been significant discussion related to the preservation of the posterior leaflet and now anterior leaflet of the mitral valve during implantation of the prosthesis. Preservation of the leaflets and the subvalvular apparatus is felt to preserve the complex twisting motion of left ventricular contraction. While the benefits to contraction may be controversial, preservation of the posterior leaflet helps to avoid the dreaded complication of atrial ventricular groove disruption. This complication has an extremely high mortality, particularly in patients with ischemic/infarcted muscle that will not adequately hold sutures. It is also important to avoid implanting an overly large valve in the relaxed heart from cardioplegic arrest during cardiopulmonary bypass, which can result in atrioventricular groove disruption. Thus many surgeons will resect the anterior leaflet but preserve the posterior one.

The regurgitant mitral valve has provided a pop-off valve for the left ventricle resulting in reduced afterload. The replacement or repair of the itral valve results in acute mitral competence with a rise in afterload. Systemic administration of vasodilators to decrease afterload for the first 3 to 5 days after surgery is commonly practiced. In addition, careful appropriate diuresis as well as prophylaxis for rapid atrioventricular conduction and ventricular response should the patient develop acute atrial fibrillation are often instituted.

The mortality rate in patients undergoing combined revascularization and mitral valve replacement in the acute setting has remained in the range of 30 to 40 per cent. As expected, the mortality rate is higher when mitral regurgitation is associated with cardiogenic shock. In patients with long-standing ischemic mitral regurgitation, prognosis is usually associated with the degree of ventricular dysfunction, as well as the presence of accompanying anatomic complications such as left ventricular aneurysm. As in the case of coronary surgery, maximizing medical therapy prior to surgery is important: administration of afterload reducing agents is particularly crucial in patients with acute mitral regurgitation. It is often impressive how aggressive medical therapy can improve pulmonary edema and thus bring the patient to a more favorable preoperative status. Sodium nitroprusside reduces peripheral as well as pulmonary vascular resistance through its direct actions on vascular smooth muscle. Afterload reduction can be further enhanced by preoperative insertion of an intra-aortic balloon pump, especially in patients with profound clinical deterioration such as those with frank papillary muscle rupture. There is no substitute, however, for appropriately timed surgical intervention, despite the operative mortality rate of approximately 50 per cent. Starr's group has reported a 5-year survival rate of 43 per cent in patients undergoing coronary artery bypass grafting and mitral valve replacement for ischemic mitral regurgitation. Survival is thus dramatically lower than that in patients with mitral regurgitation resulting from other causes.

Postinfarction ventricular septal rupture

Rupture of the ventricular septum is a catastrophic complication of myocardial infarction. This occurs in 1 to 2 per cent of patients usually within 2 weeks of the myocardial infarction. These patients often have much more localized coronary artery disease, so that they develop necrosis of the ventricular septum as the result of the relative lack of collateral vessels. The most common location for postinfarction ventricular septal rupture is the muscular anteroapical septum in the left anterior descending coronary artery distribution. This is followed by the inferior septum in the posterior descending artery territory. The anterior septal rupture is occasionally associated with ischemic mitral regurgitation as well as left ventricular aneurysm in 30 to 60 per cent of patients.

Symptoms are often related to poor peripheral perfusion and to right ventricular failure, in contradistinction to ischemic mitral regurgitation where pulmonary edema predominates. A harsh murmur can be heard in most patients. The diagnosis is established by placement of a pulmonary artery catheter and demonstration of an oxygen saturation step-up from central venous to pulmonary arterial blood. Left ventriculography and coronary angiography are helpful in localizing the lesion and the coronary anatomy.

Patients with ventricular septal rupture should undergo surgery shortly after the diagnosis is made. Prompt placement of an intra-aortic balloon may support the patient to allow diagnostic studies such as cardiac catheterization to be performed. As in patients with acute mitral regurgitation, afterload reduction is the fundamental strategy. However, profound deficits in peripheral perfusion often limit the use of vasodilatating agents. The optimal combination of afterload reduction and cardiotonic agents varies among patients and, in a given patient, changes as the severity of pump failure evolves. Delay in surgery often results only in selection of patients more likely to survive.

The surgical approach to this condition can be technically challenging. The technical principles involve ventriculotomy through the infarct and debridement of the infarcted tissue involving the left ventricular free wall as well as the septum. While uncommon, the entire septum may be infarcted requiring resection and reconstruction (Fig. 2). The ventriculotomy is made in the area of infarction without reference to the position of the coronary arteries. Following debridement the septal defect is closed with a Dacron prosthetic patch avoiding tension. With the friability of even the normal-appearing myocardium, pledgeted sutures should be used and placed from the right ventricular side (Fig. 3). Many surgeons also find it useful to place pledgets even on the prosthetic side of the suture line to cushion the patch as the sutures are tied. There is often good visualization of the mitral valve and subvalvular apparatus which if needed can be addressed through the ventriculotomy. For the anterior apical defects the free edge of the left ventricle can be approximated to the right ventricle where the free wall of the anterior patch has been anchored (Fig. 4). This is done using interrupted mattress sutures of 2-O braided polyester suture along with parallel strips of Teflon felt. After closure is completed it can be reinforced with a continuous polypropylene suture as a hemostatic stitch. During repair of these defects it can be important to maintain ventricular volume and geometry. For the posterior septal defects it is often difficult to reapproximate the ventriculotomy without significantly altering the configuration of either ventricular cavity. Thus, a small Dacron patch is usually required to reconstruct the free wall of the ventricle (Fig. 5). This patch is attached similarly with interrupted non-absorbable pledgeted mattress sutures. This is reinforced with a continuous fine reinforcing suture for hemostasis. Thus, the repair of posterior defects is often considered to be more technically demanding.

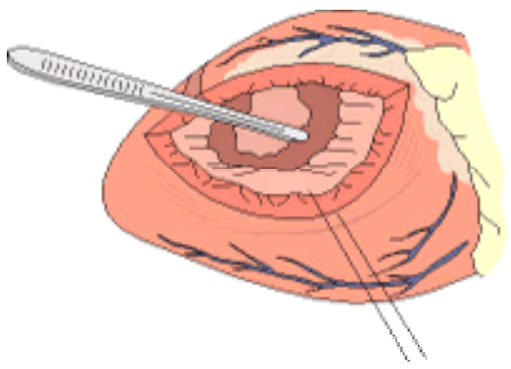


Fig. 2. Debridement of necrotic myocardium following ventricular septal rupture. Adequate exposure via incision through infarcted area is essential.

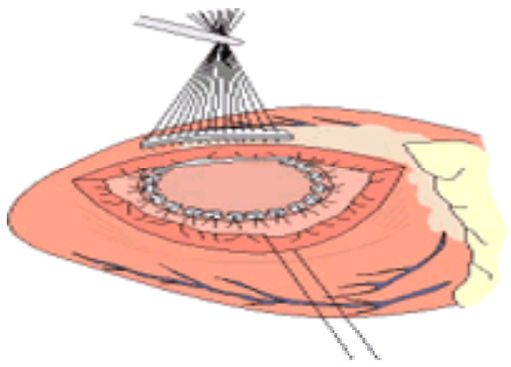


Fig. 3. Patch closure of septal rupture.

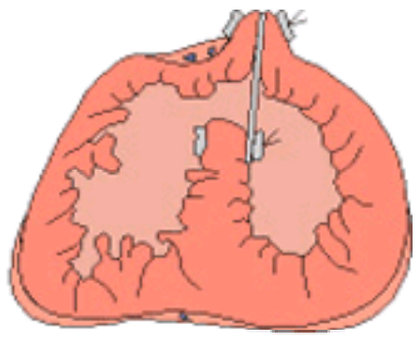


Fig. 4. Cross-sectional diagram of septal repair with primary approximation of the ventricular wall. The septal patch is incorporated within the ventricular wall repair.

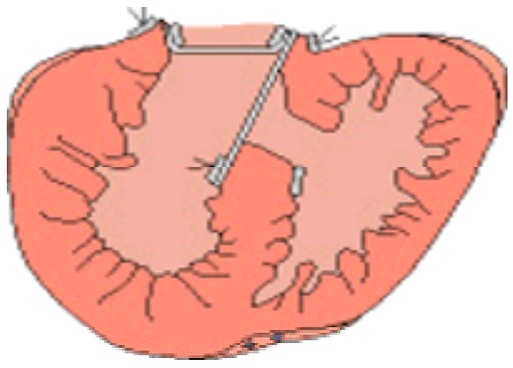


Fig. 5. Repair of posterior defect with additional patch for repair of posterior ventricular free wall.

Although its importance is not as well established as in the case of ischemic mitral regurgitation, coronary revascularization can be an important addition to the procedure. Distal anastomosis to appropriate targets can be constructed following debridement of the ischemic septum or following reconstruction of the ventricular cavity. Careful attention to myocardial preservation permits adequate time for these complex repairs. Combinations of retrograde and antegrade cardioplegia, cardioplegia via the grafts, as well as intracavity topical cooling can be used.

The mortality rate associated with this disease is high, particularly in elderly patients, exceeding 50 per cent following the first week without surgical intervention. The surgical mortality rate approaches 25 per cent and depends on the patient's hemodynamic status preoperatively. Long-term survival depends upon the extent of coronary disease as well as the residual left ventricular function.

Left ventricular rupture, pseudoaneurysm, and aneurysm

Left ventricular rupture is an often fatal complication of transmural myocardial infarction. It typically occurs 1 week after the myocardial infarction. The patient develops sudden hypotension ordinarily as the result of cardiac tamponade. Occasionally, in a few fortunate patients the rupture may be contained, allowing time for appropriate resuscitative measures and emergency surgery. As in the patient with ventricular septal rupture the free wall rupture of the left ventricle results in a defect with boundaries that consist of necrotic, friable muscle. The surgical approach involves excision of the infarcted area and primary reapproximation of the viable edges of the ventricular myocardium using interrupted non-absorbable mattress sutureswith Teflon buttresses. A second layer of continuous monofilament suture can afford additional hemostasis. A pseudoaneurysm is often described when a left ventricular rupture is contained and scar develops. Its bridging defect is often called a subepicardial aneurysm, but is functionally indistinguishable from a left ventricular aneurysm. Left ventricular aneurysm complicates 10 to 35 per cent of patients with transmural myocardial infarctions and generally develops over 2 to 8 weeks postinfarction. In 85 per cent of the patients the presenting symptoms are a combination of angina and congestive heart failure depending upon the relative proportions of scar and ischemic myocardium. The presence of significant angina suggests ische-mic myocardium which may benefit from aneurysm resection and revascularization. Ventricular dysrhythmias and embolic phenomena are less common as initial presentations of ventricular aneurysm. The presence of these symptoms in association with a significant left ventricular aneurysm is often an indication for surgery.

Although the distinction between a discrete aneurysm and an area of markedly diminished contractility is often difficult, aneurysms are generally characterized by ‘paradoxical' expansion of the segment in response to the increased intracavitary pressures generated during systole. During cardiopulmonary bypass, after placement of the ventricular sump, the aneurysm is noted to be a soft depressed area of the ventricle. The aneurysmal wall is composed of fibrous tissue which has replaced necrotic myocardium. The resulting structural weakness results in an inability to resist the high wall stress during contraction. With time, an increase in the size of the aneurysm leads to elevated surface tension, resulting in further stretching of the wall. These lesions may attain considerable size causing further reduction in cardiac output. Congestive heart failure ensues as a result of physiologic alterations during both systole and diastole. The stasis of blood in the aneurysmal sac in combination

with an abnormal and thrombogenic endocardium increases the propensity for thrombus formation and embolization.

The surgical approach to the patient with ventricular aneurysm consists of resection of the aneurysm and reconstruction of the ventricle with restoration of near-normal left ventricular configuration (Fig. 6). It is important to avoid embolization of laminated thrombus within the aneurysm. Thus, prior to significant manipulation of the heart the aortic cross-clamp is placed to prevent embolization. Aneurysmectomy is accomplished being careful to preserve the papillary muscles and the subvalvular mitral apparatus. One technique involves resection of all but 1 cm of scar for suture placement. This rim is then reapproximated with braided pledgeted horizontal mattress sutures through bolstering strips of Teflon felt. This can be followed by continuous monofilament suture for hemostasis. As in ventricular septal rupture, resection of the aneursym opens the ventricle and allows for intracavitary cooling. Many surgeons will place the coronary grafts early in the operation so that cardioplegia can be directly instilled during the aneurysm repair. Increased use of retrograde cardioplegia allows the instillation of cardioplegia throughout the procedure. Another approach to aneurysmectomy has been popularized by Cooley and Jatene. Attempting a primary closure of ventricular aneurysms often results in foreshortening and alteration of the ventricular geometry. After opening the aneurysm the junction of normal and scarred ventricle is noted. A low porosity Dacron patch is sewed to this junction in such a fashion to restore normal ventricular geometry. The excess ventricular scar is then carefully closed over this Dacron patch for reinforcement and hemostasis.

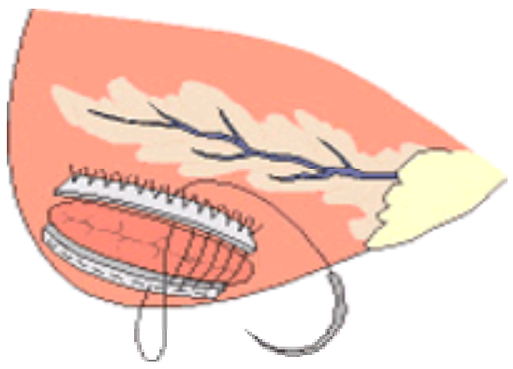


Fig. 6. Resection of left ventricular aneurysm.

Overall, the mortality rate associated with left ventricular aneurysm resection and coronary bypass grafting remains at 10 to 15 per cent. In the Coronary Artery Surgery Study, there was a 9 per cent operative mortality for revascularization and aneurysm resection. In general, hospital mortality has been correlated with the degree of preoperative congestive heart failure and the extent of coronary artery disease. Moreover, the long-term outlook has been somewhat discouraging in all centers, with Barrett-Boyes' group reporting a 30 per cent incidence of late death in an average follow-up of 39 months. The patients that do survive, however, are consistently improved in terms of symptoms. Improvement is more likely in patients with angina than in those with congestive heart failure. The Oslo group has shown that significant right ventricular dysfunction is common in patients with left ventricular aneurysm (89 per cent of patients). They have also shown that repair of the aneurysm does not generally result in early improvement in right ventricular function.

Cardiac replacement for ischemic disease

Ultimately, when congestive heart failure results from coronary artery disease, consideration must be given to replacement of the heart. At present, so-called ‘ischemic cardiomyopathy' is the most common indication for cardiac transplantation. The decision for transplantation rather than revascularization is multifactorial: consideration is given to the relative prominence of failure symptoms as opposed to angina, availability of target vessels, and candidacy for mechanical ‘bridging' to transplantation should revascularization be unsuccessful. In the current era of revascularization with advanced methods of myocardial preservation, it is possible to perform coronary bypass grafting in many patients who would formerly have undergone transplantation. In one group of patients who were candidates for transplantation but who instead underwent reparative procedures, survival was comparable with that of a cohort of patients who received transplants. In general, patients with angina pectoris are more likely to benefit from revascularization representing the presence of ischemic tissue rather than scar. The decision for cardiac transplantation is based upon the presence of markedly reduced ventricular function and the lack of feasibility of conventional surgery. Occasionally, cardiac transplantation is undertaken in patients with angina or dysrhythmias, but congestive heart failure is far more common as an indication.

The technique of cardiac transplantation is described elsewhere in this volume. The 1-year survival rate of patients receiving a cardiac transplant for ischemic heart disease is in the range of 80 to 85 per cent; 5-year survival approximates 65 per cent. Attention is being increasingly focused on the potential use of paced vascularized muscle flaps for augmentation of ventricular function and for diastolic counterpulsation in patients wtih long-standing coronary artery disease. This experience so far is small, and while physiologic effects can be shown, significant clinical improvement remains an elusive goal.

Considerable clinical attention and research effort has been focused on the use of mechanical devices for replacement of the heart. While a number of devices have been attempted, so far none has proved to be a reliable alternative to transplantation. There continue to be significant problems of embolization with the interaction of blood with the artificial surfaces. Currently implantable devices such as the Novacor Left Ventricular Assist Device are being used as a ‘bridge' for patients awaiting transplantation. In addition, increasing numbers of devices are being developed for patients who cannot be acutely weaned from cardiopulmonary bypass following conventional cardiac surgery. As technical difficulties with materials and embolus are solved, mechanical devices may gain greater use as they move through the requisite clinical trials. Concomitantly, results with transplantation are improving. It is interesting to note that a division of opinion in parallel to the discussions relating to mechanical as opposed to tissue valves is occurring. As with mechanical valves, mechanical hearts may have the advantage of durability and the disadvantage of requirement for coagulation, while biological hearts will have fewer implant-related complications but will have the potential for structural failure due to the effects of chronic rejection and have a lack of availability. It is hoped these issues will be elucidated with the continued dedicated work being pursued by many centers.

Further reading

Akins CW. Early surgical revascularization following thrombolytic therapy or PTCA failure. In: Baue AE, Geha AS, Hammond GL, Laks H, Naunheim KS, eds, *Glenn's thoracic and cardiovascular surgery*, pp. 1763–70. Appleton & Lange, Norwalk, Connecticut, 1991. [A summary of issues related to surgical revascularization after attempts at thrombolytic therapy or angioplasty.]

Aziz S, Baldwin JC, Billingham M, Jamieson SW, Shumway NE. Deleterious hemorrhagic effects of streptokinase on the ischemic myocardium. *Surgical Forum* 1986; **37**: 248–51. [A description of bleeding complications associated with the use of thrombolytics in the patients with ischemic heart disease.]

Barratt-Boyes BG, White HD, Agnew TM, Pemberton JR, Wild CJ. The results of surgical treatment of left ventricular aneurysms. *Journal of Thoracic and Cardiovascular Surgery* 1984; **87**: 87–9. [A report of a significant series of surgical repair for left ventricular aneurysms.]

Chatterjee K. Complications of acute myocardial infarction. *Current Problems in Cardiology* 1993; **5**: 79. [Review article describing the complication profiles of patients suffering acute myocardial infarction.]

Daggett WM *et al.* Improved results of surgical management of postinfarction ventricular septal rupture. *Annals of Surgery* 1982; **196**: 269. [Technical aspects and results of surgical management of post infarction ventricular septal rupture are presented.]

Epstein JI, Hutchins GM. Subepicardial aneurysms: a rare complication of myocardial infarction. *American Journal of Medicine* 1983; **75**: 639–43. [An article describing the rare complication of subepicardial aneurysm after acute myocardial infarction.]

Faxon DP *et al.* The influence of surgery on the natural history of angiographically documented left ventricular aneurysm: the coronary artery surgery study. *Circulation* 1986; **74**: 110–18. [A report from the coronary artery surgery study describing surgery and the natural history of left ventricular aneurysm.]

Hickey MS *et al.* Current prognosis of ischemic mitral regurgitation. *Circulation* 1988; **78**: 1–52. [An article describing the outcomes and prognosis of patients with ischemic mitral regurgitation based on therapy.]

Jatene AD. Left ventricular aneurysmectomy. Resection or reconstruction. *Journal of Thoracic and Cardiovascular Surgery* 1985; **89**: 321–31. [A classic article describing the approach for reconstruction of left ventricular volume after aneurysmectomy.]

Kay PH, Nunley DL, Grunkemeier GL, Pinson CW, Starr A. Late results of combined mitral valve replacement and coronary bypass surgery. *Journal of the American College of Cardiology* 1985; **5**: 29–33. [Outcomes related to patients after combined coronary artery bypass grafting with mitral valve replacement are presented in this report.]

Kriett JM, Kaye MP. The registry of the International Society for Heart and Lung Transplantation: eighth official report 1991. *Journal of Heart and Lung Transplantation* 1991; **10**: 491–8. [A special report of the international society for heart and lung transplantation.]

Lee KF, Wechsler AS. Cardiogenic shock secondary to myocardial infarction. In: Baue AE, Geha AS, Hammond GL, Laks H, Naunheim KS, eds. *Glenn's thoracic and cardiovascular surgery*, 6th edn, Vol. II, pp. 2103–13. Appleton & Lange, Stamford, Connecticut, 1996. [This chapter summarizes surgical issues related to cardiac shock after acute myocardial infarction.]

Lee L *et al.* Multi-center registry of angioplasty therapy of cardiogenic shock: initial and long-term survival. *Journal of the American College of Cardiology* 1991; **17**: 599–603. [Acute and long-term results after angioplasty therapy for patients in cardiogenic shock is presented from a multicenter study.]

Lower RN, Shumway NE. Studies on the orthotopic homotransplantation of the canine heart. *Surgical Forum* 1960; **11**: 18. [This article describes lessons learned from orthotopic homotransplantation of the canine heart.]

Madsen JC, Daggett, WM. Postinfarction ventricular septal rupture. In: Baue AE, Geha AS, Hammond GL, Laks H, Naunheim KS, eds. *Glenn's thoracic and cardiovascular surgery*, 6th edn, Vol. II, pp. 2115–29. Appleton & Lange, Stamford, Connecticut, 1996. [This chapter presents a summary of the spectrum of disease and treatment related to postinfarction ventricular septal rupture.]

Mangschau A, Geiran O, Rootwelt K, Laake B, Forfant K. Impairment in right ventricular performance after left ventricular aneurysm surgery. *Annals of Thoracic Surgery* 1988; **46**: 542–8. [An article that describes the effect of left ventricular aneurysm surgery on right ventricular geometry and performance.]

Oaks TE, Pae WE Jr, Miller CA, Pierce WS. Combined registry for the clinical use of mechanical ventricular assist pumps and the total artificial heart in conjunction with heart transplantation: fifth official report 1990. *Journal of Heart and Lung Transplantation* 1991; **10**: 621–5. [This is a report of the experience of clinical use of ventricular assist in conjunction with cardiac transplantation.]

O'Keefe JH, Bailey WL, Rutherford BD, Hartzler GO. Primary angioplasty for acute myocardial infarction in 1000 consecutive patients: results in an unselected population and high-risk subgroups. *American Journal of Cardiology* 1993; **72**: 107G–115G. [This describes a large center's experience of primary angioplasty for acute myocardial infarction.]

Pifarre R *et al.* Management of left ventricular rupture complicating myocardial infarction. *Journal of Thoracic and Cardiovascular Surgery* 1983; **86**: 441–3. [This describes a large center's experience of primary angioplasty for acute myocardial infarction]

Sanches JA, Smith CR, Spotnitz HM, Malm JR, Rose EA. High risk reparative surgery: a neglected alternative to heart transplantation. *Circulation* 1990; **82**: IV-302–5. [This article suggests the possibility of high risk cardiac surgery as opposed to direct cardiac transplant.]

Tebbe U *et al.* Will emergency coronary bypass grafting after failed elective percutaneous transluminal coronary angioplasty prevent myocardial infarction? *Thoracic and Cardiovascular Surgery* 1989; **37**: 308–12. [A review of acute coronary artery bypass grafting after failed percutaneous transluminal coronary angioplasty is presented.]

Weintraub RM, Wei JY, Thurer RL. Surgical repair of remediable postinfarction cardiogenic shock in the elderly. *Journal of the American Geriatrics Society* 1986; **34**: 389–92. [Results related to surgical management of postinfarction cardiogenic shock in the geriatric population is described.]

40.9.1 Pathology of cardiac valves

M. J. Davies

Inflammatory change
Immunological damage
Bacterial endocarditis
Degenerative change
Aortic valvular abnormalities
Normal anatomy
Anatomical variation
Mitral valve
Mitral stenosis
Mitral regurgitation
Tricuspid and pulmonary valvular disease
Further reading

Normal valve cusps are avascular and have a dense, collagenous core (fibrosa) covered on both sides by a layer of loosely arranged connective tissue (spongiosa) rich in mucins. The rate of turnover of this connective tissue is high. Dense collagen forms within the spongiosa following any abnormal impact of the cusps with each other or with the endocardium. Any derangement of cuspal architecture therefore invokes non-specific secondary fibrosis, further affecting function.

Inflammatory change

Immunological damage

Rheumatic fever causes an acute valvulitis; at a simple level this can be regarded as an example of immunological damage to tissues that incidentally share antigens with components of the streptococcal cell wall. Although antibodies to streptococcal antigens that cross-react with cardiac connective tissue are produced, there is no evidence that immune complexes are deposited, and cellular immune responses are probably involved. In the acute phase of rheumatic fever, the cusps are infiltrated with inflammatory cells (predominantly lymphocytes), and small platelet vegetations form along the apposition lines. Death in the acute stage relates to myocarditis rather than to valvulitis. Acute rheumatic fever may be followed by chronic valvular disease after an interval of some years.

Chronic rheumatic valvular disease is associated with vascularization and effacement of the cuspal architecture by dense fibrosis. Fusion of the commissures, and fibrosis, contraction and calcification of the cusps are the macroscopic hallmarks of chronic rheumatic disease and produce all combinations of stenosis and regurgitation. The mechanism underlying continuing cuspal fibrosis over years may be recurrent subclinical attacks of acute rheumatic fever, or fibrosis developing secondarily to valvular derangement.

Valves that are morphologically identical to those of chronic rheumatic disease can be found in patients who have no history of an acute rheumatic attack. It is uncertain whether these individuals have suffered subclinical streptococcal infections, or whether there are other unrelated causes of valvulitis. Both systemic lupus erythematosus and rheumatoid disease can be associated with a valvulitis that, in the endstage, is very similar to that of chronic rheumatic disease. A number of drugs, including methysergide, ergotamine and the anorectic agents fenfluramine and phentermine, are described as producing diffuse fibrosis in the spongiosa zone without vascularization or fibrosis of the fibrosa. The macroscopic appearances are similar to the late stage of aortic and mitral chronic rheumatic disease but can be distinguished histologically.

Bacterial endocarditis

Animal models of bacterial endocarditis show that platelet deposition and bacteraemia must coincide for infection to be established. Microscopic platelet deposition due to endothelial damage occurs on any human valve in which there is regurgitant flow or abnormal cusp contact. Bicuspid aortic valves, prolapsed mitral valves, and rheumatic valves are therefore at risk of bacterial endocarditis should bacteraemia occur. Small platelet thrombi are also common on the apposition lines of the aortic cusps in individuals over the age of 70 years, even when the valve has three cusps and is functionally normal.

In established bacterial endocarditis the vegetations on valve cusps show that the surface is covered by thrombus rich in platelets containing abundant dividing organisms; beneath this is a layer of thrombus rich in fibrin, containing dormant but viable organisms. The valve cusp underneath the thrombus shows an intense, mixed inflammatory infiltrate with numerous polymorphs. Fibroblastic proliferation and vascularization are also prominent. The inflammatory and repair processes are remote from the organisms ensconced in the thrombus and destruction of the organism by polymorphs does not occur. The tissue destruction in the underlying cusp is due to the activity of the polymorphs and monocytes in the inflammatory response.

Degenerative change

Fragmentation of the collagen within the cusp fibrosa produces open spaces filled with connective-tissue mucins (myxoid change); this is a feature of the valve cusps in Marfan disease. The cusps expand rather than contract, but excess cuspal movement and trauma leads to fibrosis in the spongiosa. Similar cusp changes are also found as a cause of aortic or mitral disease in the absence of any systemic manifestation of connective-tissue disease.

Age-associated calcification develops within both the cuspal fibrosa and the valve rings, and is related to mechanical stress on mature collagen. Each cusp in the aortic valve may develop a C-shaped mass of calcium at the point of maximal flexion; in the mitral valve, calcium forms across the base of the posterior cusp. Calcification is enhanced by elevation of the calcium × phosphate product in blood, and may also be accelerated by smoking and hypercholesterolaemia.

Aortic valvular abnormalities

Normal anatomy

The cusps of the aortic valve are contained within a short fibrous sleeve, the upper border of which is scalloped to form a three-pointed coronet; the commissures are attached to each point. There are two ‘rings': the upper is the junction between the aortic sleeve and the base of the aorta; the lower is where the aortic sleeve joins both the anterior cusp of the mitral valve and the interventricular septum. The upper or supra-aortic ring (sinotubular junction) is marked by a distinct ridge at the level of the commissures, and has functional significance in that dilatation leads to regurgitation. The lower ‘ring' determines the size of prosthetic valve that can be inserted. In the closed position the aortic cusps abut on their ventricular faces, thus preventing any downward displacement in diastole. Fenestrations in the portion of the cusp above the closure line commonly develop with increasing age; they are of no functional significance.

Anatomical variation

Between 1 and 2 per cent of normal individuals have bicuspid aortic valves. These range from anatomically normal cusps of equal size to unequal cusps, the larger of which may have a deep cleft in the free edge. A fibrous raphe may cross the aortic surface of the larger cusp, or a fibrous chord may support the free edge. In bicuspid valves it is more common for both coronary arteries to arise from an anterior-facing sinus representing fusion of the right and left aortic sinuses.

Aortic stenosis

The pathogenesis of aortic stenosis can best be ascertained by viewing the valve from above (Fig. 1). In adults over the age of 40 years the most common cause of isolated aortic stenosis is calcification in a bicuspid valve. The calcification is dystrophic in type. It often begins in the raphe of the larger cusp, and extends throughout both cusps as bars and nodules of calcium. In the later stages all cusp architecture is obliterated, but it is usually possible to recognize the residual opening as a transverse slit extending across the whole width of the aortic root (Fig. 2). Commissural fusion is absent. The nodular masses of calcification in the cusps may erode on to the aortic surface, and become covered by thrombus; bacterial endocarditis becomes a risk. The centres of large, calcific nodules may break down into soft, gritty material, and spontaneous or iatrogenic embolism of calcific debris can occur. Bicuspid aortic valves become calcified from the age of 30 years onward, but it is not

unusual to find a non-stenotic bicuspid valve coincidentally at autopsy in an octogenarian.

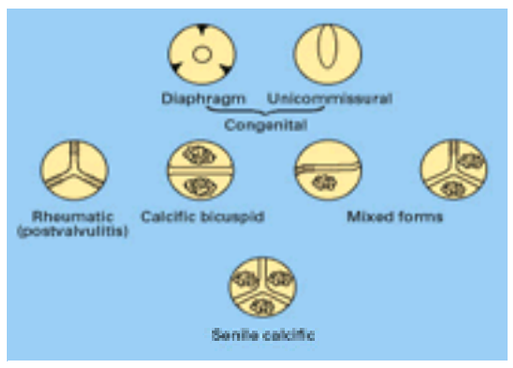


Fig. 1. Diagrammatic representation of the different forms of aortic valvular stenosis.



Fig. 2. Surgically excised cusps from bicuspid calcific aortic valvular stenosis: both cusps contain nodules of calcification; in the larger cusp this is accentuated in the raphe.

Tricuspid calcific aortic valvular stenosis is characterized by three distinct cusps without commissural fusion. Each cusp has a C-shaped nodule of calcium making it rigid and leaving a triradiate lumen. The mean age at surgery for patients with bicuspid stenotic valves is 65 years; those with tricuspid calcific aortic valves are a decade older. Patients with tricuspid calcific valves may, however, be as young as 50 years, and there is evidence that hypercholesterolaemia and smoking enhance the early calcification of the cusps by leading to inflammatory infiltration.

Concomitant coronary disease occurs in those with tricuspid calcific aortic valvular stenosis more commonly than in bicuspid stenotic valves.

Rheumatic aortic stenosis in its most typical form appears as three cusps with equal commissural fusion leaving a small, triangular, central, fixed orifice (Fig. 3).

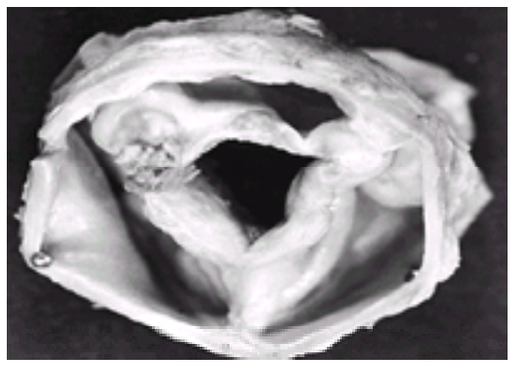


Fig. 3. Rheumatic aortic valvular stenosis: there is symmetrical fusion of all the commissures; the residual orifice is central.

True congenital aortic stenosis, where stenosis was present at or shortly after birth, ranges from a diaphragm with a central orifice and a hint of ridges where the three commissures should have formed, through what appear to be bicuspid valves with failure of separation of the commissures, to the uncommissural valve, which has a dome-shaped diaphragm and an eccentric, elliptical orifice. Aortic valvular dysplasia in infants is characterized by replacement of the whole valve by nodular masses of disorganized mesenchymal tissue.

A small number of stenotic aortic valves cannot be fitted into these categories. Bicuspid valves with commissural fusion are assumed to derive from coexistent rheumatic disease (Fig. 4). Difficulty arises over valves that show dystrophic calcification in three cusps but fusion of one commissure: these raise the question of whether asymmetrical commissural fusion occurs in chronic rheumatic disease. If typical mitral rheumatic lesions are present, the aortic lesion is usually considered to be rheumatic; if not, the question remains open. Acquired fusion of one commissure may also be confused with a bicuspid valve with a deep notch. In the acquired fusion both cusp edges can be traced up to the commissure; in the notched bicuspid valve they do not reach the commissure.



Fig. 4. Bicuspid calcific aortic valvular stenosis with superimposed rheumatic-type commissural fusion.

In any surgical series the relative proportion of each form of adult aortic valvular stenosis will depend on the selection of case material by geographic area and age.

Aortic regurgitation

Competence of the aortic valve depends on a critical relation between total cuspal area and aortic-root area at the commissural level. The total cuspal area normally exceeds the area of the root by up to 40 per cent, allowing the cusps to abut and support each other in the closed position. With age the aorta dilates, the diameter of the root at the supra-aortic ridge increases, and some compensatory increase in cuspal area occurs. Any consideration of the pathogenesis of aortic regurgitation must always take account of the relative proportions of root and cusp area.

Aortic regurgitation due to cuspal abnormalities with a normal aortic root

In chronic rheumatic disease, fibrosis leads to a drastic reduction in overall cuspal area. The presence or absence of concomitant commissural fusion determines whether coexistent stenosis is present. Perforations or tears in cusps, whether due to trauma or bacterial endocarditis, are also causes of regurgitation (Fig. 5). A small proportion of patients with bicuspid aortic valves develop regurgitation rather than stenosis. The risk is greatest where one cusp is significantly larger than the other or has a deep cleft in the free edge. Rupture of a fibrous chord supporting a large cusp may precipitate regurgitation. Tricuspid aortic valves in which the cusps are soft, redundant, and prolapsed, but in which the root is of normal size, are often designated as ‘floppy' aortic valves. The condition is regarded as analogous to the floppy mitral valve, but is very rare.



Fig. 5. Surgically excised cusps from a case of aortic regurgitation following bacterial endocarditis. One cusp has calcified vegetations on the free edge; another has a ‘windsock' cusp aneurysm, and a perforation through the body of the cusp.

Aortic regurgitation due to root disease

Any increase in cross-sectional area of the aortic root above that which can be compensated for by cuspal stretching results in regurgitation. Viewed from above the cusps fail to meet, leaving a central gap in the closed position (Fig. 6). The reduction in mutual support often allows one cusp to prolapse below its neighbour. This cusp will develop a localized, fibrous, nodular thickening of the free edge as a result of regurgitant flow (Fig. 7). Dilatation of the aortic root is a feature of Marfan disease, but the majority of patients with isolated disease of the aortic root have no systemic evidence of connective-tissue disorders. The condition is variously known as idiopathic non-inflammatory root dilatation or annuloaortic ectasia, and may be familial.

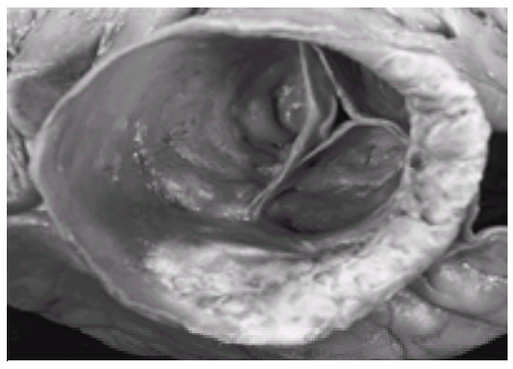


Fig. 6. Aortic-root dilatation leading to regurgitation: two cusps have prolapsed leaving a small central defect in the closed valve.

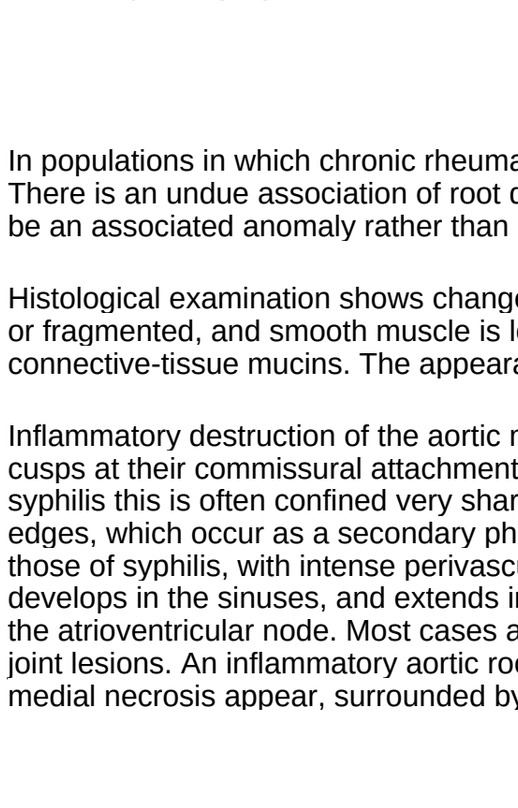


Fig. 7. Surgically excised cusps from aortic regurgitation due to root dilatation: the cusps are large without thickening apart from fibrosis along the free edge of the cusp secondary to regurgitant flow.

In populations in which chronic rheumatic valvular disease has vanished, idiopathic root dilatation is becoming the single most common cause of aortic regurgitation. There is an undue association of root dilatation with bicuspid aortic valves; the dilatation appears before and independently of regurgitation or stenosis and appears to be an associated anomaly rather than a secondary phenomenon.

Histological examination shows changes in the aortic media immediately above the supra-aortic ring. There is no inflammation. The elastic lamina appear straightened or fragmented, and smooth muscle is lost. In extreme cases, particularly in Marfan disease, elastic laminas may fragment sufficiently to leave spaces filled with connective-tissue mucins. The appearance is referred to as cystic medial necrosis, although the cell loss is likely to be due to apoptosis rather than necrosis.

Inflammatory destruction of the aortic media also leads to regurgitation. Syphilis causes both an increase in the size of the aortic root and a separation of the valve cusps at their commissural attachments. All medial destruction, whether inflammatory or non-inflammatory, leads to a wrinkled, scarred intima when viewed *en face*. In syphilis this is often confined very sharply to the first 5 cm or so of the aorta, and the coronary ostia may be narrowed. The cusps are normal, apart from rolled free edges, which occur as a secondary phenomenon. The histological changes of HLA-B27-related aortic disease in patients with ankylosing spondylitis are identical to those of syphilis, with intense perivascular infiltration of the media and adventitia by lymphocytes, plasma cells, and macrophages. In contrast to syphilis, fibrosis develops in the sinuses, and extends into the bases of the aortic cusps. Fibrosis also extends into the base of the anterior cusp of the mitral valve and into the area of the atrioventricular node. Most cases are associated with clinically expressed ankylosing spondylitis, but this valvular lesion also occurs in individuals with trivial or no joint lesions. An inflammatory aortic root and ascending aortic dilatation without cuspal involvement occurs in elderly people, particularly in women. Large areas of medial necrosis appear, surrounded by giant cells that are apparently associated with the fragmented elastic. The condition is related to, and may be associated with,

giant-cell arteritis elsewhere.

Mitral valve

Mitral stenosis

For all practical purposes the sole cause is chronic rheumatic valvular disease. In its simplest form, fusion of both commissures leads to a mobile diaphragm with a central oval aperture. Superimposed calcification may appear as nodules at the commissures (Fig. 8) or as more diffuse deposits in the cusps. Cuspal rigidity thus contributes to stenosis. Cusp calcification frequently occurs in the absence of commissural fusion. Fusion of the chords into solid pillars of collagen may superimpose an element of subvalvular stenosis. Histological examination of the atrial appendage in mitral stenosis may show residual Aschoff bodies; these are long-lasting granulomas that are a sign of previous rheumatic fever but do not indicate current activity. They may retain their characteristic histological appearances for up to 20 years before becoming small, fibrous scars.

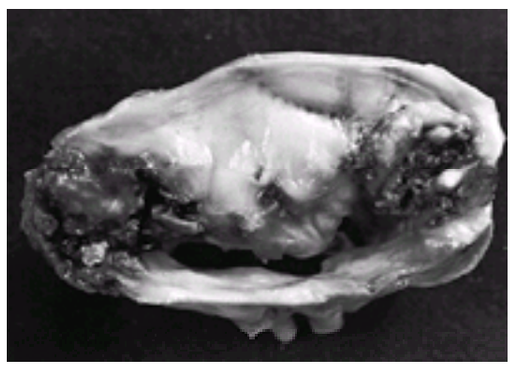


Fig. 8. Excised rheumatic mitral stenosis: both commissures are fused and ablated by nodules of calcium.

Mitral regurgitation

Fibrosis leading to retraction of the posterior cusp is typical of rheumatic disease and when associated with shortening of the chords leads to regurgitation, owing to the functional absence of a posterior cusp. The anterior cusp may be normal or even slightly expanded. Chordal fusion and shortening may also immobilize the anterior cusp (Fig. 9).



Fig. 9. Chordal fusion and shortening due to chronic rheumatic disease: the cusp is tethered to the apex of the papillary muscles by short, fibrous pillars.

Perforation through the body of the cusp is typical of the healed stage of bacterial infection. The perforation may have a well-defined edge or an aneurysmal bulge that appears blown out like a windsock. Healed bacterial infection leads to considerable fibrosis, which may obscure any pre-existing valvular lesion. Chordal rupture is also typical of bacterial endocarditis; in the healed stage, residual vegetations are often present as calcified nodules.

Cuspal expansion and chordal elongation are typical of the floppy mitral valve, the basis of which is myxomatous degeneration of the cusp fibrosa. Chordal elongation and cuspal expansion allow cusp prolapse and lead to the characteristic ballooned cusps (Fig. 10). The cusps appear to be larger than normal, but are in fact thicker since impact causes considerable surface fibrosis. The cusps feel soft and hypermobile, in contrast to rheumatic disease in which they are hard and rigid. The extent of cuspal involvement varies. The most typical form involves one or more scallops of the posterior cusp, but both cusps or the anterior cusp alone may be affected. When disease is limited the remaining cuspal tissue is often normal, and even with time it does not necessarily become abnormal. The central collagen of the chords also undergoes cystic change, and loses its tensile strength. Chordal rupture is the most significant event precipitating clinically significant regurgitation. The ruptured chords may fold back and become fused to the ventricular face of the cusp, or it may remain visible as a stump.



Fig. 10. Excised cusps of floppy mitral valve with a characteristic dome or balloon shape and long, thin chords; stumps of ruptured chords are visible.

Microscopic deposition of platelet thrombi on the apposition lines of the expanded cusps is almost universal, but, in the absence of bacterial endocarditis, larger vegetations do not occur. The microthrombi may account for the minor transient ischaemic attacks that occur in patients with prolapse of the mitral valve.

The pathogenesis of the floppy mitral valve is uncertain. Diffuse involvement of both cusps is found in patients with genetic abnormalities of connective-tissue synthesis, such as Marfan disease and the joint hypermobility syndromes. Most individuals with isolated mitral prolapse, however, have no systemic disorder. Another view of the pathogenesis is that the myxoid change is a ‘wear and tear' change in a valve that originally had minor anatomical abnormalities of the chordal support of the cusps.

There is a wide spectrum of severity of ischaemic mitral regurgitation. Necrosis of the papillary muscles occurs in up to 40 per cent of posterior–inferior acute infarcts, making ischaemic damage very common. At its most severe extreme there is rupture through either a whole papillary muscle at its base, or through one head or

portion of a head of a papillary muscle close to the chordal insertions. Rupture of a chord itself cannot be ischaemic since the chords are avascular.

Elongation (expansion) of infarcted papillary muscles may occur in the acute phase of infarction, and, although they remain intact, fibrosis perpetuates this elongation, leading to regurgitation. Fibrous scarring at the bases of the papillary muscles also alters the regulation of cusp apposition in systole and leads to mild regurgitation.

Calcification of the mitral ring is common in those over the age of 70 years, but is a misnomer because the calcification is just below the insertion of the cusp into the atrioventricular junction. Calcium splints the valve orifice, leading to a combination of mild stenosis and regurgitation. The age of the patients precludes consideration of the condition as a surgical problem.

Ring dilatation is a cause of regurgitation. The ring to which the cusps are attached in the atrioventricular junction is not a solid, fibrous structure with sufficient innate tensile strength to maintain its normal size if the ventricle dilates. Other than in patients with Marfan disease, the size of the ring is dependent more on the structural and functional state of the left ventricle.

Tricuspid and pulmonary valvular disease

Tricuspid rheumatic disease is analogous to that in the mitral valve and due to commissural fusion and cuspal retraction. Rheumatic tricuspid disease is usually concomitant with both aortic and mitral disease, and virtually never occurs in isolation. Right ventricular ring dilation rather than cuspal fibrosis is a major factor in regurgitation. Tricuspid and pulmonary stenosis and regurgitation may also develop in individuals with carcinoid tumours. The aortic and mitral valves can also be involved. The pathological processes are striking in that fibrosis is superimposed on the cusp, leaving the underlying fibrosa intact.

Further reading

Allen W, Matloff J, Fishbein M. Myxoid degeneration of the aortic valve and isolated aortic regurgitation. *American Journal of Cardiology* 1985; **55**: 439–44. [Describes myxoid degeneration of the aortic valve cusps suggesting a primary cusp abnormality analogous to the floppy mitral valve.]

Bergfeldt L. HLA-B27-associated cardiac disease. *Annals of Internal Medicine* 1997; **127**: 621–9. [Overall review of the clinical and pathological features of aortic valve disease associated with ankylosing spondylitis.]

Connolly HM *et al.* Valvular heart disease associated with fenfluramine-phentermine. *New England Journal of Medicine* 1997; **337:** 581–8. [The analysis of 24 cases of the valve lesions which follow treatment with appetite suppression drugs.]

Davies M, Moore B, Braimbridge M. The floppy mitral valve: study of incidence, pathology and complications in surgical, necropsy and forensic material. *British Heart Journal* 1978; **40**: 468–81. [A pathological study describing the spectrum of changes and complications found in floppy mitral valves due to myxoid degeneration.]

Davies M, Treasure T, Parker D. Demographic characteristics of patients undergoing aortic valve replacement for stenosis: relation to valve morphology. *Heart* 1996; **174:** 174–8. [A surgical study contrasting the pathological and clinical features of bicuspid and tricuspid calcified aortic valve stenosis.]

Edwards J. The congenital bicuspid aortic valve. *Circulation* 1961; **23**: 485–8. [A very early description describing the bicuspid aortic valve as an entity and separating it from rheumatic disease.]

Guiney T, Davies M, Parker D, Leech G, Leatham A. The aetiology and course of isolated severe aortic regurgitation: a clinical, pathological and echocardiographic study. *British Heart Journal* 1987; **58**: 358–68. [A review of the causes and prognosis of isolated aortic regurgitation as seen in patients presenting for surgery emphasizing the importance of aortic not dilatation.]

Hanson T, Edwards B, Edwards J. Pathology of surgically excised mitral valves. One hundred consecutive cases. *Archives of Pathology and Laboratory Medicine* 1985; **109**: 823–8. [An analysis of the spectrum of pathological processes which lead to mitral valve replacement.]

Lachman A, Roberts W. Calcific deposits in stenotic mitral valves—extent and relation to age, sex, degree of stenosis in a study of 164 operatively excised valves. *Circulation* 1978; **57**: 808–15. [An exhaustive account of the way in which calcification of the mitral valve develops and causes symptoms.]

Nishimura R, Schaff H, Shub C, Gersh B, Edwards W, Takik A. Papillary muscle rupture complicating acute myocardial infarction—analysis of 17 patients. *American Journal of Cardiology* 1983; **51**: 373–8. [A thorough account of the pathology and presentation of papillary muscle rupture due to ischaemic heart disease.]

Olsen L, Subramanian R, Edwards W. Surgical pathology of pure aortic insufficiency—a study of 225 cases. *Mayo Clinic Proceedings* 1984; **59**: 835–41. [A large surgical series in which the causes of pure regurgitation are described. The paper explains the declining importance of rheumatic disease.]

Passik C, Ackermann D, Pluth J, Edwards W. Temporal changes in the causes of aortic stenosis: a surgical pathological study of 646 cases. *Mayo Clinic Proceedings* 1987; **62**: 119–23. [An account of the changing patterns in the cause of aortic valve stenosis that has occurred over the last few decades.]

Perloff J, Roberts W. The mitral apparatus. Functional anatomy of mitral regurgitation. *Circulation* 1972; **46**: 227–39. [A beautiful account of the anatomy and physiology of the normal mitral valve which then describes how regurgitation develops in both structural and functional ways.]

Peterson M, Roach R, Edwards J. Types of aortic stenosis in surgically removed valves. *Archives of Pathology and Laboratory Medicine* 1985; **109**: 829–32. [A description of the spectrum of pathology which causes aortic valve stenosis emphasizing the importance of the bicuspid valve.]

Roberts W, Morrow A, McIntosh C. Congenitally bicuspid aortic valve causing severe, pure aortic regurgitation without superimposed infective endocarditis: analysis of 13 patients requiring aortic valve replacement. *American Journal of Cardiology* 1981; **47**: 206–9. [A description of the fact that one of the natural histories of the bicuspid aortic valve is to develop pure regurgitation.]

Sadee AS, Becker AE, Verheul HA, Bouma B, Hoedemaker G. Aortic valve regurgitation and the congenitally bicuspid aortic valve: a clinico-pathological correlation. *British Heart Journal* 1992; **67**: 439–42. [A further paper emphasizing that bicuspid aortic valves may become purely regurgitant without an element of stenosis.]

Waller B, Morrow A, Maron B. Etiology of clinically isolated severe, chronic, pure mitral regurgitation: analysis of 97 patients over 30 years of age having mitral valve replacement. *American Heart Journal* 1982; **104**: 276–88. [An analysis of the cause of pure mitral regurgitation emphasizing the increasing importance of non-rheumatic disease.]

40.9.2 Surgery for acquired aortic and pulmonary valve disease

Graeme L. Hammond and Michael A. Coady

Aortic valve disease
Surgical anatomy
Aortic stenosis
Aortic insufficiency
Technical considerations
Minimally invasive aortic valve surgery
Pulmonary valve surgery
Surgical anatomy
Pulmonary stenosis
Pulmonary insufficiency
Further reading

Aortic valve disease

The most common etiologic factors that result in aortic valve disease are rheumatic fever, endocarditis, congenital abnormalities, and collagen vascular disorders. The physiologic sequelae and symptomatic profile of aortic stenosis, aortic insufficiency, or combined stenotic and insufficient lesions are remarkably similar, at least initially. This chapter, however, will review acquired aortic valvular disease as seen in adults and will not examine hypertrophic cardiomyopathies, which may also mimic aortic valve disease.

Surgical anatomy

The anatomy of the aortic valve is shown in Fig. 1 and Fig. 2. This trileaflet, semilunar valve occupies the aortic orifice in the subcoronary position and is similar in structure to the pulmonary valve. Cusp nomenclature follows their relationship to the coronary arteries, therefore, one cusp is referred to as the right coronary cusp and is situated on the anterior wall, and two are referred to as the non-coronary and left coronary cusps and are located on the medial and posterior walls. The cusps are attached at their base to the annulus and are affixed to each other at the commissures. The margins are divided into two lunulas at the midpoint, referred to as the nodule of Arantius. The valve leaflets coapt during diastole, sealing the central orifice of the valve. Originating at the valve annulus are the sinuses of Valsalva. During systole, these localized dilatations ofthe aortic and pulmonary roots provide a recess for the valve leaflets that keep them out of the path of ejected blood, thereby eliminating turbulence and gradients across the valve. Because a small gap exists between the leaflets and sinuses during systole, eddy currents form in the sinuses which enable the valve cusps to separate easily from the vessel walls during early diastole thus minimizing regurgetant flow. In the aortic root, the left and right coronary arteries normally arise above the base of the cusps, and are not obstructed by the leaflets during systole.

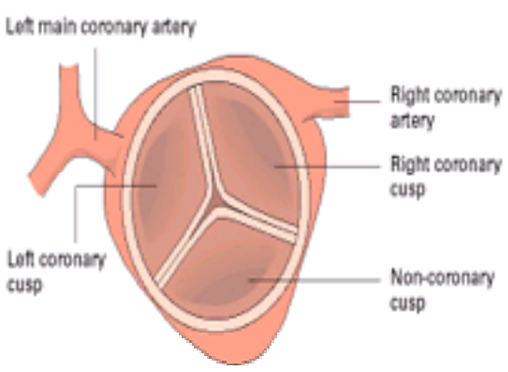


Fig. 1. Sketch of an aortic valve as seen from above.

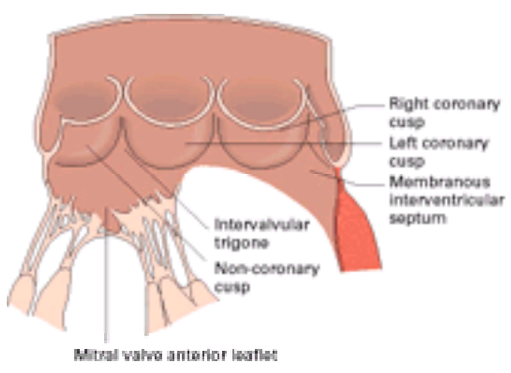


Fig. 2. Enface view of the aortic valve illustrates the subvalvular anatomy.

Despite many years of effort to develop an ideal replacement for the native aortic valve, none have been as durable or free from malfunction as the natural valve. The difficulty in designing a replacement that precisely mimics the human native aortic valve in both structure and function affirms the refined nature of biology and engineering represented by this valve.

Aortic stenosis

Aortic stenosis is characterized by a long, latent, asymptomatic period, even in the presence of a clinically detectable, systolic murmur in the aortic area. The most common causes of aortic stenosis include calcific aortic stenosis in a congenitally bicuspid aortic valve and senile calcific aortic stenosis associated with older age. Rheumatic fever is also an important risk factor for the development of aortic stenosis. The late onset of symptoms in aortic stenosis is due to the great ability of the heart to compensate for increasing pressure gradients across the aortic valve. The principal compensating measure is the development of myocardial hypertrophy. Hypertrophy can progress for many years before the volume of muscle fibers to capillaries exceeds the ratio required for oxidative metabolism. Even at this point, myocardial cells develop subtle biochemical changes, such as shifts in their lactate dehydrogenase isozyme profiles, which allow myocardial cells to contract under relatively anaerobic conditions for many years. The development of hypertrophy is also associated with unremitting fibrosis that, although insidious, eventually results in a stiff, non-compliant ventricle that is unable to pump blood effectively. This leads to decreased ventricular compliance, increased left ventricular end-diastolic pressure, and progressive pulmonary congestion. Left ventricular pressures over 300 mmHg may develop, and myocardial hypertrophy may result in hearts weighing up to 700 g. On physical examination, patients with aortic stenosis demonstrate a harsh diamond-shaped crescendo–decrescendo murmur, heard in the aortic area and radiating to the carotid arteries. The electrocardiogram may demonstrate left ventricular hypertrophy with strain.

Usually by the age of 60, the classic symptoms of aortic stenosis—angina, syncope, and congestive heart failure—begin to appear. Angina is the most common symptom, developing in approximately 50 to 75 per cent of patients; however, half of these patients also have concomitant coronary disease. These symptoms are expected with a thickened left ventricular mass manifesting high intramyocardial pressures, resisting coronary flow during diastole. In addition, there is a relative increase in the muscle mass to blood supply ratio, higher systolic ejection pressures, and increased overall myocardial demand. Syncope, occurring in 25 per cent of patients, results from brief episodes of ventricular tachycardia/ventricular fibrillation or prolonged asystoles secondary to the localized acidotic environment in which the heart must work. Finally, congestive heart failure results when severely elevated left ventricular end-diastolic pressure leads to progressive ventricular systolic

dysfunction.

Although initially the symptoms usually appear separately in any one individual, with disease progression, the three major symptoms often appear together. With the onset of symptoms, the prognosis is grave with virtually 100 per cent of patients dying within 3 years unless the aortic obstruction is corrected. Of the three symptoms, congestive heart failure carries the worst overall prognosis, with angina affecting prognosis the least. Nevertheless, the onset of any one symptom necessitates cardiologic evaluation and, usually, aortic valve replacement. Occasionally, the presence of a loud aortic ejection murmur by itself will prompt referral of an asymptomatic patient. In such a case, the patient should undergo two-dimensional echo Doppler examination to determine the presence of critical aortic stenosis and flow characteristics across the valve.

Following non-invasive evaluation, any patient in whom the severity of the stenosis is uncertain, or in whom additional information is required, should undergo cardiac catheterization. Simultaneous determination of pressure and flow (i.e. cardiac output) across the aortic valve is then used to calculate the valve area according to the Gorlin formula which, in simplified version, is expressed as:

$$\text{Valve area} = \text{cardiac output}/\sqrt{\text{gradient}}$$

The normal aortic valve has an area of 2.5 to 3.5 cm^2. An aortic valve gradient greater than 50 mmHg or a valve area of less than 1 cm^2 represents aortic stenosis. When the aortic valve area becomes less than 0.7 cm^2 this represents critical aortic stenosis.

Critical aortic stenosis is associated with a high incidence of sudden death and hence patients should undergo surgery even if there are minimal or no symptoms. The presence of angina, congestive heart failure, or syncope usually develops in association with a valve area of less than 1 cm^2. If a patient with few symptoms does not have critical aortic stenosis, it is justified to follow the patient clinically for the onset or progression of symptoms. Patients are followed by electrocardiography to evaluate left ventricular hypertrophy and ventricular strain, chest radiograph to assess cardiac size, and two-dimensional echocardiography to determine ventricular wall thickness and contractility. If these parameters show evidence of progressive change, then the patient should be considered for surgery.

Aortic insufficiency

Unlike aortic stenosis, chronic aortic insufficiency can be tolerated for many years. In contrast to the 100 per cent 3-year mortality for aortic stenosis, aortic insufficiency results in a 50 per cent mortality over a 10-year period (Fig. 3). The most common cause of aortic insufficiency is rheumatic fever. Other causes include bicuspid or unicuspid aortic valves, endocarditis, chest trauma, and annular ectasia from an aortic aneurysm or dissection. Similar to aortic stenosis, chronic aortic insufficiency also leads to left ventricular hypertrophy; however, the volume-loading strain of aortic insufficiency, as opposed to pressure-loading strain of aortic stenosis, produces distension and progressive ventricular dilatation. It is the insidious dilatation and fiber stretching that produces progressive, irreversible loss of myocardial fibers characteristic of aortic insufficiency. When fiber loss has reached a critical point, operative intervention will do little to alter the patient's functional classification. It is therefore critical to replace the aortic valve before the onset of clinically obvious congestive heart failure. This is in contrast to patients with aortic stenosis where it is justified to operate after the development of congestive heart failure.

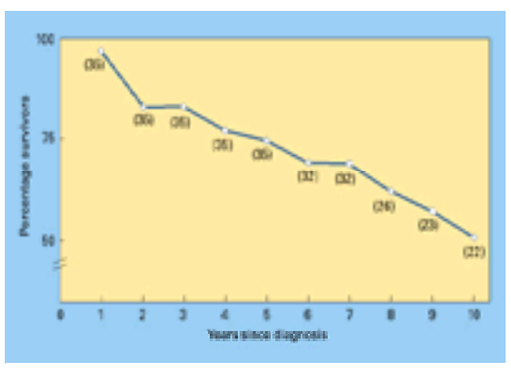


Fig. 3. Percentage survival of patients with aortic insufficiency treated medically. (From: Rapaport E. Natural history of aortic and mitral valve disease. *American Journal of Cardiology* 1975; **35**: 221.)

Patients with aortic insufficiency present with a prominent diastolic murmur at the left sternal border as well as the aortic area. An Austin Flint murmur may also be present, which is thought to be secondary to regurgitant flow and subsequent preclosure of the mitral valve. Bounding ‘water-hammer' pulses may also be present throughout the body.

In patients with aortic insufficiency, a change in exercise tolerance or the onset of shortness of breath are important premonitory signs. Patients with aortic insufficiency also develop angina and syncope, but these symptoms do not carry the same ominous prognosis as associated with aortic stenosis. In addition to accurate history taking and careful physical examination, the decision as to the timing of surgery also takes into consideration changes in cardiac silhouette by chest radiograph and chamber size by two-dimensional echocardiography. The pulmonary capillary wedge pressure (PCWP) and left ventricular end-diastolic pressure may be significantly elevated, with concomitant near equalization of the aortic and left ventricular end-diastolic pressure.

Immediate medical management for patients with aortic insufficiency includes the use of vasodilators (i.e. sodium nitroprusside) to increase forward flow and lower left ventricular volume, wall stress, and oxygen demand. Symptomatic patients should undergo valve surgery, as these individuals demonstrate an improvement in survival and clinical symptoms. Depressed left ventricular function is not a contraindication for surgery. In managing patients who are asymptomatic, the decision of operative timing is often clouded by the difficulty in predicting when left ventricular dysfunction becomes irreversible. Asymptomatic patients may be divided into two groups depending on whether there is normal or abnormal left ventricular function. Asymptomatic patients with normal left ventricular function can be managed medically until early symptoms or echocardiography demonstrates worsening left ventricular function. If the end-systolic dimension is greater than 55 mm, the patient should undergo cardiac catheterization. If the catheterization confirms the presence of severe aortic insufficiency, the patient should undergo surgery. Use of an intra-aortic balloon pump is contraindicated in patients with aortic insufficiency, as inflation during diastole exacerbates the aortic insufficiency.

Acute aortic insufficiency, such as that produced by endocarditis or collagen vascular disorders (i.e. isolated detached cusp), carries a much higher early mortality. This directly relates to the lack of compensatory mechanisms, as the left ventricle has not had a chance to hypertrophy. Accordingly, while one would like to treat patients with acute endocarditis with a full 6-week course of antibiotics, this is often impossible. The timing of aortic valve replacement secondary to acute aortic insufficiency depends primarily upon clinical symptoms. The onset of orthopnea or shortness or breath will invariably proceed within hours to acute pulmonary edema, and should be an important signal to proceed with surgery. Electrocardiographic changes of volume overload such as ST-T wave changes, atrial arrhythmias, and axis changes also invariably proceed clinical deterioration by 6 to 12 h and are very helpful indicators of the pressing need for surgery. Patients with aortic insufficiency secondary to acute type A thoracic dissections should also be operated upon urgently.

Technical considerations

Valve repair has not yet gained general acceptance as a form of surgical treatment for aortic valve disease. Therefore, in most cases, present management calls for valve replacement. There are many prosthetic devices available with various advantages and disadvantages in comparison with each other. Prosthetic heart valves are covered in more detail elsewhere. In general, mechanical valves are pre-ferred in patients who are young or have a life expectancy of more than 10 to 15 years, or who require long-term anticoagulant therapy for other reasons (i.e. atrial fibrillation). Bioprosthetic valves, on the other hand, are favored in patients who are elderly or have a life expectancy of less than 10 to 15 years, or who cannot take long-term anticoagulant therapy. A bileaflet mechanical valve (St. Jude, Carbomedics, Edwards-Duromedics) or homograft valve is preferred in patients with a small aortic annulus in whom the largest effective outlet is desired.

The technique of aortic valve replacement (Fig. 4, Fig. 5, Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, Fig. 11 and Fig. 12) for insufficiency or stenosis is very similar except for the following considerations.

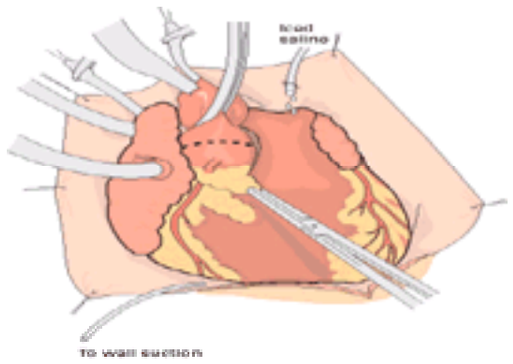


Fig. 4. Exposure of heart for aortic valve replacement, showing placement of iced saline administration and withdrawal lines and site of aortotomy.

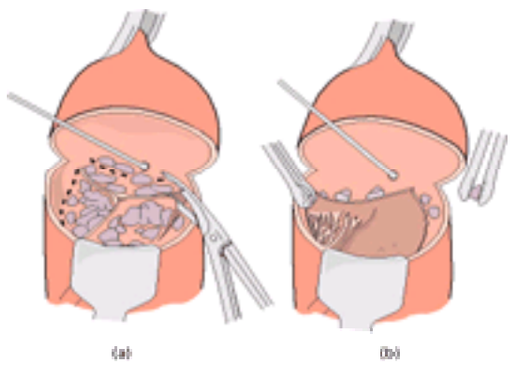


Fig. 5. Removal of valve and decalcification of annulus. If excessive calcium is left behind, perivalvular leak may result.

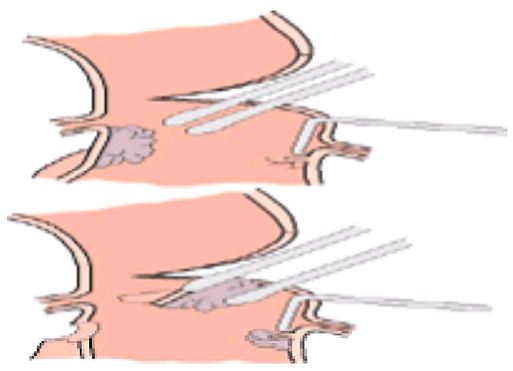


Fig. 6. Complication from decalcification. Decalcification of aortic annulus must be conducted carefully, since vigorous decalcification can result in perforation.

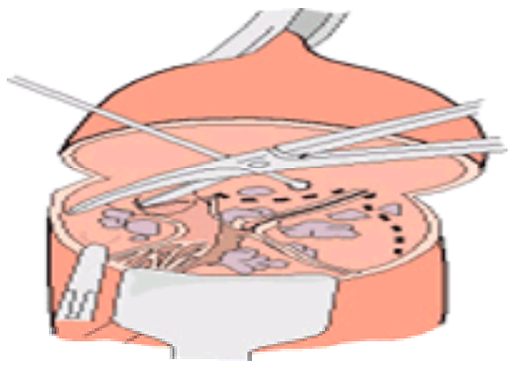


Fig. 7. Distortion of valve annulus with loss of landmarks by heavy calcification can result in accidental detachment of anterior leaflet of the mitral valve when removing non-coronary leaflet.

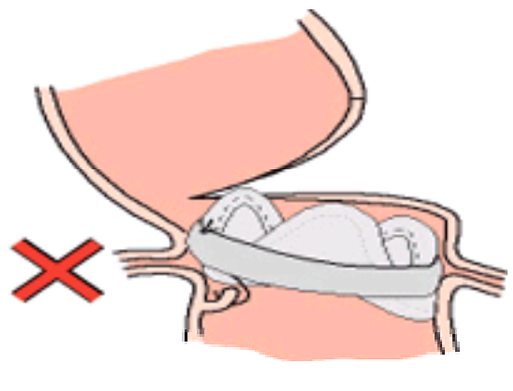


Fig. 8. Occlusion of coronary ostia can result when attempting to implant an oversized valve or by not observing relationship of sewing ring to coronary ostia when tying sutures.

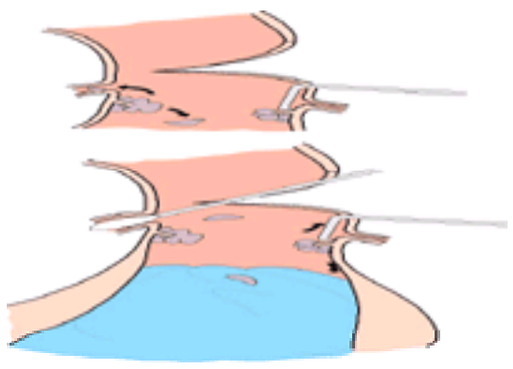


Fig. 9. Further complications from decalcification. Calcium can lodge in coronary ostia or ventricular chamber, with subsequent embolization. This can be prevented by placing a culture stick (or coronary perfusion cannula) in left coronary orifice and sponge in left ventricle during decalcification. Retractor covers right coronary orifice.

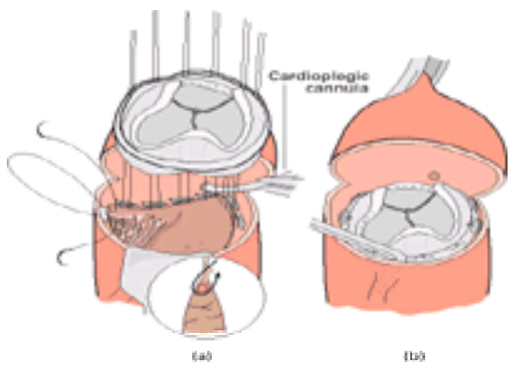


Fig. 10. Insertion of Teflon-backed mattress sutures in annulus. Inset: Location of His bundle in upper part of muscular septum. Deep sutures placed in this area can interrupt the atrioventricular conduction pathway. Sutures should be placed no deeper than the membranous septum. After sutures have been tied, seating of sewing ring below right coronary orifice is checked with a right angle clamp. The left coronary orifice can be visualized.

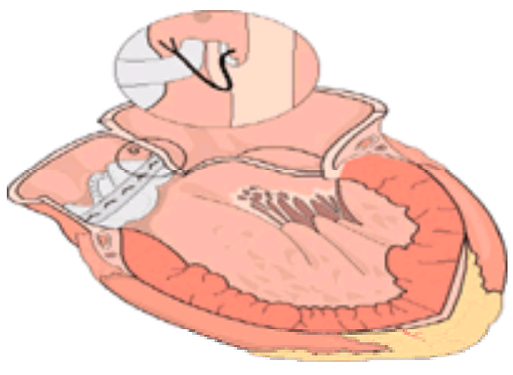


Fig. 11. Valve in place, showing anchorage with Teflon-backed mattress sutures.

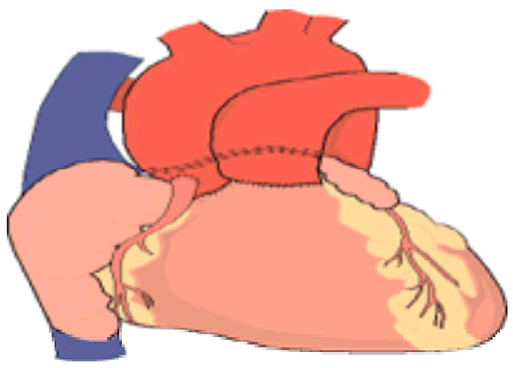


Fig. 12. The pulmonary autograft is sutured into the aortic position with the coronary arteries reimplanted; the right ventricular outflow tract has been reconstructed with an appropriate pulmonary allograft.

1. Calcification. Heavy calcification is invariably associated with aortic stenosis rather than aortic insufficiency. A calcified annulus may obliterate landmarks necessary for careful debridement such as the aortic leaflet of the mitral valve and areas adjacent to the atrioventricular node. Pieces of calcium may also fall into the coronary orifices or become entrapped in trabeculas, with subsequent cerebral or systemic embolization. Additionally, perivalvular leaks secondary to inadequate annular decalcification or iatrogenic perforations of the aortic root may occur from excessive decalcification.
2. Abscess formation. Acute endocarditis may lead to abscess formation. The abscess may burrow into the myocardium with subsequent destruction of the annulus and valve leaflets. Considerable judgment and imagination is often required in such cases to sew and properly seat the valve. It is frequently necessary to place Teflon-backed sutures from outside the aorta into the aortic lumen and through the valve sewing ring or to use the medial wall of the pulmonary artery to buttress the suture line and insure a competent valve–annular alignment.
3. Narrow aortic root. This is invariably associated with aortic stenosis or mixed stenotic and insufficient lesions. Tertiary orifice obstruction from ball and cage valves precludes the use of ball valves for the narrow aortic root. Several operations have been developed to overcome this problem. Annular enlargement procedures such as those described by Manouguian, Konno-Rastan have been developed and employed with varying degrees of success. These operations, in general, are technically difficult, carry a high incidence of mortality and morbidity, and are not always successful. Problems with the small aortic annulus have now been largely resolved with development of newer valves with greater effective orifice to annular ratios. Specifically, a 19-mm St. Jude valve will fit most narrow aortic annuli and provides satisfactory relief of outflow obstruction, even in large adults. Although it is unusual not to be able to place a 19-mm valve, it may occasionally be necessary to seat the valve at an angle to the plane of the annulus. In other words, the sutures are placed in the annulus in the usual way, but the valve is lowered into place so that the sewing ring is below the right and left coronary orifices but angled into the non-coronary sinus. The left and right coronary annular sutures are tied first, thereby abutting the sewing ring to the annulus below the coronary ostia. The non-coronary sutures are tied last, which then allows the valve to ride above the annulus in this portion of the aorta. Accordingly, annulus enlargement procedures should be reserved for those rare patients who will not accept a 19-mm St. Jude valve. Infants and children. Biologic and mechanical prosthetic valves for aortic valve replacement in infants and children are associated with significant complications, and are considered palliative. In 1967, Ross described aortic valve replacement using a pulmonary autograft. Aside from eliminating the complications associated with prosthetic valves in these patients, the autologous valve maintains potential for growth, and alleviates the need for anticoagulation. Currently, the Ross procedure is the procedure of choice for aortic valve replacement in the pediatric population. Although size mismatch between the aortic and pulmonary valve remains a relative contraindication, aortic root tailoring in cases of excessive aortic annular dilatation has extended the use of the Ross procedure.

Minimally invasive aortic valve surgery

Historically, aortic valve replacement has been performed through a median sternotomy with cardiopulmonary bypass instituted via the ascending aorta and right atrium. Reconsideration of the type of incision used for aortic valve replacement has allowed minimally invasive aortic valve replacement. Minimally invasive surgery may reduce patient pain and discomfort, and possibly improve rehabilitation. The approach uses several small incisions, reduces exposure of the patient to surgical trauma, blood utilization, and operative dissection, yet permits satisfactory visualization of the aortic root.

There have been three general operative approaches presented for minimally invasive aortic valve replacement. All three incisions are less than 10 cm in length and include the right parasternal incision (Fig. 13), the ‘mini' sternotomy (Fig. 14), and the transverse sternotomy. Although experience with these techniques is evolving, the early results in all three minimally invasive approaches for aortic valve replacement appear encouraging. At a recent Society of Thoracic Surgeons' Meeting, Cosgrove presented results from his first 50 minimally invasive aortic valve replacements using the right parasternal and transverse sternotomy. The surgical mortality rate was 2 per cent, with a 10 per cent morbidity rate (conversion to sternotomy, bleeding, and stroke). Cohn *et al.* at the Brigham and Women's Hospital recently

reported their experience with 36 patients using either the right parasternal or ‘mini' sternotomy. In their series, the surgical mortality and morbidity was 5.5 per cent and 8.3 per cent, respectively. All have shown a reduction of incisional pain, reduced rehabilitation after the hospital stay, and faster return to activities.

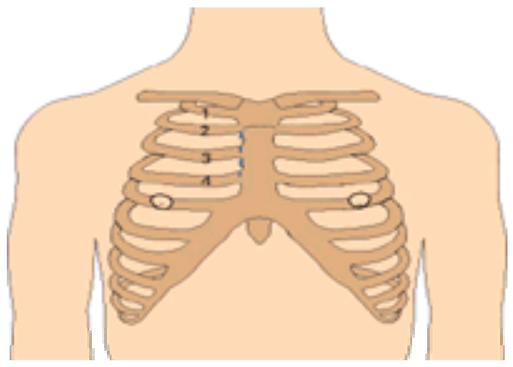


Fig. 13. Incision used in the parasternal area over the second and third costal cartilages.

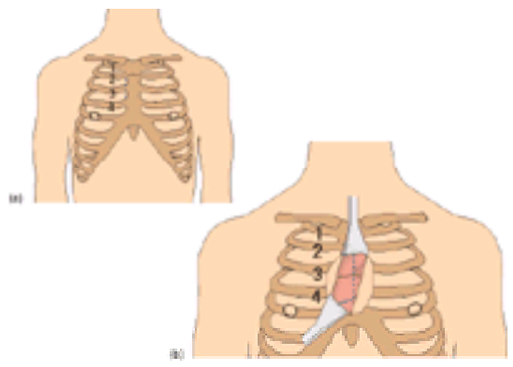


Fig. 14. Incision used in the ‘mini' sternotomy with extension of the incisions to the third intercostal space.

Pulmonary valve surgery

Surgical anatomy

Similar to the aortic valve, the semilunar pulmonary valve consists of three leaflets or cusps, and marks the dividing point between the pulmonary overflow tract and the pulmonary artery. The pulmonary valve is thinner and more tenuous than the aortic valve, and consists of the right, left, and anterior leaflets that coapt during diastole. Like the aortic valve, the sinuses of Valsalva consist of dilatations of the pulmonary root and perform the same function as in the aortic root.

Pulmonary stenosis

With the advent of balloon valvulotomy for pulmonary stenosis, surgery for pulmonary valve disease is now reserved almost exclusively for the treatment of obstructive pulmonary artery abnormalities or the secondary pulmonary insufficiency that results from surgery upon the right ventricular outflow tract. Most patients with right ventricular outflow tract obstruction are children and conditions include hypoplastic right ventricular outflow tract, absent pulmonary valve syndrome (invariably associated with annular stenosis), and double outlet right ventricle. The symptoms normally include dyspnea, fatigue, congestive heart failure, cyanosis caused by shunting through atrial septal defects or a patent foramen ovale and, in infants, poor feeding. Patients rarely become symptomatic with peak outflow tract gradients less than 50 mmHg.

Occasionally, patients may be referred for pulmonary surgery because of residual infundibular gradients following balloon valvuloplasty. Approximately 20 per cent of patients with pulmonary stenosis have secondary infundibular obstruction. In a report by Thayer and Rao, 5 of 13 infundibular gradients disappeared following valvuloplasty in 62 patients and 5 patients developed new infundibular gradients. Patients who had systemic infundibular gradients following valvuloplasty responded well to propranolol and, as is the case with surgical pulmonary valvulotomy, regression of infundibular stenosis occurred. Accordingly, surgery is not indicated for infundibular stenosis following balloon valvuloplasty.

Pulmonary insufficiency

For all intensive purposes, pulmonary insufficiency is not caused by rheumatic fever, endocarditis, or collagen vascular disorders. It is iatrogenic, and is produced by transannular repair of tetralogy of Fallot or repair of other surgical disorders involving the right ventricular outflow tract. It may be serious or inconsequential but its presence should be followed carefully for the onset of right ventricular deterioration. Major pulmonary regurgitation is often associated with peripheral pulmonary artery stenosis. The most important symptom of a deteriorating right ventricle is fatiguability. The key signs are a murmur of tricuspid insufficiency and its associated systemic physical findings. Important radiologic and angiographic signs include progressive cardiomegaly, a regurgitant fraction of 30 per cent or greater, and a transannular patch to descending aorta ratio of 2.5 or greater. In the presence of these findings, the patients should undergo surgery promptly. In a study performed by Ilbawi *et al.*, 42 patients with radiographic findings of pulmonary insufficiency (following tetralogy of Fallot repair) underwent pulmonary valve replacement. At subsequent catheterization of 18 patients, it was noted that right ventricular function improved completely in 83 per cent (5/6) of those who were operated upon within 2 years of tetralogy repair, while right ventricular function remained abnormal in all 12 patients who underwent valve replacement more than 2 years following tetralogy repair. Although we recommend that pulmonary competence be restored early after the documentation of significant pulmonary insufficiency, a review of the literature indicates that satisfactory results are obtainable many years following the surgical creation of pulmonary insufficiency.

Surgery for pulmonary insufficiency

The surgery for pulmonary insufficiency falls into three major categories:

1. valve replacement with mechanical or biologic valve prosthesis;
2. conduit replacement with aortic or pulmonary artery homografts or Dacron-valve conduits; and
3. single cusp repair with cusp bearing transannular patch.

Valve replacement must often be performed in a main pulmonary artery that is of insufficient size to accept a valve capable of allowing flow without a significant gradient. As a consequence, valve insertion into the pulmonary artery is usually performed with a Dacron patch angioplasty, as shown in Fig. 15.

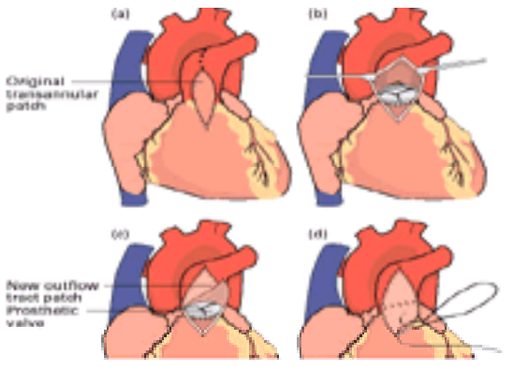


Fig. 15. Pulmonary valve replacement is begun with (A) an incision in the original transannular patch. The incision is continued into the main pulmonary artery. (B) The patch is excised and a prosthetic valve secured to the native annulus posteriorly with simple interrupted sutures. (C) A new outflow tract patch is sewn with a running suture to the pulmonary artery. (D) The prosthetic valve is secured anteriorly to the patch with interrupted horizontal mattress sutures, and patch closure of the right ventricular outflow tract is completed.

Despite the well-known accelerated degradation of biologic valves in the aortic and mitral positions in children, they appear to provide excellent long-term function on the right side of the heart. In contrast, mechanical valves placed in the tricuspid and mitral positions are associated with a high rate of thrombosis or dysfunction caused by tissue ingrowth. Fleming *et al.* reported that of 18 biologic valves inserted on the right side of the heart, only two had to be replaced over a 9-year follow-up period. On the other hand, three of four patients who received St. Jude valves required re-replacement after 36 to 56 months.

Replacement of the pulmonary valve with homografts or valve-conduits is generally reserved for conditions in which right ventricular to pulmonary artery continuity must be established, such as pulmonary atresia and truncus arteriosis. As most of these operations are performed in infancy or adolescence, reoperation is often required as a result of growth. In a series of conduit operations at the Mayo Clinic, 44 per cent of patients required reoperation at 10 years. Accordingly, the necessity of reoperation must be considered when choosing the type of conduct. Dacron graft valve conduits (Fig. 16) have the advantage of being readily available and, because they develop a surrounding fibrous peel, are easily redissected at the time of reoperation, their primary disadvantage is suture line bleeding at the time of implantation, stiffness which limits their versatility and increases the possibility of kinking, and a tendency of the xenograft valve to calcify.

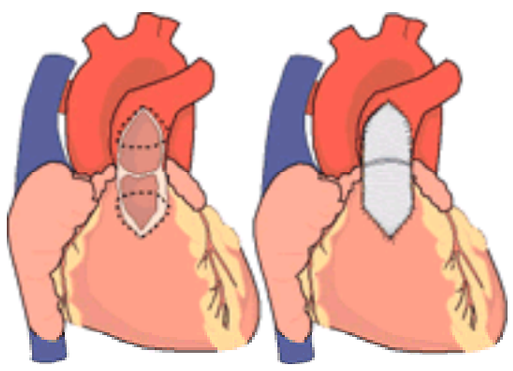


Fig. 16. Left: The incision extending from the pulmonary artery bifurcation into the outflow tract of the right ventricle. The broken lines show the sites of the two suture lines. Right: The valved conduit *in situ*, lying in the trough formed by the opened pulmonary artery.

More versatile is the fresh or cryopreserved aortic or pulmonary homograft. Homografts provide good functional results, are technically easy to use, and the valve has a low incidence of calcification. The shortage of homograft material mandates that both the aorta and pulmonary arteries be harvested from donor hearts. In the United States, Cryolife, Inc. (Marietta, Georgia) preserves postmortem intact hearts, where the aorta and pulmonary arteries are removed, prepared, and frozen under liquid nitrogen. The homografts are then made available to surgeons. The advantage of the pulmonary homograft is that it is thin walled and supple, thereby allowing a technically simple distal anastomosis to the recipient pulmonary artery and minimizing distal suture line bleeding. The proximal anastomosis of the pulmonary homograft, however, is associated with potential technical problems due to the lack of tough fibrous tissue at the base of the valve. This anatomic situation requires that donor outflow tract muscle tissue be incorporated into the proximal portion of the homograft. Muscle tissue does not cryopreserve well and, when thawed, sutures tend to tear through the friable muscle with the potential for false aneurysm development at the proximal suture line. Aortic homografts, on the other hand, have a substantial fibrous ring at the base of the valve and have the added advantage with the attachment of the anterior leaflet of the mitral valve to the non-coronary annulus of the aortic valve. Positioning the aortic homograft allows for the insertion of a transannular patch using the anterior leaflet of the mitral valve as the patch. The disadvantage of the aortic homograft is that the distal anastomosis is somewhat more difficult to construct due to the thicker diameter of the aortic wall. The general method of insertion is shown in Fig. 17. At present, the greatest disadvantage of the aortic and pulmonary homografts is their limited availability and their tendency to become very thin walled and densely adherent to surrounding tissues with the attendant increase in operative mortality during reoperation as a result of hemorrhage.

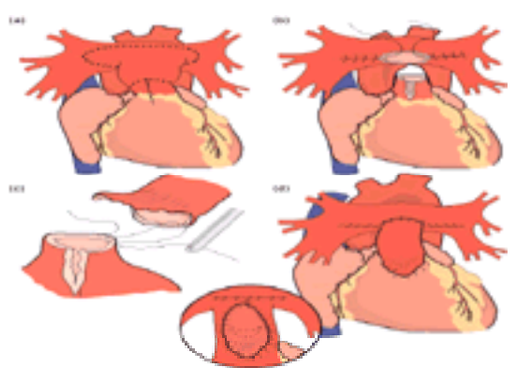


Fig. 17. (a) Line of resection of main pulmonary artery and anterior portion of main branches. Pulmonary artery is transected at annulus and an incision is made into right ventricular outflow tract (RVOT). (b) Arterioplasty of right and left pulmonary artery branches reduces their caliber. (c) Aortic homograft is sutured to annulus. Anterior mitral leaflet of homograft is used as RVOT patch. (d) Completed repair using aortic homograft. When monocusp patch is used (inset) the posterior pulmonary artery wall is preserved.

Single cusp angioplasty is an attempt to create some degree of pulmonary competence when the pulmonary annulus has been divided. One approach is to insert a transannular patch consisting of an aortic homograft with one cusp preserved or bovine or porcine pericardial patches with a monocusp. A second approach is to fold the inferior margin of the incised native pulmonary artery into the outflow tract and then reconstruct the pulmonary artery with pericardium. The portion of pulmonary artery wall in the lumen acts as a valve cusp. Both procedures seem to help in the postoperative period by decreasing pulmonary regurgitation. In a recent review of the effects of monocuspid valves on pulmonary insufficiency, however, they have not been shown to prevent short-term postoperative pulmonary insufficiency and did not improve immediate postoperative outcome for patients following surgical repair of tetralogy of Fallot. Long-term follow-up of these procedures remains to be evaluated.

Further reading

Agarwal KC, Edwards WD, Feldt RH, Danielson GK, Puga FJ, McGoon DC. Clinicopathological correlates of obstructed right-sided porcine-valved extracardiac conduits. *Journal of Thoracic and*

Cardiovascular Surgery 1981; **81**: 591–601.

Agarwal KC, Edwards WD, Feldt RH, Danielson GK, Puga FJ, McGoon DC. Pathogenesis of nonobstructive fibrous peels in right-sided porcine-valved extracardiac conduits. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 584–9.

Bigras JL, Boutin C, McCrindle BW, Rebeyka IM. Short-term effect of monocuspid valves on pulmonary insufficiency and clinical outcome after surgical repair of tetralogy of Fallot. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 33–7.

Bonow RO, Rosing DR, Kent KM, Epstein SE. Timing of operation for chronic aortic regurgitation. *American Journal of Cardiology* 1982; **50**: 325–36.

Bove EL, Kavey R-EW, Byrum CJ, Sondheimer HM, Blackman MS, Thomas FD. Improved right ventricular function following late pulmonary valve replacement for residual pulmonary insufficiency or stenosis. *Journal of Thoracic and Cardiovascular Surgery* 1985; **90**: 50–5.

Cohn LH, Adams DH, Couper GS, Bichell DP. Minimally invasive aortic valve replacement. *Seminars in Thoracic and Cardiovascular Surgery* 1997; **9**: 331–6.

Copeland JG, Griepp RB, Stinson EB, Shumway NE. Long term follow-up after isolated aortic valve replacement. *Journal of Thoracic and Cardiovascular Surgery* 1977; **74**: 875–89.

Cosgrove DM, Sabik JF, Navie J. Minimally invasive valve surgery. Presented at the Thirty-third Annual Meeting of The Society of Thoracic Surgeons, San Diego, CA, January 1997.

Cowgill LD, Campbell DN, Kelminson L, Clarke DR. Repair of pulmonary valve insufficiency using an autologous monocusp. *Annals of Thoracic Surgery* 1986; **42**: 587–9.

Durham LA, desJardins SE, Mosca RS, Bove EL. Ross procedure with aortic root tailoring for aortic valve replacement in the pediatric population. *Annals of Thoracic Surgery* 1997; **64**: 482–6.

Elkins RC, Santangelo KL, Seltzer P, Randolph JD, Knott-Craig CJ. Pulmonary autograft replacement of the aortic valve: an evolution of technique. *Journal of Cardiac Surgery* 1992; **7**: 108–16.

Finck SJ, Puga FJ, Danielson GK. Pulmonary valve insertion during reoperation for tetralogy of Fallot. *Annals of Thoracic Surgery* 1988; **45**: 610–13.

Fioretti P *et al.* Electrocardiography in chronic aortic insufficiency: Is replacement too late when left ventricular end systolic dimension reaches 55 mm? *Circulation* 1983; **67**: 216–21.

Fleming WH, Sarafian LB, Moulton AL, Robinson LA, Kugler JD. Valve replacement in the right side of the heart in children: long-term follow-up. *Annals of Thoracic Surgery* 1989; **8**: 404–8.

Gaash WH. Aortic valve disease: timing of valve replacement surgery. In: Starek PJK, ed., *Heart valve replacement and reconstruction*. Year Book Medical, St. Louis, 1987.

Geha AS, Hammond GL, Laks H, Stansel HC, Glenn WWL. Factors affecting performance and thromboembolism after porcine xenograft cardiac valve replacement. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 377–84.

Gorlin R, Gorlin SG. Hydraulic formula for calculation of area of stenotic mitral valve, other cardiac valves, and central circulatory shunts. *American Heart Journal* 1951; **41**: 1.

Gundry SR, Shattuck OH, Sardari FF, Bailey LL. Cardiac operations in adults and children by ministernotomy: facile minimally invasive cardiac surgery. Presented at the Thirty-third Annual Meeting of The Society of Thoracic Surgeons, San Diego, CA, January 1997.

Hammond GL, Nadal-Ginard B, Talner NS, Market CL. Myocardial LDH isozyme distribution in the ischemic and hypoxic heart. *Circulation* 1976; **53**: 637–43.

Hammond GL, Geha AS, Kopf GS, Hashim SW. Biological versus mechanical valves. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 182–98.

Ilbawi MN *et al.* Factors that exaggerate the deleterious effects of pulmonary insufficiency on the right ventricle after tetralogy repair. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 36–44.

Ilbawi MN *et al.* Experience with St. Jude medical valve prosthesis in children. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 73–9.

Kay PH, Ross DN. Fifteen years' experience with the aortic homograft: the conduit of choice for right ventricular outflow tract reconstruction. *Annals of Thoracic Surgery* 1985; **40**: 360–4.

Konno S *et al.* A new method for prosthetic valve replacement in congenital aortic stenosis associated with hypoplasia of the aortic valve ring. *Bulletin of the Heart Institute of Japan* 1974; **15**: 1–17.

McGoon DC, Danielson GK, Puga FJ, Ritter DG, Mair DD, Ilstrup DM. Late results after extracardiac conduit repair for congenital cardiac defects. *American Journal of Cardiology* 1982; **49**: 1741–9.

Manouguian S, Seybold-Epting W. Patch enlargement of the aortic valve by extending the aortic incision into the anterior mitral leaflet: new operative technique. *Journal of Thoracic and Cardiovascular Surgery* 1979; **78**: 402–12.

Miyamura H, Kanazawa H, Hayashi J, Eguchi S. Thrombosed St. Jude medical valve prosthesis in the right side of the heart in patients with tetralogy of Fallot. *Journal of Thoracic and Cardiovascular Surgery* 1987; **94**: 148–56.

Osbakken M, Bove AA, Span JF. Left ventricular function in chronic aortic regurgitation with reference to end-systolic pressure, volume and stress relations. *American Journal of Cardiology* 1981; **47**: 193.

Raasten H, Koncz J. Aortoventriculoplasty. A new technique for treatment of left ventricular outflow obstruction. *Journal of Thoracic and Cardiovascular Surgery* 1976; **71**: 920.

Rao, PS. Indications for balloon pulmonary valvuloplasty. *American Heart Journal* 1988; **16**: 1661–2.

Rao PS, Solymar L, Mardini MK, Fawzy ME, Guinn G. Anticoagulant therapy in children with prosthetic valves. *Annals of Thoracic Surgery* 1989; **47**: 589–92.

Ross DN. Replacement of aortic and mitral valves with a pulmonary autograft. *Lancet* 1967; **ii**: 956–8.

Sievers HH *et al.* Short-term hemodynamic results after right ventricular outflow tract reconstruction using a cusp-bearing transannular patch. *Journal of Thoracic and Cardiovascular Surgery* 1983; **86**: 777–83.

Thapar MK, Rao PS. Significance of infundibular obstruction following balloon valvuloplasty for valvar pulmonic stenosis. *American Heart Journal* 1989; **118**: 99–103.

40.9.3 Diagnosis and surgical management for tricuspid and mitral valve disease

Kenneth L. Franco and Graeme L. Hammond

Tricuspid valve surgery
Anatomy of the tricuspid valve
Tricuspid stenosis
Tricuspid regurgitation
Indications for surgery
Tricuspid valve commissurotomy
Tricuspid annuloplasty
Tricuspid valve replacement
Infective endocarditis
Ebstein's anomaly
Carcinoid heart disease
Summary
Mitral valve surgery
Anatomy of the mitral valve
Mitral stenosis
Indications for surgery
Mitral regurgitation
Indications for surgery
Surgical approaches to the mitral valve
Repair of the mitral valve
Further reading

Tricuspid valve surgery

The management of tricuspid valve disease remains controversial and frequently presents a diagnostic and therapeutic challenge for the cardiologist and cardiac surgeon. There is disagreement over whether an abnormal tricuspid valve should be left alone, repaired or replaced, which repair should be done, and which prosthesis should be implanted. Many patients have concomitant mitral and aortic valve disease, many have had one or more previous operations, and they are often in advanced stages of deterioration. It is, however, possible to improve the results obtained in the treatment of patients with multivalvular heart disease by earlier surgical intervention and by performing a concomitant procedure on the tricuspid valve, if indicated, before myocardial dysfunction develops.

Anatomy of the tricuspid valve

The tricuspid valve is composed of four anatomic elements: valve leaflets, the annulus, chordae tendineae, and papillary muscles (Fig. 1). There are three valve leaflets, the anterior being the largest of the three, separated from the septal leaflet by a small posterior leaflet. The base of the septal leaflet harbors the penetrating portion of the conducting system. The chordae tendineae of the anterior leaflet are attached to the anterior papillary muscle, which is the dominant papillary muscle in the right ventricle. The posterior and septal leaflets are attached by chordae tendineae to a small posterior papillary muscle and directly to the right ventricular myocardium.

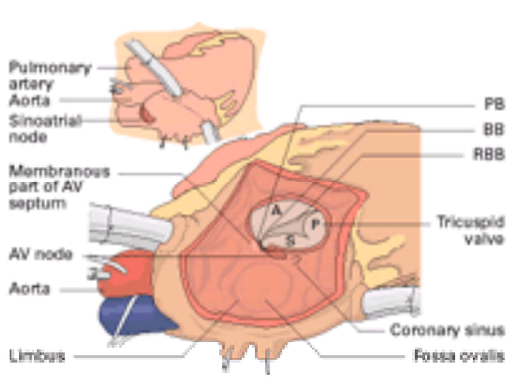


Fig. 1. Incision into right atrium and exposure of the tricuspid valve. TV-3, leaflets of tricuspid valve (S, septal; a, anterior; p, posterior); PB, penetrating bundle; BB, bundle branch; RBB, right bundle branch on the right side of the intraventricular septum; SA node, sinoatrial node.

Tricuspid stenosis

The essential problem in tricuspid stenosis is the inability of the right atrium to empty normally, thus causing the pressure in the right atrium to rise. It is usually of rheumatic origin and occurs in conjunction with mitral and aortic disease, or both. It may also be congenital, or due to carcinoid syndrome or, occasionally, right atrial myxoma.

Stenosis of the tricuspid valve is usually due to diffuse, fibrous thickening of the leaflets, with fusion of two of the three commissures. The chordae tendineae of the tricuspid valve may thicken and shorten, but chordal fusion is generally not as severe as that seen in rheumatic mitral stenosis. The leaflets in tricuspid stenosis usually show minimal calcific deposits, making them amenable to commissurotomy rather than replacement procedures. Unfortunately, most stenotic valves are also moderately regurgitant, therefore usually precluding successful commissurotomy.

Tricuspid stenosis is usually associated with a gradient of at least 5 mmHg, and a valve area that is reduced to less than 1.5 cm^2. The most frequent symptoms are dyspnea and fatigue. Elevation of the mean right atrial pressure to above 10 mmHg causes jugular venous distension, hepatosplenomegaly, ascites, and peripheral edema. The characteristic auscultatory finding is a diastolic rumble, best heard at the left lower sternal border. The influence of respiration helps to differentiate this murmur from that of mitral stenosis: the diastolic rumble of tricuspid stenosis is markedly accentuated during the inspiratory phase. The presence of atrial fibrillation, increased venous pressure, and a diastolic murmur in the classic location should strongly suggest the possibility of significant tricuspid stenosis. The most characteristic radiographic findings include right atrial enlargement, lack of pulmonary arterial enlargement, and relatively clear lung fields. M-mode and two-dimensional echocardiography may demonstrate leaflet thickening, reduced diastolic excursion of the leaflets, and characteristic doming of the valve. Doppler echocardiography can allow estimation of the diastolic gradient across the valve.

Tricuspid regurgitation

Tricuspid regurgitation is the result of an incompetent tricuspid valve allowing blood to enter the right atrium during right ventricular contraction. This is most commonly due to enlargement of the annulus secondary to congestive heart failure and pulmonary hypertension. Secondary valvular pathology of this nature is often due to rheumatic heart disease, which in turn causes pulmonary arterial hypertension, right ventricular hypertension, and eventually right ventricular dilatation with poor coaptation of the valve leaflets. The most common cause of isolated tricuspid regurgitation is infective endocarditis occurring in intravenous drug abusers. Less common causes of tricuspid regurgitation include trauma (especially valve defects caused by biopsy forceps in heart transplant recipients), the carcinoid syndrome, and congenital defects such as Ebstein's anomaly.

Symptoms of tricuspid regurgitation include fatigue, dyspnea, orthopnea, and peripheral edema. Distension of the neck veins is common, venous pulsations occurring in a large number of patients. Pulsatile hepatomegaly is usually found, as well as the characteristic auscultatory finding of a holosystolic murmur at the left sternal border that increases with respiration. The majority of patients with tricuspid insufficiency have atrial afibrillation and radiographically apparent cardiomegaly. The

degree of tricuspid regurgitation can be determined by contrast and/or Doppler echocardiography; two-dimensional echocardiography can be used to evaluate vegetations on the tricuspid valve leaflets, holes in the leaflets, flail leaflets, and abnormal coaptation. Cardiac catheterization reveals the characteristic hemodynamic criteria for tricuspid regurgitation, including a right atrial pressure contour showing ventricularization (i.e., a prominent V-wave with a steep Y-descent). In addition, right ventriculography can demonstrate regurgitation of dye into the right atrium and, in severe cases, opacification of the vena cava and hepatic veins.

Indications for surgery

The decision to proceed with surgery in patients with multivalvular heart disease is usually based on the severity of the aortic and mitral valve disease rather than the severity of tricuspid valve disease. Factors affecting the decision usually include whether a tricuspid valve procedure should be added to the mitral and/or aortic valve procedure and, if so, which tricuspid procedure should be performed, annuloplasty or valve replacement. The degree of stenosis and/or regurgitation can be estimated intraoperatively by palpation of the tricuspid valve orifice through the right atrial appendage. If surgery is not performed on the tricuspid valve as the initial operation, its orifice can again be examined by digital palpation through the right atrial appendage as bypass is discontinued after the mitral or aortic valve procedure. Intraoperative two-dimensional echocardiography provides precise information on the degree of residual valvular insufficiency after repair of either the mitral or tricuspid valve.

Tricuspid valve commissurotomy

Tricuspid stenosis is almost always accompanied by some degree of regurgitation and can be treated successfully by commissurotomy. Commissurotomy is usually done under direct vision and at the anteroseptal or posteroseptal commissures, to avoid increasing the degree of tricuspid insufficiency. The procedure may be combined with an annuloplasty ring to correct valve regurgitation. Valve replacement is occasionally necessary if disease of the leaflets and subvalvular mechanism is advanced or if severe regurgitation cannot be relieved by annuloplasty.

Tricuspid annuloplasty

When tricuspid regurgitation occurs secondarily to left ventricular failure, dilatation of the annulus occurs chiefly in the right ventricular wall, preventing coaptation of the anterior and posterior leaflets. The relative consistency in size of the septal leaflet allows it to be used as a guide to the size of the tricuspid valve required following reconstruction.

Three basic reconstructive techniques are used in the treatment of tricuspid regurgitation: annular plication, annular ring insertion, and semicircular annuloplasty. Annular plication, first described by Kay in the United States of America, converts the tricuspid into a bicuspid valve by excluding the posterior leaflet, and it is accomplished by plicating the annulus over the posterior leaflet. DeVega described a second type of annuloplasty, which consists of a semicircular suture that narrows the annulus along the anterior and posterior leaflets by a purse-string suture effect (Fig. 2). Carpentier devised a semirigid ring with an opening near the area of the septal commissure. His technique distributes the tension over the entire annulus and reduces the incidence of injury to the conduction system by eliminating sutures in the area of the bundle of His (Fig. 3). Transparent valve obturators are used to size the septal leaflet, which determines the overall size of the tricuspid annulus after a DeVega annuloplasty suture has been tied or a Carpentier ring inserted. Valve competence is checked by filling the right ventricle with cold saline from a bulb syringe. Digital palpation of the tricuspid valve when the patient is no longer on cardiopulmonary bypass can be used to substantiate the effectiveness of the repair and confirm the presence or absence of tricuspid regurgitation. Intraoperative two-dimensional echocardiography may also be used to measure any residual tricuspid regurgitation.

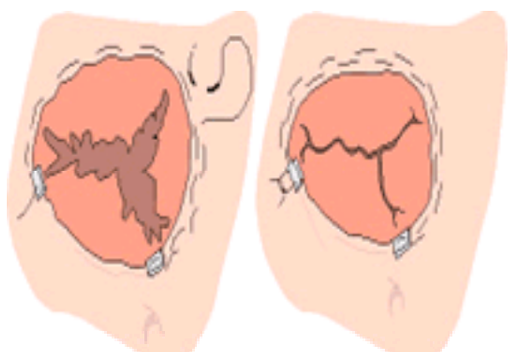


Fig. 2. DeVega tricuspid annuloplasty: suturing is done in the annulus of the anterior and posterior leaflets in order not to injure the conduction tissue.

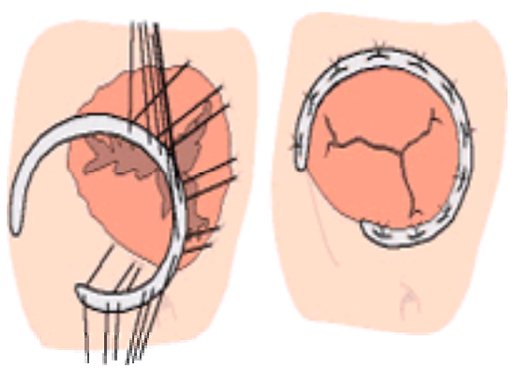


Fig. 3. Carpentier ring annuloplasty: a gap in the ring is to avoid suturing the area of the bundle of His.

Tricuspid valve replacement

The tricuspid valve should be replaced if it is not repairable (Fig. 4). When replacing the valve a 4-mm portion of the septal leaflet should be left to avoid injury to the conduction tissue when placing the valve sutures. Placement of the sutures with the heart perfused and beating facilitates recognition of conduction disturbances if the patient is in sinus rhythm and shortens rewarming after associated left-sided cardiac procedures; alternatively, the valve may be replaced under hypothermic, cardioplegic arrest. All patients undergoing replacement of the tricuspid valve should have a right ventricular epicardial lead placed at the time of surgery to prevent any conduction abnormalities occurring in the perioperative period. Many surgeons consider the porcine bioprosthesis to be the valve of choice in the tricuspid position because of the low incidence of associated thromboembolic complications, including thrombosis. The St Jude bileaflet mechanical valve, with its low profile and hemodynamics, may offer an advantage (especially in younger, good-risk patients because it is more durable than a tissue valve) in the tricuspid position than other mechanical valves, which tend to have a higher incidence of thrombosis, but it also has a higher incidence of thrombosis than with tissue valves. The higher incidence of prosthetic thrombosis in the tricuspid position is probably due to the lower pressures involved in the right-sided chambers, lower flow velocities, larger areas of stagnation, and a high incidence of valve entrapment by trabeculae. In two recent studies, preoperative risk factors predictive of hospital death after valve replacement included pulmonary hypertension and congestive heart failure.

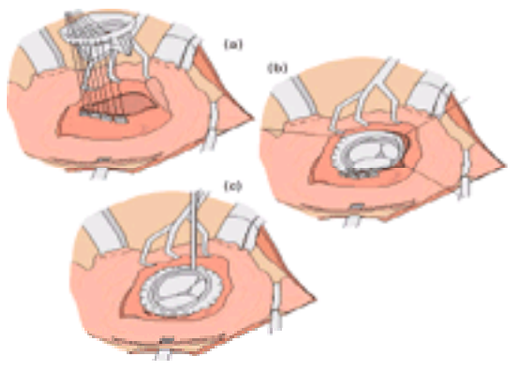


Fig. 4. Tricuspid valve replacement: everting interrupted mattress sutures are passed through the annulus and the sewing ring of a bioprosthetic valve, which is then seated into position as the sutures are tied.

Infective endocarditis

Right-sided infective endocarditis is one of the most serious complications of intravenous drug abuse, but may also occur following placement of permanent, indwelling, central venous catheters. The organisms responsible include *Staphylococcus aureus, S. epidermidis*, Gram-negative bacteria, and fungi. The majority of patients with right-sided endocarditis respond to medical therapy, but some will require surgical treatment for uncontrolled sepsis, recurrent pulmonary emboli, or congestive heart failure. Operative treatment includes vegetectomy, valvuloplasty, valve excision (provided that the pulmonary arterial and right ventricular pressures are normal), partial mitral-valve homograft, and valve replacement (should be reserved only for cases of severe valve destruction). Many of these patients are poorly compliant and will return to drug abuse: recurrence of infection is the rule rather than the exception.

Ebstein's anomaly

Ebstein's anomaly is characterized by downward displacement of the basal attachments of the septal and posterior leaflets of the tricuspid valve. This reduces the size of the right ventricular chamber, a portion of the ventricle above the valve being left as part of the right atrium (the atrialized ventricle). The entire wall of the right ventricle, both above and below the abnormally inserted tricuspid valve, is very thin and dilated. An additional feature is abnormal attachments of the free edge of the valve leaflets, which usually produce a tricuspid valve that is incompetent but may also be stenotic. Disability results from tricuspid valve incompetence and from cyanosis, due to right to left shunting of blood through an associated atrial septal defect or patent foramen ovale. There may be associated anomalies such as pulmonary stenosis or accessory conduction pathways. Surgery is recommended for patients in New York Heart Association classes 3 or 4, or those who are severely cyanotic, have paradoxical emboli, or supraventricular and ventricular arrhythmias. Surgical management includes closure of the atrial septal defect or patent foramen ovale, plication of the atrialized portion of the right ventricle, tricuspid valve repair or replacement, correction of any other associated anomalies, and division of any accessory conduction pathways that are present.

Carcinoid heart disease

Carcinoid heart disease occurs in approximately 50 per cent of patients with carcinoid syndrome and results from metastasis of carcinoid tumors of the midgut. The endocardial myofibromatous lesions usually affect the right side of the heart, causing dysfunction of the tricuspid and pulmonary valves. This may lead to volume and pressure loads on the right-sided cavities and sometimes to right ventricular failure. Valve reconstruction or replacement is recommended before extensive fibrosis and severe right heart failure have developed. Somatostatin is given prophylactically to block the effects of vasoactive substances that may be released from the tumor during the operative procedure and to reduce the complications caused by their release in the perioperative period.

Summary

In summary, because the tricuspid valve is subjected to lower pressures and lower mechanical stresses than the mitral valve, it is more forgiving by nature and therefore almost always amenable to successful repair. Because tricuspid valve replacement has an approximately 7.5 to 10 per cent complication rate, including pannus or thrombus formation on mechanical valves and primary valve failure with bioprosthesis requiring replacement at an average of 7 years, replacement should be reserved for only the most rare and unusual conditions.

Organic tricuspid insufficiency, secondary to rheumatic disease, endocarditis, or trauma, lends itself to repair. Infected, immobile, or calcified areas should be resected and the remaining valve reconstructed by transfer of leaflets and chordae or with autologous, glutaraldehyde-treated pericardium. Because autologous tissue has a tendency to calcify or stiffen when placed in the heart, the conditions of glutaraldehyde treatment are very important. The pericardium should be treated in 0.625 per cent glutaraldehyde for 15 min before used to patch holes or reconstruct new leaflets. Polytetrafluoroethylene, as described for the mitral valve, can be used as chordae if the pericardium is needed to replace the free edge of a leaflet. The valve is then tested for insufficiency in the usual way by filling the right ventricle with saline. It is not generally necessary to insert an annular ring when the leaflets are repaired primarily for acute disease, unless it is evident that there has been considerable annular dilatation.

Mitral valve surgery

Surgery for the mitral valve presents a diagnostic and therapeutic challenge for both the cardiologist and the cardiac surgeon. Recent advances in imaging techniques and a better understanding of the interaction between the mitral valve and the left ventricle have increased the number of patients being referred for reconstruction of the mitral valve. Improvements in myocardial protection and increased experience with valve reconstructive techniques, the availability of new bioprosthetic valves, and the several, low-profile, bileaflet mechanical valves have improved the survival of these patients. The use of minimally invasive incisions may further reduce the morbidity of mitral valve surgery.

Anatomy of the mitral valve

The mitral valve apparatus is a complex, coordinated mechanism that requires the functional integrity of six anatomic elements working in harmony. These elements are the anterior and posterior valve leaflets, the primary, secondary, and tertiary chordae tendineae, the anterolateral and posteromedial papillary muscles, the annulus, the left ventricular myocardium, and the left atrial endocardium (Fig. 5). There is mounting evidence that the mitral valve is an integral part of the left ventricle and that its anatomic presence plays a part in left ventricular geometry and mechanics. Papillary and chordal anchoring of the ventricular wall during systole seems essential for efficient left ventricular emptying, and preservation of these structures by either reconstructive surgery or valve replacement seems to be important in reducing the incidence of left ventricular dysfunction in the perioperative period.

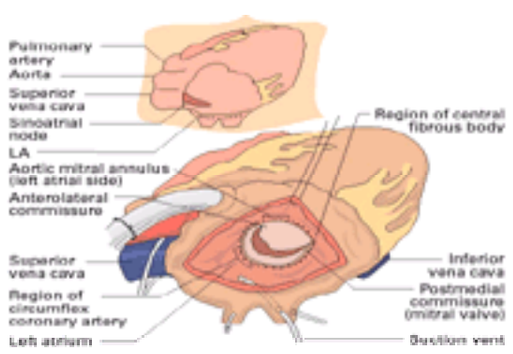


Fig. 5. Incision into left atrium and exposure of the mitral valve (PA, pulmonary artery; AO, aorta; A, anterior leaflet; P, posterior leaflet).

Mitral stenosis

With few exceptions, such as a left atrial myxoma, mitral stenosis is always due to chronic rheumatic heart disease. Rheumatic mitral stenosis is characterized pathologically by fibrous fusion of both valve commissures, fibrosis of both leaflets, and fused, thickened, and shortened chordae tendineae, all of which contribute to obstruction of the mitral valve orifice. Annular and subannular calcification, an enlarged left atrium, atrial fibrillation, and thrombus formation can occur as the disease process progresses, unless medical therapy is instituted. The valve area has to reach less than 2 cm^2 before normal diastolic blood flow begins to be obstructed. A reduction in valve area to 1 cm^2 is considered to be critical: the left atrial pressures required to achieve adequate ventricular filling through this small orifice will result in appreciable dyspnea with minimal effort. Patients with mitral stenosis suffer from dyspnea, fatigue, a gradual reduction in exercise tolerance, cough, hemoptysis, and occasionally angina pectoris, which is related to a very low cardiac output. The classical signs of a stenotic mitral valve include an opening snap, loud first heart sound, and a diastolic rumble. Chest radiographs may show enlargement of the cardiac chambers, especially the left atrium and the right ventricle, and the pulmonary venous pattern may also be pronounced. The electrocardiogram will confirm the presence of atrial fibrillation: fibrillation may increase the risk of systemic embolization in patients not receiving anticoagulants. Evaluation of mitral valve disease relies heavily on two-dimensional echocardiography, which provides excellent dynamic images of the leaflets, chordae, papillary muscles, and left ventricular myocardium. It can also define the presence and amount of calcification, valve mobility, and involvement of the subvalvular apparatus, all of which are extremely helpful in deciding whether a stenotic valve is suitable for commissurotomy. Doppler echocardiography can be used to determine the severity of mitral stenosis. Cardiac catheterization is no longer done routinely, but should be considered in selected patients in whom there are signs or symptoms suggestive of coronary arterial disease, or in those over 60 years of age.

Indications for surgery

Patients with mitral stenosis usually remain asymptomatic for many years, while the severity of mitral stenosis is increasing. It is generally accepted that patients with critical mitral stenosis and severe symptoms of effort dyspnea, pulmonary edema, hemoptysis, or right heart failure should undergo surgery. A patient with less severe symptoms might be considered for surgical intervention such as valvuloplasty if they have severe pulmonary hypertension at rest or during exercise, recurrent systemic emboli despite anticoagulant therapy, or reduced left ventricular function that would benefit from relief of mitral valve obstruction by allowing greater left ventricular filling.

Mitral regurgitation

Any disruption in the proper systolic closure of the mitral valve may result in mitral regurgitation. Many causes affect different parts of the valvular apparatus. Recent studies indicate that rheumatic heart disease is not the predominant cause of mitral regurgitation. Mitral annular dilatation may be due to chronic left ventricular dilatation, Marfan syndrome, or secondary subannular calcification. Leaflet pathology, which includes fibrous thickening, calcification, or destruction, can be secondary to rheumatic heart disease, mitral valve prolapse, endocarditis, or Marfan syndrome. The chordae tendineae may be shortened, thickened, fused, elongated, or may rupture secondarily to rheumatic heart disease, mitral valve prolapse, endocarditis, or trauma. Papillary muscle dysfunction, secondary to rupture or fibrosis, can be due to coronary arterial disease, myocardial infarction, trauma, or cardiomyopathy. The clinical symptoms and findings in patients with mitral regurgit1tion can vary, depending upon the severity of mitral regurgitation, its underlying cause, whether it is an acute or chronic process, and the status of the left ventricle. Common symptoms include fatiguability, dyspnea, a decrease in exercise tolerance, chest pain (if coexisting with coronary arterial disease), congestive heart failure, and atrial fibrillation as the disease progresses. The classic auscultatory finding for the clinical diagnosis of mitral regurgitation is a holosystolic murmur, with radiation to the axilla and the left infrascapular area. The electrocardiogram shows atrial fibrillation in 75 per cent of patients with chronic mitral regurgitation, and the chest radiograph shows cardiac enlargement and pulmonary venous congestion. As mentioned earlier, two-dimensional echocardiography provides excellent dynamic images of the mitral valve apparatus and allows the cause of the mitral incompetence to be evaluated, whether it be due to a chordal rupture or elongation, flail leaflets, prolapsing leaflets, calcification, vegetations, or papillary muscle rupture. Doppler echocardiography can be used to determine the severity of mitral regurgitation. Coronary angiography is essential in patients with angina or ventricular dilatation, and in those at risk for coronary arterial disease, as well as to confirm estimates of left ventricular function.

Indications for surgery

All patients with severe mitral regurgitation who are in New York Heart Association class 3, despite good medical management, will benefit from surgical treatment. Many of these patients have a slow, undramatic clinical course, which masks the silent but progressive left ventricular dysfunction. By the time the patient is severely symptomatic, left ventricular function can have deteriorated markedly, and the results of surgery are far from satisfactory. Patients who have severe mitral regurgitation and who are asymptomatic or minimally symptomatic should undergo noninvasive evaluation of systolic left ventricular function by two-dimensional echocardiography or radionuclide angiography: if the ejection fraction is below 55 per cent, surgery is recommended. The patient with severe mitral regurgitation, in whom echocardiography suggests that the valve can be easily repaired, will usually fare better when surgery is undertaken early, before severe symptoms develop, and the potential complications associated with mitral valve replacement can be avoided. The onset of atrial fibrillation is a controversial indication, but a recent study suggested that mitral valve repair be performed soon after the onset of atrial fibrillation to maximize the chance of postoperative sinus rhythm and to avoid the need for long-term anticoagulation.

Surgical approaches to the mitral valve

The left atrium can be exposed through a right thoracotomy, a left thoracotomy, a transverse sternotomy, or a median sternotomy. The approach through a right thoracotomy is particularly useful for the small left atrium, for reoperations following previous median sternotomies, and in the presence of patent coronary arterial bypass grafts. Left thoracotomy is useful for closed commissurotomy, but is used less today as valve repair gains in popularity. Many surgeons prefer a median sternotomy incision, which provides easy access to most cardiac structures with minimal tissue damage and postoperative pain. Cannulation for cardiopulmonary bypass is accomplished through a single aortic cannula and venous drainage is obtained through cannulation of both cava. The standard approach to the mitral valve is through a left atriotomy, posterior to the intra-atrial groove. If exposure of the mitral valve is not adequate through this incision, the left atriotomy can be extended superiorly and inferiorly around the orifices of the pulmonary veins, the atrium being retracted anteriorly and to the left. Division of the pericardial attachment to the superior and inferior vena cava also permits greater mobility of the heart during anterior retraction. Other approaches include the superior approach through the left atrium, between the superior vena cava laterally and the ascending aorta medially. When a left ventriculotomy is required for resection of a ventricular aneurysm, the mitral valve can also be easily replaced through this incision. In special cases, where the left atrium is particularly small, further exposure can be obtained by the biatrial transseptal incision of Dubost.

Repair of the mitral valve

Every effort should be made to repair the mitral valve: this avoids the problems associated with prostheses and leaves an intact mitral valve apparatus, which is important for the satisfactory functioning of the left ventricle. Clinical trials have shown that the results of mitral valve repair are superior to those of mitral valve replace-ment. The recent development of transesophageal echocardiography and color Doppler has aided the intraoperative assessment ofthe mitral valve. These techniques allow the function of the repaired mitral valve to be assessed in the operating room. The importance of proper exposure of the mitral valve during surgery cannot be overemphasized: special retractors have been developed by Carpentier and Cosgrove to provide the exposure that is needed. The choice of operative procedure must be based on direct inspection of the mitral valve apparatus, and the decision made based on the underlying pathology and the time required for the procedure when other cardiac procedures are necessary.

Commissurotomy

Closed commissurotomy or digitally dilating a stenotic mitral valve through the left atrial appendage was an important technique in the early days of cardiac surgery, but with the development of safe cardiopulmonary bypass, many surgeons now prefer open mitral commissurotomy or repair under direct vision. Use of the closed technique is generally restricted to young patients with mobile, noncalcified valves or those with severe stenosis without involvement of the subvalvular apparatus and with no evidence of thrombus, in whom closed commissurotomy by an experienced surgeon can give excellent results.

Open mitral commissurotomy begins with careful inspection of the left atrium and the mitral valve apparatus. If a left atrial thrombus is found it should be removed; the left atrial appendage should be ligated or oversewn as part of the operative procedure. Both commissures should be incised to approximately 4 mm of the mitral valve annulus. In general, patients with more advanced disease, including extensive thickening, shortening, and fusion of the chordae tendineae as well as extensive calcification, would be better served by mitral valve replacement. The operative mortality for mitral commissurotomy is less than 1 per cent. Under normal circumstances, patients may or may not be anticoagulated for the first 3 months after surgery, depending upon the surgeon's preference. However, if a thrombus is found in the left atrium, if the left atrium is extremely large or calcified, or if the patient is in atrial fibrillation, permanent anticoagulation should be instituted. The Maze procedure may be performed in selected cases at the same time as the mitral valve repair in an attempt to restore sinus rhythm.

Reconstructive surgery

Mitral valve reconstruction requires a sound knowledge of the functional anatomy of the mitral valve, and each component must be examined individually as the pathology is assessed. The mitral valve must be observed through the opened left atrium while the left ventricle is filled with blood or some other fluid. Injection of saline into the ventricle under pressure will cause the left ventricle to fill and allow the mitral valve leaflets to coapt, thereby allowing physiologic assessment of the underlying pathology. Mitral valve reconstructive techniques include extensive quadrangular resection for prolapse of the posterior leaflet caused by chordal rupture or elongation (Fig. 6) with resuturing of the remaining edges of posterior leaflet; and chordal replacement with polytetrafluoroethylene suture, triangular anterior segmental resection or transposition of a quadrangular section of posterior leaflet to anterior leaflet for prolapse of the anterior leaflet caused by chordal rupture or elongation (Fig. 7). These procedures should all be accompanied by a ring annuloplasty, as performed for patients with cardiomyopathy and annular dilatation. The size of the ring is determined by appropriate sizing of the anterior leaflet. The adequacy of mitral valve repair can be assessed by injecting cold saline through the mitral valve and watching for coaptation of the mitral leaflets, or by transesophageal echocardiography performed as the patient is being weaned from cardiopulmonary bypass. The results of mitral valve reconstruction are satisfactory in terms of both mortality and long-term survival.

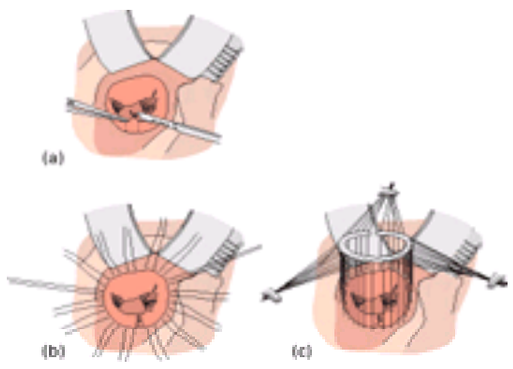


Fig. 6. A quadrilateral resection of the posterior leaflet is used to correct ruptured posterior chordae.

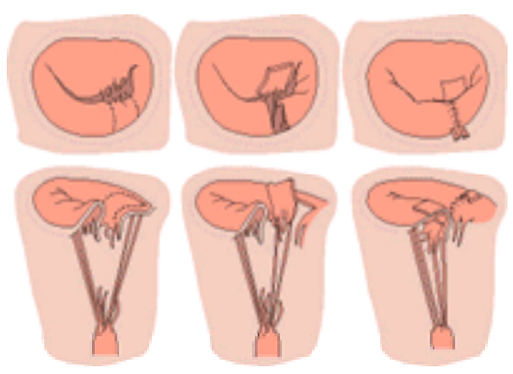


Fig. 7. When rupture of chordae to anterior leaflet occurs, a quadrilateral resection of the posterior leaflet is carried out and the intact secondary chordae are transferred to the anterior leaflet.

Cryopreserved homograft replacement of the mitral valve (entire mitral valve apparatus) is another option for reconstructive surgery in patients with acute endocarditis and severe rheumatic stenosis. All homografts are obtained from explanted hearts not used for transplantation and cryopreserved at -160°C without antibiotics. A recent paper reported on 43 patients who underwent homograft mitral valve replacement and none of the mitral valves was felt to be repairable by present techniques. Implant techniques included insertion of the homograft papillary muscle into a slit between the native papillary muscle and the wall of the left ventricle; homograft leaflet tissue was then sutured circumferentially to the mitral annulus and a ring annuloplasty was performed. Follow-up showed less than mild regurgitation in 38 patients by echocardiography and only two hospital deaths.

Cryopreserved homograft replacement of the mitral valve has several advantages because it is viable tissue, it does not require anticoagulation, and because the entire mitral valve apparatus is replaced, ventricular function should be preserved.

A tricuspid autograft (portion of the patient's own tricuspid valve) for mitral valve repair has been used in six patients from a group in Australia. The technique is based on the experience that the tricuspid valve can function very well as a bicuspid valve. Transferring a portion of the patient's own tricuspid valve to the mitral valve allows repairs of difficult lesions such as commissures destroyed by endocarditis or severe prolapse of the posterior leaflet due to multiple ruptured chordae. The selected tricuspid tissue (posterior leaflet and part of the anterior leaflet) is removed *en bloc* with leaflet, chordae, and papillary muscle. All patients survived and postoperative echocardiography showed no mitral regurgitation at the site of valve repair.

Mitral valve replacement

Mitral valve replacement carries a somewhat greater chance of operative mortality than does mitral valve repair because many of the patients undergoing this operation are old and have associated coronary arterial disease and depressed left ventricular function. Preservation of the posterior leaflet and the chordae tendineae has been shown to improve left ventricular function postoperatively, and hopefully prevents the devastating complications of left ventricular rupture, which occurs in a small number of patients, especially those who have severe calcification of the mitral valve as well as the subvalvular apparatus. Although artificial heart valves have been available for 40 years, there is still no ideal prosthesis, and within rough guidelines applicable to patient age and prosthetic reliability, the choice of valve is generally that of the surgeon. We prefer the St Jude bileaflet mechanical valve (Fig. 8) or the Carpentier–Edwards porcine bioprosthesis (Fig. 9). Both valves function well in the mitral position, have a low incidence of thromboemboli, and have good hemodynamics. The bioprosthetic valve is subject to structural degeneration and, as a rule, requires replacement after 5 to 10 years, especially in patients under the age of 60 years. The technique of mitral valve replacement is reasonably standardized, with most surgeons using interrupted, pledgeted mattress sutures around the mitral annulus (Fig. 10). It is extremely important that the sutures do not loop around the struts of a bioprosthesis and that redundant leaflet tissue does not interfere with the opening of the low-profile St Jude valve. As with all other operations on the left side of the heart, appropriate techniques to remove air must be employed at the end of cardiopulmonary bypass. All patients with mechanical valves should be anticoagulated with coumadin, as should those with biological valves and a large left atrium, atrial afibrillation, or a history of thromboembolism. The results of mitral valve replacement continue to improve owing to earlier surgery, better techniques of myocardial preservation, and concomitant repair of associated tricuspid valve and coronary arterial disease. Patients with prosthetic mitral valve dysfunction should be operated upon electively once the diagnosis is made: waiting for symptoms to develop is associated with rapidly worsening dysfunction, deterioration of cardiac function, and a marked increase in operative mortality.



Fig. 8. St Jude Medical mitral prosthesis.

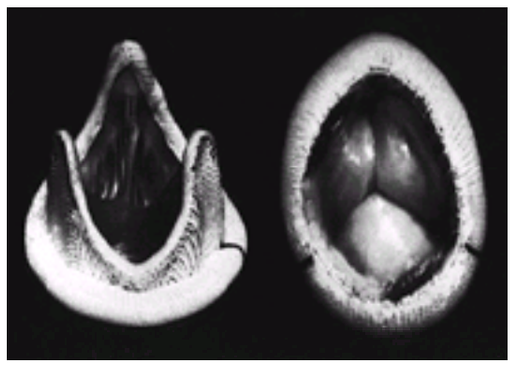


Fig. 9. Carpentier–Edwards porcine mitral prosthesis.

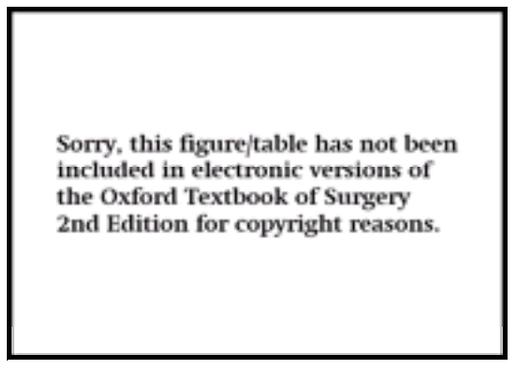


Fig. 10. Mitral valve replacement with everting interrupted mattress sutures: (a) and (b) valve excised; (c) and (d) alternate-colored, everting mattress suture with Teflon pledgets used; (c) sutures cut and tied; (f) and (g) holder removed.

Minimally invasive mitral valve surgery

Minimally invasive mitral valve surgery involves the use of smaller incisions and in some cases the videoscope, which can be introduced directly into the left atrium through a minithoracotomy while the patient is on cardiopulmonary bypass. The videoscope provides excellent close-up views of the valve projected on a monitor and allows one to videotape the operative procedure. Several approaches have recently been popularized and are summarized in Table 1. Cosgrove uses an 8-cm skin incision and an upper sternal split, which is extended into the right fourth inner space (Fig. 11). Vacuum-assisted venous drainage allows the use of smaller, less obstructive cannulas, but allows the heart to be adequately decompressed (Fig. 12). Current contraindications to this approach include severe pectus excavatum, reoperations, and a history of pericarditis. The mitral valve is approached via the intra-atrial septum. Cohn uses a 6- to 8-cm, right parasternal skin incision over the second to fourth costal cartilages (Fig. 13). Contraindications to this approach include deformities of the chest wall and extreme obesity. The mitral valve is also approached via the intra-atrial septum.

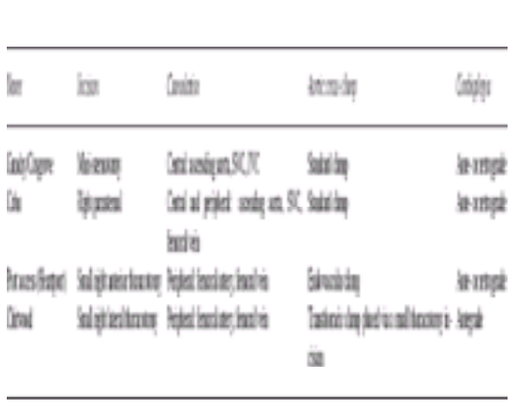


Table 1 Minimally invasive mitral valve surgery

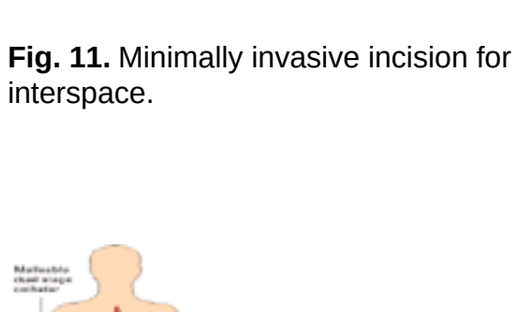


Fig. 11. Minimally invasive incision for mitral valve surgery popularized by Cosgrove; the incision involves an upper sternal split extended into the right fourth interspace.

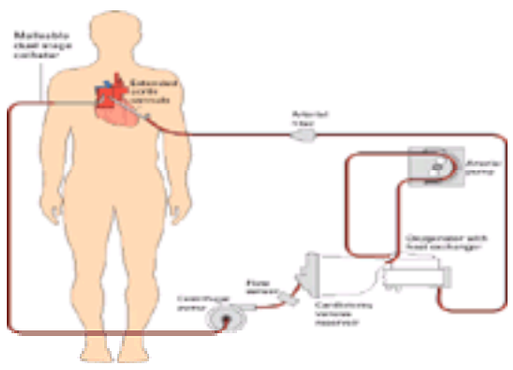


Fig. 12. Minimally invasive mitral valve surgery performed through small incisions requires the use of smaller cannulas to allow the field to be less obstructed. Improved venous drainage can be accomplished with these smaller cannulas by employing active negative pressure (vacuum-assisted) or the addition of a Biomedicus pump (by

courtesy of 3M Company.).

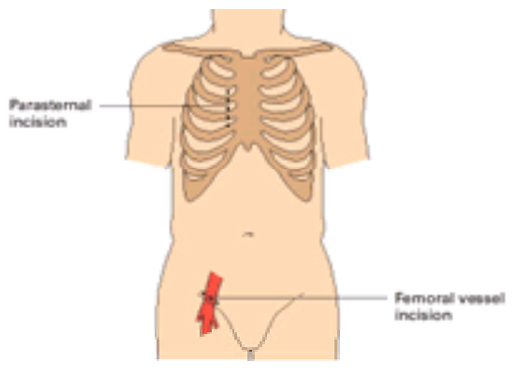


Fig. 13. Right parasternal incision used by Cohn for minimally invasive mitral valve surgery; the incision is made over the second to the fourth costal cartilages, to the right of the midline.

The port access or Heartport approach to minimally invasive mitral valve surgery uses a 5- to 7-cm, right fourth anterior minithoracotomy incision (Fig. 14). Cannulation for cardiopulmonary bypass involves the femoral artery and vein. The technique uses an endoaortic clamp to occlude the ascending aorta for delivery of antegrade cardioplegia as well as to vent the aortic root. In addition, retrograde cardioplegia can be delivered via a coronary sinus catheter placed percutaneously and positioned by fluoroscopy or transesophageal echocardiography. Also, an endopulmonary vent is placed percutaneously and positioned in the main pulmonary artery by fluoroscopy or transesophageal echocardiography. Contraindications to this approach include severe occlusive peripheral vascular disease or severe atheromatous disease of the thoracic aorta. The mitral valve is approached through an incision in the left atrium posterior to the intra-atrial groove. The technique also requires the use of specially modified surgical instruments since the surgeon is unable to get his or her entire hand into the operative incision. A variation of the Heartport approach has been described by Chitwood, where a small right lateral thoracotomy incision, usually in the fourth intercostal space, and a thoracoscope to improve exposure of the mitral valve, are used. In addition, he has developed a special transthoracic aortic cross-clamp, which is placed, via a small right thoracotomy incision (usually in the third interspace) and avoids the endovascular clamp as used in the Heartport technique.

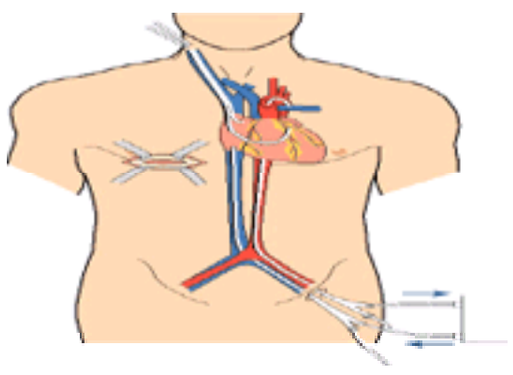


Fig. 14. Port access (Heartport) approach for minimally invasive mitral valve surgery. This technique uses endoaortic clamping, retrograde cardioplegia delivery via a coronary sinus catheter, and a cardiac vent in the pulmonary artery, all positioned fluoroscopically or by transesophageal echocardiography.

Experience with these different approaches has been limited to the pioneers in the field. These techniques appear to decrease bleeding and infection, reduce pain and discomfort, improve return of normal activity, lower the cost of the procedure, lower the length of stay, and, in addition, the incidence of atrial fibrillation is less. Postoperative impairment of pulmonary function also appears to be much less. Most of these surgeons advocate the use of transesophageal echocardiography to evaluate the repair or replacement of the valve, as well as to evacuate air at the end of the operative procedure. Further clinical studies are needed to show whether these techniques truly represent a further improvement of the care of our patients with mitral valve disease.

Further reading

Abe T *et al.* DeVega's annuloplasty for acquired tricuspid disease. *Annals of Thoracic Surgery* 1989; **48**: 670–6.

Acar C *et al.* Homograft replacement of the mitral valve: graft selection, technique of implantation, and results in forty-three patients. *Journal of Thoracic and Cardiovascular Surgery* 1996; **111**: 281–2.

Arbula A, Asfaw I. Tricuspid valvulectomy without prosthetic replacement. *Journal of Thoracic and Cardiovascular Surgery* 1981; **82:** 684–91.

Assey M-E, Spann JE. Indications for heart valve replacement. *Clinical Cardiology* 1990; **13**: 81–8.

Balasundaram SG, Duran C. Surgical approaches to the mitral valve. *Journal of Cardiac Surgery* 1990; **5**: 163–9.

Bolling SL *et al.* Intermediate-term outcome of mitral regurgitation in cardiomyopathy. *Journal of Thoracic and Cardiovascular Surgery* 1998; **115**: 381–6.

Carpentier A *et al.* Reconstructive surgery of mitral valve incompetence. Ten year appraisal. *Journal of Thoracic and Cardiovascular Surgery* 1980; **79**: 338–48.

Cerfolio RJ, Orszulak TA, Pluth JR, Harmsen WS, Schaff HV. Reoperation after valve repair for mitral regurgitation: early and intermediate results. *Journal of Thoracic and Cardiovascular Surgery* 1996; **111**: 117–84.

Cheitlin MD. The timing of surgery in mitral and aortic valve disease. *Current Problems in Cardiology* 1987; **12**: 75–149.

Chua YL, Schaff HV, Orzulak TA, Morris JJ. Outcome of mitral valve repair in patients with pre-operative atrial fibrillation: should the maze procedure by combined with mitral valvuloplasty? *Journal of Thoracic and Cardiovascular Surgery* 1994; **107**: 408–15.

Cobanoglu A, Sturr A. Tricuspid valve surgery. In Frankl WS, Brest AN, ed. *Valvular heart disease—comprehensive evaluation and management*, p. 382. Frankl FA Davis, Philadelphia, 1986.

Cohn LH. Mitral valve surgery–direct access. Paper presented at 78th Annual Meeting of American Association of Thoracic Surgeons, Boston, MA, May 1998.

Cohn LH, Kowalker W, Bhatia S, DiSesa VI, John-Sutton M, Shemin RJ, Collins JJ. Comparative morbidity of mitral valve repair versus replacement for mitral regurgitation with and without coronary artery disease. *Annals of Thoracic Surgery* 1995; **60**: 1452–3.

Cosgrove DM, Stewart WJ. Mitral valvuloplasty. *Current Problems in Cardiology* 1989; **14**: 359–415.

Cosgrove DM *et al.* Partial sternotomy for mitral valve operations. *Operative Techniques in Cardiac and Thoracic Surgery* 1998; **3**: 62–72.

Danielson GA. Ebstein's anomaly. In: Bane AE, ed. *Glenn's thoracic and cardiovascular surgery*, p. 1277. Appleton Lange, Norwalk, 1991.

David TE. A rational approach to the surgical treatment of mitral valve disease. In: Karp REB, Laks H, Wechsier A, ed. *Advances in cardiac surgery*, pp. 63–84. Mosby Year Book, Chicago, 1991.

David TE, Armstrong S, Sun Z. Replacement of chordae tendineae with Gore-Tex sutures: a ten-year experience. *Journal of Heart Valve Diseases* 1996; **11**: 136–45.

Duran CM. Acquired disease of the mitral valve. In: Bane AE, ed. *Glenn's thoracic and cardiovascular surgery*, p. 1692. Appleton and Lange, Norwalk, 1991.

Grossi EA *et al.* Port access approach for minimally invasive mitral valve surgery. *Operative Techniques in Cardiac and Thoracic Surgery* 1998; **3**: 32–46.

Hammond GL, Geha AS, Kopf GS, Hashim SW. Biological vs. mechanical valves: analyses of 1116 valves inserted in 1012 adult patients with a 4818 patient-year and 5327 valve-year follow-up. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 182–98.

Hauck AJ *et al.* Surgical pathology of the tricuspid valve. *Mayo Clinic Proceedings* 1988; **63**: 851–63.

Hetzer R *et al.* Status of patients presently living 9 to 13 years after orthotopic heart transplantation. *Annals of Thoracic Surgery* 1997; **64**: 1661–8.

Hughes CF, Noble N. Vegectomy: an alternative surgical treatment for infective endocarditis of the atrioventricular valve in drug addicts. *Journal of Thoracic and Cardiovascular Surgery* 1988; **95:** 857–61.

Ilbawi MN *et al.* Experience with St. Jude medical valve prosthesis in children. *Journal of Thoracic and Cardiovascular Surgery* 1987; **93**: 73–9.

Karlson KJ, Ashraf MM, Berger RL. Rupture of left ventricle following mitral valve replacement. *Annals of Thoracic Surgery* 1988; **46**: 590–7.

Kay JH, Mendez AM, Zubiate P. A further look at tricuspid annuloplasty. *Annals of Thoracic Surgery* 1976; **22**: 498–500.

Lange R *et al.* Tricuspid valve reconstruction—a treatment option in acute endocarditis. *European Journal of Cardio-Thoracic Surgery* 1996; **10**: 320–6.

Liao K, Frater RWM, LaPietra A, Ciuffo G, Ilardi CF, Seifter E. Time-dependent effect of glutaraldehyde on the tendency to calcify of both autografts and xenografts. *Annals of Thoracic Surgery* 1995; **60** (Suppl.): 343–7.

Lundin L *et al.* Surgical treatment of carcinoid heart disease. *Journal of Thoracic and Cardiovascular Surgery* 1990; **100**: 552–61.

McGrath LB *et al.* Tricuspid valve operations in 530 patients. *Journal of Thoracic and Cardiovascular Surgery* 1990; **99**: 124–33.

Miki, S *et al.* Mitral valve replacement with preservations of chordae tendineae and papillary muscles. *Annals of Thoracic Surgery* 1988; **45**: 28–34.

Mintz GS *et al.* Two dimensional echocardiographic evaluation of patients with mitral insufficiency. *American Journal of Cardiology* 1979; **44**: 670–8.

Miyamura H, Kanazawa H, Hayashi J, Eguchi S. Thombosed St. Jude medical valve prosthesis in the right side of the heart in patients with tetralogy of Fallot. *Journal of Cardiovascular Surgery* 1987; **94**: 148–56.

Munro, AI, Jamieson WRE, Tyers GFO, Germann E. Tricuspid valve replacement: porcine bioprostheses and mechanical prostheses. *Annals of Thoracic Surgery* 1995; **60** (Suppl.): 470–4.

Olson LI *et al.* Surgical pathology of mitral valve. *Mayo Clinic Proceedings* 1987; **62**: 22–34.

O'Rourke RA, Crawford MH. Mitral valve regurgitation. *Current Problems in Cardiology* 1984; **9**: 8–52.

Perfloff JK, Roberts WC. The mitral apparatus: functional anatomy of mitral regurgitation. *Circulation* 1972; **46**: 227–39.

Perrier P, Clausnizer B, Mistarzk K. Carpentier ‘sliding leaflet' technique for repair of the mitral valve: early results. *Annals of Thoracic Surgery* 1994; **57**: 383–6.

Poveda JJ *et al.* Tricuspid valve replacement in rheumatic disease—preoperative predictors of hospital mortality. *Journal of Heart Valve Disease* 1996; **5**: 26–30.

Pupello DF *et al.* The Carpentier Edwards bioprosthesis: a comparative study analysing failure rates by age. *Journal of Cardiac Surgery* 1988; **3**: 369–74.

Ramsheyi A *et al.* Partial mitral homograft for tricuspid valve repair. *Annals of Thoracic Surgery* 1997; **64**: 1486–8.

Ren JF, Aksut S, Lighty GW, Vigilante GJ, Sink JD, Scogal BL, Hargrove WC. Mitral valve repair is superior to valve replacement for early preservation of cardiac function: relation of ventricular geometry to function. *American Heart Journal* 1996; **131**: 974–81.

Shimada R *et al.* Diagnosis of tricuspid stenosis by M-mode and two-dimensional echocardiography. *American Journal of Cardiology* 1984; **53**: 164–8.

Singh AK *et al.* Follow-up assessment of St. Jude medical prosthetic valve in the tricuspid position: clinical and hemodynamic results. *Annals of Thoracic Surgery* 1984; **37**: 324–7.

Smedira NG *et al.* Repair of anterior leaflet prolapse: chordal transfer is superior to chordal shortening. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 287–91.

Spencer FC *et al.* Recent developments and evolving techniques of mitral valve reconstruction. *Annals of Thoracic Surgery* 1998; **65**: 307–13.

Steimle CN *et al.* Outcome of re-operative valve surgery via right thoracotomy. *Circulation* 1996; **94**: 126–8.

Ubago JL *et al.* Right ventriculography as a valid method for the diagnosis of tricuspid insufficiency. *Catheterization in Cardiovascular Diagnosis* 1981; **7**: 433–41.

Vander Salm TJ, Pape LA, Mauser JF. Mitral valve replacement with complete retention of native leaflets. *Annals of Thoracic Surgery* 1995; **59**: 52–5.

Van Nooten GJ *et al.* Tricuspid valve replacement—postoperative and long-term results. *Journal of Thoracic and Cardiovascular Surgery* 1995; **110**: 672–9.

Yee ES, Ullyot DJ. Reparative approach for right-sided endocarditis—operative considerations and results of valvuloplasty. *Journal of Cardiovascular Surgery* 1988; **96**: 133.

Zussa C, Polesel E, DaCol U, Galloni M, Valfre C. Seven-year experience with chordal replacement with expanded polytetrafluoro-ethylene in floppy mitral valve. *Journal of Thoracic and Cardiovascular Surgery* 1994; **108**: 37–41.

40.10 Surgery of the thoracic aorta

Stephen Westaby

[Historical background](#)
[Surgical anatomy of the thoracic aorta](#)
[Spinal cord blood supply](#)
[Pathology of the thoracic aorta](#)
- [Medial disease](#)
- [Atherosclerosis](#)

[Imaging technique for thoracic aortic disease](#)
[Perfusion techniques in thoracic aortic surgery](#)
- [Retrograde cerebral perfusion](#)

[Vascular prostheses](#)
[Topical haemostatic agents](#)
[Pharmacological methods to reduce perioperative bleeding](#)
[Surgery of aortic dissection](#)
- [Operative technique for type A dissection](#)
- [Type B dissection](#)
- [Descending thoracic and thoracoabdominal aneurysms](#)
- [Coarctation of the aorta in adults](#)
- [Monitoring of spinal cord function](#)
- [The thromboexclusion technique](#)
- [Aortoannular ectasia](#)
- [Aneurysms of the ascending aorta](#)
- [Aortic arch aneurysms](#)

[Further reading](#)

The history of thoracic aortic surgery is short. Operations on the aortic root and the ascending and transverse aorta were only made possible by the development of cardiopulmonary bypass in the 1950s. As in other areas of cardiovascular surgery, techniques in thoracic aortic replacement continue to improve. However, whereas surgical treatment of complex congenital cardiac anomalies, acquired valvular defects, and coronary occlusive disease has achieved very low morbidity and mortality, operations on the thoracic aorta continue to carry significant risk. Factors contributing to surgical failure include delay in diagnosis and treatment, lack of familiarity with surgical methods due to relatively low patient numbers, and failure to define the optimal approach to individual problems. Causes of death or morbidity include haemorrhage, renal and pulmonary dysfunction, mesenteric vascular occlusion, and ischaemia of the spinal cord.

Although thoracic aneurysms show a greater tendency to rupture than do abdominal aortic aneurysms, the threat of perioperative stroke, paraplegia, renal failure, or gut necrosis militates against early surgery in many centres. Inevitably, an increasingly aged population will produce more patients with aneurysm due to medial or atheromatous degenerative disease. Improved techniques are consequently needed to benefit an increasing number of patients with type A or type B dissection, and those with ascending, arch, descending thoracic, or thoracoabdominal aneurysms. Diagnostic methods, including computed tomography (**CT**) and nuclear magnetic resonance imaging (**MRI**), have already advanced the non-invasive assessment and monitoring of thoracic aortic disease. Nevertheless, patients with thoracic aneurysms are at risk from acute coronary events during aortic cross-clamping and the need for coronary angiography is not yet obsolete.

Historical background

The possibility of suturing of blood vessels was conceived more than two centuries ago, when an injured artery in the arm of a patient was successfully repaired. The concept lay dormant for more than a century until the impetus given to experimental surgery by the Listerian era. Czerny reported successful suture of an internal jugular vein and Postempski repaired a femoral arterial wound in the 1880s.

In 1881, Matas described the technique of endoaneurysmorrhaphy; in the early 1900s, experimental surgery demonstrated the feasibility of excising arterial segments and restoring continuity by end-to-end anastomosis or arterial grafts. The feasibility of vascular anastomosis led to early experiments with renal and cardiac transplantation. However, except for the individual efforts of a few surgeons, little consideration was given to the application of these principles to patients with arterial disease. This is perhaps understandable when one considers that ancillary measures such as general anaesthesia, blood transfusion, and antibiotics had not been developed, and that the technique of arteriography has not been sufficiently refined for general and safe clinical use.

Roentgen announced the monumental discovery of X-rays in 1895; within a year the feasibility of radiographic visualization of blood vessels had been demonstrated by injecting contrast into the vessels of an amputated hand. In 1929, Swick reported the successful radiological use of organic iodide solutions. Thoracic aortic surgery began in 1944, with successful excision and end-to-end anastomosis in the repair of coarctation. A few years later, human homograft tissue was first used to bridge the aortic defect resulting from excision of coarctation. In 1951, Oudot reported excision and homograft replacement for occlusive disease of the lower abdominal aorta. The following year, Dubost performed the first resection and homograft replacement of an abdominal aortic aneurysm.

The use of homograft was a considerable advance in vascular surgery, but access to such material was limited and methods of sterilization and preservation inadequate. Observation of treated patients showed that tissue elements of the graft deteriorated, calcified, and often became aneurysmal. The need for more suitable arterial substitutes became increasingly apparent. In the meantime, dos Santos observed that atheromatous occlusive disease in the femoral artery may be localized and accessible to thromboendarterectomy. The obstructive lesion could be peeled away inside the remaining wall of the artery by finding the appropriate plane of cleavage, thus restoring peripheral circulation in some patients without the need for graft replacement. An alternative approach was devised by Kunlin, who reasoned that it should be possible to support nature's collateralization of vascular obstruction by creating a vein bypass graft from above to below the lesion. The use of venous tissue for arterial reconstruction was extended by the technique of vein patch angioplasty. Although experiments with patch angioplasty had been performed as early as 1906, the procedure was thought to have limited clinical application. However, clinical experience with endarterectomy of small vessels such as the carotid, femoral, and popliteal arteries showed that luminal constriction resulting from longitudinal closure of a vascular wound could be avoided by repairing the defect with a vein patch.

Dubost's initial success with homograft replacement of the abdominal aorta was repeated by De Bakey and Cooley in 1953; this stimulated interest in the replacement of descending thoracic aneurysms. The first patient to undergo this type of surgery was a 46-year-old man who had a large fusiform aneurysm of the lower descending thoracic aorta; this was replaced via a left thoracoabdominal approach, re-establishing aortic continuity with a homograft. Ten years later the patient returned with bronchogenic carcinoma of the left lung and underwent left pneumonectomy; the aortic graft was seen to be in good condition.

Apart from saccular aneurysms of the ascending aorta and aortic arch, where tangential excision with lateral aortography may be applied, ascending aortic surgery requires cardiopulmonary bypass. The first successful resection of the ascending aorta was performed in 1956 on a 50-year-old man, a remarkable achievement considering that closure of an atrial septal defect on cardiopulmonary bypass had only been achieved the previous year. The aneurysm in this patient was extensive, originating just above the coronary ostia and involving the entire ascending aorta to the origin of the innominate artery. The distal occluding clamp was applied across the origin of this vessel, into which a small perfusion cannula was inserted. The ascending aorta was replaced with an aortic homograft and the patient remained well until his death from lung cancer 14 years later. Homograft replacement of the aortic arch followed in 1957, using separate perfusion of the head vessels.

Initial successes provided substantial support for the growing conviction that excision and graft replacement was the optimal therapy for aneurysmal disease; this was the stimulus to explore technical resources to provide a synthetic vascular graft. One of the most significant developments was the observation that a fabric woven of Vinyon ‘N' thread could function satisfactorily as an aortic substitute. Attention was then directed towards fabrics such as nylon, Orlon, Teflon, Ivalon, and Dacron, which were fashioned into tubes by sewing, braiding, heat sealing, knitting, and weaving. The Baylor group produced grafts from all these materials by sewing the edges of two sheets of fabric together and implanted them into laboratory animals: Dacron proved to be the best material.

In 1954, De Bakey implanted a Dacron aortic bifurcation graft made of two Y-shaped sheets of cloth with their edges machine-sewn together. The 54-year-old patient made an uneventful recovery and died from myocardial infarction 10 years later. A new knitting machine was designed to produce seamless Dacron tubes and bifurcation grafts. This work ultimately led to a highly satisfactory and effective arterial substitute for clinical use.

Surgical anatomy of the thoracic aorta

From a surgical standpoint, the aorta is divided into ascending, transverse, descending, thoracoabdominal, and abdominal parts (Fig. 1). The ascending aorta begins at

the aortic valve and ends at the origin of the innominate artery. Immediately above the aortic valve are the three aortic sinuses, which correspond in position to the anteriorly situated right coronary cusp and the posteriorly situated left and non-coronary cusps (Fig. 2). From the left sinus arises the left main coronary artery and from the right sinus the right coronary artery. These arteries may have anomalous origins. For example, the right coronary may arise from the left coronary sinus taking an oblique passage through the wall of the aorta. There may be separate orifices for the left anterior descending and circumflex coronaries from the left coronary sinus. Similarly, it is not unusual for there to be a separate ostium for the conus branch of the right coronary artery. The left coronary artery, and very rarely the right or both coronary arteries, may have anomalous origin from the pulmonary artery.

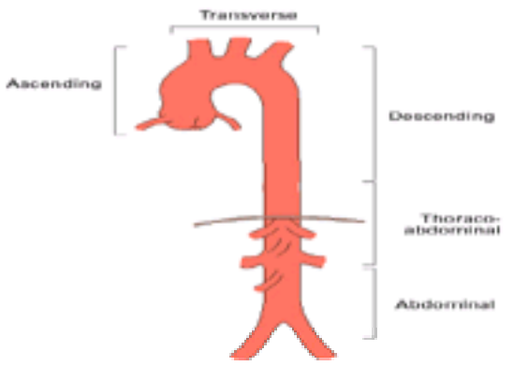


Fig. 1. Surgical anatomy of the thoracic and thoracoabdominal aorta.



Fig. 2. Normal aortogram showing the aortic sinuses, coronary arteries, and major branches of the aortic arch (investigation performed for suspected traumatic aortic laceration, which was ruled out by this picture).

Integrity of the aortic sinuses is important for normal function of the aortic valve. Loss of the aortic sinuses may occur acutely in type A aortic dissection or chronically in obliterative aortitis (e.g. Takyasu disease), resulting in aortic reflux. For descriptive purposes, in aortic surgery the left ventricular outflow tract, aortic valve, and aortic sinuses are known as the aortic root. Surgical techniques in this area depend on whether aneurysm formation involves the aortic sinuses and coronary ostia or begins above the aortic sinuses. In Marfan syndrome (Fig. 3), both the aortic annulus and the sinuses have the propensity to dilate, and surgery of the ascending aorta in patients with this condition should always include root replacement. Congenital anomalies in this area include left or right aortoventricular tunnel, aortopulmonary window, and truncus arteriosus. The ascending aorta may bifurcate to give dual aortic arches and a vascular ring. The transverse thoracic aorta begins at the root of the innominate artery and ends just after the left subclavian artery. This part of the aorta gives rise to the great vessels of the head and upper limbs. The usual sequence for these vessels is the innominate artery, then the left common carotid artery, and last the left subclavian artery. The vessels arise sequentially as the transverse aorta passes from a central position behind the manubrium towards the left side of the vertebral column. The trachea and oesophagus are situated in the midline to the right of the transverse arch. The innominate artery passes anterior to the trachea before bifurcating into the right subclavian and right common carotid arteries at the thoracic inlet. Variations in the origin of these vessels are unusual. The right subclavian and right common carotid arteries may arise separately from the transverse aorta. The right subclavian artery may arise distally from the arch and take a retro-oesophageal path to the thoracic inlet on the right; this anomaly can produce symptomatic oesophageal compression. Rarely, the descending thoracic aorta lies to the right side of the vertebral column in the thorax in association with intracardiac defects. Congenital anomalies of the great vessels may give rise to vascular rings, which cause compression of the trachea and oesophagus. They are often symptomatic in infancy or childhood.

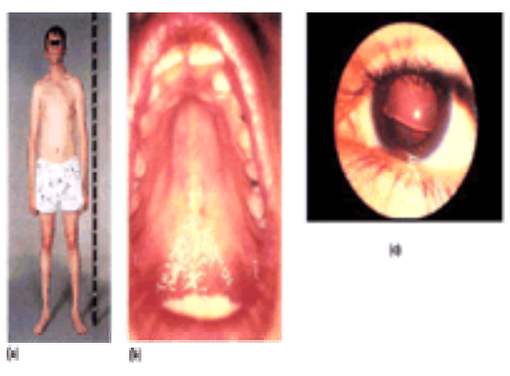


Fig. 3. Clinical features in Marfan syndrome: (a) body habitus,(b) high-arched palate, (c) dislocated lens.

The descending thoracic aorta begins just distal to the left subclavian artery where the aorta reaches the fourth thoracic vertebra. Anteriorly is the ligamentum arteriosum, the fibrous remnant of the ductus arteriosus. The ductus arteriosus rarely persists into childhood or adult life. The left recurrent laryngeal nerve arises from the vagus nerve lateral to the aorta and passes around the ligamentum arteriosum or patent ductus to take a cranial course adjacent to the left side of the trachea and oesophagus. Damage to this structure during surgery of patent ductus arteriosus or the descending thoracic aorta may cause permanent paralysis of the left vocal cord. The descending part arbitrarily meets the thoracoabdominal part at the lower end of the body of T10.

The descending thoracic aorta is fixed to the posterior wall of the chest by the segmental intercostal arteries; these in turn contribute to the vascular supply of the spinal cord (Fig. 4). Since the aorta meets the posterior chest wall at the level of T4, the upper intercostal vessels arise obliquely from the left and right sides of the aorta to reach their respective intercostal spaces. The junction of the transverse thoracic aorta with the fixed descending thoracic aorta has great strategic importance in traumatic injury. Deceleration road-traffic injuries produce shearing forces at this site that cause the aorta to tear or rupture (Fig. 5). The descending thoracic aorta gives rise to the bronchial circulation and to multiple branches to the oesophagus. Certain pathological states may alter the size and nature of the intercostal and bronchial arteries. In patients with cyanotic congenital heart disease and impaired pulmonary blood flow the bronchial arteries enlarge to form aortopulmonary collateral vessels, which can reach the size of the native subclavian artery. In coarctation of the aorta (usually situated at the level of the ligamentum arteriosum just distal to the left subclavian artery), the intercostal circulation gives rise to multiple large collateral arteries in the chest wall that supply the distal aorta. In adult life these collaterals may become aneurysmal and may cause serious bleeding during thoracotomy.

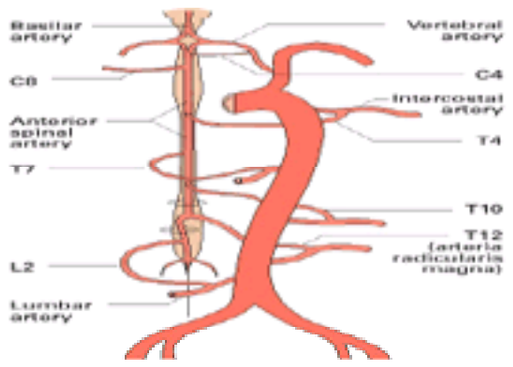


Fig. 4. Spinal cord blood supply.

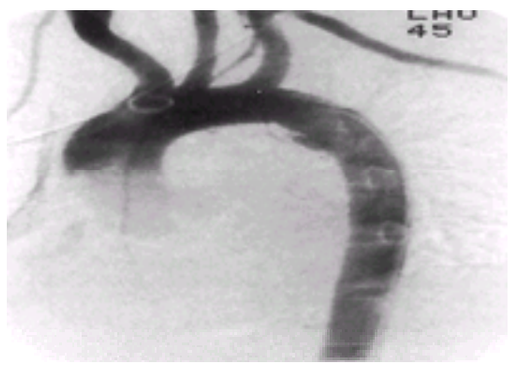


Fig. 5. Traumatic aortic transection with false aneurysm contained by overlying mediastinal tissues; usual site distal to the left subclavian artery.

The thoracoabdominal part of the descending aorta begins at the tenth intravertebral disc and is situated adjacent to the diaphragm both proximal and distal to its hiatus. The thoracic portion supplies intercostal and oesophageal branches. The abdominal portion supplies lumbar, diaphragmatic, and major visceral branches, including the coeliac axis, the superior mesenteric artery, and renal and adrenal vessels. The great radicular artery of Adamkiewicz arises from a left intercostal artery between T9 and L2 in 85 per cent of individuals; it is usually visibly larger than corresponding branches. The number and importance of the branches of the thoracoabdominal aorta make its replacement complex.

Spinal cord blood supply

The blood supply to the spinal cord (Fig. 4) has great importance in surgery of the descending thoracic and thoracoabdominal aorta. Though the causes of surgical injury to the spinal cord are multifactorial, interruption of critical intercostal arteries plays an important part. One factor that clearly contributes to the varying incidence of neurological injury relates to the unpredictable anatomical variations of blood flow to the spinal cord. The majority of 62 radicular arteries present in the embryo ultimately regress; only two to five remain in the adult. The spinal cord has anterior and posterior spinal arteries running along its length. These vessels originate inside the skull as branches of the carotid circulation, but are fed by collateral arteries from the thoracic aorta. Seventy-five per cent of the blood supply to the spinal cord is derived from the anterior spinal artery, which often becomes extremely narrow in calibre and distribution in the mid and lower portions of the cord. This vessel is reinforced at various levels. In the neck the cervical radicular arteries arise as branches of the vertebral and subclavian arteries. In the chest the intercostals supply segmental branches, though the lower part of the spinal cord relies on the great radicular artery, which arises between T5 and T8 in 15 per cent of individuals, between T9 and T12 in 75 per cent, between L1 and L2 in 10 per cent, and from L3 in 1 per cent. The vessel is large enough to be located by selective angiography when planning thoracic aortic surgery. Whenever possible it should be reimplanted into a vascular graft.

Inadequate blood flow through the lower segment of the anterior spinal artery or permanent interruption of the segmental blood supply to the dorsolumbosacral region by ligation of the critical intercostals are likely to result in ischaemic injury to the cord. The dorsolumbosacral region extends from T8 to the conus terminalus of the spinal cord and generally derives its blood supply from the single radicular vessel of Adamkiewicz (arteria radicularis magna).

In a small proportion of individuals (approximately 15 per cent), the principal spinal radicular artery arises at a high level, between T5 and T8; this is generally associated with supplemental blood supply to the lower cord arising from additional vessels in the lower dorsolumbar region. In this case, division or ligation of the intercostal arteries above T8 is unlikely to result in paraplegia; cord injury may only result from prolonged proximal aortic cross-clamping or interruption of the critical lower intercostal vessels. The importance of strategic intercostal arteries should not be underestimated, since their interruption results in a high incidence of paraplegia. High resistance is a characteristic of any blood supply based on small collateral vessels. Fluid pressure in the cerebrospinal canal represents a second barrier to blood flow: if this is higher than capillary perfusion pressure, ischaemia can result despite patent and perfused, anterior spinal collateral vessels. Distal blood pressure high enough to overcome these resistances is necessary to ensure blood flow to the spinal cord and maintain its viability.

Pathology of the thoracic aorta

Coarctation of the aorta and patent ductus arteriosus are common anomalies amenable to surgical correction from the first few hours of life. Recently, innovative techniques such as balloon dilation of coarctation and percutaneous transvascular occlusion of patent ductus arteriosus have been developed, but as yet apply to small numbers of patients. Less common anomalies, such as ruptured aneurysm of the sinus of Valsalva, aortopulmonary window, and symptomatic vascular ring, usually warrant early surgical intervention. In adults, most thoracic aortic surgery is undertaken for aneurysm repair. Congenital aneurysms do occur, though they are extremely rare. Conditions classified as congenital aortic aneurysms are more correctly identified as acquired dilatations of the aorta due to a congenital structural abnormality in the vessel wall, or due to secondary deterioration of the aorta in association with other systemic disease processes. Post-stenotic dilation occurs in congenital aortic stenosis and coarctation of the aorta. Inherited disorders of connective tissue, such as Marfan syndrome, Ehlers–Danlos syndrome, or tuberous sclerosis, may lead to aneurysm formation in childhood, though this occurs more commonly in early adult life. Rarely, bacterial invasion of the aortic wall, granulomatous inflammation, and trauma are responsible for aortic aneurysm in children. Disorders such as Takayasu disease, syphilis, and cystic medial necrosis, which have been associated with the development of aortic aneurysms in adults, are rare causes in children.

Most acquired aneurysms (trauma excepted) present later in life. A true aneurysm always contains components of the aortic wall: aortic dissection is contained by the adventitia of the aorta and is therefore classified as a true aneurysm. A false aneurysm contains no component of the true aortic wall and occurs after penetrating injuries or traumatic transection, when haemorrhage is contained by overlying pleura and haematoma. Pathological processes that lead to the formation of thoracic aortic aneurysm are shown in Table 1. The most common of these are atherosclerosis, which affects predominantly those over the age of 60 years, and medial disorders, which usually require surgery at the age of 45 to 60. Clinically, medial disease is the more important, causing early cystic medial degeneration followed by aortoannular ectasia or acute aortic dissection.



Table 1 Pathology of the thoracic aorta

Medial disease

The media of the aorta has four components: smooth muscle, which provides tone; elastic fibres, which determine the distensibility of the vessel; collagen fibres, which give strength to the media; and a mucopolysaccharide connective-tissue base (Table 2). The collagen network, particularly types 1 and 3 collagen, are primarily responsible for the tensile strength of the aorta. When deficiencies exist in the construction of this collagen, as may occur in some varieties of Ehlers–Danlos syndrome and in some families with a history of familial aortic aneurysms, it is assumed that the genetic factor predisposes to an acquired dilation. Degeneration of the media occurs with age, predominantly affecting the elastic fibres and smooth-muscle cells. Medial cystic necrosis occurs as part of the ageing process but may be accelerated in certain types of disease (Fig. 6).

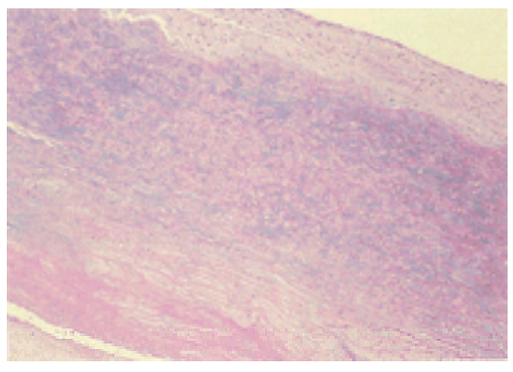


Fig. 6. Histological appearance of cystic medial degeneration.

Medial components	
Smooth muscle, provides tone	
Elastic fibres, determine distensibility	
Collagen fibres, give strength	
Mucopolysaccharide, connective tissue	
Grades of medial degeneration	
(degeneration of elastic fibres and smooth muscle cells)	
I	0–10%
II	10–25%
III	25–50%
IV	75%
Severe (grade IV) degeneration is found in:	
21% of type A dissections	
10% of type B dissections	
43% of Marfan patients	

Table 2 Aortic medial disease

Special staining techniques allow medial degeneration to be graded (Table 2). Severe degeneration (grades 3 and 4) is found in 21 per cent of patients with type A acute aortic dissection, 10 per cent of patients with type B dissection, and 43 per cent of patients with Marfan syndrome, where the elastic component of the media suffers accelerated degeneration.

Marfan syndrome is a hereditary disorder of the connective tissue producing well-known ocular, skeletal, and cardiovascular manifestations (Fig. 3). Ninety-five per cent of patients with the syndrome have cardiovascular abnormalities, including fusiform aneurysm and dissection of the aorta, aneurysm of the aortic branches, and disease of aortic and mitral valves. The most common manifestations are dilation of the aortic root and mitral valve abnormalities. Aortic root dilation begins *in utero* and invariably progresses. Cardiac events are rare before the age of 15 years, but aneurysm formation and dissection, rupture, cardiac tamponade, and sudden death occur with increasing frequency later in life as the diameter of the aortic root increases to 5 or 6 cm (Fig. 7). Insufficiency of the aortic valve usually begins when the annulus becomes dilated to 6 cm. Dissection or rupture of the aorta tends to occur before cardiac failure from aortic insufficiency (Fig. 8; Table 3). The principal indication for operation is an increase in the diameter of the aortic root to more than 5 cm. However, in patients with a strong family history of aortic dissection in the third or fourth decades of life, elective replacement of the aortic root is considered before this time. b-Adrenergic blockade retards the rate of aortic dilatation and the risk of dissection by reducing wall stress. Consequently, all Marfan patients should be prescribed b-blockers from diagnosis, even in infancy. Prolapse of the mitral valve ultimately occurs in 80 per cent of patients with Marfan syndrome. The cardiovascular manifestations of the syndrome cause death in over one-third of patients by the age of 32 and in two-thirds by the age of 50 years.

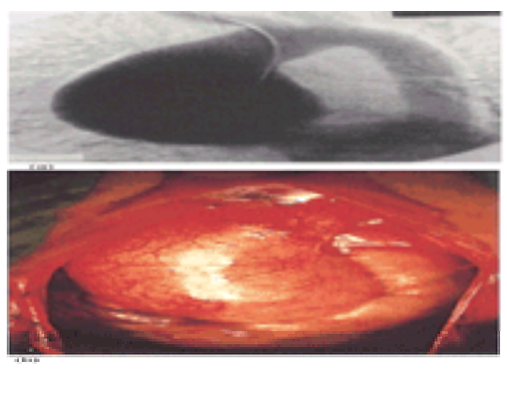


Fig. 7. Descending aortic aneurysm due to cystic medial necrosis:(a) aortogram; (b) appearance at operation.

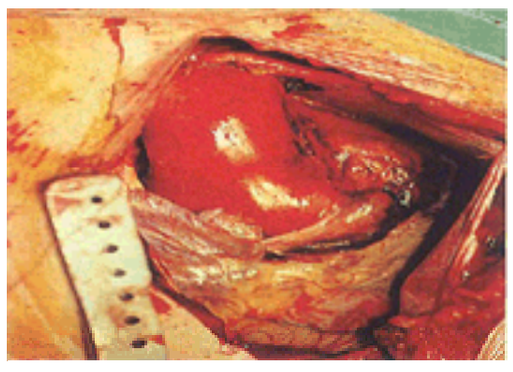


Fig. 8. Acute aortic dissection in a young patient with Marfan syndrome.

Deficiency of collagen and elastin in aortic wall
Increased circulating concentrations of collagenase and elastase
Decreased protease inhibitors
Altered protease/antiprotease activity

Table 3 The proteolysis theory of aneurysm formation

Functional changes in the Marfan aorta

Marfan syndrome occurs through variable expression of the fibrillin gene on chromosome 15q. Endothelial cells lay down fibrillin microfibrils in the vascular subendothelium. Mutant fibrillin disintegrates through an inability to bind calcium when the amino acid cysteine is replaced by other amino acids. Abnormal fibrillin causes a disturbed functional relation between blood flow and vascular endothelial-cell responses (mechanotransduction). In normal individuals, increased flow and pharmacological stimuli activate nitric oxide synthetase in the endothelial cell, causing vasodilatation. In Marfan syndrome the abnormal endothelial structural components cannot convert physical forces to nitric oxide production, resulting in failure of flow-mediated vasodilatation. Impaired mechanotransduction contributes to decreased arterial distensibility and increases in pulse-wave velocity reflect wave pressure and stress in the aortic wall; together these cause an increase in cardiac workload. Intravascular ultrasonographic measurements in the Marfan patient show changes in the diameter of the ascending aorta (from diastole to systole) to be significantly reduced with increased pulse pressure on the abnormal aortic wall during systole. Marfan patients consequently have both a structural and a functional propensity towards aortic dissection.

Atherosclerosis

Thoracic aortic aneurysms occur some three times more often in men than women. Important pathophysiological correlates in the formation of atherosclerotic aortic aneurysms are the ageing process, hypertension, smoking, hyperlipidaemia, and diabetes mellitus. The susceptibility of the aorta is variable: the thoracic aorta rarely develops obstructive atherosclerotic lesions whereas the abdominal aorta commonly develops clinically significant, stenotic and ulcerating plaques. A number of haemodynamic factors have been identified as important in the formation and localization of atherosclerotic plaques. These include wall shear stress, flow patterns and velocities, and flow disturbances. In the thoracic and the suprarenal abdominal aorta, the flow directional vector is forward throughout the cardiac cycle. The infrarenal abdominal aorta, on the other hand, experiences directional oscillation of the shear stress vector during the cardiac cycle, with forward and reverse flow. This flow characteristic may predispose the abdominal aorta and vessels of the lower extremities to increased plaque deposition. Atherosclerotic plaques preferentially form in regions of low not high wall shear stress. Low shear rates may retard the transport of atherogenic substances away from the arterial wall, resulting in an increased accumulation of lipids. The differing susceptibilities of the thoracic and abdominal aorta to atherosclerosis may also be due to differences in the architecture, composition, and nutrition of the arterial wall. The thoracic aorta is thicker and has a greater number of medial lamellar units than the abdominal aorta, reflecting its greater diameter and tangential wall tension. The thoracic aorta contains a larger proportion of elastin and a lower proportion of collagen than the abdominal aorta or peripheral vessels; this results in greater distensibility and pulse propagation in the thoracic segment. The outer two-thirds of the thoracic aortic media are well perfused by intramural vasa vasorum, whereas the inner lamellar units are nourished by diffusion from the lumen. In comparison, the abdominal aorta is nourished only from the lumen and lacks medial vasa vasorum; this may place the outer medial smooth-muscle cells in a relatively ischaemic zone and may alter metabolic function. Thus the composition and microarchitecture of the aortic media and the metabolic state of the medial smooth-muscle cells may be important factors in determining the susceptibility of different segments of the aorta to atherosclerosis. A number of haemodynamic factors act to inhibit plaque formation in the thoracic aorta, including high wall shear stress, short near-wall particle diffusion time, and unidirectional, non-oscillating flow.

Observations on the natural history of thoracic aortic aneurysm have led to the abnormal proteolysis theory of aneurysm formation (Table 3). Increased amounts of the enzyme elastase have been demonstrated in the aortic wall of patients with aneurysm and in the circulating leucocytes of smokers. The serum of smokers with abdominal aortic aneurysms contains decreased concentrations of natural protease inhibitors but unique elastolytic enzymes. This combination of factors is thought to lead to weakening of the medial component of the aortic wall and aneurysm formation.

Tobacco abuse is common among patients with aortic aneurysm, but the presence of chronic obstructive pulmonary disease, irrespective of tobacco exposure, is the single best actuarial predictor of risk of rupture. Obviously the interplay of environmental factors needs to be studied against a background of defined genetic risk. It is likely that the risk of smoking, in addition to an underlying deficiency of an antiprotease system, may be particularly deleterious, for example in individuals who smoke and who have a_1-antitrypsin deficiency and emphysema. That cigarette smoking has an effect on atherosclerosis is supported by results from a retrospective autopsy survey in which smoking histories were obtained from relatives of the dead patients to evaluate the strength of association between cigarette smoking and atherosclerosis in 1320 men aged between 25 and 64 years. Atherosclerosis in the aorta and coronary arteries was greatest in heavy smokers and least in non-smokers. Animal studies show that both acute and subacute exposure to cigarette smoke result in striking morphological changes in aortic endothelium. Smoking also promotes platelet adhesion to endothelium, together with higher concentrations of plasma fibrinogen, lower concentrations of plas-minogen and plasminogen activator, and increased concentrations of factor VII. Diabetes, atherosclerosis, and hyperlipidaemia are also closely linked: arterial disease now accounts for more than 70 per cent of deaths from diabetes.

Aortic aneurysms are occasionally caused by other localized phenomena or systemic conditions. Aneurysms associated with bacterial infection occur sporadically. Certain varieties of infected aneurysm develop as a consequence of the bacterial colonization of prosthetic grafts. The normal arterial wall is very resistant to bacterial infection, and bacterial colonization is rare unless structural disruption has occurred, either acutely through penetrating trauma or chronically in an atherosclerotic or aneurysmal area. The single exception to this rule is the predilection for infection of normal arterial wall by *Salmonella* spp. Salmonella aortitis without aneurysm formation may lead to weakening of the aortic wall, and absence of aneurysm formation does not preclude rupture. Syphilitic aneurysms still occur sporadically (Fig. 9). Spirochaetal infection of the vasa vasorum of the large arteries, especially in the thoracic aorta, leads to destruction of the aortic media and marked saccular dilatation. These aneurysms erode into the mediastinal structures, bronchial tree, or even through the chest wall. The ascending aorta and arch are sites of predilection.

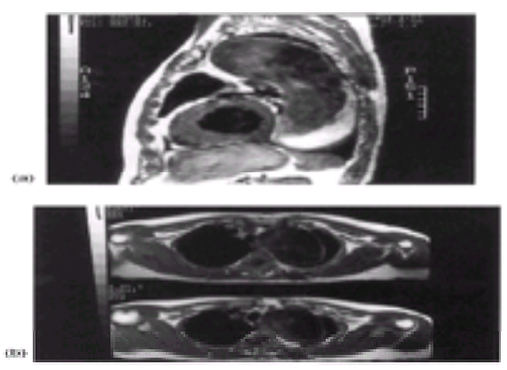


Fig. 9. (a) MRI scan of a massive, syphilitic, descending aortic aneurysm;(b) the aneurysmal ascending aorta had already been replaced for rupture.

Fibromuscular dysplasia or the hormonal changes associated with pregnancy may lead to aneurysm formation in young and middle-aged women, though this usually occurs in visceral arteries. The standard inflammatory aortic aneurysm seen usually in the elderly person occurs more frequently in the abdominal than in the thoracic aorta. Takayasu aortitis, though generally presenting with aortic occlusion, occasionally presents with thoracic aortic aneurysm (Fig. 10).

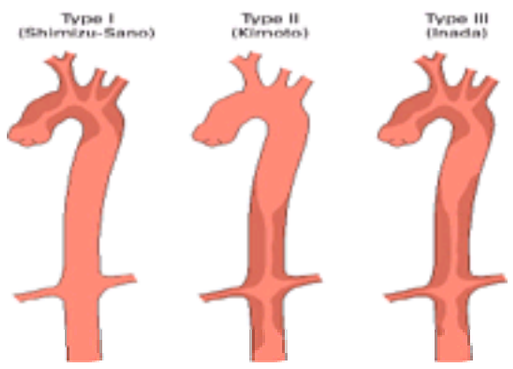


Fig. 10. Classification of Takayasu aortitis.

Imaging technique for thoracic aortic disease

Surgery of the thoracic aorta relies heavily on accurate diagnosis. Recent developments in angiography, ultrasonography, and MRI now allow thorough evaluation of the aorta with minimal invasion. Aortography has long been the standard against which other imaging techniques are compared (Fig. 7(a)). Advances in catheterization technique, together with a decrease in contrast toxicity, have lowered the risks of angiography. Digital subtraction angiography, particularly using intra-arterial injection, provides increased contrast sensitivity with excellent resolution. This method uses fluoroscopy to collect intensified images, which are converted to a digital form, stored in a computer, and processed to extract extraneous detail and amplify contrast (Fig. 5).

Angiography delineates luminal anatomy but will not demonstrate the aortic wall and periaortic tissues: the size and contour of the lumens of the aorta and branch vessels are shown best. Aortography is used to define occlusive lesions and luminal defects but will not clearly delineate aneurysms associated with occlusive disease. Accurate delineation of the dimensions of an aneurysm may be obscured by intraluminal thrombosis. Complicated aortic aneurysms, particularly contained ruptures and inflammatory aneurysms, are notoriously difficult to define by aortography. Dissecting aortic aneurysms are identified by narrowing or distortion of the contrast column or by the presence of an intimal flap between a double lumen. In many patients with disease of the thoracic aorta, coronary angiograms are, as mentioned above, desirable to determine the extent of coronary occlusive disease.

Aortography is an invasive technique and therefore has complications. Digital subtraction angiography does offer some advantages over conventional aortography: less contrast is necessary to visualize vessels, thereby reducing risk in patients with renal failure who require aortic evaluation. In practice, complications occur in fewer than 4 per cent of patients. Catheter-related complications include haemorrhage, thrombosis, pseudoaneurysm, and arteriovenous fistula formation.

Ultrasound imaging is derived from echo-ranging devices that are capable of measuring the distance to an object from the time required by an emitted pulse of sound to be reflected back to the source. In the chest, ultrasonography is used to assess the ascending aorta and aortic valve. Ultrasound can quickly differentiate aortic aneurysm from aortic tortuosity or mediastinal lymphadenopathy. Dissecting aneurysms with extramural haematoma or pericardial fluid are well visualized by two-dimensional echocardiography, and ultrasound imaging has not been shown to cause any deleterious effects. Transoesophageal echocardiography is now the investigation of choice for suspected aortic dissection.

The introduction of CT revolutionized thoracic vascular imaging. The CT scan uses a computer to process circumferentially acquired radiographic data in order to create a cross-sectional image. The image obtained actually represents a tissue slice of selected thickness between 2 mm and 1 cm: the thinner the slice the better the spatial resolution. Infusion of contrast can be used to depict vascular images and sequential analysis of CT sections can provide a three-dimensional study of the aorta. The CT scan is capable of delineating accurately images of the thoracic aorta, mediastinum, and pleural and peritoneal cavities. With contrast infusion the aortic lumen can be differentiated from wall thrombus, calcification, and aneurysm formation (Fig. 11(a),Fig. 11(b) and Fig. 11(c)). Rupture of a contained aneurysm is identified by loss of periaortic soft-tissue planes. Dissecting aortic aneurysms are easily identified when the true and false lumens are both patent (Fig. 12). If the false lumen is already occluded by thrombosis, it may be impossible to differentiate a dissection from a thrombosis-lined aneurysm. Postoperative graft complications such as pseudoaneurysms, infection, and graft–enteric fistula formation can be well defined by this technique. There are few complications of CT scanning. The level of radiation exposure is less than that for a barium enema, and contrast-related problems may be decreased by the use of a non-ionic type of contrast medium.

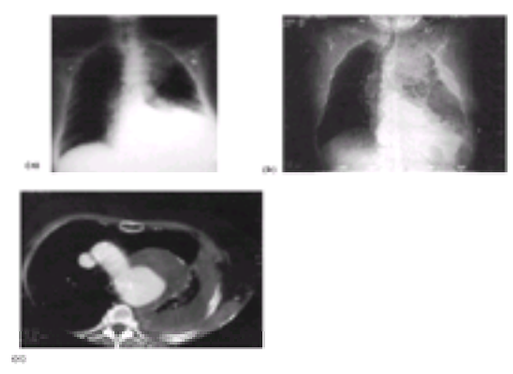


Fig. 11. Rupture of an aneurysm of the distal aortic arch (syphilitic).(a) Plain chest radiograph showing the left-sided mediastinal mass and left haemothorax. (b) Composite CT image corresponding to the chest radiograph. (c) Cross-sectional image of the aortic arch with contrast in the lumen. There is blood clot around the aneurysm and blood in the pleural cavity.

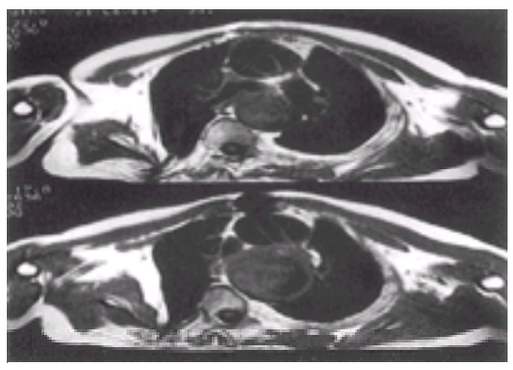


Fig. 12. CT scan in acute type A dissection; there is an associated aneurysm of the descending thoracic aorta.

MRI is a non-invasive technique that uses a radiofrequency pulse directed at a substance within a large magnetic field, inducing a transition between energy states of certain atoms. The time required for localized nuclei to return to the baseline energy state, called the relaxation time, is measured and translated by computer into an

image. Various radiofrequency pulse sequences are used in combination with gradient magnetic fields to create an optimal image. MRI does not require ionizing radiation or contrast infusion to visualize vasculature in multiple planes. Spatial resolution is similar to that obtained with CT scanning, but MRI offers better soft-tissue resolution and is not limited by air, bone, excess fat, or operator skill. The aortic lumen and wall, branch vessels, and adjacent structures are all well defined by MRI. Aortic congenital abnormalities, thoracic aneurysms, and aortic dissection are well visualized in multiple planes (Fig. 13). Congenital anomalies such as double aortic arch, coarctation of the aorta, and anomalous vessel origins can be clearly defined. Use of oblique sections may delineate vascular anatomy even better than aortography by providing images along the long axis of the vessel (Fig. 14). Resolution is such that a contained rupture can be differentiated from inflammatory fibrosis around an aneurysm (Fig. 15). Graft complications such as haematoma, pseudoaneurysm, or infection are reliably defined, though the patency of small grafts and aortoenteric fistula are not.

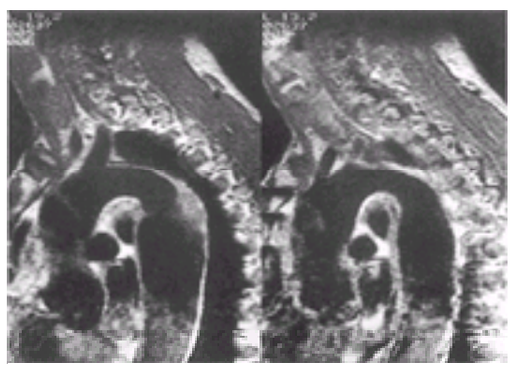


Fig. 13. MRI scan in longitudinal plane after type A aortic dissection.

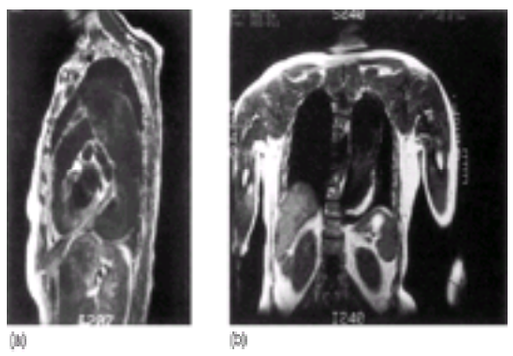


Fig. 14. Oblique MRI section showing a normal ascending aorta together with a massive chronic type B aortic dissection.

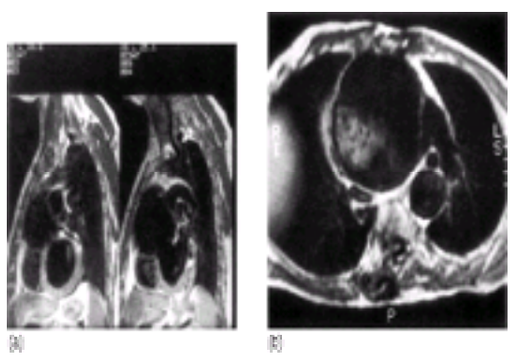


Fig. 15. MRI scan showing a massive chronic ascending aortic aneurysm containing thrombus: (a) longitudinal section; (b) transverse section.

MRI studies are contraindicated in uncooperative, haemodynamically unstable, or claustrophobic patients, and in those dependent on devices with ferromagnetic properties such as pacemakers, cerebral aneurysm clips, and ventilators. A relatively long scan time is re-quired within the strong magnetic field.

Perfusion techniques in thoracic aortic surgery

Surgical procedures on the thoracic aorta require cross-clamping of the vessel with interruption of distal flow. With the exception of saccular aneurysm, where a side clamp can be applied and distal flow maintained, the success of thoracic aortic surgery depends on preservation of function in dependent organs. The methods adopted depend on the site of cross-clamping and associated risks. The abdominal viscera, including the liver and kidneys, may tolerate 30 min of normothermic interruption of blood supply, but the brain will tolerate little more than 4 min of unprotected circulatory arrest. In addition, aortic cross-clamping performed with the heart supporting the circulation causes hypertension proximal to the clamp, increased stroke work, and subendocardial myocardial ischaemia, particularly in patients with coronary occlusive disease. For coronary patients with impaired left ventricular function, unprotected aortic cross-clamping, even of the abdominal segment, may have serious consequences. Perfusion techniques vary widely according to the site of the aortic operation. Surgery of the aortic root and ascending aorta requires whole-body perfusion with moderate or profound hypothermia while the heart and lungs are out of circuit; full cardiopulmonary bypass with a pump oxygenator and cardiotomy suction apparatus are required. To ensure continuous cerebral perfusion during resection and graft replacement of the ascending aorta, most patients require placement of the distal occluding clamp just proximal to the origin of the innominate artery. Often, however, the pathological process extends into the arch or beyond, and may involve the origin of the innominate, left carotid, and left subclavian arteries. With an aortic cross-clamp in place it is impossible to ensure perfusion to arch vessels and clamping may promote aortic rupture in aortic dissection or cerebral embolization from atherosclerotic debris. Cooley suggested the use of circulatory arrest in these patients in order to perform the distal anastomosis by an open technique (Fig. 16). Arterial return from the pump oxygenator is via a common femoral artery. The site for the venous outflow cannula depends on the anatomical findings, but is usually in the right atrium. The systemic temperature is reduced by cardiopulmonary bypass and heat exchange to 20°C. During this interval the arch vessels are mobilized and, at an appropriate time, the vessels are cross-clamped. The arterial return pump is discontinued and the venous outflow occluded. The aorta can then be opened widely and the pathological condition of the interior of the transverse arch and origin of the arch vessels appraised. In patients with aortic dissection, the relation of the true to the dissected false lumen must be established. Conducting the anastomosis with the aorta wide open and bloodless facilitates proper management of the aortic arch by resection and accurate attachment of the graft to the underside of the aorta. At the end of the procedure, arterial return is re-established to fill the aorta gradually and eliminate entrapped air from the roots of the head vessels. This open-ended technique has greatly facilitated extended resection of the ascending aorta. Meanwhile, cold potassium cardioplegic myocardial arrest together with profound systemic hypothermia protects the heart, the brain, and the abdominal viscera.

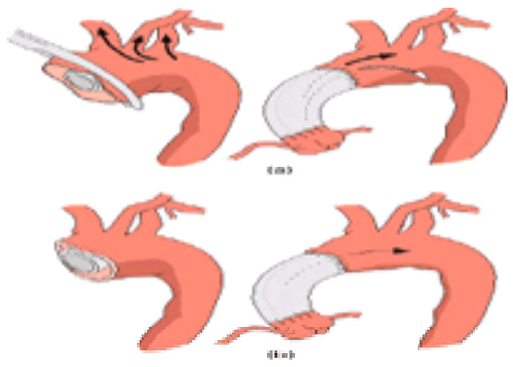


Fig. 16. Open-ended technique for aortic anastomosis (Cooley): (a) the dangers of cross-clamping; (b) the open method.

Whereas surgery of the ascending aorta and aortic arch requires hypothermic bypass in order to provide a bloodless field and protect the brain and myocardium, most operations on the descending aorta can be performed by simple cross-clamping. Shunt bypass and hypothermic cardiopulmonary bypass techniques were introduced to provide perfusion to the abdominal viscera during normothermic cross-clamping and to off-load the myocardium against an acute increase in afterload. It was hoped that the spinal cord might be protected by perfusion of the distal thoracic aorta and, consequently, of the principal spinal radicular vessels between T10 and L2. Table 4 summarizes the passive shunt, pump bypass, and hypothermic cardiopulmonary bypass techniques employed in descending aortic surgery. None can convincingly be said to protect the spinal cord if the spinal vascular anatomy is unfavourable. Hypothermic cardiopulmonary bypass methods, discussed later, require heparinization and can be used with or without total circulatory arrest. Of the passive shunt bypass methods the ascending aorta–distal aorta is preferable and can be performed with a heparin-bonded shunt, making systemic heparinization unnecessary. Shunts between the aortic arch and iliac artery, left subclavian and the distal aorta, and left ventricle–distal aorta are less reliable, with unpredictable flow. For pump bypass methods a roller or centrifugal pump may be employed. Centrifugal pumps can be used without heparin, in particular for aortoaortic bypass.

Passive shunt bypass:
1. Large cannula (> 9mm) with bonded heparin, no systemic heparinization
(a) Ascending aorta to distal aorta
(b) Aortic arch to iliac artery
(c) Left ventricle to distal aorta
(d) Left subclavian to distal aorta
Pump bypass:
2. Roller or centrifugal pump
(a) Aorto–aorta with centrifugal pump no heparin (Cooley)
(b) Left atriofemoral with heparin, reservoir, and roller pump
(c) Left atriofemoral with centrifugal pump, no heparin
(d) Venoarterial:
Superior vena cava to femoral artery
Right atrium and femoral vein to femoral artery with or without oxygenator, reservoir, and heparin
Hypothermic cardiopulmonary bypass:
3. With or without total circulatory arrest
[all with full systemic heparinization]
(a) Crawford: femoro-femoral; cannot perfuse head or heart when clamps applied
(b) Kaufmann: femoral vein and pulmonary artery to femoral artery; cannot perfuse head or heart when clamps are applied
(c) Walmsley: right atrium to a right internal jugular via aortic arch; continuous perfusion of head and heart during clamping

Table 4 Distal perfusion in surgery of the descending thoracic aorta

Left atriofemoral bypass can be performed with either a centrifugal or a roller pump: the centrifugal does not require systemic heparinization; when a roller pump is used, heparin and a blood reservoirare necessary. Various methods of venoarterial pump bypass have been employed, including from the superior vena cava to the femoral artery and from the right atrium and femoral vein to the femoral artery. Both of these methods have been used with or without an oxygenator, reservoir, and systemic heparinization. For straightforward operations, such as repair of coarctation in an adult or repair of traumatic aortic laceration, we use simple, normothermic aortic cross-clamping without bypass techniques. This type of repair can almost always be achieved in less than 30 min, the arbitrary period for alleged safe cross-clamping: ‘safe' applies more to the medicolegal aspects than to the spinal cord. For complicated excision and replacement of extensive descending thoracic and thoracoabdominal aneurysms, and in certain patients with type B dissection, we employ our own hypothermic cardiopulmonary bypass technique as described later.

Retrograde cerebral perfusion

The technique of retrograde brain perfusion via the superior vena cava was first used in 1980 by Mills and Ochsner to augment cerebral protection in patients undergoing complex replacement of the aortic arch. The method involves cannulation of the superior vena cava and retrograde hypothermic (15°C) perfusion of the cerebral veins at a flow rate of approximately 500 ml/min, with the central venous pressure kept below 30 mmHg. The probable beneficial effects include maintenance of more uniform, hypothermic brain temperatures, the prevention of rewarming of the brain during circulatory arrest, the flushing out of air and embolic material at the end of circulatory arrest, and possibly the delivery of oxygen and metabolic substrates to the brain. Metabolic products including lactic acid may be removed by continuous retrograde perfusion, thereby providing improved pH control. Whilst retrograde cerebral perfusion is thought to reduce the risk of embolic stroke, no prospective randomized trial has demonstrated efficacy. Consequently, we restrict its use to hypothermic circulatory arrest for full replacement of the aortic arch where the predicted duration of arrest exceeds 30 min. For circulatory arrest of shorter duration we use a temporary 2-min period of retrograde perfusion to evacuate air or debris from the brachiocephalic vessel before resuming cardiopulmonary bypass.

Vascular prostheses

Bleeding problems remain an important cause of morbidity and mortality in thoracic aortic surgery, particularly for aortic dissection. Operative methods have long been influenced by the relatively porous nature of proprietary vascular prostheses. Indeed the classical wraparound techniques for aortic root replacement as described by Bentall and De Bono, with modifications by Cabrol, were designed specifically to cope with oozing through porous grafts. Most commonly used grafts are made from either Dacron or Teflon (polytetrafluoroethylene, PTFE). Teflon has a smooth, non-adherent, non-wettable surface that may inhibit the sticking of platelets and fibrinous material to its surface; this allows the use of smaller-diameter grafts for peripheral vessels or for modified Blalock shunts in infants (Gore-Tex or Impra). Teflon is non-porous to blood and bleeding through the graft itself does not occur. However, Teflon grafts are relatively non-compliant and kink easily: external ribbing has been applied to help prevent this problem. Bleeding occurs through suture holes and gaps between sutures, and these grafts are not normally used in patients requiring systemic heparinization.

Dacron grafts are available as knitted and woven varieties (Fig. 17). Knitted grafts are of high porosity and are easier to sew, but leak when the vascular clamps are removed until clot forms in the graft interstices. Woven grafts are of low porosity and are more difficult to sew, but bleeding through the graft is minimized or eliminated. As a general rule, the lower the porosity of a graft the greater the stiffness and poorer the handling characteristics. Knitted grafts are designed for abdominal and peripheral vascular reconstruction in non-heparinized patients. In general they are easily sewn to pathological vessels because of their soft, compliant nature. Their deformability allows a snug fit and minimizes bleeding at the anastomotic site. Ease of suture passage is particularly helpful when the native vessel wall is heavily calcified, resulting in needle blunting or deflection from the intended suture path. However, when the aortic clamps are removed, bleeding through the graft interstices can sometimes be considerable. Knitted grafts are therefore contraindicated in heparinized patients, owing to the risk of uncontrollable haemorrhage. The standard knitted graft has been modified by the incorporation of a velour inner and outer pile to enhance the incorporation of clots in the graft interstices (Meadox double-velour graft). Knitted grafts of less than 5 mm diameter should not be used because of the high rate of thrombosis.

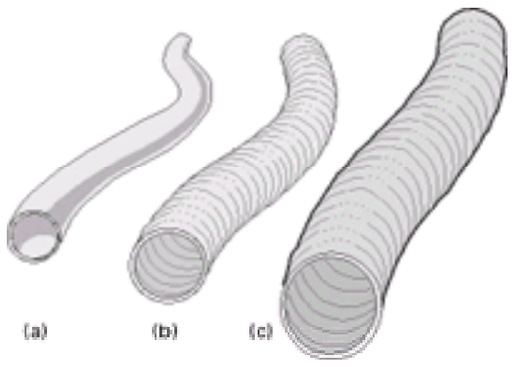


Fig. 17. Expanded polytetrafluoroethylene (a), knitted (b), and woven (c) Dacron grafts.

It was originally thought that high-porosity knitted grafts were necessary to allow endothelial cells and their subendothelial matrix to invade and adhere to the inner surface of the vascular prosthesis, reducing thromboembolism and maintaining patency. However, although prosthetic grafts may develop a neoendothelium in canine and non-human primate models, re-endothelialization occurs only sporadically or not at all in man. For practical purposes, re-endo-thelialization is, therefore, not a consideration when implanting prosthetic grafts in patients.

The tight weave of woven grafts was designed specifically for use in patients undergoing thoracic aortic replacement with systemic heparinization. These grafts are stiffer and eventually blunt fine suture needles. Graft stiffness is not usually a problem when anastomosing to the margins of thickened aorta after resection of fusiform or saccular aneurysms, where the tissues are thick, fibrous, and hold sutures well. When the aorta is thin or dissected, or in child patients, attempts to sew a stiff, non-compliant graft may be fraught with further tearing, sometimes to the point of suture-line avulsion.

There have been recent improvements in woven Dacron grafts. The Meadox low-porosity Veri-Soft graft is made of Dacron yarn that is soft enough to allow good flexibility but has a tight enough weave to prevent significant haemorrhage through the interstices. However, preclotting is advisable, even for low-porosity grafts, and this further impairs the handling characteristics. The object of preclotting is to seal the interstices of the woven or knitted material so that bleeding does not occur when the clamps are removed. The patient's own, non-heparinized, fresh blood is most effective for this purpose, but this does not apply when cardiopulmonary bypass is already established. Since it is impossible to size the native aorta accurately until the patient is systemically heparinized with the vessel open, complicated and tedious graft preparation is necessary. We have used coating with autologous or heterologous plasma followed by microwave coagulation; this certainly seals the graft, but it also impairs handling characteristics and suture passage. Fresh donor blood of the same group from a volunteer in the operating theatre is feasible but we abandoned this method when a patient contracted postoperative hepatitis for a completely different reason. Bleeding through the suture holes in Gore-Tex grafts can be limited by using small-diameter needles and Gore-Tex suture material, which expands when wet.

Experimental work has demonstrated no difference in the long-term patency of woven compared to Dacron grafts. An aortic bifurcation graft in which one limb was knitted and the other woven was implanted into 143 consecutive patients with atherosclerotic aortoiliac occlusion or abdominal aortic aneurysm: there was no difference in patency at periods of 1 month to 2 years.

For replacement of the aortic root, proprietary composites of a Carbomedics, Medtronic, or St Jude valve sewn into a woven Dacron graft are available (Fig. 18). In this case the size of the valve annulus determines the selection of the prosthesis: this can only be judged accurately with the patient heparinized and the aorta open. Preclotting with fresh frozen plasma followed by autoclaving is therefore required and wastes time. Porcine valved conduits are commonly used for reconstruction of the right ventricular outflow in children. Unfortunately, the incidence of porcine valve degradation is extremely high, making aortic homograft material far superior (Fig. 19). Homografts of the aortic root are particularly useful when enlargement is required or for the treatment of infective conditions such as endocarditis of a prosthetic valve with root abscesses. They are less applicable in very large aortic roots because of size mismatch. A non-suturable ring of Dacron felt can be incorporated into the suture line to avoid homograft dilation in patients with Marfan syndrome. Recently, the availability of the completely impervious, collagen-impregnated, woven Dacron prosthesis (Hemashield, Meadox) has eliminated the need for preclotting and provided a graft with excellent handling and suturing characteristics. Hemashield is low-porosity, woven Dacron that is impregnated with type 1 collagen derived from fresh, young calf skin. The collagen for Hemashield is prepared from calf skin by a sequence of steps including mechanical cleansing, washing, and chemical treatment. The processed skin is then homogenized to produce a paste composed of 90 per cent collagen microfibres; this is used to coat a woven, double-velour, Dacron polyester graft. Electron microscopy shows that the collagen layer completely covers the fibres and interstices of the fabric. The graft is biocompatible in terms of pyrogenicity, cytotoxicity, and antigenicity. It is soft and pliable, with excellent handling characteristics. The collagen is absorbable and rendered non-antigenic by cross-linking on the graft surface. The resulting vascular prosthesis is therefore blood compatible and does not induce platelet aggregation or intravascular clotting. Less than 10 per cent of the bovine collagen remains on the graft after 30 days and less than 2 per cent of the graft surface remains covered with collagen after 90 days.

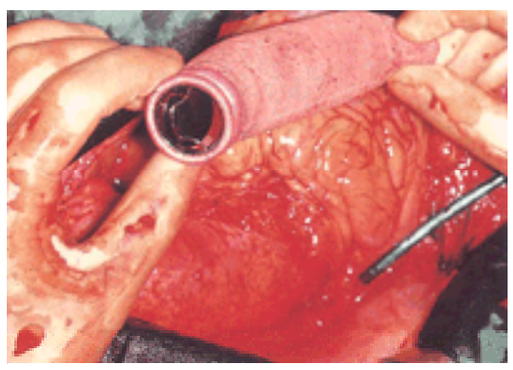


Fig. 18. Composite valved conduit for aortic root replacement.



Fig. 19. Aortic homograft for aortic root replacement.

Topical haemostatic agents

P>Successful surgery of the thoracic aorta depends on the ability to prevent both surgical and abnormal bleeding. The tenuous nature of diseased or dissected arterial wall presents a major operative risk. Medial necrosis, collagen cross-linking deficiencies, or atheroma are present in most abnormal aortic tissues, lowering the tensile strength and suture-holding capacity. This, together with high wall tension, weak adventitia, and the absence of reinforcing pleura in the ascending thoracic aorta, increases the likelihood of suture-line tearing and haemorrhage, which may be difficult to control by standard methods. Surgical bleeding prolongs cardiopulmonary

bypass, which in turn leads to coagulation problems and the potential for multisystem organ failure. Operative techniques have therefore been designed specifically to combat bleeding: these include wraparound techniques, including the fistula method of Cabrol, and the insertion of a graft with rigid ends inside the pathological aorta that is fixed with external ties to avoid an aortic suture line. Although the introduction of completely impervious vascular grafts has greatly reduced operative haemorrhage and transfused blood requirements, bleeding through suture holes in the native pathological aorta can cause problems. The use of thrombogenic materials such as oxidized cellulose and gelatin sponges is rarely effective. We routinely use Teflon strips to buttress a dissected aorta, though this does not entirely eliminate the bleeding. In patients with dissection, poor tissues, or detachment of the adventitia, we use Teflon O-rings for reimplantation of the coronaries in replacement of the aortic root. Hammond described tanning of the dissected aorta with glutaraldehyde before anastomosis. The dilute glutaraldehyde solution is carefully applied to the native aorta on cotton-wool balls, which are kept in apposition to the tissue for 4 to 5 min; this greatly strengthens the weakened tissues and promotes their suture-holding ability. Care must be taken not to allow glutaraldehyde to drip within the aortic lumen or contaminate adjacent tissue.

We occasionally employ acrylic blue glue (Braun) to seal suture lines in patients with fragile or calcified aortic walls. Cyanoacrylate adhesives were first used as topical haemostats during the late1960s, but were not adopted widely because of their lack of adhesionto moist surfaces and their potential histotoxic side-effects. Undoubtedly the profound exothermic reaction that occurs when large amounts of glue are applied has the potential to cause damage, but in practice we have not experienced any adverse reaction. When the glue is applied prophylactically before blood is admitted into the repair site there is little danger of adverse effects (Fig. 20). The glue is supplied in a plastic vial with a long, narrow, sealed vent. When the tip is removed it can be applied accurately to the native aorta and sets rapidly. If glue fails to stop bleeding from an uncontrolled tear in the native aorta wall, it can be peeled away from the surface to allow secondary suture placement. Acrylic glue can be used to occlude the path of blood loss from inaccessible areas and tamponade the bleeding area.



Fig. 20. Acrylic glue used to seal aortic suture lines.

Bachet introduced the use of gelatin–resorcin–formol glue in patients with acute dissection. Carpentier's group have successfully used this agent to stick the dissected layers together, avoiding resection of the aorta with its attendant risks of bleeding. We have found this glue to be excellent for reconstitution of the dissected aortic wall in the conventional repair of acute type A or B dissection.

Fibrin sealant is an alternative haemostatic agent that consists of completely biodegradable components. The underlying principle is local imitation of the last stages of the physiological coagulation process by the use of highly concentrated human fibrinogen and bovine thrombin. All components of fibrin sealant are completely biodegradable within 6 weeks and cause no foreign-body reaction. Fibrin sealant requires about 20 min of preparation and is applied through a needle for accurate placement. We and others have found it particularly useful in children when Dacron grafts are anastomosed to myocardium, and for bleeding surfaces at reoperation. Topical infusion of human cryoprecipitate is particularly useful for stopping diffuse oozing in patients with acute aortic dissection, disseminated intravascular coagulation, or undergoing reoperation with diffuse bleeding. The concentrated solution is poured into the pericardium over the raw surfaces. The cryoprecipitate assumes a gelatinous form and deposits fibrin strands over the surfaces. On several occasions, this method has allowed us to close patients with diffuse abnormal bleeding. Usually we combine topical and intravenous application, together with the administration of fresh frozen plasma and platelets. The risks are those associated with the use of a multidonor plasma pool, but in Britain the transmission of hepatitis or human immunodeficiency virus is extremely rare.

Pharmacological methods to reduce perioperative bleeding

Bleeding from a cut blood vessel ceases because of a combination of vasoconstriction and the formation of a platelet plug, followed by the development of a fibrin network. In the abnormal thoracic aorta, vasoconstriction does not occur, and adhesion of platelets to puncture sites and small tears is the first and crucial step in the haemostatic process. Plug formation occurs only when platelets have first been deposited on the subendothelium. Platelets possess receptors for various adhesive proteins present on the blood vessel wall or in the plasma. The high shear forces associated with cardiopulmonary bypass mean that binding of the platelets to endothelium is through the receptor glycoprotein 1b, which binds to von Willebrand factor, which in turn binds to the exposed subendothelium. Platelets are activated as they aggregate together by expression of glycoproteins IIb and IIIa, which link platelets together through fibrinogen molecules. During cardiopulmonary bypass, heparin is used to bind antithrombin III, preventing propagation of the clotting cascade. However, the fibrinolytic system and platelets are activated by exposure to the foreign materials of the perfusion circuit. Harker showed that the bleeding problems associated with cardiopulmonary bypass were commonly related to a defect in platelet function and that the degree of functional impairment was proportional to the duration of cardiopulmonary bypass and level of hypothermia. There is also evidence that fibrinolytic activity may be responsible for excessive bleeding. Increased fibrinolytic activity is primarily the result of an increase in tissue plasminogen activator initiated before cardiopulmonary bypass; this reaches a maximum 1 h after the start of perfusion. Fibrinolytic activity is also caused to a lesser extent by the production of free plasmin at the beginning of cardiopulmonary bypass induced by activation of factor XIIa, probably by the first cycle through the bypass equipment. The mechanisms underlying an increase in the release of tissue plasminogen activator during surgery are not clearly defined but include venous occlusion, hypoxia, thrombin, and adrenaline (epinephrine). A hyperadrenergic state induced by surgical stimulation may prove responsible for the initial increase in tissue plasminogen activator before bypass.

In 1984, Westaby *et al.* suggested the use of aprotinin, a non-specific protease inhibitor, to inhibit the release of elastase following complement and white-cell activation in the extracorporeal circuit. As early as 1964, aprotinin (Trasylol) had been shown to reduce fibrinolytic activity during cardiovascular operations, but this aspect was then largely forgotten. During the clinical trial of aprotinin in coronary bypass patients, bleeding was considerably reduced and the operative field appeared abnormally dry after reversal of heparin by protamine. Further trials confirmed substantial reductions in periopera-tive blood loss following aprotinin infusion. In low concentrations, aprotinin is a powerful inhibitor of plasmin; in high doses it also inhibits kallikrein, which is formed during activation of coagulation by negatively charged surfaces such as those in cardiopulmonary bypass equipment and exposed subendothelial layers. Kallikrein has a central role in the activation of the inflammatory response, complement, the angiotensin system, fibrinolysis, and coagulation. Consequently, if kallikrein is inhibited there is a decreased inflammatory response and less tendency to activate coagulation and fibrinolysis.

During cardiopulmonary bypass, aprotinin directly inhibits kallikrein production and probably the release of tissue plasminogen activator, as shown during liver transplantation. Aprotinin preserves platelet function through its antifibrinolytic effect. Free plasmin removes the von Willebrand receptor glycoprotein Ib from the platelet membrane: thus there is inhibition of platelet adhesion to subendothelial structures and a platelet plug is formed; this occurs during cardiopulmonary bypass and can be inhibited by aprotinin. In addition, aprotinin inhibits the coagulation pathway and prolongs the activated clotting time. When heparin therapy is being monitored by activated coagulation during cardiopulmonary bypass the activated clotting timer should run at a higher level than normal to ensure that adequate concentrations of heparin are used.

We have been cautious in the clinical use of aprotinin, since it is a prothrombotic agent: early in its use we suspected that it might promote coronary graft occlusion. We therefore restrict its use to patients undergoing surgery for acute aortic dissection or ruptured thoracic aneurysms who present with substantial blood loss and require transfusion of large volumes of blood. In profoundly hypothermic states of circulatory arrest, aprotinin may promote bleeding by inhibiting the protein C and tissue plasminogen activator systems that maintain the fluid status of static blood. Platelet microaggregates may form and produce a disseminated intravascular coagulation effect.

Other agents have been used to reduce perioperative bleeding. Desmopressin increases the plasma concentration and activity of von Willebrand factor, probably by inducing release of this factor from the endothelium. Clinical trials of desmopressin during cardiopulmonary bypass have demonstrated shortened bleeding times postoperatively and reduced blood loss. Plasminogen and plasmin combine to fibrin through their lysine-binding sites; this binding can be reduced by the lysine analogues, ee-aminocaproic acid and tranexamic acid, thereby delaying fibrinolysis. However, these agents are rarely used clinically and have not yet been shown to produce significant benefit in the treatment of established and excessive postoperative bleeding. Dipyridamole limits platelet aggregation and granular release by inhibiting platelet phosphodiesterase activity. This agent is used to preserve vein graft patency, but in one study of perioperative bleeding during cardiac surgery, patients receiving dipyridamole had higher platelet counts at the end of operation and less blood loss than controls.

Since bleeding remains a particularly important complication in thoracic aortic surgery, particularly for acute aortic dissection, pharmacological methods for inhibiting abnormal bleeding have an important role; aprotinin currently appears to be the most promising agent.

Surgery of aortic dissection

Acute aortic dissection occurs when an intimal tear allows blood to enter the media (Fig. 21(a),Fig. 21(b) and Fig. 21(c)). A combination of factors, including hypertension and cystic medial degeneration, allows the tear to propagate longitudinally through the aortic wall (Fig. 22). Tears occur at the site of maximal stress on the aortic wall: 66 per cent occur in the anterior wall above the aortic valve (Fig. 23) and 33 per cent on the posterior wall of the proximal descending thoracic aorta. The classification of acute aortic dissection is shown in Fig. 24. Hypertension is an important aetiological factor: dilated aortas tend to dissect more readily. Flexion of the ascending aorta and stress at the site of tethering of the descending aorta to the posterior thoracic wall may account for the site of intimal tear, which often occurs through an atherosclerotic plaque. Rigid or calcified atherosclerotic plaques alter the distensibility of the aorta. b-Blockers prescribed for patients with hypertension can increase the risk of aortic dissection by increasing systemic vascular resistance. Vasodilators reduce this effect.

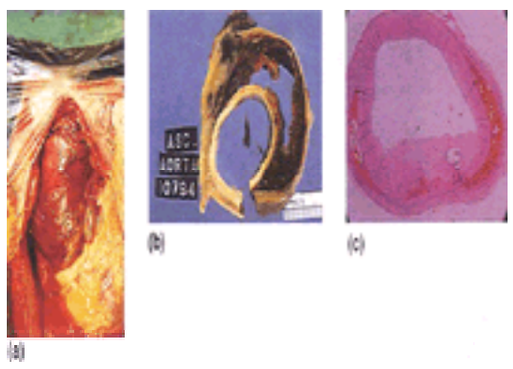


Fig. 21. Acute aortic dissection: (a) at operation; (b) cross-section demonstrating the true and false lumen; (c) histological section showing dissection of blood through the intimal layer.

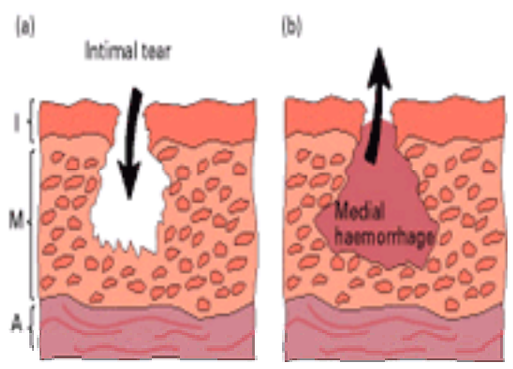


Fig. 22. Proposed mechanism of initiation of aortic dissection. In both cases, cystic medial necrosis is present: in (a) an intimal tear is the initial event allowing blood to enter the media; in (b) the primary event is haemorrhage into the media with secondary rupture of the overlying intima. I, intima; M, media; A, adventitia.



Fig. 23. Type A acute aortic dissection.

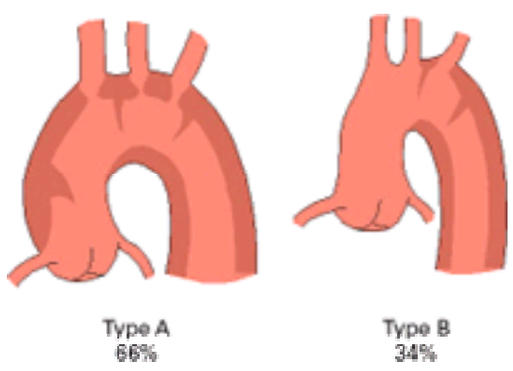


Fig. 24. Classification of acute aortic dissection.

Ninety per cent of patients with acute aortic dissection present with sudden, excruciating chest pain that migrates anteroposteriorly. There are no prodromal symptoms and the pain is distinctly different in character and mode of onset to that of acute myocardial infarction. The patient with aortic dissection can usually recall precisely their activities at the time of onset, unless stroke is the presenting feature. Seventy-five per cent of patients have systemic hypertension on admission to hospital and 25 per cent have acute aortic regurgitation with an audible murmur (Fig. 25). Ten per cent suffer an associated acute myocardial infarction, usually from occlusion of the right coronary artery by retrograde dissection down to the aortic annulus. The classical absent limb pulse or stroke occurs in relatively few patients (Fig. 26). Rupture of the aorta with sudden death from haemorrhage or cardiac tamponade occurs frequently. At presentation the chest radiograph shows mediastinal widening in at least 65 per cent of patients, often with a left pleural effusion (Fig. 27).

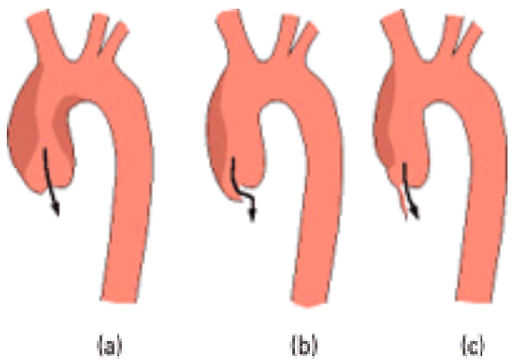


Fig. 25. Mechanism for aortic regurgitation in acute aortic dissection.

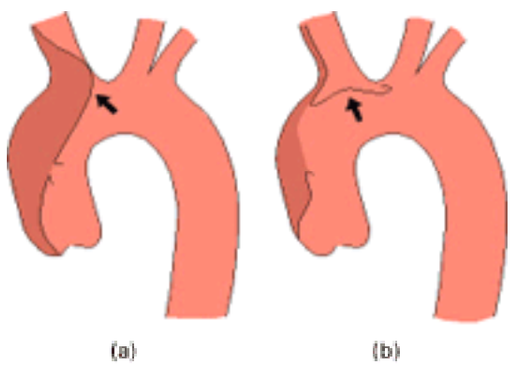


Fig. 26. Mechanism for loss of pulses in acute aortic dissection.



Fig. 27. Chest radiograph in type A aortic dissection; the patient had previously undergone double valve replacement.

These clinical features are emphasized for two important reasons. First, the outlook for patients with aortic dissection worsens with time because of the development of critical events such as stroke, renal failure, ischaemic bowel, and cardiac tamponade. Surgery provides good long-term results in patients operated upon before irreversible injury to vital organs has occurred. However, initial diagnostic error is a major problem: up to 50 per cent of patients are referred more than 36 h after the original event, often with established bowel infarction or renal failure. In our experience of 54 patients in 4.5 years, operation within 24 h of acute dissection using modern techniques carries a very low risk of death in hospital; hospital deaths occurred in patients who presented late with established pulmonary and renal failure, ischaemic gut, or coagulopathy. The second reason for emphasizing diagnostic accuracy is that a number of patients are mistakenly treated with streptokinase for suspected acute myocardial infarction; this perpetuates the tear by preventing clotting in the false lumen and creates difficulties during early operation through gross derangement of clotting and platelet function. We have operated on patients with acute dissection within 24 h of streptokinase administration by giving considerable amounts of fresh frozen plasma, platelets, and cryoprecipitate.

CT of the thorax has been acclaimed as the diagnostic method of choice in aortic dissection but is not always immediately available. CT scan may not demonstrate details of the aortic root and valve, or left ventricular function; it may even fail to demonstrate the dissection. MRI is even less likely to be immediately available, but when it is, patient monitoring is more difficult than for CT. Modern echocardiography may show the dissection flap in the ascending aorta, the anatomy of the aortic valve, and the presence of aortic regurgitation, as well as left ventricular function and regional wall-motion abnormalities. The use of the transoesophageal ‘window' raises the sensitivity and specificity to nearly 100 per cent. This technique is already available with colour-flow Doppler and biplane transducers, and satisfactory imaging is guaranteed and obtained very quickly.

Coronary angiography is desirable for patients with a history of ischaemic heart disease or previous coronary grafts, but this should not delay rapid transfer to the operating theatre. In patients with previous coronary bypass grafts the problem is addressed by excising their top ends in a single aortic button and reimplanting this into the new ascending aortic graft.

Careful medical management is imperative both pre- and postoperatively (Table 5). The aim of systemic pressure control is to reduce the mean blood pressure and dp/dt. Sodium nitroprusside alone does not reduce dp/dt. Trimetaphan camsylate (Arfonad) effectively reduces wall tension and is probably superior to most other drugs. Early presentation with stroke does not contraindicate surgery, as long as the patient is not clinically brain dead. Cerebral deficits usually resolve after prompt repair, which reopens the arch vessels and reperfuses ischaemic brain. All patients with type A dissection should be submitted for surgical repair as soon as possible (Table 6). Currently, medical management is advocated for the majority of patients with type B dissection, though improvements in surgical technique, including hypothermic protection of the spinal cord, make this more controversial. Certainly, patients with leak into the left pleural cavity or persistent pain over 2 to 3 days should be considered for operation. Surgery is the only option when dissection occludes vessels to the kidneys and gut.

Table 5 Approach to the patient with acute aortic dissection

Surgical
1. Treatment of choice for proximal dissection
2. Treatment for distal dissection complicated by:
(a) Progression with vital-organ compromise
(b) Rupture or impending rupture (saccular aneurysm formation)
(c) Aortic regurgitation
(d) Inability to control pain or blood pressure medically
Medical
1. Treatment of choice for uncomplicated distal dissection
2. Treatment for uncomplicated proximal dissection if the site of origin cannot be clearly identified
3. Treatment for stable, isolated arch dissection
4. Treatment of choice for chronic dissection, i.e., uncomplicated dissection presenting 2 weeks or later after onset

Table 6 Indications for definitive surgical and medical therapy in aortic dissection

Operative technique for type A dissection

Patients with cardiac tamponade often deteriorate in the anaesthetic room and urgent intervention is required. We routinely cannulate the femoral artery for arterial return and have the cannula in place before the pericardium is opened. Relief of tamponade often results in a blood-pressure surge, which may precipitate rupture of the false lumen. With established arterial return, bypass can be commenced using suckers while a two-stage cannula is quickly inserted into the right atrial appendage.

It is important to check the efficacy of perfusion soon after commencing bypass. The temperature of the head should be assessed: in a small proportion of patients, femoral arterial perfusion will not reach the head; it is therefore useful to have a second arterial line, which can be inserted through the apex of the left ventricle, thereby perfusing the head through the true lumen via the aortic valve (Fig. 28). We routinely use a ‘cerebral protection cocktail' consisting of the calcium channel-blocker nimodipine, the barbiturate thiopentone, and the free-radical scavenger mannitol. In one of our patients in whom the aortic cross-clamp had already been applied and cardioplegia administered, the anaesthetist reported failure to cool the head after 10 min of cardiopulmonary bypass; the temperature of the lower body had already fallen to 27°C. For this patient the cannula was urgently switched to the aortic arch, which restored cerebral perfusion after 15 min. The patient made a full neurological recovery and returned to his occupation as an accountant. We consider that pharmacological cerebral protection contributed to his recovery.

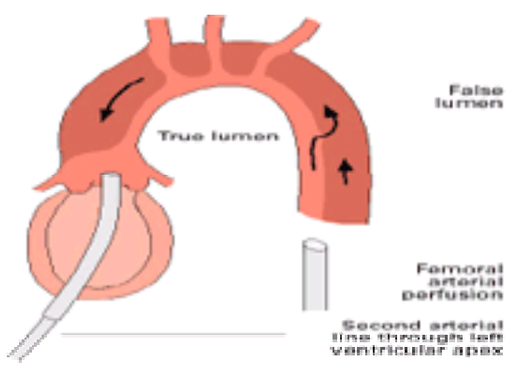


Fig. 28. Anterograde perfusion through the left ventricular apex in a patient with malperfusion through the femoral approach.

It is important to establish the extent of surgery required at an early stage. Aortic valve replacement is rarely needed: the great majority are repairable by resuspension, even in patients with non-stenotic bicuspid valves. Patients with Marfan syndrome should always undergo replacement of the aortic root since their aortic sinuses will eventually expand, with aortic regurgitation or rupture, if left (Table 7). We routinely use hypothermic total circulatory arrest at 20°C for complete replacement of the ascending aorta and proximal hemiarch with an impervious vascular graft. Both the dissected aortic sinuses and the dissected part of the aortic arch are repaired with gelatin–resorcinol–formol glue to obliterate the false lumen and improve suture holding. In 15 per cent of patients the primary tear is found within the aortic arch and cannot be addressed without total circulatory arrest and open repair. In addition, the use of a cross-clamp on a dissected aorta further damages the attenuated aortic wall and perpetuates the dissection when the cross-clamp is removed. If the repair of the aortic root is undertaken first (during the cooling period before open-ended distal anastomosis), then the site of cross-clamp application must be excised. The duration of total circulatory arrest is usually less than 20 min and simplifies rather than adds to the complexity of the operation. By using the conservative method of valve repair with glue together with open arch repair, we have reduced the operative mortality of acute type A dissection to 6.5 per cent. Whilst others advocate a more aggressive approach, with routine replacement of the aortic root, the surgical mortality for this procedure in acute type A dissection is between 12 to 15 per cent.



Table 7 Causes of aortic regurgitation

We have recently performed an MRI study on patients who had previously undergone ascending aortic replacement for type A dissection: more than 50 per cent had a persistently patent false lumen with blood flow; in two patients, aneurysm formation within the affected aorta required replacement of the aortic root and aortic arch, respectively.

Type B dissection

Reluctance to operate on patients with type B dissection stems not from the surgical difficulties of graft insertion but from a comparatively high incidence of major complications, including paraplegia and acute renal tubular necrosis. The natural history of type B dissection is for the descending aorta to become progressively enlarged and aneurysmal, with the risk of rupture and sudden death (Fig. 29). Despite this, the Mayo Clinics report satisfactory long-term results for conservative treatment. Indications for surgery in type B dissection include compromise of a major aortic branch, formation of an acute saccular-type aneurysm, haemorrhage into the pleural cavity, inability to relieve or control pain within 4 h, and progressive increase in size of the dissecting haematoma on serial chest radiographs or CT scan.

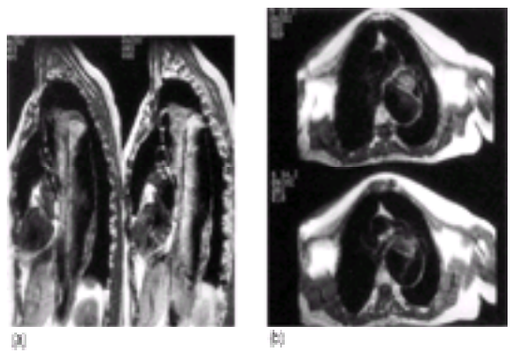


Fig. 29. (a) Longitudinal and (b) transverse MRI of chronic, aneurysmal type B dissection.

Cross-clamping of the descending thoracic aorta cuts off the segmental spinal collateral vessels from the intercostal arteries. Occlusion of the left subclavian artery causes a ‘steal' phenomenon and further reduces the blood supply to the spinal cord. When the descending thoracic aorta is open, intercostal steal takes blood away from the arteries of the spinal cord, particularly in the face of increased cerebrospinal fluid pressure (Fig. 30). Hypertension proximal to the aortic cross-clamp decreases the perfusion gradient, but the use of sodium nitroprusside to reduce hypertension causes an adverse rise in cerebrospinal fluid pressure from less than 10 mmHg to 20 mmHg, possibly by vasodilation in the intrathecal space. Nitroprusside also reduces the distal aortic pressure and may paralyse distal autonomic function. The combination of a reduced distal aortic pressure and increased intrathecal pressure decreases perfusion pressure in the spinal cord. Nitroprusside should therefore not be used to reduce afterload during aortic cross-clamping.

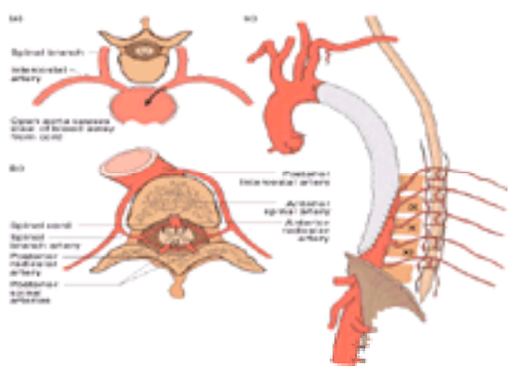


Fig. 30. Replacement of the descending thoracic aorta; preservation of spinal cord blood supply at the major radicular artery level.

In experimental studies, drainage of spinal fluid prevents the increase in its pressure during cross-clamping, and steroids appear to increase the protective effect. In clinical practice, regardless of protective efforts, the incidence of paraplegia following operations on the descending thoracic aorta ranges from 5 to 7 per cent in patients with coarctation or traumatic transection to 15 to 20 per cent in patients with extensive thoracoabdominal aneurysms. The highest rates of paraplegia (over 40 per cent) occur in patients with chronic type B dissection.

The descending thoracic aorta is exposed through the left chest with a large posterolateral incision in the fourth or fifth intercostal space. Thoracoabdominal aneurysms are exposed through a double incision, one in the fourth and one in the seventh intercostal space (Fig. 31). We are not aware of any definite evidence that normothermic shunt techniques protect against paraplegia, and consider that they prolong and complicate operative repair. At least eight different techniques have been described: this suggests that none has been entirely successful. The sites of cannulation for passive shunt bypass and roller or centrifugal pump bypass are shown in Table 4. These include left atrial–femoral bypass with or without an oxygenator, or non-pump bypass methods with a heparin-bonded shunt providing perfusion to the lower half of the body beneath the cross-clamp. For patients with extensive descending thoracic or thoracoabdominal aneurysms brought electively to the operating theatre we use hypothermic cardiopulmonary bypass with a central cannulation technique (Fig. 32).



Fig. 31. Double-incision technique for exposure of the whole descending thoracic and abdominal aorta: (a) after closure, (b) open.

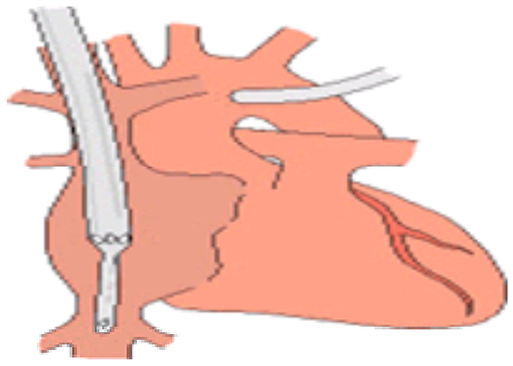


Fig. 32. Cannulation technique for descending thoracic aortic replacement with profound hypothermia.

The ascending aorta is located at the pericardial reflection and cannulated. A right-angled venous cannula is then placed into the main pulmonary artery and through the valve into the right ventricle for systemic venous return. Using this approach, standard cardiopulmonary bypass with cooling and where necessary total circulatory arrest can be employed for distal aortic arch and thoracoabdominal surgery. This method also avoids unprotected cross-clamping of the aorta, which may lead to subendocardial ischaemia and left ventricular failure in patients with ischaemic heart disease. If resection of the distal aortic arch is required, this is performed first with a short period of circulatory arrest. The vascular graft is then clamped and perfusion resumed to the brachiocephalic and coronary arteries, which allows an unhurried repair of the thoracic or thoracoabdominal aorta with the abdominal viscera protected by hypothermia. Continuous flow to the brain may also protect against paraplegia by ensuring flow to the anterior spinal artery. In contrast, other methods have distinct disadvantages. Femorofemoral bypass cannot provide blood flow to the head or

coronary circulation with the aortic cross-clamps in place, and in patients with aortoiliac disease the flow rates may be limited. Femoral perfusion may also disseminate debris from an atheromatous aorta into the brachiocephalic vessels, causing stroke.

During cooling the heart fibrillates and care is taken to check for left ventricular distension. The method is not suitable for patients with significant aortic regurgitation. When profound hypothermia is secure, the aortic cross-clamps are applied and the rate of perfusion reduced to approximately 0.8 l/min per m^2 to supply continued blood flow to the coronary and brachiocephalic vessels. With the aorta open, blood is retrieved from the operative site using a cell saver, since the depth of blood in the chest is seldom sufficient for cardiotomy suction. Short periods of low flow or circulatory arrest can be employed to control collateral intercostal bleeding. The diseased aorta is opened and those vessels requiring reimplantation are identified and mobilized. For type B dissection, the cross-clamps are usually applied between the left carotid and subclavian vessels and at the level of the fifth or sixth thoracic vertebra. The intimal tear is usually located distal to the left subclavian artery. This part of the aorta is resected and replaced with an impervious Dacron graft. Bleeding intercostal vessels must be controlled from within the aortic lumen. If the operation is performed at normothermia without bypass techniques, it should be completed so that the aortic cross-clamps can be released within 30 min to minimize the risk of ischaemia in the spinal cord. For thoracoabdominal aneurysm the aorta is replaced with appropriate lengths of Hemashield into which the visceral vessels are reimplanted. Attempts are made to identify and reimplant the principal spinal radicular artery and selected, large, proximal intercostal vessels. Profound hypothermia is extremely useful for protection of the myocardium, brain, and visceral vessels during extensive thoracoabdominal replacement. However, two of eight patients undergoing extensive thoracoabdominal replacement have suffered paraplegia despite profound hypothermia and reimplantation of the principal spinal radicular vessels; this suggests that the anatomy of the spinal cord is the key factor in paraplegia and that hypothermic protection may not be as effective as we had hoped.

Descending thoracic and thoracoabdominal aneurysms

These occur predominantly through atherosclerosis, previous aortic dissection, or syphilis. Because of the site and extent of these aneurysms, there is a high incidence of paraplegia following their repair. The methods described for type B aortic dissection apply directly to well-defined aneurysms of the descending aortic segment. However, few surgeons can replace the thoracoabdominal aorta with implantation of all visceral vessels into the graft within the safe duration of simple aortic cross-clamping (Fig. 33). Profound hypothermia with total circulatory arrest affords protection to the spinal cord and minimizes blood loss. In order to reduce the risk of paraplegia it is important to identify and reimplant the principal spinal radicular artery. In atherosclerotic disease it is often necessary to endarterectomize the coeliac axis, mesenteric artery, or renal vessels. Where the origins of two branch arteries are in close proximity, a button of aorta encompassing both may be sutured into the main graft. This technique is commonly used for the right renal artery, together with the coeliac axis or superior mesenteric artery. The left renal artery is reimplanted separately. Care must be taken not to damage the renal veins or inferior vena cava during these extensive operations. The general operative approach to suprarenal and thoracoabdominal aneurysms must be highly individualized, depending on the local anatomical circumstances. The overall morbidity and mortality associated with elective resection of thoracoabdominal aneurysm has substantially diminished with improved techniques and vascular grafts. The operative mortality is related to the age of the patient and other underlying problems such as ischaemic heart disease, carotid occlusion, chronic obstructive airways disease, and pre-existing renal failure. Infection, paraplegia, and anastomotic aneurysms have been the primary postoperative complications. The long-term prognosis after successful resection is surprisingly favourable, and depends substantially on the overall condition of the patient, many of whom have concomitant coronary, carotid, and obstructive lung disease.

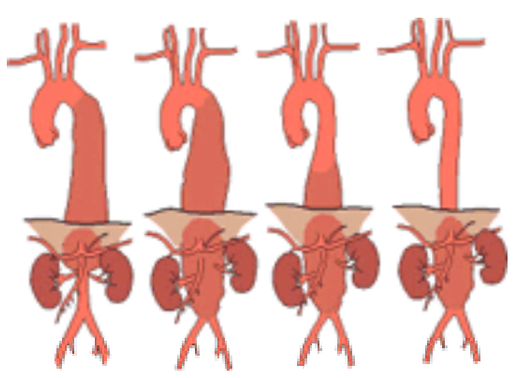


Fig. 33. Classification of thoracoabdominal aneurysms.

Coarctation of the aorta in adults

Coarctation of the aorta is a congenital narrowing of the upper descending thoracic aorta adjacent to the ligamentum arteriosum. Sometimes it is combined with more proximal aortic narrowing. Uncommonly, coarctation occurs more proximally between the left common carotid and subclavian arteries or distally in the descending thoracic aorta. Occasionally, the adult aorta may be redundant and severely kinked opposite the ligamentum arteriosum without any pressure gradient (so-called pseudocoarctation lesion). The lesion of classical coarctation is a shelf or enfolding of the aortic media into the lumen. The shelf is usually marked externally by a localized indentation or wasting of the aortic wall. The aorta beyond the narrowing usually shows post-stenotic dilatation. Collateral circulation develops between the aorta proximal to the narrowing and the distal vessel. When well developed this is responsible for some of the signs, such as parascapular pulsations and rib notching. The enlarged, tortuous, third and fourth intercostal arteries may become aneurysmal in adults. A bicuspid aortic valve occurs in association in one-third of patients. Many adolescent and young adult patients remain asymptomatic, but often have short legs. They are diagnosed when delayed or absent femoral pulses and systemic hypertension are detected on routine examination. Heart failure appears at about 30 years of age and is preceded by effort dyspnoea, cardiomegaly, and left ventricular hypertrophy on the electrocardiogram. The radiographic findings include a ‘figure 3' sign in the left upper mediastinal shadow, and almost always rib notching. There is a known association between Turner syndrome and coarctation of the aorta. The diagnosis can be confirmed by MRI, echocardiography, or invasively by catheterization and aortography. If left untreated, 35 per cent of patients die before the third decade and a further 25 per cent die during both the fourth and fifth decades; 98 per cent are dead by the age of 60 years. Causes of mortality include heart failure, bacterial endocarditis, aortic dissection, or rupture of associated intracranial aneurysms. Surgery is usually undertaken without perfusion techniques since the surgical options of resection with end-to-end anastomosis, diamond patch aortoplasty, or Dacron graft bypass can usually be done in less than 30 min. However, in adults the intercostal arteries are large and friable, and surgical dissection is hazardous. The use of controlled hypotension by the anaesthetist is important, but once the aortic clamps are in place the upper-body blood pressure is allowed to rise moderately (90–100 mmHg) to promote collateral blood flow to the spinal cord. When an aneurysm of the aorta or intercostal vessels is present, resection of the segment of the aorta along with the coarctation is required, and continuity is re-established with an interposed woven Dacron graft. Sacrifice of intercostal arteries increases the likelihood of paraplegia. Nevertheless, paraplegia occurs in less than 5 per cent of adult patients with coarctation, probably because the blood pressure in the distal aorta is maintained by collateral vessels during cross-clamping. Postoperatively, systemic hypertension persists in 50 per cent of patients and requires prolonged treatment.

Monitoring of spinal cord function

Known risk factors for injury to the spinal cord include prolonged cross-clamp time (over 30 min), increased cerebrospinal fluid pressure, which may impair cord perfusion, low distal aortic pressure, loss of critical intercostal or lumbar arteries, and the use of nitrate vasodilators. The perfusion pressure to the spinal cord below the cross-clamps depends on the value of the distal aortic pressure minus the intrathecal pressure. Theoretically, shunt techniques could protect the cord by maintaining distal aortic pressure. Drainage of spinal fluid may also prevent an increase in intrathecal pressure during cross-clamping. Blood flow to the spinal cord is partly dependent on regional vascular resistance and partly on blood pressure. Perfusion pressure can be increased by either raising the diastolic blood pressure or reducing the cerebrospinal fluid pressure. Nevertheless, Crawford and colleagues have shown a high incidence of paraplegia in patients with a high mean distal aortic pressure (80–110 mmHg) during bypass for replacement of the descending thoracic aorta.

Recording somatosensory-evoked potentials can be used to monitor conduction in the spinal cord during aortic cross-clamping. This technique uses an electrical stimulus applied to the leg or foot to evoke cerebral potential tracings, which are analysed by computer to determine the brain's electrical response (Fig. 34); this is different from conventional electroencephalographic tracing, in which spontaneous rather than provoked electrical activity is recorded from the cerebral cortex. The evoked potentials are only sensitive to the function of the dorsal (posterior) spinal cord; the ventral or anterior portion of the cord is at greatest risk of developing ischaemia. Although the use of motor-evoked potentials has been proposed as a warning sign for paraplegia during clinical surgery, the awareness of an impending cord problem is unlikely to expedite removal of the cross-clamp. In dogs, however, the method has been used to demonstrate that function of the spinal cord can be maintained for up to 1 h of normothermic cross-clamping, provided the mean distal aortic pressure is above 60 mmHg.

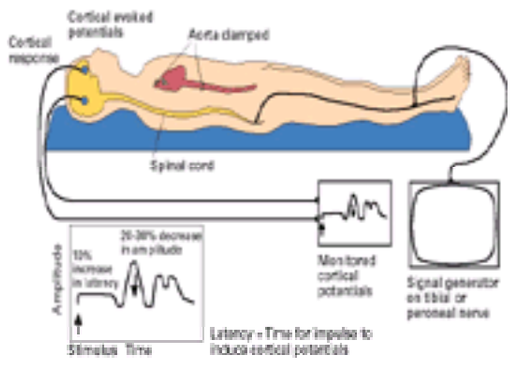


Fig. 34. Method for monitoring spinal cord integrity by cortical-evoked potentials: flow reversal causes thrombosis in the blind loop; transmission of a stimulus from the leg to the cerebral cortex confirms cord function.

Failure to reattach important blood vessels of the spinal cord after extensive thoracic or thoracoabdominal resection may cause paraplegia. The principal spinal radicular artery can sometimes be identified at operation because of its size. Methods of intraoperative identification would be of great value but those currently available are complex and not applicable to emergency operations in patients with type B dissection or ruptured thoracic aneurysms. Currently, hypothermic techniques, possibly combined with pharmacological protection from an appropriate calcium-channel blocker and barbiturate, offer the best methods of cord protection. Attempts should be made to preserve the principal spinal radicular artery: this is sometimes possible by creating a long, oblique, distal anastomosis to protect the intercostal and lumbar vessels between T10 and L2.

The thromboexclusion technique

In 1981, Carpentier's group described a thoracic aortic bypass technique for treatment of type B aortic dissection (Fig. 35). The method was designed to reduce morbidity and mortality by reducing blood loss and the risk of paraplegia. As originally described the method did not require cardiopulmonary bypass and is not subject to the problems of proximal hypertension and distal hypoperfusion through unprotected aortic cross-clamping. The operation does not involve a direct attack on the descending thoracic aorta. A median sternotomy incision extended down into the abdomen allows the descending aorta to be bypassed with a graft from the ascending to the infrarenal abdominal aorta. The native aortic lumen is then occluded just distal to the left subclavian artery. When the aorta is interrupted there is a reversal of blood flow in the descending thoracic aorta (Fig. 36); this results in retrograde flow into the aneurysm, causing considerable turbulence and progressive thrombosis of the descending thoracic aorta down to the first major abdominal branch, the coeliac axis. Thromboexclusion firstly protects the diseased aortic segment from rupture and then gradually occludes it by thrombosis. In the meantime, because the process is slow, collaterals develop and protect the spinal cord. Apart from Carpentier's initial report there has been radiological documentation of this thrombotic process in only two other patients. In general the method has not been adopted and its usefulness is underestimated.

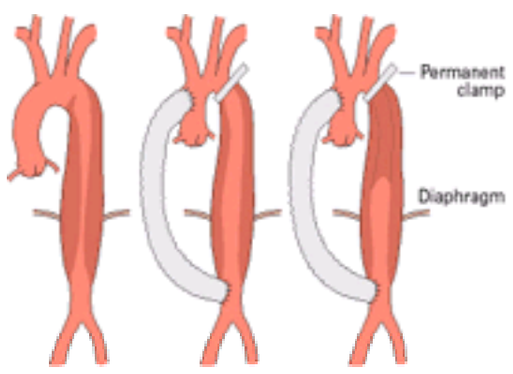


Fig. 35. Thromboexclusion technique for complex descending thoracic pathology; the patient required coronary grafts in addition (see Fig. 37).

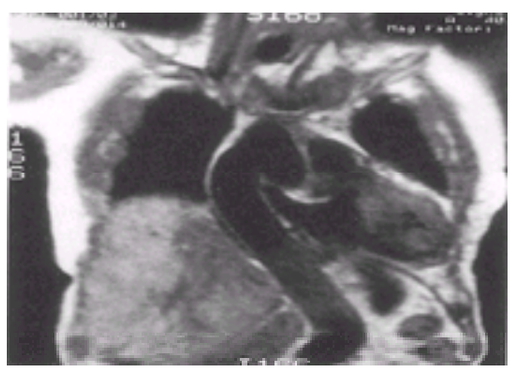


Fig. 36. MRI scan after thromboexclusion procedure; the ascending abdominal aortic graft is clearly seen.



Fig. 37. MRI of the descending aorta in a chronic type B dissection.

We have recently used thromboexclusion combined with coronary arterial bypass surgery in a 62-year-old Jehovah's witness who was restricted to a wheelchair by pain from a progressively expanding, type B dissection (Fig. 37). He had also suffered previous myocardial infarction with impairment of left ventricular function. On cardiopulmonary bypass the left anterior descending coronary was grafted and then, using reduced pressure and flow, a 30-cm Hemashield graft was anastomosed to the infrarenal abdominal aorta and brought through the diaphragm to the right side of the ascending aorta. The operation was completed by stapling the distal aortic arch. There was no significant bleeding in this patient and the postoperative haemoglobin was 10.6 g/dl. He made an uneventful recovery. Encouraged by this success we then applied the method to a 60-year-old woman with a massive, syphilitic, descending thoracic aneurysm who had previously undergone replacement of the ascending aorta for a ruptured aneurysm. In this patient the proximal end of the bypass graft was anastomosed to the Hemashield ascending aortic graft. In addition to its application for patients with acute type B dissection as advocated by Carpentier, the process for thromboexclusion may well prove applicable to patients with

massive descending thoracic aneurysms where the risks of paraplegia are considered prohibitive.

Aortoannular ectasia

These patients usually present with aortic regurgitation due to dilatation of the aortic annulus and are found to have an ascending aortic aneurysm involving the aortic sinuses (Fig. 38) (Table 7). Non-invasive diagnostic techniques such as two-dimensional echocardiography, CT, and nuclear magnetic scanning provide excellent imaging of size and extent of the aneurysm and of left ventricular function. Coronary angiography is usually used to rule out the presence of occlusive disease in patients over 40 years of age. Combined replacement of the aortic valve and ascending aorta is required, with reimplantation of the coronary ostia into the graft (Bentall procedure; Fig. 39). Composite proprietary valve conduits and antibiotic-sterilized aortic homografts are used, and with improvements in myocardial protection the operative mortality is now between 5 and 10 per cent. In Oxford the wrap-around inclusion technique has been superseded by excision of the pathological aorta and mobilization of the coronary arteries before reimplantation into the graft. Bleeding has largely been eliminated by electively sealing the suture lines with acrylic (Histoacryl; Braun) or resorcinol-based glues.

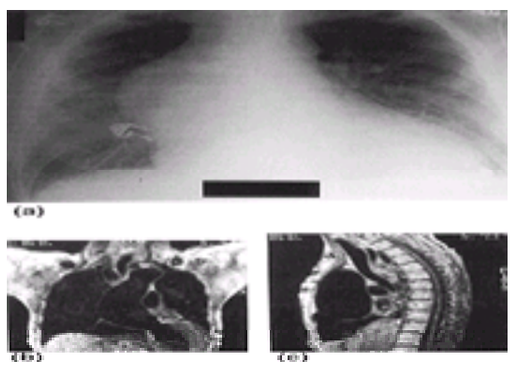


Fig. 38. Aortoannular ectasia: (a) plain chest radiograph showing a large ascending aortic aneurysm; (b) MRI scan showing an ascending aortic aneurysm extending into the arch;(c) nuclear magnetic resonance scan, lateral view.

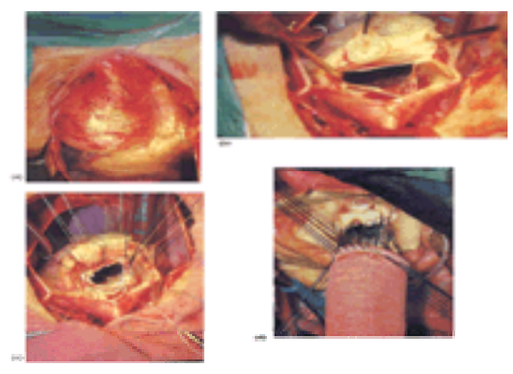


Fig. 39. Aortic root replacement for aortoannular ectasia. (a) The aneurysm. (b) The aneurysm open showing a very dilated aortic annulus; the right coronary ostium is mobilized pending reimplantation into the graft. (c) Interrupted sutures in place for implantation of the valve prosthesis composite graft. (d) Graft lowered into place.

Aneurysms of the ascending aorta

MRI and three-dimensional reconstruction of CT images have improved our understanding of the natural history of thoracic aneurysm. The second technique is able to define specific irregularities in the aortic wall that may predict rupture. Ascending aortic aneurysms are known to increase in diameter by 1.5 cm/year, three times faster than an infrarenal aortic aneurysm. Descending aortic aneurysms also increase in size faster than abdominal aortic aneurysms, but in addition expand by elongation. Risk factors for rupture of a thoracic aneurysm include size, elevated diastolic blood pressure, and the presence of obstructive pulmonary disease. Indications for surgery in thoracic aneurysms are pain, an ascending aortic dilation to more than 5 cm in patients with Marfan syndrome, a diameter above 8 cm in patients with other pathology, and for all saccular aneurysms, where the nature of the defect implies significant wall weakness. Aneurysms between 6 and 8 cm (twice normal size) can be managed conservatively, with 3- to 6-monthly CT or MRI scanning. Surgery is advised if the aneurysm reaches 8 cm.

Aneurysms that do not involve the aortic sinuses can be replaced with a straight tube graft, using cardiopulmonary bypass with or without total circulatory arrest, according to the site of the distal anastomosis. When the graft involves only the ascending aorta the surgical mortality is between 5 and 10 per cent. For operations involving the ascending aorta together with the aortic arch the mortality increases to more than 20 per cent.

Aortic arch aneurysms

Surgery for lesions of the aortic arch (Fig. 40) has been greatly simplified by the techniques of profound hypothermia and total circulatory arrest. Separate cannulation and continuous perfusion of the cerebral vessels were previously required. Using median sternotomy and total circulatory arrest without aortic cross-clamps, the arch can be replaced with a tube graft. Implantation of the innominate, left carotid, and left subclavian vessels is performed *en bloc* on an oval button of aorta. With the circulation arrested, the distal anastomosis to the descending thoracic aorta is performed first, followed by implantation of the brachiocephalic vessels. An aortic cross-clamp can then be applied to the graft and perfusion to the head and body recommenced. The proximal anastomosis to the ascending aorta is undertaken during the rewarming phase of perfusion. Care must be taken to remove air from the cerebral vessels. During the procedure the myocardium is protected by cold potassium cardioplegia.

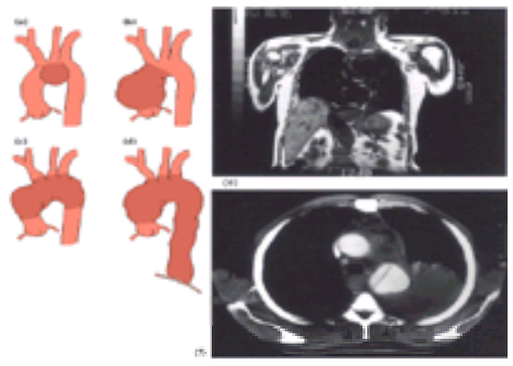


Fig. 40. Aneurysms of the aortic arch: (a) saccular; (b) atheromatous ascending aneurysm extending into the arch; (c) atheromatous arch aneurysm; (d) chronic type A dissection; (e) MRI scan of atheromatous arch aneurysm; (f) rupture of the aortic arch in chronic type B dissection.

Despite improvements in technique, the mortality rate associated with replacement of the aortic arch in most centres remains between 20 and 30 per cent. This may be due to the rarity with which the operation is performed.

Further reading

Butler J, Ormerod OJM, Giannopoulus N, Pillai R, Westaby S. Diagnostic delay and outcome in surgery for Type A aortic dissection. *Quarterly Journal of Medicine* 1991; **89**: 391–6.

Cabrol C *et al.* Long term results with total replacement of the ascending aorta and reimplantation of the coronary arteries. *Journal of Thoracic and Cardiovascular Surgery* 1986; **91**: 17–25.

Crawford ES, Crawford JL. *Diseases of the aorta including an atlas of angiographic pathology and surgical technique*. Williams and Wilkins, Baltimore, 1984.

Crawford ES, Stoe CL, Crawford JL, Titus JL, Weilbaecher DG. Aortic arch aneurysm: a sentinel of extensive aortic disease requiring sub-total and total aortic replacement. *Annals of Surgery* 1984; **199**: 742–52.

Crawford ES *et al.* Thoraco-abdominal aortic aneurysms: pre-operative and intra-operative factors determining immediate and long term results of operations in 605 patients. *Journal of Vascular Surgery* 1986; **3**: 389–404.

Crawford ES, Svenssen LG, Cosseli JS, Safi HJ, Hess KR. Aortic dissection and dissecting aortic aneurysm. *Annals of Surgery* 1988; **208**: 254–73.

Gott VL, Pyeritz RE, Magovern GJ, Cameron DE, McKusick VA. Surgical treatment of aneurysms of the ascending aorta in the Marfan syndrome. Results of composite graft repair in 50 patients. *New England Journal of Medicine* 1986; **314**: 1070–4.

Kouchoukos NT, Marshall WG, Wedige-Stecher TA. Eleven year experience with composite graft replacement of the ascending aorta and aortic valve. *Journal of Thoracic and Cardiovascular Surgery* 1986; **92**: 691–705.

Lytle BW, Mafhood SS, Cosgrove DM, Loop FD. Replacement of the ascending aorta. Early and late results. *Journal of Thoracic and Cardiovascular Surgery* 1990; **99**: 651–8.

Massimo CF *et al.* Extended and total aortic resection in the surgical treatment of type A. Aortic dissection: experience with 54 patients. *Annals of Thoracic Surgery* 1988; **46**: 420–4.

Okita Y, Franciosi G, Matsuki O, Robles A, Ross DN. Early and late results of aortic root replacement with antibiotic sterilised aortic homograft. *Journal of Thoracic and Cardiovascular Surgery* 1988; **95**: 696–704.

40.11 Surgery for cardiac rhythm disturbances

T. Bruce Ferguson, Jr.

Introduction
The Wolff–Parkinson–White syndrome and other accessory atrioventricular connections (AV re-entrant tachycardias)
Surgical anatomy
Intraoperative electrophysiologic mapping
Surgical techniques
Clinical results
Paroxysmal supraventricular tachycardia due to atrioventricular nodal re-entry (AV nodal re-entrant tachycardias)
Surgical anatomy
Surgical techniques
Clinical results
Automatic atrial (ectopic) tachycardias
Surgical anatomy
Surgical techniques
Clinical results
Atrial flutter and fibrillation
Atrial flutter: surgical anatomy, electrophysiology, and techniques
Atrial fibrillation: surgical anatomy, electrophysiology and techniques
Surgery for ischemic and non-ischemic ventricular tachycardia
Non-ischemic ventricular tachycardias
Ischemic ventricular tachycardia
Conclusions
Further reading

Introduction

The direct surgical treatment of cardiac rhythm disturbances began in 1968, when the first successful division of the accessory pathway responsible for the Wolff–Parkinson–White syndrome was performed by the ‘father' of arrhythmia surgery, Dr Will Sealy. Since then a number of surgical procedures have been developed to cure both supraventricular and ventricular tachyarrhythmias. These operations have been shown to be effective and often they are preferred over what may be a lifetime of drug therapy for many of these patients.

More recently, the development of techniques for radiofrequency ablation of re-entrant supraventricular arrhythmias has severely limited the requirement for surgical intervention. Nevertheless, the lessons derived from the anatomic, electrophysiologic, and surgical principles of these procedures are useful for the practicing cardiothoracic surgeon to understand.

The Wolff–Parkinson–White syndrome and other accessory atrioventricular connections (AV re-entrant tachycardias)

Candidates for surgical ablation of accessory pathways now include those with recurrent reciprocating tachycardia in whom one or more attempts at radiofrequency ablation have failed. A second group of candidates are those with Wolff–Parkinson–White syndrome who have symptomatic supraventricular arrhythmias and are undergoing cardiac surgery for other reasons. Previously, patients who were poorly controlled medically or had developed significant toxicity to an otherwise successful medical regimen, and those with accessory pathways and atrial fibrillation or flutter who had an excessively rapid ventricular rate, were primary candidates for surgery; the major indication for surgical intervention was refractoriness to medical treatment.

Because of the infrequency of these procedures, they should be performed by a surgeon with extensive experience of them. This is particularly true for the failed radiofrequency ablation of other, more complex, accessory atrioventricular connections (atrio–His, nodoventricular, and fasciculoventricular fibres).

Surgical anatomy

In the Wolff–Parkinson–White syndrome there is a congenital, abnormal muscular connection between the atrium and ventricle, located somewhere in the atrioventricular groove. Anatomically, the heart can be sectioned at the level of the atrioventricular groove and divided into four discrete areas defined in this horizontal plane (Fig. 1). These four areas are the left free wall, the right free wall, and the posterior septal and the anterior septal spaces.

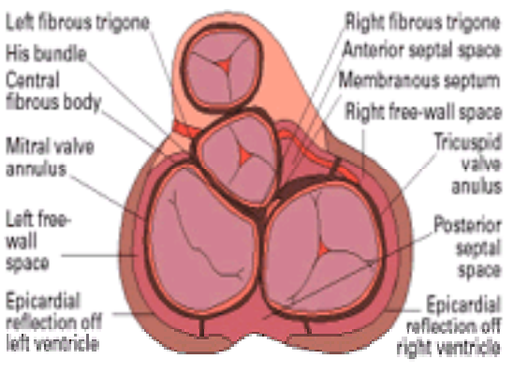


Fig. 1. Superior view of the heart with the atria cut away, demonstrating the boundaries of each of the four anatomic spaces where accessory pathways can occur in the Wolff–Parkinson–White syndrome. The left free-wall space extends from the left fibrous trigone anteriorly to the ventricular septum posteriorly and is bounded medially by the mitral valve annulus and laterally by the epicardial reflection off the free wall of the left ventricle. The posterior septal space extends from the junction of the mitral and tricuspid valves anteriorly to the epicardial reflection off the right and left ventricles posteriorly. The left lateral boundary of the posterior septal space is the mitral valve annulus and the free wall of the left ventricle. The right lateral boundaries are the tricuspid valve annulus and the free wall of the right ventricle. The right free-wall space extends from the anterior to the posterior septum and is bounded medially by the tricuspid valve annulus and laterally by the epicardial reflection off the free wall of the right ventricle. The anterior septal space extends from the membranous portion of the interatrial septum to the epicardial reflection off the anterior right ventricle. It is bounded medially by the root at the aorta and laterally by the free wall of the right ventricle. Accessory atrioventricular connections responsible for the Wolff–Parkinson–White syndrome cannot insert on to either ventricle anywhere outside the stippled areas shown in this diagram.

The annuli of the mitral and aortic valves contribute significantly to the structural integrity of the fibrous skeleton and are further strengthened at their left junction to form the left fibrous trigone (Fig. 2). The atrioventricular groove between the left and right fibrous trigones (the anterior portion of the central fibrous body) represents the site of continuity between the anterior leaflet of the mitral valve and the annulus of the aortic valve. This is the only area in the atrioventricular groove where atrial muscle is not in juxtaposition to ventricular muscle, and for this reason, accessory atrioventricular pathways are not found between the left and right fibrous trigones.

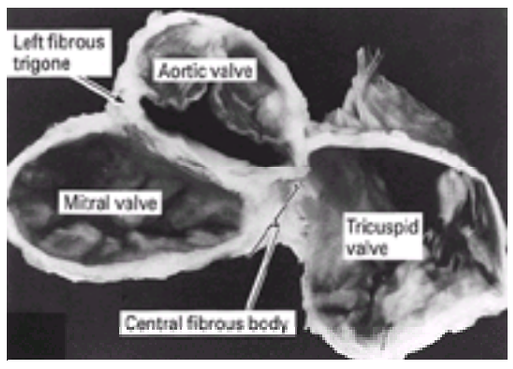


Fig. 2. Dissection showing the fibrous cardiac skeleton after it has been removed from the ventricles.

The preoperative electrophysiologic study by catheter and the intraoperative mapping are both directed towards localizing an accessory pathway to one of these four anatomic spaces. In order to ablate surgically a pathway in one of these regions with success approaching 100 per cent, the entire anatomic space in which the pathway resides must be dissected in every patient. Paradoxically, the experience with radiofrequency ablation by catheter has confirmed both the microscopic size and the variable anatomic position of these pathways, while demonstrating the efficacy of accurate localization and pinpoint ablation.

Electrophysiologically, accessory pathways are the equivalent of an electrical cable that is capable of conducting electrical impulses between the atrium and ventricle. Histologically, these pathways resemble normal atrial myocardium. Because of the anatomic limitations imposed by the valve annulus on the inside of the atrioventricular groove and the epicardium on the outside of the atrioventricular groove in the vertical plane, all accessory pathways must connect atrium to ventricle somewhere between these two boundaries (Fig. 3). In the vertical plane, therefore, accessory pathways are confined to sites adjacent to or near the valve annulus, within the fat pad of the atrioventricular groove, or just beneath the epicardium overlying that groove. The ‘depth' of the pathway within the groove in the vertical plane must be considered to be variable. In addition, when the horizontal and vertical planes are combined, it is important to understand that these accessory pathways can traverse tangentially this three-dimensional space.

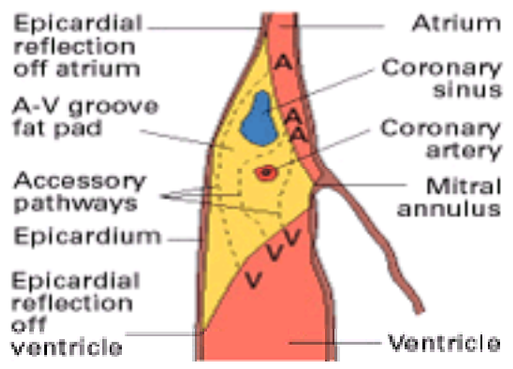


Fig. 3. Cross-section of the posterior left heart showing the different depths that left free-wall pathways can be located in relation to the mitral annulus and the epicardium overlying the atrioventricular groove. Note that regardless of the depth of the accessory pathway in the vertical plane of the groove, the atrial end of the accessory pathway must attach to the atrium somewhere between the annulus of the mitral valve and the epicardial reflection off the atrium. Similarly, the ventricular end of the accessory pathway must connect to the ventricle somewhere between the annulus of the mitral valve and the epicardial reflection off the ventricle.

Intraoperative electrophysiologic mapping

Epicardial pacing and sensing electrodes are sutured on to the atrium and ventricle near the suspected site of the accessory pathway. These electrodes are used for timing reference.

The complexity of the mapping instruments can vary from a multichannel, computerized system to a hand-held, single electrode. The principles for all types of systems are the same, however. First the ventricular side of the atrioventricular groove is mapped for the site of earliest activation in sinus rhythm and during atrial pacing to assess anterograde ventricular activation during maximal pre-excitation. Data are then recorded from the atrial side of the atrioventricular groove, and retrograde atrial activation is assessed during orthodromic reciprocating tachycardia induced with programmed electrical stimulation or with ventricular pacing. Only a few cycles of tachycardia are allowed to occur since hemodynamic compromise is common and the patients are not on cardiopulmonary bypass during mapping. Atrial data are especially important because they demonstrate unsuspected, concealed accessory pathways, undetected until this point in the mapping (Fig. 4).

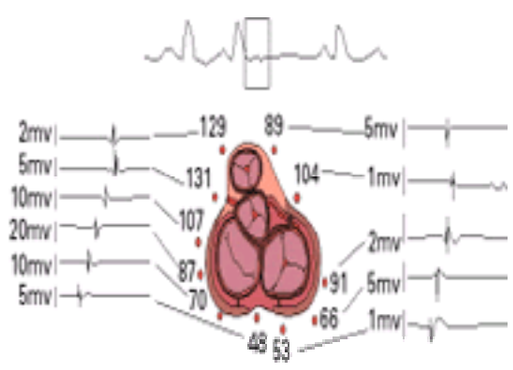


Fig. 4. Retrograde atrial epicardial activation times during single atrioventricular echo beat: an accessory pathway is located in posterior septum, where the activation time is 48 ms.

In general, rapid identification of multiple pathways and pathways manifesting intermittent conduction is greatly facilitated by computerized intraoperative mapping systems. In most circumstances, complete atrial and ventricular mapping and data analysis can be completed in approx. 10 min with these systems, and only a single beat is required to analyze ventricular or atrial activation times.

Surgical techniques

The anatomy of accessory atrioventricular connections may be classified into the four sites described earlier. In decreasing order of frequency, accessory pathways are located in the left free-wall, posterior septal, right free-wall, and anterior septal positions. Approximately 20 per cent of patients in a series accumulated at Washington University in St. Louis between 1983 and 1996 had multiple (two to four) pathways.

Two surgical approaches have been commonly employed to divide accessory atrioventricular connections. The endocardial technique is designed to divide the ventricular end of the accessory pathway, and the epicardial technique is directed toward division of the atrial end (Fig. 5). Thus the surgical objective is to divide the accessory connection, either at the atrial or the ventricular end of the pathway. Excellent results may be obtained with both techniques.

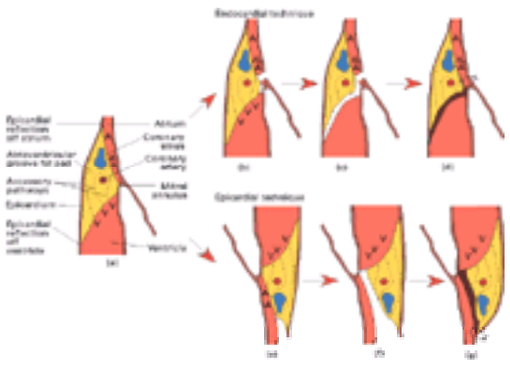


Fig. 5. Comparison of the surgical techniques used with the endocardial (b–d) and the epicardial approaches (e–g). The endocardial technique results in division of the atrial end of accessory pathways. The epicardial technique is drawn with the ventricle above and the atrium below since it is necessary to retract the apex of the heart out of the pericardium for surgical dissection on the outside of the posterior left atrium. It is for this reason that the epicardial technique requires cardiopulmonary bypass and, therefore, should not be classified as a closed-heart procedure.

Many centers have employed an endocardial approach and an anatomically based operation for division of all accessory pathways (Fig. 6, Fig. 7). The principles of this operative approach are: (1) accurate intraoperative localization of the pathway(s) to one of the four anatomic areas in the horizontal plane; (2) appreciation that the site of the pathway in the vertical plane may be variable; (3) appreciation that the endocardial dissection divides the ventricular insertion of the pathway and does nothing to its atrial insertion; (4) complete dissection of the appropriate anatomic space(s) in every patient regardless of the site of the pathway within that space as determined by intraoperative mapping; (5) appreciation that certain pathways may exist as ‘broad bands' and that when the ventricular insertion site is located at the junction of two anatomic areas, complete dissection of both anatomic spaces should be performed; (6) that isolation of the atrial rim of tissue above the annulus of the valve is important to prevent a juxta-annular pathway from activating the atrium retrogradely.

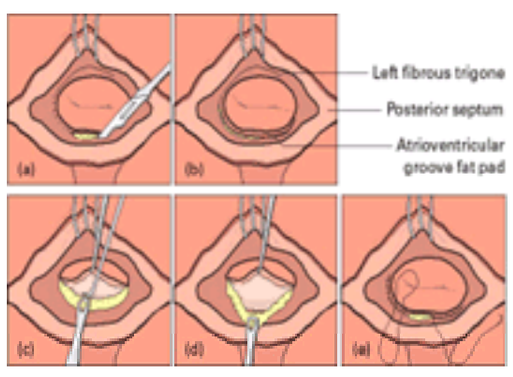


Fig. 6. Surgeon's view of the endocardial technique for dividing left free-wall accessory pathways in the Wolff–Parkinson–White syndrome. Complete dissection of the posterior interatrial groove is performed, with mobilization of the superior and inferior cava. This permits access to the superior aspect of the left atrium, necessary for adequate exposure in these patients with normal-sized left atria. A left atriotomy is performed after the aorta has been cross-clamped and the heart is arrested with cold potassium cardioplegic solution. A supra-annular incision is made 2 mm above the annulus of the mitral valve, extending from the left fibrous trigone anteriorly to the junction of the left free wall and septum, directly posterior to the medical commissure of the mitral valve. (a) A plane of dissection is established between the fat pad underlying the atrioventricular groove and the top of the posterior left ventricle throughout the length of the supra-annular incision. (b) This plane of dissection is carried to the epicardium as it reflects off the posterior left ventricle on to the fat pad of the atrioventricular groove. (c, d) The ends of the supra-annular incision are then ‘squared-off' to the level of the mitral annulus, either by surgical dissection or by placement of a cryolesion at either end of the incision; and (e) then closed.

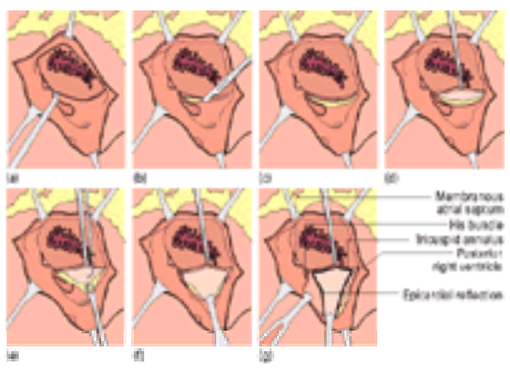


Fig. 7. Surgeon's view of the endocardial technique for dividing posterior septal accessory pathways in the Wolff–Parkinson–White syndrome. (a) After a right atriotomy, the His bundle may be identified with the hand-held probe. (b) The heart is arrested with cold cardioplegic solution after the aorta has been cross-clamped. A supra-annular incision is placed well posterior to the His bundle and continued counter-clockwise on to the free wall of the right atrium. A plane of dissection is established between the fat pad and the top of the posterior interventricular septum. This plane is developed in the anterior portion of the posterior septal space closest to the His bundle, approaching the central fibrous body from the posterior direction (c, d, e). Alternatively, this portion of the dissection may be performed before the cardioplegic arrest, during either atrial pacing or orthodromic reciprocating tachycardia, with continuous monitoring of the atrioventricular interval to avoid injury to the atrioventricular node–His bundle complex. The dissection is carried medially to the annulus of the mitral valve and posteriorly to the epicardial reflection off the posterior ventricular surface. (f, g) The left corner of the pyramidal space overlying the posterosuperior process of the left ventricle at the site where the posterior interventricular septum is juxtaposed to the posterior left ventricular free wall is carefully dissected to complete the procedure.

Clinical results

In a series of over 330 patients operated upon for the Wolff–Parkinson–White syndrome and/or other accessory pathways by Cox and Ferguson between 1981 and 1997, surgical correction of accessory atrioventricular connections using the techniques described was successful in nearly all patients at the initial operation, with an operative mortality for elective, uncomplicated cases of 0.5 per cent. There were no late recurrences after surgery using the endocardial technique. Similar results were obtained with the epicardial dissection technique as described by Guiraudon *et al.* for right free-wall, left free-wall, and certain posterior septal pathways.

Paroxysmal supraventricular tachycardia due to atrioventricular nodal re-entry (AV nodal re-entrant tachycardias)

Re-entry within the atrioventricular node is the most common cause of paroxysmal supraventricular arrhythmia in adults. The anatomic and electrophysiologic reason for this arrhythmia is the presence of ‘dual atrioventricular nodal conduction pathways', one fast and one slow, through the perinodal tissues.

Surgical anatomy

The anatomy of the triangle of Koch is the important consideration in the surgical treatment of atrioventricular nodal re-entrant tachycardia. This triangle is bounded by the tendon of Todaro superiorly, the annulus of the tricuspid valve inferiorly, and the os of the coronary sinus posteriorly. The membranous septum marks the anterior tip of the triangle, and within this area lie the compact atrioventricular node, the perinodal tissues, and the bundle of His (Fig. 8).

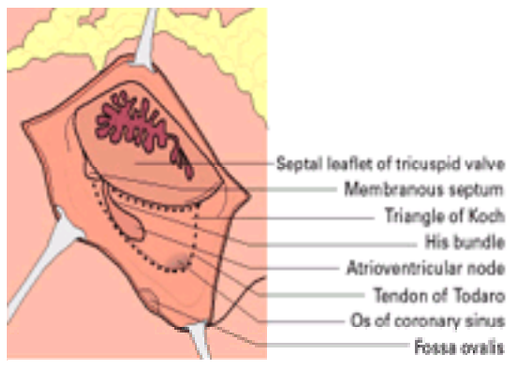


Fig. 8. The right atrial septum viewed through a longitudinal right atriotomy. The patient's head is to the left and feet are to the right. The boundaries of the triangle of Koch are the tendon of Todaro, the annulus of the tricuspid valve, and a line connecting the two at the level of the os of the coronary sinus. Within the triangle of Koch resides the atrioventricular node and the proximal portion of the His bundle, which enters the ventricular septum immediately posterior to the membranous portion of the interatrial septum.

Surgical techniques

Surgical exposure for the discrete cryosurgical technique is the same as for a dissection of a posterior septal pathway. Cryosurgical modification of the perinodal tissue to interrupt the electrophysiologic substrate for atrioventricular nodal re-entrant tachycardia is the goal of this surgical approach.

During application of the cryolesions the patient is paced atrially at a constant cycle length and atrioventricular conduction is monitored beat to beat. After setting the pacing and recording variables, a nitrous oxide cryoprobe with a 3-mm tip is used to place lesions along the tendon of Todaro at a temperature of -60°C for 2 min, or until transient heart block occurs. These four cryolesions extend from the os of the coronary sinus to the apex of the triangle (Fig. 9(a)).

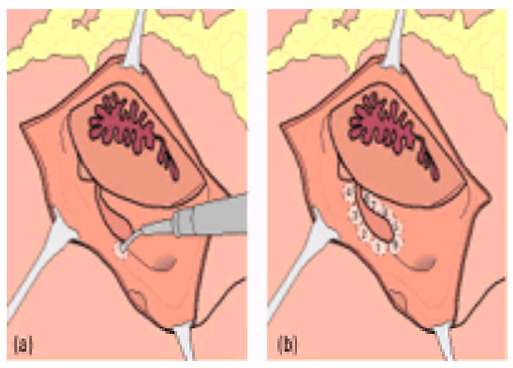


Fig. 9. (a) Discrete cryosurgical procedure for the treatment of atrioventricular node re-entry tachycardia. A 3-mm cryoprobe is used to place nine cryolesions around the periphery of the atrioventricular node, beginning at the upper edge of the os of the coronary sinus. (b) Cryolesions are placed in positions 1 to 4 as shown, then in positions 5 to 8, and finally in position 9. During the application of cryothermia at positions 7 and 8, the atrioventricular interval begins to become prolonged.

Cryolesions are then placed along the annulus of the tricuspid valve beginning just beneath the os of the coronary sinus (Fig. 9(b)). The first nine cryolesions are designed to outline the border of the triangle and to encircle the atrioventricular nodal tissue; subsequently, as many cryolesions as possible are placed within the triangle of Koch without causing block (Fig. 10). Used in this way, the cryoprobe acts as a ‘reversible knife'. In essence, the objective of this operation is to cryoablate as much of the perinodal tissue as possible without causing permanent atrioventricular conduction block.

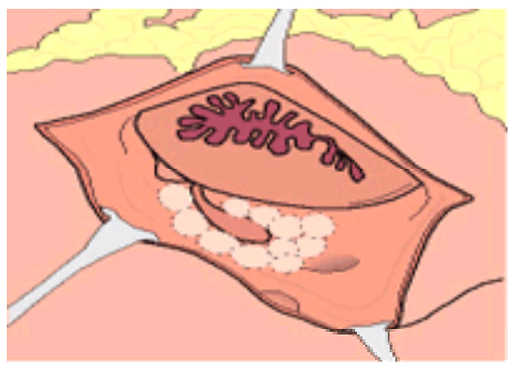


Fig. 10. After placement of the nine cryolesions around the border of the triangle of Koch, the area within the triangle is ‘filled in' with cryolesions. The tissue is cooled to -60°C for 2 min or until temporary heart block is produced; usually a further two or three lesions can be placed within the borders of the triangle.

Two other surgical techniques have been developed for the treatment of this arrhythmia. Ross and Johnson first used their surgical dissection technique for atrioventricular node re-entry in 1983; this approach has recently been modified slightly by Guiraudon.

Clinical results

The discrete cryosurgical procedure is demonstrably safe and efficacious. In all cases, postoperative electrophysiologic study has demonstrated the persistence of only a single atrioventricular conduction pathway, and the re-entrant tachycardia could not be induced. There have been no instances of permanent heart block, and no late recurrences.

As mentioned, Ross and Johnson, and Guiraudon, have described surgical dissection techniques for this arrhythmia, with good initial results. However, some cases of permanent heart block and a relatively high rate of late recurrence have been associated with these procedures.

As with atrioventricular reciprocating tachycardias, the lessons learned from surgical dissection and perinodal cryosurgery directly led the way to the current understanding of these arrhythmias and their ablation with radiofrequency catheters. The most commonly performed technique is to ablate the ‘slow' pathway, which almost always resides anatomically in the perinodal tissue inferoposterior to the compact atrioventricular node. Radiofrequency ablation in this region eliminates the re-entrant circuit while preserving conduction through the node itself.

Automatic atrial (ectopic) tachycardias

Automatic (ectopic) atrial tachycardias are often incessant. They usually originate from the body of the right or left atrium, but occasionally from the interatrial septum. The ventricular rate depends upon conduction through the atrioventricular node during the atrial tachycardia. Reversible abnormalities that may precipitate atrial tachycardias, such as hyperthyroidism, electrolyte imbalance, and digitalis toxicity, should be excluded before considering surgical treatment.

Preoperative electrophysiologic evaluation is necessary to discern the mechanism of the arrhythmia, establish the region of origin of the tachycardia, and exclude concomitant electrophysiologic derange-ments that may contribute to the rapid ventricular rate. Accurate preoperative localization is particularly important in patients with automatic atrial tachycardias if surgical ablation of the focus is contemplated: general anesthesia frequently suppresses the ectopic focus; intraoperative mapping without sophisticated, computerized, multipoint systems can be prohibitively difficult and time-consuming; and ectopic tachyarrhythmias are not inducible by standard techniques for programmed stimulation.

Surgical anatomy

Anatomically, ectopic or automatic atrial tachycardias can originate from foci occurring anywhere within the left or right atrial tissue or atrial septum. There may be multiple foci. Lowe collected data on 125 patients reported in the literature: of the 89 in whom the site was specified, 61 (68 per cent) originated in right atrial tissue, five (6 per cent) in the atrial septum, and 23 (26 per cent) in left atrial tissue (Fig. 11).

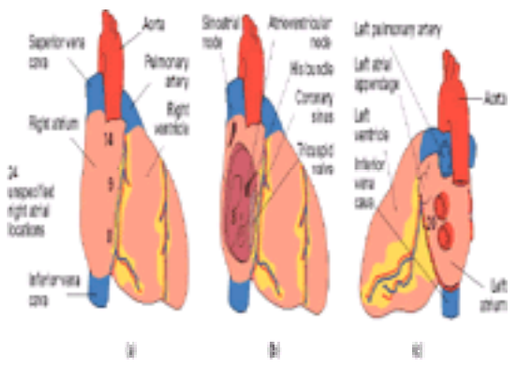


Fig. 11. Reported locations of (a) right and (b) left atrial ectopic foci: note that free-wall locations include those clustered around the pulmonary veins.

Surgical techniques

If the focus of the tachycardia can be localized, a variety of techniques has been advocated for surgical treatment. Cryoablation of the ectopic focus, with or without cardiopulmonary bypass, has been used. An alternative is wide excision of foci located on the right atrial free wall and repair with a pericardial patch. Foci on the atrial append-age may be treated by simple excision and oversewing of the line of resection. Others have used a combination of cryoablation and resection.

Foci on the left atrium tend to be near the left superior pulmonary vein; localized isolation procedures have been described for these tachyarrhythmias, but have had limited success. A significant number of these ectopic foci are localized to the left atrium on the preoperative study, but cannot be localized intraoperatively and a localized isolation procedure cannot be performed. In either instance, these ectopic foci should be excluded in the remainder of the heart by using the left atrial isolation procedure described by Williams *et al.* (Fig. 12). After that procedure the patients remain in normal sinus rhythm despite the presence of an incessant tachycardia confined to the left atrium. More recently, techniques for the ablation of left-sided tachycardias using radiofrequency catheterization have been successfully applied.

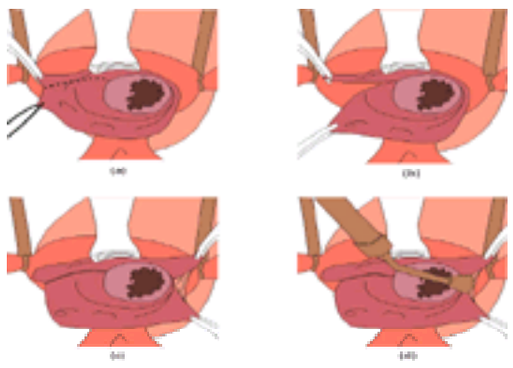


Fig. 12. Left atrial isolation procedure. (a) After incision for standard left atriotomy, the interatrial septum is retracted gently, and the atriotomy is extended anteriorly (dashed line) across Bachmann's bundle to the level of the mitral valve annulus just to the left of the right fibrous trigone. (b) Anterior extension of the standard left atriotomy has been completed; the base of the aorta and its juxtaposition with the anterior leaflet of the mitral valve are demonstrated. Note that anterior atriotomy extends across the valve's annulus. The main body of the left atrium has been separated anteriorly from remainder of heart. (c) Transmural left atriotomy is extended posteriorly to the level of the coronary sinus. The remaining part of the incision is made through endocardium and extends across the mitral valve annulus posteriorly just to the left of the interatrial septum. At this point, electrical activity continues to be propagated in 1:1 fashion between the right and left atria because of the interatrial muscular connections accompanying the coronary sinus. (d) The cryoprobe is positioned over the endocardial aspect of the posterior atriotomy, and its temperature is decreased to -60°C for 2 min. A similar cryolesion is created on the epicardial aspect of the atrioventricular groove on the opposite side of coronary sinus to ablate all remaining interatrial epicardial connections. The left atriotomy is closed with continuous 4–0 non-absorbable suture.

Right atrial tachycardias are usually confined to the atrial body, and may be multifocal. If the arrhythmia circuit or focus cannot be localized at operation, then a procedure that isolates the body of the right atrium while leaving the atrial pacemaker complex in continuity with the atrial septum and ventricles could be used. However, these right-sided lesions have proved extremely easy to ablate with radiofrequency catheters. Intracavitary ultrasonographic imaging has been of some benefit in several of these cases for correlating accurately the anatomic position with the electrophysiologic data recorded from multiple endocardial catheters.

Clinical results

When the focus of origin for ectopic atrial tachycardias could be adequately localized in the operation room, operative procedures for isolation and/or ablation of the arrhythmia have been uniformly successful. Both left and right atrial isolation procedures have had no adverse sequelae in long-term follow-up of a small number of cases.

Atrial flutter and fibrillation

The ablative treatment of atrial flutter and fibrillation is now the most exciting area in the clinical management of cardiac arrhythmias. Atrial fibrillation is the most common supraventricular arrhythmia and carries with it an increasing incidence of morbidity and mortality with age.

Atrial flutter: surgical anatomy, electrophysiology, and techniques

Experimental and clinical electrophysiologic mapping studies have shown that atrial flutter is due to ‘macro' re-entrant circuits in the atrial tissue around fixed anatomic obstacles (Fig. 13).

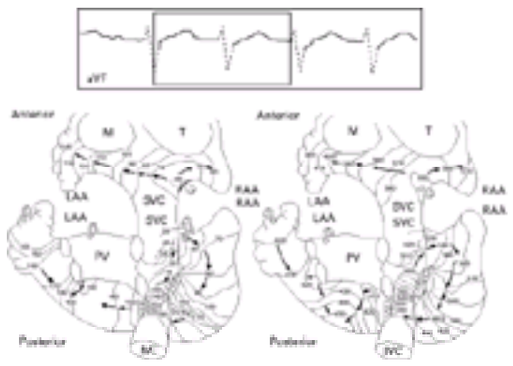


Fig. 13. Atrial flutter: two beats of flutter are shown (2:1 atrioventricular conduction). The time window from which the activation maps were constructed is shown in the upper panel, which is the standard electrocardiogram lead a VF. In both beats, the activation wave is seen to rotate around the site of the previous atriotomy on the posterior right atrium; it activates the rest of the atrium every 260 ms. The wave spreads from the posterior to the anterior right atrium, proceeding across the intra-atrial band (Bachmann bundle), and through the septum to activate the left atrium. The dotted line in the activation map of the second beat is the site of the surgical incision used to interrupt the re-entrant wave. (M, mitral valve; T, tricuspid valve; LAA, left atrial appendage; SVC, superior vena cava; RAA, right atrial appendage; PV, pulmonary veins; IVC, inferior vena cava.)

'Typical' atrial flutter occurs on the right side and the 'macro' re-entrant circuit traverses the isthmus between the eustachian valve of the superior vena cava and the os of the coronary sinus (Fig. 8). Cryolesions placed in this area of slow conduction effectively ablate the flutter circuit, which takes but a few minutes and can be safely added to any open-heart procedure. More commonly, ablation lesions can be applied to this area by radiofrequency catheter to interrupt the re-entrant circuit. Other 'atypical' circuits responsible for flutter can be more difficult to ablate in this manner.

Atrial fibrillation: surgical anatomy, electrophysiology and techniques

Chronic or intermittent atrial fibrillation results in three untoward sequelae for sufferers: loss of atrial transport function with loss of contraction; unpleasant subjective symptoms due to the irregular heart beat; and increased risk of thromboembolism. Experimental studies have confirmed that atrial fibrillation is due to intra-atrial re-entry in which multiple re-entrant wavelets are present. As atrial size increases, the number of wavefronts during atrial fibrillation and the duration of the fibrillation increase (Fig. 14). Intraoperative mapping in patients has shown that the only way to ablate atrial fibrillation surgically is to create a specific pathway of depolarization from the sinoatrial to the atrioventricular node; this pathway would have to activate all functional atrial myocardium in order for the atrium to maintain its normal transport function and to eliminate the problem of thromboembolism. The development of the Maze procedure meets these requirements (Fig. 15).

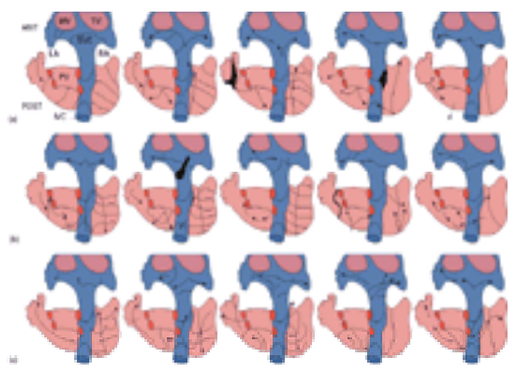


Fig. 14. Human atrial fibrillation. Activation-sequence maps of a single run of atrial fibrillation in a patient with Wolff–Parkinson–White syndrome are shown on a schematic drawing of the human atria. Each panel illustrates 1 s of data at different times during the fibrillation. The dark, thick lines represent arcs of functional block. Isochrones, the thin black lines, are shown every 20 msec. The dark, thick arrows show the direction of the spread of activation. In (a) and (b), a short, thick, black arrow shows the site of atrial insertion where an accessory pathway is conducting wavefronts from the ventricle. LA, left atrium; RA, right atrium; ANT, anterior; POST, posterior; LA, left atrial appendage; RA, right atrial appendage; SVC, superior vena cava; IVC, inferior vena cava; PV, pulmonary veins; MV, mitral valve annulus; TV, tricuspid valve annulus.

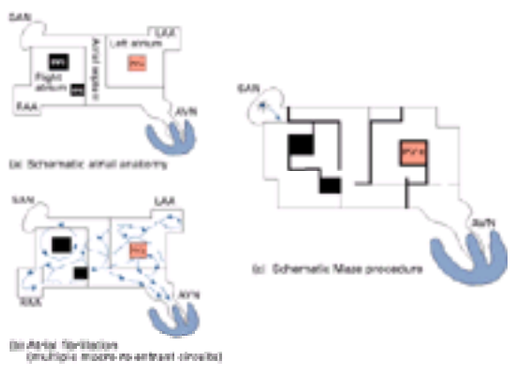


Fig. 15. Schematic depiction of the Maze procedure. (a) Schematic atrial anatomy: fixed anatomic obstacles on the right side include the SVC and IVC; and on the left side the PVs. (b) Schematic depiction of the multiple, simultaneously occurring re-entrant circuits responsible for atrial fibrillation. (c) The Maze procedure incorporates a series of incisions involving the fixed anatomic obstacles on the left and right side; the atrial appendages are removed. SVC, superior vena cava; IVC, inferior vena cava; SAN, sinoatrial node; AVN, atrioventricular node; PVs, pulmonary veins; RAA, right atrial appendage; LAA, left atrial appendage.

The five goals of this procedure are to: (i) eliminate atrial fibrillation as a clinical arrhythmia; (ii) restore normal sinus rhythm; (iii) keep atrioventricular conduction intact; (iv) restore atrial transport function on both the right and left sides; and (v) theoretically at least, to decrease the incidence of thromboembolic complications from atrial fibrillation by achievement of goals (i) to (iv).

The surgical technique of the Maze III procedure as most commonly performed by Cox *et al.* has been described in detail elsewhere. A series of right- and left-sided incisions as well as an atrial septal incision comprise the rather technically demanding procedure. Cryolesions are placed at the mitral and tricuspid annuli, and the coronary sinus atrial fibers are also cryoablated (Fig. 16).

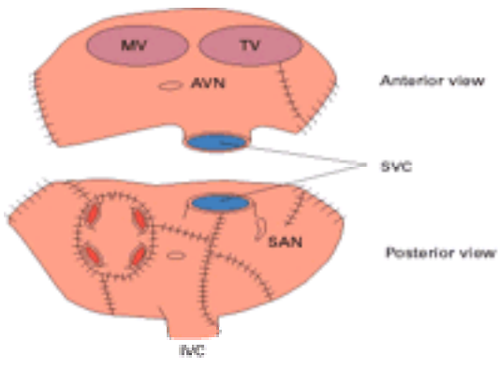


Fig. 16. The Maze III procedure on the human atrium depicted after unfolding. The upper panel shows the anterior left and right atria; the lower panel shows the left and right atria from behind. The surgical incisions are depicted as the hatched lines. The atrial appendages are removed. On the left side the pulmonary veins are encircled, and this encircling incision is connected to the appendage closure, to the incision in the atrial septum, and down to the mitral annulus. The coronary sinus is cryoablated at this location. The right-sided incisions include the one down to the tricuspid annulus as depicted in the upper panel. TV, tricuspid valve; MV, mitral valve; SAN, sinoatrial node; AVN, atrioventricular node; SVC, superior vena cava; IVC, inferior vena cava.

Clinical results with the Maze III procedure

The procedure has been performed in well over 200 patients in connection with Washington University, and in over a thousand patients worldwide, with excellent results. Virtually all patients (98 per cent) have been cured of their atrial fibrillation at late (up to 8 years) follow-up. Interestingly, the incidence of immediate postoperative atrial fibrillation is not different from that for other types of open-heart surgery. Within 2 months, however, these perioperative arrhythmias have completely disappeared. An extremely low incidence of late atrial flutter (9 per cent) has occurred, and each of these patients has been controlled with single-agent therapy. The majority have resumed sinus rhythm spontaneously; although approximately 40 per cent of patients with underlying disease of the sinus node have required atrial pacing for chronotropic support. Importantly, right- (100 per cent) and left-sided (81 per cent) atrial transport function has been demonstrated by a number of imaging techniques in late (more than 3 months) follow-up.

Following surgery, a normally generated impulse is propagated from the sinoatrial node to activate the entire atrial myocardium, apart from the excised atrial appendages and the pulmonary veins (Fig. 17). At the same time, it is impossible for a large ‘macro' re-entrant circuit to exist since the atrial impulses are precluded from turning back on themselves. Regardless of the number of ‘macro' re-entrant circuits responsible for the development and perpetuation of atrial fibrillation, they cannot occur after the Maze III procedure.

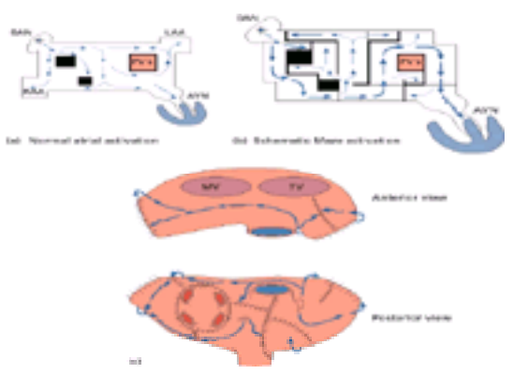


Fig. 17. Activation sequences and the Maze procedure. (a) Schematic diagram of normal atrial activation: during normal sinus rhythm the electrical impulse is generated within the SAN and propagates across the right and left atria and the atrial septum to the AVN; under normal circumstances there is a collision of two portions of the sinus impulse beneath the PV posteriorly. (b) Schematic diagram of the activation sequence following the Maze procedure: the impulse still originates in the SAN and propagates through the atrial septum and across the low right atrium to the left side where the tissue around the pulmonary veins is sequentially activated. (c) Depiction of activation following the Maze III procedure: the impulse from the SAN propagates anteriorly and inferiorly across the low right atrium, and simultaneously across the Bachmann bundle to the anterosuperior left atrium; a similar collision of activation occurs beneath the isolated pulmonary veins. Abbreviations as in Fig. 15, Fig. 16 and Fig. 17.

Surgical modifications to the Maze III procedure

A number of centers, particularly abroad, have modified the Maze III procedure in order to (i) shorten the time required for the relatively complex operation; and (ii) permit combination of the antiarrhythmic part of the procedure with other open-heart surgery, principally primary or repeat valve surgery.

As more and more surgical investigators gain experience with the Maze procedure and attempt to modify it for the above reasons, and as less invasive modifications are developed, a ‘standard' method for reporting the results of these interventions has been proposed. In this scheme, based on the intermediate postoperative atrial rhythm and the effectiveness of atrial contraction, a score of 0 equates to persistent atrial fibrillation, 1 to regular rhythm plus no atrial contraction, 2 to regular rhythm plus right atrial contraction, 3 to regular rhythm plus left and right atrial contraction, and 4 to normal sinus rhythm plus left and right atrial contraction. Ideally, this scheme should be applied to investigative, catheter-based interventions as well.

The surgical team at Osaka, Japan, have applied and studied most extensively their surgical modification, which replaces a number of the right- and left-sided incisions with cryolesions (Fig. 18). They have also published the largest series of reoperative procedures that include the Maze III modification; the majority of these patients were undergoing valve re-replacement. They achieved excellent results, with no evidence for additional incremental operative mortality. Atrial fibrillation was eliminated in approx. 80 per cent of patients, all of whom had chronic fibrillation preoperatively. While this cure rate for atrial fibrillation is somewhat less than for a primary procedure, nevertheless it is quite acceptable for this subpopulation of patients.

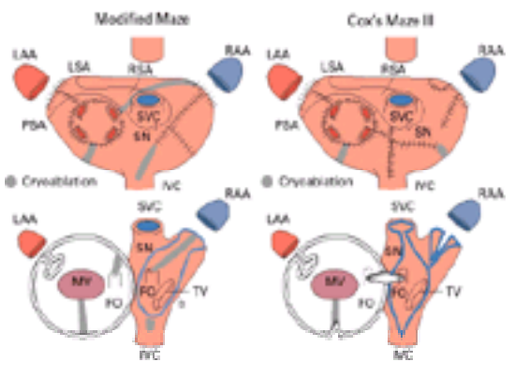


Fig. 18. Schematic drawing of the changes to the Maze III procedure made in Osaka, Japan, at the National Heart Center. As depicted, a number of the surgical incisions have been replaced by linear cryoablation lesions. Abbreviations as in Fig. 15, Fig. 16 and Fig. 17, plus LSA, left sinus-node artery; RSA, right sinus-node artery; PSA, posterior sinus-node artery.

In a larger series of patients from this institution undergoing concomitant procedures with the modification of the Maze III, three factors were recorded as risks for failure to restore normal sinus rhythm postoperatively: preoperative atrial size, duration of atrial fibrillation, and cardiothoracic ratio. Of these, atrial dimension was the most potent in predisposing to persistent fibrillation postoperatively. Interestingly, modifications to the procedure, and the pathogenesis and site of the underlying disease, were not significant predictors.

The left atrial isolation procedure described above has been applied to patients with atrial fibrillation undergoing mitral-valve replacement; 44 per cent maintained normal sinus rhythm in the long term. This operation, however, does not restore atrial transport function or decrease the incidence of thromboembolism.

Other investigators have attempted less invasive interventions in patients with atrial fibrillation. Right-sided modifications alone appear to be ineffective in controlling atrial fibrillation in the long term. Intraoperative radiofrequency ablation has been used to shorten the duration of the antiarrhythmic portion of a combined procedure, with about a 50 to 75 per cent reduction in operative time. Finally, using Heartport technology, Cox has replaced the surgical incisions with linear cryolesions in a small series of patients. It is these modifications to the basic concept of the Maze procedure that are potentially the future of antiarrhythmic surgery for atrial fibrillation.

Surgery for ischemic and non-ischemic ventricular tachycardia

Because of the electrophysiologic complexity of ventricular arrhythmias, and the difficulties with preoperative and intraoperative mapping, the development of surgical procedures for the relief of ventricular tachycardia has proceeded more slowly than for supraventricular arrhythmias. The efficacy of surgical treatments for these arrhythmias has been demonstrated, however, by analysis of operative results with 10-year follow-up, which show excellent cure rates but, in some cases, high operative mortality. Subsequent to this experience, two major advances in cardiovascular medicine occurred. The first was the development of thrombolytic therapy for acute myocardial infarction, and the second was the widespread availability of the implantable cardioverter–defibrillator. Thrombolytic therapy resulted in a decrease in the subsequent development of discrete left-ventricular aneurysms following infarcts due to early reperfusion; the discrete aneurysm and the monomorphic ventricular tachycardia were replaced with a chronic infarct and polymorphic, sustained and non-sustained, ventricular arrhythmias. As discussed below, the implantable cardioverter–defibrillator has had a substantial impact on the results and eventually the referral patterns for patients with ventricular arrhythmias.

Patients with both ischemic and non-ischemic ventricular tachycardias who are candidates for surgery should undergo complete electrophysiologic, angiographic, and ventriculographic evaluation. A catheter electrophysiologic study is performed to confirm that the arrhythmia is ventricular and not supraventricular in origin, to demonstrate, by induction and termination with programmed electrical stimulation, that the arrhythmia is re-entrant, and to identify the earliest site of origin of all morphologically distinct tachycardias using ‘catheter mapping' techniques.

Angiographic, ventriculographic, and echocardiographic evaluation is particularly helpful in the evaluation of three of the non-ischemic forms of ventricular tachycardia. In patients with diffuse cardiomyopathy due to patchy myocardial fibrosis, angiographic and hemodynamic data usually indicate some type of abnormal myocardial contractility associated with recurrent tachycardia. Ventriculography demonstrates diffuse dilation of both ventricles. This same finding on ventriculography can also be present in patients with idiopathic ventricular tachycardia due to repeated bouts of tachycardia, but in this the pathologic evidence of primary cardiac disease is absent. In arrhythmogenic right-ventricular dysplasia, caused by transmural infiltration of adipose tissue resulting in weakness and aneurysmal bulging of three areas of the right ventricle, ventriculography demonstrates diffuse dilation, depressed contractility, and delayed emptying of the right ventricle. Frank aneurysms of the infundibulum, apex, and/or posterior basilar region are seen, and hypertrophic muscular bands in the infundibulum and anterior right-ventricular wall result in a feathering appearance of the outflow tract. Since the origin of the tachycardia is the right ventricle, the 12-lead electrocardiogram shows a pattern consistent with left bundle-branch block during the tachycardia. Right ventriculography should therefore be performed in any patient with ventricular tachycardia and this pattern in the QRS complex.

Non-ischemic ventricular tachycardias

The final decision on surgical therapy for ventricular tachycardias of non-ischemic origin is based upon a variety of preoperative factors. The primary indication for surgery is refractoriness to all forms of medical therapy. Catheter ablation, for example, has proved to be extremely effective in the treatment of idiopathic, non-ischemic forms of tachycardia, especially in children.

Non-ischemic, right-sided tachycardias usually arise in the free wall or septum of the right ventricle and are, in general, extremely resistant to medical therapy. In the unlikely event of failed ablation by catheter, localized surgical isolation techniques are usually employed for tachycardias arising in the free wall, while multipoint, map-guided, cryoablative techniques are used for arrhythmias localized to the septum. In the past, more radical isolation techniques have been developed for arrhythmogenic right-ventricular dysplasia, designed to isolate the arrhythmogenic myocardium from the remainder of the heart. Postoperatively, however, the right ventricle has on occasion undergone a progressive dilatation, and cardiac transplantation in a suitable patient with this arrhythmia is currently the most feasible surgical approach.

Patients with ventricular tachycardia occurring in association with the long QT-interval syndrome frequently have *torsades de pointes* as the manifestation of their tachycardia. Medical therapy consists of b-adrenergic blockade; this has recently been coupled with permanent atrial or ventricular pacing. Surgical therapy has consisted of left cervicothoracic sympathectomy with removal of the left stellate ganglion and the first three to four left thoracic sympathetic ganglia. In general, this procedure is associated with early success but late failure. Currently, for symptomatic patients with a family history, genetic predisposition, or electrocardiographic evidence of this syndrome, implantation of a cardioverter–defibrillator, usually one capable of atrial pacing, is the recommended therapy (Fig. 19).

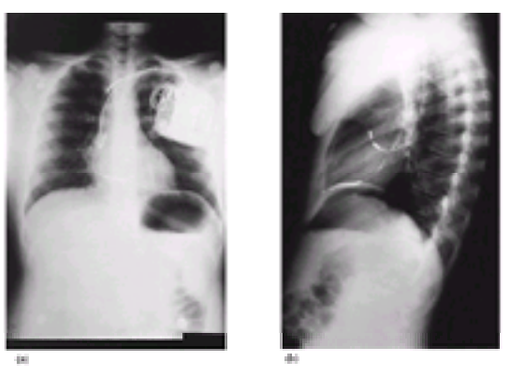


Fig. 19. Posteroanterior (a) and lateral (b) chest radiographs of an 11-year-old, 120-lb (54.4 kg) boy with long QT syndrome and syncope. A dual-chamber, implantable cardioverter–defibrillator is present in the left pectoral position. The atrial pacing lead is curled in the right atrium with the tip in the atrial appendage. The defibrillator lead tip is in the apex of the right ventricle, with the distal electrode in the right ventricular chamber and the proximal electrode in the superior vena cava and the bracheocephalic vein. The active can of the defibrillator generator makes up the third component of the defibrillation vector in this system.

Ischemic ventricular tachycardia

An algorithm for the optimal surgical treatment of refractory ischemic ventricular tachycardia is shown in Fig. 20.

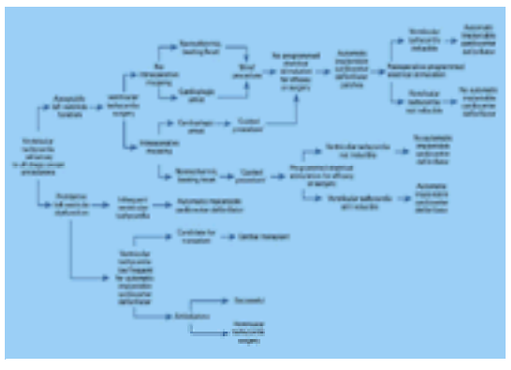


Fig. 20. Algorithm for the management of ischemic ventricular tachycardia that is refractory to all drugs except amiodarone (see text for further discussion).

The evaluation for surgical intervention should be based primarily on determining (i) that the patient has a sufficient degree of normal left-ventricular function to survive the operation, and (ii) the presence or absence of sustained (preferably monomorphic) tachycardia.

Because most patients with ischemic ventricular tachycardia have a left-ventricular aneurysm, accurate determination of the ejection fraction is difficult, and the absolute number is not an accurate predictor of operative mortality. Patients with discrete apical or posterior aneurysms and myocardium in the remainder of the heart that contracts in a normal or near-normal fashion are the best candidates from a functional point of view. The advantage of surgical intervention in appropriately selected patients is clear: a high probability of cure of the arrhythmia for an operative mortality risk that is not higher than that for surgical revascularization and left-ventricular aneurysmectomy alone (approximately 4 per cent), as compared to a lifetime of device and/or drug therapy.

The presence of polymorphic tachycardia makes intraoperative localization of the origin of the tachycardia difficult, even with complex computerized mapping systems, and decreases the likelihood of surgical cure.

Recent results from the MADIT and Antiarrhythmic versus Implantable Defibrillators (AVID) trial of implantable cardioverter–defibrillators as primary therapy for subclasses of patients with life-threatening ventricular arrhythmias and impaired left-ventricular function show that the implant should be the preferred initial therapy for patients with global ventricular dysfunction who are not direct surgical candidates.

Surgical techniques

A variety of intraoperative mapping techniques have been used in the past, and these are beyond the scope of this chapter. After completion of intraoperative mapping, with the patient on cardiopulmonary bypass, the ventricle is opened through the infarct or aneurysm. This is performed with the heart in the normothermic beating state, and preferably during ventricular tachycardia. All of the associated endocardial fibrosis is resected except that which extends on to the base of the papillary muscles (Fig. 21). About 10 per cent of patients will still have inducible tachycardia following resection of the endocardial fibrosis, indicating that their actual site of origin of the tachycardia is deeper in the myocardium than the visible border of the fibrosis. Endocardial cryolesions are applied with a 2.5-cm nitrous oxide cryoprobe to the site(s) of origin of the tachycardia(s) as determined by intraoperative mapping, thus ablating the myocardium beneath the visible fibrosis that is responsible for the tachycardia. Arrhythmogenic tissue at the base of the papillary muscles is cryoablated not resected. These techniques are applicable to both anterior and posteroinferior infarcts/aneurysms (Fig. 22).

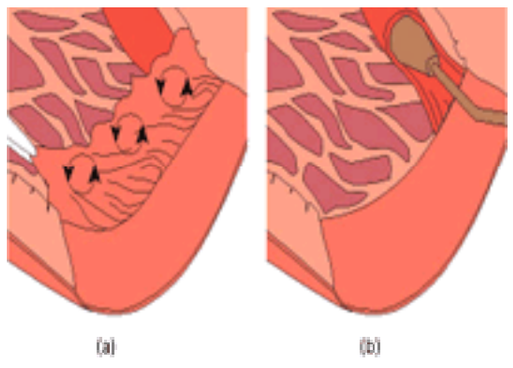


Fig. 21. (a) Diagrammatic sketch of the extended endocardial resection in an anterior left-ventricular aneurysm. The principle involved is the same as for the localized endocardial resection, but here all of the endocardial fibrosis associated with the aneurysm is resected except that involving papillary muscle. Circles: arrhythmogenic subendocardial tissue. (b) After resecting all the endocardial scar as a preliminary measure, endocardial cryolesions are placed at the sites or sites of ventricular tachycardia as determined by intraoperative mapping. In addition, any remaining scar on papillary muscle is cryoablated as shown.

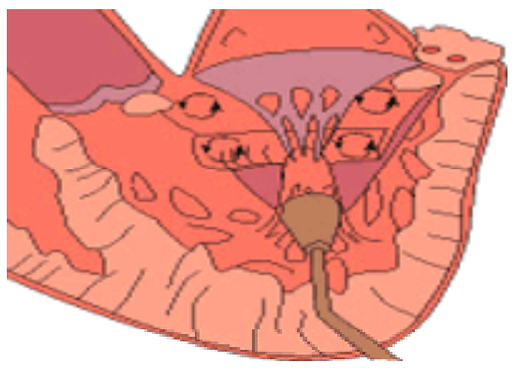


Fig. 22. Extended endocardial resection of fibrosis associated with posterior myocardial infarction or aneurysm and cryoablation of the lower two-thirds of the posterior papillary muscle. Endocardial fibrosis is resected to within 5 mm of the annuli of the aortic and mitral valves. Because the site of origin of ventricular tachycardia is frequently adjacent to the junction of these annuli, endocardial cryolesions (white circles) are applied at their bases to ablate any re-entrant circuits that might reside in remaining endocardial fibrosis immediately beneath them. In addition, endocardial cryolesions are applied to the site or sites of origin of ventricular tachycardia, as determined by intraoperative mapping, but only after removal of all endocardial scar. Circles: arrhythmogenic subendocardial tissue.

On completion of the resection and cryoablation, programmed electrical stimulation is applied in an attempt to reinduce the arrhythmia. If ventricular tachycardia is still inducible, mapping is again performed and the remaining arrhythmogenic myocardium is again cryoablated. If the arrhythmia is no longer inducible, there is a 98 per cent change that is has been permanently ablated. Coronary artery bypass grafting or other procedures that may be required are performed after completion of the antiarrhythmic part of the operation. Cardioplegic solution should probably not be administered until the antiarrhythmic part has been successfully completed: cardio-plegia may itself temporarily alter the delicate re-entry circuits causing the tachycardia. If the antitachycardia procedure is performed under cardioplegic arrest, it is impossible to determine intraoperatively whether or not it has ablated the arrhythmias.

The long-term success rate for patients who survive surgery for ventricular tachycardia is 87 per cent at 9 years in those operated upon for ischemic ventricular tachycardia. Long-term survival is excellent and parallels that for patients with ischemic heart disease and impaired ventricular function without arrhythmia (72 per cent at 9 years following successful operative intervention).

With the widespread availability of the implantable cardioverter–defibrillator, however, and the miniaturizing of these devices for pectoral implantation, the number of patients referred for antiarrhythmic surgery has declined to almost none. The defaulting to therapy with this extremely expensive device maintained over a lifetime promises to add several billion dollars a year to the cost of cardiovascular care worldwide.

Conclusions

For the most part, operations for supraventricular and ventricular arrhythmias have been replaced as front-line therapies by radiofrequency ablation and cardioverter–defibrillator implantation, respectively. They remain, however, as the building blocks for the development of these less invasive and highly effective therapies. Lessons learned from this 30-year experience are now being applied to the development of less invasive procedures for the treatment of atrial fibrillation. Finally, it is reasonable to expect that, in the unlikely event of a catheter failure for supraventricular tachycardia, or for the occasional patient with a discrete

left-ventricular aneurysm and monomorphic tachycardia, the opportunity for a surgical cure remains available.

Further reading

Allessie MA *et al.* Experimental evaluation of Moe's multiple wavelet hypothesis of atrial fibrillation. In: Zipes DP, Jalife J, ed. *Cardiac electrophysiology and arrhythmias*, pp. 265–75. Grune and Stratton, Orlando FA, 1985.

Antiarrhythmic versus Implantable Defibrillators (AVID) Investigators. A comparison of antiarrhythmic-drug therapy with implantable defibrillators in patients resuscitated from near-fatal ventricular arrhythmias. *New England Journal of Medicine* 1997; **337**: 1576–83.

Boineau JP *et al.* Mechanisms of atrial flutter as determined from spontaneous and experimental models. In: Josephson ME, Wellens HJJ, ed. *Tachycardias: mechanisms, diagnosis and treatment*, pp. 91–111. Lea and Febiger, Philadelphia, 1984.

Cain ME, Lindsay BD. The preoperative electrophysiologic study. In: Cox JL, ed. *Cardiac surgery: state of the art reviews*, pp. 1–39. Hanley and Belfus, Philadelphia, 1990.

Cox JL. Patient selection criteria and results of surgery for refractory ischemic ventricular tachycardia. *Circulation* 1989; **79** (Suppl. I): I 163–7.

Cox JL, Ferguson TB Jr. Cardiac arrhythmia surgery. In: Wells SA, ed. *Current problems in surgery*, 26th edn, pp. 193–278. Year Book Medical Publishers, Chicago, 1989.

Cox JL, Boineau JP, Schuessler RB, Jaquiss RDB, Lappas DG. Modification of the Maze procedure for atrial flutter and atrial fibrillation. I. Rationale and surgical results. *Journal of Thoracic and Cardiovascular Surgery* 1995; **110**: 473–84.

Cox JL, Boineau JP, Schuessler RB, Jaquiss RDB, Lappas DG. Modification of the Maze procedure for atrial flutter and atrial fibrillation. II. Surgical technique of the maze III procedure. *Journal of Thoracic and Cardiovascular Surgery* 1995; **110**: 485–95.

Cox JL, Ferguson TB Jr, Lindsay BD, Cain ME. Peri-nodal cryosurgery for AV nodal reentrant tachycardia in 23 patients. *Seminars in Thoracic and Cardiovascular Surgery* 1990; **99**: 440–50.

Cox JL, Gallagher JJ, Cain ME. Experience with 118 consecutive patients undergoing surgery for the Wolff–Parkinson–White syndrome. *Journal of Thoracic and Cardiovascular Surgery* 1985; **90**: 490–501.

Ferguson TB Jr. Anatomic and electrophysiologic principles in the surgical treatment of cardiac arrhythmias. In: Cox JL, ed. *Cardiac arrhythmia surgery*, pp. 41–73. Hanley and Belfus, Philadelphia, 1990.

Ferguson TB Jr. The role of the automatic implantable cardioverter-defibrillator (AICD) in the treatment of medically-refractory ventricular arrhythmias. *Trends in Cardiovascular Medicine* 1991; **1**: 131–9.

Ferguson TB Jr. The future of arrhythmia surgery. *Journal of Cardiac Electrophysiology* 1994; **5**: 621–34.

Ferguson TB Jr, Cox JL. Antiarrhythmic surgery: ventricular arrhythmias. In: Horowitz LN, ed. *Current management of arrhythmias*, pp. 382–91. Decker, Philadelphia, 1990.

Ferguson TB Jr, Cox JL. Complications related to the surgical treatment of supraventricular and ventricular cardiac arrhythmias. In: Waldhausen JA, Orringer MB, ed. *Complications in cardiothoracic surgery*, 1st edn, pp. 303–15. Mosby Year Book, St. Louis, 1990.

Ferguson TB Jr. Schuessler RB, Hand DE, Boineau JP, Cox JL. Lessons learned from computerized mapping of the atrium: surgery for atrial fibrillation and flutter. *Journal of Electrocardiology* 1994; **26**: 210–19.

Guiraudon GM *et al.* Surgery for the Wolff–Parkinson–White syndrome: the epicardial approach. *Seminars in Thoracic and Cardiovascular Surgery* 1989; **1**: 21–33.

Guiraudon GM, Klein GJ, Sharma AD, Yee R, Kaushik RJ, Fujimura O. Skeletonization of the atrioventricular node for AV node reentrant tachycardia: experience in 32 patients. *Annals of Thoracic Surgery* 1990; **49**: 565–73.

Harada A, D'Agostino HJ, Schuessler RB, Boineau JP, Cox JL. Right atrial isolation: a new surgical treatment for supraventricular tachycardia. *Seminars in Thoracic and Cardiovascular Surgery* 1988; **95**: 643–50.

Johnson DC, Nunn GR, Meldrum-Hanna W. Surgery for atrioventricular node reentry tachycardia: the surgical dissection technique. *Seminars in Thoracic and Cardiovascular Surgery* 1989; **1**: 53–7.

Kawaguchi AT *et al.* Factors affecting rhythm after the maze procedure for atrial fibrillation. *Circulation* 1996; **94** (Suppl. 9): II 139–42.

Kobayashi J *et al.* Rationale of the Cox Maze procedure for atrial fibrillation during redo mitral valve operations. *Seminars in Thoracic and Cardiovascular Surgery* 1996; **112**: 1216–21.

Lowe JE, Hendry PJ, Packer DL, Tang AS. Surgical management of chronic ectopic atrial tachycardia. *Seminars in Thoracic and Cardiovascular Surgery* 1989; **1**: 58–66.

Moss AJ *et al.* Improved survival with an implanted defibrillator in patients with coronary disease at high risk for ventricular arrhythmia. *New England Journal of Medicine* 1996; **335**: 1933–40.

Vigano M *et al.* Surgery for atrial fibrillation. *European Journal of Cardiothoracic Surgery* 1996; **10**: 490–7.

Williams JM *et al.* Left atrial isolation. New technique for the treatment of supraventricular arrhythmias. *Journal of Thoracic and Cardiovascular Surgery* 1980; **80**: 373–80.

40.12 Miscellaneous cardiac diseases

Andrew Parry

Cardiomyopathy
Hypertrophic
Dilated
Restrictive cardiomyopathy
Pulmonary embolism
Cardiac tumors
Myxoma
Fibroma
Rhabdomyoma
Lipoma
Malignant tumors
Other tumors
Pericardial disease
Acute pericarditis
Chronic pericarditis
Kawasaki disease
Further reading

Cardiomyopathy

Cardiomyopathies are diseases of cardiac muscle of unknown cause; the definition therefore excludes rare specific diseases of heart muscle. Myocarditis is sometimes included in this group but should not be as it is known to be due to an inflammatory process, although its nature remains unclear. There are three specific types of cardiomyopathy: hypertrophic, dilated, and restrictive.

Hypertrophic

Originally called hypertrophic obstructive cardiomyopathy, this cardiomyopathy is characterized by massive left ventricular hypertrophy classically causing asymmetric septal hypertrophy and gross hypertrophy of the papillary muscles (Fig. 1). Hypertrophic cardiomyopathy is generally subdivided according to the exact location of the obstruction, although this does not appear to have any prognostic significance. (A distinction should be made between the obstructive form of hypertrophic cardiomyopathy and a form of non-obstructive hypertrophic cardiomyopathy that produces apical septal hypertrophy; this has a good prognosis.) Most commonly the hypertrophy is near the base of the heart and causes obliteration of the mid-(left ventricular)cavity or subaortic obstruction. Further, systolic anterior motion of the anterior mitral leaflet caused by it being drawn into the left ventricular outflow tract by the turbulent flow (Venturi effect) exacerbates the outflow obstruction; indeed this may be the most significant element of obstruction. This systolic anterior motion may also give rise to mitral regurgitation. Hemodynamically, the asymmetry of the hypertrophy causes obstruction to part or all of the left ventricular cavity, thereby producing systolic compromise, but the hypertrophied, stiff, non-compliant walls also compromise diastolic function.

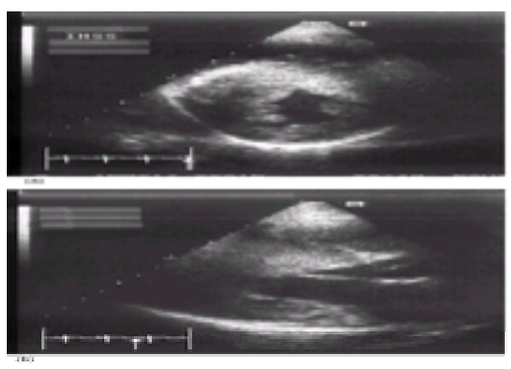


Fig. 1. Typical echocardiogram of a patient with asymmetric septal hypertrophy; the massive septal hypertrophy is evident on both the short-axis (a) and long-axis (b) views.

The cause is unknown. In some there is a strong family history and an autosomal-dominant inheritance may be seen. Further, in some individuals, point mutations in the genes coding for cardiac myosin molecules have been identified. Other research has identified links with an abnormally high density of catecholamine receptors in the outflow tract and catecholamine overproduction as a whole, though others claim instead that there is specific downregulation of the b_1-receptors independent of negative feedback from excessive circulating catecholamines. There are also links with hyperthyroidism and hyperinsulinemia, and the disease may be a reflection of a genetically determined disorder of catecholamine handling by the fetal myocardium. Forms of hypertrophic cardiomyopathy may also be part of more generalized disease processes and occur in type III glycogen-storage disease and with cytochrome oxidase deficiency, although arguably these should not be classified as cardiomyopathies.

Sudden death is common amongst these patients, with a 3.5 per cent annual mortality. Risk factors are young age at the time of diagnosis (5.9 per cent annual mortality if less than 30 years of age), positive family history, and suddenly worsening symptoms. There are no hemodynamic or morphologic predictors of worse outcome as death in this condition may be from a number of causes. It is usually due to dysrhythmias: atrial fibrillation is the most common, but ventricular tachycardia and re-entrant supraventricular tachycardia also occur. Various antidysrhythmic drugs, such as b-blockers, calcium antagonists, disopyramide, and amiodarone, may decrease the incidence of sudden death by suppressing the malignant rhythms, but some studies suggest that the prognosis for the disease is independent of medical intervention. Others have shown that there is a benefit from these medications independent of their antidysrhythmic effect, probably due to their negative inotropic side-effect resulting in improvement in left ventricular compliance and diastolic function.

Myocardial infarction is the second most common cause of death and coronary flow reserve is significantly reduced in patients with hypertrophic cardiomyopathy irrespective of the presence of classical atherosclerotic coronary arterial disease. In patients with hypertrophic cardiomyopathy the coronary arteries are commonly dysplastic, with thickening of the intima and media as well as fibroelastosis of the media, resulting in vessel narrowing. However, myocardial ischemia may also occur in the presence of normal coronary arteries for two reasons. First, the massive hypertrophy may itself lead to subendocardial ischemia or the associated increase in wall tension may prevent adequate diastolic coronary perfusion. Secondly, the hypertrophied muscle may form muscle bridges that can physically obstruct the coronary arteries. When atherosclerotic coronary arterial disease coexists with hypertrophic cardiomyopathy, the mortality is higher and cardiac dysfunction more common. The natural history of this cardiomyopathy is the development of a dilated cardiomyopathy with ventricular dysfunction; death is due to poor cardiac function.

Treatment traditionally has been primarily medical, with surgery reserved for unresponsive cases and those with more severe disease, but this is based on older reports showing a 27 per cent surgical mortality. Nowadays operative mortality is 0 to 3 per cent and, as the long-term mortality for the surgical group is 1.8 per cent per year compared to 3.5 per cent per year for the medical group, the indications for surgery may be expanded. Studies have shown that those patients more incapacitated on exercise testing, those with asymmetric hypertrophy, those with severe systolic anterior motion of the mitral valve, and those with a prolonged isovolumetric relaxation phase are most likely to benefit from surgical intervention. Those in whom surgery achieves the greatest reduction in left ventricular outflow-tract obstruction and reduction in left ventricular filling pressures early postoperatively have a better long-term prognosis.

Conservative medical treatment for those without ventricular dysrhythmias includes b-blockers and verapamil. Improvement in clinical criteria and hemodynamics have been reported with both these agents (see above), and each may be efficacious where the other has failed. Dysrhythmias are best treated with amiodarone. When congestive cardiac failure supervenes in the advanced stage the outlook is very poor and normal antifailure treatment is indicated. More recently, the concept of dual-chamber pacing has been developed as an alternative to surgery. It has long been recognized that ventricular pacing can reduce the gradient associated with subaortic stenosis, and this has now also been applied to hypertrophic cardiomyopathy. The benefit is due to a number of reasons: alteration in the activation sequence

of the left ventricle, reduction in the contact time between the mitral valve and the proximal septum, and ventricular remodeling. Positron-emission tomography has shown that pacing is associated with an improvement in septal perfusion. Clinically, dual-chamber pacing has been shown to reduce the left ventricular outflow gradient both at rest and during exercise, and later to induce regression of myocardial hypertrophy. This improvement may continue even if pacing is discontinued after several months. Overall results are inconsistent and in some, symptoms actually worsen during the period of pacing. Certainly there is no evidence as yet that continuous ventricular pacing offers any survival benefit over conventional medical therapy for these patients.

Surgery has been reserved for the 10 to 15 per cent who do not benefit from medical treatment. The aim has traditionally been to excise the obstructing muscle mass, thereby opening up the left ventricular outflow tract. Practically, by the method of Morrow, this has been accomplished by opening the aortic root and excising a mass of septal muscle from immediately below the nadir of the right coronary cusp of the aortic valve as far anticlockwise as possible without damaging the mitral valve (this should prevent damage to the conduction system) (Fig. 2). The excision is taken as inferiorly as possible to ensure that all obstruction is removed and this should be confirmed afterwards by direct palpation. The excision should be made in a single, deliberate movement so as not to drop particles of muscle into the cavity of the ventricle and thereby risk embolization. Care must be taken to avoid damage to the aortic valve and the creation of an iatrogenic ventricular septal defect during the myomectomy. An alternative approach is to enter the left ventricle at the apex below the lowest diagonal branch, taking great care not to damage the coronary arteries; though exposure may be good the incidence of late complications (for example aneurysm formation, left ventricular failure) is high and a left ventriculotomy should only be used when essential. Occasionally these two approaches may be combined to obtain wide exposure. When the obstruction is severe and subaortic, a modified Konno procedure has been used with success to relieve the gradient. In some cases this reportedly produces complete regression of the myocardial hypertrophy.

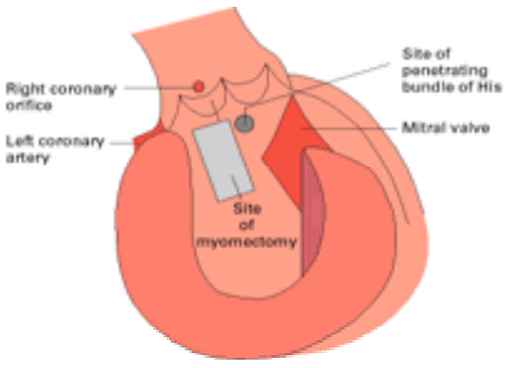


Fig. 2. The site of safe resection of ventricular septal muscle to remove the subaortic obstruction.

Although the gradient in the left ventricular outflow tract is usually well relieved by myomectomy, significant obstruction may remain due to systolic anterior motion of the anterior leaflet of the mitral valve, which has led some to advocate repair or replacement of that valve at the same time as the myomectomy. Usually the repair involves plicating the anterior leaflet of the valve to reduce its mobility, or reducing the height of the posterior leaflet, although elongating the anterior leaflet along its horizontal axis with a patch of pericardium has also been used successfully.

Myomectomy with or without surgery on the mitral valve gives excellent results, with operative fatalities of 0 to 3.2 per cent, and improvement in symptoms and exercise capacity in 86 to 91 per cent. No controlled studies have sought to determine the benefit incurred by concomitant surgery on the mitral valve, but most studies recommend that, unless there is intrinsic disease of the mitral valve or severe mitral regurgitation, simple myomectomy alone is a better choice. When atherosclerotic coronary arterial disease coexists with hypertrophic cardiomyopathy, bypass grafting combined with one of the other procedures is safe and gives greatly improved relief of symptoms and prognosis than coronary arterial grafting alone. However, in the majority of patients with normal coronary arteries but who have perfusion defects shown by thallium scanning prior to surgery, surgical relief of the obstruction in the left ventricular outflow tract results in normalization or improvement of myocardial perfusion.

Dilated

Dilated cardiomyopathy is a ‘final common pathway' for many diseases of cardiac muscle. It usually presents with congestive cardiac failure and grossly impaired systolic function, though earlier only an increase in the cardiothoracic ratio may be seen. Systemic embolization is also common, occurring at the rate of four episodes per 100 patient years. This risk can be reduced by oral anticoagulation. On electrocardiography there are non-specific changes with some delay of left ventricular conduction.

Essentially it is a diagnosis of exclusion, though many etiologic factors have been implicated [chronic severe hypertension, persistent alcohol abuse, pregnancy, and autoimmune disease (triggered by viral infection, especially coxsackie, or previous rheumatic fever)]. The heart is overweight and dilated, with ventricular hypertrophy appropriate to the degree of increased wall tension. By definition there is no coronary arterial disease but there can be extensive fibrosis of the ventricular wall and occasionally cellular infiltration suggestive of myocarditis.

The prognosis is poor: 77 per cent die within 2 years, but survivors may improve over this time. Risk factors for poor outcome include older age (97 per cent 6-year mortality if over 55 years of age compared to 66 per cent if under 55 years), larger cardiothoracic ratio (86 per cent 6-year mortality if over 0.55 compared to 40 per cent if less), and lower cardiac index (89 per cent 6-year mortality if less than 3 l/min per m^2 compared to 35 per cent if greater). Systemic embolization is common, particularly in patients in atrial fibrillation (33 per cent against 14 per cent without fibrillation).

The mainstay of treatment remains conventional medical therapy with diuretics, afterload and preload reducers, and more recently b-blockade. Those with a short history and evidence of nonspecific inflammation on biopsy on occasion benefit from immunosuppression with steroids and azothiaprine, but these are a minority. In those with severely dilated, poorly contracting hearts, especially those in atrial fibrillation, oral anticoagulation is recommended in view of the risk of embolism. Surgically the mainstay of treatment and the ‘gold standard' by which other therapies are judged is transplantation. Although this now has a 80 per cent 1-year and 65 per cent 5-year survival, with near normalization of quality of life in 50 per cent of recipients, a shortage of human donor organs has necessitated investigation of other approaches. Xenotransplantation, using organs from different species made to appear compatible to the recipient by genetic manipulation of the donor organ, is close to clinical application. However, although the major issue of hyperacute rejection can be resolved, acute and chronic rejection still occur as for heterograft transplantation. Nontransplant approaches are being developed: these include the development of chronically implantable left ventricular assist devices (**LVAD**s), ‘bridging to transplantation' (tem-porarily implanting an LVAD to allow patient survival until transplantation can be performed) and ‘bridging to recovery', ventricular remodelling/ventriculoplasty, and the training of systemic muscle to augment the failing myocardium such as in dynamic cardiomyoplasty and aortomyoplasty.

LVADs are expensive and prone to the complications of infection, air embolism, bleeding (early and late), thromboembolism, and mechanical failure. Their use in the acute situation, such as for temporary postcardiotomy ventricular failure or as temporary support prior to transplantation (occasionally for a number of months), is well established but a number of trials on the viability of chronic implantation have produced very disappointing results.

The concept of ‘bridging to recovery' has developed from observation of patients treated with the intention of ‘bridging to transplantation'. In these patients a significant improvement in left ventricular ejection fraction and cardiac index, and a significant fall in left ventricular end-diastolic dimensions and pulmonary capillary-wedge pressure, can be observed after a prolonged period of left ventricular unloading. This is associated histologically with a reduction in myocytolysis and contraction-band necrosis, although there appears to be an increase in myocardial fibrosis. These findings have led to the contention that after a prolonged period of left ventricular unloading it may be possible to remove the LVAD without the need for cardiac transplantation, and indeed this has been done in a very small number of cases. The results following LVAD removal are anecdotal and mixed; in a significant number of patients there is rapid reappearance of the dilating cardiomyopathy. However, this approach may prove to be beneficial in specific circumstances, particularly when it can reasonably be anticipated that the insult is self-limiting, such as with post partum cardiomyopathy or dilating cardiomyopathy associated with myocarditis. Here, a brief period of LVAD support during the period of myocardial insult might rest the ventricle and allow adequate function to return when the insult has ended.

In 1995 a bold new approach to dilating cardiomyopathy was reported by Batista, who claimed in an uncontrolled series to have ‘successfully' treated ‘many cases' of endstage heart failure by resecting varying amounts of left ventricular myocardium. The rationale is that as a cavity distends so wall tension increases (Law of Laplace); by reducing intracavity pressure the wall tension should also fall, with an associated improvement in cardiac function. The operation essentially involves resecting ‘nonfunctioning' segments of left ventricular myocardium, which may be any portion of the ventricular cavity. If it involves the papillary muscle(s), replacement of the mitral valve is also required. More scientific studies have supported enthusiasm for the procedure. The operative mortality was 6 per cent, with an overall 1-year survival of 86 per cent; overall, 53 per cent of patients improved from New York Heart Association (**NYHA**) class IV to class I or II. However, although there was a decrease in left ventricular dimensions as would be expected, there was little improvement in cardiac index and the reason for the clinical improvement therefore

remains unknown. The reports available to date also concern a heterogeneous group of patients, many of whom underwent valve replacement or repair and coronary revascularization at the same time as the ventricular reduction. The independent impact of the ventricular reduction surgery therefore cannot yet be ascertained until larger trials excluding these confounding factors have been conducted. However, the theory, computer modeling, and clinical results to date are encouraging.

Intuitively it would appear logical to attempt to recruit skeletal muscle to augment the failing myocardium but the associated problems have only recently been overcome. The major problem is how to convert skeletal muscle, which produces short bursts of power, into a fatigue-resistant muscle capable of continuous exercise. This transformation can be achieved by a prolonged period of low-frequency electrical stimulation, which alters the phenotypic expression of the myofibrils with the result that they become almost purely type 1 fibers with fatigue resistance. Practically, because skeletal fibers are not a syncytium, a burst stimulator is required to recruit many motor units and this is triggered off the R-wave of the electrocardiogram. The fatigue-resistant muscle can then be used in a number of ways: dynamic cardiomyoplasty, aortomyoplasty, and skeletal muscle ventricles.

Dynamic cardiomyoplasty is the most established technique, first being performed in 1985. After training, the latissimus dorsi muscle is wrapped around the ventricles and thereby augments cardiac function. The exact mechanism is unclear: the muscle might provide some contractile effort but the lack of improvement in hemodynamic variables argues against this. It might passively prevent ventricular distension by modulating cardiac remodeling, or when used in ischemic cardiomyopathy the wrap might improve myocardial perfusion. Finally, by producing iatrogenic hypertrophy, the myocardial wall tension might be reduced (Law of Laplace). Whichever mechanism is active, controlled studies comparing isolated dynamic cardiomyoplasty to conventional medial therapy have demonstrated that the operation produces significant improvements in NYHA class, quality of life, systolic function, and length of subsequent admissions to hospital. In early series there was an increased incidence of sudden cardiac death due to the burst stimulators triggering ventricular dysrhythmias and nowadays the procedure is usually combined with the implantation of an automatic defibrillator. However, as the operative mortality is high (13 per cent), the procedure should only be recommended for severely incapacitated individuals in NYHA III, not those with endstage heart failure, for whom cardiac transplantation is more appropriate.

Aortic counterpulsation is an established technique for temporary left ventricular support, reducing afterload and improving myocardial perfusion by raising diastolic pressure. Using the same muscle-training concept as above, attempts have been made to perform chronic aortic counterpulsation by wrapping the latissimus dorsi around either the ascending or descending aorta. Primarily a research tool to date, this technique has produced a significant increase in cardiac output, a decrease in systemic vascular resistance, and an improvement in subendocardial viability index. It would therefore appear attractive as a palliation of endstage cardiac failure. However, clinical experience is limited to 14 cases, with a 14 per cent operative and 36 per cent follow-up mortality (75 per cent of these were due to sudden cardiac death). Enthusiasm therefore should be cautious, although these figures are very early on the ‘learning curve'; later experience is awaited with interest.

Ventricles totally comprising skeletal muscle are at present purely investigational. Early experience was marred by their spontaneous rupture, but producing a pericardial container within the ventricle appears to have resolved this issue. Improvements in cardiac function are similar to those for aortoplasty but clinical trials have not been performed as yet.

Restrictive cardiomyopathy

This is the rarest of the cardiomyopathies and applies to three conditions: endomyocardial fibrosis, Löeffler endocarditis, and idiopathic restrictive cardiomyopathy (the most common cause in the Western world). However, some specific myocardial diseases can develop restrictive pathophysiological changes (such as amyloidosis and sarcoidosis). Endomyocardial fibrosis is the most common form of restrictive cardiomyopathy worldwide, occurring in an idiopathic form that is believed by some to follow a viral myocarditis (the mumpsvirus has recently been implicated), while in other circumstances it may accompany other ill-defined cardiac conditions, particularly noncompaction syndrome. It may be seen with a number of cardiac diseases that cause subendocardial ischemia, such as severe congenital aortic stenosis, but these are not true cardiomyopathies. Further, endomyocardial fibrosis may be associated with systemic metabolic diseases such as tissue carnitine deficiency, glutaric aciduria, and mucopolysaccharidoses I and VI. Endomyocardial fibrosis affects the endocardium and therefore may involve the cardiac valves. It can also extend into the myocardium and thereby interfere with intracardiac conduction, producing heartblock or ventricular dysrhythmias. It mainly affects young/middle-aged men, and is most common in Central Africa, less common in South America and Asia, and rare elsewhere. It is a disease of poor socio-economic circumstances and is associated with protein-poor, carbohydrate-rich diets. It is particularly associated with ingestion of cerium, a toxic element found in the soil of some regions. Animal research suggests that failure of repair due to the protein deficiency leads to fibrosis following cell necrosis. In other circumstances the cause is unknown, though it may occur with systemic diseases such as schistosomiasis (where it mainly affects the right side of the heart) and Beçhet disease (which appears to be secondary to vasculitis). Pathologically the active phase shows inflammation, edema, necrosis and cellular infiltration, later replaced by organized fibrous tissue and finally calcification. It is characterized by restriction of ventricular filling with obliteration of the ventricular inflow by the fibrotic process and associated dysfunction of the atrioventricular valve. The outflow tract of the ventricles is spared.

Endomyocardial fibrosis is most readily diagnosed by echocardiography and may involve either or both ventricles. There is thickening of the endocardium with obliteration of the apical inflow portion of the affected ventricle, calcification of the affected portion, and dilatation and hypercontractility of the outflow portion. Functionally there is severe restriction of ventricular diastolic filling, causing an elevation in preload and a fall in cardiac output. There are often associated outpouchings of the ventricular wall and filling defects due to the presence of thrombus.

Medical treatment is directed towards controlling the associated congestive cardiac failure; the 4-year survival is 77 per cent. Factors predictive of poorer outcome are a family history of the disease and worse cardiac function at the time of presentation. Aggressive immunosuppressive therapy has also been advocated, with limited success. Surgical treatment may be palliative or corrective. In those with recurrent pericardial effusions, pericardectomy, pericardial windows, and pericardioperitoneal shunts have been used with some success. If the fibrosis only involves the right ventricle, a Glenn shunt may be used to reduce the volume of blood returning to the right ventricle and augment pulmonary flow. Corrective procedures have also been devised, and, since the first report by Dubost in 1976 of stripping the fibroelastosis from the endocardium of the ventricle(s) and repairing/replacing the atrioventricular valve(s), this has become the procedure of choice. Results of the available series using a combination of the above techniques show an early operative mortality of 19 per cent (higher for those with biventricular disease). Longer-term follow-up shows a continued improvement by clinical, radiologic, and hemodynamic criteria, though in a small proportion the fibrosis returns.

Apart from this specific type of endocardial restriction, true restrictive cardiomyopathy with severe impairment of diastolic function is not amenable to conventional surgery and transplantation is the only option. This is true of the other restrictive ‘cardiomyopathies' (such as sarcoidosis and amyloidosis) but in these diseases there is the added risk of their recurrence within the newly transplanted heart.

Pulmonary embolism

Pulmonary embolism is a major risk after any form of prolonged immobility, including (post)surgical. It is estimated that 15 to 20 per cent of deaths in acute district hospitals are due to pulmonary embolism, and that 20 000 patients die each year in British hospitals while a further 40 000 have nonfatal events. The mechanism of clot development and propagation is well known and will not be reiterated here. It should be remembered, however, that venous thrombosis starts at the time the immobility starts, that is at the time of premedication, and continues until the patient is fully ambulatory again.

Clinically there is a spectrum of disease. Acute minor emboli cause tachypnea, pleuritic chest pain, and occasionally mild hemoptysis. Examination is often unremarkable, and, as the lungs have a very active fibrinolytic system, most minor emboli are lysed spontaneously. Acute major embolism causes severe dyspnea, dull central chest pain, circulatory collapse with elevation of central venous pressure, and hypoxia, hypocapnia and acidosis. Death is usually rapid and 60 per cent die within the first 2 h. However, of the survivors many suffer further pulmonary emboli over the next 72 h, and without treatment these subsequent emboli are frequently fatal. Chronic pulmonary embolism is a rare form of the disease in which recurrent minor emboli cause gradual occlusion of the pulmonary vasculature with pulmonary hypertension. Finally, subacute pulmonary embolism is an acute embolic episode occurring on the background of chronic embolic disease.

Treatment of pulmonary embolism should be twofold: adequate treatment of the existing pulmonary occlusion and prevention of further embolism. For those with minor acute events, analgesia may be all that is required. There has been much debate about when anticoagulation is appropriate and no consensus has arisen. It should be considered for those who are at greater risk of either having a further embolus or in whom a further embolus would be poorly tolerated due to pre-existing cardiorespiratory disease, those who have no risk factors preventing safe, full anticoagulation, and those with a more severe primary embolic event. The optimal duration of anticoagulation is also undetermined, but provided all risk factors have resolved it is generally agreed that 3 to 6 months of treatment is adequate.

Following massive pulmonary embolism, acute or subacute, the risk to life is great and persists for days. Aggressive treatment is therefore required and is initially medical. Traditionally, for emboli not associated with cardiovascluar compromise, the mainstay of treatment has been heparin infusion, but, although this reduces the risk of death from further emboli, it does not significantly alter the risk from the primary embolus as heparin has no thrombolytic activity. Therefore, when the insult is severe it is more appropriate to administer thrombolytics, such as streptokinase, urokinase, and tissue plasminogen activator, that can dissolve both the existing embolus and the primary thrombus, thereby reducing the risk of recurrence when given systemically. Though thrombolysis is associated with a significant rate of bleeding complications and should not therefore be used in patients with significant risk of hemorrhage, it produces a reduction in mortality compared to heparin. In an attempt to reduce the risk of hemorrhagic complications, a more recent approach has been to use low-dose (25 000 units loading dose and 10 000 units per hour infusion) streptokinase, infused directly into the embolus or into the origin of the occluded vessel, together with systemic heparin (1000 units/h infusion initially, adjusted to achieve an activated partial thromboplastin time ratio of 1.5 to 2); this is as efficacious as high-dose systemic streptokinase in producing clot lysis as judged by

clinical, pulmonary angiographic, and blood gas analysis, with significantly fewer bleeding problems. However, despite the reduction in bleeding complications this approach is still not applicable in the immediate postoperative period.

The advent of thrombolytic therapy has significantly reduced the number of patients requiring pulmonary embolectomy so that nowadays surgery is reserved for those with contraindications to, or who have failed on, thrombolytic therapy. However, for patients in cardiogenic shock after pulmonary embolization, the in-hospital survival after surgery is 72 to 77 per cent with a 1 to 2 per cent risk of re-embolization as compared to 65 to 70 per cent survival and up to 20 per cent re-embolization for medical treatment, so the pendulum may have swung too far in favour of thrombolysis. In addition, 25 to 28 per cent of patients suffer a major hemorrhage secondary to the use of thrombolytic agents. For these reasons, surgical pulmonary embolectomy should be considered for all patients who have suffered massive pulmonary embolus with cardiovascular compromise.

Two techniques for surgical pulmonary embolectomy are practised. Which is chosen depends on surgical preference and the availability of cardiopulmonary bypass. Outflow occlusion without cardiopulmonary bypass was first described by Trendelenburg in 1908 and first used successfully in 1924; the mortality may be as low as 20 per cent if cardiac arrest has not occurred prior to surgery, but rises to 75 per cent when there has been cardiac arrest. When the equipment is available, the more common approach nowadays is open pulmonary embolectomy on cardiopulmonary bypass. In this procedure the main and branch pulmonary arteries can be opened extensively and all the embolized clot carefully removed; this can easily be done with the heart beating. This approach has a mortality of 29 to 46 per cent. Older age, worse preoperative hemodynamics, more proximal location of emboli, longer duration of symptoms, and recurrent embolization are all associated with a poorer prognosis. Aggressive disobliteration of as many segmental pulmonary arteries as possible reduces the surgical mortality significantly by reducing the immediate postoperative pulmonary vascular resistance. Some have used retrograde pulmonary venous flushing to disimpact clot from the small arterioles with good result. The use of percutaneous bypass to support the circulation during the time of transport and preparation in patients with cardiovascular collapse may also reduce surgical mortality.

The prevention of further embolization is as important as the removal of any primary embolus and re-embolization occurs in 4 per cent of patients after successful embolectomy. Virtually all emboli originate in the deep leg or pelvic veins, and techniques to prevent their passage up the vena cava have been devised, either with a filter, which may be inserted percutaneously, or by interruption, plication or clipping of the vein. Indications for operating on the inferior vena cava are unclear; some advocate intervention after all pulmonary emboli and others only after recurrent embolization. Others have inserted filters prophylactically into patients deemed at high risk for pulmonary embolism, with minimal morbidity and mortality for the procedure. After the insertion of filters for therapeutic reasons the right ventricle may fail, as maintenance of an adequate output is critically dependent on a high filling pressure and this is prevented by the procedure. The overall mortality for the open procedures is 6 to 27 per cent and for percutaneous insertion, 3 to 5 per cent. Long term all these techniques are associated with complications. The long-term patency of the vena cava after clipping is only 88 per cent, with 20 per cent of patients suffering significant leg edema, but the incidence of recurrent pulmonary embolism is only 1.9 per cent. This is in contrast to an incidence of 0.5 to 7 per cent recurrent embolization after the insertion of a filter, though the incidence of significant leg edema and long-term patency are 5 to 7 per cent and 96 to 100 per cent, respectively.

Recently, less invasive techniques have been devised to break up the embolus; these involve fragmentation of the embolus using rotating catheters similar to the rotavator used for the treatment of atherosclerotic arterial disease. The studies have been successful, with a significant fall in pulmonary arterial pressure. However, although the main pulmonary arteries are cleared, the branches often become obstructed with debris and a significant amount of damage is done to the periarterial tissue. However, owing to the less invasive nature of this approach it will become a well-established treatment of acute pulmonary embolus.

An increasingly recognized cause of pulmonary hypertension is chronic pulmonary embolic disease. Prevention of further embolism is obviously necessary but, although anticoagulation may prevent further thromboembolism, it will not treat the primary disease and thrombolysis has nothing to offer as the clot is well organized. Treatment options are surgical: thromboendarterectomy or lung transplantation. The choice depends on the location and nature of the obstruction, how compromised the right heart is, and how successful it is anticipated the relief of obstruction will be. If obstruction is distal it is unlikely that thromboendarterectomy will be successful. In addition, following complete occlusion of the pulmonary artery there are irreversible atrophic changes in the distal vessels and this will limit the efficacy of the procedure. Finally, if the right ventricle is severely compromised preoperatively, any period of cardiopulmonary bypass may cause it to fail postoperatively.

Thromboendarterectomy is done on full cardiopulmonary bypass. The main pulmonary arteries and their accessible branches are opened widely and the obstructing organized thrombus is cored out. The more aggressive the operator the more successful the procedure in lowering the postoperative pulmonary arterial pressure. In centers where large numbers of this procedure are performed the operative mortality is as low as 8.7 per cent and postoperatively there is a significant improvement in clinical symptoms, fall in pulmonary vascular resistance (by up to 72 per cent), improvement in cardiac index (by up to 50 per cent), and im-provement in right ventricular function. All these improvements are evident immediately after surgery and improvement continues into the later follow-up period. No operative factors are associated with increased risk of the procedure but it has been found that a preoperative pulmonary arterial pressure of greater than 50 mmHg or pulmonary vascular resistance greater than 1100 dyn/s/cm^{-5} are highly predictive of perioperative mortality. At present there are no published long-term follow-up studies.

The alternative to thromboendarterectomy is either single or double lung transplantation. With either of these approaches the 1-year survival in relation to chronic pulmonary embolic disease is 70 per cent and the 5-year survival 60 per cent.

Cardiac tumors

The most common cardiac tumors are secondaries, which have an incidence of 1 to 2 per cent in autopsy studies. In men the most common primary sites/types are lung, esophagus, and lymphoma; in women lung, lymphoma, and breast are most common. Primary cardiac tumors are rare in all age groups, with an incidence of 1.7 to 28 per 100 000. They are mainly benign (80–95 per cent), with men having twice as many malignant cardiac tumors than women, though women have more cardiac tumors overall than men (3:2). However, even with benign tumors, early detection is essential, as successful resection is possible only if cardiac function is maintained.

Cardiac tumors, apart from myxomas, present in similar fashion. When small they usually lie intramurally and may cause any cardiac dysrhythmia from interference with the conduction apparatus; it is estimated that between 2.5 and 5 per 100 000 sudden deaths are due to small intramyocardial neoplasms. When larger they project into the cardiac cavities and cause obstruction of the valves, the cavities, and the outflow tracts. Presentation is thus often nonspecific and a high index of suspicion must be maintained as unexplained murmurs, congestive cardiac failure, and dysrhythmias may be the first signs.

Electrocardiography is usually normal or nonspecifically deranged. Chest radiography is likewise usually normal, though chamber enlargement and tumor calcification may be seen. Echocardiography is the best screening investigation, allowing assessment of the size, shape, location, attachment, and mobility of the tumor, as well as the presence or displacement of the coronary arteries by the tumor mass. However, in a number of cases echocardiography has missed large tumors in various chambers, possibly due to the relatively homogeneous state of certain tumors and to the lack of normal tissue interfaces. More recently, magnetic resonance imaging (**MRI**) has been shown to be an excellent imaging modality with the ability to discriminate between benign and malignant tumors by analysis of tumor size, quantity of the heart involved, pericardial/extracardiac extension, and tumor necrosis. MRI should therefore be considered the method of choice for the assessment of cardiac tumors.

In the fetus, cardiac tumors carry a different significance. On prenatal echocardiography the prevalence of cardiac tumors is 0.17 to 0.19 per cent, with 50 per cent of them being multiple. They are usually not hemodynamically significant, and 90 per cent are rhabdomyomas, though only 50 per cent of these are associated with tuberosclerosis. Of these tumors, 50 per cent spontaneously partially or completely regress during gestation, 30 per cent remain stable and do not require intervention, and 20 per cent require surgery in the neonatal period. Overall mortality in the neonatal period for those born with cardiac tumors is 30 to 35 per cent.

Treatment is surgical resection as far as possible, but when complete excision cannot be achieved, relief of the obstruction alone can give good long-term results. As many of the tumors are slowly progressive, ideally a cuff of normal tissue should also be removed at the primary operation. When the tumor is ventricular, enucleation or conservative resection should be undertaken; atrial tumors can be excised radically and the defect in the atrial wall repaired with pericardium. Occasionally, when the tumor involves extensively the right side of the heart, a Glenn shunt or complete Fontan may be performed; significant data on these approaches are lacking at present due to the rarity of the cardiac lesions. Overall, surgical mortality for tumor resection is around 5 per cent, with late deaths occurring in a further 3 to 7.5 per cent. Of these late deaths, 70 per cent are from recurrence and progression of malignant cardiac tumors. Ninety-five per cent of survivors are in NYHA class I or II.

Myxoma

This is the most common tumor of the heart, comprising 50 to 90 per cent of benign cardiac tumors. It is more common in females and rare before adulthood, though it has been described in a newborn. Myxoma is usually solitary but biatrial myxomas with a common attachment at the fossa ovalis have been described, as have complex and familial forms, the sufferers of which develop tumors at an earlier age and at atypical and multiple locations; recurrences are common. In the complex myxoma syndrome, cutaneous lentigo, myxomas at various sites, and endocrine tumors accompany the cardiac tumor.

Cardiac myxomas usually originate in the left atrium (75 per cent), with 20 per cent in the right atrium and 4 per cent in the ventricles. Other sites have been reported (both atrioventricular valves, inferior vena cava, and both semilunar valves) but the stalk usually originates from the inferior rim of the fossa ovalis; 84 per cent occupy

the left side of the heart, 13 per cent the right, and3 per cent are bilateral. Macroscopically they are soft, semitranslucent tumors; histologically they have large amounts of amorphous matrix with syncytial groups of elongated cells closely related to thin-walled capillaries. In the past the origin of these tumors has been questioned, some claiming that they are areas of organized thrombus not true neoplasms, and hence better called Lambl excrescences; this does not now appear appropriate.

Myxomas may present with cardiac failure due to valve obstruction or regurgitation and impingement into the cardiac chambers; this may be further manifest as pulmonary hypertension. They may also present with embolization, either systemic or pulmonary due to thrombotic embolization or fragmentation of the tumor. In addition, constitutional symptoms are common (fever, malaise, and weight loss), probably due to an autoimmune phenomenon. Physical examination is often unrewarding, though the murmur of mitral stenosis and the tumor ‘plop' are the classic findings. Diagnosis is best achieved by echocardiography; catheterization is avoided due to the risk of tumor fragmentation. When the diagnosis has been made, surgery should be undertaken urgently in view of the risk of embolization.

Since the first report of successful resection by Crafoord in 1954, the surgical mortality has fallen to 5 per cent. Care must be taken in handling the tumor to prevent embolization and all four cardiac chambers must be examined because of the risk of multiple tumors. Excision should be wide around the base of the stalk as there can be local recurrences and the tumor may become locally invasive. If this precaution is taken, recurrence is rare with single tumors but is more likely in the complex syndrome, giving an overall recurrence rate of 6.3 per cent. Occasionally they may metastasize to brain and bone. The overall 20-year survival is 89 to 91 per cent.

Fibroma

A tumor of childhood, with 85 per cent arising in infants and children, cardiac fibromas usually occur in the free wall of the left ventricle or the interventricular septum; they comprise 5 per cent of primary cardiac tumors. They are usually single and are round, firm, well-encapsulated masses with a false capsule that allows enucleation. They are entirely benign and show no malignant transformation. On cutting, the surface bulges and the fibrous tissue intermingled with muscle has a whorled appearance. Occasionally they calcify, and they often interfere with the conduction system so that sudden death is common, occurring in 30 per cent.

Rhabdomyoma

This is primarily a tumor of childhood, with 50 per cent dying within the first month of life and 80 per cent by 1 year. In 50 per cent of patients, cardiac rhabdomyoma is associated with tuberosclerosis. Rhabdomyomas are often multiple in the interventricular septum and adjacent free wall, the diffuse myocardial involvement leading to heart failure and death. The tumors are usually well-circumscribed, nonencapsulated, whitish nodules and may grow to a large size, projecting into the ventricular cavities. The tumor cells are large, vacuolated, with delicate radiating strands of cytoplasm; they are filled with glycogen. Whether these rhabdomyomas are true neoplasms or simply overgrowths of the Purkinje cells and thus hamartomas remains uncertain, but spontaneous regression has been described. Certainly they are congenital, as intrauterine diagnosis has been confirmed at neonatal operation.

Lipoma

This very uncommon tumor usually occurs in adults. Most commonly it is located subendocardially and appears as a yellow mass projecting into the cardiac chamber. Rarely it is intramural and unencapsulated. It is slow growing and may calcify following fat necrosis. Microscopically it is composed of fatty tissue with interlacing muscle fibers.

Malignant tumors

Primary malignant tumors are extremely rare and are usually angio- or fibrosarcomas. Presentation may be with hemorrhagic tamponade due to erosion of the tumor through the ventricular free wall with subsequent necrosis of the tumor bulk. Conventional excision has a poor outcome as resection is often incomplete and tumor dissemination occurs early; the 1-year survival is less than 25 per cent. Adjuvant chemo- or radiotherapy are therefore recommended. Complete cardiectomy and cardiac transplantation appears to carry a better prognosis despite the need for immunosuppressive therapy. If the cardiectomy margins are clear of tumor, prolonged survival without recurrence may be achieved; if the tumor extends to the resection margins, death usually occurs within 15 months. The exception to this is the angiosarcoma, in which embolization occurs very early and, even if the resection margins are clear, distant metastases have always developed. The presence of this tumor is therefore a contraindication to cardiac transplantation.

Other tumors

Papillary fibroelastomas comprise 7.9 per cent of benign cardiac tumors and are the most common tumors of the heart valves. They are small masses of hyalinized, hypocellular stroma with a single endothelial cell layer covering. They are usually attached to the valves, though they may occur on any endocardial surface. Even when small they should be excised, owing to the high risk of embolization. Hemangiomas and benign teratomas occur infrequently.

Pericardial disease

Pericardial disease may present in a number of ways: acute pericarditis, acute pericardial effusion, chronic restrictive pericarditis, and chronic effusive pericarditis.

Acute pericarditis

This may either be inflammatory or purulent. Acute inflammatory pericarditis occurs following myocardial infarction (Dressler syndrome), viral infections (frequently after upper respiratory-tract infections), and with autoimmune diseases (especially systemic lupus erythematosus and scleroderma). Bacterial infections (especially tuberculosis, streptococcal, and staphylococcal) are most common in nutritionally, immunologically or traumatically compromised patients. In adults, pyopericardium may be an isolated finding, but is more usually associated with pneumonia, empyema, or penetrating trauma to the pericardial cavity. In children, however, pyopericardium is most commonly spread from another source of infection such as osteomyelitis or septic arthritis. Acute pericarditis may also occur with drugs (procainamide, isoniazid, and hydralazine), malignancy, uremia, after cardiac surgery, and after radiation therapy. Presentation is with chest pain that is worse on inspiration and relieved by sitting forward, and a pericardial friction rub. Electrocardiography shows diffuse elevation of ST segments without reciprocal changes in the other chest leads, and depression of the PR segment.

Any form of acute pericarditis may progress to acute effusive pericarditis with pericardial effusion and cardiac tamponade. Symptoms at this time are of dyspnea, fatigue and orthopnea, with patients frequently observing a fall in urine output (especially after cardiac surgery) and peripheral edema. Clinical examination classically reveals pulsus paradoxus, which is exaggeration of the normal fall in systolic blood pressure that occurs during inspiration to more than 10 mmHg. The patient's neck veins are distended (due to elevated atrial filling pressures), there is narrowing of the pulse pressure (due to a decrease in left ventricular stroke volume), and, late, the patient may be in frank cardiogenic shock. On chest radiography there is a large cardiac silhouette and on electrocardiography the complexes are of low voltage.

Treatment of acute pericarditis should be directed at the cause (for example, stop the inciting drug or dialyze), and intervention should be aggressive to avoid long-term complications such as constrictive pericarditis. For nonpurulent pericarditis, treatment is with nonsteroidal anti-inflammatory drugs and occasionally steroids to reduce the inflammation. With purulent pericarditis, if early pericardiocentesis is performed with lavage of the pericardial cavity, later complications may be avoided. If presentation is late or treatment inadequate there is a high incidence of constrictive pericarditis. Even in recent series the mortality associated with pyogenic pericarditis is up to 77 per cent. The clinical features of tamponade on admission correlate closely with the subsequent development of constrictive pericarditis, even when tamponade resolves with pericardiocentesis.

The treatment of cardiac tamponade is an emergency, with specific management of the underlying condition taking secondary place. Percutaneous pericardiocentesis should be performed under echocardiographic control or an emergent subxiphisternal pericardial window created.

Chronic pericarditis

Any acute pericarditis may develop into chronic pericarditis. Classically, effusive and constrictive pericarditis have been attributed to mycobacterial infection, but nowadays other factors are more common (other infection, metabolic, post-traumatic, postsurgical, infiltration, radiation, autoimmune, and idiopathic). The diagnosis can be confirmed by echocardiography with Doppler interrogation of the venous inflow signals and flow across the atrioventricular valves. More recently, cine-computed tomography has been shown to provide anatomic as well as physiologic data, as has cine-MRI. With both these modalities the cause may be ascertained and the differential diagnosis of restrictive cardiomyopathy can be excluded. The findings on cardiac catheterization are classical, with an elevated diastolic plateau between the pressures in all chambers and prominent *x* and *y* descents in the right atrial pressure tracing. Pericarditis primarily affects diastolic function and causes impaired ventricular filling. The cardiac output is therefore highly rate dependent. For this reason, atrial pacing with rates up to 130 beats/min has been used as a temporizing measure or when other interventions are inappropriate, as this leads to a rise in cardiac output without changing the stroke volume.

Uremic pericarditis is becoming rarer since long-term hemodialysis has become more established. The incidence in patients on hemodialysis is 20 per cent, with 61 per cent of those appearing within 6 months of starting dialysis. In 50 per cent of these, medical management with aggressive hemodialysis, pericardiocentesis, and steroids is successful. Operation should be reserved for those with acute tamponade, unresponsive or recurrent effusions, or in whom constriction occurs. For all cases, except when the effusion is posterior or constriction has occurred, an anterior pericardectomy or the formation of a large pericardial window give excellent results with minimal morbidity and mortality.

Constrictive pericarditis may occur following cardiac surgery. The incidence is between 0.15 and 0.6 per cent, and it usually arises between 2 and 40 months after surgery. The mechanism of injury is unknown but both visceral and parietal layers are involved in the inflammation and infiltration. It may be caused by air-drying of the serosal layer or chemical exposure, and outbreaks have been seen following phases of using povidone–iodine solution to wash out the pericardium at the end of surgery. Alternatively, cold injury or the persistence of organized clot may be to blame. It is in all respects typical constrictive pericarditis with the usual findings on cardiac catheterization, and though medical therapy with diuretics, angiotensin-converting enzyme inhibitors, steroids, and nonsteroidal anti-inflammatory drugs may be attempted, further surgery is often required.

Pericardial infiltration by neoplasms occurs most commonly with breast and lung carcinoma, lymphoma, and leukemia. Except for lung carcinoma it is not usually a terminal event and prolonged survival may be expected. Treatment of effusive pericarditis should initially be by pericardiocentesis plus chemotherapy (both systemic and by instillation directly into the pericardium) and radiotherapy, with pericardectomy being reserved for those with good survival prospects, when medical treatment fails, or for constriction. Pericarditis may also occur after radiation, even up to 4 years after treatment. It usually arises when more than 4500 rad have been used on the mediastinum and is most common after treatment for Hodgkin lymphoma. The prognosis is primarily affected by the amount of fibrosis extending beyond the pericardium, and one should always be aware that the radiation damage may be associated with other cardiac damage such as coronary ostial stenosis, myocardial fibrosis, and valvular fibrosis causing insufficiency.

Treatment for chronic pericardial disease is initially medical. When medications fail or there is evidence of constriction, surgery is indicated. When effusions are the primary problem, pericardiocentesis under echocardiographic control is appropriate. Alternatively, especially when a diagnosis is required, an open pericardiotomy with insertion of a drainage tube may be performed as a biopsy may also be taken. If the effusions recur or drainage persists a pericardioperitoneal window may be formed between the pericardial cavity and the right side of the diaphragm immediately over the right lobe of the liver; this allows long-term drainage and has been advocated for patients with malignant pericardial effusions, as these are usually recurrent. Advocates claim good drainage without the development of carcinomatosis, though this must be considered to be a likely outcome. Others fashion a pleuropericardial window to allow free drainage of pericardial fluid into the pleural cavity; this may be done through a left anterior thoracotomy, excising as large a patch of pericardial sac anterior to the phrenic nerve as possible. More recently a similar thoracoscopic approach is gaining wide support, as these patients are usually very unstable and the insult of a full thoracotomy is avoided.

When constriction has developed, excision of the pericardium is required. Numerous surgical approaches to pericardectomy have been described (median sternotomy, anterolateral thoracotomy, bilateral thoracotomies) but none is ideal. Through a median sternotomy, exposure is somewhat limited and a subtotal pericardectomy (excising the pericardium anterior to both phrenic nerves) may be achieved. A thoracotomy usually gives better exposure. Results consequently show that, via a median sternotomy, excellent relief of right-sided restriction may be achieved but left-sided restriction often persists, while with a thoracotomy a reduction in filling pressure for both sides of the heart may be achieved. However, adequate decortication also requires adequacy of depth, and, in some cases, failure of surgery is due to failure to remove a constrictive epicardial peel. In some of these cases there is even a definite plane between the two layers. Surgery to remove the epicardial peel has increased risk of disrupting the coronary vessels and it may prove more prudent to cross-hatch the epicardium, allowing the ventricle to bulge through the gaps. Cardiopulmonary bypass is generally only required if other cardiac lesions are being corrected at the same operation but should be available on standby. Specimens should always be sent for histology, cytology, and microbiology (both routine and for tuberculosis) as in up to 50 per cent of patients the cause of the pericarditis may not be evident at the time of pericardectomy.

Following pericardectomy there may be an immediate or a delayed improvement. Incomplete pericardectomy due to retention of an epicardial peel may occasionally be blamed for a postoperative low-output state, but this may also be caused by a progressive deterioration in cardiac function due to pre-existing myocardial atrophy. Severe dilatation of the ventricles may be seen following surgery, with associated deterioration in the function of the atrioventricular valve. In such cases, improvement following surgery may continue for months. The overall surgical mortality is between 0 and 8 per cent and death is usually due to low cardiac output or bleeding. In survivors, 65 to 100 per cent improve from NYHA class III or IV to NYHA class I or II, and actuarial survival is 81 to 94 per cent at 3 years and 63 to 87 per cent at 7 years. NYHA class IV, a low-voltage electrocardiogram, preoperative ascites, and dyspnea at rest are predictors of poor outcome.

Kawasaki disease

Kawasaki disease, first described by Tomisaku Kawasaki in 1967, is an acute, multisystem vasculitic disorder affecting children and infants. It has a predilection for Japanese and other Asian populations, though it has been reported in all ‘racial' groups, and particularly in young children, with 50 per cent of cases being diagnosed by the age of 2 years and 80 per cent by the age of 4 years. The male:female ratio is 1.6:1.0. It is of unknown cause but an infectious agent is suspected from the occurrence of epidemics in different locations. Direct spread, however, has not been proved. Clinically and pathologically there is a spectrum of severity from incomplete forms to aggressive forms that resemble infantile polyarteritis nodosa. In all cases there are prominent vasculitic features.

The diagnostic criteria are detailed in Table 1. Incomplete forms are, however, common, especially in children under 6 months, and may be associated with as severe complications as in those who fulfil all diagnostic criteria for the disease. A high index of suspicion is therefore essential in children with a fever of unknown origin, and as the disease is progressive, frequently repeated examinations are required.

Principal features (five required)
1. Abrupt onset of fever unresponsive to antibiotics
2. Bilateral conjunctival injection
3. (a) Dry, chapped, fissured, or reddened lips
(b) Prominence of tongue papillae
(c) Diffuse oropharyngeal inflammation
4. (a) Reddening of palms or soles
(b) Induration of hands or feet
(c) Desquamation of fingertips or toes
5. Polymorphous truncal rash without vesicles or crust
6. Cervical lymphadenopathy (usually unilateral)
Associated features
1. Urethritis
2. Arthritis/arthralgia
3. Carditis (endocarditis/myocarditis/pericarditis)
4. Aseptic meningitis
5. Diarrhea
6. Gallbladder hydrops
7. Obstructive jaundice
Laboratory findings
1. Elevated acute-phase reactants
2. Thrombocytosis
3. Sterile pyuria
4. Negative antistreptolysin O

Table 1 Diagnostic criteria for Kawasaki disease

All layers of the heart are affected. Pericarditis is common in the acute phase of the disease and may be associated with effusions. It is usually transient and settles without long-term complications. Myocarditis and myocardial fibrosis also occur in the acute phase; clinically they are detectable in 30 per cent of patients, but by routine biopsy it is found that 100 per cent are involved. There are mild, nonspecific electrocardiographic changes but no rise in cardiac enzymes, suggesting that the dysfunction is due more to infiltration than myolysis. Diagnosis is therefore difficult, but clinical evidence of congestive cardiac failure and echocardiographic demonstration of impaired left ventricular function are probably the most reliable. The endocardium is also involved, and valvular dysfunction occurs in both acute and subacute phases of the disease. Most commonly the mitral valve is affected (10 per cent of cases), either due to leaflet inflammation or to dilatation of the annulus owing to myocarditis or myocardial ischemia. Aortic regurgitation is less common (5 per cent) and the tricuspid is rarely involved. Echocardiographic follow-up of these individuals has shown that all valvular abnormalities may resolve spontaneously.

The most dangerous of all complications is the development of arterial disease, initially aneurysmal and later occlusive (Fig. 3); this occurs in the coronary and peripheral circulations, and starts most commonly in the subacute stage of the disease in those who suffered a more severe acute phase. It is also more common in younger patients and in those treated with steroids. The administration of g-globulin has reduced the incidence of coronary abnormalities in some patients but attempts to focus treatment by identifying the at-risk subgroup have been unsuccessful. Echocardiographic evidence suggests that the occurrence of pericardial effusions or mitral regurgitation in the acute phase are reproducible predictors of later coronary disease, and aortic regurgitation has likewise been shown to be associated with coronary lesions in all cases. The aneurysms may be fusiform or saccular and have a high incidence of thrombosis leading to myocardial infarction in the subacute or convalescent phases. Two-dimensional echocardiography demonstrates 92 per cent of aneurysms shown on angiography and is the method of choice for screening. It is therefore recommended that the investigation be repeated every 4 weeks in patients shown to have aneurysms until they have all disappeared, usually within 6 months to 2 years of the onset of Kawasaki disease. Angiography can therefore be restricted to those with a history of ischemia, persistent mitral regurgitation,

coronary calcification on plain chest radiography, or persistent aneurysms on echocardiography.

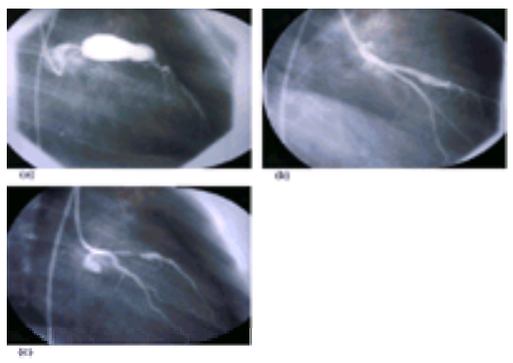


Fig. 3. A coronary arteriogram of a young patient with Kawasaki disease: the large aneurysms of the coronary artery are evident (a); this is an early stage of the disease process; the later coronary arterial stenosis has not yet occurred. (b) This later coronary angiogram demonstrates the virtually complete recovery of the arterial system, although localised stenoses have occurred at the sites of the previous aneurysms (c).

Treatment of Kawasaki disease involves frequent, thorough examination of the child both physically and by echocardiography to detect the development of cardiac complications, as well as drug therapy. Those with aneurysms of the coronary arteries are at greatest risk of developing occlusive disease and their echocardiographic assessment should continue until all the aneurysms have disappeared. Aspirin and dipyridamole are claimed to decrease chronically the risk of sudden death from thrombosis of these aneurysms, while thrombolytics are indicated in the acute phase. It has been shown, however, that stenoses of the coronary arteries develop at the sites of previous aneurysms, and long-term treatment with aspirin and regular exercise testing are therefore advocated. Steroids have no benefit in this disease and indeed have been found to increase the risk of aneurysm formation. More recently it has been shown that the administration of g-globulin reduces the incidence of abnormalities in the coronary arteries, and early identification of the at-risk group as previously described may therefore allow a reduction in this severe complication. Though not applicable to the acute phase of the disease, conventional coronary arterial surgery may be required in those with persistent aneurysms or those who develop stenoses at the location of previous aneurysms in the convalescent or chronic stage of the disease.

Further reading

Cardiomyopathy

Cherian G *et al.* Endomyocardial fibrosis: report on the hemodynamic data in 29 patients and review of the results of surgery. *American Heart Journal* 1983; **105**: 659–65.

Fighali S, Krajcer Z, Leachman RD. Septal myomectomy and mitral valve replacement for idiopathic hypertrophic subaortic stenosis: short and long term follow-up. *Journal of the American College of Cardiology* 1984; **3**: 1127–34.

Fuster V, Gersh BJ, Giuliani ER, Tajik AJ, Brandenburg RO, Frye RL. The natural history of idiopathic dilated cardiomyopathy. *American Journal of Cardiology* 1981; **47**: 525–31.

Gill CC, Duda AM, Kitazume H, Kramer JR, Loop FD. Idiopathic hypertrophic subaortic stenosis and coronary atherosclerosis. Results of coronary artery bypass alone and myomectomy combined with coronary artery bypass. *Journal of Thoracic and Cardiovascular Surgery* 1982; **84**: 856–60.

Ino T, Benson LN, Freedom RM, Rowe RD. Natural history and prognostic risk factors in endocardial fibroelastosis. *American Journal of Cardiology* 1988; **62**: 431–4.

Magovern GJ, Sr, Simpson KA. Clinical cardiomyoplasty: review of the ten-year United States experience. *Annals of Thoracic Surgery* 1996; **61**: 413–19.

Spirito P, Maron BJ, Rosing DR. Morphologic determinants of hemodynamic state after ventricular septal myotomy-myomectomy in patients with obstructive hypertrophic cardiomyopathy: M mode and two-dimensional echocardiographic assessment. *Circulation* 1984; **70**: 984–95.

Pulmonary embolism

AbuRahma AF *et al.* Therapeutic and prophylactic vena caval interruption for pulmonary embolism: caval and venous insertion site patency. *Annals of Vascular Surgery* 1993; **7**: 561–8.

Doerge HC, Schoendube FA, Loeser H, Walter M, Messmer BJ. Pulmonary embolectomy: review of a 15-year experience and role in the age of thrombolytic therapy. *European Journal of Cardiothoracic Surgery* 1996; **10**: 952–7.

Gulba DC, Schmid C, Borst HG, Lichtlen P, Dietz R, Luft FC. Medical compared with surgical treatment for massive pulmonary embolism. *Lancet* 1994; **343**: 576–7.

Hartz RS, Byrne JG, Levitsky S, Park J, Rich S. Predictors of mortality in pulmonary thromboendarterectomy. *Annals of Thoracic Surgery* 1996; **62**: 1255–9.

Leeper KV *et al.* Treatment of massive acute pulmonary embolism: the use of low doses of intrapulmonary arterial streptokinase combined with full doses of systemic heparin. *Chest* 1988; **93**: 234–9.

Cardiac tumors

Bortolotti U *et al.* Surgical excision of intracardiac myxomas: a 20-year follow-up. *Annals of Thoracic Surgery* 1990; **49**: 449–53.

Bulkley BH, Hutchins GM. Atrial myxomas: a fifty year review. *American Heart Journal* 1979; **97**: 639–43.

Goldstein DJ, Oz MC, Rose EA, Fisher P, Michler RE. Experience with heart transplantation for cardiac tumors. *Journal of Heart and Lung Transplantation* 1995; **14**: 382–6.

Holley DG *et al.* Diagnosis and management of fetal cardiac tumors: a multicenter experience and review of published reports. *Journal of the American College of Cardiology* 1995; **26**: 516–20.

Lam KY, Dickens P, Chan AC. Tumors of the heart. A 20-year experience with a review of 12, 485 consecutive autopsies. *Archives of Pathology and Laboratory Medicine* 1993; **117**: 1027–31.

Prichard RW. Tumours of the heart. *Archives of Pathology* 1951; **51**: 98–115.

Takach TJ, Reul GJ, Ott DA, Cooley DA. Primary cardiac tumors in infants and children: immediate and long-term operative results. *Annals of Thoracic Surgery* 1996; **62**: 559–64.

Pericarditis

Arsan S *et al.* Long-term experience with pericardectomy: analysis of 105 consecutive patients. *Thoracic and Cardiovascular Surgeon* 1994; **42**: 340–4.

Culliford AT, Lipton M, Spencer FC. Operation for chronic constrictive pericarditis: do the surgical approach and degree of pericardial resection influence the outcome significantly. *Annals of Thoracic Surgery* 1980; **29**: 146–52.

Lema LE. Surgical management of pyopericardium. *East Africa Medical Journal* 1993; **70**: 140–2.

Nataf P *et al.* Results of subtotal pericardectomy for constrictive pericarditis. *European Journal of Cardiothoracic Surgery* 1993; **7**: 255–6.

Suwan PK, Potjalongsilp S. Predictors of constrictive pericarditis after tuberculous pericarditis. *British Heart Journal* 1995; **73**: 187–9.

Tirilomis T, Unverdorben S, von der Emde J. Pericardectomy for chronic constrictive pericarditis: risks and outcome. *European Journal of Cardiothoracic Surgery* 1994; **8**: 487–92.

Walsh TJ, Baughman KL, Gardner TJ, Bulkley BH. Constrictive epicarditis as a cause of delayed or absent response to pericardectomy; a clinicopathological study. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 126–32.

Kawasaki disease

Burns JC *et al.* Clinical spectrum of Kawasaki disease in infants younger than 6 months of age. *Journal of Pediatrics* 1986; **109**: 759–63.

Hicks RV, Melish ME. Kawasaki syndrome. *Pediatric Rheumatology* 1986; **33**: 1151–73.

41.1 Surgical anatomy and radiology of the chest

Stephen Westaby and B. Trotman-Dickenson

[Anatomy](#)
[The thoracic cage](#)
[The thoracic cavity](#)
[Radiology of the chest](#)
[Chest wall](#)
[Thoracic cage](#)
[The diaphragm](#)
[Mediastinum](#)
[Trachea and main bronchi](#)
[Pleura](#)
[Interlobar fissures](#)
[Accessory fissures](#)
[The lung fields and pulmonary hilum](#)
[The lateral radiograph](#)
[Computed tomography of the chest](#)
[Further reading](#)

Anatomy

Effective treatment of thoracic trauma demands an in-depth knowledge of both anatomy and radiological imaging of the chest. Since relatively few major injuries are confined to a single body cavity, this knowledge must extend to the upper abdomen and root of the neck. Penetrating injuries in particular require detailed knowledge of the relationships between surface landmarks and intrathoracic and abdominal viscera. These change significantly during different phases of respiration.

The thoracic cage

The form of the thorax is that of a truncated cone, flattened in front and behind but rounded at the sides. The skeletal framework of the chest is formed anteriorly by the sternum and the costal cartilages, posteriorly by the bodies of the 12 thoracic vertebrae and the corresponding intervertebral discs, and by the ribs from their heads to their angles. On each side the chest is enclosed by the shafts of the ribs, from their angles to their cartilages. The sternum is made of three parts: the manubrium, body, and xiphoid process. The manubrium and body are not in the same plane, and thus form the sternal angle at their junction. This provides an important landmark, at which the second costal cartilage articulates with the sternum. The costal cartilages of the first to seventh ribs articulate with the sternum, whereas the costal cartilages of the eighth to tenth ribs usually attach to the cartilage of the rib above. The ventral ends of the cartilages of ribs 11 and 12 have no direct skeletal attachment. All ribs articulate dorsally with the vertebral column in such a way that their ventral end, together with the sternum can be elevated, as occurs during the inspiratory phase of respiration. The articulations of the costal cartilages with the sternum, apart from that of the first rib, are true synovial joints which allow freedom of movement.

The anterior wall of the chest is much shorter than is that of the posterior wall. During expiration the upper margin of the sternum is opposite the disc between the second and third thoracic vertebrae, and the xiphisternal joint is opposite the ninth or tenth thoracic vertebra. This relationship changes during inspiration so that the xiphisternal joint is elevated towards the eighth thoracic vertebra, and the suprasternal notch is raised to the level of the second thoracic vertebra. The ribs are, in general, directed downwards and forwards; during inspiration they become elevated and more horizontally placed by a hinge-like movement with its axis through the head and tubercle of the rib. Elevation is accompanied by lateral displacement which is inconspicuous in the uppermost ribs, but more pronounced in the fifth to ninth ribs (Fig. 1). Movement of the upper ribs increases the anteroposterior diameter of the chest, while movement of the lower ribs results mainly in an increase in transverse diameter. The short eleventh and twelfth ribs move mainly backwards in inspiration and forwards in expiration. Lateral movement of the lower ribs causes a widening of the infra-sternal angle, and permanent elevation of the ribs in the elderly produces the characteristic barrel-shaped chest. The thorax in a female is relatively shorter and more rounded than in the male, and the upper ribs are more mobile. These features provide greater movement of the upper chest, which is especially evident in the later stages of pregnancy.

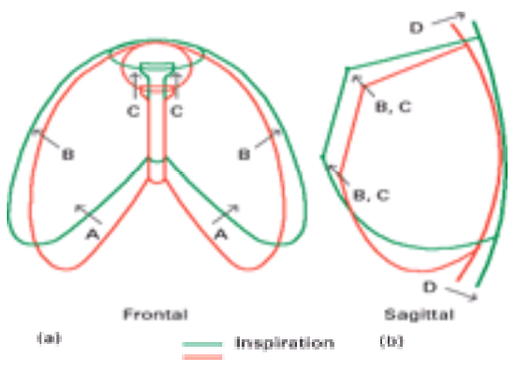


Fig. 1. (a) Shape and size of the thorax. (b) Internal aspect of the anterior chest wall and thoracic inlet. A, enlargement of the lower thorax aperture; B, lateral enlargement of the thorax; C, cranial displacement of the thorax; D, extension of the vertebral column.

The bodies of the thoracic vertebra project forward and greatly diminish the anteroposterior diameter of the chest in the median plane. The backward sweep of the posterior parts of the ribs produces a hollow on each side of the vertebral column which accommodates the greater part of the corresponding lung. Movements of the vertebral column contribute to changes in the volume of the thorax. The thoracic spine undergoes extension during inspiration. This causes still greater movement of the anterior ends of the upper ribs than would be produced by intrinsic movement relative to the vertebral column.

The inlet of the thorax is a narrow opening bounded by the body of the first thoracic vertebra, the first pair of ribs, and the upper border of the manubrium sterni. The plane of the inlet slopes very obliquely downwards and forwards, so that the anterior part of the apex of the lung is above the level of the anterior boundary of the inlet, though its posterior part only attains the level of the neck of the first rib. Patients with wounds at the root of the neck are therefore susceptible to pneumothorax. The structures that enter or leave the thorax through this inlet are the trachea, the oesophagus, the vagus and phrenic nerves on each side, the left recurrent laryngeal nerve, the thoracic duct, and the great arteries and veins carrying blood to and from the head, neck, and upper limbs. At the inlet, these are the right innominate artery and vein, the left subclavian artery, left common carotid artery, and left innominate vein (see Fig. 7). The outlet of the thorax is much larger than the inlet. It is bounded by the xiphisternal joint, the lower six costal cartilages, the twelfth rib, and the twelfth thoracic vertebra. The boundary of the outlet is curved, for it descends from the xiphisternum to the tip of the eleventh costal cartilage and then ascends towards the twelfth thoracic vertebra. The inner aspect of this lower margin gives attachment to the diaphragm, which forms a convex floor for the thorax and a concave roof for the abdomen. The upward bulging of the diaphragm greatly diminishes the vertical diameter of the thorax, so that the lower part of the chest greatly overlaps the upper part of the abdominal cavity, especially at the sides and behind (Fig. 2). There are three large openings in the diaphragm through which the aorta, oesophagus, and vena cava pass between the two cavities. The aortic opening transmits the aorta together with the thoracic duct and frequently the vena azygos as it ascends from the abdomen. The oesophageal opening transmits the oesophagus together with the vagus nerves. The vena caval opening transmits only this structure.

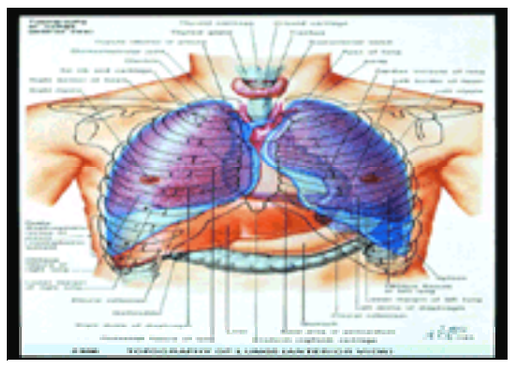


Fig. 2. Topography of the lungs and mediastinum (anterior view).

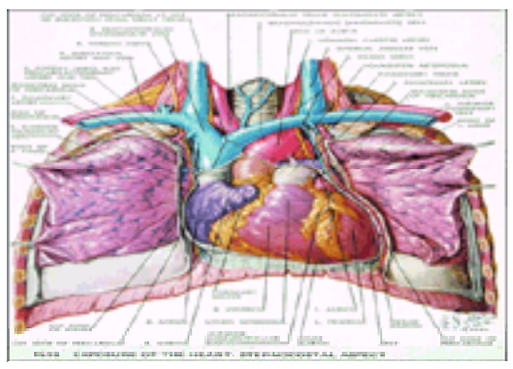


Fig. 7. The sternocostal aspect of the heart and great vessels.

The deep surface of the scapula fits against the posterolateral aspect of the thorax from the second to seventh ribs. It is held in apposition by the muscles attached to it, the only bony articulation to the chest being between the acromion process and the lateral end of the clavicle. The clavicle acts as a strut to hold the lateral angle of the scapula away from the thorax. The clavicle articulates at its medial end with the superolateral aspect of the manubrium.

Soft tissues of the chest wall

Deep to the skin and superficial fascia, which contains the mammary glands, the anterior chest wall is covered by three groups of muscles. These are the muscles of the upper extremity, those of the anterolateral abdominal wall, and the intrinsic muscles of the thorax (Fig. 3). The muscles of the upper extremity include the pectoralis major, pectoralis minor, serratus anterior, and subclavius. The pectoralis major and minor muscles are supplied by the medial and lateral anterior thoracic nerves, which are branches of the medial and lateral cords of the brachial plexus (Fig. 4). The serratus anterior muscle is supplied by the long thoracic nerve, which is a branch of the brachial plexus that courses inferiorly on the external surface of the muscle. This nerve may be damaged by penetrating wounds or chest drain insertion, causing winging of the scapula. The muscles of the anterolateral abdominal wall which originate on the chest are the external oblique muscle and the rectus abdominus muscle; both of these muscles are supplied by branches of the lower six thoracic nerves. When two or more of these branches are damaged at thoracotomy or during extensive thoracic injury, denervation of the muscles leads to ipsilateral bulging of the abdominal wall. The intrinsic muscles of the thorax include the external and internal intercostal muscles and the transversus thoracis muscle. The external intercostal muscles arise from the lower border of the rib above and insert on to the upper border of the rib below. Their fibres are directed downward and medially, extending from the tubercles of the ribs to the beginnings of the costal membranes. The internal intercostal muscles arise from the inner lip and floor of the costal groove of the rib above and from the related costal cartilage. They insert on to the upper border of the rib below. These muscles extend from the sternum to the angle of the ribs and then continue to the vertebral column as the posterior intercostal membranes. The fibres are directed downwards and laterally. Both external and internal intercostal muscles are supplied by the related intercostal nerves. The transversus thoracis muscle lines the inner surface of the anterior thorax wall as a thin sheet of muscular and tendinous fibres. The muscle arises from the posterior surfaces of the xiphoid process, the lower third of the body of the sternum, and the sternal ends of the related costal cartilages. It inserts by muscular strips into the inner surface of the second or third to the sixth costal cartilages.

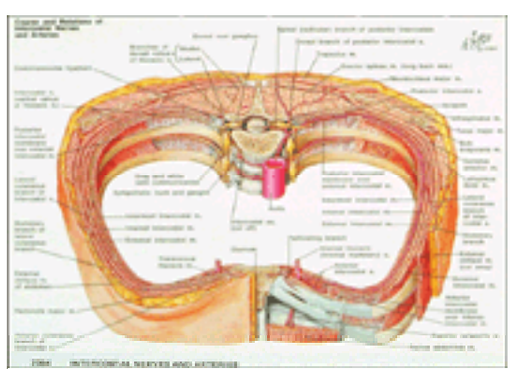


Fig. 3. Course and relations of the intercostal nerves and arteries.

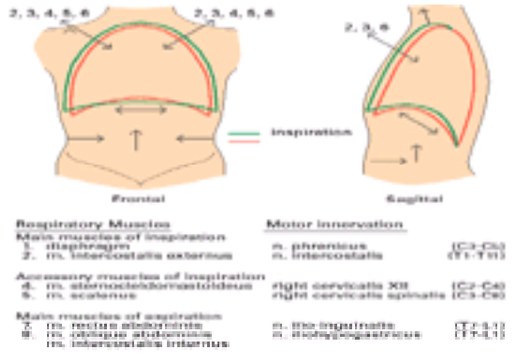


Fig. 4. Respiratory muscles.

Beneath the skin and superficial fascia of the dorsal aspect of the chest, a superficial group of muscles connect the upper limb to the vertebral column (Fig. 3). These are the trapezius, latissimus dorsi, rhomboids, and levator scapulae. The trapezius muscle arises from the posterior aspect of the skull, the posterior margin of the ligamentum nuchae which spans the spinous processes of the cervical vertebrae, and from all the thoracic vertebrae and their related supraspinous ligaments. The upper group of fibres inserts on to the lateral third of the clavicle; the middle group of fibres inserts on to the medial margin of the acromion process and upper margin of the posterior border of the spine of the scapula.

The lower fibres converge into an aponeurosis, which slides over the triangular area of the medial end of the spine of the scapula and is attached at the apex of this triangle. When the muscle contracts the scapula is pulled medially, rotates, and carries the shoulder upwards. When the shoulder is fixed the upper fibres tilt the head, so that the face turns upwards towards the opposite side. The latissimus dorsi muscle is expansive, with a broad origin from the outer lip of the iliac crest and from an extensive aponeurosis which is attached to the spinous processes of the lower six thoracic vertebrae, the lumbar and sacral vertebrae, and their related supraspinous ligaments. The muscle is inserted into the humerus and helps produce extension, adduction, and medial rotation at the shoulder joint. It also helps to depress the raised arm against resistance. The rhomboid muscles originate from the spinous processes and supraspinous ligaments of the cervical and thoracic vertebrae, and insert into the vertebral border of the scapula. These muscles draw the scapula towards the vertebral column and also slightly upwards. The lower fibres of the major rhomboid muscle rotate the scapula and depress the shoulder. The levator scapulae muscle originates from the first four cervical vertebrae, inserts into the vertebral border of the scapula, and as its name suggests, elevates the scapula, drawing it medially and rotating it, so that the tip of the shoulder is depressed.

Deep to the muscles that connect the upper limb to the vertebral column lie the serratus posterosuperior and inferior muscles. The superior muscle helps increase the size of the thoracic cavity by elevating the ribs. The inferior muscle tends to pull the last four ribs downward and outward. Just deep to the serratus posterosuperior muscle lie the thoracic portions of the splenius, cervicus, and capitus muscles. The groove lateral to the spinous processes of the thoracic vertebrae is filled by the sacrospinalis muscle. This is covered by a strong layer of lumbodorsal fascia.

Surface landmarks

Surface landmarks on the chest wall are important when siting surgical incisions or placing intercostal drains. These landmarks may be obscured to a greater or lesser extent by body fat. The upper border of the manubrium, the sternal angle, and the xiphisternal joint can be felt and often seen in the majority of individuals. The second costal cartilage articulates with the sternum at the sternal angle. This arrangement facilitates identification of the second intercostal space. The seventh costal cartilage is usually the lowest to articulate with the sternum, although on occasions the eighth also reaches it. The tenth costal cartilage forms the lowest part of the costal margin, whilst the tips of the eighth, ninth, and tenth cartilages articulate with the anterior part of the costal margin, and can usually be felt easily. The eleventh and twelfth ribs are palpable posteriorly in slender subjects, but do not articulate with the costal margin.

In clinical practice, the ribs are usually counted beginning with the second costal cartilage at the sternal angle. If the sternal angle cannot be felt, the first intercostal space is identified as the depression below the clavicle: the second rib lies immediately below it. The lower margin of the pectoralis major muscle overlaps the fifth rib and cartilage at about the same level as the xiphisternal joint. This site corresponds proximally to the level of the ninth or tenth thoracic vertebra. The upper end of the linea semilunaris of the abdomen originates at the tip of the ninth costal cartilage. This is usually about half way up the costal margin and at the level of the first lumbar vertebra. With the patient positioned for a lateral thoracotomy, the arm flexed and raised above the head, the tip of the scapula usually overlies the fifth intercostal space.

The spinous processes of all the vertebrae from the seventh cervical to the fourth lumbar can usually be felt in the midline of the back (Fig. 5). With the patient bent forward in a sitting position the lumbar spinous processes become more widely separated and easier to identify. The first thoracic vertebra is usually the most prominent process, though the seventh cervical vertebra is equally promi-nent, or more so in some individuals. Flexion in the neck increases the prominence of these spinous processes. The spinous processes are normally counted downwards from the seventh cervical or upwards from the fourth lumbar, which can be recognized by its position at the same level of the highest part of the iliac crest. In the ordinary standing position the scapula overlies the first to seventh ribs. The root of the spine of the scapula can be used as an approximate guide to the third thoracic spinous process and is usually at about the same level as the suprasternal notch. The medial border of the scapula is somewhat obscured by the rhomboid muscle, and the distance of the medial border of the scapula from the spine shows great individual variation. The upper part of the scapula is deeply placed between the thick muscle mass, comprising supraspinatous and the trapezius. The clavicles anteriorly on either side form sinuous ridges at the junction of the thorax and neck. The suprasternal notch is bounded inferiorly by the upper border of the manubrium, and separates the lower attachments of the two sternomastoid muscles. A depression, the infraclavicular fossa is found below the middle third of the clavicle and provides access for cannulation of the subclavian vein.

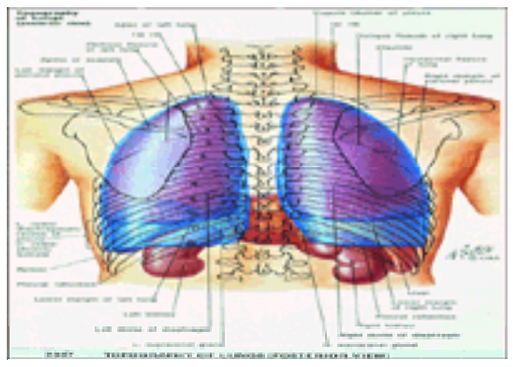


Fig. 5. Topography of the lungs, chest wall, and spine (posterior view).

The male nipple commonly lies opposite the fourth intercostal space, almost 10 cm (4 inches) from the midline, although its position varies considerably. The female breasts occupy an approximately circular area, which extends from the second rib to the sixth, and from the sternal margin to the axilla. The lower edge of the breast is marked by a pronounced, slightly curved groove, usually opposite the sixth rib in the midclavicular line. In girls younger than childbearing age, this line may be as high as the fifth intercostal space; in older or multiparous women it may be as low as the lower border of the seventh rib. The breasts are seldom exactly equal in size and are separated from each other by a cleft. There is an upward extension into the axilla, which obscures the lateral border of the pectoralis major muscle, where this enters the anterior wall of the axilla. Two-thirds of the breast overlies the pectoralis major muscle; the remaining one-third covers the serratus anterior and a small part of the external oblique and rectus abdominus muscles. The position of the female nipple varies greatly with the size, shape, and position of the breast. During pregnancy the glands greatly increase in size, whereas in old age they become atrophic.

Nerves and vessels in the chest wall (Fig. 3)

Each thoracic spinal nerve emerges from the intervertebral foramen and divides almost immediately into the dorsal ramus and ventral ramus. The dorsal ramus of the thoracic nerve passes posteriorly through the erectus spinae muscle, the trapezius muscle, and other superficial muscles of the back to reach the superficial fascia. Here it divides into medial and lateral branches, which supply the skin. The ventral ramus of the thoracic nerve forms the intercostal nerve of that particular level. The white and grey rami communicantes connect the ganglia of the sympathetic trunk and the thoracic nerve of the same level near the origin of the ventral ramus. From the seventh to the eleventh thoracic vertebrae the ventral rami continue forward from the intercostal spaces into the anterior abdominal wall. Each intercostal nerve runs forward in the thoracic wall on the inner aspect of the internal intercostal muscle, lying inferior to the intercostal vein and artery. It gives off a collateral branch to the lower part of the intercostal space, as do the corresponding vein and artery. Each nerve has a lateral cutaneous branch at the lateral aspect of the thorax that pierces the overlying intercostal muscles to reach the subcutaneous tissues. The nerve then divides into an anterior (mammary) and a posterior branch. The intercostal nerve ends by becoming the anterior cutaneous nerve at the anterior end of the intercostal space.

P>The aorta, which lies on the left anterior aspect of the vertebral bodies, gives off pairs of posterior intercostal arteries. These pass forward in the upper part of the intercostal space between the intercostal vein above and intercostal nerve below. Anteriorly they anastomose with the intercostal branches of the internal mammary and musculophrenic arteries. Collateral branches run in the inferior part of the intercostal space. The right posterior intercostal arteries cross the anterior aspect and right side of the corresponding vertebral body as they pass to reach the right-sided intercostal space. The internal mammary arteries originate from the right and left subclavian arteries then descend through the thoracic inlet to pass along the inner aspect of the costal cartilages as far as the sixth costal cartilage, where the vessel branches into the superior epigastric, gastric, and musculophrenic arteries. The intercostal and internal mammary arteries are accompanied by the corresponding veins. The anterior intercostal veins end in the musculophrenic vein and the venae commicantes of the internal mammary artery. The posterior intercostal veins, apart from the first, join the azygos vein. On the left the lower veins join the hemiazygos venous system, whereas the second and third veins form the left superior intercostal vein which crosses the arch of the aorta and ends in the left innominate vein. The first intercostal veins on both right and left sides join the innominate vein.

The thoracic cavity

The cavity of the thorax is completely divided into right and left lateral parts by a thick median partition, the mediastinum. The mediastinum is made up of a large number of important structures embedded in connective tissue extending from the sternum anteriorly to the vertebral column. The principal structures are the heart

within the pericardium, the aorta and great vessels, the oesophagus and trachea, the vagus and phrenic nerves, the remains of the thymus gland, and numerous lymph nodes. The connective tissue ensheaths these structures but is both yielding and elastic, to accommodate the dilatation and contraction of the heart and blood vessels. Anatomically, the mediastinum is described in four parts: (a) the superior mediastinum above the plane passing from the lower border of the manubrium sterni to the lower border of the fourth thoracic vertebrae. The mediastinum below that plane is subdivided into (b) the middle mediastinum occupied by the pericardium and its contents with the phrenic nerves on either side, (c) the anterior mediastinum in front of the pericardium, and (d) the posterior mediastinum behind the pericardium and diaphragm (Fig. 6).

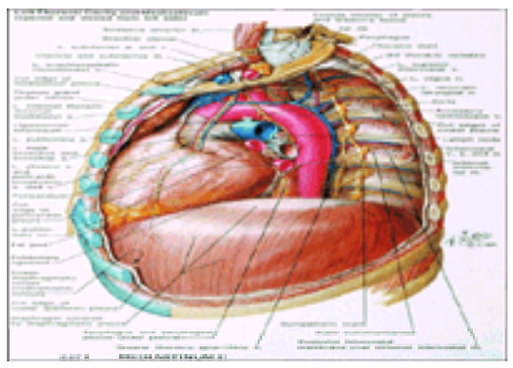


Fig. 6. The mediastinum from the left pleural cavity.

The lateral parts of the thoracic cavity are lined with a thin layer of pleura and are almost completely filled by the lung, which evaginates the pleural sac from the mediastinal side. Each lung lies freely within its pleural cavity, except along the medial surface where it is attached to the mediastinum by the root, consisting of the bronchus coming from the bifurcation of the trachea and the pulmonary artery and two pulmonary veins passing to and from the heart. When air is squeezed out of the lung, its bulk is not nearly sufficient to fill this space which it normally occupies. Under normal conditions the lung remains distended by atmospheric pressure, but when air is admitted into the pleural cavity the atmospheric pressure on the outer surface of the lung is made equal to that on the interior, so that the elasticity of the lung substance forces air out. The lung then shrinks to about one-third of its previous bulk. Similar shrinkage also occurs when blood or other types of fluid accumulate in the pleural sac.

Topography of the lungs

The apex of each lung reaches as far superiorly as the vertebral end of the first rib and usually extends about 2.5 cm above the medial one-third of the clavicle (Fig. 2 and Fig. 5). The anterior border of the right lung descends behind the sternoclavicular joint and almost reaches the midline at the sternal angle. It then continues downwards behind the sternum to the level of the sixth chrondrosternal junction. Here the inferior border curves laterally and inferiorly, crossing the sixth rib in the midclavicular line and the eighth rib in the midaxillary line (Fig. 2). It then passes posteriorly towards the midline at the level of the spinous process of the tenth thoracic vertebra (Fig. 5). These levels apply to the lung in expiration: on inspiration the levels for the inferior border are usually two ribs lower as the expanding lung descends into the costodiaphragmatic recess of the pleura.

The position of the anterior border of the left lung is similar to that of the right, though at the level of the fourth costal cartilage it deviates laterally in front of the heart, causing a cardiac notch in this border of the lung. The inferior border of the left lung is similar in position to that of the right, though it usually extends a little more inferiorly since the right lung is pushed up by the liver below the diaphragm on that side. The right lung is composed of three lobes, the upper, middle, and lower, separated by the oblique and horizontal fissures. The left lung ordinarily has only two lobes, the upper and lower, separated by the oblique fissure. There is no horizontal fissure in the left lung. Because neither lung extends as far inferiorly as the parietal pleura, some of the diaphragmatic parietal pleura remains in contact with the costal parietal pleura. This potential space, which may fill with air or fluid and whose size varies with the phases of respiration, is called the costodiaphragmatic recess. A similar, but much smaller, area is present where the anterior border of the lung does not extend to the medial limit of the pleural cavity. This area is called the costomediastinal recess. The diaphragm separates the heart and right lung from the liver. The left lung is separated by the diaphragm from the stomach and spleen. Posteriorly, the lung extends from the level of the first thoracic vertebra as far inferiorly as the diaphragm, and the base of the lung is applied to the superior surface of the diaphragm. Because of the diaphragm's domed shape the highest point on the base of the right lung is at the level of the eighth to ninth thoracic vertebra; the highest point of the base of the left lung is about 1 cm lower (Fig. 6). From these high points, the bases of the two lungs follow the curves of the diaphragm to reach the lower levels previously described. The weight of a healthy adult right lung, containing an average amount of blood, is about 620 g, and that of the left is 570 g.

The heart and great vessels

The heart is contained within a fibroserous sac, the pericardium (Fig. 7). This also surrounds parts of the great vessels as they enter and leave the heart. The fibrous pericardium is conical, the apex being pierced by the aorta, the pulmonary trunk, and the superior vena cava. The base rests on the diaphragm and is fused with the central tendon in the median plane. The right posterior part of the pericardium is pierced by the inferior vena cava. The diaphragm separates the pericardium from the liver and from the fundus of the stomach.

The pericardium lies behind the body of the sternum and the cartilages of the second to sixth ribs, inclusive. Its anterior surface is separated from them by the lungs and pleurae, except in the median plane where the anterior surface of the fibrous sac is attached to the upper and lower parts of the body of the sternum. On the left side, the pleura retreats from the sternum at the level of the fifth costal cartilage. This is the area of the cardiac notch of the left lung. Each side wall of the pericardium is in relation to the mediastinal pleura. The median part of the posterior surface of the pericardium is in front of the descending aorta and oesophagus. The lateral part on each side posteriorly is supported by the pleura and lung. Within the fibrous pericardium, the serous pericardium is a closed and evaginated sac lining the fibrous sac and enveloping the heart and parts of the great vessels.

Details of intracardiac anatomy are only of direct relevance to the cardiac surgeon. Of great importance in trauma management are the relationships between cardiac structures and surface markings, which allow damage to intracardiac structures to be predicted in patients with penetrating chest wounds. The precise position of the heart is variable. The outline of the heart as projected on the anterior chest wall in different subjects during quiet respiration, may be classified into one of three general types: oblique, which is the most usual, horizontal, and vertical. These types tend to be related to general body build and the position of the diaphragm. The horizontal heart tends to occur in individuals with a high diaphragm and a short, broad trunk, whereas the vertical occurs in those with a low diaphragm and slender trunk. Surface markings also vary according to the position of the body, the phases of respiration, and with intrinsic movements of the heart. This variability clearly makes it impossible precisely to outline the heart in any individual.

In general terms, it may be stated that in the erect position the heart reaches a little to the right of the sternum and some 7.5 to 10 cm (3 to 4 inches) to the left of it. Two-thirds of its vertical height lies behind the sternum, reaching up to the sternal angle. Up to one-third may lie below the xiphisternal joint. The heart is 2 cm higher in the recumbent position than in the erect position. The upper border of the heart is concealed by the aorta and pulmonary trunk. Its position is marked on the surface by a line drawn from the lower border of the second left costal cartilage, 4 cm from the median plane, to the upper border of the right third cartilage, 2.5 cm from the median plane. Since the pulmonary arteries run along the upper border of the heart, the right two-thirds of this line indicate the position of the right artery and the left third indicates the left artery. The right border of the heart is marked by a line slightly convex to the right, drawn from the upper border of the third right costal cartilage, 2.5 cm from the median plane, down to the sixth right cartilage, 1.25 cm from its junction with the sternum. The lower or diaphragmatic border of the heart is concave downwards, in correspondence with the upper convexity of the diaphragm on which it rests. There is a slight inclination downwards as the border is traced from right to left. This is marked by a line drawn from the sixth right costal cartilage near the sternum to the point opposite the apex in the fifth left intercostal space about 7.5 cm from the median plane. The left heart border is marked by a curved line drawn from a point opposite the apex to a point on the lower border of the second left costal cartilage, 4 cm from the median plane.

The precise level of the heart is influenced mainly by the level of the diaphragm. When the heart is situated at a low level the positions of the great arteries also differ, the arch of the aorta and pulmonary trunk being drawn downwards. When the heart is raised by upward movement of the diaphragm in expiration it becomes shorter in its vertical diameter and lies more transversely. When the diaphragm is lowered in inspiration the heart descends and appears narrower, assuming a more vertical position. The state of distension of the fundus of the stomach may affect the level of the dome of the diaphragm, and in turn modifies the position of the heart. When the fundus is greatly distended the heart tends to be pushed towards the right. Lateral displacement may occur in patients with scoliosis; displacement also occurs when pressure conditions in the pleural cavities become unequal, as in pneumothorax or haemothorax. The actual size of the heart changes during the cardiac cycle and is affected by intrathoracic pressure. When the pressure becomes relatively high at the end of the forceful expiration, the heart is, for a short period, diminished in size. The size of the heart relative to that of the thorax varies greatly in different subjects and it is therefore difficult to determine whether a slight pathological enlargement is present in a given subject. The size ratio of the heart to the thorax varies in normal subjects between one-third to one-half; this affords a rough guide to

normal size. The heart in a child is relatively larger compared to the whole thorax than is the case in the adult.

The tricuspid valve lies behind the sternum, extending from the midline at the level of the fourth costal cartilage down towards the sixth right chrondrosternal junction. The mitral valve lies behind the left side of the sternum obliquely at the level of the fourth costal cartilage. The pulmonary and aortic valves lie close to the mitral valve, the pulmonary valve lying horizontally behind the inner end of the third left costal cartilage and adjoining part of the sternum (Fig. 8). The aortic valve is placed obliquely behind the left side of the sternum, opposite the third intercostal space. The right border of the heart begins behind the second costal cartilage with the superior vena cava, this joins the right atrium at the level of the third costal cartilage. The right atrium itself extends between 1 to 2 cm lateral to the sternum as far as the xiphisternal joint. The left border of the heart starts between the first and second costal cartilages with the aortic knuckle; beneath this is the left pulmonary artery and pulmonary trunk. The left atrial appendage lies below the left pulmonary artery between the second and third costal cartilages. The left ventricle then extends obliquely laterally from the second costal cartilage to the apex of the heart, which generally lies 9 cm from the midline behind the fifth left intercostal space. The site of the apex may vary in a normal person from 6 to 10 cm lateral to the midline and in a vertical direction ranges from the fourth intercostal space to the sixth costal cartilage. The diaphragmatic border of the heart consists principally of the right ventricle to within 2.5 cm of the apex.

Fig. 8. The aorta, ventricles, and valves.

The orifices of the coronary arteries are immediately above the level of the cusps of the aortic valve. The right coronary artery passes forward between the pulmonary trunk and the right atrial appendage and then runs downwards in the atrioventricular groove to the lower part of the right margin of the heart, around which it curves before proceeding along the diaphragmatic surface. During this course, this vessel is protected by the body of the sternum on the diaphragmatic aspect of the heart; the posterior descending and left ventricular branches are vulnerable to penetrating injury passing upwards from the costal margin. The left coronary artery runs for a short distance towards the left behind the main pulmonary artery, and then passes forward between the pulmonary trunk and the left atrial appendage. The left main coronary artery divides into the circumflex branch, which curves backwards and downwards, and the left part of the atrioventricular groove reaching the lower border of the base of the heart, where its termination approximates to branches of the right coronary artery. Marginal branches of this vessel supply the posterior aspect of the heart. The left anterior descending coronary artery arises at the point where the left main stem appears between the pulmonary trunk and left atrial appendage. It then runs in the anterior interventricular groove towards the apex giving septal branches to the interventricular septum, and diagonal branches to the anterolateral aspect of the left ventricle. Both the left anterior descending and circumflex coronaries are vulnerable to penetrating wounds of the left chest.

Radiology of the chest

Recent developments in imaging methods and interventional techniques have modified the radiological assessment and management of patients requiring thoracic surgery. However, the plain chest radiograph remains the initial investigation. In acutely ill or injured patients the radiograph is first evaluated for immediate life-threatening complications such as pneumothorax, haemothorax, rib fractures, and mediastinal air or widening indicative of aortic laceration. More detailed analysis of the radiograph may then be performed. Following stabilization of the patient's condition, further radiological investigations such as computed tomography or arteriography may be indicated.

The quality of the chest radiograph is dependent on the ability of the patient to co-operate with the procedure. Supine anteroposterior views may be misleading: in particular a pneumothorax may be missed. Whenever possible an erect or semierect view should be taken. If this is not possible a lateral decubitus view with the injured side uppermost, should be performed.

A knowledge of the normal anatomy of the chest and its variations is essential for the accurate interpretation of the chest radiograph, which should be viewed in a systematic manner. Scrutiny of the radiograph should be performed at the standard viewing distance and from afar (2 m). Fixed distance viewing is associated with an increased risk of perceptual error. At the standard viewing position, lesions with sharply defined borders are readily appreciated. However, with distance viewing, masses such as lung tumours with poorly defined margins are better perceived; in particular, subtle differences in lung density are more obvious.

It is important to remember that individuals with thoracic trauma may have other important chest problems unrelated to their injuries when reviewing the radiograph. An unsuspected lung cancer may be detected, or the superior mediastinum may be widened by an intrathoracic goitre. It is therefore important to perform a careful clinical examination of the chest, neck, and abdomen before interpreting the chest radiograph. Appropriate diagnoses can only be made in the light of combined physical and radiological findings.

A recommended sequence of inspection of the chest radiograph is as follows (Fig. 9 and Fig. 10): chest wall (the soft tissues and bony thoracic cage); the diaphragm and subdiaphragmatic area; the mediastinum; the trachea and major airways; the pleura; the lung fields.

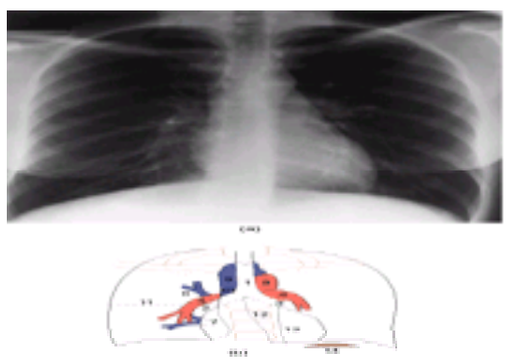


Fig. 9. (a) Normal chest radiograph: posteroanterior projection in an asymptomatic 25-year-old woman. (b) A diagram to show the normal anatomical structures: (1) trachea, (2) right main bronchus, (3) left main bronchus, (4) left pulmonary artery, (5) right pulmonary artery, (6) right upper lobe veins, (7) right middle and lower lobe veins, (8) aortic knuckle, (9) superior vena cava, (10) azygos vein, (11) horizontal fissure, (12) azygo-oesophageal line, (13) left paraspinal line, (14) gastric fundus.

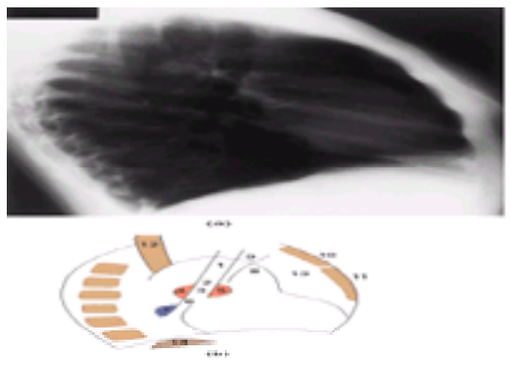


Fig. 10. (a) Normal chest radiograph: lateral projection in an asymptomatic 25-year-old woman. (b) A diagram to show the normal anatomical structures: (1) trachea, (2) right upper lobe bronchus, (3) left upper lobe bronchus, (4) left pulmonary artery, (5) right pulmonary artery, (6) right intermediate bronchus, (7) pulmonary venous confluence, (8) aortic arch, (9) brachiocephalic vessels, (10) manubrium, (11) sternum, (12) scapula, (13) retrosternal region, (14) gastric fundus.

Chest wall

Lesions in the bone or soft tissue of the chest wall may be misinterpreted as representing intrathoracic disease and may be the most important diagnostic feature. Surgical emphysema (air tracking within the tissues planes and muscle bundles) may be more obvious than an underlying pneumothorax.

Two muscles are commonly visualized. The anterior axillary fold formed by the pectoral muscles is seen curving medially and inferiorly from the axilla to the rib cage. In muscular men, the pectoralis major may appear as a continuation of the anterior axillary fold, passing obliquely across both lungs. This muscle is responsible for the increased shadowing over the middle and upper part of the lung. When absent this will give rise to an area of hypertranslucency in that region of the lung, and a misleading impression of emphysema. In women, such absence is usually the result of a mastectomy, but it may also be a congenital abnormality (Poland syndrome). The sternomastoid muscle may be visualized as a density whose lateral margins parallel the spine as it runs down vertically from the lung apices over the medial third of the clavicle. This may cause confusion in the apices of the lung, where the muscle shadow may be misinterpreted as pleural thickening.

The presence and symmetry of breast shadows should be taken into account when assessing the density of the lower lung zones. The female breast typically extends from the second to the sixth rib. Prominent breast tissue in the male suggests gynaecomastia. The nipples are identified by their size (0.5 to 1 cm), shape, and location. The lateral border of the nipple is usually sharply defined, but the medial border is poorly defined.

Companion shadows are reflections of soft tissue visualized as 2 to 3 mm shadows which parallel the superior aspect of the clavicles and the inferior and inferolateral margins of the first and second ribs. Companion shadows may be mistaken for pleural thickening. In obese individuals the companion shadows may be wider due to an abundance of extrapleural fat. The frequency with which the companion shadows are seen varies. In normal individuals the companion shadows may be bilaterally symmetrical, asymmetrical, or invisible on both sides (Fig. 11).



Fig. 11. Companion shadow: the muscle curves downwards and laterally to blend with the companion shadow on the superior aspect of the clavicle (white arrowhead).

Thoracic cage

The thoracic cage should appear symmetrical. The general shape, direction, and spacing of the ribs should be noted. The ribs run roughly parallel, the anterior ends of the ribs lying at a lower level than the posterior ends. The posterior portions of the ribs are directed downward and laterally whereas the anterior parts are directed downward and medially. Deformities of the dorsal spine distort the contour of the chest and may cause considerable difficulty in interpretation of the chest radiograph. In kyphosis, the posterior parts of the rib are crowded together. In scoliosis the ribs are more vertically inclined on the concave side of the spine.

The first rib is identified by its broader anterior end and by its articulation with the first thoracic vertebra. The latter can be distinguished from the seventh cervical vertebra by the orientation of its transverse processes which are inclined upwards and laterally, whereas those of the seventh cervical vertebra point downwards and laterally (Fig. 16). By following the ribs backwards the number of the thoracic vertebra can be ascertained, remembering that most of the ribs articulate with the upper parts of the bodies of the corresponding vertebra. The superior and inferior contours of the upper ribs are well defined. The inferior cortex of the ribs is less distinct in the mid- and lower thoracic region due to visualization of the vascular sulci. Focal superficial indentation of the inferior margin of the posterior ribs adjacent to the tubercle is a normal appearance. This indentation is distinct from pathological rib notching seen in coarctation of the aorta, which is characteristically situated more laterally near the midclavicular line. Calcification of the costal cartilages is common and of no clinical significance. Calcified costal cartilages may cause confusion, however, particularly in the line of the first rib where they may resemble intrapulmonary lesions. The first costal cartilage normally calcifies first, in the second decade, followed by calcification of the lower costal cartilages in middle age. The pattern of calcification is different between the sexes. In men the upper and lower borders of the cartilages calcify first followed by central calcification; this produces a marginal pattern. In women central calcification occurs as bands or as discrete nodules.



Fig. 16. Eventration of the left hemidiaphragm due to a congenital failure of muscular development of the diaphragm.

Congenital anomalies of the ribs are uncommon. Cervical ribs occur in 1.5 per cent of normal subjects and are usually bilateral but asymmetrical (Fig. 12(a)). Occasionally ribs may be absent or hypoplastic. Bifid or local fusion of ribs should be identified to avoid misinterpretation as representing a lung opacity. Supernumerary ribs such as an intrathoracic or pelvic rib usually occur on the right and are extremely rare (Fig. 13). Rib abnormalities may be associated with vertebral anomalies.

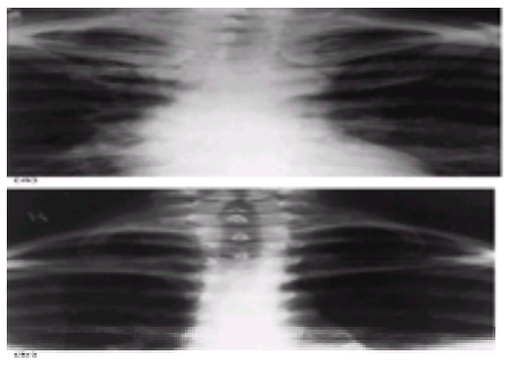


Fig. 12. (a) Cervical rib: bilateral symmetrical super numerary ribs arising from C7. The transverse processes of C7 are directed downwards. (b) Hypoplastic first rib: bilateral asymmetrical ribs arising from D1 vertebra. The transverse processes of D1 are directed upwards.



Fig. 13. Right intrathoracic rib: a rare supernumerary rib (arrow).

In injured patients the pattern of rib fractures should be noted. Subtle rib fractures are unlikely to be of clinical importance. Multiple bilateral rib fractures indicate a flail chest, while mediastinal injury should be suspected if the upper three ribs are fractured. Abdominal visceral injury should be considered if the lower three ribs are injured.

The clavicles are frequently fractured by thoracic trauma and may in turn lacerate the subclavian vessels or brachial plexus. The rhomboid fossa is an irregular notch on the inferior surface of the lateral aspect of the clavicle and is seen in 0.6 per cent of normal individuals (Fig. 14). The fossa gives rise to the costoclavicular ligament that binds the clavicle to the first rib: this is bilateral in 33 per cent of individuals in whom it is present. A foramen for the middle supraclavicular nerve may occasionally be seen either unilaterally or bilaterally. The conoid tubercle may be visualized on the inferior lateral aspect of the clavicle at the sight of insertion of the costoclavicular ligament.

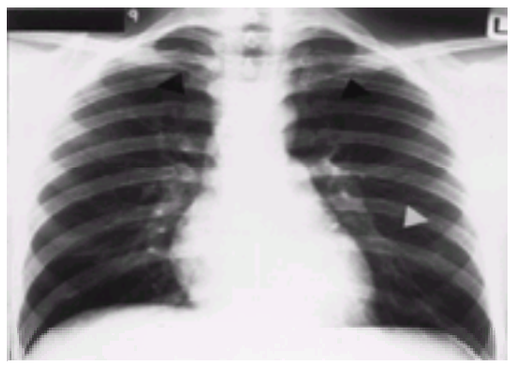


Fig. 14. Rhomboid fossa and left horizontal fissure: the fossa represents an irregular indentation on the inferior aspect of the clavicle near the sternal articulation, which gives origin to the costoclavicular ligaments (black arrowhead). The left horizontal fissure separates the lingula segments from the remainder of the upper lobe (white arrowhead).

Minor congenital anomalies of the manubrium are frequent. A wide manubrium sterni can project over the lung fields and may cause confusion, resembling a mediastinal abnormality. A depressed sternum (pectus excavatum) is common: its recognition is important to avoid erroneous interpretation of the mediastinum and lung fields. It can be confidently diagnosed on the posteroanterior radiograph by a number of features (Fig. 15). Typically, the heart appears shifted to the left with a straight left heart border and increased cardiothoracic ratio. There is loss of clarity of the right heart border and an ill-defined opacity is visualized at the right cardiophrenic area due to the soft tissues of the anterior chest wall, as they slope backwards and forwards to reach the sternum. The main pulmonary artery is prominent. The anterior ribs slope steeply downwards and there is undue clarity of the lower dorsal spine. The lateral radiograph will demonstrate the degree of sternal depression. In pectus carinatum, the anteriorly bowed sternum may be a developmental anomaly, associated with congenital heart disease, or acquired due to asthma.

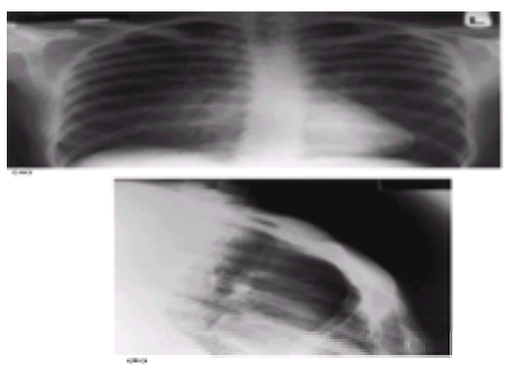


Fig. 15. Depressed sternum (pectus excavatum), posteroanterior and lateral views: the anterior end of the rib slopes sharply downwards and medially. The posterior ribs tend to be horizontal. There is an ill-defined opacity at the right cardiophrenic area due to the soft tissues of the anterior chest wall. The cardiac outline is triangular with a straight left heart border and it is displaced to the left.

On the frontal radiograph the sternum is obscured by the thoracic vertebrae and mediastinal shadows. Clinically, a sternal fracture is suggested by bruising, swelling, and tenderness over the sternum; it may be confirmed radiologically on a detailed lateral view. The scapula on a posteroanterior radiograph should not obscure the lung fields, but on an anteroposterior and lateral view, they project over the lung fields and their normal appearance must be recognized to avoid misinterpretation. This is particularly important when the scapulae are not symmetrically positioned; in this situation one lung field may appear less illuminated than the other, leading to the false diagnosis of a pathological process. By following the contour of the scapula and identifying the medial borders and inferior angles, this error should be avoided. A

fracture of the scapula is a hallmark of a very severe blow to the chest.

The diaphragm

The hemidiaphragm is seen on the frontal radiograph as a smooth curve, convex upwards with the highest point near the middle of the lung fields. Only the upper surface of the hemidiaphragm is seen because of contrast with the air-filled lungs; the undersurface blends with the opaque abdominal structures. The medial half of the left hemidiaphragm is not seen due to the heart lying immediately above and preventing the air-filled lung providing radiographic contrast. The lateral costophrenic recesses are usually clearly defined and form an acute angle. The recess is the first area to be obscured when the pleural cavity fills with blood or fluid. Exaggeration of the recess may occur with a small pneumothorax and may be the most obvious sign of a pneumothorax being present. The cardiophrenic angles are usually less well defined and less acute.

The height of the hemidiaphragms varies with the phase of respiration. On full inspiration the diaphragm lies between the fifth and seventh anterior rib space, at the level of the tenth thoracic vertebra. In 3 per cent of normal individuals it may be as high as the fourth interspace or as low as the seventh rib; it is rarely lower. If the right dome lies at the level of the seventh rib, it can only be considered normal if it has retained a curved contour and moves more than 3 cm on expiration. The height of the diaphragm is higher in women, individuals over 40 years, and in infants. Differences in height are frequently associated with body build, the diaphragm being lower in the slender type of individual and higher in those of a stocky build. In obese patients the diaphragm is displaced upwards by an increase in abdominal fat; displacement also occurs when there is excess gas in the intestinal tract. The diaphragm is commonly 3 cm higher in the supine position than in the erect position.

The right hemidiaphragm is typically higher than the left by up to 3 cm. In 10 per cent of individuals the left hemidiaphragm is at the same level as, or higher than, the right hemidiaphragm. Gaseous distension of the stomach or splenic flexure elevates the left hemidiaphragm. The height of the hemidiaphragm is determined by the systemic ventricle. The apex of the heart lies on the same side as the lowest hemidiaphragm. In the decubitus position the diaphragm is highest on the dependent side.

The range of diaphragmatic excursion is variable. On quiet respiration the range of movement is about 1 cm, while in deep respiration the excursion may be between 3 and 10 cm. In 60 per cent of individuals the right hemidiaphragm moves to a greater extent than the left, and in 10 per cent of individuals the left moves more than the right. However, the difference in excursion between the hemidiaphragms is less than 1.5 cm.

The curve of the hemidiaphragm may be assessed visually or by drawing a line across the chest, from one costophrenic recess to the other, and then a second line vertical to this from the highest point of the dome. In normal individuals it will measure 1.5 cm or more. If the vertical line measures 1 cm or less it indicates flattening of the dome, as in emphysema. Loss of definition of the diaphragmatic contour indicates poorly aerated lung adjacent to the diaphragmatic pleura and is a sign of underlying lung disease. If the clarity of the diaphragm is preserved, opacification of the lower zone is probably due to pathology in the overlying thoracic wall soft tissues. Translucencies beneath the hemidiaphragms should be identified as lying within an abdominal viscus, otherwise these lucencies represent free intraperitoneal air. If no gas is seen within the stomach below the diaphragm, a hiatus hernia or achalasia should be considered. Chilaiditi's syndrome is the interposition of the colon in front of and above the liver. It is common in children and the elderly and should not be mistaken for subdiaphragmatic free air.

Physiological variations in the diaphragmatic contour may occur. In scalloping of the diaphragm, the normal smooth convex contour is replaced by two smooth arcuate elevations. This is of no clinical significance and is usually confined to the right side. Diaphragmatic muscle slips may be visualized, originating from the lateral and posterolateral ribs as short curves, concave upwards along the lateral part of both hemidiaphragms. These are produced by exceptionally low descent of the diaphragm during inspiration and may occur in healthy young men. However, it is more commonly seen in association with severe pulmonary overinflation, as in asthma or emphysema. Muscle slips are prominent at full inspiration but disappear on expiration.

Eventration is characterized by congenital fibrous replacement of the diaphragm (Fig. 16). It is usually left-sided, but partial eventration of the anteromedial segment of the right hemidiaphragm may also occur, producing a hump. This should not be misinterpreted as a mass lesion or middle lobe consolidation. A fat pad in the anterior cardiophrenic recess may give a similar appearance. Accumulation of fat in the cardiophrenic recess is common and may occur in obese and non-obese individuals. On the left, the fat pad may simulate cardiac enlargement, but the fat is identified as having a density less than that of the adjacent heart.

Gastric herniation through the oesophageal hiatus is extremely common (Fig. 17). Herniation through the foramina of Morgagni and Bochdalek also occurs. The Morgagni hernia occurs anteriorly; liver, omentum, or bowel (especially colon) may herniate between the sternal and costal origins of the diaphragm (Fig. 17). The Bochdalek hernia is due to a posterior deficiency of the diaphragm and may contain retroperitoneal fat, part of a kidney, or the spleen.

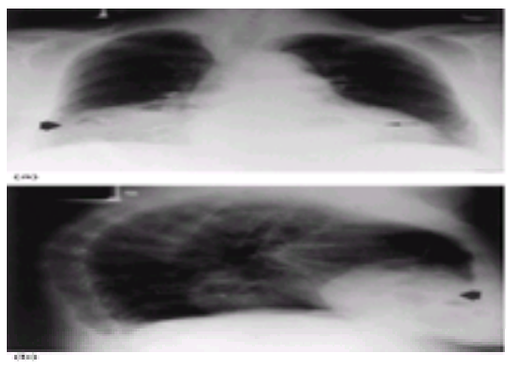


Fig. 17. Hiatus hernia and Morgagni hernia: posteroanterior and lateral projection. A retrocardiac density with an air fluid level represents the herniated stomach (small black arrow). The foramen of Morgagni is a defect in the anterior aspect of the diaphragm, almost exclusively on the right. The large hernia contains bowel and the omentum (large black arrowhead).

Diaphragmatic rupture can be difficult to diagnose, as the plain radiographic findings are non-specific, particularly in ventilated patients. However, it is an important diagnosis to consider, since early recognition may prevent the potentially lethal complication of bowel strangulation. Diaphragmatic rupture seems to be more common on the left side. Partial rupture may go unrecognized and may remain asymptomatic; alternatively the patient may present later with bowel obstruction. On the right, the appearance is of an apparently elevated diaphragm due to herniation of the liver through the defect. However, there is usually an associated haemothorax above the liver shadow and loss of the costophrenic angle (see Chapter 41.12). On the left, a ruptured diaphragm is suspected when stomach, colon, or small bowel prolapses into the chest and corresponding shadows are seen above the usual level. If necessary, this may be confirmed following oral administration of contrast medium. The shadow of an intact or ruptured spleen may be apparent above the level of the diaphragm and a haemothorax usually coexists. In patients receiving positive pressure ventilation, the supine radiograph may not reveal pathological elevation of the diaphragm. Ultrasound may be helpful, localizing the hemidiaphragm and demonstrating the presence of an intrathoracic abdominal viscus. A subpulmonic haematoma may mimic an elevated hemidiaphragm. However, the upper border of the opacity is less convex, the apex is situated more laterally, and the contour is less clearly defined compared to the hemidiaphragm. A decubitus film will confirm the presence of pleural fluid.

The infradiaphragmatic area has special importance in chest injuries, since the rib cage encompasses a substantial part of the upper abdomen and thoracic trauma frequently involves the abdominal viscera. Splenic lacerations are particularly common after blunt trauma. Fractures of the lower three ribs on the left side should alert one to the possibility of splenic trauma. The spleen lies to the left of the stomach, and when ruptured a perisplenic haematoma may occupy a considerable area under the diaphragm, displacing the gastric air bubble medially. The stomach is identified by an air bubble in the fundus which is usually larger than the bubble within the splenic flexure. The splenic flexure may contain faeces, thus identifying the viscus as colon. On the right side the subdiaphragmatic area is occupied by the opaque shadow of the liver. The diaphragm separates from the liver only when there has been rupture of a gas-filled viscus and this may readily be identified on an erect chest radiograph.

Mediastinum

The mediastinum lies in the midplane of the thorax between the lungs, from which it is separated by the visceral and parietal pleura. The thoracic inlet lies superiorly, and the diaphragm inferiorly. The conventional plain radiograph provides suboptimal demonstration of the mediastinum. However, with the use of higher voltage films (125–150 kV range) the mediastinal structures are more clearly visualized. The silhouette should be sharply defined. On the frontal radiograph the right mediastinal border is formed from above down by the right innominate vein, the superior vena cava, and the right atrium of the heart. When enlarged, the left atrium may form a double contour to the right heart border. The left atrium is situated at a higher level than the right atrium, so that the lower border of the heart consists of right atrium and inferior vena cava. The left border is formed by the left subclavian artery, the arch of the aorta, forming the so-called aortic knuckle, the pulmonary artery or its left

branch, the left atrial appendage, and the left ventricle. The pericardium itself is not normally visualized, but if air fills the pericardium after major airways injury the thin pericardial sac can be clearly seen. The appearance of the mediastinum has great important in chest injuries, since damage to the major vessels will produce mediastinal widening due to haematoma. This is usually apparent on the plain radiograph (see Chapter 41.12).

It is frequently difficult to determine whether a centrally-situated mass lesion lies in the mediastinum, the pleural space, or within the lung parenchyma. Three radiological features help to confirm a mediastinal lesion. A mediastinal mass will usually have a smooth, sharply defined contour. While a pulmonary mass may also have this appearance, a mediastinal mass will rarely have a poorly defined border. A mediastinal mass tends to form obtuse angles between its margins and contiguous lung, compared with the acute angle formed by a pulmonary lesion. Finally, a mass lesion that is in contact with, or displaces, mediastinal structures is likely to be mediastinal in origin. In convention, a mediastinal mass is classified by its location, in the anterior, middle, or posterior mediastinal compartment. This is a descriptive definition and is not related to anatomical boundaries.

The thoracic inlet is the region at the cervicothoracic junction lying in a transverse plane at the level of the first rib. It is higher posteriorly than anteriorly. The mediastinal fascia extends in continuity across the thoracic inlet where it merges with the cervical fascia. Thus the neck and mediastinum communicate freely, allowing the spread of air or disease through the cervical–thoracic continuum.

The aortopulmonary window lies with the aortic arch above and the left pulmonary artery below. It contains the ligamentum arteriosus, ductus nodes, left recurrent laryngeal nerve, and fat. The aortopulmonary line is a sharply-defined interface that extends vertically from the aortic arch to the level of the left main bronchus. Lymphadenopathy produces a prominent convex bulge in this region. End-on visualization of the left superior intercostal vein may produce a protuberance along the aortic knuckle which has been termed the aortic nipple. The nipple should not be mistaken for lymphadenopathy.

The superior mediastinal shadow is caused by the great systemic vessels. The heart hangs from these vessels, known as the vascular pedicle. The width of the pedicle is measured from the point at which the superior vena cava crosses the right main bronchus to a line drawn perpendicular to a point at which the left subclavian artery arises from the aortic arch. In the majority of normal individuals the pedicle measures less than 58 mm. In cardiogenic oedema, the vascular pedicle is increased in width while in non-cardiogenic oedema the pedicle is of normal or reduced width.

The heart lies in the midline, projecting to the left. There is a wide variation in normal cardiac size, which depends on body build. Traditionally, cardiac size is assessed by comparing the transverse cardiac diameter to the transverse diameter of the chest (cardiothoracic ratio). The value of 50 per cent is widely accepted as the upper limit of normal for this ratio. However, this is exceeded in 10 per cent of normal subjects. Individuals with a small cardiac diameter may undergo a considerable increase in cardiac size without exceeding a cardiothoracic ratio of 50 per cent. It has been suggested that an absolute value of 15.5 cm for the upper limit of normal for cardiac diameter is a more appropriate measurement. Serial radiographs are useful for monitoring cardiac size. The size and contour of the heart varies with the influence of systole and diastole, but on average, during the cardiac cycle, the cardiac diameter does not vary by more than 2 cm. The cardiac configuration is also affected by the height of the diaphragm. The greater the degree of pulmonary inflation the lower the position of the diaphragm and the longer and narrower is the cardiovascular silhouette. A poor respiratory effort will therefore result in a short squat configuration to the heart. The heart will also appear broader when the subject is supine compared to the erect position. In infants and young children, a cardiothoracic ratio of 60 per cent is accepted as normal.

Although the shape of the heart varies between long narrow, and vertical to a more rounded horizontal shape, the left heart border between the aortic knuckle and the left ventricle should maintain a concave contour. In cardiac tamponade when the pericardium is filled with blood clot, this concavity is lost and the heart adopts a globular shape. Since the pericardium is relatively non-distensible the overall size of the cardiac shadow may not perceptively increase in tamponade. Pericardial fat pads may be visualized on the frontal and lateral radiographs. These are more commonly seen on the left; the configuration is variable and may change shape with respiration (Fig. 18).

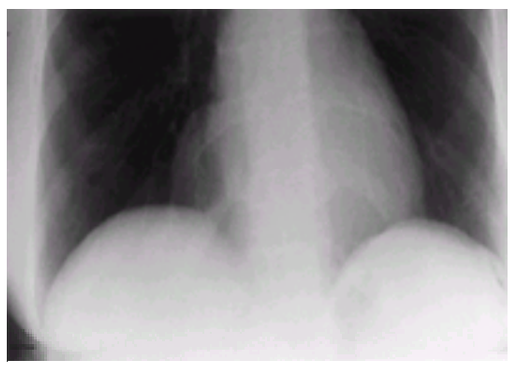


Fig. 18. Right pericardial fat pad: localized posteroanterior view of the lung bases. The fat pad occupies the anterior cardiophrenic angle. The fat is less opaque than the cardiac opacity.

Rupture of the aorta follows high speed deceleration trauma and is becoming increasingly more frequent. Approximately 95 per cent of ruptures occur in the region of the aortic isthmus, with development of a mediastinal haematoma. The radiological features of aortic injury are discussed in Section 35.12. The mediastinal haematoma may remain localized or may rupture into the left pleural cavity.

Air may enter the mediastinum, via the deep fascial planes of the neck or, rarely, by dissection from the retroperitoneal space. Pneumomediastinum is a highly significant finding. Its cause must be identified, as prompt treatment of perforation of trachea, bronchi, or oesophagus may be life-saving. Pneumothorax may occur secondary to pneumomediastinum but pneumothorax never causes pneumomediastinum. The signs of a pneumomediastinum are linear streaks of gas within the mediastinum, and displacement of the mediastinal pleura of the heart and other mediastinal structures. Retrosternal gas collections and visualization of the central portion of the diaphragm due to extrapleural dissection along the diaphragm, may occur. Pneumopericardium is rare and may be confused with pneumomediastinum. Pneumopericardium may be identified as air within the pericardium moves freely away from the dependent part of the sac, while mediastinal air shows little movement when the position of the patient is changed (Fig. 19).

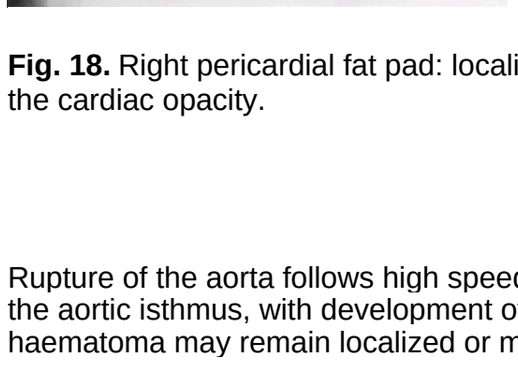


Fig. 19. Pneumopericardium in a young man following a diving accident: erect posteroanterior projection. Gas is seen within the pericardium (white arrowhead).

Trachea and main bronchi

The air column of the trachea, major bronchi, and bronchus intermedius should be visible on the frontal radiograph. The trachea extends from the cricoid to the carina and is approximately 11 cm long. It is a midline structure but deviates slightly to the right as it passes the aortic knuckle. Tracheal deviation increases with age as the aorta enlarges. The coronal diameter of the trachea is 1.5 to 2 cm and its sagittal diameter is 2 cm. A coronal diameter greater than 2.5 cm or less than 1.3 cm in men and 1.0 cm in women is abnormal. The cross-sectional shape of the normal trachea is variable but it is most commonly round or oval. The trachea is bordered on its right lateral wall by pleura: this contains a variable quantity of fat forming the right paratracheal stripe. This stripe is visualized on the majority of radiographs and has a

maximum normal width of 4 mm. Loss of the outer margin of the right paratracheal stripe suggests lymphadenopathy. The left paratracheal stripe is seldom seen.

Displacement of the trachea is an important radiological sign. However, in the young, the trachea and mediastinum are very mobile and often show a marked deviation from the midline, particularly on expiration. The lumen of the trachea should also be noted as the trachea may be narrowed, or opacified by a tumour or foreign body. The trachea divides into the two main major bronchi at the carina, which lies at the level of the fourth to fifth thoracic vertebra. The subcarinal angle shows a wide variation in normal individuals but on average measures 70°. The course of the right main bronchus is more vertical than is that of the left main bronchus in adults, which accounts for aspiration occurring more commonly on the right. In children, there is bronchial symmetry and the incidence of right- and left-sided aspiration of a foreign body is therefore similar. The transverse diameter of the right main bronchus is greater than that of the left. The left main bronchus is longer than the right. The right upper lobe bronchus arises more proximally than the left upper lobe bronchus and divides into three branches, the anterior, posterior, and apical segments. The intermediate bronchus continues distally for 3 to 4 cm, from the take-off of the right upper lobe bronchus and then bifurcates into the middle and lower lobe bronchus. The right middle lobe bronchus bifurcates into the medial and lateral segments.

The left upper lobe bronchus either bifurcates or trifurcates: the former is more usual. The superior branch divides into the apical posterior and anterior segments, while the inferior branch is the lingula bronchus, which is roughly analogous to the right middle lobe bronchus and divides into the superior and inferior segments. The left lower lobe bronchus divides in similar fashion to the right lower lobe bronchus. There is one exception: the anterior and medial portions of the lobe are supplied by a single anteromedial bronchus (Fig. 20).

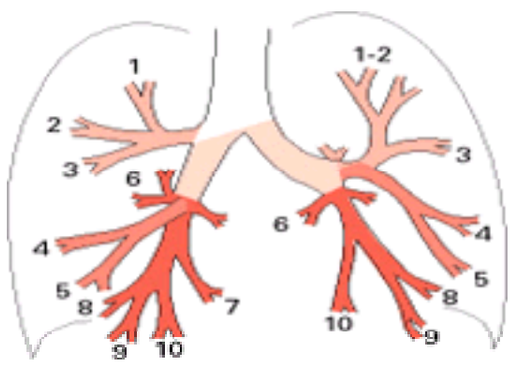


Fig. 20. Diagram illustrating the anatomy of the main bronchi and segmental divisions.

Pleura

The pleura, consisting of the visceral and parietal layers, is not normally visible over the lung apices, mediastinum, or diaphragm. Slight pleural thickening (1 to 2 mm) over the convexity of the lungs can be visualized because of the greater density of the contiguous ribs. Even moderate thickening over the mediastinal and diaphragmatic surfaces cannot be recognized as these adjacent structures have a similar radiographic density. Focal pleural thickening, as is seen in mesothelioma, can be identified by the alteration of the smooth contour of these surfaces.

Interlobar fissures

The interlobar fissures comprising two layers of visceral pleura form the contact surfaces between the pulmonary lobes. They are of hairline thickness and are visible because of the presence of air-containing lung on either side of the fissure. These fissures are only visible on the radiograph when the X-ray beam passes tangentially along their surface. All fissures have an undulating course and they are therefore never visualized in their entirety. The identification of the fissures is important in the assessment of lobar disease. Their depth varies from complete separation of the lobes to a superficial slit. The completeness of this separation determines the ease of spread of disease between the lobes and the relative ease of lung resection. The fissures should be assessed for position, configuration, and thickness.

Oblique (major) fissure

The oblique fissure separates the upper (and middle) lobes from the lower lobes. It begins at the level of the fifth thoracic vertebra and extends obliquely downward and forward running parallel with the sixth rib and ending at the diaphragm a few centimetres behind the sternum. The left oblique fissure has a more vertical course and lies posterior to the right fissure. The fissure is propeller shaped, with the upper half facing forwards and outwards and the lower half facing forward and slightly medially. A triangular opacity at the lower end of the fissure, with its base contiguous with the diaphragm and its apex tapering into the fissure, is due to fat.

Horizontal (minor) fissure

This fissure separates the anterior segment of the right upper lobe from the middle lobe and lies roughly horizontal at the level of the fourth rib anteriorly. The contour is variable and minimal displacement should not be regarded as significant. The fissure rarely reaches the mediastinum. Its medial extent is to the lateral margin of the interlobar pulmonary artery. A fissure line that projects medial to this point represents the downward projection of the oblique fissure, providing evidence of volume loss in the right lower lobe. On the frontal radiograph the horizontal fissure is seen in its entirety in 6 per cent of individuals and is visualized in part in 66 per cent.

A left minor fissure, separating the anterior segment of the upper lobe from the lingula, is an uncommon finding. On the lateral radiograph the left minor fissure is superior to the right minor fissure (Fig. 14).

Accessory fissures

Accessory fissures between segments are common and their recognition is important to avoid misinterpretation of the radiograph and to allow identification of the segment they subtend.

Azygos fissure

An azygos lobe is an uncommon anomaly, occurring in 0.4 to 1 per cent of individuals (Fig. 21). It is created by the downward invagination of the azygos vein through the apical portion of the right upper lobe. The azygos fissure is composed of a double layer of visceral and parietal pleura. The space between the layers of pleura is in communication with the general pleura cavity and may therefore contain air or fluid if a pneumothorax or pleural effusion is present. The azygos lobe should not be misinterpreted as right upper lobe collapse.



Fig. 21. Azygos fissure: the fissure is visualized as a curvilinear shadow extending obliquely across the upper part of the right lung and terminates in the teardrop shadow of the azygos vein above the right hilum.

Inferior accessory fissure

This fissure separates the medial basal segment from the remainder of the lower lobe, forming a retrocardiac lobe. This fissure is visualized in 30 per cent of chest radiographs (Fig. 22).

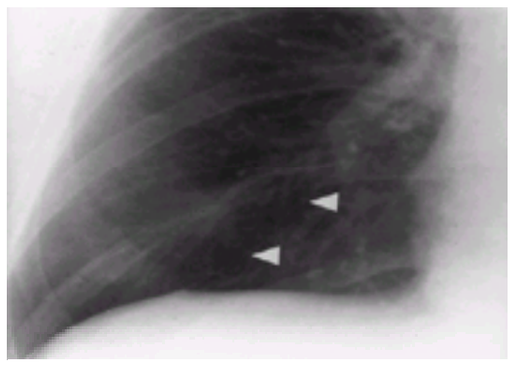


Fig. 22. Inferior accessory fissure. Localized view of the right lung base. The fissure separates the medial and anterior basal segments of the right lower lobe (arrowheads).

The superior accessory fissure

This fissure separates the superior segment from the basal segments of the lower lobe and is more commonly visualized on the right.

The lung fields and pulmonary hilum

The lungs should be reviewed individually and in comparison, matching zone by zone from apex to base. Before assessing the lung density the patient's position should be noted. Rotation of the patient is identified by observing the distance between the medial end of the clavicle and the adjacent margin of the vertebral body. It should be symmetrical in a correctly positioned patient. Rotation will alter the density between the lungs, the side closer to the film appearing blacker. Any discrepancy in density between the lungs in a patient who is not rotated must be regarded as pathological. Serial radiographs may show a variation in lung density, in the same subject depending upon the depth of respiration: the greater the degree of lung inflation, the more translucent the lungs appear.

The normal linear markings are composed of the pulmonary blood vessels and the fissures. An alteration in vasculature may be local or generalized, and should be interpreted in conjunction with an assessment of the hilar pulmonary vessels and the degree of lung inflation. Abnormal linear shadows may be identified from their compartment of origin, namely, the lung parenchyma, the interstitium, the bronchovascular bundles, or the pleura. Although it may be difficult to distinguish between parenchymal and interstitial origin, this method of evaluation is a useful aid to interpretation.

The silhouette sign is useful for identifying and localizing pulmonary disease. The cardiac and diaphragmatic outline is entirely dependent on adjacent air-filled lung for its visualization. If the lung opacifies the outline will be obscured. Any intrathoracic abnormality that obliterates the contour of the heart or diaphragm is in anatomical continuity with these surfaces. Loss of the right heart border implies middle lobe disease or, on the left, disease of the lingula. A collapsed and opaque lower lobe, lying posteriorly, will not efface the cardiac border but will obscure the adjacent diaphragmatic contour.

The pulmonary vessels should be examined from the periphery towards the hilum. This method prevents the more prominent hilar vessels from diverting one's attention from the less conspicuous vessels of the periphery. The vessels in comparable zones of the lung should be of similar size and number. The vessels should be clearly defined: marginal haziness indicates an abnormality of the perivascular connective tissue or lung parenchyma. The vessels in the upper zone of the lungs are smaller than are those of the lower zones at equivalent levels, because of preferential perfusion of the lung bases due to gravity. In the outer periphery (1 to 2 cm) of the lung, vessels are no longer visible. The pulmonary arterial system is invariably related to the bronchial tree; they branch together as they run in the bronchovascular bundles. The pulmonary veins lie within the interlobular septa and are thus separate from the bronchoarterial pathways. These vessels terminate medially into two superior and two inferior pulmonary veins. In the upper zones the veins lie lateral to the arteries, while in the lower zones they have a more horizontal course. The apical, segmental upper lobe bronchus may be visualized end-on, at the hilum. Comparison with an adjacent opaque artery is a useful assessment of the calibre of the pulmonary circulation: normally the artery is of a similar calibre to the bronchus. Redistribution of blood flow resulting in the upper lobe vessels being larger than the lower lobe vessels is a well recognized sign of raised pulmonary venous pressure.

Pulmonary hilum

Evaluation of the hilum involves an assessment of its size, position, and density. A pathological process involving the hilum causes the hilum to appear too large, too small, dense, or in an abnormal position. The hilar shadows are formed by the pulmonary arteries and veins: bronchi and normal lymph nodes do not contribute to the hilar density. The upper portion of the hilar shadow comprises the upper lobe pulmonary veins, as well as the arteries. The lower portion of the hilum consists of the basal pulmonary arteries. The inferior pulmonary veins do not contribute the hilar shadow but are visualized end-on, creating a nodular opacity posteroinferiorly to the hilum. This pseudotumour appearance can be distinguished from a mass lesion by identifying vessels converging towards the opacity, the venous confluence.

The right pulmonary artery is an intramediastinal structure, dividing at the pericardium into the ascending and descending branches. The right main pulmonary artery does not form part of the hilar shadow. The left pulmonary artery arches above and behind the left main bronchus. The left hilum is therefore at a higher level than the right (by 0.5 to 1.5 cm): in 3 per cent of individuals the hilum may be at the same level but the right hilum is never higher than the left in normal subjects. The position of the hilum can be assessed from the hilar point. This is the level at which the superior pulmonary vein meets the shadow of the basal artery. On the right the hilar point is opposite the horizontal fissure (Fig. 23).

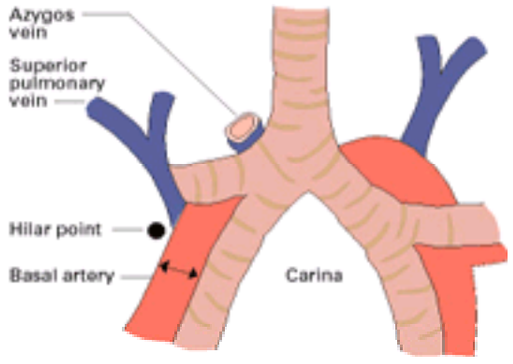


Fig. 23. Diagram to show the hilar point: this is the level at which the superior pulmonary vein (V) meets the shadow of the basal artery (a). The maximal width of this artery should not exceed 16 mm in men and 15 mm in women (Az = azygos vein).

Hilar size is initially assessed by comparison of one hilum with that of the opposite side. The size range is quite variable between individuals, but in normal individuals they should be of similar size. A reliable assessment of hilar size can be made from measurement of the diameter of the right lower lobe pulmonary artery. This vessel, when measured at its widest point, should not exceed 16 mm in men and 15 mm in women. When wider than this pulmonary hypertension should be considered. The left lower lobe artery is normally poorly visualized, and assessment of its size is therefore difficult. The diameter of this vessel is smaller than the right by 1 to 2 mm, as

blood flow is greater to the larger right lung. The inferior portion of the right hilum may have a rounded configuration rather than the more usual tubular appearance. This is due to the visualization, end-on, of a posteriorly-directed right pulmonary artery, and is a normal anatomical variant. It may simulate a hilar mass.

Many congenital and acquired conditions cause abnormality in hilar size. The most common cause for a small hilar shadow is lobar collapse, particularly of the lower lobe. The hilum appears smaller as the lower lobe artery cannot be visualized when surrounded by airless lung: only pulmonary vessels supplied by the aerated lung are radiographically visible. In Macleod's syndrome, a small hilum is associated with peripheral hypovascularity and increased transradiancy of the whole or part of one lung (Fig. 24). It has been attributed to an obliterative bronchiolitis occurring in early childhood, and it is usually discovered as a chance radiographic finding in an asymptomatic individual.



Fig. 24. Macleod's syndrome: unilateral hypolucent lung. The right lung is of normal volume but the right hilar shadow is diminutive and the lung is oligaemic.

A prominent hilum may be due to enlargement of the pulmonary artery or to a hilar mass. When the pulmonary vessels seem to arise directly from the hilar shadow, the prominent hilum is due to an enlarged pulmonary artery. However, if the pulmonary arteries arise medial to the lateral aspect of the hilar shadow then the enlargement is caused by an extravascular mass. The ‘hilum overlay' sign allows differentiation of a mediastinal mass from a prominent cardiac silhouette. The first branching of the pulmonary artery arises lateral to the cardiac shadow or just within its outer aspect in the majority of normal individuals. In the presence of an anterior mediastinal mass, the hilum is projected medial to the lateral border of the mass and the first bifurcation of the pulmonary artery will lie more than 1 cm medial to the edge of this shadow. In cardiomegaly the hilum is displaced laterally.

The density and contour of the hilar shadows should be compared. Normally, the lateral aspect of the hilum has a concave contour: this is more clearly visualized on the right. Hilar lymphadenopathy causes enlargement of the hilum and the development of a lobulated contour.

In lobar collapse, there is alteration of the hilar arterial and bronchial anatomy. With upper lobe collapse the vessels disappear, but the hilum appears to be elevated because the interlobar and lower lobe pulmonary artery remains visible. The ipsilateral main bronchus has a horizontal orientation and there is outward displacement of the bronchus intermedius on the right and the lower lobe bronchus on the left. In lower lobe collapse the hilum appears small rather than depressed. If the lower lobe pulmonary artery is visible, an opacity in the lower lobe cannot be due to collapse. The ipsilateral bronchial tree is swung medially and has a vertical orientation. The opacity of collapsed lung appears as a uniform whiteness, except in left upper lobe collapse. On the frontal radiograph, the collapsed left upper lobe appears as an area of opacification at the hilar region which fades to a more normal lucency lateral to the hilar region. As a lobe collapses, compensatory over-inflation of the adjacent lung occurs. The over-inflated lung appears blacker than usual with a decrease of vascular markings throughout the lung field. In upper lobe collapse, an over-inflated superior segment of the lower lobe may insinuate between the mediastinum and the medial aspect of the collapsed lobe to produce paramediastinal lucency. This is more commonly seen on the left.

The lateral radiograph

The lateral view is often of value in the assessment of thoracic trauma, particularly to visualize the thoracic vertebra, to define loculated collections of blood or fluid, or to locate penetrating foreign bodies. The chief landmarks on the lateral radiograph are: the bone and soft tissue shadows; the diaphragm; the mediastinum shadows; and the lung fields.

Bone and soft tissue

A fractured sternum can only be clearly defined on the lateral view. This may also disclose an associated haematoma in the anterior mediastinum that will be visualized on the posteroanterior view as mediastinal widening. The anteroposterior view may be interpreted as suggesting aortic injury. In children the manubrium and the four segments of the body of the sternum can be seen separately.

The thoracic spine is well visualized on the lateral radiograph. The alignment and vertebral body height should be noted. The spine is rigid and supported by ribs, consequently traumatic forces must be considerable to cause fractures and dislocation. Dislocations rarely present without fractures. The density of the thoracic spine on the lateral view normally becomes increasingly translucent from above downwards until the diaphragm is reached. Pleural or pulmonary pathology will alter this progressive increase in translucency and the whole length of the spine may appear equally radio-opaque. This observation may be the only evidence of a posteriorly positioned lesion which is hidden behind the heart on the frontal radiograph. A haemothorax may make it difficult to visualize the thoracic spine.

On the true lateral radiograph, a vertical opacity, the retrosternal stripe, may be visualized. It measures up to 3 mm and is composed of mediastinal fat. The parasternal stripe represents lung abutting on the anterior chest wall to the right or left, and has a wavy contour, due to the indentation of the surface of the lungs by costal cartilages. The contour may be lobulated due to anterior rib fractures.

The head of the humerus and glenoid may overlie the posterior aspect of the chest. The vertical border of each scapula is seen running downwards from the inferior margin of the glenoid which is directly behind the tracheal translucency.

The diaphragm

The domes of the diaphragm pass in a superior convex curve from the sternum backwards to the vertebra. The highest level is usually at the midaxillary line and the lowest is the posterior costophrenic recess. Localized pulmonary or pleural disease adjacent to the posterior aspect of the diaphragm, may not be recognized on the frontal radiograph but will be seen on the lateral view. Localization of the disease, therefore, requires identification of each hemidiaphragm on the lateral radiograph. The distinction is made by the following findings: the right dome is usually higher than the left and is visualized passing anteriorly through the heart; the left dome is obscured anteriorly by the heart and has the gastric air bubble lying beneath it; the inferior vena cava causes effacement of the right hemidiaphragm; visualization of the oblique fissures allows identification of the ipsilateral hemidiaphragm; the hemidiaphragm nearest the film is related to the least magnified posterior ribs.

Mediastinum

The trachea is seen as a tubular air-containing structure extending from the neck to the level of its bifurcation at the sixth thoracic vertebra. The carina is rarely identified on the lateral radiograph. The position of the trachea should be noted and the posterior tracheal stripe representing the posterior tracheal wall and surrounding soft tissue should measure between 2 and 3 mm. The tracheo-oesophageal stripe formed by the posterior tracheal wall, the anterior oesophageal wall (identified when air is contained within the oesophagus), and surrounding mediastinal soft tissue should not measure more than 5 mm. The end-on orifices of the right and left upper lobe bronchi can be identified on the majority of radiographs: each bronchus lies in the plane of the trachea. The demonstration of an abnormal position of the upper lobe bronchi, either anterior or posterior to the plane of trachea, may be an important finding. In upper lobe collapse the bronchi may be displaced forward, while in lower lobe collapse the bronchi are displaced posteriorly. Posterior displacement of the left upper lobe bronchus may be the only sign of left atrial enlargement. The right upper lobe bronchus projects above the left bronchus. The left upper lobe bronchus is readily visualized as it is surrounded by vascular structures. The right upper lobe bronchus is rarely as well defined as it is surrounded primarily by lung. If the bronchial lumen is clearly visualized on the right, the airway is likely to be surrounded by soft tissue due to lymphadenopathy. The posterior wall of the right main and bronchus intermedius is visualized extending in continuity with the posterior tracheal stripe. Anterior bronchial walls are rarely well seen. The posterior wall of the left main bronchus is seen in less than 50 per cent of patients. The walls of the central bronchi should be well defined and measure less than 3 mm in thickness. Localized thickening or lobulation suggests abnormality. The right middle lobe

bronchus is rarely seen on the lateral radiograph.

The left pulmonary artery is visualized on a plane posterior to the trachea and carina, while the right pulmonary artery lies anterior. The right superior pulmonary vein lies anterior to the right middle lobe bronchus as it enters the left atrium. The left superior pulmonary vein lies anterior to the left upper lobe bronchus. The inferior pulmonary veins lie below and behind the superior vein and end-on visualization of these veins may simulate a mass.

The oesophagus often contains considerable quantities of air, even in normal subjects, and may therefore be visualized on the radiograph.

Lung fields

Anteriorly the lungs meet the anterior chest wall above the heart forming a translucent retrosternal space. This area may be opacified by a haematoma, a thymic mass, a goitre, or enlarged lymph nodes, particularly in patients with lymphoma.

Computed tomography of the chest

CT is used routinely in the assessment of benign and malignant lesions throughout the chest wall, lung fields, or mediastinum. CT may also be used in patients with thoracic trauma, to define or confirm an abnormality detected or suspected on the standard chest radiograph. Cross-sectional imaging overcomes the difficulties of interpretation due to superimposition of structures on the chest film. CT is particularly sensitive in detecting pleural abnormalities, air, and fluid collections, which may pass unnoticed on the supine radiograph. It allows discrimination of blood from other pleural fluids due to the differences in density and is valuable in the assessment of mediastinal widening. Patients with high velocity deceleration injury and suspected trauma to the aorta or brachiocephalic arteries must always be studied by angiography to locate precisely the site of injury. However, in stable patients with mediastinal widening in whom there is a low index of suspicion of aortic injury, CT with intravenous contrast is advised. The mediastinal widening may be elucidated as being due to a fractured sternum, unfolded thoracic aorta, congenital vascular anomaly, a paramediastinal pleural collection, or a paraspinal haematoma related to a vertebral fracture. Pulmonary injury, including laceration, haematoma, and contusion is clearly defined. CT contrast studies allow differentiation of contused lung from adjacent collapsed lung or haematoma. Pulmonary abnormalities noted on CT are usually more extensive than those recognized on the plain chest radiograph. Pleural lesions, haematoma, and fractures may be demonstrated by CT but are usually apparent on clinical examination and on the plane radiograph.

CT may be helpful in the evaluation of myocardial injury. The ability to demonstrate tracheobronchial or oesophageal disruption is uncertain. Contrast examination of the oesophagus remains the investigation of choice in suspected oesophageal rupture. In patients with severe lower thoracic injury, CT scanning of the liver and spleen following intravenous contrast should be performed.

Normal anatomy

Detailed discussion of the cross-sectional anatomy of the chest is beyond the scope of this Section. The trachea and major bronchi are well visualized. The major bronchi are identified by their size and orientation. For example the right upper lobe bronchus has a horizontal course and the bronchus intermedius passes in a vertical direction. Obliquely-orientated bronchi are more difficult to evaluate. In the hilum, there is considerable variability of the vascular components. The bronchi are the most consistent structures and the veins the least consistent, in terms of their anatomical relationships. Inspection of the hilum therefore begins with the evaluation of the main lobar and segmental bronchi. The pulmonary arteries can then be distinguished from the veins by their relationship to each other and the bronchi. In the periphery it is not possible to distinguish the arteries from the veins.

The normal hilar lobulation is due to vascular structures. In particular, the superior pulmonary veins on the right are prominent components of the hilar contour and should not be mistaken for a hilar mass. The pleural fissures are rarely visualized but may be inferred by an area of avascularity. In the supine position, there is an paucity of vessels anteriorly compared to the posterior zones. This appearance is due to the effect of gravity on blood flow and may be reversed by turning the patient. The subpleural regions are relatively avascular: normal pleura are not visualized.

Vascular anomalies

Anomalies of the intrathoracic great vessels rarely result in symptoms but can cause confusing mediastinal contours on the chest radiograph. Contrast CT scanning, MRI, or angiography are used to confirm the diagnosis of a vascular malformation. An aberrant right subclavian artery occurs in one in 200 people. The artery arises from the distal part of the aortic arch as its last branch, and crosses the mediastinum obliquely upward from left to right behind the oesophagus (Fig. 25). On the lateral radiograph the vessel causes a retrotracheal opacity which may be mistaken for a mass (Fig. 26). The vessel may obscure the aortic arch and cause anterior tracheal displacement. In some patients, the artery arises from an aortic diverticulum: if large, this may be misdiagnosed as an aortic aneurysm. An aberrant left subclavian artery arises from a right-sided aortic arch in approximately one in 1000 people. About 10 per cent of these individuals have associated congenital heart disease.

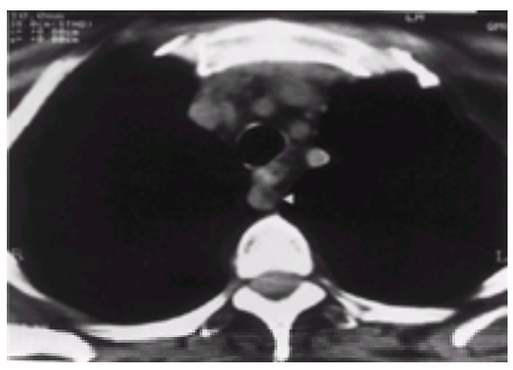


Fig. 25. CT scan through the chest at the level of the great vessels. The aberrant right subclavian artery arises from the distal part of the aortic arch as its last branch lying posterior to the oesophagus (arrowhead).

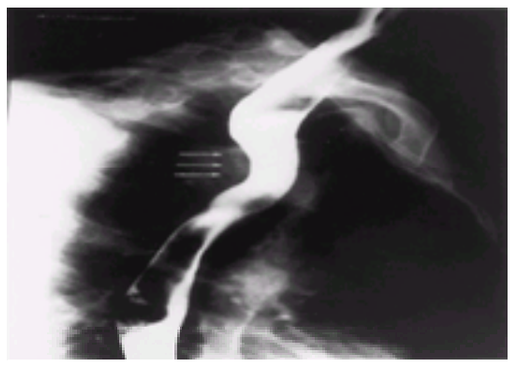


Fig. 26. Barium swallow demonstrates the posterior indentation on the oesophagus by the aberrant subclavian artery as it crosses the mediastinum obliquely upward from left to right (arrows).

The most common venous anomaly is the lateral displacement of the azygos arch associated with an azygos lobe. The azygos lobe represents the medial portion of the right upper lobe trapped by the azygos fissure and vein. CT shows the azygos arch to be abnormally high, with the azygos vein joining the superior vena cava at the level of the brachiocephalic veins. Above the carina the posterior portion of the azygos vein may simulate a pulmonary nodule on CT. The lung often protrudes into the mediastinum behind the superior vena cava to outline its medial and posterior walls. The lung may also extend into the retrotracheal space, outlining the posterior wall of the trachea and displacing the oesophagus to the left. A left-sided superior vena cava is a rare anomaly, occurring in 3 per cent of normal individuals and 5 per

cent of patients with congenital heart disease. Most have a right superior vena cava and the left brachiocephalic vein may be small or absent. The left superior vena cava lies lateral to the left common carotid artery and anterior to the left subclavian artery. It descends lateral to the aortic arch and main pulmonary artery and anterior to the left hilum to enter the coronary sinus posterior to the left atrium and ventricle. A left superior vena cava is readily identified on CT (Fig. 27). Azygos or hemiazygos continuation of the inferior vena cava have identical CT appearances. In both, the azygos arch, azygos vein, and superior vena cava are dilated due to the increased blood flow. The enlarged azygos system may be mistaken for a mediastinal mass or for posterior mediastinal lymphadenopathy.

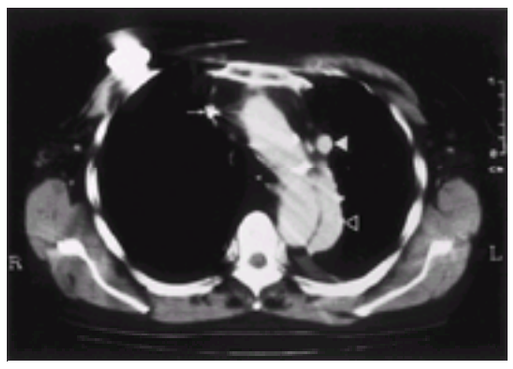


Fig. 27. Contrast enhanced CT at the level of the aortic arch shows a left superior vena cava (white arrowhead) and a right superior vena cava with a pacemaker wire *in situ* (arrow). The dilated hemiazygos vein (white lined arrowhead) is due to hemiazygos continuation of the inferior vena cava.

In hemiazygos continuation of the inferior vena cava, the dilated hemiazygos vein may not communicate with the azygos system and will course up the left side of the mediastinum to produce a retroaortic mass. The dilated azygos vein is often of equal diameter to the aorta. The suprahepatic portion of the vena cava develops normally, and an intrathoracic inferior vena cava is therefore usually present with azygos continuation (Fig. 28). Visualization of an inferior vena cava on the chest radiograph does not exclude the diagnosis of azygos or hemiazygos continuation of the inferior vena cava. Associated anomalies of the abdominal inferior vena cava are common and polysplenia and congenital heart disease may be present (Fig. 29).

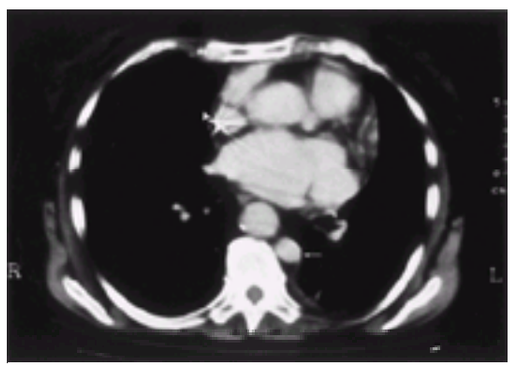


Fig. 28. Contrast enhanced CT scan of the patient with hemiazygos continuation of the inferior vena cava shows a left intrathoracic inferior vena cava (arrow). The pacemaker wire is seen in the right superior vena cava (arrowhead).

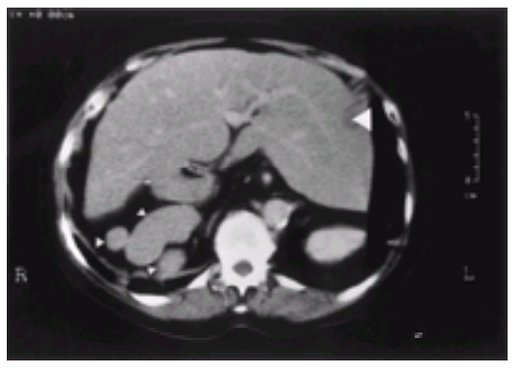


Fig. 29. Contrast enhanced CT scan through the upper abdomen in patient with the inferior vena cava anomalies. Central liver (large arrow head), right polysplenia (small arrowheads), and right-sided gastric fundus (arrow).

Further reading

Felson B. *Fundamentals of chest roentgenology*. WB Saunders, Philadelphia, 1973.

Fraser RG, Paré JAP. *Diagnosis of disease of the chest*, Vol. 1. WB Saunders, Philadelphia, 1988.

Grainger RG, Allison DJ. *Diagnostic radiology. An Anglo-American textbook*, Vol. 1. Churchill Livingstone, Edinburgh, 1992.

Lee Joseph KT, Sagel S, Stanley RJ. *Computed body tomography with MRI correlation*, 2nd edn. Raven Press, New York.

Sutton D. *A textbook of radiology and imaging*. Churchill Livingstone, Edinburgh, 1992.

41.2 The trachea

Douglas Mathisen and Hermes C. Grillo

Introduction
Surgical anatomy of the trachea
Diagnostic techniques
Anesthesia
Airway management
Preoperative evaluation
Release maneuvers
Cervical trachea
Intrathoracic trachea and carina
Midtrachea
Tracheostomy
Tracheal strictures
Congenital lesions
Trauma
External trauma
Burns
Postintubation stenosis
Clinical presentation
Diagnosis
Treatment
Tracheo-innominate artery fistula
Tracheoesophageal fistula
Persistent tracheal stoma
Postinfectious stenosis
Tracheal collapse
Miscellaneous obstructive lesions intrinsic to the airways
Idiopathic airway stenosis
Miscellaneous lesions
Tracheal tumours
Occurrence and clinical presentation
Diagnosis
Pathological features
Treatment
Surgical resection of tracheal tumours
Complications of tracheal reconstruction
Further reading

Introduction

A high index of suspicion is required to diagnose tracheal disorders. Dyspnea on exertion, wheezing, and a normal chest radiograph may lead to patients being treated for many months or years for presumed adult-onset asthma before the diagnosis of an organic problem is made. Bronchoscopy is essential to the diagnosis and management of tracheal disorders, but may provoke life-threatening airway obstruction if it occludes an already narrow trachea. Physicians treating patients with airway disorders must be aware of the indications and limitations of flexible and rigid bronchoscopy to avoid potential dangers. Successful tracheal surgery requires selecting suitable patients, appropriate timing, determining the extent of operation required, and knowing when additional procedures are needed to relieve excessive anastomotic tension. Postoperative care can be demanding because of the inability to raise secretions and the development of postoperative edema in compromised airways.

Surgical anatomy of the trachea

The trachea begins at the lower border of the cricoid cartilage, at which point the uppermost tracheal cartilage is partly inset beneath the cricoid. It terminates at the point at which the lateral walls of the right and left main bronchi branch from the lower trachea. The carinal spur is useful as a more precise landmark for the termination of the trachea, since it is clearly definable bronchoscopically and radiologically. The average adult human trachea is 11 cm long: length varies, roughly in proportion to the height of the patient. There are approximately two cartilaginous rings per centimetre of trachea, which contains from 18 to 22 rings. The subglottic laryngeal airway measures 1.5 to 2.0 cm before the trachea is reached. Except for some cases of congenital stenosis with circumferential O rings of the trachea, the only completely circular cartilage in the upper airway is the cricoid, which has a broad posterior plate.

The potential for displaying the trachea in the neck is of major importance not only for surgical access but also for ease of reconstruction after resection of any length of the trachea. In young people, particularly when there is no obesity, hyperextension of the neck delivers more than 50 per cent of the trachea into the neck. In the kyphotic, aged person, particularly if obesity is present, the cricoid cartilage may be at the level of the sternal notch, and even the most vigorous hyperextension may fail to deliver any of the trachea into the neck. The anatomical position of the trachea changes from essentially subcutaneous at the level of the cricoid cartilage to prevertebral at the carina. The course is thus obliquely caudal and dorsal when an individual is standing in an erect position. In the kyphotic, aged patient, lateral projection becomes increasingly horizontal. The slight extensibility and flexibility of the trachea in youth diminishes with increasing age. Calcification of the cartilage also occurs with age and with injury.

The blood supply is of special importance in resection and reconstruction of the trachea. The upper part is supplied principally by branches of the inferior thyroid artery. The lower portion is supplied by branches of the bronchial artery, with contributions from the subclavian, supreme intercostal, internal thoracic, and innominate arteries. These vessels send branches anteriorly to the trachea and posteriorly to the esophagus, arriving at the trachea through lateral pedicles of tissue. The longitudinal anastomoses between these vessels are very fine. Transverse intercartilaginous arteries branch ultimately into a submucosal capillary network. Excessive division of the lateral tissues by circumferential dissection of the trachea can easily destroy this blood supply, with serious and sometimes disastrous effects.

The relation of the recurrent laryngeal nerves to the trachea and the esophagus, and the point of entry of these nerves into the larynx, need not be repeated here (see Chapter 20.3.2). The relation of the trachea to the thyroid gland is also well known. The isthmus crosses the trachea at the second and third cartilaginous rings. The medial portions of both lobes of the thyroid gland adhere closely to the trachea at this same level laterally: it may be necessary to remove a lobe, or sometimes the entire thyroid gland, during removal of a tumor in the upper portion of the trachea. Posteriorly, the esophagus has a common interface through areolar tissue with the membranous tracheal wall. The blood supply of the esophagus and membranous tracheal wall are closely linked. Anteriorly, the innominate artery courses obliquely across the anterior surface of the trachea. Below this level the aorta arches backward across the left tracheobronchial angle. Here the left recurrent laryngeal nerve arrives at its place in the tracheoesophageal groove.

The lymph nodes adjacent to the trachea are stations in the pathways from the lungs and mediastinum. The lymphatics of the trachea have been less well studied. Gross observations of the clinical behavior of tumors metastatic from the trachea have been made. Metastases appear to involve the most closely adjacent groups of tracheal lymph nodes. Metastases to the nodes on the side contralateral to the primary lesion or to the carina from lower tracheal tumors are common. More remote metastases to scalene nodes or to other cervical nodes are not often seen. For primary tracheal tumors, paratracheal or subcarinal nodes may be considered as N1 stations.

Diagnostic techniques

The primary diagnostic techniques for tracheal abnormalities are radiological imaging and bronchoscopy. A plain chest radiograph is often considered to be normal, but closer inspection will reveal an abnormality of the tracheal air column. Relatively simple radiological techniques without the use of contrast media delineate tracheal disease. The location of the lesion, its linear extent, extratracheal involvement, and the amount of normal airway can be determined. In addition to standard views of the chest taken in various projections, centred high enough to show tracheal detail, anteroposterior filtered tracheal views of the entire airway from the larynx to the carina are required. Lateral neck views using soft-tissue techniques with the patient swallowing and the neck hyperextended to bring the trachea up above the clavicles are useful to define abnormalities in the upper trachea. Fluoroscopy not only demonstrates functional asymmetry of the vocal cords, if present, but also may give additional information about the extent of the lesion and collapse of the airway if malacia is present. Spot films are usually all that are required. In some cases, polytomography (anterior and lateral views) gives additional detail. Barium esophagography defines extrinsic compression or invasion of the esophagus. Computed tomography (**CT**) may detect extratracheal disease. Magnetic resonance imaging (**MRI**) adds little. Its ability to produce sagittal and coronal views, however, has been helpful in certain

cases and may give more accurate detail than the standard radiographic techniques.

Bronchoscopy is invaluable in assessing airway pathology; it should be done by the surgeon, rather than relying on the findings of others. Flexible and rigid bronchoscopy both have a role in the management of airway problems. However, flexible bronchoscopy may precipitate airway obstruction in patients with critical stenosis of the bronchus (< 4 mm) by increasing secretions, hemorrhage, edema, or by just temporarily occluding the airway. Emergency dilatation of the bronchus is required or death may ensue. It is best not to examine an airway stenosis with the flexible bronchoscope at all. Proper evaluation should be made in the operating room, with facilities available to dilate the stenosis if necessary. Endoscopy is often best done as part of the planned surgical procedure.

Bronchoscopy is invaluable in determining the extent of airway disease and, just as importantly, the amount of uninvolved airway. Measurements should be taken and recorded. The quality of the mucosa should be assessed: marked inflammation and erythema may dictate delaying definitive surgical correction until they have subsided. Tumors should be biopsied, if not excessively vascular, and the extent of disease should be determined. As a general rule, the rigid bronchoscope is much more valuable than the flexible in assessing airway disease.

Anesthesia

Most reconstructive procedures on the trachea are done under spontaneous ventilation, although assistance is given during their intrathoracic portions if needed. Prolonged paralysis of respiration is to be avoided, since it is desirable that the patient should resume spontaneous respiration postoperatively without the need for ventilatory support. After the trachea has been divided, in either the anterior or the intrathoracic approach, a sterile, flexible endotracheal tube (Tovell) with attached connecting tubing is introduced into the distal trachea for maintenance of ventilation. If carinal resection is required, ventilation is usually maintained through the opposite (usually the left) lung after resection of the carina. Such tubes are not cumbersome. The alternative of high-frequency ventilation through a catheter also works well and is used in special cases. Cardiopulmonary bypass is not needed for simple tracheal cases, and it creates hazards in complicated cases since extensive manipulation of the lung in a heparin-treated patient can lead to intrapulmonary hemorrhage.

Airway management

The ability to control the airway is crucial to the management of all problems of the trachea. Tracheal tumors and postintubation stenosis may present as emergency airway problems. Endotracheal intubation may be impossible, and even dangerous, leading to complete airway obstruction, especially in patients with high tracheal lesions. Elevation of the head, administration of cool mist and oxygen, and careful sedation may allow semielective control of the airway. This goal is best accomplished in the operating room, where an assortment of rigid bronchoscopes, dilators, biopsy forceps, and even instruments for emergency tracheostomy are available. Anesthesia, as in elective tracheal operations, is best accomplished by inhalation techniques. Patience is required to allow the patient to become anesthetized adequately: induction of anesthesia deep enough to allow rigid bronchoscopy may take as long as 20 min. Paralyzing agents should not be used if airway obstruction in an apneic patient is to be avoided.

The initial evaluation of neoplasms should be made with a rigid bronchoscope inserted through the vocal cords and stopped just proximal to the level of obstruction. Rigid telescopes can be used to assess the obstruction: these can be passed beyond most tumors, even those causing nearly total obstruction. Once the distal airway has been assessed, the tumor can be removed incompletely with biopsy forceps to determine its consistency and vascularity. The tip of the rigid bronchoscope can usually be used to ‘core out' most of the tumor, which can then be grasped with biopsy forceps and removed. The bronchoscope may be passed into the distal airway for ventilation and to tamponade any bleeding. Direct application of epinephrine (adrenaline)-soaked pledgets helps to control persistent oozing. Direct cautery (with insulated electrodes) is rarely required. Although the laser has become popular in the management of malignant strictures, it is time-consuming, costly, and rarely more advantageous than other techniques.

Postintubation stenosis poses a slightly different problem in airway control. It may be impossible to pass a large, rigid bronchoscope beyond a tough, inflammatory stricture: to do so may also result in tracheal rupture, or in total airway obstruction secondary to bleeding or edema. Jackson dilators passed through a rigid bronchoscope under direct vision can be used to dilate postintubation stenosis initially. Pediatric Jackson bronchoscopes are next inserted serially in order to dilate the stenosis adequately. Administration of racemic epinephrine and steroids in the first 24 to 48 h minimizes postdilatation edema.

Dilatation or endotracheal removal of malignant and inflammatory strictures, mechanically or by laser, is only a temporary measure: inflammatory strictures usually recur within days to weeks. However, these techniques allow time for more thorough evaluation of the patient and for elective surgery. Many patients are taking high doses of steroids at the time of presentation, having been treated for refractory asthma. Establishing an airway allows the steroid dosage to be tapered and discontinued, and surgery may then be performed without the threat of impaired healing. Dilations may have to be done repeatedly during the period of steroid tapering.

The preceding maneuvers may also be used at the time of elective surgery if the patient has presented with a stable airway. This plan allows assessment of the distal airway, placement of an endotracheal tube, and provision of an adequate lumen to prevent accumulation of carbon dioxide early in the procedure. At the time of tracheal resection, this tube can be pulled back or removed and a sterile cuffed endotracheal tube (Tovell tube) can be inserted into the distal airway. Sterile connecting tubing is connected for ventilation. The tube can be removed when necessary, for suctioning or placement of sutures. At the conclusion of the operation, the original endotracheal tube is advanced into the distal airway and the sutures are tied. The patient should be breathing spontaneously at the end of the procedure so that extubation can be done in the operating room. Intraoperative, high-frequency ventilation has been used with equal success, and is especially useful in certain complex carinal reconstructions.

Tracheostomy may be the only way to secure control of the airway. The tracheostomy tube should always be placed through the most damaged portion of the trachea, if it is accessible in the neck, preserving the maximal amount of normal trachea for subsequent reconstruction. A tracheostomy done at the completion of tracheal resection should be located at least two rings away from the anastomosis, and the anastomosis should be protected with the thyroid gland or strap muscles to avoid contamination of the suture line. This practice lessens the likelihood of subsequent dehiscence or stenosis. A tracheostomy tube should never be placed through the anastomosis. The brachiocephalic artery is also protected with a cervical strap muscle.

Preoperative evaluation

The general medical condition must be assessed individually, particularly in patients undergoing extensive intrathoracic resection and reconstruction. Patients should not be receiving steroids or require mechanical ventilation postoperatively: either of these factors could lead to serious problems with the anastomosis. Pulmonary function testing and quantitative ventilation perfusion scans are invaluable, allowing accurate prediction of postoperative pulmonary function in patients undergoing complicated intrathoracic resection with loss of pulmonary parenchyma. Anterior cervical or upper cervicomediastinal tracheal resection, on the other hand, is essentially external to body cavities and is well tolerated, even in patients with very poor pulmonary reserve, advanced age, or stable coronary disease.

Release maneuvers

Most tracheal operations can be done without release maneuvers: laryngeal release was required in only 27 (8.3 per cent) of 327 tracheal resections for postintubation stenosis and in only 18 (15 per cent) of 119 tracheal resections for primary and secondary tumors.

Judicious assessment of each patient's condition is necessary to establish the safe limits of the possible extent of tracheal resection. Previous operations (including mediastinoscopy), disease process, the extent of the lesion, age, and body habitus are important factors in deciding which patients are likely to require release maneuvers. The site of the disease is also important in determining which release procedures will be of benefit. Certain maneuvers are more effective for achieving additional length when disease is located in the cervical trachea; others are more effective for the intrathoracic trachea. A release maneuver is primarily performed to prevent unnecessary tension on the anastomosis and to avoid the need for excessive dissection of the trachea, which might jeopardize the lateral blood supply.

Cervical trachea

The simplest maneuver to minimize tension after tracheal resection is flexion of the neck and mobilization of the pretracheal plane, avoiding the lateral blood supply to the trachea. Flexion of the neck between 15 and 35° may result in downward movement of the trachea by as much as 4.5 cm, or the equivalent of seven tracheal rings. Flexion beyond this may achieve up to 1.5 cm of added length. The amount of trachea that may be removed with simple flexion varies greatly with age and physical habitus. When these simple maneuvers fail to give sufficient length, a Montgomery suprahyoid laryngeal release is performed, generally before the anastomosis is completed. The release is made by dividing the muscles that insert on the superior aspect of the hyoid bone. The hyoid bone is divided just medial to its lesser cornua; an additional 1.5 cm of length can be obtained.

Intrathoracic trachea and carina

Flexion of the neck and mobilization of the anterior surface (and posterior to a much lesser degree) of the trachea are also important for lower tracheal lesions. Even in the intrathoracic position, they will allow the trachea to devolve into the thorax. Laryngeal release is of no help in gaining additional length for intrathoracic tracheal lesions much beyond the midtrachea.

The right hilum and the inferior pulmonary ligament should be mobilized. A U-shaped incision in the pericardium below the inferior pulmonary vein allows the hilar structures and the bronchus to advance. Additional length is obtained by incising completely the pericardium around the hilar vessels (Fig. 1). A posteriorly based pedicle of tissue that includes a bronchial artery and some lymphatics should be preserved whenever possible. Attempts to ‘recreate' a carina by joining the left and right mainstem bronchi will allow little, if any, advancement of the neocarina. Length can only be obtained by advancement of the trachea from above. Carinal resections are more commonly reconstructed by end-to-end anastomosis between the distal trachea and either the left or right mainstem and end-to-side anastomosis of the remaining mainstem bronchus. Bilateral hilar release is difficult to achieve through most exposures without a drastic extension of approach.

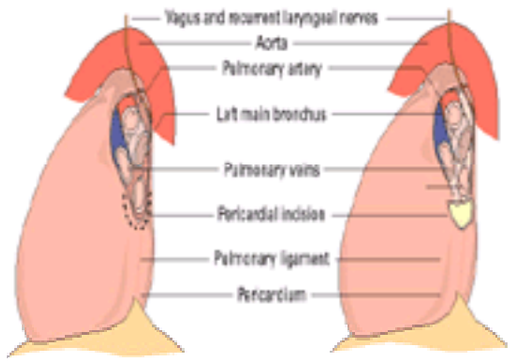


Fig. 1. The left-side intrapericardial hilar release technique showing the U-shaped pericardial incision allowing up to 2 cm of upward hilar mobility to facilitate the creation of a tension-free anastomosis.

Midtrachea

Lesions in the midtrachea can benefit from all of the release maneuvers described. Laryngeal release may predispose the patient to aspiration, particularly of thin liquids, although this problem usually resolves with time. This problem is less common when the Montgomery technique of suprahyoid release is used in preference to a thyrohyoid release.

Tracheostomy

Tracheostomy is one of the oldest surgical procedures. The ideal method limits loss of tracheal substance, to minimize the possibility of subsequent stomal stenosis. More important is avoidance of excessive leverage on the tracheostomy tube, which may erode the stoma. As the defect heals by cicatrization, stenosis may result.

The procedure is best done in the operating room with a transoral endotracheal tube in place. With the patient supine, a pack 8 to 10 cm (3–4 inches) thick is placed below the shoulders to hyperextend the neck. If local anesthesia is used [1 per cent lidocaine (lignocaine)], care should be taken to anesthetize all the soft tissues down to the pretracheal fascia. A transverse incision is made 1 cm below the cricoid cartilage and carried through the platysma muscle down to the strap muscles. These are separated in the midline and the pretracheal fascia is divided vertically.

The thyroid isthmus is usually divided, unless it can be retracted easily to expose the second and third tracheal cartilages. The trachea is cleared of all adventitial tissues. The second and third tracheal rings are incised vertically, along with part of the fourth, if necessary. The oral endotracheal tube is withdrawn until the tip is just proximal to the stoma.

A spreader holds the tracheal incision open, and a tracheostomy cannula of appropriate size is inserted into the trachea. The tracheal balloon is inflated to the point of preventing air leakage, and the tracheostomy tube is tied in place. The flange of the tube can be sutured to the skin for additional security.

The technical and anatomical pitfalls of this operation include injury to the anterior cervical veins, left innominate vein, innominate artery, internal jugular vein, and carotid artery, as well as the creation of a pneumothorax, especially in children, and placement of the stoma too low, which may result in erosion of the innominate artery. Placement too high may result in damage to the cricoid cartilage, the cricothyroid membrane, or the subglottic larynx. Although tracheostomy is often considered a ‘minor' procedure, to be delegated to a less experienced surgeon and done in far from ideal circumstances, it should be done under the guidance of a competent surgeon, ideally in an operating room, or at least in a well-equipped procedure room or intensive care unit. While recent percutaneous techniques have been made safer by use of a guide wire, there is little to recommend them.

Tracheal strictures

Strictures of the trachea may result from a number of different causes, most of which are considered separately below. Congenital stenoses are rare. Post-traumatic strictures, particularly where there has been tracheal separation, occur. Strictures are most commonly iatrogenic, resulting from intubation, usually for ventilatory support. Infections still cause stenosis of the trachea, including tuberculosis, diphtheria, histoplasmosis, scleroma, mucormycosis. A miscellaneous group of diseases that also causes strictures or, at least, tracheal obstruction, includes amyloid disease restricted to the airway, tracheopathia osteoplastica, Wegener granulomatosis, relapsing polychondritis, tracheobronchomegaly, and sarcoid. Idiopathic stenoses occur also in patients with no history of previous insult or infection and no systemic disease.

Congenital lesions

Congenital tracheal lesions are rare but a subject of interest. Tracheal agenesis or atresia is generally fatal, despite the embryological possibility of normal formation of the larynx and lungs. A communication between bronchus and esophagus may exist, and short-term survival can be achieved by using the esophagus as an airway. The more common, tracheoesophageal fistulas have been well described and are usually more an esophageal than a tracheal problem. The tracheal portion of the fistula is managed by division and repair.

Congenital tracheal stenosis may present as a web-like diaphragm, most often at the subcricoid level. More lengthy stenoses may involve the entire trachea, sometimes with normal larynx and main bronchi, or variable lengths of the trachea (Fig. 2). Funnel-like narrowing is sometimes seen, with gradual narrowing to the stenotic segment either in the upper or lower trachea. Segmental stenosis of the lower trachea may be accompanied by bronchial anomalies, such as all or part of the right upper-lobe bronchus originating from the trachea just above the stenotic segment. The main bronchi may also be stenotic. Most often the cartilaginous rings are complete circles in the area of the stenosis (Fig. 2), but the cartilages may be highly disorganized. Rarely there is segmental malacia. Other anomalies may occur in a patient with congenital tracheal stenosis.

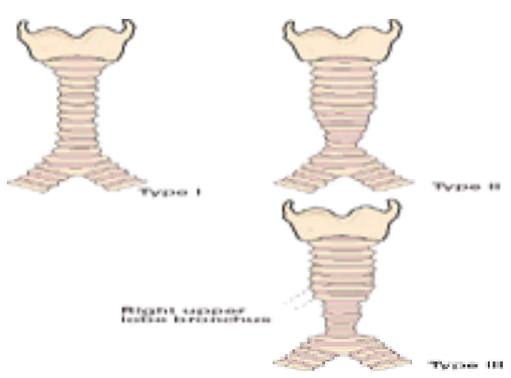


Fig. 2. Congenital tracheal stenosis. (Left) Type 1, generalized hypoplasia of the trachea: the airway has a normal caliber at the level of the cricoid cartilage and also in the main bronchi. (Center) Type II, funnel-like narrowing: the trachea has a normal caliber immediately below the cricoid cartilage, but funnels to its narrowest point most frequently just above the carina. (Right) Type III, segmental stenosis may be accompanied by bronchial anomalies; the segmental stenosis may vary in length and may be at various levels.

Aberrant pulmonary artery or ‘pulmonary artery sling' may occur in association with congenital stenosis of the lower trachea. The left pulmonary artery originates from the proximal portion of the right pulmonary artery and passes behind the trachea to the left lung. Patients with stenosis must be differentiated from those with a compressed and malacic segment of trachea adjacent to such a pulmonary artery sling.

Congenital vascular rings may also cause compression of the airway. The trachea and esophagus may be compressed without any associated tracheal anomaly, although secondary malacic changes may occur. The airway may be similarly compromised in patients with a double aortic arch, with a right aortic arch and a patent ductus arteriosus or ligamentum arteriosum, with an aberrant subclavian artery, with a diverticulum of Kummerell, or with an abnormal innominate artery.

In contrast to congenital tracheal stenosis, congenital tracheomalacia is rare. Some reported cases may well be secondary to the treatment of stenosis in infancy. A rare anomaly leading to airway obstruction is congenital agenesis of the right lung with compression and malacia of the lower trachea and left main bronchus. Malacia occurs because the heart moves into the right axillary line, rotating the aortic arch, and pulling the left pulmonary artery posteriorly against the vertebral column and aorta. The lower end of the trachea and the first part of the left main bronchus pass beneath the aortic arch at this point: the pulmonary artery lies anterior. Marked compression with secondary malacia may follow. In the adult, similar lesions are seen in ‘postpneumonectomy syndrome'. Another rare anomaly is laryngotracheoesophageal cleft, in which there is an absence of the common wall between the posterior larynx and the upper trachea and the esophagus.

Tracheobronchiomegaly (Mounier–Kuhn disease) is properly classified as a congenital disease. Despite its presentation in adult life, careful questioning reveals no clear point at which symptoms commenced. Retrospectively, patients and their families are aware that breathing has never been normal. The circumference of the tracheal wall is greatly increased, with irregularities and deformities of the cartilages. This change may lead to a reverse curvature of parts of the anterior wall of the trachea and the main bronchial cartilages, with elongation and redundancy of the membranous wall accompanied by marked thickening.

The diagnosis of congenital stenosis should be suspected when respiratory distress is recognized in an infant, often with inspiratory and/or expiratory stridor. Dyspnea may be paroxysmal rather than continuous. Feeding difficulties, failure of normal development, and frequent and stubborn respiratory infections may be present. Air tracheograms and fluoroscopy describe the anomaly more precisely: contrast medium is not generally necessary if the radiologist is skilled. A CT scan with thin sections provides comparison of the diameter and shape of the airway at various levels. Conventional radiography provides a well-delineated longitudinal picture. Such images are clearer than are sagittal views from MRI.

Bronchoscopy provides excellent delineation of the portion of the airway proximal to the stenotic lesion, and sometimes of the distal airway if the instrument can be passed safely. Rigid pediatric bronchoscopes with magnifying telescopes provide a more precise view than do flexible instruments. A fine, long telescope may be passed through a tight stenosis. Circumferential rings may be clearly seen submucosally in many patients. Disorganized cartilages are, however, less well defined. The bronchoscopist must be alert to the possibility of vascular rings. Now well defined by CT and MRI, these lesions do not necessarily require angiography for diagnosis.

A conservative therapeutic approach should be adopted to airway lesions in infants and small children. The juvenile trachea tolerates anastomotic tension less well than the adult, probably because of the delicate tissues. Furthermore, a small amount of edema in a tiny airway may be much more obstructive than a similar amount in an adult. Intubation and ventilation through the juvenile airway are more likely to injure a recent anastomosis or the native trachea. If the child can be treated conservatively, growth of the airway alone may lessen the obstructive symptoms. While the growth of a stenosis is in general proportional to that of the normal tissues and does not actually correct a congenital narrowing, improvement may be sufficient to obviate the need for correction or to permit the patient to grow, thus reducing the risks of surgical reconstruction.

Tracheal webs may be divulsed or removed bronchoscopically with biopsy forceps, cauterizing electrodes, or the laser. A limited stenotic lesion may be resected and an end-to-end anastomosis performed, using the techniques that have evolved for management of the trachea in the adult. The lower trachea in an infant is easily accessible from an anterior approach. Although only a limited amount of information has been obtained in children, animal experimentation, which is probably pertinent in this respect, shows that anastomotic sites grow on average to 80 per cent of the normal sagittal diameter and 75 per cent of the coronal diameter, achieving 80 per cent of a normal lumen. This process permits essentially normal function.

If a congenital stenosis is long and the child is unable to tolerate conservative management, widening may be achieved by linear incision and the insertion of a gusset of pericardium or cartilage. The trachea is stented for a prolonged period. The mesenchymal patch does not appear to contract with time. Dilatation alone is not useful in treating congenital stenoses. Forceful dilatation may split the circular cartilaginous rings, which cannot stretch. Attempts to manage congenital strictures by injection of steroid or by laser treatment are unavailing. More recently, ‘slide tracheoplasty' has been used successfully to correct long congenital stenosis. The stenosis is divided at its midpoint, proximal and distal stenotic segments opened vertically, on opposite sides, and the two anastomosed, thus doubling the circumference and quadrupling the cross-sectional area. Trachea only is used for the repair, bypass is unnecessary, and patients are promptly extubated.

Trauma

The trachea and carina may be damaged by blunt or penetrating trauma. The most common causes of closed tracheal trauma in the neck are impact against the steering wheel or dashboard in a motor vehicle accident, or an encounter with a cable across the path of a motorcyclist. Injury to the lower trachea and carina comes from impact against the chest wall, most often in a motor vehicle accident.

Gunshot wounds are the most frequent cause of penetrating wounds to the trachea, although rare.

External trauma

Blunt cervical injury may affect the airway at any level from the hyoid bone to the carina. Crushing and hyperextension may result in crush injuries to the larynx and trachea, separation of the larynx from the trachea, or tracheal rupture. Avulsion of the esophagus from the pharynx at the cricopharyngeal level or traumatic injury to the esophageal wall associated with similar injury to the membranous wall of the trachea may occur concomitantly. The presentation of such injuries may be subtle. Contusions and lacerations of the neck or chest may be noted, and subcutaneous emphysema may be detected in the neck. If the injury is lower, radiographs may show air dissecting mediastinal tissues. Unilateral or bilateral pneumothorax may result from tracheal injury within the thorax.

Obstruction may result from crushing of the larynx and from resultant hematoma. Patients may either die immediately of airway obstruction from tracheal transection, or they may suffer varying severity of dyspnea. A patient with an initially satisfactory airway may rapidly decompensate and develop obstruction either spontaneously or during intubation or examination of the airway. Hoarseness, inspiratory stridor, respiratory distress, or change in voice (either dysphonia or aphonia) are observed. The recurrent laryngeal nerves may be injured unilaterally or bilaterally, either temporarily or permanently. Associated injuries include subluxation of the cervical spine, with or without neurological injury, and injury to the aortic arch or great vessels. Penetrating wounds to the trachea, more frequent in the cervical region or thoracic inlet, may well be accompanied by damage to the great vessels.

When the lower trachea is damaged, prominent injuries to the chest, typical of high-impact trauma, are often seen. These include fractures of the sternum, ribs, or costal cartilages. In children and young adults, severe compression of the chest with resultant injury to the airway may occur without rib fractures. Injuries to the carina and main and lobar bronchi occur from crushing injuries to the chest; these are more common than tracheal injuries. Tracheal and carinal injuries may present with unilateral or bilateral pneumothorax, in addition to mediastinal dissection of air. Rupture at the carina causes a linear, spiralling split upward from the carina, either anteriorly or posteriorly. Bilateral or unilateral pneumothorax from tracheal injury in the mediastinum sometimes responds to tube drainage, with initial sealing of the air leak as a result of approximation of peritracheal tissues.

The site of impact in external injury or the point of entry of a penetrating injury, together with the signs and symptoms described, should announce the probability of a ruptured airway. Respiratory distress due to tension pneumothorax demands prompt diagnosis and treatment. Even if the patient shows little distress, radiological studies should be undertaken, with constant observation for the possibility of sudden deterioration of the airway.

A CT scan can show quite dramatic changes, but is not usually justified unless the patient is stable and there are special indications for the study. In moving such patients, the usual precautions against possible instability of the cervical, dorsal, and/or lumbar spine, and against possible associated injuries, are observed.

Endoscopy is important, but may precipitate acute obstruction of an already unstable airway held together only by tenuous peritracheal connective tissue. If tracheal injury is suspected, the surgeon must be prepared for immediate tracheostomy if difficulty ensues. It is permissible to attempt bronchoscopy, using either a flexible bronchoscope with an endotracheal tube threaded over it so that it may be inserted directly, or a rigid bronchoscope. Rigid bronchoscopy should only be undertaken after injury to the cervical spine has been ruled out. If there is damage to the trachea posteriorly, or if it is transected, esophagoscopy is also in order after securing the airway. An unrecognized fistula from the esophagus may produce lethal mediastinitis.

The laryngotracheal junction is frequently injured in tracheal trauma. Deceleration injuries can cause long tears in the membranous portion of the lower trachea. Vocal cord function should be assessed, if possible, although one may expect the cords not to function if there is complete transection of the cervical trachea or avulsion from the larynx.

The most important initial aspect of treating a laryngotracheal injury is the establishment and maintenance of an airway. If an endotracheal tube can be passed without excessive manipulation, it provides initial stabilization and permits a more measured approach to repair of the injury. If, however, an airway cannot be established, urgent tracheostomy must be performed. The operator encounters masses of clot and contused tissue, and the trachea may not be evident immediately. The separated distal end retracts into the mediastinum and is best identified by the exploring finger. The anterior wall of the separated trachea is grasped with a forceps and is drawn up into the lower cervical region, where the disrupted end of the trachea is intubated. The wound is debrided and the extent of damage determined after the establishment of a safe airway. A skilled operator familiar with techniques of tracheal repair may directly anastomose the trachea. If the operator is not skilled or if other life-threatening injuries demand treatment, it is preferable to anchor the distal trachea in the base of the neck and to place a tracheostomy tube in the distal, separated end. Distal tracheostomy should not be done, as this will only further injure tracheal tissue and provide no advantage over a tube placed in the already divided end. The wound is closed with drainage after adequate cleansing; secondary repair is undertaken later. If the trachea is repaired after conservative debridement, it is usually necessary to place a small, protective tracheostomy approx. 2 cm below the anastomosis, since glottic function is usually severely impaired by paralysis of the recurrent laryngeal nerve. Later procedures on the glottis will provide an adequate aperture despite the paralysis. It is frequently not recognized that a paralysed larynx, given an adequate glottic aperture, is a functional organ. The aperture permits easy respiration and produces a voice that is hoarse but completely comprehensible. Such apertures may be established by cord lateralization, arytenoidectomy, or arytenoidesis.

If the larynx has been damaged structurally by the injury, it should be reconstructed. This is often best accomplished by placement of an internal stent, such as the Montgomery moulded laryngeal stent or a T-tube that passes through the vocal cords. These injuries are best handled by a team of specialists that includes an otolaryngologist.

Partial or complete disruption of the esophagus from the pharynx at the cricopharyngeal level must be repaired or a damaging cervical fistula will result. The closure is buttressed with a pedicled cer-vical strap muscle to prevent a fistula from tracking along the reparative sutures (Fig. 3). It is not yet possible to repair the recurrent laryngeal nerves. While laryngeal function will sometimes return, permanent paralysis of the larynx often results even if a nerve has only been contused. Satisfactory function may be attained, as noted above.

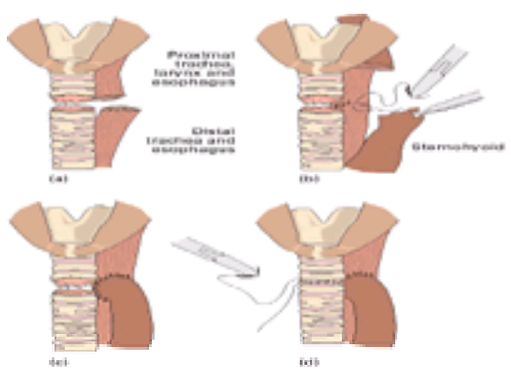


Fig. 3. Repair of traumatic injury to trachea and esophagus: (a) view of transected trachea and esophagus; (b) oesophagus closed in two layers;(c) strap muscle interposed between esophagus and tracheal suture line;(d) completed repair.

Lacerations of the lower end of the trachea can often be repaired, by those skilled in such techniques, by an anterior approach either through partial sternotomy or, in rare cases, through full sternotomy with an added approach through anterior and posterior pericardiotomy between the superior vena cava and aorta. For those less familiar with this technique, the lowermost trachea, as well as carina and bronchial ruptures, are best approached through a right thoracotomy in the fourth interspace or the bed of the fifth rib.

If tracheal rupture is not recognized or is not treated definitively and acutely, stenosis results. The stenosis will appear to be very long, since the distal, separated segment usually retracts into the mediastinum. The functional adequacy of the larynx is first assessed and an adequate glottic aperture is established. Laryngotracheal continuity is usually re-established at a second procedure. If an esophageal fistula is also present, it is repaired and supported with a tissue buttress.

Burns

Inhalation burns of the larynx, trachea, and bronchi may be particularly difficult to treat. The injurious agent may be chemical, thermal, or a combination. Patients often show little damage to the pharynx or supraglottic larynx once the acute injury has subsided. Persistent damage begins in the subglottis, below the vocal cords, and extends down the airway in a gradually diminishing range of injury. How far down is probably related to the ‘dose' as well as the injurious potential of the agent. The degree of inflammatory change, granulation, and scarring depends on the depth of mucosal injury. The tracheal rings are not generally destroyed. Resection is usually precluded, both by total involvement of the subglottic larynx from the undersurface of the vocal cords down as well, and by the length of the tracheal injury. A further caution is that burns of the airway respond to early surgery, even where limited in linear extent, in much the same way as burns of the skin and other tissues. Intense proliferation of scar tissue follows early intervention, with the probability of restenosis. A necessary ventilatory cuff may compound the injury in a localized area.

Airway obstruction caused by burns can be managed initially by a tracheostomy placed at the second or third ring, typically within the area of burn injury, followed later by placement of a silicone T-tube to span the injured area. Proximally, the tube must often extend through the vocal cords. Most patients retain a hoarse but understandable voice and, surprisingly, do not appear to aspirate if they eat carefully. In some patients the T-tube has to remain in place for years while the burn scar gradually subsides. If the original injury is to the more superficial inner layers of the trachea, there is apparent regression and maturation of the connective-tissue proliferation with re-epithelialization over time—much the same process as is seen in similar cutaneous burns. T-tubes can be removed without need for further treatment. The larynx and trachea do not return to wholly normal function or diameter, but the airway is usually adequate. If not, resection of a short segment of stenotic airway is possible. The acutely burned trachea also appears to be susceptible to injury from pressure of the cuff of the endotracheal tube used to ventilate the patient initially. Major stenosis may follow.

Postintubation stenosis

Following intubation, strictures principally occur at the level of the tracheostomy stoma or of the tube cuff (Fig. 4). The upper strictures usually occur in patients who are being maintained on ventilators. Indirect evidence suggests that the etiology is cicatricial healing of a stomal opening that had once eroded the anterior and lateral walls of the trachea. While this injury can be abetted surgically and while invasive infection may also play a part, it appears to be due most often to leverage of the tracheostomy tube against the stoma because of improperly suspended heavy equipment. Scarring is apt to occur in a patient receiving prolonged support through a tracheostomy tube. Cicatricial stenosis occurs at the anterior and lateral site of the original defect, producing an A-shaped stenosis when viewed through the bronchoscope with the patient supine. Careful suspension of connecting tubing has prevented these erosive lesions.

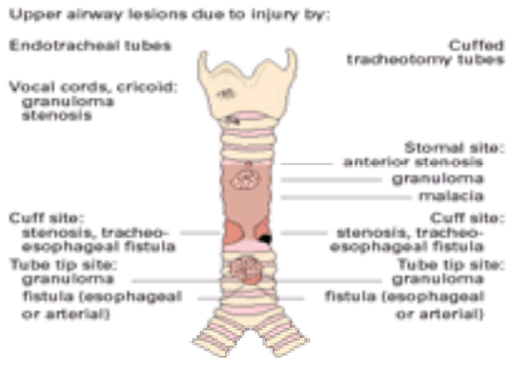


Fig. 4. Diagram of the locations and types of upper airway lesions evolving from injuries caused by endotracheal tubes and cuffed tracheostomy tubes. Note that cuff injuries may be the same.

In contrast, a high-pressure cuff or a low-pressure cuff used in a high-pressure range by overinflation will produce circumferential pressure injury in the trachea. As healing occurs, a circumferential stenosis results. If erosion has been deep and has destroyed the cartilages, the resulting stenosis will never be amenable to cure, even by the most prolonged stenting, since there is no normal architecture left in the tracheal wall. Varying depths of erosion are, of course, seen.

Cuffs on endotracheal tubes may also produce this lesion. Patients who develop stenosis after only 48 h of ventilation usually have had only an endotracheal tube in place. In many such patients the injury consists of a massive cicatricial stenosis of scar tissue within the tracheal lumen, although the cartilages are not totally destroyed, and may even be relatively intact. Endotracheal tubes may also cause glottic stenosis and stenosis at the level of the cricoid, but these are not true tracheal lesions. It is important, however, to be aware of this possibility before considering repair of a tracheal lesion. Corrective surgery on the trachea can be disastrous if there is an inadequately functioning glottis above it. Severe stenoses resulting from cricothyroidostomy may also occur in the subglottic region. In elderly, kyphotic patients a tube from a high stoma may gradually erode back through the cricoid cartilage and produce subglottic stenosis. A large-bore endotracheal tube may produce erosion and stenosis at the level of the cricoid. Repair of strictures involving the subglottic larynx is much more difficult than is repair of strictures of the trachea alone.

Inflammation may cause varying degrees of thinning of cartilages in the segment between the site of a tracheal stoma and the level of a cuff stenosis below. These areas may also become malacic. Tracheoesophageal fistula is seen most often in patients who require nasogastric feeding tubes for long periods of time in addition to an inflated cuff in the trachea (Fig. 5). The pressure of these two foreign bodies acts as an erosive pincer. Anterior erosion of the tracheal wall was seen more often when high-pressure cuffs were used routinely or when the tip of a tube was angulated forward. These sometimes caused erosion directly into the innominate artery as it crossed the trachea. A more common cause of hemorrhage from the innominate artery is the low placement of a tracheostomy tube so that the tube itself rests on the artery and erodes through it at the inferior margin of the stoma. Such lesions are seen most often in children and young adults, since their tracheas are more mobile and rise into the neck along with the innominate artery upon hyperextension. They can be avoided by placing the tracheostomy in an appropriate position at the level of the second and third tracheal rings, rather than with reference to the sternal notch.

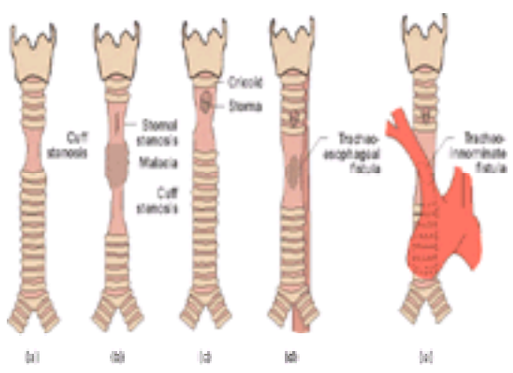


Fig. 5. Diagram of principal postintubation lesions. (a) Lesion at cuff site in a patient who has been treated with an endotracheal tube alone. The lesion is high in the trachea and is circumferential. Note that the endotracheal tube also induces glottic and subglottic laryngeal stenosis. (b) Lesions that occur with tracheostomy tubes. At the stomal level, anterolateral stenosis is seen. At the cuff level, lower than with an endotracheal tube, circumferential cuff stenosis occurs. The segment between is often inflamed and malacic. (c) Damage to the subglottic larynx. A high tracheostomy or one that erodes back by virtue of the patient's anatomy may damage the inferior cricoid and produce a low subglottic stenosis as well as an upper tracheal injury. (d) Tracheoesophageal fistula. The level of fistulization is usually where the cuff has eroded posteriorly. Occasionally, angulation of the tip of the tube may produce erosion from the tip. There is also usually severe circumferential damage at this level by the cuff. (e) Tracheo-innominate fistula. A high-pressure cuff frequently rests on the trachea directly behind the innominate artery. Erosion may occur, although rarely. The more common innominate arterial injury is from a low tracheostomy where the inner portion of the curve of the tube rests in proximity to the artery and causes direct erosion.

Clinical presentation

Most patients with postintubation stenosis have signs of upper airway obstruction. The patient initially complains of shortness of breath on exertion, followed by progressive shortness of breath even at rest, with wheezing and stridor. Occasionally, episodes of unilateral or bilateral pneumonitis are present. Cyanosis is a very late sign. Any patient who develops signs of upper airway obstruction must be considered to have an organic lesion of the trachea, particularly if they have recently undergone intubation for ventilatory support. Many of these patients are diagnosed as having adult-onset asthma despite their record of recent intubation.

Diagnosis

The clinical presentation and history give the presumptive diagnosis in most cases. This may be followed by simple tracheal radiographs of the type described in the section on diagnostic techniques. Fluoroscopy provides evidence that the glottis is functioning adequately and excludes the presence of malacia. Bronchoscopy may be required as a separate procedure in complex cases; in simpler cases it is done concurrently with the repair if the problem has been demonstrated radiologically.

Treatment

Surgical treatment of benign strictures of the trachea is standardized and is the treatment of choice if the lesion is not excessively long, does not involve other more complex anatomical structures such as the subglottic larynx, and if the surgeon has had enough experience to promise a good result. Most tracheal strictures may be managed indefinitely by reinstituting a tracheostomy, dilating the stricture, and inserting a silicone rubber T-tube (Montgomery tube). The tube requires changing at infrequent intervals.

An emergency dilation may be initiated through a rigid bronchoscope under general anesthesia, using just the tips of esophageal dilators through the bronchoscope. This is followed by serially larger pediatric bronchoscopes. The period of grace provided by such a dilation often allows further study of the patient or transport to a center where tracheal surgery is performed. For longer-term management, a tracheostomy tube or a T-tube may be introduced.

Where there is no contraindication to surgical excision and end-to-end repair, this is the treatment of choice. Morbidity rates are low and the success rate very high. In 521 tracheal resections and reconstructions for postintubation stenosis, 94 per cent had good or satisfactory results, 4 per cent failed and mortality was 2.4 per cent; the results improved in the latter half of the series. Surprisingly, very similar results were obtained in reoperations on 75 patients who previously had had unsuccessful repairs. Nearly all repairs of postintubation stricture are now undertaken through an anterior approach, using a collar incision with or without vertical, partial division of the sternum (Fig. 6). Dissection is kept close to the trachea to avoid injury to the recurrent laryngeal nerves. Stenoses are dilated before surgical repair if they are less than 6 mm in diameter in order to avoid retention of carbon dioxide and resultant arrhythmias. Division is usually made below a stricture, unless it is a low one. Intubation is carried out across the operative field, and the stricture is dissected from the esophagus (Fig. 6c). Strictures up to half the length of the trachea may be thus removed, followed by approximation. In older patients, a laryngeal release may be required to provide additional length. Tracheostomy is rarely used. If the

stricture involves the subglottic larynx, a single-stage repair of the lesion requires partial removal of the lower anterior subglottic larynx (Fig. 7, Fig. 8). If the stenosis is circumferential, the posterior scar is removed from the anterior surface of the posterior cricoid plate.

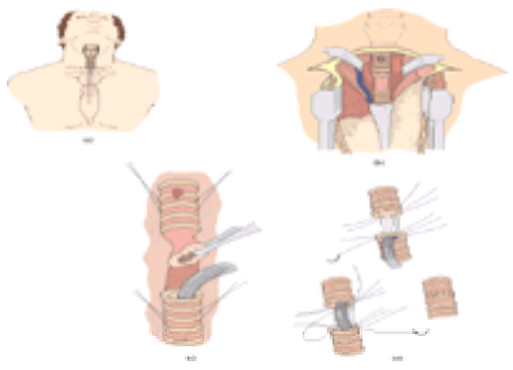


Fig. 6. Reconstruction of the upper trachea. (a) The collar incision permits exploration of the entire trachea and, in many cases, reconstruction. Division of the upper sternum just through the sternal angle widens the access for lower or more difficult lesions. (b) Sternal division in this case has made access to the lower trachea easier. Since the trachea points posteriorly, there is nothing to be gained by division of the innominate vein. The pleura is intact. (c) The trachea has been divided below a stenotic lesion and the patient intubated directly. The specimen is elevated. This allows easier dissection of a lesion that may be fused to the esophagus. Proximally, it protects the recurrent nerves. (d) The technique of anastomosis. Posterolateral sutures are first placed; the proximal tube is then advanced from above after withdrawing that in the distal trachea, and the anastomosis is completed. Sutures are generally tied from front to back, with all knots on the outside. The temporarily approximating lateral traction sutures are not shown nor is the cervical flexion that removes tension.

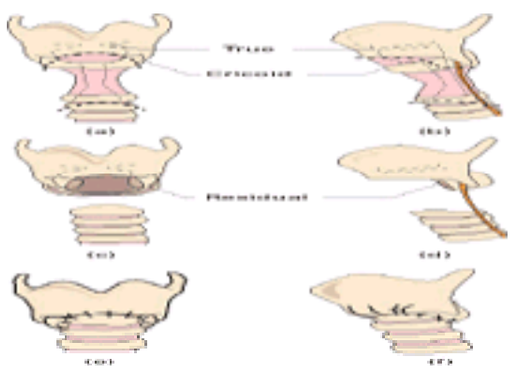


Fig. 7. Operative repair of anterolateral stenosis of the subglottic larynx and upper trachea: (a) anteroposterior view; (b) lateral view, both showing the extent of pathology and the ultimate lines of transection. The stenosis extends into the subglottic larynx well above the border of the cricoid anteriorly. There is, however, no pathology involving the posterior mucosal wall of the subglottic larynx or of the upper trachea. Proximal line of transection is centred in the midline to divide larynx from the trachea posteriorly at the lower border of the posterior cricoid lamina or cricoid plate. Inferiorly, the most proximal tracheal ring of the residual trachea is cut backward to its posterior ends. (c, d) Larynx and trachea after removal of the specimen. Recurrent nerves have been left intact but are not dissected out, as might be suggested by the diagrammatic representation. The mucous membrane of the larynx has been transected sharply at the same level of division as the cartilage. (e, f) Anteroposterior and lateral views of the reconstruction.

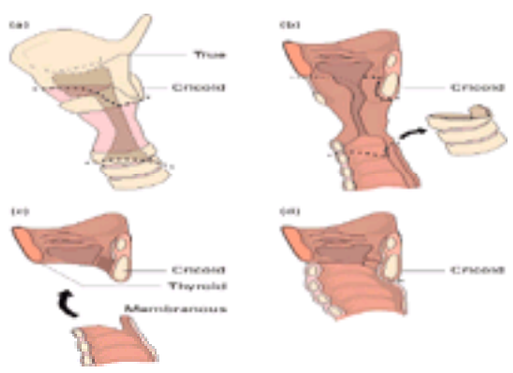


Fig. 8. Resection and reconstruction of circumferential stenosis of subglottic larynx and upper trachea. (a) External line of cartilaginous division of both larynx and trachea is the same as in the anterolateral stenosis.(b) Interior view of larynx and trachea demonstrates modifications necessary where stenosis involves the mucosa and submucosa just in front of the posterior cricoid plate. Superior dotted line indicates that the mucosa with its scarring will be cut back to within a short distance, if necessary, of the arytenoid cartilages. Inferiorly, the posterior membranous wall has been retained as a broad-based flap. (c) Resected specimen, leaving a bare area of the intraluminal portion of the lower part of the cricoid posterior lamina. The flap of membranous wall of the trachea will be fitted into this defect to provide prompt and complete mucosal coverage. (d) Mucosa of the larynx has been anastomosed to the mucosa of the membranous wall of the trachea. External to the lumen, the connective tissue of the membranous wall has been fixed with four sutures to the inferior margin of the cricoid cartilage, assuring that the flap will stay firmly applied to surface.

Appropriate tailoring of the distal segment allows repair of both of these defects in a single stage, but the technique is difficult. Of a group of 80 patients undergoing this repair, 69 had excellent or good results, eight satisfactory, two failed, and one died from myocardial infarction.

Rarely does a postintubation stenotic lesion involve over one-half of the trachea, unless there has been a previous attempt at surgery. If primary reconstruction is impossible because of the extent of damage to the trachea, the insertion of a Montgomery silicone T-tube is an excellent long-term alternative. In such a case, open surgery is not required for insertion of a prosthesis, unless the proximal and distal airways are separated by scarring.

Tracheo-innominate artery fistula

When the neck is hyperextended for tracheostomy in children or young adults, over half the trachea, and the innominate artery, rise into the neck. It follows that if the tracheal stoma is placed in relation to the sternal notch rather than the cricoid cartilage, the stoma will actually be in the midtrachea and lie just above the innominate artery. A tube may erode through the artery immediately inferior to the margin of the stoma, producing the lesion shown in Fig. 5(e). The premonitory hemorrhage that often occurs should lead to inspection of the tracheal stoma, if necessary by bronchoscopy.

In the event of massive hemorrhage, control may be obtained by finger compression down and forward against the sternum at the site of the bleeding. An endotracheal tube is then slipped into the airway and the cuff inflated to prevent blood from running into the lungs. The surgeon's finger is kept in place while the patient is moved to the operating room.

Exposure is best obtained through a collar incision at the level of the stoma and a vertical sternotomy; hemorrhage is controlled with finger pressure during the initial dissection. During proximal dissection of the origin of the artery, care must be taken to avoid extending the injury. Dissection is carried distally on the innominate artery to a level just proximal to its bifurcation; a vascular clamp or tourniquet is applied. The recurrent laryngeal nerve must be avoided as it passes around the subclavian artery. When bleeding has been controlled, the next maneuver is usually resection of the damaged segment of the artery and oversewing of the proximal and distal ends with double suture lines. Simple ligation alone is likely to result in secondary hemorrhage. The stumps of the artery proximally and distally are buried under whatever healthy tissue is available: thymus, strap muscles, or even the omentum brought up substernally. An arterial repair would lie in a potentially septic field.

Neurological sequelae are rare. It is usually best to create a new tracheal stoma at a higher, more appropriate level (second or third tracheal ring) by inserting a tube long enough to pass beyond the previous stoma. Strap muscle is sutured over the original stoma. Such an arterial fistula is best prevented by correct initial placement of the tracheostomy tube.

A tracheo-innominate artery fistula can also be caused by erosion of a high-pressure cuff anteriorly through the tracheal wall and into the overlying innominate artery (Fig. 5(e)). Occasionally the tip of a tube angulating forward produces a similar erosion. It is impossible to get a finger to the level of this fistula, which lies entirely within the mediastinum. The immediate hemorrhage is controlled by the placement of an endotracheal tube with a high-pressure cuff or with a low-pressure cuff inflated sufficiently to compress the opening.

Exposure is then achieved as described previously. The involved arterial segment is excised, and proximal and distal ends are sutured closed. In these now rare cases there is usually circumferential damage to the trachea at the level of the cuff. This damage requires tracheal resection and end-to-end anastomosis. Such patients continue to require the ventilation that they were receiving just before the hemorrhage. The tube cuff must be positioned either well above or well below the anastomotic line to avoid pressure on the tracheal anastomosis, which could cause separation.

Tracheoesophageal fistula

Tracheoesophageal fistula follows prolonged ventilatory support, usually as a result of pressure on the apposed walls of esophagus and trachea, between an inflated tracheal cuff and an in-lying esophageal feeding tube. The fistula usually lies at the level of the cuff, somewhat below that of the tracheal stoma (Fig. 5(d)). More often than not, the intratracheal balloon produces a circumferential injury, of which the fistula is just a part. If the fistula is identified while the patient still requires respiratory support, repair is postponed. The esophagus is kept free of tubes, and the lowest pressure high-volume cuff that will provide a gentle seal is used to continue the ventilation. A gastrostomy tube is inserted to drain the stomach and prevent reflux through the fistula. The patient is fed through a jejunostomy. Once the patient has been weaned from the respirator, a single-stage operation is performed.

The collar incision often circumcises the stoma (Fig. 9). Rarely is the vertical limb needed. The initial dissection is identical to that described for upper tracheal resection. After transection of the trachea below the distal margin of the stenotic segment, the specimen is elevated and the fistulous connection completely isolated circumferentially and dissected from the esophagus, removing the margin of the fistula. The anterior tracheal stoma is often close enough to the injured segment of trachea to be included conveniently in the resection. If there is a segment of fairly normal trachea between the stoma and the stenotic segment, the stoma may be left in place and allowed to close spontaneously later. If the fistulous opening is long and the tracheal wall is not destroyed anteriorly as far down as the fistula extends, the margin of the tracheal opening is excised as a V with its point aimed distally, so that the posterior membranous wall may be reconstructed with a vertical suture line before the tracheal ends are anastomosed (not shown). If necessary, some of the esophageal wall may be included at the sides of the V so that there will be enough tissue for reconstruction of the membranous wall of the trachea.

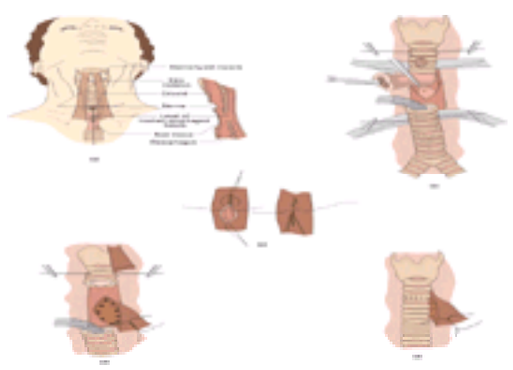


Fig. 9. (a) Exposure for most tracheoesophageal fistulas is through a low collar incision. Occasionally, a partial upper sternotomy is required for more distal exposure of the trachea. (b) Circumferential dissection above and below the fistula is very near the trachea to avoid injury to the recurrent nerves. Division of the damaged trachea gives excellent exposure of the esophageal defect. Ventilation is maintained by a sterile Tovell tube in the distal trachea. (c) The esophageal defect is closed in two layers using interrupted 4/0 silk sutures: (i) the first layer closes the esophagus; (ii) the esophageal muscle is closed over the first layer. (d) A local strap muscle is used to buttress the esophageal closure and separate it from the tracheal suture line. (e) The trachea repaired using interrupted 4/0 Vicryl sutures. Not shown are four lateral traction sutures of 2/0 Vicryl.

The esophagus is closed longitudinally with two layers of 4/0 silk. The sternohyoid muscle, or sometimes the sternothyroid, are used to create a pedicle and sutured into place over the esophageal closure, interposing healthy muscle between the vertical suture line in the esophagus and the transverse suture line of the tracheal anastomosis about to be made. The completed anastomosis of the trachea is shown in Fig. 9.

In the rare case when direct trauma has produced a tracheoesophageal fistula without circumferential damage to the trachea, repair does not require resection of the trachea. The trachea and esophagus are dissected closely until the fistula is identified fully. The recurrent laryngeal nerves are not deliberately exposed. Extra tissue from the anterior esophageal wall is used to provide sufficient material to allow closure of the membranous tracheal wall without tension. The esophagus is closed in layers, muscle is interposed, and the trachea is then repaired longitudinally. Exposure is difficult because the trachea has not been divided.

Persistent tracheal stoma

Persistent tracheal stomas usually have a mucocutaneous union of trachea and skin. The patient has often had a tracheostomy for a long time, with extended periods of ventilation. Simply drawing a strap muscle over the aperture creates the potential for formation of a granuloma from the raw tissue presented to the interior of the trachea; this is particularly likely to occur with a large stoma. It is better to cover the stomal orifice with a circular flap developed from the peristomal skin. Perioperative antibiotics help to prevent wound infection.

A circular incision is made around the stoma to define the skin that will be used for closure. Wedges of skin are excised on either side to permit linear closure of the circular defect. The collar of skin is elevated carefully, leaving sufficient attached to the trachea to preserve its blood supply. The skin is now inverted on itself by a subcuticular suture that effectively closes the stoma with an epithelial, (skin)-lined surface. Margins of the skin defect are elevated. The strap muscles are dissected free on either side and approximated in the midline in layers. The platysma and skin are closed transversely. Functional and cosmetic results are excellent.

Postinfectious stenosis

A variety of specific infections can cause stenosis of the trachea. Diphtheria was notorious for producing such injuries, though it is now rare. A small number of patients with a history of diphtheria in childhood develop stenosis later is life: it may be difficult to be certain whether the stenosis is a result of tracheostomies established for the treatment of the diphtheria, or of the diphtheria itself. Tuberculosis can cause severe stenosis, particularly of the lower trachea. Typically, the fibrosis is submucosal. An accompanying stricture of either the left or right main bronchus is also seen; upper-lobe bronchus may also be involved. If the stenosis is mature and not excessively long, resection and reconstruction may be definitive. Management becomes difficult in the presence of considerable residual inflammation. It is difficult to manage such a patient by intubation alone because of the involvement of the carina. A period of watching and waiting for the inflammation to subside with pharmacotherapy is the best cure.

Mediastinal fibrosis may be accompanied by marked narrowing and stenosis of large segments of the trachea and bronchi. At least some patients with mediastinal fibrosis have pathological or bacteriological evidence of previous histoplasmosis. Some may have problems amenable to extensive surgical resection and reconstruction to save at least some of the lung tissue. In others the mediastinum is so massively fibrotic and the length of trachea and bronchi involvedso great that excision and reconstruction are impossible. There is no known therapy except periodic redilatation and consideration of stents.

A number of other infectious diseases uncommon in the United States of America have been implicated in airway obstruction, including scleroma. Mucormycosis is another rare cause.

Tracheal collapse

Tracheal collapse is seen with a variety of lesions. There are many references to congenital tracheal malacia in the literature. Few of these cases are well documented, however, and many have complicating factors; the malacic changes may well be the result of intubation. The infant trachea is composed of very fine structures, and even slight injury may lead to softening of the already tender rings.

Segmental malacia is seen as a dominant lesion in a number of patients who have had cuff injury to the trachea following intubation. Such patients may have an apparently normal air column on static radiographs. Fluoroscopic observation, however, demonstrates a segment that collapses on cough or forced respiration. Resection is the usual treatment.

Chronic compressive lesions may also lead to collapse of the trachea. A large goitre, a cystic thymus, an aneurysm, or a congenital vascular malformation such as a vascular ring or an anomalous innominate artery may lead to compressive obstruction of the trachea. When the compressing lesion is excised or displaced so that pressure is no longer exerted, malacia may appear because the rings have been so thinned. Transitory intubation is sometimes effective treatment. Other techniques, such as splinting with local prosthetic rings placed external to the lumen or with traction sutures pulled out over buttons placed against the cervical musculature, have been employed. The problems are rare, vary greatly, and require individual solutions. T-tubes or inlying stents may be useful in some.

Two types of tracheal collapse are also seen with chronic obstructive pulmonary disease. One type consists of a softening and flattening of the trachea from front to back. When the patient coughs, the membranous wall becomes approximated to the cartilaginous wall with obstruction of the airway, particularly in the lower half of the trachea. The second deformity, known as ‘sabre-sheath trachea', which consists of an increase in the anteroposterior dimension of the trachea with side-to-side narrowing, is less common. This may, however, produce such extreme narrowing that coughing and approximation of the rigid lateral walls of the trachea cause appreciable obstruction, especially to clearance of secretions. Sabre-sheath trachea affects the lower two-thirds.

The various types of surgical procedures that have been devised to correct the first, or C-type, flattening consist of pulling the corners of the cartilages together to shorten the membranous wall. When the patient coughs or breathes forcefully, it is then not possible for the trachea to flatten out.

Relapsing polychondritis is a disease of unknown etiology that leads to destruction of cartilage in many parts of the body, including the nasal septum, ears, trachea, and bronchi. The airways soften gradually. The patient is subject not only to difficulties of moving air, but to recurrent infections. The bronchi are involved. In the upper airway, inflammatory thickening is more common. No effective surgical procedure has yet been devised.

Miscellaneous obstructive lesions intrinsic to the airways

Idiopathic airway stenosis

Stenosis of the airway occasionally occurs in a patient with no history of trauma, infection, inhalation injury, intubation, or ventilation. Gastroesophageal reflux is not prominent. Biopsy examination discloses no specific organisms or diagnostic characteristics, acute and chronic inflammation, and dense, collagenous fibrosis. These patients rarely have a history of systemic disease and do not subsequently develop systemic disease. They do not have mediastinal fibrosis or processes involving mediastinal lymph nodes. The primary manifestation is upper airway obstruction. The lesions are located principally in the subglottic larynx and upper trachea, less commonly in the upper trachea alone. Women of widely ranging age are principally affected.

When the lesions do not severely compromise the airway immediately beneath the vocal cords, meticulous laryngotracheal resection and reconstruction by the techniques described earlier have given very good results: 35 patients so treated produced 32 good or excellent results, two needed further dilations and one failed, requiring tracheostomy. In only very few have the lesions progressed later. In other patients, periodic dilation is offered. Stents are not well tolerated just below the vocal cords.

Miscellaneous lesions

Sarcoidosis may, on occasion, cause obstruction of the lower trachea and main bronchi due to a combination of massively enlarged lymph nodes and fibrotic changes in the tracheal wall itself. The circumferential stenosis so produced responds to dilatation. Resection and reconstruction are not generally feasible because of the length of airway involved. Amyloid disease may cause a similar extensive process in the airways that is not amenable to surgical reconstruction. Wegener granulomatosis may produce tracheal stenosis that occasionally is surgically treated if the disease appears to be stable and confined largely to the airway.

Tracheopathia osteoplastica is a very rare condition in which the cartilages become irregularly hypertrophied with nodular overgrowth. Obstruction may follow. The disease usually involves the entire trachea and main bronchi. In a very small number of patients, vertical tracheal division with T-tube stenting has been employed, with later removal of the T-tube accomplished. Laser treatment has been attempted without success.

In adult patients, as well as in children, right pneumonectomy may be followed by total shift of the mediastinum to the right, which also results in mediastinal rotation, overexpansion of the left lung and, sometimes, severe compression of the left main bronchus between the left pulmonary artery and either the descending aorta or the vertebral column. This has been called postpneumonectomy syndrome. It also follows left pneumonectomy with right aortic arch and, rarely, without. Correction has been achieved by mediastinal repositioning to a central site and filling the pneumonectomy space (with saline-filled breast prostheses) to prevent recurrence. In some cases, severe malacia has occurred and requires stenting or reconstruction.

Tracheal tumours

Occurrence and clinical presentation

Primary tracheal tumors are rare and are thus easily overlooked. About two-thirds of primary tracheal tumors are squamous-cell or adenoid cystic carcinomas (formerly called cylindroma). The remaining one-third are widely distributed in a heterogeneous group of malignant or benign lesions. A variety of secondary tumors is found in the trachea, including carcinomas of the larynx, thyroid, lung, and esophagus. Rarely, tumors may metastasize to the submucosa of the trachea or to the mediastinum, with secondary invasion of the trachea. Carcinoma of the breast and mediastinal lymphoma may thus invade the trachea. Incompletely removed neoplasms of the main bronchus, such as carcinoid tumors, also may invade the carina.

Tracheal tumors may appear insidiously. Their most common symptoms and signs are cough (37 per cent), hemoptysis (41 per cent), and signs of progressive airway obstruction, including shortness of breath on exertion (54 per cent), wheezing and stridor (35 per cent) and, less commonly, dysphagia or hoarseness (7 per cent). It is not commonly appreciated that wheezing may be a chief symptom of a tracheal tumor for a prolonged period. Standard chest radiographs usually show clear lung fields, causing the physician to assume that no organic mass lesion is present. Patients are often treated for adult-onset asthma. Hemoptysis may not be pursued aggressively in a patient with an apparently normal chest radiograph. Presentation may also take the form of unilateral or bilateral recurrent attacks of pneumonitis, which initially respond to antibiotic treatment.

Signs and symptoms vary with the type of tumor. Hemoptysis is prominent in patients with squamous-cell carcinoma and usually leads to earlier diagnosis. Hoarseness may signify advanced disease. Adenoid cystic carcinoma more often presents with wheezing or stridor as a main symptom, leading to delay in diagnosis. Only a little more than one-quarter of these patients have hemoptysis early in the course of their disease. Dyspnea, however, may be a prominent symptom. The mean duration of symptoms before diagnosis in patients with squamous-cell carcinoma of the trachea is only 4 months; in those with adenoid cystic carcinoma, it is 18 months. In some benign or low-grade malignant tumors of the trachea, the mean duration of an incorrect diagnosis is up to 4 years. The mean duration of symptoms of miscellaneous malignant tumors is 11 months.

Diagnosis

Endoscopy is frequently the means by which a tracheal tumor is discovered in a patient who is being studied for hemoptysis of unknown origin. A high tracheal tumor may be overlooked if a flexible endoscope is passed through a previously introduced endotracheal tube or if the endoscopist is not in the habit of looking carefully at the proximal portion of the trachea. The same hazard faces the endoscopist who uses a rigid bronchoscope. When a lesion is not obstructing or is of such radiological extent that a surgical approach seems indicated in any case, endoscopy is deferred to the time of potential resection. However, when the surgical team is not trained or experienced in the management of tracheal tumors, preliminary bronchoscopy may be done to visualize the tumor. Biopsy specimens must be obtained with care to avoid stimulation of bleeding in an excessively vascular tumor. Hemorrhage in a patient with a very vascular carcinoid tumor, for example, may be life-endangering or require emergency surgical treatment. Biopsy of the rare hemangiomatous lesion of the trachea can be lethal. If preliminary biopsy examination is not undertaken before surgery, facilities for accurate diagnosis of frozen sections must be available. Biopsy specimens may need to be taken at a distance from the tumor to determine resectability; this is particularly true for adenoid cystic carcinoma, which is notorious for submucosal spread. Endoscopic examination of the esophagus is also

appropriate in patients with extensive tumors.

Pathological features

Of 198 patients with primary tumors of the trachea treated at the Massachusetts General Hospital between 1962 and 1989, 70 had squamous-cell carcinoma, 80 had adenoid cystic carcinoma, and the rest were a mixed group. The mean age of the patients with squamous-cell carcinoma was 58 years, in contrast to 43 years for those with adenoid cystic carcinoma, among whom the age spread was also much wider.

Squamous-cell carcinoma may be exophytic or ulcerative; it may also occur as multiple lesions scattered over a considerable distance in the trachea. The tumor metastasizes to the regional lymph nodes and, in its more aggressive and late forms, invades mediastinal structures. In general, its progress is rapid in comparison to that of adenoid cystic carcinoma. Many of these patients develop a second squamous-cell carcinoma of the lung or oropharynx.

Adenoid cystic carcinoma often has a very prolonged course of clinical symptoms, sometimes extending for years. Recurrence may be seen many years after initial treatment. The tumor may extend over long distances submucosally in the airways, and also perineurally. It spreads to regional lymph nodes, although this is more characteristic of squamous-cell carcinoma. Although adenoid cystic carcinoma may invade the thyroid gland or the muscular coats of the esophagus by contiguity, one that has not been surgically interfered with frequently displaces mediastinal structures before actually invading them. Metastases to the lungs are common and may grow very slowly over a period of many years, remaining asymptomatic until they reach a large size. Metastases to bone and other organs occur.

Among the other malignant lesions seen in the trachea in the above series of 198 patients were 10 carcinoid tumors clearly originating in the trachea and not in the main bronchi, two spindle-cell sarcomas, one adenocarcinoma, one adenosquamous carcinoma, four mucoepidermoid carcinomas, one chondrosarcoma, one carcinosarcoma, one small-cell carcinoma, one primary melanoma, and one malignant fibrous histiocytoma.

The benign lesions consisted of neurofibroma, chondroma, chondroblastoma, leiomyoma, granular-cell tumor, paraganglioma, hemangioma, and pleomorphic adenoma. Several patients were seen with varieties of squamous papillomas, including solitary papillomas and papillomatosis, either widespread or of a confluent, often verrucous, type.

Secondary tumors involving the trachea were seen in 81 patients. Esophageal carcinoma (seen in 10 patients) may cause a fistula between the esophagus and the trachea or the left main bronchus. Aggressive carcinoma of the lung (29 patients) also occasionally results in fistula formation, as both trachea and esophagus are involved from the mediastinum. Some of the reported oat-cell carcinomas of the trachea have probably arisen in the lung and invaded the trachea.

Papillary and follicular carcinomas of the thyroid gland and mixed varieties of the two carcinomas invade the trachea primarily (38 patients), usually at the level of the isthmus. A patient presenting initially with hemoptysis may therefore have carcinoma of the thyroid gland. Invasion of the trachea by thyroid carcinoma is best managed by tracheal resection with airway reconstruction. Localized extension of tumor may require partial esophageal resection or radical resection, including laryngectomy with mediastinal tracheostomy. More commonly, invasion is seen following prior thyroidectomy for carcinoma in which the surgeon was aware of shaving off the tumor from the trachea. In such cases, concurrent or early resection of the involved trachea should be considered.

Bronchogenic carcinoma invading the proximal main bronchus or carina, most often squamous, may be resected by carinal pneumonectomy, most often on the right, when mediastinoscopy is negative for lymph-node metastases. When N2 nodes are involved, neoadjuvant treatment may be offered on protocol. Five-year survival varies from 20 to 40 per cent. A frighteningly high and unexplained incidence of postpneumonectomy pulmonary edema (adult respiratory distress syndrome), with extremely high mortality, has plagued such resections and properly slowed their wide application.

Treatment

When the primary tracheal tumor is circumscribed, has not metastasized remotely, does not involve an excessive length of trachea, and has not invaded the mediastinum deeply, the best primary treatment is resection with primary reconstruction of the airway. Considerable experience is required to make the judgement of whether a tumor can be resected safely with sufficient tissue to provide a potentially curative margin yet allow primary reconstruction of the airway. This judgement is even more crucial when the tumor lies in the lower portion of the trachea or at the carinal level, and when an airway has to be reconstructed finally at the time of the original surgery. The disease is particularly difficult to manage in patients with adenoid cystic carcinoma in whom frozen sections show microscopic tumor at apparently clear resection margins.

Both squamous-cell and adenoid cystic carcinoma of the trachea are usually responsive to irradiation, with varying long-term results. In general, irradiation of squamous-cell carcinoma produces variable palliation, generally extending for not much longer than a couple of years and with ultimate recurrence. Adenoid cystic carcinoma may respond for 3 to 7 years. Although some investigators have advised preoperative irradiation, particularly in the management of adenoid cystic carcinoma, many prefer to reserve radiation for postoperative treatment, particularly in patients with involved lymph nodes, or with microscopic tumor in lymphatics or nerve sheaths or at the resection margin, and when the margins appear to be too small. The same approach has been applied to other malignant primary tracheal tumors.

Experience in the management of tracheal tumors is not large enough to enable definitive statements to be made about optimal treatment. Results to date, however, strongly indicate that the above approach is soundly based. A comparison of results obtained in patients with squamous-cell and adenoid cystic carcinoma treated by the vigorous protocols described with those of patients treated before there was extirpational surgery with modern techniques of resection and reconstruction shows that remarkable progress has been achieved. Adenoid cystic carcinoma, however, may recur even 15 or 20 years after apparent cure. Excellent long-term results have been obtained in patients with benign tumors amenable to excision and also in those with low-grade malignant tumors of other varieties. Cure has rarely been achieved following resection of recurrent tumors, except for local recurrences of carcinoid tumors at the carina. Long-term palliation, however, has been achieved in patients in whom differentiated thyroid neoplasms have invaded the trachea. In a number of these patients, the ultimate cause of death is the appearance of metastases in the bones or at other remote sites, rather than locally.

When the larynx is extensively involved by tumor such as adenoid cystic carcinoma, it is not always possible to salvage the organ; laryngotracheal resection is required. Conversely, when only a portion of the larynx is involved, it is possible to salvage a functioning larynx with reconstruction of the airway.

Approximately one-third of all tumors are clearly incurable and not amenable to surgical resection. Two-thirds are amenable to primary resection with reconstruction of the airway, approximately three-quarters of the adenoid cystic and two-thirds of the squamous cancers. In the rare patient with an extensive tumor affecting a large proportion of the trachea but which has not spread distantly or invaded the mediastinum, resection and reconstruction either by staged procedures or by the use of a prosthetic device may be justified. Such patients usually have adenoid cystic carcinoma. Squamous-cell carcinomas extending over a long length of trachea too often invade the mediastinum beyond the point of possible extirpation.

When extirpation is not possible, because of the extent of the tumor or because of the age or medical condition of the patient, primary irradiation is reasonable.

Some patients with adenoid cystic carcinoma obstructing the trachea and pulmonary metastases benefit from palliative resection if reconstruction can be performed safely.

When the trachea is obstructed acutely by a tumor that is too large to be removed by primary resection, immediate palliation may be achieved by the removal of the bulk of the tumor endotracheally, followed by radiotherapy. Tumor may be cored out with a rigid bronchoscope and biopsy forceps, or a laser may be used. The laser is not applicable as a primary method of treatment of most tracheal tumors, however, since it cannot destroy the base of the tumor without also destroying the tracheal wall. It does have a role in the management of multiple squamous papillomas of the trachea.

Ultimately, as a tumor recurs and laser therapy is no longer possible because of extension through the tracheal wall, additional palliation to prevent strangulation can sometimes be given by the judicious insertion of a silicone rubber T-tube or expandable stents that span the airway and allow the patient to breathe.

When technically feasible, a trachea invaded by carcinoma of the thyroid gland should be resected and primarily reconstructed when no extensive metastases are present. In recurrence of differentiated carcinoma, resection of the involved airway and reconstruction have provided excellent, prolonged palliation, with prevention of airway obstruction and bleeding in most cases. Resection when the invasion is first identified appears to promise cure, although the notoriously prolonged course of papillary carcinoma signals caution in prediction. In carefully selected patients with massive regional disease, radical excision with laryngectomy and esophagectomy is also appropriate.

Surgical resection of tracheal tumours

Surgical resection of tracheal tumors is often made difficult by the unpredictability of the extent of a lesion.

For many years, tracheal resection was limited by the belief that only 2 cm of the trachea (about four rings) could be removed and the ends dependably anastomosed by primary suture. Various techniques of anatomical mobilization now permit the dependable and predictable resection of approximately one-half of the trachea. Simple cervical flexion, which delivers the cervical trachea into the mediastinum, is the most useful, single maneuver for extending the resection with primary repair. In a young person who is not obese and who has reasonably supple tissues, more than one-half of the trachea can be removed with primary reconstruction. With increasing age, kyphosis, obesity, and pathological changes, the portion of the trachea that can be so removed and reconstructed is reduced. If additional length is necessary, tracheal release may be employed.

In all dissections of the trachea, it is critical to preserve the lateral segmental blood supply, and to ensure gentle and precise handling of all tissues and precision of anastomosis.

Tumors of the upper portion of the trachea are generally approachedthrough a collar incision with, if necessary, a vertical extension through the upper sternum (Fig. 6). Since the extent of some tumors is not fully predictable even after preoperative radiography and bronchoscopy, it is generally wise to position a patient so that the sternal division can be carried down further and angled into the right fourth interspace to add a thoracotomy to the cervical and mediastinal exposure. Midtracheal tumors may be approached by median sternotomy, between the superior vena cava and aorta, performing anterior and posterior pericardiotomy at that level. Tumors of the lower portion of the trachea are approached most easily through a posterolateral thoracotomy. Laryngeal release adds no additional length for distal tracheal resection. Flexing the neck and freeing the anterior pretracheal plane are the most useful means of gaining additional length, as is intrapericardial release of the pulmonary vessels.

When a tumor involves the carina, various reconstructive techniques are used. Unless the tumor is very small, it is rarely adaptable to reconstruction by approximating the right and left main bronchus to form a new carina and then attaching it to the trachea. Suturing in this fashion anchors the carina low in the mediastinum. If more trachea has been excised, approximation is not possible. More commonly, either the right or left main bronchus is sutured to the trachea, and a lateral anastomosis of the other bronchus to the lower portion of the tracheal wall is performed above the initial anastomosis (Fig. 10).

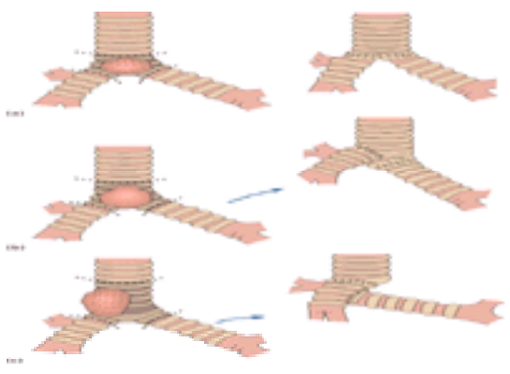


Fig. 10. Carinal excision and reconstruction. (a) Reconstruction of a neocarina after resection of a short length of trachea and bronchi. (b) Greater length of resection allows anastomosis of trachea to left main bronchus, where distance is under 4 cm. The aortic arch limits movement of left main bronchus. The right is advanced to the side of the trachea. (c) With even greater length of resection, anastomosis is performed between trachea and right main bronchus, after intrapericardial release. The left main bronchus is anastomosed to the medial wall of bronchus intermedius.

If a recurrent laryngeal nerve is involved by tumor it is sacrificed, although these nerves are usually identified and carefully saved when possible. Local paratracheal lymph nodes are excised with the specimen when possible: extensive lymph-node dissection is likely to destroy the blood supply to the residual portion of the trachea. Partial removal of the lower part of the larynx may be required for tumors high in the trachea. Individually designed procedures are necessary to preserve a functional larynx: portions of the esophagus and other adjacent structures may need to be resected (Fig. 11).

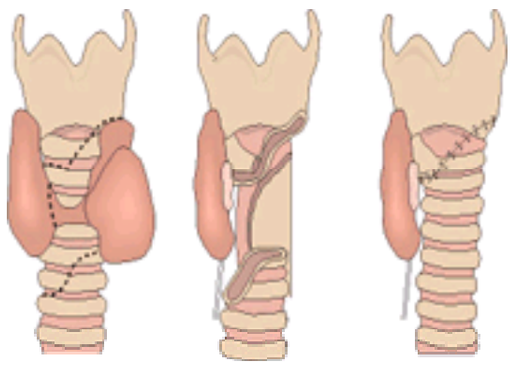


Fig. 11. Laryngotracheal resection for invasive papillary carcinoma of thyroid, with left recurrent nerve paralysis and invasion of muscular wall of esophagus. The trachea is tailored to fit the laryngeal defect.

Resection is usually controlled by immediate examination of frozen-section specimens in the laboratory to ensure clear margins. Adenoid cystic carcinoma, in particular, may be so extensive as to make resection of all microscopic disease impossible. Postoperative irradiation is then required.

Complications of tracheal reconstruction

Laryngeal edema is managed by restricting fluid intake and administering racemic epinephrine (adrenaline) and a short course of steroids (24–48 h). The edema usually regresses within a week. Pneumonia is rare after upper tracheal resections when proper attention is given to intraoperative management and to postoperative physiotherapy. All patients spend 1 day or more in a respiratory intensive-care unit whose staff are familiar with the management of such problems.

The most common late complication was the formation of granulations at the suture line (Table 1). Granulations are less of a problem in patients undergoing resections for tumor than in those who have had tracheal reconstructions for inflammatory disease, in whom residual inflammation may be present. Granulations can usually be managed by bronchoscopic removal under light anesthesia. Use of fine, absorbable Vicryl sutures (4/0 in adults, 5/0 in children) has all but eliminated this problem.

Condition	Intubation	Neoplasms
Granulations	28	10
Separation	4	6
Air leak only		1
Stenosis:		
Partial	6	3
Complete	15	
Hemorrhage	2	1
Persistent stoma	5	
Tracheo-esophageal fistula	1	
Esophagocutaneous fistula		1
Wound infection	6	
Vocal cord dysfunction	5	3
Aspiration	1	
Hypoxemia		1
Laryngeal edema	1	
Respiratory failure		2
Pneumonia		2

Table 1 Complications of tracheal reconstruction based on the indications for reconstruction at the Massachusetts General Hospital

Separation of the anastomosis is usually due to excessive tension following resection of too much trachea or when adjunctive relaxing maneuvers to lessen the tension have not been made. Excessive circumferential dissection of the trachea, particularly distal to the point of division, may destroy the blood supply and cause separation or stenosis: this is more likely to occur in patients with tumors than in those with intubation stenosis.

Tracheal separation occurring in the immediate postoperative phase suggests that there has been a serious technical error. Reoperation might be considered, to resuture the area and cover with a local muscle flap is the area was small. If the tissues do not seem appropriate for resuturing, a tracheostomy tube may be placed across the defect, to be replaced later by a Montgomery silicone T-tube. A T-tube can be placed initially if a patient does not require a sealed airway. Re-resection is often successfully feasible after a period of 4 to 6 months to allow inflammation to regress.

Results of the management of complications are listed in Table 2. Good results following complications are possible if handled properly.

	No.	Good	Satisfactory	Failed	Death
Granulations	28	24	4		
Separation	4		2		2
Restenosis	21	6	15		
Malacia	3	1		1	1
Hemorrhage	2	1			1
Tracheo-esophageal fistula	1				1
Vocal cord dysfunction	5		4	1	
Aspiration	1			1	
Wound infection	6	6			
Edema	1			1	

Table 2 The results of treatment of the benign postintubation complications of tracheal reconstruction at the Massachusetts General Hospital

Further reading

Andrews MJ, Pearson FG. An analysis of 59 cases of tracheal stenosis following tracheostomy with cuffed tube and assisted ventilation, with special reference to diagnosis and treatment. *British Journal of Surgery* 1973; **60**: 208–12.

Cooper JD, Grillo HC. The evolution of tracheal injury due to ventilatory assistance through cuffed tubes. A pathologic study. *Annals of Surgery* 1969; **169**: 334.

Couraud L, Brichon PY, Velly JF. The surgical treatment of inflammatory and fibrous laryngotracheal stenosis. *European Journal of Cardiothoracic Surgery* 1988; **2**: 410–15.

Couraud L, Velly JF, Martigne C, N'Diaye M. Posttraumatic disruption of the laryngo-tracheal junction. *European Journal of Cardiothoracic Surgery* 1989; **3**: 441–6.

Dartevelle PG *et al.* Tracheal sleeve pneumonectomy for bronchogenic carcinoma: report of 55 cases. *Annals of Thoracic Surgery* 1988; **46**: 68–72.

Donahoe PK, Gee PE. Complete laryngotracheoesophageal cleft: management and repair. *Journal of Pediatric Surgery* 1984; **19**: 143–8.

Donahue DM, Grillo HC, Wain JC, Wright CD, Mathisen DJ. Reoperative tracheal resection and reconstruction for unsuccessful repair of post intubation stenosis. *Journal of Thoracic and Cardiovascular Surgery* 1997; **114**: 934–9.

Gaissert HA, Grillo HC, Mathisen DJ, Wain JC. Temporary and permanent restoration of airway patency with the tracheal T-tube. *Journal of Thoracic and Cardiovascular Surgery* 1994; **107**: 600–6.

Gaissert HA, Lofgren RH, Grillo HC. Upper airway compromise after inhalation injury: complex strictures of larynx and trachea and their management. *Annals of Surgery* 1993; **218**: 672–8.

Grillo HC. Carinal reconstruction. *Annals of Thoracic Surgery* 1982; **34**: 356–73.

Grillo HC. Slide tracheoplasty for long segment congenital tracheal stenosis. *Annals of Thoracic Surgery* 1994; **57**: 613–21.

Grillo HC, Mathisen DJ. Cervical exenteration. *Annals of Thoracic Surgery* 1990; **49**: 401–9.

Grillo HC, Mathisen DJ. Primary tracheal tumors: treatment and results. *Annals of Thoracic Surgery* 1990; **49**: 69–77.

Grillo HC, Mark EJ, Mathisen DJ, Wain JC. Idiopathic laryngo tracheal stenosis: the entity and its management. *Annals of Thoracic Surgery* 1993; **56**: 80–7.

Grillo HC, Mathisen DJ, Wain JC. Laryngotracheal resection and reconstruction for subglottic stenosis. *Annals of Thoracic Surgery* 1992; **53**: 54–63.

Grillo HC, Moncure AC, McEnany MT. Repair of inflammatory tracheoesophageal fistula. *Annals of Thoracic Surgery* 1976; **22**: 112–17.

Grillo HC, Shepard JO, Mathisen DJ, Kanarek DJ. Postpneumonectomy syndrome: diagnosis, management and results. *Annals of Thoracic Surgery* 1992; **54**: 638–51.

Grillo HC, Suen HC, Mathisen DJ, Wain JC. Resectional management of thyroid carcinoma invading the airway. *Annals of Thoracic Surgery* 1992; **54**: 3–10.

Grillo HC, Zannini P, Michelassi F. Complications of tracheal reconstruction. *Journal of Thoracic and Cardiovascular Surgery* 1986; **91**: 322–8.

Grillo HC *et al.* Postintubation tracheal stenosis: treatment and results. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 486–93.

Herzog H, Heitz M, Keller R, Graedel E. Surgical therapy for expiratory collapse of the trachea and large bronchi. In: Grillo HC, Eschapasse H, ed. *International trends in general thoracic surgery: major challenges*, pp. 74–90. Saunders, Philadelphia, 1987.

Lawson DW, Grillo HC. Closure of a persistent tracheal stoma. *Surgery, Gynecology and Obstetrics* 1970; **130**: 995–6.

Maddaus MA, Toth JLR, Gullane PJ, Pearson FG. Subglottic tracheal resection and synchronous laryngeal reconstruction. *Journal of Thoracic and Cardiovascular Surgery* 1992; **104**: 1443–50.

Mathisen DJ, Grillo HC. Laryngotracheal trauma. *Annals of Thoracic Surgery* 1987; **43**: 254–63.

Mathisen DJ, Grillo HC. Endoscopic relief of malignant airway obstruction. *Annals of Thoracic Surgery* 1989; **48**: 469–75.

Mathisen DJ, Grillo HC. Carinal resection for bronchogenic carcinoma. *Journal of Thoracic and Cardiovascular Surgery* 1991; **102**: 16–23.

Mathisen DJ, Grillo HC. Clinical manifestations of mediastinal fibrosis and histoplasmosis. *Annals of Thoracic Surgery* 1992; **54**: 1053–8.

Mathisen DJ, Grillo HC, Wain JC, Hilgenberg AD. Management of acquired nonmalignant tracheoesophageal fistula. *Annals of Thoracic Surgery* 1991; **52**: 759–65.

Maziak DE, Todd TRJ, Keshavjee SH, Winton TL, Van Nostrand P, Pearson FG. Adenoid cystic carcinoma of the airway: 32 year experience. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 1522–32.

Mitchell JD *et al.* Clinical experience with carinal resection. *Journal of Thoracic and Cardiovascular Surgery* (in press).

Montgomery WW. The surgical management of supraglottic and subglottic stenosis. *Annals of Otolaryngology* 1968; **77**: 534–8.

Montgomery WW. Suprahyoid release for tracheal anastomosis. *Archives of Otolaryngology* 1974; **99**: 255–60.

Perelman MI, Koroleva NS. Primary tumors of the trachea. In: Grillo HC, Eschapasse H, ed. *International trends in general thoracic surgery*, Vol. 2, pp. 91–108. Saunders, Philadelphia, 1987.

Salassa JR, Pearson BW, Payne WS. Gross and microscopical blood supply of the trachea. *Annals of Thoracic Surgery* 1977; **24**: 100–7.

Weber AL. Radiologic evaluation of the trachea. *Surgical Clinics of North America* 1996; **6**: 637–73.

Wilson RS. Anesthetic management for tracheal reconstruction. In: Grillo HC, Eschapasse H, ed. *International trends in general thoracic surgery*, Vol. 2, pp. 3–12. Saunders, Philadelphia, 1987.

41.3 Benign and malignant tumours of the lung

John C. Wain

[Introduction](#)
[Neoplastic transformation and histogenesis](#)
- [Environmental factors](#)
- [Genetic factors](#)
- [Histogenesis](#)

[Specific tumor types](#)
- [Malignant epithelial tumors of the lung](#)
- [Carcinosarcoma of the lung](#)
- [Benign mesenchymal tumors of the lung](#)
- [Primary sarcoma of the lung](#)
- [Lymphoproliferative disorders of the lung](#)

[Diagnosis of lung neoplasms](#)
- [Symptoms](#)
- [Chest radiographs](#)
- [Computed tomographic (CT) scans](#)
- [Magnetic resonance imaging (MRI)](#)
- [Positron emission tomography (PET) scans](#)
- [Sputum cytology](#)
- [Bronchoscopy](#)
- [Percutaneous needle biopsy](#)
- [Excisional biopsy and other procedures](#)

[Staging](#)
- [Staging classification](#)
- [Non-invasive staging procedures](#)
- [Invasive staging procedures](#)

[Preoperative evaluation](#)
[Surgical principles and techniques](#)
- [Thoracic incisions](#)
- [Intraoperative evaluation](#)
- [Principles of pulmonary resection](#)
- [Special issues](#)

[Further reading](#)

Introduction

Benign and malignant tumors of the lung are the most common intrathoracic neoplasms. They are the most common solid tumors in North America and the United Kingdom among adults of either sex. Over 180 000 patients were diagnosed with pulmonary neoplasms in the United States in 1997. Malignant lesions, of varying metastatic potential, account for over 95 per cent of these tumors. Lung cancers are the most common fatal malignancy of the developed world. The incidence and mortality of lung cancer have risen dramatically during the twentieth century—in part due to increased life expectancy, but primarily due to increased environmental exposure to carcinogenic agents (e.g. tobacco products).

These tumors are the most common indication for pulmonary resection. For benign lesions, surgical resection provides the potential for cure in nearly all cases. For malignant lung tumors, the majority of which have metastasized at the time of initial presentation, surgery can provide important staging information and offers the greatest likelihood of cure for neoplasms that are localized to the pulmonary parenchyma.

Neoplastic transformation and histogenesis

The complexity of neoplastic transformation has been increasingly appreciated due to advances in molecular biology and molecular genetics. The importance of genetic factors in the host response to carcinogenic stimuli, the role of tumor suppressor genes, and the spectrum of neoplastic behavior related to specific oncogene expression have been incorporated into the proposed scheme of transformation.

Exposure to an environmental agent results in an internal dose of the agent, which may have a direct effect, as in the case of ionizing radiation, or more commonly, may result in activation of the agent to a carcinogen. The net effective biologic dose, determined by the internal dose offset by the rate of inactivation or excretion of carcinogen, may elicit early intracellular events. At least 10 to 20 genetic mutations are required to produce a lung cancer cell, including inactivation of tumor suppressor genes (e.g. *p53*) or expression of dominant oncogenes as well as a progressive abnormality in apoptosis, the cell's intrinsic cell death program. Ultimately, changes in structure and function can be identified histologically as the neoplastic phenotype. Carcinogens or other agents may have additive effects during the transformation process (Fig. 1).

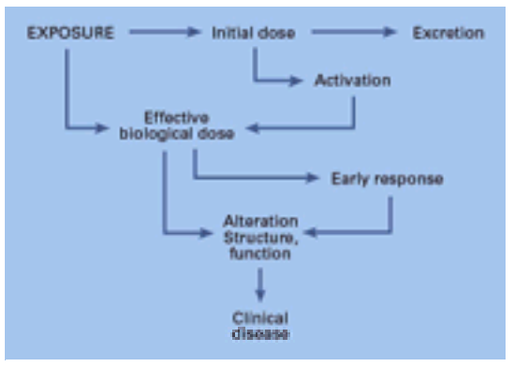


Fig. 1. Current scheme of neoplastic transformation.

Environmental factors (see also Chapter 48.2)

Environmental agents associated with malignant lung neoplasms include tobacco, radon, asbestos, arsenic, chloromethyl ethers, and nickel. Exposure to tobacco products is the most common environmental agent linked to lung cancer. Conclusive evidence of the association of cigarette smoking with an increased risk of lung cancer in men has been available from case–control studies dating from the 1950s and 1960s. Newer evidence continues to support a causal relationship between smoking and lung cancer in females and those involuntarily exposed to cigarette smoke.

Radon, an inert gas formed from radium during the natural decay of uranium, is the second most important cause of lung cancer. Radon is invariably present in indoor air. It diffuses from the soil into enclosed spaces and decomposes into short-lived products that emit a-particles. Radon exposure is causally associated with lung cancer in both uranium miners and in those who smoke cigarettes. Radon is commonplace in houses, with some houses having levels close to that of uranium mines, with an attendant lung cancer risk. Most houses, however, have levels acceptable by current guidelines. The lung cancer risk for non-smokers in these environments is unknown.

Occupational exposure to asbestos, arsenic, chloromethyl ethers, and nickel all cause lung cancer in non-smokers, and have an additive risk for people who smoke. Finally, ambient air pollution containing polycyclic hydrocarbons from fossil fuel combustion appears to pose an independent risk for lung cancer, accounting for 1 to 2

per cent of lung cancer cases.

Genetic factors (see also Chapter 48.1)

Genetic predisposition to lung cancer has been suggested by the identification of a fourfold increased risk for lung cancer among the non-smoking relatives of people who smoke and have lung cancer, compared with non-smokers without relatives with lung cancer. Other studies based on genetic models and the chronologic causal relationship of lung cancer to tobacco imply that virtually all lung cancers occur among susceptible gene carriers. Phenotypic differences in carcinogen metabolism (e.g. aryl hydrocarbon hydroxylase, debrisoquine hydroxylase) between patients with lung cancer compared with matched controls suggest one potential mechanism for a genetic predisposition to lung cancer. Identification of these factors could have a profound impact on detection and prevention of lung cancer.

Multiple genetic abnormalities characterize malignant lung neoplasms. Complex karyotypic changes are seen with greater frequency as neoplastic progression occurs. A characteristic deletion of the short arm of chromosome 3 (3p) occurs in 90 to 100 per cent of small cell lung cancers and in 25 to 50 per cent of non-small cell lung cancers. Abnormal tumor suppressor genes include the retinoblastoma gene (95 per cent in small cell lung cancer, 20 per cent in non-small cell lung cancer) and p_{53} (100 per cent in small cell lung cancer, 40 to 50 per cent in non-small cell lung cancer). Dominant oncogene abnormalities include K-*ras* and *erb* B1 in non-small cell lung cancer and *myc* and *bcl*-2 in small cell lung cancer. Mutations in K-*ras* and *myc* are associated with a decreased survival rate for their respective histologic subtypes. *bcl*-2, which prevents apoptosis, is associated with resistance to chemotherapy.

Histogenesis

All the epithelial cells of the lung derive from the endoderm, with subsequent differentiation adapted to different levels of the respiratory tract. After initial growth, the reserve cells of the epithelium are only able to replace dead cells and are looped in a terminal differentiation sequence adapted to local needs. The bronchial epithelial cells have a limited range of initial response to a variety of agents, carcinogenic or not, including basal cell hyperplasia with differentiation toward squamous or goblet (mucus) cells, or proliferation of Kulchitzky (neuroendocrine or K-type) cells of type II pneumocytes. In the absence of carcinogenesis, most of these changes are reversible. During carcinogenesis, differentiation is switched on again and is executed aberrantly by the process of neoplastic transformation described previously. Heterogeneity of the tumor cell population and clonal selection are important aspects of growth for malignant lung tumors.

For bronchogenic carcinomas, both small cell lung cancer and non-small cell lung cancer, the cell of origin, although unknown, seems to be the same. To a certain extent, endstage malignant tumors of the lung come to resemble each other histologically (Fig. 2). For less malignant lung neoplasms, such as carcinoid tumors and bronchial gland tumors, the cell of origin probably derives from the epithelium of airways or bronchial glands, respectively. The precise cell of origin for pulmonary hamartoma, the most common benign lung tumor, is unknown.

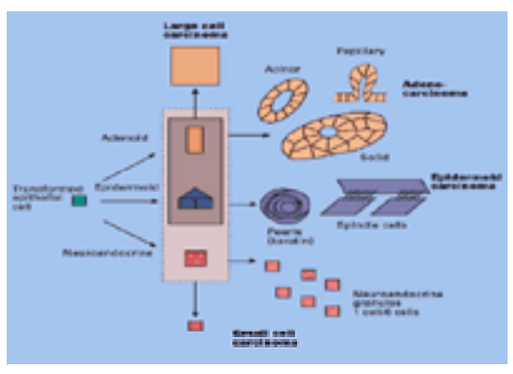


Fig. 2. The scheme of differentiation of transformed respiratory epithelial cells resulting in bronchogenic carcinoma. A reserve epithelial cell follows one of a limited number of paths of differentiation in the process of tumor evolution, resulting in the histologically classified cancer.

Specific tumor types

Tumors of the lung may be divided into: malignant epithelial tumors, including both small cell lung cancer and non-small cell lung cancer, carcinoid, and bronchial gland carcinomas; benign mesenchymal tumors, including pulmonary hamartomas; carcinosarcomas; primary sarcomas; and lymphoproliferative lesions (Table 1). The malignant epithelial group comprises 98 per cent of lung neoplasms and accounts for the majority of surgical interventions for these neoplasms.

Malignant epithelial tumors of the lung
Small cell carcinoma (neuroendocrine carcinoma)
Squamous cell carcinoma
Adenocarcinoma
Large cell carcinoma
Adenosquamous carcinoma
Carcinoid tumors
Bronchial gland carcinomas
Adenoid cystic carcinoma
Mucoepidermoid carcinoma
Miscellaneous
Carcinosarcoma of the lung
Benign mesenchymal tumors
Hamartoma
Bronchial chondroma
Granular cell tumors
Primary sarcoma of the lung
Lymphoproliferative disorders of the lung
Pseudolymphoma
Lymphocytic lymphoma

Table 1 Specific tumor types

Malignant epithelial tumors of the lung

Small cell carcinoma

The most rapidly progressive form of malignant epithelial tumor is small cell lung cancer, also termed oat cell carcinoma, which accounts for 15 to 25 per cent of all lung cancers. Biologically, small cell lung cancer has unique neuroendocrine properties and produces a wide variety of peptide hormones, including both those involved in control of cellular growth (e.g. bombesin-like peptide) and those involved in paraneoplastic syndromes (e.g. ACTH, ADH, serotonin, calcitonin). Small cell carcinoma is composed of small, uniform cells with scanty cytoplasm and nuclei containing finely stippled chromatin and inconspicuous nucleoli (Fig. 3). Foci of squamous or glandular differentiation may be seen within the larger tumor mass. Although attempts to subdivide the small cell carcinomas on the basis of nuclear morphology have been made, they are of no practical value. Ultrastructurally, membrane-bound cytoplasmic granules are found in one of every six cells. These granules, along with amine precursor uptake and decarboxylation (APUD) capabilities, identify the neuroendocrine differentiation of small cell lung cancer.

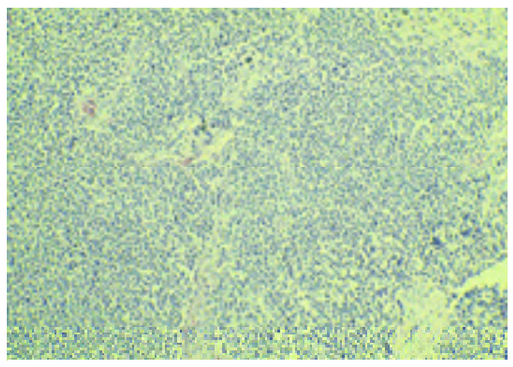


Fig. 3. Small cell carcinoma of the lung. The tumor is composed of small, uniform cells with scanty cytoplasm and inconspicuous nucleoli.

Small cell lung cancer occurs in an older-aged population, with a median age of 60 years and is more common among women than men who smoke cigarettes. Among patients less than 40 years old with lung cancer, small cell lung cancer is the most common histologic type. Small cell lung cancer usually presents symptoms at an earlier stage than other malignant epithelial tumors because of its rapid growth (mean doubling time of 33 days), propensity for early metastasis, and ectopic hormone production. Cough, weight loss, chest pain, and hemoptysis are the most common symptoms of these lesions, which manifest themselves as large, non-cavitating hilar masses with extensive nodal or mediastinal metastases in 65 per cent of cases (Fig. 4). An endobronchial lesion is found in 80 per cent of patients at the time of initial symptoms. Dysphagia, due to esophageal compression (10 per cent), hoarseness due to recurrent laryngeal nerve invasion (15 per cent), or superior vena cava syndrome (10 per cent) are more common with small cell lung cancer than with non-small cell lung cancer. Initial symptoms produced by a metastasis to a distant organ, most commonly the central nervous system, are seen in 10 per cent of patients. Clinical paraneoplastic syndromes are seen in 15 per cent of patients with small cell lung cancer, more frequently than in those with non-small cell lung cancer. The most common syndromes in small cell lung cancer are Cushing's syndrome due to ectopic ACTH production, inappropriate secretion of ADH, carcinoid syndrome, or Eaton–Lambert syndrome. Hypercalcemia or hypertrophic pulmonary osteoarthropathy are uncommon in small cell lung cancer.

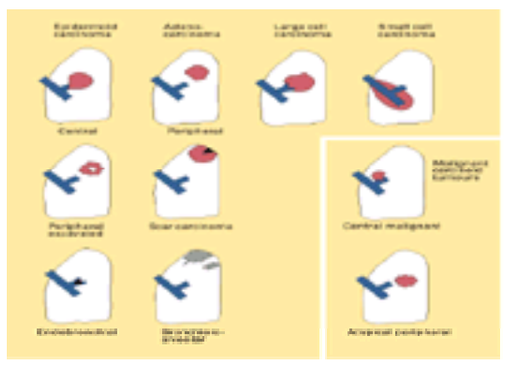


Fig. 4. A diagrammatic overview of the clinicopathologic presentation of most common types of malignant epithelial lung neoplasms.

The two most important prognostic factors for small cell lung cancer are the stage of the disease and the performance status at the time of diagnosis. Staging of small cell lung cancer has been commonly based on a two-stage system, with patients being classified as having either limited disease (tumor confined to a hemithorax, with or without pleural effusion, and with or without ipsilateral or contralateral supraclavicular or mediastinal lymph nodes) or extensive disease. An important survival difference following treatment can be identified between these two groups. However, recent data suggest that staging small cell lung cancer using the TNM system employed for non-small cell lung cancer (see below) may be useful. Although 85 per cent of small cell lung cancers are stage III or IV at initial presentation, for patients with stage I disease surgical resection and postoperative adjuvant chemotherapy provides a better prognosis than non-surgical therapy. For this reason, patients with small cell lung cancer should undergo staging procedures identical to those for non-small cell lung cancer at the time of diagnosis.

The median survival time of untreated small cell lung cancer is 3 months. For most patients with nodal metastases or extensive disease at diagnosis, combination chemotherapy is the mainstay of treatment. Multiple drugs and more aggressive regimens produce better response rates and longer duration of response (see Chapter 48.3). Concurrent radiation therapy serves to provide better control of intrathoracic disease, with a survival advantage. Therapy is typically for 12 months or less. Alternating combination chemotherapy that is not cross-resistant and the use of autologous bone marrow transplantation does not improve median survival. Complete response rates range from 20 to 50 per cent, with total response rates (complete and partial response) of 80 per cent. Median survival times range from 15 months for limited disease to 9 months for extensive disease. Two- and five-year survival rates are 29 and 12 per cent, respectively, for limited stage disease.

Radiation therapy as a primary modality, although demonstrating a better mean survival than for surgical therapy alone, does not improve median survival as compared with the use of combination chemotherapy. However, concurrent radiation with combination chemotherapy appears to provide a survival advantage. Prophylactic cranial irradiation is beneficial only for patients who demonstrate a complete remission with chemotherapy. For metastatic disease, radiation therapy provides excellent palliation of symptoms, particularly for disease metastatic to the brain.

Surgical therapy as a primary modality should be reserved for patients with stage I small cell lung cancer, that is, with no nodal metastases. Selection of such patients requires maximal use of invasive staging modalities preoperatively. With the addition of postoperative adjuvant combination chemotherapy, survival for this selected group of patients is 55 per cent at 5 years. Resection of residual pulmonary disease after combination chemotherapy does not improve survival.

Squamous cell carcinoma

The most common histologic type of malignant epithelial lung tumor is squamous cell carcinoma, a non-small cell lung cancer that accounts for 30 to 35 per cent of all lung cancers. By light microscopy, these tumors demonstrate at least one of three characteristics of differentiation: individual cell keratinization, pearl formation (the formation of small groups of concentrically oriented cells), and/or extensive intercellular bridges, similar to those noted in the prickle cell layers of the epidermis (Fig. 5). The nuclei are characteristically hyperchromatic with a jagged border. There is often a striking host reaction including both desmoplasia and acute and chronic inflammation.

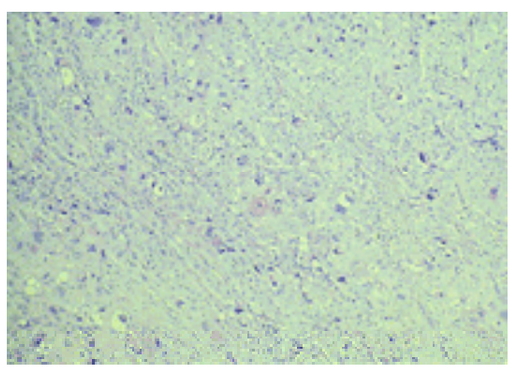


Fig. 5. Squamous cell carcinoma of the lung. A photomicrograph demonstrates individual keratin pearl formation and intracellular bridging.

Squamous cell cancers arise in the bronchial epithelium, and progress through an *in situ* phase to a preclinical phase, extending from the time the carcinoma becomes morphologically developed until it produces symptoms or is radiographically detectable, at which time it enters a clinical phase. The preclinical phase may range from 10 to 24 months, during which time it may be detected fortuitously by cytologic evaluation or directed biopsy of the airway.

Although not possessing APUD capabilities, squamous cell cancers produce PTH and ACTH, as based on immunohistochemical studies, more frequently than do other types of non-small cell lung cancer. They also produce carcinoembryonic antigen in 50 per cent of cases, which may be useful for postoperative surveillance.

Typically, squamous cell lung cancer arises in older patients (peak incidence: 70 to 79 years for men and 60 to 69 years for women). Squamous cell cancer is the most common lung cancer in men; adenocarcinoma is the most common lung cancer in women. Seventy-five per cent of squamous cancers are found in larger bronchi. The remainder present as peripheral nodules. Cavitation of the lesion is a common finding (Fig. 4). Because of their central location, sputum cytology and endoscopic biopsy may detect 50 per cent of cases. Squamous cell cancers have a mean doubling time of 136 days, but have a lower likelihood for distant metastasis than do other malignant epithelial tumors of the lung. As a result, the frequency of squamous cell cancer in surgical resections is higher (45 to 70 per cent) than in other types of lung cancer.

The majority of patients present themselves with cough, hemoptysis, obstructive pneumonitis, or positive sputum cytology. Paraneoplastic hypercalcemia, due to excessive prostaglandin production, or less frequently to excessive PTH production, is associated with squamous cell cancers more commonly than with other types of either small cell or non-small cell lung cancer.

The primary treatment of squamous cell carcinoma of the lung is surgical. The prognosis of squamous cancer of the lung is highly dependent upon its biologic stage (i.e. preclinical or clinical), its degree of differentiation (poorly, moderately, or well differentiated), and its pathologic stage (TNM classification). For patients undergoing resection in the preclinical phase, the survival rate exceeds 80 per cent. The survival rate correlates directly with histologic differentiation—longer survival with more differentiated tumors. Finally, the survival time following surgical resection for squamous cell cancers is 5 to 10 per cent longer within each pathologic TNM stage than for other similarly staged non-small cell lung cancer and small cell lung cancer.

The excellent survival rate is the primary reason for a higher incidence of second primary lung cancers in these patients. Approximately 10 per cent of patients surviving beyond 5 years develop a second primary lung cancer, not necessarily a squamous type. Autopsy studies show that areas of *in situ* carcinoma distant from the primary lesion may be found in 25 to 40 per cent of patients. In addition, 15 to 20 per cent of patients develop second primary carcinomas in other foregut sites (head and neck, esophagus), supporting the concept of a ‘field defect' of carcinogenesis induced by environmental exposure, most commonly tobacco. However, two trials of b-carotene derivatives, which reverse preinvasive epithelial changes, demonstrated an increased incidence of lung cancer among patients receiving these agents.

Extrathoracic metastases occur, in order of frequency, to the liver, adrenal gland, bone, and brain. For symptomatic metastases, radiotherapy is used most commonly, although there are data to support the use of combination resection and radiation for localized cerebral metastases. *Cis*-platinum- or taxene-containing combination chemotherapy regimens are effective for patients with metastatic disease or in protocol studies of adjuvant therapy for patients with resectable advanced stage disease.

Adenocarcinoma

The second most common type of malignant epithelial tumor of the lung and the second most common type of non-small cell lung cancer is adenocarcinoma. This is a heterogeneous group of tumors, demonstrating various aspects of glandular differentiation, with or without mucus formation, which accounts for 30 to 35 per cent of all lung cancers. Adenocarcinomas range from well differentiated subtypes, with formation of acini or papillary structures, to poorly differentiated subtypes, with only occasional intracellular mucus. Typically, these lesions arise from the epithelium of distal bronchi or bronchioles. No *in situ* component, as noted for squamous cell cancers, has been identified. Frequently, there is a prominent desmoplastic reaction associated with these tumors.

A particular subtype of adenocarcinoma is bronchoalveolar carcinoma, which accounts for 10 per cent of adenocarcinomas. Bronchoalveolar carcinoma is characterized by growth into the alveoli of cytologically bland cells, resembling mucus-secreting Clara cells or type II pneumocytes. Unlike other forms of adenocarcinoma, a desmoplastic reaction is rare. Bronchoalveolar carcinomas are multifocal in 60 per cent of patients. Whether there is a true multifocal origin of this tumor or aerogenous metastasis occurs through the airways to account for the multifocality is unknown. An ovine neoplasm known as jaagsiekte (Montana progressive pneumonia), which is associated with a transmissible agent, is histologically indistinguishable from bronchoalveolar carcinoma. There is no evidence of a transmissible agent in the etiology of the disease in humans.

Adenocarcinomas of all subtypes tend to occur with a higher frequency in women than other forms of lung cancer. In addition, adenocarcinomas are the most frequent type of peripheral carcinoma, accounting for 40 to 50 per cent of such lesions, and for 70 per cent of such lesions in women. Scar formation is found in association with adenocarcinoma in 50 per cent of cases—at times as a pre-existing lesion and at times as a manifestation of the desmoplastic response. Various explanations for this association, including interruption of local lymphatics with focal concentration of carcinogenic substances and chronic stimulation of epithelial regeneration, have been proposed; none has been proved (Fig. 4).

Cough and hemoptysis are presenting symptoms in less than one-third of patients, the majority being identified as solitary, non-cavitating nodules on chest radiograph in asymptomatic patients. Bronchoalveolar carcinomas, with their bronchiolar origin, typically manifest themselves as bronchopneumonia, which progresses to a lobar consolidation by extension to surrounding airspaces. In such patients, cough may be a more prominent symptom, and sputum cytology may demonstrate characteristic cells or psammoma bodies. Intractable bronchorrhea is an unusual but characteristic presentation for multifocal bronchoalveolar carcinoma (also termed ‘adenomatosis'). Hypertrophic pulmonary osteoarthropathy occurs most frequently with adenocarcinomas of the lung than with other histologic types and may be seen in 10 per cent of patients.

Grossly, adenocarcinomas are firm; they frequently involve or distort the overlying visceral pleura from their peripheral location. Cavitation is uncommon, although larger tumors may have regions of necrosis. Histologically, the cells are large, with abundant cytoplasm, and may demonstrate apical microvilli and intracytoplasmic secretory granules by electron microscopy. Cytokeratin and carcinoembryonic antigen expression are common. At the periphery, the tumor may invade along normal lung architecture, in a bronchoalveolar pattern, which does not imply classification into the bronchoalveolar carcinoma subtype (Fig. 6). Bronchoalveolar subtypes typically have a pneumonic appearance grossly, or demonstrate confluence of multiple bronchiolocentric nodules. Growth of cytologically bland cells along the normal lung framework, without desmoplasia, is characteristic (Fig. 7).

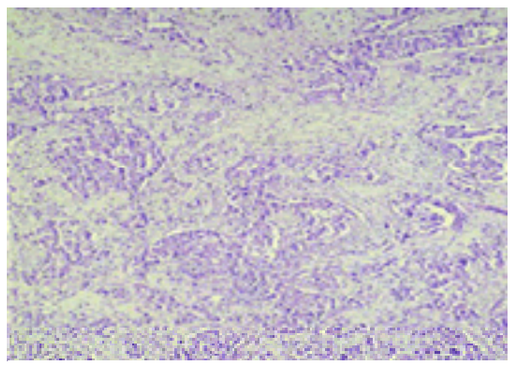


Fig. 6. Adenocarcinoma of the lung. Acini growing within a characteristic region of fibrosis (desmoplasia). At the lower margin, a pattern of bronchoalveolar growth is noted.

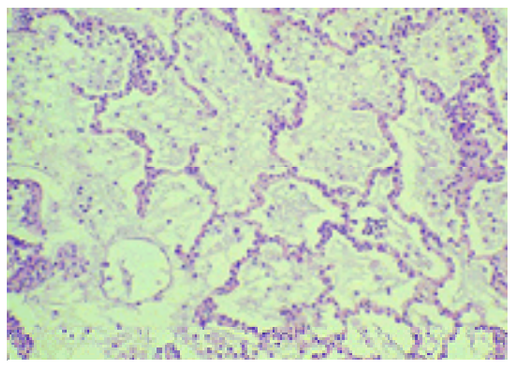


Fig. 7. Bronchoalveolar subtype, adenocarcinoma of the lung. Peg-like cells with minimal pleomorphism growing along alveolar walls are characteristic.

The only curative therapy for adenocarcinomas is surgical resection. The overall prognosis for adenocarcinoma is not as favorable as for squamous cell carcinoma. Although adenocarcinomas have the longest doubling time among lung cancers (189 days), they have an increased propensity for metastasis to the liver, adrenal gland, bone, and brain. Among resectable non-small cell cancers of the lung, adenocarcinomas have the highest incidence of asymptomatic brain metastasis (25 to 30 per cent). Prognosis for adenocarcinomas is more dependent on pathologic staging (i.e. TNM classification) than on the degree of differentiation. Expression of K-*ras* by adenocarcinomas is an independent risk factor for a decreased survival time after surgical excision. Survival from a solitary bronchoalveolar carcinoma is higher than is that for other subtypes of adenocarcinoma. Patients with solitary bronchoalveolar cancer, however, seem to be at a higher risk for metachronous primary bronchoalveolar carcinomas, perhaps reflecting the multicentric origin of this subtype. Conversely, survival for diffuse (i.e. multicentric) bronchoalveolar cancer is much lower than is that for other adenocarcinomas, with no long-term survivals reported for any form of therapy.

Finally, distinction between a peripheral primary lung adenocarcinoma and an adenocarcinoma metastatic from an extrathoracic primary can be very difficult. No reliable histologic distinction is possible. The association of the lesion with a scar, especially one with anthracotic pigment, or the presence of hilar or mediastinal nodal metastasis favors a lung primary. However, among patients with prior resected extrathoracic adenocarcinomas, 90 per cent of new solitary lung nodules will be primary lung adenocarcinomas. Thus, unless a characteristic histologic pattern is identified, all solitary lung adenocarcinomas should be considered primary pulmonary lesions for the purposes of staging and therapy.

Large cell carcinoma

A third type of non-small cell lung cancer is large cell carcinoma, an amorphous group of lung tumors accounting for 5 to 10 per cent of all lung cancers. These tumors are composed of undifferentiated cells, without evidence of squamous cell characteristics or of acini or mucus formation. The cells are larger than are those of small cell lung cancer and they lack neuroendocrine characteristics.

Histologically, the tumor cells grow in sheets, without organization or pattern. Desmoplastic reaction is rare. The cells are large, having pale cytoplasm and varied nuclear configuration. Rarely, a large number of pleomorphic multinucleated giant tumor cells or a majority of cells with clear cytoplasm may be seen, identifying the giant cell and clear cell histologic subtypes, respectively.

Large cell cancers tend to be bulky, with frequent areas of necrosis; cavitation is unusual. Subsegmental or larger bronchi are involved in 50 per cent of cases; however, lobar or main bronchial involvement tends to be secondary to growth of a tumor originating in a more distal region of the bronchial epithelium (Fig. 4). No *in situ* component is seen. Clinical presentation is evenly divided between those patients in whom the lesion is asymptomatic and those in whom symptoms of cough or distant metastasis to the liver, adrenal gland, bone, or brain are identified.

Treatment of large cell carcinomas is primarily surgical, with results of treatment being dependent upon the pathologic stage of the tumor. Seventy per cent of lesions are greater than 3 cm in diameter and lymph node metastases are present in one-half of patients at the time of diagnosis, with a concordant decrement in survival rate following resection. Even among smaller lesions, without nodal metastases, 5-year survival rates following surgical resection are only 40 to 45 per cent, lower than are the survival rates following resection of other non-small cell lung cancer. The survival rate for the giant cell and clear cell subtypes is even lower than is that for the typical large cell lung cancer.

Adenosquamous carcinoma

P>Fewer than 1 per cent of lung cancers consist of lesions with predominantly large cell undifferentiated carcinoma histology and an equal frequency of foci of squamous cell and adenocarcinoma differentiation. Such adenosquamous carcinomas are diagnosed on the basis of histologic appearance, not ultrastructural findings. Typically, these lesions arise as peripheral asymptomatic lung nodules. Their clinical behavior is similar to that of large cell undifferentiated carcinomas, with an increased propensity for metastasis compared with that of other non-small cell lung cancers. Treatment is primarily surgical, with the prognosis being stage dependent and similar to that of large cell undifferentiated lung carcinomas.

Carcinoid tumors

Carcinoid tumors are uncommon, low-grade malignant tumors with biochemical and structural features of tumors of the APUD system. Non-neoplastic cells with APUD characteristics are found in normal bronchial submucosal glands and the epithelium of major bronchi and peripheral bronchioles. Although the concept of a spectrum of biologic behavior for lung neoplasms demonstrating evidence of neuroendocrine differentiation (Kulchitzky cell carcinomas) ranging from carcinoid tumors to small cell lung cancer has been advanced, there is no conclusive evidence for a common cell of origin or for the progression from one type of Kulchitzky cell carcinoma to another. Carcinoid tumors and small cell lung cancer should be considered as separate types of malignant epithelial lung neoplasms demonstrating similar characteristics of neuroendocrine differentiation.

Carcinoid tumors demonstrate two patterns of histology. A typical carcinoid demonstrates uniform cells with abundant clear cytoplasm and regular centrally placed nuclei admixed with cells containing eosinophilic cytoplasm and smaller hyperchromatic nuclei. Occasionally, an intermingling of spindle cells is noted. Mitoses are absent or very rare. The cells are arranged in an ‘endocrinoid' pattern, consisting of a combination of cords, nests, trabeculas, and ribbons (Fig. 8). Although vascular invasion may be present, it is not an indicator of metastasis. Metaplastic bone formation may occur but is uncommon. Argyrophilic stains demonstrate numerous positive cells, while argentaffin stains demonstrate only rare foci of positive cells. Small foci of mucus production may be seen. The cells will stain for neuroendocrine substances, including chromogranin, serotonin, ACTH, catecholamines, and kinins. Ultrastructurally, abundant neurosecretory granules are seen. Typical histology is seen in 90 per cent of carcinoid tumors. An atypical carcinoid tumor does not show the same uniform histologic features. These neoplasms demonstrate increased nuclear and cytologic pleomorphism with frequent mitotic figures with microscopic evidence of necrosis and/or hemorrhage. The atypical carcinoid cells may be distinguished from small cell lung cancer cells by larger amounts of cytoplasm and by less marked nuclear abnormality.

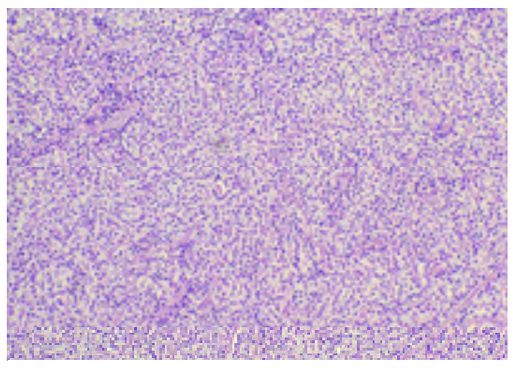


Fig. 8. Typical carcinoid tumor. Uniform cells with centrally located nuclei arranged around vessels in an ‘endocrinoid' pattern are characteristic of typical carcinoid tumors.

Typical and atypical carcinoid tumors are indistinguishable on a clinical basis. Carcinoid tumors occur in equal frequency in both sexes and have been reported from the first to the tenth decade of life, although the mean age is in the fifth decade. These tumors are not related to smoking and have no known environmental risk factors.

The clinical presentation of carcinoid tumors depends on the site of origin of the lesion within the tracheobronchial tree (Fig. 4). Central carcinoid tumors, accounting for 90 per cent of lesions, arise from bronchial submucosal glands in the wall of a subsegmental or larger bronchus. Patients frequently show cough, hemoptysis, or wheezing of long duration (months to years). The chest radiograph may be normal or may demonstrate atelectasis or pneumonitis if bronchial obstruction has occurred. A smooth polypoid mass, ranging from 2 to 4 cm in size and covered by intact mucosa, is seen at bronchoscopy. Biopsy examinations of the lesion may lead to hemorrhage due to its vascularity, although this event is rare. Peripheral carcinoid tumors (10 per cent of lesions) originate in the bronchiolar epithelium and superficially infiltrate the surrounding lung. These neoplasms present themselves as circumscribed pulmonary nodules on a chest radiograph, indistinguishable from other solitary pulmonary nodules. Typically these are asymptomatic. Peripheral carcinoids may have a more frequent spindle cell component and more nuclear variation than do central lesions. At times, multiple peripheral carcinoid tumors may be seen.

The therapy for carcinoid tumors is complete surgical resection. These tumors arise within the bronchial or bronchiolar wall, extending transmurally to involve surrounding lung parenchyma. Therefore, excision requires techniques to encompass the region of involvement, not merely the visible and/or obstructive endobronchial component. For central carcinoids, excision typically requires anatomic lobectomy or lobectomy combined with removal of the involved portion of airway and pulmonary parenchyma, reanastomosing the portions of the airway proximal and distal to the site of the lesion (‘sleeve lobectomy'). For peripheral lesions, non-anatomic parenchymal excision or anatomic lobectomy to obtain tumor-free margins, depending on the size and location of the lesion, are used for resection.

Metastases occur in 10 per cent of carcinoid tumors, most commonly to regional peribronchial lymph nodes. The incidence of metastasis correlates directly with histology. Approximately 5 per cent of typical carcinoids and 70 per cent of atypical carcinoids will have evidence of metastatic disease. Distant metastases may occur to the liver (which may be associated with a clinical carcinoid syndrome) or to bone (resulting in osteoblastic lesions).

The overall 5-year survival rate following surgical resection of carcinoid tumors is greater than 90 per cent. Survival is related primarily to histology, with a 95 to 98 per cent survival for typical carcinoids and a 70 per cent survival for atypical carcinoids. Survival is also stage dependent within respective histologic categories. However, the indolent growth of typical carcinoid tumors, the potential for endobronchial obstruction, and the lack of responsiveness to non-surgical therapies all support the concept of resection even for advanced stages, including those presenting with a pleural effusion or nodal metastases.

Bronchial gland carcinomas

Throughout the tracheobronchial tree, submucosal serous and mucus glands identical to those found in the upper airway and salivary glands are present. Neoplasms arising from these glands are morphologically and biologically analogous to salivary gland tumors. Previously, these various tumors, accounting for 1 to 2 per cent of all lung tumors, were classified as ‘bronchial adenomas' in conjunction with carcinoid tumors. The term has been discarded, however, because it ignores the biologic heterogeneity of these neoplasms and implies that they are benign, whereas the majority are, in fact, low-grade malignant tumors. Bronchial gland carcinomas are now differentiated from carcinoid tumors.

Adenoid cystic carcinoma

Adenoid cystic carcinoma is the most common bronchial gland carcinoma. This tumor is characterized by infiltrative growth, especially along neural structures, a propensity for local recurrence and a prolonged clinical course. Adenoid cystic tumors arise most commonly in the trachea and are the second most common type of primary tracheal tumor, after squamous cell cancers. Grossly, these lesions are polypoid or annular, arising within the airway wall and extending transmurally into the surrounding tissues. The mucosa is typically intact over the tumor, with a characteristic prominence of submucosal vessels. Histologically, regular glands are seen, some containing mucus, others containing a hyaline material (Fig. 9). Characteristic infiltration at the margins of the gross lesion can occur in perineural patterns.

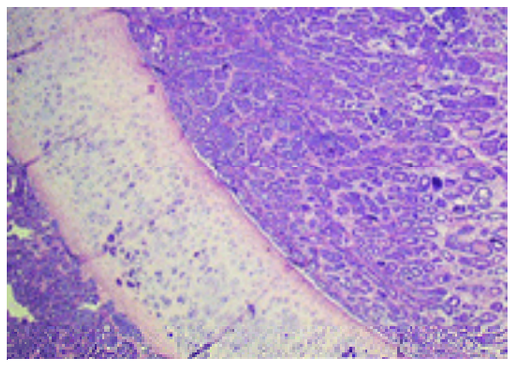


Fig. 9. Adenoid cystic carcinoma. A tumor arising in the tracheal submucosa demonstrates uniform gland formation and transmural infiltration about a tracheal cartilage.

These tumors occur in men and women equally, the average age being 45 years. Presenting symptoms are those of tracheobronchial irritation and obstruction: stridor, positional wheezing, cough, dyspnea, and hemoptysis. Duration of symptoms is usually 18 monthsor longer. Diagnosis is made by laminar tomography of the airways or by bronchoscopy with biopsy. Sputum cytology is not useful.

Treatment consists of local excision of the involved airway if technically possible. Segmental tracheal or carinal resection and reconstruction is generally required. Intraoperative pathologic examination of the margins of resection is mandatory to determine the limits of resection. However, presence of microscopic residual tumor at the margins after maximal airway excision may be satisfactorily managed by postoperative external beam radiation therapy. Primary external beam radiation therapy, augmented by endobronchial high dose-rate brachytherapy, is useful for inoperable lesions.

Overall survival for patients with adenoid cystic carcinoma is 85 per cent at 5 years and 55 per cent at 10 years. Survival following surgical resection with postoperative adjuvant radiation therapy is 85 per cent at 10 years. Adenoid cystic tumors frequently recur after a long time interval. A routine chest radiograph and bronchoscopy are required for at least 20 years following resection. Metastases, most commonly to the lung parenchyma or bones, are usually asymptomatic and slow growing.

Mucoepidermoid carcinoma

Mucoepidermoid carcinoma is the second most frequent bronchial gland carcinoma. The lesion has been subclassified into a low-grade and a high-grade subtype. Both present themselves as firm polypoid masses arising within a major bronchus, with both intraluminal and extramural extension. Histologically, low-grade mucoepidermoid cancers show an admixture of glandular elements and sheets of large cells with little or no squamous differentation (i.e. keratinization) (Fig. 10). High-grade mucoepidermoid cancers, which are extremely rare, resemble low-grade mucoepidermoid lesions except that mucus-filled cystic spaces are more common and more cellular atypia, mitoses, and foci of necrosis are seen. Differentiation of high-grade tumors from the much more common adenosquamous carcinoma may be difficult.

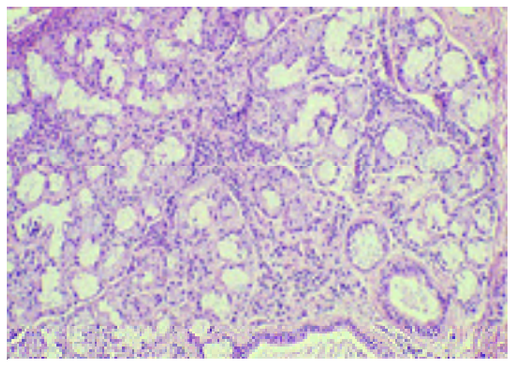


Fig. 10. Mucoepidermoid carcinoma. Glands dispersed among solid regions of tumor cells without evidence of cytologic atypia are typical of the low-grade subtype.

Mucoepidermoid carcinomas occur in men and women with equal frequency, and at a mean age of 40 years. Symptoms of bronchial irritation or obstruction, often of several years' duration, are typical. Chest radiographs to identify signs of bronchial or lobar obstruction, laminar tomograms of the airway, and bronchoscopy with biopsy examination are useful for diagnosis.

Treatment consists of excision of the involved airway and the distal portion of affected lung. Intraoperative evaluation of resection margins is critical to obtain complete excision. For low-grade mucoepidermoid cancers, the prognosis is excellent. No distant metastases have been reported with these tumors. In high-grade mucoepidermoid cancers, distant metastases occur. However, long-term survival has been reported following complete excision of these lesions.

Miscellaneous bronchial gland carcinomas

Mucus gland adenoma and pleomorphic adenoma are rare benign tumors of well-differentiated elements of the bronchial glands. Both are very slow growing and may be identified by bronchoscopy while asymptomatic, or they may present themselves with bronchial obstruction. These cancers are best treated by local excision of the involved airway. Intraoperative pathologic evaluation of resection margins is crucial to obtain a complete excision and eliminate the possibility of local recurrence. No distant metastases have been reported; complete surgical excision is curative.

Carcinosarcoma of the lung

Carcinosarcoma is a malignant tumor in which both epithelial and mesenchymal elements are present and both demonstrate metastatic potential. Presumably, these lesions represent the manifestation of the multipotentiality of the tissue of origin. In the lung, these cancers account for less than 1 per cent of malignant tumors and arise in two forms, pulmonary carcinosarcoma and pulmonary blastoma.

Pulmonary carcinosarcomas are bulky, vascular neoplasms, arising in major bronchi in 50 per cent of cases. Admixture of squamous carcinoma and fibrosarcoma account for half of these lesions histologically. Other malignant epithelial components include adenocarcinoma and large cell carcinoma; a small cell lung cancer component has not been reported. Chondrosarcoma is the second most common mesenchymal differentiation. In most cases, the mesenchymal component predominates, creating difficulty in differentiating these cancers from mixed malignant mesothelioma. Mesothelioma can be identified, however, by its hyaluronic acid production. The mean age of occurrence is 45 years; the tumor is twice as frequent in men as in women.

Pulmonary blastoma is a tumor that differentiates to produce a histologic appearance reminiscent of the lung during its pseudoglandular period of development at 3 months' gestation. These lesions are typically large, peripheral tumors, compressing major bronchi. Histologically, the epithelial elements have a primitive appearance and may condense about a small lumen or demonstrate subnuclear vacuolization that causes a resemblance to embryonic cells containing glycogen. Ultrastructural evidence supports the concept that the epithelial component is a form of poorly differentiated adenocarcinoma. The mesenchymal component has large hyperchromatic nuclei, with little differentiation in most regions. The majority of patients are men; the tumor has been reported in patients from 11 to 77 years of age.

Carcinosarcomas of both types behave as high-grade malignant tumors, although some cases of pulmonary blastoma may have a slow rate of growth. One-half of both types of carcinosarcoma present with extrapulmonary metastases. Surgical resection, if possible, is the only therapy that is curative. Radiotherapy has been used for palliation.

Benign mesenchymal tumors of the lung

Hamartoma

Hamartoma is the most common benign tumor of the lung. It is a benign mesenchymal neoplasm composed of a loose fibrous or myxoid mesenchyme, which undergoes alteration to cartilage, mature fibrous tissue, and fat in varying amounts. The term fibrochondrolipoma has also been used for these lesions. With growth, the tumor entraps bronchial and alveolar epithelium. Most hamartomas arise as isolated lesions in the periphery of the lung, although 15 per cent are endobronchial. In the parenchyma, these neoplasms are well circumscribed and irregularly lobulated. Histologically, there are wide variations in the appearance of the mesenchymal component, although cartilage in varying stages of maturation is a consistent finding. The cartilage is invariably benign, and calcification and ossification may occur. At the margin, chronic inflammation and fibrosis may be seen. A true capsule is not present. Endobronchial hamartomas demonstrate proliferation of the overlying epithelium as secondary change, not a primary neoplastic process. Rarely, multiple hamartomas occur; they tend to be more cystic and lipomatous, having a paucity of cartilage.

Hamartomas are found in 1 of 400 autopsies. Men are affected two to three times more frequently than women. They do not occur in infants or children. The median age at presentation is 35 years. Typically, parenchymal tumors are asymptomatic and detected on routine chest radiograph. Endobronchial lesions may result in symptoms of bronchial obstruction. A very slow rate of growth is characteristic, as is the determination of cartilage within the lesion by chest CT scanning or by percutaneous needle biopsy.

Treatment consists of localized parenchymal or endobronchial excision to prevent the complications of continued growth of the lesion. Malignant change is extremely rare. Recurrence results only from incomplete parenchymal excision. At times, with supporting histologic evidence of benign cartilage from needle biopsy, serial observation of these lesions may be indicated in the asymptomatic patient.

A special circumstance is that of multiple pulmonary leiomyomatous hamartomas. These occur in asymptomatic, middle-aged women with a prior history of myomectomy or hysterectomy for uterine leiomyoma. Such lesions are actually metastases from a uterine or other leiomyosarcoma with a low-grade growth potential. Histologically, the pulmonary lesions are blander than are the uterine primary lesion. A few leiomyomatous hamartomas may be treated by multiple limited pulmonary parenchymal resections. In unresectable cases, hormonal therapy should be used for control of the growth.

Bronchial chondroma

Bronchial chondroma is an unusual neoplasm that arises within pre-existing cartilage in the bronchus. These tumors are small, bosselated masses of 1 to 2 cm in diameter, found incidentally at bronchoscopy. Histologically, there is enlargement of bronchial rings within the lesion. Cartilage is mature, and ossification may be seen. Unlike hamartomas, chondromas cannot be separated from the bronchial wall; they have a propensity for malignant degeneration to a chondrosarcoma. Treatment consists of resection of the involved bronchus and associated lung tissue if necessary. Intraoperative pathologic evaluation of the margins is mandatory to ascertain the likelihood of complete removal. When a bronchial chondroma is incompletely removed, recurrence is frequent.

Granular cell tumor

Believed to be of Schwann cell origin, granular cell tumors occur rarely in the lower trachea and major bronchi. Histologically, large cells with abundant granular cytoplasm grow in an infiltrative pattern. Frequently, squamous metaplasia of the overlying bronchial mucosa is seen. Multiple tumors may occur and 15 per cent of patients have similar tumors of the tongue or skin. Treatment consists of local excision, with appropriate considerations to assure complete removal. Although local

recurrence has been reported, no reports of malignant degeneration or metastases have been made.

Miscellaneous benign mesenchymal tumors

Lipomas, fibromas, neuromas, and leiomyomas have all been reported as isolated benign tumors of the lower respiratory tract. Presentation occurs either as asymptomatic isolated parenchymal lesions or endobronchial lesions, with a greater likelihood for symptoms related to bronchial obstruction. The incidence of these lesions is low and their histology is characteristic of the specific type. Treatment consists of localized surgical excision of the involved airway or parenchyma, with curative results.

Primary sarcoma of the lung

Primary malignant mesenchymal tumors arise in the lung rarely. No known environmental exposures are associated with these tumors. Leiomyosarcoma and fibrosarcoma are the most common histologic types. These tumors may arise in the parenchyma and invade surrounding structures. An endobronchial origin is rare. Presenting symptoms include cough, hemoptysis, or chest pain. Men and women are equally represented. On chest radiography primary sarcomas present themselves as large, circumscribed parenchymal mass lesions. Diagnosis may be made by percutaneous needle biopsy. Therapy consists of surgical excision of the lesion when technically possible. Radiation therapy may be used as adjuvant therapy for an incompletely excised lesion or as primary palliative treatment for unresectable tumors. The prognosis is better for patients having the rare endobronchial sarcoma and for the uncommon lesion occurring in a pediatric patient.

Primary sarcoma of the pulmonary artery

A unique form of sarcoma found in the lung is primary sarcoma of the pulmonary artery. It arises in the intima of the major branches of the pulmonary artery. The sarcoma extends peripherally through the arterial tree and then invades the lung in an interstitial and intra-alveolar pattern. The sarcoma may also spread proximally to the right ventricular outflow tract. The tumor occurs in men and women equally, commonly in the sixth decade. Presentation consists of unexplained intractable right heart failure. Diagnosis is made by right heart catheterization with pulmonary arteriography and biopsy. Distant metastases are present in one-half of the cases at diagnosis. The neoplasm is frequently fatal, with death occurring from progressive obstruction of the pulmonary arterial tree. Surgical resection with reconstruction of the right ventricular outflow tract may result in long-term survival.

Lymphoproliferative disorders of the lung

Plasma cell granuloma

Plasma cell granuloma has also been called inflammatory pseudo-tumor, although in most cases no prior inflammation or active infection is identified. The lesion presents as a well-circumscribed solitary lung mass. Histologically, a proliferation of spindle cells admixed with plasma cells, lymphocytes, and histiocytes is noted with obliteration of normal pulmonary architecture. Most patients are less than 30 years old and are usually asymptomatic. The appropriate treatment is localized resection with preservation of pulmonary parenchyma. Recurrence is rare.

Pseudolymphoma

Pseudolymphoma is a rare polyclonal lymphoid proliferation within lung tissue that resembles benign reactive lymphoid tissue. Typically, these lesions present as subpleural parenchymal masses with ill-defined margins. Histologically, well-differentiated lymphoid cells are seen in the interstitium and infiltrating vessels and small airways. In the centre of the lesion, normal pulmonary architecture is obliterated. Plasma cells are inconspicuous, but germinal centers are present. Men and women are affected with equal frequency, typically in the sixth decade. Most are asymptomatic, but some have a cough or chest pain, while others have fever, chills, and fatigue. Radiographs demonstrate one or more parenchymal lesions without calcification or cavitation that resemble bronchoalveolar carcinoma. Diagnosis is by incisional biopsy or by excision of the lesion *in toto* if it may be encompassed by a limited parenchymal excision. Percutaneous needle aspiration reveals only non-specific lymphoid aggregates. After conservative excision, fresh tissues should be prepared for immunohistochemical studies. Overall, prognosis is good, with a 70 per cent survival rate at 10 years.

Non-Hodgkin's lymphoma

Primary pulmonary lymphoma is a malignant neoplasm consisting of a monoclonal population of well to poorly differentiated lymphoid cells, with or without regional node involvement, without evidence of lymphoma elsewhere. These lesions may present themselves as pulmonary nodules, but frequently invade the overlying pleura. They may be associated with a pleural effusion containing lymphoma cells. Clinical presentation is similar to pseudolymphoma. Excision for diagnosis should be accompanied by evaluation for disseminated disease (e.g. bone marrow aspiration). Final classification is based on immunohistochemical studies of freshly excised tissue and determines the type of chemotherapy required. Overall, the prognosis is for a 20 per cent survival rate at 10 years.

Diagnosis of lung neoplasms

The essential elements for the diagnosis of lung neoplasms include symptoms, physical examination, and chest radiography, including chest CT scans. Benign or low-grade malignant neoplasms are more likely to present in an asymptomatic state. Most malignant neoplasms present themselves with symptoms leading to radiographic evaluation. For lung cancers, the degree of manifest symptoms roughly correlates with the amount of disease. A symptomatic lung cancer is less likely to be a curable lesion.

Studies of screening chest radiographs for lung cancer in asymptomatic patients have shown that more early stage, peripheral non-small cell lung cancers are identified. Despite the inherent resectability of such lesions, randomized studies have not demonstrated reduction in mortality with these measures. Radiographic screening in high-risk individuals remains controversial.

Symptoms

Symptoms from pulmonary neoplasms can be classified according to their etiology. Bronchopulmonary symptoms are those caused by direct tumor effects within the airway or parenchyma, including cough, hemoptysis, wheezing, dyspnea, or pneumonia. Recurrent pneumonia suggests partial bronchial obstruction. Alternatively, pneumonia characterized as a persistent infiltrate indicates complete bronchial obstruction or neoplastic consolidation (e.g. by bronchoalveolar cancer). Intrathoracic symptoms are those created by tumor extension outside the parenchyma. In most cases, these are signs of extensive disease and relative unresectability. Typical intrathoracic symptoms include:

1. hoarseness due to recurrent laryngeal nerve invasion, either directly or by metastasis to mediastinal lymph nodes;
2. phrenic nerve invasion with hemidiaphragmatic paralysis;
3. chest pain or pressure, due to parietal pleural or chest wall invasion;
4. superior vena cava syndrome, due to extrinsic compression by a large hilar mass or bulky mediastinal lymph node metastases;
5. cardiac dysrhythmias, resulting from atrial or pericardial invasion;
6. dysphagia, due to compression or invasion of the esophagus in the posteroinferior mediastinum; and
7. vascular pain, typically central, severe, and unremitting, due to invasion of the aortic wall.

Extrathoracic symptoms are due, in most cases, to metastatic extension of the tumor to distant sites, with direct local effects. However, in 5 to 10 per cent of lung cancers, paraneoplastic syndromes may be identified. These remote tumor effects are not caused by metastatic deposits, but are due to polypeptide hormone production, interference with immune function, inappropriate initiation of physiologic mechanisms, or competitive receptor blockade. They are more common with malignant lung tumors and are most frequently seen with small cell lung cancer. A notable exception is the syndrome of hypertrophic pulmonary osteoarthropathy, consisting of digital clubbing and symptomatic long bone periosteal reaction, which is seen in 10 per cent of non-small cell lung cancer, especially adenocarcinomas (Table 2).

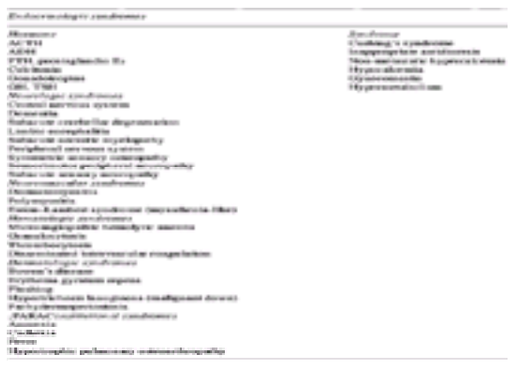


Table 2 Paraneoplastic syndromes

Chest radiographs

Chest radiography is the basic diagnostic modality for neoplasms of the lung, with a high sensitivity at a low cost and minimal risk to the patient. Standardized techniques employ a high kilovoltage source and inspiratory posteroanterior and lateral views. Indications for chest radiography include persistent or recurrent pulmonary symptoms, a past medical history or family history of lung cancer, or as an annual study in patients over 50 years of age who have a history of tobacco use. Chest radiographs should always be examined in concert with prior chest films if they are available. If a nodule or mass is identified, the pertinent diagnostic features are its duration and the presence of calcification. Nodules present for longer than 3 years without documented change in size are considered benign in nature—either benign neoplasms or granulomas. Calcification, if present in a solitary central or a lamellar distribution, signifies a granuloma. ‘Popcorn' calcification is characteristic of hamartoma. All other types of calcification are non-diagnostic. Lung cancers demonstrate calcification in 7 per cent of cases due to dystrophic deposition or to neoplastic growth about a pre-existing calcified focus in the parenchyma.

In general, asymptomatic pulmonary nodules are inflammatory or infectious in 30 per cent of cases, benign tumors in 25 per cent of cases, and malignant tumors in 45 per cent of cases. However, the proportion of malignant tumors rises to 60 per cent in patients more than 50 years of age. Any non-calcified asymptomatic nodule in a patient over 40 years of age should be considered a lung cancer until histologic evidence to the contrary is obtained.

Conventional laminar tomography of the chest has limited application in the diagnosis of parenchymal pulmonary neoplasms, as computed tomographic scanning provides a more sensitive and precise examination. However, for neoplasms involving the central airways (trachea and main and lobar bronchi), laminar tomography provides an unparalleled image. Laminar tomograms precisely define the tumor within the central airways and are essential before surgery.

Computed tomographic (CT) scans

CT scanning has greatly enhanced the imaging of intrathoracic structures. The indications for chest CT scanning include all patients with a suspected pulmonary neoplasm or with a questionable hilar abnormality. A CT scan provides contour and density characteristics of a nodule, which can differentiate inflammatory from neoplastic lesions and, occasionally, identify specific histologic subtypes. Additional parenchymal lesions, beyond the sensitivity of chest radiographs, may be identified. The mediastinum and mediastinal lymph nodes are well delineated. Nodes larger than 1 cm (short axis) are considered abnormal; they should be sampled for biopsy examinations of potential metastatic spread. Nodes smaller than 1 cm are considered normal unless they arise in the context of a known malignant neoplasm of the lung or an undiagnosed nodule greater than 3 cm in diameter. In such cases, biopsy of these ‘normal' nodes is indicated, as the incidence of nodal metastasis ranges from 15 to 25 per cent. Finally, the CT scan provides limited views of two common sites of extrathoracic metastasis, the adrenal glands and liver. Metastasis from lung cancer to the adrenal gland is a frequent event, usually occurring coincident with mediastinal lymph node metastasis. However, two-thirds of enlarged adrenal glands identified on chest CT scans in patients with lung cancer are benign adenomas. Indicators of a benign adrenal nodule include a size of less than 3 cm, a low density, and normal mediastinal lymph nodes. Biopsy data are reserved for adrenal nodules in patients with lung cancer who do not meet these criteria.

Magnetic resonance imaging (MRI)

MRI has limited application for patients with pulmonary neoplasms. It is the imaging modality of choice for the region of the thoracic inlet because of the spatial and histologic resolution obtained by MRI for the pulmonary, neural, and vascular structures in this region. MRI studies are useful for any neoplasm arising in this area. Identifying invasion of the thoracic aorta by a malignant neoplasm in a symptomatic patient is another frequent indication for MRI.

Positron emission tomography (PET) scans

PET scanning is a technique for biochemical characterization of tumors and their metastatic deposits. A substrate modified with a positron-emitting isotope, such as fluorodeoxyglucose, is injected intravenously. The substrate localizes to metabolically active cells, such as neoplastic cells. The limit of resolution for PET scanning is 8 to 10 mm. PET scanning has a role in both diagnosing and staging of lung cancer. For indeterminate pulmonary nodules, PET has 95 per cent sensitivity and 81 per cent specificity in identifying lung cancers. Whole-body PET scans may identify unsuspected metastasis in 10 per cent of patients.

Sputum cytology

Cytologic evaluation of sputum for neoplastic cells can diagnose malignant lung neoplasms in 60 per cent of cases, primarily when the tumor invades transmucosally. A lower yield is found for peripheral tumors, submucosal lesions, or when bronchial obstruction is present. Cell type can be specified in 80 per cent of cases, with a false-positive rate of less than 1 per cent. Sputum cytology should be obtained on any patient with a suspected lung tumor who has a productive cough or hemoptysis. Specimens should be collected in the morning for 3 days.

Bronchoscopy

Bronchoscopy is the endoscopic evaluation of the bronchial tree using rigid or flexible instruments. Rigid bronchoscopes, which allow visualization to the third order (segmental) bronchi, are ideal for interventional procedures, including endobronchial tumor excision for relief of obstruction (‘coring-out'), dilatation of tracheal or bronchial stenoses, and management of massive hemoptysis. Precise measurements of relative lengths of airway involved are also possible to provide an estimate of resectability. Flexible bronchoscopes allow visualization to fifth or sixth order bronchi and permit a wide range of diagnostic procedures, including directed bronchial washings, cytologic brushings, transmucosal or transbronchial biopsy, and selective transmural aspiration of paratracheal nodes. For lesions greater than 2 cm in size, located in the inner one-third of the pulmonary parenchyma, these procedures accurately diagnose neoplasms in 87 per cent of patients. Accuracy falls to 23 per cent for smaller lesions in the outer one-third of the lung.

Indications for bronchoscopy include any clinical context in which a lung neoplasm is suspected (Table 3). Rigid bronchoscopy usually requires general anesthesia, while flexible bronchoscopy can be performed with either topical or general anesthesia. Complications include hypoxia, requiring supplemental oxygen administration in all patients, laryngospasm, cardiac dysrhythmia, lidocaine toxicity, pneumothorax, or bleeding (particularly if the platelet count is less than 50 000/mm^3). The incidence of serious complications is less than 5 per cent.

Symptoms
Persistent cough
Stridor
Wheezing—positional
Hemoptysis
Radiographic findings
New pulmonary mass lesions
Persistent pulmonary infiltrate
Abnormal sputum cytology
Contemplated pulmonary resection

Table 3 Indications for bronchoscopy

Percutaneous needle biopsy

Percutaneous needle aspiration of pulmonary neoplasms is a safe, accurate method for histologic identification, particularly for those tumors located in the outer two-thirds of the pulmonary parenchyma. The technique employs a 19-gauge needle with a stylet that is positioned near the lesion using fluoroscopic or CT scan guidance. Ultrasound localization may be used for very peripheral lesions. The stylet is removed and a second needle is inserted for aspiration or core biopsy of the nodule. Diagnosis for nodular lesions is possible in 85 per cent of cases, with an accuracy of 95 to 98 per cent for malignant lesions and 85 per cent for benign lesions. Repeat biopsies may be taken from indeterminate lesions to increase diagnostic yield.

Pneumothorax occurs in 20 to 25 per cent of patients and may be lessened by positioning the patient in an ipsilateral decubitus position after the biopsy. Less than 2 per cent of patients require chest tubes. Although minor hemoptysis is common, massive hemoptysis is rare. The incidence of air embolism or tumor implantation only occurs while taking 1 in 5300 biopsy specimens. Contraindications to needle biopsy include abnormal coagulation studies, pulmonary artery hypertension, and an inability to co-operate with the examination (e.g. inability to hold respirations, uncontrollable cough).

Excisional biopsy and other procedures

Excisional biopsy of lung lesions may be used as a diagnostic modality. In most cases, excision for diagnosis should be performed only after clinical staging is complete. Accurate intraoperative pathology is required to facilitate decisions regarding the extent of resection.

Biopsy examination of the scalene fat pad or of supraclavicular nodes has been employed as a diagnostic procedure for malignant lung neoplasms. Only 15 per cent of non-palpable nodes contain metastatic disease; if nodes are clinically palpable, metastases will be identified in 90 per cent. These procedures are more accurately used as clinical staging methods when nodes are identified on physical examination.

Staging

Staging classification

Clinical staging prior to surgical intervention is primarily useful for malignant lung tumors, to allow reproducible classification, prognostication, and comparison of therapeutic modalities. Since 95 per cent of all lung neoplasms are malignant, staging of a lung lesion before resection is a frequent requirement. A modified TNM classification for lung cancers is currently in use.

The T descriptor refers to the primary lesion, in terms of its size and involvement of local tissues.

TIS—carcinoma *in situ.*
T1—a tumor 3 cm or less in greatest dimension, contained within the visceral pleura and without invasion proximal to a lobar bronchus at bronchoscopy, or a superficial lesion confined to the bronchial wall (Fig. 11(a)).
T2—a tumor more than 3 cm in greatest dimension, or a tumor with visceral pleural invasion, or bronchial obstruction extending to the hilar region; at bronchoscopy, the lesion must be 2 cm or more from the tracheal carina (Fig. 11(b)).
T3—a tumor of any size with direct extension into the chest wall (including superior sulcus tumors), diaphragm, mediastinal pleura, or parietal pericardium, or a tumor within 2 cm of the carina without direct carinal involvement or complete atelectasis of the lung.
T4—a tumor of any size with invasion of the central mediastinal structures (heart, great vessels, trachea, carina, esophagus, vertebral bodies) or a cytologically proved malignant pleural or pericardial effusion or satellite nodules within the primary tumor lobe of the lung.

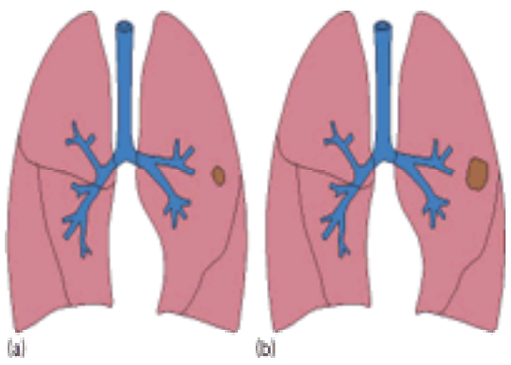


Fig. 11. TNM staging—T descriptor. (a) T1 lesions are peripheral lesions, 3 cm or smaller in size, completely contained within the visceral pleura. (b) T2 lesions are lesions greater than 3 cm in size with a similar disposition within the lung.

The N descriptor describes involvement of regional lymph nodes.

N0—no demonstrable metastasis to regional nodes.
N1—metastasis to ipsilateral peribronchial or hilar nodes, including by direct extension (Fig. 12).
N2—metastasis to ipsilateral mediastinal and/or subcarinal nodes (Fig. 12).
N3—Metastasis to contralateral mediastinal or hilar nodes or to ipsilateral or contralateral scalene or supraclavicular node.

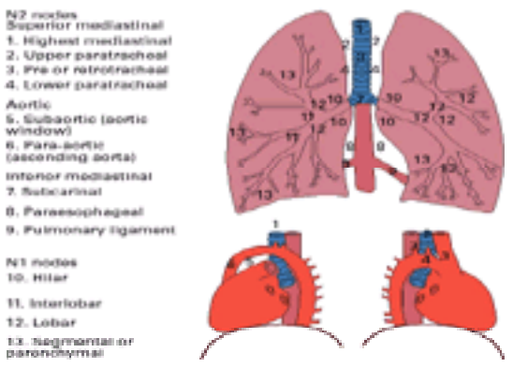


Fig. 12. TNM staging—N descriptor. N1 nodes are peribronchial or hilar nodes. N2 nodes are all mediastinal nodes arising ipsilateral to the primary tumor. N3 nodes are mediastinal nodes contralateral to the primary tumor.

The M descriptor indicates the absence or presence of dissemination to distant sites (M0 and M1, respectively). A tumor nodule in an ipsilateral non-primary tumor lobe

of the lung is classified M1.

Stage grouping of TNM subsets is demonstrated in Table 4. Stage I consists of localized tumors (T1, T2) with no evidence of lymph node metastasis. Stage IA is a T1 N0 tumor; stage IB is a T2 N0 tumor. Stage II includes patients with localized tumors and metastasis to N1 nodes (Fig. 13). Stage IIA is a T1 N1 tumor. Stage IIB is a T2 N1 or T3 N0 tumor. Stage III is subdivided essentially on the basis of surgical resectability. Stage IIIA includes T3 N1 tumors, as well as T1 to T3 tumors with N2 nodal metastasis (Fig. 14). Although surgically resectable, patients with N2 nodal disease frequently require neoadjuvant or adjuvant therapy. Stage IIIB includes patients with T4 tumors without distant metastasis and any tumor with N3 nodes. Most patients with stage IIIB lesions are not candidates for surgical resection, with the notable exception of localized carinal tumors; carinal resection and reconstruction may be possible. Stage IV includes any tumor with evidence of M1 disease. These lesions are generally considered unresectable, with the exception of those patients with isolated cerebral metastasis, in whom combined pulmonary and cerebral resection may lead to prolonged survival.

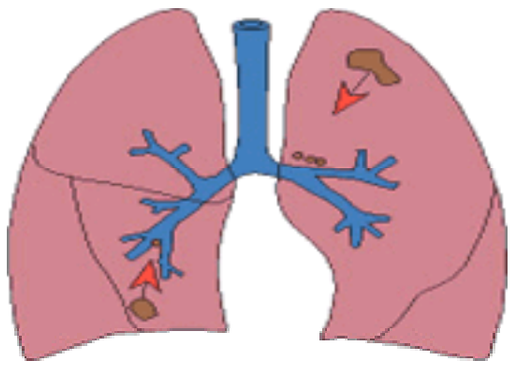


Fig. 13. TNM staging—stage II. All T1 or T2 lesions with metastasis to N1 nodes are considered as stage II lesions. T3 N0 lesions are grouped as stage IIB.

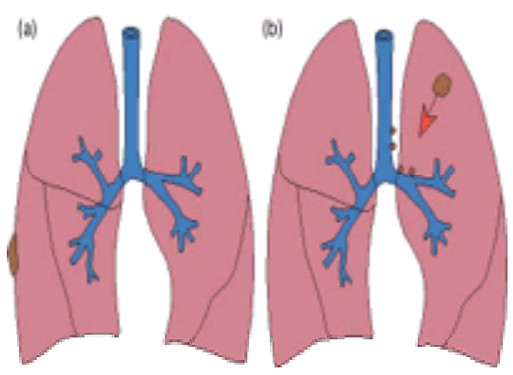


Fig. 14. TNM staging—stage IIIA. (a) Stage IIIA includes T3 lesions with involvement of the chest wall, including superior sulcus tumors and metastasis to N1 nodes. (b) Stage IIIA also includes T1 to T3 lesions with metastasis to N2 mediastinal lymph nodes.

Stage	5-year survival (%)			
IA	T1	N0	M0	67
IB	T2	N0	M0	57
IIA	T1	N1	M0	55
IIB	T2	N1	M0	39
	T3	N0	M0	
IIIA	T3	N1	M0	23
	T1—3	N2	M0	
IIIB	T4	any N	M0	5*
	T1—4	N3	M0	
IV	any T	any N	M1	<5†

*Survival for T4 tumors treated by carinal resection and reconstruction ranges from 20 to 35 per cent at 5 years.
† Survival for M1 tumors consisting of isolated cerebral metastasis treated by combined pulmonary and cerebral resection is 15 to 30 per cent at 5 years.

Table 4 TNM staging classification for lung cancer

Non-invasive staging procedures

Staging of lung tumors may be begun by a series of non-invasive studies. Chest CT scans provide important information with regards to tumor size (T descriptor), the size of mediastinal lymph nodes, and the presence of extrathoracic metastatic disease at selected sites (the cephalad one-third to one-half of the liver, the adrenal glands). Involvement of the visceral or parietal pleura or of the chest wall cannot be determined from CT scans or other non-invasive studies without direct evidence of extrapleural extension (e.g. bony destruction, pericardial effusion). A notable exception is the ability of MRI to identify the presence of vascular or neural invasion for superior sulcus tumors.

Non-invasive studies that may further elucidate extrathoracic metastases include bone scans and CT scans of the head and abdomen. Typically, these are indicated only for patients with symptomatic complaints or biochemical abnormalities related to these regions (e.g. abnormal alkaline phosphatase). However, for large tumors, poorly differentiated lung cancers, and for adenocarcinomas (which have a propensity for early cerebral metastasis) routine metastatic studies may be indicated. Bone marrow aspiration is indicated for small cell lung cancer because of the frequency of metastasis to this site.

Invasive staging procedures

Invasive staging procedures are primarily reserved for patients with intrathoracic lesions who are candidates for surgical resection. Mediastinoscopy, by a cervical incision, allows biopsy examination of common mediastinal sites of lymph node metastasis, including the paratracheal, tracheobronchial, and anterior subcarinal regions (Fig. 15). The aorticopulmonary window, posterior subcarinal, hilar, and periesophageal nodes are not accessible by this method. Mediastinoscopy should be performed in all patients with suspected or proved malignant epithelial tumors of the lung, except in cases of T1 lesions with normal mediastinal lymph nodes on chest CT scan examination. Sampling of all accessible nodal sites at mediastinoscopy is required. Mediastinoscopy has a documented sensitivity of 93 per cent and accuracy of 96 per cent. Its use has decreased the incidence of exploratory thoracotomy from 25 per cent to less than 10 per cent and increased resectability rates to greater than 90 per cent. Major complications occur in 0.3 per cent of cases, usually bleeding from the brachiocephalic artery or azygous vein. Other complications include injury to the left recurrent laryngeal nerve (1 per cent), pneumothorax, mediastinal emphysema, infection, and esophageal perforation.



Fig. 15. Mediastinoscopy. The procedure is performed through an incision in the suprasternal notch, gaining access for palpation and endoscopic visualization by the pretracheal space. Systematic sampling of all accessible nodal sites is required.

Mediastinotomy, allowing surgical access to the prevascular and hilar regions by a right or left parasternal incision, is useful for biopsy of aorticopulmonary or hilar nodes, as well as to assess direct invasion of the central pulmonary artery or thoracic aorta. In general, metastatic extension to aorticopulmonary or hilar nodes does not exclude the patient from potential surgical resection. This approach is most commonly used for left upper lobe lesions, in which lymphatic drainage proceeds from peribronchial to aorticopulmonary window nodes (Fig. 16).

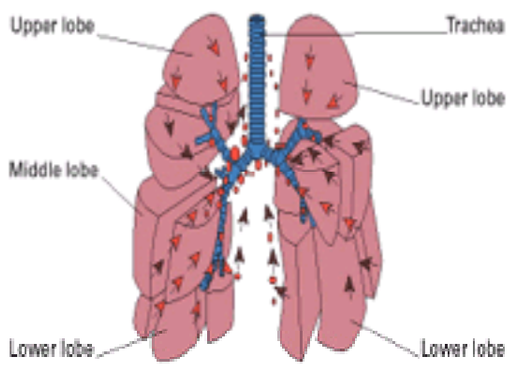


Fig. 16. Lymphatic drainage of the bronchopulmonary segments. On the left side, lymphatic drainage from the upper lobe travels to peribronchial and then subaortic lymph nodes. Lymphatic drainage from the left lower lobe follows a route to the subcarinal and right tracheobronchial lymph nodes. On the right side, drainage from all three lobes coalesces at the origin of the bronchus intermedius (sump of Borrie) prior to further dispersion to hilar and mediastinal lymph nodes.

Video-assisted thoracoscopy is an evolving technique that allows access to the entire pleural space for inspection of both parenchymal and mediastinal structures as well as for biopsy and surgical intervention. All sites not accessible to mediastinoscopy are potentially accessible to video-assisted thoracoscopy. Further experience with this technique will be required to identify its utility as a routine staging procedure.

Preoperative evaluation

The lung provides oxygen and eliminates carbon dioxide from the body. Several factors are involved in normal pulmonary function, each of which can be measured to some extent (Table 5, and see also Chapter 10.5). All thoracic surgical interventions, including bronchoscopy or thoracotomy, as well as parenchymal resection, may interfere with one or more of these components. Assessment of these factors, and approximation of the effect of the planned surgical intervention, is critical to the evaluation of patients with lung neoplasms.

Assessment	Test
Adequate gas-exchange space	Lung volume (TLC)
Adequate pulmonary ventilation	VC, MBC, FEV_1, FEV_1/FVC per cent, Pco_2
Matching of blood flow/ventilation	V/Q scans
Adequate alveolar-capillary surface	Diffusing capacity (DLco)
Pulmonary vascular pressures	Clinical examination, ECG, chest radiograph, catheterization
Respiratory control center	Hypercarbia, hypoxemia response
Systemic stress response	Exercise evaluation

Table 5 Assessment of components of pulmonary function

Many of the tests listed in Table 5 have been used to predict perioperative risk. However, the complexities of pulmonary function mean there are no absolute cut-off points. FEV_1 and arterial blood gas tests are the most important initial tests. Important information on the physiologic status of the patient is also obtained from a chest radiograph and CT scan. Evidence of coronary artery calcification on chest CT scan indicates an increased risk for cardiac dysrhythmias and myocardial ischemia in the postoperative period.

A walk with the patient up two flights of stairs can enable a clinical assessment of integrated pulmonary function. Completion of this task without extreme dyspnea signifies sufficient compensatory capability to tolerate at least a thoracotomy, which alone may diminish maximal breathing capacity by 50 per cent.

For any pulmonary resection, the proportion of functional lung tissue that will be removed can be approximated by the use of quantitative radionuclide *V*/*Q* scanning. Postoperative pulmonary function can be estimated by multiplying preoperative FEV_1 by the percentage of uninvolved lung obtained from the *V*/*Q* scan.

Resectional surgery should not be carried out in most instances in the presence of the following contraindications.

1. calculated postoperative FEV_1 < 0.8 l or < 30 per cent ideal FEV_1 value for patient's body surface area
2. Pco_2 > 45 mmHg.
3. PO_2 < 50 mmHg.
4. single breath *DL*co less than 50 per cent of predicted value
5. mean pulmonary artery pressure, after balloon occlusion of the artery to the involved lung, ³ 35 mmHg.

Surgical principles and techniques

Thoracic incisions

Various approaches have been used to access the pleural space for surgical interventions. Thoracotomy, performed with the patient in a lateral decubitus position, remains the most common incision. The standard posterolateral thoracotomy is an incision beginning midway between the medial border of the scapula and thoracic spine and continuing in a curvilinear fashion, parallel to the ribs, to the anterior axillary line. The maximum exposure of the pulmonary parenchyma and hilar structures is afforded by this incision. Optimal exposure of the tracheal carina is also afforded by this incision when performed on the right side. A disadvantage to this incision is the necessary transection of the latissimus dorsi muscle, an important accessory respiratory muscle (Fig. 17). An alternative approach is an anterolateral ‘muscle sparing' thoracotomy, extending from the tip of the scapula to the inframammary crease (Fig. 18). The operative exposure is somewhat less extensive with this incision, but transection of the chest wall musculature is avoided.

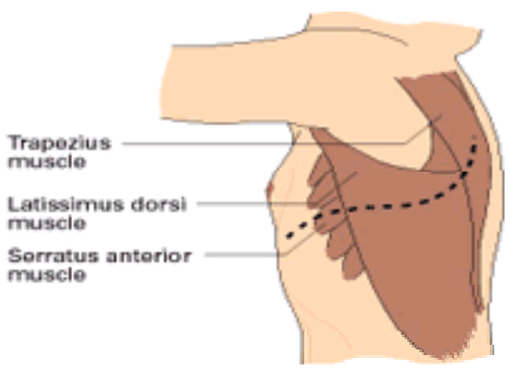


Fig. 17. Posterolateral thoracotomy. The most common incision for pulmonary resection, posterolateral thoracotomy divides the latissimus dorsi muscle and provides optimal exposure of parenchymal and hilar structures. A right posterolateral thoracotomy provides optimal exposure of the tracheal carina.

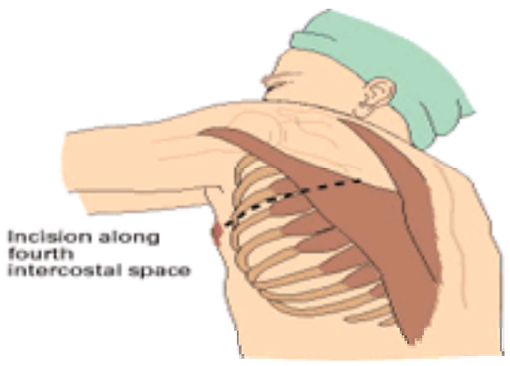


Fig. 18. Anterolateral thoracotomy. A muscle-sparing incision that preserves the chest wall musculature.

Median sternotomy affords a limited exposure of both pleural spaces and the anterior hilar structures. Surgical maneuvers involving the lower lobes or bronchi are limited by this incision. Exposure of the supracarinal trachea, however, is excellent. Median sternotomy causes the least compromise of pulmonary function in the early postoperative period of any thoracic incision. Bilateral trans-sternal anterior thoracotomy is useful when bilateral pulmonary resections are necessary. Finally, video-assisted thoracoscopy allows a minimally invasive approach for evaluation of the pleural space and for peripheral pulmonary resections. Its long-term efficacy and utility, however, remain to be elucidated.

Intraoperative evaluation

Upon entering the pleural space a systematic evaluation of the pleural surfaces, mediastinum, and pulmonary parenchyma should be completed. All intrapleural adhesions should be divided, and any fluid aspirated and submitted for cytologic evaluation. Systematic palpation and excisional biopsy of mediastinal masses or enlarged lymph nodes should be performed. Palpation of the entire pulmonary parenchyma with the lung in both an inflated and atelectatic state is necessary to identify any potential lesions. In patients with multiple pulmonary nodules, palpation at thoracotomy identifies 20 per cent more lesions than can be identified by preoperative chest CT scan.

Principles of pulmonary resection

The resection of lung neoplasms should adhere to a principle of complete excision of the tumor and appropriate lymphatic drainage. *En bloc* resection should be employed whenever possible. Margins of resection, both parenchymal and bronchial, should be free of residual neoplastic disease. Maximal attempts at conservation of uninvolved pulmonary parenchyma should be made, however, in view of the vital function of the lung as well as the incidence of second primary neoplasms (10 per cent at 5 years for patients with non-small cell lung cancer).

Vascular structures are identified and ligated. Bronchial closure is performed with an interrupted suture technique or a mechanical stapling device. In skilled hands, no difference in the rate of bronchial stump fistula is noted between these techniques. Application of a vascularized flap of pleura or pericardial fat to the closed stump diminishes the incidence of postoperative fistula, particularly when adjuvant radiation or chemotherapy is contemplated.

Typically, anatomic resection such as lobectomy or pneumonectomy is performed. Parenchyma-sparing resection is performed for localized peripheral lesions only if regional lymph nodes are demonstrated to be free of metastasis by intraoperative pathologic examination or when predicted postoperative pulmonary function is marginal. Knowledge of segmental anatomy allows control of the bronchovascular structures for more central lesions requiring parenchyma-sparing resections. The margins of such a ‘segmental' resection are determined by the limits of the neoplasm, not strictly by the anatomic boundaries of the associated bronchopulmonary segment, to assure complete excision.

Involvement of the lobar or main bronchus does not require progression to a pneumonectomy. The techniques for bronchial resection and reconstruction (‘sleeve resection') have been used for both benign and malignant lung neoplasms without adverse effects on postoperative morbidity or prognosis.

Special issues

Pleural effusion occurring in association with lung neoplasms may indicate a malignant lesion with extension beyond the limits of surgical resection. Aspiration of any pleural effusion occurring with a pulmonary neoplasm for biochemical and cytologic study is required. An exudative effusion with more than 100 000 red blood cells/mm^3 or predominance of lymphocytes suggests a malignant effusion. However, cytologic confirmation is required. Among patients with lung cancer who present with a pleural effusion, only 40 to 75 per cent will demonstrate a positive cytology; the remainders are candidates for surgical resection. Video thoracoscopic evaluation of the pleural space prior to thoracotomy is a useful staging modality in these cases.

Chest wall invasion by a pulmonary neoplasm does not preclude potential surgical resection. Survival and resectability relates more significantly to nodal status and the possibility of extrathoracic metastasis in such cases. Resection should include a margin of 5 cm about the region of chest wall invasion. Reconstruction with prosthetic material (e.g. Marlex-methacrylate, PTFE patch) is required for lateral or anterior defects with involvement of two or more ribs and their adjacent intercostal spaces. For the more common posterolateral defects, overlain by the scapula and shoulder girdle, reconstruction is rarely needed. Adjuvant radiation therapy appears beneficial only for compromised margins of resection.

Superior sulcus tumors arise in the apex of the lung cephalad to the clavicle, as seen on an anteroposterior chest radiograph. The narrow anatomic confines of the region makes early diagnosis difficult, invasion of local chest wall and neurovascular structures likely, and complete resection challenging. Many patients present themselves with symptoms of involvement of the stellate ganglion (i.e. Horner's syndrome) and of C8 or T1 nerve roots (Pancoast syndrome). MRI optimally visualizes

the region and determine the extent of invasion. Diagnosis is made by percutaneous needle biopsy. Following extrathoracic and mediastinal staging, preoperative radiation therapy (30 to 35 Gy) is used to optimize the margins of resection. Anatomic excision of the parenchymal lesion and involved chest wall, as well as the lower trunk of the brachial plexus if affected, is performed. An anterior cervicothoracic approach may facilitate resection of lesions in the anterior portion of the thoracic outlet that involve the subclavian vessels. Most patients receive postoperative radiation therapy to a total dose of 50 to 55 Gy. Survival following this approach is 35 per cent at 5 years; squamous histology has a more favorable prognosis. Concurrent chemoradiotherapy is currently under investigation as a preoperative and postoperative adjuvant modality to improve survival.

Metastasis to N2 lymph nodes implies microscopic metastatic dissemination of tumor to extrathoracic sites. For this reason, the majority of patients with N2 disease identified at mediastinoscopy undergo preoperative neoadjuvant therapy with chemotherapy or a combination of chemotherapy and mediastinal radiotherapy followed by surgical resection. Survival following this approach ranges from 20 to 40 per cent, depending on the specific regimen. In instances when N2 nodes not identified by mediastinoscopy are found at thoracotomy, anatomic parenchymal resection encompassing all lymphatic drainage will achieve a survival of 25 per cent at 5 years. For typical carcinoid tumors, N2 involvement treated by complete resection does not adversely affect long-term prognosis.

Involvement of the tracheal carina by a pulmonary neoplasm, although classified as a T4 lesion, may be managed by primary resection and reconstruction. The critical factors include the extent of involvement of the distal trachea and proximal bronchi and the possibility of mediastinal or extrathoracic metastases. For benign or low-grade malignant neoplasms (carcinoid, bronchial gland tumors), resection to the limits of the neoplasm or to the limits of reconstruction is curative in 80 to 85 per cent of cases. For non-small cell lung cancer, carinal involvement may be surgically managed by carinal resection and reconstruction alone, or by incorporation of resection of the tracheal carina with an ipsilateral pneumonectomy, depending on the anatomic disposition of the lesion. Survival in this group is 25 to 30 per cent at 5 years.

Isolated cerebral metastasis in patients with lung cancer occurs most commonly with adenocarcinoma. When the lung primary is otherwise localized or when the lung primary has been previously resected and no other metastatic disease is evident, the treatment of the cerebral lesion determines survival. Surgical resection of the cerebral metastasis is indicated for isolated superficial lesions, in the presence of a large amount of edema or mass effect, or when there is associated hemorrhage. For concurrent pulmonary and cerebral lesions, mediastinoscopy to complete clinical staging is usually followed by cerebral resection prior to pulmonary resection. Survival rates of 15 to 20 per cent at 5 years have been reported.

Further reading

Arrigoni MG, Woolner LB, Bernatz PE. Atypical carcinoid tumors of the lung. *Journal of Thoracic and Cardiovascular Surgery* 1972; **64**: 413–21. [The original description of atypical carcinoid tumors.]

Bacha EA *et al.* Surgical treatment of primary pulmonary sarcomas. *European Journal of Cardiothoracic Surgery* 1999; **15**: 456–60. [A recent report of 15 patients identifying complete resection as a requisite for long-term survival.]

Bragg DG, Chor PJ, Murray KA, Kjeldsberg CR. Lymphoproliferative disorders of the lung: histopathology, clinical manifestations and imaging features. *American Journal of Roentgenology* 1994; **163**: 273–81. [A review of a variety of non-lymphomatous disorders of the lung.]

Bria W, Kanarek DJ, Kazemi H. Prediction of postoperative pulmonary function following thoracic surgery. Value of V/Q scanning. *Journal of Thoracic and Cardiovascular Surgery* 1983; **86**: 186–92. [The initial report on the utility of V/Q scans in predicting postoperative pulmonary function.]

Carlens E. Mediastinoscopy: a method for inspection and tissue biopsy in the superior mediastinum. *Diseases of the Chest* 1959; **36**: 343–49. [The original description of the most useful method of invasive staging for lung cancer.]

Colby TV, Koss MN, Travis WD. Tumors of the lower respiratory tract. In: Rosai J, ed., *Atlas of tumor pathology*, Fascicle 13. Armed Forces Institute of Pathology, Washington DC, 1995. [An encyclopedic treatise on the pathology of lung tumors.]

Detterbeck FC. Pancoast (superior sulcus) tumors. *Annals of Thoracic Surgery* 1997; **63**: 1810–18. [A recent review of superior sulcus tumors and their management.]

Ducrocq X *et al.* Operative risk and prognostic factors of typical bronchial carcinoid tumors. *Annals of Thoracic Surgery* 1998; **65**: 1410–14. [A recent analysis of 139 patients emphasizing complete surgical resection leads to an excellent prognosis.]

Gaissert HA, Mathisen DJ, Moncure AC, Hilgenberg AD, Grillo HC, Wain JC. Survival and function after sleeve lobectomy for lung cancer. *Journal of Thoracic and Cardiovascular Surgery* 1996; **111**: 948–53. [A report documenting better outcome for sleeve resections than pneumonectomy.]

Garcia-Closas M, Kelsey KT, Wiencke JK, Xu X, Wain JC, Christiani DC. A case–control study of cytochrome p450, glutathione S-transferase M1, cigarette smoking and lung cancer susceptibility. *Cancer Causes and Control* 1997; **8**: 544–53. [A recent demonstration of a twofold risk for lung cancer based on genetic polymorphisms in carcinogen metabolism.]

Goodman GE. Prevention of lung cancer. *Current Opinion in Oncology* 1998; **10**: 122–6. [A recent review of the results of chemoprevention trials.]

Greene R *et al.* Supplementary tissue core histology from fine-needle transthoracic aspiration biopsy. *American Journal of Roentgenology* 1985; **144**: 787–92. [The initial description of the current techniques used for biopsy of lung lesions.]

Grillo HC, Mathisen DJ. Primary tracheal tumors: treatment and results. *Annals of Thoracic Surgery* 1990; **49**: 69–77. [A review of the results of surgical and palliative therapy for tracheal tumors.]

Johnson BE. Second lung cancers in patients after treatment for an initial lung cancer. *Journal of the National Cancer Institute* 1998; **90**: 1335–45. [An excellent review of the incidence and outcome of second lung cancers.]

Lowe VJ, Naunheim KS. Current role of positron emission tomography in thoracic oncology. *Thorax* 1998; **53**: 703–12. [A recent review of the utility of PET scans for thoracic neoplasms.]

McLoud TC *et al.* Bronchogenic carcinoma: analysis of staging in the mediastinum with CT by correlative lymph node mapping and sampling. *Radiology* 1992; **182**: 319–22. [A thorough comparative study of CT and mediastinoscopy in the staging of lung cancer.]

Mitchell JD *et al.* Clinical experience with carinal resection. *Journal of Thoracic and Cardiovascular Surgery* 1999; **117**: 39–53. [A large series of carinal resections for a variety of indications, emphasizing the technical factors and morbidity associated with the procedure]

Mooney LA, Perera FP. Application of molecular epidemiology to lung cancer chemoprevention. *Journal of Cellular Biochemistry* 1996; **255**: 63–8. [Discusses molecular epidemiology and lung cancer in the context of defining a high-risk population.]

Mountain CF. Revisions in the international system for staging lung cancer. *Chest* 1997; **111**: 1710–17. [The revised TNM staging system for lung cancer.]

Mountain CF, Dresler CM. Regional lymph node classification for lung cancer staging. *Chest* 1997; **111**: 1718–23. [A unifying classification of mediastinal nodes for staging.]

National Research Council, Committee on Passive Smoking. *Environmental tobacco smoke: measuring exposures and assessing health effects*. National Academy Press, Washington, DC, 1986. [A review of the risks of passive smoking.]

National Research Council, Committee on the Biological Effects of Ionizing Radiation. *Health risks of radon and other internally deposited a-emitters: BEIR IV*. National Academy Press, Washington, DC, 1988. [A review of the data supporting radon as a lung carcinogen.]

Okubo K *et al.* Bronchoalveolar carcinoma: clinical, radiologic, and pathologic factors and survival. *Journal of Thoracic and Cardiovascular Surgery* 1999; **118**: 702–9. [An analysis of factors affecting survival in 119 patients with bronchoalveolar carcinoma.]

Pearson FG *et al.* Significance of positive superior mediastinal nodes identified at mediastinoscopy in patients with resectable cancer of the lung. *Journal of Thoracic and Cardiovascular Surgery* 1982; **83**: 1–11. [A classic paper demonstrating the adverse effect of mediastinal nodal metastases on survival, therefore underscoring the need for preoperative nodal staging.]

Piehler JM *et al.* Bronchogenic carcinoma with chest wall invasion: factors affecting survival following *en bloc* resection. *Annals of Thoracic Surgery* 1982; **34**: 684–91. [A large series of chest wall resections for lung cancer.]

Pignon JP, Arriagada R, Ihde DC. A meta-analysis of thoracic radiotherapy for small cell lung cancer. *New England Journal of Medicine* 1992; **327**: 1618–24. [A meta-analysis defining the benefits of concurrent radiotherapy for small cell lung cancer.]

Rosell R *et al.* A randomized trial comparing preoperative chemotherapy plus surgery with surgery alone in patients with non-small cell lung cancer. *New England Journal of Medicine* 1994; **330**: 1553–8. [One of two randomized trials of the use of adjuvant chemotherapy in stage IIIA lung cancer, demonstrating a benefit for adjuvant therapy.]

Sethi T. Lung cancer. *British Medical Journal* 1997; **314**: 652–5. [A concise review of neoplastic transformation.]

Shaw EG, Su JQ, Eagan RT. Prophylactic cranial irradiation in complete responders with small cell lung cancer: analysis of the Mayo Clinic and North Central Cancer Treatment Group databases. *Journal of Clinical Oncology* 1994; **12**: 2327–32. [A review of the results of treatment of over 1600 patients with small cell lung cancer.]

Shepherd FA *et al.* Is there ever a role for salvage operations in limited small-cell lung cancer? *Journal of Thoracic and Cardiovascular Surgery* 1991; **101**: 196–200. [A large clinical experience with resection of early stage, small cell lung cancer.]

Strauss GM, Gleason RE, Sugarbaker DJ. Screening for lung cancer. Another look; a different view. *Chest* 1997; **111**: 754–68. [A critical analysis of the four major studies of lung cancer screening from the last two decades, suggesting that a benefit from screening may exist.]

Tisi GM. Preoperative evaluation of pulmonary function. Validity risks and benefits. *Annual Review of Respiratory Disease* 1979; **119**: 293–315. [An excellent review of the aspects of pulmonary function pertinent to surgical patients.]

US Department of Health and Human Services. *Reducing the health consequences of smoking: 25 years of progress. A report of the Surgeon General*. US Government Printing Office, Washington, DC, 1989; DHEW (CDC) publication no. 89–8411. [A review of the epidemiology of smoking behaviors.]

Wain JC. Video-assisted thoracoscopy and the staging of lung cancer. *Annals of Thoracic Surgery* 1993; **56**: 776–8. [A description of the utility of thoracoscopic staging as an adjunct to other invasive staging modalities.]

Wronski M, Arbit E, Burt M, Galicich JH. Survival after surgical treatment of brain metastases from lung cancer: a follow-up study of 231 patients treated between 1976 and 1991. *Journal of Neurosurgery* 1995; **83**: 605–16. [A large series demonstrating 1- and 5-year survival rates of 46 and 13 per cent, respectively.]

41.4 Primary chest wall neoplasms

Edwin C. McGee, Jr., Richard I. Whyte, and Douglas Mathisen

Incidence
Clinical presentation
Diagnostic evaluation
Histologic evaluation
Pathology
Benign tumors
Malignant tumors
Soft tissue sarcomas
Principles of treatment
Techniques of chest wall reconstruction
Further reading

Incidence

Primary chest wall tumors are uncommon, accounting for 1 per cent of all neoplasms found throughout the body. As such, few large series have been published. Approximately 50 per cent of chest wall tumors are malignant but incidence varies from 9 to 80 per cent, depending on the series. The distinction between a primary tumor and metastatic or locally invasive lesions may not always be clear. In a series of 100 consecutive chest wall resections from the Mayo Clinic there were 68 primary tumors and 32 metastatic or locally invasive lesions. The most common benign lesions in most series are fibrous dysplasia and enchondroma. Although rates vary among series, chondrosarcoma is the most common primary chest wall malignancy. Other common malignant tumors include malignant fibrous histiocytoma, rhabdomyosarcoma, fibrosarcoma, and Ewing's sarcoma (Table 1).

	MSKCC	M.D. Anderson	MGH	Mayo Clinic	Walter Reed
Benign					
Enchondroma	–	2	4	2	19
Osteochondroma	–	–	4	–	–
Fibrous dysplasia	–	–	8	–	8
Eosinophilic granuloma	–	–	4	–	1
Desmoid	33*	3*	5	5	–
Other	–	3	30	12	23
Total	32	8	53	19	51
Malignant					
Chondrosarcoma	96	3	14	17	10
Osteosarcoma	41	5	5	3	–
Malignant fibrous histiocytoma	11	7	22	19	–
Fibrosarcoma	17	–	12	–	19
Rhabdomyosarcoma	18	2	3	0	3
Plasmacytoma	26	–	4	2	8
Ewing's sarcoma	65	6	8	4	6
Other	111	4	22	18	13
Total	388	27	90	71	79

*Considered benign in this series.
MSKCC, Memorial Sloan-Kettering Cancer Center; MGH, Massachusetts General Hospital.

Table 1 Histology of primary chest wall tumors

Clinical presentation

Chest wall tumors frequently present as an enlarging palpable mass or as an area of localized discomfort, although signs and symptoms are absent in 10 to 20 per cent of patients. A history of distant trauma or radiation may be present. Painful tumors tend to be malignant. Sternal tumors are generally malignant. Asymptomatic lesions are often detected on radiographs obtained for routine screening or following trauma. The differential diagnosis of chest wall lesions in addition to neoplasia includes non-specific chondritis (Tietze's syndrome), costochondral separation, congenital or developmental deformities such as pectus carinatum, and infectious processes.

Diagnostic evaluation

Patients with chest wall lesions should undergo a thorough history and physical. Routine laboratory screening tests, including alkaline phosphatase, serum calcium, and tests of renal function (creatinine and urea or blood urea nitrogen) should be obtained. Accurate radiographic assessment is essential in the evaluation of chest wall masses, especially those of bony or cartilaginous origin. Some lesions, such as chondrosarcoma, have distinctive radiographic characteristics, but many lesions yield non-specific radiographic findings. Standard posteroanterior and lateral chest radiographs should be obtained and supplemented with rib views. CT is useful in determining the origin and extent of the lesion, the involvement of adjacent structures, and the presence of pleural effusions, but it has not been shown to be of great benefit in establishing an initial diagnosis. Magnetic resonance imaging (**MRI**), although an invaluable tool in the preoperative assessment of extremity sarcomas because it provides information about involvement of neurovascular structures and fascial planes, is not routinely utilized in the radiologic assessment of chest wall tumors. CT, not MRI, is the most sensitive method for detecting pulmonary metastases. Chest wall and cardiac motion limit the accuracy of MRI, although newer generation machines do filter out motion more effectively. Some feel that MRI will eventually eclipse CT as the test of choice, but at most institutions this has not yet come to pass. Radionuclide bone scans are useful in detecting multiple lesions, spread to adjacent ribs, occult extremity primaries, and multiple myeloma. When myeloma is suspected, serum protein electrophoresis is indicated. Angiography can be useful in planning myocutaneous flaps for the reconstruction of large chest wall defects, and may be useful in assessing axillary artery involvement with tumors near the thoracic outlet.

Histologic evaluation

While clinical evaluation and radiographic investigations are important in managing chest wall neoplasms, accurate histologic analysis is the single most important factor in determining the appropriate treatment. The method for obtaining tissue for histologic analysis must be individualized for each patient. Ideally, the biopsy should be done at the center where definitive therapy will take place in order to make sure that it is planned in accordance with possible future resection.

Needle biopsy, while the least invasive, is also the least accurate and can produce misleading results. Fine needle aspiration is useful only in the evaluation of lesions suspected of being metastatic. Combining image guidance with core-needle biopsy improves accuracy. The use of image-guided core-needle biopsy is helpful when chemotherapy or radiation would be indicated as primary or preoperative therapy. If core-needle biopsy is not possible or is non-diagnostic, incisional biopsy should be performed on lesions greater than 4 cm. For smaller lesions, excisional biopsy should be performed, as it is curative for most benign processes. If the tumor is malignant or incompletely resected, definitive resection is undertaken as a second procedure. When performing an excisional biopsy for a small lesion, tumor localization may be difficult; the use of bone scans to identify and localize the lesion for placement of a skin marker can be helpful. The rib can also be injected with methylene blue after needle localization. When these techniques are used, a biopsy should be taken from the patient in the same position in which the scan was obtained.

A two-stage approach for diagnosis and definitive therapy avoids the use of unnecessarily wide excisions for certain tumors and permits co-ordination between multiple specialties, such as radiotherapy and plastic surgery, before an extensive resection is undertaken. Biopsy samples should be sent for routine histopathologic analysis, although more specialized immunohistochemical techniques or electron microscopy are sometimes helpful. Frozen sections are rarely used for making an initial diagnosis although they may be helpful in determining whether the biopsy specimen contains an adequate amount of viable specimen for evaluation.

Pathology

The histology of lesions from the 35-year experience of the Massachusetts General Hospital, which is similar to most reported series, is shown in Table 2. Our series includes tumors of the ribs, chest wall soft tissue, sternum, and clavicle. Most variations from other reports can be attributed to the rarity of these tumors, the heterogeneity of tumor types, the relatively long time frames over which data were collected, geographic locations from which patients were selected, and patterns of referral to the reporting institution.

Benign	n	*Malignant*	n
Elastofibroma dorsi	10	Malignant fibrous histiocytoma	22
Fibrous dysplasia	8	Chondrosarcoma	14
Schwannoma	7	Fibrosarcoma	12
Osteochondroma	6	Ewing's sarcoma	8
Eosinophilic granuloma	4	Malignant schwannoma	7
Enchondroma	4	Osteosarcoma	5
Osteoid osteoma	1	Liposarcoma	5
Neurofibroma	3	Plasmacytoma	4
Hemangioma	2	Angiosarcoma	3
Desmoid	3	Rhabdomyosarcoma	3
Aneurysmal bone cyst	1	Hemangiopericytoma	2
Other	6	Synovial sarcoma	2
		Spindle cell sarcoma	2
		Leiomyosarcoma	1
Total	53		90

Table 2 Massachusetts General Hospital series of chest wall neoplasms (1963 to 1997)

Benign tumors

Fibrous dysplasia (bone cyst, osteofibroma, fibrous osteoma, fibrosis ossificans)

Fibrous dysplasia accounts for 30 per cent of benign chest wall lesions. It is a non-neoplastic process that often presents as an asymptomatic mass, but chest wall pain may also occur. The rib cage is the most frequent site of this slowly growing lesion. Adkins claims that fibrous dysplasia generally presents in the posterior or lateral portions of the rib. Radiographically, the lesion demonstrates cortical expansion and thinning with a trabeculated lytic center. Since the results of needle biopsy are unreliable, the diagnosis should be made by excisional biopsy. Microscopically, the lesion consists of fibrous tissue with poorly oriented and incompletely mineralized bone trabeculas. Resection of the lesion provides symptomatic relief and recurrence is rare. Polyostotic lesions occur in several bones and are a component of Albright's syndrome. There is no malignant potential.

Eosinophilic granulomatosis

The chest wall can be affected by one of the syndromes of eosinophilic granulomatosis, which are non-neoplastic diseases of the reticuloendothelial system collectively known as histiocytosis X. Eosinophilic granuloma of bone is limited to bony involvement, while Letterer–Siwe disease and Hand–Schüller–Christian disease may have systemic ‘B' symptoms and diffuse visceral organ involvement. Radiographically, eosinophilic granulomas appear as expansile, lytic processes with cortical thinning and periosteal new bone formation. The microscopic appearance is one of a reticulum of histiocytes infiltrated with eosinophilic leukocytes, which may have bilobed or hyperchromatic nuclei. The eosinophilic cells may be missing and replaced by lymphocytes, plasma cells, or neutrophils. Eosinophilic granuloma is cured by resection of the affected rib, curettage, radiation, or by one of several antimetabolic drugs. Osteomyelitis should always be in the differential diagnosis.

Elastofibroma dorsi

This rare benign tumor usually arises adjacent to the inferior angle of the scapula. The lesion, which is frequently asymptomatic, is generally considered to be a disease of the elderly, but has been reported in a patient as young as 6 years. The chest radiograph shows a soft tissue mass without rib involvement; CT often demonstrates an unencapsulated subscapular mass. Clinically, the lesion may resemble a malignant tumor being densely adherent to the muscle, rib periosteum, or scapula. On gross inspection, the lesion is a fibrofatty mass without a defined capsule; the microscopic appearance is of hypercellular fibrous tissue with irregular, thick, wavy elastic fibers that may require staining for elastic tissue to establish their identity. Since the treatment of elastofibroma dorsi consists of simple excision it must be differentiated from the more cellular and pleiomorphic low-grade fibrosarcoma or desmoid tumor, which requires a much wider resection margin.

Chondromas/osteochondromas

Chondromas, or enchondromas, are hamartomas of cartilage. They account for 15 to 43 per cent of benign rib tumors, generally occur in young adults, and are frequently located at the costochondral junction. Radiographically, they are well defined, ovoid, expansile lesions causing cortical thinning. The histologic appearance is one of cartilage with variable cellularity. Because the differentiation between chondroma and chondrosarcoma on a clinical and radiographic basis can be impossible, suspected chondromas should be managed with wide excision.

Osteochondroma is the most common benign bony neoplasm affecting the chest wall. They often present in children and adolescents as painless masses. Radiographically, a calcified rim is noted and stippled calcification appears throughout the tumor. Management is by local excision.

Desmoid tumors

Desmoid tumors originate in the fascia or muscle and are frequently located in the shoulder or chest wall. They may invade local tissue and have a high incidence of local recurrence even when negative margins are obtained. There is controversy as to whether desmoids are benign or malignant. Some feel they are a form of benign fibromatosis, while others consider them to be low-grade fibrosarcomas. Management consists of wide excision or enucleation with radiation, when vital structures preclude complete resection. Ten-year survival was 93 per cent for the 32 patients in the Memorial Sloan-Kettering Cancer Center (**MSKCC**) series. Unresectable, enlarging lesions are often responsive to irradiation by growth arrest or slow regression.

Malignant tumors

Chondrosarcomas

Chondrosarcoma is the most common primary malignant tumor of the chest wall, accounting for 20 to 50 per cent of all primary chest wall malignancies. Most chondrosarcomas arise *de novo* from the cartilaginous areas of the anterior chest wall, although malignant degeneration of previously benign lesions may occur. Chondrosarcomas are unusual in patients under 30 years of age. Ipsilateral major chest wall trauma has been reported in 12.5 per cent of patients with rib chondrosarcoma. The importance of trauma in the etiology of this condition is not clear.

A review of chest wall chondrosarcomas from the Mayo Clinic included 55 men and 41 women. The mean age was 53.5 years (range 17 to 78 years). Most of the tumors (81 per cent) arose in the ribs; 18 per cent originated in the sternum. Ninety-four per cent of patients had physical signs or symptoms with more than 50 per cent having had them for longer than 1 year. Twelve patients had a history of major ipsilateral chest wall trauma.

The diagnosis of chondrosarcoma is best made by pathologic evaluation after excision since the clinical signs, radiographic characteristics, and histologic appearance of incisional biopsies may be identical to those of benign cartilaginous tumors. The radiographic appearance of chondrosarcomas arising in flat bones such as ribs or the sternum is usually one of a soft tissue mass with bony destruction and varying amounts of new bone formation. There is often mottled calcification and the soft tissue component may be large in proportion to the size of the skeletal lesion. Mesenchymal chondrosarcoma may not be associated with calcification because of the rapid growth rate of the tumor. Grossly, chondrosarcomas appear as a cartilaginous, lobulated, firm mass. Microscopically, chondrocytes may be multinucleate or they may be atypical, with large, heavily chromatized nuclei. Areas of myxomatous change may be identified. Although difficult to do, chondrosarcomas are graded from I to III based on their histologic appearance as defined by pleiomorphism, cellularity, and nuclear atypia.

The natural history of most chondrosarcomas is one of slow growth and local recurrence, with the exception of the relatively uncommon high-grade lesions which grow quickly and metastasize early. In a study of 88 patients over 40 years at the MSKCC, improved survival was associated with absence of metastasis, age less than 50 years, complete resection, and absence of local recurrence. Tumor grade, size, site, and sex of the patient were not significant factors. The overall 5-year survival was 64 per cent. The Mayo Clinic series found that grade and tumor size are important factors (Table 3). Of the 72 patients undergoing primary resection in the Mayo series, 66 had borderline, grade I, or grade II lesions. The overall 10-year actuarial survival rate was 65 per cent. Local recurrence developed in 51 per cent of patients treated by resection.

	%
10-year survival according to tumor size	
<6cm	87
6–10cm	63
>10cm	31
10-year survival according to tumor grade	
Grade I	71
Grade II	39
Grade III	0*
10-year recurrence according to degree of resection	
Wide excision	17
Local excision	51
Palliative resection	94

* All patients died within 2 years.

Table 3 Long-term prognosis of chondrosarcoma

The treatment of chondrosarcoma is based on surgical resection. Adjunctive radiotherapy or chemotherapy have not been shown to be of benefit, but their use in patients with high-grade tumors may be warranted as the prognosis following resection of these lesions is so poor.

Osteosarcoma

Osteosarcomas rarely occur in the chest wall and comprise 10 per cent of malignant chest wall tumors in the 40-year experience of the MSKCC. Most presented as painful masses. Thirty per cent of patients had received radiation for prior malignancies such as Hodgkin's lymphoma and breast cancer. The extent of disease at presentation, resectability, and the development of metastasis were all prognostic factors. Age, sex, tumor site, and size were not of statistical significance. Wide excision is the cornerstone of treatment. Adjuvant chemotherapy protocols similar to those used to treat extremity osteosarcomas should be utilized, although prognosis does not appear as favorable. Overall 5-year survival was 15 per cent.

Ewing's sarcoma (primitive neuroectodermal tumor/Askin's tumor)

This unusual tumor, seen primarily in young patients, accounts for 8 to 24 per cent of all malignant chest wall tumors. Of all cases of Ewing's sarcoma in the combined experiences of the National Cancer Institute, the MD Anderson Hospital, and the Mayo Clinic, 6.5 per cent were primary rib lesions. In one review of 36 patients with primary Ewing's sarcoma of the ribs, 21 had localized disease at the time of diagnosis, 8 had regional disease, and 7 had metastatic disease. The patients ranged from 4 to 25 years of age (mean 13); there were 12 men and 9 women. The predominant symptom was chest pain which had been present for 5 to 235 days (mean 36) prior to evaluation. The role of surgical resection is controversial in the management of chest wall Ewing's sarcoma. Some feel that chest wall resection is not indicated, and that chemotherapy and radiation are sufficient primary therapy. Retrospective studies of extremity Ewing's sarcoma demonstrate a marked survival advantage with surgical resection. A recent series of 15 patients with chest wall Ewing's sarcoma from the Boston Children's Hospital supports the role of surgical therapy in this disease. Diagnosis was made with CT-guided core-needle biopsy, followed by preoperative chemotherapy (with vincristine, actinomycin D, cyclophosphamide, and doxorubicin), chest wall resection, and postoperative radiotherapy. Such an approach allows for a smaller chest wall resection, by reducing tumor size and vascularity, and lessens the dose of radiotherapy to the thorax. Six of ten patients without metastasis were alive and without local recurrence with median follow-up of 5 years. The overall 5-year survival in the 40-year experience of the MSKCC was 48 per cent, with the development of metastasis being the only significant predictor of survival. Local recurrence was not a significant predictor of survival, and local control was achieved equally well with surgery or radiation.

Soft tissue sarcomas

Soft tissue sarcomas account for 25 per cent of primary chest wall malignancies. In the 40-year experience of the MSKCC, 149 were reported (32 desmoids are included in this series). A painless mass was the most common presentation. Five per cent of patients had received prior chest wall radiation. The most common subtypes at the MSKCC were liposarcoma (15 per cent), rhabdomyosarcoma (12 per cent), and fibrosarcoma (11 per cent); however, percentages vary from institution to institution. Malignant fibrous histiocytoma is the most common subtype in the Massachusetts General Hospital and Mayo Clinic series.

The treatment of soft tissue sarcomas generally consists of wide local excision for low-grade lesions, supplemented with radiation and chemotherapy for high-grade tumors. An exception to this treatment regimen is rhabdomyosarcoma. This is a rapidly growing tumor of skeletal muscle that is generally seen in children and young adults. Rhabdomyosarcoma is unique in that radiation and chemotherapy produce very good long-term survival—70 per cent at 5 years, although prior to the use of combination therapy the long-term survival was poor. The role of surgery in this condition is controversial: Pairolero recommends wide excision while Pass maintains that surgery should be limited to generous biopsy and removal of residual disease after courses of radiation therapy and chemotherapy. In the MSKCC series, the overall 5- and 10-year survivals for soft tissue sarcomas are 66 and 56 per cent, respectively. Desmoids were considered malignant in this series, which may in part explain the overall good survival rate. Prognostic factors include tumor grade, the presence of metastasis, histologic subtype, and the presence of pain. In the MSKCC experience, tumors of fibrous and connective tissue subtypes were less lethal that those of muscle subtypes. The age of the patient, size of the lesion, and development of local recurrence were not statistically significant variables affecting survival.

Plasmacytomas

Solitary plasmacytomas most often present with chest wall pain and radiographically appear as ‘punched out' lytic lesions. Surgery and radiation achieve local control equally well. Overall 5-year survival at the MSKCC was 38 per cent and hinges on the development of multiple myeloma, which occurs in 75 per cent of patients. Systemic therapy is then required.

Principles of treatment

Most patients with chest wall neoplasms require surgical resection as part of their treatment. The extent of resection depends on the size and type of the neoplasm: benign lesions generally require only limited resection with clear margins. Malignant lesions usually require resection with wide margins. Opinions differ with respect to the extent of resection for malignant lesions. Pass has recommended complete resection of any involved rib with its cartilaginous articulations as tumor extension within the bone marrow cannot be predicted, while Adkins has advocated routine resection of the ribs above and below the involved ribs in order to gain adequate margins. Others recommend resection of the lesion alone, with a 4- to 5-cm margin of normal tissue; Pairolero has amended this to include removal of the entire involved rib for high-grade malignancies.

Analysis of frozen sections may be helpful in assessing resection margins, but may be difficult to interpret. For sternal or manubrial tumors, the entire bone and adjacent costal arches must be removed. Any adherent tissue, such as anterior mediastinal fat, thymus, pleura, or lung, should be removed *en bloc* with the chest wall specimen. The limits of resection should be determined by adequacy of margins rather than by attempts to preserve chest wall, as extensive chest wall resections can be performed safely and with good functional and cosmetic results.

The incision chosen for resection is determined by the size and location of the tumor, and may be a standard thoracotomy or a variant thereof. Vertical incisions for incisional biopsy should be avoided because subsequent wide resection, which must include the biopsy incision, will be difficult. The exception to this general rule is that median sternotomy is appropriate for resection of sternal lesions or when resection of bilateral pulmonary metastases is planned. At the Massachusetts General Hospital, video-assisted thoracoscopy has been used to resect selected pleural-based lesions. Great care must be taken not to violate the tumor and adequate margins must be achieved. When planning the incision, the anticipated type of wound closure, either by primary closure or chest wall reconstruction, should be taken into consideration.

Techniques of chest wall reconstruction

Reconstruction of a chest wall defect can be difficult. The primary consideration is usually the size and location of the bony and soft tissue defect, although factors such as preoperative pulmonary function, previous chest wall surgery or radiotherapy, and cosmetic considerations can be important. Reconstruction of segments of the chest wall is performed to protect underlying structures, to obtain chest wall rigidity and fixation for effective respiratory effort, as well as to prevent flail segments, reduce paradox, and prevent herniation (Table 4).

Anterior lesions
Defect greater than 5cm
Sternal defects
Patients with marginal pulmonary reserve
Posterior lesions
Defect greater than 10cm
Resection of ribs adjacent to scapular tip (5th or 6th rib)
Patients with marginal pulmonary reserve

Table 4 Indications for chest wall reconstruction

Small defects (those involving fewer than three ribs) can usually be closed primarily; those greater than 5 cm often require chest wall reconstruction as a large flail segment would otherwise result. Location is important. Posterior defects up to 10 cm in diameter can be tolerated as long as the scapula remains to provide stability. Generally, resection of the first to fourth ribs posteriorly is well tolerated, but if the resection also involves the fifth and sixth ribs, additional support is necessary to prevent the scapula from becoming caught below the lower ribs. Resections which include the costal margin can be reconstructed by suturing the diaphragm to the lowest remaining rib; this is not possible if resection extends above the fifth rib. Either the diaphragm will not reach high enough or the remaining hemithorax will be too small. Defects involving the sternum should be reconstructed to protect underlying structures and to prevent the paradoxical motion that occurs with removal of the anterior costal attachments.

Preoperative assessment of patients requiring major chest wall resection should include pulmonary function testing since a pulmonary resection may be necessary and a period of mechanical ventilation may be required postoperatively. Preoperative treatment with Adriamycin may cause impaired cardiac function. Radiation and bleomycin may cause pulmonary fibrosis. Patients with severe respiratory limitations may benefit from chest wall reconstruction following the creation of defects in any area, as the margin for error in these patients may be quite small and any further decrease in pulmonary reserve may be severely detrimental.

Patients who have undergone previous chest wall surgery or radiotherapy present special concerns in that the vascular supply to the tissue remaining after chest wall resection must be considered prior to chest wall resection. Good vascular supply is particularly important if prosthetic material is to be implanted; adjacent poorly vascularized or irradiated tissue may significantly increase the risks of infection.

Chest wall reconstruction is aimed at recreating the bony rigidity and soft tissue coverage of the native chest wall. Synthetic material and autogenous tissue, such as bone grafts and soft tissue flaps, have been used to provide both rigidity and protection of underlying structures. Historically, materials such as steel, tantalum, Ivalon, Lucite, and fiberglass have been used. Most of these create such a tissue reaction that they are rendered unusable; they have, therefore, been replaced by less reactive materials. Marlex mesh, which has been used since 1960, has been the standard synthetic material for reconstructing large defects. It can be used without additional support or can be reinforced with methyl methacrylate. This may be fashioned outside the chest wall defect or may be assembled *in situ.* The rigid methyl methacrylate component must be made smaller than the chest wall defect to provide a cuff of Marlex which can be sewn to the adjacent ribs. Too large a prosthesis may result in unnecessary pain and stiffness. Methyl methacrylate–Marlex mesh prostheses are well suited to reconstruction of sternal defects because they provide the necessary rigidity to protect the underlying heart and great vessels. Reported problems with Marlex and methyl methacrylate include a transient metabolic acidosis during the curing phase of the plastic, and seroma formation, although neither has been a problem in our experience.

More recently, 2-mm thick polytetrafluoroethene (**PTFE** or Gore-Tex™), polypropylene (Prolene™) mesh, and Dexon™ mesh have been used with good results; some surgeons prefer them to Marlex. PTFE is impermeable to both air and water, is easy to work with, but lacks rigidity. Prolene differs from Marlex in that it has multidirectional rigidity; Marlex is rigid in only one direction.

Although infection is a potential problem when any synthetic material is used, these patches rarely need to be removed for the treatment of infection.

Prosthetic materials provide a rigid base for reconstructing chest wall defects, but the overlying soft tissue must also be replaced. The pectoralis major, latissimus dorsi, and rectus abdominus myocutaneous flaps, as well as greater omentum are all available for use.

The pectoralis major myocutaneous flap is based on the pectoral branch of the thoracoacromial artery. The pectoralis, taken as a muscle flap or with its overlying skin as a myocutaneous flap, can be mobilized to provide coverage to the neck and upper sternum; it has frequently been used to provide closure for upper chest and neck wounds. It is also suitable for covering smaller defects of the anterior and lateral chest wall. The lateral tendinous attachments may be divided if additional mobility is necessary; it is also possible to mobilize the pectoralis muscle based on perforators from the internal mammary artery and to rotate the muscle medially to cover sternal defects. The pectoralis flap usually requires prosthetic material for underlying structural support and a split thickness skin graft to close the donor site.

The latissimus dorsi flap is based on the thoracodorsal artery, a branch of the subscapular artery. The flap, which is generally used as a myocutaneous flap, can support an island of skin up to 10 × 16 cm in area. It can cover larger areas than the pectoralis flap and, with a long vascular pedicle, can be mobilized to cover chest wall defects over a wide area. It may not, however, be able to reach to the inferior costal margin in the area of the xiphoid. A reverse latissimus flap has been described, based on branches of intercostal and lumbar arteries, in which the muscle is rotated medially to cover contralateral posterior defects. As with the pectoralis flap, skin grafting is usually required for closure of the donor site.

The transverse rectus abdominis flap, which is frequently used for breast reconstruction, is also excellent for covering large chest wall defects. This bipedicle flap, based on the inferior epigastric artery and superior epigastric artery, a branch of the internal mammary artery, can be mobilized on either of these vessels to cover defects as large as an entire anterior hemithorax from the costal margin to the clavicle. When used to cover defects created by removal of the sternum, the flap offers such stability that additional prosthetic material is not necessary. Three other possible muscle transposition flaps include the serratus anterior, trapezius, and external oblique flaps, which all offer coverage to only limited areas of the chest wall.

Greater omentum, based on either the right or left gastroepiploic arteries, can be brought from the abdomen to almost any location in the thorax and used to provide soft tissue coverage of prosthetic material. Split thickness skin grafts can be placed directly on the omentum. This technique is frequently less than ideal cosmetically, and is generally used when other flaps are either unavailable or have been unsuccessful. Omental flaps are especially useful in sites where residual infection may be present.

Free vascularized tissue flaps, created using microsurgical techniques, have recently been used to provide soft tissue coverage of chest wall defects. The tensor fascia lata free flap, which consists of the tensor fascia lata muscle with its overlying fascia and skin, is based on the lateral femoral circumflex vessels. These vessels are anastomosed to either the thoracoacromial or transverse cervical vessels. It is also possible to use vascularized fibular bone grafts to replace missing ribs.

Any tissue being transferred must be well vascularized. Particular consideration must be given to radiation for carcinoma of the breast which is directed to the internal mammary and axillary nodes; these are adjacent to the vascular pedicles of pectoralis and latissimus flaps, respectively. Although prior irradiation does not prohibit use of that flap, preoperative arteriography may be useful in assessing the adequacy of the blood supply.

Further reading

Adkins PC. Tumors of the chest wall. In: Shields TW, ed. *General thoracic surgery*, 2nd edn, pp. 453–63. Lea & Febiger, Philadephia, 1983.

Anderson BO, Burt ME. Chest wall neoplasms and their management. *Annals of Thoracic Surgery* 1994; **58**: 1774–71.

Bostwick J *et al.* The ‘reverse' latissimus dorsi muscle and musculocutaneous flap: anatomical and clinical considerations. *Plastic and Reconstructive Surgery* 1980; **65**: 395–9.

Boyd AD *et al.* Immediate reconstruction of full thickness chest wall defects. *Annals of Thoracic Surgery* 1981; **32**: 337–46.

Burt M. Primary malignant tumors of the chest wall: The Memorial Sloan-Kettering Cancer Center Experience. *Chest Surgery Clinics of North America* 1994; **4**: 137–54.

Burt M *et al.* Primary bony and cartilaginous sarcomas of chest wall: results of therapy. *Annals of Thoracic Surgery* 1992; **54**: 226–32.

Burt M *et al.* Medical tumors of the chest wall. Solitary plasmacytoma and Ewing's sarcoma. *Journal of Thoracic and Cardiovascular Surgery* 1993; **105**: 89–96.

Cavanaugh DG, Cabellon S, Peake JB. A logical approach to chest wall neoplasms. *Annals of Thoracic Surgery* 1986; **41**: 436–7.

Chang AE, Rosenberg SA, Glatstein EL, Antman KH. Sarcomas of soft tissue. In: DeVita VT, Hellman S, Rosenberg SA, eds, *Cancer: principles and practice of oncology*, 3rd edn, pp. 1244–92. JB Lippincott, Philadelphia, 1989.

Enzinger FM, Weiss SW. *Soft tissue tumors*, 2nd edn, pp. 122–7, 201–22. CV Mosby Co., St Louis, 1988.

Fortier M, Mayo JR, Swensen SJ, Munk PL, Vellet PA, Muller NJ. MR imaging of chest wall lesions. *Radiographics* 1994; **14**: 597–606.

Gordon MS, Hajdu SI, Bains MS, Burt ME. Soft tissue sarcomas of the chest wall. *Journal of Thoracic and Cardiovascular Surgery* 1991; **101**: 843–54.

Graeber GM *et al.* Initial and long-term results in the management of primary chest wall neoplasms. *Annals of Thoracic Surgery* 1982; **34**: 664–73.

Graham J, Usher FC, Perry JL, Barkley HT. Marlex mesh as a prosthesis in the repair of thoracic wall defects. *Annals of Surgery* 1960; **151**: 469.

King RM, Pairolero PC, Trastek VF, Piehler JM, Payne WS, Bernatz PE. Primary chest wall tumors: factors affecting survival. *Annals of Thoracic Surgery* 1986; **41**: 597–601.

McAfes MK *et al.* Chondrosarcoma of the chest wall: factors affecting survival. *Annals of Thoracic Surgery* 1985; **40**: 535–41.

McKenna RJ, Mountain CF, McMurtrey MJ, Larson D, Stiles QR. Current techniques for chest wall reconstruction: expanded possibilities for treatment. *Annals of Thoracic Surgery* 1988; **46**: 508–12.

Marin ML, Perzin KH, Markowitz AM. Elastofibroma dorsi: benign chest wall tumor. *Journal of Thoracic and Cardiovascular Surgery* 1989; **98**: 234–8.

Melawar MM, Link MP, Donaldson SS. Sarcomas of bone. In: DeVita VT, Hellman S, Rosenberg SA, eds, *Cancer: principles and practice of oncology*, 3rd edn, pp. 1293–342. JB Lippincott, Philadelphia, 1989.

Morgan RF, Edgerton MT, Wanebo HJ, Daniel TM, Spotnitz WD, Kron IL. Reconstruction of full thickness chest wall defects. *Annals of Surgery* 1988; **207**: 707–16.

Pairolero PC. Chest wall tumors. In: Shields TW, ed., *General thoracic surgery*, 4th edn. Lea & Febiger, Philadelphia, 1994.

Pairolero PC, Arnold AG. Chest wall tumors—experience with 100 consecutive patients. *Journal of Thoracic and Cardiovascular Surgery* 1985; **90**: 367–72.

Pass HI. Primary and metastatic chest wall tumors. In: Roth JA, Ruckdeschel JC, Weisenburger TH, eds, *Thoracic oncology*, pp. 546–65. WB Saunders, Philadelphia, 1989.

Rami-Porta R, Bravo-Bravo JL, Aroca-Gonzalez MJ, Alix-Treuba A, Surrano-Munoz F. Tumours and pseudotumours of the chest wall. *Scandinavian Journal of Thoracic and Cardiovascular Surgery* 1985; **19**: 97–103.

Ryan MB, McMurtrey MJ, Roth JA. Current management of chest-wall tumors. *Surgical Clinics of North America* 1989; **69**: 1061–80.

Sabanatham S, Salama FD, Morgan WE, Harvey JA. Primary chest wall tumors. *Annals of Thoracic Surgery* 1985; **39**: 4–15.

Sailer SL, Harmon DC, Mankin HJ, Truman JT, Suit HD. Ewing's sarcoma: surgical resection as a prognostic factor. *International Journal of Radiation Oncology, Biology, Physics* 1988; **15**: 43–52.

Shamberger RC, Tarbell NJ, Perez-Atayde AR, Grier HE. Malignant small round cell tumor (Ewing's-PNET) of the chest wall in children. *Journal of Pediatric Surgery* 1994; **29**: 179–85.

Thomas PRM *et al.* Primary Ewing's sarcoma of the ribs. *Cancer* 1983; **51**: 1021–7.

41.5 Mediastinal tumours

Henning A. Gaissert and Douglas Mathisen

Introduction
Anatomy
Diagnosis
Radiology
Serologic tumor markers
Biopsy
Thymic tumors
Mediastinal lymphoma
Germ cell tumors
Neurogenic tumors
Autonomic ganglion tumors
Nerve sheath tumors
Mediastinal cysts
Thoracoscopy
Further reading

Introduction

A distinct entity only by their anatomic location, mediastinal tumors comprise a variety of cysts and primary neoplasms. Correct diagnosis and treatment require accurate radiologic assessment, an understanding of the natural course of these tumors, and an established plan for evaluation and treatment. Since resection is not always the appropriate form of therapy for this diverse group of tumors, the principal goals of surgical involvement need to be stated clearly. (1) Is complete resection necessary or will biopsy without thoracotomy be more appropriate? (2) What approach should be used to accomplish the surgical procedure? (3) How will radiation therapy or chemotherapy alter the surgical management?

Anatomy

The boundaries of the mediastinum are the thoracic inlet superiorly, the diaphragm inferiorly, and laterally the medial surfaces of the parietal pleura. Three arbitrary compartments within this space have been defined (Fig. 1). The anterosuperior compartment is bounded by the sternum anteriorly and the anterior surface of the pericardium posteriorly. The middle or visceral compartment encompasses the space occupied by the heart, great vessels, trachea, main bronchi, and esophagus. The posterior compartment contains the prevertebral space and the paravertebral gutters. A separate superior compartment is occasionally defined as extending between the angle of Louis and the fourth vertebra to the thoracic inlet.

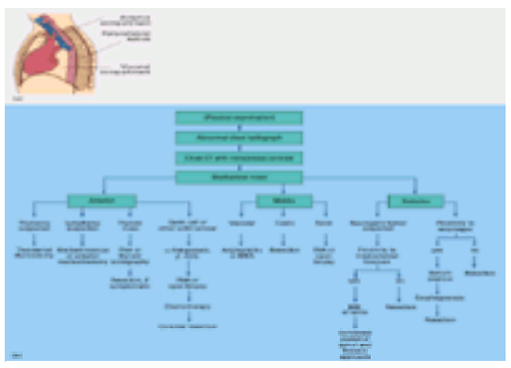


Fig. 1. (a) The three compartments of the mediastinum. (b) Diagnostic algorithm for mediastinal tumors.

This compartmentation is useful because most lesions occur only or predominantly in one of these regions. Table 1 lists the typical location of common mediastinal tumors and cysts. Thymomas, lymphomas, and germ cell tumors are most common in the anterior compartment, cysts predominate in the visceral space, and neurogenic tumors and cysts occur with equal frequency in the posterior compartment. A recent study of 400 consecutive patients found the anterosuperior mediastinum to be the site of cysts and neoplasms in 54 per cent of patients, the posterior mediastinum in 26 per cent, and the middle compartment in 20 per cent.

Anterior compartment	*Visceral compartment*	*Paravertebral sulci*
Thymoma	Enterogenous cyst	Neurilemoma–schwannoma
Germ cell tumor	Lymphoma	Neurofibroma
Lymphoma	Pleuropericardial cyst	Malignant schwannoma
Lymphangioma	Mediastinal granuloma	Ganglioneuroma
Hemangioma	Lymphoid hamartoma	Ganglioneuroblastoma
Lipoma	Mesothelial cyst	Neuroblastoma
Fibroma	Neuroenteric cyst	Paraganglioma
Fibrosarcoma	Paraganglioma	Pheochromocytoma
Thymic cyst	Pheochromocytoma	Fibrosarcoma
Parathyroid adenoma	Thoracic duct cyst	Lymphoma
Aberrant thyroid		

Table 1 Usual location of the common primary tumors and cysts of the mediastinum.

Diagnosis

The diagnostic approach is guided by the location of the mass and further inferences derived from its appearance on computed tomography (**CT**). Figure 1(b) depicts an algorithm useful in the evaluation of mediastinal tumors.

Radiology

Localization is primarily accomplished with radiologic tests. In addition to standard chest films, CT is routinely performed to give precise information about anatomic relationships. CT is used to assess whether the mass is homogenous or consists of multiple enlarged lymph nodes. In planning treatment and operative approach, the answer to two questions is of critical importance and may require additional studies.

Is the mass vascular?

CT with bolus injection of contrast may often avoid angiography by excluding aneurysms and vascular tumors. Angiography may be necessary to define arch vessel involvement, the feeding vessel of a paraganglioma, sequestration, or a mediastinal vascular malformation. The origin of the spinal blood supply, the artery of Adamkiewicz, should be angiographically demonstrated in paravertebral tumors involving the lower thoracic spine.

Is the spinal canal involved?

Approximately 10 per cent of all benign nerve sheath tumors cause extradural compression of the spinal cord due to extension of the tumor through the intervertebral foramen into the spinal canal. Because such dumb-bell-shaped tumors require a combined thoracic and neurosurgical approach for safe one-stage resection, preoperative evaluation of tumor extent with CT, linear tomography, myelography, or magnetic resonance imaging (**MRI**) is important. The three advantages associated with MRI are absence of X-ray exposure, no need for intravenous or intraspinal contrast, and the availability of sagittal cuts through the chest that provide further information about the mass. MRI and ultrasound may help to identify the cystic nature of mediastinal tumors. Venograms are occasionally useful to assess patency of the innominate vein or the superior vena cava. Fluoroscopy of the diaphragm allows functional evaluation of the phrenic nerve.

Serologic tumor markers

Various polypeptides, hormones, and enzymes occur in association with mediastinal tumors. For the purpose of this discussion, remarks are restricted to serologic markers of non-seminomatous, malignant, germ cell tumors.

Human chorionic gonadotropin (**hCG**) is a glycoprotein with a molecular weight of 45 000 and consists of a- and b-subunits. The b-chain is unique to hCG. In the normal placenta, hCG is produced by the syncytiotrophoblast and maintains the corpus luteum during early pregnancy. The oncologic usefulness of b-hCG as a tumor marker is derived from its association with choriocarcinoma and, less often, embryonal carcinoma, yolk sac tumor, and seminoma.

a-Fetoprotein is a glycoprotein with a molecular weight of 64 000 that occurs in liver, gastrointestinal tract, and yolk sac of the normal fetus. a-fetoprotein decreases to adult levels by 1 year of age. This tumor marker is increased in the serum of patients with hepatocellular cancer, embryonal carcinomas, and yolk sac tumors.

Immunohistochemical staining and localization of b-hCG and a-fetoprotein in a tumor can be employed to diagnose specific germ cell tumors.

Well over 90 per cent of patients with non-seminomatous germ cell tumors are marker positive for either b-hCG or a-fetoprotein. These two markers not only help in the diagnosis of germ cell tumors, but also allow monitoring of chemo- and radiotherapy. Return of marker levels to normal during chemotherapy before resection is a favorable prognostic factor. Elevated levels are a reliable indicator of viable tumor.

Biopsy

For some tumors, chemotherapy or radiotherapy are the primary modes of treatment. Tissue assessment by fine needle aspiration or mediastinoscopy avoids thoracotomy and reduces any delay of therapy. Specific tumors have characteristic findings on CT (multiple enlarged lymph nodes in lymphoma) or elevated serum tumor markers (non-seminomatous germ cell tumors) that are highly suggestive of the diagnosis. Considerable judgement is required in assessing a homogenous anterior mediastinal mass without obvious lymph node enlargement and negative serum markers. If thymoma is suspected and the tumor appears resectable, no further diagnostic evaluation is necessary prior to resection. Percutaneous aspiration for cytology should be reserved for patients with massive neoplasms suspected to be thymomas, in whom preoperative radiotherapy may make the lesion more amenable to resection.

Percutaneous fine needle aspiration under fluoroscopy or CT guidance provides a cytologic aspirate of tumor. The risk to the patient is minimal and consists of the development of pneumothorax and hemorrhage from a vascular tumor. A limited amount of material is available for histologic examination, making subtyping of lympho-mas difficult. The primary role of mediastinoscopy lies in the evaluation of paratracheal and subcarinal lymphadenopathy. Biopsy examination of presumed mediastinal lymph nodes is performed only after mediastinoscopic needle aspiration to exclude the possibility of a vessel appearing as a lymph node. Anterior mediastinotomy, the Chamberlain procedure, offers access to the aortic window and the anterior mediastinum. The usefulness of thoracoscopy has not been established. This technique may provide access for visual inspection and biopsy examination through the pleural space.

Thymic tumors

Neoplasms of thymic origin account for approximately one-third to one-half of all tumors in the anterosuperior compartment and occur infrequently outside the mediastinum. Thymic cysts and hyperplasia are occasionally reported in children. Carcinoid tumors, thymic lymphoma, Hodgkin's disease, and thymic carcinoma are rare disorders. Thymomas, by far the most common thymic tumors, originate from thymic epithelial cells and are divided into four subgroups according to the mixture of epithelial cells and lymphocytes: (a) predominantly epithelial, (b) predominantly lymphocytic, (c) mixed lymphoepithelial, and (d) spindle cell variety, a subtype of epithelial thymoma. A recent classification separates cell types as cortical, medullary, or mixed on the basis of immunohistochemistry and light microscopy. While some data suggest improved prognosis for medullary and mixed cell types, the accuracy of this classification in predicting relapse and survival is not established.

Various autoimmune disorders are associated with thymoma. Myasthenia gravis is the syndrome most commonly encountered. Its clinical hallmark is an abnormal fatiguability of voluntary muscle groups due to the presence of circulating antibodies to the acetylcholine receptor at the neuromuscular junction. Symptoms respond to blockage of cholinesterase inactivation of acetylcholine. Medical treatment consists of oral administration of pyridostigmine bromide and, in severe cases, steroids. A thymoma is present in 10 per cent of patients with myasthenia gravis and thymic hyperplasia in 70 to 80 per cent. Conversely, myasthenia gravis is encountered in 30 to 40 per cent of patients with thymoma. Because of improvement in the management and the favorable effects of plasmapheresis in severe myasthenia, the presence of myasthenia in patients with thymoma is no longer considered an adverse factor in survival. Removal of the thymus gland results in a partial or complete remission of myasthenia gravis documented by a reduction or elimination of medication in 60 to 80 per cent of patients.

Thymic lesions are rare in children below 16 years of age; peak incidence is between the ages of 50 and 70 years. Symptoms are present in approximately half of all patients without neuromuscular syndromes. These are chest pain, cough, dyspnea, respiratory infections, fever, and occasionally obstruction of the superior vena cava due to a large tumor. In asymptomatic patients, a chest film obtained for other reasons may demonstrate the anterior mediastinal mass.

In most patients, resection should be done without prior biopsy because of the risk of seeding the biopsy tract. Tumors confined to the thymus are best approached by median sternotomy to allow complete resection of the tumor and the gland. Careful preoperative evaluation of pulmonary function in patients with myasthenia gravis is mandatory. Plasmapheresis is used for patients with generalized myasthenia and respiratory weakness. Plasmapheresis is performed 1 week prior to surgery on alternating days for a total of three treatments. The patient should be weaned from steroids if possible and anticholinergic medications are discontinued on the evening before operation. Preoperative radiotherapy (3000 to 4000 cGy) followed by resection 2 or 3 weeks later may be indicated for extremely large tumors or those causing obstruction of the superior vena cava.

Every attempt should be made to excise the tumor and the entire gland with sufficient margin. Both pleural spaces are widely opened to assess extent of tumor, to identify both phrenic nerves, and to recognize any pleural metastasis. One phrenic nerve may be safely sacrificed if encased by tumor. Involved lung is resected in continuity, usually as a wedge with a linear stapler. If the pericardium is involved, it should be resected in continuity with the neoplasm. Involvement of the superior vena cava or its tributaries is removed by lateral resection or graft replacement.

Pathologic staging of thymoma is determined by the extent of invasion (Table 2). Although assessment by the surgeon at the time of operation is usually accurate, invasion of the capsule may not be recognized, and some stage II tumors may thus be understaged.

Stage	*Extent of invasion*
I	Macroscopic encapsulation, no microscopic invasion of the capsule
II	Macroscopic invasion into mediastinal fat or pleura or microscopic invasion
III	Macroscopic invasion into pericardium, great vessels, or lung
IVa	Pleural or pericardial dissemination
IVb	Lymphatic or hematogenous metastasis

Table 2 Pathologic staging of thymoma according to Masaoka

Survival after complete resection of stage I and II tumors is excellent and exceeds 90 per cent at 5 years. One recent study noted at 10 years a survival of 78 per cent for stage I and 74 per cent for stage II. Survival of patients with stage III disease including incompletely resected tumors is approximately 20 per cent at 10 years. Deaths from thymoma are rare in stage I. Half of all deaths within 10 years are due to thymoma, and almost half of all thymoma-related deaths occur 5 to 10 years after operation. Because of the relatively high incidence of mediastinal recurrence in patients with stage II and III disease (10 per cent), routine postoperative radiotherapy with 5000 cGy is recommended.

Mediastinal lymphoma

Tumors of the lymphatic system constitute almost half of all anterior mediastinal masses in childhood and one-quarter of mediastinal tumors in adults, with a predilection for the fourth decade of life. The mediastinum is rarely the only site of involvement. Both Hodgkin's disease and non-Hodgkin's lymphoma occur with similar frequency. Hodgkin's disease typically spreads from one lymph node region to another in a centripetal fashion. In the mediastinum, the nodular sclerosing cell type, associated with a more favorable prognosis, is more frequently encountered. Diffuse large cell or lymphoblastic lymphomas show a characteristic centrifugal spread and skip lymph node stations. For the diagnosis of both lymphatic tumors, lymph node biopsy is required rather than a cytologic aspirate to determine cell type, and mediastinoscopy or a Chamberlain procedure are performed when peripheral lymph nodes are not involved.

While treatment of Hodgkin's disease with a small mediastinal mass may consist of radiotherapy alone, bulky Hodgkin's disease is usually treated with a combination of chemotherapy and irradiation. There is anecdotal evidence that resection of localized Hodgkin's disease confined to the thorax is beneficial. Prognosis, even in advanced stages, is favorable. Non-Hodgkin's lymphomas are currently treated in most institutions with combination chemotherapy. With high-dose chemotherapy, 55 to 85 per cent of patients achieve complete remission. Half of these patients will be disease free after 2 years.

Germ cell tumors

Germ cell tumors account for approximately 10 per cent of all mediastinal masses. A recent large series noted an equal proportion of benign and malignant teratomas, including seminomas and embryonal germ cell tumors. Germ cell tumors may occur at any age between childhood and the fifth decade, but malignant germ cell tumors predominate in men between the ages of 20 and 35. Malignant tumors give rise to symptoms including chest pain, cough, dyspnea, fever, weight loss, or hemoptysis. The evaluation of young men with anterior mediastinal masses should include determination of a-fetoprotein and b-hCG and exclusion of a testicular or abdominal mass.

Once the diagnosis is confirmed by biopsy examination, treatment with cisplatin-based chemotherapy is instituted for malignant non-seminomatous tumors. Some centers advocate treatment of anterior mediastinal tumors on the basis of positive serum tumor markers alone when biopsy is inconclusive to avoid the delay imposed by thoracotomy in these rapidly growing tumors. Resection is reserved for patients with residual mediastinal tumor after completion of chemotherapy. A 5-year survival rate of 61 per cent for patients responding to chemotherapy and less than 10 per cent for non-responders has been reported.

Patients with serum marker-negative mediastinal seminomas should undergo radiation after careful evaluation for metastatic disease. Small tumors can be resected, followed by postoperative radiotherapy. In the presence of metastatic disease, treatment with cisplatin-based chemotherapy has resulted in complete response and long-term survival in some patients. Approximately 60 per cent of patients with local mediastinal disease are cured. Relapse at distant sites occurs in the remaining patients.

In benign germ cell tumors, resection is usually curative.

Neurogenic tumors

Although neurogenic tumors may arise from any neural structure in the mediastinum, they are most commonly found in the paravertebral gutters of the posterior compartment. In this region, between 50 and 80 per cent of all tumors are of neurogenic origin. They derive mainly from two cellular components of neural tissue: autonomic ganglion and nerve sheath. Other tumors, including paragangliomas, pheochromocytomas, and peripheral neuroectodermal tumors, occur with much less frequency.

Autonomic ganglion tumors

They appear almost exclusively in children and adolescents. The three variants of this cell type display a decreasing degree of maturity. Ganglioneuroma is a benign lesion, while ganglioneuroblastoma and neuroblastoma are malignant. Neuroblastoma, the most immature type, occurs most frequently in children under the age of 3 years. Ganglioneuroblastoma is of intermediate malignancy and is typically seen in the older child. Both tumors share similar radiographic features of bone destruction and calcification. Most patients are symptomatic. Horner's syndrome, chest pain, cough, dyspnea, dysphagia, and paraplegia due to cord compression may be noted. If tumors elaborate catecholamines, diarrhea, sweating, and flushing can occur. In contrast, benign ganglioneuromas are usually asymptomatic and rarely produce catecholamines. They characteristically arise in older children or adolescents.

The goal of surgical therapy is complete resection. In the benign ganglioneuromas, this goal is easily accomplished, and the prognosis after resection is excellent. In the majority of neuroblastomas and some ganglioneuroblastomas, complete resection is often not possible. Postoperative radiotherapy to the tumor bed and adjuvant chemotherapy are added when resection is incomplete. Most neuroblastomas are advanced at the time of diagnosis and overall long-term survival is less than 30 per cent. The prognosis of thoracic neuroblastomas is better in younger patients under the age of 2 years, where it approaches 90 per cent. Survival rates in older children are less favorable.

Posterior mediastinal neurogenic tumors that extend into and through the intervertebral neural foramen have been termed ‘dumb-bell' tumors (Fig. 2). They should be resected in a single stage by a combined thoracic and neurosurgical approach. A staged operation that leaves part of the tumor in the spinal canal may result in uncontrolled bleeding and paraplegia. The operative approach for resection is depicted in Fig. 3. After careful mobilization of the intrathoracic mass, a laminectomy is performed and the tumor is resected. If the dura is opened during the course of the operation, it is closed and covered with a pleural flap. The resection is done with direct visualization of the spinal cord to minimize the risk of neurologic injury.

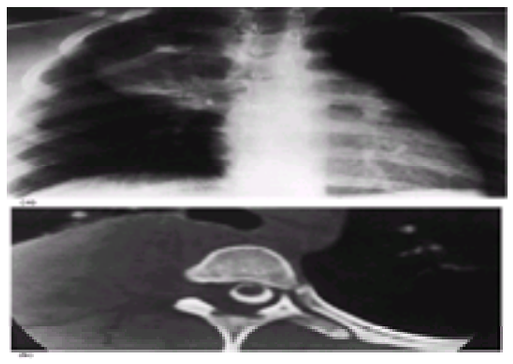


Fig. 2. (a) Right posterior mediastinal mass with splaying of the third and fourth rib. (b) Detail of CT demonstrating mass in the right paravertebral gutter extending into the spinal canal. The neural foramen is widened.

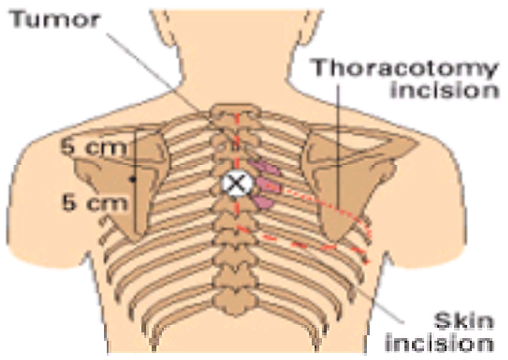


Fig. 3. Operative approach to dumb-bell-shaped tumors. The vertical part of the skin incision extends for 5 cm above and below the level of the involved foramen.

Nerve sheath tumors

Less than 10 per cent of nerve sheath tumors are malignant. They occur predominantly in young adults; mainly women. Of the two types of tumor, neurilemoma and neurofibroma, the former is the more common. Neurofibromas (schwannomas) are encapsulated and may present themselves with multiple lesions, particularly when associated with von Recklinghausen's disease. Involvement of the intervertebral foramen is found more frequently than in autonomic ganglion tumors. Symptoms occur infrequently and may indicate a malignant variant. Treatment consists of enucleation of benign tumors and resection of non-encapsulated or malignant tumors with appropriate margin. Prognosis of the benign lesion is excellent, while few patients with malignant tumors survive beyond 1 year.

Mediastinal cysts

About one-quarter of all mediastinal masses are cysts. There are two types of foregut cysts: bronchogenic cysts, which arise from the tracheobronchial tree, and enteric duplication cysts, which are derived from the esophagus. These two constitute the majority of intrathoracic cysts. Mesothelial cysts are of pericardial or pleural origin. The rare neuroenteric and gastroenteric cysts result from an embryologic defect in development. Both are associated with abnormalities of the spinal column, and gastroenteric cysts may communicate with the infradiaphragmatic gastrointestinal tract.

Mediastinal cysts come to surgical attention either as an incidental finding on radiologic studies of the chest or when they cause symptoms. In children, symptoms are typically due to mass effect in the small thorax, which can result in dyspnea, stridor, respiratory distress, or dysphagia. Cysts can cause complications at any age because of infection, hemorrhage into the cyst, erosion into adjacent structures, or, in rare instances, malignant degeneration. Surgical resection is therefore indicated even in asymptomatic patients.

Thoracoscopy

Minimally invasive and videoendoscopic techniques have made an impact on the diagnosis and treatment of mediastinal masses, although a critical evaluation of their role is not completed. Presently, most surgeons would agree that thoracoscopy is useful for biopsy of a mass and for resection of cysts and clearly benign tumors. The treatment of thymoma with thoracoscopic techniques is controversial, since exact and complete resection has not been conclusively demonstrated and the long-term outcome following this type of resection is not known. In principle, the use of less invasive surgical techniques should not compromise the aim of treatment.

Further reading

Davis RD Jr, Oldham HN Jr, Sabiston DC Jr. Primary cysts and neoplasms of the mediastinum: recent changes in clinical presentation, methods of diagnosis, management, and results. *Annals of Thoracic Surgery* 1987; **44**: 229–37.[A comparative analysis of 400 consecutive patients with mediastinal tumors.]

Demmy TL *et al.* Multicenter VATS experience with mediastinal tumors. *Annals of Thoracic Surgery* 1998; **66**: 187–92. [A multicenter study of video-assisted thoracoscopy with short-term follow-up.]

Gordon LI, Kies MS. Diagnosis and treatment of mediastinal lymphomas. In: Shields TW, ed. *Mediastinal surgery*, pp. 198–204. Lea & Febiger, Philadelphia, 1991. [A review of lymphoproliferative disorders.]

Grillo HC, Ojemann RG, Scannell JG, Zervas NT. Combined approach to ‘dumbbell' intrathoracic and intraspinal neurogenic tumours. *Annals of Thoracic Surgery* 1983; **36**: 402–7. [The original description of the combined intraspinal and thoracic approach to ‘dumbbell' tumors.]

Kaiser LR. Thoracoscopic resection of mediastinal tumors and the thymus. *Chest Surgery Clinics of North America* 1996; **6**: 41–52. [A summary of the current role of thoracoscopy.]

Shields TW, Reynolds M. Neurogenic tumours of the thorax. *Surgical Clinics of North America* 1988; **68**: 645–68. [A concise treatise of posterior mediastinal tumors of neurogenic origin.]

Wilkins EW Jr. Mediastinal tumour. In: Grillo HC *et al.*, eds, *Current therapy in cardiothoracic surgery*, pp. 122–5. BC Decker, Philadelphia, 1989. [Mediastinal tumors in a nutshell.]

Wilkins EW Jr, Grillo HC, Scannell JG, Moncure AC, Mathisen DJ. Role of staging in prognosis and management of thymoma. *Annals of Thoracic Surgery* 1991; **51**: 888–92. [An authoritative summary of the surgical treatment of thymoma emphasizing the importance of staging and long-term follow-up.]

Wright CD *et al.* Primary mediastinal nonseminomatous germ cell tumour. Results of a multimodality approach. *Journal of Thoracic and Cardiovascular Surgery* 1990; **99**: 210–17. [An important series detailing the adjuvant role of surgery in the treatment of germ cell tumors.]

41.6 The diaphragm

Daniel P. Doody and Ashby C. Moncure

Embryology
Congenital diaphragmatic hernias
Clinical presentation
Treatment
Future directions
Late-presenting posterolateral diaphragmatic hernias
Other congenital diaphragmatic defects
Eventration
Hiatal hernia
Clinical presentation
Treatment
Traumatic injuries
Diaphragmatic pacing
Further reading

Embryology

The embryonic diaphragm has six major components: septum transversum, pleuroperitoneal membrane, thoracic mesoderm, and retroglandular, aortic, and mesogastric mesoderm. In the fourth postconceptual week, the septum transversum appears as a derivative of the pericardial sac and begins to divide the peritoneal and pleuropericardial cavities. The large pleuroperitoneal canals eventually form a membrane that separates the thoracic and celomic cavities by the eighth week.

In the third month, retroglandular (perinephric) and aortic mesenchyme moves in a dorsal to ventral direction to fuse with the septum transversum, eventually with additional contributions coming from mesogastric mesoderm. Anterolaterally, costal myoblasts separate as an inner muscular layer from the thoracic wall (perhaps explaining the dual musculature of the thorax as opposed to the treble layers of the abdomen) to move centrally and join with the transverse septum. It is uncertain whether the lumbocostal trigone is the last remnant of this pleuroperitoneal membrane or is a junction of the anterolateral and dorsal musculature; current thought favors the latter explanation. By the time of parturition, the neonatal diaphragm has four major regions defined by the central tendon, the dorsal musculature through which the various foramina pass, the posterolateral lumbocostal triangle, and the anterolateral muscle.

The cervical innervation of the diaphragm by the phrenic nerve is related to the muscular contribution to the diaphragm derived from the infrahyoid mesoderm as myoblasts migrating to the transverse septum. Embryologically, a muscular component of the septum is not identified. A more likely explanation may be found in the gradual relocation of the dorsal alignment of the transverse septum from a cervical to a lumbar orientation. During this period of accelerated craniothoracic growth, the diaphragm appears to ‘descend'. During these fetal stages and by the eighth embryonic week, the phrenic innervation is complete.

Although a deficiency in any of the mesodermal contributions may create a hernia, the most frequent congenital defect is through the posterolateral aspect of the diaphragm (the hernia of Bochdalek) (Fig. 1). The closure of the pleuroperitoneal canal at the eighth week only shortly precedes the rapid return of the intestine from the yolk sac to the abdomen at the 10th embryologic week. Moreover, during these fetal stages the muscular formation of the diaphragm is proceeding. An imbalance between the speed of closure of the diaphragm (as a membrane and as a muscle) and the increasing pressure in the peritoneal cavity may provide an opportunity for the intestine to intrude into the newly forming hemithorax. The left hemidiaphragm closes at a slightly later fetal stage than does the right diaphragm. A combination of earlier closure and the protection to the right hemidiaphragm afforded by the hepatic anlage may explain the overall predominance of left-sided posterolateral diaphragmatic defects as a congenital problem.

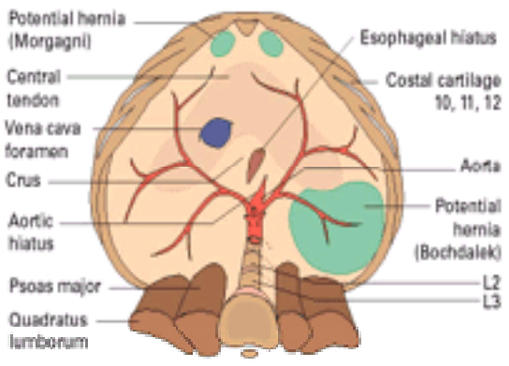


Fig. 1. Underside of the diaphragm demonstrates the anteriorly located space of Larrey where the potential hernia of Morgagni can occur. The posterolateral defect at the lumbocostal triangle is where the more common and life-threatening Bochdalek hernia occurs. The central tendon is the most common site of traumatic diaphragmatic rupture in children and adults.

Congenital diaphragmatic hernias

One of the most challenging experiences for the practising physician is the infant with a congenital diaphragmatic hernia who demonstrates immediate respiratory distress. Despite advances in neonatal intensive care and respiratory support, infants who present with respiratory distress in the first 6 h have a predicted 40 to 60 per cent mortality, with pulmonary hypoplasia and a reactive pulmonary arterial vasculature accounting for this significant risk. The hernia of Bochdalek, or posterolateral diaphragmatic hernia, has an incidence of 1 in 4400 live births and a predicted incidence of approximately 1 in 2200 conceptions. Many of these fetuses have associated congenital and chromosomal anomalies that are incompatible with life and result in early fetal loss or stillbirth. Many argue that the severity and frequency of these associated congenital anomalies should be weighed in evaluating new therapies designed to address the problem of congenital diaphragmatic defects, in particular those demanding *in utero* intervention.

A left-sided diaphragmatic defect occurs in 80 to 90 per cent of newborns with respiratory distress. Right-sided defects are found in 10 to 15 per cent of cases; bilateral hernias are rare. The male:female ratio is 1:1.36. In diaphragmatic defects that manifest themselves late the right:left ratio tends towards an equal distribution.

Clinical presentation

There must be a high index of suspicion for diaphragmatic hernia when an infant has respiratory distress (Fig. 2). While other causes need to be eliminated, the classic clinical presentation of a scaphoid abdomen and barrel chest with a right-sided cardiac impulse suggests the presence of an underlying diaphragmatic defect. Many children with diaphragmatic hernias are term infants with no other obvious anomaly.

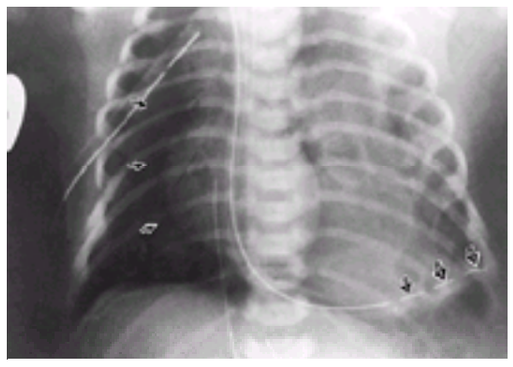


Fig. 2. Chest radiograph of neonate with severe respiratory distress reveals a left-sided diaphragmatic hernia with contralateral pneumothorax and collapse of the right lung (closed arrows). There is marked gaseous distension of the intestinal contents in the left hemithorax and the nasogastric tube is visualized in the left chest (open arrows).

Once the diaphragmatic hernia has been recognized, intubation is mandatory. Continued mask ventilation worsens matters by causing distension of the intrathoracic abdominal viscera, which aggravates hypoxia and hypercarbia. The underlying pulmonary hypoplasia prevents normal chest excursion. If there is a possibility of a diaphragmatic defect, low-pressure ventilation is recommended at peak inflation pressures of 30 cmH_2O and respiratory rates of 60 to 120 breaths/min. Once the child has been intubated, a nasogastric tube should be passed to prevent further distension of the stomach and distal bowel, which may compromise pulmonary compression.

Treatment

Intravenous fluid administration is started. Acidosis needs to be corrected aggressively with a sodium bicarbonate drip (0.5–2 mEq/kg per hour) or tromethamine. An infant in respiratory distress should be paralysed with pancuronium and sedated with fentanyl (which may have some beneficial pulmonary vasodilatory effects).

Immediate operative reduction does not improve the mortality rate. Evisceration of the infant in the delivery room has been universally fatal. A period of stabilization appears to be of benefit and is currently accepted as the preferred treatment of infants with diaphragmatic hernias. Predictors of survival have been proposed, based on radiographic criteria, level of preoperative acidosis, and on the arterial CO_2 and the ventilatory index (respiratory rate × mean airway pressure). Despite these predictors, the newer treatments, including extracorporeal membrane oxygenation, have a produced survival in infants with ‘nonsurvivor' indices.

After stabilization, which may require several days, the infant is taken to the operating room, where the viscera are removed from the thoracic cavity through a subcostal incision. The majority of neonates do not have a sac if a large diaphragmatic defect is found. A posterior rim of muscle is typically found: this needs to be freed from the retroperitoneum, the adrenal gland, and the posterolateral thoracoabdominal wall, and approximated to the anterior diaphragmatic leaf. Although not usually possible, a two-layer closure is performed using a nonabsorbable suture. Occasionally, the defect is so large that synthetic material must be used: any of the prosthetic materials that are currently available are suitable, although thin Gore-Tex is our preference. Muscular flaps have also been used to close a large defect, but if extracorporeal membrane oxygenation is being considered, extensive muscular dissection should be minimized and a patch should be used.

Nonrotation of the gastrointestinal tract is a frequent finding in the neonate with a left-sided diaphragmatic defect. If the child's condition is sufficiently stable, division of duodenocolic bands (Ladd's procedure) is performed to open the small-bowel mesentery. An incidental appendectomy is performed.

Postoperative care is directed to minimizing the onset of pulmonary hypertension of the newborn. The pulmonary vasculature of a newborn with a posterolateral diaphragmatic defect has a diminished cross-sectional area and extension of medial hypertrophy out to terminal arterioles. The best means to prevent the return of a persistent ‘fetal' circulation appears to be a combination of mild alkalosis with normal oxygenation. The use of alkalinizing agents, including sodium bicarbonate or tromethamine, is of some benefit. If a respiratory alkalosis can be induced, there may be a benefit to reaching that CO_2 tension at which pulmonary arterial hypertension reverts. The combination of acidosis and hypoxia can cause severe, unrelenting pulmonary hypertension and right-to-left shunting through the patent foramen ovale and ductus arteriosus.

Ventilator management is crucial. Changes usually tolerated by a normal infant or adult can result in a reversal of the pulmonary vasodilatation and a return to the ‘fetal' vasoconstricted state. The improved oxygenation with this normally perfused pulmonary bed before the return of a ‘fetal' circulation, with increased right-to-left shunting, has been termed the ‘honeymoon' period. Initial ventilator changes gradually lower the FIO_2 and then decrease the mean airway pressure to a normal range. High peak inflation pressure increases barotrauma and eventually results in the formation of interstitial emphysema and late bronchopulmonary dysplasia.

High-frequency jet ventilation and oscillatory ventilation have been used in the medical management of neonates with persistent pulmonary hypertension; these depend on an outward diffusion of CO_2 in small increments while oxygenation is maintained at much lower mean airway pressures. Although early experience has failed to demonstrate significant success, there is continued investigation of this therapy because of the apparent diminution of barotrauma.

A benefit in reversing pulmonary hypertension has been shown with tolazoline, but this pharmacologic support is effective in only 30 per cent of cases. With tolazoline therapy, there is appreciable gastrointestinal morbidity, including massive bleeding and infarction. Other agents, including nitroglycerine, nitroprusside, and isuprel, have not had much effect on pulmonary hypertension.

Like tolazoline, inhaled nitric oxide can relax the pulmonary vasculature. Although frequently efficacious in the treatment of pulmonary hypertension of the newborn, it has not been as effective in treating the severe pulmonary hypertension that complicates diaphragmatic hernia. While this therapy improves and stabilizes the severe hypoxemia in newborns with diaphragmatic hernias, the need for extracorporeal support and the mortality in the treatment of this problem are equivalent in those infants treated with inhaled nitric oxide and those treated conventionally.

If medical therapy fails, extracorporeal oxygenation has been used in infants, with improved survival over the previous 100 per cent mortality. The criteria for use of this therapy vary from hospital to hospital, but commonly include a prolonged AaO_2 gradient greater than 78 kPa for over 12 h, an oxygenation index (**OI**) greater than 40 [OI = (FIO_2 × MAP/postductal PaO_2 (in Torr)) × 100] or evidence of severe barotrauma. An infant whose condition meets these criteria has a predicted mortality of from 80 to 90 per cent. Extracorporeal membrane oxygenation, which consists of bypass through a membrane oxygenator, can reverse pulmonary hypertension. In the neonate, extracorporeal support is commonly provided by placing a venous drainage catheter into the right atrium and an arterial catheter in the right common carotid artery. Once the catheters are in place, the blood is passed through a membrane oxygenator, through a heat exchanger, and is then returned to the infant's arterial side. Venovenous support can also be used, although the mediastinal shift may cause mechanical complications, with ‘kinking' and occlusion of the cannula. The infant needs to be anticoagulated for this therapy, and there is a risk of bleeding. Intracranial hemorrhage occurs in 10 to 15 per cent of neonates: extracorporeal bypass needs to be urgently terminated, as extension of the intracranial bleeding can be fatal.

Fifty-nine per cent of neonates with diaphragmatic hernia and unrelenting persistent pulmonary hypertension survive after treatment with extracorporeal membrane oxygenation. There is less of an impact on the incidence of bronchopulmonary dysplasia.

Future directions

The use of *in utero* surgery to correct diaphragmatic defects is a new avenue for addressing the problem of pulmonary hypoplasia and vascular reactivity that results in the significant morbidity and mortality associated with congenital diaphragmatic hernias. If a subgroup of fetuses with a high risk for a fatal outcome can be identified, *in utero* intervention might be indicated. The presence of an intrathoracic stomach and evidence of maternal polyhydramnios are antenatal indicators. Diaphragmatic hernia identified by ultrasonography before 24 weeks' gestation is frequently associated with a fatal outcome.

Although successful *in utero* intervention with reduction of incarcerated viscera and repair of left-sided diaphragmatic hernias has been reported, the outcome and mortality at this time are insignificantly different from those of the group treated by conventional therapy and extracorporeal membrane oxygenation. Tracheal occlusion to promote pulmonary hyperplasia has been used clinically in selected cases, but the outcome for the fetuses so treated has not been as good as in infants treated conventionally. While these methods remain experimental, with refinement of techniques and improvement in tocolytic therapy, they may become major treatments for

the extreme morbidity and mortality associated with pulmonary hypoplasia and hypertension of the newborn.

Late-presenting posterolateral diaphragmatic hernias

Infants and children with diaphragmatic hernias who do not have respiratory distress in the first 6 to 12 h of life have a more favorable prognosis. Late-presenting posterolateral diaphragmatic hernias do not have the complicating feature of pulmonary hypoplasia, and their perioperative mortality in this group is low, and should approach zero. The differential diagnosis includes pneumatoceles secondary to staphylococcal pneumonia (Fig. 3) and the macrocystic variant of cystic adenomatoid malformation. Contrast studies can help delineate the presence of intestine in the thoracic cavity.

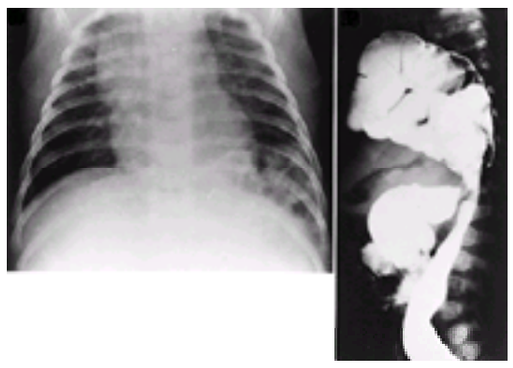


Fig. 3. Six-month-old child with apparent staphylococcal pneumonia and pneumatoceles of the left lower lobe. Contrast enema demonstrates presence of colon through a left posterolateral diaphragmatic defect. The child underwent uneventful repair with an uncomplicated postoperative course as pulmonary hypoplasia did not complicate this late-presenting diaphragmatic defect.

Surgical repair can be accomplished through the abdomen in almost all instances. Even right-sided diaphragmatic defects are generally easily approached and repaired through the abdomen, although a thoracic extension is occasionally necessary. A thoracic approach can be used successfully, but reduction of the intestine may not be easily accomplished.

Of special note is the presentation of right-sided diaphragmatic hernias with group B streptococcal infection. While streptococcal infections are common in the neonatal period (estimated at 1 in 300 births), the presence of a right-sided ‘pneumonia' that fails to respond to appropriate antibiotic coverage should alert the physician to the presence of a right-sided diaphragmatic defect.

Other congenital diaphragmatic defects

Isolated anterior diaphragmatic defects through the retrosternal space of Larrey create the Morgagni hernia. The defect is most frequently noted as an unexpected radiographic finding. Repair is easily accomplished and should be performed to avoid the potential complications of incarceration or gastric volvulus. The anterior diaphragmatic defect associated with the pentalogy of Cantrell is more complex. This defect is a remnant of the pericardioperitoneal canal and is associated with epigastric omphalocele, crescentic anterior diaphragmatic defect, pericardial defect, ectopia cordis, distal sternal cleft, and ventriculoseptal defect.

Paraesophageal hernia in the neonate is secondary to the failure of dorsal musculature to form from the posterior mesoderm. These defects are rare; they generally present because the children eat poorly and do not grow. Pulmonary hypoplasia does not complicate either the Morgagni or the congenital paraesophageal hernia.

Eventration

Diaphragmatic eventration can be congenital or acquired. The congenital form presents with moderate respiratory distress. Although attempts at medical management are occasionally successful, additional ventilatory support is often indicated. If the child is dependent on a respirator for more than 5 days, the eventration is best addressed by diaphragmatic plication. Plication usually increases the functional residual capacity and allows rapid weaning from the ventilator.

The acquired forms of eventration are secondary to damage or entrapment of the phrenic nerve. If respiratory distress does not intervene, observation alone is indicated. If symptoms are present, plication through a low thoracic approach can be beneficial. In the special instance of an intraoperative injury to the phrenic nerve, immediate plication of the diaphragm, particularly in infants, seems to avoid postoperative respiratory problems. In over 80 per cent of patients treated with immediate plication, diaphragmatic motion appears to return to normal when studied fluoroscopically.

Hiatal hernia

The most common type of diaphragmatic hernia is the hiatal hernia. Three types of hiatal hernias have been described. The classic sliding hiatal hernia (type I) is the passage of the gastroesophageal junction and variable amounts of the cardia above the diaphragm. In type II hiatal hernia, commonly referred to as paraesophageal or rolling hernia, the gastroesophageal junction remains below the diaphragm but the fundus rolls up, generally in an anterior location, next to the esophagus. The type III hiatal hernia is a combination of a sliding and paraesophageal hernia.

Clinical presentation

Hiatal hernias are frequently asymptomatic. Increased intra-abdominal pressure during routine upper gastrointestinal contrast studies frequently discloses a small, unsuspected sliding hernia. Nonetheless, 20 per cent of hiatal hernias are symptomatic, primarily because of problems related to esophageal reflux.

In contrast, paraesophageal hernias can be associated with upper abdominal discomfort. Occasionally, fatigue and malaise are symptoms of chronic blood loss, which may be secondary to engorgement of the entrapped gastric mucosa. Reports of incarceration followed by strangulation have appeared.

Treatment

Asymptomatic sliding hiatal hernias do not require surgical exploration. Patients with symptomatic type I hiatal hernias and esophageal reflux should initially be managed medically with antacids or H_2-receptor antagonists to diminish the acidity of the gastric efflux, and with metoclopramide or cisapride, which promote gastric emptying and tighten the lower esophageal sphincter. Simethicone tablets may relieve symptoms. If medical management is unsuccessful or if complications such as esophageal ulceration, stricture, stenosis, or Barrett's esophagus intervene, surgical repair is indicated. Repair can be accomplished satisfactorily by a transthoracic or transabdominal approach. Nissen fundoplication, Belsey–Mark IV fundoplication, and the Hill gastropexy all produce excellent results when properly performed.

Large paraesophageal hiatal hernias in otherwise healthy individuals should be repaired, even when asymptomatic. Incarceration, strangulation, or hemorrhage of such a hernia is an indicator for urgent repair.

Traumatic injuries

Traumatic diaphragmatic injuries may result from either penetrating or blunt trauma, usually of the lower chest or upper abdomen. The location of the injury in the diaphragm is variable and is associated with injuries to other anatomic structures, particularly the ribs, spleen, long bones, the cranium and its contents, and the liver (Fig. 4).



Fig. 4. Traumatic diaphragmatic rupture with presence of liver in right hemithorax of an adult.

Diaphragmatic injury is frequently difficult to diagnose and is often an unexpected finding at exploration in a patient who has suffered penetrating trauma. A high index of suspicion for the presence of diaphragmatic injury is necessary. Symptoms directly related to this injury are due to the effects of encroachment and entrapment of herniated viscera within the affected hemithorax; dyspnea and pain related to the injured thorax or abdominal wall are common. Physical findings relate to thoracic or abdominal injury and to those specific to the diaphragmatic injury (dullness and diminished breath sounds over the affected hemithorax). The initial chest film is often normal or shows no specific findings. Radiographic nonspecificity is more frequent after blunt injury: apparent elevation or irregularity of the diaphragm or extraneous shadows above it suggest the diagnosis. In a stable patient, a contrast study of the stomach may establish the diagnosis, as may ultrasonographic examination or a computed tomographic scan with visceral contrast. Since diaphragmatic rents do not heal spontaneously, all injuries should be surgically repaired. The operative approach is largely determined by the condition of the patient and the nature of the suspected injury. Hemodynamically stable patients may be treated as semi-urgent, with the abdominal approach being used because of the frequency of associated occult intra-abdominal injuries. If suspected injuries to thoracic organs dictate thoracotomy, diaphragmatic repair can be accomplished through either it or via a laparotomy. Most diaphragmatic injuries can be closed with simple, nonabsorbable sutures.

The chronic presentation of diaphragmatic rupture represents the effect of entrapment of stomach, colon, or small bowel within the thorax, or the consequence of long-standing pulmonary collapse from encroachment on the affected lung. Early satiety, postprandial distress, productive sputum, or dyspnea are the presenting symptoms. A chest radiograph with subsequent contrast examination of the gastrointestinal tract will usually secure the diagnosis. A thoracotomy is usually appropriate to allow separation of adhesions between the trapped viscera and lung, with subsequent repair of the diaphragmatic injury through this exposure.

Diaphragmatic pacing

Diaphragmatic pacing in adults is usually performed for quadriplegia and secondary paresis of diaphragmatic muscles. In infants, congenital central alveolar hypoventilation syndrome (Ondine's curse) and Arnold–Chiari II malformation respond well to the use of the pacemaker. An electrode is placed about the phrenic nerve in the chest of an infant and in the neck of an adult: it can partially free over 90 per cent of patients from ventilator dependence.

Further reading

Ashcraft KW *et al.* Diagnosis and treatment of right Bochdalek hernia associated with group B streptococcal pneumonia and sepsis in the neonate. *Journal of Pediatric Surgery* 1983; **18**: 480–5. [Early report noting the unusual association of group B streptococcal pneumonia and right-sided congenital diaphragmatic hernia.]

Bohn D, Tamura M, Perrin D, Barker G, Rabinovitch M. Ventilatory predictors of pulmonary hypoplasia in congenital diaphragmatic hernia, confirmed by morphologic assessment. *Journal of Pediatrics* 1987; **111**: 423–31. [An institutional review to establish factors predicting survival in infants with immediate respiratory distress complicating congenital diaphragmatic defects.]

Gray SW, Skandalakis JE. *Embryology for surgeons: the embryological basis for the treatment of congenital defects*. Saunders, Philadelphia, 1972. [Excellent review of the embryonic and fetal formation of the diaphragm.]

Harrison MR *et al.* Successful repair *in utero* of a fetal diaphragmatic hernia after removal of herniated viscera from the left thorax. *New England Journal of Medicine* 1990; **322**: 1582–4. [Early report of successful *in utero* correction of diaphragmatic hernia with a poor outcome predicted by antenatal ultrasound.]

Harrison MR *et al.* Correction of congenital diaphragmatic hernia in utero VII: a prospective trial. *Journal of Pediatric Surgery* 1997; **32**: 1637–42. [Comparative trial of *in utero* intervention and conventional therapy in the management of left-sided congenital diaphragmatic hernias, which fails to demonstrate benefit of *in utero* surgery.]

Harrison MR *et al.* Correction of congenital diaphragmatic hernia in utero VIII: response of the hypoplastic lung to tracheal occlusion. *Journal of Pediatric Surgery* 1996; **31**: 1339–48. [Early clinical application of tracheal occlusion *in utero* to try to reverse the severe pulmonary hypoplasia associated with congenital diaphragmatic hernia.]

Heiss K *et al.* Reversal of mortality for congenital diaphragmatic hernia with ECMO. *Annals of Surgery* 1989; **209**: 225–30. [Institutional review demonstrating significant survival benefit using extracorporeal membrane oxygenation to treat severe respiratory distress complicating congenital diaphragmatic hernia.]

Reickert CA *et al.* Congenital diaphragmatic hernia survival and the use of extracorporeal life support at selected level III nurseries with multimodality support. *Surgery* 1998; **123**: 305–10. [Multicenter survey to determine the use of extracorporeal membrane oxygenation and the current survival of infants with congenital diaphragmatic hernias.]

Symbas PN, Vlasis SE, Hatcher C. Blunt and penetrating diaphragmatic injuries with or without herniation of organs into the chest. *Annals of Thoracic Surgery* 1986; **42**: 158–62. [A review of the diagnosis and management of traumatic diaphragmatic injury treated at the Grady Memorial Hospital, Atlanta GA.]

41.7 Lung volume reduction surgery and pulmonary transplantation

Joseph I. Miller, Jr.

Lung volume reduction surgery
Surgical technique
Pulmonary transplantation
History
Donor harvest and technique
Single lung transplant
Choice of side
Bilateral sequential lung transplantation
Complications
Mortality
Further reading

Chronic obstructive pulmonary disease affects approximately 15 million people in the United States. It is the fourth leading cause of death and approximately 1.9 million individuals in the United States have been diagnosed with pulmonary emphysema. Whereas the usual patient with pulmonary emphysema can be managed medically with bronchodilators, antibiotics, corticosteroids, oxygen, and pulmonary rehabilitation, there are certain select individuals who may benefit from surgical procedures. There are currently three surgical procedures that may benefit the patient with pulmonary emphysema. These are pulmonary bullectomy of which few patients meet the defined criteria; lung volume reduction surgery; and pulmonary transplantation. This chapter will focus on the two that are most widely applied in the United States: lung volume reduction surgery and pulmonary transplantation.

Lung volume reduction surgery

Lung volume reduction surgery was originally proposed by Otto Brantigan in 1955. He suggested that by resecting targeted zones in lungs with severe emphysema that the elastic recoil of the remaining lung would be improved and, therefore, better ventilation achieved. Whereas his initial results were encouraging, the operative mortality was significant and an enthusiasm for this operation waned.

Brantigan originally proposed that by reducing the lung volume of non-perfused areas of lung, the elastic recoil of the remaining lung would be improved. He proposed bilateral stage procedures which resulted in a 16 per cent operative mortality. There were few pulmonary function criteria by which to judge objective success and the operation was abandoned in 1960. It was in 1993 that Cooper reapplied Brantigan's concept utilizing bovine pericardium loaded on thoracic staplers. The initial report was on 20 patients and their initial data presented at the American Association of Thoracic Surgery in March 1994. He demonstrated that with improved anesthetic techniques, pain control, pulmonary rehabilitation, and improvement in instrumentation, lung volume reduction surgery could be performed with low mortality and good physiologic results.

Since that time institutions in the United States, Canada, and Europe have begun utilizing this operation with varying degrees of success.

Lung volume reduction has the potential advantages of offering a select group of the population with pulmonary emphysema, improved functional results for a significant period of time. This means that the number of patients that would require immunosuppression, if they had undergone transplant, would be significantly reduced. In addition, there is much less waiting time than transplantation and the procedure may also be used as a possible bridge to transplantation.

The physiologic mechanisms of improvement in lung volume reduction surgery are due to: (a) increased elastic recoil of the lung following resection of heterogeneous lung; (b) improvement in ventilation perfusion mismatch; (c) restoration of chest wall mechanics and diaphragmatic position; and (d) improvement in cardiovascular hemodynamics.

At the present time lung volume reduction surgery offers significant palliation with benefits extending for at least 2 years. Improvement has been shown by quantitative dyspnea on exertion and quality of life indicators; measurement of pulmonary function; and objective physiologic parameters such as airways obstruction, elastic recoil, exercise capacity, and tolerance.

The patient selection criteria have been improved based on the experience of several reporting groups. Increased arterial PCO_2 and advanced age have been confirmed as indicators of an adverse outcome.

From the experience of a number of groups certain criteria and parameters have been established in regards to selection criteria, pulmonary rehabilitation, operative technique, postoperative care, and outcomes analysis. Our selection criteria for lung volume reduction surgery are as listed in Table 1. The patient must have advanced generalized emphysema and failed maximum medical therapy. Ideally, the patient should be less than 75 years of age, and have an forced expiratory volume in 1 s (FEV_1) less than 30 per cent of predicted and a PCO_2 less than 45 mmHg. There should be no significant history of coronary artery disease, no major life-threatening illnesses, and the patient should have abstinence from tobacco for a minimum of 6 months. In addition, he/she should be able to perform pulmonary rehabilitation and be on less than 10 mg of prednisone per day and have no significant spinal osteoporosis. Exclusion criteria for lung volume reduction surgery are predominant bullous emphysema; the patient may be physiologically too good; significant coronary artery disease; and the presence of pulmonary hypertension with a mean pulmonary arterial pressure greater than 35 mmHg or a systolic pressure greater than 45. If the patient cannot perform pulmonary rehabilitation for a minimum of 30 consecutive min on the bike or treadmill, that indicates a poor adverse outcome. Prednisone dosages greater than 10 mg/day, the presence of significant bronchitis or asthma, or a history of previous pulmonary surgery or sclerosis contraindicate lung volume reduction surgery.

1. Failure of maximum medical therapy for generalized emphysema
2. FEV_1 20–30 per cent of predicted (ideal)
3. Age <70 years
4. Radiographic evidence of generalized emphysema with "target zones of heterogeneity" (predominantly upper lobe distribution)
5. Absence of significant coronary artery disease
6. Ability to participate in pulmonary rehabilitation
7. Absence of tobacco for a minimum of 6 months
8. Lack of osteoporosis
9. Prednisone dosage <15 mg/day

FEV_1, forced expiratory volume in 1 s

Table 1 Selection criteria for lung volume reduction surgery

The ideal patient would have the following characteristics. The patient would be less than 70 years of age and have an FEV_1 less than 30 per cent of predicted. They would have advanced generalized emphysema and have failed all routine medical therapy. They should have the radiographic presence of significant hyperinflation with targeted areas and a PCO_2 less than 45 mmHg. They should have no underlying pulmonary hypertension and be able to achieve pulmonary rehabilitation.

All patients undergoing lung volume reduction surgery undergo a standard workup with a complete history and physical examination; inspiratory and expiratory chest radiographs; spirometry with arterial blood gases; split perfusion lung scan with lobar counts; standard 6 min physiologic walk test; high resolution computed tomography; and a pulmonary rehabilitation consult. In addition, they undergo a dobutamine stress echocardiogram and if there is any evidence of pulmonary hypertension or left ventricular dysfunction or coronary artery disease, they undergo both right and left heart catheterization, respectively, based upon the dobutamine echocardiography.

Studies to date have shown that an age greater than 70 years, PCO_2 greater than 50 and an FEV_1 less than 0.5 l/min are predictive of an adverse outcome. In addition, the presence of significant obesity, a diffusion capacity less than 15 per cent of predicted, steroids greater than 10 mg/day; uncontrolled panic attacks; the absence of

target areas; and the inability to achieve pulmonary rehabilitation goals are all associated with adverse outcomes and patients should not be considered for surgery, generally, when these situations are present.

Pulmonary rehabilitation has played a significant part in the management of patients undergoing both lung volume reduction surgery and pulmonary transplantation. The benefits of pulmonary rehabilitation are an improvement in exercise capacity, improvement in endurance, and an improved sense of well being and quality of life. This is achieved through increased aerobic capacity, increased muscle strength, desensitization to dyspnea, and improved techniques of performance. Patients are enrolled in a 6-week to 2-month period of upper extremity arm orgometry, in addition to exercise training on the treadmill and stationary bike. The individual must be able to achieve 30 consecutive min on either the bike and/or treadmill before being declared a viable candidate for lung volume reduction surgery.

The following are the recommended pulmonary function guidelines: (a) FEV_1 20 to 30 per cent predicted; (b) total lung capacity greater than 115 per cent of predicted; (c) residual volume greater than 200 per cent of predicted; (e) diffusion capacity greater than 30 per cent of predicted; and (f) arterial blood gases PCO_2 less than 45 and PO_2 greater than 50.

Surgical technique

In the past 4 years surgical and anesthetic techniques have been well standardized. It is now accepted that the approach may be by either the open (median sternotomy) or closed (thoracoscopic) technique. By far, the most important factor is the experience and skill of the operating surgeon and team. The consensus is that stapling with bovine buttressed pericardium is the method of choice. A few groups have reported stapling without buttressing reinforcement, but these groups are few.

From the anesthetic standpoint all patients are monitored with central venous lines and arterial pressure monitoring. All patients have the placement of a double lumen endotracheal tube at the time of surgery and a working epidural catheter is a prerequisite. Whenever possible, one should utilize the same anesthetic team and anesthetic technique. It is imperative that all patients be extubated early and their pain control well standardized. From a surgical technique point, the procedure may be done through either an open or closed surgical approach. Pertinent aspects of surgical technique include double lumen anesthesia; epidural for pain control; and a quick operating time of 1.5 to 2 h for a bilateral procedure. It is felt that each lung should be done under a single collapse of the lung with 20 to 30 per cent of the targeted lung volume resected. There should be no suction placed on the chest tubes and bovine pericardium should be utilized for a continuous reinforced suture line. There should be meticulous control of air leaks so that, essentially, no air leaks exist at the time of surgery.

Our technique for lung volume reduction surgery is as shown in Fig. 1. The patient is prepped and draped in routine fashion and a midline sternotomy incision made. The mediastinal pleura is opened with care to avoid injury to the phrenic nerve. Beginning on the medial aspect of the right upper lobe, a continuous ‘U' shape inverting buttress bovine staple suture line is made in an inverting ‘U' fashion resecting approximately 20 to 30 per cent of the upper lobe. Following this, the inferior pulmonary ligament is released. The lung is re-expanded and if no apical space exists, then a pleural tent is not dropped. If there is an apical space, a pleural tent is then dropped at this time. The lung is examined under water and if any air leaks are present meticulous attempt is made to close these. If only minor leaks exist from the fissure line, experience has shown that these will generally stop within 24 to 48 h. The right lung is then re-expanded and the left pleural space is then opened in a similar manner. A similar lung volume reduction surgery procedure is then carried out on the left upper lobe in a routine manner. At the completion of the procedure both mediastinal pleura's are closed in a continuous running 2-0 absorbable suture to separate the pleural spaces and prevent cross air leaks through the mediastinum. Two pleural chest tubes are left at each pleural space and a single mediastinal tube placed and pulled at 48 h.

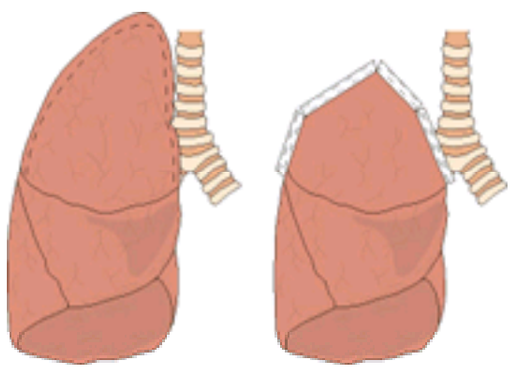


Fig. 1. (a) Inverted ‘U' with planned resection. (b) Stapled resection with 30 per cent of upper lobe removed.

Pitfalls in surgical technique that have been learned through a series of trial and error include: there should be no parallel suture lines within 2 cm of each other as these tend to result in traction tears; the lungs should be handled with gentleness in order to avoid traction tears of the lung; minor suture line leaks will generally close within 24 to 48 h. One should be particularly aware of a_1-antitrypsin patients and anytime the right lower lobe is resected a right angle dependent low chest tube should be placed. No suction is placed on the chest tubes.

A thoracoscopic approach can be performed as a bilateral sequential simultaneous procedure. It is a consensus that equal results can be achieved by either technique. The results reported by both the open and closed technique are given in Table 2.

Author	Approach	Preoperative FEV_1 (l/min)	Preoperative 6MW (ft)	LOS	FEV_1 per cent increase	6MW per cent increase	Mortality
Keenan	Endo VATS	0.82	784	17	27	34	5.5
McKenna	Endo VATS	0.78	-	13	33	-	2.5
Wakefield	Endo VATS	0.73	864	13	35	38	4.0
McKenna	Bilat. VATS	0.69	-	10.9	52	-	1.0
Cooper	Bilat. open	0.78	1175	15.1	57	59	5.0
Miller	Bilat. open	0.80	1106	-	53	29	11
Daniel	Bilat. open	0.78	-	13.4	48	-	3.8
Miller	Bilat. open	0.56	705	-	97	104	5.1
Argenziano	Bilat. open	0.59	589	17	61	82	7.0

LOS, length of stay; FEV_1, forced expiratory volume in 1 s; preop., preoperative; endo, endobronchial; bilat., bilateral; VATS, video-assisted thoracoscopy; 6MW, examine walk.

Table 2 Results of lung volume reduction surgery

Each of the series reported shows short-term improvement of physiologic function based on pulmonary function criteria and 6 min walk. Cooper *et al.* recently reported a 2-year follow up for the original 20 patients who had undergone the procedure by the open technique. The original group had a mean FEV_1 of 1.4 l at 6 months and a mean value of 1.25 l at 30 months. Residual volume had increased slightly from 3.9 l at 6 months to 4.4 l at 30 months. The 6 min walk did not decrease and indicated that long-term results may be lasting in the majority of patients; however, there appears to be a fall-off of 5 to 10 per cent in some individuals.

The most common reported postoperative complications are postoperative air leaks greater than 7 days reported in 30 to 40 per cent of patients. Other complications are pneumonia in 9 to 22 per cent; gastrointestinal disturbances in 15 per cent; and respiratory failure in 2.5 to 13 per cent. Early and late mortality rates are reported from 2.4 to 7 per cent and from 4 to 17 per cent, respectively. Morbidity and mortality do not appear to depend on which type of surgical approach is used.

Outcomes analysis may be measured in terms of physiologic function based on pulmonary function criteria and 6 min walk, as well as the subjective criteria of dyspnea on exertion and evaluation of quality of life indicators. All reports to date with follow up to 1 year report acceptable mortality and morbidity.

From the reported series a number of preoperative predictors of poor clinical outcome and/or death have been indicated. These are increasing age greater than 70 years, a PCO_2 greater than 45 mmHg, a diffusion capacity less than 30 per cent of predicted, and a 6 min walk less than 700 feet after pulmonary rehabilitation.

Serious consideration should be given before performing lung volume reduction surgery when any of the above criteria are present.

To date early analysis of outcomes from 12 to 18 months have clearly documented the efficacy of lung volume reduction surgery whether performed by the opened or closed technique. It is documented that a bilateral procedure yields better functional result than a unilateral procedure and stapling yields results superior to that of laser therapy.

Questions that still need to be addressed and what outcome parameters are most appropriate to balance risks and benefits of surgery include the following: (a) How long the benefits of surgery last and do they justify the expense of the procedure? (b) Which clinical, physiologic and radiographic parameters define patients at highest risks with greatest benefits? (c) What level of improvement warrants the expense of the procedure? (d) What is the risk/benefit ratio? and (f) Should there be centers of excellence or should this only be performed in university centers?

Pulmonary transplantation

Since the first successful human lung transplant in 1980 reported by Cooper *et al.*, remarkable progress has been made in the field of pulmonary transplantation. Pulmonary transplantation is now performed in major medical centers throughout the world. There have been significant advances made in donor and recipient selection, surgical technique, advantageous immunologic strategies, and newer antibiotic regimens. The operative mortality now ranges from 10 to 15 per cent and 2-year survival rates are reported at 80 per cent.

History

Hardy reported the first human lung transplant in 1963. The patient died after 18 days, but initial early success had been achieved. At that time there were no immunosuppressive drugs. The development of the immunosuppressive drug, cyclosporin, enabled the initial advance in pulmonary transplantation to be achieved. The first successful pulmonary transplant was carried out in 1983 by the Toronto Lung Transplant Group on a 58-year-old man with idiopathic pulmonary fibrosis. In 1988, Patterson *et al.* described an *en bloc* double lung replacement technique which enabled the application of bilateral lung replacement for patients for whom a single lung transplant was not appropriate. In 1990, Pasque *et al.* described a bilateral sequential replacement technique which has now become the standard throughout the world for bilateral pulmonary transplantation.

The indications for pulmonary transplantation are listed in Table 3. These include: (a) pulmonary vascular disease which consists primarily of primary pulmonary hypertension or pulmonary hypertension that may be associated with other cardiac diseases or as is found in Eisingmenger's syndrome; (b) chronic obstructive pulmonary disease which predominantly consists of generalized emphysema and in a few select patients with a_1-antitrypsin deficiency; (c) the presence of restrictive lung disease manifested by idiopathic pulmonary fibrosis. Suppurative diseases of the lung such as cystic fibrosis and/or bronchiectasis show an increasing need for transplantation. More recently, some patients have been re-transplanted because of the presence of chronic rejection.

1.	Pulmonary vascular disease
	• primary pulmonary hypertension
	• pulmonary hypertension secondary to systemic or cardiac disease
	• Eisenmenger syndrome
2.	Obstructive lung disease
	• generalized emphysema
	• a_1-antitrypsin deficiency
3.	Restrictive lung disease
	• idiopathic pulmonary fibrosis
	• fibrosis—secondary to other lung disease
4.	Suppurative disease of the lung
	• cystic fibrosis
	• bronchiectasis
5.	Chronic rejection

Table 3 Indications for pulmonary transplantation

The types of pulmonary transplantation that may be carried out are as follows: (a) heart/lung transplant; (b) bilateral sequential lung transplant; (c) single lung transplant; and (d) lobar transplant (Table 4). Single lung transplantation is the procedure most frequently performed for emphysema, primary pulmonary hypertension, interstitial lung disease, and Eisingmenger's syndrome when the heart defect can be closed at the same surgical procedure. Bilateral sequential single lung transplantation is most commonly performed for suppurative lung disease such as cystic fibrosis and combined heart/lung performed for disease that cannot be surgically corrected at the time of single lung transplantation. There have been a few reported series of lobar transplantation from a single lobe or bilateral single lobar from living donors or candidates facing imminent death without access to a conventional transplant. Experience has been confined to a few centers of this type.

1.	Heart–lung transplant
2.	Bilateral sequential lung
3.	Single lung transplant
4.	Lobar transplant

Table 4 Types of pulmonary transplantation

Chronic obstructive pulmonary disease noted as emphysema is the most common indication for pulmonary transplantation. More than one-third of all transplant procedures reported through the St Louis International Lung Transplant Registry have been performed for chronic obstructive pulmonary disease. The presence of cystic fibrosis is becoming the most common cause of bilateral sequential pulmonary transplantation.

Selection criteria for pulmonary transplantation include end-stage and physiologically severe disease (Table 5). There should be no alternative medical therapy or surgical therapy available. Life expectancy is extremely limited and in most cases to less than 1 year. Adequate cardiac function without significant coronary artery disease is an absolute prerequisite. The patient must have an adequate nutritional and rehabilitational status and must fulfill a correct psychologic profile and have an adequate emotional support system.

1.	Clinically and physiologic severe disease
2.	Medical therapy ineffective or unavailable
3.	Substantial limitation in activity of daily living
4.	Limited life expectancy
5.	No significant coronary artery disease
6.	Ability to perform pulmonary rehabilitation
7.	Acceptable nutritional status
8.	Satisfactory psychologic profile

Table 5 Patient selection criteria

Contraindications (Table 6) to pulmonary transplantation include:(a) ventilatory dependency; (b) current tobacco use; (c) malignancy within 5 years; (d) significant medical illnesses such as cirrhosis or hepatitis or severe osteoporosis; (f) sputum with pan-resistant bacteria; (g) significant obesity or cachexia; (h) significant coronary artery disease; (i) age greater than 60 years; and (j) an excessive corticosteroid therapy greater than 10 mg/day.

1.	Ventilatory dependence
2.	Current tobacco use
3.	Malignancy within 5 years (experimental exceptions)
4.	Cirrhosis
5.	Hepatitis
6.	Sputum with pan-resistant bacteria
7.	Significant obesity or cachexia
8.	Severe osteoporosis
9.	Significant coronary artery disease
10.	Excessive corticosteroid therapy (prednisone >10 mg)
11.	Age (>60)

Table 6 Contraindication to pulmonary transplant

Lung transplant donor criteria (Table 7) include the following: (a) the donor should be less than 55 years of age and have adequate ABO compatibility; (b) the chest radiograph should be clear and the donor must achieve an arterial PO_2 greater than 300 mmHg on 100 per cent FiO_2 and 5 cm of PEEP. There should be no previous surgery on the side of harvest and no history of aspiration or sepsis. One must have the absence of purulent secretions confirmed at fiberoptic bronchoscopy at the time of donor harvest and a sputum gram stain free of bacteria and fungus.

1.	Age <55
2.	ABO compatibility
3.	Clear chest radiograph
4.	Arterial PO_2 >300 on 100 per cent FiO_2 and 5 cm PEEP
5.	No previous surgery on side of harvest
6.	No aspiration or sepsis
7.	Absence of purulent secretions at bronchoscopy
8.	Sputum gram stain free of bacteria or fungus

PEEP, positive end expiratory pressure.

Table 7 Lung transplant donor criteria

Donor harvest and technique

The performance of successful pulmonary transplantation requires extreme cooperation between the donor harvesting team and the implanting team. The donor lung should be harvested by an experienced surgeon and kept in a hyperinflated state with a cold slush solution at 1 to 4°C. Utilizing this technique, donor presentation up to 6 h has been achieved. The donor extraction technique was reported by Sundaresen *et al.* in 1993. It is beyond the scope of this chapter to describe in detail the technique of donor extraction.

Single lung transplant

Choice of side

The Barnes group prefer to perform the transplant on the side with the least pulmonary function as judged by preoperative nuclear perfusion scans. It was originally felt that the right side was preferable for patients with chronic obstructive pulmonary disease; however, in studies done by Patterson and Cooper (1995) to date it does not appear that there is any difference in long-term outcome among single lung recipients regardless of the transplanted side. If a patient has an associated cardiovascular condition or has primary pulmonary hypertension with severe pulmonary fibrosis, then the right side is the preferred transplanting side. In patient's with Eisenmenger's syndrome the preferred side is also the right. A posterior lateral thoracotomy on the side of implant is the preferred incision assuming that a combined cardiovascular procedure is not contemplated. If a combined correction of a cardiac defect and/or abnormality of the implant is required, then this can be performed through a median sternotomy. A recipient pneumonectomy is then carried out after adequate exposure has been obtained. The pulmonary veins are divided between either transfixtional ligatures or staples distal to their exit from the right atrium in order to leave as long a proximal cuff as possible. The right main pulmonary artery is then divided between staples just proximal to the take-off of the trunchus anterior. The bronchus is then divided just proximal to the take-off of the upper lobe bronchus to leave as long a bronchial stump as possible. Once the recipient pneumonectomy has been completed the cavity is completely washed in appropriate antibiotic solution. The donor lung is then taken from its ice storage and placed on the back table and prepared for implant by an assistant. The bronchial anastomosis is performed first utilizing 4-0 absorbable monofilament suture. The posterior membranous wall is closed with a 4-0 running suture and the anterior cartilaginous wall closed with interrupted 4-0 absorbable monofilament suture. Sutures are placed using a modified mattress technique to intercept the smaller donor bronchus into the larger recipient bronchus. The bronchial anastomosis is generally covered utilizing either local peribronchial tissue or pedicled pericardium or thymic fat. The proximal pulmonary artery on the recipient is then clamped, the ends of the donor artery are trimmed to match it, and this anastomosis then is performed with a running 5-0 prolene suture. The atrial cuff containing the pulmonary veins is then trimmed and the recipient pulmonary veins are then anastomosed to the donor veins utilizing a running 4-0 prolene suture. Two chest tubes are left in the pleural space and the chest closed. A fiberoptic bronchoscopy is then performed to ensure an adequate caliber of the tracheobronchial tree.

Bilateral sequential lung transplantation

Bilateral sequential lung transplantation is performed through a bilateral anterior thoracotomy through the fourth or fifth intercostal space. This has been termed the ‘clam-shell' incision. Once exposure has been obtained and retractors placed, the recipient pneumonectomy is then carried out in a sequential fashion. In general, the left side is performed first and then the right side. The recipient pneumonectomy and implantation techniques are identical to those described under unilateral

implantation. If cardiopulmonary bypass is required, it can be instituted at this time.

Postoperatively the patients are watched in the intensive care unit for several days and are treated with immunosuppressive drugs generally consisting of cyclosporin, azothiaprine, and corticosteroids. The exact management of these is beyond the scope of this chapter.

Complications

The potential complications that can occur are: (a) errors in technical management and making the bronchial anastomosis too small or making the arterial and venous anastomosis too small, and (b) one can have early graft dysfunction seen in as many as 20 per cent of patients and potential infections of bacterial or fungal infection may occur. Causes of early and late death in pulmonary transplantation are as listed in Table 8 and Table 9, respectively.

Cause	*Percentage*
Infection	35
Graft failure	13
Heart failure	9
Bleeding	6
Rejection	5
Anastomosis	5
Other	27

Cooper—International Registry Report, April 1996

Table 8 Causes of early death in lung transplant

Cause	*Percentage*
Infection	30
BOS	29
Malignancy	6
Respiratory failure	5
Bleeding	4
Other	26

Cooper—International Registry Report, April 1996

Table 9 Causes of late death

Mortality

Lung transplantation has now evolved in the past 10 years until it is a widely accepted procedure. If one considers all forms of pulmonary disease for which transplantation is carried out, there is an overall 71 per cent 1-year survival, and a 45 per cent, 5-year survival, if all recipient groups are combined. There does not appear to be any significant difference between patients who undergo isolated single lung transplant or combined heart–lung transplantation. Cooper *et al.* from the St. Louis International Lung Transplant Registry has reported the following. For patients with emphysema there is an 80 per cent 1-year survival and 45 per cent 5-year survival. For patients with cystic fibrosis there is approximately 75 per cent 1-year survival and a 45 per cent 5-year survival. For patients with pulmonary fibrosis there is a 60 per cent 1-year survival and a 40 per cent 5-year survival. For patients with primary pulmonary hypotension there is a 60 per cent 1-year survival and approximately a 42 per cent 5-year survival. Patients who are retransplanted have a 45 per cent1-year survival and approximately 38 per cent 3-year survival.

In summary, a modified algorithm for the overall management of chronic obstructive pulmonary disease is shown in Fig. 2. The role of lung volume reduction surgery in the management of chronic obstructive pulmonary disease remains controversial, but, hopefully, studies being performed by the National Institutes of Health will help define its role. Pulmonary transplantation remains a viable alternative in highly selected patients with endstage pulmonary disease.

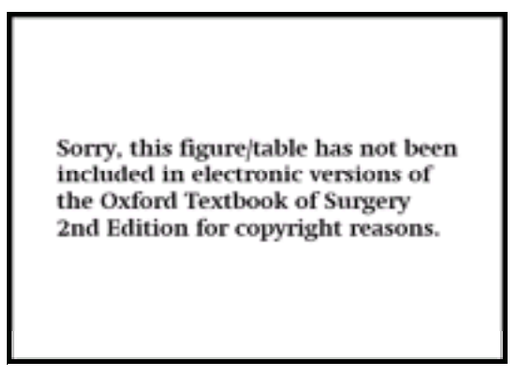


Fig. 2. Algorithm for chronic obstructive pulmonary disease.

Further reading

Brantigan OC, Muallee E, Kress MB. A surgical approach to pulmonary emphysema. *American Revue of Respiratory Disease* 1959; **80**: 194–202.

Cooper JD. *St. Louis International Lung Transplant Registry*, 1993.

Cooper JD *et al.* Bilateral pneumectomy (volume reduction) for chronic obstructive pulmonary disease. *Journal of Thoracic and Cardiovascular Surgery* 1995; **109**: 106–19.

Cooper JD *et al.* Results of 150 consecutive bilateral lung volume reduction procedures in patients with severe emphysema. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**: 1319–30.

Hardy JD. Lung homotransplantation in man. *Journal of the American Medical Association* 1963; **186**: 1065–74.

Lefrak SS. Lung volume reduction surgery. *Annual Revue of Medicine* 1997*a*; **48**: 307.

Lefrak SS. The LVRS debate: where are we, how we got there and where are we going. *Pulmonary Perspectives* 1997*b*; **14**: 1–3.

Miller JI, Lee R, Mansour KA. Lung volume reduction surgery: lessons learned. *Annals of Thoracic Surgery* 1996; **61**: 1464–9

Naunheim KS, Fergason MK. The current status of lung volume reduction operations for emphysema. *Annals of Thoracic Surgery* 1996; **66**: 601–12.

Pasque MK *et al.* Single-lung transplantation for pulmonary hypertension: single institution experience in 34 patients. *Circulation* 1995; **92**: 2252–8.

Patterson GA *et al.* Technique of successful clinical double-lung transplantation. *Annals of Thoracic Surgery* 1988; **45**: 626–32.

Patterson GA, Cooper JD. Lung transplantation. In: Pearson FG, Deslauries J, Ginsberg RJ, Hiebert CA, McKneally M, Vrschel HC, eds. *Thoracic surgery*, pp. 931–60. New York, Churchill Livingstone, 1995.

Rogers RM, Sciurba FC, Keenan RJ. Lung volume reduction surgery in chronic obstructive pulmonary disease. *Medical Clinics of North America* 1996; **80**: 623–44.

Sezkely LA *et al.* Preoperative predictors of operative morbidity and mortality in COPD patients undergoing bilateral lung volume reduction surgery. *Chest* 1997; **111**: 550–8.

St. Louis International Lung Transplant Registry Report. Washington University School of Medicine, St. Louis, MO, 1996.

Sundaresen S *et al.* Donor lung procurement: assessment and operative technique. *Annals of Thoracic Surgery* 1993; **56**: 1403.

Toronto Lung Transplant Group. Unilateral lung transplantation for pulmonary fibrosis. *New England Journal of Medicine* 1986; **314**: 1140–5.

Trulock EP. Recipient selection. *Chest Surgical Clinics of North America* 1993; **3**: 1–18.

William TS, Snell GI. Early and long term functional outcomes in unilateral, bilateral, and living-related transplant recipients. *Clinics in Chest Medicine* 1997; **18**: 245–57.

Yosen RD, Lefrak SS. Evaluation of patients with emphysema for lung volume reduction surgery. *Seminars in Thoracic and Cardiovascular Surgery* 1996; **8**: 83–93.

Yusen RD, Lefrak SS, Trulock EP. Evaluating and preoperative management of lung volume reduction surgery candidates. *Clinics in Chest Medicine* 1997; **18**: 199–224.

41.8 Pneumothorax

Ashby C. Moncure

Aetiology and classification
Primary spontaneous pneumothorax
Secondary spontaneous pneumothorax
Neonatal spontaneous pneumothorax
Catamenial spontaneous pneumothorax
Traumatic pneumothorax
Iatrogenic pneumothorax
Diagnostic pneumothorax
Diagnosis
Management
Further reading

The presence of gas within the pleural space, termed a pneumothorax, is always secondary to disruption of the parietal, visceral, or mediastinal pleura. It may occur as a result of trauma (traumatic pneumothorax) or from of a ruptured subpleural bleb (spontaneous pneumothorax). When pleural gas collects under pressure within the pleural space, consequent upon a one-way flap valve mechanism, a tension pneumothorax is produced.

Aetiology and classification

A classification of pneumothorax is given in Table 1.

Spontaneous pneumothorax
Primary (no identifiable pathology)
Secondary (underlying pulmonary disorder)
Catamenial
Neonatal
Traumatic pneumothorax
Blunt or penetrating thoracic injury
Iatrogenic: postoperative, mechanical ventilation, thoracocentesis, central venous catheterization
Diagnostic pneumothorax

Table 1 Classification of pneumothorax

Primary spontaneous pneumothorax

This is usually found in young, healthy, adult men: 85 per cent of patients are less than 40 years of age and the male:female ratio is 6:1. The disease has an annual incidence of nine cases per 100 000 population.

Pneumothorax is caused by rupture of acquired subpleural blebs; these are usually found in the pulmonary apices, but occasionally occur in the superior segment of the lower lobes. These blebs have no epithelial lining, and possibly arise from rupture of alveolar walls in the area of the bleb. Apical blebs are found in 85 per cent of adults who undergo thoracotomy. Spontaneous pneumothorax is bilateral in 10 per cent of patients. This is potentially life threatening, though not usually so. No abnormalities are found at thoracotomy in 15 to 20 per cent of patients, but active pulmonary tuberculosis may subsequently develop in 2 to 3 per cent. The frequency of primary spontaneous pneumothorax increases after each episode: the chance of a second episode is 50 per cent, a third 62 per cent, and a fourth 80 per cent. Most recurrences occur within 2 years of the initial episode. Three-quarters of them are ipsilateral.

Secondary spontaneous pneumothorax

Spontaneous pneumothorax is secondary to underlying pulmonary disease in 10 to 20 per cent of patients. Chronic obstructive pulmonary disease, usually with bullae formed by progressive destruction of alveolar walls, is the most commonly encountered. These patients tend to be older than average, have compromised pulmonary function, and, because of their limited pulmonary reserve, may experience respiratory failure from the pneumothorax.

Other underlying pulmonary diseases that may cause pneumothorax include interstitial lung disease, infections, primary and metastatic neoplasms, and emboli. Pneumothorax also occurs in Ehlers–Danlos syndrome, Marfan syndrome, and endometriosis.

There is a 50 per cent recurrence rate for secondary spontaneous pneumothorax. Because of the severe underlying pulmonary disease, the disease is associated with a mortality rate of 16 per cent.

Neonatal spontaneous pneumothorax

Pneumothorax in the neonate may be associated with hyaline membrane disease, renal malformation, Potter syndrome, and meconium aspiration. It is also present in children with cystic fibrosis.

Catamenial spontaneous pneumothorax

A spontaneous pneumothorax may occur during menstruation, usually 48 to 72 h after the beginning of menses. The right hemithorax is affected in 90 per cent of patients. The underlying cause of this condition is unknown.

Traumatic pneumothorax

Blunt or penetrating injuries to the chest wall may produce a sucking wound or may cause fractures of the ribs, which then lacerate the lung. Penetrating missiles or a knife may also cause such lacerations. Laryngotracheobronchial or esophageal injury, whether produced by blunt or penetrating trauma, can produce mediastinal emphysema, pleural rupture, and pneumothorax.

Iatrogenic pneumothorax

Pneumothorax is common after pulmonary resection, and is universal for some time after pneumonectomy. It is also frequently present after resection of smaller amounts of pulmonary parenchyma. Placement of an intercostal catheter to vent the pleural space is routine after such resection. Pneumothorax is occasionally seen after thoracocentesis, either as a consequence of air entering the pleural space through the needle or through the catheter used to drain the pleural space, or secondary to penetration of the visceral pleura when the needle is introduced into the pleural space. Mechanical ventilation, particularly with high airway pressures, may rupture pulmonary parenchyma. Placement of central venous lines for monitoring of parenteral nutrition may also produce pneumothorax if their associated needles and guidewires penetrate the pleura.

Diagnostic pneumothorax

Rarely, air may be purposely introduced into the pleural space to assess pleural disease or suspected pleural involvement in patients with pulmonary parenchymal disease.

Diagnosis

The dominant symptoms of pneumothorax from any cause are acute pleuritic chest pain on the affected side, and dyspnea due to pulmonary compression. The severity of symptoms is generally proportional to the magnitude of the pneumothorax, although in patients with multiple trauma, associated injuries may dominate the clinical picture. Mental obtundation from any cause may mask the presence and severity of the pneumothorax.

Absence of breath sounds over the affected hemithorax and, in the presence of a tension pneumothorax, shift of the trachea to the contralateral side, are the physical findings that mark the presence of a pneumothorax.

Although a chest radiograph will usually allow the diagnosis to be clearly established, it may not clearly demonstrate the underlying cause of the pneumothorax. Computed tomographic scanning of the chest, with ‘lung windows', is extremely useful in the assessment of the pulmonary parenchyma, chest wall, and mediastinum.

Management

The management of pneumothorax is based on observation, both clinically and with regular chest radiographs, thoracocentesis, and the introduction of a catheter into the pleural space through the fifth or sixth intercostal space. Blunt penetration of the intercostal space with a Kelly clamp is less likely to injure the lung than insertion of a sharp trochar. The pleural space should be entered on the upper margin of the rib, to avoid injury to the intercostal neurovascular bundle. All holes in the tube must be located within the pleural space: this must be verified by chest radiographs. Thereafter, the tube is placed to water-seal suction drainage, and the patient is observed until absence of an air leak from the pleural cavity and expansion of the lung have occurred.

Patients who have a post-traumatic pneumothorax and who need a general endotracheal anesthetic, or in whom tracheal intubation is required for the purpose of intermittent positive-pressure ventilation, should have an intercostal catheter placed within the affected pleural space promptly. In the presence of a ‘sucking' wound of the chest, one must assume the presence of pulmonary parenchymal injury. Venting of the pleural space with an intercostal catheter must precede temporary closure of the wound in the chest wall, in preparation for later operative repair of the defect.

Although a spontaneous pneumothorax may be no more than that, signs of a tension pneumothorax (hypotension, shift of the mediastinum, engorged jugular vein) should lead to venting with an intercostal catheter immediately, prior to all diagnostic studies. On the other hand, if the spontaneous pneumothorax is the first such episode, a minimal pneumothorax in an asymptomatic patient may be managed by observation and by assessment with periodic chest radiographs. Moderate pneumothoraces (greater than 20 per cent lung collapse) should be managed by insertion of an intercostal catheter connected to water-seal suction drainage. If there is a minimal air leak, but with complete lung expansion, a one-way flutter valve (Heimlich valve) may facilitate evacuation of intrathoracic air until cessation of the air leak allows removal of the tube.

Resection of pulmonary bullae responsible for the air leak, operative abrasion of the pleural lining, or parietal pleurectomy are indicated in patients with an air leak that persists for more than 10 days, in those in whom an intercostal catheter with water-seal suction drainage fails to expand the lung because of a massive air leak, and usually in recurrent spontaneous pneumothorax, or with complications such as hemothorax, empyema, and chronic pneumothorax. Previous contralateral or bilateral simultaneous pneumothorax is best managed by pleurodesis, with access through a median sternotomy. If a specific pulmonary disease amenable to operative management is responsible for the pneumothorax, an appropriate operation should be undertaken. The recent development of thorascopic video-assisted techniques have allowed application of these techniques to resection of apical bullae and pleurabrasion in treatment of spontaneous pneumothorax for the above indications.

Poor-risk patients requiring pleurodesis need to have the lung expanded by closed-tube thoracostomy, followed by chemical pleurodesis with tetracycline hydrochloride (20 mg/kg in 50 ml of saline solution). Ascorbic acid in the tetracycline preparation is the active agent, not tetracycline itself.

Further reading

Baumann MH, Strange C. Treatment of spontaneous pneumothorax. *Chest* 1997; **112**: 789–804. [A report advocating early aggressive treatment of spontaneous pneumothorax.]

Deslauriers J. Transaxillary pleurectomy for treatment of spontaneous pneumothorax. *Annals of Thoracic Surgery* 1980; **30**: 569–74. [A report detailing the management of 409 patients with spontaneous pneumothorax by transaxillary pleurectomy.]

Gobbel WG, Rhea WG, Nelson IA, Daniel RA. Spontaneous pneumothorax. *Journal of Thoracic and Cardiovascular Surgery* 1963; **46**: 331–45. [A review of the natural history of the disease and its management at the Vanderbilt University Medical Center in 135 patients.]

Mercier C, Page A, Verdant A. Outpatient management of intercostal tube drainage in spontaneous pneumothorax. *Annals of Thoracic Surgery* 1976; **22**: 163–4. [A large series of patients with spontaneous pneumothorax, 54 per cent of whom were successfully managed by intercostal tube drainage and a flutter valve.]

Rhea JT, DeLuca SA, Greene RE. Determining the size of pneumothorax in the upright patient. *Radiology* 1982; **44**: 733–6. [A study describing a simple and radiographically reproducible measurement that accurately predicts size of pneumothorax.]

Shields TW, Oilschlager GA. Spontaneous pneumothorax in patients 40 years of age and older. *Annals of Thoracic Surgery* 1966; **2**: 377–83. [A paper describing management in the difficult subgroup of patients with spontaneous pneumothorax (elderly, infirm).]

Stephenson LW. Treatment of pneumothorax with intrapleural tetracycline. *Chest* 1985; **88**: 803–4. [A description of the use of chemical pleurodesis in the treatment of spontaneous pneumothorax.]

41.9 Chylothorax

Ashby C. Moncure

Introduction
Anatomic and biochemic considerations
Etiology of chylothorax
Diagnosis
Management
Further reading

Introduction

The presence of lymphatic fluid within the pleural space is termed chylothorax. Because of the high concentration of fat, protein, lymphocytes, and antibodies in the lymph of the thoracic duct, continuous loss of this fluid can produce serious nutritional and immunologic deficiencies. Its accumulation within the pleural and pericardial spaces may encroach upon the lung and heart, seriously compromising their function.

Anatomic and biochemic considerations

The lymph originating in the intestine (chyle) contains 0.4 to 5 g per cent total fat, 65 to 220 mg per cent total cholesterol, 1.2 to 4.1 g per cent albumin, 1.1 to 3.6 g per cent globulin, and electrolytes in similar concentrations to those of plasma. It also contains 400 to 6800 lymphocytes and 50 to 600 erythrocytes per mm^3.

Of the thoracic lymph, 95 per cent originates in the liver and intestinal tract. Some 60 to 70 per cent of ingested fat is transported to the bloodstream by the thoracic duct; only fatty acids with fewer than 10 carbon atoms in the chain are absorbed directly into the portal venous system.

The thoracic duct (Fig. 1) is a muscular tube that communicates with an extensive collateral circulation. Its multiple, unidirectional valves conduct lymph in a cephalad direction. The thoracic duct usually originates from the cisterna chyli on the anterior surface of the vertebral column at about the level of L2 and ascends through the diaphragmatic aortic hiatus behind the esophagus and between the aorta and the azygous vein on the surface of the right anterior vertebral column to about the level of T5. There the duct crosses behind the aorta and aortic arch, ascending through the left posterior mediastinum along the esophagus into the base of the left neck, and emptying into the venous system in the area of the confluence of the left jugular and subclavian veins.

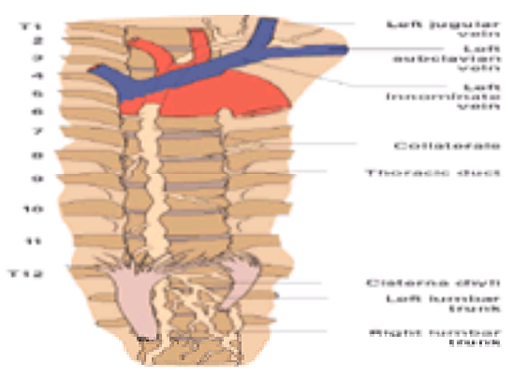


Fig. 1. Schematic drawing of the usual pattern and course of the thoracic duct. The single duct that enters the mediastinum through the aortic hiatus between T12 and T10 is a relatively consistent finding and the usual site for surgical ligation.

After the ingestion of a fatty meal the volume of intestinal lymph may increase by 10 times over its basal flow rate; starvation and rest markedly diminish the volume.

Etiology of chylothorax

DeMeester has published a helpful classification of the causes of chylothorax (Table 1).

Congenital
Atresia of thoracic duct
Thoracic duct–pleural space fistula
Birth trauma
Traumatic
Blunt
Penetrating
Surgical:
Cervical:
excision of lymph node
radical neck dissection
Thoracic:
ligation of patent ductus arteriosus
excision of coarction of aorta
esophagectomy
resection of thoracic aortic aneurysm
left pneumonectomy
Abdominal:
sympathectomy
radical lymph-node dissection
Diagnostic procedures:
Translumbar arteriography
Subclavian vein catheterization
Neoplasms
Miscellaneous

Table 1 Aetiology of chylothorax

Congenital chylothorax is rare, and due to birth trauma or congenital defects or malformations. External trauma, usually hyperextension of the spine and, rarely, penetrating missile injuries, and intraoperative trauma cause approximately 50 per cent of cases of chylothorax. Malignant neoplasms, usually lymphoma, bronchogenic carcinoma, or other metastatic mediastinal malignancies, also cause approximately 50 per cent. Other rare causes include tuberculosis, aortic aneurysm, filariasis, and pulmonary lymphangiomatosis.

Diagnosis

The rapidity with which a chylous pleural effusion accumulates is reflected in the symptoms associated with it. Gradual chylothorax is associated with the insidious onset of dyspnea; rapid accumulation causes marked respiratory distress and shock. The clinical context, such as a postoperative or post-traumatic state or the presence of an antecedent malignancy, may suggest the diagnosis.

The chest radiograph demonstrates an opacified hemithorax or suggests the presence of a pleural effusion. Computed tomography of the chest confirms the presence of a pleural effusion and may reveal its underlying cause if an abnormality is identifiable within the chest.

The recovery of nonclotting, milky fluid from the pleural space, with analysis revealing free microscopic fat and a fat content higher than that of plasma, is strongly suggestive of chylothorax. Other causes of milky pleural fluid are pseudochyle (present with malignant tumors or infection), which contains only a trace of fat, and cholesterol pleural effusions (associated with tuberculosis or rheumatoid arthritis), which contain a high concentration of cholesterol crystals. If the milky pleural fluid has a triglyceride concentration of more than 110 mg/100 ml, there is a 99 per cent likelihood that the fluid is chyle; if the concentration is less than 50 mg/100 ml, there

is only a 5 per cent likelihood.

Management

The management of chylothorax depends upon the underlying cause and the extent of nutritional depletion. Lampson and associates in 1948 applied ligation of the thoracic duct in the management of chylothorax and decreased the mortality from 50 to 15 per cent. Surgical treatments include supradiaphragmatic ligation of the thoracic duct, shunting into the peritoneal cavity or into one hemithorax or the other, and pleurectomy.

Conservative management includes control of the encroachment upon the lung by the chylous pleural effusion using closed drainage by intercostal catheter. Total parenteral nutrition is required, with medium-chain triglycerides being administered orally.

Radiation therapy is occasionally successful in patients with mediastinal lymphoma and carcinoma.

In the neonate, idiopathic chylothorax generally responds to thoracocentesis. Traumatic chylothorax may initially be treated conservatively as outlined above. However, if pleural fluid losses exceed 1500 ml/day for 1 week, or 500 ml/day for 2 weeks, operative management (right thoracotomy and ligation of the supradiaphragmatic thoracic duct) should be employed. If conservative management is unsuccessful after 2 weeks in those patients with underlying malignancy, irradiation to the tumor bed may be useful.

Further reading

Bessone LN, Ferguson TB, Burford TH. Chylothorax: a collective review. *Annals of Thoracic Surgery* 1971; **12**: 527–50. [A review of the world literature on chylothorax, the embryology, anatomy and physiology of the thoracic-duct system, and a classification of types of chylothorax.]

Boyd A. Chylothorax. In: Hood RM, ed. *Surgical disease of the pleura and chest wall*, pp. 172–83. Saunders, Philadelphia, 1986. [An excellent review of the anatomy, physiology, diagnosis, and management of chylothorax.]

Cerfolio RJ, Allen MS, Deschamps C, Trastek VF, Pairolero PC. Postoperative chylothorax. *Journal of Thoracic and Cardiovascular Surgery* 1996; **112**(5): 1361–5. [A recent report on the results of a large and clinically managed series.]

DeMeester TR, Lafontaine E. The pleura. In: Sabiston DC, Spencer FC, ed. *Surgery of the chest*, 4th edn, pp. 455–8. Saunders, Philadelphia, 1990. [An excellent treatment of the diagnosis and management of chylothorax.]

Lampson RS. Traumatic chylothorax: a review of the literature and report of a case treated by mediastinal ligation of the thoracic duct. *Journal of Thoracic Surgery* 1948; **17**: 778–91. [An early report advocating operative management of traumatic chylothorax by mediastinal ligation of the thoracic duct.]

Miller JI, Jr. Diagnosis and management of chylothorax. *Chest Surgery Clinics of North America* 1996; **6**(1): 139–48. [A well-organized and perceptive review of the topic.]

Milson JW, Kron IL, Rheuban KS, Rodgers BM. Chylothorax: an assessment of current surgical management. *Journal of Thoracic and Cardiovascular Surgery* 1985; **89**: 221–7. [A review of the management of chylothorax in infants and adults.]

Patterson GA, Todd TR, Delarue NC, Ilves R, Pearson FG, Cooper JD. Supradiaphragmatic ligation of the thoracic duct in intractable chylous fistula. *Annals of Thoracic Surgery* 1981; **32**: 44–9. [A report of experience with operative management of adult chylous fistulas with use of supradiaphragmatic mass ligature technique.]

Staats BA, Ellefson RD, Budahn LL, Dines DE, Prakash UB, Offord K. The lipoprotein profile of chylous and nonchylous pleural effusion. *Mayo Clinic Proceedings* 1980; **55**: 700–4. [A biochemical analysis of the pleural fluid in chylothorax and nonchylous pleural effusions.]

41.10 Respiratory infections

Malcolm K. Benson

Introduction
Empyema thoracis
Clinical features
Investigations
Management
Bronchiectasis
Pathology and pathogenesis
Clinical features
Investigations
Treatment
Tuberculosis
Epidemiology
Clinical features
Diagnosis
Management
Further reading

Introduction

Respiratory-tract infections are common and can range from mild infective bronchitis to potentially fatal pneumonia. They represent a significant risk for postoperative morbidity and mortality; these problems are considered elsewhere. There are, however, certain infective conditions affecting the lungs and pleural cavity that have surgical relevance and may require operative treatment.

1. A thoracic empyema is a collection of infected fluid in the pleural space: it is one of those conditions in which a variety of therapeutic options results in diverse views on the most appropriate management in a particular individual.
2. Bronchiectasis is a morphological disorder in which dilated and damaged bronchi result in recurrent bronchial sepsis. Although the role of surgical resection of the affected area has diminished, operative treatment may be appropriate in selected patients.
3. Tuberculosis is a chronic infection, which, although mainly affecting the lung, can result in disease elsewhere. It may, therefore, present as a diagnostic problem to physicians or surgeons and on occasions the diagnosis is only made after histological examination of resected material. Surgical treatment of tuberculosis is now largely a matter of historical interest, although may occasionally have a role for specific complications or in the management of multidrug-resistant disease.

Empyema thoracis

A thoracic empyema is a consequence of pleural infection, usually from adjacent lung. Community-acquired pneumonias, including tuberculosis or a lung abscess, are the most common sources. Infection can also result from penetrating trauma, thoracic surgery, a subdiaphragmatic abscess, or oesophageal perforation. Specific factors that may contribute to the development of an empyema include general debilitation, aspiration of oropharyngeal secretions, or a bronchial obstruction. It is a relatively common condition, with estimates of approx. 4000 cases per annum in the United Kingdom. Despite current treatment, empyema carries a significant mortality, although this may in part reflect associated pathology that might have contributed to its development.

The pleural response to inflammation and infection may go through a number of phases, all of which may resolve spontaneously or after treatment. ‘Dry' pleurisy results from inflamed pleural surfaces but there is clinically insignificant pleural fluid. A parapneumonic effusion is an exudative reaction resulting in straw-coloured fluid; this can then evolve into a loculated collection of turbid fluid with a high cellular content. The final phase is represented by multiple loculations and extensive fibrosis with a grossly thickened pleura. At this stage, the lung is unlikely to re-expand even after pus has been evacuated, and decortication becomes necessary.

Clinical features

The most common presentation is as an acute infective illness in association with a pneumonia. Symptoms are those of an infection, with either rigors or malaise, together with specific respiratory symptoms of pleuritic chest pain, cough, or breathlessness. The failure of a pneumonia to resolve rapidly despite appropriate antibiotics, together with clinical and radiological evidence of a pleural effusion, should alert the clinician to the possibility of an empyema. Less commonly the presentation is insidious, with malaise, sweats, anorexia and weight loss, features that are occasionally misdiagnosed as due to a malignancy.

Investigations

A standard chest radiograph showing evidence of an effusion is often the first clue to an empyema. Pleural collections are likely to be loculated and may contain an air–fluid level due to the development of a bronchopleural fistula or, more rarely, as a result of a gas-forming organism. Ultrasonographic examination or a computed tomographic scan can be useful in confirming the presence of fluid, demonstrate loculi, and help guide pleural aspiration.

A definitive diagnosis of a suspected empyema is usually made from pleural aspiration. However, as already indicated, the spectrum of pleural responses from a free-flowing exudate to multiloculated collections of purulent fluid constitutes a dilemma about the precise defining characteristics of an empyema as well as uncertainty over the need for pleural drainage. Pleural fluid should be sent for cytological, microbiological, and biochemical analysis. Purulent fluid contains large numbers of white cells, usually in excess of 10 000/mm^3. Neutrophils are associated with pyogenic infections and a predominant lymphocytosis should alert the clinician to the possibility of tuberculosis. Biochemical changes associated with infection include an acidosis, a low glucose, and an elevated lactic dehydrogenase in pleural fluid; a pH of less than 7.2 has been identified as one of the best indicators of the need for pleural drainage, although the size of the effusion and the severity of the illness need to be taken into consideration. A positive microbiological diagnosis can be made in a majority of patients, but may be difficult in those who have received prior treatment with antibiotics. The spectrum of organisms most frequently encountered in the United Kingdom is listed in Table 1.

Single organisms (75%)	
Gram-positive aerobes:	
Streptococcus milleri	+++
Streptococcus pneumoniae	++
Staphylococcus aureus	+
Gram-negative aerobes:	
Escherichia coli	+
Haemophilus influenzae	+
Proteus	+
Anaerobes:	
Bacteroides melaninogenicus	++
Streptococci	+
Fusobacterium	+
Fungi:	
Candida spp.	+
Multiple organisms (25%)	
Streptococcus milleri plus anaerobes	

+, rare; ++, common.

Table 1 Microbiology of empyema thoracis

Management

Treatment of an empyema requires a combination of appropriate antibiotics together with early and effective drainage. The initial choice of antibiotic is largely determined by clinical circumstances, but may require modification in the light of microbiological cultures. Following a community-acquired pneumonia, any antibiotic regimen must provide a cover against pneumococci. High-dose ampicillin or cefuroxime (750 mg–1.5 g 8-hourly) are reasonable choices. Anaerobic infection should be

suspected when there is dental sepsis or the possibility of aspiration, and aspirated pus is often foul smelling. Metronidazole is a reasonable addition to cefuroxime. Gram-negative organisms may require the addition of an aminoglycoside.

Perhaps the area of greatest uncertainty and debate is how best to achieve adequate pleural drainage and thus allow the lung to re-expand. The range of options extends from simple needle aspiration through to thoracotomy and decortication. Rarely there may be a need for more complex surgical procedures such as the creation of an Eloesser flap or thoracoplasty. New drainage techniques utilizing radiographic guidance, the use of intrapleural fibrinolysis, and video-assisted thoracoscopic surgery have broadened the range of therapeutic options and may reduce the need for thoracotomy. In general, needle aspiration is only appropriate for a parapneumonic effusion that is not loculated and is easy to aspirate. Where there is pus in the pleural space, intercostal drainage requires the insertion of a catheter with strict aseptic technique, ideally under radiological guidance. Flexible, fine-bore catheters are undoubtedly less traumatic to the patient and observational studies report encouraging outcomes provided there is good care of the catheter with adequate irrigation. There are, however, acknowledged experts who continue to advocate the use of wide-bore catheters. The use of intrapleural fibrinolytics (streptokinase, urokinase) seems to improve pleural drainage and may result in a breakdown of adhesions and fibrinous loculi. Initial reports need to be treated with some caution as there is limited evidence of benefit in mortality rates or need for surgery. The precise role of video-assisted thoracoscopic surgery in the management of a pleural empyema has yet to be established, but it may reduce the need for thoracotomy. It is unlikely to be appropriate in the late stages of an empyema when there is organization with fibrosis and a grossly thickened pleura.

In summary, initial management requires antibiotics with intercostal drainage, with or without fibrinolysis. Surgical intervention is more likely to be needed where there has been chronic infection, where there is radiological evidence of gross pleural thickening and multiloculation, or when initial measures fail to achieve rapid resolution.

Bronchiectasis

Bronchiectasis is a morphological disorder that represents end-organ damage to the bronchi. Inflammation, usually due to infection, combined with localized obstruction, results in destruction and dilatation of parts of the bronchial tree, which may be localized and isolated to a particular segment or more generalized. The consequence of this anatomical disruption is failure of normal mucociliary clearance, pooling of secretions, and recurrent bronchial sepsis. Cough and excessive production of sputum are the major symptoms. If there is more generalized disease, breathlessness and respiratory failure can ensue. The true incidence is uncertain, although there is the impression that it has become less common in the developed world. The clinical symptoms of cough and sputum production are common to a number of bronchial disorders, so without resort to complex radiological techniques such as computed tomographic scans or bronchography, bronchiectasis is probably underdiagnosed.

Pathology and pathogenesis

Gross inspection of bronchiectatic lungs reveals dilated, inflamed bronchi that are distorted and have collapsible walls. They often contain mucus and there is variable fibrosis of the surrounding parenchyma. The extent of the disease is variable, ranging from a limited area confined to subsegmental bronchi through to extensive disease involving much of the bronchial tree. Whitell's classic description in 1952 defines three main types from pathological features: saccular, follicular, and atelectatic. Of these, follicular is the most common, but the value of this classification in understanding the pathogenesis is uncertain.

Factors that might cause bronchiectasis are poorly understood, but a combination of infection and bronchial obstruction could be the initial event. The extent to which childhood infection is a causative factor is unclear; a number of studies have suggested an association, but these were largely based on retrospective surveys. Such postinfective damage now seems to be less likely, probably because of the benefits of immunization and antibiotic treatment. Increasingly, the focus has shifted to identifying host factors that increase susceptibility to inflammation and lung damage. These include abnormalities of ciliary function, immune defects, abnormalities of mucus production, and hyperimmune states (Table 2). As yet, only a minority of patients have a clearly identifiable abnormality and it seems likely that a combination of factors results in the clinical disease.

Intrinsic host defence defects
Immune defects
Hypogammaglobulinaemia
Selective immunoglobulin deficiencies
Functional antibody deficiencies
Complement deficiencies
Phagocyte defects
Chronic granulomatous disease
Mucociliary clearance defects
Cystic fibrosis
Immotile cilia
Kartagener's syndrome
Young's syndrome
Bronchial obstruction
Recurrent aspiration
Extrinsic factors
Infective
Tuberculosis, measles, whooping cough, adenovirus, HIV infection
Inflammatory
Gastric aspiration
Toxic gases
Inhaled foreign body
Miscellaneous
Inflammatory bowel disease
Rheumatoid arthritis
α_1-Antitrypsin deficiency
Yellow nail syndrome
Post-lung transplant
Allergic bronchopulmonary aspergillosis
Anatomical
Defective bronchial cartilage
Pulmonary sequestration

Table 2 Disorders predisposing to bronchiectasis

Clinical features

At one end of the clinical spectrum are patients whose only manifestation is occasional infective exacerbations with cough and sputum production. In between, they remain well and asymptomatic for months or years. At the other extreme are patients who are severely disabled, with cough producing large volumes of purulent sputum, breathlessness, and general debility. Pleuritic chest pain and haemoptysis are episodic features, usually associated with infective exacerbations. Chronic sinusitis and otitis are present in a significant proportion of patients; whilst these could contribute to the development of bronchiectasis, they may simply reflect an abnormality of defence mechanisms in both the upper and lower respiratory tracts.

Physical signs are equally variable and examination may be normal. Coarse crackles can be heard over the affected area, and clubbing is sometimes a feature in patients with chronic sepsis. If there is generalized disease with airflow obstruction, there may be features of respiratory failure and cor pulmonale. Other complications, such as secondary amyloidosis or metastatic spread of infection to such sites as the brain, are now rarely seen.

Investigations

Diagnostic confirmation of bronchiectasis depends on radiographic abnormalities. A posteroanterior chest radiograph is often normal, although it may show thickened bronchial walls—'tramlining'. Gross cystic changes, often with fluid levels, are easily visible but are rare. Bronchography, in which a radiopaque contrast medium is instilled into the bronchial tree, has now been superseded by computed tomography, which will show the presence of bronchiectatic areas of lung and the extent of the disease and is a prerequisite to any consideration of surgery.

Microbiological investigation of a sputum sample may help to guide the choice of antibiotic. Usually there is a mixed growth and often the organisms are those normally found in the upper respiratory tract. Organisms that are likely to be significant include *Haemophilus influenzae* (unencapsulated), *Streptococcus pneumoniae, Moraxella catarrhalis*, and *Pseudomonas aeruginosa*. Sputum should also be examined for mycobacteria.

Lung function should be tested in patients with more extensive disease and those who are breathless. Spirometry usually shows an obstructive defect, although often the picture is mixed, with some restriction in addition.

Final consideration should be given to possible factors predisposing to bronchiectasis, especially those that might lead to improved treatment. Measurement of immunoglobulins occasionally reveals panhypogammaglobulinaemic states and such patients may benefit from immune replacement. Subclass deficiencies, especially of IgA, IgG2, and IgG4, are more common, but their clinical importance is uncertain. Abnormalities of mucociliary clearance are inherited as an autosomal recessive with a frequency of about 1:20 000. A simple test of function involves placing a saccharine tablet on the inferior nasal turbinate and measuring time to sense for taste.

Cystic fibrosis is usually obvious from childhood but can occasionally present in early adult life with respiratory symptoms. Aside from the clinical features, the skin sweat test is the single most useful investigation, although genetic studies are helpful diagnostic adjuncts.

Other possible investigations include assessment of atopy and evidence of allergic bronchopulmonary aspergillosis with positive *Aspergillus* precipitins. a_1-Antitrypsin

deficiency is more commonly associated with emphysema but may predispose to bronchiectasis.

Treatment

The mainstay of medical treatment is a combination of physiotherapy and antibiotics. Postural drainage has long been regarded as a useful way of clearing accumulated mucus, although there is relatively little documented evidence of benefit. The patient is taught the most appropriate position that will enable gravity to assist expectoration; this can furthered by percussion techniques and the use of a forced expiratory ‘huffing' manoeuvre. Ideally, physiotherapy should be undertaken regularly, once or twice daily, although compliance is often poor.

Antibiotics

There is little doubt that antibiotics are beneficial in acute infective exacerbations, both in reducing the volume and purulence of sputum and in treating the constitutional symptoms. Many patients simply require occasional short courses of antibiotics for 1 or 2 weeks during these episodes. The first choice of antibiotic is likely to be amoxycillin (amoxicillin), a tetracycline, or cefaclor. Patients with more frequent infective exacerbations or those whose condition relapses once antibiotics have been stopped are likely to require antibiotics on a more regular basis. Two options exist. One is to use continuous antibiotics, which are changed every few weeks to reduce the development of resistant organisms. The alternative is to give ‘pulsed' courses of intravenous antibiotics for short periods, repeated every 2 or 3 months. This approach has been successfully used in some patients with cystic fibrosis. In practice, treatment tailored to suit an individual patient is preferable to strict adherence to a fixed regimen.

The antibiotic chosen should have a sufficiently wide spectrum to cover the organisms that most commonly colonize the bronchial tree. If the initial choice fails, further treatment should be based on the results of sputum culture and sensitivity testing. Colonization with *Pseudomonas* is common, and its eradication virtually impossible. Treatment usually requires parental administration of antibiotics, but the decision to treat should be taken on clinical grounds rather than simply on the basis of a microbiological isolate. Maintenance treatment with nebulized antibiotics can, in some instances, be beneficial.

Patients with generalized disease may develop airway obstruction and respiratory failure. Any reversal from treatment with inhaled bronchodilators is likely to be small, but even a limited improvement in airway function can be beneficial. It may also aid mucociliary clearance. Steroids and mucolytics have been used, but no consensus on their benefit has been reached.

Surgical treatment

Surgical resection of bronchiectatic lung was popular in the middle part of the twentieth century and offered a potential cure. The fact that it is now undertaken rarely may be due to a number of factors. The prognosis has undoubtedly improved with antimicrobial treatment. In addition, it is uncommon to have disabling symptoms that are due to localized disease, and even in the best reported series a significant proportion of patients remain symptomatic after apparently successful resection. Thus, when assessing for possible surgery, it is important to consider the severity of the symptoms, the anatomical extent of the disease, the degree of respiratory reserve, and, perhaps most importantly, to exclude any pre-existing disorder that might predispose to continued infection. Patients who do best after surgery are those in whom there is localized disease that can be attributed to a specific event. The other possible indication for surgery is haemoptysis. Small haemoptyses are common and major bleeding is rare. If this cannot be controlled bronchoscopically or by bronchial arterial embolization, surgical resection may be necessary.

Tuberculosis

Tuberculosis results from infection with *Mycobacterium tuberculosis*, a rod-like bacillus first identified by Robert Koch in 1882. Identification using the Ziehl–Nielsen stain reveals organisms that after staining with carbol fuschin resist decolorization with acid. The generic term, tubercle bacilli, incorporates several species including *M. tuberculosis, M. bovis*, and *M. africanum*, all of which are human pathogens. In addition, there are a number of non-tuberculous ‘atypical' mycobacteria that are of low pathogenicity. Thus a definitive diagnosis still depends on culture techniques, which together with sensitivity testing help to determine clinical management.

Epidemiology

Tuberculosis is one of world's major public-health problems. It is estimated that approximately one-third of the world's population has been infected with tuberculosis, although only a fraction will develop clinical disease: of one hundred million newly infected individuals each year, approximately eight million will develop disease and three million will die as a result. The major burden of disease still falls on the developing countries and is of particular concern in sub-Saharan Africa. This is largely due to the impact of human immunodeficiency virus (**HIV**) infection, which greatly increases the risk of developing disease in individuals who have been previously exposed to infection. In the industrialized Western countries, where the incidence of disease had been steadily falling since the beginning of the nineteenth century, the rate of decline has slowed or the incidence has even started to rise. The reasons for this are complex and include HIV infection, migration, social deprivation, and a relaxation of public-health measures.

An additional major concern is the emergence of organisms that are resistant to standard drug regimens. Strains resistant to at least one drug are common in many non-industrial countries, but remain rare in previously untreated patients born in the United Kingdom. Resistance to isoniazid is considerably higher in the Indian subcontinent, South-East Asia, and parts of Africa, ranging from 10 to 30–per cent. In the past few years there have been several outbreaks of multiresistant infection in the United States and, to a lesser extent, in the United Kingdom.

Clinical features

In considering the clinical manifestations of tuberculosis, it is important to recognize that only a proportion of patients infected with *M. tuberculosis* develop clinical disease. Initial exposure is usually by inhalation of expectorated droplets containing bacilli. This infection may resolve completely with no clinical manifestations other than the development of an immune response. Alternatively, it may result in disease at the time of initial exposure (primary disease), or it may be contained with the potential to reactivate at some time in the future (postprimary disease).

Primary tuberculosis results in a focal area of infection with dissemination to regional nodes, the primary complex. Patients may be asymptomatic and the disease found on routine contact tracing. Cough, fever, or anorexia may result; less common complications include bronchopneumonia, pleural effusions, and miliary dissemination. Microbiological confirmation may be difficult but tuberculin conversion occurs within about 6 weeks of initial infection.

Postprimary disease develops in patients previously exposed to tuberculosis and may result from re-infection or reactivation. Pulmonary disease is common and may evolve slowly over weeks or months. Thus the presentation is often subacute, with non-specific symptoms such as malaise, sweats, anorexia, and weight loss. Respiratory symptoms include cough and haemoptysis, and less commonly chest pain or breathlessness. Lung signs are variable and depend on the extent of the disease. There may be focal crackles or bronchial breathing. More chronic disease results in scarring with volume loss.

Diagnosis

Although the diagnosis can be suspected on clinical and radiological grounds, diagnostic confirmation depends on microbiological culture of *M. tuberculosis*. Radiological changes are sensitive but non-specific. The classical picture is one of apical consolidation, often with soft, nodular shadows and associated cavitation. More widespread consolidation occurs with tuberculous bronchopneumonia. Calcification and volume loss due to fibrosis indicate more chronic or healed disease. Occasionally, suspected disease is treated without microbiological confirmation, although concerns about the emergence of resistant organisms mean that, where possible, samples should be obtained for culture and sensitivity testing. Sputum, bronchoscopic lavage, gastric lavage, and pleural aspiration may all yield positive results. The finding of acid-fast bacilli on direct microscopy provides rapid presumptive evidence of tuberculosis, but the technique is not specific and fails to distinguish *M. tuberculosis* from other mycobacteria. Cultures, which take 6 to 8 weeks, provide a definitive diagnosis, but treatment may need to be started before there is this microbiological confirmation. As a result there is increasing interest in more rapid diagnostic techniques using molecular biological methods. Tests that identify RNA or DNA from *M. tuberculosis* are available but their lack of sufficient specificity and sensitivity means that results should be interpreted with caution. This, together with their cost, means that their impact on the global problem of tuberculosis is likely to be minimal for the foreseeable future.

Non-respiratory tuberculosis

Although pulmonary disease accounts for the majority of tuberculosis, in a significant minority of cases presentation is in other sites, including lymph nodes, bones and joints, genitourinary structures, pericardium, skin, and meninges. Nodal disease is more common among individuals from Asian subcontinent and in immunocompromised patients. Confirmation of the diagnosis is often made after surgical biopsy. Histological findings of granulomas, giant-cell inflammation, and

caseation suggest tuberculosis. Acid-fast bacilli may be present, but are often sparse and difficult to find.

Management

The management and control of tuberculosis require a combination of public-health measures and specific treatment for infected patients. Such treatments should be supervised by physicians with full training in the management of tuberculosis and with direct access to tuberculosis health visitors or nurse specialists. Whilst the majority of patients can be treated with a standard drug regimen, issues such as monitoring and recognition of side-effects, the use of directly observed therapy, the management of drug-resistant disease, prophylactic chemotherapy, and the care of patients with HIV infection and tuberculosis, all require specialist expertise.

Drug treatment of active disease

The treatment of tuberculosis depends on the use of appropriate combination chemotherapy. Supervision and monitoring are necessary, since successful treatment is crucially dependent on the patient's compliance. A combination of antituberculous drugs is required to minimize the risk of emergence of resistant organisms and, because of the slow growth of mycobacteria, a prolonged course of treatment is also essential for therapeutic success. In the last 20–years there have been a number of modifications in treatment regimens, with the evolution of short-course regimens for a period of 6 months.

The standard drugs used in the treatment of tuberculosis are listed in Table 3. For the majority of disease, both pulmonary and extrapulmonary, a 6-month course comprising an initial 2 months of rifampicin, isoniazid, pyrazinamide and ethambutol, followed by a further 4 months of rifampicin and isoniazid is recommended. Use of ethambutol in the initial phase may be unnecessary in patients in whom there is a low risk of isoniazid resistance. In patients who may not take their medication reliably, supervised chemotherapy can be given three times a week using all four drugs initially and continuing with high-dose isoniazid and rifampicin.

Drug	Adult dose	Main side-effects	Other comments
Isoniazid	300 mg daily	Peripheral neuropathy prevented by pyridoxine; Hepatitis	Primary resistance high (20% in Far East, over 4% in United Kingdom)
Rifampicin	450–600 mg daily	Hepatitis; transient 'flu-like reactions; Rashes	Enzyme induction; increases metabolism of: oestrogen-type pill, anticonvulsants, digoxin, oral steroids
Pyrazinamide	35 mg/kg daily (maximum 3 g)	Gastrointestinal intolerance; Hepatitis	Efficacy limited to first 2 months of treatment
Ethambutol	25 mg/kg daily	Retrobulbar neuritis (dose-related)	Test visual acuity
Streptomycin, Capreomycin, viomycin	15 mg/kg (maximum 1 g) daily	Tinnitus, ataxia, vertigo, Renal impairment	Given by intramuscular injection, Reduce dose in renal impairment
Ethionamide	375–500 mg, twice daily	Gastrointestinal, Hepatitis	Avoid in pregnancy
Cycloserine	250 mg, twice daily	Depression, Fits	
Thiacetazone	150 mg daily	Gastrointestinal Vertigo Rash	Avoid in pregnancy
Ciprofloxacin	750 mg, twice daily	Headache	
Ofloxacin	400 mg, twice daily	Abdominal discomfort	
Rifabutin	3 g, twice daily	Gastrointestinal, Rash, Fever, Hepatitis	
Clarithromycin	500 mg, twice daily	Gastrointestinal	

Table 3 Drugs used in treatment of tuberculosis

Although there have been fewer controlled trials in the treatment of non-respiratory tuberculosis, available evidence suggests that the 6-month short-course regimen is an appropriate and effective treatment for nodal disease, bone and joint disease and genitourinary infections. Tuberculous meningitis may require a more prolonged course of treatment for up to 12 months, and corticosteroids are recommended for patients with more severe disease. As the non-respiratory disease is relatively rare, there are no good randomized trials.

Treatment in association with other diseases

Patients with renal failure requiring treatment can be given standard doses of rifampicin, isoniazid, and pyrazinamide. Both streptomycin and ethambutol are best avoided, although they can be used in reduced dosage with monitoring of serum concentrations.

Patients with liver disease can be treated with standard tuberculous chemotherapy. However, since rifampicin, isoniazid and pyrazinamide are all potentially hepatotoxic, careful monitoring of liver function is mandatory.

Standard treatment is safe in pregnancy. None of the first-line drugs is teratogenic. Streptomycin can be ototoxic to the fetus, and ethionamide and prothionamide are best avoided.

Surgical treatment

Before the advent of chemotherapy for tuberculosis, various surgical procedures provided one of the mainstays of treatment. The consequences of thoracoplasty, plombage, a phrenic crush or artificial pneumothoraces are still apparent in some elderly patients. As successful drug treatment emerged, the need for surgery declined. Currently, surgical resection of solitary nodules for diagnostic purposes continues to reveal occasional cases of tuberculosis. Therapeutic surgery has, until recently, largely been reserved for rare complications such as massive haemoptysis, bronchopleural fistula, or tuberculous empyema. The recent emergence of multidrug resistance has rekindled interest in combining medical and surgical treatments, which includes resection of all severely damaged lung and cavitating disease.

Prevention and control

A variety of methods have been adopted to prevent and control the spread of tuberculosis. They involve identifying individuals and groups at particular risk of disease, instituting appropriate screening measures, and, where appropriate, offering chemoprophylaxis. It is essential that all patients with active tuberculosis should be notified to facilitate contact tracing: this is an effective case-finding method; approx. 10–per cent of close contacts of patients with smear-positive pulmonary tuberculosis develop the disease. Patients with smear-negative or non-pulmonary disease can be regarded as non-infectious. Close contacts are usually people living in the same household, although contact at work is occasionally close enough to represent a significant risk. Casual contacts run a low risk, although a highly infectious index case can occasionally cause an outbreak of tuberculosis.

Contacts, as well as immigrants (to the United Kingdom) from areas with a high incidence of tuberculosis, are investigated by Heaf testing and chest radiography. Interpretation of the Heaf test depends on knowledge of prior vaccination with the bacillus of Calmette–Guèrin (**BCG**). Chemoprophylaxis using a combination of rifampicin and isoniazid for 3 months is designed to reduce the chances of developing active disease. It should be considered in all children found at immigrant or contact screening who have a positive (grade II–IV) Heaf test. New-born babies of mothers with sputum-positive pulmonary tuberculosis should also be given chemoprophylaxis. For individuals not receiving drug treatment, a 2-year follow-up with periodic radiographic examination is recommended. More detailed guidelines relating to the control of prevention of tuberculosis in Britain have been issued by the British Thoracic Society.

BCG vaccination

This is given to uninfected, tuberculin-negative individuals to increase host resistance. It offers partial protection only and, although an active vaccination policy has been used in the United Kingdom, it has not been adopted in North America and some European countries. The effectiveness of BCG vaccination varies in different countries, with some surveys showing virtually no protective effect. The vaccination of schoolchildren between the ages of 10 and 14–years has been standard practice in most parts of the United Kingdom since the early 1950s. Whether this policy should be abandoned in favour of more selective vaccination of specific high-risk groups is the subject of considerable debate.

Further reading

Empyema

Davies RJO, Traill ZC, Gleeson FV. Randomised controlled trial of intra-pleural streptokinase in community acquired pleural infection. *Thorax* 1997; **52**: 416–21. [Small randomized controlled study indicating potential benefits of intrapleural streptokinase.]

Delikari PG, Conlan AA, Abramer E, Hurnitz SS, Studi R. Empyema thoracis—a prospective study on 73 patients. *South African Medical Journal* 1984; **65**: 47–9. [Description of clinical features,

pathogenesis and outcome in patients with empyema thoracis.]

Ferguson AD, Prescott RJ, Selkon JB, Watson D, Swinburn CR. The clinical course and management of thoracic empyema. *Quarterly Journal of Medicine* 1996; **89**: 285–9. [Prospective multicentre audit of clinical features, treatment and outcome in 119 patients with thoracic empyemas.]

Hamm H, Light RW. Parapneumonic effusion and empyema. *European Respiratory Journal* 1997; **10**: 1150–6. [Review of diagnostic and therapeutic options with guidelines for treating the various stages of parapneumonic effusions and empyemas.]

Landreneau RJ, Keenan RJ, Hazelrigg SR, Mack MJ, Naunheim KS. Thoracoscopy for empyema and haemothorax. *Chest* 1996; **109**: 18–24. [Description of clinical experience with the VATS in management of empyema.]

Wait MA, Sharma S, Hohn J, Dal Nogare A. A randomised trial of empyema therapy. *Chest* 1997; **111**: 1548–51. [Small comparative study of pleural drainage of fibrinolysis versus video-assisted thoracoscopic surgery.]

Wells FC. Empyema thoracis: What is the role of surgery? *Respiratory Medicine* 1990; **84**: 97–9. [Brief review of indications for surgery and operative procedures in the management of empyema.]

Bronchiectasis

Borrie J, Lichter I. Surgical treatment of bronchiectasis: ten year survey. *British Medical Journal* 1965; **2**: 908–12. [Historical review of surgical management and clinical outcome following resection for bronchiectasis.]

Cole P. A new look at the pathogenesis and management of persistent bronchial sepsis: a vicious circle hypothesis and its logical connotations. In: Davies RJ, ed. *Strategies for the management of chronic bronchial sepsis*, pp. 1–16. Medicine Publishing Foundation, Oxford, 1984. [Review of mechanisms of lung damage in patients with bronchiectasis.]

Ellis DA, Thornley PE, Whiteman AJ, Walker M, Chalmers J, Crofton JW. Present outlook in bronchiectasis: clinical and social study and review of factors influencing prognosis. *Thorax* 1981; **36**: 659–64. [Description of clinical features and natural history in 116 patients with bronchiectasis.]

Hill SL, Burnett D, Hewitson KA, Stockley RA. The response of patients with purulent bronchiectasis to antibiotics for four months. *Quarterly Journal of Medicine* 1988; **66**; 163–73. [Clinical and pathophysiological study of prolonged antibiotic treatment in 10 patients with bronchiectasis.]

Stockley RA. Bronchiectasis—a management problem? *British Journal of Diseases of the Chest* 1988; **82**; 209–9. [Review of treatment options in patients with bronchiectasis.]

Whitwell F. A study of the pathology and pathogenesis of bronchiectasis. *Thorax* 1952; **7**: 213–39. [A classical description of morbid anatomy and histopathology.]

Tuberculosis

American Thoracic Society. Rapid diagnostic tests for tuberculosis. *American Journal of Respiratory and Critical Care Medicine* 1997; **155**: 1804–14. [Workshop review of direct amplification techniques used in the diagnosis of pulmonary tuberculosis.]

Crofton J, Horn N, Miller F. *Clinical tuberculosis*. Macmillan, London, 1992. [Comprehensive coverage of all aspects of tuberculosis.]

Joint Tuberculosis Committee of the British Thoracic Society. Control and prevention of tuberculosis in the United Kingdom: code of practice 1994. *Thorax* 1994; **49**: 1193–200. [Guidelines with respect to notification, contact tracing and protection for at risk groups.]

Joint Tuberculosis Committee of the British Thoracic Society. Chemotherapy and management of tuberculosis in the United Kingdom: recommendations 1998. *Thorax* 1998; **53**: 536–48. [Current recommendations for treatment of tuberculosis, chemoprophylaxis, and management of drug resistant disease.]

Pomerantz M, Brown JM. Surgery in the treatment of multi-drug-resistant tuberculosis. In: Iseman MD, Huitt GA, ed. *Clinics in chest medicine—tuberculosis*, pp. 123–30. Saunders, Philadelphia, 1997. [Review of authors' experience in the surgical management of drug-resistant tuberculosis.]

41.11 Principles and practice of thoracic drainage

Stephen Westaby

Physiology of the pleural cavity
Pathophysiology of pneumothorax and intrapleural fluid accumulation
The pathology of thoracic drainage
Simple pneumothorax
Pleural effusions
Practical aspects of thoracic drainage
Practical aspects of chest drain insertion
Technique for insertion of the true apical chest drain by the posterior intercostal approach
Care of the chest drain *in situ*
When to remove the chest drain
Drain removal
Complications of chest drainage and how to avoid them
Failure to enter the pleural cavity
Penetration of the lung
Penetration of the peritoneal cavity
Major haemorrhage
Insertion of the drain into a bulla
Blocked drains
Pleural sepsis
Unexpected results from drainage

Pleural drainage by intercostal tube insertion is a simple, safe, and effective therapeutic manoeuvre which is well within the scope of any physician or emergency medical technician. However, no other simple technique generates more anxiety, technical failure, or iatrogenic morbidity. Chest drain insertion requires detailed knowledge of the relationships between surface landmarks and intrathoracic and abdominal viscera as described in an earlier section (see Chapter 41.1). As for other practical procedures, an unsatisfactory outcome usually follows inadequate tuition or misunderstanding of the principles of thoracic drainage.

Physiology of the pleural cavity

The surface of the pleura consists of mesothelial cells lying on a basement membrane, beneath which is a connective tissue layer containing elastic and collagen fibres. The pressure in the pleural space is lower than atmospheric pressure because of the elastic recoil forces of lung tissue. Despite the negative pressure, the potential volume of the pleural space is kept to a minimum by constant reabsorption. In a normal, healthy adult the quantity of pleural fluid is about 2 ml; this contains electrolytes and proteins. The protein content is 1.5per cent providing a colloid osmotic pressure of about 0.8 kPa (8 cmH_2O). The colloid osmotic pressure gradient between blood and pleural fluid is therefore 2.6 kPa (26 cmH_2O). This causes fluid flux from the pleural space into the adjacent blood plasma of both visceral and parietal pleura (Fig. 1).

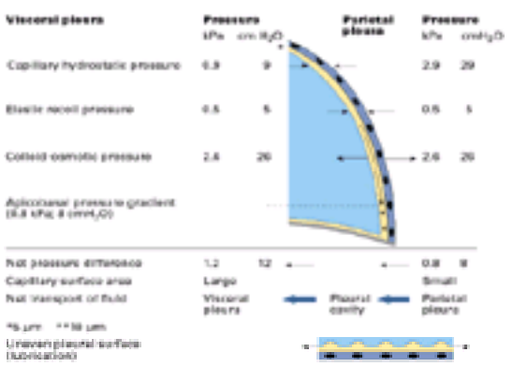


Fig. 1. Transport of fluid in the pleura.

Gravity and regional differences in lung deformation cause an apicobasal pressure gradient in the pleural cavity of 0.8 kPa (8cmH_2O). The hydrostatic pressure in the capillaries of the lung amounts to about 0.9 kPa (9 cmH_2O); and this is considerably lower than the hydrostatic pressure in the capillaries of the systemic circulation of the parietal pleura, which amounts to some 2.9 kPa (about 28 cmH_2O). The hydrostatic pressure in the pleural space is, therefore, lower than atmospheric pressure (due to the retraction forces of the lung) and is lower than the hydrostatic pressure in both pulmonary and systemic capillary systems. The average net pressure gradient resulting from these factors causes transport of fluid from the parietal pleura to the pleural cavity and then to the visceral pleura and the lung parenchyma. The reserves of fluid transport are large since the contact surface formed by the capillaries of the visceral pleura is greater than that in the parietal pleura. Extra fluid in the pleural cavity is absorbed, the rate of transport being determined by the hydrostatic pressure gradient, the colloid osmotic pressure gradient, the elastic recoil pressure of the lung, and tissue permeability. Absorption of pleural fluid continues until only the minimum quantity required for lubrication of the pleural sheets (about 2 ml) remains. In practice, the thickness of the fluid layer is about 5 μm at the apical area and 10 μm in the basal area because of gravity.

In practical terms, the pleura is a closed collapsible space which can be filled by air or fluid. The pleural space is gas-free because the gas pressure in the tissues is subatmospheric (about 705 mmHg, 94 kPa), roughly the same as that in venous blood. The gas pressure in oxygenated blood differs little from barometric pressure (760mmHg, 101 kPa). The difference results from cell metabolism and is associated with transport properties of the blood for oxygen and carbon dioxide.

During quiet breathing the pulmonary recoil pressure is about 5 cmH_2O; if air is introduced into the pleural cavity the lung shrinks by an equivalent volume. However, pleural tissue is easily penetrated by most gases and reabsorption of gas from a pneumothorax occurs progressively. The partial pressures of oxygen and carbon dioxide initially rapidly assume the values of the surrounding pleural tissue. However, the total gas pressure in the pleural cavity must remain equal to the barometric pressure minus the pulmonary recoil pressure. This is achieved by an increase in the partial pressure of nitrogen, which produces a nitrogen gradient. Consequently, nitrogen diffuses out of the pleural cavity into the pleural tissue, resulting in an increased carbon dioxide and oxygen tension. These gases then also diffuse into the tissues. Excretion of carbon dioxide is rapid, whereas oxygen transport depends on the O_2/N_2 ratio. While breathing room air the partial pressure of oxygen is comparatively low and a steady state for O_2 is achieved rapidly (Fig. 2).

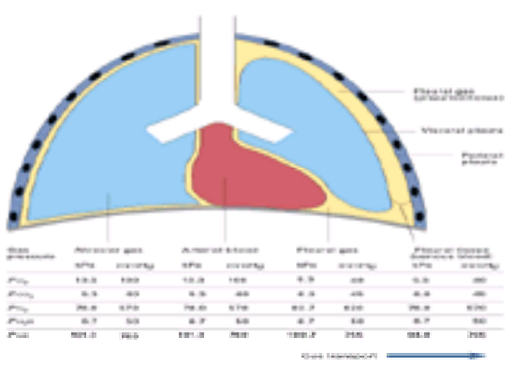


Fig. 2. Gas transport in the pleura.

Pathophysiology of pneumothorax and intrapleural fluid accumulation

Most injuries or diseases which cause accumulation of air or fluid in the pleural cavity impair ventilation. Different pathological states may cause the pleural cavity to fill with air, blood, pleural exudate or transudate, chyle, pus, bile, or gastric contents. Loss of lung volume occurs first by elastic recoil, and then by compression of the lung. Large volumes of intrapleural gas or fluid cause mediastinal shift with volume loss in the contralateral lung. Ventilation of both the collapsed and partially collapsed lung is then seriously compromised. The mechanics of breathing are disturbed and ventilation perfusion imbalance follows (Fig. 3).

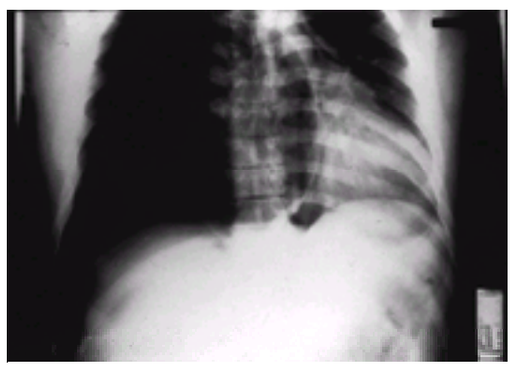


Fig. 3. Tension pneumothorax with impaired ventilation of both the collapsed and potentially collapsed contralateral lung.

Oxygenation of the blood and elimination of carbon dioxide depends on the balance between ventilation of the lung and its blood supply. Distribution of inspired air is influenced by regional variations in airway resistance and compliance. The latter is governed by gravity-dependent intrapleural pressure gradients. On this account, gas distribution in the lung is uneven, even during normal resting tidal ventilation. Normally, in the lower or more dependent pleural space the pressure is closest to atmospheric (least negative). Intrapleural pressure becomes increasingly negative towards the apex or non-dependent region of the lung. When a normal inspiration is taken from end expiration (functional residual capacity), expansion of the initially smaller dependent alveoli is governed by the steep portion of the compliance curve. Consequently, they expand more for each unit of pressure change than do those at the apices which are influenced by the upper, flatter portion of the curve. These differences result in preferential distribution of inspired gases to the areas of greater expansion in the dependent portions of the lungs. However, if terminal air spaces have collapsed through pneumothorax or pleural effusion, inspiration results in preferential distribution to already expanded areas because of the influence of the compliance curve. The maldistribution of ventilation is worsened by airflow obstruction in terminal airways from external compression, and elevated intrapleural pressure. When a whole lung or lobe collapses there is perfusion of non-ventilated lung, and a serious veno-arterial shunt effect. Ineffective oxygenation of the venous blood is reflected by widening of the alveolar arterial oxygen tension difference and systemic hypoxia.

Movement of blood through the lungs is also influenced by gravity, and the pressure gradient between the pulmonary arteries and left atrium. Blood flow is normally directed preferentially to the dependent parts of normal lung where ventilation is most efficient. Collapse of these areas of lung causes ventilation–perfusion imbalance and hypoxia. The importance of early evacuation of intrapleural air, blood, or effusion, together with re-expansion of the atelectatic lung is readily apparent.

The pathology of thoracic drainage

Simple pneumothorax (Fig. 4)

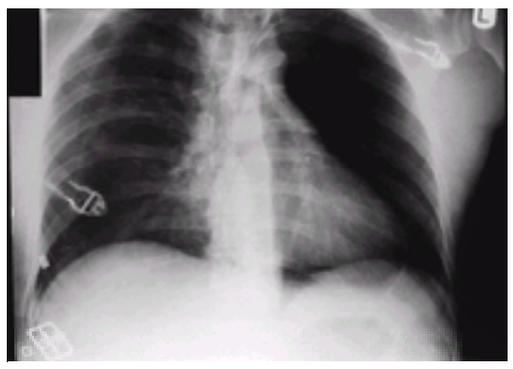


Fig. 4. Simple pneumothorax (right side) together with pneumopericardium after closed chest trauma.

Pneumothorax is defined as the presence of air between the visceral and parietal pleurae (see Chapter 41.8). It may be induced by trauma or thoracic surgery or occur spontaneously. Simple pneumothorax occurs in association with a great number of underlying lung disease such as asthma, chronic obstructive airways disease, and bullous emphysema. Infective processes, particularly tuberculosis, pyogenic lung abscesses, and fungal disease also cause simple pneumothorax. Pneumothorax occurs in children with cystic fibrosis and in babies with hyaline membrane disease. The most common cause of spontaneous pneumothorax in young adults (16–26 years) is a ruptured congenital bulla, commonly apical in position. In elderly patients bullae develop as a consequence of emphysema and chronic obstructive airways disease (Table 1). Mechanical ventilation increases the risk of pneumothorax, which occurs commonly in patients requiring long-term support for adult respiratory distress syndrome. Pneumothorax is common in patients with chest injuries, and after needle aspiration of a pleural effusion. Air may enter through the needle, or laceration of the visceral pleura may cause an air leak from the surface of the lung. Cannulation of the subclavian vein and external cardiac compression for cardiopulmonary resuscitation may also cause pneumothorax.

1. Idiopathic (majority)
2. Connective tissue disorder:
e.g. Marfan's, Ehlers-Danlos syndromes
3. Air space disorder:
e.g. bullae, emphysema, cysts
4. Airways obstruction
(important differential diagnosis in exacerbations of asthma and chronic bronchitis)
5. Generalized lung disease:
e.g. pulmonary fibrosis, pneumocystosis, eosinophilic granuloma
6. Localized lung disease:
e.g. tuberculosis, lung abscess, tumour, hydatid disease
7. Secondary to mediastinal emphysema:
e.g. decompression of divers

Table 1 Causes of spontaneous pneumothorax

Introduction of air into the potential pleural space allows the lung to collapse by elastic recoil. The effect of the pneumothorax depends on the volume of air. No serious disturbance results unless continuous leakage of air causes intrathoracic pressure to rise above atmospheric pressure. Severe reduction in lung volume or a rise in intrathoracic pressure above atmospheric pressure causes occlusion of intrathoracic airways during expiration and air trapping. Continued ventilation then requires forceful contraction of the inspiratory muscles to overcome the elastic forces of the distended thoracic cage. Tension pneumothorax occurs where an air leak from the lung surface causes progressive rise in intrapleural pressure because of a ball valve mechanism which permits air to enter but not leave the pleural space. This commonly occurs after penetration of the lung parenchyma by a fractured rib or penetrating instrument. Compression of the lung causes atelectasis and reduced pulmonary blood flow. Eventually, mediastinal shift distorts the venae cavae, reducing venous return, stroke volume, and cardiac output. Clinically the patient complains of dyspnoea and chest pain (Fig. 3).

On examination, the patient is tachypnoeic, agitated, and has hyper-resonance and decreased breath sounds on the affected side. There is decreased motion of the chest wall and mediastinal shift to the opposite side, as shown by the deviated trachea. In severe cases the neck veins are distended and the patient is peripherally cold and clammy.

The diagnosis of pneumothorax is made clinically. Unless the patient is severely compromised by pneumothorax the diagnosis should be confirmed by plain chest radiographs and the need for intervention should be determined from the findings. Patients with tension pneumothorax and respiratory or haemodynamic distress must be treated before confirmation by chest radiography. Uncertainty about pneumothorax in a patient presenting with acute respiratory distress can be resolved by needle aspiration of the pleural cavity and relieved by prompt scalpel incision through the chest wall allowing air under pressure to escape through the track. Having relieved the acute event, the chest drain with underwater seal drainage is inserted secondarily.

Therapeutic needle aspiration of pneumothorax alone is usually inadvisable since the lung may be lacerated by the needle point. When a pneumothorax is large enough to warrant evacuation this is best performed by chest drain insertion, underwater seal drainage, and mechanical suction.

The extent of a pneumothorax is determined on the chest radiograph by measuring the distance from the lung border to the inside of the thoracic wall at the level of the anterior bony end of the third rib. Pneumothoraces more than 1.5 cm wide in an adult are defined as significant and drained. Smaller pneumothoraxes are drained if bilateral as a safety measure against simultaneous collapse of both lungs, or if the patient is to undergo intermittent positive pressure ventilation for any reason. This prevents the development of iatrogenic tension pneumothorax. In patients with chronic obstructive airways disease, restrictive or destructive lung disease, or a spinal cord injury, smaller pneumothoraces should be drained.

Pleural effusions

Pleural effusions may complicate obvious pre-existing inflammatory or malignant disorders, or may arise without evidence of an underlying lung or pleural lesion. A comprehensive list of the causes of pleural effusion is presented in Table 2. The fluid may freely fill the pleural space or be loculated between the lobes, against the chest wall or mediastinum, or above the diaphragm. A large volume of fluid may displace the mediastinum. Signs of pleural effusion include diminished chest movement on the affected side, stony dullness to percussion, diminution of breath sounds, and decreased vocal resonance. If the underlying lung is consolidated, bronchial breathing may be heard at the upper level of dullness.

Transudate	Cardiac failure
	Constrictive pericarditis
	Hypoalbuminaemia (due to nephrotic syndrome, liver failure)
Exudate	Infective tuberculosis or bacterial empyema
	Inflammatory
	Secondary to underlying pneumonia
	Subphrenic abscess
	Pancreatitis
	Collagen disorders
	SLE
	Rheumatoid arthritis
	Meig's syndrome (associated with ovarian fibroma)
	Infection
	Pulmonary embolus
	Malignant
	Primary mesothelioma
	Secondary bronchial, breast

Table 2 The causes of pleural effusion

Pleural fluid causes a dense homogeneous opacity on chest radiographs, often seen as a triangular shadow based on the diaphragm with the apex curving up towards the axilla (Fig. 5). The upper level of the fluid is only seen as horizontal when air is introduced into the pleural space (hydropneumothorax) (Fig. 6). The chest radiograph cannot differentiate between serous fluid, pus, chyle, or blood: this is only determined by needle aspiration or intercostal drainage. Even then differentiation may require cytological or biochemical analysis. It used to be the practice to differentiate between exudates and transudates by estimation of protein content, though this information is rarely of practical importance. Exudates have a protein content greater than 30 g/l, whereas transudates have a protein level less than 30 g/l. Transudates occur in cardiac failure, where the effusion tends to occur more often on the right. Other transudates occur with hypoproteinaemia from liver disease or from constrictive pericarditis and idiopathic mediastinal fibrosis.

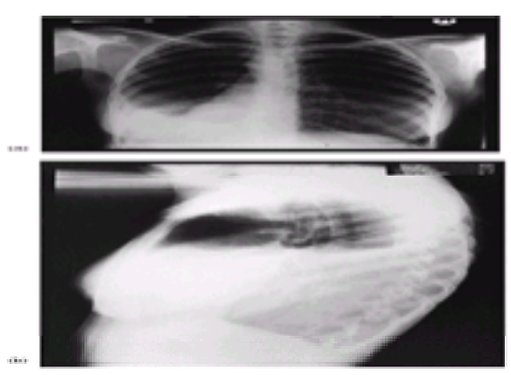


Fig. 5. Left-sided pleural effusion. (a) Posteroanterior view. The patient has carcinoma of the breast. (b) Lateral view. The patient has carcinoma of the breast.

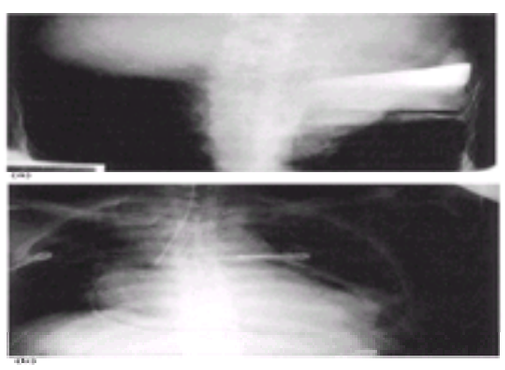


Fig. 6. (a) Left sided empyema with fluid level after initial aspiration and instillation of dye to outline the cavity before rib resection and surgical drainage. (b) Gas in the pericardium after perforation by carcinoma of the oesophagus.

Infective pleural effusions

Postpneumonic pleural effusions tend to remain serous if the patient has received antibiotics or to form an empyema if not. When the fluid is serous, the initial cell content is predominantly polymorphic; lymphocytes are present 7 days later. Most pneumonic effusions follow bacterial pneumonia and at an early stage respond to simple needle aspiration. At a later stage, chest drain insertion or eventually surgical decortication may be required. Aggressive early treatment reduces the likely need for surgery. Tuberculous effusions are uncommon but may complicate miliary tuberculosis or paravertebral abscess. Active pulmonary tuberculosis sometimes involves the pleura, causing a tuberculous empyema. The pleural fluid is usually straw-coloured and occasionally bloodstained. The predominant cell is the lymphocyte, and tubercle bacilli are cultured though rarely seen on microscopy.

In tuberculosis, the value of aspiration is uncertain unless the effusion is causing respiratory distress. Tuberculous empyema can be managed by aspiration or intercostal drainage followed by instillation of streptomycin. Surgical decortication is often required, particularly after rupture of a tuberculous cavity where the hydropyothorax contains both tubercle bacilli and pyogenic organisms.

Malignant pleural effusions

Bronchial carcinoma in men and carcinoma of the breast in women are the most common causes of malignant pleural effusion (Fig. 7). Other primary tumours, particularly ovarian, also metastasize to the pleura. Primary pleural tumours may be diffuse, as in malignant mesothelioma commonly with a history of asbestos exposure (Fig. 8), or localized, as in fibroma of the pleura which almost invariably produces the syndrome of hypertrophic pulmonary osteoarthropathy. Reticuloses, such as Hodgkin's disease and lymphosarcoma, may present as pleural effusion but almost always involve other organs by this stage.



Fig. 7. Left-side bronchogenic carcinoma, with pleural effusion.



Fig. 8. Malignant mesothelioma. Thoracotomy was performed for biopsy.

Malignant tumours cause accumulation of fluid in the pleura in different ways. Visceral pleural deposits may stimulate fluid production and additionally block lymphatics and compromise capillary reabsorption (Fig. 9). Parietal pleural deposits may also damage lymphatic integrity and increase capillary permeability. Impaired reabsorption of pleural fluid may follow a rise in systemic pressure due to superior vena caval obstruction or cardiac tamponade secondary to pericardial tumour deposits (Fig. 10 (a, b)). In patients with malignant ascites, overspill of protein-laden lymph through transdiaphragmatic lymphatic anastomoses causes protein-rich pleural effusion.



Fig. 9. Lymphangitis carcinomatosa with left pleural effusion.

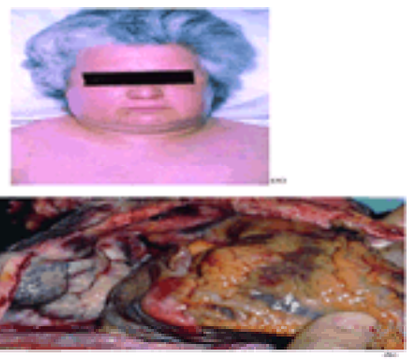


Fig. 10. (a) Swollen head and neck in superior vena caval obstruction. (b) Extensive pericardial tumour deposits from thyroid carcinoma.

Clinical presentation is usually with shortness of breath associated with a non-productive cough, or pleuritic chest pain. The diagnosis is made radiologically with a plain chest radiograph and then CT scan for detailed assessment of the pleural surfaces. The nature of the effusion can be confirmed by needle aspiration and cytological examination. Definitive treatment depends on the nature of the tumour. Symptomatic improvement is achieved rapidly by needle aspiration or intercostal drainage. Effusions from metastatic breast carcinoma may resolve after appropriate hormonal or chemotherapeutic treatment. Intrapleural anti-tumour agents have proved disappointing for most lesions. Patients with potential life expectancy of longer than 1 to 2 months should be considered for talc pleurodesis or pleurectomy.

To achieve adhesion of the visceral to the parietal pleura with permanent obliteration of the pleural space various conditions must be met. First, pleural fluid must be completely drained, preferably by insertion of an intercostal drain into the most dependent part of the pleural space. This is followed by continuous application of negative intrapleural pressure by mechanical suction. It is imperative to maintain apposition of the parietal and visceral pleural surfaces continuously during the process to pleurodesis, while fibrinous adhesions bond the two. Intercostal drains should therefore remain in place for at least 4 days whether the patient has undergone pleural sclerosis with blood, tetracycline, a cytotoxic agent, or insufflated talc. Drainage for more than 5 days carries the risk of infection through the drainage track and empyema formation.

Clinical trials have demonstrated talc insufflation to be the most effective of the conservative methods of pleurodesis. This procedure is performed under general anaesthesia with the patient in a lateral position. Pleural fluid is evacuated through a trocar and cannula, and the pleural cavity is inspected with the thoracoscope. This allows biopsies to be taken. Approximately 10 g of talc can then be insufflated by moving the top of the insufflator around the pleural cavity, ensuring uniform distribution. At the end of the procedure, one or two intercostal drains are inserted and attached to an underwater seal. The drains are placed on suction with a high pressure/volume pump (Thompson or Tubbs–Barrett) (Fig. 11). A negative pressure of 15 to 20 cmH_2O applied continuously to the pleural cavity promotes pleurodesis and prevents accumulation of further malignant or inflammatory exudate.

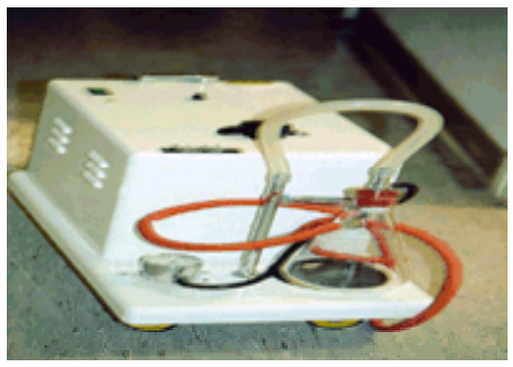


Fig. 11. Tubbs–Barrett high volume suction pump.

For patients with persistent chylothorax, complete parietal pleurectomy is preferable through a small lateral thoracotomy. This can be reinforced by talc application. Patients with long-standing large pleural effusions are occasionally referred late for pleurodesis, when the lung is already encased in fibrin and will no longer expand. This should be recognized during evacuation of the fluid at thoracoscopy since suction will draw the mediastinum across towards the space evacuated by the fluid.

Pleural effusions also occur in patients with collagen disorders such as acute rheumatic fever, rheumatoid arthritis, systemic lupus erythematosus, and polyarteritis nodosa. These seldom require more than simple needle aspiration. Pleural effusions may follow subphrenic disease, including subphrenic abscess where a reactionary sterile, straw-coloured effusion is common. Alternatively, a pyogenic or amoebic hepatic abscess may burst through the diaphragm into the right pleural cavity. Acute pancreatitis may be complicated by a left-sided serous effusion which may be haemorrhagic. This has a high amylase content and is only eradicated by effective treatment of the pancreatitis. Amylase content of the effusion is much higher than that of the serum, and fat necrosis occurs in the pericardial fat. Small bilateral pleural effusions are common in association with ascites, constrictive pericarditis, or cardiac failure.

Chylothorax (see Chapter 41.9)

Chylous pleural effusions may result from abnormalities of the lymphatics. Chylothorax is an unusual condition, classically caused by rupture of the thoracic duct. Occasionally this follows non-penetrating injury where the thoracic duct is stretched over the vertebral bodies by jumping, vomiting, or coughing. There is often a latent interval between injury and chyle filling the pleural cavity. Chyle initially accumulates beneath the mediastinal pleura, then ruptures into the pleural space, causing the patient to become short of breath. Chylous effusion may also occur from rupture of distended lymphatic tributaries which cause multiple small fistulae that weep chyle (Fig. 12) This difficult problem can be treated successfully with obliteration of the pleural cavity by pleurectomy together with ligation of the thoracic duct at the diaphragm. Injury of the thoracic duct during thoracotomy, particularly for palliation of congenital heart disease, may cause troublesome persistent lymph leak (Fig. 13). Initial management is by dietary fat restriction, though re-exploration and ligation of the thoracic duct may be required.

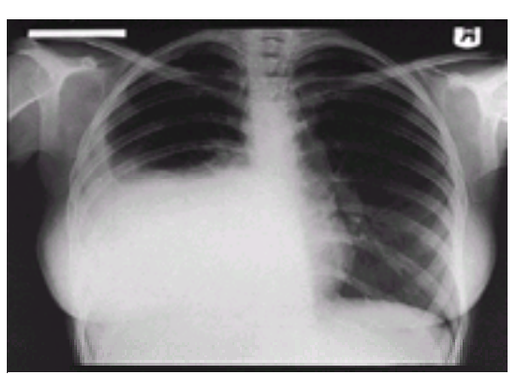


Fig. 12. Left-sided chylothorax in a female patient with metastatic breast carcinoma occluding the thoracic duct.



Fig. 13. Right-sided chylothorax in a child following surgical injury to the thoracic duct.

Empyema (see Chapter 41.10)

There are many causes of empyema including lung abscess, pneumonia, and tuberculosis. Empyema may also follow inadequately treated traumatic haemothorax, bronchopleural fistula formation, lung resection, and subphrenic abscess. Aggressive, early treatment by intercostal drainage reduces the need for surgical decortication. In practice, reliance on needle aspiration which fails to clear thick pus and debris often sentences the patient to prolonged antibiotic treatment followed by surgery. If necessary, children should be given a general anaesthetic for intercostal drain insertion. Suction on the drain helps evacuate the pus and obliterate the pleural space.

Traumatic haemothorax

This subject is covered comprehensively in Chapter 41.12.

In patients who present moribund with tension pneumothorax or an extensive haemothorax, a chest drain should be inserted rapidly on the basis of physical findings. Torrential haemorrhage through the drain may demonstrate the need for immediate accident department thoracotomy as a life-saving manoeuvre. In the stable patient, a plain chest radiograph is the most important investigation and should be taken before chest drain insertion (Fig. 14).

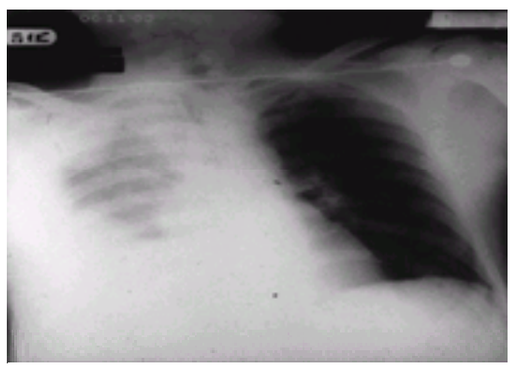


Fig. 14. Right-sided traumatic haemothorax.

Fifteen per cent of patients with fractured ribs, pulmonary contusion, and surgical emphysema in the chest wall will have no pneumothorax. This is because the pleural cavity has been obliterated by previous pneumonic adhesions. In this circumstance insertion of an intercostal drain is extremely dangerous (Fig. 15).



Fig. 15. CT scan showing direct penetration of the right lung by an ill advised intercostal drain. Old pneumonic adhesions caused the lung to adhere to the chest wall. The CT scan shows tell-tale loss of volume and a posteriorly situated pleural plaque on the right side.

Evacuation of a haemopneumothorax is carried out with a carefully-sited, large-bore chest drain connected to an underwater seal and placed immediately on high volume suction at a negative pressure of 20 cmH_2O. Indications for thoracotomy based on the results of drainage are discussed in Chapter 41.12. Thoracic drainage aims to relieve hypoxia and acidosis secondary to hypoventilation and atelectasis. When inserting the drain care must be taken not to confuse the dilated stomach or gut for pneumothorax when the diaphragm is ruptured (Fig. 16). On the right side, herniation of the liver through a lacerated diaphragm may mimic haemothorax.



Fig. 16. Ruptured left hemidiaphragm with stomach in the left chest.

When the chest radiograph is taken after chest drain insertion further decisions must be made. First, if drainage of blood is inadequate a second large drain may be required. If the lung fails to expand despite technically satisfactory drainage, bronchoscopy may be required to clear the airways of blood, secretions, or teeth in order to promote re-expansion.

With the pleural cavities cleared of blood and air, a better impression of the extent of injuries is obtained. Defects such as ruptured bronchus, lacerated diaphragm, or transected thoracic aorta with a wide mediastinum may then become apparent. Contralateral bleeding or pneumothorax may require chest drain insertion on the opposite side. Failure to evacuate a clotted haemothorax should be followed by thoracotomy. Failure to do so is followed by resolution of the haemothorax with development of a fibrothorax which will progressively reduce the size of the lung and the compliance of the chest wall. Inadequate resolution of haemothorax after early discharge from hospital may require decortication if the vital capacity diminishes to 75 per cent of predicted. This is usually undertaken 5 to 6 weeks after the accident, when organized fibrous tissue can be peeled off the underlying lung. Infection in an inadequately drained haemothorax produces an empyema which again requires thoracotomy. Fibrinolytic agents have been recommended by some to help liquefy haemothorax or empyema. These are seldom effective and are not recommended.

Practical aspects of thoracic drainage

The intercostal drain is simply a conduit used to remove air, blood, pus, or fluid from the pleural cavity, thereby preventing compression of the lung and allowing it to re-expand. Besides draining air and fluid out of the pleural space the drainage system must also prevent their return, encouraged by negative intrapleural pressure created by breathing in. For this dual purpose three basic components are required: an unobstructed chest drain of adequate diameter, a collecting container below

chest level, and a one-way mechanism—a water seal or valve to prevent the return of air or fluid.

The patient may have one or two drains in place depending on the extent of pathology. Because air rises and liquid sinks, the tip of a tube to drain air is traditionally placed in the apex of the pleural space and that for liquid at the base posterolaterally in the costophrenic angle. Although these relative positions are used in thoracic surgery where drains are placed accurately with the chest open, percutaneous drain insertion usually results in a single tube sited towards the apex posteriorly (Fig. 17). In a free pleural cavity blood or air will find the tube, particularly when negative pressure is applied.

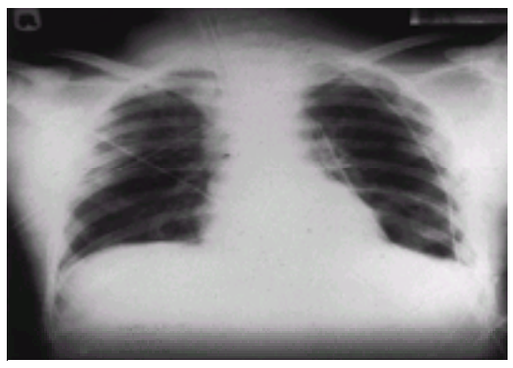


Fig. 17. Plain chest radiograph after left thoracotomy to repair a ruptured diaphragm. The patient has surgically placed apical and basal drains on the left and a percutaneously situated drain on the right.

The chest drain is joined to about 1.8 m of connecting tubing which leads to a bottle or container placed several feet below the patient's chest level. This tubing allows the patient to turn and move and minimizes the chance that a deep breath could suck liquid back into the chest. The position of the bottle and underwater seal takes advantage of gravity by establishing the patient's pleural space as the area of higher pressure and the collecting bottle as the area of lower pressure. The non-patient end of the tubing is led below the surface of sterile water which serves to create a ‘water seal' and to establish a closed system by sealing off the open end of the tubing from the atmosphere (Fig. 18).



Fig. 18. Complete collapse of the left lung after rupture and dehiscence of the left main bronchus. Pneumothorax under tension but decompressed into the bronchial tree.

The positive pressure in the chest during exhalation can push air and liquid out of the pleural cavity. The air bubbles out through the water and out of the vent tube of the drainage bottle while liquid mixes with the water, raising its level. The length of tube drain below the water surface should be short so that the resistance to air escape is no more than 2 to 3 cmH_2O pressure. If the drainage bottle was full of liquid, such as blood from chest trauma, the resistance to air passage through the underwater part of the tube (say 15–20 cmH_2O) could prevent air escape and produce a pneumothorax even in the presence of a ‘functioning chest drain'. Note that any time the tube is clamped no air or liquid can escape from the pleural space and this puts the patient with a pleural air leak in real danger of developing a tension pneumothorax.

A pneumothorax can take two forms—open and closed. If a chest drain becomes disconnected, opening the pleural cavity to the atmosphere, the lung collapses away from the chest wall by elastic recoil. A relatively small amount of air (about one-fifth of the lung volume) enters the pleural cavity and impairs ventilation since the lung no longer expands sufficiently. As long as the hole in the chest wall is smaller than the trachea an open pneumothorax is well tolerated since the other lung continues to function normally. Reconnection of the chest tube with a bout of coughing causes complete re-expansion of the lung. In contrast, if the chest drain is clamped in a patient with an air leak from the surface of the lung, air enters the pleural space, but cannot get out. Each time the patient breathes in more air escapes, builds up in the pleural cavity, and compresses the lung. In tension pneumothorax the pressure may build up rapidly and the lung collapses totally (Fig. 18). The mediastinum is pushed over to the opposite side and the opposite lung is then also compressed. Compression of the mediastinum by air under pressure also interferes with venous return to the heart, and reduces cardiac output. Eventually impairment of ventilation and reduced cardiac output can cause death.

Traditionally, chest drain clamping has been used on the slight pretext that the drainage bottle may become disconnected or upset and broken. To this end two haemostats with rubber tubing over their tip are commonly kept within easy reach. In practice, complete disconnection of the drain with entry of a small volume of air is a completely trivial and easily reversible event. It is therefore reasonable never to clamp chest tubes in patients with pneumothorax, either when the patient gets out of bed, walks, or is taken to some other part of the hospital, or indeed, is transferred from one hospital to another. The drainage bottle should be handled carefully as one would any piece of equipment attached to a patient. It should be kept below chest level, making sure the water seal remains intact. If the drainage tubing comes adrift from the bottle, the end must be cleaned, reconnected and the patient should be asked calmly to cough a few times. This will rid the pleural cavity of the pneumothorax.

For a patient who has never had an air leak, clamping for a short time when changing the drainage bottles does no harm. It is just so much better to be in the ‘no-clamping' than the ‘clamping' habit. Drains to the pericardium or mediastinum after cardiac surgery do not usually communicate with the pleural cavity and need not drain to an underwater seal. In practice, however, the underwater seal is usually employed for all thoracic drains and for patients with thoracoabdominal incisions after oesophagectomy.

In all except emergency circumstances, it is important to obtain a posteroanterior and lateral chest radiograph to determine the size, extent, and morphology of the pneumothorax or effusion. Parts of the pleural space may be obliterated, so that precise placement of the drain is required. A CT scan is then useful to define areas of loculation of pus or blood. It is imperative to make certain that either air or fluid is interposed between the chest wall and lung. In about 15 per cent of patients with postpneumonic fibrous intrapleural adhesions, insertion of a trocar through the chest wall will result in penetration of the lung and intrapleural haemorrhage.

Practical aspects of chest drain insertion

Consult the plain chest radiograph unless the drain is to be inserted in an absolute emergency. Loculations of fluid or air may determine the precise site of insertion. It is a fallacy that the drain must be in a basal position to drain blood or an apical position to drain air. Blood, fluid, or air will find the drain in a free pleural cavity particularly when suction is applied.

Choose the site of insertion carefully. For most conditions this will be between the fourth and seventh intercostal spaces between the mid-axillary and anterior axillary lines. The level at which the anterior axillary fold meets the chest wall is a useful guide. Insertion of the drain more posteriorly than the mid-axillary line may cause the patient to compress and obstruct the tube whilst lying on his back. There is no contraindication to inserting the drain through an area of injury, but if there is a possibility of a ruptured diaphragm with viscera in the chest the drain should be inserted higher towards the axilla (Fig. 19 (a, b)). The anterior approach through the second interspace is best avoided. This route transfixes the two major accessory respiratory muscles, pectoralis major and minor, and in females the scar may be unsightly. If an apical drain is required because of intrapleural adhesions towards the base then the true apical approach above the scapula and into the first interspace posteriorly

is preferred (Fig. 20).

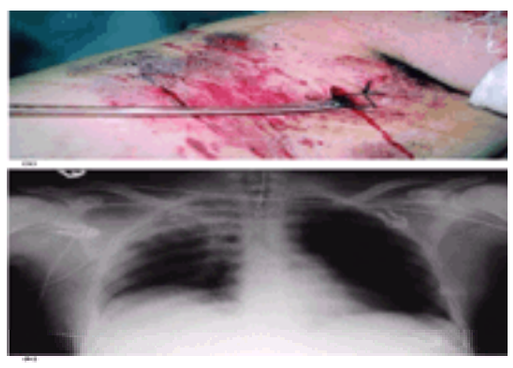


Fig. 19. (a) Chest drain inserted at the level of the axillary fold in a patient with right chest trauma. (b) Plain chest radiograph showing a well-sited drain at the right apex with complete evacuation of the haemopneumothorax.



Fig. 20. True ‘suprascapular' apical chest drain.

Choose a chest drain of appropriate size. For adult patients with traumatic haemopneumothorax the drain should always be 28 French gauge or larger. Smaller drains will rapidly occlude with blood clot. The same applies to an empyema with thick pus or a chylous effusion. In patients with clotted haemothorax or those with gelatinized pus, drainage will fail and thoracotomy is required. Smaller bore drains are suitable for simple pneumothorax although with time fibrin generated by the inflammatory reaction to the drain will cause blockage. For this reason drains of less than 20 French gauge should not be used in adults. In children drain size depends on age and the distance between the ribs.

In a conscious patient, infiltrate 10 to 15 ml of 1 per cent lignocaine through to the periosteum on the upper border of the rib below the chosen interspace. Advancing the needle above the rib should infiltrate the pleura. Passage of the needle into the pleural cavity confirms the presence of free air, fluid, or blood on aspiration. Liberal local anaesthesia facilitates scalpel incision through the full thickness of the chest wall so that the drain can be inserted smoothly without pushing. The needle used for infiltration can be left *in situ* so that the precise skin area and direction of anaesthetic infiltration is not lost when the skin is cleansed with povidone iodine solution.

Aseptic technique is important since a sterile effusion or haemothorax can easily be turned into an empyema when infection is introduced through the drain. A wide area of skin around the proposed site is cleaned with povidone iodine or chlorhexidine in spirit. The operator should scrub as for a surgical procedure and wear a mask and gloves.

For trauma patients it is wise to cover the procedure with a single dose of prophylactic intravenous antibiotic. In an emergency situation many of these precautions will be over-ruled by the degree of urgency. For life-threatening tension pneumothorax a stab wound through the chest wall without preparation provides immediate relief and should not await aseptic technique or the availability of a drain. An antibiotic should be administered after completing the procedure.

Use a scalpel to incise the chest wall skin and subcutaneous fat 1 to 2 cm beneath the proposed site of pleural incision so that the drain track leads the drain towards the apex of the pleural cavity. The underside of the rib above should be avoided so as to preserve the intercostal nerve and vessels. The scalpel should find the rib below the interspace to be breached; the remainder of the track is then completed by blunt dissection through to the pleural cavity with artery forceps. If there is any question about the adequacy of this incision a gloved finger should be passed through the track. The finger tip should enter the pleural cavity easily without resistance and on removing the finger air, blood, fluid, or pus should exit through the drain site. This precaution should always be taken, when a partially obliterated pleural cavity is expected (Fig. 21).

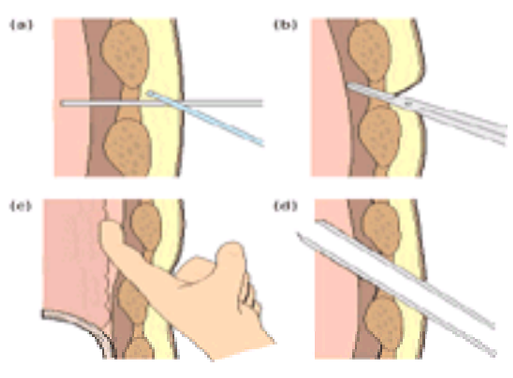


Fig. 21. Chest drain insertion. (a) Needle and local anaesthetic infiltrating the proposed drain track. (b) Blunt dissection of the parietal pleura. (c) Digital examination along the tract into the pleural space. (d) Withdrawal of central trocar and positioning of drain.

The drain should slide easily through the track and into the pleural cavity when blood, fluid, or air will flash fill the tube. When the expected contents of the pleural cavity do not appear in the drain it is important to ascertain why. First, there is the possibility that the drain has not passed through the intercostal space but has deflected into the tissues of the chest wall. Second, the radio-opaque contents of the pleural cavity may be solid instead of liquid, as in clotted haemothorax or empyema. Drains have been inserted into mesothelioma and lung tumours. Do not be concerned about allowing air to enter along the drain: this will be evacuated immediately when the lung expands and underwater sealed drainage is established.

When the drain is in good position it is secured with at least zero gauge silk or nylon suture to prevent displacement. Once inserted a drain should never be allowed to fall out. The suture material must therefore be strong enough to ‘bite into' the plastic tube and grip firmly. A purse string suture is placed around the incision to close the track when the time comes for drain removal.

Attach the intercostal tube to the drainage apparatus and underwater seal. Reusable glass bottles were employed historically but have been replaced by lighter, portable containers with scales permitting accurate measurement of blood or fluid loss (Fig. 22). This is particularly important in trauma and following cardiac and thoracic surgery since blood replacement in these patients is determined by blood loss. Modern drainage devices combine cylinders for autotransfusion of shed pleural or mediastinal blood. A small isolated underwater seal chamber prevents the fluid drained from reaching a depth which presents significant resistance to free drainage.

This can occur in a glass bottle which fills progressively eventually to provide an obstruction of 15 to 20 cm of blood. There is no arbitrary fixed limit to the volume of fluid which may be drained, though rapid evacuation of more than 1 litre of blood or pleural fluid may cause acute mediastinal shift with pain or discomfort. ‘Re-expansion' pulmonary oedema is rare but occurs unpredictably after drainage of large effusions.

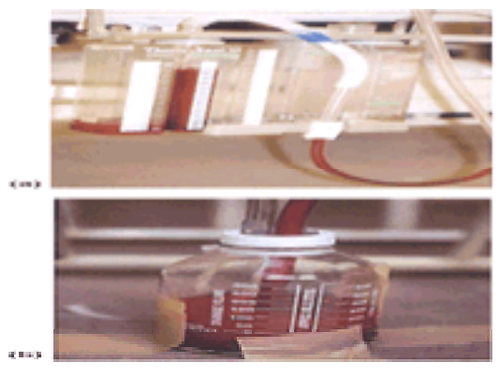


Fig. 22. (a) Modern thoracic underwater seal apparatus with facility for careful measurement and autotransfusion. (b) Simple glass bottle underwater seal.

Apply negative pressure suction to the chest drain to ensure evacuation of fluid or air and to assess the rate of continuing haemorrhage or air leak in trauma patients. A Thompson or Tubbs–Barrett suction machine is necessary to ensure large volume suction at a negative pressure of 15 to 20 cmH_2O. The Roberts suction pump should not be used in patients with pneumothorax and persistent air leak as it will shift only small volumes of air and may constitute an obstruction in the presence of a rapid air leak. If suction causes pain or increased breathlessness it should be decreased or temporarily stopped. Always disconnect the suction system when not in use. Suction must never be applied to a pneumonectomy space or when a completely collapsed lung will not re-expand. This will cause mediastinal shift and haemodynamic deterioration. Suction is used after partial lung resection.

Check the position and effect of the drain by plain chest radiograph. Is the drain in the expected position or are readjustments necessary? Has the drain cleared the pneumothorax or fluid accumulation? Note the initial volume of blood or fluid evacuated and the continued rate of drainage during suction. The need for thoracotomy following trauma is determined by the continued rate of bleeding. Assess the volume of air leak in patients with pneumothorax in whom drainage of air does not cease in the first few minutes.

When these steps are followed thoracic drainage should be safe and effective.

Technique for insertion of the true apical chest drain by the posterior intercostal approach

Loculated apical pneumothoraces occur in patients in whom the pleural space is partially obliterated by inflammatory adhesions, and after thoracotomy and lung resection. In these circumstances the best approach is via the apical posterior intercostal approach. The pleural space to be drained should be clearly defined by plain chest radiographs. The drain is then inserted with the patient in a sitting position preferably on a chair facing and resting on the chair back. Both arms are folded to draw the scapulae away from the midline. The operator stands behind the patient and prepares the skin over the upper posterior thorax. The point of entry is midway between the spinous process of the seventh thoracic vertebra in the midline and the upper medial border of the scapula adjacent to the spine of the scapula. The position is approximately one hand's width away from the thoracic spine (Fig. 23 (a, b)). The skin and subcutaneous layers in this area are infiltrated with 1 per cent lignocaine.



Fig. 23. (a) Landmarks for insertion of the suprascapular apical chest drain. (b) Drain in place (see text).

The number one needle is then changed for a long spinal needle which is aimed downwards towards the posterior aspect of the thoracic inlet. This is roughly perpendicular to the skin and the needle is advanced through the trapezius and rhomboid muscles until the first rib interspace is located. Aspiration of the needle produces air or fluid as the apical thoracic space is entered. The skin to pleural space depth can be estimated by judging the length of needle outside the skin. The needle is left in place as a guide to the direction of the scalpel incision. A 1.5 cm transverse skin incision is then made over the reference point and the track to the pleural space developed by advancing the scalpel vertically through the trapezius muscle for a variable distance. This facilitates passage of the trocar and intercostal drain. During this procedure care must be taken not to allow the trocar to deviate medially towards the structures in the midline. The drain is aimed towards the ipsilateral nipple. The posterior upper ribs tend to overlap each other from above downwards and when the ribs are reached it is sometimes necessary to deviate the trocar tip anteriorly to pass through the intercostal space.

When the drain reaches the pleural cavity it should only be advanced for a further 5 to 7.5 cm (2 to 3 inches). When a suitable position is obtained the drain is sutured into place and attached to the underwater seal. The drain is usually looped anteriorly over the shoulder and fixed to the upper chest. The drain then exits the thoracic cavity in a gentle curve which avoids kinking. This seems to be a particularly comfortable drain for the patient. Its position is verified by chest radiograph.

Care of the chest drain *in situ*

Once the drain is inserted in the required position and securely fixed no further adjustments should be made. In particular, the temptation to advance the drain further into the chest should be avoided since this will introduce infection into the track. The wound should be sealed around the drain with povidone iodine gel or the wound sprayed with clear plastic dressing. Dressings are unnecessary although a thin square of gauze around the drain will absorb exudate. The ill-advised application of large quantities of adhesive tape should be avoided. Many patients suffer blistered skin from sheets of Elastoplast. This practice also hinders expansion of the chest wall and physiotherapy.

The chest drain should never be clamped in the presence of an air leak because of the risk of tension pneumothorax. In contrast, disconnection of the drain from its underwater seal merely allows air to enter the pleural cavity through elastic recoil of the lung. Reconnection of the drain to its underwater seal and resumed suction will immediately clear the pneumothorax.

For patients with simple pleural effusion, chylothorax, or empyema, the drain can be clamped safely prior to changing the underwater seal drainage bottle. During resolution of the acute problem the patient may be taken off continuous suction and allowed to walk around the ward carrying the drainage bottle. For patients with a long-term air leak after pulmonary surgery, the drainage bottle may eventually be replaced with a Heimlich flutter valve. This contains a long, thin-walled rubber tube which allows release of air but collapses when inspiration transmits negative pressure through the drain (Fig. 24).



Fig. 24. The Heimlich ‘flutter' valve.

When transporting the patient with a chest drain, clamps are usually unnecessary and clamps must never be used for patients with a continuous or intermittent air leak. The drainage bottle is carried at a level below the patient's chest to avoid syphon of the water seal back into the chest. It is advisable to clamp the intercostal drain briefly when the bottle must be lifted across the patient during positioning on the operating table, but only in patients with haemothorax or pleural effusion and not those with a brisk air leak. For air leak patients it is safer to disconnect the drain from the underwater seal system and reconnect as soon as possible. The dangerous practice of indiscriminant chest drain clamping has led to the death of air leak patients during interhospital transfer for pleurectomy or pleurodesis.

When to remove the chest drain

Any surgical drain acts as a conduit for bacterial infection. The chest drain should therefore be removed as soon as its purpose is accomplished. Long-term drainage may be required for chronic air leak in patients with obstructive airways disease or bullous emphysema and in some patients with chylothorax and empyema. Problems which remain unresolved after 2 to 3 weeks should be treated surgically. An intercostal drain should not be left in one site for more than 3 weeks since erosion of the intercostal vessels may cause haemorrhage.

The presence of a drain in the pleural cavity promotes a local inflammatory response and pleural adhesions. This is part of the therapeutic process of chest drainage but is not an indication for prolonged drainage times, which are associated with increased infection rates and morbidity. In patients with simple spontaneous pneumothorax and traumatic haemopneumothorax, it takes an average of 12 h to achieve full lung expansion and cessation of air leak. Suction on the drains promotes clearance of the pleural cavity and approximation of the visceral and parietal layers.

In practice many patients are left with a drain *in situ* for 5 to 7 days, usually without a definite indication. With pressure on hospital beds and resources this duration of intercostal drainage is rarely necessary. In approximately 80 per cent of patients drained for pneumothorax the air leak stops within 12 h. Recent studies have shown that tube removal shortly afterwards does not lead to a higher redrainage rate than for patients whose tubes are kept for longer periods. Mobilization and physiotherapy promotes full re-expansion of the lung. Patients who continue to leak air after re-expansion require drainage until the air leak stops. Suction is applied continuously to keep the parietal and visceral pleura in close apposition, thus promoting the development of inflammatory adhesions. When the lung remains expanded when suction is stopped, suction is suspended for mobility during the day but reconnected at night. After lung resection the drain can usually be removed at 2 weeks, even when bubbling persists. As an interim measure the underwater seal can be replaced with a Heimlich valve to allow free mobility. The drain can eventually be removed 12 h after the last bubble has been observed.

Drains may be retained temporarily to cover anaesthesia for extrathoracic problems or in patients requiring positive pressure ventilation or continuous positive airway pressure management. For patients with pleural effusion, empyema or chylothorax, the drain is kept *in situ* until the fluid leak stops or blood turns to serous fluid. Failure of resolution within 3 weeks usually leads to surgical intervention.

Drain removal

To avoid pneumothorax through aspiration of air, tube removal is performed during a sustained Valsalva manoeuvre which forcibly inflates the lung against the chest wall. Breathing is suspended until the purse string suture is tied. Meanwhile the drain track can be occluded by finger pressure 1 cm above the skin incision.

The skin suture should not be overtightened or rapidly snapped closed since this can cause ischaemia and necrosis within the area of constriction. Sterile technique is not necessary for drain removal but the tip of the drain should be cut off with sterile scissors and sent for culture. The wound is sprayed with clear plastic dressing and covered lightly with a gauze swab or left uncovered. Extensive Elastoplast dressings are not useful and irritate the skin.

Many units ask for a plain chest radiograph after drain removal. Auscultation of breath sounds gives the same information and even in the presence of a small residual pneumothorax redrainage is unnecessary. Chest radiograph is therefore an unnecessary expense and is more useful 24 h later, prior to hospital discharge.

Complications of chest drainage and how to avoid them

Failure to enter the pleural cavity

This occurs commonly through poor technique. The chest radiograph may be deceptive when the drain passes forwards or backwards within the soft tissues of the chest wall. The problem is recognized when the expected pleural contents, air, blood, pus, or chyle, do not enter the drain and no clear respiratory swing is obtained on deep inspiration. This may occur with lateral, anterior, or posterior approaches to the pleural cavity. We have seen trocar insertion directly into the mediastinum, brachial plexus, and subclavian vessels during attempted entry of the second intercostal space anteriorly. With the lateral approach the drain may end in the soft tissues of the axilla. When the expected result is not achieved the drain should be removed and the procedure repeated.

Penetration of the lung

Minor lung lacerations are common after trocar drain insertion. When good technique has been used the damage is slight and resolves spontaneously. The most important lung injuries occur when a drain is inserted inappropriately into an obliterated pleural cavity. Postpneumonic adhesions tether the lung to the chest wall in 15 to 20 per cent of patients. This can rarely be recognized from the plain chest radiograph alone but is suspected in patients with a past history of pleurisy, pneumonia, or tuberculosis (Fig. 25). Major lung laceration usually occurs in trauma patients who present with surgical emphysema in the chest wall. This occurs in the absence of pneumothorax because the lung is unable to collapse and lacerated ribs cause air to leak directly in to the extrathoracic tissues. If the chest radiograph shows no blood or air in the pleural cavity then a drain must not be inserted without digital confirmation of a pleural space. We have seen serious intrapulmonary bleeding after anaesthetists have asked for ‘prophylactic' pleural drainage prior to positive pressure ventilation in this type of patient.

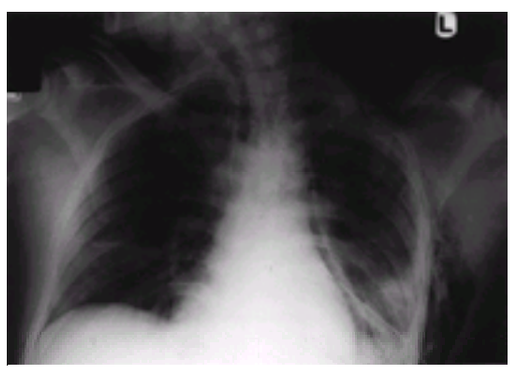


Fig. 25. Chest drain (left side) penetrating directly through the lung into the mediastinum. The drain was inserted ‘prophylactically' in a patient about to undergo laparotomy for ruptured spleen. The result is intrapulmonary bleeding and subcutaneous and mediastinal emphysema.

Penetration of the peritoneal cavity

Liver, spleen, stomach, and bowel have been transfixed by chest drains. This occurs through poor technique when the drain is inserted beneath the sixth intercostal space, and penetrates through the lower part of the pleural cavity and diaphragm into the upper abdomen. This also occurs after traumatic rupture of the diaphragm if siting of the drain has not taken into account the radiological appearance (Fig. 26). It is avoided by siting the drain above the sixth interspace, or higher when a large pleural effusion obscures the position of the diaphragm.

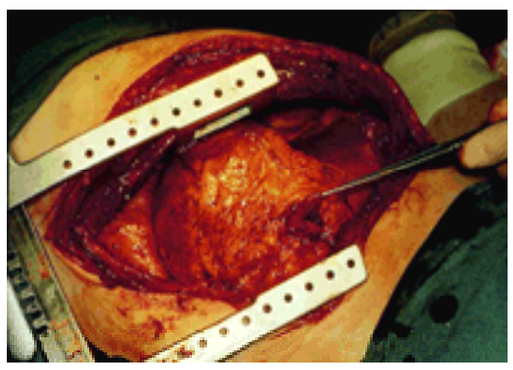


Fig. 26. Traumatic rupture of the left hemidiaphragm. The chest drain penetrated the intrathoracic gastric fundus.

When deep liver penetration occurs we advocate leaving the drain *in situ* for 48 h until resolution of blood clot prevents major haemorrhage. Penetration of the colon requires early laparotomy for peritoneal toilet. Penetration of the spleen or stomach may be managed conservatively, though clinical deterioration will often precipitate laparotomy.

Major haemorrhage

Penetration of the heart is rare but may occur in patients with cardiomegaly, where the left ventricle is in apposition to the chest wall laterally. Exsanguination has occurred through left ventricular penetration followed by removal of the drain. If entry of a cardiac chamber is suspected the drain must be left *in situ* and the tube clamped pending urgent thoracotomy. Laceration of a major pulmonary vessel occurs particularly when a collapsed lung is deeply penetrated (Fig. 27). The drain should be left within the pleural cavity and the rate of bleeding determined. When immediate drainage exceeds 1000 ml or the subsequent rate 500 ml/h for three consecutive hours thoracotomy should be performed. Substantial haemorrhage may also occur through penetration of an intercostal or internal mammary vessel. This complication can usually be managed conservatively but occasionally requires exploration.

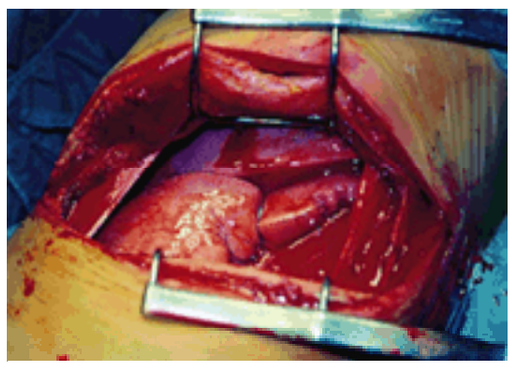


Fig. 27. Urgent thoracotomy for haemothorax following transfixion of the left pulmonary artery.

Insertion of the drain into a bulla

Large bullae mimic pneumothorax on the plain chest radiograph. Failure to recognize the bulla may result in insertion of a drain and copious air leak. When suction is applied to the drain the air leak increases and results in significant loss of tidal volume with acute exacerbation of breathlessness. The situation is usually recognized because of the continuous large volume air leak. Removal of the drain may result in tension pneumothorax and clamping of the drain causes extensive surgical emphysema. The drain should be withdrawn from the bulla and the patient observed carefully. Tension pneumothorax may require drainage, though frequently the lung already has inflammatory adhesions to the chest wall. Thoracotomy may be undertaken to resolve this situation, although we have successfully passed a Foley catheter through the shortened drain and used the balloon to plug the hole in the bulla and traction to hold the bulla against the chest wall.

Blocked drains

Occlusion of a chest drain may lead to recurrent pneumothorax or reaccumulation of blood or fluid. The smaller the chest drain the more likely blockage is to occur. Suction promotes drainage and clearance of the pleural cavity. Clamping of the chest drain promotes occlusion. Even in pneumothorax a fibrinous exudate caused by the inflammatory reaction to the drain will occlude the side holes if suction is interrupted. A great number of pleurectomies for recurrent pneumothorax and decortications for empyema could be avoided if measures were employed to ensure continuous drainage.

Pleural sepsis

Failure to adhere to aseptic technique and unnecessarily prolonged intercostal drainage both promote pleural sepsis. Except for trauma patients, prophylactic antibiotic treatment is not recommended, but a therapeutic course of antibiotics may be required if a patient with a chest drain becomes febrile. The cutaneous drain site may be inflamed or clear serous drainage may become purulent. The drain at fault should be removed and antibiotic treatment continued. Accumulation of pus in the pleural cavity may require needle aspiration or further drainage through a separate site.

Unexpected results from drainage

Occasionally intercostal drainage produces a surprise. The pleural effusion may be chylous, may contain bile, or contain gastric contents. Chylous effusion may result from trauma or malignant occlusion of the thoracic duct. Continued drainage results in emaciation of the patient. Conservative methods of treatment by fat restriction are seldom successful and surgical pleurodesis is usually required. Bile in the right pleural cavity may follow rupture of a liver abscess through the diaphragm. This situation requires specialized treatment of the underlying pathology. Gastric contents fill the pleural cavity after spontaneous or instrumental perforation of the oesophagus. In the Boerhaave syndrome, an acute rise in intragastric pressure splits the lower oesophagus and forces the stomach contents into the left, or more commonly the right, chest. The condition is often mistaken for an acute vascular event such as myocardial infarction or aortic dissection, and is seldom diagnosed within 24 h. Early surgical repair and pleural drainage are necessary to avoid mortality or serious morbidity.

41.12 Thoracic trauma

Stephen Westaby

Mechanism of major visceral injury in thoracic trauma
- High velocity impact
- Low velocity impact
- Crush injury
- Penetrating chest wounds

Pathophysiology of trauma to the lungs and chest wall
- Altered mechanics of breathing
- Ventilation/perfusion imbalance
- Impairment of gas transfer

Practical aspects of resuscitation in major thoracic trauma
- Procuring the airway
- The mechanics of breathing
- Cardiovascular system
- Radiology

Decision making and early intervention
- Penetrating injuries
- Blunt cardiac trauma
- The ruptured aorta
- Blunt injury to the major airways
- The lacerated diaphragm
- Treatment of chest wall injuries

Further reading

The cause of traumatic injury to the chest varies greatly in different parts of the world. Currently, in large American cities and parts of South Africa, black males have a 1 in 20 chance of being fatally shot or stabbed before the age of 30. Although terrorism and civilian violence are on the increase throughout Europe, the absolute numbers of victims remains small in comparison. In England and Wales, the annual death rate from stabbing and gunshot wounds is less than 200. Britain and Europe, nevertheless, have increasing figures for road casualties. Road traffic accidents in England and Wales account for 60 000 hospital admissions per year, and London and south-east England have 57 fatal or serious road traffic accidents per 100 km of road. Blunt thoracic injury is almost exclusively caused by rapid deceleration in motor vehicle collisions. A small number of injuries follow crushing industrial accidents.

Although less than 15 per cent of patients with chest trauma require surgical intervention, many needless deaths occur through inadequate or delayed treatment of an easily remediable injury. The majority of chest injuries are confined to the thoracic cage. These consist of rib fractures with underlying pulmonary contusion, haemothorax, or pneumothorax, which can usually be dealt with simply and effectively by chest drain insertion and fluid restriction. When ignored, underestimated, or inadequately treated, chest injuries may cause the death of a patient during surgical intervention for seemingly more pressing intracranial or abdominal haemorrhage. The basis for successful management of thoracic trauma is effective cardiopulmonary resuscitation followed by early detection and treatment of life-threatening injuries. The former is based on the ABC principle:

A. establish a reliable Airway;
B. restore the mechanics of Breathing;
C. stabilize the Cardiovascular system.

The most serious intrathoracic injuries often occur in the absence of significant chest wall damage. Recognition must depend upon exclusion rather than direct manifestation of injury. The latter depends upon prediction of likely lesions according to the mechanism of injury. Important injuries to rule out are aortic transection or dissection, major airways disruption, ruptured diaphragm, and severe cardiac contusion or valvular regurgitation. Although all are relatively rare, they may coexist after major trauma. A high index of suspicion is the key to early diagnosis.

Mechanism of major visceral injury in thoracic trauma

The existence of serious visceral injury can usually be predicted with knowledge of the type of accident or assault and should be confirmed or excluded by appropriate investigation such as chest radiographs and CT scan. There are three broad categories for the mechanism of injury associated with blunt thoracic trauma (Table 1). The patterns of injuries sustained are different in each case.

Mechanism of injury	Chest wall injury	Possible thoracic visceral injuries	Common associated injuries
High velocity impact (deceleration)	Chest wall ± sternum intact or fractured sternum or bilateral rib fractures with anterior fluid (pneumaviscera)	Ruptured aorta Cardiac contusion Major airways injury Ruptured diaphragm	Head and facial maxillary injuries Fractured cervical spine Lacerated liver or spleen Long bone fractures
Low velocity impact (direct blow)	Lateral Unilateral fractured ribs Anterior Fractured sternum	Pulmonary contusion Cardiac contusion	Lacerated liver or spleen if ribs 9–12 are involved
Crush injury	Anteroposterior Bilateral rib fractures ± anterior flail	Ruptured bronchus Cardiac contusion	Fractured thoracic spine Lacerated liver or spleen
	Lateral Ipsilateral fractures ± flail Possible contralateral fractures	Pulmonary contusion	Lacerated liver or spleen

Table 1 Mechanisms of injury in blunt injuries to the thorax

High velocity impact

Sudden profound deceleration (Fig. 1) produces the so called intrathoracic ‘bell clanger' effect. The chest wall suffers direct impact. The aorta, a heavy column of blood with inertia, swings like the clanger of a bell within the thorax, producing severe shear forces. These forces may tear the vessel at its fixed points above the aortic valve or, more frequently, at the posterior chest wall, causing transection. The main bronchi situated beneath the aortic arch are subject to similar forces and may rupture. Impact of the neck may transect the trachea, and compression of the abdominal viscera may rupture the diaphragm, the spleen, or the liver. There may be little or no injury to the bony chest wall, though bilateral clavicular fractures or a fractured sternum must arouse suspicion. Anterior chest wall contusion is an indicator of underlying myocardial injury.



Fig. 1. Deceleration type automobile accident.

Low velocity impact

Low velocity impact causes direct damage to the bony thorax, with or without contusion of the underlying lungs or myocardium (Fig. 2). This type of injury does not usually create stress or compression forces sufficient to damage the aorta, bronchi, or diaphragm, although the liver or spleen may be ruptured by a direct blow over the lower part of the thoracic cage.



Fig. 2. Chest radiograph after a low velocity blow which caused rib fractures and pulmonary contusion.

Crush injury

Following compression, multiple bilateral rib fractures are likely. However, in the young the sternum may be forced backwards to touch the spine without fracture (Fig. 3). Such low velocity forces seldom damage the aorta or myocardium, but may rupture the diaphragm or lacerate the bronchi by a different mechanism from that above (Fig. 4). When sudden, forceful compression of the thoracic cage decreases the anteroposterior diameter and produces a widening of the transverse diameter, the negative intrapleural pressure ensures that the lungs remain in contact with the chest wall. Lateral motion pulls the two lungs apart thus producing traction on the trachea at the carina. Rupture occurs when the elasticity of the tracheobronchial tree is exceeded. If the glottis is closed at the moment of impact, intrabronchial pressure may rise suddenly. The greatest tension develops in the larger bronchi and increases the tendency to rupture.

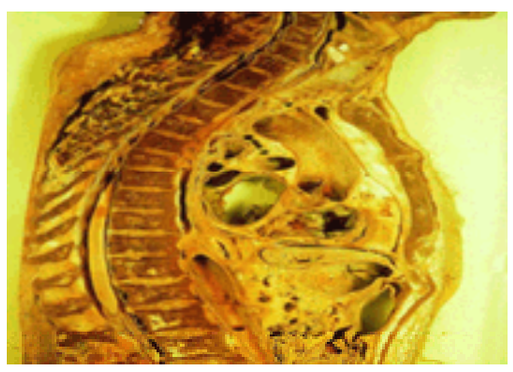


Fig. 3. Longitudinal section of the chest showing the relationships of thoracic and upper abdominal viscera.

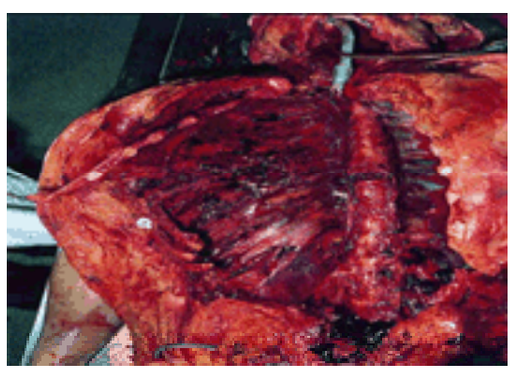


Fig. 4. Severe crush injury with multiple bilateral rib fractures and fractured spine.

Penetrating chest wounds

Knife and gunshot wounds are now more commonplace in British accident departments and in some North American cities they are responsible for up to 40 per cent of trauma admissions. The extent of damage inflicted by a penetrating agent depends on the size, shape, stability, and, above all, the velocity of the missile (Fig. 5). The vast majority of civilian gunshot wounds and accidental industrial or road traffic penetrating injuries occur with low velocity. Such missiles core out a hole through the body and damage only those tissues with which they are in direct contact. They cause death by damage to vital structures or exsanguinating haemorrhage. Because of their low kinetic energy, the path taken by a hand gun bullet through the body is unpredictable, deflection being caused by bone or even parenchymal organs such as the liver or spleen. Chest wounds may be accompanied by abdominal injury, and vice versa. The greater the mass of the penetrating object the greater the damage inflicted.



Fig. 5. High and low velocity bullets.

In contrast, high velocity rifle bullets and fragments from explosive devices over a short range have a large amount of energy which causes damage remote from the path of the missile itself. High velocity bullets cause extensive tissue damage by cavitation and shock waves (Fig. 6). Energy from the bullet is dissipated into the surrounding tissues, which are violently accelerated forwards and outwards. This creates a large temporary cavity 30 to 40 times the diameter of the missile. The greatest extent of cavitation occurs only after the missile has passed through the tissues. This has subatmospheric pressures and is open at both entrance and exit holes. The cavity then collapses in a pulsatile fashion sucking air, debris, and bacteria into the wound and producing a large amount of dead or devitalized, contaminated tissue.

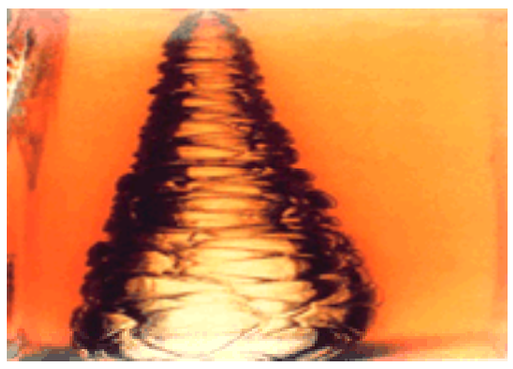


Fig. 6. Cavitation caused by a high velocity bullet.

Damage is directly proportional to the density of the tissue: homogeneous structures such as brain, liver, spleen, or muscle are very sensitive, whereas light tissues such as lung, which mainly consists of air, are resistant. The destruction is also inversely proportional to the proportion of elastic fibres present: skin and lung are resistant whereas bone is shattered. The external appearance of a bullet wound in the chest is therefore deceptive. The elasticity of the skin produces tiny entry and exit holes that disguise extensive internal destruction.

Pathophysiology of trauma to the lungs and chest wall

Chest injuries adversely affect pulmonary function by three separate mechanisms: altered mechanics of breathing, ventilation/perfusion imbalance, and impairment of gas transfer (see Chapter 11.3 and Chapter 11.5).

Altered mechanics of breathing

The great majority of blunt injuries to the thoracic cage and those penetrating injuries that cause haemothorax or pneumothorax impair ventilation (Fig. 7). Even relatively minor trauma with fractured ribs but no underlying pathology may cause pain sufficient to lead to hypoventilation, atelectasis, failure to clear secretions, pneumonia, septicaemia, respiratory failure, and even death in an elderly or bronchitic patient. More serious problems which cause severe impairment of the mechanics of breathing include pneumothorax (particularly tension pneumothorax), haemothorax, ruptured diaphragm, multiple rib fractures with unstable segments, and injuries to the major airways (Fig. 8). The most extensive disruption of the chest wall tends to occur with crush injuries where multiple bilateral rib fractures, ruptured diaphragm, or fractures of the spine or sternum coexist. A single rib fracture is associated with a blood loss of 150 ml. Multiple fractures may cause substantial blood loss into the chest wall and pleural cavity; this may be increased by laceration of the underlying lung by sharp edges.

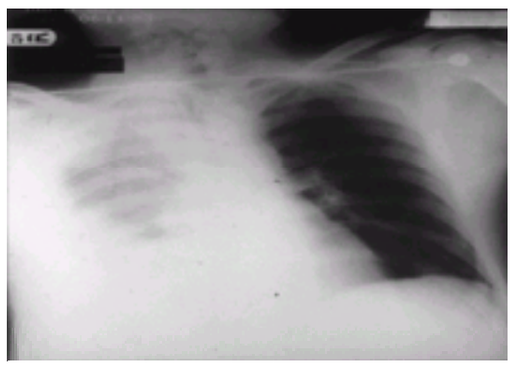


Fig. 7. Haemothorax following a bullet wound to the right chest.



Fig. 8. Tension pneumothorax after blunt chest injury with ruptured bronchus. The drain failed to enter the pleural cavity.

There are two main types of chest wall derangement, a functionally important traumatic defect (sucking chest wound) or a flail segment (Fig. 9). The latter may be unilateral with double fractures of three or more ribs, or anterior with fractures of three or more ribs on both sides of the sternum (Fig. 10). Some fractures may occur through the costochondral junctions, and are therefore invisible on plain chest radiographs. The unstable segment moves inwards on inspiration (paradoxical movement) and consequently compromises ventilation by reducing tidal volume. Following diaphragmatic rupture, the abdominal contents are similarly sucked into the chest on inspiration (Fig. 11). If the pleural cavity is filled with air or blood, ventilation of partially collapsed lung is similarly compromised. Injuries to the major airways or inhalation of foreign material including teeth, windscreen glass, or stomach contents may physically occlude the large or small airways, thereby obstructing air entry.

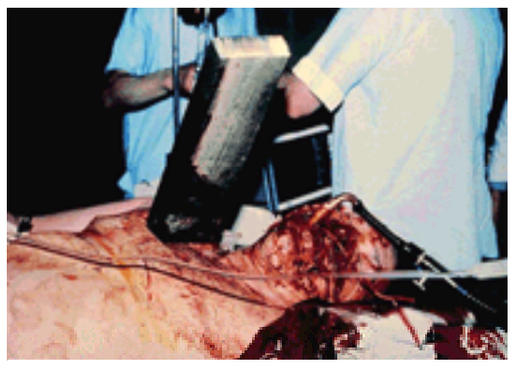


Fig. 9. Defect in the anterior chest wound after transfixion injury. This caused a ‘sucking chest wound'.

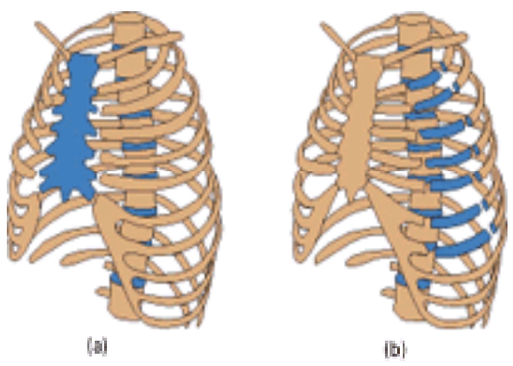


Fig. 10. Flail segments due to multiple unilateral or bilateral rib fractures.

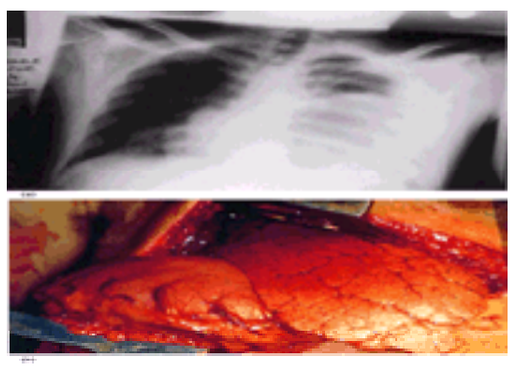


Fig. 11. Ruptured left hemidiaphragm after blunt chest trauma. (a) Plain radiograph. (b) Stomach in chest at thoracotomy.

Ventilation/perfusion imbalance

Effective oxygenation of the blood and elimination of CO_2 depends on a balance between ventilation of the lung and its blood supply. Thoracic injuries compromise ventilation perfusion/balance by a number of different mechanisms. Mechanical obstruction of the airway is an obvious cause of impaired ventilation, but in practice other mechanisms predominate. Distribution of ventilation in the lung is influenced by regional variations in airways resistance and compliance. The latter is governed by gravity-dependent intrapleural pressure gradients, and gas distribution in the lung is therefore uneven during normal resting tidal ventilation. In the lower or more dependent pleural space the pressure is closest to atmospheric (least negative). It becomes increasingly negative towards the apex or non-dependent region of the lung. When a normal inspiration is taken from end-expiration (functional residual capacity) expansion of the initially smaller dependent alveoli is governed by the steep portion of the compliance curve. Consequently, they expand more for each unit of pressure change than do those at the apices which are influenced by the upper, flatter portion of the curve. These differences result in preferential distribution of inspired gases to the areas of greater expansion in the dependent portions of the lungs. However, if terminal air spaces have collapsed due to haemopneumothorax or adult respiratory distress syndrome, inspiration results in preferential distribution to already expanded areas because of the influence of the compliance curve. The maldistribution of ventilation is worsened by airflow obstruction in terminal airways due to external compression, elevated intrapleural pressure, or interstitial oedema fluid. When a whole lobe or lung eventually collapses, there is perfusion of non-ventilated lung and a serious veno-arterial shunt effect. Ineffective oxygenation of the venous blood is reflected by widening of the alveolar–arterial oxygen tension difference and systemic hypoxia.

Movement of blood through the lungs is also influenced by gravity and the pressure gradient between the pulmonary arteries and left atrium. Blood flow is normally directed preferentially to the dependent parts of normal lung where ventilation is also most efficient. However, perfusion is often impaired by thrombosis of vessels in contused lung or widespread pulmonary microembolism by fat from bone marrow or platelet/neutrophil microemboli in patients with disseminated intravascular coagulation or adult respiratory distress syndrome.

In many patients after even minor thoracic trauma, the effects of ventilation/perfusion mismatch may lead to unsuspectedly severe hypoxia which is seldom recognized without blood gas analysis. Pulmonary contusion, intrapulmonary haemorrhage, and haemothorax or pneumothorax are invariably associated with serious deterioration in pulmonary function.

Impairment of gas transfer

Passive diffusion of gas across the alveolar capillary barrier is dependent on the surface area available, the width of the membrane, certain plasma and erythrocyte enzymic factors, and the partial pressure gradient between the alveolar and vascular spaces. Following thoracic trauma a number of factors, including injury to the pulmonary parenchyma by contusion, damage to the alveolar capillary barrier by inhalation of gastric contents or smoke, impaired cardiac output, and interstitial pulmonary oedema due to overtransfusion of crystalloid, colloid, or blood (elevated left atrial pressure), may adversely affect gas exchange. However, the most sinister process involved is that which begins with the pathophysiological effects of shock and, if uninterrupted by prompt resuscitation, may progress to adult respiratory distress syndrome (Fig. 12). It is not appropriate to discuss the detailed humoral and cytological changes that culminate in this syndrome (see Chapter 11.3). Nevertheless, they are important, are probably triggered by activation and interaction of the complement, coagulation, kallikrein, and plasminogen cascades, and result in trapping of ‘activated' neutrophils in the pulmonary microvasculature. Here they release protease enzymes and generate oxygen free radicals with the potential to damage the alveolar capillary membrane. The clinical and radiological stages of adult respiratory distress syndrome are summarized in Table 2. When full-blown adult respiratory distress syndrome occurs in a patient with thoracic trauma, particularly following multiple injuries, the chances of survival decrease markedly. Gas exchange is impaired by extension of the diffusion pathway by the presence of hyaline membrane and oedema fluid in the alveoli; accumulation of interstitial fluid within the septum and of proliterated type II cells along its alveolar border; reduction in the surface for diffusion because of terminal air space collapse and closure of capillary channels; and the detrimental influence of consequent hypoxaemia, hypercapnia, and acid–base shifts in erythrocyte enzyme kinetics. Some of the consequences of this process precipitate further deterioration in pulmonary dysfunction. Pulmonary arterial hypertension develops because of hypoxic arteriolar vasoconstriction. Higher flow resistance in small vessels can be made worse by increased interstitial fluid pressure. Intravascular coagulation may exacerbate these problems and contribute to ventilation/perfusion mismatch.

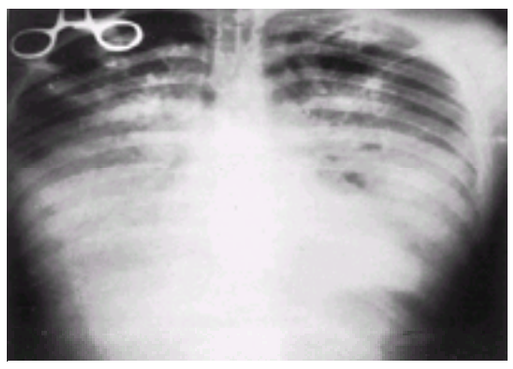


Fig. 12. Plain chest radiograph in a patient with adult respiratory distress syndrome.

Phase	Clinical evidence of respiratory distress	Chest radiograph	Whole breathing gas mixture (F_IO_2 0.21) Pa_{O_2} mmHg	Pa_{CO_2} mmHg	$A\text{-}aDO_2$ mmHg	Q_S/Q_T (%)
Normal			80–100	35–45	5–10	5
I Injury and resuscitation	None	Normal	70–90	30–40	20–40	10–20
II Subclinical respiratory distress	Mild to moderate tachypnoea	Minimal or no infiltrate	60–80	25–35	30–50	15–25
III Established respiratory distress	Increasing tachypnoea and increasing dyspnoea	Confluence of infiltrates	50–60	35–50	40–60	25–40
IV Severe respiratory failure (later stages)	Obvious respiratory labour	Increasing opacification of the lungs	35–50	35–60	60–80	40–75

Abbreviations: Pa_{O_2}, arterial oxygen tension; Pa_{CO_2}, arterial carbon dioxide tension; A-aDO$_2$, alveolar-arterial oxygen difference; Q_S/Q_T, physiological shunting in the lungs as a result of various conditions

Table 2 Clinical and radiological stages in adult respiratory distress syndrome

Hypoxaemia is the first objective sign of the onset of adult respiratory distress syndrome, and is the cardinal index of its progressive severity. At a later stage CO_2 retention develops, with its consequent disturbances of acid–base balance. Unless blood gases are monitored continuously in patients with thoracic trauma the primary effects in the lung may be misinterpreted. The manifestations of respiratory insufficiency are reflected principally in deterioration of cardiovascular and central nervous system dysfunction.

Practical aspects of resuscitation in major thoracic trauma

Advances in prehospital transportation have provided an opportunity to sustain life in patients who would previously have died (Fig. 13). In the United States, Europe, and Japan, helicopter retrieval with resuscitation en route by positive pressure ventilation, external cardiac massage, and blood volume replacement are employed routinely for patients with major trauma. Relief of cardiac tamponade, con-trol of exsanguinating haemorrhage, stabilization of respiratory dynamics, and the use of mechanical circulatory assistance should therefore be integral steps in accident department resuscitation rather than sequels to resuscitation. The decision to employ such procedures must be undertaken on clinical judgement and without recourse to radiology or special investigations.



Fig. 13. On-site resuscitation (a) and helicopter transportation (b) of a patient with major injuries.

More than 80 per cent of all patients with ruptured aorta, trachea, or bronchus die rapidly at the site of injury. Those who reach hospital alive are a self-selected group who should survive with appropriate treatment. Once in hospital, resuscitative manoeuvres are best carried out with caution and in an unhurried way, since most survivors have reached an impaired but balanced respiratory and haemodynamic status compatible with life. More drastic measures, such as thoracotomy in the accident department, are reserved strictly for those with penetrating injuries or severe chest wall disruption who arrive moribund and in circulatory arrest. The admitting department should have an area set aside for patients with severe injuries. This should contain a trolley that can be tipped, and that has adequate space around it on all sides for manoeuvre. An anaesthetic trolley should be readily at hand with chest drains, peritoneal lavage cannulas, and a variety of intravenous and central venous cannulas for use in the large veins. An overhead radiograph machine and CT scanner should be available and an autotransfusion system is desirable (Fig. 14). Once in the accident department, the airways, breathing, and cardiovascular status are reviewed and take priority over other procedures as appropriate (Table 3). A brief but thorough history is obtained from the patient, witnesses, or police officers. The time and mechanism of injury, type of weapon, and condition of the victim during transport are noted. Nursing and medical staff remove all the patient's clothes but, in the case of penetrating wounds, must preserve them carefully for future forensic or medicolegal investigation. Orientation of the garment on the patient, traces of gunpowder, or missile fragments may constitute important evidence, and even tissue debrided from wound edges should be retained. The vital signs, pulse and respiratory rate, blood pressure, and central venous pressure are measured, and the haemodynamic state and volume of overt or concealed haemorrhage assessed. A central venous catheter and one or more large bore peripheral cannulas are inserted.



Fig. 14. The trauma unit CT scanner is used for rapid diagnosis of head, chest, and abdominal injuries.

A. Establish a reliable airway
Oropharyngeal airway
Endotracheal tube
Bronchoscope
Tracheostomy
B. Restore the mechanics of breathing
Artificial respiration
Evacuation of haemopneumothorax
Stabilization of unstable chest wall
Mechanical ventilation
C. Resuscitate the cardiovascular system
Intravenous infusion of crystalloid colloid or blood
Restore acid-base status and electrolytes
Inotropic support
External or internal cardiac massage
Immediate surgery to stem haemorrhage

Table 3 The ABC principle of cardiopulmonary resuscitation

In the collapsed patient a Ryle's nasogastric tube, quickly inserted by cutting down into an antecubital vein and passing into the superior vena cava, is excellent for rapid blood transfusion, venous pressure measurement, and repeated venous sampling (Fig. 15). An arterial line facilitates blood pressure recording and blood gas analysis. Blood volume replacement is initially carried out using crystalloid or colloid substitutes, but after massive blood loss or continued rapid bleeding, oxygen carrying capacity and coagulation must be preserved with group O Rhesus negative blood or the bleeding stopped by immediate surgical intervention. No attempt should be made to attain normal blood pressure at this stage: this may precipitate fatal haemorrhage from lacerated major vessels, intracranial bleeding, or cardiac tamponade. Controlled hypotension is desirable until the extent of injury is determined or surgical access to damaged structures is obtained. Whenever feasible, initial evaluation (including radiology) and resuscitation should be completed within 5 to 10 min. A patient whose vital signs cease or continue to deteriorate should undergo rapid endotracheal intubation and should be removed immediately to an area where operative resuscitation and repair can begin. For some moribund patients operation is a co-ordinated part of resuscitation (Fig. 16).

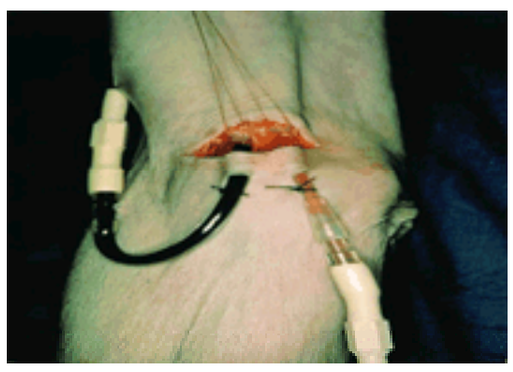


Fig. 15. Ryle's tube used for rapid transfusion, venous pressure measurement, and venous sampling.



Fig. 16. Emergency room thoracotomy and laparotomy for control of exsanguinating haemorrhage in a moribund patient. There was a successful outcome.

External cardiac massage is seldom successful in restoring cardiac activity in patients who have suffered penetrating trauma. Immediate thoracotomy for control of haemorrhage, cardiac compression, preservation of available blood volume, and direct repair of critical injuries can be life saving in patients who present in extremis with injuries that preclude transportation to the operating theatre, and in those whose condition deteriorates during the first few minutes following arrival, suggesting impending death despite resuscitation. ‘Do not meddle with an obviously dead patient' is an important dictum: patients with detectable vital signs can usually be sustained as far as the operating theatre. However, within these constraints there have been encouraging results from the use of immediate anterolateral thoracotomy for release of cardiac tamponade, suture repair of great vessels and cardiac wounds, internal cardiac massage, and cross-clamping of the descending thoracic aorta (Fig. 17). This last manoeuvre improves perfusion of the coronary and cerebral circulation in patients with profound hypovolaemia and stems exsanguinating haemorrhage in the abdomen.

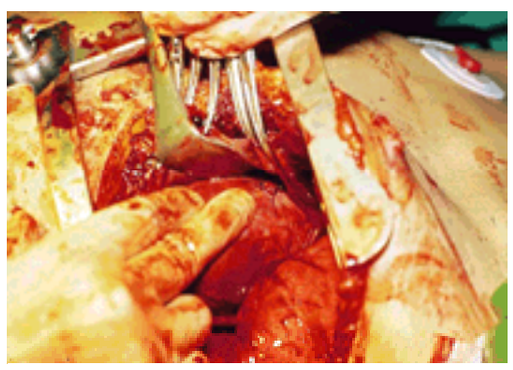


Fig. 17. Immediate thoracotomy for relief of tamponade and suture of a penetrating cardiac wound.

In one group of patients with no immediately discernible vital signs or with systolic blood pressure less than 60 mmHg, and with preterminal respiratory pattern and cerebral activity, who did not respond to volume replacement and ventilatory support, there was a 66 per cent survival rate following immediate thoracotomy when the injuries were limited to intrathoracic organs (Fig. 18). This contrasted strongly with only 20 per cent survival when thoracotomy was performed for cardiac resuscitation in patients whose injuries were totally extrathoracic.



Fig. 18. Patient with bullet wound involving the left main pulmonary artery. Immediate thoracotomy and clamping of the pulmonary hilum ensured survival.

The blood pressure level at 60 mmHg is important: young and fit individuals compensate for blood loss by vasoconstriction and pressure is maintained until critical hypovolaemia and acidosis occur. The extent of blood loss is disclosed by insertion of a central venous pressure line, though false elevation of venous pressure occurs with cardiac tamponade, shivering, straining, or poor position of the catheter. Measurement of base deficit is a sensitive index of the efficacy of resuscitation. Typing and cross-matching for at least six units of blood is performed at the same time.

In victims of assault or attempted homicide, the entire body should be examined for signs of penetration, each site being marked with a radio-opaque marker so that the track of the missile can be established. Decreased or absent breath sounds, subcutaneous emphysema, tracheal displacement, and neck vein distension are sought. The girth of the abdomen is measured and the presence of abdominal distension, localized and rebound tenderness, guarding, rigidity, frank dullness on percussion, and hyperaesthesia over the shoulder (Kehr's sign indicative of subdiaphragmatic irritation by blood) are sought. When the patient is unconscious it is easy to miss serious injuries such as paraplegia. A catheter is inserted into the bladder to monitor urine flow. A nasogastric tube is used to empty the stomach.

Procuring the airway

After basic steps, including clearance of blood and mucus from the pharynx, conventional endotracheal intubation can usually be used to secure an airway. This may prove impossible or undesirable in patients with direct injury to the major airways or severe faciomaxillary trauma (Fig. 19). Clues to disruption of the larynx, trachea, or major bronchi include aphonia, stridor, haemoptysis, and free air in the subcutaneous tissues or thorax (radiologically, pneumothorax, pneumomediastinum, or pneumopericardium). The severity of such injuries may be deceptive (Fig. 20).



Fig. 19. Severe faciomaxillary trauma may complicate tracheal intubation.

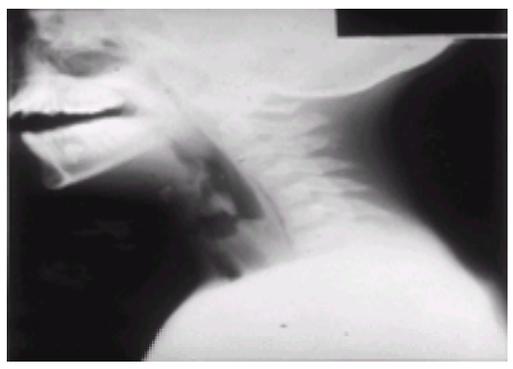


Fig. 20. Transected larynx after blunt neck injury. Initially the patient could breathe and talk normally.

Direct negotiation of the airway by rigid bronchoscopy or flexible intubating fibrescope is safer than blind intubation, which may completely obliterate a critical narrowing. If a patient with stridor is breathing spontaneously a helium–oxygen mixture may alleviate asphyxia until the bronchoscope is passed. Severe laryngeal injuries usually require tracheostomy.

In patients with tracheal trauma, tracheostomy adds a second tracheal injury and may fail to secure an airway when the lesion is intrathoracic. In those with tracheal transection, an endotracheal tube can be passed distal to the injury over a guide bougie inserted at bronchoscopy. Alternatively, high frequency jet ventilation can be used via a narrow catheter inserted into the distal segment pending direct surgical exposure. In desperate circumstances, immediate exploration of the neck is required. The distal trachea is located by a finger passed into the superior mediastinum, grasped in forceps, delivered, and intubated directly.

The mechanics of breathing

Arterial blood gases should be analysed in all patients with chest injury since apparently minor chest wall injuries may cause major degrees of hypoxaemia. In those with uncomplicated chest wall injuries, pain and anxiety are usually the biggest barrier to adequate respiratory excursion; once these are relieved ventilation improves substantially. It is imperative to restore both gas exchange and acid–base balance as soon as possible, by whatever means, since these abnormalities seriously compromise cardiovascular function.

The mechanics of breathing may be compromised by disruption of the bony chest wall or diaphragm, accumulation of blood or air in the pleural cavities, or haemorrhage and oedema in the lung itself, which reduces compliance (Fig. 21). A neurological injury may alter the respiratory drive and pattern of breathing. Early resuscitative measures include stabilization of a flail segment or covering a chest wall defect (sucking chest wound), intercostal drainage of a haemopneumothorax, and, sometimes, positive pressure ventilation from the outset. A large plastic adhesive membrane such as ‘Opsite' is useful for covering or stabilizing the chest wall; this can be supported by dressings and strapping until surgical debridement. A tension pneumothorax is always readily apparent on clinical examination and should never wait for radiological confirmation. A needle inserted through the chest wall will confirm the diagnosis but is useless therapeutically. A scalpel incision will relieve tension and restore both ventilation and venous return; an intercostal drain can then be inserted with aseptic technique. At this point remember that by lowering intrathoracic pressure venous return will increase, as will bleeding from damaged structures. Ventilatory support is initiated promptly when respiratory efforts are inadequate or obstructed. If a bronchial or pulmonary air leak coexists, it is important to allow air to escape continuously, otherwise tension pneumothorax results. A large volume suction pump (Tubbs–Barrett) should be applied to chest drains inserted for haemopneumothorax (Fig. 22). Negative pressures of –15 to 20 cmH_2O are applied unless a large bronchial air leak results in extraction of a significant proportion of the tidal volume. Low volume (Roberts) pumps should never be used in

trauma patients because they may obstruct air flow.

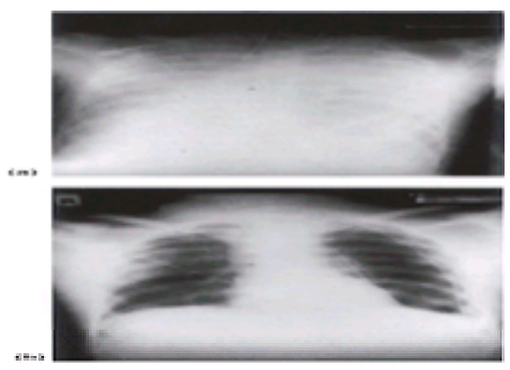


Fig. 21. Ventilatory impairment by extensive intrathoracic injuries. (a) Before and (b) after bilateral thoracotomies.

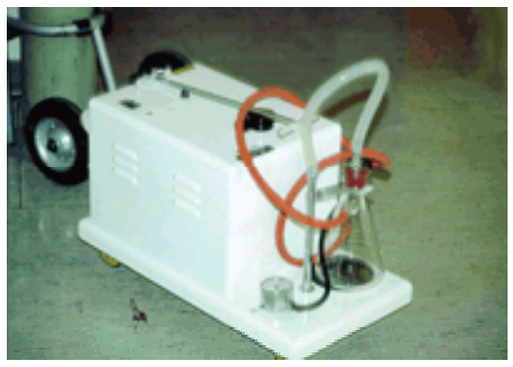


Fig. 22. Large volume (Tubbs–Barrett) suction pump for evacuation of haemopneumothorax.

Cardiovascular system

Assessment of external or concealed blood loss is particularly difficult in patients with chest trauma accompanied by multiple long bone fractures or abdominal injury. Physical signs and measurement of systemic and central venous pressure may be deceptive for the following reasons. First, cardiac contusion (predominantly of the right ventricle) or cardiac tamponade produces spuriously elevated right atrial pressure. After substantial haemorrhage this may fall within normal limits even in the presence of severe tamponade. Second, major vascular injuries such as aortic transection separate the baroreceptors from the vessel lumen and reflexly produce systemic hypertension. After serious blood loss the systolic blood pressure can still be normal. Lastly, fit young adults compensate for blood loss remarkably well. For these reasons it is imperative to measure and continue to monitor central venous pressure in every patient who is thought to require volume replacement. Peripheral pulses are checked and the presence of bruits sought in the neck. The possibility of injury to the heart or great vessels must be considered, however unlikely, since it is the unexpected rather than the obvious which most often leads to disaster. When there is no exit wound in the chest an abdominal injury may coexist. In patients surviving with either cardiac tamponade or a ruptured aorta, a balance is established between systemic hypotension and occlusion of the traumatic defect by blood clot or haematoma, which can create pressure beneath the surrounding tissues (Fig. 23). Over-aggressive blood volume replacement raises intravascular pressures, upsets this equilibrium, and may cause fatal bleeding or tamponade. In these circumstances, controlled hypotension is maintained until surgical repair. A systolic pressure of 80 to 100 mmHg should not be exceeded, especially in the presence of satisfactory acid–base balance and urine flow.

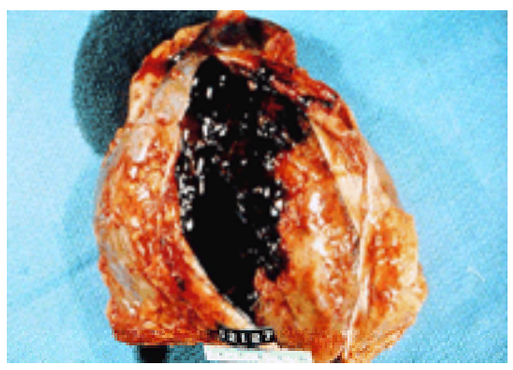


Fig. 23. Cardiac tamponade postmortem specimen. Note that clotted blood is not amenable to needle aspiration.

Needle pericardiocentesis is usually a waste of valuable time in trauma patients: cardiac tamponade is diagnosed clinically and should be relieved surgically. Except in the moribund patient, this procedure should be performed in the operating theatre to enable control of the source of haemorrhage.

In patients with severe myocardial contusion or unexplained haemodynamic deterioration, a Swan–Ganz catheter is used both to measure the right ventricular, pulmonary arterial, and left atrial (wedge) pressures, and to monitor the need for, and efficacy of, inotropic support. Contused lungs are very sensitive to volume overload and often contain fat emboli. The goal of fluid and blood replacement in such cases is to restore the intravascular volume, tissue perfusion, and oxygen carrying capacity as rapidly as possible, without causing pulmonary dysfunction or reactionary haemorrhage. The debate regarding composition of such an infusion is self-perpetuating. There is general agreement that acute, massive blood loss requires resuscitation, at least in part, with whole blood and/or blood components. When packed cell volume falls below a critical value, oxygen delivery to the tissues is severely impaired as a result of insufficient circulating red cells. In healthy adults, compensatory cardiac and pulmonary mechanisms are effective when packed cell volume falls below 0.20 but generally if this falls below 0.30 red cells should be infused. Thus, in acute loss of 1 to 1.5 litres, whole blood is not essential and blood volume can safely be restored with a plasma substitute. If the volume of haemorrhage is larger than this, whole blood should be available early in order that fluid volume and oxygen carrying capacity can be simultaneously replaced.

Radiology

When the patient is haemodynamically stable an upright chest and abdominal radiograph (preferably posteroanterior) is obtained in full inspiration. The patient may require physical support and care must be taken not to disconnect catheters, since air embolism may occur on inspiration in the presence of low venous pressure. Supine radiographs are of limited value for assessment of intrapleural haemorrhage or intra-abdominal injury but may be the only alternative in the patient with multiple injuries. Sometimes there is surprisingly little abnormality, especially when direct injury to the mediastinum leaves the pleural cavities unscathed. Heart size may prove deceptively normal despite cardiac laceration and tamponade, and this should not cast doubt on a clinical diagnosis based on circulatory collapse with raised venous pressure (Fig. 24(a), Fig. 24(b)). Widening of the mediastinum occurs after damage to the major vessels, and if ruptured aorta is suspected an aortogram (not CT scan) must follow. Mediastinal air, subcutaneous emphysema, pneumothorax, and pneumopericardium are good evidence for injury to the bronchial tree or lung parenchyma (Fig. 25). Haemopneumothorax is the most common finding after penetration of the pleural cavity. The CT scan provides little information that cannot be derived from plain chest radiographs but is excellent for identifying associated head and abdominal injuries.



Fig. 24. Cardiac tamponade after a stab wound. (a) Plain chest radiograph. (b) Distended neck veins confirm the diagnosis.

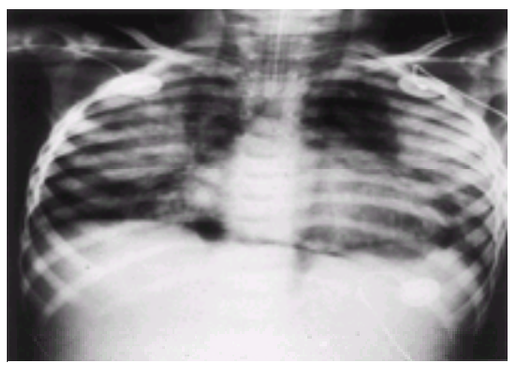


Fig. 25. Extensive mediastinal free air in a patient with tracheal transection.

Decision making and early intervention

Penetrating injuries

In practice, insertion of an intercostal drain (size 28–36 F) is the only intervention required in up to 85 per cent of patients with penetrating thoracic wounds (Fig. 26). This allows evaluation of blood loss and re-expansion of the lung. The usual approach is via the sixth or seventh interspace in the mid-axillary line. The tube is directed towards the apex of the chest posteriorly, and is immediately placed on suction. The initial blood loss and presence of an air leak are noted so that ongoing drainage can be assessed. Subsequent management is based upon the results of chest drainage.



Fig. 26. Intercostal drainage of a stab wound which transected the internal mammary vessels. Thoracotomy was not required.

When the wound is several hours old the blood is dark in colour. Initial drainage may be large in volume, due to reactive pleural effusion stimulated by blood in the pleural cavity. Continued bleeding and the increasing or decreasing hourly trend is more important than the initial value and determines what further intervention must be undertaken. A persistent air leak associated with bleeding and haemopneumothorax indicates major pulmonary laceration, though in a stable patient this may not require surgical treatment. Damage to a large bronchus usually produces a massive air leak with surgical emphysema (Fig. 27); this does require early intervention. With blood transfusion and intercostal drainage, most pulmonary lacerations resolve. When the chest tube drainage yields less than250 ml/h without a large or continuous air leak, and if diagnostic studies show no major structural damage to the trachea, bronchial tree, oesophagus, or cardiovascular system, chest drainage and supportive measures are sufficient. If the patient's haemodynamics are stable and an arterial bruit, absent pulse, widened mediastinum, or a wound track passing across the mediastinum is observed an aortogram should be performed to assess major vessel injury, bleeding from which may be deceptively controlled by a haematoma under tension. Signs of cardiac tamponade or failure to stabilize the haemodynamics mitigate against this course and prompt early surgical exploration (Table 4). When a missile path is known to traverse the mediastinum and the patient remains stable bronchoscopy and oesophagoscopy are useful before exploration, particularly if mediastinal emphysema is seen on the chest radiograph or subcutaneous emphysema is felt in the neck.



Fig. 27. Stab wound involving the trachea and causing extensive subcutaneous emphysema of the face and neck.

1.	[illegible]
2.	[illegible]
3.	[illegible]
4.	[illegible]
5.	[illegible]
6.	[illegible]
7.	[illegible]

Table 4 Criteria for thoracotomy in open or closed chest injury

Well-defined protocols exist for thoracotomy either on an urgent or delayed basis (Table 4). The most frequent indications for emergency surgery are exsanguinating haemorrhage via the intercostal drain from pulmonary hilar wounds, or tamponade from cardiac laceration. Unless the patient is moribund and thoracotomy is required for resuscitation, it is preferable to maintain a state of controlled hypotension, insert the appropriate lines for transfusion, correct acid–base deficit and move quickly towards the appropriate facilities in the operating theatre.

Penetrating cardiac wounds

Approximately 80 per cent of patients with penetrating cardiac wounds are dead on arrival in the accident department or are taken directly to the coroner. Major intracardiac injuries such as valve disruption, coronary artery transection, or septal defects are therefore rare in patients but more frequent in autopsy series. Autopsy series have a preponderance of gunshot wounds whilst, with few exceptions, clinical papers discuss predominantly stab wounds. This reflects the relative severity of these modes of injury. The triad of a mediastinal entrance wound, clinical evidence of cardiac tamponade, and a pulseless patient with profound systemic hypotension are pathognomonic of important cardiac injury. Though a posterior, subxiphoid, or subcostal penetrating wound may also produce cardiac injury, a mediastinal site of penetration is responsible for 57 to 85 per cent of cases. The diagnosis must be made rapidly from the clinical condition and will seldom await elective diagnostic studiesother than the plain chest radiograph or two-dimensional echocardio-graphy in a few stable and equivocal cases with tamponade. Many patients appear lifeless but have not sustained irreversible brain damage. The first question, therefore, is whether the patient is dead or alive. Those patients who reach hospital alive are a self-selected group who survive through arrest of haemorrhage by cardiac tamponade, or who are conveyed rapidly with bleeding from potentially fatal injuries. Cardiac wounds, consequently, present with one or two distinct syndromes—pericardial tamponade or haemorrhagic shock. The key to successful management of penetrating cardiac injuries is prompt diagnosis and immediate resuscitation and repair (Fig. 28(a), Fig. 28(b)).



Fig. 28. (a) Transfixion of the heart with a screwdriver. (b) Access is imperative before removal.

Haemorrhagic shock

Haemorrhagic shock follows cardiac wounding when the pericardial laceration is large enough to allow free exit of blood so that tamponade cannot occur. Those patients who arrive moribund with haemorrhagic shock should undergo immediate thoracotomy on the side of injury, with the objective of controlling haemorrhage from the cardiac laceration and restoring blood volume and acid–base status. Thoracotomy is performed in the accident department only when the patient arrives warm with reactive pupils but with circulatory arrest, or when acute deterioration, uncontrolled haemorrhage, or cardiac arrest occurs in a patient whose injuries suggest cardiac involvement. In this situation, transfer to an operating theatre is not feasible, since a 5- to 10-min delay is not compatible with survival. The patient is rapidly intubated before thoracotomy and infused with a crystalloid solution such as Ringer's lactate. Unmatched type-specific or non-specific universal donor blood is given as soon as possible. Arterial pH and blood gas measurements are taken early and bicarbonate administered to correct profound acidosis. Anaesthesia is not required for the incision in a moribund patient: survivors will not recall the events.

An anterolateral thoracotomy through the 5th intercostal space is used most frequently. The incision can be taken across the sternum for access to the right atrium and venae cavae or through the costal cartilages on the left side superiorly for involvement of the great vessels. This approach is also useful for patients with blunt trauma or flail or crushed chest who have suffered cardiopulmonary arrest and cannot undergo external cardiac compression safely. Moribund or profoundly hypotensive patients who have suffered blunt or penetrating abdominal trauma can be stabilized by resuscitative thoracotomy and cross-clamping of the descending thoracic aorta. This arrests intra-abdominal haemorrhage, facilitates internal cardiac massage, and allows perfusion of the brain and myocardium in preference to other organs. Patients with ruptured abdominal aortic aneurysms can be revived in the same way. Median sternotomy gives a better access to all cardiac chambers but sternal saws are not readily available in most United Kingdom accident departments.

Surgical management of cardiac laceration depends on the location. Control of haemorrhage can usually be obtained by digital or manual compression, and the wound can be sutured directly or over felt pledgets. A small number of patients who survive laceration of a major coronary artery or valvular or septal disruption may require cardiopulmonary bypass after urgent thoracotomy. Left or right atrial lacerations may be controlled by inserting a Foley catheter and inflating the balloon. This facilitates direct suture repair of the defect. Autotransfusion is a valuable adjunct in patients with a major cardiac laceration, though few of these patients survive to reach hospital. Frequently, the lacerated heart has already arrested or is fibrillating. Control of the penetrating wound must then be carried out in association with cardiac massage and defibrillation.

All patients who respond to treatment are transferred to an operating theatre for formal exploration, and more secure cardiorrhaphy if necessary. Cleansing and closure of the chest are then undertaken in sterile surroundings. Intravenous antibiotics, such as gentamicin, flucloxacillin, and penicillin, are administered postoperatively for 5 days or more. Treatment with steroids, mannitol, calcium channel blockers, or hypothermia may be considered for patients with signs of cerebral hypoxia. Complications of the procedure often relate to the uncontrolled operative circumstances and the need for rapid intervention. Iatrogenic lacerations of the heart, coronary arteries, and lung may contribute to failure. Considerable experience is essential in deciding which patient should be explored in the accident department. The heart is relatively easy to resuscitate. The outcome depends more on the cerebral status.

Reviews of large series of patients undergoing immediate thoracotomy in the United States demonstrate two groups with clear prognostic differences. Patients who suffer circulatory arrest more than 4 min prior to admission to the accident department and who are virtually dead on arrival rarely survive. Those patients who are admitted with some minimal signs of life have a better prognosis. Survival correlates well with neurological status immediately after resuscitation. Those patients who show prompt improvement in neurological status with resuscitation usually have a satisfactory outcome. Young age, a brief period of cerebral hypoxia, and lack of major intracardiac structural injury are good prognostic features. Factors associated with high mortality in penetrating cardiac injuries include shotgun wounds (100 per cent mortality), left ventricular injury (47 percent mortality), unconsciousness at the time of arrival (86 per cent mortality), and the absence of measurable blood pressure on arrival (66 per cent mortality). The mortality associated with knife wounds is usually lower than that associated with gunshot wounds.

Cardiac tamponade

Tamponade occurs when the pericardial laceration is small so that blood accumulates within the sac and clots, thus arresting haemorrhage. Patients presenting with cardiac tamponade are easily recognized by the combination of an appropriate entrance wound, systemic hypotension and tachycardia, and distended neck veins despite haemorrhage (Fig. 24). The majority of these patients remain conscious but are extremely anxious, cold, clammy, and cerebrally obtunded. The cardiac wound is usually small, and bleeding stops when intracardiac pressures fall and intrapericardial pressure and blood clot in the wounds arrests haemorrhage. The distended pericardium usually contains blood clot equivalent to between 500 and 1500 ml of blood. Bleeding of this degree may not lower the systemic pressure and the patient's overall condition may not suggest a cardiac wound.

On arrival in the accident department, large bore central and the peripheral venous cannulae are inserted, though it is important at this stage to avoid transfusion, which may elevate the intracardiac and systemic pressures beyond 100 mmHg. This upsets the homeostatic mechanisms of cardiac tamponade and may cause fatal reactive haemorrhage. Full volume replacement and administration of a muscle relaxant must wait until the tamponaded patient is prepared and draped on the operating table with sternotomy or thoracotomy underway.

For patients whose circulatory status remains good and where there is doubt as to whether the heart has been damaged, plain chest radiographs or two-dimensional echocardiography can be carried out once venous lines have been inserted. In tamponade patients, the cardiac shadow is typically globular shaped. Two-dimensional echocardiography shows a pericardial effusion with compression of the cardiac chambers and abnormal ventricular filling.

Patients with tamponade who survive a prolonged journey to hospital can usually survive the further short transfer to an operating theatre. Most centres who treat large numbers of penetrating cardiac wounds have abandoned pericardiocentesis in favour of early thoracotomy since the pericardial blood has usually clotted. Some advocate surgical transdiaphragmatic pericardotomy through a small subxiphoid incision for diagnosis and relief of suspected tamponade. This may produce temporary clinical improvement until median sternotomy or thoracotomy can be carried out. However, it may also disturb the blood clot in the cardiac wound and precipitate fatal haemorrhage before adequate surgical access is possible. If the knife or wounding implement is still in place it is important not to withdraw this until exposure of the damaged structures has been obtained (Fig. 29).

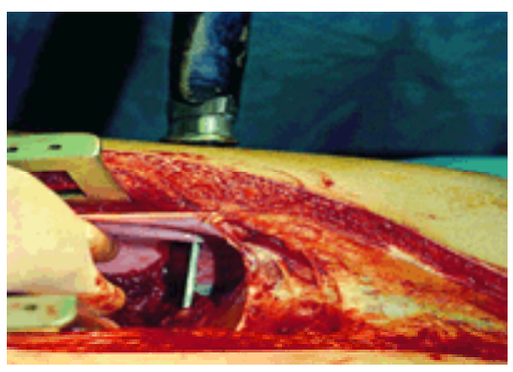


Fig. 29. Knife wound of the chest passing through the pericardium. A near miss.

Surgical intervention

Median sternotomy is the incision of choice for cardiac wounds. This provides access to all aspects of the heart and both pleural cavities, and the pulmonary hila, mediastinal trachea, and upper oesophagus area are also accessible. Penetrating wounds involving the heart can be approached via lateral thoracotomy, the position of which depends upon the site of penetration and predicted track of the weapon. Wounds above the nipple line require an incision through the fifth interspace. Those below, with possible diaphragmatic and abdominal damage, should be approached through the seventh interspace.

Opening of the sternum and evacuation of blood clot is usually followed by fresh bleeding from the site of penetration. Most bleeding areas in the ventricles or great vessels can be controlled with finger pressure followed by suture. It is important to inspect the rest of the heart carefully for an exit wound. Only rarely is cardiopulmonary bypass required at this stage.

Excessive bleeding from the lung can be arrested by a clamp across the hilum. In the profoundly hypotensive patient, bleeding is minimal; once haemorrhage is stopped, blood or crystalloid infusion can be used to restore the intravascular volume and blood pressure. If ventricular fibrillation occurs during thoracotomy, cross-clamping of the descending aorta allows perfusion of the cerebral and coronary vessels while blood volume is restored. Severe pulmonary laceration or contusion may require debridement or resection, whereas cardiac, bronchial, or oesophageal lacerations can usually be repaired by direct suture. Suture of cardiac lacerations should avoid the coronary arteries. If a major coronary artery is divided proximally, a saphenous vein or internal mammary arterial graft may be required.

Damage to intracardiac structures such as the interventricular septum or valves may require open repair at a later stage but cardiopulmonary bypass is rarely required in the acute phase. In difficult situations such as inaccessible hole in the left atrium, balloon tamponade with a Foley catheter can allow time to organize appropriate equipment and personnel.

Gunshot or stab wounds below the nipple line may produce combined thoracoabdominal injuries. The abdominal component may be recognized by preoperative peritoneal lavage or examination of the diaphragm during thoracotomy. Diaphragmatic injury must be repaired carefully since the morbidity associated with strangulated diaphragmatic herniation is considerable.

Blunt cardiac trauma

Any patient who has sustained high-speed deceleration trauma with head, thoracic, or abdominal injuries may also have cardiac contusion. Clinical and autopsy studies report a 15 to 75 per cent incidence of myocardial contusion in patients with blunt chest trauma, and life-threatening cardiac injury can occur with minimal or absent external signs. If cardiac contusion is not excluded by deliberate investigation and treated appropriately if present, the first clinical event may be cardiogenic shock, ventricular fibrillation, or complete heart block. Cardiogenic shock may cause death during surgery or rehabilitation for other injuries.

The spectrum of blunt cardiac injury

High speed deceleration impact of the anterior chest wall, for example against a steering wheel, and severe crush injuries cause acute depression of the sternum and costal cartilages which compress the heart against the vertebral column posteriorly. In children and young adults the chest wall may spring back into position, leaving no external evidence of the severity of injury. Older adults may sustain a transverse fracture of the sternum, anterior costochondral dislocation, or multiple rib fractures, producing an anterior flail segment. These findings guarantee some degree of blunt cardiac injury. Pulmonary contusion, ruptured diaphragm, liver, or spleen, head injury, faciomaxillary trauma, aortic transection, and long bone fractures may coexist.

The range of potential cardiac lesions is shown in Table 5. Pericardial effusion and partial or full thickness myocardial contusion with conduction disturbances or dysrhythmia are common. Valve disruption, cardiac rupture, and ventricular septal defect are rare in patients who reach hospital alive. In order of frequency, traumatic cardiac rupture affects the right ventricle, left ventricle, right atrium, and left atrium. Rupture of the ventricular septum occurs as a delayed event, recognized after acute recovery; the patient develops cardiac failure with a loud systolic murmur. Injuries to the tricuspid, mitral, and aortic valves are usually apparent at an early stage, although papillary muscle ischaemia may lead to late development of tricuspid and mitral regurgitation. Rupture of the cordae tendineae and mitral or tricuspid leaflets is extremely rare. Aortic valve tears usually occur at the base where the cusps attach to the annulus. Acute occlusion of a non-atherosclerotic left anterior descending coronary artery may cause acute myocardial infarction, cardiogenic shock, or late left ventricular aneurysm formation. Right ventricular aneurysm formation has been described following severe right ventricular contusion.

A. Myocardium
1. Contusion
2. Laceration
3. Rupture
4. Septal perforation
5. Aneurysm, pseudoaneurysm
6. Haemopericardium, tamponade
7. Thrombosis, systemic embolism
B. Pericardium
1. Pericarditis
2. Postpericardiotomy syndrome
3. Constrictive pericarditis
4. Pericardial laceration
5. Haemorrhage
6. Cardiac herniation
C. Endocardial structures
1. Rupture of papillary muscle
2. Rupture of chordae tendineae
3. Rupture of atrioventricular and semilunar valves
D. Coronary artery
1. Thrombosis
2. Laceration
3. Fistula

Table 5 Types of cardiac injury from blunt trauma

Although cardiac contusion is usually considered as equivalent to acute myocardial infarction, there are important morphological and clinical differences between the two. Coronary artery injury is rare in patients with cardiac contusion, and coronary flow may be normal or increased in the contused myocardium. Nevertheless, in severe cases, transmural injury causes the same sequelae as infarction. Most patients show limited areas of necrosis restricted to portions of muscle bundles rather than complete coagulation necrosis. Myocardial haemorrhage and oedema produce a decrease in compliance, though the transition from normal to abnormal myocardium is more abrupt than in an acute myocardial infarction. Healing is by patchy scarring rather than fibrosis.

Diagnosis of blunt cardiac injury

Physical signs

The possibility of cardiac trauma should be considered in every patient who sustains a high-velocity deceleration injury. Precordial bruising, and tyre or seatbelt marks reinforce suspicion of cardiac injury. Petechial haemorrhages over the chest and neck, together with subconjunctival haemorrhage, suggest traumatic asphyxia due to severe anteroposterior crush injury (Fig. 30). Cardiac contusion should be assumed to coexist in patients with aortic transection, blunt tracheobronchial injuries, lacerated diaphragm, and ruptured liver or spleen, all of which indicate the severity of trauma. A pericardial friction rub, cardiac murmur, or irregular pulse suggest acute cardiac injury in a patient with no previous history of cardiac disease.



Fig. 30. Traumatic asphyxia with subconjunctival haemorrhage.

Elevated jugular venous pressure prior to volume replacement is worrisome, especially when combined with blood loss from other injuries. Central venous pressure should therefore be measured from the outset. An inappropriately raised venous pressure suggests either cardiac tamponade or severe right ventricular dysfunction due to myocardial contusion or tricuspid regurgitation. It is extremely difficult to assess the severity of myocardial contusion, the spectrum of which ranges from barely perceptible to that sufficient to cause complete myocardial disruption.

The plain chest radiograph

The appearance of the heart shadow on chest radiographs may be entirely normal, but radiological evidence of pulmonary contusion or oedema implies an element of cardiac contusion. More severe cardiac trauma causes the radiograph to show cardiac enlargement or pulmonary venous congestion. Apparent cardiac dilatation may result from a haemopericardium; true dilatation results from an acute valve injury with left ventricular failure. Pneumopericardium may occur after severe pulmonary contusion, when air tracks from the lungs through the sheath around the pulmonary veins or when a lacerated pericardium allows free communication from a pneumothorax to the pericardial cavity. A large amount of air in the pericardium is often associated with rupture of a main bronchus, and tension pneumopericardium has been reported. Pneumopericardium can be distinguished from pneumomediastinum by tipping the patient into the left lateral decubitus position (Fig. 31).

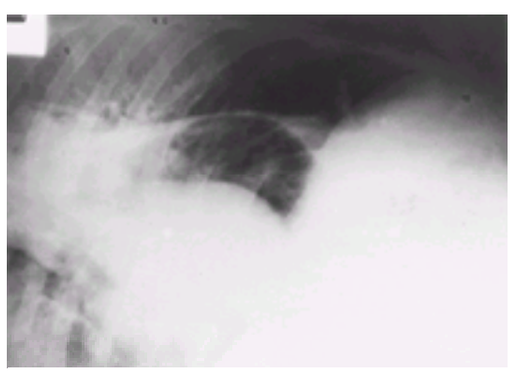


Fig. 31. Pneumopericardium.

The electrocardiogram

Electrocardiographic findings are variable. The ECG may show non-specific ST and T wave changes, sinus tachycardia, supraventricular tachycardia, and conduction abnormalities such as right bundle branch block. Bundle branch block, left anterior hemiblock, or first or second degree heart block may progress to complete heart block and sudden death, with worsening of the oedema around the conduction tissue. The most frequent clinical manifestation of cardiac contusion is ventricular tachycardia, the likelihood of which seems unrelated to the size of contusion or the extent of morphological damage. Electrical instability at sites of unevenly perfused tissue may initiate re-entry responses conducive to tachyarrhythmia. These electrical events often occur in the convalescent period, when no chest or cardiac injury has been suspected.

Quantifying myocardial injury

It is difficult to assess the degree of myocardial injury following trauma in the same way as acute myocardial infarction. In patients who do not have clinically obvious cardiac dilatation, raised filling pressures, pulmonary oedema, or electrocardiographic changes myocardial contusion can be reliably confirmed by two-dimensional

echocardiographic imaging of ventricular wall motion abnormalities, and by the measurement of the myocardial isoenzyme of creatine phosphatase. Echocardiographic demonstration of pericardial effusion and either right or left ventricular dyskinesia indicates serious myocardial damage. Septal akinesia is often associated with electrocardiographic conduction abnormalities and ST segment and T wave changes. A rise in creatine phosphatase isoenzyme levels to more than 6 per cent of total creatine kinase is a sensitive and accurate index of important myocardial injury. Such patients have a high incidence of conduction defects and dysrhythmia on ECG, and of associated pulmonary contusion, which constitutes the best radiological correlate of important myocardial contusion. In the future, troponin-T levels may provide a better quantitative assessment of direct myo-cardial injury.

Treatment of blunt cardiac trauma

Early intervention depends on the extent of associated injuries. Hypoxia and acidosis due to blood loss, haemopneumothorax, or airways obstruction seriously compound a head or cardiac injury and must be corrected at the earliest opportunity. There should be a low threshold for intubation and ventilation, and care must be taken not to overtransfuse the patient. There is an early and consistent depression of cardiac output following myocardial contusion. The damaged heart has less reserve with which to compensate for abnormal haemodynamic states. Meticulous fluid management is therefore necessary to avoid the sequelae of hyper- or hypovolaemia.

With careful treatment most patients with myocardial contusion make a complete functional recovery within a relatively short period (4–6 weeks), with no residual disability. Patients with severe damage, characterized by wall motion abnormalities, raised creatine kinase, and ECG changes, are subject to the same consequences as those suffering acute myocardial infarction, including ventricular septal rupture, left or right ventricular scar formation with dysrhythmia, and ventricular aneurysms. Cardiogenic shock is rarely seen following myocardial contusion alone but occurs in patients with cardiac tamponade, valve disruption, or ventricular septal defect. The thin-walled right ventricle which lies directly behind the sternum may fail acutely. This may contribute to death from chest trauma of apparently moderate severity (particularly when subject to volume overload).

Patients who sustain cardiac injury should be managed similarly to those suffering acute myocardial infarction. Such a history should cause concern during prolonged surgery for associated abdominal trauma or long bone injuries. Although operations for associated injuries are mandatory, they must be undertaken by experienced surgeons and anaesthetists who can minimize the duration of surgery. In particular, dysrhythmia prophylaxis should be discussed beforehand, if necessary with a cardiologist. Bed rest with continuous ECG monitoring, serial creatine kinase measurement, and repeat two-dimensional echocardiography should be employed until the risks of complete heart block, ventricular fibrillation, or delayed cardiac rupture are considered extinct.

In patients with cardiogenic shock, it is important to use two-dimensional echocardiography to diagnose or rule out cardiac tamponade, pericardial laceration with cardiac herniation, or acute laceration of the tricuspid, mitral, or aortic valves. We have undertaken successful immediate thoracotomy in the accident department for pericardial laceration and herniation of the ventricles, and for blunt left ventricular rupture (Fig. 32). Injury to the cardiac valves can usually be managed conservatively until the patient's condition is stabilized. These then require cardiopulmonary bypass for valve replacement or repair.

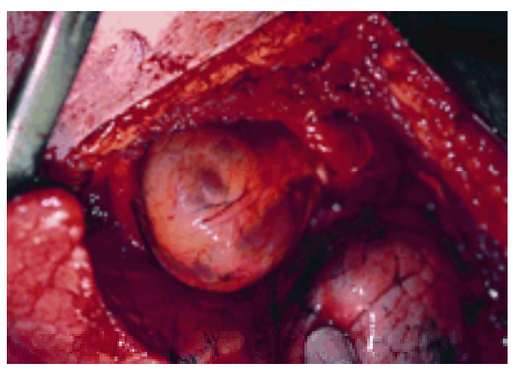


Fig. 32. Immediate thoracotomy for blunt pericardial dehiscence and herniation of the ventricles.

Patients with established cardiogenic shock may require early treatment with inotropic drugs. However, inotropic support in isolation increases cardiac work and oxygen consumption and provides an ideal setting for fatal dysrhythmias. Maintenance of adequate perfusion pressure to uninjured areas of cardiac muscle is vital, and the intra-aortic balloon pump provides an effective means of increasing coronary perfusion and reducing afterload.

If multiple injuries have been sustained, cardiac filling pressure must be optimized by invasive monitoring. Following severe right ventricular contusion, the right atrial pressure may substantially exceed the left atrial or pulmonary arterial wedge pressure. Tricuspid regurgitation due to papillary muscle dysfunction, ruptured cordae, or laceration of the anterior leaflet will result in a characteristic right atrial pressure trace. In cardiac tamponade, both left and right atrial pressures are elevated and equal. Pericardial effusion may compound the effects of myocardial contusion and lead to late deterioration. Post-traumatic pericardial effusions may be reactive and serous or traumatic (frank blood). Serous effusions can be successfully tapped with a needle and, if necessary, an indwelling catheter. Whole blood in the pericardium may clot and require surgical decompression. Blood clot that is not evacuated will eventually liquefy, causing an increase in volume due to osmosis and the risk of progressive cardiac compression. Two-dimensional echocardiography should be repeated daily, and more frequently if haemodynamic deterioration occurs. Delayed haemodynamic deterioration may also follow worsening oedema in the myocardium of the right or left ventricles, thrombosis of the left anterior descending coronary artery with myocardial infarction, increasing papillary muscle dysfunction, or ventricular septal or free wall rupture following haemorrhage into ischaemic myocardium. Cardiac catheterization may then be necessary to check the status of the coronary arteries or the significance of a new murmur.

A lacerated aortic valve requires early repair or replacement. The tricuspid valve can be repaired satisfactorily, even following leaflet laceration (Fig. 33). Care should be taken when assessing a severely regurgitant mitral valve: if the leaflets and cordae are intact the valve will appear normal and the problem lies in an infarcted papillary muscle. Under these circumstances the mitral valve should be replaced. Traumatic ventricular septal defects have been repaired successfully in a manner similar to postmyocardial infarction ventricular septal defect.

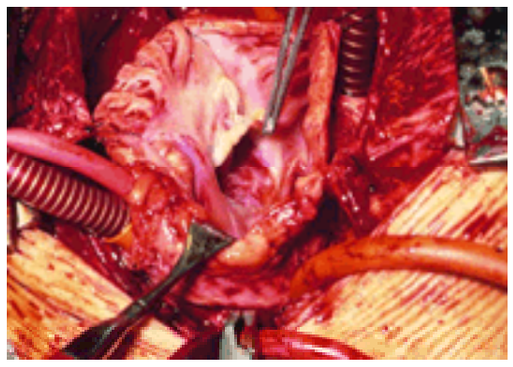


Fig. 33. ‘Bursting' type injury of the tricuspid valve which was successfully repaired.

Experience has shown that patients with normal ECG, creatine kinase, and echocardiographic findings will not deteriorate subsequently. However, all patients with evidence of myocardial injury should be monitored carefully in hospital and subsequently as an outpatient for 2 years.

The ruptured aorta

Sixteen per cent of patients who die in automobile accidents have some form of aortic injury. The most common location of this injury is just distal to the origin of the left subclavian artery and ligamentum arteriosum (Fig. 34). The aortic tear may be partial (usually on the posteromedial aspect of the aorta), complete, or spiral. The exact physical mechanism of the injury is not completely understood but is probably a combination of a shearing effect at the ligamentum attachment caused by

displacement of the aorta during deceleration due to mass–inertia effect and its high density, and profound intra-luminal hypertension during impact.

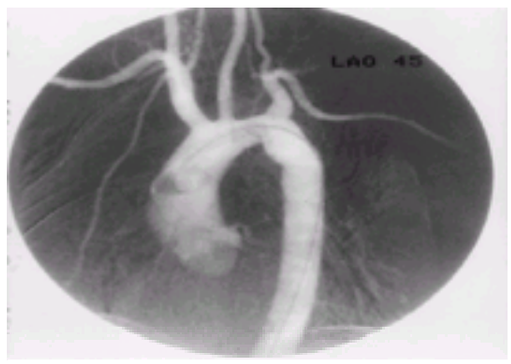


Fig. 34. Aortogram showing blunt injury to the descending thoracic aorta.

Immediate survival depends upon the formation of an acute false aneurysm. The intima and media rupture but the adventitia and adjacent mediastinal structures contain the bleeding. Only 10 to 20 per cent of patients with this type of injury survive to reach hospital.

The classical physical findings in patients with aortic transection are normal or elevated blood pressure in the right arm with absent or low pressure pulses in the legs. Urine flow is absent or decreased. In rare cases where dissection has occurred, there may be asymmetry of the pulses, a pericardial rub, and the diastolic murmur of aortic regurgitation.

In practice the chest radiograph provides the first evidence of aortic injury (Fig. 35). The widened mediastinum has been widely discussed in this context, though several other signs are also of importance. Marsh and Sturm highlight six radiographic findings strongly predictive of aortic rupture:

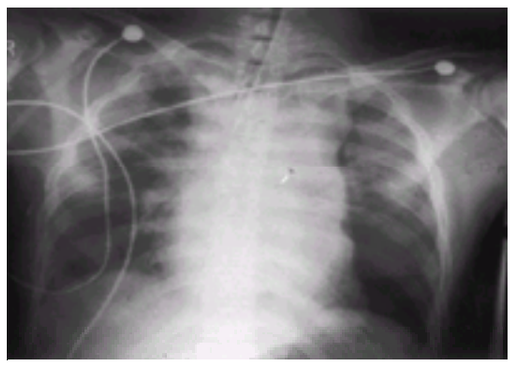


Fig. 35. The plain chest radiograph in blunt aortic rupture.

1. widening of the superior mediastinum to 8 cm on the 100-cm anteroposterior supine radiograph;
2. tracheal shift to the right;
3. blurring of the aortic outline;
4. obliteration of the medial aspect of the left upper lobe;
5. opacification of the clear space between the aorta and left pulmonary artery;
6. depression of the left main bronchus below a 40° angle with the trachea.

Although fracture of the first and second ribs is a hallmark of severe trauma, patients with this injury are no more likely to have a damaged aorta than patients with other rib fractures or no bony injury at all.

The natural history of aortic rupture is immediate death in 82 per cent of patients; death within 24 h in a further 6 per cent, within 2 to 7 days in 6 per cent, and during the second week in 4 per cent. Chronic traumatic aneurysm persists in 2 per cent of patients although stability is doubtful; delayed deaths occur without surgical intervention. Primary repair is indicated in all patients unless the extent of associated injuries clearly mitigates against survival.

Surgery for aortic transection

The patient should be taken directly from the aortogram to the operating room, though in some cases laparotomy for exsanguinating intra-abdominal haemorrhage or craniotomy for a blown pupil may be needed first.

During preparation and transfer to the operating theatre, care is taken to keep the blood pressure below 100 mmHg. Use of nitroprusside is avoided due to its adverse effects on the spinal cord. When transection of the descending thoracic aorta is confirmed, the patient is positioned for left thoracotomy with the left groin exposed to allow access to the femoral vessels. Surgical access is obtained through the left fourth intercostal space and dissection is performed to gain exposure for cross-clamping of the aortic arch between the carotid and left subclavian arteries: the left subclavian artery proximal to the vertebral origin, and the descending thoracic aorta distal to the transection. Only after clamps have been placed is the pleura over the transection opened and the injury examined. The tear may be amenable to repair by direct suture. More often a woven Dacron tube graft is inserted by continuous Prolene suture anastomosis (Fig. 36).

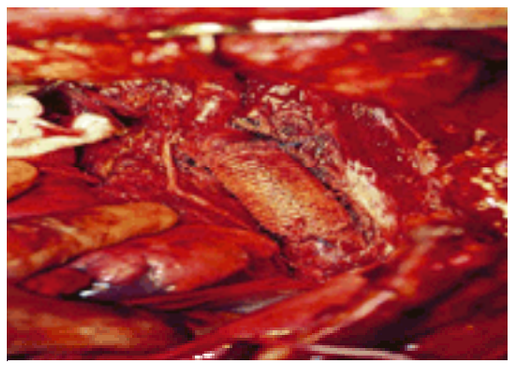


Fig. 36. Dacron tube graft repair of transected aorta.

Heparin-bonded shunts, partial left heart bypass, and femoral vein to femoral artery bypass are used by some surgeons for treatment of both acute transection and traumatic aneurysms. These are used to convey blood from the aortic arch or apex of the left ventricle to the descending aorta, thereby bypassing the injured area. However, the need for bypass procedures is questionable. Cardiopulmonary bypass requires systemic heparinization, which is clearly dangerous in patients with head

injuries and multiple fractures.

With the exception of tears in the aortic arch or ascending aorta, surgery of this lesion can be accomplished safely and expeditiously without the use of shunts or extracorporeal bypass. Bypass techniques have not been shown to decrease the incidence of paraplegia (5 to 7 per cent) or renal failure compared with simple cross-clamping of the thoracic aorta. Interruption of proximal intercostal branches associated with the aortic transection itself may produce sufficient spinal cord ischaemia to cause paraplegia before surgery is undertaken. Paraplegia results from unfavourable spinal vascular anatomy and has been reported with as little as 15 min cross-clamping. Some 20 to 40 min cross-clamping is usually required for graft anastomoses and up to 110 min has been reported without paraplegia.

Postoperatively, these patients should be treated as if it were certain that progressive post-traumatic respiratory insufficiency will develop. Blood gases should be monitored frequently and appropriate changes in respiratory and metabolic therapy should be made rapidly and aggressively. Microfiltration of transfused blood should be routine. Since cardiac contusion frequently accompanies this injury, serial electrocardiograms should be evaluated. Unexplained cardiovascular instability may be the hallmark of severe cardiac contusion.

Blunt injury to the major airways

All levels of the trachea or main bronchi may be involved. More than 80 per cent of injuries occur within 2.5 cm of the carina, equally distributed between right and left sides. Mainstem bronchi are injured in 80 per cent of patients and the distal bronchi in only 9 per cent. The type of lesion varies from simple linear mucosal lacerations to extensive full thickness tears involving the trachea, main bronchi, and branch bronchi. Following complete transverse rupture and separation of the trachea, some continuity is usually maintained by the intact elastic mucosa or loose peribronchial tissue. In some patients the bronchial cartilage and muscle fractures, leaving an unsupported mucosal tube with a ‘flap valve' effect.

Pathogenic features of airways injury are free air in the soft tissues, signs of airways obstruction, and haemoptysis (Table 6). When the chest is crushed, petechial haemorrhage of the upper chest and face may develop. This traumatic asphyxia syndrome is caused by retrograde venous flow from the right side of the heart.

High index of suspicion from mechanism of injury
Free air
Subcutaneous or deep cervical emphysema
Pneumothorax
Pneumomediastinum
Pneumopericardium
Haemoptysis
Airways obstruction
Stridor
Aphonia
Difficult intubation
Presence of common associated injuries:
Sternal or first rib fracture
Rupture aorta

Table 6 Physical signs suggesting the presence of major airways injury

Patients with intrathoracic tracheal or bronchial disruption show distinct clinical patterns, depending on whether or not there is free communication between the site of disruption and the pleural cavity (Fig. 37). In the first group (70 per cent) the damaged bronchus opens into the pleural cavity causing a large (often tension) pneumothorax which continues to leak air after insertion of an intercostal drain. As a rule the lung fails to re-expand. The usual signs of injury are dyspnoea, haemoptysis, subcutaneous and mediastinal emphysema, and, in severe cases, cyanosis. Pericardial laceration in proximity to a bronchial rupture causes pneumopericardium.



Fig. 37. Distribution of free air in traumatic bronchial rupture. Combination of right pneumothorax, and mediastinal and subcutaneous emphysema.

In the second group rupture occurs proximal to the pleural sheath and there is little or no communication between the site of injury and the pleural cavity, even when disruption is complete. There is usually no pneumothorax; if one is present it is small and does not recur after chest drainage. Small pleural lacerations become sealed by fibrin or blood clot, allowing the lung to re-expand providing bronchial continuity is not immediately lost. Air generally escapes into the mediastinum and tracks under the deep cervical fascia. Mediastinal emphysema may be insufficient to be clinically detectable, or may be massive. Positive pressure ventilation worsens the subcutaneous emphysema or pneumothorax, and complete respiratory obstruction may occur.

Diagnostic techniques

Chest and neck radiographs are essential since they may disclose free air not detectable on physical examination. Pneumomediastinum is an early diagnostic sign but may easily be overlooked because of the technical quality of the film. The deep cervical fascia is in direct continuity with the mediastinum and air almost invariably tracks into the cervical region. Since this area is easily penetrated by X-rays and is not obscured by overlying soft tissue structures, deep cervical emphysema is easily recognized on lateral radiographs of the neck. Other important radiological signs include hyoid bone elevation (indicating tracheal transection), pneumothorax, subcutaneous air, pneumopericardium, fractures limited to the upper rib cage or ster-num, air surrounding the bronchus, and obstruction in the course of an air-filled bronchus. Air trapping distal to a ‘flap valve' causes over-distension of the ipsilateral lung with mediastinal shift to the opposite side or herniation of the hyperinflated lung across the anterior mediastinum. When a main bronchus is transected within its pleural sheath, the characteristic radiographic appearance is of the affected lung dropped down onto the diaphragm (Fig. 38). This picture contrasts with the findings in patients with pneumothorax without transection, in whom the lung collapses towards the mediastinum.



Fig. 38. ‘Fallen lung' in complete transection of the left main bronchus.

Definitive diagnosis depends on bronchoscopy. This is best performed in a thoracic surgical unit with experienced anaesthetic support and facilities for complex intubation, thoracotomy, and respiratory support. However, it is often better for a thoracic surgeon to take his equipment to a referring hospital rather than risk interhospital transfer for bronchoscopy in circumstances where complete respiratory obstruction may occur en route.

Bronchoscopy is performed only when initial resuscitative measures such as restoration of circulatory blood volume, correction of acid–base balance, and stabilization of the mechanics of breathing (e.g. insertion of an intercostal drain) have been carried out. An exception to this rule occurs when bronchoscopy is required to establish an airway after intrathoracic tracheal transection. A rigid Negus or Stortz instrument is preferable since bronchoscopy must also be used to remove blood clot, broken teeth, and secretions.

Treatment

The site, nature, and extent of injury must be carefully defined. The airways may be full of blood clot and mucus. Torn bronchial mucosa and oedema may obscure the true extent of the injury. Care must be taken not to displace the ends of a transected trachea or bronchus if a satisfactory airway exists. Details of surgical techniques are beyond the scope of this chapter. A successful outcome depends on prompt diagnosis, procurement of a reliable airway, stabilization of the cardiovascular status and the mechanics of breathing, correction of acid–base balance, and early primary repair.

Conservative treatment is permissible only in certain well-defined circumstances. The first is when a longitudinal tear involves only a short length of posterior tracheal membrane, following thoracic compression against a closed glottis. This ‘bursting' injury self-seals when the intratracheal pressure is normal. ‘Minitracheostomy' can be used to maintain a low intratracheal pressure and discourage air leakage into the mediastinum whilst the mucosal laceration heals spontaneously (Fig. 39). The second is when bronchoscopy shows the tear to be less than one-third the circumference of the bronchus and chest tube drainage re-expands the lung completely with early cessation of air leak. Repair is then usually unnecessary.



Fig. 39. Minitracheostomy.

When the acute injury is undiagnosed, and the patient survives, symptoms of respiratory distress usually follow in 7 to 10 days. These are caused by an ingrowth of granulation tissue, displacement of the injured segment, oedema, and surrounding haematoma. Untreated partial rupture of a mainstem bronchus leads to bronchopulmonary suppuration, atelectasis, and fibrosis. Reconstructive surgery with excision of the granulation tissue is indicated at this stage, otherwise stenosis will develop at the site. In complete bronchial rupture the separated lung is atelectatic and uninfected, and is often capable of circulatory and ventilatory function after being re-expanded. Normal pulmonary function has been observed following reanastomosis 2 months after the initial injury. However, the chances of performing successful, delayed primary repair for the missed bronchial injury complicated by infection, are poor and pneumonectomy is often required. Bronchial stenosis has a similarly poor outlook unless the segment is short and resection with reconstruction can be performed.

The lacerated diaphragm

Considerable compression forces are required to rupture the diaphragm. Such forces are usually encountered during deceleration road traffic accidents, when associated visceral and musculoskeletal injuries are the rule. In the absence of respiratory distress, delay in diagnosis is almost invariable.

In one large series, the injury remained undiagnosed in 94 per cent of patients, the average interval until recognition being 4 to 5 years. Positive pressure ventilation, often undertaken on an urgent basis before the chest radiograph, masks respiratory distress and signs such as bowel sounds in the left chest. This may also replace some of the herniated contents back into the abdomen and prevent gastric distension.

Rupture is said to be more common on the left side than on the right, which is protected by the liver. This has not been our experience, but right-sided rupture is diagnosed less often. Bilateral rupture is very rare, but is seen, even in the absence of visceral injury (Fig. 40). Diagnosis of left-sided rupture usually follows radiological detection of bowel in the left chest. On the right side, appearances are those of a raised right hemidiaphragm (Fig. 41); other non-invasive means of diagnosis such as CT or ultrasound scanning are notoriously unreliable. It is easy to overlook diaphragmatic rupture during laparotomy for hepatic or splenic rupture.

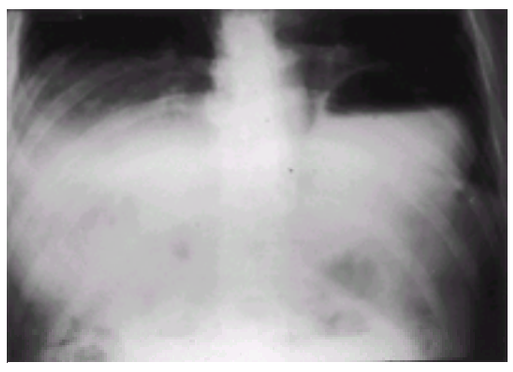


Fig. 40. Bilateral traumatic rupture of the diaphragm. There were no other visceral injuries.

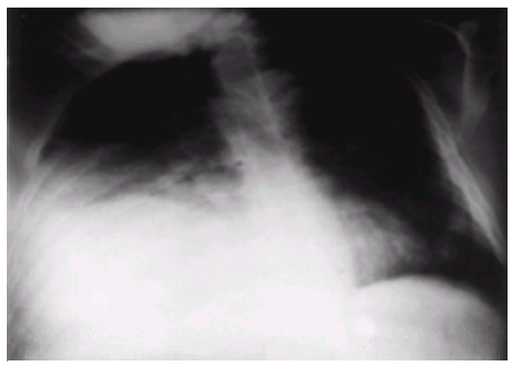


Fig. 41. Ruptured right hemidiaphragm.

The surgical approach depends on the stage of recognition and the presence of associated injuries. If the tear is recognized early in a patient with intra-abdominal injuries, the transabdominal approach is acceptable, although access to the right hemidiaphragm is difficult. Epstein and Lempke reported 36 cases of right-sided rupture in which the diagnosis was made only six times preoperatively. They concluded that an approach through the right chest wall was necessary for accurate and safe repair. This also applies for left-sided rupture diagnosed late or when the diagnosis is made early and no intra-abdominal injury is suspected. Repair is undertaken with a two-layer closure since dehiscence of the suture line occurs frequently. If this cannot be performed without tension, particularly after delayed diagnosis, Marlex mesh or a pericardial flap are employed.

Treatment of chest wall injuries

In most patients with thoracic trauma, injuries are limited to the thoracic cage, with or without underlying pulmonary contusion, pneumothorax, or haemothorax. The extent of chest wall derangement varies considerably. The most extensive disruption occurs in patients with severe crush injuries, in whom multiple bilateral rib fractures and fractures of the spine and sternum coexist.

A flail segment moves inwards on inspiration (paradoxical movement) and consequently compromises ventilation by reducing tidal flow. Physical examination is more valuable than radiology in defining the nature and extent of such an injury. Both pleural cavities act as single bellows. If the bellows are damaged and cannot produce sufficient negative pressure in the presence of a large flail segment, then to and fro movement of the segment may equal the attempted tidal volume of that side of the chest and ventilation is seriously compromised.

It is important to restore full expansion of the lungs as soon as possible by drainage of a haemopneumothorax. The two almost always coexist to some extent after trauma, and the size and position of the intercostal drain is important. This must be sufficiently large to drain blood and some fresh blood clot. A size 32 to 36 F Argyle tube is appropriate and is usually inserted laterally between the anterior and mid-axillary lines. The intercostal space and precise site chosen depends upon the site of injury and the location of blood or air. It is important not to enter the bony thorax too low or penetration of the diaphragm and abdominal viscera may occur: the fifth to seventh interspaces laterally at the anterior axillary line are usually safe. It is unnecessary for the drain to be in the ‘basal' position to drain blood, or in the apical position for air. A safe insertion is the first consideration, especially when radiological landmarks are obscured by a haemothorax and the chest wall itself is bruised or deformed. Liquid, blood, and air will ‘find' the drain as the lung expands. Established blood clots will not drain and an extensive clotted haemothorax requires surgical evacuation. It is important to incise the chest wall through to the pleura with a scalpel and to complete the entry with scissors or an artery forcep. Blood or air will then drain through the entry wound and the large drain can be inserted and directed to its required position without difficulty. Blood loss is measured carefully, since the need for thoracotomy is determined on the basis of the initial volume lost and the continued rate of bleeding. A guide for this is given in Table 4. There is no contradiction against insertion of the drain through the area of injury. Three commonly performed manoeuvres are inadvisable.

These are, first, insertion of a chest drain ‘prophylactically' after injury on the pretext of preventing pneumothorax during positive pressure ventilation. If there is no pneumo- or haemothorax the underlying lung may be damaged by the procedure, thus creating an air leak or bleeding. This applies particularly to the 15 per cent of patients whose pleural cavity has been obliterated by fibrous adhesions from previous ‘pleurisy'. In the presence of a pneumothorax, a drain should always be inserted prior to positive pressure ventilation.

Second, insertion of a chest drain anteriorly in the second intercostal space. This causes pain and transfixes the principal accessory respiratory muscle, the pectoralis major. If an apical drain is specifically required then the posterior suprascapular route to the second interspace can be used.

Third, clamping of a chest drain at any time in the presence of an airleak. This produces serious risk of tension pneumothorax. The alternative, without clamping, is disconnection of the drain from its underwater seal. This merely produces a simple pneumothorax which is rapidly resolved by reconnection.

When the chest drain has served its purpose, it should be removed directly. It will otherwise act as a conduit for bacterial infection and hamper chest wall movement and mobility during physiotherapy. Further management is aimed not at the chest wall itself but at preservation of respiratory function.

Two or three simple fractures may lead to the death of a patient with chronic bronchitis: pain causes decreased respiratory excursion and failure to ventilate the basal segments. Atelectasis supervenes. Pain may also inhibit cough, so that secretions are not cleared from the bronchial tree. Pneumonia follows and profuse retained secretions cause respiratory obstruction. The patient then rapidly deteriorates into acute respiratory failure.

Pain treatment

Pain relief is of paramount importance from the outset. This allows the most affective therapeutic option, physiotherapy, to be carried out frequently and with the full co-operation of the patient. A few stable rib fractures are usually manageable with potent antipyretic oral or intravenous opiate analgesia. Patients with multiple fractures, or those with instability, require specialized attention.

Intercostal or paravertebral nerve blocks provide temporary relief for some patients with unilateral rib fractures. Nevertheless, the need for multiple injections is a distinct disadvantage. Thoracic epidural anaesthesia is an effective alternative that has radically improved the overall management of patients with chest wall trauma. Recent studies show a decreased mortality in patients with extensive rib fractures following epidural anaesthesia without ventilatory support, compared to those treated without routine mechanical ventilation. This applies equally to those who have undergone surgery for concomitant abdominal or orthopaedic injuries, those with flail segments, and those with pulmonary or cardiac contusion as long as arterial po_2 exceeds 50 mmHg (with 50 per cent inspired oxygen) and vital capacity exceeds 10 ml/kg.

Analgesia is planned to provide a sensory block of the dermatome area of up to, but no higher than, the level of T4. Only if the arterial po_2 falls below 50 mmHg on supplemental oxygen or if there is an increase in the respiratory rate to more than 40/min with an inability to cough, is intubation and ventilation considered for further management. In borderline patients, clearance of secretions by minitracheostomy (Portex) may tip the balance in favour of continued conservative management. There are potential complications attached to the thoracic epidural anaesthesia, however. The procedure should be carried out by experts and managed in the intensive care unit. Patients are excluded if they have spinal damage, pre-existing chronic neurological disease, skin infection overlying the area for insertion of the epidural catheter, or are unconscious, unco-operative, or unwilling to undergo a spinal anaesthetic procedure. Autonomic nerve block may produce hypotension, and motor blockade may decrease maximum ventilatory capacity in a few patients. However, in one excellent study, the requirement for mechanical ventilation in patients with ‘severe' chest wall injury was reduced from 100 per cent to 9 per cent. Pain relief with epidural anaesthesia was dramatic and sedation was not required, thereby providing an alert, co-operative patient who was able to comply with physiotherapy demands.

Mechanical ventilation

While a conservative approach should be used in the first instance, mechanical ventilation has an important role to play in managing patients who are unconscious and unco-operative, those in severe respiratory failure, or with worsening respiratory failure despite adequate analgesia and clearance of secretions by physiotherapy and minitracheostomy.

The disadvantages of prolonged intermittent positive pressure ventilation include the need for prolonged intubation or formal tracheostomy (with the risks of accidents), prolonged sedation, inadequate nutrition, and immobility with the risk of respiratory infections and pressure sores. Effective physiotherapy is more difficult and repeated passage of suction catheters causes trauma to the major airways. Patients with chronic obstructive airways disease and those with severe pulmonary contusion are more likely to require mechanical ventilation. A flail segment in itself is not an indication for mechanical ventilation: it is he functional rather than the anatomical consequence of injury which determine the necessity for ventilatory support.

Surgical treatment of chest wall injury

This is only rarely undertaken for fracture fixation and stabilization of the bony chest wall, or correction of serious deformity with functional sequelae. Chest wall fixation may also be performed during evacuation of an extensive clotted haemothorax or for haemostasis in a severely lacerated lung.

The vogue for nail or clip fixation of fractured ribs in flail chest waned with widespread use of positive pressure ventilation for these patients. However, it has recently

been resurrected, particularly in the United States, with the trend towards ‘conservative', non-ventilatory management. Provided that rib fractures are not too comminuted, internal fixation may eliminate the need for artificial ventilation in a few patients and lessen the risks of pulmonary infection.

When a clotted haemothorax is left *in situ* the long-term complications are debilitating. The haematoma deposits fibrin on the chest wall and lung and fibrotic contraction eventually causes crowding of the ribs, with severe restriction of lung compliance and expansion. There is also a significant risk of infection and emphysema. Thoracotomy should therefore be performed early when intercostal drainage fails to evacuate a substantial haemothorax. Respiratory function and recovery rate are improved by this procedure. Surgical morbidityand mortality are negligible when the patient is haemodynamically stable beforehand.

The sternum

Substantial force is required to fracture the sternum. These fractures are usually transverse and may overlap, causing agonizing pain. Internal fixation is required to correct deformity and relieve the discomfort of instability. This is performed by making an incision in the xiphisternum and passing a Steinmann pin up through the medulla of both sternum fragments from underneath. A separate incision over the fracture itself aids alignment of the bone ends.

Similarly, stabilization of an anterior flail segment can be performed by passing a flat metal plate behind the sternum with its ends resting on the ribs lateral to the fracture. Both of the above procedures can be performed to great effect through small skin incisions. This is in contraindication to the extensive dissection required for multiple rib fixation.

Further reading

Blegvad S, Lippeot H, Lund OK, Kromann Hansen O, Christensen T. Acute or delayed surgical treatment of traumatic rupture of the descending aorta. *Journal of Cardiovascular Surgery* 1989; **30**: 559–64.

Evans J, Gray LA, Rayner A. Principles for the management of penetrating cardiac wounds. *Annals of Surgery* 1979; **189**: 777–84.

Fernandez G, Fontan F, Deville G, Madonna F, Thibaud D. Long term evaluation of direct repair of traumatic isthmic aortic transection. *European Journal of Cardiothoracic Surgery* 1989; **3**: 327–34.

Ivatury R, Shah PM, Katuski I. Emergency room thoracotomy for resuscitation of patients with ‘fatal' penetrating injuries of the heart. *Annals of Thoracic Surgery* 1981; **32**: 377–85.

Millikan JS, Moore EE. Outcome of resuscitative thoracotomy and descending aortic occlusion performed in the operating room. *Journal of Trauma* 1984; **24**: 387–90.

Newman RJ, Jones IS. A prospective study of 413 consecutive car occupants with chest injuries. *Journal of Trauma* 1984; **24**: 129–35.

Shackford SR, Virgilio RW, Peters RM. Selective use of ventilator therapy in flail chest injury. *Journal of Thoracic Cardiovascular Surgery* 1988; **81**: 194–201.

Snow N, Richardson JD, Flint LM. Myocardial contusion: Implications for patients with multiple traumatic injuries. *Surgery* 1982; **92**: 744–50.

Spencer JA, Rogers CE, Westaby S. Clinico-radiological correlates in rupture of the major airways. *British Journal of Radiology* 1991; **43**: 371–6.

Sturm JT, Marsh DG, Kenton CB. Ruptured thoracic aorta: evolving radiologic concepts. *Surgery* 1979; **85**: 363–7.

Westaby S. Injury to the major airways. *British Journal of Hospital Medicine* 1985; **34**: 210–20.

Westaby S. Blunt injuries to the chest. In: Westaby S, ed. *Trauma: pathogenesis and treatment*. Heinemann Medical Books, London, 1989, pp. 118–41.

Westaby S. Penetrating wounds of the chest and abdomen. In: Westaby S, ed. *Trauma: pathogenesis and treatment*. Heinemann Medical Books, London, 1989, pp. 159–80.

Westaby S, Kobayashi K. Shock and blood volume replacement. In: Westaby S, ed. *Trauma: pathogenesis and treatment*. Heinemann Medical Books, London, 1989, pp. 42–55.

Westaby S. The injured heart. *Clinical Intensive Care* 1990; **5**: 210–9.

42.1 General care of the paediatric patient, especially the newborn infant

Andrew R. Wilkinson

Introduction
Admission procedure
General care of the newborn
Growth in relation to gestational age
Surfactant deficiency
Temperature control
Apnoea
Jaundice
Feeding, fluids, and electrolytes
Infection
Periventricular haemorrhage and ischaemia
Polycythaemia
General care of older children
Clinical assessment
Further reading

Introduction

The last decade of the twentieth century saw further improvements in the prognosis for children who undergo surgical procedures. Added to the advancements made in the 1970s and 1980s, the overall outlook has been transformed although there are still many challenges to be faced. This is a reflection of continued improvement in our understanding of the physiology of newborn babies and of children and an awareness of how their needs differ from adults. Advances in anaesthesia, surgical procedures, and postoperative care, as well as better pharmacological agents, have all made it possible to perform complex surgery successfully. The differences between children and adults are most marked immediately after birth, when the infant is adapting to extrauterine life. However, such differences exist throughout childhood and must be clearly understood and acted on if a satisfactory outcome is to be achieved.

All staff involved in the care of ill children should have specialized training and appropriate experience. A child-orientated team approach should be developed which should include not only expert nurses, paediatric surgeons, and paediatricians, but pharmacists, laboratory staff, radiographers, dieticians, physiotherapists, and social workers, who should all understand the important contribution that they can make.

As well as having unique physiological characteristics, sick children have special psychological needs which the hospital environment must take into account. The family must be involved in the care of their child: staff must, therefore, be able to communicate with parents, siblings, and others who are close to the patient. The physical environment must provide space for private consultation with parents as well as sleeping accommodation and bathrooms. There must be arrangements for eating and access to a telephone. Some parents will need financial assistance with the costs incurred in travelling from home to hospital.

Admission procedure

Irrespective of the condition that leads to a child's admission to hospital, a number of important principles must be understood if hazards are to be avoided. Critically ill children are not ‘small adults': unnecessary mistakes in management can occur if all the differences are not appreciated. Surgical procedures undertaken in the first few days after birth carry particularly high risks; the complex process of adaptation to extrauterine life must not be adversely affected.

On admission, the child should undergo careful assessment from a medical and nursing perspective. The details of the presenting condition should be documented in full. Photographs taken with a Polaroid camera can be useful, particularly if a condition is developing rapidly and if the patient is being transferred to another hospital. Information concerning the maternal obstetric history and place of birth must not be overlooked, even in an older child. Perinatal events and the state of the baby at birth must be recorded: evidence of birth asphyxia, possible aspiration, and the need for resuscitation should be noted. The family and social history should also be obtained immediately: these have implications for the impact of the child's illness.

Body temperature, heart rate, respiratory rate, and blood pressure must be recorded. Haematological and biochemical indices should be measured routinely and the urine analysed. Relevant specimens, including blood, should be sent for microbiological culture. Accurate measurement of body weight is of vital importance. Fluid therapy is adjusted precisely depending on weight; drug dosages are also related to weight as well as to age.

General care of the newborn

Growth in relation to gestational age

Neonates are classified by their birthweight and by the length of gestation. The majority of babies are appropriately grown for their gestational age. If the birthweight lies below the tenth centile for gestational age the baby is classified as small for gestational age, and if greater than the 90th centile, as large for gestational age. Centile charts derived from local cohorts of babies should be used wherever possible, because of ethnic differences in birthweight.

Preterm babies are those born before 37 completed weeks of gestation, term babies are born at 37 to 42 weeks, and post-term babies are born after 42 weeks. Babies who suffer intrauterine growth retardation are often born to mothers with complicated pregnancy. Pre-eclampsia toxaemia, smoking, and drug addiction are the most common factors associated with intrauterine growth retardation in ‘developed' countries, while poor nutrition and infection are more common in other parts of the world. The clinical problems seen in babies in these categories differ and those that commonly occurin low birthweight babies are listed in Table 1.

Preterm (<37 weeks)	*Small for gestational age (<10%)*
Surfactant deficiency	Intrauterine death
Impaired temperature control	Congenital malformations
Apnoeic attacks	Birth asphyxia
Jaundice	Meconium aspiration
Immature sucking and swallowing	Persistent pulmonary hypertension
Hypocalcaemia	Impaired temperature control
Functional intestinal ileus	Hypoglycaemia
Infection	Polycythaemia
Periventricular haemorrhage/ischaemia	
Necrotizing enterocolitis	
Patent ductus arteriosus etc.	

Table 1 Common clinical problems in low birthweight babies

Preterm babies have conditions which reflect the immaturity of most systems of the body. A baby may be delivered preterm because of fetal and maternal illness and be of low birthweight (<2500 g), very low birthweight (<1500 g), or extremely low birthweight (<1000 g). The preterm baby must be supported to allow the process of maturation to take place after delivery. Clearly, a baby may be born growth retarded as well as preterm.

The surgery planned for the management of congenital malformation soon after birth must take into account these problems. The approach to the patient by surgeon, neonatologist, anaethetist, and nurses should ensure that an operation is carried out at the optimal time.

Surfactant deficiency

Surfactant, a complex mixture of phospholipids and specific proteins, is released by type II cells in the lining of the alveoli. This substance reduces the surface forces in the terminal bronchioles and alveoli and prevents airway collapse. A deficiency of surfactant due to immaturity of the lung gives rise to the respiratory distress syndrome. The clinical signs of tachypnoea (<60 breaths/min), intercostal or sternal retraction, expiratory grunting, and central cyanosis in air may be present soon after birth in most babies born at less than 32 weeks gestation; these may require tracheal intubation and mechanical ventilation. Artificial surfactant preparations are now available; they are either synthetic or natural (derived from animal surfactant). Meta-anaysis of the results of large, randomized, controlled trials show the benefits of giving artificial surfactant to all babies less than 32 week gestation as soon as they are intubated. This may be after resuscitation or when artificial ventilation is required. These studies show that overall mortality, pneumothorax, and pulmonary intestitial emphysema are reduced by about 40 per cent. The animal-derived products, which contain some of the specialized surfactant proteins, have a more rapid onset of action and randomized, controlled trials of synthetic and natural surfactants suggest that mortality and morbidity (particularly retinopathy of prematurity) are reduced by a factor of one in every 45 babies treated. In babies born between 30 and 34 weeks' gestation the signs of respiratory distress syndrome may take a few hours to develop; these babies must be observed carefully.

The severity of any respiratory disease and the underlying cause must be taken into account when planning the timing of surgery soon after birth. Ideally, surgery should be delayed for as long as is necessary for the signs of respiratory distress to resolve. All newborn infants, but particularly those who have suffered a period of hypoxaemia, are at risk of reverting to a fetal state in which high pulmonary vascular pressure leads to right-to-left shunting of blood. This occurs within the lungs and through the foramen ovale and ductus arteriosus. This may herald a very unstable state in which further episodes of severe hypoxaemia may occur.

Temperature control

Babies lose body heat very quickly unless nursed in an environment with an optimal temperature at which oxygen consumption is minimal. In a cold environment, catecholamine secretion is stimulated through hypothalamic receptors. In an attempt to maintain body temperature brown fat stores, predominantly alongside the scapula and aorta, are broken down by oxidative phosphorylation to release heat. This process of non-shivering thermogenesis, which increases oxygen consumption, is limited during the first 12 h after birth and may be impaired after asphyxia or hypoxia and maternal sedation; it may be non-existent in the most immature infants. Ideally, therefore, babies should be nursed in conditions in which oxygen consumption is kept to a minimum while the core (rectal) temperature is maintained at 37°C (within the range 36.7–37.3°C) and skin temperature at a range of 36.2 to 36.8°C. Infants with cardiorespiratory disease, or impaired nutrition as a result of gastrointestinal disorders, and preterm infants with respiratory disease are at greatest risk of cold stress. Despite the existence of compensatory mechanisms, heat loss occurs rapidly because of the large surface area to body weight ratio and inadequate fat stores. The four different routes through which heat loss occurs and the approaches that can be taken to minimize this risk are shown in Table 2.

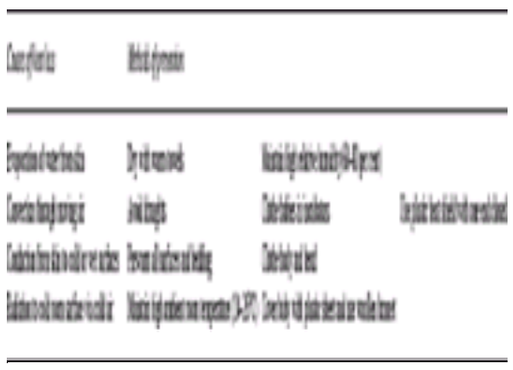


Table 2 The mechanisms of heat loss in new-born infants and standard procedures to conserve heat

The skin must be dried with warmed towels and the baby well covered. If body temperature is normal, further losses by radiation can be minimized by an aluminium foil wrap over the blanket. This is not a substitute for warming the baby first if heat loss has already occurred, as can happen during resuscitation, but may prevent further losses while the baby is being moved to the nursery. An infant undergoing an operation is subjected to additional thermal stresses due to starvation and sedation which blunt the physiological response to cold stress. Before transport from the intensive care nursery the baby should be well wrapped and insulated by cotton wool padding. Increasing the operation theatre temperature to 28°C, avoiding draughts, and the use of a warming pad below the body can minimize heat loss. The part of the body exposed should be kept to a minimum, and intravenous fluids and blood products should be allowed to warm to room temperature before use.

Apnoea

Immaturity of respiratory control is particularly common in preterm babies, but apnoea can occur in any sick infant. Respiration must be monitored and heart rate and oxygen tension should be measured continuously in severely ill children. Apnoea may reflect other problems, such as hypothermia, sepsis, or deranged biochemical states, particularly hypoglycaemia. After these abnormalities have been investigated and treated, and if attacks do not respond to tactile stimulation, caffeine citrate (10 mg/kg intravenously repeated after 12 h and then followed by 10 mg/kg per day intravenously or orally) should be given. Artificial ventilation, first by continuous positive airway pressure and finally by intermittent positive pressure, may be necessary. Immature respiratory control may be responsible for a slower than expected postoperative recovery and should be considered when other complications have been excluded.

Jaundice

After birth bilirubin is excreted in bile in the water-soluble conjugated form. In the first few days after birth, while the liver enzymes are maturing, clinical jaundice may become apparent when the serum bilirubin rises above 85 µmol/l (5 mg/dl). Bruising during delivery is a common cause of jaundice; this may also occur if there is polycythaemia. If jaundice occurs on the first day of life, a pathological process must be suspected and systemic infection or a haemolytic process must be ruled out. A level greater than 150 µmol/l on the second day of life or 250 µmol in the first week in a term baby also requires further investigation. Jaundice associated with lethargy and poor feeding may be the first sign of infection and should be investigated and treated without delay. Rare metabolic causes such as galactosaemia must be considered if jaundice persists.

A low serum albumin level and the presence of hypoxia contributes to the risk of kernicterus, a disorder in which free unconjugated bilirubin enters the lipid fraction of brain cells and causes an encephalopathy. Jaundice may complicate many surgical conditions in which bowel obstruction occurs below the ampulla of Vater, such as intestinal atresia, Hirschsprung's disease, and meconium ileus. Conjugated hyperbilirubinaemia occurs in biliary atresia, choledochal cyst, or other causes of obstruction to the bile duct such as inspissated bile syndrome.

Management of neonatal jaundice depends on a careful history and identification of the cause. The majority of well babies born at term need no treatment. Preterm babies and those who have suffered asphyxia or extensive bruising need phototherapy at lower serum bilirubin levels. If jaundice develops within the first 24 h after birth, investigation into the cause and phototherapy should be started immediately. When haemolysis is present due to rhesus incompatibility, the need for exchange transfusion will depend on the rate of rise of the serum bilirubin concentration, the absolute level, and the gestational age of the baby. Haemolysis due to ABO or other blood group incompatibility may also require active treatment (Table 3).

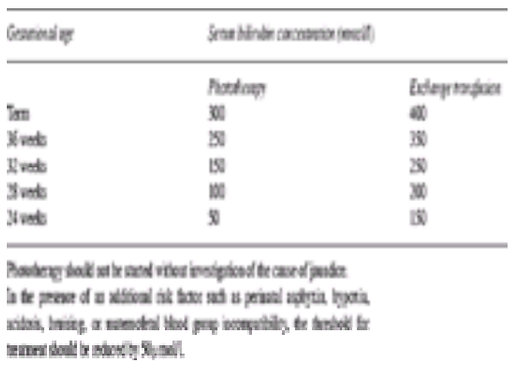


Table 3 Management of neonatal jaundice

Many drugs displace bilirubin from albumin and their use should be avoided in the newborn (Table 4).

Ampicillin
Chlorothiazide
Cloxacillin
Digoxin
Erythromycin
Frusemide
Sodium benzoate (carrier for diazepam)
Hydrocortizone
Oxacillin
Penicillin
Rifampicin
Sodium salicylate
Sulphonamides

Table 4 Drugs which displace bilirubin from albumin and increase the risk of kernicterus

Feeding, fluids, and electrolytes

Term babies should ideally receive breast milk. Breast feeding cannot be fully established if a neonate requires emergency surgery, and is interrupted when elective surgery is necessary. In both situations arrangements must be made for the mother to express her milk until feeding can recommence. Low birthweight or sick babies should be fed through an orogastric tube if enteral feeds can be given. Specific formulas are now available for the low birthweight baby if breast milk is not available: careful attention must be paid to the volume of fluid given, which will depend on the state of hydration, as well as the volume and specific gravity of urine. Any increase in insensible fluid losses, such as those which can occur if a preterm baby is nursed under a radiant heat warmer or when internal surfaces are exposed (as occurs in gastroschisis and exomphalos), must be taken into account. Urine volume is usually in the range 1 to 3 ml/kg per h and urine specific gravity is below 1012. The normal daily maintenance fluid requirements are shown in Table 5. Adjustments need to be made on the basis of body weight, serum electrolyte concentrations, plasma and urine osmolality, and the measured and insensible water losses. Inappropriate secretion of antidiuretic hormone leading to abnormally low serum sodium and osmolality, with concentrated urine, can occur after severe infections, pulmonary and neurological disorders, and neonatal surgery. Treatment is by fluid restriction. Extra intake of electrolytes through fluids administered intravenously must be taken into account and other sources of electrolytes such as antibiotics (e.g. sodium penicillin) may need to be included. The standard electrolyte intake is shown in Table 6. Sodium intake will need to be adjusted if there are abnormal losses or if secretion of ADH is inappropriate, and potassium intake must be monitored carefully if the baby has any bruising. Intravenous electrolyte supplementation can usually be discontinued when a baby is taking more than half the fluid requirements in the form of milk.

Day	Term	Preterm[a]
1	40	60
2	60	90
3	80	120
4	100	150
5	120	160
6	140	170
7	160	180
8	180	180

If a preterm baby is nursed under a radiant warmer add an extra 40ml/kg. day but in those of very low and extremely low birthweight the total volume may need to be restricted to 75 per cent to avoid a symptomatic patent ductus arteriosus. After 7days the volume may need to be increased further if weight gain is unsatisfactory.

Table 5 Standard fluid intake (ml/kg per day)

Sodium	3–5
Potassium	2–3
Calcium	0.5–1.0
Magnesium	0.25–0.5
Phosphate	0.25–1.0

Table 6 Standard electrolyte intake (mmol/kg per 24 h)

Infection

Babies, particularly those born preterm, are at increased risk of infections because of deficiencies in their defence mechanisms. Colonization with organisms acquired from the birth canal and the immediate hospital environment is rapid. The thin, easily-damaged skin of the newborn and the umbilical stump can be sources of infection. The transplacental passage of maternal IgG is the main source of humoral immunity in the neonate; because passage occurs mostly in the last trimester, preterm babies are at special risk. Cellular immunity is also deficient. Leucocytes have impaired chemotaxis and phagocytosis, and their absolute number is reduced.

Obsessional attention to hand washing by all staff and parents is essential if cross-infection is to be minimized. Gloves should be worn during any procedure where contact with blood may occur. Equipment such as stethoscopes should be provided for each baby or carefully disinfected before and after use. The wearing of gowns and masks does not reduce the risk of cross-infection and is unnecessary, except when staff need to be protected.

Investigation and treatment of infants with infection

Laboratory investigations should be carried out as soon as infection is suspected or when the maternal history, particularly during the labour, suggests that the risk of infection is high (Table 7). The clinical signs of infection in the newborn are often non-specific. Poor feeding with vomiting, temperature instability, hypotonia, apnoea, jaundice, tachypnoea, and hypoglycaemia are some of the early signs.

Prolonged rupture of membranes (<18 h)
Offensive amniotic fluid
Maternal fever (>37.5°C)
Spontaneous onset of preterm labour

Table 7 Perinatal risk factors for neonatal sepsis

A ‘sepsis screen' must be carried out promptly, and antibiotic treatment started without delay. The specimens to be collected and sent for bacterial and viral culture are shown in Table 8. Other investigations should be performed on the basis of clinical indications. Such investigations may include chest and abdominal radiographs, urinalysis and measurement of blood glucose, serum urea and electrolytes, and arterial pH and blood gases.

Early-onset sepsis (0–48 h after birth)	
Surface cultures	External ear canal
Systemic cultures	Blood
	CSF (except when it is judged that lumbar puncture would exacerbate respiratory distress)
Late-onset sepsis (after 48 h)	
Surface cultures	Clinically indicated surface swabs (e.g. umbilicus, septic spots)
Systemic cultures	Blood
	CSF
	Urine (suprapubic aspiration if strong clinical suspicion of urinary tract infection)

Table 8 Specimens to be cultured in the investigation of infection in the neonate

In addition to minimizing the problems associated with cold stress and inadequate feeding, babies suffering from infection may need ventilatory support for apnoea and respiratory failure. Inadequate peripheral circulation and hypotension must be treated with appropriate volume replacement (10 ml/kg over 20–60 min); this may need to be repeated. Anaemia should be treated by blood transfusion; this may be urgent if oxygen requirement is increasing. Hypoglycaemia should be treated immediately, while serum electrolyte abnormalities must be corrected gradually.

Antibiotic treatment

Initial therapy should be with antibiotics active against the spectrum of bacteria prevalent in the nursery. The organisms causing sepsis and the efficacy of the standard antibiotic policy should be continuously monitored.

In Oxford, penicillin G and an aminoglycoside have been found to be satisfactory empirical treatment for early-onset sepsis. Ampicillin, instead of penicillin, is favoured by many, but monotherapy with a third-generation cephalosporin is inadvisable since these drugs are ineffective against *Listeria monocytogenes*. The current standard regimen for suspected late-onset sepsis is flucloxacillin and netilmicin. Therapy can be changed on the basis of the results of cultures and clinical conditions. Systemic infection with *Staphylococcus epidermidis* is a particular hazard in infants with centrally placed silastic cannulae or ventriculoperitoneal shunts. Treatment of such sepsis with intravenous vancomycin may be unsuccessful until the foreign plastic is removed.

Periventricular haemorrhage and ischaemia

Frequent ultrasound scans of the brain, particularly in babies born at less than 33 weeks' gestation, will detect periventricular haemorrhage. This may occur in the first few days after birth and be followed by ventricular dilation. Periventricular leucomalacia as a result of ischaemic damage to the brain may be detected by the characteristic cystic formation seen on ultrasound scanning of the brain during the second and third weeks after birth.

The cerebral blood flow in sick newborn babies is unstable and the adverse effects of sudden changes in cardiac output and blood pressure must be taken into account during surgical procedures. Hypoxia and acidosis as well as hyper- and hypocapnoea must be avoided.

Polycythaemia

A venous haematocrit above 65 per cent increases blood viscosity. This is most common in growth retarded infants, those born postmature, or those born to mothers with pre-eclampsia toxaemia or diabetes. Delayed cord clamping and twin-to-twin transfusion lead directly to polycythaemia. Symptoms are attributable to inadequate blood flow through the capillary bed of various organs. Respiratory distress and cyanosis, hypoglycaemia, irritability, hypotonia, and lethargy may occur. Seizures, necrotizing enterocolitis, renal vein thrombosis, and peripheral gangrene occasionally occur.

If the haematocrit is between 65 and 70 per cent and any symptoms are present a dilutional exchange transfusion with fresh frozen plasma or human albumin solution should be carried out. A haematocrit above 75 per cent should be reduced even if there are no symptoms.

General care of older children

The level of anxiety, distress, and pain that a child experiences will depend on the circumstances and the clinical condition. All staff must recognize and respect the special needs of children and make an approach which is appropriate to their age and psychosocial development.

Children at greatest risk of subsequent psychological problems after an admission to hospital are those aged 6 months to 4 years who have either endured a long hospital stay or who have repeated admissions. Their response depends on the stage of development, the diagnosis, the family environment, and personal attitudes. Infants and toddlers may show regressive behaviour, such as clinging and decreased interest in play and physical contact. Slightly older children may regard their illness as a punishment and enuresis may occur. School-aged children may become physically and verbally aggressive; they need to be given the opportunity to have their questions answered and to be actively involved in their treatment. When a surgical procedure is planned, a visit to the hospital prior to admission can be a great advantage to the child and parents. This applies to minor as much as to major operations. Adolescents are more able to deduce the sequence of events and can analyse the experience independently providing they are given opportunities to discuss their treatment. Respecting their need for privacy and encouraging involvement in their care can contribute to a successful outcome. Clearly the diagnosis and subsequent treatment play a major part in influencing a child's behaviour. In addition, a full understanding of a child's social circumstances, including other illness in the family and possible experience of a grandparent's death, for example, are important factors.

Effective communication between all the staff involved, as well as consistent and frequent discussions with the parents and child, are an integral part of the treatment.

Clinical assessment

The ability to recognize illness in a child is an important asset for any member of the clinical team. Examination should involve an immediate assessment of the level of

consciousness, degree of distress, and, if possible, localization of pain. The state of peripheral perfusion, hydration, and presence or absence of anaemia or cyanosis must be noted. Lack of eye contact, hypotonia, and irritability when roused are often signs of illness.

Normal heart rate decreases from an average of 145 beats/min at the age of 1 month to 70 beats/min at 14 years. Tachycardia will normally increase cardiac output and should be expected to a certain degree when cardiorespiratory disease is present. Blood pressure increases gradually throughout infancy and childhood, reaching adult levels by 14 years.

A child's blood volume is approximately 80 ml/kg. In addition to any traumatic blood loss, blood taken for laboratory analysis must be recorded and a cumulative total kept. Laboratories should use micromethods, as frequent samples often need to be taken.

In childhood, lung compliance increases from 5 ml/cmH_2O to 70 ml/cmH_2O by 7 years; this is approximately half the adult value. The chest wall becomes more rigid throughout this time, increasing the efficiency of ventilation. Because of the small tidal volume, mechanical ventilators for children must have characteristics that enable low pressures and short inspiratory times to be used.

The assessment of neurological integrity in a sick child requires knowledge of the normal stages of development throughout childhood. Management of severe encephalopathy after trauma or infection should only be carried out in a centre which has full monitoring facilities.

Fluid must be administered to a sick child via a volume-controlled infusion pump, and must be monitored on an hourly basis. All sources of fluid intake and loss should be taken into account when fluid requirements are calculated. Hypoglycaemia is common in children, and an intravenous infusion should be maintained until they are able to feed normally. Caloric intake must be maintained and enhanced in all conditions that lead to a higher metabolic rate. These include cardiorespiratory distress, sepsis, pyrexia, and trauma, particularly after burns and during the postoperative phase.

A neutral thermal environment, whilst most important in the care of the newborn, can also contribute to the well being and recovery of older children. Unnecessary distress from overheating as well as from underheating must be avoided, particularly in children who are unable to move fully.

The risk of nosocomially acquired infection increases postoperatively, particularly when ventilatory support and central catheters continue to be necessary. Meticulous attention to hand washing before and after examination of the child is essential.

Further reading

Chwals WJ. The metabolic response to surgery in neonates. *Current Opinion Pediatrics* 1994; **6**: 334–40. [Over the past 35 years surgery-associated mortality has declined from >60 per cent to <10 per cent due to a better understanding of the pathophysiological changes during the perioperative period. This article highlights recent developments in our understanding of the acute metabolic response to injury and how this knowledge might be used to improve further the outcome.]

Halliday HL. Overview of clinical trials comparing natural and synthetic surfactants. *Biology of Neonate* 1995; **67** (suppl. 1): 32–47. [Meta-analysis of 10 randomized, controlled trials showing more rapid onset of action and reduced risk of neonatal mortality and morbidity with natural as compared to synthetic surfactants.]

Isaacs D, Moxon ER. *Neonatal infections*. Butterworth-Heinemann, Oxford, 1991. [A monograph covering all aspects of the pathogenesis, epidemiology, clinical presentation, and treatment of infection in the newborn with particular emphasis on surveillance and clinical pharmacology. Appendices on definitions of infections and guidelines for treatment.]

Lipsky K (ed.). Psychological, social, and legal issues. In: Levin DL, Morriss FC, eds. *Essentials of pediatric intensive care*, 2nd edn. Quality Medical Publishing, St Louis, 1998. [Psychosocial, cultural, and legal aspects of paediatric intensive care from the view point of the parents, patient, and staff.]

Morley CJ. Systematic review of prophylaxis *vs* rescue surfactant. *Archives of Disease in Childhood* 1997; **77**: F70–4. [Analysis of six large studies showing 39 per cent reduction in mortality in babies less than 32 weeks treated with surfactant at birth (prophylactic), equivalent to saving seven more lives than treatment later (rescue) for every 100 babies treated.]

Sinclair JC, Bracken MB, Horbar JD, Soll RF. Introduction to neonatal systematic reviews. *Pediatrics* 1997; **100**: 892–5. [Evidence-based practice requires valid evidence of efficacy and safety of treatments. This article presents the methodology of systematic reviews and the role of the Cochrane Collaboration.]

Spitz L. Neonatal surgery. *Journal of the Royal College of Surgeons Edinburgh* 1995; **40**: 84–7. [Advances in the sophistication of operative and anaesthetic techniques and pre- and postoperative care have contributed to improved survival of infants undergoing surgery. Neonatal surgery must be concentrated in regional centres with a full range of expertise to contribute to research and training.]

42.2 Thoracic surgical problems

Robert C. Shamberger and W. Hardy Hendren

Chest wall
Congenital lesions
Chest wall tumors
Diaphragm
Bochdalek hernia
Morgagni hernia
Eventration of the diaphragm
Tracheobronchial tree
Congenital lesions
Tracheobronchial and esophageal foreign bodies
Pulmonary
Congenital lesions
Infectious lesions
Pulmonary complications of cystic fibrosis
Pulmonary neoplasms
Pleura
Chylothorax
Pneumothorax
Mediastinum
Congenital lesions
Neoplastic lesions
Esophagus
Congenital lesions
Acquired lesions
Aortic arch anomalies
Further reading

Chest wall

Congenital lesions

Congenital deformities of the anterior thorax can be divided into four groups: pectus excavatum, pectus carinatum, Poland's syndrome, and sternal clefts and defects, including ectopia cordis.

Pectus excavatum

Pectus excavatum (funnel chest) appears as a depression of the sternum and lower costal cartilages. The first and second ribs and their costal cartilages, as well as the manubrium, are usually normal (Fig. 1). The sternal and cartilaginous deformity is variable in extent and is frequently asymmetrical: the right side is often more depressed than the left, with corresponding rotation of the sternum.

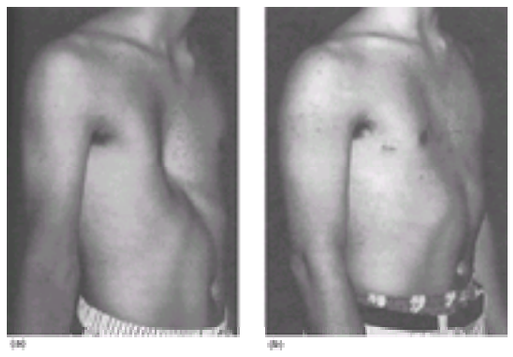


Fig. 1. (a) Preoperative photograph of a 15-year-old boy with a symmetric pectus excavatum deformity. (b) The excellent postoperative result utilizing a retrosternal strut.

The etiology of pectus excavatum is unknown, and there is little evidence to support the hypothesis that it arises from an abnormality of the diaphragm. The condition occurs once in every 300 to 400 live births, and is usually noted within the first year of life. In less than 5 per cent of cases the diagnosis is made at adolescence. A family history of anterior thoracic deformity is present in 37 per cent of patients. Scoliosis affects 26 per cent of patients and 11 per cent of these have a family history of scoliosis. Pectus excavatum is common in patients with abdominal musculature deficiency syndrome (prune belly syndrome) and in patients with Marfan's syndrome, in whom it is often accompanied by scoliosis.

Pectus excavatum is well tolerated in infancy. Older children may complain of pain in the area of the deformed cartilage or of precordial pain during exercise. A few patients suffer from palpitations or syncope, presumably due to transient atrial arrhythmias: mitral valve prolapse may be identified in these patients. Since transient deformity of the chest wall commonly occurs in infants during vigorous breathing or crying, pectus excavatum should not be corrected before the age of 2 years.

Pathophysiology

Although some authors maintain that pectus excavatum causes no cardiovascular or pulmonary impairment, many patients seem to have increased stamina after surgical correction of the deformity. Abnormalities of pulmonary function are difficult to demonstrate, since ‘normal' parameters vary widely and are heavily dependent on physical training and body habitus. Abnormalities suggested by several different studies of patients have included an increased work of breathing, abnormal ventilation to perfusion ratios, subnormal cardiac response to upright exercise, and cardiac compression. Exercise tolerance increases following surgery in many patients.

Surgical repair

The earliest surgical repairs of pectus excavatum were reported by Meyer in 1911 and Sauerbruch in 1920. In 1949 Ravitch reported a technique that included excision of all deformed costal cartilages with the perichondrium, division of the xiphoid from the sternum, division of the intercostal bundles from the sternum, and a transverse sternal osteotomy overcorrecting the sternum anteriorly. Kirschner wire fixation was used in the first two patients, and silk suture fixation in later patients. There is no conclusive evidence that internal fixation with Kirschner wires or metallic struts provides better long-term results than those achieved in patients without metal fixation as no randomized study has been performed.

In 1958 Welch reported a technique that allowed safe and satisfactory correction of pectus excavatum. He emphasized total preservation of the perichondrial sheaths of the costal cartilage, preservation of the upper intercostal bundles, and anterior fixation of the sternum with silk sutures. This technique has been used at The Children's Hospital, Boston, to treat more than 700 patients in the past 30 years. Nevertheless, we use struts devised by Adkins and Blades to hold the sternum in an anterior position, in the belief that these provide stability and improved initial results, and greater comfort and less chance of recurrence. The struts are removed 6 to 12 months after the repair.

Our operative technique is illustrated in Fig. 2. In girls, particular attention is taken to place the incision within the inframammary fold, avoiding the complications of breast deformity and abnormal development. Skin flaps are mobilized primarily in the midline to the angle of Louis superiorly and to the xiphoid inferiorly. Bovie electrocautery minimizes blood loss. Starting medially, the chest wall muscles are detached and retracted to expose the depressed costal cartilages. The lateral extent of muscle elevation is the costochondral junction of the third to fifth ribs, and occasionally the second. Particular care is taken to avoid injury to the intercostal bundles: this can result in major bleeding. The rectus muscles may be detached from the lower sternum and xiphoid, allowing blunt dissection behind the sternum. This is often not required if retrosternal struts are used. If the xiphoid projects forward, it should be removed. Subperichondrial resection of the costal cartilages is performed, removing the third, fourth, and fifth cartilages, but preserving the costochondral junctions. Longer segments of the sixth and seventh cartilages (5 to 6 cm) are resected to the point where they flatten to join the costal arch. Bridges joining the fifth and sixth costal cartilages are often present lateral to the sternum. Transverse osteotomy of the anterior table of the sternum is performed. The osteo-tomy must be large enough to allow the sternum to be brought forward after fracture of the posterior table. The osteotomy is then closed with either heavy silk sutures while the sternum is held forward in a deliberately overcorrected position or with strut fixation. Struts can be used to hold the sternum forward, and are especially helpful in patients with extensive sternal rotation or a severe depression. If required, the lowest one or two sets of intercostal bundles (sixth and seventh) can be divided from the sternum to allow it to be brought forward without excessive tension. The number of perichondrial bundles divided should be minimized to avoid an unsightly local depression just inferior to the tip of the sternum and division is rarely required with strut fixation.

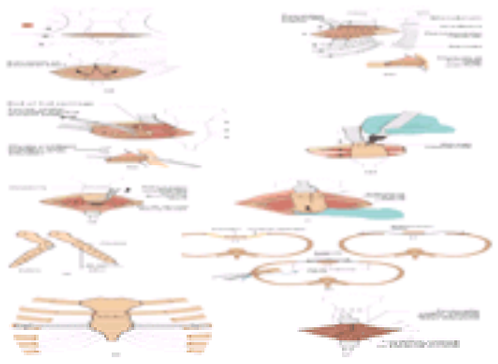


Fig. 2. (a) A transverse incision is placed below and well within the nipples at the site of the future inframammary crease. The pectoralis major muscle is detached from the sternum along with portions of the pectoralis minor and serratus anterior muscles and retracted forward and laterally to expose the depressed costal cartilages (usually the third to seventh). (b) Subperichondrial resection of the costal cartilages is achieved by incising the perichondrium anteriorly. It is then dissected from the costal cartilages in the bloodless plane between perichondrium and costal cartilage. Cutting the perichondrium 90° in each direction at its junction with the sternum (inset) facilitates visualization of the back wall of the costal cartilage. (c) The cartilages are divided at the junction with the sternum with a Welch perichondrial elevator held posteriorly to elevate the cartilage and protect the mediastinum (inset). The divided cartilage can then be held with an Allis clamp, elevated, and divided leaving 5 to 10 mm of cartilage on the costochondral junction. (d) The sternal osteotomy is created above the level of the last deformed cartilage and the posterior angulation of the sternum, generally the third cartilage but occasionally the second. Two transverse sternal osteotomies are created through the anterior cortex 2 to 4 mm apart with a Hall air drill. (e) The base of the sternum and the rectus muscle flap are elevated with two towel clips and the xiphoid can be divided from the sternum with electrocautery, allowing entry into the retrosternal space. This step is not always required with the use of a retrosternal strut. Preservation of the attachment of the perichondral sheaths avoids an unsightly depression which can occur below the sternum. (f) When a strut is not used, the osteotomy is closed with several heavy silk sutures as the sternum is being elevated with the assistant's thumb. (g) Correction of the abnormal position of the sternum is achieved by creation of a wedge-shaped osteotomy which is then closed bringing the sternum anteriorly into an overcorrected position. (h) This figure demonstrates the use of both retrosternal struts and the Rehbein struts. The Rehbein struts are inserted into the marrow cavity (inset) of the third or fourth ribs and are then joined to each other medially to create a metal ‘arch' anterior to the sternum. The sternum is sewn to the arch to secure it in a forward position. The retrosternal strut is placed behind the sternum and is secured to the rib ends laterally to prevent migration. (i) Anterior view of the retrosternal strut. The perichondrial sheath to either the third or fourth rib is divided at its junction with the sternum, and the retrosternal space is bluntly dissected to allow passage of the strut behind the sternum. An adequate space must be created to avoid injury to the pericardium. The strut is secured with two pericostal sutures laterally to prevent migration. (j) The pectoral muscle flaps are secured to the midline of the sternum advancing the flaps to obtain coverage of the entire sternum. The rectus muscle flap is then joined to the pectoral muscle flaps.

This technique allows precise resection of costal cartilages with preservation of perichondrial sheaths, without injury to mediastinal structures, and usually without pleural or pericardial entry. If the pleura is entered, the hole should be made large enough to prevent tension pneumothorax and the lungs should be ventilated with enough positive pressure to maintain inflation. Any pleural air is aspirated through a temporary catheter, which is withdrawn when the defect is covered by the pectoralis muscle closure. The wound is flooded with warm saline solution and cefazolin to remove clots. A single-limb, medium, Hemovac drain is brought through the inferior skin flap with the suction ports in a parasternal position to the level of the highest costal cartilage resection. The pectoralis muscle flaps are sutured to the sternum, advancing the muscles medially and inferiorly to cover the underlying sternum with muscle. The rectus muscles if divided are then reattached to the lower sternum medially and to the pectoralis muscles laterally. A postoperative chest radiograph is obtained in the recovery room. Antibiotic prophylaxis consists of cefazolin, given once before surgery and for three doses afterwards. Blood loss is well below that at which transfusion is required. Safe and effective repair is possible at any age: we have experienced no deaths and few complications (4.4 per cent) in 704 patients.

Complications

Complications of pectus excavatum repair should be few and minor. Limited pneumothorax, which may or may not require aspiration, occurs in 2 per cent of patients. Tube thoracostomy was required in only four patients in our series. Wound infection is rare with the use of perioperative antibiotic coverage.

The most distressing complication following surgical correction of pectus excavatum is major recurrence years after the original repair (17 of 704 patients). It is difficult to predict which patients will be affected, but they often have an asthenic or ‘marfanoid' habitus with poor muscle development and a narrow anterior/posterior chest wall diameter.

Martinez first described in 1990 a deficiency in thoracic growth in children following repair of pectus excavatum which was most noticeable in children repaired early during the preschool years. Haller more recently reported on three boys who have presented in their teens with apparent limited growth of the ribs following resection of the costal cartilages at an early age producing a band-like narrowing of the mid-chest (Fig. 3). In some cases the first and second ribs in which the costal cartilages have not been resected have apparent relative overgrowth producing anterior protrusion of the upper sternum. This occurrence has been attributed by Haller to injury during surgical repair of the costochondral junctions which are the longitudinal growth centers for the ribs and to decreased growth of the sternum resulting from injury to its growth centers or vascular supply.



Fig. 3. A 13-year-old boy is shown 8.5 years after his pectus excavatum repair at 4 years of age. His chest demonstrates ‘overgrowth' of the upper portion of the chest and ribs which were not resected during his repair of a pectus excavatum and a relative narrowing of the mid-chest.

Martinez demonstrated experimentally in 6-week-old rabbits that resection of the costal cartilages produced a marked impairment in chest growth, particularly the anterior/posterior diameter during a 5.5 month period of observation. Less severe impairment occurred if only the medial three-quarters of the costal cartilage was

resected leaving the growth centers at the costochondral junction. This impairment was attributed to fibrosis and scarring within the perichondrial sheaths. Perichondrial sheaths, bone, or other prosthetic tissues which cannot grow should also not be joined posterior to the sternum as they will form a band-like stricture across the chest. This complication of delayed thoracic growth was described primarily in children repaired in early childhood and can be avoided by delaying surgery until the children are older. Preservation of the costochondral junction leaving a segment of the cartilage on the osseous portion of the rib may also minimize growth impairment. The author currently delays surgery until the children have begun their pubertal growth spurt. Weber has described a procedure for widening the sternum with an interposition graft in a child with severe restrictive respiratory impairment from this complication which was successful in improving respiratory function.

Pectus carinatum

Pectus carinatum describes anterior protrusion deformities of the chest wall: these are less common than depression deformities (16.7 per cent of our series). The most common form of pectus carinatum consists of symmetrical anterior displacement of the sternum with concavity of the costal cartilages laterally: a chondrogladiolar defect (Fig. 4). Asymmetrical deformities, with anterior displacement of the costal cartilages on one side and normal cartilages on the contralateral side and a normally positioned or oblique sternum, are less common. ‘Mixed' lesions are characterized by a carinate deformity on one side and a depression or excavatum deformity on the other. The upper or chondromanubrial deformities (‘pouter pigeon' deformities), with protrusion of the manubrium and second and third costal cartilages with relative depression of the short body of the sternum are least common (Fig. 5).

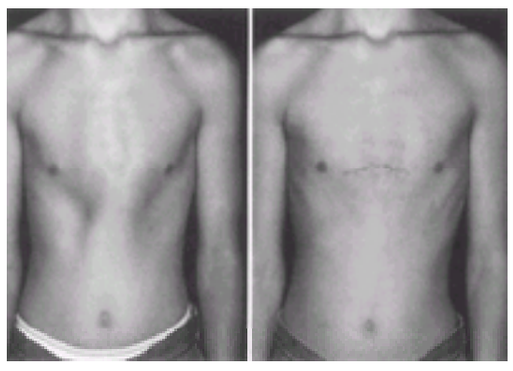


Fig. 4. Preoperative photograph (left) of a patient with a symmetric pectus carinatum deformity demonstrating symmetric anterior protrusion of the body of the sternum as well as of the costal cartilages. The postoperative result (right) shows marked improvement in the chest wall contour.

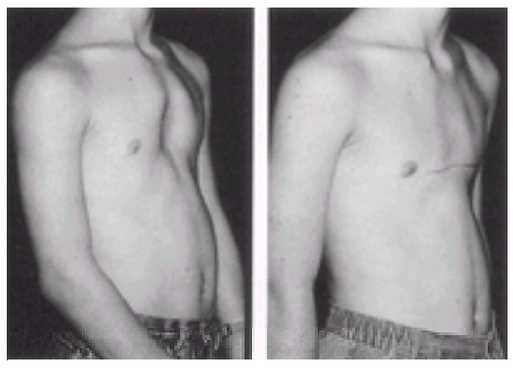


Fig. 5. Preoperative photograph (left) of a patient with the upper form of pectus carinatum deformity with marked anterior protrusion of the manubrium and second and third costal cartilages along with depression of the body of the sternum (pouter pigeon deformity). The postoperative result (right) is shown with correction of both the depression and superior protrusion components of the deformity.

The etiology of pectus carinatum is no better understood than that of pectus excavatum. It appears as an overgrowth of the costal cartilages with forward buckling and anterior displacement of the sternum. A clear-cut familial tendency suggests a genetic basis for this deformity. In a recent review of 152 patients, 26 per cent had a family history of chest wall deformity. A family history of scoliosis was obtained in 12 per cent of the patients. The condition is more common in boys than in girls. Scoliosis and other deformities of the spine are the most common associated musculoskeletal deformities.

Pectus carinatum usually appears in childhood; in almost half of our patients the deformity was not identified until the children were at least 11 years of age. The deformity may appear in a mild form at birth and often progresses, particularly during the period of rapid growth at puberty. The chondromanubrial deformity is noted at birth and is associated with a short truncated sternum with absent sternal segmentation or premature obliteration of sternal sutures. In a prospective review of 135 patients with sternal fusion anomalies, 18 per cent had documented congenital heart disease.

Surgical repair

Surgical repair of carinate deformities was first reported by Ravitch in 1952. He resected multiple costal cartilages and performed a double sternal osteotomy to correct chondromanubrial deformity. In 1953 Lester reported two methods of repair for chondrogladiolar deformity. The first, involving resection of the anterior portion of the sternum, was abandoned because of excessive blood loss and unsatisfactory results. The second, although no less radical, involved subperiosteal resection of the entire sternum. Chin advanced the transected xiphoid and attached the rectus muscles to a higher site on the sternum (xyphosternopexy). This method produced posterior displacement of the sternum in younger patients with a flexible chest wall. Howard combined this method with subperichondrial costal cartilage resection and a sternal osteotomy. In 1973 Welch reported the approach which I use today.

As in the pectus excavatum repair, a transverse incision is made just below and within the nipples, with mobilization of the pectoral muscle flaps and subperichondrial resection of the deformed costal cartilages. A sternal osteotomy allows the sternum to be fractured and displaced posteriorly into an orthotopic position (Fig. 6(a)). A second osteotomy is occasionally required to enable the lower portion of the body of the sternum to be displaced posteriorly. For correction of the upper or chondromanubrial deformity, the costal cartilages must be resected from the second cartilage inferiorly. A generous wedge osteotomy is then performed at the point of maximal protrusion of the sternum. The superior segment of the sternum can then be displaced posteriorly as the osteotomy is closed, advancing the inferior segment anteriorly (Fig. 6(b)). This corrects both components of the deformity. Mixed pectus carinatum/excavatum deformities are managed with a transverse wedge-shaped osteotomy, allowing anterior displacement and rotation of the sternum (Fig. 6(c)). Closure and postoperative drainage are performed as in pectus excavatum.

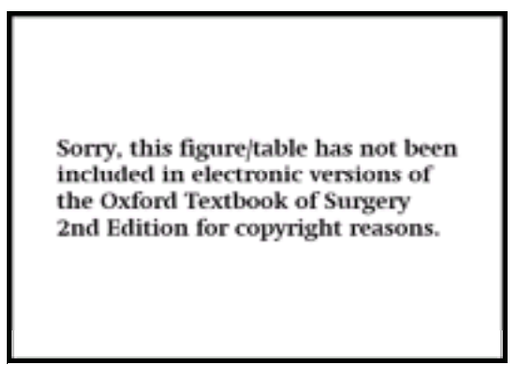


Fig. 6. (a) Chondrogladiolar deformity (90.8 per cent of carinate deformities; symmetrical or asymmetrical) is managed with a single or double osteotomy after resection of the costal cartilages. This allows posterior displacement of the sternum to an orthotopic position. (b) Chondromanubrial deformity is rare. The broad superior osteotomy achieves anterior displacement of the body of the sternum and posterior displacement of the manubrium. (c) The mixed pectus deformity is corrected by symmetrical resection of the third to seventh costal cartilages followed by transverse offset (0 to 10°) wedged-shaped sternal osteotomy. Closure of this defect

achieves both anterior displacement and rotation of the sternum.

Operative results

Surgical correction of pectus carinatum is generally successful and postoperative recovery uneventful. Blood transfusions are rarely required and there is a 3.9 per cent complication rate. Complications include pneumothorax, wound infection, recurrence, and postoperative pneumonitis. Rare patients require revision, usually in the form of resection of additional lower costal cartilages to correct for persistent unilateral malformation of the costal arch.

Poland's syndrome

In 1841 Poland described congenital absence of the pectoralis major and minor muscles associated with syndactyly. The clinical manifestations of this condition cover a wide spectrum, often with chest wall and breast deformities. Thoracic deformity ranges from hypoplasia of the sternal head of the pectoralis major muscle and the pectoralis minor muscle with normal underlying ribs to severe hypoplasia of the ribs with complete absence of the anterior portions of the second to fourth ribs and cartilages (Fig. 7). There may be mild hypoplasia of the breast, or complete absence of the breast (amastia) and nipple (athelia). There is often minimal subcutaneous fat and no axillary hair on the involved side. Hand deformities include hypoplasia (brachydactyly), fused fingers (syndactyly), and mitten or claw deformity (ectromelia). There is no correlation between the extent of hand and thoracic deformities.

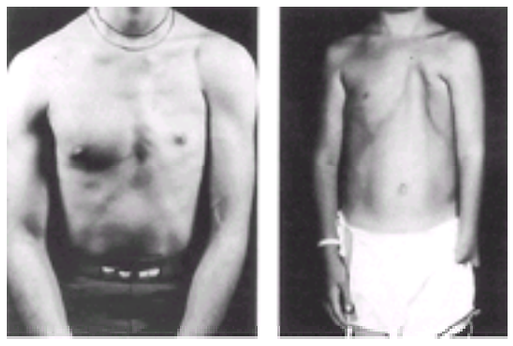


Fig. 7. (a) A 16-year-old boy with Poland's syndrome with absent pectoralis major and pectoralis minor muscles. Ribs are intact and in a normal contour. (b) A 10-year-old boy with Poland's syndrome in whom the anterior third to fifth left ribs are hypoplastic and do not reach the sternum. The left nipple is also hypoplastic and displaced superiorly. This child also had a severe hand deformity with absence of most of the bony elements of the hand and wrist.

This condition is present from birth and has an estimated incidence of 1/30 000 to 1/32 000 live births. Breast abnormalities can be identified at birth by the absence of the underlying breast bud and hypoplasia or aplasia of the nipple, which is often superiorly displaced. The etiology is unknown and cases are sporadic.

Surgical repair

Assessment of the extent to which various musculoskeletal components are involved is critical for optimal thoracic reconstruction. If deformity is limited to the sternal component of the pectoralis major and minor muscles, there is little functional deficit and repair is not necessary, although breast augmentation should be performed in females when growth is complete. If the underlying costal cartilages are depressed or absent, chest wall repair must be considered to minimize the concavity, eliminate the paradoxic motion of the chest wall, and to provide an optimal base for breast reconstruction in females. Ravitch reported correction of posteriorly displaced costal cartilages by unilateral resection of the cartilages, a wedge osteotomy of the sternum allowing rotation of the sternum, and fixation with Rehbein struts and Steinmann pins. Good results can often be achieved with bilateral costal cartilage resection and an oblique osteotomy. This method allows correction of the rotational deformity. The sternum is displaced anteriorly correcting the posterior displacement of the costal cartilages. An unappreciated carinate deformity is often present on the contralateral side, accentuating the ipsilateral concavity.

Absence of the medial portion of the ribs can be managed with split rib grafts taken from the contralateral side. These grafts must be secured to the sternum medially and to the hypoplastic rib ends laterally, and then can be covered with a prosthetic mesh if further support is needed. In these patients there is little tissue between the endothoracic fascia and the fascial remnants of the pectoral muscles. Augmentation of this area by transfer of a latissimus dorsi muscle flap is particularly helpful in girls, who will require breast augmentation. Latissimus transfer is seldom if ever required in boys and has the disadvantage of adding a second thoracic scar, as well as removing one of the major functional muscles of the shoulder and arm.

Sternal defects

Deformities due to failure of ventral fusion of the thoracic wall can be divided into thoracic ectopia cordis, thoracoabdominal ectopia cordis, and cleft sternum without ectopia cordis.

Cleft sternum

Cleft sternum with an orthotopic heart is the simplest of these deformities. The cleft may be complete or incomplete and results from failure of ventral fusion of the sternal bars, which normally occurs at the eighth week of gestation. In cleft sternum the midline sternal separation is covered and the pericardium, pleura, and diaphragm are intact; omphaloceles do not occur in conjunction with this anomaly. Crying or a Valsalva maneuver produces a dramatic increase in the apparent severity of the deformity. Intrinsic congenital heart disease is rare.

Treatment is recommended in the newborn period, when the malformation can be closed primarily without prosthetic materials or cardiac compression. Primary closure in older patients produces excessive cardiac compression. Reconstruction of the anterior chest wall using multiple oblique chondrotomies lengthens the costal cartilages and decreases cardiac compression in older patients, who have a less flexible chest wall. Autologous grafts including costal cartilages, split ribs, and resection of the costal arch complex have been described, but repair undertaken in the newborn period is optimal and avoids the need for these methods.

Thoracic ectopia cordis

While management of isolated cleft sternum is uniformly successful, few patients survive surgical treatment of thoracic ectopia cordis. These patients have a high incidence of associated intrinsic cardiac anomalies, in addition to the abnormal position of the heart and surrounding somatic structures. In thoracic ectopia cordis the heart is external to the thorax, protruding at the upper to mid-thoracic level. The apex of the heart points anteriorly. Attempts to return the heart to the thorax occlude the great vessels and are not tolerated.

Thoracoabdominal ectopia cordis (Cantrell's pentalogy)

The features of the Cantrell pentalogy are a cleft lower sternum, an anterior diaphragmatic defect due to failure of development of the septum transversum, absence of the parietal pericardium, an adjacent or completely separate omphalocele, and in most patients, a cardiac anomaly, frequently tetralogy of Fallot.

Immediate closure of the abdominal wall is required in neonates with an omphalocele and lower sternal cleft. An aggressive approach should be taken if survival rates are to be improved. Repair of the cardiac defect can be performed after somatic closure.

Chest wall tumors

Chest wall tumors are a rare occurrence in infants and children. Only 1.8 per cent of the solid childhood tumors admitted to St. Jude's Hospital (Memphis, Tennessee)

were located in the thorax. Thoracic neoplasms, both benign and malignant, are primarily of mesenchymal origin and comprise a broad spectrum of tumors (Table 1). In adults, secondary involvement of the chest wall by pulmonary carcinoma and breast carcinoma occurs, but this is very rare in children.

Benign lesions
Chondroma
Osteochondroma
Fibroma
Lipoma
Mesenchymal hamartoma
Eosinophilic granuloma
Hemangioma
Aneurysmal bone cyst
Fibrous dysplasia
Malignant lesions
Chondrosarcoma
Osteosarcoma
Osteochondrosarcoma
Fibrosarcoma
Malignant mesenchymoma
Ewing's sarcoma (Askin tumor)
Rhabdomyosarcoma
Leiomyosarcoma
Lymphoma (Hodgkin's and non-Hodgkin's)

Table 1 Chest wall tumors in infants and children

Children with a chest wall tumor most frequently present with either a palpable mass, cough, or pain. Rarely a child with a malignant neoplasm will present with respiratory distress from a large malignant pleural effusion or an extensive intrathoracic mass (Fig. 8). Pectus carinatum appearing in the pubertal adolescent must be distinguished from a neoplasm. In pectus carinatum the protruding costal cartilages can each be palpated and the chest radiograph is normal. Unnecessary radiographic studies frequently are obtained in these children prior to referral to a pediatric surgeon.

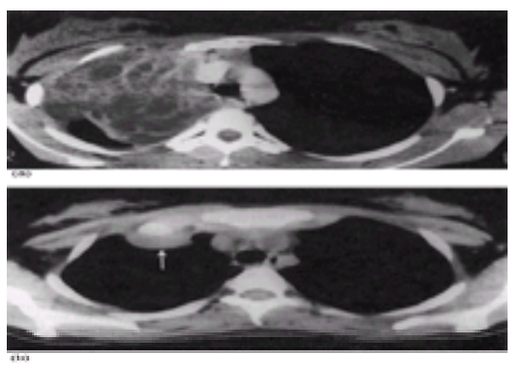


Fig. 8. (a) CT scan of a 14-year-old girl who presented with a cough. Her chest radiograph showed an apical thoracic mass. The CT scan shows a large mass of varying densities without erosion of the ribs. (b) The chest radiograph was normal after 2 months of chemotherapy and the CT scan shows a small residual soft tissue mass (arrow). The pathologic examination after resection revealed no viable tumor in this specimen.

Surgical management of chest wall tumors is similar to that of sarcomas at other sites. Evaluation of a suspected chest wall tumor should begin with standard chest radiographs. These will define the presence of bony involvement, the extent of the soft tissue mass, and the presence of a pleural effusion or pulmonary nodules. A computed tomogram (**CT scan**) will provide additional useful information. Many tumors involving the chest wall are highly malignant, so the presence of metastatic disease must always be assessed. Before major surgical resection, a bone scan is appropriate to exclude osseous metastases. The bone scan further defines the extent of primary disease within the rib and presence of intramedullary ‘skip' metastasis. Bone marrow aspirate and biopsy and brain scan may also be appropriate.

The initial therapeutic intervention is usually an incisional or needle biopsy of the mass. Proper placement of the incision to avoid compromise of future resection is of paramount importance. Adequate diagnostic material for standard histologic, cytogenetic, and biologic studies can usually be obtained with multiple needle biopsies.

The majority of tumors are in the Ewing's sarcoma/primitive neuroectodermal family (malignant, small, round cell tumors of the chest wall) followed by rhabdomyosarcomas and then tumors of osseous origin including osteosarcoma and chondrosarcoma. Primary resection should be limited to the benign lesions and rarely, the small malignant tumors. In most cases, after biopsy, chemotherapy is administered prior to secondary resection of the tumor. Initial chemotherapy will generally produce a more resectable tumor which is smaller, less vascular, and less friable and prone to rupture during resection.

Benign lesions

A growing mass can rarely be ignored because of the high preponderance of malignant lesions in the chest wall. Eosinophilic granuloma, aneurysmal bone cyst, and fibrous dysplasia produce intraosseous lesions within the ribs. Biopsy, either incisional or excisional, may be required to establish a diagnosis. Similarly, a hemangioma growing within the muscles of the chest wall in a young infant may increase dramatically in size. Unless characteristic dermal markings are present, an incisional biopsy is required to exclude rhabdomyosarcoma or other soft tissue sarcoma. Once the diagnosis of a hemangioma is established, observation is appropriate. A chondroma, osteochondroma, or osteoma of the ribs will often have characteristic radiographic findings and if small, will not necessarily require biopsy.

Mesenchymal hamartoma

Mesenchymal hamartoma is rare in infants and usually is identified at birth. This lesion involves focal overgrowth of normal skeletal elements and histologically is benign. The radiographic appearance of a mesenchymal hamartoma is of a partially mineralized extrapleural mass involving one or more ribs (Fig. 9). Surgical ablation has been utilized in the past, but may result in major blood loss and the long-term complication of scoliosis. No reports of malignant degeneration of these tumors has been published. One patient had a biopsy taken of the lesion with no subsequent surgery and was followed for 24 years at which time she was alive and well. The size of the chest wall tumor had decreased proportionately from when she was an infant. The only symptoms attributed to these tumors have been respiratory. One infant of nine reported by McLeod and Dahlin suffered respiratory death 30 min after delivery, presumably due to severe pulmonary hypoplasia produced by the tumor which filled an entire hemithorax. The term ‘mesenchymoma' that is occasionally used to describe these lesions is probably incorrect in that it suggests a neoplasm. Since these are self-limited, non-neoplastic processes, the label mesenchymal hamartoma is more appropriate.



Fig. 9. Chest radiograph of a newborn girl who presented with respiratory distress. Multiple masses were palpable in the right axilla. The radiograph demonstrates a large mass nearly filling the right hemithorax. Multiple ribs were involved and calcification was seen within the mass. Extensive bleeding occurred during an incisional biopsy. Because of the extensive bleeding and vascular nature of the mass, this lesion was followed expectantly. As the child grew it became progressively smaller and smaller in relative size and the girl lived to a normal adulthood.

Mesenchymal hamartomas are generally well circumscribed, compressing the surrounding tissue rather than invading it. Histologically they are largely composed of a mixture of proliferating elements. Other areas within these lesions are similar to the aneurysmal bone cyst with large vascular channels. These cyst-like areas filled with blood may compose a major portion of the mass explaining the extensive blood loss encountered with resection. Other areas may have primarily fibroblastic or chondroblastic tissues with evidence of endochondral ossification. Cellular hyperplasia is without malignant characteristics. Multiple ribs are often involved in these lesions.

Malignant lesions

Chondrosarcoma may occur in the pediatric age range. As in adults, this is primarily a locally invasive tumor which responds poorly to any form of chemotherapy or radiotherapy. Aggressive surgical resection following initial biopsy of chondrosarcoma is appropriate.

Osteosarcoma may occur in the ribs. As in extremity lesions, initial incisional biopsy followed by preliminary or ‘up-front' chemotherapy is utilized prior to secondary surgical resection. Survival is directly related to two factors: the presence of metastatic disease, in which case survival is limited, and complete resection of the primary tumor. Local control in osteosarcoma can rarely be achieved by chemotherapy and radiotherapy alone.

Congenital fibrosarcoma

Congenital fibrosarcoma will generally present at birth. It must be distinguished from mesenchymal hamartoma. These tumors are responsive to treatment with vincristine, actinomycin-D, and cyclophosphamide (**VAC**) or doxorubicin (**Adriamycin**) in conjunction with VAC (**Adria-VAC**). In infants with large congenital fibrosarcoma, initial treatment with chemotherapy has been effective. It can decrease the size of the mass dramatically allowing a much less extensive resection. Complete resolution of these tumors with chemotherapy alone has been reported with the Adria-VAC regimen, but if resection is feasible it should be considered.

Ewing's sarcoma—malignant, small, round cell tumor

The predominant tumor in all pediatric series of chest wall lesions is Ewing's sarcoma (malignant, small, round cell tumor). Most are of osseous origin. Ewing's sarcoma is an extremely aggressive tumor with frequent metastatic spread and local recurrence. Up to 6.5 per cent of primary Ewing's sarcomas arise in the chest wall. It occurs primarily in children and young adults with a median age of 13 to 16 years. Survival in children who present with metastatic disease remains dismal despite the most aggressive chemotherapeutic treatment including use of autologous bone marrow transplantation.

The systemic nature of Ewing's sarcoma is supported by the treatment results achieved prior to the development of chemotherapy. Radiotherapy achieved local control in 50 to 85 per cent of patients treated with more than 40 Gy, but only 15 to 20 per cent of these patients would survive 2 to 3 years and only 10 per cent would survive 5 years because of distant relapse. The usual pattern of relapse is hematogenous, primarily to the lung and bone. Hence, the treatment of these tumors must achieve control of both local and systemic disease. Dissemination must be assumed at diagnosis.

In 1979 Askin reported a group of tumors of soft tissue origin arising in the thoracopulmonary region. Characteristic features included a high incidence of local recurrence, lower risk of dissemination than in Ewing's sarcoma, and neuroepithelial derivation suggested by electron microscopy. It has been demonstrated that the classic Ewing's tumors and the small, round cell tumors with neural differentiation have similar responses to therapy. The German Society of Paediatric Oncology performed a retrospective evaluation of 30 patients in their Ewing's sarcoma trial in whom characteristics of primitive neuroectodermal tumors were identified. Nineteen of the 30 were in the chest wall and 20 had bony involvement. The disease-free survival for the patients with localized disease was 52 per cent, comparable with that of Ewing's sarcoma. In this study ‘radical surgery' with complete removal of the tumor in conjunction with chemotherapy was of prognostic significance. All 13 patients with complete excision remain disease free. These findings are in contrast with the earlier report by Askin which suggested an increased risk for local recurrence. Review of the Askin study, however, shows that the majority of the tumors in that report had only partial resection (only three of 20 patients had complete resection). Others have found a similar response to treatment of the these tumors.

The absence of our ability to differentiate these tumors extends to the cytogenetic level. Both contain the same balanced translocation between chromosomes 11 and 22 [t(11, 22) (q24, q12)]. The translocation point of these tumors has now been cloned and is identical. Finally, both tumors contain high concentrations of the protein product of *mic* 2 gene on their surface. Since these two entities cannot be distinguished histologically, clinically, or cytogenetically, most consider them as the single entity of malignant, small, round cell tumors.

Local control of Ewing's sarcoma traditionally has been managed by radiotherapy or surgical resection. Ewing, in his early descriptions of this tumor, defined its radiosensitivity. However, local treatment requires relatively high-dose radiation. Limited tumor size, diameter (< 8 cm), and volume (< 100 ml) correlates with improved local control. No randomized study has been performed between surgical resection and radiotherapy. In non-randomized trials the improved survival in the surgical resection group has been attributed to the inclusion of larger tumors in the radiation group.

In contrast to some primary sites of Ewing's sarcoma, chest wall tumors can generally be removed. Preliminary treatment with chemotherapy will often significantly shrink the tumor and lead to the resection of fewer ribs (Fig. 8). As the ribs are generally ‘expendable', the entire rib from the costochondral junction to the vertebral insertion can be removed. In cases where the lesion is very lateral or anterior, preservation of the posterior segment of the ribs may decrease the risk or extent of scoliosis. One must be aware, however, that Ewing's sarcoma can have ‘skip lesions' within the marrow cavity. This possibility should be excluded by either bone scan or MRI.

Rhabdomyosarcoma

Rhabdomyosarcoma is a common solid tumor in infants and children. In the first Intergroup Rhabdomyosarcoma Study (**IRS-I**) it occurred in the trunk in only 7.4 per cent of cases. Of those, half involved the chest wall, one-third involved the paraspinous region, and one-sixth involved the abdominal wall. In this study there were several cases classified as Ewing's sarcoma and undifferentiated sarcoma in addition to the standard rhabdomyosarcomas. Alvelolar histology (seven cases) predominated over embryonal (two cases) and undifferentiated histology (two cases). Survival in these cases correlated with the stage at presentation, but was generally less favorable than in other sites. Patients with metastatic disease (group IV) did poorly and none survived.

Specific considerations for the chest wall are no different than for rhabdomyosarcoma in other locations. Efforts should be made to remove all visible and microscopic tumor whenever possible, frequently after shrinkage by chemotherapy.

More recent analysis of children with thoracic lesions treated on IRS-II and IRS-III documents the continued difficulty in curing these children. Sixty children (71 per cent) achieved a complete response, seven a partial response, three showed improvement, and 14 had no response or progressed. Of the sixty children with a complete response, 28 had a local recurrence and 11 had a distant recurrence.

In a multivariate analysis only clinical group ($p < 0.002$), local ($p < 0.006$), and distant recurrence ($p < 0.001$) showed statistical significance. Ironically, a higher rate of recurrence occurred in group I (54 per cent) than group II (27 per cent) suggesting difficulties in accurate pathologic assessment of margin status. As patients in group I were felt to have complete tumor resection, most did not receive radiation. In contrast, all but two of the patients in group II received 4000 to 4500 cGy to the primary site. These findings suggest that the adequacy of local tumor resection must be very carefully evaluated and may require re-resection for confirmation of margin status and to achieve complete resection.

Surgical resection

The importance of surgical resection has been stressed in prior discussions, particularly for Ewing's sarcoma, osteosarcoma, and chondrosarcoma. Incisions must be appropriately placed to allow preservation of overlying skin and if possible muscle flaps. In many cases, resection of one or multiple ribs is required. Multiple techniques have been used for reconstruction. Prosthetic materials including Marlex mesh (Bard Vascular Systems Division, Billerica, MA) and Gortex tissue patch (W.L. Gore, Co., Flagstaff, AZ) have been used to replace missing sections of the chest wall. In larger lesions, a ‘sandwich' created using two sheets of Marlex mesh and methyl methacrylate can be used. This composite can be shaped to match the contour of the chest wall and produce a solid reconstruction. The major deterrence to use of these prosthetic materials in the young infant or growing child is that they do not expand as the child grows and may result in more severe scoliosis than occurs in defects closed with living tissue. The goals of chest wall resection and reconstruction as defined by Grosfeld include complete removal of the local tumor, restoration of adequate protection of the thoracic viscera, restoration of physiologic function providing for adequate lung and chest wall growth, and an acceptable chest wall

appearance. These should be the aims of all surgical resections, but local control must remain the primary goal.

Diaphragm

Bochdalek hernia

Congenital absence of a portion of the diaphragm results in herniation of the abdominal viscera into the thoracic cavity. Diaphragmatic defects occur about once in every 4000 live births and once in every 2200 births if stillbirths are included. Ninety per cent of these herniations are through the posterior part of the diaphragm, the foramen of Bochdalek.

Most patients suffer respiratory distress in the immediate neonatal period. This results from pulmonary hypoplasia and compression of the pulmonary parenchyma by abdominal viscera in the thoracic cavity. Despite significant advances in surgical and anesthetic techniques, and in postoperative ventilatory management, the mortality rate remains high, and treatment of congenital diaphragmatic hernia is one of the major unsolved problems in pediatric surgery.

Congenital diaphragmatic hernia has historically been considered as an isolated defect. However, in one recent study, up to 40 per cent of infants had one or more severe malformations involving the heart, brain, genitourinary system, or craniofacial region. Cardiac defects far outnumbered all others accounting for 65 per cent of all associated malformations. Survival of patients with additional anomalies is particularly poor.

Embryology

The septum transversum grows posteriorly to meet the dorsal mesentery of the foregut, forming the central portion of the diaphragm during the fourth to eighth weeks of fetal life. The lateral folds of peritoneum and pleura develop simultaneously, completing the separation between the thorax and the abdomen. Disruption of this process results in a posterolateral defect in the diaphragm (foramen of Bochdalek). This closure of the pleuroperitoneal canal is completed on the right side before the left, which may explain why 90 per cent of the diaphragmatic defects present on the left side.

As the diaphragm is developing, the midgut undergoes elongation and development outside the celom. The midgut normally returns to the abdominal cavity and undergoes rotation and fixation at about the tenth week of fetal life. If closure of the diaphragm is incomplete, the intestine herniates into the chest as it returns to the celom, inhibiting development of the lung and preventing the normal process of intestinal rotation and fixation. The spleen, stomach, and left lobe of the liver as well as the bulk of the intestine often reside within the thoracic cavity.

Pathophysiology

Herniation of the abdominal viscera into the thoracic cavity seriously impairs development of the ipsilateral, and to some extent the contralateral, lung (Fig. 10). This hypoplasia of the pulmonary parenchyma accounts for the major physiologic problems encountered in these infants. Creation of a diaphragmatic hernia in fetal lambs results in a decrease in the total size of the lung, fewer vessels per unit area of lung, and increased muscularization of the pulmonary arterial tree. These findings are identical to those in the affected human infant (Fig. 11). The degree of ‘pruning' or hypoplasia of the pulmonary vessels correlates with the adequacy of ventilation. If the pulmonary parenchyma is inadequate, mechanical ventilation will not prevent severe hypoxia, hypercarbia, and acidosis. Occasionally, adequate ventilation can be achieved initially (the so-called honeymoon period), indicating that enough pulmonary parenchyma is present to sustain life. Later, arterial blood gas balance deteriorates, with the development of pulmonary hypertension and right to left shunting through the patent ductus arteriosus (persistent fetal circulation). In this situation a ‘reactive' pulmonary circulation with progressive vasoconstriction and pulmonary hypertension decreases perfusion of the lung and results in clinical deterioration. Traditional vasodilating agents do not alter this pattern of persistent fetal circulation. They dilate the systemic as well as the pulmonary circulation, causing systemic hypotension; this creates a need for additional fluid administration, which in turn results in further pulmonary edema and deterioration. Recent studies utilizing inhaled nitric oxide demonstrate selective pulmonary vasodilatation which has been very clinically useful in these infants. However, in the only prospective randomized trial of nitric oxide immediate short-term improvements in oxygenation were seen in some infants, but the need for extracorporeal membrane oxygenation and overall survival was not improved.

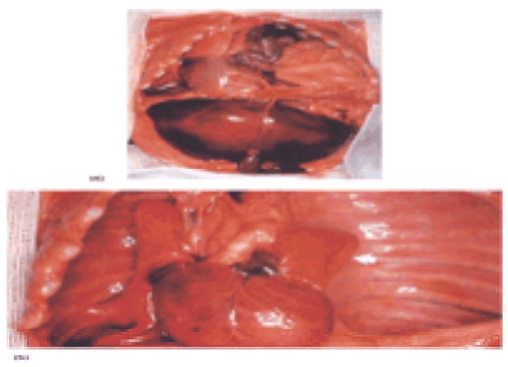


Fig. 10. (a) Photograph from an autopsy of an infant with congenital diaphragmatic hernia demonstrates the usual intrathoracic location of the abdominal viscera entering through a defect in the left diaphragm with only the liver remaining in the abdominal cavity. (b) With the abdominal viscera removed, the markedly hypoplastic lung bud (solid arrow) on the left side can be seen. Hypoplasia of the right lung (open arrow) accounted for this infant's rapid death.

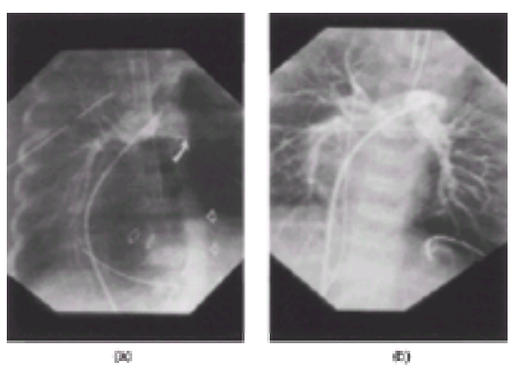


Fig. 11. (a) Pulmonary arteriogram in an infant with congenital diaphragmatic hernia with severe hypoplasia of the left lung (solid arrow). Rapid filling of the aorta (open arrow) demonstrates a right to left shunt resulting from pulmonary hypertension. This infant could never be adequately ventilated and died. (b) Pulmonary arteriogram in a second infant with congenital diaphragmatic hernia with limited ‘pruning' of the pulmonary vessels of the left side and normal vessels on the right side. This infant was readily weaned from the ventilator following surgical repair of the defect.

Clinical picture

Most infants develop respiratory symptoms in the first 24 h after birth. Although a spectrum of respiratory distress exists, many children immediately become severely cyanotic. Physical examination discloses a scaphoid abdomen, displacement of the cardiac apex away from the side with the defect, and decreased or absent breath sounds on the affected side. After birth, air entering the gastrointestinal tract distends the herniated bowel, shifting the mediastinum toward the contralateral side and compressing the contralateral lung. This process can be largely prevented by sump catheter decompression of the stomach (Fig. 12). Rarely, bowel sounds can be heard in the chest. The occasional infants who become symptomatic weeks to months after birth respond well to surgical treatment.

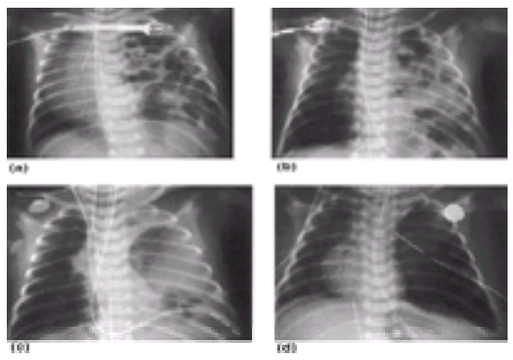


Fig. 12. (a) Initial chest radiograph obtained shortly after birth in an infant with congenital diaphragmatic hernia. The infant required intubation in the delivery room. Note air-filled intestines in the left thoracic cavity and shift of the mediastinum and heart to the right following resuscitation. (b) One hour later after the stomach has been decompressed and the infant ventilated with small tidal volumes the mediastinum has returned to a more normal midline position. (c) The infant was managed with delayed surgery. Forty-eight hours after birth, continued improvement in decompression of the intestines is seen as well as appearance of some pulmonary tissue on the left side. This infant suffered severe deterioration in his respiratory status by conventional ventilation resulting from development of pulmonary hypertension. He required extracorporeal membrane oxygenation for 5 days, after which he was returned to standard ventilation and underwent surgical repair without difficulty. (d) Chest radiograph following repair in this infant who was then readily weaned from the ventilator.

Diagnosis

The diagnosis can almost always be established on an upright thoracic radiograph (Fig. 12), which shows multiple loops of intestine in the thoracic cavity and no diaphragmatic outline. Differential diagnoses include cystic adenomatoid malformation, eventration of the diaphragm, pneumatoceles from staphylococcal pneumonia, and pulmonary agenesis or hypoplasia. The radiographic appearances of all of these entities include presence of the diaphragm and a normal intra-abdominal bowel gas pattern. As antenatal ultrasound is applied more widely a growing number of infants with congenital diaphragmatic hernia are being identified *in utero*. Antenatal diagnosis has been considered an adverse prognostic indicator, and recent evaluations by Wilson and colleagues have elucidated the basis for this finding. A high incidence of associated anomalies were identified in this cohort of infants, over 50 per cent incidence in infants diagnosed before 25 weeks gestation. The increased mortality was related to the associated anomalies, and infants with isolated diaphragmatic hernia diagnosed antenatally had no greater risk of mortality than postnatally diagnosed infants.

Treatment

Initial resuscitation is mandatory if these infants are to survive. Intubation for respiratory distress is often required before a radiograph is available. A nasogastric tube should be inserted and placed on suction to prevent further distention of the intestine. Because the hypoplastic lungs are susceptible to barotrauma, ventilation with high pressure must be avoided. A tension pneumothorax on the contralateral side often proves fatal. Historically, pulmonary management utilized rapid ventilation (60–150 breaths/min) with short inspiratory times, lower pressure (mean 45 cmH_2O) and 100 per cent oxygen. Alkalosis, hypocarbia, and oxygenation all decrease pulmonary artery pressures and the right to left shunt seen in persistent fetal circulation. Goals of ventilation in the past were a PaO_2 above 100 torr, $PaCO_2$ below 30 mmHg, and pH above 7.4. Vacanti and colleagues have stressed the importance of sedating these infants preoperatively and postoperatively with fentanyl and pancuronium combined with rapid frequency ventilation and moderate fluid restriction. Sedation seems to benefit those with very reactive pulmonary vasculature, but it does not help those with inadequate pulmonary tissue, who develop early pulmonary hypertension and acidosis. The infant must be kept warm during transport to avoid peripheral vasoconstriction and acidosis: the latter can elevate pulmonary artery pressure. Wung and colleagues have promoted a different approach in these infants. They stress the need to avoid barotrauma and maintain the peak inspiratory pressures at 20 to 25 cmH_2O. Infants are allowed to breath spontaneously and $PaCO_2$ up to 60 mmHg was accepted. Recently, Wilson and colleagues noted a 25 per cent improvement in survival and a smaller proportion of children required extracorporeal membrane oxygenation when the traditional hyperventilation/alkalosis protocols were abandoned in favor of the permissive hypercapnea model proposed by Wung. This has now become the new standard of care.

High-frequency ventilatory support may improve oxygenation while minimizing barotrauma to the lung. Bohn and colleagues studied 58 infants with congenital diaphragmatic hernia presenting within the first 6 h of life. Infants with CO_2 retention and severe preductal shunting were unresponsive to hyperventilation and had a 90 per cent mortality rate. The second group of infants responded well to hyperventilation and had reversible ductal shunting: 97 per cent survived. High-frequency ventilation (375 to 1800 cycles/s) has been unsuccessful in infants with inadequate response to lower-frequency ventilation. Inhaled nitric oxide is now utilized in children who develop pulmonary artery hypertension and declining PaO_2. It has been helpful in stabilizing some of these infants.

Extracorporeal membrane oxygenation has been used in infants in whom all ventilatory efforts to achieve adequate oxygenation and alkalosis have failed. Criteria for the use of this technique vary significantly between centers, but have traditionally been based on arterial–alveolar oxygenation differences (A-a DO_2) or on oxygenation index (OI = mean airway pressure × FiO_2/PO_2). More recently, with the wider adoption of permissive hypercapnea and greater reliance on preductal PO_2s, the indications for extracorporeal membrane oxygenation have been inadequate oxygen delivery to tissues despite optimized ventilation. Inadequate delivery is defined as preductal saturations below 90 per cent on 100 per cent FiO_2 with an evolving metabolic acidosis. Infants with any evidence of intracranial hemorrhage on cranial ultrasound or a gestational age of under 34 weeks were traditionally excluded since they were felt to be at increased risk for cerebral bleeding during heparinization. Wilson and colleagues have subsequently demonstrated the efficacy of the use of aminocaproic acid to reduce intracranial hemorrhage and other hemorrhagic complications of extracorporeal membrane oxygenation, thereby allowing smaller infants and infants with either grade I or II intracranial hemorrhage to also benefit from this therapy.

While extracorporeal membrane oxygenation has been successful in infants with persistent fetal circulation following meconium aspiration, neonatal respiratory distress syndrome, and sepsis, it has been much less successful in infants with congenital diaphragmatic hernia. In one study, extracorporeal membrane oxygenation produced a survival rate of 90 per cent among infants with medical causes of acute respiratory failure, but of only 38 per cent for those with congenital diaphragmatic hernia. Presumably this treatment cannot be maintained for long enough to allow lung growth in infants whose pulmonary hypoplasia is of a degree incompatible with life. It may be useful in infants with mild hypoplasia whose condition deteriorates after birth because of pulmonary artery hypertension and fetal circulation. Reports by Wilson and colleagues have demonstrated that occlusion of the trachea *in utero* will produce increased growth of the lung and will in fact overcome the impairment of the hernia on pulmonary growth. Occlusion of the trachea by a plug in the fetus has also been attempted by Harrison and colleagues based on the work of Wilson. It remains investigational at this time and is plagued by problems with premature labor and abortion. More recent attempts at endoscopic tracheal occlusion have shown slightly improved results and a 50 per cent overall survival for isolated congenital diaphragmatic hernia.

Surgical repair

Although surgical repair was traditionally performed urgently, recent studies have shown that ventilatory compliance decreases appreciably in infants after surgery, producing a decline in the arterial gases. The current practice is to stabilize these infants with ventilation, sedation, and intestinal decompression, deferring repair of the defect until 36 to 72 h after birth. Delay may avoid the development of pulmonary hypertension and persistent fetal circulation due to surgical stress.

A transverse upper abdominal incision is created on the side of the diaphragmatic defect and the abdominal viscera are withdrawn from the chest by gentle traction. A retractor elevating the upper diaphragmatic leaf facilitates this maneuver by allowing air to enter the thorax as the intestines are withdrawn. The defect in the diaphragm is closed with interrupted mattress sutures of 3–0 non-absorbable synthetic material. The posterior leaf of the diaphragm must be dissected free from the peritoneum and retroperitoneal tissues that tether it posteriorly. The diaphragm is often better developed than is first thought. In the rare situation where no posterior leaf exists, sutures are placed around the tenth and eleventh ribs to provide solid fixation. If the diaphragm is inadequate for primary closure, a prosthetic patch of tissue Gore-Tex is used. Occasionally a sac will be present: this should be excised prior to closure of the diaphragm. In these cases the defect is not a congenital Bochdalek hernia, but rather a large eventration of the central diaphragm. Almost every child with a diaphragmatic hernia who requires extracorporeal membrane oxygenation will require a patch repair. Wilson and associates identified that most of the postoperative bleeding came from mobilization of the posterior leaf of the diaphragm. Hence, they suggest that repair on extracorporeal membrane oxygenation be modified by not dissecting up the posterior leaf of the diaphragm, and using a slightly larger patch. This modification along with aminocaproic acid has resulted in essentially no bleeding in their patients with diaphragmatic hernia repaired on extracorporeal membrane oxygenation.

Although some surgeons prefer a thoracic incision for repair, the abdominal approach allows the intestines to be withdrawn from the abdominal cavity as the diaphragmatic defect is closed. It also allows placement of a gastrostomy tube for intestinal decompression and early feeding. Malrotation is generally present, and the

duodenum should be inspected for obstructing Ladd's bands. The abdominal cavity may be small in babies in whom the bowel has resided in the thorax *in utero*; the abdominal wall can be stretched manually to facilitate closure. If primary closure produces excessive pressure, a ventral hernia can be created by closing only the skin or by the use of a silon pouch, but these techniques are rarely required. If a prosthetic closure is necessary, final closure of the abdominal wall can usually be achieved in several days. A chest tube may be inserted into the pleural cavity prior to closure. It is placed on water seal and not to suction to avoid any excessive shift of the mediastinum, which may produce kinking of the mediastinal vessels. If the condition of the infant is unstable, a chest tube on the contralateral side will prevent fatal tension pneumothorax. However, some centers have adopted Wung and Stolar's approach to avoid the routine use of a chest tube, which may drain significant amounts of serum complicating fluid management. Extralobar pulmonary sequestration in association with congenital diaphragmatic hernia is rare: when present the sequestration should be removed.

Postoperative care

Although the diaphragmatic hernia can usually be repaired, if both lungs are markedly hypoplastic, adequate oxygenation will never be achieved. Survival may be compromised in babies with relatively good lungs if pulmonary artery hypertension occurs. Extracorporeal membrane oxygenation can decrease the degree of pulmonary artery hypertension by producing adequate oxygenation, alkalosis, and hypocarbia. All three factors will lower pulmonary artery pressures, and this technique may save some infants who previously have succumbed to persistent fetal circulation.

Ventilatory support is nearly always needed following repair. It must be regulated to limit inspiratory pressures and minimize barotrauma to the lungs. The hypoplastic lung bud should not be distended artificially. Close monitoring of arterial blood gases is essential: deterioration often indicates rising pulmonary artery pressures. We go out of our way not to alkalinize the patients allowing the pH to be as low as 7.25 and the PCO_2 as high as 60 mmHg. The preductal saturations are followed, completely ignoring the shunt and the postductal saturations. The critical factor is adequate oxygen delivery to the tissues and the lack of progressive metabolic acidosis. While a right to left shunt may have some negative impact, the therapy to address it (hyperventilation, alkalosis, barotrauma, and fluid overload) are far worse.

Acute deterioration may be due to a pneumothorax on the contralateral side: this should be confirmed radiographically and treated with intercostal tube drainage on 10 to 15 cmH_2O suction.

Results

Survival rates vary widely, but most major centers are reporting around 70 per cent overall survival for patients symptomatic in the first hours of life; however, survival has been reported to be as high as 85 to 95 per cent for isolated diaphragmatic hernia. The apparent failure of results to improve can be attributed to the fact that severely affected babies who died before referral in the past are now surviving, due to antenatal diagnosis and rapid transport.

Late pulmonary function studies have shown no abnormality in total lung capacity, vital capacity, or diffusion, but have demonstrated reduced blood flow and a permanent reduction in pulmonary vessels in the ipsilateral lung.

Morgagni hernia

Failure of fusion of the central and lateral portions of the diaphragm anteriorly results in a defect known as the foramen of Morgagni. This is far less common than the foramen of Bochdalek hernia (less than 2 per cent) and has less physiologic significance. These retrosternal hernias usually have a true sac and cause few respiratory symptoms: intestinal obstruction is a more common mode of presentation.

The diagnosis is suggested by the presence of air or fluid levels in the retrosternal region on an upright thoracic radiograph (Fig. 13). Barium contrast studies may demonstrate the presence of intestine in the sac, but are rarely necessary. The hernia is best repaired through a transabdominal approach, with excision of the sac and primary closure of the defect in the diaphragm. Because there is little compression of the lungs during fetal development, pulmonary hypoplasia or respiratory compromise after repair are rare.

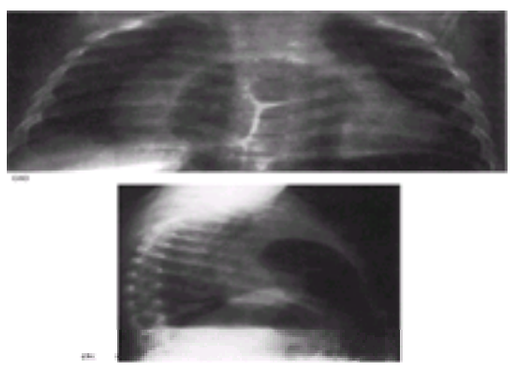


Fig. 13. (a) Posterior–anterior chest radiograph obtained in a 4-month-old infant because of a cold. To the surprise of the pediatrician an air-filled structure was identified anterior to the heart. (b) The lateral chest radiograph confirmed the presence of an air-filled structure within the mediastinum. This child was explored through a transverse upper abdominal incision. A diaphragmatic hernia of Morgagni was found with a sac containing transverse colon. The sac was excised and the hernia readily closed. The pericardium was separate from the hernia sac.

Eventration of the diaphragm

Eventration of the diaphragm may be congenital, resulting from failure of normal ingrowth of muscle into the developing diaphragm or it may be acquired, resulting from phrenic nerve injury. Whatever the cause, there is an abnormally redundant and attenuated diaphragm which has little or no ability to contract. Ninety per cent of these lesions are on the left side. Symptoms vary greatly: some patients suffer severe respiratory compromise, while in others it is an asymptomatic incidental finding. The diagnosis is best confirmed by fluoroscopy, which demonstrates paradoxical motion of the diaphragm. Flexibility of the mediastinum in infants permits transmission of the paradoxical motion of the diaphragm to the contralateral side.

The need for repair is determined primarily by the symptoms produced by the eventration, and not by the radiographic findings. A small eventration can be repaired easily through a limited, low thoracotomy, the defect being reefed up by plication with non-absorbable sutures until the diaphragm is taut. A large eventration has a similar radiographic appearance to a Bochdalek hernia, the thorax being filled with abdominal viscera. This is best repaired through an abdominal incision.

Tracheobronchial tree

Congenital lesions

Tracheal agenesis is a uniformly fatal lesion in which a segment of the trachea is absent or contains no lumen. Severe respiratory distress invariably occurs at delivery. Intubation via the larynx is unsuccessful, but satisfactory ventilation following esophageal intubation when an esophagotracheal fistula is present may allow the infant to survive for several hours. Diagnosis is confirmed by endoscopy. Associated complex cardiovascular anomalies are often present. Reconstruction has been attempted using the esophagus as a trachea, with division of the esophagus proximally to create a ‘spit fistula' and division of the stomach distally. There have been no survivors.

Tracheal stenosis is more amenable to surgical repair. If only a short segment of trachea is affected, dilatation of a membranous web or its resection may be successful. Extensive lesions involving a greater length of the trachea were first treated by a tracheoplasty with a costal cartilage graft in 1982; this has produced favorable results since then. Up to 50 per cent of these infants have associated cardiovascular malformations.

Infants with tracheomalacia have abnormal tracheal compliance due to increased pliability or hypoplasia of the tracheal cartilages. Stridor is usually present shortly after birth, but in most infants respiratory embarrassment is limited. Fluoroscopy demonstrates collapse of the trachea on vigorous breathing or crying. The normal trachea collapses if increased airway resistance is present. All of these infants must be evaluated endoscopically to exclude extrinsic obstructions. Bronchoscopy performed while the infant is breathing spontaneously will demonstrate tracheal collapse with inspiration. Tracheomalacia occurs most frequently in association with

tracheo-esophageal fistula. Compression of the trachea by the distended proximal esophageal pouch *in utero* may produce this flexibility; alternatively, it may result from a diffuse abnormality of tracheal development. Abnormal separation of the esophagus and trachea and hypoplasia of the tracheal rings are both components of development. Tracheomalacia may produce episodes of cyanosis following coughing spells or crying. In most infants the condition can be managed expectantly since the trachea becomes larger and the cartilage stiffens as the child grows.

Laryngotracheo-esophageal cleft is a rare anomaly in which the posterior wall of the larynx and a variable length of trachea fail to separate from the esophagus, resulting in a proximal communication between the airway and the esophagus. It presents with recurrent episodes of aspiration and coughing on feeding. Diagnosis is often delayed. Contrast studies often have the appearance of contrast medium ‘spilling over' from the esophagus into the trachea. The cleft is best demonstrated endoscopically, with a catheter in the esophagus and an endoscope in the trachea. Distension of the esophagus with air blown into the catheter causes separation of the cleft beginning at the arytenoids and extending distally. Repair is performed through the neck, closing the trachea without narrowing it by using a small strip of esophagus in the closure, interposing a strap muscle between the esophageal and tracheal closures to prevent recurrence. Extensive lesions involving much of the trachea are repaired through a combined cervical and thoracic approach and have only recently been successful.

Tracheobronchial and esophageal foreign bodies

An inhaled foreign body is commonly misdiagnosed. The sudden onset of coughing should suggest aspiration of a foreign body, even when a physical examination and chest radiograph are negative. Fluoroscopy and/or inspiratory and expiratory films increase the sensitivity of radiographic examinations, but normal results cannot entirely exclude the presence of a foreign body. Air trapping with hyperinflation of the lung, shift of the mediastinum, or decreased movement of the diaphragm support the presence of a retained foreign body. Most objects aspirated into the airway are not radio-opaque. In trained hands, bronchoscopic examination carries little risk and should be performed in any infant with a history suggestive of a foreign body. The sudden onset of wheezing in an infant or child should also raise the question of foreign body aspiration. Pneumonia after an episode of coughing is an indication for bronchoscopy after initiation of antibiotic treatment.

Fiber-optic endoscopic equipment of suitable size for infants and children must be available, along with a set of foreign body instruments. Friable material, such as the frequently encountered peanut, can be removed with gentle foreign body forceps. The Fogarty arterial embolectomy catheter (3 or 4 French) can be used to remove an object that is likely to crumble if grasped by forceps. The catheter is inserted through the side port of the endoscope and it is passed beyond the object under direct vision. After inflation of the catheter balloon, the catheter and foreign body are withdrawn into the pharynx. Occasionally, if the endoscope is large enough, the foreign body can be pulled into the endoscope and extracted with it. Lobectomy may be required if successful extraction of the foreign body is not achieved and if the lung tissue distal to the foreign body has been destroyed by recurrent infection and bronchiectasis. Peanuts are especially troublesome, and exuberant granulation tissue develops around them.

Swallowed foreign bodies lodged in the esophagus can cause respiratory symptoms as well as symptoms of esophageal obstruction. They usually lodge just below the level of the cricopharyngeus, behind the larynx, or at the level of the aortic arch. The history generally suggests ingestion and, as most objects swallowed are radio-opaque, a chest radiograph will usually confirm the diagnosis.

A patient with a foreign body in the esophagus should be intubated to secure the airway before endoscopic removal of the foreign body is attempted. Coins, the foreign body most commonly lodged in the esophagus, can be readily extracted with toothed foreign body forceps. Esophagoscopy can be facilitated by gentle insufflation of oxygen through a catheter inserted through the side port of the esophagoscope. Insufflation distends the esophagus and allows better visualization. Retained secretions can also be aspirated through this catheter.

Respiratory symptoms may also occur if a foreign body perforates the esophagus, producing an abscess and granulation tissue between the trachea and esophagus. Management is generally limited to endoscopic extraction of the foreign body if possible. The inflammatory mass will resolve along with the respiratory symptoms and cervical or thoracic drainage is rarely required.

Pulmonary

Congenital lesions

A variety of pulmonary lesions, ranging from agenesis of a lung to the more frequent cystic or hamartomatous malformations, result from abnormal parenchymal development. These malformations present primarily as space-occupying lesions in an infant with decreased pulmonary function. Less often, recurrent pulmonary infections may be the mode of presentation.

Pulmonary agenesis

In pulmonary agenesis the bronchus and pulmonary tissue are absent. The etiology is unknown but agenesis frequently occurs in association with other complex anomalies, particularly of the heart and great vessels. Many affected children are asymptomatic, but tachypnea may occur with crying or exercise. Pulmonary agenesis is associated with predisposition to recurrent pulmonary infections, although the mechanism for this is not understood. Death usually results from associated lesions: the combination of pulmonary agenesis and tracheo-esophageal fistula has a particularly poor prognosis. Tracheal stenosis also occurs in association with agenesis of the lung. The diagnosis can be established on plain radiographs which demonstrate a shift of the mediastinum to the affected side, which is entirely opacified (Fig. 14). Distinction must be made between this lesion and a pleural effusion or empyema, which would shift the mediastinum to the contralateral side.

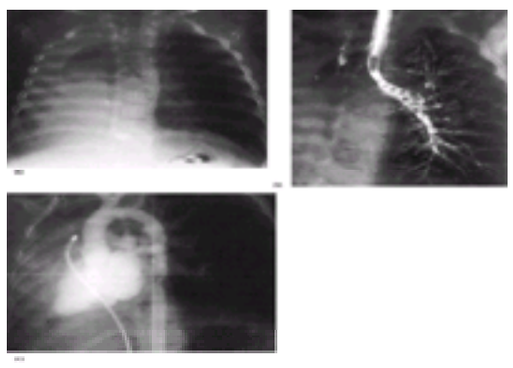


Fig. 14. (a) Chest radiograph obtained because of mild respiratory distress in the immediate newborn period. On physical examination the heart was found in the right chest and no breath sounds were audible on that side. The chest radiograph revealed a shift of the mediastinum into the right chest which was felt to be consistent with pulmonary agenesis. (b) At 3 months of age a bronchogram was performed which revealed complete absence of the right bronchus. (c) An arteriogram was subsequently obtained after the child had suffered multiple pulmonary infections. It demonstrated the heart occupying the right hemithorax and a single pulmonary artery to the left lung.

Pulmonary hypoplasia

In pulmonary hypoplasia the mainstem bronchus and proximal divisions are normal, but the amount of parenchyma of one or both lungs is decreased. Pulmonary hypoplasia is most common in patients with congenital diaphragmatic hernia, in whom compression of the lung impairs normal development. The severity of symptoms correlates with the degree of hypoplasia.

Pulmonary cyst and congenital cystic adenomatoid malformation

The simplest pulmonary abnormality is an isolated pulmonary cyst. Historically, many of these were thought to be acquired lesions secondary to infection; this is not true in most cases. Lung cysts occur in young infants in the absence of infection and are lined with tall, ciliated columnar epithelium, appropriate for the respiratory tree (Fig. 15). They generally communicate with the bronchial tree, can occur in either lung, and affect only one lobe. Other pulmonary abnormalities may be present; the arterial blood supply for the affected lobe may originate from the aorta rather than from the pulmonary artery. In some cases hamartomatous development is evident, and abnormal epithelium is present; the distinction between lung cyst, intralobar pulmonary sequestration, and cystic adenomatoid malformation is often blurred. Patients with cystic lesions may present with secondary infection. It can be difficult to distinguish between a lung abscess, resolving staphylococcal pneumonia with a

pneumatocele, or a loculated empyema. An underlying anatomic abnormality must be suspected in a patient with ‘pneumonia' that does not resolve completely or that recurs in the same anatomic location.

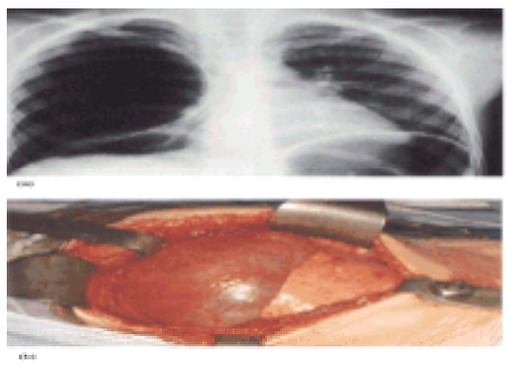


Fig. 15. (a) Chest radiograph obtained because of croup in a 4-year-old female. It revealed a large air-filled pulmonary cyst occupying at least half of the hemithorax. (b) Operative findings of a large thin-walled air-filled cyst within the right upper lobe. A limited amount of normal pulmonary parenchyma of the upper lobe was ‘splayed' out over the superior surface of the cyst. Right upper lobectomy was performed.

Congenital cystic adenomatoid malformation covers a broad spectrum of disease, with varying proportions of cysts and adenomas. Cystic adenomatoid malformations have been classified into three types: type I represents single or multiple large cysts (more than 2 cm in diameter) with prominent smooth muscle and elastic tissues in the wall; type II lesions are composed of multiple small cysts (less than 1 cm in diameter) lined by ciliated cuboidal to columnar epithelium—these are frequently associated with other congenital anomalies; the type III lesion is a large, bulky, non-cystic mass that often produces a mediastinal shift and which has a poor prognosis. The mass effect caused by these lesions *in utero* results in pulmonary hypoplasia, which may be severe (Fig. 16).

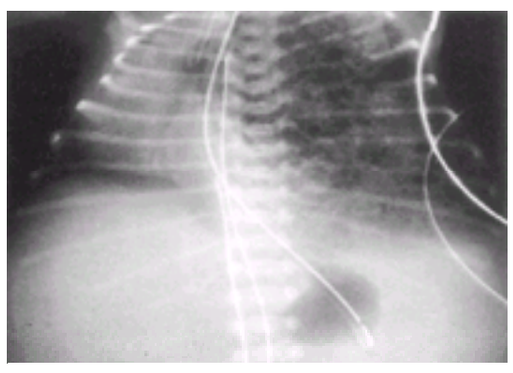


Fig. 16. Chest radiograph in a newborn infant with cystic adenomatoid malformation of the left lung first identified on an antenatal ultrasound. It was a very large lesion producing a marked shift of the mediastinum to the right side. The infant developed severe respiratory distress in the delivery room requiring intubation and emergency surgical intervention. Despite resection of the mass the infant could not be adequately ventilated because of severe hypoplasia of the contralateral right lung as well as of the left lung. In this case the cystic adenomatoid malformation was so large that it produced bilateral pulmonary hypoplasia.

Many infants present in the first month of life with respiratory distress; others present late, with cough and fever due to secondary infection of the lesion. Treatment requires resection of the abnormal pulmonary tissue. Prognosis is related directly to the presence of associated anomalies and the amount of normal pulmonary parenchyma. Antenatal diagnosis of cystic adenomatoid malformation is possible, and prognosis correlated with the type of lesion. Microcystic lesions are often associated with fetal hydrops and have a poor prognosis (80 per cent mortality), while the macrocystic lesions, which are rarely associated with hydrops, have a much better prognosis (29 per cent mortality). Early diagnosis allows delivery in a specialist unit and early surgical intervention. *In utero* intervention with resection has been successful in some of these poor prognosis lesions. Rhabdomyosarcoma has been reported as a rare occurrence in children with congenital cystic adenomatoid malformation.

Pulmonary sequestration

Pulmonary sequestration is characterized by failure of the pulmonary tissue to communicate with the tracheobronchial tree, and the presence of an aberrant arterial supply to the pulmonary tissue. This often arises directly from the thoracic aorta (85 to 90 per cent of cases) and less frequently from the abdominal aorta. Multiple systemic arteries supply the sequestration in 15 per cent of patients. The venous drainage from a sequestration may be through systemic (azygos) or pulmonary veins.

An extralobar sequestration occurs when an isolated segment of pulmonary tissue is surrounded by pleura and is distinct from the normal lobes. It has no communication with the bronchus and is not susceptible to contamination by surrounding ventilated pulmonary tissue: such sequestrations rarely become infected. The systemic arterial supply to the segment can, however, result in high-output cardiac failure from shunting through the sequestration (Fig. 17). Extralobar sequestration is most commonly present in the posterior mediastinum at the costophrenic sulcus, and is present on the left side in two-thirds of affected patients.

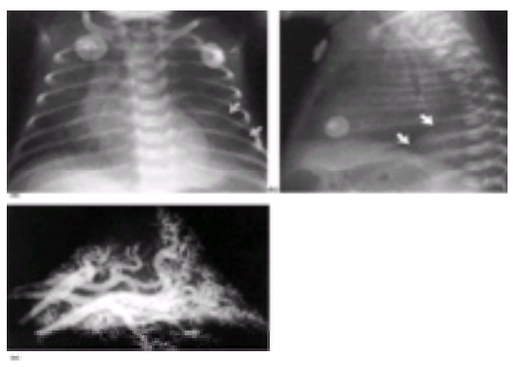


Fig. 17. (a) Posterior–anterior chest radiograph in a newborn full-term infant in congestive heart failure with a systolic ejection murmur best heard posteriorly on the right side. The right diaphragm is obscured by a density in the lower chest (arrows). (b) Lateral chest radiograph demonstrates a posterior triangular defect most consistent with an extralobar pulmonary sequestration (arrows). The child was in borderline congestive heart failure. A magnetic resonance image (MRI scan) showed several large systemic vessels feeding into the sequestration. At surgery four large vessels were identified coming directly from the aorta to the sequestration. (c) Arteriogram of the resected specimen shows the large dilated vessels which can produce a significant systemic shunt in these infants. After surgery the murmurs were no longer audible and digoxin and diuretics were stopped.

Intralobar sequestrations are more common (85 per cent of cases), and occur within a pulmonary lobe, usually a lower lobe that has a normal bronchial supply (Fig. 18). The intralobar sequestration lacks normal bronchial communication, but often has microscopic communications with the surrounding ventilated tissue, and can become secondarily infected. These lesions frequently present as recurring localized pulmonary infections. The arterial supply of the intralobar sequestrations generally arises from the aorta. Perfusion from both the pulmonary artery and aorta distinguishes sequestrations from other pulmonary lesions such as pulmonary cystic disease,

localized bronchiectasis, or lung abscess.

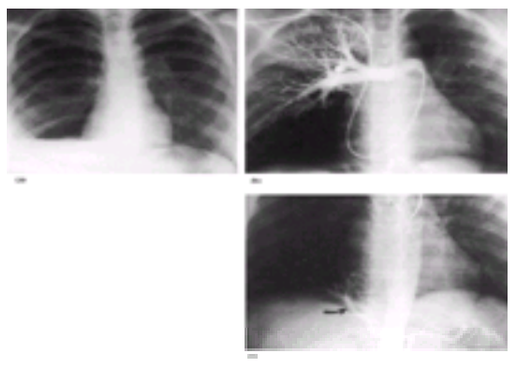


Fig. 18. (a) Chest radiograph of a 12-year-old girl referred for evaluation of the right lower lobe. She had an episode of pleuritic pain and right lower lobe pneumonia 5 months before, treated with penicillin. After recovery the chest radiograph showed a remaining pulmonary effusion, an area of ‘scarring', and hyperlucency of the right lower lobe, all identified on pulmonary tomograms as well. An underlying anatomic anomaly was suspected. (b) Pulmonary arteriogram demonstrates perfusion of the right upper and middle lobe but no perfusion of the lower lobe by the pulmonary artery. (c) Aortogram demonstrates perfusion of the right lower lobe by arteries arising from the aorta (arrow). At surgery this was confirmed and a right lower lobectomy was performed.

Extralobar sequestration is treated by resection. The systemic arterial supply of these lesions is easily defined. Lobectomy is generally required for an intralobar lesion: both the pulmonary and systemic arterial supply must be identified before this can be undertaken safely.

Lobar emphysema

Lobar emphysema (massive overdistension of one lobe) is an uncommon cause of respiratory distress in infancy. It produces compression and atelectasis of adjacent lobes, and can also shift the mediastinum, compressing the contralateral side. Respiratory symptoms appear early, and half of the patients present within the first 2 days of life. The left upper lobe is most frequently involved, followed by the right middle lobe; the lower lobes are rarely affected. Lobar emphysema is generally apparent on plain radiographs (Fig. 19), and can be distinguished from pneumothorax, in which a rim of collapsed lung is visible. Placement of a chest tube following an incorrect diagnosis of a pneumothorax can injure the distended lung. Bronchovascular markings within the emphysematous lobe, although they are abnormally far apart, allow this condition to be distinguished from a distended pulmonary cyst. Bronchography can confirm the diagnosis, but is generally not undertaken as it may impair pulmonary function in an already stressed infant (Fig. 20). An anatomic cause of lobar emphysema is found in less than 50 per cent of resected specimens. Pathologic findings include bronchial cartilaginous dysplasia and intrinsic obstruction of the bronchus. The polyalveolar lobe syndrome has recently been shown to be a cause of lobar emphysema. Patients with this condition have normal airways and arteries, but the resected lobe shows a three- to fivefold increase in the number of alveoli, representing a gigantic alveolar unit. If other congenital lesions are present, they usually involve the heart.

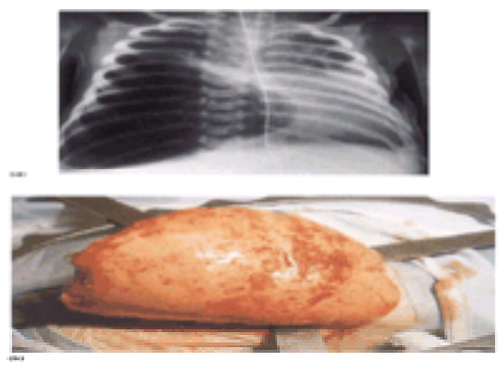


Fig. 19. (a) Chest radiograph of an infant with congenital lobar emphysema involving the right lower lobe. Distension of this lobe had progressed over several days with worsening pulmonary function because of compression of the remaining normal lobes. (b) At surgery the emphysematous lobe protruded through the thoracotomy incision immediately improving arterial gases.

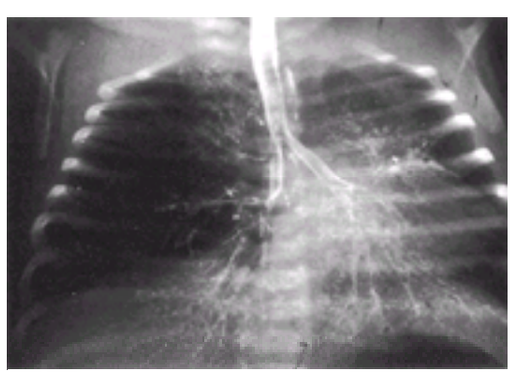


Fig. 20. Bronchogram in another infant with congenital lobar emphysema of the right middle lobe. It demonstrates passage of contrast into the middle lobe as well as compression and displacement of the upper and lower lobes.

Treatment of lobar emphysema by resection of the affected lobe was first performed successfully by Robert Gross in 1945. The severity of respiratory distress determines the urgency of resection: profound respiratory distress improves immediately when the chest is opened and emergence of the distended lobe through the incision allows expansion of the atelectatic normal lobes (Fig. 19).

Patients have no long-term symptomatic impairment following resection for lobar emphysema. One long-term follow-up study demonstrated normal total lung capacity in all 15 patients and normal vital capacity in 13 of 15 patients following lobectomy for congenital lobar emphysema. Radionuclide lung scans showed nearly equal lung volumes on both sides, and symmetrical pulmonary perfusion.

Infectious lesions

Major pulmonary infections have become less common since the introduction of antibiotics. Staphylococcal pneumonia still occurs, however, and its pulmonary and pleural complications must be identified. Chronic infection in a patient with cystic fibrosis produces progressive pulmonary damage and bronchiectasis. Although surgery is generally avoided, it may be required to control local areas of severe destruction and chronic infection.

Empyema and pneumatoceles

Staphylococcus aureus is the most frequent organism associated with empyemas in childhood followed in frequency by *Haemophilus influenzae* and *Streptococcus pneumoniae*. Infants under 6 months of age are most commonly affected. A high fever is usually followed by respiratory symptoms. If pneumothorax occurs or if significant empyema is present, an intercostal tube should be inserted immediately. This allows drainage of a large proportion of the infected material, with expansion

of the lung. Drainage and appropriate antibiotics will control the infection in most cases and the mortality rate is low.

Pneumatoceles often form after staphylococcal infections, paradoxically at a time when the infant is improving clinically (Fig. 21). Rupture of a pneumatocele into the pleural space should be treated by chest tube drainage. Resection is not necessary: pneumatoceles invariably resolve. Pulmonary decortication is rarely needed when empyema is drained, but if required may be achieved by a limited thoracotomy or thoracoscopic approach.

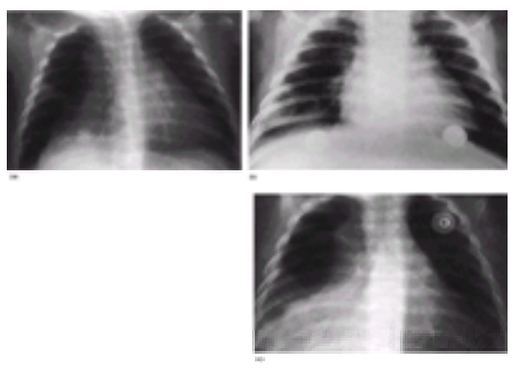


Fig. 21. (a) Chest radiograph of a 2-month-old infant with tachypnea and fever of 12 h duration. Consolidation of the right lower lobe was present, obscuring the diaphragm. Broad-spectrum antibiotics were started immediately. (b) Chest radiograph obtained 24 h later shows the development of pneumatoceles in the right lung. (c) At 72 h they had increased in size, although the infant had become afebrile. No significant pleural effusion developed to require placement of a chest tube and the child was managed with only parenteral antibiotics. The pneumatoceles resolved spontaneously over several months without development of a pneumothorax.

It can be difficult to determine whether a cystic lesion of the lung which is identified following a pulmonary infection is a congenital lung cyst or a pneumatocele. Differential diagnosis may be aided by the anatomic location of the lesion, a history of recurring localized infection, or a pre-existing radiographic abnormality. A known staphylococcal pneumonia is strongly suggestive of a pneumatocele. In most instances observation of the lesion is appropriate: pneumatocele will resolve progressively, albeit sometimes slowly, while a congenital cystic lesion will persist and should be resected.

Polymicrobial lung abscesses occur in children following aspiration. As in adults, the superior segment of the lower lobes and the posterior segment of the upper lobes are most commonly affected. An underlying immunodeficiency or neurologic impairment is generally present in these patients. Primary treatment is with broad-spectrum antibiotics and chest physiotherapy.

Pulmonary complications of cystic fibrosis

Pneumothorax is the most common surgical complication of cystic fibrosis. Chest tube drainage is usually required, since pulmonary reserve is limited in these patients. Most adjunctive treatments aim to decrease the 80 per cent recurrence rate. Sclerosis induced with tetracycline, silver nitrate, quinacrine, and talc poudrage obliterates the pleural space and prevents recurrence. Pleurectomy or surgical pleural abrasion have a limited role in children who tolerate thoracotomy poorly, although these methods have the lowest rates of recurrence. Thoracoscopic methods avoid a thoracotomy and have a similar efficacy.

Pulmonary resection has limited indications in patients with cystic fibrosis. Hemorrhage from dilated bronchial vessels can produce massive hemoptysis. This can usually be controlled by angiographic embolization of the bleeding vessel, but if embolization fails resection of the affected segment may be required. Pulmonary resection is required more often for severe bronchiectasis and recurrent infection in a destroyed pulmonary lobe than for hemoptysis. Resection is reserved for lesions which fail to respond to medical treatment and which are a recurrent source of infection of the remainder of the lung. Radioisotope scans allow surgical resection to be planned such that loss of functional tissue is avoided. Aggressive postural drainage, chest physiotherapy, and treatment with broad-spectrum parenteral antibiotics should be undertaken before resection. Resection can then be employed safely in the patient with cystic fibrosis, but it should be limited to the severely destroyed pulmonary segment with little functional capacity (Fig. 22).

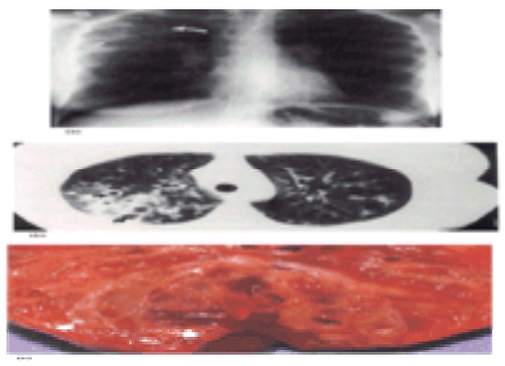


Fig. 22. (a) Chest radiograph of a 16-year-old girl with cystic fibrosis. Chronic hyperinflation is seen throughout her lung fields, but the most dramatic finding is the scarring and bronchiectasis seen in the right upper lobe (arrows). (b) The CT scan demonstrates severe bronchiectasis (arrows) which was limited to the upper lobe. Pulmonary radionuclide scans showed little function in this lobe. She developed hemoptysis, and it was elected to proceed with resection. Her recurrent pulmonary infections were attributed to this non-functional lobe. (c) The surgical specimen showed a severely scarred and consolidated lobe with little functional tissue. Severe bronchiectasis extended almost to the pleural surface (arrows).

Pulmonary neoplasms

Primary pulmonary neoplasms are rare in infancy and childhood, and the spectrum of tumors is different than that seen in adults. These lesions often present with cough, hemoptysis, wheezing, or localized areas of pulmonary collapse. Respiratory failure is rare. One-third of these lesions are benign consisting primarily of inflammatory pseudotumors or hamartomas. Malignant lesions comprise the misnamed bronchial ‘adenomas', which include carcinoids, mucoepidermoid carcinomas, adenoid cystic carcinomas, and mucous gland adenomas (Table 2). All are capable of dissemination, except for the extremely rare mucous gland adenomas. Adenomas generally occur in a primary or secondary bronchus. These lesions can usually only be partially resected by an endoscopic approach and attempts at such resection are often associated with heavy blood loss. Sleeve resection of the bronchus preserves pulmonary tissue and is currently used for the treatment of lesions that by their location lend themselves to this resection. Lobar resection should be performed if it is needed to remove the lesion entirely. Bronchiectasis may occur beyond an obstruction which has been present for many months: resection avoids recurrent pulmonary infections.

	Number of patients	*Percentage of patients*
Benign lesions (79)		
Inflammatory pseudotumor	45	19.6
Hamartoma	15	6.5
Neurogenic tumor	9	3.9
Leiomyoma	6	2.6
Mucous gland adenoma	2	0.9
Myoblastoma	2	0.9
Malignant lesions (151)		
Bronchial adenoma	65	28.3
Bronchogenic carcinoma	47	20.4
Pulmonary blastoma	14	6.1
Leiomyosarcoma	9	3.9
Rhabdomyosarcoma	6	2.6
Hemangiopericytoma	3	1.3
Lymphoma	3	1.3
Teratoma	2	0.9
Plasmacytoma	1	0.4
Myxosarcoma	1	0.4

Adapted from Hartman GE, Shochat SJ. *Annals of Thoracic Surgery* 1988; 45: 108.

Table 2 Primary pulmonary tumors in 230 infants and children

Bronchogenic carcinoma is the second most common malignant lesion after adenomas. Undifferentiated carcinoma and adenocarcinoma account for 80 per cent of these lesions; squamous cell tumors are less common in children (12 per cent) than in adults. Affected patients often have disseminated disease at presentation and the average survival is only 7 months. Pulmonary blastoma, a tumor which resembles fetal lung histologically, and other types of sarcomas account for the remainder of malignant primary lesions of the lung in infants and children. The pulmonary blastoma has been described in multiple reports arising from congenital pulmonary cysts, providing an additional basis for their routine removal.

Metastatic tumors are more common than primary pulmonary tumors in children. Most are metastatic sarcomas: these are the most frequent solid malignancies in infants and children, and carcinomas are rare. CT scanning aids in the identification of lesions, which would not be detected on routine chest radiographs. Treatment of the pulmonary metastasis is tumor specific, and may involve chemotherapy, radiotherapy, or resection of the primary lesion. Resection of pulmonary metastases, especially for osteogenic sarcoma, with adjunctive treatments may provide cure or long-term survival in patients previously thought to be ‘hopeless'. Pulmonary lesions which appear during treatment of a solid tumor must not be assumed to be metastatic; in one series of 37 children with known primary tumors, 33 per cent of pulmonary nodules were found to be benign lesions.

Pleura

Chylothorax (see also Chapter 41.9)

Chylothorax is a rare condition caused by the leakage of chyle from the thoracic duct or its tributaries. It may cause respiratory distress due to accumulation of lymph in the pleural space, most frequently the right side. In neonates spontaneous chylothorax is frequently related to an acute rise in venous pressure during delivery, which causes rupture of the thoracic duct. The most common cause of chylothorax is surgical trauma to the thoracic duct during mediastinal dissection for aortic arch anomalies and congenital heart disease. Its occurrence after repair of congenital diaphragmatic hernia results from injury to the thoracic duct by one of the sutures placed in the posterior diaphragmatic leaf. Chylothorax also occurs in premature infants with superior vena caval thrombosis.

The chest radiograph demonstrates opacification of one hemithorax with a shift of the mediastinal structures to the contralateral side (Fig. 23). It must be distinguished from pulmonary agenesis, in which the mediastinum will be shifted toward the affected side (Fig. 14).

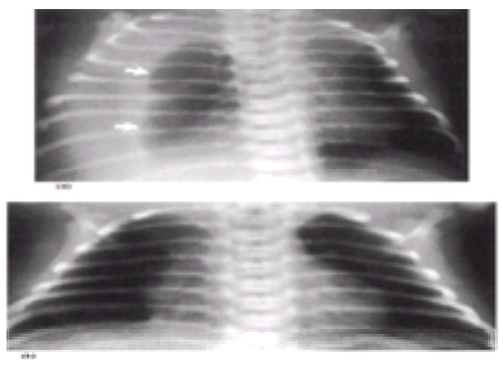


Fig. 23. (a) Chest radiograph of a newborn infant with dyspnea and decreased breath sounds on the right side. Note collapsed right lung (arrows) caused by a chylothorax. Pleural drainage was instituted by a chest tube and the baby improved rapidly. (b) The tube was removed after 5 days as the drainage resolved. No further collection developed.

Pleural drainage with a chest tube will allow many of these leaks to close. Limitation of fatty foods generally decreases lymphatic drainage. If high-volume drainage persists, parenteral feeding will decrease lymphatic fluid production by the intestines, often resulting in decreased drainage and closure. Parenteral nutrition also avoids the nutritional wasting that formerly resulted in an almost 50 per cent mortality rate in infants with chylothorax. When lymphatic drainage persists, surgical intervention is required. Identification of the site of the leak is facilitated by administration of a fat-laden meal prior to the surgery. Closure should be performed with non-absorbable sutures.

Malignant pleural effusions are rare in children. Obstruction of lymphatic drainage rarely occurs in association with Hodgkin's disease or lymphoma: chemotherapy directed at the primary tumor allows the effusion to resolve.

Pneumothorax (see also Chapter 41.8)

Air in the pleural space, especially when under tension, causes severe respiratory compromise. It is most common in very premature neonates receiving treatment with high ventilatory pressure for hyaline membrane disease. It is particularly dangerous in infants with congenital diaphragmatic hernia, in whom ventilation is already marginal.

Pneumothorax is rare after the neonatal period until the teenage years, when it may occur in an otherwise healthy youngster, generally due to rupture of a small apical bleb. Pneumothorax may also be the first sign of pulmonary metastasis in patients with osteogenic sarcomas.

A pneumothorax can be confirmed by chest radiograph (Fig. 24); differential diagnosis includes lobar emphysema and congenital cystic adenomatoid malformation. In a supine neonate air is often trapped anteriorly; this must not be overlooked. Treatment, which requires placement of a chest tube, is determined by the size of the pneumothorax and the degree of respiratory compromise. In the infant with a congenital diaphragmatic hernia and sudden respiratory decompensation, a chest tube should be placed on the contralateral side before chest radiography.

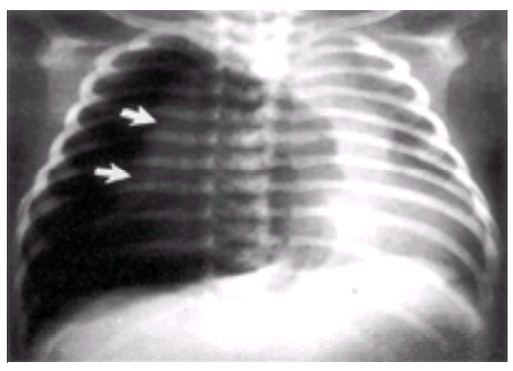


Fig. 24. Chest radiograph of an infant with a spontaneous pneumothorax. The collapsed lung can be readily seen (arrows) when a pneumothorax is present, in contrast with a pulmonary cyst or congenital lobar emphysema when the normal lung tissue is compressed by the air-filled cyst or lobe.

Spontaneous pneumothorax recurs in 50 per cent of teenage patients. If a pneumothorax recurs more than once, oversewing or resection of the blebs should be performed, in conjunction with abrasion of the parietal pleura with gauze sponges to promote adhesion of the lung to the chest wall. The intrathoracic exposure needed

is not great. Traditionally a vertical skin incision in the midaxillary line was used, since it is not very noticeable, being hidden by the arm. A fourth interspace incision is created to enter the chest. Currently most cases are managed by thoracoscopic exposure, resection, and abrasion.

Mediastinum

Mass lesions within the mediastinum can be divided into two broad categories: congenital cystic lesions and neoplastic tumors (Table 3). Symptoms are related more to their size and location than to their cystic or neoplastic nature.

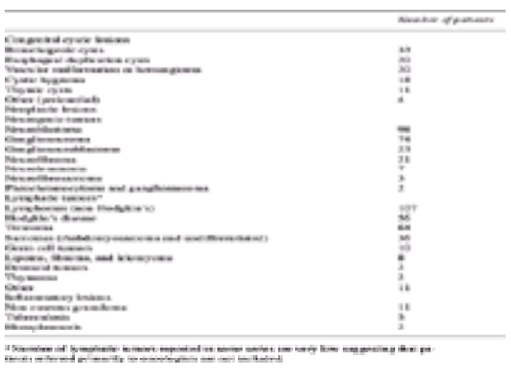


Table 3 Mediastinal masses in infants and children: summary of seven reported pediatric series

Congenital lesions

Cystic lesions occur throughout the mediastinum, and may present with respiratory symptoms including cough, wheezing, recurrent pneumonias, or stridor resulting from compression of the tracheobronchial tree. Infants may present in acute respiratory distress, and posterior mediastinal cysts occasionally cause esophageal compression, resulting in dysphagia. Approximately one-third of these cystic lesions are identified on routine chest radiographs. CT scan performed with intravenous contrast can differentiate solid neoplastic lesions from cysts (Fig. 25 and Fig. 26).

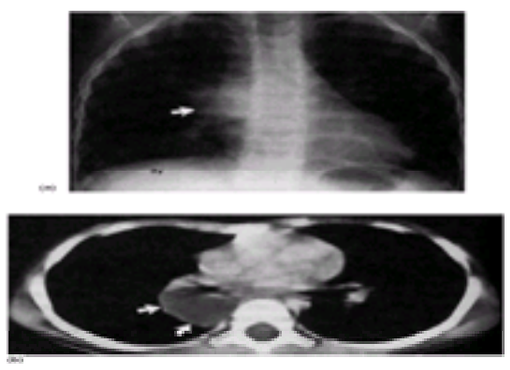


Fig. 25. (a) Chest radiograph of a 3-year-old boy obtained to evaluate the severity of a pectus excavatum deformity. A density was noted in the right retrohilar region (arrow). (b) A CT scan performed with intravenous contrast confirmed its presence directly posterior to the hilum and below the tracheal bifurcation (arrows). The darker lucency of the mass relative to muscle suggested that it was cystic, not a solid tumor. At surgery a bronchogenic cyst was found with muscle surrounding the cyst which was adherent to the back wall of the mainstem bronchus.

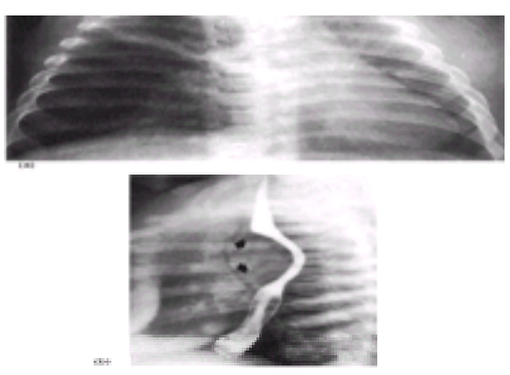


Fig. 26. (a) Chest radiograph of a 6-month-old boy admitted with stridor and respiratory distress. Tachypnea and substernal retractions had been noted at birth. This study revealed hyperinflation of the right lung and shift of the mediastinum to the left. (b) A lateral chest radiograph and barium swallow demonstrated a mass between the narrowed trachea (arrows) anteriorly and the esophagus posteriorly. At thoracotomy a bronchogenic cyst was found adherent to the posterior wall of the carina but without communication with the trachea.

Most cysts are of bronchopulmonary origin, and they are found in close proximity to the tracheobronchial tree, often in a perihilar or subcarinal location where radiographic findings may be limited. Their presence is suggested by emphysema or air trapping of one lung due to incomplete bronchial obstruction. Partial collapse of one lung results from high-grade bronchial obstruction.

Esophageal duplications are found in the posterior mediastinum. They are generally in contact with the esophagus but do not communicate with its lumen. A barium swallow may demonstrate ex-trinsic compression of the esophagus. The lining of these duplication cysts is usually squamous epithelium, but respiratory epithelium also may occur; rarely they are lined with gastric mucosa and present with perforation or hemorrhage. Esophageal duplications are generally isolated lesions, but they occasionally occur in patients with esophageal atresia, vertebral anomalies, or intraspinal extensions (neuroenteric cysts). Back pain, sensory or motor deficits, or gait disturbances may be the first symptoms of a neurenteric cyst. All patients with vertebral anomalies or neurologic symptoms and posterior cystic lesions should undergo preoperative myelography, or MRI scanning as an intraspinal component of the cyst is often present.

Surgical resection is the treatment of choice, even when the lesion is asymptomatic. An untreated cyst may enlarge from accumulation of secretions, causing compression of the trachea or bronchus, or it may become infected. Malignant lesions can arise in these cysts, and if the cyst is lined with gastric mucosa, ulceration or perforation may occur. Esophageal duplications share a common muscular wall with the esophagus. Resection of only the mucosal portion of the cyst in the area of contact with the esophagus avoids entry into the esophageal lumen.

Cystic hygromas and thymic cysts occasionally arise in the mediastinum. Cystic hygromas are multilocular, thin-walled cysts of lymphatic origin which should be considered in infants with cervical or upper thoracic lesions. Their extent can be determined from routine chest radiographs. Occasionally a cystic hygroma is confined entirely to the mediastinum and causes severe respiratory distress, resulting from tracheobronchial compression. Surgical treatment is required, although complete resection is often impossible.

Cystic lesions of the thymus appear as an anterior mediastinal mass on chest radiographs. Their nature and origin will be best defined by CT scan with intravenous

contrast. Major respiratory symptoms may be produced by large lesions: these are readily removed surgically.

Neoplastic lesions

Division of the mediastinum into anterior, middle, and posterior compartments will aid in the differential diagnosis of these lesions. Anterior mediastinal lesions consist of teratomas, thymomas, cystic hygromas, germ cell tumors, Hodgkin's disease, and non-Hodgkin's lymphoma. Middle mediastinal lesions are primarily lymphatic in origin and include lymphomas and Hodgkin's disease. Posterior mediastinal lesions located in the paravertebral sulcus are primarily neural in origin. They include neuroblastoma, ganglioneuroma, and neurofibroma. An extralobar sequestration can also present as a wedge-shaped, solid lesion in this area. Anterior mediastinal tumors may cause tracheal compression and respiratory symptoms. Posterior mediastinal lesions are more likely to present as asymptomatic masses, unless they are massive or produce neurologic symptoms due to intraspinal extensions.

Neurogenic tumors

Tumors of neural crest origin generally arise from the sympathetic chain and are the most common masses affecting the posterior mediastinum. The histologic spectrum of neurogenic tumors ranges from benign ganglioneuromas to malignant neuroblastomas: neuroblastoma is the most common and occurs primarily in infants or young children.

Ganglioneuromas occur in older children and teenagers, are frequently asymptomatic, and are often discovered when spindle-shaped lesions are seen along the spine on chest radiographs obtained following trauma or for respiratory symptoms unrelated to the mediastinal mass (Fig. 27). They can occur anywhere from the thoracic inlet to the diaphragm and may contain areas of calcification. Both ganglioneuromas and neuroblastomas may extend along the nerve roots into the spinal canal. Any child with neurologic symptoms should undergo a myelogram, CT scan, or MRI. Surgical resection is the treatment of choice and is curative for benign lesions. Ganglioneuromas are encapsulated, firm, and generally readily resected (Fig. 27(c)). Histologically they contain ganglion cells and fibrous tissue.

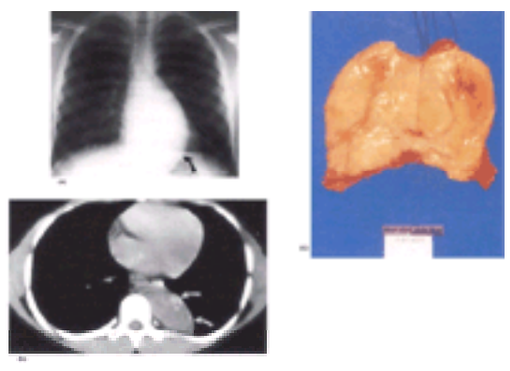


Fig. 27. (a) Chest radiograph obtained in a 14-year-old boy after a blow to his chest during a hockey game. It demonstrated a large spindle-shaped mass present behind the heart in the posterior vertebral sulcus (arrow). (b) CT scan revealed calcification within the mass (arrows) and no evidence of extension into the spinal cord. (c) At thoracotomy the mass was firm and could be dissected free from surrounding structures. The cut specimen demonstrates the fibrous and relatively avascular nature of mature ganglioneuroma.

Neuroblastoma in infants under 1 year of age is more benign than that affecting older children. In these infants, treatment comprises near total xcision of the lesion and observation. No further treatment is required unless osseous metastases are present. Adjunctive chemotherapy and radiotherapy may be required in children over 1 year of age who have distant metastasis or residual tumor. The prognosis of intrathoracic neuroblastomas is more favorable than that of abdominal lesions of comparable stage (Fig. 28).

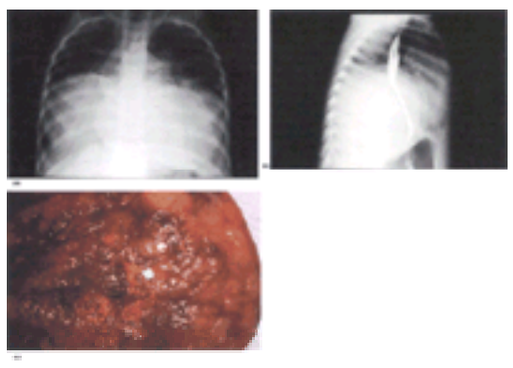


Fig. 28. (a) Chest radiograph of a 14-month-old infant who presented with failure to thrive, tachypnea, and tachycardia. Bilateral thoracic masses were noted posterior to the heart. (b) The lateral chest radiograph and barium swallow revealed anterior displacement of the esophagus and other mediastinal structures. CT scan showed areas of calcification and irregular perfusion suggesting necrosis. A neuroblastoma was suspected based on the age and location. The child underwent staged resections of these masses first on the left and then on the right side. (c) Cut section of the mass reveals its ‘fleshy' vascular nature as well as areas of calcification (arrow). Histologic examination confirmed the diagnosis of neuroblastoma. The child received chemotherapy after resection of the masses and is free of disease 4 years later.

Although variable, classic neuroblastoma consists of many small round cells (neuroblasts) which have a large nucleus and limited cytoplasm. Clustering of neuroblasts around a tangle of fine nerve fibers (rosette formation) is seen in tumors showing early differentiation. Ganglioneuroblastoma is a more differentiated tumor made up of mature ganglion cells, along with primitive neuroblasts in rosette patterns and other neural elements.

Plexiform neurofibroma in the mediastinum may be an isolated lesion or may be associated with neurofibromatosis (von Recklinghausen's disease). Resection of isolated lesions allows a histologic diagnosis to be obtained. Resection is also required in patients with neurofibromatosis if the lesions become symptomatic. Rapid growth of these lesions follows their malignant degeneration to neurofibrosarcoma. Pheochromocytoma can also occur along the sympathetic chain in the neck and mediastinum, producing symptoms of compression as well as flushing and hypertension.

Lymphatic tumors

Tumors in the anterior mediastinum may not produce symptoms until they become large. Cough, wheezing, dyspnea on exertion, and orthopnea are suggestive of important tracheal compression. CT is the best means of defining the extent of the lesion and the degree of tracheal compression (Fig. 29). Lymphoma and Hodgkin's disease do not usually occur until adolescence.

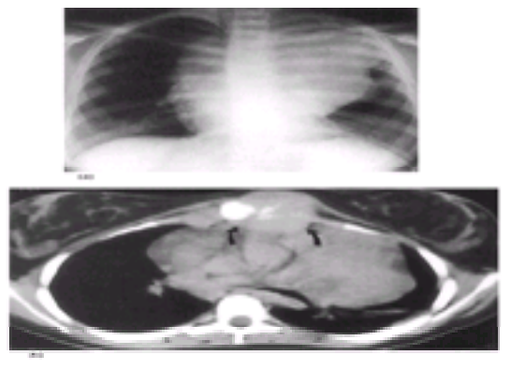


Fig. 29. (a) Chest radiograph obtained in a 12-year-old girl because of mild left pleuritic chest pain shows a large anterior mediastinal mass. (b) The CT scan revealed that the mass was eroding through the chest wall and sternum (arrows). Because of significant tracheal compression, a biopsy was obtained under local anesthesia, and Hodgkin's disease was diagnosed.

Patients with Hodgkin's disease or non-Hodgkin's lymphoma require biopsy specimens before initiation of treatment. Resection should not be attempted. Disseminated disease is often present, but the tumor shows an excellent response to radiotherapy and chemotherapy. Tracheal compression by a large mass can produce cardiorespiratory collapse upon induction of anesthesia. If the measured tracheal area on CT scan is less than 50 per cent of that expected for age, general anesthesia must be avoided. Recent studies have also shown that if the peak expiratory flow rate is greater than 50 per cent of predicted, general anesthesia will be safe. A cervical biopsy can often be performed under local anesthesia. Anterior thoracotomy also can be performed under local anesthesia, entering through the perichondrium of the resected second costal cartilage. Preliminary radiotherapy if used must spare an area of the tumor for biopsy confirmation of the diagnosis.

Teratomas

Mediastinal teratomas, the second most common tumor of the anterior mediastinum, are generally benign and may present at any age. They contain derivatives of all three germ cell layers and may be cystic. CT defines the extent of the lesion and any tracheal compression (Fig. 30). Calcification and varying densities of the tissues in the mass suggest the diagnosis. Surgical resection is usually curative, although malignant elements in the tumor will require further therapy. Elements of choriocarcinoma carry the worst prognosis.

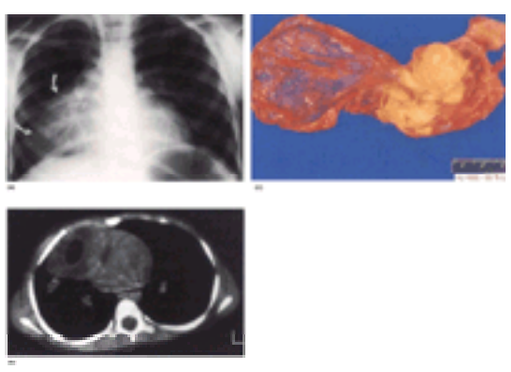


Fig. 30. (a) Chest radiograph obtained in a 7-year-old girl followed for an asymptomatic mediastinal mass since she was 2 years old. It shows a large mass extending into the right pleural cavity (arrows). (b) CT scan showed that the mass seen anterior and to the right of the heart was inhomogeneous with several densities present, including that of fat and water. (c) The surgical specimen resected through a right thoracotomy revealed a benign teratoma with no immature elements. It had a large cystic area, shown opened to the left, and a solid area containing fat, skin and skin appendages, and bronchial epithelium.

Thymomas

Thymomas are rare in infants and children. They are usually benign and arise in the upper anterior mediastinum or at the base of the neck. Although they may be massive, they generally compress adjacent structures rather than invade them. They are occasionally associated with myasthenia gravis. Surgical resection is required and recurrence is rare (2 per cent).

Malignant thymomas are epithelial in nature and invasive. Aggressive surgery is required: resection of lung, pleura, diaphragm, superior vena cava, and pericardium may be necessary to remove the lesion completely. Their response to chemotherapy and radiotherapy is limited.

The thymus may be the primary site of Hodgkin's disease. The diagnosis is established after resection of the mass, and additional chemotherapy or radiotherapy is required (Fig. 31).

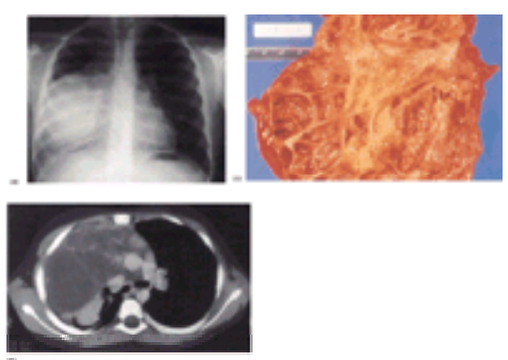


Fig. 31. (a) Chest radiograph obtained in a 10-year-old boy who had a brief episode of right pleural chest pain. On evaluation he had absent breath sounds on the right leading to the study which demonstrated a large right thoracic mass. He had had an entirely normal chest radiograph 18 months before this study. (b) CT scan showed a cystic mass with multiple septas present. (c) Resection of the cystic mass was performed through a right thoracotomy. The mass was found to arise from the thymus and it extended into the right pleural space. The walls of the mass were thick and irregular, with areas of nodularity which raised concern about malignancy. Pathologic gross examination of the opened specimen shown here revealed a thymic cyst with extensive involvement by Hodgkin's disease.

Germ cell tumors

Seminomas, teratocarcinomas, embryonal carcinomas, choriocarcinomas, and endodermal sinus tumors can all arise within the anterior mediastinum. They present in the adolescent years or later with respiratory symptoms, and the correct diagnosis is rarely made before biopsy. These lesions respond well to chemotherapy.

Sarcomas

Undifferentiated sarcoma and rhabdomyosarcoma may occur throughout the mediastinum. Complete resection is difficult: they tend to be extensive and to invade vital

structures. Long-term survivors are rare, despite adjunctive therapy with radiation and chemotherapy.

Esophagus

Congenital lesions

Esophageal atresia and tracheo-esophageal fistula

The first successful primary repair of esophageal atresia was performed in 1941. Today the vast majority of infants with esophageal atresia survive and lead normal lives. In infants without a tracheo-esophageal fistula the gap between the proximal and distal ends of the esophagus is long, making primary repair difficult. Classification of the various forms of esophageal atresia and tracheo-esophageal fistula are shown in Fig. 32 and Table 4. In 1967 Waterston devised a prognostic classification for infants based on their birth weight and the presence of sepsis or associated congenital anomalies. With advances in neonatal care a new system was developed in Montreal by Poenaro and associates and reported in 1993. The only two factors found to be predictive of survival were preoperative ventilator dependence and associated major anomalies. In most current series over 90 per cent of infants survive.

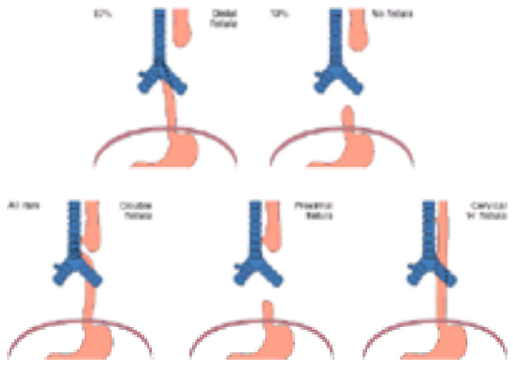


Fig. 32. Types of esophageal atresia and tracheo-esophageal fistula: a distal fistula is the most common (about 85 to 87 per cent of cases); the second most common (about 8 to 10 per cent) is pure atresia without a fistula. Tracheo-esophageal fistula without atresia occurs in around 4 per cent of cases. The other types are rare.

Esophageal atresia with distal fistula	85.5
Esophageal atresia without fistula	7.7
Tracheo-esophageal fistula without atresia	4.2
Esophageal atresia with proximal fistula	0.8
Esophageal atresia with proximal and distal fistula	7

Data from Holder TM *et al.* Esophageal atresia and tracheoesophageal fistula: a survey of its members by the Surgical Section of the American Academy of Pediatrics.

Table 4 Frequency of anatomic types of esophageal atresia

A fetus with esophageal atresia is unable to swallow amniotic fluid *in utero* and polyhydramnios results. Prenatal ultrasound examination allows early identification of these infants. Polyhydramnios causes premature labor in one-third of these pregnancies. The most common associated anomalies are imperforate anus and congenital heart disease.

The diagnosis of esophageal atresia should be considered in any infant producing excessive mucus or showing difficulty swallowing. It is confirmed by passing an 8 or 10 Fr catheter through the nose: the catheter will not reach the stomach, but meets a point of obstruction about 10 cm from the nostril. Air introduced through the catheter distends the pouch, and a chest radiograph demonstrates the level of obstruction without the use of contrast agents. If a radiocontrast agent is used, the study must be undertaken with fluoroscopic control, using a small volume of the agent and with the infant in the upright position to avoid aspiration. The contrast material should be removed immediately after the study, while the infant is held upright. Aspiration causes major pulmonary complications, and delays definitive surgery. Occasionally the contrast material enters the trachea even when the proximal pouch is not overfilled, suggesting a proximal fistula. A fistula between the trachea and the distal esophageal segment allows air to fill the stomach and intestine. If no air is seen in the stomach on an abdominal radiograph, a distal tracheo-esophageal fistula is absent. In these infants the distal esophageal segment is usually rudimentary, making primary repair difficult.

Infants in whom the diagnosis of esophageal atresia is delayed develop severe pneumonitis due to aspiration of gastric secretions through a distal fistula. If pneumonitis occurs, primary repair should be delayed and a Stamm gastrostomy performed under local anesthesia. The gastrostomy will vent air blown through the fistula into the stomach and prevent further reflux of gastric secretions. A sump suction catheter should be placed in the proximal pouch once the diagnosis is suspected, to prevent aspiration of saliva.

Surgical repair

A right thoracotomy is performed through the fourth intercostal space. Although a transpleural approach can be used, a retropleural approach allows separation of the pleura from the chest wall and its reflection medially and anteriorly. This approach prevents complete collapse of the lung during operation and if an anastomotic leak occurs, drainage is confined to the limited extrapleural space. Empyema does not then develop in the pleural cavity.

The distance between the two ends of the esophagus is variable, and is greatest in isolated esophageal atresia. The esophagus can be lengthened by mobilization of the proximal esophageal pouch, which has an excellent blood supply in the submucosal layer. Application of pressure to the proximal pouch with either a rubber bougie or a Bâkes common-duct dilator facilitates the mobilization and avoids the need to grasp the pouch, which may produce trauma.

The distal esophageal pouch is ligated at its junction with the trachea and divided. The distal segment should not be mobilized: it has a tenuous segmental blood supply and ischemia is easily produced. The two ends are anastomosed in a single-layer fashion. The two-layer anastomosis of Haight produces more strictures and does not prevent leaks. A segment of the azygous vein or mediastinal adventitia is interposed between the esophageal anastomosis and the fistula closure to minimize recurrence. An extrapleural chest tube is left in close proximity to the anastomosis, sutured to the chest wall to prevent direct contact.

Undue tension often results in a leak or complete disruption of the anastomosis within 1 week of repair. If extensive mobilization of the proximal pouch does not allow anastomosis without undue tension, several techniques can be used to increase esophageal length further. A circular myotomy through the muscular portion of the proximal pouch lengthens the proximal pouch but maintains a viable blood supply in the submucosa. A spiral myotomy is an alternative technique to increase esophageal length.

If all these methods fail, the distal esophageal end should be closed and sewn to the chest wall to prevent its retraction. Following closure of the chest, a cervical esophagostomy is placed anterior to the sternocleidomastoid muscle on the left. Esophageal reconstruction is then performed at about 1 year of age, when the infant reaches 10 kg. Colonic replacement of the esophagus can be performed retrosternally, although we prefer the left thoracic approach. (Fig. 33). Others have used a tube created from the greater curvature of the stomach, rotated through the chest to bridge the gap; this carries a risk of late peptic esophagitis. Rarely is it suitable in infants to mobilize the entire stomach into the chest to bridge the esophageal gap.

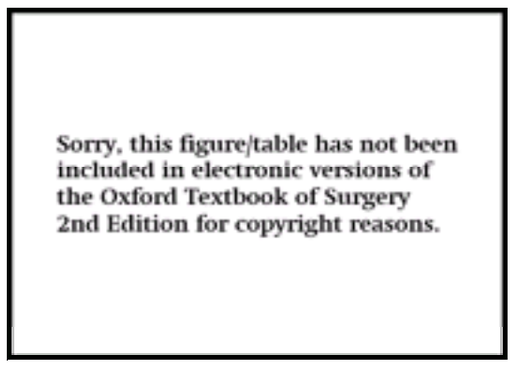


Fig. 33. (a) The patient is positioned for left ninth interspace thoracoabdominal incision, with arm wrapped so that it can be retracted downward to give access to the neck. The diaphragm is incised at its periphery, which avoids denervation. (b) The colon is mobilized, isolating transverse colon on pedicle of left colic artery. The point of transection should be carefully measured to be sure that it is long enough to reach the upper anastomotic site with absolutely no tension. In some this will be distal to the hepatic flexure, and in others it may extend well into the right colon. The bowel is cleansed thoroughly before surgery for elective cases. In two instances it was cleansed intraoperatively when the operation was done as an emergency procedure. (c) The colon is brought through the diaphragm. The location of the hiatus should be medial, to avoid angulation of the conduit as it goes from the thoracic gutter through the diaphragm. Note that the spleen and tail of pancreas have been reflected medially, so that the vascular pedicle runs straight upward anterior to the kidney and adrenal. Before the two lower anastomoses are performed, a short segment is sacrificed as shown. This facilitates doing the cologastric and colocolic anastomoses. The vascular pedicle to the conduit is preserved with great care. (d) The upper anastomosis is completed, using two layers of fine non-absorbable interrupted sutures. It is important to make the opening at the thoracic inlet of ample size to avoid extrinsic compression of the colon conduit. The dissection is carried immediately behind the clavicle to the apex of the thorax, anterior to the subclavian vessels. The sternocleidomastoid muscle is detached enough to make a generous opening. The bowel is tacked to this hiatus to prevent any tugging on the upper anastomosis. While the upper anastomosis is in progress, a sump suction tube is manually directed through the nose down into the conduit. This provides a means for deflating it postoperatively for several days. This tube should be anchored securely. Should it be pulled out inadvertently, it is not safe to reintroduce the tube blindly after surgery. The conduit is tacked also to the endothoracic fascia of the thoracic gutter, to guard against its twisting when the patient is moved postoperatively. Note anastomosis of colon to back wall of the stomach, with about 5 cm of intra-abdominal colon to protect against reflux. Note also the pyloroplasty and gastrostomy. Observation should be made that spleen and tail of pancreas do not compress pedicle of colon conduit. (e) Method of providing exposure for performing upper esophagocolic anastomosis in the chest. The chest muscles are reflected upward, exposing a higher interspace, and opening it to provide excellent exposure of the upper thorax. The esophagus is mobilized from beneath the aortic arch, and brought laterally as shown. This facilitates performing the upper anastomosis. (f) Options for lower anastomosis. (1) Anastomosis of colon to stomach, bringing colon conduit through new hiatus which should be close to the normally placed hiatus. Conduit can be put into either front or back wall of the stomach, according to how it lies in a particular case. (2) Anastomosis to the defunctioned distal esophagus in cases of esophageal atresia. Theoretically this should be better, but we have not noted any clinical distinction between that method and direct anastomosis to the stomach. (3) Use of original hiatus in patients where simultaneous esophagectomy is performed. In cases with extensive scarring of the esophagus, it may be better to leave esophagectomy until later, since that can add greatly to the length and difficulty of the procedure; for instance, combining esophagectomy with colon replacement to a high level.

Successful repair is achieved in more than 95 per cent of infants with uncomplicated esophageal atresia. Congenital heart disease or prematurity significantly worsen these results. In the past, very small infants were treated with placement of an initial gastrostomy tube under local anesthesia. If the infant's respiratory status was satisfactory, thoracotomy was performed several days later. The tracheo-esophageal fistula was divided, closing the distal segment and tacking it to the chest wall to prevent retraction. Gastrostomy feedings were then given without risk of reflux into the tracheobronchial tree. A sump suction catheter was maintained in the upper pouch to prevent aspiration. Major problems can be encountered, however, ventilating very premature infants with hyaline membrane disease after gastrostomy tube placement. The distal fistula, stomach, and gastrostomy tube provide a low-pressure ‘vent' that impairs adequate ventilation of the poorly compliant lungs. Some surgeons prefer initial ligation and division of the fistula, with delayed primary repair. Primary repair is now possible in many infants previously treated with the staged approach.

Special consideration must be given to infants with isolated esophageal atresia in whom the esophageal segments are too short to allow primary repair. Most of these infants have a rudimentary distal segment. The length of the segments can be defined radiographically, avoiding an unsuccessful thoracotomy. Aortic arch anomalies, including a left arch with an aberrant right subclavian artery originating from the descending aorta and a right-sided aortic arch with an aberrant left subclavian artery, have also been associated with wide gap atresias. Most surgeons place a gastrostomy tube initially and begin feedings with the hope that reflux into the distal pouch will dilate this segment. Others use intermittent dilatation of the proximal pouch with Maloney dilatators, with variable success. Simultaneous electromagnetic stretching of both esophageal ends elongates the esophageal segments and allows delayed primary anastomosis. Difficulties with nursing these infants led to abandonment of this technique.

Swallowing is abnormal in children after repair of esophageal atresia: fluoroscopic and manometric studies demonstrate a significant incidence of dyskinesia of the proximal and distal segments. Uncoordinated peristaltic activity is common, and gastroesophageal reflux is seen in about 20 per cent of patients. Food reaches the stomach mainly by the effect of gravity in colonic and gastric conduits.

Surgical complications

The anastomosis will leak if the repair is performed with excessive tension or if the distal esophageal segment is mobilized. Small leaks are well tolerated if adequately drained, and most will close without intervention. If complete disruption of the anastomosis occurs, the chest must be explored and a cervical esophagostomy (‘spit fistula') created, with ligation of the distal esophageal segment. Later esophageal substitution will be required.

Strictures can occur after repair, with or without a preceding esophageal leak. Most can be managed by dilatation, and injection of steroids into the scar prevents recurrence. The occurrence of a stricture, particularly if it recurs, must suggest the possibility of gastroesophageal reflux and lead to its evaluation. The use of filiform and followers pulled through the gastrostomy is the safest method for dilatation. A refractory stricture may require resection and reanastomosis.

Recurrent tracheo-esophageal fistula may be fatal: reoperation is always necessary. Closure of the fistula and reanastomosis of the esophagus may be possible. It is important to place well-vascularized tissue over the tracheal closure (pericardium, pleura, or intercostal muscle), to prevent recurrence. If the esophageal tissues are too inflamed, only tracheal closure is possible.

Gastro-esophageal reflux occurs in about 20 per cent of patients with repaired esophageal atresia and experimentally is related to the degree of tension on the anastomosis. The majority of symptomatic infants respond to non-surgical treatment. If reflux produces recurrent strictures, a fundoplication is often required.

Tracheomalacia is occasionally seen in association with esophageal atresia. In most instances the symptoms will improve with time, but in severe cases with life-threatening anoxic spells aortopexy is required as described by Filler and others.

Isolated tracheo-esophageal fistula (H-type)

Isolated tracheo-esophageal fistula is uncommon, and its diagnosis requires a high index of suspicion in an infant with recurrent pneumonitis. The fistula runs in an oblique downward direction from the trachea to the esophagus, and it can be missed if a barium swallow is not performed correctly. A small feeding tube should be placed in the cervical esophagus. With the infant in a lateral position a small amount of contrast medium is injected under fluoroscopic control (Fig. 34). If a suspected fistula is not seen on radiographic examination, endoscopy is performed under general anesthesia using a 30° lens. If a fistula is seen, a small Fogarty catheter can be passed through the endoscope and through the fistula into the esophagus to facilitate its identification at surgical repair.

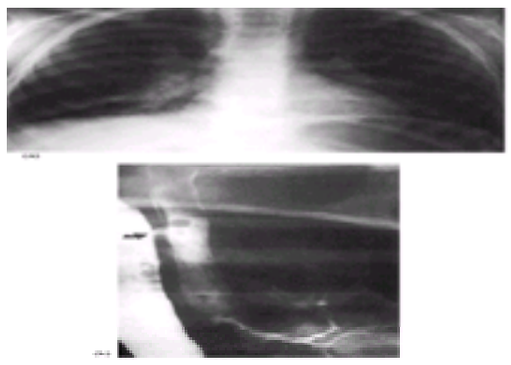


Fig. 34. (a) Chest radiograph obtained in a 30-month-old boy who presented with recurring pulmonary infections. Each responded to therapy but soon recurred. They involved primarily the right lower lobe. (b) Barium swallow demonstrated an H-type tracheo-esophageal fistula (arrow). Extreme caution must be taken in performing this study when looking for this type of fistula. The patient must be in the lateral position and barium slowly instilled into the esophagus under fluoroscopic control to see the fistula.

An H-fistula is approached through the right side of the neck taking special care to avoid injury to the recurrent laryngeal nerve. A strap muscle interposed between the trachea and esophagus after the fistula is divided and ligated protects against recurrence.

Esophageal tracheobronchial remnants

Tracheobronchial remnants in the esophageal wall produce esophageal stricture. They may contain abnormal epithelial lining or respiratory ducts within the wall. This form of congenital esophageal stricture does not respond well to dilatation: segmental resection with primary anastomosis or interposition of a short esophageal substitute is generally required.

Hiatal hernia and gastro-esophageal reflux

Gastro-esophageal reflux occurs in children with or without hiatal hernia. It spontaneously disappears in the majority of infants as esophageal sphincter tone, which is low at birth, increases with age. Infants with severe neurologic disease uniformly fail to show such improvement: spasticity, abnormal pharyngeal function and gag reflex, aerophagia, and chronic supine position all exacerbate the esophageal reflux.

Infants with gastro-esophageal reflux present with a variety of symptoms, including vomiting, failure to thrive, recurrent pneumonia, and anemia. Pneumonia is characterized by its migratory nature, lack of predominating organisms, and refractory nature. Apneic spells in infants may be dramatic and fatal. These spells typically occur when the infant is horizontal, usually after feeding and frequently during sleep. Reflux is often occult and unsuspected in the absence of vomiting. Older children and adolescents may develop symptomatic peptic esophagitis, with retrosternal burning as the primary symptom. Less frequently dysphagia results from esophagitis, spasm, or stricture.

Radiographic contrast studies often demonstrate reflux, but they do not provide an accurate assessment of its severity or significance. Radionuclide ‘milk scans' can provide information about the rate of gastric emptying, the frequency of reflux, and the rate of esophageal clearance of the isotope. Collection of isotope in the lungs documents aspiration. Radionuclide studies are also more sensitive than barium studies. Esophageal pH probes allow assessment of the frequency of episodes of reflux as well as the length of time acid is present in the esophagus.

Manometric studies of the esophagus exclude primary esophageal motility disorders as the cause of reflux. Surgical repair increases lower esophageal sphincter pressure and can produce major obstructive problems in patients with esophageal motility disorders. Esophagoscopy with biopsy will also provide an accurate assessment of the severity of the esophagitis, although the histology is more accurate than the gross findings at endoscopy unless ulceration or mucosal slough is present. Eosinophilic infiltration of the esophageal wall is often seen in these patients with reflux.

Initial treatment of gastro-esophageal reflux consists of administration of H_2-receptor antagonists, which decrease the acidity of the stomach. Metoclopramide can increase the rate of gastric emptying as well as increase the pressure at the gastro-esophageal junction. Thickening of foods with cereal and giving smaller and more frequent feedings can decrease the degree of reflux in infants. A sitting position has traditionally been employed to decrease reflux, although recent studies suggest that the prone position on a 30° incline is optimal.

Indications for surgical intervention include severe apneic episodes, failure to thrive, and recurrent pneumonia despite medical therapy. Emesis alone is rarely an indication. In older children the indications include persistent pain and active esophagitis despite medical therapy, Barrett's esophagus, and esophageal stricture. Boix-Ochoa reported that only 4.2 per cent of 1525 patients under 14 months of age required surgical treatment. Surgery was required in 20 per cent of 475 patients over 14 months of age.

All surgical methods for correction of gastro-esophageal reflux involve reducing the stomach and distal esophagus into the abdominal cavity and increasing the pressure at the gastro-esophageal junction (Fig. 35). Most repairs also restore the acute angle of entry of the esophagus into the stomach. Extensive esophageal strictures due to chronic reflux fail to respond to simple dilatation and require an antireflux procedure combined with intraoperative and postoperative dilations. Esophagectomy and substitution may be required if severe stricture is present because of neglected reflux.

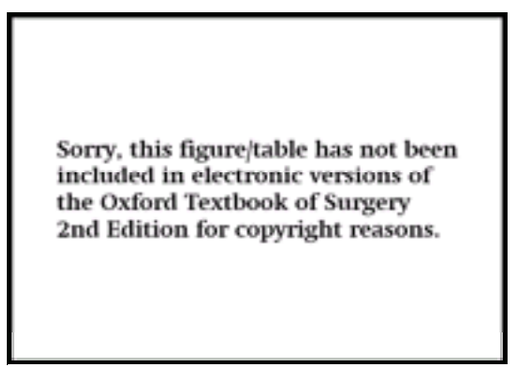


Fig. 35. Surgical technique for fundoplication in infants and children. (a) Abdominal incision and mobilization of the left lobe of the liver. (b) Incising the phreno-esophageal peritoneum to expose the esophagogastic junction. (c) Identifying the vagus nerves and mobilizing the lower esophagus. (d) Mobilizing a portion of the greater and lesser curvatures and identifying the margins of the esophageal hiatus and the median arcuate ligament. (e) Approximating the crura posteriorly. (f) Creating a partial fundoplication and tacking the esophagus to the hiatus. (g) Anchoring the lesser curvature of the stomach to the median arcuate ligament. (h) The completed repair.

Achalasia

Achalasia of the esophagus is seen occasionally in children. It is characterized by failure of normal relaxation of the lower esophageal sphincter, resulting in dysphagia, regurgitation of undigested food, retrosternal pain, weight loss, and pulmonary complications. Infants present with symptoms often attributed to gastro-esophageal reflux, including choking, apnea, pneumonia, and failure to thrive. The terminal esophagus appears narrow on radiographs, ending in a ‘bird's beak' with dilatation above (Fig. 36). Manometric studies of the lower esophageal sphincter demonstrate resting pressure two to three times normal, and absence of relaxation with swallowing. The pathology in the lower esophagus is similar to that of Hirschsprung's disease of the rectum, with absence of intramural ganglion cells.

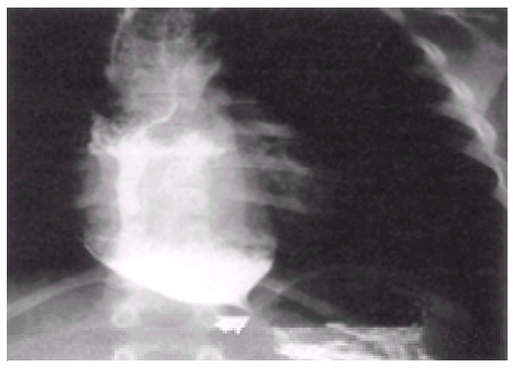


Fig. 36. Barium swallow of a child with dysphagia and retrosternal pain when swallowing. It reveals a persistent area of narrowing, (arrow) the ‘birds beak', of the distal esophagus which did not open with passage of the bolus of barium. This is consistent with achalasia.

Treatment may include dilatation, esophageal resection, and various plastic procedures. Dilatation generally provides only temporary symptomatic improvement, although it is often the initial treatment for adults. The modified Heller operation is satisfactory. Recent thoracoscopic myotomy has also been described.

Acquired lesions

Alkali ingestion

Although the incidence of ingestion of caustic agents has declined, it continues to occur. Alkalis, more frequently found in the household than strong acids, are the most commonly ingested. Most patients are about 2 years old, and generally under 5. Ingestions by adolescents are usually suicide attempts.

Strong acids primarily damage the stomach and duodenum and may produce perforation. They are less often the cause of serious esophageal injury since they produce coagulation necrosis that limits the penetration of acid into the deeper tissues of the esophagus. Strong alkalis, on the other hand, penetrate deeply into the muscle layers of the esophagus by a process of liquefaction necrosis which continues until the alkali is neutralized. Serious injury to the stomach is uncommon. Ingestion of ammonia and household bleach (sodium hypochlorite) seldom causes severe esophageal or gastric injury.

Clinical assessment of caustic damage to the esophagus cannot be based reliably on the presence of burns in the mouth. Some children with obvious burns to the lips and mouth spit out the substance, avoiding esophageal injury. Others have no evidence of oral burns, but have extensive esophageal burns that, if left untreated, can result in severe stricture. Solid alkaline material causes severe burns of the mouth and/or pharynx but is expelled without being swallowed. Liquid agents are more rapidly swallowed, causing less tissue injury within the mouth and more injury to the esophagus.

A barium swallow is not a sensitive means of identifying those children with early or limited esophageal injury. Demonstration of an atonic dilated esophagus has been related, however, to diffuse muscular necrosis and a high likelihood of a severe stricture. An atonic rigid esophagus appearing as a narrow rigid tube with no peristaltic activity is also likely to develop a stricture.

Esophagoscopy and laryngoscopy should be performed in all children with suspected caustic ingestion as signs and symptoms are poor predictors of injury. This procedure is usually performed within 48 h of admission, depending on the general condition of the child. A 24-h delay before endoscopy may allow a mild esophageal injury to become apparent. No attempt should be made to pass the esophagoscope beyond the most proximal site of injury because of the risk of perforation.

Various combinations of antibiotics, bougienage, esophageal stents, and steroids have been used to treat experimentally produced burns of the esophagus. A combination of these seems to be most effective, although a recent large prospective study failed to demonstrate benefit from the use of systemic steroids.

Esophageal injury can be classified as a first-degree burn (mucosal hyperemia, edema, and superficial sloughing), a second-degree burn, involving all layers of the esophagus (characterized by exudate, ulceration, and loss of mucosa), or third-degree burn, involving the entire esophageal wall with erosion into the pre-esophageal tissues. The last is rare, and requires an esophagectomy. Treatment with steroids and antibiotics is continued for at least 2 to 3 weeks. Patients may begin taking liquids when they are able to swallow their saliva. Children with severe pharyngeal and/or laryngeal burns may require intubation or an emergency tracheostomy several hours after injury as airway edema progresses. If vomiting and aspiration occur at the time of the initial injury, severe respiratory distress and pneumonitis can occur.

Esophageal dilatation is often instituted once the acute injury has resolved and mucosal edema has subsided (2 to 3 weeks after injury). Others await radiographic evidence of esophageal narrowing before attempting dilatation. Extreme caution must be taken with the first dilatations performed to avoid the risk of perforating the friable esophagus. Dilations performed daily as the mucosa regenerates prevents some strictures. Direct injection of steroids (triamcinolone diacetate) in conjunction with dilatation is helpful in preventing recurrence of some isolated strictures.

Esophagectomy and replacement with colon is required in some patients with severe injuries or delayed treatment (Fig. 37). If an esophageal replacement is required, it is best to remove the injured esophagus which carries a long-term risk of carcinoma, estimated to be 1000-fold that in the normal population with a latency period of 20 to 40 years.

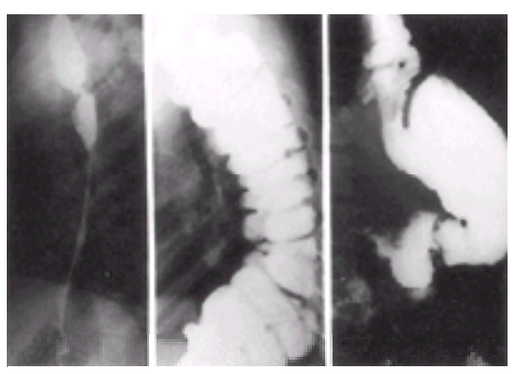


Fig. 37. Barium swallow with stricture in a 5-year-old child 18 months following ingestion of sodium hydroxide solution (left). Multiple esophageal dilatations had been performed but without lasting benefit. Barium swallow after replacement of the scarred oesophagus with a colon conduit (center), showing flow of barium through the proximal anastomosis and down the colon into the stomach. Cologastric anastomosis (right) is widely patent as shown in this spot view from the barium swallow.

Aortic arch anomalies

Vascular anomalies of the aorta that compress the trachea are among the surgically correctable causes of respiratory distress in infants and children. The aorta and its major branches develop from the six embryonic pairs of aortic arches corresponding to the branchial clefts. The third pair forms the carotid arteries, the fourth the aortic arch, the sixth the pulmonary artery and ductus arteriosus. The right side of the fourth arch generally disappears, producing the normal course of the aorta arching to the left and descending to the left of the spine. If the left fourth arch disappears and the right persists, the individual has a right aortic arch, a common anomaly. If both arches persist, they form a double arch or vascular ring encircling the trachea and esophagus. Gross first described the operative correction of this problem in an infant.

Many anatomic variations exist (Table 5): the most important and common are shown in Fig. 38. Some of the anomalies are asymptomatic; others may produce

compression of the airway with severe respiratory symptoms from birth. Infants with a true vascular ring encircling the trachea and esophagus have the most severe symptoms which are often aggravated by feedings. Other infants have a history of ‘noisy breathing' but no actual distress until they develop an upper respiratory infection. These infants commonly lie in a hyperextended position which stretches the trachea taut, reducing the constricting effect of a compressing vessel. Flexing the head forward can totally occlude the airway. This contrasts with infants with laryngomalacia or other causes of upper airway stridor, in whom position of the head and neck does not affect the severity of the symptoms.

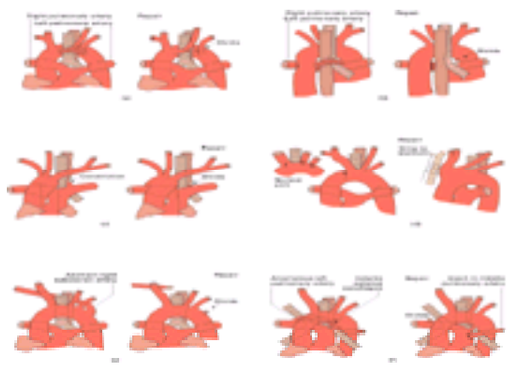


Fig. 38. (a) Double aortic arch forming a vascular ring, which compresses trachea and esophagus. This ring is the most common variety of vascular ring. Its correction requires division of the smaller, anterior arch at an appropriate point after a thorough dissection and identification of all of the vascular anatomy. (b) Vascular ring formed by double aortic arch, with the smaller component lying behind the esophagus. The smaller arch is usually short; accurate ligation and positive control of the proximal end are imperative before dividing the short arch, lest it slip into the depths of the wound when it is cut. (c) Right aortic arch with left ligamentum arteriosum, or patent ductus arteriosus, compressing trachea and esophagus. (d) Anomalous innominate artery. The origin of the innominate artery is closer than usual to the left common carotid, so that the trachea is caught in the crotch between these vessels. This problem can be accentuated by a large thymus gland in the neonate. This most common type of vascular compression of the trachea is also the most difficult to diagnose. (e) Aberrant right subclavian artery originating distal to the origin of the left subclavian, and crossing through the mediastinum posterior to the esophagus. (f) Anomalous left pulmonary artery, a rare lesion.

Anatomic lesions	Number of patients	Percentage of patients
Double aortic arch	136	59
Right aortic arch and left ligamentum arteriosum	52	23
Anomalous innominate artery	20	9
Anomalous subclavian artery	11	5
Pulmonary artery sling	10	4
Left aortic arch, right descending aorta, and right patent ductus arteriosus	1	1

Table 5 Aortic anomalies requiring surgical intervention in 230 infants and children

Diagnosis of a vascular anomaly is often established on plain chest radiographs, which demonstrate narrowing of the tracheal air column, and a barium swallow, showing esophageal indentation (Fig. 39). When the barium swallow is normal and the plain films are inconclusive, endoscopy should be performed to exclude tracheal and esophageal compression. Ten per cent of patients with aortic anomalies requiring surgical treatment have associated cardiac lesions.

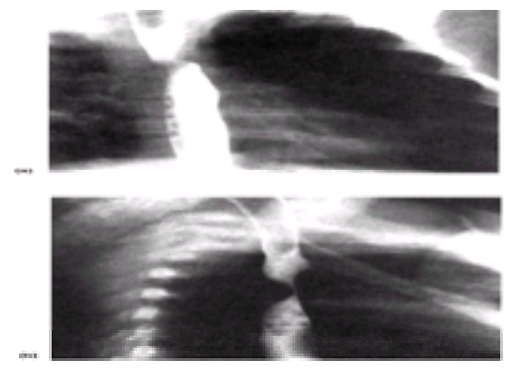


Fig. 39. (a) This 6-month-old infant presented with severe stridor. The barium swallow showed marked indentation of the esophagus from a posterior lesion. Near occlusion of the trachea was seen on the lateral radiograph as well. (b) A double aortic arch was suspected on the basis of this study and was confirmed by an echocardiogram. Bronchoscopy immediately prior to repair demonstrated near occlusion of the trachea.

The *sine qua non* of surgical correction in these anomalies is accurate and complete identification of the vascular anatomy in the mediastinum before division of any vessels. A left thoracotomy through the third or fourth intercostal space is performed. Most of the thymus is removed to expose the aortic arch and its branches. The most common of the vascular ring anomalies is shown in Fig. 38(a). The large arch lies posterior and the smaller anterior. The ring is divided at an appropriate point, usually between the left carotid and left subclavian artery. Division of the ligamentum arteriosum may further open the constricting ring. It is important to dissect the segment of trachea and esophagus that had been constricted to make certain that they are completely free from compression.

Figure 1(b) shows a double arch in which the smaller limb lies behind the trachea and esophagus. This limb is frequently very short and can slip away when divided, resulting in fatal hemorrhage. It is best to ligate the far end of this vessel securely, leaving an ample cuff beyond the ligature before it is divided. A clamp is used to control the end of the vessel nearer the operator, oversewing it flush with the descending aorta rather than relying on a ligature.

The right aortic arch (Fig. 38(c)) is a common variant. In association with a left ligamentum arteriosum or patent ductus arteriosus, it can entrap the trachea and esophagus, producing clinical symptoms identical to those of a double arch.

Anomalous innominate artery is difficult to diagnose with certainty. The trachea is compressed by impingement in a crotch formed by the innominate and left common carotid artery (Fig. 38(d)). Several variations are possible: the innominate can be displaced far to the left and actually cross the tracheal wall, or the left carotid can originate more proximally on the arch than normal and can be the principal cause for compression.

The barium swallow is completely normal in infants with an anomalous innominate artery, in contrast to the true vascular ring anomalies. The plain chest radiograph may also be inconclusive. Bronchoscopy shows a pulsating vessel indenting the anterior wall of the trachea. One must be prepared to proceed to thoracotomy and repair following bronchoscopy, which may acutely aggravate the infant's respiratory distress. Repair is performed through a left thoracotomy. The thymus is subtotally removed and sutures are placed in the adventitia of the obstructing vessels to suspend them forward to the sternum (Fig. 38(d)). This procedure can produce dramatic results in some infants. Some infants with mild respiratory symptoms do not require surgery.

The aberrant subclavian artery (Fig. 38(e)) is common, being seen in about 1 per cent of subjects at postmortem examination. The right subclavian originates in an anomalous position from the aortic arch beyond the origin of the left subclavian artery. A pseudodiverticulum of the aorta is often present at its origin. The artery passes

posteriorly, traversing the mediastinum posterior to the esophagus in most instances. It may occasionally produce enough esophageal compression to produce dysphagia. The symptoms are similar to those caused by an esophageal stricture. Barium swallow studies show a characteristic oblique indentation of the barium column in the anteroposterior projection. The clinical symptoms are difficulty in swallowing, often with regurgitation and aspiration when the upper esophagus fills and spills over into the trachea. Esophagoscopy demonstrates narrowing of the lumen by a pulsating vessel compressing the posterior wall of the esophagus.

Surgical correction, when required for symptoms, is performed through a left thoracotomy. The artery is divided at its origin from the aorta and the distal end is allowed to retract into the mediastinum releasing the compression of the esophagus. The blood supply to the right arm is adequate with proximal division through collateral circulation at the shoulder.

Anomalous left pulmonary artery (vascular ‘sling' Fig. 38(f)) is a rare but interesting cause of vascular compression of the trachea. The left pulmonary artery may cross anterior to the trachea, but can also be located between the trachea and the esophagus. Angiography may be necessary to establish the diagnosis, although anterior indentation at the level of the pulmonary hilum is diagnostic if present. Surgical correction is accomplished by moving the origin of the left pulmonary artery to an appropriate location anterior to the trachea and dividing the ligamentum arteriosum. It may be associated with severe intrinsic stenosis of the distal trachea or right mainstem bronchus, which must also be treated. Cardiopulmonary bypass offers the safest approach to the repair of this anomaly.

Anesthetic management of patients with these anomalies must allow for the constriction in the lower trachea. The endotracheal tube must pass beyond this point to maintain effective ventilation.

Residual symptoms may be present for a time after relief of obstruction due to persistent deformity of the softened cartilages of the previously compressed trachea. Ventilation may be required for several days after surgery. Oral feedings should not be attempted until the children are free from respiratory distress.

Further reading

Adkins PC, Blades B. A stainless steel strut for correction of pectus excavatum. *Surgery, Gynecology, and Obstetrics* 1961; **113**: 111–13.

Adzick NS *et al.* Diaphragmatic hernia in the fetus: prenatal diagnosis and outcome in 94 cases. *Journal of Pediatric Surgery* 1985; **20**: 357–61.

Adzick NS *et al.* Correction of congenital diaphragmatic hernia *in utero*. IV. An early gestational fetal lamb model for pulmonary vascular morphometric analysis. *Journal of Pediatric Surgery* 1985; **2**: 673–80.

Adzick NS *et al.* Fetal surgery for cystic adenomatoid malformation of the lung. *Journal of Pediatric Surgery* 1993; **28**: 806–12.

Altman RP, Randolph JG, Shearin RB. Tracheal agenesis: recognition and management. *Journal of Pediatric Surgery* 1972; **7**: 112–18.

Anderson, KD, Randolph JG. Gastric tube interposition: a satisfactory alternative to the colon for esophageal replacement in children. *Annals of Thoracic Surgery* 1978; **25**: 521–5.

Anderson KD, Rouse TM, Randolph JG. A controlled trial of corticosteroids in children with corrosive injury of the esophagus. *New England Journal of Medicine* 1990; **323**: 637–9.

Andrassy RJ *et al.* Thoracic sarcomas in children. *Annals of Thoracic Surgery* 1998; **227**: 170–3.

Asher MI *et al.* Primary lung abscess in childhood. *American Journal of Diseases of Children* 1982; **136**: 491–4.

Askin FB *et al.* Malignant small cell tumor of the thoracopulmonary region in childhood: a distinctive clinicopathologic entity of uncertain histogenesis. *Cancer* 1979; **43**: 2438–51.

Atkinson JB *et al.* The impact of extracorporeal membrane support in the treatment of congenital diaphragmatic hernia. *Journal of Pediatric Surgery* 1991; **26**: 791–3.

Beiser GD, Epstein SE, Stampfer M, Goldstein RE, Noland SP, Levitsky S. Impairment of cardiac function in patients with pectus excavatum, with improvement after operative correction. *New England Journal of Medicine* 1972; **287**: 267–72.

Bishop PJ *et al.* Transpleural repair of esophageal atresia without a primary gastrostomy: 240 patients treated between 1951 and 1983. *Journal of Pediatric Surgery* 1985; **20**: 823–8.

Black PR, Welch KJ. Pulmonary agenesis (aplasia), esophageal atresia, and tracheoesophageal fistula: a different treatment strategy. *Journal of Pediatric Surgery* 1986; **21**: 936–8.

Brereton RJ, Rickwood AMK. Esophageal atresia with pulmonary agenesis. *Journal of Pediatric Surgery* 1983; **18**: 618–20.

Buchino JJ, Meagher DP Jr, Cox JA. Tracheal agenesis: a clinical approach. *Journal of Pediatric Surgery* 1982; **17**: 132–7.

Butler C *et al.* Morphologic aspects of experimental esophageal lye strictures. II. Effect of steroid hormones, bougienage, and induced lathyrism on acute lye burns. *Surgery* 1977; **81**: 431–5.

Butler N, Claireaux AE. Congenital diaphragmatic hernia as a cause of perinatal mortality. *Lancet* 1962; i: 659–63.

Cahill JL, Lees GM, Robertson HT. A summary of preoperative and postoperative cardiorespiratory performance in patients undergoing pectus excavatum and carinatum repair. *Journal of Pediatric Surgery* 1984; **19**: 430–3.

Castile RG, Staats BA, Westbrook PR. Symptomatic pectus deformities of the chest. *American Review of Respiratory Disease* 1982; **126**: 564–8.

Cohen M *et al.* Pulmonary pseudometastases in children with malignant tumors. *Radiology* 1981; **141**: 371–4.

Coln D, Chang JHT. Experience with esophageal stenting for caustic burns in children. *Journal of Pediatric Surgery* 1986; **21**: 588–91.

Crist WM *et al.* Intrathoracic soft tissue sarcomas in children. *Cancer* 1982; **50**: 598–604.

Csendes A *et al.* A prospective randomized study comparing forceful dilatation and esophagomyotomy in patients with achalasia of the esophagus. *Gastroenterology* 1981; **80**: 789–95.

D'Agostino JA *et al.* Outcome for infants with congenital diaphragmatic hernia requiring extracorporeal membrane oxygenation: the first year. *Journal of Pediatric Surgery* 1995; **30**: 10–15.

deLorimier AA, Harrison MR, Hardy K, Howell LJ, Adzick NS. Tracheobronchial obstructions in infants and children. *Annals of Surgery* 1990; **212**: 277–89.

DiFiore JW, Fauza DO, Slavin R, Peters CA, Fackler JC, Wilson JM. Experimental fetal tracheal ligation reverses the structural and physiological effects of pulmonary hypoplasia in congenital diaphragmatic hernia. *Journal of Pediatric Surgery* 1994; **29**: 248–57.

Eigen H, Lemen RJ, Waring WW. Congenital lobar emphysema: long-term evaluation of surgically and conservatively treated children. *American Review of Respiratory Disease* 1976; **113**: 823–31.

Eraklis AJ, Griscom NT, McGovern JB. Bronchogenic cysts of the mediastinum in infancy. *New England Journal of Medicine* 1969; **281**: 1150–5.

Filler RM, Rossello PJ, Lebowitz RL. Life-threatening anoxic spells caused by tracheal compression after repair of esophageal atresia: correction by surgery. *Journal of Pediatric Surgery* 1976; **11**: 739–48.

Filston HC, Ferguson TB, Jr, Oldham HN. Airway obstruction by vascular anomalies: importance of telescopic bronchoscopy. *Annals of Surgery* 1987; **205**: 541–9.

Freire-Maia N, Chautard EA, Opitz JM, Freire-Maia A, Quelce-Salgado A. The Poland syndrome—clinical and genealogical data, dermatoglyphic analysis, and incidence. *Human Heredity* 1973; **23**: 97–104.

Gaissert HA, Mathisen DJ, Grillo HC, Vacanti JP, Wain JC. Tracheobronchial sleeve resection in children and adolescents. *Journal of Pediatric Surgery* 1994; **29**: 192–8.

Gaudreault P *et al.* Predictability of esophageal injury from signs and symptoms: a study of caustic ingestion in 378 children. *Pediatrics* 1983; **71**: 767–70.

Geggel RL *et al.* Congenital diaphragmatic hernia: arterial structural changes and persistent pulmonary hypertension after surgical repair. *Journal of Pediatrics* 1985; **107**: 57–64.

Göbel V *et al.* Malignant peripheral neuroectodermal tumors (MPNT) of childhood and adolescence: retrospective analysis of treatment results in 30 patients (PTS). In: *Proceedings of the Twenty-Second Annual Meeting of the American Society of Clinical Oncology* 1986; **5**: 206.

Goldman LP, Weigert JM. Corrosive substance ingestion: a review. *American Journal of Gastroenterology* 1984; **79**: 85–90.

Grosfeld JL *et al.* Chest wall resection and reconstruction for malignant conditions in childhood. *Journal of Pediatric Surgery* 1988; **23**: 667–73.

Gross RE. Congenital hernia of the diaphragm. In: *Surgery of infancy and childhood*. WB Saunders, Philadelphia, 1953, pp. 428–44.

Gross RE. Arterial malformations which cause compression of the trachea or esophagus. *Circulation* 1955; **11**: 124–34.

Haller JA. Severe chest wall constriction from growth retardation after too extensive and too early (< 4 years) pectus excavatum repair. An alert. *Annals of Thoracic Surgery* 1995; **60**: 1857–64.

Haller JA Jr, Mazur DO, Morgan WW Jr. Diagnosis and management of mediastinal masses in children. *Journal of Thoracic and Cardiovascular Surgery* 1969; **58**: 385–93.

Haller JA, Peters GN, Mazur D, White JJ. Pectus excavatum: a 20 year experience. *Journal of Thoracic and Cardiovascular Surgery* 1970; **60**: 375–83.

Haller JA Jr *et al.* Pathophysiology and management of acute corrosive burns of the esophagus: results of treatment in 285 children. *Journal of Pediatric Surgery* 1971; **6**: 578–83.

Haller JA Jr, Colombani PM, Miller D, Manson P. Early reconstruction of Poland's syndrome using autologous rib grafts combined with a latissimus muscle flap. *Journal of Pediatric Surgery* 1984; **19**: 423–9.

Han M-T *et al.* Aggressive thoracotomy for pulmonary metastatic osteogenic sarcoma in children and young adolescents. *Journal of Pediatric Surgery* 1981; **16**: 928–33.

Hartman GE, Shochat SJ. Primary pulmonary neoplasms of childhood: a review. *Annals of Thoracic Surgery* 1983; **36**: 108–19.

Hecker WCh, Procher G, Dietz HG. Results of operative correction of pigeon and funnel chest following a modified procedure of Ravitch and Haller. *Zeitschrift für Kinderchirurgie* 1981; **34**: 220–7.

Hemalatha V *et al.* Intrathoracic foregut cyst (foregut duplication) associated with esophageal atresia. *Journal of Pediatric Surgery* 1980; **15**: 178–80.

Hendren WH III, Haggerty RJ. Staphylococcic pneumonia in infancy and childhood: analysis of seventy-five cases. *Journal of the American Medical Association* 1958; **168**: 6–16.

Hendren WH, Hendren WG. Colon interposition for esophagus in children. *Journal of Pediatric Surgery* 1985; **20**: 829–39.

Hendren WH, Kim SH. Pediatric thoracic surgery. In: Scarpelli EM, Auld PAM, Goldman HS, eds, *Pulmonary disease of the fetus, newborn and child*. Lea & 38; Febiger, Philadelphia, 1978, pp. 166–234.

Hendren WH, McKee DM. Lobar emphysema of infancy. *Journal of Pediatric Surgery* 1966; **1**: 24–39.

Hopkins RA, Postlethwait RW. Caustic burns and carcinoma of the esophagus. *Annals of Surgery* 1981; **194**: 146–8.

Hümmer HP, Willital GH. Morphologic findings of chest deformities in children corresponding to the Willital–Hümmer classification. *Journal of Pediatric Surgery* 1984; **19**: 562–6.

Kim SH, Hendren WH, Donahoe PK. Gastroesophageal reflux and hiatus hernia in children: experience with 70 cases. *Journal of Pediatric Surgery* 1980; **15**: 443–51.

Kimura K *et al.* A new approach for the salvage of unsuccessful esophageal atresia repair: a spiral myotomy and delayed definitive operation. *Journal of Pediatric Surgery* 1987; **22**: 981–3.

King RM *et al.* Primary mediastinal tumors in children. *Journal of Pediatric Surgery* 1982; **17**: 512–20.

Kirsh MM *et al.* Treatment of caustic injuries of the esophagus: a ten year experience. *Annals of Surgery* 1978; **188**: 675–8.

Kosloske AM *et al.* Drainage of pediatric lung abscess by cough, catheter, or complete resection. *Journal of Pediatric Surgery* 1986; **21**: 596–600.

Lam CR, Kabbani S, Arciniegas E. Symptomatic anomalies of the aortic arch. *Surgery, Gynecology and Obestetrics* 1978; **147**: 673–81.

Langer JC *et al.* Timing for congenital diaphragmatic hernia: Is emergency operation necessary? *Journal of Pediatric Surgery* 1988; **23**: 731–4.

Leape LL, Bhan I, Ramenofsky ML. Esophageal biopsy in the diagnosis of reflux esophagitis. *Journal of Pediatric Surgery* 1981; **16**: 379–84.

Lester CW. The etiology and pathogenesis of funnel chest, pigeon breast, and related deformities of the anterior chest wall. *Journal of Thoracic Surgery* 1957; **34**: 1–10.

Levy MS, Share JC, Fauza DO, Wilson JM. Fate of the reconstructed carotid artery after extracorporeal membrane oxygenation. *Journal of Pediatric Surgery* 1995; **30**: 1046–9.

Liechty E *et al.* Primary *Hemophilus influenzae* lung abscesses with bronchial obstruction. *Journal of Pediatric Surgery* 1982; **17**: 281–4.

Loeff DS *et al.* Congenital tracheal stenosis: a review of 22 patients from 1965 to 1987. *Journal of Pediatric Surgery* 1988; **23**: 744–8.

Louhimo I, Lindahl H. Esophageal atresia: primary results of 500 consecutively treated patients. *Journal of Pediatric Surgery* 1983; **18**: 217–29.

McBride JT *et al.* Lung growth and airway function after lobectomy in infancy for congenital lobar emphysema. *Journal of Clinical Investigation* 1980; **66**: 962–70.

Manning PB *et al.* Fifty years' experience with esophageal atresia and tracheoesophageael fistula: beginning with Cameron Haight's first operation in 1935. *Annals of Surgery* 1986; **204**: 446–53.

Marmon LM *et al.* Vascular rings and slings: long-term follow-up of pulmonary function. *Journal of Pediatric Surgery* 1984; **19**: 683–92.

Martinez D, Juame J, Stein T, Peña A. The effect of costal cartilage resection on chest wall development. *Pediatric Surgery International* 1990; **5**: 170–3.

Meyers WF *et al.* Value of tests for evaluation of gastroesophageal reflux in children. *Journal of Pediatric Surgery* 1985; **20**: 515–20.

Murphy JJ *et al.* Rhabdomyosarcoma arising within congenital pulmonary cysts: report of three cases. *Journal of Pediatric Surgery* 1992; **27**: 1364–67

Nakayama DK *et al.* Pneumatic dilatation and operative treatment of achalasia in children. *Journal of Pediatric Surgery* 1987; **22**: 619–22.

Neilson IR *et al.* Congenital adenomatoid malformation of the lung: current management and prognosis. *Journal of Pediatric Surgery* 1991; **26**: 975–81.

Newman B. Real-time ultrasound and color-Doppler imaging in pulmonary sequestration. *Pediatrics* 1990; **86**: 620–3.

Nguyen L *et al.* The mortality of congenital diaphragmatic hernia: is total pulmonary mass inadequate, no matter what? *Annals of Surgery* 1983; **198**: 766–70.

Niitu Y *et al.* Lung cancer (squamous cell carcinoma) in adolescence. *American Journal of Diseases of Children* 1974; **127**: 108–11.

Nishina T, Tsuchida Y, Saito S. Congenital esophageal stenosis due to tracheobronchial remnants and its associated anomalies. *Journal of Pediatric Surgery* 1981; **16**: 190–3.

Nobuhara KK, Gorski YC, LaQuaglia MP, Shamberger RC. Bronchogenic cysts and esophageal duplications: common origins and treatments. *Journal of Pediatric Surgery* 1997; **32**: 1408–13.

O'Neill JA Jr *et al.* Surgical management of reflux strictures of the esophagus in childhood. *Annals of Surgery* 1982; **196**: 453–60.

O'Neill JA Jr, Holcomb GW Jr, Neblett WW III. Recent experience with esophageal atresia. *Annals of Surgery* 1982; **195**: 739–45.

O'Neill JA Jr, Holcomb GW Jr, Neblett WW. Management of tracheobronchial and esophageal foreign bodies in childhood. *Journal of Pediatric Surgery* 1983; **18**: 475–9.

O'Rourke PP *et al.* The use of the postductal Paosub2; as a predictor of pulmonary vascular hypoplasia in infants with congenital diaphragmatic hernia. *Journal of Pediatric Surger,* 1988; **23**: 904–7.

O'Rourke PP, Lillehei CW, Crone RK, Vacanti JP. The effect of extracorporeal membrane oxygenation on the survival of neonates with high-risk congenital diaphragmatic hernia: 45 cases from a single institution. *Journal of Pediatric Surgery* 1991; **26**: 147–52.

Peterson RJ, Young WG Jr, Godwin JD, Sabiston DC Jr, Jones RH. Noninvasive assessment of exercise cardiac function before and after pectus excavatum repair. *Journal of Thoracic Cardiovascular Surgery* 1985; **90**: 251–60.

Poenaru D, Laberge JM, Neilson IR, Guttman FM. A new prognostic classification for esophageal atresia. *Surgery* 1993; **113**: 426–32.

Pohlson EC, Schaller RT, Tapper D. Improved survival with primary anastomosis in the low birth weight neonate with esophageal atresia and tracheosophageal fistula. *Journal of Pediatric Surgery* 1988; **23**: 418–21.

Poland A. Defiency of the pectoralis muscles. *Guy's Hospital Report* 1841; **6**: 191–3.

Puri P, O'Donnell B, Guiney EJ. Delayed primary anastomosis following spontaneous growth of esophageal segments in esophageal atresia. *Journal of Pediatric Surgery* 1981; **16**: 180–3.

Raney RB Jr *et al.* Soft-tissue sarcoma of the trunk in childhood: results of the Intergroup Rhabdomyosarcoma Study. *Cancer* 1982; **49**: 2612–16.

Ravitch MM. Atypical deformities of the chest wall—absence and deformities of the ribs and costal cartilages. *Surgery* 1966; **59**: 438–49.

Ravitch MM. *Congenital deformities of the chest wall and their operative correction*. WB Saunders, Philadelphia, 1977.

Reyes HM *et al.* Experimental treatment of corrosive esophageal burns. *Journal of Pediatric Surgery* 1974; **9**: 317–26.

Roesler M *et al.* Surgical management of vascular ring. *Annals of Surgery* 1983; **197**: 139–46.

Ryan DP, Muehrcke DD, Doody DP, Kim SH, Donahoe PK. Laryngotracheoesophageal cleft (type IV): management and repair of lesions beyond the carina. *Journal of Pediatric Surgery* 1991; **26**: 962–70.

Saenz NC *et al.* Posterior mediastinal masses. *Journal of Pediatric Surgery* 1993; **28**: 172–6.

Sakai H *et al.* Effect of surgical repair on respiratory mechanics in congenital diaphragmatic hernia. *Journal of Pediatrics* 1987; **111**: 432–8.

Schaller RT Jr *et al.* Improved survival in children with osteosarcoma following resection of pulmonary metastases. *Journal of Pediatric Sugery* 1982; **17**: 546–50.

Selle JG, Snyder WH III, Schreiber JT. Chylothorax: indications for surgery. *Annals of Surgery* 1973; **177**: 245–9.

Shamberger RC, Welch KJ. Surgical correction of pectus carinatum. *Journal of Pediatric Surgery* 1987; **22**: 48–53.

Shamberger RC, Welch KJ. Surgical correction of pectus excavatum. *Journal of Pediatric Surgery* 1988; **23**: 615–22.

Shamberger RC, Welch KJ. Cardiopulmonary function in pectus excavatum. *Surgery, Gynecology and Obstetrics* 1988; **166**: 383–91.

Shamberger RC, Welch KJ. Surgical correction of chondromanubrial deformity (Currarino Silverman syndrome). *Journal of Pediatric Surgery* 1988; **23**: 319–22.

Shamberger RC, Welch KJ. Sternal defects. *Pediatric Surgery International* 1990; **5**: 156–64.

Shamberger RC, Welch KJ, Upton J III. Surgical treatment of thoracic deformity in Poland's syndrome. *Journal of Pediatric Surgery* 1989; **24**: 760–6.

Shamberger RC *et al.* Chest wall tumors in infancy and childhood. *Cancer* 1989; **63**: 774–85.

Shamberger RC, Holzman RS, Griscom NT, Tarbell NJ, Weinstein HJ. CT quantitation of tracheal cross-sectional area as a guide to the surgical and anesthetic management of children with anterior mediastinal masses. *Journal of Pediatric Surgery* 1991; **26**: 138–42.

Shamberger RC, Holzman RS, Griscom NT, Tarbell NJ, Weinstein HJ, Wohl ME. Prospective evaluation by computed tomography and pulmonary function tests of children with mediastinal masses. *Surgery* 1995; **118**: 468–71.

Skarsgard ED, Meuli M, VanderWall KJ, Bealer JF, Adzick NS, Harrison MR. Fetal endoscopic tracheal occlusion ('Fetendo-PLUG') for congenital diaphragmatic hernia. *Journal of Pediatric Surgery* 1996; **10**: 1335–8.

Smith MB, Hardin WD Jr, Dressel DA, Beckerman RC, Moynihan PC. Predicting outcome following pulmonary resection in cystic fibrosis patients. *Journal of Pediatric Surger,* 1991; **26**: 655–9.

Snyder ME *et al.* Diagnostic dilemmas of mediastinal cyst. *Journal of Pediatric Surgery* 1985; **20**: 810–15.

Spitz L, Kiely E, Brereton RJ. Esophageal atresia: five year experience with 148 cases. *Journal of Pediatric Surgery* 1987; **22**: 103–8.

Stocker JT, Madewell JE, Drake RM. Congenital cystic adenomatoid malformation of the lung: classification and morphologic spectrum. *Human Pathology* 1977; **8**: 155–71.

Stolar CJ, Snedecor SM, Bartlett R. Extracorporeal membrane oxygenation and neonatal respiratory failure: experience from the extracorporeal life support organization. *Journal of Pediatric Surgery* 1991; **26**: 563–71.

Tagge EP, Mulvihill D, Chandler JC, Richardson M, Uflacker R, Othersen HB. Childhood pleuropulmonary blastoma: caution against nonoperative management of congenital lung cysts. *Journal of Pediatric Surgery* 1996; **31**: 187–90.

Tapper D *et al.* Polyalveolar lobe: anatomic and physiologic parameters and their relationship to congenital lobar emphysema. *Journal of Pediatric Surgery* 1980; **15**: 931–7.

Templeton JM Jr *et al.* Management of esophageal atresia and tracheoesophageal fistula in the neonate with severe respiratory distress syndrome. *Journal of Pediatric Surgery* 1985; **20**: 394–7.

The Neonatal Inhaled Nitric Oxide Study Group. Inhaled nitric oxide in full-term and nearly full-term infants with hypoxic respiratory failure. *New England Journal of Medicine* 1997; **336**: 597–604.

The Neonatal Inhaled Nitric Oxide Study Group (NINOS). Inhaled nitric oxide and hypoxic respiratory failure in infants with congenital diaphragmatic hernia. *Pediatrics* 1997; **99**: 838–45.

Tribble CG, Selden RF, Rodgers BM. Talc poudrage in the treatment of spontaneous pneumothoraces in patients with cystic fibrosis. *Annals of Surgery* 1986; **204**: 677–80.

Tsugawa C *et al.* Congenital stenosis involving a long segment of the trachea: further experience in reconstructive surgery. *Journal of Pediatric Surgery* 1988; **23**: 471–5.

VanderWall KJ, Skarsgard ED, Filly RA, Eckert J, Harrison MR. Fetendo-Clip: a fetal endoscopic tracheal clip procedure in a human fetus. *Journal of Pediatric Surgery* 1997; **32**: 970–2.

Vane DW *et al.* Late results following esophagomyotomy in children with achalasia. *Journal of Pediatric Surgery* 1988; **23**: 515–19.

Weber TR, Kurkchubasche AG. Operative management of asphyxiating thoracic dystrophy after pectus repair. *Journal of Pediatric Surgery* 1998; **33**: 262–5.

Whittaker LD Jr, Lynn HB. Mediastinal tumors and cysts in the pediatric patient. *Surgical Clinics of North America* 1973; **53**: 893–904.

Wilcox DT, Glick PL, Karamanoukian HL, Morin FC, Fuhrman BP, Leach C. Partial liquid ventilation and nitric oxide in congenital diaphragmatic hernia. *Journal of Pediatric Surgery* 1997; **32**: 1211–15.

Wilson JM, Boer LK, Lund DP. Evolution of the technique of congenital diaphragmatic hernia repair on ECMO. *Journal of Pediatric Surgery* 1994; **29**: 1109–12.

Wilson JM, Fauza DO, Lund DP, Benacerraf BR, Hendren WH. Antenatal diagnosis of isolated congenital diaphragmatic hernia is not an indicator of outcome. *Journal of Pediatric Surgery* 1994; **29**: 815–19.

Wilson JM, Bower LK, Fackler JC, Beals DA, Bergus BO, Kevy SV. Aminocaproic acid decreases the incidence of intracranial hemorrhage and other hemorrhagic complications of ECMO. *Journal of Pediatric Surgery* 1993; **29**: 536–41.

Wung JT, Sahni R, Moffitt ST, Lipsitz E, Stolar CJH. Congenital diaphragmatic hernia: survival treated with very delayed surgery, spontaneous respiration, and no chest tube. *Journal of Pediatric Surgery* 1995; **30**: 406–9.

42.3 Evaluation of acute abdominal pain in children

Richard R. Ricketts

History
Associated conditions
Physical examination
Laboratory evaluation
Etiology of acute abdominal pain in infants and children
Acute appendicitis in children
Summary
Further reading

Acute abdominal pain is one of the more common complaints in childhood. There are many causes of abdominal pain in children, some of which originate in the abdomen while others do not, some of which require surgical treatment while some can be treated medically, and some of which involve the gastrointestinal tract while others do not. Whereas many causes of abdominal pain in children mirror their counterparts in adults, there are some conditions that are rather unique to childhood which will be discussed here. Furthermore, the evaluation and treatment of acute abdominal pain in children presents certain unique challenges to the practicing physician. This chapter will provide a brief overview of the evaluation of abdominal pain in children and will focus on those causes requiring surgery, particularly appendicitis.

History

Pain is a subjective feeling of discomfort. Although an older child can indicate specific complaints related to abdominal pain, an infant or younger child cannot. These patients may indicate the presence of abdominal pain by crying, fussiness and irritability, refusal to eat, emesis, or drawing up the legs and clinching the fists. If a history can be elicited, it is important to determine the site where the pain was initially noted. The initial pain is generally visceral in nature and is caused by distention of a hollow viscus or by ischemia. The site of the initial abdominal pain is influenced by the embryologic origins of the viscera involved. Foregut-derived structures (stomach, liver, duodenum, gallbladder, and pancreas) generally present with epigastric pain, midgut-derived structures (small bowel and colon to the mid-transverse colon) present with periumbilical pain, and hindgut-derived structures present with hypogastric pain. Referred pain also relates to the embryologic origins of the structures involved. Flank pain caused by renal diseases may be associated with scrotal pain because of the embryologic origin of the testis from the mesonephros. Subdiaphragmatic processes can be referred to the supraclavicular regions since these are the sensory dermatomes corresponding with the third, fourth, and fifth cervical nerves. Pain becomes localized once the somatic nervous system in contradistinction to the autonomic nervous system becomes stimulated.

The characteristics of the pain are also important in determining the etiology. Pain of gradual onset usually indicates an inflammatory process whereas that of sudden onset generally indicates a perforation or a sudden vascular compromise, such as torsion of the testis. Cramping, intermittent pain is characteristic of pain originating from obstruction of a hollow viscus (such as an early acute appendicitis, biliary colic, or ureteral colic), whereas steady pain occurs once the inflammatory process has been established or once infarction has occurred. Steady pain is also present with generalized peritonitis or with blood in the abdominal cavity.

A thorough history for associated symptoms is important in evaluating abdominal pain in childhood. Abdominal pain from adenitis may accompany an upper respiratory tract infection. Similarly, intussusception frequently occurs during the peak seasons for upper respiratory tract infection (fall and spring). An acute hydrocele mimicking an incarcerated hernia can also be associated with a recent upper respiratory tract infection. Abdominal pain associated with respiratory symptoms is frequently indicative of lower lobe pneumonitis. Abdominal pain associated with gastrointestinal symptoms can frequently be localized to the gastointestinal tract although this is not always the case. Pyelonephritis, salpingitis, and pancreatitis can be accompanied by gastrointestinal tract symptoms such as nausea, vomiting, and abdominal pain. Hepatobiliary symptoms usually localize the source of the pain to the liver, gallbladder, common duct, or pancreas. Symptoms referable to the internal and external genitalia are important in evaluating abdominal pain in children, particularly in girls. Salpingitis, tubal pregnancies, and other ovarian problems can certainly occur in the adolescent female. Acute abdominal pain in a boy with an undescended testis may represent torsion of an intra-abdominal testis. Urinary tract symptoms such as dysuria, frequency, and hematuria help to localize the source of the abdominal pain to the genitourinary tract.

Associated conditions

Certain associated conditions in childhood may help determine the source of the abdominal pain. Certain systemic diseases are frequently associated with abdominal pain. In sickle cell disease, an abdominal crisis can mimic acute cholecystitis or acute appendicitis. The patient with an abdominal sickle cell crisis will generally respond to intravenous hydration, oxygen, and perhaps transfusion therapy whereas patients with acute inflammatory processes within the abdominal cavity will not. Patients with spherocytosis can have abdominal pain secondary to rapid splenic enlargement. Children with collagen vascular diseases can present with vasculitis involving the abdominal viscera, peptic ulcer disease, or complications from steroid use. Children with leukemia can have abdominal pain related to chemotherapy *per se* or could develop pain secondary to typhlitis (neutropenic colitis).

Children who have abdominal pain in association with an abdominal mass most likely have constipation. However, other causes could include intussusception, ovarian or testicular torsion, an incarcerated inguinal hernia, or an abdominal tumor such as Wilms' tumor or neuroblastoma.

Children with abdominal pain associated with diarrhea generally have gastroenteritis. Other potential causes include hemolytic uremic syndrome, inflammatory bowel disease, or the occasional patient with appendicitis who has diarrhea.

Abdominal pain in association with blood in the stools would most likely indicate an intussusception in a toddler. Other causes of abdominal pain associated with bloody stools include hemolytic uremic syndrome, malrotation with midgut volvulus, gastroenteritis, Meckel's diverticulitis, and Henoch–Schölein purpura.

Physical examination

The physical examination begins with careful observation of the child. This is often done while speaking with the child, getting to know something about the child, and while taking the medical history. Is the child agitated or lying quietly? Is the child lying supine on the examining bed or is he/she in a flexed position? Is the abdomen distended or not? Are you able to elicit a smile from the child? Children with peritonitis generally do not smile, they lie very quietly in a semiflexed position, and they generally have abdominal distention.

The elicitation of tenderness is the most important factor in determining whether or not a child has significant intra-abdominal pathology which will require surgical intervention. In examining children, one must have warm hands and a gentle touch. The child should be distracted during the physical examination by talking with the child about activities at school, girlfriends or boyfriends, their hobbies, their pets, and so on. When one is speaking with a child, the abdomen is gently palpated repeatedly, staying away from painful areas until the end of the examination. When observing and speaking with the child, one can detect grimaces or other involuntary indicators of true abdominal tenderness. Often in small children a non-narcotic sedative (such as chloral hydrate) can calm a child down so that an examination for tenderness can be accomplished. The non-narcotic sedative will not mask true abdominal tenderness. The elicitation of ‘rebound tenderness' is not performed in the same manner as it is in an adult. In children, percussion of the abdomen can elicit ‘rebound tenderness' thereby indicating peritoneal irritation. Additionally, elicitation of unilateral involuntary rigidity has significant diagnostic implications. A rectal examination can be traumatic for the child unless performed properly. During the examination, one should speak with the child and explain what is being performed. The lubricated gloved finger is then inserted so that the thin part of the finger is oriented in an anterioposterior direction so that it goes easily through the oval anal sphincter. One should then sweep the examining digit posteriorly on the sacrum because if this maneuver causes ‘pain', the remaining examination will not be useful. If palpating the sacrum does not elicit the complaint of pain, then the lateral aspects of the rectum and finally the anterior aspect of the rectum are evaluated for the elicitation of tenderness. Additionally, the rectal examination can be used to evaluate the internal genitalia in preadolescent females; a standard pelvic examination is not necessary in these young children.

Bowel sounds are not helpful in evaluating abdominal pain in children with the exception of a mechanical bowel obstruction. Children can have normal bowel sounds in the presence of acute gastrointestinal pathology or can have absent bowel sound in diseases having nothing to do with the gastrointestinal tract, such as pyelonephritis with ileus.

Other components of the abdominal examination include evaluation for the presence of masses since children with large intra-abdominal malignancies often present with abdominal pain. One should look for the presence of erythema of the abdominal wall indicating the likelihood of a perforation and for dullness indicating the presence of fluid.

A careful chest examination is necessary to evaluate for the possibility of lower lobe pneumonia, which frequently presents as abdominal pain. In the adolescent population, a pelvic examination evaluating for ovarian or tubal diseases must be performed. In boys, a careful examination of the groin and scrotum evaluating for incarcerated hernia, undescended testes, orchitis, testicular torsion, or epididymus must be performed.

Laboratory evaluation

Certain laboratory and radiologic studies are appropriate for most children with abdominal pain. These would include a complete blood count and differential, a urinalysis, an amylase level, and perhaps a chest radiograph and abdominal plain film. In children of African heritage, a sickle cell preparation would also be indicated. In selected patients, depending upon the history and physical findings, additional laboratory and radiographic tests would be appropriate. If the patient has a history of vomiting or diarrhea and appears dehydrated, serum electrolytes should be analyzed. Persons with vomiting and diarrhea should also have a fecal smear for organisms, white blood cells, and ova and parasites performed. If the patient has a history suggestive of hepatobiliary disease, tests of hepatic function and hepatic enzymes should be obtained. An abdominal ultrasound is appropriate for patients having suspected renal pathology, adnexal pathology, or in the evaluation of a patient with an abdominal mass. Ultrasound is also useful in patients with pyloric stenosis, appendicitis, pancreatitis, and hepatobiliary disease. If intussusception is suspected, an air or contrast enema would be appropriate. If malrotation is suspected, an upper gastrointestinal series should be performed unless the patient requires immediate surgery.

Etiology of acute abdominal pain in infants and children

The causes of acute abdominal pain in infants and children which require surgery can conveniently be divided by age groups. Certain diagnoses occur in some age groups but not in others. The life of a child can be conveniently divided into the following periods: newborn (0 to 2 weeks), infant (2 weeks to 6 months), toddler (6 months to 2 years), child (2 to 12 years), and adolescence (12 to 18 years). Newborns generally present with intestinal obstruction, although they may present with perforation (spontaneous gastric or ileal perforation, necrotizing enterocolitis) or bleeding. The common causes of neonatal intestinal obstruction and clues to making the diagnosis of these conditions are shown in Table 1. These entities have been discussed elsewhere and will not be reiterated here.



Table 1 Causes and evaluation of neonatal intestinal obstruction

The most common causes of acute abdominal pain requiring surgery in children in the other age groups is shown in Table 2. Pyloric stenosis, midgut volvulus, and incarcerated inguinal hernias most commonly occur in infancy. Intussusception most commonly occurs in the toddler age group. The most common cause of acute abdominal pain requiring surgery during the childhood years is appendicitis. In addition to appendicitis, adnexal problems become apparent during adolescence. These entities have been discussed in elsewhere and will not be reiterated here.

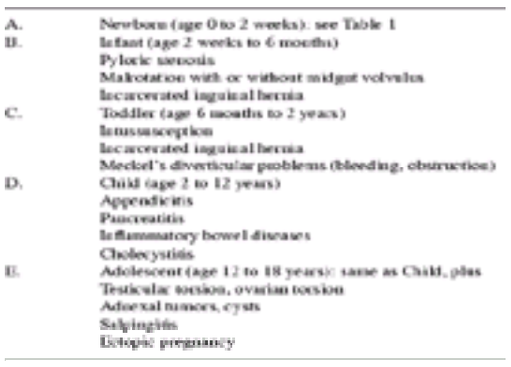


Table 2 Causes of abdominal pain requiring surgery in children

Acute appendicitis in children

Appendicitis is the most common acute surgical emergency in childhood. The typical symptoms include periumbilical cramping pain followed by vomiting, anorexia, a shift in the pain to the right lower quadrant, and a change in its character to a dull ache. In children, anorexia is not a frequent finding as it is in adults. The key to the diagnosis is the shift in the abdominal pain from the mid-abdominal region to the right lower quadrant and the change in its character from a cramp to a dull, steady ache. This pain shift occurs within about 4 to 6 h in children. As the inflammatory process progresses, perforation generally occurs in children within 36 to 48 h after the onset of symptoms. Diarrhea, vomiting, and signs of intestinal obstruction are more frequent in children, particularly young children, than in adults.

The typical signs of appendicitis include a low-grade fever, localized tenderness, and wincing or involuntary rigidity of the muscles of the right lateral abdominal wall. To elicit tenderness in children requires a gentle touch and patience as previously discussed. When the child is distracted, valid signs of peritoneal irritation can be elicited. Tenderness to light percussion (a ‘mini rebound') reliably predicts peritoneal irritation. A rectal examination, as previously described, is useful in differentiating appendicitis from gastroenteritis and is also useful to conduct the pelvic examination in preadolescent girls.

The laboratory findings of acute appendicitis are similar to those in adults. A useful adjunctive laboratory test is abdominal ultrasound, which can indicate the presence of appendicitis. The characteristic ultrasonographic features include a visible, non-compressible, tender tubular mass in the right lower quadrant with a total transverse diameter of 6 mm or greater. Ultrasound is also useful in demonstrating a periappendiceal abscess or pelvic pathology in girls.

The treatment of appendicitis in children is generally the same as in adults. However, the majority of children present with perforated appendicitis. With an acute perforation, the treatment is appendectomy, lavage of the right lower quadrant, closure of the peritoneum and fascia without closure of the skin, and avoidance of drains. An appendiceal abscess is generally treated non-operatively with a 2- to 3-week course of intravenous antibiotics, usually administered at home, followed by an elective ‘interval' appendectomy 6 or 8 weeks later. If the patient does not respond to non-operative management, operative or CT-directed drainage of the abscess is appropriate.

The differential diagnosis of appendicitis in children is extensive. The most common cause of abdominal pain in children that must be differentiated from acute appendicitis is acute gastroenteritis. Historical and physical findings which can aid in making the correct diagnosis are shown in Table 3. Some of the other entities which must be differentiated from appendicitis in children are shown in Table 4.

	Gastroenteritis	*Appendicitis*
Onset of pain	After vomiting	Before vomiting
Diarrhea	High volume; watery	Low volume; mucoid
Emesis	High volume; frequent	Low volume; infrequent
High fever	Likely	Unlikely
Peristalsis	High frequency	Low frequency
Rectal tenderness	Absent	Present
Peritoneal irritation	Absent	Present

Table 3 Acute gastroenteritis compared with acute appendicitis

Acute gastroenteritis	Ovarian cyst or tumor
Cholecystitis	Ovarian torsion
Constipation	Pancreatitis
Diabetic ketoacidosis	Pharyngitis, tonsillitis
Ectopic pregnancy	Primary peritonitis
Hemolytic-uremic syndrome	Renal calculi
Hemophilia	Rheumatic fever
Henoch–Schönlein purpura	Salpingitis
Incarcerated hernias	Sickle cell crisis
Inflammatory bowel diseases	Testicular torsion
Lead poisoning	Typhlitis
Lower lobe pneumonia	Typhoid fever
Meckel's diverticulitis	Urinary tract infection
Mittelschmerz	Yersinia enterocolitis

Table 4 Differential diagnosis of appendicitis in children

Summary

Acute abdominal pain (or manifestation of pain) is second only to fever as the most common complaint presenting to pediatric emergency services. By taking into account elements of the medical history—especially associated symptoms and conditions—and by considering the child's age, the appropriate diagnosis can usually be made after a careful physical examination. Laboratory and radiologic examinations serve to support the clinical diagnosis. With an accurate diagnosis in hand, appropriate therapy follows.

Further reading

Caty MG, Azizkan RG. Acute surgical conditions of the abdomen. *Pediatric Annals* 1994; **23**: 192–201. [This article reviews the common surgical illnesses of the abdomen in children arranged according to their initial presenting symptoms: pain, vomiting, and bleeding.]

Grosfeld JL, ed. *Common problems in pediatric surgery*, Part II: *Acute abdominal conditions*, pp. 69–138. Mosby-Year Book, Inc., St. Louis, 1991. [This section of the book reviews various causes of acute abdominal pain in children including appendicitis, intussusception, malrotation with midgut volvulus, pelvic inflammatory disease, ovarian torsion, inflammatory bowel disease, Hirschsprung's disease, Meckel's diverticulitis, and necrotizing enterocolitis.]

Hutson JM, Beasley SW. *The surgical examination of children: an illustrated guide*, pp. 16–34, 71–84. Heinemann Medical Books, Oxford, 1988. [This excellent book clearly describes and illustrates the common causes of abdominal pain in infants and children and follows each section with ‘Golden Rules' which should not be overlooked.]

Neilson IR *et al.* Appendicitis in children: current therapeutic recommendations. *Journal of Pediatric Surgery* 1990; **25**: 1113–16. [This article reports the results of a prospective protocol followed in 420 children with appendicitis. It looks at the results of their antibiotic protocol with respect to lowering the incidence of infectious complications in perforated and non-perforated appendicitis.]

Weiner DJ, Katz A, Hirschl RB, Drengowski R, Coran AG. Interval appendectomy in perforated appendicitis. *Pediatric Surgery International* 1995; **10**: 82–5. [This article compares the results of primary appendectomy with interval appendectomy for perforated appendicitis in childhood.]

42.4 Disorders of the gastrointestinal tract in children

D. Mervyn Griffiths

Neonatal problems
Neonatal intestinal obstruction
Necrotizing enterocolitis
Older children
Infantile hypertrophic pyloric stenosis
Intussusception
Constipation
Rectal bleeding
Inflammatory bowel disease
Further reading

Neonatal problems

Neonatal intestinal obstruction

Introduction

Intestinal obstruction in the neonate can be due to a wide variety of conditions with very varied pathology, associated conditions, treatment, and prognosis. However, the presenting features are very similar. Bile-stained green vomit is never normal in a neonate and implies the presence of an obstruction. The degree of abdominal distension varies with the level of obstruction. Ninety-five per cent of normal neonates pass meconium within 24 h of birth; failure to pass meconium at all or to stop after an initial defecation is abnormal. Pain is not usually a feature. Functional obstruction, as in Hirschsprung's disease, may also occur in septicaemia, premature babies with immature ganglion cells, in association with severe hypoglycaemia, or following hypotension and hypoxia. It is obviously important to avoid laparotomy in these babies.

Pylorus

Congenital aplasia, atresia, membranes, and webs of the pylorus are all very rare (0.5 per cent of all atresias). Embryologically they may reflect failure of recanalization, but this is uncertain. Polyhydramnios may be present. Complete obstruction presents with non-bile-stained vomiting and eventual metabolic alkalosis. Partial obstruction may not be detected until adult life.

Plain radiographs show a distended stomach, and a barium meal may demonstrate a web. Following nasogastric decompression, gastric lavage, and correction of any metabolic abnormalities a laparotomy is needed to demonstrate the exact pathology.

Gastroduodenostomy is required for the treatment of aplasia, while atresia, membranes, and webs are all treated by a Heineke–Mikulicz pyloroplasty. Prompt diagnosis and treatment should ensure a good long-term prognosis.

Duodenal atresia and stenosis

Intrinsic obstruction of the duodenum may be complete or incomplete (with a ratio of 2:1) and occurs in about 1 in 10 000 live births.

Anatomy

Duodenum The stomach and proximal duodenum are very hypertrophied and dilated, while the distal bowel is collapsed. The site of obstruction is most commonly in the second part of the duodenum, around the level of the ampulla of Vater. About 40 per cent of obstructions are preampullary, and 60 per cent postampullary. Less than 5 per cent are actually at the ampulla. The type of obstruction varies from a complete gap between the two ends, through connection with a fibrous cord, to apposition. A stenosis may be a short narrow segment or a perforate diaphragm, the small orifice of which may be central or eccentric. A thin diaphragm may stretch along the lumen like a windsock.

Bile duct The biliary tree is usually normal. The common bile duct can terminate on or very close to the attachment of a diaphragm. The terminal portion may be bifid, with orifices in both parts of the atresia. This may be impossible to detect clinically unless there is a complete atresia with air in the distal bowel.

Pancreas An annular pancreas is present in about 25 per cent of patients with duodenal atresia. It encircles the second part of the duodenum, which may be atretic or stenotic.

Embryology

Between 30 and 40 days of gestation, the duodenal mucosa proliferates, occluding the lumen. Duodenal atresia is due to failure of recanalization (see ileal atresia). The embryological cause of annular pancreas is unclear, but it is probably related to failure of rotation of the ventral anlage.

Clinical features

Antenatal At least 50 per cent of duodenal atresias and stenoses are associated with polyhydramnios, and at least 60 per cent of such pregnancies will be complicated or end prematurely. The hypertrophied, dilated stomach and duodenum are easy to see on antenatal ultrasound scans, but may be normal at an 18-week booking scan. Antenatal volvulus with infarction of the midgut, but survival of the fetus, will produce a jejunal atresia which mimics duodenal atresia on antenatal scans.

Associated malformations Down's syndrome is present in 30 per cent of babies with duodenal atresia, and antenatal karyotyping is justifiable. Sixty-five per cent of both Down's and normal babies with duodenal atresia have at least one other abnormality, ranging from major cardiac defects through malrotation (9 per cent) to biliary atresia and exomphalos. Oesophageal atresia is present in 10 per cent.

Symptoms These are those of a high gastrointestinal obstruction. Copious vomiting or regurgitation of feeds occurs early, but this may be delayed if there is only a stenosis. If the obstruction is preampullary, the vomit will not be bile stained and diagnosis tends to be delayed. Postampullary obstruction usually produces classic bile-stained vomit, but if not, then the diagnosis is delayed. Some blood may be present, due to oesophagitis, and severe dehydration may develop. Jaundice is usually not due to anatomical abnormality but to biliary atresia or surgical trauma. Colic, distension, and constipation are not common features.

Management

Preoperative A large nasogastric tube needs to be passed urgently to empty the stomach and duodenum and prevent aspiration of vomit. Fluid replacement may be needed, depending on the degree of dehydration. The presence of major cardiac and other defects needs to be excluded. Surgery should only be considered when the baby is fit for theatre.

If 40 to 50 ml of air are blown down the nasogastric tube, a subsequent plain abdominal film should be diagnostic. In duodenal atresia, a classic ‘double bubble' will be seen due to the massive dilatation of the first part of the duodenum, with air/fluid levels in the duodenum and stomach and no distal gas. Duodenal stenosis is characterized by the presence of air distally. Alternatively, ultrasound will confirm the diagnosis.

If the baby has Down's syndrome then the ethics of corrective surgery may need to be discussed with the parents.

Operation All babies with duodenal atresia should be treated by duodenoduodenostomy following Kocher's manoeuvre of the duodenum. One layer of absorbable sutures is sufficient. An annular pancreas is left alone and an anterior duodenoduodenostomy created.

Diaphragms can be incised or excised superolaterally to avoid damage to the ampulla. The presence of windsock diaphragms and additional diaphragms (4 per cent) must be excluded by passing a catheter proximally and distally through the duodenotomy. A grossly ectatic duodenum may function more quickly if it is tapered or plicated to a more normal size, but this is unproved.

Gastrostomies are not necessary if intravenous feeding is available. A soft silicone transanastomotic tube (a Silk tube) can be passed into the midjejeunum. Any other intra-abdominal pathology, such as malrotation, should be corrected.

Postoperatively The nasogastric tube must be aspirated regularly to avoid the possibility of aspiration pneumonia. After 24 h the transanastomotic tube can be used for replacement of nasogastric losses directly into the distal bowel and for feeding. Presence of the tube is unlikely to hasten feeding or bowel movements. The bowel may take 1 or 2 weeks to start to function normally, and if feeding via the transanastomotic tube is unsuccessful, intravenous feeding may be required.

Results

Less than 10 per cent of affected babies die due to their associated congenital abnormalities, including major cardiac problems, the reduced resistance of babies with Down's syndrome to infection, or the complications of prematurity. Most of the rest become asymptomatic, though bile reflux into the stomach has been demonstrated in adults.

Jejunal, ileal, and colonic atresia and stenosis

Atresia and stenosis of the bowel are less common distal to the duodenum than in the duodenum; estimates of the incidence vary widely from 1 in 1500 to 1 in 20 000 live births.

Embryology

Atresia distal to the duodenum is not an early event, as it occurs after the production of bile (11th week). No single cause explains all cases: some are due to intussusception with necrosis, while others are caused by a vascular accident which may or may not be identifiable (such as volvulus or cocaine abuse). Multiple atresia appears to be due to hereditary maldevelopment of the bowel.

Classification

There are four main types. Type I has a diaphragm with continuity of the bowel wall. Type II has a gap between the two ends (which may be connected by a fibrous cord), together with a mesenteric defect. Type III has multiple atresias. Type IV is ‘apple peel' atresia with loss of the dorsal mesentery and a blood supply to the ileum from the middle colic artery.

Clinical features

Antenatal ultrasound examination may show polyhydramnios or dilated bowel. Postnatally, the abdomen becomes very distended and bile-stained vomiting occurs early, especially in those with high obstruction. Meconium may be passed rectally. Stenoses commonly present after a delay of days to years, depending on the size and level of the stenosis.

Plain radiographs of the abdomen disclose hugely dilated bowel and a gas-free rectum (Fig. 1(a)). It is often possible to identify the level of obstruction preoperatively by the size, position, and number of bowel loops. Distal air is present in patients with stenosis and further contrast studies are required for accurate preoperative diagnosis.

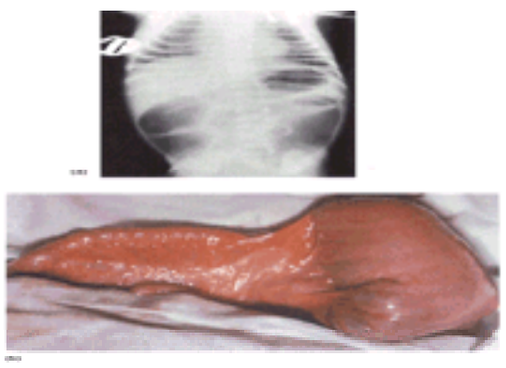


Fig. 1. (a) Plain abdominal radiograph in ileal atresia. (b) Ileal atresia type II.

The proximal bowel is always dilated and hypertrophied and the distal bowel collapsed but containing meconium.

Treatment

A nasogastric tube needs to be passed to decompress the stomach and avoid aspiration pneumonia. Dehydration must be corrected preoperatively.

At laparotomy, the hugely dilated proximal bowel should be resected if possible, or tapered, to reduce postoperative ileus and to make the anastomosis easier (Fig. 1(b)). The initial 5 to 10 cm of distal bowel should also be resected if there is any question of a poor vascular supply. After cut back of the distal bowel to ensure two lumens of approximately the same size, a single layer end-to-back anastomosis with interrupted absorbable sutures is performed. Multiple atresias can often be excised, leaving only one anastomosis. In apple peel atresia the blood supply of the ileum is precarious and great care is required to ensure viability of the distal bowel. Second-look laparotomy may be a wise precaution. Stenoses require excision and end-to-end anastomosis.

Postoperatively

Postoperative ileus is reduced by excising the dilated proximal bowel, but it may still occur. Intravenous feeding may rarely be necessary. Lactose intolerance is common and a lactose-free milk containing medium-chain triglycerides (such as Pregestemil) should be tried initially.

Results

The mortality rate of 100 per cent seen early this century has been replaced by a survival rate of over 85 per cent. Death is usually due to associated abnormalities.

Meconium ileus

Meconium ileus is the most common cause of neonatal intraluminal intestinal obstruction and is usually (75 to 80 per cent) due to cystic fibrosis. Cystic fibrosis occurs in about 1 in 2000 live births and is inherited as an autosomal recessive trait.

Pathogenesis

In cystic fibrosis, the pancreatic secretions are abnormally viscid and cause obstruction within the pancreas itself, leading to autodigestion of the acinar cells. This event causes abnormalities in the composition of meconium which becomes inspissated in the lower ileum, due to the absence of proteolytic enzymes, an excess of albumin, and hyperviscid ileal mucus. The obstructed ileum becomes filled with putty-like lumps of meconium and classically tapers from the grossly dilated empty proximal ileum to collapsed ileum distal to the obstructing meconium. There is a microcolon (Fig. 2). Complications such as volvulus, perforation, and atresia occur in 30 to 50 per cent of patients. The liver, sweat glands, epididymis, and lungs are also affected.

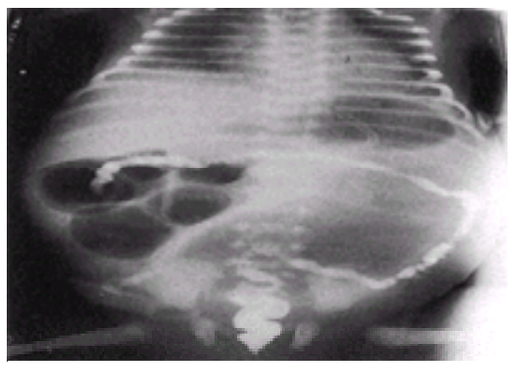


Fig. 2. Meconium ileus: the Gastrografin enema demonstrates the microcolon. The ileal distension is obvious on the right. There is free air in a large left-sided pseudocyst.

Clinical features

The features of low intestinal obstruction are usually apparent on the first day of life, with bile-stained vomiting and marked abdominal distension. Meconium is not passed, though bile-stained vomiting *in utero* can mimic meconium-stained liquor. Palpable, distended loops of bowel with indentable meconium may be felt. Radiology shows distended ileal loops, but few fluid levels. An intraluminal ‘ground-glass' appearance due to air and abnormal meconium may be present, especially in the right upper quadrant (Neuhauser's sign). Calcification is due to meconium peritonitis and a resulting inflammatory mass.

Treatment

Non-operative Gastrografin enemas refluxed into the terminal ileum will be successful in about 50 per cent of patients with uncomplicated obstruction (15 to 25 per cent overall). The Gastrografin must be diluted as its hyperosmolality is dangerous. It contains a wetting agent (Tween 80) which can free the viscid meconium. The baby must be well hydrated and carefully monitored, and an operating theatre must be available immediately in case of failure or complications (especially perforation).

Surgical The classic treatment was a Bishop–Koop or vented ileostomy. However, these require a second operation to close the stoma. Ileostomy has now been superseded by enterotomy and lavage using either Gastrografin or acetylcysteine (Parvolex), which liquefy the meconium. All of the tenacious meconium is removed. Any twisted or ischaemic bowel has to be excised and any atresia treated as above with a primary anastomosis, thus avoiding the complications of stomas in the neonate. The colon is flushed intraoperatively to exclude any distal obstruction. To prevent any postoperative meconium ileus equivalent, Pancrex can be administered immediately postoperatively via the nasogastric tube, as all modern enzyme preparations are too big to go down a nasogastric tube.

Results

Ninety per cent of babies survive the neonatal period, though their long-term prognosis will depend on the severity of their pulmonary disease (median survival in 1998 is 31 years). If cystic fibrosis is not the cause, and long-segment Hirschsprung's disease has been excluded, the postoperative course should be uneventful with no long-term problems.

Meconium plug

Some babies fail to pass meconium until a rectal examination or wash-out has stimulated the passage of a meconium plug. This consists of a clear mucous lower third (tip) with a green meconium upper two-thirds (body). The baby usually vomits and has a distended abdomen. The plug may be formed by over-reabsorption of water from the meconium by a colon with immature motility. Oral feeds can be started immediately. Most of these babies have normal bowel function, but cystic fibrosis and Hirschsprung's disease should be excluded routinely.

Meconium peritonitis

Meconium peritonitis (Fig. 3) is a result of antenatal perforation from any cause (such as meconium ileus). An intense chemical peritonitis following the presence of the meconium may result in calcification, dense adhesions, or the formation of a pseudocyst. Postnatally, if the perforation is still present, bacterial peritonitis will ensue. Perforation, adhesion, obstruction, and pseudocysts all require surgical treatment.

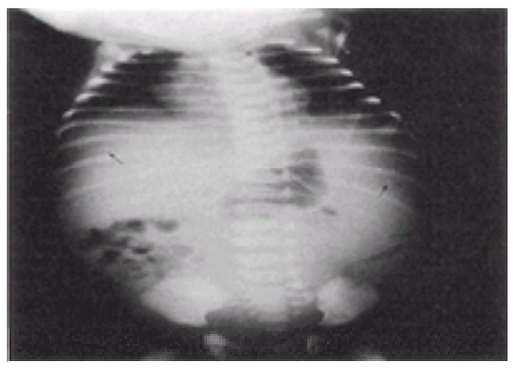


Fig. 3. Meconium peritonitis calcification (arrowed).

Malrotation

Embryology

Between 4 and 10 weeks of development, the intestines herniate into the umbilical cord. When they return to the abdomen, they usually rotate anticlockwise, the duodenum by 270° so that the duodenojejunal flexure lies to the left of the superior mesenteric artery and the caecum by 360° so that the transverse colon is anterior to the superior mesenteric artery, and the caecum is in the right iliac fossa. Complete failure of this complex process will result in the ‘universal mesentery' present in patients with congenital diaphragmatic hernia or exomphalos, whilst partial failure results in ‘malrotation'. The most common result is a caecum lying very close to the duodenojejunal flexure, with abnormal adhesions binding them both to the posterior abdominal wall. The resulting midgut mesentery is abnormally narrow, allowing volvulus to occur.

Clinical features

Malrotation presents in two main ways. A baby with intermittent bile-stained vomiting and possible abdominal distension who has passed meconium normally is often referred late. Alternatively, volvulus occurs, the baby collapses, and may be moribund, severely acidotic, and hyponatraemic, with an ischaemic and prenecrotic midgut. There may be very little to feel abdominally in either case.

Radiological examination after putting 40 to 50 ml of air into the stomach may show the distended first and second parts of the duodenum, or may show the ileum lying on the right (Fig. 4). Often there is no specific feature. Contrast studies may be necessary to confirm the diagnosis: a contrast meal tends to produce more diagnostic results than an enema. Non-ionic contrast should be used because of the possibility of perforation. Ultrasound examination may demonstrate an abnormal relationship between the superior mesenteric artery and the portal vein.



Fig. 4. Malrotation: the ileal corkscrew loop is clearly visible on this barium meal.

Treatment

Because of the risk of volvulus and subsequent midgut ischaemic necrosis, laparotomy should be performed as soon as possible after adequate urgent resuscitation and nasogastric decompression. Any volvulus will need to be unwound, usually in an anticlockwise direction. All abnormal adhesions between the caecum and the duodenum need to be divided (Ladd's procedure). By making the midgut mesentery as wide as possible, the completely freed caecum and colon can be placed on the left. A kink-free duodenum will lie on the right of the midline with the ileum. Inversion appendicectomy should be performed to avoid peritoneal contamination and the diagnostic difficulty of appendicitis with a left-sided caecum.

If there is massive necrosis, obviously dead bowel should be resected. All doubtfully viable bowel should be left *in situ* and both ends either exteriorized or closed with Ligaclips and left intraperitoneally. A second-look laparotomy the next day may show more recovery than had been expected. A primary anastomosis of viable bowel may be possible, avoiding the morbidity associated with ileostomies in neonates.

Results

Provided that there are no major associated abnormalities and no resection is necessary, survival should approach 100 per cent. Massive intestinal loss (short bowel syndrome) and associated abnormalities account for most of the deaths in recent series.

Necrotizing enterocolitis

Necrotizing enterocolitis is one of the most common single presentations on a surgical neonatal unit. It was almost unknown before the mid-1960s, but is now seen as increasingly intensive management of sick premature neonates means that more of these babies survive. Although premature babies are the predominant group with necrotizing enterocolitis (90 per cent), term babies who are otherwise well may develop the condition, especially following surgery either for gastrointestinal or completely unrelated conditions.

Pathogenesis

Necrotizing enterocolitis appears to be the final common pathway or end result of a wide variety of stresses in a susceptible baby. Usually these stresses are multiple, though a single cause may occasionally be identified. The risk of necrotizing enterocolitis is inversely proportional to birth weight and is often associated with episodes of hypoxia, hypothermia, hypotension, hyperviscosity, acidosis, or the presence of free oxygen radicals. Umbilical artery cannulas, exchange transfusion, hyperosmolar feeds, packed cell transfusion, or overdosage with calcium antagonists are also known risk factors. All of these could cause mesenteric ischaemia which will allow bacterial invasion of the mucosa and bowel wall leading to necrosis. This ischaemia may mimic the ‘dive reflex' in seals when the blood is diverted away from the bowel during episodes of hypoxia.

Pathology

Necrotizing enterocolitis can affect any part of the bowel, but the terminal ileum and caecum, the left colon, and the entire colon are predominately affected. Macroscopically the bowel is dilated and friable, with areas of haemorrhage and necrosis. The mucosa is ulcerated and sloughed. Microscopically, there is a complete spectrum from coagulation necrosis of the mucosa, through ulceration and haemorrhage, to transmural necrosis which is histologically identical to Hirschsprung's enterocolitis. Healing is by epithelialization but fibrosis may cause stricture formation.

Presentation and clinical features

Most babies with necrotizing enterocolitis present with bilious vomiting, abdominal distension, colour change, bradycardia, apnoea, lethargy, or poor response to handling. Gross or occult blood in the stools is seen in 75 per cent of patients. Initially the abdomen is soft, but erythema and oedema of the anterior abdominal wall often coincide with a mass of inflamed bowel. Thrombocytopenia and neutropenia indicate disseminated intravascular coagulation and septicaemia, often associated with acidosis.

The classic diagnostic finding is pneumatosis intestinalis on plain abdominal radiographs. This event appears as linear gas shadows in the bowel wall which are most dramatic when the bowel is seen end on (Fig. 5). The presence of portal vein gas increases the mortality by 20 per cent, though the exact mechanism of this is unclear. Free intraperitoneal air implies perforation, except in the presence of pneumomediastinum. A persistent gas-filled loop in a series of radiographs implies the presence of infarcted bowel.

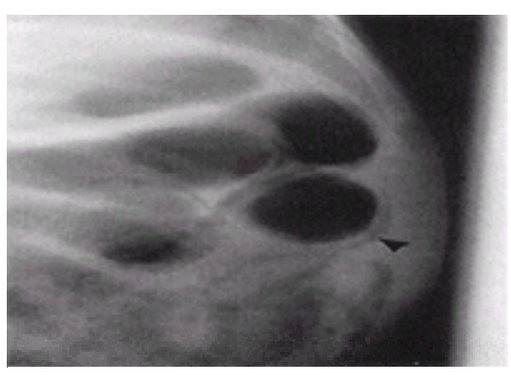


Fig. 5. Necrotizing enterocolitis: the intramural gas, seen end on, is clearly visible (arrowed).

Management

The baby is aggressively resuscitated with plasma, morphine, blood transfusion if necessary, additional oxygen or ventilation, and, following blood cultures, antibiotics. Antibiotics active against a broad spectrum, including anaerobes, must be used: most centres use metronidazole, an aminoglycoside (e.g. gentamicin), and a penicillin or third-generation cephalosporin.

All feeds are stopped for a minimum of 7 to 10 days, a nasogastric tube is passed to allow gastric decompression, and intravenous feeding is started. Regular clinical monitoring of the baby together with measurement of the haemoglobin, platelets, and blood gases allow correction of any abnormality. Hypovolaemia is readily detected using the toe-core temperature difference, which should be less than 3°C. If there are coagulation problems, fresh frozen plasma should be used rather than simple plasma. Serial radiographs are costly and unnecessary. The best method of pain relief is opiate analgesia by continuous intravenous infusion, with ventilation if required.

The optimal time for operation is not in the acute period, though whether it is better to operate after stabilization for a few days or electively for complications arising during convalesence is not known. Pneumoperitoneum is no longer an absolute indication for surgery as silent perforations are frequently recognized. The desperately ill, but stable neonate can withstand the presence of much more intestinal necrosis than would be possible in an adult, but suffers the stress of laparotomy poorly. In a baby who is deteriorating despite maximal support, laparotomy is the only option. However, in a stable neonate, non-surgical management should be maintained to allow as much recovery as possible. Neonates with very low birth weight and extensive disease may benefit from percutaneous drainage under local anaesthesia. Despite extensive clinical or radiological disease not all neonates with necrotizing enterocolitis require laparotomy, either acutely or at a later date.

At laparotomy, all necrotic bowel must be excised, but provided the two bowel ends are viable and bleed well, primary anastomosis is safe and avoids the morbidity associated with stomas in neonates. Multiple areas of necrosis may require multiple anastomoses. Delayed surgery following prolonged medical management, usually for colonic stricture, is safer than emergency laparotomy.

Postoperatively many babies seem to do better on a homogenized lactose-free milk formula such as pregestemil. If a stoma has been created, the sodium loss from this may cause failure to thrive and additional supplementation is required (an additional 0.13 mmol sodium/ml ileostomy loss or 1 ml of 8.4 per cent sodium hydrogen carbonate/3- to 4-hourly feed).

Results

The mortality rate of 40 to 50 per cent in the 1960s has been reduced to 25 per cent with more aggressive medical support and the introduction of metronidazole. Between 10 and 30 per cent of survivors develop a colonic stricture, whether they were operatively or conservatively managed, so this must be looked for clinically. Short gut syndrome may result from extensive resection.

Older children

Infantile hypertrophic pyloric stenosis

Although this condition can occur within the first week of life, the age of presentation is usually between 10 days and 10 weeks, with a peak at about 4 weeks, independent of gestational age. Therefore the condition should probably be called infantile rather than congenital hypertrophic pyloric stenosis.

Incidence

More boys than girls are affected, in a ratio of 4:1. Over half the infants are firstborn. The incidence varies throughout the world, but infantile hypertrophic pyloric stenosis affects approximately 2 per 1000 live births in the United Kingdom.

Aetiology

The aetiology must be multifactorial. There is a strong familial element, with a 5 per cent incidence in children with affected mothers. There is a seasonal variation (more common in the winter) and a relation to social class. Many hypotheses have been put forward to explain the hypertrophy, but whether the factors are cause or effect is uncertain. Increased stress causing a hormonal effect both in pregnancy and postnatally has been suggested, with gastrin as the possible hormone.

Immunocytochemical studies show a reduced number of nerve fibres containing vasoactive intestinal peptide and encephalin, but the pyloric ganglia otherwise seem normal.

Pathology

Macroscopically there is an olive-shaped tumour up to 2.5 cm in diameter which blends into the antral musculature. At the duodenal end, the pylorus bulges into the lumen like the cervix into the vagina, creating a circumferential fornix. The stomach is markedly dilated.

Microscopically there is marked hypertrophy of the circular muscle of the pylorus and some hypertrophy of the longitudinal muscle, together with marked mucosal oedema. There is lengthening of the pyloric canal.

Clinical features

Most babies with infantile hypertrophic pyloric stenosis feed normally initially and then gradually vomit with increasing frequency and force. The vomit is not bile-stained, consists of milk curds and mucus, and becomes projectile. Following a vomit, the baby will generally feed again greedily. Twenty per cent of patients have haematemesis due to oesophagitis. The production of free fatty acids in the stomach causes gastritis.

Most babies have lost weight by the time of presentation, either due to dehydration or to loss of subcutaneous fat. The absence of weight loss puts the diagnosis in doubt. It is crucial to examine the baby in the correct position to maximize the chance of the correct diagnosis. The baby should have a test feed from the mother's left breast or be held in the left arm. Milk is better than dextrose. The mother and baby should be seated between a window and the examiner, who sits on the baby's left side. The entire abdomen is exposed and the legs supported to relax the abdominal muscles. Talcum powder on the abdomen allows the fingers to move around more easily. Visible peristalsis of the hypertrophied stomach may be seen through the anterior abdominal wall coursing from the left towards the midline (Fig. 6(a)).

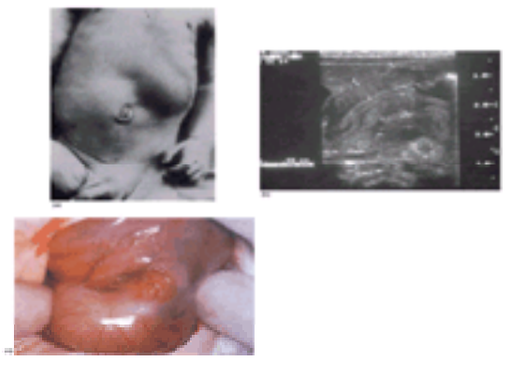


Fig. 6. (a) Pyloric stenosis: visible peristalsis in a thin boy. (b) Pyloric stenosis: the elongated thick pylorus is clearly seen on this longitudinal ultrasound. The muscle

thickness (1) is 0.55 cm. (c) Pyloric stenosis: the enlarged pylorus at laparotomy to show the relatively avascular superior part.

The examiner, using his left hand, attempts to feel the pylorus, or pyloric ‘tumour', by trapping it against the vertebrae, either round the edge of the right rectus muscle or through the linea alba. It has been likened to ‘feeling a walnut through a blanket'. Projectile vomiting, weight loss, and a palpable pylorus make the diagnosis certain.

If the initial test feed is negative, it should be repeated 4 to 6 h later. If still negative, a more experienced clinician needs to repeat it again. If there is still doubt then an ultrasound should be diagnostic (Fig. 6(b)).

Preoperative management

Because of the obstruction at the pylorus, these babies produce the classic metabolic picture of hydrogen and chloride ion loss. This hypochloraemic metabolic alkalosis with dehydration is complicated by the renal response, which involves potassium excretion in exchange for hydrogen ions. These babies must therefore be resuscitated with at least 0.5 N saline solution (0.45 per cent sodium chloride) in 10 per cent dextrose plus 2 g of potassium chloride per 500 ml bag initially to avoid profound hypokalaemia and failure to correct the alkalosis. Only when the baby is electrolytically normal (bicarbonate less than 25 mmol/l) should the child be taken to theatre. The stomach should be lavaged to ensure that all the milk curds are washed out.

Treatment

Medical therapy using an anticholinergic drug is no longer practised.

Infantile hypertrophic pyloric stenosis is probably the only condition for which all the surgeons in the world use the same operation: Ramstedt's pyloromyotomy. Through a right upper transverse rectus-cutting incision, the greater curve of the stomach is delivered, followed by the pylorus. A supraumbilical incision is cosmetically superior, but may have a greater complication rate. The pylorus is rotated so that the relatively avascular superior part (Fig. 6(c)) is uppermost and the serosa is incised from near the pyloroduodenal junction at least 2 cm on to the antrum. The muscle fibres are split by spreading them using an artery clip or a purpose-made spreader proceeding proximally until the fibres appear to tear tangentially. The mucosa bulges up into the defect. Distally, the muscle fibres are split almost to the pyloroduodenal junction, though a 100 per cent split of the tumour is not necessary, as a 95 per cent split works just as well. Because of the duodenal fornix, perforation is easy, but is simply oversewn with an omental buttress. Air (40 to 50 ml) is syringed down the nasogastric tube by the anaesthetist and squeezed through into the duodenum. Any leak is obvious. Haemostasis if usually unnecessary. Careful closure in a well-resuscitated baby should ensure that the incidence of wound infection and dehiscence is minimal.

Feeds are withheld for 24 h postoperatively or until the following morning, when full feeds are restarted. There is no need for complex feeding regimens.

Results

There should be no mortality, though wound dehiscence and infection (5 to 10 per cent) still occur.

Intussusception

Intussusception is the most common cause of intestinal obstruction in infants aged 6 to 18 months.

Pathology

An intussusception is most common in the terminal ileum and can only occur in the presence of a mobile ‘partly malrotated' caecum and terminal ileum. When a lead point exists in the bowel wall, peristalsis causes that part of the bowel to invaginate and progress along the lumen, through the ileocaecal valve and along the colon. Although most intussusceptions only reach the transverse colon, if there is sufficient mobility they may present at the anus. Lead points may be an inflamed Peyer's patch, a Meckel's diverticulum, or a polyp (e.g. in Peutz–Jeghers syndrome). Obstruction occurs due to oedema of the intussuscepted bowel, which may become ischaemic and necrotic.

Clinical features

Infants with intussusception present classically with intermittent attacks of intestinal colic lasting a few minutes, characterized by a high-pitched scream, pallor, curling up their knees, and vomiting. The vomit is usually bile stained. Between bouts of colic the infant is quiet, listless, and anorexic. There may be a history of recent immunization or a viral illness. Rectal bleeding (redcurrant jelly stool) (Fig. 7(a)) is said to be a late sign, but this is by no means always the case.

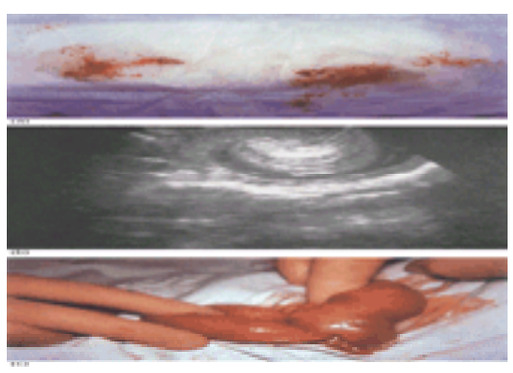


Fig. 7. (a) Intussusception: redcurrant jelly stool. (b) Ultrasound of intussusception to show ‘doughnut' appearance (c) Intussusception: terminal ileum intussuscepted within the caecum at laparotomy.

Examination between bouts of colic often reveals a soft abdomen which is remarkably non-tender. A sausage-shaped mass in the right upper quadrant or epigastrium is palpable in about 30 per cent of cases. Rectal examination may reveal the apex of the intussusception or blood on the glove. Bowel perforation is a rare event but leads to widespread peritonism.

Investigation

A supine plain abdominal radiograph should reveal absence of the caecal gas shadow from the right iliac fossa. A soft tissue mass may be obvious. The apex of the intussusception may be visible in the transverse colon, outlined by the distal gas. Any free gas due to perforation is best demonstrated on an erect chest radiograph. If there has been excessive vomiting, serum urea and electrolyte levels should be measured. An ultrasound will demonstrate the ‘doughnut' of the intussusception (Fig. 7(b)).

Management

All infants with intussusception need resuscitation prior to treatment. A minimum of 10 ml/kg of plasma should be given routinely; even clinically ‘well' infants may need as much as 30 to 40 ml/kg eventually. All patients with intussusception have considerable ‘third space' losses due to oedema of the bowel wall, fluid loss into the ileal lumen, and peritoneal exudation.

Reduction by enema should be attempted in all except very dehydrated and sick infants, in whom bowel necrosis has already occurred, or in those with free

intraperitoneal air. A long (more than 48 h) history of intussusception is not a contraindication to enema reduction. The exact period of obstruction is difficult to define and, in a well baby, enema reduction may be successful despite a ‘long' history. Intussusception into the rectum implies extensive bowel mobility and an easy enema reduction, not the reverse.

Premedication with Omnopon (papaveretum) 0.2 mg/kg and scopolamine 0.004 mg/kg intramuscularly 1 h prior to the enema is helpful as Omnopon inhibits the desire to defecate and scopolamine is a smooth muscle relaxant. Hydrostatic reduction using barium sulphate has been replaced by air insufflation under ultrasound control with up to a 75 per cent success rate. It has the advantage that should perforation occur, there is no soiling of the peritoneum with barium. The apex of the intussusception can usually be identified and watched as it reduces. Free reflux of air into the terminal ileum defines a successful reduction. Failure of reflux may be due to an incomplete reduction or an oedematous ileocaecal value; clinical re-examination is mandatory. If the infant is well, relaxed, and no longer having bouts of colic, then the intussusception has been reduced. If there is still clinical evidence of intussusception, then a urther enema, 4 to 6 h later, when the oedema of the bowel has reduced, has a high success rate. An unnecessary laparotomy for a reduced intussusception should be avoided.

If there is evidence of necrosis, or if the intussusception is irreducible, laparotomy is performed through a right iliac fossa muscle-splitting incision, just below the level of the umbilicus (Fig. 7(c)). Reduction is achieved by gently squeezing the mass to reduce the oedema rather than by pulling on the ileum, which will cause tearing. Any necrotic bowel, Meckel's diverticulum, or polyp will need to be resected. If possible, an appendicectomy should also be performed as this will encourage fixation of the caecum in the right iliac fossa and avoid any future confusion as to the reason for the abdominal scar.

Ileoileal intussusception requires laparotomy and resection of the offending polyp, the usual pathology, together with the minimum amount of normal ileum.

Results

There is a 5 to 10 per cent recurrence rate following either enema reduction or laparotomy without resection. A repeat enema should be performed, but at the third recurrence, laparotomy is justified. In children over 3 years old, enlarged Peyer's patches are rare and laparotomy is more likely to be required.

Constipation

Ten per cent of babies pass meconium at birth and a further 85 per cent pass meconium within the first 48 h. After that, breast-fed babies usually pass more motions than bottle-fed babies, though the absolute frequency is very variable. Normality for adults is between three stools per day and once every 3 days.

Constipation may present at birth due to conditions such as anal stenosis and Hirschsprung's disease, but it more commonly arises after 2 years of age. The most common cause is an anal fissure. A hard motion tears the anal verge, causing pain and spasm. The child holds on because it knows that defecation will be painful, and so a further hard motion is formed which eventually passes with more tearing, pain, and spasm in a vicious circle. A distended colon peristalses poorly and increases bowel transit time. The rectum is always full and soiling may occur. Behavioural changes, aggression, and moodiness are common and are due to absorption of toxins from the colon. Even after the fissure has healed the abnormal bowel habit remains.

Early aggressive treatment of fissure in children with a combination of stool softeners such as lactulose and stimulant laxatives such as Senokot (in adequate dosage) works well, together with a local anaesthetic ointment such as lignocaine 5 per cent. An anal stretch under general anaesthetic which disrupts the pecten bands and abolishes the spasm may allow healing, but may have long-term complications. Lateral internal sphincterotomy may be safer.

Established constipation requires plenty of time and a full explanation to a family which is often completely fixated on the child's bowel habit. Regular rectal emptying allows the re-establishment of rectal sensation, the abolition of soiling, and the promotion of a regular bowel habit. Family interactions should improve once the parents can be convinced that they can control their child's bowels, and improvement is seen.

If meconium is not passed within 48 h of birth, or if the bowel problem began in the neonatal period, the diagnosis of Hirschsprung's disease needs to be excluded by a full-thickness rectal biopsy which is sent to a specialist paediatric pathologist..

Rectal bleeding

The reasons for rectal bleeding in a child vary with age (Table 1) and are usually diagnosable on history and examination. General causes such as haemorrhagic disease of the newborn are rare. Children rarely present with massive severe rectal bleeding. Brisk bright red bleeding is usually due to bleeding from an ulcer opposite a Meckel's diverticulum. Chronic, undiagnosed rectal bleeding despite a series of negative investigations has a 50 per cent chance of diagnosis at laparotomy.

Neonate	Toddler	Child
Necrotizing enterocolitis	Anal fissure	Juvenile polyp
Milk allergy	Intussusception	Anal fissure
Duplication cyst	Juvenile polyp	Meckel's diverticulum
Anal fissure	Gastroenteritis	Rectal prolapse
Midgut volvulus	Meckel's diverticulum	Inflammatory bowel disease
	Rectal prolapse	External haemorrhoidal venous plexus

Table 1 Causes of rectal bleeding

Inflammatory bowel disease

Crohn's disease and ulcerative colitis are more common in adults than in children and are covered in detail elsewhere. Anyone treating children with inflammatory bowel disease needs to remember the psychological stresses that occur within the family of a child with a severe, chronic illness which may affect growth.

Ulcerative colitis is rare before the early teens and usually presents with diarrhoea, abdominal pain, and mucus and blood *per rectum*. Toxic megacolon is very rare (less than 2 per cent), weight loss and fever are common (80 per cent), and arthritis is unusual (10 per cent). An endoscopic examination under general anaesthetic allows a biopsy to be obtained for histological confirmation. An early barium enema is often normal. Treatment ranges from Salazopyrin for mild cases through rectal steroids to systemic prednisone (1 to 2 mg/kg.day). Severe ulcerative colitis requires admission to hospital, maximal fluid support, antibiotics active against anaerobes, such as metronidazole, and intravenous hydrocortisone. Intravenous feeding is not helpful in inducing remission or predicting a need for colectomy. Regular endoscopic review is required for carcinoma surveillance. Colectomy is curative, but the psychological effect of a permanent ileostomy is variable. Continent intra-abdominal pouches are still improving, though the risk of a major complication is still formidable. Early diagnosis and treatment is associated with improved growth and general health. Chronic symptoms are present in 60 to 70 per cent of patients. Pancolitis is more likely to result in carcinoma, but the risk is not related to age of diagnosis. Although the chance of carcinoma is probably less than was originally thought, total colectomy is still performed in about 25 per cent of children with ulcerative colitis.

Fifteen per cent of patients with Crohn's disease present before the age of 15 years. The symptoms may mimic those of ulcerative colitis, or may be more subtle, with failure to thrive, anorexia, mouth ulcers, malabsorption, fistula formation, or perianal disease. Weight loss and fever are common, but arthritis is uncommon (15 per cent). The erythrocyte sedimentation rate and C-reactive protein level are usually raised (90 per cent) and a small bowel meal may be diagnostic. Infectious diseases, including yersinia, must be excluded. The treatment of Crohn's disease has not been so well studied in children as in adults. Prednisone is probably better than Salazopyrin but has a greater effect on growth potential. Because of their oncological risk, 6-mercaptopurine and azathioprine are reserved for use in steroid-dependent children with severe disease. Enteral feeding has a major role in decreasing symptoms, inducing remission, closing fistulas, improving postoperative growth, and in the treatment of simple growth failure. At least 70 per cent of patients will require surgery within 7 years of diagnosis, usually for obstruction, fistulas, or growth failure. To improve the chance of normal growth, operation should be undertaken earlier than is usual in adults, though only the minimum surgery should be performed (minimal resection and primary anastomosis). Fifty per cent of children think that their quality of life is worse than that of their peers. Although eventual

operation is likely, the mortality from the disease is low.

Further reading

Atwell JD, ed. *Paediatric surgery*. Arnold, London, 1998. [Registrar/senior registrar grade textbook of paediatric surgery.]

Claydon GS, Agnarsson U. *Constipation in childhood*. Oxford Scientific, Oxford, 1991. [Excellent user-friendly guide to constipation.]

Freeman NV *et al.*, eds. *Surgery of the newborn*. Churchill Livingstone, London, 1994. [More detailed description of neonatal surgical problems with some operative guidelines.]

Welch KJ *et al. Pediatric surgery*, 5th edn, (Vols 1 and 2). Year Book Medical Publishers Inc., Chicago, 1998. [Major reference source for all paediatric surgical problems.]

42.5 Duplications of the gastrointestinal tract and mesenteric cysts

Robert C. Shamberger

Duplications of the gastrointestinal tract
Esophageal duplications
Thoracoabdominal duplications
Gastric duplications
Pyloric duplications
Duodenal duplications
Peripancreatic duplications
Intestinal duplications
Colonic duplications
Rectal duplications
Mesenteric cysts
Further reading

Duplications of the gastrointestinal tract

Duplications of the gastrointestinal tract occur from the tongue to the anus. Classification was simplified by Ladd in 1937, when he pointed out that lesions throughout the gastrointestinal tract, previously termed enteric cysts, enterogenous cysts, diverticula, giant diverticula, ileum duplex, jejunum duplex, and unusual Meckel's diverticula, shared common characteristics and treatment.

Duplications have three common features regardless of their location. They are hollow structures, lined with gastrointestinal-tract epithelium, which have a wall of smooth muscle often shared with the contiguous intestinal tract. The epithelial lining of the duplication may be different (heterotopic) from that of the adjacent bowel. Symptoms vary depending upon the location of the duplication, but are produced by the mass effect of the cyst, which causes obstruction of the intestinal tract or airway, or by hemorrhage or perforation caused by acid secretions from heterotopic gastric mucosa within the cyst. Rarely, volvulus of an affected segment of bowel may occur. While duplications may present themselves at any age, they generally are identified in infancy or childhood.

In considering the cause of duplications, Lewis and Thyng identified diverticula of the intestinal tract budding into the subepithelial connective tissues during development. These diverticula, they postulated, may become isolated from the intestinal lumen to form cysts; however, the diverticula occur around the entire circumference of the bowel and are not limited to the mesenteric border, where duplications are found. Incomplete resolution of the solid phase of the intestinal tract, when the lumen of the intestine is entirely filled with epithelial cells during early development, may produce duplications. Such cysts, however, would not be limited to the mesenteric border of the bowel nor surrounded by a separate smooth-muscle wall. In human embryos the solid stage of enteric development has not been identified distal to the duodenum. A more acceptable explanation for some duplications involves the persistence of the neuroenteric canal, the embryonic communication between neural ectoderm and the gastrointestinal endoderm. Such an event, resulting in the canal extending through the mesodermal layer that forms the spine (the notochord), will result in a dorsal enteric fistula or communication between the intestine and the dorsal skin surface, passing through a bifid spinal cord. Lesser degrees of this process with only partial resolution of the canal may result in cystic duplications; in some cases, duplications are associated with abnormalities of the spine.

Extensive colonic duplications appear to be a result of partial ‘caudal twinning' because they are often associated with duplication of the urinary tract and external genitalia. Their cause may be distinct from that of other forms of duplication. Duplications are occasionally seen in association with atresia of the intestine: this was noted in 11 per cent of 68 cases reported by Gross and associates. Multiple atresias are seen in 15 per cent of patients.

The frequency of duplications throughout the intestinal tract is shown in Fig. 1. Duplications occur most frequently in the ileum; the second most common site is the esophagus. Although Ladd stated that none of the patients in his initial report was diagnosed before surgery, most are now identified by imaging before exploration. Thoracic lesions are often discovered incidentally on radiographs of the chest; a barium swallow or CT scan may be required to define the lesion fully. Abdominal duplications are reliably defined on ultrasonographic examination, because of their cystic nature. The radionuclide [$^{99}Tc^{m}$]pertechnetate macroaggregated albumin scan can demonstrate the presence of heterotopic gastric mucosa within the duplication.

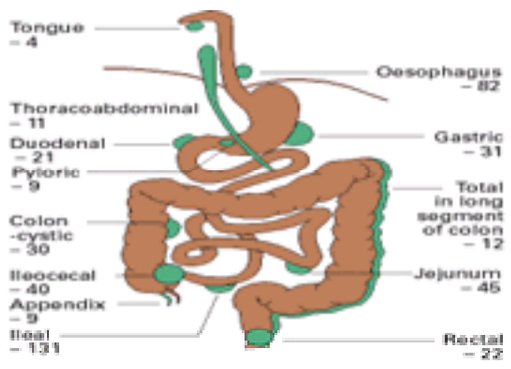


Fig. 1. The frequency of gastrointestinal duplications shown by location, as summarized from series reported by 10 institutions.

Resection of duplications is recommended once they have been identified, since large cystic lesions of the colon and ileum may perforate after progressive distension by mucosal secretions or may cause intestinal obstruction. Neoplasia has been described at almost every site of duplication, but usually occurs in patients 26 to 65 years old. Both squamous carcinoma and adenocarcinoma have been reported.

Esophageal duplications

Esophageal duplications may produce dysphagia or respiratory distress, particularly when present in the cervical esophagus or between the trachea and esophagus in the upper mediastinum. Many, however, are asymptomatic and are identified on chest radiographs obtained for other reasons (Fig. 2).

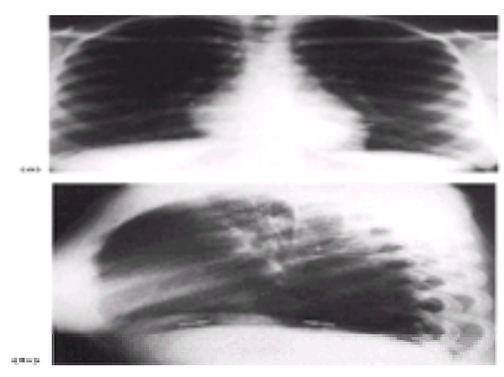


Fig. 2. (a) Chest radiograph obtained because of a cough in a 14-year-old girl revealed an unrelated retrocardiac shadow (arrows); (b) the lateral radiograph confirmed its presence anterior to the spine (arrows).

Esophageal duplications are resected through a thoracotomy. Recognition of the shared wall between the cystic duplication and the esophagus is critical to avoid injury to the esophagus. The muscular layer overlying the cyst is opened and the mucosal portion of the cyst is dissected from the shared esophageal wall. The epithelial lining may be squamous, ciliated respiratory, or, less frequently, ectopic intestinal or gastric mucosa. Communication with the esophageal lumen is rare. Bronchogenic cysts and esophageal duplications are felt to arise from the same process, the embryonic separation ofthe esophagus from the trachea. Bronchogenic cysts may have a squamous epithelial lining similar to the esophagus since they share a common embryological anlagen. They are distinguished from esophageal duplication cysts only by their lack of proximity to the esophagus.

Thoracoabdominal duplications

Large cystic duplications within the thorax may communicate distally with the duodenum or intestine. Patients harbouring these duplications may present with respiratory symptoms secondary to the large size of the duplication, sepsis from intestinal colonization, or hemorrhage from peptic erosion arising from heterotopic gastric mucosa. Large thoracoabdominal duplications extending into the thorax through the diaphragm are often adherent to the spine and are frequently associated with spinal anomalies. Complete resection is required and may be done through a thoracoabdominal incision or through separate thoracic and abdominal incisions.

Gastric duplications

Gastric duplications are generally identified within the first year of life. Presenting symptoms and signs include a palpable abdominal mass, vomiting, weight loss, abdominal pain, anemia, melaena, hematemesis, or fever. Of the 83 cases reviewed by Pruksapong most occurred along the greater curvature (54), with fewer along the lesser curve (7), anterior wall (5), posterior wall (9), or in the region of the pylorus (1). In seven cases the location was not well described. Communication with the gastric lumen is unusual unless it occurs secondary to peptic erosion. Rudimentary pancreatic duplications are also associated with gastric duplications, into which their ducts may drain.

Gastric duplications require surgical excision. A limited partial gastrectomy of the shared wall is often done for lesions of the greater curve, but in lesions of the lesser curve a submucosal resection can be performed.

Pyloric duplications

Pyloric duplications are very rare. Gastric-outlet obstruction is often the presenting symptom and resection of the duplication may be curative.

Duodenal duplications

Duodenal duplications are rare and occur primarily in the first and second portions of the duodenum, on the pancreatic margin. In 50 cases of duodenal duplication summarized by Leenders and associates, 45 were posterior to the duodenum; only four communicated with the lumen. In contrast to what are termed peripancreatic duplications, they do not communicate with the pancreatic duct. Because of their location, submucosal resection to remove gastric mucosa is often required to avoid pancreatic injury. A cystoduodenostomy may be performed for lesions without gastric mucosa, but resection is preferred and marsupialization into the peritoneal cavity is to be avoided.

Peripancreatic duplications

Peripancreatic duplications present with symptoms of nausea, vomiting, weight loss, and abdominal pain. The majority (62 per cent) occur within the head of the pancreas, and most communicate with the pancreatic duct. Most contain gastric mucosa (90 per cent). Intermittent bathing of the pancreatic duct by acid secretions produces pancreatitis and the presenting symptoms. Identification of these lesions may be difficult, but they should be sought in every patient presenting with ‘idiopathic' pancreatitis in childhood. Ultrasonographic examination shows a cystic area within or adjacent to the pancreas. Endoscopic retrograde cholangiopancreatography may demonstrate communication with the pancreatic duct, but if inflammation is present the cyst may not fill.

At surgery the mucosal lining of the cyst must be removed if the cyst itself cannot be resected completely. Pancreatoduodenectomy has been performed in some patients in whom lesser options were not suitable to achieve complete removal of the cyst or its lining.

Intestinal duplications

Intestinal duplications are found within the leaves of the mesentery, with the blood supply to the intestine draped over their surface. Smaller lesions may be missed if the mesentery is not closely inspected. They occur as either a long, tubular duplication extending along a significant length of small intestine, or as small, spherical cysts. Signs and symptoms include obstruction caused by the spherical lesions and, occasionally, intussusception. Less frequently, intestinal bleeding, perforation, or abdominal pain occur when ectopic gastric mucosa is present.

Limited spherical duplications, which are generally lined with intestinal mucosa, are treated by resection of the cyst and its overlying intestinal segment, with primary anastomosis. Long, tubular lesions present a greater challenge. Their removal is important because they often contain heterotopic gastric mucosa and communicate distally with the intestinal lumen. Acid secretions produce ulceration within the duplication or in the ileum distal to its communication with the duplication. Wrenn described a technique for submucosal resection of the duplication through a series of transverse incisions in the seromuscular wall of the cyst placed between mesenteric vessels (Fig. 3). This technique allows the entire mucosal lining of the duplication to be excised without resection of any intestine. Resection of the entire cyst between the leaves of the mesentery can be done only if the muscular wall is not shared with the intestine, an unusual occurrence. An innovative approach to this problem is the creation of an anastomosis between the stomach and the distal end of a complete intestinal duplication containing gastric mucosa. This diversion of the acid secretions of the duplication into the stomach prevents ileal or colonic peptic ulceration, and long-term follow-up has shown no evidence of blind-loop syndrome or malabsorption.

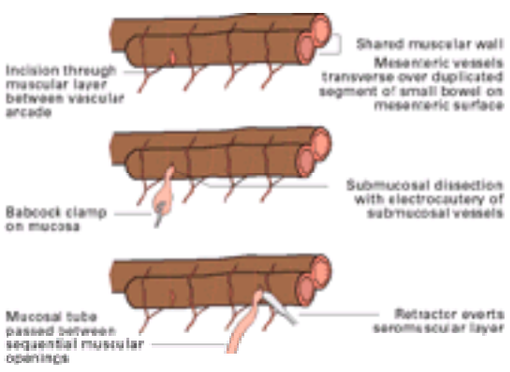


Fig. 3. Technique for removal of the mucosa in extensive tubular intestinal duplications as described by Wrenn. Multiple transverse myotomies into the duplication are made between mesenteric vessels, with submucosal resection of the mucosa. The shared muscular wall between duplication and intestine is depicted in transverse section.

Colonic duplications

Colonic duplications also occur in two forms. Limited spherical duplications occur most frequently at the cecum and are managed by partial colectomy (Fig. 4). Long, tubular duplications may extend from the ileum to the anus, and are often associated with other anomalies (80 per cent), including spina bifida and duplications of the bladder, urethra, or external genitalia. The proximal end of a tubular colonic duplication invariably communicates with the bowel lumen; half of all tubular duplications have a fistula at the distal end, connecting to the urinary tract, vagina, vulva, or perineum. Duplicated anus is rare. Tubular duplications without a distal fistula become

massively distended in the distal portion of the duplication.

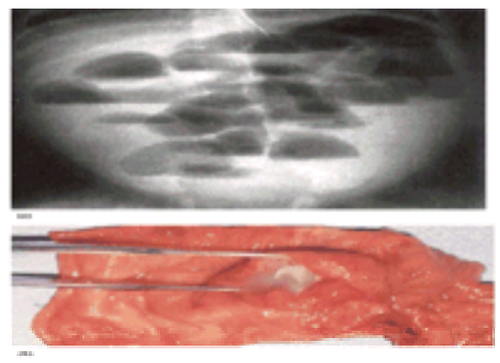


Fig. 4. (a) Radiograph of a 10-week-old infant who presented with intestinal obstruction. (b) A cecal duplication was identified preoperatively by ultrasound performed to evaluate a palpable right lower-quadrant mass. The shared wall between the cecum and the duplication has been opened, demonstrating the mucus filling the duplication and producing obstruction by compressing the ileocecal valve.

Tubular duplications of the colon occur on its mesenteric or antimesenteric surface, suggesting a different embryological origin from that of the more proximal duplications, which invariably occur on the mesenteric surface. The mucosal lining of these lesions is generally colonic mucosa. A radionuclide scan will exclude the presence of ectopic gastric mucosa in the duplication, which would require resection. Surgical management of total colonic duplications lined by colonic mucosa involves division and proximal dissection of the distal fistula to the urinary tract or perineum. A fenestration is then placed between the abnormally ending duplication and the colon at a site where they share a common wall (Fig. 5). The duplication segment should be excised beyond the fenestration to avoid distension of a blind-ending segment.

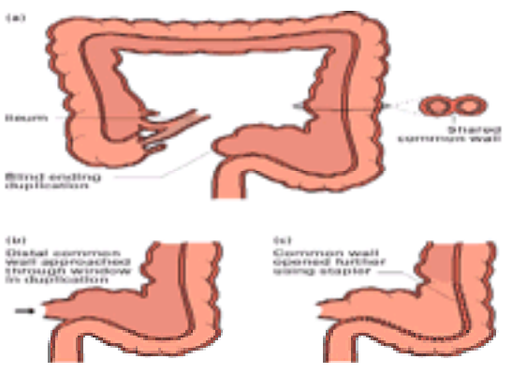


Fig. 5. (a) Depiction of a total colonic duplication without a distal fistula. Note the shared wall between the duplication and the colon (inset). (b) The distal portion of the duplication is opened to provide exposure of the shared wall, which is opened. (c) This common wall is then opened proximally to create wide communication between the duplication and the colon.

Rectal duplications

Rectal duplications are spherical and occur between the rectum and the sacrum. The differential diagnosis of these lesions includes anterior myelomeningocele and sacrococcygeal teratoma. Rectal duplications are best managed by total excision from a transanal or perineal approach. Simple excision of the common wall between the rectum and duplication leaving the duplication *in situ* is not recommended because stool in the diverticulum produced by the duplication is not effectively evacuated.

Mesenteric cysts

Mesenteric cysts are cystic lesions of the intestinal mesentery lined with endothelium. Symptoms caused by them are primarily pain resulting from traction on the mesentery or partial intestinal obstruction, vomiting, and failure to thrive; obstruction arises due to recurring segmental volvulus or chronic partial obstruction (Fig. 6). Because the cysts are soft, a mass is rarely palpable. Although radiographs reveal only a gasless area in the abdomen, ultrasonography will demonstrate a large cystic lesion.

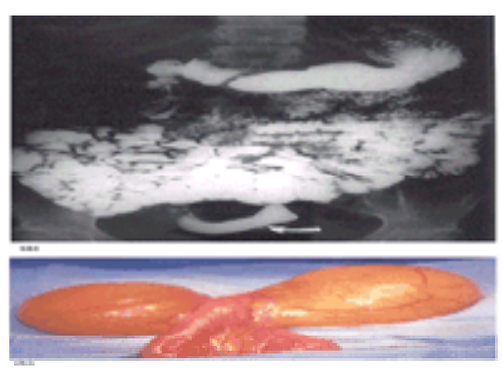


Fig. 6. (a) A barium upper-gastrointestinal series from an 18-year-old woman with symptoms of intermittent intestinal obstruction and a weight loss of 11.4 kg (25 lb); an abnormal loop of intestine is present in the distal ileum (arrow). (b) Mesenteric cyst removed at exploration. Preoperative ultrasound examination also demonstrated a cystic mass in the lower abdomen. The thin wall of the cyst reveals the serous colour of its contents. The cyst was intimately adherent to the bowel, and its removal required a segmental resection of the involved intestine with the cyst.

Only 700 cases had been reported by 1975. The primary cause of these lesions is probably abnormal development of the lymphatic system of the mesentery, resulting in stasis and progressive distension of lymphatic channels.

Mesenteric cysts are unilocular or multilocular. In contrast to the intestinal duplications they lack a muscular wall and are lined with endothelial cells rather than intestinal mucosa. Their contents are serous or chylous and rarely bloody, unless associated with recent trauma. Most mesenteric cysts are found in association with small intestine, less frequently in association with the mesocolon of the transverse, sigmoid, or right colon.

Treatment of mesenteric cysts is by surgical resection. A segmental resection of the involved loop of intestine is often required because of the intimate proximity of the cysts to the bowel wall.

Further reading

Bentley JFR, Smith JR. Developmental posterior enteric remnants and spinal malformations: the split notochord syndrome. *Archives of Diseases of Childhood* 1960; **35**, 76–86.

Black PR, Welch KJ, Eraklis AJ. Juxtapancreatic intestinal duplications with pancreatic ductal communication: a cause of pancreatitis and recurrent abdominal pain in childhood. *Journal of Pediatric Surgery* 1986; **21**: 257–61.

Bower RJ, Sieber WK, Kiesewetter WB. Alimentary tract duplications in children. *Annals of Surgery* 1978; **188**: 669–74.

Bremer JL. Diverticula and duplications of the intestinal tract. *Archives of Pathology* 1944; **38**: 132–40.

Gross RE, Holcomb GW, Jr, Farber S. Duplications of the alimentary tract. *Pediatrics* 1952; **9**: 449–68.

Jewett TC, Jr, Walker AB, Cooney DR. A long-term follow-up on a duplication of the entire small intestine treated by gastroduplication. *Journal of Pediatric Surgery* 1983; **18**: 185–8.

Kurtz RJ, Heimann TM, Beck AR, Holt J. Mesenteric and retroperitoneal cysts. *Annals of Surgery* 1986; **203**: 109–12.

Ladd WE. Duplications of the alimentary tract. *Southern Medical Journa*, 1937; **30**: 363–71.

LaQuaglia MP, Feins N, Eraklis A, Hendren WH. Rectal duplications. *Journal of Pediatric Surgery* 1990; **25**: 980–4.

Leenders EL, Osman MZ, Sukarochana K. Treatment of duodenal duplication with international review. *Annals of Surgery* 1970; **36**: 368–71.

Lewis FT, Thyng FW. The regular occurrence of intestinal diverticula in embryos of the pig, rabbit, and man. *American Journal of Anatomy* 1907; **7**: 505–19.

Moss RL, Ryan JA, Kozarek RA, Hatch EI. Pancreatitis caused by a gastric duplication communicating with an aberrant pancreatic lobe. *Journal of Pediatric Surgery* 1996; **31**: 733–6.

Nobuhara KK, Gorski YC, LaQuaglia MP, Shamberger RC. Bronchogenic cysts and esophageal duplications: common origins and treatment. *Journal of Pediatric Surgery* 1997; **32:**, 1408–13.

Orr MM, Edwards AJ. Neoplastic change in duplications of the alimentary tract. *British Journal of Surgery* 1975; **62**: 269–74.

Pfeiffer VM, Lambrecht W, de Heer K. Mesenteric cysts. *Zentralblatt für Chirugie* 1981; **106**: 724–33.

Pruksapong C, Donovan RJ, Pinit A, Heldrich FJ. Gastric duplication. *Journal of Pediatric Surgery* 1979; **14**: 83–5.

Ravitch MM. Hind gut duplication—doubling of colon and genital urinary tracts. *Annals of Surgery* 1953; **137**: 588–601.

Schwesinger WH, Croom III, RD, Habibian MR. Diagnosis of an enteric duplication with pertechnetate ^{99m}Tc scanning. *Annals of Surgery* 1975; **181**: 428–30.

Siddiqui AM, Shamberger RC, Filler RM, Perez-Atayde AR, Lillehei CW. Enteric duplications of the pancreatic head: definitive management by local resection. *Journal of Pediatric Surgery* (in press).

Soper RT. Tubular duplication of the colon and distal ileum: case report and discussion. *Surgery* 1968; **63**: 998–1004.

Teele RL, Henschke CI, Tapper D. The radiographic and ultrasonographic evaluation of enteric duplication cysts. *Pediatric Radiology* 1980; **10**: 9–14.

Wrenn EL, Jr. Tubular duplication of the small intestine. *Surgery* 1962; **52**: 494–8.

Yousefzadeh DK, Bickers GH, Jackson JH, Jr, Benton C. Tubular colonic duplication—review of 1876–1981 literature. *Pediatric Radiology* 1983; **13**: 65–71.

42.6 Paedatric hepatobiliary disorders—surgical aspects

Eric L. Lazar and R. Peter Altman

Surgical disease of the biliary tract
The jaundiced neonate
Extrahepatic biliary atresia
Congenital biliary dilatation
Spontaneous bile duct perforation
Portal hypertension
Prehepatic portal hypertension
Surgical disease of the liver
Liver trauma
Hepatic abscess
Hepatic tumors
Further reading

Surgical disease of the biliary tract

Identification of the pediatric patient with biliary tract disease in whom obstructive pathology can be defined and corrected is crucial. Early intervention is imperative in order to minimize the impact of cholestasis on developing cirrhosis.

The jaundiced neonate

Jaundice is the hallmark of hepatobiliary disease. The evaluation of the jaundiced neonate begins by ascertaining the nature of the hyperbilirubinemia. Indirect hyperbilirubinemia in the neonate is usually physiologic and clears spontaneously with feeding or with the application of phototherapy. Rarely, plasmapheresis is required. Physiologic jaundice accounts for the vast majority of jaundice appearing after the first day or two of life. Other causes of unconjugated hyperbilirubinemia are non-obstructive and medically managed; breast milk jaundice, hemolysis, familial non-hemolytic icterus, and Gilbert's syndrome are all in the differential diagnosis of unconjugated hyperbilirubinemia. A discussion of these entities exceeds the scope of this chapter.

Persistence of jaundice, the appearance of jaundice after the immediate neonatal period, or jaundice in the setting of dark urine and clay-colored stools is pathologic and suggests an obstructive etiology. The presence of these signs or a significant elevation in the direct fraction of bilirubin mandates prompt evaluation. The cholestasis of direct hyperbilirubinemia may have a number of causes that are not correctable by surgery: maternally acquired intrauterine infection, metabolic disorders such as a_1-antitrypsin deficiency, familial disease syndromes such as Dubin-Johnson or Rotor's, and the cholestasis of total parenteral nutrition. Each of these must be differentiated from the important causes of jaundice with direct hyperbilirubinemia that are treatable by surgery: biliary atresia, choledochal cyst, spontaneous perforation of the bile ducts, and less commonly, biliary tumor and gallbladder disease.

Important biliary lesions may also present with less obvious manifestations than jaundice: failure to thrive and malnutrition, spontaneous hemorrhage secondary to elevated prothrombin time, and ascites may herald surgical biliary tract disease.

Extrahepatic biliary atresia

Atresia of the biliary tract is a progressive, sclerotic, obliterative process of uncertain and probably heterogeneous etiology. Wide-ranging hypotheses have been advanced for genetic, autoimmune, ischemic, and infectious causes, each of which have supporting but inconclusive evidence. For example, reovirus 3, type C rotavirus, and cytomegalovirus have each been suggested as an etiological agent in recent investigations but their role is now uncertain and the definitive cause of biliary atresia remains obscure. An autoimmune basis is gaining favor in light of recent reports that intracellular adhesion molecule 1 (**ICAM-1**) is found in association with the major histocompatibility complex class I antigens on affected ductal epithelium. Dillon and Krummel used immunoperoxidase staining to demonstrate the presence of the inflammatory mediator ICAM-1 on ductal epithelium from six patients with biliary atresia while this molecule was absent from five control specimens from patients with other biliary duct pathology. This is now a fertile area of investigation.

Biliary atresia is the most common surgical hepatobiliary disorder in the neonatal period, affecting about 1/14 000 neonates. It is also the most common index diagnosis leading to liver transplant in the pediatric population. Affected patients most often have isolated biliary atresia, although as many as 20 per cent of patients may have atresia as part of the polysplenia syndrome or other associated vascular anomalies such as a preduodenal portal vein.

Pathology

The obliterative cholangiopathy of biliary atresia can involve the common duct with patent proximal ducts; the common hepatic duct; the common hepatic, cystic, and common bile ducts; or the entire extrahepatic biliary tree including the ductules at the porta hepatis that give rise to the main left and right hepatic ducts. Luminal obliteration of the extrahepatic tree is associated with cannulicular bile stasis and plugging within the hepatic parenchyma. There is cholestasis within the hepatocyte as well. In addition, there is periportal edema and a periductal inflammatory cell infiltrate. Giant multinucleated hepatocytes, usually seen in hepatitis, often accompany the cellular infiltrate. Histologically, this stage may be indiscernible from other inflammatory conditions. Somewhat helpful is that the monocytic infiltrate of biliary atresia is concentrated in the portal areas, whereas a diffuse lymphocytic infiltrate is characteristic of neonatal hepatitis.

The cannulicular bile plugging results in ductular cell degeneration with compensatory biliary ductule proliferation and it is this proliferative response which distinguishes biliary atresia from the paucity syndromes. As the inflammation progresses, there is fibrosis of the portal tracts, which can extend from portal tract to portal tract. This bridging fibrosis is the beginning of cirrhosis. Unless early adequate surgical drainage is established, cirrhosis will progress.

Clinical features

The infant with biliary atresia is anicteric at birth, thrives, and may pass normal meconium for the first few weeks. Jaundice appears usually in the first month but may be subtle and intermittent, obscuring the diagnosis. There is the gradual onset of dark urine and then acholic or clay-colored stools. On physical examination, the liver is palpably firm but not hard early in the course. As cirrhosis worsens, the liver will feel hard and nodular. The surgeon should examine a fresh stool specimen to verify its acholic nature since pigmented stool makes the diagnosis of biliary atresia unlikely. The additional physical findings of a pulmonic murmur and characteristic facial features may suggest the alternative diagnosis of Alagille's syndrome.

The natural history of an untreated biliary atresia is one of nutritional failure and progressive liver failure with ascites, portal hypertension, and bleeding. Death is inevitable in the untreated patient.

Diagnosis

A newborn infant with the clinical features described must undergo an expeditious laboratory, imaging, and biopsy evaluation in order to effect prompt surgical drainage. No single test is definitive and most tests merely suggest the diagnosis. Therefore, each test must be interpreted within the constellation of the clinical features and other available test results. The investigational strategy is directed at differentiating biliary atresia from the other diagnoses that may masquerade as biliary atresia—neonatal hepatitis, paucity syndromes, and metabolic diseases. Hence, the laboratory investigation begins with exclusion of known microbial agents of hepatitis. Since the hyperbilirubinemia of biliary atresia is conjugated, a high total/low direct determination suggests an alternative diagnosis such as hemolysis. A serum determination of a_1-antitrypsin should be obtained to exclude deficiency of this protease inhibitor.

Abdominal sonography is performed in the fasting neonate and can be helpful in excluding the diagnosis by demonstrating intrahepatic duct dilatation or establishing choledochal cyst as the cause of the obstruction. The sonogram can be misleading, however, if certain pitfalls are encountered. An ‘absent' gallbladder may be the result of a post-prandial examination or ‘patent ducts' may be a perfect delineation of the hepatic arterial anatomy.

Hepatobiliary excretion scintiscanning is extremely helpful. This test utilizes one of a variety of technetium-labeled isopropyl diacetic acids. Pretreatment with

phenobarbital (5 mg/kg) for 3 to 5 days preoperatively enhances the discrimination between atresia and non-obstructive causes of intrahepatic cholestasis. Isotope detected in the duodenum excludes the diagnosis of biliary atresia, whereas failure of the radiopharmaceutical to appear in the small intestine merely keeps atresia in the differential diagnosis.

Percutaneous liver biopsy and the pathologic features typical of atresia have been described above. Evaluation by an experienced histopathologist will differentiate biliary atresia from a paucity syndrome. The obstructive phase of hepatitis may be more difficult to distinguish.

The final diagnostic maneuver may be laparotomy, at which time the extrahepatic biliary tree can be inspected. The experienced hepatobiliary surgeon will recognize the fibrotic extrahepatic biliary tree making an operative cholangiogram unnecessary in the majority of cases; intraoperative cholangiograms have been all but abandoned at our center. An intraoperative cholangiogram may be helpful in the case of a biopsy and nuclear scan consistent with obstruction, but an extrahepatic biliary system that does not appear fibrotic at exploration. A cholangiogram might demonstrate diminutive but patent biliary radicles (biliary hypoplasia) consistent with a paucity syndrome which does not warrant any intervention.

Treatment

Following the reports of Morio Kasai in 1959 and 1968, portoenterostomy has offered dramatic improvement in survival for the infant with biliary atresia. The authors favor a modified Kasai operation, which has the following features. Laparotomy via a generous transverse right upper abdominal incision commences with an examination of the spleen which alerts the surgeon to a polysplenia syndrome. The presence of polysplenia often indicates other anatomic abnormalities such as a preduodenal portal vein or malrotation, findings that must be considered during surgical reconstruction. The color and texture of the liver is noted and a liver biopsy obtained.

The goal of the dissection is to define the biliary plate where the fibrous cone of atretic ducts emerges from the substance of the liver at a level superior to the bifurcating portal vein. It is at this and only this precise location that the anastomosis will be fashioned. The porta is exposed and the gallbladder or its fibrous remnant is dissected from the liver bed. The common duct remnant is divided and the gallbladder, cystic duct, and proximal common duct are used to define the plane of the fibrous cone, anterior to the vessels of the hepatoduodenal ligament. These structures are usually readily identified and may be densely adherent to the hepatic and portal vessels. As the dissection proceeds, the portal vein and its left and right branches are defined as the posterior border of the field. Often there are tiny ‘middle' branches that enter the caudate lobe between the left and right branches of the portal vein—these are identified and divided to avoid troublesome bleeding. As the fibrotic biliary tree is dissected further into the porta, the cone-shaped confluence of the right and left hepatic ducts is defined to a level higher than the bifurcation of the portal vein. It is at the margin of the liver parenchyma that proximal transection of the fibrotic extrahepatic biliary tree is accomplished. The cone is incised from its lateral edges toward the middle, fashioning a diamond-shaped cut at the surface of the liver. The specimen is passed off the field. The exposed biliary plate now has an anterior and posterior margin that will serve as the anastomotic surface for the Roux limb.

Attention is then turned to fashioning the Roux limb, which is constructed as follows. A convenient loop of jejunum near the ligament of Treitz is selected and the bowel divided. The distal end is oversewn and the proximal end is anastomosed end-to-side 45 cm from this point. The long segment decreases reflux of enteric contents into the portal region. This maneuver keeps the incidence of subsequent cholangitis at a minimum, obviating the need for intussuscepted valves or externalized conduits, which are no longer employed. The Roux limb is then brought to the porta through the mesentery of the transverse mesocolon and incised along the antimesenteric border near the oversewn end. An end porta-to-side jejunal anastomosis is formed using a single layer of non-absorbable suture, placing first the posterior row and then the anterior row. Lack of a second layer in the anastomosis is essentially the only modification of the original operation described by Kasai. The jejunal mesentery is reapproximated with fine suture to prevent internal herniation. The mesenteric defect in the transverse mesocolon is also closed to prevent herniation. A penrose drain is split along its length for a few centimeters with one limb placed posterior and the other anterior to the anastomosis; the distal end is brought out through a separate stab wound in the right lower abdomen below the incision.

Postoperative complications include cholangitis, primary non-function, and bowel obstruction. We discharge patients on antibiotics to prevent cholangitis and with choleretic, chenodeoxycholine, and fat-soluble vitamin supplements. Steroids are not routinely used; however, they are occasionally prescribed if primary non-function is evidenced by continued acholic stools and elevated direct bilirubin after discharge from the surgical admission.

Results

Outcomes after portoenterostomy have been well studied. Age at operation, the gross anatomy of the bile ducts at operation, and the diameter of the ductules at the cut surface of the biliary plate independently predict failure of portoenterostomy. Children operated upon older than 70 days, those with atretic ducts at the porta, or with ductules at the cut surface measuring less than 150 microns in diameter are at greater risk for failure. Surprisingly, the presence of cirrhosis on liver biopsy does not portend failure as one might predict. In the large series by Altman and Lilly, the median survival of patients in whom bile flow could be established was 15 years; for those who required liver transplant, the mean age at that surgery was 5.4 years. The 10-year survival in the group requiring transplant was 71 per cent. Thus, even in ‘failure', the portoenterostomy provides good palliation. Primary portoenterostomy and rescue transplant are complimentary and we conclude that portoenterostomy should remain first-line therapy for atresia of the extrahepatic biliary tree.

Congenital biliary dilatation

Usually referred to as choledochal cyst, the term congenital biliary dilatation conveys two important aspects of this entity, not recognized by the older name. First, it is a congenital lesion, increasingly diagnosed by antenatal sonography. Second, the disease process involves more than the choledochus—the intrahepatic ducts are involved in Alonso–Lej types IV and V. The majority of cases, however, feature a fusiform dilatation of the common bile duct that may present with biliary obstruction, abdominal mass, or may remain occult.

Congenital biliary dilatation is rare but there is considerable international variation in its incidence, ranging from 1:1000 in Japan to 1:15 000 in the United States. The worldwide incidence is probably in the order of 1:100 000. The overwhelming majority of cases, perhaps 80 per cent, occur in females.

The widely accepted classification system for biliary duct dilatation differentiates five pathologic types (Fig. 1). Type I, the most com-mon, is a dilatation of the common bile duct, usually inferior to the confluence of the left and right hepatic ducts, and is subdivided into cystic and fusiform types depending upon the shape of the dilatation. Type II is the most rare and is a diverticulum of the common duct. Type III describes intraduodenal dilatation of the common duct and is uncommon. Type IV denotes intra- and extrahepatic involvement and type V, Caroli's disease, is intrahepatic biliary dilatation, which may involve the entire liver but often preferentially the left lobe. In all types, the cystic duct is usually normal and does not dilate despite communication with the cyst. The gallbladder is often normal.

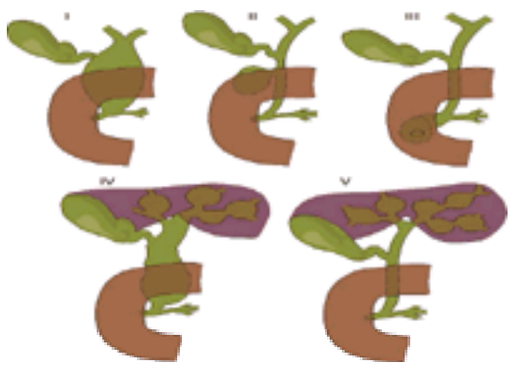


Fig. 1. Classification of congenital biliary dilatation. Note in all cases, the gallbladder and cystic duct are unaffected. Type I is fusiform dilatation and is the most commonly encountered type. Type II is diverticularization of the common duct. This is rather uncommon. Type III is a choledochocele in the intramural portion of the duodenum. Type IV represents both intra- and extrahepatic disease. Type V is intrahepatic disease, also known as Caroli's disease.

Etiology

The cause of biliary dilatation is unproved but the likely cause is related to the anatomy of the common channel demonstrated in the majority of patients with this entity. Seventy-six per cent of the 78 children studied at King's College Hospital had the long common channel anatomy, which is thought to permit pancreatic enzyme reflux

into the common bile duct. The proteolytic action of the enzymes is thought to weaken the duct wall which progressively dilates. The high amylase content of bile in patients with cystic biliary dilatation supports this concept. Since this anomaly is not present in all patients with the disease, this model cannot fully explain the etiology of congenital biliary dilatation. Lilly suggests that choledochal cyst may be merely one manifestation of a wider hepatobiliary disorder complex since a small number of patients present with multiple hepatobiliary anomalies.

Regardless of the etiology, the dilatation of the common duct allows for sludging of bile with a high amylase content causing mucosal ulcerations. There is a submucosal inflammatory infiltrate with cyst wall thickening and ultimately fibrosis. The cholestatic state is a risk factor for episodic cholangitis with subsequent liver injury. The spectrum of liver injury ranges from periportal infiltrates to biliary cirrhosis and consequent portal hypertension, which can be severe. Long-standing mucosal inflammation places these patients at risk of adenocarcinoma which occurs in 6 to 10 per cent of cases.

Clinical features

Congenital biliary dilatation has a bimodal distribution of age at presentation. The early clinical presentation is in infancy and may be not unlike that of biliary atresia—direct hyperbilirubinemia and progressive jaundice. Indeed, cystic dilatation is one of the primary diseases to be excluded in the differential diagnosis for biliary atresia. Historically, the classic triad of jaundice, pain, and an abdominal mass is described, but at most 15 per cent of children present this way. There is increasing presentation in the antenatal period with routine ultrasound and hence the diagnosis presents before symptoms. In the older child, intermittent pain, pancreatitis, and cholangitis are the features of presentation; on physical examination, an abdominal mass may be felt and features of portal hypertension might be noted. Uncommonly, bile peritonitis from a spontaneously ruptured cyst may be the presenting symptom.

Diagnosis

Abdominal sonography is the main diagnostic modality and is obtained in the evaluation of right upper quadrant pain or jaundice. The diagnosis is usually obvious from this test alone but CT scanning is sometimes used as an auxiliary imaging method to define the regional anatomy further. Scintiscanning can be helpful in that the diagnosis is confirmed by demonstrating an area of decreased ion intensity compared to the rest of the intrahepatic tree. The caveat in studying patients with suspected cystic biliary dilatation is in the setting of severe liver failure in which the labeled diacetic acid may not be cleared by impaired hepatocytes. Laparoscopy has been used but is unnecessary. Endoscopic retrograde cholangiopancreatography (**ERCP**) might demonstrate the common channel or presence of stones but is likewise not vital to making the diagnosis (Fig. 2). Increasing in popularity is magnetic resonance cholangiopancreatography (**MRCP**), which is non-invasive yet provides all of the information that ERCP does (Fig. 3). As noted for biliary atresia, the diagnosis is made preoperatively with a high degree of certainty and intraoperative diagnostics are usually not needed. Miyano and colleagues have suggested intraoperative cholangioscopy to exclude stones in the proximal hepatic ducts or the intraduodenal portion of the common duct.



Fig. 2. Endoscopic retrograde cholangiopancreatogram. This study demonstrates a fusiform dilatation of the distal common bile duct with some dilatation of the left main duct, seen near the spine. Note the gallbladder and cystic duct are relatively normal.



Fig. 3. Magnetic resonance image cholangiogram (lateral view). The bright globiform mass represents the cystic dilatation of the common duct with some dilatation of the left and right main ducts seen as accessory dilatations on the superior surface. The less dense form is the gallbladder.

Treatment

Cyst enterostomy has been abandoned in favor of complete cyst excision and biliary reconstruction, which is the treatment of choice for type I biliary cystic dilatation. Simple cyst enterostomy invites inevitable stone formation in the retained cyst cavity which does not obliterate. Furthermore, the malignant potential of the cyst lining is not addressed. In the few cases where pericystic inflammation or portal hypertension is so severe as to preclude safe excision of the entire cyst, the outer portion of the cyst wall may be left in apposition to the portal vein. In these cases, the cyst lining is removed in its entirety but an intramural plane is developed so that the outer layers of the cyst are intact in the region overlying the portal vein. Reconstruction is via a long Roux limb as noted for atresia of the extrahepatic biliary tree.

Type II cysts are very rare and so a standard treatment is not established. Primary reconstruction of the duct may be possible in these patients but the Roux limb is safe and effective. Type III cysts may require duodenotomy and sphincteroplasty or bypass by choledochoduodenostomy. Choledochoduodenostomy should probably accompany any partial cyst resection in order to insure complete biliary drainage.

The intrahepatic cysts in patients with type IV disease are not directly addressed by extrahepatic cyst excision and duct reconstruction, but this none the less ameliorates most of the symptoms. Lobe resection when feasible is the treatment of Caroli's disease, which is palliated by long-term antibiotics. Bilobar disease is not treatable except by liver transplant.

Results

The outcome after biliary reconstruction is excellent in most cases and depends upon the degree of established liver injury and cirrhosis. Cholangitis and pancreatitis are postoperative sequelae but are minimized by the long (45 cm) Roux reconstruction. As in atresia, nipple valve or intussuscepted antireflux constructions are meddlesome, prone to complication, and do not guarantee against the complications that they were designed to prevent.

Spontaneous bile duct perforation

As ‘spontaneous' implies, there is no known cause for this rare, unusual entity. The clinical presentation can be confusing. In some infants, it is similar to biliary atresia with jaundice, acholic stools, dark urine, and direct hyperbilirubinemia. There may be ascites and bile peritonitis has been reported. At laparotomy, the perforation is seen virtually exclusively at the junction of the cystic and common ducts. Chardot's recent report of 10 cases, however, identified at least four cases of perforation at other extrahepatic biliary tree sites. The treatment may vary depending upon the operative findings with primary repair of the perforated duct in selected cases.

However, for the majority, simple drainage of accumulated bile in the region of the common duct is advisable. This results in a controlled biliary fistula which is preferable to leak or stricture secondary to primary repair in an inflamed field. Primary, Roux-type reconstruction to drain any accumulated bile is contraindicated because any apparent cyst is not a true anatomic structure but rather an inflammatory shell.

Miscellaneous benign ductal obstruction

Cholelithiasis and acute cholecystitis occur in children, albeit less frequently than in adults. Predisposed patients have hemolytic anemia, cystic fibrosis, or have been on long-term parenteral nutrition. The presentation is not unlike that in the adult and the treatment is the same. Laparoscopic cholecystectomy is safe in young children and we have performed this procedure in children as young as 2 years. If common duct involvement is suspected preoperatively, ERCP with sphincterotomy and/or stenting is recommended.

Patients with cystic fibrosis merit special mention. Improved pulmonary care has resulted in a cohort of patients with cystic fibrosis living into adulthood and the spectrum of hepatobiliary disease seen in these patients is coming into focus. Inspissated bile, choledocholithiasis, and strictures may complicate portal hypertension and impaired synthetic function. The spectrum of operations to deal with ductal obstruction is similar to those utilized in the adult and the reader is referred to the relevant text. Aggressive operation may improve liver function and is indicated in the patient whose pulmonary disease is controlled.

Tumors of the extrahepatic biliary tree

Tumors of the extrahepatic biliary tree are extraordinarily rare with non-rhabdosarcomatous lesions occasioning only scattered case reports. Even among childhood rhabdomyosarcomas, biliary tract primaries represent less than 1 per cent. The prognosis is grim because complete resection is usually not possible at presentation. There is limited experience with neoadjuvant chemotherapy that promotes tumor shrinkage and anatomic resection. Hepaticoportoenterostomy is the reconstruction of choice for a proximal, small, well-localized tumor. Complete resection, particularly in the case of a distal lesion, requires a pancreaticoduodenectomy in order to achieve appropriate margins. Despite poor outcome, an aggressive approach is warranted because cure is not possible without complete surgical resection. Adjuvant chemotherapy usually consists of doxorubicin, actinomycin D, or vincristine based regimens.

Portal hypertension

Portal hypertension in children is characterized on the basis of intra-, post-, or prehepatic portal venous occlusion, as in the adult population. Intrahepatic disease is the final common pathway for many of the biliary disorders previously described in which biliary cirrhosis follows long-standing cholestasis. Parasitic infestation (schistosomiasis) in endemic areas is a common cause of post-sinusoidal intrahepatic obstruction. The presentation, diagnosis, and treatment of portal hypertension in this setting or for the Budd–Chiari syndrome (posthepatic) does not materially differ from adults and the reader is referred to the appropriate chapter.

Prehepatic portal hypertension

Of particular interest in children is prehepatic portal hypertension. The presumed etiology is portal vein thrombosis on the basis of antecedent neonatal omphalitis. Also, umbilical vein catheterization, a common event in the modern neonatal care unit, may be an important inciting event. In either case, there is thrombosis of the portal vein with resultant portal hypertension, splenomegaly, and occasionally ascites. The diagnosis is suspected at the onset of a herald bleed from esophageal varices. These bleeds are often self-limited and while potentially life-threatening, are seldom fatal. Aggressive resuscitation, the use of octreotide or pitressin, and the ready availability of sclerotherapy have further reduced the likelihood of a fatal bleed. Tamponade employing the Blakemore tube is seldom needed and must be used with caution because of the attendant risks to the airway. The essential distinction between these patients and a patient with cirrhosis and bleeding esophageal varices is that synthetic liver function is preserved and the feared triad of bleeding, encephalopathy, and coagulopathy is avoided. In time, cavernous transformation of the portal vein allows portal decompression resulting in fewer and less severe bleeding episodes with time. This pattern of gradual improvement even without shunting is in direct contradistinction to the patient with cirrhosis whose likelihood of rebleeding increases with time and in whom each bleeding episode has the potential for spiralling out of control and resulting in death from exacerbated liver failure, encephalopathy, and hepatorenal syndrome. Notwithstanding these differences between children and adults, both Maksoud and Bismuth have advocated decompression via portosystemic shunting even for pediatric prehepatic portal hypertension.

Shunt surgery is useful in treating portal hypertension in children though it is performed less frequently than in the past. This is probably due to the success of sclerotherapy and the transjugular intrahepatic portosystemic shunt that can be placed without the morbidity of laparotomy. This rigid endoprosthesis is placed via jugular puncture into a hepatic vein and forcibly passed into the liver parenchyma. This creates a traumatic fistula from the portal venous system to the systemic venous system. As long as the prosthesis remains patent, this is a reliable shunt and may temporize a patient long enough for the development of collaterals or for a liver transplant in the case of patients with intrahepatic-based portal hypertension. With these modalities, abdominal shunt surgery, though of unquestioned effectiveness, is often not required.

Non-shunt abdominal surgery for the treatment of the complications of portal hypertension has all but vanished in pediatric practice. The lengthy and morbid esophageal devascularization (Sugiura) is rarely indicated and the results are inferior to any of the shunt procedures.

Surgical disease of the liver

Blunt trauma, abscess, and tumor represent the spectrum of common conditions encountered in pediatric hepatic surgery. Metabolic abnormalities and endstage liver disease, formerly not amenable to surgical treatment, are now addressed by orthotopic liver transplantation. The availability of reduced and living-related donor transplants offers the hope that many more children will survive previously fatal diseases. The reader is referred to Chapter 16.6 for a review of these topics.

Liver trauma

The mechanism of injury in the child sustaining hepatic trauma is similar to the adult—vehicular accidents account for the vast majority of blunt injury followed by falls, sports-related accidents, and child abuse. Penetrating injury to the chest or abdomen can result in liver injury. Injury can take the form of small capsular tears, non-bleeding lacerations, major parenchymal disruption, and hepatic vein/retrohepatic caval injury. The treatment of these injuries depends upon the patient's hemodynamic status—the well-perfused patient with mild to moderate injury noted on CT scan may be safely observed; the hypotensive patient with major crush injury and bleeding noted at laparotomy may require abdominal packing. In any case, the range of diagnostic tests and therapeutic modalities are not unlike that of the adult patient.

Two particular forms of liver injury are unique to the pediatric patient—child abuse and birth injury. Child abuse may present insidiously with multiple visits to the emergency department for abdominal pain and vomiting. There may be bruising of the abdominal wall but often there is none. Radiographs can reveal multiple, aging fractures. Once the diagnosis of abuse is made, a CT scan will delineate the extent of the hepatic injury. The treatment is no different than for accidental blunt hepatic injury as previously outlined. The difficulty lies in suspecting the diagnosis since the history of abdominal trauma will usually not be offered.

Hepatic injury can accompany a difficult delivery and the mortality in this context is high. Risk factors for this unusual injury are post-maturity, hepatomegaly, breech presentation, and the use of significant obstetrical force. The diagnosis is entertained in the infant with any of the above features who is poorly perfused, has abdominal distension, and a falling hematocrit. The presence of free peritoneal blood may be suspected by scrotal discoloration or Cullen's sign about the umbilicus. At laparotomy, the injury may be difficult to identify and the newborn liver holds sutures poorly. Therefore, tamponade with abdominal packing and temporary closure is recommended. Reoperation follows in 24 to 48 h after the infant is stable, the blood volume is replete, and the coagulopathy is corrected.

Hepatic abscess

Pyogenic liver abscess is rare in the immunocompetent host in the modern antibiotic era. Formerly, pylephlebitis as a consequence of advanced appendicitis was commonly encountered. Currently, abscess formation may accompany umbilical vein catheterization or a course of necrotizing enterocolitis. In the immunocompromised host with chronic granulomatous disease, however, pyogenic abscess still represents a formidable challenge. The management of pyogenic abscess in children is similar to that in the adult, namely antibiotic coverage specific to the identified organism and drainage. Increasingly, percutaneous drainage is practised.

Amebic abscess is of concern in Asia, Africa, and parts of South America though it can occur in other parts of the world. Metronidazol alone may be sufficient in the patient with an uncomplicated abscess in an endemic area. Percutaneous aspiration is used to establish the etiology of an abscess in the patient in whom a pyogenic process is suspected; its routine use for diagnosis in endemic areas is not justified. Open drainage is reserved for the patient with multiple or complicated abscesses, or abscesses that do not respond to drug therapy.

Hepatic tumors

The most common liver tumor occurring in the neonate is metastatic neuroblastoma. These patients are either born with hepatomegaly and distension or develop these symptoms in the first months of life. Although the oncologic prognosis for patients with this stage-IV S variant is favorable, the large abdominal mass may adversely impact ventilation requiring prolonged mechanical support. Treatment is supportive awaiting the spontaneous regression characteristic of stage IV-S disease. Radiation therapy and daily low dose cyclophosphamide have been used to treat massive hepatomegaly.

Benign tumors

Benign tumors are rare and hemangioma and hemangioendothelioma comprise most of this group. These lesions may be large or multiple. Though cellularly benign, a large hepatic lesion can trap platelets causing thrombocytopenia and subsequent exsanguinating hemorrhage (Kassabach–Merritt syndrome). Larger lesions also may represent a significant shunt in the small patient and lead to congestive heart failure. For these reasons, and despite a natural history of regression, surgical resection has been used in the past and may be required on occasion (Fig. 4). More recently, in lesions requiring aggressive treatment, embolization has been employed with quite satisfactory results. The drug therapy of these lesions includes corticosteroids and interferon-a. Corticosteroids have produced a dramatic clinical response in some cases, though concerns remain about effects on growth and metabolism. Interferon-a is effective though it may have an unacceptable side-effect profile in infants, limiting its clinical usefulness. It is important to recognize that multiple strategies, including radiotherapy, may be required to treat the symptomatic lesion.

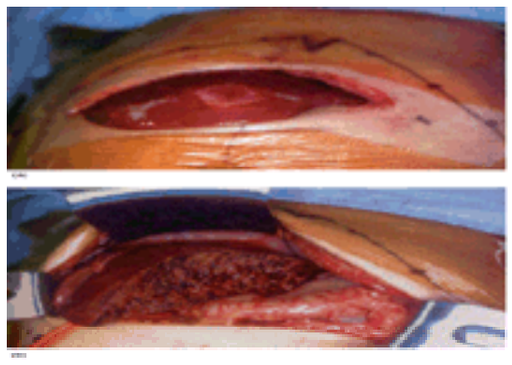


Fig. 4. (a) Hemangioma of left lobe. This infant had manifestations of congestive failure with a modest-sized hemangioma in the left lateral segment. (b) The left medial lobe after resection. The cut surface is inspected carefully for leaking biliary ducts and bleeding.

Hepatoblastoma

Hepatoblastoma is the third most common solid intra-abdominal tumor after neuroblastoma and nephroblastoma and it is the most common primary hepatic malignancy in children. None the less, it is a relatively rare tumor with an incidence of less than 1.0 per million. The characteristic presentation is in a child of less than 3 years with an asymptomatic, enlarging abdominal mass discovered by a caretaker or pediatrician. Weight loss may occur in about 25 per cent of cases but may be masked by the enlarging abdomen.

There is an imprecisely characterized association between hepatoblastoma and the Beckwith–Wiedemann syndrome, non-syndromic Wilms' tumor, and other embryonal tumors (rhabdomyosarcoma). This commonality suggests a shared genetic etiology and may involve loss of heterozygosity of a suppressor gene in the region of the short arm of chromosome 11. Environmental factors such as maternal exposure to certain paints, oils, pigments, and dyes may have an independent role.

The diagnostic evaluation proceeds with an abdominal sonogram and a determination of serum a-fetoprotein. The sonogram establishes the liver as the origin of the mass and can differentiate this primary liver lesion from a Wilms' tumor or neuroblastoma in the majority of cases. The integrity of the hepatic veins and vena cava can be assessed with Doppler-enhanced sonography.

Serum levels of a-fetoprotein should be interpreted cautiously in the infant less than 1 year of age since there is a physiologic elevation in the newborn that does not return to ‘normal' levels until about the eighth month. Furthermore, the half-life of a-fetoprotein is variable and age dependent in the first year of life making this a sensitive marker for recurrent disease only after the first year. After that time, any a-fetoprotein level over 500 ng/ml should be regarded as elevated. Though not specific for hepatoblastoma, elevation of serum levels of a-fetoprotein in the setting of a hepatic mass suggests the diagnosis.

A CT scan is often obtained for the further assessment of the resectability of the lesion and to provide a baseline examination prior to operation or neoadjuvant therapy. The initial study usually includes the chest to assess for pulmonary metastases. In general, further imaging is unnecessary although there is increasing experience with MRI, particularly MR-angiography to delineate the vascular anatomy. MRI may ultimately replace CT scanning as the initial examination.

Biopsy firmly establishes the diagnosis and there are a variety of strategies available. Percutaneous needle biopsies are helpful for the large, palpable mass. In the alternative, a limited right upper quadrant incision with open biopsy is used. Laparoscopic biopsies may also be obtained. Each of these ‘lesser' procedures has its advocates but as a group they are recommended when the imaging suggests an unresectable lesion (extensive bilobar disease or portal node involvement). Thus, a tissue diagnosis is secured and catheter access may be established prior to neoadjuvant therapy. This strategy is preferred to the scenario of undertaking an unsafe resection and acknowledges the variable application of the term ‘unresectable' to include variations of ‘formidable' (Fig. 5). The aforementioned notwithstanding, two-thirds to three-quarters of hepatoblastomas are resectable at initial staging and every attempt to do so should be made. It is clear that the best chance for survival is in the setting of a complete surgical resection.

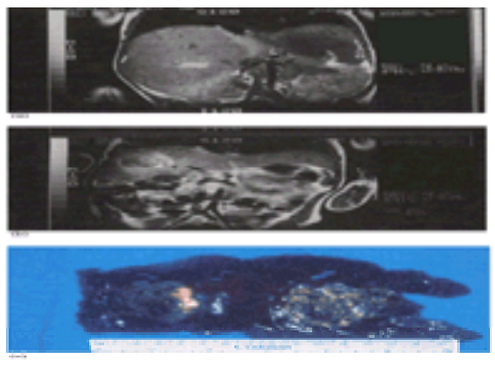


Fig. 5. Management of ‘unresectable' hepatoblastoma. This 3-year-old child has a large right lobe hepatoblastoma (5A). An extended resection might be possible but would certainly be formidable with a high morbidity. After 3 months, and eight cycles of chemotherapy, the tumor is much smaller on the follow-up MRI (5B). The resection specimen (5C) still represents a large hepatectomy, however, the tumor does not abut the margins.

The post-surgical therapy of hepatoblastoma is stage dependent. Stage I is a completely resected tumor and stage II tumors show microscopically positive margins. These stages are usually treated postoperatively with four courses of a cisplatin-based regimen (cisplatin/vincristine/5-flurouracil). If the stage I tumor demonstrates pure fetal histology without an alveolar component, then four courses of doxorubicin may be sufficient. Survival in these two stages should be better than 90 per cent. Neoadjuvant therapy is recommended for stage III (bilobar disease or portahepatis involved) and stage IV (metastatic) disease. Stage III is also reserved for an incomplete or aborted resection. Stages III and IV carry a 40 to 50 per cent mortality rate.

Hepatocellular carcinoma

This tumor is less common than hepatoblastoma and occurs in the cirrhotic and post-hepatitic liver. The staging and treatment are similar to that described above, but survival is far worse. Whereas the inclusion of platinum-based regimens has dramatically improved survival for hepatoblastoma, the relatively chemoresistant hepatocellular carcinoma has a 90 per cent mortality in stages III and IV.

Further reading

Altman RP. Long term results after the Kasai procedure. In: Daum F, ed., *Extrahepatic biliary atresia*, p. 96. Marcel Dekker, New York, 1983. [This monograph outlines the outcome in the early American experience with the Kasai portoenterostomy.]

Altman RP, Abramson SJ. Potential errors in the diagnosis and surgical management of neonatal jaundice. *Journal of Pediatric Surgery* 1985; **20**: 529–34. [A comprehensive evaluation of the various tests used to diagnose biliary atresia.]

Altman RP, Hicks BA. The jaundiced neonate. In: Altman RP, Stylianos S, eds, *Pediatric clinics of North America*, p. 1166. Saunders, Philadelpha, 1993. [A practical clinical approach to diagnosis in the jaundiced infant is presented.]

Altman RP, Lilly JR, Greenfeld J, Weinberg A, van Leeuwen K, Flanigan LM. A multivariable risk factor analysis of the portoenterostomy (Kasai) procedure for biliary atresia: twenty-five years experience from two centers. *Annals of Surgery* 1997; **226**: 348–53. [This recent work represents the largest American series.]

Bleacher JC, Newman KD. Hepatoblastoma. In: Andrassy R, ed., *Pediatric surgical oncology*, p. 214. WB Saunders, Philadelphia, 1998. [Contemporary look at hepatoblastoma in text of pediatric surgical oncology.]

Chardot C *et al.* Spontaneous perforation of the biliary tract in infancy: a series of 11 cases. *European Journal of Pediatric Surgery* 1996; **6**: 341–6. [The continental experience with this unusual entity.]

Davenport M, Howard ER. Macroscopic appearance at portoenterostomy—a prognostic variable in biliary atresia. *Journal of Pediatric Surgery* 1996; **31**: 1387–90. [This report emphasizes the prognostic value of the operative assessment of the porta.]

Davies MR. Iatrogenic hepatic rupture in the newborn and its management by pack tamponade. *Journal of Pediatric Surgery* 1997; **32**: 1414–19. [A discussion of this important modality in pediatric liver trauma.]

Dillon P, Belchis D, Tracy T, Cilley R, Krummel T. Increased expression of intracellular adhesion molecules in biliary atresia. *American Journal of Pathology* 1994; **145**: 263–7. [This work hypothesizes a molecular mechanism for the etiology of biliary atresia.]

Evans A, Chatten J, D'Angio GJ, Gerson JM, Robinson J, Schnaufer L. Review of 17 VI s neuroblastoma patients at the Children's Hospital of Philadelphia. *Cancer* 1980; **45**: 833–9. [This is a classic paper from a highly respected group.]

Fieber SS, Nance FC. Choledochal cyst and neoplasm: a comprehensive review of 106 cases and presentation of two original cases. *American Surgeon* 1997; **63**: 982–7. [Large review of malignant change in choledochal cyst.]

Fok TF, Chan MS, Metrweli C, Ng PC, Yeung CK, Li AK. Hepatic haemangioendothelioma presenting with early heart failure in a newborn: treatment with hepatic artery embolization and interferon. *Acta Paediatrica* 1996; **85**: 1373–5. [Case report emphasizing a multimodal approach to this lesion.]

Iyer CP, Stanley P, Mahour GH. Hepatic hemangiomas in infants and children: a review of 30 cases. *American Surgeon* 1996; **62**: 356–60. [Multiple treatment strategies are outlined in this large series of hepatic hemangiomata.]

Kasai M, Kimura S, Asakura Y, Suzuki H. Surgical treatment of biliary atresia. *Journal of Pediatric Surgery* 1968; **3**: 665–75. [The original American publication suggesting a novel treatment for biliary atresia.]

Lefkowitch JH. Biliary atresia. *Mayo Clinic Proceedings* 1998; **73**: 90–5. [Excellent pathologic review by a recognized hepatobiliary pathologist.]

Miyano T. Congenital biliary dilatation (choledochal cyst). In: Freeman NV, Burge DM, Griffiths DM, Malone PSJ, eds, *Surgery of the newborn*, p. 281. Churchill Livingstone, Edinburgh, 1994. [Current review of the basic principles of diagnosis and therapy of congenital biliary dilatation.]

Miyano T, Yamataka A, Kato Y, Kohno S, Fujiwara T. Choledochal cysts: special emphasis on the usefulness of intraoperative endoscopy. *Journal of Pediatric Surgery* 1995; **30**: 482–4. [Diagnostic strategy emphasizing ERCP in the evaluation of choledochal cyst.]

Nako Y *et al.* Successful interferon therapy in a neonate with life threatening Kasabach–Merritt syndrome. *Journal of Perinatology* 1997; **17**: 244–7. [One of the earliest reports of successful interferon therapy in the treatment of hepatic hemangioma.]

Samuel M, Spitz L. Choledochal cyst: varied clinical presentations and long term results of surgery. *European Journal of Pediatric Surgery* 1996; **6**: 78–81. [This report emphasizes the continental experience with varied presentations of choledochal cyst and the results of therapy.]

Sanz N, de Mingo L, Florez F, Rollan V. Rhabdomyosarcoma of the biliary tree. *Pediatric Surgery International* 1997; **12**(2–3): 40–3. [This uncommon malignancy is reviewed.]

Weintraub W, Lilly JR, Altman RP. Spontaneous perforation of extrahepatic biliary ducts and bile peritonitis in infancy. *Surgery* 1974; **75**: 664–73. [Report of this unusual problem.]

Yamataka A *et al.* Initial experience with non-breath holding magnetic resonance cholangiopancreatography: a new noninvasive technique for the diagnosis of choledochal cysts in children. *Journal of Pediatric Surgery* 1997; **32**: 1560–2. [A novel approach to imaging in choledochal cyst is reviewed.]

42.7.1 Hirschsprung's disease

Richard R. Ricketts

History
Etiology
Pathology and pathophysiology
Incidence
Clinical features
Diagnosis
Treatment
Outcome
Hirschsprung's disease in adolescents or adults
Other neuroendocrine disorders
Further reading

Congenital aganglionic megacolon, or Hirschsprung's disease, is the most common cause of intestinal obstruction in the neonate. Since its original description in 1886 and since the first curative operation was described in 1948, steady progress has been made in understanding the etiology and pathophysiology of the disease. The disease is now almost always diagnosed in early infancy leading to earlier treatment and improved outcomes, with a current mortality rate of 1 to 3 per cent.

History

Although first described by Ruysch in 1691, the Danish pediatrician Harald Hirschsprung has been credited with the disease which bears his name since his report to the Society of Paediatrics in Berlin in 1886. He described two infants with massive colonic distension who died with features of Hirschsprung's-associated enterocolitis. Hirschsprung postulated that the disease was congenital in origin and that it affected the entire colon. There followed numerous theories to explain the disease including that it resulted from a congenital abnormality of the dilated colon, a distal mechanical obstruction causing proximal dilatation, an infection, and a neurogenic imbalance. Swenson proposed a relationship between aganglionosis of the bowel and the failure of relaxation of the bowel, and called attention to the fact that the distal bowel, not the dilated proximal bowel, was the site of the defect. Swenson and Bill described their technique for resection of the rectum and rectosigmoid with preservation of the anal sphincter in 15 dogs and 3 children in 1948. Swenson and Neuhauser described the importance of the barium enema in making the diagnosis. Duhamel modified the original procedure to avoid the anterior dissection. Duhamel's procedure was subsequently modified by Martin to eliminate the problem of an obstructing fecaloma developing in the retained rectal segment. Soave described the endorectal pull-through without anastomosis in 1964. In the same year, Boley proposed the same procedure but with a primary coloanal anastomosis. Ravitch and Sabiston had already done this operation in dogs in 1947 and, unbeknownst to Soave and Boley, Yancey had already performed it in an adult with Hirschsprung's disease in 1952.

The diagnosis of Hirschsprung's disease depends upon the demonstration of aganglionosis in a rectal biopsy. Noblett introduced the suction biopsy forcep in 1968 making the rectal biopsy a safe bedside procedure. A histochemical staining technique for detection of acetylcholinesterase in the biopsy specimens was introduced by Meier-Ruge in 1972.

Martin introduced the ‘extended Duhamel' procedure for treatment of total colon Hirschsprung's disease in 1972. A long side-to-side anastomosis between the left colon and small bowel is performed after the bowel has been pulled-through utilizing the Duhamel technique. Since the right colon would be expected to absorb water better than the left, Kimura and subsequently Boley, designed pull-through procedures utilizing a right colon patch in 1981 and 1984, respectively. Currently, surgical advances including one-stage procedures performed in newborns and minimally invasive techniques continue to evolve.

Etiology

The absence of ganglion cells in the submucosal (Meissner) and myenteric (Auerbach) plexuses of the distal bowel is pathognomic for Hirschsprung's disease. Okamoto and Ueda postulated that this results from failure of migration of the cervical vagal neural crest cells from the esophagus to the anus during the 5th to 12th weeks of gestation. Current theories propose that the neuroblasts may have been present but failed to develop into mature functioning ganglia or that they were hindered in their migration or destroyed by elements in the microenvironment of the bowel wall. Factors which could adversely affect the migration, proliferation, differentiation, and colonization of these cells may have their basis in genetic, immunologic, vascular, or other mechanisms.

The increased risk to siblings and other family members in developing the disease (4 to 8 per cent) compared with the incidence in the general population (0.02 per cent), the increased frequency of Hirschsprung's disease in patients with chromosomal abnormalities such as trisomy 21, and the association of Hirschsprung's with other genetically determined diseases such as Waardenburg syndrome, all implicate a genetic basis for the disease. Badner studied the genetics of 487 patients and their families. She found that for patients with aganglionosis extending beyond the splenic flexure (‘long-segment' disease), the mode of inheritance was compatible with a dominant gene with incomplete penetrance. For shorter-segment disease, inheritance was equally likely to be either multifactorial or due to a recessive gene with very low penetrance, or as an X-linked recessive trait. The risk of siblings having the disease ranges from 1 to 5 per cent for short-segment disease up to 9 to 33 per cent for long-segment disease; the highest risk is for a male sibling of a female with long-segment disease (33 per cent). Approximately 4 to 8 per cent of patients have ‘familial' Hirschsprung's disease in which more than one family member is affected. In these families, long-segment disease is more frequent (61 compared with 27 per cent), total colonic Hirschsprung's is more common (39 compared with 5.6 per cent), and risk to subsequent family members is much higher when compared with non-familial Hirschsprung's disease.

Mutations in the RET proto-oncogene, located at chromosome 10q11.2, have been found in association with long-segment and familial Hirschsprung's disease. RET mutations may result in the loss of signals at the molecular level that are necessary for cell growth and differentiation of the enteric ganglia. Another susceptibility gene for Hirschsprung's disease is the endothelin-B receptor gene (**EDNRB**) located at chromosome 13q22. The signals from this gene are necessary for the development and maturation of the neural crest cells that innervate the colon. Mutations in this gene are most often found in non-familial and short-segment disease. The endothelian-3 gene has recently been proposed as another susceptibility gene. The effect of these genetic mutations is impaired or abrogated signaling necessary for normal development of the enteric nervous system. Mutations in the RET proto-oncogene are inherited in an autosomal dominant manner with 50 to 70 per cent penetrance and are found in about 50 per cent of familial cases and in only 15 to 20 per cent of sporadic cases. Mutations in the EDNRB gene are inherited in a pseudodominant mode and are seen in only 5 per cent of cases, usually sporadic ones.

Mutations in the RET proto-oncogene have also been found in over 90 per cent of families with the MEN-2 syndrome and in patients with familial medullary thyroid carcinoma. The association of Hirschsprung's disease with MEN-2a has been reported. Decker recommends that patients with MEN-2a be counseled regarding the potential risk of Hirschsprung's disease in offspring, and that genetic screening for MEN-2a be considered in children with familial Hirschsprung's disease in order to identify potential occult malignancies. The risk of MEN-2a for the patients with Hirschsprung's disease is at least 250 times that of the general population.

Abnormalities in the microenvironment of the bowel wall may prevent normal neural crest cell migration or differentiation. A marked elevation of class II major histocompatibility complex antigens has been shown to be present in the aganglionic bowel of patients with Hirschsprung's disease, but not to be present in the ganglionic bowel or in normal controls, suggesting an autoimmune mechanism for development of the disease.

Extracellular matrix proteins are important for cell adhesion and movement during early development. High levels of the matrix glycoproteins laminin and collagen type IV have been found in the aganglionic bowel. These alterations in the microenvironment of the bowel may prevent normal neural crest cell migration and have a role in the etiology of Hirschsprung's disease.

Pathology and pathophysiology

Hirschsprung's disease is characterized by aganglionosis of the submucosal and myenteric plexuses and hypertrophy of the nerve fibers in the aganglionic bowel wall. The neuropathophysiology of Hirschsprung's disease is complex and not fully understood. Normal bowel receives extrinsic innervation from the parasympathetic (cholinergic) and the sympathetic (adrenergic) nervous systems. The cholinergic fibers are excitatory to the colon (contraction) and inhibitory to the anal sphincter, whereas the adrenergic fibers are inhibitory to the colon (relaxation) and excitatory to the sphincter. Additionally, there is a vast intrinsic enteric nervous system within the bowel wall itself composed of a variety of ‘non-adrenergic non-cholinergic (**NANC**) inhibitory fibers' which function in the regulation of intestinal secretion, motility, mucosal defense, and immune response. The ganglion cells co-ordinate the muscular activity of the bowel by balancing the signals they receive from the adrenergic and cholinergic fibers, and from the intrinsic (enteric) NANC inhibitory fibers. In Hirschsprung's disease, the ganglion cells are absent so that the co-ordinated contraction and relaxation of the bowel cannot occur. Cholinergic excess may be responsible for the spasticity of the aganglionic segment. An excess of acetylcholine is released that stimulates the overproduction of acetylcholinesterase, which can be detected histochemically and used in making the diagnosis of Hirschsprung's

disease.

Probably more important than adrenergic or cholinergic abnormalities in causing spasm of the bowel is the absence of the NANC inhibitory fibers of the enteric nervous system and their neuropeptide transmitters. Vasoactive intestinal peptide (**VIP**) is the major relaxant of the internal anal sphincter; VIP-containing nerve fibers are absent in the aganglionic bowel of patients with Hirschsprung's disease. Nitric oxide (**NO**) is another potent neurotransmitter in the NANC inhibitory nerves, mediating relaxation of the bowel. NO synthase is normally present in the enteric plexus of the gut. NO synthase and thus NO activity is absent in the aganglionic bowel in patients with Hirschsprung's disease. A lack of NO-and VIP-containing nerve fibers in the aganglionic bowel in patients with Hirschsprung's disease may be the main factor in the pathophysiology of the disease.

Incidence

Hirschsprung's disease occurs in about 1 in 5000 live births. Approximately 7 per cent of patients are born prematurely. The overall sex ratio favors males 3.4:1. However, in short-segment disease the ratio is 4:1, while in long-segment disease the ratio drops to 1:1, and for total colon Hirschsprung's disease the ratio favors females 1.6:1. Aganglionosis extends to the rectosigmoid in 75 per cent of patients, up to the ascending colon in 17 per cent, and involves the total colon and terminal ileum in 8 per cent of patients. Total intestinal ganglionosis has rarely been reported. Isolated aganglionic areas between two segments of ganglionated bowel, are exceedingly rare. Up to 8 per cent of patients have ‘familial' Hirschsprung's disease. Ryan found that 20 per cent of infants with long-segment disease had affected relatives compared with only 3 per cent of those with short-segment disease.

Associated anomalies occur in about 20 per cent of patients, the most common being Down syndrome, occurring in between 8 and 16 per cent of patients. Patients with Down syndrome have an unsatisfactory outcome following treatment because of increased Hirschsprung's-associated enterocolitis (in about 50 per cent), poorer functional results with respect to fecal continence, and increased mortality (up to 38 per cent).

Other associated abnormalities include congenital heart defects in 7.8 per cent, genitourinary abnormalities in 5.6 per cent, and other gastrointestinal abnormalities in 3.9 per cent of patients. Another rare condition, termed ‘neurocristopathy', is the association of Hirschsprung's disease (usually total colon involvement) with central hypoventilation syndrome (Ondine's curse) and neuroblastoma.

Clinical features

Patients may present in the neonatal period with a complete intestinal obstruction or in later childhood or as adults with chronic constipation. Currently, 80 to 90 per cent of all patients are diagnosed in the neonatal period. The first clue to the diagnosis is the time of passage of the first meconium stool. Whereas 94 per cent of normal full-term infants pass their first stool before 24 h of age, 94 per cent of patients with Hirschsprung's disease do not. Neonates present with obstipation, abdominal distension, and bilious emesis within the first few days of life. A rectal examination may alleviate the obstruction temporarily. Breast milk feeds may delay the onset of symptoms because of the soft stool produced; symptoms will emerge when formula feeds are instituted.

Approximately 6 to 14 per cent of infants will present with Hirschsprung's-associated enterocolitis. Clinical findings suggesting this diagnosis include abdominal distension, diarrhea, fever, hematochezia, emesis, and signs of sepsis. Explosive diarrhea follows a rectal examination. Prompt recognition and treatment of this condition is imperative since Hirschsprung's-associated enterocolitis is the leading cause of death in this disease, with mortality rates as high as 30 per cent. Treatment consists of decompression with rectal irrigations, volume resuscitation, and administration of broad-spectrum antibiotics if signs of sepsis are present. In some cases an emergency colostomy may be required.

In Swenson's series of 501 patients treated over a 26-year period, 3.4 per cent of patients, all neonates, presented with perforations. Analysis of these plus patients reported subsequently reveals that 90 per cent of the perforations were in the proximal colon, cecum, or ileum and that total colon Hirschsprung's disease was present in 47 per cent.

Diagnosis

The diagnosis of Hirschsprung's disease can be suggested by a barium enema or anorectal manometry, but the definitive diagnosis requires a rectal biopsy. The most useful initial study in neonates is a barium enema of the unprepared rectum and colon. Typically a narrow or normal-sized rectum leading to a dilated proximal colon will be seen (Fig. 1). In an unaffected individual, the diameter of the rectum is greater than that of the sigmoid colon; in Hirschsprung's disease, this ratio is reversed. The barium enema is highly diagnostic in infants, toddlers, and older children but is only 80 per cent accurate in neonates. Retention of barium in the colon on a 24-h delayed film supports the diagnosis. Infants with total colon Hirschsprung's disease may demonstrate a normal-appearing colon or a microcolon by barium enema.

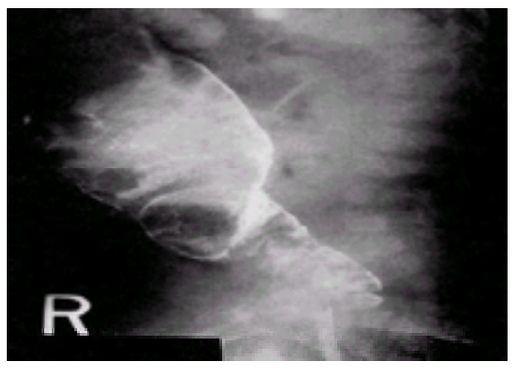


Fig. 1. Barium enema demonstrating narrow rectum and dilated rectosigmoid colon in a patient with typical short-segment Hirschsprung's disease.

The barium enema can also exclude other conditions which present with intestinal obstruction in the newborn. Intestinal atresias and meconium ileus are characterized by a microcolon. Small left colon syndrome, associated with maternal diabetes, and meconium plug syndrome are characterized by an obstructing bolus of meconium which is usually evacuated after the enema. Persistence of obstructive symptoms after the barium enema in these conditions requires that a rectal biopsy be done to exclude Hirschsprung's disease.

Anorectal manometry is also useful in the diagnosis of Hirschsprung's disease. In normal individuals, distension of the rectum causes reflex relaxation of the internal anal sphincter; in Hirschsprung's disease, relaxation does not occur. This finding is highly diagnostic for Hirschsprung's disease in toddlers and older children (90 per cent), but its diagnostic accuracy in neonates is questioned. However, if relaxation of the internal sphincter is demonstrated, the diagnosis of Hirschsprung's disease is excluded.

A suction or full-thickness rectal biopsy demonstrating absence of ganglion cells in the submucosal and/or myenteric plexus of the rectal wall establishes the diagnosis of Hirschsprung's disease (Fig. 2). Hypertrophied nerve fibers in the bowel wall are also present, although in long-segment disease, the myenteric plexus nerves will be hypoplastic or absent. The biopsy must be taken at least 2 cm proximal to the pectinate line, since a zone of hypoganglionosis and aganglionosis normally exists distal to this point, especially in older patients. The presence of acetylcholinesterase activity in the biopsy can aid in establishing the diagnosis. This test, however, does not mitigate against the requirement for a histologic diagnosis, since in both Hirschsprung's disease and neuronal intestinal dysplasia (see later) acetylcholinesterase activity will be increased; the former will not have ganglion cells whereas the later will have hyperganglionosis. Using acetylcholinesterase staining in conjunction with standard histologic evaluation yields a diagnostic accuracy rate of 99 per cent.

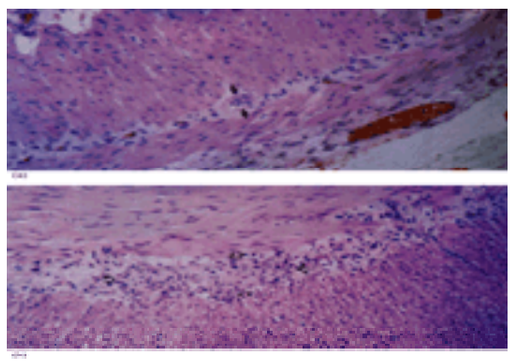


Fig. 2. (a) Normal rectal biopsy demonstrating ganglion cells in the myenteric plexus (closed arrows). Hematoxylin and eosin, 250 ×. (b) Rectal biopsy from a patient with Hirschsprung's disease demonstrating hypertrophied nerve trunks (open arrows) and no ganglion cells in the myenteric plexus. Hematoxylin and eosin, 250 ×.

Treatment

A treatment algorithm for standard ‘short-segment' Hirschsprung's disease based upon the time of diagnosis is shown in Fig. 3. The initial goal of treatment in neonates is decompression of the colon, either by a colostomy situated in the ganglionated bowel proximal to the transition zone (‘leveling' colostomy), or by a program of rectal irrigations. Rectal irrigations consist of administering saline high into the colon at a dose of 20 ml/kg two or three times per day. While either option is satisfactory, a colostomy is more appropriate if the social situation is such that the caretakers cannot reliably perform rectal irrigations or cannot return regularly for follow-up. Additionally, neonates who present with enterocolitis are more appropriately managed with an initial colostomy.

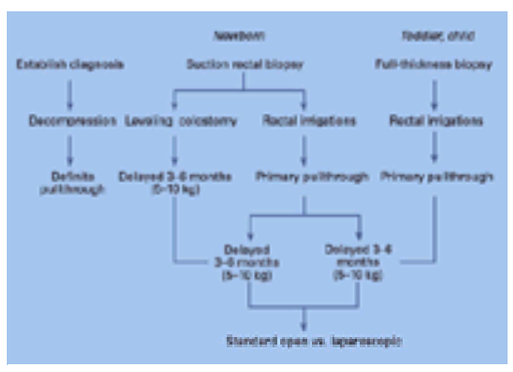


Fig. 3. Treatment algorithm for standard short-segment Hirschsprung's disease based on the time of diagnosis.

If a leveling colostomy is performed, the definitive pull-through procedure is delayed for 3 to 6 months or until the child weighs 5 to10 kg. If rectal irrigations are instituted, the definitive procedure can be delayed or performed in the newborn period. The outcome, with respect to morbidity, mortality, and final functional status is equivalent for all three treatment regimens.

Toddlers and children are initially managed with rectal irrigations followed by a definitive primary pull-through when the colon has been adequately decompressed. The success of a primary pull-through largely depends upon the experience of the pathologist, who must determine the site of the normal ganglionated bowel based on frozen section examination of seromuscular biopsies.

The basic principle of all operations used to treat Hirschsprung's disease is to bring normally ganglionated bowel down to the anus. The three most commonly used procedures are the rectosigmoidectomy with abdominal perineal pull-through of Swenson and Bill, the retrorectal transanal pull-through of Duhamel, and the endorectal pull-through credited to Soave. A schematic depiction of each operation is shown in Fig. 4; the technical details are beyond the scope of this chapter. These procedures are now being done laparoscopically, even in neonates. The laparoscopic procedure described by Georgeson is a transabdominal extramuscular dissection (Swenson) from above and a transanal mucosectomy (Soave) from below.

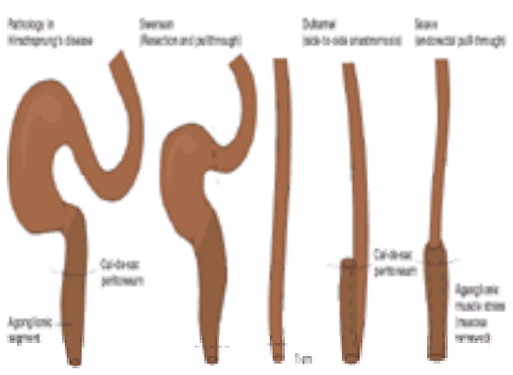


Fig. 4. Schematic representation of the three most commonly used operations to treat Hirschsprung's disease.

Infants with total colon Hirschsprung's disease should be managed initially with a leveling ileostomy. Severe fluid and electrolyte losses and nutritional disturbances can occur resulting in mortality rates of 20 to 65 per cent. Application of a right colonic patch to the normally ganglionated small bowel as described by Kimura and Boley can ameliorate these fluid and electrolyte losses and improve functional results and survival.

Aganglionosis extending into the proximal jejunum or even more proximal has rarely been described and was uniformly fatal until 1987 when Ziegler described his technique of extended myectomy–myotomy of the aganglionic bowel. This procedure takes advantage of the fact that the aganglionic bowel maintains normal absorptive function despite having abnormal motility.

Ultrashort-segment Hirschsprung's disease, which is more appropriately termed anorectal achalasia, is characterized by the presence of ganglion cells on rectal biopsy, but lack of a transition zone on barium enema and lack of reflex internal sphincter relaxation to transient rectal distension on anorectal manometry. These patients are effectively treated with an internal anal sphincter myomectomy.

Outcome

Table 1 presents the published results from large series of patients operated on for Hirschsprung's disease utilizing the three most common operations. These are not prospective studies comparing the operations, the follow-up is variable, and the data span a time period beginning in 1948 until the present. Even with these limitations, it is evident that the procedures are comparable with respect to morbidity, mortality, and long-term functional outcome. In all procedures, the functional results improve with time, so that by 10 years' follow-up about 90 per cent of patients will have ‘normal' bowel function.

Reference date	No. of patients	Anastomotic leak	Rectal obstruction	Constipation	Soiling	Postoperative enterocolitis	Mortality	Normal bowel function	Impotence	Mortality
Swenson procedure										
1. Glenn/Ament (1976)	300	11.7	9.5	N.A.	3.2	15.6		N.A.	N.A.	2.5
2. Boley (1984)	89	7.9	8.9	N.A.	9.0	32.7		N.A.	N.A.	2.2
3. Swenson (1989)	800	1.7	7.6	N.A.	12.9	24.0		9.1% (@ 10 years)	0	2.4
Duhamel procedure										
1. Glenn/Ament (1976)	209	2.4	5.3	N.A.	1.1	5.9		N.A.	N.A.	1.0
2. Boley (1984)	73.5	6.8	3.0	N.A.	2.5	13.9		N.A.	N.A.	1.5
3. Ortenzio (1992)	0.0	N.A.	N.A.	0	0	12.5		9.1% (@ 10 years)	N.A.	6.2
4. Swenker (1997)	24.0	2.2	0.7	0	5.3	5.3		97.9	N.A.	1.6
Soave procedure										
1. Glenn/Ament (1976)	282	4.2	11.3	N.A.	1.4	9.9		N.A.	N.A.	1.8
2. Boley (1984)	208	7.3	5.1	N.A.	2.5	12.1		N.A.	N.A.	2.1
3. Soave (1985)	27.1	–	0.7	2	N.A.	5		88	0	6.6

N.A. = not available

Table 1 Compiled data on morbidity, mortality, and functional results after surgery for Hirschsprung's disease (figures are percentages)

The unique and most deadly complication associated with Hirschsprung's disease, both preoperatively and postoperatively, is enterocolitis. It occurs in 20 to 58 per cent of patients and is the presenting symptom in up to 14 per cent of patients. Factors which increase the risk of developing enterocolitis include delay in diagnosis, long-segment disease, a positive family history for Hirschsprung's disease, and presence of associated anomalies, particularly Down syndrome. Enterocolitis occurs in approximately 50 per cent of patients with Down syndrome and may be related to intrinsic immunologic deficiencies present in these patients.

A survey of the members of the Surgical Section of the American Academy of Pediatrics, published in 1979, reported a 30 per cent mortality rate associated with enterocolitis. Currently the mortality rate is less than 1 to 3 per cent, yet it remains the leading cause of death in patients with Hirschsprung's disease, accounting for up to 50 per cent of all deaths.

The etiology of Hirschsprung's-associated enterocolitis is not fully understood. It may be caused by a functional obstruction at the level of the internal sphincter. This mechanical theory is supported by clinical findings since the mainstay of treating enterocolitis is rectal decompression and irrigation. Furthermore, patients with repeated episodes of enterocolitis can be successfully treated with forceful dilatation of the anal sphincter, sphincterotomy or myomectomy, or intrasphincteric botulinum toxin injection.

Impaired gut immunity and IgA transport, and complex infective mechanisms have been implicated as causes for enterocolitis. In Hirschsprung's disease, the protective mucus barrier on the luminal surface is deficient and there is reduced turnover of mucin glycoproteins. The bowel wall is thus susceptible to infection by enterocyte-adherent organisms (e.g. *Escherichia coli* and *Clostridium difficile*) which could release toxins resulting in the systemic and local symptoms of enterocolitis. Furthermore, these defects have been found in ganglionated bowel of patients with Hirschsprung's disease, perhaps explaining why enterocolitis can occur after the abnormal aganglionic bowel has been resected.

There are no significant deficiencies of immunoglobulin-containing cells or T and B lymphocytes in the rectal mucosa of patients with Hirschsprung's disease when compared with normal individuals. However, impaired transport of secretory IgA into the lumen of the bowel and diminished levels of serum and salivary IgA have been demonstrated in patients developing enterocolitis.

There are no significant differences in the microbiological flora of the colon in patients with uncomplicated Hirschsprung's disease or in those with enterocolitis compared with controls. Viral studies, however, have isolated rotavirus from the stools of 78 per cent of patients with enterocolitis suggesting an etiologic role in enterocolitis.

Hirschsprung's-associated enterocolitis should be suspected in any postoperative patient (colostomy or pull-through) who presents with abdominal distension and explosive watery diarrhea. Prompt recognition and treatment of this disorder with intravenous fluid resuscitation, rectal irrigations, and broad-spectrum antibiotics (including vancomycin or metronidazole) if fever is present, will almost always result in a successful outcome.

Other causes for persistent or recurrent postoperative constipation and distension include resection of an inadequate amount of aganglionic colon, anal sphincter achalasia, neuronal intestinal dysplasia, or acquired postoperative aganglionosis. These disorders can be differentiated by repeat full-thickness rectal biopsy and manometry. The specimen should be submitted for histologic evaluation for the presence or absence of ganglion cells and for acetylcholinesterase staining. Treatment may be accomplished medically or may require an internal sphincter myomectomy or a repeat pull-through procedure.

Hirschsprung's disease in adolescents or adults

Nearly all patients with Hirschsprung's disease are diagnosed in the neonatal period with only 5 per cent of patients having a delayed diagnosis beyond 5 years of age. The clinical features and diagnostic procedures which help to differentiate a patient with idiopathic chronic constipation from one with Hirschsprung's disease are shown in Table 2. Normal relaxation of the internal sphincter with transient rectal distension during anorectal manometry rules out the diagnosis of Hirschsprung's disease in older patients. However, a full-thickness rectal biopsy taken at least 2 cm above the dentate line is still required to definitively establish or exclude the diagnosis.

	Hirschsprung's disease	*Chronic constipation*
Onset of symptoms	Neonatal period	'Potty' training period
Constipation	Present	Present
Distension	Present	Absent
Soiling	Absent	Present
Rectum	Empty	Loaded with stool
Rectal exam	Explosive stool	Impaction
Barium enema	Transition zone; sigmoid diameter > rectal diameter	No transition zone
Manometry	Internal sphincter failure of relaxation	Normal relaxation to rectal distention
Rectal biopsy	No ganglion cells in Meissner's/Auerbach's plexuses/thi	Ganglion cells present

Table 2 Hirschsprung's disease compared with chronic constipation

The management of an older patient with Hirschsprung's disease is depicted in Fig. 5. If the colon is massively dilated beyond the splenic flexure, irrigations are unlikely to be successful in cleaning it out and reducing its diameter sufficiently to allow for a pull-through procedure to be done. In this case, an ileostomy rather than a ‘leveling' colostomy should be done since the latter are huge and tend to retract or prolapse. If the disease appears to involve only the rectosigmoid and the proximal colon is of reasonable size to allow for a pull-through, irrigations and laxatives can usually achieve adequate decompression so that a primary procedure can be done. Any one of the major operations for Hirschsprung's disease can be done in adults with excellent results.

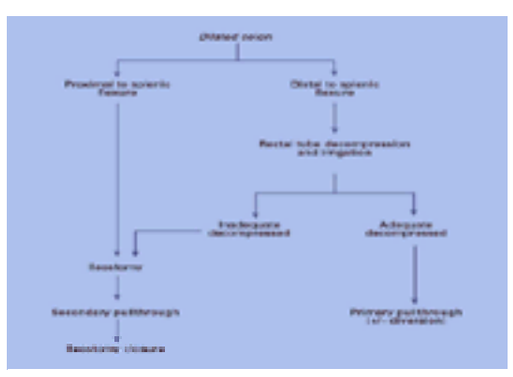


Fig. 5. Algorithm for the management of adolescents or adults with Hirschsprung's disease.

Other neuroendocrine disorders

Several conditions resemble Hirschsprung's disease clinically but have ganglion cells on the rectal biopsy. The most common of these are neuronal intestinal dysplasia, internal sphincter achalasia, smooth muscle disorders, hypoganglionosis, and immature ganglia.

There are two types of neuronal intestinal dysplasia. Type A, accounting for only 5 per cent of the cases, is characterized histologically by hypoplasia or aplasia of the sympathetic innervation of the bowel, inflammation of the colonic mucosa, and hyperplasia of the myenteric plexus. It presents in the neonatal period with intestinal obstruction, diarrhea, and colitis. Type B is characterized histologically by hyperplasia of the submucosal nerve plexus, giant ganglia containing more than seven nerve cells, buds of nerve cells on parasympathetic fibers, and increased acetylcholinesterase staining in nerve fibers in the adventitia of submucous arteries. An abnormal or absent anorectal reflex may be demonstrated manometrically. While the etiology is not fully understood, an elevation of class II major histocompatibility antigens has been demonstrated in neuronal intestinal dysplasia and in Hirschsprung's disease, perhaps pointing to a common autoimmune mechanism. Type B neuronal intestinal dysplasia has a clinical picture which is indistinguishable from Hirschsprung's disease and is coincidentally present in 20 to 75 per cent of patients with Hirschsprung's disease, perhaps accounting for the postoperative constipation seen following pull-through procedures. Patients with neuronal intestinal dysplasia are managed medically with enemas, laxatives, and prokinetic agents. There is often spontaneous improvement suggesting neuronal maturation. If symptoms persist beyond 6 months, especially in patients having Hirschsprung's disease, internal sphincter myomectomy should be considered.

Patients with internal anal sphincter achalasia (ultrashort-segment Hirschsprung's disease) present clinically like patients with Hirschsprung's disease, but they have ganglion cells on rectal biopsy and absence of the rectosphincteric reflex. They are treated successfully with internal anal sphincter myomectomy.

Smooth muscle disorders causing functional bowel obstruction, such as the megacystis–microcolon–intestinal hypoperistalsis syndrome, can only be diagnosed using electron microscopy. They are characterized by absent peristalsis throughout the gastrointestinal tract and are usually fatal.

Hypoganglionosis is very rare and is characterized histologically by sparse and small myenteric ganglia, absence of or low acetylcholinesterase activity in the lamina propria, and hypertrophy of the muscularis mucosa and circular muscle. Treatment is by resection of the affected bowel and a pull-through procedure.

Finally, immature ganglia are found in biopsy specimens from premature infants presenting with functional obstruction. These are small ganglia with small ganglion cells. The treatment is conservative.

Further reading

Bourdelat D, Vrsansky P, Pages R, Duhamel B. Duhamel operation 40 years after: a multicentric study. *European Journal of Pediatric Surgery* 1997; **7**: 70–6. [This is an international retrospective study of 2430 patients treated for Hirschsprung's disease with the Duhamel technique.]

Hirschsprung H. Stuhltragheit neugeborener in Folge von dilatation und hypertrophie des Colons. *Jahrbuch für Kinderheilkonde* 1887; **27**: 1–7. [This is the original description of congenital aganglionic megacolon by Professor Harold Hirschsprung.]

Ikeda K, Goto S. Diagnosis and treatment of Hirschsprung's disease in Japan. An analysis of 1628 patients. *Annals of Surgery* 1984; **199**: 400–5. [This is a report of treating 1628 patients of Hirschsprung's disease in Japan between 1978–1982.]

Kimura K, Nishijima E, Muraji T, Tsugawa C, Matsutmo Y. Extensive aganglionosis: further experience with the colonic patch graft procedure and long-term results. *Journal of Pediatric Surgery* 1988; **23**: 52–6. [This is a report of the results in patients with total colon Hirschsprung's disease treated with the right colonic patch technique described by Dr Kimura.]

Kleinhaus S, Boley SJ, Sheran M, Sieber WK. Hirschsprung's disease. A survey of the members of the Surgical Section of the American Academy of Pediatrics. *Journal of Pediatric Surgery* 1979; **14**: 588–97. [This report details the management of 1196 patients with Hirschsprung's disease treated in the United States between 1975 and 1976.]

Langer JC, Birnbaum E. Preliminary experience with intrasphincteric botulinum toxin for persistent constipation after pull-through for Hirschsprung's disease. *Journal of Pediatric Surgery* 1997; **32**: 1059–62. [This describes the use of botox injection for patients with persistent constipation following pull-through procedures for Hirschsprung's disease.]

Meier-Ruge W, Lutterbeck PM, Herzog B, Morger R, Moser R, Scharli A. Acetylcholinesterase activity in suction rectal biopsies of the rectum in the diagnosis of Hirschsprung's disease. *Journal of Pediatric Surgery* 1972; **7**: 11–17. [The utility of the acetylcholinesterase stain in diagnosing Hirschsprung's disease is described.]

Puri P. Variant Hirschsprung's disease. *Journal of Pediatric Surgery* 1997; **32**: 149–57. [This is an excellent discussion of neuronal intestinal dysplasia and other variants of Hirschsprung's disease.]

Rescorla FJ, Morrison AM, Engles D, West KW, Grosfeld JL. Hirschsprung's disease. Evaluation of mortality and long-term function in 260 cases. *Archives of Surgery* 1992; **127**: 934–42. [This report outlines the long-term functional outcome in 260 patients treated for Hirschsprung's disease at a single center and also provides an excellent review of Hirschsprung's disease in the ‘Comment' section of the paper.]

Ryan DP. Neuronal intestinal dysplasia. *Seminars in Pediatric Surgery* 1995; **4**: 22–5. [This is an excellent discussion of neuronal intestinal dysplasia.]

Sherman JO *et al.* A 40-year multinational retrospective study of 880 Swenson procedures. *Journal of Pediatric Surgery* 1989; **24**: 833–8. [This reviews the experience in treating Hirschsprung's disease in North America and Western Europe utilizing the Swenson procedure in 880 patients.]

Skinner MA. Hirschsprung's disease. *Current Problems in Surgery* 1996; **33**: 389–468. [This represents the most thorough monograph on Hirschsprung's disease ever read by this author.]

Soave F. Endorectal pull-through: 20 years experience. Address of the guest speaker, APSA, 1984. *Journal of Pediatric Surgery* 1985; **20**: 568–79. [This is a nice description of the Soave endorectal pull-through technique and the results achieved by Dr Soave.]

Swenson O, Rheinlander HF, Diamond I. Hirschsprung's disease: a new concept of the etiology. Operative results in thirty-four patients. *New England Journal of Medicine* 1949; **241**: 551–6. [Dr Swenson proposed a relationship between aganglionosis and the failure of relaxation of the bowel in patients with Hirschsprung's disease and reported his results utilizing the first curative operation for Hirschsprung's disease which Dr Swenson described.]

Teitelbaum DH, Coran AG, Weitzman JJ, Ziegler MM, Kane T. Hirschsprung's disease and related neuromuscular disorders of the intestine. In: O'Neill, Jr. JA, Rowe MI, Grosfeld JL, Fonkalsrud EW, Coran AG, eds, *Pediatric Surgery*, 5th edn, pp. 1381–424. Mosby, St. Louis, 1998. [An excellent up to date discussion of Hirschsprung's disease including descriptions of the major surgical procedures to treat the disorder.]

Wilcox DT, Bruce J, Bowen J, Bianchi A. One-stage neonatal pull-through to treat Hirschsprung's disease. *Journal of Pediatric Surgery* 1997; **32**: 243–7. [This report shows the feasibility of performing a pull-through procedure in the neonatal period to treat Hirschsprung's disease.]

Wulkan ML, Georgeson KE. Primary laparoscopic endorectal pull-through for Hirschsprung's disease in infants and children. *Seminars in Laparoscopic Surgery* 1998; **5**: 9–13. [This illustrates a technique for laparoscopic pull-through procedure for Hirschpsrung's disease in the neonatal period.]

Ziegler MM, Royal RE, Brandt J, Drasnin J, Martin LW. Extended myectomy–myotomy. A therapeutic alternative for total intestinal aganglionosis. *Annals of Surgery* 1993; **218**: 504–11. [This report describes a technique of treating extensive aganglionosis, that which extends up into a majority of the small bowel.]

42.7.2 Other anorectal problems

Richard R. Ricketts

Perianal abscess and fistula-in-ano
Anal fissures
Hemorrhoids
Juvenile polyps
Rectal prolapse
Condyloma acuminatum
Anterior ectopic anus
Further reading

There is a wide spectrum of minor anorectal disorders in children. Although a great source of anxiety to parents, these disorders, excluding anorectal malformations, Hirschsprung's disease, and inflammatory bowel disease, can usually be managed non-operatively or with relatively simple surgical procedures.

Perianal abscess and fistula-in-ano

Perianal abscesses occur more frequently in males and they are associated with a fistula-in-ano in about one-third to one-half of infants. Superficial abscesses are treated with warm sitz baths until the area is fluctuant; incision and drainage is then usually curative. If the abscess recurs in the same location, a fistula-in-ano is present. Deep perirectal and ischiorectal abscesses do not occur in children, except in those with immune deficiencies or inflammatory bowel disease.

Fistulas-in-ano in infants are different from their adult counterparts. Over 90 per cent of the cases occur in males; most occur in infants less than 12 months old; fistulas are direct and do not follow Goodsall's rule; multiple but separate tracts are common (about 20 per cent); the openings are usually lateral; and recurrence is rare. The current consensus is that they arise from congenitally abnormal (deep) anal crypts which predispose to infection that proceeds to perianal abscess and subsequently to fistula formation. They are practically always intrasphincteric and are treated effectively with simple fistulotomy; fistulectomy is rarely if ever required. Abnormally deep crypts should also be unroofed.

Anal fissures

Superficial acute posterior anal fissures are the most common cause of hematoschezia in childhood. They are caused by a firm stool which tears the delicate anoderm distal to the dentate line. The pain associated with the fissure potentiates the pattern of constipation which in turn perpetuates the problem. The vicious cycle of hard stool—tear—pain—spasm—constipation—hard stool can be managed with gentle anal dilatation, stool softeners, laxatives, and sitz baths.

Chronic anal fissures, particularly those located laterally, should raise the suspicion for Crohn's disease or for immunodeficiency states. Up to one-half of children with Crohn's disease have perianal involvement as one of the initial manifestations of the disease. Management of chronic anal fissures in children is similar to that in adults and may include anal dilatation, lateral internal sphincterotomy, or injection of botulinum toxin.

Hemorrhoids

Hemorrhoids are rare in children unless associated with portal hypertension. Other sources of bleeding (polyps or fissures), protrusions from the anus (polyps or prolapse), or itching (*Enterobius vermicularis* infection) should be sought. Thrombosed external hemorrhoids in teenagers are treated with incision and evacuation of the thrombus, if acute, or hemorrhoidectomy.

Juvenile polyps

Juvenile polyps are benign hamartomatous lesions which are usually solitary and located in the rectum, although they can be multiple or present as juvenile polyposis. They are a frequent source of minor bright red bleeding from the rectum and they may prolapse through the anus and have the appearance of a hemorrhoid. Treatment consists of transanal polypectomy and proctosigmoidoscopy to exclude the presence of additional polyps.

Rectal prolapse

Rectal prolapse in children occurs equally in boys and girls, usually between the ages of 1 and 3 years (mean 32 months). The prolapse may be only mucosal (having radial folds) or full thickness (‘rectal procendentia', having circular folds). The most frequent underlying causes are constipation (28 to 54 per cent), diarrhea (20 per cent), and cystic fibrosis (15 to 20 per cent). All children with rectal prolapse having loose stools should be tested for parasites and cystic fibrosis (quantitative pilocarpine iontophoresis sweat chloride test) since rectal prolapse will be the initial manifestation of cystic fibrosis in 18 to 23 per cent of patients. Children with neuromuscular disorders or anatomic abnormalities such as myelomeningocele or bladder exstrophy frequently have full-thickness prolapse.

After reduction of the prolapse, treatment is directed toward the underlying cause (antidiarrheal agents, antiparasitic agents, or bulk laxatives). The parents are instructed to reduce the prolapse as soon as it is noticed in order to help prevent edema, bleeding, and soilage. Adequate pancreatic enzyme replacement in children with cystic fibrosis is usually curative.

In children with neuromuscular or anatomic abnormalities, conservative measures may fail. In these instances, a wide variety of surgical procedures such as injection sclerotherapy, levator repair and posterior suspension, posterior saggital anorectoplasty, transanal mucosal sleeve resection, or the modified Thiersch procedure can be used. The simplest effective treatment involves the injection of 5 per cent phenol in almond oil (up to 10 ml) or 50 per cent glucose in water (1 ml/kg) circumferentially into the rectal submucosa 6 to 8 cm above the anal verge. Success rates up to 100 per cent following one or two injections are reported.

Condyloma acuminatum

Condyloma acuminatum is an infection caused by the human papilloma virus (**HPV**), usually types 6 or 11. In children the lesions are usually perianal, non-confluent, and located on the skin rather than on the mucosa. Possible modes of transmission of HPV in children include sexual (including sexual abuse), transmission to an infant from an infected mother either transplacentally or via the birth canal, or transmission of a non-genital type of virus to the genital area. Non-sexual transmission is common in children under 3 years of age, whereas sexual abuse accounts for up to 71 per cent of cases in older children.

Topical treatment with podophylline is effective if few warts are present on the skin. For numerous warts or for those located on the mucosa, fulguration with electrocautery or ablation with a carbon dioxide laser is indicated. Eye protection and a suction evacuator to remove aerosolized virus particles must be used. Intralesional injection with natural interferon-a has been used successfully in adults but has not been reported in children.

Anterior ectopic anus

A rare syndrome of constipation which begins in early infancy and is characterized by straining on defecation and anterior displacement of an otherwise normal-appearing anus associated with a posterior rectal shelf, detected on physical examination or radiologically, has been described in some children, usually females. Careful inspection shows that the anus is eccentrically located in the pigmented skin of the perineum. Treatment consists of posterior displacement of the anus after division of the posterior rectal shelf and external sphincter. Anal dilatations are required postoperatively to prevent stenosis. A procedure which is carried out too aggressively could result in incontinence.

Further reading

Al-Salem AH, Laing W, Talwalker V. Fistula-in-ano in infancy and childhood. *Journal of Pediatric Surgery* 1994; **29**: 436–4. [This is a retrospective study of 36 children with fistula-in-ano treated during a 3-year period.]

Ashcraft KW, Garred JL, Holder TM, Amoury RA, Sharp RJ, Murphy JP. Rectal prolapse: 17-year experience with the posterior repair and suspension. *Journal of Pediatric Surgery* 1990; **25**:

992–95. [Forty-six children treated for rectal prolapse with posterior repair and suspension over a 17-year period at a children's hospital are discussed.]

Chan WK-Y, Kay SM, Laberge J-M, Gallucci JG, Bensoussan AL, Yazbeck S. Injection sclerotherapy in the treatment of rectal prolapse in infants and children. *Journal of Pediatric Surgery* 1998; **33**: 255–8. [A 16-year period of treating 28 children with rectal prolapse with injection sclerotherapy is reported.]

Cohen BA, Honig P, Androphy E. Anogenital warts in children—clinical and virologic evaluation for sexual abuse. *Archives of Dermatology* 1990; **126**: 1575–80. [This is a thorough discussion of anogenital warts in children based on treatment of 73 patients over a 2-year period.]

Hendren WH. Constipation caused by anterior location of the anus and its surgical correction. *Journal of Pediatric Surgery* 1978; **13**: 505–12. [An experience with 134 patients with an anteriorly dislocated anus treated between 1960 and 1977 is reported.]

Stafford WS. Other disorders of the anus and rectum, anorectal function. In: O'Neill Jr JA, Rowe MI, Grosfeld JL, Fonkalsrud EW, and Coran AG, eds, *Pediatric surgery*, 5th edn, pp. 1449–60. Mosby, Philadelphia, 1998. [This is a thorough discussion of anorectal problems in children including constipation, rectal prolapse, anal fissure, perianal and perirectal abscesses, fistula-in-ano, hemorrhoids, and condyloma acuminatum.]

Zempsky WT, Rosenstein BJ. The cause of rectal prolapse in children. *American Journal of Diseases of Children* 1988; **142**: 338–9. [This details the etiology of rectal prolapse in 54 pediatric patients.]

42.8 The abdominal wall

P. K. H. Tam

Exomphalos and gastroschisis
Embryology
Incidence
Associated conditions
Antenatal management
Postnatal management
Definitive management
Prognosis
Midline syndromes
Lower midline syndrome
Upper midline syndrome (pentalogy of Cantrell)
The umbilicus
Embryology
Umbilical abnormalities
Umbilical and paraumbilical hernia
Epigastric hernia
Divarication of recti
Prune belly syndrome
Inguinal hernia and hydrocele
Embryology
Inguinal hernia
Hydrocele
Femoral hernia
Further reading

Exomphalos and gastroschisis

Exomphalos (omphalocele) and gastroschisis are the most common forms of congenital defects of the abdominal wall. Exomphalos (Fig. 1) is characterized by herniation of intra-abdominal contents through a full-thickness umbilical defect that is covered by a sac composed of peritoneum internally and amnion of the umbilical cord externally. The umbilical vessels are usually spread out over the sac. The defect varies between 4 to 12 cm in size and often allows evisceration of both liver and bowel. The bowel is non-rotated but is otherwise normal. The abdominal cavity is underdeveloped (Table 1). Small defects (less than 4 cm) are sometimes termed hernias of the cord: they are easily repaired, have a good prognosis, and are considered to be a separate entity.

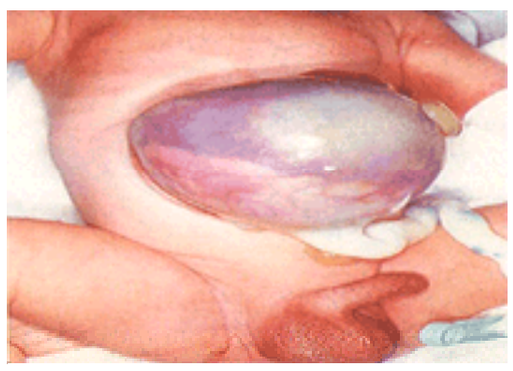


Fig. 1. A large exomphalos with liver and small bowel visible within the sac.

	Exomphalos	Gastroschisis
Location	Umbilical ring	Right of umbilicus
Defect	4–12 cm	< 4 cm
Sac	Present	Absent
Cord	On to sac	Normal
Liver herniation	Yes	No
Bowel:		
Non-rotation	Yes	Yes
Appearance	Normal	Matted, thickened
Function	Normal	Dysmotile
Atresia	Rare	Common
Associated anomalies and syndromes	Common	Rare

Table 1

Gastroschisis is characterized by a relatively small defect in the abdominal wall, 1 to 4 cm in diameter, usually just to the right of the umbilicus (Fig. 2). The herniated viscera are not covered by a sac and seldom include the liver. The non-rotated bowel is often matted, thickened, foreshortened, and covered by a fibrinous ‘peel' as a result of exposure to amniotic fluid. Prolonged ileus is common. Constriction at the defect in the abdominal wall in the intrauterine period may lead to intestinal atresia (10–15 per cent); postnatally, the bowel is at risk for strangulation.



Fig. 2. Gastroschisis: note the defect is beside the umbilicus and the loops of small intestine are matted together.

Embryology

The anterior abdominal wall is developed from four folds (cephalic, caudal, and lateral) that meet in the centre to form the umbilical ring, a process completed by the fourth week of gestation. Rapid growth of the midgut results in herniation through the umbilicus to form the ‘physiological hernia'. The intestine returns to the abdominal cavity between the tenth and twelfth week, when normal intestinal rotation takes place. The recti and oblique muscles are seen within the abdominal wall by day 52.

Duhamel suggested that exomphalos represents a failure of the lateral folds of the body wall to fuse during formation of the body of the embryo. De Vries proposed that the abnormality is caused by persistence of the body stalk of the germinal disc, associated with arrest of development of the musculature of the abdominal wall.

The embryogenesis of gastroschisis is more controversial. According to Duhamel's theory, gastroschisis occurs as a result of a localized teratogenic incident that affects the normal differentiation of the periumbilical mesoderm. De Vries argued that a defect in involution of the right umbilical vein causes local ischaemia, resulting in necrosis of the periumbilical abdominal wall. Hoyme *et al.* suggested that gastroschisis results from an early gestational vascular accident involving the omphalomesenteric artery. Shaw proposed that gastroschisis is the result of early prenatal rupture of a small exomphalos.

Incidence

Exomphalos occurs in between 1 in 4000 and 1 in 10 000 live births. The incidence of gastroschisis seems to be increasing, with a reported incidence of 1 in 17 000 in 1952 and an estimated present incidence of about 1 in 5000. The reasons for this apparent increase are unclear. Exomphalos is more common in males than females, with a sex ratio of 3:2, but gastroschisis occurs equally among the sexes. Gestational age and birth weight tend to be subnormal to some extent in exomphalos and more so in gastroschisis.

Associated conditions

Associated congenital malformations are frequently seen in children with exomphalos; such abnormalities will occur in about two-thirds of those with large defects. Up to 10 per cent of patients have the Beckwith–Wiedemann syndrome, with macroglossia and organomegaly. The principal danger in these patients is postnatal hypoglycaemia (due to pancreatic islet hyperplasia) causing cerebral damage. Associated major cardiac anomalies are seen in 25 to 33 per cent, and about 25 per cent have gastrointestinal abnormalities. Trisomy (usually 21 or 18) is seen in 20 per cent of patients. Exomphalos may also occur as part of lower and upper midline syndromes (see below).

Gastroschisis is rarely associated with abnormalities outside the gastrointestinal system. Intestinal atresia, often resulting from local ischaemia at the neck of the sac, is by far the most common abnormality associated with gastroschisis, with an incidence of between 11 and 23 per cent. A significant number of these are colonic (which are otherwise rare).

Antenatal management

An elevated maternal serum a-fetoprotein may provide the first clue to the fetal presence of a defect of the anterior abdominal wall. More often, exomphalos and gastroschisis are diagnosed antenatally by maternal ultrasonography. For exomphalos, amniocentesis is advisable to exclude associated chromosomal abnormalities. Gastroschisis and non-syndromic exomphalos in themselves are not indications for termination of pregnancy since in most instances the prognosis is excellent. Termination should be considered for exomphalos associated with severe anomalies or potentially lethal syndromes. Parental counselling should involve a paediatric surgeon. Delivery at a specialist centre avoids the unnecessary risks of neonatal transport and allows planning of surgery. Neither early delivery nor caesarean section have been shown to confer any additional advantage.

Postnatal management

If the defect in the abdominal wall has not been detected antenatally, diagnosis will be obvious at birth. Exomphalos and gastroschisis present many difficulties in management and should be treated in a paediatric surgical unit. The principal immediate problems are profound heat and fluid loss from the eviscerated bowel and intestinal distension from swallowed air (especially if the baby is ventilated). Prompt postnatal management is therefore of great importance. A large nasogastric tube must be inserted and left on free drainage with regular aspiration. Wrapping up the whole of the baby's trunk in ‘cling film' or a plastic bag minimizes both heat and fluid loss. Fluid deficit is corrected by intravenous infusion of large volumes of crystalloid solution; often two to three times the maintenance requirement is needed. The rate of infusion is adjusted according to clinical response. Every baby with exomphalos must be examined for signs of the Beckwith–Wiedemann syndrome since severe hypoglycaemia may occur shortly after birth, resulting in cerebral damage. A thorough clinical cardiac assessment should be made preoperatively. Broad-spectrum antibiotics are given. After proper resuscitation, infants with defects of the abdominal wall not born in a tertiary centre should be transferred there in a warm incubator accompanied by appropriate personnel for definitive management.

Definitive management

Non-operative treatment is rarely undertaken in infants with giant exomphalos, for whom surgical repair is unlikely to be successful or is impossible because of medical (for example, life-threatening cardiac anomalies) or non-medical reasons. Topical treatment with antiseptic dressings allows gradual epithelialization over many weeks but leaves behind a large ventral hernia requiring later repair.

Operative repair of exomphalos and gastroschisis may be primary or staged. For exomphalos, the umbilical vessels are divided and the sac excised. For both conditions the abdominal cavity is forcibly stretched to allow the herniated viscera to be returned (after division of Ladd's bands, if present). It is normally possible to effect a primary closure, but closure may result in such an increase of intra-abdominal pressure that ventilation or venous return are compromised. If this is the case a prosthetic ‘silo' may be constructed from Silastic sheeting. The gut is then gradually returned to the abdominal cavity by reducing serially the size of the sac over the following week, after which the wound is repaired. Associated intestinal atresias are dealt with by primary anastomosis, the creation of stomas or by delayed repair, depending on individual clinical conditions.

When operating for gastroschisis a central venous feeding line may be sited, since normal intestinal activity often takes two or more weeks to become established, necessitating total parenteral nutrition. Postoperative ventilation is frequently necessary.

Prognosis

Infants with gastroschisis and those with exomphalos uncomplicated by associated abnormalities have excellent survival and may grow to lead a completely normal life.

Midline syndromes

Lower midline syndrome

Failure of development of the caudal fold results in this rare association of somatic and splanchnic abnormalities: hypogastric omphalocele/abdominal-wall defect, exstrophy complex, absence of hindgut, neural tube and vertebral defects. Exstrophy complex consists of a spectrum of anatomical variants including cloacal exstrophy, which is the most severely affected form, and bladder exstrophy, which is less rare. Cloacal exstrophy affects between 1 in 200 000 and 1 in 400 000 live births, and occurs more frequently in males. The bladder and hindgut are exstrophic and the short small bowel terminates as a stoma.

If untreated, this severe abnormality is usually rapidly fatal due to ileal fluid loss. Repair should be carried out in the new-born period and only in a specialist unit. Even after successful closure of the defect, many potential problems remain: major associated abnormalities are found in about 85 per cent of cases, affecting the gut, spine, and urinary tract.

Upper midline syndrome (pentalogy of Cantrell)

Failure of development of the cephalic fold results in a pentalogy comprising an upper midline exomphalos, an anterior diaphragmatic hernia with an associated sternal cleft, ectopia cordis, and intracardiac defects. It is rare and frequently fatal, despite attempts at operative correction. Operation may be avoided in patients with lethal

cardiac defects if preoperative echocardiography is performed.

The umbilicus

Embryology

The vitellointestinal duct connects the primitive midgut to the yolk sac. Spontaneous obliteration usually occurs before the sixth week of embryonic life. The urachus (connecting the bladder to the allantois) develops from the urogenital compartment of the cloaca. It is usually obliterated by the time of birth.

Umbilical abnormalities

Anatomical abnormalities and discharge at the umbilicus are frequent reasons for referral to the paediatric surgeon. Discharge of fluid or mucus in the new-born period is likely to be caused by local infection, which usually responds to antibiotics and improved hygiene. Inadequately treated omphalitis may rarely lead to portal venous sepsis and thrombosis. Umbilical granuloma results from incomplete separation of the umbilical cord. Application of a silver nitrate stick is usually effective as a rapid treatment.

Anatomical abnormalities are related either to the gastrointestinal or to the urinary tracts, and should be considered when umbilical discharge presents outside early infancy. An umbilical polyp is a cherry-red nodule that represents failure of complete obliteration of the vitellointestinal duct. The lesion, which usually contains small-intestinal mucosa, should be excised. A search should be made for remnants of vitellointestinal duct. An incomplete vitellointestinal duct may be represented by an umbilical sinus, a cyst, a fibrous band connecting the umbilicus and terminal ileum, or Meckel's diverticulum. The intra-abdominal band may cause internal hernia or volvulus of the small intestine at any age. Meckel's diverticulum could result in recurrent abdominal pain, infection, bleeding, intussusception, or intestinal obstruction. A completely patent vitellointestinal duct is uncommon and may open into a normal umbilicus or an exomphalos. There could be faeculent discharge or, rarely, prolapse of ileum via the duct. At surgery, malrotation must be treated if present.

Patent urachus is very rare and is usually suspected when there is passage of urine via the umbilicus during the first 3 weeks of life. Obstruction of bladder outflow also may be present. Other urachal remnants include urachal sinus (umbilical), cyst, and diverticulum (bladder). Urachal cysts may become infected, giving rise to abscesses; these present as a tender midline swelling and may point to the bladder or umbilicus.

Umbilical and paraumbilical hernia

Umbilical hernia is common in infants (32 per cent in black infants, 4 per cent in white infants) and represents a delay in closure of the fascial ring resulting in a circular orifice with intact skin and peritoneum. The condition is particularly common in preterm infants and children with Down's syndrome, hypothyroidism, mucopolysaccharidosis, and Beckwith–Wiedemann syndrome. The majority of umbilical hernias resolve spontaneously within the first 5 years of life and complications are extremely rare. A hernia that persists to this age is treated surgically. Non-absorbable or slow-absorbing sutures should be used to repair the defect. Paraumbilical hernia represents a fascial defect that is transverse elliptical and located just above or occasionally below the umbilicus. The lesion does not resolve spontaneously and operative repair is required.

Epigastric hernia

Epigastric hernia represents a small, midline defect in the linea alba located usually one-third or one-half the distance from the umbilicus to the xiphisternum above the umbilicus. This allows extraperitoneal fat to herniate through into the subcutaneous tissues causing discomfort. The defect should be localized by careful examination and repaired.

Divarication of recti

This occurs commonly in infants and requires no treatment.

Prune belly syndrome

This condition of unknown aetiology is otherwise known as Eagle–Barrett syndrome or triad syndrome. The triad of abnormalities consists of a deficiency of the musculature of the anterior abdominal wall, which results in a wrinkled skin giving the appearance of a prune, cryptorchidism, and malformations of the urinary tract. The incidence is between 1 in 29 000 and 1 in 40 000 live births, and 95 per cent of affected children are male. Plication operations designed to treat laxity of the abdominal wall yield little or no improvement to the function or appearance of the abdomen. Historically, the prune belly syndrome had a very high infant mortality rate, due principally to urinary abnormalities; recent improvements in urological management have reduced this considerably.

Inguinal hernia and hydrocele

Embryology

During the third month of gestation the processus vaginalis, a peritoneal diverticulum, is formed at the internal inguinal ring. The processus vaginalis extends to the scrotum in association with descent of the testis. The processus closes in the last few weeks of term gestation when testicular descent is completed, leaving a small remnant (tunica vaginalis) attached to the testes. The factors responsible for obliteration of the processus are unknown. Inguinal hernia and hydrocele in infants and children result from failure of this obliterative process (Fig. 3).

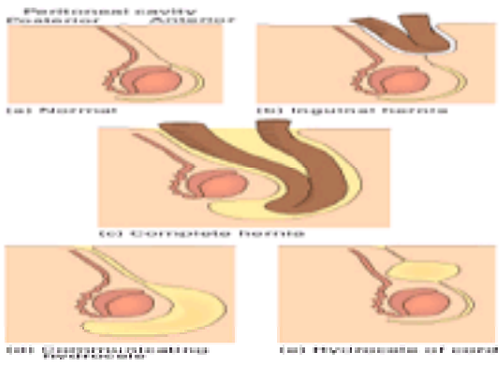


Fig. 3. Inguinal hernia and hydrocele.

Inguinal hernia

Incidence

Inguinal hernia is one of the most common surgical conditions in childhood, affecting 1 to 5 per cent of children. The incidence is higher in infants and this is further increased with prematurity: 10 per cent of preterm infants and 35 per cent of infants less than 28 weeks' gestation are affected. The condition is four to ten times more common in boys than girls. The right side is more often involved and bilateral occurrence is common (10 per cent).

Types

A patent processus vaginalis, which is usually wide, only becomes a hernia when it contains viscera, forming an indirect inguinal hernia. A blind-ending sac may be

found or the processus may remain patent throughout its whole length.

Diagnosis

The presenting symptom is a reducible groin swelling. The diagnosis is often made from the history alone, especially in younger children, since the hernia may not be visible. In the presence of an empty sac, the cord is described as thickened (silk glove sign). In the acute situation an incarcerated hernia has to be differentiated from acute hydrocele (painless), inguinal lymphadenitis (signs of inflammation), or torsion of an undescended testis (empty scrotum).

Complications

Incarceration occurs in 2 to 18 per cent of patients; the incidence rises to 30 per cent or more in the first 8 weeks of life. This poses two dangers, intestinal gangrene and gonadal infarction. The risk of gonadal infarction secondary to incarceration appears to be much higher (30 per cent) in infants less than 12 weeks of age than in older children.

Management

Herniotomy should be done as an early, elective procedure because of the risk of incarceration. In the case of very small premature babies in special care units, however, the benefits of waiting for the lungs to mature outweigh the risk of incarceration and herniotomy is done just before the baby is ready for discharge from hospital.

The inguinal ligament is identified in a groin-crease incision and followed medially until the spermatic cord is seen at the external ring. In children less than 18 months old the operation may be performed through the external ring since the inguinal canal is short, but in older children access is best obtained by making a small 'buttonhole' incision in the external oblique aponeurosis through which the cord may be delivered. After separating the patent processus from the rest of the cord contents, and ensuring that it is empty, the processus is divided and tied off proximally as close as possible to the internal ring, using a slowly absorbable material. The distal end should be left open to avoid the complication of iatrogenic hydrocele. If the external oblique aponeurosis has been opened, it is repaired with absorbable sutures. The skin is closed with a subcuticular suture. Routine exploration of the contralateral groin, advocated by some surgeons because of the presence of a patent processus vaginalis contralaterally in 30 to 60 per cent, is not necessary as the majority will not develop into a hernia.

Management of incarcerated hernia

Provided the child is well, with no evidence of intestinal gangrene, an attempt should be made to reduce the hernia by taxis. Sedation with morphine considerably aids this procedure and it is essential that an intravenous infusion be sited before the attempt. After reduction, observations must be made and recorded in case of accidental reduction of gangrenous gut or *en masse* reduction. In this way the majority of incarcerated hernias may be reduced: operation should be performed 24 to 48 h later when tissue oedema has subsided. Failure of manual reduction is an indication for emergency operation. This may be an extremely difficult procedure because the processus is very friable.

Postoperative complications

Haematoma and sepsis are occasional complications. The hernia may recur at any time postoperatively, and this is much more common after emergency operation. The testis may become atrophic or remain high if it has been drawn up and not brought back to the scrotum during herniotomy.

Hydrocele

A hydrocele results when the narrowly patent processus allows peritoneal fluid to accumulate inside. It occurs most frequently in infants. It is seen more often on the right side but the condition is frequently bilateral. It presents as a soft, cystic, scrotal swelling. A hydrocele transilluminates brilliantly; however, an infantile hernia containing thin-walled bowel may also transilluminate. A hydrocele may fluctuate in size as more fluid accumulates towards evening but, unlike hernia, it is not reducible. Occasionally, a tense hydrocele arises acutely. Encysted hydroceles are less common: these are termed hydrocele of the cord in boys and hydrocele of the canal of Nuck in girls. In the majority of infants, a hydrocele will resolve spontaneously by 1 year of age and operation is necessary only if it persists until this time. The operation consists of high ligation of the patent processus vaginalis, which is exactly the same procedure as for inguinal hernia, and fluid in the tunica vaginalis is emptied by aspiration.

Femoral hernia

This extremely rare condition is seen most frequently in 5- to 10-year-old girls. Strangulation is uncommon. Operation is by one of the standard approaches.

Further reading

Cilley RE, Krummel TM. Disorders of the umbilicus. In: O'Neill JA Jr, Rowe MI, Grosfeld JL, Fonkalsrud EW, Coran AG, ed. *Pediatric surgery*, 5th edn, pp. 1029–44. Mosby-Year Book, St Louis, 1998. [Chapter on umbilical pathology from the standard reference textbook of paediatric surgery.]

Cooney DR. Defects of the abdominal wall. In: O'Neill JA Jr, Rowe MI, Grosfeld JL, Fonkalsrud EW, Coran AG, ed. *Pediatric surgery*, 5th edn, pp. 1045–69. Mosby-Year Book, St Louis, 1998. [Chapter on abdominal wall defects from the standard reference textbook of paediatric surgery.]

Dykes EH. Prenatal diagnosis and management of abdominal wall defects. *Seminars in Pediatric Surgery* 1996; **5**:90–4.

Grosfeld JL. Current concepts in inguinal hernia in infants and children. *World Journal of Surgery* 1989;**13**:506–15. [Comprehensive review article on inguinal hernia.]

Lloyd DA, Rintala RJ. Inguinal hernia and hydrocele. In: O'Neill JA Jr, Rowe MI, Grosfeld JL, Fonkalsrud EW, Coran AG, ed. *Pediatric surgery*, 5th edn, pp. 1071–86. Mosby-Year Book, St Louis, 1998. [Chapter on hernia and hydrocele from the standard textbook of paediatric surgery.]

Misra D, Hewitt G, Potts SR, Brown S, Boston VE. Inguinal herniotomy in young infants, with emphasis on premature neonates. *Journal of Pediatric Surgery* 1994; **29**:1496–8. [A review of special problems of management of inguinal hernia in young and premature infants.]

Nuchtern JG, Baxter R, Hatch EI Jr. Nonoperative initial management versus silon chimney for treatment of giant omphalocele. *Journal of Pediatric Surgery* 1995; **30**:771–6. [Revisiting the role of conservative treatment of giant omphalocele in the modern era.]

Shuster SR. A new method for the staged repair of large omphaloceles. *Surgery, Gynecology and Obstetrics* 1967; **125**:837–50. [Classic paper on the introduction of 'silo' as treatment for large abdominal wall defects.]

Tam PKH, Lister J. Femoral hernia in children. *Archives of Surgery* 1984; **119**:1151–64. [A large series of femoral hernia, highlighting difficulties of diagnosis and management in children.]

Tam PKH. Inguinal hernia. In: Lister J, Irving IM, ed. *Neonatal surgery*, 3rd edn, pp. 367–75. Butterworths, London, 1990. [Chapter on inguinal hernia in neonates - a single institution's experience over several decades.]

Wiener ES *et al.* Hernia survey of the section on surgery of the American Academy of Pediatrics. *Journal of Pediatric Surgery* 1996; **31**:1166–9. [Pattern of current practice on inguinal hernia among paediatric surgeons in North America.]

42.9 Surgery of the urinary tract in children

W. Hardy Hendren and Joseph G. Borer

Introduction
Vesicoureteral reflux
Ureteropelvic junction obstruction
Megaureter
Ureterocele
Urethral valves
Ureteral reimplantation
Transureteroureterostomy
Bladder augmentation
Psoas hitch
Bladder neck repair
Urinary diversion
Continent diversion
Urinary undiversion
Epispadias
Exstrophy of the bladder
Ectopic ureter
Neuropathic bladder
Hypospadias
Undescended testicle
Intersex states
Cloacal malformations
Cloacal exstrophy
Tumors
Further reading

Introduction

Unrecognized obstruction can lead to urinary infection, gradual loss of renal function, and ultimately death from renal failure. When recognized early an obstruction can be repaired, even in the newborn, giving normal renal function and longevity. Urinary incontinence, which accompanies certain urologic disorders, can adversely affect social development of the child. Similarly, genital malformations, when repaired properly, can produce a good cosmetic result and self-image for the child. Functional and social disasters await the child who grows without expert surgical correction of genital malformations. There are tumors unique to the pediatric urinary tract, which although highly malignant, have a high rate of cure when treated by modern therapy.

Table 1 outlines the categories of disease to be described and the major conditions to be included. Several problems may occur in one condition: for example, ectopic ureters can be associated with both obstruction and incontinence. Similarly, epispadias is both a congenital malformation of the genitals and a cause of urinary incontinence because the bladder neck is usually deficient. Furthermore, a child's urinary tract may have multiple associated problems, such as urethral valves, ureterovesical junction obstruction, and ureteropelvic junction obstruction.

Vesicoureteral reflux	
Obstructions	
Ureteropelvic junction obstruction (the upper ureter)	
Megaureter (the lower ureter)	
Ureterocele	
Urethral valves	
Incontinence	
Epispadias	
Exstrophy of the bladder	
Ectopic ureter	
Neuropathic bladder	
Genital malformations	
Hypospadias	
Epispadias	
Undescended testicle	
Intersex states	
Cloacal malformations	
Cloacal exstrophy	
Tumors	
Kidney	
Adrenal	
Bladder	
Genital tract	
Gonads	

Table 1 Pediatric urologic disorders

Table 2 lists the most common surgical maneuvers in pediatric urologic surgery.

Ureteral reimplantation
Transureteroureterostomy
Bladder augmentation
Small bowel
Colon
Stomach
Psoas hitch
Bladder neck repair
Urinary diversion
Continent urinary diversion
Urinary undiversion

Table 2 Surgical maneuvers in urologic repair

Vesicoureteral reflux

Vesicoureteral reflux, the backwash of urine from the bladder up the ureters to the kidney, is the most common cause of serious urinary tract infection in young patients. Reflux is usually caused by congenital maldevelopment of the ureterovesical junction. Normally, a ureter enters the bladder trigone at an oblique angle and has a submucosal segment of ureter with muscle behind it. This anatomy produces a valve-like mechanism that prevents backflow of urine during voiding. If the ureter is located more laterally than normal, often having a straight course through the bladder wall, this valve mechanism is deficient. Figure 1 contrasts normal ureterovesical junction anatomy with that usually seen with reflux, and illustrates how reflux can be corrected surgically by reimplanting the ureter.

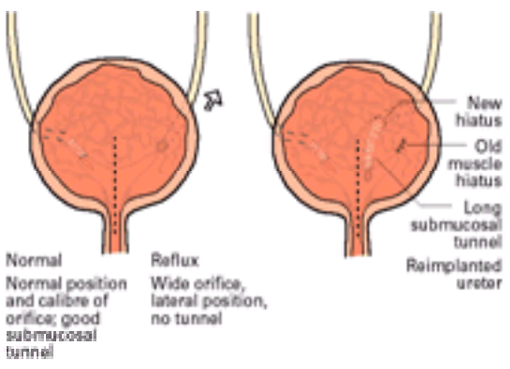


Fig. 1. Anatomy of normal ureter, ureter with vesicoureteral reflux, and ureteral reimplantation to correct reflux.

Since reflux is often familial, asymptomatic siblings of an affected child should undergo screening tests. These include ultrasound examination of the kidneys to exclude cortical scarring and caliceal dilatation, together with voiding cystourethrography to visualize the urethra, bladder neck, the process of bladder emptying, and the absence of backwash. A simple static cystogram of the full bladder is insufficient: reflux frequently occurs only during voiding. A voiding study in males also excludes the presence of urethral valves, an obstruction that can cause high pressures in the bladder during voiding and may cause a relatively normal ureterovesical junction to backwash. An experienced endoscopist can predict the presence of vesicoureteral reflux by the configuration of the ureteral orifice: a normal ureteral orifice is small and elliptically shaped. A refluxing orifice is sometimes so dilated that the endoscope can be passed from the bladder up into the ureter.

Every congenital malformation occurs in a wide spectrum of severity, ranging from almost normal anatomy to obvious abnormality. The treatment of vesicoureteral reflux also varies with its severity, and depends on whether the upper urinary tracts are normal or scarred, and whether the abnormality is minimal or severe. Some reflux will disappear in time: a congenitally short ureterovesical tunnel can elongate as a child grows.

Reimplantation of the ureter is a well established operation that should correct reflux without creating obstruction, in virtually all cases. A poor operation can have disastrous results, causing obstruction of the ureter or persistent reflux. A child with an unscarred upper urinary tract and mild reflux can be treated by observation and antibiotic therapy, with frequent monitoring of urinary cultures. Children with dilatation and/or scarring of the upper urinary tract or very abnormal anatomy which is not likely to improve require surgical treatment.

Subureteric injection of Teflon paste to provide an artificial backing to the ureter can stop reflux in some patients. This technique has not met universal approval because its long-term safety is not established.

Three anatomic conditions predispose to vesicoureteral reflux: paraureteral diverticulum, in which the lower ureter lacks muscle backing; duplex ureters, in which the lower pole orifice, which lies more craniad in the bladder, has too short a tunnel; and ureteral ectopia, in which the ureter enters the bladder neck or urethra and has a straight course that allows reflux. These anomalies are best managed surgically. Figure 2 illustrates how severe renal atrophy can occur when reflux and infection are allowed to persist.

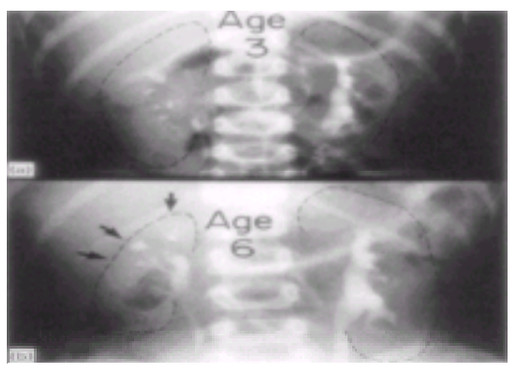


Fig. 2. Intravenous pyelograms from a young girl to show devastating effects of reflux. (a) Intravenous pyelogram at age 3 years. Reflux was not corrected, and she was not maintained on antimicrobial therapy and frequent urine cultures. (b) Severe atrophy of right kidney evident at age 6 years caused by continuing reflux with infection. Successful correction of reflux prevents such damage.

Ureteropelvic junction obstruction

Obstruction at the upper end of the ureter is one of the most common surgical problems in infants and children occurring twice as often in boys as in girls (Fig. 3). Histologic examination of the narrow ureteropelvic junction segment usually discloses disordered smooth muscle bundles and excessive collagen in the wall of the ureter. In about 10 per cent of cases an aberrant artery to the lower pole of the kidney crosses anterior to the point of obstruction, and 20 per cent of cases are bilateral. Since the advent of antenatal ultrasound examination, a large proportion of these cases are detected antenatally. If the condition is not detected before birth, symptoms leading to its clinical detection include urinary infection, abdominal pain, palpation of an abdominal mass, and hematuria. For some patients the condition is not diagnosed until adulthood, when long-term stasis may lead to stone formation. Relatively minor trauma can rupture a hydronephrotic kidney and lead to the diagnosis of congenital obstruction.

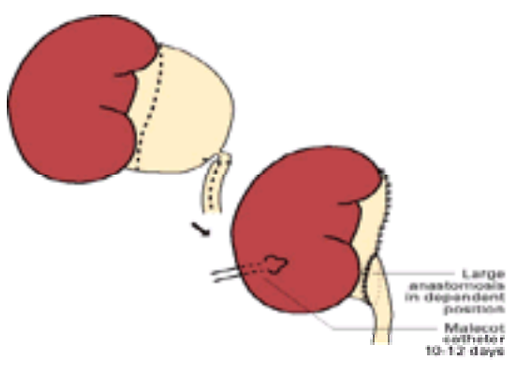


Fig. 3. Ureteropelvic junction obstruction and its repair by dismembered pyeloplasty. The renal pelvis should be reduced in size only when very large. Temporary nephrostomy drainage is only used occasionally.

The diagnosis of hydronephrosis due to ureteropelvic junction obstruction is not difficult. Hydronephrosis can be detected by ultrasound examination of the kidneys, radionuclide scan with delayed emptying, or intravenous pyelogram. If the contrast agent is not seen on early films, delayed films should be obtained after 4 to 8 h, or even later. A voiding cystogram should also be obtained to rule out the presence of associated vesicoureteral reflux. Massive vesicoureteral reflux produces tortuosity of the ureteropelvic junction, which can mimic primary ureteropelvic junction obstruction. In some cases, correcting reflux allows the ureteropelvic junction tortuosity and apparent obstruction to disappear. In other cases, however, tortuosity is permanent and both ends of the ureter require corrective surgery. The ureter is often not seen in an intravenous pyelogram, and a retrograde pyelogram should be performed to define accurately the extent of narrowing and rule out the presence of a second level of narrowing.

Figure 4 shows the goal to attain in correcting ureteropelvic junction obstruction. The kidney is approached through a flank incision. The area of narrowing is sometimes reached more easily from behind, but in other cases access is easier from an anterior approach. The ureter is not widely mobilized to avoid injury to its blood supply. The ureteropelvic junction itself is divided, and the dilated renal pelvis is mobilized widely; if the renal pelvis is extremely large, it can be subtotally resected. A moderately enlarged renal pelvis does not require reduction since relief of obstruction is followed by reduction in volume of the renal pelvis. The ureter is spatulated along its lateral aspect to allow a wide anastomosis to be made in a dependent position on the renal pelvis. A water-tight anastomosis of adequate caliber does not require stenting, and a nephrostomy tube is only used to drain the kidney when there is extreme hydronephrosis. The majority of patients do not require temporary nephrostomy drainage, but the flank is always drained in case there is leakage of the anastomosis, although this is rare when the operation has been performed properly.

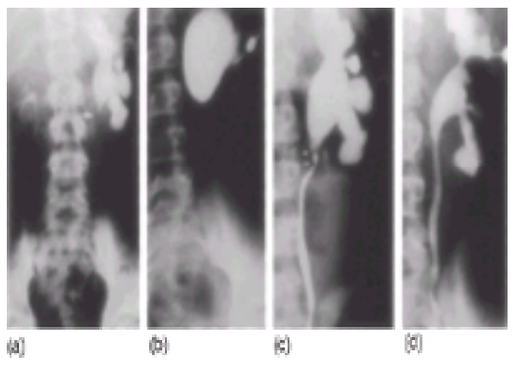


Fig. 4. Ten-year-old boy with typical ureteropelvic junction obstruction. (a) Intravenous pyelogram which shows normal right kidney, and obstruction at upper end of left ureter (arrow). (b) Left antegrade pyelogram which showed no flow through ureteropelvic junction and high pressure. (c) Preoperative combined retrograde pyelogram with antegrade filling of kidney to show two discrete areas of narrowing (arrows). (d) Antegrade pyelogram 1 year after dismembered pyeloplasty showing repaired ureteropelvic junction. The patient originally presented with left flank pain, which disappeared after repair.

Giant hydronephrosis is defined as a kidney that occupies the hemi-abdomen, crossing the midline, extending vertically the length of five vertebrae. At least 50 per cent of those kidneys can be salvaged: they have a surprisingly large degree of function. Radionuclide scans can be helpful in assessing the amount of function present. Temporary decompression with a percutaneously placed nephrostomy tube may be used to decompress the kidney and to measure its function by creatinine clearance determination. When a kidney provides less than 10 per cent of the total renal function, and if the other kidney is normal, nephrectomy is usually the best option. An aberrant vessel, if present, should never be divided, since this can devascularize that segment of the kidney and lead to hypertension. The ureteropelvic junction should be divided, and reanastomosed with a spatulated pyeloplasty, which is usually placed anteriorly to the artery.

Figure 4 illustrates moderately severe hydronephrosis of the left kidney in a 10-year-old boy who presented with left flank pain. He had a normal right kidney and moderately severe hydronephrosis of the left kidney. Percutaneous antegrade pyelography disclosed no drainage through the ureteropelvic junction and reproduced the patient's pain. There were two areas of narrowing in the upper ureter. Study after dismembered pyeloplasty showed free flow of contrast through the ureteropelvic junction.

Megaureter

Megaureter refers to a dilated and sometimes tortuous ureter resulting from intrinsic obstruction at the lower end of the ureter or from massive vesicoureteral reflux (Fig. 5). Wide ureters may be a primary problem caused by maldevelopment of the ureterovesical junction, but with an otherwise normal bladder and urethra; it can also be a secondary condition. Neurogenic bladder, in which emptying is impaired or pressures are high, may produce ureteral dilatation, as may urethral valves with severe outlet obstruction. A ureter that is ectopic in the bladder neck or urethra may be functionally obstructed at rest, but demonstrate massive reflux during micturition. A ureter can also be dilated as a result of extrinsic compression by a pelvic mass or scarring by previous surgery. Extremely abnormal ureters are seen in rare conditions such as ‘prune belly syndrome', which appears to be due to a primary muscular maldevelopment.

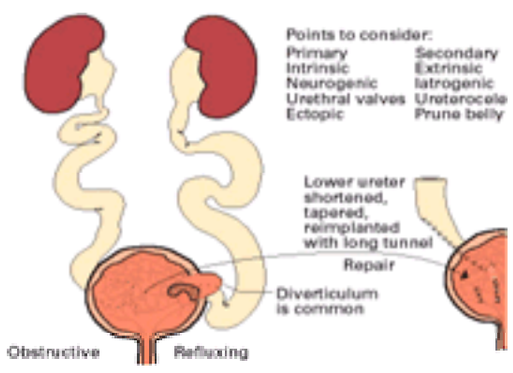


Fig. 5. Principal types of megaureter and method of repair by shortening, tapering, and reimplanting ureter into bladder with non-refluxing tunnel.

The function of the ureter is to transport urine from the kidney to the bladder: it cannot function properly if the ureteral walls do not coapt during peristalsis. If the ureters contain a large volume of urine, stasis can promote urinary infection and stone formation. Surgery can benefit many of these patients, the majority of whom are infants and children who present with urinary tract infection. When the megaureter is clearly not a secondary problem, its surgical repair should be considered, with shortening and tapering of the lower ureter, which is reimplanted into the bladder with a long tunnel to prevent reflux. The tortuosity of the upper ureter often improves spontaneously. In some patients, however, persisting upper ureteral tortuosity can produce partial obstruction at the ureteropelvic junction. Tortuosity can be assessed by serial radiographic examinations, antegrade pressure–perfusion studies, and radionuclide scans with furosemide washout, to measure the speed of emptying of the dilated upper system. If drainage is impaired, the upper ureter can be shortened, straightened, and sometimes tapered to enhance its emptying, as is performed in patients with ureteropelvic junction obstruction. The blood supply of the ureter must be maintained when surgery is undertaken. The striking improvement that can result from surgery for megaureter is illustrated in Fig. 6.

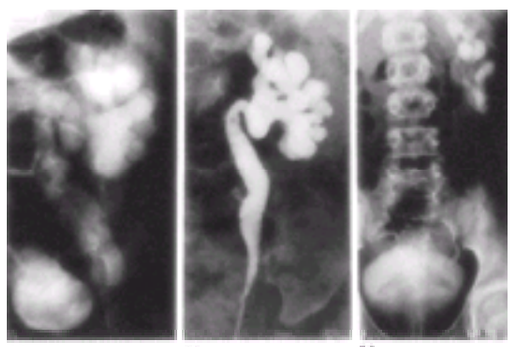


Fig. 6. Obstructive megaureter, solitary left kidney, in a female infant age 4 months. (a) Preoperative intravenous pyelogram. Lower megaureter was repaired by resecting the obstruction, tapering the lower ureter, and reimplanting it. (b) Retrograde pyelogram 6 months later. Note normal caliber of the lower ureter which was no longer obstructed. Tortuousity of the upper ureter did not recede, and, therefore, the upper ureter was shortened, straightened, and tapered. (c) Intravenous pyelogram at age 12 years showing normal ureter. Interestingly, the original blood urea nitrogen (BUN) was 16 mg/100 ml and the serum creatinine was 0.6 mg/100 ml, deceptively normal. However, the creatinine clearance was extremely abnormal, at only 15 l/m^2 body surface area per day. After repair of the obstruction, creatinine clearance was consistently in the range of normal, over 100 l/m^2 body surface area. This patient is now 36 years old. Current BUN 9 mg/100 ml, serum creatinine 0.6 mg/100 ml, and creatinine clearance 108 l/m^2.

Ureterocele

Ureterocele is a term that describes a sacular, cyst-like dilatation at the lower end of a ureter. It occurs in several anatomic forms (Fig. 7), the least severe of which is a small ureterocele at the end of a single ureter that is moderately dilated. These patients are usually boys. More commonly, there is a duplex collecting system in which

the ureterocele is actually the lower end of the ureter to the upper aspect of the kidney which itself is dysplastic or hydronephrotic. This form is more common in girls. A ureterocele is ‘normotopic' when it is confined to the bladder, although a large normotopic ureterocele can obstruct the bladder outlet. An ectopic ureterocele is one in which the muscle defect extends down into the bladder neck and urethra. The entire back wall of the urethra may be affected, and the ureterocele can rupture into the vagina, causing total urinary incontinence.

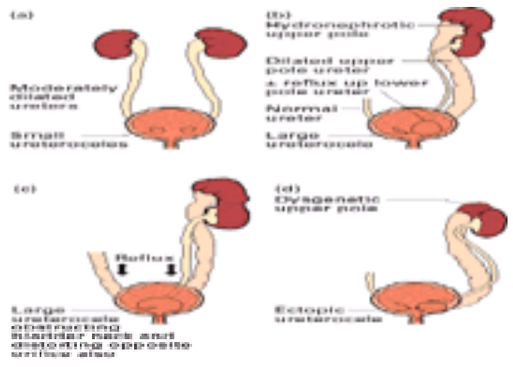


Fig. 7. Types of ureterocele. (a) Small ureteroceles of single ureters with mild obstruction of ureters. (b) Large ureterocele with duplex collecting system. The ureterocele is associated with the ureter to the upper renal moiety. There is often reflux up the ipsilateral lower pole ureter. (c) Large normotopic ureterocele extending down into bladder outlet causing bladder outlet obstruction, and reflux into ipsilateral ureter and contralateral single ureter. (d) Ectopic ureterocele, so called because the muscle defect associated with the ureterocele extends into the bladder neck and urethra. Sometimes an ectopic ureterocele erodes into the vagina, causing incontinence.

The diagnosis of a ureterocele can be made from an intravenous pyelogram (Fig. 8). A small ureterocele is revealed as a round accumulation of contrast in the lower ureter, termed a ‘cobra head' or ‘spring onion'. These patients are usually older children. A duplex collecting system typically appears as a large filling defect in the bladder, which does not contain contrast medium because there is poor renal function of the upper pole moiety associated with the ureterocele. The lower pole of the kidney is usually displaced laterally and downward by the hydronephrotic upper pole. Cystography should be performed in these patients, in whom there is often vesicoureteral reflux along the adjacent lower pole ureter which is located on the summit of the ureterocele. The ureterocele lacks muscle backing, as do patients with primary reflux. There is often reflux along a contralateral single ureter. In the worst cases bilaterally duplicated systems with two large ureteroceles fill the bladder outlet.

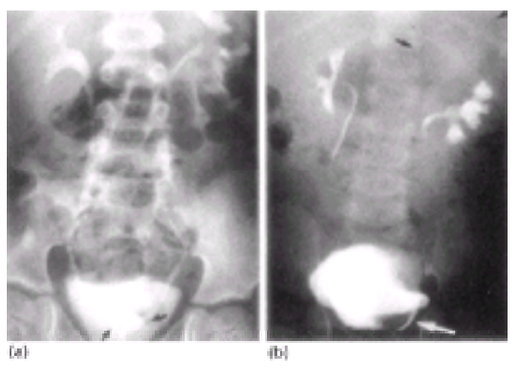


Fig. 8. Radiographs of patients with ureterocele. (a) Seven-year-old boy with two small ureteroceles (arrows). (b) Nine-month-old girl with large left ureterocele. Note displacement of adjacent lower pole downward and laterally by faintly visualized, hydronephrotic upper pole (arrows). Upper pole nephroureterectomy was performed at age 9 months. Reflux persisted to lower pole of left kidney. Therefore, the left lower pole ureter was reimplanted at age 3 years. At age 5 years 45 per cent of the patient's total renal function was generated by the left kidney.

The treatment of a child with a ureterocele varies greatly. A small ureterocele in a single system is best managed by resecting the ureterocele and reimplanting the ureter to create a non-refluxing tunnel, after relieving its obstruction (Fig. 9(a)). Many years ago these patients were managed by ‘unroofing' the ureterocele to relieve its obstruction: this practice merely substituted reflux for obstruction. Surgical correction in the duplex system most often involves removing the ureterocele, its ureter, and the upper pole of the kidney, while repairing the defect in the bladder trigone and reimplanting the adjacent ipsilateral lower pole ureter (Fig. 9(b)). Occasionally, the upper pole of the kidney contains enough renal parenchyma to merit its salvage by pyelopyelostomy.

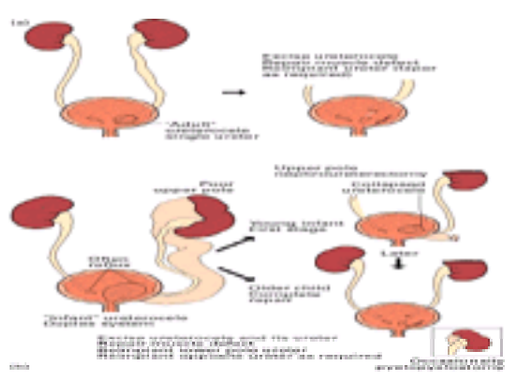


Fig. 9. (a) Repair of small ureterocele of a single ureter. (b) Repair of ureterocele with duplex collecting system. Usually the upper pole and its ureter are removed, repairing the defect in the bladder trigone and bladder neck created by the ureterocele and reimplanting the lower pole ureter. Occasionally, there is enough function in the upper pole to warrant draining it into the adjacent lower pole.

These patients are frequently small infants severely ill with urosepsis. Initial therapy may be drainage of the bladder with a small plastic catheter, and antibiotic treatment of urinary tract infection. Upper pole partial nephroureterectomy can remove the large reservoir of urine that predisposes to infection, causing the ureterocele to collapse. Alternatively, the obstructed upper pole can be temporarily relieved by inserting a percutaneous catheter or puncturing the ureterocele endoscopically to drain it, followed by upper pole nephroureterectomy a few days later. Sometimes a child improves dramatically following removal of the upper pole and its ureter. If there is no associated reflux, it may be possible to delay the lower reconstruction for several years. Although some surgeons would not undertake the lower reconstruction at all in these circumstances, late complications may arise from the defunctionalized ureterocele. These include stone formation within it, or bladder outlet obstruction caused by the diverticulum-like defect associated with the ureterocele.A ureterocele occasionally prolapses from the urethra, causing total urinary obstruction.

Urethral valves

Most urethral valves are an accentuation of folds seen in the normal male urethra. The principal types of urethral valves were described more than 70 years ago (Fig. 10). The clinical picture in boys with urethral valves varies with the degree of obstruction caused. In some boys this obstruction is minimal and allows the bladder to empty normally, although at slightly higher pressures than normal. At the other end of the spectrum is high-grade obstruction with extreme hydronephrosis and greatly compromised renal function. Symptoms in boys with valves include bed wetting, urinary infection, daytime wetting, urinary frequency, dribbling, and sometimes a poor

urinary stream. Curiously, a history of a poor stream is often not obtained, even in patients with high-grade urethral obstruction. Residual urine is not always found in boys with valves because any hollow viscus can compensate for partial obstruction by hypertrophy, which allows emptying despite obstruction, up to a point. The intravesical pressure in such cases may be 80 to 100 cmH_2O, which is three to five times higher than normal. Back pressure from obstruction causes inadequate development of the fetal kidney and results in persistent hydronephrosis, tubular ectasia, and dysplastic morphology in extreme cases.

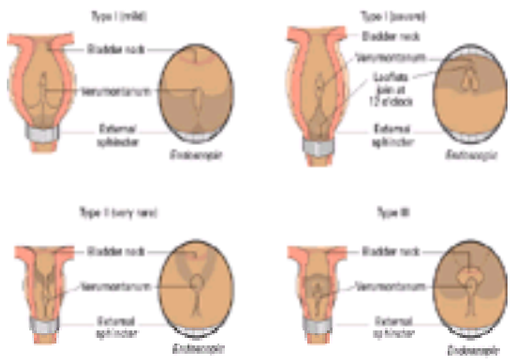


Fig. 10. Types of urethral valves. (a) Mild type I valves, common in boys with voiding disturbances and/or urinary infection if there is vesicoureteral reflux. (b) Severe type I valves in which the leaflets encircle the urethra and join at 12 o'clock, forming a diaphragm. There is usually hydronephrosis. (c) Type II valves, which are extremely rare and seldom cause obstruction. (d) Type III valves with a diaphragm. These are usually located at the level of the verumontanum, but can be located at any level of the urethra. Some valves are hybrid in appearance, that is, intermediate between type I and type III.

The diagnosis of urethral valves is made by voiding cystourethrography (Fig. 11). Even in the newborn baby, satisfactory voiding films can be obtained by slowly filling the bladder through a small plastic urethral catheter or through a needle inserted suprapubically, obtaining spot films when the patient voids. Treatment of boys with urethral valves depends on the age of the patient, whether there is impaired renal function, and whether the patient has infection and uremia at the time of presentation. An ill, acidotic, uremic neonate is often best treated initially by passing a small, non-reactive, plastic catheter to drain the bladder, and treating infection and acidosis with appropriate antibiotics and fluid replacement.

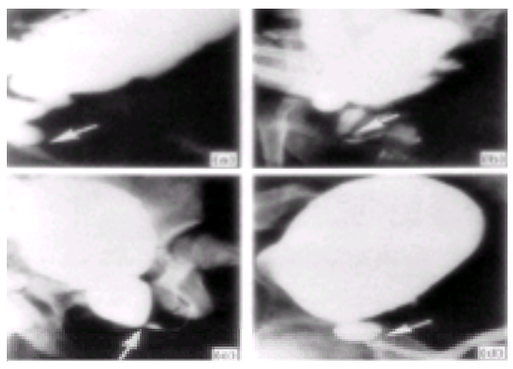


Fig. 11. Voiding cystograms of four boys with severe valves (arrows). Note the variable appearance of the urethra in these four patients. (a) Three-month-old boy with urosepsis. The bladder neck is prominent and there are cellules of the bladder trigone reflecting chronic obstruction.(b) Five-year-old boy with urethral valves. Note widening and elongation of the prostatic urethra, hypertrophy of the bladder neck, and massive reflux. (c) Eight-year-old boy with bed wetting but no other symptoms. There is enormous widening of the prostatic urethra but its shape is different from the two previous cases. (d) Type III urethral valve, a diaphragm obstructing at the level of the verumontanum in a 5-year-old boy with bed wetting, day wetting, and urinary frequency. There is no reflux, the bladder is not trabeculated, and there is good filling of the penile urethra during voiding. Symptoms were cured by endoscopic incision of the diaphragm.

Endoscopic relief of valve obstruction can be accomplished in all but the smallest premature babies, whose urethra is too small to accept a 7.5 Fr endoscope. In this circumstance, a temporary vesicostomy to drain the bladder may be advisable. Valves can be incised using a small wire electrode, taking care not to touch the nearby urethral sphincter, the verumontanum, or the bladder neck (Fig. 12). When obstruction is mild and the upper urinary tract is normal or near normal, valve resection alone will usually suffice. Vesicoureteral reflux is common in these patients. It may disappear when normal intravesical voiding pressures are achieved by relieving the urethral obstruction. Spontaneous improvement of hydronephrosis may be observed following relief of obstruction (Fig. 13). When there is massive vesicoureteral reflux or ureteral obstruction without reflux despite valve ablation, a major reconstruction may be required to preserve renal function. The newborn kidney has a creatinine clearance capacity about 20 per cent of that seen in the adult. Normally, clearance increases in the first year of life to attain adult clearance values of about 100 liter/m^2 body surface area per day. There is some evidence to suggest that early relief of obstruction allows recovery of maximum function in an obstructed renal unit. Late relief of obstruction can give a disappointing degree of improvement in renal function of that kidney. This fact has led us to pursue an aggressive attack on massive reflux or ureteral obstruction in infant boys with seriously compromised upper tracts (Fig. 14).

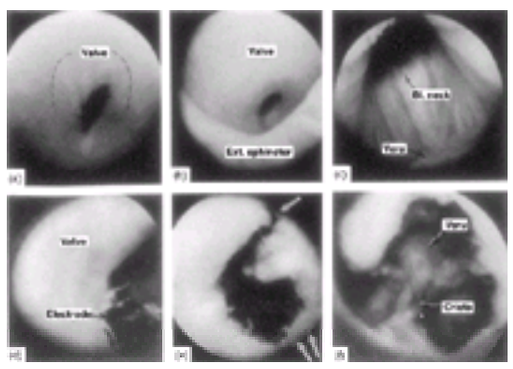


Fig. 12. Endoscopic resection of severe urethral valves in a 6-year-old boy referred to the endocrinology service for ‘diabetes insipidus'. Symptoms included urinary frequency, day and night wetting, and polyuria. There was severe hydronephrosis and massive reflux secondary to severe urethral valves. (a) Close up view of urethral valve, formed by confluence of valve leaflets. (b) Endoscope pulled slightly more distally to show position of the external sphincter with relation to the urethral valve. At 6 o'clock the sphincter is clearly apart from the base of the valve leaflets which arise from the crista urethralis. At 12 o'clock, however, the dorsal confluence of the valve leaflets lies very close to the external urethral sphincter.(c) Moderately hypertrophied bladder neck. (d) Cutting electrode visible in foreground. (e) The 12 o'clock cut is visible (arrow) as well as the cut in the valve leaflet at 4 to 5 o'clock (arrows). The right leaflet is still intact.(f) After incision of the right leaflet. Note clover leaf opening where the valve had been located. There is a clear view of the verumontanum. The external sphincter and bladder neck remained intact.

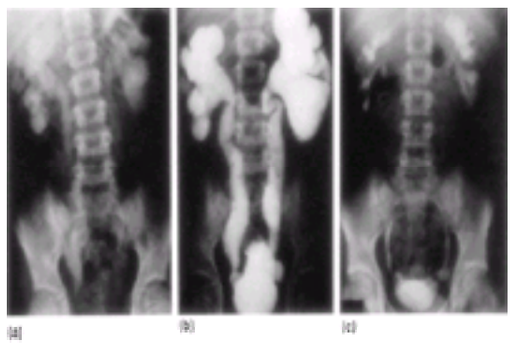


Fig. 13. Nine-year-old boy with severe urethral valves and hydronephrosis that improved after endoscopic resection of urethral valves. Clinical symptoms included daytime and night-time urinary wetting and a less forceful urinary stream than normal; no infection. (a) Intravenous pyelogram age 8 years showing bilateral hydronephrosis and poor visualization of contrast medium. (b) Bilateral percutaneous antegrade perfusion of contrast into kidneys to visualize upper tracts and measure perfusion pressures. The differential pressure between the kidney and bladder was normal, ruling out lower ureteral obstruction. There was no vesicoureteral reflux. Therefore, the valves were destroyed and no ureteral surgery was performed. The initial serum creatinine of 1.1 mg/100 fell to 0.8 mg/1001 week after the valves were destroyed. (c) Intravenous pyelogram 5 years later showing marked spontaneous improvement of hydronephrosis, normal appearing ureters, and better concentration of contrast medium. Bladder diverticula are evident, but there was no vesicoureteral reflux. Now 20 years later, the patient is well. This patient illustrates marked improvement in the urinary tract which can occur after endoscopic resection of severe valves.

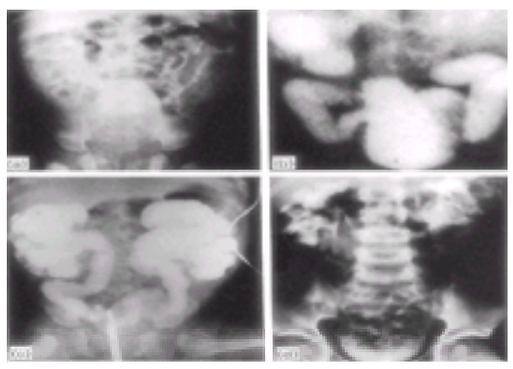


Fig. 14. Radiographs of 2-month-old infant with severe urethral valves, uremia, and acidosis, treated by resection of valves and immediate reconstruction 1 week after admission to the hospital. (a) Intravenous pyelogram before admission showing poor visualization bilaterally.(b) Cystogram with massive bilateral reflux and hydronephrosis. If valves are likely, a cystogram should be done first because if reflux is present the upper tracts will be visualized, obviating the need for an intravenous pyelogram. (c) Retrograde pyelogram taken at time of reconstructive procedure. The baby was first treated with drainage of the bladder through a small plastic catheter *per urethra*, intravenous fluid, bicarbonate, and antibiotics, which greatly improved his clinical state. At endoscopy, ureteral orifices were massively dilated. It was believed that with massive reflux and this degree of hydronephrosis it would be safest to repair the ureters rather than wait. (d) Intravenous pyelogram age 6 years. This patient is now aged 27 years. He is 6ft 4in tall, weighs 200 pounds and has a stable intravenous pyelogram. The serum blood urea nitrogen is 9 mg/100, and serum creatinine 1.4 mg/100. Renal function is two-thirds of normal, measuring 66 l/m^2 body surface area.

Even with aggressive therapy about one-third of these boys develop renal failure because they lack the renal reserve to support the increased body muscle mass attained at or before puberty. Prenatal relief of urinary obstruction has been attempted by intrauterine placement of drainage catheters or ureterostomy to divert the urine. Results to date have been disappointing, however. We believe that the principal benefit of antenatal identification of urinary obstruction is the knowledge of its presence, which allows therapy to begin immediately after birth and before urinary infection occurs. Some boys with severe urethral valves develop a thick-walled non-compliant bladder, which may be improved by augmentation of the bladder with bowel. A urethral valve-like obstruction is rarely encountered in girls. These valves are best demonstrated by passing a malleable probe bent into a hook. The probe is passed through the urethra alongside the endoscope, hooking the leaflets and pulling them into view out of the urethral orifice.

Ureteral reimplantation

Reimplantation of the ureter has been described briefly in association with correcting vesicoureteral reflux, which is the principal cause of pyelonephritis in children. Ureteral reimplantation is also necessary in patients with ureterocele and ureteral ectopia, and in some of those with urinary incontinence or tumors. Whatever the indication, the general principle remains the same; the ureter must be joined to the bladder or a bowel augmentation in such a fashion that there is a submucosal tunnel which prevents backwash and does not permit angulation of the ureter when the bladder fills. Prevention of reflux, whichever technique is used, requires a tunnel length to ureter diameter ratio of about 5:1. The most common mistakes made when reimplanting a ureter are angulation of the ureter where it re-enters the bladder and devascularization of the ureter by dissecting too close to it during mobilization. These errors result in fibrosis and obstruction.

Transureteroureterostomy

This is a useful procedure for achieving drainage when two satisfactory ureteral reimplantations cannot be performed or when one ureter is too short to reach the bladder. The ureter can be drained across the abdomen retroperitoneally and joined to the opposite ureter (Fig. 15). Devascularization of the ureter to be drained can be avoided if it is mobilized with all of its periureteral tissue, and often with the gonadal vessels adjacent to the ureter. Infarction of an ovary or testicle following ligation of the ovarian or testicular vessels in the lower abdomen and their retention with the ureter to be mobilized is rare. It is important to avoid angulating the ureter beneath a mesenteric vessel; there must be no tension on the anastomosis.

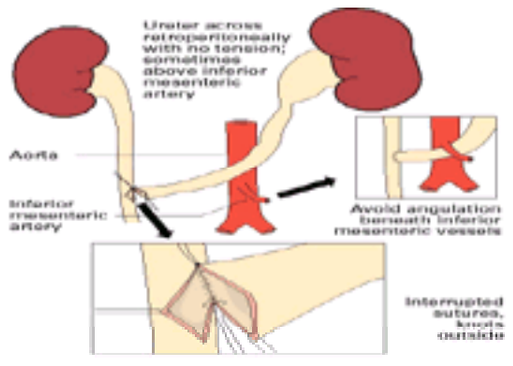


Fig. 15. Technique of transureteroureterostomy.

Bladder augmentation

Increasing the size of the bladder was originally used for the treatment of vesical contracture secondary to genitourinary tuberculosis. The bladder can be augmented using colon (Fig. 16) or a similar segment of small bowel. The bowel must be reconfigured and not used as an intact segment, which can cause incontinence if pressures produced by peristaltic waves exceed the resistance of the bladder neck and external sphincter. Cecum can also be used (Fig. 17), the terminal ileum being inverted and sewn to the adjacent wall of the cecum as a type of nipple to prevent reflux. This method is only used when neither ureter can be tunneled into the bowel

wall for reimplantation. When the bladder is augmented, it must be opened widely, to avoid the augmented segment behaving as a diverticulum and trapping urine.

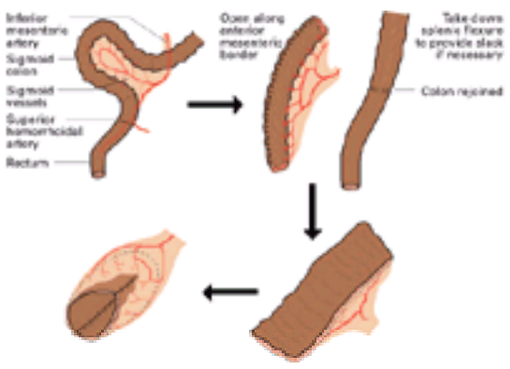


Fig. 16. Use of colon to augment the bladder. The bowel must be reconfigured to prevent colonic peristaltic activity which can exert high pressures in the augmented bladder, causing leakage.

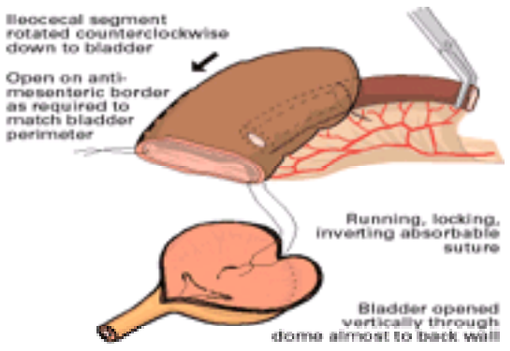


Fig. 17. Augmentation of the bladder using a cecum with intussuscepted and fixed terminal ileum to prevent reflux. Note wide opening of the bladder to prevent augmentation from functioning as a diverticulum.

Stomach has increasingly been used for augmentation of the bladder (Fig. 18) and has certain metabolic advantages over the small bowel and colon. Small bowel and colon both absorb chloride and excrete fixed base, aggravating metabolic acidosis in patients with poor renal function. Gastric segments, however, actively excrete chloride and can actually improve acidosis in some patients. A patient with a gastrocystoplasty, however, can develop hypochloremic alkalosis rapidly, with vomiting and consequent inadequate salt intake. Patients in whom small bowel and colon are used for augmentation have mucus in the urine, which may serve as a nidus for stone formation. This complication has not been seen in patients with a gastrocystoplasty. Bacilluria is more easily controlled in patients with gastrocystoplasty: the urine tends to have an acid pH, which discourages growth of bacteria. Some of these patients have complained of pain following bladder augmentation by stomach. This is sometimes relieved by administration of H_2 receptor antagonists.

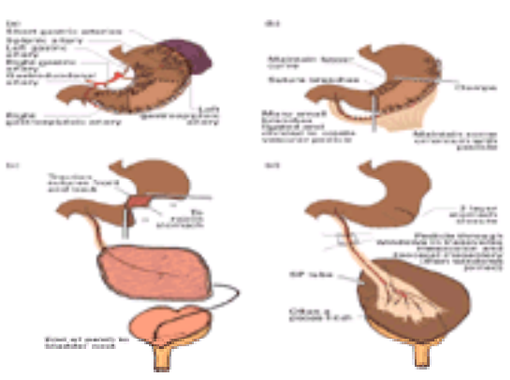


Fig. 18. Use of stomach to augment the bladder.

Bladder augmentation has made urinary tract reconstruction possible in many patients who were previously treated by urinary diversion. Augmentation provides adequate urine storage capacity and bladder compliance. The ability to empty the bladder spontaneously is governed by the degree of neurologic control of the bladder neck and/or external urethral sphincter. If the patient has neurologic impairment, due to conditions such as spina bifida and myelomeningocele, or if there is a fixed resistance at the bladder outlet as is seen in many patients following bladder neck narrowing procedures, spontaneous voiding may be impossible. Any patient for whom augmentation is planned should understand that intermittent self-catheterization may be required to empty the bladder. Failure to empty the augmented bladder leads to overdistension and to spontaneous perforation.

Psoas hitch

Hitching the bladder to the psoas muscle (Fig. 19) prevents angulation of the ureter when the bladder fills. It also allows longer tunneled ureteral reimplantation into the bladder to be achieved.

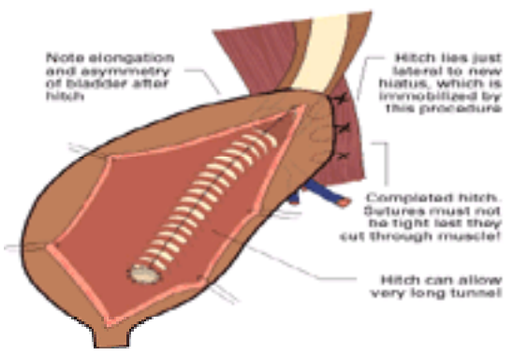


Fig. 19. Psoas hitch. This technique allows a longer ureteral tunnel to be made. It also fixes the ureteral hiatus so that when the bladder fills it does not allow angulation of the ureter at its point of entry into the bladder.

Bladder neck repair

The normal bladder outlet has a bladder neck or internal sphincter, which opens involuntarily during voiding, and an external, voluntary urethral sphincter. When normally innervated these two sphincters provide continence. Generally, continence is still possible after the loss of one sphincter, but not following the loss of both. For example, fracture of the pelvis in males often disrupts the voluntary external sphincter, yet the patient can usually maintain continence with the bladder neck. The bladder neck may be abnormally wide or absent in patients with some congenital malformations, including congenital epispadias and ectopic ureters (Fig. 20). Both sphincters are often deficient, particularly in those with complete epispadias or bladder exstrophy. Outlet resistance can be improved by enlarging the abnormal bladder outlet and urethra (Fig. 21). This exposure can be improved by splitting the pubic symphysis. The ureters are usually moved upward in the bladder and are reimplanted by cross-trigone advancement. This technique allows mucosa to be removed from the wide bladder outlet on both sides and a strip of mucosa to be retained in the center. A tube is created from this and muscle is closed over it to create a new bladder neck with greater outlet resistance.

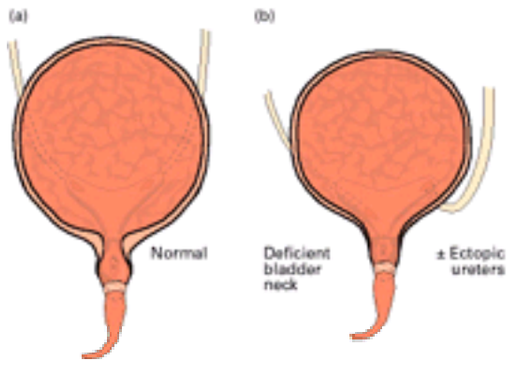


Fig. 20. (a) Normal bladder neck anatomy and (b) deficient bladder neck which may be seen in patients with ureteral ectopia, congenital epispadias, and other malformations.

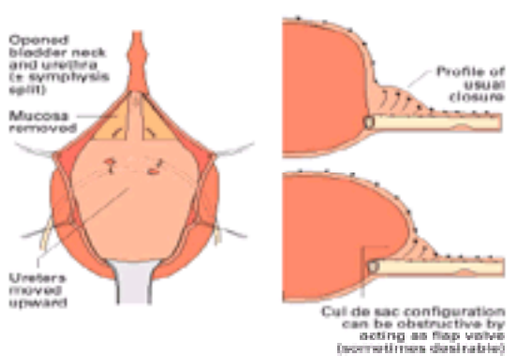


Fig. 21. Narrowing of bladder neck to create more outlet resistance. This procedure usually requires moving the ureters to a higher point in the trigone. The bladder is often augmented in conjunction with a narrowing procedure when the bladder is too small or non-compliant.

Urinary diversion

Diversion of the urinary tract, both temporary and permanent, is being performed less frequently today than in the past as emphasis is put on the reconstruction of malformations and intermittent catheterization for bladder emptying. Temporary loop cutaneous ureterostomy or pyelostomy (Fig. 22) was often used to drain the kidneys in boys with urethral valves. Fortunately, this practice has largely waned. End cutaneous ureterostomy has also been used extensively. This often does not provide good drainage and makes final reconstruction difficult.

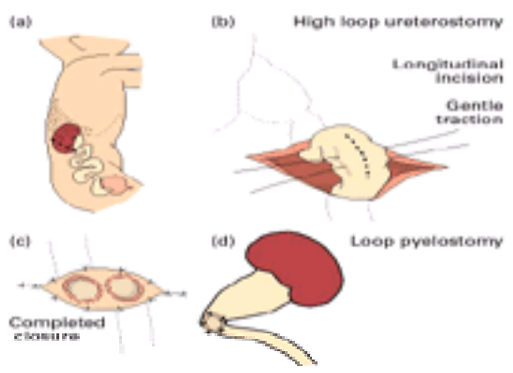


Fig. 22. Temporary urinary diversion by high loop ureterostomy or pyelostomy. The authors believe this operation has been used far too often in the past.

The ileal loop urinary diversion was introduced as a means for draining the urinary tract in adults undergoing cystectomy for bladder cancer. It was also widely used in children with various obstructive uropathies in whom reconstructive surgery had not been successful. The ileal loop was a simple segment of small bowel into which the ureters were joined end to side. This was then extended to the surface where urine was collected in an external appliance worn on the abdominal wall. Reflux in these loops results in ascending pyelonephritis, stones, strictures, and renal failure, and the non-refluxing colon conduit was therefore introduced (Fig. 23). Non-refluxing tunnels can be constructed more effectively in the colon wall than in the wall of the small bowel. Although the ileal loop urinary diversion is still used widely in many patients with bladder cancer, especially when life expectancy is short, it is not suitable for young patients with a long life ahead of them. In the rare circumstances when a conduit is required, a non-refluxing technique should be used. A conduit can also be constructed using cecum, following preparation as for the cecal bladder augmentation (Fig. 17).

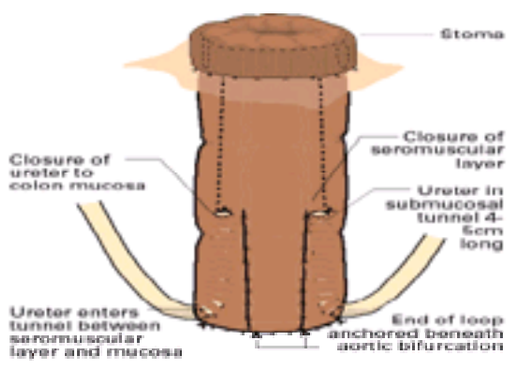


Fig. 23. Urinary diversion by colon conduit, incorporating non-refluxing anastomoses of the ureters into the conduit.

Cutaneous vesicostomy is a form of urinary drainage used when long-term, but temporary, drainage of the lower urinary tract is required. The opening in the bladder must be high in its dome, where it will drain adequately. An opening in the front wall of the bladder maybecome occluded by prolapse of the back wall of the bladder against the cutaneous opening.

Chronic drainage of the bladder by an inlying suprapubic tube was a common surgical procedure several decades ago and results in chronic bladder infection, stones, and fibrosis with contracture of the bladder. Short-term percutaneous tube drainage of the bladder is useful after reconstructive bladder outlet surgery or pelvic trauma with urethral disruption.

Continent diversion

In recent years, continent urinary diversion has been employed in adult patients following radical extirpative pelvic surgery for malignancy, and in pediatric patients whose bladders were considered to be non-reconstructible. This technique involves the creation of an internal reservoir to hold a volume of urine approximately equal to that of normal bladder capacity for that age, preventing reflux to the kidneys, and preventing leakage at the abdominal stoma. Emptying is accomplished by intermittent catheterization of the continent reservoir, several types of which are in vogue (Fig. 24, Fig. 25, and Fig. 26). Complications common to these pouches include chronic low-grade bacilluria, stone formation, leakage, solute resorption, and spontaneous perforation if the pouch is not emptied regularly. In the past, too many urinary diversions were performed for reconstructible urinary tract disease and it is possible that more continent diversions are currently being performed than may be necessary. Certainly, when functional repair is possible, any type of urinary diversion, continent or otherwise, is a poor second choice.

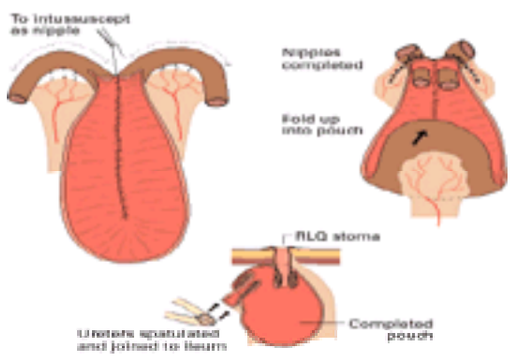


Fig. 24. Continent urinary diversion using the Kock pouch.

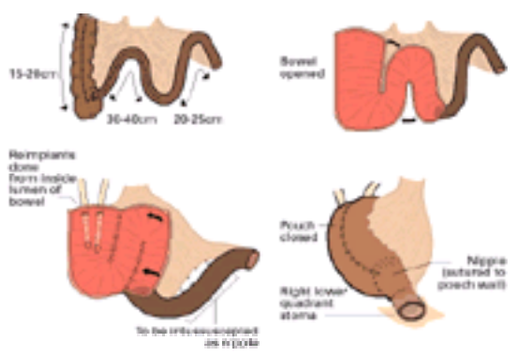


Fig. 25. Continent urinary diversion using the right colon.

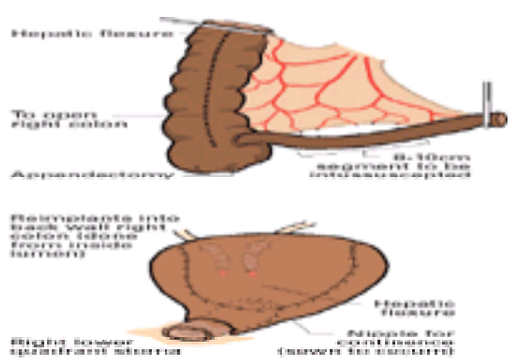


Fig. 26. Continent urinary diversion using the Mainz pouch.

Urinary undiversion

Many patients who underwent urinary tract diversion in the past are suitable for urinary tract reconstruction. Figure 27 shows the anatomy of an ileal loop urinary diversion and the surgical options that might be possible for reversal of this diversion. Obviously, this procedure depends on the length of the ureters and the condition of the bladder. Experience with nearly 200 reversals has emphasized several important concepts. First, the duration of a previous diversion is not important: reconstruction can be achieved in patients who have worn a bag on their abdomen for nearly three decades. Second, poor urinary tract function is not a contraindication to reconstruction: patients with urinary diversions should be re-evaluated in the light of the surgical advances made since their original diversions were performed.

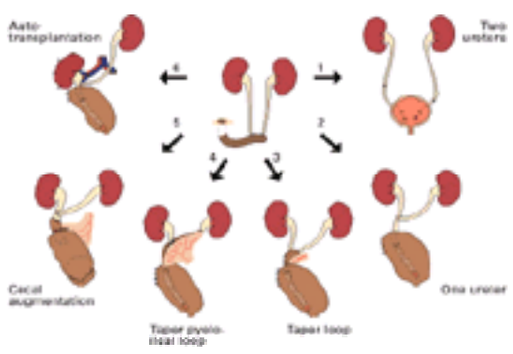


Fig. 27. Options for reconstruction of a patient with an ileal loop.

Epispadias

Epispadias refers to a dorsal opening of the urethra—the opposite of hypospadias (Fig. 28). Like all malformations, epispadias occurs in various degrees of severity. In males there may be a simple cleft in the dorsum of the glans; in more severe cases a widely open urethra may extend back through the bladder neck, urinary incontinence resulting from the open external urethral sphincter and the internal urethral sphincter (the bladder neck). Although most malformations are more common in a mild form, epispadias is commonly severe, with complete urinary incontinence. Incontinence in these patients is restored by increasing bladder outlet resistance, often also moving the ureters upward to lengthen the urethra (Fig. 21). The wide urethra is closed. In the male, the penile shaft and the glans penis must be reconstructed. In females, the two clitori are reapproximated to restore a normal anatomic appearance. Some patients with epispadias require bladder augmentation.

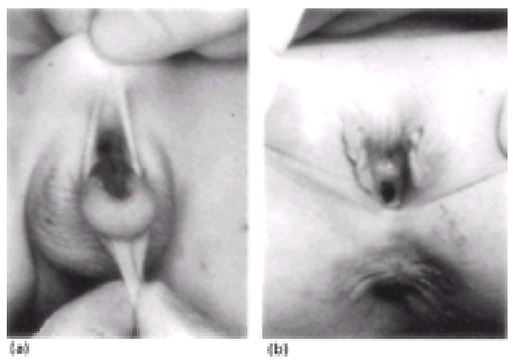


Fig. 28. Congenital epispadias. (a) Male and (b) female. The severe form of this malformation is more commonly associated with incontinence than minor degrees of epispadias.

Exstrophy of the bladder

This is one of the most devastating abnormalities in the newborn baby. If a sound were introduced into the normal urethra of the male or female, passing the sound up to the dome of the bladder, then cutting down upon the sound from the navel to the urethral meatus, exstrophy of the bladder with complete epispadias would be reproduced. This malformation is shown in Fig. 29.

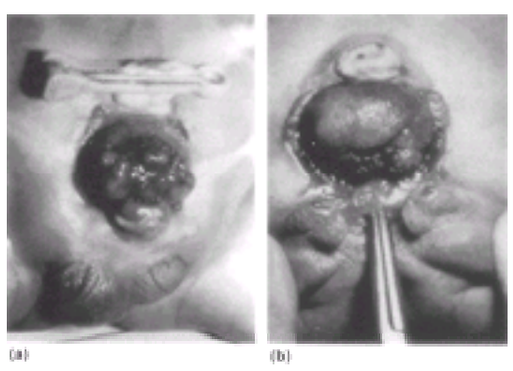


Fig. 29. Exstrophy of the bladder. (a) Male. Note epispadiac penis which is tethered upward toward base of bladder. (b) Female. Note surgical clamp in vaginal orifice just behind bladder base. There is a labium and hemiclitoris on each side.

Until the mid-1950s the majority of patients with exstrophy of the bladder were treated by internal diversion of the urine into the colon (ureterosigmoidostomy). Many patients succumbed to urinary infection ascending from the fecal stream to the kidney. There is also a 10 per cent risk of carcinoma of the colon developing at the site of the ureterosigmoidostomy anastomosis. Success can only be achieved by tunneling of the ureters into the colon (Fig. 23).

In the past three decades there has been increasing emphasis on functional reconstruction of the bladder in these patients. When the bladder size is adequate, relatively normal bladder function may be possible; however, for many patients the bladder is too small. Bladder augmentation with intermittent catheterization is appropriate for many patients with bladder exstrophy and urinary diversion is now rarely required. Surgical repair can produce functional genitalia with much better success than might be imagined possible (Fig. 30): this is generally performed in a series of carefully planned stages. The bladder is closed surgically as soon after birth as possible, before infection or metaplasia of the exposed bladder surface occurs. The wide apart pubic bones can be brought together with strong sutures in the first few days of life. After the first week an iliac or pubic osteotomy may be useful to approximate the halves of the pubic symphysis. Later the ureters are reimplanted, a bladder neck is reconstructed, and the epispadias malformation is corrected. Reoperation is often needed to gain greater outlet resistance if insufficient continence is achieved and augmentation is necessary if the bladder remains small. A series of operations producing an ultimate success is preferable to a urinary diversion. Recently, we have performed total reconstruction in neonates with satisfactory outcomes. This included ureteral reimplantation, bladder closure, bladder neck and epispadias repair.

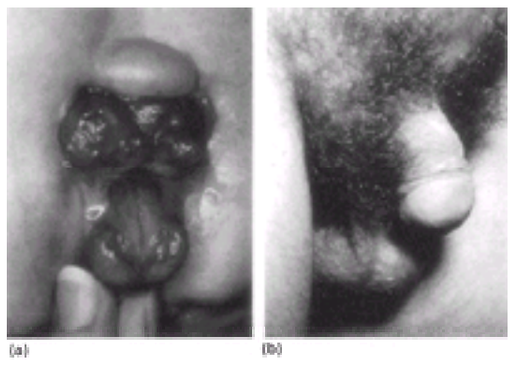


Fig. 30. A male with exstrophy of the bladder and complete epispadias.(a) At age 1 month. The bladder is rudimentary and the penis is being held downward to display its epispadiac appearance. (b) Age 16 years, following initial reconstruction of the penis as a baby, and secondary revision at age 15 years to lengthen it.

Artificial sphincters are now available. These consist of an inflatable cuff placed around the bladder neck or proximal urethra, a reservoir of fluid, and a mechanism for transfer of fluid under pressure from the reservoir to the cuff to provide outlet resistance. They do, however, have a high incidence of late complications, including erosion of tissues compressed by the sphincter. Such complications are particularly likely in patients who have undergone extensive surgery to the bladder neck and urethra, such as those with exstrophy and epispadias. The best indication for use of an artificial sphincter in young patients is a neurologically impaired bladder with

relatively normal tissues of the bladder neck and urethra.

Ectopic ureter

There are many ectopic positions of the ureter, all distal to the normal location of the interureteric ridge. In males, the ureter may empty into the bladder neck, a seminal vesicle, or even into the vas deferens (Fig. 31). Crossed ectopia refers to the condition where the ureter crosses the midline and drains a kidney on the opposite side. The ectopic ureter is always above the level of the external urethral sphincter. These patients, however, are not incontinent, and usually present with urinary infection. Paradoxically, an ectopic ureter in the bladder neck can show both obstruction and reflux: in the resting state the muscle of the bladder neck compresses the lower ureter and obstructs it; during voiding, when the bladder neck opens, the same ureter may exhibit massive reflux.

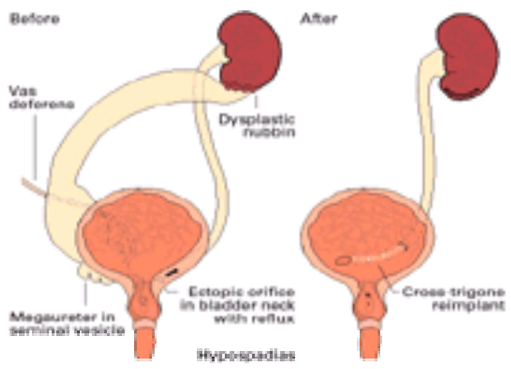


Fig. 31. Patient with complex ureteral ectopia. Note crossed renal ectopia, ectopic position of right ureter into seminal vesicle, and ectopic position of left ureter in bladder neck.

In females an ectopic ureter can be located in the bladder neck, the urethra, or just outside of the urethra. In the latter case constant moisture typically appears from the ectopic orifice. The renal tissue with an ectopic ureter is usually dysplastic and is not visualized by intravenous pyelography. If its presence is suspected, however, careful study of the renal shadow may disclose a small amount of non-functional tissue medial to the upper pole of the kidney. Intravenous contrast media such as indigocarmine or methylene blue dye is not concentrated sufficiently to reveal the blue dye at the ectopic orifice. The ectopic orifice can sometimes be located by observing the perineum after administration of a large water load. In the female an ectopic orifice may rarely be located in the vagina, cervix, or even the uterus. Removal of the dysplastic tissue and the ectopic ureter cures the associated poor continence. When an ectopic ureter is located within the bladder, the renal segment associated with it may have good function. Reimplantation of the ureter to prevent reflux is the proper treatment.

‘Bilateral single ectopic ureters' is a rare condition in which both ureters are located distally to the bladder neck. These patients also have an inadequate bladder neck and must be treated in the same way as patients with complete epispadias and incontinence.

Neuropathic bladder

Spina bifida and myelomeningocele are common in infancy and are frequently associated with neuropathic changes in the bladder. An apparently normal urinary tract can soon deteriorate if its innervation is faulty. Figure 32 shows a typical neuropathic bladder with irregular contour and bilateral vesicoureteral reflux: the upper urinary tract can deteriorate quickly when the bladder is non-compliant and allows reflux, resulting in infection.

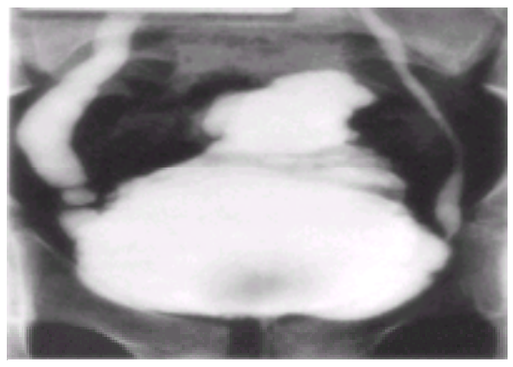


Fig. 32. Cystogram in a 14-year-old male with typical neuropathic bladder showing ‘Christmas tree' shape of bladder, severe trabeculation with cellular formation, and bilateral vesicoureteral reflux.

Many of these children were previously treated by urinary diversion, which cured incontinence but often at the expense of gradual deterioration of the kidneys. Today, more emphasis is placed on intermittent catheterization of the neuropathic bladder if it is of adequate size and its pressures are low. Many of these bladders, however, have high pressures and a small capacity and augmentation cystoplasty is required. Although ureteral reimplantation is frequently indicated, this is a technical challenge for the surgeon: the bladder wall is often thick and inflamed. Increasing bladder outlet resistance is needed to control poor continence in some patients. Urodynamic evaluation should be performed to measure their bladder volume, sensation, pressures, and outlet resistance.

Unfortunately, some children with spinal defects are confined to a wheelchair; they are unsatisfactory candidates for intermittent self-catheterization of the urethra. Creating a continent reservoir may offer the best means of achieving continence and of preserving the upper tracts with a catheterizable abdominal stoma.

Hypospadias

In hypospadias the urethral opening lies on the ventral aspect of the penis: the opposite of epispadias. This condition occurs in various degrees (Fig. 33), and is often associated with chordee, or ventral curvature of the penis. In minimal hypospadias the urethral meatus may be just a little proximal to the tip of the penis, without chordee. Perineal hypospadias represents the severe end of the disease spectrum, the urethral opening is at the base of the scrotum, which may be bifid, and the penile body is bent so that the glans is tethered to the perineum. The management of hypospadias has changed drastically in the past two decades. Early repair in the first year or two of life is now common practice. The goal is to straighten the penis, construct a urethra that extends to its tip, reshape the glans penis when it is abnormally splayed open, and provide a normal cosmetic appearance. The majority of these repairs can be achieved in a single operation.

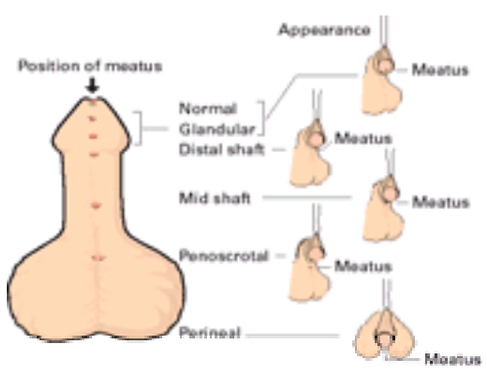


Fig. 33. Degrees of hypospadias.

A variety of techniques have been described and are in current use. The most severe cases may require a staged repair. At the first operation the penis is straightened and the skin is rearranged to cover the shaft, deferring urethroplasty until 6 to 12 months later. A urethra can be constructed using a graft of extragenital skin or bladder mucosa. The repair of hypospadias and chordee should not use skin of the perineum or scrotum, which will ultimately grow hair, since this predisposes to formation of stones in the adult. There is no reason today why a child born with hypospadias and chordee should not grow up with a functionally and cosmetically satisfactory penis if repair is performed by a surgeon conversant with modern surgical techniques. Hypospadias may occur in multiple male members of a family, as may vesicoureteral reflux. Severe hypospadias, where testicles are not palpable in the scrotum, may be an intersex state.

Undescended testicle

The treatment of cryptorchidism has changed in recent years. A testis that has not descended into the scrotum by 1 year of age seldom descends spontaneously. If it is located in the inguinal canal, or higher, it is subject to a higher than normal ambient temperature which impedes its development. Therefore, undescended testes are treated much earlier than was customary two decades ago. The testis and its spermatic cord are mobilized in the inguinal canal through an inguinal incision, separating cremasteric muscle fibers. In the majority of patients there is an inguinal hernia sac that is dissected free from the cord and is ligated at a high level. The spermatic cord is usually then dissected up into the retroperitoneal region to obtain enough length of spermatic vessels and vas deferens to bring the testicle into the scrotum without tension.

Some undescended testes are abnormal and have failed to descend *in utero* for that reason. Orchidopexy will not improve the long-term prognosis for testicular function in such patients. The classic example of this is the prune belly syndrome, in which the testes are located intra-abdominally and uniformly show abnormal development microscopically. An impalpable testis is often non-functional as a result of torsion *in utero*. The spermatic vessels and vas deferens will be found to end blindly, usually at the level of the internal inguinal ring. A prosthetic testis ensures that the child will not grow up with an empty scrotum: the prosthesis can be upgraded to adult size at puberty. There is a slight increase in the incidence of malignancy in patients with cryptorchidism, and this does not seem to be altered by orchidopexy.

Intersex states

The birth of an infant with ambiguous genitalia should be regarded as an emergency; it is a social disaster for parents not to know whether their new baby is a boy or a girl. Evaluation of these babies includes family history, physical examination, cytogenic characterization, radiographic examination, biochemical evaluation, and, sometimes, endoscopy and laparotomy with gonad biopsy. Figure 34 shows a scheme of genital development. If the undifferentiated gonad becomes a testis it secretes müllerian-inhibiting substance, causing the müllerian ducts to involute. Simultaneously, androgen from the testis stimulates the wolffian ducts to differentiate as male. If the undifferentiated gonad becomes an ovary or if there is an absent gonad, the wolffian ducts disappear and the müllerian duct structures develop. In experimental animals these phenomena are unilateral: a testis on one side causes the expected male development on that side, whereas an ovary or no gonad on the other side results in female development of those structures. Simultaneously, the external genitalia develop as male or female, depending on the presence of androgen and androgen receptors in the genital anlagen.

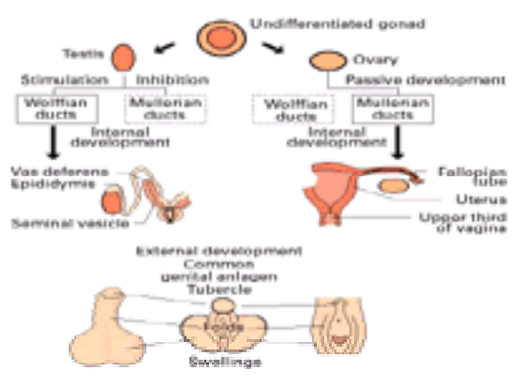


Fig. 34. Genital development (after Federman).

Some of these conditions are familial, especially the adrenogenital syndrome with salt loss. Affected infants have adrenal hyperplasia and, in the severe form, have the appearance of a male with a well developed phallus, but an empty scrotum. There is a wide spectrum in the degree of masculinization of the genitalia in these babies. An infant with the salt-losing syndrome may develop adrenal crisis with shock, usually at about 1 week of age. Immediate administration of cortisone and electrolyte solution is life saving. Some of the intersex states are inherited as sex-linked recessive traits. On physical examination an attempt should be made to feel the gonads. The immature uterus can be felt, if present, on rectal examination. Chromosome study should be performed. However, the most important fact in guiding gender choice is the anatomy with which the surgeon must deal. The surgeon should enlist the help of a pediatric endocrinologist in evaluating these babies. The decision for gender assignment is a group decision based on the information that can usually be gathered in 3 to 4 days. If there is no phallus, a genetic male should not be raised as a male, since an anatomically inadequate male is a social disaster. A child reared as a female is more able to cope socially even if anatomic structures are incomplete. Feminizing genitoplasty and vaginal construction are surgically possible.

Cystourethrography can prove useful in delineating the anatomy of the lower urinary tract and disclosing whether a vagina and cervix are present (Fig. 35). Biochemical testing includes determination of 17-keto steroids, and 17-hydroxycorticoids to detect adrenocortical hyperplasia syndromes. Levels of serum testosterone, pituitary gonadotropin follicle-stimulating hormone, and luteinizing hormone should also be determined.

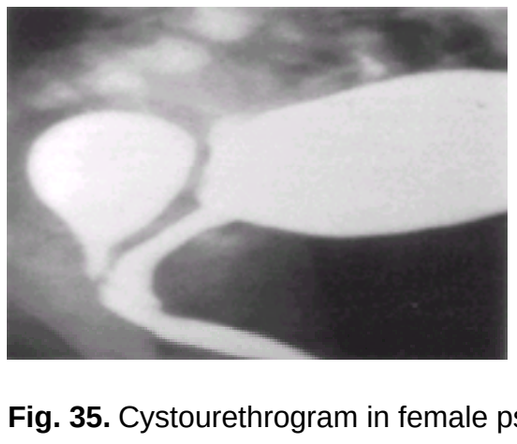


Fig. 35. Cystourethrogram in female pseudohermaphroditism with vagina communicating with urethra just above the external urethral sphincter.

The adrenogenital syndrome is the most common condition presenting with genital ambiguity. These children are genetically normal females with normal ovaries, fallopian tubes, uterus, and upper vagina. The external genitalia, and sometimes the lower urinary tract, show masculinization. In extreme cases there may be a well formed ‘penis' with a urethra which comes to its tip; in these the vagina will communicate with the urethra between the external urethral sphincter and the bladder neck.

Mixed gonadal dysgenesis is the second most common type of intersex state. The internal and external structures are variable. The characteristic karyotype is 45,X/46,XY, but there are several chromosome variants in this condition. There is usually a dysplastic testis on one side and a streak ovary on the other. Gonads

should eventually be removed when the Y chromosome is present since malignant gonadoblastoma may develop in later years.

There are several other male and female pseudohermaphroditism states which must be considered. One of the most interesting is the so-called feminizing testis syndrome. These patients have a male chromosome karyotype and intra-abdominal testes but a phenotypically female appearance. Sometimes the diagnosis will be made when repairing an inguinal hernia in an apparently female child upon finding a testis in the hernia sac.

Surgical treatment of intersex states usually entails feminizing genitoplasty. An enlarged phallic structure should be recessed, not amputated as was customary a generation ago. Vaginoplasty techniques allow a high vagina to be exteriorized, or a vagina to be created. Figure 36 shows the genitalia of a child with true hermaphroditism before and after feminizing genitoplasty.

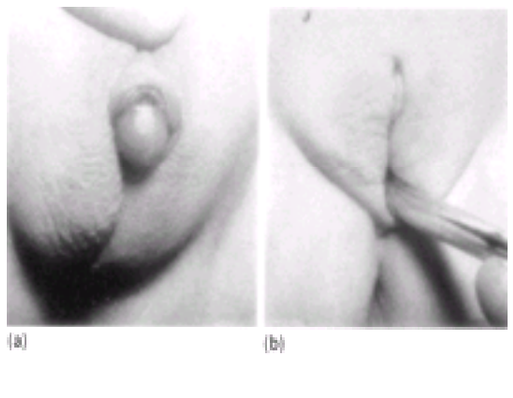


Fig. 36. True hermaphroditism. (a) Genitalia during infancy. There is a well formed phallus. On the right there is a dysplastic testis in the scrotum and on the left an ovary in the pelvis. (b) Postoperative appearance 2 years after genital reconstruction. This repair consisted of subtotal clitorectomy with recessing of the glans, exteriorizing the vagina into which a sound is being passed, and reconfiguring the ‘scrotum' to form labia majora.

There is clearly male psychologic ‘imprinting' in some children with the Y chromosome who have been raised as female. Some genetic females subjected to high levels of testosterone *in utero* also show male behavioral characteristics.

Cloacal malformations

Cloaca is the Latin word for sewer. Anatomically it denotes confluence of the urinary, genital, and gastrointestinal tracts into a common chamber. The 7.5-mm human embryo passes through a cloacal stage. Subsequently, the urorectal septum descends, dividing the structure into a posterior rectal passage and an anterior urogenital sinus. Many aberrations can occur during this complex development, but their anatomy should not be confused with intersex states; they are female babies with imperforate anus and a persisting urogenital sinus (Fig. 37, Fig. 38). Laparotomy is required at birth to vent the colon, preferably by a right-transverse divided-loop colostomy. Decompression of the obstructed urinary tract can usually be accomplished by intermittent catheterization of the urogenital sinus. Drainage of the urine-filled vagina prevents the development of hydronephrosis and urinary tract infection. In the majority of these patients the defect can be successfully repaired by a posterior sagittal approach; some also require laparotomy.

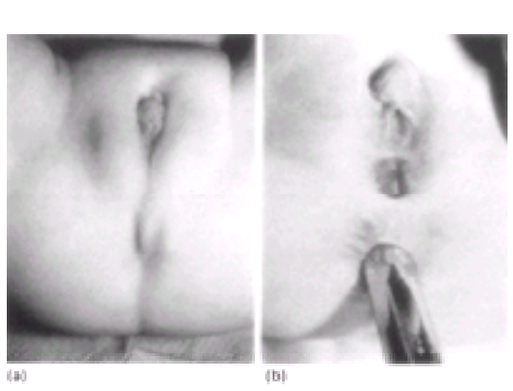


Fig. 37. Cloacal malformation. (a) Preoperative appearance as an infant. Note imperforate anus and no vaginal opening. Colostomy was performed at birth. Intermittent catheterization was used during the first year to empty urine from the vagina. (b) Postoperative appearance. Note Hegar dilator in anus, perineal body which has been created between anus and vagina, and separate urethral opening beneath the clitoris.

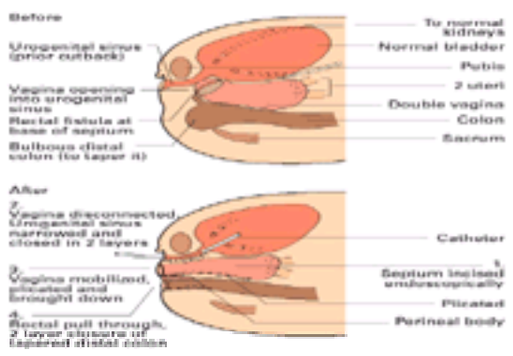


Fig. 38. Anatomy of patient shown in Fig. 37. (a) Preoperative showing high confluence of the vaginas into the urogenital sinus and colon fistula near vaginas. (b) After reconstruction, which was done from the posterior sagittal approach with separation and pullthrough of the vagina and rectum. It was not necessary to enter the abdomen.

Cloacal exstrophy

This rare condition, which occurs only once in 250 000 births, was once uniformly fatal. Today a satisfactory surgical outcome can usually be accomplished. Magnetic resonance imaging of the spine should be performed in these babies because all have a tethered spinal cord. Some will benefit by neurosurgical release of the tethered spinal cord to avert progressive neurologic sequelae. A typical patient with cloacal exstrophy has omphalocele, just below which is a bowel opening, which is typically the cecum (Fig. 39). The terminal ileum may prolapse out through the ileocecal junction as an ‘elephant trunk'. On each side of the bowel fistula is characteristically an exstrophic hemibladder, and the anus is imperforate. Although half of these infants are boys and half are girls, it may be difficult to determine the gender if the observer is not familiar with the condition. Boys usually have a testis in widely separated scrotal sacs, and two rudimentary penises which are far apart. It is seldom appropriate to rear a child with this malformation as a male, since males rarely have anatomy from which a satisfactory penis can be made. As shown in Fig. 40 the colon is usually rudimentary, hanging down in the pelvis as a blindly ending tube. This rudimentary colon should never be discarded: it has the ability to grow and function as a colon if given the chance.

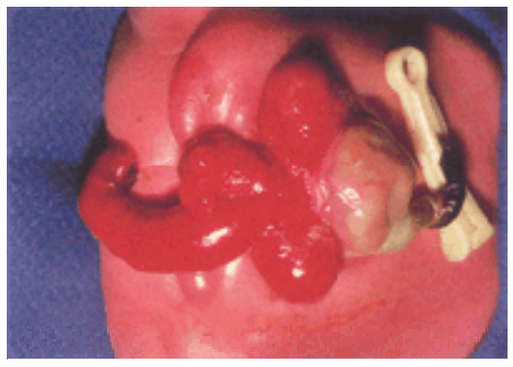


Fig. 39. Exstrophy of the cloaca in a girl. Note omphalocele, two hemibladders, and bowel prolapsing through the cecum which is located between the two small bladders.

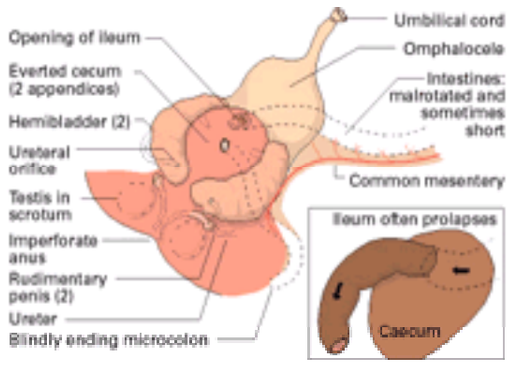


Fig. 40. Anatomy of cloacal exstrophy in the male.

The omphalocele is repaired at birth, and the cecum and terminal ileum are separated from the hemibladders and closed, thereby producing a functional colon, the end of which can be exteriorized as a temporary end colostomy. Because the colon is small, a decompressive temporary ileostomy may be desirable to protect the cecal closure from leakage. An immediate pullthrough of the colon may be considered if the baby does not have a caudal meningocele. The two hemibladders are joined and closed, as for any newborn with exstrophy of the bladder. Further urinary tract reconstruction must be undertaken later, at the age of 1 to 2 years, when a catheterizable urinary reservoir is created and the bladder is augmented.

Tumors

Wilms' tumor, the most common solid tumor of childhood, is a malignant tumor of the kidney that often presents as a large palpable, abdominal mass in a young child (Fig. 41). It may also present as hematuria. The diagnosis is usually made by intravenous pyelography, which discloses a renal mass with excretory function, but with distorted caliceal architecture. The extent of these tumors should be assessed by CT scanning: they may extend into the renal vein and along the vena cava. Occasionally Wilms' tumor occurs bilaterally.



Fig. 41. Large left Wilms' tumor removed via thoracoabdominal approach in a 1-year-old male.

The mainstay of therapy is radical surgical extirpation. However, the tumor is sensitive to chemotherapy (especially actinomycin D and vincristine), and also usually responds well to radiation therapy, which is used in selected patients. The current overall cure rate of patients presenting with Wilms' tumor is approximately 85 per cent, a figure that includes some patients who have pulmonary metastases. Other renal tumors that rarely occur in children include angiomyolipoma, renal cell carcinoma, transitional cell carcinoma, and mesoblastic nephroma.

Neuroblastoma is the second most common solid tumor of childhood. It commonly occurs in the adrenal glands and it can arise from neural crest cells at any level adjacent to the spine, from the neck to the pelvis. Afflicted children with a large tumor frequently appear pale and cachectic. Although some neuroblastomas are encapsulated and surgically removable, many are locally infiltrative and defy surgical excision. Chemotherapy is also less effective for neuroblastoma than for Wilms' tumor. In general, the prognosis is better when neuroblastoma is discovered in the first year of life rather than later in childhood. One special category of neuroblastoma is characterized by a small primary in an adrenal gland and diffuse metastatic disease through the liver. Although this situation would seem hopeless, there is a high rate of spontaneous disappearance of the tumor.

Rhabdomyosarcoma is the most common malignant tumor seen in the genital tract of children. It occurs in the bladder, vagina, uterus, and in the prostate in the male. Once uniformly fatal, rhabdomyosarcoma of the genitourinary tract can now be cured with a combination of chemotherapy, radiation therapy, and surgical excision. ‘Sarcoma botryoides' is a term used to describe the gross appearance of this tumor as it presents in the genitourinary tract: typically the tumor has the appearance of a bunch of grapes with multiple fronds of tissue (Fig. 42). Functional reconstruction of the urinary tract is possible in some of these patients, despite the presence of large tumors which need major surgery for removal (Fig. 43).



Fig. 42. Typical grape-like mass of sarcoma botryoides protruding from the vagina of a 1-year-old baby.

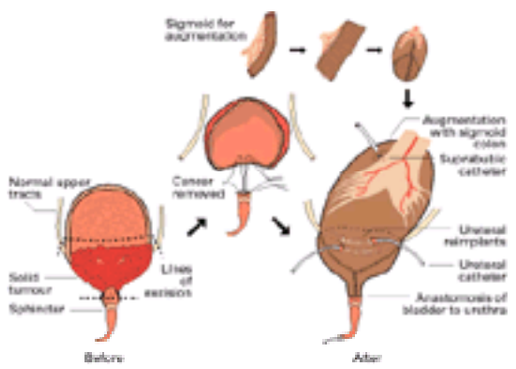


Fig. 43. Surgical treatment of rhabdomyosarcoma of prostate and bladder in an 11-year-old boy. A partial response was obtained to 6 months of chemotherapy and radiation therapy. Two years later a second operation was needed to achieve complete urinary continence; this included splitting the symphysis pubis, narrowing the bladder outlet, and adding a second augmentation using stomach. There is no residual tumor today and the patient is currently dry. He empties the bladder by intermittent self-catheterization.

The treatment of gonadal malignancies in children is based on the same principles as those applied to adults. A testicular mass in a child should not be examined by biopsy through the scrotum. As in the adult, an inguinal incision is made, isolating the spermatic cord, occluding it with a non-crushing clamp, and delivering the testis for inspection. If the tumor is malignant the cord is divided high and the testis is removed. Evaluation of these patients includes CT scan of the retroperitoneum and chest, and measuring levels of tumor markers, human chorionic gonadotropin, and a-fetoprotein. If lymph nodes are involved, as is common in patients with rhabdomyosarcoma, retroperitoneal lymphadenectomy should be performed as a secondary procedure, usually in conjunction with chemotherapy.

Most ovarian masses in young girls are benign cysts, but teratoma and adenocarcinoma can occur. Ultrasound and CT scan can help to define the nature of an ovarian mass in a child. Treatment is surgical excision.

Further reading

Vesicoureteral reflux

Arant B. Vesicoureteric reflux and renal injury. *American Journal of Kidney Diseases* 1991; **17**: 491.

Borer JG, Hendren WH. Disorders of the bladder and urethra. In: Burg FD, Ingelfinger JR, Wald ER, Polin RA, eds, *Gellis and Kagan's current pediatric therapy*, 16th edn, pp. 843–9. WB Saunders, Philadelphia, 1999.

Connolly LP *et al.* Vesicoureteral reflux in children: incidence and severity in siblings. *Journal of Urology* 1997; **157**: 2287.

Puri P, Cascio S, Lakshmandass G, Colhoun E. Urinary tract infection and renal damage in sibling vesicoureteral reflux. *Journal of Urology* 1998; **160**: 1028.

Van den Abbeele AD. Vesicoureteral reflux in asymptomatic siblings of patients with known reflux: radionuclide cystography. *Pediatrics* 1987; **79**: 147.

Ureteropelvic junction obstruction

Berstein GT, Mandell J, Lebowitz RL, Bauer S, Colodny AH, Retik AB. Ureteropelvic junction obstruction in the neonate. *Journal of Urology* 1988; **140**(2): 1216.

Cockrell SN, Hendren WH. The importance of visualizing the ureter before performing pyeloplasty. *Journal of Urology* 1990; **144**(2): 588.

Crooks KK, Hendren WH, Pfister RC. Giant hydronephrosis in children. *Journal of Pediatric Surgery* 1979; **14**: 844.

Hendren WH, Radhakrishnan J, Middleton AW Jr. Pediatric pyeloplasty. *Journal of Pediatric Surgery* 1980; **15**: 133.

Lebowitz RL, Johan BG. The coexistence of ureteropelvic junction obstruction and reflux. *American Journal of Radiology* 1982; **140**: 231.

Peters CA. Urinary tract obstruction in children. *Journal of Urology* 1995; **154**: 1874. [Review.]

Takla NV, Hamilton BD, Cartwright PC, Snow BW. Apparent unilateral ureteropelvic junction obstruction in the newborn: expectations for resolution. *Journal of Urology* 1998; **160**: 2175.

Megaureter

Hendren WH. Operative repair of megaureter in children. *Journal of Urology* 1969; **101**: 491.

Hendren WH, Carr MC, Adams MC. Megaureter and prune-belly syndrome. In: O'Neill JA Jr, Rowe MI, Grosfeld JL, Fonkalsrud EW, Coran AA, eds, *Pediatric surgery*, pp. 1631–51. Mosby-Year Book, St. Louis, 1998.

Peters CA *et al.* Congenital obstructed megaureter in early infancy: diagnosis and treatment. *Journal of Urology* 1989; **142**: 641.

Ureterocele

Ashcraft KW, Hendren WH. Bladder outlet obstruction after operation for ureterocele. *Journal of Pediatric Surgery* 1979; **14**: 819.

Coplen DE, Duckett JW. The modern approach to ureteroceles. *Journal of Urology* 1995; **153**: 166.

Decter RM, Roth DR, Gonzales ET. Individualized treatment of ureteroceles. *Journal of Urology* 1989; **142**(2): 535.

Gross RE, Clatworthy HW. Ureterocele in infancy and childhood. *Pediatrics* 1950; **5**: 68.

Hendren WH, Mitchell ME. Surgical correction of ureteroceles. *Journal of Urology* 1979; **121**: 590.

Kroovand RL, Perlmutter AD. A one-stage surgical approach to ectopic ureterocele. *Journal of Urology* 1979; **122**: 367.

Stephens FD. Caecoureterocele and concepts on the embryology and aetiology of ureteroceles. *Australian and New Zealand Journal of Surgery* 1971; **40**: 239.

Ectopic ureter

Johnson JH, Davenport TJ. The single ectopic ureter. *British Journal of Urology* 1969; **41**: 428.

Mandell J, Bauer SB, Colodny AH, Lebowitz RL, Retik AB. Ureteral ectopia in infants and children. *Journal of Urology* 1981; **126**: 219.

Posterior urethral valves

Hendren WH. A new approach to infants with severe obstructive uropathy: early complete reconstruction. *Journal of Pediatric Surgery* 1970; **5**: 184.

Hendren WH. Posterior urethral valves in boys. A broad clinical spectrum. *Journal of Urology* 1971; **106**: 298.

Hoover DL, Duckett JW Jr. Posterior urethral valves, unilateral reflux and renal dysplasia: a syndrome. *Journal of Urology* 1982; **128**: 994.

Kim YH, Horowitz M, Combs AJ, Nitti VW, Borer JG, Glassberg KI. Management of posterior urethral valves on the basis of urodynamic findings. *Journal of Urology* 1997; **158**: 1011.

Peters CA *et al.* The urodynamic consequences of posterior urethral valves. *Journal of Urology* 1990; **144**(2): 122.

Young HH, Frontz WA, Baldwin JC. Congenital obstruction of the posterior urethra. *Journal of Urology* 1919; **3**: 289.

Techniques

Ureteral reimplantation

Cohen SJ. Ureterozystoneostomie. Eine Neue Antireflux Technik. *Aktuelle Urologie* 1975; **6**: 1.

Hendren WH. Ureteral reimplantation in children. *Journal of Pediatric Surgery* 1968; **3**: 649.

Hendren WH. Reoperation for the failed ureteral reimplantation. *Journal of Urology* 1974; **111**: 403.

Hendren WH. Reoperative ureteral reimplantation: management of the difficult case. *Journal of Pediatric Surgery* 1980; **15**: 770.

Politano VA, Leadbetter WF. An operative technique for the correction of vesicoureteral reflux. *Journal of Urology* 1958; **79**: 932.

Puri P. Ten year experience with subureteric Teflon (polytetrafluoroethylene) injection (STING) in the treatment of vesico-ureteric reflux. *British Journal of Urology* 1995; **75**: 126.

Transureteroureterostomy and psoas hitch

Hendren WH, Hensle TW. Transureteroureterostomy: experience with 75 cases. *Journal of Urology* 1980; **123**: 826.

Hodges CV, Moore RJ, Lehman TH, Behnam AM. Clinical experience with transureteroureterostomy. *Journal of Urology* 1963; **90**: 552.

Bladder neck reconstruction

Kaefer M, Bauer SB. The surgical correction of incontinence in myelodysplastic children. In: King LR, ed., *Urologic surgery in infants and children*, pp. 119–37. WB Saunders, Philadelphia, 1998.

Leadbetter GW Jr. Surgical correction of total urinary incontinence. *Journal of Urology* 1964; **91**: 261.

Peters CA, Hendren WH. Splitting the pubis for exposure in difficult reconstructions for incontinence. *Journal of Urology* 1989; **142**(2): 527.

Young HH. An operation for the cure of incontinence associated with exstrophy. *Journal of Urology* 1922; **7**: 1.

Bladder augmentation

Adams MC, Mitchell ME, Rink RC. Gastrocystoplasty: an alternative solution to the problem of urological reconstruction in the severely compromised patient. *Journal of Urology* 1988; **140**: 1152.

Atala A, Bauer SB, Hendren WH, Retik AB. The effect of gastric augmentation on bladder function. *Journal of Urology* 1993; **149**: 1099.

Goodwin WE, Winter CC, Barker WF. Cup-patch technique of ileocystoplasty for bladder enlargement or partial substitution. *Surgery, Gynecology, and Obstetrics* 1959; **108**: 240.

Hendren WH, Hendren RB. Bladder augmentation: experience with 129 children and young adults. *Journal of Urology* 1990; **144**(2): 445.

Kaefer M *et al.* Reservoir calculi: a comparison of reservoirs constructed from stomach and other enteric segments. *Journal of Urology* 1998; **160**: 2187.

Kurzrock EA, Baskin LS, Kogan BA. Gastrocystoplasty: long-term followup. *Journal of Urology* 1998; **160**: 2182.

Leong CH. Use of stomach for bladder replacement and urinary diversion. *Annals of the Royal College of Surgeons of England* 1978; **60**: 283.

Urinary diversion

Athausen AF, Hagen-Cook K, Hendren WH. Non-refluxing colon conduit: experience with 70 cases. *Journal of Urology* 1978; **120**: 35.

Bricker EM. Bladder substitution after pelvic evisceration. *Surgical Clinics of North America* 1990; **30**: 1511.

Duckett JW Jr. Cutaneous vesicostomy in childhood. The Blocksom technique. *Urologic Clinics of North America* 1974; **1**: 485.

Goodwin WE, Harris AP, Kaufman JJ, Beal, JM. Open transcolonic ureterointestinal anastomosis: new approach. *Surgery, Gynecology, and Obstetrics* 1953; **97**: 295.

Middleton AW Jr, Hendren WH. Ileal conduits in children at the Massachusetts General Hospital from 1955 to 1970. *Journal of Urology* 1976; **115**: 591.

Retik AB, Perlmutter AD, Gross RE. Cutaneous ureteroileostomy in children. *New England Journal of Medicine* 1967; **277**: 217.

Richie JP, Skinner DG, Waisman J. The effect of reflux on the development of pyelonephritis in urinary diversion: an experimental study. *Journal of Surgical Research* 1974; **16**: 256.

Richie JP, Skinner DG. Urinary diversion: the physiological rationale for nonrefluxing colonic conduits. *British Journal of Urology* 1975; **47**: 269.

Smith ED. Follow-up studies on 150 ileal conduits in children. *Journal of Pediatric Surgery* 1972; **7**: 1.

Urinary tract refunctionalization

Hendren WH. Urinary tract refunctionalization after long-term division. A 20-year experience with 177 patients. *Annals of Surgery* 1990; **212**: 478.

Hendren WH. Urinary undiversion: refunctionalization of the previously diverted urinary tract. In Walsh PC, Retik AB, Vaughan ED, Wein AJ, eds, *Campbell's urology*, 7th edn, pp. 3247–73. WB Saunders, Philadelphia, 1998.

Continent urinary diversion

Duckett JW Jr, Snyder HM III. Continent urinary diversion: variations on the Mitrofanoff principle. *Journal of Urology* 1986; **136**: 58.

Kaefer M *et al.* Continent urinary diversion: the Children's Hospital experience. *Journal of Urology* 1997; **157**: 1394.

Rowland RG *et al.* The Indiana continent urinary reservoir. *Journal of Urology* 1987; **137**: 1136.

Skinner DG, Lieskovsky G, Boyd SD. Continuing experience with the continent ileal reservoir (Kock pouch) as an alternative to cutaneous urinary diversions: an update after 250 cases. *Journal of Urology* 1987; **137**: 1140.

Artificial urinary sphincter

Levesque PE *et al.* Ten-year experience with the artificial urinary sphincter in children. *Journal of Urology* 1996; **156**: 625.

Hypospadias

Borer JG, Retik AB. Current trends in hypospadias repair. *Urologic Clinics of North America* 1999; **26**: 15.

Devine CJ, Horton CE. Hypospadias repair. *Journal of Urology* 1977; **118**: 188.

Hendren WH, Ceasar RE. Chordee without hypospadias: experience with 33 cases. *Journal of Urology* 1992; **147**: 107.

Hendren WH, Keating MA. Use of dermal free graft and free urethral graft in penile reconstruction. *Journal of Urology* 1988; **140**(2): 1265.

Epispadias

Diamond DA, Ransley PG. Male epispadias. *Journal of Urology* 1995; **154**: 2150. [Review.]

Hendren WH. Penile lengthening after previous repair of epispadias. *Journal of Urology* 1979; **121**: 527.

Mitchell M, Bagli D. Complete penile disassembly for epispadias repair: the Mitchell technique. *Journal of Urology* 1996; **155**: 300.

Peters CA, Gearhart JP, Jeffs RD. Epispadias and incontinence: the challenge of the small bladder. *Journal of Urology* 1988; **140**(2): 1199.

Bladder exstrophy

Avolio L, Koo HP, Bescript AC, Snyder HM, Canning DA, Duckett JW Jr. The long-term outcome in men with exstrophy/epispadias: sexual function and social integration. *Journal of Urology* 1996; **156**: 822.

Gearhart JP, Mathews R, Taylor S, Jeffs RD. Combined bladder closure and epispadias repair in the reconstruction of bladder exstrophy. *Journal of Urology* 1998; **160**: 1182.

Grady RW, Mitchell ME. Newborn exstrophy closure and epispadias repair. *World Journal of Urology* 1998; **16**: 200.

Lepor H, Shapiro E, Jeffs RD. Urethral reconstruction in males with classical bladder exstrophy. *Journal of Urology* 1984; **131**: 512.

Cryptorchidism

Cisek LJ, Peters CA, Atala A, Bauer SB, Diamond DA, Retik AB. Current findings in diagnostic laparoscopic evaluation of the nonpalpable testis. *Journal of Urology* 1998; **160**: 1145.

Gross RE, Replogle RL. Treatment of undescended testes. Opinions gained from 1967 operations. *Postgraduate Medicine* 1963; **34**: 266.

Hendren WH, Black PR. Inguinal hernia repair, orchidopexy, and umbilical hernia repair in children. *Problems in General Surgery* 1986; **3**: 329.

Huff DS, Hadziselimovic F, Snyder HM III, Duckett JW Jr, Keating MA. Postnatal testicular maldevelopment in unilateral cryptochidism. *Journal of Urology* 1989; **142**(2): 546.

Neuropathic bladder

Edelstein RA *et al.* The long-term urological response of neonates with myelodysplasia treated proactively with intermittent catheterization and anticholinergic therapy. *Journal of Urology* 1995; **154**: 1500.

Fernandes ET, Reinberg Y, Vernier R, Gonzalez R. Neurogenic bladder dysfunction in children: review of pathophysiology and current management. *Journal of Pediatrics* 1994; **124**: 1.

Spindel MR *et al.* The changing neurourlogic lesion in myelodysplasia. *Journal of the American Medical Association* 1987; **258**: 1630.

Intersex states

Donahoe PK, Hendren WH. Perineal reconstruction in ambiguous genitalia in infants raised as females. *Annals of Surgery* 1984; **200**: 363.

Hendren WH. Surgical approach to intersex problems. *Seminars in Pediatric Surgery* 1998; **7**: 8.

Hendren WH, Crawford JD. The child with ambiguous genitalia. *Current Problems in Surgery* 1972; **Nov**: 1.

Cloacal anomalies

Hendren WH. Urogenital sinus and cloacal malformations. *Journal of Pelvic Surgery* 1995; **1**: 149.

Hendren WH. Cloaca, the most severe degree of imperforate anus. Experience with 195 cases. *Annals of Surgery* 1998; **228**: 331.

Cloacal exstrophy

Diamond DA, Jeffs RD. Cloacal exstrophy: a 22-year experience. *Journal of Urology* 1985; **133**: 779.

Hurwitz RS, Manzoni GAM, Ransley PG, Stephens FD. Cloacal exstrophy: a report of 34 cases. *Journal of Urology* 1987; **138**: 1060.

Lund DP, Hendren WH. Cloacal exstrophy: experience with 20 cases. *Journal of Pediatric Surgery* 1993; **28**: 1360.

Mathews R, Jeffs RD, Reiner WG, Docimo SG, Gearhart JP. Cloacal exstrophy—improving the quality of life: the Johns Hopkins experience. *Journal of Urology* 1998; **160**: 2452.

Pediatric urologic tumors

Green DM *et al.* Wilms tumor. *Ca: A Cancer Journal for Clinicians* 1996; **46**: 46. [Review.]

Matthay KK. Neuroblastoma: biology and therapy. *Oncology* 1997; **11**: 1857. [Review.]

Snyder HM, D'Angio GJ, Evans AE, Raney RB. Pediatric oncology. In: Walsh PC, Retik AB, Vaughan ED, Wein AJ, eds, *Campbell's urology*, 7th edn, pp. 2210–56. WB Saunders, Philadelphia, 1998.

42.10 Paediatric solid tumours

Robert C. Shamberger

Wilms tumor
- Molecular biology
- Nephrogenic rests and nephroblastomatosis
- Histology
- Treatment
- Surgery of Wilms tumor

Neuroblastoma
- Histology
- Staging
- Clinical presentation
- Molecular biology
- Treatment
- Antenatal diagnosis
- Neuroblastoma screening trials

Malignant liver tumors
- Histology
- Presentation and evaluation
- Treatment

Rhabdomyosarcoma
- Pathology
- Clinical features
- Treatment
- Prognosis
- Specific primary tumor sites
- Genitourinary

Germ cell tumors
- Pathology
- Staging
- Treatment
- Teratoma—other extra-gonadal sites
- Treatment—chemotherapy

Further reading

Major advances in the management and cure of solid tumors in infants and children have occurred in the last several decades. These have resulted from collaborative efforts by national and international cooperative groups. Because of the relative rarity of pediatric solid tumors, no single institution can accumulate an adequate number of cases to perform randomized treatment protocols. Cooperative efforts by the National Wilms Tumor Study Group (NWTSG), International Society of Pediatric Oncology (SIOP), Pediatric Oncology Group, Children's Cancer Group, and others have accumulated adequate cases for essential randomized studies. Advances in understanding the molecular biology and genetics of solid pediatric tumors have also resulted from these collaborative efforts which provide large numbers of tumor specimens for evaluation and study.

The common pediatric tumors will be presented in this section. Surgery plays two major roles in their management. The first is to provide histologic diagnosis and staging of the tumor; the second is to achieve resection. It is critical for the surgeon to work in a collaborative fashion with the pediatric oncologist and radiation oncologist, because resection may be best accomplished after initial chemotherapy and/or radiotherapy. Initial treatment of some tumors may decrease both the potential risks of resection and long-term morbidity. Similarly, it is critical that the surgeon be involved from the onset in the care of a child presenting with a solid tumor, because an inappropriate biopsy of the tumor may complicate later resection. Questions regarding resectability should be determined by the surgeon who is best equipped to address these issues.

Advances in anesthetic management of infants and children have contributed greatly to the successful surgical resection of solid tumors. Procedures such as hepatectomy and extensive retroperitoneal resections performed in the past with significant morbidity and mortality are now accomplished on a routine basis with limited risks. Current principles of anesthetic management include maintenance of the body temperature of the child, adequate blood volume replacement including fresh frozen plasma and platelets to avoid development of a coagulopathy, and ventilation that avoids barotrauma to the lung. Intraoperative monitoring of hemoglobin oxygen saturation and urine output and in extensive resections monitoring both arterial and central venous pressures allows rapid identification and correction of any physiologic derangements.

Wilms tumor

The most frequent presenting complaint of a child with a renal tumor is the identification of an abdominal mass. These are often quite large before they are noted during a bath or while dressing a child. Hematuria, which is a frequent symptom of renal cell carcinoma in adults, is much less commonly associated with the pediatric tumors. Pain is also much less frequent. The initial evaluation of an infant or child presenting with an abdominal mass is shown in Fig. 1. Evaluation of urinary catecholamines is important to facilitate the differential diagnosis between Wilms tumor and neuroblastoma which may be difficult in some cases. The catecholamines will be elevated in 90 to 95 per cent of cases of neuroblastoma, but will be normal in Wilms tumor. Wilms tumor is the most frequent tumor of the kidney in infants and children occurring with a reported incidence of eight cases per million children less than 15 years of age.

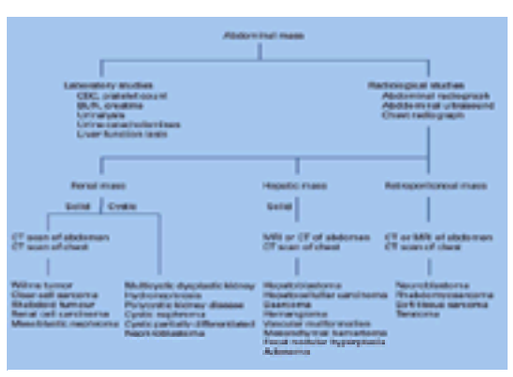


Fig. 1. Algorithm for diagnosis of abdominal masses in children.

Molecular biology

The association between Wilms tumor and several congenital syndromes including sporadic aniridia, isolated hemihypertrophy, the Denys–Drash syndrome (nephropathy, renal failure, male pseudohermaphroditism, and Wilms tumor), genital anomalies, Beckwith–Wiedemann syndrome (visceromegaly, macroglossia, omphalocele, and hyperinsulinemic hypoglycemia), and the WAGR complex (Wilms tumor with aniridia, genitourinary malformations, and retardation) led to early speculation of a genetic predisposition to this tumor. Subsequent molecular studies of children with Wilms tumor have demonstrated in some cases a deletion within the short arm of chromosome 11 occurring at band p13. This deletion has subsequently been shown to include the aniridia gene and a putative Wilms tumor suppressor gene (WT1). The protein product of this gene is a transcriptional factor which regulates the expression of other genes including growth inducing genes such as those encoding early growth response, insulin-like growth factor II, and platelet-derived growth factor A chain. Suppression of these growth-associated genes may

explain the tumor suppressor role of WT1. The WT1 gene product has also recently been found to bind physically to the p53 protein.

Abnormalities at the 11p15 locus have been associated with Beckwith–Wiedemann patients, although the locus for a Wilms tumor gene has not been defined. Some children with Beckwith–Wiedemann syndrome have overexpression of a gene normally expressed by only one of the parental alleles. In some cases a constitutional duplication of the paternal 11p15 chromosomal fragment (trisomy at 11p15) has been identified. In other cases where the children appear to have two grossly normal copies of chromosome 11, it has been demonstrated that both copies are from the father with none from the mother (uniparental isodisomy). These findings led to speculation that the Beckwith–Wiedemann gene is expressed only by the paternal allele. These genetic abnormalities which led to the presence of two paternal alleles would double the expression of this gene and may result in the overgrowth. The insulin-like growth factor 2 gene is present at the 11p15 locus. It is an embryonal growth-inducing gene with expression restricted to the paternal allele. However, there is no direct evidence to link the insulin-like growth factor 2 gene to the causation of the Beckwith–Wiedemann syndrome. It is of note, that those children, whose manifestations of the Beckwith–Wiedemann syndrome include hemihypertrophy, appear to have a greater risk for the occurrence of malignancy. In one series cancer was reported in 7.5 per cent of all children with the syndrome, but in more than 40 per cent of those with the syndrome and hemihypertrophy.

Recent studies have suggested that loss of heterozygosity on chromosome 16q in Wilms tumors was associated with a 3.3-fold increase in relapse and a 12-fold increase in mortality when compared with children without these changes. A similar trend was seen for children with loss of heterozygosity for 1p, but these trends were not statistically significant.

Routine radiographic screening of children with syndromes associated with Wilms tumor has been suggested with ultrasonograms every 3 months until the children are 5 years of age. No prospective studies, however, have been performed to evaluate the cost-effectiveness or efficacy of this recommendation. A retrospective review suggested that routine ultrasonographic screening did not alter either the stage distribution at presentation or the outcome of the children.

Nephrogenic rests and nephroblastomatosis

The occurrence of nephrogenic rests, persistent metanephric tissue in the kidney after the 36th week of gestation, has been associated with the occurrence of Wilms tumor and these lesions are considered premalignant. They may occur in a perilobular or intralobular location and may be single or multiple. In children with aniridia or the Denys–Drash syndrome the lesions are primarily intralobular whereas, children with hemihypertrophy or the Beckwith–Wiedemann syndrome have predominantly perilobular lesions. The presence of multiple or diffuse nephrogenic rests is termed nephroblastomatosis.

In an autopsy series of infants under 3 months of age, nine of 1035 infants (0.87 per cent) had perilobular nephrogenic rests while intralobular nephrogenic rests are a much rarer occurrence, two of 2000 cases (0.1 per cent). The vast majority of these lesions will spontaneously resolve without producing a tumor as the incidence is about 100 times that of Wilms tumor (approximately one per 10 000 infants will develop a Wilms tumor during childhood). Despite their relatively rare occurrence nephrogenic rests are frequently found in association with Wilms tumor. In a review of cases of unilateral Wilms tumor in the National Wilms Tumor Study-4 (NWTS-4), 41 per cent of unilateral Wilms tumors were associated with rests. In children with synchronous bilateral Wilms tumor, the incidence in NWTS-3 was 99 per cent. These were primarily perilobular nephrogenic rests, which may be due to the fact that these lesions are much more prevalent than the intralobular nephrogenic rest.

Histology

The importance of the histology of Wilms tumor for prognosis and treatment is increasingly recognized. During the course of randomized multi-institutional studies the aggressive behavior and poor outcome of several pathologic subgroups has been identified. Wilms tumors are currently divided into those with ‘favorable' histology and those with ‘unfavorable' histology. Included in the latter group are children with ‘focal' or ‘diffuse' anaplasia. While clear cell sarcoma of the kidney and malignant rhabdoid tumors of the kidney were initially grouped with the Wilms tumors, they are now considered as distinct entities and receive quite different chemotherapy.

Treatment

Wilms tumor was the first malignancy in which the importance of adjuvant treatment of the tumor was recognized. From 1931 through 1939, survival from surgical resection alone involving ligation of the renal pedicle before removal was 32 per cent at the Children's Hospital in Boston. Beginning in 1940, most of the patients received postoperative radiation to the renal fossa. This treatment decreased the local recurrence rate, but did not significantly impact the frequency of pulmonary metastases or improve the long-term survival. Of the 53 children who had no demonstrable metastases on admission treated from 1957 to 1964 with combined therapy of surgery, local irradiation, and actinomycin-D, an 89 per cent 2-year disease-free survival was reported. In children with metastases identified at presentation, 18 of 31 (58 per cent) were alive and free of disease 2 years later. The need for adjuvant treatment ‘was based upon the supposition that in the children with Wilms tumor who died, the tumor must have metastasized already at the time of discovery of the primary tumor', although no evidence of spread was available.

Optimal chemotherapy regimens have been defined in a series of randomized studies by two large multi-institutional groups, the NWTSG in the United States and Canada and SIOP in Europe. Despite the advances in chemotherapy, surgery continues to play a major part in the treatment of this tumor; surgical staging and resection of Wilms tumor are major components of its treatment. Rarely are radiotherapy and chemotherapy employed solely for local control. The importance of staging cannot be underestimated (Table 1). The form of adjuvant treatment is determined by such factors as regional lymph node involvement and penetration of the renal capsule by tumor which cannot be determined radiographically.

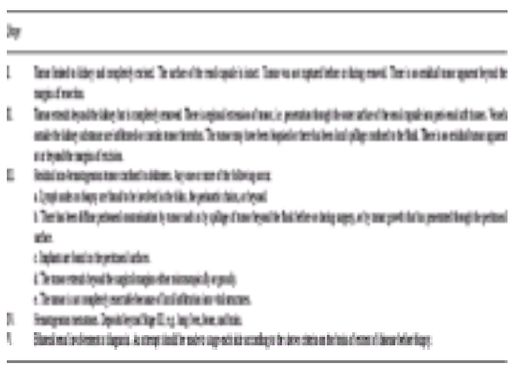


Table 1 National Wilms' Tumor Study staging system

Surgery of Wilms tumor

Initial resection of the tumor has been the policy supported by the NWTSG through all of its protocols. Despite the fact that most Wilms tumors present as large masses, resection is generally feasible (Fig. 2). Wilms tumor, in contrast with neuroblastoma, is much less likely to invade surrounding organs directly, which would complicate resection. Children undergoing initial nephrectomy in the NWTS-3 cohort demonstrated a complication rate of 19.8 per cent in a group very closely followed and evaluated. The most frequent complication was intestinal obstruction occurring in 6.9 per cent of the children followed by extensive intraoperative hemorrhage (>50 ml/kg of body weight) in 5.8 per cent. Injuries to other visceral organs (1 per cent) and extensive vascular injuries (1.4 per cent) were much less frequent. Nine deaths in the series were attributed to surgical complications (0.5 per cent), only one of which was intraoperative. Factors, which correlated with increased risk of surgical complications, were advanced stage local disease, intravascular extension of the tumor, and resection of other organs. The latter were often found not to be invaded by the tumor, but rather compressed, distorted or adherent without actual tumor infiltration. Extensive resection involving removal of other organs or procedures which are of a magnitude to be life threatening, should probably be aborted and a biopsy obtained of the tumor and of regional lymph nodes, followed by administration of chemotherapy before a second attempt at resection. Following this algorithm, 93 per cent of 131 children enrolled in NWTS-3 who were initially judged unresectable at surgery or by imaging studies were successfully resected after initial chemotherapy and/or irradiation. Only eight children with tumors which grew or failed to respond, did not undergo nephrectomy.

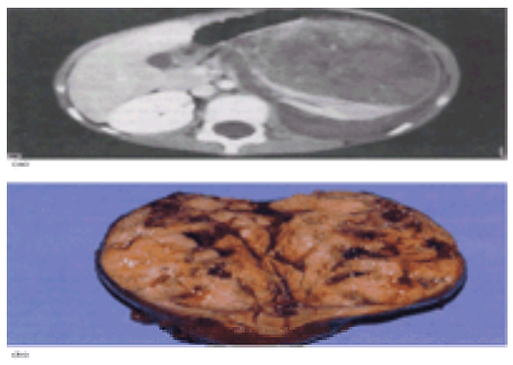


Fig. 2. (a) Computed tomogram of a 5-year-old girl with a large anaplastic Wilms tumor reveals the thin capsule of normal renal parenchyma surrounding the primary tumor. There is also extension of the tumor into the retroperitoneal tissues posterior to the kidney. (b) Transected Wilms tumor demonstrates margin of normal parenchyma and the characteristic heterogeneous Wilms tumor.

SIOP has promoted the use of preoperative treatment of children with Wilms tumor with radiotherapy or chemotherapy since the early 1970s. Histologic confirmation of the diagnosis before therapy is not routinely recommended by SIOP. This approach has several risks: (a) the potential for administration of chemotherapy for benign disease; (b) modification of the tumor histology; (c) the loss of staging information; and (d) treatment for the incorrect tumor, such as rhabdoid tumor of the kidney. Treatment without an initial diagnosis is difficult to sustain when NWTSG and SIOP studies have demonstrated a 7.6 to 9.9 per cent rate of benign or altered malignant diagnosis in children with a prenephrectomy diagnosis of Wilms tumor.

A major factor driving the use of preoperative therapy by SIOP was the high rate of operative tumor rupture in their early series without preliminary treatment. The rupture rate decreased from 33 per cent (20 of 60) to 4 per cent (three of 72) when preoperative abdominal irradiation (20 Gy) was implemented; however, it must be noted that 33 per cent is an extremely high frequency of this complication. This did not affect survival and there is no mention of the incidence of local recurrence. Operative rupture occurred in NWTS-1 and NWTS -2 in 22 per cent and 12 per cent of children respectively. In a subsequent SIOP randomized study, the rate of rupture was essentially the same for children receiving abdominal irradiation (20 Gy) and actinomycin D (9 per cent, seven of 76) and those receiving vincristine and actinomycin D (6 per cent, five of 88). In two consecutive but non-randomized studies the proportion of stage II lymph node negative children versus stage II lymph node positive and stage III, changed from 45 to 32 per cent and 33 to 19 per cent suggesting that the preoperative treatments significantly decreased the apparent stage of the children.

The histologic diagnosis following preoperative treatment in a group of children followed on NWTS did not appear to have been distorted by treatment, but it is less certain that factors of staging have not been altered. The third SIOP study of Wilms tumor randomized the use of local radiotherapy (20 Gy) in children treated preoperatively with chemotherapy (vincristine and dactinomycin) who had stage II lymph node negative disease at resection. All children received vincristine and dactinomycin for 3 weeks. A high rate of stage I tumors (52 per cent) was found in this study and a low frequency of ruptures (7 per cent). The study was terminated after randomization of 123 children because of an increased incidence of abdominal recurrence during the first year of follow-up in the children not receiving radiation (six of 59 versus zero of 64). This suggested that prenephrectomy treatment altered the pathologic findings which would have led to a diagnosis of stage III disease (i.e. lymph node involvement or capsular penetration) and to the standard administration of local irradiation.

The efficacy of preoperative chemotherapy in allowing the safe performance of partial nephrectomy for Wilms tumor has been evaluated by several centers. A group in Toronto obtained percutaneous biopsy in 37 children with Wilms tumor and then administered multiagent chemotherapy for 4 to 6 weeks. A partial nephrectomy was then performed in nine children (four with unilateral and five with bilateral tumors). Two children suffered intra-abdominal relapse. Only four of the 30 unilateral tumors (13.3 per cent) were amenable to a partial nephrectomy. Another analysis of the feasibility of partial nephrectomies was performed at St Jude Children's Research Hospital. Preoperative computed tomography (CT) scans of 43 children with non-metastatic unilateral Wilms tumor were reviewed retrospectively. Criteria to allow partial nephrectomy were involvement by the tumor of one pole and less than one-third of the kidney, a functioning kidney, no involvement of the collecting system or renal vein, and clear margins between the tumor and surrounding structures. Utilizing these criteria partial nephrectomy was feasible identified in only two of 43 scans (4.7 per cent). The primary concerns regarding use of preoperative chemotherapy to create ‘resectable' small tumors is that these children may be curable by surgical resection alone without subjecting them to the toxicity of additional treatments.

Preoperative treatment of Wilms tumor is generally accepted in certain circumstances. These include the occurrence of Wilms tumor in a solitary kidney, bilateral renal tumors, tumor in a horseshoe kidney, and respiratory distress from extensive metastatic tumor (Fig. 3). In most instances pretreatment biopsy should be obtained. The aim of treatment prior to surgical resection in the bilateral tumors and tumor in a solitary kidney is to preserve the maximum amount of the kidney possible and hence the greatest renal function.

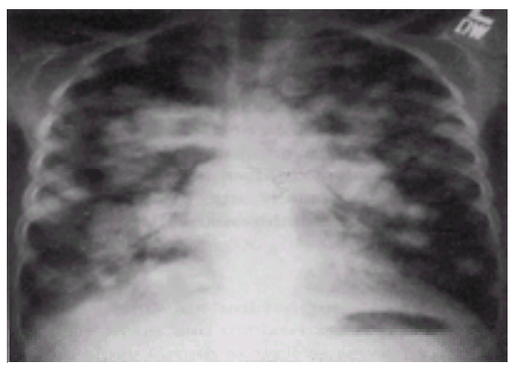


Fig. 3. A chest radiograph of a 5½-year-old girl who presented with respiratory distress resulting from extensive pulmonary metastases from a standard histology Wilms tumor.

Growth of the remaining kidney after unilateral nephrectomy has been well established and a 180 per cent volume augmentation has been demonstrated. Focal segmental glomerulosclerosis has been reported in children with a unilateral kidney. The incidence of renal failure, however, following unilateral nephrectomy in the NWTS-1 to NWTS-4 population was only 0.25 per cent. Studies from Europe on pretreatment of unilateral Wilms tumor have demonstrated that in most instances a nephrectomy is still required because of the extent of tumor involvement with the kidney at presentation.

In children with bilateral tumors, preservation of renal parenchyma is a critical issue. In the NWTS review of renal failure in 55 children from NWTS-1 to NWTS-4, 39 children had bilateral tumor involvement. Increasing efforts to preserve renal parenchymal in bilateral cases in the progression of the NWTS studies resulted in a decline in the incidence of renal failure from 16.4 per cent in NWTS-1 and NWTS-2 to 9.9 per cent in NWTS-3 and 3.8 per cent in NWTS-4. Preliminary treatment in most cases following biopsy and staging will produce shrinkage of the tumor and facilitate resection of the tumor with preservation of a portion of the kidney. Ninety-eight children with bilateral Wilms tumors underwent a partial nephrectomy of 134 kidneys during NWTS-4. One hundred and one kidneys underwent initial biopsy followed by multi-agent chemotherapy prior to their resections. Complete resection of gross disease was accomplished in 118 (88 per cent) of the 134 kidneys. A higher incidence of positive surgical margins (16 per cent) and local tumor recurrence 8.2 per cent (11 children), was seen in this group of children than in children with unilateral tumors which was justified by the attempt to preserve renal tissue and avoid renal failure. Overall, portions of 72 per cent of the kidneys were preserved and the 4-year survival rate was 81.7 per cent. Of note, local recurrence occurred in only one of the children with gross residual disease or positive margins.

Vascular extension of a tumor thrombus is not a rare event occurring in 4 per cent of children with Wilms tumor. Identification of vascular extension by radiographic studies or early in the surgical exploration is critical to avoid a tumor embolus during mobilization of the kidney. Ultrasonography will define vascular invasion well and is probably more sensitive than CT. Traditionally, vascular extension has been managed by nephrectomy with resection of the tumor thrombus into the renal vein or vena cava. Cardiopulmonary bypass has been utilized for children with atrial extension of the tumor thrombus. This approach has been associated with a significant incidence of complications. In a recent NWTS report, 30 children with vascular extension (caval in 15 children and atrial in 15 children) were treated initially with chemotherapy after biopsy of the renal mass. In 23 children a decrease in the size of the intravascular extension occurred and complete resolution of the tumor thrombus was seen in seven children. Of the 15 children with tumor initially extending into the atrium, one child succumbed to the adult respiratory distress syndrome

before surgery. Only four had residual atrial involvement at resection requiring sternotomy and bypass. In 10 children with atrial extension a complete or marked response occurred and the tumor was removed transabdominally. Tumor embolism did not occur during chemotherapy. Fibrosis of the caval wall developed in some cases and in two IVC occlusion occurred postoperatively. This study suggests that many, although not all children with atrial or caval involvement, will require significantly less extensive resection with preliminary treatment. A similar review of children treated in the United Kingdom with extensive vascular involvement also demonstrated a decrease in the extent of vascular involvement in 16 of 21 children and showed that children receiving preoperative chemotherapy had a better outcome.

Adequate biopsy of lymph nodes in the renal hilum and along the vena cava or aorta is critical for adequate staging. Even in children with stage IV disease local staging is critical, as it will determine local treatment, particularly the utilization of radiotherapy. The surgeon must always consider stage III disease and obtain adequate tissue for its diagnosis. Recent studies by the author and coworkers have demonstrated that children in whom the lymph nodes are not sampled have a higher incidence of local recurrence. This suggests that they are being understaged and as a result, receive inadequate treatment for their local disease. The increased incidence of local recurrence occurred primarily in the stage I patients. While grossly involved lymph nodes are generally resected, this approach should not be extrapolated into a recommendation for an extensive retroperitoneal lymph node resection as this has not been demonstrated to improve local control.

Gross hematuria in children with Wilms tumor is infrequent, but its occurrence suggests extensive involvement of the renal pelvis with possible extension into the ureter. Cystoscopy should be considered in these children to identify extension of the tumor into the bladder and to avoid transection of the tumor thrombus during division of the ureter. Ureteral extension, which is recognized and entirely resected, does not increase the local stage of the tumor.

Exploration of the contralateral kidney has been recommended by the NWTSG based on the 5 per cent occurrence of synchronous lesions. In NWTS-2 and NWTS-3 approximately one-third of the children with bilateral tumors were not detected by intravenous pyelography or CT before surgical exploration. Opening of Gerota's fascia to examine directly the anterior and posterior aspects of the kidney is suggested. A recent review of children with bilateral tumors treated on NWTS-4 identified nine of 122 children in whom the diagnosis of bilateral disease was missed by the preoperative imaging studies (CT, ultrasonography, or magnetic resonance imaging (MRI)). All but one of these lesions were small with five being less than 1 cm and three were 1 to 3 cm in diameter. CT was found to be more accurate for diagnosing contralateral disease than ultrasonography. Thus, 7 per cent of the 5 per cent of children presenting with bilateral disease would have been missed if exploration had not been performed, or an incidence of 0.35 per cent among all children with Wilms tumor. This low incidence explains why recent studies with smaller numbers of children have drawn other conclusions.

A small group of children with Wilms tumor may require only resection of the primary and no adjuvant treatment. This step is taken very cautiously as adjuvant treatment produced such a dramatic improvement in the survival of children with Wilms tumor. The outcome in children less than 2 years of age with stage I tumors weighing less than 550 g registered in the NWTS studies has been reviewed. The 4-year relapse-free survival for children meeting these criteria exceeded 90 per cent suggesting that they could be selected for treatment with surgery alone. A prospective pilot study of eight children with stage I disease who were under 2 years of age with unilateral, favorable histopathology and combined tumor and kidney weight under 550 g was carried out. Resection and close follow-up was performed. Metachronous recurrence in one child was cured by resection and chemotherapy. Continued evaluation of this series shows no episodes of local recurrence. Children meeting these criteria will be followed closely without adjuvant treatment on the NWTS-5 protocol.

The role of surgery in treatment of pulmonary relapse in 211 patients has been evaluated by Green and colleagues. While diagnostic confirmation of relapse may be required, there was no therapeutic benefit identified to resection of a solitary pulmonary metastasis in addition to pulmonary radiotherapy and chemotherapy alone. Four-year survival was identical in the two groups.

Wilms tumor occurs rarely in the neonate. Only 27 infants 30 days old or less with renal tumors (0.8 per cent) were identified in the 3340 children entered into the NWTS from 1969 to 1984. Over half of the infants (18) had mesoblastic nephroma and four others had non-neoplastic lesions. One infant had a malignant rhabdoid tumor and four had Wilms tumors for an incidence of 0.1 per cent of all Wilms tumors appearing at this early age. All of these children had favorable histology tumors without metastasis. They did well receiving a variety of treatments ranging from surgery alone to 15 months of three-drug therapy. A subsequent report of 15 cases of Wilms tumor occurring in infants in the first 30 days of life again demonstrated a universal occurrence of favorable histology tumors and absence of metastatic disease. Ten of these infants received postoperative chemotherapy and five were followed without additional treatment. Only one of these five children recurred locally and with pulmonary disease and ultimately succumbed to her disease at 16 months of age. The other children are all disease-free at a median follow-up of 31 months.

Neuroblastoma

Neuroblastoma is the most common extracranial solid tumor in infants and children. The incidence is 9.7 cases per million children younger than 15 years of age in the United States. Neuroblastoma occurs primarily in the infant and toddler and the median age at diagnosis is 2 years. Metastases are present in 60 to 65 per cent of children at presentation and despite advances in treatment of many other pediatric solid tumors, the outlook for children with advanced stage neuroblastoma remains dismal. Neuroblastoma accounts for a disproportionately high number of cancer deaths in children.

Neuroblastoma arises from the primitive neural crest cells and presents as a solid and generally very vascular mass. It may occur anywhere along the sympathetic chain within the neck, thorax, retroperitoneum, pelvis, and in the adrenal gland. Seventy-five per cent arise in the retroperitoneum; 50 per cent in the adrenal and 25 per cent in the paravertebral ganglia. Twenty-two per cent occur in the thorax with the remaining 8 per cent divided primarily between cervical and pelvic sites. The incidence of *in-situ* neuroblastoma identified at autopsy in infants less than 3 months of age is 1 in 259. This 400-fold higher autopsy incidence compared with the clinical incidence suggests that the vast majority of these *in situ* tumors spontaneously regress. Neuroblastoma has a generally favorable course when it occurs in infants under a year of age, but a much more aggressive behavior in older children.

Histology

Microscopically, neuroblastoma is in the family of ‘small round blue cell tumors' seen in children which includes rhabdomyosarcoma, Ewing's sarcoma, and primitive neuroectodermal tumors. Over 90 per cent of children with neuroblastoma will have elevated levels of urinary catecholamines. Neuroblasts have a small amount of cytoplasm and may be arranged in rosettes with fine nerve fibers in the center. The degree of differentiation in these neural tumors varies from the benign ganglioneuroma to the malignant neuroblastoma. Several grading systems have been developed to provide a prognostic index for neuroblastoma. One classification has been developed based upon patient age and the separation of tumors into two large groups; stroma-rich and stroma-poor tumors based on the presence or absence of Schwannian spindle cell stroma. The stroma-poor tumors were further subdivided into those that were differentiated and those that were undifferentiated. Undifferentiated tumors have a high mitosis karyorrhexis index. This index is derived from the sum of necrotic tumor cells, cells with mitosis, and cells with malformed, lobulated, or pyknotic nuclei per 5000 cells examined. The cut-off value for a high index is age related; for infants under 18 months it is greater than 200/5000 cells while for children over 18 months it is 100/5000 cells. Children over 5 years are all defined as having unfavorable histology. The unfavorable stroma-poor and nodular stroma-rich histology groups have a poor prognosis (<10 per cent survival). Children with favorable stroma-poor and well-differentiated tumors in contrast have excellent survival (>78 per cent). It has been demonstrated that both histology and N-myc amplification provide prognostic information which is independent of stage.

Staging

Multiple staging systems are utilized for neuroblastoma including those developed by Evans and the later systems developed for use by the Pediatric Oncology Group as well as several European methods. This makes comparison of treatment results between various centers difficult. Recently, an international neuroblastoma staging system has been created and accepted by investigators working worldwide. It is hoped that generalized acceptance of this system will both enhance collaborative studies and facilitate comparison of treatment results (Table 2).

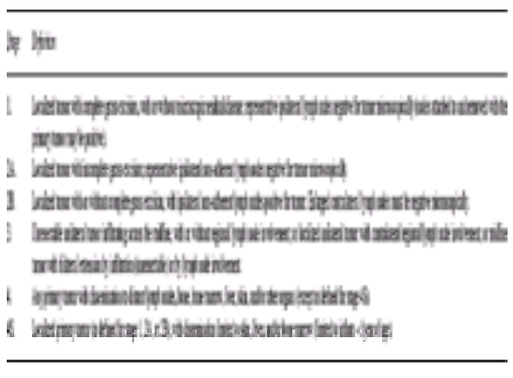


Table 2 International neuroblastoma staging system

Clinical presentation

Infants and children with neuroblastoma present with a myriad of symptoms. Infants may present with a localized cervical or abdominal mass. An unusual presentation which occurs primarily in association with mediastinal lesions is opsoclonus (bursts of rapid involuntary eye movements) and myoclonus (motor incoordination from frequent irregular, jerky contractions of muscles of the limbs and trunk), which may be the only manifestation of the disease and allow diagnosis when the tumor is extremely small. The mechanism of this encephalopathy remains poorly understood, presumably from an autoimmune phenomenon. Although survival is very favorable in these children, the neurologic outcome including learning disabilities and attention deficits, is less certain despite successful extirpation of the tumor and treatment with adrenocorticotropic hormone. In older children, symptoms from metastatic disease predominate. These include bone pain which may produce gate abnormalities, pallor, lethargy, weight loss, respiratory symptoms, irritability, and thoracic or abdominal pain. Unexplained skeletal and joint pain must always raise concern regarding this tumor. The retro-orbital area is a frequent site of spread and children may present with exophthalmos or infraorbital ecchymosis ‘panda eyes', which may initially lead to concerns regarding child abuse. Catecholamine production by the tumor produces hypertension in 25 per cent of children. It can also produce flushing, sweating, and irritability. A small number of children will have watery diarrhea resulting from increased levels of vasoactive intestinal polypeptide generated by the tumor. Paravertebral or adrenal lesions may extend into the spinal canal and produce cord compression and lower extremity neurologic signs and symptoms.

Evaluation for the presence of metastatic disease in a child with neuroblastoma should include a bone marrow biopsy, radioisotope bone scintigraphy, and radiographic studies of the primary area of the tumor to evaluate for lymph node metastases. A chest radiograph should be obtained to identify the occasional child with pulmonary metastases.

Molecular biology

Investigation of the molecular biology of neuroblastoma has been extremely rewarding. Deletion of the distal short-arm of chromosome 1 was the genetic mutation first described in neuroblastoma. This abnormality is found more frequently in patients with advanced disease and it implies a poor prognosis.

The N-myc proto-oncogene is derived from the C-myc viral oncogene and is located in the distal arm of chromosome 2. Amplification of the N-myc oncogene in neuroblastoma correlates very closely with advanced stage disease and with a poor outcome. It has been demonstrated that in infants under a year of age the biologic variables are predominantly favorable, whereas in older children unfavorable variables predominate. These findings explain the long recognized difference in survival based on age. DNA content or ‘ploidy' of neuroblastoma cells is a predictor of response to chemotherapy of unresectable neuroblastoma. Neuroblastoma was one of the first tumors in which molecular biologic characteristics determined the treatment received.

Treatment

Resent developments in identification of the prognostic criteria (Table 3) has allowed children to be categorized into more appropriate risk categories, beyond the previously utilized staging systems. It is expected that these systems utilizing multiple factors will define those children, regardless of age, with the most aggressive tumors and worst prognosis which require the most extensive treatments for cure. Similarly, those with the lowest risk will presumably require little more than surgery, and an intermediate risk group will fall midway between these two. Treatment protocols based on these multiple parameters are in progress.

	Unfavorable if:
Tumor DNA content and ploidy	Normal ploidy
N-myc oncogene	Amplified oncogene
Deletion of 1p	Present
Abnormalities in nerve growth factor	Lack high-affinity receptor (TRK) gene
Neuron-specific enolase	Elevated levels (> 100ng/ml)
Serum ferritin levels	Elevated levels (> 150ng/ml)

Table 3 Biologic markers in neuroblastoma

Stage 1 and 2 disease

Surgery has traditionally played a major part in the treatment of low-stage neuroblastoma. The surgical approach varies depending upon the site of the primary. Most thoracic, pelvic, and cervical primaries and abdominal lesions that do not extend across the midline to involve the great vessels will be resectable at presentation. Surgery should include complete excision of the localized tumor as well as sampling of lymph nodes adjacent to the tumor and at contralateral and distant sites. Lymph nodes grossly involved with the tumor should be resected if possible. Surgery alone is curative in many children with low-stage disease now recognized to be children with favorable biologic markers. Survival is achieved in over 90 per cent of these cases.

Stage 3 disease

Improved survival has been reported in children with stage 3 disease (those with extension across the midline) who have achieved complete resection at initial or subsequent exploration in comparison with those children who were not resectable. Similar results were reported from the European Neuroblastoma Study Group. The Italian Cooperative Group for neuroblastoma reported that complete resection after chemotherapy of extensive initially unresectable neuroblastoma was associated with improved survival when compared with those children undergoing only partial resection. Similar results have been found by others. Children previously treated on CCG protocols were reclassified using the international neuroblastoma staging system system. This study demonstrated a strong correlation between survival and the extent of residual tumor after resection, although unfortunately biologic markers were not known in these children. These studies reported before all of the current biologic markers were available are difficult to interpret because resectability may have been an expression of the invasive biology of the tumor. Recent studies have suggested that incomplete surgical resection alone may be adequate therapy for biologically favorable tumors, but these reports need confirmation.

Histologically viable neuroblastoma is found in most tumors resected following preliminary chemotherapy, supporting the role of resection in this setting. No difference in survival was found by the author and associates between those children undergoing resection at presentation versus those receiving preliminary chemotherapy. The biologically favorable lesions may be more resectable. Recent studies have correlated the extent of neovascularity within the tumor evaluated by histologic studies and the overall prognosis and N-myc amplification of the tumor.

Thoracic lesions appear to have a favorable prognosis compared with other sites of primary disease. This has recently been correlated with a higher incidence of favorable biologic markers in tumors that occur at this site. Large tumors are also more readily resected in the chest than in the abdomen where they surround the aorta, vena cava, renal vessels, celiac axis, and the superior mesenteric artery. Intraspinal extension of the tumor presents a challenging problem (Fig. 4). Children may present with significant neurologic deficits. Modalities of treatment in this setting are emergent radiotherapy, laminectomy, and decompression, or chemotherapy. Laminectomy may be required if the tempo of the progression of symptoms is rapid or if the patient fails to improve following radiotherapy or chemotherapy. Spinal instability may result following laminectomy and require later spinal fixation. The French Society of Paediatric Oncology has reported excellent response to chemotherapy with both improvement in the neurologic symptoms and regression of tumor from the spinal canal.



Fig. 4. Thoracic MRI scan of a 5-year-old girl who presented with back pain. The chest radiograph revealed a right paravertebral mass and this scan demonstrated extension of the tumor through four neural foramen into the spinal canal. Tumor can be seen compressing the spinal cord to the left.

The chemotherapy utilized in neuroblastoma has been based on alkylating agents (cyclophosphamide) and doxorubicin or cisplatin and has been quite varied depending upon the institution or group conducting the trial. Many combinations have produced a decrease in the size of the primary and its vascularity, but comparison of the results is hampered by the multiple staging systems utilized and absence of information on the biologic markers in earlier studies (Fig. 5).

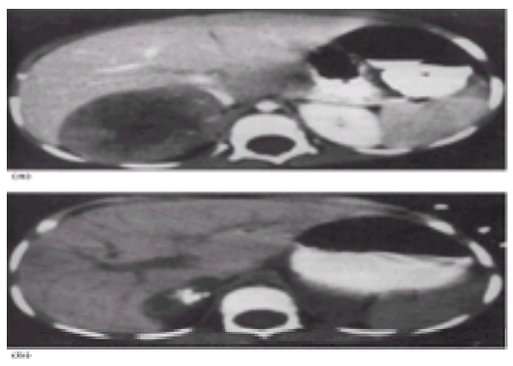


Fig. 5. (a) CT scan of a right adrenal neuroblastoma in a 1½-year-old girl presenting with stage D disease. (b) Follow-up study demonstrates marked resolution of the tumor and the development of calcification within the mass.

Stage 4 disease

Regrettably, the majority of infants and children present with metastatic disease and the initial role of surgery is to provide diagnostic tissue. Traditionally, this has been obtained by laparotomy or thoracotomy. Improved methods have been developed for percutaneous biopsy. When these latter methods are utilized, it is critical that adequate tissue be obtained for evaluation of the biologic markers and to provide complete histologic grading. Injudicious attempts at debulking extensive primaries which are often very friable and vascular may produce life-threatening hemorrhage or injury to vital organs.

Multiple single agents have been effective in neuroblastoma including cyclophosphamide, vincristine, imidazole carboxamide (DTIC), doxorubicin (Adriamycin), etoposide (VP-16), teniposide, ifosfamide, cisplatin, and carboplatin. Attempts at curing children with disseminated disease have utilized increasingly more aggressive combinations and dosages of these agents. Cell kinetic studies by Hayes have demonstrated a large proportion of non-proliferating tumor cells. This may explain the failure to achieve greater than a 10 per cent survival in children over a year of age with stage 4 (metastatic) disease.

Recent attempts to extend chemotherapy further have utilized ablative courses of agents and total abdominal irradiation followed by autologous bone marrow transplantation. This was first performed with monoclonal antibody purged bone marrow which was preserved and stored during ablation. Current trials obtain peripheral blood stem cells obtained by apheresis techniques during intervals of white cell recovery from preliminary chemotherapy. Stored cells are then infused after ablation. Although early complete remission rates of 20 to 30 per cent were reported, late relapses continue to occur and survival rates decline.

Most children with distant disease have large primaries. Those in the abdomen arising from adrenal or paravertebral primaries often encircle the celiac axis and superior mesenteric arteries (Fig. 6). Several studies have demonstrated that resection of extensive lesions can be best accomplished after initial treatment with chemotherapy. Preliminary treatment decreases the vascularity and friability of the tumor which facilitates its dissection from vessels and normal organs allowing safe resection. The evidence is conflicting in stage 4 disease between various studies that report improved survival after complete resection of the primary tumor and other studies which refute these findings. Consistent evidence does support a lower incidence of local recurrence after complete resection, not surprising given the high incidence of viable tumor on pathologic exam of tumors resected after initial chemotherapy. But children with stage 4 disease are plagued by distant relapse of persistent tumor to which the majority will succumb. Attempts should be made to preserve the ipsilateral kidney during the resection of the primary tumor in children with stage 4 neuroblastoma. Many of these children will receive either a bone marrow transplant or nephrotoxic agents making preservation of maximal renal function an important goal. Children with adrenal or paravertebral primaries generally have metastatic tumors in lymph nodes which wrap around the renal vessels. These must be dissected free to preserve the kidney. The risk of nephrectomy has been defined by the author in a cohort of children treated on Pediatric Oncology Group protocols to be greatest in children resected prior to the administration of chemotherapy.

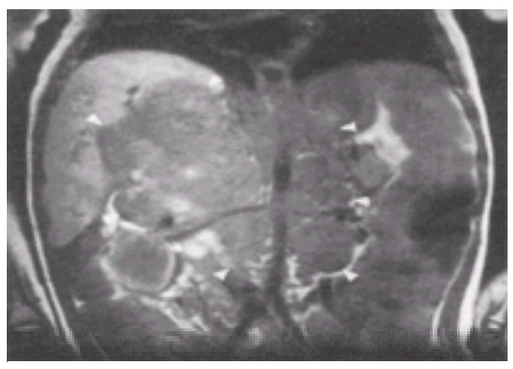


Fig. 6. MRI scan of a 30-month-old child who presented with a large right adrenal neuroblastoma with extensive lymph node metastases (arrows) surrounding the aorta, vena cava, and the origin of the celiac axis and the superior mesenteric artery.

Survival rates in children over 1 year of age with metastatic disease remain 10 to 20 per cent, a stark contrast with children with Wilms tumor. As treatment of distant sites of disease improves with biologic modifiers, such as 13 *cis*-retinoic acid, or immunologic techniques with specific monoclonal antibodies against the GD2 ganglioside, local control will become more critical.

Stage 4 (S) disease

Infants with stage 4 (S) disease require biopsy for diagnosis and analysis of biologic factors. In most cases, resection is not necessary unless the most readily obtained tissue for diagnosis is a small, easily resected primary. The tumor in infants with 4 (S) disease will spontaneously resolve in the vast majority of cases and long-term

survival is 75 to 80 per cent (Fig. 7). Tumor in infants with progressive disease has been particularly unresponsive to chemotherapy. There is no convincing evidence that resection of the primary will accelerate or insure resolution of metastatic disease. A recent evaluation of a large cohort of children with stage 4 (S) neuroblastoma showed an overall estimated 3-year survival of 85 per cent plus or minus 4 per cent. Survival was closely correlated with biologic markers as well as age. Children at 2 months of age or younger at presentation, had an overall survival of 71 per cent plus or minus 8 per cent, while survival in those with diploid ploidy is 68 plus or minus 12 per cent. Survival was most adversely affected by N-myc amplification (44 per cent plus or minus 33 per cent) and unfavorable histology (33 per cent plus or minus 19 per cent). Survival rates were comparable in this group whether or not the children receive adjuvant chemotherapy. There was no statistical difference in survival rates for infants who underwent complete resection of their primary tumor compared with those who underwent only partial resection or biopsy only, refuting claims to the contrary in smaller studies. If a mass remains at the primary site after resolution of the distant disease, resection is often performed and frequently demonstrates a ganglioneuroma or neuroblastoma with maturation.

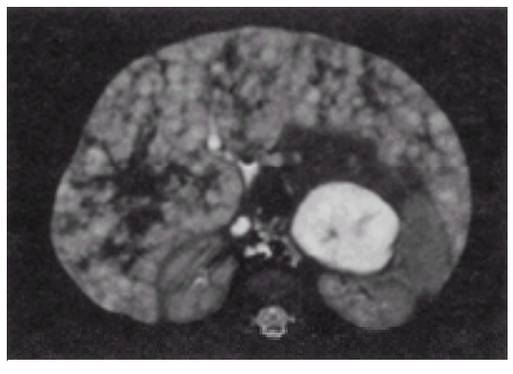


Fig. 7. MRI scan of a 3-month-old infant who presented with a large upper abdominal mass demonstrates extensive involvement of the liver with the tumor shown to be a neuroblastoma by biopsy. A follow-up scan 8 months later demonstrated complete resolution of the tumor.

Surgical techniques have been developed for abdominal expansion in infants presenting with extensive hepatomegaly that seriously compromises respiratory function. The abdominal fascia is divided and prosthetic materials are inserted to increase the volume of the abdominal cavity. In infants where spontaneous resolution has not occurred, the outcome is generally fatal.

Antenatal diagnosis

Antenatal diagnosis of neuroblastoma appears to identify a particularly favorable population of infants. Biologic markers on these tumors are uniformly favorable, and infants have done very well following resection of the primary tumor. It is not known whether spontaneous regression would occur if these infants were observed, but a recent report of four infants with antenatal diagnosis of adrenal masses (two solid and two cystic) demonstrated resolution of the abnormalities by 2.5 to 8 weeks of age.

Neuroblastoma screening trials

Infants have been screened for neuroblastoma in Japan since 1985 based on urinary catecholamine excretion (vanillylmandelic acid and homovanillic acid). While these trials have almost doubled the frequency of neuroblastoma identified in infants under a year of age, the majority of these infants have low-risk tumors which presumably would have undergone spontaneous regression. Similar results were reported from the Quebec screening project. Regrettably, the incidence of older children with biologically unfavorable tumors has not decreased leading a consensus conference reported by Murphy and coauthors to recommend against wide adoption of screening at this time.

Malignant liver tumors

Malignant tumors of the liver comprise only 2 per cent of malignancies in childhood with a population incidence of 1.5 cases per 1 million children in the United States. Hepatoblastoma and hepatocellular carcinoma are the predominant liver tumors in children comprising 65 and 25 per cent of the cases. Hepatoblastoma comprises the bulk of tumors in children younger than 15 years of age while hepatocellular carcinoma occurs in older children (Fig. 8). The median age at diagnosis for hepatoblastoma is approximately 1 year of age with 80 per cent of the cases in children under 3 years of age while for hepatocellular carcinoma the median age is approximately 12 years. Both tumors occur more frequently in boys with a 3:1 ratio. Hepatoblastoma-like Wilms tumor is associated with several clinical syndromes including hemihypertrophy, familial adenomatous polyposis, congenital absence of a kidney or adrenal gland, Beckwith–Wiedemann syndrome, and fetal alcohol syndrome. Just as hepatocellular carcinoma occurs with an increased frequency in adults with cirrhosis and chronic hepatitis, it is associated in pediatric patients with several underlying diseases, producing fibrosis and chronic inflammation including tyrosinemia, galactosemia, biliary atresia, progressive familial cholestatic cirrhosis, and giant cell hepatitis of infancy, Fanconi anemia, glycogen storage disease, a_1-antitrypsin deficiency and chronic hepatitis B. The incidence of hepatocellular carcinoma varies widely by geographic location. It is the third most common abdominal cancer in Japan, and it is seen with increased frequency in Africa and Asia where hepatitis B is endemic. Cytogenetic abnormalities identified in hepatoblastoma include trisomy for chromosome 2 or 20 and loss of heterozygosity of the short arm of chromosome 11.

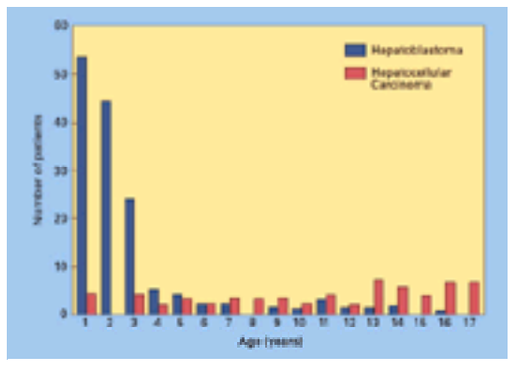


Fig. 8. This graph of the age distribution of hepatoblastoma and hepatocellular carcinoma collected from multiple pediatric series demonstrates the preponderance of hepatoblastoma in children under 3 years while HCC occurs more evenly throughout the 18 years.

Histology

Hepatoblastoma is divided into several subtypes including the epithelial tumors and the mixed epithelial and mesenchymal teratoid and non-teratoid tumors, fetal, embryonal, microtrabecular, and small cell undifferentiated. Fetal histology has the best prognosis in children with complete resection of these groups, although this finding was not confirmed by others. The histology of hepatocellular carcinoma (HCC) is identical with that occurring in adults. The fibrolamellar variant of HCC is generally felt to follow a more indolent course than the standard HCC.

Serum a-fetoprotein (AFP) is a useful marker for hepatoblastoma and is elevated in the majority (70 per cent) of cases. It is also elevated in 50 per cent of cases with HCC. Serial measurements of AFP are helpful to evaluate children for their response to therapy, as well as for residual or recurrent disease after resection. Rising AFP levels must stimulate a vigorous search for recurrent disease.

Presentation and evaluation

Most children with hepatic tumors present with an intra-abdominal mass or abdominal distension. Fever and failure to thrive with weight loss and anorexia, abdominal

pain, and vomiting are also seen, but it is rare for children to present with liver failure or jaundice.

Ultrasound is the appropriate first diagnostic study to identify the origin of the mass and whether it is solid or cystic. Identification of an intrahepatic mass should then be followed by CT or MRI of the liver to define further the anatomic involvement within the liver by the tumor, as well as a CT scan of the chest to evaluate for distant metastases. Bone marrow aspirate should also be obtained to evaluate for metastases. Angiography frequently used in the past is currently rarely required with the quality of other imaging modalities.

Radiographic imaging does not provide absolutely reliable assessment of resectability. The German Cooperative Liver Tumor Study demonstrated an error rate of 21 per cent using a combination of ultrasonography with either CT or MRI to determine resectability.

Comparison of treatment results is complicated in hepatoblastoma, as it is in neuroblastoma, by multiple staging systems. The German Cooperative Study Group, the Japanese Society of Pediatric Surgeons, SIOP all have different methods of staging while the United States Cooperative Groups have used a system based on the outcome of initial surgical procedure.

Treatment

Resection plays a major part in the treatment of hepatoblastoma and HCC, because long-term survival is rare without complete resection. Hepatoblastoma is generally unifocal, arising at one site within the liver parenchyma, whereas HCC is often multifocal. HCC has an invasive pattern of spread along the anatomic planes of the biliary system and within hepatic vessels. Complete resection of HCC is frequently difficult because of extensive involvement of the liver (Fig. 9). Hepatoblastoma remains localized in one lobe, but progressively distorts the liver anatomy as it grows. The resectability of a primary liver tumor is determined by its anatomic location within the liver and by its size. Surgical considerations are complex and best addressed by surgeons intimately familiar with hepatic anatomy and resection. To achieve a successful resection, perfusion and biliary drainage for an anatomic segment of the liver must be preserved. Initial surgical resection is preferred if it can be achieved safely. The surgical complications are more frequent in second-look than primary resections. However, the SIOP protocol recommends initial chemotherapy before attempted resection, which is consistent with their approach to other pediatric solid tumors.

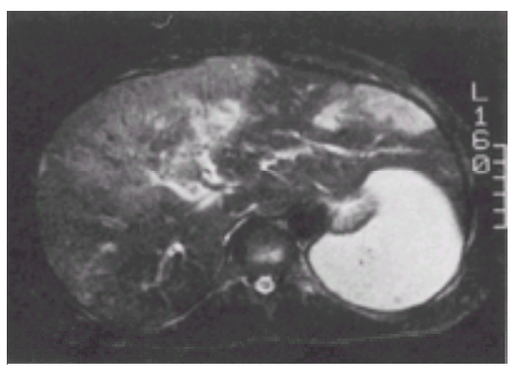


Fig. 9. MRI scan of a 7-year-old boy with HCC (which appears white in this scan) demonstrates the multifocal nature of the tumor with involvement of both the right and left lobes.

Combination chemotherapy has been defined to be particularly helpful in shrinking massive hepatoblastoma greatly facilitating surgical resection and as adjuvant therapy of completely resected tumors (Fig. 10). An early cooperative group study by CCG evaluated a 12-month course of doxorubicin, cyclophosphamide, vincristine, and 5-fluorouracil after surgical resection and showed an 83 per cent disease free-survival at 2 years in comparison with a 37 per cent survival in a preceding study where children received no adjuvant therapy. In a subsequent CCG trial initiated in 1986, the efficacy of continuous infusion doxorubicin and cisplatin was evaluated in 46 children with liver tumors (32 hepatoblastoma and 14 HCC). All children had either residual microscopic disease (10 children), gross residual tumor (25 children), or distant metastasis (11 children). A critical finding was that 20 of the 34 children with initially unresectable tumors had a favorable response to treatment and underwent delayed resection. Also 21 children (46 per cent of the entire group) remained disease free at 2 years and all except one remained disease free at 5 years. Survival was much better in the hepatoblastoma than the HCC patients; 63 versus 17 per cent survival.

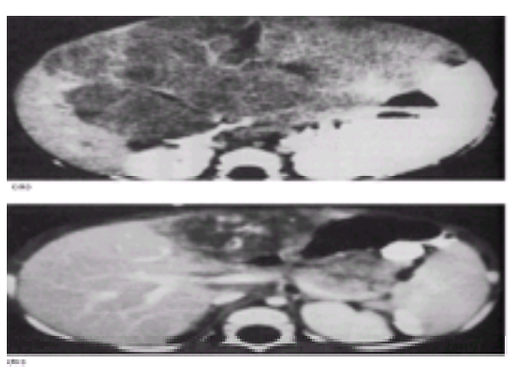


Fig. 10. (a) CT scan of an 8-month-old female who presented with a massive hepatoblastoma. Only the margin of the right lobe appears normal. (b) Dramatic response of the tumor shown on the follow-up scan 4 months later allowed successful resection of the tumor and the child remains disease free 10 years later.

Similar results were reported in a Pediatric Oncology Group study utilizing cisplatin, vincristine, and 5-fluorouracil. Delayed resection was achieved in 26 of 37 children (70 per cent) initially deemed to have unresectable tumors. A randomized comparison between cisplatin, vincristine, and 5-fluorouracil with that of doxorubicin and cisplatin was begun in 1989 by CCG and the Pediatric Oncology Group. It showed that the two combinations produced similar relapse-free and overall survival percentages. The cisplatin, vincristine, and 5-fluorouracil protocol had significantly less morbidity and fewer toxic deaths than in the doxorubicin/cisplatin group. The German Cooperative Pediatric Liver Tumor Study demonstrated the development of drug resistance by the tumors with six of 11 showing tumor growth if the children were treated with four or more courses before resection.

Response of HCC to current treatments remains dismal in pediatric as well as in adult patients as demonstrated by the United States and German Cooperative groups. Radiotherapy is not curative for hepatoblastoma or HCC because of dose limitations for hepatic tolerance. Excellent regeneration and function of the liver in children following resection has been confirmed in long-term follow-up studies by the author and associates.

Hepatic transplantation has been utilized by several centers for children with unresectable liver tumors with a survival rate of 50 per cent for 12 children with hepatoblastoma treated at multiple institutions. Of the children who died, three deaths resulted from progressive disease and three from postoperative complications. A single institution experience from Pittsburgh has been reported with 15 children transplanted over a 10-year period with HCC (nine) and hepatoblastoma (six). While five of six children with hepatoblastoma were alive and free of disease at an average follow-up of 21 months, only four of the nine with HCC were alive. Success with this treatment approach for patients with HCC has been extremely challenging. Disease must be limited to the liver for these efforts to be successful. Lymph node involvement is a contraindication for efforts at resection and transplantation.

Rhabdomyosarcoma

Rhabdomyosarcoma is the most common malignant tumor of the soft tissues in children comprising 50 per cent of all soft tissue sarcomas in infants and children. It accounts for 8 per cent of all cancers in children under 15 years of age and has an incidence of 4.5 cases per million. The median age at diagnosis is approximately 4 years. Rhabdomyosarcoma occurs with increased frequency in children from families with the Li-Fraumeni syndrome and in children with neurofibromatosis type I. The age of appearance depends upon the anatomic site. The most common primary site is the orbit accounting for over 10 per cent of all cases in children. Other characteristic sites are parameningeal, nasal, pharynx, extremity, and genitourinary tract including the bladder, prostate, vagina, uterus, cervix, and paratesticular

subgroups. There are two age peaks at 2 to 6 years and 15 to 19 years with the younger children having predominately head and neck and genitourinary primaries and the adolescents having primarily extremity, truncal, and paratesticular primaries.

Pathology

The traditional classification of rhabdomyosarcoma included divisions of pleomorphic, alveolar, and embryonal subtypes. Several other classifications have been proposed and a new system which is the result of an international collaborative study divides rhabdomyosarcoma into six subtypes including botryoid rhabdomyosarcoma, spindle cell rhabdomyosarcoma, embryonal rhabdomyosarcoma, alveolar rhabdomyosarcoma, undifferentiated sarcoma, and rhabdomyosarcoma with rhabdoid features. This system reported by Newton and colleagues appears to have greater reproducibility than the other classifications. Rhabdomyosarcoma tumors characteristically show some muscular differentiation which may be confirmed by either positive staining with periodic-Schiff techniques or with electron microscopy revealing intracytoplasmic filaments and Z-band material, or antibody staining against myosin, myoglobin, or desmin.

Clinical features

The multiplicity of sites of origin of rhabdomyosarcoma produces a great variety of presenting symptoms related primarily to the site of the tumor. While mass effect is the most common, ptosis and other neurologic symptoms are encountered from the parameningeal sites, and either urinary or bowel dysfunction are frequently produced by the genitourinary sites. Diagnostic studies are dictated by the site of origin, but evaluation of metastatic spread should include a chest radiograph and CT as well as a bone marrow aspiration and biopsy and radioisotope bone scintiscans. Bone marrow aspirate is reported at the time of diagnosis to be positive in approximately 10 per cent of children.

The staging systems are complicated based on the multiple sites of origin. A tumor staging (TNM) system has been found to be most effective and the Intergroup Rhabdomyosarcoma Study (IRS) protocols in the United States have been based on both that system as well as clinical group classification (Table 4).

Group	Residual tumor after resection
I	None (gross or microscopic)
IIa	Microscopic residual tumor at primary site; pathologic negative nodes
IIb	No residual tumor at primary site; pathologic positive nodes
IIc	Microscopic residual tumor at primary site and pathologic positive nodes
III	Gross residual tumor
IV	Distant metastasis

Table 4 IRS clinical group classification

Treatment

Rhabdomyosarcoma is a classic example of the role of multimodal therapy and treatment to cure solid tumors in children. The algorithms for treatment are tumor site specific. Chemotherapy, radiation, and often surgical resection play critical parts in therapy and their sequence is dependent upon the site of origin and initial staging. In some sites, such as the orbit, resection is rarely required and treatment with radiation and chemotherapy are often curative. In other sites, such as the bladder, prostate, and uterus, cure often requires a complete surgical resection. In the trunk and extremities, complete resection can obviate the need for radiation and will often enhance the chances of cure. Biopsy is required of tumors at all locations, but progressive improvement and response to chemotherapy and radiation has obviated the need for immediate surgical resection particularly in the orbit. The biopsy site and direction of the incision should always be planned with future resection in mind. Inadequate or inappropriate attempts at initial resection of an extremity or truncal lesion which leave positive margins can greatly complicate future resection.

The need for lymph node biopsy depends upon the primary site, but all enlarged lymph nodes should be biopsied. Children presenting without distant metastases have an overall 10 per cent incidence of lymphatic spread. The frequency is highest for the prostate (41 per cent), paratesticular (26 per cent), and genitourinary sites (24 per cent). Extremity lesions have an intermediate frequency of 12 per cent while orbit (0 per cent) and non-orbital head and neck sites (7 per cent) and truncal sites (3 per cent) have the lowest frequency.

Rhabdomyosarcoma is only moderately radiation sensitive, so high doses are required for tumor cell kill. The required dose is not modified by tumor size or the patient's age. The chemotherapy agents with efficacy in rhabdomyosarcoma are cyclophosphamide, actinomycin D, vincristine, doxorubicin, cisplatin, melphalan, ifosfamide, and etoposide (VP-16). The most reliable and well established regimen is VAC (vincristine, actinomycin D, and cyclophosphamide).

Prognosis

Collaborative studies, particularly those of the IRS have defined treatment parameters for rhabdomyosarcoma. Its multiplicity of sites of origin and stages have often produced confusing results. Overall survival rates have risen from a 2-year survival of approximately 20 per cent to over 65 per cent in the last 2 decades. Survival is based primarily on stage and also site. Tumors of the distal genitourinary tract and orbit have the best prognosis, while tumors in the extremities, trunk, and mucosal sites have the worse prognosis as do advanced stage or group tumors.

Specific primary tumor sites

Head and neck

Eye and orbit

Children with orbital rhabdomyosarcoma present with proptosis, due to a retrobulbar tumor, or swelling of the eyelid (Fig. 11). The radiographic evaluation must demonstrate whether the tumor is confined to the orbit or has extended anteriorly into the maxillary sinus or posteriorly into the ethmoid sinus and cranial cavity. CT provides excellent definition of the soft tissue mass, and can demonstrate the presence of bone estruction.

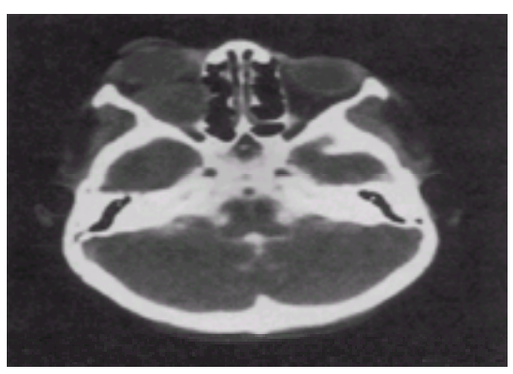


Fig. 11. A 6-year-old boy developed proptosis of the right eye and decreased visual acuity. The CT scan showed a large soft tissue mass in the right orbit. Biopsy demonstrated embryonal rhabdomyosarcoma.

Biopsy is required before beginning treatment of orbital tumors. Because of the excellent response to radiotherapy and chemotherapy, orbital exenteration is rarely required except for recurrent disease. Survival of more than 90 per cent of patients is standard.

Parameningeal

The parameningeal sites include the middle ear, nasopharynx, and adjacent areas. Children with a primary tumor in the middle ear present with a peripheral facial nerve palsy or a mass in the external auditory canal, and may have been misdiagnosed as having chronic otitis media or Bell's palsy. Those with a rhabdomyosarcoma of the nasopharynx present with signs of upper airways obstruction, which is frequently associated with cranial nerve palsies. On examination, they are found to have a soft tissue mass that is depressing the palate and extending into the retropharyngeal space.

Surgical resection is infrequently required at this site because of the excellent response to chemotherapy and radiotherapy. In children who have persistent disease after completion of radiotherapy or who have tumor recurrence, surgery should be considered. In some cases no viable tumor will be found within the residual mass. Modern methods of resection and reconstruction of these sites make surgery much less mutilating than in the past, and long-term functional results are good. Prior to the initiation of therapy, a lumbar puncture should be performed. A cytocentrifuge preparation of the cerebrospinal fluid should be examined for the presence of tumor cells.

Trunk

Primary tumors of the trunk may occur in the chest wall, abdominal wall, and paraspinal areas. These tumors often present as a localized swelling. The radiographic evaluation of a patient with a primary tumor of the trunk should include a plain chest radiograph and CT of the lesion and draining lymph nodes.

Lesions in the trunk arising in the chest or abdominal wall and paraspinal or retroperitoneal sites are best resected if possible. Tumor in these sites have a less favorable response to chemotherapy and radiotherapy than at other locations. They are also much less likely to maintain a durable response to chemotherapy than most other sites. Complete surgical removal is limited by the frequent large size of these lesions at diagnosis as well as their involvement of vital structures within the abdomen or chest. Resections of the chest or abdominal wall often require prosthetic mesh reconstruction. Primary re-excision should be considered if pathologic examination of the initial specimen demonstrates microscopic residual disease as tumors of the trunk have a higher rate of recurrence than extremity lesions. Of 76 chest wall tumors treated in IRS II and III, 70 per cent achieved a complete response. In the abdomen, 82 per cent achieved complete remission, and 46 per cent developed recurrent disease.

Extremity

Patients with rhabdomyosarcoma of an extremity usually present because of localized swelling or pain. Occasionally, a patient with rhabdomyosarcoma of an extremity will present because of symptoms caused by metastases, such as spinal cord compression secondary to vertebral body metastases, or pain, due to other bone metastases.

The radiographic evaluation of a patient with a rhabdomyosarcoma of the extremity should include plain radiographs of the primary tumor site and a bone scan. The presence of increased radionuclide uptake in the adjacent bone, although generally not associated with frank invasion of the bone by tumor, is correlated with the presence of inflammatory adhesions between the tumor and adjacent bone. Local recurrence of the tumor is likely if the tumor is not removed *en bloc* with the adjacent bone. CT and MRI are useful for defining the extent of the soft tissue mass and the presence of bone destruction. Arteriography may be necessary to evaluate the relationship between the tumor, contiguous muscle compartments, and their vascular supply.

Complete resection of the tumor with no microscopic margins is the goal in extremity sarcomas. There is no advantage to amputation or muscle group excision compared with local excision with an adequate surrounding rim of normal tissue, provided the resection results in negative microscopic margins. The extent of the resection is often tempered by attempts to minimize functional impairment. In extremity tumors, consideration of the initial biopsy site and direction of the incision are particularly important, because an imprecise biopsy can greatly complicate later resection. Extremity lesions should rarely be resected without an initial biopsy because the surgical approach when resecting a malignant lesion will be quite different from a benign lesion.

Extensive local lesions with invasion of vital structures are often treated first with chemotherapy and subjected to delayed surgical resection. The goal of resection is to render the child free of gross residual disease and microscopic residual disease. It has been demonstrated that microscopic disease can be controlled with a lower dose of radiotherapy (4000 cGy versus 5500 cGy) than can gross residual disease. This multidisciplinary approach will result in good local control and minimize morbidity of amputation or a more extensive initial resection.

Lymph node sampling is important in extremity rhabdomyosarcoma because of the significant risk of involvement (12 per cent). A representative sample of lymph nodes from the draining nodal group should be biopsied. A lymph node resection should not be performed, however, because lymphedema will complicate radiotherapy and subsequent surgical resection of the primary lesion.

Recent reports from IRS III show that the estimated 5-year survival percentage was directly related to the clinical group with a 95 per cent survival in group I versus 67 per cent in group II, 58 per cent in group III, and 33 per cent in group IV and was independent of histology or site. Multivariate analysis of pretreatment factors showed that lymph node metastasis, age over 10 years, and distant metastases predicted worse survival. A survival rate of 80 per cent in patients with negative lymph nodes contrasted with a 46 per cent rate when they were positive. The difference between the survival percentage of 51 group II and III patients with lymph node negative distal extremity lesions who could have been made group I with extensive resection and that of patients with group I distal extremity tumors approached statistical significance in univariate analysis ($p < 0.06$). All four group III patients who had re-excision of margins survived while only 18 of 31 without re-excision survived.

Genitourinary

Patients with rhabdomyosarcoma of the paratesticular tissue present with painless enlargement of the scrotum. The age distribution is bimodal within the pediatric age range with peaks at 5 and 16 years of age. Children with rhabdomyosarcoma of the prostate present most frequently with acute urinary retention or with difficulty in voiding (Fig. 12). Hematuria, dysuria, and abdominal swelling are also observed. Those with rhabdomyosarcoma of the bladder usually present because of hematuria, urinary frequency, or dysuria (Fig. 13). The physical examination of children with a prostate tumor is unremarkable except for the presence of a mass between the bladder and rectum that is palpable on rectal examination. The radiographic evaluation should include a voiding cystourethrogram and CT of the abdomen and pelvis. These studies will demonstrate displacement of the bladder, in the case of a prostatic primary tumor, or the presence of multiple polypoid masses within the bladder with thickening of the bladder wall in the case of a primary bladder tumor. Children with sarcoma botryoides of the vagina frequently present with a polypoid mass protruding from the vaginal orifice or with vaginal bleeding. These children should have an AFP level obtained prior to tumor biopsy or excision, because yolk sac tumor also may present as a vaginal mass.

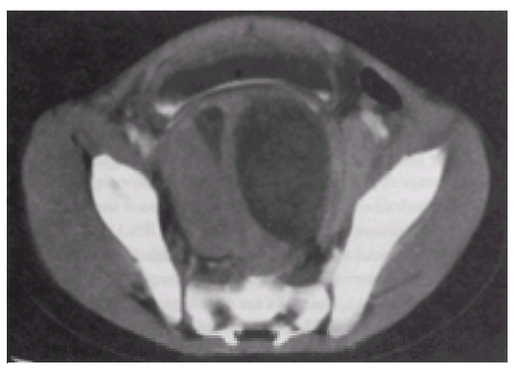


Fig. 12. This 5-year-old boy presented with a 2-week history of leg pain and constipation and 1-day of dysuria. On physical exam, lower abdominal fullness was noted with a prerectal mass. Ultrasound and this CT scan revealed a large heterogeneous mass in the region of the prostate indenting the bladder base. The cystic area suggested necrosis. A cystoscopic biopsy confirmed the diagnosis of embryonal rhabdomyosarcoma.

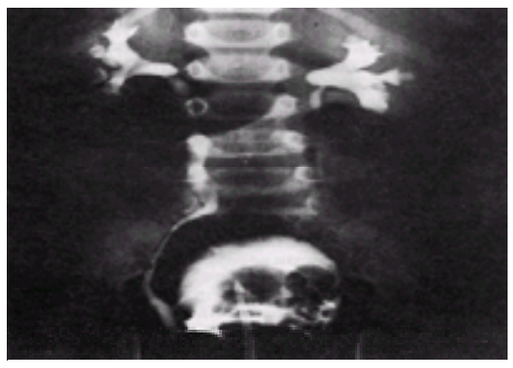


Fig. 13. A 3-year-old boy developed dysuria and gross hematuria. Ultrasound and this urogram demonstrated many smooth rounded intraluminal filling defects in the bladder arising from its base. Sarcoma botryoides with embryonal histology was confirmed on biopsy.

Anterior pelvic exenteration is rarely required today. Bladder, prostate, and vaginal primaries are first biopsied and lymph node extension is defined. Chemotherapy and radiotherapy are then utilized in an effort to minimize residual disease. In the IRS I and II studies there were 28 vaginal and 10 uterine lesions. Vaginal lesions generally arise from the anterior vaginal wall in the area of the embryonic vesico-vaginal septum. The mean age of children with vaginal tumors was under 2 years. Of the 21 children with localized disease initially treated with chemotherapy, 17 had tumor resection following 8 to 48 weeks of therapy. Hysterectomy with partial vaginectomy was required in 13 children and vaginectomy (partial or total) alone in four children. There were five local failures in the group.

Girls in the IRS III study were treated with cytoreductive chemotherapy after initial biopsy. Of 24 girls with vaginal primaries, 20 had an initial biopsy with gross residual disease. Three had resection with positive margins and one had metastatic disease. Following chemotherapy, seven had resection by partial or complete vaginectomy, and six had no viable tumor identified. Resection required cystectomy in one and hysterectomy in five. Seventeen girls are free of disease 66 to 108 months after diagnosis. Ten of these 17 girls were treated by biopsy and chemotherapy alone; four received radiotherapy.

The uterine primary tumors are distinct from the vaginal lesions because they occur in older girls and have a propensity toward local recurrence. Uterine primaries were seen in 10 girls in IRS I and II. The tumor mass was a single, large polyp on a stalk protruding through the cervices of six girls. Four had intramural involvement of the cervix or body of the uterus. The former frequently responded to local therapy of dilatation and curettage with preservation of the uterus. The latter were more aggressive and required hysterectomy for cure in those girls without metastatic disease.

Fourteen girls were treated on the IRS III and IV pilot protocols for uterine lesions. Three presented with cervical lesions that were completely excised. Eight girls had initial biopsy with gross residual disease (group III), and three had metastatic disease at presentation (group IV). Five girls received primary chemotherapy or chemotherapy and radiotherapy. Two of these five had delayed hysterectomy and vaginectomy, one had no further surgery, and two had exploration with no evidence of disease. There were no deaths or relapses in this group. One child underwent initial resection of a broad ligament mass, but experienced an early recurrence while receiving chemotherapy. She then progressed to metastatic disease and death. Four girls died of chemotherapy toxicity or sepsis, emphasizing the risks of this aggressive chemotherapy. However, 89 per cent of these girls survived without evidence of disease 1.5 to 6 years after diagnosis (excluding those who died of treatment-related causes).

Preservation of the bladder and prostate was one of the primary goals of the IRS II study in which surgical therapy was shifted from initial resection to primary biopsy with subsequent radiotherapy or surgery. The 3-year survival rate of patients on IRS II (70 per cent) was similar to that of IRS I (78 per cent), but the 3-year disease-free survival rate was lower (52 per cent versus 70 per cent). In IRS II there were 95 children with bladder/prostate tumors. Twenty-two per cent retained their bladder and were alive 3 years after presentation. This was comparable with the 23 per cent survival figures for the 66 children with bladder or prostate tumors in IRS I. Sequential treatment on this protocol failed to improve bladder salvage. The rate of retention of a functional bladder at 4 years from diagnosis was 60 per cent and the mortality, excluding children presenting with disseminated disease, was 10 per cent among children treated on IRS III.

Paratesticular rhabdomyosarcoma should be resected by a transinguinal radical orchiectomy. Testicular masses should never be approached through the scrotum because of the risk of producing spread through the inguinal lymph nodes (the drainage route of the scrotum) and the inability to achieve a high inguinal ligation of the spermatic cord. A total of 121 boys with paratesticular rhabdomyosarcoma treated on IRS III had retroperitoneal lymph node dissection to evaluate nodal status. Lymph nodes were negative based on the finding of CT in 81 per cent of the boys, but 14 per cent with negative CT scans had positive nodes when biopsy or retroperitoneal lymph node dissection was performed. Of the clinically positive boys, 94 per cent were confirmed to be positive pathologically. Retroperitoneal relapse occurred in only two of the 121 boys, one of whom had pathologically negative lymph nodes and did not receive radiotherapy. While CT was very accurate if lymph node abnormalities were identified, it was not extremely sensitive in identifying nodal involvement.

Germ cell tumors

The term germ cell tumor encompasses a spectrum of tumors which can arise in multiple anatomic locations. The unifying factor in this family of tumors is their presumed origin from the totipotent primitive germ cell which explains both the multiple sites and histologies of the tumors. Germ cells arise in the yolk sac endoderm and can be identified in the caudal portion of the yolk sac by the 24th day. They migrate by ameboid movement to the fetal hindgut by the end of the fifth week of gestation. From there, they continue to travel along the hindgut mesentery to the gonadal ridge which will form the ovary or testes. Residual germ cells deposited along this migratory pathway account for the distribution of germ cell tumors from the base of the brain to the neck, anterior mediastinum, retroperitoneum, and coccyx as well as within the gonads. Origin from totipotent germ cells also explains the broad diversity of histologic patterns which these tumors demonstrate. Germ cell tumors as a group account for 3 per cent of malignant tumors in children and adolescents with an annual incidence of 4.2 cases per million children under 15 years of age. A two- to fourfold predominance is seen in girls.

Tumors arising in extra gonadal sites are often identified at birth while the gonadal tumors appear later. Symptoms at presentation may result from the size of the tumor producing pain or a palpable mass. Occasionally, hormone-producing lesions such as the dysgerminomas will produce precocious development of secondary sexual characteristics or masculinization. Serum markers are very helpful in some tumors to assess both clinical response, extent of resection, and recurrence of tumor. AFP is the primary serum protein in the fetus. It is produced initially by the fetal yolk sac and in later development by the liver. It has a peak concentration at 12 to 14 weeks of gestation and then declines to adult levels by 12 months. It is frequently elevated in the yolk sac tumors as well as previously mentioned in hepatoblastoma and HCC. b-Chorionic gonadotropin is a glycoprotein produced by the placental syncytiotrophoblasts. It is composed of a- and b-protein subunits which resemble the structure of the pituitary gonadotropins. The beta subunit can be quantitatively assayed. Its elevation suggests the presence of syncytiotrophoblasts in seminomas, dysgerminomas, choriocarcinoma, and occasionally embryonal carcinoma.

Pathology

Germ cell tumors include multiple pathologic types (Table 5). Teratomas contain embryonic elements derived from more than one of the three primary germ cell layers and are frequently arranged in a chaotic manner. The tissues may range from mature well differentiated to immature to those with malignant elements (Table 6). Teratomas contain tissue elements foreign to the anatomic site or organ in which they are located. The most primitive germ cell tumors are the germinomas often referred to as seminomas when arising in the testes, dysgerminomas when in the ovary, and germinoma when arising from extra-gonadal sites. They comprise 10 per cent of all ovarian tumors and 15 per cent of all germ cell tumors and often present as bulky encapsulated solid tumors. It is the principal malignant tumor found in dysgenetic gonads. They are bilateral in 5 to 30 per cent of cases. Histologically, germinomas are composed of uniform large round or polyhedral cells with distinct borders and clear or granular cytoplasm resembling primordial germ cells with a large centrally located nucleus. The cells are supported by delicate connective tissue stroma. Endodermal sinus tumors (yolk sac tumors) can occur alone or as a component of mixed germ cell tumors. They comprise the most common of the malignant germ cell tumors in children and adolescents. In malignant sacrococcygeal tumors, it is the most frequent component and in older children the ovary is the most frequent site. Testicular endodermal sinus tumors show two peaks of incidence, one in infancy and the other in adolescence. Yolk sac tumors have intracellular hyaline droplets which are periodic-acid Schiff-positive and resist digestion with diastase. The droplets contain AFP as well as other proteins. The most common microscopic arrangement of cells is a papillary pattern which may display characteristic perivascular sheaths of cells, the so-called endodermal sinus structure, or Schiller–Duval bodies. Embryonal carcinoma is often found in association with other germ cell tumors in a mixed lesion and rarely occurs alone as a pure lesion. Embryonal carcinoma is composed of cells that resemble epithelial cells. They have considerable variation in their size, shape, and arrangement and may be large and pleomorphic without distinct cell borders. Frequently small or large acinar, tubular and papillary structures are formed, but the cells may occur as solid sheets. Hemorrhage and necrosis are frequent. Choriocarcinoma is rare in the pediatric age range, but is highly malignant when it occurs. Young infants may present with disseminated disease resulting from a placental primary that has invaded villous vessels and the fetal blood stream. In older children it is seen primarily in the mediastinum and gonads. Histologically, cytotrophoblasts and syncytiotrophoblasts are seen adjacent to dilated vascular sinusoids. Immunoperoxidase staining is positive for beta chorionic gonadotropin.

Gonadoblastoma arises primarily in children with dysgenetic gonads. Scully reported that the majority arise in phenotypic females who are chromatin negative and the most common karyotypes were a 46 XY or 45 X/46 XY.

Germinoma	
	dysgerminoma
	seminoma
Embryonic differentiation (embryonal carcinoma)	
Somatic differentiation	
	mature teratoma
	immature teratoma
	malignant teratoma
Extra-embryonic differentiation	
	endodermal sinus tumor, yolk sac tumor
	trophoblast choriocarcinoma

Table 5 Classification of germ cell tumors based on embryologic relationships

Grade I	All tissues mature with no mitotic activity
Grade I	Neoplasms with some immature tissue but absent or rare neuroepithelial components
Grade II	Neoplasms where immaturity and neuroepithelial elements present to a greater degree than in grade I
Grade III	Neoplasms with prominent immaturity and neuroectodermal elements

Table 6 Grading system for immature teratomas

Staging

Children with germ cell tumors must be evaluated for metastases to regional lymph nodes, liver, lungs, bones, or brain. Chest radiographs and CT scans as well as bone scan and bone marrow aspirate and biopsies are appropriate. Tumors of the ovary and testis are often evaluated by ultrasound and CT of the abdomen to define the primary mass and evaluate for lymph node involvement (Table 7, Table 8, and Table 9).

Stage
I Tumor confined to testis, epididymis, and spermatic cord; complete resection by radical inguinal orchiectomy; no tumor rupture or biopsy; scrotum not violated; hilar vessels not involved; tumor markers normal after appropriate half-life decline
II Microscopic residual disease; gross total resection with microscopic residual disease at margins of resection and/or regional lymph node involvement; hilar vessels not involved; tumor markers normal after appropriate half-life decline
III Retroperitoneal lymph node involvement, gross or microscopic residual disease
IV Distant metastatic disease

Table 7 Intergroup (Pediatric Oncology Group and CCG) staging system for testicular germ cell tumors

Stage	Extent of disease
I	Limited to ovaries; complete resection; tumor markers fall to normal
II	Microscopic residual or positive retroperitoneal nodes < 2 cm in diameter; markers ±
III	Gross residual or biopsy only; positive nodes > 2 cm; contiguous visceral involvement; or positive peritoneal washings
IV	Distant metastatic disease

Table 8 Clinicopathologic staging of ovarian germ cell tumors by Pediatric Oncology Group and CCG

Stage	Extent of disease
I	Complete resection by any appropriate procedure; negative tumor margins; tumor markers negative after half-life decline; lymph nodes not positive by CT or MRI; negative diagnostic laparotomy and/or negative biopsy
II	Microscopic residual disease; negative tumor markers; positive regional nodes
III	Gross residual disease; or biopsy only; or unresectable disease; negative or positive tumor markers; positive regional nodes
IV	Tumor metastatic, including liver

Table 9 Staging for extragonadal tumors

Treatment

Testes

All scrotal masses should be explored through an inguinal incision after an initial ultrasound examination and blood samples are obtained for tumor markers. The differential diagnosis includes epididymitis (rare in prepubertal males), hydrocele, testicular torsion, orchitis, leukemia, lymphoma, and paratesticular rhabdomyosarcoma or a sex cord tumor. Of the germ cell tumors, the majority (80 per cent) are malignant. Most frequently the endodermal sinus tumor. The tumor is confined to the testes in the majority (80 to 85 per cent) of children under 4 years of age. Transcrotal biopsy contaminates the scrotum and its lymphatic drainage to the inguinal lymph nodes and prevents high ligation of the spermatic cord. High ligation of the spermatic cord should be performed at the internal ring if the tumor is clearly malignant. In cases where it is uncertain, the cord is clamped at the internal ring with a non-crushing vascular clamp and the testicle is elevated out of the scrotum on the spermatic cord. The testicle is walled off and an incisional biopsy performed. The cord should remain clamped until the diagnosis is obtained by frozen section. If malignancy is present, then a radical orchiectomy should be performed without releasing the clamp. Infants have a predominance of early stage lesions which are primarily endodermal sinus tumors (yolk sac tumors) in contrast with teenage boys in whom embryonal carcinoma or mixed germ cell tumors predominate. Teenagers frequently have a delayed presentation, resulting in a higher proportion of advanced stage disease. In most series infants with clinical stage I endodermal sinus tumors with no evidence of spread and complete resection are treated by radical orchiectomy and close follow-up. In the United Kingdom Children's Cancer Group Study of malignant germ cell tumors, 87 per cent of the boys presented with stage I testicular tumors and were treated by orchiectomy alone. These were predominately yolk sac tumors. In seven boys a rising serum AFP was the only evidence for incomplete resection and all responded to chemotherapy. Survival using this protocol was 100 per cent. A smaller series from the Institute Gustav-Roussy, reported recurrence in two of 12 boys followed without adjuvant chemotherapy or retroperitoneal lymph node resection. Both were cured by surgery and chemotherapy. Children with yolk sac tumors of the testis rarely require a bilateral retroperitoneal lymph node dissection for adequate staging. The procedure has a substantial risk of producing impotence and retrograde ejaculation and is usually negative. Only two of 49 bilateral retroperitoneal lymph node dissections and four of 29 unilateral retroperitoneal lymph node dissections performed in children with yolk sac tumors of the testis were positive. The percentage of positive retroperitoneal lymph node dissections was not influenced by the age of the patient. The procedure infrequently provides information that will change the stage or influence therapy. Retroperitoneal lymph node biopsies should be considered in boys with a normal AFP level before orchiectomy, boys with prolonged elevation of the AFP level following orchiectomy and normal chest CT and bone scan, and in boys with extensive residual disease. Although adequate data are lacking, this subgroup of patients may be one in which there will be a greater yield from abdominal exploration. In a modified retroperitoneal lymph node dissection, the lymph nodes on the ipsilateral side are resected from the inguinal canal to above the level of the renal vessels and the dissection is carried to the contralateral side at that level, avoiding bilateral dissection below the pelvic brim.

Embryonal carcinoma in a teenage boy is very similar to that in the adult. Adults with clinical stage I embryonal carcinoma of the testes will have positive retroperitoneal lymph nodes in approximately 30 per cent of cases. Adults after radical orchiectomy with stage I disease have been observed. Twenty-eight per cent relapsed, which was consistent with the expected incidence of lymph node involvement. A similar approach of observation in clinical stage I children was employed at the St. Jude Children's Research Hospital. Relapse with a pulmonary metastasis occurred in one of eight boys. In many current protocols, children with clinical stage I disease (i.e. no radiographically identifiable retroperitoneal tumor and falling serum markers) are followed after radical orchiectomy. Children with identifiable retroperitoneal tumor frequently receive initial chemotherapy with surgery reserved for residual masses or persistently elevated tumor markers. Residual masses in adults resected after chemotherapy contained necrotic tissue in 45 per cent of the cases. Forty-two per cent had teratoma, and only 12.5 per cent contained viable germ cell tumor. The completed Intergroup Germ Cell Study of the Pediatric Oncology Group and the Children's Cancer Group followed this approach.

Ovary

Girls with an ovarian mass present most frequently with an abdominal mass or pain. It may become severe if torsion occurs. Germ cell tumors account for the vast majority of benign and malignant ovarian tumors in children with a small component (10 to 15 per cent) composed of sex-cord tumors. This is in marked contrast with adults where the epithelial neoplasms predominate. Preoperative evaluation includes ultrasound or abdominal CT scans (Fig. 14 and Fig. 15). A plain radiograph will demonstrate presence of calcification frequently seen in teratomas. Evaluation for metastatic disease is appropriate once the tumor is known to be malignant. Surgical exploration of an ovarian mass must accomplish two goals: resection of the primary tumor and adequate staging. The peritoneal fluid should be aspirated for cytology, although it is much less likely to be positive in children and adolescents than in older women with epithelial tumors. The omentum is rarely involved and routine removal is unnecessary. The iliac, low periaortic or pericaval lymph nodes, and nodes at the level of the renal vessels should be biopsied. Lymph node biopsies should be obtained on all girls in whom an ovarian mass is removed unless it is clearly cystic and benign as the gross appearance of a benign and malignant teratoma can be quite similar. Careful inspection of the contralateral ovary is important as 10 per cent of teratomas will be bilateral.

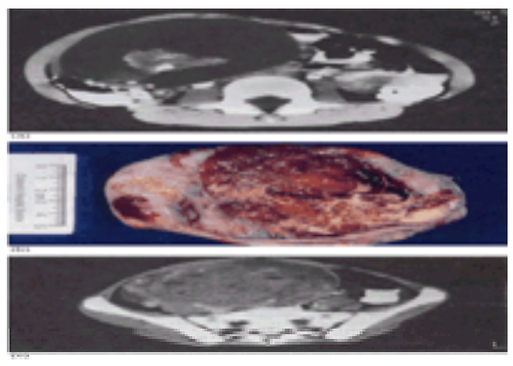


Fig. 14. (a) CT scan of a 5-year-old girl who presented with increasing abdominal girth noted by her father showed a primarily cystic tumor with a solid component with calcification. Pathology revealed a mature teratoma with various tissue types present (b). (c) CT scan of an 8-year-old girl with similar complaints reveals a primarily solid ovarian mass, also a mature teratoma.

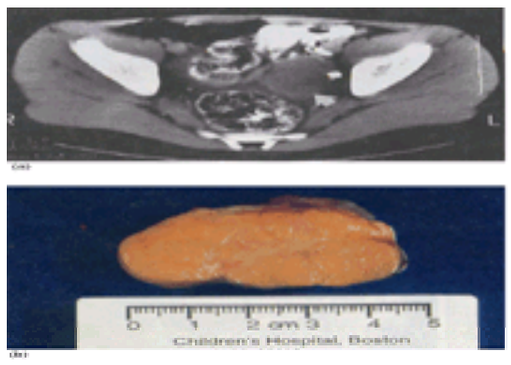


Fig. 15. An ultrasound obtained on an 8-year-old girl who presented with a 1-year history of abdominal pain demonstrated a left adnexal mass. (a) The CT scan demonstrated a homogeneous mass (arrow). (b) Cross-section of the ovarian mass demonstrated the characteristic lobulated pattern without necrosis of a dysgerminoma.

Sacrococcygeal teratomas

The sacrococcygeal teratoma comprises 40 to 50 per cent of all germ cell tumors and is frequently identified in the newborn child. They have a 2.3:1 female prevalence. The differential diagnosis of presacral lesions includes neuroblastoma, chordoma, ependymoma, anterior meningocele, and rectal duplication. Sacrococcygeal teratomas are classified according to their anatomic location (Fig. 16). Congenital anomalies of the vertebra, genitourinary system, or anorectum occur in children with sacrococcygeal teratomas. The vast majority of sacrococcygeal teratomas are benign when identified in the newborn. The frequency of malignancy increases as they are identified in older children rising to 50 per cent at 2 years of age. The incidence of malignancy in the four types of sacrococcygeal teratomas correlates with the age at which they are most frequently diagnosed. The type I tumor is exophytic, invariably diagnosed in the newborn, and has a low rate of malignancy (8 per cent). The type II tumors, which have internal as well as external components, have a 21 per cent incidence of malignancy. Those with very

extensive intra-abdominal components (type III) have a 34 per cent incidence and those with type IV, which often have a delayed diagnosis, will have a 38 per cent incidence (Fig. 17).

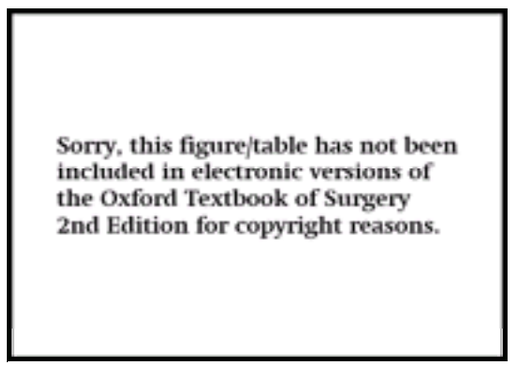


Fig. 16. Schematic depiction of the anatomic classification of sacrococcygeal teratomas.

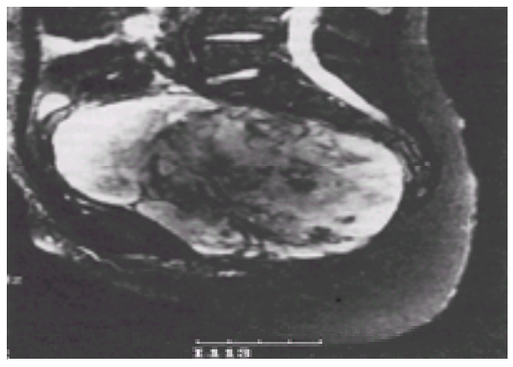


Fig. 17. MRI scan of an 8-year-old girl who presented with months of constipation and then acute urinary retention. A large pelvis mass was shown to be an immature sacrococcygeal teratoma (type IV).

The goal for surgery of sacrococcygeal teratomas is complete removal of the tumor (Fig. 18). CT scans may be helpful in some cases to define the extent of the tumor as well as possible invasion of adjoining structures. The coccyx from which these tumors arise must be excised to minimize local recurrence. Rates as high as 35 per cent local recurrence are reported if it is left unresected. Approximately 10 per cent of infants will require a combined perineal and abdominal approach to resect the tumor.



Fig. 18. This figure demonstrates a typical sacrococcygeal teratoma in a newborn infant. Note the displacement of the anus by the mass. This was an entirely mature teratoma.

In malignant lesions, resection is most readily accomplished following initial chemotherapy. The efficacy of current cisplatin-based therapy will often result in remarkable resolution of the tumor and result in only residual ‘benign' teratomatous remnants at the time of resection. The serum AFP level may be elevated and provide a good marker for recurrence in children with malignant tumors.

Teratoma—other extra-gonadal sites

The anterior superior mediastinum is the second most common extragonadal site. Teratomas at this site have a male predominance and children present primarily with respiratory symptoms of a cough, wheezing, dyspnea, or pain. A chest CT scan will best define the extent and resectability of the lesion and its relation to surrounding structures. Tumor markers will help distinguish germ cell tumors of the mediastinum from the other tumors which occur at this anatomic site. The retroperitoneum is the third site in order of frequency and teratoma at this site are identified most frequently infants under 2 years of age. Children present primarily with a mass or gastrointestinal symptoms. Intracranial and cervical lesions are less frequent and are identified by neurologic symptoms or as a mass. Intracranial teratomas occur primarily in the pineal region, but are also seen in suprasellar sites. The majority are benign in infancy, but malignant in older children.

Treatment—chemotherapy

Major advances have occurred in the chemotherapy of germ cell tumors. Whereas children cannot be universally cured, the vast majority of children are, even with advanced stage disease. Current therapy was developed founded on the adult experience. Treatment is based on cisplatin, etoposide and bleomycin.

Further reading

Wilms tumor

Beckwith JB. Wilms' tumor and other renal tumors of childhood: A selective review from the National Wilms' Tumor Study Pathology Center. *Human Pathology* 1983; **14**: 481–92.

Beckwith JB, Palmer NF. Histopathology and prognosis of Wilms tumor. Results from the First National Wilms' Tumor Study. *Cancer* 1978; **41**: 1937–48.

Beckwith JB, Kiviat NB, Bonadio JF. Nephrogenic rests, nephroblastomatosis, and the pathogenesis of Wilms' tumor. *Pediatric Pathology* 1990; **10**: 1–36.

Bennington JL, Beckwith JB. Tumors of the kidney, renal pelvis, and ureter. *Atlas of tumor pathology*, Second Series, Fascicle 12. Armed Forces Institute of Pathology, Washington DC, 1975.

Blute ML *et al.* Bilateral Wilms tumor. *Journal of Urology* part 2 1987; **138**: 968–73.

Burger D *et al.* The advantages of preoperative therapy in Wilms' tumour. A summarized report on clinical trials conducted by the International Society of Paediatric Oncology (SIOP). *Zeitschrift für Kinderchirurgie* 1985; **40**: 170–5.

Coppes MJ, Haber DA, Grundy PE. Genetic events in the development of Wilms' tumor. *New England Journal of Medicine* 1994; **331**: 586–90.

Cozzi F *et al.* Enucleative surgery for stage I nephroblastoma with a normal contralateral kidney. *Journal of Urology* 1996; **156**: 1788–93.

Craft AW, Parker L, Stiller C, Cole M. Screening for Wilms' tumour in patients with aniridia, Beckwith syndrome or hemihypertrophy. *Medical Pediatric Oncology* 1995; **24**: 231–4.

D'Angio GJ *et al.* The treatment of Wilms' tumor: Results of the National Wilms' Tumor Study. *Cancer* 1976; **38**: 633–46.

D'Angio GJ *et al.* The treatment of Wilms' tumor. Results of the Second National Wilms' Tumor Study. *Cancer* 1981; **47**: 2302–11.

D'Angio GJ *et al.* Treatment of Wilms' tumor. Results of the Third National Wilms' Tumor Study. *Cancer* 1989; **64**: 349–60.

DeChiara TM, Robertson EJ, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 1991; **64**: 849–59.

Dinkel E *et al.* Renal growth in patients nephrectomized for Wilms tumour as compared to renal agenesis. *European Journal of Pediatrics* 1988; **147**: 54–8.

Farber S. Chemotherapy in the treatment of Wilms' tumor. *Journal of the American Medical Association* 1966; **198**: 826–36.

Faria P *et al.* Focal versus diffuse anaplasia in Wilms tumor-new definitions with prognostic significance. A report from the National Wilms Tumor Study Group. *American Journal of Surgical Pathology* 1996; **20**: 909–20.

Goleta-Dy A, Shaw PJ, Stevens MM. RE: The necessity of contralateral surgical exploration in Wilms tumor with modern noninvasive imaging technique. *Journal of Urology* 1992; **147**: 171–74.

Green DM *et al.* The role of surgical excision in the management of relapsed Wilms' tumor patients with pulmonary metastases: A report from the National Wilms' Tumor Study. *Journal of Pediatric Surgery* 1991; **26**: 728–33.

Green DM, Breslow NE, Beckwith JB, Norkool P. Screening of children with hemihypertrophy, aniridia, and Beckwith–Wiedemann syndrome in patients with Wilms tumor: A report from the National Wilms Tumor Study. *Medical Pediatric Oncology* 1993*a*; **21**: 188–92.

Green DM *et al.* Treatment outcomes in patients less than 2 years of age will small, stage I, favorable-histology Wilms' tumors: A report from the National Wilms' Study Group. *Journal of Clinical Oncology* 1993*b*; **11**: 91–5.

Green DM *et al.* Treatment of children with stages II to IV anaplastic Wilms' tumor: A report from the National Wilms' Tumor Study Group. *Journal of Clinical Oncology* 1994*a*; **12**: 2126–31.

Green DM *et al.* Treatment of children with clear-cell sarcoma of the kidney: A report from the National Wilms' Tumor Study Group. *Journal of Clinical Oncology* 1994*b*; **12**: 2132–7.

Grundy P *et al.* Chromosome 11 uniparental isodisomy predisposing to embryonal neoplasms. *Lancet* 1991; **338**: 1079–80.

Grundy PE *et al.* Loss of heterozygosity for chromosomes 16q and 1p in Wilms' tumors predicts an adverse outcome. *Cancer Research* 1994; **54**: 2331–3.

Haas JE, Palmer NF, Weinberg AG, Beckwith JB. Ultrastructure of malignant rhabdoid tumor of the kidney. A distinctive renal tumor of children. *Human Pathology* 1981; **12**: 646–57.

Henry I *et al.* Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature* 1991; **351**: 665–7.

Horwitz JR *et al.* Renal salvage procedures in patients with synchronous bilateral Wilms' tumors: A report from the National Wilms' Tumor Study Group. *Journal of Pediatric Surgery* 1996; **31**: 1020–5.

Hrabovsky EE *et al.* Wilms' tumor in the neonate: A report from the National Wilms' Tumor Study. *Journal of Pediatric Surgery* 1986; **21**: 385–7.

Koo AS *et al.* RE: The necessity of contralateral surgical exploration in Wilms tumor with modern noninvasive imaging technique: A reassessment. *Journal of Urology*, part 2 1990; **144**: 416–17.

Larsen E *et al.* Surgery only for the treatment of patients with stage I (Cassady) Wilms' tumor. *Cancer* 1990; **66**: 264–6.

Lemerle J *et al.* Preoperative versus postoperative radiotherapy, single versus multiple courses of actinomycin D, in the treatment of Wilms' tumor. Preliminary results of a controlled clinical trial conducted by the International Society of Paediatric Oncology (SIOP). *Cancer* 1976; **38**: 647–54.

Lemerle J *et al.* Effectiveness of preoperative chemotherapy in Wilms' tumor: Results of an International Society of Paediatric Oncology (SIOP) Clinical Trial. *Journal of Clinical Oncology* 1983; **1**: 604–9.

Maheswaran S *et al.* Physical and functional interaction between WT1 and p53 proteins. *Proceedings of the National Academy of Sciences, USA* 1993; **90**: 5100–4.

Marsden HB, Lawler W, Kumar PM. Bone metastasizing renal tumor of childhood. Morphological and clinical features, and differences from Wilms' tumor. *Cancer* 1978; **42**: 1922–8.

McLorie GA *et al.* Reduction in tumor burden allowing partial nephrectomy following preoperative chemotherapy in biopsy proved Wilms tumor. *Journal of Urology* 1991; **146**: 509–13.

Miller RW, Fraumeni JF, Manning MD. Association of Wilms' tumor with aniridia, hemihypertrophy and other congenital malformations. *New England Journal of Medicine* 1964; **270**: 922–7.

Moorman-Voestermans CG *et al.* Is partial nephrectomy appropriate treatment for unilateral Wilms' tumor? *Journal of Pediatric Surgery* 1998; **33**: 165–70.

Mushtaq I, Carachi R, Roy G, Azmy A. Childhood renal tumours with intravascular extension. *British Journal of Urology* 1996; **78**: 772–6.

Paulino AC *et al.* Local control in synchronous bilateral Wilms tumor. *International Journal of Radiation, Oncology, Biology, and Physics* 1996; **36**: 541–8.

Rauscher FJ, III. The WT1 Wilms tumor gene product: a developmentally regulated transcription factor in the kidney that functions as a tumor suppressor. *FASEB Journal* 1993; **7**: 896–903.

Riccardi VM, Sujansky E, Smith AC, Francke U. Chromosomal inbalance in the aniridia-Wilms' tumor association: 11p interstitial deletion. *Pediatrics* 1978; **61**: 604–10.

Ritchey ML *et al.* Surgical complications after nephrectomy for Wilms' tumor. *Surgery, Gynecology and Obstetrics* 1992; **175**: 507–14.

Ritchey ML *et al.* Small bowel obstruction after nephrectomy for Wilms' tumor. A report of the National Wilms' Tumor Study-3. *Annals of Surgery* 1993*a*; **218**: 654–9.

Ritchey ML *et al.* Preoperative therapy for intracaval and atrial extension of Wilms tumor. *Cancer* 1993*b*; **71**: 4101–10.

Ritchey ML *et al.* Management and outcome of inoperable Wilms tumor. A report of National Wilms Tumor Study-3. *Annals of Surgery* 1994; **220**: 683–90.

Ritchey ML *et al.* Neonatal Wilms' tumor. *Journal of Pediatric Surgery* 1995*a*; **30**: 856–9.

Ritchey ML *et al.* Accuracy of current imaging modalities in the diagnosis of synchronous bilateral Wilms' tumor. A report from the National Wilms' Tumor Study Group. *Cancer* 1995*b*; **75**: 600–4.

Ritchey ML *et al.* Renal failure in Wilms' tumor patients: A report from the National Wilms' Tumor Study Group. *Medical Pediatric Oncology* 1996; **26**: 75–80.

Shamberger RC, Roger MM, Sallan SE. Recent experience with Wilms' tumor: 1978–1991. *Annals of Surgery Oncology* 1994; **1**: 59–65.

Shamberger RC *et al.* Surgery-related factors and local recurrence of Wilms tumor in National Wilms Tumor Study-4. *Annals of Surgery* 1999; **229**: 292–7.

Thorner PS, Arbus GS, Celermajer DS, Baumal R. Focal segmental glomerulosclerosis and progessive renal failure associated with a unilateral kidney. *Pediatrics* 1984; **73**: 806–10.

Tournade MF *et al.* Results of the Sixth International Society of Pediatric Oncology Wilms' tumor trial and study: A risk-adapted therapeutic approach in Wilms' tumor. *Journal of Clinical Oncology* 1993; **11**: 1014–23.

Urban CE *et al.* Partial nephrectomy in well-responding stage I Wilms' tumors: Report of three cases. *Pediatric Hematology and Oncology* 1995; **12**: 143–52.

Voute PA *et al.* Preoperative chemotherapy (CT) as first treatment in children with Wilms' tumor. Results of the SIOP nephroblastoma trials and studies. *Proceedings of ASCO* 1987; **6**: 223.

Waziri M, Patil SR, Hanson JW, Bartley JA. Abnormality of chromosome 11 in patients with features of Beckwith–Wiedemann syndrome. *Journal of Pediatrics* 1983; **102**: 873–6.

Weeks DA, Beckwith JB, Mierau GW, Luckey DW. Rhabdoid tumor of kidney. A report of 111 cases from the National Wilms' Tumor Study Pathology Center. *American Journal of Surgical Pathology* 1989; **13**: 439–58.

Wiedemann HR. Tumours and hemihypertrophy associated with Wiedemann-Beckwith syndrome. *European Journal of Pediatrics* 1983; **141**: 129.

Wilimas JA *et al.* Is renal salvage feasible in unilateral Wilms' tumor?: Proposed computed tomographic criteria and their relation to surgicopathologic findings. *American Journal of Pediatric Hematology and Oncology* 1990; **12**: 164–7.

Zuppan CW *et al.* The effect of preoperative therapy on the histologic features of Wilms' tumor. An analysis of cases from the Third National Wilms' Tumor Study. *Cancer* 1991; **68**: 385–94.

Neuroblastoma

Azizkhan RG, Shaw A, Chandler JG. Surgical complications of neuroblastoma resection. *Surgery* 1984; **97**: 514–17.

Brodeur GM *et al.* Revision of the international criteria for diagnosis, staging, and response to treatment in patients with neuroblastoma. *Journal of Clinical Oncology* 1988; **6**: 1874–881.

Cecchetto G *et al.* The role of surgery in non-localized neuroblastoma. Analysis of 59 cases. *Tumori* 1983; **69**: 327–9.

Garaventa A *et al.* Localized but unresectable neuroblastoma: Treatment and outcome of 145 cases. *Journal of Clinical Oncology* 1993; **11**: 1770–9.

Haase GM *et al.* Improvement in survival after excision of primary tumor in Stage III neuroblastoma. *Journal of Pediatric Surgery* 1989; **24**: 194–200.

Haase GM *et al.* Aggressive surgery combined with intensive chemotherapy improves survival in poor-risk neuroblastoma. *Journal of Pediatric Surgery* 1991; **26**: 1119–24.

Haase GM *et al.* Surgical management and outcome of locoregional neuroblastoma: Comparison of the Childrens Cancer Group and the International Staging System. *Journal of Pediatric Surgery* 1995; **30**: 289–95.

Hata Y *et al.* Kidney-preserving radical tumor resection in advanced neuroblastoma. *Journal of Pediatric Surgery* 1989; **24**: 382–5.

Hayes FA, Green AA, Mauer AM. Correlation of cell kinetics and clinical response to chemotherapy in disseminated neuroblastoma. *Cancer Research* 1977; **37**: 3766–770.

Hoffer FA, Chung T, Diller L, Kozakewich H, Fletcher JA, Shamberger RC. Percutaneous biopsy for prognostic testing of neuroblastoma. *Radiology* 1996; **200**: 213–16.

Ho PT *et al.* Prenatal detection of neuroblastoma: A ten year experience from the Dana-Farber Cancer Institute and Children's Hospital. *Pediatrics* 1993; **92**: 358–64

Ikeda H *et al.* Sites of relapse in patients with neuroblastoma following bone marrow transplantation in relation to preparatory ‘debulking' treatments. *Journal of Pediatric Surgery* 1992; **27**: 1438–41.

Katzenstain HM *et al.* Prognostic significance of age, MYCN oncogene amplification, tumor cell ploidy, and histology in 110 infants with Stage D(S) neuroblastoma: The Pediatric Oncology Group experience. A Pediatric Oncology Group Study. *Journal of Clinical Oncology* 1998; **16**: 1–13.

Kiely EM. The surgical challenge of neuroblastoma. *Journal of Pediatric Surgery* 1994; **29**: 128–33.

Lee EW, Applebaum H. Abdominal expansion as a bridging technique in Stage IV-S neuroblastoma with massive hepatomegaly. *Journal of Pediatric Surgery* 1994; 29: 1470–71.

Le Tourneau JN *et al.* Evaluation of the role of surgery in 130 patients with neuroblastoma. *Journal of Pediatric Surgery* 1985; **20**: 244–9.

Look AT *et al.* Clinical relevance of tumor cell ploidy and n-myc gene amplification in childhood neuroblastoma: A Pediatric Oncology Group Study. *Journal of Clinical Oncology* 1991; **9**: 581.

Martinez DA *et al.* Resection of the primary tumor is appropriate for children with Stage IV-S neuroblastoma: An analysis of 37 patients. *Journal of Pediatric Surgery* 1992; **27**: 1016–21.

Matsumura M *et al.* An evaluation of the role of surgery in metastatic neuroblastoma. *Journal of Pediatric Surgery* 1988; **23**: 448–53.

Matthay KK. Neuroblastoma: Biology and therapy. *Oncology* 1997; **11**: 1857–70.

Matthay KK *et al.* Successful treatment of Stage III neuroblastoma based on prospective biologic staging: A Children's Cancer Group Study. *Journal of Clinical Oncology* 1998; **16**: 1256–64.

Seeger RC *et al.* Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastoma. *New England Journal of Medicine* 1985; **313**: 1111–16.

Shamberger RC *et al.* Surgical management of Stage III and IV neuroblastoma: Resection before or after chemotherapy. *Journal of Pediatric Surgery* 1991; **26**: 1113–18.

Shamberger RC *et al.* The risk of nephrectomy during local control in abdominal neuroblastoma. *Journal of Pediatric Surgery* 1998; **33**: 161–4.

Shimada H *et al.* Identification of subsets of neuroblastoma by combined histopathologic and N-myc analysis. *Journal of the National Cancer Institute* 1995; **87**: 1470–6.

Sitarz A *et al.* An evaluation of the role of surgery in advanced neuroblastoma: A report from the Children's Cancer Group. *Journal of Pediatric Surgery* 1983; **18**: 147–51.

Tsuchida Y *et al.* Therapeutic significance of surgery in advanced neuroblastoma: A report from the study group of Japan. *Journal of Pediatric Surgery* 1992; **27**: 616–22.

Malignant liver tumors

Conran RM, Hitchock CL, Waclawiw MA, Stocker JT, Ishak KG. Hepatoblastoma: The prognostic significance of histologic type, *Pediatric Pathology* 1992; **12**: 167–83.

Douglass EC, Reynolds M, Finegold M, Cantor AB, Glicksman A. Cisplatin, vincristine and fluorouracil therapy for hepatoblastoma: a pediatric oncology group study. *Journal of Clinical Oncology* 1993; **11**: 96–9.

Evans AE, Land VJ, Newton WA, Randolph JG, Sather HN, Tefft M. Combination chemotherapy (vincristine, adriamycin, cyclophosphamide, and 5-fluorouracil) in the treatment of children with malignant hepatoma. *Cancer* 1982; **50**: 821–6.

Filler RM, Ehrlich PF, Greenberg ML, Babyn PS. Preoperative chemotherapy in hepatoblastoma. *Surgery* 1991; **110**: 591–7.

Haas JE *et al.* Histopathology and prognosis in childhood hepatoblastoma and hepatocarcinoma, *Cancer* 1989; **64**: 1082–95.

Habrand JL, Pritchard J. Letter to the Editor. Role of radiotherapy in hepatoblastoma and hepatocellular carcinoma in children and adolescents. *Pediatric Oncology* 1991; **19**: 208.

Hata Y. The clinical features and prognosis of hepatoblastoma: follow-up studies done on pediatric tumors enrolled in the Japanese Pediatric Tumor Registry between 1971 and 1980, part I, *Japanese Journal of Surgery* 1990; **20**: 498–502.

King DR *et al.* The surgical management of children with incompletely resected hepatic cancer is facilitated by intensive chemotherapy. *Journal of Pediatric Surgery* 1991; **26**: 1074–81.

Koneru B *et al.* Liver transplantation for hepatoblastoma. *Annals of Surgery* 1991; **213**: 118–21.

Ni YH *et al.* Hepatocellular carcinoma in childhood: Clinical manifestations and prognosis. *Cancer* 1991; **68**: 1737–41.

Ortega JA *et al.* Effective treatment of unresectable or metastatic hepatoblastoma with cisplatin and continuous infusion doxorubicin chemotherapy: a report from the Children's Cancer Study Group. *Journal of Clinical Oncology* 1991; **9**: 2167–76.

Reynolds M, Douglass EC, Finegold M, Cantor A, Glicksman A. Chemotherapy can convert unresectable hepatoblastoma. *Journal of Pediatric Surgery* 1992; **27**: 1080–4.

Tagge EP *et al.* Resection, including transplantation, for hepatoblastoma and hepatocellular carcinoma: Impact on survival. *Journal of Pediatric Surgery* 1992; **27**: 292–7.

Von Schweinitz D, Bürger D, Mildenberger H. Is lapartotomy the first step in treatment of childhood liver tumors?-The experience from the German Cooperative Pediatric Liver Tumor Study HB-89. *European Journal of Pediatric Surgery* 1994; **4**: 82–6.

Weinblatt ME, Siegel SE, Siegel MM, Stanley P, Weitzman JJ. Preoperative chemotherapy for unresectable primary hepatic malignancies in children. *Cancer* 1982; **50**: 1061–4

Rhabdomyosarcoma

Abramson DH *et al.* The treatment of orbital rhabdomyosarcoma with irradiation and chemotherapy. *Ophthalmology* 1979; **86**: 1330.

Andrassy RJ *et al.* Conservative surgical management of vaginal and vulvar pediatric rhabdomyosarcoma: A report from the Intergroup Rhabdomyosarcoma Study III. *Journal of Pediatric Surgery* 1995; **30**: 1034–7.

Andrassy RJ *et al.* Extremity sarcomas: an analysis of prognostic factors from the Intergroup Rhabdomyosarcoma Study (IRS) III. *Journal of Pediatric Surgery* 1996; **31**: 191–6.

Brodeur GM *et al.* Revisions of the International criteria for neuroblastoma diagnosis, staging, and response to treatment. *Journal of Clinical Oncology* 1993; **11**: 1466–77.

Corpron CA *et al.* Conservative management of uterine pediatric rhabdomyosarcoma: A report from the Intergroup Rhabdomyosarcoma Study III and IV Pilot. *Journal of Pediatric Surgery* 1995; **30**: 942–4.

Crist W *et al.* The third Intergroup rhabdomyosarcoma study (IRS) III. *Journal of Clinical Oncology* 1995; **13**: 610–30.

Forbes GS, Ernest F IV, Waller RR. Computed tomography of orbital tumors, including late-generation scanning techniques. *Radiology* 1982; **142**: 387–94.

Hays DM. Bladder/prostate rhabdomyosarcoma: results of the multi-institutional trials of the Intergroup Rhabdomyosarcoma Study. *Seminars on Surgical Oncology* 1993; **9**: 520

Hays DM *et al.* Clinical staging and treatment results in rhabdomyosarcoma of the female genital tract among children and adolescents. *Cancer* 1988; **61**: 1893–903.

Hays DM *et al.* Primary reexcision for patients with 'microscopic residual' tumor following initial excision of sarcomas of the trunk and extremity sites. *Journal of Pediatric Surgery* 1989; **24**: 5–10.

LaQuaglia MP. Extremity rhabdomyosarcoma. Biological principles, staging, and treatment. *Seminars on Surgical Oncology* 1993; **9**: 510–19.

Lawrence W *et al.* Lymphatic metastases with childhood rhabdomyosarcoma: A report from the Intergroup Rhabdomyosarcoma Study. *Cancer* 1987; **60**: 910–15.

Lawrence W Jr, Hays DM, Heyn R, Beltangady M, Maurer HM. Surgical lessions from the intergroup rhabdomyosarcoma study (IRS) pertaining to extremity tumors. *World Journal of Surgery* 1988; **12**: 676–84.

Lee FA. Rhabdomyosarcoma. In: Parker BR, Castellino RA, eds. *Pediatric oncologic radiology*, p. 417. CV Mosby, St. Louis, 1977.

Lobe TE *et al.* The argument for conservative, delayed surgery in management of prostatic rhabdomyosarcoma. *Journal of Pediatric Surgery* 1996; **31**: 1084–7.

Pratt CB *et al.* Factors leading to delay in diagnosis and affecting survival of children with head and neck rhabdomyosarcoma. *Pediatrics* 1978; **61**: 30–4.

Raney RB Jr *et al.* Paratesticular rhabdomyosarcoma in children. *Cancer* 1978; **42**: 729–36.

Raney RB *et al.* Primary chemotherapy with or without radiation therapy and/or surgery for children with localized sarcoma of the bladder, prostate, vagina, uterus, and cervix. *Cancer* 1990; **66**: 2072–81.

Simon MA, Enneking WF. The management of soft-tissue sarcomas of the extremities. *Journal of Bone and Joint Surgery* 1976; **58A**: 317.

Weiner ES. Head and neck rhabdomyosarcoma. *Seminars on Pediatric Surgery* 1994; **3**: 203.

Wiener ES *et al.* Retroperitoneal node biopsy in paratesticular rhabdomyosarcoma. *Journal of Pediatric Surgery* 1994; **29**: 171–8.

Germ cell tumors

Ablin AR *et al.* Results of treatment of malignant germ cell tumors in 93 children: a report from the Children's Cancer Study Group. *Journal of Clinical Oncology* 1991; **9**: 1782.

Altman RP, Randolph JG, Lilly JR. Sacrococcygeal teratoma: American Academy of Pediatrics Surgical Section Survey-1973. *Journal of Pediatric Surgery* 1974; **9**: 389–98.

Aprikian AG *et al.* Resection of postchemotherapy residual masses and limited retroperitoneal lymphadenectomy in patients with metastatic testicular nonseminomatous germ cell tumors. *Cancer* 1994; **74**: 1329–34.

Brosman SA. Testicular tumors in prepubertal children. *Urology* 1979; **13**: 581–8.

Crist W *et al.* The Intergroup Rhabdomyosarcoma Study-III, *Journal of Clinical Oncology* 1995; **13**: 610.

Ein SH, Darte JM, Stephens CA. Cystic and solid ovarian tumors in children: A 44 year review. *Journal of Pediatric Surgery* 1970; **5**: 148–56.

Exelby PR. Testicular cancer in children. *Cancer* 1980; **45**: 1803–9.

Fernandes ET *et al.* Two decades of experience with testicular tumors in children at St Jude Children's Research Hospital. *Journal of Pediatric Surgery* 1989; **24**: 677–82.

Flamant F *et al.* Optimal treatment of clinical Stage I yolk sac tumor of the testis in children. *Journal of Pediatric Surgery* 1986; **21**: 108–11.

Gribbon M, Ein SH, Mancer K. Pediatric malignant ovarian tumors: A 43-year review. *Journal of Pediatric Surgery* 1992; **27**: 480–4.

Hawkins EP *et al.* Nongerminomatous malignant germ cell tumors in children: A review of 89 cases from the Pediatric Oncology Group. *Cancer* 1971–84; **58**: 2579–84.

Hawkings E *et al.* Occult malignancy in neonatal sacrococcygeal teratomas: A report from a Combined Pediatric Oncology Group and Children's Cancer Group Study. *American Journal of Pediatric Hematology and Oncology* 1993; **15**: 406–9.

Hoskin P *et al.* Prognostic factors in Stage I non-seminomatous germ-cell testicular tumors managed by orchiectomy and surveillance: Implications for adjuvant chemotherapy. *Journal of Clinical Oncology* 1986; **4**: 1031–6.

Kaplan GW *et al.* Genitourinary tumors: Prepubertal yolk sac testicular tumors—Report of the Testicular Tumor Registry. *Journal of Urology*, part 2 1988; **140**: 1109–12.

Mann JR *et al.* Results of the United Kingdom Children's Cancer Study Group's malignant germ cell tumor studies. *Cancer* 1989; **63**: 1657–67.

Marina N *et al.* Treatment of childhood germ cell tumors. *Cancer* 1992; **70**: 2568–75.

Newton WA Jr *et al.* Classification of rhabdomyosarcomas and related sarcomas. Pathologic aspects and proposal for a new classification—An Intergroup Rhabdomyosarcoma Study. *Cancer* 1995; **76**: 1073.

Norris HJ, Zirkin HW, Benson WL. Immature (malignant) teratoma of the ovary: A clinical and pathologic study of 58 cases. *Cancer* 1976; **37**: 2359–72.

Noseworthy J *et al.* Sacrococcygeal germ cell tumors in childhood: An updated experience with 118 patients. *Journal of Pediatric Surgery* 1981; **16**: 358–64.

Pappo AS, Shapiro DN, Crist WM, Maurer HM. Biology and therapy of pediatric rhabdomyosarcoma. *Journal of Clinical Oncology* 1995; 13: 2123–139.

Raney RB *et al.* Improved prognosis with intensive treatment of children with cranial soft tissue sarcomas arising in nonorbital parameningeal sites. A report from the Intergroup Rhabdomyosarcoma Study. *Cancer* 1987; **59**: 147.

Rescorla FJ, Sawin RS, Coran AG, Dillon PW, Azizkhan RG. Long-term outcome for infants and children with sacrococcygeal teratoma: A report from the Children's Cancer Group. *Journal of Pediatric Surgery* 1998; **33**: 171–6.

Scully RE. Gonadoblastoma: A gonadal tumor related to the dysgerminoma (seminoma) and capable of sex-hormone production. *Cancer* 1953; **3**: 455–63.

Scully RE. Gonadoblastoma: A review of 74 cases. *Cancer* 1970; **25**: 1340–55.

42.11 Fetal surgery

Craig T. Albanese and Michael R. Harrison

The risks and benefits
Fetal surgical techniques
Open hysterotomy procedures
Minimally invasive procedures (fetal endosurgery)
Postoperative management of mother and fetus
Preterm labor
Fetal problems amenable to surgical correction before birth
Urinary tract obstruction
Congenital cystic adenomatoid malformation
Congenital diaphragmatic hernia
Sacrococcygeal teratoma
Fetal airway obstruction
Myelomeninogocele
Summary
Further reading

Over the past two decades, sophisticated ultrasonographic imaging and fetal sampling techniques have had a tremendous impact on the study of abnormal fetal development. Since the first reports of *in utero* ultrasonographic diagnosis of congenital anomalies in the 1970s, increasingly accurate prenatal ultrasonography has identified a growing number of disorders that are amenable to prenatal surgical correction. As a result, we can now view the fetus as a patient.

Although most prenatally diagnosed malformations are best managed by appropriate medical and surgical therapy after planned delivery near term, an increasing number of simple anatomic abnormalities with predictably devastating developmental consequences have been successfully corrected before birth. In the 1980s, the pathophysiology of several potentially correctable fetal lesions was elucidated, and the feasibility of fetal repair established using a variety of animal models. Concomitantly, the natural history of these abnormalities was determined by serial ultrasonographic observation of human fetuses; selection criteria for prenatal intervention, and anesthetic, tocolytic, and surgical techniques for hysterotomy and fetal surgery, were developed. Prenatal diagnosis has defined a ‘hidden mortality rate' for some lesions, such as congenital diaphragmatic hernia, bilateral hydronephrosis, sacrococcygeal teratoma, and congenital cystic adenomatoid malformation of the lung. These lesions, when first evaluated and treated postnatally, demonstrate a favorable selection bias. The most severely affected fetuses often die *in utero* or immediately after birth. In the 1990s, this investment in basic and clinical research has benefited an increasing number of fetal patients.

The risks and benefits

For the fetus, the risk of the procedure is weighed against the benefit of correction of a lethal or debilitating defect. However, the risks and benefits for the mother are more difficult to assess. Maternal safety is paramount, since most fetal malformations do not directly threaten the mother's health, yet she must bear significant risk and discomfort from the surgical procedure and the postoperative tocolytic therapy.

Since open fetal surgery has rarely been attempted elsewhere, the 75 cases performed at the University of California, San Francisco, Fetal Treatment Center through June 1997 provide the best data on maternal outcome (Table 1). There have been no maternal deaths and few postoperative maternal complications but considerable morbidity, primarily related to preterm labor and its treatment. There were no infections. Eleven patients required blood transfusions. In our early experience, two patients developed leaks of amniotic fluid through the hysterotomy incision requiring repair, and five patients developed such leaks from the vagina. All patients experienced labor after hysterotomy. Eleven patients developed pulmonary edema while receiving high doses of tocolytic drugs. Although reversible, this complication emphasized the need for close monitoring under intensive care. Since the midgestation hysterotomy is usually not in the lower uterine segment, delivery after fetal surgery and all future deliveries should be by cesarean section. Despite this recommendation, five women went into labor against medical advice and had a dehiscence through the uterine scar; uterine closure and neonatal outcome were excellent in all cases. Finally, the ability to carry and deliver subsequent pregnancies does not appear to be jeopardized by fetal surgery. Since 1981, we have follow-up data from 35 mothers who have attempted pregnancy after fetal surgery: 32 conceived and delivered 30 normal children (two have yet to deliver); of the three who could not conceive, two had a strong history of infertility and the third had only attempted for 6 months.

Variable		Median	Range
Maternal age (years)		27.0	18–43
Gestational age of fetus (weeks)		26.0	17–38
Operative time, total (min)		135.4	69–303
Operative time, fetal repair (min)		32.0	5–92
Blood loss (ml)		400.0	150–2500
Interval to delivery (weeks)		4.5	1–15
Subsequent pregnancy history	(No.)		
Term cesarean delivery, healthy	(32)*		
Currently pregnant	(3)		
Not attempted, not desired, too early, etc.	(29)		

*University of California, San Francisco, Fetal Treatment Center, 75 cases through June 1997.
**Three women had two subsequent pregnancies each.
Adapted from: Harrison MR. Fetal surgery. *American Journal of Obstetrics and Gynecology* 1998; 178: 1188–94.

Table 1 Maternal outcome with open and endoscopic fetal surgery*

Fetal surgical techniques

Numerous technical aspects of fetal surgical procedures have evolved over 15 years of experimental and clinical work. In the operating room, the mother is positioned with left uterine displacement to avoid compression of the inferior vena cava by the gravid uterus, and she and her baby are anesthetized with a halogenated agent. Maternal monitoring is accomplished with routine noninvasive monitors plus central venous and arterial catheters. There are two surgical approaches to the fetus: one involves opening the uterus and delivering the fetal part to be repaired; the other employs minimally invasive techniques.

Open hysterotomy procedures

The uterus is exposed through a low, transverse abdominal incision. Ultrasonography is used to localize the placenta and inject the fetus with a narcotic and a paralytic agent. Depending on the position of the placenta, the uterus is opened either anteriorly or posteriorly, using absorbable staples that provide hemostasis and seal the membranes. Specially designed back-biting retractors, along with the staples, provide uterine hemostasis and aid in exposure of the pertinent fetal part. A pulse oximeter records the fetal pulse rate and oxygen saturation intraoperatively. Warm lactated Ringer's solution is continuously infused around the fetus. After fetal repair, a radiotelemeter is implanted in a submuscular infra-axillary pocket before closing the fetal incision; this provides postoperative monitoring of the fetal heart rate, temperature, and amniotic pressure. The uterine incision is closed with two layers of absorbable sutures and fibrin glue. Amniotic fluid is restored with warm lactated Ringer's solution.

Minimally invasive procedures (fetal endosurgery)

The mother is placed in low lithotomy position with left uterine displacement (Fig. 1). Intraoperative sonography maps the position of the placenta and guides trocar placement. Usually three balloon/compression trocars or radially expanding sleeves are used to perform fetal endosurgical procedures; 3- and 5-mm instruments are used. To optimize visibility, the amniotic fluid is exchanged for lactated Ringer's solution and the operative field is continuously washed using a heated pump irrigation system via the sheath of the 12° 5-mm telescope. The fetus is monitored by transuterine ultrasonography. A miniaturized telemeter that can be placed into the amniotic cavity via a trocar site has been developed and is presently being tested. Each uterine puncture site is closed with one or two absorbable sutures and fibrin glue.

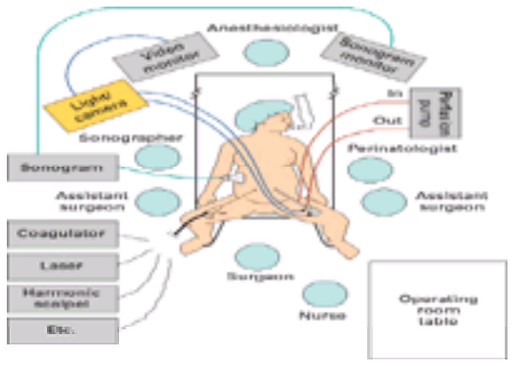


Fig. 1. Maternal positioning and videoscopic set-up for fetal endoscopic surgery.

Postoperative management of mother and fetus

Maternal arterial pressure, central venous pressure, urine output, and oxygen saturation are continuously monitored. Fetal well-being and uterine activity are recorded externally by tocodynamometer and by a radiotelemeter, implanted at surgery, that continuously records the fetal pulse, temperature, and maternal intra-amniotic pressure. Maternal arterial and venous catheters are not required since high-dose tocolytic agents are rarely necessary. Patient-controlled analgesia and/or continuous epidural analgesics ease maternal stress and aid tocolysis. When labor is controlled and the fetus stable (mean 5 days for endoscopic procedures and 13 days for open procedures), the patient is discharged. Outpatient monitoring and tocolysis continue; fetal sonograms and stress testing are performed at least weekly. Cesarean delivery is performed when the membranes rupture or labor cannot be controlled.

Preterm labor

Breeching the uterus, whether by puncture or incision, incites uterine contractions. In spite of technical advances, preterm labor is the Achilles' heel of fetal therapy. Although lessened by fetal endoscopic procedures, it still occurs in 100 per cent of patients. The regimen of preoperative indomethacin, intraoperative deep halogenated inhalation anesthesia, and postoperative indomethacin, magnesium sulfate, and b-mimetics, which was first perfected in sheep and monkeys, has proved inadequate for extensive procedures in humans. Although halogenated inhalation agents provide satisfactory anesthesia for mother and fetus, the depth of anesthesia required to achieve intraoperative uterine relaxation can produce fetal and maternal myocardial depression and affect placental perfusion. Indomethacin can constrict the fetal ductus arteriosus; the combination of magnesium sulfate and b-mimetics can produce maternal pulmonary edema. Fluid restriction to avoid this complication can compromise the maternal–placental–fetal circulation and contribute to recalcitrant preterm labor. Despite a vast experience with tocolysis in this setting, the average time to delivery after a fetal surgical procedure is approximately 4 weeks. Certainly a more thorough understanding of the pathophysiology, coupled with improved therapy for preterm labor, will allow us to expand the indications for fetal surgery to include select non-life-threatening disorders.

Fetal problems amenable to surgical correction before birth

Presently, the only anatomic malformations that warrant consideration for fetal intervention are those lethal defects that interfere with fetal organ development and that, if alleviated, would allow normal development to proceed. At present, only a small number of life-threatening malformations have been successfully corrected. A few others should be considered for treatment as their prenatal pathophysiology (natural history) is unraveled. As minimally invasive interventional techniques continue to be developed and proved safe, along with improvements in the treatment and/or prevention of preterm labor, prenatal treatment of select nonlethal anomalies may be considered.

Urinary tract obstruction

Fetal urethral obstruction produces pulmonary hypoplasia and renal dysplasia; these often fatal consequences can be ameliorated by decompression of the urinary tract before birth. The natural history of untreated obstruction of the fetal urinary tract is well documented; selection criteria based on fetal urine electrolyte and B_2-microglobulin concentrations (Table 2), and the sonographic appearance of the fetal kidneys, have proved reliable. Out of all the fetuses with dilatation of the urinary tract, as many as 90 per cent do not require intervention. However, those with bilateral hydronephrosis due to urethral obstruction (e.g. posterior urethral valves in males) who develop oligohydramnios do require treatment. If the lungs are mature, the fetus can be delivered early for postnatal decompression. If the lungs are immature, the bladder can be decompressed *in utero* by a catheter shunt placed percutaneously under sonographic guidance. Experience of treating several hundred fetuses in many institutions suggests that selection is good enough to avoid inappropriate intervention, and that restoration of amniotic fluid can prevent the development of fatal pulmonary hypoplasia. It is not clear that decompression can reverse renal functional damage.

Predicted function	[Sodium (mEq/l)]	Chloride (mEq/l)	Osmolarity (mOsm)	Output (ml/h)
Poor	>100	>90	>210	<2
Good	<100	<90	<210	>2

*Fetal urine composition and volume.

Table 2 Prognostic criteria* for the fetus with bilateral obstructive uropathy

Congenital cystic adenomatoid malformation

Although this malformation often presents as a benign pulmonary mass in infancy or childhood, some fetuses with large lesions die *in utero* or at birth from hydrops and pulmonary hypoplasia. The pathophysiology of hydrops and the feasibility of resecting the fetal lung have been studied in animals. Experience of managing over 100 cases suggests that most lesions can be successfully treated after birth, and that some lesions resolve or significantly regress before birth. Although only a few fetuses with very large lesions will develop hydrops, nearly all cases progress rapidly and die *in utero*. Careful sonographic surveillance of large lesions is necessary to detect the first signs of hydrops because fetuses who develop hydrops (fewer than 10 per cent of all those with congenital cystic adenomatoid malformations) can be successfully treated by emergency resection of the cystic lobe *in utero*. Advanced hydrops, heralded by placentomegaly and signs of maternal pre-eclampsia, is a contraindication to fetal intervention; early in our experience, this clinical scenario was responsible for some of the postoperative deaths. Fetal pulmonary lobectomy has proved to be surprisingly simple and quite successful. Twelve fetuses have undergone open surgical resection of the massively enlarged pulmonary lobe. Six of these had rapid resolution of hydrops, impressive *in utero* lung growth bilaterally, and normal postnatal growth and development on follow up of 52 to 84 months. For the rare lesions with single, large cysts, percutaneous thoracoamniotic shunting has also been successful.

Congenital diaphragmatic hernia

This is an anatomically simple defect that is correctable after birth by removing the herniated viscera from the chest and closing the diaphragm. Although less severely affected babies survive with modern postnatal surgical care, including extracorporeal membrane oxygenation, the majority die despite all intervention because their lungs are underdeveloped (hypoplastic) and there is associated pulmonary hypertension. The babies who die *in utero* and soon after birth contribute to a substantial

‘hidden mortality' of approximately 58 per cent, despite the most sophisticated postnatal care. Salvage of these severely affected babies remains an unsolved problem.

We have shown experimentally that repair before birth, when the lungs can grow while the fetus remains on placental support, is physiologically sound and technically feasible. Repair *in utero* has proved to be a formidable challenge. Many technical problems associated with this difficult repair led to the development of the ‘congenital diaphragmatic hernia two-step', a carefully orchestrated approach that allows the reduction of viscera, reconstruction of the diaphragm, and enlargement of the abdomen to accept the returned viscera. This procedure can only be performed in those fetuses in which a large segment of liver is not incarcerated in the hemithorax. When the left lobe of the liver is incarcerated in the chest, adequate repair cannot be made, since reduction of the liver compromises the umbilical blood flow. The efficacy, safety, and cost-effectiveness of the complete *in utero* repair has been evaluated prospectively in a National Institutes of Health-sponsored trial. The results indicate that children in whom the liver is not incarcerated in the left hemithorax (‘liver down') can be successfully treated by the *in utero* repair. However, *in utero* repair in this group has not led to better survival than in matched controls who received standard postnatal therapy.

Further refinements of our prenatal prognostic criteria have established the highest-risk group: those diagnosed before 24 weeks' gestation, those with a major portion of the liver in the hemithorax (‘liver up'), and those with a low lung:head ratio. The lung:head ratio is determined by measuring the right lung volume corrected for gestational age by dividing by the head circumference; the greater the mediastinal shift, the greater the ipsilateral and contralateral lung compression, and the lower the lung:head ratio. A lung:head ratio of less than 1.0 is highly predictive of a high postnatal mortality. Fetuses with herniated liver have never been successfully repaired using the two-step technique *in utero*, despite extensive efforts using a variety of techniques. Indeed, it took many years to recognize the significance of liver herniation and to be able reliably to predict it sonographically. For fetuses deemed unfixable by virtue of liver herniation, we have developed experimentally and now tested clinically a new approach to improving fetal lung development. We have shown in lamb fetuses that impeding the normal egress of fetal lung fluid by controlled tracheal obstruction enlarges the hypoplastic lungs and pushes the viscera back into the abdomen. Initial experience with this new procedure, which we call ‘PLUG' (*p*lug the *l*ung *u*ntil it *g*rows) suggests that temporary occlusion of the fetal trachea accelerates lung growth and ameliorates the often fatal pulmonary hypoplasia associated with severe congenital diaphragmatic hernia. We initially occluded the trachea by using open fetal surgical techniques. Presently, we have developed a fetoscopic approach that has proved safe, efficacious, and has resulted in less preterm labor and maternal morbidity. The trachea is occluded externally with a titanium clip (Fig. 2). The clip is removed using the *ex utero* intrapartum treatment (**EXIT**) strategy in which a hysterotomy is made, the fetal head and shoulders are delivered, and the umbilical cord is clamped only after the clip has been removed and the airway secured (Fig. 3).

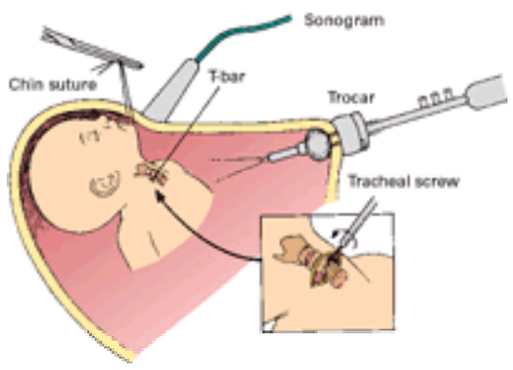


Fig. 2. Under sonographic guidance (1) the fetus's neck is exposed and the head stabilized by placing a transuterine chin suture, and (2) a T-bar is placed in the fetal trachea to aid in localizing the midline fetal neck. After anterior tracheal dissection, a tracheal ‘screw' (inset) is placed in the tracheal wall and anterior traction applied, allowing safe posterolateral tracheal dissection. Subsequently, the trachea is occluded with a titanium clip.

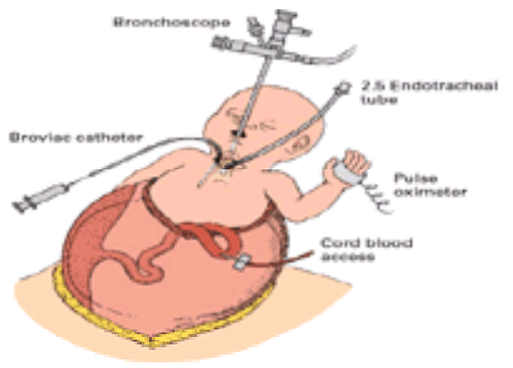


Fig. 3. Essential elements of the *ex utero* intrapartum treatment (EXIT) procedure.

Sacrococcygeal teratoma

Most neonates with sacrococcygeal teratoma survive and malignant invasion is unusual. However, the outcome for sacrococcygeal teratoma diagnosed prenatally is less favorable. A subset of fetuses (fewer than 20 per cent) with large tumors develop hydrops from high-output failure secondary to high blood flow through the extremely vascular tumor. Because hydrops progresses very rapidly to fetal demise, frequent sonographic follow up is mandatory for those tumors diagnosed prenatally. Excision of the tumor reverses this pathophysiology, and has been successful in two patients. Recently, percutaneous radiofrequency coagulation of the tumor's blood supply has been successfully performed.

Fetal airway obstruction

The fetal lungs secrete fluid under positive pressure (with respect to the amniotic fluid pressure) that escapes the upper airway between fetal breathing movements. This normal egress of fluid is prevented in fetuses with congenital high airway-obstruction syndrome (CHAOS) intrinsically due to laryngeal/tracheal atresia or extrinsically caused by a large cervical teratoma or lymphangioma. This experiment of nature formed the basis for our present treatment of fetuses with severe congenital diaphragmatic hernia. Sonographically, fetuses with airway obstruction are characterized by large, fluid-filled, echogenic lungs that compress the mediastinum, evert the hemidiaphragms, and may cause hydrops due to compression of the heart and great vessels. Fetal tracheostomy or early delivery using the EXIT strategy may prevent development of fetal hydrops and death.

Myelomeninogocele

Although it has been assumed that the spinal cord is intrinsically defective, recent studies suggest that neurologic impairment after birth may be due to exposure of the spinal cord *in utero*. We have shown in fetal lambs that exposure of the spinal cord causes neurologic damage that can be prevented or ameliorated by repairing the anatomic defect *in utero*. Although we are developing techniques for repair experimentally, clinical application for this nonfatal defect is not yet justified. The natural history of myelomeningocele in human fetuses must first be elucidated, particularly the time in gestation when the neurologic damage to the lower extremities occurs.

Summary

The great promise of fetal therapy is that for some diseases the earliest possible intervention (i.e., before birth) will produce the best possible outcome, that is, the best quality of life for the resources expended. However, the promise of cost-effective, preventive fetal therapy can be subverted by misguided clinical applications, for example a complex *in utero* procedure that ‘half saves' an otherwise doomed fetus for a life of intensive (and expensive) care. Enthusiasm for fetal intervention must be tempered by reverence for the interests of the mother and her family, by careful study of the disease in experimental fetal animals and untreated human fetuses, and by a willingness to abandon therapy that does not prove effective and cost-effective in properly controlled trials.

Further reading

Albanese CT *et al.* Endoscopic fetal tracheal occlusion procedure: evolution of techniques. *Pediatric Endosurgery and Innovative Technique* 1998; **2**: 47–53. [The first paper to summarize the technical details required to perform operative fetoscopy.]

Albanese CT, Harrison MR. Surgical treatment for fetal disease: state of the art. *Annals of the New York Academy of Science* 1998; **847**:74–85. [General overview of fetal surgical therapy.]

Farrell JA, Albanese CT, Kilpatrick S, Jennings RW, Harrison MR. Maternal fertility is unaffected after fetal surgery. *Fetal Diagnosis and Therapy* 1999; **14**: 190–2. [A retrospective outcomes study for those mothers whose fetuses underwent *in utero* surgery.]

Harrison M, Danty M. *Atlas of fetal surgery*. Chapman & 38; Hall, Florence, KY. [A comprehensive compilation of drawing, photographic images, and scientific outlines spanning an 18-year history of the development of the fetus as a patient.]

Harrison MR, Golbus MS, Filly RA, ed. *The unborn patient: prenatal diagnosis and treatment*, 2nd edn. Saunders, Philadelphia, 1990. [The most comprehensive text detailing fetal disease and its prenatal management.]

42.12 Paediatric otolaryngology

Ellen M. Friedman and Ronald W. Deskin

Disorders of the ear
Infection
Acute otitis media
Otitis media with effusion
Chronic otitis media
External otitis
Hearing problems
Ear trauma
Disorders of the nose and paranasal sinuses
Sinusitis
Nasal obstruction
Choanal atresia
Congenital nasal masses
Nasal foreign bodies
Epistaxis
Nasal trauma
Disorders of the oral cavity and oropharynx infection
Congenital malformations
Salivary gland disorders
Caustic ingestion
Esophageal foreign bodies
Disorders of the larynx and trachea
Hoarseness
Stridor
Inflammatory disorders of the airways
Congenital malformation
Foreign bodies
Tracheostomy
Disorders of the neck
Inflammatory neck masses
Congenital masses
Sternocleidomastoid muscle tumor of infancy
Neoplasms of the neck
Further reading

Pediatric otolaryngology has become a mature surgical subspecialty in the last decade. The management of otolaryngologic disorders in infants and children requires specific expertise in both otolaryngology and in the care of children. In addition, as the survival rates of small premature infants has improved, a team approach to their care has developed involving neonatologist, pediatrician, otolaryngologist, and other pediatric specialists.

Disorders of the ear

Infection

An initial discussion of the nomenclature of ear infections is important. Otitis media refers to inflammation of the middle ear, without reference to pathogenesis. Acute otitis media is synonymous with suppurative or purulent otitis media, while otitis media with effusion is inflammation of the middle ear with fluid and no tympanic membrane perforation. If the fluid persists for less than 3 weeks it is acute; if it lasts for 3 weeks to 3 months it is referred to as subacute, while chronic otitis media with effusion persists for longer than 3 months. Serous media or secretory otitis media are synonyms for otitis media with effusion. The term chronic otitis media indicates a perforated eardrum and should be qualified as being with or without cholesteatoma.

Otitis media is the most common illness causing a child to visit a physician. One-third of children will have had at least one episode of otitis media by the age of 1 year, two-thirds will have had at least one episode by the age of 3 years, and one-third will have had over three episodes by this age. Fluid in the middle ear frequently persists after the completion of 10 days of antibiotic treatment; however, 50 per cent of treated patients still have fluid after 4 weeks, 20 per cent after 8 weeks, and 10 per cent after 12 weeks.

The pathogenesis of otitis media is primarily eustachian tube dysfunction. Children under the age of 6 years have an increased incidence of otitis media because the eustachian tube is more horizontal and shorter than in the adult. Nasopharyngeal contaminants may lead to a complex inflammatory response in the middle ear and a decrease in mucociliary transport. Functional obstruction (adenoid tissue, allergic edema) may cause a high negative pressure in the middle ear which may result in a sterile effusion. This effusion frequently becomes secondarily infected.

Factors which appear to increase the incidence of otitis media include immunodeficiency, use of day-care centres, and exposure to cigarette smoke in the home. Patients with craniofacial anomalies such as cleft palate, and those with Down syndrome have a significantly increased incidence of otitis media.

Resistant strains of *Streptococcus pneumoniae* have become an increasing problem in all areas of the world. Day-care centers and other institutional settings along with the indiscriminate use of antibiotics appear to be implicated in these problems.

Acute otitis media

Ear pain, irritability, increased ear pressure, and decreased hearing acuity are symptoms of acute otitis media. Fever may or may not be present in infants. The diagnosis is made on pneumatic otoscopy, which reveals a full or bulging tympanic membrane, usually red but sometimes white or yellow, which moves poorly.

The microbiology of acute otitis media has been well studied. The causative organisms are: *S. pneumoniae* (previously *Diplococcus pneumoniae*), 29 per cent; *Haemophilus influenzae* (non-typable), 23 per cent; *Branhamella catarrhalis* (previously *Neisseria catarrhalis*) 8 to 15 per cent; group A b-streptococcus, 4 per cent; and *Staphylococcus aureus*, 2 per cent. Negative cultures may represent viral infections or infections with anaerobic bacteria. Up to 30 per cent of the *H. influenzae* and almost all of the *B. catarrhalis* isolates are b-lactamase producers and resistant to ampicillin drugs. The initial treatment is most commonly amoxicillin (40 mg/kg per day in three doses for 10 days). Myringotomy and aspiration is indicated if there is no improvement in the acute symptoms and findings over 48 h. If resistant organisms are suspected, amoxicillin with clavulanic acid, cefaclor, or erythromycin/sulphonamide should be administered). In addition, third-generation cephalosporins (cefpodoxime, cefprozil), macrolide antibiotics (clarithromycin, azithromycin) may all be useful for these more resistant bacteria. Ceftriaxone administered intramuscularly as a single dose may be an alternative when there is a concern that the child will not be given the medication or will not take it.

Children with recurrent acute otitis media, defined as three episodes in 6 months or five episodes in 12 months may benefit from prophylactic antibiotics for 1 to 3 months (amoxicillin 20 mg/kg per day in a single dose or sulfroxizole 25 mg/kg twice daily). The role of the prophylactic regimen is being re-examined in view of the high rate of the development of resistant organisms.

Otitis media with effusion

Persistent middle ear effusion may cause more subtle symptoms and signs, such as ear fullness, hearing loss, loss of appetite, ear tugging, and malaise. Despite the absence of acute symptoms or physical findings, 25 per cent of these ears have bacteria present. The causative organisms are the same as those responsible for acute otitis media. For this reason, another trial of a different antibiotic may be useful.

The indications for insertion of a ventilation tube (grommet) are listed in Table 1. Such tubes are usually inserted under general anesthesia and are left in place for 6 to

12 months. Potential complications are persistent tympanic membrane perforation, cholesteatoma formation, tympanosclerosis, tube otorrhea, and recurrence of otitis media when the tube has extruded.

Otitis media with effusion for 3 months despite antibiotic therapy
Recurrent acute otitis media: three episodes in 6 months or five episodes in 12 months with failure to remain free of symptoms
Persistent negative pressure with pain, vertigo, conductive hearing loss 30 dB, and retraction pocket in the tympanic membrane
Suppurative complications of otitis media, i.e. mastoiditis, acute labyrinthitis, subperiosteal abscess
Cleft palate child, other craniofacial anomalies associated with increased incidence of otitis media

Table 1 Indications for ventilation tube insertion

Adenoidectomy may reduce the incidence of recurrences of acute otitis media and otitis media with effusion, particularly in patients in whom ventilation tube treatment has not been successful and in whom symptoms of upper airway obstruction such as snoring or mouth breathing are apparent. The size of the adenoid may not be as important as lymphatic stasis and pooling of infection in and around the eustachian tube resulting from the presence of the adenoid tissue.

Chronic otitis media

Frequent otorrhea in the presence of a tympanic membrane perforation usually indicates either mucosal infection or ingrowth of a benign skin-lined white sac-like cyst (cholesteatoma). Most acquired cholesteatomas develop from a perforation or a deep retraction pocket containing squamous epithelial debris, especially in the superior part of the tympanic membrane (confined to the attic area of the middle ear). Surgical removal of the skin lining, reconstruction of the ossicular chain, and reconstruction of the tympanic membrane with a fascia graft is required, and mastoidectomy is often necessary to rid the middle ear/mastoid of the disease.

Congenital cholesteatoma (epithelial rest behind intact tympanic membrane) is less common than acquired cholesteatoma. These most commonly arise in the anterior superior quadrant of the middle ear space and require surgical removal.

External otitis

External otitis is active inflammation of the ear canal and sometimes the pinna; moisture in the ear canal is the usual cause. Swimming, showers, washing the hair, and even humid climates may predispose the ear canal to infection. Local trauma with a fingernail or cotton swabs allow bacterial entry through the skin of the ear canal. Localized conditions such as eczema and psoriasis may also alter the canal skin and predispose to infection. *S. aureus* is the most common cause of acute external otitis while *Pseudomonas* and fungi (*Aspergillus* and *Candida*) may cause a chronic infection.

Symptoms range from mild itching to fullness and throbbing pain. Movement of the ear canal on opening the jaw or moving the pinna exacerbates the pain and allows otitis externa to be distinguished from otitis media. The external auditory canal is frequently oedematous and may be filled with foul-smelling debris.

Treatment begins with elimination of water from the ear canal and warnings against the introduction of swabs and other instruments. Prophylactic use of alcohol/vinegar drops in a 1:1 mixture after swimming is helpful if the tympanic membrane is intact. Careful, gentle cleaning of the ear canal with suction or a loop under microscopic guidance is essential. Prompt administration of topical antibiotic or steroid drops and avoidance of water are the main principles of treatment. Occasionally, placement of a wick of cotton (or commercially prepared wick) may be required for 2 to 4 days to help resolve ear canal swelling and allow drops to enter the canal. Broad-spectrum antibiotics are rarely used unless there is inflammation of the pinna, cervical adenopathy, or fever.

Hearing problems

Hearing loss in children is usually first noticed by parents or teachers. In infants, a subtle lack of response to sudden noises and unusual behavior indicate hearing loss. Delayed speech and language development may be the initial signs in a toddler or older child. A unilateral hearing loss may go undetected for years as communication skills may initially appear to be only slightly affected.

Hearing loss is classed as mild (30 to 40 dB hearing level), moderate (40 to 50 dB), severe (50 to 70 dB), and profound (>70 dB). Loss of greater than 60 dB is usually sensorineural or mixed (conductive and sensorineural). Hearing loss may be stable, progressive, or fluctuating, depending on the etiology. The specific tone frequencies lost are important; the result determines the effect on speech and language development. Conductive hearing loss, the most common hearing problem in children, is usually due to middle ear fluid or infection. Congenital malformation of the ossicular chain, trauma, or ossicular erosion by chronic infection may also result in a conductive hearing loss. Impaction of cerumen rarely causes a significant conductive hearing loss.

Sensorineural hearing loss is usually congenital and may be genetic or acquired *in utero* following infection of the fetus by agents such as rubella virus, *Toxoplasma*, and cytomegalovirus, exposure to drugs, or metabolic disorders. Genetic causes are usually inherited in an autosomal recessive manner and may also produce malformations in other systems. Infantile or childhood meningitis remains a significant cause of acquired sensorineural hearing loss.

No child is too young to undergo a hearing test. Auditory brain-stem response (ABR, BSER, BAER) may be carried out in the neonate, uncooperative, or retarded child (Fig. 1, Fig. 2). Otoacoustic emissions consist of faint acoustic energy originating in the inner ear which are measurable by an acoustic probe placed in the external auditory canal. This holds promise as a rapid sensitive technique for infant hearing screening. It is widely used in research, and its clinical usefulness is currently being evaluated. Cochlear implantation in children is being increasingly used for bilateral profound sensorineural hearing loss when there is minimal help with conventional amplification. This is a very labor-intensive process involving extensive cooperation between the otologic surgeon, deaf educators, speech pathologists, and psychologists. Factors carrying a high risk for hearing loss in infants are listed in Table 2.



Fig. 1. Two-month-old patient undergoing an auditory brainstem response evoked potential test.



Fig. 2. Audiologist at the computer, performing an auditory brainstem response evoked potential test.

Family history of hearing loss in childhood
Congenital infections, toxoplasmosis, rubella, cytomegalovirus, herpes, syphilis
Craniofacial anomalies
Birth weight <2500 g
Bacterial meningitis
Hyperbilirubinemia
Birth anoxia, hypoxia, or low Apgar scores
Maternal infections during pregnancy

Table 2 High risk factors for hearing loss in the newborn infant

Ear trauma

External ear

The pinna is exposed to trauma due to its location on the lateral surface of the skull. Blunt trauma may cause ecchymosis of the soft tissue or a hematoma, causing interruption of the blood supply. Cartilage damage may result and a significant cosmetic deformity may occur (cauliflower ear). Participants in contact sports should wear protective headgear to prevent auricular damage. Hematomas or seromas should be drained before cartilage damage occurs. Frostbite of the pinna may follow exposure to extreme temperatures. The ear should be warmed rapidly with moist cotton soaked in saline at 37 to 40°C. Excessive manipulation should be avoided.

Foreign bodies in the ear canal are commonly encountered in children. Swelling of vegetable matter causes infection, odor, drainage, and pain. Plastic or metal objects and rocks may be found on routine examination in patients with minimal symptoms. The shape and size of the foreign body directs how it should be removed; irregular edges may be grasped with alligator forceps, while an ear loop or wax curette or right angle hook inserted behind a small, round object may roll it out. A foreign body located near the tympanic membrane may be dislodged by gentle irrigation with water at body temperature or moved to a location in the ear canal where it can be reached with a loop or forceps. Insects may be killed with mineral oil or alcohol (provided the eardrum is intact) and then removed by forceps or irrigation. Antibiotic or steroid ear drops applied after foreign body removal hasten healing of the canal if abrasion or irritation has occurred.

Tympanic membrane and middle ear

Traumatic perforation of the tympanic membrane may be caused by a penetrating foreign body (including cotton applicators), a slap to the side of the head, or activity such as diving or water skiing. Most small or moderate perforations will heal spontaneously in 6 to 8 weeks, if infection is prevented, water contamination is avoided, and the tympanic membrane edges are not folded into the middle ear. Folded edges should be teased up under microscopic control to promote healing and to avoid inward migration of squamous epithelium (cholesteatoma). Barotrauma during flying or scuba diving rarely results in traumatic perforation: hemotympanum (blood in the middle ear) is more common. the latter may become infected, in which case antibiotics are required. Myringotomy is recommended if the blood or fluid persists for longer than 3 weeks.

All of the above injuries may also cause traumatic ossicular chain disruption and/or perilymph fistula.

Inner ear

Blunt trauma to the head may fracture the temporal bone. Transverse fractures usually result from occipital or frontal blows and may cause hemotympanum, sensorineural hearing loss, vertigo, and facial nerve paralysis. Longitudinal fractures may occur from more lateral blows and may disrupt the tympanic membrane, ossicles, and the external ear canal. These are less likely to cause sensorineural hearing loss and facial nerve paralysis. Computed tomography (CT) of the temporal bones may help to delineate the presence and location of the fracture.

Clear rhinorrhea or otorrhea after temporal bone fracture suggests a cerebrospinal fluid leak. Repair using a neurosurgical otologic or combined approach may be necessary. Delayed facial nerve paralysis due to swelling may be treated in the same way as Bell's palsy, but immediate paralysis requires surgical exploration, decompression, or repair.

Sudden hearing loss and/or vertigo following straining, lifting, or strenuous exercise may be due to acute perilymph fistula formation. Exploration and fat or fascia patching of the leak at the oval and round windows is required if spontaneous healing does not occur.

Disorders of the nose and paranasal sinuses

Sinusitis

Acute bacterial sinusitis is usually secondary to viral upper respiratory infection, anatomical deformity of the nasal passages, or allergic rhinitis. The causative organisms are similar to those of otitis media (*S. pneumoniae, H. influenzae*, and *B. catarrhalis*); therefore, choice of antibiotic is also similar. Most acute episodes are treated for 21 days. Physical examination discloses purulent rhinorrhea in the middle meatus. Nasal congestion and cough are commonly associated with sinusitis in children, while halitosis, fever, facial pressure, and pain, which are common in adults, are less common in pediatric patients. If sinusitis is persistent, or unresponsive to treatment, plain radiographs may show air/fluid levels with acute mucosal thickening or total opacification (Fig. 3). If there is failure of antibiotic therapy or recurrence a culture may be obtained by antral puncture, which is more informative than a swab of the nose or nasopharynx. Puncture is performed through the inferior meatus and usually requires general anesthesia in children. Careful attention to the high positioning of the floor of the maxillary sinus in children as well as the location of tooth buds is important. Frontal sinus trephination through a small medial eyebrow incision may occasionally be required.



Fig. 3. Water's view radiograph demonstrating left maxillary sinusitis with mucosal thickening.

Complications of sinusitis include orbital and intracranial spread of infection. Periorbital (preseptal) cellulitis causes edema of the upper eyelid and is treated with intravenous antibiotics active against *Staphylococcus aureus* and *H. influenzae*. Orbital cellulitis causes chemosis, proptosis, and limitation of eye movement; aggressive intravenous antibiotics and topical nasal decongestant drops usually resolve these symptoms. Reduced visual acuity requires immediate drainage of sinuses and orbit. Orbital abscess is treated in a similar fashion, but is more likely to result in permanent visual loss. Cavernous sinus thrombus is extremely rare and manifests itself with bilateral signs of orbital abscess. This grave condition is treated with intravenous antibiotics, anticoagulants, and drainage of the infected sinus. Mortality rate is 15 per cent.

Intracranial complications of sinusitis include meningitis, and epidural, subdural, and brain abscesses; these are becoming less common as ability to diagnose disease rapidly and to administer aggressive therapy improves. Surgical treatment for sinusitis in children may involve indirect procedures such as adenoidectomy or more direct procedures such as limited irrigation procedures and perhaps endoscopic sinus surgery for persistent clinical disease despite adequate medical management, pulmonary complications of active sinus disease, and complications of sinusitis such as intracranial spread and orbital complications. External procedures for frontal sinus and ethmoid sinus complications may also be indicated when endoscopic approach is deemed unwarranted (constricted nasal space in small children and operator inexperience).

Nasal obstruction

The nose is a passageway for air, which it also warms, filters, and humidifies. Nasal obstruction in children may result from anatomic deformity, an obstructive mass, or an alteration in the physiologic function of the nose. Epistaxis, anosmia, hyponasal speech, and rhinorrhea may accompany obstructive symptoms. Intranasal examination with a head mirror or headlight and nasal speculum is essential, and a flexible fiberoptic endoscope may also provide valuable information. Intranasal examinations are usually enhanced by the use of topical intranasal vasoconstrictive agents, such as 0.25 per cent Neo-Synephrine. Paranasal sinus radiographs, CT scan, or magnetic resonance imaging scan may be required.

Choanal atresia

Congenital nasal obstruction may be secondary to choanal atresia due to the failure of the nasobuccal membrane to rupture in the fourth embryonic week. This condition occurs in 1/8000 births, is usually unilateral, and is more common in females. Most cases (90 per cent) are due to bony rather than membranous atresia.

As the infant is an obligate nasal breather for the first weeks to months of life, bilateral choanal atresia may cause severe early respiratory distress. The diagnosis is confirmed by failure to pass nasal catheters (no. 6 and no. 8 French); CT examination will further define the nature of obstruction (bony versus membranous; thin versus thick). Keeping the infant's mouth open with a McGovern nipple or an oral airway may allow mouth breathing to be learned (Fig. 4). Feeding by orogastric tube may be necessary. If the infant learns to mouth breathe and gains weight, transpalatal repair is usually undertaken at 6 to 12 months of age, although earlier intervention is frequently necessary. Transnasal rupture of a membranous bilateral atresia may be performed early in life if required. Although the use of stents is controversial, stenting for 4 to 6 weeks with endotracheal tube modifications is generally used in combination with either intranasal or transpalatal repair. Over 50 per cent of affected infants have other congenital deformities, and this condition may be part of the CHARGE association (Coloboma, Heart defects, Atresia of the choana, Retarded growth and development, Genitourinary abnormalities, and External ear deformities).

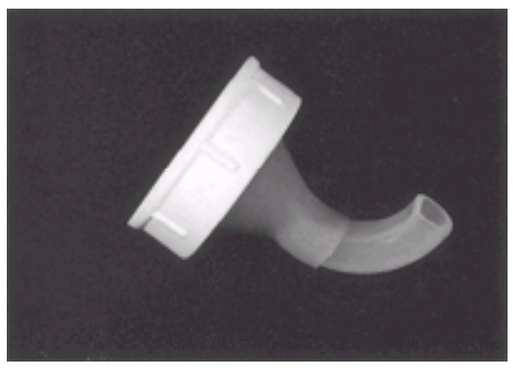


Fig. 4. McGovern nipple used to maintain patient oral airway in neonates with choanal atresia.

Anterior or posterior nasal stenosis may occur and cause similar but milder symptoms of atresia. Initial conservative management, with brief use of nasal topical decongestants, frequent saline drops, and suctioning may be sufficient until growth occurs. Surgical repair is rarely necessary.

Congenital nasal masses

Dermoid cysts are congenital epithelium-lined rests at lines of embryonic fusions. The nasal midline, between the nose and orbit, and lateral supraorbital ridge are usual sites of these cysts. They may present with external pits at the nasal dorsum. Intracranial connections are possible, but uncommon.

Encephaloceles and gliomas are epithelial and/or glial extensions of the central nervous system into the nasal and facial area. Encephaloceles, which may connect with the central nervous system and contain cerebrospinal fluid, are usually pulsatile and enlarge with crying. Gliomas, unlike encephaloceles, have no central nervous system connection. CT and magnetic resonance imaging scans are helpful in evaluating the possibility of intracranial extension in these cases. Excision is frequently carried out for safety and cosmetic reasons, using a combined neurosurgical and otolaryngologic approach.

Nasal foreign bodies

Intranasal foreign bodies may cause obstructive symptoms and, if undetected, will cause unilateral malodorous, purulent rhinorrhea. The differential diagnoses of unilateral foul-smelling nasal drainage include foreign body, unilateral choanal atresia, and unilateral intranasal tumor. Suction and a topical nasal decongestant helps to establish the diagnosis. When a foreign body is seen, removal with suction, blunt hook, or forceps may be possible. However, removal under general anesthesia is often required in order to avoid dislodging the foreign body posteriorly into the airway.

Epistaxis

Epistaxis is relatively common in childhood, and most episodes are secondary to inflammation or trauma. A history of bleeding, easy bruising, trauma, or possible associated illness (such as leukemia) must be obtained. The anterior septum (Kiesselbach's triangle or Little's area) is the most common site because of the presence of a plexus of vessels. Careful examination and use of a topical nasal decongestant is important. Single finger pressure to squeeze the nostrils together usually controls

active bleeding. Cauterization with silver nitrate, packing with absorbable collagen or gelatin, or even anterior packing with antibiotic-impregnated gauze may be necessary. Posterior packing with lamb's wool or a Foley catheter with the balloon inflated is very rarely needed in children. If this is the case, admission is required while the pack is in place. Subsequent rebleeding can be prevented by the use of nasal saline drops, humidification, and topical vaseline applications. Nasal septal deviation, sinusitis, tumor (angiofibroma), foreign body, and blood dyscrasia are uncommon causes of childhood epistaxis, but deserve consideration if epistaxis is persistent.

When exterior and posterior nasopharyngeal packing fails to control hemorrhage, ligation of the internal maxillary artery may be necessary.

Nasal trauma

Newborns may acquire a septal deviation during birth, indicated by a tilted columella, tip deviation, and alar asymmetry. If the septum is dislocated from the vomer groove immediate closed reduction with an intranasal instrument such as an elevator should be attempted. If dislocation has not occurred, spontaneous correction is likely to occur over 6 to 8 weeks.

Trauma to the nose in a child may cause a hematoma of the septum. This must be diagnosed and drained intranasally with needle aspiration or incision and drainage to prevent loss of cartilage support and resultant saddle nose deformity. Closed (or open) reduction of nasal fractures should wait until swelling has resolved over 5 to 7 days, but should be undertaken before fibrous healing makes reduction impossible. Closed reduction is optimal within 7 days following trauma. Unreduced nasal fractures may produce significant functional and cosmetic abnormalities which may require septorhinoplasty during later years.

Disorders of the oral cavity and oropharynx infection

Most throat infections are viral and therefore require no antibiotics. Throat cultures are necessary for the diagnosis of bacterial infections. Pharyngitis or tonsillitis due to b-hemolytic streptococci is treated with penicillin for 10 days; erythromycin or cefaclor may be used for those allergic to penicillin.

Peritonsillar abscess or quinsy is localization of pus in the space between the tonsil capsule and the superior constrictor muscles of the pharynx. The tonsil and soft palate bulge medially and the uvula deviates away from the abscess. Intravenous administration of penicillin, with excision and drainage, is required. Many otolaryngologists recommend tonsillectomy 6 weeks later.

Retropharyngeal abscess may develop from suppurative adenitis of retropharyngeal nodes, and is best seen on a lateral neck film taken in inspiration and neck extension. Parapharyngeal abscess from the parapharyngeal lymph nodes is best seen as pharyngeal fullness and neck swelling. Intravenous antibiotics with incision and drainage are required for the treatment of both of these conditions.

The indications for adenotonsillectomy relate primarily to the frequency and severity of infection and the extent of airways obstruction. Most otolaryngologists recommend adenotonsillectomy if tonsillitis occurs four to six times per year or if airways obstruction occurs secondary to tonsillar and adenoidal hypertrophy, presenting as mouth-breathing, snoring, restless sleep, or full sleep apnea syndrome. Midface hypoplasia, dental malocclusion, and severe orthodontic problems have been described secondary to prolonged mouth-breathing. Snoring and sleep disturbance can be assessed informally from a cassette tape recording of a patient's breathing during sleep. A formal sleep laboratory study is helpful in quantifying the degree and severity of sleep apnea. Enuresis may occur in the child with sleep disturbances secondary to upper airways obstruction and severe obstruction may lead to cor pulmonale.

Congenital malformations

A common abnormality is a short lingual frenulum (ankyloglossia) or ‘tongue-tie.' There is usually no need for surgical correction unless eating, speech, or dental development is affected. In this case surgical release with simple incision or Z-plasty may be done. Care must be taken to avoid damage to the sublingual salivary ducts.

A bifid uvula may be associated with a submucous cleft palate or absence or weakness of the midline palate aponeurosis, and a notch in the posterior edge of the hard palate. Adenoidectomy is contraindicated in these children to avoid development of significant velopharyngeal incompetency, resulting in nasal regurgitation and hyponasal speech. Infants with cleft palate may require a special large-opening nipple to enable feeding. Cleft lip usually is repaired at the age of 3 to 4 months and palate repair is performed at 12 to 18 months. Most children with cleft palate develop otitis media with effusion at an early age and require ventilation tube insertion as early as 3 to 6 months of age, in conjunction with plastic surgery.

Salivary gland disorders

Infection in the salivary gland (sialoadenitis) may be viral or bacterial. Viral mumps is uncommon: immunity produced by administration of live mumps virus vaccine is widespread. Bacterial infection presents with signs and symptoms similar to cervical adenitis—fever, swelling, pain, and tenderness. Massage of the affected parotid or submandibular gland will produce pus at the duct opening. Intravenous antibiotics, heat, massage of the gland, increased fluid intake, and sialogogues such as lemon drops are helpful. Recurrent or chronic sialoadenitis may be caused by ductal stricture, stones, immunodeficiency, or autoimmune disorders. If continued painful swelling occurs, drainage or gland excision may be necessary.

While salivary gland tumors are uncommon in children, the parotid gland is most commonly affected. The most common parotid neoplasm in children is the benign hemangioma, which may be detected early in life and is frequently associated with cutaneous lesions. Most enlarge over 6 to 12 months and then undergo spontaneous regression over several years. Systemic corticosteroids may help the tumor to regress. Lymphatic malformations may present as a soft, painless, compressible mass in the parotid region. They rarely resolve spontaneously and may cause significant airway obstruction if cervical or supraglottic involvement is prominent.

Excluding hamartomatous malformations, over 50 per cent of salivary gland tumors in children are malignant. CT scan may be helpful and fine needle aspiration, where available, may help as an initial procedure. Local pain, rapid growth, facial paralysis, and systemic changes are suggestive of malignancy. Malignant parotid tumors are treated with total parotidectomy; the facial nerve is sacrificed only if the tumor involves the nerve. Neck dissection is recommended for palpable positive regional nodes. Chemotherapy and radiotherapy may be helpful in high-grade malignancy or when total excision is not possible. Submandibular tumors are more likely to be malignant than are those in the parotid.

Caustic ingestion

Ingestion of caustic substances by young children is usually accidental, but may be the result of intentional child abuse. Teenagers may attempt suicide in this manner. Drooling and dysphagia are the most common presenting symptoms. Physical examination may reveal obvious injury; however, the lack of oral burns does not rule out esophageal injury. Esophagoscopy should be performed within 48 hours of the ingestion but carried out only to the level of the first burn.

The classification of burns is shown in Table 3. Grade I burns are limited to superficial mucosa and will not result in scar formation; grade II burns penetrate the mucosa and may lead to strictures. Steroids appear to help these injuries. Grade III burns, penetrate deeper into the mucosa and should not be treated with steroids for fear of esophageal penetration. A barium swallow is obtained at 3 weeks and gentle periodic dilatation is started if stricture formation has begun. The role of early barium swallow is being re-examined. Endoscopy is critical for the documentation of esophageal injury, but barium swallow may help with the predication of eventual stricture formation.

Agent	pH	Pathology	Most common sites affected
Lye	>7	Liquefaction, necrosis	Oral cavity, upper esophagus
Sulfuric acid	<7	Coagulation, necrosis	Lower esophagus, stomach
Bleach	?	Irritation	Oral cavity, upper esophagus

Table 3 Classification of caustic injuries of the esophagus

Esophageal foreign bodies

A foreign body ingested by a child may lodge in the esophagus, most commonly at the cricopharyngeal inlet, aortic knob, and the esophagogastric junction. Small, round objects usually pass through the gastrointestinal tract. Endoscopic removal under general anesthesia with airway intubation is the safest way to retrieve an esophageal foreign body. Foreign bodies in the esophagus may cause airway symptoms with stridor, croupy cough, and wheezing due to tracheal compression through the adjacent wall.

Disorders of the larynx and trachea

Hoarseness

Hoarseness is best defined as an alteration of the normal voice; this may range from a weak, breathy cry in an infant with vocal cord paralysis to a hoarse, scratchy voice in an adolescent due to voice abuse at an athletic event. Edema and inflammation due to upper respiratory tract infection is the most common cause and will resolve as the infection improves. If hoarseness persists beyond resolution of the acute illness, evaluation is necessary.

Vocal cord nodules begin as polyp-like edema at the junction of the anterior one-third and middle one-third of the true vocal cords and are secondary to vocal cord abuse or poor voice habits. Indirect (mirror) laryngoscopy or flexible fiberoptic scope examination usually confirms the diagnosis. Referral for speech therapy is made in children over the age of 8 when they can begin to comply with instruction. Direct laryngoscopy and removal of nodules with cup forceps or carbon dioxide laser is indicated if a 6- to 12-month trial of speech therapy fails, or if fibrous nodules develop. Vocal abuse must be corrected or surgical removal will help only temporarily.

Chronic cough and throat clearing in a child with allergy and chronic sinusitis may cause vocal cord edema and thickening. The basic cause must be controlled. Other causes of hoarseness include vocal cord paralysis, congenital laryngeal malformation (e.g. web, cysts), laryngeal papilloma, and gastro-esophageal reflux. Trauma from endotracheal intubation may cause development of a granuloma on the vocal cord or arytenoid area.

Stridor

Stridor is an audible sound from the airway indicating turbulent airflow due to partial obstruction. The level of obstruction determines the quality and phase of respiration of the stridor (Table 4). Nasal or nasopharyngeal obstruction causes snoring or snorting sounds on inspiration. Supraglottic (epiglottis, aryepiglottic folds, false vocal cords) obstruction causes inspiratory stridor, while lower (distal trachea, bronchi, and bronchiolus) obstruction causes expiratory stridor or wheezing. The subglottic area and immediate upper trachea display a biphasic (inspiratory and expiratory) stridor when obstruction is present.

Supraglottic	
	Stridor—inspiratory
	Cry/voice—muffled
	Swallow—abnormal
Glottic	
	Stridor—inspiratory, possible biphasic
	Cry/voice—abnormal
	Swallow—normal
Subglottic	
	Stridor—biphasic
	Cry/voice—normal
	Swallow—unaffected
Tracheal	
	Stridor—expiratory
	Cry/voice—unaffected
	Swallow—unaffected

Table 4 Characteristics of airways obstruction by anatomic level

A careful history is mandatory including onset, fluctuation, possible presence of a foreign body, and any factor that alters the stridor, such as position change, feeding, or treatment. Physical examination is vital but should be limited if stridor is acute, progressive, or if severe obstruction has occurred or is imminent. Radiographic studies should only be requested in the non-urgent case; these should include anterior, posterior, and lateral chest and neck films. Fluoroscopy allows the dynamic effects of respiration of the airways to be studied and may be helpful, especially if the child is unable to cooperate for inspiratory and expiratory films. Barium swallow examination can rule out vascular ring, gastroesophageal reflux, tracheoesophageal fistula, and neuromuscular swallowing problems. Flexible fiberoptic examination of the airway may provide information on glottic and supraglottic conditions, but this technique should be used with caution if a foreign body or progressive obstruction is expected. Direct laryngoscopy, bronchoscopy, and esophagoscopy are the definitive procedures, and should be performed under general anesthesia with spontaneous ventilation in the operating room. Careful planning and cooperation with the anesthesiologist is essential so that the airways can be shared safely during the examination.

Inflammatory disorders of the airways

Epiglottitis (supraglottis) is acute inflammatory swelling of the supraglottic larynx, including the epiglottis, aryepiglottic folds, and false vocal cords, usually caused by *H. influenzae* type B. The most common age is 2 to 6 years, and patients present with rapid acute onset of fever, throat pain, drooling, and airway distress with inspiratory stridor. Retractions may occur late in the process as more shallow, rapid breathing causes less supraglottic collapse with inspiration. A careful look at the throat with a tongue depressor may be difficult, and this is recommended only if symptoms are mild and intubation equipment and personnel are available. Lateral neck radiographs will demonstrate an enlarged epiglottis, but should not be obtained in severely ill children (Fig. 5). A team approach has been shown to decrease mortality associated with this diagnosis; otolaryngologist, anesthesiologist, pediatrician, and emergency room, operating room, and pediatric intensive care unit staff should be present. The examination is undertaken in the operating room under general anesthesia and shows a large, red, swollen epiglottis; the airway below the vocal cords is usually normal. The initial oral endotracheal tube is changed to a nasal tube and meticulously secured. Blood cultures and surface culture swabs are usually positive for *H. influenzae*. Concomitant extralaryngeal *H. influenzae* infection, including pneumonia and meningitis, may also occur. After 36 to 48 h in the pediatric intensive care unit an air leak is usually noted around the endotracheal tube and the child is extubated. The initial antibiotic of choice is ampicillin/chloramphenicol or a cephalosporin, pending the results of culture and antibiotic sensitivity of the blood samples. Oral ampicillin or amoxicillin treatment should be continued for 10 to 14 days unless resistance is shown. With the increased availability of the new *H. influenzae* type B vaccine, administered at 2 to 3 months, acute epiglottitis has become less frequent or may appear in younger children.

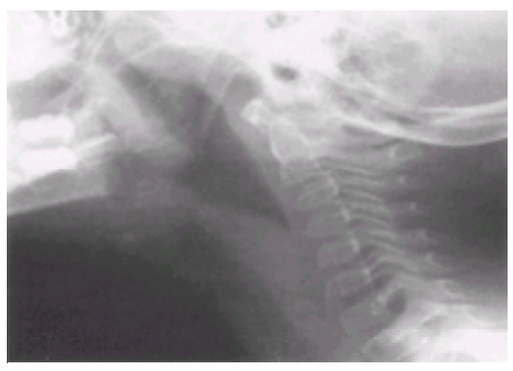


Fig. 5. Six-year-old with respiratory distress and supraglottitis.

Croup (laryngotracheobronchitis) usually occurs in patients under 2 years old, and is usually due to parainfluenza virus infection. The barking cough and biphasic stridor is characteristic. Onset is gradual with general symptoms of upper respiratory tract infection. Significant obstruction rarely develops, although some episodes can be severe. Radiographs show typical anterior, posterior, and lateral narrowing of the subglottic space (Fig. 6). Mild cases are managed with cool mist and increased fluid intake. If racemic epinephrine is needed, hospital admission is recommended to observe the patient for rebound. Corticosteroid administration is recommended if the condition worsens to avoid the need for an artificial airway. If required, intubation with a smaller than calculated endotracheal tube for 3 to 5 days may be acceptable, although some clinicians prefer tracheostomy. Antibiotics are used only if the condition follows a prolonged course or if intubation is required.



Fig. 6. Six-month-old with croup, whose neck film shows the characteristic pencil sign in the subglottic area.

Bacterial tracheitis is fairly common and is heralded by a history characteristic of a viral illness with a sudden rapid deterioration over 8 to 10 h. The patient has thick secretions, high fever, and tracheal crust formation which often is seen on radiographs as density in the tracheal air column with scalloping of the tracheal margins. Fifty per cent of cases are due to *S. aureus*. Bronchoscopy to establish the diagnosis and to suction meticulously the desquamated mucosa with crust cleaning, heavy mist, and antistaphylococcal antibiotics are the mainstay of treatment (Table 5).

Croup	3–6 months	Respiratory syncytial virus, parainfluenza	Radiographs	Mist, racemic epinephrine
Bacterial tracheitis	All ages	*S. aureus*	Variable, pseudomembranes with tracheal obstruction	Bronchoscopy and suction, IV antibiotics
Laryngomalacia	3–6 weeks	Not relevant	Typical laryngoscopy	Supportive care, IV antibiotics

Table 5 Common inflammatory diseases of the pediatric airways

Congenital malformation

Laryngomalacia is the most common cause of stridor in the infant. It is thought to be caused by soft, collapsible supraglottic laryngeal structures, although there are many theories concerning its etiology. At laryngoscopy, redundant arytenoids and a floppy, omega-shaped epiglottis are drawn into the larynx on inspiration, resulting in a high-pitched inspiratory stridor. Laryngomalacia usually begins shortly after birth, and increases when the baby is active, agitated, feeding, or lying supine. When the child is relaxed and prone, stridor is decreased or absent. Laryngoscopy and bronchoscopy should be undertaken in children who fail to thrive, or who suffer from cyanotic spells, apnea, progressive worsening, or who fail to improve by the age of 9 to 12 months. Atypical stridor may indicate a second problem: 10 to 15 per cent of infants with stridor have a second abnormality. Rarely intubation or tracheostomy may be necessary; carbon dioxide laser division of the aryepiglottic folds to reduce the tissue in folding with inspiration may be attempted to avoid the need for tracheostomy.

Vocal cord paralysis is the next most common cause of neonatal stridor. In neonates this may be due to neck stretching during delivery, and usually spontaneously resolves. In older children, chest and skull radiographs and metabolic studies including blood sugars and thyroid function studies, are required. Unilateral vocal cord paralysis causes early hoarseness with breathy, weak cry and possibly aspiration. Arytenoid dislocation from intubation trauma must be ruled out as the arytenoid may be repositioned more easily when an early diagnosis is made. Bilateral vocal cord paralysis causes severe airway obstruction but a near-normal voice. Prompt intervention with intubation or tracheostomy is usually required. Bilateral vocal cord paralysis is usually due to central nervous system problems. In children with the Arnold–Chiari malformation, a ventriculoperitoneal shunt procedure and posterior fossa decompression may correct the vocal cord paralysis and avoid the need for a tracheostomy. Cardiac anomalies may be present, especially in those with left unilateral vocal cord paralysis. Other causes are metabolic (diabetes, thyroid), lead poisoning, trauma from neck or cardiac surgery, central vascular lines, and cervical chest or mediastinal masses.

Congenital subglottic stenosis may cause early stridor or recurrent croup-like episodes in infancy. The stridor is biphasic and may be associated with a barking cough if inflammatory swelling occurs during the course of an upper respiratory tract infection. Failure to pass a 3.5 mm bronchoscope (outer diameter) through the subglottis of a full-term newborn, is suggestive of subglottic stenosis. Mild to moderate stenosis (up to 50 per cent compromise) may respond to conservative management as growth alone may resolve the difficulty. Tracheostomy, dilatation, carbon dioxide laser resection and open laryngotracheal reconstruction may be required in more severe cases.

Acquired subglottic stenosis in neonates, especially the premature, is more likely to be due to prolonged intubation than to congenital causes. Treatment is required for all but the mildest cases. Most cases present as a failed attempt at intubation after the infant has been weaned from the ventilator and the primary need for intubation has resolved. Most reconstruction techniques involve an anterior costal cartilage graft, with or without stenting of the grafted area. Costal cartilage has been used most extensively but auricular, nasal septal, and thyroid ala cartilage have also been used. Stenting may be required if there is high-grade stenosis, a long segment of stenosis, posterior stenosis requiring posterior cricoid split or any condition where there is extensive loss of cartilage structure. For a single-stage laryngotracheoplasty, a previous placed tracheostoma is excised at the time of the repair and an endotracheal tube is left in place for approximately 7 days. Accidental extubation in this situation is a potential risk of the procedure and careful titration of sedation and paralysis may be required.

Subglottic hemangioma may also present as recurrent croup. Fifty per cent of patients with subglottic hemangiomas have other cutaneous hemangiomas. Anterior, posterior, and lateral neck radiographs show asymmetrical subglottic swelling. There may be mild stridor (biphasic) between ‘croup' episodes. This usually becomes obvious at the age of 3 to 6 months and progresses for 1 to 2 years, while the hemangioma increases in size. The typical appearance at laryngoscopy is a smooth posterolateral swelling with a bluish tint. Patients with large hemangiomas usually require tracheotomy to maintain an adequate airway. They frequently show spontaneous involution by the age of 3 to 4 years. Small hemangiomas may cause only mild symptoms and intervention is not required. Staged segmental carbon dioxide laser excision may be used to avoid tracheostomy.

Laryngeal webs and stenosis are the result of failure of the larynx to recanalize during embryonic development. Most webs are anterior at the vocal cord level, and they are an early cause of hoarseness and stridor. Congenital cysts are usually supraglottic in the vallecula and aryepiglottic folds and cause inspiratory stridor. Endoscopic removal of webs and cysts with microinstruments or carbon dioxide laser is usually possible.

Laryngotracheal clefts are uncommon and represent a form of high tracheoesophageal fistula. These clefts cause early stridor, hoarseness, aspiration, and choking.

Early correction is necessary. Except for very small shallow clefts, a midline or lateral neck approach is usually required.

Vascular anomalies such as anomalous innominate artery, right aortic arch, double aortic arch, anomalous right subclavian artery, and pulmonary artery sling may cause respiratory symptoms including stridor, wheezing, coughing, pneumonia, and feeding problems. Chest radiographs and barium swallow are helpful, but often the diagnosis is made at endoscopy. Magnetic resonance imaging or CT scan of the chest may confirm the diagnosis; however, an arteriogram is rarely necessary. Aortoplexy or correction of the vessel abnormalities may be necessary.

Foreign bodies (Fig. 7)



Fig. 7. Chest radiograph of a 4-year-old male who has aspirated a peanut: inspiration.

Foreign bodies that lodge in the larynx may be rapidly fatal unless coughing clears the material from the laryngeal inlet. If there is absolutely no air passage and no phonation, a Heimlich maneuver or suspending the child by its feet and back slaps to dislodge the foreign body may be lifesaving. Emergency cricothyroid space tracheostomy is rarely successful in small children. Foreign bodies in the tracheobronchial tree cause symptoms related to the specific area. Vegetable matter may swell and fragment, causing early symptoms, but obstruction due to plastic and metal objects may be asymptomatic for a prolonged period of time. Unilateral wheezing, air trapping on the affected side, and mediastinal shift away from the affected side on expiration is diagnostic of partial airways obstruction (Fig. 8). Fluoroscopy is usually necessary to show the phases of respiration in a small child, and direct laryngoscopy and bronchoscopy are required. An esophageal foreign body should not be overlooked as a cause of airways obstruction due to tracheal compression through the party wall which is particularly compliant in children.

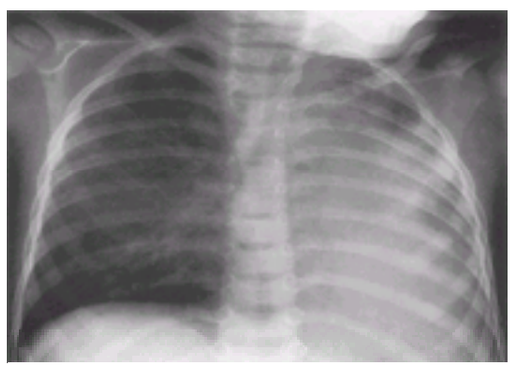


Fig. 8. Same patient as in Fig. 7 on expiration, showing the shift of the mediastinum associated with a foreign body in the left bronchus.

Tracheostomy

Tracheostomy is usually performed in a child for prolonged assisted ventilation secondary to neurologic or cardiac problems, failure to extubate a premature infant with bronchopulmonary dysplasia, pulmonary secretion toilet in a neurologically disabled patient, and, occasionally, for congenital or acquired airway obstruction. Ideally, the procedure should be undertaken over an endotracheal tube or bronchoscope in the operating room. A vertical skin incision assures that the surgeon remains in the midline, and the cosmetic result is quite acceptable after the stoma has matured around the tube and formed a circular scar. Traction sutures of 2-0 silk are used laterally along the vertical tracheal incision through tracheal rings two and three. These traction sutures make reinsertion of the tube easier if accidental extubation occurs in the early postoperative period. The smallest tube that will allow adequate suction and air exchange is used if the child is not on a ventilator; for those on a ventilator, a larger tube with a good seal is necessary. The tracheostomy tube is changed after the first 7 days and then every 7 to 14 days.

Disorders of the neck

Inflammatory neck masses

Cervical masses in children require systematic evaluation and management. A thorough history, complete examination of ear, nose, throat, head, and neck are necessary. Radiographic studies, including sinus films and neck CT are often helpful. Skin tests for tuberculosis are also applied. When antibiotic therapy does not cause adenopathy to resolve or if diagnostic studies are inconclusive, an excisional biopsy is recommended.

Inflamed lymph nodes account for the majority of neck masses in children. The nodes are tender and usually multiple. Cervical node infections may break down and form an abscess, indicated by fluctuance and shiny red changes of the overlying skin, although chronic infection may result in multiple firm nodes that are non-tender and without skin changes. Cat-scratch fever, atypical mycobacteria, and toxoplasmosis may require biopsy for diagnosis. *S. aureus* is the most common cause of an abscess and is treated with intravenous antibiotics. When fluctuance develops, needle aspiration may hasten recovery and allow an organism to be cultured. Larger and more persistent abscesses require incision and drainage in the operating room. CT helps to locate and define the abscess. Parapharyngeal abscesses lying between the pharyngeal constrictor muscles and the vascular compartment, cause a stiff neck and fever, and the child is usually septicemic. Dysphagia is common. Incision and drainage through the neck is required, followed by intravenous administration of antibiotics active against staphylococci. Retropharyngeal abscess may develop in a similar fashion, between the retropharyngeal fascia and prevertebral fascia. This space connects the base of the skull to the mediastinum and infection may spread up or down if not controlled. Lateral neck radiographs taken in full extension and inspiration show a bulge of the space, sometimes with air formation. Intravenous antibiotics and intraoral (and sometimes external neck) incision and drainage are required.

Congenital masses

Congenital masses are most common after inflammatory masses of the neck. Persistence of the embryonic branchial cleft causes branchial cleft anomalies of the head and neck. If the lateral cleft is obliterated, a cyst may develop; this may connect to the pharynx if the medial cleft persists and to the neck if the lateral cleft persists. A sinus connects from cyst to pharynx or cyst to skin surface, while a fistula connects pharynx and skin surface. Lymphatic tissue and cartilage rests may lie in the cyst or tract. A first cleft cyst and tract may present as a cyst in a position posterior and inferior to the ear canal and connect to the skin of the ear canal and into the parotid gland and facial nerve area. The second cleft anomaly is the most common, presenting as a cystic mass and/or a sinus opening in the neck anterior to the sternocleidomastoid muscle. Collections of epithelial debris and infection may take years to develop; this accounts for the variable age at which the congenital anomaly presents. An infection may be the first indication of a problem. If a tract persists, it passes from the cyst between the internal and external carotid arteries, superior to the hypoglossal nerve (XII) and inferior to the glossopharyngeal nerve (IX), to terminate in the tonsillar fossa. If a tract extends up and down from the cyst, multiple step-ladder neck incisions may be required for its removal. Third and fourth cleft cysts are more unusual.

Thyroglossal duct cyst is an epithelial-lined cyst persisting along the course of descent of the thyroid gland from the foramen cecum at the base of tongue to the lower midline of the neck. The most common site is at or below the hyoid bone, where it usually moves up and down with swallowing and tongue extension and retraction.

Infection and collection of mucoid material cause the cyst to present as an enlarging or inflamed mass. There is no outward extension through the skin unless previous aspiration, incision and drainage, surgery, or spontaneous rupture has occurred. Differential diagnoses include cervical node, dermoid, and neoplasm. A thyroid scan may be performed to ensure that the mass is not the sole functioning thyroid tissue. Surgical excision involves removal of the central segment of the hyoid bone with the specimen (Sistrunk procedure). Extension to the base of the tongue often requires a core of tongue tissue to be removed in order to achieve tract excision.

Sternocleidomastoid muscle tumor of infancy

This tumor is an organizing hematoma which arises as a result of perinatal trauma. It presents as a mass in the sternocleidomastoid muscle and torticollis. Fibrosis develops in the hematoma, typically causing symptoms at the age of 3 to 6 weeks. Conservative treatment with heat and passive motion usually resolves the problem. Surgery to resect the fibrosis and lengthen the muscle is rarely necessary.

Neoplasms of the neck

Hemangioma of the head and neck (including the parotid gland) is the most common head and neck neoplasm in children. Rapid growth for 12 to 18 months is usually followed by slow resolution over 1 to 2 years. Observation is the preferred management strategy, unless airway obstruction, hemorrhage, high-output cardiac failure, or platelet trapping occurs. Lymphatic malformations (cystic hygroma) consist of multiple cystic spaces filled with lymph. A large transilluminating neck mass is the most common presentation. Progressive enlargement usually occurs; a rapid increase in size may be due to hemorrhage or infection. Airway compromise may occur. Spontaneous resolution is rare and surgical excision is recommended for all but the smallest of lesions. Surgical procedures may be staged for large lesions, the first procedure being performed at 9 to 12 months of age. Respiratory or feeding problems in a neonate may require insertion of a tracheostomy or gastrostomy tube.

Fortunately, malignant tumors are not common in childhood. The most common malignant neoplasm in children is lymphoma, Hodgkin's and non-Hodgkin's lymphoma being roughly equally common. Multiple firm, rubbery nodes are present, and excisional biopsy is required for diagnosis. Radiation therapy is used to treat localized disease, with the addition of chemotherapy for systemic or diffuse disease. Well-differentiated Hodgkin's lymphoma has a better prognosis than non-Hodgkin's lymphoma.

Rhabdomyosarcoma is the most common soft tissue head and neck malignancy in childhood, the diagnosis being suggested by a rapidly growing painless neck mass. The diagnosis is confirmed by biopsy, and treatment consists of excision of as much of the tumor mass as is reasonably possible, followed by radiation therapy and chemotherapy.

Squamous cell carcinoma of the nasopharynx, melanoma, neuroblastoma, and thyroid malignancy may initially present as cervical adenopathy.

Further reading

Ear

Bevan RC. *Hearing impaired children*. Charles C Thomas, Springfield, IL, 1988.

Bluestone CD, Stool SE. *Pediatric otolaryngology*. WB Saunders, Philadelphia, PA, 1983.

Breiman RF, Butler JC, Tenover FC *et al.* Emergence of drug resistant pneumococcal infections in the United States. *Journal of the American Medical Association* 1994; **271**: 1831–5.

Cochlear Implants in Adults and Children. NIH Consensus Statement, 1995, Bethesda, MD, May 15–17, **13**(2): 130.

Gates GA, Avery CA, Prihoda TJ. The effect of adenotonsillectomy upon children with chronic otitis media with effusion. *Laryngoscope* 1988; **98**: 58–63.

Hough JVD. Otologic trauma. In Paparella M, Shumrich D, eds. *Otolaryngology*, 2nd edn. WB Saunders, Philadelphia, 1980: 1656–79.

Joint Committee on Infant Hearing, American Academy of Pediatrics Statement 1982. *Pediatrics* 1982; **70**: 496–7.

Rosenfeld RM, Vetrees JE, Carr J, *et al.* Clinical efficacy of antimicrobial drugs for acute otitis media: Meta-analysis of 5400 children from thirty three randomized trials. *Journal of Pediatrics* 1994; **124**: 355–67.

Simmons FB. Patterns of deafness in newborns. *Laryngoscope* 1980; **90**: 448–53.

Teele DW, Pelton SU, Klein JO. Bacteriology of acute otitis media in response to initial antimicrobial therapy. *Journal of Pediatrics* 1981; **98**: 537–9.

Teele DW *et al.* Otitis media with effusion during the first three years of life and development of speech and language. *Pediatrics* 1984; **74**: 282–7.

Nose and paranasal sinuses

Benjamin B. Evaluation of choanal atresia. *Annals of Otology, Rhinology, and Laryngology* 1985; **94**: 429.

Bluestone CD, Stool SE, Arjona SK. *Pediatric otolaryngology*, pp. 719–28. WB Saunders Co., Philadelphia, 1983.

Lembo RC. Sleep apnea and related disorders in infants and children. *Ear, Nose, and Throat Journal* 1985; **64**: 56–69.

Parsons DS, ed. Pediatric sinusitis. *Otolaryngologic Clinics of North America* 1996; **29**: 1–9.

Richardson MA, Osguthorpe JD. Surgical management of choanal atresia. *Laryngoscope* 1988; **98**: 915.

Shapiro EP *et al.* Bacteriology of the maxillary sinuses in patients with cystic fibrosis. *Journal of Infectious Diseases* 1982; **146**: 509–93.

Oral cavity and oropharynx

Birck, HG, Adams JS. Pediatric caustic ingestions. *Annals of Otology, Rhinology, and Laryngology* 1982; **91**: 650–8.

Gates GA. Salivary gland disease in children. In Healy GB, ed. *Common problems in pediatric otolaryngology*, pp. 289–95. Yearbook Medical Publishers, Chicago, IL, 1990.

Hawkins DB, Demeter MJ, Barnett TE. Caustic ingestion: controversies in a review of 214 cases. *Laryngoscope* 1980; **90**: 98–109.

Kuhn JR, Tunell WP. The role of initial cineesophagography in caustic esophageal injury. *American Journal of Surgery* 1993; **146**: 804–6.

Paradise JL *et al.* Efficacy of tonsillectomy for recurrent throat infection in severely affected children. *New England Journal of Medicine* 1984; **310**: 674–83.

Potsic WP *et al.* Relief of upper airway obstruction by adenotonsillectomy. *Otolaryngology, Head and Neck Surgery* 1986; **94**: 476–80.

Larynx and trachea

Cohen SR. Unusual presentations and problems created by management of foreign bodies in the aerodigestive tract in the pediatric patient. *Annals of Otology, Rhinology, and Laryngology* 1981; **90**: 316–22.

Cotton R. Management of subglottic stenosis of the larynx in infants and children: Progress report. *Annals of Otology, Rhinology, and Laryngology* 1972; **81**: 508–13.

Friedman EM. Foreign bodies. In Healy GB, ed. *Common problems in pediatric otolaryngology*. Yearbook Medical Publishers, Chicago, IL, 1990; 401–5.

Healy GB, McGill TJ, Friedman EM. Carbon dioxide laser in subglottic hemangiomas: an update. *Annals of Otology, Rhinology, and Laryngology* 1984; **93**: 370–3.

McGill TJ. Congenital disease of the larynx. *Otolaryngology Clinics of North America* 1984; **7**: 57–62.

Myer CM, Villaret DB. Subglottic stenosis. In: Hotaling AJ, Stankiewicz JA, eds. *Pediatric otolaryngology for the general otolaryngologist*, pp. 49–59. IGAKU—Shoin, New York, 1996.

Myers E, Stool SE, Johnson J. Technique of tracheotomy. In: Myers E, Stool SE, Johnson J, eds. *Tracheotomy*. Churchill Livingstone, New York, 1985: 113–24.

Neck

Tom LC *et al.* The sternocleidomastoid tumor of infancy. *International Journal of Pediatric Otorhinolaryngology* 1987; **13**: 2455–255.

Torsiglieu AJ *et al.* Pediatric neck masses: guidelines for evaluation. *International Journal of Pediatric Otorhinolaryngology* 1988; **16**: 199–210.

42.13 Spina bifida

Gordon Tang and Daniel L. Barrow

Classification
Etiology
Prenatal diagnosis
Management of myelomeningocele
Management of occult spina bifida
Outcome
Summary
Further reading

Spina bifida refers to a broad range of congenital defects resulting from improper closure of the developing neural tube. Expression of spina bifida can be as simple as a bony anomaly of the spine found in up to 30 per cent of the general population and bears no clinical significance. On the other extreme, defects involving the spinal cord can lead to lifelong disabilities such as paraplegia, hydrocephalus, genitourinary dysfunction, skeletal deformities, and mental impairment. In the United States, spina bifida is the most common major birth defect among live-born infants. The prevalence and multisystem involvement of this anomaly warrant its familiarity by pediatricians and pediatric surgeons.

Classification

Spina bifida can be divided into defects due to disorders of primary neurulation and defects due to disorders of caudal neural tube formation (Table 1). Because defects of primary neurulation generally lack a dermal covering and those of caudal neural tube formation remain covered by skin, these categories can be simplified as open and occult. Open spina bifida usually refers to myelomeningocele. In myelomeningocele, incompletely formed neural tissue pouches dorsally through a deficient spinal column covered by thin membrane (Fig. 1). Associated central nervous system (**CNS**) abnormalities are common. Hydrocephalus is present in 90 per cent and most infants harbor the craniocerebral abnormalities of the Chiari malformation. Skeletal, genitourinary, cardiopulmonary, and gastrointestinal defects frequently accompany myelomeningocele (Table 2).



Fig. 1. Dorsal view of newborn baby with myelomeningocele.

Disorders of primary neurulation (open)		*Disorders of secondary neurulation (occult)*	
Malformation	*Mechanism*	*Malformation*	*Mechanism*
Myelomeningocele	Failure of posterior neural tube closure	Lipomyelomeningocele	Premature dysjunction
		Dermal sinus	Incomplete dysjunction
		Split cord malformation	Failure of midline integration
		Thickened filum	Disordered caudal cell mass development
		Myelocystocele	Caudal cell mass disorder

Table 1 Classification of spina bifida

Skeletal	Hip dislocation, clubfoot, kyphoscoliosis, sacral malformations
Genitourinary	Hydronephrosis, hydroureter, bowel/bladder, endocrine/sexual issues, hydrocele
Gastrointestinal	Malrotation, omphalocele, Meckel's diverticulum, inguinal hernia

Table 2 Anomalies associated with spina bifina

Occult spina bifida encompasses a range of anomalies including myelocystocele (cystic dilation of the central canal of the caudal neural tube), split cord malformation, lipomeningocele, lipoma, teratoma, and dermal sinus (track extending inward and occasionally into the cord). Nearly all disorders are unified by the finding of an abnormal conus and filum (Fig. 2). These abnormalities frequently ‘tether' or fix the distal cord at its caudal end thereby impairing normal mobility. Eighty per cent of cases are heralded by a dermal lesion in the lumbosacral region such as dimples, tufts of hair, cutaneous tracts, a subcutaneous mass (Fig. 3), or a hemangioma. Other associated CNS abnormalities are uncommon. Due to the overlying skin and variable neurologic deficits, occult spina bifida often goes undetected.

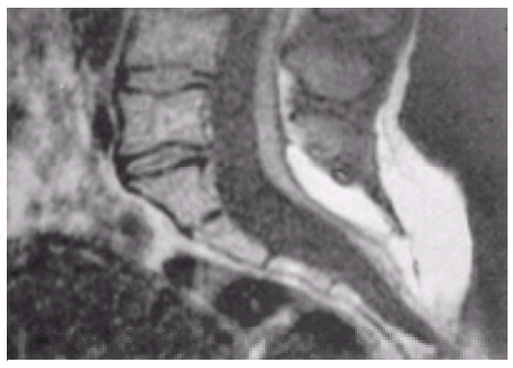


Fig. 2. Mid-sagittal magnetic resonance image of lipomeningocele with elongated tethered filum terminale. Lipomatous structure extends into subcutaneous space.



Fig. 3. Lipoma.

Etiology

Clinically significant spina bifida affects 0.15 per cent of the Caucasian population and 0.04 per cent of the black population. Observations of female preponderance, ethnic differences, parental consanguinity, twin concordance, and increased incidence in siblings favor a genetic role. Multifactorial inheritance, single mutant genes, and chromosomal abnormalities have been related to myelomeningocele. For counseling purposes, families with one child with spina bifida have a 5 per cent risk of having a subsequent affected sibling. The risk of spina bifida in a child of one affected parent is roughly 3 per cent.

Large geographic, seasonal, and economic variations had long suggested an environmental influence. A large-scale, randomized, double-blinded, placebo-controlled trial in 1991 conclusively demonstrated the preventive role of periconceptional folate supplementation. Folate deficiency is now understood to account for half of neural tube defects. The Centers for Disease Control currently recommend folate-enriched foods or 0.4 mg/day for all women contemplating pregnancy. Maternal exposure to hyperthermia, valproic acid, and carbamazepine have been identified as potential teratogens of uncertain quantitative significance.

Prenatal diagnosis

Prenatal diagnosis of spina bifida is based on levels of a-fetoprotein in amniotic fluid. a-Fetoprotein can be detected 30 days after conception and peaks at 10 to 13 weeks of gestation. Raised a-fetoprotein can also be due to fetal blood contamination and multiple gestation as well as gastrointestinal and genitourinary anomalies such as omphalocele, gastroschisis, and renal agenesis. Improved ultrasonography has a sensitivity approaching 100 per cent in the diagnosis of myelomeningocele and may supplant amniocentesis.

Management of myelomeningocele

Management of myelomeningocele begins with its prenatal diagnosis. Parental counseling begins at this point. Experimental and clinical evidence that neurologic injury in myelomeningocele is due to chronic mechanical injury and amniotic fluid-induced chemical trauma has prompted attempts to ‘rescue' neurologic function by *in utero* repair of spina bifida. Although initial results have been encouraging, the technique is still regarded as experimental. Additional prenatal consideration should also be given to the optimal mode of delivery. In a recent review of 160 myelomeningocele cases, infants delivered by cesarian section before the onset of labor had a mean level of paralysis 3.3 segments below the anatomic level of the spinal lesion as compared with 1.1 for infants delivered vaginally.

Following delivery, the exposed neural placode should be covered with a sterile, saline-soaked gauze. The infant should be kept on his or her abdomen to reduce trauma to the neural tissue. One-third of these patients subsequently gain motor function not previously detected and the exposed neural tissue should be regarded as functional. Closure of the back lesion within the first 72 h limits the risk of infection. Prophylactic antibiotics are weakly supported by the literature. Immediate closure is not necessary. Time is available to search for coexistent anomalies and to allow parents the time to adjust to the new issues before them. Although most anomalies are not life threatening, it is important to appreciate that myelomeningocele may have fatal associated malformations that may alter care plans. Hydrocephalus accompanies myelomeningocele in up to 90 per cent of cases and a concomitant ventriculoperitoneal shunt is generally preferred.

Beyond the perinatal period, management centers on treatment of attendant orthopedic and genitourinary anomalies while maintaining vigilant monitoring for evidence of neurologic deterioration. Shunt malfunction in this population can be the root cause for any deterioration ranging from signs of raised intracranial pressure such as headache and nausea to subtle changes in behavior, tone, or bowel habits. Consequently, shunt function should always be checked before entertaining other treatments for clinical deterioration. Neurologic decline may also be due to cord tethering, syringomyelia, and hindbrain dysfunction related to a Chiari malformation. In addition to neurologic decline, urinary tract complications are the major source of mortality after the first year of life and early urodynamic evaluation can aid management.

Recent years have brought increasing awareness of the high incidence of latex sensitivity in patients with spina bifida. The prevalence of a positive skin prick test for latex ranges from 2.97 to 32 per cent. Symptoms range from urticaria and conjunctivitis to catastrophic anaphylaxis. Accordingly, latex-free products should be used in medical and surgical procedures in all patients with spina bifida. A prior history of atopy and a large number of prior operations correlate to latex sensitization.

Management of occult spina bifida

Incidental identification of suggestive cutaneous stigmas often initiates consideration of occult spina bifida. Neurologic deficits in infants with occult spina bifida are unusual. Work-up begins with radiographic evaluation. Poor ossification of perinatal posterior spinal elements makes ultrasonography possible. If both ultrasound and plain radiographs are normal, no further work-up is needed. Abnormalities can be further evaluated by magnetic resonance imaging. Beyond infancy, the most common clinical presentations for occult spina bifida include developmental delay of sphincter control or walking, lower extremity asymmetry, and back or leg pain. On rare occasions, spina bifida will present in adulthood with sudden neurologic decline following minimal trauma, patient positioning in the operating room, and even following vaginal delivery.

Surgery in occult dysraphism aims to prevent neurologic decline. Whether the lesion is a lipoma, split cord malformation, or a thickened filum terminale, surgery focuses on untethering. Although there is no consensus on timing of surgery, early surgery may minimize the potential for further decline. Progressive symptomatology warrants urgent repair. Normal development generally follows appropriate repair.

Outcome

In patients with myelomeningocele, ‘conservative' therapy in the 1950s led to an 85 to 90 per cent 10-year mortality. Thereafter, unselected early closure and

aggressive management of hydrocephalus improved survival but at the expense of quality of life. The 40 to 50 per cent who survived at 16 years had a mean IQ of 77 and only 55 per cent were ambulatory. These experiences bred an early nihilistic attitude. In the modern era, improved vigilance, aggressive therapy with refined technology, and infection control has lowered early mortality to approximately 10 per cent with 80 per cent demonstrating normal cognitive development. In patients with occult spina bifida, normal development typically follows appropriate surgical repair.

Summary

Spina bifida is a common anomaly with a broad spectrum of clinical disease. Periconceptional folate supplementation may reduce the incidence of spina bifida by half. Improved screening programs and modern ultrasound allow reliable intrauterine diagnosis. The introduction of *in utero* repair has the potential for further reduction of postnatal neurologic deficits. The universal need for lifelong neurologic, orthopedic, and urologic care and surveillance strongly favor co-ordinated care by a multidisciplinary team. Meticulous vigilance and aggressive management of multisystem defects can be expected to provide long-term survival with a high quality of life. In occult spina bifida, awareness of cutaneous stigmas, radiographic work-up, and early surgical repair can be expected to provide normal development.

Further reading

Adzick NS, Sutton LN, Cromleholme TM, Flake AW. Successful fetal surgery for spina bifida. *Lancet* 1998; **352**: 1675–6.

Carter CO, Evans K. Children of adult survivors with spina bifida cystica. *Lancet* 1973; **ii**: 924–6.

Luthy DA *et al.* Cesarean section before the onset of labor and subsequent motor function in infants with meningomyelocele diagnosed antenatally. *New England Journal of Medicine* 1991; **324**: 662–6.

McLone DG. Results of treatment of children born with myelomeningocele. *Clinical Neurosurgery* 1983; **30**: 407–12.

McLone DG. Care of the neonate with myelomeningocele. *Neurosurgery Clinics of North America* 1998; **9**: 111–20.

Medical Research Council. MRC randomized controlled trial use of multivitamins and folic acid for prevention of recurrence of neural tube defects. *Lancet* 1991; **338**: 1–7.

Reigel DH, Rotenstein D. Spina bifida. In: Cheek WR, Marlin AE, McLone DG, Reigel DH, Walker ML, eds, *Pediatric neurosurgery*, pp. 51–76. WB Saunders, Philadelphia, 1994.

Volpe, JJ. *Neurology of the newborn*, pp. 9–21. WB Saunders, Philadelphia, 1995.

Werler WM, Shapiro S, Mitchell AA. Periconceptional folic acid and risk of occurent neural tube defects. *Journal of the American Medical Association* 1993; **269**: 1257–61.

43.1 Surgery of the spleen

P. Jane Clarke and Peter J. Morris

Introduction
Embryology
Congenital abnormalities
Anatomy
Relations
Surface anatomy
Splenic vasculature
Microscopic anatomy
Splenic function and physiology
Phagocytosis
Immune response
Erythrocyte storage
Cytopoiesis
Indications for splenectomy
Splenic trauma
Splenectomy for haematological disease
Splenectomy for other reasons
Operations on the spleen
Historical background
Operations
Complications of splenectomy
Further reading

Introduction

The spleen was a source of intrigue to ancient physicians and philosophers alike. Galen described the spleen as an organ of mystery, and believed that it extracted ‘melancholy' from the circulating blood and from the liver, and returned it to the stomach after a process of purification. Although it was noted by Aristotle that the asplenic state is compatible with life, the complexity of its function and anatomy are still a major topic of medical and scientific study. Thus the care of patients with splenic disorders often requires a multidisciplinary team of haematologists, pathologists, and surgeons.

Embryology

At 5 weeks of gestation, the part of the gut tube destined to become the stomach appears as a fusiform swelling in the foregut. It is attached to the dorsal body wall by a peritoneal fold, the dorsal mesogastrium. During the following weeks the stomach rotates in both its longitudinal and anteroposterior axes. As a result, the pyloric end moves to the right and the cardiac to the left; the left side faces anteriorly and the right posteriorly. These rotations result in the dorsal mesogastrium being pulled out and pouched to the left of the median plane. The pouch developed in this way lies behind the stomach and forms the omental bursa. At the same time, the fetal splenic tissue develops from condensations of mesoderm in the dorsal mesogastrium. This condensation has the effect of dividing the mesogastrium into two parts, that between the stomach and the fetal splenic tissue forming the gastrosplenic ligament and that between it and the kidney becoming the lienorenal ligament. The mesenchymal condensations then fuse to form the spleen.

Congenital abnormalities

Simple splenic agenesis is rare as an isolated abnormality but is found in 4.0 per cent of children with congenital cardiac disease. Because the spleen develops from independent collections of mesoderm that then fuse, both accessory spleens (splenunculi) and polysplenia can occur. Splenunculi are found in approximately 10 to 30 per cent of the population, and are therefore one of the most frequent congenital anomalies. They usually occur in the gastrosplenic ligament and the greater omentum. The spleen may retain its fetal, lobulated form or show deep notches on its diaphragmatic surface. In the rare condition of polysplenia, two to nine distinct parts are found, owing to failure of splenic fusion.

Anatomy

Relations

The spleen is a lymphatic organ situated in the left hypochondrium between the gastric fundus and the left hemidiaphragm. Its long axis is in the line of the tenth rib; the hilum is in the angle between the stomach and the left kidney and makes contact with the tail of the pancreas. It is invested by visceral peritoneum. The diaphragmatic surface is moulded into a reciprocal convexity and the visceral surface has impressions from the stomach, left kidney, pancreas, and splenic flexure. Its size and weight vary depending on age and in different conditions, but a normal adult spleen weighs approximately 150 g (range 80–300 g), and measures 12 × 7 × 3 cm.

Surface anatomy

The long axis of the spleen lies along the tenth rib. The lower pole does not normally project beyond the midaxillary line. When the spleen enlarges the long axis extends down along the tenth rib and the anterior border approaches the costal margin. The spleen must at least treble in size before becoming palpable, which it does by passing in front of the splenic flexure of the colon. During clinical examination the spleen is palpated as a mass in the left upper quadrant that is ballotable, and is identified by the splenic notch and the lack of resonance to percussion over the mass (Fig. 1; see also Fig. 4).

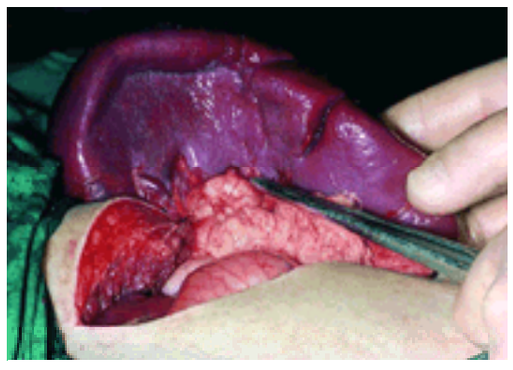


Fig. 1. An enlarged spleen delivered through a subcostal incision in a patient with Hodgkin's disease, showing the splenic notch.

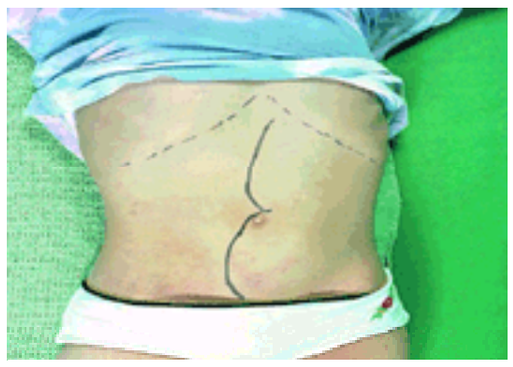


Fig. 4. An adolescent girl with b-thalassaemia who was transfusion dependent; the border of an enormous spleen is outlined, showing also the splenic notch.

Splenic vasculature

The arterial supply to the spleen passes through the splenic artery, a tortuous branch of the coeliac axis (usually from a common stem with a hepatic artery), running along the superior surface of the body and tail of the pancreas. The short gastric and left gastroepiploic branches of the splenic artery pass between the layers of the gastrosplenic ligament. The splenic artery divides at the splenic hilum into superior and inferior branches, which then further divide into four or five segmental branches, each serving one segment of spleen. Radiopaque dye injection and corrosion casting techniques demonstrate each segment to be distinct, with a separate arterial supply and venous drainage. Such studies have clarified the existence of pyramidal segments in the polar regions and wedge-shaped segments in the central part of the spleen. Further subsegmentation has also been demonstrated, being similar in appearance and blood supply to accessory spleens or splenunculi. Vascular anastomoses between segments are uncommon. This knowledge of the arrangement of the segmental vasculature is important when considering splenic salvage surgery (see below).

The splenic vein forms from five or more tributaries leaving the hilum. It then runs behind the pancreas to join the superior mesenteric vein, thus forming the portal vein. Its tributaries correspond to the arterial branches. The lymphatic drainage of the spleen comprises extensive efferent vessels in the white pulp that run with the arterioles and emerge from nodes at the hilum. These nodes drain through the retropancreatic nodes to those at the coeliac axis. There are no afferent vessels. Sympathetic nerve fibres run from the coeliac plexus and mainly innervate branches of the splenic artery.

Microscopic anatomy

The spleen is surrounded by serosa and a collagenous capsule from which trabeculae penetrate the parenchyma. The trabeculae are dense connective-tissue fibres, rich in collagen and elastic tissue. In many mammals they contain non-striated myocytes but in man these are few in number. Between the trabeculae there is a network of reticular fibres supporting the splenic parenchyma, which consists of the red and white pulp separated by the marginal zone.

On macroscopic examination of the cut surface, the white pulp appears as discrete, grey–white nodules embedded in a red matrix, the red pulp. The white pulp consists of periarteriolar lymphatic sheaths and lymphoid follicles, whereas the red pulp (which comprises approximately 75 per cent of the splenic volume) consists of the venous sinusoids and the splenic cords (of Billroth).

The microscopic anatomy of the red and white pulp is based on a complex vascular arrangement (Fig. 2). Blood flow from the splenic artery enters the trabecular arteries from which central arteries supply the white pulp. The adventitia of the central arteries is then replaced by a periarteriolar lymphoid sheath (predominantly of T cells), and at various points these enlarge to form the splenic follicles. When stimulated by antigen the follicles form germinal centres, composed principally of B lymphocytes, that are similar in structure to the lymphoid follicles in lymph nodes. The central arterioles lie to one side of the germinal centres; they then lose their periarteriolar lymphoid sheath and continue through the transitional zone into the red pulp as several, very straight branches or penicillary arterioles (penicilli). The red pulp is composed of numerous, thin-walled venous sinusoids separated by the splenic cords, sponge-like structures consisting mainly of macrophages that are loosely connected via dendritic processes to form a filter through which blood can seep slowly. The blood entering the red pulp from the penicillary arterioles can take two routes to reach the splenic veins. The ‘open' route is via the splenic cords, with blood slowly filtering out to the sinusoids; the ‘closed' route involves direct drainage from the penicillary arterioles to the veins without passing via the cords. It is thought that the majority of the circulation in health is via the closed, more rapid, circulatory pathway. The pulp venules collect the blood from both the sinuses and the cords and carry it to the trabecular veins, and thence to the hilum and the splenic veins.

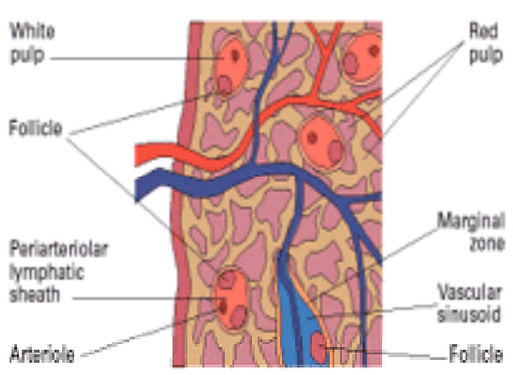


Fig. 2. A simple diagram of the anatomy of the spleen showing the white pulp, which comprises the arterioles, the periarteriolar lymphatic sheath, and the lymphoid follicles. The white pulp is embedded in the much larger mass of red pulp made up of the erythrocyte-filled sinusoids.

Splenic function and physiology

Phagocytosis

A major function of the spleen is phagocytosis. Effete and damaged red cells are removed daily from the circulation, as well as particulate foreign matter, microbes, antigens, and cellular debris. This process occurs in the sinusoids and the splenic cords by the action of the endothelial macrophages. In addition, intact cells are held up and siderotic granules, as well as Howell–Jolly and Heinz bodies (nuclear remnants and precipitated haemoglobin or globin subunits, respectively), are removed before the red cells are returned to the circulation.

Immune response

The spleen comprises the largest single accumulation of lymphoid cells in the body, containing 25 per cent of the total T-lymphocyte population and 10 to 15 per cent of the B-lymphocyte population. Blood-borne cells and antigens are trapped in the periarteriolar lymphatic sheaths and presented to immunocompetent cells, leading to antibody production by plasma cells and an increase in size of the germinal centres in the lymphoid follicles. The spleen is particularly involved in the clearance of particulate antigens, immune complexes, and the antibody-coated cells generated in autoimmune responses. After splenectomy, patients have a decreased ability to generate an IgM response, a decreased capacity to respond to polysaccharide antigens, there is a decrease in the activity of their alternative complement pathway, and their production of tuftsin is defective. It is possible that these deficiencies in immune function explain the problem of overwhelming postsplenectomy infection that sometimes complicates the asplenic state (see below).

Erythrocyte storage

This function is less marked in man than other species but the spleen does contain a large volume of blood (approximately 8 per cent of the red cell mass) either in the venous sinuses or in the reticular meshes of the cords. During emergencies such as anoxia, large volumes of blood can be discharged into the circulation. Enlarged

spleens may contain a much larger proportion of the blood volume (up to 40 per cent), including platelets and white cells.

Cytopoiesis

The red pulp contains groups of myelocytes, erythroblasts, and megakaryocytes. From the fourth month of intrauterine life some degree of haemopoiesis occurs in the human fetus within the spleen, although this is a minor site compared with the liver. Although a large bulk of lymphoid tissue, the mature spleen is not a major site of lymphopoiesis. However, stimulation in the white pulp by antigenic challenge does result in the proliferation of germinal centres, which contributes to the circulating pool of competent T and B cells and macrophages. This may also occur in disease states, especially in myeloproliferative disorders, thalassaemias, and chronic haemolytic anaemias.

Indications for splenectomy (Table 1)

Indication	*Pathology*
Bleeding	Trauma
	Iatrogenic
	Spontaneous
As part of a surgical resection	During total gastrectomy, distal pancreatectomy
Diagnostic	Splenomegaly of unknown cause
	Pyrexia of unknown origin
Staging	Hodgkin's disease
Therapeutic	Hypersplenism
	Idiopathic thrombocytopenic purpura
	Haemolytic anaemia
	Giant spleen with symptoms
Other	Splenic abscess
	Hydatid disease

Table 1 Indications for splenectomy

Splenic trauma

Diagnosis

Damage to the spleen may result from accidental or iatrogenic trauma, or may present as ‘delayed' or ‘spontaneous' rupture. It has been estimated that more than 20 per cent of all splenectomies are for damage to the spleen occurring during the course of another abdominal operation. The most common type of non-iatrogenic injury to result in splenic trauma is non-penetrating, blunt trauma to the upper abdomen, often associated with fractures of the left lower ribs. Delayed rupture is said to occur when the clinical signs develop after a delay of at least 48 h from the initial injury. It is generally believed to result from tearing of the capsule by the expansion of a subcapsular haematoma. The clinical management and presentation are the same as for immediate rupture. ‘Spontaneous' rupture of a diseased spleen usually results from trivial trauma, often forgotten by the patient. It occurs most commonly as a complication of malaria, but is the most common cause of death in infectious mononucleosis.

If the trauma is localized to the spleen, in the absence of other intraperitoneal damage, physical signs depend on the degree of blood loss; they vary from minimal tenderness in the left upper quadrant in an otherwise well patient to signs of peritonitis accompanied by evidence of shock. Peritoneal lavage can aid the diagnosis of intra-abdominal bleeding, especially in cases that are otherwise difficult to assess, such as an obtunded patient and those with multiple potential sites of blood loss. After clinical assessment of the degree of blood loss and appropriate resuscitation with colloid or blood, a management plan as suggested by the algorithm in Fig. 3 can be followed. The stable patient should be investigated with abdominal ultrasonography, computed tomography (**CT**), or splenic scintiscanning.

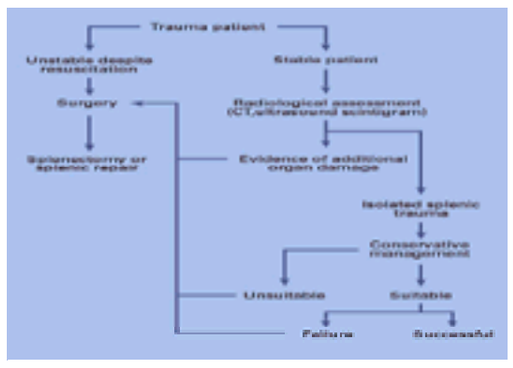


Fig. 3. An algorithm for the management of a patient with splenic trauma.

Operative management

Laparotomy is indicated for splenic trauma if there is obvious evidence of continuing blood loss despite adequate resuscitation, or if there is clinical suspicion of additional trauma to other organs such as the liver, pancreas, or bowel. In several series of adult patients with abdominal trauma, the incidence of major intra-abdominal injuries associated with blunt splenic trauma is approximately 30 to 60 per cent. If there is any doubt about the presence of an associated intra-abdominal injury, laparotomy should not be delayed. The surgical treatment of these injuries is dictated by the operative findings. At laparotomy the decision to perform splenectomy or splenic repair (splenorrhaphy), the techniques of which are discussed later, will depend largely on the degree of trauma and the expertise available. A simple grading system with suggested management is shown in Table 2.

Grade I:	Capsular injury not actively bleeding
Grade II:	Capsular or parenchymal injury requiring only topical haemostatic agents
Grade III:	Parenchymal injuries requiring suturing + haemostatic agents
Grade IV:	Parenchymal injuries requiring partial splenic resection
Grade V:	Injuries requiring splenectomy

Table 2 A grading system of splenic injuries and their management

The proportion of splenic injuries suitable for repair will vary with the type of trauma seen in an individual centre and ranges from 30 to 90 per cent. The presence of extensive hilar injury makes it unlikely that the spleen will be suitable for repair, and avulsion or extensive fragmentation requires splenectomy, as does failure to achieve haemostasis following attempted splenorrhaphy. If repair is undertaken in conjunction with resection of splenic tissue, consideration should be given to the

amount of spleen that will remain. The ‘critical mass' required to confer substantial protection from bacterial infection in experimental models seems to be approximately one-quarter to one-third. Repair in the presence of significant peritoneal contamination is controversial, as is the repair of a diseased, ruptured spleen, but is probably inadvisable.

After splenic salvage surgery, careful postoperative monitoring is necessary to alert the surgeon to rebleeding. If the clinical condition deteriorates and there is evidence of continuing blood loss, re-exploration of the abdomen should not be delayed. This assessment may be difficult, or indeed impossible in the presence of multiple other injuries, and such circumstances may make splenectomy preferable to attempted repair.

The complications of splenic repair are few, but reported series are small. True ‘failures' seem rare and the majority represent inadequate assessment of the damage at the time of laparotomy, missed injuries, or overambitious attempts to repair severely damaged spleens.

Non-operative management

The suspicion of splenic trauma in a patient whose cardiovascular system is stable can be investigated with ultrasound, CT, or scintiscanning. If the radiological investigation confirms the presence of a splenic injury, a decision should be taken on whether it is appropriate to manage the patient conservatively. The proportion of cases judged to be suitable for non-operative management varies from 20 to 45 per cent. Patients are ideally cared for in a high-dependency or intensive care unit, where careful cardiovascular and haematological monitoring with repeated clinical examination is performed. Bed rest is imposed, and, assuming the patient remains stable, discharge to the ward can occur after 24 to 48 h of observation, after which activity can be increased up to discharge. It is prudent to restrict activity for 4 to 6 weeks and to avoid contact sports for up to 6 months. Regular rescanning (by either ultrasonography or isotope scinti-graphy) will monitor resolution of the trauma, which is complete at 3 months in 90 per cent of cases. Deterioration, or evidence of bleeding not controlled by transfusion, is an indication for surgery.

Up to nearly one-third of cases managed in this way will ‘fail' and require surgery. Variations between series (failure rates of 0 to 30 per cent) reflect the differences in clinical judgement about which patients should be treated non-operatively. Most instances in which conservative management fails and a laparotomy is needed result in the patient losing their spleen. Indeed, the protagonists of splenic repair would argue that more spleens would be preserved overall if early laparotomy were to replace attempts at conservative management.

Conservative (non-operative) management of splenic trauma has been most widely practised in children, because the increased incidence of overwhelming postsplenectomy infection in the child behoves the surgeon to attempt to conserve the spleen if possible. Fortunately, there are various reasons why conservative management of splenic injury in children is more likely to be successful than in adults. The type of injury suffered by children is less likely to be of the penetrating type (as associated with crimes of violence), and multiple injuries are also less common. Both criteria make non-operative management safer in the child. Furthermore, children have a better capacity for haemostasis, an increased resilience of the cardiovascular system to hypovolaemia, and an increased compliance of the splenic capsule and septa. All these factors improve the chances of a successful outcome for conservative management.

Splenectomy for haematological disease

Splenectomy may be required in the management of a variety of haematological disorders. Indications include the need to alleviate complications, to prevent repeated episodes of infarction, to prevent massive red-cell pooling and as part of the management of hypersplenism.

Hypersplenism is a clinical syndrome with many causes (Table 3). It is characterized by splenic enlargement, haematological cytopenia (reduction in one or more cellular components of the blood as a result of splenic pooling), maturation arrest in the marrow, and the premature release of immature cells into the circulation. Red cell and/or platelet survival may be decreased. Anaemia, neutropenia, and thrombocytopenia can all occur, either alone or in any combination, their relative severity depending on the nature of the underlying disease. Usually the decision whether the patient will benefit from splenectomy is made on the basis of knowledge about the particular disease process and the severity of the haematological cytopenia, but in difficult cases it may be necessary to use special tests of splenic function. These may include isotope studies of the splenic red-cell pool, splenic erythropoiesis and tests of red-cell survival, but despite their use the haematological outcome of splenectomy may be difficult to predict and the decision to operate is made by a multidisciplinary team that includes the haematologist, surgeon, immunologist, and microbiologist.

Lymphoma (Hodgkin's disease and non-Hodgkin's lymphoma)
Chronic lymphatic leukaemia
Hairy-cell leukaemia
Portal hypertension (Banti's syndrome)
Rheumatoid arthritis (Felty's syndrome)
Infection (e.g. malaria and kala-azar)
Infiltrative diseases (e.g. sarcoid)
Lipid storage disease (e.g. Niemann–Pick syndrome, Gaucher's disease)

Table 3 Causes of hypersplenism

Idiopathic thrombocytopenic purpura

Idiopathic thrombocytopenic purpura mainly affects females between the ages of 15 and 50 years. It presents with bruising, usually after trauma or pressure, and examination reveals variable numbers of petechial haemorrhages in the skin. The clinical course is often intermittent and chronic. The platelets of affected patients become sensitized by antiplatelet IgG autoantibodies and are then removed from the circulation. Many patients need no treatment if their platelet count remains over 50×10^9/l and no spontaneous bleeding occurs. If the count falls below 20×10^9/l the risk of bleeding increases. Spontaneous recovery occurs in under 10 per cent of patients and treatment is initially directed at decreasing the autoantibodies and the rate of platelet destruction. Eighty per cent of patients go into remission on steroid therapy (60 mg prednisolone/day). If this does not result in remission, or continuing high doses are needed, or the platelet count remains low (under 20–30 $\times 10^9$/l), splenectomy is indicated, the spleen being the initial and major site of antibody production and also a major site of platelet destruction. Good results are usually obtained, with approximately 80 per cent of patients having a satisfactory outcome, no longer requiring steroids to maintain an adequate number of platelets. However, in more heavily sensitized patients, platelets can continue to be destroyed elsewhere in the reticuloendothelial system, including the liver, and after an initial good response a relapse may also occur secondarily to an acute bacterial or viral illness. A good response is most likely in patients under the age of 45 years, in those in whom the thrombocytopenia is less severe, and in those who have shown at least an initial response to steroid therapy. At operation, the spleen is usually of normal size and no special technique is needed for its removal, although a careful search must be made to ensure that no splenunculi remain. Although the preoperative platelet count may be very low, platelet transfusion is not begun until the splenic vessels have been ligated, when 6 to 10 units of platelets may be transfused. In addition, some patients may benefit from high doses of intravenous gammaglobulin preoperatively to raise the platelet count to an acceptable number (greater than 20×10^9/l).

Haemolytic anaemias

Hereditary spherocytosis

This is an autosomal-dominant hereditary disorder characterized by small, spherocytic red cells. Abnormalities in the cell membrane transform the red cells to a spherical shape, which results in a reduced ability to be deformed and therefore to entrapment in the spleen. The clinical presentation is commonly in childhood but may be delayed until later in life. Mild, intermittent jaundice is associated with mild anaemia, splenomegaly, and gallstones. The diagnosis is made on examination of a peripheral blood film, in which reticulocytes and spherocytes are seen. Serum bilirubin is commonly elevated.

If the disease is severe, splenectomy is the treatment of choice but surgery is normally delayed until over the age of 6 years (and always until the age of 2 years), to minimize the risk of postsplenectomy infection. The red cells retain their spherical appearance but in the absence of the spleen they have a near-normal survival. The long-term results of the operation are excellent, relapses being attributed to missed splenunculi, but with no very convincing evidence.

The decision to operate for mild to moderate disease is more difficult and the treatment of patients with asymptomatic anaemia is controversial. The efficacy of partial splenectomy to relieve haemolysis but maintain phagocytic function is an attractive possibility currently under study.

Hereditary elliptocytosis

This condition is characterized by oval-shaped red cells caused by an abnormality in the cell membrane. There are three types, two inherited as an autosomal-dominant and one as a recessive trait. Most heterozygous carriers require no treatment, but if the anaemia is symptomatic, partial relief can be obtained from splenectomy.

b-Thalassaemia

The management of this condition involves regular blood transfusion, chelating agents to minimize the problem of iron overload, and the judicious use of splenectomy if hypersplenism occurs and the patient becomes transfusion dependent. Marked splenomegaly can occur (Fig. 4, Fig. 5). If indicated, surgery should be delayed until at least 5 years of age.



Fig. 5. The mobilized spleen of the patient shown in Fig. 4.

Sickle-cell disease

In this condition, splenic infarction can occur, and only in rare cases is removal of the spleen indicated. In cases of sequestration crisis there is rapid enlargement of the spleen or liver that may be fatal. If the infant or baby survives, these episodes have a tendency to recur and splenectomy may be required. In addition the spleen may enlarge to such a degree that hypersplenism develops, a problem most common in areas in which malaria is endemic.

Lymphoma

The surgeon may be required to remove the spleen for a variety of reasons in the management of patients with lymphoma (Fig. 6). Occasionally, splenectomy may be needed to achieve a diagnosis (in the absence of palpable nodes for biopsy) and this is becoming more frequent in human immunodeficiency virus-related disease, in which splenic lymphoma may develop. Splenectomy may also be indicated in hypersplenism (see above), or to relieve the symptoms of gross splenomegaly (such as satiety and discomfort in the left upper quadrant). Staging laparotomy for either Hodgkin's disease or non-Hodgkin's lymphoma is now rarely performed. In Hodgkin's disease, accurate staging is important as it has a bearing on both prognosis and the selection of treatment for an individual patient, but no clear survival advantage has been demonstrated in patients who have been surgically staged. This probably relates both to the success of salvage chemotherapy for disease that relapses after radiotherapy, and an increased tendency to use chemotherapy for early, bulky disease. Hence CT is increasingly replacing surgical staging, although the technique is not without its limitations. It cannot detect disease in normal-sized nodes, differentiate diseased from reactive nodes, or demonstrate micronodular involvement of the spleen (Fig. 7). Magnetic resonance imaging can be useful when there is a suspicion of focal bone or marrow involvement.

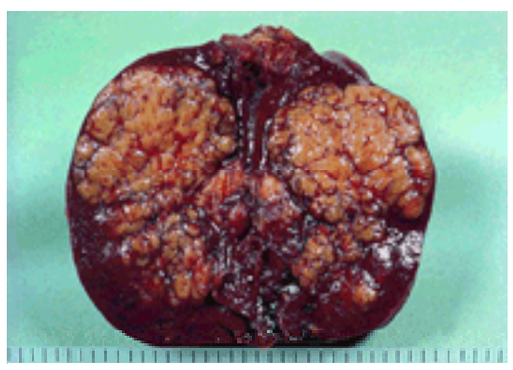


Fig. 6. An enlarged spleen showing extensive involvement with disease in a patient with stage IV Hodgkin's disease. The patient had a pancytopenia due to hypersplenism, precluding chemotherapy. Splenectomy led to haematological remission allowing the initiation of chemotherapy, which produced a complete remission in disease.

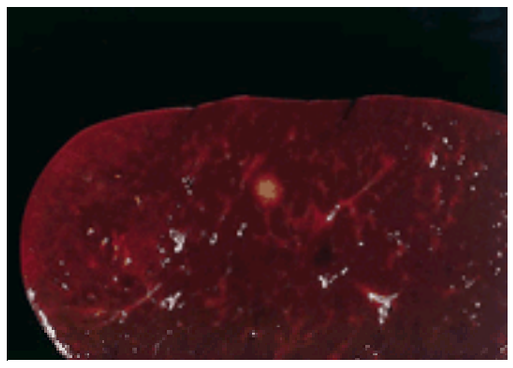


Fig. 7. The spleen of a patient with clinical stage 1A Hodgkin's disease who presented with nodes in the neck. No disease was evident in the abdomen at laparotomy but one positive nodule of disease was found on one cut surface of the spleen.

In low-grade non-Hodgkin's lymphoma, splenectomy may be helpful if the disease is largely confined to the spleen, especially if hypersplenism is a feature.

Chronic lymphatic leukaemia

In chronic lymphatic leukaemia, 50 per cent of patients will have splenomegaly of variable degree, and splenectomy can be of value in selected cases. Indications for surgery include those cases resistant to medical management in which splenomegaly is present, when hypersplenism develops, or if the patient develops autoimmune

complications (such as haemolytic anaemia or thrombocytopenia) and this fails to respond to medical treatment. In patients with chronic lymphatic leukaemia where the spleen is the main organ affected, splenectomy can downstage the disease and can lead to an improvement in survival. It is worthwhile to remember that these patients have poor humoral immunity and so postsplenectomy prophylaxis against infection will rely mainly on antibiotics rather than immunization.

Hairy-cell leukaemia

In hairy-cell leukaemia, a low-grade, B-cell condition of middle-aged men, two-thirds of patients have splenomegaly, and premature destruction of blood cells occurs during their passage through the spleen. In the past, splenectomy has been the treatment of choice, but is now reserved for patients with a very large spleen. Interferon-a improves the blood count and the bone marrow but rarely induces a complete remission; new drugs, including deoxycoformycin, lead to remission rates of approximately 75 per cent.

Myelosclerosis

This is a disease for which there is no definitive treatment. It may present with splenomegaly, which may be massive, but the place of splenectomy is controversial. If the spleen is removed early in the disease it will prevent the development of hypersplenism, but the operation may be technically difficult and followed by a massive thrombocytosis that may prove resistant to treatment. Splenectomy is probably best reserved for cases in which splenic pain or hypersplenism are major features.

Splenectomy for other reasons

1. Abscess: a relatively rare condition, generally requiring a splenectomy.
2. Hydatid disease (see Chapter 47.2.4)
3. As part of a cancer operation: because of the local extension of malignant tumours, the spleen may need to be resected as a part of operations to remove tumours of the greater curve of the stomach, the splenic flexure, or the tail of the pancreas. Splenectomy as part of radical gastric surgery for malignancy is no longer routinely practised.
4. Felty's syndrome (rheumatoid arthritis, splenomegaly, leucopenia): splenectomy may lead to a temporary rise in the neutrophil count but does not seem to provide reliable protection against recurrent infection. It is best reserved for when severe anaemia is secondary to hypersplenism.
5. Systemic lupus erythematosus: this may be associated with immune thrombocytopenic purpura and the role of splenectomy is controversial; it is most successful when combined with immunosuppression.
6. Human immunodeficiency virus infection: in the later stages of this disease, mild to moderate splenomegaly may develop and splenectomy may be necessary to exclude lymphoma or to diagnose opportunistic infection.
7. Immunosuppression: splenectomy will prolong renal allograft survival in experimental models in the rat, and has been shown to produce improved survival of human renal allografts in association with conventional immunosuppressive drugs. However, as time progresses there is an increased death rate from overwhelming infection; as a result, splenectomy is no longer practised as part of an immunosuppressive protocol in clinical renal transplantation.

Operations on the spleen

Historical background

The observation that the asplenic state was compatible with life was made by Aristotle and was confirmed by experiments in the seventeenth and eighteenth centuries by Wren and Morgagni. It is unknown when the first splenectomy was done but splenectomy as treatment for a disease process was possibly first performed in 1549 by Adriana Zaccarello (although there is dispute as to whether the removed organ was in fact an ovarian mass!). The first splenectomy for trauma was in 1678 by Nicholas Matthias. In 1928, William Mayo reported a series of 500 splenectomies, with a mortality rate of 10 per cent. In the absence of a clear understanding of splenic function, subsequent reports of healthy survivors led to the concept that there were no ill effects from splenectomy. Despite the fact that in 1919 Morris and Bullock had reported that asplenic rats were susceptible to infection and had a shorter lifespan than other rats, the caution advised by these investigators went unheeded for over 30 years. In 1953, a report from King and Schumacker demonstrated an increased susceptibility to infection and death from sepsis in infants who had undergone splenectomy for congenital spherocytosis. It had been noted by Billroth that haemostasis could occur in a traumatized spleen without removal and that non-operative management of splenic trauma was therefore feasible. Early in the twentieth century, initial attempts at non-operative management and splenorrhaphy had poor results and splenectomy was the rule for trauma. In the latter half of the twentieth century, with the recognition of the risk of overwhelming postsplenectomy infection, especially in children, non-operative management has become an option once more and the current management of trauma is usually dictated by the age of the patient, the experience of the institution, the individual surgeon, and the type of trauma.

Operations

Open splenectomy

General technique

Under general anaesthesia with endotracheal intubation and muscle relaxation, the patient is positioned supine on the operating table. After the intravenous administration of a prophylactic antibiotic (a penicillin or second-generation cephalosporin), the abdomen is opened through an upper midline, a left paramedian, a subcostal, or, rarely, a thoracoabdominal approach. Having opened the peritoneal cavity, a laparotomy is performed appropriate to the indication for surgery. Before mobilization of the spleen it is advisable to divide, if necessary after diathermy or ligation, any adhesions between its lower pole and the greater omentum or splenic flexure. The colon can then be retracted down with the aid of a moistened abdominal pack. The right-handed surgeon then retracts the spleen medially with the left hand in order to facilitate division of the posterior layer of the lienorenal ligament behind the spleen by sharp dissection. This is usually avascular unless splenomegaly has resulted in vascular adhesions at this site. The gastrosplenic ligament is then divided between ligatures on the short gastric vessels passing to the upper pole of the spleen, care being taken not to include any gastric tissue in the ligatures. Having divided these two ligaments the spleen can be delivered medially into the wound and the splenic pedicle identified. The splenic artery and vein are then double ligated separately, without damaging the tail of the pancreas in the splenic hilum. After checking for haemostasis the abdomen is closed and a drain is not usually needed. In the presence of oozing, a suction drain can be left in the left upper quadrant for 24 to 48 h. If there has been suspected or definite damage to the tail of the pancreas, a tube drain should be inserted to drain a potential pancreatic fistula.

Special problems

When splenectomy is done for traumatic injury, an upper midline incision is recommended in order that a full laparotomy may be performed as necessary. In such cases the lienorenal ligament may have been avulsed and the spleen may deliver into the wound with little mobilization. However, if salvage surgery is planned (see below), then mobilization must be done with extra care.

When performing splenectomy for massive splenomegaly, a subcostal incision can be used if the costal margin is wide and if the spleen does not extend below the umbilicus, in which event a midline incision is preferable. In the absence of adhesions between the diaphragm and the spleen, the size of the spleen stretches the ligaments and mobilization is not usually difficult. However, in the presence of dense adhesions, these should ideally be divided under direct vision, if necessary after ligation of the splenic artery.

Laparoscopic splenectomy

Laparoscopic splenectomy was first described in 1992; in some centres, especially those with a large paediatric practice, it is now used for the elective removal of the spleen in selected cases, especially in idiopathic thrombocytopenic purpura. The operation involves the use of either four or five ports, usually with the patient in the lateral position. Whether the main splenic vessels should be identified and stapled as the first procedure or after mobilization of the spleen is a matter of debate, but following successful splenectomy the organ is either removed in a bag via the umbilical port (after maceration) or via a small Pfannensteil incision. A specific problem is that splenunculi may be missed and left *in situ*, which may result in a poor operative result when the splenectomy was for a haematological disease such as idiopathic thrombocytopenic purpura. The size of the spleen must also be assessed preoperatively, with most series quoting an upper limit of 20 cm for successful removal. It is an operation with a considerable ‘learning curve', but in common with other laparoscopic techniques, advocates of the procedure report reductions in analgesia requirements and length of hospital stay, and a more rapid return to normal activities. To date there have been no randomized studies to compare morbidity between the open and laparoscopic operation.

Operations for splenic salvage

As the spleen is a delicate, easily traumatized organ, splenic trauma can result in haemorrhage. For the reason discussed below, salvage of the spleen is preferable to

splenectomy in certain situations. Various techniques have been described.

Haemostatic agents

For trivial trauma, pressure can be combined with topical agents to arrest bleeding. Fibrin glue is a highly concentrated form of human fibrinogen and clotting factors. It is used by spraying a thin layer directly on to the injury or injecting it into or over a fractured surface. For knife or bullet tracts it can be injected deep into the base of the injury and slowly withdrawn. This distends the tract slightly to allow for both a haemostatic and tamponade effect. The glue can also be used in association with sutures or to seal the splenic surface after a partial resection. It has advantage over cyanoacrylate adhesive in that it is less histiotoxic. Microfibrillar collagen acts by trapping and then activating platelets. It forms a firm, adherent coagulum with an affinity for moist surfaces. It has been shown to be hypoallergenic, exciting little in the way of tissue reaction, and is absorbed in 3 to 6 weeks. It should be applied with dry instruments and in sufficient amounts to cover the bleeding surface to a depth of several millimetres. Following application, pressure should be applied for about 5 min. Linear or stellate cracks can be packed and it is most successful if the surface blood flow can be reduced to a mild to moderate ooze. Other topical agents include gelatin foam, thrombin, and bovine collagen. These techniques are applicable to minor superficial lacerations or capsular tears without parenchymal damage.

Splenic artery ligation

The main splenic artery can be ligated in continuity in an attempt to reduce blood loss from a traumatized spleen; this can be done without inevitable splenic infarction and loss of function, provided the short gastric vessels are intact. Arterial ligation may also be used in addition to the other techniques described below. The branches of the splenic artery can be dissected at the hilum. The most constant is the superior polar artery, which is the first branch from the main artery before entering the hilum. The lower polar vessels may not be obvious outside the spleen and it may be necessary to ligate the main vessel distal to the superior polar branch. The branches to bleeding segments may be ligated individually, either with or without partial splenectomy. Ligation of the short gastric vessels may be indicated for bleeding from the upper pole. After ligating any of the vessels, the spleen should be carefully examined so that any devitalized tissue can be resected by sharp dissection, diathermy, or blunt finger dissection. During this technique, as vessels are encountered they may be clipped, tied, or under-run. When under-running with a stitch the parenchyma of the spleen may cut through, but the vessel wall will hold and residual bleeding can be controlled with the aid of haemostatic agents as described above. Stitches can be tied over pledgets of absorbable gelatin sponge (Gelfoam), oxidized regenerated cellulose (Surgicell), or compressed microfibrillar collagen.

Repair techniques (splenorrhaphy)

Before any repair is undertaken the spleen must be fully mobilized to assess the extent of damage. During this procedure, extra care must be taken not to strip the posterior capsule, as consequent blood loss may preclude a successful repair. Capsular avulsion is most likely to occur if the incision in the posterior peritoneum is made too close to the spleen. Any peritoneal folds to the lower pole of the spleen from the greater omentum must be carefully divided. Following this incision, blunt dissection posteriorly is continued to elevate the spleen and the tail of the pancreas. If time permits, isolation of the splenic artery with a vascular sling is ideal to allow occlusion with a vascular clamp if necessary. The veins are fragile and no attempt should be made to isolate them.

Splenorrhaphy should aim both to achieve control of bleeding and to avoid causing or leaving behind infarcted splenic tissue. Simple suture of the torn spleen often results in further bleeding, but the use of a buttress technique with collagen, omentum, or Teflon pledgets can be effective. Haemostatic agents can be used in addition to the sutures (see above). The spleen may be wrapped in the greater omentum to assist haemostasis, and Dexon and Vicryl mesh can be used to achieve a similar effect.

In view of the segmental nature of the vascular supply, if one pole or segment of the spleen has been extensively damaged, partial splenectomy can be successfully performed. Haemostatic agents or an omental buttress can then be applied to the raw splenic surface after resection.

In general, it is appropriate to drain the left upper quadrant after a repair procedure. Drainage, in addition to clinical assessment, can be used to determine the success of the procedure. These procedures are often technically more demanding than a simple splenectomy, and it must be constantly remembered that the short- and long-term morbidity of such a procedure must be less than that resulting from splenectomy in order for it to be an acceptable part of the treatment.

Complications of splenectomy

General

Certain complications are specific to the operation of splenectomy, as opposed to those seen after any abdominal operation. Atelectasis of the left lower lobe is common, and all patients should receive active physiotherapy from the first postoperative day. Gastric ileus is usually short lived and a nasogastric tube is not needed routinely. Postsplenectomy fever is described and cannot always be attributed to atelectasis or subphrenic haematoma; in the absence of a definite cause it is a self-limiting feature.

Thrombocytosis and thrombosis

The platelet count often increases to between 600 and 1000 × 10^9/l postoperatively, usually peaking between days 7 to 12. This rise is usually transitory but may last up to 3 months. If a count of 750 × 10^9/l is reached it is a sensible precaution to administer aspirin (150 mg/day) to prevent deep venous thrombosis.

Overwhelming postsplenectomy infection

Definition and cause

Splenectomy is the traditional treatment for the traumatized spleen, but it is now well recognized that there are disadvantages to the asplenic state, the most important of these being the potential for the development of overwhelming postsplenectomy infection, a term coined by Diamond in 1969. The clinical course consists of a fulminant bacteraemia, a frequent absence of a septic focus, coma, shock, consumptive coagulopathy, and adrenal haemorrhage. The bacteraemia is most commonly associated with the encapsulated organism *Streptococcus pneumoniae*, and less commonly with *Neisseria meningitidis, Escherichia coli*, and *Haemophilus influenzae*.

Removal of the spleen renders the patient more susceptible to infection for two reasons. Firstly, the absence of the red pulp results in impairment of phagocytosis and clearance of exogenous organisms. Secondly, antibody production and subsequent bacterial opsonization are reduced by the removal of the white pulp, and poorly opsonized bacteria are cleared less effectively by the liver.

Incidence

Estimates of the incidence of overwhelming postsplenectomy infection and its associated mortality vary, but large overviews found an overall incidence of approximately 1 to 5 per cent, with all series showing the incidence in children and adolescents to be consistently higher than in the adult population. The incidence of overwhelming postsplenectomy infection varies according to the indication for which the splenectomy is performed. Some 1.5 to 2.5 per cent of children develop overwhelming postsplenectomy infection when the spleen is removed for trauma, but the greatest risk is seen in patients undergoing splenectomy for congenital anaemias, portal hypertension, or lymphoreticular tumours. The time of development of overwhelming postsplenectomy infection is variable in relation to the splenectomy. Although it has been described up to 30 years after loss of the spleen, it usually occurs within 3 years after the operation. It is of particular importance, as this infection is associated with a very high mortality rate (25–75 per cent reported in many series). It is primarily to avoid this complication that a renewed interest in splenic preservation has arisen, especially in children. Attempts to preserve the spleen can be achieved by conservative, non-operative management, or by one or other of the splenic salvage operations described in the previous section.

Prophylaxis

Antibiotics The effect of prophylactic antibiotics on overwhelming postsplenectomy infection is controversial. Traditionally, oral penicillin has been given postoperatively for periods that have varied from 6 weeks to 5 years, or until the onset of puberty if splenectomy is performed in a child. As the time of greatest risk is during the first 2 or 3 years after splenectomy, it would not be unreasonable to give all patients some form of antibiotic prophylaxis at least during that time. However, many septic episodes in asplenic patients are secondary to organisms not sensitive to penicillin and patient compliance is a problem. As a compromise, it seems appropriate to treat patients with prophylactic antibiotics only if they are at high risk, such as children and immunosuppressed patients. For the prevention of infection, antibiotics may be given in reduced dosage to avoid undesirable side-effects. Penicillin V (250 mg daily) is a recommended prophylactic dose. It is likely that antibiotics are of more use when given promptly and appropriately when infection develops, and clinicians should ensure that asplenic patients are supplied with amoxycillin (amoxicillin) to take at the first sign of a respiratory illness or fever. Amoxycillin is recommended firstly because of the higher blood concentrations achieved with oral

administration, and secondly because of its activity against *H. influenzae*.

Vaccines Vaccines are available to the pneumococci, *H. influenzae*, and meningococci. Splenectomy results in a reduced antibody responsiveness and this is most marked in the young and those with malignancy or on immunosuppressive drugs. Although vaccination failures have been reported, in the case of elective surgery, vaccination is recommended and should be given as soon as a decision to operate has been taken. In a trauma victim, vaccination can be given in the postoperative period and the resulting amounts of antibody will be protective in the majority of cases, although they are less than 50 per cent of those achieved if vaccination is given in the presence of an intact spleen. Following vaccination, antibody titres remain high for at least a year before slowly declining, but this rate is increased in immunosuppressed patients and those with lymphoma. Although the incidence of pneumococcal infection has been shown to be reduced following vaccination, protection is not guaranteed. The reasons for this include the fact that about half the organisms causing postsplenectomy sepsis are not pneumococcal, and further more, 20 per cent of subtypes are not covered. Antigenic types also vary and therefore protection from a particular vaccine may be incomplete.

The pneumococcal vaccine consists of purified, capsular polysaccharide antigens of the 23 most prevalent serotypes of *S. pneumoniae*. Most healthy adults show at least a twofold rise in antibody titre within 2 weeks. Antibody titres of 256 ng immunoglobulin nitrogen/ml or more are indicative of immunity. Raised antibody titres have been demonstrated 5 years after vaccination but the exact duration of protection is not known. Current advice is to recommend revaccination every 6 years.

The vaccine against *H. influenzae* consists of bacterial polysaccharide conjugated to protein. The response to this preparation and its long-term efficacy are greater than the response to the polysaccharide vaccine alone and so its use is particularly important in children and when vaccination has to be given postoperatively.

Implantation of splenic tissue

Splenosis, or spontaneous regrowth of splenic tissue following splenectomy for trauma, possibly accounts for the lower incidence of overwhelming postsplenectomy infection in patients who have had splenectomy for trauma rather than for other reasons. Detection of splenosis by scintiscanning is complicated by the presence of splenunculi. The precise incidence of splenosis is unknown but may be as high as 50 per cent following trauma. Function of the splenic tissue can be assessed using red cell morphology, tests of phagocytic function, the presence of Howell–Jolly bodies, assays of IgM, IgG, and IgA, and scanning. It is clear that the presence of splenosis does not eliminate the risk of developing overwhelming postsplenectomy infection, as there have been documented reports of deaths from this cause occurring in patients in whom splenosis has been demonstrated at autopsy. It has been estimated that 25 to 30 g of tissue is needed for protection against overwhelming postsplenectomy infection.

In an attempt to conserve functioning splenic tissue in cases in which splenunculi are not present and where splenorrhaphy is not possible, splenic tissue can be autotransplanted at the time of surgery by dicing the spleen into small pieces and implanting them into the abdominal wall or an omental pouch. Both techniques can be shown to result in functioning splenic tissue and it is clear from reports that such transplants can be shown to survive, but experimental work has shown a variable potential for decreasing the incidence of infection in asplenic animal models. Further evidence suggests that the mere presence of functioning splenic tissue in itself is insufficient to protect against overwhelming postsplenectomy infection (whether from splenosis or as a result of autotransplantation), and it seems likely that the normal splenic vasculature is crucial for maximum protection. Overwhelming postsplenectomy infection affects only 2.5 per cent of patients after splenectomy for trauma and clinical reports of ‘successful' transplants are anecdotal, so it is unclear whether it should be routine practice, but as it is not associated with significant morbidity there seems little to be lost by performing it.

Further reading

Bickerstaff KI, Morris PJ. Splenectomy for massive splenomegaly. *British Journal of Surgery* 1987; **74**: 346–9. [Outcome review of 47 patients.]

Cooper MJ, Williamson RCN. Splenectomy: indications, hazards and alternatives. *British Journal of Surgery* 1984; **71**: 173–80. [Excellent review article.]

Foss Abramsen A *et al.* Systemic pneumococcal disease after staging splenectomy for Hodgkin's disease 1969–1980 without pneumococcal vaccine protection: a follow up study 1994. *European Journal of Haematology* 1997; **58**: 73–7. [Demonstrates 20-fold increased risk compared with the general population.]

Ghosh S, Symes JM, Walsh TH. Splenic repair for trauma. *British Journal of Surgery* 1988; **75**: 1139–40. [Thirteen cases of repair from a single centre.]

Holdsworth RJ, Irving AD, Cushchieri A. Postsplenectomy sepsis and its mortality rate: actual versus perceived risks. *British Journal of Surgery* 1991; **78**: 1031–8. [Literature review of postsplenectomy sepsis 1952–1987.]

Katkhouda N *et al.* Laparoscopic splenectomy: outcome and efficacy in 103 consecutive patients. *Annals of Surgery* 1998; **228**: 568–78. [Good account of operative technique.]

King H, Schumaacker HB. Splenic studies. I. Susceptibility to infection after splenectomy performed in infancy. *Annals of Surgery* 1952; **136**: 239–42. [Historical interest.]

Kram HB, del Junco T, Clark SR, Ocampo HP, Shoemaker WC. Techniques of splenic preservation using fibrin glue. *Journal of Trauma* 1990; **30**: 97–101. [Good account of use of glue.]

Malangoni MA, Levine AW, Droege EA, Aprahamian C, Condon RE. Management of injury to the spleen in adults. *Annals of Surgery* 1984; **200**: 702–5. [Review of 77 patients.]

Morgenstern L, Shapiro SJ. Techniques of splenic conservation. *Archives of Surgery* 1979; **114**: 449–54. [Good review of available techniques.]

Morris DH, Bullock FD. The importance of the spleen in resistance to infection. *Annals of Surgery* 1919; **70**: 513–21. [First experimental demonstration of overwhelming postsplenectomy infection.]

Morris PJ, Cooper IA, Madigan JP. Splenectomy for haematological cytopenias in patients with malignant lymphomas. *Lancet* 1975; **ii**: 250–3. [Review of 86 patients.]

Mucha P, Daly RC, Farnell MB. Selective management of blunt splenic trauma. *Journal of Trauma* 1986; **26**: 970–9. [Good discussion.]

Redmond HP, Redmond JM, Rooney BP, Duignan JP, Bouchier-Hayes DJ. Surgical anatomy of the human spleen. *British Journal of Surgery* 1989; **76**: 198–201. [A useful article when splenic salvage is to be considered.]

Rescoria FJ, Breitfeld PP, West KW, Williams D, Engum SA, Grosfeld JL. A case controlled comparison of open and laparoscopic splenectomy in children. *Surgery* 1998; **124**: 670–5. [Non-randomized comparison and good discussion.]

The Splenic Injury Study Group. Splenic injury: a prospective multicentre study on non-operative and operative treatment. *British Journal of Surgery* 1987; **74**: 310–13. [Multicentre study of splenic salvage.]

43.2 Lymph node diagnosis

Armando E. Giuliano and Pond R. Kelemen

Further reading

The most common indication for lymph node biopsy is to determine the etiology of lymphadenopathy or as a staging technique for malignancy. While the rate of non-diagnostic biopsies or reactive hyperplasia is approximately 40 to 55 per cent, of these cases 25 to 60 per cent will have disease revealed within a few months. The results of diagnostic lymph node biopsies depend upon the patient population. In Western countries just over one-half of patients with enlarged lymph nodes will have metastatic tumors. Meanwhile in a review of 2194 lymph node specimens from Zimbabwe, only 28.4 per cent showed metastatic disease, lymphoma, or Kaposi's sarcoma. In children, granulomatous lymphadenitis is most often found, and malignancy is seen in less than 20 per cent of excisional biopsies. A wide variety of disorders can result in lymphadenopathy (see Table 1).

I	Neoplastic	
	• Solid tumors	Melanoma, breast, lung, head, and neck, etc.
	• Hematologic	Lymphoma, leukemia, myeloproliferative disorders, and histiocytoses
II	Inflammatory	
	• Infectious	Bacterial, viral, mycobacterial, and fungal
	• Autoimmune	Rheumatoid arthritis, lupus, and sarcoidosis
	• Miscellaneous	Angiofollicular hyperplasia, dermatopathic lymphadenitis, and immunoblastic adenopathy
III	Lipid storage diseases	
	• Gaucher's and Niemann-Pick disease	

Table 1 Diseases associated with lymphadenopathy

For patients with human immunodeficiency viral (**HIV**) infection, an enlarging or tender lymph node is an indication for a lymph node biopsy. In fact, before the advent of serologic testing, open lymph node biopsy was commonly used to diagnose HIV infection. A biopsy by fine needle aspiration can detect most abnormalities, with excisional biopsy reserved for equivocal cases. Infection with HIV results in three types of histologic patterns that correlate with overall prognosis. In addition, lymph nodes may show some of the common complications of HIV infection, which are found in about 10 per cent of cases. These include infections with mycobacterium and HIV-associated malignancies such as B-cell lymphoma and Kaposi's sarcoma.

Benign lymph node enlargement can accompany systemic bacterial or viral infections. Patients with sepsis frequently have associated adenopathy. Mesenteric adenitis in children with viral infections can present as an acute appendicitis. Primary infections of nodal tissue can also occur and provide both diagnostic and therapeutic challenges. Tuberculous infection of lymph nodes is the most common manifestation of extrapulmonary disease. Cervical and axillary node enlargement is frequently found and shows caseating granulomas and acid-fast bacilli on histologic examination. Acutely inflamed suppurative nodes frequently require excision or drainage and are commonly found in pediatric patients. While malignancy is rarely the cause of such a presentation, it does need to be excluded in cases with no resolution after adequate therapy. A wide variety of agents can present with suppurative regional lymphadentitis. Purulent nodal drainage may be found complicating staphylococcal and streptococcal superficial skin infections, but is also associated with lymphogranuloma venereum, cat-scratch disease, and tularemia.

The most important information when assessing a patient with enlarged lymph nodes is the associated symptoms and signs. In children, benign lymphadenopathy from reactive hyperplasia is common. Viral upper respiratory infections can cause lymph node enlargement, but the constitutional symptoms of fever, malaise, and weight loss may also accompany lymphoma. A long history of lymphadenopathy is more likely to be malignant than an acute development of a tender node. The location of adenopathy can be a clue to the diagnosis. Enlarged posterior auricular, supraclavicular, epitrochlear, popliteal, mediastinal, or abdominal lymph nodes are considered abnormal. However, patients with wounds or infections on an extremity will commonly have palpable epitrochlear, axillary, or inguinal lymph nodes. Likewise, cervical adenopathy may be seen in patients with dental disease. Lymph nodes with metastatic carcinoma tend to be firm, matted, and sometimes fixed, while lymphomas tend to be rubbery, discrete, and symmetrical.

Most conditions from enlarged lymph nodes can be adequately diagnosed by fine needle aspiration, with the exception of hematopoietic malignancies that require architectural detail to type them properly. The success of fine needle aspiration depends on carefully explaining to the patients the procedure so as to alleviate anxiety. In addition, the best results occur when the clinician doing the aspiration is the same person who reads the cytology. The procedure consists of cleaning the skin with an alcohol swab and securing the node in question with the non-dominant hand. Without applying suction, a 23- to 25-gauge needle is percutaneously inserted into the lymph node. Continuous suction is then applied as multiple passes are made through the substance of the node without exiting the skin. Suction is then stopped and the needle is withdrawn from the skin. The contents of the needle are blown onto a clean slide and smeared onto several other slides. These slides are then divided into two groups; one for air drying and staining with Diff-Quik® and the other for fixing in alcohol and Papanicolaou staining.

For equivocal results of fine needle aspiration in suspicious nodes or when hematopoietic malignancy is suspected, excisional lymph node biopsy is appropriate. It has been shown for squamous cell carcinoma and melanoma that open biopsy of metastatic lymph nodes before therapeutic dissection does not adversely affect survival. Most open biopsies can be performed with local anesthesia in an outpatient setting. Incision lines should be placed along Langer's lines and be oriented so that they can be incorporated into a lymph node dissection incision if that should become necessary. Sometimes an incisional biopsy is best when a biopsy is needed from a group of matted or fixed nodes. Special care must be made when removing cervical lymph nodes. Injury to the spinal accessory nerve can occur in as many as 10 per cent of cases and results in significant morbidity. Nodal tissue is best delivered to the pathology department fresh. In this way samples can be obtained for culture, touch preps, flow cytometry, molecular diagnosis, and cell culture as well as routine processing with hematoxylin and eosin staining.

In addition to the evaluation of clinically abnormal lymph nodes, open biopsy is being used to detect clinically occult metastatic disease. It has been shown in patients with melanoma and breast cancer that while elective lymph node dissection provides valuable prognostic information based on nodal tumor status, there is no clear evidence for improved overall survival. In addition, many systemic therapies are based upon the nodal status. Complete node dissections are morbid procedures with many short- and long-term sequelae. In more than half of melanomas and breast cancers, the lymph node basins will be without nodal metastases and therefore without any conceivable benefit from a complete node dissection. Due to the complications associated with lymph node dissections combined with the lack of evidence showing improved survival, many clinicians forego lymph node dissections and base systemic therapy upon primary tumor characteristics.

In 1992 Morton *et al.* reported on the use of intraoperative lymphatic mapping and sentinel lymph node dissection in melanoma. This procedure is based on the hypothesis that the sentinel node is the first lymph node draining the primary tumor and that the lymphatic drainage of the tumor is an orderly process. By mapping the lymphatic drainage of a tumor and removing the first lymph node encountered in that drainage the tumor status of the basin can be predicted. In 1994 Giuliano *et al.* published its use in breast cancer with breast-conserving surgery. For melanoma and breast cancer, its use is increasing and is currently being studied in large multicenter trials. In fact, this technique can been used in all solid neoplasms that primarily spread via the lymphatic system. It has been reported for head and neck, Merkel cell, thyroid, and colon cancers. Other tumors where this procedure may prove useful are lung, gastric, and pancreatic cancers. While sarcomas generally are not considered to spread via the lymphatic system, with overall nodal metastases rates of less than 5 per cent, some histologic variants such as angiosarcoma, epitheliod sarcoma, embryonal rhadomyosarcoma, synovial, and alveolar soft part sarcoma have nodal tumor metastases in more than 10 per cent of cases and may benefit from sentinel lymphadenectomy.

Lymphatic mapping commonly involves the use of a vital dye such as isosulfan blue and a radioisotope such as sulfur colloid. Patients may first undergo a preoperative lymphoscintigraphy to determine which basin the lymphatic vessels drain to (Fig. 1). This is particularly important in trunk melanomas and medial breast tumors which have ambiguous drainage patterns. Between 4 and 24 h later the patient is brought to the operating theatre and after induction of anesthesia, which is usually a general anesthetic, 1 to 5 ml of 1 per cent isosulfan blue is injected. The injection is one of the most critical features of the procedure. In melanoma the dye is injected into the surrounding dermis of the previous biopsy site. In breast cancer the dye is injected either into the periphery of the tumor or into the wall of the biopsy cavity. The nodal basin is explored 3 to 8 min later for the sentinel node. A limited incision is then made in the area with the greatest counts as determined by a hand-held g-ray probe. Using careful blunt dissection a blue lymphatic vessel can be identified coursing to a blue lymph node (Fig. 2) that is also ‘hot'. After removal of the sentinel node, the

probe can help identify any other nodes that are ‘hot'. If desired by the clinician, a frozen section can be obtained at that time to determine if a complete node dissection is necessary.

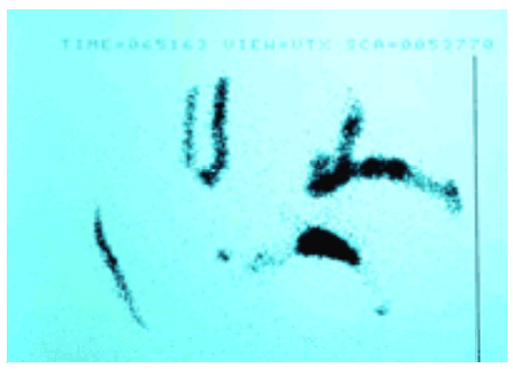


Fig. 1. Lymphoscintigram performed on right breast showing bright spot at initial injection site near primary tumor. A lymphatic vessel is seen coursing to axilla and ending in a sentinel node.



Fig. 2. Intraoperative photograph of left axilla in a patient with breast cancer showing blue lymphatic vessel heading into a sentinel node.

Sentinel lymphadenectomy for melanoma and breast cancer has been reported by many authors as having an identification rate of more than 90 per cent and an accuracy rate approaching 100 per cent. In a study of 103 patients with breast cancer who underwent sentinel lymphadenectomy followed by full axillary dissection, only one metastasis was found out of 1087 non-sentinel lymph nodes from patients who had negative sentinel nodes (false-negative rate =0.97 per cent). All non-sentinel nodes had undergone serial sectioning and the lone metastasis was identified only by immunohistochemical stains. This confirmed the sentinel lymph node hypothesis and shows how accurate the procedure can be.

One of the interesting features of sentinel lymphadenectomy is the focused study one can perform on the nodal tissue. In node-negative cases as determined by routine hematoxylin and eosin stains, many authors perform immunohistochemistry on serial sections of lymph node tissue. For melanoma S-100 and HMB 45 antibodies, and for breast cancer a mixture of anticytokeratin stains of low and intermediate molecular weight are used. This allows the detection of occult tumor cells and upstages approximately 10 to 20 per cent of patients. Sentinel nodes are also being studied with tumor-specific markers for mRNA using reverse transcriptase-polymerase chain reaction (RT-PCR). This technique is capable of detecting one cell in the entire node. This can further upstage an additional 20 to 40 per cent of patients. The clinical significance of these micrometastases detected by the most sensitive of tests remains to be determined.

Further reading

Balch CM *et al.* Efficacy of an elective regional lymph node dissection of 1 to 4 mm thick melanomas for patients 60 years of age and younger. *Annals of Surgery* 1996; **224**: 255–66. [Major trial showing no survival advantage in patients with melanoma of elective lymph node dissection except in a subset of patients.]

Burton F, Patete PL, Goodwin, Jr. WJ. Indications for open cervical node biopsy in HIV-positive patients. *Otolaryngology—Head and Neck Surgery* 1992; **107**: 367–9. [Study of indications for lymph node biopsy in patients infected with HIV.]

Fisher B *et al.* Ten year results of a randomized clinical trial comparing radical mastectomy and total mastectomy with or without radiation. *New England Journal of Medicine* 1985; **312**: 674–82. [Major study showing no therapeutic value of axillary lymph node dissection in patients with breast cancer.]

Fong Y, Coit DG, Woodruff JM, Brennan MF. Lymph node metastasis from soft tissue sarcoma in adults; analysis of data from prospective database of 1772 sarcoma patients. *Annals of Surgery* 1993; **217**: 72–7. [Study of the incidence of lymph node metastasis in soft tissue tumors.]

Giuliano AE, Kirgan DM, Guenther JM, Morton DL. Lymphatic mapping and sentinel lymphadenectomy for breast cancer. *Annals of Surgery* 1994; **220**: 391–401. [Review and technical description of sentinel lymph node dissection for breast cancer.]

Lake AM, Oski FA. Peripheral lymphadenopathy in childhood. *American Journal of the Diseases of Childhood* 1978; **132**: 357–9. [Review of experience with lymph node biopsy in children.]

London J, London NJ, Kay SPJ. Iatrogenic accessory nerve injury. *Annals of the Royal College of Surgeons of England* 1996; **78**: 146–50. [Study revealing incidence and morbidity of accessory nerve injury with cervical node biopsy.]

Morton DL *et al.* Technical details of intraoperative lymphatic mapping and sentinel node biopsy in the management of primary melanoma. *Archives of Surgery* 1992; **127**: 392. [First report of sentinel node dissection with excellent description of technique.]

Sibanda EN, Stanczuk G. Lymph node pathology in Zimbabwe: a review of 2194 specimens. *Quarterly Journal of Medicine* 1993; **86**: 811–17. [Review of experience with lymph node biopsy in Zimbabwe.]

Siegel RJ. Surgical pathology of lymph nodes in cancer staging: routine and specialized techniques. *Surgical Oncology Clinics of North America* 1996; **5**: 25–31. [Review of special techniques to identify micrometastatic disease.]

Sriwatanawongsa V, Cardoso R, Chang P. Incidence of malignancy in peripheral lymph node biopsy. *American Surgeon* 1985; **51**: 587–90. [Review of experience with lymph node biopsy in the United States.]

43.3 Surgery in patients with leukaemia

P. Jane Clarke

Introduction
Incidence
Indications
Diagnostic
To allow adequate treatment
Surgery for the complications of the leukaemic process
Complications of treatment
Difficulties in surgical management
A strategy for surgical management
Conclusions
Further reading

Introduction

Improved remission rates in patients with leukaemia mean that elective surgical procedures, which formerly would not have been considered, may now be undertaken just as they would in healthy patients. Thus the major surgical challenge remains the patient with acute leukaemia requiring an emergency operation. Operative mortality rates in patients with acute leukaemia have been prohibitively high (up to 75 per cent), but the option of conservative treatment of an acute surgical problem has been associated with mortality rates approaching 100 per cent. The decision to perform major surgery in patients with acute leukaemia can be very difficult and should be taken only after discussion with other members of the multidisciplinary team.

Incidence

The demand for surgery in these patients is increasing for two reasons. First, as a result of improved chemotherapy and medical management of leukaemia, complications requiring surgery are becoming more frequent. The true incidence is unknown but several studies report a surgically remediable complication rate of 5 to 10 per cent. Second, advances in supportive techniques have facilitated a more aggressive approach in dealing with complications requiring surgery.

Indications

Patients with leukaemia are susceptible to several conditions and complications, an awareness of which will enhance early diagnosis and appropriate management. In addition to these specific prob-lems, patients with leukaemia are susceptible to the same surgical problems as the general population, and these conditions may also demand urgent surgery. Surgery for these unrelated conditions need not be withheld if the patient is in remission and there is no sign of drug-induced bone marrow suppression as there is no evidence that these patients are more prone to complications than the general population. Table 1 classifies the indications for surgery.

Diagnostic procedures
Lymph node biopsy
Soft tissue/lung biopsy
Central nervous system biopsy
Testicular biopsy
Laparotomy
Surgical procedures to facilitate treatment
Insertion of Hickman line/portacath
Insertion of Ommaya reservoir
Surgical procedures to treat complications of the leukaemia process
Splenectomy (spontaneous rupture, infarction, hypersplenism, symptomatic enlargement)
Operations for perianal disease/sepsis
Biliary procedures
Prostatectomy
Fixation of pathological fractures
Surgical procedures to treat complications of treatment
Laparotomy for intestinal perforation/bleeding (necrotizing enterocolitis)
Tracheostomy (pharyngeal haematoma, severe laryngotracheobronchitis)
Treatment of urinary tract calculi

Table 1 A classification of indications for surgery in patients with leukaemia

Diagnostic

Surgery has little to offer in the diagnosis of acute leukaemia. Chronic lymphatic leukaemia is diagnosed by lymph node biopsy, and a surgical biopsy may be necessary to diagnose relapse in the central nervous system, testicle, or spleen.

To allow adequate treatment

The surgeon may be involved in providing venous access for either chemotherapy or antibiotics (either a Hickman line or Portacath device), and these techniques are discussed in Chapter 48.6. Central nervous system involvement, especially in children, involves injections into the intrathecal space for up to a year. Repeated lumbar punctures can be avoided by the use of a subcutaneous reservoir of cerebrospinal fluid described by Ommaya. A small plastic reservoir is installed beneath the scalp and connected to the ventricle of the non-dominant hemisphere through a single occipital or frontal burr hole.

Surgery for the complications of the leukaemic process

Abdominal pain in the patient with leukaemia may require laparotomy; abdominal pain is not uncommon in children with the disease and may be the mode of presentation. The gastrointestinal manifestations of leukaemia are becoming increasingly common, especially in the more aggressive forms of the disease. Leukaemic infiltrates are usually multifocal and can affect any part of the gastrointestinal tract. They vary in size from microscopic to several centimetres in diameter and can form either a plaque-like thickening, or raised nodular or polypoid lesions which can intersussept. Obstruction secondary to the tumour itself can occur but is uncommon, but obstruction may be precipitated by chemotherapy which can cause oedema and fibrosis. Diffuse mucosal and submucosal infiltration with or without ulceration is a source of haemorrhage, and perforation can occur when extensive necrosis is present, which may be increased by treatment and infection. Patients who have had a bone marrow transplant may develop graft-versus-host disease, which can be associated with inflammation of the stomach and small intestine.

The association between anorectal disease and acute leukaemia is well recognized. It manifests as ulcers, abscesses, fissures, and fistulas. Rectal ulcers due to infiltration, with or without associated abscesses, are uncommon, but are said to be distinctive complications of this disease. Anorectal infections may be characterized by rapid local spread of inflammation, and early surgery has been associated with uncontrolled, massive sloughing of the perianal tissues. The safest approach in these patients is to administer high-dose antibiotics when the picture is one of inflammatory induration, while a small well-placed incision should be used when there is an obviously fluctuant abscess. Wide excision of inflamed tissues should be avoided.

Infiltration of organs outside the gastrointestinal tract occurs most commonly in the spleen, testicle, gall bladder, and occasionally, the prostate. Splenectomy may be indicated for several reasons, including painful enlargement, infarction, hypersplenism, or spontaneous rupture. Approximately 13 per cent of boys with leukaemia relapse in the testis, which presents as a painless swelling on one or both sides. The diagnosis can be confirmed by biopsy. The gall bladder may become diffusely infiltrated, and the common bile duct or common hepatic duct may be obstructed by nodes and need some form of stenting or surgical bypass. Prostatectomy may be necessary to relieve bladder outlet obstruction, but should be postponed until the patient is in remission. These complications most commonly occur in patients with one of the chronic leukaemias.

Complications of treatment

The development of abdominal pain during the pancytopenic period after chemotherapy is well recognized. In addition to coincidental causes, a specific syndrome of

necrotizing enterocolitis has been recognized and is estimated to have an incidence of about 5 per cent. Other terms which have been used in the past are typhlitis and ileocaecal syndrome. These patients have two common features. First, they have usually been treated with cytosine arabinoside and an anthrocycline (the standard treatment for AML), and second, they demonstrate necrotizing lesions affecting the ileum, appendix, and right colon. The patients present with nausea, diarrhoea, and peritoneal signs 1 or 2 weeks after chemotherapy. Plain abdominal radiology may either be non-specific or show signs of ileus, and at laparotomy there is non-specific haemorrhagic or gangrenous bowel with oedema. The pathogenesis is unclear but is thought to be a primary gut necrosis with secondary infection. Why the right colon is most frequently affected is unknown. Survival from this condition depends on well planned surgery with recovery of bone marrow function, with antibiotic, platelet, and granulocyte support.

Other complications of treatment which may involve the surgeon are the abdominal pain and constipation associated with vincrstine, and pancreatitis is a rare complication of asparginase treatment. In addition, acute peptic ulceration may occur in any patient treated with corticosteroids.

Difficulties in surgical management

The patient with acute leukaemia presents the surgeon with several problems. Poor quality white blood cells and immunosuppression, induced by chemotherapy and steroids, predispose to infection and abscess formation. Thrombocytopenia, abnormal platelet function, and clotting disorders increase the risks of perioperative haemorrhage.

Symptoms and signs of serious pathology, for example peritonitis, are frequently mild or non-specific because of the inability of these patients to mount an adequate host response to infection. The host response is further suppressed by steroid therapy. This suppression may be reflected by an inappropriately low-grade temperature, white cell count, and levels of inflammatory markers such as C-reactive protein or erythrocyte sedimentation rate. In addition, signs suggestive of an acute abdomen, such as abdominal pain, nausea, vomiting, and even malaena are common in patients with leukaemia and are frequent symptoms of toxic chemotherapy. Careful consideration by the multidisciplinary team and the use of CT and MRI have increased the accuracy of preoperative diagnosis and reduced the occurrence of unnecessary laparotomy for a possible acute abdomen.

A strategy for surgical management

A high index of suspicion is required when assessing these patients because of their minimal signs and symptoms. CT and MRI should be readily used when the diagnosis is in doubt. Surgery must not be undertaken without the availability of the necessary blood product support. Emergency surgery can usually be postponed overnight to ensure a full blood transfusion and haematological laboratory service. Emergency surgery, even for a relatively simple procedure such as appendicectomy, should be performed by an experienced surgeon and not delegated to a junior. Aggressive antibiotic cover is used and the clinical microbiologists are notified and involved in management at an early stage. When platelet support is anticipated, it should be started after induction of anaesthesia and continued for several hours after surgery. Surgical incisions must be planned to give generous exposure, especially when there is any diagnostic doubt. Muscle-cutting incisions should be avoided if possible. Long midline incisions are generally suitable. Meticulous haemostasis must be obtained during surgery, and when bleeding has been a problem, suction drains are advisable to facilitate postoperative monitoring. Postoperative management should ideally be in an intensive care or high dependency unit where accurate monitoring is possible. A high index of suspicion for haemorrhagic complications is essential and a policy of early relook laparotomy, before the onset of coagulopathies, is prudent.

Conclusions

In the patient with leukaemia, problems with preoperative diagnosis are common and early imaging is advised. An awareness of the conditions and complications to which these patients are predisposed will aid diagnosis. With platelet support, meticulous surgical technique, and careful postoperative management, major surgery is usually feasible, and the decision to operate on these patients will be based increasingly on surgical indications.

Further reading

Bishop JF, Schiffer CA, Aisner J, Matthews JP, Wiernik PH. Surgery in acute leukemia: a review of 167 operations in thrombocytopenic patients. *American Journal of Hematology* 1987; **26**: 147–55.

Dewar GJ, Lim CNH, Michalyshyn. Gastrointestinal complications in patients with chronic and acute leukemia. *Canadian Journal of Surgery* 1981; **24**: 67–71.

Kies MS *et al.* Neutropenic enterocolitis. *Cancer* 1979; **43**: 730–4.

Rasmussen BL, Freeman JS. Major surgery in leukemia. *American Journal of Surgery* 1975; **130**: 647–51.

Sedhev MK, Dowling MD, Seal SH, Stearns MW. Perianal and anorectal complications in leukemia. *Cancer* 1973; **31**: 149–52.

Spiers ASD. Surgery in management of patients with leukaemia. *British Medical Journal* 1973; **3**: 528–32.

44.1 General introduction to neurosurgery

C. B. T. Adams

Introduction
Diagnosis
Structural and functional diagnosis
Anatomical diagnosis
Pathological diagnosis
Raised intracranial pressure
Aetiology
Pathology
Features of raised intracranial pressure
Treatment
Meningeal irritation
Subarachnoid haemorrhage
Meningitis
Carcinomatous ‘meningitis'
Chemical meningitis
Investigations
Plain radiography of the skull
Lumbar puncture
CT scanning
Angiography
Magnetic resonance imaging
Electroencephalography
Further reading

Introduction

The delicate function of the central nervous system, the need for minimal brain retraction, and therefore for exact operative exposure, as well as the overwhelming requirement for good surgical and neurological judgement, have determined the development of the neurosurgery. Despite the advent of modern methods of investigation, a careful history and examination are still paramount in the assessment of patients.

The response of the central nervous system to surgery is essentially the same as it is to other parts of the body and the neurosurgeon must apply the basic principles of surgical technique: careful handling of tissue, minimal tissue necrosis, elimination of dead space, and avoidance of haematoma formation by careful haemostasis. Like surgery elsewhere, neurosurgery is concerned with removing lumps, cysts, and diverticula; with relieving obstruction, eliminating sinuses and fistulas; with stopping haemorrhage, removing pus, and improving or altering function. Knowledge of these basic surgical disorders can be applied to all areas of surgery including neurosurgery.

Diagnosis

The diagnosis of intracranial disorders is made in three distinct phases (Table 1).

1. STRUCTURAL AND FUNCTIONAL DIAGNOSIS. —from the history and examination
2. ANATOMICAL DIAGNOSIS. —i.e., evidence of:
Focal neurological lesion (or lesions)
Raised intracranial pressure
Meningeal irritation
3. PATHOLOGICAL DIAGNOSIS
(a) Possibilities:
Congenital
Traumatic
Inflammatory
Neoplastic
Vascular
Degenerative
Metabolic
Toxic
(b) Factors to consider:
The anatomical diagnosis
The number of focal lesions
The age and sex of the patient
The speed of onset and consequent course of the symptoms and signs
Other factors—history, elsewhere

Table 1 The diagnosis of intracranial disorders

Structural and functional diagnosis

These are made from the history and examination. When dealing with patients who have disorders of intellect or speech, relatives must be interviewed. The history and examination will indicate whether the patient's weakness, for instance, is upper or lower motor neurone in origin or due to a muscle disorder, and whether disorders such as dysphasia or dysarthria are present. Problems that do not have an anatomical or pathological basis, such as migraine or idiopathic epilepsy, are best classified as structural or functional. One must also remember to examine the head for a lump, which might indicate an underlying meningioma. Surgeons should be aware of a midline ‘sebaceous cyst'; this may be a dermoid cyst tracking into the posterior fossa. The contents of the cyst may create a fistula in the occipital or lumbar region, and cause meningitis.

Anatomical diagnosis

Intracranial disorders present with either one or more of raised intracranial pressure, meningeal irritation, and focal neurological signs. The absence of these features in the history or examination usually indicates the absence of serious disease. However, caution should be exercised when examining an individual who is receiving high doses of steroids, which may mask raised intracranial pressure.

One decides which of these anatomical disorders is present on the basis of the disorders of structure and function. Diagnosis of focal neurological signs does not usually require a profound knowledge of neuroanatomy (Fig. 1). It is not difficult to decide whether the focal signs arise from a focal lesion, multiple lesions, or a diffuse lesion.

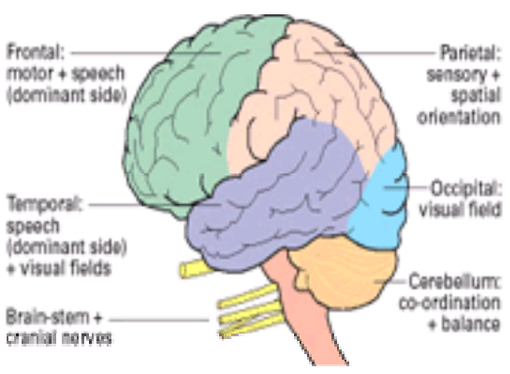


Fig. 1. Diagram to show the basic functions of the major subdivisions of the brain.

Pathological diagnosis

Only when the other diagnostic groups have been considered should one consider the pathological diagnosis. The possibilities are congenital, traumatic, inflammatory, neoplastic, vascular, degenerative, metabolic, and toxic. Factors that should be considered in reaching this diagnosis include structural and functional diagnosis, anatomical diagnosis, age and sex of the patient (middle-aged women are prone to meningiomas; children to cerebellar astrocytomas or medulloblastomas), speed of onset of the disorder (sudden or progressive), and the subsequent course of the illness (progressive almost always suggests an enlarging mass; fluctuating suggests a cystic lesion). Multiple lesions are suggestive of metastases, cerebral lymphomas, or multiple sclerosis. The presence of disorders elsewhere, such as primary carcinoma or neurofibromatosis, should also be considered.

Raised intracranial pressure

Aetiology

The normal intracranial pressure measured at lumbar puncture is less than 200 mm of water. The causes of raised intracranial pressure are illustrated in Fig. 2.

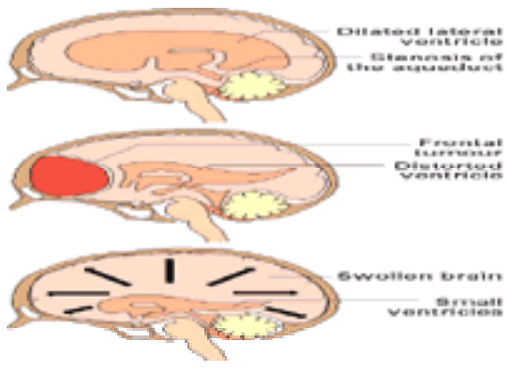


Fig. 2. Diagram illustrating the causes of raised intracranial pressure.

Accumulation of cerebrospinal fluid (hydrocephalus)

This is caused by obstruction to the normal flow of cerebrospinal fluid either in the ventricular system (obstructive hydrocephalus) or in the subarachnoid space (communicating hydrocephalus) secondary to any persistent blood or pus in that space. Occasionally, excessive production of cerebrospinal fluid may be responsible, as is seen in patients with a papilloma of the choroid plexus (Fig. 3).

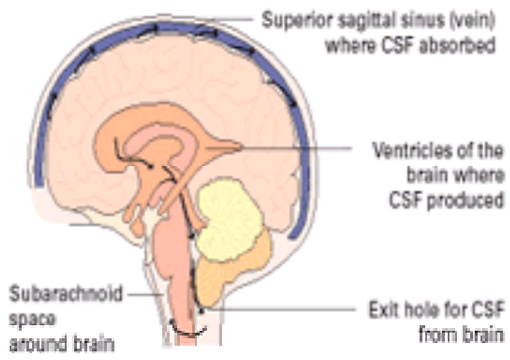


Fig. 3. Diagram to illustrate the production, circulation, and absorption of cerebrospinal fluid and the likely sites of interference of flow.

Increased bulk

Any space-occupying lesion, tumour, abscess, cyst, or blood clot will raise intracranial pressure by reason of its own bulk or by causing a blockage of the cerebrospinal fluid pathways. Benign intracranial hypertension due to cerebral swelling has an unknown cause, but is usually seen in overweight young women, and may cause severe papilloedema. This syndrome may also be caused by obstruction to venous drainage of the brain.

Increased volume of intracranial blood

Accumulation of carbon dioxide due to respiratory insufficiency causes vasodilation of cerebral blood vessels and raised intracranial pressure—hence the importance of maintaining the airway in head-injured patients. Obstruction of the large venous sinuses will cause impaired cerebral venous drainage and brain swelling. The ventricular size will be normal, unlike that found in hydrocephalus. Benign intracranial hypertension is characterized by severe papilloedema with retinal haemorrhages in a patient who is otherwise well, apart from possibly showing a VIth nerve palsy and an extensor plantar response. Benign is a misnomer: this condition may cause blindness unless treated early.

Relative diminution in the size of the skull

Craniosynostosis is characterized by premature fusion of the cranial sutures, which prevents normal enlargement of the skull *pari passu* with brain rowth. The treatment of this condition is considered elsewhere (Chapter 50.2).

Patients not infrequently present with raised intracranial pressure without focal signs. Hydrocephalus due to an intraventricular mass such as a colloid cyst of the third ventricle or a tumour of the fourth ventricle should be considered. A tumour in the silent part of the brain (the non-dominant frontal lobe or in the region of the tentorium) is also a possibility.

Multiple tumours, typically metastases surrounded by oedema, may cause raised pressure but little in the way of focal signs. Bilateral subdural haematomas can present with raised pressure without focal signs. Finally, benign intracranial hypertension due to venous sinus obstruction or cerebral oedema should be considered if there is no evidence of a mass or hydrocephalus, though a diffusely infiltrating astrocytoma can mimic this condition quite closely.

Pathology

The most important effect of raised intracranial pressure is the distortion and shift of the brainstem (Fig. 4). Supratentorial lesions force the ipsilateral temporal lobe through the tentorial notch, applying pressure on the horizontally directed, ipsilateral IIIrd cranial nerve. This causes a dilated pupil, and the pressure on the midbrain

causes drowsiness. Eventually the contralateral cerebral peduncle is pressed against the opposite margin of the tentorial notch, causing pyramidal signs on the side of the lesion and eventually damage to the contralateral IIIrd cranial nerve and bilateral fixed, dilated pupils. Severe distortion of the brainstem causes decerebrate posturing, characterized by extension of the limbs, hyperpronation of the hands and wrists, and hyperventilation in response to any stimulus. The delicate perforating arteries entering the brainstem become stretched and eventually rupture, causing irreversible damage and death.

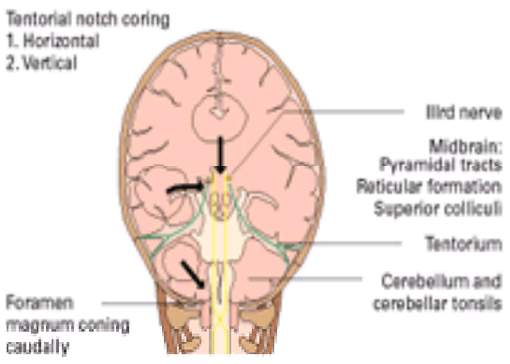


Fig. 4. Diagram to show brainstem shift from supratentorial-occupying lesions or posterior fossa tumours; these cause tentorial or foramen magnum coning, respectively. Compare with Fig. 5 and Fig. 6.

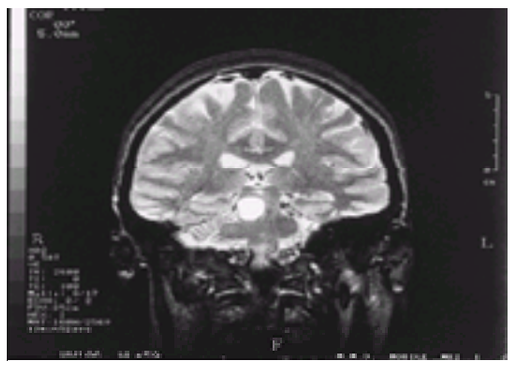


Fig. 5. Coronal section, magnetic resonance scan to show the tentorium separating the cerebral hemispheres above from the cerebellum below (the folia can be well seen). The brainstem contains a benign cyst (shown as white) just at the level of the tentorial notch. On either side of the midbrain, posterior cerebral arteries can be seen (as a black dot). This scan shows how herniation of the temporal lobe at the tentorial notch may obstruct the posterior cerebral artery and cause occipital infarction. The cyst was evacuated by elevating the temporal lobe, identifying the posterior cerebral artery, and incising the midbrain adjacent to this vessel.

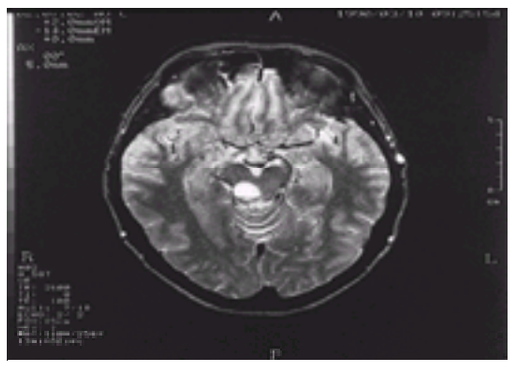


Fig. 6. Horizontal section, magnetic resonance scan of the same patient as Fig. 5 to show the posterior cerebral arteries (black) arising from the basilar artery, encircling the midbrain. The folia of the vermis of the cerebellum can be seen posteriorly. The internal carotid arteries and branches (black) can be seen well. The cyst is clearly visible. The cerebrospinal fluid is white.

The posterior cerebral artery may also become compressed as it passes around the midbrain by the ipsilateral temporal lobe, distorting the artery against the tentorium (Fig. 5 and Fig. 6). This produces infarction of the occipital lobe and a homonymous hemianopia, which is permanent if the patient recovers.

Bilateral lesions, such as bilateral chronic subdural haematomas, force both temporal lobes through the tentorial notch, compressing the brainstem and pushing it downwards. This causes bilateral ptosis and limitation of upward gaze due to compression of the superior colliculi as well as, eventually, dilated and fixed pupils. More slowly progressive lesions cause bilateral VIth nerve palsies as these vertically directed nerves become stretched across the petrous bone when the brainstem descends.

Lumps in the posterior fossa cause ‘tonsillar coning' when the cerebellar tonsils are forced through the foramen magnum, thus squeezing the medulla oblongata. This produces neck stiffness with slowing, irregularity, and finally cessation of respiration. Breathing may stop very suddenly, although the blood pressure is maintained. The correct treatment is an emergency burr hole and ventricular drainage to relieve the hydrocephalus.

Any sudden rise in intracranial pressure causes reflex slowing of the pulse and elevation of blood pressure (Cushing reflex). Patients with long-standing hydrocephalus (i.e., aqueduct stenosis) show gross dilatation of the lateral ventricles. This causes compression on the skull vault, which becomes thin, particularly over the cerebral gyri (Fig. 7). The infundibular recess of the third ventricle becomes dilated and herniates into the pituitary fossa. Hypopituitarism may develop, with stunted growth and incomplete sexual development.

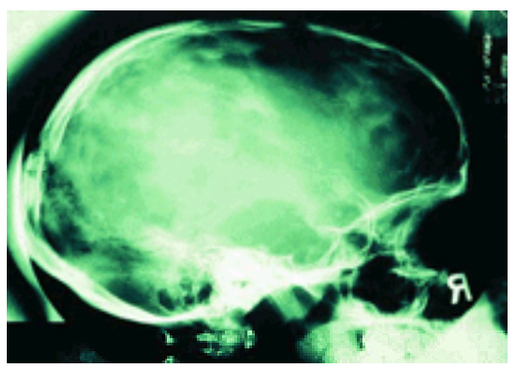


Fig. 7. Lateral skull radiograph showing ‘copper beating' of the vault of the skull due to thinning by distended adjacent cortex. The dorsum sella is eroded and flattened. The pituitary fossa is misshapen due to herniation of the distended third ventricle into the pituitary fossa. This patient had aqueduct stenosis and consequently a shallow posterior fossa.

Features of raised intracranial pressure

The most important symptom of raised intracranial pressure is headache that is worse in the morning and may wake the patient up at night. The site and quality of the headache are less helpful, but any severe, unusual headache should be taken seriously. The headache is often associated with vomiting and the diagnosis of migraine is made all too often. ‘Acute migraine'—that is, headache of only a few months' duration—is an extremely dangerous diagnosis; it is usually associated with raised intracranial pressure. Vomiting unassociated with or preceding the headache may be due to a lesion (such as an ependymoma) of the floor of the fourth ventricle and may well defy gastrointestinal diagnostic efforts.

The most important sign of raised intracranial pressure is drowsiness; this is much more important than papilloedema, though this, when present, is of great importance. However, patients may lapse into coma and die from raised intracranial pressure before papilloedema has time to develop (Fig. 8): its absence does not exclude the diagnosis. Severe and sudden raised intracranial pressure may cause retinal haemorrhages, which may rupture into the subhyloid space, producing a subhyloid haematoma. The Glasgow coma scale (see Table 3 in Chapter 44.2) a well-established clinical method of observing and recording the depth of coma. This scale should be used rather than vague terms such as stupor, semicoma, and coma.



Fig. 8. Papilloedema due to raised intracranial pressure. However, the most important sign of raised intracranial pressure is drowsiness.

Brainstem distortion causes slowing of the pulse, elevation of the blood pressure, and so-called false localizing signs of tentorial herniation or foramen magnum coning, depending on the site of the mass.

Infants show vomiting and drowsiness, as well as an increase in circumference of the head, which has a ‘crackpot' sound when percussed, due to spreading of the sutures (Fig. 9). Bulging of the anterior fontanelle is an important sign. Papilloedema is less common, and although retinal haemorrhages may occur, they usually indicate a subdural haematoma in this age group.

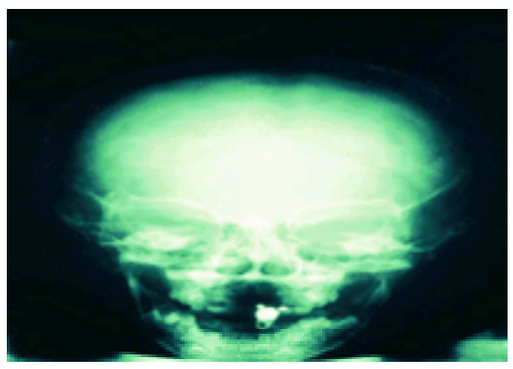


Fig. 9. Radiograph of the skull to show separation of the skull sutures due to raised intracranial pressure in childhood.

Treatment

This is directed to the cause. Any space-occupying lesion is removed if possible; obstructions to cerebrospinal fluid circulation are removed or bypassed. Any sudden rise in pressure may be countered by high-dose steroids, or in a more acute situation, 20 per cent mannitol intravenously. Of course it is essential to maintain the airway at all times, particularly when the patient is drowsy.

Meningeal irritation

Meningeal irritation is a clinical entity similar to peritoneal or pleural irritation, with the same causes, including blood, pus, cancer cells, or chemicals. The features of meningeal irritation are headache, which may be associated with neck ache, backache, and pain extending down the backs of the legs. These pains reflect the distribution of blood or pus within the subarachnoid space. Vomiting is common, and photophobia is more severe than one would expect with the degree of headache. The signs are those of limitation of neck flexion (also caused by foramen magnum ‘coning' from a mass in the posterior fossa and, of course, by something wrong with the neck), together with limitation of straight-leg raising.

Limited straight-leg raising is also a feature of any lump pressing on a lumbar or sacral nerve root, usually a prolapsed lumbar disc. The limitation is then usually unilateral. A fever may occur with meningeal irritation of any cause. A lumbar puncture and examination of the cerebrospinal fluid is necessary to elucidate the cause.

Subarachnoid haemorrhage

This is one diagnosis that can be made over the telephone. The headache has two characteristic features: it is sudden, and unusual. It is important to be alert to this diagnosis, since these patients require urgent hospital admission even if the symptoms are mild or improving (although the headache usually lasts 24 h or more).

A lumbar puncture is necessary to establish the diagnosis. Computed tomographic (**CT**) scanning will show subarachnoid blood but will not exclude a minor haemorrhage: if a CT scan is normal then a lumbar puncture is mandatory. Cerebrospinal fluid must be spun down: xanthochromia of the supernatant fluid demonstrates that blood has been present in the cerebrospinal fluid for at least a few hours before the lumbar puncture, differentiating it from bloodstained fluid produced immediately by a traumatic puncture. Subarachnoid haemorrhage needs full investigation by carotid and vertebral angiography. The most common cause is a ruptured aneurysm; ruptured arteriovenous malformation, a primary cerebral haemorrhage, and bleeding tumours are other possibilities. Sometimes no cause is found, even after full investigation.

Meningitis

This diagnosis indicated by an increased white-cell count in the cerebrospinal fluid in association with meningeal irritation. It is of surgical importance as it may occur after head injuries, particularly a fracture involving the paranasal air sinuses. This cause should be considered with any case of pneumococcal meningitis, even if it occurs some years after a head injury. Such meningitis may cause coma within a few hours. Lumbar puncture is a better method of investigating meningeal irritation than is CT scanning, which will never diagnose meningitis. Meningitis may also be associated with a dermoid fistula (often marked by a lump or dimple over the vertex of the skull) connecting the scalp with the posterior fossa. A skull radiograph shows a bony groove and channel in the region of the torcular; such an examination

should be undertaken before excision of a midline ‘sebaceous cyst' from the back of the head.

Carcinomatous ‘meningitis'

This occurs when a secondary deposit adjacent to the ventricle seeds into the cerebrospinal fluid. Malignant gliomas, particularly medulloblastomas in children, may also behave in a similar fashion. The patient is ill and often in severe pain due to deposits on the nerve roots, and the cerebrospinal fluid contains malignant cells, large amounts of protein, and little sugar. The diagnosis is best made by magnetic resonance imaging with gadolinium enhancement.

Chemical meningitis

This is usually iatrogenic, and follows introduction of substances into the subarachnoid space daring lumbar puncture. Occasionally a ‘cholesterol' meningitis may occur secondary to leakage of cholesterol from a dermoid or an epidermoid cyst and very occasionally, from a craniopharyngioma. This may follow surgery, and very occasionally occurs spontaneously.

Investigations

Investigations should only be undertaken after careful clinical assessment has produced a likely clinical diagnosis. Talking to the patient may have profound implications for the surgeon and may determine, for instance, a more cautious excision of a tumour to cause less risk of disability in certain domestic circumstances.

Patients with intracranial lesions should routinely undergo chest radiography to exclude carcinoma of the lung, as well as routine blood and urine tests.

Plain radiography of the skull

This is still a useful investigation and may show evidence of raised intracranial pressure by erosion of the dorsum sella or thinning of the skull vault by cerebral convolutions (copper beating). In children and infants, spreading of the sutures and enlargement of the head may be apparent (Fig. 7, Fig. 9). Local bone erosion, bony metastases, hyperostosis due to a meningioma, or expansion of the sella turcica indicating a pituitary tumour can also be seen on the plain radiographs of the skull. Shifting of a calcified pineal gland, vertically or horizontally; is useful information, and calcification may be apparent in tumours. Radiographs of the spine may show bone collapse or erosion of the pedicles of the vertebral bodies by secondary deposits, while scalloping of the posterior surface of a vertebral body or enlargement of an intravertebral foramen usually indicates a benign neoplasm such as neurofibroma.

Lumbar puncture

This is indicated when there is ‘meningeal irritation', but is contraindicated in patients with evidence of raised intracranial pressure, especially if a mass lesion of the posterior fossa is likely. Evidence of cord compression requires a lumbar puncture and myelography, but this investigation is best done in a neurosurgical unit as function may deteriorate thereafter and may require urgent surgery, Lumbar punctures are also done to obtain cerebrospinal fluid for assessment of proteins, for instance in patients with suspected multiple sclerosis.

CT scanning

The introduction of CT scanning has dramatically changed the ease with which neurosurgical investigation cm be performed. It demonstrates intracranial blood and calcium (Fig. 10, Fig. 11), and the ventricles as well as the brain and skull. It also demonstrates most space-occupying lesions (Fig. 12,Fig. 13 and Fig. 14), although an isodense subdural haematoma may be misinterpreted as brain swelling. CT scanning is not as sensitive as is magnetic resonance imaging for showing small or subtle intracerebral lesions.

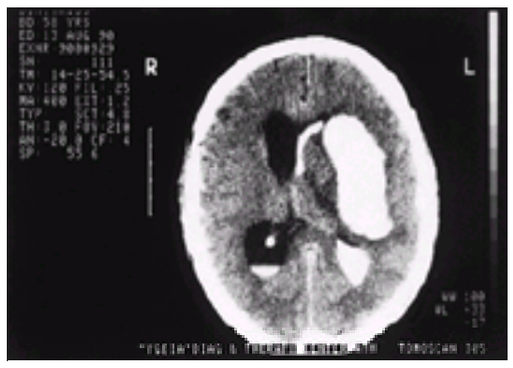


Fig. 10. CT scan showing a large primary intracerebral bleed in a 58-year-old hypertensive man. This haematoma has ruptured into the ventricle and blood can be seen layered in the right occipital horn. There is secondary hydrocephalus due to shift of the midline structures preventing egress of cerebrospinal fluid from the right lateral ventricle.

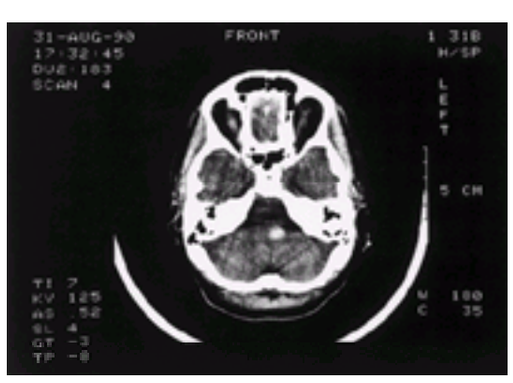


Fig. 11. CT scan without contrast showing a brainstem haematoma in a 32-year-old woman, 30 weeks pregnant.

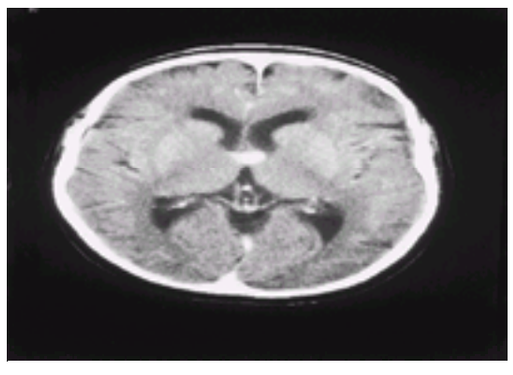


Fig. 12. CT scan showing a typical hyperdense colloid cyst of the third ventricle causing hydrocephalus. The patient presented having had five episodes of unconsciousness due to sudden hydrocephalic attacks from intermittent raised intracranial pressure.

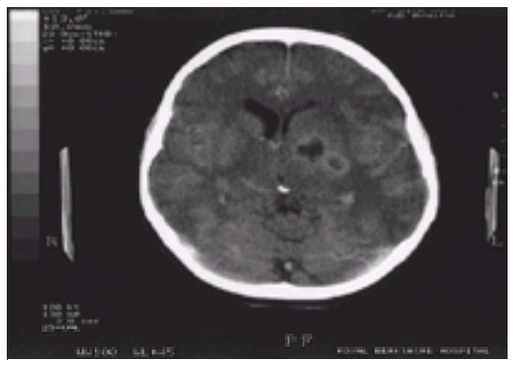


Fig. 13. CT scan with contrast showing an infiltrating malignant glioma; this was biopsied stereotactically.

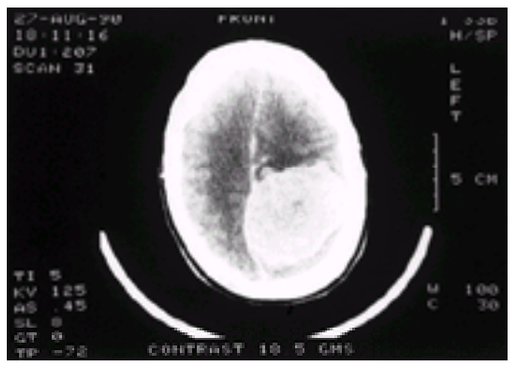


Fig. 14. CT scan with contrast showing a typical (parasagittal) meningioma; Fig. 23 shows the blood supply to this tumour.

Angiography

The main indication for angiography is the investigation of subarachnoid haemorrhage following its confirmation by CT scanning or lumbar puncture. Angiography demonstrates intracranial aneurysms (Fig. 15, Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21), arteriovenous malformations, and the pathological circulation of some tumours which, if very vascular, can be embolized preoperatively (Fig. 22, Fig. 23).

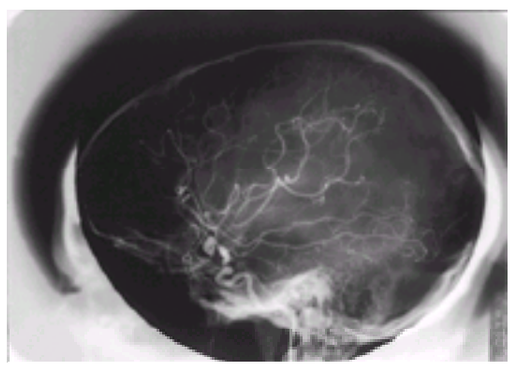


Fig. 15. Right lateral internal carotid angiogram to show an aneurysm of the posterior communicating artery and a bilocular middle cerebral aneurysm (see Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21).

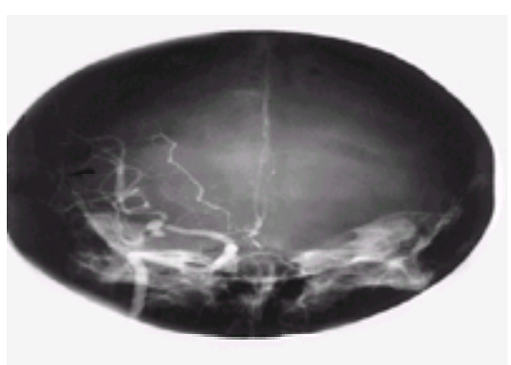


Fig. 16. Right internal carotid angiogram to show the bilocular middle cerebral artery aneurysm (Fig. 15). There is a hypoplastic right anterior cerebral artery which is why the aneurysm on the anterior communicating artery does not fill from the right. The aneurysm on the posterior communicating artery is ‘hidden' behind the internal carotid artery and can only be seen on the lateral view.

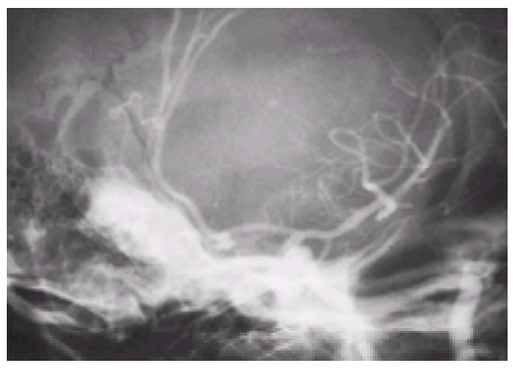


Fig. 17. The left internal carotid angiogram to show an aneurysm of the anterior communicating artery filling just from the left side. This is an oblique view to demonstrate the anterior cerebral artery and its branches. The same patient is shown in Fig. 15, Fig. 16, Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21.

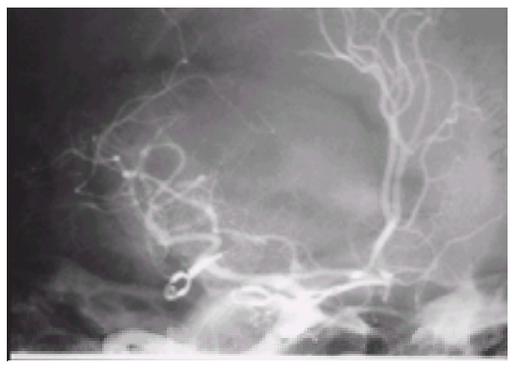


Fig. 18. Postoperative angiogram to show the three aneurysms illustrated in this series of figures have been clipped via a right frontotemporal craniotomy.

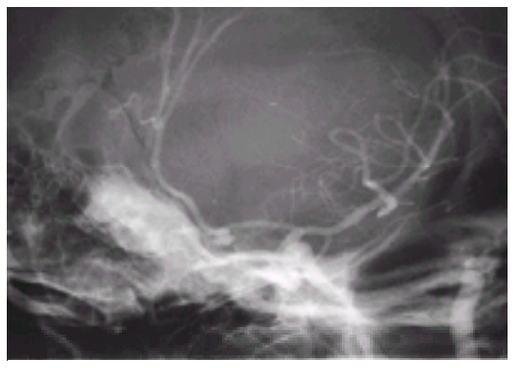


Fig. 19. Four years later, vertebral angiography showed an aneurysm arising at the junction between the right posterior inferior cerebellar artery and the vertebral artery; this was not seen at the previous investigation.

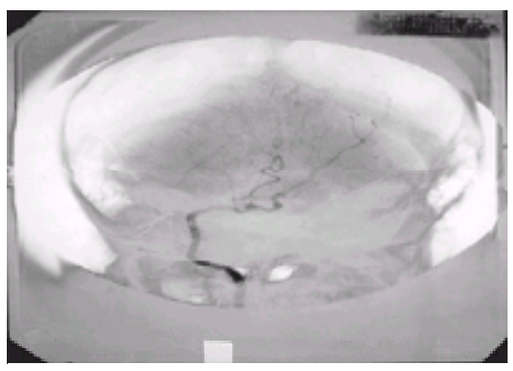


Fig. 20. Subtraction angiogram showing the aneurysm on the posterior inferior cerebellar artery more clearly. This technique subtracts the bone away and allows a clearer view of the angiographic contrast.

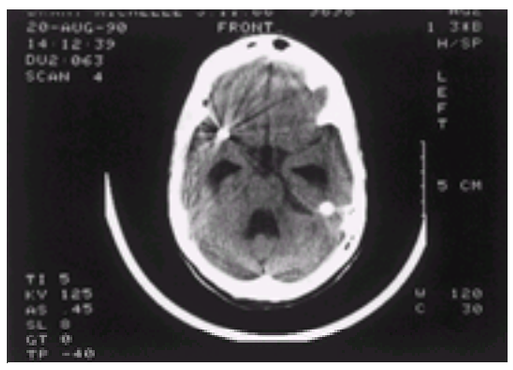


Fig. 21. CT scan showing a communicating hydrocephalus in the patient illustrated in Fig. 15, Fig. 16, Fig. 17, Fig. 18, Fig. 19 and Fig. 20. This is not surprising after two subarachnoid haemorrhages and two operations to clip four aneurysms. She made a good recovery after a shunt.

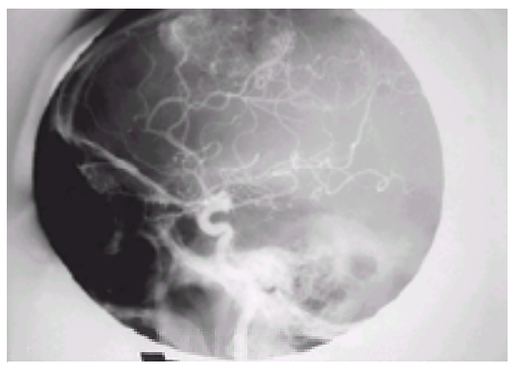


Fig. 22. Internal carotid angiogram showing a meningeal blush typical of a meningioma.

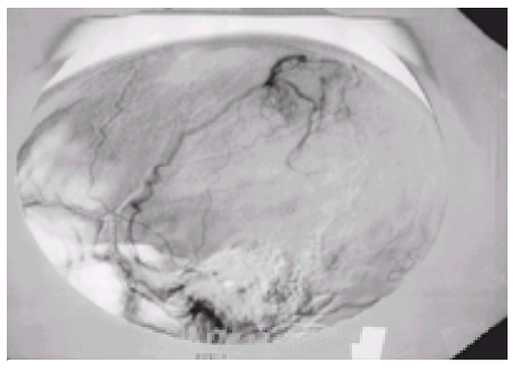


Fig. 23. External carotid angiogram showing the middle meningeal supply to a parasagittal meningioma; this is the typical source of blood supply to a meningioma.

Magnetic resonance imaging

This method of imaging depends on exciting molecules in a magnetic field. Its inability to show calcification, bone, or blood clot is offset by the lack of bone artefact. It is the diagnostic method of choice for small acoustic neuromas within the internal auditory meatus or small pituitary tumours in the pituitary fossa. It is also superior to CT scanning in demonstrating small lesions in the hemisphere, brainstem, and spinal cord. Cavernous angiomas (Fig. 24, Fig. 25) are particularly well shown, as are brainstem or spinal cord (intramedullary) gliomas (Fig. 26, Fig. 27). The multiple paraventricular lesions of multiple sclerosis are well demonstrated. Magnetic resonance imaging (**MRI**) has disclosed for the first time multifocal, low-density lesions associated with comparatively mild head injuries. These are not seen on CT scanning but confirm that mild injuries can have profound effects on brain function. Affected patients require careful sympathetic counselling.

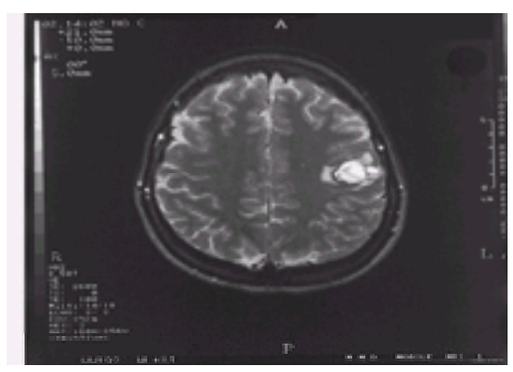


Fig. 24. MRI showing typical appearances of a cavernous angioma in a 23-year-old girl presenting with a sudden onset of right facial weakness diagnosed as Bell's palsy. She then developed focal motor seizures. At operation, old haemorrhage and haemosiderin staining was found around the lesion, which typically looks like a large blackberry.

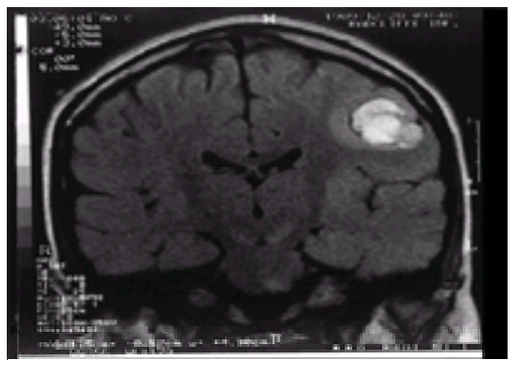


Fig. 25. Coronal MRI of the same patient shown in Fig. 24.

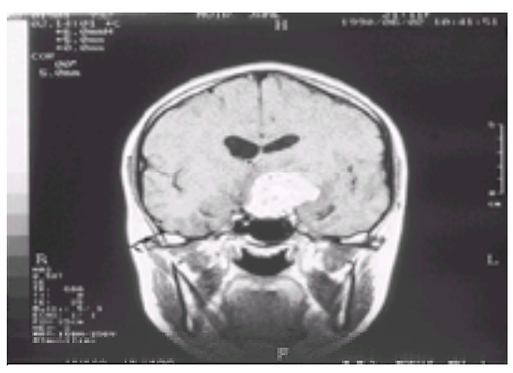


Fig. 26. MRI to show an intrinsic (glioma) tumour of the thalamus and hypothalamus. The vessels of the circle of Willis contain (gadolinium) contrast material. The temporal horns are seen (as black) on either side.

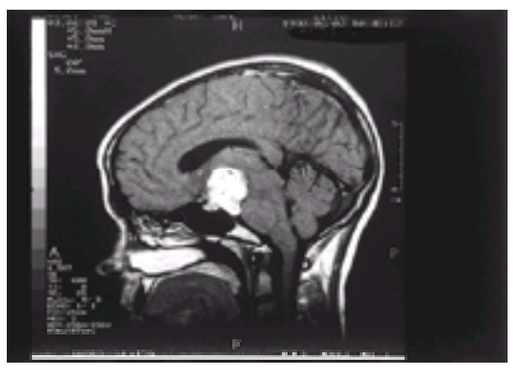


Fig. 27. MRI, sagittal view of the same patient as Fig. 26 to show the extent of the tumour. A biopsy examination showed this to be a glioma.

Electroencephalography

This has little value in neurosurgery except in the investigation of patients with epilepsy.

Neurophysiological peroperative monitoring of cranial nerve or spinal cord function is likely to be used more often in the future.

Further reading

Adams C. *A neurosurgeon's notebook*. Blackwell Scientific, Oxford, 1998.

Apuzzo MLJ. *Brain surgery, complication avoidance and management*. Churchill Livingstone, Edinburgh, 1993.

Congress of Neurological Surgeons. *Clinical neurosurgery*. Williams and Wilkins, Baltimore. [Annual volumes.]

Dudley H, Carter D, Russell RCG, ed. *Rob and Smith's operative surgery: neurosurgery*, 4th edn. Butterworth, London, 1989.

Jennett WB, Galbraith S. *An introduction to neurosurgery*, 4th edn. Heinemann, London, 1983.

Kaye AH. *Essential neurosurgery*. Churchill Livingstone, London, 1991.

Miller JD, ed. *Northfield's surgery of the Central nervous system*, 2nd edn. Blackwell Scientific, Oxford, 1987.

Oxbury JM, Adams CBT. Neurosurgery for epilepsy. *British Journal of Hospital Medicine*, 1989; **41** : 372–8.

Rengachary SS, Wilkins RH. *Principles of neurosurgery*. Mosby-Wolfe, St Louis 1994.

Schmideck HH, Sweet WH, ed. *Operative neurosurgery technique*, 2nd edn. Saunders, Philadelphia, 1988.

Yarsargil G. *Microsurgery*. George Thieme, Stuttgart, 1984.

Youmans JR, ed. *Neurological surgery*, 2nd edn. Saunders, Philadelphia, 1982.

44.2 Head injury

L. T. Dunn and G. M. Teasdale

Introduction
Epidemiology
Incidence
Mechanisms
Prevention
Classification and pathophysiology
Classification
The pathology of head injury
Cellular and molecular mechanisms
Assessment and investigation
Pre-hospital
Resuscitation
Assessment
Investigations
Guidelines
Management
Mild injury
Transfers
Non-surgical management
Surgical management
The sequelae of head injury
Cranial nerve deficits
Delayed complications
Outcome measures
Outcome after head injury
Evidence-based practice
Therapies under development
Further reading

Introduction

Head injuries are common and cause substantial disability, often in relatively young adults. Operative intervention is usually the province of neurosurgeons, but surgeons of almost all specialties are likely to have contact with patients with a head injury as part of their clinical practice. A sound knowledge of the assessment, resuscitation, and initial management of these patients is an important component of surgery in general because prompt recognition and treatment of secondary complications can lead to improved outcome.

Epidemiology

Incidence

Each year in the United Kingdom approximately one million patients attend accident and emergency departments with a head injury. About one patient in five is admitted, an annual incidence of 270/100 000. Few of those admitted in the United Kingdom are transferred to a regional neurosurgical unit (5–10 per cent). Many victims are young adult males, many injuries are related to alcohol intake, and many survivors experience prolonged sequelae.

Mechanisms

Among all admissions, of whatever severity, most are due to a fall or are the result of an assault. The most frequent cause of severe head injuries is motor vehicle accidents, responsible for between 48 and 53 per cent. Between 20 and 28 per cent of severe injuries are the result of a fall, and between 3 and 9 per cent result from sport and recreation activities.

Prevention

Public education and other measures to reduce risk are attractive means of reducing the incidence and cost of head injury. Alcohol is commonly involved in road traffic accidents, and in many falls and assaults. There is good evidence that the risk of an accident increases with increasing blood alcohol concentrations and appropriate legislation in a number of countries has led to a fall in the incidence of serious road traffic accidents. Better road design, speed control, and vehicle design have also contributed to a decrease in the risk of head injury and death after road traffic accidents. The evidence is most compelling for seat-belt use, one study showing that unbelted persons are 8.4 times more likely to sustain a head injury and 2.7 times more likely to sustain a fracture. Similar evidence supports the use of motorcycle and bicycle helmets.

Classification and pathophysiology

Classification

Head injuries can be classified according to mechanism and severity. Mechanistically the distinction is between missile and non-missile injuries. In a non-missile injury, far and away the most common in civilian practice, rapid deceleration and acceleration cause the brain to move within the cranial cavity and to come into contact with bony protuberances within the skull. This results in contusions, lacerations, and shearing strains within the brain substance (particularly when there is a rotational element to the acceleration). In this type of injury the acceleration/deceleration forces rather than actual impact against the skull are the critical factors in producing brain injury. In missile injuries, damage is the result of an object entering the cranial cavity and dissipating energy through the brain. Damage is often focal and the extent relates to the velocity of the missile.

The brain damage from head injury can be classified morphologically as focal or diffuse, and temporally as primary or secondary. Primary injury occurs at the time of impact and is distinguished from secondary brain injury, which occurs as a result of events after the initial injury. The distinction between these two types of injury is becoming less clear as more is learned about the underlying mechanisms responsible for damage after traumatic brain injury. It now appears that the initial injury sets in train a cascade of cellular and molecular events whose ultimate effect depends also on interactions with events subsequent to the injury. These processes may be amenable to treatment. (Table 1)

Primary brain injury
Cerebral contusions
Diffuse axonal injury and tissue tears
Subarachnoid haemorrhage
Diffuse vascular damage
Secondary brain injury
Intracranial haematoma
Secondary to raised intracranial pressure
Brain swelling
Ischaemic brain damage
Infection
Seizures

Table 1 A classification of primary and secondary brain injury

The pathology of head injury

Primary damage

Lesions of the scalp and skull

Scalp lacerations occur in about 40 per cent of patients who attend hospital with a head injury; they are useful diagnostically in helping to establish that a head injury has occurred if a history is not available. A skull fracture indicates the force of the injury, but many fractures are sustained without associated brain damage and, conversely, many fatal head injuries have no skull fracture. Skull fractures can be described in terms of location, shape, whether or not the fragments are depressed, and whether or not the wound is compound.

Cerebral contusions

This type of focal brain damage occurs at the time of injury as the result of contact between the brain and bony protuberances of the skull base. Consequently, contusions have a characteristic distribution (orbital surface of frontal lobes, frontal poles, around the Sylvian fissure, temporal poles, and undersurface of temporal lobes). They occur on the crests of gyri but commonly extend into subcortical white matter.

Diffuse axonal injury

This occurs at the time of injury and is the most common cause of coma after head injury. Widespread axonal injury occurs as a consequence of shear and tensile strains. It is now widely accepted that there is a spectrum of diffuse axonal injury. At the least severe level it occurs primarily in the parasagittal white matter of the cerebral hemispheres. At more severe levels there are, in addition, focal lesions in the corpus callosum; finally, at the most severe level, discrete lesions are found in the dorsolateral quadrants of the rostral brainstem. These discrete lesions may be visible macroscopically. Diffuse, microscopic axonal damage leads to the early occurrence of axonal swellings, which progress to the formation of axonal bulbs and later to collections of microglia. There is a low incidence of associated intracranial haematoma, skull fracture, or raised intracranial pressure.

Secondary damage

Intracranial haematoma

These are the most common cause of deterioration and death in patients who have experienced a lucid interval following injury. An extradural haematoma usually complicates a skull fracture that has disrupted a meningeal artery and occurs in the space between the dura and the inner table of the skull. If untreated, progressive midline shift, tentorial herniation, and midbrain compression may occur. There is often little evidence of other brain damage associated with the injury. Intradural haematomas may be subdural, intracerebral, or a combination of the two (burst lobe). Subdural haematomas may result from the tearing of bridging veins crossing the subdural space or from haemorrhage from severe cerebral contusions. Subdural haematomas spread more diffusely over the hemisphere than extradural and are often associated with diffuse swelling of the underlying hemisphere. They may be classified as acute, subacute or chronic, depending on their age. Intracerebral haematomas may occur in association with contusions or be more deep-seated, often in the basal ganglia where they tend to be multiple and indicative of severe injury. A burst lobe consists of subdural haematoma, contusion and intracerebral haematoma, and occurs most commonly in the frontal and temporal lobes. It generally indicates a severe injury.

Cerebral oedema and brain swelling

Ischaemic cerebral oedema occurs around contusions and haematomas. Damage to the blood–brain barrier allows extravasation of water, sodium, and protein molecules into the brain parenchyma. Diffuse swelling of one or both hemispheres may also occur and has been termed congestive oedema. The pathogenesis of this type of swelling in brain injury is not well understood.

Brain damage secondary to raised intracranial pressure

Raised intracranial pressure may result in brain herniation. In tentorial herniation the uncus of the temporal lobe herniates through the tentorial incisura. Cingulate herniation occurs when the cingulate gyrus is forced under the falx. In tonsillar herniation (coning) the cerebellar tonsils are forced down into the foramen magnum. Raised intracranial pressure can also result in ventricular displacement (contralateral obstructive hydrocephalus), flattening of cortical gyri, compression of the posterior cerebral arteries and infarction of the occipital lobe, and compression of the anterior choroidal arteries and infarction in the globus pallidus.

Ischaemic brain damage

This is common in severe head injury and occurs in 80 to 90 per cent of fatal head injuries. It is often found adjacent to contusions and haematomas, secondary to raised intracranial pressure, at the boundary zones of arterial territories, or as diffuse, hypoxic damage. It probably occurs soon after injury, and there is a relation with episodes of systemic hypotension and hypoxia.

Infection

Infection after head injury is relatively rare (4 per cent of fatal head injuries). Meningitis may occur as the result of spread of micro-organisms via open skull fractures or fractures of the base of the skull in non-missile head injuries. Brain abscess is more commonly a complication of missile head injuries.

Cellular and molecular mechanisms

The cellular and molecular events surrounding traumatic brain injury have been extensively studied through experiments in animal models and intensive investigation in patients. Cerebral hypoperfusion and hypoxaemia result in a lack of oxygen essential for cellular functioning and survival. The resultant energy failure then leads to a failure in membrane homeostasis and a cascade of further events contributing to cell damage and dysfunction (Fig. 1). Mechanically injured neurones have a heightened susceptibility to these effects and there is a complex interaction between events at a cellular and a systemic level (Fig. 2).

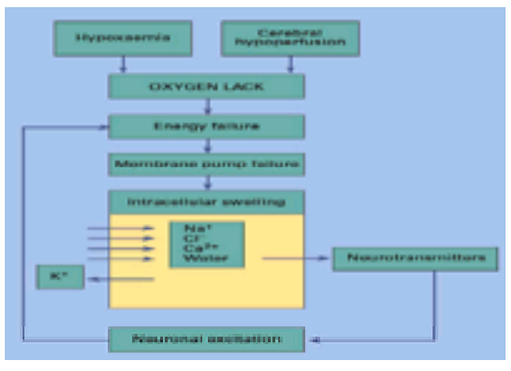


Fig. 1. Metabolic factors contributing to cell damage after head injury.

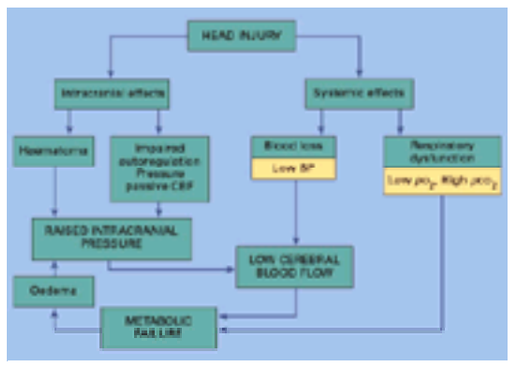


Fig. 2. The interaction between systemic and intracranial factors after head injury.

At a cellular level a variety of different mechanisms have been identified that may contribute to neuronal damage (Fig. 3). The relative importance of these mechanisms in traumatic brain injury in man is not fully established, although it seems likely that raised intracellular calcium and increased release of excitotoxic transmitters such as glutamate are important. Much current interest focuses on which if any may offer useful points for therapeutic intervention.



Fig. 3. Molecular and biochemical mechanisms leading to cell death after trauma or ischaemia. Derived from: Siesjo (1993)(see Further Reading).

Assessment and investigation

Pre-hospital

Information about the mechanism of injury and the condition of the victim immediately following the accident are of considerable value in assessing severity, predicting the type of damage, and assisting in planning management. Such information is frequently available from paramedics and witnesses at the scene but is not always collected. Relevant medical and drug history may also have a bearing on subsequent management and this information should be obtained if at all possible. Many patients with severe head injury will be transferred from the hospital of initial admission to a tertiary care unit, so accurate documentation, communication, and resuscitation are particularly important.

Resuscitation

Immediate resuscitation

A number of reports over recent years have highlighted the incidence of undiagnosed and untreated injuries in trauma victims, particularly in relation to the chest and abdomen, and it seems clear that a number of deaths after trauma are preventable. The Advanced Trauma Life Support system developed by the American College of Surgeons has become widely accepted as an effective and workable system of early trauma care. This system emphasizes early detection and treatment of life-threatening injuries, particularly in relation to the airway, ventilation, and circulation. These principles apply particularly to victims of a head injury.

Importance and incidence of extracranial injuries

The presence of systemic hypotension results in a doubling of fatality after severe head injury. Such hypotension is almost always the result of extracranial blood loss. Extracranial injury is common in association with severe head injury (Table 2).

Extracranial injury	%
Pulmonary	78
Musculoskeletal	43
Abdominal	53
Cardiovascular	3
Endocrine	23
Spine	6

Table 2 The incidence of extracranial injury in association with severe head injury

Following head injury, cerebrovascular autoregulation is often impaired and the brain therefore becomes even more vulnerable to the effects of hypotension and hypoxia (Fig. 4).

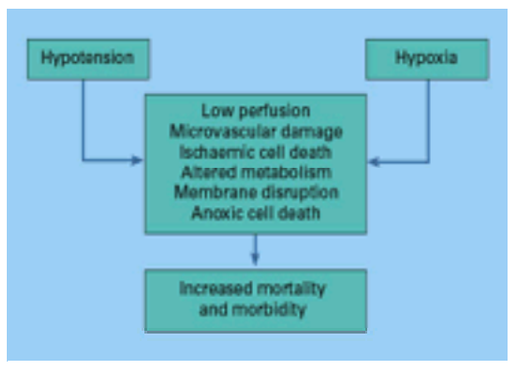


Fig. 4. The effects of hypoxia and hypotension on the injured brain.

Assessment

The Glasgow coma scale

The most important aspect of clinical assessment of head injury is repeated assessment of conscious level. This can be most reliably, consistently, and accurately done using the Glasgow coma scale (Table 3).

Eye opening	*Best motor response*	*Best verbal response*
1 None	1 None	1 None
2 To pain	2 Extension	2 Sounds only
3 To speech	3 Abnormal flexion	3 Words
4 Spontaneously	4 Flexion	4 Confused speech
	5 Localizing	5 Orientated
	6 Obeys commands	

After Teasdale and Jennett (1974)

Note. The scale refers to the best motor response. If there is asymmetry in the response then the best response is scored. The most useful painful stimuli are pressure to the finger tip to assess arousal and to the supraorbital margin. For a patient to localize the hand must be lifted above the level of the chin. If a particular component of the scale cannot be scored (e.g. eyes swollen shut, endotracheal tube in place, limb splinted etc.), this should be recorded. More information is conveyed if the score for each of the three components is reported separately. Coma has a specific meaning in terms of the scale: inability to obey commands, open eyes or utter words. Consequently the vast majority of patients in coma will have a total score of 8 or less.

Table 3 The Glasgow coma scale

Observations for the Glasgow coma scale are usually recorded sequentially on a coma chart such as that shown in Fig. 5, along with readings for pulse and blood pressure, pupil equality and reactivity, limb strength, and sometimes oxygen saturation. These additional measurements are important in assessment and in the early detection of secondary injuries such as hypotension, hypoxia, and developing intracranial mass lesions. Clinically the hallmark of rising intracranial pressure is a decreasing conscious level. The severity of injury is usually classified according to the score on the Glasgow coma scale after resuscitation. A total score of 8 or less signifies a severe injury; a score between 9 and 12 a moderate injury, and a score of 13, 14 or 15 a mild injury.

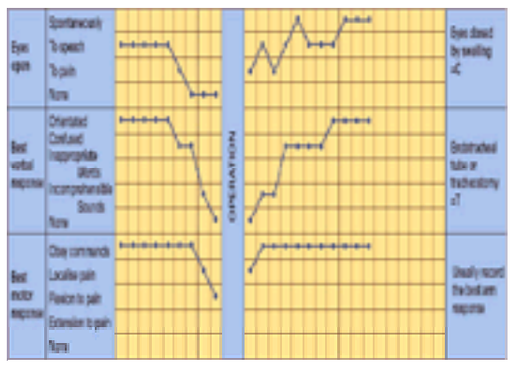


Fig. 5. A typical coma chart in a patient with head injury. Hospital charts should also include the facility to chart pulse, blood pressure, pupil reactivity, and limb strength in addition to the Glasgow coma scale.

Pupils and lateralizing neurological signs

The presence of pupillary asymmetry and/or asymmetry of motor response is useful clinically in predicting the presence of intracranial mass lesions. Expanding mass lesions result in a rise in intracranial pressure and a fall in cerebral perfusion pressure. They may also result in brain herniation within the skull, often with characteristic clinical signs as a result (Fig. 6).

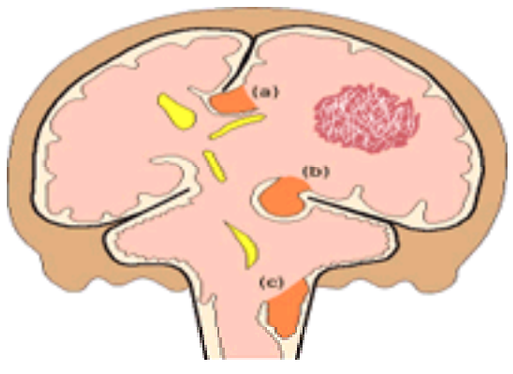


Fig. 6. Intracranial herniation. (a) Note cingulate herniation may compromise the anterior cerebral artery against the falx but is not associated with specific signs or symptoms. (b) Uncal herniation, the most common herniation observed clinically. The uncus of the temporal lobe herniates between the rostral brainstem and tentorial edge into the posterior fossa, resulting in a clinical syndrome of progressively impaired consciousness, dilated ipsilateral pupil, and contralateral hemiplegia. Impaired consciousness results from compression of the reticular activating system in the rostral brainstem. The dilated pupil results from compression of the parasympathetic pupillomotor fibres carried in the third nerve, and the hemiplegia from compression of the cerebral peduncle carrying corticospinal fibres to the opposite side. In cases where there is compression of the contralateral cerebral peduncle, hemiplegia may occur on the side ipsilateral to the lesion—a so-called false-localizing sign. In cases where a posterior cerebral artery is compressed against the edge of the tentorium, occipital infarction may follow. (c) Tonsillar herniation. The tonsils of the cerebellum herniate through the foramen magnum into the upper spinal canal compressing the medulla. Clinically this may result in cardiorespiratory impairment, hypertension, high pulse pressure, Cheyne–Stokes respiration (a cycle of gradually increasing amplitude of respiration followed by a tailing off and a period of apnoea), neurogenic hyperventilation, deep coma, and brainstem death. The combination of bradycardia and hypertension is known as Cushing's response and occurs in about one-third of

cases of tonsillar herniation.

Investigations

Imaging

The imaging investigation of choice in the acute stages of head injury is unenhanced computed tomographic (**CT**) scanning of the head. This facility is becoming increasingly available in hospitals admitting head-injured patients and can provide all the information necessary for early management. Plain radiographs of the skull remain important in the triage of minor injuries and as a screening method in centres where a CT scanner is not available. Lateral and frontal views should be obtained. The main value of plain films is in the detection of skull fracture because this is an important index of the risk of intracranial haematomas (Table 4).

GCS	*Risk*	*Other fractures*	*Risk*
15	1 in: 3615	None	1 in: 31300
		PTA	1 in: 6700
		Skull fracture	1 in: 81
		Skull fracture and PTA	1 in: 29
9–14	1 in: 51	No fracture	1 in: 180
		Skull fracture	1 in: 5
3–8	1 in: 7	No fracture	1 in: 27
		Skull fracture	1 in: 4

GCS, Glasgow coma scale; PTA, post-traumatic amnesia.
Data derived from Teasdale *et al.* (1990).

Table 4 The risk of an operable intracranial haematoma in head-injured patients: assessment of conscious level made at first hospital attendance

Some of the common CT abnormalities that may follow a head injury are illustrated in Fig. 7 and the indications for CT scanning in head-injured patients are discussed below.

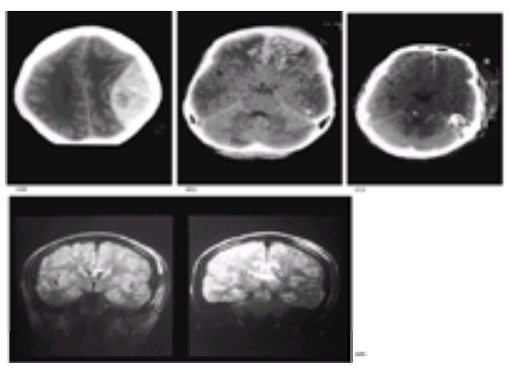


Fig. 7. Examples of CT and magnetic resonance imaging (MRI) abnormalities seen after blunt head injury. (a) CT of an extradural haematoma; (b) CT of bilateral frontal contusions; (c) CT of deep-seated parenchymal haematomas including a haematoma in the brainstem; (d) coronal MRI scan showing the type of focal lesion seen after severe, diffuse axonal injury.

A patient who has suffered blunt head trauma and has a decreased conscious level should have imaging of the cervical spine (there is an incidence of spinal-cord injury of 10 per cent associated with severe, blunt head injury) as well as any other imaging investigations appropriate for any extracranial injuries. Adequate radiographs of the cervical spine should include C7/T1. If plain films are inadequate, additional imaging (e.g. CT scanning through the relevant levels) should be undertaken.

Other investigations

A victim of multiple trauma should have radiographs taken of the cervical spine, chest and pelvis, and other imaging according to the pattern of injury. Arterial blood gases should be determined in patients with head injury because of the deleterious effects of hypoxia and hypercarbia on the injured brain. Blood should be cross-matched as part of the initial evaluation and resuscitation. Patients with a history of bleeding disorder or who are on anticoagulant medications should have clotting factors determined and any coagulopathy should be corrected. Coagulopathy is particularly a risk in patients with a history of alcohol abuse.

Guidelines

Guidelines for the admission, investigation, and transfer of head-injured patients have proliferated in recent years. The Society of British Neurological Surgeons has recently endorsed guidelines applicable to United Kingdom practice and these are detailed in Fig. 8.



Fig. 8. Guidelines for the management of blunt head injury.

Management

Mild injury

Most patients attending hospital with head injury have a mild injury. Those meeting the indications for hospital admission detailed in Fig. 8 should be admitted. Their

Glasgow coma scale should be checked and charted on arrival, and these patients should then undergo a period of observation aimed at early detection of signs of increasing intracranial pressure or intracranial infection. Before discharge, further medical assessment should be undertaken. Patients who are fully orientated when assessed and who have no skull fracture and no other indication for admission can usually be discharged home in the company of a responsible individual. Most accident and emergency departments have head-injury instructions detailing symptoms and signs that should prompt a return to hospital.

Transfers

Many moderately or severely injured patients will require transfer from the hospital of their initial admission to a neurosurgical unit for definitive care. A number of studies have shown that a high proportion of such patients suffer potentially preventable secondary insults during this process, especially as a result of hypoxia and hypotension. Guidelines for the transfer of head-injured patients have been published in the United Kingdom by the Neuroanaesthesia Society and the Association of Anaesthetists. Agreed protocols for the transfer of head-injured patients between district general hospitals and regional neurosurgical units that take account of particular local circumstances should be established and known to all those involved in the care and transfer of these patients.

Non-surgical management

Management in the intensive-care unit

Continuing care of the head-injured patient in the intensive-care unit follows the same principles as those guiding initial resuscitation, namely the detection, treatment, and prevention of secondary insults to the injured brain. Systemic resuscitation and the maintenance of adequate perfusion and oxygenation are priorities. Measures to reduce intracranial pressure are only employed when there is evidence from monitoring of intracranial hypertension.

Approaches to sedation and the use of neuromuscular blockade vary widely. There have been no studies on the effect of sedation on outcome after head injury. The routine use of neuromuscular blockade appears to carry no benefit and results in a more prolonged course in intensive care and a higher incidence of pneumonia. Neuromuscular blockade should be reserved for specific indications such as the treatment of intracranial hypertension and transport.

Systolic blood pressure of less than 90 mmHg and PaO_2 of less than 60 mmHg (approximately 8 kPa) are associated with increased morbidity and mortality and should be vigorously treated.

Hypermetabolism and nitrogen loss occur in the week after head injury. If feeding is not instituted patients lose approximately 15 per cent body weight per week. Full nutritional support by either the enteral or parenteral route should be introduced at the latest by the end of the first week after injury; early enteral feeding may reduce infection.

Monitoring of intracranial pressure and medical management of raised intracranial pressure

Intracranial pressure is routinely monitored in severely head-injured patients in many neurosurgical centres. Although there is a clear association between raised intracranial pressure and increased mortality and morbidity after head injury, there has not yet been a randomized, prospective, controlled trial to demonstrate that present treatments of raised intracranial pressure improve outcome. Nevertheless, the logic of it having a role in producing secondary damage means that there are a variety of extradural, subdural, intraparenchymal, and intraventricular devices available to enable continuous monitoring of intracranial pressure and many of these can be inserted at the bedside, in the intensive care unit.

Indications for intracranial pressure monitoring have been suggested by the American Association of Neurological Surgeons as follows.

1. Comatose head-injured patients (Glasgow coma scale 3–8) with an abnormal CT scan.
2. A comatose head-injured patient with a normal CT scan and two or more of:
   a. age over 40–years,
   b. unilateral or bilateral motor posturing,
   c. systolic blood pressure less than 90 mmHg.

Considerable debate continues about the appropriate threshold at which to commence treatment for raised intracranial pressure. It is unlikely that a uniformly applicable threshold exists and the value of the intracranial pressure must be considered in the context of the cerebral perfusion pressure. However, what data are available support an upper threshold of between 20 and 30 mmHg.

A number of measures are available that may reduce or prevent intracranial hypertension.

General measures include: the control of pyrexia; the treatment of seizures; elevation of the head of the bed; avoidance of jugular venous outflow obstruction; sedation with or without pharmacological paralysis; maintenance of adequate arterial oxygenation; volume resuscitation.

Specific measures (in the first tier) include: mannitol; hyperventilation to low normal $PaCO_2$(30–35 mmHg) with concomitant monitoring of jugular venous oxygen saturation; repeat CT to exclude any surgically treatable mass lesion. Specific measures (in the second tier; value unproved) include: barbiturate therapy; hypothermia; hyperventilation to $PaCO_2$ < 30 mmHg; decompressive craniectomy; hypertensive therapy.

Many of these measures have been shown to lower intracranial pressure but evidence that they improve outcome is lacking. The second-tier measures are associated with a significant incidence of complications and side-effects.

Prophylactic antibiotics and anticonvulsants

A number of prospective, controlled studies have shown that the prophylactic use of phenytoin, carbamazepine, and phenobarbital does not reduce the incidence of late post-traumatic seizures and there is thus no indication for their routine use. There is some evidence to suggest that prophylactic anticonvulsants may prevent early post-traumatic seizures but there is no evidence that preventing such seizures will improve outcome in the longer term.

Evidence on the use of prophylactic antibiotics for cerebrospinal fluid rhinorrhea and otorrhoea is more conflicting. There is an overall risk of meningitis of 30 per cent associated with these conditions, most commonly caused by the pneumococcus. Only one recent retrospective study has demonstrated a significant benefit of prophylactic antibiotics, and a recent meta-analysis of studies found no difference between patients treated with prophylactic antibiotics and those not. Persistent cerebrospinal fluid fistulas should be repaired surgically if possible.

Surgical management

Haematomas

An intracranial haematoma is the most common cause of secondary deterioration after head injury. Early operative evacuation of a haematoma in selected patients reduces mortality and improves outcome. Craniotomy is not, however, required in all patients with an intracranial mass lesion on the CT scan.

Accepted indications for surgical removal are as follows (adapted from Brain Trauma Foundation, 1995—see Further Reading).

1. The presence of a mass lesion of more than 40 ml (suggested by midline shift of more than 5 mm, effacement of the third ventricle and basal cisterns, or dilatation of the contralateral temporal horn.
2. In conscious, non-ventilated patients:
   a. a decline in conscious state,
   b. the development of focal signs,
   c. severe or worsening headache, nausea or vomiting.
3. In unconscious ventilated patients:
   a. a decline in neurological state (the development of brainstem signs may be the only indication of this),
   b. increase in intracranial pressure, e.g. consistently greater than 25 mmHg.

4. An increase in haematoma size on serial CT scans.

Non-operative management may be considered:

1. in patients who are fully conscious;
2. when a thin, extra-axial mass lesion is the only abnormality;
3. when there are no features of mass effect on the CT scan.

Exploratory burr holes

There are very few indications for the use of exploratory burr holes either outside or inside a neurosurgical unit. A properly organized system of care for head injuries, properly applied in the individual patient, should make such circumstances extremely rare. Exploratory surgery may rarely be merited if a patient has shown marked deterioration in conscious level, with clear, localizing neurological signs, and a CT scan and/or definitive surgery cannot be arranged expeditiously. Such surgery should only be performed by a surgeon adequately trained in the technique and there are a number of potential pitfalls:

1. the majority of comatose head-injured patients do not have a haematoma;
2. a burr hole placed as little as 1 cm away from a haematoma may not locate it;
3. only a small portion of even an extradural haematoma can be evacuated through a burr hole;
4. the procedure itself may cause further brain damage or intracranial bleeding;
5. placing a burr hole may take as long as transferring the patient to a neurosurgical centre;
6. the procedure is at best a temporizing manoeuvre;
7. the additional time to perform a CT scan, if available, is only a few minutes;
8. ‘blind' exploratory surgery may actually delay definitive decompression.

Technique Under endotracheal intubation and anaesthesia, the whole head is shaved, prepared, and draped to allow access to the frontal, parietal, and temporal areas bilaterally. The most valuable clinical signs in determining the side of an extradural haematoma are a unilateral dilated pupil, which will be ipsilateral to the haematoma, an ipsilateral skull fracture, and a hemiparesis, which is usually contralateral to the haematoma. The next most reliable sign is a hemiparesis, which is usually contralateral to the haematoma. The first burr hole is placed over the fracture at the impact site, or in the absence of local signs, in the temporal region just above the zygoma and a finger-breadth anterior to the tragus, as this is the site of most extradural haematomas unless a fracture is present in a different location. A vertical skin incision is made and the temporalis muscle incised down to the skull. Bleeding from the muscle or superficial temporal artery is controlled with ligation, diathermy, or self-retaining retractors. The periosteum is scraped off the temporal bone and a burr hole is then drilled using a Hudson brace and perforator until an irregular wobbling is felt indicating perforation of the inner table. The opening is then completed with the burr. If an extradural haematoma is identified the hole can be enlarged, preferably by a neurosurgeon, using bone rongeurs to form a small craniectomy and allow more complete evacuation of the haematoma. If no abnormality is detected then further burr holes may be made in the ipsilateral frontal and posterior parietal regions but arrangements for CT scanning should not be further delayed. Identification and evacuation of a significant subdural haematoma can be very difficult through a burr hole and intradural exploration should be undertaken only by neurosurgeons.

Extradural haematomas

Craniotomy is the preferred operation for evacuation of extradural haematomas. Most extradural haematomas are temporal and can be approached through a standard, ‘question-mark' trauma flap. More posterior haematomas may require an inverted horseshoe flap, and frontal and subfrontal haematomas can be approached through a bicoronal skin incision.

Technique The position of the haematoma should be transposed on to the CT scout film to assist in flap planning. The head is placed on a horseshoe headrest and the scalp shaved and prepped with povidone iodine. A myocutaneous flap is turned and then a free bone flap fashioned using a high-speed drill and craniotome attachment. Medially the bone flap should be taken no closer than 2 cm to the midline to avoid damage to the sagittal sinus. The inferior margin of the flap should allow easy access to the middle cranial fossa. Once the bone flap has been removed the haematoma is sucked and irrigated away. Bleeding dural vessels are coagulated with bipolar diathermy and bleeding bone edges controlled with bone wax. Dural hitch sutures are placed around the margins of the craniotomy to control bleeding from beneath the bone edges and to prevent recurrent haematoma formation. The dura is not routinely incised but may be if it remains tense and blue or if a subdural clot was shown on the CT scan. The bone flap is replaced and held in position with miniplates, or metal or nylon sutures. A suction drain is placed in the subgaleal space and left *in situ* for 12 to 24 h. The muscle and galea are closed with Vicryl and the skin with nylon or staples. Frequently an intracranial pressure monitor will be placed at the end of the procedure to assist with postoperative management in the intensive-care unit.

Subdural haematomas

Surgical management of acute subdural haematomas follows similar principles. The appropriate area as identified on the preoperative CT scan is exposed. The dura is opened using a sharp hook and scalpel. Several linear, unconnected incisions in the dura are often an effective means of evacuating the haematoma and at the same time preventing brain herniation. The haematoma should be suctioned and irrigated gently away whilst avoiding suctioning beyond the margins of the craniotomy and in areas not readily visible or accessible. Visible bleeders should be controlled with bipolar diathermy and Surgicell. Care should be taken not to miss a parasagittal venous bleeder (present in 40 to 50 per cent of acute subdural haematomas) as this is a common cause of recurrent haematomas. Obviously contused and necrotic brain can be removed by gentle suction and irrigation. Following removal of the haematoma the dura should be closed, with the assistance of a duroplasty if necessary. The bone flap is replaced and held in position as described above unless uncontrolled brain swelling makes this impossible. The remainder of the wound is closed as described above over a subgaleal suction drain. An intracranial pressure monitor is inserted at the end of the procedure.

Intraparenchymal lesions

When CT scanning shows significant mass effect due to lesions in the brain substance, operative evacuation is required and is more effective and definitive than medical methods of controlling intracranial pressure. The mass may be due to coalescence of contusions, the haemorrhage accompanying laceration of the brain surface, or to subcortical intraparenchymal clots resulting from shearing of perforating vessels. The temporal and frontal lobes are the usual sites involved and operation is guided by the CT findings.

The merits of decompression by evacuation of smaller lesions or craniectomy are controversial and clear evidence of benefit is lacking.

Depressed fractures

Surgical treatment of depressed fractures is undertaken either for wound toilet and debridement in the case of open injuries or to minimize cosmetic deformity. Depression of the outer table of the depressed fragment below the level of the inner table of the surrounding skull is usually considered an indication for operative elevation. An attempt is made to incorporate any scalp laceration into the surgical incision and obviously devitalized scalp tissue excised. A burr hole is usually fashioned a little distance away from the depressed fragment and sufficient bone rongeured away to allow elevation of the depressed fragment without causing any further damage to the underlying brain. If a dural laceration is identified, non-viable underlying brain is excised and the laceration repaired. Fragments of dirt, hair, and other foreign bodies are removed and the wound thoroughly irrigated with saline. The bone fragments may be replaced as a mosaic in areas where appearance is important, and the wound is then closed in layers. The scalp can usually be mobilized sufficiently to cover small scalp defects.

Cerebrospinal fluid fistulas

There are varying views about precisely when surgical repair of a cerebrospinal fluid fistula should be undertaken, but accepted indications are:

1. a cerebrospinal fluid leak that persists for 7 to 10–days;
2. a cerebrospinal fluid leak that ceases but recurs after 7 to 10–days;
3. when there is clinical evidence of a large dural defect, for example a large aerocele or escape of brain tissue through the nostrils;
4. if meningitis or brain abscess occurs.

Repair is usually undertaken electively; acute transcranial repair is often difficult because of brain swelling. If there are associated facial fractures, a combined procedure is undertaken with maxillofacial surgeons. Some fractures involving the sphenoid or ethmoid bones can be approached and repaired subcranially with the

advantage that brain retraction is not necessary.

The sequelae of head injury

Cranial nerve deficits

Cranial nerve palsies are found in about one-third of survivors of a severe, closed head injury. Loss of smell is reported by up to 20 per cent of patients with a history of loss of consciousness. A fracture is found in less than half of these cases. Recovery occurs in 15 to 50 per cent of cases but is unusual if anosmia persists at 3 months after the injury.

Injury to any part of the visual pathway may occur after head injury. Optic nerve lesions occur in 4 per cent of patients after blunt head injury, most commonly affecting the intracanalicular portion of the nerve. This results in monocular blindness with an unreactive pupil, usually appearing immediately after injury.

Diplopia is common after head injury and often reflects orbital rather than cranial nerve pathology. Ocular nerve palsies after head injury affect the VIth nerve in 34 per cent of cases, the IIIrd nerve in 30 per cent, and the IVth in 15 per cent. More than one nerve is involved in 22 per cent of cases. A IIIrd nerve palsy may result from impact damage or transtentorial herniation. Recovery usually occurs. VIth nerve palsies are usually associated with a fracture of the petrous temporal or sphenoid bone and also often recover spontaneously.

Damage to the VIIth and VIIIth nerves often accompanies a fracture of the petrous temporal bone. A transverse fracture may disrupt the facial nerve and produce a facial nerve palsy that is immediate and usually permanent. A longitudinal petrous fracture is more often associated with a palsy that appears after a delay, is incomplete, and usually recovers over 6 to 8 weeks. A fracture that disrupts the labyrinth and utricle often causes severe vertigo and nystagmus, which can last 6 to 12 weeks until central compensation occurs. Hearing loss is a frequent complaint after head injury, is most commonly of a conductive type, and associated with a haemotympanum or lacerated ear drum. This often resolves. Sensorineural type hearing loss may also occur in association with a petrous fracture or as a result of concussive damage to the organ of Corti.

Delayed complications

Hydrocephalus

Hydrocephalus after head injury is usually the result of atrophy of damaged white matter and is related to the severity of injury (*ex vacuo* hydrocephalus). Occasionally it is a consequence of obstruction of cerebrospinal fluid flow following bleeding (obstructive hydro-cephalus); or as a chronic communicating hydrocephalus (normal-pressure hydrocephalus). Those with obstructive or normal-pressure hydrocephalus may benefit from shunting.

Post-traumatic epilepsy

This is the most common delayed complication after head injury and occurs in about 5 per cent of non-missile head-injured patients admitted to hospital. Some types of injury are associated with a higher risk (Table 5). An operated intracranial haematoma, early seizures, and compound depressed fractures of the vault all increase the risk of late seizures. Seizures during the first week after injury often do not recur and are regarded separately from late-onset seizures. In approximately 80 per cent of patients who develop post-traumatic epilepsy the first seizure is within 2 years of the injury, but some have a first seizure more than 4 years after their injury.

Risk factor	*Incidence of late seizures (%)*
Penetrating missile injury	53
Intracerebral haematoma—laceration	39
Focal brain damage on early CT	32
Early seizures	25
Depressed fracture—torn dura	25
Extradural or subdural haemorrhage	20
Focal signs	20
Depressed skull fracture	15
Loss of consciousness > 24h	5
Linear fracture	5
Mild concussion	1

Table 5 Risk factors for late post-traumatic seizures

Mental and cognitive sequelae

Mental and cognitive problems and changes in personality are usually the most disabling sequelae of head injury for the patient and his or her family. The presence and severity of these problems is related to the severity of the initial injury as assessed by duration of coma and length of post-traumatic amnesia. Changes in personality are more common than alterations in intellectual function. The mental changes seen after head injury are most often related to widespread rather than focal damage, although some patients develop patterns of deficit consistent with predominantly frontal lobe damage and in others the predominance of memory problems may be related to greater damage in the temporal lobes. Neuropsychological deficits can be correlated to some extent with diffuse damage seen on MRI 15 to 18 months after injury but not with focal lesions on imaging studies done sooner after injury.

Outcome measures

The most commonly used measure of outcome after head injury is the Glasgow outcome scale, devised by Jennett and Bond (Table 6).

1	Dead
2	Persistent vegetative state. Patient exhibits no obvious cortical function
3	Severe disability (conscious but disabled). Patient dependent upon others for daily support (some mental or physical disability or both)
4	Moderate disability (disabled but independent). Patient is independent so far as daily life is concerned. The disabilities found include varying degrees of hemiplegia, aphasia, memory, or intellectual deficit, personality changes
5	Good recovery. Resumption of normal activities even though there may be minor neurological or psychological deficits

Table 6 The Glasgow outcome scale

The scale has the advantage of being simple and easy to apply, and its reliability and validity have been established in a number of studies. It is most useful for broad classifications, for example, in series of patients, and does not aim to reflect small improvements occurring late in recovery that are not large enough to result in a change in category but are nevertheless significant for patients and their families. The Disability Rating Scale was devised by Rappaport to ‘track' patients from the initial to the late stages of recovery but is even less sensitive to late changes in mental sequelae. Self-administered questionnaires such as the ‘Short Form 36' give a measure of the patient's own perception of their health in relation to the general population.

Prediction of outcome (Table 7)

Age is a major determinant of outcome after head injury. With increasing age above 5 years old, outcome becomes significantly worse. Elderly patients also have a greater predisposition to an intracranial mass lesion. The post-resuscitation Glasgow coma scale score is also a powerful predictor of outcome. The motor component of the score is particularly important in severe head injury. Patients with no motor response after resuscitation have a mortality of 76 per cent. Pupillary size and reactivity are also important. In patients with normal pupils throughout their hospital course, 10 per cent are dead or vegetative at discharge compared with a 61 per cent mortality in patients who develop unilateral pupil abnormality after resuscitation and 86 per cent mortality in patients both of whose pupils remain fixed and dilated after resuscitation. Arterial hypotension (systolic blood pressure < 90 mmHg) is also associated with poorer outcome, and this, in association with a severe head injury, leads to a doubling of mortality. Raised intracranial pressure is also associated with worse outcomes. The extent to which this is a direct effect of the intracranial pressure and the extent to which it is because intracranial pressure reflects established damage is still uncertain.

	Dead or vegetative (%)	*Good recovery or moderate disability (%)*
Age (years):		
0–29	39	50
30–59	49	34
>60	81	11
Coma score:		
3–5	84	11
6–7	56	29
8–10	28	58
11–15	16	72
Pupils:		
Both fixed	86	6
One or both reacting	16	72

Table 7 Determinants of outcome after head injury

Outcome after head injury

Mild head injury

Mortality in the year after admission to hospital with mild head injury (Glasgow coma scale 13–15) is between 5 and 10 per cent. The long-term sequelae in survivors are often underestimated (Table 8). Headache, dizziness, vertigo, irritability, inability to concentrate, impaired memory and easy fatiguability are frequent complaints, and many patients reporting these symptoms have measurable neuropsychological deficits. In most patients these symptoms improve over the months following the injury, but they persist in a substantial proportion.

	At discharge	*At 1 year*
Headache	36	18
Dizziness	17	14
Depression	8	18

Table 8 Symptoms reported after mild head injury

Moderate head injury

Few studies have looked in detail at outcome following moderate head injury (Glasgow coma scale 9–12). Patients in this group tend to be slightly older, to have a higher incidence of pre-injury problems such as alcohol abuse, and a higher incidence of focal intracranial lesions. At 3 months only 38 per cent make a good recovery as assessed by the Glasgow outcome scale, and 66 per cent of those previously employed have not returned to work.

Severe head injury

Patients with a severe injury have a Glasgow coma scale of 3 to 8 after resuscitation. More than half of this group have a focal lesion on their CT scan. Mortality in this group ranges between 28 and 51 per cent. Recent data from the United Kingdom show that about 45 per cent make a good recovery or have moderate disability, and mortality is 37 per cent (Table 9).

	Follow-up	Dead (%)	Vegetative (%)	Severe disability (%)	Moderate disability (%)	Good recovery (%)
A	1037	6 months	45	15	20	18	16
B	1253	6 months	37	13	18	18	23

Cohort A, coma > 6h; cohort B, includes patients with coma < 6h.

Table 9 Outcome after severe head injury in the United Kingdom

Evidence-based practice

The most comprehensive review of evidence-based practice in the management of severe head injury has been published by the American Association of Neurological Surgeons. They have systematically reviewed clinical trials relating to a variety of topics in head injury management and weighted evidence according to the quality of trial design. Based on this review they have produced a series of recommendations of differing degrees of certainty. Standards represent principles of patient management with a high degree of clinical certainty. Guidelines represent recommendations associated with a moderate degree of clinical certainty. Options are defined as strategies for patient management for which there is unclear clinical certainty. It is apparent from this assessment that there is a paucity of good clinical studies and in particular a shortage of prospective, randomized trials to guide clinical practice. Of the 12 clinical management topics assessed there were sufficient data

to allow the recommendation of standards in only three, and each of these was negative:

1. In the absence of increased intracranial pressure, chronic prolonged hyperventilation therapy (*P*aCO_2 of 25 mmHg or less) should be avoided after severe traumatic brain injury.
2. The use of glucocorticoids is not recommended for improving outcome or reducing intracranial pressure in patients with severe head injury.
3. Prophylactic use of phenytoin, carbamazepine, or phenobarbital is not recommended for preventing late post-traumatic seizures.

Guidelines for aspects of management of head-injured patients have been produced by several organizations. Despite the relative shortage of definitive, ‘Class I' evidence there is an immense coherence, and certainly no major conflict, in the recommendations derived from ‘consensus' in different groups. The use of these guidelines can be recommended strongly in order to achieve consistency and to avoid confusion in the management of head injury.

Therapies under development

Major advances have been made in understanding the pathophysiological mechanisms underlying traumatic brain injury in animal models and a variety of mechanisms identified that may be amenable to pharmacological treatment. A number of agents have shown promise in experimental studies and several of these, including free-radical scavengers, synthetic corticosteroids, calcium-channel blockers, calcium antagonists, amino steroids and *N*-methyl-D-aspartate antagonists, have undergone clinical evaluation.

So far the results of clinical trials have been disappointing. With the exception of nimodipine in patients with a traumatic subarachnoid haemorrhage, none of the agents subjected to clinical trials so far has been shown to have a robust clinical effect. Future clinical trials may need to be targeted at mechanisms that may be important in particular subgroups of the head-injured population. Nearly all trials so far have been in severe head injury. Moderate and mild head injury cause considerable morbidity but relatively little mortality and are considerably more common than severe head injury. The development of agents effective in these patients remains as a challenge for the future.

Further reading

Anderson ID *et al.* Retrospective study of 1000 deaths from injury in England and Wales. *British Medical Journal* 1988; **296**: 1305–8. [Review of trauma deaths in the United Kingdom.]

Bartlett J, Mendelow D, Pickard J, Teasdale G. Guidelines for the initial management of head injuries. Recommendations from the Society of Neurological Surgeons. *British Journal of Neurosurgery* 1998;**12**: 349–52. [Recently developed guidelines for head injury management applicable to the United Kingdom.]

Blumbergs PC. Pathology. In: Reilly P, Bullock R, ed. *Head injury. Pathophysiology and management of severe closed head injury*, pp. 39–70. Chapman and Hall Medical, London, 1997. [Review of pathological changes after head injury.]

Bracken MB *et al.* A randomized controlled trial of methylprednisolone or naloxone in the treatment of acute spinal cord injury. *New England Journal of Medicine* 1990;**322**: 1405–11

Bullock MP, Chesunt RM, Clifton G, *et al.* Guidelines for management of severe head injury. *Journal of Neurotrauma* 1996; **13**: 639–734. [A systematic analysis of clinical evidence available to guide management of severe head injury.]

Cartledge NEF. Post-concussional syndrome. *Scottish Journal of Medicine* 1977;**23**: 103.

Chestnut RM *et al.* The role of secondary brain injury in determining outcome from severe head injury. *Journal of Trauma* 1993; **34**: 216–22. [Prospective study of 717 patients showing that hypotension and hypoxia adversely influence outcome independently of other prognostic factors.]

Committee on Trauma. *Advanced trauma life support for doctors, student course manual*, 6th edn. American College of Surgeons, Chicago, 1997. [The ATLS trauma system.]

Doppenberg EMR, Bullock R. Clinical neuro-protection trials in severe traumatic brain injury: lessons from previous studies. *Journal of Neurotrauma* 1997; **14**: 71–80.

Dunn LT. Head injury therapies. *Expert Opinion on Investigational Drugs* 1997; **6**:1511–26. [Review of clinical trials in head injury therapy.]

Gentleman D, Jennett B. Audit of transfer of unconscious head-injured patients to a neurosurgical unit. *Lancet* 1990; **335**: 330–4.

Graham DI *et al.* Ischaemic damage is still common in fatal non-missile head injury. *Journal of Neurology, Neurosurgery and Psychiatry* 1989; **52**: 346–50. [A recent paper confirming the importance of ischaemic damage in human head injury.]

Group of Neurosurgeons. Guidelines for initial management after head injury in adults. *British Medical Journal* 1984; **288**: 983–5.

ISRA safety restraint assessment—Iowa, 1978–1988. *Journal of the American Medical Association* 1989; **262**: 2804–5. [Effect of seatbelts on road traffic accident injuries.]

Jennett B. *Epilepsy after non-missile head injuries*, 2nd edn. Heinemann, London, 1975. [The most comprehensive study of post-traumatic epilepsy published to date.]

Jennett B. Epidemiology of head injury. *Journal of Neurology, Neurosurgery and Psychiatry* 1996; **60**: 362–9. [Review of head injury epidemiology world-wide.]

Jennett B, Bond M. Assessment of outcome after severe brain damage. *Lancet* 1975; **I**: 480–4. [The original description of the Glasgow outcome scale.]

Jennett B, MacMillan R. Epidemiology of head injury. *British Medical Journal* 1981: **282**: 101–4. [Epidemiology of head injury in the United Kingdom.]

Jennett B, Teasdale G, Braakman R, Minderhoud J, Heiden J, Kurze T. Prognosis of patients with severe head injury. *Neurosurgery* 1979; **4**: 283–9. [Factors influencing outcome after severe, closed head injury.]

Maas AIR *et al.* EBIC—Guidelines for Management of Severe Head Injury in Adults. *Acta Neurochirurgica* 1997; **139**: 286–94.

McIntosh TK, Smith DH, Meaney DF, Kotapka MJ, Gennarelli TA, Graham DI. Neuropathological sequelae of traumatic brain injury: relationship to neurochemical and biomechanical mechanisms. *Laboratory Investigation* 1996; **74**: 315–42. [Recent review of neuropathological and biochemical changes associated with experimental brain injury.]

McQueen JK *et al.* Low risk of late posttraumatic seizures following severe head injury: Implications for clinical trials of prophylaxis. *Journal of Neurology, Neurosurgery and Psychiatry* 1983; **46**: 899–904. [The use of prophylactic anticonvulsants after head injury.]

Mendelow AD *et al.* Risks of intracranial haematoma in head injured adults. *British Medical Journal* 1983; **287**: 1173–6. [The risk of intracranial haematoma after blunt head injury in relation to conscious level and the presence of skull fracture on presentation to hospital.]

Murray LS *et al.* Does prediction of outcome alter patient management? *Lancet* 1993; **34**: 1487–91.

Ommaya AK, Thibault LE, Bandak F. Mechanics of impact head injury. *International Journal of Impact Engineering* 1994; **15**: 535–60. [The biomechanics of head injury.]

Pagni CA. Posttraumatic epilepsy, incidence and prophylaxis. *Acta Neurochirurgica* 1990; Suppl. 50: 38–47. [A review of the factors influencing the risk of post-traumatic epilepsy.]

Potter JM, Braakman R. Injury to the facial nerve. In: Vinken PJ, Bruyn GW, ed. *Handbook of clinical neurology*, Vol. 24, pp. 105–117. North Holland, Amsterdam, 1976.

Rappaport M *et al.* Disability rating scale for severe head trauma: coma to community. *Archives of Physical Medicine and Rehabilitation* 1982; **63**: 118–23.

Rathmore MH. Do prophylactic antibiotics prevent meningitis after basilar skull fracture?*Paediatric Infectious Disease Journal* 1991; **10**(2): 87–8. [Meta-analysis of trials of the use of prophylactic antibiotics for traumatic cerebrospinal fluid fistulas.]

Rucker CW. The causes of paralysis of the third, fourth and sixth cranial nerves. *American Journal of Ophthalmology* 1966; **61**: 1293–8.

Seelig JM, Becker DP, Miller JD, Greenberg RP, Ward JD, Choi SC. Traumatic acute subdural haematoma. Major mortality reduction in patients treated within 4 hours. *New England Journal of Medicine* 1981; **304**: 1511–18. [One of the first studies showing that early surgical treatment of traumatic haematomas improves outcome.]

Siesjo J. Basic mechanisms of traumatic brain damage. *Annals of Emergency Medicine* 1993; **22**: 959–9.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. Lancet 1974; **ii**: 81–4. [The original description of the Glasgow coma scale.]

Teasdale G *et al.* Risks of an acute intracranial haematoma in children and adults: implications for managing head injury. *British Medical Journal* 1990; **300**: 363–7.

Teasdale GM, Thornhill S, Murray GD, McGillivray C, McLeod D, McEwen JD. How great is the burden of disability after hospital admission for head injury? Proceedings of the 132rd meeting of the

Society of British Neurological Surgeons. *British Journal of Neurosurgery* 1999; **13**: 235–45.

Temkin NR, Dikmen SS, Wilensky AJ, Keihm J, Chabal S, Winn HR. A randomised double blind study of phenytoin for the prevention of posttraumatic seizures. *New England Journal of Medicine* 1990; **323**(8): 497–502.

Ware JE. Measuring patients' views: the optimum outcome measure. SF 36: a valid reliable assessment of health from the patient's point of view. *British Medical Journal* 1993; **306**: 1429–30.

Wilson JTL *et al.* Early and late magnetic resonance imaging and neuropsychological outcome after head injury. *Journal of Neurology, Neurosurgery and Psychiatry* 1988; **51**: 391–6.

Working Party of the Neuroanaesthesia Society and Association of Anaesthetists. *Recommendations for the transfer of patients with acute head injuries to neurosurgical units*. The Neuroanaesthesia Society of Great Britain and Ireland and the Association of Anaesthetists of Great Britain and Ireland, London,1996.

Zusho H. Post-traumatic anosmia. *Acta Otolaryngologica* 1982; **108**: 90–2.

44.3 Intracranial tumours

Jeffrey J. Olson, Daniel L. Barrow, Mark R. Gilbert and Nelson Oyesiku

Tumors of the nervous system
Introduction
Primary brain tumors
Glioblastoma multiforme and anaplastic astrocytoma
Astrocytoma
Oligodendroglioma and anaplastic oligodendroglioma
Ependymoma
Medulloblastoma
Neuronal and mixed neuronal and glial tumors: ganglioglioma and neurocytoma
Tumors of the skull base that arise separately from the nervous system
General comments
Chordoma
Esthesioneuroblastoma
Nasopharyngeal carcinoma
Glomus tumor
Neurofibromas and schwannomas
Neurofibromatosis
Diagnosis
Treatment
Pineal tumors
Introduction
Clinical presentation
Diagnosis
Tumor types
Treatment
Meningiomas and choroid plexus tumors
Meningiomas
Choroid plexus tumors
Pituitary tumors
Clinical problems
Diagnosis and treatment
Endocrine-inactive tumors
Prolactinomas
Acromegaly
Cushing disease
Gonadrotropin-secreting adenomas
TSH-secreting adenomas
Future directions in pituitary surgery
Intracranial epidermoid tumors
Craniopharyngioma
Brain metastases
Introduction
Diagnostic evaluations
Treatment
Outcome
Primary lymphoma of the central nervous system
Introduction
Diagnosis
Management
Further reading

Tumors of the nervous system

Introduction

General classification

Neoplasms of the nervous system are a very heterogeneous group of diseases. Though often grouped together in one discussion, the range of surgical techniques used to deal safely and effectively with the variety of tissue consistencies, anatomic locations, and adjacent tissues is necessarily extensive and provides some of the greatest challenges in neurosurgery. The World Health Organization (**WHO**) classification allows one to see this more directly. The summary in Table 1 demonstrates the various types and sites of origin of these lesions.Table 1 New information on cytogenetics, molecular biology, and proliferative capacity offers to extend our appreciation of the differences in the origin and progression of brain tumors. Though many other classifications are available, the discussions below will rely primarily on the WHO system. Consistent usage of classifications allows comparison between reports of different series and coherent discussions of lesions between clinicians.

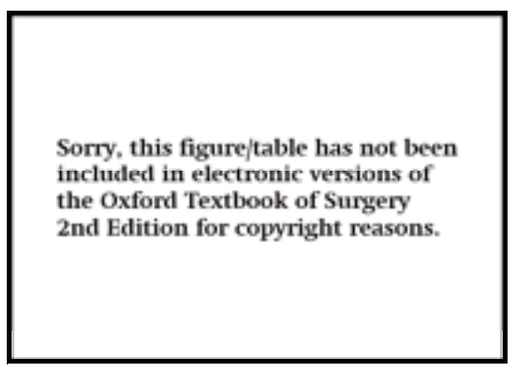


Table 1 Histologic classification of tumors of the central nervous system

The WHO classification refers to primary brain tumors as those that arise from the cells of the central nervous system. Brain tumors are a small component of cancer overall, with an autopsy frequency of 1 to 2 per cent. The neuroepithelial tumors are those of the glia and neurones; they account for 50 to 60 per cent of primary intracranial tumors in adults. Of these, the relative incidence of tumor types is: glioblastoma 50 per cent, anaplastic astrocytoma 30 per cent, oligodendroglioma 6 per cent, nonanaplastic astrocytoma 5 per cent, ependymal tumor 4 per cent, medulloblastoma 2 per cent; the neurocytomas, pineal-cell tumors, subependymal giant-cell tumors, pilocytic astrocytomas, and choroid plexus tumors account for 1 per cent or less each. Meningiomas comprise about 20 per cent of intracranial tumors. Overall, metastases account for approximately one-half of all intracranial neoplasms.

Initial management

More often than not, the individual with a tumor affecting the nervous system presents to the physician they consider their primary caregiver, be they a family physician,

internist, or surgeon. The majority present with new neurologic deficits, changes in mental status, headaches, or deformities that are gradual in onset, often over days to months, as is described in greater detail below. An example of a more urgent presentation would in the emergency room after a new seizure. The initial steps in the management of these patients are important. In each of them a simple neurologic exam will guide further evaluation. Imaging studies of the appropriate portion of the central nervous system are based on the findings of the physical examination and history. Where obtundation, rapidly progressive deficits, or difficulty in controlling seizures are present, admission to hospital is necessary. If images show a significant mass effect and this appears to be due to vasogenic edema, corticosteroids, usually in the form intravenous dexamethasone, are begun. H_2-blockers are started prophylactically. When seizures are present, anticonvulsants such as phenytoin or phosphenytoin are begun. Where the individual cannot drink or eat, intravenous fluids with half-normal saline are necessary. Free water or dextrose in water is not advisable, as these may enhance brain edema. The majority of patients are neurologically stable and their admission is unnecessary. Oral corticosteroids, H_2-blockers, and anticonvulsants can be used in preference to intravenous preparations. Whether or not admission is required, a key component in early management is to find a neurosurgeon or neurologist willing to manage the patient's immediate problems and plan methods of diagnosis and treatment.

Aspects of clinical presentation

Because of the diversity of these lesions, blanket statements about their signs and symptoms cannot be made. In general, though, symptoms arise from malfunction of that part of the nervous system where the tumor is arising or from compression of adjacent structures. The common general symptoms at presentation include changes in mental status, headaches, vertigo, dizziness, and alteration of level of consciousness. Changes in mental status, such as altered concentration, memory, affect, personality, initiative, abstract reasoning, and confusion, are often indolent and may occur in 15 to 20 per cent of patients with brain tumors at presentation. Other symptoms, such as weakness, sensory changes, visual abnormalities, aphasias, gait instability and focal seizure activity, are examples of symptoms that have localizing value. Approximately 10 to 20 per cent of adults with the new onset of seizures have brain tumors.

Headache is not universal in those with brain tumors. Contrary to popular expectation, only 70 per cent of patients with primary brain tumors will have a headache of some sort at the time of diagnosis: in 35 per cent it the first symptom; only about 15 per cent have all of the usual characteristics of headache supposedly related to raised intracranial pressure, such as increased intensity in the morning, association with nausea, exacerbation by postural change, and pain that arouses one from sleep.

General diagnostic methods

The clinical examination yields clues to the localization of a lesion within the central nervous system but falls short of providing the detailed information necessary to plan a surgical approach. Visualization of central nervous tumors is centered around computed tomography (**CT**) and magnetic resonance imaging (**MRI**). Where the presentation is a precipitous decline in neurologic status, suggesting acute hydrocephalus or intracranial hemorrhage, a CT scan is indicated. Once these patients have been stabilized, and in most others who have a more gradual onset of symptoms, MRI without and with contrast-enhancing agents will yield the most information about the lesion. The multiplanar representation of the lesion seen in various imaging sequences allows the surgeon to appreciate the tumor mass, its surrounding edema, and its relation to normal structures.

Angiography can be of use in selected circumstances: these include preoperative evaluation of tumor vasculature, the assessment of normal arterial and venous anatomy adjacent to or within the tumor, preoperative embolization, and ruling out the presence of arteriovenous malformations and aneurysms in patients presenting with hemorrhage. Other diagnostic techniques, such as electroencephalography, positron-emission tomography and nuclear brain scanning, do not often assist the surgeon in determining how to proceed. Although imaging studies are of great value in localizing the lesion, actual diagnosis is dependent on obtaining adequate tissue for histologic evaluation. Studies that can be diagnostic include cytology of cerebrospinal fluid and endocrine evaluation, which may be diagnostic in carcinomatosis metastatic to the central nervous system and pituitary adenoma, respectively.

Primary brain tumors

In adults the most common form of primary brain tumor, known as the glioblastoma multiforme, is also the most malignant. Less commonly encountered is the anaplastic astrocytoma, which is not as aggressive but is still ultimately lethal. The astrocytoma and oligodendroglioma are more indolent, but could be considered benign with malignant potential as they often go on to devastating progression years after diagnosis. The less common histologic types must be recognized, as some have extraordinarily benign courses while others require aggressive intervention to obtain even temporary control.

The discussions of the tumors below include aspects of management beyond surgery alone. Because many brain tumors require surgical intervention or at least surgical decision-making, a neurosurgeon will often supervise the overall care of these patients. Thus, the neurosurgeon must be familiar with all the aspects of their care.

Glioblastoma multiforme and anaplastic astrocytoma

Clinical presentation

The glioblastoma multiforme is generally a rapidly growing lesion that makes itself known by causing malfunction of the central nervous system at its site of origin and, because it can attain significant size rapidly, by compression of surrounding normal tissue. The less common anaplastic astrocytoma may be somewhat slower in its proliferation, but often has a similar set of presenting characteristics. In contrast to the neurologic deficits from processes such as ischemia, the onset of which is relatively sudden, those from brain tumors tend to be gradually progressive. Fewer than 20 per cent of patients have had symptoms for less than 1 month and fewer than 10 per cent for longer than a year. Symptoms of increased intracranial pressure, including lethargy, confusion, headache, nausea and vomiting, may be the first presentation. On neurologic examination the site of origin can sometimes be discerned. For example, a lesion arising in the left frontal lobe may be associated with expressive aphasia and a right hemiparesis. A lesion of the temporal lobe may present with seizure activity and a superior quadrantanopia. Parietal lesions can be subtle, with only minor cognitive abnormalities and contralateral sensory changes. Occipital lesions are almost always associated with some degree of deficit in the visual field, even if not appreciated by the patient. Lesions of the diencephalon, brainstem, and cerebellum may be symptomatic when small because of the compact nature of the functionally important neurones and fibers in these regions. Increased intracranial pressure can be discerned by ophthalmoscopic examination to determine whether or not papilledema is present. Seizures are an occasional mode of presentation in these more malignant tumors, but not so frequently as with less aggressive types of the same family of tumors. These symptoms and signs are not pathognomonic for brain tumors, but rather signify the presence of any destructive or mass-producing intracranial lesion. One must keep in mind that, in the appropriate clinical setting, brain abscess, inflammatory disease, or even ischemia can produce such a presentation.

Diagnosis

The clinical examination is an interesting and rewarding exercise in neurologic localization but will not necessarily be diagnostic by itself. In a patient with objective signs of neurologic changes and/or increased intracranial pressure, MRI is the optimal method for characterizing the anatomy of the lesion. The glioblastoma multiforme tends to have the more bizarre appearance, with irregularly enhancing margins, cystic and necrotic areas associated with mass effect, and varying amounts of edema (Fig. 1). Anaplastic astrocytomas can have any of these characteristics but are, in general, more homogeneous in appearance. Importantly, it is not uncommon to see an anaplastic astrocytoma with no contrast enhancement (Fig. 2). Thus enhancement by itself is not the essential feature of malignancy. As mentioned above, some individuals will present with sudden changes in consciousness, or new seizure activity. In these cases CT will be the most useful in determining if the process needs emergency surgical or medical intervention. Angiography, electroencephalography, and imaging based on radionuclides rarely provide information that alters the surgical approach or subsequent therapy. Lumbar puncture is rarely diagnostic and, especially when increased intracranial pressure is shown by imaging, may be dangerous. Ultimately, diagnosis is based on histologic evaluation, in which the WHO system serves as the standard. The presence of necrosis differentiates glioblastoma from anaplastic astrocytoma. In recurrent masses in patients who have had radiation therapy, positron-emission tomography may assist in differentiating tumor recurrence from radiation necrosis. The most common histologic correlate is the presence of necrosis within the lesion, though some grading systems outside of the standard WHO system may allow this diagnosis to be made without the presence of necrosis. When there is mesenchymal differentiation within the lesion, the tumor may be designated a gliosarcoma (Fig. 3); these sometimes may actually follow a more aggressive course than glioblastoma.

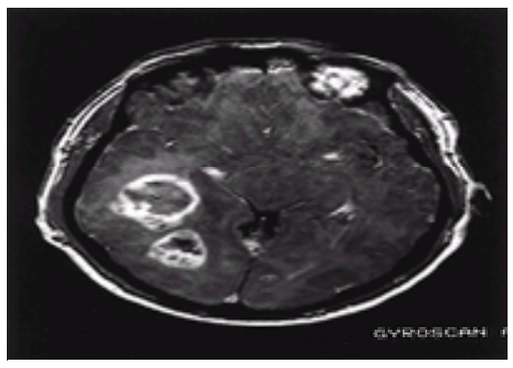


Fig. 1. Glioblastoma multiforme. Axial T_1-weighted gadolinium-enhanced MRI showing a right temporo-occipital lesion with irregular rim enhancement, central hypointensity suggestive of necrosis, and edema extending anteriorly into the temporal lobe. The two apparent components of this tumor on this image merged lower in the temporal lobe.

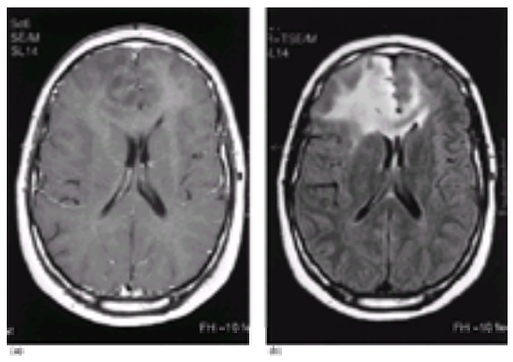


Fig. 2. Anaplastic astrocytoma. (a) Axial T_1-weighted gadolinium-enhanced MRI demonstrating a right mesial frontal mass with faint gyriform enhancement and decreased signal intensity of the associated gyral white-matter tracts. (b) Axial FLAIR (TR 7000,TE 130) MRI better demonstrating the edema and infiltrative nature of this lesion, with involvement of not only the corpus callosum but also the left frontal lobe.

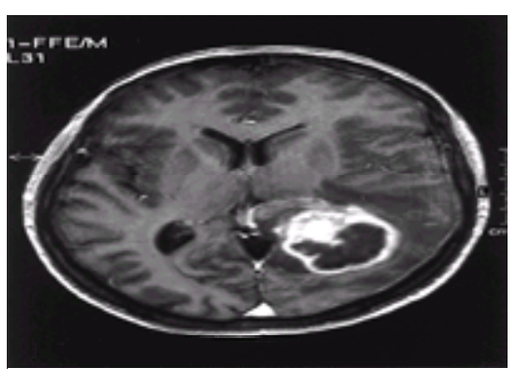


Fig. 3. Gliosarcoma. Axial T_1-weighted gadolinium-enhanced MRI of a left parieto-occipital tumor with mass effect on the corpus callosum and obliteration of the normal contours of the lateral ventricle on that side. There is an irregular rim of enhancement with decreased central signal, suggestive of extensive necrosis. The radiographic appearance is similar to that of glioblastoma.

Treatment

Though various therapies are available for these difficult lesions, it is important for the surgeon to recognize that the most important prognostic factors for survival are the histology of the tumor, the patient's age, and their performance status at the time of diagnosis. Knowing this helps the surgeon keep the risks and benefits of the proposed therapies in proper perspective.

Upon finding an intracranial lesion that one suspects is a tumor it is the surgeon's responsibility to obtain a diagnosis and, when necessary and deemed safe, decompress surrounding normal tissue. Small lesions without mass effect and lesions in the thalamus, hypothalamus, and upper brainstem can usually be diagnosed safely with stereotactic methods. The needle-biopsy specimens can be informative, with the key to success with these procedures being an experienced neuropathologist. Lesions compressing the surrounding normal tissue that contain significant cystic fluid, necrotic material, or solid tumor can benefit from craniotomy and resection. It is with the resection of the margins, where visually imperceptible infiltration into the normal tissue occurs, that new neurologic deficits are most likely to be induced. Measures to maximize the accuracy of tumor identification and avoid entering the adjacent normal tissue include the use of awake craniotomy, preoperative electrocorticography, intraoperative sensory-evoked potentials for mapping, image-guided surgical techniques, and frame-based stereotactics. Standard bipolar cautery and suction account for most of the resection in these lesions, but speed can be added by the judicious use of ultrasonic dissection devices. It is important to maintain hemostasis as the resection proceeds and to confirm it upon completion of the resection of all tissues that can be clearly identified as tumor. The surgeon must keep in mind that cure is not possible with surgery alone. Though, under selected circumstances, outcome is associated with the degree of resection, the primary goal of surgery is to obtain diagnosis and decompression.

External radiation is the single most effective form of therapy for these lesions. Multifractioned X-irradiation in fractions of 1.8 to 2 Gy/day to a total dose of approximately 60 Gy to primary region of the tumor can usually be offered to all individuals with this diagnosis. Toxicity does occur and needs to be balanced with prognosis. Radionuclide brachytherapy and stereotactic radiation can be used as additional methods of treatment for recurrent tumors.

Cytotoxic chemotherapy has a role in the treatment of these tumors, particularly in newly diagnosed cases in younger individuals with good performance status. No single agent has demonstrated anything near curative potential, but **BCNU** (carmustine; 1,3-bis(2-chloroethyl)-1-nitrosourea) and cisplatin have reliable, though small, response profiles in glioblastoma multiforme. For anaplastic astrocytoma a combination of procarbazine, **CCNU** (lomustine; 1-(2-chloroethyl)-3-cyclohexyl-nitrosourea), and vincristine has been advocated, but prospective data suggest that the ability of this combination to add value to radiation may be limited. In selected cases these agents may have a greater impact when given before, rather than during or after, radiation. On recurrence, the implantation of biodegradable polymers that release BCNU into a tumor bed may augment the value of surgical debulking.

Because these standard forms of therapy are associated with poor prognosis, this group of patients has provided fertile ground for investigational treatment agents and routes of administration. A wide variety of alternative cytotoxic agents has been evaluated, demonstrating no clear advantage over the agents mentioned above. Treatments based on molecular biologic information and using immunologic agents, biologic-response modifiers, RNA (oligonucleotides), and DNA (genes or fragments thereof) are under active study. It is encouraging that patients with glioblastoma had a three times greater 5-year survival (12 per cent compared to 4.5 per cent) when treated as part of an investigative rather than standard protocol.

It must be kept in mind that these lesions are frequently heterogeneous histologically, with some areas appearing relatively low grade while others are completely consistent with glioblastoma multiforme. It is recommended that the lesion be treated as indicated by its most aggressive-appearing portion.

The surgeon must become an advocate for the patient's best interests. Knowing the patient's age, overall health, and the tumor's location and probable histologic type,

the surgeon can inform the patient and the family about treatment options and the likelihood of side-effects and success.

Upon completion of treatment, vigilance is necessary to detect recurrence or progression. Neurologic examination and imaging, usually MRI without and with gadolinium-based contrast agents, should be performed at regular intervals.

Astrocytoma

Clinical presentation

The growth pattern of these lesions tends to be more indolent than that of their more malignant counterparts. Because of this, the brain may be more tolerant of their presence and this alters their presentation. Rather than symptomatic increases in intracranial pressure and focal neurologic deficits, generalized or focal seizures are a much more common form of presentation. Seizure activity may antedate the clinical diagnosis by months or years. Gradually progressive cognitive changes can also develop and signal the presence of astrocytoma.

Diagnosis

In the presentation with new seizure activity or neurologic deficits, MRI without and with contrast enhancement is the most sensitive method for detecting this tumor. Frequently, the lesion does not enhance with contrast administration; in fact, the best imaging is often with T_2-weighted sequences (Fig. 4). CT is of less value, and usually shows an ill-defined lesion without contrast enhancement. Confirmatory imaging studies, including positron-emission tomography, may be of assistance.

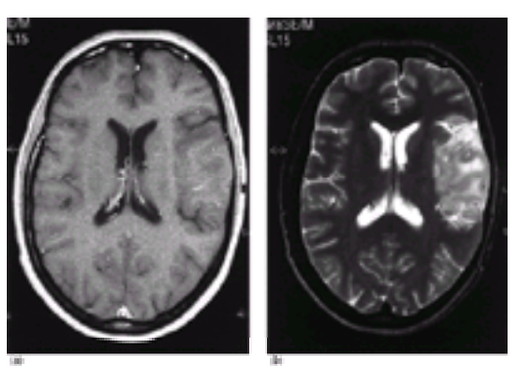


Fig. 4. Astrocytoma. (a) Axial T_1-weighted gadolinium-enhanced MRI of a left frontotemporal tumor. There is no contrast enhancement or mass effect. The cortex of the sylvian fissure and insula have decreased signal intensity compared to the right side. (b) T_2-weighted axial image of the same tumor with a well demarcated area of increased signal in the lateral insula and temporal lobe.

As with all tumors, diagnosis hinges on accurate histologic analysis of properly representative specimens. The lesions can be recognized by the presence of an increased number of cells, often with increased size but with only minor nuclear atypia. A general increase in staining for glial fibrillary acid protein may be demonstrable immunohistochemically in the background of these lesions. Three histologic patterns of astrocytoma may be seen, termed fibrillary, gemistocytic, and protoplasmic. Based on autopsy and retrospective clinical series, as many as half of all astrocytomas may have a malignant or anaplastic component at the time of diagnosis. Two-thirds will go on to recur, with clinical behavior and a pathologic appearance consistent with malignancy. The molecular events associated with this progression are under active investigation and appear to include loss of chromosome 19q or 10. Additionally, loss, alteration, or amplification of platelet-derived growth factor, retinoblastoma protein, and even epidermal growth factor may be steps on the way to anaplastic change in an astrocytoma.

The juvenile pilocytic astrocytoma is not commonly seen in adults, and when found is often associated with good prognosis. Astrocytomas may be mixed in the sense that they can have a combination of astrocytic and oligodendroglial cells. In mixed tumors, prognosis is dependent on the most aggressive component. Variants are recognized, such as the pleomorphic xanthoastrocytoma, a tumor usually occurring superficially in the cerebrum, with pleomorphic neoplastic astrocytes, some of which have lipid inclusions. Though the cells of the pleomorphic xanthoastrocytoma appear aggressive, the prognosis is better than for anaplastic astrocytoma. The subependymal giant-cell astrocytoma arises almost exclusively in tuberous sclerosis.

Treatment

Although these lesions may be discovered by imaging, a tissue diagnosis is still obligatory in most cases for the most accurate possible prediction of prognosis. No prospective analysis of the extent of surgical resection is available to guide the surgeon in making a decision to biopsy, resect subtotally, or attempt to resect completely. Thus the surgeon's judgement and experience are essential in these decisions. In lesions involving functionally important areas it is reasonable to use stereotactic biopsy to obtain a diagnosis and proceed with a treatment plan. For lesions with mass effect and in more quiescent regions of the brain, more aggressive resection can be entertained. Complete resection of small lesions seems to be associated with better tumor control. Difficult-to-control seizure activity can also be improved with resection. Whether in eloquent or silent areas of the brain, accuracy of resection can be augmented with stereotactic techniques, electrophysiologic mapping, and electrocorticography.

When a significant amount of tumor is left behind after biopsy or attempted resection, radiation therapy may be added to control the residue, but there is a risk of inducing cognitive deficits and an alternative is to delay radiation until there is objective imaging evidence of tumor progression. Where surgical resection is demonstrably complete on postoperative imaging, one option is to withhold radiation until there is objective imaging evidence of tumor progression. In a prospective study the timing of the addition of radiation, either at the time of diagnosis or of progression, did not seem to affect survival. That study, as well as retrospective reviews, provide evidence that radiation early after subtotal resection may prolong survival. Additional histologic investigation, such as on the tumor's proliferative capacity as measured by 5-bromodeoxyuridine or Ki-67 labeling, may detect more aggressive tumors and may justify the use of radiation early after incomplete resection. Chemotherapy has been investigated in the treatment of these lesions, but it has been difficult to show any benefit from its addition and it is not considered standard care. Agents under investigation that induce tumor differentiation may ultimately prove to be useful alternatives to standard cytotoxic chemotherapy.

In some of the more unusual variants, such as pleomorphic xanthoastrocytoma or subependymal giant-cell tumor, surgery with the goal of complete resection is the primary therapy, with repeat resection or radiation the options for treatment of any recurrence.

Upon completion of the initial phase of treatment, vigilance is necessary for recurrence or progression. Neurologic examination and imaging, usually MRI without and with gadolinium-based contrast agents, should be performed at regular intervals. Dedifferentiation can be expected at the time of progression. Surgery may be indicated for decompression and to determine whether the lesion has taken on the characteristics of an anaplastic astrocytoma or glioblastoma, thus guiding further treatment planning.

Oligodendroglioma and anaplastic oligodendroglioma

Clinical presentation

The oligodendroglioma arises from its myelin-producing namesake. Though of biologic interest, this fact does not seem to alter the presentation much from that of an astrocytoma. Oligodendrogliomas tend to be lesions of the frontal lobes and of younger adults (peak incidence, 26 to 46 years of age). They present more frequently with seizures and headaches than do astrocytomas. Many of these lesions may have been present for a significant period of time before becoming symptomatic. Therefore, due to the plasticity of the brain, they can occasionally attain large size without the symptoms of increased intracranial pressure. Ultimately, cognitive fall-off will occur, and headache or other focal deficits will be the reason for seeking medical attention in at least 50 per cent of patients. In some instances, possibly due to the neoplastic capillary bed, presentation may be with spontaneous intracranial hemorrhage.

The anaplastic counterpart of the oligodendroglioma can present much like the anaplastic astrocytoma. Symptoms of increased intracranial pressure are more likely to be seen here. In general, the onset of symptoms is faster than with its less aggressive counterpart. Also, focal neurologic deficits, rather than just seizures or changes

in mental status, will be a part of presentation in many patients.

Diagnosis

The clinical presentation of new seizures without obvious cause in an adult justifies the use of sophisticated imaging. MRI will show these lesions well, particularly those with little or no contrast enhancement. The T_2-weighted images tend to have well-delineated margins, implying that some of these lesions may be less infiltrative than their astrocytic counterparts. The oligodendroglioma has a significant propensity for calcification; this can be seen well in a CT scan, but, of itself, calcification does not differentiate oligodendroglioma from astrocytoma and therefore a search for this finding alone does not justify obtaining a CT.

A nidus of contrast enhancement in a lesion with otherwise homogeneous, nonenhancing characteristics suggests an oligodendroglioma with an anaplastic component. Cystic components as well as calcification can be seen here.

Histologic analysis of oligodendrogliomas shows monotonous fields of cells somewhat larger than normal oligodendrogliocytes, with round nuclei. Formalin-fixed tissue undergoes acute swelling and dissolution of cytoplasmic contents giving an artefactual honey-comb appearance, sometimes thought of as characteristic of this tumor type. Microcalcifications and a branching capillary network are often seen. Anaplastic lesions have a more heterogeneous cell population with frequent mitotic figures. Not uncommonly, oligodendrogliomas may be mixed with astrocytic tumor cells, termed oligoastrocytoma or mixed glioma.

Seizures are an important component of the natural history. Once the lesion has been detected by imaging, the surgeon learns little else by the addition of electroencephalography. On the other hand, patients with known residual oligodendroglioma after initial treatment who are experiencing poor seizure control may benefit from electroencephalography. Identification of a residual seizure focus in or adjacent to the tumor may indicate the need for additional resection, possibly with intraoperative electrocorticography, with the goal being seizure control.

Treatment

It has been suggested that the central component of an oligodendroglioma is more purely tumorous than that of an astrocytoma, whose infiltrative nature is more likely to encompass some normal brain. Because of this, homogeneous, nonenhancing brain lesions in safely approachable areas, consistent with oligodendroglioma, may be good candidates for surgical resection. Complete resection of these lesions, when possible, results in improved long-term control. As with the astrocytoma, a small deep lesion may best be diagnosed by stereotactic biopsy. The addition of radiation at initial diagnosis or on progression is an option that should be decided from the aggressiveness of the presentation, the amount of residual tumor, and possibly the proliferative index. When the histologic appearance is of oligodendroglioma without anaplasia and total resection is confirmed by postoperative images, careful observation with regularly spaced MRI may be warranted as follow-up management. Tumors known to be anaplastic oligodendroglioma always warrant radiation therapy.

The anaplastic oligodendroglioma does seem to have a greater propensity to grow. When this diagnosis is obtained, the use of standard or intensive doses of procarbazine, CCNU, and vincristine in addition to radiation has been advocated. Their use prior to radiation has been proposed, and documented responses have occurred. These lesions are also sensitive to radiation and this is recommended as a standard part of their treatment, whether or not chemotherapy is added.

Upon completion of treatment, vigilance is necessary for recurrence or progression. Neurologic examination and imaging, usually MRI without and with gadolinium-based contrast agents, should be performed at regular intervals.

Ependymoma

Clinical presentation

These lesions arise from the ependymal cells lining the cerebrospinal fluid-containing spaces in the central nervous system. They are relatively uncommon, accounting for about 5 per cent of tumors in the adult central nervous system and 10 per cent in children. Approximately half of these lesions develop in the first two decades of life. The majority occur in the cranium, predominantly in the posterior fossa in children and supratentorially in adults, but they are certainly seen also in the spinal canal.

Focal neurologic deficits related to the location of the lesion may occur. There may be extension into normal periventricular tissue but it is important to recognize that many of these lesions frequently extend into the ventricular system, obstructing cerebrospinal fluid dynamics and causing symptoms of hydrocephalus and increased intracranial pressure. Infratentorially, in addition to symptoms of hydrocephalus, there may be new cranial-nerve palsies or cerebellar dysfunction. Ependymoma can seed the subarachnoid space, both intracranially and along the spinal canal, and present with symptoms localizing to sites remote from that of its origin.

Primary spinal lesions, or lesions that develop by seeding from cranial lesions, tend to have indolent myelopathies. These are often painless presentations including gait instability, incoordination in the upper or lower extremities, and sensory alterations that do not follow strict dermatomal patterns. Some are associated with syrinx formation above and below the lesion, whereby the patients may have suspended sensory losses. Hyper-reflexia or clonus may be detected, as well as up-going toes to various maneuvers. Ependymoma in the spine, particularly in the cauda equina, may be associated with pain. If the neurologic examination demonstrates any more than expected from a simple radiculopathy, then imaging studies are warranted to assess the possible presence of this neoplasm.

Diagnosis

As with other primary tumors of the central nervous parenchyma, MRI is the most sensitive method for detecting any abnormality. Symptomatic lesions can be rather small when sited near the foramen of Monroe, aqueduct of Sylvius, or foramen magnum. Any extension into the ventricular system suggests that ependymoma should be placed higher in the differential diagnosis of the lesion. Once the diagnosis is made, MRI of the spinal canal, and lumbar puncture if necessary, are indicated to look for leptomeningeal spread.

On MRI, intraspinal ependymomas tend to show a very focal expansion of the spinal cord (one to two segments) and the enhancement will be fairly intense and homogeneous (Fig. 5). In patients over 20 years of age the probability that a spinal intramedullary tumor with these findings will be an ependymoma is about 60 per cent. A more diffuse widening, with or without cystic changes and heterogeneous enhancement, suggests astrocytoma. There may be extension proximally and distally through the central canal. The myxopapillary variant of ependymoma can attain huge proportions and encompass the entire cauda equina before detection. Often, intraoperative inspection reveals that these tumors are actually arising from the conus or filum terminale.



Fig. 5. Spinal ependymoma. Sagittal T_1-weighted gadolinium-enhanced image of the lumbar spine: an enhancing lesion occupying the majority of the spinal canal from L1 to L4 is associated with a cystic component in its rostral end. At surgery this lesion was found to arise at the conus and followed the filum terminale.

Treatment

Small intracranial lesions can be diagnosed by stereotactic biopsy. More often, symptoms from hydrocephalus and mass effect are present; craniotomy is then indicated for diagnosis and resection. Often the lesion can be approached through an adjacent sulcus in a functionally less important region. The site of origin must be visualized in order to attain total resection. When the lesion arises from the floor of the fourth ventricle, some may be left behind if differentiating between tumor and

normal parenchyma proves difficult and there is a chance of causing mechanical damage to the brainstem. Local radiation therapy is nearly always used as an adjuvant, with or without total resection. If histologic evidence of anaplasia is present, radiation is more strongly advised. When tumor regrowth occurs, chemotherapy with BCNU, dibromodulcitol, ifosfamide, VP-16, or platinum-based compounds has been described, though no standard regimen with a reliable response rate has been developed.

The first choice in the treatment of leptomeningeal spread is craniospinal irradiation; this is a particularly difficult option in children because of its extensive toxicity. Chemotherapy, either systemic or intrathecal, has been used in children with this form of the disease, but again, no standard regimen has been clearly established.

Primary intramedullary ependymomas of the spinal cord can be approached surgically, both for diagnosis and for decompression of the surrounding normal tissue. The surgical technique involves identification of the site where the lesion has become most superficial and its exposure via a midline myelotomy. Microscopic assistance is essential, and monitoring of somatosensory or motor-evoked potentials helps in identifying, but does not guarantee the preservation of, the functional integrity of the cord during the procedure. Piecemeal resection of these lesions from the inside out minimizes the manipulation and traction of normal tissue. Microscopic extensions of the tumor rostrally or caudally, or poorly defined extensions into the spinal cord, may not be resectable, requiring the addition of radiation in the early postoperative period or at the time of recurrence. The myxopapillary variant of these tumors, with its characteristic myxoid and mucinous background, is benign in character, found primarily in the conus, filum and cauda equina, and best treated by total surgical resection. Also, the subependymoma, usually seen in the fourth ventricle, is a low-grade lesion that, once recognized histologically, is treated surgically without additional radiation (Fig. 6).

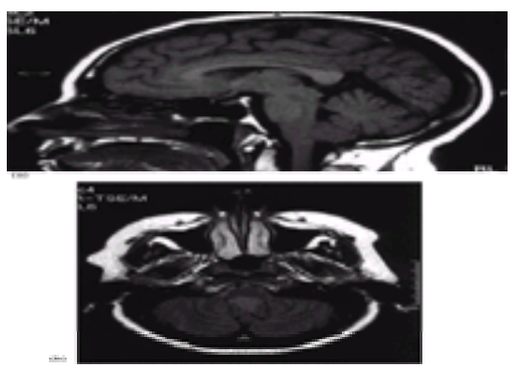


Fig. 6. Subependymoma. (a) Sagittal T_1-weighted MRI demonstrating a mass at the caudal end of the fourth ventricle extending to the top of the cisterna magna with mild mass effect on the medulla and the inferior vermis. (b) Axial T_1-weighted gadolinium-enhanced MRI of a minimally enhancing lesion in the caudal fourth ventricle; note the relatively clear delineation of the tumor from the floor of the fourth ventricle.

Medulloblastoma

Clinical presentation

Seventy per cent of medulloblastomas occur in individuals younger than 16 years of age. The peak occurrence is at 7 years of age. In adults, 80 per cent develop between 21 and 40 years of age, and more often in men than women. Medulloblastoma is the most common and best characterized of the primitive neuroectodermal tumors. It is most often a lesion in the midline of the posterior fossa, involving the vermis, and inducing symptoms of increased intracranial pressure, such as headaches, nausea, occasional vomiting, and lethargy, by its shear size and sometimes from hydrocephalus. Also, symptoms due to cranial-nerve, brainstem or cerebellar malfunction, such as diplopia, altered facial motion and sensation, difficulty with swallowing, ataxia or dysmetria, may occur. Because of their malignant nature, medulloblastomas tend to show symptoms that have been present over a period of only weeks to months.

Diagnosis

In patients with symptoms of cerebellar, cranial-nerve, and brainstem malfunction and/or increased intracranial pressure suggesting disease of the posterior fossa, MRI with and without gadolinium or CT with and without contrast are indicated. Medulloblastomas are generally solid, intensely enhancing lesions, in or adjacent to the vermis, and compressing the surrounding normal tissue. As the age of patient at presentation increases, there is the likelihood of greater involvement of the cerebellar hemispheres. Images confirm the presence of a lesion in the posterior fossa that appears to be relatively separate from the cerebellum itself. Additionally, approximately one-third of patients will have nodular or diffuse leptomeningeal and ependymal spread at the time of diagnosis. Because of this, proper treatment planning requires that all patients with this diagnosis should have spinal imaging to look for leptomeningeal spread and cytologic sampling of cerebrospinal fluid if the imaging is negative. In younger patients, systemic involvement is not uncommon and bone-marrow studies are indicated.

Treatment

Surgery for diagnosis and the relief of intracranial pressure by tumor debulking is the initial step in nearly all patients. At the same sitting, the surgeon may choose to place a ventriculostomy to avoid intra- or postoperative intracranial hypertension or hydrocephalus. Fortunately, in those with complete resections, less than one-third will need permanent ventriculoperitoneal shunts. Though radical resections can be accomplished, surgical removal is not curative. In adults, once the wound is well healed, radiation therapy supplemented with chemotherapy can be added. Regular follow-up with neurologic examination and imaging is necessary in view of the propensity of these lesions to recur.

Considerable experience has been gained with these tumors in children. Essential to their treatment is the use of chemotherapy alone in children of less than 2 years of age with good risk. A number of agents have been used in this effort, with no single regimen being superior in every clinical situation. Therefore, the enrolment of such patients into research protocols where available will speed the delineation of the best regimens. Radiation is delayed if possible in order to protect the immature nervous system from its toxicity. Radiation is used more aggressively, in doses on par with those for malignant gliomas in older children and adults.

Neuronal and mixed neuronal and glial tumors: ganglioglioma and neurocytoma

Clinical presentation

Presentation tends to be nonspecific, but a significant proportion of gangliogliomas present with seizure activity, either generalized or focal, with symptoms appropriate to the site of origin. The epilepsy may be new or long standing. The neurocytoma is more likely to be associated with headache and symptoms of increased intracranial pressure. The neurocytoma and ganglioglioma occur most often in those below 30 years of age. These lesions generally grow slowly and may also be associated with gradual alterations in mental status appreciated only in retrospect.

Diagnosis

An individual with new onset of seizures or an unexplained fall off in cognitive abilities should undergo a neurologic examination to determine if any localizing abnormalities are present; this should be followed by MRI with and without gadolinium to evaluate the site of the abnormality. Neurocytomas will usually be associated with the lateral or third ventricles, often deforming the septum pellucidum. When in this location they are sometimes called central neurocytomas. Those that occur away from the ventricles may be called cerebral neurocytomas (Fig. 7). These lesions are usually homogeneous in density and may not enhance.



Fig. 7. Cerebral neurocytoma. Axial T_1-weighted gadolinium-enhanced MRI of a left frontopolar tumor with mild homogeneous enhancement. A small amount of mass effect is present on the ipsilateral ventricle. No apparent association with the ventricle is present.

Gangliogliomas tend to be seen best on T_2-weighted images and are most often localized to the temporal or frontal lobes. Some are cystic, and the more aggressive gangliogliomas may enhance more. Within the differential diagnosis are hamartomatous changes: in general, individuals with hamartomatous changes will have long-standing changes in mental status and seizure disorders, though this clinical information is only sometimes of value in separating these entities.

Other less common diagnoses must be kept in mind when neurocytoma and ganglioglioma are suspected. The patient's age and the anatomic location, histologic pattern, immunohistochemical and ultrastructural features of the tumor will differentiate the gangliocytoma, dysplastic gangliocytoma of the cerebellum, desmoplastic infantile ganglioglioma and dysembryoplastic neuroepithelial tumors from the neurocytoma and ganglioglioma. The subependymal giant-cell astrocytoma can have divergent cell differentiation including neuronal development.

Treatment

Suspected neurocytomas that have mass effect and are in safely reachable sites should be approached directly for biopsy and resection; these include lesions with significant extension into the lateral or third ventricles. Those that are small or with extensive involvement of the basal ganglia may be diagnosed by stereotactic biopsy. The ganglioglioma is often without mass effect and therefore stereotactic biopsy is often used. On examination of preliminary or frozen sections the neurocytoma appears to be composed of a homogeneous population of cells. On this basis, and because fine calcification is sometimes present, a preliminary diagnosis of oligodendroglioma may be made. The ganglioglioma will show a more heterogeneous field of cells with greater fibrillary background. The cells of these tumors can be identified as having neuronal, ganglionic, and astrocytic characteristics by immunohistochemistry and electron microscopy. The malignancy of the glial components of the ganglioglioma determines the propensity for recurrence.

Once diagnosis and/or resection has been accomplished, the next step depends on the proliferative nature of the lesions. The neurocytoma tends to grow slowly and can have an excellent prognosis after complete resection, and therefore its postoperative management is often simple observation with regular imaging and neurologic examinations. However, even in the neurocytoma, there can be significant postoperative residue in a tumor with mitotic activity and a high proliferative index as measured by MIB-1. In these there may be a role for more aggressive options, such as early repeat resection or more frequent postoperative imaging to detect progression early. X-irradiation is generally reserved for lesions that fail a second attempt at resection, detected by biopsy at recurrence as having some component of malignant glioma.

Regular electroencephalography may have a role in the chronic management of those with difficult-to-control seizures. In such a patient with ganglioglioma, craniotomy with electrocorticography to guide the resection can assist successfully in improving the seizure disorder. In the ganglioglioma with an aggressive or anaplastic glial component, radiation may be indicated at initial diagnosis.

Tumors of the skull base that arise separately from the nervous system

General comments

The skull base, particularly including the clivus, the petrous portions of the temporal bones, the sphenoid, the ethmoid, and the orbital roofs, serves as an interface between diseases treated by surgeons in other disciplines and those usually the province of the neurosurgeon. The anatomy of this location is heterogeneous in the sense that numerous important structures lie in and adjacent to these bones. Thus the surgeon must be familiar with the location and course of the primary vascular structures, including the internal carotid, vertebral and basilar, and meningeal arteries, the sphenoparietal, superior and inferior petrosal, transverse and sigmoid sinuses, the jugular bulb, and the cavernous sinus. Any of the 12 cranial nerves can become involved in operations at these sites. Additionally, the pituitary gland and its stalk should be respected. The paranasal sinuses, parts of the auditory and vestibular apparatus, and of the visual apparatus, are all found in association with tumors of this location. The preservation of these normal structures, and the maintenance of a reasonable cosmetic appearance but eradicating a tumor, are the primary challenges to working in this area.

Each type of tumor has a particular propensity to arise in certain regions and a characteristic tissue consistency. Approaches to these areas and techniques for their removal vary widely and have been the subject of much innovation. Tumor control is often obtained with multimodal management utilizing not only surgery but also radiation and embolization. The purpose here is to discuss considerations that are common in the care of patients with these lesions rather than provide an exhaustive list of possible approaches in each case.

Chordoma

Clinical presentation

Cranial and spinal chordomas putatively develop from the remnants of the embryonic notochord. Clinically, they arise in the ventral midline along the clivus, upper vertebral bodies, and sacrum (the most common site of origin). They occur most often in the third to fifth decades of life. Forty per cent arise from the clivus and the single most common complaint is headache. Presentation may be with new cranial neuropathy or radiculopathy, depending on the site of tumor origin. Motor symptoms and endocrine involvement are less common. These are unencapsulated lesions that compress and deform neurologic structures but remain separate from the nervous system. They are not separate from bone and are locally invasive. The symptoms, with or without pain, are often slow to develop, though sudden new symptoms can occur after local trauma.

Diagnosis

Though not always obtained, skull films may show abnormal bone density in the area of the clivus. CT will confirm the presence of areas of bone destruction or abnormal bone density extending into the posterior fossa, sella turcica, paranasal sinuses, and/or nasopharynx. The matrix of the lesion is often isodense with the brain. MRI usually provides the most detailed information about the relation of the lesion to the surrounding bony, vascular, and nervous structures such as the sella and cavernous sinus (Fig. 8). In addition to the solid components of the tumor, regions of mucin accumulation and hemorrhage are commonly seen. However, because of the multiple densities in the skull base on MRI, the exact margins cannot always be delineated well until surgery.

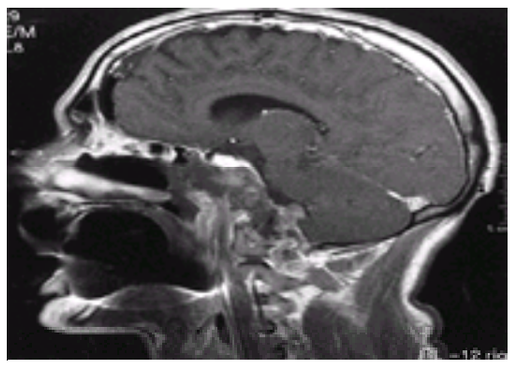


Fig. 8. Chordoma. Parasagittal T_1-weighted gadolinium-enhanced MRI showing a hypointense lesion obliterating the normal contours of the clivus and extending into the posterior fossa and abutting the pons and cerebellum. A portion of the lesion extends into the posterior nasal cavity. The ill-defined nature of the boundaries of this lesion are appreciated on this image.

No single imaging characteristic can differentiate chondrosarcoma from chordoma and often only a histologic sample for will provide the definitive answer. Other tumors that can invade the skull base and are associated with calcific or bony changes include pituitary adenoma, mucinous adenocarcinoma, craniopharyngioma, meningioma, schwannoma, nasopharyngeal carcinoma, and salivary gland tumors. One histologic subtype of the chordoma, termed chondroid chordoma, composed of extensive cartilaginous matrix with smaller foci of chordoma, reportedly has better survival with less locally aggressive behavior.

When the tumor abuts the internal carotid it may be necessary to sacrifice that artery in order to obtain resection or intraoperative hemostasis, so a preoperative angiogram with balloon-test occlusion is useful. Additionally, a preoperative study of the venous anatomy will alert the surgeon to the dominant venous drainage pathways should it become necessary to consider sacrificing veins during the surgical procedure.

Unusually, where an exophytic component of the lesion involves the posterior nasopharynx, the diagnosis may be made via an endoscopic biopsy done by an otolaryngologist.

Treatment

Surgical management remains the primary form of therapy for these lesions; the clearest indications are for lesions with compression of the spinal cord, brainstem, or supratentorial structures. The approach to the lesion depends on the predominant location. Location can be classified by the tumor's primary nidus as upper, middle and lower clival. The majority are best approached anteriorly via extended frontal, transnasal, transmaxillary, or transoral approaches; not infrequently a combination of these must be used to reach an entire lesion; less frequently, a component of the lesion extending laterally can best be approached by a temporal, transpetrous, transcondylar, suboccipital, or retropharyngeal pathway. The tumor will insinuate itself into many of the bony recesses in the skull base, repositioning or enveloping vascular and neurologic structures. Thus microscopic dissection is necessary, with supplemental image-guided techniques if possible. Often the majority of the operation is extradural. Intradural extensions may require removal of a component of a portion of the dura, with repair by grafting and fibrin glue at completion of the resection. Temporary diversion of cerebrospinal fluid may be necessary. In anteriorly extending lesions with no significant alteration of the contours of the spinal canal or posterior fossa, transnasal endoscopic biopsy of the mass can be carried out.

In small lesions with no clearly defined site for needle or endoscopic biopsy, a surgical approach for open biopsy is indicated. Plans for resection must be made at the same time. Accurate knowledge of histologic type is important, as metastatic carcinomas and aggressive primary bone tumors can arise in these areas.

Where observation by scanning is chosen, the surgeon must be careful not to allow too long an interval between scans in the early stages of follow-up. The patient must be alerted to symptoms that may develop which would suggest the need for earlier follow-up imaging. When progression is noted, tissue diagnosis is still necessary even if resection is not used, to confirm the appropriateness of X-irradiation or proton-beam radiation.

The principles of care for spinal chordomas are the same: complete surgical resection when possible, including reconstruction of the spine as necessary, with supplemental irradiation where there is incomplete resection or progression. Metastasis reportedly occurs in up to 30 per cent of spinal chordomas, with widespread sites of involvement.

It is important to obtain images in the early postoperative period to establish a baseline for future reference. Not infrequently, resection of all but the smallest of midline lesions is incomplete, and the residue will be prone to slow but inexorable progression. To deal with this likelihood, nonsurgical options such as standard X-irradiation, stereotactic or proton-beam radiation can then be considered to control the lesion. Conventional radiation, though it does not seem to improve overall survival time, may prolong disease-free survival. Proton-beam radiation, though not commonly available, seems to have the most benefit in these lesions over the long term. Even with surgery and radiation the patient is committed to a lifetime of follow-up in view of this tumor's propensity for recurrence. No reliable chemotherapy regimens have been identified as yet.

Esthesioneuroblastoma

Clinical presentation

This tumor, sometimes referred to as olfactory neuroblastoma, is thought to arise from the olfactory neuroepithelium at the top of the nasal cavity, which explains its primary site of central nervous involvement, the midline of the anterior fossa. The portion of the neoplasm extending into the ethmoid sinuses, frontal sinus, and superior portions of the nasal airway will often present with symptoms of sinusitis. The tumor can also extend superiorly, lifting or transgressing the basal frontal dura mater, and this is where the neurosurgeon's input is most useful. As the compression usually occurs at the base of the frontal lobes, symptoms can include changes in mental status or an alteration or decrease in the sense of smell or taste. As some of these lesions invade the medial aspect of the orbit, diplopia, exophthalmos, or alterations in visual acuity may be noted.

Diagnosis

These individuals most often present to an otolaryngologist. The symptoms of sinusitis are unresponsive to standard conservative measures and medications. Skull films are often normal or will show fullness of the ethmoid and frontal sinus. CT will demonstrate lytic lesions in the cribriform and ethmoids on ‘bone windows'. Additionally, there will be a homogeneously enhancing soft-tissue mass in this region, usually asymmetric with variable amounts of tumor above and below the cribriform. MRI, particularly in the coronal plane, generally shows the tumor to better advantage (Fig. 9). As mentioned above, the persistent nasal mass will often mean that biopsy is done via this route, resulting in a histologic diagnosis prior to neurosurgical involvement. The histologic appearances are not uniform and, by utilizing the presence of lobar architecture, mitotic activity, nuclear pleomorphism, rosettes, and necrosis, a pathologic grading system has been designed that predicts reasonably well both outcome and the need for adjuvant therapy.

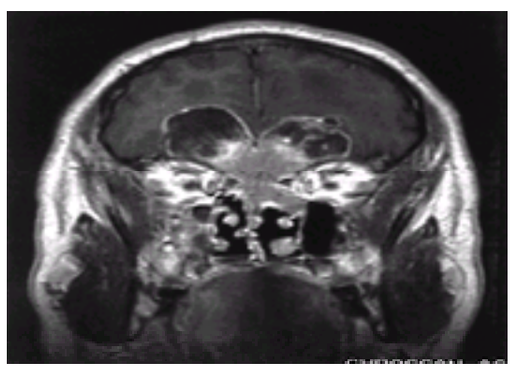


Fig. 9. Esthesioneuroblastoma. Coronal T_1-weighted gadolinium-enhanced MRI of a large lesion with homogeneous enhancement in the superior nasal, ethmoid, and

cribriform region. The usual thin, hypointense line showing the bony base of the olfactory groove is missing. The tumor extends bilaterally into the anterior fossa, with hypointense portions extending well into the frontal lobes.

Treatment

Surgery is the primary treatment for these lesions. In small lesions lying entirely below the cribriform, endoscopic resection is optimal. With clear evidence of defects in the skull base in a lesion primarily involving the nose and sinus, it is still prudent to be prepared for intracranial inspection and repair if necessary. Certainly, with a clearly intracranial tumor mass, a frontal craniotomy is necessary to address dural involvement or transdural invasion. The size of this approach depends on the size and location of the tumor. The surgeon must be able to see the normal structures of the skull base around the lesion. The tumor is usually separate from the brain, but large lesions are often infiltrative, requiring careful dissection between them and adjacent edematous brain. A graft of temporalis fascia, pericranium, or fascia lata can be used to repair the dura. Circumferential osteotomies will allow removal of the tumor. An attached and vascularized pericranial graft is then laid over the anterior skull base, covering the exenterated and occluded frontal sinus and the tumor defect in the ethmoids and cribriform, and anchored to the basal frontal dura behind the repair of the primary defect. The space between the pericranial flap and primary dural repair may be filled with fibrin glue. A brief period of cerebrospinal fluid diversion via a ventriculostomy or lumbar drain will assist in allowing the repair to heal well.

Gross total resection of a small esthesioneuroblastoma can be ascertained by the surgeon and the patient may be followed with CT or MRI at regular intervals. If microscopic or gross tumor is clearly still left behind, or if significant malignant changes are present on histologic analysis, then fractionated radiation doses of 55 to 65 Gy over 6 to 7 weeks can be used. Lesions with a significant degree of anaplasia may be treated with intravenous cytotoxic agents. No specific regimen is clearly superior but a combination of Adriamycin and cisplatin has some efficacy. Regular follow-up examinations and imaging should be maintained to observe for recurrence at or near the tumor bed.

Nasopharyngeal carcinoma

Clinical presentation

As many as 25 per cent of nasopharyngeal carcinomas affect the skull base. They tend to occur more often in males, and with a mean age of onset of 62 years in men and 72 years in women. As with other neoplasms of this region, presentation is usually associated with nasal congestion, nasal and pharyngeal drainage, epistaxis, and facial or frontal pain that may have been acute or, more commonly, indolent in onset. Lesions with which the neurosurgeon is involved are often invasive and may have spread into the orbit and intracranial space. Orbital involvement may be associated with exophthalmos, diplopia, erythema and chemosis, and visual obscuration. Intracranial invasion is most often frontal, but may be via the cavernous sinuses into the middle fossa or even through the clivus to the posterior fossa. The symptoms induced by intracranial invasion depend on the site of brain involvement. In general, frontal involvement will be associated with changes in mental status. Involvement of the middle fossa may be preceded by alterations of function in the nerves in the cavernous sinus. Once temporal damage has occurred, the patient may develop seizure activity or alterations in mental status. Invasion of the posterior fossa is uncommon, but when it occurs, cranial-nerve symptoms may be the first evidence, as these structures are injured as they leave the posterior fossa. Difficulties with balance and coordination may occur, owing to cerebellar dysfunction or hydrocephalus.

Diagnosis

An irregularly enhancing, nasal or sinus mass on CT, usually with variable degrees of bone destruction seen on properly ‘windowed' images, will suggest this diagnosis. The most common route of intracranial extension is through the foramen lacerum. From here the tumor has direct access to the cavernous sinus and orbital apex. During this process, dysfunction of the IIIrd, IVth, Vth, and VIth cranial nerves can be seen. It is believed that another route of tumor spread is along the nerves in and around the skull base. Therefore, enlargement and erosion of normally occurring foramina containing abnormal soft-tissue densities can also suggest this diagnosis. MRI will best delineate any involvement of the cranial nerves, cavernous sinus, sella, and brain. The tumor tissue is usually homogeneous on T_1-weighted images and hyperintense on T_2-weighted sequences (Fig. 10). MRI can also assist in differentiating inspissated mucus from tumor within sinuses. Direct transnasal biopsy of the lesion to confirm the diagnosis will be possible in the majority of cases.

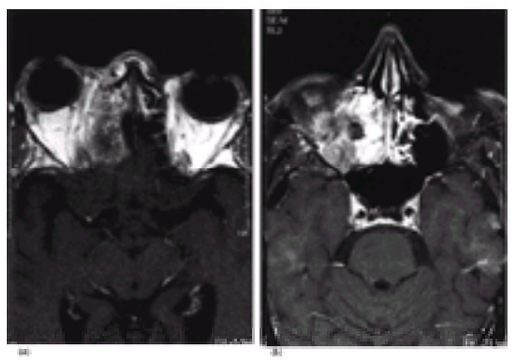


Fig. 10. Nasopharyngeal carcinoma. (a) Axial t_1-weighted gadolinium-enhanced MRI of a tumor arising the right ethmoid, deforming the medial aspect of the orbit and growing up against the gyri of the basal frontal lobe. (b) A parallel axial T_1-weighted gadolinium-enhanced image of the same tumor at a more inferior level, showing invasion of the orbit.

Histologic types of cancer in this location, in descending order of incidence, include squamous-cell carcinoma, adenocarcinoma, anaplastic carcinoma, salivary or adenoid cystic carcinoma, and metastatic tumors such as melanoma or lymphoma. Imaging by itself does cannot separate these types easily, indicating the need for biopsy to define more exactly the management and prognosis.

Treatment

If the patient is otherwise healthy and without debilitating systemic metastases, surgical resection is indicated. The goal of surgery is to obtain complete resection with tumor-free margins. Surgical planning includes determination of the approach, method of resection, and reconstruction.

Approaches to the anterior skull base include the trans-septal (trans-sphenoidal), transethmoidal, transmaxillary, extended frontal, anterior craniofacial, facial translocation, transoral, and the transmandibular–transcervical routes. Lateral extensions of the tumor may require approaches via the middle fossa, infratemporal fossa, or orbitozygomatic routes. In many cases the invasive nature of the tumor will have resulted in it becoming intertwined with structures of the orbit or cavernous sinus in a manner that is not amenable to complete surgical resection. The dura serves as an effective barrier to many of these tumors. Where the dura is crossed and there is clearly involvement of brain parenchyma, there is cerebral edema and mass effect. Neurosurgical involvement is indicated to resect this portion of the lesion if aggressive management is chosen. Resection by craniotomy or in a combined craniofacial resection may add to the palliative effort in these cases. Extensive invasion of the cavernous sinus with involvement of the carotids may preclude anything more than a biopsy. Also, orbital exenteration is indicated only if it is fairly certain that complete resection is being obtained. Large resections of the skull base will require repair with vascularized tissue to avoid leakage of cerebrospinal fluid.

As these lesions are rarely cured surgically, subsequent radiation to the tumor bed is necessary. In fact, part of the neurosurgeon's role is often to determine that no, or only limited, surgical intervention is warranted and to follow-up and perform well-documented neurologic examinations in an attempt to detect clinical signs of early tumor progression. Stereotactic radiation provides a method for focal radiation of tumor residues after initial surgery or at the time of recurrence. Chemotherapy may have a role; cisplatin and 5-fluorouracil have been used.

Survival depends on the stage or extent of involvement of the surrounding tissue. Greater extension is associated with shorter survival. Recurrences are frequent, with progression-free survival at 10 years being as low as 7 per cent for adenoid cystic carcinoma. With multimodal management, 5-year survival is approximately 20 to 40 per cent, depending on the histologic type of tumor.

Glomus tumor

Clinical presentation

Glomus jugulare tumors arise from cells in the adventitia of the jugular bulb and along Jacobson's and Arnold's nerves within the tympanic cleft. They are the most common tumor of the middle ear and second to the acoustic neuroma for tumors of the temporal bone. The mean age at presentation is 52 to 55 years and there is a female predominance. It must be kept in mind that these lesions can be bilateral or multicentric and, in rare cases, familial, where a modified autosomal-dominant transmission may be present. It appears that the genetic abnormality may be localized on chromosome 11q at 11q23 and 11q13.1.

Because these tumors often involve the middle ear, conductive hearing loss, ear pain, bloody discharge, and tinnitus can occur. As an alternative to tinnitus, the patient may note a pulsatile sound that corresponds with a mastoid bruit found on examination and reflects the vascular nature of these lesions. The surgeon may even be able to observe the lesion behind or transgressing the tympanic membrane or the external auditory canal. The glomus jugulare involves the jugular foramen and its contents, resulting in symptoms due to dysfunction of the glossopharyngeal and vagal nerves, and occlusion of the jugular vein and sigmoid sinus. Problems with swallowing and an alteration in vocal quality are the main clinical changes related to the glomus tumor. Interestingly, palsies of the IXth and Xth nerves are seen with another tumor in the differential diagnosis, schwannoma of the jugular foramen. Glomus tumors can involve the VIIth and VIIIth nerves or the contents of the middle ear, causing facial weakness and hearing loss. Shoulder weakness or aching from involvement of the accessory nerve can occur. Additionally, with large lesions, compression of the cerebellum or of vestibular nerves, or invasion of the vestibular apparatus, will result in ataxia and incoordination. Because glomus tumors grow slowly, the symptoms from venous occlusion tend to be minimal.

These lesions are also called paragangliomas, implying origin from paraganglionic cells of the neural crest. Early classifications mention chromaffin and nonchromaffin paragangliomas, based on tissue reaction with chromic acid, but this reaction is not always reliable and terminology based on location is clearer. Glomus tumors have been identified at various locations. The glomus tympanicum is not uncommon and arises from cells on the promontory of the cochlea in the hypotympanum; therefore it is more likely to result in altered hearing. Also, there is the glomus vagale, arising from receptor cells along the course of the vagus nerve, and the glomus caroticum (carotid body tumor or chemodectoma), arising at the division of the common carotid artery.

Diagnosis

Clinical examination will confirm the presence of abnormalities in voice quality, palatal movement, pharyngeal sensation, and gross alterations in hearing. Laryngoscopy and audiometry can better confirm these findings. CT using a ‘soft-tissue window' with and without contrast and ‘bone windows' will delineate these lesions well. MRI may assist in sorting out the relation between the tumor and structures outside of the jugular foramen and within the posterior fossa, which becomes important with larger lesions. On T_1-weighted images, mixed hypointense and isointense signals will be noted. The hypointense components are the flow voids of vascular channels and the isointense portions reflect solid tumor. This mixture of intensities, sometimes called a salt-and-pepper pattern, is the result of the tumor's extensive vascularity. Gadolinium enhancement is usually heterogeneous. As implied, glomus jugulare tumors can be very vascular and an angiographic blush of the jugular foramen region fed by the external carotid is characteristic (Fig. 11).

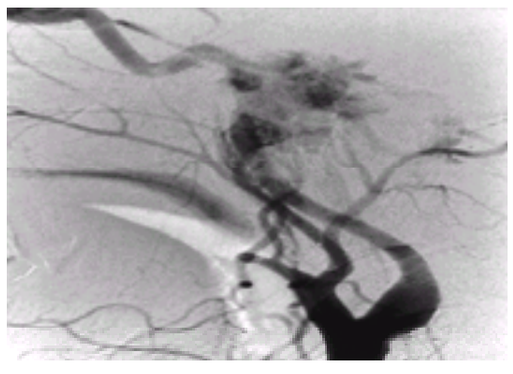


Fig. 11. Glomus jugulare. Left common carotid angiogram in modified lateral view showing contributions from branches of the external carotid artery to a tumor blush in the region of the jugular foramen; this pattern is characteristic.

As many of these symptoms will first lead the patient to the otolaryngologist, they will usually be seen by the neurosurgeon in consultation. Sometimes, larger lesions will have been biopsied via the external auditory canal, though most otolaryngologists recognize these and understand that this can result in significant hemorrhage in some cases.

Some of these tumors (approximately 1 per cent) are hormonally active, and urine and serum assays for catecholamines, vanillylmandelic acid, and metanephrine should be part of the evaluation; positive results will influence the anesthetic management of the patient.

Treatment

Because of their vascularity, preoperative embolization 1 to 2 days before surgery may ease dissection and decrease blood loss in larger lesions. This vascular intervention will also clarify the relation between the tumor and the carotid. In these larger lesions, upper cervical dissection to isolate the jugular, carotids, and IXth to XIIth cranial nerves serves as an initial step, followed by a mastoidectomy with exposure of the VIIth nerve, sigmoid sinus, and semicircular canals. Monitoring of the facial nerve and auditory-evoked potentials is useful during the dissection. The resection of the lesion, usually along with the sigmoid sinus and jugular bulb, is then carried out. An attempt is made to preserve the cranial nerves, and any attachments to the carotid artery are carefully dissected. Unresectable attachments to the carotid are best left and coagulated rather than risk injury to the artery. Intracranial portions of the lesion are resected via presigmoid or retrosigmoid approaches. Dural closure is important to avoid leakage of cerebrospinal fluid. In some cases, larger defects or sites that have already been radiated may need closure with rotational or vascularized muscle grafts.

In smaller lesions without intracranial extension, transpetrous resection without invasion of the posterior fossa is reasonable; the work within the petrous bone must preserve the VIIth nerve and the middle and inner ear. Where there is compression of the contents of the posterior fossa, neurosurgical intervention is necessary to stop progression of the neurologic deficit. Imaging studies will often show that, even with lesions having significant involvement of the posterior fossa, cure can only be obtained by removal of the portions of the tumor in the jugular foramen, mastoid, middle ear, and upper cervical soft tissues.

These lesions are radiosensitive. Its use is controversial in some circles, but offers a good alternative in patients who are unable to undergo surgery, have significant postoperative residues, or who suffer recurrence. Fractionated radiation in larger tumors and stereotactic radiation in lesions less than 3 cm in diameter can arrest growth and even result in regression. The patient must be told of the possibility of injury to the surrounding structures with this form of therapy, even if there is no incision. This method is particularly appealing in patients with other medical infirmities and in the presence of measurable recurrence.

Even after resection, the surgeon must be aware of the possibility of recurrence. Therefore postoperative imaging, at least annually, is necessary.

Neurofibromas and schwannomas

Neurofibromas and schwannomas are histologically benign neoplasms that arise from nerve sheaths, most commonly from sensory nerve roots. Neurofibromas appear grossly as a fusiform enlargement of the nerve whereas schwannomas are more clearly delineated from the nerve root. Therefore, surgical resection of schwannomas may be accomplished without sacrifice of the associated nerve root, a result that is almost impossible to achieve with a neurofibroma.

In the absence of neurofibromatosis, nerve-sheath tumors are almost invariably schwannomas and only rarely neurofibromas. Within the spinal axis, nerve-sheath tumors are distributed fairly evenly, being more prevalent in the thoracic region. They occur most commonly in an intradural, extramedullary location and present with pain in the distribution of the involved nerve root, followed by progressive myelopathy with continued growth.

Intracranially, schwannomas most commonly arise from the vestibular nerve. Although the most accurate term for these neoplasms is vestibular schwannoma, the term ‘acoustic neuroma' is firmly entrenched in the neurosurgical literature. The most common presentation of the acoustic neuroma is progressive, unilateral hearing loss due to compression of the adjacent cochlear nerve. Tinnitus is another common early symptom. With progressive growth, the tumor may compress the adjacent trigeminal nerve and brainstem. Nerve-sheath tumors may occur on other cranial nerves such as the trigeminal or glossopharyngeal.

Neurofibromatosis

The neurofibromatoses are genetic disorders associated with a high incidence of tumors of the nervous system including neurofibromas, schwannomas, meningiomas, and gliomas. Two types of neurofibromatosis have been well characterized and the genetic abnormalities identified. Both neurofibromatosis type I and type II are transmitted in an autosomal-dominant manner with high penetrance. The genes for each type have been cloned and protein sequences identified.

Type I is the most common, occurring approximately once in every 4000 births. Previously known as peripheral neurofibromatosis or von Recklinghausen's disease, its most common manifestations are multiple cutaneous neurofibromas, *café au lait* marks, axillary freckling, Lisch nodules of the iris, skeletal abnormalities, and an increased incidence of childhood gliomas including optic gliomas, ependymomas as well as meningiomas, hamartomas, and primitive neuroectodermal tumors. Tumors of spinal nerve roots in patients with type I neurofibromatosis are commonly asymptomatic, multiple, and more often neurofibroma than schwannoma. Plexiform neurofibromas are peripheral nerve tumors composed of Schwann cells, fibroblasts, and connective tissue; these may degenerate into malignant neurofibrosarcomas in patients with type I.

Neurofibromatosis type II is much less common than type I, occurring once in every 100 000 births. Previously called central neurofibromatosis, the hallmark of this condition is bilateral acoustic neuromas. Meningiomas, sporadic unilateral acoustic neuromas, sporadic spinal schwannomas, and ependymomas are found with higher incidence in patients with type II.

Diagnosis

Both intracranial and intraspinal nerve-sheath tumors are best seen by MRI, with and without gadolinium. In spinal lesions these studies demonstrate the relation of the tumor to the spinal cord, to bony structures, and, in high cervical lesions or lesions in the foramen magnum, its relation to the vertebral arteries (Fig. 12). MRI elegantly demonstrates the extraspinal component of neurofibromas that extend through the neural foramina in a dumbbell shape. Nerve-sheath tumors usually enhance homogeneously after the administration of gadolinium.

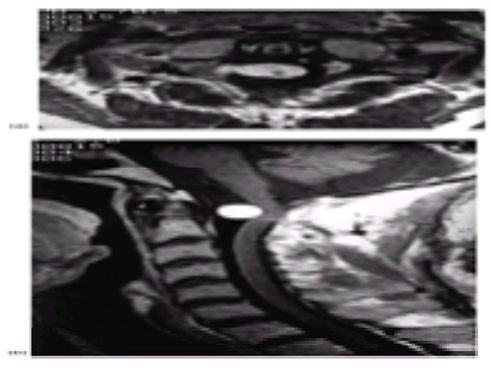


Fig. 12. (a) Axial and (b) sagittal MRI demonstrates a schwannoma ventral to the upper cervical spinal cord causing significant compression; this study illustrates the relation of the tumor to surrounding important structures.

Intracranial nerve-sheath tumors such as acoustic neuromas are best visualized by MRI with and without contrast; this is very helpful in determining the relation of the tumor to important structures and in planning surgical approaches (Fig. 13).

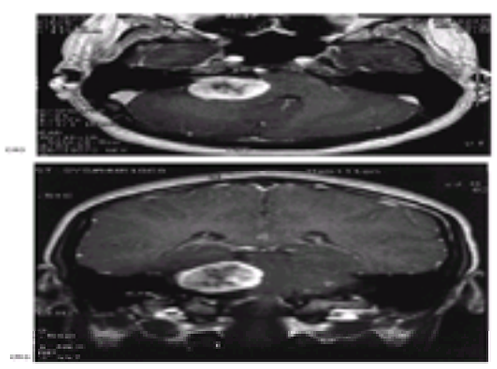


Fig. 13. (a) Gadolinium-enhanced axial T_1-weighted MRI of a right-sided acoustic neuroma. There is compression of the pons and cerebellum and the fourth ventricle is shifted to the left. The tumor extends into the internal auditory canal and has a heterogeneous signal internally. (b) Gadolinium-enhanced coronal T_1-weighted MRI of the same lesion showing that the tumor extends all way up to the tentorium and occupys the usual location of the middle cerebellar peduncle.

Treatment

Most tumors of spinal nerve sheaths are benign and complete surgical excision is potentially curative. For symptomatic patients with a reasonable life expectancy, surgical treatment offers a high likelihood of success and an opportunity to eliminate a potential source of progressive myelopathy. Frequently, patients with neurofibromatosis will have multiple asymptomatic tumors that can be simply observed by serial MRI.

For intracranial tumors, such as acoustic neuroma, there is a variety of treatment options. Small, asymptomatic or minimally symptomatic tumors may be observed by serial MRI, particularly in the elderly patient. Different surgical approaches may be considered for acoustic neuromas, chosen on the basis of several factors including tumor size and configuration, opportunity for preservation of hearing, and surgeon's experience. The translabyrinthine operation is a versatile approach for managing acoustic tumors: one of its primary advantages is early and accurate visualization and preservation of the facial nerve, but it is always associated with hearing loss on the side of the tumor. Translabyrinthine approaches are often used for the resection of small to medium-sized tumors in patients without useful hearing. A retromastoid suboccipital craniotomy is an excellent approach for tumors of all sizes and is the procedure of choice when preservation of hearing is attempted. Small lesions within the internal auditory canal in patients with functional hearing may also be resected via the middle fossa.

Stereotactic radiosurgery is another option in the management of acoustic tumors for selected, high-risk patients. It involves the delivery to the tumor, in a single session, of a high dose of targeted radiation such that the surrounding neural structures receive a low dose. Its primary advantage is that it is minimally invasive. Its disadvantages are the long interval to the desired effects of the radiation, the relatively low cure rate, and the low but definite risk of radiation injury to the brainstem and adjacent cranial nerves.

Pineal tumors

Introduction

The pineal gland is a small, midline organ situated in the quadrigeminal plate cistern attached to the posterior roof of the third ventricle. The gland is divided by connective tissue septa into lobules consisting of pineocytes, which are specialized neuroepithelial cells with neurosecretory properties. The physiologic importance of the pineal gland remains unclear, although it is believed to have a role in circadian rhythms and reproduction. Despite its small size, the diversity of cell types that constitute the gland and adjacent structures results in a large variety of neoplasms. Although many different neoplastic and non-neoplastic mass lesions occur in this region, pineal tumors represent only about 1 per cent of all brain tumors and 3 to 10 per cent of brain tumors in childhood.

Clinical presentation

Because of the strategic anatomic location of the pineal gland, the presenting symptoms and signs of pineal mass lesions are similar regardless of the histopathologic features. Growth of a mass lesion in this region frequently results in compression of the aqueduct of Sylvius, resulting in obstructive hydrocephalus. Lesions causing hydrocephalus present with the features of increased intracranial pressure, including headache, nausea and vomiting, and lethargy. Infiltration or compression of the midbrain results in Parinaud's syndrome, the classic syndrome associated with tumors of the pineal region. Parinaud's syndrome includes paralysis of upward gaze and ocular convergence, and pupillary abnormalities characterized by non-reactivity to light but constriction to accommodation. Endocrine disturbances, including precocious puberty, hypogonadism, diabetes insipidus, and anterior pituitary insufficiency, are frequently associated with tumors of the pineal region.

Diagnosis

The diagnostic investigation of masses in the pineal region has been revolutionized by MRI. This noninvasive study clearly demonstrates all the mass lesions that arise in this region, the extent of tumor compression or invasion, and the precise relation to surrounding anatomic structures in multiple planes. Because certain pineal tumors have a predilection for metastasis within the cerebrospinal fluid spaces, evaluation of the spinal axis by MRI is recommended to look for these ‘drop metastases'.

In general, the precise diagnosis of a tumor of the pineal region cannot be made from imaging studies. Histologic diagnosis is essential, since each different tumor arising in this region has an optimal method of treatment.

Tumor types

Tumors arising within the pineal region can be classified into five major groups of mass lesions: pineal parenchymal tumors, tumorsof the supporting glial cells, germ-cell tumors, mesenchymaltumors (meningiomas), and non-neoplastic masses.

Pineal parenchymal tumors

Tumors arising from the pineal parenchyma constitute approximately 15 to 30 per cent of all tumors of the pineal region. They include:

1. *pineoblastomas*: highly malignant, primitive neoplasms with a propensity for local infiltration and dissemination in cerebrospinal fluid;
2. *pineocytomas*: tumors with primarily benign cytologic features and little tendency to invade the surrounding brain or spread within the neuraxis (Fig. 14);

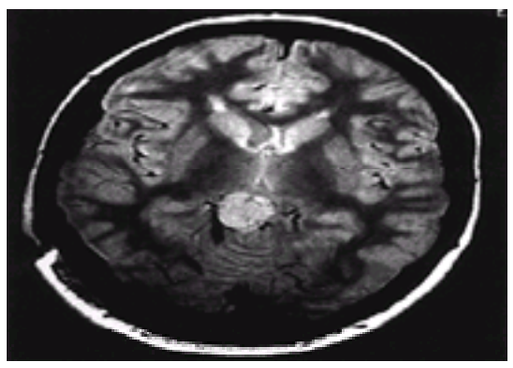


Fig. 14. Axial, primarily T_2-weighted MRI of a pineocytoma: this tumor is located posterior to the third ventricle, anterior to the cerebellar vermis, and just to the right of the midline; note the significant vascular flow voids adjacent to the lesion and the relatively clear delineation between tumor and normal parenchyma.

3. *mixed pineocytoma–pineoblastomas*: neoplasms that exhibit features of both of the aforementioned lesions.

Glial tumors

Gliomas, particularly astrocytomas, arise from the supporting glial cells present within the pineal gland or from structures adjacent to the pineal gland; they account for 10 to 35 per cent of tumors in the pineal region. Glial tumors arising here include benign astrocytomas, anaplastic astrocytomas, glioblastomas, ependymomas, oligodendrogliomas, and papillomas of the choroid plexus.

Meningiomas

Meningiomas of the pineal region typically arise from the falx, tentorium, or the velum interpositum, a fold of arachnoid that forms at the roof of the third ventricle. These lesions typically expand into the pineal region and posterior third ventricle.

Germ-cell tumors

Germ-cell tumors that occur in the region of the pineal gland are indistinguishable from those arising in the testicle and ovary; they comprise the largest group of neoplasms in the pineal region, accounting for nearly 50 per cent. They occur most commonly during the first two decades of life and are more common in males than females. The six major types of germ-cell tumors in approximate order of increasing malignancy are mature and immature teratomas, germinomas, malignant teratomas, embryonal carcinomas, endodermal sinus tumors, and choriocarcinomas.

Germinomas, which are similar histologically to testicular seminomas, are the most common germ-cell tumor affecting the pineal region. They are locally invasive and tend to spread through the ventricular system and subarachnoid spaces via the cerebrospinal fluid. Although these tumors have malignant characteristics, they are exquisitely radiosensitive and long-term survival or cure can be accomplished. Other germ-cell tumors are highly malignant. Although they may show an early response to radiation therapy, they are usually fatal.

Non-neoplastic lesions

Cysts of the pineal region are relatively common findings at autopsy. With the advent of MRI, these lesions have been diagnosed with greater frequency. They are typically benign and almost always an incidental finding. On rare occasions they will enlarge and result in obstructive hydrocephalus.

Treatment

Prior to the advent of microsurgical techniques, surgical approaches to the pineal region were fraught with high morbidity and mortality. Standard treatment typically consisted of ventricular shunting to relieve hydrocephalus, followed by an empiric course of radiation. Because of the high incidence of germinomas, this approach was associated with some successes.

In the modern microsurgical era, tumors of the pineal region can be approached directly with a very low operative morbidity. Those patients who present with symptomatic hydrocephalus should undergo placement of a ventriculoperitoneal shunt to relieve the hydrocephalus. Lesions are most commonly approached by an infratentorial, supracerebellar method that exposes the pineal region through the posterior fossa over the superior surface of the cerebellar vermis. Approximately

one-third of pineal tumors are benign and curable with surgery alone. If the intraoperative biopsy demonstrates a germinoma, the treatment of choice is radiation therapy. Other highly malignant germ-cell tumors are typically treated by subtotal resection, followed by radiation therapy. Chemotherapy is now often instituted, based on experience with tumors of gonadal origin arising elsewhere in the body.

Certain malignant germ-cell tumors are associated with tumor markers in cerebrospinal fluid. Concentrations of a-fetoprotein, normally produced by the fetal yolk sac, may be elevated in patients with endodermal sinus tumors or embryonal carcinomas. Human chorionic gonadotrophin is normally produced by syncytiotrophoblasts of the placenta and is commonly elevated in the cerebrospinal fluid of patients with choriocarcinoma or embryonal carcinoma. Although cytologic evaluation of cerebrospinal fluid, tumor markers, and radiographic features can inform predictions of a diagnosis, this can be accomplished accurately only by pathologic examination of tumor tissue. These markers in cerebrospinal fluid are useful in following patients for evidence of recurrence of disease or response to treatment.

Despite the advantages of direct surgical approaches to pineal tumors, there are several circumstances in which a stereotactic biopsy is the preferred treatment. It is most appropriate for patients who are elderly or in whom significant medical complications increase the risk of open craniotomy. Patients with disseminated tumors at the time of presentation are often best managed by stereotactic biopsy to establish a diagnosis, followed by adjuvant therapy including radiation and/or chemotherapy.

Meningiomas and choroid plexus tumors

Meningiomas are neoplasms that arise from the meningeal coverings of the brain or spinal cord. They originate from the arachnoidal cap cells, commonly in association with the arachnoid villi at the dural venous sinuses and their tributaries, the foramina of the cranial nerves, the cribriform plate, and the medial middle fossa. The overwhelming majority of meningiomas are benign and most grow very slowly.

Meningiomas

Etiology

Various environmental and genetic factors are proposed as potential contributors to the pathogenesis of meningiomas. Prior head trauma has been associated with these tumors, but several case–control studies have failed to show a correlation. A clear connection between radiation and meningiomas has been established. Significant evidence suggests that the growth of meningiomas may be influenced by female sex hormones, providing a rationale for clinical trials of estrogen and progesterone inhibitors.

Clinical features

Although meningiomas may arise anywhere within the intracranial or intraspinal cavities, there are specific locations where they are more common. The clinical presentation of meningioma depends upon the location, the rate of tumor growth, and the extent of peritumoral edema. Compression of adjacent brain and neural structures may produce global or focal neurologic symptoms and signs. A meningioma occurring in a relatively silent area of the brain may become rather large before producing symptoms, whereas a relatively small tumor may produce significant neurologic deficits if it is in a functionally important area.

Convexity meningiomas arising from the dura over the surface of the brain cause symptoms by compressing the adjacent cerebral hemispheres and may produce headache, focal neurologic deficits, or seizures.

Parasagittal meningiomas involve the superior sagittal sinus and adjacent dura and falx; patients frequently present with focal motor or sensory seizures, or leg weakness. Progressive enlargement of these lesions may eventually occlude the superior sagittal sinus.

Falx meningiomas arise from the falx cerebri and only secondarily invade the superior sagittal sinus; their presentation is similar to that of parasagittal meningiomas.

Meningiomas of the sphenoid wing may arise anywhere along the sphenoid ridge. In the lateral sphenoid wing they usually present with slowly progressive, painless exophthalmos. In the mid-portion of the sphenoid wing they may compress the frontal and temporal lobes, and present with headaches and/or seizures. Meningiomas of the medial sphenoid wing may involve the anterior clinoid process and present with unilateral loss of vision. Secondary involvement of the superior orbital fissure or cavernous sinus may produce cranial-nerve palsies.

Meningiomas of the cavernous sinus may originate within the sinus itself or invade this structure secondarily from other adjacent locations. Diplopia and facial numbness or pain are manifestations of involvement of the intracavernous cranial nerves. Meningiomas of the optic sheath present with progressive, painless, ipsilateral visual loss.

Meningiomas of the olfactory groove may reach large size before clinical detection because of slow growth compressing the frontal lobes; this may result in changes in mental status, decreased visual acuity, or seizures. Anosmia is often present.

Meningiomas arising from the tuberculum sellae typically present with visual symptoms due to lateral displacement of the optic nerves; these lesions commonly present with headache and changes in mental status.

Meningiomas may arise from any site along the tentorium. If the growth is primarily supratentorial, patients most commonly present with headaches and/or seizures. Primarily, infratentorial growth will result in cerebellar, brainstem or cranial-nerve deficits, or hydrocephalus.

Meningiomas may arise from any level of the clivus. Clival meningiomas may grow along the petrous bone or into the cavernous sinus and middle fossa. Symptoms and signs result from cranial-nerve dysfunction, cerebellar and/or brainstem compression, and intracranial hypertension.

Meningiomas of the cerebellopontine angle present with cerebellar symptoms and signs, and hearing loss.

Meningiomas of the foramen magnum are most commonly located anterolateral to the medulla and typically present with progressive quadriparesis and dysfunction of the lower cranial nerves.

Intraventricular meningiomas are rare and most commonly occur in the trigone of the lateral ventricle. Headache is the most common symptom of these lesions. With progressive enlargement, they may produce intracranial hypertension due to direct compression of the adjacent brain and development of hydrocephalus.

Spinal meningiomas typically present with focal back pain and progressive paresis below the level of the lesion. As compression of the spinal cord advances, patients develop sphincter disturbances.

Diagnosis

Meningiomas have a characteristic appearance on both CT and MRI. MRI has become the procedure of choice for diagnosis of most meningiomas: this noninvasive study elegantly demonstrates the lesion in multiple planes, its relation to surrounding neural structures, and any associated edema (Fig. 15). Some meningiomas are associated with significant hyperostosis or frank invasion of the Haversian canals of adjacent bone, best seen on CT.

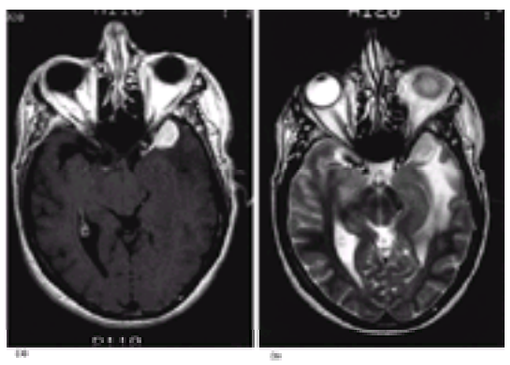


Fig. 15. (a) Contrast-enhanced axial T_1-weighted MRI of a meningioma of the medial sphenoid wing; (b) T_2-weighted MRI of the same case—note the extensive hyperintense signal abnormality in the temporal lobe adjacent to the tumor.

Magnetic resonance angiography is useful in defining the relation of meningiomas to the arterial circulation and in disclosing the presence of invasion or occlusion of venous sinuses. Occasionally, catheter angiography is indicated to define better the vascular relations of a meningioma, or for preoperative embolization of highly vascular lesions.

Treatment

With modern, noninvasive imaging many meningiomas are diagnosed when they are minimally symptomatic or as incidental findings during an evaluation for unrelated symptoms. Particularly in elderly patients these lesions may not require treatment, and serial MRI or CT can be used to follow tumor growth or the progression of symptoms.

The primary treatment for most symptomatic meningiomas is surgical resection. Factors influencing safety, ease, and outcome of surgical removal include tumor location, size, consistency, and vascular and neural involvement. To minimize the likelihood of recurrence the neurosurgeon must resect not only the entire neoplasm but also all involved dura, soft tissue, and bone. Meningiomas of the convexity have the greatest potential for total removal and cure because of the relative ease of exposure and the opportunity to resect a wide margin of dura. Meningiomas involving the anterior sagittal sinus may be managed by ligation and excision of the sinus, if involved. Involvement of the posterior two-thirds of the superior sagittal sinus often requires extensive efforts to reconstruct the sinus and maintain its patency if not already occluded by the tumor.

Variuos approaches that involve more extensive removal of bone to minimize retraction of the brain are useful in the management of meningiomas involving the anterior, middle, and posterior skull base. Spinal meningiomas are generally approached through a laminectomy with dural opening, with careful microsurgical resection of the lesion and its dural attachment.

Although the surgical goal for meningiomas is complete removal of the tumor and involved dura, this objective should be tempered with good judgement and the primary aim of preserving and improving neurologic function. In some patients, total removal of the meningioma carries the risk of significant morbidity, and extensive but subtotal removal may be followed by resolution of symptoms for many years without significant recurrence.

A number of studies have shown the effectiveness of radiation therapy in reducing the likelihood and rapidity of recurrence following subtotal resection. Although less conclusive, there are reports of the effectiveness of radiation therapy in meningiomas considered inoperable because of their location or the patient's poor health. More recently, stereotactic radiosurgery has been used with promising results on relatively small tumors considered inoperable or high risk, and for residual tumor in functionally important areas.

Although little information is available on the efficacy of traditional chemotherapeutic agents for meningiomas, the antiestrogen tamoxifen and the antiprogesterone mifepristone have produced some encouraging early results in selected cases.

Choroid plexus tumors

Tumors of the choroid plexus may occur anywhere within the ventricular system where choroid plexus exists. The majorityoccur in the atrium of the lateral ventricle, a smaller number within the fourth ventricle, and rarely in the third ventricle. Choroid plexus tumors characteristically occur in young children. Although they account for 10 to 20 per cent of all brain tumors occurring during the first year of life, they represent only 3 per cent of all brain tumors of childhood. Tumors of the choroid plexus include benign papillomas, or malignant carcinomas that invade the brain and tend to disseminate within the cerebrospinal fluid.

Tumors of the choroid plexus most commonly present with hydrocephalus, which may be due to an overproduction of cerebrospinal fluid by the tumor, direct obstruction of the fluid pathways, or impairment of the fluid absorption due to episodes of subarachnoid hemorrhage.

These lesions are best diagnosed by MRI or CT. On CT, papillomas of the choroid plexus are often hyperdense with calcification, and demonstrate homogeneous enhancement after contrast injection. Carcinomas are more heterogeneous in density, with homogeneous contrast enhancement and evidence of tumor necrosis or cyst formation. These lesions and their relation to the ventricle and adjacent neural structures are elegantly demonstrated on MRI.

Surgery is the treatment of choice for papillomas and carcinomas of the choroid plexus. Complete removal of either of these types of lesion may be associated with long disease-free survival. Because of the tendency for leptomeningeal spread, these patients must be monitored closely, and those with carcinomas may require radiotherapy and chemotherapy for control of disease.

Pituitary tumors

Clinical problems

Pituitary adenomas may present in one of three ways: (1) as an endocrinopathy, i.e. the abnormal release or suppression of hormones, causing an alteration in the body's hormonal function; or (2) as a mass lesion with growth of the tumor compressing neural or vascular structures; or (3) as a combination of those two.

Tumors that arise in the pituitary may be described as endocrine active (functional) or endocrine inactive (non-functional). The most common endocrine-active tumors are those that secrete prolactin, and the next most common are growth hormone- and adrenocorticotrophic hormone (**ACTH**)-producing tumors. Only occasional tumors produce luteinizing, follicle-stimulating, or thyroid-stimulating (**TSH**) hormones. The most common adjacent neural structure affected is the optic chiasm, causing bitemporal hemianopia.

Diagnosis and treatment

Pituitary adenomas may be treated surgically, medically, or with radiation. In many instances the therapeutic end-point is control rather than cure of the endocrinopathy. The goals of treatment in patients with functional pituitary adenomas are (1) to decompress any neural structures compromised by the tumor mass and (2) to restore and maintain normal hormone secretion.

Endocrine evaluation

Pituitary endocrine function can be assessed systematically. A normal morning serum cortisol usually reflects an intact hypothalamic–pituitary–adrenal axis. The pituitary–thyroid axis can be assumed intact if the serum thyroxine (T4) (total or free) is normal. Laboratory assessment may also include measurement of serum luteinizing and follicle-stimulating hormones and sex steroids (estradiol in women; testosterone in men). Prolactin should be assayed in all patients with pituitary tumors, especially apparently nonfunctional tumors. Growth hormone and somatomedin-C levels are required if there is clinical suspicion of acromegaly. Growth hormone deficiency may be clinically significant in adults and should be evaluated in all children with pituitary or hypothalamic lesions.

Imaging studies

MRI has become the primary modality for diseases of the sellar and parasellar region. The most definitive study using MRI involves thin coronal, sagittal, and axial slices combined with administration of gadolinium. With this technique, the normal pituitary gland, cavernous sinus, and infundibulum enhance immediately. Microadenomas do not enhance immediately and therefore show up as areas of decreased signal intensity.

Treatment

Surgery is the preferred treatment for most pituitary tumors. The two most common procedures used are trans-sphenoidal adenomectomy and craniotomy. The goals

of surgery are: (1) to decompress neural structures, particularly the optic nerves; (2) to correct the endocrine dysfunction; (3) to obtain a histologic diagnosis; and (4) to remove the tumor completely, if feasible.

The trans-sphenoidal approach is favored by most surgeons because it is safer and more effective than craniotomy. Trans-sphenoidal adenomectomy is the procedure of choice for pituitary adenomas that cause Cushing disease and acromegaly, and for some prolactinomas. Morbidity and mortality are low when the trans-sphenoidal procedure is done by an experienced surgeon.

Hydrocortisone (100 mg) is given intravenously before surgery and tapered postoperatively. The primary considerations in the postoperative period are fluid and electrolyte balance, and corticosteroid coverage. Most instances of postoperative diabetes insipidus after trans-sphenoidal surgery are transient and resolve spontaneously.

Currently, there are no cytotoxic drugs that cure pituitary adenomas. Therefore the options for treating these lesions medically are limited to attempts to (1) reduce the excessive amounts of hormone associated with certain hypersecretory tumors, and (2) reduce the size of certain others.

Because trans-sphenoidal adenomectomy has produced such excellent results in the treatment of pituitary adenomas, radiation usually is reserved for adjunctive treatment with recurrent or residual tumors. Use of radiation as the primary treatment usually is reserved for patients who cannot undergo surgery for medical reasons or for those who refuse surgery.

Endocrine-inactive tumors

Clinical presentation

The symptoms of endocrine-inactive tumors usually relate to loss of pituitary hormonal function and mass effect because patients commonly have tumors that have grown to considerable size before being recognized.

A careful history taken from a patient with an endocrine-inactive tumor who is being examined because of loss of vision almost always reveals symptoms consistent with loss of endocrine function that considerably predate the onset of that loss. In women these symptoms may include early menopause or infertility, and in men the symptoms include lack of potency or infertility. Both men and women may experience excessive fatigue, inability to tolerate temperature extremes, and decreased hair growth. Headache is not an uncommon symptom and when present has one of two basic components: pain referred to the vertex of the skull, reflecting pressure on to the diaphragm of the sella, or pain referred to the retro-orbital and supraorbital areas from pressure on the ophthalmic division of the trigeminal nerve.

Prolactinomas

Clinical presentation

Women with prolactinomas usually see a physician because they have experienced amenorrhea, galactorrhea, or infertility. Hyperprolactinemia also suppresses gonadal function in men, but because of insidious onset, denial, or reluctance, decreased libido or erectile impotence do not often bring men to their physicians. This delay may account in part for the disproportionate number of macroprolactinomas (larger than 10 mm in diameter) in men, whereas these account for only 5 to 10 per cent of all prolactinomas in women.

Physical examination in sexually developed women reveals galactorrhea in more than 85 per cent of patients. Abnormal physical findings in men include galactorrhea, gynecomastia, and testicular atrophy. If the serum prolactin is above 200 ng/l the patient almost certainly has a prolactinoma.

Management

Surgical

The most successful cure rates (60–70 per cent) are reported by centers in which an experienced surgeon performs the surgery, and when the tumors are microadenomas and the prolactin is below 200 ng/ml. When surgery is successful, galactorrhea stops, normal menstrual cycles resume, and the prolactin normalizes below 25 ng/ml.

Medical

Various hormone treatments have been used for prolactin-secreting tumors, but the most common drug treatment is the ergot derivative bromocriptine. The most common side-effects are nausea and vomiting, dizziness, and sedation. More recently, carbegoline (Dostinex) has become available; this has the advantage of weekly or twice-weekly dosing compared with daily dosing for bromocriptine.

Acromegaly

Acromegaly is an uncommon condition, with an incidence of 500 to 1000 cases each year in the United States. It accounts for approximately 10 per cent of all pituitary tumors. Acromegaly poses a significant risk to health and may shorten life expectancy if left untreated.

Clinical presentation

The most common manifestations of acromegaly are swelling of soft tissue and skin thickening. Headache, excessive perspiration (hyperhidrosis), altered menses, decreased libido, and dysesthesias (e.g. carpal tunnel syndrome) resulting from nerve entrapment are all common complaints.

Endocrine evaluation

Basal concentrations of growth hormone, and of somatomedin-C or insulin-like growth factor-I, are elevated. Somatomedin-C provides an essential means of confirming the diagnosis of acromegaly.

Management

Surgical

The optimal surgical treatment is selective, total removal of the growth hormone-secreting pituitary adenoma.

Medical

Currently the mainstay of drug therapy for acromegaly is octreotide, which is given by subcutaneous injection.

Cushing disease

Cushing disease is a pituitary-dependent excess secretion of ACTH. It affects women more often than men, and poses a significant threat to health and life expectancy. Adult sufferers complain of weakness, weight gain, easy bruising, pyschologic changes, excess body hair, oligomenorrhea, change in facial shape and color, and bone pain.

Physical examination reveals centripetal obesity, muscle wasting, thin skin with ecchymoses and telangiectasias, intertriginous fungal infections, and facial plethora. Often a fine, downy hirsutism in the ‘sideburn' area, violaceous stretch marks, and acne are present. Men have a reduction in testicular size, impotence, and loss of body hair. Diabetes and hypertension are very common, and signs of congestive heart failure with edema may be present.

Diagnostic studies

Measuring free cortisol in a 24-h sample (i.e., urinary free cortisol) from an unstressed patient provides a reliable indication of cortisol production. Another useful screening test for Cushing syndrome is the overnight dexamethasone suppression test. Plasma ACTH is elevated in patients with Cushing disease and in those with an ectopic source of ACTH.

Sampling from the petrosal sinus was developed to establish a pituitary origin for Cushing disease in patients with negative MRI and to lateralize the tumor within the gland by establishing a differential in ACTH concentrations between the left and right sinuses draining the pituitary gland.

Medical management

Currently, ketoconazole is the most effective drug available for treating Cushing disease.

Gonadrotropin-secreting adenomas

Clinical presentation

Because they cause relatively few symptoms, gonadotropin-secreting adenomas frequently go undetected. Most of these tumors occur in middle-aged men; they are uncommon in women of reproductive age. Headaches and visual problems are common symptoms.

TSH-secreting adenomas

Clinical presentation

Tumors that secrete TSH are the least common adenomas, accounting for less than 1 per cent. They are usually macroadenomas and manifest with compressive symptoms. Some patients have hyperthyroidism with goiter.

Future directions in pituitary surgery

Endoscopy is being used in surgery to improve visualization. Intraoperative ultrasonography and MRI are also available for real-time assessment of tumor resection. Molecular analysis of pituitary tumorigenesis is an active area of research.

Intracranial epidermoid tumors

Epidermoid tumors are congenital neoplasms that grow through desquamation of keratin, cholesterol, and cellular debris. They are histologically benign, but rare malignant degeneration does occur. Intracranial epidermoid tumors account for approximately 1 to 2 per cent of all brain tumors; the most common locations are parachiasmal and in the cerebellopontine angle; they may also occur in the calvarium. The natural history of the intracranial tumors is a slow, relentless progression of symptoms. They are rarely asymptomatic. Specific signs and symptoms depend upon the location of the tumor. Seizures are a common presenting symptom. They may also present with recurrent episodes of aseptic meningitis.

MRI has revolutionized the diagnosis of these tumors. In T_1 images the epidermoid is well defined and separable from normal tissues; in T_2 images the tumors have high signal intensity. They rarely enhance.

Supratentorial tumors

Patients with suprasellar tumors typically present with visual complaints.

Infratentorial tumors

Tumors of the cerebellopontine angle may present with hearing loss, dizziness, or trigeminal paresthesias.

Treatment

The goal of surgery is total removal: the real issue is the difficulty of removing every bit of capsule; if the entire capsule is not removed, recurrence is the rule. The second major problem with surgery is the noxious nature of the cholesterol/keratin material that constitutes the bulk of the tumor. If this material escapes into the subarachnoid space, the result can be a meningitis that may be highly disabling and can produce permanent cranial-nerve deficits.

Craniopharyngioma

Craniopharyngiomas are epithelial tumors derived from the vanishing hypophyseal duct during the formation of Rathke's pouch. They may be intrasellar, suprasellar, intraventricular, or in combinations of these locales. Because of their central location they may be anatomically related to many neural and vascular structures. Inferiorly, the pituitary gland and stalk may be in close proximity. Anteriorly, the suprasellar cistern contains the optic nerves and chiasm, the carotid arteries and their branches. Superiorly, the hypothalamus, and posteriorly, the midbrain and cranial nerves III, IV, V, and VI may be encountered, as well as the basilar artery and its branches. These tumors comprise 2.5 to 4 per cent of all brain tumors and about half occur in childhood.

Clinical features

The majority of patients with craniopharyngiomas present with ophthalmologic symptoms. Other clinical symptoms include headaches, seizures, or neurologic deficits.

Surgical treatment

A number of surgical routes have been used to reach these tumors, including the subfrontal, transcortical–transventricular, transcal-losal, trans-sphenoidal, subtemporal, pterional, and stereotactic approaches. The route selected depends on the location and anatomic relations of the tumor, and the preference of the surgeon.

Endocrinologic compromise and its management

The activity of the hypothalamic–pituitary axis should be assessed to determine the extent of hormonal disturbance before and after surgery. Endocrinologic compromise may be neurohypophyseal or adenohypophyseal. Neurohypophyseal compromise is usually manifested by disturbances in fluid and electrolyte balance; specifically, the disorders of diabetes insipidus and the syndrome of inappropriate antidiuretic hormone secretion.

Adenohypophyseal

Deficiencies may be of ACTH or thyrotropin (TSH); these must be replaced in the preoperative and perioperative period. Long-term replacement is also frequently necessary.

Radiation

Postoperative radiation may be useful in treating residual tumor. In children this should be used judiciously because of deleterious effects on intelligence. Radiation may be delivered as interstitial intracavitary therapy or conventional external radiotherapy. Total removal of craniopharyngiomas is often not achievable; therefore,

radiation serves as an important adjuvant treatment.

Brain metastases

Introduction

The spread of systemic cancer to the nervous system has been estimated to occur in 20 to 40 per cent of patients. Direct spread into the brain parenchyma is by far the most common, although nearly 10 per cent of patients have extension of tumor into the leptomeninges (neoplastic meningitis) or dura. The actual incidence of central nervous metastases has been difficult to determine because of variability in the methods of assessment (clinical, radiographic, or autopsy) used to make the diagnosis. More recent studies, using modern imaging techniques, are yielding the highest incidences.

The frequency of spread to the central nervous system varies with the type of cancer (Table 2). Some, such as malignant melanoma, have a distinct predilection to metastasize to brain; in many series malignant melanoma is the third most common histologic type of brain metastasis despite the relatively low incidence of this tumor. Lung cancer, particularly small cell, is the most common primary. Breast cancer, a very common cancer with an often prolonged survival, frequently spreads to the central nervous system. Prostate cancer, despite its high incidence and prolonged survival, rarely produces brain metastases.

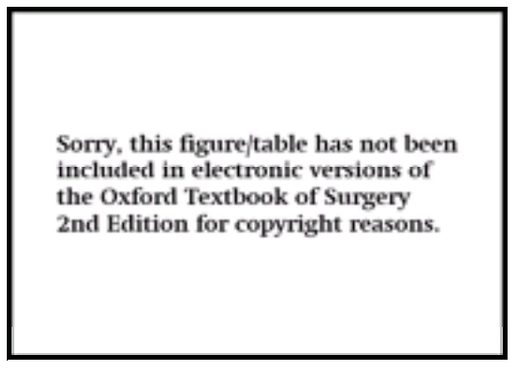


Table 2 Brain metastases by primary tumor type

Brain metastases are most commonly hematogenous, which provides a reasonable explanation for their frequent development at the gray–white junction, a ‘watershed' zone in the cerebral vasculature (Fig. 16). The location of metastatic lesions generally follows the volume of blood flow, with 90 per cent in the supratentorial region and 10 per cent developing infratentorially. The exceptions to this are the retroperitoneal cancers, such as uterine, cervical and bladder, which frequently spread to the posterior fossa, attributed to tumor infiltration of Batson's venous plexus in the pelvis. Renal-cell carcinoma, although retroperitoneal, does not preferentially spread to the posterior fossa. A second potential mechanism of spread to the brain is via the subarachnoid space (leptomeningeal dissemination). The detection of tumor nodules adjacent to cerebrospinal fluid, such as in brain sulci or in the wall of a ventricle, suggests that the tumor is from a deposit evolving from leptomeningeal dissemination (Fig. 17).

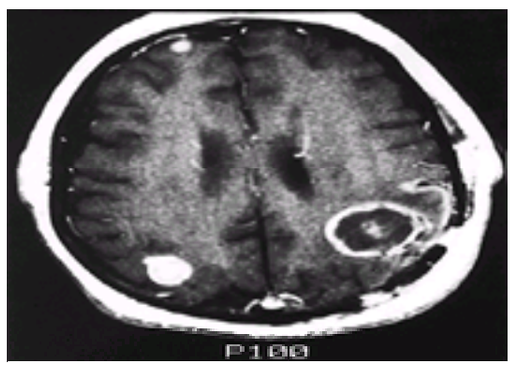


Fig. 16. Multiple metastases in a 52-year-old man with metastatic malignant melanoma; the lesions are at the gray–white junction.

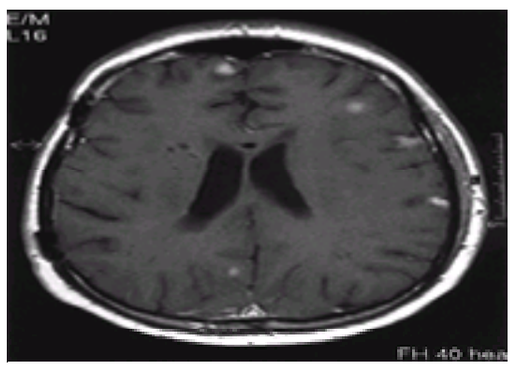


Fig. 17. Multiple brain metastases in a 54-year-old man with metastatic melanoma; the tumor nodules arise from sulci, indicating that these lesions arise from a deposit of tumor from the leptomeninges.

Diagnostic evaluations

Clinical symptoms and signs are dependent upon the location of the lesion(s) and the rapidity of growth or change in size. A slow-growing tumor may cause only subtle neurologic or cognitive changes despite achieving a large size, whereas a rapidly growing tumor or acute hemorrhage may cause sudden neurologic symptoms. Multiple lesions are more likely to cause changes in the sensorium or cognitive function.

Radiologic imaging is essential in diagnosing brain metastases. Contrast-enhanced MRI has proved to be the most effective modality (Fig. 16). Double-dose contrast CT, although less sensitive, can also be used. Nearly all brain metastases will enhance with contrast material, although very small lesions may be detected only as a T_2 signal abnormality on MRI.

The need for pathologic confirmation of metastatic tumor depends upon the clinical setting. Patients with multiple enhancing brain lesions and a known systemic malignancy generally do not require biopsy. Patients without known cancer should have pathologic confirmation prior to treatment. Patients with established systemic cancer and a single brain lesion should have either a resection or stereotactic biopsy to confirm the diagnosis. A recent series demonstrated that 11 per cent of patients with known systemic cancer and a solitary enhancing brain lesion did not have a brain metastases.

Treatment

Supportive care includes the use of corticosteroids to decrease peritumoral edema. Anticonvulsants are generally reserved for patients with known seizures or with

metastatic lesions from melanoma, renal-cell carcinoma or choriocarcinoma, tumors that have a high likelihood of spontaneous hemorrhage.

The treatment of brain metastases is determined by several critical factors: (1) number of lesions; (2) extent of the systemic malignancy; and (3) performance status. Several other factors, including the responsiveness of the tumor to chemotherapy and/or radiation, are also potentially important.

Patients with a solitary metastases (one brain lesion, no active systemic cancer) benefit from surgical resection of the tumor and would likely benefit from radiosurgery if the lesion is not surgically accessible. Current recommendations include postoperative whole-brain radiotherapy. Patients with a single brain metastasis but active systemic disease may benefit from resection or radiosurgery if the anticipated survival from the systemic disease exceeds 4 to 6 months.

Patients with multiple brain metastases are commonly treated with whole-brain radiotherapy. Surgical resection of one of the lesions, if it is large and causing mass effect or compromising neurologic function, may be indicated. No consensus exists on the optimal treatment schema, although most treatment programs administer 3000 cGy to the whole brain in 300-cGy fractions. Some radiotherapists recommend a more prolonged course (4400–5000 cGy) in 200-cGy fractions for patients with a better prognosis (age under 60 years, limited systemic disease, Karnofsky performance status above 70). Although these treatment strategies have similar efficacy, the convenience of fewer treatments is offset by an increase in risk for neurotoxicity, especially in patients with a better prognosis. Radiosurgery is now being investigated as a treatment for patients with three or fewer metastatic lesions. There are reports of surgical resection of multiple metastases, but the efficacy has been variable.

Chemotherapy is reportedly effective in patients with sensitive tumors. There are reports of response rates greater than 50 per cent for combined chemotherapy treatment of germ-cell tumors, lymphomas, breast cancer, and small-cell lung cancer. Delivery of the chemotherapeutic drugs to the brain lesions has been adequate, likely because of the local disruption of the blood–brain barrier, evident by contrast enhancement.

Outcome

The recent advances in imaging modalities as well as radiotherapy and surgical techniques has improved the outcome for patients with brain metastases. Early reports state an average survival of 1 month with no treatment, 2 months with corticosteroids and supportive care. Current treatment approaches have resulted in improved control of brain metastases; less than half of the patients die as a result of the brain lesions. Newer approaches to treatment, such as local chemotherapy or brachytherapy, and expanding the role of radiosurgery, will likely further improve treatment and outcome for these patients.

Primary lymphoma of the central nervous system

Introduction

Primary lymphoma of the central nervous system was once considered a rare neoplasm, but there has been a marked increase in its incidence in both immunocompetent and immunocompromised patients. Current epidemiologic studies report a threefold increase in incidence in immunocompetent individuals over the past 10 years, and a 4 to 6 per cent incidence in patients with acquired immune deficiency syndrome (**AIDS**).

The origin of the tumor cells is uncertain. Hypotheses include the malignant transformation of a lymphocyte clone that has cell-surface adhesion molecules specific for central nervous infiltration or that systemic immunity limits an otherwise systemic lymphoma to the central nervous system. Rarely does the tumor spread outside the central nervous system.

The majority of primary lymphomas of the central nervous system are of B-cell origin. The histologic classifications commonly seen include diffuse large-cell, immunoblastic, lymphoblastic, or small non-cleaved cells. The pathogenesis of the tumor is uncertain and no genetic predisposition has yet been established. In patients with immunodeficiency, a high percentage of tumors have a portion of the Epstein–Barr virus in the tumor cell genome; this is only rarely found in primary lymphoma of the central nervous system from immunocompetent patients.

Diagnosis

Patients often present with global neurologic symptoms including headache and alterations in mentation or level of conscious. The signs and symptoms at presentation in immunocompetent patients is somewhat different from patients with AIDS-related primary lymphoma of the central nervous system (Table 3). The symptoms often develop subacutely, over days to weeks, underscoring the rapid growth of the tumor. A complete physical examination is essential to exclude evidence of systemic lymphoma, which would preclude the diagnosis of primary lymphoma of the central nervous system.

Symptom	Immunocompetent patients (%)	Patients with AIDS (%)
Focal neurologic deficits	56	51
Mental status changes	34	53
Seizures	11	27
Increased intracranial pressure	32	14

Table 3 Signs and symptoms of lymphoma of the central nervous system

Brain imaging by contrast-enhanced CT or MRI is essential for making the diagnosis. More than 80 per cent of these tumors are enhancing; 50 per cent show a homogeneous enhancing pattern in contradistinction to metastatic cancer or primary glial neoplasms, where a ring enhancing pattern is most common (Fig. 18). The tumor is often multifocal and, although it can develop in all regions of the central nervous system, the most common location is periventricular.

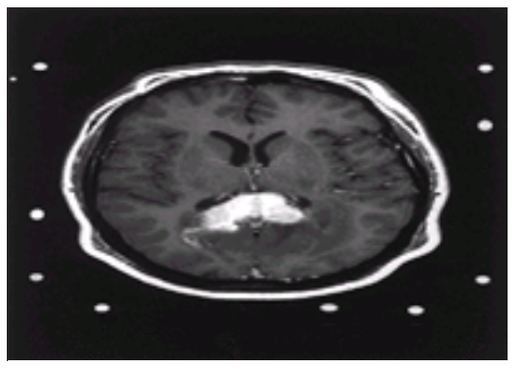


Fig. 18. Axial T_1-weighted gadolinium-enhanced MRI showing a homogeneously enhancing, primary lymphoma of the central nervous system presenting in the splenium of the corpus callosum.

Systemic evaluation for evidence of tumor outside the central nervous system is controversial. Some advocate CT scans of the chest, abdomen and pelvis, along with examination of bone marrow. Others argue that the yield is extremely low and therefore that extensive testing is unnecessary.

Pathologic diagnosis is critical to confirm the presence of lymphoma. The tumor is characteristically vasocentric, with infiltration of the blood-vessel wall (Fig. 19). Currently, stereotactic biopsy is the preferred method of obtaining tissue. The diagnosis can be confirmed from the small sample of tissue using immunohistochemistry to prove the monoclonal nature of the cells. Extensive resection has not improved outcome, but is associated with loss of neurologic function. These tumors can be extremely responsive to corticosteroids; reports indicate a 20 per cent rate of complete tumor resolution with steroid treatment. Subsequent biopsy then reveals only necrosis and gliosis. Therefore, in suspected cases of primary lymphoma of the central nervous system, steroids should be withheld until the biopsy is complete unless tumor-associated mass effect is severe.

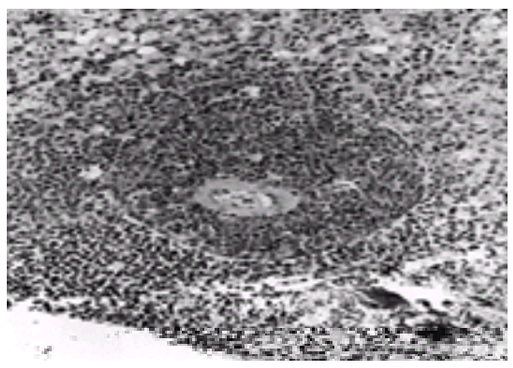


Fig. 19. Histopathology of lymphoma of the central nervous system. The tissue section is stained with hematoxylin–eosin, demonstrating the infiltration of malignant lymphocytes into the wall of a blood vessel and surrounding brain parenchyma.

Staging requires determination of the extent of involvement of the central nervous system. The tumor has a high propensity to infiltrate into the leptomeninges (60–100 per cent) and, although cranial-nerve palsies and other signs of leptomeningeal tumor are unusual at the time of diagnosis, failure to treat this tumor adequately will result in future relapse. In addition, ocular involvement is common, with infiltration of the vitreous. Therefore, all patients should undergo slit-lamp examination with possible vitreal biopsy to evaluate for tumor.

Management

As mentioned above, surgical resection has little role in the management of primary lymphoma of the central nervous system. External-beam radiotherapy has been proved to benefit patients; median survival increases from 2 to 3 months with supportive care to 12 months with radiotherapy. Many treatment algorithms exist, although most recommend whole-brain radiotherapy using 4000 to 5000 cGy. Some treatment plans also include a boost of 1200 to 1500 cGy to the tumor bed.

Chemotherapy has been shown to benefit patients with primary lymphoma of the central nervous system. The initial studies utilized chemotherapy for patients with recurrent disease after radiotherapy; more recently, preradiation chemotherapy has been utilized. All studies reported to date are small series using a variety of regimens. No prospective, randomized (phase III) study has been completed. The current regimens incorporate treatment for tumor behind the blood–brain barrier and tumor in the leptomeninges. Often, therapy requires direct intrathecal delivery by lumbar puncture or intraventricular delivery via a reservoir and intraventricular catheter system. Treatment of ocular involvement is not well established. Some chemotherapeutic agents (methotrexate, cytosine arabinoside) penetrate the vitreous, but most protocols recommend subsequent radiotherapy to the posterior two-thirds of the eye.

The optimal treatment for primary lymphoma of the central nervous system is evolving. Multiple studies suggest that there are several prognostic factors such as age (under 50 years), absence of immunodeficiency state, and good performance status that correlate with better outcome. Several series report survival in excess of 5 years in some patients with good prognostic factors. Some current studies are examining the role of chemotherapy, withholding radiotherapy until there is tumor relapse in an effort to reduce the neurotoxicity of the radiation treatment.

Further reading

General information

Burger PC. Classification, grading and patterns of spread of malignant gliomas. In: Apuzzo MLJ, ed. *Malignant cerebral glioma*, pp. 3–18. American Association of Neurological Surgeons, Park Ridge IL, 1990.

Jaeckle KA. Clinical presentation and therapy of nervous system tumors. In: Bradley WG, Darhoff RB, Fenichel GM, Marsden CD, ed. *Neurology in clinical practice*, pp. 1008–30. Butterworth-Heinemann, Boston, 1991.

Kleihues P, Burger PC, Scheithauer BW. *World Health Organization international histological classification of tumours. Histological typing of tumours of the central nervous system*, 2nd edn. Springer, Berlin, 1993.

Kleihues P, Cavenee WK. WHO classification of tumors of the nervous system. In: Kleihues P, Cavenee WK, eds. *World Health Organization classification of tumours. Pathology and genetics. Tumours of the nervous system*. IARC Press, Lyon, 2000

Kraemer DL, Bullard DE. Clinical presentation of the brain tumor patient. In: Morantz RA, Walsh JW, eds. *Brain tumors: a comprehensive text*, pp. 183–212. Marcel Dekker, New York, 1994.

Leeds NE, Kumar AJ. Diagnostic imaging: magnetic resonance imaging and computed tomography. In: Levin VA, ed. *Cancer in the nervous system*, pp. 13–49. Churchill Livingstone, New York, 1996.

Wen PY, Schiff D. Clinical evaluation of patients with astrocytomas. In: Black PM, Schoene WC, Lampson LA, ed. *Contemporary issues in neurological surgery: astrocytomas: diagnosis, treatment, and biology*, pp. 26–36. Blackwell Scientific, Boston, 1993.

Primary brain tumors

Evans RG. The role of radiation therapy in the treatment of adult brain tumors. In: Morantz RA, Walsh JW, ed. *Brain tumors*, pp. 643–57. Marcel Dekker, New York, 1994.

Loeffler JS, Alexander E, III. The role of radiotherapy in the management of astrocytomas. In: Black PM, Schoene WC, Lampson LA, ed. *Contemporary issues in neurological surgery. astrocytomas: diagnosis, treatment and biology*, pp. 73–85. Blackwell Scientific, Boston, 1993.

Maciunas RJ. Adjunctive management of gliomas. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 671–82. Williams and Wilkins, Baltimore, 1996.

Roth JG, Elvidge AR. Glioblastoma multiforme: a clinical survey. *Journal of Neurosurgery* 1960; **17**: 736–50.

Salcman M. Glioblastoma and malignant astrocytoma. In: Kaye AH, Laws ER, Jr. ed. *Brain tumors: an encyclopedic approach*, pp. 449–78. Churchill Livingstone, Edinburgh, 1995.

Astrocytoma

Berger MS, Leibel SA, Bruner JM, Finlay JL, Levin VA. Primary central nervous system tumors of the supratentorial compartment. In: Levin VA, ed. *Cancer in the nervous system*, pp. 57–126. Churchill Livingstone, New York, 1996.

Kepes JJ, Louis DN, Paulus W. Pleomorphic xanthoastrocytoma. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 34–6. International Agency for Research on Cancer, Lyon, 1997.

Kleihues P, Davis RL, Ohgaki H, Cavenee WK. Low-grade diffuse astrocytomas. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 10–13. International Agency for Research on Cancer, Lyon, 1997.

Weistler OD, Lopez BS, Crino PB. Tuberous sclerosis and subependymal giant cell astrocytoma. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 182–4. International Agency for Research on Cancer, Lyon, 1997.

Oligodendroglioma and anaplastic oligodendroglioma

Bullard DE. Surgical aspects and general management of oligodendrogliomas. In: Apuzzo MLJ, ed. *Benign cerebral glioma*, Vol. II, pp. 397–412. American Association of Neurological Surgeons, Park

Ridge IL, 1995.

Reifenberger G, Kros JM, Schiffer D, Collins VP. Oligodendroglioma, anaplastic oligodendroglioma. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 38–44. International Agency for Research on Cancer, Lyon, 1997.

Shaw EG, Scheithauer BW, O'Fallon JR, Tazelaar HD, Davis DH. Oligodendrogliomas: the Mayo Clinic experience. *Journal of Neurosurgery* 1992; **76**(3): 428–34.

Ependymoma

Constantini S, Epstein FJ. Primary spinal cord tumors. In: Levin VA, ed. *Cancer in the nervous system*, pp. 127–37. Churchill Livingstone, New York, 1996.

Duncan JA, III, Hoffman HJ. Surgical aspects and general management of ependymomas. In: Apuzzo MLJ, ed. *Benign cerebral glioma*, Vol. II, pp. 413–26. American Association of Neurological Surgeons, Park Ridge IL, 1995.

Medulloblastoma

Berger MS, Magrassi L, Geyer R. Medulloblastoma and primitive neuroectodermal tumors. In: Kaye AH, Laws ER, Jr, ed. *Brain tumors*, pp. 561–74. Churchill Livingstone, Edinburgh, 1995.

Giangaspero F, Bigner SH, Giordana MT, Kleihues P, Trojanowski JQ. Medulloblastoma. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 96–103. International Agency for Research on Cancer, Lyon, 1997.

Wisoff JH, Epstein JH. Management of pediatric brain tumors. In: Morantz RA, Walsh JW, ed. *Brain tumors: a comprehensive text*, pp. 581–611. Marcel Dekker, New York, 1994.

Neuronal and mixed neuronal and glial tumors: ganglioglioma and neurocytoma

Chen TC, Gonzalez-Gomez I, McComb JG. Uncommon glial tumors. In: Kaye AH, Laws ER, Jr, ed. Brain tumors, pp. 525–57. Churchill Livingstone, Edinburgh, 1995.

Silver JM, Rawlings CE, Rossitch ER, Zeidman SM, Friedman AH. Gangliomas: a clinical study with long-term follow-up. *Surgical Neurology* 1991; **35**: 261–6.

Chordoma

Eriksson B, Gunterberg B, Kindblom LG. Chordoma, a clinicopathologic and prognostic study of a Swedish national series. *Acta Orthopaedica Scandinavica* 1981; **52**: 49–58.

Forsyth PA *et al.* Intracranial chordomas: a clinicopathological and prognostic study of 51 cases. *Journal of Neurosurgery* 1993; **78**: 741–7.

Gay E, Sekhar LN, Wright DC. Chordomas and chondrosarcomas of the cranial base. In: Kaye AH, Laws ER, Jr, ed. *Brain tumors*, pp. 777–94. Churchill Livingstone, Edinburgh, 1995.

Sundaresan N, Krol G, Hughes JEO, Hough L. Tumors of the spine: diagnosis and management. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 1303–22. Williams and Wilkins, Baltimore, 1996.

Esthesioneuroblastoma

Ebersold MJ, Olsen KD, Roote RL, Buckner JC, Quast LM. Esthesioneuroblastoma. In: Kaye AH, Laws ER, Jr, ed. *Brain tumors*, pp. 825–38. Churchill Livingstone, Edinburgh, 1995.

Hyams VJ, Batsakis JG, Michaels L. Tumors of the upper respiratory tract and ear. *Atlas of tumor pathology*, 2nd series, Fascicle 25, pp. 240–8. Armed Forces Institute of Pathology, Washington DC, 1988.

Nasopharyngeal carcinoma

Cacayorin ED, Petro GR, Kieffer SA. Radiology of skull-base neoplasms. In: Long DM, ed. *Contemporary issues in neurological surgery: surgery for skull base tumors*, pp. 1–30. Blackwell Scientific, Boston, 1992.

Danks RA, Kaye AH. Carcinoma of the paranasal sinuses. In: Kaye AH, Laws ER, Jr, ed. *Brain tumors*, pp. 809–24. Churchill Livingstone, Edinburgh, 1995.

Rousseau GL, Sekhar LN. Surgery of cranial base tumors. In: Morantz RA, Walsh JW, ed. *Brain tumors: a comprehensive text*, pp. 451–91. Marcel Dekker, New York, 1994.

Glomus tumor

Batnitzky S, Eckard DA. The radiology of brain tumors: general considerations and neoplasms of the posterior fossa. In: Morantz RA, Walsh JW, ed. *Brain tumors: a comprehensive text*, pp. 213–72. Marcel Dekker, New York, 1994.

Ebersold MJ, Morita A, Olsen KD, Quast LM. Glomus jugulare tumors. In: Kaye AH, Laws ER, Jr, ed. *Brain tumors*, pp. 795–807. Churchill Livingstone, Edinburgh, 1995.

P>**Holliday MJ.** A two-staged planned removal of glomus tumors. In: Long DM, ed. *Contemporary issues in neurological surgery: surgery for skull base tumors*, pp. 247–53. Blackwell Scientific, Boston, 1992.

Soffer D. Paraganglioma. In: Kleihues P, Cavenee WK, ed. *Tumours of the nervous system. Pathology and genetics*, pp. 80–2. International Agency for Research on Cancer, Lyon, 1997.

Neurofibromas and schwannomas

Eisenberg MB, Catalano PJ. Management of acoustic schwannomas. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 995–1004. Williams and Wilkins, Baltimore, 1996.

Freeman TB, Cahill DW. Management of intradural extramedullary tumors. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 1323–34. Williams and Wilkins, Baltimore, 1996.

Pineal tumors

Apuzzo MLJ. *Surgery of the third ventricle*. Williams and Wilkins, Baltimore, 1987.

Bruce JN, Stein BM. Management of pineal tumors. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 875–87. Williams and Wilkins, Baltimore, 1996.

Neuwelt EA, O'Neill OR, Roman-Goldstein S, Shults T. Clinical features and imaging of pineal region tumors. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 857–74. Williams and Wilkins, Baltimore, 1996.

Meningiomas and choroid plexus tumors

Al-Mefty, O. *Operative atlas of meningiomas*. Lippincott Raven, Philadelphia, 1998.

DeMonte F, Al-Mefty, O. Management of meningiomas. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 691–714. Williams and Wilkins, Baltimore, 1996.

Halliday Al, Harsh GR, IV. Meningiomas: clinical features. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 683–90. Williams and Wilkins, Baltimore, 1996.

Brain metastases

Berk L. An overview of radiotherapy trials for the treatment of brain metastases. *Oncology* 1995; **9**: 1205–19.

Black PM. Solitary brain metastases. Radiation, resection or radiosurgery? *Chest* 1993; **103** (Suppl.): 367–9S.

Buckner JC. The role of chemotherapy in the treatment of patients with brain metastases from solid tumors. *Cancer and Metastases Reviews* 1991; **10**: 335–341.

Coia LR, Aaronson N, Linggood R, Loeffler J, Priestman TJ. A report of the consensus workshop panel on the treatment of brain metastases. *International Journal of Radiation, Oncology, Biology, Physics* 1992; **23**: 223–7.

DeAngelis LM. Management of brain metastases. *Cancer Investigation* 1994; **12**: 156–65.

Patchell RA. The treatment of brain metastases. *Cancer Investigation* 1996; **14**: 169–77.

Shiau C-Y *et al.* Radiosurgery for brain metastases: relationship of dose and pattern of enhancement to local control. *International Journal of Radiation, Oncology, Biology, Physics* 1997; **37**: 375–83.

Primary central nervous system lymphoma

DeAngelis LM. Primary central nervous system lymphoma. *Recent Results in Cancer Research* 1994; **135**: 155–69.

DeAngelis LM. Current management of primary central nervous system lymphoma. *Oncology* 1995; **9**: 63–71.

Ferreri AJM, Reni M, Villa E. Primary central nervous system lymphoma in immunocompetent patients. *Cancer Treatment Reviews* 1995; **21**: 415–46.

Fine HA, Mayer RJ. Primary central nervous system lymphoma. *Annals of Internal Medicine* 1993; **119**: 1093–104.

Reni M, Ferreri AJM, Garancini MP, Villa E. Therapeutic management of primary central nervous system lymphoma in immunocompetent patients: results of a critical review of the literature. *Annals of Oncology* 1997; **8**: 227–234.

44.4 Surgery for epilepsy

C. B. T. Adams

Surgical procedure
Further reading

The possibility of surgery should be considered in any patient whose epilepsy is not controlled by medication. Many patients whose epilepsy could be successfully treated are not referred for surgery. This under-referral is not due to the ignorance of neurologists and paediatricians only. There is a widespread feeling that such surgery is complex, daunting for the patients, and that the results are uncertain. However, modern imaging techniques have made recognizing epileptogenic but normal cortex or ‘spike chasing' a thing of the past. Good results depend on finding and removing focal pathology; a surgical scar is not usually epileptogenic.

Patients suitable for consideration of epileptic surgery include those resistant to drug therapy, and those with a consistent type of epileptic pattern. If a seizure always follows a constant type of warning aura, the clinician should seriously consider referring the patient for specialist evaluation. Constancy of the initial phase of the seizure is suggestive of a focal origin to the seizure. Surgical treatment is also indicated if a lesion is found on imaging. Most surgery for drug-resistant epilepsy is performed on the temporal lobe. Scanning in the line of the temporal horn discloses lesions not seen with conventional scans that cut through the temporal lobe. Thus ‘temporal orientated' computed tomographic (**CT**) scanning is an important investigation. Magnetic resonance imaging may also reveal small epileptogenic lesions, and is being increasingly used in evaluating patients for epileptic surgery. Electroencephalography also has a place in the investigation of patients, particularly when no lesion is seen on imaging. Specialized techniques using implanted electrodes are, however, of debatable value. They are probably applicable to only a small percentage of patients, in whom the results of surgery are less certain.

The results of surgery are best in young patients treated before schooling has ceased. Teenagers with epilepsy can become recluse; surgery renders approx. 60 to 70 per cent of patients seizure-free, and allows social and scholastic rehabilitation.

H3>Surgical procedure

Four types of procedure are carried out: temporal lobectomy (70 per cent) (Fig. 1, Fig. 2); extratemporal cortical excisions (25 per cent); hemispherectomy (2 per cent); and callosotomy (3 per cent). Seizures arising from the temporal lobe are due either to a small, indolent glioma (Fig. 3) or sclerosis of Ammon's horn. The latter is thought to arise following a prolonged (more than 30 min) febrile convulsion in a young child (up to 4 years of age). Prompt treatment of febrile convulsions is therefore important. Sclerosis of the amygdala and hippocampus (Ammon's horn) gives rise to psychomotor or temporal-lobe epilepsy several years later (about the age of 6 or 7 years). The amygdala and hippocampus seem to be particularly involved with the genesis of epilepsy, and excision of these structures together with any pathological areas seems to produce better seizure relief. More recently, amygdalohippocampectomy has been introduced. This entails removing the structures without removing the remaining neocortex of the temporal lobe (Fig. 4, Fig. 5). Extratemporal cortical excisions must be performed under electrocorticographic guidance.

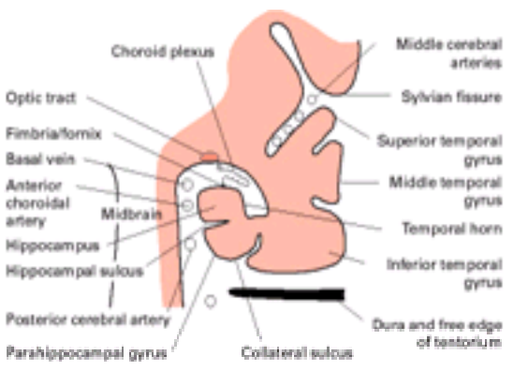


Fig. 1. The relevant anatomy adjacent to the temporal lobe: it is essential to preserve the arachnoid covering the brainstem and associated vessels when removing the medial structures of the lobe.

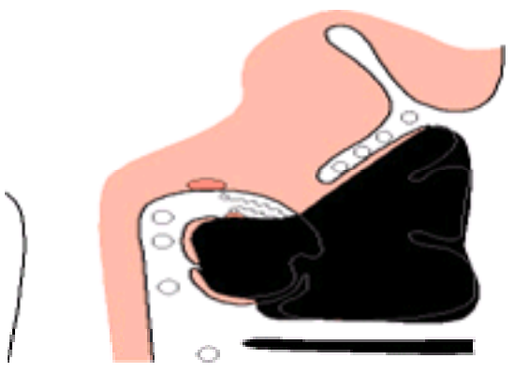


Fig. 2. A coronal section showing the extent of removal of the temporal lobe during *en bloc* removal.

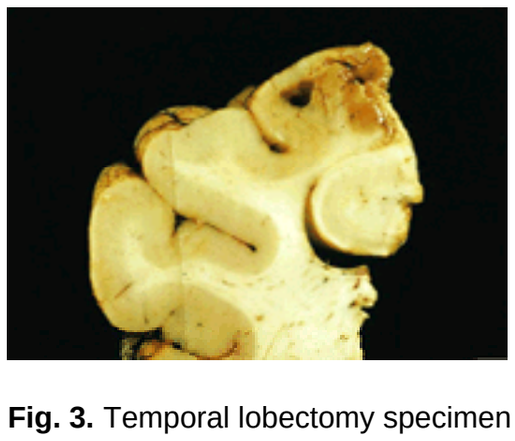


Fig. 3. Temporal lobectomy specimen showing the hippocampus projecting into the temporal horn. There is an astrocytoma in the parahippocampal gyrus.

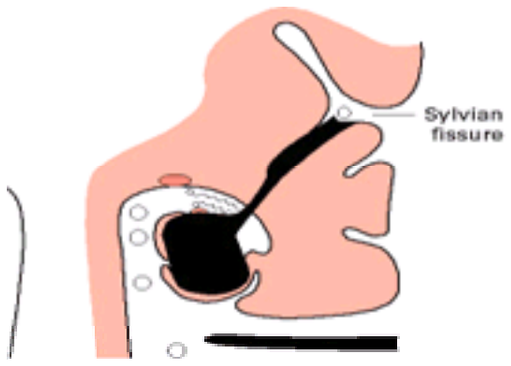


Fig. 4. Selective amygdalohippocampectomy using Yasugil's pterional approach. Incising the cortex is avoided, but the approach is restrictive.

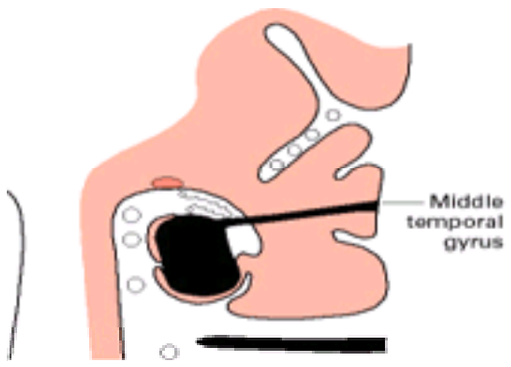


Fig. 5. The selective amygdalohippocampectomy after Neimyer. The relative advantages and disadvantages of these approaches are still unclear.

Hemispherectomy has been described as the best operation for epilepsy and is indicated for children with the epilepsy of infantile hemiplegia. Long-term bleeding complications curtailed its use, but the procedure has been modified and is now being performed more often (Fig. 6, Fig. 7 and Fig. 8).

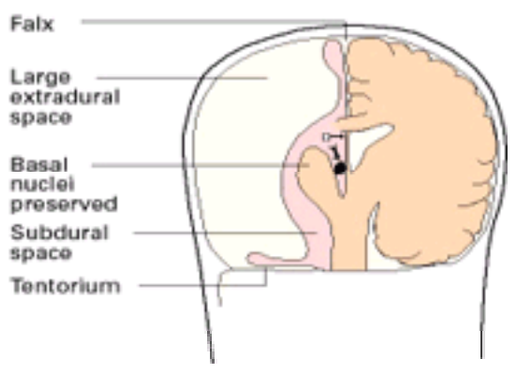


Fig. 6. Modification of hemispherectomy by creating a large extradural space. Boxed arrow emphasizes the importance of preserving the septum pellucidium. Double arrow shows the foramen of Monro obstructed.

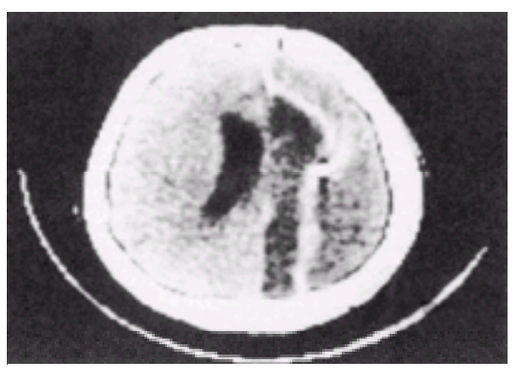


Fig. 7. Postoperative CT scan to show the different fluid densities in the lateral ventricles, operative (subdural) cavity, and the extradural cavity. The dura is highlighted because of a lining membrane of blood.

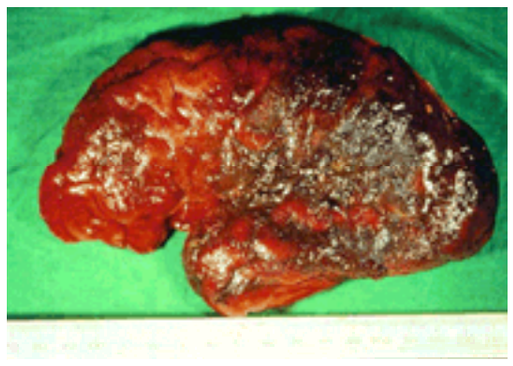


Fig. 8. Hemispherectomy specimen showing Sturge–Weber angiomatosis of the cortex.

The indications for sectioning of the corpus callosum are much less certain. This procedure, although fashionable, will probably not survive the test of time, in common with other operations for epilepsy that do not entail finding and removing focal pathology.

Further reading

Adams C. *A neurosurgeon's notebook*. Blackwell Scientific, Oxford, 1998. [Philosophy of surgery, surgical judgement, and technical tips.]

Apuzzo MLJ. *Brain surgery, complication avoidance and management*. Churchill Livingstone, London, 1993. [Two-volume detailed account of surgery.]

Luders HO. *Epilepsy surgery*. Raven, New York, 1992. [Detailed account of epilepsy surgery.]

Oxbury JM, Adams CBT. Neurosurgery for epilepsy. *British Journal of Hospital Medicine* 1989; **41**: 372–8. [A practical approach to epilepsy surgery.]

Rengachary SS, Wilkins RH. *Principles of neurosurgery*. Mosby-Wolfe, St Louis, 1994. [General account of neurosurgery for beginners.]

44.5 Surgery for Parkinson's disease and other movement disorders

Roy A. E. Bakay, Mahlon R. DeLong and Jerrold L. Vitek

Definitions
Parkinson's disease
Non-parkinsonian tremor
Dystonia
Hemiballismus
Huntington's disease
History
Physiology of movement disorders: basal ganglia model
Ablative surgery: indications and results
Pallidotomy for Parkinson's disease
Pallidotomy for other movement disorders
Thalamotomy for tremor
Thalamotomy for other conditions
Deep brain stimulation
Indications and results
Comparison with lesioning
Restorative therapy
Adrenal chromaffin autografts
Fetal ventral mesencephalic allografts
Dopamine-secreting xenografts
Genetically engineered autologous tissue
Summary
Further reading

There are currently three major foci in the surgery of movement disorders: ablative procedures, deep brain stimulation, and ‘restorative' therapies using cell transplantation or the intracerebral delivery of growth factors. All of these alter function in the basal ganglia and/or cerebellothalamic pathways. Ablative surgery and deep brain stimulation attempt to compensate for, rather than correct, the cellular or biochemical brain abnormality causing the disordered movement. ‘Restorative' therapies attempt to correct the cellular or biochemical brain defect that is the basis for the disorder, by replacing or promoting the survival of degenerating cell populations.

Movement disorders that may be considered for surgery of the central nervous system include Parkinson's disease, non-parkinsonian tremor, dystonia, hemiballismus, Huntington's disease, and tardive syndromes, i.e., tardive dystonia and tardive dyskinesias.

Definitions

Parkinson's disease

Parkinson's disease is a progressive, neurodegenerative disease involving the basal ganglia. The cardinal clinical signs are resting tremor, rigidity, and bradykinesia. The fundamental biochemical defect is loss of dopamine in the striatum and in other basal ganglia nuclei, due to loss of the dopaminergic cells of the substantia nigra. Recently, genetic defects have been identified in familial forms of Parkinson's disease.

Non-parkinsonian tremor

Non-parkinsonian tremors include essential tremor and the cerebellar-outflow tremors, which are commonly due to trauma, stroke, or multiple sclerosis. Essential tremor is the most common of these and the most amenable to surgical treatment. It is characterized by a postural tremor, predominantly including the upper extremities. Its pathophysiologic basis is unknown, but presumably involves cerebello-thalamic pathways. Many cases appear to be inherited in an autosomal-dominant pattern.

Dystonia

The term dystonia may refer to a symptom or a particular disease. Dystonia is characterized by sustained involuntary muscle contractions leading to twisting, repetitive movements, and abnormal postures, with underlying cocontraction of agonist and antagonist muscles, and ‘overflow' into other muscle groups. Primary (idiopathic) forms of dystonia are often due to single-gene defects, and their genetic bases are under active investigation. Dystonia may also be secondary to traumatic, neoplastic, or infectious lesions involving the basal ganglia (usually the putamen) or less commonly the motor thalamus.

Hemiballismus

Hemiballismus is characterized by violent, uncontrollable ballismistic movements of the proximal arm and leg on one side of the body. It is usually due to ischemic lesions involving the subthalamic nucleus and/or its efferent pathways.

Huntington's disease

Huntington's disease is an autosomal-dominant degenerative disorder. In early stages, its cardinal motor sign is chorea, whereas later stages are often characterized by an akinetic rigid state. Depression, psychotic behavior, and dementia often result in greater disability than the motor signs. Its pathologic hallmark is degeneration of striatal-output neurones. Huntington's disease is due to expansion of a trinucleotide repeat region on chromosome 4 that encodes for the protein huntingtin, but it is not know how this gene defect produces the observed pathology and clinical signs.

History

The earliest treatments of movement disorders involved surgical lesioning of motor pathways. Beginning in the 1940s, before a solid scientific basis for these procedures was established, surgical ablations of regions of the globus pallidus (pallidotomy) and motor thalamus (thalamotomy) were performed empirically. Pallidotomy has a long history in the treatment of Parkinson's disease and other movement disorders. Russell Myers pioneered surgery on the basal ganglia, but results were inconsistent and often associated with significant morbidity. In the 1950s, the introduction of stereotactic techniques decreased the morbidity and increased the effectiveness of pallidotomy. Leksell identified the posterior pallidum as a more effective target than the anterior target used by other neurosurgeons. Following Irving Cooper's and R. Hassler's demonstration that lesions in the cerebellar receiving area of the thalamus (nucleus ventralis intermedius) were more effective than pallidotomy in stopping tremor, pallidotomy was largely replaced by thalamotomy. In the 1960s, L-dopa therapy for Parkinson's disease became widespread and was highly successful in treating the motor symptoms. As a result, surgical therapy of all types fell out of favor. However, the effectiveness of L-dopa typically declines with time, and debilitating motor fluctuations and drug-induced dyskinesias often ensue after 3 to 5 years. The response to medical therapy becomes highly unpredictable and symptoms are poorly controlled. As a result of advances in surgical techniques and in the theoretical understanding of the rationale for different surgical interventions, there has been a re-exploration of surgery for the treatment of movement disorders.

In considering current surgical options, historical perspective is of limited value. It is difficult to compare surgical studies from the time before the availability of L-dopa with contemporary procedures. Before L-dopa, standard parkinsonian rating scales to assess clinical outcome were not available, rigorous long-term follow up was rare, and documentation of lesion location in the brain was possible only by autopsy, since computed tomography and magnetic resonance imaging (**MRI**) did not exist. Also, the patient population before L-dopa was different in several respects. It included many patients with postencephalitic parkinsonism, which is now rare. Moreover, patients with Parkinson-plus syndromes were not differentiated from those with idiopathic Parkinson's disease. Finally, patients with Parkinson's disease before L-dopa did not suffer from drug-induced, on–off fluctuations or dyskinesias, which are now a major source of morbidity. Another factor in comparing results is that the major focus of surgery was once tremor rather than akinesia/bradykinesia. Therefore, any discussion of the surgical options for Parkinson's disease must dwell upon contemporary reports, even though some of the procedures reviewed here (pallidotomy and thalamotomy) have been performed for many years.

There are currently three subcortical targets for ablative surgery and deep brain stimulation: the globus pallidus, the subthalamic nucleus, and the motor thalamus. Pallidotomy has seen a remarkable resurgence since the demonstration of its effectiveness for most parkinsonian motor signs, including bradykinesia. Thalamotomy continues to be performed primarily for tremor but is being replaced by deep brain stimulation of the nucleus ventralis intermedius of the thalamus, which has well-documented efficacy for tremor, similar to that of thalamotomy. More recently, the subthalamic nucleus has attracted attention as a possible stimulation target for Parkinson's disease. Deep brain stimulation in the internal segment of the globus pallidus and the subthalamic nucleus appears to treat most parkinsonian motor signs and may supplement or replace ablative therapies, given the reversibility of its side-effects and its safety even when used bilaterally. The optimal target locations and indications for each of these procedures have not been standardized.

Physiology of movement disorders: basal ganglia model

The cortex, basal ganglia, and thalamus form multiple segregated circuits controlling motor, limbic, and associative functions. The basal ganglia–thalamocortical motor circuit has a key role in regulating motor behavior. The cortical regions participating in the motor circuit are the motor, premotor, and supplementary motor cortices. The structures in the basal ganglia that participate in this circuit include portions of the striatum (putamen), the internal and external segments of the globus pallidus, the subthalamic nucleus, the motor thalamus, and the substantia nigra pars compacta and pars reticulata. The portions of these nuclei that participate in the motor circuit contain neurones whose discharge rates are modulated by passive joint movements or which have cells whose discharge is modulated by active movements, i.e., sensorimotor properties. Sensorimotor regions are anatomically separate from the regions that regulate non-motor functions. Because of this, it is possible to design surgical interventions that alter function in the motor circuit without directly affecting non-motor circuits.

The current model of the basal ganglia–thalamocortical motor circuit is illustrated schematically in Fig. 1(a). Cortical efferent projections enter the basal ganglia via the putamen. The major output nuclei in the basal ganglia are the internal segment of the globus pallidus and substantia nigra pars reticulata, which project to subdivisions of the motor thalamus and brainstem. Within the basal ganglia, motor function is thought to be regulated by two major pathways, both of which are modulated by striatal dopamine. The ‘direct' pathway is a monosynaptic projection from the putamen to the internal segment of globus pallidus/substantia nigra pars reticulata. In the ‘indirect' pathway, the putamen projects to the internal segment/pars reticulata via intermediate nuclei, the external segment of the globus pallidus and the subthalamic nucleus. Dopaminergic innervation of the putamen by the pars compacta inhibits transmission over striatopallidal fibers in the indirect pathway, but facilitates transmission over striatopallidal fibers in the direct pathway. The nondopaminergic projections between the basal ganglia nuclei are g-aminobutyric acid (**GABA**)-ergic (inhibitory), except for the projection from the subthalamic nucleus to the internal segment/pars reticulata, which is glutamatergic (excitatory). Balance between the direct and indirect pathways is considered important for facilitating cortically initiated movement (direct pathway) and suppressing (indirect pathway) unintended movements, and for scaling movement amplitude. An imbalance between the two pathways results in the abnormal activity in the basal ganglia–thalamocortical motor circuit, which is thought to be responsible for many of the motor abnormalities associated with movement disorders. Greater understanding of these alterations has allowed specialists in movement disorders to begin building a theoretical basis for surgical intervention.

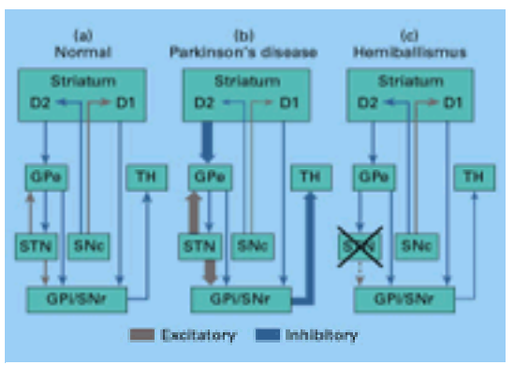


Fig. 1. Schematic diagram of the basal ganglia–thalamocortical motor circuit in the normal state and in movement disorders. *Abbreviations*: SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; GPe, globus pallidus, external segment; GPi, globus pallidus, internal segment; STN, subthalamic nucleus; HB, hemiballismus; PD, Parkinson's disease. Open arrows, excitatory connections; filled arrows, inhibitory connections. Changes in the widths of the arrows between (A), (B), and (C) indicate changes in the activity of the pathway represented. Wider lines indicate increased activity, narrower lines indicate decreased activity. (A) normal;(B) Parkinson's disease; (C) hemiballismus.

Movement disorders are often classified as hyperkinetic or hypokinetic, although the symptom complexes of a given disorder may not fit neatly into these categories. Figure 1(b), Figure 1(c) illustrates a model of how alterations in the basal ganglia–thalamocortical motor circuit may lead to hyper- or to hypokinetic movement disorders. Hemiballismus is illustrated as an example of a hyperkinetic disorder (Fig. 1(c)). Damage to the subthalamic nucleus results in decreased activity of the output nuclei in the basal ganglia and disinhibition of the thalamocortical pathway. Huntington's disease also fits into the hyperkinetic model. In early Huntington's disease, there is selective degeneration of neurones that give rise to the indirect pathway; this results in hypoactivity of the subthalamic nucleus, hypoactivity of the output nuclei, and hyperactivity of the thalamocortical pathway. The resulting cortical hyperactivity is manifested by hyperkinetic motor signs. In idiopathic torsion dystonia, discharge rates in the internal segment of the globus pallidus are also decreased, suggesting that dystonia may also be grouped with the hyperkinetic disorders.

Abnormalities in hypokinetic movement disorders, such as Parkinson's disease, are illustrated in Fig. 1(b). Loss of dopaminergic cells in the substantia nigra pars compacta, the fundamental defect in Parkinson's disease, results in abnormalities in neuronal activity in the basal ganglia–thalamocortical motor circuit. In the indirect pathway, loss of striatal dopamine leads to increased activity in the subthalamic nucleus and output nuclei, the internal segment of globus pallidus and substantia nigra pars reticulata, which in turn leads to excessive inhibition of the thalamocortical and brainstem pathways. The deficiency in cortical excitation is manifested by the hypokinetic motor signs of parkinsonism. This model of Parkinson's disease is supported by metabolic, pharmacologic, and electrophysiologic studies of the basal ganglia in the parkinsonian [1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine (**MPTP**)-treated] monkey and in patients with Parkinson's disease. In this model, neuronal activity in the internal segment of globus pallidus and the subthalamic nucleus is greater than normal, and inactivation of the subthalamic nucleus, whether via lesioning or injections of the GABA agonist muscimol, ameliorates all of the signs of Parkinson's disease in the contralateral limbs.

Not all movement disorders, however, can be described in terms of alterations in the basal ganglia–thalamocortical motor circuit. Projections from the output nuclei in the basal ganglia to brainstem nuclei such as the pedunculopontine nucleus may also be important in the pathophysiology of movement disorders, though the function of these projections is less well understood. Furthermore, other forms of tremor, i.e., post-traumatic or essential, reflect derangements in cerebellothalamocortical pathways.

Ablative surgery: indications and results

Pallidotomy for Parkinson's disease

The recent resurgence of interest in pallidotomy as a treatment for Parkinson's disease is due to multiple factors: (1) recognition that long-term medical therapy for Parkinson's disease is often unsatisfactory, with patients eventually suffering from drug-induced dyskinesias, motor fluctuations, and variable responses to medication; (2) greater understanding of the pathophysiology of Parkinson's disease, providing a rational theoretical basis for pallidotomy; and (3) the use of newer techniques, including computed tomographic- and MRI-guided stereotaxis and microelectrode recording, making surgical intervention in the basal ganglia more precise and safe (Fig. 2).

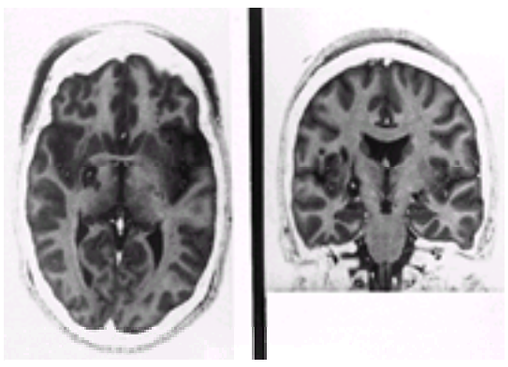


Fig. 2. Axial and coronal MRI of the brain, obtained 4 h after pallidotomy. Fast spin echo inversion recovery sequence; TI, 200 ms; TR, 3000 ms; TE, 40 ms; slice thickness, 2 mm; interslice spacing, 0. (A) Axial image at the level of the AC–PC line; (B) coronal image at the middle of the lesion–edema complex (dark outer ring and central white area). Note the edema crossing over to the thalamus representing the destruction of this pathway.

Indications

Candidates for pallidotomy should have a diagnosis of idiopathic Parkinson's disease, which requires the presence of at least two of the three cardinal motor signs of the disease (bradykinesia, rest tremor, and rigidity) and a well-documented beneficial response to L-dopa. Most candidates are moderately disabled, with Höhn and Yahr scores of 3 or more when off medication, have disabling drug-induced dyskinesias, and severe motor fluctuations. The majority of patients have rigidity and bradykinesia as predominant symptoms, although patients with tremor-dominant Parkinson's disease are also considered good candidates. There is no specific age cut-off. Contraindications to pallidotomy include dementia, extensive brain atrophy, and Parkinson-plus syndromes.

Outcome

Most contemporary studies cite significant improvements in parkinsonian motor signs following pallidotomy of the internal segment, with benefits sustained over the period of evaluation (varying from 6 months to 4 years). In one study of microelectrode-guided pallidotomy using blinded investigators (based on videotaped examinations), motor scores in the ‘off' state (examined before morning dose) had improved by 30 per cent at 6 months postoperatively. The bradykinesia score improved by 33 per cent; the gait score by 15 per cent. There was a nearly complete elimination of ‘on' (peak drug effect) dyskinesias on the side opposite to the lesion, although the patients continued to take the same doses of dopaminergic medications as before the surgery. Other centers report similar improvement in the ‘off' state and lack of ‘on' dyskinesias at 12 to 24 months following surgery. In addition, others report that tremor improved in the vast majority of patients (85 per cent) and ‘on' periods were prolonged. The long-term outcome from pallidotomy is less well documented. It is clear that some patients experience some loss of benefit over time; this may be related to inadequate lesion size or suboptimal location (discussed below), or to disease progression on the unoperated side.

Complications

In one of the first modern pallidotomy series there was a 14 per cent incidence of visual-field deficits due to injury to the optic tract, which passes below the target. Face or limb weakness is usually transient, but permanent paralysis has been reported, due to injury to the corticospinal tract, which passes medial to the target. Other major complications include symptomatic intraparenchymal hemorrhages along the lesioning tract, subdural empyema, subdural hematomas, and unexplained postoperative encephalopathy, especially in elderly patients. Death is rare. Other reported complications of pallidotomy include hypophonia or increased difficulty with speech articulation, which are uncommon after unilateral procedures but are common after bilateral pallidotomy. The overall incidence of serious, permanent complications is 1 to 2 per cent.

Bilateral pallidotomy

Following simultaneous or staged bilateral pallidotomy, the likelihood of significant speech deterioration (based on questionnaire data) is reported to be as high as 60 per cent. One group has reported that bilateral pallidotomy can produce major cognitive deficits. Even with staged bilateral pallidotomies, the majority of patients are hypophonic to varying degrees. Others have not observed significant cognitive or speech changes in these patients, although they have not performed formal neuropsychologic evaluation or speech testing. The safety of bilateral pallidotomy needs closer scrutiny before it can be routinely recommended.

Pallidotomy for other movement disorders

While medically intractable Parkinson's disease is the predominant indication for pallidotomy of the internal segment, the operation has been applied to other movement disorders as well. The first use of stereotactic surgery for a movement disorder was for chorea in a patient with Huntington's disease. The procedure was effective in decreasing choreiform movements contralateral to the lesion. It is unclear whether this could provide functional benefit to patients, and the brain atrophy associated with Huntington's disease increases the risk of postsurgical complications. Isolated case reports indicate that pallidotomy may be effective in alleviating the involuntary movements associated with hemiballismus. Recently, several centers have reported significant improvement in patients with idiopathic or secondary generalized dystonia. Careful clinical follow up is needed to document long-term effectiveness before the use of pallidotomy in the treatment of dystonia can be routinely recommended.

Thalamotomy for tremor

Thalamotomy was first shown to be effective for tremor by Cooper, who erroneously placed a lesion in the lateral thalamus while performing a pallidotomy. Hassler experimented with lesions of a variety of sizes and shapes within the thalamus. The motor thalamus is subdivided into several nuclei. From posterior to anterior, using the Hassler nomenclature, these are: ventralis intermedius, ventralis oralis posterior, and ventralis oralis anterior. Ventralis oralis anterior and posterior are the receiving areas for the basal ganglia, while ventralis intermedius is the cerebellar receiving area. For many years, thalamotomy was the mainstay of surgical treatment for a variety of symptom complexes in several movement disorders, including parkinsonian tremor and rigidity, non-parkinsonian tremor, and dystonias.

The most accepted contemporary target of thalamotomy for tremor is the nucleus ventralis intermedius (Fig. 3). Microelectrode recording in that nucleus in patients with parkinsonian, essential, or cerebellar tremor has revealed cells that discharge in bursts and are time-locked to the patient's tremor, suggesting that tremor of several causes is associated with abnormal discharge in the cerebellothalamic pathway and that interruption of this pathway leads to resolution of tremor by interfering with transmission of synchronous bursting activity from the thalamus to the motor cortex.

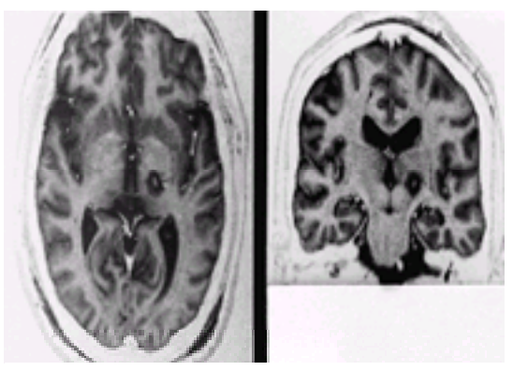


Fig. 3. Axial and coronal MRI of the brain, obtained 4 h after ventralis intermedius thalamotomy. Fast spin echo inversion recovery sequence: TI, 200 ms; TR, 2000 ms; TE, 24 ms; slice thickness, 2.0 mm. There is a ring of edema (black) around the coagulum of the lesion (white). The corticospinal tract is lateral to the lesion.

Indications and results

Appropriate candidates for ventralis intermedius thalamotomy are patients with essential tremor, or tremor-dominant Parkinson's disease, whose tremor has become debilitating despite optimal medical management. Several recent reviews have documented long-term (longer than 3 years) efficacy of 70 to 90 per cent for alleviation of tremor in these diseases. Cerebellar-outflow tremors, such as those due to multiple sclerosis or cerebellar trauma, are the most resistant to treatment. Long-term memory or speech deficits are reported complications of unilateral thalamotomy, particularly in the dominant hemisphere. These complications are much more common after bilateral thalamotomy, which, therefore, is rarely recommended. There are reports that extending the thalamotomy lesion anteriorly into ventralis oralis posterior/anterior improves rigidity and may alleviate drug-induced dyskinesias. However, rigidity, bradykinesia, and motor fluctuations appear to be better controlled by pallidotomy of the internal segment. Thus, for Parkinson's disease, pallidotomy is preferable to thalamotomy in all but the minority of patients with Parkinson's disease for whom tremor is the dominant and only feature of the disease.

Thalamotomy for other conditions

In the 1950s, in the course of performing thalamotomy for Parkinson's disease, it was noted that an associated dystonia was often relieved by these lesions. Thus, in addition to tremor, thalamotomy is preferred by many neurosurgeons for primary dystonia as well. Since then, several groups have erformed thalamotomy for non-parkinsonian dystonia. The results have been variable, which may reflect heterogeneous patient populations and variations in the site of thalamic lesioning. Published series have included various forms of dystonia, including idiopathic, secondary, focal, and generalized. The largest series of thalamotomy for idiopathic torsion dystonia is from Cooper, who reported marked improvement in 24.5 per cent and mild improvement in 45.2 per cent of patients, with a mean follow-up of 7.9 years. In Andrew's series, only four of 16 patients with generalized dystonia had sustained benefit, while 62 per cent of patients with focal or segmental dystonia had sustained benefit. Tasker's most successful results in thalamotomy for dystonia were from lesions involving both nucleus ventralis intermedius and nucleus ventralis oralis posterior. In another series, lesions were placed in ventralis oralis posterior, and if ineffective, were extended anteriorly into ventralis oralis anterior with a second operation. It would appear that lesions to treat dystonia must be much larger than those that are effective for tremor. No published study has been made prospectively with blinded evaluation, documentation of lesion location, or use of detailed, standardized quantitative rating scales of motor disability. Thus, the true long-term outcome of thalamotomy for dystonia is unknown. The neurosurgical treatment of dystonia is currently being re-evaluated and is likely to evolve rapidly.

Since hemiballismus is a rare condition and may resolve spontaneously, thalamotomy is rarely performed for hemiballismus. However, several recently reported cases of ventralis intermedius thalamotomy for hemiballismus showed excellent control.

Deep brain stimulation

Intraoperative electrical stimulation of subcortical nuclei has long been performed during ablative surgery for movement disorders to provide confirmation of target location prior to lesioning. Investigators found empirically that deep brain stimulation at high frequency (greater than 100 Hz) in the motor thalamus arrested tremor similarly to that observed after lesioning, but that the effect reversed when stimulation was stopped. It appeared as if deep brain stimulation created a reversible inactivation of a focal brain region near the electrode. Deep brain stimulation for motor signs of Parkinson's disease was first reported by Siegfried in the motor thalamus and later the pallidum. Benabid also pioneered the development of chronic thalamic stimulation for the treatment of tremor and, more recently, stimulation in the subthalamic nucleus for motor symptoms of Parkinson's disease. The rationale for selecting the nucleus ventralis intermedius, the internal segment of the globus pallidus, or the subthalamic nucleus as targets for stimulator placement is similar to that for selecting these sites for lesioning, since the effects appear similar. The advantages of deep brain stimulation are the reversibility and adjustability of the effect, and that bilateral procedures can be performed with less risk of permanent complications than from lesioning.

The mechanism underlying the beneficial effects of deep brain stimulation is poorly understood. The evidence that chronic high-frequency stimulation inactivates neighboring neurones is circumstantial: in experimental animals as well as in humans, deep brain stimulation in several subcortical nuclei can produce similar physiologic effects to those of lesioning. The stimulation may directly inactivate a region of neurones or indirectly inactivate them via activation of inhibitory projections to that region. However, it is also possible that cells are activated, rather than inactivated, by deep brain stimulation, depending upon the stimulation variables used and upon the cells' morphology, membrane properties, distance from the stimulating electrode, and baseline discharge rate. Positron-emission tomographic studies of patients undergoing stimulation of the nucleus ventralis intermedius for tremor support this possibility. Those studies show changes in cerebellar metabolism during clinically effective, but not during ineffective, stimulation. Since ventralis intermedius has no known direct projection to the cerebellum, these changes in cerebellar metabolism appear to be due to antidromic activation of cerebellar afferents to the nucleus. Stimulation-induced activation of neuronal discharge could have similar effects on motor function as lesioning, by ‘jamming' or desynchronizing abnormally oscillating patterns of discharge.

Indications and results

Thalamic stimulation

Chronic stimulation of the nucleus ventralis intermedius is effective for medically intractable parkinsonian or essential tremor (Fig. 4). In a series of 100 patients who received 147 stimulations for tremors, 88 per cent of patients with Parkinson's disease and 61 per cent of patients with essential tremor had complete or near complete resolution of their tremor at the longest follow up, ranging from 6 months to 8 years. These results are similar to the best for thalamotomy, and stimulation of ventralis intermedius may be done bilaterally with fewer side-effects than from thalamotomy. At present, stimulation of ventralis intermedius is the procedure of choice when contralateral thalamotomy has been performed or is anticipated. Whether it should completely replace unilateral thalamotomy of ventralis intermedius is unclear at this time and requires further study.

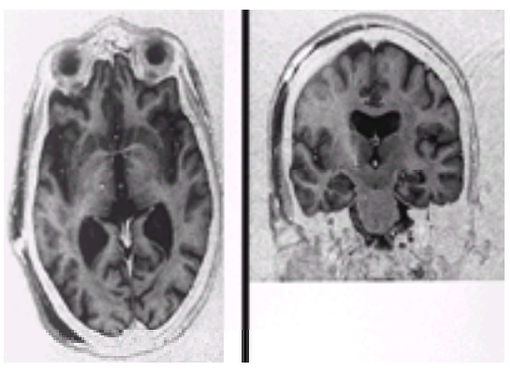


Fig. 4. MRI obtained 1 day after the placement of a Medtronic quadripolar deep-brain stimulation electrode into the ventralis intermedius (Vim) of the thalamus for parkinsonian tremor. Fast spin echo inversion recovery sequence; TI, 200 ms; TR, 2000 ms; TE, 24 ms; slice thickness, 2.0 mm.(A) Axial scan at the level of the PC; (B) coronal scan 20 mm posterior to the AC. There is artefact around the lead, but its position in the ventral lateral thalamus 5 mm anterior to the PC can be seen.

Stimulation of the internal segment of globus pallidus

A small number of case reports suggest that chronic stimulation of the internal segment has similar effects to those of pallidotomy for rigid/bradykinetic Parkinson's disease, and may be safe for bilateral use. It is possible that pallidal stimulation contralateral to a previous pallidotomy would benefit patients with bilateral disease, while avoiding the potential complications associated with bilateral pallidotomy (Fig. 5), but this remains to be confirmed in clinical trials.

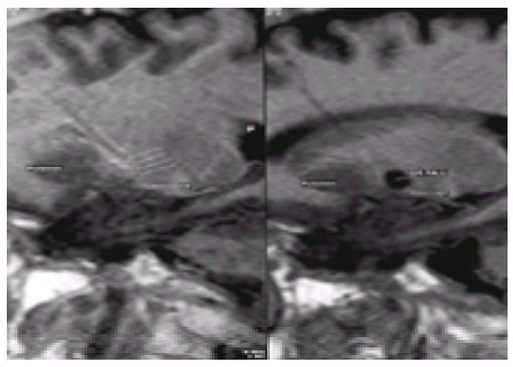


Fig. 5. Parasagittal MRI of the brain 6 h after placement of deep brain stimulation in the internal segment of globus pallidus (GPi). Technique is the same as in Fig. 4. The four contacts can be identified (arrows) above the optic tract (OT) in GPi. On the opposite side is a pallidotomy more than 6 months postoperatively.

Stimulation of the subthalamic nucleus

Based on work in parkinsonian monkeys showing substantial benefit from lesions in the subthalamic nucleus, Benabid and colleagues have implanted bilateral stimulators in that nucleus in several patients with predominantly rigid/bradykinetic Parkinson's disease. In the first three patients, motor scores were improved by 42 to 82 per cent 3 months after surgery. Akinesia and rigidity were the motor signs that improved the most. Dyskinesias could be induced by stimulation of the subthalamic nucleus at higher current intensities, but were reversible when the current intensity was reduced. Dyskinesias can also be decreased if the medication can be reduced. There appeared to be a ‘therapeutic window' of stimulation variables within which parkinsonian motor signs could be ameliorated without inducing involuntary movements. Thus this technique appears promising, and is now being tested at multiple centers.

Comparison with lesioning

The major advantages of deep brain stimulation compared to lesioning are that it is (1) reversible, (2) permits adjustment in the magnitude of its effect at any time during treatment, and (3) may be safer for bilateral use. Also, it lends itself readily to objective evaluation since, by varying the status of the stimulator (on or off), its effects can be easily studied in a double-blind fashion. The disadvantages of deep brain stimulation, as compared with lesioning, are: (1) it requires implantation of a foreign body with concomitant risks of infection and migration, (2) batteries must be replaced at regular intervals, or an external power source must be worn against the skin, and (3) a single stimulator is limited in the volume of tissue it can affect, thus rendering it potentially less practical in regions where the physiologic target is large.

Restorative therapy

Restorative therapy is highly experimental but is currently being studied in patients with Parkinson's and Huntington's disease. Theory, results, and techniques will be briefly reviewed. Idiopathic Parkinson's disease is the neurodegenerative disorder most amenable to a transplantation approach because it is a relatively focal disease, which, in its early or middle stages, involves mainly the loss of one cell type, the dopaminergic cells of the substantia nigra pars compacta. Since the major projection from the pars compacta innervates the striatum, the strategy of all clinical transplantation efforts thus far has been to place dopamine-secreting tissue within the striatum.

Adrenal chromaffin autografts

Intrastriatal autografting of adrenal chromaffin tissue was the first technique attempted in clinical transplantation for Parkinson's disease. Denervated adrenal medulla secretes dopamine, and thus offered the theoretical promise of serving as an autologous ‘dopamine pump' within the brain. This technique avoided the immunologic and ethical issues associated with transplantation of nonautologous fetal tissues. However, studies showed only very modest improvement, along with significant surgical morbidity. Autopsy studies of transplanted patients revealed minimal survival of chromaffin cells. Although laboratory work continues in an effort to find means of enhancing the survival of adrenal chromaffin grafts, currently there are no clinical studies in the United States.

Fetal ventral mesencephalic allografts

In the developing mammalian fetus, the ventral mesencephalon contains the dopaminergic cell bodies that are the precursors of the mature dopaminergic cells of the substantia nigra pars compacta. In both rodent and nonhuman primate models of Parkinson's disease, grafts of fetal ventral mesencephalon tissue, placed heterotopically into the striatum, have been shown to survive and reinnervate portions of the striatum, coincident with improvement in abnormal motor behaviors. Anatomic and physiologic studies of such intrastriatal grafts suggest that synaptically mediated host–graft interactions occur, resulting in a more complex level of function than would be expected from implantation of an unregulated dopamine pump.

Patient characteristics and clinical outcomes

As of 1997, at least 11 groups have reported over 200 patients who have undergone transplantation of human fetal ventral mesencephalon cells to the striatum for parkinsonism. Most transplanted patients suffered from idiopathic Parkinson's disease, although two suffered from MPTP-induced parkinsonism. All had relatively advanced disease, with Höhn and Yahr stages of about 3 when ‘on' and 4–5 when ‘off'. Most patients were 40 to 60 years old and had suffered from Parkinson's disease for from 5 to 20 years prior to surgery. Patients were primarily bradykinetic and rigid, with little or no tremor.

Outcomes have been highly variable; improvements were rarely immediate, usually taking several months. There was a worsening in many patients' motor symptoms during the first 4 to 6 weeks after surgery and sometimes transiently increased dyskinesias, as well as several transient psychiatric complications. There were several complications of immunosuppressive therapy requiring its cessation. In most studies, modest benefit was reported in most, but not all, patients. Commonly reported were reductions in the amount of time spent ‘off', reductions in drug-induced dyskinesias (which may reflect lowering the medications), and some improvements in fine motor tasks. The majority of patients did not have a significant change in their functional status, and rarely are patients able to be taken off medications.

Survival of grafted fetal dopaminergic cells can be assessed noninvasively by positron-emission tomography. A tracer of dopaminergic metabolic activity, such as [6-^{l8}F]fluoro-L-dopa, is used as the positron-emitting source. The concentration of that source in the brain, following intravenous administration, reflects the metabolic activity in dopaminergic nerve cells and terminals. In Parkinson's disease, uptake of fluoro-L-dopa in the striatum is less than in normal controls. In patients followed for the longest time, there is evidence for graft survival at 3 years. Uptake of fluoro-L-dopa by positron-emission scanning cannot distinguish between the survival of transplanted tissue and the sprouting of host dopaminergic tissue, which is know to occur with lesioning of the striatum. Short of autopsy studies, positron-emission tomography is currently the best method for following graft survival.

One autopsy report, from a patient transplanted at the University of South Florida, demonstrated excellent graft survival. This patient died 18 months after grafting, from a pulmonary embolism following orthopedic surgery. He had received bilateral grafts into the putamen from seven fetuses, age 6.5 to 9 weeks. He had been immunosuppressed for 6 months after transplantation. Many tyrosine hydroxylase-positive cells were observed in the grafts, up to 1000 per section. Neuronal processes extending up to 7 mm into surrounding normal brain were seen. There was no evidence of sprouting of tyrosine hydroxylase -positive fibers from the host brain. This study demonstrates that human fetal tissue implanted in a human can survive and robustly reinnervate host tissues. In addition, it shows that long-term immunosuppression (greater than 6 months post-transplant) may not be needed for survival of human neural tissue transplants, even if multiple donors are used.

Dopamine-secreting xenografts

The use of porcine xenografts for neuronal cell transplantation, as an alternative to human fetal allografts, is under active investigation. Immune-mediated rejection of xenografts is, in general, even more vigorous than that of allografts and this could limit the clinical utility of xenografts. Since transplant rejection is mediated mainly by antigens of the major histocompatibility complex (**MHC**), one strategy employed to circumvent rejection is to use an antibody fragment directed against MHC antigens to induce tolerance of the foreign tissue in the host environment. Treating pancreatic and hepatic xenografts with the variable region fragment F(ab´)$_2$ of a monoclonal antibody against MHC antigens, just prior to transplantation, induces tolerance in the host. This tolerance is long lasting even though the transplant is treated only once. By using antibody fragments from which the constant region (Fc) has been removed, activation of the complement cascade and antibody-mediated destruction is avoided while cellular immunity is modified. This MHC-rnasking technique has been applied to porcine neural xenografts and, in 1995, a clinical trial of porcine fetal

neuronal transplantation for Parkinson's disease began, using F(ab´)$_2$-treated cells.

Genetically engineered autologous tissue

To completely circumvent the immunologic complexity of xeno- or allotransplantation, a patient's own tissues could be removed, genetically engineered to adopt a useful neuronal phenotype, then reimplanted into the host brain. Primary fibroblasts have been engineered to express tyrosine hydroxylase using retrovirus-mediated transfection; these cells can then secrete L-dopa. Implantation of autologous L-dopa-secreting fibroblasts in the 6-hydroxydopamine model of Parkinson's disease results in behavioral improvement for at least 8 weeks post-implantation, but the improvement declines after 2 weeks and the implanted fibroblasts show some signs of neoplasia.

The use of plasmid-transfected, primary cultured muscle cells, in the same experimental approach, has resulted in more stable behavioral amelioration, with evidence of continued L-dopa secretion up to 6 months post-transplant. A disadvantage of using genetically engineered non-neuronal tissues is that it is not possible for them to form functional synapses with host tissue. In addition, long-term expression of foreign genes in transfixed cells is difficult to achieve with present gene-transfer methods.

Summary

There are a number of procedures available to patients with movement disorders. The ablative procedures are most common, but the use of stimulation is rapidly increasing. The future is being explored with transplantation and gene therapy.

Further reading

Baron MS *et al.* Treatment of advanced Parkinson's disease by posterior GPi pallidotomy: 1-year results of a pilot study. *Annals of Neurology* 1996; **40**: 355–66.

Benabid AL *et al.* Chronic electrical stimulation of the ventralis intermedius nucleus of the thalamus as a treatment of movement disorders. *Journal of Neurosurgery* 1996; **84**: 203–14.

Blond S *et al.* Control of tremor and involuntary movement disorders by chronic stereotactic stimulation of the ventral intermediate thalamic nucleus. *Journal of Neurosurgery* 1992; **77**: 62–8.

Cooper IS, Bergman LL, Caracalos A. Anatomic verification of the lesion which abolishes parkinsonian tremor and rigidity. *Neurology* 1963; **13**: 779–87.

Deacon T *et al.* Histological evidence of fetal pig neural cell surgical after transplantation into a patient with Parkinson's disease. *Nature Medicine* 1997; **3**: 350–3.

Fazzini E, Dogali M, Stereo D, Eidelberg D, Beric A. Stereotactic pallidotomy for Parkinson's disease: a long-term follow-up of unilateral pallidotomy. *Neurology* 1997; **48**: 1273–7.

Giller CA, Dewey RB, Ginsburg MI, Mendelsohn DB, Berk AM. Stereotactic pallidotomy and thalamotomy using individual variations of anatomic landmarks for localization. *Neurosurgery* 1998; **42**: 56–62.

Hassler R, Mundingr F, Reichert T. Correlation between clinical and autoptic findings in stereotaxic operations of parkinsonisn. *Confinia Neurologica* 1965; **26**: 282–90.

Hutchison, WD, Lozano AM, Tasker RR, Lang AE, Dostrovsky JO. Identification and characterization of neurons with tremor-frequency activity in human globus pallidus. *Experimental Brain Research* 1997; **113**: 557–63.

Jankovic J, Cardoso F, Grossman RG, Hamilton WJ. Outcome after stereotactic thalamotomy for parkinsonian, essential, and other types of tremor. *Neurosurgery* 1995; **37**: 680–7.

Laitinen LV, Bergenheim AT, Hariz MI. Leksell's posteroventral pallidotomy in the treatment of Parkinson's disease. *Journal of Neurosurgery* 1992; **76**: 53–61.

Lenz FA *et al.* Methods for microstimulation and recording of single neurons and evoked potentials in the human central nervous system. *Journal of Neurosurgery* 1988; **68**: 630–4.

Limousin, P, Greene J, Pollak P, Rothwell J, Benabid AL, Frackowiak R. Changes in cerebral activity pattern due to subthalamic nucleus or internal pallidum stimulation in Parkinson's disease. *Annals of Neurology* 1997; **42**: 283–91.

Masterman D *et al.* Motor, cognitive, and behavior performance following unilateral ventroposterior pallidotomy for Parkinson's disease. *Archives of Neurology* 1998; **55**: 1201–8.

Olanow CW, Kordower JH, Freeman TB. Fetal nigral transplantation as a therapy for Parkinson's disease. *Trends in Neuroscience* 1996; **19**: 102–9.

Rye DB, Turner RS, Vitek JL, Bakay RAE, Crutcher MD, DeLong MR. Anatomical investigations of the pallidotegmental pathway in monkey and man. In: Ohye C, ed. *The basal ganglia*, Vol. 5, pp. 59–75. Plenum, New York, 1996.

Siegfried J, Lippitz B. Chronic electrical stimulation of the VL–VPL complex and of the pallidum in the treatment of movement disorders: Personal experience since 1982. *Stereotactic and Functional Neurosurgery* 1994; **62**: 71–5.

Starr, PA, Vitek JL, Bakay RAE. Ablative surgery and deep brain stimulation for Parkinson's disease. *Neurosurgery* 1998; **43**: 989–1015.

Svennilson, E, Torvik A, Lowe R, Leksell L. Treatment of parkinsonism by stereotactic thermolesions in the pallidal region: a clinical evaluation of 81 cases. *Acta Psychiatrica et Neurologica Scandinavica* 1960; **35**: 358–77.

Vitek JL, Bakay RAE. The role of pallidotomy in Parkinson's disease and dystonia. *Current Opinion in Neurology* 1997; **10**: 332–9.

Vitek JL *et al.* Microelectrode-guided pallidotomy: technical approach and its application in medically intractable Parkinson's disease. *Journal of Neurosurgery* 1998; **88**: 1027–43.

44.6 Trigeminal neuralgia

C. B. T. Adams

Further reading

Trigeminal neuralgia is one of the most painful conditions known to mankind. Its cause is unknown, but the current hypothesis, that it is due to microvascular compression of the trigeminal nerve adjacent to the brainstem by blood vessels, has not been proved. Patients are usually over 40 years of age, unless the condition is associated with multiple sclerosis, and individuals of black African origin are rarely affected. There is no diagnostic test; diagnosis depends on the clinical features.

The pain is usually stabbing in quality and in the distribution of the trigeminal nerve. It is almost always unilateral, and is precipitated by many factors including touch, chewing, and talking. Remissions occur and may last for weeks, or even years; however, the pain inevitably returns. If carbamazepine fails to produce clear relief, then the diagnosis is in doubt. A few patients develop an idiosyncratic rash and are unable to take the drug.

The surgical treatment is debated. If carbamazepine fails to control the pain, ‘numbing' the trigeminal nerve is effective. This procedure can be performed anywhere in the distribution of the pain—in the skin, mucosa, peripheral nerve, nerve root, or brainstem. Although most other pain-relieving operations depend on cutting the nerve fibres between the cause of the pain and the brain, this does not seem to be the case with trigeminal neuralgia. Rubbing the nerve will stop the pain in two-thirds of patients.

Glycerol injections, balloon compression, radiofrequency lesions, ‘microvascular decompression', or partial sectioning (Fig. 1) of the trigeminal sensory root in the posterior fossa have all been used: the less the root is damaged, the greater the recurrence rate. Although many patients want overwhelmingly to be rid of the pain permanently, care must be taken in counselling: a numb face is not normal, and the patient must understand fully the effect of any proposed procedure.

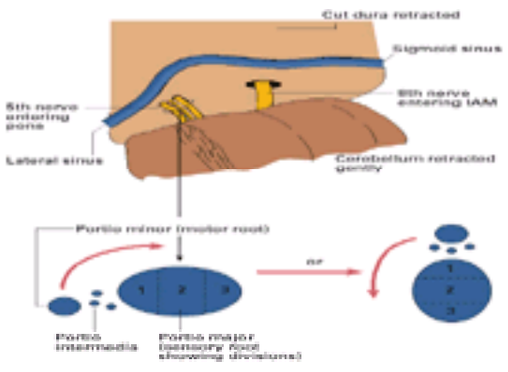


Fig. 1. The varying degrees of rotation of the trigeminal roots as they enter the brainstem. The majority of the sensory fibres from the third division of the trigeminal nerve maintain a constant relation to the part of the sensory root most removed from the motor roots or roots.

Glossopharyngeal neuralgia has similar features to trigeminal neuralgia, but occurs in the distribution of the glossopharyngeal nerve.

Further reading

Adams CBT. Microvascular compression: an alternative view and hypothesis. *Journal of Neurosurgery* 1989; **57**: 1–12. [A sceptical view of the microvascular compression.]

Gybels JM, Sweet WH. *Neurosurgical treatment of persistent pain*. Karger, Basel, 1989. [An overview of the treatment of trigeminal neuralgia.]

44.7 Intracranial infection

C. B. T. Adams

Extradural pus
Subdural pus
Cerebral abscess

Features
Investigation
Treatment

Intracranial pus may be extradural, subdural, or intracerebral.

Extradural pus

Extradural pus is associated with osteomyelitis due to frontal sinusitis or middle-ear disease, or to compound fractures. Local swelling (Pott's puffy tumour) occurs, and treatment requires removal of the infected bone (Fig. 1). The dura is a remarkable barrier to infection: intradural extension of an infection associated with osteomyelitis is rare, but can occur.

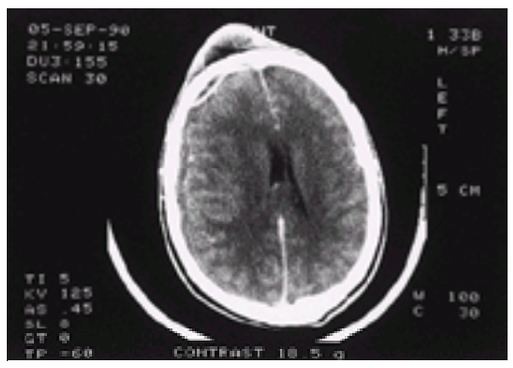


Fig. 1. CT scan, with contrast, showing an extradural abscess and Pott's puffy tumour. The infected dura is highlighted by contrast and is stripped away from the skull. In addition there is a subdural abscess showing as a low-density shadow displacing the cerebral hemisphere from the skull. The source of infection was the right frontal sinus.

Subdural pus

Subdural pus usually arises from frontal sinusitis. Once pus enters the subdural space it diffuses widely, not only over the hemisphere, but also alongside the falx (Fig. 1 and Fig. 2). Thrombosis of the draining cortical veins produces the characteristic picture of severe epilepsy and profound hemiplegia (and aphasia if on the dominant side). The causative organism is usually *Streptococcus milleri*. Strict anaerobes, such as bacteroides, may occasionally be found if appropriate laboratory methods are used.

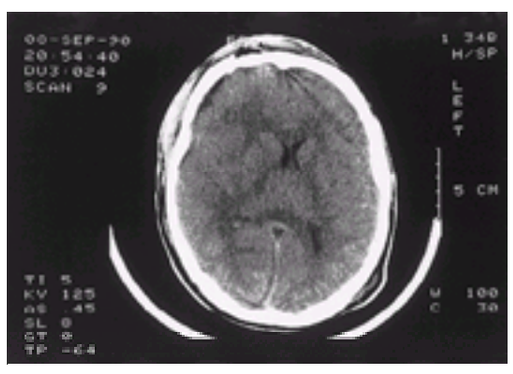


Fig. 2. CT 3 days later of patient shown in Fig. 1, showing subdural pus situated along the right side of the falx posteriorly. Parafalcine pus often occurs with subdural abscess. The extradural pus and convexity subdural pus have been evacuated.

The clinical features are those of severe, progressive ‘sinus' headache; the patient is febrile and unwell, and epilepsy and severe, focal, neurological signs develop within 24 or 48 h. Intracranial pressure becomes raised later in the course of the illness. Computed tomographic (**CT**) scanning usually shows a parafalcine collection of pus, but if the diagnosis is in doubt a burr hole should be made.

An extensive craniotomy extending laterally from the midline enables parafalcine pus to be evacuated. High doses of parenteral penicillin (20 Mu/day) and metronidazole (400–600 mg twice daily) should be given before the results of bacteriological investigations are known. The frontal sinus may also require drainage.

Cerebral abscess

The majority (75 per cent) of intracerebral abscesses are caused by spread of an infection in a nearby site, usually in the frontal sinuses or middle ear. The remainder are caused by haematogenous spread of micro-organisms, particularly in patients with congenital cyanotic heart disease (Fig. 3).

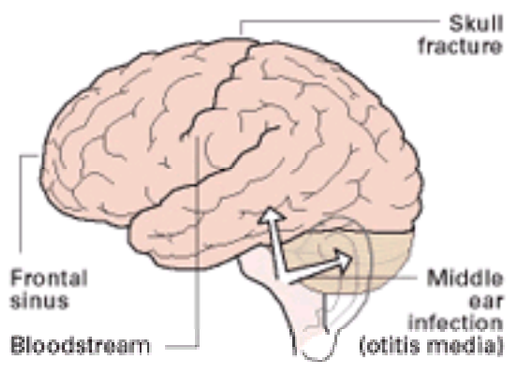


Fig. 3. Diagram to show the modes of spread in the formation of a cerebral abscess.

Initially there is a zone of suppurative encephalitis; this is followed by necrosis, with frank pus formation surrounded by granulation tissue (acute abscess). After 5 to 6 weeks a zone of fibrous tissue develops (chronic abscess). Frontal sinusitis produces a frontal-lobe abscess, while middle-ear infection passes through the roof of the middle-ear cavity into the temporal lobe; haematogenous spread produces multiple and often deep abscesses. Cerebellar abscess is now rare; the causative organism is usually *S. milleri* (microaerophilic) or bacteroides (anaerobic).

Features

Raised intracranial pressure occurs early in the course of the illness because of the considerable oedema surrounding the abscess. Papilloedema is often present without other features of raised pressure; frequent fundoscopic examination is important not only for the diagnosis but also the assessment of patients during treatment (Fig. 4). Focal signs are sometimes difficult to elicit in a drowsy patient, but if carefully looked for these will be found. A frontal-lobe abscess may cause a lack of facility or a drift of an outstretched arm, suggestive of mild weakness. A temporal-lobe abscess causes a homonymous quadrantopia, which can be demonstrated by asking the patient to take hold of the examiner's hands: the hand on the side of the defect is ignored.

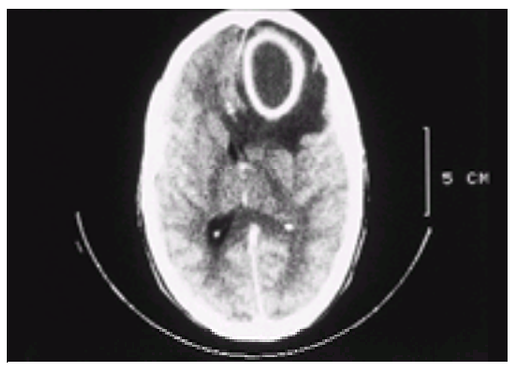


Fig. 4. CT scan to show a frontal-lobe abscess: the regular ring shadow with marked surrounding oedema is characteristic.

Investigation

The CT scan is the best method of continuing the diagnosis (Fig. 4). This shows a cystic lesion with a fine but regular capsule that enhances with contrast due to the zone of granulation tissue. There is considerable surrounding oedema. Cystic gliomas sometimes have a similar appearance: if there is doubt a diagnostic burr hole and aspiration (performed stereotactically, with CT guidance for deep and inaccessible lesions) is indicated. The oedema surrounding a cerebral abscess is dramatically reduced by treatment with steroids (dexamethasone). This may improve the patient's headache and drowsiness and mask the CT appearances by reducing the oedema and compression of adjacent structures. A false-negative diagnosis may be made if the patient has received steroid treatment before CT scanning has been performed.

Treatment

The treatment of a cerebral abscess is the same as that for collections of pus anywhere in the body. The inflammation is treated by antibiotics. Penicillin (20 Mu/day intravenously) and metronidazole (600 mg twice daily) are reasonable choices until the bacteriological diagnosis and antibiotic sensitivities are known. Removal of the pus is best achieved by burr-hole aspiration. This should be repeated each day until no further pus is obtained. The oedema persists for several weeks or months and repeat CT scans are advisable to confirm that there is no recrudescence of pus. If aspiration fails because of reinfection or because the capsule is too thick, craniotomy and excision of the abscess is necessary. Drainage is sometimes used, but this has not found general favour.

The effects of the abscess must be treated. If the patient has greatly raised intracranial pressure and shows signs of tentorial coning, judicious administration of high doses of dexamethasone may be life-saving, especially if emptying the abscess of pus fails to reduce the intracranial pressure rapidly. The comatose patient requires appropriate nursing case with maintenance of the airway, and care of the skin, bladder, and bowels. About one-third of patients develop epilepsy, and anticonvulsants should be given to all patients with an intracerebral abscess. If chronic epilepsy results despite adequate drug treatment, this can be sometimes effectively treated, for instance by excising the frontal lobe containing the abscess scar. The cause of infection also requires treatment; this is often frontal-sinus disease, middle-ear infection, or congenital heart disease.

44.8 Hydrocephalus

Daniel L. Barrow

[Physiology and definition of terms](#)
[Etiology](#)
[Clinical presentation](#)
[Radiological examination](#)
[Treatment](#)
[Further reading](#)

Physiology and definition of terms

Cerebrospinal fluid is a clear, colorless fluid produced within, and circulated through, the cerebral ventricles. The choroid plexus is its primary source, although some extrachoroidal sources are well documented. Under normal circumstances, cerebrospinal fluid is produced at a rate of 20& ml/h. It circulates through the ventricular system, from the paired lateral ventricles, through the foramina of Monro into the midline third ventricle, through the aqueduct of Sylvius to the fourth ventricle, out the paired foramina of Luschka and single foramen of Magendie, through the basal cisterns and spinal subarachnoid space, and is reabsorbed through the arachnoid villi, which are specialized organs at the level of the sagittal sinus (Fig. 1). Equal amounts of cerebrospinal fluid are produced and absorbed each day.

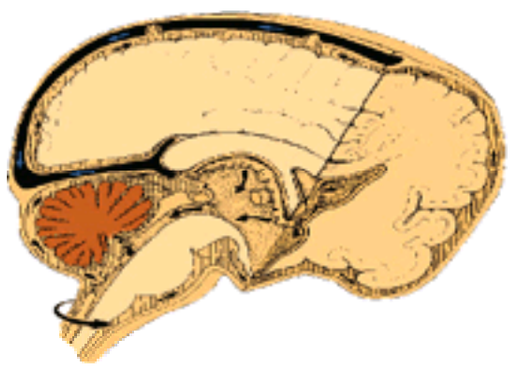


Fig. 1. Coronal and sagittal cross-section of brain illustrating anatomy of cerebrospinal fluid pathways.

Hydrocephalus is a clinical condition with a multitude of causes that results from an inability to absorb all of the cerebrospinal fluid produced. Except in the rare instances of overproduction of cerebrospinal fluid by a papilloma of the choroid plexus, hydrocephalus occurs when there is an obstruction to the flow of fluid from one compartment to another. Hydrocephalus may be defined in anatomic terms as a condition that results in enlargement of the intracranial ventricles beyond accepted norms. This definition would include compensatory enlargement of the ventricles due to cerebral atrophy (hydrocephalus *ex vacuo*) or large ventricles in asymptomatic patients, two conditions that do not require treatment. A functional definition is more useful for clinical purposes. Hydrocephalus is a clinical syndrome characterized by enlarged ventricles, seen on imaging studies, in which therapy is directed toward the alleviation of symptoms arising from increased intracranial pressure or the enlarged ventricles.

In clinical practice, most cases of hydrocephalus occur as a result of an obstruction along some point in the circulation pathway of the cerebrospinal fluid. An anatomic obstruction of the ventricular system, such as a tumor within or adjacent to it, results in non-communicating hydrocephalus, whereas an impairment of circulation within the subarachnoid spaces or abnormal absorption of cerebrospinal fluid results in communicating hydrocephalus. The accumulation of cerebrospinal fluid proximal to the point of obstruction will cause enlargement of the ventricles (ventriculomegaly), and will usually cause symptoms by elevating intracranial pressure or by distension of the ventricles.

Etiology

In infancy and childhood, the most common causes of hydrocephalus are related to congenital abnormalities of the brain. Approximately 80 per cent of infants with myelomeningocele will have hydrocephalus due to the Chiari malformation, a congenital deformity of the hindbrain that impairs the flow of cerebrospinal fluid into the subarachnoid space from the fourth ventricle. Aqueductal stenosis, a congenital narrowing of the aqueduct of Sylvius, accounts for many cases of hydrocephalus in infancy or childhood. Other congenital conditions associated with hydrocephalus are listed in Table 1.

CNS conditions	Non-CNS conditions
Spina bifida:	Trisomy 13 and 18
Chiari malformation	Achondroplasia
Myelomeningocele	Pulmonary hypoplasia
Aqueductal stenosis	Renal hypoplasia
Dandy–Walker malformation	Transposition of great vessels
Encephalocele	Endocardial cushion defects
Holoprosencephaly	Hydronephrosis
Intracranial hemorrhage	Esophageal atresia
Cerebral infarction	Omphalocele
Porencephalic cyst	Intestinal malrotation
Vein of Galen aneurysm	
Congenital tumors:	
Teratomas	
Ependymomas	
Astrocytomas	
Infections:	
Toxoplasmosis	
Cytomegalovirus	
Bacterial meningitis	
Agenesis of corpus callosum	
Septo-optic dysplasia	

CNS, central nervous system.

Table 1 Conditions associated with hydrocephalus

A pathological process that results in inflammation and fibrosis or scarring of the subarachnoid spaces or other pathways for cerebrospinal fluid may result in hydrocephalus in children or adults. Intrauterine infections with toxoplasma or cytomegalovirus cause hydrocephalus by this mechanism. Bacterial meningitis, at any age, may cause obstruction of the scarred subarachnoid spaces, impair absorption of cerebrospinal fluid, and produce hydrocephalus. Intracranial hemorrhage is another common cause of hydrocephalus. In premature infants with respiratory distress, intraventricular hemorrhage from the fragile germinal matrix at the ependymal surface of the ventricles is a common cause of hydrocephalus. Traumatic subarachnoid hemorrhage or spontaneous hemorrhage from an intracranial aneurysm, vascular malformation or hypertension may be complicated by the development of hydrocephalus.

A variety of intracranial tumors may cause hydrocephalus in children and adults, and are the most common cause of non-communicating hydrocephalus. Brain tumors are the most common solid neoplasm in childhood. Half of all brain tumors affecting children are medulloblastomas or cerebellar astrocytomas that block the fourth ventricle and cause hydrocephalus. Tumors of the pineal region and midbrain gliomas usually compress the aqueduct of Sylvius and produce obstructive hydrocephalus, often as the presenting symptom. Colloid cysts of the third ventricle obstruct the foramina of Monro, and hydrocephalus is usually responsible for all the presenting symptoms and signs. Tumors of the skull base, such as meningiomas, acoustic neuromas, craniopharyngiomas, or pituitary adenomas, may impair the pathways of cerebrospinal fluid and cause hydrocephalus.

A syndrome peculiar to adults is normal-pressure hydrocephalus. First described in 1965, the syndrome most commonly occurs in the sixth to eighth decades of life and represents one of the treatable causes of dementia. Normal-pressure hydrocephalus can occur idiopathically, after trauma, or after subarachnoid hemorrhage, and

is thought to be due to an impairment of the absorption of cerebrospinal fluid.

Clinical presentation

The patient with hydrocephalus may have symptoms and signs of intracranial hypertension due to the hydrocephalus or clinical manifestations dominated by its underlying cause. Patients with obstructive hydrocephalus due to certain mass lesions, such as a colloid cyst, may have symptoms entirely attributable to the associated hydrocephalus. Patients with hydrocephalus as a consequence of severe bacterial meningitis, brain trauma with subarachnoid hemorrhage, or hypertensive hemorrhage, may have symptoms and signs of their primary neurological disorder that overshadow any clinical manifestations of the hydrocephalus.

The patient's age will also influence the clinical presentation of hydrocephalus. In infants and young children, before the closure of the fontanelles and cranial sutures, the head will enlarge, sutures will split, and the fontanelle will bulge in response to the accumulation of cerebrospinal fluid and ventriculomegaly. In addition to the measurable macrocephaly, infants will frequently present with irritability, poor feeding response, and MacEwen's sign (cracked pot sound on percussion of the head). Once the cranial sutures fuse, hydrocephalus usually presents with signs and symptoms of increased intracranial pressure, including headache, nausea, vomiting, papilledema, impairment of upward gaze, and deterioration in level of consciousness. Occasionally, the intracranial hypertension associated with hydrocephalus will cause cranial nerve palsies. The abducens nerve is most commonly affected because it has the longest intracranial course.

Older adults may, as outlined above, develop the syndrome of normal-pressure hydrocephalus. This more chronic condition results in mechanical stretching and disruption of frontal cortical axons, a potentially irreversible cause of axonal dysfunction. This long-standing ventricular dilatation can initially cause mild behavorial changes, forgetfulness, and apathy. Chronic hydrocephalus in the adult is characterized by dementia, urinary incontinence, and a gait disturbance manifested by a wide-based, shuffling walk.

Radiological examination

In the infant with an open fontanelle, cranial ultrasonography is a reliable diagnostic technique for the detection of hydrocephalus. By using real-time, high-resolution, B-mode ultrasonography, hydrocephalus may even be diagnosed *in utero*. In older children and adults, hydrocephalus is usually imaged with computed tomography or magnetic resonance imaging. These methods record the ventriculomegaly and also demonstrate associated pathology that may be the cause of the hydrocephalus such as subarachnoid or intracerebral hemorrhage, tumors, or other mass lesions (Fig. 2).

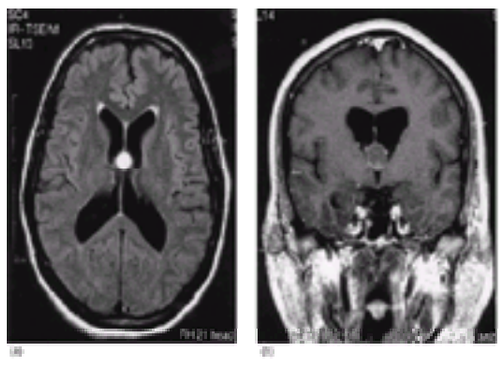


Fig. 2. (a) Axial and (b) coronal magnetic resonance images demonstrate a colloid cyst occluding the foramina of Monro and causing obstructive hydrocephalus.

Radionuclide cisternograms may be of some benefit in revealing abnormalities of the flow and absorption of cerebrospinal fluid in patients with normal-pressure hydrocephalus.

Treatment

Not all cases of ventriculomegaly require treatment. Therefore, the first step is to formulate a management plan to decide whether treatment is indicated, and if that treatment should be temporary or permanent. Asymptomatic ventriculomegaly should stimulate a search for its cause and may require careful clinical and radiologic follow-up, but it usually does not require treatment. In some circumstances, obstructive hydrocephalus will resolve after treatment of the underlying condition, such as removal of a tumor or cyst obstructing the ventricular system.

Medical treatment

The nonsurgical treatment of hydrocephalus includes medical therapy to decrease the production of cerebrospinal fluid, increase its absorption, or decrease the intracranial pressure. The primary objectives of all nonsurgical treatments are to avoid or delay the need for placement of a permanent shunt. The drugs most commonly used to decrease production of cerebrospinal fluid are acetazolamide and furosemide. They have been used individually or in combination, primarily as a temporizing measure in infants and children.

Heparin, urokinase, and hyaluronidase have been used to increase absorption of cerebrospinal fluid. Hyaluronidase is an enzyme that hydrolyzes fibrin. It has been used to treat patients with hydrocephalus following tuberculous meningitis. Heparin and urokinase have been injected into the subarachnoid or ventricular space following hemorrhage in an effort to lyse clot and improve absorption of cerebrospinal fluid.

Mannitol, urea, glycerol, and isosorbide have been used as osmotic diuretics to decrease intracranial pressure associated with hydrocephalus. Metabolic side-effects usually preclude their long-term use.

Temporary treatments

Hydrocephalus has been treated by intermittent removal of cerebrospinal fluid through serial lumbar puncture or external ventricular drainage. Lumbar puncture should only be used after proving that the hydrocephalus is communicating and not obstructive. Serial lumbar puncture has been used most commonly in infants with posthemorrhagic hydrocephalus of prematurity to allow the child to grow and develop before placing a permanent shunt.

Aneurysmal subarachnoid hemorrhage will cause at least transient hydrocephalus in virtually all patients. Only 20 per cent require permanent shunting for treatment of hydrocephalus. Many, however, are treated by external ventricular drainage to measure and control cerebrospinal fluid pressure, until the hydrocephalus resolves.

Ventricular shunting

Most patients with symptomatic hydrocephalus require surgical shunting to divert cerebrospinal fluid away from or around the site of obstruction. Most commonly, this involves a system that drains cerebrospinal fluid from the enlarged lateral ventricles to another site in the body that absorbs fluid. Current systems generally involve three components: a silicone ventricular catheter, a one-way valve, and a distal silicone catheter. Other devices such as on/off valves, antisiphon devices, and flushing chambers are used at the surgeon's discretion for special circumstances.

When performing a shunt, the neurosurgeon places the ventricular catheter into the dilated lateral ventricle through a frontal or occipital burr hole using surface landmarks for accurate placement. The cerebrospinal fluid pressure is measured and a sample collected for analysis. A ventricular catheter is attached to a unidirectional valve that regulates pressure and ensures flow of cerebrospinal fluid only in the desired direction. The valve is placed in a subcutaneous pocket in the scalp and its distal end is attached to the distal catheter, which is tunneled subcutaneously into the area of the body to which the cerebrospinal fluid is diverted. Cerebrospinal fluid is most commonly shunted to the peritoneal cavity (ventriculoperitoneal shunt), which is entered through a small incision over the anterior abdominal wall or by using a split trochar. Other body sites used less commonly for distal shunting include the right atrium of the heart (ventriculoatrial shunt) accessed through the jugular vein via a small neck incision, or the pleural space (ventriculopleural shunt) inserted through a small incision between the ribs on the anterior chest wall.

Patients with communicating hydrocephalus may also be candidates for a lumbar–peritoneal shunt. A Silastic tube is placed into the lumbar subarachnoid space

through a Tuohy needle, via a small, midline lumbar incision. The catheter is tunneled around the flank to the abdomen, where it is placed into the peritoneal cavity through a small abdominal incision.

Patients with nontumoral aqueductal stenosis may be treated by stereotactic or endoscopic third ventriculostomy to open the aqueduct and avoid the need for an indwelling shunt.

Shunt complications

Ventricular shunting is considered by most neurosurgeons to be a routine procedure, but if not performed in a consistent manner with attention to detail, it will result in unnecessarily high rates of infections and complications. The most common complications of shunts are malfunction and infection. Shunt malfunction may result from either underdrainage or overdrainage of cerebrospinal fluid. Underdrainage occurs if the shunt system becomes obstructed or disconnected. Most commonly, a shunt malfunction results from an obstruction of the ventricular catheter due to misplacement, migration, occlusion by choroid plexus, collapse of the ventricle, or growth of the patient's head. Malfunction may also result from disconnection of the system or distal obstruction. Distal obstruction may be due to an abdominal pseudocyst or loculated collection of cerebrospinal fluid around the peritoneal catheter, or the inability of the pleural space to absorb the fluid adequately.

Overdrainage can result in persistent headache from low cerebrospinal fluid pressure or it may predispose the patient to the development of subdural hematomas or hygromas. These complications may require placement of a higher-pressure valve to limit the amount of cerebrospinal fluid drained from the ventricular system.

Shunt infections make up 2 to 10 per cent of complications in reported series. Most shunt infections are caused by indolent normal skin flora including *Staphylococcus epidermidis* and *Propionibacterium acnes*. Treatment usually requires externalization of the shunt, systemic antibiotics until the cerebrospinal fluid is sterile, and replacement with a new, sterile shunt system. Other complications of shunting include hemorrhage, seizure disorders, and shunt metastases.

Further reading

Rekate HL, Cherny WB. Pathophysiology, diagnosis and clinical features of hydrocephalus in infants and children. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 2689–705. Williams and Wilkins, Baltimore, 1996. [Discusses pathophysiology and manifestations of hydrocephalus in adults and children.]

Robertson DP, Simpson RK. Cerebrospinal fluid diversion in adults. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 1267–78. Williams and Wilkins, Baltimore, 1996. [Gives details of surgical managment of hydrocephalus.]

Roth PA, Cohen AR. Management of hydrocephalus in children. In: Tindall GT, Cooper PR, Barrow DL, ed. *The practice of neurosurgery*, pp. 2707–28. Williams and Wilkins, Baltimore, 1996. [Focuses on hydrocephalus in the pediatric patient.]

Scott RM, ed. *Concepts in neurosurgery* Vol. 3, *Hydrocephalus*. Williams and Wilkins, Baltimore, 1990. [Covers the subject in detail.]

44.9 Intracranial aneurysms and arteriovenous malformations

Thomas A. D. Cadoux-Hudson and Andrew Molyneux

Introduction
Intracranial aneurysm
Intracranial arteriovenous malformations
Subarachnoid haemorrhage
Cerebral aneurysm and subarachoid haemorrhage
Cerebral aneurysm
Subarachnoid haemorrhage
Further reading

Introduction

Intracranial aneurysm

Intracranial or ‘berry' aneurysms are usually saccular but may occasionally be fusiform; they arise as an outpouching from the vessel wall, usually at a branching point. The saccular form comprises a neck and fundus. The neck is usually narrower than the fundus, which can be simple or lobulated and measure from a few millimetres up to several centimetres in diameter. Lobulated aneurysms can be complex and involve nearby branching vessels. Aneurysms are more commonly found on the major vessels of the anterior cerebral circulation (arteries of the circle of Willis), with the region of the anterior communicating artery, the middle cerebral bifurcation, and the internal carotid at the origin of the posterior communicating artery being the most common sites. The posterior circulation (vertebral or basilar arteries) typically accounts for 10 to 15 per cent of cerebral aneurysms that rupture.

Intracranial arteriovenous malformations

Arteriovenous malformations in the central nervous system can be considered as vascular hamartomas; they are not neoplastic. The vascular mesenchyme is split by the development of the brain and so arteriovenous malformations can affect all layers of the brain, skull, and scalp. Several groups of these malformations are recognized, including the more common intracererbral or brain arteriovenous malformation. Other types include cavernous angiomas (cavernoma), venous malformations or capillary telangiectasis; these seldom bleed.

Brain arteriovenous malformations

Brain arteriovenous malformations consist of abnormal communications between arteries and veins and sometimes direct arteriovenous fistulas; the abnormal communications are 50 to 250 µm, with a nidus fed by one or more enlarged feeding arteries and draining veins. The nidus has a sponge-like appearance. They vary very widely in size from tiny lesions of less than 1 cm to large lesions of 6 cm or more. Berry aneurysms can form on feeding arteries to brain arteriovenous malformations; they have an increased risk of rupture and greater complexity of treatment.

Brain arteriovenous malformations may present with either haemorrhage, epilepsy or neurological deficit, or they may be found incidentally. Haemorrhage can be into the subarachnoid layer, intraparenchymal, or intraventricular. Ischaemia can be due to a ‘steal' phenomenon whereby the malformation acts as an arteriovenous fistula and blood is diverted from nearby cerebral tissue, causing ischaemia. Compression and gliosis around these malformations can cause epilepsy.

Subarachnoid haemorrhage

Subarachnoid haemorrhage occurs when an intracranial aneurysm ruptures, with haemorrhage into the subarachnoid layer around the brain. However, aneurysms can present with pure subdural, intraparenchymal (intracerebral), or intraventricular haemorrhage (Fig. 1). Subarachnoid haemorrhage may be caused by other intracranial lesions such as in head injury, haemorrhage stroke, and brain arteriovenous malformations.

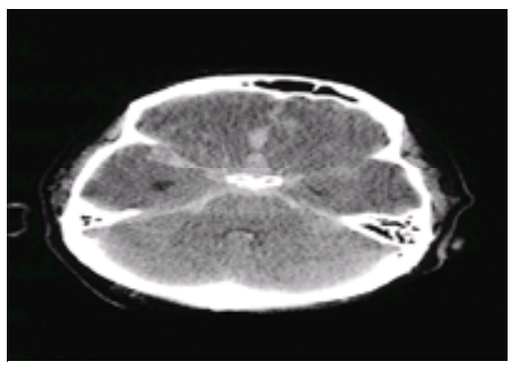


Fig. 1. Large intraventricular haemorrhage.

Cerebral aneurysm and subarachoid haemorrhage

Cerebral aneurysm

Incidence

Intracranial aneurysms can be found at any age, but more commonly in adults, with 1 to 3 per cent of adults found to be affected at autopsy. The yearly incidence of rupture is low. Recent evidence suggests that the risk of bleeding from a previously unruptured aneurysm (less than 10 mm in diameter) is less than 0.1 per cent per annum. However, rupture of an intracranial aneurysm (subarachnoid haemorrhage) is a devastating event with a 30 per cent immediate mortality. The 70 per cent who survive the initial ictus may suffer neurological damage ranging from minimal to severe. The risk for further rupture of a previously ruptured aneurysm is substantially higher in the first 2 weeks (approximately 20 per cent), with a 60 per cent mortality, than after 6 months (3 per cent per annum). The frequency of aneurysmal subarachnoid haemorrhage is estimated to be between 6 to 12 per 100 000 per annum in most Western populations.

Aetiology

The cause of intracranial aneurysm remains unknown. Extensive genetic studies have failed to find a common molecular biological basis. Some inherited conditions increase the risk of intracranial aneurysm formation, most notably adult polycystic kidney disease. However, the vast majority as yet do not have a known genetic basis, apart from epidemiological evidence that consanguineous first-degree relatives have a fourfold increase in suffering a subarachnoid haemorrhage. We do not as yet understand enough about the natural history to justify the risks of investigation and treatment following screening; at present, screening for intracranial aneurysm is not justified.

Pathology

Intracranial aneurysms may cause neuronal damage by compression or rupture. Compression may cause pain, dysfunction, or epilepsy. Pain is usually due to meningeal irritation, but vascular stretching may also be a component. Dysfunction can result from the compression of cranial nerves, nearby vasculature, or brain.

Brain compression may produce focal neurological changes or, more rarely, epilepsy. Rupture usually results in subarachnoid haemorrhage, but may bleed directly into the brain (intracerebral; see Fig. 1) or occasional into other intracranial layers such as subdural or intraventricular.

Subarachnoid haemorrhage

Subarachnoid haemorrhage leads to a chemically induced meningitis, vasospasm of the cerebral arteries, and acute obstructive or subacute communicating hydrocephalus. The cause of the cerebral vasospasm is unknown; it is the most common cause of further neurological deterioration in the presence of adequate treatment of the aneurysm and hydrocephalus. Cerebral vasospasm occurs in 60 per cent of survivors and causes delayed neurological deterioration in 25 to 35 per cent, of whom some will slowly recover.

Clinical presentation

Typically, patients present with a sudden (within a fraction of a second) and unusual (never experienced before) headache, which is frequently described as ‘like being hit on the back of the head' and many sufferers accuse people nearby of assault. Migraine sufferers struck down by subarachnoid haemorrhage can usually distinguish between their familiar migrainous pain and the ictal headache of the haemorrhage.

The level of consciousness may be reduced and there may be focal neurological signs associated with neck stiffness from the chemical meningitis. Vomiting usually accompanies the headache. Thirty per cent of patients will die at the ictus. The clinical status of the patients should be record according to the World Federation of Neurosurgeons (**WFNS**) scale, in which the Glasgow coma scale (**GCS**) is used to measure consciousness: grade 1 = headache + GCS 15; grade II = GCS 13–14, no major neurological deficit; grade III = GCS 13–14 with focal deficit; grade IV = GCS 7–12; grade V = GCS 6–3.

Rebleeding (approximately 4 per cent in the first 24 h and 16 per cent in next 2 weeks) is the most dangerous sequel of aneurysmal subarachnoid haemorrhage. It is associated with a mortality of the order of 60 per cent and significant higher risk of poor outcome. Other causes of deterioration include hydrocephalus (obstructive or communicating) and delayed ischaemic neurological deficits or vasospasm may cause secondary deterioration.

Clinical management is aimed at maintaining the patient's airway, blood pressure, and pain control, since poor ventilation, myocardial complications, and severe head pain all cause further deterioration. The next step is to establish the diagnosis by appropriately timed investigations. Any patient with sudden, unusual headache must be investigated to exclude subarachnoid haemorrhage as an emergency.

Investigation

The timing of appropriate tests is important in preventing overinvestigation of headaches not due to subarachnoid haemorrhage. Computed tomography (**CT**) is the technique of choice in establishing the diagnosis. A CT scan within the first 24 h of the ictus (Fig. 2) carries a 95 per cent sensitivity, falling to 50 per cent at 3 days; urgent access to CT scanning is therefore essential. If the CT is normal or equivocal, lumbar puncture is mandatory to look for xanthochromia (preferably after 3 h of ictus to allow bilirubin to accumulate). The sample should be spun (3000 rev/min for 10 min) soon after collection to separate out the red blood cells, which otherwise may continue to lyse, supply further oxyhaemoglobin, and allow bilirubin to form *in vitro.* The presentation of acute bacterial meningitis can simulate that of subarachnoid haemorrhage, so microbiological examination of the cerebrospinal fluid should also be requested.

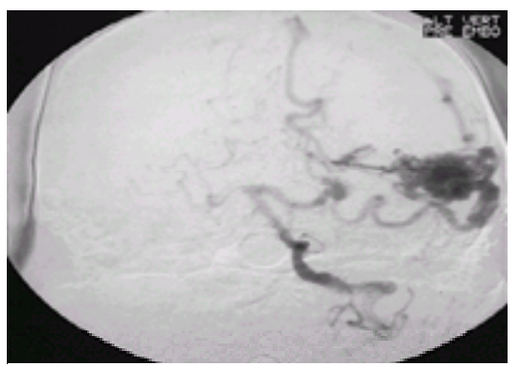


Fig. 2. CT scan of acute subarachnoid haemorrhage.

Cerebral angiography (Fig. 3) remains the investigation of choice for establishing the cause of the subarachnoid haemorrhage. Magnetic resonance or CT angiography do not yet have equivalent resolution, although recent developments with contrast agents may improve the spatial resolution. The most common cause of subarachnoid haemorrhage is a ‘berry' aneurysm in patients over 40 years of age, with arteriovenous malformation more common in patients under 40 years.

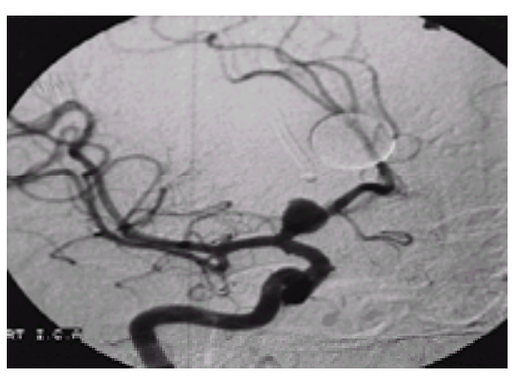


Fig. 3. Angiogram of internal carotid artery aneurysm before treatment.

Negative cerebral angiography presents a challenge to further management. False-negative angiography (CT scan positive) has been described systematically. Some patterns of haemorrhage in the region of the pons, known as perimesencephalic haemorrhage, are consistently associated with negative angiography. However, some negative angiograms are due to temporary thrombosis of a ‘berry' aneurysm or vasospasm of feeding vessels, with subsequent risk of repeated haemorrhage and death. Angiography should be repeated where clinical doubt remains, but must be balanced against the risks of the investigation (less than 0.1 per cent risk of neurological complications when done by an experienced neuroradiologist).

Treatment of patients with aneurysmal subarachnoid haemorrhage

Medical management includes bed rest, adequate intravenous fluids (3 litres normal saline/24 h), pain relief, and monitoring of blood pressure, tissue oxygen saturation and neurological status.

Prevention of rebleeding

The greatest risk to a patient with subarachnoid haemorrhage from an established berry aneurysm is rebleeding, which carries about a 60 per cent chance of fatality. Intervention is aimed at stopping the rebleeding; this can be achieved by the surgical application of a clip around the neck of the aneurysm, and has been the standard treatment for many years. Recently, endovascular techniques using detachable platinum coils (Fig. 4) have been introduced to occlude the aneurysm and prevent

rebleeding, thus avoiding the risks associated with craniotomy.

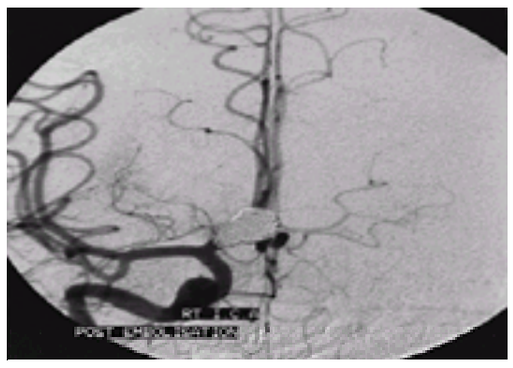


Fig. 4. Angiogram of aneurysm after treatment with platinum coils.

Surgical treatment of aneurysms

Timing Having decided that open intervention via a craniotomy is the appropriate treatment, the next controversial decision revolves around timing. Early surgery (within 3 days) reduces the incidence of early rebleeding, avoids operating with vasospastic arteries, but is technically more demanding as the surgeon is operating around swollen, friable brain. Delayed surgery (after 10 days) allows the brain and arteries to settle after the initial ictus, but the patient must accept the additional 10 per cent mortality from rebleeding. Various neurosurgeons have investigated outcomes from either approach but there has been no rigorous, randomized prospective trial to clarify any difference between early and late surgical intervention.

Approach A multitude of different approaches, both traditional and ‘micro', are described for aneurysms of both the anterior and posterior circulations. Standard approaches to aneurysms of the anterior circulation include a frontal incision with the raising of a pterional free or vascularized bone flap and exposure of the sylvian fissure.

The main principles governing the intradural component of the operation are to secure proximal control of the feeding artery(s), remove cerebrospinal fluid either directly or via a lumbar/ventricular drain, and dissect microscopically the arachnoid layers to expose the aneurysm neck.

Clipping the neck Applying the clip (magnetic resonance imaging-compatible where possible) to the neck of the aneurysm requires knowledge of the feeding artery and its subsequent branches. Incomplete exclusion of the aneurysm for the circulation may lead to rebleeding or reformation. Incorporating branches of the parent vessel may be lead to ischaemic injury and death from brain swelling. It can be extremely difficult to exclude some aneurysms from the circulation, and a compromise between cure and neurovascular injury must be accepted. Some giant aneurysms have thick, calcified walls and external clipping may not be possible; resection of these lesions requires significant periods of operative ischaemia. Brain cooling and other methods of cerebral protection have been tried with varying success.

Incomplete exclusion of the aneurysm can be supported by various preparations such as skeletal muscle (temporalis), acrylic cement, and cottonoid wool. These additional methods have not be shown to alter significantly the incidence of rebleeding and can produce comorbidity from granulation/fibrotic reactions affecting nearby cranial nerves.

Postoperative care This requires rigorous attention to fluid balance, blood pressure, and neurological status. Neurological deterioration should be investigated, with suitable blood tests to exclude hyponatraemia, estimation of ventilatory status (blood gases), and CT scanning to exclude postoperative haematoma. Transcranial Doppler may be helpful following the development of vasospasm, but can be difficult to interpret in the postoperative period. Postoperative cerebral angiography can help in demonstrating occlusion of the aneurysm and the presence of vasospasm.

Patients can be mobilized when the risk of vasospasm (usually after 5 days) has subsided. Nimodipine, 60 mg 4-hourly, where started preoperatively, should be continued for 21 days. Neurological outcome can be assessed between 6 months and 1 year postoperatively.

Endovascular treatment of intracranial aneurysm

Description of technique The endovascular treatment of intracranial aneurysms is achieved by the use of detachable platinum coils. These are introduced via a series of catheters from the femoral artery, with placement of a guiding catheter in the internal carotid or vertebral artery, under digital fluoroscopy using a modern digital angiographic system. A microcatheter (1.8–2 F or 1.5 mm) and fine guidewire (0.010–0.014) are introduced into the cerebral circulation and navigated under fluoroscopic guidance into the aneurysm. A series of platinum coils is then placed, commencing with one approximately to the size of the aneurysm. These coils come in a helical configuration with a diameter ranging from 2 to 20 mm. When it has been satisfactorily positioned, the coil it is then detached from the introducing wire by electrolysis of the junction (Guglielmi detachable coil; Target Therapeutics). When a ‘basket' has been formed in the centre it is packed with small coils until dense occlusion of the aneurysm is achieved, producing physical obstruction of blood flow into it.

Angiography is used during and at the end of the procedure (see Fig. 3 and Fig. 4) to confirm the occlusion. Further follow-up angiography is performed, usually about 6 months, later to confirm that the occlusion is stable.

Selection of patients for treatment

Aneurysms in the posterior fossa (basilar derived) are best treated with intravascular methods using detachable platinum coils if possible because of the relatively higher surgical risks in this location. Intervention via craniotomy or the intravascular route for intracranial aneurysms is currently equipoised and forms the basis for a prospective randomized study, the International Subarachnoid Aneurysm Trial (ISAT), which is under way.

Hydrocephalus

Communicating hydrocephalus can be treated by repeated lumbar puncture and usually resolves over 10 days from the ictus. Obstructive hydrocephalus is best treated by a ventricular drain or ventriculoperitoneal (or atrial) shunt.

Haematoma

Haematoma associated with midline shift and reduce consciousness, particularly when outside the parenchyma, can be considered for craniotomy and evacuation; this can either be accompanied by clipping of the aneurysm or preceded by intravascular occlusion.

Management of vasospasm

Cerebral vasospasm following subarachnoid haemorrhage is an important cause of additional and potentially avoidable neurological deterioration. Increasing the amount of blood in the subarachnoid space increases the risk of vasospasm. Overall 25 to 35 per cent of patients with subarachnoid haemorrhage will suffer delayed ischaemic neurological deficits due to vasospasm. With current medical management the majority of these patients will suffer permanent neurological deficits as a result.

Medical management of cerebral vasospasm in subarachnoid haemorrhage relies on maintaining adequate blood pressure with intravenous fluids (mixture of crystalloid and colloids, 3.5 l/day), since cerebral perfusion is related to mean arterial pressure and intracranial pressure. The haematocrit can be allowed to fall and inotropes used to push the mean arterial pressure to supraphysiological levels—triple H therapy. A number of other complex metabolic responses also occur during this period (usually day 3 to day 10 post ictus), such as natriuresis, hyponatraemia, and a variety of cardiac arrhy-thmias.

Prognosis

The prognosis after subarachnoid haemorrhage is largely determined by the initial clinical grade of the patient (WFNS), the amount of intracranial blood (Fisher grade), age, and comorbidity such as hypertension. Patients below the age of 70 years with an initial clinical grade II or better should have a 60 to 80 per cent chance of good or excellent outcome, such as return to work. At the other end of the clinical grades, 70 per cent mortality can be expected in grade V patients, with a few per cent achieving excellent outcomes.

The additional factors that cause deteriorating outcome include rebleeding, cerebral vasospasm, and hydrocephalus; these are potentially avoidable and are best managed in specialist units by those with an interest in neurovascular conditions.

Despite considerable improvements in early diagnosis, prompt intervention to stop rebleeding, and intensive medical management the outcome from subarachnoid haemorrhage, when input grade is compared to outcome, is still worse than for other types of cerebrovascular accident.

Treatment of brain arteriovenous malformations

The risks of rebleeding from brain arteriovenous malformations are much lower than from an aneurysm; thus there is less urgency about treatment. The rebleeding rate is about 2 to 4 per cent per annum. Treatment is aimed at preventing rebleeding or, occasionally, improving focal neurological deficits. The surgical resection of extensive arteriovenous malformations in eloquent areas of the brain is impossible, but these can be treated by a combination of intravascular embolization and stereotactic radiosurgery using a precisely focused radiotherapy or g-ray beam. Some arteriovenous malformations are untreatable by any technique.

The management of patients with brain arteriovenous malformations is complex and depends on a multitude of factors: age, presentation, location and size of the malformations, and particularly whether it is in an eloquent area of the brain. If a decision is made to treat, then the options areas follows.

1. Surgery, usually for accessible and small arteriovenous malformations.
2. Embolization using microcatheters in feeding arteries and tissue adhesive to obliterate the residue of the malformation (Fig. 5 and Fig. 6). Only about 15 per cent of arteriovenous malformations can be completely obliterated by this method but their size can often be reduced significantly, which makes the lesion more accessible to stereotactic radiation or surgical removal.

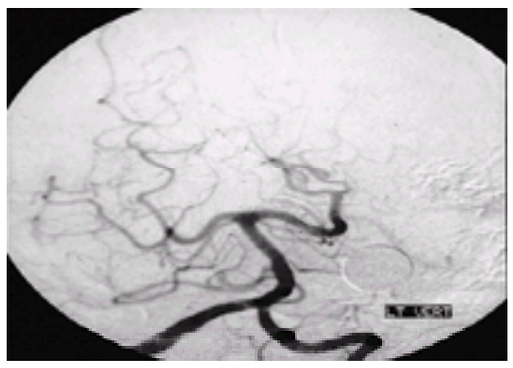


Fig. 5. Angiogram of brain arteriovenous malformation, left temporal lobe, pretreatment.



Fig. 6. Brain arteriovenous malformation embolized, post treatment.

3. Stereotactic radiosurgery using a gamma knife unit; this consists of about 200 separate cobalt sources and provides highly controlled, focused radiation treatment of brain lesions. In small lesions (less than 1 cm), radiosurgery is over 90 per cent successful but takes 2 years for full effect; for larger lesions the success rate falls. This type of radiation can also be performed by stereotactically adapted linear accelerators, although data on whether the technique is as effective are lacking.
4. A combination of the above treatments is frequently used.

Further reading

Bromberg JE, Rinkel GJ, Algra A, Limburg M, van Gijn J. Outcome in familial subarachnoid hemorrhage. *Stroke* 1995; **26**(6): 961–3.

Caplan LR. Should intracranial aneurysms be treated before they rupture. *New England Journal of Medicine* 1998; **339**: 1774–5.

Fisher CM, Kistler JP, Davis JM. Relation of cerebral vasospasm to subarachnoid haemorrhage visualised by computerised tomographic scanning. *Neurosurgery* 1980;**6**: 1–9.

International Study of Unruptured Intracranial Aneurysms. Unruptured intracranial aneurysms—risk of rupture and risks of surgical intervention. *New England Journal of Medicine* 1998; **339**: 1725–33.

Shaw MDM. Chs 11 and 12. In: Miller JD, ed. *Northfield's surgery of the central nervous system*, 2nd edn, p. 378.

44.10 Spinal surgery

C. B. T. Adams

Introduction—principles of diagnosis
Local features
Root features
Spinal cord features
Other signs
Spinal investigations
Plain radiographs
Myelography
Computed tomography
Magnetic resonance imaging
Spinal angiography
Spinal tumours
Extradural tumours
Intradural extramedullary tumours
Intramedullary tumours
Cauda equina tumours
Spondylosis and disc prolapse
Introduction
Cervical spondylosis
Acute cervical disc protrusion
Lumbar disc protrusion
Lumbar spondylosis
Thoracic disc prolapse
Sciatica and back pain
Postoperative pain
Further reading

Introduction—principles of diagnosis

A thorough knowledge of fairly basic anatomy and neurology is necessary to enable disorders of nerve roots or the spinal cord to be understood. Spinal disorders, including tumours, may present with combination of the following four groups of features.

Local features

Disorders of the spinal column often produce localized pain. If the disorder actually arises from the spinal column, such as a metastatic tumour in a vertebral body, localized pain may well predate other clinical features such as disorders of the nerve roots or spinal cord. A metastatic tumour in the spine should be suspected in any patient with a known primary tumour who develops backache, often between the shoulder blades. Such patients often present late with profound and irreversible damage to the cord. Tenderness on palpation or percussion of the spinous processes is often a very helpful physical sign. Localized deformity or absence of a spinous process on palpation are other useful signs.

Root features

A characteristic of root pain is that it is often worse by night than by day and may wake the sufferer. It is relieved by walking or sitting upright in a chair. Nerve-root pain is often thought of as a shooting pain radiating in a dermatomal distribution, often associated with pins and needles, and both being made worse by coughing, sneezing, or straining. This may be so, for instance, with a lumbar disc protrusion compressing the L5 or S1 root, but root pain due to a tumour more commonly produces an aching pain that is worse on inactivity.

The features of a lower-neurone lesion are wasting, weakness, hypotonia, and associated reduction of the tendon reflexes. It is important to be able to test the relevant myotomes; the outline in Table 1 has withstood the test of time and busy clinical practice, although other clinicians' lists may differ. Thus a C5 root lesion will cause weakness of shoulder abduction and a C6 lesion weakness of elbow flexion and an absent or impaired biceps jerk. A C7 root lesion causes weakness of elbow extension and reduction or absence of the triceps jerk, while weakness of finger flexion denotes a C8 root lesion. A TI lesion produces weakness of the small hand muscles. A series of tests can therefore be used to work from C5 to T1 quickly and accurately.

Arm	
Shoulder abduction	C5 (deltoid jerk)
Elbow flexion	C6 (biceps jerk)
Elbow extension	C7 (triceps jerk)
Finger flexion	C8 (finger jerk)
Small hand muscles	T1
Leg	
Hip flexion	L2/3
Knee extension	L3/4 (quadriceps jerk)
Ankle inversion	L4/5
Plantar flexion	S1/2 (ankle jerk)
All others	L5/S1
	Hip extension
	Knee flexion
	Ankle eversion
	Dorsiflexion of the foot and toes

Table 1 Important myotomes

Testing the myotomes in the legs is a little more complicated, but hip flexion is dictated by the L2/3 myotomes, knee extension by the L3/4 myotomes, ankle inversion by L4/5, and plantar flexion by S1/2. All of the other commonly tested manoeuvres [hip extension (gluteus maximus), knee flexion (hamstrings), ankle eversion, mid-dorsiflexion of the foot and toes] are controlled by the L5/S1 myotomes. An LS root lesion can be differentiated from a lesion in S1 by examining the ankle reflex; this is equivalent to plantar flexion and contains no L5 myotome. Thus absence of the ankle jerk indicates an S1 lesion; weakness of the L5/S1 myotomes and an intact ankle jerk is due to an L5 root lesion.

The maximum circumference of the calves should be measured. Any difference of more than 1 cm is suggestive of a lower motor-neurone lesion, provided there are no other obvious abnormalities.

A long-established lower motor-neurone lesion will produce fasciculation of the muscles of that myotome. Pain may also arise as a result of a motor root being affected, the pain following the myotome distribution of the root. This is an intense, aching type of pain, more ill defined than sensory nerve-root pain. Thus a C7 motor-root pain is in the region of the pectoralis major and radiates down the back of the arm to the triceps muscle; on the left side it is not infrequently misdiagnosed as pain from myocardial ischaemia.

Sensory root symptoms, predominantly those of pain in a dermatomal distribution, are often associated with pins and needles. Patients may complain of hyperalgesia—an intense, unpleasant, and unusual pain in response to the slightest painful stimulation of the skin. Hyperaesthesia, the heightened and unpleasant appreciation of light touch, may also occur. Hyperaesthesia is well known in association with herpes zoster or postherpetic neuralgia; affected patients are usually hypoalgesic in the hyperaesthetic area. The distribution of dermatomes is best tested where there is least overlap from adjacent dermatomes. Since three adjacent nerve roots need to be cut before dermatomal sensory loss is appreciated, because of the degree of overlap, it is more important to know where to test for each

dermatome than to know their rather variable geographic boundaries. Table 2 provides a list of the more important dermatomes.

C4	Top of shoulder
C5	Lateral side of upper arm (over deltoid insertion)
C6	Thumb
C7	Middle finger
C8	Little finger
T1	Medial (ulnar) forearm
T2	Medial (ulnar) upper arm
T3	Axilla
T4	Nipple line
T6	Xiphoid process
T10	Umbilicus
T12	Groin
L1	Upper 1/3 of front of the thigh
L2	Middle 1/3 of front of the thigh
L3	Lower 1/3 (knee cap) of front of the thigh
L4	Medial malleolus
L5	Middle three toes
S1	Lateral malleolus extending along anterior border of foot
S2	Back of leg
S3	Gluteal fold

N.B. Proprioceptive sensory loss in other areas before perianal sensory loss with cauda equina compression and should be sought for; 'saddle anesthesia' sensation is always associated with, at least, proprioceptive sensory loss.

Table 2 Important dermatomes: areas of least overlap

Spinal cord features

Compression of the spinal cord is associated with disorders of motor function, causing an upper motor-neurone lesion, impaired sensation, and abnormal bladder and bowel function below the level of the lesion.

Pain rarely results from pressure on the spinal cord. Any such pain that occurs is described more as an altered and unpleasant sensation (dysaesthesia). Numbness or tingling of the legs is a common and early symptom. Patients frequently describe a feeling of a tight band around the trunk or limb; this feeling often indicates a lesion of the posterior columns of the spinal cord. Impairment of pain and temperature sensation may produce symptoms of burning or coldness, but there are often no symptoms other than complaints of painless burns or other injuries. Careful enquiry about sensory symptoms will allow the clinicians to focus sensory testing on the appropriate area.

The first motor symptom to appear is often that of ‘stiffness' of the legs: scuffing may cause the patient to trip on uneven surfaces. Weakness and disability follows, and when a marked degree of spasticity exists the patient may develop involuntary jerking of the legs. Spontaneous ankle clonus may also develop: this is characteristically aborted by plantar flexion of the foot.

Symptoms of bladder and bowel dysfunction should be sought. The most reliable symptom is lack of a normal feeling of a full bladder; the patient also fails to feel urine passing down the urethra, and may know when the bladder is empty only by cessation of the noise of urine flowing into the receptacle. Patients may also be unable to decide whether they are about to pass flatus or faeces, and men will develop impotence. These symptoms also occur in patients with cauda equina (nerve root) disorders.

Examination of a patient with damage to the spinal cord discloses an upper motor-neurone lesion of the legs, trunk, and arms, depending on the level of the lesion. The deltoid jerk, which is not a tendon reflex but a muscle reflex, should be tested. Tapping the normal deltoid muscle does not elicit a reflex contraction; in the presence of an upper motor-neurone lesion a reflex response occurs, indicating that the lesion is at C4 or above.

Examination of sensation in these circumstances should be directed at finding a sensory level. This level may be abrupt and obvious where profound lesions exist, when the diagnosis has been made too late. In early disease, considerable persistence may be required before a sensory level can be determined. Normally there is heightened appreciation of pinprick on passing from the abdomen to the costal margin. Patients often find it easier to appreciate a change in sensation when a pin is dragged lightly up the body. In the thoracic region a sweat level is often present, due to the body being sympathectomized below the lesion: this may be useful in an unconsciousness patient with suspected damage to the spinal cord. When looking for suspected dermatomal sensory loss it is better to dissuade patients from sensory signs, for if too assiduously looked for, they will describe a normal variation of skin sensitivity. However, when looking for a sensory level one has to be most obsessive. If a possible level is found, testing further up the trunk will show whether it is genuine or just a variation in degree of sensory loss beneath the real sensory level. The finding of a root lesion above the level denotes the true pathological level: the sensory level does not attain the pathological level until the cord compression is profound.

In early stages of cord compression the sensory loss is in the legs; this then gradually ascends up the trunk. Initially there is spacing in the buttock region, unless the tumour is at the level of the conus medullaris or cauda equina.

Intramedullary tumours or syringomyelia produce a different picture from that of other spinal tumours. They first affect the spinothalamic fibres crossing in the anterior commissure, causing a ‘suspended dissociated sensory loss'. There is loss of pain and temperature sensation, but not light touch, at the level of the lesion. Progression of the intramedullary lesion causes caudal extension of the sensory loss, eventually producing a typical sensory level. As an intramedullary lesion is often elongated; lower motor-neurone signs, especially absent tendon reflexes, are common and obvious at the level of the lesion.

It is unusual to find a pure Brown–Sequard syndrome, but the incomplete form is usually caused by a soft cervical disc protrusion, although neurologists more often see this syndrome in patients with multiple sclerosis.

Other signs

It is important to examine the rest of the body. Tumours elsewhere may metastasize to the spinal column, and may also extend intradurally. Careful examination of the skin may show evidence of neurofibromatosis, while a port-wine stain, especially in a dermatomal distribution, should suggest spinal angioma. Auscultation over a spinal angioma may reveal a bruit.

The dissociation of the vertebral and the neurological levels that occurs because the spinal cord ends at about the lower border of the L1 vertebral body should be remembered. The ‘thoracic cord' (the T12 segment of the spinal cord) ends at the lower border of the ninth thoracic vertebral body. The lumbar plexus therefore arises in the spinal cord opposite the T10 and T11 bodies, while the sacral plexus arises opposite the T12 and L1 bodies. Failure to appreciate this dissociation will prevent investigations being made at the appropriate spinal level (usually not high enough), and the diagnosis will not be reached. It is also worth knowing that the C8 neurological (or myelomere) level is opposite the C6/7 disc space.

Spinal investigations

Careful clinical assessment is essential, not only to determine the most relevant investigation but to ensure appropriate correlation with any abnormality seen. This warning is especially relevant in evaluation of disorders of the back or neck, where ‘bulging discs' are normal and osteophytes the rule in patients over the age of 40 years (Fig. 1 and Fig. 2).

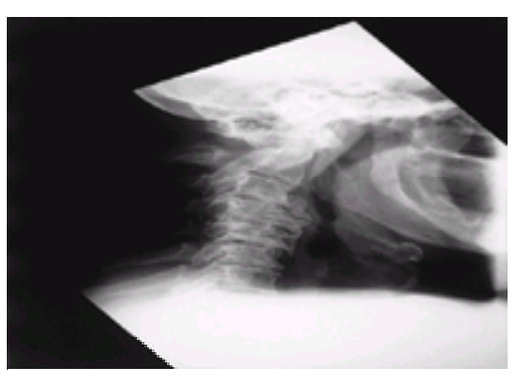


Fig. 1. Lateral cervical radiograph showing severe cervical spondylosis. However, this patient had no symptoms relating to this. Hence the importance of a clinical assessment by taking a history and examining the patient. The cervical spine has a very low range of movement on flexion–extension, which is probably why no neurological complications developed.



Fig. 2. Myelogram: oblique view showing apparent compression of the right L5 and S1 roots crossing the L4/5 and the L5/S1 disc spaces, respectively. However, the patient was asymptomatic and had no back pain or right-sided sciatica. This is shown to emphasize the importance of treating patients rather than their radiographs or scans.

Plain radiographs

It is important to obtain plain radiographs at the appropriate level, after careful neurological assessment, bearing in mind the dissociation between the vertebral and neurological levels (see above). Collapse, erosion, and scalloping of vertebral bodies (especially the pedicles) may be seen. Expansion of the intervertebral foramen or the diameters of the spinal canal indicates a slow-growing, expansive lesion. Lateral radiographs of the cervical spine with the neck fully extended and flexed are particularly helpful when assessing patients with possible cervical spondylotic radiculopathy or myelopathy.

Myelography

Myelography is no longer the standard method of investigation but may still be used if a water-soluble dye is injected into the subarachnoid space by lumbar or cisternal puncture. Myelography demonstrates compression of the spinal cord; very occasionally, rapid deterioration of function may follow myelography and require immediate neurosurgical intervention. Failure to obtain cerebrospinal fluid should suggest the possibility of a tumour of the cauda equina, especially if the patient experiences severe root pain during the attempted lumbar puncture. The most common error during myelography is failure to run the dye high enough. Myelography to investigate back pain and sciatica should routinely include examination of the midthoracic region to exclude a spinal angioma or a high intradural tumour. Use of the term ‘radiculogram', implying mere examination of the nerve roots, is inadvisable; such an examination should never be performed. Despite the advent of magnetic resonance imaging (**MRI**), myelography is still the best investigation for cervical spondylotic radiculopathy.

Cerebrospinal fluid obtained during the lumbar puncture should be sent to the laboratory for analysis. If it is yellow (xanthochromic), either the patient has suffered a subarachnoid haemorrhage or the protein concentration is abnormally high, indicating a complete or near-complete obstruction to fluid flow in the spinal canal.

Computed tomography

Scanning by computed tomography (**CT**) is of particular value for demonstrating abnormalities in the vertebrae, such as the extent of metastatic tumour and associated paraspinal spread. CT-guided biopsy of neoplastic or inflammatory vertebral disease is extremely useful and avoids many fruitless operations (Fig. 3).



Fig. 3. CT scan after a myelogram to show contrast in the subarachnoid space (white) and an extradural lesion compressing the theca from the right. The transverse process is expanded by tumour. The appearances are typical of a lymphoma. CT-guided biopsy avoids open surgery.

CT scanning alone is of more limited value for demonstrating pathology within the spinal canal, although its value is much enhanced by administration of intrathecal contract medium. Postmyelography CT scanning is the best method of investigation for some lesions of the spinal canal and disc protrusions (especially thoracic disc protrusions). CT scanning alone may show a pathological prolapse of the lumbar disc but false-positive and false-negative results are common. MRI is much more accurate and should be used in preference to CT for lumbar disc disease (Fig. 4).



Fig. 4. MRI: horizontal cut through the lumbar spine to show a L5/S1 disc protrusion compressing the left S1 nerve root. The right S1 nerve root can be clearly seen outlined by (white) fat. This patient's pain subsided spontaneously and surgery was not therefore advised.

Magnetic resonance imaging

MRI is the investigation of choice for spinal disorders, including lumbar disc disease and intradural extramedullary and intramedullary lesions of the spinal cord, especially when used with gadolinium enhancement. It shows normal, degenerative, and protruding discs particularly well, and is also useful in demonstrating Arnold–Chiari malformations at the spinal–occipital junction. Calcification is not shown by MRI, and this limits its use in the assessment of bony (vertebral) disease (Fig. 4).

Spinal angiography

Spinal angiography is used to delineate spinal angiomas that have been previously demonstrated by myelography (Fig. 5). Before intravascular embolization of spinal angiomas is undertaken the exact vascular anatomy of the feeding arteries, the fistula, and the draining veins must be determined, Angiography is the only investigation able to do this. It is rarely used in the investigation of other spinal conditions.

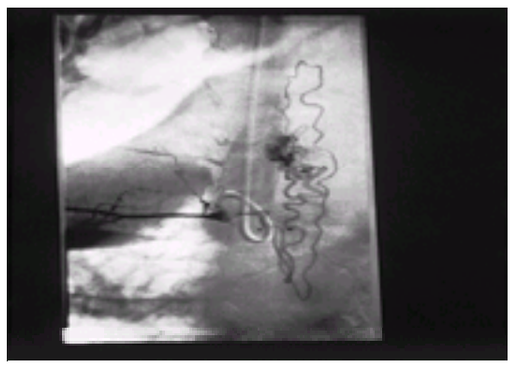


Fig. 5. Spinal angiography demonstrating a spinal arteriovenous malformation. These can be diagnosed on myelography but the precise anatomy of the feeding vessels, fistula, and draining veins can only be elucidated on angiography.

Spinal tumours

Extradural tumours

The most common cause is a metastatic deposit from the lung, breast, or prostate. Primary tumours of bone, typically myeloma, or of the reticuloendothelial system such as lymphoma, are not rare (Fig. 6 and Fig. 3). Less common primary tumours, such as chordoma, aneurysmal bone cyst, and osteoblastoma (in children), also occur. Inflammatory diseases, such as tuberculosis or staphylococcal osteomyelitis with extradural pus, are still common in certain parts of the world. A thoracic disc protrusion may present as progressive compression of the spinal cord.

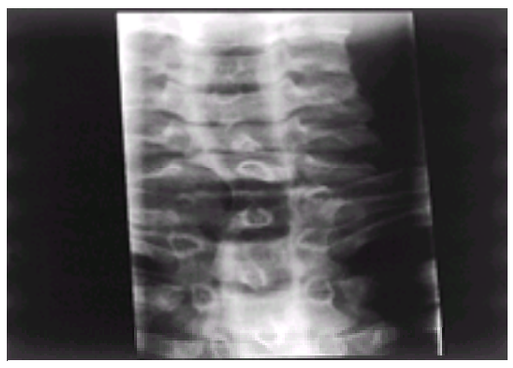


Fig. 6. Myelogram showing an extradural tumour; see Fig. 3 for the postmyelogram CT scan, which is more informative.

Intradural extramedullary tumours

These lie within the subarachnoid space (Fig. 7 and Fig. 8). Meningiomas (which are particularly common in the thoracic region of middle-aged women) and neurofibromas are the most common tumours in this group. A neurofibroma arising from any nerve root may extend through to the extrathecal space and on through the intervertebral foramen (dumb-bell tumour). Arachnoid cysts and dermoid cysts also occur. Arteriovenous malformations are partly extramedullary and partly intramedullary, and may also cause progressive cord damage, as well as sudden damage by haemorrhage.

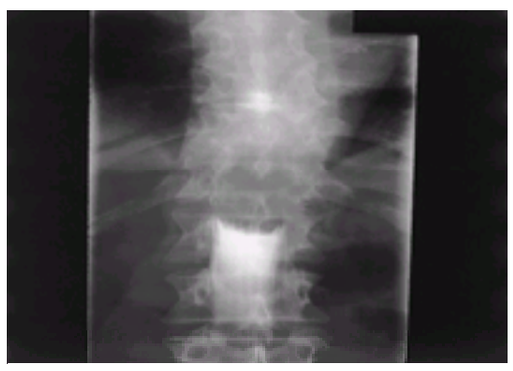


Fig. 7. Myelogram showing an intradural extramedullary tumour.



Fig. 8. Myelogram: lateral view showing an intradural extramedullary tumour pushing the lower extremity of the spinal cord and the roots of the cauda equina anteriorly; the semilunar shadow of the lower pole of the tumour in the ‘expanded' subarachnoid space is characteristic.

Intramedullary tumours

These rare and slow-growing tumours are usually astrocytomas and ependymomas (Fig. 9); both types may be cystic and require differentiation from syringomyelia.



Fig. 9. Myelogram showing an intramedullary tumour (ependymoma) in an 8-year-old girl who presented with severe back pain; the tumour was totally removed without neurological deficit.

Clinical features

The essential features that should alert the clinician to compression of the spinal cord are symptoms or signs indicating a sensory level, especially if with a history of progression. The clinical features in any one patient are a combination of vertebral or local, nerve root, and spinal cord.

Which of these groups of symptoms predominates depends on the cause of the compression. Extradural (vertebral) causes present with back pain. Intradural extramedullary tumours present often with nerve-root pain before cord compression, while vertebral pain is uncommon or late. Intramedullary lesions present with a dissociated suspended sensory loss, like syringomyelia, but further progression then causes typical features of cord compression.

Differential diagnosis

Cervical spondylotic myelopathy and acute cervical disc protrusions may mimic compression of the spinal cord; they will be discussed later. Spinal angiomas are extremely rare, but their diagnosis is important, since they are treatable. Acute transverse myelitis due to demyelinating disease may occasionally cause compression, with a block seen on the myelogram due to the acute necrosis and swelling of the spinal cord.

Investigation

Plain radiographs (including the chest) often show evidence of collapse of vertebral bodies or pedicular erosion if a metastatic lesion is present. Expansion of the spinal canal or intervertebral foramen indicates a benign tumour. Myelography is necessary for definitive diagnosis, and CT delineates any tumour extension into the vertebral bodies or paraspinally (Fig. 3). MRI is the best method to delineate intramedullary (cord) lesions.

Treatment

Results of treatment depend on the severity of compression; early diagnosis is essential for good recovery. The stability of the spinal column must be preserved: further damage to the cord can occur if the spine is rendered unstable.

Stability of the spine

It is convenient to consider the stability of the spine as depending on two columns. The anterior column consists of vertebral bodies, discs, and ligaments; the posterior column consists of the facet joints, neural arch, and ligaments. Chronic instability may result from damage to just one column. Laminectomy in children, especially when done before the growth spurt, may produce severe deformity and subsequent cord compression; all children treated in this way should be carefully monitored. Early bracing will prevent severe deformity. ‘Acute' instability occurs if both columns are disrupted, for instance following decompressive laminectomy in a patient with a metastatic tumour causing collapse of a vertebral body.

The management of cord compression

The management of cord compression relies on the maintenance of spinal stability. Thus extradural tumours causing a collapsed vertebral body should not be treated by laminectomy. A CT-guided biopsy will establish the diagnosis and radiotherapy will control the pain. If a vertebral body is affected by a slow-growing and reasonably confined tumour, and the patient is otherwise fit and with some residual function of the spinal cord, the vertebral body may be approached from the anterior. Excision of the tumour and replacement of the vertebral body with either bone, acrylic, or a metal prosthesis should be combined with a stabilization procedure. Such major surgery allows reasonable recovery of neurological function.

An extradural tumour confined to the spine and lamina and producing posterior compression is properly treated by laminectomy and excision: this will not cause acute instability.

Intradural extramedullary tumours require treatment by microneurosurgical techniques, (including bipolar coagulation). Even tumours anterior to the cord can be removed without cord damage by the use of proper techniques. Remarkable recovery can follow removal of such tumours, and surgery should always be offered, even to patients with a long history of severe cord compression.

Intramedullary tumours are rare: microsurgical excision of ependymomas is being undertaken increasingly often with minimal cord damage, but radiotherapy for such tumours is also particularly effective (Fig. 9). Astrocytomas, often cystic, are less easily removed in adults but may be removed in children. These tumours grow slowly, and the choice of treatment must depend on factors such as their site and extent, the degree of cord function, and, not least, the patient's wishes.

Cauda equina tumours

Causes

Neurofibromas and ependymomas (arising from the conus medullaris or the filum terminale) are the usual tumours (Fig. 7 and Fig. 8). Metastatic tumours are common and are generally excluded from this group, as local pain and plain radiographs of the lumbar spine quickly establish the diagnosis.

Features

Cauda equina tumours present in different ways to different specialists. The characteristic pain is sciatica (or root pain), which wakes the patient at night, forcing them

to pace about or sit in a chair. The presence of progressive nocturnal pain should suggest a tumour rather than the usual disc protrusion. Painless weakness and wasting of the legs may occur, especially if the tumour is anterior to the conus. Sensory ataxia is caused by severe loss of the sense of joint position. If sphincter function is disturbed the bladder is usually affected before the bowels. Subarachnoid haemorrhage causing sudden unusual back pain, then headache and vomiting, is due to the propensity of ependymomas to bleed. The cause of papilloedema is debatable: communicating hydrocephalus following subarachnoid haemorrhage is a possibility, but is more likely to be due to a disturbance of retinal drainage associated with high concentrations of protein in the cerebrospinal fluid.

P>**Treatment**

The main difficulty lies in making the clinical diagnosis. Myelography confirms the diagnosis and microneurosurgical excision produces excellent results, with relief of excruciating pain.

Spondylosis and disc prolapse

Introduction

'Osteoarthritis' denotes degenerative arthropathy of a synovial joint, but as the only synovial joints of the spine are the apophyseal or facet joints, the term spondylosis is used generally to denote the degenerative changes seen between the vertebral bodies. Such degeneration is caused by the decrease in water content of the discs that is seen after the age of about 20 years. The nucleus pulposus no longer acts as a ‘ball bearing' allowing vertebrae to rotate on each other (Fig. 10): the disc space narrows and sliding motion occurs between the vertebrae. Osteophytes or bony spurs develop in response to this abnormal movement, seemingly to reduce it: if this movement is prevented then the osteophytes resorb. The presence of osteophytes, in conjunction with the normal movement of the spine, may produce intermittent trauma to either the nerve root (spondylotic radiculopathy) or the spinal cord (spondylotic myelopathy). A disc protrusion occurs when the nucleus pulposus herniates through a tear in the annulus: the spinal level and the position of the extrusion determine whether cord or root compression, or both, occur.

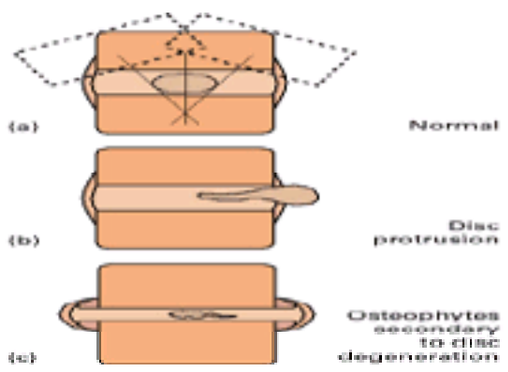


Fig. 10. Diagram to show (a) normal disc space with normal rotatory movement of one vertebra upon the adjacent one; (b) a disc protrusion;(c) osteophytes developing secondarily to disc degeneration (or disc protrusion). Note the different origin of the disc protrusion and osteophytes.

Spondylosis and disc prolapse commonly occur at the most mobile section of the spine (cervical) or the point of maximum weight bearing (low lumbar region). Cervical spondylosis is more common than cervical disc prolapse, whereas lumbar disc prolapse more commonly produces symptoms of nerve-root compression than of lumbar spondylosis.

Cervical spondylosis

Biomechanics

The cervical spine is a remarkably mobile structure with a normal range of movement between flexion and extension of 100° to 110°. The posterior contour of the cervical spinal canal changes by 5 cm between these extremes, and the cervical dura, cord, and nerve roots need to adapt to these changes.

This is mainly achieved by folding or stretching like an accordion, but also by moving up and down from and into the thoracic spinal canal (Fig. 11).

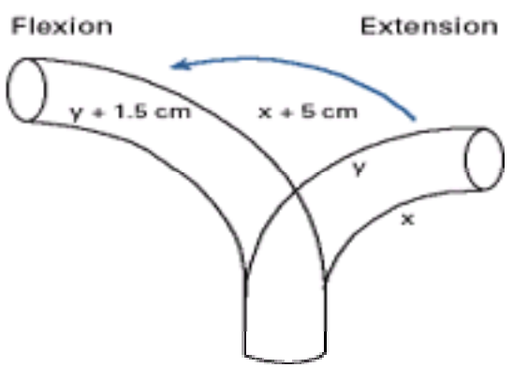


Fig. 11. Diagram illustrating the significant elongation of the cervical spinal canal during flexion. The dura, cord, and nerve roots need to adapt to this movement and elongation, imposing important physiological requirements on these structures. This movement is an important basis for the development of cervical spondylotic myelopathy and radiculopathy.

Extremes of cervical flexion cause the conus to move cranially, which exacerbates sciatica, and the brainstem to move slightly caudally, while the cervical dura becomes taut. Limitation of neck flexion therefore occurs in patients with meningeal irritation. Extension produces a slack, thicker spinal cord and nerve roots, a narrower spinal canal, due to bulging forwards of the ligmenta flava, and narrower intervertebral foramina. The degree of narrowing depends on the individual size of the spinal canal, the size of any encroaching osteophytes, and the degree of extension and range of movement of the cervical spine. Thus the pathogenesis of cervical spondylosis lies in a complex of factors, but it is basically due to the presence of osteophytes and spinal movement.

Clinical features

Cervical spondylotic myelopathy

The typical patient is over 50 years of age, and often nearer 60 or 70. The history is usually one of 1 to 2 years of progressive lack of use of the hands, with either numbness or tingling extending up the forearm in a glove distribution. The legs feel stiff and often scuff the ground, and there is often a sense of constriction around the knee or trunk. Neck pain is rare. The symptoms may worsen steadily or intermittently, or may become static as spondylosis gradually immobilizes the neck.

The signs are those of a spastic tetraparesis with sensory signs in the legs and superimposed lower motor-neurone signs in the arms, reflecting diffuse damage to the cervical cord. Careful neurological testing, especially of the arm reflexes, will allow correlation with the radiographic investigations. The bladder becomes affected late in the course of the disease.

Differential diagnosis Causes of cord compression, especially in the region of the foramen magnum, need to be considered. Instability of the cervical spine and

subsequent damage to the cord also occur in patients with rheumatoid arthritis, ankylosing spondylitis, or Down's syndrome. Motor-neurone disease is usually obvious.

Cervical spondylotic radiculopathy

This typically presents as a stiff neck, either on waking or after a minor twist of the neck. Pain radiates down the arm, especially during extension of the neck, producing the characteristic symptoms and signs of a root lesion.

The C6 or C7 roots are commonly affected, since these pass through the most mobile joints of the spinal column (C5/6 and C6/7). Weakness, tingling, and numbness occur in the appropriate root distribution, and testing the arm reflexes usually allows accurate localization. Ninety per cent of patients recover within 5 or 6 weeks, but episodes commonly recur.

Differential diagnosis Other causes of nerve-root compression need to be considered, especially if the symptoms do not subside. A tumour at the foramen magnum may inexplicably cause wasting of the small hand muscles or tingling fingers. A cervical rib (fibrous band) causes a TI syndrome with numbness in the forearm, as does an apical lung cancer (Pancoast tumour). Ulnar neuropathy is usually characteristic. Carpal-tunnel syndrome causes tingling in the thumb and fingers, except for the little finger. The aching that extends up the arm with this syndrome may be confusing, but the symptoms typically waken the patient at night or appear during gripping (e.g. driving, holding a newspaper, knitting),

Investigations

Since radiographic spondylosis is common and often asymptomatic (Fig. 1), no investigation can substitute for a careful history and examination. Electromyograms are unreliable, since irrelevant or unimportant abnormalities are often shown.

Plain lateral radiographs, taken flexion and extension, are important in excluding other diseases as well as allowing assessment of canal size, the range of movement of the cervical spine and intervertebral joints, and any instability (Fig. 12). Confirmation of cord or root compression is best obtained by MRI (Fig. 13).

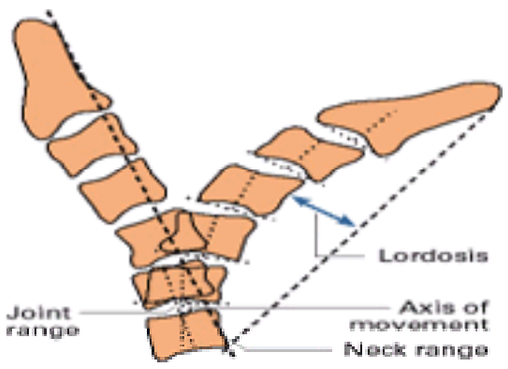


Fig. 12. Diagram to show the measurements obtainable from superimposing flexion and extension lateral radiographs of the cervical spine.

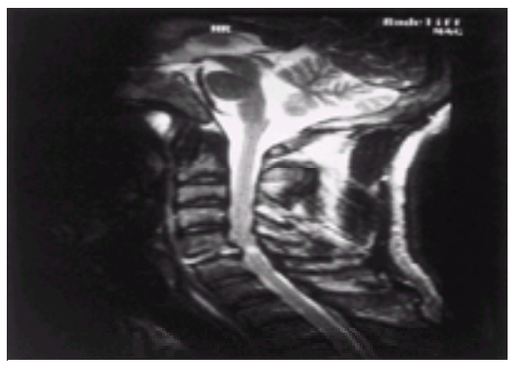


Fig. 13. Cervical spondylotic myelopathy: lateral (sagittal) view to show the area of the cord damage (white). The MRI scan shows us, for the first time, the cord damage and the diagnosis is no longer one of exclusion.

Treatment

Cervical spondylotic myelopathy is often helped by wearing a collar to restrict movement. Severely affected or rapidly worsening patients require surgery: the choice of operation should be tailored to the biomechanics encountered in the individual patient. There is evidence to suggest that a ‘decompressive laminectomy' performed in a mobile neck may itself cause later neurological damage; the dura (or cord if the dura is left opened) becomes adherent to the cervical muscles and the normal physiological movement of the dura and cord is disrupted, producing abnormal traction forces on the cord.

Laminoplasty to enlarge the spinal canal should be performed only on a patient with a diffusely narrow canal that is relatively immobile (less than 30° of head and neck movement between flexion and extension). More localized spondylosis is best treated by anterior cervical fusion with microsurgical removal of the intervertebral disc and replacement with an iliac bone graft. Many surgeons attempt to drill away osteophytes but this is probably unnecessary and may be dangerous: osteophytes are resorbed if the movement ceases. If indicated, laminectomy may follow anterior fusion, the fusion having produced a relatively immobile neck.

The results of such operations are difficult to evaluate. In general, about 60 per cent of patients show neurological improvement, and the remainder cease to deteriorate. Total internal fixation of the head and neck, producing total immobilization, creates remarkable neurological improvement in elderly and bedridden patients, emphasizing the importance of movement in the pathogenesis of this condition and the importance of eradicating this movement in its successful treatment.

Cervical spondylotic radiculopathy usually subsides spontaneously, but in patients suffering severe or frequent episodes of pain, anterior cervical fusion may be performed with or without a bone graft, depending on the instability or otherwise of the joint. The results of fusion without removal of osteophytes are excellent in properly selected patients, over 95 per cent of whom are relieved of their root pain.

Acute cervical disc protrusion

Acute cervical disc protrusion may occur spontaneously after a minor twisting injury or following a major flexion injury, such as diving into a shallow pool (Fig. 14); this usually causes a radiculopathy that is often indistinguishable clinically and radiologically from osteophytic compression. Severe root pain is likely to be due to a disc protrusion, and this may coexist with an osteophyte. The treatment is the same, and the surgeon often sees a vertical rent in the posterior longitudinal ligament during microsurgical discectomy. Gentle exploration with a blunt hook allows the disc material to be teased out, using the operating microscope



Fig. 14. MRI: lateral view of the cervical spine to show an acute disc protrusion from the C6/7 disc space, herniating upwards behind the body of the C6 vertebra. This protrusion was removed by the interior approach using an operating microscope but without the need to drill away the vertebral body.

More centrally placed disc protrusions cause compression of the spinal cord: a ‘partial Brown–Sequard syndrome' may be seen if the protrusion is paracentral. Surgery via the anterior route allows the disc to be removed without difficulty. The results of surgery for cervical disc protrusions are, in general, excellent.

Lumbar disc protrusion

Lumbar disc protrusions (Fig. 15) are common, yet the results of surgery leave some patients dissatisfied. Removal of a lumbar disc protrusion relieves over 95 per cent of patients of their sciatica, but such results are only obtained if a definite disc protrusion compressing the nerve root is found and removed. Selection of appropriate patients is therefore the single, most important factor determining the result. Bulging discs are normal and found in asymptomatic patients (Fig. 2).

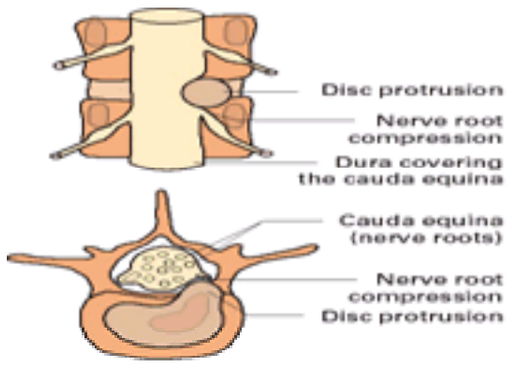


Fig. 15. Diagram to show the relations of a lumbar disc prolapse.

Clinical features

The usual history is that the patient bends and twists, and then suddenly feels their back ‘go'. They are locked in a flexed position with the disc prolapse compressing the pain-sensitive posterior longitudinal ligament. After a few hours or days, further extension of the prolapse causes nerve-root compression and the patient develops pain down the posterior or lateral aspect of the leg, radiating below the knee into the calf, ankle, or foot. Pins and needles and numbness develop in the appropriate dermatome. Physical signs comprise those of nerve-root damage, either L5 or S1 (common levels for lumbar disc prolapse being L4/5 and L5/S1), limitation of straight-leg raising on the ipsilateral side and pain on the affected side when raising the contralateral leg, and local tenderness and stiffness of the back, with scoliosis due to muscle spasm.

The most important signs indicating a lumbar disc protrusion are pain on raising the contralateral leg, calf wasting (measured with a tape measure), impairment or absence of the ankle reflex, weakness of dorsiflexion of the foot, and scoliosis of the lumbar spine.

Management

The vast majority of patients recover with time; if this is to occur, it is obvious by 5 or 6 weeks from the onset. For a patient seen at this time with a history of sciatica (rather than back pain) and with definite signs, it is reasonable to advise further investigation.

MRI (Fig. 16) has now replaced myelography as the investigation of choice (Fig. 4). There is no place for discography in the investigation of sciatica. CT may be used but it is undoubtedly inferior to MRI.

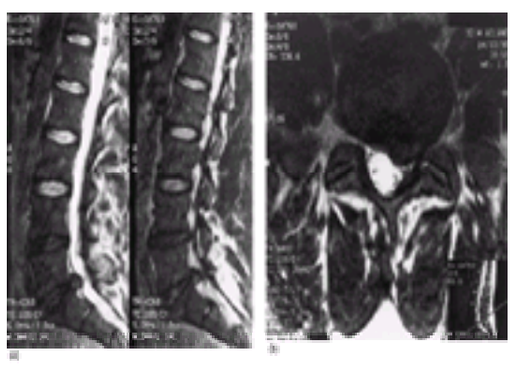


Fig. 16. MRI: sagittal (a) and axial (b) views to show a lumbar disc prolapse.

Surgical treatment

Surgery should be advised if the patient can no longer cope with sciatic pain and if there is definite clinical and radiographic evidence of a disc protrusion compressing the nerve root. The usual indication is therefore pain, but occasionally a painless foot drop in a young, active person may best be treated by removal of the prolapse, allowing the nerve root to recover. Back pain without significant sciatica should not, in general, be treated by removal of a disc protrusion, although a central disc protrusion (rather than a bulge) may occasionally cause back pain without sciatica. There is usually a history of trauma, and the patient finds sitting particularly difficult.

A central disc protrusion may also cause compression of the cauda equina, which is often associated with bilateral sciatica, perineal (pericoccygeal) numbness and tingling, and retention of urine. This is a surgical emergency requiring urgent myelography and removal of the disc prolapse; bladder, bowel, and genital function will only recover following prompt and gentle surgical intervention.

The most important factor determining a good result is patient selection. Some surgeons advocate microsurgical removal, but this has no advantage, and the results are probably less good because adequate decompression of the nerve root is less likely with this restricted approach. There is also a danger of operating at the wrong

level.

Insertion of an epidural catheter during surgery allows postoperative epidural diamorphine to be given. Encouraging patients to take adequate analgesia and keep moving allows mobilization within 48 h and discharge, on average, 3 or 4 days postoperatively. The patient is by then taking oral analgesics. Patients are instructed in back extension exercises: it is essential to re-educate the sacrospinalis muscles, which are often weak, particularly if the patient has been resting in bed or using a corset.

Epidural fat grafts are pointless; every patient develops epidural scarring and this does not cause pain. Fusing the back in conjunction with removing a disc is wholly unnecessary. Percutaneous disc aspiration is inappropriate treatment for patients with a disc prolapse compressing a nerve; this is usually performed on patients with ‘bulging discs' who should not be offered surgery anyway.

Complications

The most common complication is probably failure to cure the patient. If no disc protrusion compressing the nerve root is found, the patient is likely to continue to have pain. Unfortunately, a failed operation leaves the patient worse off, and they enter a group referred to as having a ‘failed back'. The failure in fact is that of the surgeon, who was insufficiently careful about selecting the patient. Six weeks' relief of pain is common, even after inappropriate surgery: this appears to be due to a placebo effect. ‘Facet injections' for back pain produce a similar pain-free period.

Disc prolapse may recur. The patient has clear-cut relief of sciatica (for longer than 6 weeks), followed by pain described as ‘exactly the same as before'. Proof of prolapse is not always clear on reinvestigation, due to postoperative scarring, and it may be necessary to explore the disc space.

Discitis may occur some days or weeks postoperatively; this appears to be a low-grade infection in the disc space producing back pain without sciatica. Radiographs or CT show erosion of the cortical margins and the erythrocyte sedimentation rate is elevated. Bed rest is necessary, but the condition usually settles, often with spontaneous bony fusion of the disc space and excellent pain relief.

Peroperative nerve-root damage should be very rare: the usual cause is inadequate exposure of the root, especially if it is tucked under the facet joint and is compressed by a lateral disc prolapse. Failure to expose and decompress the nerve root adequately leads to excessive root retraction.

A central disc prolapse also requires generous bony decompression before removal; attempts to remove a central protrusion through a fenestration (by just removing the ligamentum flavum) lead to severe and irreversible damage to the cauda equina, with disastrous consequences for the patient.

Lumbar spondylosis

Lumbar spondylosis and facet-joint degenerative hypertrophy may cause symptoms of nerve-root compression. Two syndromes, one well defined and the other less so, occur.

Cauda equina claudication

Cauda equina claudication is a syndrome of sciatica with or without paraesthesia, which affects the patient when standing or walking but is completely relieved by sitting or lying down. The distance that can be walked gradually lessens with time; patients can walk further if they flex slightly and can usually bicycle without pain (which obviates the ischaemic theory of this condition). The symptoms can be bilateral or unilateral, and the only sign is that one or both ankle jerks are absent, especially if tested immediately after walking. Unlike patients with vascular insufficiency of the legs, those with the cauda equina syndrome have to sit down to relieve pain; the patient with vascular insufficiency can obtain relief by standing.

The usual cause of this syndrome is stenosis of the lumbar canal due to spondylosis superimposed on a congenitally narrow canal. The ligamenta flava bulge forward in extension, hence the relative relief obtained by flexing the spine. Patients are usually over the age of 60 years; they may be imprisoned by pain, being unable to walk beyond the front door of their homes. Decompressive laminectomy produces excellent results, provided that the symptoms are exactly as described as above, the ankle jerk (or jerks) is absent, and MRI or myelography shows a complete or near-complete block (Fig. 17). These may show a near-complete block in asymptomatic individuals; the history and clinical findings are essential in making the diagnosis. A spinal angioma may mimic this syndrome and should be considered if there are doubts.

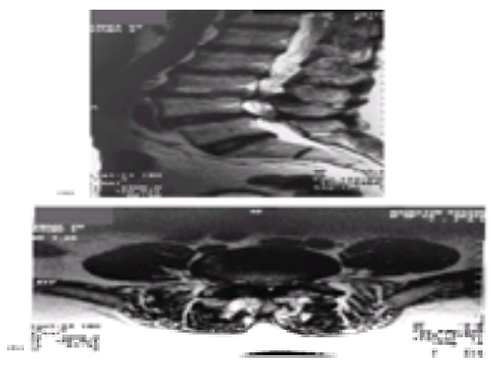


Fig. 17. MRI scan to show stenosis of the lumbar canal due to facet-joint hypertrophy and osteophyte formation: sagittal (a) and axial (b) views.

Younger patients with a disc protrusion may also present with symptoms of cauda equina claudication; it has been suggested that the adoption of a vertical posture causes further protrusion to occur, but this seems a somewhat facile explanation.

'Lateral gutter syndrome'

This is an ill-defined syndrome of intermittent sciatic pain that is worse on standing and walking, but in a less well-defined and consistent manner than is seen in patients with cauda equina claudication. Appropriate nerve-root signs are found and CT myelography reveals the root trapped in the lateral gutter; that is, between, anteriorly, an osteophyte and, posteriorly, a hypertrophic, degenerative facet joint. There is, however, no appreciable canal stenosis (Fig. 18). This is a less well-defined syndrome, and the radiographic appearances are often seen in asymptomatic individuals.

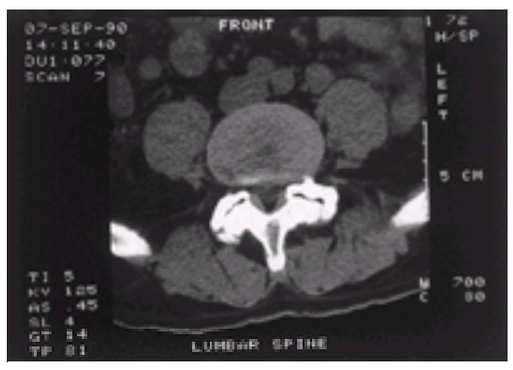


Fig. 18. CT scan through the L4/5 disc space; note the irregular facet joints particularly on the left. The L5 nerve root was trapped between this hypertrophied facet joint posteriorly and the vertebral bodies and disc space anteriorly; this produces the ‘lateral gutter syndrome'.

Decompression of the nerve root is justified in patients with sciatica (rather than back pain), with signs of nerve-root damage, and with appropriate radiographic appearances. Surgery produces relief of sciatica pain in 60 to 70 per cent of patients. Such surgery does not weaken the spine and by 3 months after the patients can be encouraged to lead an entirely normal life, as long as back extension exercises are performed to counteract the mechanical abnormalities of the spine that arise from the previous disc prolapse or spondylosis.

Thoracic disc prolapse

This is rare, but important. The patient presents with progressive compression of the spinal cord, and often has a history of previous trauma or a fall from a height. Back pain may predate the onset of cord or root features. Thoracic disc protrusions are unlike protrusions elsewhere: they are often calcified and may erode through the dura. Very occasionally they are liquefied. The tumour-like, progressive clinical picture may be a result of the immobility of the thoracic spine, due to its splintage by the ribs.

A thoracic disc prolapse is best demonstrated by CT myelography (Fig. 19 and Fig. 20). Surgical treatment is required, since once neurological symptoms or signs develop, progressive deterioration follows. The surgical approach to a central protrusion should be transthoracic. If the protrusion is laterally placed, a posterolateral approach with removal of the transverse process and pedicle is satisfactory. A laminectomy should never be performed, as this will almost certainly cause cord damage. Whichever approach is used, great care is necessary to identify the correct level by radiographic marking of the rib or transverse process. The use of microsurgical techniques allows cord damage to be avoided.

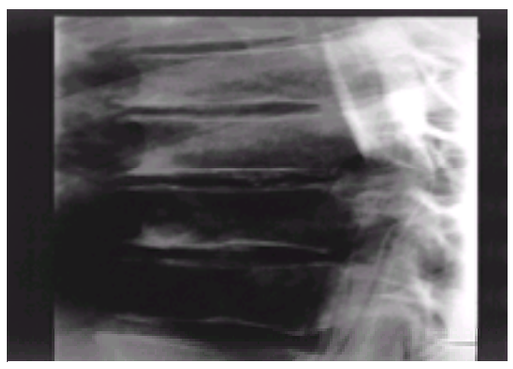


Fig. 19. Myelogram showing posterior displacement of the cord by a calcified, central disc protrusion arising from the mushrooming out of the thoracic disc space. The postmyelogram CT scan of this same patient shows the abnormality more clearly and is the investigation of choice.



Fig. 20. CT scan showing a calcified, central thoracic disc protrusion compressing the theca (containing contrast) and the spinal cord (shown as a thin black crescent adjacent to the calcified disc).

Sciatica and back pain

The surgical relief of nerve-root compression is beneficial to the patient with incapacitating sciatica. Back pain is not usually so amenable to surgery. Pain down the front of the thigh indicates a high lumbar root lesion (L1, 2, 3) and is not due to sciatic nerve or root compression (L4, 5, and S1). Referred pain from the back may extend down the back of the thigh to the knee, but pain extending beyond the knee, especially when associated with paraesthesiae, is due to nerve-root compression.

There are three clinically well-defined ‘sciatica syndromes' (Table 3) that should be considered while taking a history. Myelography (taken to the midthoracic area) is still a good method of investigating sciatica but MRI is preferable.

Pathology	*Syndrome*
1. Disc prolapse	Sudden severe persistent sciatic pain (with paraesthesiae)
2. Cauda equina claudication	Sciatica on standing or walking relieved completely by sitting or lying down
3. Cauda equina tumour	Sciatica worse at night forcing the patient out of bed, either to pace around or spend the rest of the night sitting in a chair
4. Other causes	
(a) Lateral gutter syndrome	Ill-defined, but some worsening pain on standing and walking
(b) Spinal angioma	Cauda equina claudication that's 'not quite right'
(c) Metastatic tumour	Known primary; obvious on radiographs
(d) Pelvic tumour	Often missed; requires pelvic CT scan

Table 3 Causes of sciatica

Back pain is not amenable to surgical treatment except when it is (rarely) caused by central lumbar disc protrusion, or when it occurs in association with spondylolisthesis (congenital as opposed to degenerative). This may cause severe low back pain in adolescents, and may require lumbar fusion. Back pain due to lumbar spondylosis is not generally relieved by fusion. The results are uncertain, but back pain that is completely relieved by bed rest might benefit from such treatment. The problem is twofold: first, the origin of degenerative back pain is obscure (rigid anterior and posterior fusion often fails to relieve the pain); secondly, radiography and MRI invariably show disc degeneration after the age of 40 years. The surgeon faced with a patient complaining of back pain may have to operate on radiographic appearances rather than on any clinical syndrome—a very unsatisfactory basis indeed. It is in the patient's best interest if the surgeon does not advise surgery for degenerative back pain.

Provocation discography, which involves injecting contrast medium into the disc space in an attempt to reproduce the pain, assumes incorrectly that back pain arises from the disc space. Clearly it does not, and this investigation should not be undertaken.

Postoperative pain

The causes of postoperative pain are outlined in Table 4. The vast majority of patients with pain following back surgery should never have been operated on in the first place. Patients with such a ‘failed back' invariably suffer worsening symptoms and become crippled. Their lives (and those of their relatives) become totally dominated by the back pain, and compensation claims often exacerbate the situation. Unrealistic expectations of surgery must be corrected preoperatively. Although the patients' psychological characteristics must be considered, it is equally important to realize that long-standing pain, often causing sleep deprivation, is often associated with depression. There is a tendency to label patients with cauda equina tumours as hysterical or malingering, because their pain seems out of all proportion to their physical signs: it is (Fig. 21).

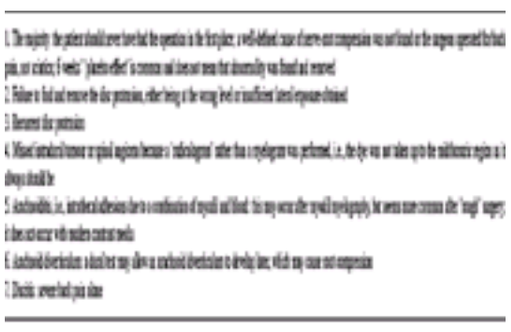


Table 4 Causes of ‘the failed back'

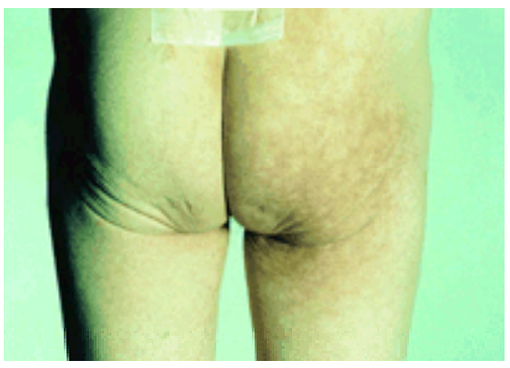


Fig. 21. Severe, genuine back pain is shown by pigmentation (erythema ab igne) secondary to the application of repeated hot-water bottles to the painful area.

The patient's reaction to ‘back trouble' may be analysed into three components. The first is the organic disease, perhaps the disc prolapse compressing the nerve. The second is the neurotic reaction, such as depression or anxiety. A certain amount is normal but in some patients this achieves undue proportions and, however effective the surgery to the nerve root, the anxiety or depression may, if excessive, mar the result. The third reaction is that of ‘illness behaviour'. This means abnormal behaviour in reaction to the disease, more explicitly a desire not to get better, because of either domestic or financial gain; the patient gains either domestic attention that otherwise would not be available (pain becomes an ‘old friend') or gains financially. In the latter the patient might act quite normally at home and reserve their symptoms for the assessing doctors. Illness behaviour in these circumstances has now become frank malingering.

The features of patients who maintain an illness behaviour need to be understood. They are remarkably well defined: pain is constant, present at all times, and spreads to other parts of the body. The pain gradually worsens, especially after a therapeutic intervention. Examination should start before the patient is seen. Do they sit easily in the waiting room? How do they walk in? Do they sit without grimacing? Do they move like a patient in pain? Can they undress and get on and off the couch? A patient may produce a dramatic entry, and show great difficulty undressing or getting on and off the couch. However, while getting on the couch the patient may inadvertently sit at a right angle, demonstrating full straight-leg raising, although this is inevitably much reduced on formal testing. To any experienced clinician it is obvious that the patient does not behave like one with real pain. There is usually a marked lack of effort on testing motor power. Sensory loss is in a non-organic distribution, and usually extends up to the knee or thigh in a stocking distribution. Yet there are no definite neurological signs. The back is tender to the slightest touch and when the patient is asked to sit up to allow its examination, this is (usually) done easily with relative enthusiasm to show the doctor the bad back, again demonstrating full straight-leg raising. Pressure on the top of the head often produces a complaint of low back pain (which is not possible in mechanical terms).

If frank malingering is suspected, it is often rewarding to observe the patient leaving the hospital.

Clearly, surgery in these patients is doomed to fail, but it is surprising how many have been previously offered spinal fusion, suggesting illness behaviour is not as widely recognized as it should be.

Further reading

Box CR. A case of invasion of the cauda equina by tumour with demarcation of all the sensory root areas of the lower limbs. *Lancet* 1903; **ii**: 1566–72.

Feamside MR, Adams CBT. Tumours of the cauda equina. *Journal of Neurology, Neurosurgery, and Psychiatry* 1978; **41**: 24–31.

Herkowitz HN *et al.*, ed. *Rothman–Simeone, the spine*. Saunders, Philadelphia, 1982.

Jayson MIV, ed. *The lumbar spine and backpain*, 3rd edn. Churchill Livingstone, London, 1987.

Kerr RSC, Cadoux-Hudson TA, Adams CBT. The value of accurate clinical assessment in the surgical management of lumbar disc protrusion. *Journal of Neurology, Neurosurgery, and Psychiatry* 1988; **51:** 169–73.

Tuloy E, D'Costa D. *Back attack*. Little Brown, Boston, 1985.

Vinken PI, Bruyn GW, ed. *Handbook of clinical neurology*, Vol. 26, Part 11, *Injuries of spine and spinal cord*. Chapter 9 Cervical spondylotic radiculopathy and myelopathy. American Elsevier, New York, 1971: 97–112.

44.11 Neurosurgery for intractable pain

Peter J. Teddy

[Introduction](#)
[Operations for interrupting nociceptive pathways](#)
[Operations upon peripheral nerves, dorsal roots, and ganglia](#)
[Operations on the spinal cord](#)
[Dorsal root entry zone operations](#)
[Anterolateral cordotomy](#)
[Commissural myelotomy](#)
[Sympathectomy](#)
[Operations on cranial nerves](#)
[Operations on the brain](#)
[Electrical stimulation of peripheral nerves, spinal cord, and brain](#)
[Spinal cord stimulation](#)
[Deep brain stimulation](#)
[Drug delivery to the spinal fluid](#)
[Further reading](#)

Introduction

The best way of managing severe pain is to identify and eliminate the causal agent. This objective is not always possible; the cause of the pain may not be identifiable or it may not be amenable to treatment. Intractable pain has then to be treated symptomatically.

Optimal surgical treatment for intractable pain should be confined in its effect to the painful area, be simple and inexpensive to perform, be associated with a low mortality and morbidity, and carry a low risk of neurological deficit. It should also be effective and long lasting. Most forms of neurosurgical intervention for pain relief involve creating a lesion within the nervous system, with inevitable risk to neurological function. For this reason neurosurgeons are often called in much too late in the course of treatment of patients with intractable pain.

There are three principal neurosurgical techniques used in the treatment of pain.

1. Operations that interrupt nociceptive pathways by creating lesions in peripheral nerves, nerve roots or ganglia, the spinal cord, various parts of the brain and brainstem, and the sympathetic nervous system.
2. Electrical stimulation of pain suppressive systems or blocking pain pathways (peripheral nerve, spinal cord, or brain).
3. Administration of various drugs to the intraspinal or intraventricular compartments of the cerebrospinal fluid pathways.

The choice of treatment will be dictated by the most appropriate therapy for the particular pain syndrome, whether the pain is of benign or malignant aetiology, the age and general medical condition of the patient, life expectancy, the mental state and, to some extent, the intelligence of the patient if some form of patient-operated implant is being considered. The balance between the cost of surgical and conservative forms of treatment may also be a deciding factor.

Only a few of the more common forms of intractable pain (Table 1) are discussed here. The treatment of trigeminal neuralgia is reviewed elsewhere.

Benign
Amputation stump and phantom limb pains
Arachnoiditis
Brachial and lumbosacral plexus avulsion injuries
Causalgia (from peripheral nerve damage)
Pain associated with paraplegia (spinal cord injury)
Post-herpetic neuralgia
Post-thoracotomy syndrome
Spastic syndromes with pain
Spinal surgery pain ("the failed back")
Syringomyelia
Thalamic pain (usually from stroke)
Trigeminal neuralgia
Malignant
Bony metastases
Pancoast syndrome
Radiation induced damage to brachial and lumbosacral plexus
Sacrococcygeal and pelvic pains (carcinoma of the prostate and anorectal tumours)
Generalized lower limb pain: flow in malignant disease

Table 1 The more common forms of intractable pain

One should start with the simplest and most peripheral procedure and progress centrally, since more central operations may be associated with serious complications or side-effects. Other procedures such as spinal cord stimulation may be innocuous but the cost involved may exclude patients with pain due to malignancy and a limited life expectancy.

A careful history and examination must first be undertaken to establish whether there is an identifiable and treatable cause for the pain. The exact area affected must be determined, along with the type of pain (burning, stabbing, nagging, tingling) and any altered sensation in the affected area so as to establish whether the pain is somatic or neurogenic. The former is probably due to chronic activation of nociceptors responsible for acute pain, while the latter (central, deafferentation, or dysaesthetic pain) is probably a result of changes within some part of the transmission system. Deafferentation pain is more difficult to treat and is frequently associated with neurological injury. Those operations which seek to relieve pain by causing numbness in the affected area are generally inappropriate if that area is already numb.

Operations for interrupting nociceptive pathways

The surgeon may interrupt or alter the transmission of painful stimuli anywhere along the pathways shown in classical form in Fig. 1.

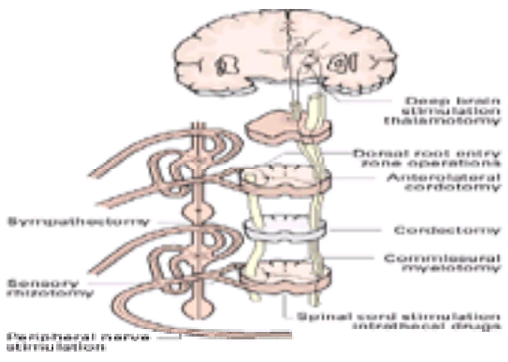


Fig. 1. Pathways for the transmission of painful stimuli and sites of surgical pain-relieving procedures.

Operations upon peripheral nerves, dorsal roots, and ganglia

Among the more successful surgical procedures performed on peripheral nerves are those used in treating pain caused by compression, such as carpal tunnel syndrome and ulnar nerve entrapment at the elbow (except, perhaps, in its chronic form). The dorsal root or dorsal root ganglia may be approached by an open operation (usually intradural) or percutaneously using a radiofrequency generator and suitable electrode. These procedures are most commonly used in the treatment of pain in the trunk (e.g. post-thoracotomy syndrome). The results are variable, and although pain relief is often achieved there may be a high recurrence rate and unpleasant side-effects such as post-rhizotomy dysaesthesia. To ensure effective analgesia in any given dermatome, the affected root is cut together with two roots above and two below. This produces a widespread total deafferentation, and the method is therefore inappropriate for limb pains.

Selective posterior rhizotomy (rhizidiotomy) makes use of the fact that at the spinal cord/rootlet junction (around 1 mm from the cord) large and small fibres are dissociated anatomically. The small fibres are found on the undersurface of the rootlet and then regroup in the ventrolateral part of the cord to enter the tract of Lissauer. It is possible to cut just the small-diameter nociceptive fibres at their entrance to the spinal cord, activating pain inhibitory mechanisms while preserving propioception, thus making it suitable for treating pain syndromes affecting the limbs.

Operations on the spinal cord

Dorsal root entry zone operations

Following laminectomy or hemilaminectomy with exposure of the spinal cord, small lesions are produced with radiofrequency generators or lasers at the point of entry of the dorsal roots into the spinal cord within the intermediolateral sulcus. The operation was designed to treat central pain thought to be due to hyperactivity of cells within the dorsal horn which relay via the spinoreticular and spinothalamic tracts to supraspinal structures. Dorsal root entry zone lesions destroy either the hyperactive neurones or the cells of origin of the spinothalamic and spinoreticular tracts.

Dorsal root entry zone operations were first used for the treatment of brachial plexus avulsion injuries, and produced long-term pain relief in around 70 per cent of patients treated. They have also been used less successfully in the treatment of the central pain of paraplegia, post-herpetic neuralgia, atypical facial pain, and pain associated with failed back syndrome and arachnoiditis. Although few serious postoperative side-effects are reported, there is probably a morbidity rate of at least 10 per cent.

Anterolateral cordotomy

The main pain-carrying tract in humans is traditionally considered to be the anterolateral spinothalamic pathway; however, only a small number of these fibres reach the thalamus without synapse. Spinothalamic tractotomy (anterolateral cordotomy) has been practised to good effect since the beginning of this century. It may be performed as an open or percutaneous procedure on the opposite side to the painful area and at least four vertebral segments of the cord above the affected dermatomes. The axons of the spinothalamic neurones cross the midline at the levels of the anterior commissures over several segments. Many surgeons prefer to perform all such operations in the high cervical region, where the crossover is certain to have been completed.

The operation should produce relief of pain and loss of temperature sensation in the affected area, with good analgesia. It is most appropriate for unilateral malignant pain and for treating superficial or deep somatic pains rather than visceral pain. Bilateral operations increase the risks of neurological deficits. Provided the patient is fit for surgery and has a life expectancy of more than a few months, this is a good method of relieving pain which is often overlooked in treatment of the terminally ill. Operative mortality in the latter group of patients is high (about 10 per cent) and the principal risks are those of respiratory failure, motor weakness, dysaesthesia, and sphincter dysfunction.

Commissural myelotomy

Commissural myelotomy is an alternative to bilateral anterolateral cordotomy and is appropriate for the treatment of malignant pain in both lower limbs or in the perineal region. Pain and temperature fibres decussate in the spinal cord whereas light touch and motor fibres decussate in the lower brainstem: a midline myelotomy which divides the commissures at an appropriate level can be used to treat bilateral pains of various types. An extensive laminectomy is necessary at open operation. Central myelotomy using stereotaxic guidance requires detailed knowledge of stereotactic techniques.

Cordectomy is rarely effective or appropriate even in patients with pre-existing complete spinal cord damage.

Sympathectomy

Sympathetic denervation has been used in the treatment of limb, cardiac, and abdominal visceral pain. Open sympathectomy has largely been replaced by the use of oral or intravenous sympatholytic drugs, percutaneous radiofrequency lesions, and chemical procedures. The most common indications are those of causalgia, Sudeck's atrophy, and ischaemic pain in the limbs due to peripheral vascular disease. Cardiac sympathectomy in the treatment of angina has largely been superseded by improved medication and coronary artery bypass grafts, although percutaneous radiofrequency cardiac sympathectomy may still have a place in the treatment of medically intractable pain. Malignant pain affecting the pancreas, liver, gallbladder, and stomach together with painful chronic relapsing pancreatitis has been treated by chemical sympathectomy of the coeliac plexus and splanchnic nerves using an injection of 50 per cent alcohol or phenol.

Operations on cranial nerves

The treatment of trigeminal neuralgia is discussed elsewhere. Dorsal root entry zone lesions made in the spinal nucleus of the fifth nerve and Lissauer's tract extending from the C2 dorsal root to the level of the obex may be one of the few successful treatments for chronic facial post-herpetic neuralgia.

Operations on the brain

Stereotaxic thalamotomy employs a stereotaxic frame applied to the head, and with the guidance of MRI or CT scanning, suitable targets for a radiofrequency-generated burn or for the tip of a stimulating electrode can be readily identified. Those targets include the centromedian nucleus and more rostral intralaminar nuclei of the thalamus and the parafascicular nuclei.

Such operations may be expensive (in the case of neurostimulatory techniques), have potentially serious side-effects such as hemiplegia or speech deficits, and are somewhat unpredictable in producing pain relief. Thalamotomy is most appropriate for the relief of pain due to malignancy in the head and neck.

Injection of alcohol into the intrasellar region has been used to destroy the pituitary gland in patients with widespread pain from advanced carcinoma of the breast and prostate. Between 50 and 70 per cent of patients may gain complete pain relief following this procedure but endocrine disorders inevitably ensue.

Focal stereotaxic lesions within the frontal lobes and their projections which have found more general acceptance include cingulotomy, subcaudate tractotomy, and inferior frontomedial leucotomy. Their use is controversial owing to possible alterations of the psyche.

Electrical stimulation of peripheral nerves, spinal cord, and brain

Modern devices for stimulating the central nervous system to produce pain relief comprise an electrode or set of electrodes applied to either a peripheral nerve, the dura overlying the back of the spinal cord, or a chosen target within the brain. The electrodes are linked to a receiver system implanted under the skin and activated by an external radiotransmitter, or are connected directly to a totally programmable implanted activating device. Peripheral nerve and spinal cord stimulators are inserted by a small operation, and are associated with few serious complications. The stereotaxic implantation of a deep brain-stimulating system is complex and more prone to unwanted sequelae.

The likely benefits are commonly assessed by temporary stimulation prior to implantation. One of the chief problems with this form of treatment is the expense involved; in addition the life expectancy of some systems may not exceed 5 years. Peripheral nerve stimulators have been used for pain following peripheral nerve damage in the limbs, and they act either by blocking nociceptive afferents or by more central inhibition of spinothalamic tract neurones.

Spinal cord stimulation

Spinal cord stimulation evolved from gate control theory such that stimulation in the region of the dorsal columns would mainly inhibit C-fibres—hence the original term dorsal column stimulation. However, there are many ways, both electrical and neurochemical, by which spinal cord stimulation might produce pain relief. Electrodes are inserted into the epidural space, either percutaneously or through a small laminectomy. The latter allows more accurate placement of the electrode and it can be secured to the dura. Spinal cord stimulation is mainly used in the treatment of syndromes such as the ‘failed back'. Some believe that this is the only form of worthwhile treatment in such cases. It may prove beneficial in patients with peripheral vascular disease, by easing the pain associated with ischaemic limbs and improving peripheral circulation.

Deep brain stimulation

The anatomical and physiological postulates underlying the choice of targets within the brain to effect pain control are complex. Two main target areas have been chosen. Stimulation of lateral margin of the periaqueductal and periventricular grey matter is thought to affect a pathway running from the midbrain to the dorsal horn, inhibiting nociceptive neurones. The second target area is the VPM/VPL nuclei of the thalamus, which ultimately inhibit spinothalamic tract neurones. Electrodes, implanted into the brain at these sites using stereotaxic guidance systems, are connected subcutaneously to a suitable stimulating device. The parameters of stimulation may be altered by radiotelemetry. A newer technique which looks promising in the treatment of central pain syndromes is that of motor cortex stimulation.

Drug delivery to the spinal fluid

Opiate drugs injected into the spinal fluid produce analgesia by their direct action on the spinal cord, and opioid receptors have been identified in both the spinal cord and the brain. Chronic administration of opiates and other analgesic substances into the cerebrospinal fluid, via lumbar or intraventricular catheters, has been used to produce long-term pain relief. The doses employed are a fraction of those required when the same drug is administered orally or intravenously. Powerful opiates such as morphine may cause respiratory depression and urinary retention, and tolerance may develop. These disadvantages may be offset by the reduction in those adverse effects associated with administration of opiates by other routes.

Substances other than opiates have been administered intrathecally to relieve pain: clonidine has been one of the more successful. Most are thought to moderate nociceptive information to neurones in the dorsal horn. The drugs may be delivered by continuous infusion or in carefully controlled bolus doses through a catheter implanted into the ventricle or into the lumbar sac and connected to a reservoir and pump system implanted under the skin of the abdominal wall. This system has been most widely used to alleviate pain in patients with cancer who have a reasonable life expectancy. Pain associated with spasticity following spinal injury can effectively be treated with intrathecal baclofen. Disadvantages other than side-effects of the drugs are the cost, risks of infection, and the frequency of pump and catheter failures. These recent advances in neurostimulation and in delivery systems which allow small aliquots of drug to be administered to exact loci within the central nervous system point the way toward the development of more precise and effective methods of pain relief while reducing the incidence of neurological deficit.

Further reading

Gybels JM, Sweet WH. *Neurosurgical treatement of persistent pain*. Basel: Kayser, 1989.

Wall PD, Melzack R. *Textbook of pain*, 3rd edn. Edinburgh: Churchill Livingstone, 1994.

White JC, Sweet WH. *Pain and the neurosurgeon*. Illinois: Charles C. Thomas, 1969.

45.1 Eye emergencies

Bradford J. Shingleton and Mariana D. Mead

Eye examination
History
Examination techniques
Radiology and computed tomography (CT) of the orbit
Treatment of traumatic ocular emergencies
Chemical burns
Thermal burns
Corneoscleral rupture or laceration/intraocular foreign body
Hyphema
Orbital trauma/fractures
Lid lacerations
Corneal abrasions and foreign bodies
Treatment of non-traumatic ocular emergencies
Red eye
Conjunctivitis
Iritis
Corneal infection
Acute glaucoma
Orbital cellulitis
Acute visual loss
Further reading

Emergency eye problems are common. In the United States alone, over 1.5 million people suffer eye injuries each year and over 40 000 of these are associated with some degree of permanent visual loss. When injuries are coupled with eye infections and other non-traumatic eye emergencies, these numbers are more than quadrupled. In the acute setting, differentiation between the potentially blinding eye injury and a less severe problem must be made. All physicians and surgeons who see emergency patients must be able to identify the true ocular emergency, initiate appropriate care, and know when to obtain further consultation.

Eye examination

History

A good history is as important in evaluating eye emergencies as it is with any other medical problem. The chief complaint of each patient is explored to determine the presence of altered vision, pain, swelling, hemorrhage, or discharge. Any history of trauma is fully reviewed to establish the exact mechanism of injury. The past ocular history may also be very important, as it relates to pre-existing visual acuity loss, previous eye surgery, disease, or eye trauma.

Examination techniques (Table 1 and Table 2)

1.	Vision charts
2.	Metal shields
3.	Eye patches
4.	Tape
5.	Scissors
6.	Polyethylene sheets (Saran wrap)
7.	Sterile irrigation fluid
8.	Intravenous tubing
9.	Squeeze bottle
10.	Cotton tip applicators
11.	pH paper
12.	Fluorescein paper strips
13.	Lid retractors
14.	25-gauge needle and syringe
15.	Foreign body removal spud
16.	Focal illuminator or pen lights
17.	Cobalt blue filter
18.	Direct ophthalmoscope
19.	Schiotz tonometer
20.	Magnifying loupes or slit lamp

Table 1 Basic equipment for emergency eye evaluation

The emergency eye examination is divided into six sections: vision, lids/orbit, anterior segment, pupils and extraocular movement, posterior segment, and intraocular pressure. Since some serious eye problems are not readily apparent, a systematic progression through all six sections of the eye examination is required in every emergency evaluation.



Table 2 Basic medications for emergency use in evaluation and treatment of eye disorders

Vision

Notation of visual acuity is required for each eye separately. This is crucial for proper diagnosis, case management, and medical/legal documentation. The only exception to delaying visual acuity assessment is in the case of a chemical burn when vision is tested after completion of emergency eye irrigation. Vision testing begins with asking the patient to read a standard eye chart. If the patient cannot read the letters or numbers on the chart, the patient is asked to count the fingers on the examiner's hand. If the patient is unable to count fingers, determine the distance at which the patient can detect hand movement. If hand movement is not appreciated, determine if the patient can perceive light. The precise level of visual acuity performance is recorded. Visual field testing by the confrontation method is appropriate for patients with altered vision, head trauma, or other neurologic complaints.

Lids/orbit

The lids are assessed for completeness of elevation and closure, marginal integrity, swelling, and color. Foreign bodies on the ocular surface can be hidden under the

inner surface of the upper lid or in the cul-de-sac formed by the reflection of upper lid and globe conjunctiva. Examiners must be able to evert the upper lids to assess for such foreign bodies. Simple eversion of the upper lid is performed by instilling a topical anesthetic and asking the patient to look down. A cotton applicator tip is placed on the orbital rim and the eye lashes are grasped with the other hand. The eyelid is flipped over the applicator stick, which serves as a fulcrum. Foreign bodies may often be found on the everted lid surface. Examination of the orbit in cases of trauma includes palpation for subcutaneous emphysema and defects in the bony orbital rim. Areas of hypesthesia are established, as well as the presence of exophthalmos or enophthalmos.

Anterior segment

The anterior segment includes the sclera/conjunctiva, cornea, anterior chamber, iris, and lens.

The sclera/conjunctiva is inspected for foreign bodies, lacerations, discharge, swelling, and vascular infection.

The cornea should be clear and lustrous. The light reflex from a hand-held flashlight should be crisp and sharp. Any alteration in the light reflex implies corneal pathology. Topical fluorescein is helpful to confirm the presence of a corneal abrasion.

The anterior chamber is the aqueous-filled space between the cornea and iris/lens diaphragm. It is inspected for the presence of blood (hyphema) or pus (hypopyon).

The iris is evaluated for alteration in shape or contour. Iris may protrude from the eye in full-thickness globe lacerations.

The lens is situated in the pupillary axis behind the iris diaphragm. It is normally clear and any opacification implies cataract formation.

Pupils and extraocular movement

The pupils should be black, round, equal, and reactive to light. Any non-black pupil implies media opacification in the anterior chamber, lens, or vitreous. Pupillary eccentricity may occur after trauma and often is associated with serious ocular damage, such as a ruptured globe. Abnormal pupillary light reactivity or size may be important clues to the presence of intracranial pathology in the setting of head trauma. Critical assessment for an afferent pupillary defect (Marcus–Gunn pupil) is important in any case of trauma or reduced vision.

Extraocular movement is checked by determining the extent of excursion in the four cardinal positions of gaze—up, down, right, and left.

Posterior segment

The posterior segment includes the vitreous, retina, and optic nerve. The direct ophthalmoscope is utilized to assess the posterior segment. Pupillary dilation facilitates examination and can best be accomplished by topical application of tropicamide 1 per cent eye drops in combination with phenylephrine 2.5 per cent eye drops. Dilation should be used with caution in patients with a history of head trauma because of the need to monitor pupillary signs. The risk of precipitating angle closure glaucoma is very small with pupillary dilation and should not prevent the use of mydriatic drops if important information can be gained by their use.

Intraocular pressure

Intraocular pressure is measured if acute glaucoma is in the differential diagnosis of the presenting complaint. The intraocular pressure should not be measured if there is a suspicion of a ruptured globe. Intraocular pressure is most accurately measured by slit lamp applanation tonometry or hand-held Schiotz tonometry. In addition, the extremely firm eye of a patient with acute glaucoma and markedly elevated intraocular pressure can readily be detected by comparing the tactile pressures of the globes. Tactile intraocular pressures are established by digitally balloting the globes through closed eyelids.

Radiology and computed tomography (CT) of the orbit

Radiographic evaluation of the orbit is helpful in suspected cases of ruptured globes, intraocular or intraorbital foreign bodies, orbital fractures, orbital hemorrhage, orbital abscess, or abnormal globe position. For fractures and radiopague foreign bodies, the preferred radiographic series includes posterioanterior and lateral views, with Caldwell, Waters, and base projections. CT scanning has superseded plain radiographic evaluation in many cases and should be utilized whenever possible. Magnetic resonance imaging (MRI) should be avoided in any case of suspected metallic intraocular foreign body.

Treatment of traumatic ocular emergencies

Chemical burns

The one ocular emergency that requires immediate treatment is a chemical burn of the eye. Immediate irrigation is the treatment of choice. Although the full effects of chemical burns may take days or weeks to become manifest, the damage itself occurs within the first minutes to hours after the chemical injury. Alkaline burns are often more dangerous than acid burns, because of more rapid penetration of ocular tissues by alkaline agents. The most effective treatment for chemical burns occurs when irrigation is initiated at the site of injury. Physicians and their entire staff should be aware of the emergency nature of a chemical eye burn and when a telephone call is received regarding such an injury, the caller should be instructed to start irrigation immediately with the nearest irrigation source available. If trained emergency personnel are not present at the site of injury, copious irrigation should be performed for 5 min and the patient then instructed to seek emergency center evaluation.

When the patient presents to the physician for treatment, instillation of topical anesthetic drops will make the patient more comfortable so that more effective irrigation can be performed. The eyelids are retracted and any particulate chemical matter is removed. A continuous flow of normal saline via an intravenous tubing set is directed across the ocular surface. Neutralizing solutions to negate the alkalinity or acidity of the offending chemical are not required. After completion of irrigation and removal of particulate matter, the ocular surface pH is checked and confirmed as normal. A topical cycloplegic agent to reduce ciliary muscle spasm and a topical antibiotic drop are instilled. This is followed by application of a sterile eye patch. Prompt referral to an ophthalmologist is recommended as patients with severe burns (Fig. 1) may require admission to hospital for medical and/or surgical management.

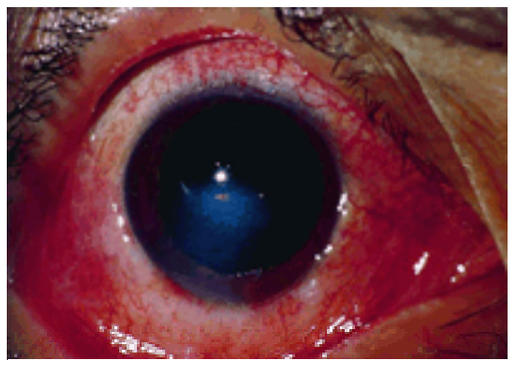


Fig. 1. Alkali chemical burn (lime) with conjunctival infection, swelling, and corneal haze. The corneal haze confirms alkali penetration into the corneal stroma and the severity of the injury.

Thermal burns

Thermal injuries commonly involve the face and ocular adnexa. Fortunately, the blink reflex and Bell's reflex, which turns the eye upward when the lids close, usually protect the globe. Occasionally a thermal corneal abrasion may occur and these injuries are best treated with a topical cycloplegic agent and topical antibiotic. Routine burn care for the skin of the lids and periorbital tissues is provided. Emphasis is directed towards minimizing scarring during the healing process so that subsequent cicatricial changes in the lid do not result in ocular exposure or misdirection of eyelashes. If lid closure is compromised on an acute or chronic basis, a moist chamber

must be provided to protect the exposed cornea. Failure to maintain hydration of the cornea in the setting of a corneal abrasion can resultin perforation within hours to days. To achieve moist chamber pro-tection, lubricate the surface of the eye with ophthalmic oint-ment or drops. If necessary, apply sterile Vaseline to the skin of the orbital rim and place a polyethylene sheet (Saran wrap) overthe orbital area. This technique will provide a moist, closed chamber for corneal protection. Skin grafting and lid tarsorrhaphy may be needed at a later date to establish proper lid position and adequate globe protection.

Corneoscleral rupture or laceration/intraocular foreign body

A full-thickness corneoscleral laceration is a surgical emergency. It is of utmost importance that the evaluating physician identify this problem and protect the eye from further damage prior to the time of operative repair.

It is not always easy to make the diagnosis of a corneoscleral laceration. Vision may be excellent with small lacerations and the external appearance of the globe may be normal. The history often provides clues to the diagnosis of an occult globe rupture or an intraocular foreign body. Suspect a ruptured globe whenever a patient has been engaging in activities involving metal-on-metal contact. A history of penetrating trauma or significant blunt trauma should raise the suspicion of globe rupture. The sclera is thinnest underneath the rectus muscles and blunt injury can cause rupture and scleral dehiscence in these areas.

Important signs may herald the presence of a ruptured globe (Fig. 2). A large subconjunctival hemorrhage with swelling may overlie a scleral rupture. Uveal prolapse, indicated by the presence of pigmented iris or ciliary body on the surface of the globe, confirms globe rupture or laceration. A traumatic hyphema should be considered an open globe until proven otherwise. An irregular or pear-shaped pupil often points in the direction of a corneoscleral laceration or rupture. Any focal opacification of the lens may represent foreign body penetration. Such focal changes may stand out as white changes against the black pupillary space when examined with a pen light or as a dark silhouette against the brighter red reflex of the fundus when the eye is examined with a direct ophthalmoscope.

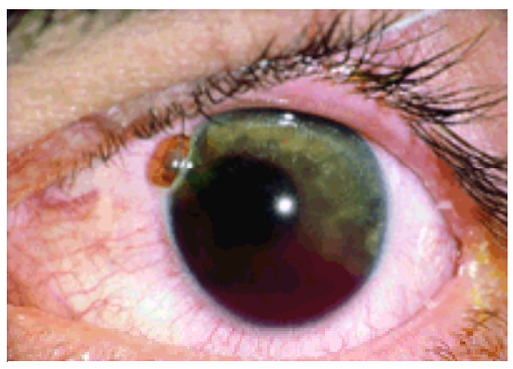


Fig. 2. A ruptured globe with uveal prolapse, hyphema, and oval pupil.

If a ruptured globe is suspected, stop the examination and place a protective shield over the eye. Do not patch the eye because the pressure of an eye patch may forcefully extrude intraocular contents. Avoid any further ocular manipulation and do not proceed with periocular surgery. Tetanus prophylaxis is indicated. An immediate referral to an ophthalmologist is indicated for urgent surgical repair. If there will be any delay in surgical repair, topical antibiotic drops (not ophthalmic ointments) may be placed in the eye. Intravenous antibiotic therapy providing Gram-positive and Gram-negative coverage is indicated. Radiographic and CT studies should be performed if intraocular foreign bodies are suspected. Microsurgical repair represents definitive treatment. Repair aims to restore normal ocular architecture and minimize scarring in the visual axis. Prompt treatment restores useful vision in many eyes.

One rare, but important, potential problem of corneoscleral lacerations is sympathetic ophthalmia. Sympathetic ophthalmia is a bilateral granulomatous uveitis that develops rarely after unilateral penetrating trauma. Patients suffering penetrating ocular trauma are carefully evaluated during the first 14 days following injury. If there is no potential for restoration of vision in the injured eye, it is often preferable to remove the injured eye in order to guard against the development of sympathetic ophthalmia in the fellow eye.

Standardized classification of ocular trauma

An internationally standardized classification of ocular trauma terminology is gaining increased worldwide acceptance. A common language will improve both clinical care and research. The entire globe is used as the tissue of reference. An open globe injury implies a full-thickness wound and can be caused by a rupture (blunt object) or laceration (usually a sharp object). A penetrating injury implies a single laceration of the eyeball whereas a perforating injury involves two full-thickness lacerations (entrance and exit) of the eyewall.

Hyphema

Blunt or penetrating trauma may tear iris vessels resulting in hyphema (a layering of blood in the anterior chamber). The presence of hyphema confirms serious eye damage. Associated glaucoma, lens injury, and posterior segment damage are present in 25 to 30 per cent of these patients. Globe rupture must also be considered. Protection of the eye with a shield and immediate ophthalmic consultation are mandatory. Elevation of the patient's head promotes dependent settling of the blood and facilitates further evaluation. Surgical exploration may occasionally be needed to rule out rupture of the globe.

Many patients with hyphema are treated with topical cycloplegic agents, topical steroids, and antifibrinolytic agents. Activity restriction and globe protection are required during the first 5 days because this is the critical time for rebleeding. Rebleeding may be associated with glaucoma and corneal blood staining and is associated with reduced visual prognosis.

Orbital trauma/fractures

Blunt orbital trauma commonly results in lid swelling and ecchymosis. It occasionally causes orbital fractures, retrobulbar hemorrhage, and globe rupture.

The uncomplicated black-eye characterized by swelling, ecchymosis, and subconjunctival hemorrhage is best treated with cold compresses and pain relief that does not contain aspirin. Rarely, hemorrhage will occur in the retrobulbar space resulting in proptosis and elevated intraocular pressure. A retrobulbar hemorrhage with compromised visual function is treated with topical b-blockers, a-agonists, and intravenous carbonic anhydrase inhibitors to reduce intraocular pressure. Patients may also benefit from intravenous osmotic agents. Surgical decompression of the orbit may be needed if medical measures are not effective.

Fractures resulting from blunt trauma most commonly involve the thin inferior or medial orbital walls. Signs of possible orbital fracture include restricted eye motility, enophthalmos, epistaxis, hypesthesia over the infraorbital nerve distribution, palpable orbital rim stepoff fractures, orbital emphysema with crepitance, and cerebrospinal fluid rhinorrhea.

Definitive diagnosis requires radiographic confirmation. A routine radiographic series is indicated (Fig. 3). CT scanning has largely replaced X-ray tomography in determining the extent of inferior blow-out fractures.

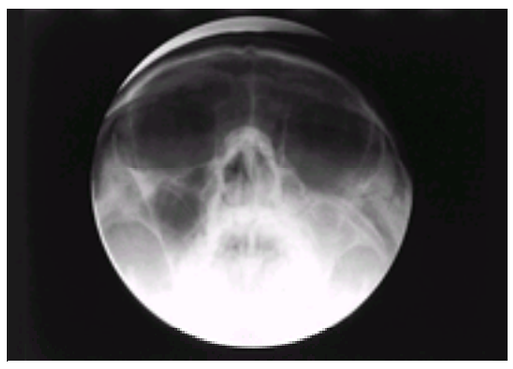


Fig. 3. Water's radiographic view demonstrating right-sided zygomaticomaxillary fracture with orbital fracture.

Supportive treatment of orbital fractures includes cold compresses and pain relief not containing aspirin. Systemic antibiotics may be indicated because of the associated sinus involvement. Eye and ear, nose, and throat (**ENT**) consultation is indicated. Orbital fractures are rarely surgical emergencies and if surgery is indicated, it is often best performed 4 to 10 days after the time of injury.

A number of surgical approaches are available to gain exposure for repair of orbital floor fractures. A subciliary incision can be used and coupled with inferior fornix or transconjunctival cul-de-sac incisions (Fig. 4). If required, a Caldwell–Luc approach via the gingival sulcus can also be used. Once exposure to the fracture is achieved, a malleable retractor is used to elevate the orbital contents superiorly and enhance exposure of the fracture site (Fig. 5). Entrapped tissues are freed with care taken to avoid trauma to the infero-orbital neurovascular tissues. Autologous or alloplastic implant material is then placed over the fracture site. Autologous tissue may be harvested from the iliac bone, rib, or calvarium. Alloplastic material includes bone cement, supramid, Silastic, and other mesh-like materials.

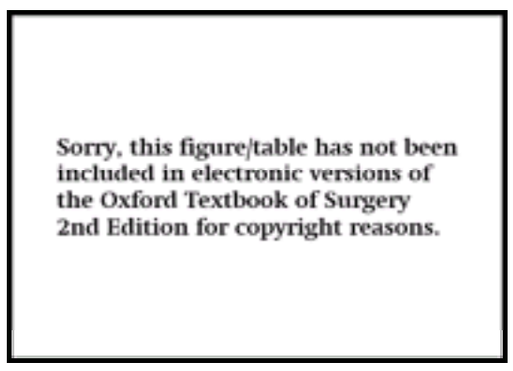
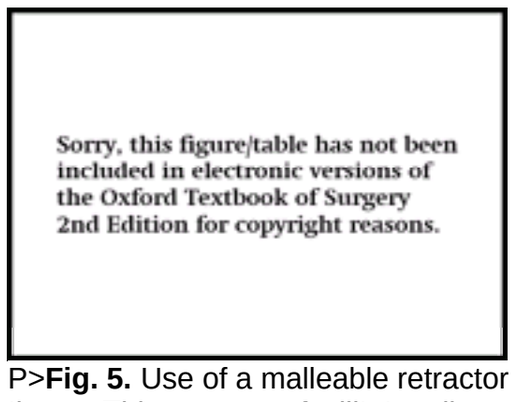


Fig. 4. Use of periosteal elevator to expose the fracture site after incising the conjunctiva, capsulopalpebral fascia, and periosteum.



P>**Fig. 5.** Use of a malleable retractor to retract the periorbita and orbital contents upward. This gives excellent exposure of the fracture site and any incarcerated tissue. This exposure facilitates dissection of entrapped tissues and placement of orbital implant material.

Profound visual loss in the setting of blunt orbital trauma may be caused by optic nerve injury. Such patients may benefit from intensive steroids or immediate surgical decompression of the optic canal.

Lid lacerations

As with orbital trauma, the physician evaluating lid lacerations must keep in mind the possibility of deeper penetration. In many cases, the diagnosis and treatment of occult globe injury takes precedence over the more readily apparent lid laceration.

Superficial lacerations not involving the lid margin should be cleaned and debrided of foreign material. Tetanus prophylaxis is indicated. Surgical repair is directed at maintaining full lid function and good cosmetic appearance. Traction on the lid margin should be avoided to minimize lid margin distortion.

Several types of lacerations call for particular care and handling (Fig. 6). First, any laceration extending into the medial third of the upper or lower lid may involve the tear drainage canalicular system. Surgical repair must include coaptation of the severed canaliculi. Second, deep lacerations of the superior portion of the orbit may involve the lid elevator (levator aponeurosis) resulting in an inability to raise the lid. To avoid ptosis, the levator aponeurosis tear must be recognized and repaired at the time of initial injury. Third, lacerations with fat prolapse imply deep penetration through the orbital septum. These lacerations are associated with a higher incidence of globe penetration, foreign bodies, lid functional impairment, and infection. Fourth, full-thickness lacerations involving the lid margin must be repaired in layers: tarsus, subcutaneous tissues/orbicularis muscle, and skin. Proper alignment of the lashline and mucocutaneous junction is essential. Improper closure may lead to notching of the lid margin, impaired lid function, and long-term breakdown of the ocular surface. Fifth, avulsion injuries often require repair based upon skin flap mobilization and skin grafts. Any piece of avulsed lid tissue should be saved for potential use in free grafting.



Fig. 6. Deep upper lid laceration through lid margin with loss of tissue and involvement of the superior lacrimal canalicular system.

Corneal abrasions and foreign bodies

Corneal abrasions are rarely vision threatening, but result in significant patient complaints due to foreign body sensation, watering, and light sensitivity. A topical anesthetic promptly relieves the pain and topical fluorescein highlights the area of corneal involvement (Fig. 7). Foreign bodies must be searched for by everting the upper lid. A foreign body on the conjunctival surface may simply be irrigated or removed with a cotton swab. Foreign bodies on the surface of the cornea are best removed using the magnification of loupes or a slit lamp. A 25-gauge needle or spud is used to flip the foreign body off the surface of the cornea. Residual rust rings may require more manipulation.

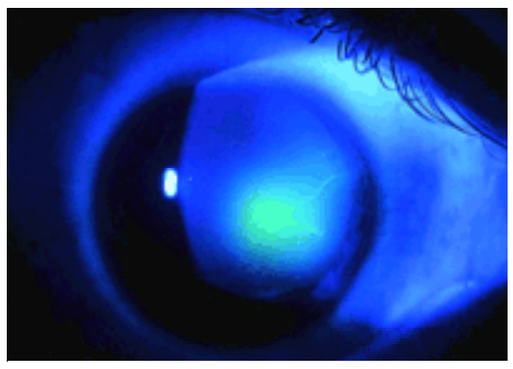


Fig. 7. Linear corneal abrasion caused by a fingernail highlighted with the use of topical fluorescein and cobalt blue illumination.

Patients with corneal abrasions and removed foreign bodies are treated with topical antibiotic and cycloplegic drops. A pressure patch is applied to maintain lid closure. Most corneal defects heal by epithelial migration under the closed lid within 24 h. Some abrasions may be managed by use of topical antibiotics and topical non-steroidal anti-inflammatory agents. Systemic analgesics may be necessary. Topical anesthetics should never be prescribed for home use by the patient.

Treatment of non-traumatic ocular emergencies

Red eye

The most important considerations in the differential diagnosis of the red eye are conjunctivitis, iritis, corneal infection, and acute glaucoma.

Conjunctivitis

Conjunctivitis or ‘pink eye', is the most frequent cause of red eye presenting to the physician (Fig. 8) Infectious conjunctivitis is most often caused by viruses and less commonly by bacteria: both types of conjunctivitis are associated with watering, ocular infection, discharge with crusting of the lids, and mild discomfort. More exuberant mucopurulent discharge is characteristic of bacterial infections and palpable preauricular nodes are common with viral infections. The corneas are clear on pen-light or slit-lamp examination. Gram-positive organisms are the most common etiologic agents in bacterial conjunctivitis and antibiotic selection must keep this point in mind. Viral conjunctivitis is best treated symptomatically; bland antibiotic ointments can provide soothing comfort. Neonatal conjunctivitis represents a potentially serious concern because of the possibility of gonococcal infection. *Neisseria gonorrhoeae* is the one bacterial species that can penetrate intact corneal epithelium. A minor external conjunctivitis can turn into a virulent endophthalmitis that is vision threatening. Topical and systemic therapy, including hospital admission, must be co-ordinated by the ophthalmologist and pediatrician.

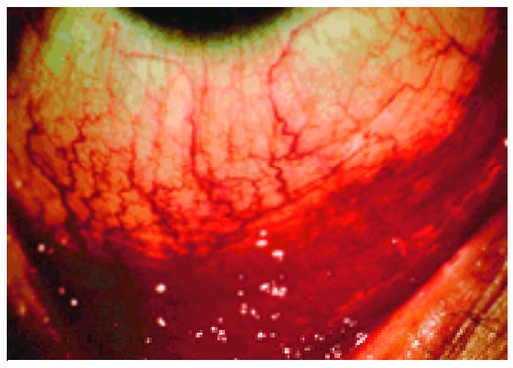


Fig. 8. Viral conjunctivitis with mild ocular infection and follicular changes of the lower lid.

Allergic conjunctivitis is typically seasonal in presentation and associated with itching, watering, and ocular infection. Discharge tends to be less prominent. These patients are best treated with cold compresses and a topical vasoconstrictor combining an ocular decongestant and antihistamine. Mast cell stabilizers can also be helpful in more prolonged cases of ocular allergies.

Iritis

Iritis is a term used for intraocular inflammation, in contrast to an external infection such as conjunctivitis. Patients often complain of blurred vision and light sensitivity. The eye is red but the redness tends to be accentuated in the circumcorneal area (Fig. 9). The pupil is often constricted when compared with the fellow eye. The cornea is generally clear but may be hazy if the intraocular pressure is elevated. Definitive diagnosis requires a slit lamp to detect inflammatory cells in the anterior chamber. Treatment includes topical cycloplegics and topical steroids. Topical steroids should never be initiated without discussion with an ophthalmologist. Short-term steroid treatment can unmask latent herpes simplex keratitis; in the long term it can cause glaucoma or cataract.

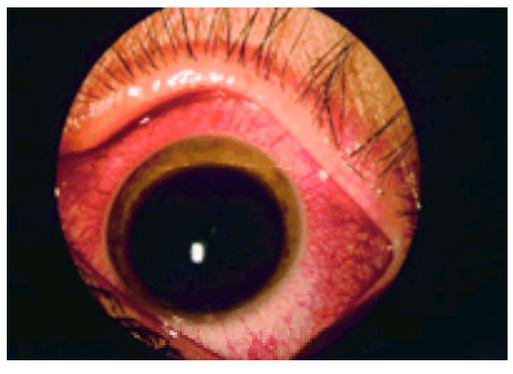


Fig. 9. Circumcorneal redness and inflammation with iritis.

Corneal infection

Corneal infection manifests as a focal infiltrate (Fig. 10) in the normally clear and lustrous cornea. A corneal ulcer by definition is an infiltrate with an overlying epithelial defect. The patient presents with a foreign body sensation and an altered corneal reflex on pen-light examination. Contact lens wear is a major predisposing factor. This is a vision-threatening emergency. Cultures are often taken of the involved area to determine etiology and antibiotic sensitivities.



Fig. 10. A focal corneal infiltrate confirms the presence of a potentially vision-threatening corneal infection.

Bacterial corneal infections may progress rapidly and require institution of topical antibiotic therapy, at a frequency up to every 30 min. The incidence of corneal infections has increased with the increasing popularity of extended-wear contact lenses. Gram-negative infections are more common in these patients. If there is any doubt as to the type of corneal infection, broad-spectrum topical antibiotics effective against both Gram-positive and Gram-negative organisms should be provided.

Herpes simplex keratitis constitutes a special type of corneal infection. Classically, patients have a dendritic pattern of fluorescein staining on the cornea (constitutes live virus) when viewed with a cobalt blue light. These patients require topical antiviral therapy. Topical steroids may exacerbate the problem. Fungal ulcers require topical antifungal therapy. An ophthalmologist should be actively involved in the care of patients with corneal infections.

Acute glaucoma

Acute angle closure glaucoma results from a sudden obstruction of aqueous outflow within the eye. Iris blocks the outflow structures and intraocular pressure rises dramatically. Patients have deep-seated aching pain that is often severe. This pain may even be referred to the chest and abdomen with associated nausea and vomiting. Vision is reduced due to corneal edema. There is a mottled light reflex from the edematous cornea. Elevated intraocular pressure is confirmed by tactile testing or tonometry.

The ophthalmologist should be involved as quickly as possible. Initially, medical therapy is instituted, including topical b-blockers, topical a-agonists, topical pilocarpine, and intravenous carbonic anhydrase inhibitors. An oral or intravenous hyperosmotic agent may also be indicated. Definitive therapy requires creation of a laser or surgical iridectomy (Fig. 11) to relieve iris pupillary block.

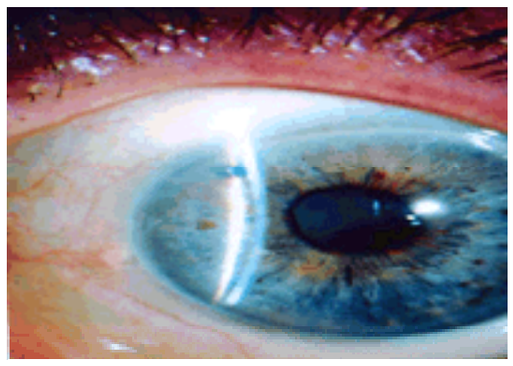


Fig. 11. Patent laser iridectomy created to relieve acute angle closure glaucoma.

In most cases, the ophthalmologist can create a laser iridectomy utilizing the argon/krypton laser or the neodymium YAG laser. Efforts are initially made to clear the cornea of edema to enhance visualization. This is accomplished with the glaucoma medications noted above as well as topical glycerin drops. Topical miotic agents help reduce the size of the pupil and so stretch the iris. A contact lens is then placed on the surface of the eye by the ophthalmologist at the laser. This lens has a convex button applied to the surface with high diopter plus power that allows magnification and focusing of the laser energy on an iris crypt. If the argon laser is used, a 50 μm spot size is preferred, along with a duration of 0.1 to 0.5 s and a power setting of 1000 mW or higher. It is often easier to achieve full-thickness penetration utilizing the neodymium YAG laser. The power settings are generally 5.0 mJ with 1 to 3 pulses per burst. It is important to examine the patient with a gonioscope following the creation of the laser iridectomy to establish whether the angle has opened. If opening has not been achieved, argon or krypton laser energy can be delivered to the peripheral iris to facilitate pulling of the iris away from the angle structures. Glaucoma medications are continued postoperatively in addition to topical anti-inflammatory drops. The patient is followed closely until the intraocular pressure is brought under satisfactory control.

Orbital cellulitis

Cellulitis may occur anterior or posterior to the orbital septum. The majority of these infections are due to contiguous sinus disease or associated dacryocystitis. Anterior preseptal cellulitis is associated with swelling, redness, and occasional discharge. Orbital cellulitis, with penetration deep to the orbital septum, may be heralded by decreased vision, impaired ocular motility, an afferent pupillary defect, increasing proptosis, or optic nerve swelling.

All orbital infections should be cultured and Gram stained as appropriate. Preseptal cellulitis generally responds to warm compresses, topical antibiotics, and oral antibiotics. Orbital cellulitis requires more aggressive systemic treatment including intravenous antibiotics and more extensive ophthalmologic, ENT, and radiographic evaluation to determine the underlining cause. Patients with cellulitis who are diabetic, chronically ill, or immunologically compromised represent particularly high-risk groups for developing a fungal infection (phycomycosis) that may be fulminating and life threatening.

Acute visual loss

Acute, non-traumatic loss of vision generally arises from a problem with the optic nerve or retina. The most important causes of painless loss of vision are central retinal artery occlusion, central retinal vein occlusion, vitreous hemorrhage, and retinal detachment. Painful loss of vision occurs with optic neuritis and temporal arteritis. Optic neuritis tends to occur in younger patients and is associated with painful movement of the eye. Temporal arteritis is associated with a more elderly population and a characteristic unilateral headache (jaw claudication).

The two causes of acute loss of vision that require immediate treatment are central retinal artery occlusion and temporal arteritis. Central retinal artery occlusion usually presents as a sudden, profound loss of vision in one eye of an elderly patient; it may be intermittent at first. Vision is reduced to the level of counting fingers or perception of light, and an afferent pupillary defect is present. Ophthalmoscopic examination reveals narrowed retinal arterioles and a ‘cherry red spot' macula (Fig. 12). Emergency treatment is directed at maximizing circulation to the optic nerve and retina. Circulation may be enhanced with rebreathing of carbon dioxide, topical b-blockers, intravenous carbonic anhydrase inhibitors, and intermittent digital massage if retinal emboli are documented. A surgical paracentesis for immediate

decompression of the globe may be performed by an ophthalmologist. All patients require neurologic evaluation to determine the source of the emboli.



Fig. 12. A ‘cherry red spot' macula characteristic of a central retinal artery occlusion.

Temporal arteritis is potentially blinding to both eyes. Patients present with decreased vision and exquisite tenderness over the forehead or scalp. An afferent pupillary defect is present and there may be hemorrhagic optic nerve swelling. Generalized symptoms of polymyalgia rheumatica may be present. If the erythrocyte sedimentation rate is elevated, oral systemic steroids are required. Immediate steroid therapy minimizes the chance of bilateral optic nerve involvement. A temporal artery biopsy can subsequently be performed on a more elective basis to confirm the diagnosis.

Further reading

Deutsch TA, Feller DB. *Paton and Goldberg's management of ocular injuries*, p. 312. WB Saunders, Philadelphia, 1985. [Concise, well written comprehensive textbook on ocular trauma.]

Eagling EM, Roper-Hall MJ. *Eye injuries, an illustrated guide*. JB Lippincott, Philadelphia, 1996. [Well illustrated guide to eye injuries.]

Grant WM. *Toxicology of the eye*, 2nd edn. Charles C. Thomas, Springfield, IL, 1974. [Encyclopedic review of agents with ocular toxicity.]

Kuhn F *et al.* A standardized classification of ocular trauma. *Ophthalmology* 1996; **103** (2): 240–3. [Proposes an international classification system for ocular trauma.]

Mead, MD. Evaluation and initial management of patients with ocular and adnexal trauma. In: Jakobiec FA and Albert D, eds, *Principles and practice of ophthalmology*, pp. 3362–82. WB Saunders, Philadelphia, 1994. [This book chapter thoroughly discusses the evaluation of ocular trauma and an appropriate diagnostic workup as well as delineates initial management.]

Mead MD, Shingleton, BJ. *Eye trauma and emergencies. A slide-script program*. American Academy of Ophthalmology, San Francisco, 1996. [Recently updated slide-script review of ocular trauma.]

Piermaci DJ *et al.* Perspective: A system for classifying mechanical injuries of the eye (globe). *American Journal of Ophthalmology* 1997; **123**(6): 820–31. [Discusses development and theory of an international classification system for ocular trauma.]

Roper-Hall MJ. *Eye emergencies*. Churchill Livingston, London, 1987. [A practical accompaniment to multispecialty care of the patient with an ocular emergency.]

Shingleton BJ, Frederick AR, Hutchinson BT. Ocular emergencies. In: Wilkins EW, ed., *Emergency medicine*, 3rd edn, pp. 876–91. Williams & Wilkins, Baltimore, 1988. [A cogent review of eye emergencies for the emergency department physician.]

Shingleton BJ, Hersh PS, Kenyon KR. *Eye trauma*. CV Mosby, St. Louis, MO, 1990. [A comprehensive textbook for management of the patient with eye trauma.]

Spoor TC, Nesi FA. *Management of ocular, orbital and adnexal trauma*. Raven Press, New York, 1988. [Well illustrated book emphasizing multispecialty approach to patients with eye trauma (facial plastics, neuroradiology, ophthalmology).]

45.2 Nose and sinuses

Richard R. Orlandi and David W. Kennedy

Introduction
Anatomy
Physiology
Evaluation
Congenital anomalies
Inflammatory conditions
Nasal airway obstruction
Septal perforation
Epistaxis
Trauma
Neoplasms
Nasal manifestations of systemic disease
Conclusion
Further reading

Introduction

The anatomy of the external nose, nasal cavity, and paranasal sinuses and the diseases that affect them are of interest to a number of surgical subspecialties including otorhinolaryngology, neurosurgery, ophthalmology, oral and maxillofacial surgery, plastic surgery, and general surgery. Diseases of this area afflict a large number of individuals, with chronic rhinosinusitis alone affecting 34 million people in the United States and accounting for at least $2.4 billion annually in direct medical costs. While most patients with nasal complaints suffer from conditions effectively managed without surgery, others may harbor far more serious disease. The proximity of the frontal and sphenoid sinuses to the intracranial cavity and of the ethmoid sinuses to the orbit may give rise to serious complications in unrecognized infections. Neoplastic processes affecting these areas typically present with seemingly innocuous symptoms that mimic uncomplicated epistaxis or rhinosinusitis. Whereas intranasal tumors may result in symptomatic nasal obstruction at a relatively early stage in the disease process, the accessory sinuses are silent areas, and late diagnosis of sinus neoplasia is the general rule. It is therefore incumbent upon the treating physician to be aware of the breadth of conditions affecting the nose and paranasal sinuses so that subtle early signs and symptoms of more serious pathology do not go unnoticed.

Anatomy

The nasal cavity and paranasal sinuses are related to the anterior and middle cranial fossae, the orbits, the oral cavity, and the soft tissues of the face. The nasal cavity is formed by the paired maxillary and palatine bones laterally and inferiorly, with the vault capped by the ethmoid complex superiorly. Posteriorly the nasal cavity and nasopharynx are bordered by the sphenoid body and basiocciput. The nasal cavity opens anteriorly as the piriform aperture, formed by the maxillary bones and the nasal bones. The external nasal complex is completed by the upper and lower lateral nasal cartilages which, combined with the nasal septum, account for the majority of the esthetic appearance of the nose (Fig. 1).

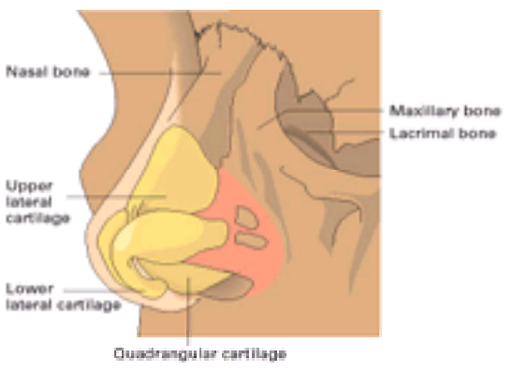


Fig. 1. Bone and cartilage of the external nose.

The nasal cavity is divided in the midline (or nearly so) by the nasal septum which is comprised of both bone and cartilage (Fig. 2). The posterior-inferior bony portion is derived from the vomer while most of the remaining bone is from the perpendicular plate of the ethmoid. The quadrangular cartilage completes the nasal septum more anteriorly and inferiorly and sits atop the anterior nasal spine formed by the maxillary bones.

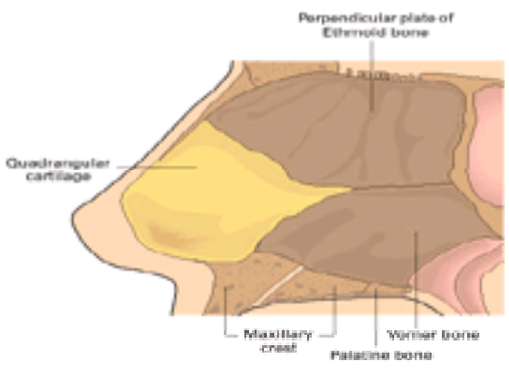


Fig. 2. Lateral view of nasal septum.

The nasal cavity contains three pairs of bony inferomedial projections from the lateral wall called turbinates (Fig. 3). Named for their position in the direct coronal plane, the inferior turbinate is a separate bone while the middle and superior turbinates are part of the ethmoid complex. The turbinates increase surface area within the nasal cavity and act as battens to direct the airflow in a laminar stream posteriorly into the nasopharynx. Between each of these shelf-like structures is a meatus named for the turbinate more superior to it and it is into these meati that the paranasal sinuses drain. The maxillary sinus, or antrum (of Highmore), drains into the middle meatus via a narrow infundibulum, as do the anterior ethmoid cells. The frontal sinus, often variably developed, drains into the middle meatus as well via the frontal recess of the ethmoid complex. It is because of these narrow common drainage pathways that inflammation in the anterior ethmoid sinuses may contribute to sinus disease in adjacent sinuses. The posterior ethmoid sinuses drain into the superior meatus, inferior and lateral to the superior turbinate, while the sphenoid sinus drains into the sphenoethmoidal recess, between the superior turbinate and the nasal septum. Into the inferior meatus, under the cover of the inferior turbinate, drains the nasolacrimal duct.

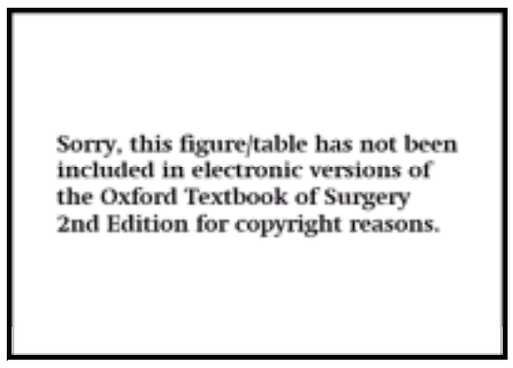


Fig. 3. Lateral nasal wall.

The paranasal sinuses as well as the majority of the nasal cavity are lined by pseudostratified ciliated columnar (respiratory) epithelium, which provides for mucus production as well as mucus transport. The nasal vestibule is lined by stratified squamous epithelium and contains numerous vibrissae, hairs that serve to filter large particles out of the airstream entering the nose. Posteriorly, as the nasopharynx transitions into the oropharynx, the histology also changes from respiratory epithelium to stratified squamous epithelium.

Physiology

The primary roles of the nose are respiration, conditioning of inspired air, and olfaction. The purpose of the sinuses has not been conclusively established but is hypothesized to include lightening of the skull, protection of the cranial contents during facial trauma, providing resonance to the voice, and, probably most importantly, mucus production.

The nose processes approximately 10 000 to 20 000 liters of air per day, warming and humidifying the air as well as trapping particulate matter. Processing of inspired air is accomplished through complex mechanisms involving mucociliary transport as well as the nasal microvasculature. Contained within the respiratory epithelium are goblet cells that secrete mucus. This mucus forms a biphasic or two-layered mucus blanket that covers the entire lining of the nasal cavity and paranasal sinuses. The upper layer, called the ‘gel' layer, is thick and is separated from the mucosa by the ‘sol' layer, which is much more thin and watery. The gel layer traps particulate material and protects the underlying mucosa from drying. Cilia from the mucosa project into the sol layer and, through co-ordinated beating move the gel layer along the underlying sol layer (Fig. 4). Within the sinuses the mucus blanket is transported toward the natural ostium and into the nasal cavity where it is moved posteriorly toward the nasopharynx. From the surgical standpoint, it is important to realize that the direction of mucociliary clearance is predetermined and is not changed by the creation of an additional opening into the sinus. Thus, in order to be functional from the standpoint of mucociliary clearance, any opening made into the sinus to improve drainage must include the natural ostium.

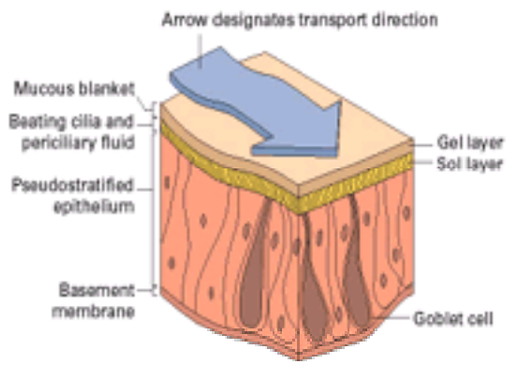


Fig. 4. Diagram of respiratory epithelium with its overlying mucus blanket. The arrow indicates the direction of mucociliary flow.

The mucus blanket moves at approximately 8 mm/min, resulting in a complete turnover every 10 min. Mucus and trapped particulate matter are transported into the nasopharynx and from there swallowed and eliminated via the gastrointestinal tract. Congenital defects in mucociliary clearance, either due to defects in ciliary motility or due to abnormalities in mucus viscosity (e.g. cystic fibrosis) can lead to mucus stasis and infection, typically presenting as frequent episodes of sinusitis at a relatively young age. Impaired mucociliary clearance within the lungs will lead to increased pulmonary infections as well.

The surface area exposed to the airstream is controlled by a complex system of blood vessels that allow for selective thinning or engorgement of the nasal cavity lining. This microvascular system is similar to erectile tissue located elsewhere in the body and is most fully developed along the anterior nasal septum and on the inferior turbinates. The vascular network consists of arterioles with precapillary sphincters that constrict and relax in response to certain neural and humoral transmitters. These arterioles lead to capillary networks and venous sinusoids or to arteriovenous shunts. When the inspired air is warm and moist, selective arteriovenous shunting of blood keepsthe lining thin. During exposure to cold or dry air, the blood preferentially flows through the capillaries and venous sinusoids, leading to engorgement of the erectile tissue and swelling of the nasal lining. The complex system of humoral and neural factors that control the nasal mucosa includes adrenergic and cholinergic input as well as vasoactive intestinal peptide and nitrous oxide. Most nasal decongestants are selective a-agonists that cause precapillary vasoconstriction, resulting in selective arteriovenous shunting of blood and decreased nasal edema.

Olfaction is an often overlooked nasal function. It not only provides for enjoyment of one's environment but also functions as a safety mechanism, warning of smoke, spoiled food, and the presence of noxious gases. High within the nasal vault, the cribriform plate of the ethmoid bone separates the anterior cranial fossa from the nasal cavity and is perforated by numerous olfactory fila. These bipolar projections from the olfactory bulbs synapse with chemoreceptors within the olfactory neuroepithelium, located on the roof of the nasal cavity and along the superior turbinate and superior portion of the nasal septum. Damage to the delicate olfactory fila as they pierce the cribriform plate, particularly by shearing forces from head trauma, may account for anosmia (loss of the sense of smell). Obstructions that prevent delivery of odor molecules to the olfactory neuroepithelium or damage to this specialized area (e.g. viral insult) can also result in anosmia.

Evaluation

Evaluation of a patient with nasal and sinus complaints begins with a complete history and physical examination with special attention to the head and neck. New-onset nasal obstruction, facial pain or pressure, and discolored nasal discharge suggest an infectious etiology. Malaise and fatigue may be more subtle presenting symptoms of a chronic infection. More systemic symptoms should also be sought, including allergic complaints and manifestations of systemic disease. New-onset unilateral nasal obstruction or recurrent unilateral epistaxis may be early symptoms of a neoplasm and should receive prompt attention. Hypesthesia of the maxillary division of the trigeminal nerve, while sometimes a symptom of inflammation, is more often the initial presentation of a maxillary sinus neoplasm and again deserves immediate evaluation. Conditions arising in the nose may spread to the orbit and present as changes in visual acuity, diplopia, pain with eye motion, epiphora, proptosis, or eyelid edema. Prompt diagnosis and treatment may avert permanent visual deficits.

Headache is a non-specific finding that can be caused by a long list of conditions, some of which are referable to the nose and sinuses. Certainly, acute sinusitis may give rise to headache or pain over the involved sinus. Evaluation of chronic headache often includes investigation of the sinuses as a potential source as well, but whereas facial pressure and pain may accompany chronic sinusitis, headache is less common as a primary symptom. However, there is some evidence to suggest that inflammation within the sinuses may be a trigger of another process (e.g. ‘rhinogenic migraine'). Contact points between the turbinates and nasal septum have also been hypothesized as sources of chronic headache. Unfortunately, procedures to relieve the pain by eliminating the contact have met with only mixed success.

Physical examination of the nose begins with an examination of the external structure of the nose. Subtle findings may give clues to underlying conditions, such as a transverse skin crease in the supratip region of the nasal dorsum caused by persistent rubbing of the nose during allergic episodes (so-called ‘allergic salute'). Examination of the interior of the nose requires illumination, either from a head mirror or from another source such as an otoscope. The condition of the mucosa is determined and the quality of nasal secretions should be noted. Anatomic abnormalities are best evaluated after application of topical nasal decongestants. The nasal

septum, inferior turbinates, and middle turbinates should be assessed with each nasal examination. Tenderness during percussion of the frontal or maxillary bones may indicate acute inflammation within these sinuses. Transillumination of the sinuses is of limited usefulness because of the potential for false-positive and false-negative results. Nevertheless, it may be of some help in confirming a clinical suspicion, assuming that the significant limitations of the technique are recognized.

Nasal endoscopy has greatly enhanced the clinician's diagnostic capabilities by providing exceptional visualization of the nasal cavity and its many recesses. The telescope's illumination and magnification allow for a thorough examination of the inferior and middle meati as well as the sphenoethmoidal recess and nasopharynx. When compared with the view provided by traditional examination methods, it is easy to see why nasal endoscopy has revolutionized the evaluation and treatment of nasal and sinus diseases (Fig. 5).

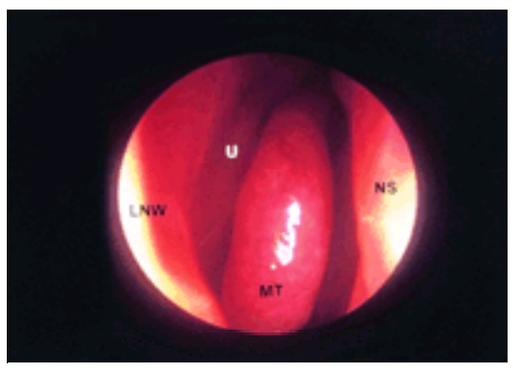


Fig. 5. View of the middle turbinate (MT), uncinate process (U), lateral nasal wall (LNW), and nasal septum (NS) obtained with a 30° nasal telescope.

The physical examination should also include a thorough head and neck examination. Rhinologic processes will often present signs in other areas. Nasopharyngeal neoplasms often present with neck metastases and typically result in serous otitis media due to eustachian tube dysfunction. Postnasal drip from nasal inflammation can often be seen in the oropharynx while oroantral and oronasal fistulas can be seen intraorally. Loose teeth or poor-fitting dentures may be relatively early signs of a malignancy of the maxillary antrum. Infraorbital nerve and orbital findings may also be seen in this disease.

Evaluation of the nose and paranasal sinuses may often include imaging studies. These are indicated for the assessment of space-occupying lesions and of mucosal changes within the sinuses. Imaging studies may also be indicated in the trauma setting to assess the integrity of the midface. Radiologic assessment of isolated nasal fractures is unnecessary, as it does not alter management. Plain radiographs of the nose and sinuses have limited usefulness when compared with computed tomography (**CT**) and have largely been replaced by the latter. Typically done in the direct coronal plane, these scans may be complemented by axial images as well. CT will delineate most space-occupying lesions within the nose and sinuses and is particularly well suited to demonstrate bony changes from expansion and extension outside the nasal vault. The principal shortcoming of using CT to evaluate nasal neoplasms is its inability to differentiate soft tissue from mucus. Lesions within the nose and sinuses will frequently lead to sinus obstruction and collection of mucus within the blocked sinus cavity. While this collection of fluid is difficult to differentiate from the tumor on CT, magnetic resonance imaging (**MRI**) can typically discern between the two, allowing for more accurate mapping of the tumor (Fig. 6). MRI is also of critical importance when there is sinus opacification adjacent to a skull base defect, as it allows differentiation of a nasal mass causing bone erosion from a meningocele or encephalocele.

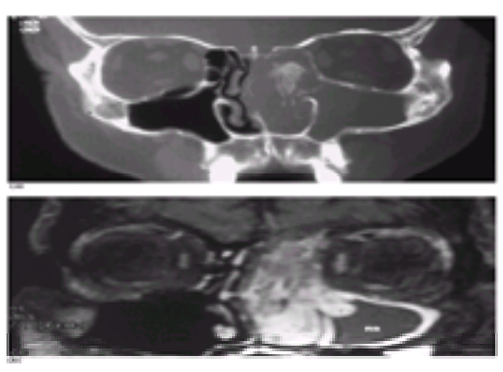


Fig. 6. Inverted papilloma. (a) CT demonstrates opacification of the maxillary sinus. (b) MRI (T_1-weighted contrast-enhanced with fat suppression applied) better delineates the extent of the tumor, differentiating between the retained secretions (RS) and the more brightly enhancing tumor (asterisk) at the maxillary ostium. (Images generously provided by David M. Yousem, M.D.)

CT is often used to assess chronic inflammation in the paranasal sinuses. It is typically performed in the direct coronal plane with 3- to 5-mm contiguous cuts imaged with a bone algorithm. Timing of the scan is critical so as not to overestimate the extent of disease. Patients with suspected chronic sinus infections should receive aggressive medical treatment for at least 4 to 6 weeks and then, if there is little or no symptomatic improvement, undergo CT imaging. It is essential that any acute inflammation be treated and allowed to resolve so that the patient's underlying chronic disease can be assessed with the scan. Patients imaged too close to an acute inflammatory episode may have striking changes on CT that typically resolve once the acute episode has passed.

It is important to remember that CT alone is insufficient for diagnosing chronic sinusitis. Some evidence of mucosal thickening in the sinuses can be seen in 30 to 40 per cent of the asymptomatic population. Conversely, it is possible to have evidence of chronic sinusitis on endoscopy along with symptoms, yet have an absence of significant CT changes. CT images are only one component of a complete evaluation and must be viewed in the context of a thorough history and physical examination, often including nasal endoscopy.

Congenital anomalies

The neural tube develops in the fourth week of embryogenesis and results from closure of the neural folds, proceeding from the midportion of the embryo toward the rostral and caudal ends. The anterior neuropore represents the last connection of the neural tube to the remaining ectoderm and typically closes at 25 days. Faulty closure or abnormal retention of the neural tube's connection to the overlying ectoderm can result in midline nasal masses. These typically present early in the postnatal period as either a widened nasal dorsum or a midline intranasal mass. If the intracranial connection is patent, the masses will often pulsate or enlarge with crying. Gliomas represent solid remnants of neural tissue that may connect with the intracranial contents. Meningoceles maintain patency with the subarachnoid space while meningoencephaloceles contain herniations of brain tissue as well. Dermoids may also present as midline nasal masses. They represent abnormal collections of embryonic tissue but do not have a relation to the anterior neuropore or intracranial contents.

Due to the potential for intracranial connection of any midline nasal mass, it is mandatory to establish the nature of the mass with imaging prior to any invasive manipulation, including fine-needle aspiration biopsy. Depending on the nature of the lesion, treatment may be local excision alone or excision combined with repair of the intracranial connection.

Choanal atresia represents a failure of the bucconasal membrane to break down during development of the posterior choanae. When bilateral, this condition becomes immediately manifest in the postnatal period, as infants are obligate nasal breathers. Typically the child will have adequate respiration while crying (oral breathing) but will demonstrate obstruction during quiet breathing and especially during feeding (nasal breathing). When unilateral, the lesion may go undiagnosed for some time but will often present later as unilateral foul discharge resulting from impaired clearance of nasal secretions. Treatment is directed toward widening the atretic segment, which is usually composed of both membranous and bony elements. The approach for this procedure has traditionally been transpalatal, but recent endoscopic repairs have demonstrated good results.

Congenital midfacial clefting can have a substantial impact on the nose. Palatal clefting leaves an obvious oronasal fistula that complicates alimentation and necessitates special feeding strategies. Unilateral cleft lip yields a characteristic nasal deformity consisting of a weakened and splayed lower lateral cartilage, short

columella, asymmetric nasal tip with depression on the side of the cleft, and horizontally oriented nostril. These factors must be taken into account during repair of the lip and often must be addressed in a subsequent rhinoplasty.

Inflammatory conditions

Inflammation within the nose has been traditionally referred to as rhinitis while that in the paranasal sinuses has been called sinusitis. As understanding of the physiology and pathophysiology of the nose and sinuses has improved, however, it has become clear that the two conditions are interrelated. CT imaging has shown that inflammation from a cold or allergies affects the sinus lining as well and rarely is sinusitis seen without a preceding or concurrent rhinitis. The Task Force on Rhinosinusitis has therefore proposed the term ‘rhinosinusitis' as a more accurate term for this disease entity.

Allergic rhinosinusitis results from an IgE-mediated degranulation of mast cells, releasing preformed mediators such as histamine. These chemicals cause vasodilatation and increased endothelial permeability, resulting in nasal edema. Sneezing as well as nasal and ocular itching are other common manifestations. The nasal mucosa typically appears pale and edematous and there may be clear nasal discharge. Smears of the nasal mucus usually reveal a large number of eosinophils, typical of an allergic reaction. Other manifestations of respiratory mucosal hyperreactivity may be present, including asthma and nasal polyposis. Allergen avoidance is the most effective treatment and identification of the offending allergens through formal allergy testing may be helpful. Topical nasal corticosteroids play a significant role in decreasing mucosal inflammation and antihistamines may be useful. Desensitization for patients who are severely affected or who respond poorly to medical regimens may offer significant relief.

Bacterial rhinosinusitis occurs when mucosal edema and/or anatomic abnormalities obstruct sinus ostia, resulting in pooling of secretions. Mucosal edema can arise from allergies, chronic infection, or acute viral upper respiratory infections. Hypoventilation of the blocked sinus cavity combines with the retained mucus to create an environment that favors bacterial growth. While the symptoms of acute sinusitis, with localized pain, acute onset of nasal congestion, and nasal discharge combined with systemic signs of infection, are relatively easy to diagnose, the symptoms of chronic sinusitis are muchmore subtle. Additionally, because the onset may be insidious, chronic sinusitis may significantly interfere with quality of life before it is diagnosed definitively. Chronic sinusitis typically presents with nasal congestion, nasal obstruction, and postnasal discharge. Pressure or pain does not usually occur in the involved sinuses until they have become obstructed. Subtle systemic changes may be present including fatigue, malaise, sleep disturbance, and intermittent low-grade fever.

Whether the infection is acute or chronic, treatment is directed at re-establishing ostial patency and mucociliary clearance. Nasal steroids once again play a major role and treatment of any underlying allergic conditions with antihistamines or environmental manipulation may be necessary as well. Oral steroids may be necessary in patients with significant airway hyperreactivity but should be used as infrequently and in as small a dose as possible. Oral decongestants can be helpful unless contraindicated and mucus-thinning agents or nasal irrigations may promote mucociliary clearance.

Often patients with chronic or recurrent acute rhinosinusitis are effectively managed with medical therapy. Some patients will fail this treatment, however, and should be considered for surgical therapy. Like medical therapy, surgical treatment of rhinosinusitis is aimed at re-establishing ostial patency and resolving inflammation. Formerly, the inflamed mucosa within the sinuses was thought to be irreversibly diseased and sinus procedures were designed to remove the mucosa and establish gravity-dependent drainage. The Caldwell–Luc procedure accomplished these goals and was the mainstay of surgical treatment for nearly a century.

Described by George Caldwell in 1893 and Henri Luc in 1897, the procedure involves a sublabial incision and stripping out the mucosa using forceps and curettes. Typically, a puncture through the medial wall of the sinus into the inferior meatus is made and widened. The mucosa that relines the cavity usuallly has disorganized ciliary transport, with resultant mucus pooling with intermittent infection and drainage through the inferior meatal antrostomy. Despite these drawbacks, this procedure was the standard treatment for what was thought to be irreversibly diseased mucosa. Using nasal telescopes, Messerklinger dispelled this belief by demonstrating reversal of chronic mucosal inflammation when mucociliary clearance was re-established. Furthermore, Hilding and subsequently Messerklinger showed that mucus is consistently transported within the sinus toward the natural ostium, even in the presence of artificially created ostia.

Many of the sinus drainage pathways converge on an area adjacent to and including the middle meatus which is called the ostiomeatal complex (Fig. 7). The ostiomeatal complex contains a number of narrow passages that may be blocked by mucosal edema. Focal inflammation in this critical area can lead, therefore, to obstruction of the maxillary, anterior ethmoid, and frontal sinuses and result in chronic mucosal inflammation in these areas. Inflammation may spread to the adjacent posterior ethmoid and sphenoid sinuses. Limited resection of the obstructing tissue in the ostiomeatal complex can lead to ventilation and drainage of the affected sinuses, allowing the inflamed mucosa to recover.



Fig. 7. Coronal CT image of the ostiomeatal complex. Note the narrowness of the drainage pathway bounded by the ethmoid infundibulum, ethmoid bulla, and uncinate process. O, orbit; M, maxillary sinus; B, ethmoid bulla; U, uncinate process; MT, middle turbinate; IT, inferior turbinate; *, ethmoid infundibulum.

Improved understanding of the pathophysiology of chronic rhinosinusitis has led to a significant change in its surgical treatment, whereby sinus drainage is re-established surgically while secondarily inflamed mucosa is left undisturbed. This procedure, made possible by the nasal telescope's increased visualization capabilities, has been termed ‘functional endoscopic sinus surgery' by Kennedy. Long-term follow-up data have validated the concepts behind this approach and it has become the mainstay for the surgical treatment of patients who fail medical therapy.

The increased magnification and illumination provided by the nasal telescope have extended its use to treatment of more complex conditions. Mucoceles, cerebrospinal fluid leaks of the anterior and middle cranial fossa, thyroid-related orbitopathy, traumatic optic neuropathy, and epiphora from lacrimal stenosis have all been effectively treated using endoscopic approaches. As experience increases and new instruments are developed, indications have and will continue to expand.

While the capabilities and potential of the nasal telescope are great, the potential complications are equally significant. The details of the various endoscopic procedures are outside the scope of this chapter but attention to the anatomic relationships surrounding the nose and paranasal sinuses reveals areas of possible injury. Injury to the floor of the anterior cranial fossa, which is as thin as 50 μm in some areas, can result in cerebrospinal fluid rhinorrhea or intracranial bleeding. Perforation of the medial orbital wall can lead to direct optic nerve injury or to orbital hematoma, both of which may result in blindness. Diplopia may result from medial rectus muscle injury and epiphora can occur if the nasolacrimal duct is disrupted. Fortunately, less serious complications such as bleeding, intranasal scarring, and hyposmia are more common.

Fungal sinus infections are relatively rare and can be classified based on the presence or absence of histologic mucosal invasion. Non-invasive (extramucosal) fungal sinusitis can be divided into fungal ball (mycetoma) and allergic fungal sinusitis. Allergic fungal rhinosinusitis occurs as a non-invasive form of fungal disease in which the host's immune response is triggered, leading to a vicious cycle of mucosal edema, sinus hypoventilation, and production of an extremely thick, tenacious mucus. This mucus is laden with fungus and eosinophils and sometimes referred to as ‘peanut butter' or ‘axle grease' secretions. Most frequently due to *Aspergillus* and *Bipolaris* species, the disease results in chronic sinusitis with possible sinus expansion and bone erosion. Orbital findings such as proptosis are not uncommon. Treatment is geared toward sinus ventilation and arrest of the allergic reaction. Oral corticosteroids combined with ventilation of the sinus and extensive debridement of the thick fungus-containing mucus are the mainstays of treatment.

Invasion of fungus into the mucosa is uncommon in immunologically competent patients in developed countries, but is common in the Sudan and some other regions. In patients who are immunologically normal, invasive fungal sinusitis is a slowly progressive disease.

In contrast, fungal infections in immunocompromised individuals are rapidly fulminant and may represent life-threatening complications of underlying disease or immunosuppressive therapy. Often these patients present with sudden onset of orbital or neurologic findings and imaging reveals focal sinus disease. Intracranial

extension may occur relatively early in the course of the disease, often involving the cavernous sinus. Patients with poorly controlled diabetes and those treated with chemotherapy or renal dialysis appear to have the greatest risk, with *Aspergillus* species and the *Zygomycetes* class (*Mucor, Rhizopus*, and *Absidia* genera) being the most common causative agents. Intranasal examination reveals gray or black necrotic tissue and biopsy will demonstrate the fungal hyphas. Treatment is debridement of the entire necrotic area and reversal of the underlying immunosuppression. Either of these two goals, or in some cases both, may not be attainable. Systemic antifungal agents are commonly utilized and topical antifungal treatment has also been tried. Even with aggressive therapy in these patients, the prognosis of invasive fungal sinusitis is poor, especially if the immunosuppression is not fully reversed.

Nasal airway obstruction

Diminished ability to breathe through the nose is a common complaint and can be attributed to a number of factors. Airflow through the nose normally follows a laminar pattern that passes through the vestibule and nasal valve area and then along the inferior and middle portions of nasal septum. The nasal valve is formed by the junction of the upper lateral cartilage and the nasal septum and may be an area of collapse, particularly in patients who have undergone partial resection of the upper lateral cartilage during rhinoplasty. Anatomic abnormalities in the inferior turbinate, middle turbinate, or nasal septum or in the mucosa covering these structures can result in interruption of the normal laminar airflow. Patients typically perceive the turbulence as an obstruction of flow. Structural irregularities may be congenital or traumatic and typically affect the nasal septum, presenting as deviations from the ideal midline placement. Additionally, absence of the turbinates or septum with the creation of a very large, empty nasal cavity will give rise to a feeling of nasal obstruction as a result of the loss of the normal laminar airflow. However, the sensation of the nasal obstruction is not just related to nasal resistance or airflow through the nose. Patients with obstruction of the sinus ostia will also frequently complain of a sensation of nasal obstruction, even in the absence of any intranasal airflow obstruction.

Numerous other factors can also lead to nasal obstruction from mucosal edema and congestion, including a cold or dry climate, allergies, irritants, antihypertensive medication, pregnancy and oral contraceptives, and overuse of topical decongestants.

A thorough history and physical examination will allow the physician to decipher the causative factors and treat the patient's underlying condition(s) accordingly. Strengthening the nasal valve often relieves symptoms in patients with collapse of this area and septoplasty may be helpful in patients with deviations in the quadrangular cartilage, perpendicular plate of the ethmoid, or vomer. Septoplasty can be performed under local or general anesthesia and is usually approached through an anterior incision just within the nasal cavity. A plane is developed beneath the mucoperichondrium and mucoperiosteum and the cartilage just anterior to the deviation is carefully incised. The same tissue dissection is performed on the contralateral side and the deflected bone and cartilage are addressed. Because the cartilage relies entirely upon the overlying mucosa for its blood supply, care must be taken so as not to create mucosal disruptions on opposed sides of the septum. This can lead to breakdown of the cartilage and a septal perforation. Because the nasal septum forms the majority of the support for the lower two-thirds of the nose, care must be taken not to weaken the caudal and dorsal portions. Closure of the incision with absorbable suture is commonly augmented by through and through sutures in order to keep the mucosal flaps opposed and prevent formation of a postoperative septal hematoma. Elevation of the mucosa off of the cartilage by a septal hematoma can cause cartilage destruction and also result in loss of dorsal or caudal support. Loss of septal structural integrity typically causes a saddle deformity of the nasal dorsum or drooping of the nasal tip.

In general, turbinate hypertrophy is a reflection of underlying local or systemic factors, although the inferior turbinates also undergo cyclical congestion and decongestion, known as the nasal cycle. In this physiologic process the predominant airflow alternates from one side of the nose to the other. This cycle can result in unilateral intermittent nasal obstruction or, more rarely, the sense of blockage may change from side to side. In patients with severe turbinate hypertrophy who fail medical therapy directed at the causative factors, surgical intervention is sometimes performed. However, a partial turbinectomy aimed at improving the nasal airway must be balanced with the sequellae of lost mucosal surface area, dryness, crusting, and potential worsening of the turbulent airflow.

Septal perforation

Perforations of the nasal septum most often result from surgical manipulation, exposure to environmental irritants, or cocaine use. Rarely, necrotizing inflammatory processes, neoplasia, chronic digital trauma, or inappropriate use of nasal sprays may be the cause. Loss of the integrity of the mucoperichondrium leads to necrosis of the underlying cartilage and a full-thickness tissue loss. Symptoms include a sensation of nasal airway obstruction due to turbulent flow, whistling during inspiration or expiration, crusting, and epistaxis. While the history is most helpful in arriving at the etiology of the septal perforation, biopsy may be warranted to rule out a granulomatous or malignant destructive process. Symptomatic treatment can often be accomplished with nasal saline spray and ointments to moisturize the mucosa and decrease the crusting and epistaxis. If whistling or a sensation of nasal obstruction is the primary symptom, many patients will benefit from a silicone ‘button' prosthesis that can be placed under local anesthesia. Patients who object to this device or tolerate it poorly may be candidates for surgical closure of the perforation. This procedure involves raising pedicled mucosal flaps within the nose and rotating them to cover the defect and has a significant failure rate, especially in large perforations.

Epistaxis

The rich vascular network of the nasal cavity provides for the many functions of the nose but also makes it a common source of bleeding. This rich blood supply reaches the nose from the anterior and posterior ethmoidal arteries, derived from the internal carotid artery, and the sphenopalatine artery, the terminal branch of the internal maxillary artery. Smaller contributions also come from the facial artery and other branches of the internal maxillary artery. Branches of these arteries arborize throughout the nasal cavity and a number of them coalesce on the anterior septum in a region called Kisselbach's plexus (Fig. 8).

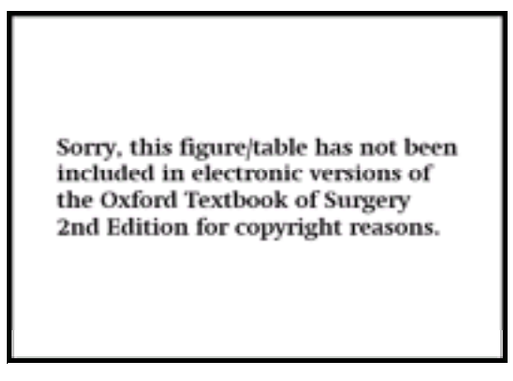


Fig. 8. Vascular supply to the nasal septum.

Kisselbach's plexus is the most frequent site of nasal bleeding, probably due to a combination of its rich vascular supply and its accessibility to digital trauma. Mucosal drying, seen in winter months as a result of indoor heating, and inflammation may contribute to epistaxis. Septal deviations and perforations can also cause focal drying due to airflow alterations. Neoplasms and foreign bodies are less common local factors. Systemic factors often play a role in epistaxis and undiagnosed disorders may initially manifest themselves as recurrent nasal bleeding. Aspirin use is a common causative factor and other forms of coagulopathy should also be considered. Hypertension may also be a factor.

Recurrent, severe epistaxis may be caused by hereditary hemorrhagic telangiectasia. This disease may also present with bleeding from other telangiectases within the gastrointestinal and respiratory tracts. Also known as Osler–Weber–Rendu disease, it is inherited as an autosomal dominant trait. Histologic examination of the telangiectases reveals arterioles within the nasal submucosa that lack adventitial and muscular coats, leading to fragile, dilated vessels that bleed easily. The traditional approach to this disease sought to replace the mucosa with a skin graft (septodermoplasty). More recently, lasers have been used to target the offending vessels and have met with moderate success in controlling the frequent bleeding.

The primary treatment for epistaxis, like all bleeding, is direct pressure. As most bleeding originates on the anterior nasal septum, this can be accomplished by pinching the nose. Should this initial maneuver fail to resolve the bleeding, anterior rhinoscopy after vasoconstriction can often isolate the source of the bleeding. Focal application of topical or injected vasoconstrictors may slow or stop the bleeding. Topical cocaine solution provides effective anesthesia and vasoconstriction and may be augmented by injection of a lidocaine and epinephrine solution. Cautery of the bleeding site can be accomplished with application of silver nitrate sticks or, if necessary, using electrical cautery.

Failure of more conservative measures may necessitate packing of the nose to control the bleeding. In many cases, endoscopic techniques for localizing and

cauterizing the bleeding vessels are making packing a less frequent necessity. Nevertheless, patients with coagulopathies, particularly those undergoing myelosuppressive chemotherapy with resulting thrombocytopenia, may have diffuse bleeding. Others may have profuse hemorrhage or nasal anatomy that prevents visualization of the bleeding site. Numerous materials have been used for nasal packing, including oxidized cellulose (Surgicel), gelatin sponge (Gelfoam), hydroxylated polyvinyl acetate (Merocel), latex balloons, frozen salt pork, and ointment-impregnated gauze. Whatever the material, the principle is the same. The nasal cavity should be thoroughly packed so that direct pressure is placed on the maximum amount of mucosa. Packing should be performed after the patient has been fully anesthetized with topical and, if necessary, injected agents. Intravenous narcotics and anxiolytics may decrease the patient's discomfort, but should be used judiciously in view of the patient's risk of aspiration and respiratory compromise.

Occasionally, the bleeding will not resolve with traditional anterior packing and is then operationally defined as a posterior bleed. Posterior epistaxis, usually arising from branches of the sphenopalatine artery, may require a posterior pack if it cannot be controlled with endoscopic-directed cautery. Posterior packs, which are uncomfortable and require inpatient monitoring, can be created with either a saline-filled balloon or rolled gauze. Some custom saline-filled epistaxis balloons have a posterior component that can be individually inflated as needed. Alternatively, a urinary catheter with a 30 ml balloon can be adapted for use. In anticipation of securing the pack at the end of the procedure, it is necessary to pass a 2 to 3 cm silicone collar around the catheter before placing it. A cut segment of an endotracheal tube satisfies this need quite nicely. The anterior packing is removed on one side and the catheter is passed into the nasopharynx. The balloon is then inflated while the oropharynx is carefully observed. When the soft palate begins to bulge, a small amount of saline is removed (so the palate no longer bulges) and the catheter is pulled anteriorly, snug against the posterior choanas. The anterior, ointment-impregnated gauze packing is then replaced. In order to keep the pressure against the posterior choanas it is necessary to secure the catheter against the anterior edge of the packing. This is most simply done with a C-clamp or umbilical clamp separated from the anterior packing by the silicone collar. This collar is crucial as it prevents the clamp from causing pressure necrosis on the nasal ala or columella.

The packing is left in place and the balloon inflated for 24 to 48 h, during which time any underlying factors predisposing the patient to epistaxis are corrected. The clamp is then removed, the balloon deflated, and the pharynx is examined for any fresh bleeding. If none is seen, the catheter is left in place with the balloon deflated for an additional 24 h whereupon, if there is no further bleeding, the catheter is removed. The anterior packing is subsequently removed after 1 to 2 days. The time-frames are not rigid and must obviously be adapted to the individual patient's overall status and response to treatment.

Nasal packing carries with it some risks and is painful both during placement and while in place. The risks include additional mucosal trauma and bleeding; toxic shock syndrome (estimated at 16 of 100 000 patients); alar, columellar, septal, or palatal necrosis; hemotympanum; sinusitis; and nasal airway obstruction with resultant hypoventilation and hypoxia. Previously, hypoxia following posterior pack placement was thought to be due to a nasopulmonary apneic reflex. However, studies have failed to demonstrate any changes in lung volumes, flow, or alveolar gas exchange with posterior nasal packing. Pre-existing lung disease, aspiration, sedation, and obstructive sleep apnea induced by the packing are more likely causes of the hypoxia. Regardless of the etiology, this alteration in oxygenation seen with posterior packing necessitates administration and monitoring of oxygen while the pack is in place.

Because of the discomfort and other problems associated with nasal packing, surgeons increasingly tend to favor direct intervention if local measures or the placement of a combined anterior and posterior nasal balloon fails to the control epistaxis. In addition to cauterization or ligation of the bleeding vessel under endoscopic control, open ligation or selective percutaneous vascular embolization should be considered when packing fails or as an alternative to prolonged posterior packing. Typically, epistaxis not arising from Kisselbach's plexus arises posteriorly from branches of the sphenopalatine artery. It may also arise superiorly in the nose from the anterior or posterior ethmoidal artery.

Open surgical ligation of the anterior and posterior ethmoidal arteries may be performed through a Lynch incision, medial to the medial canthus. The incision is carried down to the periosteum of the nasal bone and the lacrimal sac is carefully dissected out of its fossa. The periorbita is then elevated off of the lamina papyracea while seeking the anterior ethmoidal artery within the frontoethmoid suture. The anterior ethmoidal artery is then ligated with clips or treated with bipolar cautery. When ligation of the posterior ethmoidal artery is also necessary, the anterior ethmoidal artery is divided and the posterior ethmoid artery is then sought along the same suture line, approximately 12 mm posteriorly. Great care should be taken in the region of the posterior ethmoidal artery as it is 6 mm anterior to the optic nerve and may be absent in some cases. Once the arteries have been ligated, the skin incision is meticulously closed over a small drain.

Bleeding from the sphenopalatine artery may be controlled with ligation of the internal maxillary artery as it courses through the pterygopalatine fossa, posterior to the posterior wall of the maxillary sinus. This can be performed by endoscope transnasally or through a transantral approach utilizing a sublabial incision and opening in the anterior antral wall, as for a Caldwell–Luc procedure. To approach the internal maxillary artery, the posterior wall of the antrum is opened with a small osteotome and the opening is widened with punch forceps. The tortuous course of the artery can be seen within the fat of the pterygopalatine fossa and multiple clips are applied to the artery as distally as possible. Because of the rich anastomotic network involving the branches of the internal maxillary artery, the clips should be applied as distally as possible and the descending palatine branch of the artery should be ligated as well.

In many centers, endovascular embolization of the distal branches of the internal maxillary artery has largely replaced transantral ligation. The procedure was applied to epistaxis by Sokoloff in 1974 and has been shown to have a success rate of 70 to 90 per cent with a low incidence of complications. A variety of materials can be used for the embolization, including absorbable particles (Gelfoam, Avitene), non-absorbable particles (polyvinyl alcohol, detachable balloons, coils), and non-absorbable liquid (N-butyl cyanoacrylate). A pre-embolization arteriogram is typically performed to identify the bleeding point and also to identify any potentially dangerous intracranial anastomoses. The embolization material is delivered slowly with superselective angiographic technique so as to avoid reflux of the material into collateral vessels. While transient pain and trismus after the procedure is common, neurologic injury is rare, occurring in 0.1 to 1 per cent of cases.

More recently, direct transnasal endoscopic cautery and ligation of the sphenopalatine artery as it enters the nasal cavity have been described and are increasingly utilized for the control of epistaxis. These procedures rely upon endoscopic localization of the bleeding to the posterior nasal cavity. Often injection of the greater palatine canal with 1.5 ml of 1 per cent lidocaine with 1:100 000 epinephrine will sufficiently slow the bleeding and provide the patient with the necessary anesthesia. The anesthesia and vasoconstriction may be supplemented with topical cocaine. Endoscopic cautery or ligation may be performed under local or general anesthesia and, when performed promptly, may be used in lieu of posterior packing, decreasing patient discomfort and shortening hospital stays.

Trauma

Fracture of the nasal bones occurs commonly and may accompany other fractures within the face. Patients presenting after a traumatic injury should therefore receive a complete assessment, with attention to the bones of the midface as well as other areas as dictated by the history and physical findings. Often fractures of the nasal bones result in asymmetry of the nasal dorsum and may be obvious, especially before edema sets in. A palpable step-off or mobility of the bones may be appreciated on examination and severely depressed nasal bones may significantly narrow the nasal airway. Physical examination of the nose should also include a thorough evaluation of the nasal septum, as a septal hematoma may be present. This condition must be promptly recognized and treated in order to avoid septal cartilage necrosis and resulting saddle deformity of the nasal dorsum. Deviation of the nasal septum due to fracture or displacement off of the maxillary crest may also be seen on endonasal examination.

Radiographs of the nasal bones are not necessary to diagnose a nasal fracture. The treatment of nasal fractures is entirely based on history and physical examination. If a thorough examination fails to reveal any deformity or functional deficit, no additional treatment is necessary even if there is a non-displaced fracture. The patient should avoid additional trauma to the nose and the examination may be repeated 5 to 7 days later after edema resolves.

Treatment of nasal fractures is directed at re-establishing the pretraumatic appearance and, where necessary, correcting nasal airway obstruction. Correction of deviations in the nasal dorsum is best performed immediately, before edema makes assessment of the postreduction result impossible. More commonly, however, reduction of the fractures is performed 5 to 10 days after the initial insult when the edema has largely resolved. A fibrous union may restrict the mobility of the bones after 10 to 14 days, making an adequate reduction difficult. Timely reduction of the fracture should therefore be arranged when the diagnosis is first made.

Closed reduction of the nasal fracture is typically performed under local anesthesia, often supplemented with intravenous sedation. Anesthetic blockage of the infraorbital nerve as it exits the infraorbital foramen can be combined with intranasal injections and topical anesthesia. Any septal dislocation is first addressed and the fractured nasal bones are then manipulated and placed into their pretraumatic position. Photographs of the patient taken before the trauma may help with this goal but frequently palpation of the bones and visual inspection from the frontal and lateral aspects can sufficiently assess the success of the reduction. Tape and an external nasal splint are applied and left in place for 1 week to maintain alignment and protect the dorsum from minor trauma.

Septal fractures as well as severely depressed nasal fractures may require more extensive procedures including septoplasty or formal septorhinoplasty. Rhinoplasty may also be indicated in patients who do not achieve a satisfactory result from closed reduction of the nasal fracture. Such revisions should wait until edema and healing have sufficiently stabilized, usually at least 6 months.

Fractures of the nasoethmoid complex result from severe frontal impact and cause loss of the integrity of the nasal dorsum and medial orbital walls, sometimes with collapse of the entire nasoethmoid complex into the nasal cavity. Traumatic telecanthus is commonly seen, due to avulsion of the medial canthal tendon or lateral displacement of a fragment to which the tendon attaches. Patients with this injury will also demonstrate epiphora and blunting of the inner angle of the eye. Patients with nasoethmoid complex fractures often have accompanying intracranial, orbital, and other facial injuries and a comprehensive evaluation of such patients is

mandatory. Treatment of these fractures is aimed at re-establishing integrity of the medial canthal tendon attachment and disimpacting the complex from the nasal cavity. Associated frontal sinus fractures, orbital fractures, and cerebrospinal fluid leak are also repaired when present.

P>Trauma to the midface may result in fractures of the maxilla with extension through the piriform aperture, nasal dorsum, or fronto-nasal suture. These Le Fort fractures often require open reduction and internal fixation of the maxillary buttresses. Such treatment often will stabilize the nasal component of the fracture.

Neoplasms

Both benign and malignant neoplasms can affect the nasal cavity, nasopharynx, and paranasal sinuses. Benign sinonasal neoplasms include squamous papillomas, inverted papillomas, and nasopharyngeal angiofibromas. Malignancies in this region account for approximately 3 per cent of all upper aerodigestive tract cancers and up to 44 per cent are related to industrial exposures. Workers exposed to nickel, organic fibers from wood and leather, and volatile hydrocarbons have an increased risk of sinonasal malignancy.

Neoplasms of the nasal cavity, nasopharynx, and paranasal sinuses often present with vague, non-specific symptoms that can mimic less serious nasal diseases. Typical symptoms are unilateral nasal obstruction, epistaxis, facial pain, facial or palatal swelling, nasal discharge, infraorbital hypesthesia, and unilateral serous otitis media. On average, 6 to 8 months will pass from the time of initial presentation to arrival at the correct diagnosis. Often extension to the brain or orbit results in new symptoms that prompt the physician to re-examine the problem, usually with imaging studies that reveal the disease. Because the presenting symptoms are subtle and mistaken for rhinosinusitis or uncomplicated eustachian tube dysfunction, a high index of suspicion is needed to diagnose these lesions in their early stages. Many experts have advocated the concept that unilateral serous otitis media in an adult or unilateral sinusitis should be considered a neoplasm until proven otherwise.

Papillomas within the nasal cavity are benign neoplasms that can be divided into squamous and schneiderian types. The term schneiderian papilloma refers to a papillomatous growth of the schneiderian or respiratory epithelium and includes two types of papillomas, cylindrical cell and inverted papillomas. The latter is more common and typically arises from the lateral nasal wall, whereas cylindrical cell papillomas typically arise from the nasal septum.

Squamous papillomas typically are sessile warty lesions that affect the nasal septum. They may also be found in the nasal vestibule and are analogous to common skin warts. They may present as nasal airway obstruction or frequent epistaxis and are treated by localized excision, either with conventional techniques or with a laser. Like analogous epidermal lesions, they have a tendency to return but have little tendency toward malignant degeneration.

Inverted papillomas, on the other hand, have a 10 to 15 per cent rate of malignancy. Most arise from the middle meatus in the area of the ethmoidal bulla, and may affect drainage pathways within the ostiomeatal complex. Therefore, these lesions typically present as unilateral maxillary and/or ethmoid sinusitis. Like other papillomas they have a propensity to recur at a rate typically quoted in the 10 to 20 per cent range, but rates as high as 70 per cent have been reported. Complete excision is therefore mandatory and has been traditionally accomplished by performing a medial maxillectomy. In this procedure, the maxilla is exposed through a lateral rhinotomy or midfacial degloving approach. The resection includes the medial walls of the maxillary sinus and the orbit to the frontoethmoid suture, the floor of the orbit medial to the infraorbital nerve, the entire inferior turbinate, and many of the ethmoid sinus cells. Employing this procedure, the recurrence rate for inverted papilloma has been reported to be 15 to 30 per cent. More recently, endoscopic transnasal resection of inverted papilloma has been reported and, with selected lesions, has yielded equivalent recurrence rates.

Nasopharyngeal angiofibroma almost exclusively affects adolescent males. The lesion arises within the sphenopalatine foramen adjacent to the nasopharynx and grows into the nasal and sinus cavities, usually presenting with unilateral nasal obstruction and epistaxis which can be severe and is typically recurrent. Examination reveals a smooth-walled submucosal mass within the posterior nasal cavity or nasopharynx. Such a lesion found in an adolescent male should prompt the physician to obtain an imaging study before biopsy. The findings on CT and MRI are nearly pathognomonic, demonstrating a nasopharyngeal lesion involving the pterygopalatine fossa causing anterior bowing of the posterior wall of the maxillary sinus (Fig. 9). Subsequent angiography can confirm the diagnosis and obviate the need for a biopsy. The lesion is composed of vessels without a muscular coat so that biopsy can lead to extensive hemorrhage. Angiography also allows preoperative embolization, which has been shown to decrease intraoperative blood loss dramatically.



Fig. 9. CT image of a nasopharyngeal angiofibroma, extending into the pterygopalatine fossa. Note the pathognomonic anterior bowing of the posterior wall of the maxillary sinus. (Image generously provided by David M. Yousem, M.D.)

Nasopharyngeal angiofibromas may extend intracranially through the foramina that open into the pterygopalatine fossa, including the foramen rotundum, foramen ovale, foramen lacerum, and superior orbital fissure. They can also extend directly through the sphenoid sinus into the middle cranial fossa. These lesions obtain their principal blood supply from the terminal branches of the internal maxillary artery, often with a contribution from the ascending pharyngeal artery. Those lesions with intracranial extension often develop feeders from the internal carotid system. Treatment is surgical removal, with irradiation reserved for lesions not amenable to surgical resection such as those with substantial intracranial extension. Surgical excision may be accomplished through a medial maxillectomy (via a midfacial degloving or lateral rhinotomy) or, less commonly, through a transpalatal approach. The transpalatal technique employs a Wilson flap, a mucoperiosteal flap based on the greater palatine arteries. After the flap is elevated, part of the bone of the hard palate is removed on the side of the lesion and the mucosa of the nasal floor is incised, exposing the lesion. After the resection is completed, the flaps are meticulously replaced so as to avoid an oronasal fistula. Endoscopic techniques have recently been applied to this lesion and enable the complete excision of limited lesions which are well embolized preoperatively with minimal morbidity. However, because of the potential for severe bleeding, an endoscopic approach requires excellent interventional radiology, advanced skills, and specialized instrumentation.

Squamous cell carcinoma of the nasal cavity and paranasal sinuses comprises 70 to 80 per cent of malignancies in this area. Nickel workers are particularly predisposed to this lesion but, unlike squamous cell carcinoma in other areas of the upper aerodigestive tract, tobacco use has not been shown to be a significant risk factor. They most commonly present in the maxillary sinus and, because of their non-descript symptoms, typically are diagnosed in more advanced stages. Adenocarcinoma is the second most common malignancy of the nasal cavity and sinuses. Wood and leather workers have an increased risk of developing adenocarcinoma, thought to be due to organic fiber exposure. Minor salivary glands can be found throughout this area and other salivary malignancies, including adenoid cystic carcinoma, arise within these cavities. Esthesioneuroblastomas, also called olfactory neuroblastomas, arise from olfactory neuroepithelium although the exact cell of origin is unknown. They typically originate high in the nasal cavity or ethmoid sinus and, because of their position just below the cribriform plate, often extend intracranially. Melanoma and sarcoma are rare lesions of the nasal cavity and sinuses and carry a poor prognosis.

Lymphomas rarely occur within the nasal cavity and nasopharynx, comprising 0.17 per cent of extranodal sites. They are much more common in Asia, representing the second most frequent site of extranodal lymphoma after the gastrointestinal tract. Most nasal lymphomas are T-cell lesions and are often referred to as angiocentric immunoproliferative lesions. This term has replaced former descriptions of this lesion known as polymorphic reticulosis, pseudolymphoma, and idiopathic midline destructive disease. Angiocentric immunoproliferative lesions, as well as other localized lymphomas of the nose and sinuses, respond well to irradiation. B-cell lymphomas are more commonly seen in the nasopharynx, often in the lymphoid tissue of Waldeyer's ring.

Treatment of sinonasal malignancies, with the exception of lymphomas, commonly combines surgery and radiation. Because of the proximity to the orbit and brain, wide margins are usually not possible but complete excision with clear microscopic margins remains the goal of the resection. Extension into the orbit may necessitate orbital exenteration if the periorbita is breached. Similarly, transgression of the floor of the anterior cranial fossa often indicates a craniofacial resection, with an otolarnygologist resecting the lesion from below through a facial incision or by an endoscopic approach and a neurosurgeon completing the resection from above through a frontal craniotomy.

Adenoid cystic carcinoma, because of its propensity to invade and extend along nerves, can be particularly difficult to resect with a clear margin as the tumor often has

microscopic extension along the cranial nerves through the skull base. Radiation to the operative field is often combined with surgery and is also indicated for unresectable lesions. Patients with antral malignancies treated with surgery and postoperative radiation therapy exhibit a 45 per cent cure rate. Recent reports have indicated promise with adding chemotherapy to the treatment regimen for some lesions, particularly with esthesioneuroblastoma and sinonasal undifferentiated carcinoma. This clearly will be an area of future pursuit.

Nasopharyngeal carcinoma is a relatively rare lesion, comprising 0.25 per cent of cancers in North American males. However, in the Chinese population living in Taiwan, it is the leading cancer in men and the third leading malignancy in women. Dietary factors have been theorized to play a role and N-nitrosamines in salted fish have received much attention. Nevertheless, genetic factors are also thought to be involved because nasopharyngeal carcinoma is much more prevalent in the Chinese population living in North America (18 per cent of malignancies) than in the Caucasian and African American populations. Hearing loss from serous otitis media and the presence of a neck mass are the most common presenting symptoms. The serous otitis media stems from eustachian tube dysfunction from the nasopharyngeal mass and may be the only symptom of an early malignancy. For this reason, unilateral serous otitis media in an adult hould be considered a neoplasm until proven otherwise. Such a finding mandates thorough examination of the nasopharynx. Based on differentiation and presence of keratin, nasopharyngeal carcinoma has been classified by the World Health Organization into three types with prognostic significance. Type 1 lesions are similar to keratinizing squamous cell carcinomas and carry a 5-year survival of less than 20 per cent. Type 2 lesions are non-keratinizing and are also known as transitional cell carcinomas due to their resemblance to genitourinary tract malignancies. Undifferentiated tumors fall into type 3 and may be confused with large cell lymphomas. Type 2 and 3 lesions carry a 60 per cent 5-year survival. Radiation is the mainstay of treatment of nasopharyngeal carcinoma, which includes both sides of the neck because of the propensity for cervical metastasis. Surgery is not usually considered in primary treatment of this lesion but has successfully salvaged some cases that have failed radiation.

Metastatic disease must also be considered within the differential diagnosis of a sinonasal malignancy. Renal cell carcinoma is the most common cancer to metastasize to the nasal cavity and sinuses. It is theorized to spread via the vertebral venous system into the pterygoid plexus of veins during episodes of increased intra-abdominal pressure. It is sometimes difficult to differentiate between primary processes within the sinus and metastastic lesions, particularly with melanoma and adenocarcinoma. Often an exhaustive search for the primary lesion must be performed before metastasis can be ruled out and the sinonasal lesion can be considered primary.

Nasal manifestations of systemic disease

Diseases of immunosuppression typically cause recurrent bouts of bacterial rhinosinusitis. Immunoglobulin deficiencies, T-cell abnormalities, and acquired immune deficiency syndrome (AIDS) should be considered in patients with otherwise unexplained frequent infections. As noted previously, invasive fungal infections are serious sequellae of patients with depressed immunity and mandate urgent aggressive management. Frequently, recurrent infections in children and adolescents who are found to have nasal polyps on examination may be the initial presentation of cystic fibrosis. The infections result from inspissated mucus that becomes infected, often by *Pseudomonas aeruginosa*, and are often accompanied by pulmonary infections. Genetic screening has been developed recently and complements previously performed sweat testing. These tests should be strongly considered in any patient who developed their nasal polyps prior to age 20 years, even if the patient is in the fourth or fifth decade of life and still without other manifestations of the disease.

Wegener's granulomatosis is a necrotizing vasculitis that affects the upper and lower respiratory tracts and the kidneys. The vast majority of patients present with symptoms of nasal and sinus involvement, including crusting, epistaxis, purulent discharge, and saddle nose deformity. Systemic symptoms such as malaise, weight loss, arthralgias, and fever are common. Intranasal examination reveals mucosal ulceration with progressive destruction of the septum and turbinates. Laboratory studies show an elevated erythrocyte sedimentation rate and antineutrophil cytoplasmic antibody (**ANCA**) titer. Definitive diagnosis relies upon tissue biopsy, which demonstrates necrotizing vasculitis and granulomas. Unfortunately, the nasal cavity often does not yield tissue with these classic findings and definitive diagnosis may require biopsy from the lung and/or kidney. Treatment for Wegener's granulomatosis consists of cytotoxic agents, typically cyclophosphamide, but methotrexate has been used in some cases. Oral steroids and trimethoprim-sulfamethoxazole have also been included in treatment protocols for this disease.

Chronic nasal infections that demonstrate friable or ulcerated mucosa may represent *Mycobacterium tuberculosis* or *Klebsiella rhinoscleromatis* infection. These infections cause granuloma formation and on gross examination appear similar to Wegener's granulomatosis or other granulomatous processes such as sarcoidosis or leprosy. Diagnosis can be made by biopsy and culture of the nasal mucosa with additional laboratory evaluations, including ANCA and angiotensin-converting enzyme (ACE) levels, erythrocyte sedimentation rate, serum calcium, purified protein derivative (PPD) testing, and chest radiograph.

Rhinosporidium seeberi is another infectious agent that has a predilection for the nose. Infection occurs after swimming in contaminated water and presents as a large painless polypoid growth. Congenital syphilis may initially present as a mucoid rhinorrhea often referred to as snuffles.

Conditions resulting from immunologic dysfunction, such as polyarteritis nodosa and relapsing polychondritis, may also affect the nose. Polyarteritis nodosa causes a necrotizing vasculitis that can be seen affecting the mucosa of the nasal cavity. The absence of granulomas separates this disease from Wegener's granulomatosis. Relapsing polychondritis can cause nasal septal resorption and typically affects the cartilage of the external ear as well.

Conclusion

The nose and sinuses play host to a variety of pathologic processes, including congenital, infectious, traumatic, autoimmune, and neoplastic diseases. Knowledge of the anatomy and physiology of the nasal cavity and paranasal sinuses increases one's ability to diagnose and treat this variety of disorders effectively. Advances in technology have led to a better understanding of conditions such as rhinosinusitis, cystic fibrosis, and angiocentric immunoproliferative lesions of the nasal cavity. Improvements in treatment have and will continue to flow from such progress.

Further reading

Adams PF, Marano MA. *Current estimates from the National Health Interview Survey. 1994*. National Center for Health Statistics, Hyatsville, Maryland, 1995. (Vital and Health Statistics, Series 10, No. 193.) (DHHS Publication No. (PHS) 96–1521). [This article reports on the results of a survey of self-reported health problems.]

Cleary KR, Batsakis JG. Sinonasal lymphomas. *Annals of Otology, Rhinology, and Laryngology* 1994; **103**: 911–14. [Review of sinonasal lymphomas, including discussion of previous nomenclature.]

Eden BV *et al.* Esthesioneuroblastoma. Long-term outcome and patterns of failure—the University of Virginia experience. *Cancer* 1994; **73**: 2556–62. [This article discusses advances in the treatment of these nasal tumors, including chemotherapy and radiation.]

Kennedy DW. Functional endoscopic sinus surgery. Technique. *Archives of Otolaryngology* 1985; **111**: 643–9.

Mathog RH. *Atlas of craniofacial trauma*. WB Saunders, Philadelphia, 1992. [Comprehensive text describing pathophysiology and treatment of craniofacial trauma.]

Messerklinger W. *Endoscopy of the nose*. Urban & 38; Schwarzenberg, Baltimore, 1978. [Atlas of endoscopic examination and treatment for a variety of nasal and sinus diseases.]

Orlandi RR, Kennedy DW. Surgical management of rhinosinusitis. *American Journal of the Medical Sciences* 1998; **316**: 29–38. [A review of pertinent paranasal sinus anatomy as well as the pathophysiology and treatment of rhinosinusitis.]

Report of the Rhinosinusitis Task Force Committee Meeting. Alexandria, Virginia, August 17, 1996. *Otolaryngology—Head and Neck Surgery* 1997 Sep; **117**: S1–S68. [This comprehensive report discusses the evaluation, diagnosis, staging, and medical and surgical management of rhinosinusitis.]

45.3 The evaluation of a neck mass in the adult patient

C. M. Norris, Jr. and Marvin P. Fried

History
Physical examination
Additional studies
Benign congenital conditions
Inflammatory conditions
Malignancies
Treatment of neoplastic disease
Further reading

Concern about a neck mass is one of the most frequent reasons for which a specialty consultation is sought. While it is now fairly well-established in the public psyche that a ‘lump' in the neck could mean cancer, the route through which a patient ends up with a definitive diagnosis is still often quite tortuous. Depending on age, 80 to 90 per cent of non-thyroidal masses in the adult neck are neoplastic. Over two-thirds are malignant, and most of the malignancies are metastatic lymph nodes, usually from a source in the upper aerodigestive tract, but possibly from the lungs or gastrointestinal tract. By contrast, the vast majority of pediatric neck masses are benign, and the malignancies are most often lymphomas. These observations notwithstanding, the possible differential diagnosis of any given adult neck mass is extensive and diverse, comprising infectious, neoplastic, inflammatory, congenital, immunologic, cystic, and variant anatomic conditions from a multitude of locations and histologic sources. The history, physical examination, and some well-selected additional studies can usually narrow the diagnostic spectrum considerably.

History

The interview of a patient with a neck mass should not only probe for the possible source of the mass, but also elicit symptoms of a possible primary site of malignancy and those of systemic illness.

The history should include specific questions relating to the duration of the mass and whether it has changed in size or character, fluctuations in its size, and associated local symptoms of pressure or pain, as well as possible changes in characteristics of the skin overlying the mass. Associated symptoms of voice or speech changes, shortness of breath, cough, stridor, dysphagia, pain, odynophagia, or hemoptysis are important. Weakness of motion of the face, lower lip, shoulder, or tongue should be investigated as they suggest nerve impairment. Otalgia, hearing loss, nasal obstruction or bleeding, change in the sense of smell, and any self-identified or previously treated skin or mucosal lesions should be identified and characterized. Systemic symptoms of weight loss (and its putative cause), fever, night sweats, chills, malaise, anorexia, or pruritus can be determinate clues, as well as, in some cases, thoracic or gastrointestinal complaints.

Any history of the use of tobacco or alcohol should also be pursued as an estimate of cancer risk, though the absence of this history does not rule out the possiblility of a mucosal malignancy of the head and neck. The past medical history must identify the existence of prior malignancy and its treatment, as well as any biopsies or samples of skin or mucosa that may have been taken, even in the distant past. While rarely contributory, the family history can be determinate in cases of neurofibromatosis, tuberculosis, carotid body tumors, and other congenital and neoplastic disorders. The occupational history may uncover a specific cancer or infectious risk.

Physical examination

The anatomy of the neck is compact and complex, and includes a broad range of different structures. The location of a mass, and its relation to nearby known, normal structures can be extremely helpful in paring down the differential diagnosis. For example, the distinction between an enlarged upper jugular lymph node and a parotid gland mass can make all the difference between a likely malignant and a likely benign diagnosis.

Some normal or variant anatomic structures might be mistaken for a pathologic neck mass, such as an elongated transverse process of the second cervical vertebra (Fig. 1). Similarly, an elongated styloid process may present itself underneath the angle of the mandible. Other skeletal structures such as the hyoid bone may be splayed laterally and palpated in the upper lateral neck. The lateral process of the sixth cervical vertebra, the tip of the mastoid bone, and an anomalous cervical rib can be palpated as a firm mass in the neck. The critical finding in assessment of all of these circumstances is the bilateral symmetry of these masses. A dilated carotid bulb or distended jugular vein can also be confused with an abnormal mass. The skin of the head and neck and the fundi should also be examined, because of the possibility of malignant melanoma.

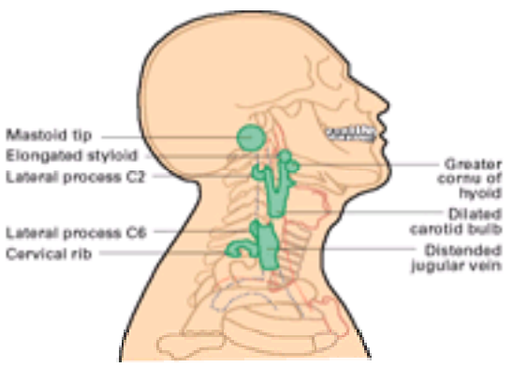


Fig. 1. Normal anatomy. Normal structures may be confused with pathologic processes, especially if the structure is prominent, or the patient is thin.

The physical examination must include a complete description of the mass itself. Size, firmness, mobility, tenderness, fluctuance, and any change in surrounding structures, such as the overlying skin, are germane. A thorough examination of the accessible mucosa of the head and neck, as well as the skin, ears, and cranial nerves, is essential. Selectively, evaluations of the chest and abdomen may be appropriate. The most limiting factor in making a diagnosis of head and neck cancer is an inadequate examination.

The ear examination may reveal a middle-ear effusion, indicating a process in the nose or nasopharynx, or a skin lesion of the canal or auricle, or, rarely, a middle ear neoplasm or infection.

The presence of a unilateral mass within a nasal passage, particularly in an adult, is of concern. Although the nasal cavities and nasopharynx may be difficult to examine, every attempt should be made to do so, and evidence of mucoid, purulent, or bloody drainage, mass lesion or asymmetry, or mucosal outcropping or ulceration should be noted. Referral for specialist evaluations by fiberoptic instrumentation, if not otherwise available, can be critical.

A thorough examination of the oral cavity should note the presence of any mucosal lesions, their size, character, and ‘feel' since palpation is an integral part of the examination. The lateral and posterior aspects of each side of the tongue must be examined carefully. Lingual carcinomas arising in this region give minimal symptoms of pain or discomfort. Motion of the tongue should also be checked, and both digital and bimanual examination of the oral cavity and tongue base carried out. Size and symmetry of the tonsils should be assessed and examinations performed in conditions with adequate lighting. The hard palate and posterior floor of the mouth are not easily seen on routine examination and must be consciously investigated.

The hypopharynx and larynx will require a mirror examination or transnasal fiberoptic endoscopy, an examination that can be performed simply with the appropriate equipment and with minimal topical anesthesia. Excellent results can also be achieved with rigid endoscopes that have high-quality optical characteristics. Attention should be placed on the configuration of the epiglottis and the piriform sinuses, which lie on either side of the aryepiglottic folds. Particularly the tongue base, vallecula, and subglottic larynx can be elusive, even with fiberoptic equipment in experienced hands.

The cranial nerves and skin of the face, neck, and scalp should also be consciously assessed, the latter particularly for evidence of active lesions obscured by hair, or

scars representing forgotten biopsy or excision sites.

Palpation of the neck should be performed to assess neck symmetry. The mass in question should be described in terms of location, size, consistency, and whether pain or tenderness can be elicited. The size should be measured so that any change can be detected in subsequent examinations. Not all palpable lymph nodes are necessarily abnormal, nor are small masses necessarily benign. Fixation of the mass to surrounding structures may be difficult to assess, but degree of mobility should be noted, particularly if the mass moves with swallowing. True fixation almost always represents advanced malignancy if a normal bony structure has been ruled out. The location of multiple or bilateral masses in the neck may be essential to the refining of a diagnosis.

Many masses in the adult neck entail pathologies of lymph nodes. For purposes of lymph node or mass localization, the neck itself can be considered to be divided into anatomic units, or ‘levels', representing the submental/submandibular (level I), upper jugular (level II), middle jugular (level III), lower jugular (level IV), posterior cervical and supraclavicular (level V), and anterior midline or paratracheal (level VI) regions (Fig. 2). A diagram should be entered in the patient's chart, along with a written description of the mass.

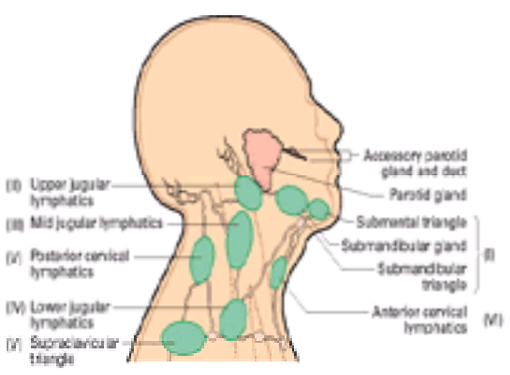


Fig. 2. Normal lymphatic and glandular structures. The major regions and the location of the parotid and submandibular glands are designated. ‘Level' nomenclature indicated in parentheses.

Additional studies

Additional laboratory studies which may be required for diagnosis include blood tests, various imaging modalities, needle aspiration cytology, and sometimes, ultimately, endoscopic mucosal or excis-ional neck mass biopsies. As in all aspects of medicine, the interpretation of test results is critically tied to the context in which they are performed.

With the advent of fine-needle aspiration ‘biopsy', more correctly, fine-needle aspiration cytology, much of the blood work previously relevant to and, at times, helpful in the actual diagnosis of a neck mass has been rendered non-contributory. Blood tests may, of course, still be very important in the management of the patient, the furtherance of a patient's treatment, or the identification of relevant comorbidities.

Conventionally, a complete blood count and biochemical profile, including liver function tests and calcium, and a urinalysis are performed. These may support a developing diagnosis of an inflammatory or infectious entity, lymphoma or leukemia, disseminated malignancy, or other entities reflected in various organ dysfunctions. Occasionally, studies for mononucleosis, toxoplasmosis, Epstein–Barr virus, or cytomegalovirus may be warranted if the condition is felt to be inflammatory. The relevance of a positive HIV test is now well-established, particularly in the context of opportunistic infections and lymphomas.

A chest radiograph looking for a single nodule or multiple densities may disclose a primary tumor or metastases. An apical lordotic view should be ordered to assess the upper lobes for both malignancy or clues for active tuberculosis. Sinus radiographs, particularly the lateral views of the nasopharynx, may be helpful in selected cases, as also may be barium contrast esophagography.

When anatomically warranted, the thyroid gland may be examined by ultrasound with or without concomitant fine-needle aspiration under guidance. The traditional radionucleotide thyroid scan is not diagnostic of a single, dominant mass and is generally held in abeyance, lest the diagnosis of thyroid malignancy suggest a more important use for iodine-mediated imaging and treatment.

A computed tomographic (**CT**) scan or magnetic resonance imaging (**MRI**) scan may be invaluable in determining the location of a neck mass, particularly in reference to surrounding structures, but will rarely, by itself, be diagnostic. Such so-called ‘cross-sectional imaging', combined with a good history and complete physical examination, increases diagnostic accuracy, but one should not supplant the other. Imaging may also reveal some physical characteristics of the mass, whether solid or cystic. An MRI scan is particularly useful for soft tissue evaluation and gadolinium enhancement will help define the vascularity of the mass in question. Consultation with a radiologist for the best techniques to use in the study of any individual lesion is always advisable, lest inappropriate studies be unnecessarily performed or their interpretation out of context be misleading.

In addition to a barium esophagogram, in patients with a supraclavicular mass, further radiographic studies of the stomach, pancreas, and colon are appropriate in searching for a primary disease process.

The vast majority of neck masses can be assessed on the basis of history and physical examination alone. Additional studies are required only where appropriate. In patients over 40 years of age, and particularly in those with a history of cigarette smoking and alcohol ingestion, the primary concern must be a metastatic malignancy until proved otherwise. The primary tumor is most often identified as arising from the mucosa of the head and neck. Masses presenting in the anterior midline are usually thyroid-related, congenital cysts or neoplasms.

Over the last two decades, in most medical centers in which the study is available, fine-needle aspiration cytology (**FNA**) has emerged as the single most important ancillary test to be performed in the evaluation of an adult neck mass, not infrequently in combination with one or another form of neck imaging. The FNA diagnosis of metastatic squamous cell carcinoma of the head and neck will pre-empt most other tests and focus the investigator on the need to identify the primary source of malignancy in the upper aerodigestive tract mucosa, usually through multiple endoscopic examination under general anesthesia, with directed mucosal biopsies.

For many other entities, an FNA diagnosis is sufficiently accurate to permit a move to the treatment phase of a patient's condition. An increasing array of immunohistochemical, microbiologic, and other diagnostic studies may now be done on cytologic preparations. At times, surgical excision of the mass will be diagnostically necessary for confirmation and/or more elaborate testing. For congenital cystic and benign neoplastic conditions, as well as some malignancies of the thyroid or major salivary glands, excisional or resectional surgery will also be therapeutic. In the case of metastatic head and neck malignancy, ritual excision of the mass without other investigations is, however, disfavored, and should only be performed as a ‘last resort'.

Benign congenital conditions

Although most congenital lesions present before adolescence, it is not unusual for a congenital neck mass to become apparent in adulthood (Fig. 3). Though several unusual presentations have been described, branchial cleft cysts are most often asymptomatic, manifesting as a painless, relatively soft, somewhat discrete mass in one of several characteristic locations. An exception would be the occasional case in which a deep neck infection or abscess developed around or within a branchial cyst. First-cleft branchial cysts present near the angle of the mandible, while those on the second and third pouch are found anterior to the mid-portion of the sternocleidomastoid muscle, sometimes associated with a draining or blind-ended sinus tract. The appearance on cross-sectional imaging and needle cytology characteristics are typical and fairly well-established. Treatment entails complete surgical excision of the cyst along with any associated sinus tract. However, particularly in older individuals with a risk of head and neck cancer, cervical lymph node metastases can be single, large, and cystic, and risk confusion with a branchial cleft cyst if a circumspect clinical approach to the diagnostic algorithm is not followed. Metastatic thyroid cancer and low-grade mucoepidermoid carcinomas of the major salivary glands may also have a significant cystic component and risk a mistaken diagnosis.

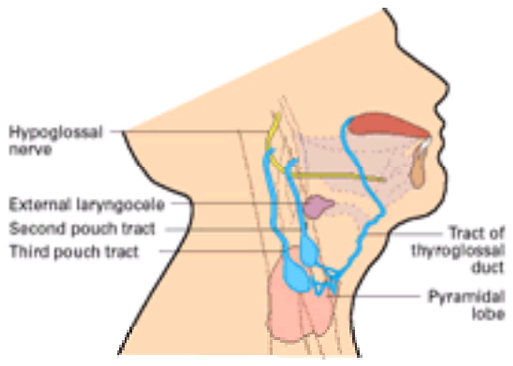


Fig. 3. Embryologic cervical vestiges found in the adult. Second branchial sinus tracts arise in the tonsillar fossa, coursing between the internal and external carotid arteries and over the hypoglossal nerve, presenting anterior to the sternocleidomastoid muscle (SCM). Third branchial sinuses arise from the piriform sinus, travel posterior to the carotid vessels and over the hypoglossal, and are anterior to the sternocleidomastoid muscle. Thyroglossal duct cysts are found near the midline, laryngoceles in the thyrohyoid region.

Thyroglossal duct cysts are located in the anterior midline of the neck in the vicinity of the thyrohyoid membrane where they can be confused with other submental masses such as epidermal inclusion cysts and salivary duct abnormalities. Following the embryologic route of the thyroglossal duct, these lesions can arise anywhere from the midline tongue base, at the foramen cecum, to the hyoid bone, or as far inferiorly as the thyroid gland itself. The duct remnant can be up to a couple of centimeters paramedian, as can a cyst arising from it. The tract or associated cyst can become infected, giving rise to an inflammatory mass or abscess which might require appropriate antibiotic therapy before the diagnosis is clarified. To prevent recurrence, the midline body of the hyoid bone must be excised along with the cyst and any duct remnant, possibly all the way to the tongue base (a ‘Sistrunk procedure').

Less common congenital lesions include laryngoceles, which arise from the laryngeal ventricle and may protrude through the thyrohyoid membrane as a mass anterior to the sternocleidomastoid muscle. Occasionally, an internal component is also present as well and this can impinge on the airway. Therapy is surgical excision.

Inflammatory conditions

Any infection of the upper respiratory tract or teeth can be associated with cervical lymphadenitis and can evolve such that a neck mass is the predominant clinical manifestation at some point in the course of the illness. A putative causative agent, or agents, can be suspected on the basis of associated clinical findings and sometimes confirmed by cultures of the appropriate site or blood, or, if an abscess forms, by culture and sensitivity testing of an aspirate. In the immunocompromised host, all pathogens should be considered. When an abscess is suspected, a CT scan may be of value since the presence of fluctuance on physical examination tends to occur as a late manifestation. Incision and drainage is generally required, particularly if compromise of the airway or great vessels of the neck is likely to occur.

In the adolescent, infectious mononucleosis may begin with diffuse adenopathy and associated pharyngitis, gingival hyperplasia, and, occasionally, unilateral tonsillar hypertrophy or tonsillitis. The diagnosis can be confirmed by the presence of more than 10 per cent atypical lymphocytes in a white blood cell count, and the presence of antibodies against Epstein–Barr virus. Infections by *Toxoplasma gondii* and cytomegalovirus may have a similar clinical picture; the detection of specific antibodies is the mainstay of differential diagnosis. Cat scratch disease, a usually self-limited infection manifesting as lymphadenopathy, generally requires a tissue biopsy for definitive diagnosis.

Chronic granulomatous diseases are endemic in various parts of the world and are making a resurgence in others. In particular, cervical tuberculosis (scrofula) and non-tubercular mycobacterial infection should be considered in young adults with a solitary painless neck mass. Pulmonary disease does not need to be active for the cervical lymph nodes to become involved, but appropriate skin testing is often indicative. A neck mass occurs in two-thirds of patients with extrapulmonary lymphadenitis due to *Mycobacterium tuberculosis*. Histoplasmosis, coccidiodomycosis, and actinomycosis can also produce cervical lymphadenopathy as a dominant feature.

Salivary gland inflammation or infection of either the parotid gland or the submandibular glands can produce a neck mass, usually related to some form of obstruction of the salivary duct (possibly by a stone) and secondary bacterial infection. The presence of swelling, pain, tenderness, and purulence expressed from the duct help suggest the diagnosis. Obstruction alone can result in a less dramatic clinical picture, sometimes entailing a waxing and waning mass in the respec-tive gland. Neoplastic disease should be suspected when the physical findings of inflammation are absent and the mass in question is more discrete, firmer, or, in particular with malignancies, when neurologic defects are present. A mass in the tail of the parotid gland can be difficult to distinguish from an upper cervical lymph node, and neither can invariably be separated from the angle of the mandible by palpation.

Sarcoidosis often affects the mediastinal and tracheal lymph nodes but dominant, single cervical adenopathy can be an associated, presenting, or sole site of the condition. These enlarged neck nodes are often non-tender and discrete, measuring up to 3 cm in diameter, and tending to resolve spontaneously.

Reactive lymphoid hyperplasia may also be caused by sinus histiocytosis with massive lymphadenopathy, a disorder that affects young adults or children. This causes enlarged lymph nodes, often in the neck and in extranodal sites such as the nasal cavity, tongue, oral mucosa, salivary glands, pharynx, trachea, and orbit. Sinus histiocytosis with massive lymphadenopathy is a self-limited disorder which eventually regresses. Lymphoid hyperplasia can be induced by the hydantoin family of anticonvulsant drugs, resolving upon cessation of the medication. These enlarged nodes may mimic lymphoma histologically, but are easily distinguished as benign by the flow cytometry and surface-marker analysis that is currently routine in the pathologic analysis of lymph nodes.

Other benign conditions that may present as masses in the neck are benign tumors, including lipomas, schwannomas, neurofibromas, thyroid adenomas, and paragangliomas. Aneurysms and other vascular abnormalities, foreign bodies, amyloidosis, and lymphangiomas may also present as a mass in the neck.

Malignancies

Hayes Martin, one of the century's foremost head and neck cancer surgeons, said that ‘asymptomatic enlargement of one or more cervical nodes in the adult is almost always cancerous'. Midline masses are almost always associated with thyroid disorders and usually represent a benign process or low-grade malignancy. Some masses are inflammatory and related to an identifiable infectious process. However, a firm mass over 2 cm in diameter demands further investigation and must be assumed to be malignant until proved otherwise.

Excluding those arising from the thyroid gland, nearly 90 per cent of malignant neoplasms in the neck are epithelial in origin. Most malignant masses in the cervical region are metastatic to cervical lymph nodes or primary lymphomas. An exceedingly rare primary process seen in this area is a carcinoma arising from a branchial cleft cyst, a diagnosis occasionally mistakenly offered in the case of a cystic metastatic lymph node. Only a few cases have been reported which fulfill the criteria for primary cervical branchiogenic carcinoma. These include an epithelial neoplasm arising from a cyst wall, located in a region compatible for branchial tissue which is otherwise histologically compatible with the diagnosis, no identification of a primary (mucosal) cancer, no treatment with radiation therapy to the head and neck, and follow-up for 5 years without manifestation of a primary site of malignancy.

If one assumes that the neck mass under investigation is metastatic, the primary source will often be found during a thorough subspecialty examination of the head and neck. The location of the metastasis frequently suggests the likely site of the primary tumor (Fig. 4). Tumors of the oral cavity often metastasize to the submandibular triangle (level I), whereas tumors of the nasopharynx produce adenopathy in the posterior cervical region (level V). A metastatic mass in the upper half of the neck can be due to an ‘unknown primary' tumor arising from the nasopharynx, tongue base, hypopharynx, larynx, or tonsillar fossa. In the supraclavicular region, however, the primary neoplasm arises from below the clavicles in 75 per cent of cases, from a source in the lung, gastrointestinal tract, or genitourinary system.

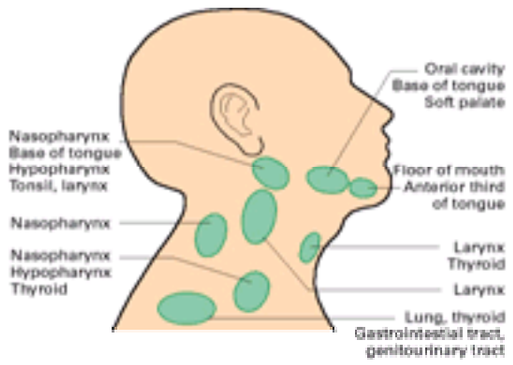


Fig. 4. The primary sites of cervical metastatic nodal disease can be suggested by the location of the neck mass.

Certain clues should raise the suspicion of a neoplastic process. Patients with a history of heavy smoking and alcohol ingestion, particularly men over 50 years of age, are at high risk. A history of prior excision of even small lesions of the head and neck mucosa and skin must be pursued. Certain areas such as the tonsil, the tongue base, the nasopharynx, and the piriform sinuses may contain very small, asymptomatic primary cancers that give rise to disproportionately large, single (and sometimes cystic) cervical metastases. However, the same sites that are traditionally incriminated as harboring so-called ‘unknown primaries', the tissues of Waldeyer's ring (tonsils, tongue base, and nasopharynx), are also potential sites for extranodal lymphomas.

A neck MRI or CT scan should generally be performed prior to any invasive procedure. A fine-needle aspiration biopsy using a 25-gauge needle has become a time-saving and critical asset in focusing further evaluation and treatment (Fig. 5). The FNA technique is simple and safe, and conventional epithelial tumors, which comprise over 90 per cent of head and neck mucosal malignancies, can be diagnosed accurately. A strong suspicion of lymphoma can also be con-firmed by an experienced cytologist. In no circumstance should a premature excisional biopsy of a neoplastic lymph node be undertaken. Such excision is traditionally thought to carry an inappropriate risk of wound necrosis, tumor recurrence, and distant metastasis. Open biopsies should only be performed at the time of a diagnostic endoscopy, when the surgeon is prepared to proceed with definitive surgical therapy on the neck, for instance, by radical neck dissection, if metastatic squamous cell carcinoma is confirmed in the absence of an identifiable primary site of disease.

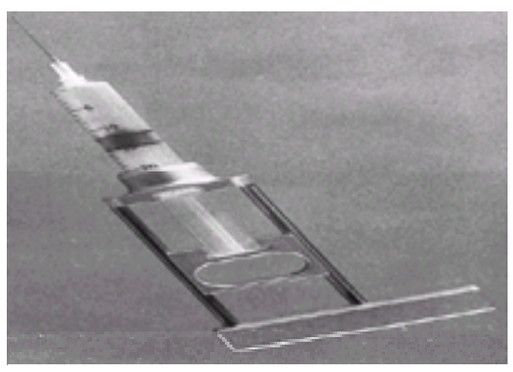


Fig. 5. A fine-needle (25-gauge) aspiration device using a 20-ml syringe enables the physician to perform aspiration with one hand while he stabilizes the neck mass with the other. (Handle: Cameco Syringe Pistol, Enebyberg, Sweden).

If a primary head and neck neoplasm is suspected by examination and by a positive cytologic aspiration of the neck mass, a complete endoscopic examination should be performed under anesthesia. This should include examination of the nasopharynx, hypopharynx, and larynx. Esophagoscopy is more accurate in the detection of esophageal disease than is a barium swallow. Although bronchoscopy with washings can be performed, its diagnostic yield is low if a routine chest radiograph is normal. Biopsies should be taken from any suspicious areas seen on endoscopy. If none is found, ipsilateral ‘directed' biopsies of the risk areas indicated by the location of the metastatic mass may be warranted. If a primary neoplasm is found, a staging diagram should depict the primary as well as metastasis (Fig. 6). The tumor should be staged according to established criteria, and concomitant second primaries should be sought, since they occur in up to 15 per cent of patients with head and neck tumors.

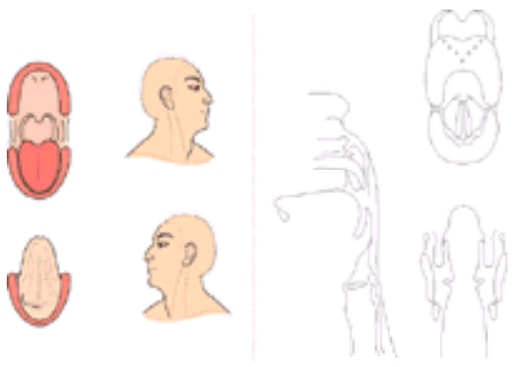


Fig. 6. This diagram format is an example of several that may be used to stage and document both the primary cancer and the location and size of the nodal disease in the head and neck region. Staging is determined according to the American Joint Committee on Cancer Staging.

Treatment of neoplastic disease

No single treatment is applicable to all patients. If a thorough search for a primary tumor of the head and neck is unsuccessful, an excisional biopsy can be performed through an incision that could be incorporated into a radical or modified radical neck dissection. If metastatic squamous cell carcinoma is confirmed, the options exist either to complete a neck dissection, modified appropriately for the site, or to close the incision and treat with radiation therapy if no primary site is discovered. In this setting, radiation treatment is delivered to all potential primary sites, including the nasopharynx, the base of the tongue, and the piriform sinuses, as well as to the neck. Generally, a neck with large-volume, multiple, or multilevel metastates would warrant surgery, followed by radiation, while an isolated, less than 3-cm metastasis high in the neck might reasonably be treated with excision alone and radiation as described. Repeated examinations performed during the course of radiation therapy may disclose the primary site as an area of unusually severe reaction.

Patients presenting with masses greater than 3 cm in diameter have a low survival rate, as do those with metastatic adenocarcinoma arising from a site other than the head and neck. In the current era, strategies incorporating chemotherapy, either alone as a preliminary treatment or combined concomitantly with radiation, are popular for advanced squamous cell carcinomas of the head and neck, including those presenting as large-volume metastasis with an unknown primary site.

Adenocarcinomas may arise from the major or minor salivary glands. If identified, these regions must be resected at the time of neck dissection. Metastatic adenocarcinoma arising from the thyroid gland, particularly that from a well-differentiated thyroid neoplasm, often carries a good prognosis and should be treated along the lines of traditional thyroid algorithms.

Further reading

Adams GL. Decisions and management of metastatic cancer with an unknown primary site. In: McQuarrie DG, Adams GL, Shors AR, Browne GA, eds, *Head and neck cancer—clinical decisions and management principles*, pp. 441–8. Yearbook, Chicago, 1985.

Agaton-Bonilla FC, Gay-Escoda C. Diagnosis and treatment of branchial cleft cysts and fistulae. A retrospective study of 183 patients. *International Journal of Oral and Maxillofacial Surgery* 1996; **25**: 449–52.

Ahuja A, Ng CF, King W, Metreweli C. Solitary cystic nodal metastasis from occult papillary carcinoma of the thyroid mimicking a branchial cyst: a potential pitfall. *Clinical Radiology* 1998; **53**: 61–3.

Bataini JP, Rodriquez J, Jaullery C, Brugere J, Glossein VA. Treatment of metastatic neck nodes secondary to an occult epidermoid carcinoma of the head and neck. *Laryngoscope* 1987; **97**: 1080–4.

Batsakis JG. The pathology of head and neck tumors: the occult primary and metastases to the head and neck. *Head and Neck Surgery* 1981; **3**: 409–23.

Conley J. How to examine the oro-naso-laryngo-pharynx for cancer. *Journal of the American Medical Association* 1971; **215**: 456.

Damion J, Hybels R. The neck mass: general concepts and congenital causes. *Postgraduate Medicine* 1987; **81**: 75–6, 81–8, 93.

Damion J, Hybels R. The neck mass: inflammatory and neoplastic causes. *Postgraduate Medicine* 1987, **81**: 97–103, 106–7.

Davidson BJ, Morris MS, Kornblut AD, Macher AM. Lymphadenopathy in the HIV-seropositive patient. *Ear, Nose, and Throat Journal* 1990; **69**: 478–80, 483–6.

deBraud F *et al.* Metastatic squamous cell carcinoma of an unknown primary localized to the neck. Advantages of an aggressive treatment. *Cancer* 1989; **64**: 510–15.

Finn DG. Lymphoma of the head and neck and acquired immunodeficiency syndrome: clinical investigation and immunohistological study. *Laryngoscope* 1995; **105**: 1–18.

Foucar E, Rosai J, Dorfman R. Sinus histiocytosis with massive lymphadenopathy (Rosai–Dorfman disease): review of the entity. *Seminars in Diagnostic Pathology* 1990; **7**: 19–73.

Frable MAS, Frable WJ. Fine-needle aspiration biopsy revisited. *Laryngoscope* 1982; **92**: 1414–18.

Fried MP. Evaluation of the adult neck mass. *Medical Times* 1982; January.

Fried MP, Diehl W, Brownson R, Sessions DG, Ogura JH. Cervical metastasis from an unknown primary. *Annals of Otology, Rhinology, and Laryngology* 1975; **84**: 152–7.

Ghaneim A, Atkins P. The management of thyroglossal duct cysts. *International Journal of Clinical Practice* 1997; **51**: 512–13.

Godin DA, Guarisco JL. Cystic hygromas of the head and neck. *Journal of the Louisiana State Medical Society* 1997; **149**: 224–8.

Greinwald JH Jr, Leichtman LG, Simko EJ. Hereditary thyroglossal duct cysts. *Archives of Otolaryngology—Head and Neck Surgery* 1996; **122**: 1094–6.

Harnsberger HR *et al.* Non-Hodgkin's lymphoma of the head and neck: CT evaluation of nodal and extranodal sites. *American Journal of Roentgenology* 1987; **149**: 785–91.

Kempf HG, Becker G, Weber BP, Ruck P, Lenarz T. Diagnostic and clinical outcome of neurogenic tumours in the head and neck area. *ORL; Journal of Oto-Rhino-Laryngology and Related Specialties* 1995; **57**: 273–8.

Komp DM. The treatment of sinus histiocytosis with massive lymphadenopathy (Rosai–Dorfman disease). *Seminars in Diagnostic Pathology* 1990; **7**: 83–6.

Mancuso AA, Hanafee WN. Elusive head and neck carcinomas beneath intact mucosa. *Laryngoscope* 1983; **93**: 133–9.

Martin H, Morfit HM. Cervical lymph node metastasis as the first symptom of cancer. *Surgery, Gynecology, and Obstetrics* 1944; **78**: 133–7.

Martin H, Romieu C. The diagnostic significance of a ‘lump in the neck'. *Postgraduate Medicine* 1952; **11**: 491–500.

McQuirt WF, McCabe BF. Significance of node biopsy before definitive treatment of cervical metastatic carcinomas. *Laryngoscope* 1978; **88**: 594–7.

Mitchell RO, Richardson JD, Lambert GE. Characteristics, surgical management, and outcome in 17 carotid body tumors. *American Surgeon* 1996; **62**: 1034–7.

Miyamoto RT, Roos KL, Campbell RL, Worth RM. Contemporary management of neurofibromatosis. *Annals of Otology, Rhinology, and Laryngology* 1991; **100**: 38–43.

Molinari R. A statistical approach to detection of the primary cancer based on the site of neck lymph node metastases. *Tumori* 1977; **63**: 267–82.

Neel HB III, Pearson GR, Weiland LH, Taylor WF, Goepfert HH. Immunologic detection of occult primary cancer of the head and neck. *Otolaryngology—Head and Neck Surgery* 1981; **89**: 230–4.

Patt BS, Schaefer SD, Vuitch F. Role of fine-needle aspiration in the evaluation of neck masses. *Medical Clinics of North America* 1993; **77**: 611–23.

Penfold CN, Revington PJ. A review of 23 patients with tuberculosis of the head and neck. *British Journal of Oral and Maxillofacial Surgery* 1996; **34**: 508–10.

Rafaty FM. Cervical adenopathy secondary to toxoplasmosis. *Archives of Otolaryngology* 1977; **103**: 547–9.

Rajagopalan K, Adelstein DJ, Lavertu P. Extra-nodal non-Hodgkin's lymphoma of the head and neck (ELHN): a 20 year review. *Proceedings of the Annual Meeting of the American Society of Clinical Oncology* 1996; **15**: A1302.

Regauer S, Gogg-Kamerer M, Braun H, Beham A. Lateral neck cysts—the branchial theory revisited. A critical review and clinicopathological study of 97 cases with special emphasis on cytokeratin expression. *APMIS* 1997; **105**: 623–30.

Schwartz R, Chan NH, MacFarlane JK. Fine needle aspiration cytology in the evaluation of head and neck masses. *American Journal of Surgery* 1990; **159**: 482–5.

Shedd DP, Arias JD, Glunk RP. Familial occurrence of carotid body tumors. *Head and Neck* 1990; **12**: 496–9.

Slaughter DP, Majarkis JD, Southwick HW. Clinical evaluation of swellings in the neck. *Surgical Clinics of North America* 1956; **36**: 3–28.

Sobol SM, Dailey JC. Familial multiple cervical paragangliomas: report of a kindred and review of the literature. *Otolaryngology—Head and Neck Surgery* 1990; **102**: 382–90.

Stewart MG, Schwartz MR, Alford BR. Atypical and malignant lipomatous lesions of the head and neck. *Archives of Otolaryngology—Head and Neck Surgery* 1994; **120**: 1151–5.

Tuzuner N, Dogusoy G, Demirkesen C, Ozkan F, Altas K. Value of lymph node biopsy in the diagnosis of acquired toxoplasmosis. *Journal of Laryngology and Otology* 1996; **110**: 348–52.

Weaver DK. Atypical lymphadenopathies of the head and neck. *Critical Reviews in Clinical Laboratory Sciences* 1981; **15**: 1–24.

Winegar LK, Griffin W. The occult primary tumor. *Archives of Otolaryngology* 1973: **98**: 159–63.

45.4 Salivary glands

Joseph Califano and David W. Eisele

[Anatomy and physiology](#)
[Evaluation](#)
[Bacterial sialadenitis](#)
[Viral salivary gland disease](#)
[Granulomatous disease](#)
[Sjögren's syndrome](#)
[Sialadenosis](#)
[Trauma](#)
[Acquired and congenital cysts](#)
[Salivary gland neoplasms](#)
[General considerations](#)
[Benign neoplasms](#)
[Malignant neoplasms](#)
[Therapy for salivary gland neoplasms](#)
[Further reading](#)

The evaluation and treatment of salivary gland disease requires knowledge of the varied pathologic processes that affect the salivary glands in particular, and knowledge of general systemic disease states that affect the salivary glands. Treatment of salivary gland pathology often requires surgical management, therefore it is imperative that a proper head and neck surgical evaluation be performed on most forms of salivary gland pathology.

In this chapter, basic salivary gland anatomy and physiology, specific disease processes, and methods of evaluation and treatment of the salivary gland pathology will be presented.

Anatomy and physiology

The salivary glands consist of the paired, major salivary glands, including the parotid, submandibular, and sublingual glands. In addition, approximately 600 to 1000 small, minor salivary glands are distributed thoughout the upper aerodigestive tract.

The parotid gland is the largest salivary gland, located beneath the preauricular skin and subcutaneous tissues. It is composed almost entirely of serous acinar cells and accounts for the majority of salivary flow in an active state, but only provides approximately 25 per cent of the total volume of saliva. It is bounded posteriorly by the external auditory canal, inferiorly by the styloid process and the styloid musculature, and superiorly by the zygoma. Portions of the gland extend anteriorly over the masseter muscle, posteriorly overlapping the sternocleidomastoid muscle, and can extend medially into the parapharyngeal space. The facial nerve exits from the stylomastoid foramen, turns anterolaterally to enter the parotid gland, and branches at the *pes anserinus*, dividing into the frontal, zygomatic, buccal, marginal mandibular, and cervical branches that innervate the mimetic facial musculature. The facial nerve serves to divide the gland into a superficial and deep lobe, which are not embryologically or anatomically distinct lobes but are produced by surgical dissection. The parotid gland is drained by Stensen's duct, which exits from the anterior border of the gland, 1.5 cm below the zygoma. The duct courses anteriorly and superficial to the masseter muscle, turning medially to enter through the buccinator muscle and opens intraorally in the buccal mucosa opposite the second maxillary molar.

The submandibular gland, composed of both mucous and serous secretory cells, is the second largest salivary gland, supplying approximately 70 per cent of the total volume of saliva. It is located in the submandibular triangle, bounded inferiorly by the anterior and posterior bellies of the digastric muscle, superiorly by the body of the mandible and the floor of mouth, and superficially by the facial vein and marginal mandibular branch of the facial nerve. It extends anteriorly with portions that extend superficial and deep to the mylohyoid muscle. Wharton's duct drains the gland, exiting medially before traveling anteriorly between the mylohyoid and hyoglossus muscles onto the genioglossus muscle and deep to the sublingual gland, eventually opening a few millimeters lateral to the midline lingual frenulum in the anterior floor of the mouth.

The sublingual gland is the smallest of the major salivary glands, produces predominantly mucous secretions, and lies just beneath the mucosa of the anterior floor of the mouth. Approximately 10 small ducts exit the superior aspect of the gland directly into the oral cavity. The gland is bordered laterally by the mandible and genioglossus, inferiorly by the mylohyoid muscle, with the lingual nerve and Wharton's duct traveling between the genioglossus and the sublingual gland.

The minor salivary glands, approximately 600 to 1000 in number, are simple, submucosal glands, composed of predominantly mucous secretory cells. They are located on practically all surfaces of the oral cavity, as well as the nasopharynx, oropharynx, and other upper aerodigestive mucosal surfaces. They account for a small percentage of total saliva production.

Evaluation

As with other organ systems, salivary gland disease diagnosis and evaluation relies on an adequate history and physical examination, supplemented by appropriate laboratory and imaging studies.

Pain, fever, hypersecretion, hyposecretion, nerve disfunction, purulent discharge, aberrant taste sensation, swallowing and voice dysfunction, gland enlargement, trismus, and evidence of mass in extrasalivary head and neck sites are important factors. Any evidence of concurrent systemic disease, in particular infectious (human immunodeficiency virus, or HIV), autoimmune, or lymphatic disease (lymphoma, Sjögren's syndrome) is of particular interest.

A complete examination of the head and neck, including otoscopy, rhinoscopy, indirect laryngoscopy, examination of all mucosal surfaces, and bimanual palpation of head and neck structures as well as a complete cranial nerve examination is needed. All major salivary glands should be palpated bimanually with one hand in the oral cavity to define suspected masses precisely. Sialolithiasis may be detected by palpation of stones in Wharton's or Stensen's duct. Salivary gland secretions may be milked from parotid and submandibular glands, and their character noted (cloudy, purulent, bloody). Particular attention is paid to facial, lingual, and hypoglossal nerve function, because of their proximity to the major salivary glands.

Laboratory studies may be obtained to evaluate systemic diseases with salivary gland manifestations, including immunologic (e.g. Sjögren's syndrome), infectious (e.g. tuberculosis, cat scratch, HIV), and other diseases.

Fine needle aspiration may be used to evaluate suspected neoplasms or abscess. Plain film radiography may be used to demonstrate salivary calculi. Sialography, once the mainstay of salivary gland imaging, is now only rarely used. Computed tomography (**CT**) or magnetic resonance imaging (**MRI**) of the neck with intravenous contrast may be used to define the location of salivary gland masses, define masses as cystic or solid, look for evidence of invasion of adjacent structures, and define significant cervical lymphadenopathy associated with infectious and neoplastic processes.

Bacterial sialadenitis

Bacterial infections of the major salivary glands are typically the result of mechanical blockage of salivary ducts or due to stasis of salivary secretions. Historically, this has occurred infrequently in patients in whom dehydration and salivary stasis has resulted from prologed anorexia and uncompensated fluid losses in the postoperative period. Subsequent retrograde migration of intraoral bacteria results in acute bacterial sialadenitis. Sialolithiasis, the formation of calculi in the ductal system of salivary glands, may result in blockage of salivary flow and predispose to episodes of acute and chronic sialadenitis in a non-debilitated patient population.

Symptoms of acute bacterial sialadenitis include pain, swelling, tenderness, and induration of the involved gland with erythema of the overlying skin or oral mucosa. Fever and leukocytosis are also commonly present. The involved gland or glands may be milked to yield purulent discharge from the duct orifice that may be cultured, and a stone in the duct of the involved gland may or may not be palpable. Plain radiography may be able to confirm the presence of a salivary stone (Fig. 1). Etiologic bacterial species include *Staphylococcus aureus, Streptococcus* spp., *Haemophilus* spp., and anaerobic species. Many of these isolates produce b-lactamase.

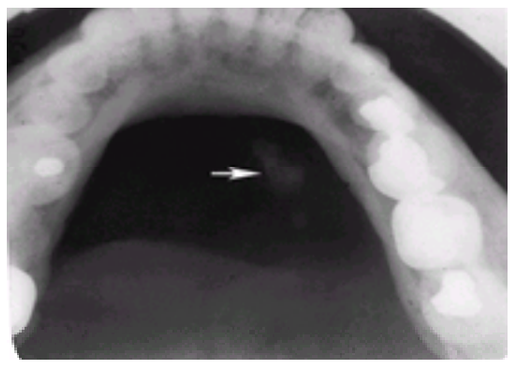


Fig. 1. A plain radiograph demonstrating a salivary calculus (arrow) in Wharton's duct.

Initial management includes administration of appropriate parenteral or oral antibiotics, adequate hydration, sialagogues, oral hygiene, warm compresses, and bimanual massage to milk purulent drainage from the gland. If a stone is located close to a distal duct orifice, it may be expressed manually, or removed after enlarging the duct orifice with a small incision placed over a lacrimal probe which is used to cannulate the duct (Fig. 2).



Fig. 2. Salivary gland calculus protruding from Stensen's duct (arrow).

Failure to respond to the above therapy may indicate abscess formation, which can be confirmed by CT or ultrasonography. Surgical drainage is indicated if an abscess is present. Drainage of the parotid gland is accomplished by elevation of a standard anteriorly based flap over the superficial parotid fascia, incision of the parotid fascia parallel to the branches of the facial nerve, blunt dissection in the gland parenchyma parallel to the direction of the facial nerve branches to drain any loculated abscess, and placement of an external drain.

Chronic sialadenitis is characterized by repeated episodes of salivary gland pain and inflammation, with or without overt, acute bacterial infection. Obstruction of salivary flow from ductal stricture, mucous plugging, or calculi is the usual etiology of this entity. A vicious cycle of obstruction, followed by inflammation, followed by worsening obstruction is characteristic of the clinical course. To date, the only reliable method of treatment of chronic sialadenitis is surgical removal of the offending gland.

Viral salivary gland disease

Mumps is caused by a paramyxovirus and is the most common cause of viral acute sialadenitis. It is characterized by bilateral parotid gland swelling and pain, low-grade fever, arthralgia, malaise, and headache, and occurs predominantly in children. Diagnosis is usually made on clinical grounds, but can be confirmed by hemagglutination inhibition and complement fixation tests. Treatment is by supportive measures.

Infection with the HIV virus is associated with lymphoproliferative disorders that may affect the salivary glands, and with cystic enlargement of the major salivary glands. Gradual, painless enlargement of one or more salivary glands is noted. These lesions usually appear as epithelial-lined simple cysts (Fig. 3), but may also be HIV-associated lymphoma. Fine needle aspiration or imaging studies such as CT or MRI that are consistent with a benign cyst can allow management of these lesions by conservative, clinical observation. Evidence of lymphoma or other neoplasm requires surgical management for diagnostic and therapeutic purposes.

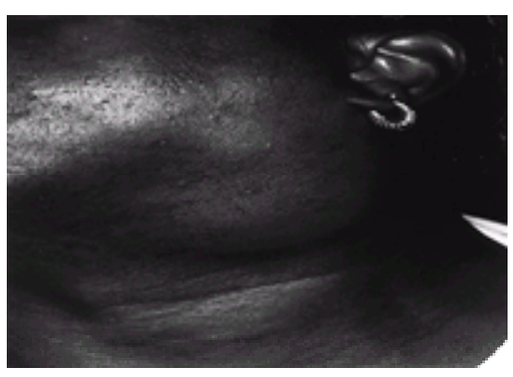


Fig. 3. Typical clinical appearance of an HIV-associated parotid cyst.

Granulomatous disease

Chronic granulomatous diseases can affect the salivary glands via the lymphatic network intrinsic to the parenchyma of the parotid gland and the lymphatics adjacent to all of the major salivary glands.

Tuberculous mycobacterial infection of the salivary glands has been uncommon in the past, but the incidence of extrapulmonary tuberculosis has been increasing, in part due to the contribution of tuberculosis associated with HIV infection. Infection of the salivary glands by *Mycobacterium tuberculosis* may occur as a primary infection or as a secondary infection related to a primary pulmonary focus. The parotid gland may be diffusely involved or may be involved with a solid, slow growing, painless mass. Most cases of parotid tuberculosis are diagnosed after parotidectomy to remove an infected mass, despite attempts at diagnosis by fine needle aspiration.

Atypical (non-tuberculous) mycobacterial infections of the salivary glands are commonly seen in children from 1 to 3 years of age, presenting as a mass with a rapid increase in size. When mature, these will often erode into the skin and drain spontaneously. Pulmonary and constitutional symptoms are usually absent. Diagnosis is performed based on clinical presentation, and treatment consists of excision and/or curettage of involved tissue, with simultaneous culture of purulent cyst contents to confirm diagnosis.

Actinomyces include many species of Gram-positive anaerobic bacteria, commonly found in the oropharynx. Poor oral hygiene, recent oral trauma, and an immunosuppressive state are associated with infection. Clinical presentation includes painless induration and enlargement of an involved gland, without constitutional

symptoms. Lymph node necrosis may lead to skin erosion and spontaneous drainage. Diagnosis is confirmed by needle aspiration or swab that demonstrates the presence of sulfur granules and filamentous Gram-negative rods. Treatment consists of penicillin with surgery reserved for diagnosis.

Cat-scratch disease is a granulomatous lymphadenitis resulting from inoculation via skin trauma from a domestic cat. Systemic symptoms are usually absent, but can be severe. Diagnosis is usually made by history and clinical presentation in the context of a sterile abscess culture. Fine needle aspiration or exisional biopsy may aid in diagnosis. A typical histopathologic appearance is demonstrated in involved lymph nodes. Treatment includes supportive measures.

Toxoplasmosis is caused by a parasite found in undercooked meat or cat feces, resulting in a disseminated form of disease and a form resulting in localized lymphadenopathy. Diagnosis is usually made by measurement of acute and convalescent titers. Tularemia is caused by a Gram-negative organism, *Francisella tularensis*, transmitted by insect vectors and other routes. A facial insect bite results in a local erythematous papule, followed by local lymphadenopathy and constitutional symptoms. Diagnosis is made by serologic titers. Aggressive, early drainage of involved lymph nodes in the acute stages of infection may result in systemic dissemination of disease. Following appropriate antibiotic therapy, remaining cysts may be incised or drained.

Sjögren's syndrome

Recurrent parotid gland swelling associated with keratoconjunctivitis sicca and xerostomia are the hallmarks of primary Sjögren's syndrome, a connective tissue disease thought to be of autoimmune origin. When associated with another connective tissue disorder, it is termed secondary Sjögren's syndrome. Diagnosis may be aided by serologic studies, but minor salivary gland excisional biopsy is usually performed to provide for histopathologic analysis to demonstrate characteristic lymphocytic infiltration of the gland. Approximately 5 per cent of patients with Sjögren's syndrome eventually develop a lymphoproliferative neoplasm.

Sialadenosis

Sialadenosis describes recurrent, bilateral, non-tender parotid swelling, usually associated with an underlying disease state or nutritional deficiency, or pharmacologic side-effect. Cirrhosis, diabetes, alcoholism, malnutrition, bulimia, ovarian insufficiency, hypothyroidism, pancreatic insufficiency, and multiple drugs may cause sialadenosis.

Trauma

Management of salivary gland trauma comprises treatment of the injured glands and ductal structures of the major salivary gland themselves, and management of trauma to important structures anatomically related to the glands. Physical examination must evaluate the integrity of the glands, major salivary ducts, major vascular structures, and associated cranial nerves, including the facial, hypoglossal, and lingual nerves, as well as all major head and neck structures. Injuries to the facial nerve lateral to an imaginary line drawn between the lateral canthus and the mental foramen should be explored and surgically repaired with a tension-free primary reanastomosis or grafting of severed or damaged nerve fibers. Injuries to the facial nerve medial to this line are not usually treated by neural anastomosis. The ability to stimulate distal, severed nerve endings electrically is retained for approximately 36 h after injury. Stensen's duct injuries may be repaired using microvascular techniques over a Silastic stent to prevent subsequent stricture.

Failure to repair ductal injuries may result in post-traumatic salivary fistula or sialocele. Management consists of repeated aspiration and compression, with further surgical or pharmacologic therapy reserved for persistent complications.

Acquired and congenital cysts

Congenital cysts of the first branchial cleft may occur as a duplication of the membranous (type I) or membranous and cartilaginous (type II) external auditory canal. These branchial cleft cysts present as preauricular or parotid masses, may be intimately associated with the parotid gland and facial nerve, and usually mandate identification of the facial nerve and superficial parotidectomy to facilitate complete excision in a safe manner.

Acquired cysts of the submandibular and sublingual glands may present as a translucent, cystic mass in the floor of the mouth (ranula, Fig. 4), or they may extend inferiorly and present as a neck mass (plunging ranula). Excision of the entire cyst as well as the entire associated major salivary gland, usually the sublingual, is necessary to ensure a low risk of recurrence.



Fig. 4. A ranula in the floor of the mouth arising from the sublingual gland.

Salivary gland neoplasms

General considerations

Salivary gland neoplasms are a diverse group of tumors, and appropriate treatment depends on an adequate understanding of the pathophysiologic behavior of each tumor type.

Approximately 80 per cent of salivary gland neoplasms originate in the parotid gland (Fig. 5), 15 per cent develop in the submandibular gland (Fig. 6), and the remaining tumors arise in the sublingual and minor salivary glands (Fig. 7). Eighty per cent of parotid neoplasms are benign, 50 per cent of submandibular neoplasms are benign, and fewer than 40 per cent of sublingual and minor salivary gland tumors are benign. Most salivary gland neoplasms (95 per cent) occur in adults. The most common benign pediatric salivary gland tumor is the hemangioma, followed by the pleomorphic adenoma; the most common malignant salivary gland tumor in children is the mucoepidermoid carcinoma.



Fig. 5. Clinical appearance of a typical parotid neoplasm (pleomorphic adenoma).

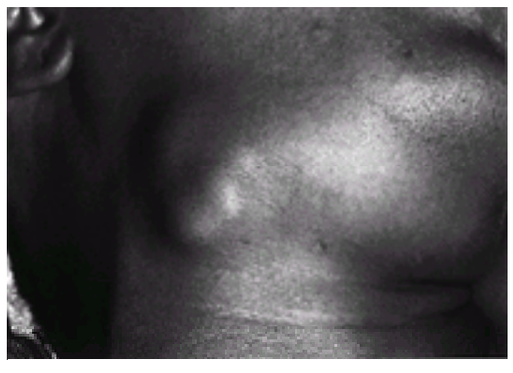


Fig. 6. A submandibular gland neoplasm (pleomorphic adenoma).



Fig. 7. A tumor of minor salivary gland origin (pleomorphic adenoma) presenting as a mass on the hard palate.

Most salivary gland neoplasms present as asymptomatic masses. Benign neoplasms typically enlarge slowly. Pain may occur with both benign and malignant neoplasms and is not an indicator of malignancy, but when found in association with malignancy there is a poorer prognosis. Minor salivary gland tumors may present as painless masses almost anywhere in the upper aerodigestive tract.

A thorough head and neck examination as outlined above should be performed. Special attention is paid to nerve function, fixation to local structures, and associated cervical adenopathy. Biopsy by fine needle aspiration may be used to aid in accurate diagnosis. CT and MRI scans do not differentiate benign from malignant tumors, and rarely alter the therapeutic approach to major salivary gland neoplasms. These imaging modalities are useful, however, for known malignant or recurrent neoplasms, large tumors, evaluation of parapharyngeal space involvement, suspected carotid artery involvement, or involvement of structures that suggest inoperability. Imaging is routine for assessment of minor salivary gland malignancies involving the nose or paranasal sinuses or other minor salivary gland tumors to assess location and extent of invasion.

Benign neoplasms

Pleomorphic adenomas comprise 65 per cent of all salivary gland neoplasms. They are most frequently found in the parotid gland. Although this tumor grossly has a well defined capsule, microscopically it demonstrates incomplete encapsulation and transcapsular growth with pseudopod extensions. This accounts for the high rate of recurrence after simple enucleation of these tumors. Pleomorphic adenomas must be excised by parotidectomy with an adequate margin of normal parotid tissue around the neoplasm to ensure complete resection. Resection is advocated because, if observed, pleomorphic adenomas invariably continue to enlarge, and may undergo malignant transformation (at a rate of approximately 5 per cent).

Warthin's tumor is the second most common benign neoplasm of the parotid gland and accounts for 6 to 10 per cent of all parotid tumors, but is rarely found in other salivary gland sites. Most tumors are found in older men, with an increased risk for occurrence in smokers. Warthin's tumors are often cystic, can occur multicentrically, and approximately 10 per cent occur bilaterally. These tumors rarely undergo malignant transformation and are usually treated by parotidectomy.

Oncocytomas occur almost exclusively in the parotid gland, comprise less than 1 per cent of all salivary gland neoplasms, and commonly occur in patients who are in the sixth decade of life. These tumors are solid and are removed by parotidectomy with facial nerve preservation.

Monomorphic adenomas describe a group of rare salivary gland neoplasms, including basal cell adenoma, clear cell adenoma, glycogen-rich adenoma, and other rare tumors. The most common of this group is the basal cell adenoma, which is commonly found in the minor salivary glands, followed by the parotid gland. The monomorphic adenomas are considered benign, non-aggressive neoplasms, and are treated by parotidectomy with a margin of normal tissue.

Malignant neoplasms

Mucoepidermoid carcinoma is the most common malignant tumor involving the parotid gland and the second most common malignant tumor of the submandibular gland, after adenoid cystic carcinoma. Between 6 and 9 per cent of all major salivary gland neoplasms are mucoepidermoid carcinomas and the majority are located in the parotid gland. These tumors are classified into low-grade and high-grade types. Low-grade tumors behave like benign neoplasms but may invade locally and metastasize. High-grade neoplasms are aggressive malignancies with a high propensity for metastasis, resemble squamous cell carcinomas histopathologically, and may require special mucin staining to differentiate them from squamous cell carcinoma.

Adenoid cystic carcinoma accounts for 6 per cent of all salivary gland neoplasms and is the most common malignancy of the submandibular gland and minor salivary glands. It usually presents as an asymptomatic mass in patients in the fifth decade of life. Perineural invasion is a typical feature of this tumor, accounting for the difficulty in eradication of this tumor despite wide surgical excision, and the tendency for late recurrence, sometimes after a decade or more of disease-free status. Treatment consists of surgical resection with adjuvant radiation therapy.

Acinic cell carcinoma represents 1 per cent of all salivary gland neoplasms and 2 to 4 per cent of parotid gland neoplasms. Almost all acinic cell carcinomas (95 per cent) are found in the parotid gland and are rarely found in the submandibular gland. Patients are usually women in the fifth decade of life. Bilateral involvement occurs approximately 3 per cent of the time. At early follow-up intervals, acinic cell carcinoma exhibits a benign course, but 20-year follow-up studies have shown a 50 per cent survival rate.

Adenocarcinoma occurs most commonly in the minor salivary glands, followed by the parotid gland, and represents 15 per cent of malignant parotid neoplasms. These are aggressive tumors that are likely to recur and metastasize.

Polymorphous low-grade adenocarcinoma is the second most common malignancy of the minor salivary glands, presenting as firm, painless masses on the palate, buccal mucosa, and upper lip during the sixth decade of life. This neoplasm is characterized by perineural spread. Treatment consists of complete resection, which results in a favorable prognosis.

Carcinoma ex-pleomorphic adenoma describes a malignant tumor that has arisen from a pre-existing pleomorphic adenoma, and represents 2 to 5 per cent of salivary gland tumors. This tumor presents as a slow growing mass that has been present for 10 to 15 years which suddenly increases in size. Local and distant metastases are common for this tumor type, and it is associated with a very poor prognosis. Surgery followed by postoperative radiation is recommended.

Squamous cell carcinoma is a rare malignancy of the salivary glands that constitutes less than 2 per cent of salivary gland tumors and occurs most commonly in the submandibular gland. Diagnosis of a true, primary squamous cell carcinoma of the salivary glands requires exclusion of contiguous spread of an adjacent squamous cell carcinoma into the gland, metastases to the gland from a cutaneous primary, or a high-grade mucoepidermoid carcinoma. These tumors present in the seventh decade of life and are more common in males. They are associated with a high rate of regional and distant metastasis, and surgical excision followed by radiation therapy is recommended.

Undifferentiated carcinoma accounts for 3 per cent of all tumors to the major salivary glands and 1 to 5 per cent of all parotid malignancies and occurs late in life. This malignancy is extremely aggressive, demonstrates local invasion, early distant metastasis, and low survival rates.

Sarcomas of the salivary glands are rare. Diagnosis requires exclusion of metastatic spread of a sarcoma from another primary site and invasion of the gland from an adjacent soft tissue sarcoma. Behavior of these tumors is analogous to that of sarcomas found in other anatomic sites, depending on size, tumor type, and grade.

Primary lymphoma of the salivary glands is rare. This diagnosis requires proof of no known extrasalivary lymphoma, proof that the lymphoma arose from salivary gland parenchyma, and standard pathologic confirmation of the diagnosis. The prognosis is usually good for this neoplasm and surgical intervention is usually reserved for diagnosis.

Therapy for salivary gland neoplasms

A mass in the region of the parotid gland would be considered a neoplasm until proved otherwise. Surgical incisional biopsy, excisional biopsy, or tumor enucleation of parotid neoplasms is to be avoided, due to the risks of tumor spillage, tumor recurrence, or facial nerve injury. The proper approach to most parotid neoplasms is to perform a parotidectomy, with identification and preservation of the facial nerve and its branches, so that tumor is removed with an adequate margin of normal tissue. Superficial parotidectomy is curative for most benign parotid neoplasms and low-grade malignant neoplasms. Total parotidectomy, in which both superfical and deep lobes are resected, is recommended for high-grade malignancies. The facial nerve is preserved unless grossly invaded with tumor. If resected, the facial nerve is immediately reconstructed with a nerve graft intraoperatively.

A standard parotidectomy incision is shown in Fig. 8(a). The portion of the incision that extends over the mastoid tip inferior to the lobule of the pinna should curve gently, so as not to create a narrow, avascular portion of the flap that may necrose. The skin flap is dissected in a plane immediately superficial to the parotid fascia. The greater auricular nerve is usually divided so that the tail of the parotid gland may be dissected free from the sternocleidomastoid muscle, the mastoid tip, and the auricular cartilage. Theposterior belly of the digastric muscle is identified deep to the anterior border of the sternocleidomastoid muscle, and a second plane of dissection is developed in the pretragal space using blunt dissection (Fig. 8(b)). With the knowlege that the facial nerve exits the stylomastoid foramen 1 cm medial and inferior to the tip of the tragal cartilage ‘pointer', the remaining bridge of parotid tissue lateral to the digastric muscle that remains attached to the sternocleidomastoid and pretragal cartilage is divided. This allows for wide exposure that is needed to identify the main trunk of the facial nerve. Blunt dissection with a fine hemostat parallel to the course of the facial nerve in the remaining tissue allows for identification of the facial nerve (Fig. 8(c)). Once the main trunk of the nerve is identified, individual branches are followed peripherally, elevating the superfical portion of the gland. This is performed by placing a small hemostat in the plane superficial to the nerve, elevating the hemostat laterally, gently spreading the tips of the hemostat, and then incising the tissue between the tips of the hemostat while maintaining direct visualization of the nerve. This is performed on successive branches until the entire superficial portion of the gland containing the mass is removed *en bloc* (Fig. 8(d)). A drain is placed and the wound is closed after meticulous hemostasis is obtained (Fig. 8(e)).

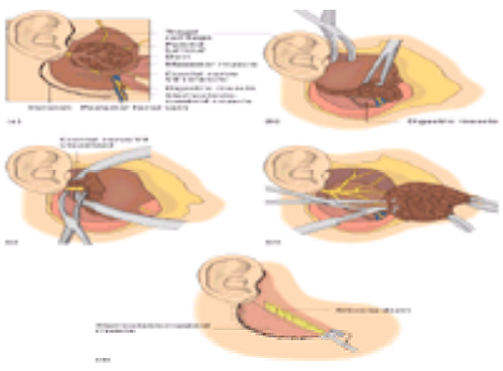


Fig. 8. The surgical technique for parotidectomy.

Neck dissection is indicated for clinically positive cervical metastases or if suspicious nodes found during the primary resection are positive on frozen section examination. There is no proven benefit for elective neck dissection for a clinically negative neck. Complications of parotidectomy include facial nerve injury, hematoma, infection, flap necrosis, salivary fistula, and Frey's syndrome. Frey's syndrome is present symptomatically in 10 per cent of patients after parotidectomy and is related to aberrant regeneration of severed postganglionic parasympathetic nerve fibers supply-ing the parotid gland. These fibers regenerate to innervate the postganglionic sympathetic nerves of the cutaneous sweat glands, resulting in gustatory sweating from the skin overlying the parotid bed.

Parapharyngeal space tumors may be removed via a submandibular approach, removing the submandibular gland, if necessary, to allow access to the anterior compartment of the parapharyngeal space. Neoplasms involving the submandibular gland are usually limited to the gland itself, and surgical resection is usually confined to the submandibular triangle. All nerves are preserved unless involved by tumor.

Radiation therapy in combination with surgery has improved locoregional control and survival for patients with carcinoma of the major salivary glands, and benefits patients with minor salivary gland malignancies who are at high risk for local failure. Chemotherapy has been shown to be of little benefit in the treatment of salivary gland malignancy and is usually reserved for palliation of unresectable, previously irradiated tumors.

Further reading

Batsakis JG. Tumors of the major salivary glands. In: Batsakis JG, ed., *Tumors of the head and neck, clinical and pathological considerations*, p. 1–75. Williams & 38; Wilkins, Baltimore, 1979. [Comprehensive coverage of pathology of the major salivary glands with clinicopathological correlations.]

Batsakis JG. Neoplasms of the minor and ‘lesser'major salivary glands. In: Batsakis JG, ed., *Tumors of the head and neck, clinical and pathological considerations*, p. 65–99. Williams & 38; Wilkins, Baltimore, 1979. [Comprehensive coverage of pathology of the minor salivary glands with clinicopathological correlations.]

Eisele DW, Johns ME. Salivary gland neoplasms. In: Bailey BJ, ed., *Head and neck surgery—otolaryngology*, 2nd edn, p. 1485–1508. Lippincott-Raven, Philadelphia, 1998. [Chapter covering the management of benign and malignant salivary gland neoplasms, emphasizing a thorough understanding of the pathology and behavior of each tumor type in order to guide appropriate treatment.]

Gayner SM, Kane WJ, McCaffrey TV. Infections of the salivary glands. In: Cummings CW, ed., *Otolaryngology—head and neck surgery*, 3rd edn, p. 1234–46. Mosby Year Book, St. Louis, 1998. [Chapter covering various infections of the salivary glands.]

Haller JR. Trauma. In: Cummings CW, ed., *Otolaryngology—head and neck surgery*, 3rd edn, p. 1247–54. Mosby Year Book, St. Louis, 1998. [Chapter devoted to the evaluation and management of trauma to the salivary glands and adjacent structures.]

Rice DH, Eisele DW, eds. Salivary gland diseases. *Otolaryngologic Clinics of North America* 1999; **32**: 779–964. [Volume covering non-neoplastic and neoplastic diseases of the salivary glands with emphasis on diagnostic evaluation, pathology, and treatment approaches.]

45.5 Oral and oropharyngeal cancer

C. M. Norris, Jr., James W. Lucarini and Marshall Posner

Anatomy
Epidemiology
Presentation
Oral cancer: clinical features by subsite
Oropharyngeal carcinoma: clinical features by subsite
Diagnosis and associated examinations
Staging
Pathology
Treatment
Chemotherapy
General prognosis
Further reading

Anatomy

By convention and widespread practice, the oral cavity includes the mucosa, submucosal adnexae (minor salivary glands, sublingual glands, ducts, regional nerves, and intrinsic muscles) and immediately adjacent bone and teeth of the mandible and maxilla, the mobile tongue, the floor of the mouth, the alveolar ridges, the buccal surfaces, the hard palate, the retromolar trigone (the mucosa apposed to the medial surface of the angle of the mandible), and the unexposed mucosa of the lips. The distinction between the oral cavity and oropharynx has not only anatomic but also clinical and prognostic significance in head-and-neck cancer. The oropharyngeal structures include the base of the tongue, the soft palate, the tonsillar pillars and fossa, and the pharyngeal walls from the level of the soft palate to the pharyngoepiglottic fold. The demarcation between the oral cavity and oropharynx is taken as the junction of the hard and soft palates, the circumvallate papillae of the tongue, and a vertical line just anterior to the anterior tonsillar pillars bilaterally. The oropharynx is demarcated from the nasopharynx by the horizontal plane of the hard palate, and from the supraglottic larynx by the pharyngoepiglottic folds laterally and the vallecula medially. Stensen's duct, draining the parotid gland, and Wharton's duct, draining the submandibular gland, empty through small papillas in the posterosuperior buccal area and anterior floor of mouth, respectively.

The mucosa of the oral cavity is made up of a stratified squamous epithelium with varying thicknesses of submucosa. Multiple minor salivary glands and taste buds are scattered throughout the mucosa and submucosa. Submucosal lymphoid aggregates, including the tubal, palatine (‘adenoids'), pharyngeal (‘tonsils'), and lingual tonsils, form a circular aggregate known as Waldeyer's ring.

The oral cavity and oropharynx are supplied by cranial nerves V, VII, IX, X, and XII. The predominant blood supply is from branches of the external carotid artery, especially the lingual and facial arteries. Venous drainage is via direct tributaries to the internal jugular vein. Lymphatic drainage is variable, generally more extensive in the posterior oral cavity than anteriorly, and most extensive in the oropharynx, an anatomic feature with direct oncologic, clinical, and therapeutic impact.

By some convention, though not universally adapted, the groups of cervical lymph nodes relevant to the spread of head-and-neck cancer are designated by Roman numerals, as follows (Fig. 1):

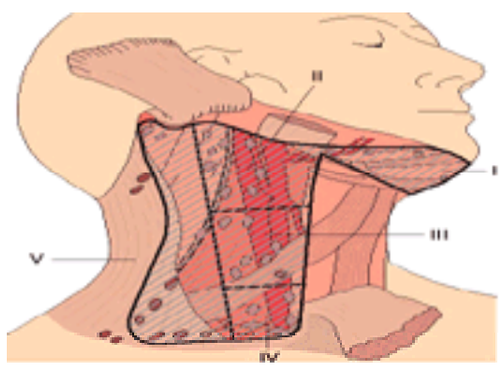


Fig. 1. Cervical lymph-node regions.

I. Submandibular triangle and submental
II. Upper deep jugular (above the carotid bulb)
III. Middle deep jugular (between carotid bulb and omohyoid muscle tendon)
IV. Lower deep jugular (below the omohyoid muscle tendon)
V. Posterior triangle, upper and lower (posterior to the sternocleidomastoid muscle).

Epidemiology

The oral cavity is the second most common site for cancer in the head and neck, after the larynx. In the United States of America, approximately 30 000 new cases of oral and oropharyngeal malignancy were identified in 1992, with carcinoma of the oral cavity occurring approximately two-and-a-half times more frequently than oropharyngeal carcinoma. In the same year, almost 8000 deaths were attributable to oral and oropharyngeal cancer in the United States. Worldwide, cancer of the oral cavity is the seventh most common malignancy, after lung, stomach, breast, colon–rectum, prostate, and the uterine cervix. Tobacco use of any kind, but particularly cigar, pipe and chewing tobacco, are causally significant factors. The concomitant use of alcohol in a smoker increases the risk of oral and oropharyngeal cancer significantly over that of either a nontobacco-using drinker or and nonalcohol-using smoker. Alcohol probably acts directly, both as an independent carcinogen and as a facilitator for tobacco toxins, and indirectly, through nutritional, immunologic, and other lifestyle influences. Poor dentition and/or chronic oral inflammation due to mechanical or infectious causes are frequently identified concomitants in patients with oral and oropharyngeal cancer.

Although syphilis was implicated in the etiology of oral cancer in the past, this was probably a lifestyle coincidence, as more recent studies cast doubt on its role in causing oral malignancy.

In Asia (including the IndiaN subcontinent) and the Far East (China and Japan), oral cancers make up as many as half of all malignancies. The chewing of *pan* and the reverse smoking of *chuttas* predispose to oral carcinomas in this part of the world. Patients with head-and-neck cancer often show depressed cellular immunity, although the importance of this in predisposing to malignancy remains ill defined.

Ninety per cent of oral and oropharyngeal cancers arise directly from the mucosal epithelium—squamous-cell carcinoma (also referred to as epidermoid carcinoma). Occasionally, an epithelial malignancy will arise from a duct structure or from invaginated epithelium of the tonsil or tongue base, giving rise to a submucosal lesion. Malignancies of the minor salivary glands (mucoepidermoid carcinoma, adenocarcinoma, adenoid cystic carcinoma) make up roughly 5 per cent of oral and oropharyngeal cancers, also arise submucosally, and can be difficult to detect. As an anatomic subsite, the soft palate is the most common location for nonepithelial malignancies, which comprise some 40 per cent of cancers of that area. In the oral cavity and oropharynx overall, the lateral tongue and floor of mouth are the most common sites for squamous-cell carcinoma. Other histologic types including lymphoma, melanoma (primary and metastatic) and variant squamous-cell malignancies (verrucous carcinoma, lymphoepithelioma, spindle-cell carcinoma) also occur, but rarely.

Presentation

The spectrum of presentation of oral cancers differs somewhat from that of oropharyngeal lesions. Both can be remarkably surreptitious, however, and an advanced stage at presentation is frequent, particularly in the oropharynx. Confusion of associated symptoms with more common and less threatening entities contributes to the delay in diagnosis. Early lesions of the oral cavity are usually painless and may arise in areas of pre-existing leukoplakia, a condition in which the mucosa is thickened

and whitish in appearance, occasionally roughened and palpable.

Nonmalignant leukoplakia is very common, occurring in areas of mechanical, dental, or chemical irritation and at times produced by other conditions (e.g. oral lichen planus). Persistent reddish lesions (erythroplakia), ulcerations, and exophytic growths present an increasing risk of malignant change. Local or referred pain (especially otalgia), bleeding, dysarthria, and odynophagia may attend the diagnosis of cancer. Particularly in the oropharynx, a neck mass, representing cervical metastasis, may be the single presenting concern.

Routine medical or dental examination may identify early lesions of the oral cavity, but the tongue base and most of the oropharynx are inaccessible to simple examination. An understanding and awareness of incriminating symptoms or associated findings are essential for promoting timely referral to a head-and-neck physician for complete examination. Oral and oropharyngeal cancers are notoriously poorly demarcated and subject to understaging. In some patients, particularly smokers, multiple areas of abnormal mucosa can confound simple physical examination and require detailed and systematic evaluation by biopsy. The possible occurrence of multiple primary malignancies, either synchronous or asynchronous, must be kept in mind.

In addition to, or instead of, a visible lesion, visible or palpable asymmetries in the normal anatomic contours beneath apparently normal mucosa may be present, in the tonsil region for example, representing submucosal infiltration of an epithelial cancer or the submucosal origin of a minor salivary tumor. Deep infiltration of the tongue or pterygoid muscles, or adherence of the tumor to jaw structures, may manifest as poor mobility of the tongue or palate, or trismus with dysphagia. Functional asymmetries or distant patterns of dysesthesia may also occur, due to cranial-nerve involvement by tumor. Direct bony invasion by tumor may be apparent from the presence of loosened teeth or a palpable defect, but can also be occult, requiring imaging studies or even surgical exploration to confirm. Bimanual palpation is essential to appreciate the extent of a tumor and also the presence of cervical nodal metastasis. Conventional examination of the oropharynx, hypopharynx, and larynx by mirror or fiberoptics is important, in part to identify associated findings, second primary or occult primary tumors, but may actually provide less useful information than careful palpation for tumors in the oral cavity and oropharynx. The tonsillar fossa and the base of tongue are two of the most common sites for hidden squamous-cell carcinomas and should be examined carefully in patients who present with cervical metastatic disease and an ‘unknown' primary tumor. Both locally and regionally, the midline of the oral cavity/oropharynx provides no barrier to or protection against tumor spread. Contralateral metastasis from even a relatively small cancer of the tongue base may be the sole presenting feature.

Oral cancer: clinical features by subsite

Carcinoma of the lip occurs most commonly on the lower lip and least frequently at the oral commissures. Lip cancer can be caused by tobacco use, but also by sun exposure of the outer aspects of the lips, resulting in a lesion indistinguishable from that caused by smoking and entailing essentially the same clinical considerations. The early bleeding, ulcerative, or exophytic lesion may eventually progress to cervical metastasis, involvement of the mental nerve, and mandibular invasion. Carcinomas of the upper lip tend to metastasize earlier and grow more rapidly than lesions of the lower lip. Early (T1) tumors can be managed by local wedge resection, including at least a 0.5-cm margin of normal tissue confirmed histologically. More advanced lesions usually require combined radical surgery and adjuvant radiation therapy. Overall 5-year survival is between 65 and 80 per cent.

Worldwide, carcinoma of the tongue is second only to cancer of the lip in frequency. As in most head-and-neck cancer sites, this is more often a disease of older men. Tobacco use, both chewed and inhaled, alcohol intake, and chronic dental inflammatory disease are implicated. However, tongue cancer and that of other sites within the oral cavity and oropharynx is now found with increasing frequency in younger individuals with no predisposing factors. The diagnosis of malignancy should therefore not be discounted in any individual, irrespective of tobacco/alcohol use or age. The lateral border of the mobile tongue is the most common site, but lesions can occur in any location.

As in all squamous-cell cancers of the head and neck, cervical nodal metastasis, occult or overt, occurs in proportion to the size/stage of the primary tumor and its location. Lymphatic spread from lesions of the anterior tongue and floor of mouth will tend to manifest in level I and II cervical nodes, but can also be a source of spread within the oral cavity. As a general rule, on proceeding posteriorly from the tip of the mobile tongue to the tongue base the stage at presentation advances, the incidence of cervical nodal metastasis, including contralateral disease, increases, and the overall prognosis worsens. Notable exceptions occur, but are unusual. Likewise, on proceeding posteriorly from the oral cavity to oropharynx the morbidity of surgical treatment increases. Particularly when a primary tumor has spread to the contralateral side of the tongue, the morbidity of surgical treatment can be extreme.

The anatomy beneath the floor-of-mouth mucosa is complex, compact, and difficult to evaluate clinically. Carcinomas of the floor of mouth often begin anteriorly but can progress to invade along the lingual nerve or submandibular ducts, into the deep musculature of the tongue, on to and through the periosteum of the mandible, and/or into the mandibular cortex itself. Mandibular invasion can occur through preformed anatomic routes without direct cortical invasion, particularly in the posterior floor of mouth and retromolar trigone, where the perineural spread of tumor along the mandibular nerve can result in bone involvement. Cancer of the retromolar trigone tends to be less well differentiated than cancer located more anteriorly in the oral cavity. Occult invasion through prior dental extraction sites can also occur and is difficult to detect unless advanced bone erosion from within has occurred. Carcinomas of the floor of the mouth and mobile tongue will present with palpable nodal metastases in between 30 and 40 per cent of cases, with the incidence of occult nodal disease even higher.

The alveolar ridge is a less common site of origin for cancer and is more often involved by extension from the floor of the mouth. There is often a history of local trauma from a denture, in addition to smoking and alcohol use. Most tumors are located on the lower gingiva posteriorly. Because of the direct apposition of the mucoperiosteum to the mandible, bone involvement is likely, even with relatively small primary tumors. The risk of osteoradionecrosis of the mandible with radiation therapy is a major concern, prompting a treatment plan that usually involves preliminary surgery.

The buccal area is fortunately an uncommon site of carcinoma. Deep extension or infiltration of other sites in the oral cavity is common, especially with the infiltrative variety of squamous carcinoma. The treatment of advanced lesions in this region can be disfiguring as well as functionally debilitating.

Cancers of the hard palate are about equally divided between squamous-cell and salivary gland malignancies. Squamous-cell carcinoma of this site is relatively rare, except in parts of the East where reverse smoking is practised. The tumor tends to spread tangentially as it grows, since the periosteum acts as a barrier to bone destruction early in the disease process. Cervical metastasis is uncommon.

A small proportion of tumors in the oral cavity will be verrucous carcinomas—exophytic, relatively low-grade malignancies with low metastatic potential and a deceptively benign papillary appearance histologically and clinically. This variant of squamous-cell carcinoma is generally radioresistant and requires surgical management regardless of site or stage, but has a generally good prognosis unless dedifferentiation to a more anaplastic type has occurred.

Oropharyngeal carcinoma: clinical features by subsite

The tonsillar fossa is the most common site of oropharyngeal malignancy. The term ‘carcinoma of the tonsil' can be misleading as it is used, regardless of the presence or not of the palatal tonsil tissue itself, as an anatomic descriptor. The majority are squamous carcinomas, but lymphomas of the non-Hodgkin type can also arise in the lymphoid tissue of the tongue base and palatal tonsils. Along with the nasopharynx and its adenoidal tissue, these two regions comprising Waldeyer's ring are notoriously common sites for initially occult primary cancers. Carcinomas of the tonsillar fossa often extend to the base of tongue, soft palate, retromolar trigone, and pterygoid muscles. As in the tongue base, unless uncovered during a search for an ‘unknown' primary prompted by the presence of a metastatic cervical node, tonsillar cancer often presents at an advanced stage. Occasionally a mass high in the cervical area is evidence not of metastasis but of direct extension into the parapharyngeal space. Conversely, a submucosal deformity in the tonsil region may be an indication of a primary parapharyngeal tumor, an important distinction as most of these are benign.

Cancers of the soft palate tend to be diffuse tumors and may be mixed in with premalignant ‘field' changes as well as nicotinic stomatitis. There is a propensity for bilateral metastasis to the neck, often occult, via the retropharyngeal pathway, thereby suggesting the need for prophylactic irradiation of both cervical regions and the retropharyngeal lymphatics. Malignancies of minor salivary glands of the oral cavity and oropharynx arise most often in the soft palate, and include mucoepidermoid and adenoid cystic carcinomas. These can also occur at other sites, but rarely.

Cancers of the base of the tongue tend to present late in their course, often with a paucity or very short duration of incriminating symptoms. Subtle otalgia, vague and nonspecific dysphagia or dysarthria, or a neck mass may be present at the time of the diagnosis. Even earlier-stage lesions have a high rate of cervical nodal metastasis, including bilateral disease. Almost always, radiation must be included in the treatment plan. Cancer of the tongue base may extend into the tonsil region, retromolar trigone, and posterior floor of mouth, but it is its extension across the midline or into the supraglottic larynx that is particularly morbid and a prime argument for nonsurgical therapy if feasible.

Diagnosis and associated examinations

Cancers of the oral cavity, though accessible to self-examination, are identified by the patient surprisingly infrequently or are too readily passed off as traumatic or

dental in origin. They may occasionally be identified on routine dental or medical examination, but a simple tongue blade and flashlight examination generally precludes a good view of the floor of mouth, hard palate, inner lips, and certainly the tongue base.

The most important aspect of diagnosis of an oral or oropharyngeal mucosal or mass lesion is a careful and complete examination by an appropriate specialist. A biopsy could often be performed under local anesthesia, but is best deferred at least until after any necessary associated studies have been performed.

An examination under general anesthesia is traditionally advised as a separate evaluation prior to the formulation of a treatment plan and should be omitted in only the rarest of cases. A general anesthetic permits multiple endoscopic examinations at one sitting, multiple biopsies under safe and comfortable conditions, and more accurate anatomic tumor mapping and, therefore, staging, all by a member of the treatment-planning team, typically the head-and-neck surgeon. The so-called triple endoscopy or panendoscopy consists of a rigid direct laryngoscopy, rigid and/or flexible esophagoscopy, and flexible bronchoscopy. Selectively, an endoscopic fiberoptic or direct examination of the nasal cavity and nasopharynx is added as well. In addition to characterizing completely the index tumor and documenting its histologic variety, this extensive examination is intended to detect second or multiple areas of primary mucosal malignancies, or, equally importantly, to ensure their absence prior to treatment planning.

Tumor mapping entails the development of a rigorous and accurate anatomic description of the tumor and its extent, particularly of any involvement of adjacent structures that would bear significantly on treatment decisions. As an adjunct the tumor mass may be outlined by tattooing the surrounding normal mucosa with India ink, which offers the surgeon a permanent demarcation of the original tumor distribution for determining the surgical margins following chemotherapy and/or radiation.

Biopsy usually consists of a punch or wedge of tissue taken from the periphery of the mass, both to allow comparison with the surrounding normal tissue and to avoid sampling only the potentially devitalized tissue at the center of the tumor. The specimen should be submitted for pathologic examination in an appropriate fixative such as formalin. Immediate frozen-section evaluation of a portion of the biopsy material is advisable in order to ensure that an adequate amount and quality of tissue has been provided for subsequent histologic analysis, and to specify the need for special handling if a lymphoma or non-squamous tumor is suggested. Especially in the oral cavity, so-called field cancerization can occur, wherein larger areas of premalignant or early malignant change (carcinoma *in situ*) may exist in proximity to, or separate from, the index cancer. Survey biopsies of all suspicious areas should be considered, especially if a surgical treatment plan is under review.

Given the number of examinations, staging issues, and the potential for multiple biopsies, it is clear to most practitioners that an adequate evaluation of a head-and-neck cancer cannot be completed in an office setting or under local/topical anesthesia. Furthermore, when appropriate, in preparation for a particular treatment plan, dental extractions, the placement of tracheotomy or gastro-stomy tubes, and/or the insertion of an indwelling intravenousaccess device can all be done during the same general anesthetic, permitting a higher measure of efficiency, comfort, and safety for the patient.

Biopsy of suspicious cervical adenopathy should be limited initially to fine-needle aspiration for cytology. Core-needle and open surgical biopsy may be associated with shortened survival, presumably due to tumor dissemination. Abnormal adenopathy in the presence of any head-and-neck carcinoma should be assumed to represent metastasis, so that diagnosis of cervical tissue is rarely necessary. In patients with suspicious cervical adenopathy and no identified primary tumor, open biopsy should be preceded by a careful triple endoscopy and examination under anesthesia, with selective biopsy of any suspicious mucosal abnormalities or, in their absence, consideration for directed biopsies of potential sites of occult cancer (base of tongue, tonsillar fossa, nasopharynx, and piriform sinus). An open biopsy should be preceded by fine-needle cytologic aspiration and performed only if the latter is inadequate. Every effort should be made to excise completely the mass in question rather than perform a partial, incisional sampling, lest tumor spillage and seeding occur. Immediate neck dissection may be worthwhile for some histologic types if confirmation of malignancy can be derived from frozen section at the time of the diagnostic excisional lymph-node biopsy.

The local and regional extent of tumor, and the possibility of distant metastasis or associated malignancy, can be further clarified by appropriate imaging studies, which should be performed prior to any manipulation that may artefactually distort the images. A chest radiograph is standard. If not already done, any area suggested by suspicious symptoms or past history should be appropriately investigated; blood tests are required to evaluate both the patient's comorbid medical status and candidacy for treatment.

Orthoplanograms (panorex radiographs) are useful in assessing bony invasion by tumors originating close to the mandible. These are also useful in differentiating between primary bony tumors and cysts. A negative bone scan of the mandible is very reassuring in ruling out early, subclinical cortical erosion. Barium-contrast esophagography may be useful in evaluating the esophagus as well as the inferior extent of oropharyngeal or hypopharyngeal tumors. Computerized tomographic (**CT**) scanning can define bony invasion, extension into the nose or maxillary sinuses, prevertebral invasion, disease of the pterygoid and parapharyngeal spaces, and, to a certain degree, the extent of cervical metastasis. In comparison to CT scans, magnetic resonance imaging (**MRI**) offers superior resolution of soft-tissue detail for delineating the extent of primary tumors and nodal disease, but is limited in evaluating bony involvement. CT and MRI may be useful in determining the resectability of advanced tumors in certain instances, and may allow assessment of response to therapy in certain deeply invasive malignancies. Bone scanning is also useful if disseminated disease is suspected, but in general, other than the chest radiograph, empiric total-body imaging does not contribute sufficiently to treatment planning to be warranted routinely. Positron-emission tomography is being investigated for its potential to distinguish between recurrent cancer and scar tissue.

Staging

The clinical staging of oral and oropharyngeal tumors is listed in Table 1. In general, larger tumors of the oral cavity and any oropharyngeal cancers are associated with a significant risk of cervical metastasis, with some notable case-by-case exceptions. The staging system according to the American Joint Committee is useful for general prognostication, treatment planning, and reporting consistency, but is neither perfect nor infallible. In the simplest of generalities, stage I and II disease is treatable with either surgery or radiation. Advanced disease in stages III and IV requires combined therapy with both surgery and radiation for maximum rates of cure. Advanced techniques of radiation and chemotherapy have more recently been utilized both to improve cure rates and to attempt cure without certain surgical complications. Some strategies are new, others novel combinations of established therapies; these include hyperfractionation radiation, brachytherapy, neoadjuvant chemotherapy, and combined chemotherapy/radiation.

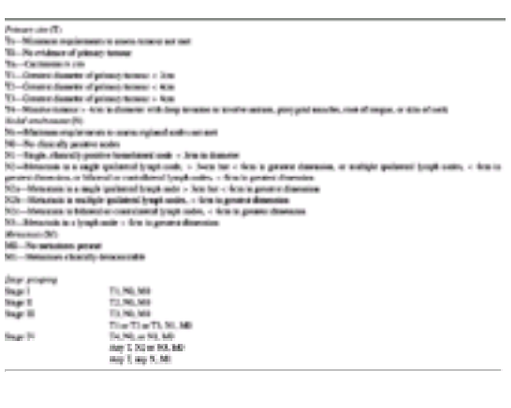


Table 1 TNM staging of oral and oropharyngeal cancer

Pathology

Ninety per cent of malignancies in the oral cavity are squamous-cell carcinomas. Histologically, most squamous carcinomas at this site exhibit relatively little pleomorphism, few mitoses, and significant keratinization (grades I and II); more poorly differentiated cancers (grades III and IV) are less frequent. There is a trend toward the more poorly differentiated types in oropharyngeal locations, particularly the tongue base. Adenocarcinomas of minor salivary-gland origin make up the bulk of the remaining malignancies. Other, unusual cancers include lymphomas, sarcomas, melanomas and metastatic tumors. Oral Kaposi sarcoma has been noted with increased frequency in patients with acquired immune deficiency syndrome.

Squamous-cell carcinomas are most commonly ulcerative, with relatively early metastasis and higher histologic grade. The infiltrative form shows less surface change but is similarly aggressive. Exophytic tumors are less common; they tend to grow along mucosal surfaces and metastasize later in their course. The least common form is verrucous carcinoma, a relatively benign-appearing, papillary lesion found most commonly along the buccal mucosa. This proliferative, well-differentiated lesion may be mistaken for benign hyperplasia, leading to underdiagnosis. Although it carries the best prognosis of the various types of squamous-cell carcinoma it can be locally aggressive. Its diagnosis requires a high index of suspicion by the head-and-neck surgeon.

Field cancerization refers to the occurrence of synchronous or metachronous premalignant mucosal lesions or frank second malignancies related to the chronic exposure of multiple areas of the upper aerodigestive tract to carcinogenic influences. In a Mayo Clinic series, second primary cancers of the oral cavity, pharynx, or esophagus were found in 16.3 per cent of 732 patients with squamous-cell carcinoma of the oral cavity. Thus, initial panendoscopy (direct laryngoscopy, esophagoscopy, bronchoscopy, and nasopharyngoscopy) to search for synchronous second primary tumors as well as careful follow-up monitoring for metachronous second malignancies is essential.

In the oropharynx, squamous-cell carcinoma is often nonkeratinizing and poorly differentiated, although the histologic features of the tumor at a specific site appears to make little difference prognostically. This observation emphasizes the current trend in molecular biologic evaluation of head-and-neck cancer which seeks to identify biologic behavioral correlates that may be more accurate prognosticators than crude anatomic staging or simple, traditional histology. In addition there are more unusual variants of squamous-cell carcinoma including spindle-cell and adenoid squamous-cell carcinoma, which are treated in the same fashion. Verrucous carcinoma is rare in the oropharynx. Lymphomas, especially in the tonsillar and base-of-tongue areas, are more common than in the oral cavity.

Over 80 per cent of tumors in minor salivary glands of the oral cavity are malignant. These are adenoid cystic carcinoma, mucoepidermoid carcinoma, and other adenocarcinomas. They occur mostly within the posterior hard palate and soft palate, and tend to present as painless, nonulcerated, submucosal masses. Cervical nodal metastasis is rare, as is a primary origin in the oropharynx.

Treatment

While there are now ever more possible treatment combinations, when a complete diagnostic and staging evaluation has been completed, one or two strategies generally come to the fore. The ideal and currently expected forum through which to develop a treatment plan is multidisciplinary, in which representatives from the several relevant specialties involved in the treatment of head-and-neck cancer review the pertinent information and contribute to a sanctioned recommendation. Beyond this, an informative interaction with the patient is of the utmost importance, to solicit at least their understanding of, if not preference for, a given treatment proposal. While it is arguably inappropriate to allow the patient to dictate their own course of treatment without the benefit of a specific recommendation and rationale from the consulting team, a number of patients will be burdened with erroneous preconceptions about their disease, perhaps logical in lay terms but none the less inaccurate. These must be reviewed and discussed, and an appropriate, long-term perspective generated collectively.

A detailed discussion of all the possible treatment schemes and their rationale for given sites and stages of disease is beyond the scope of this introductory chapter, as are descriptions of the many surgical and radiation techniques applicable in the current practice of head-and-neck oncology. Some generalizations are quite reasonable, however. Since early-stage primary lesions of both the oral cavity and oropharynx are controlled equally well by either radiation or surgery, treatment decisions are made on the basis of perceived surgical morbidity and/or the risk of occult cervical nodal metastasis. The decision may be heavily weighted by patient factors, including their subjective preference and medical candidacy. To a reasonable extent, as site of the lesion moves posteriorly in the oral cavity and then into the oropharynx, surgical morbidity, at least short term, increases and radiation may be increasingly favored. Likewise, more posterior lesions will tend to more often be associated with occult cervical metastasis, the treatment of which, particularly if bilateral, would most often involve external-beam radiation. Consequently, early posterior cancers will tend to be treated with radiation to both primary and nodal sites, while anterior lesions may lend themselves to surgical resection, possibly with adjunctive neck dissection for nodal assessment.

Advanced-stage primary lesions of the oral cavity and oropharynx, particularly tumors of large volume or those associated with deep invasion of muscle and/or bone, are poorly curable by radiation, and potentially at the expense of osteoradionecrosis. With or without cervical nodal metastasis, traditional treatment has involved radically ablative surgery and neck dissection, with reconstruction and postoperative external radiation therapy. While modern-day reconstructive techniques, particularly those involving microvascular free-tissue transfer, offer clearly superior function, cosmesis, and healing, such radical surgery is perceived by many a patient and practitioner alike to be excessively morbid for a disease whose curability has changed little in several decades. Consequently, nonsurgical treatment strategies have evolved, purporting to maintain existing rates of tumor control while improving function and appearance through the avoidance of any planned radical surgery.

While it is a general rule that patients will tend to underestimate the deleterious side-effects of radiation and overestimate those of radical surgery, certainly the avoidance of laryngectomy or major jaw resections is a reasonable goal. While for resectable disease entities nonsurgical treatments have not been conclusively shown to increase survival, they have been shown to decrease morbidity with unchanged rates of disease control, which is perceived as worthwhile. Consequently, in advanced-stage disease, a nonsurgical treatment option through preliminary (‘induction' or ‘neoadjuvant') chemotherapy followed by one of various radiation-based techniques, or one of several other combinations of chemotherapy and radiation, exists. Determining a patient's medical candidacy for these advanced treatment protocols can be every bit as challenging as preparing one for, and recovering one from, radical surgery, however.

One of the more useful radiation-based techniques especially applicable in the oral cavity and oropharynx is brachytherapy, or interstitial radiation by implantation of a radioactive source. Usually used in combination with external-beam therapy, brachytherapy allows the dose of radiation to be significantly boosted in the accessible region of the primary tumor while lessening the exposure of the surrounding uninvolved tissues, as compared to an entirely external beam-based approach. Intraoral, focused-cone radiation can accomplish the same purpose, but with some limitations in geometry.

In oral and oropharyngeal carcinoma of the head and neck, clinically gross metastatic cervical adenopathy generally requires classical radical or modified radical neck dissection (Fig. 2). This *en bloc* resection of all (levels I–V) of the cervical lymph nodes may include the removal of the sternocleidomastoid muscle, internal jugular vein, and the spinal accessory (XI) nerve. Supraomohyoid neck dissections (levels I–III), unilateral or bilateral, are appropriate for earlier-stage lesions in the oral cavity in which the risk of occult nodal metastasis is perceived as significant (Fig. 3). These less-than-complete lymphadenectomies are conceived as diagnostic procedures. In general, should metastatic adenopathy be identified intraoperatively, a more radical procedure would be performed and postoperative radiation planned. If metastasis were identified on postoperative histologic analysis only, postoperative radiation would be considered.

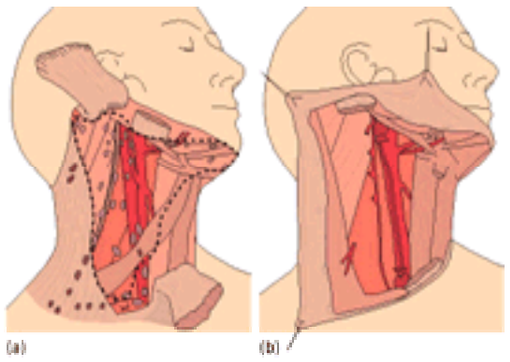


Fig. 2. (a) Complete neck dissection; (b) classic, radical neck dissection.

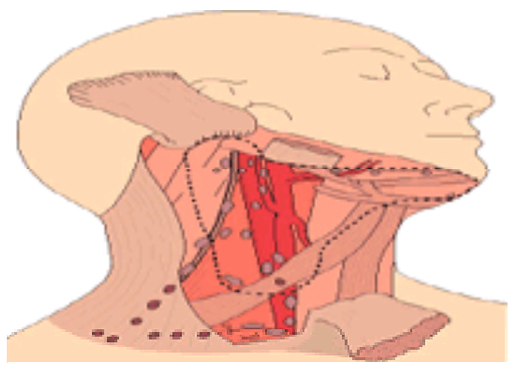


Fig. 3. Supraomohyoid neck dissection.

Primary resection of the site of earlier lesions of the anterior oral cavity can often be performed transorally, especially if the patient is edentulous. A gross margin of 2 cm of normal tissue, with frozen-section confirmation, is desirable but not always possible in this compact and complex region. Primary closure is often possible, but reconstruction using a skin, dermal, or free mucosal graft may be needed where mucosa cannot be advanced or rotated, or where tethering of the mobile tongue may result.

Lip lesions can usually be removed by ‘V' or wedge resection, with either primary closure or the use of nasolabial flaps for reconstruction. More advanced lesions intraorally and in the oropharynx require better exposure, through median mandibulotomy, for access to the tumor for resection and reconstruction (Fig. 4 and Fig. 5). Exposure can also be accomplished through a lower lip splitting approach with elevation of a cheek flap, or through a submandibular degloving approach to the mandible.

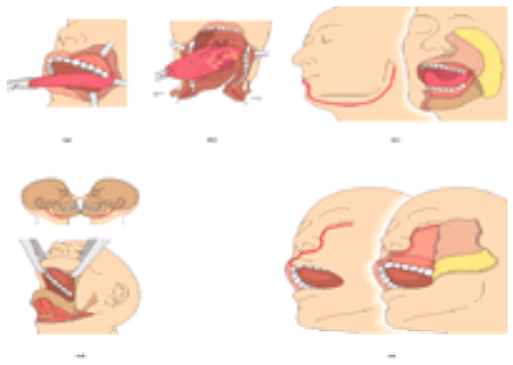


Fig. 4. Surgical approaches to the oral cavity and oropharynx: (a) peroral; (b) mandibulotomy with paralingual extension; (c) lower cheek flap; (d) ‘visor' flap; (e) upper cheek flap.

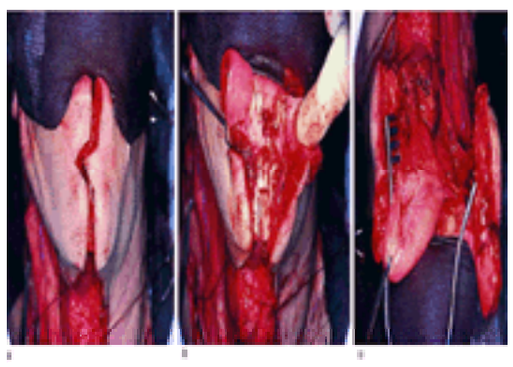


Fig. 5. Median mandibulotomy approach to the oropharynx: (a) lip-splitting incision; (b) median mandibulotomy; (c) paralingual approach exposing tongue base, oropharynx, and hypopharynx.

Complete tumor resection may require segmental or hemimandibulectomy, creating both a soft-tissue and a bony deficiency. Rehabilitation of speech and deglutition requires reconstruction of the resultant defect, using local and regional tongue, cutaneous, and myocutaneous flaps (Fig. 6). Mandibular reconstruction techniques using stainless-steel plating and pedicled myocutaneous flaps (usually from the pectoral region), or particularly a microvascular osteomyocutaneous free tissue transfer, are the more current approaches to functional and cosmetic rehabilitation after radical ablative surgery for cancer of the oral cavity or oropharynx.

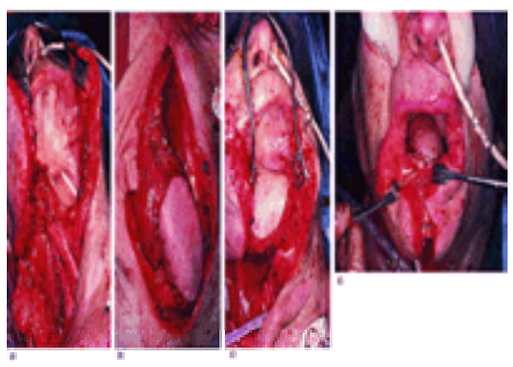


Fig. 6. Large oral defect following tumour extirpation requires reconstruction: (a) treatment resulting in total glossectomy; (b) pectoralis major myocutaneous flap developed to reconstitute the defect; (c) the flap is designed to fill the defect appropriately; (d) resultant tension-free closure of the oral cavity and reapproximation of the mandible with stainless-steel wire or fixation plate.

Chemotherapy

The role of chemotherapy in the treatment of oral and oropharyngeal carcinoma has evolved considerably over the last two decades. New agents and a better appreciation of the biology of squamous-cell cancer of the head and neck have led to substantial advances in the integrated treatment of this disease. Chemotherapy was initially reserved for the treatment of recurrent and incurable disease. Single-agent therapy and combination treatments were given in an effort to palliate patients by improving their quality of life and extending life expectancy. Randomized trials demonstrated that combination chemotherapy with the most active agents, cisplatinum and 5-fluorouracil, is more effective than treatment with single drugs in eliciting responses. Among other agents tested, carboplatinum and methotrexate are known to be active as single agents and have been compared to combination cisplatinum/5-fluorouracil. More recently, the taxanes, docetaxel and paclitaxel, have been found to be effective single-agent treatments for palliation. Randomized clinical trials are now comparing combinations of the taxanes with cisplatinum/5-fluorouracil to determine which combination might be best suited to palliative treatment. While combination therapy gives a higher response rate in patients with recurrent disease, it is associated with more side-effects. Efficacy, palliation, quality of life, and life expectancy must be balanced. Occasionally, single-agent therapy as first-line treatment may be better for a patient than an intensive course of chemotherapy.

Chemotherapy has a more important role in the treatment of locally advanced, previously untreated oral and oropharyngeal carcinoma. Although the first studies of induction chemotherapy and chemoradiotherapy showed no advantage to the use of chemotherapy, there is now substantial and compelling evidence that they should be part of the initial treatment for patients with advanced disease, or in whom organ preservation would be an important consideration. Two trials have demonstrated that induction chemotherapy with cisplatinum/5-fluorouracil can replace surgery in the treatment of resectable hypopharyngeal and laryngeal cancer. In both trials, chemotherapy resulted in the preservation of speech and swallowing function in 60 per cent of survivors. Survival was not altered when compared to surgical treatment alone. Another trial demonstrated that treatment with induction chemotherapy can improve survival in unresectable patients compared to radiotherapy alone. Although the tongue base and other sites of oral and oropharyngeal cancer have not been directly assessed in these studies, the results strongly support the use of induction chemotherapy in this disease, both for organ preservation and primary treatment. More recent phase II trials have suggested that improved response rates and locoregional control might be achieved using new agents such as the taxanes or modifications of cisplatinum/5-fluorouracil such as cisplatinum/5-fluorouracil with leucovorin.

The timing of chemotherapy and radiotherapy is very important. Based on the growth patterns of tumor cells, it is now believed that if chemotherapy is to be part of a combined-modality program, radiation should follow immediately after, and any surgery should then follow radiotherapy.

Chemoradiotherapy has also been shown to improve survival in unresectable patients, although its impact on organ preservation has not been demonstrated. Several studies have demonstrated the superiority of chemoradiotherapy over radiotherapy alone, including a recent study on chemotherapy delivered with hyperfractionated radiotherapy. Of interest, several chemoradiotherapy programs also included adjuvant chemotherapy with modified cisplatinum/5-fluorouracil. Two meta-analyses have also demonstrated the superiority of chemoradiotherapy and induction chemotherapy over standard treatment. The choice between induction chemotherapy and chemoradiotherapy is still unclear. Each has advantages and disadvantages. One can now state that chemotherapy and organ preservation can be considered standard care for patients with locally advanced disease and can result in reduced morbidity as well as improved survival.

General prognosis

In the absence of clinical cervical nodal metastasis, early cancers, T1 and T2 primary lesions of the oral cavity and T1 primaries of the oropharynx, have what is considered in context to be a good chance of cure, approximately 80 to 100 per cent. In the long term, second primary cancers are a particular problem in the oral cavity, an observation favoring the surgical approach for initial treatment.

In advanced disease, especially T4 in either region, the prognosis worsens precipitously. With any stage of primary, the concomitant presence of nodal disease, especially N2 or N3, roughly halves the chance of cure. For both advanced primary stage and advanced nodal metastasis, or both, the risk of distant metastasis (about 20 per cent), most often to the lung, becomes an additional significant factor determining treatment failure. By virtue of their systemic effect, chemotherapeutic treatment strategies can influence the incidence and/or timing of distant metastatic disease as well as assist with local and regional control. A complete response to induction chemotherapy, particularly a histologically proved complete response, has been associated with a marked improvement in both disease-free and overall survival, essentially independent of presenting stage, when compared with partial responders, nonresponders, and historical controls treated by surgery and/or radiation. Whether this observation is attributable to the effect of chemotherapy or its selection of otherwise favorable, presumably radiosensitive, tumors had not yet been conclusively proved or disproved by prospective trials.

With the possible exception of delayed regional metastasis to cervical nodes, recurrent cancer of the oral cavity or oropharynx is infrequently salvaged, particularly if radical surgery or induction chemotherapy have been used as part of the initial treatment. For all stages, overall survival at 5 years is approximately 40 to 50 per cent for sites in the oral cavity and 25 to 35 per cent for the oropharyngeal region. Moreover, in some cases, the sequelae of treatment may impose potentially life-threatening morbidity because of aspiration, airway compromise, and nutritional complications.

Further reading

Batsakis JG. *Tumors of the head and neck. Clinical and pathological considerations*, 2nd edn. Williams and Wilkins, Baltimore, 1979.

Beahrs OH. Surgical anatomy and technique of radical neck dissection. *Surgical Clinics of North America* 1977; **57**: 663–700.

Beahrs OH, Henson DE, Hutter RVP, Myers MH, ed. *Manual for staging of cancer*, 3rd edn. Lippincott, Philadelphia, 1988.

Brizel D *et al.* Hyperfractionated radiation with or without concurrent chemotherapy for locally advanced head and neck cancer. *New England Journal of Medicine* 1998; **338**: 1798–804.

Byers RM *et al.* Can we detect or predict the presence of occult nodal metastases in patients with squamous carcinoma of the oral tongue? *Head and Neck* 1998; **20**: 138–44.

Clark J *et al.* Induction chemotherapy with cisplatin, fluorouracil and high-dose leucovorin for squamous cell carcinoma of the head and neck: long-term results. *Journal of Clinical Oncology* 1997; **15**: 3100–10.

Clark JR, Fallon BG, Frei E, III. Induction chemotherapy as initial treatment for advanced head and neck cancer: a model for the multidisciplinary treatment of solid tumors. In: DeVita VT, Jr, Hellman S, Rosenberg SA, ed. *Important advances in oncology*, pp. 175–95. Lippincott, Philadelphia, 1987.

Close LG, Burns DK, Reisch J, Schaefer SD. Microvascular invasion in cancer of the oral cavity and oropharynx. *Head and Neck Surgery* 1987; **113**: 1191–5.

Colevas A *et al.* Induction chemotherapy with docetaxel, cisplatin, 5-fluorouracil and leucovorin (TPFL5) for squamous cell carcinoma of the head and neck: a phase I/II trial. *Journal of Clinical Oncology* 1998; **16**: 1331–9.

Cox J *et al.* Interruptions adversely affect local control and survival with hyperfractionated radiation of carcinomas of the upper respiratory and digestive tracts. *Cancer* 1992; **69**: 2744–8.

Cunningham MJ, Johnson JT, Myers EN, Schramm VL, Jr, Thearle PB. Cervical lymph node metastasis after local excision of early squamous cell carcinoma of the oral cavity. *American Journal of Surgery* 1986; **152**: 361–6.

Cusumano RJ, Persky MS. Squamous cell carcinoma of the oral cavity and oropharynx in young adults. *Head and Neck Surgery* 1988; **10**: 229–34.

de Visscher JG, van den Elsaker K, Grond AJ, van der Waal JE, van der Waal I. Surgical treatment of squamous cell carcinoma of the lower lip: evaluation of long-term results and prognostic factors—a retrospective analysis of 184 patients. *Journal of Oral and Maxillofacial Surgery* 1998; **56**: 814–21.

Donegan JO, Gluckman JL, Crissman JD. The role of suprahyoid neck dissection in the management of cancer of the tongue and floor of mouth. *Head and Neck Surgery* 1982; **4**: 209–12.

Dreyfuss A *et al.* Docetaxel: an active drug for squamous cell carcinoma of the head and neck. *Journal of Clinical Oncology* 1996; **14**: 1672–8.

El-Sayed S, Nelson N. Adjuvant and adjunctive chemotherapy in the management of squamous cell carcinoma of the head and neck region: a meta-analysis of prospective and randomized trials. *Journal of Clinical Oncology* 1996; **14**: 838–47.

Fein DA *et al.* Carcinoma of the oral tongue: a comparison of results and complications of treatment with radiotherapy and/or surgery. *Head and Neck* 1994; **16**: 358–65.

Fletcher GH. Elective irradiation of subclinical disease in cancers of the head and neck. *Cancer* 1972; **29**: 1450–4.

Foote RL *et al.* Tonsil cancer. Patterns of failure after surgery alone and surgery combined with postoperative radiation therapy. *Cancer* 1994; **73**: 2638–47.

Forastiere A *et al.* Randomized comparison of cisplatin plus fluorouracil and carboplatin plus fluorouracil versus methotrexate in advanced squamous cell carcinoma of the head and neck: a Southwest Oncology Group study. *Journal of Clinical Oncology* 1992; **10**: 1245–51.

Forastiere A, Shank D, Neuberg D, Taylor S, CeConti R, Adams G. Final report of a phase II evaluation of paclitaxel in patients with advanced squamous cell carcinoma of the head and neck. *Cancer* 1998; **82**: 2270–4.

Frazell EL, Lucas JC. Cancer of the tongue: report of the management of 1,554 patients. *Cancer* 1962; **15**: 1085–99.

Gluckman JL *et al.* Prognostic indicators for squamous cell carcinoma of the oral cavity: a clinicopathological correlation. *Laryngoscope* 1997; **107**: 1239–44.

Goffinet DR *et al.* ^{192}Ir pharyngoepiglottic fold interstitial implants. The key to successful treatment of base of tongue carcinoma by radiation therapy. *Cancer* 1985; **55**: 941–8.

Goldblatt LL, Ellis GL. Salivary gland tumors of the tongue. *Cancer* 1987; **60**: 74–81.

Hicks WL, Jr *et al.* Surgery versus radiation therapy as single-modality treatment of tonsillar fossa carcinoma: the Roswell Park Cancer Institute experience (1971–1991). *Laryngoscope* 1998; **108**: 1014–19.

Hinerman RW, Parsons JT, Mendenhall WM, Stringer SP, Cassisi NJ, Million RR. External beam irradiation alone or combined with neck dissection for base of tongue carcinoma: an alternative to primary surgery. *Laryngoscope* 1994; **104**: 1466–70.

Horiot J *et al.* Hyperfractionation versus conventional fractionation in oropharyngeal carcinoma: final analysis of a randomized trial of the EORTC Cooperative Group of Radiotherapy. *Radiotherapy and Oncology* 1992; **25**: 231–41.

Jacobs C. Head and neck cancer in 1994: a change in the standard of care. *Journal of the National Cancer Institute* 1994; **86**: 250–2.

Jacobs C *et al.* A phase III randomized study comparing cisplatin and fluorouracil as single agents and in combination for advanced squamous cell carcinoma of the head and neck. *Journal of Clinical Oncology* 1992; **10**: 257–63.

Jesse RH, Barkley HT, Jr, Lindberg RD, Fletcher GH. Cancer of the oral cavity: is elective neck dissection beneficial? *American Journal of Surgery* 1970; **120**: 505–8.

Jovanovic A, Schulten EA, Kostense PJ, Snow GB, van der Waal I. Squamous cell carcinoma of the lip and oral cavity in The Netherlands: an epidemiological study of 740 patients. *Journal of Craniomaxillofacial Surgery* 1993; **21**: 149–52.

Jovanovic A *et al.* Risk of multiple primary tumors following oral squamous cell carcinoma. *International Journal of Cancer* 1994; **56**: 320–3.

Kligerman J *et al.* Supraomohyoid neck dissection in the treatment of T1/T2 squamous cell carcinoma of the oral cavity. *American Journal of Surgery* 1994; **168**: 391–4.

Kotelnikov V *et al.* Cell kinetics of head and neck cancer. *Clinical Cancer Research* 1995; **1**: 527–7.

Kowalski LP *et al.* Lateness of diagnosis of oral and oropharyngeal carcinoma: factors related to the tumor, the patient and health professionals. *European Journal of Cancer Biology and Oral Oncology* 1994; **30**: 167–73.

Kraus DH, Vastola AP, Huvos AG, Spiro RH. Surgical management of squamous cell carcinoma of the base of the tongue. *American Journal of Surgery* 1993; **166**: 384–8.

Leedy DA, Trune DR, Kronz JD, Weidner N, Cohen JI. Tumor angiogenesis, the p53 antigen, and cervical metastasis in squamous carcinoma of the tongue. *Otolaryngologic and Head and Neck Surgery* 1994; **111**: 417–22.

Lefebvre J, Chevalier D, Luboinski B, Kirkpatrick A, Collette L, Sahmoud T. Larynx preservation in pyriform sinus cancer: preliminary results of a European Organization for Research and Treatment of Cancer phase III trial. *Journal of the National Cancer Institute* 1996; **88**: 890–8.

Lefebvre JL, Coche-Dequeant B, Castelain B, Prevost B, Buisset E, Ton Van J. Interstitial brachytherapy and early tongue squamous cell carcinoma management. *Head and Neck* 1990; **12**: 232–26.

Lindberg RD. Distribution of cervical lymph node metastasis from squamous cell carcinoma of the upper respiratory and digestive tracts. *Cancer* 1972; **29**: 1446–50.

Medina JE. A rational classification of neck dissections. *Otolaryngologic and Head and Neck Surgery* 1989; **100**: 169–76.

Medina JE, Byers RM. Supraomohyoid neck dissection: rationale, indications, and surgical technique. *Head and Neck* 1989; **11**: 111–22.

Medina JE, Dichtel W, Luna MA. Verrucous–squamous carcinomas of the oral cavity. *Archives of Otolaryngology* 1984; **110**: 437–40.

Merlano M *et al.* Five year update of a randomized trial of alternating radiotherapy and chemotherapy compared with radiotherapy alone in treatment of unresectable squamous cell carcinoma of the head and neck. *Journal of the National Cancer Institute* 1996; **88**: 583–9.

Moose BD, Kelly MD, Levine PA, Constable WC, Cantrell RW, Larner JM. Definitive radiotherapy for T1 and T2 squamous cell carcinoma of the tonsil. *Head and Neck* 1995; **17**: 334–8.

Munro A. An overview of randomised controlled trials of adjuvant chemotherapy in head and neck cancer. *British Journal of Cancer* 1995; **71**: 83–91.

Nakanishi Y *et al.* p53 expression in multicentric squamous cell carcinoma and surrounding squamous epithelium of the upper aerodigestive tract. Immunohistochemical analysis of 95 lesions. *Cancer* 1995; **75**: 1657–62.

National Cancer Institute Head and Neck Contracts Writing Committee. Adjuvant chemotherapy for advanced head and neck carcinoma. Final report of the Head and Neck Contracts Program. *Cancer* 1988; **60**: 301–11.

Oral Health Coordinating Committee. Toward improving the oral health of Americans: an overview of oral health status, resources, and care delivery. *Public Health Reports* 1993; **108**: 657–72.

Paccagnella A *et al.* Phase III trial of initial chemotherapy in stage III or IV head and neck cancers: a study by the Gruppo di Studio sui Tumori della Testa e del Collo. *Journal of the National Cancer Institute* 1994; **86**: 265–72.

Parsons JT, Cassisi NJ, Million RR. Results of twice-a-day irradiation of squamous cell carcinomas of the head and neck. *International Journal of Radiation Oncology, Biology, Physics* 1984; **10**: 2041–51.

Robbins KT, Medina JE, Wolfe GT, Levine PA, Sessions RB, Pruet CW. Standardizing neck dissection terminology. Official report of the Academy's Committee for Head and Neck Surgery and Oncology. *Archives of Otolaryngology and Head and Neck Surgery* 1991; **117**: 601–5. [Comment in: *ibid*. 1995; **121**: 352–3.]

Shaha AR, Hoover EL, Mitrani M, Marti JR, Krespi YP. Synchronicity, multicentricity, and metachronicity of head and neck cancer. *Head and Neck Surgery* 1988; **10**: 225–8.

Shibuya H, Hisamitsu S, Shioiri S, Horiuchi J, Suzuki S. Multiple primary cancer risk in patients with squamous cell carcinoma of the oral cavity. *Cancer* 1987; **60**: 3083–6.

Shikhani AH, Matanoski GM, Jones MM, Kashima HK, Johns ME. Multiple primary malignancies in head and neck cancer. *Archives of Otolaryngology and Head and Neck Surgery* 1986; **112**: 1172–9.

Sigal R *et al.* CT and MR imaging of squamous cell carcinoma of the tongue and floor of mouth. *Radiographics* 1996; **16**: 787–810.

Slaughter DP, Southwick HW, Smejkal W. 'Field cancerization' in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer* 1953; **6**: 963–8.

Son YH, Kapp DS. Oral cavity and oropharyngeal cancer in a younger population. Review of the literature and experience at Yale. *Cancer* 1985; **55**: 441–4.

Spiro RH, Koss LG, Hajdu SI, Strong EW. Tumors of the minor salivary origin: a clinico-pathologic study of 492 cases. *Cancer* 1973; **31**: 117–29.

Teichgraeber JF, Clairmont AA. The incidence of occult metastases for cancer of the oral tongue and floor of mouth: treatment rationale. *Head and Neck Surgery* 1984; **7**: 15–21.

Thawley SE, Panje WR, ed. *Comprehensive management of head and neck tumors*, p. 555. Saunders, Philadelphia, 1987.

Veterans Affairs Laryngeal Cancer Study Group. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *New England Journal of Medicine* 1991; **324**: 1685–9.

Wang CC. Radiotherapeutic management and results of T1N0, T2N0 carcinoma of the oral tongue: evaluation of boost techniques. *International Journal of Radiation Oncology, Biology, Physics* 1989; **17**: 287–91.

Wey PD, Lotz MJ, Triedman LJ. Oral cancer in women nonusers of tobacco and alcohol. *Cancer* 1987; **60**: 1644–50.

Zeitouni AG, Yamamoto YL, Black M, Gjedde A. Functional imaging of head and neck tumors using positron emission tomography. *Journal of Otolaryngology* 1994; **23**: 77–80.

45.6 Nasopharyngeal cancer

Walter W. K. King and Arthur K. C. Li

Introduction
Epidemiology
Geographical and ethnic distribution
Sex and age pattern
Aetiology
Genetic factors
Environmental factors
Epstein–Barr virus
Epstein–Barr virus and nasopharyngeal carcinoma
Pathogenesis
Diagnosis
Anatomy of the nasopharynx
Clinical presentation
Diagnostic approach
Staging
Histopathology
Other nasopharyngeal tumours
Treatment
Radiation as the primary treatment
Role of chemotherapy
Surgical treatment of recurrent nasopharyngeal carcinoma
Radical neck dissection for regional recurrence
High dose re-irradiation versus nasopharyngectomy
Further reading

Introduction

Malignant tumours of the nasopharynx comprise lymphomas, adenocarcinomas, sarcomas, and neuroendocrine carcinomas, but the predominant nasopharyngeal carcinoma occurring in endemic areas (e.g. North Africa and southern China) is the undifferentiated type. The World Health Organization (WHO) classification divides the undifferentiated carcinomas (showing neither squamous nor glandular differentiation) into non-keratinizing carcinomas (WHO type II) and undifferentiated carcinomas (WHO type III). Differentiated or keratinizing squamous cell carcinoma (WHO type I) accounts for only a small percentage of nasopharyngeal carcinoma (Table 1). The undifferentiated nasopharyngeal carcinomas (WHO types II and III) are considered poorly differentiated forms of squamous cell carcinomas because they show epithelial differentiation with cellular junctions including desmosomes and tonofilaments. They are more closely associated with the Epstein–Barr virus, more responsive to radiotherapy, and have a better prognosis.

Non-keratinizing carcinomas	236 (90.4 per cent) Non-keratinizing (WHO II) = 21 (8.0 per cent) Undifferentiated (WHO III) = 215 (82.4 per cent)
Keratinizing squamous cell carcinomas (WHO I)	7 (2.7 per cent)
Malignant lymphoma (non-Hodgkin's and polymorphic reticulosis group)	13 (5.0 per cent)
Adenocarcinoma	1 (0.38 per cent)
Granulocytic sarcoma	1 (0.38 per cent)
Sinonasal undifferentiated carcinoma	1 (0.38 per cent)
Olfactory neuroblastoma	1 (0.38 per cent)
Embryonal rhabdomyosarcoma	1 (0.38 per cent)
Total	261 (100 per cent)

Table 1 Nasopharyngeal malignant tumours in Prince of Wales Hospital (January 1986–September 1987)

While nasopharyngeal carcinoma affects predominantly southern Chinese and South-east Asians, it is gaining worldwide attention for two reasons. First, the geographical movement of ethnic groups, either as travellers or migrants, over the last decades has increased the incidence with which clinicians encounter nasopharyngeal cancer. Second, the strong causal relationship between nasopharyngeal carcinoma and Epstein–Barr virus makes it a valuable model for the study of virus-related tumours.

Epidemiology

Geographical and ethnic distribution

In most parts of the world, the incidence of nasopharyngeal carcinoma for either sex is less than 1 per 100 000 persons per year. The highest incidence of nasopharyngeal carcinoma is in southern China, especially in the provinces of Guangdong, Guangxi, Fujian, Hunan, and Jiangxi (approximately 20 per 100 000 persons per year), where this tumour accounts for 4 to 14 per cent of all cancer deaths. In the highest risk areas of Guangdong, the incidence of nasopharyngeal carcinoma in males between 45 and 55 years of age reaches 120 per 100 000 persons per year. The incidence is also high in other areas of South-east Asia heavily populated by southern Chinese (Singapore, Hong Kong, northern Thailand), at 20 per 100 000 persons per year. A similarly high incidence of nasopharyngeal carcinoma has been observed in southern Chinese who have migrated to America and also in native Alaskans. Elsewhere, an intermediate incidence of 1.5 to 9 per 100 000 persons per year is observed in North Africa (Algeria, Tunisia, Morocco, Sudan), while Europe, the United States of America, Canada, and South America have an incidence of less than 1 per 100 000 persons per year. Registry data collected in Guangdong and Singapore suggest that the annual incidence of nasopharyngeal carcinoma during the 1970s has shown little fluctuation.

Sex and age pattern

Nasopharyngeal carcinoma has a male preponderance of 3:1. Although it has been reported in patients ranging from 3 to 86 years of age, it is uncommon under the age of 20 years (less than 1 per cent). In high-risk areas, there is a sharp rise in incidence after the age of 20 reaching a plateau between 40 and 60 years of age before declining with age thereafter. The mean age at diagnosis of this disease is about 50. The age distribution pattern suggests an induction time of 20 years and a plateauing period that is generally 10 years earlier than squamous carcinoma of other head and neck sites, suggesting that aetiological factors other than alcohol and smoking may be implicated in nasopharyngeal carcinoma.

Aetiology

The marked ethnic differences and geographical variations in the incidence of nasopharyngeal carcinoma have stimulated much research into the aetiology of this cancer. Recent studies have linked this cancer with a genetically determined susceptibility associated with specific human leucocyte antigen (HLA) haplotypes, early latent infection by the ubiquitous Epstein–Barr virus and its reactivation and exposure to environmental factors, in particular, tumour-promoting chemicals in food products, especially early age consumption of salted preserved fish.

Genetic factors

Epidemiological surveys have established that southern Chinese are at high risk of developing nasopharyngeal carcinoma, suggesting an inherited genetic predisposition to nasopharyngeal carcinoma. Familial aggregation of the disease has also been reported. In a case–control study carried out in China, 33 of 200

patients had a family history of cancer, including 14 families with nasopharyngeal carcinoma, while 10 of 200 matched controls had a family history of cancer with only two cases of nasopharyngeal carcinoma. Nasopharyngeal carcinoma was significantly more common in the siblings of patients with the disease (14.6 per cent) than in the controls (1.8 per cent). Whether the increased risk of developing nasopharyngeal carcinoma observed among siblings of patients represents genetic or environmental influences remains to be elucidated.

Specific haplotypes (whole chromosome or a large portion of chromosome) in the HLA region have been found to be associated with a two- to threefold increase in risk of nasopharyngeal carcinoma. The HLA loci involved are situated on the short arm of chromosome 6 and include A2, BW46, BW58, and DR9. HLA molecules perform tasks by presenting antigenic peptides to T cells which are responsible for immune responses. Polymorphic HLA molecules may vary in their ability to bind peptides. Therefore, the susceptibility to nasopharyngeal carcinoma may reflect the different capability of HLA haplotypes in mounting cellular-mediated immunity to Epstein–Barr virus infection. Furthermore, the existence of a highly significant recessive nasopharyngeal carcinoma susceptibility gene linked to the HLA region *has* also been reported.

Environmental factors

Nasopharyngeal carcinoma may be associated with the southern Chinese habit of feeding salted preserved fish to children. Traces of nitrosodimethylamine, found in salted fish, induced paranasal and nasal tumours when fed to weaning rats. Case–control studies carried out in Malaysia and Hong Kong also show a highly significant association between childhood consumption of salted fish and nasopharyngeal carcinoma.

The relationship between certain medicinal herbs and nasopharyngeal carcinoma has also been under recent investigation. A Chinese medicinal herb extract (*Wikstroemia indica*, found in southern China) promoted the induction of nasopharyngeal carcinoma in rats by the action of dinitrosopiperazine, a nitroso compound with relative predilection for nasopharyngeal epithelium. Chemicals present in the environment may therefore act as strong promoting factors in the development of nasopharyngeal carcinoma.

Recent reports also incriminate occupational exposure to combustion products as non-dietary risk factors for nasopharyngeal carcinoma but no obvious association between the disease and dust or exposure to chemical fumes have been observed. A significant threefold risk in smokers of 30 or more cigarettes per day have been reported, while living with a smoker during childhood also constitutes a significant and independent risk factor for nasopharyngeal cancer.

Epstein–Barr virus

Epstein–Barr virus is a herpesvirus discovered by Epstein in 1964 in a culture derived from a Burkitt's lymphoma biopsy. It is a ubiquitous virus with DNA of about 106 $\times 10^6$ Da. Latent infection by Epstein–Barr virus is present in many populations around the world, but the age at which primary infection occurs varies. The virus is transmitted horizontally by saliva. Early infection in childhood is associated with Burkitt's lymphoma and nasopharyngeal carcinoma in tropical Africa, and with nasopharyngeal carcinoma in southern China. Late infection in adolescence is associated with infectious mononucleosis in North America and Europe. Infection is accompanied by seroconversion with the development of specific antibodies to virus-determined antigens and the establishment of permanent immunity to reinfection.

Epstein–Barr virus and nasopharyngeal carcinoma

Epstein–Barr virus is associated with both Burkitt's lymphoma and nasopharyngeal carcinoma. The association of Epstein–Barr virus with nasopharyngeal carcinoma is strong and consistent. Sera derived from patients with nasopharyngeal carcinoma exhibit a characteristic pattern of antibodies to Epstein–Barr virus. In 1966, precipitating antibodies to Epstein–Barr virus-related antigen were demonstrated in sera from patients with nasopharyngeal carcinoma. Virus-associated nuclear antigen was demonstrated in tumour cells of patients with nasopharyngeal carcinoma in 1974. By 1976, the IgA class of antibodies to viral capsid antigen and early antigen had been identified in sera and were used clinically for the early diagnosis of nasopharyngeal carcinoma. More recently an antibody-dependent cellular cytotoxicity assay for the detection of antibodies against Epstein–Barr virus-specific membrane antigen has also been under investigation for use as serological test for early diagnosis and prognosis. The current and potential clinical applications of selected serological tests for the detection of Epstein–Barr virus-related nasopharyngeal carcinoma is summarized in Table 2. The aetiological relationship between Epstein–Barr virus and nasopharyngeal carcinoma have also been further strengthened by the discovery of viral DNA sequences in epithelial cells of these tumours by *in-situ* hybridization.

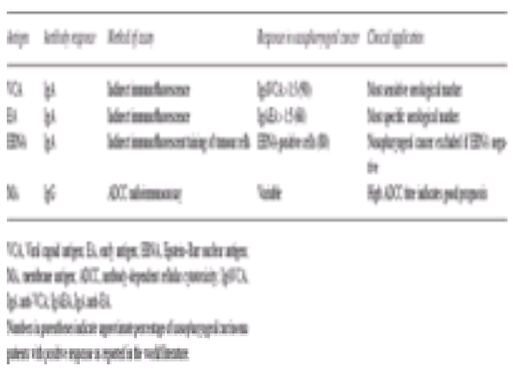


Table 2 Clinical application for selected Epstein–Barr virus specific antigen–antibody responses of patients with nasopharyngeal carcinoma

Pathogenesis

Patients with nasopharyngeal carcinoma have latent Epstein–Barr virus infection in both their tumours and blood lymphocytes. The Epstein–Barr virus genome and gene products are usually found in the differentiated as well as the undifferentiated types of nasopharyngeal carcinoma. All nasopharyngeal carcinomas express the Epstein–Barr virus encoded latent protein, Epstein–Barr nuclear antigen 1 and 60 to 90 per cent of them express latent membrane protein 1. Latent membrane protein 1 is a viral oncogene of Epstein–Barr virus, which is important in the pathogenesis of nasopharyngeal carcinoma as it induces epidermal hyperplasia and alters cellular growth control mechanisms.

Nasopharyngeal carcinoma is thought to develop as a result of multiple genetic changes induced by environmental factors and Epstein–Barr virus infections. Recent molecular genetic studies by the loss of heterozygosity studies and multiplex polymerase chain reaction analysis have demonstrated that the deletion of short arm of chromosome 3 is the most frequent genetic change in nasopharyngeal carcinoma (67 to 100 per cent). It is suggested that multiple tumour suppressor genes in this chromosome are inactivated during the development of nasopharyngeal carcinoma. A high percentage of loss of heterozygosity have also been demonstrated at chromosomes 9p and 11q (61 per cent and 54 per cent, respectively). Overexpression of oncogenes (e.g. bcl-2, an inner mitochondrial membrane protein which inhibits apoptosis or programmed cell death) as well as reactivation of telomerase activity are other mechanisms that have recently been proposed to be important in the tumorigenesis of nasopharyngeal carcinoma.

In a recently proposed conceptual tumorigenesis model (Fig. 1), genetic alterations is believed to occur in low grade precancerous lesions (nasopharyngeal intraepithelial neoplasia grades I and II) in which Epstein–Barr virus infection is absent. With latent Epstein–Barr virus infection, there is progression of low grade to high grade precancerous lesions of nasopharynx (nasopharyngeal intraepithelial neoplasia grade III). It is proposed that only the dysplastic cells infected with Epstein–Barr virus will further progress to form clinically important tumours and the cells without Epstein–Barr virus infection may eventually regress through apoptosis.

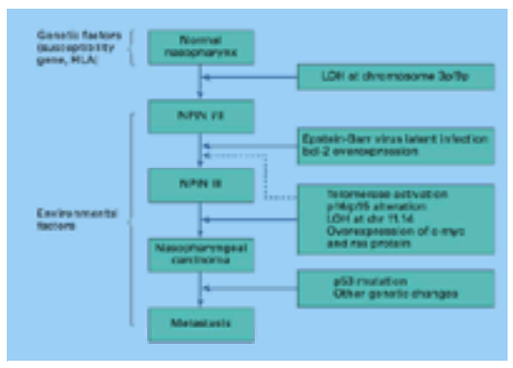


Fig. 1. A conceptual model for the tumorigenesis of nasopharyngeal carcinoma

Diagnosis

Anatomy of the nasopharynx

The nasopharynx is the uppermost portion of the aerodigestive tract. It functions as a conduit for air and for nasal and paranasal sinus secretions, and is lined by squamous, respiratory, and transitional epithelium. Cuboidal in shape, it lies above the level of the soft palate and posterior to the nasal septum and the posterior choana (Fig. 2). The sphenoid sinus, the lower clivus, the atlas of the cervical spine, and prevertebral muscles make up its posterosuperior wall. This is covered by pharyngeal mucosa, superior constrictor muscle, and the tough pharyngobasilar fascia. Laterally, the clinically important lateral wall consists of the eustachian tube opening, the lateral pharyngeal recess (fossa of Rosenmüller), and the cartilagenous medial end of the eustachian tube (torus tubarius) upon which the levator veli palatini muscle attaches. The fossa of Rosenmüller is a cleft-like pouch reaching to the parapharyngeal space through which runs the internal carotid artery, internal jugular vein, cranial nerves IX to XII, and sympathetic nerves (Fig. 3). The adjacent skull base contains the foramen lacerum and foramen ovale which provide the potential spread of tumour to the brain via the cranial nerves III, IV, V, and VI.

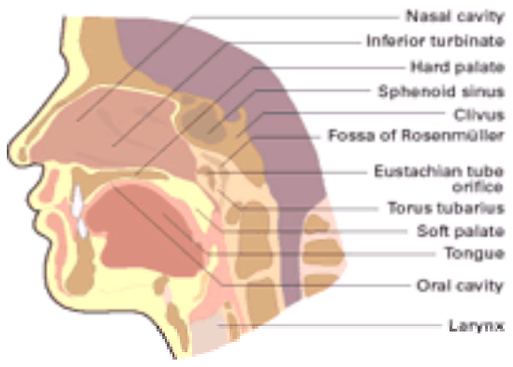


Fig. 2. Sagittal view of the nasopharynx.

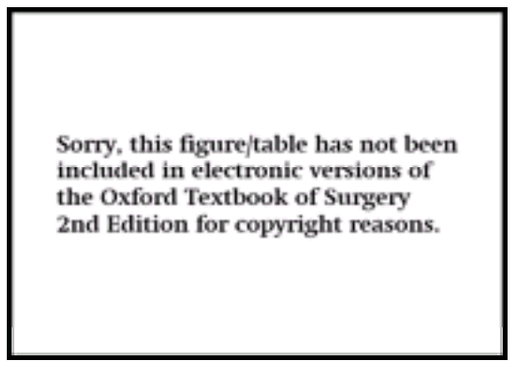


Fig. 3. Axial view of the nasopharynx showing the parapharyngeal space.

The nasopharyngeal mucosa overlying the posterior wall of the nasopharynx contains a collection of lymphoid tissue which enlarges before age 10 to become the adenoids. The lymphatic system include lateral retropharyngeal lymph nodes of Rouviere situated in the retropharyngeal space anterior to the pre-vertebral fascia. Clinically important first echelon of lymphatic drainage system are the upper deep cervical chain of nodes which lie along the upper internal jugular vein at or above the level of the digastric muscle. Lower down, the lower deep jugular nodes as well as nodes along the spinal accessory nerve are important lymphatic drainage system of the nasopharynx (Fig. 4).



Fig. 4. Metastatic cervical nodes in a 70-year-old man with nasopharyngeal carcinoma.

Clinical presentation

Common signs and symptoms associated with early nasopharyngeal carcinoma include nasal symptoms (blood-stained anterior or postnasal discharge, epistaxis, obstruction) and aural symptoms (serous otitis media, tinnitus, conductive hearing loss). Because the nasopharynx is not readily visible or accessible, the first sign ofnasopharyngeal carcinoma in more than 50 per cent of patients is enlarged metastatic cervical node(s) (Fig. 4). Owing to the proximity of the nasopharynx to the skull base, neurological symptoms such as diplopia, facial numbness, pain, ptosis, and hoarseness due to involvement of single or multiple cranial nerves III to XII and sympathetic nerves may be present in advanced disease. Headaches occur with intracranial disease, and trismus from oropharyngeal extension indicates extensive

local disease.

In general, metastasis first occurs by the lymphatics to jugulodigastric and posterior triangle cervical lymph nodes (Fig. 4). Bilateral cervical metastasis can occur. Distant metastasis to bone (frequently the spine), liver, and lungs is a sign of late or recurrent disease. Rarely, spread of disease to the axilla and inguinal lymph nodes may mimic lymphoma.

The diagnosis of nasopharyngeal carcinoma must therefore be considered in an adult presenting with any one of cervical lymphadenopathy, blood-stained nasal discharge, and serous otitis media.

Diagnostic approach

Following a history and physical examination with emphasis on the head and neck area, patients suspected of, or at risk of, developing nasopharyngeal carcinoma should undergo flexible fibreoptic nasopharyngoscopic examination as a routine, along with blood tests for IgA antibodies to Epstein–Barr virus capsid antigen and early antigen (Fig. 5).

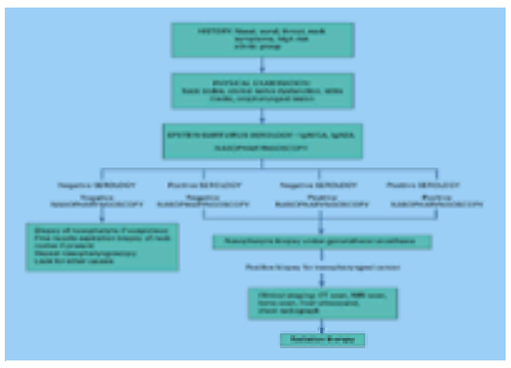


Fig. 5. Algorithm for the diagnosis of nasopharyngeal carcinoma.

Nasopharyngoscopy

The flexible fibreoptic nasopharyngoscope has made easy the complete examination of the nasopharynx in the doctor's office with minimal discomfort to the patient. Direct visualization of the nasopharynx by this means allows simultaneous biopsy of any tumours seen and close inspection of the nasopharyngeal mucosa for easy bleeding or presence of inflamed, haemorrhagic, granular mucosa, which may be the only subtle abnormalities present in a nasopharynx with submucosal infiltration by early nasopharyngeal carcinoma. Figure 6 shows the endoscopic view of an ulcerative and nodular tumour arising from the right posterolateral wall. As a rule, nasopharyngeal carcinoma is best diagnosed by punch biopsy of the primary tumour. The biopsy procedure can be done under topical anaesthesia with the endoscope and the biopsy forcep targeting the nasopharynx on different sides of the nasal septum. Excisional node biopsy can usually be avoided if the nasopharynx has been thoroughly examined and biopsied.

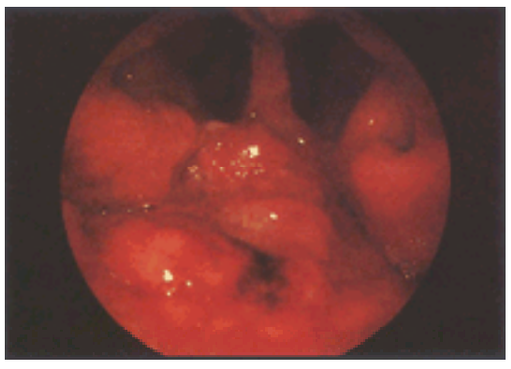


Fig. 6. Endoscopic view of nasopharyngeal carcinoma involving the right posterolateral wall.

Serology

The use of IgA antibodies to Epstein–Barr virus-related antigens is extremely useful in the diagnosis of nasopharyngeal carcinoma. Symptomatic patients with elevated levels of antibody to viral capsid antigen (IgAVCA) and/or early antigen (IgAEA) and an apparently normal nasopharynx should be admitted for examination under general anaesthesia and multiple endoscopic-guided biopsies of the nasopharynx. IgA to viral capsid antigen is a sensitive screening marker for nasopharyngeal carcinoma (97 per cent sensitivity, 67 per cent specificity) (Table 2), but levels may be falsely elevated in patients with no clinical evidence of nasopharyngeal carcinoma or other head and neck cancer. There is a growing belief that reactivation of Epstein–Barr virus infection, as indicated by elevated levels of this antibody, is a prerequisite for the development of nasopharyngeal carcinoma. The relative risk of developing this tumour is 50 to 80 times higher for people with an elevated level of IgA against viral capsid antigen than in the average population. These antibody levels can be elevated many months before the onset of clinical symptoms of nasopharyngeal carcinoma: patients with elevated antibody levels but a negative biopsy result should be kept under continuous surveillance. Only 60 per cent of patients with nasopharyngeal carcinoma have elevated levels of IgA to the viral early antigen, but nearly all of those in whom this antibody is present have the disease (sensitivity 79 per cent; specificity 97 per cent); hence, symptomatic and asymptomatic patients with increased levels of IgA against the early antigen must undergo extensive biopsy of the nasopharynx to confirm the diagnosis. Measurement of levels of IgA against viral capsid antibody has largely replaced the assay of non-specific IgG antibodies to capsid and early antigens as the most sensitive and reliable serological screening for nasopharyngeal carcinoma. The IgG antibody levels are usually non-specifically elevated in patients with a past history of Epstein–Barr virus infection.

The antibody-dependent cellular cytotoxity assay is occasionally used to measure levels of antibody against Epstein–Barr virus-related membrane antigen. High antibody-dependent cellular cytotoxity assay litre correlates with better prognosis. When lymph nodes are affected by metastases from an unknown primary tumour and nasopharyngeal carcinoma is suspected, tissue smears from fine needle aspiration smears of the node or from a biopsy can be stained for Epstein–Barr-related nuclear antigen. This confirms the presence of a neoplasm related to this virus when positive (Table 2). Nasopharyngeal carcinoma can be excluded when the stained tissue is negative for this antigen.

When the results of antibody tests become available, usually in a few days, sequential diagnostic steps including repeat nasopharyngoscopy whenever necessary, biopsy of the nasopharynx under local or general anaesthesia, and fine needle aspiration biopsy of any enlarged cervical node are undertaken to confirm or exclude the diagnosis.

Diagnostic imaging

Plain radiographic examination of the nasopharynx is only of historical interest. It has no role in the early diagnosis of nasopharyngeal carcinoma. Computed tomography (CT) scanning cannot replace direct examination of the nasopharynx by endoscopy for detection of early disease, as mild asymmetry of the nasopharynx is common. Once the histological diagnosis has been made, however, CT scanning and, in recent years, magnetic resonance imaging (MRI) is required for clinical staging and planning for radiotherapy. High resolution CT scanning with contrast is valuable in delineating the local extent of the tumour in relation to the parapharyngeal space, the retropharyngeal space, and the skull base. The CT scan appearance of a large, polypoid nasopharyngeal carcinoma extending to the choana is shown in Fig. 7. Both CT scan and MRI are also useful in detecting recurrent primary disease following radiotherapy. MRI is particularly helpful in distinguishing between fluid and tumour in opaque sinuses and between tumour and muscle. MRI has the additional benefit of providing sagittal as well as axial and

coronal images.

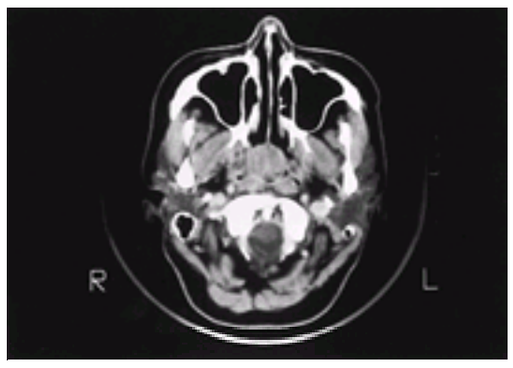


Fig. 7. CT scan appearance of a large, polypoid nasopharyngeal carcinoma extending to the choana in a 42-year-old woman.

High resolution ultrasound examination of the neck with its superior sensitivity over clinical examination (92 per cent versus 70 per cent) and its specificity when combined with a guided fine needle aspiration biopsy makes it an ideal investigation for differentiating benign from malignant nodes in patients suspected to have recurrent nasopharyngeal carcinoma.

Staging

There are several stage classifications for nasopharyngeal carcinoma. The old UICC (1987) and the new improved AJCC/UICC (1997) staging system for nasopharyngeal carcinoma are shown in Table 3. The new AJCC/UICC staging system takes into account the prognostic significance of parapharyngeal disease and large size (greater than 6 cm) metastatic cervical nodes which are important prognostic factors.

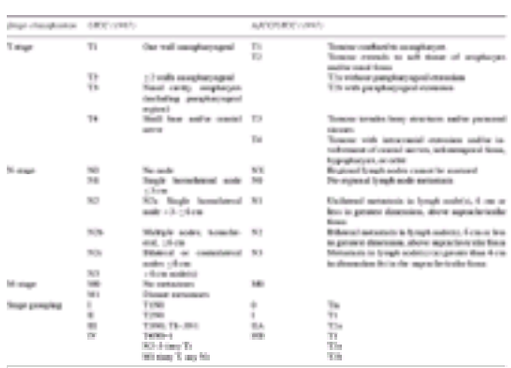


Table 3 Stage-classifications for nasopharyngeal carcinoma

Histopathology

The term nasopharyngeal carcinoma encompasses any carcinoma arising in the nasopharynx. However, in endemic regions where nasopharyngeal carcinoma is common, unless otherwise specified, the term nasopharyngeal carcinoma refers to undifferentiated carcinoma of nasopharynx and its variants. Undifferentiated carcinoma include non-keratinizing carcinoma (WHO type II) and undifferentiated carcinoma (WHO type III). There is increasing evidence that the two subdivisions represent a continuous spectrum with no meaningful borders. Squamous cell carcinoma of the nasopharynx (WHO type I) constitutes a greater proportion of malignant tumours of the nasopharynx in non-endemic areas. A recent report suggests that the squamous or keratinizing carcinoma contains Epstein–Barr virus DNA but to a lesser extent than the undifferentiated type.

Other nasopharyngeal tumours

Malignant nasopharyngeal tumours also include adenocarcinoma, adenoid cystic carcinoma, neuroendocrine carcinoma, lymphoma, and melanoma. Adenocarcinomas show glandular differentiation and usually arise from submucosal mucous and serous glands or their ducts. Specific morphological types include adenoid cystic carcinoma. Neuroendocrine carcinomas display neuroendocrine differentiation. They can be small cell carcinoma or sinonasal undifferentiated carcinoma. They are rare and aggressive neoplasms. Sinonasal lympho-mas are commonly diffuse large cell non-Hodgkin's lymphomas. They can be B-cell or T-cell lymphomas. Radiotherapy is the main treatment. Mucosal melanoma have been reported to occur in the nasopharynx. Melanomas generally stain positive for S-100 protein and negative for cytokeratin while nasopharyngeal carcinoma reacts in the opposite pattern.

Other tumours that may occupy the nasopharynx as a polypoid mass include olfactory neuroblastoma and angiofibroma. Olfactory neuroblastoma is a malignancy that requires surgery and radiotherapy to prevent recurrences or metastases. Angiofibroma is a histologically benign but locally aggressive tumour primarily affecting male adolescents. Nasopharyngeal angiofibroma is vascular and surgical removal usually requires a surgical approach to the skull base similar to that for nasopharyngeal carcinoma.

Treatment

Radiation as the primary treatment

External radiotherapy is the primary treatment for biopsy-proven nasopharyngeal carcinoma of all stages. Nasopharyngeal carcinoma is more radioresponsive than is squamous carcinoma of other head and neck sites, but a dose of more than 60 Gy is still required for curative treatment. At the Prince of Wales Hospital in Hong Kong, the entire nasopharynx and adjacent soft tissues and bones are usually treated to 60 Gy/24 fractions per 6 weeks by a three-field technique. Significant parapharyngeal disease receives an extra boost of 20 Gy in 2 weeks. High radiation doses to the spinal cord, pituitary gland, and the temporomandibular joint must be avoided in order to prevent radiation myelitis and trismus. Prophylactic as well as therapeutic radiation of the neck glands is routinely administered. An additional cervical irradiation is delivered with an electron beam to any palpable nodes that remain after external radiotherapy. Persistent localized disease in the nasopharynx is treated with intracavitary brachytherapy with iridium-192 to 24 Gy in three fractions over 2 weeks. Patients with nodes more than 5 cm in size receive adjuvant chemotherapy with cis-platinum and 5-fluorouracil for two to three courses, 3 weeks apart, prior to radiotherapy. In an effort to improve the results of radiotherapy, a multidisciplinary approach to the treatment of nasopharyngeal carcinoma is frequently utilized (Fig. 8).



Fig. 8. Multidisciplinary approach to the treatment of nasopharyngeal carcinoma at the Prince of Wales Hospital, Hong Kong.

From 1984 to 1989, 903 patients with localized nasopharyngeal carcinoma received their first treatment with radiotherapy at the Prince of Wales Hospital, The Chinese University of Hong Kong. Over 95 per cent of the patients had undifferentiated carcinoma (WHO III). Intracavitary brachytherapy with iridium-192 was required for 99 local persistences. With a mean and median follow up of 5.3 years, the 5-year acturial survival and disease-free survival rates were 66 per cent and 57 per cent respectively. The male sex, skull base and cranial nerve involvement, advanced Ho N-level, presence of fixed or partially fixed nodes, or presence of contralateral nodes significantly determined survival and distant metastasis rates. Parapharyngeal tumour, skull base, and cranial nerve involvement, advanced age, and male sex worsened local control.

Role of chemotherapy

Treatment of nasopharyngeal carcinoma results in an overall 5-year survival rate of over 80 per cent for Ho's stage I and over 70 per cent for Ho's stage II disease. However, the majority of patients presented with neck nodes have advanced stage disease. The presence of bulky cervical lymph node metastases (*greater than* 3 cm) or supraclavicular lymph node metastases have been observed to be strongly associated with a high rate of distant metastases, local failure after radiotherapy, and a 5-year survival rate of only 10 to 40 per cent.

The use of cisplatin-based neoadjuvant and adjuvant chemotherapy to treat locoregionally advanced nasopharyngeal carcinoma have resulted in consistently high response rates of 38 to 91 per cent; however, few prospective randomized trials have demonstrated an improvement in overall survival. A retrospective study of 422 patients with nasopharyngeal carcinoma with cervical nodal metastases treated between 1984 and 1987 showed no apparent difference in survival between the 169 patients who received neoadjuvant chemotherapy (cisplatin and 5-fluorouracil for two to three courses prior to radiotherapy) and the 253 patients who were treated by radiotherapy alone. However, a recent Head and Neck Intergroup study for stages III and IV nasopharyngeal carcinoma (UICC/AJCC) showed significantly better survival at 2 years for patients who received concurrent cisplatin and adjuvant cisplatin-5-fluorouracil as compared with those who received radiotherapy alone. The radiotherapy technique was 2 Gy per fraction to 70 Gy in 7 weeks. In the chemotherapy–radiotherapy arm, cisplatin 100 mg/m^2 was given every 3 weeks for three doses during radiotherapy, and cisplatin 80 mg/m^2 on day 1 and 5-fluorouracil 1 g/m^2 on days 1 to 4 was given for three cycles after concurrent chemotherapy–radiotherapy was completed. One hundred and thirty-eight patients were evaluable for survival at a median follow up of 4 years. The median progression-free survival was 13 months in the radiotherapy arm compared with 52 months in the concurrent arm. The 2-year survival was 55 per cent and 80 per cent, respectively.

Surgical treatment of recurrent nasopharyngeal carcinoma

Although radiotherapy is the primary treatment for nasopharyngeal carcinoma of all stages, local, regional, and distant metastases often remain after such therapy: approximately 50 per cent of patients have recurrent disease 5 years after treatment. For early stage disease (T1,T2), the 5-year relapse-free survival is 75 per cent; detection and treatment of recurrent disease is therefore important.

Re-irradiation at the primary site renders 10 to 25 per cent of patients disease-free, but is associated with complications, including radiation myelitis, radiation encephalopathy, cranial nerve palsy, otitis media, trismus, and osteoradionecrosis. In 1988, Fee reported surgical treatment of nine patients with recurrent nasopharyngeal carcinoma, using a transpalatal approach to *en bloc* nasopharyngectomy with transcervical protection of the internal carotid artery, with good early results.

Since 1986 we have carried out more than 30 nasopharyngectomies for recurrent nasopharyngeal carcinoma. Initially we used a transoropalatal approach in which the internal carotid artery in the parapharyngeal space at the lateral limit of resection is protected by transoropharyngeal dissection (Fig. 9). We raised the entire soft and hard palate mucoperiosteal flap pedicled on a single greater palatine artery to obtain maximum exposure of the nasopharynx after removal of the posterior two-thirds of the palatal bone (Fig. 10). In a typical nasopharyngectomy, the entire nasopharyngeal mucosa and the underlying superior constrictor muscle, pharyngobasilar fascia, prevertebral fascia and muscles, and the cartilaginous portion of the eustachian tube are resected *en bloc*. The resultant defect is skin grafted. Figure 11 shows the CT appearance after transoropalatal right nasopharyngectomy for recurrent nasopharyngeal carcinoma. Subsequently, we increasingly used the mandibular swing approach (mandibulotomy) to improve the access to the nasopharynx. The mandibular swing approach allows easier resection of tumours with parapharyngeal extension and better protection of the carotid vessels from the neck. Figure 12 shows that the mid-line mandibulotomy allows division of the soft tissues of the floor of mouth posteriorly to reach the oropharynx. Figure 13 shows that retraction of the unilaterally pedicled palatal flap to the right permits good access to the nasopharyngeal cancer. In recent years, we have also included the maxillary swing approach for access to the nasopharynx. The maxillary swing approach utilizes the standard Weber–Fergusson incision (Fig. 14) with osteotomies carried out as for subtotal maxillectomy. However, the anterior wall of the maxilla is left attached to the cheek flap which is swung out of the way as a composite flap (Fig. 15). After nasopharyngectomy, the bone cuts are reattached and fixed with miniplates (Fig. 16). The maxillary swing allows excellent access to one side of the nasopharynx and bone healing is not compromised by previous radiotherapy. In general, curative nasopharyngectomy is not possible in patients with extensive parapharyngeal disease and intracranial extension. The presence of cervical lymph node metastasis requires a combined nasopharyngectomy and radical neck dissection. The results of nasopharyngectomy for recurrent nasopharyngeal carcinoma by various groups using different surgical approaches are listed in Table 4.

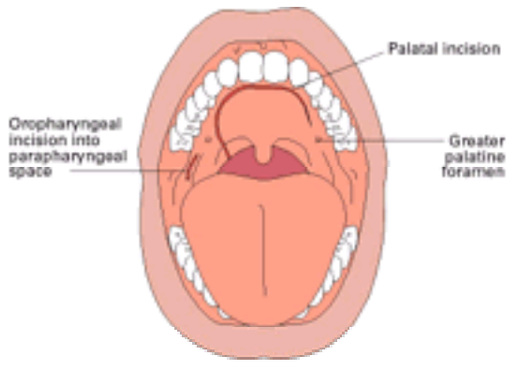


Fig. 9. Incisions for transoropalatal nasopharyngectomy.

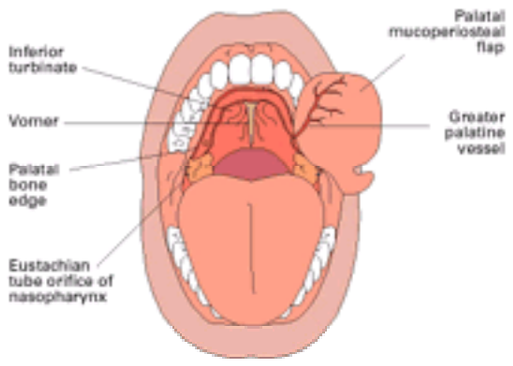


Fig. 10. Access to the nasopharynx via a unilaterally pedicled mucoperiosteal palatal flap.

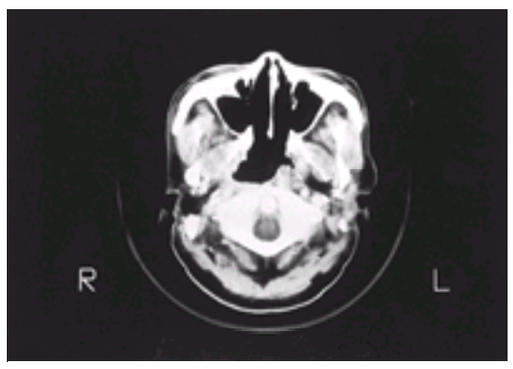


Fig. 11. CT scan appearance after transoropalatal right nasopharyngectomy for recurrent nasopharyngeal carcinoma in a 33-year-old man.

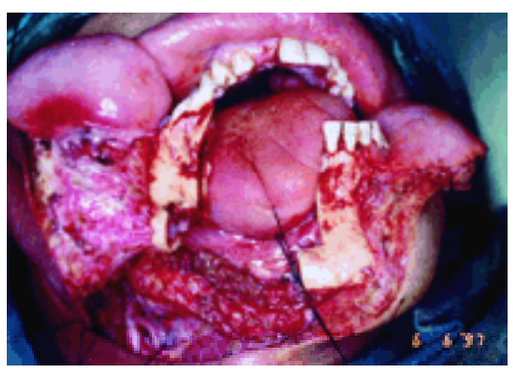


Fig. 12. The mandibular swing approach with median mandibulotomy.

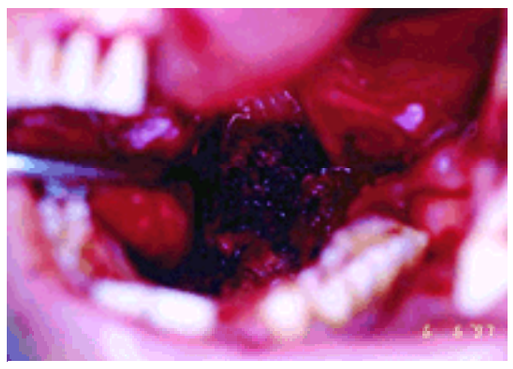


Fig. 13. Retraction of the unilaterally pedicled palate flap allows access to the nasopharyngeal cancer.

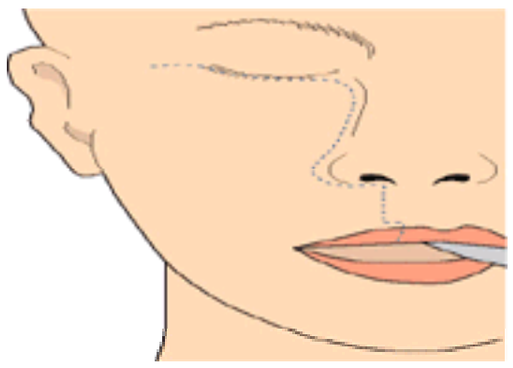


Fig. 14. Incision for the maxillary swing approach.



Fig. 15. The maxillary swing approach with maxilla attached to the cheek flap.

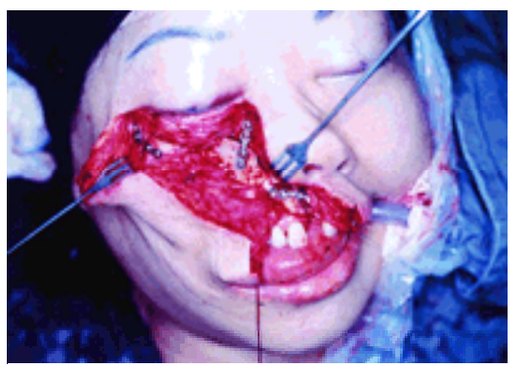


Fig. 16. Reattachment and fixation of maxilla osteotomies with miniplates.

Author	*Surgical approach*	*Local control*
Fitch 1985, 1988	Infratemporal	0/7 (0 per cent)
Fee 1988, 1990, 1991	Transpalatal	0/2 (0 per cent)
Panje, Gross 1987	Transtemporal	0/7 (0 per cent)
Fish 1985, 1988	Infratemporal	5/6 (83 per cent)
Sooy 1956	Transoral	3/6 (50 per cent)
To 1988	Transpalatal	4/9 (44 per cent)
Fee 1988, 1990, 1991	Transpalatal	3/12 (25 per cent)
Wei 1991, 1995	Transmaxillary	(42 per cent)
Marcus 1996	Transmandibular	4/6 (67 per cent)
King 1999	Transpalatal, trans-mandibular, trans-maxillary	15/28 (54 per cent)

Table 4 Reported results of nasopharyngectomy for recurrent nasopharyngeal carcinoma

Radical neck dissection for regional recurrence

Regional recurrence after radiotherapy presenting as persistent or recurrent cervical lymphadenopathy is more common in patients with bulky cervical disease. Recurrent cervical disease can be confirmed by fine needle aspiration biopsy or by excisional biopsy of a suspicious cervical node. In the absence of distant metastasis and when biopsy reveals absence of disease at the nasopharynx, radical neck dissection for cure is indicated.

Between 1986 and 1991, we carried out radical neck dissection in 38 patients with nasopharyngeal carcinoma and recurrent cervical metastasis after radiotherapy. There were no in-hospital deaths. Three of 38 patients required a myocutaneous flap for neck coverage. As pectoralis major and latissimus dorsi myocutaneous flaps become commonly available for soft tissue coverage of the neck whenever cervical skin becomes deficient due to surgery or infectious complications, the risk of carotid artery exposure and rupture after radical neck dissection in heavily irradiated necks can be minimized.

Less than a formal radical neck dissection (such as modified radical neck dissection with preservation of the spinal accessory nerve) is not recommended for these patients because of the possible involvement of the spinal accessory chain of lymph nodes and extranodal disease. In our experience, radical neck dissection results in minimal shoulder dysfunction and is well tolerated, even in heavily irradiated patients. Figure 17 shows a patient with well-healed parallel incisions (McFee) after a left radical neck dissection for recurrent cervical metastasis after radiotherapy for nasopharyngeal carcinoma.

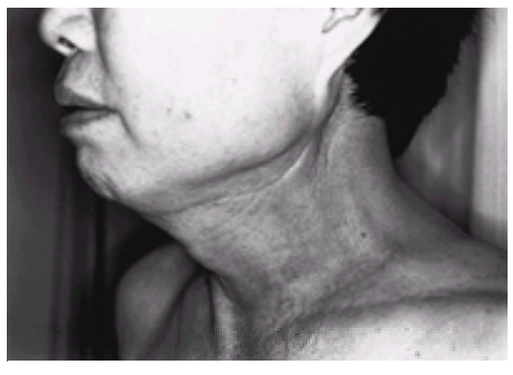


Fig. 17. Appearance of a patient after radical neck dissection by McFee incision.

High dose re-irradiation versus nasopharyngectomy

High-dose (³60 Gy) re-irradiation using mainly external beams versus nasopharyngectomy in salvaging local failures of nasopharyngeal carcinoma was retrospectively reviewed in 903 patients with non-disseminated nasopharyngeal carcinoma treated with radiotherapy between 1984 and 1989 at the Prince of Wales Hospital. One hundred and seventy-six had local failures comprising nine persistences and 167 recurrences. In 10 patients local failures were accompanied by distant metastases within 2 months. Forty-three patients received only palliative treatments. The remaining 123 patients were treated with either re-irradiation to high dose (³60 Gy) using mainly external photon beams (*n* = 103) or nasopharyngectomy with/without radical neck dissection with/without postoperative radiotherapy (*n* = 20). Nasopharyngectomy was performed via the transoro-palatal approach, the mandibular swing approach or the maxillar swing approach. With a median follow-up of 20 months (range 2.5 to 81 months) since the diagnosis of local failure, the actuarial 5-year overall survival, further relapse-free survival and free-from-local-tumour rates were 9.4, 11.5, and 18.7 per cent, respectively, for the 123 patients treated by either high-dose re-irradiation (*n* = 103) or nasopharyngectomy (*n* = 20). Re-irradiation to high dose (³60 Gy) mainly by external photon beams achieved a 5-year overall survival of 7.6 per cent and 5-year local control of 15.2 per cent with significant complications. Nasopharyngectomy (with/without neck dissection, with/without postoperative radiotherapy) was associated with earlier recurrent T-stages (mostly rT1 and rT2) and better survival and local control than re-irradiation. In a further review of 31 patients treated at Prince of Wales Hospital by nasopharyngectomy, the 5-year acturial overall survival, acturial disease-free survival and tumour control were 47 per cent, 42 per cent, and 43 per cent, respectively. More long-term follow-up of these surgically treated patients will be required for valid conclusions. Currently, we prefer nasopharyngectomy and postoperative radiotherapy over re-irradiation in the treatment of suitable patients with recurrent nasopharyngeal carcinoma.

Further reading

Chan ATC, Teo PML, Leung TWT, Johnson PJ. The role of chemotherapy in the management of nasopharyngeal carcinoma. *Cancer* 1998; **82**: 1003–12. [An excellent review of chemotherapy in nasopharyngeal carcinoma.]

Chen J-Y *et al.* Antibodies to Epstein-Barr virus-specific DNase in patients with nasopharyngeal carcinoma and control groups. *Journal of Medical Virology* 1987; **23**: 11–21. [EBV and serum

antibodies.]

Fee WE, Jr, Gilmer PA, Goffinet DR. Surgical management of recurrent nasopharyngeal carcinoma after radiation failure at the primary site. *Laryngoscope* 1988; **98**: 1220–6. [First detailed report of surgery for nasopharyngeal carcinoma.]

Henle G, Henle W. Epstein-Barr virus-specific serum antibodies as an outstanding feature of nasopharyngeal carcinoma. *International Journal of Cancer* 1976; **17**: 1–7. [EBV and serum antibodies.]

Ho JHC. An epidemiologic and clinical study of nasopharyngeal carcinoma. *International Journal of Radiation, Oncology, Biology and Physics* 1978; **4**: 181–98. [A good review on nasopharyngeal carcinoma.]

Huang DP, Ho JHC, Henle W, Henle G, Saw D, Lui M. Presence of EBNA in nasopharyngeal carcinoma and control patient tissues related to EBV serology. *International Journal of Cancer* 1978; **22**: 266–74. [EBNA and nasopharyngeal carcinoma.]

King WK, Teo PML, Li AKC. Patterns of failure after radical neck dissection for recurrent nasopharyngeal carcinoma. *American Journal of Surgery* 1992; **164**: 599–602. [Good report on neck dissection in nasopharyngeal carcinoma.]

King WWK, Ku PKM, Mok CO, Teo PTL. Nasopharyngectomy in treatment of recurrent nasopharyngeal carcinoma—a twelfth-year experience. *Head and Neck* (in press). [An update on surgery for nasopharyngeal carcinoma.]

Klein G, Giovanella BC, Lindahl T, Fialkow PJ, Singh S, Stehlin J. Direct evidence for the presence of Epstein-Barr virus DNA and nuclear antigen in malignant epithelial cells from patients with poorly differentiated carcinoma of the nasopharynx. *Proceedings of the National Academy of Sciences, USA* 1974; **71**: 4737–41. [EBNA and nasopharyngeal carcinoma.]

Lee JCK. Precancerous changes. In: Van Hasselt CA, Gibb AG, eds. *Nasopharyngeal carcinoma*, pp. 37–46. The Chinese University Press, Hong Kong, 1991. [An excellent look at the histopathology of nasopharyngeal carcinoma.]

Liu MY, Chou WH, Nutter L, Hsu MM, Chen JY, Yang CS. Antibody against Epstein-Barr virus DNA polymerase activity in sera of patients with nasopharyngeal carcinoma. *Journal of Medical Virology* 1989; **28**: 101–5. [EBV and serum antibodies.]

Lu SJ *et al.* Linkage of a nasopharyngeal carcinoma susceptibility : locus to the HLA region. *Nature* 1990; **346**: 470–47. [Molecular studies on nasopharyngeal carcinoma.]

Lucarini JW. Nasopharyngeal cancer. In:Morris PJ, Malt RA, eds. *Oxford textbook of surgery. Head and neck surgery*, Vol. 2, 39.6, pp. 2233–6. Oxford University Press, 1997. [Good review on nasopharyngeal carcinoma.]

McGuire LJ, Lee JCK. The histopathologic diagnosis of nasopharyngeal carcinoma. *Ear, Nose and Throat Journal* 1990; **69**: 229–36. [Good review on the histopathology of nasopharyngeal carcinoma.]

Morton RP, Liavaag PG, McLean M, Freeman JL. Transcervico-mandibulo-palatal approach for surgical salvage of recurrent nasopharyngeal cancer. *Surgical Salvage of Nasopharyngeal Cancer* 1996; **18**: 352–8. [Very good report on surgery for nasopharyngeal carcinoma.]

Neel HB III, Taylor WF. New staging system for nasopharyngeal carcinoma. Long-term outcome. *Archives of Otolaryngology and Head Neck Surgery* 1989; **115**: 1293–303. [Staging system in nasopharyngeal carcinoma.]

Nonoyama M, Huang CH, Pagano JS, Klein G, Singh S. DNA of Epstein-Barr virus detected in tissue of Burkitt's lymphoma and nasopharyngeal carcinoma. *Proceedings of the National Academy of Sciences, USA* 1973; **70**: 3265–8. [EBV and nasopharyngeal carcinoma.]

Old LJ *et al.* Precipitating antibody in human serum to an antigen present in cultured Burkitt's lymphoma cells. *Proceedings of the National Academy of Sciences, USA* 1966; **56**: 1699–704. [EBV and serum antibodies.]

Tang WP, Huang PG, Zhao ML, Liao SL, Zeng Y. *Wikstroemia indica* promotes development of nasopharyngeal carcinoma in rats initiated by dinitrosopiperazine. *Journal of Cancer Research and Clinical Oncology* 1988; **114**: 429–31.[Promoting chemicals in nasopharyngeal carcinoma.]

Teo P *et al.* A clinical study of 407 cases of nasopharyngeal carcinoma in Hong Kong. *International Journal of Radiation, Oncology, Biology and Physics* 1989; **17**: 515–30. [Excellent report on radiation for nasopharyngeal carcinoma.]

Teo PML, Leung TWT, Chan ATC, Yu P, Lee WY, Leung SF, Kwan WH, Johnson PJ. A retrospective study of the use of cisplatinum-5-fluorouracil neoadjuvant chemotherapy in cervical-node-positive nasopharyngeal carcinoma (NPC). *Oral Oncology European Journal of Cancer* 1995; **31B**(6): 373–9. [Role of chemotherapy in nasopharyngeal carcinoma.]

Teo P, Lee WY, Yu P. The prognostic significance of parapharyngeal tumour involvement in nasopharyngeal carcinoma. *Radiotherapy and Oncology* 1996*a*; **39**: 209–21. [Significance of parapharyngeal disease in nasopharyngeal carcinoma.]

Teo PML, Kwan WH, Yu P, Lee WY, Leung SF, Choi P. A retrospective study of the role of intracavitary brachytherapy and prognoic factors determining local tumour control after primary radical radiotherapy in 903 non-disseminated nasopharyngeal carcinoma patients. *Clinical Oncology* 1996*b*; **8**: 160–6. [Role of brachytherapy in nasopharyngeal carcinoma.]

de The G. Cancer of the nasopharynx. Viral relationships. In: Ariyan S, ed. *Cancer of the head and neck*, pp. 317–24. CV Mosby, St Louis, 1987. [Role of virus in nasopharyngeal carcinoma.]

Van Hasselt CA. Surgery. In Van Hasselt CA, Gibb AG, eds. *Nasopharyngeal carcinoma*, pp. 235–9. The Chinese University Press, Hong Kong, 1991. [An excellent textbook on nasopharyngeal carcinoma.]

Yu MC, Garabrant DH, Huang TB, Henderson BE. Occupational and other non-dietary risk factors for nasopharyngeal carcinoma in Guangzhou, China. *International Journal of Cancer* 1990; **45**: 1033–9. [Environmental risk factors for nasopharyngeal carcinoma.]

45.7 Cancer of the larynx and hypopharynx

Mark C. Weissler

Laryngeal cancer
Pertinent laryngeal anatomy and physiology
Etiology and demographics of laryngeal cancer
Growth patterns of laryngeal cancer
Signs and symptoms of laryngeal cancer
Evaluation of patients with suspected laryngeal cancer
Staging of laryngeal cancer
Treatment of laryngeal cancer
Hypopharyngeal cancer
Anatomy and physiology of the hypopharynx
Etiology and demographics of hypopharyngeal cancer
Growth patterns of hypopharyngeal cancer
Signs and symptoms of hypopharyngeal cancer
Evaluation of patients with suspected hypopharyngeal cancer
Treatment of hypopharyngeal cancer
Prognosis of patients with cancers of the hypopharynx
Treatment of the neck in patients with cancers of the larynx and hypopharynx
Follow-up of patients with cancer of the larynx and hypopharynx
Further reading

Laryngeal cancer

Pertinent laryngeal anatomy and physiology

The larynx serves three main functions. Its primary function at the junction of the respiratory and digestive tracts is to protect the airway from food and saliva. It also serves some respiratory function, opening and closing with respiration, and affording the ability to cough and perform the Valsalva maneuver. Finally the larynx allows phonation. Protection of the airway is the most vital and basic laryngeal function.

The larynx is divided into three portions: the glottis, the supraglottis, and the subglottis. The supraglottic larynx consists of the epiglottis, aryepiglottic folds, false vocal cords (ventricular bands), superior aspect of the laryngeal ventricles, and the arytenoids. It stops inferiorly at the level of the vestibular apex. The suprahyoid posterior surface of the epiglottis, along with the aryepiglottic folds and arytenoids, is sometimes referred to as the epilarynx. The glottic larynx consists principally of the true vocal cords, and anterior and posterior commissures. The subglottic larynx extends from a plane 1 cm inferior to the ventricular apex to the inferior edge of the cricoid cartilage.

The larynx contains three paired cartilages: the arytenoid, cuneiform, and corniculate cartilages. Of these only the arytenoids are clinically important; the other two pairs act simply as batons in the free edge of the aryepiglottic fold. There are also three unpaired cartilages: the thyroid, cricoid, and epiglottic cartilages. The thyroid and cricoid cartilages are composed of hyaline cartilage and begin to ossify in early adulthood. The epiglottic cartilage consists of elastic cartilage.

The cricoarytenoid and cricothyroid joints are true synovial joints. The larynx is covered by pseudostratified, columnar, ciliated respiratory epithelium, except for the free edge of the true vocal fold and the edges of the epiglottis and aryepiglottic folds which are exposed to constant trauma and become covered with a non-keratinizing squamous epithelium.

Blood is supplied to the larynx through branches of the superior and inferior thyroid arteries that anastomose with each other. The supraglottic larynx is richly supplied with lymphatics, which drain bilaterally through the thyrohyoid membrane to the upper and middle jugular chain of lymph nodes. The vocal cord is a complex multilayered structure consisting of an overlying epithelium, a lamina propria, and the underlying vocalis muscle. The lamina propria in turn is composed of three layers: the superficial layer, or Reinke's space, is composed of loose connective tissue, the middle layer is composed mostly of elastic fibers, while the deep layer is composed of dense connective tissue. Together, the intermediate and deep layers compose the vocal ligament. Clinically the epithelium covering the true vocal cord is separated from the underlying vocal ligament by the thin potential space of loose connective tissue known as Reinke's space. Classically it has been taught that the glottic larynx has very sparse lymphatic supply and this accounts for the tendency of glottic cancer to develop regional metastases only at more advanced stages. The lymphatics of the subglottic larynx tend to drain by way of the cricothyroid membrane to the middle and lower jugular nodes, as well as to the paratracheal and tracheo-esophageal nodes.

The motor innervation of the intrinsic laryngeal musculature is supplied by two branches of the vagus nerve: the recurrent laryngeal and superior laryngeal nerves. The superior laryngeal nerve has two branches: the internal and external laryngeal nerves. All the intrinsic muscles except for the cricothyroideus are supplied by the recurrent laryngeal nerve. The cricothyroideus is supplied by the external laryngeal branch of the superior laryngeal nerve.

Sensory innervation to the larynx is also supplied by the vagus nerve. The region below the true vocal folds is supplied by the recurrent laryngeal nerve, while the supraglottic larynx is supplied by the internal laryngeal branch of the superior laryngeal nerve. These two nerves communicate with each other via the anastomosis of Galen. On the left, the recurrent laryngeal nerve passes around the aortic arch lateral to the ligamentum arteriosum. On the right the recurrent nerve passes around the subclavian artery. Rarely, the right recurrent nerve is non-recurrent, passing directly from the vagus to the larynx. In such cases, the origin of the right subclavian artery from the aorta is anomalous.

The only true abductor of the true vocal folds is the posterior cricoarytenoid muscle, which is supplied by the recurrent laryngeal nerve. The remaining intrinsic musculature adducts or tenses the true vocal folds.

Etiology and demographics of laryngeal cancer

Laryngeal cancer is largely a preventable disease, related to the use of tobacco products. The following statistics were collected by the American Cancer Society and apply to the United States of America in 1998. Laryngeal cancer will account for approximately 11 100 new cancer cases and 4300 cancer deaths in 1998. This yields a gross mortality rate of 39 per cent. Laryngeal cancer accounts for about 0.9 per cent of new cancers and 0.8 per cent of cancer deaths each year. To put this in perspective, there will be 180 300 new cases of breast cancer with 43 900 cancer deaths and 171 500 new cases of lung cancer with 160 100 cancer deaths in 1998. The gross mortality rates for breast and lung cancer in the United States are 24 and 93 per cent, respectively.

Survival rates for laryngeal cancer have remained fairly stable over the past 20 years in the United States. The 5-year relative survival rate for laryngeal cancer was 66 per cent in 1974 to 1976 and 66 per cent in 1989 to 1994.

Laryngeal cancer occurs 1.3 times more frequently in men than in women, and this ratio is decreasing. Ten years ago, laryngeal cancer occurred 4.4 times more frequently in men than women. This change is due to a change in the smoking habits of American men and women.

Although smoking is by far the greatest risk factor for the development of laryngeal cancer, others include alcohol abuse, radiation exposure, asbestos exposure, and genetic factors. Genetic predisposition to the development of a variety of cancers is becoming more and more clear. This may take the form of inheritance of altered tumor suppressor genes such as p53 and p16, or the inheritance of specific carcinogen-metabolizing enzyme systems which either are less efficient at metabolizing carcinogens or result in more carcinogenic metabolic products. This is an area of active research.

Growth patterns of laryngeal cancer

Cancers confined to the vocal ligament have an excellent prognosis, partly because of the sparse lymphatic drainage of the area, and partly because of the early clinical symptom of hoarseness which often alerts patients to see a physician during early stages of disease. The clinical assessment of vocal fold mobility is mandatory in the assessment of glottic cancers. Even early, superficial tumors that spread to the anterior commissure tend to remain confined to the cord without invading the

cartilage, as long as there is no superior or inferior extension. A dense condensation of connective tissue (Broyle's ligament) consisting of the confluence of the vocal ligament, upper end of the conus elasticus, thyroepiglottic ligament, and the internal perichondrium of the thyroid cartilage helps to prevent direct involvement of thyroid cartilage. Lesions of the anterior commissure which spread superiorly to the inferior aspect of the epiglottis usually invade the thyroid cartilage. Involvement of the anterior commissure may limit the effectiveness of radiation therapy. Fixation of the vocal cord is an ominous clinical finding; it is usually due to direct invasion of the thryoarytenoid (vocalis) muscle by cancer. Such tumors have free egress to the paraglottic space and tend to spread to the cervical lymph nodes. Transglottic tumors, those involving both the supraglottic and glottic larynx by continuous spread across the apex of the ventricle, spread outward to involve the paraglottic space through the deficiency at this level between the conus elasticus and quadrangular ligament. Once the paraglottic space is involved, cancer may invade the thyroid cartilage or spread outside the larynx through the cricothyroid membrane. Most such neoplasms require total laryngectomy for their complete removal.

Supraglottic cancers have a propensity for bilateral cervical metastases. Those arising below the level of the hyoid bone may penetrate the multiple small dehiscences in the epiglottic cartilage and invade the pre-epiglottic space. Those arising above the level of the hyoid bone are above the level of the pre-epiglottic space and tend to have a better prognosis.

Solely subglottic cancers are rare; however, glottic cancer may spread inferiorly to involve the subglottis. Such cancers may spread outside the larynx through the cricothyroid ligament, with direct tumor extension into the thyroid gland. A hemithyroidectomy on the side of the lesion should be performed at the time of surgical resection. Subglottic neoplasms tend to metastasize to the paratracheal and tracheo-esophageal lymph nodes, which must be removed at the time of neck dissection. Subglottic cancers, or others with subglottic extension, tend to have a greater incidence of peristomal local recurrence after total laryngectomy.

Signs and symptoms of laryngeal cancer

Glottic cancers cause hoarseness early in the course of the disease; only minimal irregularity of the vocal fold results in a marked degree of voice change. This, together with the sparse lymphatic drainage of the glottic larynx with little early potential for regional or distant metastasis, accounts for the good prognosis of early glottic cancer. Early glottic cancers appear as red (erythroplakia) or white (leukoplakia) patches on the vocal folds. There may be a mass or ulceration of the vocal folds. Although early glottic cancers do not impair gross vocal fold motion, stroboscopy shows impairment of the superficial mucosal wave pattern. Advanced glottic cancers present with symptoms of dyspnea, dysphagia, odynophagia, pain, bleeding, and a neck mass. Impaired vocal fold motion with a fixed hemilarynx and a large exophytic and/or ulcerated mass may be seen. There may be evidence of spread to cervical lymph nodes, including the prelaryngeal or delphian node. Most commonly, glottic cancers involve the middle and inferior jugular lymph nodes.

In some patients, the first sign of disease is acute airway obstruction. Controversy has surrounded the proper handling of these patients. Some authorities feel that a tracheotomy increases the risk of a stomal recurrence of the cancer secondary to direct tumor seeding in the tracheotomy wound. Others have suggested that patients with large tumors, who are likely to present with acute airway obstruction, are more likely to have subglottic extension and involvement of the paratracheal and tracheo-esophageal lymph nodes, and that these factors are more directly related to stomal recurrence.

Emergency laryngectomy, intubation in the intensive care unit until laryngectomy can be undertaken, and laser debulking of the tumor have all been advocated as methods by which tracheotomy can be avoided in patients with an obstructing laryngeal cancer. Certainly, if tracheotomy is undertaken, laryngectomy should follow within a week or two, and postoperative irradiation should include a boost dose to the tracheostoma and upper mediastinum.

Supraglottic cancers may be more silent than glottic cancers and, for that reason, present at a more advanced stage. Early symptoms are more likely to relate to painful deglutition (odynophagia) than to hoarseness or dyspnea. Supraglottic cancers have a marked tendency to metastasize bilaterally through the thyrohyoid membrane to the jugular lymph nodes.

Primary subglottic cancers are rare. They tend to present with dyspnea.

Evaluation of patients with suspected laryngeal cancer

All physicians responsible for the evaluation of patients with possible laryngeal cancer should become adept in the techniques of indirect laryngoscopy. The alternative technique of fiberoptic flexible laryngoscopy is simple to learn, although the equipment is expensive and fragile. Indirect laryngoscopy and fiberoptic laryngoscopy are the basis for the diagnosis and monitoring of patients with laryngeal cancer.

A complete head and neck examination should be undertaken as there is a significant incidence of second primary cancers in this group of patients. Most second cancers will be smoking and alcohol related and will be found in the upper aerodigestive tract, lung, or esophagus. Twenty per cent of these patients will harbor a second synchronous or metachronous cancer of the aerodigestive tract. Another commonly quoted figure is that 7 per cent per year will develop a second cancer of the upper aerodigestive tract. Careful inspection of the nasal cavity, nasopharynx, oral cavity, oropharynx, hypopharynx, larynx, and both sides of the neck should be performed in an attempt to define the limits of the disease, and to detect second primary cancers and regional cervical metastases. Patients should be questioned about recent bone pain, neurologic changes such as seizure disorders, dyspnea, hemoptysis, dysphagia, odynophagia, and weight loss.

A chest radiograph and liver function studies (aspartate aminotransferase, alanine aminotransferase, g-glutamyl transferase, alkaline phosphatase, bilirubin, lactic dehydrogenase) should be obtained in all patients. If there is a history of recent unexplained bone pain or elevated serum alkaline phosphatase or serum calcium level, a bone scan should be obtained. A history of recent central nervous system derangement is an indication for CT scan of the brain. If the chest radiograph is abnormal, chest fluoroscopy or a CT scan of the chest may be indicated. If the liver function studies are markedly abnormal, a CT scan of the liver is indicated. These studies remain extremely non-specific and are only useful for gross screening. Abnormal liver function studies are most frequently related to underlying alcoholic liver disease, and mild abnormalities do not warrant further work-up. In one recent study, a chest radiograph was 50 per cent sensitive and 94 per cent specific for pulmonary metastases, alkaline phosphatase measurements were 20 per cent sensitive and 98 per cent specific for bony metastases, and liver function tests were 50 per cent sensitive and 81 per cent specific for liver metastases.

The fundamental means of evaluation of laryngeal malignancy remains direct laryngoscopy and biopsy in the operating room. Here the larynx can be palpated, structures obstructing the view can be moved out of the way, the operating microscope and/or rigid quartz Hopkin's rods can be used to obtain a magnified, illuminated view, and appropriate biopsies can be obtained. Although direct laryngoscopy was commonly undertaken using local anesthesia in the past, it can be performed safely and more adequately under general anesthesia. Direct laryngoscopy could be combined with esophagoscopy, bronchoscopy, and careful intraoperative evaluation of the oral cavity, oropharynx, and nasopharynx. Chest radiography and bronchoscopy complement each other; chest radiographs detect tumors of the peripheral airways, while bronchoscopy detects non-obstructing lesions of the major airways.

Second primary cancers of the esophagus are among the most frequently reported, and for this reason esophagoscopy is recommended. If this is not undertaken, an esophagogram should be obtained before endoscopy. Esophagoscopy and esophagography may be complementary: the former is better at diagnosing lesions of the proximal esophagus where multiple overlapping shadows make radiologic interpretation more difficult, while the latter is better for lesions of the distal esophagus.

Staging of laryngeal cancer

Cancer of the larynx is staged according to a TNM system. The importance of staging lies in its ability to allow the surgeon to document the original extent of the disease, to assess the prognosis, and to compare his results with those of others. All patients should undergo formal staging at the time of endoscopy. Appropriate drawings of the tumor should be made and included in the patient's permanent record.

The most important factor in the surgical approach to these cancers is a complete understanding of the true three-dimensional extent of the tumor. Only then can an appropriate operation be conceived and planned. Local control of these cancers through surgery means obtaining clear surgical margins, and this goal can only be achieved if the surgeon has a good idea of where those margins lie. Figure 1 shows the American Joint Committee on Cancer staging system for laryngeal cancer.

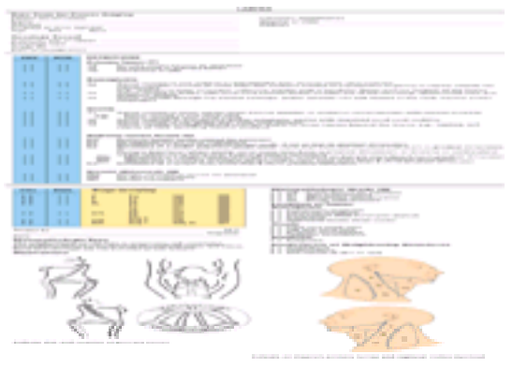


Fig. 1. American Joint Commission on Cancer staging system for laryngeal cancer.

Treatment of laryngeal cancer

Appropriate treatment of laryngeal cancer is dependent on the extent of disease. As noted above, appropriate treatment planning requires pretreatment evaluation of each individual tumor. In general, early T1 cancers of the glottic larynx are treated with radiation alone. T2 lesions are often treated with radiation; if there is some limitation of vocal cord motion without complete fixation, a vertical hemilaryngectomy may be required.

The decision about whether to use postoperative radiation therapy after partial laryngectomy for T2 disease is based on the final pathologic findings. Indications for postoperative radiation therapy in this setting are positive surgical margins, the presence of marked vascular or perineural invasion, and the presence of lymph node metastases in any neck dissection performed concomitantly.

T3 and T4 glottic lesions are generally treated with a combination of total laryngectomy and postoperative radiation therapy (5000 to 6000 Gy). Supraglottic cancers not extending to the true vocal cords or piriform sinus, or to within 1 cm of the circumvallate papillas of the tongue may be treated with a supraglottic laryngectomy. Clinically positive lymph nodes are removed by neck dissection, and both sides of the neck receive postoperative radiotherapy. If the neck is clinically and radiologically free of metastases, bilateral supraomohyoid neck dissections (lymph node levels I, II, and III) for staging may be performed in the hope of finding no disease. If the primary cancer has been removed with clear surgical margins and without extensive perineural or vascular invasion, and if the neck is pathologically free of metastases, postoperative radiation therapy can be withheld. Alternatively, the clinically cancer-free neck could be left undissected and radiation given to each side of the neck after resection of the primary cancer.

Advanced supraglottic cancers with major extension in the base of the tongue are best treated with planned combined surgery and postoperative radiation therapy. If appreciable amounts of the tongue base have to be resected, postoperative deglutition may be so severely affected as to mandate either an initial total laryngectomy, if the problem is anticipated, or a delayed completion laryngectomy, if it is not. If total glossectomy in addition to total laryngectomy were necessary for complete tumor removal, most physicians would opt for non-surgical treatment with combined chemotherapy and radiation. Supraglottic cancers fixing the hemilarynx or extending into the piriform sinus are usually best treated with total laryngectomy and postoperative radiation therapy. In some instances supracricoid or near-total laryngectomy may be safely used. Subglottic cancers almost always require a total laryngectomy for safe removal.

Radiation therapy

Million and Cassisi reported a 93 per cent local control rate for 171 patients with T1 glottic cancer treated with radiation alone. With surgical salvage, the control rate rose to 97 per cent. For T2 glottic cancers there was a 70 per cent local control rate with radiation alone. This rose to 94 per cent with surgical salvage. Wang reported a 93 per cent 5-year local control rate in 665 patients treated with once daily radiation for T1 glottic cancer. He found a significantly improved control rate with twice daily radiation for T2 and T3 glottic lesions. With twice daily radiation 5-year local control rates were 78 per cent for T2 and 67 per cent for T3 glottic cancers. When radiation therapy is used in combination with surgery for advanced glottic cancers, it is generally given postoperatively. This plan reduces the operative complications and yields results equally as good as with preoperative radiation. Wang has reported that 73 per cent of 25 patients with T3, N0 glottic cancer treated with combined surgery and postoperative radiation showed no evidence of disease after 5 years. In most centers, this is the preferred treatment for T3 and T4 glottic cancer.

Local control was achieved in 84 and 61 per cent of patients with T1 and T2 supraglottic cancers, respectively, 5 years after twice daily radiation alone. Local control rates of 71 and 84 per cent are reported by Wang for T3 and T4 supraglottic cancers respectively, after treatment with twice daily radiation.

Primary subglottic laryngeal cancer is rare and meaningful data on the results of irradiation of these lesions is lacking.

Combined chemotherapy and radiation therapy

A recent study from the Veterans Administration compared standard surgery and postoperative radiation with a combination of chemotherapy and radiation in patients who would otherwise have required a total laryngectomy. The chemotherapy was delivered in a neoadjuvant manner using cisplatin and 5-fluorouracil followed by radiation for patients with a good response to the initial chemotherapy. Non-responders were operated on. Of 166 patients originally assigned to the organ preservation arm, 59 (36 per cent) ultimately required laryngectomy. At the time the paper was written, with a median follow-up of 33 months, 101 of the patients in the organ preservation group remained alive, 63 with a functioning larynx (38 per cent of the original group). Disease-free survival was decreased in the organ preservation arm, but ultimate 2-year survival was equivalent. This study, and several others like it, have led some investigators to recommend this organ preservation approach as mainstream for the treatment of resectable T3 and T4 laryngeal cancer in an attempt to preserve the larynx. The study has been criticized for lacking a radiation alone arm. It can be very difficult to follow patients after combined chemoradiation and determine if the changes seen are a result of radiation effect or represent residual cancer. Biopsy may lead to radionecrosis and/or airway comprise. It has been shown that patients treated in this manner have a better outcome when treated at a medical facility experienced in the use of such combined protocols. Delays in the overall radiation treatment period are associated with poorer response.

Not infrequently after primary treatment with combined chemoradiation, a mass will remain in the neck at the conclusion of therapy. This is usually removed via some form of neck dissection. Some treatment centers routinely perform neck dissections after combined chemoradiation for patients who initially present with bulky neck disease. Others make treatment decisions based on the clinical and radiographic response to therapy. This remains an area of research interest. New imaging techniques that could clearly differentiate persistent tumor from treatment effect would be extremely useful and are under investigation. PET scanning and MRI spectroscopy are the two techniques most intensively under investigation at this time.

Endoscopic surgery

Early T1 cancers of the glottis which are confined to the mid-portion of a mobile vocal cord may be treated endoscopically with either microdissection instruments or the CO_2 laser. Following laser treatment, recurrence was seen in 25 per cent of 21 patients with T1 glottic cancers and carcinoma *in situ.* Three responded to radiation therapy, while the other responded to repeat laser excision. The anterior commissure is the most frequent site of recurrence: neoplasms at this site should not be treated by laser.

Laryngofissure and cordectomy

In this procedure the larynx is opened vertically, and the soft tissues of one true and false vocal cord are removed. This procedure is indicated only for early tumors confined to a single mobile vocal cord. Cure rates are greater than 90 per cent. It is an alternative to primary radiation for these tumors.

Vertical hemilaryngectomy

This procedure has many variations. It essentially differs from laryngofissure and cordectomy in that the underlying thyroid cartilage is removed with the specimen. The ipsilateral arytenoid cartilage is usually also removed (Fig. 2). Variations on the theme include the anterolateral and anterior partial laryngectomy, in which both arytenoids may be spared. Despite the thyroid cartilage being removed in this procedure, it is not appropriate for cancers invading cartilage, fixing the vocal cord, or invading the paraglottic space, which are best treated with a total laryngectomy.

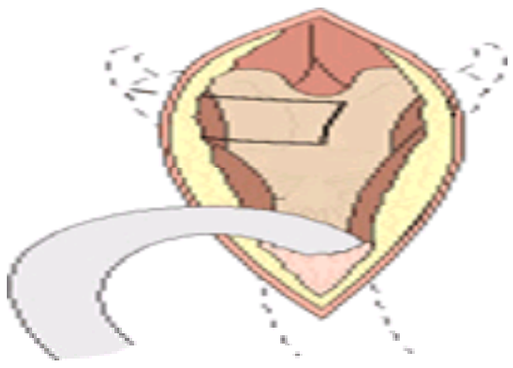


Fig. 2. Vertical hemilaryngectomy. The ipsilateral thyroid cartilage, true and false vocal cord, and arytenoid cartilage are removed.

After vertical partial laryngectomy, deglutition generally remains good because of the remaining epiglottis and sensory innervation of the supraglottic larynx, but the voice is usually hoarse. Some attempts have been made to lessen this hoarseness by insinuating soft tissue into the laryngeal defect, usually in the form of strap muscle. This procedure must be performed with great care to avoid compromising the airway.

Anterolateral laryngectomy

Anterolateral laryngectomy is a variation of the hemilaryngectomy discussed above, in which the anterior portions of the true vocal cords, together with the underlying cartilage, are removed. The arytenoids are spared.

Supraglottic laryngectomy

Supraglottic laryngectomy is indicated for cancers of the supraglottic larynx not involving the true vocal cords or arytenoid cartilages: it entails removing the supraglottic larynx down to the level of the laryngeal ventricles (Fig. 3). Neoplasms situated above the level of the hyoid bone usually have a better prognosis. The region of the anterior commissure must be carefully evaluated at the time of endoscopy. Involvement of the medial wall of the piriform sinus is a relative contraindication to this procedure, although some tumors barely spilling over the aryepiglottic fold into the most superior aspect of the medial piriform sinus might still be considered appropriate for such an operation.

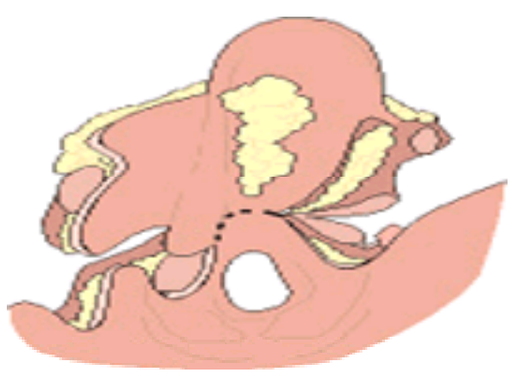


Fig. 3. Supraglottic laryngectomy. The supraglottic larynx including the epiglottis and false vocal cords are removed down to the level of the laryngeal ventricle. The arytenoid cartilages are left *in situ.*

Although the tongue base can be resected to within 1 cm of the circumvallate papillas, the risk of serious functional morbidity is greatly increased. Morbidity associated with tongue base resection usually relates to deglutition. Because the epiglottis is resected and the innervation is disturbed, food tends to slide over the tongue base and immediately into the glottic larynx, causing aspiration. If pulmonary reserve is good, most patients will compensate. However, if pulmonary reserve is poor, the patient may be unable to feed himself, or worse, may suffer chronic aspiration, ultimately requiring a secondary procedure to complete the laryngectomy. The voice is often good after this procedure, as opposed to the breathy voice heard after vertical hemilaryngectomy.

Supracricoid laryngectomy

The supracricoid laryngectomy is a relatively new procedure most commonly used for supraglottic cancers which spread inferiorly to the glottis. In this procedure, the laryngeal structures superior to the cricoid are excised, sparing the posterior glottis, arytenoids, and interarytenoid space. One arytenoid can be included in the resection. The entire pre-epiglottic and paraglottic spaces are included in the resection. The wound is closed by sewing the cricoid to the hyoid bone (cricohyoidopexy). Significant subglottic involvement is a contraindication to this procedure (Fig. 4).

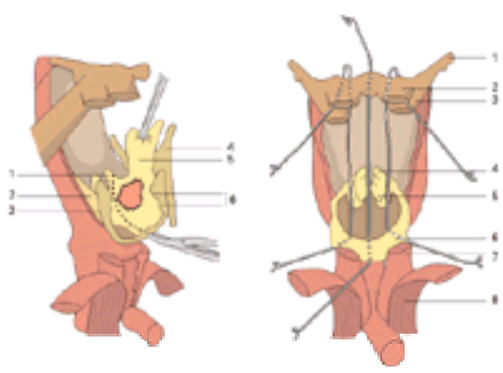


Fig. 4. Supracricoid laryngectomy. The laryngeal structures above the cricoid are resected, preserving one or both arytenoid cartilages. The wound is closed by suturing the cricoid cartilage to the hyoid bone.

Near-total laryngectomy

The near-total laryngectomy removes the thyroid and cricoid cartilage, leaving only a single innervated arytenoid and a ‘speaking shunt'. Unlike other partial laryngectomy procedures, near-total laryngectomy assumes that the patient will be dependent on a tracheostomy for the rest of his or her life. The advantages of this procedure, compared with a laryngectomy and voice rehabilitation with a tracheo-esophageal puncture, are the quality of the voice and the ease of care afterwards. Large tumors that affect one vocal cord or which spill over into the upper aspects of one piriform sinus are suitable for this procedure.

Total laryngectomy

Total laryngectomy is the standard against which other procedures must be weighed in the treatment of advanced laryngeal cancers. Vocal cord fixation, cartilage

invasion, and involvement of the piriform sinuses have been traditional indications for this procedure.

Total laryngectomy involves removal of the larynx from the remaining pharynx and transection of the trachea (Fig. 5). The remaining pharynx is then closed on itself, usually in a ‘T' fashion; the proximal end of the trachea is sewn to the skin as an end tracheostoma.

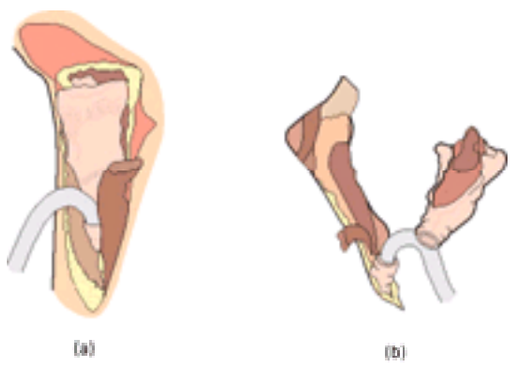


Fig. 5. (a) Total laryngectomy. The larynx is removed from the pharynx. Entrance to the pharynx is usually, though not always, made through the piriform sinus opposite the side of the tumor. (b) Total laryngectomy. The larynx has been removed. The proximal end of the distal trachea will be sewn to the skin as an end tracheostoma. The pharynx will be closed, thus separating the digestive and respiratory systems.

Extended total laryngectomy

Extended total laryngectomy is a procedure usually used for tumors involving the piriform sinus. In these cases more than the usual amount of piriform and pharyngeal mucosa must be removed, narrowing the resulting neopharynx. In general, most of the mucosa of a single piriform sinus can be resected leaving enough remaining mucosa to afford an adequate closure. If larger amounts of mucosa are resected, the pharynx can still be closed successfully, but at the expense of pharyngeal stenosis and subsequent morbidity in the form of inability to eat substances other than liquids. Tight closure over a nasogastric tube is too tight. It is better to supplement the closure with additional tissue, such as a pectoralis major myocutaneous flap, or to proceed with a total laryngopharyngectomy.

Total laryngopharyngectomy

In this procedure, the pharynx is removed circumferentially down to the prevertebral fascia along with the larynx. Total laryngopharyngectomy is used for cancers that spread circumferentially around the pharynx, or nearly so, in which total safe extirpation of the cancer dictates circumferential pharyngectomy; it is also recommended for tumors spilling over into the postcricoid region of the hypopharynx. After this procedure the patient is left with a circumferential, tubular defect, which must be reconstructed. In general, immediate reconstruction with a free jejunal flap is preferred. Patients with advanced cancers requiring total pharyngectomy have a poor outlook for prolonged survival. They should not be subjected to multiple reconstructive procedures, which often require lengthy hospital stays. Alternative reconstructive techniques include other tubed cutaneous free flaps such as the lateral thigh, myocutaneous flaps such as the tubed pectoralis major flap, and gastric pull-up. If the latter is chosen, a total esophagectomy is generally performed. If the technical ability is not available to perform one of these procedures, the wound ay be closed by creating an end pharyngostome and esophagostome through the neck skin flaps in addition to the tracheostome. The defect could then be reconstructed at a later date after postoperative radiation.

Total laryngopharyngo-esophagectomy

Total laryngopharyngo-esophagectomy involves blunt dissection of the esophagus from the posterior mediastinum, in conjunction with a laryngopharyngectomy (Fig. 6). This procedure is required for the treatment of tumors spreading inferiorly into the cervical esophagus beyond the thoracic inlet. Some surgeons feel that the propensity for skip lesions and submucosal spread of advanced tumors involving the hypopharynx make this procedure the best treatment for all such tumors.

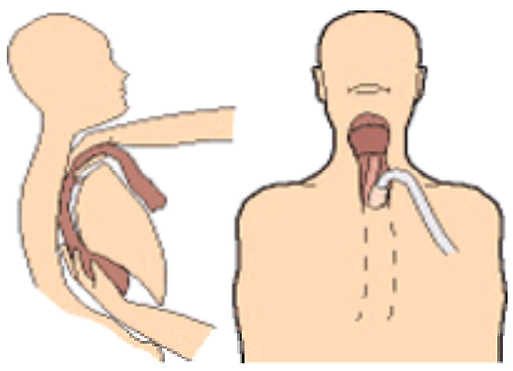


Fig. 6. Total laryngopharyngo-esophagectomy and gastric pull-up. The esophagus, still attached to the larynx and pharynx, is removed from the posterior mediastinum circumferentially by blunt dissection. The stomach is pulled up through the posterior mediastinum and anastomosed to the remaining pharynx.

Voice rehabilitation

All patients undergoing total laryngectomy initially lose their voice, but most of them can be taught to speak in an understandable and socially acceptable manner. Anyone performing laryngectomies must work closely with a speech therapist.

In the past, most patients relied on esophageal speech, in which air was swallowed and then belched, to cause vibration within the pharynx, producing a noise which was then articulated into speech using the teeth, tongue, and palate. A number of electrical devices (electrolarynges) that produce a more fluid tone are now available. These devices are either held up against the skin of the neck overlying the pharynx or are placed in the mouth using a small reed. The most recent advance combines tracheo-esophageal puncture with insertion of a speaking valve prosthesis to allow air to pass from the trachea back into the pharynx when the tracheostome is occluded. The valve allows air to pass from the trachea into the esophagus, but prevents saliva from running in the opposite direction (Fig. 7). Use of this device requires some manual dexterity and self-motivation on the part of the patient. The newest devices need be changed only every 6 months or so in the clinic.

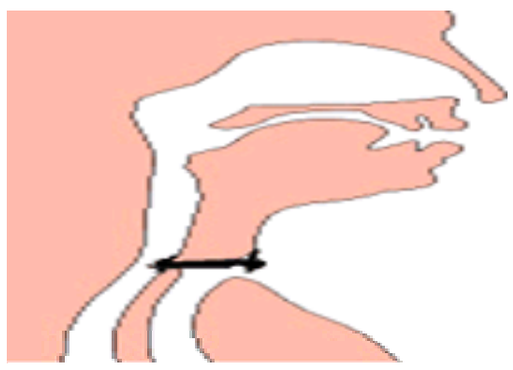


Fig. 7. Tracheo-esophageal puncture voice prosthesis. A small one-way valve which will allow the flow of air from the trachea into the pharynx, but prevent the flow of

saliva from the pharynx into the trachea, is inserted through a small puncture created in the common wall between the trachea and esophagus or pharynx.

In addition to assisting the patients to acquire some means of speech, the speech therapist is invaluable in helping with problems in deglutition after partial laryngectomy. Clear liquids are most difficult to manage because they run into the larynx before the patient can mobilize compensatory mechanisms. Soft but thicker substances such as puddings, fruit sauces, or purées are thus the best foods to begin with after surgery. If continued difficulties are encountered, a modified barium swallow, in which substances of various consistency and size are swallowed under the fluoroscopic observation of a trained observer, may define the abnormalities so that appropriate remedies can be invoked. Rarely a patient requires a gastrostomy for long-term alimentation.

Hypopharyngeal cancer

Anatomy and physiology of the hypopharynx

The hypopharynx is a tube interrupted anteriorly by the larynx. It is divided anatomically into three subsites: the piriform fossas on either side of the larynx funneling down into the esophagus, the postcricoid area consisting of the mucosa on the posterior aspect of the posterior cricoid lamina overlying the posterior cricoarytenoid muscles, and the posterior hypopharyngeal wall.

Etiology and demographics of hypopharyngeal cancer

Tobacco and alcohol abuse remain the major etiologic agents in the development of hypopharyngeal cancer. Women suffering from the Plummer–Vinson syndrome (sideropenic dysphagia, Patterson–Brown–Kelly syndrome) are also at increased risk of developing squamous cancers of the postcricoid hypopharynx and cervical esophagus. This syndrome is defined by iron deficiency anemia, mucosal atrophy, and dysphagia due to poor esophageal motility and cervical esophageal web. It has been seen most commonly in fair-skinned women of Scandinavian and northern European descent, but seems to be waning in incidence for unclear reasons. Statistics available for the United States include all pharyngeal sites (nasopharynx, oropharynx, and hypopharynx). There will be about 8600 new cases of pharyngeal cancer and 2100 cancer-related deaths due to pharyngeal cancer in the United States in 1988. This yields a gross mortality rate of 24 per cent. Hypopharyngeal cancers are the most lethal of the pharyngeal cancers. Pharyngeal cancer is 3.1 times more common in men than women. These cancers account for 0.6 per cent of new cancers and 0.4 per cent of cancer-related deaths each year in the United States. Again, genetic predisposition to these cancers is assumed, and is an area of great research interest.

Growth patterns of hypopharyngeal cancer

Squamous cell cancer of the hypopharynx is reputed to have a greater propensity for submucosal spread and skip lesions than similar cancers in other areas of the upper aerodigestive tract. Cancers involving the apex of the piriform fossa may spread to involve the thyroid gland. In addition to the jugular lymph nodes, the retropharyngeal, paratracheal, and tracheo-esophageal lymph nodes are at increased risk of involvement. Bilateral cervical metastases are common.

Signs and symptoms of hypopharyngeal cancer

Unlike laryngeal glottic cancer, in which hoarseness is present early, hypopharyngeal cancers tend to remain asymptomatic until they reach an advanced stage. Patients often present with only a neck mass, indicative of cervical metastases. Dysphagia and odynophagia are symptoms of more advanced cancer. Ipsilateral otalgia is frequently present and is due to referred pain to the ear from the vagus nerve which innervates both the pharynx and a small portion of the external ear. Advanced cancers of the hypopharynx can present with airway obstruction or hoarseness. This is often a result of fixation of a vocal cord by cancer infiltrating the larynx from the piriform sinus.

Evaluation of patients with suspected hypopharyngeal cancer

Assessment is similar to that for laryngeal cancer. Laryngoscopy, bronchoscopy, and esophagoscopy are even more important to the preoperative evaluation of patients with hypopharyngeal cancer. These cancers often cannot be adequately seen in the clinic because of a tendency for the mucosa of the hypopharynx to infold. Because submucosal spread and skip lesions are common, careful examination for separate lesions in the lower esophagus is required to delineate the inferior extent of the cancer. Since decisions such as whether or not the entire pharynx or esophagus need to be removed cannot be made until the position and extent of the cancer are known, endoscopy should be carried out as a separate procedure to permit discussion of available options with the patient.

As with laryngeal cancer, hypopharyngeal cancer is staged according to a TNM system devised by the American Joint Commission on Cancer (Fig. 8).



Fig. 8. American Joint Commission on Cancer staging system for hypopharyngeal cancer.

Treatment of hypopharyngeal cancer

Hypopharyngeal cancer is best treated with a combination of surgery and postoperative radiation. Although cancers of the posterior pharyngeal wall can occasionally be approached surgically with some form of pharyngotomy or median labiomandibulotomy, cancers of the postcricoid area and piriform sinus have traditionally been treated with an extended total laryngectomy. Ogura introduced the concept of conservative laryngeal surgery for selected early piriform sinus cancers in the 1960s, but these procedures are still controversial.

Radiation therapy

Radiation therapy is rarely used alone for the treatment of hypopharyngeal cancers. These tumors tend to present late because of the difficulties encountered in examining the hypopharynx and identifying such tumors. However, some early T1 cancers may be treated with radiation alone. Wang reports a 74 per cent 5-year local control rate for T1 and T2 piriform sinus cancers and 51 per cent for T3 and T4 cancers treated with twice daily radiation.

Treatment of even small cancers often requires extensive surgery, including total laryngectomy. Extensive surgery of small T1 lesions may often be avoided by the use of primary radiotherapy, but possibly at the expense of some decreased long-term survival. A combination of surgery and postoperative radiation produces the best results. Adjuvant chemotherapy is offered to these patients if an experimental protocol is available. Combinations of chemotherapy and radiation therapy, sometimes given concomitantly, are being studied and may one day prove superior to radiation alone. As discussed earlier, when to perform elective neck dissection in this setting is controversial. The most common approach is to remove clinically or radiographically persistent neck disease if the primary site is controlled after combined chemoradiation.

Transhyoid or lateral pharyngotomy

This technique is useful for the removal of small lesions of the posterior pharyngeal wall, which are rare. The pharynx is entered either from the anterior over the hyoid bone and through the vallecula, or from the side behind the lateral extent of the thyroid cartilage ala. The resulting defect may be left to granulate, closed primarily, or, more usually, closed with a skin graft. Occasionally a T1 tumor confined to the lateral aspect of the piriform sinus may be amenable to resection via a lateral pharyngotomy including the lateral thyroid cartilage ala. These wounds can usually be closed primarily or with a pectoralis major myocutaneous flap. A temporary tracheostomy is required for all transhyoid and lateral pharyngotomy approaches.

Median labiomandibulotomy

This approach is occasionally useful for the treatment of early neoplasms of the posterior hypopharyngeal wall. The midline raphe of the tongue affords a bloodless plane of dissection. The mandible is cut in the midline and is reapproximated with a plate. The inferior alveolar nerves should be spared.

Partial laryngopharyngectomy

An extended supraglottic laryngectomy or near-total laryngectomy may permit extirpation of selected cancers of the superior piriform sinus. Destruction of thyroid cartilage or involvement of the piri-form apex is a relative contraindication to such an approach. Involvement of the postcricoid region is a relative contraindication to any type of conservation laryngeal surgery.

Extended total laryngectomy

This remains the standard operation for the surgical treatment of hypopharyngeal cancers, against which others must be compared. In the treatment of piriform sinus cancers, the dissection often only requires extension to include the mucosa of the piriform sinus. If the piriform apex is involved, the ipsilateral hemithyroid should be re-moved with the specimen. Reconstitution of the pharynx can be achieved with as little as 2 cm of horizontal width of remaining pharyngeal mucosa. There is no absolute guideline as to when supplemental tissue is required for pharyngeal closure; a tight closure around the nasogastric tube is too tight. If the closure is so tight as to compromise blood flow or cause irritation from the nasogastric tube, deglutatory morbidity or postoperative fistulization results. A closure may be successful in that it heals, but it is a functional failure if the patient can only tolerate a liquid diet. It is easier to supplement the closure with additional tissue, usually a pectoralis major myocutaneous flap or free radial forearm flap, than to perform reconstruction later.

Total laryngopharyngectomy

This is often necessary to ensure complete removal of large hypopharyngeal cancers. The larynx and pharynx are removed circumferentially. Reconstruction is discussed above and is usually accomplished with a free jejunal flap, free tubed cutaneous flap, or tubed pectoralis major myocutaneous flap. Both inferior and superior margins must be carefully assessed. High nasopharyngeal involvement may limit the success of surgical resection and may be better treated with combined chemotherapy and irradiation. Involvement of the esophagus beyond the thoracic inlet would necessitate esophagectomy. The retropharyngeal nodes of Rouviere may be involved and prevent successful surgical resection.

Total laryngopharyngo-esophagectomy

Some surgeons consider that submucosal spread and skip lesions indicate total esophagectomy for the treatment of many advanced hypopharyngeal cancers. Others prefer reconstruction by gastric pull-up, together with total esophagectomy, for all circumferential pharyngeal defects. In some patients the cancer visibly spreads down into the thoracic esophagus. Others have an obvious skip lesion or primary cancer of the more distal esophagus. In such cases total esophagectomy is the only adequate surgical approach. The prognosis of patients with cancers of the thoracic esophagus that invade all the way through the wall is so dismal following surgical therapy alone that some other approach is indicated. Initial therapy with radiation and chemotherapy, perhaps concomitantly, is reasonable, surgery being undertaken later, and only if a good response is achieved.

Reconstructive techniques

The tubed pectoralis major myocutaneous flap, gastric pull-up, free jejunal flap, and free tubed cutaneous flap are the most widely used methods for reconstruction of circumferential pharyngeal defects. The tubed pectoralis flap is a difficult procedure in women with large breasts and in those with a large amount of subcutaneous tissue. Most patients with pharyngeal cancer are quite cachectic, however, easing this reconstructive problem. Although flaps may become stenotic, especially at the lower anastomosis to the cervical esophagus, the problem can often be avoided by making a vertical cut inferiorly in the anterior wall of the cervical esophagus and cutting the upper end of the pectoralis flap on the chest wall in a fusiform manner such that, when it is tubed, it has a slanted end. When the flap is transposed into the neck, this becomes the lower end, which can then be sewn to the cervical esophagus on a bias, avoiding a perfectly circumferential suture line.

The free jejunal flap (cut from its native blood supply) or another free flap may also be reliably used for reconstruction. The rate of vascular anastomotic failure in any free flap is about 10 per cent. This may prove fatal if a severe infection develops around great vessels in the neck with direct communication with the mediastinum. A free jejunal flap can also be cut on a bias on what will be the distal end; the anterior wall of the remaining esophagus can be incised to avoid a perfectly circular anastomosis and to help prevent stricture. The bias on the jejunum must be cut with the longer side toward the mesentery to avoid devascularizing a portion of the jejunum. In addition, at the superior end of the jejunum, a longitudinal incision is made in the antimesenteric wall to provide a longer perimeter to which to sew the pharynx, which has a much larger natural circumference than the jejunum. The status of the free graft can be monitored postoperatively by leaving a small portion of the flap exteriorized through the skin at a suture line so that it can be seen. The patency of the vascular anastomosis can then be followed by visual inspection of the piece of exteriorized jejunum and with a Doppler probe applied to its mesentery.

Gastric pull-up normally entails blunt dissection of the esophagus from the posterior mediastinum, pulling the stomach up through this space to be anastomosed to the pharynx. The stomach can normally only be pulled up to the level of the circumvallate papillas of the tongue. Previous abdominal surgery may compromise the vascularity of the stomach. There are inherent risks in bluntly dissecting the esophagus out of the posterior mediastinum: rupture of the azygous vein, chylous fistula, and disruption of the posterior tracheal wall are the most common. The physiologic derangements seen after gastric pull-up or free jejunal flap reconstruction are greater than those incurred in simple circumferential reconstruction of the pharynx with a pectoralis major myocutaneous flap.

Prognosis of patients with cancers of the hypopharynx

The experience at Memorial-Sloan Kettering Cancer Center has re-cently been published. They compared 26 patients with advanced, re-sectable squamous cell cancer of the hypopharynx treated with induction chemotherapy and radiation with 30 patients treated with surgery and postoperative radiation. The two groups were comparable, though there were more T4 lesions in the chemoradiation arm. The 5-year local recurrence-free survival was 50 and 69 per cent, respectively. The 5-year disease-free survival was 30 and 42 per cent, respectively, while the 5-year overall survival was 15 and 22 per cent, respectively. None of these differences were statistically significant, but the small numbers make statistical comparisons suspect. Taken together, the majority of studies have shown surgery and radiation to be superior to radiation alone for cancers of the hypopharynx. Data from the M.D. Anderson Cancer Institute in the 1970s showed control rates for pharyngeal wall cancers of 49 per cent with radiation alone, compared with control rates of 72 per cent with radiation plus surgery.

Treatment of the neck in patients with cancers of the larynx and hypopharynx

Physical examination predicts the presence or absence of metastatic disease in the neck poorly. CT scans and MRI studies can visualize enlarged lymph nodes in the neck, but cannot always differentiate reactive lymph nodes from those containing cancer.

Radical neck dissection refers to the complete removal of lymph node-containing tissue from the neck, along with the sternocleidomastoid muscle, internal jugular vein, and spinal accessory nerve. Modifications of the radical neck dissection generally preserve some combination of the sternocleidomastoid muscle, internal jugular vein, or spinal accessory nerve, while still attempting total removal of the lymph node-containing compartments. Of these three structures, preservation of the spinal accessory nerve is by far the most important in preventing postoperative morbidity. The spinal accessory nerve is a posterior neck structure only in its inferior extent. More proximally, it crosses the internal jugular vein at the lateral process of the first cervical vertebra; here it is intimately related to the junctional or jugulodigastric level II lymph nodes, which are common to both the jugular and spinal accessory chains of lymph nodes. It is unsafe to preserve this nerve if these lymph nodes are involved. Initially, the nerve was preserved only when there were no clinically positive nodes in the neck, but there has been a recent move to preserve it as long as there are no clinically positive nodes directly adherent to the nerve.

Partial neck dissections remove those lymph nodes considered to be at greatest risk of harboring metastases from a given cancer. These procedures are generally used in the treatment of the clinically disease-free neck as a staging procedure to determine whether further therapy is indicated. The lymph nodes in the neck are divided into seven levels as shown in Fig. 9. The most commonly performed partial neck dissections are: the lateral neck dissections (for cancers of the larynx and pharynx) removing levels II, III, and IV; the supraomohyoid neck dissection (for cancers of the oral cavity) removing levels I, II, and III; and the posterolateral neck dissection (for skin cancers of the posterior scalp) removing levels VII, V, II, III, and IV. Radiation therapy is effective in the treatment of occult neck disease. The greatest value of elective partial neck dissections is in the treatment of the clinically disease-free neck when the primary cancer can be effectively treated with surgery alone. In this instance, a partial neck dissection disclosing no pathologic evidence of cervical metastases indicates that no further therapy is needed. The added morbidity of immediate radiation therapy can be avoided, such treatment being reserved for any future recurrences or new cancers.

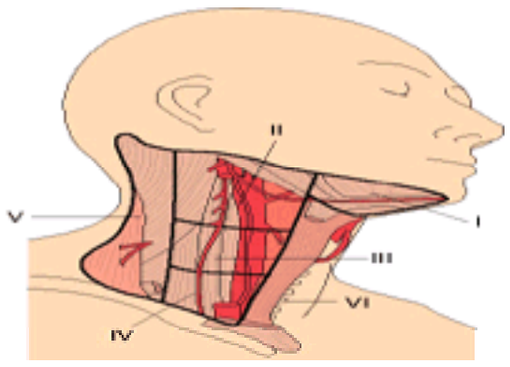


Fig. 9. Lymph node levels of the neck. Level 1, submental and submandibular group; level II, upper jugular group; level III, middle jugular group; level IV, lower jugular group; level V, posterior triangle group; level VI, anterior compartment group.

If clinical or radiographic evidence of disease persists in the neck after successful treatment of the primary cancer site with radiation or combined chemotherapy and radiation, it should be removed by neck dissection.

Follow-up of patients with cancer of the larynx and hypopharynx

After completing treatment, patients with squamous cell carcinomas of the head and neck should be seen about every month for the first year, every 2 months for the second year, every 3 months for the third year, every 6 months for the fourth year, and annually thereafter. A chest radiograph should be obtained at 6 months and annually thereafter. If the patient has received radiation therapy to the neck, thyroid-stimulating hormone level should be tested annually.

Further reading

Brockstein BE, Vokes EE. Chemotherapy for head and neck cancer. In: Devita VT, Hellman S, Rosenberg SA, eds, *Principles and practice of oncology, updates*, Vol. 10, pp. 1–19. Lippincott-Raven Publishers, Philadelphia, 1996. [A good comprehensive medical oncology text.]

Carpenter RJ, De Santo LW, Devine LW, Taylor WF. Cancer of the hypopharynx. *Archives of Otolaryngology* 1976; **102**: 716–21. [A review of surgical treatment of cancers of the hypopharynx.]

El Badawi SA, Goepfert H, Fletcher GH, Herson J, Oswald MJ. Squamous cell carcinoma of the pyriform sinus. *Head and Neck Surgery* 1983; **5**: 508–13. [A landmark article showing the advantage of combined therapy for cancers of the pyriform sinus.]

Harrison DFN, Thompson AE. Pharyngolaryngoesophagectomy with pharyngogastric anastomosis for cancer of the hypopharynx: review of 101 operations. *Head and Neck Surgery* 1986; **8**: 418–28. [The experience of a master surgeon with total pharyngolaryngoesophagectomy.]

Laccourreye H, Laccourrreye O, Weinstein G, Menard M, Brasnu D. Supracricoid laryngectomy with cricohyoidopexy: a partial laryngeal procedure for selected supraglottic and transglottic carcinomas. *Laryngoscope* 1990; **100**: 735–41. [A landmark article describing the supracricoid laryngectomy.]

Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1998. *Ca: A Cancer Journal for Clinicians* 1998; **48**: 6–29. [Compendium of United States of America cancer statistics.]

Lingeman RE, Shellhamer RH. Surgical management of tumors of the neck. In: Thawley SE, Panje WR, Batsakis JG, Lindberg RD, eds, *Comprehensive management of head and neck tumours*, pp. 1325–50. WB Saunders, Philadelphia, 1987. [An up-to-date review of surgery of the neck as it relates to head and neck cancers.]

Million RR, Cassisi NJ, eds. *Management of head and neck cancer: a multidisciplinary approach*, 2nd edn. JB Lippincott, Philadelphia, 1994. [A good comprehensive text on the management of head and neck cancer.]

Million RR, Cassisi NJ, Mancuso AA. Larynx. In: Million RR, Cassisi NJ, eds, *Management of head and neck cancer: a multidisciplinary approach*, 2nd edn. JB Lippincott, Philadelphia, 1994. [A good comprehensive text on the management of head and neck cancer.]

Ogura JH, Marks JE, Freeman RB. Results of conservation surgery for cancers of the supraglottis and pyriform sinus. *Laryngoscope* 1983; **90**: 591–600. [Review of subtotal larygectomy for cancers of the supraglottic larynx.]

Silver CE. *Surgery for cancer of the larynx and related structures*. WB Saunders, Philadelphia, 1996. [An excellent comprehensive text on the treatment of larynx cancer.]

Singer MI. Tracheoesophageal speech: vocal rehabilitation after laryngectomy. *Laryngoscope* 1980; **93**: 1454–65. [An excellent reference on tracheo-esophageal puncture speech rehabilitation after total laryngectomy.]

Thawley SE, Panje WR, Batsakis JG, Lindberg RD, eds. *Comprehensive management of head and neck tumours*. WB Saunders, Philadelphia, 1987. [Another good comprehensive text on the management of head and neck cancer.]

The Department of Veterans Affairs Laryngeal Cancer Study Group. Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. *New England Journal of Medicine* 1991; **324**: 1685–90. [One of the first quality prospective studies of conservation treatment of T3 laryngeal cancers with chemotherapy and radiation therapy in an attempt to avoid total layngectomy. Methods are now somewhat dated.]

Wang CC. *Radiation therapy for head and neck neoplasms*. Wiley-Liss, New York, 1997. [A comprehensive text on primary radiation therapy for head and neck cancer.]

Wetmore SJ, Key JM, Suen J. Laser therapy for T1 glottic carcinoma of the larynx. *Archives of Otolaryngology—Head and Neck Surgery* 1986; **112**: 853–5. [A review of endoscopic laser therapy for early glottic cancers.]

45.8 Surgery of the ear and temporal bone

Andrew P. Freeland and Martin J. Burton

Surgical anatomy
Anatomical surgical issues
Trauma to the ear and temporal bone
External ear
Middle ear
Temporal bone
Barotrauma
Infection of the middle ear and mastoid
Acute suppurative otitis media
Acute mastoiditis
Chronic suppurative otitis media
Secretory otitis media
Neoplasia
Acoustic neuroma (vestibular schwannoma)
Other benign tumours
Malignant tumours
Otosclerosis
Non-surgical treatment
Stapedotomy
Menière's disease
Medical treatment
Surgery
Management of conductive deafness
Medical treatment
Surgical treatment
Bone-anchored hearing aids
Management of sensorineural deafness
Cochlear implants
Further reading

Surgical anatomy

The ear comprises three distinct parts: the external ear, middle ear, and inner ear. The external ear includes the pinna and external auditory canal. At the medial end of the canal the tympanic membrane separates the external ear from the middle ear. The middle ear consists of a narrow, air-containing cleft within which are the three ossicles, the malleus, incus and stapes. The footplate of the stapes occupies the oval window. Deep to this is the inner ear containing the organs of hearing (the cochlea with organ of Corti) and balance (the vestibular system with saccule, utricle, and semicircular canals). Apart from the lateral portion of the external auditory canal, which is cartilaginous, the remainder of the structures described are contained within the temporal bone.

The anatomy of the temporal bone is complex but a thorough knowledge of it is vital if diseases of this region and their management are to be understood. The temporal bone forms the inferolateral limit of the skull base (Fig. 1). Laterally the tympanic portion forms part of the external auditory canal. Anteriorly the zygomatic portion articulates with the zygoma and superolaterally the squamous portion articulates with the parietal and occipital bones. The mastoid portion is situated posteriorly; the petrous portion passes medially as a pyramid with its apex pointing anteromedially. The superior surface of this pyramid forms the floor of the middle cranial fossa, and the posterior surface forms the anterior wall of the posterior fossa.

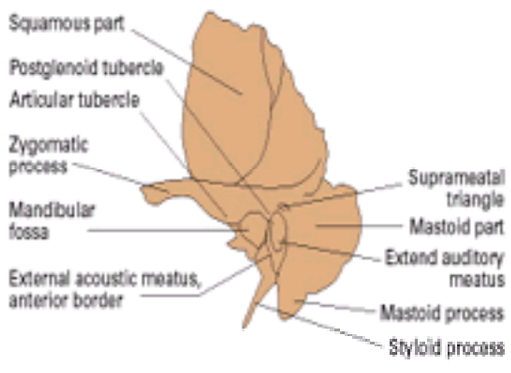


Fig. 1. Lateral view of temporal bone

The external canal passes from laterally to medially as it enters the temporal bone. In line with it, the internal auditory canal passes from the inner ear to the posterior surface of the temporal bone. It contains the cochleovestibular nerve (cranial nerve VIII comprising the cochlear nerve, superior and inferior vestibular nerves), and the facial nerve (VII).

Both the VIIth and VIIIth cranial nerves cross the cerebellopontine angle as they pass between the brainstem and the internal auditory canal. The facial nerve passes from medially to laterally through the canal. At the fundus it turns posteriorly at the first genu and runs in the medial wall of the middle ear before turning inferiorly at the second genu to pass to the stylomastoid foramen, where it exits the skull base. The fibres of the VIIIth nerve contact the sensory epithelial cells of the organ of Corti and the vestibular end-organs, and carry impulses back to the cochlear and vestibular nuclei in the brainstem.

The eustachian tube leaves the anterior wall of the middle ear; as it opens and closes with swallowing, it equilibrates pressure between the middle ear and nasopharynx. High in the posterior wall of the middle ear is the aditus, an opening into the antrum that in turn leads to the complex array of air cells within the mastoid. These are highly variable. In some patients the whole temporal bone is well pneumatized, air cells being present in all parts including the petrous apex. In others there are few cells and the mastoid is sclerotic. The basal turn of the cochlea protrudes into the middle-ear cleft and this forms the promontory. The round window is situated posteroinferiorly and the membrane that closes it separates the middle ear from the interior of the cochlea.

The sigmoid sinus forms a groove on the inner surface of the temporal bone posteriorly as it passes inferiorly to exit the jugular foramen. The internal carotid artery passes into the carotid canal in the base of the petrous portion of the temporal bone. Closely related to the floor of the middle ear, it turns forwards and runs towards the petrous apex. It is also in close relation to the eustachian tube before entering the middle fossa.

Anatomical surgical issues

The principal surgical challenges within the temporal bone relate to presence of the facial nerve and the extremely delicate inner-ear structures. Facial palsy and total sensorineural deafness (with or without balance dysfunction) are potential complications of surgery in this region.

The facial nerve is at risk in the internal auditory canal (during removal of an acoustic neuroma), middle ear and mastoid (during surgery for chronic suppurative otitis media), and in surgery to the external ear when congenital anomalies exist. The nerve may be dehiscent as it passes through the middle ear but usually lies in the bony fallopian canal. Several landmarks help the surgeon to locate the nerve: these include the processus cochleariformis (around which the tendon of the tensor tympani muscle turns), the lateral semicircular canal, the oval window, and the digastric notch. The best landmark is the nerve itself; the surgeon who identifies the nerve and is

constantly aware of its actual position is less likely to damage it.

The inner ear may be damaged if it is entered accidentally or intentionally (as in stapedotomy). In patients with cholesteatoma, the disease itself may have eroded bone of the otic capsule and there may be a fistula into the inner ear (usually via the lateral semicircular canal). Removal of the cholesteatoma matrix overlying such a fistula may result in hearing loss. The appropriately trained surgeon should be able to avoid entering the canals unintentionally, although normal anatomy may be distorted by advanced disease. The short process of the incus is closely related to the lateral canal and the second genu of the facial nerve, and serves as a pointer to these structures.

Other anatomical structures that may be exposed either during mastoid surgery or by disease within the temporal bone are the dura (especially that of the middle fossa) and the sigmoid sinus.

Trauma to the ear and temporal bone

External ear

The pinna is in an exposed location and therefore prone to trauma. However, the framework is elastic cartilage, which, unlike hyaline cartilage, does not fracture.

Blunt trauma

A cauliflower ear is the most common result of blunt trauma. A haematoma forms between the perichondrium and the cartilage of the pinna, resulting in an ugly defect. The blood supply of the elastic cartilage relies on the perichondrium. If infection complicates the haematoma, cartilage may die, giving a severe deformity. Any external laceration requires a broad-spectrum antibiotic. Aspiration may be tried in the acute stage, but the haematoma frequently recurs and open drainage, antibiotics, and tight packing may be the only way to control it.

Sharp trauma

Lacerations, or even total avulsion, occasionally occur in accidents, and bites, either from animals or humans, are not infrequently seen. The pinna has an excellent blood supply, which aids reconstruction, but bites by definition are infected and need to be treated with antibiotics and tetanus toxoid. If cartilage is exposed and damaged, it should be resected.

Thermal trauma

Frostbite causes ischaemia from arteriolar vasoconstriction. Ice crystals may form within the cells. At this stage the ear is much at risk from mechanical damage. Gentle rewarming is the only treatment.

Middle ear

Direct trauma from cotton buds or foreign bodies pushed into the ear by children may rupture the tympanic membrane and disrupt the ossicular chain, usually the incudostapedial joint.

Pressure from the blast of an explosion, a very loud noise, or a physical slap across the ear can be forcible enough to rupture the tympanic membrane or even dislocate the ossicles.

Temporal bone fracture (discussed in more detail later) may also disrupt the tympanic membrane and the ossicles.

Most acute ruptures occur in the posterior part of the tympanic membrane and most heal spontaneously. It is essential that the ear be kept clean and free of water, and that an audiogram be performed. A conductive deafness up to 30 dB occurs if the tympanic membrane alone has been ruptured. However, if the ossicular chain has been damaged, a 60-dB deafness can result. If the tympanic membrane closes spontaneously, leaving a persistent conductive deafness, reconstruction of the ossicular chain may be necessary.

Dislocation of the incudostapedial joint is the most common injury. Repositioning a bone graft or artificial prosthesis is the usual repair; this is approached via a per-meatal tympanotomy, elevating the posterior ear canal skin down to the tympanic membrane, the posterior half of which is then turned forward to reveal the middle-ear contents.

Temporal bone

Head injury remains one of the most frequent injuries. It is surprising that the temporal bone, which includes the dense petrous bone (*petros* = rock), is ever fractured. However, it is traversed by many foramina and interconnections, which probably weakens it, allowing trauma to the semicircular canals, cochlea, facial, auditory and vestibular nerves, and, more laterally, the middle and external ear.

Labyrinthine concussion

Concussional deafness and vestibular symptoms can occur after head injuries even without a fracture. Most cases recover slowly. Benign positional paroxysmal vertigo, probably due to displacement of the otolith organ within the posterior semicircular canal, is a frequent complication of head injury, but time and repositioning manoeuvres (Epley) usually aid recovery.

Fractures

The two classic fracture lines in the temporal bone are longitudinal and transverse (Fig. 2). Longitudinal fractures make up about 80 per cent and can be bilateral in up to a quarter of patients. They cross the posterior superior roof of the external auditory canal and middle ear, avoiding the internal ear and extending anterior to the carotid canal. The tympanic membrane, the roof of the middle ear and the ossicles may be disrupted, causing a conductive deafness, but if there is a sensorineural component, inner-ear concussion is the usual cause, which improves with time.

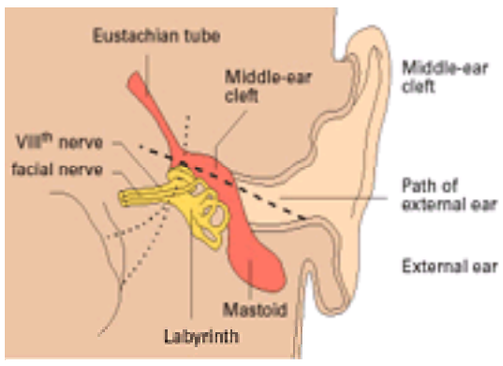


Fig. 2. Longitudinal and transverse fracture lines.

Much greater force is needed to cause a transverse fracture and is usually associated with other severe head injuries, making diagnosis of deafness and facial palsy difficult until consciousness has been regained. The fracture line involves the bony labyrinth and a sensorineural deafness, either from damage to the labyrinth or the

auditory nerve, may occur. An associated facial paralysis, due to the fracture line being perpendicular to the course of the facial nerve, occurs in 50 per cent of cases.

The sign of a temporal bone fracture is a bloody aural discharge that may contain cerebrospinal fluid. Cerebrospinal fluid otorrhoea usually settles spontaneously, but antibiotics are required to prevent meningitis.

When the patient's clinical state allows, radiology and hearing assessment should be undertaken. Even with high-resolution computed tomography (**CT**), fracture lines can be difficult to detect. Middle-ear damage can be repaired but sensorineural damage will only recover if it is concussional. Facial nerve paralysis after trauma is an indication for surgical exploration, provided the patient's condition permits. If electromyography shows severe neuronal degeneration, then decompression, elevation of bony fragments, or grafting may be indicated.

Barotrauma

Most barotrauma occurs during descent in aircraft because of the middle ear pressure being less than atmospheric at ground level. Efficient opening of the eustachian tube will prevent this.

Diving injuries can be due to trapped air in the external auditory canal, occasionally because of ear plugs, allowing the ambient air pressure to increase as the trapped air pressure decreases (so-called ‘reversed ear squeeze') and resulting in haemorrhage to the skin of the ear canal and even rupture of the tympanic membrane. ‘Middle ear squeeze' is experienced on descent because the diver fails to equalize the middle-ear pressure with the increasing ambient pressure. Mucosal haemorrhage and rupture of the tympanic membrane can occur, and, if cold water enters the middle ear, a caloric effect may result causing disorientation and vertigo, which is dangerous under water.

Infection of the middle ear and mastoid

The middle-ear cleft includes the eustachian tube, the middle ear, and the mastoid air cells. Since there is a natural communication between these systems, spread of infection occurs easily. The mastoid is related superiorly to the middle cranial fossa and posteriorly to the posterior cranial fossa. Thus intracranial complications are a possible sequelae of inadequate treatment. Other complications include sigmoid sinus thrombosis, labyrinthine fistula, and facial nerve palsy.

Inflammatory disease includes suppurative otitis media and secretory otitis media.

Acute suppurative otitis media

This is mainly seen in young children and is the most frequent infection in the first year of life. Ninety per cent of children under 5 years of age have had at least one attack, but by the age of 7 years it is rare. This is due to the developing eustachian tube changing from an immature, short, straight tube to a longer, more angled one. Increased upper respiratory infections and an immature immune system are present in infancy. Thirty per cent are viral and 70 per cent bacterial, usually *Streptococcus pneumoniae* and *Haemophilus influenzae*.

Clinical features

Early dull earache gives way to severe pain as the pressure of the exudate builds up, and a high fever along with a conductive deafness occurs. Natural resolution results if the tympanic membrane ruptures with mucopurulent discharge. The membrane normally heals spontaneously.

Treatment

There is controversy about whether or not to give antibiotics for otitis media. Some studies have shown that antibiotics make little difference to the outcome. However, if they are given, a full 7-day course is essential even though symptoms often clear within 48 h. The choice of antibiotic is difficult because of the rise in many countries of b-lactamase-producing organisms; the most logical drug at the moment is a mixture of amoxycillin (amoxicillin) and clavulanate. If resolution has not occurred within 48 h, myringotomy and drainage of the effusion should be considered.

Acute mastoiditis

Acute otitis media implies infection throughout the middle-ear cleft and a potential complication, though rare now, is acute mastoiditis. The mastoid air cells are almost fully developed by the age of 2 years. Pus under pressure within the bony air cells of the mastoid causes destruction and coalescence of the cells so that a pus-filled cavity results and eventually the outer cortex of the mastoid breaks down, forming a subperiosteal abscess. Occasionally, intracranial suppuration may follow. Tenderness over the mastoid process, and particularly over McEwen's triangle (the surface marking of the mastoid antrum), a bulging, red tympanic membrane, and sag of the posterior–superior deep meatal wall with the pinna pushed forward makes the diagnosis obvious. CT scanning shows opacity of the air cells with breakdown of their walls and may show other intracranial complications. Lack of response to systemic intravenous antibiotic therapy is the indication for a myringotomy and cortical mastoidectomy.

Surgical treatment of acute suppurative otitis media and acute mastoiditis

Myringotomy

This is indicated when symptoms, or complications such as a facial nerve palsy, are present after failed antibiotic therapy. Ten per cent of facial nerves in the middle ear are not in a bony canal and the pressure of pus within the middle-ear cavity may affect the exposed nerve.

General anaesthesia is necessary in childhood. The incision (Fig. 3) is in the anterior–inferior quadrant either radially or circumferentially. On no account it should it be carried out posterosuperiorly for fear of damaging the incudostapedial joint. The pus is aspirated through the incision and cultured. Systemic antibiotics are continued and an instant relief of symptoms is to be expected. The incision heals within a week.

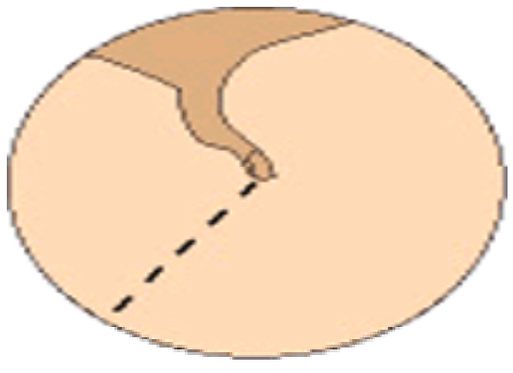


Fig. 3. Radial myringotomy incision in the tympanic membrane.

Cortical mastoidectomy

Under general anaesthesia a postauricular incision is made about 1 cm behind the postauricular groove. In infants it is important to place the incision superiorly since the facial nerve is superficial over the developing mastoid tip. Pus will be encountered from the subperiosteal abscess. The periosteum is elevated off the cortex of the mastoid bone, which is opened with a cutting burr. It is important that the mastoid antrum be exposed, but a complete removal of mastoid air cells is usually difficult due to the bleeding associated with an acute infection. The landmark of the antrum is the lateral semicircular canal, in front of which is the facial nerve. The ideal result of mastoidectomy will be a cavity bounded about by the tegmen separating the mastoid from the middle cranial fossa, behind by the bone covering the sigmoid sinus and

the contents of the posterior cranial fossa, anteriorly the posterior meatal wall, and more medially the antrum leading into the middle ear (Fig. 4). A needle is inserted into the sigmoid sinus to assess patency and rule out thrombosis. The wound is closed in layers and drained for a few days. Antibiotics should be continued until symptoms have resolved.

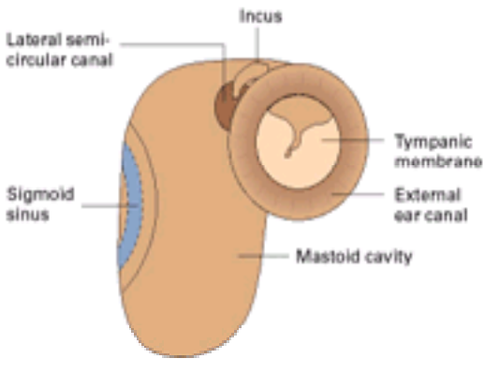


Fig. 4. Landmarks of right cortical mastoidectomy.

Chronic suppurative otitis media

Chronic suppurative otitis media is divided into ‘safe' and ‘unsafe' ear disease. ‘Safe' ear disease is characterized by a central perforation of the pars tensa and is associated with inadequate function of the eustachian tube, and is often called tubotympanic otitis media. ‘Unsafe' ear disease is associated with cholesteatoma and characterized by a perforation of the posterior superior part of the pars flaccida or the margin of the pars tensa (Fig. 5), and is often called atticoantral otitis media. ‘Unsafe' refers to the frequency of complications.

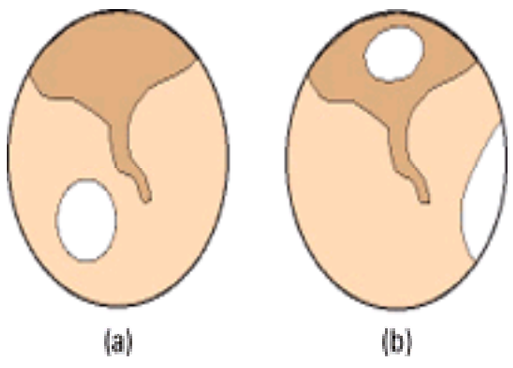


Fig. 5. (a) ‘Safe' perforation; (b) ‘unsafe' perforations—both attic and marginal.

‘Safe' chronic suppurative otitis media

Perforations of the pars tensa, with or without disruption of the ossicular chain, occur following recurrent acute otitis media and may present with intermittent or continuous otorrhoea. It is a chronic mucosal disease often associated with poor function of the eustachian tube.

Surgery may be indicated for hearing loss, waterproofing the ear for swimming, or recurrent otorrhoea following upper respiratory-tract infections. Repair of the tympanic membrane is called myringoplasty and of the ossicles, ossiculoplasty. The term tympanoplasty combines the two. A basic requirement is to freshen the edges of the perforation so that a raw surface is produced to encourage epithelialization over a graft, which acts as a bridge or scaffold.

Myringoplasty

A ‘fat plug' myringoplasty is only suitable for small perforations. Fat is obtained from the ear lobe and inserted into the perforation, the edges of which are freshened to encourage epithelium to migrate and close the perforation.

Larger perforations require a more radical approach involving incisions and skin flaps. The essential requirement is to provide a vascular bed for successful graft ‘take'. The ‘take' of the graft relies heavily on it being accurately placed under a vascular edge of ear-canal skin or tympanic remnant. Temporalis fascia is the common graft material and is usually obtained via a postauricular or endaural incision. The fascia is dried to make handling easier. A posterior meatal-wall flap is created, elevating the edge of the rim of the tympanic membrane forwards. The fascia is inserted deep to the tympanic membrane to lie on its medial surface, the edge of the perforation having been de-epithelialized. The meatal skin is returned to the posterior canal wall with the graft deep to it. The graft is checked for accurate positioning in relation to the perforation and materials such as Gelfoam in the middle ear may help pack the graft against the medial surface of the tympanic membrane. The expected graft ‘take' rate is 80 to 90 per cent at 12 months.

Although ossiculoplasty can be carried out at the same time as myringoplasty, it is frequently staged. A perforation with intact ossicles gives a conductive deafness of about 30 dB, whereas with ossicular discontinuity a 60-dB loss may be expected.

Ossiculoplasty

Reconstructive materials include autograft bone, homograft bone from cadavers (which, because of transmission of viral diseases is no longer popular) or non-biological prostheses, which are made from plastics, ceramics, and various other materials chosen for their bio-compatibility. The major problem with these non-biological materials is rejection but the advantage is their availability and ease of use.

Unless surgery improves hearing to within 20 dB, the patient may not benefit and sometimes the best advice is a hearing aid. Reconstruction of the ossicular chain is most satisfactory when the only abnormality is discontinuity between the incus and stapes. The long process of the incus has a poor blood supply and, in chronic inflammatory disease, frequently disintegrates. A repair with autograft bone glued into place between the head of the stapes and the long process of the incus is very satisfactory. Alternatively, the remnant of the incus is removed and refashioned to fit between the handle of the malleus and the head of the stapes (‘malleus–stapes assembly') (Fig. 6).

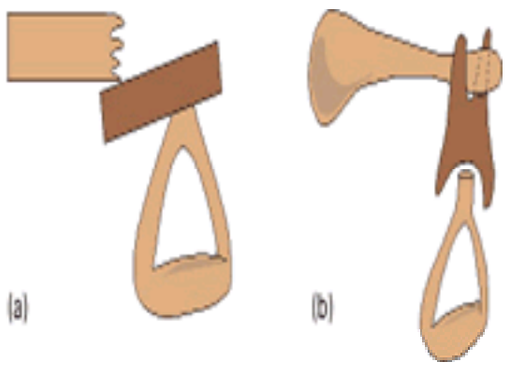


Fig. 6. (a) Bone graft and fibrin glue to repair necrosed long process of incus; (b) malleus–stapes assembly using refashioned incus.

A mobile stapes footplate without superstructure can be connected to the tympanic membrane using a strut made from the remnant of the incus to connect the malleus to the oval window. A prosthetic total ossicular replacement prosthesis (Fig. 7) can also be used. Artificial prostheses tend to extrude, and autograft cartilage between the head of the prosthesis and the undersurface of the tympanic membrane lessens the risk. A partial ossicular prosthesis is also useful to connect the head of the stapes directly to the tympanic membrane. Although short-term results are often satisfactory, no prosthesis is free from rejection and extrusion, and they are expensive. Satisfactory hearing results are occasionally seen in patients when a collapsed tympanic membrane adheres to the head of the stapes, the so-called type III mechanism. There is no point in reconstruction if the hearing is adequate. The future probably lies in the electronics industry with minute, implantable hearing aids, which are currently under clinical trial.

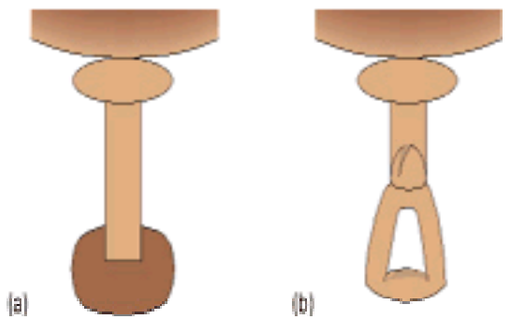


Fig. 7. (a) Total ossicular replacement prosthesis (TORP) to connect tympanic membrane to footplate of stapes; (b) partial ossicular replacement prosthesis (PORP) to connect tympanic membrane to stapes head.

‘Unsafe' chronic suppurative otitis media (cholesteatoma)

The term cholesteatoma suggests a tumour associated with cholesterol. It is neither but more an inside-out bag of skin with a propensity to cause bone erosion. It is still a frequent cause of intracranial infection.

There is debate as to how cholesteatoma forms. The simplest mechanism is the ‘suck-in' theory due a negative pressure caused by dysfunction of the eustachian tube (Fig. 8). Unlike the pars tensa, the pars flaccida has no supporting fibrous layer and chronic middle-ear pressure may cause it to retract and then form into a pocket lined by keratinizing squamous epithelium. Eventually, the neck of the pocket becomes relatively smaller than the sac and keratinized debris is trapped. The sac expands into the mastoid antrum eroding bone, which may include the malleus and incus heads, the lateral semicircular canal, the facial nerve and middle and posterior cranial fossa. Marginal perforations allow migration of squamous epithelium from the ear canal into the middle ear, often causing a destructive middle-ear cholesteatoma.

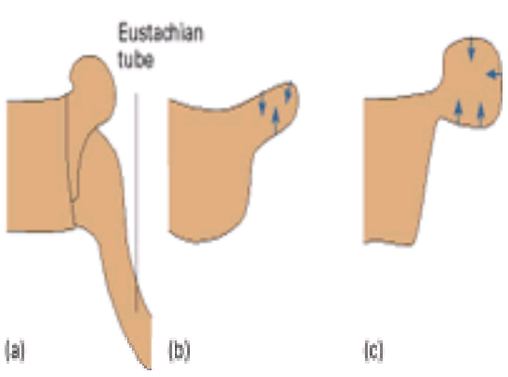


Fig. 8. ‘Suck-in' theory of cholesteatoma formation: (a) attic retraction;(b) attic pocket; (c) attic cholesteatoma (arrows represent keratin collection from desquamated epithelium).

In the early stages, white keratin may be seen in an attic pocket but more often the presentation is a progressive conductive deafness and smelly discharge (from bone erosion) infected with *Pseudo-monas*. Vertigo, on compressing the external auditory canal (fistula sign), suggests erosion of the lateral semicircular canal. Facial nerve palsy should be excluded. Audiology is necessary and CT scanning may demonstrate potential complications such as erosion of bone over the lateral semicircular canal, the facial nerve, or middle or posterior cranial fossa.

Cholesteatoma is a surgically treated disease. Children often have aggressive cholesteatoma in large, pneumatized cavities whereas adult disease occurs slowly in sclerotic cavities. The recurrence rate is greatest in children, complications more in adults.

Surgery for cholesteatoma

Modified radical mastoidectomy is the classic operation. It aims to exteriorize the mastoid system while preserving all possible middle-ear function. The older radical mastoidectomy removed all middle-earcontents and sealed the eustachian tube. ‘Closed' techniques, without leaving an open mastoid cavity, are associated with recurrence in all but a few expert hands. Only modified radical mastoidectomy will be described here but if disease is limited to the attic, atticotomy may be appropriate. Via an endaural incision, ear-canal skin is elevated to the attic defect and the lateral wall of the attic burred away until the cholesteatoma pocket can be excised intact. The resulting defect may be grafted with temporalis fascia.

Modified radical mastoidectomy A postauricular incision is usual, and the posterior ear-canal wall is removed down to the level of the descending portion of the facial nerve. The cholesteatomatous sac, which tracks from the attic deep to the posterior canal wall, through the aditus and into the mastoid system, is thereby exteriorized (Fig. 9). The simplest technique is to open the mastoid cortex and expose the antrum, finding the lateral semicircular canal, which is the landmark for the second bend in the facial nerve between the horizontal and descending portions; staying lateral to this and the short process of the incus, the posterior canal wall is removed from behind down to the level of the vertical section of the facial nerve. Facial nerve monitoring is necessary. The cholesteatoma sac can now be peeled away from the underlying cells and removed completely with a variable amount of tympanic membrane and involved ossicles (usually the incus). If the stapes is present and disease eradicated, ossiculoplasty may be possible. Temporalis fascia grafting to the mastoid cavity is usual to aid epithelialization. Meatoplasty, to enlarge the external auditory canal, allows for careful monitoring and cleaning of the mastoid cavity during outpatient review, which, in satisfactory cases, is on an annual basis to remove wax and debris. Syringing causes a caloric effect from cold water on the exposed lateral semicircular canal and should not be attempted.

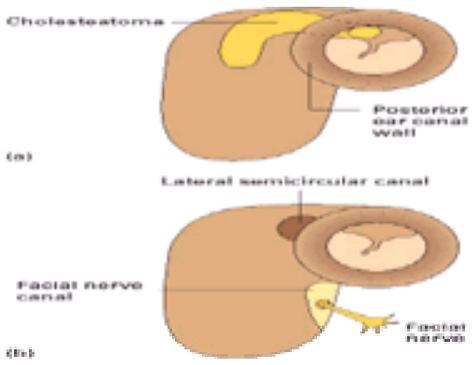


Fig. 9. (a) Cholesteatoma sac deep to posterior meatal wall; (b) end result of right modified radical mastoidectomy with posterior ear canal wall removed down to the level of the facial nerve canal.

Secretory otitis media

Secretory otitis media (glue ear) is the presence of a chronic, non-purulent fluid (serous or mucoid) within the middle-ear cleft. It predominantly affects young children (40 per cent of 2-year-olds have transient bouts but 5 per cent have persistent fluid for over a year). It causes a conductive deafness, but the effect on speech and language development is uncertain.

It is more common in boys, the norm in children with cleft palate, and frequent in craniofacial abnormalities including Down's syndrome. Parental smoking and low birth weight are also associated. A seasonal variation is often seen. Dysfunction of the eustachian tube is the cause, the strongest evidence being the association with cleft palate. Nasopharyngeal infection and obstruction from adenoids almost certainly play a part, and improvement comes with age, regression of adenoids, and elongation of the eustachian tube. Most chronic cases resolve spontaneously but often not until the age of 10 or 11 years, by which time educational difficulties and speech and language problems may have occurred.

The pathology includes mucosal changes with an increase in goblet cells producing a protein-rich exudate (glue ear). The exudate may act as a culture medium for acute otitis media. The under-2-year-old may present with delayed speech and language or recurrent otitis media. Children of nursery-school age may exhibit lack of concentration and difficult behaviour. Children of school age may have obvious hearing loss and learning delay.

Huge variations in screening and detection of deafness exist and parental suspicion is often best.

Hearing assessment

Under the age of 4 years, behavioural observation and distraction tests are used, and rely on the experience of the tester. Pure-tone audiometry should be possible over the age of 4 years and a conductive loss of between 20 to 40 dB is usual.

Tympanometry is an objective way of detecting glue ear but does not assess the degree of hearing loss. It detects movement of the tympanic membrane in response to changing pressure in the ear canal. Figure 10 shows the difference between a normal trace and the flat trace of glue ear. Otoscopy shows a dull, slightly yellow, and often retracted tympanic membrane and the light reflex is usually absent. Once diagnosed, a 3-month wait should occur in the hope of resolution before further testing and treatment are instituted.

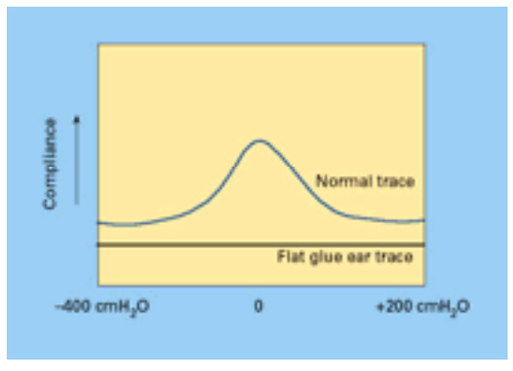


Fig. 10. Tympanogram showing normal and glue-ear traces.

Medical treatment is disappointing. There is no statistical evidence that topical and systemic decongestants, antihistamines, or mucolytic drugs are useful. Antibiotics are only effective in about 20 per cent of patients and then only for short periods; resistant organisms may occur from the use and abuse of antibiotics. Inflation of the eustachian tube with a nasal balloon (Otovent) depends on compliance and the age of the child, but there may be short-term benefits. Hearing aids help deafness but do not cure the disease. *Haemophilus influenzae* and *Pneumococcus* vaccine may prove to be helpful in the future.

Surgical treatment relies on grommet (ventilating tube) insertion, with or without adenoidectomy. A grommet functions as an artificial eustachian tube, allowing air entry to the middle ear and the mucosa to return to normal. Indications are the non-clearance of glue ear over a period of watchful waiting of at least 3 months and which is affecting the child's development, or for chronic atelectasis due to retraction from a negative middle-ear pressure.

Grommet insertion

A general anaesthetic is necessary in children. A radial myringotomy incision in the anterior inferior quadrant of the tympanic membrane is usual and the exudate aspirated. A grommet is inserted through the myringotomy incision. The inner flange is placed in the middle ear, the waist in the incision and the outer flange on the external surface of the tympanic membrane (Fig. 11). Extrusion occurs naturally in 6 to 12 months by the forces of external migration of the squamous epithelium of the tympanic membrane. Once extruded, the defect normally heals spontaneously, but there is a 3 per cent perforation rate. Thirty per cent of children need a second grommet; if there is further recurrence, a long-term grommet may be considered, but these are associated with an increased perforation rate. Another complication is tympanosclerosis. This is hyaline degeneration of the fibrous layer of the tympanic membrane, in which a white plaque is seen on otoscopy, but rarely causes hearing loss. Intermittent otorrhoea, often from the ingress of water through the grommet, may also occur.

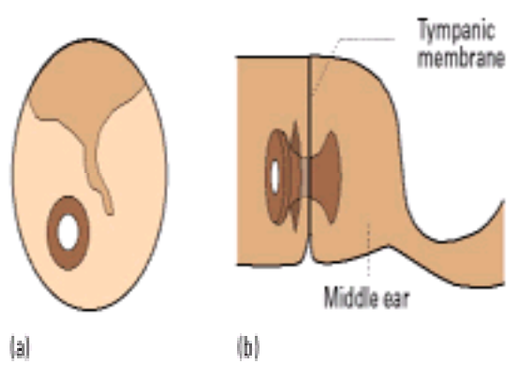


Fig. 11. (a) Otoscopic view of grommet in left tympanic membrane;(b) lateral view.

Swimming with grommets is permitted but not diving, since underwater pressure is likely to force water through the grommet. The surface tension of the water, however, prevents non-pressurized water entering the lumen of the grommet.

Adenoidectomy

Adenoidectomy has a beneficial effect on the function of the eustachian tube and the clearance of middle-ear fluid. Rare complications are postoperative bleeding and velopharyngeal incompetence leading to an abnormal voice. Since only 30 per cent of children have recurrence of glue ear after grommet insertion alone, this should be the first line of treatment, but a grommet insertion plus adenoidectomy should be considered for recurrent disease.

Neoplasia

Tumours may arise from all parts of the ear and temporal bone. Benign lesions of the auricle and external ear will not be considered here. The management of tumours of the middle ear and temporal bone present significant challenges to the neuro-otologist and skull-base surgeon. Whilst the majority of lesions are histologically benign, they can give rise to serious problems because of their effects on surrounding structures. This is particularly true of lesions of the cerebellopontine angle.

Acoustic neuroma (vestibular schwannoma)

The term ‘acoustic neuroma' is a misnomer; these lesions are schwannomas arising from the vestibular nerve within the internal auditory canal. They grow slowly and usually present with unilateral sensorineural hearing loss or tinnitus. Imbalance may occur. Facial weakness is rare. As the tumours grow out of the internal canal into the cerebellopontine angle, the Vth cranial nerve may be involved. Untreated, brainstem compression, hydrocephalus, and death may ensue. The diagnosis is made on the basis of magnetic resonance imaging (**MRI**) (Fig. 12 and Fig. 13). With modern scanning techniques the use of contrast agents such as gadolinium is not always necessary.

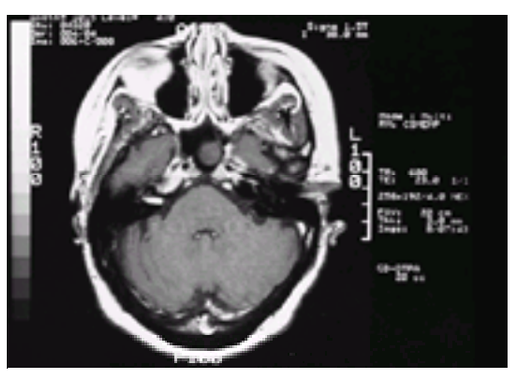


Fig. 12. Small, right-sided, intracanalicular acoustic neuroma *(left of photograph)* on MRI scan.

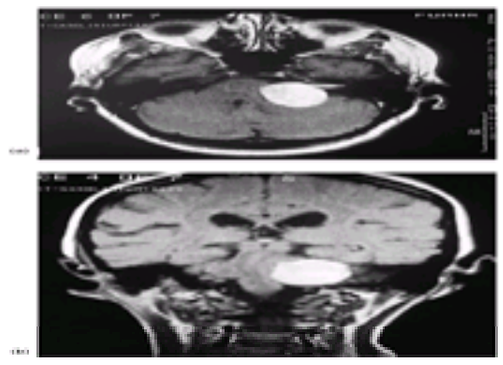


Fig. 13. Large acoustic neuroma visible as a bright spherical lesion on the right of (a) axial MRI scan; (b) coronal MRI scan.

Several treatment strategies have been proposed. Important issues include (a) the size of the tumour and (b) the medical status of the patient. In patients who represent a poor surgical risk, some authorities advocate a policy of ‘watchful waiting', with serial scanning to follow growth. Some tumours fail to grow at all or regress. The argument against this policy is that surgical risk is likely to increase as the tumour grows and the patient will be older, and perhaps less fit, as time progresses.

Radiosurgical treatment with the ‘gamma knife' is another non-surgical option. The merits or otherwise of this treatment method are still being debated.

Several surgical approaches are available for the primary aim of surgery—tumour removal. Preservation of facial movement and of ser-viceable hearing are secondary aims. The emphasis here is on retaining hearing that is, or could be, useful. The hearing in the contralateral ear is important in this regard.

Two factors are important in selecting a surgical approach. First of these is the size of the lesion. A small lesion confined to the most lateral part of the internal canal presents quite a different scenario to the large tumour extending far medially into the cerebellopontine angle. Secondly, the experience of the surgical team is important. A team approach is recommended because of the complementary skills of the neurosurgeon, neuro-otologist, neuroanaesthetist, and electrophysiological monitoring team.

Approaches to cerebellopontine angle

Of the several approaches available to access the cerebellopontine angle, the following are those most usually employed.

Translabyrinthine

The internal auditory canal can be reached via the mastoid if the bony labyrinth is removed. This forms the basis of the translabyrinthine approach. Hearing is inevitably destroyed because the bony labyrinth is removed. This approach is suitable for many acoustic neuromas.

Retrosigmoid

The retrosigmoid approach provides excellent exposure to the cerebellopontine angle. It is possible to preserve hearing using this approach but it is a more invasive procedure involving a craniotomy with retraction of the cerebellum.

Middle fossa

The internal canal can be approached from above via the middle cranial fossa. Through a craniotomy superior to the pinna the temporal lobe is gently retracted and the internal auditory canal deroofed. Hearing preservation is said to be superior with this approach but difficulty reaching the cerebellopontine angle make it unsuitable for larger tumours.

Other benign tumours

Glomus tumours arise from paraganglia within the middle ear (glomus tympanicum tumours) or jugular bulb (glomus jugulare tumours). The former classically presents with pulsatile tinnitus. Diagnosis is confirmed by MRI scan (Fig. 14). Treatment may be expectant (following growth with serial scanning in the elderly or infirm patient), or radiotherapy or surgery may be employed alone or sequentially. Surgery for a glomus tympanicum can be via a per-meatal tympanotomy or mastoidectomy approach. Surgical treatment of larger glomus jugulare lesions is made more challenging by the proximity of important and vital structures such as the facial nerve, jugular bulb, carotid artery, and lower cranial nerves. Several surgical approaches to the lateral skull base are well described.

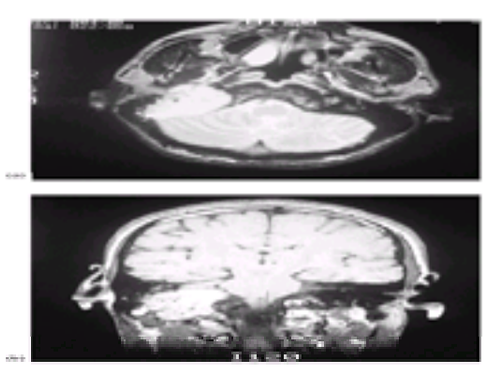


Fig. 14. Glomus jugulare tumour (left of photograph) on (a) axial MRI scan; (b) coronal MRI scan.

Malignant tumours

Thankfully, malignant tumours of the ear and temporal bone are uncommon. The pinna can be affected by squamous-cell carcinoma, basal-cell carcinoma, and malignant melanoma. Squamous-cell carcinoma is by far the most common malignant lesion of the external canal and spread from this site is the usual source of carcinoma of the temporal bone. Malignant tumours of almost all other tissue elements of the temporal bone have been described. Metastatic disease at this site has been reported but is uncommon.

Squamous carcinoma, basal-cell lesions, and malignant melanoma of the pinna are managed by excision, in some cases combined with neck dissection and postoperative radiotherapy. Resection of the lateral temporal bone may be required. Surgical treatment of squamous carcinoma of the middle ear and temporal bone is a major procedure involving subtotal or total resection of the temporal bone and should not be embarked upon lightly. Palliative radiotherapy or less extensive surgery may be more appropriate.

Otosclerosis

Otosclerosis is a familial disorder (probably inherited as an autosomal-dominant characteristic with variable penetrance) in which the compact bone of the otic capsule surrounding the inner ear is replaced by bone that is much more vascular and spongy. The most common site is the fissula ante fenestram just anterior to the oval window. This abnormal bone fixes the stapes within the window and produces a conductive hearing loss (Fig. 15). As the disease progresses a sensorineural hearing loss may develop. Tinnitus is common in these patients but dizziness is rare. Clinical examination reveals a normal ear drum. Audiometry and tuning-fork tests will confirm the presence of a conductive hearing loss. Typically a bone-conduction audiogram shows an increase in thresholds at 2000 Hz referred to as Carhart's notch.

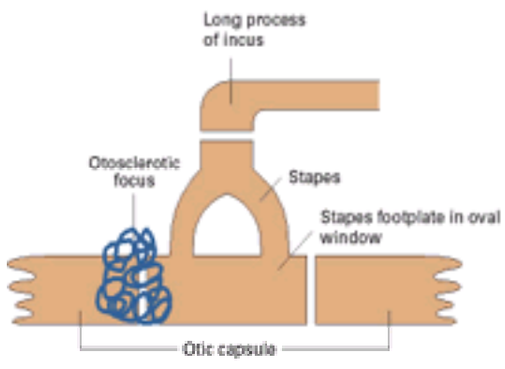


Fig. 15. New abnormal otosclerotic bone fixes the stapes footplate in the oval-window niche.

Non-surgical treatment

The patient may elect to do nothing, especially if the disease is unilateral. An appropriately fitted hearing aid will satisfactorily correct the hearing deficit in many cases. Some find this cosmetically unacceptable.

Stapedotomy

The modern operation of stapedotomy has largely replaced the original stapedectomy procedure developed for this condition. The ear drum is lifted forwards under local or general anaesthesia, and the diagnosis confirmed by palpation of the ossicular chain. The stapes superstructure (head and crura) is removed. A small, 0.8-mm diameter hole (the stapedotomy) is made in the stapes footplate with a microdrill and/or laser. This is covered by a small piece of vein graft. The distal end of a stapes prosthesis is placed on the graft in the stapedotomy and the proximal end attached to the long process of the incus (Fig. 16). The ear drum is replaced.

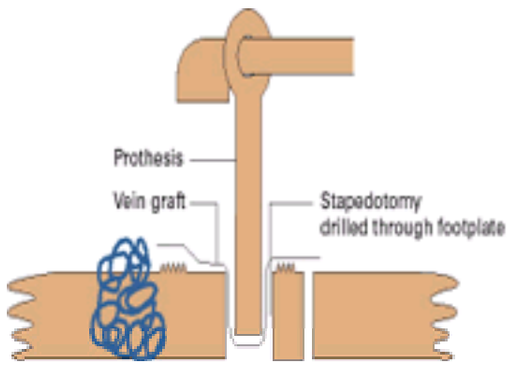


Fig. 16. A stapedotomy hole is drilled in the footplate and covered with a vein graft. The prosthesis sits in the stapedotomy and is attached to the long process of the

incus.

Surgery for otosclerosis should only be undertaken when the patient is fully aware of the risks and complications that may ensue. The risks of this minimal (‘small fenestra') surgery and vein graft are less than those experienced when the whole or even part of the footplate were removed and the prosthesis was positioned in the ensuing defect surrounded by pieces of fat or gelatin sponge. The principal risk of surgery is total sensorineural hearing loss. Whilst the exact risk of this is dependent on the particular surgeon, a figure of 2 to 3 per cent is often quoted. Total sensorineural hearing loss means that future use of a hearing aid on this ear is impossible. Such a loss may occur at the time of surgery or months or years later. Other important risks include balance disturbance and late failure due to detachment of the prosthesis.

Menière's disease

Menière's disease is one of many causes of episodic ‘dizziness'. It is often overdiagnosed (especially in primary care). Characteristic features should be present before the diagnosis is made. The symptoms required are (a) episodes of true vertigo (often a spinning sensation), associated with (b) hearing loss and (c) tinnitus in the affected ear. An attack is often preceded by (d) a sense of fullness in the ear. Guidelines produced by the American Academy of Otolaryngology—Head and Neck Surgery require tinnitus or fullness or both to be present in addition to vertigo and hearing loss to establish the diagnosis. In the typical attack the fullness is followed by the onset of tinnitus and hearing loss. Then comes the attack of vertigo, which lasts 1.5 to 2 h, and may be accompanied by nausea or vomiting. It is followed by a feeling of imbalance for the rest of the day. Attacks may occur in clusters over a period of weeks or days. Over time, the hearing loss (which is sensorineural in type and predominantly low frequency) increases. The disease often ‘burns out' after several years. It is believed to be associated with an increase in pressure within the endolymph of the inner ear and this process is known as ‘endolymphatic hydrops'.

Medical treatment

The aim of medical treatment is to reduce fluid pressure within the inner ear. The patient is encouraged to follow a low-salt diet. In Europe, betahistine hydrochloride is prescribed. A mild diuretic can be added. The acute attacks may be aborted by the use of a vestibular sedative such as prochlorperazine; sublabial versions are available to avoid the problems of swallowing whilst nauseous. In some patients, oral steroids may have a role.

If symptoms are poorly controlled, more invasive treatment should be considered. In doing so, it must be accepted that Menière's disease is a disorder with a high rate of natural resolution and that several placebo-controlled studies have shown some placebos to be remarkably effective in controlling symptoms. It has even been suggested that simply contemplating surgery may lead to a clinical improvement.

Surgery

Surgery for Menière's disease may aim to preserve hearing whilst modulating or abolishing vestibular function, or it may be destructive, destroying both hearing and balance in the affected ear.

Endolymphatic sac decompression

The endolymphatic sac is decompressed in the belief that, in so doing, endolymphatic pressure is reduced. A cortical mastoidectomy is performed and the sac identified within the posterior fossa dura on the posterior surface of the temporal bone. It is incised and a drain may be inserted. The procedure does not affect hearing. If an initial improvement is not maintained the procedure can be revised.

Vestibular nerve section

The vestibular nerve can be sectioned (preserving the cochlear nerve and hence the hearing) to prevent the vertiginous symptoms of Menière's. This can be performed via a retrolabyrinthine route or via the middle cranial fossa using an extradural approach to the internal auditory canal. There is a risk of hearing loss and facial palsy with this procedure, and if all the fibres are not divided the results may not be wholly satisfactory.

Labyrinthectomy

If hearing is poor and unserviceable a formal labyrinthectomy may be performed. This can be undertaken via the middle ear, with opening of the oval and round windows and disruption of the membranous labyrinth. Alternatively, via a mastoidectomy approach, the bony labyrinth can be removed entirely. All hearing is lost.

Medical labyrinthectomy with hearing preservation

The tendency of some ototoxic antibiotics to be more vestibulotoxic than cochleotoxic has led to the use of intratympanic gentamicin to perform a medical labyrinthectomy. In this technique the antibiotic is instilled at regular intervals into the middle-ear cleft. It diffuses into the inner ear via the round-window membrane and destroys vestibular function. Hearing is monitored and treatment discontinued if significant hearing loss occurs. This procedure has several merits, not least a low morbidity profile and the avoidance of any risks associated with major ear surgery.

Management of conductive deafness

There are several causes of conductive deafness in addition to otosclerosis described above. Among the most common are the sequelae of chronic otitis media; perforations of the tympanic membrane, and losses of portions of the ossicular chain, most often the long process of the incus.

Medical treatment

Several options exist for non-surgical management of conductive hearing loss. The status of the opposite ear is of fundamental importance when considering all of these. If the patient has a normal ear on the contralateral side they will often manage reasonably well in many situations; they may even become unaware of their loss. However, in some situations such a loss is problematic and patients with bilateral losses often have a significant disability. A hearing aid represents a good option for management in such cases and is free of risk. Patients with normal levels of sensorineural hearing and speech discrimination can make excellent hearing-aid users since the amplified sound is ultimately transduced by a normal cochlea. They should be distinguished from those with a degree of sensorineural loss in whom amplification of sound may not be accompanied by a commensurate improvement of clarity because of poor discrimination. Having said this, it must be accepted without irritation that many patients find hearing aids cosmetically unsatisfactory. In addition, those with abnormally shaped ear canals and/or discharging ears may be unable to wear an aid.

Surgical treatment

Some forms of conductive hearing loss are amenable to reconstructive surgery. Stapedectomy, myringoplasty, and ossiculoplasty have already been discussed. As with all ear operations they carry a small risk of sensorineural hearing loss. It is pertinent to bear in mind that on occasions an increase in auditory thresholds measured by the surgeon is not reflected by improvements in the quality of a patient's hearing. Neither the patient nor surgeon will be disappointed if both parties have realistic expectations about the outcome of surgery.

Bone-anchored hearing aids

Patients with abnormal or chronically discharging ears have in the past been provided with bone-conduction hearing aids. A vibrator attached to a head-band sends impulses through the bone of the mastoid to the inner ear. The advent of osseointegrated titanium implants led to the development of a percutaneous osseointegrated abutment to which a hearing aid could be more effectively attached (Fig. 17). Not only are such devices cosmetically more acceptable than the old aids, the quality of the hearing produced is improved.

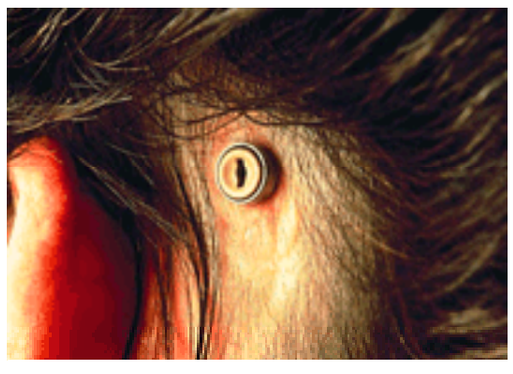


Fig. 17. The percutaneous osseointegrated abutment to which a bone-anchored hearing aid is attached.

Management of sensorineural deafness

Sensorineural hearing loss is common. The most common cause is the ageing process and the result is termed presbyacusis. A full discussion of sensorineural hearing loss is beyond the scope of this chapter, but there are a number of disorders with surgical implications.

Sudden sensorineural deafness is a medical emergency. In many cases no specific cause is identified but empirical treatment with medications designed to improve cochlear blood flow, reduce inflammation, and eradicate viral infection are thought to be effective. One important cause is a perilymph fistula. This condition can arise as a result of violent straining or of sudden pressure changes. The round-window membrane or oval-window ligament are ruptured and perilymph leaks into the middle ear. There may be associated dizziness and/or tinnitus. Many leaks settle spontaneously with rest. However, it may be necessary to perform a tympanotomy, elevating the ear drum and entering the middle ear to inspect the regions mentioned. If perilymph leakage is encountered, the leak may be plugged with fat or fascia. This is a controversial area for several reasons. Some authorities believe in spontaneous perilymph fistulas (leakage occurring without an antecedent history of straining or pressure change); others do not. It has been shown that when a middle ear is examined by several surgeons they do not always agree on whether or not a leak is present; it is not uncommon for tissue fluid to be found in the middle ear when there is no question of a leak.

Cochlear implants

Surgery now has a much greater part to play in the management of sensorineural deafness than in the past, thanks to the development of cochlear implants. These devices have an established place in the management of adults and children with severe or profound deafness. An electrode array is inserted into the basal turn of the cochlea via a mastoidectomy procedure. The array is connected to a receiver/stimulator package sited in a bed drilled in the cortical bone posterosuperior to the ear (Fig. 18). This is buried beneath the skin. The patient wears a device behind the ear that looks like a small hearing aid. This picks up the auditory signals and conveys them to a processor, where they are analysed by a customized programme. The processor sends signals to a coil that overlies the embedded receiver/stimulator and through which both the signals and power are sent to the implanted device. The intracochlear electrodes are then stimulated in a pattern determined by the processor. Until recently, the processor was worn on the belt. There are now much smaller devices available that are worn behind the patient's ear, providing cosmetically much more acceptable results.

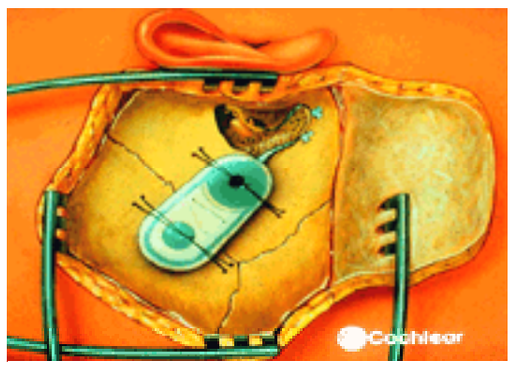


Fig. 18. A multichannel cochlear implant with the receiver/stimulator package and intracochlear electrode array. (Surgical position; right ear.)

The ability of a patient to derive benefit from a cochlear implant varies. Whilst a few manage to communicate freely on the telephone, the majority find the device useful as an aid to lip reading. A few derive benefit even though they may simply be made aware of environmental sound. This may seem a modest gain, but for the profoundly deaf patient it can be a remarkable improvement.

Further reading

Jackler RK, Brackmann DE, ed. *Neurotology*. Mosby-Year Book, St. Louis, 1994. [Comprehensive review of neurotological diagnosis and management.]

Tos M. *Manual of middle ear surgery*. Thieme, Stuttgart, 1993 (Vol. 1), 1995 (Vol. 2), 1997 (Vol. 3).

45.9 Temporomandibular joint dysfunction

Richard P. Juniper

Anatomy and physiology
Aetiology
History
Examination
Extraoral examination
Intraoral examination
Investigations
Treatment
Appliance therapy
Drug therapy
Surgery
Further reading

Temporomandibular joint dysfunction (myofascial pain dysfunction) consists of pain in and around the temporomandibular joint, joint sounds, and limitation of opening the mouth. A more recent term, facial arthromyalgia, describes the condition more precisely. The symptoms can range from a minor click to a debilitating pain. In its varying forms it is very common, affecting up to 25 per cent of the population, with a female:male predominance of up to 5:1. Most patients present at the age of 15 to 25 years, although a few will complain of trouble up to the age of about 45 years. Beyond this age the condition is uncommon, but in recent years there has been a tendency for adolescents under the age of 15 years to present.

Anatomy and physiology

The temporomandibular joint is a diarthrodial joint with articulation between the mandibular condyle and the skull base. Interposed between the bony structures is a meniscus or disc that envelops the head of the condyle and has a similar shape to that of a jockey's cap (Fig. 1). It separates the bones, creating two synovial spaces, which allow rotation in the lower space and a gliding movement in the upper. Its shape spreads load from the convex condyle on to the varying shape of the glenoid fossa and eminence.

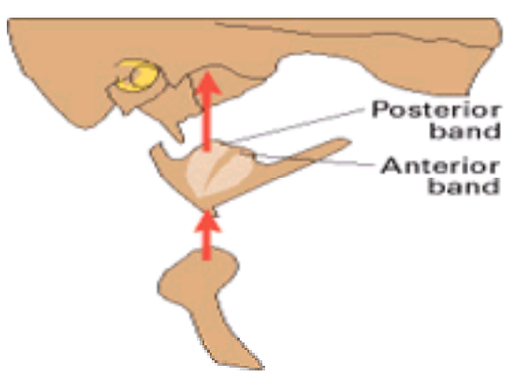


Fig. 1. An expanded view of the temporomandibular joint.

The movement of the condyle as the mouth opens is complex (Fig. 2). With the teeth in occlusion the condyle lies approximately in the centre of the fossa. When the mouth is opened, the condyle rotates and translates downwards and forwards across the backward facing slope of the articular eminence. On wide opening, the condyle may pass forward until it leaves the eminence completely. Closing follows a reverse path. As the condyle translates, the meniscus does so too at approximately half the rate, much as the slide does on a filing-cabinet drawer. The shape and elasticity of the meniscus allow it to accommodate the ever-changing shape of the articular surfaces. It has a posterior band approx. 3 mm thick lying immediately over the condyle in the closed position, an anterior band, and an intermediate zone between the two. The two bands converge and fuse at the medial and lateral poles of the condyle and in effect make an elongated ring into which the anterior part fits (Fig. 3). This ring maintains its position abreast the condyle, held by insertions at the medial and lateral pole, by the close apposition of the joint and by the joint capsule. Behind the posterior band, elastic fibres insert into the squamotympanic fissure. Under the load of mastication, the joint is stabilized against the eminence by contraction of the upper head of lateral pterygoid, which is inserted partly into the anterior band and partly into the condylar neck. It contracts as the mouth closes.

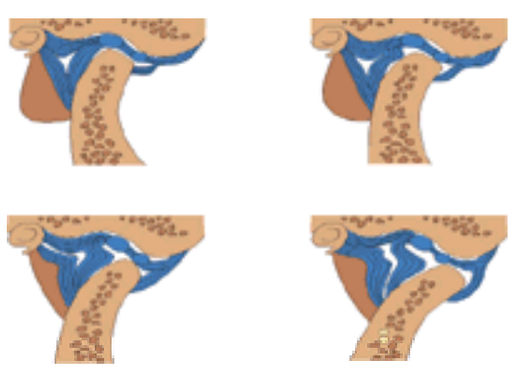


Fig. 2. Mode of condylar movement as the mouth opens.

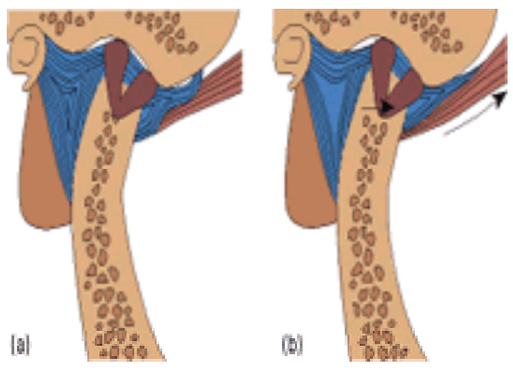


Fig. 3. The two bands of the meniscus converge at each pole. The bands represent an elongated ring into which fits the head of the condyle.

For the joint to move smoothly and without dysfunction, there must be well-coordinated muscle action, an uninflamed synovium, and a correctly positioned meniscus.

Where these fail, temporomandibular joint dysfunction (facial arthromyalgia) may result.

Aetiology

The cause of temporomandibular joint dysfunction has been a contentious issue over many years. Many hypotheses have been put forward, from occlusal disharmony to psychosomatic illness. In recent years the aetiological background has become clearer. The painless click is created by sudden movement of the posterior band of the meniscus. Throughout life the posterior band migrates forwards and medially as a normal process of ageing; this is a subluxation of the meniscus. Changes in the bony contour normally allow smooth movement to continue to occur, but in time this does not happen. In these cases the posterior band has a sudden movement backwards over the condyle on opening and forwards on closing. Whiplash injuries are said to damage the posterior ligamentous structures and allow the meniscus to sublux acutely, although the evidence for this is sparse. Progressive subluxation may result from joint overload. Joint overload causes degeneration of the fibrocartilaginous articular surfaces, and results in increasing friction and tension in the posterior ligamentous insertions of the meniscus. These may weaken, stretch, and fail, allowing the subluxation to occur. Chronic compressive stress may cause the joint surfaces to change shape, particularly where the meniscus is subluxed; this leads to radiographic changes and the ectopic meniscus will be detected on magnetic resonance imaging (**MRI**). However, the deviation from normal has little prognostic value in most cases of facial arthromyalgia, as the joint has great powers of healing and adaptation. It is generally considered that the changes in the joint are caused by overload, but there is disagreement as to how the original overload occurs.

In the past there has been an emphasis on the changes in the joint in the condition of facial arthromyalgia. While the joint may give pain (arthralgia), the majority of patients have pain spreading over the side of the face and head. This is caused by pain from the muscles of mastication (myalgia), notably the masseter, medial pterygoid and temporal. The muscles will be tender to palpation. The mechanism for this pain is uncertain, but again it is thought that it is due to overload. If the joint becomes inflamed and painful from parafunctional habits, such as nail biting, clenching, and bruxing, the muscles become inflamed too.

A popular hypothesis, held for many years, is that occlusal discrepancy can lead to stress on the temporomandibular joint. Occlusal interferences encourage the patient to bite in an awkward way and stimulate bruxism. This has never been proved to be a primary cause of facial arthromyalgia. Lack of posterior teeth may be associated with facial arthromyalgia and it is thought that this creates overload on the joint structures and muscle simply because the individual has to protrude their jaw to masticate their food. Protrusive habits, particularly nail biting and lip chewing, are predisposing factors that seem to confirm this suggestion.

There is no doubt at all, however, that stress plays a very important part in this condition. Some people respond to stress by clenching their teeth during the day, and by bruxing at night. Patients suffering from facial arthromyalgia and exposed to acute stress will tend to show hyperactivity of the masticatory muscles over a number of subsequent nights. This hyperactivity, if it persists, overloads the joint and muscles. Overload produces synovitis and changes in the fibrocartilage, and a degenerative sequence may begin in the joint. The muscles ache. If parafunctional habits can be reduced and stress-related clenching and bruxing can be reduced, the joint surfaces may heal and the muscles recover, but the bone is permanently remodelled. There is much evidence to suggest that facial arthromyalgia is a self-limiting condition, particularly when stress is eliminated, or where the parafunctional habits produced by that stress are corrected. The changes in the joint seen on imaging still allow a painless, fully functional joint in most cases.

Any condition in the mouth that encourages the patient to find a more comfortable position in which to occlude their teeth may produce the pain of arthromyalgia. Such intraoral anomalies should be sought and their elimination included in the treatment plan. Anomalies found in the occlusion should not be changed irreversibly (by grinding/reshaping the teeth) until the muscles of mastication are painless and functioning normally. Various forms of appliance (biteguards) placed between the teeth have been shown to help to eliminate pain.

History

Some people presenting with facial arthromyalgia have experienced trauma to their masticatory apparatus in the past, such as a blow to the jaw or a wide yawn. This may have led to an acute pain with limitation of jaw movement. A painless click or joint noise may follow initial recovery. Some months or years later a further episode may occur, with pain over the joint and over the side of the face and head, and further limitation of opening. The pain may become severe at the limit of opening and the click may cease. Recovery follows, only to return after an inadvertent wide yawn, or a prolonged visit to a dental practitioner. The condition may arise without a history of trauma, however.

Facial arthromyalgia may arise and have exacerbations at times of stress: there is often some joint pain, but mainly an ache is felt in the muscles of mastication. These patients fall mainly into two groups: a large group that has pain and trismus on waking, and a smaller group in whom the pain comes on later in the day. Simple analgesics help but do not abolish the pain and, in a few patients, the discomfort is so severe and prolonged that enjoyment of life is impossible. Many will be found to have parafunctional habits such as bruxing, clenching the teeth in response to stress, lip biting, and nail biting. One side tends to be affected predominantly. Rarely, younger patients may present with very severe acute symptoms that are incapacitating.

Examination

The examination starts and the treatment begins as the patient enters the room. It is important continually to assess their emotional state as the history is taken and they are examined. Specific questions should be directed to such as family problems and school examinations.

Extraoral examination

The general demeanour of the patient is noted, with any abnormalities of facial contour and swellings. The temporomandibular joints should be palpated bilaterally over the lateral poles of the condyles, both when stationary and in motion. Tenderness suggests local inflammation. Any clicks identified should be timed in relation to opening and closing. From this the position of the meniscus may be assessed: an early opening click suggests a minor displacement, a late one a major displacement. Joint crepitus may also be detected. The muscles of mastication should be palpated with the teeth clenched. Tenderness suggests hyperactivity. The distance between the upper and lower incisor teeth on maximal opening should be measured (normal, 40–55 mm) and any deviation from the midline on opening noted. This is the most sensitive test for temporomandibular joint dysfunction. Tenderness of the sternomastoid and suboccipital group of muscles suggests a more generalized hyperactivity, and indicates a disorder beyond the masticatory apparatus.

Intraoral examination

All patients should undergo general intraoral examination. The mucous membranes of the tongue and floor of mouth should be inspected; the teeth and gums should be examined for caries, periodontitis, and pericoronitis in relation to an unerupted third molar. The patient should be asked to tap the back teeth together: a normal occlusion produces a solid single click, while discrepancies will be recognized as an irregular sound. Lack of posterior teeth may be noted. Patients should be asked to go into each lateral excursion of the lower jaw and any painful areas in the occlusion noted. An edentulous patient with old dentures may be ‘overclosed' due to loss of alveolar bone. This is a common cause of temporomandibular joint pain in elderly people.

Investigations

The most potent investigation of patients with temporomandibular joint dysfunction is the examination. Plain radiographs, either in the form of orthopantomograms or even close-cut corrected tomograms of the joint, have been shown to be of no value except in patients with extreme and chronic disease. Computed tomographic scans may be of value, especially where serious disease in the bone is suspected, and MRI will demonstrate soft tissues and can ascertain the position of the meniscus. Arthrography is now rarely used, but assesses the condition and mobility of the various elements of the joint. Perforations in the meniscus and adhesions may be seen. Arthroscopy, using arthroscopes of around 2-mm diameter, can visualize joint structures, mainly of the upper space. It can be used therapeutically as a means of lysis and lavage, and for release of adhesions. In selected cases, and in experienced hands, joint structures may be sampled, curetted, injected, and even sutured.

Where other causes of joint pain are suspected (rheumatoid arthritis, gonococcal arthritis, for example), appropriate blood tests and investigations should be instituted.

Treatment

Patients may present with varying degrees of arthralgia and myalgia. Arthralgia alone is uncommon: limitation of opening (locking) with the pain restricted to the joint suggests a subluxed meniscus with inflammation and fibrosis. Crepitus may indicate inflammatory exudate or arthrotic degeneration but is not of prognostic significance. However, most patients will have a large element of myalgia (painful, tender muscles). Where this is evident, treatment is directed primarily at the causes

of muscle hyperactivity.

At least 90 per cent of sufferers from temporomandibular joint dysfunction respond to conservative treatment. Treated with consideration, 40 per cent will get better within 6 to 8 weeks with little more than reassurance. Explanation of the cause of the pain helps them to identify parafunctional habits, such as clenching or nail biting, and to eradicate them. A simple exercise regimen (Table 1) has been found most beneficial. It is probable that much of the improvement arises as a result of reassurance that nothing is seriously wrong.



Table 1 Example of an instruction sheet for exercises to improve the function of the temporomandibular joint

Those who fail to improve on that regimen should be questioned to explore the possibility of an underlying cause—stress at home or work. While this may need to be addressed, additional treatment can be instituted, aimed at reducing muscle hyperactivity. There are two main approaches to this stage: appliance therapy and drug therapy. Some dental practitioners feel that there is benefit in equilibrating the occlusion by spot grinding (coronoplasty), but, as outlined above, this must be reserved until muscle pain and tenderness is relieved by other methods.

Where appliance or drug therapies fail, or where the patient has had intractable pain over a number of years with failed conservative treatment, then there is benefit in investigating the joint further by arthrography, arthroscopy, or MRI (see above).

Appliance therapy

Many types of intraoral appliance have been constructed for the management of facial arthromyalgia. They may be fitted to the upper or lower jaw, worn continuously, or only at night. Research has shown that it is the parting of the teeth which seems to bring the benefit; an opening of 3 to 4 mm is generally found to be the most satisfactory, although wider openings seem to bring improvement more rapidly. Wearing an appliance at night seems to be sufficient. Proponents of this method report considerable success. The appliances should not be used for more than 3 to 4 months as they can produce unwanted changes in the occlusion.

Drug therapy

Tricyclic drugs are most successful, usually in doses well below those for the treatment of depression. Sedative forms such as dothiepin are said to impart more benefit than less sedative forms such as nortriptyline. They act to reduce anxiety, improve sleep patterns so reducing bruxism, and may function as a centrally acting analgesic. Doses start at between 25 and 50 mg at night, and are increased every 3 to 4 weeks until a good response is obtained. It is important to explain to the patient why the drugs are being given, to discuss the side-effects, and to explain that there may be a slow response. With a very agitated or anxious patient, some mild sedative such as fluphenazine or trifluoperazine may be added, but never prescribed for more than 3 months. Diazepam does not seem to be effective and can lead to dependence. With this regimen a further 40 or 50 per cent of patients will get better in 2 to 3 months.

Surgery

The aim of surgery is to reconstruct the joint. The conservative forms of surgery have a reported success rate for facial arthromyalgia of up to 90 per cent.

Meniscoplasty

The meniscus is identified through a preauricular incision and arthrotomy. It is mobilized by incising adhesions, and then sutured into a more posterior and lateral position. Postoperative MRI has shown that the repositioning of the meniscus is not permanent, and the success of the procedure probably relates to removal of adhesions.

Menisectomy

This is reserved for occasions when the meniscus cannot be mobilized satisfactorily, or where it is perforated or irretrievably damaged. A temporal muscle flap may be mobilized and sutured across the joint surfaces to make a satisfactory substitute for the meniscus, although recent research suggests that menisectomy alone is just as satisfactory.

Artificial materials

Many artificial materials have been used to replace the meniscus, the most popular in recent years being a bilaminate of Teflon and Proplast. There have been so many complications of this procedure that most surgeons have abandoned it.

Arthroscopic surgery

Recently, techniques have been developed using arthroscopy with a second puncture to introduce minute surgical instruments. The joint is inspected with up to ×15 magnification, lavage can be performed to remove inflammatory substances where synovitis is identified, anti-inflammatory drugs may be injected directly into the inflamed tissues, and adhesions can be incised.

Joint replacement

Materials have been developed for total replacement of the temporomandibular joint. Generally these are reserved for gross joint destruction such as post-traumatic ankylosis, or where the joint is severely damaged by rheumatoid disease.

Further reading

Norman JE de B, Bramley P, ed. *A textbook and colour atlas of the temporomandibular joint*. Wolfe Medical, London, 1990. [A multiauthor book covering all aspects of temporomandibular joint disorders.]

45.10 Oral and maxillofacial surgery

Joseph W. Wilkes, David A. Keith, Carol A. Lorente and John A. Buehler

General preparation
Common maxillofacial surgical approaches
Extraoral approach to the mandible
Approach to the submandibular gland
Intraoral open approach to the mandible
Intraoral open approach to the maxilla
Extraoral open approach to the maxilla
Orthognathic surgery
Maxillary osteotomy (Le Fort I)
Maxillary osteotomy (anterior segmental)
Bilateral intraoral oblique osteotomy of mandible
Bilateral sagittal osteotomy of the mandible
Genioplasty
Preprosthetic surgery
Surgery of the temporomandibular joint
Surgical approaches
Internal derangement of the joint
Joint reconstruction
Reconstruction of mandibular discontinuities
Treatment of mandibular fractures
Uvulopalatopharyngoplasty
Further reading

General preparation

Oral and maxillofacial procedures are usually done in the operating room with the patient under nasoendotracheal general anesthesia and in a supine position, with a specially designed plastic dish or towel roll for occipital head support. The trunk, head, and neck can be elevated slightly to reduce blood loss.

The surgical field is prepared by perioral swabbing with a soap solution for intraoral procedures, but oral irrigation with antiseptic solution is not usually necessary. Surgical sites outside the mouth are prepared with povidone iodine solution. Drapes are placed around the surgical site, or around the entire head and neck, depending on the size of the surgical field.

Oroendotracheal anesthesia can be used for some procedures if nasoendotracheal intubation is impossible for anatomic or other reasons. Examples are the reduction of zygomatic fractures, enucleation of jaw cysts, and tumor resections. Procedures that require intermaxillary fixation are not undertaken with oroendotracheal intubation unless the fixation can be applied after the patient has fully awakened from anesthesia. Tracheostomy is used where intermaxillary fixation is required as part of the surgical procedure if nasoendotracheal intubation is impossible.

Common maxillofacial surgical approaches

Extraoral approach to the mandible

Access to any portion of the inferior border of the mandible is gained by a skin incision 2 cm inferior, and parallel to, that border (Fig. 1). The length of the incision depends on the amount of access needed. Typically, an incision of 5 cm is sufficient for open reduction and internal fixation of a mandibular fracture. Neoplastic or large cystic lesions require additional length for access.

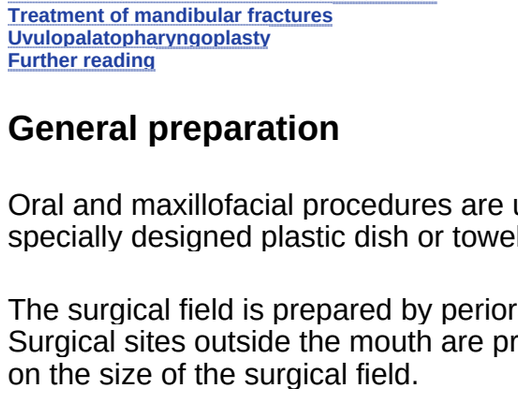


Fig. 1. The extraoral approach to the mandible is made 2 cm inferior to the mandibular lower border (I). This generally allows avoidance of the mandibular branch of the facial nerve (F), which can be identified deep to the platysma and retracted superiorly. The facial vessels (V) might be encountered, depending upon the portion of the mandible approach. They can be retracted or tied and divided. The submandibular gland (G) is deep to the mandible and can also be approached this way.

Dissection begins with sharp incision of skin and subcutaneous tissues to the platysma muscle. Sharp dissection continues very carefully through this muscle layer until the superficial layer of deep cervical fascia is encountered. Within this layer is the marginal mandibular branch of the facial nerve. The position of the nerve is tested with an electrical nerve stimulator.

Once the facial nerve has been retracted superiorly within the flap, dissection continues deep towards the inferior border of the mandible. The facial artery and vein are encountered along the mandibular body at the anterior attachment of the masseter muscle. These vessels may be cross-clamped and divided. Periosteum at the inferior border is incised if direct access to the bone is desired. For neoplasms and cysts involving soft tissue, that tissue is resected with blunt and sharp dissection assuring adequate margins. For large, perforating, or malignant lesions, the mandible is resected with its overlying soft tissue. A radical neck dissection is completed in continuity with the mandibular resection for malignant diseases.

The wound is closed in layers, with approximation of the periosteum whenever practical. Closure of the platysma permits accurate subcutaneous and skin closure. Drains are unnecessary except when resection of a large portion of muscle or bone makes point hemostasis incapable of controlling bleeding. In such a case, a suction drain to the tissue bed is placed through a separate puncture.

Approach to the submandibular gland

The approach to the submandibular salivary gland is similar to the extraoral approach to the mandible with the exception that the incision is made 1 cm lower in the neck, is curvilinear, and is optimally placed in a skin crease.

Dissection continues deep through the platysma muscle. The marginal mandibular nerve is carefully isolated as it courses through the superficial layer of deep cervical fascia. The facial vein and artery are cross-clamped, divided, and retracted superiorly. The capsule of the submandibular gland is exposed by blunt dissection. With careful posterior retraction of the gland, the mylohyoid muscle can be visualized. This forms the deep portion of the gland bed, and it is here that the lingual and hypoglossal nerves as well as the submandibular duct come into view. The lingual nerve is deep as it courses anteriorly and then runs superior to the submandibular duct. The gland is freed by double ligation and division of the duct as well as blunt dissection along the posterior surface. The facial artery is again encountered along

the undersurface of the gland and is ligated and divided, allowing complete removal of the gland.

Hemostasis is best achieved with bipolar electrocautery. Closure is made in layers after meticulous hemostasis. This will eliminate the need for a drain. Closure of the platysma allows for good approximation of the subcutaneous and skin layers.

If malignancy or disease clearly extend beyond the capsule of the gland, resection is performed with sufficient tissue margins. A radical neck dissection is part of the operation for malignancy, with removal of the gland as part of the composite resection.

Intraoral open approach to the mandible

The mandible can be approached intraorally for dentoalveolar surgery, the placement of dental implants, biopsy, removal of cysts and benign neoplasms, and for the open reduction and internal fixation of certain types of fractures.

In the dentate patient the incision is confined to the gingival sulcus between the marginal gingiva and the teeth. A vertical releasing incision may be necessary at one or both ends of the sulcus incision. A mucoperiosteal flap is reflected, providing good exposure of the mandible. The extent of the incision is determined by the amount of exposure needed to accomplish the procedure with adequate access and visibility and to allow the incision line to heal over solid bone. The incision in edentulous patients is through the attached gingiva along the crest of the alveolar ridge, with or without a releasing incision. A mucoperiosteal flap is reflected.

The approach for open reduction of symphyseal fractures differs in that the incision is in the mandibular labial vestibule, allowing 1 cm of alveolar mucosa below the line of the attached gingiva (the junction between mucosa that is firmly attached to alveolar bone and gingiva that is mobile). This incision gives more convenient access to the bony chin. The mental nerves, which are anterior extensions of the inferior alveolar sensory nerves to the lower lip, leave through foramina inferior to the bicuspid (premolar) teeth and then branch under the incised mucosa. Superficial dissection through the vestibular mucosa identifies these fine white filaments, which can then be isolated by sharp and blunt dissection and retracted for protection. The mentalis muscle is divided and its attachment left on bone so that, during closure, reattachment prevents a soft-tissue chin deformity. The mucosa is closed as the final layer. This approach can also be used for various types of genioplasty.

The intraoral approach to the mandibular ramus for osteotomy is accomplished by a 5 cm incision starting posterolateral to the last molar tooth over the midpoint of the external oblique ridge and extending anteriorly and lateral to the first and second molars. A 1-cm margin of gingival tissue is allowed between these teeth and the incision. A subperiosteal dissection then exposes the mandibular ramus, which is carried lateral to the ramus for vertical subcondylar osteotomy when a mandibular setback is desired. The dissection is carried medial to the ramus, superior to the inferior alveolar nerve, for a sagittal osteotomy if mandibular advancement is desired. Closure is accomplished at the mucosal level only. Drains are unnecessary.

Intraoral open approach to the maxilla

Access to the maxilla for osteotomy, biopsy, sinus curettage, and resection of small benign lesions is obtained by an incision in the maxillary vestibule parallel with, and 1 cm superior to, the mucogingival line (Fig. 2). The length of the incision is governed by the area of the maxilla that needs to be visualized. The incision is carried directly to bone, and the tissue is reflected superiorly in the subperiosteal plane as a mucogingival flap. The infraorbital nerve, a branch of the trigeminal nerve supplying sensation to the skin of the cheek, lateral portion of the nose, upper lip, and the mucosa of the upper lip and buccal alveolus, is identified and preserved as it leaves from the infraorbital foramen on the anterior wall of the maxilla.

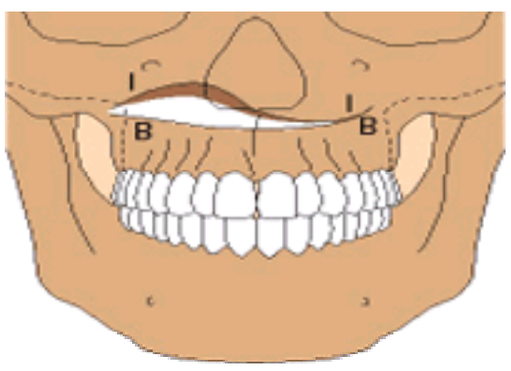


Fig. 2. Incision (I) for the intraoral approach to the maxilla. In this case the incision extends from the zygomatic buttress (B) on the right, across the midline to the buttress (B) on the left. This is for access to the complete maxilla. For access to a portion of the maxilla, such as in unilateral sinus surgery, an appropriate portion of this incision can be used.

For access to the maxillary sinus an incision is made over the canine fossa, which is on the lateral wall of the maxilla just posterior to the canine tooth. Once the flap has been elevated a drill is used to create a window into the sinus through the fossa, taking care to remain 4 to 5 mm above the apices of the teeth. This allows access to the sinus for curettage, removal of a foreign body, or removal of a mucocele.

Before closure the sinus is irrigated and, if indicated, an antrostomy can be made on the medial wall into the floor of the nasal cavity. To provide dependent drainage of the sinus, a Penrose drain is placed from the sinus, leaving the nose through the antrostomy, for the first 24 to 48 h. The intraoral mucosal incision is closed without a drain in a running fashion with resorbable suture.

There are several approaches to obtain access for an anterior maxillary osteotomy. If superior repositioning of the anterior segment is desired, a horizontal vestibular incision is made extending one tooth posterior to the planned osteotomy site. A vertical incision is made through the buccal mucoperiosteum in the interdental space immediately anterior to the planned osteotomy. If there is a major concern about blood supply, it is possible to preserve both buccal and palatal pedicles, but direct visualization of the underlying bone will be compromised. Vertical incisions are made on the buccal, with the inferior extent in the interdental space immediately anterior to the proposed osteotomy and the superior extent curving anteriorly as it approaches the height of the vestibule. A vertical, midline incision over the nasal septum is also made and then subperiosteal tunnels are made connecting the incisions. If the anterior maxillary osteotomy is combined with a Le Fort I osteotomy, these approaches are unnecessary, as the segmental osteotomies can be done from above.

All incisions are closed without drains with running resorbable sutures after an initial, interrupted midline suture.

Extraoral open approach to the maxilla

This approach is used when the lesion being excised not only involves the maxillary alveolus but also the antrum. If the lesion is benign, the resection may be conservative. If the lesion is malignant, the resection must take into account soft- and hard-tissue margins. An effort is made to preserve the bony floor of the orbit. If this is impossible, then preservation of Lockwood's suspensory ligament is essential to maintain the position of the globe. If there is concomitant involvement of the globe and exenteration or enucleation is indicated, an ophthalmologic surgeon should assist. Preoperative evaluation by both computed tomography and magnetic resonance imaging helps to determine the degree of resection required.

The Weber–Fergusson incision is used to access the maxilla (Fig. 3). This incision begins at the medial canthus of the eye and is brought inferiorly along the lateral aspect of the nares to the alar base. It continues under the nostril to the midline of the philtrum of the lip. The upper lip is divided along the midline. This incision may or may not include a subciliary lateral release. The incision is carried down to bone with the scalpel. The infraorbital nerve is preserved unless it is affected by the pathology.

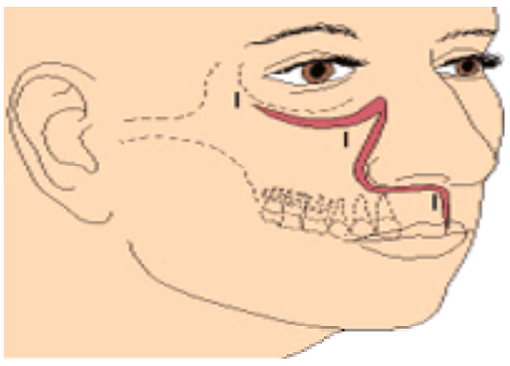


Fig. 3. Incision (I) for the extraoral approach to the maxilla (unilateral).

The underlying maxilla is exposed by subperiosteal dissection. If there is periosteal or soft-tissue involvement, then the required amount of overlying tissue is left in continuity with the specimen. Using a drill, saw or osteotomes, bone cuts are made in the maxilla where necessary. Bleeding occurs predominately from the greater palatine artery in the posterior maxilla, and is controlled with a clip.

Closure of the extraoral tissue is accomplished in layers. Drains are unnecessary. When overlying skin and subcutaneous tissues have been removed as part of the resection, a pedicled rotational flap or free-tissue transfer may be necessary for closure.

The intraoral defect is allowed to remain open. It may be covered with a split-thickness skin graft from the lateral thigh or buttock. This technique requires a stent to facilitate the close contact between the skin graft and the underlying vascular bed needed for proper healing. A prefabricated obturator is very helpful during the initial healing phase. It can be customized in the operating room from thermoplastic materials and serves as the stent for the split-thickness skin graft as well as assisting the patient in swallowing and speech during the postoperative period. These obturators are supported on any remaining teeth, or on the maxillary alveolar ridge.

Orthognathic surgery

The oral and maxillofacial surgeon must first determine whether a skeletal deformity of the maxilla, mandible, or both is present. This requires physical examination of the patient, and measurements from cephalometric and panoramic radiographs and dental casts. The lateral and frontal planes are both considered, with asymmetries noted. Orthodontic analysis and treatment are required before surgery to ensure creation of a stable dental occlusion.

Maxillary osteotomy (Le Fort I) (Fig. 4)

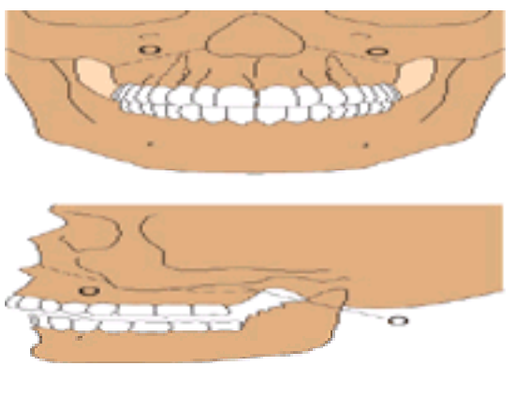


Fig. 4. The osteotomy (O) in frontal and lateral views for the Le Fort I procedure. The maxilla can be mobilized from the remainder of the skeletal midface, while remaining pedicled by the greater palatine arteries and palatal soft tissues. The maxilla can be moved superiorly (with bone removal in the osteotomy), inferiorly with free bone grafts in the osteotomy, anteriorly or posteriorly.

The operation can be used to close an anterior open bite, or to correct a maxillary cant or an anterior–posterior discrepancy. The maxilla is exposed through a vestibular incision from the pterygomaxillary fissure on one side to that on the opposite side. The nasal mucosa is dissected from the lateral and inferior walls of the nasal cavity and protected by insertion of a malleable retractor. A sagittal saw is used with irrigation to create an osteotomy from the maxillary buttress to the pyriform aperture on both sides, staying 4 to 5 mm above the apices of the teeth. Osteotomes with guarded blades are used to separate the maxilla from the medial antral wall. The vomer is separated from the nasal floor with a notched osteotome. The pterygomaxillary junction is separated with a curved osteotome to create a ‘greenstick' type of fracture. Care is taken not to drive the osteotome deeply, since bleeding from the pterygoid plexus may be hard to control. The maxillary downward fracture is completed by pressure on the anterior maxilla inferiorly with a bone hook and on the lateral walls with laminar spreaders. As the maxilla moves inferiorly, it remains attached to its blood supply through the soft palate and the greater palatine vessels, which are seen in the posterior aspect of the surgical site, coursing from the pterygoid plates to the maxilla. If intact, these vessels should be preserved; if bleeding occurs, they can be clipped.

The maxilla is mobilized with Rowe disimpaction forceps and bony interferences are removed with a rongeur or burr to allow repositioning of the maxilla superiorly, anteriorly, posteriorly and/or laterally, or to correct a cant. Interpositional bone grafts may be required to position the maxilla inferiorly. These can be obtained from the patient's iliac crest, rib or the outer table of the cranium, or from the bone bank. The maxilla is positioned with a prefabricated splint and is stabilized with transosseous wires or bone plates. The usual points of fixation are the zygomatic buttresses posteriorly and the lateral wall of the maxilla just lateral to the pyriform aperture anteriorly.

The maxilla can be cut into segments from above with a reciprocating saw or a drill under continuous irrigation with saline. Care is taken not to cut through the palatal mucosa. Multiple segments place the maxilla more at risk for vascular compromise. An occlusal splint is needed to stabilize the segments.

Maxillary osteotomy (anterior segmental)

Using the incisions described previously, it is possible to perform an isolated anterior segmental osteotomy, moving the anterior six or eight teeth. Orthodontic preparation of the teeth is critical so that there is sufficient room between their roots to perform the osteotomy without damaging them. The procedure is the same as for a Le Fort I except that vertical cuts are made in the alveolus posterior to the segment being moved and then a transpalatal cut is made with a saw or burr to allow mobilization of the anterior segment. Care must be taken not to tear the soft-tissue pedicle and thus compromise the vascular supply to the segment. The bony cuts and mucoperiosteal incisions should not coincide, so that there is intact soft tissue over the osteotomy sites. The segment is fixed with an occlusal splint or preformed archwire that can be ligated into the orthodontic brackets. Bony fixation may not be necessary.

Bilateral intraoral oblique osteotomy of mandible

This is used to correct a prognathic mandible (Fig. 5). Access is gained by the intraoral approach to the mandible previously described. The incision is confined to 5 cm over both external oblique ridges.

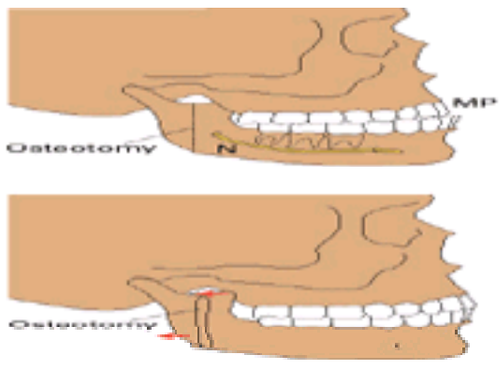


Fig. 5. Oblique osteotomy of the mandible (typically performed bilaterally) posterior to the entry of the sensory inferior alveolar nerve into the mandible, accompanied by vessels (N). In the lower illustration the tooth-bearing portion of the mandible has been set back to correct the mandibular prognathism (MP) shown in the upper illustration.

A subperiosteal dissection exposes the lateral aspect of the mandibular ramus, including the sigmoid notch, condylar neck, posterior border, and angle. Fiberoptic retractors are placed into the sigmoid notch (Bauer type) and posterior border (Le Vasseur Merrill type) to protect the adjacent tissues and to provide illumination. The lingula, which marks the approximate location of the mandibular foramen on the lateral aspect of the mandible, is located. Using a 135° oscillating saw under copious saline irrigation, a cut is made from the sigmoid notch to the mandibular angle, making sure to remain posterior to the lingula. The proximal segment is elevated laterally and stripped of its medial pterygoid muscle sling with a periosteal elevator. This movement allows the proximal segment to rest lateral to the ramus. After the bilateral osteotomies, the mandible is moved posteriorly to its planned position.

Fixation is accomplished by closing the patient's teeth into occlusion and connecting surgical orthodontic hooks with elastics or 24G wire. Intermaxillary fixation is maintained for 6 weeks after surgery to allow osseous healing.

The significant, though rare, complication of this osteotomy is damage to the inferior alveolar nerve, resulting in altered sensation to the ipsilateral lip, chin, and teeth. This may be avoided by careful examination of the planned osteotomy site for the previously described lingula and remaining posterior to this landmark while doing the osteotomy. The risk is not entirely eliminated because there is variation in normal anatomy.

Bilateral sagittal osteotomy of the mandible

This operation is used for advancement of the mandible to correct retrognathism (Fig. 6), and can be used to correct prognathism and asymmetries. The intraoral incision extends for approximately 5 cm over the external oblique ridges of the mandible.

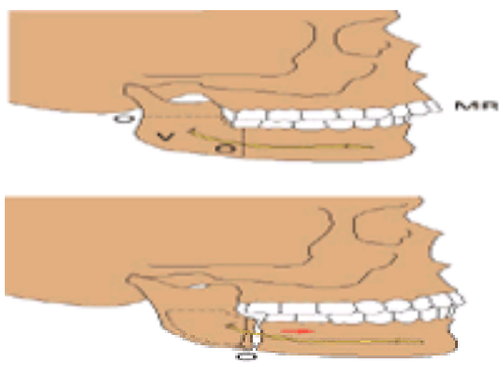


Fig. 6. Sagittal osteotomy of the mandible (O) for advancement of the tooth-bearing portion to correct mandibular retrognathism (MR). Care is taken to split the mandible slowly so that the inferior alveolar neurovascular bundle (V) can be identified and protected.

A subperiosteal dissection is made on the medial surface of the mandible from the base of the coronoid process to the posterior border. The subperiosteal pocket formed is kept superior to the entry of the inferior alveolar neurovascular bundle, which is protected by retracting it inferior to the level of the bone cut with a Cushing periosteal elevator. A Lindemann burr in a straight handpiece or a reciprocating saw, either with continuous saline irrigation, is used to cut through the medial mandibular cortex only, in a line from posterior border to the base of the coronoid.

Subperiosteal dissection is undertaken next on the lateral surface of the mandible from the base of the coronoid process to the inferior border, and anteriorly to the first molar. The Lindemann burr or reciprocating saw is again used, with continuous irrigation, to cut through the lateral mandibular cortex from the inferior to superior border at the posterior edge of the second molar. The facial vessels are protected by inferior retraction with a channel retractor hooked under the inferior border. Care is taken to avoid entering the inferior alveolar canal. The lateral and medial cuts are then connected over the external oblique ridge of the mandible using a straight fissure burr under continuous saline irrigation.

Osteotomes are used to develop the split, which may be completed with an osteotome or bone spreaders. The inferior alveolar neurovascular bundle is identified in the osteotomy site, and is dissected so that it remains with the tooth-bearing segment of the mandible.

Completion of the osteotomy through both mandibular rami allows advancement of the tooth-bearing portion of the mandible to the planned position, while maintaining good bony contact for healing.

After advancement of the mandible, fixation can be achieved with a superior-border wire or bicortical position screws. Intermaxillary fixation is used for 6 weeks postoperatively for the wire fixation; position-screw fixation requires limited intermaxillary fixation.

Genioplasty

Dentofacial deformities often include a retrusive, prominent, or asymmetric chin. Cephalometric radiographs that take into account the planned mandibular and/or maxillary movement are required for proper correction

Genioplasty is best achieved with horizontal osteotomies for both augmentations and reductions. Use of alloplastic implants often results in migration of the implant with accumulation of fibrous scar tissue.

The approach to the symphyseal region is intraoral. The incision is made along the mucobuccal fold between the position of the mental foramina as judged from the panoramic radiograph. Leaving 10 to 15 mm of mucosa along the gingival aspect facilitates closure. Dissection continues though mentalis muscle to underlying periosteum and bone. The mental nerves are isolated bilaterally, freed of their periosteal attachments, and protected with retractors.

Horizontal chin augmentation is accomplished by making an osteotomy with a reciprocating saw beneath the teeth apices and extending this cut laterally. Vertical reduction is accomplished by making two horizontal osteotomies and removing a wedge of bone. The segment is advanced anteriorly and secured with either lag screws or a miniplate. If a smaller or asymmetric reduction is needed, using a burr in a rotary instrument may achieve the required results.

Augmentation can be achieved by using free bone grafts to the bony chin, but the ease of working with the native chin makes this unnecessary for all but extremely large augmentations.

Closure is in layers, insuring that the mentalis is reattached to its bony origin to prevent ptosis of the chin and lip. The mucosa is closed with a running absorbable suture.

The mental nerve is rarely severed in this procedure. If repair is required, 9/0 nylon sutures are placed under magnification along the epineurium. If the nerve is severed at the foramen, enlarging the mental foramen with a rotary instrument and saline irrigation will help access the proximal aspect of the nerve.

Preprosthetic surgery

Preprosthetic surgery (Fig. 7) is the preparation of the intraoral hard and soft tissues for dental prostheses. Patients may have insufficient height of the alveolar ridge or underlying bony protuberances that preclude a comfortable and functional prosthesis. Redundant soft tissue may also prevent prosthetic rehabilitation. Insufficient height of the alveolar ridge is most commonly treated with titanium dental implants. Bone-graft augmentation in conjunction with dental implants may be used if there is insufficient residual bone for implants.

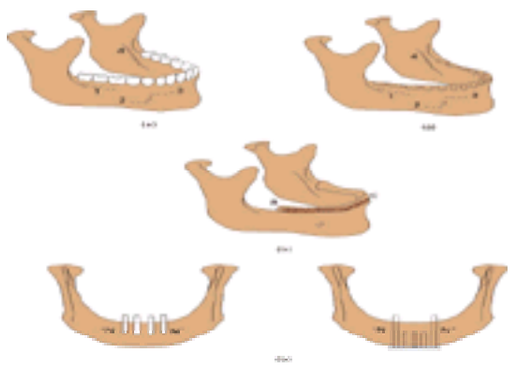


Fig. 7. (a) A mandible with a full complement of teeth. If the teeth are removed for disease, the alveolar bone around the roots gradually resorbs, leaving an atrophic mandible. The numbers refer to muscle attachments (1, buccinator; 2, depressor anguli oris; 3, depressor labii inferioris; 4, mylohyoid). As the ridge resorbs, these attachments become relatively closer to the crest of the ridge. This, combined with ridge atrophy, tends to unseat removable prostheses. (b) These three figures show methods to improve stabilization of a removable prosthesis on the atrophic mandibular ridge. First is a free autogenous rib graft (R) shaped to the lingual surface of the mandible with cancellous chips from autogenous iliac crest (C) packed around it. Second are implants into the alveolar bone medial to the mental foramen and nerves (N). Third is the ‘staple bone plate', also placed so that its transosseous pins are medial to the mental foramina and nerves (N).

Placement of dental implants requires careful preoperative planning. Location and angulation are determined by the amount of existing bone and the prosthetic needs as shown by study models. An intraoperative stent that predetermines location and position is essential. A full-thickness mucoperiosteal flap exposes the underlying bone. The dental implants are placed under copious saline irrigation and the flaps are closed with interrupted, nonabsorbable sutures. The implants require 3 to 6 months to allow for osseointegration. When osseointegration is complete, the area is re-exposed and abutments are placed for the final prosthetic rehabilitation. Coordination with a prosthodontist who can construct dental prostheses on the implants is essential.

An alternative to conventional dental implants is the ‘staple' implant, which is placed through the extraoral approach to the inferior border of the mandibular symphysis. This implant has two to four transosseous pins that enter the mouth through the alveolar ridge, driven through holes drilled from below. It is secured with screws that attach to the inferior border of the mandible. The dental prosthesis is attached to the transosseous pins within the mouth. The advantage to the ‘staple' implant is that the alveolar height needed for placement is decreased; however, it does require an extraoral approach.

Implants are contraindicated in patients where the quality and quantity of bone is insufficient. Significant systemic disease may also preclude placement of dental implants.

Surgery of the temporomandibular joint

Surgical intervention should be considered in patients with the following: chronic dislocation of the mandibular joint(s) in which no psychological or neuromuscular factors are contributing; disabling ankylosis; internal derangement, when noninvasive techniques have failed to increase range of motion and relieve pain; degenerative arthritis, when noninvasive treatment has failed to resolve the symptoms; loss or absence of the mandibular condyle because of congenital malformation, trauma, or surgical ablation; fractures of the mandibular condyle or condylar neck with dislocation or significant displacement; infections of the joint where drainage is necessary; and tumours of the joint.

Surgical approaches

The main approach to the temporomandibular joint (Fig. 8) is by way of a 2.5-cm vertical incision in a skin fold immediately anterior to the pinna and extending into the temporal region. The dissection is carried down sharply through the subcutaneous layer. The superficial temporal artery and vein are encountered, and are retracted posteriorly or tied and divided. The dissection is carried down to the temporal fascia superior to the zygomatic arch. Care is then taken to avoid the frontalis branch of the VIIth cranial nerve as the incision is deepened along its entire length just anterior to the cartilage of the external auditory meatus. The skin and soft-tissue flap are dissected forward along the zygomatic arch to the articular eminence. The joint capsule can then be seen.

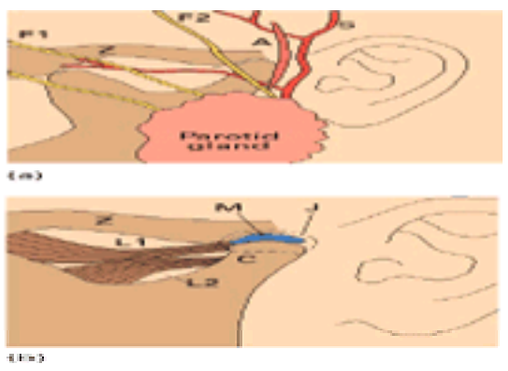


Fig. 8. (a) Incision (A) for preauricular approach to the temporomandibular joint. Also shown are the zygomatic arch (Z), zygomatic (F1) and frontal (F2) branches of the facial nerve, and the superficial temporal artery (S). (b) A schematic diagram of the temporomandibular joint showing the condyle (C), zygomatic arch (Z) posteriorly forming the glenoid fossa, the superior (L1) and inferior (L2) heads of the lateral pterygoid muscle, the meniscus (M), and joint capsule (J). The superior joint space is superior and the inferior joint space inferior to the meniscus.

The zygomatic arch and lateral pole of the condyle can be approached by two incisions, one along the lateral surface of the zygomatic arch, down to bone, and the second vertically through the lateral capsule on the lateral pole of the condyle. These two incisions allow access to the superior and inferior joint spaces, the meniscus, and the condyle. The incision along the zygomatic arch is useful for access to the articular eminence of the zygomatic arch when reduction of the eminence is needed. Endaural incisions have been described, and the hemicorneal incision can be used for better access in surgery for ankylosis.

Internal derangement of the joint

Either as a result of gross trauma or microtrauma from bruxism the meniscus can become displaced, causing pain and limitation of mandibular motion. Conservative

measures, including dental splints, nonsteroidal, anti-inflammatory drugs, muscle relaxants, physical therapy and behavioral modification, are usually successful in reducing the patients discomfort and improving their mandibular range of motion. If the patient's symptoms persist, arthroscopy or arthrocentesis will usually be helpful. If the derangement is associated with significant pain, limitation, or arthritic change that have not responded to less invasive therapy, an open surgical technique may be warranted.

Joint reconstruction

The temporomandibular joint may be so diseased that it is unable to function: this may be due to degenerative change, trauma, infection, tumor, previous surgery, or radiotherapy. Joint replacement may be necessary and autogenous materials are preferred. The costochondral graft is widely used to replace the mandibular condyle (Fig. 9). The rib is harvested from the side of the body opposite to the temporomandibular joint being reconstructed as its natural curvature approximates that of the portion of the mandible being replaced.

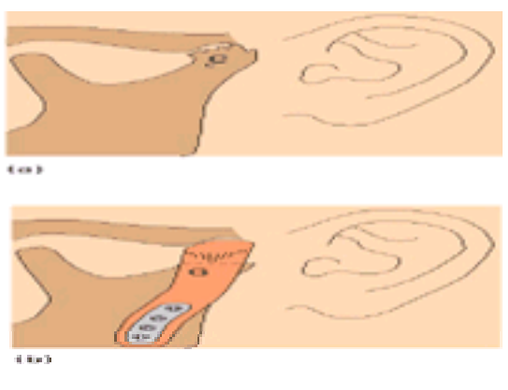


Fig. 9. (a) Degenerative disease of the temporomandibular joint affecting the condyle (C); fibrous or bony ankylosis can accompany this. (b) The condyle is excised and replaced with autogenous costochondral graft (G). The costal cartilage is placed in the glenoid fossa. The rib is secured to the ramus of the mandible with a miniplate.

The surgical approach to the temporomandibular joint is by way of preauricular and submandibular incisions. The combination of these approaches provides access to the joint and convenient access to the lateral surface of the mandible, where internal fixation devices are placed to hold the graft in place.

When the joint meniscus is damaged or absent, it is replaced with a temporalis fascia or muscle flap, either pedicled on the coronoid process and rotated under the articular eminence or based inferiorly and the vertical flap brought over the lateral aspect of the glenoid fossa. The free end of the flap is secured to the soft tissues anterior to the external auditory meatus or deep to the joint. Care is taken to place the flap so that it covers the medial surface of the natural or reconstructed condyle. If this is not achieved, the medial surface can ankylose to the cranial bone.

Bony reconstruction of the temporomandibular joint requires intermaxillary fixation during surgery to maintain the proper maxillomandibular relation. This fixation is maintained for a variable period of time after surgery depending on the method of internal fixation. Wire internal fixation requires 4 to 6 weeks of intermaxillary fixation; rigid internal fixation with screws or plates will allow release of intermaxillary fixation as early as 1 week after surgery. Physical therapy may be started soon thereafter.

Reconstruction of mandibular discontinuities

Mandibular discontinuity defects occur from resections for the treatment of tumors and from osteomyelitis unresponsive to debridement and antibiotics. Malignant tumors require sufficient margins to ensure eradication of disease. Certain nonmalignant lesions (such as ameloblastoma) are locally aggressive; their treatment often leaves large defects in hard and soft tissues. Reconstruction should restore preoperative maxillomandibular relations, occlusion, masticatory function, allow dental prosthetic rehabilitation, allow the patient to manage their salivary secretions adequately, and maximize speech. Whenever possible, mandibular reconstruction (Fig. 10) is best achieved in the setting of primary reconstruction, at the time of initial ablation.

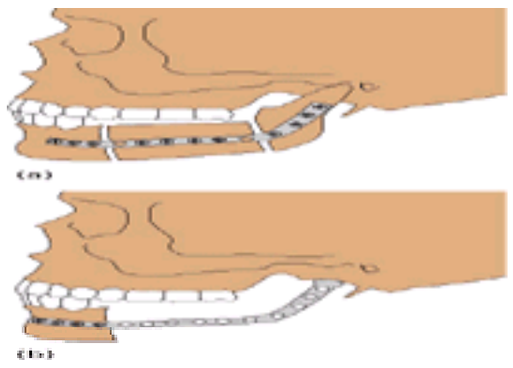


Fig. 10. (a) Sections of resected mandible can be replaced with autogenous iliac bone graft, secured by a reconstruction plate (upper diagram). If the condyle is also removed, a plate alone with integral condyle can be used (b).

The first stage in reconstruction involves the maintenance of the original occlusal relation. If the patient is dentate, intermaxillary fixation should be applied. The approach to the mandible is dictated by the extent of disease. The surgical margins need to be defined and, before the ablation of hard tissue, a rigid reconstruction plate should be adapted passively to the unresected mandible. The plate is secured to the bone with the appropriate screws. The reconstruction plate and screws are then removed, carefully labeled, and placed aside. The resection now proceeds as needed. The reconstruction plate is then replaced and secured to the native mandible both proximally and distally. This now re-establishes the exact interarch relation that existed before the resection.

The individual techniques for reconstruction are numerous. They include reconstruction plates without osseous reconstruction, nonvascularized free bone grafts, pedicled osteomyocutaneous flaps, and osteomyocutaneous free flaps. Examples of nonvascularized free bone grafts include autogenous rib or iliac crest. Cadaveric mandible, if available, provides an excellent anatomic substitute. Examples of pedicled osteomyocutaneous flaps include pectoralis major and rib, pectoralis major and split-thickness sternum, and trapezius and scapula. Examples of osteomyocutaneous free flaps include revascularized fibula, ilium, and scapula. The bone grafts most often used are the nonvascularized free grafts and, more recently, the osteomyocutaneous free flaps. The advantage of free tissue transfer is the ability to transfer adequate viable bone as well as soft tissue to the area in a single procedure.

The desired bone is inset and attached to both the proximal and distal aspects of the native mandible as well as the reconstruction plate. If a curve is desired, as would be needed along the mandibular symphysis, sequential osteotomies are performed along the length of harvested bone. If a free flap is used, the vessels are anastomosed, and the soft-tissue component is placed in the areas of deficit and contoured as necessary. It is imperative to not compromise the osseous vascular supply. Doppler monitoring of the flap is essential during the perioperative period. The intraoral closure must be meticulous to preclude salivary contamination. The extraoral closure is in layers using suction drains through a separate puncture site.

Intermaxillary fixation is not necessary during healing. If the patient is dentate, guiding elastics attached to appliances bonded to the teeth are helpful in correcting the minor occlusal discrepancies that sometimes result.

Treatment of mandibular fractures

Accurate diagnosis must precede treatment; in many instances, plain radiographs are usually sufficient. Views from at least two directions are necessary to assess

alignment and displacement. The principal mandibular views are lateral, posterior–anterior, Towne's, panoramic, and occasionally occlusal. The occlusal view is useful for fractures in the symphyseal region. If condylar fractures are present, a computed tomographic study may be necessary to determine whether the condyle is in the glenoid fossa or whether the fracture is comminuted or intracapsular.

Teeth in the line of fracture are usually not removed unless they are diseased, extremely mobile, or have a root fracture.

Fractures of the mandible are treated (Fig. 11) in a closed fashion if they are not displaced and if there is a functional occlusion with good intercuspation. For undisplaced condylar fractures, if the patient is able to maintain the premorbid occlusion, a liquid diet and decreased use of the jaw, with careful observation, are usually sufficient. For fractures in other areas of the mandible or for condylar fractures that result in a malocclusion, intermaxillary fixation is used. Arch bars are attached to the teeth in both the maxilla and the mandible with stainless-steel, circumdental 26G wire. Elastics or 26G wires are applied to the lugs of the arch bars to hold the upper and lower teeth together, thus splinting and immobilizing the mandible. The patient is checked weekly to observe the occlusion and adjust the intermaxillary fixation as needed. The fixation is usually removed at 10 to 14 days for condylar fractures, and at 4 weeks in adults and 2 to 3 weeks in children for other fractures. After releasing the intermaxillary fixation, guiding elastics may be needed to maintain a stable occlusion. One or two elastics are placed on each side between the arch bars to guide the teeth into the correct bite.

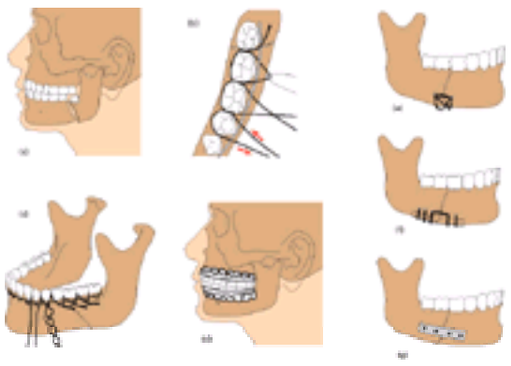


Fig. 11. Mandible fractures (a) are most often first treated by intermaxillary fixation. Circumdental wiring (b) attaches arch bars (c) with lugs to allow elastic intermaxillary fixation (d). If internal fixation is denied, intraosseous wires (e), a K-wire attached with wires to the lower border in a prepared osseous groove (f), or a bone plate (g) can be used.

Open reduction is used when there is a major displacement or mobility at the fracture site. It is also used if a functional occlusion is absent or if the fracture is displaced and in the area of the wisdom teeth (third molars) or angle of the mandible. If there are teeth present that can help stabilize the fracture and establish an occlusion, arch bars are placed as previously described. The teeth just adjacent to the fracture should be loosely ligated so that the fracture can be manipulated into an anatomic position. The open approach to the mandible, described earlier, by either the extraoral or intraoral route is used to gain access to the fracture. If possible, intermaxillary fixation is applied; internal fixation of the bone using either wires or plates is then accomplished during open reduction so that stabilization can be assured. Once the fracture is stabilized, the occlusion is checked to determine if it is correct. With most plating systems, intermaxillary fixation is not necessary, although guiding elastics may be used for several weeks. If wire fixation or nonrigid plates are used, then intermaxillary fixation is maintained for 4 weeks.

Open reduction of condylar fractures is indicated if the fractured condyle is dislocated, mandibular motion is limited by the displaced condyle, malocclusion remains 1 or 2 weeks after an attempt at closed reduction, or if there are bilateral condylar fractures and it is necessary to re-establish the vertical dimension of the mandible. In the last instance, only one condyle needs to be opened. The approach is the same as the open approach to the temporomandibular joint described earlier. As with temporomandibular joint reconstruction, an open approach to the inferior border of the mandible, with tunneling in the subperiosteal plane to the fractured condylar neck, is sometimes needed to allow the convenient placement of wire, screws, or miniplates to stabilize the fracture. An alternative approach would be a hemicoronal incision.

Uvulopalatopharyngoplasty

Uvulopalatopharyngoplasty is one of several surgical options available for the treatment of sleep apnea syndrome. This procedure was the first surgical approach to target specifically an anatomic deformity implicated in the etiology of this condition. Obstructive sleep apnea is characterized by repeated episodes of apnea and hypopnea during the sleeping state. Clinical symptoms include excessive snoring with cessation of breathing. These apneic episodes may last from 10 s up to 3 min, but generally they last approximately 30 s. Patients often complain of excessive daytime somnolence, difficulty concentrating, headaches, and impotence. Systemic ramifications include systemic hypertension, pulmonary hypertension, cor pulmonale, and cardiac arrhythmias.

Sleep apnea has several variants; these include central, obstructive, and mixed. Uvulopalatopharyngoplasty is effective in the obstructive form and to a lesser extent in the mixed form. Treatment depends on the variant of sleep apnea present, and if obstructive in nature, the level of obstruction.

Preoperative evaluation should consist of a thorough history and physical examination, fiberoptic nasopharyngoscopy, cephalometric analysis, and polysomnography. Physical examination should identify anatomic factors, which may contribute to airway impedance, such as adenoid or tonsillar hypertrophy, excessive soft palate, and redundant pharyngeal mucosa. Fiberoptic nasopharyngoscopy identifies the position and location of the soft palate, and posterior and lateral pharyngeal walls. The degree of collapse of the pharyngeal wall upon inspiration against a closed mouth and nares simulates the negative pressures encountered during obstructive episodes (Muller maneuver). Cephalometric analysis allows identification of skeletal and soft-tissue deformities that may contribute to obstruction of the airway.

Polysomnography is important to differentiate central and obstructive causes of sleep apnea. This takes place in a sleep laboratory where the patient's sleep patterns are evaluated by electroencephalogram, electromyelogram, electro-oculogram, and electrocardiogram. Oxygen desaturation are carefully monitored as well as respiratory effort and chest-wall movement.

Most patients that present with symptoms of obstructive sleep apnea are obese males over 40 years of age. Initial treatment should encourage weight loss and any sedating medications, including alcohol, should be stopped. A trial of nasal continuous positive airway pressure is helpful to determine the benefits of decreasing airway resistance. When the area of obstruction is diagnosed to be isolated to the palatal and tonsillar regions, a uvulopalatopharyngoplasty is undertaken.

The surgical technique first described has been modified slightly to include full resection of the uvula. Tonsillectomy is done first; if already done, the mucosa of the existing tonsillar fossa is excised and mobilized. An incision is made with needle-tip electrocautery in the soft palate approximately 10 mm anterior to the uvula. This incision is carried laterally to connect to the anterior pillars. An incision is then made along the posterior soft palate and carried laterally to connect to the posterior pillars. A small amount of palatal muscle is excised. The nasopharyngeal mucosa along the posterior aspect of the soft palate is released and advanced. This allows for primary closure. The incision is closed with interrupted 3/0 chromic sutures. To eliminate redundant tissue along the lateral and posterior pharyngeal walls, the tonsillar fauces are closed with interrupted 3/0 chromic sutures. The inferior aspects of the fauces remain open to prevent hematoma formation. It is imperative that meticulous hemostasis be achieved.

Complications of uvulopalatopharyngoplasty include airway obstruction, early and late bleeding, and infection. Velopharyngeal incompetence may occur as well as palatal stenosis. Postoperative pain is significant and narcotics must be used judiciously in patients with a history of sleep apnea. Success of this procedure is directly related to the preoperative diagnosis. If there is more than one site of obstruction, concomitant procedures must be undertaken to ensure a successful outcome.

Further reading

Bailey BJ, ed. *Surgery of the oral cavity*. Year Book Medical, St Louis, 1989. [General oral surgical text.]

Bell WH. *Modern practice in orthognathic and reconstructive surgery*, Vols 1, 2, 3. Saunders, Philadelphia, 1992. [Classic text for orthognathic surgery.]

Block MS, Kent JN. *Endosseous implants for maxillofacial reconstruction*. Saunders, Philadelphia, 1995.

Branemark P. Osseointegration and its experimental background. *Journal of Prosthetic Dentistry* 1983; **50**: 399. [Review of several investigations, which led to the theory of osseointegration.]

Donoff RB. *Massachusetts General Hospital manual of oral and maxillofacial surgery*. Mosby, St Louis, 1996. [Handy reference manual for oral and maxillofacial surgery.]

Enlow DH, ed. *Facial growth*, 3rd edn. Saunders, Philadelphia, 1990. [Classic craniofacial growth and development text.]

Epher BN, Fish LC. *Dentofacial deformities: integrated orthodontic and surgical correction*. Mosby, St Louis, 1986. [Orthodontic and surgical correction of dentofacial deformities.]

Fonseca RJ, Davis WH. *Reconstructive preprosthetic oral and maxillofacial surgery*. Saunders St Louis, 1986. [Classic text on reconstructive surgery for oral and maxillofacial region.]

Fonseca RJ, Walker RV. *Oral and maxillofacial trauma*, Vols 1, 2. Saunders, Philadelphia, 1991. [Surgical text on trauma to oral and maxillofacial region.]

Keith DA, ed. *Surgery of the temporomandibular joint*. Blackwell Scientific, Oxford, 1992. [Background to temporomandibular Joint surgery, history and complications.]

Keith DA, ed. *Atlas of oral and maxillofacial surgery*. Saunders, Philadelphia, 1992. [Step-by-step atlas of oral and maxillofacial surgical procedures.]

Kuriloff D, Sullivan M. Mandibular reconstruction using vascularized bone grafts. *Otolaryngologic Clinics of North America* 1991; **24**(6). [Overview of vascularized bone and free tissue transfer techniques with emphasis on reconstructive goals.]

Proffitt WR, White RP. *Surgical-orthodontic treatment*. Mosby, St Louis, 1991. [Text of combined surgical/orthodontic treatment of orofacial deformities.]

Rowe NL, Williams JLI. *Maxillofacial injuries*, Vols 1, 2. Churchill Livingstone, New York, 1985. [Classic maxillofacial trauma surgery text.]

Shelton DW, Irby WB, ed. *Current advances in oral and maxillofacial surgery: orthognathic surgery*. Mosby, St Louis, 1986. [Orthognathic surgery text.]

Small IA, Misek D. A 16-year evaluation of the mandibular staple bone plate. *Journal of Oral and Maxillofacial Surgery* 1986; **44**: 60–6. [Retrospective study of 1516 patients with mandibular staple implants from 1968–1984.]

Speise B, ed. *New concepts in maxillofacial bone surgery*. Springer, Berlin, 1976. [Bone grafting in the maxillofacial region.]

Waite PO, ed. Oral and maxillofacial treatment of obstructive sleep apnea. *Oral and Maxillofacial Surgery Clinics of North America* 1995; **7**(2). [In-depth description of work-up and management of the patient with sleep apnea.]

45.11 Facial injuries

R. J. Gates and S. A. Wall

Introduction
Initial management
Airway
Circulation
Associated injuries
Soft tissue injuries
Principles of management of major facial fractures
Timing of surgery
Surgical access
Mandibular fractures
Diagnosis
Investigation
Treatment
Nasal fractures
Zygomatic fractures
Diagnosis
Investigation
Treatment
Maxillary fractures
Diagnosis
Investigation
Treatment
Complex/panfacial injuries
Orbital fracture
Clinical
Investigation
Treatment
Complications
Nasoethmoido-orbital (NOE) fractures
Fractures of the frontal sinus
Further reading

Introduction

Facial injuries most commonly result from road traffic accident or assault. The facial skeleton consists of a series of strong horizontal and vertical bony buttresses, separating cavities and thin-walled sinuses, thus the face acts as a crumple zone, absorbing the energy of an impact and reducing injury to the brain.

Initial management

Airway

The initial management of any patient with major facial injuries should follow the established ATLS (advanced trauma life support) system guidelines. Anticipation of airway compromise and early establishment of an airway are the keys to management of these injuries. Airway obstruction may be due to foreign bodies, oedema, loss of tongue support, or haematoma. Blood clots, vomitus, and foreign bodies such as fractured teeth or dentures must be sought and removed from the oropharynx. Progressive airway compromise due to oropharyngeal swelling must be anticipated and treated by early intubation. Continuing bleeding, either into the upper airway or the interstitium of the neck, may also compromise the airway. Bilateral fracture of the mandible may result in loss of anterior support for the tongue, which may prolapse into the upper airway. Manual reduction and support of the fracture may be needed until definitive treatment is given. Endotracheal intubation, usually by the oral route and with a cuffed tube, should be carried out prior to the first suggestion of airway compromise. Awaiting the onset of stridor, dyspnoea, or hypoxia courts disaster. If an airway cannot be established by this method, tracheostomy or cricothyroidotomy may be required.

Circulation

Haemorrhage from facial and scalp injuries can be significant. The standard treatment techniques of direct pressure, elevation of the head (once cervical spine injury has been excluded), and early, haemostatic suturing of lacerations or ligation of superficial vessels should be performed. Persistent haemorrhage may require anterior and posterior nasal packing, reduction and fixation of fractures, ligation of the external carotid arteries, and/or selective embolization of bleeding vessels for control.

Associated injuries

Up to 10 per cent of patients with major facial fractures have cervical spine injuries and all should be appropriately assessed for such injuries. Between 5 and 17 per cent will have associated eye injuries and an ophthalmological examination is essential for those with periorbital injuries. Intracranial injury also occurs frequently in association with facial injuries.

Soft tissue injuries

The facial soft tissues have a rich blood supply and their debridement should generally be conservative. Important structures at risk in lacerations of the face are the facial nerve, parotid duct, and lacrimal duct system. Facial nerve branches divided anterior to the level of the lateral canthus do not require repair, as the affected muscles tend to reinnervate spontaneously. Posterior to this level, the nerve branches should be repaired microsurgically.

The parotid duct lies along the middle third of a line drawn between the tragus of the ear and the midportion of the upper lip. Failure to repair the damaged duct may result in an external salivary fistula or chronic parotitis from duct stenosis. The intraoral end can be identified by probing the duct orifice, opposite the upper second molar tooth. The proximal end is identified by expressing saliva from it. Lacerations should be repaired over a stent, which is left *in situ* for 2 weeks. Untreated injury to the lacrimal apparatus, near the medial canthus, may result in epiphora or dacryocystitis. Any laceration to the canaliculi, sac, or duct should be probed and repaired over a Silastic stent, which is left *in situ* for 6 weeks.

Particular anatomical borders of the face require accurate apposition to avoid a visible step and produce optimal cosmetic results. These include the eyelid, nostril and helical margins, the eyebrow, and the vermilion border of the lip.

Principles of management of major facial fractures

The developments of CT scanning, rigid plate and screw fixation, and the application of craniofacial surgical techniques have advanced the treatment of facial fractures over the past two decades. The principles of management of major facial fractures are:

i. accurate and comprehensive clinical and radiological diagnosis of all fractures followed by early, one-stage repair;
ii. this requires direct exposure of all fracture fragments, precise anatomical reduction, rigid internal fixation, and primary bone grafting, where appropriate;
iii. the objective is to re-establish facial height, width, and projection, by reconstructing the horizontal and vertical facial buttresses. This is achieved by working from areas of stability to areas of instability, progressing from inferiorly (mandible) and laterally (zygoma) to superiorly and medially (nasoethmoido-orbital region); and
iv. definitive soft tissue management should be performed at the same procedure.

Timing of surgery

Primary definitive treatment of facial fractures should be performed once all appropriate investigations have been performed and the patient's condition allows. The treatment of intracranial, cervical spine, or other injuries may take priority. Treatment may be at the time of presentation, or as a delayed primary repair once swelling has subsided. Secondary repair should be avoided as excessive delay allows the soft tissues to contract and scar over a displaced facial skeleton and makes subsequent reconstruction more difficult.

Surgical access

Virtually the entire craniofacial skeleton can be accessed by a combination of coronal scalp, lower lid, and upper and lower buccal sulcus incisions. Lacerations may afford direct access to fracture sites. Other incisions such as Dingman's lateral brow, Gillies' temporal, and Risdon's retromandibular are used when specific or more limited access is appropriate.

Mandibular fractures

Diagnosis

Specific signs of mandibular fracture include malocclusion, trismus, crepitus or mobility at the fracture site, a step in the dentition or at the inferior border of the mandible, and numbness in the distribution of the inferior alveolar nerve. Fractures may be classified according to site, as dentoalveolar, symphyseal, parasymphyseal, body, angle, ramus, or condylar. Predictive classifications based on fracture orientation also exist. Fractures are multiple in up to 50 per cent of cases and common combinations include angle and contralateral body, parasymphysis and contralateral subcondyle, symphysis and bilateral subcondyles. Mandibular fractures are regarded as compound if intraoral mucosa or skin is breached or the fracture line involves a tooth socket.

Investigation

Radiological investigation includes orthopantomogram and lateral oblique and anterior views of the mandible. CT scanning is usually of little additional benefit in isolated mandibular injuries.

Treatment

The aims of treatment are to achieve anatomical reduction and immobilization of the fracture, in order to restore preinjury functional occlusion and facial appearance. Available treatment options include conservative management, intermaxillary fixation, open reduction and internal fixation, and external fixation.

Conservative treatment is appropriate if the occlusion is normal, the fracture undisplaced and stable, and the patient in minimal discomfort. A soft diet is given and the occlusion observed for 6 weeks.

Intermaxillary fixation involves the application of either eyelet wires or arch bars to the teeth of both jaws (avoiding attachment to the incisors and canines, which are easily avulsed, and to loose teeth adjacent to the fracture for the same reason). The two appliances are then connected with wires or elastic bands to hold the reduction in a normal occlusion. Elastic bands allow some further correction of the initial post-reduction occlusion if required. Intermaxillary fixation is maintained for 3 to 6 weeks. Predictive classifications, based on the site and direction of the fracture line and the presence and number of teeth, allow decisions to be made regarding the likelihood of success of treatment by intermaxillary fixation.

Indications for open reduction and internal fixation include inability to achieve reduction by the closed method or to maintain stable reduction with intermaxillary fixation alone. This is often the case in comminuted fractures, displaced fractures in edentulous mandibles and those with teeth on only one side of the fracture, and in displaced fractures of the angle behind the last molar. However, the advantages to the patient of avoiding intermaxillary fixation, with its attendant eating difficulties and weight loss, possible permanent reduction in range of mouth opening, risks of airway obstruction, and poor oral hygiene have made open reduction and internal fixation an increasingly popular form of definitive treatment. Complication rates with open reduction and internal fixation are of the order of 3 per cent. Wherever possible, intraoral mucosal rather than external skin incisions should be used to access the fracture site.

Plate fixation of fractures of the body of the mandible is achieved with one plate applied to the thickest part of the bone along the lower border and a second fixation point, either a monocortical plate applied along the upper border or an arch bar or interdental wire applied to the teeth, to resist the forces of tension along the upper border. Fractures anterior to the mental foramen (parasymphyseal) require plate fixation along both upper and lower borders to resist torsional force in this region.

Fractures at the angle of the mandible may require open reduction and internal fixation, as it is otherwise difficult to control the position of the posterior fragment. The third molar may need to be removed if it is unerupted and prevents adequate reduction of the fracture. Access for screw placement is facilitated by use of a transcutaneous trocar. Fractures of the ramus are usually undisplaced, due to the splinting effect of the masseter muscle and can usually be managed by intermaxillary fixation alone.

The management of condylar fractures remains controversial. Intracapsular fractures are prone to ankylosis, particularly if immobilized. If the occlusion is normal, conservative management with soft diet is appropriate. If malocclusion exists, closed reduction and application of intermaxillary fixation for 7 days, followed by a period of elastic traction and active mobilization is given. For extracapsular fractures in which closed reduction achieves normal occlusion, intermaxillary fixation should be applied for 2 to 3 weeks followed by a period of elastic traction and jaw mobilization. Open reduction and internal fixation, through a combination of pretragal and retromandibular incisions, is indicated if:

i. the condyle is laterally displaced 30° or more from the axis of the ascending ramus;
ii. bilateral fractures are associated with ascending ramus shortening and an anterior open bite;
iii. the condylar head is dislocated from the glenoid fossa; or
iv. adequate occlusion cannot be achieved with closed reduction.

Other methods of fixation are rarely use nowadays. External fixation may be appropriate for massive bone and soft tissue loss following high velocity gunshot wounds. Splints may be used to aid the application of intermaxillary fixation in the partially edentulous, and these may be augmented by the use of circumoral wires in the edentulous. Teeth in the line of the fracture should be retained, unless they are non-viable or prevent reduction of the fracture. All compound fractures require antibiotic cover.

Nasal fractures

The nasal bones are the most commonly fractured facial bones. Lateral impact creates a deviation of one or both nasal bones and the nasal septum to the opposite side. Frontal impact causes splaying of the nasal bones and buckling or dislocation of the nasal septum with depression of the dorsum. More severe impact may result in fractures extending to involve the piriform aperture, nasofrontal junction, or orbit. Clinical examination will demonstrate deformity and tenderness locally. Septal haematoma should be excluded by intranasal examination and if found, drained, to prevent pressure necrosis and subsequent septal perforation. Radiology adds little to clinical examination in simple fractures. Simple nasal fractures are usually reduced closed, by relocation of the cartilaginous septum with Ash's forceps and repositioning of the nasal bones using Walsham's forceps or the blunt end of a scalpel handle. Anterior nasal packing is inserted and a dorsal nasal splint applied. Rhinoplasty is often required at a later stage to correct residual deformity.

Zygomatic fractures

The zygoma forms the eminence of the cheek and the inferolateral part of the orbit. In addition to isolated arch fractures, the zygoma tends to fracture in a ‘tetrapod' fashion at or near its junctions with adjacent bones, namely at the lateral orbital margin (zygomati-cofrontal suture), lateral orbit (with greater wing of sphenoid), inferior orbital margin, orbital floor and anterior wall of the maxillary sinus (with maxilla), and the zygomatic arch (with temporal bone).

Diagnosis

Clinical signs of zygomatic fracture include periorbital swelling and ecchymosis, ecchymosis of the conjunctiva (which stays bright red due to the continued transfer of oxygen through the conjunctiva to the underlying haemoglobin), and upper buccal sulcus. Unilateral epistaxis may occur due to blood draining from the disrupted maxillary sinus. Malar depression or flattening of the cheek is best observed by looking from above the patient's forehead. A bony step may be palpable at the inferior

orbital margin, in the upper buccal sulcus, along the arch, and occasionally at the zygomaticofrontal suture. Numbness of the anterior cheek, lateral nose, upper lip, and upper dentition is due to trauma to the infraorbital nerve in the floor of the orbit or at its foramen, and of the posterior cheek due to damage to the zygomaticofacial branch as it exits the bone. Diplopia, particularly on upward gaze, is due to either contusion or entrapment of the extraocular musculature (Fig. 1). Significant displacement of the zygoma may cause distortion of the orbital floor and lateral wall, resulting in orbital dystopia or enophthalmos. Displacement of the inferior orbital rim may cause lower lid shortening as the attachment of the orbital septum along the rim pulls the lid down. Displacement of the attached lateral canthal ligament may give an antimongoloid slant to the palpebral fissure. As this fracture, by definition, involves the orbital floor and lateral wall, a detailed eye examination must be performed to exclude ocular pathology. Medial displacement of the arch may impinge on the masseter or temporalis muscles or the coronoid process of the mandible causing trismus.



Fig. 1. Patient with left malar fracture demonstrating subconjunctival haemorrhage, loss of malar prominence, and hypoaesthesia of the cheek.

Investigation

Fractures of the zygoma are best seen on Waters' (oblique posteroanterior) view where opacification of the maxillary sinus, fracture of the inferior orbital rim and floor, and fracture of the anterior wall of the maxillary sinus (zygomaticomaxillary buttress) may be seen (Fig. 2). The Caldwell (anteroposterior 15° flexion) view shows separation of the zygomaticofrontal suture. Submentovertex views will identify isolated arch fractures. CT scanning is indicated if clinical examination suggests intraorbital pathology, including entrapment. Both axial and coronal views are needed (Fig. 3).



Fig. 2. Radiograph demonstrating left zygomatic fracture with partially opacified sinus and bony disjunction at the zygomatico-frontal suture, inferior orbital rim, zygomatico-maxillary suture and zygomatic arch.



Fig. 3. Radiograph demonstrating isolated right zygomatic arch fracture.

Treatment

Stable, undisplaced fractures require only conservative management, soft diet and regular review for 6 weeks to check against later displacement. Isolated arch fractures can usually be reduced via a Gillies temporal incision. If the fracture remains unstable, a bicoronal incision approach and plate fixation is required.

Reduction of tetrapod fractures can be achieved by several approaches, either using a bone hook placed through the cheek skin or upper buccal sulcus, or by passing a heavy periosteal elevator, inserted through a Gillies temporal incision, into the plane deep to the deep temporal fascia. Open reduction and internal fixation is indicated if closed reduction cannot be achieved, or remains unstable. This may be anticipated if there is demonstrable separation of the zygomaticofrontal suture. Open reduction and internal fixation/orbital exploration is also indicated if the orbital contents are displaced or if there is comminution of the orbital floor. The zygomaticofrontal suture is approached through a lateral brow incision, the inferior orbital rim and floor via a lower lid or transconjunctival incision, and the zygomaticomaxillary buttress via an upper buccal sulcus incision. Plate fixation is required at two points, the zygomaticofrontal suture and at least one other of the infraorbital rim or zygomaticomaxillary buttress. If wire fixation is required, all three sites must be secured to prevent rotational deformity. The lateral orbital wall must also be inspected, as malalignment of fragments here will translate into orbital volume changes and secondary deformity.

Maxillary fractures

The maxilla forms a significant part of the middle third of the face. A system of strong buttresses reinforce this structure. They include the vertical nasomaxillary buttress lying along the lateral piriform margin, the zygomaticomaxillary laterally, and the pterygomaxillary buttress posteriorly. The horizontal buttresses include the alveolus and the infraorbital rims and the zygomatic arches. Reconstruction of the buttresses is the key to restoration of midfacial form.

Diagnosis

Le Fort classified typical patterns of midfacial fractures (Fig. 4). The Le Fort I (Guerin) fracture line passes transversely from the floor of the piriform aperture across the base of the maxillary sinus to the pterygoid plates, leaving a mobile fragment containing the entire maxillary alveolar bone, palate, and the pterygoid plates (floating palate). The Le Fort II (pyramidal) fracture line passes across the nasal bones into the medial part of the orbit (so that the lacrimal crests are included in the mobile segment), diagonally across the maxilla, below the zygoma, and into the pterygoid plates. The segment consists of a pyramidal piece of bone, including the piriform aperture, the maxillary alveolus, and lacrimal region. The Le Fort III facture (craniofacial dysjunction) passes through the nasofrontal suture, across the orbital floor to

the zygomaticofrontal suture and lateral orbital wall, and to the zygomatic arch and pterygoid plates, leaving the entire midfacial skeleton detached from the skull base. Sagittal fractures of the maxilla and isolated fractures of the maxillary alveolus may also occur. Fracture patterns in the maxilla are often asymmetrical, and multiple levels of fracture may coexist (Fig. 5).

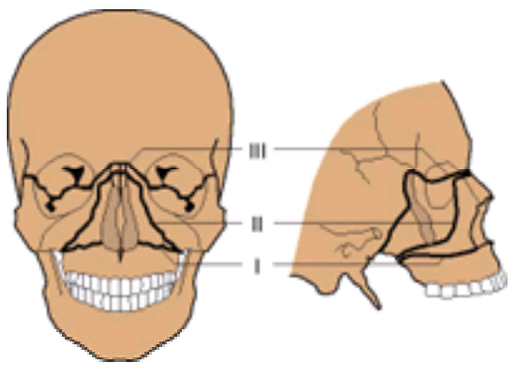


Fig. 4. Diagram showing patterns of Le Fort facial fractures.

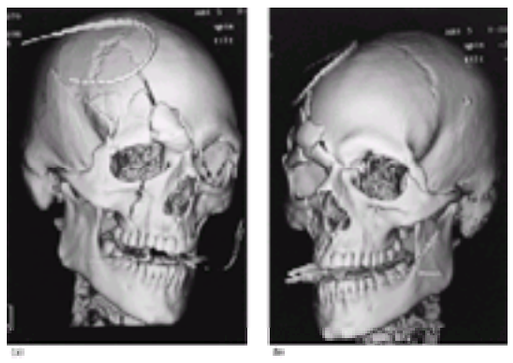


Fig. 5. Three-dimensional CT scans demonstrating asymmetrical craniofacial fractures. Left Le Fort III fracture and right zygomatico-orbito-frontal fracture.

Depending on the level of fracture, the patient may manifest signs of epistaxis, ecchymosis of the periorbital and subconjunctival tissues, and significant facial oedema. Displacement of the fracture fragments is often posteriorly and downward. This gives the midface an elongated and retruded appearance, the ‘donkey facies' or dish face deformity. Malocclusion is the most sensitive symptom of midfacial fracture, often manifest as an anterior open bite. Palpable bony deformity should be sought along fracture lines both extra- and intraorally. Mobility of the maxillary segment is the pathognomonic sign of Le Fort fractures, and is demonstrated by stabilizing the glabellar region with one hand while applying a distracting force to the maxillary alveolus with the other. The relative motion at the nasofrontal and zygomaticofacial sutures and palate demonstrate the fracture pattern clinically. The test may, however, be negative in impacted or greenstick fractures. Laceration, haematoma, or mobility of the hard palate may indicate sagittal fracture of the palate. Cerebrospinal fluid rhinorrhoea or otorrhoea may occur if the fracture communicates with the anterior or middle cranial fossas.

Investigation

According to the fracture pattern, plain films (Waters' view) may demonstrate opacity of the maxillary sinus, fracture lines at the nasofrontal suture, orbital floor, zygomaticofrontal suture, and arch with vertical elongation of the orbital margins in Le Fort III fractures or inferior orbital margin and zygomaticomaxillary buttress fractures in Le Fort II fractures. Axial and coronal CT scans remain the definitive investigation.

Treatment

Maxillary fractures are approached through a combination of the previously mentioned incisions. The nasofrontal suture, medial and lateral orbital walls, and zygomatic arches are approached through a coronal incision. The inferior orbital margin, orbital floor, and anterior zygoma are accessed through lower lid incisions. The lower maxilla is exposed through an upper buccal sulcus incision. Reduction involves mobilization of the maxillary fragment, often with Rowe's disimpaction forceps, and anatomical reduction with particular attention to re-establishing preinjury functional occlusion. The occlusion is held with intermaxillary fixation, and fixation of fractures is achieved with plates and screws, placed along the axes of the major facial buttresses. Bone grafting may be required to reconstitute the buttresses in cases of severe comminution, to bridge bone gaps greater than 5 mm, or to recreate normal bony contour. Intermaxillary fixation is usually removed at the completion of the procedure, reducing the need for tracheostomy.

Complex/panfacial injuries

The principles of treatment of panfacial fractures are as outlined in the introduction. The mandible is reduced and fixed first, as it is the most substantial of the facial bones and the easiest to align correctly, and will re-establish facial height and lower facial width and anterior projection. The maxilla is then fixed to it with intermaxillary fixation, to re-establish an occlusion. The zygomas are fixed, starting with the arches and zygomaticofrontal sutures to establish midfacial width, and fixation then proceeds to the other maxillary buttresses. Nasoethmoido-orbital and central fractures are then addressed.

Orbital fracture

Orbital fractures may be isolated or occur in association with fractures of the zygoma, nasoethmoido-orbital region, or high Le Fort fractures. Isolated orbital floor fractures, ‘blowout fractures', occur when pressure transmitted through the globe or orbital margin disrupts the weakest part of the orbit, the orbital floor and associated medial wall (lamina papyracea). The fracture is typically comminuted.

Clinical

The signs and symptoms of isolated blowout fracture include periorbital and subconjunctival ecchymosis, diplopia, enophthalmos, surgical emphysema of the periorbital tissues, and numbness in the distribution of the infraorbital nerve including the upper teeth and gums. A thorough ophthalmological examination must be performed on all patients with suspected orbital fractures to exclude injury to the globe or its supporting structures. Diplopia may be mechanical or functional in origin. Mechanical causes include entrapment of the inferior rectus or oblique muscles or, more commonly, periorbital fat and fascial attachments, within the fracture line. This in turn limits globe movement, most commonly on upward gaze (Fig. 6). Non-mechanical causes include extraocular muscle or nerve contusion. Enophthalmos occurs when part of the orbital floor or wall is disrupted, either increasing the volume of the orbit or allowing intraorbital fat to escape beyond the normal boundaries of the bony orbit, in turn allowing the globe to sink back. It may also be due to entrapment of periorbital tissues in the fracture line, when the globe is posteriorly displaced, at the moment of impact on the globe. This then tethers the globe in a posterior position. Lastly, fat atrophy may cause enophthalmos of delayed onset. Enophthalmos can, in turn, cause descent of the upper lid, pseudoptosis, and deepening of the supratarsal fold.

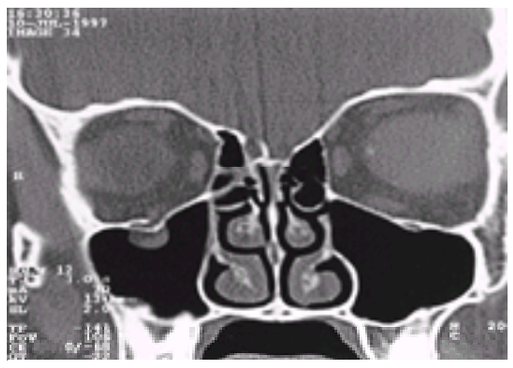


Fig. 6. Coronal CT scan demonstrating small right sided blow-out fracture withentrapment of inferior rectus muscle. The patient complained of diplopia on upward gaze.

Investigation

The diagnosis of blowout fracture is often difficult to make clinically. The forced duction test may differentiate between mechanical and non-mechanical diplopia, however oedema in the periorbital tissues may make early results equivocal. Enophthalmos may be masked by soft tissue swelling in the first week after the injury. Plain radiographs, particularly the Waters' view, may show disruption of the orbital floor, an air–fluid level within, or opacity of the maxillary sinus. The presence of prolapsed soft tissue below the orbital floor may be seen as the teardrop sign. The definitive investigation is coronal CT scan of the orbit.

Treatment

The principles of treatment of a blowout fracture are to release incarcerated soft tissues, accurately define the extent of the fracture, replace orbital contents into the confines of the normal bony cavity, and reconstruct the bony orbit in its normal position, restoring the normal orbital cavity volume.

The indications for surgical exploration of the orbital floor include:

i. symptomatic diplopia either with a positive forced duction test or which does not resolve within 10 days;
ii. radiological evidence of entrapment;
iii. early enophthalmos of 3 mm or more;
iv. a large bony defect likely to result in late enophthalmos; or
v. the presence of other craniofacial fractures requiring exploration.

The orbital floor can be approached through a lower lid skin or transconjunctival incision. The full extent of the fracture is defined and trapped tissues released. The medial wall must also be inspected to exclude concomitant fracture. The fracture fragments are reduced if possible and if stable reduction is achieved, this is adequate. Otherwise the orbital floor must be reconstituted to achieve a normal orbital cavity size. This can be achieved by use of bone graft, either split calvarium or rib, iliac crest, or maxillary antral wall, or by alloplastic materials.

Complications

These include ocular injury and impaired vision, diplopia, infraorbital anaesthesia, enophthalmos, infection, implant extrusion, and scleral show.

Nasoethmoido-orbital (NOE) fractures

Trauma to the interorbital region can result in a fracture pattern involving the nasal root, medial orbital wall, and ethmoid air cells. The fractures are characteristically comminuted. The nasal root is often depressed, giving nasal shortening and loss of dorsal prominence. The medial orbital wall is often displaced, usually in a blowout pattern with the potential for enophthalmos, but occasionally laterally causing exophthalmos. If the fragment of the medial orbital rim to which the medial canthal tendon is attached becomes laterally displaced, traumatic telecanthus may result.

NOE fractures generally require open reduction. Access is gained via overlying lacerations, coronal incision, or others such as the Lynch (medial orbital) incision. The nasal bones are elevated and a dorsal nasal bone graft inserted if necessary. The nasomaxillary buttress is reconstituted. In most cases the medial canthal tendon remains attached to a bone fragment. If this is large, anatomical reduction of the fragment repositions the canthus. If smaller, transnasal bony fixation is required and if minute or avulsed, tendon reattachment by transnasal wiring or by reanchoring using bone fixation devices can be performed. The medial orbital wall is reconstructed with a bone graft where necessary. The lacrimal system is not explored concurrently, as most cases of obstruction settle over 6 weeks, but is reconstructed at a later time if necessary.

Fractures of the frontal sinus

The potential complications of inadequately treated fracture of the frontal sinus include meningitis, mucocele, cerebral abscess, and osteitis. Clinical findings include overlying laceration or bruising, palpable bony depression, and cerebrospinal fluid rhinorrhoea. The Caldwell and lateral views may show fracture lines, an opaque sinus, or air–fluid level or pneumocephalus. Any of these findings should prompt axial and coronal CT scanning of the region.

The indications for operative treatment include anterior table displacement causing contour deformity of the forehead, nasofrontal duct involvement or obstruction, and displaced posterior wall fracture with significant displacement or cerebrospinal fluid leak. Exploration is performed either through an overlying laceration or coronal incision. Isolated anterior wall fractures are reduced and fixed with plates and screws. Involvement of the nasofrontal duct may be suspected with associated ethmoidal fractures and if present, anterior obliteration is recommended. This involves complete removal of all sinus mucosa under magnification, removal of the inner cortex of the bone using a high-speed burr, occlusion of the nasofrontal duct with a bone plug, and sinus obliteration by spontaneous osteogenesis or autogenous bone graft. Posterior table fractures may result in communication between the sinus and intracranial cavity. Minimally displaced fractures (less than one thickness of the posterior table) may be treated conservatively. However, in the presence of a cerebrospinal fluid leak with assumed dural tear, the frontal sinus should be cranialized by removing the posterior wall of the sinus, allowing brain and dural expansion to fill the resulting dead space, stripping all lining mucosa, and plugging the nasofrontal duct. The dura is repaired and a pericranial patch or flap applied to reline the anterior fossa. Long-term follow-up is required to exclude late development of frontal mucocele. The treatment of frontal fractures is summarized in Table 1.

Anterior wall	*Drainage system involved*	*Posterior wall fracture*	*Cerebrospinal fluid leak*	*Treatment*
Undisplaced	No	No	No	Nil
Displaced	No	No	No	ORIF
Displaced	Involved	No	No	Obliteration procedure/ spontaneous osteogenesis + ORIF
Displaced/undisplaced	Involved/uninvolved	Displaced more than/thickness of inner wall	Yes or no	Cranialization + ductal repair if/ not present
Displaced/undisplaced	Involved/uninvolved	Displaced or undisplaced	Yes	Ductal repair + cranialization
Undisplaced	Uninvolved	Undisplaced	No	Nil

ORIF, open reduction and internal fixation.

Table 1 Summary of treatment of frontal fractures

Further reading

David DJ, Simpson DA. *Craniomaxillofacial trauma*. Edinburgh: Churchill Livingstone, 1995. [A comprehensive volume with broader coverage of dental, ophthalmological, and anaesthetic topics.]

Ellis E, Dean J. Rigid fixation of condylar fractures. *Oral Surgery, Oral Medicine, Oral Pathology*, 1993; **76**:6-15.

Gruss JS. Nasoethmoidoorbital fractures: classification, fixation and role of primarybone grafting. *Plastic and Reconstructive Surgery,* 1985; **75**: 303.

Gruss JS, Bubak PJ, Egbert MA. Craniofacial fractures: An algorithm to optimize results. *Clinics in Plastic Surgery* 1992; **19**: 195–206.

Hanson DH. Nasal injuries: their pathogenesis and treatment. *British Journal of Plastic Surgery,* 1979; **32:** 57-64.

Manson PN, Dark N, *et al.* Subunit principles in midface fractures. *Plastic and Reconstructive Surgery*, 1999; **103**: 1287-306. [Up-to-date summary of midface fracture treatment including soft tissue management.]

Rohrich RJ, Hollier LH. Management of frontal sinus fractures. Changing concepts. *Clinics in Plastic Surgery* 1992; **19**: 219–32. [An excellent review from an issue entirely devoted to the topic of facial skeletal injury.]

Rowe NL, Williams JL. *Maxillofacial injuries.* Edinburgh: Churchill Livingstone, 1994.[A comprehensive two-volume text covering ail aspects of facial trauma.]

45.12 The surgical treatment of snoring and obstructive sleep apnoea

Robert J. O. Davies, Grant Bates and John R. Stradling

Introduction
The epidemiology of snoring and sleep apnoea
The assessment of snoring and sleep apnoea
History
Examination
Sleep study
The treatment of snoring
Mandibular advancement devices
Surgery for snoring
Surgery for obstructive sleep apnoea
Tracheostomy
Tonsillectomy
Uvulopalatopharyngoplasty
Other surgical procedures
Surgical techniques
Septoplasty and turbinate resection
Uvulopalatopharyngoplasty
Palatal stiffening
Laser-assisted palatoplasty
Conclusions
Further reading

Introduction

Snoring and obstructive sleep apnoea are due to partial or complete collapse of the muscular pharyngeal airway during sleep. Patency of this section of the airway is maintained by a balance of forces across its muscular wall as there is no bony or cartilaginous support at this site. The forces tending to collapse the pharynx are the negative inspiratory pressure within the pharyngeal lumen and the positive extrinsic tissue pressure. Airways patency is maintained by airways muscle tone and collapse results from any factors which narrow the airways or disturb the balance of these forces. These include nasal obstruction (which increases negative intrapharyngeal inspiratory pressure), excessive tonsillar tissue or generally narrow airways, which is usually due to obesity or an abnormal facial skeleton. Snoring and sleep apnoea develop at sleep onset with the normal reduction in pharyngeal muscle tone which occurs at this time.

The sound of snoring is due to the partially obstructed pharyngeal airway and is not directly due to the pharyngeal wall vibration. Obstructive sleep apnoea ensues when the balance of pharyngeal forces leads to complete pharyngeal collapse. Once the airway is totally obstructed surface tension adds to the forces holding the airway closed until arousal from sleep (triggered by increasing respiratory effort) reopens the airway by returning muscle tone. Treatment for snoring and obstructive sleep apnoea (including surgery) is aimed at improving airway patency by optimizing the balance of forces across the airway wall.

The epidemiology of snoring and sleep apnoea

Simple snoring and severe obstructive sleep apnoea with ventilatory failure represent extremes of a continuous spectrum of abnormality. It is therefore not appropriate to consider snoring and sleep apnoea as ‘separate diseases'. Approximately 20 per cent of adult men snore and at least 0.3 per cent of the adult male population have severe obstructive sleep apnoea which occurs in all positions of sleep.

The assessment of snoring and sleep apnoea

Both medical and surgical therapies are available to treat snoring and sleep apnoea and the selection of the most appropriate therapy requires the following information.

1. Whether snoring is present and is intrusive enough to warrant treatment.
2. The anatomical cause for the upper airway problem.
3. The severity of any obstructive sleep apnoea.

History

Treatment is indicated for simple snoring to reduce the social distress it causes. It is therefore helpful to interview the spouse as well as the patient. Questions are asked about the frequency of the noise and whether it varies with posture or alcohol intake. A partner who is consistently forced to sleep in another room is a useful marker of a substantial problem. However, marital disharmony should be considered as minor snoring can become an inappropriate focus of discontent. Questions are also asked about features suggesting sleep apnoea. These include excessive daytime sleepiness, restless sleep, breath-holding episodes during sleep, and nocturia. These features are not sufficiently sensitive and specific to predict reliably the presence of obstructive sleep apnoea without a sleep study and daytime sleepiness and restless sleep can occasionally be due to very heavy snoring alone.

Examination

Physical examination aims to identify the anatomical causes for upper airway collapse. It should document the presence of obesity and particularly neck obesity (which can be quantified as the neck circumference) and the mandibular and maxillary structure (identifying a small jaw which crowds the retroglossal airway). The nose and throat examination aims to identify the likely site of airway obstruction. This may be the nose, oropharynx, or hypopharynx, or a combination of these sites.

The nasal airway is examined with a thudicum speculum and then a fibreoptic telescope. The septum may be significantly deviated, the middle turbinate may be expanded by excessive pneumatization, or there may be general mucosal swelling including polyps (rhinitis).

The pharynx is assessed with the flexible fibrescope which is passed through the nose into the nasopharynx. From this position the closure of the soft palate and the lateral pharyngeal walls can be assessed (Fig. 1). The Mueller manoeuvre, in which the patient attempts to inspire against a closed glottis and with the nose sealed around the telescope, emphasizes the site of greatest airways collapsibility. ‘Sleep nasendoscopy', in which the patient is sedated with benzodiazepines during the fibrescope examination, is controversial. This technique may disclose a site of maximal airway collapse, although the muscle relaxation associated with the sedation may overemphasize apparent abnormality and thus not represent the situation during normal sleep. Lateral cephalometry, computed tomography, and magnetic resonance imaging can identify masses impinging on the airways and quantify any retrognathia. Newer techniques can also image dynamic airway collapse, although the clinical value of these techniques has yet to be established.

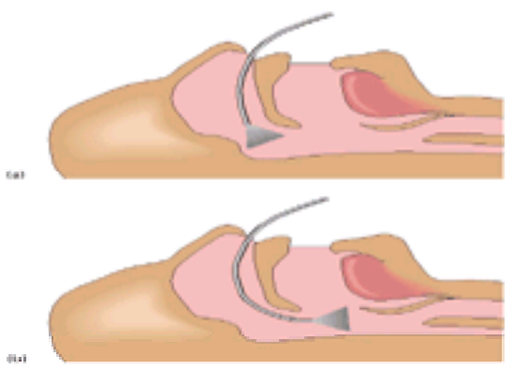


Fig. 1. The positioning of the fibrescope for nasendoscopy to first assess palatal closure and then the tongue base and larynx.

Sleep study

All patients undergoing surgery for snoring or obstructive sleep apnoea should have a sleep study. This is to confirm the presence of snoring and to quantify the severity of any obstructive sleep apnoea. An appropriate study should document the presence of snoring, sleep disturbance (the cause of daytime sleepiness), and episodes of apnoea. The documentation of snoring is important to confirm an indication for treatment and because nocturnal stridor and vocalization during dream sleep maybe miss-classified as snoring.

The treatment of snoring

All patients seeking help for snoring should be advised to achieve an optimal body weight and avoid alcohol after 6 p.m., if possible. Rhinitis should be treated according to its cause, such as atopy, and in most cases topical nasal steroids are effective. Where such measures are ineffective, then therapeutic options include a mandibular advancement device or surgery.

Mandibular advancement devices

Mandibular advancement devices reduce snoring and control moderate obstructive sleep apnoea by protruding the mandible during sleep. This displaces the tongue forward, enlarging the retroglossal airway and so reducing the likelihood of turbulent airflow. These devices are attached to the dentition and good teeth are a necessity for their use. They are removable, relatively inexpensive and free from irreversible side-effects which frequently makes them preferable to surgical intervention.

Surgery for snoring

The most commonly performed surgical procedure for snoring is uvulopalatopharyngoplasty, with smaller roles for nasal surgery and tonsillectomy. Despite the frequency with which they are performed, evidence for the efficacy of these approaches is poor—a deficit emphasized in systematic reviews. The literature includes no randomized trials and most reports only describe the patient's subjective perception of their benefit and not objective sound recordings. The considerable pain associated with uvulopalatopharyngoplasty results in a substantial placebo response and observations based only on subjective reports are therefore unreliable. There are two reports of objective snoring measurements before and after uvulopalatopharyngoplasty. In both, subjects reported substantial subjective benefit which was not confirmed by the objective recordings. In one study there was no objective improvement at all and in the other there was only minor benefit. New techniques using laser technology to stiffen, resect, or scar the soft palate have been developed. There is no sound evidence to define the safety and efficacy of these techniques, although early data suggest some improvement in snoring.

Given the limited evidence for efficacy and the side-effects of soft palate surgery, it should be reserved for severe cases where simpler therapy is ineffective and the patient should be aware of the problems and limited proven efficacy of these techniques. There is a compelling need for controlled trials to define the safety and efficacy of these procedures.

Surgery for obstructive sleep apnoea

Tracheostomy

Tracheostomy is effective in correcting obstructive sleep apnoea by bypassing the collapsing pharyngeal segment and was the standard therapy for this disorder prior to the advent of nasal continuous positive airway pressure. Because of the efficacy of nasal continuous positive airway pressure, tracheostomy is now rarely needed but may be useful in patients with severe obstructive sleep apnoea who are unable to use nasal continuous positive airway pressure (less than 0.5 per cent of patients).

Tonsillectomy

Tonsillectomy is effective where tonsillar enlargement is the primary cause of obstructive sleep apnoea, as is usual in children where clinical trials have confirmed its efficacy. Tonsillectomy is also effective in the occasional adult obstructive sleep apnoea patient with large tonsils.

Uvulopalatopharyngoplasty

Uvulopalatopharyngoplasty is sometimes suggested for the treatment of obstructive sleep apnoea in adults. Compared with control, this produces some improvement in sleep apnoea severity, although this improvement is modest when compared with nasal continuous positive airway pressure. Nasal continuous positive airway pressure is more than 90 per cent effective in correcting obstructive sleep apnoea whereas uvulopalatopharyngoplasty produces a reduction in sleep apnoea severity of 50 per cent in only 50 per cent of treated cases. Nasal continuous positive airway pressure should be the first-line treatment for patients who have moderate/severe obstructive sleep apnoea as meta-analysis shows the efficacy of uvulopalatopharyngoplasty declines with increasing sleep apnoea severity, and soft palate surgery reduces the ability of patients to tolerate nasal continuous positive airway pressure treatment. Obesity is a relative contra-indication to uvulopalatopharyngoplasty as it is associated with retroglossal as well as velopharyngeal airway collapse which predicts poor uvulopalatopharyngoplasty outcome. Currently, uvulopalatopharyngoplasty should only be used for obstructive sleep apnoea in patients who cannot tolerate nasal continuous positive airway pressure and who have been informed of its limited efficacy and side-effects.

Other surgical procedures

More radical surgical procedures have been used in small groups of patients. Procedures in this group include genioglossal advancement and hyoid myotomy, and mandibular and maxillary osteotomy with advancement. Data describing the efficacy of these procedures is limited to short case series. Mandibular and maxillary osteotomy appears to be capable of controlling even severe obstructive sleep apnoea whereas the efficacy of genioglossal advancement and hyoid myotomy remains contentious.

Surgical techniques

Septoplasty and turbinate resection

The septoplasty operation is to reposition a displaced septum in the mid-line of the nasal cavity. An incision is made in front of the columella and the attachments of the quadrilateral cartilage are freed until a mid-line position is obtained.

Endoscopic middle turbinate resection is without significant risk, but partial or complete inferior turbinectomy can cause severe epistaxis. For this reason lasers are now used to reduce the size of the inferior turbinates, to reduce the frequency of bleeding.

Uvulopalatopharyngoplasty

The principle of uvulopalatopharyngoplasty is to increase the oropharyngeal airway without jeopardizing the functions of the soft palate. The operation is performed under general anaesthesia using an oral endotracheal tube and a Boyle–Davis gag. The position is the same as for tonsillectomy and if the tonsils are present they are removed. Meticulous attention is then paid to haemostasis of the tonsil bed. If the patient has had a tonsillectomy the mucosa between the tonsil pillars is removed.

In order to decide how much soft palate to resect, the soft palate is pushed back against the posterior pharyngeal wall, and the point of contact noted (Fig. 2). The incision in the soft palate is started below this point so that pharyngeal competence can be maintained. ‘An incision is made and carried from this point in a horizontal direction to the anterior pillars on both sides (Fig. 3). In this way the uvula and a strip of soft palate are resected. A section of the anterior pillar is resected, but it is advisable to preserve most of the posterior pillars. The aim should be to produce a ‘box-like' resection with conservation of the majority of the middle of the palate (Fig. 4). The resection can be carried out with a scalpel, cutting diathermy, CO2, Nd:YAG, KTP or diode lasers. In general lasers are more accurate, coagulate better and

may reduce postoperative pain. Once haemostasis has been established the posterior pillar is advanced and sewn to the anterior pillar. In this way the incision lines are pulled forward, away from the posterior pharyngeal wall. A 3/0 chromic catgut is used and a solid bite is required through the muscle layers. It is not necessary to sew up the leading edge of the soft palate. All sutures should be cut short to avoid irritation.

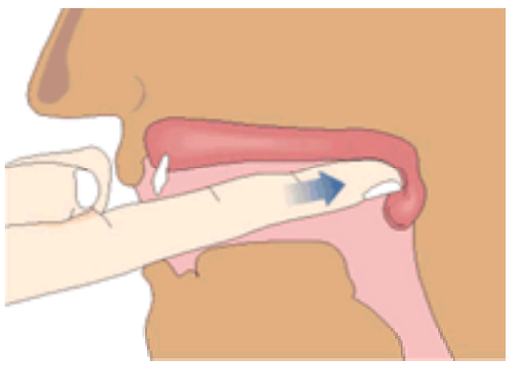


Fig. 2. Using a finger to judge how much soft palate to resect.

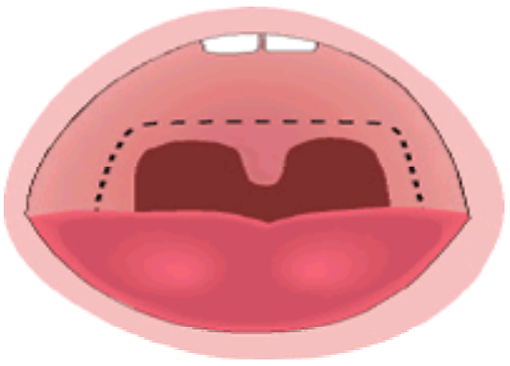


Fig. 3. The technique for uvulopalatopharyngoplasty.



Fig. 4. The pharynx of a ‘typical snorer'.

At the end of the procedure it will be noted that the lateral dimension and the anterior/posterior diameter of the oropharyngeal airway has been increased (Fig. 4 and Fig. 5). Marcaine, 0.5 per cent, with 1 in 200 000 adrenaline is then slowly injected into the lateral margins of the soft palate. It is sensible to leave a nasopharyngeal airway in place, trimmed so that it hangs down just behind the palate. This can be removed once the patient is fully recovered from anaesthesia.



Fig. 5. The postoperative result after palatoplasty. There is an increase in the velopalatine airway.

The postoperative management of uvulopalatopharyngoplasty patients is particularly important. The majority of postoperative complications are minor, but occasionally significant morbidity and even mortality can occur. The immediate postoperative period can be associated with airway obstruction due to the combined effects of sedation, muscle relaxation, oedema, and/or haemorrhage. If primary haemorrhage occurs it is managed in the same way as following tonsillectomy. However, postoperative oedema and the presence of blood in the upper airway can make reintubation extremely difficult and if there is any doubt a temporary tracheostomy should be performed prior to intervention. Secondary haemorrhage (typically 10 days) occurs as a result of postoperative infection, and usually requires antibiotics only.

Uvulopalatopharyngoplasty is the most painful operation in the ear, nose, and throat repertoire. A combination of analgesics are required; perioperatively this should include a small dose of opiate, local anaesthetic infiltration, and an anti-inflammatory suppository. As local anaesthesia and opiate sedation reduce upper airway tone and so worsen obstructive sleep apnoea, the need for this sedation regimen emphasizes the importance of excluding obstructive sleep apnoea pre-operatively. The patient should be encouraged to start a soft bland diet as soon as possible.

The exact side-effect profile of uvulopalatopharyngoplasty is not clearly described as few studies have systematically reported complication rates in a sizeable cohort of patients. In a retrospective review of 101 cases having surgery for frank obstructive sleep apnoea—and hence high risk cases: 11 per cent of patients experienced upper airway obstruction (one fatal); 5 per cent suffered postoperative haemorrhage requiring a return to theatre; 24 per cent reported some persistent nasal regurgitation of fluids on swallowing; 31 per cent reported a ‘dry throat'; 10 per cent described non-specific swallowing disturbance; and 5 per cent non-specific ‘breathing difficulty'. Temporary velopharyngeal incompetence with nasal regurgitation occurs in most patients. This usually subsides after 2 to 3 weeks. Hyponasal speech is a manifestation of velopharyngeal incompetence and occurs if to much palate has been excised. The loss of the uvula may be missed by individuals who speak languages requiring pharyngeal sounds.

Palatal stiffening

Patients thought to have palatal flutter on sleep nasendoscopy are selected for this operation. A diamond section of mucosa is removed from the palatal surface of the soft palate and a uvulectomy is performed. The scarring that results stiffens the mid-section of the soft palate and reduces vibration. Promising initial subjective results are suggested but snoring may recur later.

Laser-assisted palatoplasty

Laser-assisted palatoplasty is usually performed under local anaesthetic with the patient sitting upright. Using a laser, vertical incisions are made through the palate on each side of the uvula (Fig. 6). The uvula is shortened and re-shaped. Laser-assisted palatoplasty may prove to be useful in severe snorers who do not have obstructive tonsils, although there are no controlled trials to confirm this.

Fig. 6. Laser-assisted palatoplasty technique.

Conclusions

For simple snoring, nasal surgery and tonsillectomy are relatively free of long-term complications and are effective in selected patients. Soft palate surgery is associated with significant complications and evidence for its therapeutic efficacy is poor. Where soft palate surgery is used, pre-operative consent should include a warning of the postoperative pain, the limited evidence concerning efficacy, and possible side-effects.

In obstructive sleep apnoea, tonsillectomy is effective for subjects with substantial tonsillar tissue, which is the commonest cause of obstructive sleep apnoea in children. Soft palate resection is partially effective but this is less than the benefit seen with nasal continuous positive airway pressure and, therefore, surgery is used as second-line therapy following failed nasal continuous positive airway pressure.

Further reading

Ali NJ, Pitson D, Stradling JR. Sleep disordered breathing: effects of adenotonsillectomy on behaviour and psychological functioning. *European Journal of Pediatrics* 1996; **155**: 56–62. [Controlled study of the efficacy of tonsillectomy for children with sleep apnoea due to tonsillar hypertrophy.]

Croft CB, Pringle M. Sleep nasendoscopy: a technique of assessment in snoring and obstructive sleep apnoea. *Clinical Otolaryngology* 1991; **16**: 504–9. [Description of sleep nasendoscopy.]

Ellis PD. Laser palatoplasty for snoring due to palatal flutter: a further report. *Clinical Otolaryngology* 1994; **16**: 504–9. [Description of laser palatoplasty.]

Miljeteig H, Mateika S, Haight JS, Cole P, Hoffstein V. Subjective and objective assessment of uvulopalatopharyngoplasty for treatment of snoring and obstructive sleep apnea. *American Journal of Respiratory Critical Care Medicine* 1994; **150**: 1286–90. [Objective measurements of sound changes after soft palate surgery for snoring.]

Mortimore IL, Bradley PA, Murray JA, Douglas NJ. Uvulopalatopharyngoplasty may compromise nasal CPAP therapy in sleep apnea syndrome. *American Journal of Respiratory Critical Care Medicine* 1996; **154**: 1759–62. [The adverse effects of soft palate resection on tolerability of positive airway pressure therapy in sleep apnoea..]

Prichard AJ, Smithson AJ, White JE, Close PR, Drinnan MJ, Griffiths CJ, Marshall HF, Gibson GJ. Objective measurement of the results of uvulopalatopharyngoplasty. *Clinical Otolaryngology* 1995; **20**: 495–8. [Methodologically sound assessment of the efficacy of soft palate resection for snoring.]

Sher AE, Schechtman KB, Piccirillo JF. The efficacy of surgical modifications of the upper airway in adults with obstructive sleep apnea syndrome. *Sleep* 1996; **19**: 156–77. [Meta-analysis of the efficacy of upper airway surgery on snoring and sleep apnoea.]

45.13 Pharyngeal pouch surgery

Grant Bates

Presenting features
Investigations
Treatment
Historical perspective
Surgical technique
Postoperative care
Complications
Alternative endoscopic techniques
Further reading

A pharyngeal pouch is a pulsion type of diverticulum arising from the posterior wall of the pharynx between the two components of the inferior constrictor muscle (thyropharyngeus and cricopharyn-geus). It forms gradually and is thought to be caused by incoordination of the pharyngeal muscles during swallowing. The incoordination produces a high pressure above the thyropharyngeus muscle and, as a result, the pharyngeal mucosa herniates backwards through a potential area of weakness known as Killian's dehiscence (Fig. 1). The pouch develops posteriorly and is then deflected to one side or the other. Patients with a pharyngeal pouch often have associated gastric reflux, which may impair the function of the inferior constrictor muscle.

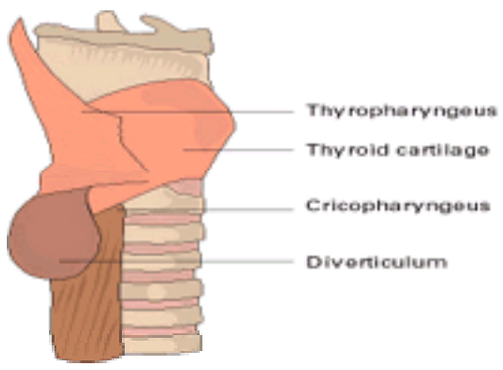


Fig. 1. A pharyngeal pouch is a pulsion diverticulum that presents through an area of weakness known as Killian's dehiscence.

The incidence of the pharyngeal pouch in the population of the United Kingdom is 1 per 100 000 per year.

Presenting features

The patient is usually elderly. The symptoms consist of progressive dysphagia, regurgitation of undigested liquids and solids, halitosis and, frequently, a cough. The patient may notice gurgling noises in the neck. The initial presentation may be with pneumonia or recurrent chest infections that have resulted from ‘silent' aspiration.

On examination the patient is often thin and pooling of saliva can usually be seen in the pyriform fossas on nasendoscopy.

Investigations

The investigation of choice is a video swallow or, if not available, a barium swallow (Fig. 2). A chest radiograph and basic blood tests should be obtained.

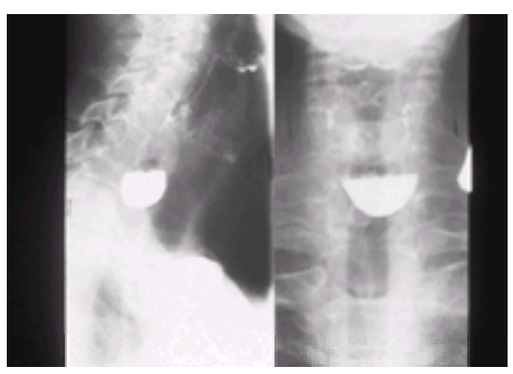


Fig. 2. Barium swallow, lateral and anteroposterior, showing a large pouch.

Treatment

A patient with a pharyngeal pouch usually has significant symptoms and, if left untreated, these progress. As the patient is usually elderly, at first presentation and in most cases it is therefore sensible to offer treatment as soon as possible. The only effective treatment is surgery. Until recently, the choice of operation was controversial.

Historical perspective

In 1764 Ludlow, a Bristol surgeon, first described a patient with a pharyngeal pouch. Wheeler, in 1886, performed the first successful pouch excision. An endoscopic diverticulotomy operation to divide the cricopharyngeal bar that lies between the oesophagus and pouch was devised by Mosher in 1906.

Both the external excision technique and the diverticulotomy have been refined throughout the twentieth century. Modifications of the external approach include the use of a stapling gun to seal the pharyngeal lumen, or complete inversion of the pouch. The technique of endoscopic diverticulotomy was improved by Dolman, who introduced a special beaked endoscope and used diathermy to divide the cricopharyngeal bar. Despite these technical refinements, pouch excision and endoscopic diverticulotomy using diathermy both have a high morbidity and significant mortality. The main complications include fistula formation, mediastinitis, and recurrent laryngeal nerve palsy. The external operation, in particular, is also a major undertaking for an elderly person.

In 1992, Callard (Belgium) and Newbegin and Van Hirsch (United Kingdom) simultaneously and independently described a new endoscopic technique that used a stapling gun to divide the cricopharyngeal bar. This technique has now been taken up widely and has been evaluated over the last 6 years. The key message from the recent United Kingdom Enquiry into Perioperative Deaths was that the endoscopic stapling technique should become the method of choice and that it should be performed by surgeons with a special interest in pouch surgery.

Surgical technique

Preoperative preparation

The patient's general medical condition should be optimized before surgery. Access with the endoscope is made difficult by prominent teeth, retrognathism, and/or a stiff neck (Fig. 3). The author warns patients whose anatomy looks difficult that endoscopic resection may not be possible. In practice (the first 100 consecutive patients), endoscopic resection has always been possible, albeit difficult at times.



Fig. 3. These ferocious teeth belonged to a senior London solicitor! They made access difficult but not impossible.

The operation

A general anaesthetic is required. The patient is supine and, generally, the best head position is with the cervical spine flexed and the head extended on the atlanto-occipital joint (Fig. 4). The teeth, or gums, are protected with a guard or a wet swab. Special endoscopes are available for pouch surgery [the Weerda scope (Stortz Co.) is shown in Fig. 5]. The endoscope is gently inserted so that the top blade lies above the oesophageal lumen. The depth and angle between each blade of the scope can be adjusted so that an optimum view of the cricopharyngeal bar can be obtained (Fig. 6).



Fig. 4. The patient lies supine and the endoscope has a special support allowing the surgeon to work with two hands.

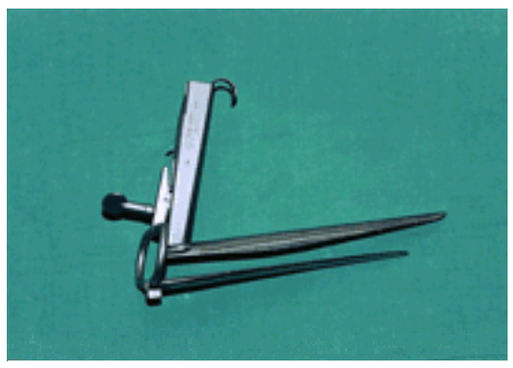


Fig. 5. A specially designed, two-beaked endoscope can be used in most patients.



Fig. 6. The endoscopic appearance: the anterior beak of the endoscope lies in the oesophageal lumen and the wide-necked pouch can be seen posteriorly.

There is often debris in the pouch, which needs to be carefully removed with a rubber-tipped sucker. (Mucosal trauma should be avoided because any bleeding impairs the view.) The walls of the pouch should then be inspected carefully. Any mucosal abnormality should be biopsied, although this is a rare event. (The instance of carcinoma in a pouch is very low; three in 1000 consecutive cases reported from the Mayo Clinic.)

The stapling gun (e.g. Ethicon TSB 35) is passed down the scope. The beaks of the gun are passed over the cricopharyngeal bar and then locked in position. It is important to avoid taking an excessive bite of tissue, which may jam the gun. The operator then fires the gun, which cuts through the cricopharyngeal bar and simultaneously seals the edge of the incision with three rows of titanium staples (Fig. 7). Two cartridges of staples are usually required, although a large pouch may take more. Unless the procedure is traumatic, a nasogastric tube or prophylactic antibiotics are not required.

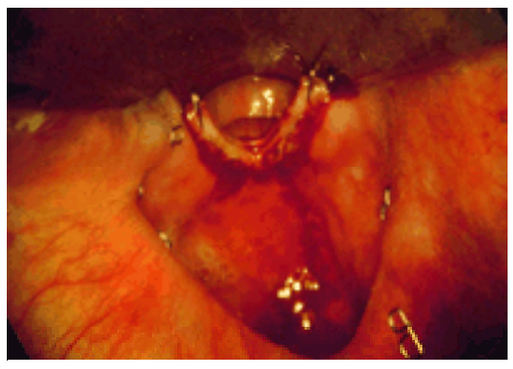


Fig. 7. The stapling gun divides the cricopharyngeal bar and seals the edges with three rows of titanium staples.

When there is difficult access

In approximately 10 per cent of patients it will not be possible to insert a wide, bivalved endoscope. Under these circumstances a Negus pharyngoscope can be used (Fig. 8). The Negus is narrower but does permit a view of the cricopharyngeal bar as well as allowing the staple gun to pass down the lumen. To aid positioning of the jaws of the gun, a fibreoptic telescope can be passed down the light-channel guide of the endoscope.

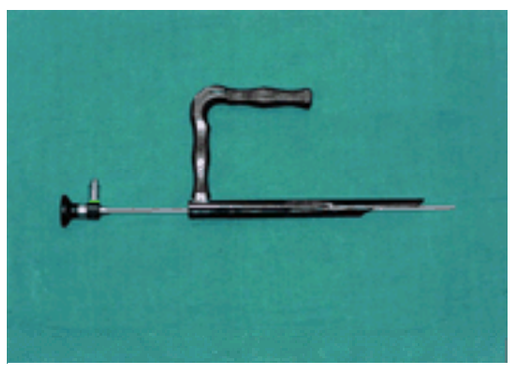


Fig. 8. The Negus pharyngoscope with a fibreoptic telescope in place, a good combination for patients in whom access is difficult.

Small oesophageal lumen

Occasionally it may be hard to identify the oesophageal lumen, which sometimes appears as a tiny pit amongst oedematous mucosa. In these circumstances, gentle dilatation with boogies will distend the lumen and eventually reveal a bar, which can be stapled.

Postoperative care

The patient should be kept ‘nil-by-mouth' for 4 h. If there are no untoward signs, they should be allowed to drink on the first evening and should resume a normal soft diet on the following day. Provided the patient's general condition permits, it is usual to discharge them on the first postoperative day.

If a tachycardia or significant neck or back pain develop, then a perforation should be suspected. If surgical emphysema is in the neck tissues, it means that a perforation has occurred.

Complications

Perforation

This should be an extremely rare occurrence with the endoscopic stapling technique. A small perforation could be treated conservatively, for example feeding via a nasogastric tube with a contrast study after 5 days. A large perforation requires exploration of the neck and either excision of the pouch or a repair of the oesophageal segment, whichever is appropriate.

Residual or recurrent symptoms

A proportion of patients (approximately 10 per cent) may have residual or recurrent symptoms. Investigation with a barium swallow is unhelpful because, in all patients who have had endoscopic stapling, a lax segment of pouch wall persists. Therefore, the patient should be judged solely on symptoms and, if significant symptoms persist, it is usually worth endoscoping them again. Under these circumstances a residual bar can usually be identified and stapled.

Previous pouch surgery

This is not a contraindication to endoscopic stapling. If a pouch has recurred after excision it has the same appearance as a new pouch and stapling should be straightforward.

Alternative endoscopic techniques

In experienced hands, division of the cricopharyngeal bar using a laser has produced good results with low morbidity. However, the laser does not seal the edges of the diverticulotomy and so perforation, and possible mediastinitis, are more likely to occur with laser division than with stapling.

At present, endoscopic pouch stapling is the method of choice for patients with a pharyngeal pouch. Pharyngeal pouch is a relatively uncommon condition and a good case can be made for subspecialization, which would allow a surgeon to operate on at least 10 pouch patients a year.

Further reading

Baldwin D, Toma A. Endoscopic stapled diverticulotomy: a real advance in the treatment of hypopharyngeal diverticulum. *Clinics in Otolaryngology* 1998; **23**(3): 244–7.

Koay C, Sharp H, Bates G. Current practice in the surgical management of pharyngeal pouch in England and Wales. *Annals of the Royal College of Surgeons of England* 1997; **79**: 190–4.

Royal College of Surgeons of England. *The report of the National Confidential Enquiry into Perioperative Deaths 1996/1997* (published 3 November 1998). Available from: Royal College of Surgeons of England, 35–43 Lincoln's Inn Fields, London WC2A 3PN.

Van Overbeek JJM. Meditation on the pathogenesis of hypopharyngeal (Zenker's) diverticulum and a report of endoscopic treatment in 545 patients. *Annals of Otology, Rhinology and Laryngology* 1994; **103**: 178–85.

45.14 Surgical technique of tracheostomy, minitracheostomy, and cricothyroidotomy

M. J. Burton

Tracheostomy
Emergency tracheostomy
Emergency cricothyroidotomy
Elective tracheostomy in adults
Elective tracheostomy in children
Postoperative care
Complications of tracheostomy
Percutaneous tracheostomy
Further reading

Tracheostomy

The technique of tracheostomy is slightly different when in an emergency or elective. In an emergency it is usually undertaken when endotracheal intubation is impossible or has failed. Speed is of the essence, the outcome is critical, and the procedure is not straightforward.

Emergency tracheostomy

The patient is supine, the head extended on the neck, and the neck flexed on the trunk by placing a sandbag under the shoulders (Fig. 1). A long, vertical incision is made in the midline from the thyroid notch to the suprasternal notch. It is vital to stay in the midline. The cricoid is identified by palpation below the thyroid cartilage and if possible a cricoid hook is passed under its lower border and lifted vertically upwards. The deeper soft tissues are incised, again via a midline vertical incision, from the lower border of the cricoid downwards. This will be bloody because the thyroid isthmus is often divided. The vertical incision is extended into the trachea (Fig. 2(b)). Tracheal dilators (or any other suitable ‘spreading' instrument) are put into the trachea. They are opened and a tracheostomy tube inserted. In an emergency, almost any other tube can be used. Once the airway has been secured in this way, haemostasis can also be secured and the patient anaesthetized if necessary. Not infrequently it is appropriate to perform a direct laryngoscopy to ascertain the cause of the upper airway obstruction necessitating the tracheostomy.

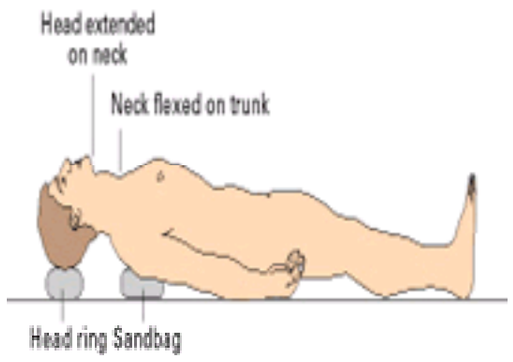


Fig. 1. Position for tracheostomy.

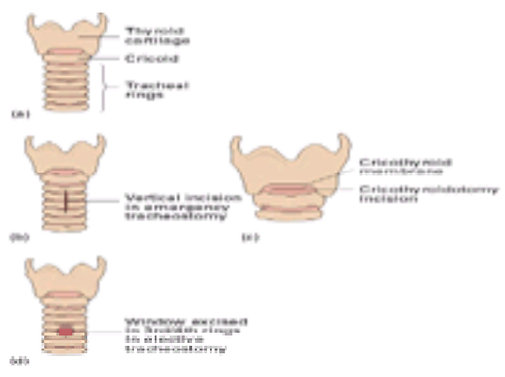


Fig. 2. (a) Anatomy of external laryngeal skeleton and trachea. (b) The vertical incision in emergency tracheostomy. (c) In a cricothyroidotomy procedure a horizontal incision is made in the cricothyroid membrane. (d) In elective tracheostomy a ‘window' is created in the anterior tracheal wall by excising part of the third and fourth tracheal rings.

Emergency cricothyroidotomy

Cricothyroidotomy is an alternative to emergency tracheostomy. The trachea is entered through the cricothyroid membrane, between the inferior border of the thyroid cartilage and the superior border of the cricoid. With the patient in the position described above, a vertical skin incision is made from the superior thyroid notch to a point below the cricoid cartilage. The cricothyroid membrane is palpated and a horizontal incision made next to the upper border of the cricoid cartilage (Fig. 2(c)). As before, a suitable instrument is used to spread open the opening into the trachea and a tube inserted.

Emergency minitracheostomy

Several commercially produced ‘kits' are available for performing a cricothyroidotomy. These comprise an instrument to incise or puncture the skin and a cannula with an introducer.

Elective tracheostomy in adults

An elective tracheostomy can be undertaken with the usual care and attention to detail given to other surgical procedures. However, some of the principles of emergency surgery, including the importance of positioning the patient correctly and staying in the midline, are still applicable. It is not a trivial procedure.

The patient is positioned supine on the operating table, the head extended on the neck, and the neck flexed on the trunk by placing a sandbag under the shoulders. A head ring is useful to keep the head stable, in particular preventing rotation away from the midline. Avoid trying to do this procedure on a bed outside the normal operating theatre environment if at all possible. The rigid surface of the operating table, the operating lighting, and all the other facilities present in a fully functional operating theatre make the procedure much easier and consequently safer for the patient. A headlight can be useful in a patient with a thick neck or when particular difficulties are anticipated.

The skin of the anterior neck is prepared from the undersurface of the chin to below the suprasternal notch. The area is draped with sterile drapes. If head towels are used they can be wrapped around the head in such a way that the mouth can be accessed easily by the anaesthetist. This is an important consideration. If the procedure is being undertaken under local anaesthesia, the patient's mouth is not covered and oxygen can be administered via a facemask. More usually, general

anaesthesia is used and this arrangement allows the endotracheal tube to be withdrawn in a controlled fashion later in the procedure.

Local anaesthetic with a vasoconstrictor [such as 1 per cent lignocaine (lidocaine) with 1 in 200 000 adrenaline (epinephrine)] is infiltrated into the operative field. The skin and deeper subcutaneous tissues are infiltrated. A small quantity may be injected into the lumen of the trachea to help prevent coughing when the lumen is entered.

A horizontal incision is made through the skin and superficial subcutaneous tissues midway between the thyroid prominence and the suprasternal notch. The skin edges are retracted and the remainder of the procedure continues in the midline. A vertical incision is made in the fascia between the strap muscles, which are retracted laterally. The cricoid cartilage and thyroid isthmus are identified; the isthmus usually covers the third to fifth tracheal rings. If the isthmus is anything other than extremely thin and mobile, it should be formally divided. To do this the fascia above it is divided and a tunnel developed between its back and the surface of the trachea. The isthmus is then clamped with large artery clamps, divided, and transfixed with a suture. The anterior wall of the trachea should now be clearly visible. A small swab may be used to clean away any remaining soft tissue.

At this point the situation is a stable one. Before proceeding with incision into the trachea it is important to ascertain that the appropriate tracheostomy tube is available, that the cuff will inflate and remain so, and in particular that the correct connectors are available to connect the tube to the apparatus being used by the anaesthetist. This is the surgeon's responsibility and should not be delegated. When the anaesthetist is ready, and whilst an assistant has suction and the tracheostomy tube at hand, the trachea is incised. Several options are available; the author prefers to excise a window of cartilage including the third and fourth tracheal rings. The first and second tracheal rings should not be touched. As the trachea is incised the cuff of the endotracheal tube may be punctured. If not, it is deflated and the tube withdrawn slowly until the tip passes the window in the trachea. As soon as this happens, the tracheostomy tube is inserted, the cuff inflated, a sterile connector attached and the end passed back under the head drapes to the anaesthetist. The surgeon does not leave go of the tracheostomy tube. Once it is clearly established that the patient is being ventilated satisfactorily the procedure may continue.

The skin edges are closed loosely with sutures. The closure should not be tight, otherwise surgical emphysema may result. It is imperative that the tracheostomy tube be well secured. Accidental decannulation may be fatal in patients with upper airway obstruction. This may seem less important in patients who are undergoing tracheostomy for long-term ventilation purposes, who could be reintubated with relative ease. Nonetheless, it is good practice to ensure that the tube is well secured in all patients. Sutures are placed through the skin and the body or flanges of the tube, as close as possible to the central axis of the lumen. Tapes are also passed round the patient's neck. These should only be tied when the neck is in the normal flexed position, otherwise they will prove to be too loose. To achieve this, the head must be lifted from the operating table and flexed on the neck whilst the tapes are tied.

Elective tracheostomy in children

There are a number of specific problems in undertaking tracheostomy in children and this is a specialist area even for the otolaryngologist. Anatomically there are important differences between the child and adult. In children, extension of the neck results in the great vessels (especially the left innominate vein), the thymus, and the apices of the lungs being pulled upwards into the neck. The trachea is softer and can be more difficult to palpate.

The child is placed in a similar position to that of the adult. A 1- to 2-cm horizontal skin incision is used. The soft tissues are divided in the standard fashion. Before incising the trachea, non-absorbable stay sutures are passed through the second and third or third and fourth tracheal rings, lateral to the midline. After the preparatory steps described above, a vertical incision is made in the trachea, between the sutures, and the tracheostomy tube inserted. The stay sutures remain in place postoperatively. If accidental decannulation occurs, the sutures are grasped and gentle traction will allow identification of the tracheal lumen and facilitate recannulation.

In children a postoperative chest radiograph is undertaken to establish the position of the tip of the tracheostomy tube.

Postoperative care

Appropriate postoperative care is essential for both children and adults. The patient should be nursed by staff who are experienced in the management of tracheostomies. Humidification is important, since the normal humidifying functions of the nose have been bypassed. Regular suction is necessary to clear secretions. Care should be taken to prevent repeated trauma to the carina. Equally, the diameter of the suction catheter is important. This should not exceed half the internal diameter of the tracheostomy tube or pulmonary collapse may ensue. The first tube change should be undertaken by the surgeon at 1 week.

Complications of tracheostomy

The early complications of tracheostomy include bleeding, pneumothorax, surgical emphysema, and infection. The most important, however, are obstruction of the tube—by secretions or, in children, by the soft tissues of the chin and neck—and accidental decannulation. Facilities for recannulation and intubation should be available at the bedside at all times.

Obstruction and accidental decannulation are also potential long-term complications. Additional problems include granuloma formation (at the tip of the tube or around the tracheostome), bleeding from erosion of a vessel due to pressure necrosis, tracheal stenosis, or tracheo-oesophageal fistula.

Percutaneous tracheostomy

Percutaneous tracheostomy has become increasingly popular in the last few years. It can be undertaken at the bedside in the intensive care unit and is frequently performed by intensive care staff rather than the surgical team. Several commercially produced kits are available. Local anaesthetic is injected below the level of the cricoid cartilage and into the deeper tissues. An inner-needle cannula with an outer sheath is inserted into the trachea through the skin. The inner needle is removed and a guidewire passed through the outer sheath. The sheath is removed. A guiding catheter is then placed into the trachea and successively larger dilators passed over the catheter. Once the track has been dilated to a suitable size, a tracheostomy tube is advanced over the dilator and the dilator and guidewire removed.

Complications of this procedure include the development of false passages outside the trachea, puncture of the posterolateral tracheal wall, and bleeding. The long-term complications are uncertain.

Further reading

Freeland AP. Tracheostomy, cricothyroidotomy and percutaneous tracheostomy. In: Garrard C, Foëx P, Westaby S, ed. *Principles and practice of critical care*, pp. 734–47. Blackwell Science, Oxford, 1997.

46.1 Sports medicine and athletic injuries

Eric W. Carson and Brant Bair

The knee
Anatomy of the knee
Physical examination of the knee
Imaging studies of the knee
Injuries about the knee and their treatment
The shoulder
Anatomy of the shoulder
Physical examination of the shoulder
Imaging studies of the shoulder
Injuries about the shoulder and their treatment
Elbow
Anatomy
Physical examination
Radiographic evaluation
Ligament injuries
Articular injuries
Leg
Compartment syndrome
Stress fractures
Medial tibial stress SYndrome
Ankle injuries
Ankle sprains
Foot injuries
Tendinitis
Turf toe
Hand and wrist injuries
Wrist fractures
Wrist ligamentous injuries
Wrist tendinitis
Finger injuries
Further reading

The knee

Anatomy of the knee

Arthrology

The ‘knee joint' comprises three bones—the tibia, femur, and patella. These three bones create three primary articulations—tibiofemoral, tibiofibular, and patellofemoral joints. The tibiofemoral joint is the primary weight-bearing joint of the knee. It consists of a medial and lateral joint space corresponding to the tibial plateau and femoral condyles (Fig. 1). There is little inherent stability of the knee secondary to the relatively incongruous bony articulation. The menisci improve the congruency of the knee joint. The menisci also function in the dispersion of forces across the knee. The medial meniscus is c-shaped and is tightly attached at its periphery to the medial joint capsule, also known as the deep medial collateral ligament. The lateral meniscus is more circular in shape. The lateral meniscus is not attached to the lateral collateral ligament but is loosely anchored to the joint capsule (Fig. 2).

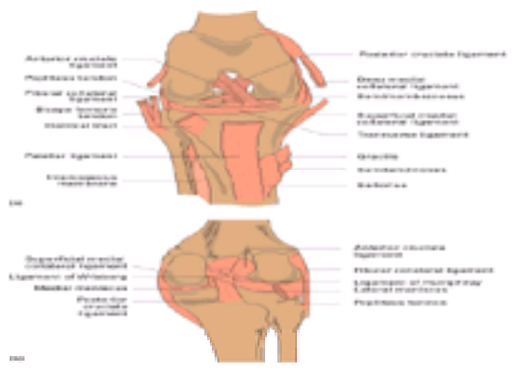


Fig. 1. Knee joint. (a) Anterior, (b) posterior views.

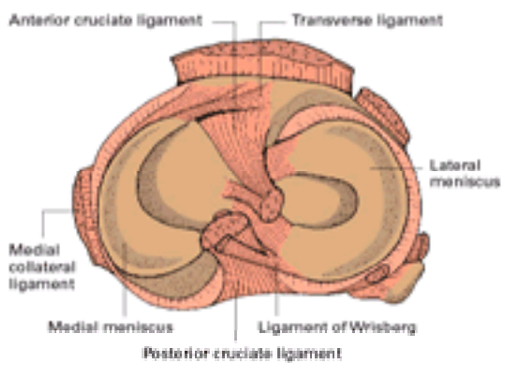


Fig. 2. Medial and lateral meniscus.

The patellofemoral joint is formed by the patella gliding in the trochlear groove of the distal femur. The patellofemoral joint is important because it increases the mechanical advantage of the quadriceps musculotendinous unit. Significant forces are applied through the patellofemoral joint, especially during activities such as ascending or descending stairs. The stability of this joint is determined by the depth of the trochlear groove and the balanced muscle pull of the quadriceps and hamstring tendons.

Ligamentous structures

As mentioned above, the knee has little inherent stability owing to the lack of congruency of the articular surfaces. The knee relies heavily on the ligamentous structures to maintain stability. The four main ligaments of the knee are the anterior cruciate, posterior cruciate, medial collateral, and lateral collateral ligaments.

The anterior cruciate ligament has its femoral attachment on the medial/posterior aspect of the lateral femoral condyle in the intercondylar notch. The path of this

ligament is anterior and medial as it passes from proximal to distal and inserts on the tibial plateau lateral to the medial tibial spine. The anterior cruciate ligament is the primary restraint against anterior translation of the tibia with respect to the femur. The ligament is composed of two bundles: anteromedial and posterolateral. The importance of this ‘two bundle' system is that different portions of the ligament are taut depending on the position of the knee. Therefore at any given position, one portion of the anterior cruciate ligament is under tension.

The posterior cruciate ligament is attached to the anterior/lateral aspect of the medial femoral condyle within the intercondylar notch. Tibial attachment of this ligament is posterior and approximately 1 cm inferior to the level of the tibial plateau. The primary function of the ligament is to resist the posterior translation of the tibia with respect to the femur. The ligament is also composed of a two-bundle system: the larger (95 per cent) anteriolateral and the smaller posteriomedial. The majority of the posterior cruciate ligament fibers are taut with the knee in a flexed position.

The medial collateral ligament (MCL), also known as the tibial collateral ligament, is the primary restraint to valgus stress on the knee. This ligament is composed of a superficial and deep component. The superficial ligament is a broad structure that originates on the medial femoral condyle, passes distally along the joint line, and inserts 4 to 5 cm distally below the joint line on the tibial metaphysis. The deep ligament is a consolidation of the capsule and well attached to the medial meniscus. The superficial ligament is the primary restraint to valgus stress with the knee in 30 to 40° of flexion. A secondary function of the medial collateral ligament is resistance to external rotation of the tibia on the femur.

The lateral collateral ligament (LCL) or fibular collateral ligament (much smaller than MCL) is part of the posterolateral complex (posterolateral capsule and popliteus) that provides the primary resistance to varus force on the knee. The origin of the this ligament is on the lateral femoral condyle and the insertion is on the proximal and lateral portion of the fibular head.

Musculotendinous structures

Many large muscles cross the knee joint to provide power, mobility, and stability to the knee. The quadriceps mechanism lies on the anterior thigh and functions primarily to extend the knee. It consists of four individual muscle units—the rectus femoris, vastus medialis, vastus lateralis, and vastus intermedius. Of these four muscles, the rectus femoris is the only one to cross both the hip and the knee. The rectus originates from the anterior inferior iliac spine and inserts distally with the remainder of the quadriceps. Because it crosses the hip it also functions as a hip flexor. The remainder of the quadriceps (vastus medialis, vastus lateralis, and vastus intermedius) originates from the anterior half of the femur and courses distally to insert into the patella. The vastus medialis and lateralis additionally function to stabilize the patella from medial and lateral stresses. The extensor mechanism is completed with the patellar tendon, which extends from the inferior pole of the patella to the tibial tubercle.

The hamstrings consist of a muscle mass that primarily functions to flex the knee but also functions to extend the hip. The individual muscle units that make up the hamstrings are the semitendinosus, semimembranosus, and biceps femoris. The origin of the hamstrings is the ischium and the insertion is individual. The biceps femoris lies laterally and inserts on the fibular head above the LCL. The semitendinosus inserts on the anterior medial tibia with the sartorius and gracilis, forming the pes anserine. The semimembranosus inserts on the posterior aspect of the tibia primarily, but does have numerous soft-tissue insertions in the posterior aspect of the knee.

Physical examination of the knee

History

Evaluation of knee injuries must first start with the physician obtaining a detailed and accurate history. In most cases, the history will guide the physician to the correct diagnosis. Knee injuries can occur with a variety of mechanisms. Key questions that the physician must ask include: What was the mechanism of injury? Did the symptoms develop over time or was there a specific event leading to the symptoms? What position was the leg in at the time of the injury? Was it a contact or non-contact injury? Is there associated swelling, localized pain, popping, locking, or giving away? What activities make the symptoms worse? Is there any previous history of injury? When an accurate history combined with a comprehensive physical examination is obtained, the majority of knee injuries can be properly diagnosed.

Inspection

The examination of the knee begins by observing the patient walk into the examination room, taking careful note of gait and presence of a limp. The knees should be inspected visually; comparing the two sides for differences in alignment, soft-tissue swelling, and atrophy. Effusion of the knee can be estimated by ballotting the patella and noting the size of the suprapatellar pouch when the knee is in full extension. Bursal swelling will manifest as a more local swelling in the area of the involved bursa.

The entire knee and its surrounding structures should then be palpated systematically. Areas to palpate include the medial and lateral joint lines, tibial tubercle, patella, patellar tendon, medial and lateral tibial plateaus, medial and lateral femoral condyles, quadriceps tendon, hamstring tendons, iliotibial band, medial and lateral collateral ligaments, and popliteal fossa.

As with all joint examinations it is important to evaluate and document the active and passive range of motion of the knee. The primary range of motion of the knee is in a flexion/extension plane; however, there is a small degree of internal and external rotation. Normal range of motion is from 0° of extension to 135° of flexion. About 10° of internal and external rotation are also present.

The strength of the quadriceps and hamstrings should be tested and scored using the standard 0 to 5 rating scale. The lower extremity reflexes, sensation, and vascular status should also be examined and documented (see Table 1).

Grade	*Degree of muscle strength*
0 (Zero)	No palpable contraction
1 (Trace)	Muscle contracts, no movement of extremity even without gravity
2 (Poor)	Muscle moves the extremity but not against gravity
3 (Fair)	Muscle moves the extremity through a range against gravity
4 (Good)	Muscle moves extremity even against resistance
5 (Excellent)	Normal strength against full resistance

Table 1 Grading of muscle strength

After completing the above basic examination, there are numerous special and provocative tests that can be used for better evaluation of specific structures. Varus and valgus stress should be applied to the knee at 0° and 30° of extension (Fig. 3). These two examinations evaluate the lateral and medial collateral ligaments, respectively. In full extension, it should be remembered that the anterior and posterior cruciate ligaments contribute to the varus/valgus stability.

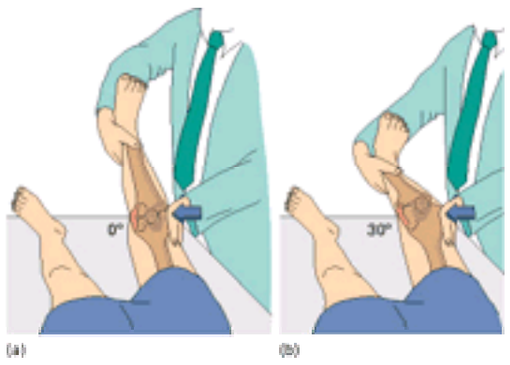


Fig. 3. Valgus stress examination of the medial collateral ligament (0°/30°).

Examination of the anterior cruciate ligament (ACL) can be performed in a numbers of different ways. The most reliable and sensitive of the various methods is the Lachman test. The Lachman test is performed by stabilizing the femur with one hand and attempting to translate the tibia anteriorly with the other hand, with the knee in about 30° of flexion (Fig. 4). A firm endpoint should be appreciated with an intact anterior cruciate ligament. The side-to-side difference in the amount of translation also helps to decide the integrity of this ligament. Other tests for the ligament include the anterior drawer (Fig. 5), which is performed with the knee at 90° and trying to translate the tibia anteriorly. This test is less specific than the Lachman examination.

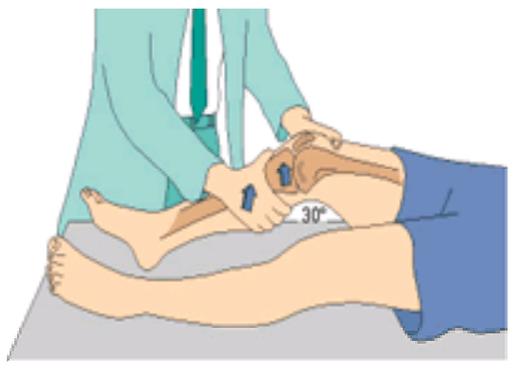


Fig. 4. Lachman test.



Fig. 5. Anterior drawer test.

Examination of the posterior cruciate ligament is slightly more difficult than the ACL. The most commonly utilized examination is the posterior drawer test (Fig. 6). This is essentially the opposite of the anterior drawer test. An important point about this test, however, is to note the starting position of the tibia. With the knee flexed to 90°, the tibia should be slightly anterior to the femoral condyle. If it is not, this represents a ‘posterior sag' and a posterior cruciate ligament disruption. This posterior sag can be accentuated by flexing both the hip and knee to 90°.



Fig. 6. Posterior drawer test.

Numerous maneuvers have been described to elicit pain due to tears of the medial or lateral meniscus. The McMurray test is performed by applying first valgus and internal rotation stress to the knee while flexing and extending the knee, then switching to varus and external rotation forces through the same range of motion (Fig. 7). Apley's grind test is performed with the patient in a prone position, the knee flexed 90°, and a downward load supplied to the foot. Other tests including hyperextension, ‘bounce home', and hyperflexion of the knee are also used in the assessment of meniscal pathology. These tests are considered positive if pain is reproduced with the specific maneuver.

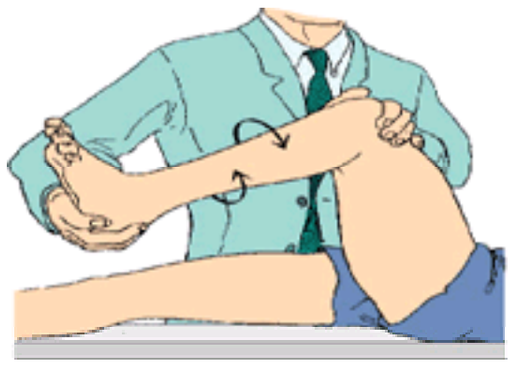


Fig. 7. McMurray test.

The patellofemoral joint must be examined for evidence of instability or degeneration. The patella apprehension test is performed with the patient supine and the knee fully extended (Fig. 8). A laterally directed force is then applied to the patella. If the patella begins to sublux the patient will experience pain and anxiety. The patellofemoral grind test should also be performed. This is carried out in the same position as the apprehension test. A downward force is applied to the patella and then the patient is asked to contract the quadriceps actively. Pain will be elicited if there are any degenerative changes on the undersurface of the patella.



Fig. 8. Patella apprehension test.

Imaging studies of the knee

Plain radiographs

Plain radiographs are a standard examination of an acute knee injury after the history and physical examination. The bony anatomy of the knee can be evaluated initially with four views of the knee: standing anterioposterior and lateral, merchant patellar, and notch views. The anterioposterior and lateral views are important for detecting subtle or gross fractures of the femoral condyles or tibial plateau. The merchant patellar view is important for evaluation of the patellofemoral articulation. The presence of fractures or patella tilt can be appreciated. The notch view is important for evaluating the morphology of the notch, presence of any loose bodies, osteochondral lesions, or avulsions of the insertion of the anterior cruciate ligament on the tibial plateau.

Arthrogram

Arthrography is performed by injecting radio-opaque contrast into the knee and taking plain radiographs. This technique was used to help diagnose meniscal tears, articular cartilage abnormalities, and popliteal cysts; however, it has fallen out of favor because it is invasive and not as accurate as an MRI of the knee.

Computed tomography

CT has a limited but important role in evaluation of acute knee injuries. It is important in defining the bony anatomy. This is especially important in the case of a tibial plateau fracture in which the fracture planes are difficult to appreciate on plain radiographs.

Nuclear imaging

Bone scans also have a limited but important role in the evaluation of sports injuries. They are useful for the evaluation of occult stress fractures about the lower extremity.

Magnetic resonance imaging

MRI became widely available in the 1980s and has significantly enhanced the evaluation of the patient with a suspected knee injury. It has allowed more accurate diagnosis of injuries and injury patterns which previously were underappreciated conditions. The advantages of MRI are numerous and include exquisite anatomic detail of the meniscus, articular cartilage, ligaments, and muscles utilizing multiplanar imaging and lack of ionizing radiation. As good as MRI can be, it is not always necessary or accurate. False-positive results do occur. In fact an MRI should only confirm what is expected based on a thorough history and physical. For this reason it is felt that the MRI is often best ordered in those patients with multiple ligament injury or those patients in whom the diagnosis is unclear.

Injuries about the knee and their treatment

Fractures

Fractures about the knee can occur with a variety of mechanisms including valgus or varus stress, axial load, a direct blow, or a combination of forces. Patella fractures are usually the result of a direct blow to the anterior aspect of the knee. Diagnosis is based on presence of tenderness over the patella, an associated effusion, and loss of active knee extension. Radiographs will show a fracture line in the patella, usually in the transverse plane. Non-displaced fractures can be treated in a cast with the knee in full extension. However, displaced fractures require open reduction and internal fixation.

Fractures of the tibial plateau occur during sporting events. The lateral plateau is more commonly injured than the medial tibial plateau. Tibial plateau fractures are generally high-energy injuries and suspicion must be high for associated injuries such as ligament disruptions, knee dislocations, and vascular injuries. Fractures of the tibial plateau are generally treated operatively. Various methods of fixation, including open reduction and internal or external fixation, are available based on the fracture pattern.

Dislocations

Dislocations about the knee are of two types: tibiofemoral or patellofemoral dislocations. Traumatic dislocation of the knee (tibiofemoral joint) is uncommon; however, it is a limb-threatening orthopedic emergency. The classification of knee dislocations is based on the relationship of the tibia to the femur. The traumatic knee dislocation is often misdiagnosed or underdiagnosed for a couple of reasons. First, the knee may spontaneously reduce at the scene because of the severe soft-tissue injury. Second, there may be other injuries that are distracting attention from the knee. The most important part of the assessment is the documentation of a complete neurovascular examination of the involved limb prior to any reduction attempts. This neurovascular examination should be repeated and documented following the reduction of the knee dislocation. Popliteal artery injury is present in 30 per cent of all knee dislocations. It is important to remember that the collateral circulation is

insufficient to sustain limb viability and the presence of pedal pulses does not eliminate the possibility of vascular injury. The knee dislocation must be reduced promptly, the soft tissues stabilized, and the ligaments reconstructed on an elective basis.

Patella dislocation or subluxation is a common problem in the young athlete. Usually it is a non-contact injury that occurs as the patient has the affected foot planted and cuts to the opposite direction. It is often associated with an audible ‘pop' and a sensation of giving way. There are occasionally two pops, one with the dislocation and the other with the spontaneous reduction. An effusion usually occurs within 12 h and resolves over the course of a week. Physical examination findings associated with patella instability include: quadriceps atrophy, patella alta (high-riding patella), lateral patellar tilt, and excessive valgus alignment (increased Q angle). Other injuries may be associated with a patella dislocation, such as an osteochondral fracture of the femoral trochlea or patella. The first line of treatment is physical therapy directed at strengthening the quadriceps, especially the vastus medialis obliquus. Patella stabilizing braces may also be used to help stabilize the patella. The exercises are more important than the braces, but they are usually used concomitantly. Occasionally, the patient who experiences chronic patellar dislocations will require operative reconstruction. The procedures are aimed at realigning the mechanical forces that are pulling the patella laterally.

Ligamentous injuries

The medial collateral ligament is injured when the mechanism of injury is a force applied to the lateral or posterolateral side of the knee producing a valgus stress. This ligament is extracapsular and should have localized swelling and tenderness. However, if there is a large joint effusion, this implies there are other injuries to the intracapsular structures of the knee. An isolated injury of the medial collateral ligament will demonstrate laxity to valgus stress with the knee flexed 30°. The injuries are graded by the amount of instability: grade I is consistent with a sprain and minimal laxity, grade II is a torn ligament with laxity of less than 5 mm, and grade III is a complete tear of the ligament with no endpoint and laxity of greater than 5 mm. The treatment of an isolated grade I or II injury traditionally has been non-operative. The patient is treated with a hinged knee brace and protective weight-bearing for 4 to 6 weeks. Treatment of grade III injury has been controversial. The current trend is for non-operative treatment. Those patients who develop chronic medial instability or have associated ligament injuries are indicated for surgical repair.

The lateral collateral ligament is ruptured by a combination of hyperextension and varus force. Many of these injuries occur in association with injuries of the posterolateral complex, popliteus, posterolateral capsule, or biceps femoris. The patient may hear a pop but have little swelling after the injury. The integrity of the ligament is examined by applying a varus force with the knee flexed 30°. The treatment of isolated injuries of the lateral collateral ligament is similar to those of the medial collateral ligament; grade I or II injuries are treated conservatively, while grade III require surgical intervention. High-performance athletes and those patients with a significant varus alignment of the lower extremities may benefit from an operative repair.

The anterior cruciate ligament is a commonly injured knee ligament. An isolated tear of this ligament is usually the result of a non-contact twisting injury. The most common mechanism is deceleration, valgus, and external rotation. The patient may hear a pop, feel the knee sublux, and develop an effusion within 1 to 2 h. A clinician who obtains a good history and performs a skilled physical examination can make the correct diagnosis. Plain radiographs are part of the initial examination to rule out a subtle fracture. In most cases MRI is not part of the routine examination. However, MRI is useful in some patients with a tear of the anterior cruciate ligament in whom other injuries such as osteochondral fractures, meniscal tears, or multiple ligament tears are suspected. Initial treatment includes ice and a short period of immobilization, followed by a physical rehabilitation program which focuses on maintaining quadriceps strength. Not all patients with disruption of the anterior cruciate ligament will have functional instability. The long-term sequelae and natural history of knees with deficiency of this ligament are unpredictable. Data do suggest that there is an increased icidence of menical teas following ligament deficiency. However, the development of osteoarthritis secondary to instability has not been proved as once thought. The choice of whether to reconstruct the ligament or not must be based on the individual patient according to age, activity level, secondary ligamentous restraints, and meniscus status. Current indications for surgery include a young healthy patient, a multiple ligament injury, and instability with activities of daily living or sports. The anterior cruciate ligament is commonly reconstructed arthroscopically using the central third of the patella tendon (Fig. 9) with a bony block from the patella on one end and a bony block from the tibial tubercle or semitendinosis and gracilus hamstring tendons. The patients undergo an extensive physical therapy program with focus on regaining normal range of motion, quadriceps strength, and the ability to return to sports. The return to sports usually requires 6 to 9 months of therapy.

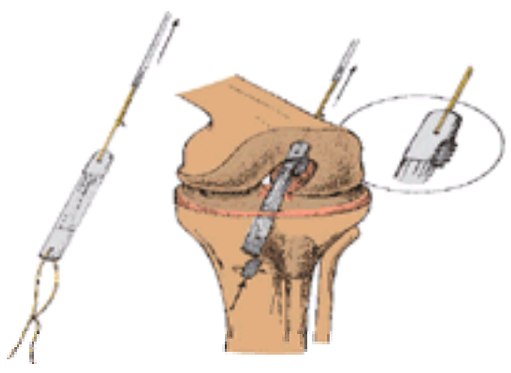


Fig. 9. Technique demonstrating endoscopic reconstruction of the anterior cruciate ligament.

Injury of the posterior cruciate ligament is a lot less common than of its anterior counterpart. Most injuries of the posterior cruciate ligament are not isolated and are commonly associated with multiple ligament disruptions. The mechanism of injury is a flexed knee sustaining an anterior force to the anterior tibia, ‘dashboard' injury, or a fall onto a hyperextended knee. The typical presentation is that of pain, particularly patellofemoral, not instability. On physical examination the patient will demonstrate posterior sag or a posterior drawer sign. Careful examination of the neurovascular status must be completed and documented, as other soft-tissue injuries are associated with tears of this ligament. A knee without the posterior cruciate ligament will have altered biomechanics. The tibia will tend to sublux posteriorly with respect to the femur, thus producing increased forces on the patellofemoral joint. Non-operative treatment of these tears has long been the mainstay; however, newer reconstruction techniques assisted by the arthroscope have been used for those patients with significant instability.

Meniscal injuries

Tears of the medial and lateral menisci are very common sports injuries (Fig. 10). The most common mechanism is a twisting non-contact movement that produces a sheer stress across the meniscus. The patient may complain of intermittent swelling, locking, popping, or giving way. Physical diagnosis is complemented by MRI in cases where there is any doubt. MRI has a sensitivity in most studies well above 90 per cent and as high as 98 per cent.

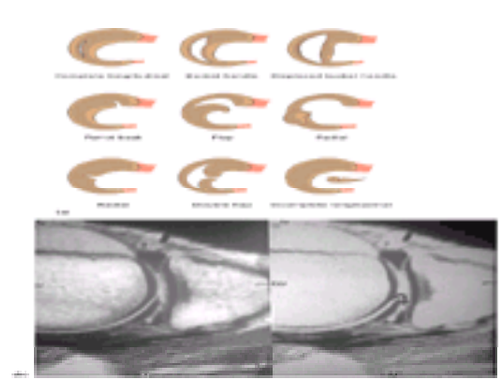


Fig. 10. (a) Classification of meniscal tears. (b) MRI of medial meniscal knee tear.

Non-operative management is indicated for those patients with symptoms that are consistent with a meniscal tear but are improving or do not have functional limitations. A program of rehabilitation exercises for the quadriceps and hamstrings as well as non-steroidal anti-inflammatory drugs may help alleviate the patient's symptoms. If the symptoms persist and are interfering with activities, operative intervention may be indicated. Arthroscopy is the mainstay for treatment of meniscal pathology. During arthroscopy several patterns of meniscal tears may be encountered. It must be decided whether the tear is repairable or requires partial menisectomy. In order to be repairable the tear must occur in the peripheral third of the meniscus. This is the vascular area of the meniscus near the meniscosynovial

junction that will allow healing to take place. A variety of techniques for meniscal repair have been described including: open, arthroscopic outside-in, inside-out, or all-inside techniques.

Those meniscal tears that are not amenable to repair require debridement and partial menisectomy. The goal of the partial menisectomy is to remove the torn unstable portion of the meniscus that is creating the mechanical symptoms while leaving behind a stable contoured rim of meniscus.

In the past, the options for the patient who sustains a meniscal injury that results in a total or near total menisectomy were limited. However, now meniscal allograft transplants are being performed. The early results of this technique have been promising in those patient with minimal degenerative changes. The meniscal allograft is able to transmit the forces across the knee joint and delay further degenerative changes.

Osteochondral injuries

Osteochondral injuries, known as osteochondritis dissecans, most commonly occur on the articular surface of the femoral condyle. Considerable debate exists about the etiology of osteochondritis dissecans. Proposed theories include: ischemia, trauma, and pre-existing bony abnormality. Regardless of the etiology, the result is that the subchondral bone undergoes avascular necrosis and the overlying articular cartilage delaminates from the subchondral bone. This results in either a chondral flap or a free-floating loose body in the knee. The patient will complain of swelling, pain, catching, or giving way. The majority of osteochondritis dissecans lesions of the knee occur on the lateral aspect of the medial femoral condyle involving the weight-bearing aspect. Plain radiographs most often confirm diagnosis. MRI and arthroscopy aid in the status of osteochondritis dissecans, as to whether the lesion is stable or unstable (Fig. 11).

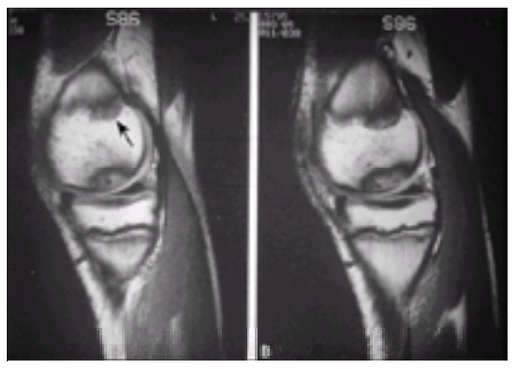


Fig. 11. Osteochondritis dissecans of medial femoral condyle.

The goal of treatment of osteochondritis dissecans is prevention of detachment, encouragement of healing, and prevention of further degeneration of the articular surface. Initial treatment includes brief immobilization followed by modification of activity. If the osteochondritis dissecans is detached, the options include excision or repair. The free fragment may be repaired if it is large enough to hold fixation and has subchondral bone attached to it to support bony healing. However, if the osteochondral fragment is small or has no bony attachment, fixation will be difficult and most likely unsuccessful. In this case excision is indicated. If excision is performed, the base of the lesion can be drilled to stimulate fibrocartilage healing.

The natural history of a large, non-repairable, osteochondral defect has been progressive degenerative changes of the articular surface. For this reason new techniques are being developed for the treatment of these lesions, especially on the weight-bearing surface. One of the new techniques is autogenous chondrocyte transplantation. In this technique, cells from the calcified matrix are harvested, grown in culture, and then replanted onto the articular surface under a periosteal covering. The cells then proliferate and form ‘new' articular cartilage. The second of the new techniques is osteochondral autograft transplantation. In this technique a plug of articular cartilage and subchondral bone is taken from a non-articular portion of the femoral condyle such as the notch or lateral trochlea and then transplanted into the lesion. The long-term efficacy of these two techniques is unknown, but there have been promising early results in this otherwise difficult problem.

Soft-tissue injuries about the knee

Soft-tissue injuries about the knee are common in all levels of athletic participation. A bursa is a flat fluid-filled sac that forms from synovial tissue in response to friction. There are several bursas about the knee, namely the pes anserinus, prepatellar, infrapatellar, and semimembranosus, that may become inflamed with overuse injuries. As the bursa becomes inflamed, a process known as bursitis, the walls become thickened and the sac becomes filled with fibrinous exudate. The patients present with localized swelling and pain. Generally, regardless of location, the treatment of bursitis is rest, compression, ice, and anti-inflammatory medication. Occasionally, the swollen bursa may become secondarily infected, usually by a staphylococcus species. This requires antibiotic therapy, aspiration, and occasionally operative debridement.

Tendinitis is an inflammatory condition of the tendon sheath and the contained tendon. It develops as a physiologic response to repetitive loading. Any of the musculotendinous structures about the knee may be involved; however, some of the most common are the semimembranosus, pes anerine tendons, popliteus, and the patellar tendon. Patellar tendinitis, also known as jumper's knee, is characterized by pain localized at the inferior pole of the patella associated with activities. The patient usually presents with localized swelling. The treatment of patellar tendinitis is generally non-operative. The regimen includes rest, anti-inflammatory medication, adequate warm-up and stretching, isometric strengthening, and occasionally a patella restraining brace. If the condition is allowed to progress untreated, frank rupture of the patella tendon may occur secondarily to degeneration of the tendon. Treatment will require surgical debridement of the degenerative tissue and repair of the remaining tendon.

Soft-tissue structures about the knee may also rub across a bony prominence and cause a friction syndrome. The common soft tissues to be involved in this syndrome are the iliotibial band and a mediopatellar plica. The iliotibial band is a tendinous continuation of the fascia that envelops the gluteus maximus and tensor fasciae lata. With knee flexion and extension, the iliotibial band glides anteriorly and posteriorly over the lateral femoral condyle. Iliotibial band syndrome is a common injury in long-distance runners. The continuous friction may result in inflammation and the onset of lateral knee pain. Physical examination with the patient in a lateral position and the hip abducted and extended causes pain to be reproduced. This provocative test is referred to as Ober's test (Fig. 12). Treatment is conservative and is aimed at reducing the friction. The usual rest, ice, and anti-inflammatory medication are utilized as well as a vigorous stretching program aimed at the iliotibial band. Injection of corticosteroids can be of benefit to those patients who are slow to improve with physical therapy. Operative treatment is rarely indicated.

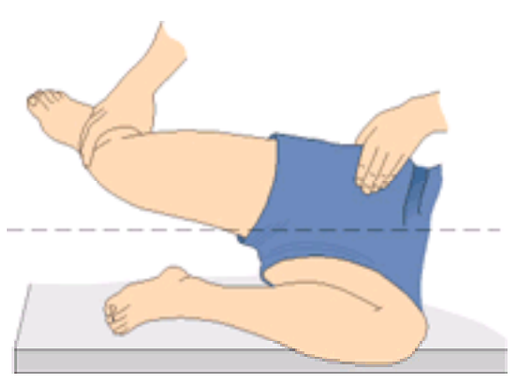


Fig. 12. Ober's test for iliotibial band syndrome.

A plica represents an embryologic remnant of the synovial divisions of the knee (Fig. 13). The most consistently present and occasionally pathologic is the medial patellar plica. This plica extends from the medial suprapatellar pouch to the synovium of the infrapatellar fat pad. The plica can extend across and impinge on the articular cartilage of the medial femoral condyle. These patients present with complaints of pain in the region of the medial femoral condyle. Injury can be the result of a helmet striking the medial aspect of the knee or overuse as seen in cyclists. Initial treatment is non-operative, consisting of activity modification and anti-inflammatory

medication. Should the pain continue, arthroscopic resection of the plica may be indicated.

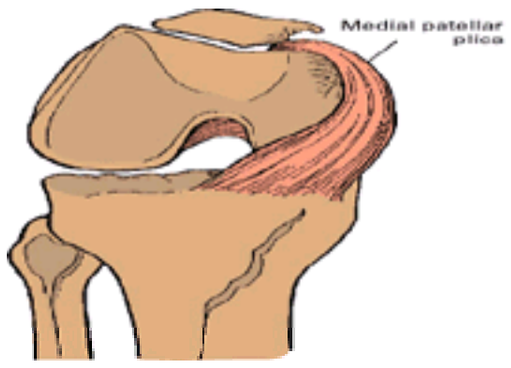


Fig. 13. Knee medial plica.

The shoulder

Anatomy of the shoulder

Arthrology

The shoulder is a complex structure that involves four ‘joints'—the glenohumeral joint, acromioclavicular joint, sternoclavicular joint, and the scapulothoracic articulation. The bones of the shoulder joint complex are the humerus, scapula, clavicle, and sternum (Fig. 14).

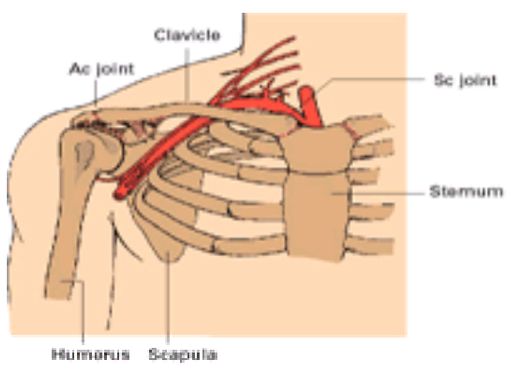


Fig. 14. Shoulder articulation.

The clavicle is a long tubular bone that connects the appendicular skeleton of the upper extremity to the axial skeleton. It is a subcutaneous structure and easily palpable. The medial end of the clavicle articulates with the sternum and forms the sternoclavicular joint. This synovial joint has little inherent bony stability. The stability of the subcutaneous joint is provided by the sternoclavicular ligaments, joint capsule, and the articular disk. The lateral end of the clavicle articulates with the acromion and forms the acromioclavicular joint. This joint also has little inherent bony stability and relies on the ligamentous structures to provide stability.

The scapula is a large flat bone that is an integral component of the shoulder girdle complex. The body of the scapula serves as the origin for the rotator cuff muscles. The acromion is the anterior continuation of the spine of the scapula, which serves as the ‘roof' of the shoulder and articulates with the clavicle. The coracoid is an anterior projection of the scapula. The conjoined tendon (coracobrachialis and short head of the biceps) attaches at the coracoid base. The glenoid is the lateral projection of the scapula, which articulates with the humeral head. The labrum is a continuation of the capsular tissue that is firmly attached to the rim of the glenoid which deepens and increases the congruity of the glenohumeral joint. The proximal humerus is the remaining bony structure of the shoulder joint complex. The relatively spherical humeral head articulates with the shallow glenoid fossa. Lacking bony constraints, the glenohumeral joint has multiple degrees of freedom, allowing a tremendous range of motion. The proximal humerus is composed of two tuberosities, greater and lesser, which are separated by the bicipital groove, site of the long head of the biceps tendon. The tuberosities function as insertion points for the rotator cuff musculature. The supraspinatus, infraspinatus, and teres minor insert on the greater tuberosity from superior to inferior, respectively. The subscapularis inserts on the lesser tuberosity anteriorly.

Ligamentous structures

The shoulder relies heavily on ligamentous structures and rotator cuff musculature to maintain stability and joint congruity. The acromioclavicular joint is formed at the lateral end of the clavicle and the acromion. The stability of this joint is enhanced by an articular disk, the surrounding joint capsule, and ligamentous structures including the acromioclavicular and the coracoclavicular ligaments. The acromioclavicular ligament resists posterior translation of the clavicle. It is relatively weak when compared with the coracoclavicular ligament. The coracoclavicular ligament is formed by two individual bands—the conoid ligament and the trapezoid ligament. These ligaments are important in resisting vertical translation of the clavicle. The coracoacromial ligament is a broad taut ligament, which courses from the coracoid to the underside of the acromion with a broad expansion. The coracoacromial ligament functions to depress the humeral head (Fig. 15).

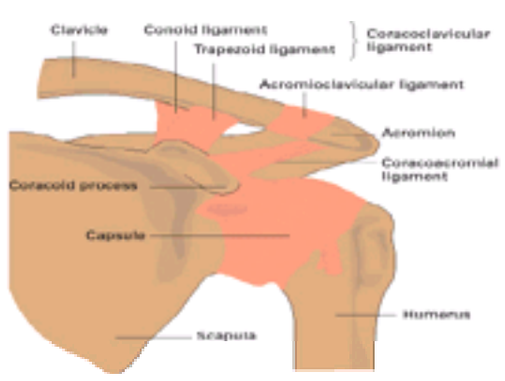


Fig. 15. Acromial ligaments (CA, coracoacromial ligament; CC, coroacoclavicular ligament— trapezoid, conoid; AC, acromioclavicular ligament, superior/inferior).

The glenohumeral joint also relies on ligamentous structures to help maintain static stability. The labrum, as mentioned above, increases the articular congruity of the glenohumeral joint. Three glenohumeral ligaments provide static stability; these are the superior glenohumeral ligament, the middle glenohumeral ligament, and the inferior glenohumeral ligament composed of anterior and posterior bands (Fig. 16). The most important of these three ligaments is the inferior glenohumeral ligament. This ligament, specifically the anterior band, provides stability to the abducted and externally rotated arm and provides the main resistance to anterior translation in the position of throwing.

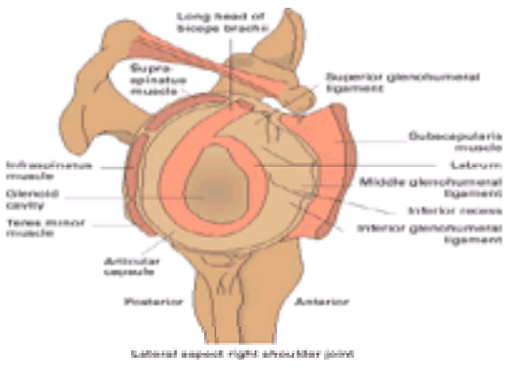


Fig. 16. Glenohumeral ligaments (GHL) (IGHL, anterior/posterior inferior GHL; MGHL, middle GHL; SGHL, superior GHL).

Musculotendinous structures

The muscles surrounding the shoulder joint are powerful and provide for a tremendous range of motion. They also provide dynamic stability for the glenohumeral joint. Two muscles attach the axial skeleton to the humerus—the latissimus dorsi and the pectoralis major. The latissimus dorsi originates from the vertebral column and inserts on the floor of the bicipital groove. It is innervated by the thoracodorsal nerve and functions to adduct, internally rotate, and extend the shoulder. The pectoralis major originates from the clavicle, sternum, and upper thorax and inserts on the lateral rim of the bicipital groove. It is innervated by the medial and lateral pectoral nerves and functions to adduct and internally rotate the shoulder.

Another group of muscles attaching the scapula to the humerus include the deltoid, teres major, and the rotator cuff musculature (supraspinatus, infraspinatus, teres minor, and subscapularis). The deltoid is composed of three heads that originate from the clavicle (anterior), acromion (middle), and spine of the scapula (posterior). They all insert on the deltoid tuberosity on the anterolateral aspect of the proximal humerus and are innervated by the axillary nerve. The anterior head functions to provide forward elevation, the middle head provides abduction, and the posterior head provides extension of the shoulder. The teres major originates from the lower border of the scapula and inserts on the medial rim of the bicipital groove. It is innervated by the lower subscapular nerve and functions to rotate the shoulder internally.

The rotator cuff muscles are dynamic stabilizers of the humeral head and work in an intricate fashion to center the humeral head in the glenoid throughout the range of motion of the shoulder. Additionally they function to resist the strong forces of the deltoid, latissimus dorsi, and teres major. The rotator cuff contributes to the motion of the shoulder. The primary function of the supraspinatus is abduction of the shoulder, the infraspinatus and teres minor function to rotate the shoulder externally, and the subscapularis functions to rotate the shoulder internally (Fig. 17).

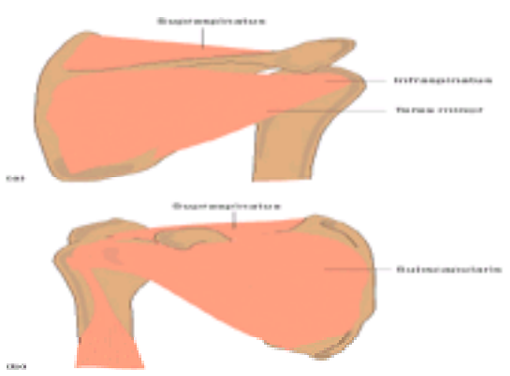


Fig. 17. (a, b) Shoulder rotator cuff musculature.

Physical examination of the shoulder

Inspection

The examination first begins with inspection of the patient's shoulder with attention paid to any asynchronous, abnormal, decreased, or painful motion. The shoulder is visually inspected for any asymmetry or the presence of atrophy of the shoulder musculature.

The entire shoulder girdle should be palpated in a sequential fashion for any areas of tenderness or crepitus within the bony structures comprising the acromioclavicular joint, clavicle, and subcutaneous joint, and the soft-tissue structures comprising the biceps tendon, and the anterior and posterior rotator cuffs. It is important to evaluate both the patient's active and passive range of motion. The shoulder has six directions of motion that should be evaluated: abduction, adduction, extension, flexion, internal rotation, and external rotation. When evaluating abduction the glenohumeral and scapulothoracic joints contribute to the motion in a 2:1 ratio, respectively. Normal range of motion is: abduction 180°, adduction 45°, extension 45°, flexion 180°, internal rotation 55°, and external rotation 45°. Strength testing should be the next component of the shoulder examination. The strength of each muscle should be isolated and individually graded using the standard 0 to 5 rating scale in Table 1.

The neck should always be included as part of the examination of the shoulder. Neck pathology, such as a herniated cervical disk or congenital cervical stenosis, can be confused with disorders of the shoulder.

Several special provocative tests of the shoulder may aid in diagnosis. The apprehension test is a test for shoulder instability (Fig. 18). The patient's arm is brought into an abducted and externally rotated position in both a sitting and supine position. If the patient has a sense of shoulder apprehension and states that the shoulder is ready to pop out, the findings are consistent with a positive test. A posteriorly directed force on the humeral head may relieve this unstable feeling. This is termed a positive relocation test. Rotator cuff tendinitis or impingement can be elicited using certain maneuvers. The Neer impingement test brings the patient's arm into a forward elevation position while stabilizing the spine of the scapula. The reproduction of pain is consistent with a positive test (Fig. 19). Hawkin's impingement test, another provocative test for impingement, is performed by bringing the patient's forward flexed arm into an adducted internally rotated position (Fig. 20). This similarly reproduces pain associated with rotator cuff tendinitis or impingement. The cross-adduction test is specific for acromioclavicular joint injuries. This test can be performed in both a sitting and supine position. With the arm flexed forward 90°, it is maximally adducted. Patients will describe localized pain over the acromioclavicular joint. Provocative tests for tendinitis of the long head of the biceps include Yergason's and Speed's tests. Yergason's test is performed with direct palpation over the long head of the biceps within the bicipital groove and resistance to supination with the elbow flexed at 90° and a pronated forearm. Speed's test is performed with a fully extended arm and pain is reproduced on forward elevation against resistance.



Fig. 18. Shoulder instability apprehension test.



Fig. 19. Neer impingement test.



Fig. 20. Hawkin's impingement test.

Imaging studies of the shoulder

Plain radiographs

Plain radiographs of the shoulder are essential in the clinical setting of trauma, instability, or degenerative joint disease. The routine views that should be obtained are anteroposterior views in internal, external, and neutral rotation and an axillary view. These views profile the acromioclavicular joint, glenohumeral joint, humeral head, and the tuberosities. They provide information about alignment, arthritis, and calcific deposits around the joint. Other views that can be helpful include the Stryker notch view to demonstrate a humeral head defect, Hill Sachs lesion, and West Point view which demonstrates an avulsion fracture of the glenoid, Bankart lesion, associated with shoulder dislocation. The supraspinatus outlet view described by Bigliani can be obtained for the best evaluation of the shape of the acromion and in those patients with rotator cuff pathology and impingement (Fig. 21).

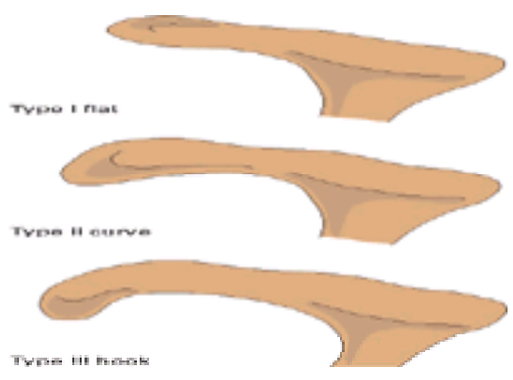


Fig. 21. Types of acromion.

Arthrography

Arthrography of the shoulder, single or double contrast, is helpful for diagnosing full-thickness rotator cuff tears, with a sensitivity of 95 per cent. The test is performed by injecting radio-opaque contrast into the glenohumeral joint and taking plain radiographs. Those patients with full-thickness rotator cuff tears will exhibit extravasation of contrast into the subacromial space. Arthrography can also be utilized to evaluate other shoulder conditions such as adhesive capsulitis, ‘frozen shoulder', which demonstrates a significant decreasein capsular volume (< 10 ml), absence of the axillary recess, and labral pathology.

Computed tomography

CT is useful in the shoulder to aid in the diagnosis of bony disorders such as fractures. In trauma it is useful in the definition of fracture patterns of the scapular body, proximal humerus, or glenoid. Software is available for the creation of three-dimensional reconstructions as an aid to preoperative planning. CT scans of the shoulder may be enhanced by the injection of contrast for better delineation of subtle osseous or periarticular abnormalities such as capsular redundancy, loose bodies, glenoid labral lesions, and full-thickness rotator cuff tears.

Magnetic resonance imaging

MRI provides exquisite anatomic detail (Fig. 22). The advantages of MRI include the high level of soft-tissue detail, multiplanar images, and absence of bone artifacts, combined with the lack of ionizing radiation and non-invasiveness. Unlike plain arthrography, the size of the rotator cuff tear may be quantified in relation to the amount of retraction and the quality of the muscle. Those patients with chronic rotator cuff tears will display fatty degeneration of the muscle. The intracapsular structures can also be enhanced further by injecting a contrast medium, such as gadolinium or saline, into the glenohumeral joint prior to imaging. An MRI arthrogram is beneficial in assessing capsular and labral pathology. The disadvantages of MRI are its present high cost, reliance on expert radiologic interpretation, and possible claustrophobia.

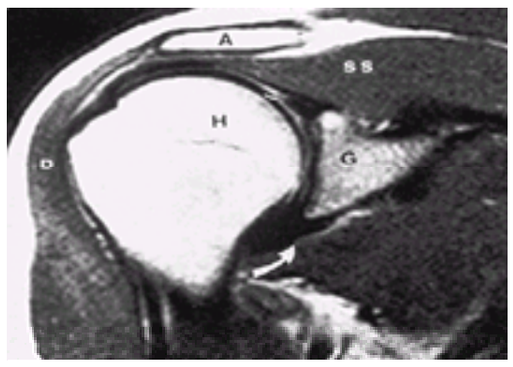


Fig. 22. MRI of shoulder demonstrating normal anatomy (arrow, labrum; H, humerus; D, deltoid; A, acromion; SS, supraspinatus; G, glenoid).

Ultrasonography

Ultrasound is utilized in some institutions and is another non-invasive modality that can be used to evaluate shoulder anatomy and pathology. It allows comparison with the other side for the detection of pathology. Studies have reported a 91 per cent sensitivity and specificity in the detection of rotator cuff tears. Ultrasound may also be useful in the diagnosis of rotator cuff tears in patients who have had previous repairs and ruptures of the biceps tendon. The primary disadvantage of ultrasound is that the interpretation of the images obtained is difficult and very user dependent.

Injuries about the shoulder and their treatment

Fractures

Clavicle fractures are common in the athlete. They usually result from either direct trauma or a fall onto an outstretched hand. The majority of these fractures can be treated with a simple sling and swathe even when significant displacement exists. Indications for operative fixation include underlying vascular or brachial plexus injuries and open fractures.

Fractures of the proximal humerus are less common in the younger athletic population and are more commonly seen in the older patient with low-energy trauma. Displacement and angulation of the fracture will dictate the treatment plan.

Scapular body fractures require a tremendous amount of energy; thus other injuries are commonly present and may include damage to major vessels, thorax, or cervical spine. These fractures are usually treated non-operatively.

Glenoid fractures are a result of a dislocation or other high-energy trauma. Non-displaced fractures can be treated with brief sling immobilization and early motion. However, fractures that involve the articular surface and have a significant step-off may require open reduction and internal fixation.

Dislocations and subluxations

Acute dislocation of the shoulder in an athlete is often in an anterior direction, but can be posterior, inferior, or any combination (multidirectional). The mechanism typically includes indirect forces to the upper extremity. These forces generally place the shoulder in a position of extreme external rotation combined with abduction and hyperextension. The diagnosis of an anterior dislocation is usually obvious on physical examination. The arm is held in a position of slight abduction and external rotation. The humeral head is palpated on the anterior aspect of the shoulder beneath the coracoid. The younger the patient at the time of initial dislocation, the higher the risk of redislocation. In a recent study, patients younger than 20 years of age at initial dislocation had a 90 per cent recurrence rate. Initial treatment of the dislocation is a closed reduction after the neurovascular status of the extremity has been carefully examined and documented. After the closed reduction, plain radiographs should be obtained to rule out associated fracture of the glenoid (Bankart or Perthes lesion) or impaction fracture of the humeral head (Hill Sachs lesion). The patient should wear a sling for about 2 weeks and then be referred for a program of physical therapy to regain motion and strength. The athlete may return to competitive events when the shoulder is no longer symptomatic and has regained full strength. Those patients who suffer recurrent dislocations will require operative repair, either arthroscopically or open.

Acute anterior subluxation of the shoulder is common in the throwing athlete. Usually the patients are unaware that their shoulder has subluxed but experience pain during the cocking and acceleration phase of throwing. The athlete may describe a sharp pain when the arm is placed in such a position, loses control of the extremity, and drops the ball from the hand. This is consistent with ‘dead arm syndrome'. After the acute episode, the severe pain usually subsides quickly, but the shoulder may remain sore and weak. Recurrent subluxation may be the result of a traumatic anterior dislocation or generalized ligamentous laxity. All the signs of instability, such as apprehesion and relocation tests, are positive. The athlete may also have signs of secondary rotator cuff tendinitis or impingement. Treatment includes anti-inflammatory medication, rest from throwing activities, and strengthening the rotator cuff musculature. Those patients who do not improve on a conservative therapy program are candidates for surgical reconstruction. This involves tightening of the attenuated capsule and anterior inferior capsular shift. A newer approach includes shrinkage of the capsule with a thermal probe to decrease capsular volume and improve stability.

The acromioclavicular joint is commonly injured during contact sporting activities. The mechanism of injury involves either a direct force, fall onto the point of the shoulder, or less commonly an indirect force to the shoulder when the athlete falls on an outstretched arm. Tenderness, swelling, and obvious bony deformity are localized over the acromioclavicular joint. On physical examination the outer end of the clavicle is displaced superior to the acromion to the acromioclavicular and the coracoclavicular ligaments, depending on the severity of the injury. These injuries are classified into six types: (I) sprain of the acromioclavicular ligament, (II) disruption of the acromioclavicular ligament and a sprain of the coracoclavicular ligament, (III) disruption of the acromioclavicular and coracoclavicular ligaments with 100 per cent superior migration of the distal clavicle, (IV) disruption of the acromioclavicular and coracoclavicular ligaments with posterior displacement of the distal clavicle through the trapezius muscle, (V) disruption of the acromioclavicular and coracoclavicular ligaments with 300 per cent superior displacement of the distal clavicle, and (VI) disruption of the acromioclavicular and coracoclavicular ligaments with inferior displacement of the distal clavicle below the coracoid (Fig. 23). Grades I, II, and III are generally treated non-operatively with a sling and the athlete may return to sports when pain free. However, some orthopedists prefer to make surgical repairs of type III injuries, particularly in throwing athletes. In those athletes managed non-operatively, the treating physician should warn of the risk of developing degenerative arthritic changes of the acromioclavicular joint that may require excision of the distal clavicle in the future. Injuries of types IV, V, and VI are commonly treated with open reduction and internal fixation.

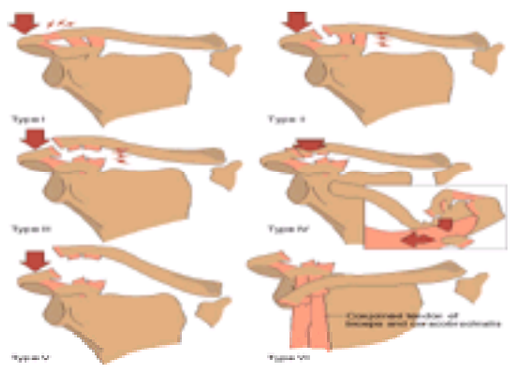


Fig. 23. Classification of acromioclavicular joint dislocations.

Rotator cuff injuries

Impingement syndrome or rotator cuff tendinitis is one of the most common causes of shoulder pain in athletes. The majority of the patients are involved in sports requiring forceful overhead throwing motions such as baseball, tennis, and volleyball. As mentioned previously, the primary function of the rotator cuff is to center the humeral head in the glenoid throughout the range of motion of the shoulder by resisting the superiorly directed forces applied bythe much larger deltoid. When this balance of forces is disruptedby acute injury or repetitive stress, the deltoid will overpower the rotator cuff, forcing the humeral head to migrate in a superior direction and impinge on the undersurface of the acromion. Subtle shoulder instability or direct injury to the rotator cuff such as a contusion may lead to the superior migration of the humeral head and eventual impingement. Traditionally, impingement has been described in three progressive stages leading to eventual rotator cuff degeneration: (I) inflammation and edema; (II) degeneration and fibrosis; (III) frank tear, either partial or full. Stage I rotator cuff impingement is reversible if activities are modified and there is a well-structured rehabilitation program that focusus on decreasing inflammation with anti-inflammatory medications, occasional subacromial injection (impingement test) of corticosteriods, and strengthening the rotator cuff musculature. Subacromial injection may serve both a diagnostic and therapeutic function. Those patients who do not respond to an aggressive therapy regimen may require surgical intervention. Most young patients have an underlying shoulder instability, which leads to secondary impingement/rotator cuff tendinitis. The primary pathology, attenuated capsule, must be addressed first prior to performing impingement procedures such as an acromioplasty, particularly in those patients with underlying shoulder instability. Concurrent rotator cuff tears must be repaired along with an anterior acromioplasty, either open or arthroscopically. Return to athletics, particularly in overhead throwers such as baseball pitchers, is unpredictable as a full recovery may be lengthy.

Biceps tendon injuries

Injuries that involve the long head of the biceps tendon are a common cause of anterior shoulder pain. The spectrum of injury includes tendinitis and complete or partial rupture of the tendon. Biceps tendinitis is often seen in association with impingement syndrome. Pain in the young athlete is reproduced with overhead activities such as throwing. Yergason's and Speed's tests on physical examination are positive (refer above). Treatment includes rest, anti-inflammatory medication, physical therapy, and occasional corticosteriod injection. If the injury does not improve and the biceps tendon is continually subjected to pressure and impingement on the underside of the acromion, it becomes frayed, flattened, and eventually ruptures. The rupture of the biceps tendon proximally is generally treated non-operatively. There is a minimal lose in strength (10 to 15 per cent with supination, near normal strength with elbow flexion). In athletes such as weightlifters where this loss is unacceptable, operative intervention may be considered.

Labral injuries

SLAP (superior labrum anterior posterior) lesions described by Synder involve injury of the origin of the biceps tendon at the superior aspect of the glenoid (Fig. 24). This injury pattern has been diagnosed with increased frequency in throwers and overhead activities. The mechanism of injury involves a fall on an outstretched arm with the shoulder in abduction and slight forward flexion. Symptoms include pain that worsens with overhead activity and a sensation of ‘catching' or popping in the shoulder. The ‘clunk test' may aid in the diagnosis of SLAP lesions. With the patient in the supine position, one of the examiner's hands is placed posterior to the humeral head and used to apply an anteriorly directed force to the humerus while the opposite hand rests on the humeral condyles and is used to rotate the humerus. A clunk or grind noted when the arm is in the full overhead abducted position is suggestive of a labral tear. These lesions begin posteriorly and extend anteriorly and have a spectrum ranging from fraying of the labrum to a complete bucket-handle tear of the labrum involving the origin of the biceps tendon. MRI can be helpful in assessing SLAP tears. An unstable superior labrum thatis causing pain or a catching sensation may be arthroscopically reattached to the rim of the glenoid, while degenerative and frayed labral tears are best treated with arthroscopic debridement.

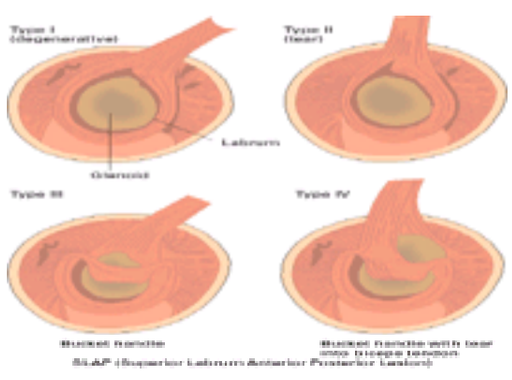


Fig. 24. Classification of SLAP lesions.

Elbow

Anatomy

Anatomically the elbow is a loose hinge joint between the distal humerus and proximal radius and ulna (Fig. 25). The distal humerus is divided into the capitellum and trochlea. The capitellum articulates with the radial head. The trochlea articulates with the olecranon and coronoid of the proximal ulna. Elbow stability is provided through the bony conformity between coronoid / trochlear articulation and radiocapitellar joint composed of the annular ligament and capsule. The primary ligament support of the elbow to valgus forces is the medial collateral ligament complex composed of three distinct bands— anterior oblique, posterior oblique, and transverse (Fig. 26a). The anterior oblique band of the medial collateral ligament is the most important component in restraint to valgus forces. The lateral collateral ligament, which is less developed, provides restraint to varus stresses (Fig. 26b). Motion of the elbow includes flexion and extension (0 to 150°) and pronation and supination (80 to 90°). The elbow in full extension has a valgus alignment (10 to 15°) known as the carrying angle.

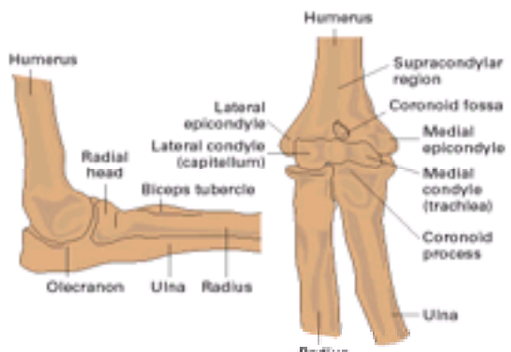


Fig. 25. Elbow joint.

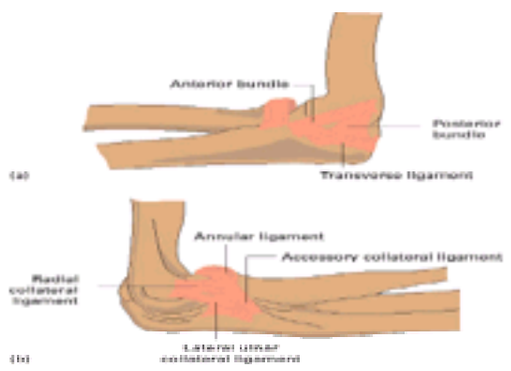


Fig. 26. (a) Elbow medial collateral ligament. (b) Elbow lateral stabilizing structures.

Physical examination

The examination begins initially with inspection. Muscle hypertrophy, atrophy, or swelling about the elbow may be present. Range of motion should also be documented. Loss of motion may be related to a intra-articular loose body or muscle tightness secondary to tendinitis or muscle strain. The elbow should be palpated to localize areas of tenderness. Integrity of the medial and lateral collateral ligaments is best assessed with varus and valgus stress applied at 25 to 30° flexion to unlock the olecranon (Fig. 27). Various provocative tests can be used to assess the spectrum of elbow disorders.

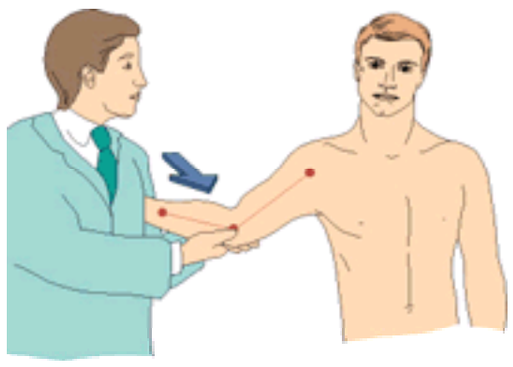


Fig. 27. Valgus stress examination of the medial collateral ligament of the elbow.

Radiographic evaluation

Standard elbow radiographs include anteroposterior, lateral, and external and internal oblique views for the evaluation of most elbow disorders. Close evaluation of the lateral projection of the elbow after acute trauma may reveal a posterior fat pad sign that is indicative of intra-articular fracture. Stress radiographs are helpful in evaluating medial and lateral collateral ligament injuries. CT is a proven adjunct to plain radiographs in demonstrating subtle elbow injuries. Radionuclide bone scan can be helpful in documenting and localizing bone stress fractures and occult post-traumatic fractures not diagnosed by conventional radiographs. MRI has gained acceptance in the evaluation of both.

Ligament injuries

The medial aspect of the elbow is exposed to high repetitive valgus stresses during overhead throwing sports. These forces may lead to acute or chronic failure of the ulnar collateral ligament. Patients have pain localized over the medial elbow and the ulnar collateral ligament, especially distally. Pain and instability are also noted when a valgus stress is applied to a slightly flexed (25°) elbow. Plain radiographs may reveal loose bodies in chronic cases. Gravity stress radiographs may reveal medial opening. MRI is a new technique being utilized to evaluate the integrity of the ulnar collateral ligament. Non-surgical treatment including anti-inflammatory medication, rest, modification of activities, and cold compression to reduce swelling and pain is successful in most cases. In elite athletes or those with very unstable elbows, surgery may be indicated.

Articular injuries

Medial epicondyle fracture is common in the adolescent. This condition is commonly referred to as ‘little league elbow'. Repetitive valgus stresses in young athletes who are throwers can result in a stress fracture of the medial epicondylar apophysis. Localized elbow pain at the medial epicondyle is associated with inability to throw. Local treatment, rest, and avoidance of throwing activities allow for adequate healing. Continued stresses will result in epicondylar avulsion fracture or medial elbow instability and require surgical treatment.

Lateral elbow pain in the adolescent athlete is associated with osteochondritis dissecans or osteochondrosis of the radiocapitellar joint. The basic pathologic state is a vascular insufficiency that is most often of the capitellum but can involve the radial head. The physician must have a high index of suspicion. Patients often present with pain on the lateral side of the elbow with a flexion contracture. Treatment consists of rest and withdrawal from throwing activities. Continued throwing will deform the articular surfaces and produce permanent joint irregularities or exfoliate loose bodies. Surgery is indicated in those patients who develop fragmentation resulting in intra-articular loose bodies. These can be removed either arthroscopically or by open arthrotomy

Leg

Compartment syndrome

Much research has centered on the diagnosis, etiology, and treatment of acute compartment syndrome related to trauma. Chronic exertional or exercise-induced compartment SYndrome in the athlete has been described over the past decade. This chronic form is more common than the acute form. The clinical picture of chronic compartment syndrome differs from the acute form, although the pathophysiology is essentially the same. Compartment syndrome has been defined as a condition in which an elevated tissue pressure exists within a closed fascial space, resulting in reduced capillary blood perfusion and compromised neuromuscular function. Unlike the acute form, the changes in the muscle are completely reversible between episodes. The anterior and deep compartments of the leg are most frequently involved. Athletes, particularly long-distance runners, will complain of pain that is gradual in onset during exercise and relieved by rest. They will also describe leg swelling, tightness, numbness, and weakness. Physical findings are often unimpressive with mild muscle soreness on deep palpation. Measurement of compartment pressures before, during, and immediately following exercise can help confirm the diagnosis. Normal resting tissue pressure in the muscle ranges form 0 to 14 mmHg at rest. Most authorities agree that resting pressures greater than 15 mmHg or delay in the return to normal after exercise (pressure greater than 20 mmHg 5 min after exercise) are consistent with chronic exertional compartment syndrome. Non-operative treatment, consisting of alterations to activity and training, physical therapy, anti-inflammatory medications, and shoe modifications, can improve the symptoms. For those patients who do not respond to these measures, fasciotomies are required.

Stress fractures

Stress fractures, also described as fatigue or insufficiency fractures, are common in athletes. Stress fractures are defined as partial or complete fractures of bone caused by repetitive stresses beyond what the bone is able to withstand. Incidence of stress fractures in runners has been estimated to range from 4 to 16 per cent. Women have been found to be up to 12 times more susceptible to the development of a stress fracture than their male counterparts. This fact appears to be attributed not only to smaller bone size but also menstrual irregularities. The tibia is the most commonly involved bone (49.1 per cent) (Fig. 28). The development of stress fractures often coincides with an increased level, duration, or intensity of training. Other factors associated with stress fractures include running on hard surfaces, inadequate footwear with poor impact-absorbing quality, and lower limb malalignment. Patients typically present with early stress fracture and will experience pain after training; but as the condition advances, pain is exacerbated during the activity. Eventually, pain is present with activities of daily living. Tenderness and periosteal thickening associated with stress fracture is usually well localized, unlike medial tibial stress syndrome. Callus formation may also be palpable. Findings on plain radiographs are variable. Usually 2 to 3 weeks are required for changes to occur. Early radiographic changes will reveal periosteal reaction. As the process matures, periosteal new bone formation occurs with eventual cortical fracture. Technetium bone scan has been the most reliable and sensitive means of detecting a stress fracture. The characteristic focal fusiform appearance is readily distinguished from the linear activity seen with medial tibial stress syndrome. Stress fractures generally resolve with activity modification and rest. To enhance healing, electrical or ultrasound stimulation has been used. Recalcitrant fractures may require surgical treatment. Persistence of the ‘dreaded black line', cortical fracture, with a positive bone scan for more than 6 months is an indication for surgical treatment and possible bone grafting.

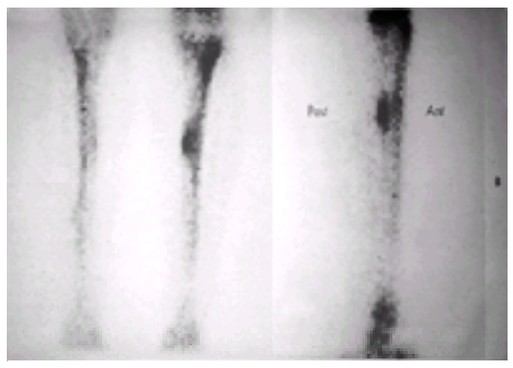


Fig. 28. Stress fracture of tibia.

Medial tibial stress SYndrome

Medial tibial stress syndrome, or more commonly referred to in layman's terms as ‘shin splints', is a common overuse injury of the lower extremity in the runner. Patients with medial tibial stress syndrome present with pain. The pain is initially felt upon exertion and may be relieved upon completion of running. The pain may recur toward the end of the workout or after running. At times the pain may be described as a dull ache or soreness. As the condition advances, the pain may be perceived as sharp, penetrating, and severe in nature. The athlete may start developing severe pain during the workout, limiting the athlete's ability to train. With time and continued abuse, the leg pain may be present with activities of daily living. The physical examination reveals tenderness along the posteromedial border of the tibia, usually about 4 cm proximal to the medial malleolus, that extends proximally for a variable distance. Slight swelling may be present. Percussion of the tibia elsewhere does not elicit pain. Active resistance plantar flexion and toe raises may elicit pain. Routine radiographs of the leg are usually normal. In chronic cases, there may be hypertrophy of the posterior cortex of the tibia or faint periosteal reaction. These findings are consistent with periostitis related to the insertion of the posterior tibialis and soleus. Technetium bone scan will distinguish medial tibial stress syndrome from stress fracture. Uptake in stress fractures is typically fusiform and localized compared with a more diffuse uptake in medial tibial stress syndrome. Treatment includes the use of anti-inflammatory medications, rest, and heel cord stretching. Those patients who fail to improve on a conservative regimen are candidates for surgery. Surgical treatment involves release of the superficial investing fascia around the soleus insertion on the posteromedial tibia.

Ankle injuries

Ankle sprains

Ankle sprains are the most common athletic injury constituting 16 to 25 per cent of all injuries. They account for the greatest loss of time from training and competition. Ankle stability is related to both bony architecture and ligament support. The shape of the talus and its unique fit between the tibia and fibula maintains the ankle joint, forming a mortise. Ligamentous stability of the ankle is formed by a complex of ligaments. The deltoid is a broad-based medial ligament composed of superficial and deep components running from the medial malleous to the talus. The lateral ligament complex is composed of the anterior talofibular ligament, calcaneofibular ligament, and posterior talofibular ligament (Fig. 29). The calcaneofibular ligament is the strongest of the lateral ligaments. The anterior talofibular and calcaneofibular ligaments are synergistic with range of motion of the ankle. With plantar flexion of the ankle, the anterior talofibular ligament becomes taut while the calcaneofibular ligament is relaxed. With dorsiflexion the opposite occurs, the calcaneofibular ligament becomes taut hile the anterior talofibular ligament is relaxed. The tibiofibular syndesmosis is the ligament that maintains the relationship between the tibia and fibula. It consists of the anterior and posterior inferior ligaments and the interosseous membrane.

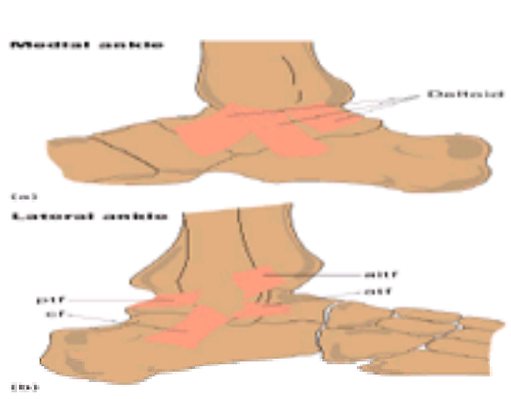


Fig. 29. (a) Ankle deltoid ligament. (b) Lateral ankle ligaments (atf, anterior talofibular; cf, calcaneofibular; ptf, posterior talofibular; aitf, anterior inferior tibiofibular).

Obtaining a complete history is of the utmost importance. The history should include eliciting the mechanism of injury. Radiographic evaluation of ankle injuries includes anteroposterior, lateral, mortise of the ankle, and proximal fibula views. These films should be critically reviewed for any osteochondral or avulsion fractures. Comparison stress radiographs are also very helpful in the assessment of ligament injuries. CT is useful in evaluating articular lesions such as osteochondral fractures while MRI is best at demonstrating soft-tissue damage.

Lateral ligament injuries account for the majority of ankle sprains. The mechanism of lateral ankle sprains involves a plantar flexed inverted ankle. Patients often remember a distinct pop at the time of the injury. The swelling and tenderness are localized over both the anterior talofibular and calcaneofibular ligaments. The swelling usually develops immediately in acute injury and is often delayed in chronic injuries. In lateral ligament injuries, clinical stability must be assessed. The anterior drawer test is a test to assess the integrity of the anterior talofibular ligament (Fig. 30(a)). The test is performed with the foot in a neutral position and 10° of plantar flexion. The talar tilt test is used to assess the integrity of the calcaneofibular ligament (Fig. 30(b)). The ankle is placed in a neutral position or 10° dorsiflexion and an inversion stress is applied. The amount of displacement with all these tests is recorded as grade I (mild), grade II (moderate), and grade III (marked). Grade I injuries correlate with mild stretching of the anterior talofibular ligament; grade II injuries involve partial tear of the anterior talofibular and calcaneofibular ligaments; while grade III injuries involve complete rupture of the anterior talofibular and calcaneofibular ligaments and marked instability. Non-operative treatment is recommended for acute grade I, II, and III sprains. For patients with little or no instability (grade I or II), treatment consists of immobilization of the ankle in a neutral position and protected weight-bearing until symptoms resolve, followed by peroneal strengthening, range of motion exercises, and proprioceptive training. Grade III sprains are treated with immobilization in a short leg walking cast or brace for 3 to 6 weeks. Thosepatients who go on to develop chronic instability may require surgical reconstruction.

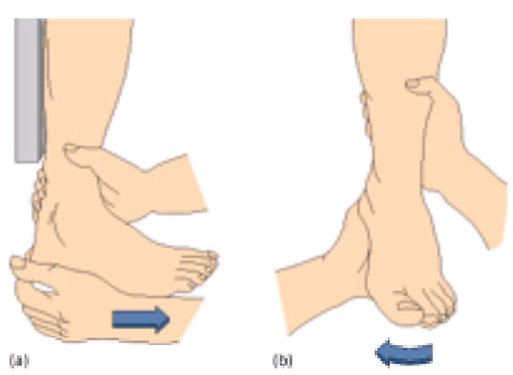


Fig. 30. (a) Anterior drawer test. (b) Talar tilt test.

Isolated deltoid ligament injuries account for less than 5 per cent of all ankle sprains and usually occur in association with other ligament, including syndesmotic, injury or fracture. The mechanism involves an eversion type of force. Pain, tenderness, and swelling are usually present on the medial side. Ecchymosis may be present after 1 or 2 days. The deltoid ligament is evaluated utilizing the valgus (eversion) test in which the examiner applies a strong eversion force with the foot in a neutral position. Stable injuries without disruption of the mortise can be treated non-operatively. Those grade III sprains with disruption of the deltoid and widening of the mortise will require examination under anesthesia with stress radiographs. If the mortise closes easily, a cast can be applied with the foot in plantar flexion and inversion for 6 weeks. Operative repair of the deltoid ligament consisits of closure of the mortise with a syndesmotic screw.

Syndesmotic ankle injuries are misdiagnosed often as a routine ankle sprain. A high index of suspicion is required to make the correct diagnosis. Analysis of the mechanism will often lead the clinician to the proper diagnosis. In SYndesmotic injuries, the foot is dorsiflexed and externally rotated. Pain and tenderness are located on the anterior aspect of the interosseous membrane. Active movement by external rotation of the foot is painful. Syndesmotic injuries are associated with prolonged recovery and delay return to a competitive level. The cotton and squeeze tests assess the syndesmotic injuries. The squeeze test is considered positive when proximal compression produces distal pain in the area of the interosseous membrane (Fig. 31). The cotton test is positive when pain is reproduced and instability observed when a mediolateral force is applied to the talus. Treatment includes primary repair of anterior inferior tibiofibular and fixation of syndesmosis.

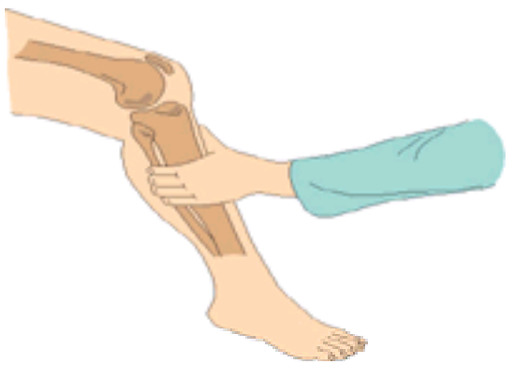


Fig. 31. Syndesmotic squeeze test.

Foot injuries

Tendinitis

Tendons are relatively avascular, which accounts for their white appearance, and are composed of 30 per cent collagen (mainly type I), 2 per cent elastin, and 69 per cent water. Collagen resists tensile forces and elastin provides flexibility to the tendon. The tendinous insertion of the muscle–tendon unit around the foot and ankle bears much of the stress and force that occurs during vigorous athletic activities. These repetitive activities can lead to inflammation of the tendon and surrounding tissue. In some cases there will be degeneration of the tendon and eventually complete rupture. Poor training habits, such as a marked increase in distance or speed, or improper footwear and excessive foot pronation have been implicated in the etiology of tendinitis. Tendons often affected include the Achilles, posterior tibilalis, and peroneal tendons.

Achilles tendinitis is a common entity found in runners. The paratendon surrounding the Achilles tendon becomes inflamed and thickened with repetitive loads. This leads to impairment of the gliding of the Achilles tendon and further inflammation. The patient presents with pain and throbbing with varying degrees of activity. Physical findings in patients with Achilles tendinitis include soft-tissue swelling, local tenderness, and crepitus. In more chronic cases, the area of tenderness is larger and often thickened and nodular raising suspicion of tendinosis or partial rupture of the tendon. Most cases of Achilles tendinitis can be successfully treated non-operatively. The basic modalities of non-operative treatment are rest or decreased mileage, use of heel lift, anti-inflammatory medications, use of orthotics to control excessive foot pronation, ultrasound therapy, and stretching. Immobilization of the ankle for a period of 7 to 10 days may be indicated in individuals with severe acute symptoms. Those chronic cases that fail to respond to conservative treatment may require surgical intervention. The surgery involves release and excision of the inflamed and scarred paratendon and, if necessary, debridement of degenerative tendon. Chronic cases may result in partial or complete rupture of the Achilles tendon. In the athlete, operative intervention with primary repair of the Achilles tendon is the treatment of choice.

Posterior tibial tendinitis is most commonly seen in athletic activities requiring quick change in direction such as basketball, soccer, and tennis. As the posterior tibial tendon passes behind the medial malleolus, repetitive microtrauma occurs causing inflammation that may eventually lead to degeneration and rupture. Physical examination reveals a more flat foot and mildly positive heel-rise test (lack of inversion of the heel when patient stands on toes). There is tenderness and swelling along the tendon as it courses behind the medial malleolus. Passive pronation and active supination reproduce painful symptoms. Treatment consists of rest, ice, anti-inflammatory medication, and orthotics. In those young athletes with acute inflammation, a period of immobilization in a short leg cast that is non-weight-bearing and in slight inversion, for 10 days to 2 weeks, often relieves the symptoms dramatically. Steroid injections have been advocated by some to decrease the inflammation. However, there is a risk of tendon rupture. Those cases that do not respond to conservative treatment may require surgery. Release of the tendon sheath, debridement of scar tissue, and synovectomy have been shown to be successful in such cases.

Flexor hallucis longus tendinitis is a common entity in athletes, such as classical ballet dancers, who perform repetitive push-off maneuvers. Tremendous stress is transmitted across the tendon causing inflammation. Dancers with this tendinitis often note insidious onset of pain at the posterior aspect of the medial malleolus. The chronic inflammation may lead to stenosis of the fibro-osseous causing triggering and snapping of the flexor hallucis longus tendon. On clinical examination the tendon may be tight. Tightness of this tendon can be checked by the amount of passive extension of the great toe metatarsophalangeal joint with the foot and ankle in both neutral and plantar flexed position. Treatment consists of ice, anti-inflammatory medication, and modification of activity. Chronic cases may require a short period of immobilization. Operative intervention with release of the tendon sheath and SYnovectomy should be considered when there is persistent synovitis or triggering of the tendon.

Plantar fasciitis is a very common overuse injury in runners. The plantar fascia functions to absorb shock from heel strike. Patients predisposed to plantar fasciitis include those with excessive foot pronation, poor footwear, high-arched feet, and a tight gastocnemius–soleus complex. Patients present with gradual pain along the inside of the heel. The pain is worse in the morning and then decreases with increased activity. Physical examination reveals point tenderness along the medial and plantar aspect of the heel. Stretching of the plantar fascia by passive dorsiflexion of the ankle may be painful. There may be soft-tissue swelling and a nodular feel to the plantar fascia when palpated. Treatment of plantar fasciitis has two facets: control of pain and inflammation, and modification of mechanical factors. The former is usually accomplished by use of anti-inflammatory medication, ice, orthotics/heel cups/night splinting, rest, and a heel cord stretching regimen with modalities such as ultrasound and iontophoresis. Local steroid injection is indicated for those patients who fail to improve with this initial conservative regimen. Injudicious use of steroid injection may result in heel pad atrophy. The majority of cases respond to a conservative treatment protocol. Patients who remain symptomatic may be candidates for release of the plantar fascia.

Turf toe

Turf toe is a common injury seen in American football players who compete on artificial turf. The injury is caused by forced hyperextension of the metatarsophalangeal joint of the great toe. The injury pattern ranges from a minor sprain to dislocation of the first metatarsophalangeal joint. There can be associated fracture of the metatarsal head or sesmoids. Predisposing factors include hard artificial playing surfaces, flexible soft-sole athletic shoes, and pre-existing limited motion of the first metatarsophalangeal joint. Physical examination and radiographs are important to define the injury pattern and to rule out other conditions of the forefoot. Depending upon the magnitude of the injury, different options are available. For minor sprains, which are characterized by local tenderness, ice, compression, and non-steroidal anti-inflammatory medications may be used. Taping of the toe to reduce dorsiflexion and insertion of rigid forefoot insoles to stiffen the shoe may decrease discomfort. The patient may continue to participate in athletics if pain is minimal. Those patients with more severe pain may have significant capsuloligamentous injury. Dislocation of the first metatarsophalangeal joint may occur. The joint may spontaneously reduce or require closed reduction. The joint should be immobilized for 3 to 4 weeks. Fractures of the metatarsal head may require open reduction and internal fixation. If the athlete has chronic instability or fails to improve with a conservative protocol, capsular repair or sesamoid excision may be required.

Hand and wrist injuries

Hand and wrist injuries are one of the most common injuries sustained in competitive sporting events from the young recreational athlete to the elite professional athlete. Finger injuries occur in ball handling sports such as basketball, baseball, and volleyball. Wrist injuries primarily occur in contact sports such as football, but also in racket sports, and gymnastics. These injuries can include fractures, dislocations, and sprains.

Wrist fractures

Fractures of the scaffold are one of the most common and problematic fractures of the carpal bones. These fractures usually occur in sports secondary to falls or contact with another player. There are a number of patients who are diagnosed late with a scaphoid fracture. These patients sustain minor trauma and are told they have a ‘minor wrist sprain'. Patients present with wrist pain localized to the snuffbox. This physical finding is enough to justify immobilization until a definitive diagnosis is made. Radiographs of the wrist should include posteroanterior views in neutral and ulnar deviation, lateral, and clenched fist view (Fig. 32). Bone scan after 72 h can help identify a fracture not seen on plain radiographs. There are several factors that adversely affect healing capacity of scaphoid fractures. Fractures of the proximal pole of the scaphoid are at risk of developing delayed union, non-union, and avascular necrosis. Displaced scaphoid fractures (greater than 1 mm) are at a significant risk of non-union. This fact is related to the retrograde blood supply to the scaphoid. Treatment for non-displaced scaphoid fractures includes cast immobilization with the wrist slightly plantar flexed and radial deviations for 6 weeks. Displaced scaphoid fractures (greater than 1 mm) require open reduction and internal fixation.

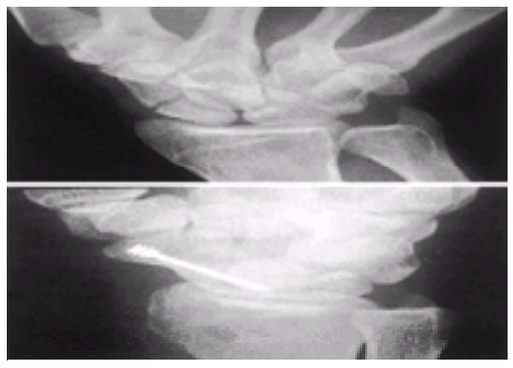


Fig. 32. Scaphoid fracture.

Fractures of the hamate usually occur at the hook and are commonly misdiagnosed. In athletes this injury commonly results from the direct force of a bat or tennis racket on the hypothenar eminence. The diagnosis should be suspected in those athletes who present with ulnar wrist pain. Examination reveals tenderness over the hook of the hamate. Plain radiographs are usually negative. A carpal tunnel view of the wrist best demonstrates this fracture. Patients with untreated cases of displaced hook fracture may develop a non-union and persistent hypothenar pain (Fig. 33). Additional complications include flexor tendon rupture and ulna nerve dysfunction. Excision of the hook through the fracture site is an effective treatment and allows return to sports in 6 to 8 weeks.

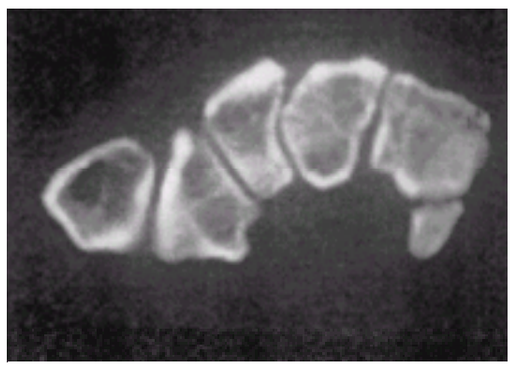


Fig. 33. Hook of hamate fracture.

Wrist ligamentous injuries

Ligament injuries, commonly known as ‘wrist sprain', can be a source of significant disability in the athlete whose sport requires a stable wrist with good mobility such as racket sports, gymnastics, and weightlifting. The single most useful means of achieving the correct diagnosis is by taking a detailed history. One needs to ascertain both the mechanism (if the wrist was in an extended or flexed position) and hand position at the time of injury and the current functional complaints (what motion or activity reproduces the pain). Next a careful examination of the wrist is performed with attention paid to swelling, ecchymosis, crepitus, areas of localized tenderness, and localizing clicks/snaps which may be associated with pain. Finally, the examiner attempts to reproduce and localize the patient's symptoms by stress testing specific joints and comparing them with the uninjured side. Once the affected joint has been identified, it may be selectively injected with Xylocaine. Radiographic evaluation of the athlete with a suspected ligament injury of the wrist includes standard anteroposterior, lateral, and oblique views and lateral and anteroposterior stress views as well as a clenched fist view. Bone scintigraphy and a wrist arthrogram may be indicated in cases where ligament injury is suspected but plain radiographs are normal. Treatment of wrist ligament injuries is systematic. Those athletes who present with relatively minor wrist sprains can be evaluated clinically, have radiographs taken, be splinted for a 10- to 14-day period, and then re-examined.

Scapholunate instability is the most common form of carpal instability. There is often a delay in diagnosis secondary to subtle findings in these injuries. The proposed mechanism of injury is a fall on or blow to an outstretched hand, which impacts the thenar eminence. This leads to disruption of a complex of ligaments: the volar radioscapholunate ligament, which is the strongest and most important ligament, the dorsal scapholunate ligament, and the interosseous scapholunate ligament. Pain, swelling, and tenderness are noted over the dorsoradial aspect of the wrist. Radiographic findings may be subtle. A widening (less than 2 mm) of the scapholunate interval, ‘Terry Thompson' sign, is noted on the anteroposterior view. The lateral radiograph reveals a characteristic instability pattern. Acute injuries of the scapholunate are treated according to the presence or absence of diastasis. Those patients with acute injury and pain localized to the scapholunate without a scaphoid fracture should be treated with 2 to 3 weeks of cast immobilization. Diastasis is treated with either open or closed reduction and pinning. Chronic injuries are problematic and require more extensive surgery.

Ligamentous injuries also commonly occur to the distal radioulnar joint. There is a complex of ligaments, which are the major stabilizers of the distal ulnar joint, known as the triangular fibrocartilage complex. This complex is located on the ulnar side of the wrist and is composed of the ulnar collateral ligament, the dorsal and volar radioulnar ligaments, a meniscus, the ulnolunate and ulnatriquetral ligaments, and the extensor carpi ulnaris tendon sheath (Fig. 34). In the athlete, injuries of the triangular fibrocartilage complex may occur as a result of acute trauma or secondary to repetitive activities. Acute tears of the disk may occur in association with a fall on an outstretched wrist causing a proximal shift of the radius with tearing of the triangular fibrocartilage complex. Sudden excessive pronation or supination may also cause disruption of the triangular fibrocartilage complex. A physical examination shows tenderness over the ulnar aspect of the wrist. Pain is reproduced with pronation and supination of the forearm. Standard radiographs of the wrist along with a wrist arthrogram aid in the diagnosis of a triangular fibrocartilage complex tear. MRI is gaining acceptance in the evaluation of these injuries. Most tears of this complex will heal with immobilization for 4 to 6 weeks. If symptoms persist despite a period of immobilization, arthroscopic excision or repair of the torn complex is indicated.

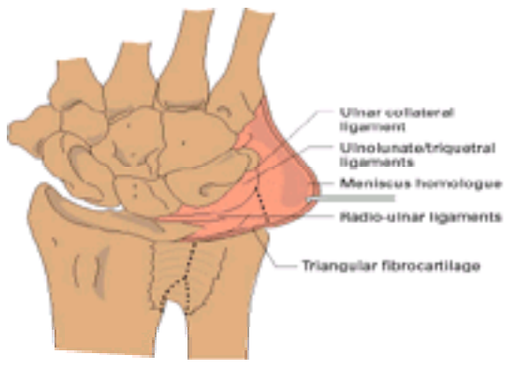


Fig. 34. Triangular fibrocartilage complex.

Wrist tendinitis

Tendinitis of the wrist is very common in the athlete. DeQuervain's syndrome or tenosynovitis of the first dorsal extensor compartment, composed of the abductor pollicis longus and extensor polllicis brevis, is common in those sports requiring repetitive ulna deviation such as golf and racket sports (Fig. 35). Physical findings include local swelling and tenderness at the radial styloid. The Finkelstein test, which is performed with the thumb fully adducted and ulnar deviation, reproduces pain. Treatment includes splinting, anti-inflammatory medications, and avoidance and modification of activities. Local corticosteriod injection of the first compartment can aid in resolution of symptoms. Resistant cases of DeQuervain's syndrome may require surgical release. Other forms of wrist tendinitis include intersection syndrome, where the abductor pollicis longus and extensor pollicis brevis cross over the underlying extensor tendons, and extensor carpi ulnaris tendinitis. These conditions are often seen in oarsmen. Treatments for these conditions are similar to those for other tendinitis.

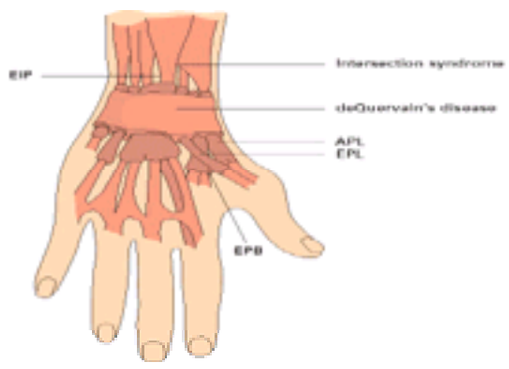


Fig. 35. Wrist extensor tendons (EPL, extensor pollicis longus; EPB, extensor pollicis brevis; APL, abductor pollicis longus; EIP, extensor indicis proprius), de Quervain's syndrome, intersection SYndrome.

Finger injuries

Mallet finger is an injury to the distal interphalangeal joint in which there is a stretching or rupture of the terminal extensor tendon (Fig. 36). The usual mechanism involves an axial load injury to an extended finger in sports such as basketball. The patient complains of pain, swelling, tenderness over the dorsum of the finger, and a notable flexed position, ‘drop finger'. Plain radiographs may show a fracture at the base of the distal phalanx. Treatment for the acute injury is splinting of the distal phalanx in extension or slight hyperextension for 6 weeks. Joint instability manifested by persistent or recurrent volar subluxation may require surgical reduction and repair.

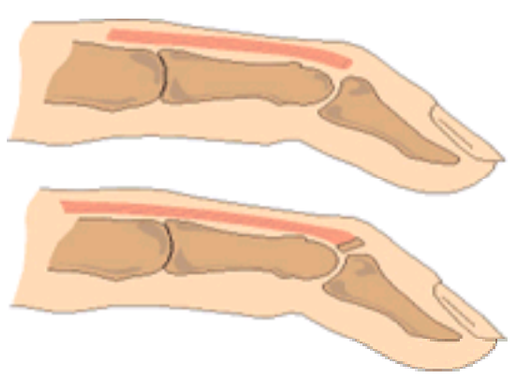


Fig. 36. Mallet finger, rupture of extensor tendon.

Avulsion of the flexor digitorum profundus tendon, or ‘jersey finger', is an injury sustained by athletes (Fig. 37). Two mechanisms of the injury are grabbing the jersey of a ball carrier in football or rugby, and catching a finger on the rim while slam dunking. As the athlete attempts to free the finger, it is suddenly extended, rupturing the flexor digitorum profundus tendon. The ring finger is the most commonly injured digit. Frequently the athlete does not appreciate the seriousness of the injury. The patient presents with a variable amount of pain. There is immediate swelling and ecchymosis along the tendon sheath. The physician must have a high index of suspicion and awareness of this injury, because a fair number of these injuries are missed. Plain radiographs may reveal a small avulsion from the distal phalanx. Early operative repair of the ruptured tendon is required for all injuries in order to achieve a good result.

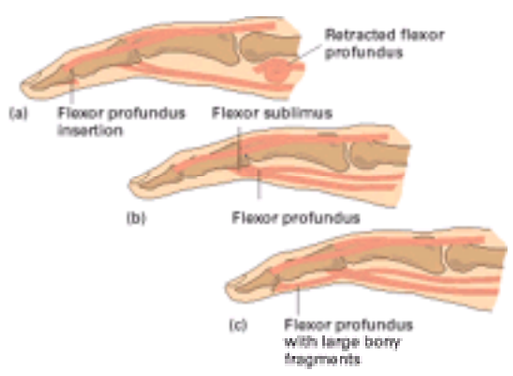


Fig. 37. Rupture of the flexor digitorum profundus tendon.

Injuries to the proximal interphalangeal joint include sprains and strains of the radial and ulnar collateral ligaments, dislocation, and fractures. Collateral ligament injuries are common in contact sports such as football, wrestling, and basketball. The cause of the injury is an adduction or abduction force with the proximal interphalangeal joint in extension. The radial collateral ligament is more commonly injured than the ulnar, and tears occur more commonly in a proximal position. Most proximal interphalangeal collateral injuries can be treated with functional splinting.

Instability following an injury to the ulnar collateral ligament of the thumb is common. The ligament is torn when an acute valgus (abduction) stress is applied to the metacarpophalangeal joint resulting in partial or complete disruption of the ligament. Injury to the ligament is common in such sports as skiing (‘skier's thumb'), football, and wrestling. The patient presents with a painful, swollen metacarpophalangeal joint of the thumb. Usually, the point of maximal tenderness can be localized to the ulnar aspect of the joint. The integrity of the ulnar collateral ligament should be examined in both extension and flexion, grading the opening and quality of the endpoint to differentiate between sprain (partial tear) and rupture (complete tear). This can often be difficult. Routine radiographs that should be obtained include posteroanterior and lateral views prior to stressing the metacarpophalangeal joint of the thumb. Close examination of the radiographs may reveal a small avulsion fracture at the base of the proximal phalanx representing a ‘Stenner lesion'. Avulsion fractures with significant displacement are typically complete tears of the ulnar collateral ligament and require primary surgical repair. Partial stable injuries should be treated non-operatively with a well-molded thumb spica cast for 3 to 6 weeks.

Proximal interphalangeal dislocations are also a very common injury in the athlete. The majority of these dislocations are in the dorsal direction as a result of an axial force and hyperextension of the digit. The volar plate is torn in such an injury. Volar dislocations of the proximal interphalangeal joint are rare and result from varus and valgus forces rupturing the collateral ligaments. A coach or trainer on site usually performs reduction. After reduction, a careful examination should be performed to rule out the possibility of associated injuries. Additionally, stability of the joint should be documented. Static splinting in 20° of flexion for 2 to 3 weeks is indicated, followed by buddy taping. Dislocations with large fracture fragments or that are irreducible may require open reduction and surgical repair of the soft tissues.

Further reading

Shoulder

Ada J, Miller M. Scapular fractures: analysis of 113 cases. *Clinical Orthopaedics and Related Research* 1991; **269**: 174–80.

Altchek D, Warren R, Wickiewiez T. Arthroscopic acromioplasty. Technique and results. *Journal of Bone and Joint Surgery* 1990; **72A**: 1198–207.

Arciero R, Wheeler J, Ryan J, McBride J. Arthroscopic Bankart repair versus non-operative treatment for initial anterior shoulder dislocation. *American Journal of Sports Medicine* 1994; **22**: 589–94.

Burkhead W, Rockwood C. Treatment of instability of the shoulder with an exercise program. *Journal of Bone and Joint Surgery* 1992; **74A**: 890–6.

Burns T, Turba J. Arthroscopic treatment of shoulder impingement in athletes. *American Journal of Sports Medicine* 1992; **20**: 13–16.

Cofield R, Irving J. Evaluation and classification of shoulder instability. With special reference to examination under anesthesia. *Clinical Orthopaedics and Related Research* 1987; **223**: 32–43.

Cox J. Current method of treatment of acromioclavicular joint dislocations. *Orthopedics* 1992; **15**: 1041–4.

Goss T. Current concepts review: Fractures of the glenoid cavity. *Journal of Bone and Joint Surgery* 1992; **74A**: 299–305.

Hawkins R, Kennedy J. Impingement syndrome in athletes. *American Journal of Sports Medicine* 1980; **8**: 151–8.

Hovelius L. Anterior dislocation of the shoulder in teenagers and young adults: five-year prognosis. *Journal of Bone and Joint Surgery* 1987; **69A**: 393–9.

Iannotti J, Zlatkin M, Esterhai J, Kressel H, Esterhaj J. Magnetic resonance imaging of the shoulder: sensitivity, specificity, and predictive value. *Journal of Bone and Joint Surgery* 1991; **73A**: 17–29.

Larsen E, Bjerg-Nielsen A, Christensen P. Conservative or surgical treatment of acromioclavicular dislocation: a prospective, controlled, randomized study. *Journal of Bone and Joint Surgery* 1986; **68A**: 552–5.

Lyons A, Tomlinson J. Clinical diagnosis of tears of the rotator cuff. *Journal of Bone and Joint Surgery* 1992; **74B**: 414–15.

Neer C. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. *Journal of Bone and Joint Surgery* 1972; **54A**: 41–50.

Nuber G, Bowen M. Acromioclavicular joint injuries and distal clavicle fractures. *Journal of the American Academy of Orthopedic Surgeons* 1997; **5**: 11–18.

Post M. Current concepts in the treatment of fractures of the clavicle. *Clinical Orthopaedics and Related Research* 1989; **245**: 89–101.

Rowe C. Prognosis in dislocation of the shoulder. *Journal of Bone and Joint Surgery* 1956; **38A**: 957–77.

Rowe C, Patel D, Southmayd W. The Bankart procedure: a long-term end result study. *Journal of Bone and Joint Surgery* 1976; **60A**: 1–16.

Speer K, Deng X, Borrero S, Altchek D, Warren R. Biomechanical evaluation of a simulated Bankart lesion. *Journal of Bone and Joint Surgery* 1994; **76A**: 1819–26.

Synder S, Karzel R, Del Pizzo W, Karzel R. SLAP lesions of the shoulder. *Arthroscopy* 1990; **6**: 274–9.

Tibone J. Surgical treatment of tears of the rotator cuff in athletes. *Journal of Bone and Joint Surgery* 1986; **68A**: 887–91.

Van Holsbeeck E, DeRycke J, Declercq G. Subacromial impingement: open versus arthroscopic decompression. *Arthroscopy* 1992; **8**: 173–8.

Wojtys E, Nelson G. Conservative treatment of grade III acromioclavicular dislocations. *Clinical Orthopaedics and Related Research* 1991; **268**: 112–19.

Zarins B, Andrews J, Carson W. *Injuries to the throwing arm*. Saunders, Philadelphia, 1998.

Elbow

Bauer M, Jonsson K, Josefsson P. Osteochondritis dissecans of the elbow: a long-term follow up study. *Clinical Orthopaedics and Related Research* 1992; **284**: 156–60.

Bennett J, Green M, Tullos H. Surgical management of chronic medial elbow instability. *Clinical Orthopaedics and Related Research* 1992; **278**: 62–8.

Conway J, Jobe F, Glousman R, Pink M. Medial instability of the elbow in throwing athletes: treatment by repair or reconstruction of the ulnar collateral ligament. *Journal of Bone and Joint Surgery* 1992; **74A**: 67–83.

Fuss F. The ulnar collateral ligament of the elbow joint: anatomy, function and biomechanics. *Journal of Anatomy* 1991; **175**: 203–12.

Jobe F, Stark H, Lombardo S. Reconstruction of the ulnar collateral ligament in athletes. *Journal of Bone and Joint Surgery* 1986; **68A**: 1158–63.

Morrey B, An K. Articular and ligamentous contributions to the stability of the elbow joint. *American Journal of Sports Medicine* 1983; **11**: 315–19.

Nirschl R. Soft tissue injuries about the elbow. *Clinics in Sports Medicine* 1986; **5**: 637–52.

Nirschl R, Pettrone F. Tennis elbow: The surgical management of lateral epicondylitis. *Journal of Bone and Joint Surgery* 1979; **61A**: 832–9.

Vangsness C, Jobe F. Surgical treatment of medial epicondylitis: results in 35 elbows. *Journal of Bone and Joint Surgery* 1991; **73B**: 409–11.

Wrist and hand

Cooney W, Dobyns J, Linscheid R. Fractures of the scaphoid: a rational approach to management. *Clinical Orthopaedics and Related Research* 1980; **149**: 90–7.

Culver J. Sports-related fractures of the hand and wrist. *Clinics in Sports Medicine* 1990; **9**: 85–109.

Dell P. Traumatic disorders of the distal radioulnar joint. *Clinics in Sports Medicine* 1992: **11**: 141–59.

Fox J, Del Pizzo W, eds. *The patellofemoral joint*. McGraw-Hill, New York, 1993.

Harvey F, Harvey P, Horsley M. DeQuervain's disease: surgical and non-surgical treatment. *Journal of Hand Surgery* 1990; **15A**: 83–7.

Herbert T. *The fracture of the scaphoid*. Quality Medical Publishing, St Louis, MO, 1990.

Koman L, Mooney J, Poehling C. Fractures and ligamentous injuries of the wrist. *Hand Clinics* 1990; **6**: 477–91.

Louis D, Huebner J, Hankin F. Rupture and displacement of the ulnar collateral ligament of the metacarpophangeal joint of the thumb: preoperative diagnosis. *Journal of Bone and Joint Surgery* 1986; **68A**: 1320–6.

Mack G, Bosse M, Gelberman R. The natural history of scaphoid non-union. *Journal of Bone and Joint Surgery* 1984; **66A**: 504–9.

Mooney J, Siegel D, Koman L. Ligamentous injuries of the wrist in athletes. *Clinics in Sports Medicine* 1992; **11**: 129–39.

Pedowitz R, Hargens A, Mubarak S, Gershuni D. Modified criteria for the objective diagnosis of chronic compartment syndrome of the leg. *American Journal of Sports Medicine* 1990; **18**: 35–40.

Rettig A. Closed tendon injuries of the hand and wrist in the athlete. *Clinics in Sports Medicine 1990; 9: 85–109*.

Riester J, Baker B, Mosher J, Lowe D. A review of fracture scaphoid healing in competitive athletes. *American Journal of Sports Medicine* 1985; **13**: 159–61.

Rorabeck C, Fowler P, Nott L. The results of fasciotomy in the management of chronic exertional compartment syndrome. *American Journal of Sports Medicine* 1988; **16**: 143–6.

Ruby L, Stinson, Belsky M. The natural history of scaphoid non-union: a review of 55 cases. *Journal of Bone and Joint Surgery* 1985; **67A**: 428–32.

Stark H, Jobe F, Boyes J. Fractures of the hook of the hamate in athletes. *Journal of Bone and Joint Surgery* 1977; **59A**: 575–82.

Stener B. Displacement of the ruptured ulnar collateral ligament of the metacarpophangeal joint of the thumb: a clinical and anatomic study. *Journal of Bone and Joint Surgery* 1962; **44B**: 869–79.

Styf J. Diagnosis of exercise induced pain in the anterior aspect of the lower leg. *American Journal of Sports Medicine* 1988; **16**: 165–9.

Wehbe M, Schneider L. Mallet fractures. *Journal of Bone and Joint Surgery* 1984; **66A**: 658–69.

Whalen J, Bishop A, Linscheid R. Non operative treatment of acute hamate hook fractures. *Journal of Hand Surgery* 1992; **17A**: 507–11.

Knee

Blazina M, Kerlan R, Jobe F, Tibone R. Jumper's knee. *Orthopedic Clinics of North America* 1973; **4**: 665–78.

Cash J, Hughston J. Treatment of acute patellar dislocation. *American Journal of Sports Medicine* 1988; **16**: 244–9.

Cawley P, France E, Paulos L. The current state of functional knee bracing research: a review of the literature. *American Journal of Sports Medicine* 1991; **19**: 226–33.

Fairbanks T. Knee joint changes after menisectomy. *Journal of Bone and Joint Surgery* 1948; **30B**: 664–70.

Fowler P, Messieh S. Isolated posterior cruciate ligament in athletes. *American Journal of Sports Medicine* 1987; **15**: 553–7.

Garrick J, Requa R. Prophylactic knee braces. *American Journal of Sports Medicine* 1987; **15**: 471–6.

Gillquist J, Odensten M. Arthroscopic reconstruction of the anterior cruciate ligament. *Arthroscopy* 1988; **4**: 5–9.

Halbrecht J, Douglas W. Acute dislocation of the patella. In: Fox J, Del Pizzo W, eds, *The patellofemoral joint*. McGraw-Hill, New York, 1993.

Hardin G, Farr J, Bach B. Meniscal tears: diagnosis, evaluation, and treatment. *Orthopaedic Reviews* 1992; **21**: 1311–17.

Jackson D. The results of arthroscopic partial menisectomy in patients over 40 years of age. *Journal of Bone and Joint Surgery* 1982; **64B**: 481–5.

Jakob R. The arthroscopic meniscal repair: techniques and clinical experience. *American Journal of Sports Medicine* 1988; **16**: 137–42.

Johnson R, Benyon B, Nichlos C. Current concepts review: The treatment of injuries of the anterior cruciate ligament. *Journal of Bone and Joint Surgery* 1992; **74A**: 140–51.

Marks P, Fowler P. Imaging modalities for assessing the anterior cruciate ligament deficient knee. *Orthopedics* 1993; **16**: 417–24.

Martens M, Libbrecht, Burrssens A. Surgical treatment of the iliotibial band friction syndrome. *American Journal of Sports Medicine* 1989; **17**: 651–4.

Nottage W, Sprauge N, Auerbach B. Medial patellar plica syndrome. *American Journal of Sports Medicine* 1983; **11**: 211–13.

Shelbourne K, Nitz P. Accelerated rehabilitation after anterior cruciate ligament reconstruction. *American Journal of Sports Medicine* 1990; **18**: 292–9.

Taunton J, Clement D, Smart G. Non-surgical treatment of overuse knee injuries in runners. *Canadian Journal of Sports Science* 1987; **12**: 11–18.

Leg

Matheson G, Clement D, McKenzie D. Stress fractures in athletes: a study of 320 cases. *American Journal of Sports Medicine* 1987; **15**: 46–58.

Mubarak S. The medial tibial stress syndrome (a cause of shin splints). *American Journal of Sports Medicine* 1982; **10**: 201–5.

Foot and ankle

Attarian D. Biomechanical characteristics of human ankle ligaments. *Foot and Ankle* 1985; **6**: 54–8.

Black H, Brand R, Eichelberger M. An improved technique for evaluation of ligamentous injury in severe ankle sprains. *American Journal of Sports Medicine* 1978; **6**: 276–7.

Chrisman O, Snook G. Reconstruction of lateral ligament tears of the ankle. *Journal of Bone and Joint Surgery* 1969; **51A**: 904–12.

Cox J. Surgical and nonsurgical treatment of acute ankle sprains. *Clinical Orthopaedics and Related Research* 1985; **198**: 118.

Kannus P, Renstrom P. Treatment of acute tears of the lateral ligaments of the ankle. Operation, cast, or early controlled mobilization. *Journal of Bone and Joint Surgery* 1991; **73A**: 305–12.

Nelen G, Martens M, Burssens, A. Surgical treatment of chronic Achilles tendinitis. *American Journal of Sports Medicine* 1989; **17**: 754–9.

Nistor L. Surgical and nonsurgical treatment of Achilles tendon. *Journal of Bone and Joint Surgery* 1981; **63A**: 394–99.

Schepsis A, Leach R. Surgical management of Achilles tendinitis. *American Journal of Sports Medicine* 1987; **15**: 308–15.

Schepsis A, Leach R, Goryca J. Plantar fasciitis: etiology, treatment, surgical results, and review of the literature. *Clinical Orthopaedics and Related Research* 1991; **266**: 185–96.

46.2.1 Musculoskeletal trauma

Andrew H. Schmidt and Marc F. Swiontkowski

Introduction
Trauma center organization
Mechanism of trauma as predictor of injury
Systemic effects of skeletal injuries
Fractures and hemorrhagic shock
Early intervention for fractures
Head-injured patients
Femoral fractures and chest injury
Skeletal healing following trauma
Indirect reduction of fractures
Open fractures
Gunshot wounds
Vascular injuries
Limb salvage
Complications
Compartment syndrome
Thromboembolism
Blood-borne pathogens
Outcomes of musculoskeletal trauma
Upper extremity injuries
Shoulder girdle
Proximal humerus
Humeral shaft
Distal humerus
Elbow
Forearm
Wrist
Pelvis
Pelvic ring
Acetabulum
Lower extremity injuries
Hip dislocation
Femoral head fracture
Proximal femur
Femoral shaft
Distal femur
Knee
Tibia plateau
Tibial shaft
Distal tibia
Ankle
Foot
Further reading

Introduction

Musculoskeletal injury is a frequent accompaniment of blunt or penetrating trauma. Although brain, chest, or abdominal visceral injury may be life threatening, the long-term function of the patient is often determined by the severity of the associated spine, pelvis, or extremity injuries and the way that these injuries are managed. Furthermore, appropriate initial management of musculoskeletal injuries has a profound influence on the overall recovery of a multiply injured patient and reduces the rate of serious complications, including adult respiratory distress syndrome, fat embolism, and thromboembolic disease.

Trauma center organization

Multiply injured patients with musculoskeletal trauma should be managed at designated trauma centers. In the United States, patients transferred to trauma centers from other facilities have higher complication rates and longer hospital stays. This may be due to delay in treatment, higher rates of medical co-morbidities in indigent patients, and a higher frequency of psychiatric disease or substance abuse in transferred patients. In the United Kingdom, trauma systems have been less well developed than in the United States or Germany, but recent improvements in trauma system organization have led to better outcomes. Trauma centers in the United States are burdened by the continuing increase in patients while reimbursement is declining. Lack of public support is a major problem for the continued viability of trauma centers in both North America and the United Kingdom.

Alcohol, marijuana, and cocaine use are seen in over 50 per cent of orthopedic trauma patients, and substance abuse should be suspected in high-risk patients (males between 31 and 40 years, and those with injuries related to gunshots, altercations, and auto-pedestrian accidents). Drug-using patients tend to have more severe injuries and require longer periods in hospital. Pediatric and elderly patients require special attention, including specialized monitoring, nursing care, and rehabilitation. Despite the more frequent morbidity and mortality in geriatric patients following severe injury, the majority of those who survive return to independent living.

Mechanism of trauma as predictor of injury

There is a clear relationship between the mechanism of trauma and the resultant injury that is produced. In motor vehicle accidents, victims of lateral impact have more chest and abdominal injuries. Musculoskeletal injuries following vehicular accident are distinctly related to the type of trauma as well, with forearm fracture, pelvis fracture, and lower extremity fracture significantly more common after lateral impacts.

Systemic effects of skeletal injuries

Skeletal injuries have a profound impact on multiply injured patients, effecting cardiopulmonary physiology, pain levels, and the likelihood of complications. Early initiation of extracorporeal life support is a lifesaving procedure in patients with fractures who develop refractory pulmonary failure. The performance of splenectomy appears to increase the likelihood of infectious complications following internal fixation of fractures; splenic salvage should be attempted whenever possible.

Fractures and hemorrhagic shock

Pelvic and long-bone fractures are associated with potential blood loss of 1000 to 4000 ml or more. However, patients with closed femoral shaft fractures and hypotension should be evaluated for occult visceral injury; it should not be assumed that the femur fracture is the cause of blood loss. Similarly, in patients with pelvic fractures the major threat of hemorrhage is from non-pelvic sources.

Nevertheless, following traumatic pelvic disruption blood loss may be significant, and occurs from two sources: cancellous bone surfaces exposed by the fracture and pelvic vasculature, primarily the pre-sacral venous plexus. Hemorrhage is the primary cause of death after pelvic fractures in patients who do not have associated head trauma. Delay in the diagnosis and treatment of severe bleeding after pelvic fracture is life-threatening. The emergent management of a patient with a pelvic fracture and hypotension demands a multidisciplinary approach typically utilizing both angiography and orthopedic stabilization.

Understanding the likelihood of severe bleeding begins with classifying the pelvic injury based on its mechanism. Pelvic injuries have been classified into three types depending on the nature of the applied forces: lateral compression, anteroposterior compression, and vertical shear. Each major type is subdivided into more specific patterns depending on the severity of the injury (Table 1). Each subtype of fracture may be related to a specific risk of major bleeding, based on anatomic relationships

between the large vessels of the pelvis and the disrupted osseous and ligamentous structures.

Type	Anterior lesion	Posterior lesion
LC-I	Transverse pubic rami fractures	Sacral compression fracture
LC-II		Iliac wing (crescent) fracture
LC-III		Contralateral open-book injury
APC-I	Symphysis diastasis	Slight widening of sacroiliac joint
APC-II	Symphysis diastasis or vertical pubic ramus fracture	Widened sacroiliac joint, and:
APC-III		Posterior ligaments intact
		Posterior ligaments torn
VS	Symphysis diastasis or vertical pubic ramus fracture	Vertical displacement of both anterior and posterior ring
CM	Combined	Combined

Table 1 Classification of pelvic ring injuries (see Burgess et al., 1990)

The greatest risk of arterial injury occurs in anteroposterior compression type II and III injuries, involving the superior gluteal and internal pudendal arteries. A significant risk of exsanguination is also present following vertical shear and combined injury patterns.

Intra-abdominal and intrathoracic injury accompany pelvic fractures in up to 55 per cent of patients. Patients with pelvic fracture have an incidence of ruptured thoracic aorta that is several times greater than seen in the general population of trauma patients, especially after anteroposterior compression type injuries.

Peritoneal lavage produces false-positive results in up to 29 per cent of cases of pelvic fracture. In order to make this maneuver more accurate, diagnostic peritoneal lavage should be performed above the umbilicus in patients with pelvic fractures. Laparotomy may actually exacerbate bleeding from retroperitoneal sources in these patients, and is not advised unless other life-threatening intra-abdominal injury is present. Finally, early stabilization of the pelvis with external fixation lessens bleeding by stabilizing bony surfaces and reducing intrapelvic volume. Specialized pelvic resuscitation clamps are available that can be rapidly applied in the field, emergency department, or operating room to a patient with hemodynamic collapse. They address displacement of the posterior pelvic ring that cannot be dealt with using standard anterior external fixation.

An understanding of these principles leads to a logical treatment algorithm for patients with hypotension and pelvic fracture (Fig. 1). First, the pelvic injury is classified as high risk or low risk based on mechanism and displacement. In high-risk injuries with suspected catastrophic abdominal or thoracic injuries, time must not be wasted and emergent angiography of the pelvis, abdomen, and possibly chest must be performed. Embolization of bleeding pelvic vessels is performed and associated intra-abdominal and thoracic vascular injuries may be identified quickly. Aggressive fluid resuscitation, treatment of associated extrapelvic injuries, and selective angiography with embolization leads to a low overall mortality rate, with deaths usually due to associated injuries. If possible, temporary external fixation of the pelvis is performed as an adjunct, but should not delay angiography. This can be accomplished in the angiography suite under fluoroscopy during the preparations for angiography. In high-risk fractures without associated chest or abdominal injury, immediate anterior external fixation (for anterior ring displacement) or posterior pelvic clamp application (for posterior pelvic ring displacement) often stops the bleeding and angiography is not necessary. Definitive pelvic fixation, usually by open reduction and internal fixation, is delayed until the patient is stable.

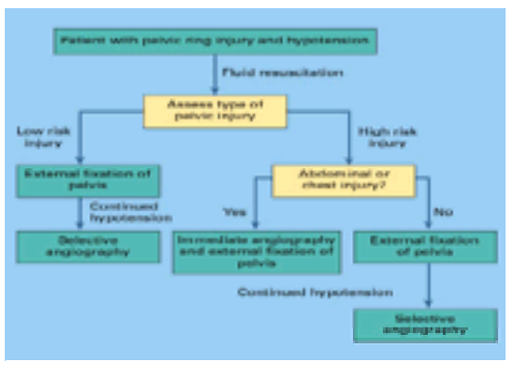


Fig. 1. Algorithm for management of hypotension associated with pelvic ring disruption.

Early intervention for fractures

Patients with lower extremity and pelvic fractures are at high risk of developing complications related to immobilization, including thromboembolism, fat embolism, and sepsis. Pulmonary manifestations of this phenomenon vary from mild hypoxia to adult respiratory distress syndrome. Closed femoral fractures cause significant immunosuppression and altered gastrointestinal permeability in a murine model. Many comparative studies now demonstrate that these complications are best prevented by the early, definitive stabilization of fractures. Early fracture fixation is considered safe in burn patients, when bacterial wound colonization is lowest.

Head-injured patients

Patients with severe brain injuries and fractures are challenging to manage. Many neurosurgeons are reluctant to allow patients with brain injury to undergo orthopedic procedures because of concerns about the prolonged duration of anesthesia, increased blood loss, hypotension, and pain contributing to further brain injury. Comparative studies have provided conflicting results. One series reported lower mortality and better functional outcome in patients operated for fractures on the day of injury, while another series documented lower discharge Glasgow Coma Scores compared with similar patients undergoing fixation 24 h or more after injury. The ability to predict survival following head trauma is difficult, with 46 per cent of patients expected to die surviving and 20 per cent making a full recovery. Such patients must not have appropriate, early, definitive care of their musculoskeletal injuries withheld. If secondary brain injury due to hypotension and hypoxia are avoided, early operative intervention for fractures adds little additional risk, and should be expected to decrease the likelihood of complications and improve outcome. Patients with severe head injuries and fractures, for whom intraoperative hypotension may not be avoidable, should have definitive fracture stabilization delayed until they are considered stable.

Femoral fractures and chest injury

Investigators have determined that intramedullary nailing has profound effects on pulmonary physiology. Intramedullary nailing increases the intramedullary pressure, leading to extravasation of fat and bone marrow. In addition, polymorphonuclear leukocyte activation and increased capillary permeability are seen. These effects may be exacerbated by associated pulmonary trauma. A clinical series demonstrating increased mortality in patients with chest injury following reamed femoral nailing led some surgeons to suggest that femoral fractures should be treated by unreamed nails or plate fixation in these patients. Subsequent animal and clinical studies have clarified that the severity of the pulmonary injury determines mortality and not the method of treatment of the femoral fracture. Nevertheless, in patients with severe chest injury such as bilateral pulmonary contusion, reaming should be minimized in order to lessen the additional pulmonary insult. Use of a pneumatic tourniquet for lower leg surgery after reamed femoral nailing increases the incidence of pulmonary morbidity, attributed to additional pulmonary microvascular injury following limb reperfusion.

Skeletal healing following trauma

Significant advances have been made in our understanding of the biology and mechanics of fracture healing. Within the last decade, the importance of the soft tissues and the maintenance of bone vascularity have been appreciated, leading to marked changes in the approach to the internal fixation of fractures. Since the publication of the open fracture classification by Gustilo *et al.* in 1976, it is widely recognized that healing, complications, and function are directly related to the severity of the soft-tissue injury. The recognition that closed fractures may be associated with serious injury to the soft tissues as well has led to a useful classification of closed

fractures. Both closed and open fractures may be subdivided into low-energy and high-energy injuries (Table 2). Low-energy injuries, whether open or closed, follow a benign clinical course and have a good prognosis. Closed, high-energy injuries are prone to all of the complications usually associated with open fractures. This recognition of this has been another important advance in fracture management.

Low-energy injuries	*High-energy injuries*
Closed	Closed
Tscherne types 0–1	Tscherne types 2–3
Open	Open
Gustilo grades 1–2	Gustilo grades IIIA, IIIB, IIIC

Table 2 High-energy versus low-energy injuries

The initial techniques of open reduction and internal fixation, advanced by the AO/ASIF (Arbeitsgemeinschaft für Osteosynthesefragen/Association for the Study of Internal Fixation), emphasized anatomic reduction and rigid internal fixation of all fragments. Primary bone healing was possible under these circumstances. Difficulty was encountered when adapting these principles to high-grade closed and open fractures, where the associated soft-tissue injury led to wound complications and delayed fracture healing. Concomitantly, the success of intramedullary nailing and of closed functional treatment of diaphyseal fractures demonstrated that open, anatomic, and absolutely rigid fixation of many fractures is not necessary. Therefore, the concept of fracture fixation has evolved from one of maximal stability to one of optimal fixation.

A common observation in patients with head injuries is the rapid and abundant formation of callus that occurs at fractures and at other sites of blunt or surgical injury. Patients with head injuries exhibit both a more rapid time to union as well as a greater healing response. Moreover, in head-injured patients, the callus appears to be more mature at the periphery, suggesting that the underlying mechanism is different than in normal fracture healing and is akin to the process of heterotopic ossification. A humoral substance released by injured brain tissue may be responsible for this phenomenon.

Indirect reduction of fractures

The soft tissues adjacent to an injured bone are extremely important in maintaining the blood supply to fracture fragments, and that in many cases, soft-tissue attachments may facilitate the reduction of a fracture when traction is applied—the principle of ligamentotaxis. The reduction requirements for specific fractures have been further elucidated (Table 3). For diaphyseal injuries, restoration of length and alignment (in three planes) is all that is necessary. Internal fixation must produce limited additional soft-tissue injury. Metaphyseal injuries may require internal fixation in order to restore the supportive buttress of this area of bone, but again, anatomic reduction of all fragments is not required. Only articular fractures require complete anatomic reduction, and this must be done without causing further devascularization.

Diaphysis	*Metaphysis*	*Articular*
Length	Support	Anatomic
Alignment		
Rotation		

Table 3 Reduction requirements of fracture types

Using these principles, complex fractures are now treated by limited internal fixation of articular injuries, with closed or open reduction of the joint surface done through carefully planned, direct incisions that do not require stripping of periosteum or retraction of skin. Metaphyseal injuries are supported by external fixation or low-profile buttress plates. Diaphyseal fractures are treated by intramedullary nails or limited contact plates, utilizing indirect reduction maneuvers (Fig. 2).



Fig. 2. (a–c) Comminuted intertrochanteric and subtrochanteric femur fracture treated with open reduction and internal fixation using a dynamic condylar screw and indirect reduction of the comminuted segment of the shaft. (a) Comminuted proximal femur fracture. (b) Anteroposterior view taken 18 months after injury, showing union with complete consolidation of the comminuted segment. Rapid union occurred as a result of the indirect reduction of this area by restoration of length and stable fixation, with preservation of soft-tissue attachments and vascular supply. (c) Lateral view taken 18 months after injury.

Open fractures

Open fractures are a surgical emergency. Immediate intervention consists of surgical debridement, bone stabilization, and appropriate antibiotic administration. Delayed soft-tissue closure and/or reconstruction are performed in the first week. Secondary bone reconstruction depends on the adequacy of the soft-tissue envelope.

Open fractures are classified according to the criteria of Gustilo *et al.* The risk of infection, non-union, and amputation is related to the severity of the soft-tissue injury (Table 4). Emergency room management consists of documenting the neurovascular status of the injured limb, covering the wound with a sterile dressing, reducing and splinting the fracture, and administering tetanus prophylaxis and intravenous antibiotics. Despite the fact that most infections following open fracture are nosocomial, early antibiotic administration reduces the risk of infection. The choice of antibiotic depends on the severity of the injury and is predicated on knowledge of the likely pathogens. Early wound cultures are not predictive of late infection and are not recommended unless special circumstances are present, such as a soil-contaminated wound or a marine injury.

Type	Sepsis	Amputation	Non-union
I	0–7 per cent	0	3 per cent
II	1–11 per cent	0	9 per cent
IIIA	0–18 per cent	0	9–27 per cent
IIIB	12–52 per cent	16–36 per cent	20–43 per cent
IIIC	5–42 per cent	8–87 per cent	31–100 per cent

Table 4 Complications versus open fracture grade

Immediate surgical management is crucial to avoid infectious complications. Debridement must be done systematically, removing all non-viable skin, fascia, muscle, and bone. There does not appear to be any value in providing antibiotic prophylaxis for longer than 24 h after each procedure. In fact, prolonged administration of prophylactic antibiotics to patients with both open fractures and blunt chest trauma increases the incidence of pneumonia and the incidence of resistant bacteria.

Skeletal stabilization must be achieved at the initial operation. In most cases, definitive internal fixation with plates or intramedullary devices may be safely performed following adequate debridement. In cases of severe contamination or high-energy soft-tissue injury, provisional external fixation is carried out. In these cases, definitive internal fixation is delayed until the soft tissues have stabilized. In all cases of open fracture, traumatic wounds should be left open and subsequent closure performed in 2 to 3 days. The antibiotic bead-pouch technique has been effective in limiting deep wound infection (Fig. 3). In severe injuries, early flap coverage reduces the incidence of sepsis.



Fig. 3. (a–c) Comminuted open distal tibia fracture/dislocation treated with limited internal fixation of the articular surface, external fixation of the metaphysis, and bead-pouch technique. (a) Clinical photograph of grade IIIA open ankle distal tibia fracture/dislocation. (b) Anteroposterior radiograph showing the displacement of the distal tibia articular surface and the metaphyseal comminution. (c) Postoperative view showing the reduction of the ankle roof with limited internal fixation, external fixation of the metaphysis, and placement of tobramycin-impregnated methacrylate beads.

Gunshot wounds

Gunshot injuries are increasing rapidly in urban regions. A recent audit of the orthopedic surgery service at a major North American trauma center revealed that patients with gunshot wounds comprised 24 per cent of all admissions, 33 per cent of the average inpatient census, and accounted for 14 per cent of all surgeries performed. Fortunately, due to stricter gun control, European countries are not experiencing this problem to the same degree.

Gunshot wounds are classically divided into low-velocity and high-velocity injuries. Although this distinction is simplistic, it remains useful as a rough guide to treatment. However, knowledge of the muzzle velocity or caliber of the weapon used does not supplant careful examination of the patient prior to deciding upon treatment. Low-velocity injuries result from handguns, while rifles and shotguns cause high-velocity injuries. Injuries caused by low-velocity gunshot wounds are considered benign, with a low risk of infection. Low-velocity gunshot injuries, with little tissue damage and simple fractures, are treated by local debridement of the entrance and exit wounds. If present, associated fractures are treated closed or with internal fixation as dictated by the fracture pattern. Oral antibiotics appear to be sufficient in a randomized study comparing 3 days of intravenous cephapirin and gentamicin to 3 days of oral ciprofloxacin. The infection rate was 2 per cent in both groups. Antibiotics should be started promptly.

High-energy gunshot injuries result in severe soft-tissue damage and highly comminuted fractures. The zone of injury may extend well beyond the wounds due to cavitation effects within the soft tissue. The degree of tissue injury depends on many factors, including the velocity and size of the missile, the degree of deformation of the bullet, and whether the bullet yaws or fragments within the tissues. In high-energy injuries, the bullet will appear fragmented on radiographs, the wounds will be larger, the fractures more severely comminuted, and the soft tissues more swollen. Significant contamination and necrosis of skin and muscle may be present. These injuries demand serial debridements and open fracture care as would be provided to any high-grade open fracture.

Vascular injuries

Vascular injuries may occur with virtually any type of extremity trauma. The incidence of vascular trauma varies from up to 40 per cent following knee dislocation to 3 per cent following long bone fracture. Gunshot wounds result in a 38 per cent incidence of arterial injury. Vascular injuries of the lower extremity more often result in limb-threatening ischemia because of fewer collateral vessels and greater muscle mass.

The treatment of a vascular injury involves the control of hemorrhage, prompt treatment of ischemia, and the prevention of associated complications such as compartment syndrome, aneurysm, arteriovenous fistula, and thrombosis. The outcome following vascular injury is directly related to the duration of ischemia, with 6 h being the upper limit. After this time, irreversible changes occur within the ischemic tissues.

Prompt diagnosis depends on a high index of suspicion and careful neurovascular examination. Many patients present with cold, pulseless extremities following trauma; the majority do not have vascular injury. The patient must be warmed and resuscitated with fluids, and associated fractures grossly aligned, before a loss of pulse is attributed to arterial injury. Doppler measurements of pulse and blood pressure are important adjuncts. A Doppler-derived ankle–brachial systolic pressure ratio of less than 0.9 has an accuracy of 95 per cent for diagnosing vascular injury. Although arteriography may be considered the ‘gold standard' for the diagnosis of vascular lesions, it has a very low yield following extremity trauma in the absence of Doppler pulse deficits and delayed capillary refill.

When vascular injury is diagnosed by absent Doppler pulses and the site of injury is known, immediate arterial exploration and repair are warranted. Generally, restoration of circulation takes precedence over orthopedic interventions, although a temporary shunt may be utilized. Rapid stabilization of the limb with an external fixator is useful to protect the vessels during and after repair, until definitive fixation can be performed. If a partial pulse deficit exists or an occult lesion is suspected, either ultrasonography or arteriography may be performed. Presently, lesions such as intimal tears may be followed serially in the absence of clinical compromise.

The late outcome following vascular repair in patients with extremity trauma is related primarily to the associated soft-tissue injuries. Following tibia fracture, the incidence of delayed or non-union is increased in patients with associated posterior tibial artery injury.

Limb salvage

Whether to proceed with salvage of a severely traumatized extremity or to proceed with amputation is a difficult clinical decision. Unsuccessful attempts at limb salvage

may doom a patient to multiple surgeries, numerous complications, and years of disability, chronic pain, and financial and emotional hardship. Amputation and use of a limb prosthesis may result in stump problems, phantom limb pain, and the need for expensive prostheses.

Objective criteria have been proposed to help clinicians decide when amputation is appropriate. Absolute criteria for amputation of the traumatized leg include warm ischemia time greater than 6 h, tibial nerve laceration, and irreparable vascular lesions. Relative indications for early amputation include severe crushing injuries of the foot associated with severe leg trauma, segmental tibia fractures associated with vascular lesions, and severe associated life-threatening injuries. For femoral or popliteal artery lesions, greater than two long-bone fractures is predictive of amputation. For tibial arteries, one-, two-, or three-vessel injuries are associated with amputation rates of 20, 22, and 100 per cent, respectively.

Recently, several limb-salvage scoring systems have been proposed to assist in clinical decision making. The most widely used system is the Mangled Extremity Severity Score, or MESS (Table 5). In the initial report, a MESS score greater than 7 was 100 per cent accurate in predicting the need for eventual amputation. Subsequent work by other investigators has suggested that a MESS of 9 may have a better ability to discriminate unsalvageable injuries. Although each of these scores are claimed by their proponents to have the ability to discriminate salvageable injuries from those destined to require amputation, prospective comparative evaluation has not confirmed that they perform consistently enough to be relied upon. Thus, decisions regarding amputation must be done on an individual basis.

Criteria		*Score*
Skeletal/soft tissue injury		
	Low energy	1
	Medium energy	2
	High energy	3
	Very high energy	4
Limb ischemia		
	Pulse reduced or absent but normal perfusion	1
	Pulseless; paresthesias; reduced capillary refill	2
	Cool, paralysed, insensate, numb (score doubled for ischemia >6 h)	3
Shock		
	Systolic blood pressure maintained >90 mmHg	0
	Transient hypotension	1
	Persistent hypotension	2
Age (years)		
	<30	0
	30–50	1
	>50	2

Table 5 Mangled Extremity Severity Score (see Johansen et al., 1990)

Long-term functional outcome studies have been reported that provide useful information about the results of limb salvage and amputation. There are significant advantages and disadvantages to each approach. Patients that undergo amputation require fewer surgeries and have a shorter rehabilitation time. The functional differences between amputees and patients with salvaged limbs are less clear. In one report from Switzerland, patients with salvaged limbs had fewer lifestyle changes, were more likely to return to their job, and drew less disability income. A similar study performed in the United States reported different conclusions, with successfully salvaged patients being less willing or able to work. A quality of life evaluation, using the Nottingham Health Profile and the General Well-Being Schedule, showed that significantly more patients who had limb salvage considered themselves disabled. The long-term financial costs appear to be similar; limb salvage is more expensive in the short-term, but amputees incur significant long-term costs associated with prosthetic wear.

Complications

When treating patients with trauma, physicians must not only deal with the injuries present, but must care for ensuing complications. The treatment of complications starts with prevention, and knowledge of the complications that a trauma patient is at risk for is of paramount importance. In addition to complications related directly to the injury, trauma patients are susceptible to numerous medical complications, including shock, respiratory distress syndrome, multiple organ system failure, thromboembolism, clotting disorders, and immunologic, endocrine, and metabolic dysfunction.

Proper care of a trauma patient begins with a complete assessment of all injuries. Four to 6 per cent of musculoskeletal injuries are diagnosed late or missed altogether during the initial assessment. Missed injuries occur most often in the most severely injured patients and are most often located in the hands or feet. However, delayed diagnosis of injuries about the hip and knee often is an important cause of poor outcomes. The most common cause of missed diagnoses are problems related to radiographs: radiographs are not performed, are of poor quality, or are misinterpreted.

The maintenance of adequate nutrition is important for the trauma patient. Nutritional requirements are several-fold greater following trauma; patients with multiple fractures may need 6000 cal/day. Paradoxically, many trauma patients are unable to eat, and rapidly develop signs of malnourishment. Protein and calorie malnutrition may affect wound and fracture healing, as well as reduce immunocompetence. Malnutrition is avoided by early assessment and monitoring, with early enteral or parenteral supplementation. A strong association between decreased serum ionized calcium levels and the development of severe musculoskeletal complications such as sepsis and fat embolism exists.

Compartment syndrome

Compartment SYndrome is myoneural ischemia due to elevated intracompartmental pressures. The sequelae of untreated compartment syndrome are neural dysfunction and contractures. Compartment syndrome is a potentially limb-threatening complication of extremity injury. It is most common in association with both closed and open tibia fractures, but may also occur following crush injuries without fracture, prolonged ischemia followed by reperfusion, or extrinsic compression such as from a cast or medical anti-shock trousers. Post-ischemic swelling due to reperfusion is associated with compartment syndrome that develops following repair of vascular injury, with ischemia due to prolonged pressure, and following surgery in the hemilithotomy position. Compartment syndrome has been described in the deltoid, arm, forearm, hand, gluteal compartments, thigh, leg, and foot.

The diagnosis of compartment syndrome depends upon a high degree of suspicion and careful, repeated clinical evaluation. The hallmark is pain out of proportion to the injury, and pain with passive stretching of the involved muscles. The involved compartment is usually quite tense when palpated. Neural dysfunction manifested by sensory loss represents a later finding, and when present is indicative of an advanced degree of pathology.

Compartment SYndrome is caused by elevated intracompartment pressures, which may be directly measured to confirm the diagnosis. Real-time monitoring of compartment pressures is also advocated in high-risk patients who are sedated, head-injured, or intoxicated, in whom physical examination is unreliable. In the past, absolute pressures greater than 30 to 40 mmHg were thought to indicate compartment syndrome. Presently, a differential pressure (diastolic or mean arterial pressure minus compartment pressure) of less than 30 to 40 mmHg is considered more specific than an absolute pressure for the diagnosis of compartment syndrome.

Once diagnosed, the only acceptable treatment is urgent fasciotomy of involved compartments (Fig. 4). Medical therapies, including hyperbaric oxygen and intermittent plantar compression, may serve to limit myonecrosis, but are not in themselves adequate treatment. Adequate skin release must be performed in addition to myofascial release. Wounds are left open until swelling resolves, and are then closed or skin grafted.

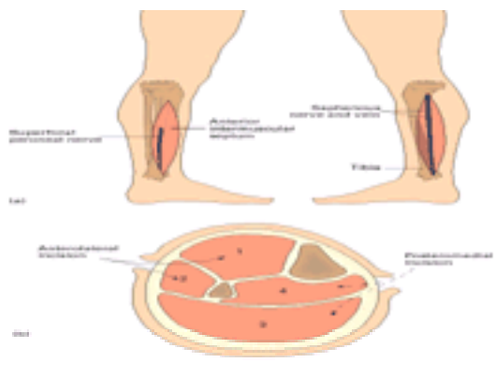


Fig. 4. Two incision technique of fasciotomy in the lower extremity for management of compartment syndrome. (a) Long lateral and medial incisions are required, with care to avoid neurovascular structures at risk. (b) Cross-section of the leg depicting the release of two compartment from each incision. Through the lateral incision, the anterior and lateral compartments may be released on each side of the intermuscular septum. On the medial side, the superficial and deep posterior compartments are

readily released as shown.

Thromboembolism

Orthopedic trauma patients are at a very high risk for thromboembolic complications. Over 30 years ago, post-mortem examinations of patients who had died following injuries or burns revealed iliofemoral thrombosis in 65 per cent and popliteal vein thrombosis in 30 per cent. The highest risk seems to be in patients with pelvic and lower extremity trauma, followed by spine injury, head injury, and face/chest/abdominal injury. Proximal deep vein thrombosis occurs in 25 per cent to 35 per cent of patients with pelvic fractures. Symptomatic pulmonary embolism is seen in 2 per cent to 10 per cent, and fatal pulmonary embolism happens in 0.5 per cent to 2 percent of cases of pelvic trauma.

Given that trauma patients are at high risk for the development of deep-vein thrombosis and its complications, prevention of thromboembolism is of key importance. Various methods have been used, including anticoagulation, mechanical devices, patient/limb mobilization, and vena cava filters. All of these methods may be difficult to manage in the polytrauma patient. Pharmacologic anticoagulation may be contraindicated because of the risk of further hemorrhage in patients with head or visceral injuries, with fractures that are prone to bleeding complications, or in patients awaiting major surgery. When anticoagulation is possible, the method of choice remains debatable. Low-molecular weight heparin compounds are proven to be effective at reducing the incidence of thromboembolism, but are associated with a mild increased risk of bleeding complications. Other choices include low-dose heparin, adjusted-dose heparin, and coumadin.

For many trauma patients, mechanical devices such as intermittent pneumatic compression stockings or plantar compression devices are the best choice. A prospective, randomized study compared pneumatic compression alone with no prophylaxis in over 300 patients with hip and pelvic fractures. This regimen was effective in the patients with hip fractures, but not in those with pelvic fractures. In select high-risk patients, placement of vena cava filters may be performed prophylactically. In a study of 35 patients with pelvic or lower extremity fractures, vena cava filters were placed preoperatively and patients were maintained on low-dose coumadin for 6 to 12 weeks after surgery. There were no instances of symptomatic pulmonary embolism. Complications related to the vena cava filters occurred in one-fourth of the patients, but none caused symptoms. In another study, a significant decrease (from 2.5 per cent to 0.2 per cent) in the incidence of pulmonary embolism in orthopedic trauma patients was documented after the institution of a protocol calling for the use of vena cava filters in high-risk patients. The long-term patency rate of the filters was 94 per cent.

Thus, a reasonable approach to the management of thromboembolic complications in orthopedic trauma patients is to provide immediate mechanical prophylaxis for all. Those patients that are not at significant risk for bleeding may be started on a regimen of low-dose coumadin or low-molecular weight heparin. High-risk patients may be screened with duplex ultrasonography, and vena cava filters used in selected patients.

Blood-borne pathogens

Human immunodeficiency virus (HIV) infection is thought to be more prevalent in trauma patients than in the general population. Studies have shown that the prevalence ranges from 0.8 per cent in San Antonio, Texas to 14.6 per cent in Baltimore. Exposure to HIV may be considered an occupational risk for orthopedic surgeons, whose annual risk of seroconversion has been estimated to be from 0.025 per cent to 12 per cent. The risk of seroconversion after a single exposure is estimated to be approximately 0.3 per cent for hollow-bore needles, and less for solid needles such as those used for suturing. Despite theoretical risk, to date there have been no instances of documented HIV seroconversion following solid needle puncture or aerosol transmission. Following so-called ‘Universal Precautions' best prevents HIV transmission, in which all blood and body fluids are considered to be contaminated with potentially infectious agents. Specific guidelines have been published by the American Academy of Orthopedic Surgeons regarding risk prevention. In addition to HIV, orthopedic surgeons are at risk for exposure to viral hepatitis and tuberculosis. Routine hepatitis B vaccination and annual testing for exposure to tuberculosis are recommended. Despite the wide availability of appropriate barrier precautions in the United States, a disturbing high percentage of orthopedic surgeons reported poor compliance with these guidelines in a survey done by the Orthopedic Trauma Association.

Asymptomatic patients with HIV are at increased risk of infectious complications following trauma and orthopedic surgery. The risk is especially severe after open fractures, when seropositive patients had a 55.6 per cent infection rate compared with 11.3 per cent in seronegative controls.

Outcomes of musculoskeletal trauma

The orthopedic literature has been criticized for its lack of well-designed studies and frequent use of physician-derived, unvalidated assessment tools. This results in a lack of agreement among orthopedic surgeons regarding the efficacy of different methods of treatment, and is partly responsible for the regional variation in rates of hospitalization for various conditions.

Increasing emphasis is being placed on outcomes research in order to understand the effectiveness of treatment rendered. Numerous articles are now appearing that document the ‘outcome' of trauma using accepted and validated assessment tools and appropriate methodology. Several validated musculoskeletal assessment tools now exist for analysis of outcome. The Short Form 36 (SF-36) Health Survey has been used in several studies to assess the outcome of pelvic injuries norms. Two to 4 years after injury, patients with severe pelvic fractures have poor SF-36 scores in regard to physical functioning, physical role, bodily pain, and general health, compared with age-specific norms. Fourteen per cent impairment in physical outcome and 5 per cent impairment in mental outcome were documented, compared with normative data.

The Sickness Impact Profile (SIP), another validated health assessment tool, has been used in several studies reporting the outcome of trauma. Although the extent of disability after lower extremity trauma improved between 6 and 12 months following injury, significant impairment persisted, even in psychosocial functioning.

Upper extremity injuries

Arm injuries may occur in isolation following minor trauma such as a fall, or in association with other injuries in a polytrauma patient. In general, evaluation and treatment of life-threatening injuries to the head, chest, abdomen, and pelvis are done before most upper extremity injuries. However, the initial survey of a trauma patient must identify possible life-threatening or limb-threatening injuries that involve the upper extremity. Once the patient is stabilized, a careful follow-up examination must be done in order to diagnosis occult injuries, typically to the carpus and hand that may be missed initially.

Definitive treatment of upper extremity injuries is initiated following complete evaluation. Limb-threatening injuries assume the highest priority. Dislocations are reduced as soon as possible. Open fractures must be urgently irrigated, debrided, and appropriately stabilized. Closed fractures can be splinted until definitive treatment can be rendered.

Shoulder girdle

A potentially life-threatening upper extremity injury is scapulothoracic dissociation, which represents a closed avulsion of the arm and is usually associated with axillary artery and brachial plexus injury. The hallmark of this injury is lateral displacement of the scapula on radiographs, which may be missed if not specifically looked for. Occasionally, simple fractures of the clavicle and scapula or dislocations of the shoulder are associated with neurovascular injury as well.

Scapular fractures indicate significant trauma, often associated with serious thoracic injuries. Isolated scapula body fractures are of little importance and do not require specific treatment other than early shoulder motion exercises. Intra-articular fractures of the glenoid process and unstable fractures of the glenoid neck should be considered for open reduction and internal fixation. Computed tomography (CT) is often beneficial in determining the presence and severity of articular injury.

The concept of the superior suspensory complex of the shoulder has been proposed. This provides a framework for evaluating glenoid neck fractures and associated injuries to the clavicle, coracoid, acromioclavicular joint, or scapula. Double disruptions of the superior shoulder suspensory mechanism are unstable injuries and should be treated by internal fixation of some or all components of the injury.

Dislocations of the sternoclavicular joint may be difficult to diagnose on radiographs. CT is recommended when this injury is suspected. Posterior dislocations may cause respiratory or vascular compromise, and should be identified and reduced promptly. Closed reduction is performed with the patient supine and a rolled towel between the scapulae. Posterior pressure on both anterior shoulder regions distracts the dislocated joint. If this method does not reduce the deformity, anterior traction on the medial clavicle with a percutaneous towel clip may be required. The reduction is usually stable. Anterior dislocations do not compress the mediastinum, but are less stable. Recurrence of anterior sternoclavicular subluxation is common after closed reduction. Any deformity resulting from the prominent anteriorly displaced

medial clavicle is usually asymptomatic. As the medial clavicular growth plate fuses in the early twenties, many apparent sternoclavicular dislocations are actually physeal injuries.

Fractures of the clavicle are common, and the outcome seems to be related to the location of the fracture and the degree of displacement. Following middle-third fractures, healing occurs in over 95 per cent of patients, although frequently with some deformity. The effect of malunion due to shortening on functional outcome is controversial. Surgery is generally performed only for open clavicle fractures, those associated with impending skin necrosis, or those associated with neurovascular injury. Recently, open reduction and internal fixation of displaced middle-third fractures has been recommended on the basis of superior functional results and a non-union rate of 20 per cent in the subgroup of fractures with greater than 2 cm of shortening at presentation. However, non-operative management with a figure-of-eight strap remains the traditional treatment. Clavicle fractures caused by high-energy trauma have a higher rate of non-union, which should be treated by plate fixation and bone grafting if symptomatic. Fractures of the distal third of the clavicle have a higher incidence of problems. Displaced fractures that occur between the coracoclavicular ligaments have an especially high rate of non-union and require close clinical follow-up. Surgical stabilization of displaced fractures in this region results in a high rate of union.

Acromioclavicular joint separations are graded by the degree of displacement of the distal clavicle, which is related to the extent of disruption of the capsular and coracoclavicular ligaments. Completely displaced injuries are further subdivided depending on the nature and severity of displacement, so that six types of acromioclavicular joint separation are recognized. Surgical intervention, including reduction of the acromioclavicular joint and repair of the capsule and deltotrapezial fascia, is recommended for types IV, V, and VI. The concomitant use of internal fixation to augment the repair, even temporarily, is controversial. For the lesser grades of injury (types I, II, and III) early functional rehabilitation is emphasized, and a good functional result expected in spite of residual distal clavicular prominence.

Anterior shoulder dislocations are usually obvious because of pain and deformity. Radiographs (anteroposterior and axillary lateral) are essential to diagnose associated fractures. Posterior dislocations are less common and missed in up to 50 per cent of cases. In cases of posterior dislocation, the anteroposterior radiograph may appear unremarkable, as medial displacement of the humeral head is slight. An axillary lateral or trans-scapular lateral radiograph will confirm or disprove this diagnosis, which must be considered whenever a patient lacks normal external rotation of the shoulder. Shoulder dislocations must be reduced promptly after evaluation and radiographic confirmation, using relaxation and traction in line with the humeral shaft. Neurologic injury, particularly to the axillary (circumflex) nerve is common. Lateral shoulder sensation and contraction of the deltoid muscle must be evaluated. Avulsion fracture of the greater tuberosity may be associated with anterior shoulder dislocation. If residual superior or posterior displacement of the tuberosity persists after reduction, the tuberosity and attached rotator cuff tendon must be repaired. Patients with recurrent shoulder dislocations merit surgical repair, involving reattachment of the anterior–inferior labrum (Bankhart lesion) to the glenoid.

Proximal humerus

Eighty-five per cent of proximal humerus fractures are minimally displaced and are treated with early passive motion exercises. The management of displaced fractures is complex and predicated on a thorough understanding of the fracture and its effects on rotator cuff function, shoulder biomechanics, and humeral head viability. Four typical fragments are seen: humeral head, humeral shaft, and greater and lesser tuberosities. Associated glenohumeral dislocations must be identified. Plain radiographs (including axillary lateral) are essential; CT is useful in difficult cases.

Proximal humerus fractures are classified according to Neer, recognizing that this scheme, like most others, is subject to poor intra-observer and inter-observer variability. The fractured parts must be identified, and their displacements determined, based on imaging studies. Displaced fractures of the tuberosities must be repaired to limit subacromial impingement and resulting pain. Fractures of the surgical neck with varus angulation result in poor functional outcome, and should be reduced. Two- and three-part fractures have a low incidence of avascular necrosis, and may be fixed by percutaneous pins, tension band wires, buttress plates, intramedullary devices, or a combination of these. The fixation method chosen for a given patient depends on the severity of the fracture, bone quality, patient demands, surgeon experience, and device availability. Four-part fractures have a high risk of complete avascular necrosis of the humeral head, and are generally treated by early prosthetic replacement in older patients. The exception is the valgus-impacted fracture, which is amenable to limited internal fixation. Prosthetic replacement is also recommended for displaced anatomic neck fractures (except in the young, active patient), and head-splitting fractures.

Humeral shaft

Isolated humeral shaft fractures are best managed non-operatively, because of union rates greater than 95 per cent. Radial nerve injuries complicate fractures of the middle and distal thirds of the shaft in 10 to 15 per cent of cases. Most such injuries are neurapraxias and nerve function recovers spontaneously. Even secondary palsies that occur during fracture manipulation usually recover, and operative exploration is not warranted. Exceptions that require early nerve exploration are open fractures, and closed fractures with bayonet apposition. The results of late nerve decompression or repair of the radial nerve are quite good.

Successful non-operative management of closed humeral fractures requires following several principles. The patient must be kept in an upright position so that gravity can align the fracture. The fractured humerus should be restrained to the chest with sling and swath to control rotation. Healing is expedited by early functional use of an adjustable fracture brace, which must be tightened as swelling and discomfort diminish. Healing with acceptable, although rarely anatomic alignment usually occurs within 6 to 8 weeks.

Surgical fixation of humeral shaft fractures is recommended in special situations: obese patients, open fractures, multiply injured patients with pelvic, lower extremity, or other upper extremity injuries, associated vascular injuries, pathologic fractures, and non-unions. Currently, rigid interlocking nails or plate fixation is recommended. Both seem to be acceptable for weight-bearing on the injured arm. Intramedullary devices are best indicated in pathologic fractures or segmental fractures, and cannot be used in distal third fractures. Nails may be inserted either antegrade or retrograde. Antegrade insertion is associated with shoulder pain in up to 40 per cent of cases. Retrograde insertion is associated with a risk of iatrogenic supracondylar humerus fracture, elbow pain, and mild loss of elbow extension. Plates are suitable for most other fractures, and are generally applied via an anterolateral approach for proximal to middle-third fractures, and a posterior approach for distal fractures.

Distal humerus

Distal humerus fractures are classified as extra-articular (type A), partial articular (type B), and complete articular (type C). Two types of injuries are seen: high-energy injuries in young individuals with normal bone, and low-energy injuries in osteopenic elderly patients. Elbow function may be significantly impaired following these injuries due to stiffness, deformity, or arthrosis. Anatomic reduction, rigid fixation, and early motion provide the best results (Fig. 5). Operative complications must be avoided: poor fixation that does not allow early joint motion may yield a worse result than non-operative management.



Fig. 5. (a–c) Example of open reduction and stable internal fixation of a displaced, T-type fracture of the distal humerus. (a) Anteroposterior view showing a displaced, low supracondylar humerus fracture with intra-articular extension. (b) Lateral radiograph of injury. (c) Postoperative radiographs after union showing maintenance of reduction and excellent range of motion.

Neurovascular status and the condition of the surrounding skin and soft tissues should be assessed prior to treatment. Undisplaced fractures require brief immobilization in a removable splint or hinged fracture brace and gentle active exercises. Extra-articular fractures that are well aligned in a cast or brace may be managed non-operatively, but significant loss of motion is possible if immobilization is continued for more than 3 weeks. Displaced articular fractures require open reduction and internal fixation. A triceps-splitting incision may be used for fractures above the olecranon fossa. Distal supracondylar and all intercondylar fractures are best visualized by performing osteotomy of the olecranon process. Anterior transposition of the ulnar nerve is often appropriate. Fixation should be performed with two

plates placed at right angles to each other. One plate is generally placed on the posterior surface of the lateral pillar; the other on the medial aspect of the medial pillar.

Complications include loss of fixation, infection, neurologic injury, and heterotopic ossification. These are best avoided by careful surgery and closely supervised rehabilitation. Postoperative care necessitates a brief period of immobilization for wound healing. Active range of motion exercises are begun within 2 weeks. Slight loss of terminal elbow extension is expected and generally well-tolerated.

Elbow

The elbow is a complex joint that allows both flexion/extension of the arm and rotation of the forearm. Elbow stability is provided by congruence of the articular surfaces, the collateral ligaments, and the tension of muscles crossing the joint. Injury to one or more of these structures can compromise function and comfort of the elbow. Formation of heterotopic bone around an injured elbow may be associated with severe soft-tissue injuries, delayed or traumatic reduction attempts, and traumatic brain injury. Elbow dislocations involve disruption of the capsule and ligaments, and may be associated with fractures. Closed reduction of an elbow dislocation with distal traction and correction of deformity should be performed promptly. Active motion exercises should be started within a few days, unless instability prevents this. In such cases immobilization in flexion for up to 3 weeks may restore adequate stability; active exercises can then be instituted. Medial collateral ligament repair may be necessary in a very unstable joint. Whenever elbow instability is suspected, joint alignment must be monitored with periodic radiographs so that redislocation is not missed.

Isolated radial head dislocation may occur, by itself or together with a fracture of the ulna. This injury complex, known by the eponym ‘Monteggia fracture', may be missed if inadequate radiographs are accepted. Radiographs of the elbow must be obtained whenever there is a forearm fracture. Careful evaluation of the radiocapitellar joint is necessary; on any view, the long axis of the proximal radius should bisect the capitellum.

While closed management of Monteggia fractures is usually adequate for children, adults require anatomic open reduction and internal fixation of the ulnar fracture. Once this is done, the radial head usually is reduced spontaneously. If this does not occur, open reduction is required, often with repair of the orbicular ligament.

Fractures of the proximal radius may involve the head or neck. If undisplaced, only early motion is required. Internal fixation of displaced radial head and neck fractures is recommended. Severely comminuted, displaced fractures may be treated by radial head excision. Preservation of the radial head is paramount in cases where there is associated injury to the medial collateral ligaments or the interosseous ligament of the forearm. In this situation, a radial head prosthesis may improve stability.

Forearm

Diaphyseal fractures of the radius and ulna occur singly or in combination. Isolated, minimally displaced fractures of the ulna are common, and are frequently due to a direct blow to the subcutaneous surface of the ulna (‘nightstick' fracture). Such injuries may be treated with limited immobilization for comfort. An apparently isolated, displaced fracture of one bone must be associated with a dislocation of either the proximal or the distal radioulnar joint. Displaced fractures of the radius with an intact ulna are associated with disruption of the distal radioulnar joint. This pattern of injury is known as a ‘Galeazzi fracture'. Fractures of both radius and ulna are commonly displaced in adults, and require anatomic alignment and stabilization of both bones for restoration of normal function.

A fractured forearm should be examined carefully for open wounds, which may be quite small. Compartment SYndromes occasionally occur, and need urgent fasciotomy. Because individual forearm muscles have significant investing fascia, each muscle must be assessed for adequacy of decompression, and all potentially closed myofascial spaces must be released. Unless an open wound or a compartment syndrome is present, forearm diaphyseal fractures can be grossly aligned and splinted until definitive treatment can be performed.

Non-operative treatment of displaced fractures by closed reduction and casting is an effective treatment in children, but has poor results in adults. Following closed or open forearm fractures, internal fixation with compression plating gives a union rate of 98 to 100 per cent, and an excellent or satisfactory functional result in 92 per cent. Bone grafting is no longer recommended, even for comminuted and open fractures. As elsewhere, aggressive, serial debridement and early soft-tissue coverage are important steps in the management of open forearm fractures. Badly contaminated wounds with extensive soft-tissue defects and open fractures with significant delays before treatment may be better managed initially with external skeletal fixation. Once the soft-tissue envelope has healed, delayed internal fixation may be considered.

Wrist

Fractures of the distal forearm may involve the radiocarpal and distal radioulnar joints, as well as the distal ulna and the triangular fibrocartilage complex. The distal radius is a common site of injury in older, osteoporotic patients following simple falls. Severe distal forearm fractures occur in younger patients as a result of high-energy trauma. Prognosis is determined by the injury pattern, in particular the degree of involvement of the radiocarpal and radioulnar joints. In particular, residual articular surface incongruity of more than 2 or 3 mm is associated with the development of osteoarthritis. Potential complications of distal forearm fractures include median nerve compression, forearm compartment SYndrome, flexor or extensor tendon damage, ligament disruptions, and reflex sympathetic dystrophy.

Distal forearm fractures are classified as extra-articular (type A), partial articular (type B), and complete articular (type C). Several specific eponymic classifications exist as well. Important prognostic factors include the degree of radial shortening, loss of normal volar tilt, dorsal or volar comminution, presence of articular incongruity, and associated ulnar styloid fracture. Plain radiographs and occasionally plain or CT are required to evaluate these factors and decide upon treatment.

Treatment of distal forearm fractures begins with a closed reduction, unless surgery is required because of an open wound or compartment syndrome. Analgesia and muscle relaxation can be obtained with general or regional techniques. If the injury is acute, sterile aspiration of the fracture hematoma and injection of 5 ml of 1 per cent lidocaine is often adequate. Distraction, manipulation of displaced distal radius fragments, and moulding into position are the usual stages of reduction, followed by splinting. Post-reduction radiographs will document restoration of normal anatomic landmarks. Inadequate reduction may be remanipulated. Loss of reduction within the first 2 weeks may be treated by remanipulation, often with adjunctive pin fixation.

Extra-articular fractures may be treated with closed reduction as described. If radial shortening greater than 5 to 10 mm or dorsal comminution greater than 50 per cent are present, the fracture is considered unstable. Adjunctive percutaneous pin fixation or external fixation is recommended for unstable fractures to prevent loss of reduction. Fractures with associated volar or dorsal subluxation, or significant articular displacement and/or incongruity require open reduction and internal fixation with buttress plates. Bone grafting may be required for metaphyseal defects. High-energy fractures of the distal radius may require combined dorsal and volar approaches.

Pelvis

The pelvis may be considered as an osteoligamentous ring that transfers load from the spine to the lower extremities and houses genitourinary and visceral structures as well as neurovascular conduits. Functionally, it provides a level platform for sitting, and includes the acetabular portion of the hip joint. Injuries to the pelvis are conventionally divided into those that disrupt the pelvic ring and those that involve the acetabulum, although they may coexist. Pelvic injuries are frequently occult: all injured patients must be screened with an anteroposterior pelvic radiograph. As discussed earlier, pelvic ring injuries are potentially lethal and may be a marker for other serious injury. Problems associated with pelvic injuries include hypovolemic shock due to bleeding from the pelvic fracture site or veins, arterial injuries (rarely), associated head and thoracoabdominal injuries, urologic injuries, including torn urethra and ruptured bladder, neurologic injuries to lumbosacral nerve roots, plexus, or peripheral nerves, genital injuries, rectal injuries, and open wounds around the fracture or affecting the pelvic contents. Open wounds may be small, and involve the rectum, vagina, or perineum. Open wounds must be identified, as the mortality rate of open pelvic fracture approaches 50 per cent.

The initial evaluation of a pelvic injury involves careful documentation of the neurovascular status of the lower limbs and assessment of the patient's vital signs. The inlet (40-degree caudad) and outlet (30-degree cephalad) radiographic views document anteroposterior and vertical displacement of the pelvic ring, respectively. The Judet views allow proper classification of acetabular fractures. Bladder catheterization is performed after retrograde urethrogram if blood from the urethral meatus or prostate gland displacement on rectal exam is noted. A cystogram is required to exclude the possibility of bladder rupture. Finally, CT is done when necessary to assess further any sacral or acetabular fracture and to evaluate widening or displacement of the sacroiliac joint.

The severity of bleeding, mechanical instability, deformity, pattern of injury, and associated injuries determine the treatment of pelvic ring trauma. By definition, unstable pelvic injuries involve at least two disruptions of the ring. Anteriorly there may be failure of the pubic symphysis or pubic rami. Posteriorly, the injury may involve the sacrum, sacroiliac joint, ilium, or a combination. The sacroiliac joint may be totally disrupted or merely opened anteriorly with the strong posterior ligaments acting as a hinge.

Most hemorrhage from pelvic injuries is from fracture surfaces and venous lesions, and responds to fluid resuscitation. Temporary external fixation is advisable in cases of persistent hemodynamic instability, as discussed previously. If laparotomy is required to address abdominal or pelvic visceral injury, plate fixation of symphyseal disruption may be easily performed. Skeletal traction can be helpful if vertical instability is present. Open pelvic ring injuries require thorough irrigation and

debridement, skeletal stabilization (usually external), and open wound management. Selective fecal diversion is performed for open wounds that involve the perineum, but is not necessary for wounds in the posterolateral buttock. Early definitive fixation is of benefit in the multiply injured patient.

Pelvic ring

Pelvic ring injuries are divided into three types based on the pattern of instability. These are stable patterns (type A), rotationally unstable (type B), or vertically unstable (type C). Anteroposterior radiographs alone, with or without CT scans, are not enough to rule out vertical displacement of the hemipelvis. Adequate inlet and outlet views must be obtained. Rarely, so-called ‘push–pull' stress views are helpful to assess instability. Stress radiographs should not be performed in cases of sacral fracture due to the risk of neurologic injury.

Stable pelvic ring injuries are treated symptomatically with limited weight-bearing for 6 weeks. Rotationally unstable injuries require either anterior external or internal fixation and restricted weight-bearing for 6 to 8 weeks. Complete posterior pelvic ring disruption requires both anterior and posterior pelvic fixation (Fig. 6). Sacroiliac joint dislocations may be treated by either iliosacral screws or anterior plating. Iliosacral screws may be placed percutaneously using fluoroscopic guidance or placed using an open posterior approach. Percutaneous iliosacral screw placement is technically demanding and can result in nerve or vascular injury. Some authors prefer open reduction and screw insertion. Displaced sacral fractures require posterior open reduction and internal fixation with iliosacral screws, plates, or posterior interiliac bolts (‘sacral bars'). Sacroiliac joint fracture–dislocations are treated by plate fixation of the iliac wing. Associated genitourinary injuries may be safely treated at the time of pelvic fixation.



Fig. 6. Example of an unstable pelvic ring injury. (a) Anteroposterior radiograph of the pelvis showing a pubic symphysis diastasis and bilateral iliac wing fractures. (b) Postoperative view showing anatomic reduction and fixation of the symphysis and iliac wing fractures using plates.

Acetabulum

Acetabular fractures are caused by the forceful impaction of the femoral head. The fracture pattern depends on the amount of energy involved, the direction of the applied force, and the position of the leg. Malunion frequently results in disabling arthritis; therefore, anatomic reduction and rigid internal fixation are recommended. When an anatomic reduction is achieved at surgery, the likelihood of later arthritis is low. If articular surface congruity is present, the fracture pattern is relatively stable, and there are no interposed osteocartilaginous fragments in the hip joint, non-operative treatment with restricted weight-bearing is adequate. When appropriate expertise is available, virtually all acetabular fractures should be stabilized to allow the patient to be mobilized and avoid complications relatedto prolonged recumbency. Operative treatment of acetabular fractures requires particular skill. If an experienced surgeon is not available to treat a displaced acetabular fracture, consideration should be given to transferring the patient, or choosing non-operative treatment.

Patients with acetabular fractures must be as carefully evaluated as any other pelvic injury. Documentation of neurovascular status is mandatory. Necessary imaging studies include the 45-degree oblique (Judet) views of the involved innominate bone. These views should be sufficient to classify the fracture according to the schemas of Letournel (Table 6). CT provides additional valuable information, especially regarding degree of comminution, presence of articular impaction, femoral head lesions, or loose bodies within the joint. Criteria for displacement and congruency are important for deciding whether or not open reduction and internal fixation is warranted. Congruent coverage of the femoral head by the weight-bearing dome as determined by roof arc measurements on all three standard radiographic views is the minimum criterion for acceptable alignment. Fractures of the posterior wall require special assessment because roof arc measurements do not reliably indicate adequate femoral head coverage in this region. These fractures are better evaluated by CT and the obturator oblique radiograph. A displaced posterior wall fragment that is large or permits femoral head subluxation requires open reduction and internal fixation. Stress radiographs taken with the patient under anesthetic may be needed to assess instability.

Elementary patterns	*Associated patterns*
Posterior wall	T-shaped fractures
Posterior column	Posterior column and posterior wall
Anterior wall	Transverse and posterior wall
Anterior column	Anterior column and posterior hemitransverse
Transverse	Both column fractures

Table 6 Letournel classification of acetabular fractures

Initial treatment consists of reduction of persistent dislocation of the femoral head by closed reduction. Skeletal traction is advisable if the hip is subluxated; 10 kg or more force may be required. Open reduction and internal fixation, when necessary, is usually delayed until the patient is hemodynamically stable and preparations for surgery are complete. The surgical approach chosen depends upon proper patient assessment and fracture classification. The posterior Kocher–Langenbeck is chosen for the management of posterior wall fractures, posterior column fractures, low-transverse fractures, and T-type fractures. The ilioinguinal approach is best for both column fractures, anterior column or wall fractures, and anterior column–posterior hemitransverse patterns. In addition, the ilioinguinal approach should be used whenever possible in patients with traumatic brain injury in order to minimize the risk of heterotopic ossification. Ideally, the acetabulum is repaired through a single incision, although combined approaches may also be used. When necessary, greater exposure can be obtained by trochanteric osteotomy, or by extensile triradiate or iliofemoral incisions. Such approaches may be required for delayed treatment. Prophylaxis against heterotopic ossification is necessary following posterior or extended iliofemoral approaches with either Indocin 25 mg orally three times a day (for 6 to 12 weeks), or low-dose irradiation (700 to 1000 R, single dose).

Lower extremity injuries

Several useful guidelines regarding the assessment of leg injuries must be remembered. Apparent neurologic deficits may be the result of spinal cord, nerve root, or lumbosacral plexus injuries rather than lower extremity peripheral nerve injury. Associated pelvic injuries are common: examination of the pelvis and a routine anteroposterior radiograph of the pelvis are essential. Multilevel injuries must always be considered. Examples include knee injuries associated with acetabular fractures, ipsilateral femoral neck fractures (often occult) and femoral shaft fractures, and knee ligament injuries combined with femoral or tibial fractures. Foot fractures may be associated with more proximal injuries.

Essential steps in the care of leg injuries is the documentation of distal pulses, evaluation of all open wounds, assessment of motor and sensory function, and observation for tense swelling or severe progressive pain, which may indicate ischemia. An unrecognized intimal flap tear of the popliteal artery carries a significant risk of delayed thrombosis and may be associated with dislocations and fracture–dislocations of the knee. Fractures, dislocations, and ligament injuries must be identified.

Once the entire limb has been evaluated, priorities for treatment can be established.

Hip dislocation

Hip dislocation requires appreciable force, and usually is the sequelae of high-energy trauma. Associated musculoskeletal trauma may be seen in 30 to 40 per cent of cases. Pure dislocations occur either anteriorly or posteriorly. Anterior dislocations represent 15 per cent of all hip dislocations, and are often associated with femoral head impaction fractures. Most dislocations (85 per cent) are posterior and are caused by force applied to the anterior knee while the hip is flexed and somewhat adducted. If the hip is more abducted, the posterior lip of the acetabulum or femoral neck is fractured. In these cases, hip stability needs to be assessed and the size of the posterior wall fragment determined.

The principles of treatment are well outlined and involve immediate reduction of the hip under appropriate anesthesia. Any delay in treatment of a hip dislocation increases the risk of avascular necrosis of the femoral head. This leads to secondary segmental collapse during the revascularization phase, and often results in disabling post-traumatic arthritis. Anterior dislocations are reduced by traction, extension, and internal rotation of the femur. Posterior dislocations are reduced by traction in adduction and flexion, followed by gentle abduction and extension of the hip. Post-reduction imaging is required to assess the reduction. Operative intervention is necessary if congruent closed reduction cannot be obtained, if an associated posterior wall fracture greater than 25 per cent of the wall is identified, or if osteochondral debris is present within the joint space. If present, these fragments require removal. The size and location of any fractures of the acetabular rim or femoral head needs to be determined. For this reason, post reduction CT scanning is recommended.

Following uncomplicated reduction, the patient is treated with protected weight-bearing for 4 to 6 weeks. Although the functional outcome after reduction of simple posterior hip dislocations is variable, excellent and good results are obtained in 70 to 80 per cent of patients. The majority of patients have minimal stiffness, slight pain, and minimal arthritic changes. Avascular necrosis may result in 15 per cent, and is usually evident within the first year (but may appear up to 3 years after injury). Degenerative arthritis may be the result in up to 75 per cent of patients at long-term follow-up. Recurrent dislocations are rare in adults.

Femoral head fracture

Combined fractures of the femoral head or neck occur in 7 per cent of posterior hip dislocations. These injuries have been further categorized into four subtypes by Pipkin: type I, fracture of the femoral head caudad to the fovea; type II, fracture of the femoral head cephalad to the fovea, type III, femoral head plus femoral neck fracture; and type IV, femoral head plus fracture of the acetabulum. CT scanning is recommended for identification and sizing of the femoral head fragment.

Femoral head fractures associated with persistent hip dislocation are best treated by closed reduction for Pipkin types I, II, and IV. Type III injuries, involving an associated fracture of the femoral neck, require open reduction. Unsuccessful closed reductions of the associated hip dislocation should be treated by open reduction. Femoral head fractures usually require open reduction from an anterior approach and internal fixation with countersunk cancellous or Herbert screws. In the very elderly patient with low functional demands, prosthetic replacement is indicated.

Proximal femur

Proximal femur fractures are typically divided into extra-capsular fractures involving the trochanteric region or intracapsular femoral neck fractures. Because of their differing prognosis and therapeutic needs, each of these injuries should be considered separately.

Femoral neck

Femoral neck fractures are common in the elderly, and are occasionally seen in young individuals after high-energy trauma. The treatment of femoral neck fracture depends primarily on the age of the patient and the degree of displacement. Elderly patients with unstable medical comorbidities should have their condition optimized. The timing for internal fixation of femoral neck fractures remains controversial. Rates of avascular necrosis and non-union are decreased by fixation within 12 h after injury. However, some studies have not shown an increase in the rate of avascular necrosis or non-union with delayed fixation up to a week. Because of the biologic advantages, we advocate stabilization of the fracture as soon as the patient has been evaluated medically and is stable.

For undisplaced fractures, *in-situ* fixation is advisable as non-operative treatment carries a 10 to 14 per cent risk of displacement. Displaced femoral neck fractures are treated according to the age and functional demands of the patient. These fractures are at high risk for avascular necrosis and non-union, averaging 30 per cent and 15 per cent, respectively. In healthy patients under 75 the treatment of choice is reduction and internal fixation. Gaining an anatomic reduction is important for a good result. Femoral neck fractures reduced anatomically are best fixed with three pins or screws. A closed reduction may suffice, but the capsule should be opened to decompress the hemarthrosis that may impair head perfusion. If a satisfactory closed reduction is not achieved, an open reduction is indicated. Patients physiologically over 75 should be considered for prosthetic replacement to avoid a secondary operation. A prosthetic replacement may also be desired in a patient with hip arthritis in conjunction with a femoral neck fracture.

Pertrochanteric

Intertrochanteric fractures heal readily, but often with varus angulation. The extended bed-rest required until union occurs is not well tolerated, so these are usually internally fixed with a sliding hip-screw or sliding nail device. Anatomic reduction is the most important factor in avoiding complications. Placement of the hip screw into the center of the femoral head within 5 mm of subchondral bone, on both anteroposterior and lateral radiographs, is also important to avoid fixation failure. Elderly, osteoporotic people with subtrochanteric involvement pose potential problems for fixation, especially when the fracture line obliquity is reversed, with its lateral end more distal. In these instances, fixed low-angle devices or intramedullary nails are good choices.

Subtrochanteric

Subtrochanteric fractures occur in the most highly stressed region of the femur and may be difficult to manage. Non-operative treatment has a high rate of malunion, non-union, and prolonged disability. The choice of operative treatment depends on the type of fracture. So-called high subtrochanteric fractures involve the lesser trochanter. If the piriformis fossa is intact, high fractures may be treated with a second-generation intramedullary nail. If the intertrochanteric region is comminuted, then a sliding hip screw, 95-degree condylar blade plate or 95-degree condylar screw may be used. Low fractures that do not involve the lesser trochanter are best treated with standard nailing techniques. Bone grafting is not necessary when plating is performed if indirect reduction techniques are used and the medial fragments not exposed.

Femoral shaft

Femur fractures usually follow high-energy injuries, and may be associated with significant and possibly occult injuries to other organ systems. In one series, over half of the patients with thoracolumbar spine injuries and femoral shaft fractures did not have their spine injury diagnosed until after fixation of the femur fracture. The femur fracture itself must be promptly immobilized, optimally with a traction splint, as failure to do so increases local bleeding and soft-tissue damage. Prompt surgical stabilization of femoral shaft fractures offers many SYstemic benefits for the patient. If definitive fixation must be delayed, skeletal traction or external fixation should be applied as soon as possible.

Femoral shaft fractures unite readily due to the mass of surrounding muscles. However, because of the forces that these muscles exert, problems with malunion and knee dysfunction result from non-operative treatment. The risk of adult respiratory distress syndrome is similarly increased by conservative treatment.

Closed intramedullary nailing is the treatment of choice for closed and open femoral shaft fracture. It offers the lowest incidence of complications and the fastest rehabilitation. Although femoral nailing is usually performed on a fracture table, the procedure may be performed on a standard table safely using either manual traction or the femoral distractor. These techniques are especially valuable in the patient with an unstable spine or pelvis injury, for whom positioning on the fracture table would be dangerous. Interlocking screws should be used routinely. In a severely injured patient, especially one with associated vascular injury, plate fixation may be considered. When plating is performed, indirect reduction techniques should be followed to minimize the risk of non-union and the need for bone grafting. External fixation is usually used only as temporary stabilization.

Although it is considered that the results of femoral nailing are superior, the late outcome following femoral nailing has received little attention. In one series, a group of patients was evaluated at a minimum of 12 months after femoral nailing with an abbreviated functional assessment. Thirty-nine per cent had some limitation in ability to ambulate or stand, and 37 per cent had some pain. Nine per cent had to modify or change their jobs.

Distal femur

Fractures of the distal femur are divided into those that involve the supracondylar region alone and those that have intra-articular extension. Also of importance are the presence of coronal fractures of the anterior or posterior condyles, and the severity of comminution. Radiographic assessment permits classification and planning of management. Open wounds, knee ligament injury, proximal tibial fractures, and nerve or arterial injuries may complicate the management of these injuries.

Non-displaced, stable fractures can be treated in a hinged cast-brace. Displaced extra-articular fractures may be treated with a blade plate, dynamic condylar screw, or intramedullary nail (antegrade or retrograde). Displaced articular fractures require anatomic reduction, rigid fixation, and early motion of the knee in order to restore alignment, mobility, muscle strength and endurance, and avoid post-traumatic degenerative arthritis. This may be done by reassembling the displaced condyles with interfragmentary screws, and then using either a dynamic condylar screw or buttress plate for fixation.

Knee

Soft-tissue injuries of the knee may occur in association with other fractures of the femur or tibia, or occur as isolated injuries. Careful examination can elucidate most of these injuries. Magnetic resonance imaging is useful for confirmation when the diagnosis is uncertain.

Knee dislocations are associated with rupture of several ligaments and may have associated fractures. They are important to diagnose early because of the high incidence of associated, often occult popliteal artery injury. Frequently these are intimal tears that have the potential for delayed thrombosis, occlusion, and irreversible distal ischemia requiring amputation. A high-quality arteriogram is essential to evaluate such injuries. If ischemia is detected, urgent exploration and repair of the popliteal artery are required. If limb perfusion is not compromised, intimal flaps may be followed clinically. Peroneal and/or tibial nerve injuries are frequent as well.

Isolated injuries of the medial collateral ligament of the knee usually recover with non-operative treatment using a hinged brace and vigorous rehabilitation. Injuries of the fibular (lateral) collateral ligament may result in persistent laxity and functional disability, especially when combined with injuries of the posterolateral corner ligament complex. These injuries should be considered for early repair. Delayed repair is more difficult and generally requires reconstruction with allograft tissues.

Torn cruciate ligaments do not heal with closed treatment or primary suture repair, except in the rare instance when they are avulsed from the tibia or femur. Isolated anterior or posterior cruciate ligament injuries are usually not repaired acutely. If persistent anterior cruciate ligament laxity results in functional instability, and if the patient will co-operate with a rehabilitation program, delayed anterior cruciate ligament reconstruction may be performed. Techniques for posterior cruciate ligament repair are improving, but the indications for posterior cruciate repair remain poorly defined.The timing of surgery is critical to avoid knee stiffness. Full knee motion without significant swelling is required to assure optimum results. Either immediate reconstruction is performed, or surgery is delayed until soft tissues stabilize, and range of motion returns. Choice of graft for reconstruction depends on the extent of associated injuries, with allograft tissues sometimes used to eliminate donor site morbidity.

Meniscus tears that interfere with normal motion or cause pain are best treated arthroscopically. Tears of the meniscus are usually treated by partial meniscectomy; preservation of the meniscus is important to the longevity and function of the joint. Peripheral meniscus tears have potential for healing and should be re-attached, and protected until the suture line has healed. A small undisplaced tear may heal with protection alone.

Dislocations of the patella (usually laterally) may occur following trauma, or may be associated with developmental malalignment of the knee extensor mechanism. Osteochondral or marginal avulsion fractures may be present, and should be looked for on radiographs or magnetic resonance imaging. An initial acute patellar dislocation usually tears the medial retinaculum: this should be protected by maintaining knee extension during the 6 or so weeks required for healing.

Patella fracture

Patellar fractures occur from direct blows to the anterior knee, or indirectly as a result of sudden, severe loading to the knee extensor mechanism. The ability to raise the extended leg is an important test for quadriceps continuity, and may also reveal rupture of the quadriceps tendon or patellar ligament when the patella is intact. A comminuted fracture that is undisplaced and has intact quadriceps function can be treated non-operatively, with early range of motion exercises, splinting during ambulation to prevent loading the flexed knee, and weight-bearing as tolerated with the knee fully extended. The occasional vertical fracture can be similarly managed, unless it is displaced enough to affect articular surface congruity. Transverse, displaced patellar fractures indicate quadriceps mechanism disruption and surgical repair is required to restore adequate knee extension strength. Displaced fractures should be fixed anatomically, preserving some or the entire patella if at all possible. Biomechanical studies demonstrate that screw fixation, supplemented by an anterior tension band wire, provides the greatest strength of repair. If a small proximal or distal fragment is excised, the quadriceps tendon or patellar ligament is reattached as indicated. Repairs of the patellar ligament are protected with a cerclage suture or wire. Transverse tears in the medial and/or lateral quadriceps expansions must be repaired concomitantly. A severely comminuted patellar fracture is often best excised to avoid disabling pain or post-traumatic arthritis, even though absence of the patella weakens knee extension.

Tibia plateau

Tibial plateau fractures are classified by Schatzker into six types. Injuries are described by the involved plateau, presence of a split fracture, and whether any joint depression is present. The recognized types are: type I, undisplaced lateral plateau fracture; type II, central depression fracture; type III, split-depression fracture; type IV, medial plateau fracture; type V, bicondylar fracture; and type VI, bicondylar fracture with shaft dissociation. Uncorrected articular surface displacement or depression increases the risk of post-traumatic osteoarthritis, as well as chronic instability.

Undisplaced fractures usually need only to be protected during healing. Following closed management of fractures with less than 10 degrees of instability in full extension or less than 10 mm of joint depression, 90 per cent of patients have good long-term results. Displaced fractures are best managed by reduction and fixation according to fracture type. Split fractures of the lateral or medial plateau are managed by reduction and fixation with 7-mm screws. Buttress plating is added when the bone quality is poor. Depressed segments of the plateau causing instability should be elevated and internally fixed; the resulting defect is filled with autogenous bone graft or graft substitute. Fixation should be secure enough to permit early motion of the joint. An injured meniscus should be repaired and saved if at all possible. Torn collateral ligaments may be repaired or protected in a hinged cast-brace, but cruciate ligaments are not repaired primarily, unless an avulsion permits reattachment to bone and early motion is permissible. While early motion helps by preventing stiffness and promoting cartilage recovery, weight bearing must be delayed for several months, until fracture healing is mature.

Displaced bicondylar tibial plateau fractures are associated with severe soft-tissue damage, so that open reduction and internal fixation risks slough of the soft-tissue flaps. Careful judgement regarding operability and timing of surgery is essential. Minimally traumatic techniques are necessary. Limited open reduction and internal fixation of the articular surface components, with external fixation using a wire-pin hybrid device, has become popular and seems to avoid the soft-tissue complications related to plate fixation in these injuries.

Tibial shaft

Tibial diaphyseal fractures are common and are often associated with fibular shaft fracture. Fractures may extend into the knee or ankle joints. Severe injuries have a greater risk of complications such as neurovascular compromise, delayed union, non-union, malunion, and infection.

Treatment should be chosen with regard to the severity of soft-tissue and bone injury. Patients with low-energy injuries usually do well with functional treatment in a cast or brace. Displaced tibial fractures benefit from a more aggressive approach. Operative care is mandatory if the fracture is open, if compartment syndrome exists, or if limb-threatening ischemia present. Open wounds require prompt and aggressive surgical debridement with copious irrigation, bone stabilization, and delayed closure. Compartment syndromes require fasciotomy. Vascular injuries causing ischemia require repair. Early amputation should be considered when functional limb salvage is unlikely because of severe soft-tissue injury, tibial nerve injury, or irreparable vascular injury.

The severity of a tibial fracture is related to the energy absorbed by the soft tissues and bone of the leg. Factors that need to be assessed include the extent of devitalized skin and muscle, periosteal stripping from bone, neurovascular injury, and wound contamination. Swelling sufficient to produce a compartment SYndrome usually indicates a severe injury. Associated fibular fractures may indicate more severe injuries, but do not necessarily affect the prognosis. In fact, an intact fibula may make non-operative management more difficult, with tendencies for varus malalignment and delayed union. Obvious diastasis between the two shafts requires longitudinal disruption of the interosseous membrane, and indicates a severe soft-tissue injury.

Initially, overall alignment should be restored with a well-padded long leg splint, capable of being loosened to accommodate swelling. Closed reduction, which may prove definitive for less severe injuries, is best achieved with the aid of gravity. Gentle cast moulding usually corrects any angulation. The initial below-knee plaster is

used to hold the limb horizontally, maintaining correct rotation, while the plaster is extended about two-thirds the way up the thigh, with the knee flexed 10 to 15 degrees.

Closed treatment for tibial shaft fractures is usually undertaken by external support in a weight-bearing cast. Although well suited to the common fractures caused by low-energy trauma, this approach may fail to maintain alignment and to achieve timely union for moderate and severe tibial fractures. Initial immobilization is provided by a long leg cast as described. Early weight bearing is generally advocated. Once swelling has resolved, a patellar-tendon-bearing cast or brace may replace the long leg cast, for more stable fractures, to permit knee, and possibly ankle, motion. The majority of tibial fractures heal with satisfactory alignment and function within 4 or 5 months. Limited ankle motion has been identified as the greatest source of disability following closed treatment of tibia fractures.

Intramedullary nailing may be the treatment of choice for managing displaced tibial diaphyseal fractures (Fig. 7). Early function is often possible, long before full radiographic healing is complete. Bone et al. (1989) matched 25 patients with displaced tibial fractures treated with a cast to 25 patients with similar fractures treated by nailing. When evaluated a mean of 4.4 years later with the Iowa Knee Evaluation, the Ankle-Evaluation Rating SYstem, and the SF-36, the patients who had nailing had superior results.

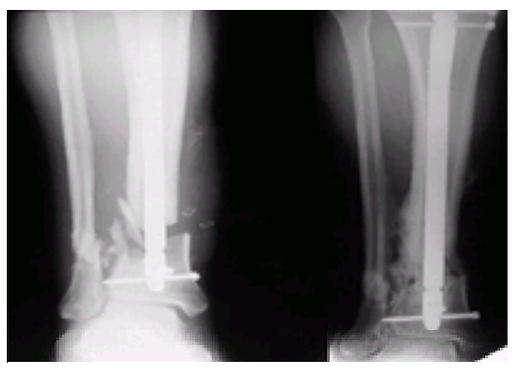


Fig. 7. Example of a displaced, grade IIIB open fracture of the tibia, managed by intramedullary nailing, free tissue transfer, and staged bone grafting. On the left, an anterioposterior radiograph of the injured tibia immediately after intramedullary nailing. With the nail in place, management of the associated soft-tissue injury was facilitated. Note the segmental bone loss present. On the right an anteroposterior radiograph taken 1 year after injury, following staged posterolateral and anteromedial autogenous bone grafts. The fracture has united and the patient is able to walk normally.

The technique of tibial nailing has evolved considerably. Interlocking nails are now routinely used. As in the femur, most fractures should be treated initially with static interlocking, as loss of reduction is common when dynamic locking is performed. Reaming of the medullary canal was initially done to obtain a longer zone of purchase by the nail in the medullary canal as well as to permit the use of larger diameter nails. Later, reaming became less favored because of concerns that reaming devascularizes the tibial cortex, increases the possibility of compartment syndrome, and increases the risk of infection in open fractures. A decade ago, a new generation of small diameter ‘unreamed' nails with interlocking capability was introduced in the hope of solving these problems. The use of small diameter nails is associated with nail or screw failure in up to 40 per cent of cases, and that delayed union is more common. The results of recent prospective comparative studies have swung the pendulum back in favor of reamed nailing, which is safe and more effective in both open and closed fractures.

Largely because of the success of primary intramedullary nailing, external fixation of tibial shaft fractures is now less commonly performed. Although external fixation has classically been recommended as definitive treatment for open fractures, it is now used primarily as temporary fixation for severe injuries. When possible, open tibia fractures initially treated by external fixation are converted to nails within the first 2 weeks in order to avoid possible infection from contaminated pin tracts.

Again, with the success of nailing techniques, the fixation of tibial fractures with plates is generally confined to the metaphyseal regions. Use of plates in the treatment of open fractures carries an increased risk of infection. When necessary, and as stated before, indirect reduction techniques should be employed whenever plating of tibial fractures is performed.

Distal tibia

The distal tibial metaphysis is also referred to as the tibial ‘pilon'. Fractures involving this region often extend into the roof of the ankle, or ‘plafond'. These injuries are often associated with severe soft-tissue injuries. Anatomic restoration of axial alignment, articular surface, and mortise improves outcome, but only if it can be achieved without complications, primarily soft-tissue sloughs.

Initial treatment of a tibial pilon fracture involves aggressive debridement of associated open wounds. Generally, restoration of length and realignment of the extremity should be done initially by external fixation, commonly supplemented by plate fixation of the distal fibula. Definitive treatment involves use of a so-called hybrid external fixator, or delayed open reduction and internal fixation using meticulous indirect reduction techniques and low profile implants. The combined use of limited internal fixation of the articular surface with screws, bone grafting of metaphyseal defects, and external fixation of the metaphysis is a widely accepted management protocol.

Ankle

The ankle region is frequently injured and trauma patients must be evaluated for history of injury, pain, swelling, impaired function, or deformity. Localized tenderness, instability, and radiography are the keys to the proper definition of pathology. Integrity of skin, tendons, ligaments, and vascularity, as well as function of motor and sensory nerves must be assessed. Combinations of injuries are typical: the foot and leg may be involved in injuries affecting the ankle.

Most fractures are apparent on routine radiographs, which should include anteroposterior, mortise, and lateral views. Some injuries, such as osteochondral lesions or lateral process fractures of the talus, or fractures of the foot (anterior process of calcaneus, base of fifth metatarsal, tarsometatarsal joint injuries) may be overlooked without appropriate radiographic projections. Injured ligaments are usually tender, painful, and/or unstable when stressed. Ruptured tendons are often tender, and a careful examination should demonstrate loss of specific function. It should be noted that active plantar flexion is typically present despite rupture of the tendocalcaneus, which is best diagnosed by squeezing the calf of the prone patient and noting absence of ankle movement (Thompson test).

The anatomy of the ankle joint determines common injury patterns. The ankle joint is a congruent hinge that permits dorsi and plantar flexion of the talus around an axis that connects the tips of medial and lateral malleoli. Inversion and eversion occur at the flexed knee and subtalar joint. The talus articulates with an osteoligamentous socket (‘mortise') formed by the tibial plafond and medial malleolus of the tibia, the tibiofibular syndesmotic ligaments, and the lateral malleolus. Restoring and maintaining the function of the ankle mortise is the goal of treatment for malleolar fractures. Ankle injuries that disrupt the normally congruent fit between talus and mortise pose a significant risk of post-traumatic osteoarthritis with pain, limited motion, and impaired gait. Most frequently, incongruence is due to a malleolar fracture caused by indirect forces that produce rotation of the talus relative to the tibia. Inversion (adduction) or abduction (eversion) in the frontal plane, and external rotation of the talus in the transverse plane are the movements that produce malleolar fractures. These usually occur when the foot is on the ground and the body's momentum forces the tibia to move in an abnormal direction.

Ankle fractures and sprains may be assessed by the mechanistic schemas of Lauge-Hansen. Undisplaced fractures of the lateral malleolus are stable when the medial side of the ankle is not injured. There is no disturbance of the mortise relationships of the talus to the distal tibia, nor of the lateral malleolus to either distal tibia or to talus. Their prognosis is good with non-operative treatment.

Displaced lateral malleolar fractures should raise suspicion of an occult medial ligament disruption with ankle instability and loss of support for the talus. Bimalleolar injuries are always unstable. The most severe malleolar fractures are fracture dislocations, in which the talus is completely displaced from the tibial articular surface; this is most evident on the lateral radiograph. Prompt reduction of talus to tibia is necessary to minimize soft-tissue compromise. A major posterior tibial lip fracture (greater than 25 to 30 per cent) may render this reduction unstable, with recurrent posterior talar subluxation on lateral radiographs taken after reduction, in spite of apparently satisfactory plaster immobilization. Lateral malleolar length, rotation, and close, stable approximation to the distal tibia are all important. A mortise radiograph demonstrating a symmetric clear space about the talus indicates a satisfactorily congruent mortise. Displaced malleolar fractures are best treated by anatomic open reduction and stable internal fixation. An additional benefit of internal fixation is that immobilization in a non-functional position is avoided, and ankle motion and function can be rapidly initiated. Open reduction and internal fixation should be performed within a few hours after injury, before soft-tissue swelling becomes severe. If this is not possible it is better to wait until soft-tissue swelling has resolved, and skin wrinkles are again visible, even though this may require several days. The best possible closed reduction and well-padded cast or splint immobilization are essential during this waiting period. Elevation and observation for

neurovascular compromise are also required.

Open fractures require prompt irrigation and debridement, antitetanus therapy and antibiotics, fracture reduction, and stabilization, with delayed wound closure. Primary internal fixation is effective and safe for most of these injuries. In severe cases, internal fixation of accessible fractures, combined with external fixation of the foot and ankle, may be safer. The external fixator may be retained through fracture healing, or exchanged for staged internal fixation once the soft tissues permit.

Medial malleolar fractures are approached through a straight or curved incision, through which the medial ankle joint is visualized and irrigated. Fractures are fixed with lag screws perpendicular to the fracture line or with K-wires and ‘tension-band' wires if small and/or comminuted. Avulsion fractures are generally transverse. Screws are placed obliquely. Injuries with vertical fracture lines require horizontal screws, and may need reduction of impacted plafond components. If the medial component of an unstable malleolar injury is ligamentous, surgical repair is not required, as anatomic fibular repair is both necessary and sufficient. Occasionally medial exposure may be required, however, as medial tendon interposition prevents reduction of the talus in the mortise.

The distal fibula is generally repaired with either lateral or posterior plates. The posterior antiglide plate has theoretical biomechanical advantages. Lateral fibular plating is done most commonly, but often results in prominent hardware. Small posterior malleolar fractures do not need additional treatment if they minimally involve the articular surface. If they are larger, anatomic reduction and stable lag screw fixation provide stability and a congruent articular surface of the plafond. Either the lateral or medial incision can be used to visualize reduction. Lag screws may be inserted from back to front, or from front to back through stab wounds if required, but should be preceded by provisional K-wire fixation and an intra-operative radiograph to confirm a satisfactory reduction.

After reduction and fixation of malleolar fractures, the ankle is initially splinted in neutral dorsiflexion to permit soft-tissue healing. It may then be treated with early motion, but resumption of unprotected weight bearing is delayed until after fracture healing. With such treatment, satisfactory long-term results should be anticipated in 80 per cent or more of injuries. Poorer outcomes are typical of more severe fractures, especially dislocations and those affecting more of the articular surface.

Inversion sprains of the lateral ankle ligament complex are common injuries. The lateral collateral ligament of the ankle has three parts: the anterior and posterior talofibular ligaments, and the calcaneofibular ligament. Sprains usually involve the anterior talofibular and calcaneofibular components, which may be partially or completely torn. Diagnostic features include a history of acute inversion stress with lateral ankle pain, swelling, ecchymosis, and tenderness, usually best localized over the involved ligaments. Passive inversion usually increases the pain. Radiographs are essential to exclude other injuries, including osteochondral fractures of the talus and lateral process fractures of the talus. Although inversion sprains interfere with walking and sports acutely, they rarely result in chronic instability. Protection from weight bearing, elevation, ice, and a supportive posterior or U-shaped well-padded splint provide appropriate initial treatment for all but the most trivial sprains. Functional rehabilitation includes peroneal muscle strengthening, balance exercises, and progressive resumption of activity, perhaps with a less restrictive stirrup splint. Rarely, chronic ankle inversion instability may require surgical ligament reconstruction.

Syndesmotic injuries may occur alone or in association with ankle fractures. When associated with diastasis of the distal tibiofibular joint, these injuries are readily apparent. Occult injuries are more difficult to diagnose, and may be responsible for persistent pain following a seemingly simple ankle sprain. Complete rupture of the syndesmosis sometimes occurs with fracture of the proximal fibula. This injury, the so-called ‘Maisonneuve' fracture, is treated by closed or open reduction of the syndesmosis and placement of a 4.5-mm syndesmosis screw. Controversy persists regarding the need for routine removal of SYndesmosis screws.

Achilles tendon ruptures usually occur in middle-aged individuals during a push-off or jump, as in basketball or racket sports. Pain, inability to continue the precipitating activity and some degree of gait impairment are noted. Tenderness, a palpable defect in the tendon that becomes obscured by hematoma and swelling, and weak but persistent plantar flexion are noted. Compressing the calf while a patient kneels with his toes hanging freely normally produces passive plantar flexion (Thompson test). This response is absent or reduced when the tendocalcaneus is ruptured. Treatment can be either by immobilization in gentle equinus for 8 weeks, or by surgical repair with any of several techniques. Repair may have more complications, but produces better function with a lower rate of recurrence.

Foot

Foot injuries are often missed, ignored, or deferred ‘until later', especially when other more impressive fractures are present. This results in a chronically painful, deformed foot that is a significant cause of chronic disability, due to pain on weight bearing and impaired gait. Foot trauma is a significant source of persistent disability in multiple trauma patients. Precise diagnosis of injuries and prompt, secure anatomic repairs offer the best chance of successful rehabilitation for patients with serious foot fractures and dislocations.

Talus fractures occur in several patterns. Osteochondral fractures of the articular surface may occur, and should be excised if the lesion is small. If they are large, reduction and internal fixation may be successful. A displaced fracture of the lateral process of the talus involves the subtalar joint, and generally should be repaired or excised. Talar neck fractures often extend into the talar body. Displaced talar neck fractures are associated with subtalar subluxation or dislocation. Tenuous skin cover and deformity may result in wound problems. Closed reduction is urgent, but often requires an anesthetic; if unsuccessful, an open reduction is required. Anatomic reduction and rigid internal fixation with lag screws decreases the risk of avascular necrosis of the talar body, a more frequent complication of closed treatment. The incidence of this problem increases with the degree of displacement of the talar body from its sources of blood supply in the talar neck, the sinus-tarsi vessels, and the medial collateral ligament of the ankle. For optimal results, displaced talar fractures should be treated with anatomic alignment and delay of weight bearing until union has occurred.

Calcaneus fractures result usually from a fall from a height, with the patient landing on his heel. Associated axial loading injuries often include compression fractures of the thoracolumbar spine. Typically, the talus is forced downward into the os calcis, driving the subtalar joint surface plantarward into the substance of the calcaneus, decreasing the height and increasing the width of the bone. Subtalar comminution and deformity are present to a greater or lesser degree. Soft-tissue damage may be severe enough to produce a skin slough, from swelling alone or in combination with skin tension over extreme bony deformity. Plantar nerve damage and even compartment syndromes of the foot may develop. Lateral, oblique, and axial radiographs are necessary for assessment of the injury. CT scans are helpful for planning surgical repair, which is becoming more popular for these disabling injuries. However, comparative studies have not shown clear functional benefit beyond that obtained with non-operative care emphasizing rest, elevation, active motion, and acceptance of deformity. Pain on weight bearing and loss of subtalar motion, especially noticeable when walking on uneven surfaces, are the typical sequelae of a calcaneus fracture. Closed reduction is rarely successful, and deformity persists without surgery. Symptoms tend to decrease gradually over many years following non-operative treatment.

Tarsometatarsal fractures and dislocation (so-called Lisfranc joint injuries) frequently cause pain and disability. The best results are provided by precise, anatomic reduction with screw fixation. Instability and tenderness of the forefoot or the midfoot must be sought whenever a foot injury is suspected. Slight loss of alignment of the medial three metatarsals with their cuneiforms, especially the recessed second metatarsocuneiform joint, is important, as is the alignment of the more mobile lateral two metatarsals with the cuboid. The overall configuration of the forefoot should be compared with normal, and especially with the patient's uninjured opposite foot. Arthrodesis of chronically painful tarsometatarsal joints may improve SYmptoms and function. This is far easier if the foot is normally aligned.

Displaced metatarsal fractures may need open reduction and internal fixation with K-wires or small plates and screws. Restoration of the normal distal plantar surface is necessary to provide normal weight distribution on the sole of the foot, and to avoid metatarsalgia resulting from a prominent metatarsal head or bony deformity.

Toe dislocations should be reduced promptly with closed techniques, or if these are unsuccessful, open techniques. Displaced, intra-articular great toe fractures may benefit from open reduction and internal fixation to preserve overall alignment and dorsiflexion range of the first metatarsophalangeal joint. Lesser toes are less exposed to weight bearing. As long as fractures do not produce a deformity that will interfere with shoe wear or cause pressure on an adjacent digit, gross realignment is usually all that is required. Loose taping to adjacent toes is usually sufficient. Percutaneous K-wire fixation may be helpful, especially to stabilize an open injury.

Further reading

Bandyk DF. Vascular injury associated with extremity trauma. *Clinical Orthopaedics and Related Research* 1995; **318**: 117–24.[A comprehensive review of the diagnosis and management of vascular injuries associated with orthopaedic trauma.]

Ben-Menachem Y, Coldwell DM, Young JWR, Burgess AR. Hemorrhage associated with pelvic fracture: causes, diagnosis, and emergent management. *American Journal of Radiology* 1991; **157**: 1005–14. [A review of the roles of angiography and stabilization are provided in the context of a logical treatment algorithm for the management of bleeding associated with pelvic fracture.]

Bone LB, Johnson KD, Weigelt J, Scheinberg R. Early versus delayed stabilization of femoral fractures. A prospective randomized study. *Journal of Bone and Joint Surgery* 1989; **71A**: 336–40. [A landmark study that proves the benefit of early fracture fixation in multiply injured patients.]

Bosse MJ *et al.* Adult respiratory distress syndrome, pneumonia, and mortality following thoracic injury and a femoral fracture treated either with intramedullary nailing with reaming or with a plate. A comparative study. *Journal of Bone and Joint Surgery* 1997; **79A**: 799–809. [An important study providing evidence that the severity of pulmonary injury, and not the type of femoral fracture fixation, determines pulmonary morbidity after fracture of the femur.]

Brumback RJ, Jones AL. Interobserver agreement in the classification of open fractures of the tibia. The results of a survey of two hundred and forty-five orthopaedic surgeons. *Journal of Bone and*

Joint Surgery 1994; **76A**:1162–6. [A review of the current use and limitations of the Gustilo classification of open fractures.]

Burgess AR *et al.* Pelvic ring disruptions: effective classification SYstem and treatment protocols. *Journal of Trauma* 1990; **30**: 848–56. [The results of a classification and treatment protocol based on injury mechanism is presented for a series of 210 patients.]

Chou L, Reynolds MR, Esterhai JL Jr. Hazards to the orthopaedic trauma surgeon: occupational exposure to HIV and viral hepatitis. (Review article.) *Journal of Orthopaedic Trauma* 1996; **10**: 289–96. [Reviews the guidelines for risk prevention of HIV and hepatitis B and C, recommended by the Centers for Disease Control and Prevention (United States).]

Gerber C, Mast JW, Ganz R. Biological internal fixation of fractures. *Archives of Orthopaedic and Trauma Surgery* 1990; **109**: 295–303. [A thorough summary, with examples, of the principles of indirect fracture reduction and biologic osteosynthesis.]

Herscovici D, Sanders R, DiPasquale T, Gregory P. Injuries of the shoulder girdle. *Clinical Orthopaedics and Related Research* 1995; **318**: 54–60. [A succinct, thorough review of shoulder girdle injuries.]

Hooper GJ, Keddell RG, Penny ID. Conservative management or closed nailing for tibial shaft fractures: a randomised prospective trial. *Journal of Bone and Joint Surgery* 1991; **73B**: 83–5. [A well-designed study demonstrating superior functional results after nailing of closed, displaced tibial fractures compared with cast treatment.]

Johansen K, Daines M, Howey T, Helfet D, Hansen ST. Objective criteria accurately predict amputation following lower extremity trauma. *Journal of Trauma* 1990; **30**: 568–73. [MESS is described, and is evaluated in a combined retrospective/prospective study. A MESS value greater than or equal to 7 predicted amputation in both studies with 100 per cent accuracy.]

Jupiter JB, Lipton H. The operative treatment of intraarticular fractures of the distal radius. *Clinical Orthopaedics and Related Research* 1993; **292**: 48–61. [The results of surgical management of 13 complex intra-articular distal radius fractures are presented in a thorough review of the anatomy, classification, fixation techniques, and aftercare.]

Letournel E. Acetabulum fractures: classification and management. *Clinical Orthopaedics and Related Research* 1980; **151**: 81–106. [A must-read description of the anatomy, classification, radiology, and surgical treatment of acetabular fractures, written by the foremost expert.]

Matta JM, Saucedo T. Internal fixation of pelvic ring fractures. *Clinical Orthopaedics and Related Research* 1989; **242**: 83–97. [Discusses the diagnosis, classification, and rationale for treatment of pelvic ring injuries in three groups of patients: treated non-operatively, by anterior external fixation, and by internal fixation.]

Müller ME, Nazarian J, Koch P, Schatzker J. *Comprehensive classification of fractures of long bones*. Springer-Verlag, Berlin, 1990. [The current academic standard for fracture classification.]

Ochsner MG Jr *et al.* Associated aortic rupture-pelvic fracture: an alert for orthopedic and general surgeons. *Journal of Trauma* 1992; **33**: 429–34. [Aortic rupture occurs eight times more often in patients with anteroposterior compression pelvic fracture patterns than in the general blunt trauma population.]

Pape HC, Auf'm'kolk M, Paffrath T, Regel G, Sturm JA, Tscherne H. Primary intramedullary femur fixation in multiple trauma patients with associated lung contusion—a cause of posttraumatic ARDS? *Journal of Trauma* 1993*a*; **34**: 540–8. [The paper that caused international interest in the issue of femoral reaming in polytrauma patients. Careful reading of the paper reveals that the incidence of an important clinical factor—bilateral pulmonary contusion—is not equal in the two subgroups.]

Pape HC, Regel G, Dwenger A, Sturm JA, Tscherne H. Influence of thoracic trauma and primary intramedullary nailing on the incidence of ARDS in multiple trauma patients. *Injury* 1993*b*; **24** (Suppl. 3): 82–103. [A comprehensive review of many basic science and clinical investigations of the physiologic effects of femoral reaming.]

Swiontkowski MF. Outcomes measurement in orthopaedic trauma surgery. *Injury* 1995; **26**: 653–7. [A review of the problems of regional practice variation, evidence-based medicine, and outcomes assessment as t(ey relae to orhopaedi trauma.]

Tscherne H, Gotzen L. *Fractures with soft tissue injuries*. Springer-Verlag, Berlin, 1984. [The standard reference for the classification and treatment of soft-tissue injuries.]

Tscherne H, Regel G, Pape HC, Pohlemann T, Krettek C. Internal fixation of multiple fractures in patients with polytrauma. *Clinical Orthopaedics and Related Research* 1998; **347**: 62–78. [A very thorough algorithm covering all aspects is explained in detail, providing references for clinical decision making.]

46.2.2 The adult shoulder

A. J. Carr

Introduction
Tissue structure and function
The glenohumeral joint
Capsule
Subacromial space
Acromioclavicular joint
Sternoclavicular joint
Biomechanics
Glenohumeral mobility
Shoulder stability
Examination
Investigation
The unstable shoulder
Classification
Impingement and rotator cuff disease
Non-surgical treatment
Surgical treatment
Complications
Massive rotator cuff tears
Cuff tear arthropathy
Fractures
Proximal humeral fractures
Clavicle fractures
Arthritis of the shoulder
Surgical management of rheumatoid arthritis
Osteoarthritis of the shoulder
Frozen shoulder
Further reading

Introduction

Shoulder disorders are common and shoulder pain is the most frequent musculoskeletal problem after back pain to present in primary care. The shoulder is also the commonest joint in the human body to dislocate. Fractures to the shoulder joint complex span all age groups from clavicular fractures in children to four part proximal humeral fractures in the elderly. Shoulder problems constitute an important area of orthopaedic practice and this chapter aims to summarize the diagnosis and management of these disorders.

Tissue structure and function

The shoulder girdle comprises three joints: glenohumeral, acromioclavicular, sternoclavicular; and two main spaces the subacromial and scapulothoracic.

The glenohumeral joint

This comprises the humerus and glenoid of the scapula. The radius of curvature of the humeral head in the coronial plane is 24 mm (plus or minus 1.2 mm). The radius of curvature of the glenoid is 2.3 mm greater than that of the humeral head. The articular cartilage of the humeral head is thickest at its centre and of the glenoid at its periphery. The humeral head is retroverted to the humeral shaft to an average of 30 degrees (plus or minus 11 degrees). The glenoid is tilted superiorly by 5 degrees and posteriorly by 7 degrees in relation to the long axis of the scapula. The glenoid is surrounded by a fibrous labrum and is attached through a fibrocartilaginous transition zone. The long head of biceps is attached to the superior part of the labrum. Also attached to the labrum are glenohumeral ligaments of which the inferior is the most substantial.

Capsule

The glenohumeral ligaments form part of the shoulder joint capsule and are shown in Fig. 1. The inferior ligament consists of a thick anterior band and thinner and much more inconsistent posterior band. This ligament becomes taut as the arm abducts and externally rotates. The long head of biceps passes through the glenohumeral joint and acts as a joint stabilizer and as a humeral head depressor.

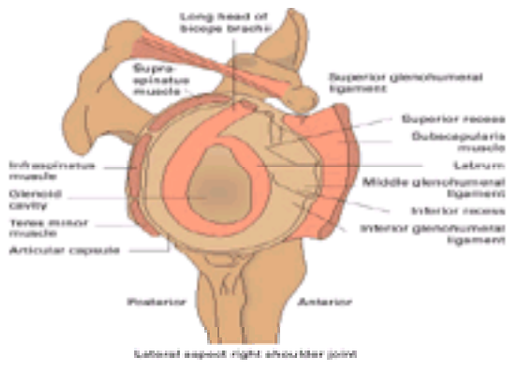


Fig. 1. Demonstrating the glenoid of the scapula with its labrum, capsule, and ligaments.

Subacromial space

The glenohumeral joint is surrounded by the tendons of the rotator cuff muscles. The subscapularis muscle attaches to the lesser tuberosity and the supraspinatus, infraspinatus, and terres minor muscles to the greater tuberosity. The tuberosities are separated by the groove of the long head of biceps tendon. Around the superior part of the cuff is a bursa. This lies beneath the acromion and coraco-acromial ligament. These two structures together form an arch above the shoulder joint. The morphology of this space differs between individuals. In some people the space is relatively small and this pre-disposes to rubbing or attrition of the superior surface of the cuff on the undersurface of the acromion. The relationship of the rotator cuff muscles and the acromion are shown in Fig. 2.

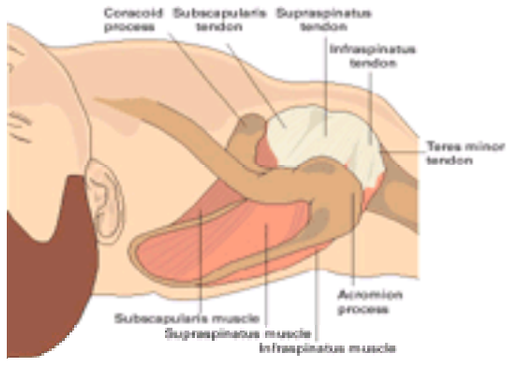


Fig. 2. View of the shoulder from above demonstrating the relationship of the supraspinatus tendon to the anterior edge of the acromion. The coracoid and the acromion are connected by a ligament.

Acromioclavicular joint

The acromioclavicular joint is triarthrodial. It has a shallow concave facet on the acromion and a convex surface on the distal clavicle. A meniscus is attached to the superior capsule and is usually incomplete.

The joint is stabilized by strong ligaments that pass from the coracoid process of the scapular to the undersurface of the clavicle medial to the joint.

Sternoclavicular joint

This small joint relies heavily on its stability from ligamentous structures surrounding the joint. The capsular ligaments allow 30 degrees of upward and downward movement along with 45 degrees of rotation along the long axis.

Biomechanics

The glenohumeral joint has a greater range of motion than any other joint in the human body, this can produce problems of instability.

Glenohumeral mobility

The primary role of the shoulder is to allow forces to transfer between the arm and the torso. These forces can be carried through: (a) contact between articular surfaces; (b) soft tissues such as ligament and joint capsule; and (c) the active control of the 18 muscles that cross the four articulations of the shoulder complex.

Clinically, movement of the shoulder may be described as abduction, adduction, forward flexion, extension internal, and external rotation. This system does not adequately describe the complex three-dimensional movements that occur. As a consequence other systems have been tried such as the Eulerian co-ordinate system and the screw displacement axis system. These involve three-dimensional descriptions of the relative positions of the humerus in space and require three angular rotations and three linear translations. Investigations of these movements can be accomplished by means of *in vitro* and *in vivo* studies. These assessments are often accompanied by quantification of the actions of the various muscle groups around the shoulder using electromyography.

These studies were first described by Poppen and Walker in the 1970s and demonstrated various patterns of movement at the shoulder. One particularly interesting feature is the relative rhythm of the glenohumeral joint and the scapula thoracic articulation. The scapular moves inconsistently during the first 30 degrees of upward motion and then in a constant ratio of 2:1 until maximal elevation. Experimental analysis of this simple movement is so difficult that the precise relationship of humerus, scapula, and thorax remains controversial. What does appear clear from simple clinical observation is that the rhythm of the shoulder is often affected fairly early on in any shoulder pathology.

Shoulder stability

Stability of the shoulder joint is provided by passive and dynamic stabilizers. It is an interesting observation that, although the joint surface area of the shoulder is small, the shoulder joint remains in position despite loss of muscle tone in anatomical specimens or anaesthetized patients. Table 1 lists the passive mechanisms that are involved in maintaining joint stability.

1.	Bony features and joint conformity
2.	Joint pressure
3.	Adhesion and cohesion
4	Ligaments and capsules
5.	Glenoid labrum

Table 1 Passive stabilizers of the shoulder

The bony contributors to stability are small. The effect of the glenoid labrum is to improve the containment of the humeral head within its socket. There is a negative pressure within the shoulder joint of approximately 4 mm of mercury. If this is equalized with atmospheric pressure then the shoulder will begin to sublux. The SYnovial fluid of the joint provides viscose and intermolecular forces which allow for a low coefficient of friction but a high surface tension. These factors significantly improve the stability of the shoulder.

Perhaps the most significant stabilizers of the shoulder are its ligaments. Of these the most substantial is the inferior glenohumeral ligament which originates from the glenoid labrum at the 2 to 4 o'clock position. The function of the ligament is described in Fig. 3.

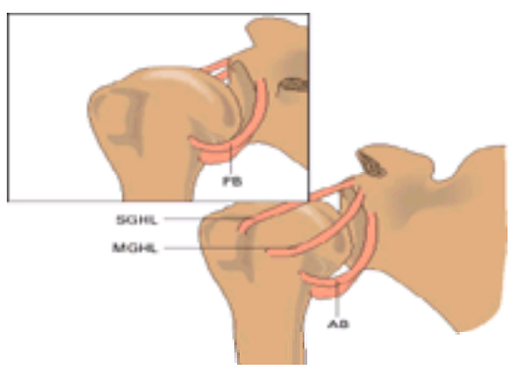


Fig. 3. The main section of this illustration reveals the anterior ligaments of the shoulder. SGHL, superior glenohumeral ligament; MGHL, middle glenohumeral ligament; AB, anterior band of the inferior glenohumeral ligament. The inset illustration reveals the posterior band of the glenohumeral ligament (PB).

The dynamic stabilizers of the shoulders are described in Table 2.

1. Passive muscle tension from the size effect of the muscle
2. Active contraction causing compression of the articular surface
3. Joint movement causing secondary tension in the ligamentous constraint
4. The barrier effect created by the contraction of muscle

Table 2 Dynamic factors contributing to stability of the shoulder joint

Shoulder instability is a complex problem and often involves failure of more than one of the elements of the shoulder which contribute to its stability.

Examination

Good history taking and examination are of paramount importance in the diagnosis of disorders of the shoulder. Most shoulder disorders present with a combination of pain, stiffness, instability, and weakness, and can be evaluated without the need for expensive investigation. Some general medical conditions may be of relevance for example rheumatoid arthritis and diabetes. Some conditions may produce symptoms in the shoulder and upper limb such as disorders of the cervical spine and polymyalgia rheumatica.

The age of the patient may provide significant clues to a possible diagnosis. Between the age of 10 to 30 years trauma and instability are the commonest problems. Between the ages of 40 and 60 the commonest diagnoses are impingement syndrome, adhesive capsulitis, inflammatory joint disease, and trauma. In the older age group (60 to 80 years) full thickness rotator cuff tears, osteoarthritis, and trauma most commonly occur. Shoulder instability is a common problem and almost invariably the patient will volunteer that their shoulder feels unstable. Instability is rarely a symptom described by a patient if the problem is anything other than shoulder subluxation or dislocation.

On the other hand patients do not frequently describe weakness unless there is a reasonably profound loss of strength. The commonest symptom to present by far is that of pain and it is important to determine in what circumstances and in what positions the shoulder gives rise to pain.

Examination should always begin with an assessment of the cervical spine. Then move to observation, paying careful attention to any evidence of asymmetry, swellings, and muscle wasting. The patient may describe the site of their pain. This is often not well localized with disorders from the shoulder joint or rotator cuff. Pain from the acromioclavicular joint is usually much better localized.

Then assess the range of movement. This is classically described as being forward flexion and backward extension, abduction and adduction, and internal and external rotation (Fig. 4, Fig. 5, Fig. 6, Fig. 7, and Fig. 8). Test for weakness of these movements. It is particularly important to test for strength of external rotation as this will be markedly reduced in cases of rupture of the rotator cuff involving the infraspinatus. Specific tests of rotator cuff disorders involve impingement signs and rotator cuff weakness signs. A large number of different methods of examining for impingement are described. Perhaps the most common is to elevate the arm to 80 degrees in the scapula plane, the thumb is pointed to the ground so that the greater tuberosity comes to lie beneath the acromion and the patient is asked to resist adduction. This draws the greater tuberosity on to the undersurface of the acromion and at the same time stresses the supraspinatus. The patient with a rotator cuff or subacromial bursa disorder will complain of pain. This sign maybe abolished after the injection of local anaesthetic into the subacromial space known as the impingement test of Near. Figure 9 demonstrates a positive impingement test. Figure 10 demonstrates weakness of the infraspinatus by testing external rotation with the arm by the side.

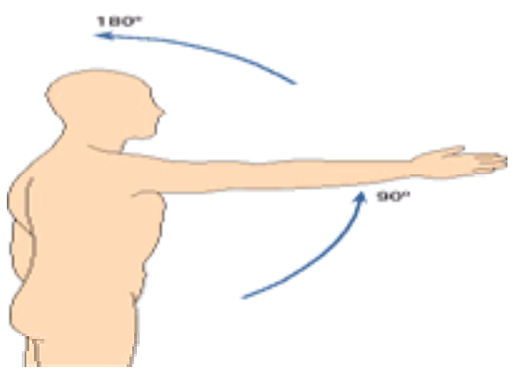


Fig. 4. Describing flexion at the shoulder.

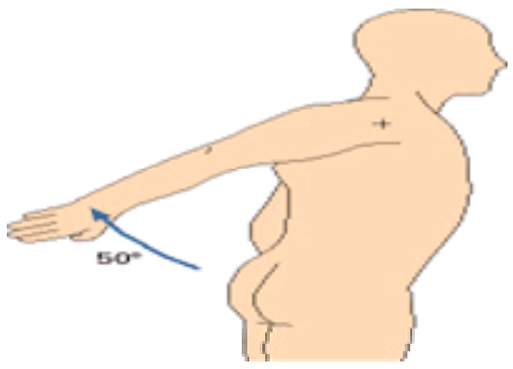


Fig. 5. Extension at the shoulder.

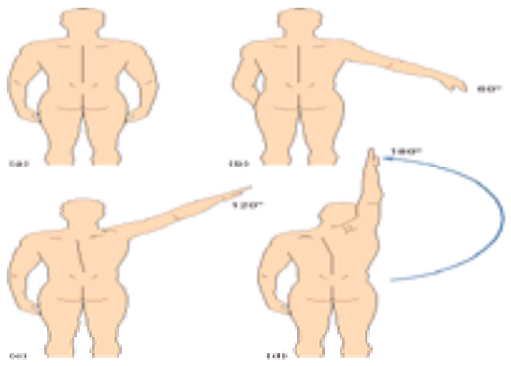


Fig. 6. Describing abduction of the shoulder.

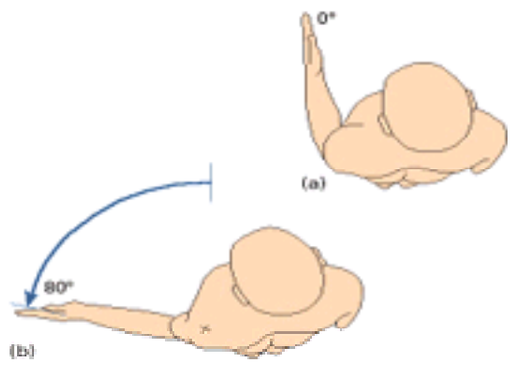


Fig. 7. External rotation of the shoulder with the arm adducted.

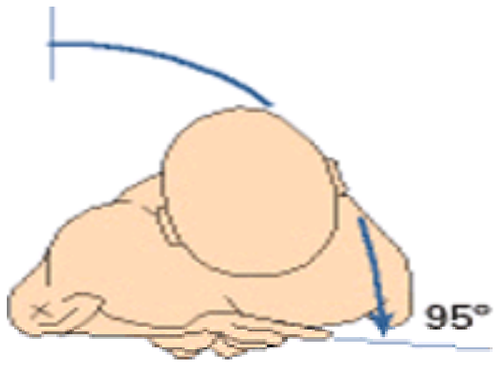


Fig. 8. Internal rotation with the arm reaching behind the back.



Fig. 9. Testing for impingement. The arm is placed at 60 degrees of abduction with the thumb pointing towards the ground. The patient is asked to resist adduction of the arm. If this produces pain beneath the acromion then the test is positive.

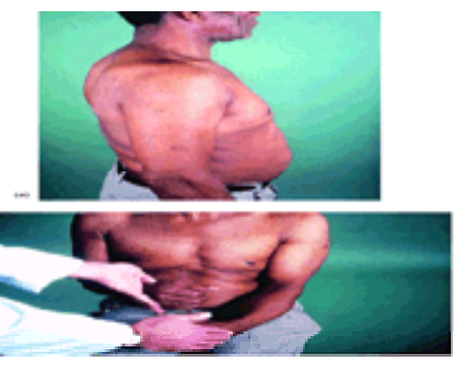


Fig. 10. (a) Demonstrating wasting of the infraspinatus and supraspinatus muscles. (b) Demonstrating weakness of external rotation of the humerus due to rupture of the infraspinatus tendon.

Assessment of the biceps tendon is with palpation of the bicipital groove and with stressing of the biceps tendon by attempting to overcome flexion and supination. If the long head of biceps is ruptured then the muscle will recoil into the distal part of the upper arm producing the characteristic swelling in this region (Fig. 11).



Fig. 11. Rupture of the long head of biceps produces a lump of biceps muscle in the distal arm when the muscle is contracted.

Examination for instability of the shoulder should begin with testing for generalised joint laxity of all joint groups. In individuals with generally lax joints the shoulders may be hypermobile. In children with this type of shoulder instability the shoulder joint may be inferiorly subluxed with gravity. (The Sulcus sign Fig. 12.)



Fig. 12. In multi-directional shoulder instability the humerus can be subluxed inferiorly in the glenohumeral joint producing a Sulcus sign.

Traumatic forms of instability can be tested for by means of an apprehension test. The commonest direction of instability is anterior and inferior and this is tested in 90 degrees of abduction and external rotation as shown in Fig. 13.



Fig. 13. The anterior apprehension test for anterior instability of the shoulder. The glenohumeral joint is placed in 90 degrees of abduction and the humerus is externally rotated. The patient will describe apprehension if the joint is unstable.

Investigation

Good clinical evaluation should be accompanied by basic investigation of the shoulder. This usually involves a true anteroposterior and axillary radiograph. If the axillary radiograph is not obtainable due to pain or loss of movement then a scapula lateral is a useful alternative.

Some special views are also possible. These include the West Point axillary view for glenoid rim fractures. An anteroposterior view with the shoulder in internal rotation, or a Stryker notch for the assessment of Hill–Sachs lesions. It is important to rule out associated fractures or loose bodies in conditions of instability. Figure 14 demonstrates an anteroposterior radiograph of a dislocated shoulder and an axillary view showing the humeral head lodged over the glenoid producing a Hill–Sachs defect in the humeral head.

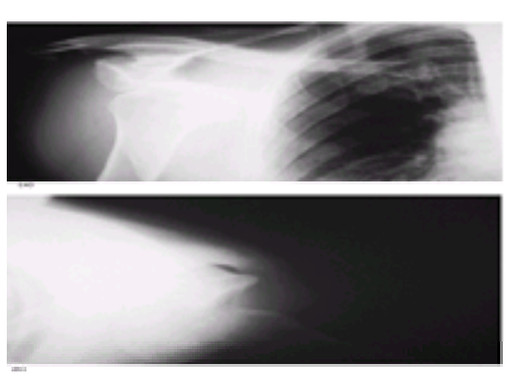


Fig. 14. (a) A radiograph demonstrating anterior glenohumeral dislocation. (b) An axillary view of an anterior dislocation of the shoulder with notching of the posterior humeral head. The indentation is described as a Hill–Sachs lesion.

The further investigation of rotator cuff disorders is an area that has changed dramatically over recent years. Twenty years ago the most common investigation for this problem was an air contrast arthrogram. These investigations produced an accuracy of between 80 and 90 per cent at determining full thickness rotator cuff tears but were complicated by being invasive and time consuming. More recently investigation of the cuff has been changed by the use of magnetic resonance imaging (MRI) and ultrasonography. Both these investigation have accuracies of between 90 and 98 per cent. The advantage of ultrasound is that it is relatively inexpensive and quick to perform, the disadvantage is that it requires an experienced ultrasonographer to be certain of such levels of accuracy. MRI probably has slightly higher accuracy than that of ultrasound and is less user dependent. It is, however, more expensive and time consuming.

The unstable shoulder

Shoulder instability is relatively common and accounts for 45 per cent of all joint dislocations in the human body (17 per 100 000 population per year). Some of these are recurrent dislocations and the incidence of primary dislocations is 12 per 100 000 per year. There are two age-related peaks one in men between the age of 21 and 30 years and the second in women between 61 and 80 years. Approximately 7 per cent have associated nerve lesions with a 3 per cent incidence of axillary nerve damage. A random sample of the Swedish population between the ages of 18 and 70 years found a prevalence of 1.7 per cent in the general population and 8 per cent in ice hockey players. Once an anterior dislocation has occurred there is a significant risk of further episodes of instability. For those aged under 25 years the recurrence rate is 60 per cent which drops to 25 per cent for those aged over 34 years. If dislocation occurs over the age of 40 years then it is often associated with damage to the rotator cuff mechanism.

Shoulder instability has been described as being of two patterns, the first being traumatic, unidirectional, and associated with a Bankart lesion and the second being atraumatic, multi-directional, and bilateral.

This is almost certainly an oversimplification of a condition that represents a continuum between the congenitally lax and poorly co-ordinated and those involved in significant trauma.

Closer inspection of individuals suffering traumatic anterior dislocations of the shoulder reveals that they often have a family history of the problem and approximately 20 per cent will have an episode of instability in the opposite shoulder at some time.

Classification

The classification of shoulder instability is described in Fig. 15. Essentially, this involves determining whether or not the disclocations are involuntary or voluntary. Whether their origins are traumatic or atraumatic and then whether they are anterior posterior or multi-directional. It is also useful to determine whether or not the instability is recurrent and whether or not it involves simply subluxation or complete dislocation of the joint.

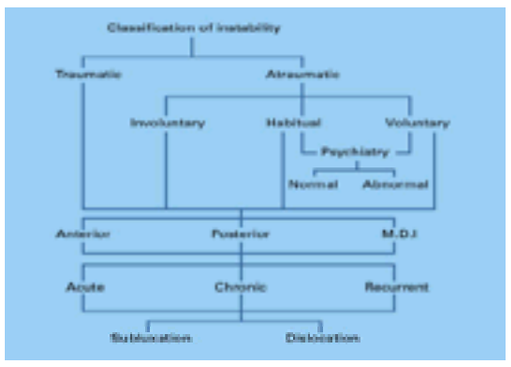


Fig. 15. Classification of shoulder instability.

Anterior instability

This is by far the commonest direction of shoulder instability. It usually occurs after some form of trauma involving forced abduction and external rotation of the arm. During the dislocation the capsule attachment to the glenoid is disrupted creating a Bankart lesion. The posterior surface of the humeral head maybe damaged by the anterior edge of the glenoid during the dislocation producing a Hill–Sachs lesion. In younger individuals nearly two-thirds will have further episodes of instability. These episodes of instability may be either subluxations or true dislocations of the joint. The history is usually clear and is confirmed on clinical examination. These patients are often involved in further investigations but it is doubtful whether or not these investigations add to the information received from good clinical examination and history taking.

The treatment of these conditions is usually with surgery. Reports of the use of physiotherapy are described but in true recurrent instability it seems unlikely that physiotherapy has a great part to play. Methods of surgical stabilization of the shoulder are described in Table 3.

Arthroscopic methods	*Capsular procedures*	*Bankart repair*
		Capsular tightening
Open methods		
	Capsular procedures	Capsular shift
		Capsular overlap (Putti–Platt)
		Bankart repair
	Muscular procedures	Overlap of the subscapularis (Putti–Platt)
		Lateral advancement of the lesser tuberosity (Magnuson–Stack)
		Coracoid transfers (Bristow procedure)
	Humeral osteotomy	
	Glenoid osteotomy	

Table 3

Surgical treatment for shoulder instability was first popularized in the 1950s with the Bankart repair and the Putti–Platt operation. The recurrence risks following surgery varies dramatically between 0 and 30 per cent. Modern methods of stabilizing the shoulder involve a combination of Bankart repair and capsular shift. These procedures are described in Fig. 16 and Fig. 17. The recurrence risk for these procedures is between 2 and 10 per cent. Recently, attempts have been made to stabilize the shoulder with arthroscopic methods. A relatively small number of series are published in the literature but many of these demonstrate dramatically worse recurrence risks of between 30 and 50 per cent. The advantage of the capsular shift and Bankart operations over other procedures such as the Putti–Platt operation is that there is generally less restriction of external rotation which is important for those individuals wishing to return to sport.

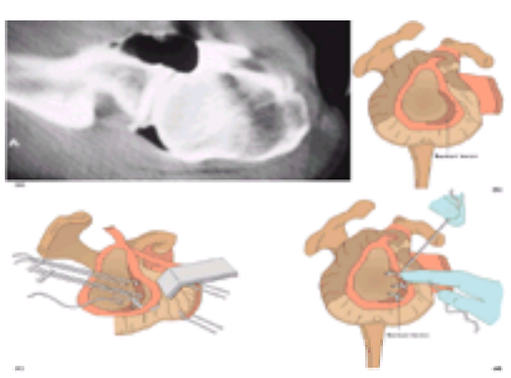


Fig. 16. (a) Transverse computed tomography scan of a shoulder with recurrent anterior instability. The anterior labrum can be seen to be stripped from the glenoid. This is a Bankart lesion. (b) This schematic illustration describes the typical position of a Bankart lesion with stripping of the antero-inferior labrum from the glenoid. (c)

Demonstrates the re-attachment of the anterior labrum to the glenoid with sutures through holes in bones. (d) Shows these sutures being tightened and with restoration of the normal anterior labrum of the shoulder.

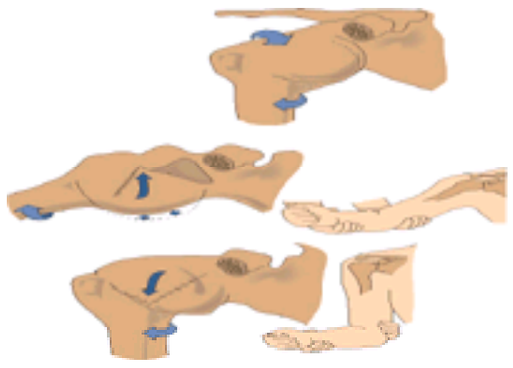


Fig. 17. Describes the principles of the capsular shift. A T-shape opening is made in the anterior capsule of the shoulder joint. The inferior flap of capsule is raised superiorly and sutured into position. The upper flap is then sutured over the lower flap. This procedure causes circumferential shortening of the capsule restricting antero-inferior glide of the humeral head in external rotation and abduction. Its advantage is that there is no excessive shortening of the anterior structures to prevent external rotation.

Posterior instability

This is a much rarer pattern of instability and is often associated with a more multi-directional pattern. A number of surgeons are now attempting to stabilize the shoulder by means of capsular shrinkage using either laser or cautery. If these methods fail then posterior stabilization of the shoulder may be necessary. This surgery can involve either glenoid-osteotomy tilting the glenoid anteriorly or posterior capsular reefing and infraspinatus reefing. There are relatively few reports in the literature of outcome of these procedures but failure rates as high as 50 per cent are reported. Complications include damage to the axillary and suprascapular nerve.

Multi-directional instability

This is a much rarer pattern of instability and much more complex. Generally there is no single cause for multi-directional instability and often there is a predominant direction of instability, usually posterior and inferior. These patients often present during adolescence at the time of the adolescent growth spurt. Investigation should always include a generalized assessment of the patient to exclude other conditions such as Ehlers–Danlos syndrome. Strength of the shoulder should be assessed carefully as this may be improved by selective physiotherapy. Investigation with complex imaging of the shoulder is usually not helpful. Treatment is almost exclusively with the use of rehabilitation programmes. This often requires specialist help from physiotherapists used to dealing with this problem. Should the physiotherapy fail to be of benefit then shrinkage of the capsule using arthroscopic techniques including laser or cautery may be beneficial. Such techniques have not been in practice for long and the long-term outcomes are not known. An alternative method of stabilization is with an inferior capsular shift. Recurrence rates following this type of surgery are high and have been reported at about 50 per cent.

Impingement and rotator cuff disease

Shoulder pain is an increasingly recognized problem in the ageing population. In Dutch studies the incidence is reported at being between 12 and 25 per 1000. In national surveys from England and Wales the incidence has been reported as 6.6 per 1000. It is the commonest musculoskeletal complaint after back pain to present to general practitioners and at present only a small proportion are referred to a specialist. Much of this pathology presenting in general practice are disorders of the rotator cuff. Many patients have long-term SYmptoms and of those patients presenting for the first time to a general practitioner nearly 50 per cent still have symptoms at 12 months.

The precise basis and natural history of disorders of the rotator cuff is still unclear. There does appear to be a relationship between this condition and the morphology of the coraco-acromial arch. Neer believed that the impingement syndrome was caused by the congenital shape of the acromion. If it was hooked it caused attrition on the supraspinatus leading to impingement and rotator cuff tear. It appears clear that the shape of the acromion does differ between individuals. However, the truly hooked or type III acromion is acquired and not congenital. The intrinsic problem causing impingement syndrome and rotator cuff failure lies within the rotator cuff itself. This is especially so of the leading or anterior edge of the supraspinatus just behind the long head of biceps tendon. Acromial morphology is shown in Fig. 18. It should be noted that the configuration of the coraco-acromial ligament is ‘Y' shaped. It has a complex attachment to the anterior and inferior surface of the acromion. The leading edge of the supraspinatus tendon lies just beneath this intersection of the ligament and acromion. During abduction the supraspinatus tendon may rub this undersurface producing impingement. The clinical presentation of impingement is with a classic painful arc. The patient notices pain on abduction beyond 70 degrees this pain may disappear on full abduction as the greater tuberosity moves underneath the acromion.

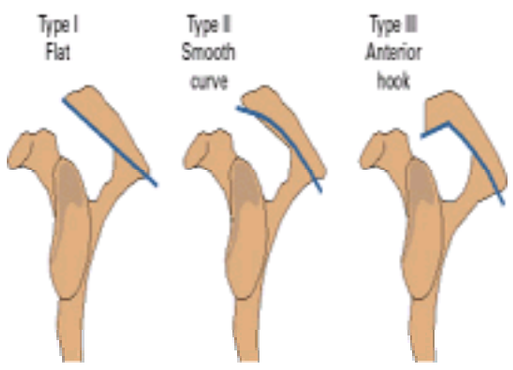


Fig. 18. Describing the three shapes of acromion: type I, flat; type II, smooth curved; and, type III, anterior hook.

The pathogenesis of rotator cuff failure is probably vascular in origin. The supraspinatus tendon is relatively poorly supplied with blood and begins to degenerate through an age-related process. As the tendon degenerates it swells and may begin to impinge on the undersurface of the acromion. The inconsistent element of this process is that some patients complain of pain but others do not. The zone of the supraspinatus tendon which is most likely to degenerate is known as the critical zone. This intrinsic failure of the rotator cuff is then superimposed on by extrinsic factors involving impingement on the undersurface of the acromion. This produces attrition damage to the superior surface of the cuff. Symptoms from this pathology are variable and may improve with rest. They may also be improved by the use of physiotherapy and hydrocortisone injection. Recent symptomatic reviews of the use of cortisone injections have not produced any conclusive evidence that they are better than injection of saline alone. Further studies are needed in this area. It is clear that impingement syndrome overlaps with true tears of the rotator cuff. These have been estimated to develop in up to 80 per cent of people aged 60 years and older. What is inconsistent is the relationship of symptoms to the presence of tears. Many tears of the supraspinatus are asymptotic and these have been observed in MRI studies of asymptotic volunteers. This evidence has been backed up by data from cadaveric studies. The frequency of complete and partial rotator cuff tears ranges from 5 to 39 per cent in cadaveric studies and 13 to 37 per cent in MRI studies of adults in the general population. This clearly has a bearing on the management of this condition and a decision to operate on patients with impingement and rotator cuff tear should not be based on imaging alone.

Non-surgical treatment

By far the majority of patients presenting with impingement and rotator cuff tear are aged over 40. Physical examination of the patient includes testing for impingement and also assessment of the integrity of the rotator cuff. This examination may be supplemented by the injection of local anaesthetic into the subacromial bursa to see if

symptoms improve. Investigation of the rotator cuff with plain radiographs may demonstrate narrowing of the acromiohumeral interval. The features of progressive failure of the rotator cuff are demonstrated in Fig. 19 which shows progression from a normal acromiohumeral space to advanced cuff tear arthritis. Plain radiographs have an accuracy of approximately 80 to 85 per cent at determining full thickness rotator cuff tear. This is substantially less than the accuracy of ultrasound and MRI (Fig. 20).

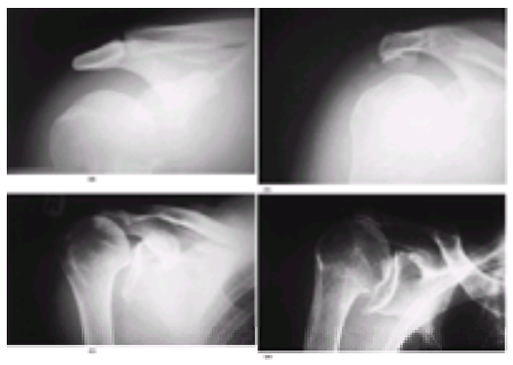


Fig. 19. (a) A radiograph of the subacromial space showing a normal appearance of the undersurface of the acromion and acromioclavicular joint. (b) Demonstrating formation of anterior osteophytes on the undersurface of the acromion. (c) With loss of the rotator cuff muscles the humeral head subluxes superiorly causing abnormal articulation of the glenohumeral joint and physical contact between the humerus and the undersurface of the acromion. (d) If superior subluxation of the glenohumeral joint persists then arthritis may develop in the glenohumeral joint. This is known as cuff tear arthropathy.

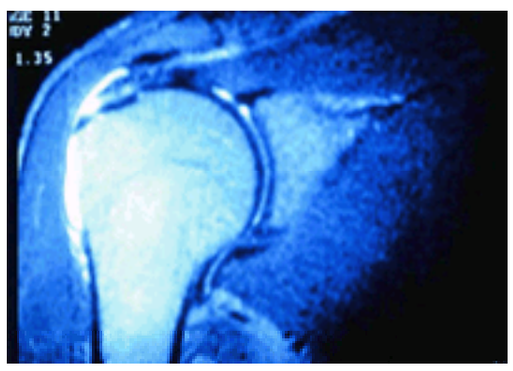


Fig. 20. An MRI scan of the shoulder demonstrating a small tear of the supraspinatus tendon close to its insertion into the greater tuberosity.

The non-surgical treatment of impingement and rotator cuff tears includes the use of non-steroidal anti-inflammatory medication and cortisone injection. It may also be supplemented by appropriately directed physiotherapy rehabilitation programmes. An excellent systematic review of randomized control trials for non-surgical treatment of shoulder pain by Green et al. (1998) found that there was little evidence to either support or refute the efficacy of common interventions for shoulder pain.

Surgical treatment

Generally speaking surgical intervention is indicated only for patients with chronic symptoms which have not settled with rest and with conservative methods. The available surgical techniques are subacromial decompression and coraco-acromial ligament excision with and without repair of any rotator cuff tear. Anterior acromioplasty was first described by Neer in 1972 and since then several series have reviewed the outcome of this procedure. An average success of 81 per cent has been described for open acromioplasty compared with 76 per cent for the arthroscopic procedure. Complications are infrequent. The principle of the procedure is shown in Fig. 21.

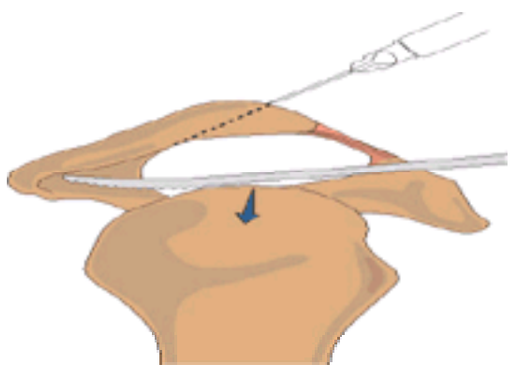


Fig. 21. Demonstrating the principles of an anterior acromioplasty. The coraco-acromial ligament is released and the hooked anterior edge of the acromion resected.

Full thickness rotator cuff tears begin at the critical zone in the supraspinatus. They often start as a horseshoe-shaped tear. Surgical interventions for repair of full thickness tears are good and excellent in between 85 and 90 per cent of patients. Longer-term follow-up of these patients demonstrate that the majority of patients with good results will maintain these good results at between 5 and 10 years. Relatively few of these studies have followed the patients using ultrasound or MRI to determine whether or not the repair remains intact. There does appear to be a correlation in outcome between those patients that remain intact in the longer term.

Complications

A number of complications arise as a result of surgery to the rotator cuff. This may involve mistakes in diagnosis of the underlying pathology. Further complications arise as a result of failure of the deltoid re-attachment. This is particularly so if extended approaches are made to the deltoid and if it is detached from the lateral border of the acromion. Further failures may arise as a result of inadequate decompression of the acromion or excessive decompression of the acromion, including acromion fracture. In some circumstances hetero-topic ossification may occur. Excessive excision of the distal clavicle may result in clavicular instability.

Perhaps the commonest complication is re-rupture of the rotator cuff. This appears to be particularly common in attempted repair of large and massive tears (measuring 3 to 5 cm and greater than 5 cm). Occasionally, shoulder surgery is complicated by postoperative stiffness or a form of frozen shoulder. This condition usually resolves with time and operative intervention is very rarely required.

Massive rotator cuff tears

These constitute a special problem. A massive rotator cuff tear is defined as one being greater than 5 cm and will therefore involve the supraspinatus, infraspinatus, and possibly part of the subscapularis and terres minor muscles. Unfortunately, the presentation of rotator cuff disease is often with this type of problem. The patient has either only relatively recently developed symptoms or has had SYmptoms for many years but the rotator cuff disease has gone unrecognized and possibly untreated. The precise reasons why rotator cuff tears do not present until such a late stage in the pathology of degeneration is not known and further investigation in this area is required. Some 5 cm tears are relatively easy to mobilize and the tendon does not appear to have atrophied and become dysfunctional. It is often difficult to determine this prior to surgery. Increasing use of imaging of the shoulder may allow the quality of the remaining muscle to be determined with higher accuracy. The

surgical options for the massive rotator cuff tear are:

1. simple debridement and decompression
2. local muscle repairs of transfers
3. distant muscle transfers
4. free tissue grafts
5. synthetic grafts
6. arthroplasty
7. arthrodesis

The role of simple subacromial decompression is controversial. It appears that whatever pain relief may be afforded is relatively short lasting and after 5 years only 50 per cent of patients demonstrate any improvement. Where possible the local transfer of either the infraspinatus or subscapularis muscle is recommended. However, long-term results from this type of technique are uncertain. Somewhat more controversially distant muscle transfers using the latissimus dorsi muscle have been tried. It appears that the results from this type of surgery are reasonable with regard to pain relief but very little active humeral movement is achieved from the muscle transfer. The results of free tissue grafts have been universally poor and a static non-viable graft, particularly if synthetic materials are used, does significantly increase the risk of infection.

Arthroplasty has been used with some success in painful massive rotator cuff tears. Improvement in movement is usually only minimal but significant improvement in pain can be achieved by use of a hemi-arthroplasty. This has considerable advantages over arthrodesis where, although pain is relieved, no movement at all occurs at the glenohumeral joint.

Cuff tear arthropathy

Once the rotator cuff tear becomes advanced and the glenohumeral joint develops incongruity arthritis may develop. It is interesting to speculate why some shoulders with massive rotator cuff tears develop arthritis and some do not. The surgical treatment of this condition involves:

1. surgical debridement and acromioplasty
2. glenohumeral arthrodesis
3. constrained shoulder arthroplasty
4. semi-constrained shoulder arthroplasty
5. bipolar arthroplasty
6. hemi-arthroplasty

Surgical debridement and acromioplasty has some limited short-term benefit but probably long-lasting benefit is minimal. Constrained and semi-constrained shoulder replacement may produce some improvement in pain but because of the absence of the rotator cuff muscle there is significant risk of loosening of the implant particularly the constrained device and the glenoid component of the semi-constrained device. Recently, bipolar arthroplasty has become popular but some recent reports suggest that it is less beneficial than hemi-arthroplasty. The principle of hemi-arthroplasty is simply to interpose tissue between the undersurface of the acromion and the humerus and this operation does appear to produce good relief of pain.

Fractures

Proximal humeral fractures

Fractures of the proximal humerus account for about 5 to 7 per cent of all fractures. They are more frequent in the elderly. Most are associated with no or minimal displacement and can be treated without surgery with physiotherapy. Accurate diagnosis of the fracture pattern and type is essential and requires radiographs to be taken in two planes either anteroposterior and axillary or anteroposterior and scapular lateral. The most widely used classification of proximal humeral fractures is that described by Neer and is shown in Table 4 (1970 *a*). This classification is based on the position of four fragments: the head, the lesser tuberosity, the greater tuberosity, and the shaft. The classification describes the number of fracture fragments and is subdivided to describe the displacement of these fracture fragments. The evaluation of the fracture may be significantly improved in complex cases by the use of computed tomography.

Fracture	Treatment
One-part minimal displacement	Conservative treatment, early mobilization after 2 weeks
Two-part head (anatomical neck) displacement	Open reduction and internal fixation
Two-part shaft (surgical neck) with displacement	
(a) Impacted	If there is more than 45 degrees of angulation closed reduction is preferred. Internal fixation is only rarely necessary
(b) Unimpacted	Closed reduction and careful observation. If closed reduction fails and fracture still unstable percutaneous pin fixation. If closed reduction unsuccessful and soft tissue interposed open reduction and internal fixation.
(c) Comminuted	If significant comminution occurs then conservative treatment maybe preferred.
Two-part greater tuberosity with displacement	Open reduction and internal fixation is preferred.
Three part with displacement	Open reduction and internal fixation. If blood supply to the head is ensured then hemi-arthroplasty.
Four-part fracture with dislocation	Some may be managed with internal fixation but most require hemi-arthroplasty

Table 4

The principal goal of fracture management is to allow healing of the bone without loss of function. Eighty-five per cent of proximal humeral fractures are minimally displaced and can be treated conservatively. A relatively shortened period of immobilization between 2 and 3 weeks is advised and physiotherapy can commence soon after this. Many of these fractures are in elderly relatively osteoporotic women and in approximately 80 per cent a good result will be obtained. In unstable displaced fractures surgical treatment needs to be considered. Some two-part fractures can be managed with closed reduction. This may be supplemented by percutaneous pinning under radiographic control. Surgical neck fracture dislocations may require internal fixation. Open reduction and internal fixation is the usual treatment for most three- and four-part fracture dislocations. If the head fragment is dislocated and the bone very porotic then replacement of the humeral head and re-attachment of the tuberosities is the preferred treatment.

A significant number of methods of internal fixation have been described. One commonly applied method is to use intramedullary wires or rush pins supplemented with tension band wire around the tuberosities and sutures. An alternative is to use a plate fixation and screws. Staples have also been used. Mal-union of these fractures is common and the associated soft tissue damage caused by the fracture and the subsequent surgery often produces scarring with resultant loss of function and some times pain.

If the humeral head has been dislocated or has its soft tissue attachments significantly disrupted then there is a high rate of avascular necrosis. In cases where the vascular supply to the humeral head is seriously doubted then a primary hemi-arthroplasty is preferred.

Isolated fractures of the greater tuberosity are best treated surgically. These fractures often displace and mal-union of this fracture will produce loss of function and pain. Figure 22 shows an isolated fracture of the greater tuberosity with displacement that has been fixed with screws and supplemented with suture. Table 4 describes the common treatment for fractures of the proximal humerus.

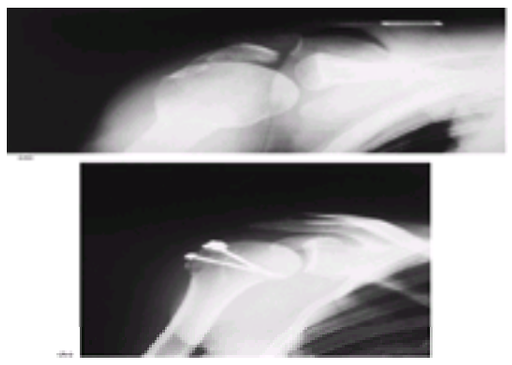


Fig. 22. (a) Demonstrating subluxation of the glenohumeral joint with fracture of the greater tuberosity. (b) The greater tuberosity fragment has been re-attached using screws and sutures. The glenohumeral joint is now in a more normal position.

The complications of these fractures include nerve injury which often occurs at the time of the fracture but may occur at the time of surgical fixation. One common complication of this type of fracture management is mal-union. This may cause bony blocks to movement. Avascular necrosis of the humeral head may occur particularly in three- and four-part fractures. Scarring and damage to the rotator cuff is also common and may cause abnormalities of movement and pain. Heterotopic bone formation may occur particularly after more displaced fractures.

Clavicle fractures

Clavicle fractures represent about 10 per cent of all fractures and are the most common fracture in childhood. They are usually easy to diagnose and with very few exceptions unite without complications. Clinical evaluation should include an assessment of the site of the fracture and deformity. Any excessive stretch of the skin should be noted. Three per cent of clavicle fractures will present with an associated injury of either the brachial plexus, vascular injury, or pneumothorax. These are almost invariably high energy injuries. Plain radiographs will delineate the fracture in most instances but sometimes a computed tomography scan is required for medial injuries. Middle third clavicle fractures represent 80 per cent of all clavicle fractures. These usually heal without intervention. Lateral third clavicle fractures account for 15 per cent of injuries and they are classified into type I distal to the coracoclavicular ligament, type II medial to the coracoclavicular ligament, and type III is an intra-articular fracture of the acromioclavicular joint.

Type II distal clavicle fractures have the highest rate of non-union. However, 70 per cent of these will heal without internal fixation and a conservative method is employed initially. If the fracture fails to unite then treatment with plate and bone graft is usually recommended.

Arthritis of the shoulder

The shoulder is commonly involved in inflammatory and rheumatoid arthritis. Over 90 per cent of patients with rheumatoid arthritis will describe shoulder pain and approximately a third will have severe problems with the shoulder associated with their disease. Generally speaking, the initial symptoms are due to intense synovitis and this can be managed with medical treatments and possibly with radionucleotide synovectomy. In more advanced disease the shoulder joint becomes degenerate and cartilage is destroyed. There is often associated disease of the rotator cuff which, although intact, may not function efficiently as a head depressor. This causes upward migration of the humeral head with associated impingement. Almost invariably the shoulder is part of a polyarthritic presentation and with rheumatoid patients the function and limitations of other joints must be taken into consideration before surgery is contemplated.

Surgical management of rheumatoid arthritis

Synovectomy

Isolated SYnovectomy of the shoulder is relatively infrequently performed. It has a place in younger patients with florid synovitis which is not controlled by medical means.

Subacromial decompression and rotator cuff repair

In a subset of patients the predominant symptoms are not from the joint itself but from associated rotator cuff disease. In these patients subacromial decompression and rotator cuff repair is indicated. Great care should be taken, however, in resection of the coraco-acromio ligament as this is contra-indicated in patients with arthritis of the joint.

Shoulder arthroplasty

Arthrodesis of the shoulder and resection arthroplasty have demonstrated extremely poor results. This is particularly when compared with the more conventional treatment in modern practice of shoulder replacement. Often in rheumatoid arthritis the rotator cuff is defective and hemi-arthroplasty is preferred in these patients.

Over 90 per cent of patients will achieve satisfactory relief of pain. Only modest improvement in movements are achieved. Complications of the operation include infection. The rate of infection is difficult to calculate from the published series but is probably less than 5 per cent. Fracture may also occur particularly in weak osteoporotic bone of rheumatoid patients. Instability is also described and is particularly associated with previous surgery to the coraco-acromial ligament.

Osteoarthritis of the shoulder

This may be primary or secondary to fracture or trauma. It presents with pain and restriction of movement. Assessment is with plain radiographs. Further imaging is rarely necessary. Management of the condition is with analgesic medication and many patients will cope without the need for surgery. In more advanced cases of arthritis with severe symptoms shoulder replacement is considered. In primary osteoarthritis of the shoulder the rotator cuff is often intact and the ideal treatment is with a total shoulder replacement. Figure 23 demonstrates an osteoarthritic joint with a total joint replacement for treatment. If the shoulder arthritis is secondary to tear of the rotator cuff or following trauma where the rotator cuff is damage then hemi-arthroplasty is currently advised. As many as 10 per cent of all shoulder replacements are complicated in some way. Early complications include infection, nerve damage, and dislocation. Later complications include loosening and peri-prosthetic fracture.

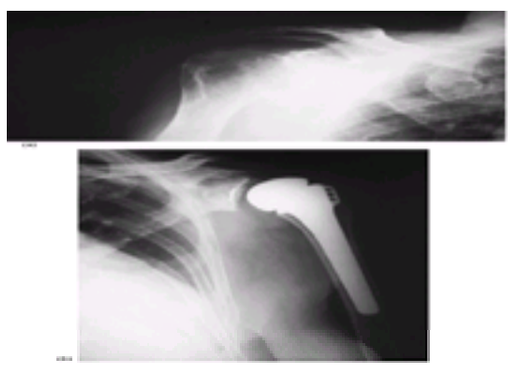


Fig. 23. (a) Osteoarthritis of the glenohumeral joint with loss of articular surface peri-articular sclerosis and osteophyte formation. (b) The shoulder has been replaced with a total shoulder replacement. This is a modular shoulder with a separate humeral head and humeral shaft. A polyethylene glenoid is cemented into the scapula and is revealed with a marker pin.

Generally speaking, the results are excellent with 90 per cent of patients obtaining good relief of pain. This is particularly so for the treatment of primary osteoarthritis. In the management of osteoarthritis secondary to rotator cuff tear pain relief is less reliable and is of the order of 70 to 90 per cent. If the rotator cuff is intact then the range of movement is usually good but is poor if the rotator cuff is deficient.

Frozen shoulder

Frozen shoulder has been recognized since the early part of the twentieth century and remains something of a mystery in terms of its aetiology. Twenty-two per cent of all shoulder problems presenting in general practice are defined as being capsular in origin and most of these are probably frozen shoulder. The condition has been described in three phases. The first freezing phase is both painful and stiff, in the second thawing phase the pain reduces but stiffness persists. It is in the third resolving phase that movement tends to return. It is generally a condition characterized by global loss of both active and passive movement initially associated with pain. It does tend to be self-resolving, although it may take 2 years for the condition to run its course.

The typical patient presenting with the condition is in middle age. There maybe a history of minor trauma but many present with no apparent cause. The condition is associated with diabetes and also with the use of anti-epileptic medication. Plain radiographs of the shoulder are normal. Arthrograms of the shoulder will demonstrate a contracted joint with small volume. Histologically, the capsule has been described as demonstrating an inflammatory process with some fibroblast proliferation and possible fibromatosis. The condition has been associated with Dupuytren's disease, although it differs from this condition in being generally self-limiting.

The treatment of the condition is initially with pain relief. Physiotherapy in the first phase of the condition will exacerbate the symptoms and is best avoided. However, in the later stages of the condition exercises and physiotherapy may help to improve movement. Steroid injections have been given in this condition but are of doubtful value. In the resistant later phase of the condition manipulation under anaesthetic has been attempted and controlled studies have demonstrated improvement. This management is now being supplemented by arthroscopic distension and release. The long-term results of this type of procedure have yet to be evaluated.

Further reading

Arntz CT, Jackins S, Mattsen FA. Prosthetic replacement of the shoulder for the treatment of defects of the rotator cuff and the surface of the glenohumeral joint. *Journal of Bone and Joint Surgery* 1993; **75A**; 485–91. [Review of 21 shoulders followed for 9 years. Some improvement in range of movement was noted with significant improvement in pain in 15 of 21 shoulders.]

Bankart ASB. Recurrent or habitual dislocations of the shoulder joint. *British Medical Journal* 1933; **2**: 1133. [The classic description of the essential lesion of anterior shoulder instability with a lesion of the anterior capsule.]

Bigliani LU, Morrison DS, April EW. The morphology of the acromion and rotator cuff impingement. *Orthopedic Transactions* 1986; **10**: 288. [A classic description of three types of acromion morphology with a type III or hooked acromion being most commonly associated with rotator cuff tears in a cadaver study.]

Biglianni LU *et al.* Operative treatment of failed repairs of the rotator cuff. *Journal of Bone and Joint Surgery* 1992; **74A**: 1505–15. [Retrospective review of repeat rotator cuff surgery in 31 patients. Only 52 per cent had successful outcome.]

Biggliani LU *et al.* Shift of the postero-inferior aspect of the capsule for recurrent posterior glenohumeral instability. *Journal of Bone and Joint Surgery* 1995; **77A**: 1011–20. [Thirty-four patients treated with posterior inferior capsular shift with 80 per cent satisfactory results.]

Carr AJ. Biomechanics of shoulder stability. *Clinical Orthopaedics* 1996; **10**: 146–51. [Overview of biomechanical aspects of shoulder stability.]

Clark JM, Harriman DT. Tendons ligaments and capsule of the rotator cuff, gross and microscopic anatomy. *Journal of Bone and Joint Surgery* 1992; **74A**: 713–25. [Cadaveric tendons examined grossly and microscopically. The microstructure of the rotator cuff tendons is described showing distinct layers with interdigitations.]

Cofield RH. Physical examination of the shoulder: Effectiveness in assessing shoulder stability. In: Mattson FA, Foo FUFH, Hawkins RJ, eds. *The shoulder: A balance of mobility and stability*, pp. 331–43. American Academy of Orthopaedic Surgeons, Rosemont, IL, 1993. [This is a review of examination of the glenohumeral joint including assessment of stability and movement.]

Cooper DE *et al.* Anatomy histology and vascularity of the glenoid labrum: An anatomical study. *Journal of Bone and Joint Surgery* 1992; **74A**: 46–52. [Twenty-three fresh frozen shoulders were examined. The anatomy of the biceps tendon, glenoid labrum, and capsular described. They also examine the arterial supply of the glenoid labrum showing it to be limited to the periphery of this structure.]

Green D *et al.* Systematic review of randomised control trials of intervention for painful shoulder: Selection criteria, outcome assessment and efficacy. *British Medical Journal* 1998; **71**(28): 354–60. [Comprehensive metra-analysis of non-surgical intervention demonstrating that there is little evidence to either support or refute the use of common interventions for shoulder pain. A need for further well designed clinical trials is emphasized.]

Harriman DT *et al.* Repairs of the rotator cuff. Correlation of functional results with integrity of the cuff. *Journal of Bone and Joint Surgery* 1991; **73A**: 982–9. [Describing the association between the integrity of the repair and subsequent results. The worse results were found in those patients in whom the rotator cuff had re-ruptured.]

Hawkins RJ, Misselmore GW, Pehobeika. Surgery for full thickness rotator cuff tears. *Journal of Bone and Joint Surgery* 1985; **67A**: 1349–55. [Classic paper describing the success of rotator cuff tendon repair. Large tears are much more successful if repaired then if treated with decompression and acromioplasty alone.]

Hovelius L, Falling J, Fredin H. Recurrent anterior dislocation of the shoulder. Results after the Bankart and Putti-Platt operations. *Journal of Bone and Joint Surgery* 1979; **61A**; 566–9.

Iannotti JP *et al.* Magnetic resonance imaging of the shoulder: Sensitivity specificity and predicted values. *Journal of Bone and Joint Surgery* 1991; **73A**: 17–29. [Demonstrating that MRI has a sensitivity of 100 per cent and specificity of 95 per cent for full thickness rotator cuff tear. Imaging for instability is less good with a sensitivity of 88 per cent and specificity of 93 per cent.]

Iannotti JP *et al.* The normal glenohumeral relationships an anatomical study of 140 shoulders. *Journal of Bone and Joint Surgery* 1992; **74A**: 491–500. [One hundred and forty-one shoulders were examined in cadavers showing the relationships of the humerus and glenoid; MRI scans also performed.]

Iannotti JP *et al.* The normal glenohumeral relationships: An anatomical study of 140 shoulders. *Journal of Bone and Joint Surgery* 1992; **74A**: 491–500. [This study reviews cadavers and magnetic resonance imaging. Measurements of orientation of the humeral head and glenoid are provided.]

Kelly IG. Unconstrained shoulder arthroplasty in rheumatoid arthritis. *Clinical Orthopedics* 1994; **307**: 94–102. [Excellent review of the possibilities for different surgical modalities in rheumatoid arthritis.]

Lohr JF, Uhthoff HK. The microvascular pattern of the supraspinatus tendon. *Clinical Orthopedics* 1990; **254**: 35–8. [Describing the microvascular pattern within the supraspinatus tendon and its critical zone. This appears to be a watershed area.]

Lyons AR, Thomlinson JE. Clinical diagnosis of tears of the rotator cuff. *Journal of Bone and Joint Surgery* 1992; **74B**: 414–15. [Examination of 42 patients prior to surgery and compared with surgical findings. Clinical examination demonstrated sensitivity of 91 per cent and specificity of 75 per cent.]

Neer CS. Fractures of the distal third of the clavicle. *Clinical Orthopedics* 1968; **58**: 43–50. [Classic description of fractures of the distal third of the clavicle and their classification.]

Neer CS. Displaced proximal humeral fractures. Part I: Classification and evaluation. *Journal of Bone and Joint Surgery* 1970*a*; **52A**: 1077–89. [Classic description of the classification of the proximal humeral fractures.]

Neer CS. Displaced proximal humeral fractures. Part II: Treatment of three part and four part displacement. *Journal of Bone and Joint Surgery* 1970*b*; **52A**: 1090–103. [Classic paper describing an approach to the management of three and four part fractures with displacement.]

Neer CS. Replacement arthroplasty for glenohumeral osteoarthritis. *Journal of Bone and Joint Surgery* 1974; **65A**: 1–13.

Neer CS, Watson KC, Stanton FJ. Recent experience in total shoulder replacement. *Journal of Bone and Joint Surgery* 1982; **64A**: 319–37. [Classic article reporting 50 rheumatoid shoulders with greater than 2-year follow-up. The majority have satisfactory relief of pain.]

Neer CS. Anterior acromioplasty for the chronic impingement syndrome in the shoulder. Preliminary report. *Journal of Bone and Joint Surgery* 1972; **54A**: 4150. [The classic description of anterior acromioplasty for the impingement syndrome. Neer believed that rotator cuff tears where caused by attrition from the anterior acromion.]

Nordqvist A, Peterson C, Redlund-Johnelli J. The natural course of lateral clavicle fractures, 15 year follow-up of 110 cases. *Acta Orthopaedica Scandinavica* 1993; **64**: 97–1. [Fifteen-year follow-up of 110 patients. Ninety-five (86 per cent) were asymptomatic including eight with non-union, 15 had moderate pain and dysfunction, and two had painful non-union.]

Ozaki J *et al.* Recalcitrant chronic adhesive capsulitis of the shoulder role of contracture of coraco-humeral ligament and rotator interval in pathogenesis and treatment. *Journal of Bone and Joint Surgery* 1989; **71A**: 1511–15.

Palmer WE, Caslowitz PL. Anterior shoulder instability: Diagnostic criteria determined from prospect analysis of 121 MRI arthrograms. *Radiology* 1995; **197**: 819–25. [Conclusion that MRI arthrograms without Bankart lesions were not useful in predicting instability.]

Poppen NK, Walker PS. Normal and abnormal motion of the shoulder. *Journal of Bone and Joint Surgery* 1976; **58A**: 195–201. [Classic article describing the motion of the shoulder.]

Rockwood CA, Lyons FR. Shoulder impingement SYndrome. Diagnosis, radiographic evaluation and treatment with a modified Neer acromioplasty. *Journal of Bone and Joint Surgery* 1993: **75A**: 409–24. [Seventy-one patients were treated with open acromioplasty with good results.]

Rockwood CA, Williams GR, Birkhead WZ. Debridement of degenerative irreparable lesions of the rotator cuff. *Journal of Bone and Joint Surgery* 1995; **77A**: 857–66. [Retrospective review of 50 patients treated for irrepairable cuff tears. Satisfactory results in 83 per cent where the deltoid was undamaged and the long head of biceps intact.]

Sher JS *et al.* Abnormal findings on magnetic resonance images of asymptomatic shoulders. *Journal of Bone and Joint Surgery* 1995; **77A**: 10–15. [Ninety-six normal volunteers were examined with MRI scans. Complete and partial rotator cuff tears were found in 14 and 20 per cent of individuals, respectively. In people aged over 60, 28, and 26 per cent had complete and partial tears, respectively. This has implications for the management of symptomatic cuff tears.]

Turkel SJ *et al.* Stabilising mechanisms preventing anterior dislocation of the glenohumeral joint. *Journal of Bone and Joint Surgery* 1981; **63A**: 1208–17. [Another classical study demonstrating the importance of the inferior glenohumeral ligament in external rotation and abduction.]

Zukerman JD, Cuomo F, Rokito S. Definition and classification of frozen shoulder; A consensus approach. *Journal of Shoulder and Elbow Surgery* 1994; **3**: 72. [This article under the umbrella of the American Shoulder and Elbow Surgeons defines frozen shoulder.]

46.2.3 The hand and wrist

Edward Akelman and Arnold-Peter C. Weiss

Fractures and dislocations of the hand and wrist
Fractures of the distal phalanx
Fractures of the middle phalanx
Fractures of the proximal phalanx
Intracondylar fractures
Intra-articular fracture dislocations of the proximal interphalangeal joint
Distal interphalangeal joint dislocations
Proximal interphalangeal joint dislocations
Fractures of the metacarpals
Carpal fractures
Scaphoid fractures
Triquetral fractures
Ligament injuries of the hand and wrist
Pathophysiology of wrist ligament injuries
Ligament injuries of the thumb
Flexor tendon injuries of the hand
Anatomy
Diagnosis
Flexor tendon injury management
Extensor tendon injuries
Anatomy
Injury management
Dupuytren's contracture
Anatomy
Pathobiology
Incidence
Treatment
Hand infections
General principles
Cellulitis
Paronychia
Felon
Infections of the web space
Pyogenic flexor tendon tenosynovitis
Thenar space, midpalmar space, and hypothenar space infections
Infections of the radial and ulnar bursa
Human and animal bites
Herpes simplex infections
Amputations of the hand
Fingertip amputations
Thumb amputations
Middle phalanx amputations
Amputations through the proximal phalanx
Metacarpal and carpal amputations
Further reading

Fractures and dislocations of the hand and wrist

Fractures of the distal phalanx

The most common fracture in the hand is that of the distal phalanx. These injuries are commonly associated with pulp compartmental crush, nail bed injuries, and soft tissue losses. Most trauma results in comminuted fractures, without displacement. Immobilization of the fingertip and the distal interphalangeal joint with a dorsal or volar splint allows these fractures to heal: most closed fractures heal in 14 to 21 days. Irrigation, debridement, and careful repair of the nail bed should be accomplished at the time of the initial injury to allow cosmesis of the nail and functional return. Subungual hematomas should be drained through the intact hard nail using electrocautery or a heated needle, allowing relief of severe pain. When the hematoma covers 50 per cent or more of the nail area, the nail should be removed and nail bed injury sought.

Most displaced distal phalanx fractures are reduced by closed reduction methods with the use of a digital nerve block and local anesthesia. Flexion of the distal fragment with a dorsal splint over the distal interphalangeal joint in slight flexion overcomes the pull exerted by the flexor digitorum profundus on the proximal distal phalanx fragment. Unstable displaced fractures occasionally require fixation with a small Kirschner wire, placed longitudinally with the distal interphalangeal joint in extension. Displacement or gapping between the two distal phalanx fracture fragments usually indicates entrapment of the nail matrix. Open reduction and internal fixation with repair of the nail bed matrix is required.

Fractures of the middle phalanx

Extra-articular fractures of the middle phalanx are uncommon. Middle phalanx fractures may be displaced, depending on the level of the fracture and its relationship to the flexor digitorum superficialis insertion. An extra-articular fracture of the base of the middle phalanx may collapse and angulate dorsally. Fractures of the neck of the phalanx distal to the superficialis insertion angulate in a volar direction. A fracture located in the midshaft between these two levels may angulate in either a dorsal or volar direction.

Stable non-displaced fractures of the middle phalanx should be treated with a dorsal splint allowing active distal interphalangeal phalanx motion when tolerated. Timing of motion can be determined by careful palpation of the fracture site. Most patients do not complain of pain after 2 to 4 weeks, are assumed healed, and motion may then be encouraged.

Displaced fractures should be treated with closed reduction. The hand and wrist should be placed in an intrinsic plus position with the wrist resting at 20° of dorsiflexion, the metaphalangeal joints flexed to 70°, and the interphalangeal joints held in extension. If the middle phalanx fracture reduction is unstable, unacceptable, or if it is irreducible, it may be treated by either closed or open reduction and internal fixation. The best functional results after these injuries are obtained in patients who are allowed to move the proximal and distal joints: early motion keeps both extensor and flexor tendons from becoming adherent to the fracture.

Fractures of the proximal phalanx

Fractures of the proximal phalanx of the hand are common. Stable fractures of the proximal phalanx with minimal displacement may be treated much in the same manner as stable middle phalanx fractures, in a cast with the wrist and hand placed in an intrinsic plus position. However, many proximal phalanx fracture patterns are inherently unstable. Following high-energy injuries, transverse fractures in the proximal third to middle regions of the proximal phalanx may collapse into an apex volar angular deformity. With this displacement, closed reduction is obtained by flexing the metacarpophalangeal joint to 90°, releasing the deforming forces of the intrinsic musculature. Spiral oblique proximal phalanx fractures may be unstable and are reduced by correcting distal bony deformities. Fractures at the base of the proximal phalanx can usually be reduced by this same maneuver.

After reduction of the proximal phalanx fracture, the digit should be held in an intrinsic plus position. As these fractures have a high probability of displacement, radiographs should be obtained weekly for 3 to 4 weeks to check alignment.

Unstable fractures, fractures that displace, and reductions felt to be unacceptable require either closed reduction or open reduction and internal fixation. Both methods

have provided excellent results. The long-term functional result of these common fractures seems to be improved when treatment allows early motion of the injured digit.

Intracondylar fractures

Fractures of the middle and proximal phalanx intracondyles are usually associated with articular cartilage displacement in the form of unicondylar, bicondylar, or osteochondral ‘slices'. These are common injuries in athletes. Closed reduction should be attempted, even when there is significant joint displacement, although open reduction and internal fixation is usually required. Depending on the size of the fracture fragments, multiple Kirschner wire or small screw fixation may be applicable. Early protected motion improves excellent functional results.

Intra-articular fracture dislocations of the proximal interphalangeal joint

Fractures at the base of the middle phalanx may cause dorsal subluxation or dorsal dislocation of the proximal interphalangeal joint. Correct treatment of these injuries is most dependent on recognition of the injury pattern: anteroposterior and true lateral radiographs of the joint are required. If only one-third of the articular surface is involved, these injuries are generally stable. Fractures affecting more than one-third of the articular surface involved are usually unstable.

Distal interphalangeal joint dislocations

True dislocations of the distal interphalangeal joint without fracture are uncommon. They are usually dorsal and frequently open, with a tell-tale transverse laceration at the level of the distal interphalangeal joint flexion crease. They should be treated as open injuries with irrigation and debridement of the joint. Reduction is usually stable, but fixation may be necessary if the joint is truly unstable. The joint is held in 10 to 20° of flexion for approximately 4 weeks, and active motion of the metaphalangeal and proximal interphalangeal joints encouraged.

Proximal interphalangeal joint dislocations

Proximal interphalangeal joint dislocations, particularly dorsal dislocations, are common. These are usually injuries to the volar plate of the joint at its insertion on the middle phalanx. Most can be reduced with digital block anesthesia. Treatment includes 2 weeks of dorsal proximal interphalangeal joint splinting in approximately 15 to 20° of flexion. Early motion is encouraged by taping the injured finger to an adjacent finger for the next 6 weeks. This encourages motion while using the adjacent digit as a splint. Follow-up of these injuries within10 days is needed to ensure that the reduction is stable. True lateral radiographs are essential. Overzealous treatment of these injuries, such as 4 to 6 weeks of splinting or casting, gives poor results, with limitation of joint motion. Splinting for longer than 2 weeks gives poorer results.

Volar dislocations of the proximal interphalangeal joint are rare. When reducible they may be treated with a dorsal extension splint, with the dislocated joint held in extension for 6 weeks. These injuries may be associated with rupture of the central slip of the extensor ten-don or herniation of the proximal phalanx through the extensor mechanism, causing irreducible dislocation of the joint. Injury usually occurs through the interval between the central slip and the ipsilateral lateral band. The condyle may become trapped, causing the joint to be irreducible: open exploration is then required. Repair of the acute central extensor insertion is required, and Kirschner wire pinning of the joint in extension for 4 weeks should be carried out. If the lateral band is entrapped, the extensor mechanism is repaired. A postoperative extension splint is worn for 6 weeks.

Fractures of the metacarpals

As in most fractures, initial treatment depends on the type of fracture, the fracture pattern, whether the fracture is open or closed, and whether the fracture is intra- or extra-articular. The index and long finger metacarpals are the stabilizing central (fixed) unit of the hand. The thumb on the radial side and the ring and small finger on the ulnar side of the hand move about this central fixed unit; they have been described as the mobile units of the hand. Angulation, shortening, and displacement of these three metacarpals have less impact on the function of the hand because of their mobility. For this reason, only minimal displacement should be allowed in metacarpal fractures of the index and long fingers, whereas more displacement and angulation may be allowed in the thumb, ring finger, and small finger metacarpals.

Extra-articular metacarpal fractures

Fractures of the neck of the metacarpal are the most common of the extra-articular metacarpal fractures. These injuries are caused by a direct blow to the metacarpal heads. Depending on the level of energy imparted to the hand at the time of injury, these fractures may be unstable because of palmar comminution of fracture fragments. Radiographs are required to monitor the fracture reduction. Most of these fractures may be held using a short arm cast holding the wrist in extension, the metacarpophalangeal joints in 70° of flexion, and the proximal interphalangeal joints in full extension. The acceptable degree of residual angulation is controversial. In general, 0 to 10° of angulation is allowed in the index and long and 30 to 40° of angulation is allowed in the ring and fifth finger metacarpals.

Metacarpal shaft fractures are usually the result of direct blows to the dorsum of the hand, longitudinal compression, or torsional stresses. Angulatory or rotatory deformities may occur, depending on the pull of different muscular forces. Angulation is less acceptable in the index and long metacarpals than in the ring and small fingers. Most fractures can be reduced using longitudinal traction, wrist extension, metacarpophalangeal joint flexion, and dorsal three-point molding. Unstable or malrotated fractures may be treated by closed or open reduction and internal fixation.

Metacarpophalangeal joint dislocations

Dorsal dislocations of the metacarpophalangeal joint are the result of a hyperextension injury with injury to the volar plate. These dislocations may be reduced using wrist block anesthesia to allow relaxation of intrinsic musculature. The most gentle reduction is usually accomplished by flexing the wrist to relax the extrinsic flexor tendons. The metacarpal head is then reduced into the sulcus of the normal metacarpophalangeal joint with volar pressure exerted onthe distally displaced proximal phalanx. Longitudinal traction will not assist, and may actually impede, reduction.

Some dorsal metacarpophalangeal dislocations may be irreducible. These complex dislocations are indicated by the classic volar skin change of puckering, most commonly at the level of the distal palmar crease of the hand. The metacarpal head is usually prominent in the palm. The structure usually preventing reduction is the volar plate. Open reduction is performed through a curving incision over the palmar portion of the metacarpophalangeal joint. The neurovascular structures are usually held in close apposition to the underlying skin, and care must be taken to avoid injury to this complex when making the incision.

Intra-articular metacarpal fractures

These fractures involve the border digits of the hand, the index finger and the little finger being most commonly involved. They are usually caused by a direct blow that causes axial compression, and are frequently associated with human bite injuries. Infection of the joint should always be considered. If large articular fragments are involved, open reduction and internal fixation with restoration of joint surface congruity is of paramount importance.

Intra-articular fractures of the base of the metacarpals are most common in the thumb and in the little finger. In these injuries, there is subluxation of either the thumb or little finger carpometacarpal joint. When this injury occurs at the base of the thumb, it is called a Bennett's fracture. The volar ulnar tubercle of the thumb metacarpal remains attached to the volar beak ligament. With continued pull of the abductor pollicis longus tendon, the large remnant thumb metacarpal subluxes dorsoradically and into relative supination. The adductor pollicis muscle pulls the shaft of the thumb metacarpal towards the second metacarpal. As in all articular fracture subluxations, it is of prime importance to restore articular congruity. This may be accomplished by either a closed or open reduction and internal fixation. Closed reduction may be carried out by pulling longitudinally on the thumb metacarpal, followed by radial abduction and pronation, placing the thumb metacarpal in relative palmar abduction.

Fractures of the little finger metacarpal base simulate Bennett's fracture dislocation of the thumb. The little finger metacarpal subluxes because of the pull of the extensor carpi ulnaris tendon. It is also relatively adducted by the hypothenar intrinsic musculature. Again, attention must be paid to careful reduction of the joint. Reduction is carried out by relative longitudinal traction, ulnar abduction, and supination. Percutaneous pinning can then be carried out.

Carpometacarpal joint dislocations

Dislocations of the carpometacarpal joints of the metacarpals of the hand are often missed. These injuries, caused by longitudinal compression combined with flexion of the affected metacarpals, may be dorsal or volar. They are more commonly dorsal and involve the fourth and fifth metacarpal bases because of the inherent instability of the anatomy of this region. The key to this diagnosis is a ‘pronated 30° lateral' where the hand is actually supinated 60° from the true lateral on a

radiograph.

Dislocations of the finger carpometacarpal joints are usually easy to reduce using wrist nerve block anesthesia and finger trap traction of approximately 4.5 kg (10 pounds). Reduction may be aided by direct finger pressure on the metacarpal base, combined with extension of the metacarpal head. Maintenance of these reductions is controversial. Percutaneous Kirschner wire fixation with short arm casting for approximately 4 weeks may be required. In these casts, the metacarpophalangeal joint is usually flexed 70° with the wrist extended 20°. The pins are removed at 4 to 6 weeks and early motion is begun after that time. Pain and weakness may result from unrecognized and unreduced fracture dislocations. Treatments reported for late arthrosis have included arthrodesis or interposition arthroplasty.

Carpometacarpal dislocations of the thumb are rare and usually occur in a dorsoradial direction. These may be treated with a careful reduction confirmed by radiograph and a short arm thumb spica cast for 4 weeks, with interphalangeal motion allowed. If the joint subluxes after reduction, most patients develop pain, deformity, and pinch weakness.

Carpal fractures

The scaphoid is the most common carpal bone to be fractured.

Scaphoid fractures

Second in incidence only to fractures of the distal radius, scaphoid fracture s most frequent in young adult males at the level of the scaphoid waist. The scaphoid bone is almost completely covered with hyaline cartilage. The scaphoid interosseous blood supply to the proximal two-thirds of the scaphoid enters in the dorsal ridge region. Vascular injection studies of the scaphoid have shown that approximately 13 per cent of specimens would lose retrograde blood supply following a scaphoid waist fracture. Because of vascular variations from patient to patient, it is difficult to determine whether avascular necrosis will occur after a fracture.

Scaphoid fractures occur when a load is applied to the radial side of the palm with the wrist dorsiflexed 95 to 100°. With the wrist in radial deviation, there is a bending moment in the waist of a volar flexed scaphoid between the radius, capitate, and volar capsular ligaments. This causes a fracture between the supported and the unsupported bones.

The diagnosis of scaphoid fractures depends on an index of suspicion, a careful clinical physical examination, and good radiographs. A scaphoid fracture should be suspected if the patient complains of dorsoradial ‘snuff box' tenderness, pain in ulnar and radial deviation in the same area, or tenderness in the palmar pole of the scaphoid. A standard wrist radiographic series should include a posteroanterior view, a lateral view, a posteroanterior view with the wrist in ulnar deviation, and two oblique views—posteroanterior, with the wrist pronated 45° from the lateral, and posteroanterior with the wrist sup-inated 45° from the lateral. Any patient with tenderness on clinical examination or suspicion of fracture on a radiograph should be treated as if a fracture is present. Radiographs should be repeated in 2 to 4 weeks if there is still doubt regarding the injury.

The anatomic location of a scaphoid fracture may help determine the possibility of union because of the anatomy of the blood supply. Seventy per cent of these fractures occur in the middle third or waist of the scaphoid, 20 per cent occur within the proximal third, and approximately 10 per cent occur distally. Fractures in the proximal and to a somewhat lesser extent the middle third of the scaphoid are at higher risk for delayed union and/or avascular necrosis. If clinical examination 3 months after fracture shows pain, increasing instability, angulation, or radiographic evidence of delay in healing, open reduction and bone grafting should be carried out.

Triquetral fractures

The triquetral fracture is more commonly associated with a hyperextension or hyperflexion injury of the wrist. It is frequently an avulsion fracture in the area of insertion of the extensor carpi ulnaris. The injury may be caused by compression or impingement. Oblique and lateral radiographs are helpful in making a diagnosis. Short arm casting for 4 to 6 weeks gives excellent results. In the rare patient who experiences worsening pain during use, surgical removal of a small dorsal fragment may be required when a non-union is noted.

Ligament injuries of the hand and wrist

The strong ligaments of the hand and wrist complement the bony architecture and provide stability during functional activity. The most important ligaments of the human wrist joint are volar, thick, and intracapsular.

Pathophysiology of wrist ligament injuries

Perilunate dislocation

Perilunate dislocations are the result of a violent force exerted with the wrist in hyperextension, ulnar deviation, and intercarpal supination. Patients may present with dorsal wrist discomfort, mild swelling, and numbness. Range of motion is usually limited. True lateral radiographs will confirm the diagnosis, the capitate being dislocated dorsally on the lunate. Under either regional or general anesthesia, the hand is suspended using finger traps with the elbow flexed 90°. Weights of 4.5 to 7 kg (10 to 15 pounds) may be needed for reduction: traction may be continued for 5 to 10 min. There is often an audible ‘clunk' on reduction. Alternatively, the hand may be dorsiflexed while maintaining longitudinal traction. The hand is then palmar flexed, reducing the capitate–lunate articulation. Preliminary radiographs in traction usually demonstrate satisfactory reduction, with a normal capitolunate relationship and scaphoid position without rotatory subluxation.

Radiographs are assessed weekly for the first 4 to 6 weeks. The capitolunate relationship, normally 0 to 15°, is monitored. Radiographs of the opposite wrist may be used for comparison.

Transcaphoid perilunate dislocations

Transcaphoid perilunate dislocation is uncommon. The only difference between this injury and dorsal perilunate dislocation is the location of the direction of force, which occurs through the scaphoid rather than between the scaphoid and the lunate.

Rotary subluxation of the scaphoid bone

The scaphoid is the key link between the proximal and distal carpal rows. The radial and proximal position of this carpal bone controls carpal stability of the wrist. Injury to this carpal bone predisposes the wrist and hand to functional disability. The chances of successful closed treatment depend on suspicion and early recognition of the injury. Patients frequently complain of wrist pain after an apparently trivial injury. On examination, the patient appears to have pain and tenderness in the radiocarpal area at the level of articulation between the radius and the proximal scaphoid. Watsons' test is useful. In this test, the patient's elbow rests on his lap. At all times the forearm is pronated. The examiner uses the opposite hand to take the patient's wrist into full ulnar deviation. The patient's hand is then radially deviated while the examiner pushes against the distal pole of the scaphoid with his thumb, attempting to prevent the scaphoid's normal palmar flexion. This maneuver forces the distal pole to sublux if the scapholunate joint is unstable. This test is positive if painful on the injured side and if it reproduces SYmptoms.

Acute rotary subluxation of the scaphoid should be treated by open repair. A dorsal approach is used to repair the injured scapholunate ligament, with the repair stabilized by the use of Kirschner wires from the scaphoid to the capitate and lunate. A thumb spica cast is placed, with the wrist in slight flexion. Pins are removed at 6 weeks; immobilization is continued for 6 more weeks.

Ligament injuries of the thumb

The unique versatility of the thumb makes any injury to this digit extremely disabling. Injuries to the ligament complexes of the three thumb joints may cause instability during abduction and adduction stresses, diminishing pinch.

The metacarpophalangeal joint of the thumb has strong radial and ulnar collateral ligaments. Stability is enhanced by the volar plate, which inserts on the proximal phalanx. Injuries at the metacarpophalangeal joint of the thumb are common in athletes, particularly skiers and football players, and are caused when the joint is forcibly pulled into abduction or adduction.

Patients with acute injuries present with local tenderness and swelling along the injured side of the metacarpophalangeal joint. When seen 6 or more weeks after injury, patients may complain of weakness of thumb-index pinch, instability of the metaphalangeal joint, or pain on abduction stress.

Flexor tendon injuries of the hand

Function after a flexor tendon injury depends on many factors including the type of injury, the initial surgical treatment, the healing potential of the tendon ends, the experience of the surgeon, knowledge of finger and hand anatomy, and the type of rehabilitation.

Anatomy

The flexor tendons originate in the forearm at the musculotendinous junction. Four flexor digitorum superficialis tendons lie anterior to the four flexor digitorum profundus tendons. The superficialis muscle tendon units have separate, independently acting muscles, while the flexor profundus muscle belly pulls all of the flexor profundus tendons *en masse*.

The fibrous retinacular sheath, beginning at the area of the metacarpal neck and ending at the distal phalanx, comprises both annular and cruciate pulleys. The pulley system is important in flexor tendon function. Bowstringing is prevented, because the pulleys hold the flexor tendons in close apposition to the phalanges during digital flexion.

Diagnosis

Although at times difficult to perform in an uncooperative patient, careful digital examination will diagnose all flexor tendon injuries. Assessment of both flexor tendon systems depends upon the unique anatomy of the profundus and superficialis SYstems. Testing of the flexor digitorum profundus tendon recreates the function of the individual tendon unit. Each distal interphalangeal joint in the index, ring, long, and small fingers is tested to determine whether individual flexion is intact. Injuries to the profundus tendon system are present if the distal interphalangeal joint does not flex. Individual superficialis tendons, which normally actively flex the proximal interphalangeal joints, are tested by placing the other three fingers in a position with each distal and proximal interphalangeal joint extended while asking the patient to flex the finger being examined. The pull of the profundus musculotendinous unit is negated by the passive extension of the non-tested fingers. The flexor pollicis longus muscle is tested by asking the patient to flex actively the interphalangeal joint of the thumb.

Partial flexor tendon lacerations are difficult to diagnose. Partial tendon injury should be suspected in all patients with pain against resistance during an examination. If the wound is not explored, splinting should be carried out for 2 to 3 weeks to prevent rupture of the damaged tendon. If there is any doubt about the amount of tendon lacerated, exploration is recommended.

Flexor tendon injury management

Operative repair is indicated for flexor tendon lacerations. All such surgery should be accomplished with a sterile and bloodless field, a rested surgeon, regional or general anesthesia, and optical magnification. Surgical incisions should be based on principles of an extensile approach without crossing skin creases at a right angle. The volar oblique transaxial zigzag incision using interaxial digital lines is preferred. The best surgical results are obtained after early direct repair of complete tendon injuries. The definition of ‘early' is controversial, but is generally less than 2 to 3 weeks after injury.

Treatment of partial tendon injuries remains controversial. The only strong evidence available suggests that debridement and early motion are appropriate for partial injuries involving less than 60 per cent of a tendon.

Extensor tendon injuries

The treatment of extensor tendon injuries demands the same knowledge of anatomy, surgical technique, and postoperative rehabilitation as does treatment of flexor tendon injuries. Unfortunately, it is usually the least experienced surgeon who is called on to repair extensor tendon injuries in the emergency room, with expectation of excellent results when two tendon ends are sutured together. These injuries are not easy to treat, and poor postoperative results may cause worse impairment of function than do flexor tendon injuries. Extensor tendon injuries are common because of their superficial location on the dorsum of the hand. Injuries may be caused by lacerations, skin loss, or abrasion, and can be associated with lacerations at multiple joint levels. Bacterial contamination of joints may increase the complexity of the treatment plan; this is common when these injuries are caused by human bites.

Anatomy

The forearm extensor mechanism arises from the lateral portion of the distal humerus to form the multiple extensor muscle–tendon units. Independent extension is provided through the extensor pollicis longus, the extensor pollicis brevis, the extensor indicis proprius, and the extensor digiti quinti. The extensor digitorum communis tendons have limited independent action and have four distinct tendons. The extensor digitorum communis tendons enter the hand through fibro-osseous tunnels at the level of the wrist. Tendons at this level are covered with a synovial sheath. The extensor retinaculum is a fibrous band which prevents bowstringing of extensor tendons across the radiocarpal joint. It consists of two layers—the supratendinous and the infratendinous. Six dorsal compartments are separated by septas that arise from the supratendinous retinaculum and insert onto the radius. The communis tendons are joined by tendinous interconnections proximal to the metaphalangeal joint. Lacerations of an individual finger communis tendon proximal to the juncture may result in only incomplete extension loss of that finger, as extension may be provided through an adjacent extensor tendon.

Extensor tendons at the metacarpophalangeal joint level are held in place over the dorsum of the joint by the conjoined tendons of the intrinsic muscles and the sagittal bands. An injury to the extensor hood at this level may result in subluxation or dislocation of the central extensor tendon.

The six separate dorsal compartments are normally numbered from the radial to ulnar side of the wrist. The first dorsal compartment contains the abductor pollicis longus and the extensor pollicis brevis. Occasionally the extensor pollicis brevis has a partially separate sheath. The second dorsal compartment contains the tendons of the extensor carpi radialis longus and brevis. The third dorsal compartment contains the extensor pollicis longus, which runs around Lister's tubercle, angling towards the thumb. The extensor indicis proprius and the extensor digitorum communis are contained within the fourth dorsal compartment. The fifth dorsal compartment contains the extensor digiti quinti, which occasionally has a double tendon. The sixth dorsal compartment contains the extensor carpi ulnaris.

In the thumb, the distal phalanx is the distal attachment of the oblique interosseous fibers and the central slip of the extensor pollicis longus. The extensor pollicis brevis attaches to the metacarpophalangeal joint capsule and the abductor pollicis longus attaches to the base of the thumb metacarpal.

Injury management

Zone I injuries occur at the level of the distal insertion of the lateral bands at the distal phalanx. These cause loss of distal interphalangeal joint extension and are commonly due to a laceration at the level of the distal interphalangeal joint. They may also be due to direct trauma to an extended fingertip. Such injuries may be associated with a closed tendon injury or a distal phalanx fracture. All open lacerations should be treated by irrigation of the joint. The extensor tendon should be sutured and the joint and distal phalanx should be splinted in extension full time for at least 6 weeks. Active motion is then encouraged. Night splinting may be required for an additional 6 weeks.

Closed injuries may be treated by extension splinting for 6 to 8 weeks with additional night splinting for up to 3 months. Open treatment of closed injuries is reserved for those injuries with large fracture fragments or subluxation of the distal interphalangeal joint.

Zone IIa injuries are extremely difficult to treat. They involve the region along the insertion of the central tendon into the middle phalanx, and may be open or closed. Open injuries overlying a joint need to be aggressively irrigated and debrided. Primary tendon repair and immobilization, with the proximal interphalangeal joint splinted or pinned in extension for 4 to 6 weeks, is required. Active mobilization of the distal interphalangeal and metacarpophalangeal joints is encouraged. Treatment of closed injuries may be difficult: there is a delay in diagnosis in most cases, with subsequent development of boutonnière deformities. Treatment should be directed towards immobilizing the proximal interphalangeal joint by extension splinting or pinning for 4 to 6 weeks. Early motion of the distal interphalangeal and metacarpophalangeal joints is encouraged at 2 to 4 weeks to prevent extensor tendon adherence.

Zone IIb injuries occur over the proximal phalanx proper. Tendon lacerations in this area do not usually retract after being injured because of the joint tethers. Direct

repair followed by immobilization with the metacarpophalangeal joint in flexion and the proximal interphalangeal joint in extension is the treatment for most acute lacerations. Splinting should be carried out for 4 to 6 weeks with careful mobilization.

Zone IIc injuries occur at the level of the metacarpophalangeal joint. Human bites cause injury to the central extensor tendon and the metacarpophalangeal joint: if indicated, the joint should be irrigated. If exploration reveals early inflammatory changes or infection, this should be treated first; the tendon injury is then repaired later. Most tendon injuries at this level can be treated by early controlled mobilization.

Zone III extensor tendon injuries should be treated by direct primary repair. In some cases, especially those with independent tendon function, the proximal tendon may retract up through the fibro-osseous sheath. The tendon then needs to be found and sutured primarily. Postoperatively, treatment is by early controlled mobilization or immobilization of wrist in extension, with the metacarpophalangeal joints flexed 20°.

Zone IV is a zone of fibro-osseous pulleys with multiple tendons in close proximity. Repair of all injured tendons in this area is recommended. Only compartments that hold injured tendons should be opened.

Lacerations in the first, second, and third dorsal compartments should be repaired primarily. A small portion of the dorsal retinaculum should be excised at the tendon repair site to prevent the lacerated retinaculum interfering with tendon gliding.

Injuries to extensor tendons in zone V should be treated by primary repair of all injured structures. Most treated injuries have an excellent outcome. Postoperatively, zone V tendon injuries should be protected for 4 weeks with the wrist held in extension. Controlled assisted mobilization is then begun.

Dupuytren's contracture

In Dupuytren's contracture, a form of nodular fibromatosis develops in the palmar fascia of the hand causing progressive fixed-flexion contractures of the thumb and fingers. Some authors have suggested that although little progress has been made in the basic understanding of Dupuytren's contracture in the past 100 years, results of treatment have improved.

The disease appears to be an active cellular process that results in the production of oriented collagen, which is subjected to intrinsic forces and external stresses. The cellular activity takes place in nodules, some of which are large enough to be palpated, some of which are microscopic, scattered throughout the diseased tissue. The removal of a nodule or incision of a contracted cord does not prevent progression of the disease because the cellular activity is widespread.

Anatomy

Dupuytren's contracture always involves the palmar fascia. The ring and small fingers are more frequently involved than are the index and long fingers, although the index finger, the long finger, and the thumb index web may be involved in some patients.

Pathobiology

Gabbiani first showed that palmar fascia nodules were composed of myofibroblasts. He noted that these cells displayed structures which allowed contractile abilities. Proliferation of these cells is slowly progressive. Much recent investigative research has been directed toward discovering the cellular origin of the myofibroblast. The two most widely held theories are that the myofibroblast originates subdermally (extrinsic theory) or is actually a differentiated fascial fibroblast (intrinsic theory).

Incidence

Dupuytren's disease is more common in Northern European countries and in patients originating from Northern Europe. The disease incidence increases with the age of the population studied. Dupuytren's contracture has been reported in association with epilepsy, diabetes mellitus, and alcoholism.

Treatment

The treatment of Dupuytren's contracture is surgical. Indications for operative treatment include flexion contractures of the metacarpophalangeal joint at 30° or more, or any degree of contracture of the proximal interphalangeal joint. Surgical treatment of flexion deformities of the proximal interphalangeal joint does not always provide full correction: this is more likely when the contracture is small. Patients should be advised that palmar nodules alone and/or pain is not an indication for surgery. Hand function may improve with surgery, and surgery may slow the progression of the disease; it does not cure Dupuytren's disease. Closed fasciotomies have high rates of complications. Total fasciectomy has high complication rates. Complications include large hematomas causing infection, skin loss, edema, and delay in healing.

Each patient should be carefully educated in the role of postoperative hand therapy and splinting. Orthoplast extension splints are of great help, placing the postoperative hand and finger in a position of maximum extension. Early range of motion exercises are begun, and post-surgical edema is treated aggressively.

Hand infections

General principles

Hand infections continue to create serious problems for patients and surgeons.

Staphylococcus aureus and streptococci are the most common causes of hand infections. Other organisms such as anaerobic streptococci and clostridia may cause severe infections in the upper extremity. Gram-negative rods such as *Escherichia coli, Proteus* spp., and *Pseudomonas* spp. may cause infections in a trauma victim or immunologically compromised host. Viral and fungal infections, although uncommon, may have serious sequelae if not considered in specific patients.

Several factors are important in determining whether an individual hand infection can be managed without operative intervention. The time from onset of pain and discomfort to the patient seeking treatment is important. Most patients treated within 24 to 48 h of initial onset of pain and swelling improve when treated with high doses of intravenous antibiotics. Immunocompromised patients, including those with diabetes, those receiving chemotherapy for cancer or connective tissue disease, and those receiving oral steroids, are less likely to improve with non-operative therapy alone. The choice of intravenous antibiotics should be based on the organism likely to be responsible for the infection. The hand should be immobilized in a splint and elevated.

Cellulitis

In the early phase, infections of the hand present themselves as cellulitis. Patients present with a hand or finger showing diffuse swelling and erythema, with or without accompanying lymphangitis. Hemolytic streptococci most commonly cause such cellulitis. Radiographs should be obtained to rule out any underlying bony infection. These infections usually respond to intravenous antibiotics.

Paronychia

Paronychia is the most common of all hand infections, involving the radial and ulnar sides of the nail and surrounding tissue. If the infection also involves the eponychium as well as the lateral fold, it is called an eponychia. Extension to the opposite side of the fingernail, which is uncommon, is called a run-around abscess. The early stages of infection can be treated by soaks and warm saline solution with systemic oral antibiotics for 5 to 7 days. Pus is usually visible in more extensive or deeper paronychial infections. When such pus is visible, drainage is preferred. In these cases, the paronychia is compressed along the nail edge, trapping the abscess. All procedures that successfully treat paronychia separate it from the hard nail. If the infection is limited to less than one-half of the eponychium, a single incision placed to drain the paronychium and to elevate the eponychial fold for excision of the proximal one-third of the nail is satisfactory. If the entire eponychium is involved, two incisions are required. The entire eponychial fold is elevated with excision of at least the proximal one-third of the nail. Xeroform gauze can be placed to prevent premature closure of this area.

All patients require oral antibiotics. The gauze is removed after 72 h and gauze dressings are applied for 24 to 48 h. Hand washings are then performed, and the patient is instructed to dress the wound with sticking plaster and perform early range of motion exercises.

Felon

A felon is a potentially permanently disabling infection of the pulp tissue of the tip of a digit. It involves the multiple vertical fibrous septas that divide the fingertip pulp into several small compartments, and often follows a puncture wound, although many patients do not remember a penetrating injury. The major symptom in these patients is rapidly developing pain, usually of 12 to 24 h duration. Because of the lack of expansion through the skin, this abscess may extend proximally into the flexor tendon sheath, or into the distal phalanx, causing osteomyelitis.

All felons should be drained before any pus is seen. Many techniques are available. Once drained, a Xeroform or appropriate gauze should be packed in the wound after cultures have been obtained, and a bulky dressing should be applied. The drain is removed at 72 h. Oral antibiotics are given for 5 to 7 days.

Dressing changes with wound repacking, plus irrigation to keep the wound open in the early post-surgery period, helps prevent recurrence.

Infections of the web space

The web space of the hand is defined by webbed skin dorsally, vertical septas ulnarly and radially, and volarly by the transverse palmar fascia. A laceration in the area of the finger web may place infected material in this area, causing a ‘collar button' abscess. Patients present with web space swelling both palmarly and dorsally; this must be drained both palmarly and dorsally. Incisions should be placed longitudinally. A through-and-through drain can be placed, and the wounds packed open. Cultures are taken, and attention is directed toward keeping the wounds open and encouraging early range of motion to improve functional outcome. Antibiotics, chosen on the basis of culture results, are given for 24 to 48 h intravenously, and are continued orally for 5 days.

Pyogenic flexor tendon tenosynovitis

Infections in the flexor tendon sheath usually follow open wounds or crush injuries. Infection within the flexor tendon sheath causes severe inflammation of both the flexor tendon and sheath structures. Chronic changes limit flexor tendon motion and cause significant functional impairment. Early recognition with surgical drainage is the treatment of choice. Historically, Kanavel was the first to recognize the four signs of flexor tendon sheath infections: tenderness over the involved flexor tendon sheath, pain on passive extension of the involved digit, flexed attitude of the involved digit, and fusiform swelling of the involved digit. The ring, middle, and index fingers are most commonly affected. Coagulase-positive *Staphylococcus aureus* is the most common cause of infection. In the early stages, most of these infections are a flexor tendon and sheath cellulitis; intravenous antibiotics, elevation, and splinting may be curative. These patients should be admitted to hospital and placed on antibiotics. If there is no change in clinical signs, predominantly the pain on extension, the sheath must be opened and irrigated. Surgical drainage of the sheath can be carried out using open techniques, single incision techniques for antibiotic instillation, distal drainage with proximal instillation of antibiotics, through-and-through intermittent antibiotic irrigation, or closed tendon sheath irrigation. Early active range of motion exercises may improve the significant functional impairment that accompanies these infections.

Thenar space, midpalmar space, and hypothenar space infections

The three major potential deep compartments in the hand are the thenar space, the midpalmar space, and the hypothenar space. Bacterial contamination of these spaces follows direct lacerations or communications from expanding flexor tendon infections. Anatomically, the midpalmar space is bounded dorsally by the third, fourth, and fifth metacarpals, and the fascia of the volar interosseous muscles, and palmarly by the flexor tendons. Its radial border is the vertical septum between the sheath of the long finger flexor digitorum profundus and the third metacarpal. The ulnar border of this space is the fascia of the hypothenar muscles. The vertical septa of the palmar fascia are the distal margin of this space, with the proximal end bounded by the thin fascial layer at the distal end of the carpal tunnel. In these infections the palm may become flattened, although the dorsum of the hand swells because of the lymphatic drainage system. Swelling and tenderness are present over the space itself. In the later stages, erythema and fluctuation is noted, and any motion of the flexor tendons causes pain.

Treatment is incision and drainage. Multiple incisions have been recommended. An incision along the distal palmar flexion crease provides a bloodless field and avoids the contraction of a vertical scar. The abscess cavity is irrigated and left open, with a drain for 24 to 48 h. Early active assisted motion is recommended. Empirical antibiotics are started, and changed on the basis of specific culture results.

The thenar space is located radially to the midpalmar septum and extends radially to the lateral edge of the fascia of the adductor pollicis longus muscle. Infections in this space lead to swelling towards the palm in the thenar region; this pushes the thumb into abduction. Such infections arise from puncture wounds or from extension of an infection in the midpalmar space, in the radial bursa, or in flexor tendon sheaths. The clinical presentation is one of marked thenar swelling. Incision and drainage is performed through a combined dorsal and volar approach to the space. Antibiotics are given intravenously and the wound is kept open with a drain. Early active assisted motion is encouraged.

Infections of the radial and ulnar bursa

The radial and ulnar bursas represent fascial compartments that enclose the flexor tendons. The radial bursa is the proximal extension of the flexor pollicis longus tendon sheath extending through the carpal tunnel into the forearm. The ulnar bursa is the proximal extension of the flexor sheath of the flexor digitorum profundus of the small finger. Infection in these spaces follow direct inoculation, or are extensions of an infected tendon sheath. Swelling, erythema, and tenderness are noted along the anatomic boundaries. As in all closed space infections of the hand, treatment is by either open or closed irrigation and drainage. Drains must be placed through the skin in the proximal wounds. Antibiotics are continued until the infection has cleared.

Human and animal bites

Bites of the finger and hand can be caused by humans or animals. In human beings, most injuries are caused by clenched fist blows to another person's mouth. Cultures of human bite wounds usually grow *Staphylococcus aureus*, streptococci, and *Eikenella corodense*, an aerobic Gram-negative rod. There is usually no evidence of joint involvement; however, examination is usually carried out with the finger extended, a position that masks the usual extensor tendon and encapsular injury which occurs in flexion. Admission to hospital for operative irrigation debridement and exploration of the metacarpophalangeal is essential. Administration of penicillin and a second-generation cephalosporin is recommended until final cultures are available. *E. corodense* is sensitive to both penicillin and ampicillin.

Dog and cat bites or scratches can produce cellulitis, lymphangitis, and tendon or joint infections. The chief pathogens are streptococci and *Pasteurella multocida*, a small Gram-negative coccus. Most cellulitic infections respond well to treatment with intravenous penicillin. *P. multocida* is responsible for 50 per cent of infections associated with dog bites and 80 per cent of infected cat bites. If treatment with antibiotics alone is unsuccessful, wounds should be irrigated.

Herpes simplex infections

Herpes simplex virus causes herpetic whitlows, an infection frequently seen in medical and dental personnel. It is also sometimes seen in small children. Treatment is SYmptomatic. Topical acyclovir can be used in patients who complain of severe pain in the tips of their fingers. The diagnosis of herpetic whitlow can be made clinically by history and physical examination. Laboratory documentation of infection may be obtained by finding serum antibodies to viral antigens.

Amputations of the hand

Fingertip amputations

Fingertip amputations are the most commonly treated hand amputation. Usually there is skin or pulp tissue loss only. The size and position of the defect helps to determine appropriate treatment. Skin loss may be transverse or oblique, more radial-sided or ulnar-sided, or may involve a loss of more skin on either the volar or dorsal side. When no bone is exposed the decision must be made as to whether soft tissue cover needs to be undertaken or whether secondary healing should be allowed to occur: the fingertip has excellent self-healing and regenerative capacities, and many untreated defects become less visible as the fingertip continues to remodel. Some surgeons prefer to treat digital tip injuries with no exposed bone by conservative means with multiple dressing changes. Others resurface the distal pulp with split- or full-thickness skin grafts, or with proximal undamaged tissue.

If skin loss extends to the palmar surface of the flexor tendon and/or distal phalanx, cross-finger pedicle flaps may be useful. A dorsal flap of undamaged skin and subcutaneous tissue is elevated from the adjacent uninjured finger and attached to the defect in the injured finger. The volar surface can be repadded with excellent soft tissue cover. The donor site on the dorsal aspect of the finger is then covered with a free skin graft. Complications include stiffness of interphalangeal joints, but this can be controlled with assisted active range of motion exercises. A soft tissue defect on the palmar surface of a finger down to bone and flexor tendon can also be

covered using the thenar flap. In this operation the injured finger is flexed and attached to a thenar pedicle flap raised from the skin and subcutaneous tissues. Stiffness, although reported as a complication, can be minimized by early division of the thenar pedicle. The proximal end of the flap should be allowed to settle into its bed by its own means.

Larger palmar defects can be treated with a Moberg volar advancement flap, a method that is more appropriate for the thumb than for the fingers. Complications have been reported, including fixed flexion deformity of the distal joint and necrosis of the entire volar flap.

Thumb amputations

The length of the thumb is of critical importance in the functioning of the hand. Because of its functional independence, there is a greater need for the preservation in its length and restoration of its sensitivity than is the case for other digits. Treatment of soft tissue losses should be directed toward healing by secondary intention or free skin grafting. If bone is exposed, preservation of length is critical: use of the Moberg advancement flap is recommended. In this surgical procedure, the volar tissue of the thumb is elevated along the midaxial plane with preservation of the neurovascular bundles. Attention must be paid to dissecting this volar flap free from the flexor pollicis longus sheath. The distal extent of the thumb, including the interphalangeal and metacarpophalangeal joints, is then flexed to allow this volar flap to cover the distal-most tip that is exposed. After healing, actively assisted range of motion and stretching will minimize scar contractures. This type of advancement flap can cover up to 1 cm of lost palmar tissue.

Amputations with loss of more than 1 cm of thumb are challenges to the reconstructive hand surgeon. Although there has been recent interest in microvascular transfers of distant tissues, these procedures carry significant morbidity. A less sophisticated method may solve reconstructive problems, with less inherent risk. Matev has described a distraction/augmentation stage reconstruction of the thumb when the amputation is at the level of the metacarpophalangeal joint. After healing of the primary amputation, a transverse osteotomy at the level of the thumb metacarpal is made as the first stage of the procedure. This osteotomy is stabilized by transverse metallic pins which are attached to an external fixation device. After primary closure of the wounds, 1 to 2 mm of distal distraction per day is carried out. Careful attention must be paid to the vascular status of the tip of the thumb remnant during distraction. After the appropriate length of metacarpal has been obtained, a second operation is carried out. Using the same incisions, an autogenous iliac bone graft is fashioned and placed in the space available between the two thumb metacarpal fragments. This graft may be stabilized by Kirschner wire or internal fixation techniques. With length, the thumb function is significantly improved. Deepening of the first web space may be carried out to improve function of the hand.

Several options are available for the treatment of amputations at levels proximal to the metacarpophalangeal joint. Pollicization of the index finger provides excellent function. In this procedure, the normal index finger is transferred, using appropriate reconstructive principles, to the position of the amputated thumb. The procedure appears to work better in young patients. Other microsurgical techniques, including the wrap-around flap, have been developed to try to treat this special problem.

Middle phalanx amputations

Amputations of the middle phalanx are functional if the amputation is distal to the insertion of the flexor digitorum superficialis. In these amputations, length is not so important. The patient should be treated with procedures that allow primary coverage of the amputation.

Amputations through the proximal phalanx

Amputations through the proximal phalanx diminish the function of the finger remnant significantly. At this level there is no insertion of the flexor digitorum superficialis or profundus to provide power in grip. Ray resections, transposition, and/or reconstruction of the contiguous transverse metacarpal ligament should be performed. Ray amputation of an injured index finger leads to an excellent cosmetic result with good function. There may be functional disabilities, including weakness of pinch and grip after these operations. Ray amputation of the long finger may be treated in two ways. Reconstruction of the volar intermetacarpal ligament can be carried out after amputation of the long finger, suturing this anatomic structure between the metacarpals of the index and ring fingers in an attempt to close dead space and to stabilize both metacarpals. The second procedure is transposition of the index metacarpal to the position of the long finger metacarpal. In the case of ring finger resection, the fifth ray can be transposed to the ring position.

Metacarpal and carpal amputations

Carpal amputations or midpalmar amputations are difficult to treat. All attempts should be made to save the radiocarpal joint to allow wrist flexion/extension, which may be helpful in function. Patients with severe length loss may retain function by using residual carpus and metacarpals as a terminal device.

Disarticulations at the radiocarpal joint allow pronation and supination of the forearm and the hand unit.

Further reading

Abu-Jamra FN, Khuri S. The treatment of finger tip injuries. *Journal of Trauma* 1971; **11**: 749–57.

Adler JB, Shalton GW. Fractures of the capitate. *Journal of Bone and Joint Surgery* 1962; **44A**: 1537–43.

Amadio PC, Jaeger SH, Hunter JM. Nutritional aspects of tendon healing. In: Hunter JM, Schneider LM, Mackin EJ, Callahan AD, eds, *Rehabilitation of the hand*, 2nd edn, pp. 255–60. Mosby, St Louis, 1984.

Amido PC. Tendon injuries in the upper extremity. In: Dee R, ed., *Principles of orthopedic practice*, pp. 699–718. McGraw Hill, New York, 1989.

Atasoy E, Ioakimidis E, Kasdan ML, Kutz JE, Kleinert HE. Reconstruction of the amputated finger tip with a triangular volar flap. A new surgical procedure. *Journal of Bone and Joint Surgery* 1970; **52A**: 921–6.

Barrack RL, Mogabgab WJ, Edmunds JO, Skinner HB. Management of bite wounds. *Orthopedic Reviews* 1983; **12**: 83–8.

Beasley RW. Reconstruction of amputated finger tips. *Plastic and Reconstructive Surgery* 1969; **44**: 349–52.

Beasley RW. Local flaps for surgery of the hand. *Orthopedic Clinics of North America* 1970; **1**: 219–25.

Beasley RW. Principles of managing acute hand injuries. In: Littler JW, ed., *Reconstructive plastic surgery*, Vol. 6, pp. 3000–102. Saunders, Philadelphia, 1977.

Blue AI, Spira M, Hardy SB. Repair of extensor tendon injuries of the hand. *American Journal of Surgery* 1976; **132**: 128–32.

Bora FW, Didizian NH. The treatment of injuries to the carpometacarpal joint of the little finger. *Journal of Bone and Joint Surgery* 1974; **56A**: 1459–63.

Boyes JH, Start HH. Flexor tendon grafts in the fingers and thumb: a study of factors influencing results in 1000 cases. *Journal of Bone and Joint Surgery* 1971; **53A**: 1332–42.

Brailliar F, Horner RL. Sensory cross-finger pedicle graft. *Journal of Bone and Joint Surgery* 1969; **51A**: 1264–8.

Brody GS, Cloutier AM, Woodhouse FM. The finger tip injury—an assessment of management. *Plastic and Reconstructive Surgery* 1960; **26**: 80–90.

Bryan RS, Dobyns JH. Fractures of the carpal bones other than lunate and navicular. *Clinical Orthopaedics and Related Research* 1980; **149**: 107–11.

Callahan M. Dog bite wounds. *Journal of the American Medical Association* 1980; **244**: 2327–8.

Carroll RE. Transposition of the index finger to replace the middle finger. *Clinical Orthopaedics and Related Research* 1950; **15**: 27–34.

Carter PR. *Common hand injuries and infections*. Saunders, Philadelphia, 1983.

Chase RA. Functional levels of amputation in the hand. *Surgical Clinics of North America* 1960; **40**: 415–23.

Citron DM, Goldstein EJ. Role of anaerobic bacteria in bite wound infections. *Review of Infectious Diseases* 1985; **6**(suppl 1): 361–7.

Cooney WP, Dobyns JH, Linscheid RL. Non-union of the scaphoid: analysis of the results from bone grafting. *Journal of Hand Surgery* 1980; **5**: 343.

Doyle JR. Extensor tendons—acute injuries. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 2045–72. Churchill Livingstone, New York, 1988.

Dray GJ, Eaton RG. Dislocations and ligament injuries in the digits. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 777–811. Churchill Livingstone, New York, 1988.

Dunn AW. Fractures and dislocations of the carpus. *Surgical Clinics of North America* 1972; **52**: 1513.

Eaton RG, Littler JW. Joint injuries and their sequelae. *Clinics in Plastic Surgery* 1976; **3**: 85–98.

Eddeland A, Eiken O, Hellgren E, Ohlsson NM. Fractures of the scaphoid. *Scandinavian Journal of Plastic and Reconstructive Surgery* 1975; **9**: 234–9.

Ejeskar A. Flexor tendon repair of no-man's-land: results of primary repair with controlled mobilization. *Journal of Hand Surgery* 1984; **9**: 171–7.

Ferraro MC, Coppola A, Lippman K, Hurst LC. Closed functional bracing of metacarpal fractures. *Orthopedic Reviews* 1983; **12**: 49–56.

Fraker WH, Wray R, Weeks PM. Factors influencing final range of motion in the fingers after fractures of the hand. *Plastic and Reconstructive Surgery* 1979; **63**: 82–7.

Freiberg A, Maktelow R. The Kutler repair of finger tip amputations. *Plastic and Reconstructive Surgery* 1972; **50**: 371–5.

Gelberman H, Panagris JS, Taleisnik J, Baumgaertner M. The arterial anatomy of the human carpus. Part 1: the extraosseous vascularity. *Journal of Hand Surgery* 1983; **8**: 367.

Graham WC, Brown JB, Cannon B, Riordan DC. Transposition of fingers in severe injuries of the hand. *Journal of Bone and Joint Surgery* 1947; **29**: 998–1004.

Green DP. *Operative hand surgery*. Churchill Livingstone, New York, 1982.

Gunther SF. The carpometacarpal joints. *Orthopedic Clinics of North America* 1984; **15**: 259–77.

Harkins PD, Rafferty JE. Digital transposition in the injured hand. *Journal of Bone and Joint Surgery* 1972; **54A**: 1064–7.

Hoskins HD. The versatile cross finger flap. A report on 26 cases. *Journal of Bone and Joint Surgery* 1960; **42A**: 261–77.

Hurst LL, Badalamante MA. Dupuytren's contracture. In: Dee R, ed., *Principles of orthopedic practice*, pp. 775–80. McGraw Hill, New York, 1989.

Hurst LL, Nathan J. Infections in the upper extremity. In: Dee R, ed., *Principles of orthopedic practice*, pp. 741–51. McGraw Hill, New York, 1989.

Idler RS. Anatomy and biomechanics of the digital flexor tendons. *Hand Clinics* 1985; **1**: 3–11.

Jahss SA. Fractures of the proximal phalanges: alignment and immobilization. *Journal of Bone and Joint Surgery* 1936; **18**: 726–31.

James JP, Wright TA. Fractures of the metacarpals and proximal and middle phalanges of the finger. *Journal of Bone and Joint Surgery* 1966; **48B**: 181–2.

Keim HA, Grantham SA. Volar flap advancement for thumb and finger tip injuries. *Clinical Orthopaedics and Related Research* 1969; **66**: 109–12.

Keyser JJ, Littler JW, Eaton RG. Surgical treatment of infections and lesions of the perionychium. *Hand Clinics* 1990; **6**: 137–57.

Kleineret JM, Zenni EJ. Non-union of the scaphoid—review of literature and current treatment. *Orthopedic Reviews* 1984; **13**: 125–41.

Kleinman WB, Grantham SA. Multiple volar carpo-metacarpal joint dislocation. *Journal of Hand Surgery* 1978; **3**: 377–82.

Leddy JP. Flexor tendons—acute injuries. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 1935–68. Churchill Livingstone, New York, 1988.

Leppard JJ, Mirza MA. Traumatic amputations of the hand and wrist. In: Dee R, ed., *Principles of orthopedic practice*, pp. 728–34. McGraw Hill, New York, 1989.

Lie KK, Margargle RK, Posch JL. Free full thickness skin grafts from the palm to cover defects of the fingers. *Journal of Bone and Joint Surgery* 1970; **52A**: 559–61.

Lister G. Pitfalls and complications of flexor tendon surgery. *Hand Clinics* 1985; **1**: 133–46.

Lister GD. Incision and closure of the flexor sheath during primary tendon repair. *Hand* 1983; **15**: 123–35.

Louis DS. Amputations. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 61–119. Churchill Livingstone, New York, 1988.

Manske PR, Gelberman RH, Lesker PA. Flexor tendon healing. *Hand Clinics* 1985; **1**: 25–34.

Mayfield JK. Wrist ligamentous anatomy and pathogenesis of carpal instability. *Orthopedic Clinics of North America* 1984; **14**: 209–16.

McCue FC, Baugher WH, Dulund DN, Gieck JH. Hand and wrist injuries in the athlete. *American Journal of Sports Medicine* 1979; **7**: 275–86.

McElfresh EC, Dobyns JH. Intra-articular metacarpal head fractures. *Journal of Hand Surgery* 1983; **8**: 383–93.

McElfresh EC, Dobyns JH, O'Brien ET. Management of fracture-dislocation of the proximal interphalangeal joints by extension-blocking splinting. *Journal of Bone and Joint Surgery* 1972; **54A**: 1705.

McFarlane RM. Dupuytren's contractures. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 553–89. Churchill Livingstone, New York, 1988.

Micks JE, Wilson JN. Full thickness of sole-skin grafts for resurfacing the hand. *Journal of Bone and Joint Surgery* 1967; **49A**: 1128–34.

Neviaser RJ. Closed tendon sheath irrigation for pyogenic flexor tenosynovitis. *Journal of Hand Surgery* 1978; **3**: 462–6.

Nevaiser RJ. In: Green DP, ed., *Operative hand surgery*, p. 77. Churchill Livingstone, New York, 1982.

Newmeyer WL, Kilgore ES. Finger tip injuries: a simple, effective method of treatment. *Journal of Trauma* 1974; **14**: 58–64.

Nunley JA, Urbaniak JR. Treatment of extensor tendon injuries of the hand. *Orthopedic Surgeon (Update Series)* 1982; **2**: 2–8.

O'Brien ET. Acute fractures and dislocations of the carpus. *Orthopedic Clinics of North America* 1984; **15**: 237–58.

O'Brien ET. Fractures of the metacarpals and phalanges. In: Green DP, ed., *Operative hand surgery*, 2nd edn, pp. 709–75. Churchill Livingstone, New York, 1988.

Osterman AL, Bora FW. Injuries of the wrist. In: Heppenstall RB, ed., *Fracture treatment and healing*, p. 504. Saunders, Philadelphia, 1980.

Rand JA, Linscheid RL, Dobyns JH. Capitate fractures. *Clinical Orthopaedics and Related Research*, 1982; **165**: 209–16.

Russe O. Fracture of the carpal navicular. *Journal of Bone and Joint Surgery* 1960; **42A**: 759.

Sampson S, Akelman E. Fractures and dislocations of the hand, wrist and forearm. In: Dee R, ed., *Principles of orthopedic practice*, pp. 536–53. McGraw Hill, New York, 1989.

Schlenker JD, Lister GD, Kleinert HE. Three complications of untreated partial laceration of flexor tendon—entrapment, rupture and triggering. *Journal of Hand Surgery* 1984; **6**: 392–6.

Schneider LH, Hunter JM, Norris TR, Nadeau PO. Delayed flexor tendon repair in no man's land. *Journal of Hand Surgery* 1977; **2**: 452–5.

Sehayik RJ, Bassett FH. Herpes simplex infections involved in the hand. *Clinical Orthopaedics and Related Research* 1982; **166**: 138–40.

Spinner M, Choi BY. Anterior dislocation of the proximal interphalangeal joint: a cause of rupture of the central slip of the extensor mechanism. *Journal of Bone and Joint Surgery* 1970; **52A**: 1329–36.

Strickland JW. Opinions and preferences in flexor tendon surgery. *Hand Clinics* 1985; **1**: 187–91.

Strickland JW *et al.* Phalangeal fractures: factor influencing digital performance. *Orthopedic Reviews* 1982; **11**: 39–50.

Webbe MA, Schneider LH. Mallet fractures. *Journal of Bone and Joint Surgery* 1984; **66A**: 658–69.

Wilson JN, Rowland SA. Fracture dislocations of the proximal interphalangeal joint of the finger. *Journal of Bone and Joint Surgery* 1986; **48A**: 293.

Woods GL, Burton RI. Avoiding pitfalls in the diagnosis of the acutely injured proximal interphalangeal joint. *Clinics in Plastic Surgery* 1981; **8**: 95–105.

46.2.4 The arthritic hip

James R. Roberson

Anatomy
Evaluation
Nonsurgical management
Surgical management
Alternatives to total hip replacement
Total hip replacement
Further reading

The hip is a ball-and-socket joint mating the lower extremity and the pelvis. Activities of daily living cause heavy loads across the joint, and hip arthritis has been a common cause of disability since the days of early man. Preserved skeletons of prehistoric man frequently have arthritic hip joints. Effective treatment of this painful disorder is a relatively recent accomplishment.

Anatomy

The normal anatomy of the pelvis and proximal thigh is complex, and managing disorders of the hip joint requires a thorough understanding of three-dimensional anatomic relations. The acetabulum located at the confluence of the three portions of the innominate bone, faces 45° lateral and 20° anterior. The ilium makes up the superior 40 per cent of the socket and comprises the weight-bearing dome. The ischium forms the posterior 40 per cent and the pubis the anterior 20 per cent of the socket. With the exception of the acetabular fossa, an oval depression in the inferior medial acetabulum, the entire inner surface of the acetabulum is lined with articular cartilage. The transverse ligament defines the inferior border of the fossa, and the acetabular branch of the obturator artery pierces the transverse ligament at its posterior attachment. The ligamentum capitis is a SYnovial-covered ligament originating from the margin of the fossa and the transverse ligament. The ligamentum capitis attaches to a pit-like depression on the medial surface of the femoral head and contains vessels from the acetabular branch of the obturator artery. In children, these vessels provide circulation to the subchondral bone of the femoral head around the attachment of the ligament. There is, however, no significant blood flow through the ligament in adults. The ligament provides some joint stability as it becomes taut during adduction of the hip. The floor of the acetabular fossa defines the medial wall of the acetabulum. The labrum of the acetabulum is located between articular cartilage and joint capsule, and forms a ring attaching to the entire rim of the acetabulum. A thick, fibrous capsule encloses the joint.

The acetabulum contains the spherical femoral head. Articular cartilage covers approximately two-thirds of the surface of the femoral head. The femoral neck connects the femoral head and the shaft of the femur. The neck produces an offset of the femoral head from the shaft, resulting in a mechanical advantage for muscles controlling the hip joint. The greater and lesser trochanters provide points of attachment for these muscles to the proximal femur. The depth of the acetabulum, the strength of the surrounding muscles, the acetabular labrum, and the ligamentum capitis contribute to joint stability.

The amount of subcutaneous fat about the hips is variable and often does not correlate with total body fat. Specifically, obese men have a tendency toward increased abdominal girth with less subcutaneous fat around the hip and proximal thigh. Obese women, in contrast, have more subcutaneous fat over the hips and thighs than the abdomen. This ‘pear-shaped' body habitus makes surgical exposure of the hip joint more difficult and requires longer skin incisions and more retraction.

The tensor fascia lata is the fascial doorway to the deeper anatomic structure of the hip joint. This thick connective-tissue layer provides insertion for the tensor fascia muscle and the proximal two-thirds of the gluteus maximus muscle. The interval between the tensor fascia and gluteus maximus insertions is located over the anterior margin of the greater trochanter. Some fibers of the gluteus medius muscle also originate from the deep surface of the tensor fascia, although most of it originates from the iliac crest.

The gluteus medius is the primary hip abductor and represents the second muscular layer encountered during posterolateral exposure of the hip joint. The strongest insertion of the gluteus medius to the greater trochanter is through its tendinous posterior portion. Although the primary function of the gluteus medius is abduction, its anterior fibers have a more horizontal orientation and serve to internally rotate the femur. The gluteus minimis muscle has a broad origin from the iliac crest distal to that of the gluteus medius and inserts on the anterior greater trochanter deep to the medius. Its fibers are closely intertwined with the superficial surface of the hip-joint capsule, from which it derives part of its origin. The piriformis tendon overlaps the posterior margin of the gluteus minimis muscle. The superior gluteal nerve, artery, and vein exit the pelvis via the greater sciatic foramen and run in an anterior direction about 5 cm above the tip of the greater trochanter. The gluteus medius, gluteus minimis, and tensor fascia muscle are supplied by the superior gluteal nerve.

The distal one-third of the gluteus maximus muscle inserts through a thick tendon to the posterior lateral portion of the femur distal to the lesser trochanter. This insertion is approximately 4 cm long and is easily visualized after the tensor fascia has been incised. The gluteus maximus nerve supply is from the inferior gluteal nerve, which exits the pelvis through the distal part of the greater sciatic foramen below the piriformis muscle.

The short external rotators of the hip are a group of six muscles inserting along the posterior border of the greater trochanter and upper femur. Their fibers blend with the posterior capsule of the hip joint. The piriformis muscle originates from the anterior surface of the lower sacral vertebrae and exits the pelvis through the greater sciatic foramen. Its long, distinct tendon passes over the posterior hip-joint capsule, under the corner of the greater trochanter, and inserts in a depression at the junction of the posterior femoral neck with the greater trochanter. The tendon of the piriformis is sandwiched between the posterior margins of the gluteus medius and minimis muscles. The sciatic nerve exits the pelvis below the piriformis muscle, although the nerve may also have an anomalous exit over or through the piriformis. The piriformis muscle primarily functions to externally rotate the femur but also assists in abduction of the hip. The more caudal, short external rotators are the obturator internus, gemellus superior, gemellus inferior, quadratus femoris, and obturator externus. These muscles insert on the posterior margin of the greater trochanter and have indistinct, often conjoined tendons. The medial circumflex artery runs deep to the quadratus femoris muscle and, with the lateral femoral circumflex artery, forms an extracapsular vascular ring. Smaller branches pierce the posterior capsule, progress along the periosteal surface of the femoral neck, and enter the femoral head through tiny foramina at the osteochondral junction.

The iliacus muscle originates from the medial wing of the ilium. The psoas major muscle arises from the bodies and intervertebral disks of the last thoracic and all five lumbar vertebrae. As these muscles exit the pelvis over the midportion of the superior pubic ramus; their fibers coalesce into a single tendon. This conjoined tendon curves around the medial aspect of the hip-joint capsule to insert on the posterior aspect of the upper femur at the lesser trochanter. Contraction of the iliopsoas flexes and externally rotates the hip joint.

Evaluation

Hip pain is a common complaint from patients seeking medical help. Since extra-articular problems often elicit the complaint of hip pain, the early evaluation should determine if the ‘hip pain' is from the hip joint. In most cases, the initial history and physical examination reliably differentiate intra- from extra-articular pathology. Pain from problems within the hip joint most commonly results in groin pain that may radiate to the anterior and medial thigh, and to the knee. Occasionally, patients with hip arthritis complain only of knee pain without groin or thigh SYmptoms. Hip pain is rarely perceived in the buttock or lateral thigh, and never in the sacroiliac area or below the knee. Extra-articular sources are usually the cause of pain localized to these areas.

Common intra-articular causes of hip pain include osteoarthritis, osteonecrosis, and inflammatory arthritis. Osteoarthritis is the most common cause of hip degeneration and is frequently a secondary process. Developmental dysplasia and a slipped capital femoral epiphysis cause an abnormal distribution of forces, resulting in focal overload of articular cartilage and secondary osteoarthritis (Fig. 1). Trauma, obesity, and previous sepsis also cause secondary osteoarthritis by either direct injury or overload of articular cartilage (Fig. 2). Genetic factors play a strong part in the development of primary osteoarthritis. Pain associated with osteoarthritis of the hip most commonly has a gradual onset and is usually associated with degenerative changes in other joints such as the lumbar spine, knee, and distal interphalangeal joints of the fingers.



Fig. 1. Osteoarthritis secondary to developmental dysplasia.

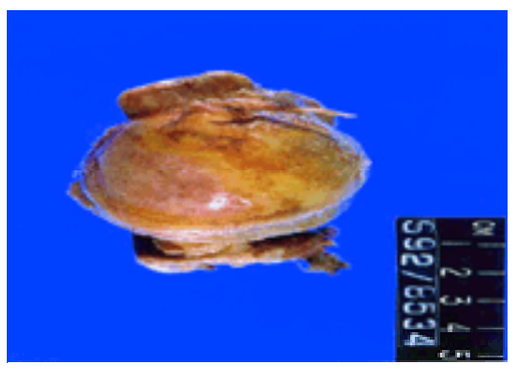


Fig. 2. Eburnated bone and complete loss of femoral head articular cartilage at the time of hip replacement for osteoarthritis.

Osteonecrosis, unlike other causes of hip arthritis, is initially a bone rather than cartilage problem. Bone infarction precedes the onset of hip pain by a variable but often long time, and articular cartilage is normal until late stages of the disease. The onset of pain occurs with fatigue fracture of the necrotic subchondral bone, and is often sudden and severe. Radiographic changes at this stage can be subtle as the articular cartilage space is well preserved, and the diagnosis may not be apparent (Fig. 3). A detailed history usually elucidates an underlying cause such as alcohol abuse, corticosteroid use, hemoglobinopathy, or Gaucher disease. The diagnosis of osteonecrosis should be considered when a patient with these risk factors has the acute onset of groin pain. Alcohol or corticosteroid use are the most common causes of osteonecrosis of the femoral head. There appears to be an individual susceptibility to alcohol and steroids as there is no clear threshold dose for either. SYstemic corticosteroids at any dose should only be prescribed with caution.



Fig. 3. Osteonecrosis of the femoral head.

Inflammatory arthritis affecting the hip produces an aggressive injury to articular cartilage associated with chronic synovitis and inflammation. More rapid joint destruction than with osteoarthritis or osteonecrosis usually results. Rheumatoid arthritis and hemophilic arthropathy are common examples.

Extra-articular causes of hip pain can be separated into four categories: neurogenic, vascular, soft-tissue inflammation, and bone. Sciatica from lumbar nerve-root impingement is the most common cause of neurogenic pain around the hip. Spinal stenosis and hip osteoarthritis frequently occur in the same patient; detailed evaluation is necessary to identify the cause of symptoms. Although the pain of both hip degeneration and spinal stenosis increases with activity, the location and character of the pain differs. The patient with spinal stenosis has back and buttock pain that may extend in the distribution of the sciatic nerve to the posterior and lateral thigh, calf, and foot. Paresthesias may also be present in the same distribution. As with hip osteoarthritis, pain associated with spinal stenosis worsens with standing or walking, but is relieved within minutes of sitting or lying down. Activity-related pain from the arthritic hip typically has a more rapid onset with the first steps of standing and walking, and is more gradually relieved by sitting and lying down. Unlike spinal stenosis, hip arthritis most commonly causes groin and anterior thigh rather than buttock pain. The passive and active range of hip motion produces pain with hip arthritis but not with spinal stenosis. Active straight-leg raising in the supine position generates very large reactive loads at the hip joint. The patient who can do this maneuver without pain rarely has significant symptoms from an arthritic hip.

Abductor tendinitis is one of the most common causes of pain in the hip area. The term trochanteric bursitis is incorrect for most patients since soft-tissue degeneration and inflammation occur within the tendons of the gluteus medius and minimis muscles rather than in a bursa overlying the greater trochanter. The disorder is associated with attrition of the abductor tendon in older people, or with overuse in the very active or the obese patient. Patients with hip arthritis will also frequently have abductor tendinitis. Physical examination easily differentiates abductor tendinitis from intra-articular hip degeneration. The patient with abductor tendinitis and a normal joint will usually have a normal range of hip motion, no limp, and no pain from a supine straight-leg raise. Palpation of the proximal lateral thigh over the greater trochanter always causes pain. Relief of pain following injection of the peritrochanteric soft tissues with a local anesthetic is both diagnostic and therapeutic. Nonsteroidal anti-inflammatory medications along with activity modification, hydrotherapy conditioning exercises, and weight loss are helpful for the long-term management of abductor tendinitis. The disorder rarely incapacitates and, in most cases, spontaneously improves. There is no role for surgical treatment.

Pain from fractures, including stress fractures, of the pelvic rami can be confused with intra-articular hip pain. These injuries are common in elderly patients with osteopenia, who may not remember a fall. Like hip pain, the pain from these injuries is localized to the groin and anterior medial thigh, and is worse with activity. Unlike hip arthritis, ramus fractures are often associated with ecchymosis and pain to palpation over the pubis. Nondisplaced fractures are difficult to see and require scrutiny of the radiograph. Neoplasm involving either the proximal femur or periacetabular bone can also present as hip pain.

A simple anteroposterior pelvis radiograph with a lateral view of the involved proximal femur are sufficient to evaluate most cases of hip pain. When taking the anteroposterior pelvis view, the X-ray beam should be centered low to include an adequate amount of the upper femur. Inlet, outlet, and oblique views of the pelvis are useful when there is a need to delineate clearly the acetabular lips and pelvic rami. Radiographs may be normal at the onset of osteonecrosis, but early marrow changes are clearly seen with magnetic resonance imaging (**MRI**) (Fig. 4). MRI is an expensive study and is rarely needed to diagnose osteonecrosis when infarcts of the femoral head are visible on radiographs. MRI also clearly illustrates other, difficult-to-see marrow processes such as transient osteoporosis of the proximal femur or infiltrative marrow-cell malignancies. Scintiscans are a useful means of identifying bone metastases and musculoskeletal infection. The bone scan is also used as a screening study in the patient with hip pain and no obvious diagnosis: the study is very sensitive but not specific (Fig. 5); therefore, occult pathology is rare with a negative scan result. Aspiration and culture of joint fluid is the best means of diagnosing joint sepsis. Although most investigators have reported a relatively high rate of false-positive results, false-negative results are uncommon.

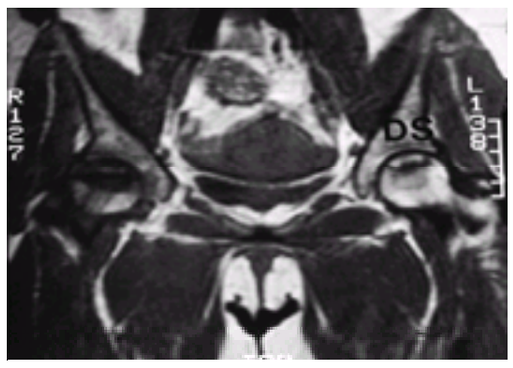


Fig. 4. Femoral head infarcts on MRI.



Fig. 5. Articular cartilage flap from femoral head at the time of hip replacement for osteonecrosis.

Nonsurgical management

The primary goal of treating hip arthritis without surgery is to decrease inflammation-associated joint pain. For many patients this can be accomplished effectively by decreasing joint loads or using anti-inflammatory medications. Cumulative hip-joint load is decreased by activity modification, weight loss, and use of a cane (walking stick).

Patients should be encouraged to play an active part in the management of their hip arthritis. Patient participation in conditioning is important to maintain ideal body weight and muscle tone for optimal joint function. However, conditioning exercises must be modified to avoid overloading worn joint surfaces and increasing pain. Frequency rather than intensity of exercise should be emphasized. A cross-training program of diverse exercises helps prevent both overuse injuries and boredom. Aerobic conditioning can be well maintained with a combination of water exercise, cycling, and light weight–high repetition, resistive exercises. Walking can be an important part of conditioning, but only at a distance or intensity that do not increase hip pain. Therefore, with severe joint degeneration, the role of walking for exercise is limited. The use of a cane in the hand opposite the arthritic hip can dramatically increase comfortable walking distance and the patient who walks for exercise should be encouraged to use one. Patients should avoid such impact-loading activities as running or jumping. Continuous standing also increases the pain of hip arthritis and simple modifications in the workplace, when possible, are helpful.

Nonsteroidal anti-inflammatory drugs are the primary means of medical management of arthritis. These medications work by inhibiting cyclo-oxygenase, the enzyme that converts arachidonic acid to prostaglandins, a mediator of inflammation. Although they are often effective at decreasing inflammation, nonsteroidals have significant possible side-effects. Especially in the older patient, their use should be closely monitored. The most frequent side-effect is gastrointestinal injury and bleeding caused by a decrease in prostaglandin-mediated protective coating of the gastric mucosa. The drugs can also decrease renal, hepatic, and platelet function. Nonsteroidal anti-inflammatory drugs should be discontinued in advance of elective operations, as the associated platelet dysfunction can cause increased surgical blood loss.

Medical management of inflammatory arthritis is much more complex. In addition to nonsteroidals, methotrexate, azathioprine, corticosteroids, and gold are commonly used. These are potent drugs with numerous deleterious effects, and should be prescribed only by a physician who is well versed in their use and can closely monitor the patient.

Surgical management

Alternatives to total hip replacement

The limitations of hip replacement relate primarily to component wear. Although, of the available surgical options, the results of total hip arthroplasty most closely approximate normal hip function, other operations are appropriate for patients at risk for long-term, wear-related complications. Alternative procedures should be considered, in an attempt to delay the need for total replacement, for the young, heavy, or highly active patient.

Osteotomy

The goal of osteotomy in the management of the arthritic hip is to increase the surface area available for the transmission of weight-bearing forces. The ideal patient for osteotomy is younger and has a correctable anatomic deformity involving the proximal femur or acetabulum. There should be no more than mild damage to articular cartilage and an acceptable range of hip motion. Osteotomy can be performed on the proximal femur or periacetabular area.

Intertrochanteric osteotomy was once the most commonly performed operation for the patient with an arthritic hip. Now the procedure is used less frequently and the indications are controversial, as good results are more predictable after total hip replacement. The procedure is most appropriate for the young, active patient with early loss of articular cartilage and an identifiable deformity of the proximal femur such as coxa vara. Nonunion of a femoral-neck fracture is another clear indication for proximal femoral osteotomy. Careful preoperative planning, close attention to detail, and rigid internal fixation of the osteotomy are critically important. Disadvantages of a varus-producing osteotomy include shortening of the resting fiber length of the abductors, and the creation of a limb-length discrepancy and limp. Although the failed osteotomy can be converted to a total hip replacement, the creation of significant deformity in the proximal femur compromises the result of subsequent arthroplasty and should be avoided.

Pelvic osteotomy is not associated with the limb shortening and abductor weakness that occurs with varus intertrochanteric osteotomy. Pelvic osteotomy also creates no proximal femoral deformity with later compromise of hip replacement. The triple innominate, Ganz, and Dial osteotomies all increase containment of the femoral head by redirecting the orientation of the acetabulum. The procedures are indicated for young adults with developmental dysplasia and minimal or no loss of articular cartilage.

Arthrodesis

The indications for hip arthrodesis became limited when the long-term results of total hip replacement improved. In general, the ideal patients for arthrodesis are younger, heavier, and more active. They should have unilateral hip disease, and no back or knee problems. A patient's need to continue in a vocation of manual labor is also a reason to consider arthrodesis.

Gait after a successful hip fusion can be relatively normal at slower walking speeds. Since, during walking, the lumbosacral spine and ipsilateral knee move in the same plane of motion as the hip, these joints can compensate for lack of hip motion after fusion. At faster walking speeds, however, the limp associated with a fused hip becomes more pronounced. Cramped spaces such as airplane seats and small cars are difficult to enter and exit. Although the fusion itself is quite durable, patients

with hip arthrodesis have long-term risk of low-back and ipsilateral knee degeneration.

Many different techniques have been described for hip arthrodesis. These are either intra-articular, extra-articular, or a combination of both. Most current techniques utilize some form of internal fixation. The technique selected should allow a predictably high rate of fusion without compromising the results of a later hip replacement if needed. The surgeon should accurately position the leg and avoid damage to the abductor muscles or distortion of the proximal femur. Although there is lack of consensus on the exact best position of fusion, it is approx. 40° of flexion, neutral abduction, and slight external rotation. Excessive hip abduction causes abnormally high loads in the medial compartment of the knee, while excessive adduction increases loads in the lateral compartment.

Callaghan and Sponseller both reported good long-term results at more than 20 years after hip fusion in approx. 80 per cent of patients. Best results were in patients younger than 18 years at the time of surgery. No patient had incapacitating back pain. Back pain was entirely absent in 16 per cent and only mild after strenuous activity in an additional 27 per cent. Despite these favorable reports, incapacitating back pain is the most common reason for conversion of the successful hip arthrodesis to total hip replacement.

Although the fused hip can be converted, this is a technically difficult procedure with a high rate of complications. Even after successful conversion to hip replacement, the majority of these patients have a restricted range of motion and residual abductor muscle weakness with limp. Conversion does effectively improve back pain in most patients.

Resection arthroplasty

The indications for resection arthroplasty, or removal of the hip joint, are limited. Treatment of the infected hip replacement is the most common indication, but the procedure may also be useful in the management of the paralytic patient with a painful hip dislocation. Loss of the fulcrum provided by the hip joint results in a large limb-length discrepancy, muscle weakness, and increased oxygen consumption during walking. Most patients can work short distances with a walker or crutches but need a wheelchair for longer distances. Although the time to maximum improvement is prolonged, the procedure usually affords reasonable pain relief. Most patients report improving strength and function for 1 to 2 years before reaching maximum improvement. Patient satisfaction with resection arthroplasty is poor when used for other than chronic, severe hip problems. In most cases, after eradication of infection, the prosthetic hip can be reimplanted.

Total hip replacement

Total hip replacement is the mainstay of surgical management of hip arthritis and dramatically improves quality of life for most patients. The ideal patient for total hip replacement is older, healthy, close to an ideal body weight, and has full-thickness loss of articular cartilage, with pain and limited hip function. Hip replacement affords this patient a high probability of normal hip function with a single operation for their remaining lifetime. Unfortunately, hip arthritis often occurs in other than the ideal patient, and patient selection and education are important for durable results. The young, heavy, active patient is likely to have an excellent early result but presents a difficult long-term challenge. When possible, these less-than-ideal candidates for hip replacement should be managed without surgery or with alternative surgical procedures. They should have their hips replaced only after nonsurgical treatment fails to provide reasonable relief of symptoms. They should also be carefully selected and thoroughly counseled about the limitations of the prosthetic hip and the possible need for repeat operations.

Preoperative planning

Once the decision to replace the hip has been made, careful preoperative planning is important to facilitate the operation and later rehabilitation. General health should be assessed, with particular attention to any history of deep venous thrombosis, or recurring dental or urinary infections. The physical examination should note any limb-length discrepancy or hip-joint contracture. Family and home support should be assessed to insure the availability of adequate assistance during the period of relative dependence that follows the stay in hospital. If inadequate support is available, plans for inpatient rehabilitation or nursing-home stay can be initiated before surgery. Thorough patient education about the expected results and possible complications of the procedure is important for later patient satisfaction. Most patients with significant hip arthritis are deconditioned and a preoperative exercise program without overloading the hip enhances early rehabilitation after surgery. Hydrotherapy exercises and upper-body resistive exercises with light weights are ideal for this purpose.

The optimal choice of prosthetic components is equivocal, as there are literally hundreds of them. The decision should be based on patient factors such as life expectancy and activity level, anatomic factors such as bone mass and the presence of deformity, surgeon factors such as experience and familiarity with components, and finally economic factors. The surgeon has an obligation to select components based on an expected best result. When expected results are equivalent, the surgeon also has an obligation to control costs. There is no consensus on the appropriate indications for the use of methylmethacrylate, porous ingrowth, hydroxyapatite, and ceramics at this time.

Radiographs and templates should be carefully scrutinized before the operation. The routine pelvic radiograph excludes the portion of the upper isthmus critical for accurate selection of the site for the prosthetic stem. The radiograph should be centered low to include the femur to the level of the distal tip of the prosthetic stem. The lateral radiograph is needed to assess the amount of posterior bow at the proximal femur, anterior bow at the diaphysis, and the diameter of the medullary canal. The oval cross-section of the medullary canal in most patients is oriented such that the measured diameter is greater on the lateral view of the femur than on the anteroposterior view. When limb-length discrepancy is present before surgery, templating the contralateral hip illustrates the correct level of femoral resection and prosthetic neck length required to reproduce equal limb lengths.

The primary purpose of templating the acetabulum is to establish the medial and proximal position of the center of rotation of the prosthetic hip, as cup size is more accurately determined during reaming. The acetabular template should be placed 2 to 3 mm from the medial wall of the acetabulum. The anatomic teardrop provides a reliable landmark for proximal/distal cup position, and the ileopectineal and ileoischial lines delineate the medial wall for medial/lateral position. Any protrusio or lateralization of the femoral head on preoperative radiographs should be noted. When the hip is laterally subluxed, osteophytic bone often fills the medial floor of the acetabulum. Thus, the distance between the medial wall of the acetabulum and the medial aspect of the femoral head is increased. Failure to medialize placement of the prosthetic socket to an appropriate position causes an increase in the length of the bodyweight moment arm and a shallow socket with compromised bone support for the cup.

After the planned center of hip rotation has been established with acetabular templates, femoral templates determine the appropriate level of resection of the femoral neck, and the size and offset of the femoral component. The size of stem selected differs for cemented and uncemented stems. The stem used without bone cement should fill the intramedullary canal while good cementation technique requires a smaller-diameter stem to allow for an adequately thick mantle of methacrylate. The femoral-component template is positioned initially so that the longitudinal axis of the stem should exactly match the axis of the medullary canal of the femur on the anteroposterior radiograph. In the presence of malunion of an old fracture, bowing of the femur on the anteroposterior view, or excessive bowing on the lateral view, it may be necessary to plan for an osteotomy of the femoral shaft. Vertical placement of the template and the level of the femoral-neck resection are established by placing the axis of the prosthetic femoral neck parallel to, and centered within, the anatomic femoral neck. The appropriate length of modular head places the center of the prosthetic head at the same level as the center of the acetabular component. A ruler adjusted for the radiographic magnification is used to measure the distance from the lesser trochanter to the prosthetic collar. This measurement is repeated on the femur during surgery to recreate the planned level of femoral-neck resection.

Offset is the distance from the long axis of the prosthetic stem to the center of the prosthetic head on the anteroposterior radiograph. Limb length and offset are closely related but not synonymous. For any given limb length, greater offset results in an increased abductor moment arm, lower joint reactive forces, and more tension on joint capsule and muscle across the hip. When possible, components that produce equal limb lengths without decreasing offset should be selected.

Surgical exposures

Total hip arthroplasty is performed for patients with abnormal hip anatomy. Optimum surgical management of diverse hip problems requires familiarity with different surgical exposures of the joint. There is no consensus among surgeons on the best surgical exposure for routine total hip replacement, and the posterolateral, anterior, anterolateral, direct lateral, and transtrochanteric exposures have all been advocated. Each has unique advantages and disadvantages.

The anterior approach, often referred to as the Smith–Peterson approach, is performed with the patient supine. The anterior joint capsule is exposed through the interval between the sartorius and tensor fascia muscles. Adequate acetabular exposure and access to the femoral canal for hip replacement are difficult with this approach and it is more valuable for procedures on the anterior column of the acetabulum. The Smith–Peterson approach is often used for periacetabular osteotomy and internal fixation of fractures of the anterior column of the acetabulum.

The anterolateral or Watson–Jones approach is also performed with the patient in the supine position. The surgical dissection is between the gluteus medius and tensor fascia muscles. The acetabulum is well visualized through this approach, but a direct line of access to the femur for reaming, instrumentation, and placement of components is difficult to achieve without traumatizing the anterior portion of the gluteus medius. If exposure of the femur is inadequate, the femoral stem will be

oriented with the tip of the prosthesis against the posterior and lateral cortex.

The Southern approach described by Moore provides easy access to the posterior hip joint and proximal femur by splitting the gluteus maximus muscle fibers. It provides less-extensile exposure and limited visualization of the acetabulum, and its use should be limited to hemiarthroplasty insertion.

The direct lateral approach, described by Hardinge, modifies the anterolateral approach by detaching the anterior tendon of the gluteus medius and minimis muscles as a continuous sleeve with the vastus lateralis. The exposure is done with the patient in the lateral decubitus position, the hip dislocation is anterior, and the leg is placed in a sterile bag in the figure of four position to allow femoral preparation. The anterior sleeve of abductor and vastus lateralis tendon is reattached during closure. The direct lateral approach provides excellent exposure of both the proximal femur and acetabulum. The incidence of dislocation is decreased because the posterior capsule and muscles are not detached. Because of the intramuscular abductor dissection there is an increased incidence of heterotopic ossification within the abductor muscle. Also, if the abductor tendon reattachment is not secure, an internal rotation lag and ‘toeing out' result. If the gluteus medius is split too proximally, injury to the superior gluteal nerve causes dysfunction of the anterior portion of the gluteus medius. When done correctly, the direct lateral exposure does not compromise abductor muscle function.

The transtrochanteric approach allows circumferential visualization of the hip joint, but is associated with increased operative time and blood loss. The approach is typically used only for difficult cases. If secure fixation of the trochanteric fragment against adequate bone of the lateral femur is not achieved, the rate of trochanteric nonunion is high. The trochanteric slide, a modification of the exposure described by Glassman, leaves the vastus lateralis and the abductor muscles attached to the respective ends of the trochanter. This muscle–bone–muscle sleeve is retracted anterior to the joint. The vastus lateralis attachment preserves blood flow to the trochanter and provides a tether against proximal migration in the event of nonunion.

The superior exposure and minimal surgical trauma to tissues provided by the posterolateral approach, described separately by Kocher and Langenbeck and later popularized by Gibson, make it an excellent choice for the majority of hip replacements.

Surgical technique

The procedure may be performed with a general anesthetic, subarachnoid block, continuous epidural, or a combined technique. Maintaining continuous epidural anesthesia during the days immediately after surgery decreases the need for narcotic analgesics and their associated nausea, drowsiness, and respiratory depression. There is evidence that continuous epidural anesthesia for total hip replacement decreases the incidence of deep venous thrombosis. Disadvantages of the epidural technique include the possibility of dural puncture and spinal headache, urinary retention requiring bladder catheterization, decreased control of lower-extremity muscles, and orthostatic hypotension during early postoperative mobilization. Unless coexisting medical problems contraindicate, maintaining slight hypotension during the operation decreases blood loss and surgical time.

When performing total hip replacement through the posterolateral approach, the patient should be firmly secured in the lateral decubitus position. Shifting the patient's position during the operation makes it more difficult to determine pelvic orientation and component anteversion when implanting the cup. After draping, the iliac crest should be accessible for the palpation of anatomic landmarks during placement of the acetabular component or for harvesting bone graft if needed. Sterile adhesive drapes effectively seal the perineal area from the field, and cover the lateral thigh.

The skin incision should be relatively straight and centered over the greater trochanter with the hip flexed about 45°. Sharp posterior angulation of the proximal incision should be avoided as it makes the exposure less extensile and places the weight of the buttocks on the incision after surgery. A straight incision avoids these problems whilst still affording excellent visualization of posterior structures.

The fascial incision should be slightly oblique, with the inferior portion positioned more anterior and the posterior portion more posterior than the skin incision. This orientation avoids the tendon of the gluteus maximus in the distal part of the incision and allows additional posterior exposure at the proximal part of the incision if needed. The fascia should be divided between the insertions of the gluteus maximus and tensor fascia muscles (Fig. 6). Posterior visualization may be limited by tightness of the posterior part of tensor fascia lata, which can be remedied by extending the proximal fascial incision in line with the fibers of the gluteus maximus muscle. Bleeding with this maneuver can be avoided by incising only the fascia, as spreading the gluteus maximus muscle belly too far proximally tears difficult-to-coagulate branches of the inferior gluteal vessels. Although not routinely necessary, releasing part or all of the gluteus maximus insertion on the femur also increases posterior exposure. Bleeding can be avoided by releasing only the tendinous portion of the insertion, as several large vessels are located within the muscle fibers attaching to the femur deep to the tendon.

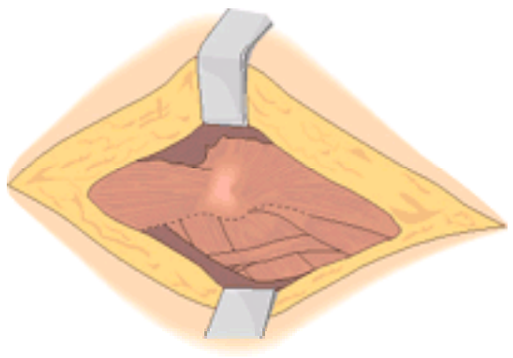


Fig. 6. Surgical exposure after fascial incision during the Kocher–Langenbeck approach.

The thick bursa covering the greater trochanter and posterior hip muscles should be incised along the posterior margin of the greater trochanter to expose the short external rotator muscles. The piriformis tendon is then easily palpated deep to the posterior border of the gluteus medius tendon. The piriformis and its underlying joint capsule can then be transected close to their insertion at the base of the femoral neck. Care should be taken to divide these structures at their attachment to bone in order to preserve length for reattachment without creating an external rotation contracture. After mobilization of the piriformis tendon, the posterior margin of the gluteus minimis muscle can be bluntly elevated to incise the joint capsule. Using the electrocautery, the remainder of the short external rotators and the joint capsule are detached from the posterior margin of the greater trochanter (Fig. 7). Progressive internal rotation of the hip keeps these tissues under tension and facilitates their release. Bleeding can be limited by coagulating the multiple branches of the medial femoral circumflex artery as they are transected. The short external rotators should not be elevated from the capsule, as this single layer provides stronger tissue for reattachment at the end of the procedure. The capsulotomy should extend from the inferior aspect of the quadratus femoris muscle to the superior margin of the piriformis tendon. The lesser trochanter and iliopsoas tendon should be easily visualized.

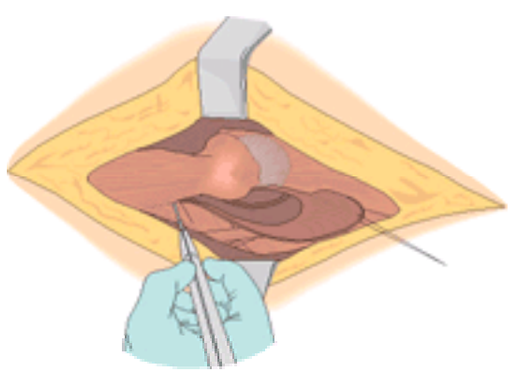


Fig. 7. Release of posterior soft-tissue attachment to the greater trochanter.

The hip joint is then dislocated by adduction, flexion, and internal rotation. A cobra retractor placed under the posterior femur distal to the lesser trochanter elevates the

femur from the wound as it retracts the posterior tissues. The desired level of femoral-neck osteotomy, established during preoperative templating, is then marked by measuring from the lesser trochanter, and the bone is cut with an oscillating saw. The lateral femoral neck is removed to prevent varus reaming and malposition of the femoral component.

A single assistant can provide excellent exposure by standing opposite the surgeon on the distal side of the flexed, adducted, and internally rotated leg. The medial side of the tibia is braced against the assistant's side when the knee is flexed 90°. In this position, both the assistant's hands are free. Excessive internal rotational force through the long lever arm of the femur and tibia must be avoided, as the insertion of the gluteus maximus creates an unyielding tether and a spiral fracture of the femur may result. Additional internal rotation can safely be accomplished, if needed, by releasing the gluteus maximus tendon from the femur.

After femoral instrumentation, the canal is packed and the acetabulum exposed. A large hip retractor under the femur and over the anterior rim of the acetabulum will translate the femur anterior to the acetabulum. Incision of the posterior inferior and superior lateral joint capsule with the origin of the reflected head of the rectus femoris muscle facilitates mobilization and retraction of the femur. This capsulotomy should continue until a circumferential view of the acetabulum is achieved (Fig. 8). Branches of the obturator artery are located adjacent to the extra-articular surface of the inferior joint capsule, and care should be taken to avoid dividing these vessels, as they are troublesome to coagulate. During insertion of the socket, adequate anterior femoral translation is necessary to prevent the femur from pushing the acetabular instrumentation into a retroverted position. Excellent exposure without capsulectomy can be attained. Repair of the preserved posterior capsule improves stability of the prosthetic joint during the early time after surgery. The capsule and muscles should be sutured as one layer to drill holes through the posterior margin of the greater trochanter (Fig. 9).

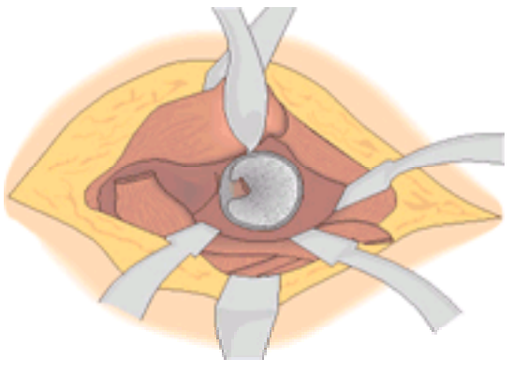


Fig. 8. Acetabular exposure by femoral retraction during the Kocker–Langenbeck approach

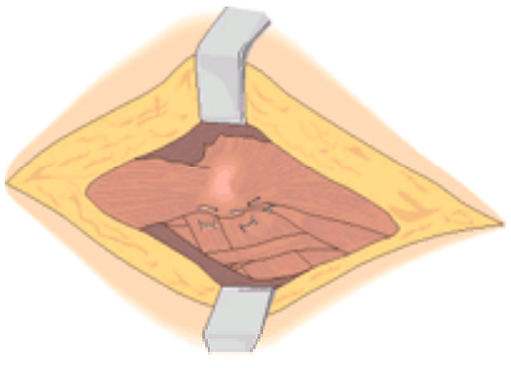


Fig. 9. Posterior soft-tissue repair during the Kocker–Langenbeck approach

The posterolateral approach described above allows excellent visualization with minimal trauma to soft tissues from intramuscular dissection or excessive retraction. Adequate anterior translation of the femur allowing accurate placement of the acetabular component and secure posterior soft-tissue repair lessen the increased incidence of dislocation. The majority of routine and more difficult hip replacements can be well managed through this exposure.

Rehabilitation following hip replacement

Because the physical capacity of patients undergoing hip replacement varies greatly, the postoperative rehabilitation program should be flexible. The otherwise healthy, well-conditioned, middle-aged patient with a single arthritic hip, the patient with multiarticular rheumatoid arthritis, and the elderly, inactive patient have different abilities and needs for rehabilitation. Common needs for all patients immediately after surgery include instructions in bed exercises for regaining early motor control of the extremity, assistance with early ambulation, and proper techniques for getting in and out of beds and vehicles. An overhead frame and trapeze attached to the bed facilitate bed exercises, starting on the day of surgery. Early muscle-setting exercises of all lower-extremity groups and frequent position changes with assistance enhance early pain control and mobility. Early correction of shifts in fluid volume after surgery decreases the likelihood of a delay in the ability to participate in out-of-bed activities the day after surgery. Ambulation supervised by a physical therapist increases daily until the patient is adept in the use of a walker or crutches. The amount of weight borne on the operated leg depends on multiple factors, but after the routine hip replacement placing as much weight as is comfortable is best. Early weight bearing facilitates balance and return of motor control, and decreases strain on upper extremities for patients with multiple arthritic joints. Some problems such as periprosthetic fractures or large bone grafts require more limited weight bearing.

The timing and selection of conditioning exercises after hip replacement should be tailored to the individual patient. Ideally, conditioning is not limited to the initial postoperative period, but becomes a fixture of the patients lifestyle. Too aggressive exercise before adequate soft-tissue healing and return of motor control increases discomfort and swelling, and is counterproductive. Over-enthusiastic therapists and patients should be counseled to avoid pain either during or immediately after exercise in the early postoperative period. The average patient has sufficient strength and balance to use a cane by 2 to 4 weeks. Use of the cane should continue until there is minimal limp. Walking for extended distances or on unlevel surfaces requires the use of a cane for longer periods. The patient can begin walking in chest-deep water when the incision is healed. Swimming or hydrobic classes are excellent for long-term aerobic conditioning for patients after hip replacement and can begin when the patient begins walking without the cane.

A continuing, independent exercise program is beneficial for long-term, optimal hip function. However, the patient should avoid repetitive activities likely to decrease the longevity of the arthroplasty. This is especially important for younger patients. An aerobic cross-training program can utilize swimming-pool exercises, ergometers, and light weight–high repetition resistive exercises for the upper body. The use of weights for resistive strengthening of the muscle groups across the hip affords limited benefit and such maneuvers as squats and leg presses should be avoided. Walking can be an important part of the exercise program, especially for older patients. Encourage the patient who wants to include walking longer distances for exercise to use a cane even if they don't need to during their daily routine. Using the cane provides the same level of exercise without increasing component wear. The patient at risk for long-term, wear-related failure and reoperation is wise to avoid making weight-bearing activities a major part of their aerobic conditioning. Recreational activities resulting in excessive loads on the prosthetic surfaces should be avoided. Tennis and snow skiing are recreations frequently in question. Running during tennis should be avoided. The patient who skis after hip replacement should do so only at a level of aggressiveness below their ability. Swimming, golf, dancing, and cycling are very acceptable recreations for the patient with a replaced hip.

Complications of total hip replacement

Although total hip replacement is highly successful for a large majority of patients, the surgeon and patient must know the operation is complex and there is a chance of major complications.

Infection

Joint sepsis following hip replacement is a devastating complication, usually requiring removal of the prosthesis for treatment. Infection rates following the introduction of total hip replacement were unacceptably high. Charnley and Wilson reported deep sepsis in 7 and 11 per cent, respectively. The use of prophylactic antibiotics

lowered the risk dramatically, and today deep sepsis is generally estimated to occur in approx. 1 per cent of cases. Sepsis after revision hip replacement occurs two to three times more commonly than after first-time surgery. Elimination of sources of infection before surgery, a controlled, clean operating-room environment, and meticulous surgical technique can decrease the incidence of infection.

Fitzgerald categorized deep sepsis into three stages based on the time of initial symptoms and the clinical cause of the infection. Stage I infections are deep infections that become apparent during the early period after surgery. However, at times it is difficult to differentiate a superficial from a deep infection even during wound debridement in the operating room. An indolent infection that usually becomes apparent 6 to 24 months postoperatively is a stage II infection. Stage III infections are hematogenous infections that develop in a previously asymptomatic hip.

Diagnosing a stage I or stage III infection is usually simple. The patient with a stage III, or hematogenous, infection will usually present with recent onset of fever and hip pain, and have an obvious distant source of the infection. The diagnosis is best confirmed by aspiration of purulent material from the joint. Advice to seek early medical management of any bacterial infection is an important part of patient education after total hip replacement. Even relatively minor bacterial infections such as an infected ingrown toenail can have disastrous consequences if microbes find their way to the prosthetic joint. Patients should also receive prophylactic antibiotics during intra-abdominal, gynecologic, or urologic surgery. Traditionally, the use of antibiotics for routine dental prophylaxis has been recommended. Recent review of the issue prompted a policy statement by the American Dental Association and the American Academy of Orthopedic Surgeons narrowing the indications for antibiotics. Now antibiotics are recommended for dental work in the absence of infection only when the patient is immunocompromised, is within 2 years of a major joint replacement, or major gum bleeding is anticipated with the procedure.

Stage II infections are usually caused by less virulent organisms and are frequently difficult to diagnose as they rarely cause systemic signs of sepsis. The history and examination are the most valuable diagnostic tools for the experienced clinician. Multiple previous hip operations, delayed wound healing, unanticipated return to the operating room, and lengthy operating times are frequently part of the history. Patients with stage II infections typically never achieve the usual relief of arthritic pain after their hip is replaced. Radiographic signs of progressive lucency at the bone–cement or bone–prosthesis interface, component migration, periosteal reaction, and lytic areas within bone become apparent with time. Joint aspiration and culture is a valuable tool for making the diagnosis of infection and identifying the offending organism. Reports of the accuracy of aspiration and culture vary. False-negative results are uncommon in most series but false-positives occur frequently. Some have recommended repeat aspiration when there is a positive culture without a strong clinical suspicion for infection. Most surgeons now perform aspiration and culture before surgery only when clinical findings suggest infection.

The appropriate management of infection after total hip replacement depends on several host-, infecting organism-, and operation-specific factors. The most appropriate treatment regimens remain controversial. Antibiotics alone are often adequate treatment for superficial infection such as cellulitis. Wound debridement without removal of components may be effective when done very soon after the onset of deep infection. This approach should only be considered when the infection has been present just a few days, the patient has a competent immune system, and the hip replacement was functioning well before the onset of infection.

The majority of cases of deep infection are treated by removal and replacement of infected implants. The optimal timing for reimplantation of prosthetic components is not clear. Immediate, one-staged exchange is done more commonly in Europe, while in the United States a two-staged exchange is more common. Antibiotic-impregnated bone cement is used when immediate exchange is performed. Recent investigators have also advocated the use of sterilized components coated with antibiotic-impregnated bone cement as a prosthetic spacer for staged exchanges. The spacers are implanted loosely within bone to maintain limb length and muscle tension, and to deliver high local concentrations of antibiotic. Since these are not intended to be permanent implants, larger amounts of antibiotic are included without concern for weakening the cement. At a second stage a revision to well-fixed components is performed.

Deep venous thrombosis

Thrombosis of the large veins of the lower extremity and pelvis is one of the more common, significant complications following total hip replacement and is probably the most common cause of procedure-related death. Dependent edema of the operated extremity is a common early finding after hip replacement and is difficult to differentiate from thrombosis without diagnostic studies. Even large thrombi usually do not result in total occlusion of the lumen of the vessel, and phlebitis often causes minimal clinical symptoms. Pulmonary embolism may be the first sign of deep venous thrombosis.

Contrast venography is the most accurate means of diagnosing such thrombosis. However, venography is an invasive study that has some risk of allergic reaction and inducing thrombosis. As with most studies, skill is required to perform and interpret the test.

Nuclear venography uses injected radioactive isotopes and a gamma camera to provide an image of the venous system. The chance of allergic reaction to dye is eliminated and patient discomfort from the procedure is minimal. Unfortunately, the accuracy of the procedure is less than that of contrast venography.

Duplex ultrasonography uses high-frequency sound waves to image changes in venous compressibility and flow. The duplex scan is safe and causes minimal patient discomfort. The study accurately identifies clots within larger veins but is less accurate for thrombi below the knee. The duplex scan is useful for screening the high-risk patient before their discharge from the hospital.

Pulmonary embolism, like deep venous thrombosis, is difficult to diagnose on clinical findings alone, as many emboli cause no symptoms. When present, chest pain, shortness of breath, hemoptysis, and tachycardia suggest pulmonary embolus. The electrocardiogram and arterial blood gas may also be abnormal.

Prophylaxis effectively decreases the incidence of deep venous thrombosis after hip replacement. Currently, sequential compression stockings, low-dose warfarin, and low molecular-weight heparin are most commonly used to prevent clots. Which regimen and duration of use are the most effective is currently unresolved.

Anticoagulation is the treatment of choice when deep venous thrombosis occurs. Immediate heparinization followed by the prolonged use of warfarin decreases the chance of propagation of thrombus, pulmonary embolism, and chronic edema of the lower extremity. Placement of a vena cava filter effectively prevents pulmonary embolus when anticoagulation is ineffective or contraindicated.

Wear

A prosthetic hip is a heavily loaded mechanical device with moving parts. The generation of small particles from wear on the articulating surfaces is unavoidable and has significant consequences (Fig. 10). These particles elicit a cellular and humoral response causing bone resorption and loosening of prosthetic components. The lysis or resorption of bone can be an extensive and progressive process resulting in massive bone deficiency (Fig. 11). Decreasing the cumulative load on articulating surfaces, improving the resistance of materials to wear, preventing access of debris particles to bone, and decreasing the patient's response to particles are means of addressing the wear problem.

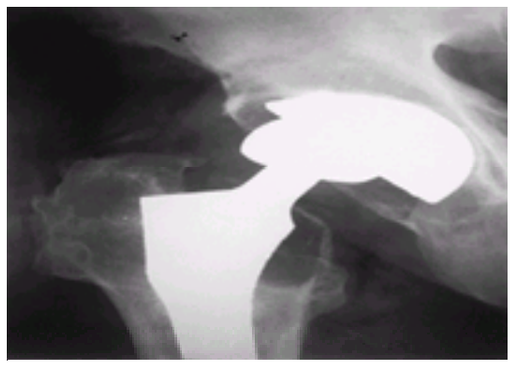


Fig. 10. Polyethylene wear following total hip replacement; note the osteolysis within the proximal femur.

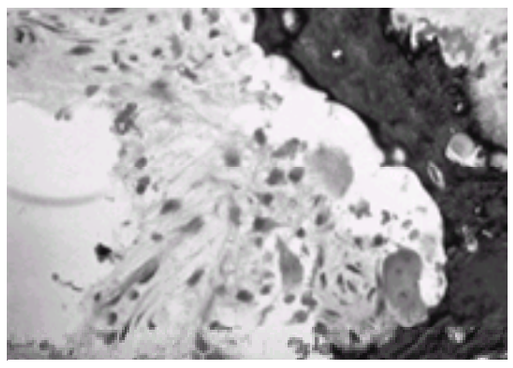


Fig. 11. Histology of aggressive lytic response to wear debris particles; note multinucleated osteoclasts.

The amount of debris generated from articulating surfaces is proportional to the cumulative loads applied to those surfaces. Weight loss and the avoidance of high-end activity are patient-controlled means of decreasing wear. The surgeon should emphasize the importance of wear-related problems in discussions with younger, heavier, and more active patients before surgery.

Surgical technique and component design also influence the amount of wear debris generated. Larger-diameter prosthetic heads result in larger volumes of debris particles. Sockets with an excessively vertical position wear faster and component impingement also causes more wear.

Polyethylene is used as the articulating surface in most prosthetic sockets, and is the material most responsible for osteolysis. Wear of polyethylene, although not preventable, can be minimized. Oxidation causes a decrease in the length of the polymer chains and increased wear. Oxidation of polyethylene accelerates with increasing component shelf-life before implantation. Sterilization by exposure to gamma radiation in open air also increases the rate of oxidation. Mechanical properties are improved with alternative methods of sterilization and packaging in inert gases

In vitro studies indicate superior wear properties of ceramics against polyethylene and ceramic prosthetic heads may provide more durability for younger patients. The disadvantages of ceramics include higher cost and the possibility of breakage. Ceramic rather than polyethylene socket inserts have theoretical promise and are currently being used with ceramic prosthetic heads in clinical trials. Metal/metal articulations are also being used in clinical trials.

Limb-length discrepancy

Limb-length discrepancy following hip replacement is a common initiating complaint in malpractice suits. Although careful preoperative planning and operative technique minimize the chance of unequal lengths, the problem is impossible to eliminate. Despite great attention to a comprehensive approach to limb lengths, Woolson reported greater than 0.6 mm of limb lengthening in 2.5 per cent of patients. Others have reported a much higher incidence of lengthening. Reckling reported clinical symptoms requiring a lift on the contralateral side in 27 per cent of 150 patients.

Preoperative planning with radiographic templates is the most effective means of reproducing equal limb lengths. The accuracy of devices used during surgery is limited, as any change of position of the leg significantly alters the measurements. The surgeon must consider both limb length and offset during planning before surgery and assessment during surgery, as alterations in offset affect the surgeon's ability to judge soft-tissue tension. A decreased offset may result in inadvertent limb lengthening to produce appropriate soft-tissue tension. The surgeon must also clearly understand the definition of real and apparent length discrepancies. An apparent discrepancy results from secondary pelvic obliquity despite an actual equality in the lengths of the component bones. Causes of pelvic obliquity and apparent discrepancies include lumbar scoliosis and fixed hip abduction or adduction contractures.

Although limb lengthening can contribute to sciatic nerve palsy and shortening can contribute to dislocation and limp, discrepancies rarely significantly effect the results of hip replacement. When the problem is noted the patient should be reassured and a lift prescribed if needed. Approximately 0.5 cm of lift can be worn as a heel pad inside the shoe. Additional lift must be applied under the heel of the shoe. Clear discussion before surgery of the possibility of leg-length discrepancy is an effective means to avoid having a disgruntled patient afterwards.

Periprosthetic femur fracture

Periprosthetic fracture of the femur can occur during or after hip replacement. The hoop stresses generated when implanting a tight-fitting, uncemented stem result in an increased incidence of proximal femoral cracks. These proximal splits are inconsequential when recognized during surgery, as they are easily managed by circlage wire placement. If unrecognized, the split can propagate during weight bearing, causing loss of fixation of the femoral component.

Fractures of the femoral shaft during surgery usually occur through weak points or perforations in the cortex and are much more common in revision operations. Inadvertent perforation of the cortex during removal of cement must be addressed. Fractures of the femoral shaft usually require the use of longer stems to stabilize them. The surgeon should expose the shaft of the femur by reflecting the vastus lateralis muscle from its origin along the linea aspera, and carefully identify and ligate the multiple perforating vessels. After exposure, the surgeon should stabilize the fracture with circlage wires before implanting the long stem.

Although management of the shaft fracture occurring after surgery also usually requires the use of a longer stem, internal fixation alone can be used when the fracture is distal to a well-fixed stem. Either long metal or cortical allograft bone plates fixed by circlage wires are effective for fixation of these fractures. Allograft cortical bone plates have a high rate of union, supplement thin host cortex, and add rotational stability to fractures managed with long-stem components.

Dislocation

Unlike other complications following hip replacement, dislocation often occurs despite an otherwise excellent result. For these patients, dislocation represents a painful, expensive inconvenience that may require additional surgery. The reported incidence of dislocation varies but in the best of hands is probably 2 to 3 per cent. The rate of dislocation after revision hip replacement is at least two to three times that of primary procedures. Since patients usually have several dislocations before additional surgery to correct the problem, a patient with a dislocated hip replacement is a frequent occurrence in a hospital with a large volume of hip surgery.

To effectively treat a patient with a dislocation, the surgeon must be able to identify the cause of instability. The most common cause is self-correcting and resolves spontaneously. Early after surgery, a degree of instability is inherent in every prosthetic hip. During this period, control and tone of muscles about the hip are markedly decreased. The repaired, soft-tissue surgical entry to the joint takes several weeks to heal and contract. Even without other predisposing factors, if the patient exceeds a safe arc of motion before return of muscle tone and capsular healing, the hip will dislocate. Morrey reported that 40 per cent of initial dislocations occurred within the first month; 70 per cent were within that period in Williams' series. Multiple dislocations are less likely if the initial episode occurs within a month of surgery than if it occurs longer than 3 months after surgery. These patients with early initial dislocations are likely to not have predisposing causes other than inadequate healing time. Instructions on safe zones of motion, how to dress, enter an automobile, and arise from sitting are important components of rehabilitation in the early period after surgery. Patient education and compliance are the most effective tools for preventing dislocation within the early, high-risk period.

Component malposition, especially the acetabular component, is the most common cause of recurrent dislocations. Lewinnek defined a safe zone of socket orientation to be between 5° and 25° of anteversion and between 30° and 50° of abduction. Patients with a socket positioned outside this safe range had a 6 per cent dislocation rate compared to a 1.5 per cent rate for those within the safe zone. Malposition of the femoral component is rarely a cause of dislocation. The exception is the patient with excessive femoral anteversion associated with developmental dysplasia. Anterior dislocation may result if the condition is not recognized.

Tissue tension of capsule, muscle and fascia crossing the hip joint influence the degree of stability of the prosthetic hip. Loss of adequate soft-tissue tension may result from multiple operations, shortening of the femoral neck, proximal socket placement, loosening and migration of components, and trochanteric fracture or nonunion. The amount of femoral component offset also influences soft-tissue tension. A decrease in offset for any given limb length results is less soft-tissue tension.

Dislocation of a prosthetic hip is treated by closed reduction under conscious sedation or general anesthesia. After returning the prosthesis to its proper position, the surgeon may elect to apply a cast or brace for a time to limit range of hip motion. With initial dislocations and reliable patients, reinforcement of the positions to avoid is frequently the only management required after reducing the hip.

After the second or third episode, the likelihood of repeated dislocations is very high, and additional surgery should be considered. Accurate identification of the cause of repeated dislocations is important during surgery. The hip should be placed through a range of motion to determine the point and direction of dislocation. Component orientation should be checked and changed as needed. Most current designs of prosthetic cup are modular with polyethylene inserts that can be changed without removing the metal shell. Placing an insert with an extended 10° or 20° lip in the direction of instability may be enough to compensate for poor socket orientation.

At extremes of hip motion, impingement of the prosthetic femoral neck on the cup rim causes a levering of the head from the acetabulum. The use of a polyethylene insert with an extended rim also increases the likelihood of impingement. For example, if the extended rim is placed on the posterior side of the cup, the posterior femoral neck may contact the polyethylene rim, causing an anterior dislocation when the hip is extended.

Femoral components with a small head diameter:neck diameter ratio decrease the range of hip motion before impingement occurs. Thus, larger-diameter heads are more stable than smaller diameter. Modular prosthetic heads are mated to the femoral neck through a morse taper. Decreasing the taper within the head results in locking before the neck seats as deeply within the head. This effectively increases limb length, offset, and tissue tension. However, longer neck tapers require protective skirts to decrease the chance of fatigue failure of the femoral neck. The skirt increases the likelihood of impingement by decreasing the head diameter:neck diameter ratio. If these skirted prosthetic heads are used, there must be no impingement when range of motion is checked before completion of the procedure.

Inadequate tissue tension can be addressed by increasing the neck length of a modular prosthetic head, using a femoral stem with increased offset, or by advancing the greater trochanter. Constrained sockets that capture the prosthetic head are available. If no identifiable cause of dislocation is found, the use of these components should be considered.

Further reading

Abraham WD, Dimon JH. Leg length discrepancy in total hip arthroplasty. *Orthopedic Clinics of North America* 1992; **23**, 201–10. [Describes methods to avoid and correct leg length discrepancy.]

Amstutz HC, Friscia DA, Dorey F, Carney BT. Warfarin prophylaxis to prevent mortality from pulmonary embolism after total hip replacement. *Journal of Bone and Joint Surgery* 1989; **71A**: 321–6. [Retrospective review describing efficacy of warfarin in prevention of the fatal pulmonary embolus.]

Barrack RL, Harris WH. The value of aspiration of the hip joint before revision total hip arthroplasty. *Journal of Bone and Joint Surgery* 1993; **75**: 66–77. [Describes limited accuracy of preoperative joint aspiration for diagnosing infection.]

Blount WP. Don't throw away the cane. *Journal of Bone and Joint Surgery* 1956; **30A**: 695–708. [Excellent description of components of joint reactive forces on the hip.]

Bourne RB, Hunter GA, Rorabeck CH, McNab JJ. A six-year follow-up of infected total hip replacements managed by Girdlestones's arthroplasty. *Journal of Bone and Joint Surgery* 1984; **66**: 340–3. [Review describing results of pain and function following resection arthroplasty.]

Callaghan JJ, Brand RA, Pedersen DR. Hip arthrodesis. a long term follow-up. *Journal of Bone and Joint Surgery* 1985; **67**: 1328–35. [Long-term patient satisfaction and function after fusion.]

Capello WN. Preoperative planning of total hip arthroplasty. In: Anderson LD ed. American Academy of Orthopaedic Surgeons *Instructional Course Lectures*, Vol. XXXV, pp. 249–57. Mosby, St. Louis, 1986. [Excellent review of preoperative planning for total hip replacement.]

Evans BG, Cuckler JM. Evaluation of the painful total hip arthroplasty. *Orthopedic Clinics of North America* 1992; **23**: 303–11. [Algorithm for evaluation of the painful hip replacement.]

Ficat P, Raymond P. Treatment of avascular necrosis of the femoral head. In: The Hip Society *The hip*, p. 280. Mosby, St Louis, 1983. [Describes techniques for avoiding replacing the hip in patients with avascular necrosis.]

Fitzgerald RH, Nolan DR, Ilstrup DM. Deep wound sepsis following total hip arthroplasty. *Journal of Bone and Joint Surgery* 1977; **59**: 847–55. [Comprehensive review of management of the infected hip replacement.]

Garvin KL, Hanssen AD. Infection after total hip arthroplasty; *Journal of Bone and Joint Surgery* 1995; **77**: 1576–86. [Management of the infected hip replacement.]

Glassman AH, Engh CA, Bobyn JD. A technique of extensile exposure for total hip arthroplasty. *Journal of Arthroplasty* 1987; **2**: 2–11. [Describes technique for sliding trochanteric osteotomy.]

Hardinge K. The direct lateral approach to the hip. *Journal of Bone and Joint Surgery* **64B**: 17–19. [Describes transgluteal anterolateral surgical exposure.]

Jacobs JJ, Shanbbag A, Glant TT, Black J, Galante JO. Wear debris in total joint replacements. *Journal of the American Academy of Orthopedic Surgeons* 1994; 212–20. [Describes sources and results of particulate debris after arthroplasty.]

Jones RE, Cadambi A, Maale GE. Antibiotic-cement–implant composites: a protocol for staged revision of infected total hip and knee arthroplasties. *Orthopaedics* 1995, 45–54. [Describes a method for maintaining function during two-staged revisions for infection.]

Kavanagh BF. Femoral fractures associated with total hip arthroplasty. In: Roberson JR, Nasser S, ed. *Complications of total hip arthroplasty*, pp. 249–58, *Orthopedic Clinics of North America*. Saunders, Philadelphia, 1975. [Comprehensive review of treatment of periprosthetic femoral fractures.]

Li S, Burstein AH. Ultra-high molecular weight polyethylene, the material and its use in total joint implants. *Journal of Bone and Joint Surgery* 1994; 1080–90. [Description of manufacturing and material properties of polyethylene.]

Lieberman JR, Geerts WH. Prevention of venous thromboembolism after total hip and knee arthroplasty. *Journal of Bone and Joint Surgery* 1994; 1239–49. [Review of methods of prevention of deep vein thrombosis.]

Millis MB, Murphy SB, Poss R. Osteotomies about the hip for the prevention and treatment of osteoarthrosis. *Journal of Bone and Joint Surgery* 1995; 626–47. [Comprehensive review of osteotomy techniques.]

Mont MA, Hungerford DS. Non-traumatic avascular necrosis of the femoral head. *Journal of Bone and Joint Surgery* 1995; 459–73. [Describes diagnosis and management of different stages of osteonecrosis.]

Sponseller PD, McBeath AA, Perpich M. Long-term follow-up of hip arthrodesis performed in young patients. In: *The hip*, pp. 43–53. Mosby, St. Louis, 1984. [Excellent long-term outcome of hip arthrodesis.]

Tigges S, Stiles RG, Meli RJ, Roberson JR. Hip aspiration: a cost effective and accurate method of evaluating the potentially infected hip prosthesis. *Radiology*, 1993; 485–8. [Describes accuracy of aspiration for evaluation of infection.]

Woo RY, Morey BF. Dislocations after total hip arthroplasty. *Journal of Bone and Joint Surgery* 1982; **64A**: 692–7. [Incidence and risk factors for dislocation after hip replacement.]

Zimlich RH, Fulbright BM, Friedman RJ. Current status of anticoagulation therapy after total hip and total knee arthroplasty. *Journal of the American Academy of Orthopedic Surgeons* 1996; **4**(1): 54–62. [Comprehensive review of techniques for deep vein thrombosis prophylaxis.]

46.2.5 The knee

Richard S. Limbird and Michael Belanger

Introduction
Anatomy
Patient evaluation
Physical examination
Radiographic examination
Osteoarthritis
Osteonecrosis
Surgical procedures
Arthroscopy
Osteotomy
Arthroplasty
Total knee arthroplasty
Patellofemoral joint
Further reading

Introduction

The patient's age frequently influences the reason that he or she seeks medical attention for evaluation of a knee complaint. Younger patients' complaints focus on pain and stability, while older patients' complaints are primarily pain and impaired function. Mechanical failure of the knee joint can result from meniscal tears, ligamentous injuries, and patellofemoral syndromes. Pain may result from any of these problems, as well as from inflammatory conditions, articular chondrosis, degenerative arthritis, and osteonecrosis. A careful history and physical examination, with ancillary studies when indicated, usually provide the clinician with a diagnosis and a treatment plan appropriate for the disease process.

Anatomy

The knee depends on its ligamentous structures for stability. The location of the attachments of the ligaments stabilize the knee in sagittal, rotational, and axial planes. The medial and lateral meniscus aid in stabilizing the knee in a rotational plane and function as secondary restraints in the sagittal plane as well as distributing force along the longitudinal axis of the knee. As the knee moves from flexion to terminal extension, the tibia externally rotates.

In addition, there is a translational posterior motion of the femur on the tibia as flexion begins and after 30 degrees of flexion. The relationship between the two articular surfaces remains static in the translational plane.

The quadriceps muscle extends the knee. The four muscles merge into a common tendon and insert into the patella. Extensions of the central part of this common tendon are the patellar tendon, which inserts into the tibial tubercle. The importance of the quadriceps' mechanism to the function of the knee is significant. The patella improves the mechanical advantage of the quadricep by moving its insertion further anterior from the axis of rotation of the knee. This mechanical gain increases from 15 to 45 degrees of knee flexion and decreases as the knee reaches full flexion. Flexion of the knee is performed primarily by the semimembranous, semitendinous, and biceps femoris muscles and secondarily by the gastrocnemius and popl-iteus muscles. Contraction of the extensors and flexors of the knee provide dynamic stabilization of the knee in response to stress placed on the knee.

The femoral nerve innervates the quadriceps muscle. It is formed by the posterior rami of L2, L3, and L4 spinal nerves. After the nerve leaves the femoral triangle, it becomes the saphenous nerve, which is a sensory nerve, and descends the leg with the greater saphenous vein to give off intrapatellar branches to the knee before terminating near the medial malleolus.

The obturator nerve is formed by the anterior rami of L2, L3, and L4. It innervates the adductors of the thigh; in addition it supplies the capsule of the knee, entering the joint with the middle geniculate artery.

The sciatic nerve is formed by the anterior rami of L4, L5, S1, S2, and S3. It divides in the upper part of the popliteal fossa into the common peroneal and tibial nerve. The tibial component of the nerve innervates the semimembranous, semitendinous, and long head of the biceps femoris. Innervation of the knee joint itself is by branches of the femoral and obturator nerves and by geniculate branches of the peroneal and tibial nerves.

The popliteal artery leaves Hunter's canal, crosses through the popliteal fossa, and divides into the anterior and posterior tibial arteries. Prior to its division, the artery gives off five articular branches and multiple muscular branches. The anastomosis around the knee is formed by the superior medial and lateral geniculate arteries, the inferior medial and lateral geniculate arteries, and the middle geniculate arteries from the popliteal artery, descending branch of the lateral circumflex artery, and the descending geniculate artery from the femoral artery.

The knee joint is a diarthrodial joint divided into three compartments. The articular cartilage varies in thickness; the thickest portion being the surface of the patella at 5 mm. The role of articular cartilage minimizes stress on subchondral bone and reduces the friction on the joint surface. Injury, degeneration, and inflammatory tissue can affect the integrity of this surface.

When considering the reconstructive procedures, attention to the alignment of the lower extremity is critical. Axial, sagittal, and coronal alignment is of particular importance in the performance of knee arthroplasty and osteotomies.

Normal alignment has been evaluated in several anthropometric studies. It is accepted by most that in stance the knees are closer to the midline than the hips and the ankles are even closer to the midline than the knees. A line drawn from the center of the femoral head to the center of the ankle is the mechanical axis of the extremity and is approximately 3 degrees varus. The anatomic axis of the lower extremity is determined by the intersection of the anatomic axis of the femur, which is valgus, and the axis of the tibia. The anatomic axis is within a range of 5 to 9 degrees of valgus.

Normal alignment would imply that, if an individual has no abnormalities of the cartilage, the joint can be loaded in the stance phase of gait so that forces are evenly distributed throughout the knee.

Alignment is an important consideration for the surgeon. Alteration in the alignment beyond the norm can accelerate cartilage deformation, and wear. Re-establishment of a more favorable alignment may improve the function of the knee.

The ideal alignment in total knee replacement is important to the longevity of the implant. If the implant can be loaded ‘favorably' with optimal implant stability and design there is the potential for a durable arthroplasty.

Patient evaluation

The adult patient with knee pain can be of any age. The clinical history of the patient with knee pain and dysfunction is important. A history of injury should be carefully explored as intra-articular fractures either involving the distal femur, proximal tibia, or patella can lead to post-traumatic arthritis via either a significant chondral injury or postfracture malalignment (Fig. 1). Anterior cruciate ligament injuries after initial recovery from the injury can be associated with minimal pain and minimal instability, yet with the resulting chronic anterolateral instability most patients develop meniscal tears and subsequent degenerative disease.



Fig. 1. Varus deformity 3 years after bicondylar fracture of the proximal tibia.

Patellofemoral syndromes are common and associated with varying degrees of knee dysfunction. Isolated degenerative arthritis involving the patellofemoral joint, although not common, is extremely disabling. Articular injury to the patella occurs as a result of a chronic malalignment syndrome of the patellofemoral joint or trauma with or without fracture of the patella.

From the beginning, the physician should be open to an extensive differential diagnosis and evaluation of the patient with knee pain. It is especially important as a surgeon to avoid the trap of limiting diagnostic possibilities and arriving at a premature conclusion as to causal relationship. A knee problem can be differentiated by the clinician as either a monarticular disease or manifestation of systemic disease. Familiarity with these patterns is important for a thorough evaluation. Pathologic processes involving the hip joint and the lumbosacral spine not uncommonly produce referred pain to the knee. Synovitis of the knee with other joint involvement is likely to be associated with a rheumatic disorder or other systemic disease. A painful or swollen knee, as a monarticular complaint, may be the manifestations of infection, trauma, or gout and pseudogout.

There has been described a syndrome of reflex SYmpathetic dystrophy of the knee. Although reflex sympathetic dystrophy has been recognized to occur in the upper extremity following trauma, its prevalence with involvement of the knee has not been well-described. The characteristics of intense pain, vasomotor disturbance, delay in recovery of function, and trophic changes in reflex sympathetic dystrophy of the knee vary in their presentation. Almost all cases of reflex sympathetic dystrophy of the knee involve the patellofemoral joint and have been reported to occur after arthroplasty surgery.

Ultimately, an objective history is a means of assessing disability and a reasonable measure of the patient's needs and response to a given treatment.

Physical examination

The examination of the knee begins with inspection of the joint. Asymmetry between the two extremities is important to note and differentiation should be made from a knee effusion and enlargement from synovial capsular thickening. Alignment may be more clearly assessed on the basis of radiographs but axial, sagittal, and coronal alignment can be evaluated by physical examination with a patient, supine, standing, sitting, and in gait. An examination of the patella, tibial tubercle, and malleoli are helpful in assessing torsional abnormalities. Observation of stance and gait are important in the evaluation of the degenerative knee.

Malalignment and angular thrust in gait, subtle at times, are objective measures of the pathologic process present. The active and passive range of motion are noted at the time of examination. The normal arc of flexion (–5 to 135 degrees) is frequently altered in the diseased knee. The location of pain can sometimes be isolated but not uncommonly is diffuse. The patella is easily palpated. In most people both the medial and lateral compartments, at least anteriorly, can be palpated. The superficial location of the knee allows assessment of the articular cartilage, presence of osteophyte, effusion, and capsular thickening. The articular cartilage of the patella is most easily evaluated with the knee in approximately 30 degrees of flexion. Side-to-side motion of the patella within the femoral trochlea with a moderate degree of pressure may produce a grinding sensation (grind test), suggesting pathology of articular cartilage. Lateral displacement of the patella with moderate pressure may produce a quadriceps contraction (apprehension test), a provocative test for patella instability.

Varus and valgus testing of the knee in extension, in 30 degrees, and in 90 degrees of flexion are important in evaluating ligament stability. Examination for anterior and posterior cruciate stability with anterior and posterior drawer test, Lachman, pivot shift, and reverse pivot shift complete the examination for stability. Examination of the ipsilateral hip, ankle, and foot are important and results influence treatment of knee problems. Neurologic and vascular examination of both lower extremities are essential for the patient examination.

Radiographic examination

Radiographic examination of the knee is an objective means of assessment of the joint, articulation, soft tissue, and alignment. The basic evaluation of the knee must include anterior/posterior, lateral, tunnel, and patellofemoral views.

Because the anterior/posterior view is influenced by weight bearing, an anterior/posterior standing view is not necessarily an accurate study if the extremity is too painful to bear weight (Fig. 2). A supine anterior/posterior stress view can confirm clinical suspicion of unicompartmental articular loss as well as ligament stability. A useful adjunct to the anterior/posterior view is the flexion weight-bearing radiograph. This radiograph is taken with a posterior/anterior view with the patient's knee in 30 to 40 degrees of flexion and the X-ray beam at 90 degrees to the long axis of the tibia. The flexion weight-bearing radiograph has a greater sensitivity to articular cartilage loss especially in the areas subject to early degenerative loss.



Fig. 2. Standing anteroposterior view of the knee demonstrating loss of medial joint space as well as a pagetoid distal femur.

Tunnel views of the knee are invaluable in evaluation of the tibial plateau and posterior femoral condyles. Osteochondritis dissecans and osteonecrosis of the femoral condyles are observed more clearly in this view. The Merchant or Lauren views have limitations but are useful in axial evaluation of the patellofemoral joint (Fig. 3).

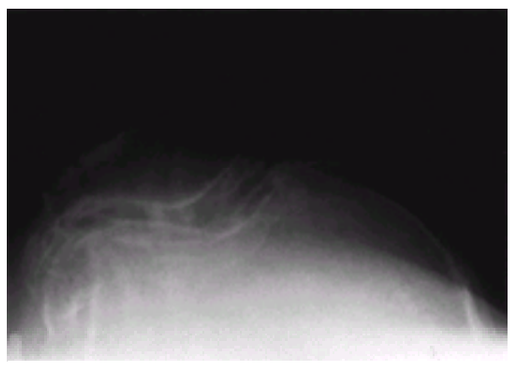


Fig. 3. Axial view of patella femoral joint demonstrating significant arthrosis.

Pathologic changes on plain radiographs can demonstrate effusions, joint space narrowing, subchondral sclerosis, osteophyte formation, subchondral cysts, loose bodies, cortical erosions, and deformity. Neoplastic processes, both benign and malignant, have a predilection for occurring about the knee; therefore, changes in the bony architecture should not be ignored.

Long-leg standing films are used for evaluation and planning of reconstructive surgery in a patient with a degenerative knee. Achieving proper knee alignment is paramount in knee arthroplasty and osteotomy surgery. Standing anterior/posterior on a 3-foot cassette is ideal for evaluation of the axial alignment. Inherent problems exist with this study. Increasing internal rotation of the knee will accentuate the degree of valgus and external rotation of the knee will accentuate the degree of varus. It is helpful for the surgeon to be present when this study is being performed.

In stance, because the ankles are closer to the midline, the mechanical axis drawn from the center of the femoral head to the center of the talus is one of varus (between 2 and 3 degrees). The anatomic axis is the tibial femoral angle and could be measured directly from a radiograph by the intersection of lines drawn through the shafts of the tibia and femur. This angle is variable as well, with a range of normal between 5 and 9 degrees of valgus. The tibial–femoral angle is dependent on the geometry of the femur because the hip, shaft, and knee axes are the result of the inter-relationship of the length of the femur and the length, angle, and rotation of the femoral neck. Joint line position is dependent on the position of the lower extremity. Influenced by abduction or adduction of the hip, variations could be expected to exist from the normal to the pathologic condition. Thigh girth, ipsilateral joint disease, and an aging broad-based gait with secondary abduction of the hip will place the joint line in a secondary valgus position.

Interpretation of radiographs of a prosthetic knee should take into account its design and type of fixation. Axial and sagittal alignment are adequately assessed where coronal alignment, although no less important, may be difficult to interpret on plain radiographs. Depending on fixation, an interface may or may not be clearly visualized. In a cemented implant, a lucency less than 2 mm can occur at the bone–cement interface whereas with a bone ingrowth prosthesis no lucency is expected surrounding the implant. Assessment of the implant requires serial radiographs accurately to evaluate an implant for subsidence and loosening. Failure of implant fixation may be either an aseptic or septic process and plain radiographs will not separate the two (Fig. 4). Fracture and dislocation of the components can occur in the postoperative period.

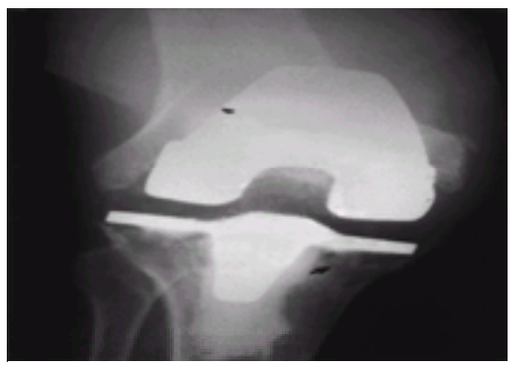


Fig. 4. Alteration in the femoral component position with the area of osteolysis beneath the tibial component led to a diagnosis of sepsis.

Radionuclear imaging of the knee is a reasonable adjunct to plain radiographs of the knee. Technetium-99m phosphate bone scans are helpful in the diagnosis of osteonecrosis, especially when plain radiographs have not yet demonstrated a lesion. The diagnosis of an infectious process is more accurate using either indium or gallium compared with technetium scanning, but should not be relied upon solely for diagnosis (Fig. 5, Fig. 6).

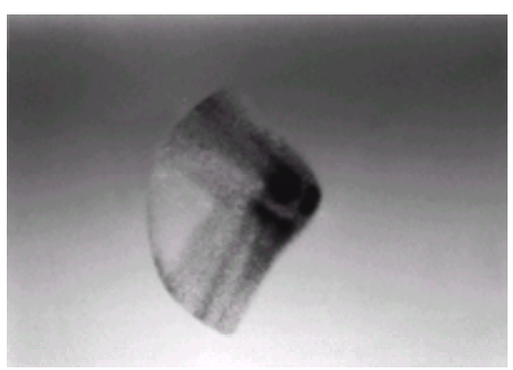


Fig. 5. Bone scan 3 years after total knee arthroplasty with increased uptake at all interfaces.

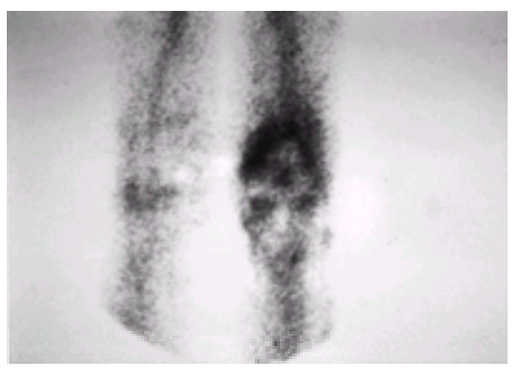


Fig. 6. Gallium scan in the same patient as in Fig. 5.

Arthrography continues to be useful in combination with aspiration for evaluation of a joint prosthesis. More accurate information using this technique can be gained with radiologic subtraction studies. With the advent and availability of magnetic resonance imaging, arthrography has otherwise become an outmoded means of evaluation of the native knee. Magnetic resonance imaging is the primary non-surgical imaging modality of the knee. Accuracy of magnetic resonance imaging examinations range from 64 to 95 per cent for the medial meniscus, and from 83 to 94 per cent for the lateral meniscus. Both computed axial tomography and magnetic resonance imaging are used in the evaluation of the patellofemoral joint (Fig. 7). The computed tomography scan with its excellent detail of bone may be superior as an imaging modality of the patellofemoral joint to assess tilt and subluxation.

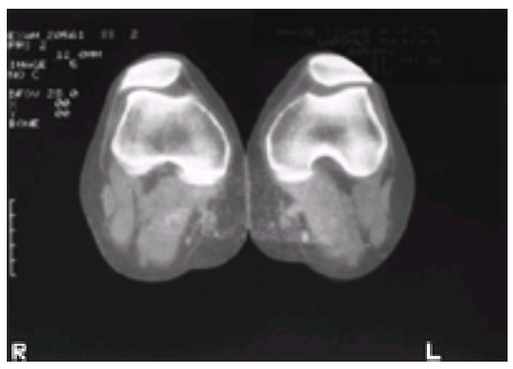


Fig. 7. Computed tomography scan with knees in 20 degrees of flexion demonstrating chronic tilt and subluxation of patellofemoral joints.

Osteoarthritis

The adult patient with knee pain, but without a history of significant trauma is likely to have an inflammatory or degenerative process of the knee. Different stages of the disease process, the patient's age, and postoperative expectation for function and pain relief influence the ultimate decision for treatment. Basic management of all patients with a painful knee from arthritic conditions should include non-steroidal anti-inflammatory medication, orthotics when indicated, rest, localized corticosteroid injections, and weight loss. The response to non-operative treatment is variable, but many patients experience some improvement in their pain and function ability.

Debate exists as to the etiology of osteoarthritis of the knee. Several mechanisms initiating cartilage destruction are known, but ultimately destruction of the matrix supersedes the ability of chondrocytes to repair a local defect. Malunion of the tibia or femur after fracture that alters alignment and leads to pathologic loading of the joint will, in most cases, progress to osteoarthritis with time.

When primary or secondary osteoarthritis exists, an underlying process of matrix destruction occurs as load is in excess of the physiologic limit of the cartilage, and deformation occurs. A self-perpetuating cycle of destruction continues as the force generated from loading the diseased knee produces asymmetric wear of cartilage and subsequent bone loss. This loss of cartilage and bone increases the deformity and the extent of malalignment. Unfortunately, this process increases the force and deformation of the joint surface and secondary asymmetric loading. In the valgus knee, patients can compensate this by alterations in their gait better than patients with varus alignment.

Osteonecrosis

In idiopathic spontaneous osteonecrosis, typical patients are women over the age of 60. The etiology of this disorder is unknown and can involve the osteochondral surface of either side of the knee joint but most commonly involves the medial femoral condyle. Diagnosis can be confirmed by a positive bone scan with focal uptake in the condyle. Magnetic resonance imaging is helpful in defining osteonecrosis of the knee, but the T_1 weighted image not uncommonly exaggerates the apparent size of the lesion generated on plain radiographs. Stages of osteonecrosis of the knee have been described to define the sequence of progression of this pathologic process. In stage I disease, radiographs are normal and the bone scan is positive. Spontaneous resolution is possible if the disease does not progress beyond this stage. As the disease progresses to stage II, plain radiographs demonstrate only slight flattening of the condyle in association with a positive bone scan. If the size of the lesion remains small, SYmptoms frequently resolve. In stage III, radiographs demonstrate an area of lucency surrounding sclerotic subchondral bone. In stage IV, osteonecrosis of the femoral or tibial condyle produces a segment of necrotic bone in the weight-bearing portion of the condyle which can progress to subchondral fracture and collapse. Progression to stage V with joint space narrowing is the end-point of the disease process (Fig. 8, Fig. 9).

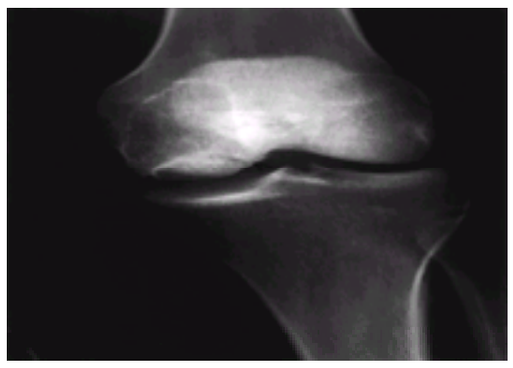


Fig. 8. Stage IV osteonecrosis anteroposterior view of the knee with sequestrum of medial femoral condyle.

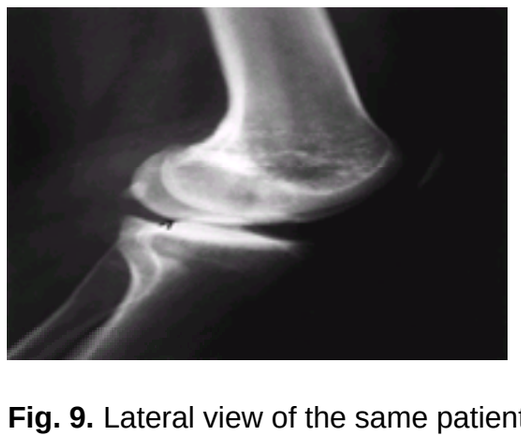


Fig. 9. Lateral view of the same patient as in Fig. 8 demonstrating lesion (2.8 × 3.5 cm).

Patients usually have intense localized pain, and the clinical progression of the disease is highly dependent on the size of the lesion. Lesions that occupy an area less than 3.5 cm in diameter frequently are asymptomatic within a year. Lesions that occupy more than 50 per cent of the width of the medial femoral condyle have progressive pain and secondary destruction of the joint.

The treatment of osteonecrosis is initially conservative until the size of the lesion is elucidated. When degenerative changes develop, surgical treatment is considered. Prosthetic replacement is the recommended treatment in the older age groups and tibial osteotomy, with or without bone grafting, is recommended in the younger age

groups.

H3>Surgical procedures

Arthroscopy

Arthroscopy is a widely accepted modality for treatment of disorders of the knee. In selected patients, arthroscopy is a valuable tool in the treatment of the arthritic knee. While realizing that the surgeon cannot manage all intra-articular disease of the knee, there are definite advantages to arthroscopic procedures. The evaluation of intra-articular pathology is accurate, and operative procedures have a low complication rate with rapid return to full activity. Preoperative mechanical symptoms suggestive of loose bodies, torn meniscus, or impinging osteophyte are associated with more successful arthroscopic management of the degenerative knee. Single compartment disease and a normally aligned knee have a higher success rate of arthroscopic management with a torn meniscus. Worse results have been reported in the valgus knee and chondrocalcinosis. Between 60 and 80 per cent improvement have been seen at 1 year, which deteriorates over time (Fig. 10).

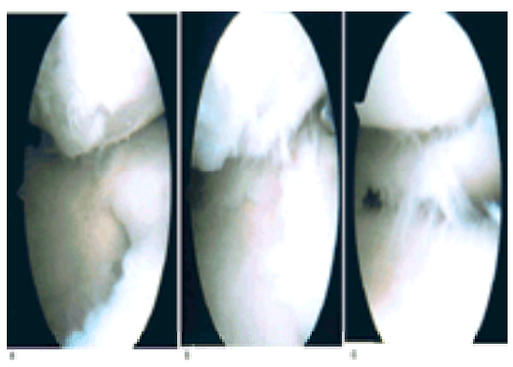


Fig. 10. Arthroscopic view of the medial compartment of the knee before and after partial meniscectomy and debridement.

Abrasion arthroplasty has not been proven to be beneficial to the arthroscopic management of the degenerative knee and in one series as many as 33 per cent of patients were made worse with the procedure and total knee replacement was subsequently performed as a salvage procedure. Newer cartilage resurfacing or healing stimulation procedures are becoming more commonplace. Long-term results of these techniques are yet to be seen.

Synovial proliferative disorders are suitable for arthroscopic synovectomy. Included in this group are rheumatoid arthritis, pigmented villonodular synovitis, hemophilia, and synovial chondromatosis. Arthroscopic synovectomy has generally replaced open SYnovectomy in the rheumatoid and hemophiliac patient because of its reduced morbidity and ease of access to the entire knee joint. Indications for synovectomy in the rheumatoid knee would be failure of medical management with a persistent synovitis and minimal radiographic changes. The indication for synovectomy in the hemophiliac knee is recurrent, frequent hemathrosis, despite adequate hemostatic therapy, with a presence of a synovitis and minimal radiographic changes. With synovectomy, fewer episodes of bleeding can be expected in the hemophiliac, but loss of motion is still a problem. Adequate analgesia and constant passive motion devices can minimize the problem. Although radiographic progression of the disease is not changed, relief can be expected in both groups for as much as 70 per cent of patients at 7 years.

Arthroscopic SYnovectomy alone or in combination with open synovectomy for treatment of pigmented villanodular synovitis has been reported. Reoccurrences of pigmented villanodular synovitis are seen in diffuse disease with incomplete synovectomy (Fig. 11). Reoccurrences are associated with articular cartilage changes. Treatment with radiation after a reoccurrence has been demonstrated to be beneficial in controlling proliferation of the disease.

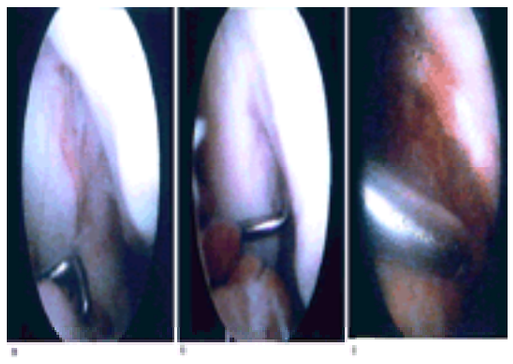


Fig. 11. Arthroscopic view of a knee with pigmented villanodular synovitis before and after synectomy.

Osteotomy

In considering a patient for osteotomy, several considerations must be made. Generally, the patient is younger, active, and the disease is confined to one compartment. Significant loads across the knee are accentuated by malalignment. Calculated static loads are inaccurate in defining stresses in the diseased knee. Separate dynamic studies have shown that the medial compartment can bear the majority of load even with valgus alignment. The goal of osteotomy is correcting the mechanical axis and decreasing loads in the diseased compartment of the knee.

Proximal tibial osteotomy may be performed for both varus and valgus deformities. The most common osteotomy performed is a valgus osteotomy proximal to the tibial tubercle for varus deformity. The advantages of performing an osteotomy proximal to the tibial tubercle are that the bone is cancellous and will heal more rapidly, the contractual forces of the quadriceps are compressive to the site of osteotomy, and, finally, the site of osteotomy is closer to the deformity.

A variety of guides are available. The use of intraoperative radiographs are prudent to improve reproducible postoperative alignment and to avoid complications. Different techniques are used for fixation; these include postoperative immobilization in a cylinder cast, staples, and contemporary fracture fixation techniques. Advantages of rigid fixation are the decreased necessity for postoperative immobilization and the shorter postoperative recovery. Disadvantages are increased surgical exposure, and increased difficulty in removing the devices if total knee replacement is required.

Surgical exposure of the proximal tibia for tibial osteotomy should take into consideration the future possibilities of additional procedures, especially total knee arthroplasty, which is usually performed through a midline incision. A slightly longer near-midline incision can be used for exposure of the proximal lateral tibia taking this into consideration. Depending on the surgeon's preference, other incisions can be considered without compromise of skin for future surgery. Either a transverse incision made laterally 2 cm below the joint line, or a longitudinal incision made slightly anterior to the fibula and continued proximally parallel to the thigh have been described for proximal tibial osteotomy.

The ideal patient for a proximal tibial osteotomy should be younger than 60, may be heavier than his ideal weight, and have failed other conservative measures. Range of motion is important and a flexion contracture greater than 15 degrees or flexion less than 90 degrees is less than ideal for the knee for intended osteotomy. Lateral subluxation is a relative contraindication to osteotomy. If the varus deformity is greater than 15 degrees a barrel vault type of osteotomy as described by Maquet may be necessary for correction.

Preoperative planning is germane to all osteotomies for the correction of tibial–femoral alignment or mechanical axis. Three-foot long-leg films are ideal for making computations for intraoperative surgical correction. These films may require the presence of the surgeon to assure their accuracy as variations in internal and external rotation may respectively accentuate valgus and varus deformity. In the varus knee, correction to 9 to 10 degrees of anatomic valgus is recommended—either that or a mechanical axis with 2 to 4 degrees of valgus beyond a neutral mechanical axis. In the valgus knee, correction to a neutral mechanical axis plus 4 degrees, or an

anatomic axis of 0 degrees, is felt to be ideal. Slight overcorrection is superior to undercorrection, especially in the correction of the varus knee (Fig. 12, Fig. 13).



Fig. 12. Standing preoperative alignment tibiofemoral angle of 13 degrees varus.



Fig. 13. Standing postoperative alignment tibiofemoral angle of 9 degrees valgus after valgus osteotomy.

Complications of osteotomies are usually technical. The proximal fragment should be 2 cm below the joint line and parallel to the joint to prevent fractures into the joint and non-union. Undercorrection, overcorrection, and recurrence of deformity are not uncommon findings in the postoperative course. Peroneal nerve palsy, non-union, infections, deep vein thrombosis, and compartment syndrome are all complications of this procedure.

The results of valgus osteotomy can be misleading. They may be excellent in 80 to 90 per cent of patients at 5 years but a decline in satisfaction and function occurs by 10 years. Slight overcorrection of the mechanical axis is associated with better long-term results. Patients with high adduction moments have been identified as at risk for early recurrence of the varus deformity and an unfavorable long-term response to this procedure. Potential compromise of subsequent arthroplasty is a consideration for the older patient and with low failure rates reported in total knee arthroplasty, a more prudent choice would be to avoid osteotomy even for those patients with unicompartmental disease. Satisfactory results of total knee arthroplasty after osteotomy in the majority of patients is comparable with those patients having only an index arthroplasty. The number of excellent results is lower and the procedure can be technically more difficult.

Varus osteotomy may be performed on the proximal tibia or distal femur. Indications for varus osteotomy on the tibia are limited to those deformities less than 12 degrees. The joint line will be altered and tilted medially after varus osteotomy of the tibia with a tendency for subluxation of the knee. An additional disadvantage of performing the varus osteotomy through the proximal tibia is resulting relaxation of the medial collateral ligament which requires plication to avoid medial laxity.

Supracondylar varus osteotomy of the distal femur (Fig. 14, Fig. 15) to an axial axis of 0 degrees may be performed with either a medial or lateral closing wedge and stabilization with an AO-ASIF 90 degrees osteotomy plate anticipating satisfactory results in 90 per cent of the patients with a diagnosis of osteoarthritis. The complication rate is similar to that of proximal tibial osteotomy.



Fig. 14. Standing tibiofemoral angle of 17 degrees valgus.

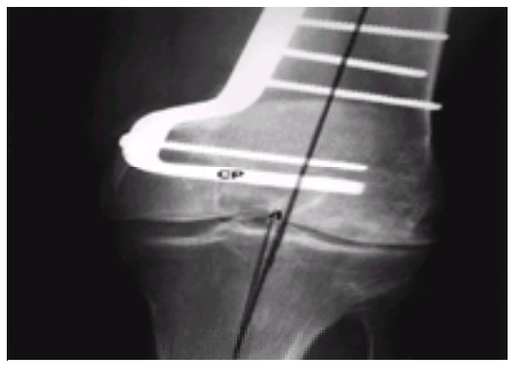


Fig. 15. Standing tibiofemoral angle of 4 degrees valgus postoperative after distal femoral varus osteotomy.

Arthroplasty

The major indication for an operative resurfacing procedure of the knee, either hemiarthroplasty, unicompartmental arthroplasty, arthroplasty, or total knee arthroplasty, is disabling pain. The most conservative of these procedures is the hemiarthroplasty. Indications for this procedure are disease isolated to one compartment, passively corrected angular deformity, and a young patient with post-traumatic arthritis or obesity. Results are not always predictable; the rate of revision is 15 to 20 per cent at 8

years.

Unicompartmental replacement is an excellent procedure in the well selected patient. Ideally, the patient should be sedentary, not overweight, and have an angular deformity less than 15 degrees, non-inflammatory arthritis being restricted to one compartment of the knee. Although patients with osteoarthritis can manifest inflammatory characteristics in their disease, the initiating event to the pathologic process of osteoarthritis is abnormal loading of the articular cartilage beyond its yield point with deformation and loss of matrix. An inflammatory response can ensue as a result of degradative enzymes released by cellular elements of the knee. This distinction should be made from knees with inflammatory or crystalline arthropathy as unicompartmental arthroplasty performed in these conditions will fail with the presence of remaining articular cartilage subject to an ongoing pathologic process.

The McKeever hemiarthroplasty is used without cement. Unicompartmental replacement is available in a cemented or uncemented design with experience limited in the uncemented designs and long-term results of uncemented unicompartmental knee arthroplasties poor. The initial Brigham unicompartmental replacement demonstrated good to excellent results in approximately 90 per cent of patients with a failure rate of only 1 per cent per year. Marmor in an 11-year follow-up had survivorship of 91.4 per cent. Additional reported series of unicompartmental replacements in the literature are available with wide-ranging results. Individual surgeons experienced in unicompartmental replacement continue to be enthusiastic about the results of the procedure. Failures in the past could usually be related to errors in surgical technique, patient selection, or component design. With this knowledge gained, continued use of unicondylar arthroplasty is favorable in the well-selected patient.

The advantages of unicompartmental arthroplasty over total knee arthroplasty is preservation of bone, the cruciate ligaments, and the patellofemoral articulation. Patients recover from the surgery rapidly and with preservation of anatomy, proprioception, and kinematics. Postoperative recovery and complication rate is lower for unicompartmental knee arthroplasty compared with osteotomy. Full weight-bearing and functional motion are present within a month after surgery. The short-term complication rate is much less after a unicompartmental arthroplasty, compared with a total knee arthroplasty. If failure occurs, total knee arthroplasty may be performed readily without undue technical difficulty. Failure of a unicompartmental arthroplasty is caused by aseptic loosening of the prosthesis, tibial subluxation, patellar impingement on the femoral component, component breakage, progression of disease in other compartments of the knee, and polyethylene failure (Fig. 16).

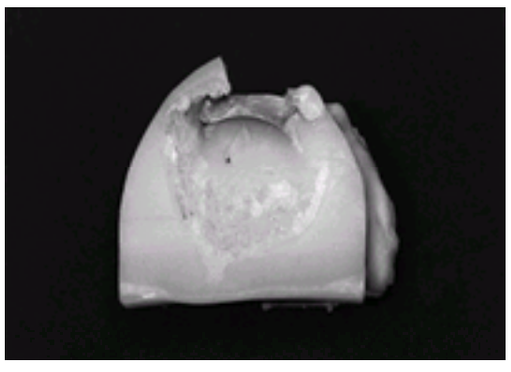


Fig. 16. Retrieval of failed metal-backed tibial component of a unicompartmental knee arthroplasty demonstrating wear of polyethylene to the metal backing.

Total knee arthroplasty

For the patient with disabling bicompartment or tricompartment arthritis, total knee arthroplasty, with improved design, instrumentation, and technique, has become a predictable procedure with satisfactory long-term results. Controversies exist in several aspects of the surgery, but it is agreed that ideal postoperative alignment and proper soft tissue management are paramount to the success of total knee arthroplasty.

Optimal goals for postoperative axial alignment are the mechanical axis which passes through the center of the knee with a tibial femoral angle between 5 and 9 degrees of valgus. Sagittal alignment on the femur should be anatomic (plus or minus 2 degrees) and on the tibia, anatomic or a posterior tilt is ideal.

Instrument systems primarily depend on an intramedullary femoral guide and extramedullary tibial guide for reproducible component position. There have been advocates of extramedullary femoral guides, but accuracy is less using either the femoral head or palpable shaft as a guide. Intramedullary tibial guides are attractive but extramedullary tibial guides have documented accuracy and are used predominantly.

Fixation has been controversial in recent years but the trend has been to use cement fixation. Low failure rates with a well designed and stable total knee arthroplasty have been documented in clinical studies with posterior cruciate sacrificing, posterior substituting, and posterior cruciate retaining condylar implants. Survivorship of cemented total knee arthroplasty beyond 10 years is approximately 97 per cent regardless of implant design.

There continues to be interest in uncemented fixation of total knee arthroplasty. Arguments for biologic fixation in knee arthroplasty are difficult to justify. Long-term studies have not been available to justify the increased cost of the prosthesis and associated problems that exist with their design. Retrieval studies quantifying bone ingrowth have been even in stable implants less than 20 per cent.

Wear of polyethylene of the tibial and patella components have been troubling. Osteolysis does occur as a result of polyethylene, metal, and cement particulate debris (Fig. 17, Fig. 18). Increasing the conformity of the articulation and the thickness of polyethylene can reduce high stresses in the polyethylene.



Fig. 17. Axial view of patella and femoral components of an uncemented arthroplasty. Osteolysis is adjacent to the patella component.

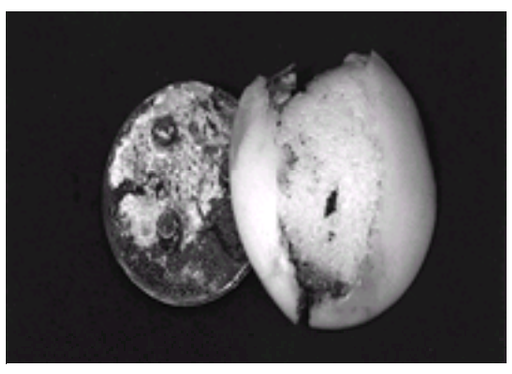


Fig. 18. Retrieval of failed well fixed failed patella component of Fig. 17.

Bone defects encountered at the time of arthroplasty may be reconstructed with autogenous bone, metal wedges, or a custom implant. Biomechanically, the custom implant is superior to the other choices but its intraoperative use is difficult to always predict, even with CAD-CAM designs. Incorporation of bone graft in peripheral defects has a success rate of 90 per cent and its use is recommended when the peripheral defect involves 50 per cent or more of the bony support of either tibial plateau and the defect is greater than 5 mm in depth.

Patellar resurfacing is recommended in the presence of inflammatory arthritis but is controversial in osteoarthritis. Patellofemoral problems are the most frequent source of complications in total knee arthroplasty. Their occurrence ranges from 1.5 to 10 per cent and they include instability, fracture of the patella, wear, patella ligament rupture, and soft tissue impingement. Maintenance of quadriceps function and avoidance of infection are key factors in the management of these problems.

Failure of total knee replacement when indicated, is treated with revision surgery. Careful preoperative evaluation of soft tissue, presence of infection, stability, and bone loss are important in the planning of revision total knee arthroplasty (Fig. 19). Complications are similar to primary arthroplasty with a higher infection rate, wound problems, instability, and extensor mechanism problems. The satisfactory results for revision total knee arthroplasty, with limited follow-up, are approximately 70 per cent.

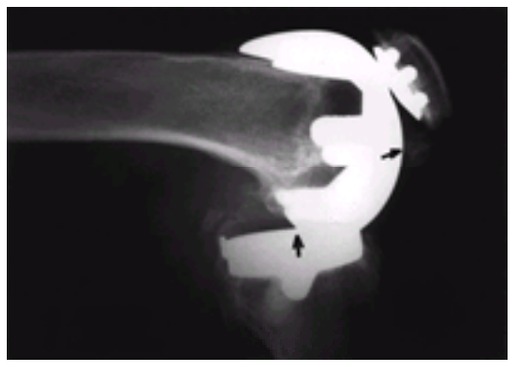


Fig. 19. Lateral view of knee after total knee arthroplasty with posterior instability and fracture of inferior pole of patella.

Infection after total knee arthroplasty is between 1 and 2 per cent. Perioperative antibiotics and clean air operating rooms have made the incidence of this complication low. Increased infection rates are associated with revision surgery, the use of a hinge prosthesis, immunocompromised patients, and patients with chronic skin ulcerations on their lower extremities. Treatment of infections in the immediate postoperative period is with early aggressive surgical debridement and antibiotics; treatment is successful in 90 per cent of patients. Chronic infections are treated most effectively by two-stage reimplantation. An antibiotic impregnated cement spacer has gained popularity as an adjunct to the management of the infection. Retraction of soft tissues is prevented to some degree with the use of the spacer and enough stability is achieved to allow ambulation (Fig. 20). Successful treatment of prosthetic infections is more likely in gram-positive infections. Reimplantation should be considered on an individual basis after initial aggressive debridement and an adequate course of appropriate antibiotics.



Fig. 20. Anteroposterior and lateral view of a knee with a temporary antibiotic loaded prosthesis.

The diagnosis of prosthetic joint infection is not always so definite. On clinical grounds painful swelling of the knee may be present. Plain radiographs may demonstrate osteitis. Three-phase bone scan, and elevated erythrocyte sedimentation rate, and aspiration have a sensitivity of between 33 and 67 per cent and a specificity of between 65 and 96 per cent. Intraoperative cell count of fluid with a white blood cell count greater than 25 000 and a frozen section with polymorphonuclear cells greater than five per high power field are strongly indicative of infection.

Surgical alternatives to reimplantation are arthrodesis, resection arthroplasty, and rarely amputation. Suppressive antibiotics alone are used in patients who are not surgical candidates. The major disadvantage to resection arthroplasty is a high incidence of significant pain and functional limitation. In view of this, treatment by pseudoarthrosis is limited to patients with limited ambulatory status. Arthrodesis can provide a stable and painless extremity, although fusion rates are variable. Removal of a constrained prosthesis with associated bone loss has a 56 per cent rate of fusion. Less constrained prostheses have a higher fusion rate of approximately 80 per cent.

Peroneal nerve palsy is a rare complication following total knee arthroplasty. Severe valgus and flexion deformities have a greater likelihood of having this complication. Prophylactic decompression of the peroneal nerve does not prevent its occurrence postoperatively.

Supracondylar fracture of the femur following total knee arthroplasty has an incidence approaching 2.5 per cent. This complication is associated with osteoporosis and notching of the femur. Improved instrumentation, an increased inventory of sizes of prosthesis, and alertness of the surgeon should decrease notching of the femur. Periprosthetic supracondylar fracture of the femur is optimally treated to correct significant displacement and instability associated with the fracture. Methods of fixation are individualized by fracture pattern and type of prosthesis. Rigid plate fixation has a high success rate, but intramedullary fixation has gained popularity with the availability of retrograde rods. Careful preoperative planning to stabilize the fracture is mandatory for a good outcome.

The incidence of thromboembolic disease is common after surgery on the lower extremity. Deep vein thrombosis without any treatment after total knee arthroplasty is between 40 and 80 per cent. Proximal or thigh deep vein thrombosis is between 9 and 20 per cent. The incidence of pulmonary embolism is 1.8 to 7 per cent and fatal pulmonary embolism is 0.7 per cent. Therapeutic choices for treatment which have proven to be efficacious include external pneumatic compression stockings worn continuously except while ambulatory and in the high-risk patient an inferior vena caval filter. Drug treatments proven to be efficacious are coumadin or low molecular weight heparin. In prospective studies comparing these two drugs, low molecular weight heparin with a twice daily dosage proves to be superior in reducing the incidence of venogram-proven deep vein thrombosis over warfarin. Risks of bleeding complications were not significantly different in the two groups. The length of prophylaxis following surgery has not been clearly defined.

Arthrodesis

Improving longevity of total knee arthroplasty and the limited life of an arthrodesis has made arthrodesis of the knee as a primary procedure uncommon. The most common indication for arthrodesis of the knee is septic failure of a total knee replacement. Additional indications for arthrodesis include post-traumatic arthritis in the

young, painful ankylosis, neuromuscular disease, and the neuropathic joint. Arthrodesis in the neuropathic joint has a high failure rate, and in selected cases, total knee arthroplasty has been successful. Stabilization of the knee with a hinged knee brace may be effective but if significant recurvatum is present, arthrodesis may be an effective consideration for surgical management. Surgical methods of arthrodesis include external, intramedullary, and plate fixation.

Complications of arthrodesis of the knee include non-union, recurrent sepsis, fracture, and patient dissatisfaction. Non-union following attempts at arthrodesis of the knee may be as high as 50 per cent if significant bone loss is present. The development of a fibrous ankylosis may not always be painful, but if pain is present, a secondary procedure will be required. Persistent infection can result even with achieving fusion, after septic failure of total knee arthroplasty. Dissatisfaction with knee arthrodesis is high and conversion to total knee replacement has not been recommended in the past. Recent reports of conversion to total knee arthroplasty have shown favorable outcomes.

Patellofemoral joint

Disorders of the patellofemoral joint are manifested by either pain, instability, or a combination of both. Patellofemoral arthralgia may be the result of chondromalacia patellae, abnormalities of patella alignment, or pathology of the retinaculum.

Abnormalities of the cartilage have been graded in a number of different classification systems. A new system of describing articular cartilage abnormalities is based on a description of the cartilage, the depth of involvement, the diameter, and the location of the lesion. Pathologic change of the patellofemoral joint is a result of repeated loading of the cartilage beyond its ability to withstand deformation. This is usually the result of abnormal joint mechanics involving the patellofemoral articulation. Pain can be extremely disabling with associated functional limitations. Instability patterns of the patellofemoral joint result in lateral subluxation or dislocation of the patella. Chronic instability may occur as a result of an acute traumatic dislocation, rotational malalignment to the lower extremity, ligamentous laxity, excessive quadriceps angle, patella alta, neuromuscular disorders, and either patella or femoral anatomic variation.

Evaluation of the patient with patellofemoral pain and instability should include in the physical examination, concentration of the location of pain, the rotational and axial alignment of the extremity, patellar mobility, and pronation of the foot. Radiographic evaluation should include either a Lauren or Merchant axial view and supplementary computed tomography or magnetic resonance imaging scanning with midpatella views at 0, 10, 20, and 30 degrees are helpful in diagnosing subtleties of malalignment.

Treatment initially is usually non-operative emphasizing rehabilitation, strengthening of the quadriceps, and stretching of the hamstrings, iliotibial band, and retinaculum. Elastic knee supports, orthotics, and anti-inflammatory medication are additionally helpful for the symptomatic patient. Operative treatment consists of realignment procedures for malalignment disorders.

Further reading

ACCP Consensus Conference on Antithrombotic Therapy. *Chest* 1995; **108**: 312–34. [All studies demonstrate low molecular weight heparin to be superior to low-dose warfarin in efficacy with a twofold increase in the risk of bleeding complications.]

Altcheck D, Sculco TP. Autogenous bone grafting for severe deformity in total knee arthroplasty. *Journal of Arthroplasty* 1989; **4**: 151–5. [A clinical review of the outcomes of bone grafting with severe angular deformity in total knee arthroplasty. At 4-year follow-up the autogenous grafts used on the tibia head healed without evidence of component subsidence.]

Angari S, Worwick P. Incidence of fatal pulmonary embolism after 1390 knee arthroplasties without routine anticoagulation except in high risk cases. *Journal of Arthroplasty* 1997; **12**: 599–602. [A consecutive series of 1600 arthroplasties were performed from 1980 to 1994. Autopsy confirmed pulmonary embolism as a cause of death in two patients (0.22 per cent). The authors concluded that the risk of clinical non-fatal thromboembolic disease was low and the indication for routine prophylaxy was not warranted after total knee arthroplasty.]

Bartel DL, Bicknell LL, Wright TM. The effect of conformity, thickness, and material on stresses in UHMW polyethylene components for total joint replacement. *Journal of Bone and Joint Surgery* 1986; **68A**: 1041–51. [An analysis of contact stresses as a function of thickness of the polyethylene insert for tibial components. Recommendation of more than 8 mm whenever possible. Stresses were reduced when the articulating surface was more conforming in medial-lateral dimension. Maximum shear stress in the tibial polyethylene was subsurface.]

Bert JM, Maschka K. The arthroscopic treatment of unicompartmental gonarthrosis: A five year follow-up study of abrasion arthroplasty plus arthroscopic debridement and arthroscopic debridement alone. *Arthroscopy* 1989; **5**: 25–32. [A retrospective review of 126 patients with a mean follow-up at 5 years. The results of abrasion arthroplasty were inferior to debridement alone.]

Broebaum RD, Rubman MH, Hofman AA. Bone ingrowth into porous coated tibial components implanted with autograft bone chips. *Journal of Arthroplasty* 1992; **7**: 483–93. [Retrieval study of 10 consecutive same design knees with the use of bone surgery placed at intervals. The time from implantation to retrieval was 1 week to 48 months. The ingrowth was progressive over that time from 8 to 20 per cent.]

Callahan CM, Drake BG, Heck DA. Patient outcomes following tricompartmental total knee replacement. *Journal of the American Medical Association* **271**(17): 1349–1357. [A literature review of 130 studies of total knee arthroplasty. The total number of patients was 9879. A percentage of 89.3 patients showed good to excellent results. The revision rate was 3.6 per cent at 4 years. Knees done with posterior cruciate ligament retaining total knee arthroplasty had fewer complications.]

Conventry MB, Ilstrup PM. Proximal tibial osteotomy: A critical long term study of 87 cases. *Journal of Bone and Joint Surgery* 1993; **75A**: 196–201. [A 10-year follow-up of proximal tibial osteotomies. A survival rate of the procedure was 94 per cent at 5 years and at 10 years the correction was 8 per cent or greater and if the patient was not overweight.]

Diduch DR, Insall JN, Scott N, Scuderi GR. Total knee replacement in young active adults. *Journal of Bone and Joint Surgery* 1997; **79A**: 575–82. Long-term follow-up with a range of 3 to 18 years (average 8). [One hundred and fourteen replacements in patients less than 55. Cemented posterior cruciate ligament —sacrifice 94 per cent good to excellent results and survival 94 per cent at 18 years for the femoral and tibial components.]

Feldman DS, Lonner JH, Desai P, Zuckerman JD. The role of intraoperative frozen section in revision total joint arthroplasty. *Journal of Bone and Joint Surgery* 1995; **77A**: 1807–13. [Intraoperative frozen section of periprosthetic tissue with more than five polymorphonuclear cells per high-power field is a reliable indicator of infection during revision joint arthroplasty.]

Font-Rodriguez DE, Scuderi GR, Insall JN. Survivorship of cemented total knee arthroplasty. *Clinical Orthopaedics and Relative Research* 1997; **345**: 79–86. [The survivorship analysis of 2629 cemented primary total knee arthroplasty, a posterior cruciate ligament sacrifice implant, over a 22-year period. With a metal-backed tibial component the annual failure rate was:14 per cent and a 14-year success rate of 98.1 per cent. Overall the success rate was not influenced by gender, age, diagnosis, or percentage of ideal body weight.]

Grelsomer RP. Unicompartmental osteoarthritis of the knee. *Journal of Bone and Joint Surgery* 1995; **77A**: 278–92. [Current concepts review of unicompartmental arthritis. Excellent review of the pathophysiology, mechanics, natural history, and treatment of this disorder.]

Hanson, A, Rand J. The treatment of infected total knee arthroplasty with insertion of another prosthesis. *Clinical Orthopaedics* 1994; **309**: 44–55. [Eighty-six patients with 89 infected total knee arthroplasty were reviewed. The use of antibiotics in the polymethyl methacrylate at the time of re-implantation had a lower reoccurrence rate at 4 years (28 per cent versus 4.7 per cent).]

Healy WL, Auglin JO. Distal femoral versus osteotomy. *Journal of Bone and Joint Surgery* 1988; **70A**: 102–09. [Twenty-three knees with valgus angulation underwent distal femoral varus osteotomy with a 90-degree blade plate. Three of four knees with rheumatoid arthritis failed. Results at 4 years in patients with osteoarthritis were 93 per cent good to excellent.]

Healy WL, Siliski JM. Operative treatment of distal femoral fracture proximal total knee replacement. *Journal of Bone and Joint Surgery* 1993; **75A**: 27–34. [Syndracondylar fracture after total knee arthroplasty are associated with osteopenic bone and surgical treatment is recommended for displaced and unstable fractures.]

Heck DA, Marmor L. Unicompartmental knee arthroplasty: A multicenter investigation. *Clinical Orthopaedics and Related Research* 1993; **286**: 154–59. [A review of 245 unicompartmental replacement at 10 years demonstrated a survivorship of 91.4 per cent. The non-metal backed tibial components were used. Survivorship was equal between medial and lateral compartment unicompartmental knee arthroplasty. Obesity was associated with increased failure.]

Jacobs MA, Hungerford DS, Krackow KA. Revision TKA for aseptic failure. *Clinical Orthopaedics Related Research* 1988; **226**: 78–85. [Twenty-four patients were reviewed at 2 to 4 years postoperatively. Eighty-three per cent of the patients with defined mechanical problems achieved good to excellent results. Those patients revised for pain or without a clear objective problem were not improved.]

Kim H, Pelker R. Rollback in PCL retaining total knee arthroplasty. *Journal of Arthroplasty* 1997; **12**(5): 553–61. [A radiographic analysis of 49 posterior cruciate ligament retaining total knee arthroplasty in 38 patients, which demonstrated that there is no consistent femoral rollback with knee flexion.]

Levitsky KA, Hozack WJ. Evaluation of the painful prosthetic joint: Relative value of bone scan, ESR, and joint aspiration. *Journal of Arthroplasty* 1991; **6**: 237–44. [Seventy-two joint arthroplasties were studied prospectively. Three-phase bone scan had a sensitivity of 33 per cent and a specificity of 86 per cent. Using 30 as the limit for infection, erythrocyte sedimentation rate had a sensitivity of 60 per cent and a specificity of 65 per cent. Preoperative aspiration had a sensitivity of 67 per cent and a specificity of 96 per cent.]

Lieberman J, Sung R. Low dose warfarin prophylaxis to prevent symptomatic pulmonary embolism after total knee arthroplasty. *Journal of Arthroplasty* 1997; **12**(2): 180–4. [Between 1984 and 1993, 815 primary and revised total knee arthroplasties received low-dose warfarin. There were three patients with symptomatic pulmonary embolism (0.3 per cent); eight patients with symptomatic deep vein thrombosis (1 per cent). No deaths were attributed to pulmonary embolism. Seventeen patients developed hematoma (2.5 per cent) and two of these required draining of the knee.]

Lotke PA, Ecker ML. Osteonecrosis of the knee. *Journal of Bone and Joint Surgery* 1988; **70A**: 470–3. [Current concepts and review article on osteonecrosis of the knee.]

Mathews LS, Goreskin SA. Proximal tibial osteotomy: Factors that influence the duration of satisfactory function. *Clinical Orthopaedics Related Research* 1988; **229**: 193–200. [A retrospective review of 40 patients treated with proximal tibial osteotomy. Factors which led to failure included preoperative obesity, older age, severe axial deformity. Good to excellent results were 86 per cent at 1 year, 50

per cent at 5 years, and 28 per cent at 9 years.]

Mont MA, Waldman B. Multiple irrigations, debridement, and retention of components in infected total knee arthroplasty. *Journal of Arthroplasty* 1997; **12**: 426–433. [A prospective study of 24 infected total knee arthroplasties treated by irrigation, debridement, and retention of components. Patients included in the study were either less than 30 days postoperative or in the hematogenous group were symptomatic for less than 30 days. Treatment was irrigation and debridement and 6 weeks parenteral antibiotics. At follow-up of 48 months there was no reoccurrence in the acute surgical infection group and 10 reoccurrence (71 per cent) in the hematogenous group.]

Moseley JB, Wray NP, Kaykendau D. Arthroscopic treatment of osteoarthritis of the knee: A prospective, randomized, placebo-controlled trial. *American Journal of Sports Medicine* 1996; **24**: 28–34. [A study of a small group of patients randomized and double blinded, which highlights the possibility of a placebo effect from arthroscopic debridement for osteoarthritis.]

Noyes FR, Stabler CL. A system of grading articular cartilage lesions at arthroscopy. *American Journal of Sports Medicine* 1989; **17**(4): 505–13. [A new grading system of articular cartilage damage is described addressing four variables: a description, the depth of involvement, the diameter, and location of the lesion.]

Olgivie-Harris DJ, Mclean J. Pigmented villonodular synovitis of the knee: the results of total arthroscopic SYnorectomy, partial arthroscopic synorectomy, and arthroscopic local excision. *Journal of Bone and Joint Surgery* 1992; **74A**: 119–23. [Twenty-five patients with pigmented villanodular synovitis were followed after arthroscopic management. No reoccurrences were identified in the patient with localized lesions. Diffuse pigmented villanodular synovitis was associated with reoccurrences.]

Rand JA. Role of arthroscopy in osteoarthritis of the knee. *Journal of Arthroscopic and Related Surgery* 1991; **7**(4): 358–63. [Two groups of patients. Group I: debridement, removal of loose bodies and partial meniscectomy; 80 per cent improvement at 1 year, 67 per cent at 5 years. Group II: in addition had an abrasion arthroplasty; 40 per cent improvement at 1 year and 50 per cent underwent total knee arthroplasty at a mean of 3 years.]

Rand JA, Ilstrup DM. Survivorship of total knee arthroplasty. *Journal of Bone and Joint Surgery* 1991; **73A**: 397–409. [Cumulative rates of survival of 9200 total knee arthroplasty. Survivorship at 10 years was 97 per cent. Failure lowest in primary total knee arthroplasty, diagnosis of rheumatoid arthritis, age greater than 60, and condylar prosthesis with a metal-backed tibia.]

RD Heparin Arthroplasty Group. RD heparin compared with warfarin for prevention of venous TED following total hip and knee arthroplasty. *Journal of Bone and Joint Surgery* 1994; **76**: 1174–85. [Low molecular weight heparin (twice daily dosage) and warfarin were compared in total hip arthroplasty and total knee arthroplasty patients. In total knee arthroplasty patients there was a significant reduction in thromboembolic disease and no statistical increased incidence in bleeding complications with the use of low molecular weight heparin.]

Ritter MA, Herbst SA, Keating EM. Long term survival analysis of a posterior cruciate-retaining total condylar knee arthroplasty. *Clinical Orthopaedic Related Research* 1994; **309**: 136–45. [A clinical study retrospectively reviewing the result of posterior cruciate ligament- retaining condylar knee with a minimum follow-up of 10 years. All failure were of the tibial component (7). Survival rate at 12 years was 97 per cent.]

Rorabeck CH, Bourne RB, Nott L. The cemented Kinematic-II and the non-cemented PCA for TKA: A prospective evaluation. *Journal of Bone and Joint Surgery* 1988; **70A**: 483–90. [A prospective evaluation of a total of 160 total knee arthroplasties. Minimum follow-up of 2 years showed that the cemented knees had higher knee scores and a better range of motion.]

Scott RD, Cobb AG. Unicompartmental knee arthroplasty: 8–12 year follow-up with survivorship analysis. *Clinical Orthopaedics and Related Research* 1991; **271**: 96–100. [The results of 100 consecutive unicompartmental knee arthroplasties were reviewed. 87 per cent of patients had no pain. Average flexion was 115 degrees. Revision had been performed on 13. Survivorship analysis was 85 per cent at 10 years.]

Sumner DR, Kienapfer H, Jacobs J. Bone ingrowth and wear debris in well fixed cementless porous coated tibial components removed from patients. *Journal of Arthroplasty* 1995; **10**: 1–11. [A study of 13 primary porous total knee arthroplasty components retrieved from 3 to 30 months postoperatively had ingrowth of 2–18 per cent.]

Wilson MG, Kelly K, Thornhill TS. Infection as a complication of TKA risk factors and treatment in 67 cases. *Journal of Bone and Joint Surgery* 1990; **72A**: 878–83. [The incidence of infection at the Brigham and Women's Hospital was 1.6 per cent. The failure of treatment of prosthetic joint infection by debridement, prosthesis retention, and antibiotics was 45 per cent. *Staphylococcal aureus* was associated with the failure of treatment.]

Windsor RE, Bono JV. Infected total knee replacements. *Journal of the American Academy of Orthopaedic Surgeons* 1994; **2**(1): 44–53. [A review article on the incidence, diagnosis, and treatment; and the results of the management of this devastating problem. The most successful results follow a two-stage re-implantation with a cure rate average of 90 per cent.]

46.2.6 Degenerative disorders of the spine

Philip R. Lucas

Introduction
Degenerative changes of the lumbar spine
The clinical categories
Disorders of the cervical spine
Mechanical neck pain
Further reading

Introduction

The spine or axial skeleton begins to undergo degenerative changes soon after reaching skeletal maturity. This process, the conversion or replacement of normal tissue, is a universal one that occurs at a variable rate in different individuals. Nowhere is the degeneration better seen than in the intervertebral disc. The disc is the largest avascular structure in the body and functions, in the simplest sense, as a shock absorber. Many authorities feel that changes within the disc structure begin the process of spinal degeneration.

The intervertebral disc is composed of an outer ring or anulus and a central portion, the nucleus pulposus; these structures are sandwiched between two hyaline cartilage end-plates (Fig. 1). The anulus is made up predominantly of coarse collagen fibers arranged in parallel sheets that interconnect adjacent vertebral bodies. The nucleus is a gelatinous central core made up of loosely bound collagen and an abundant ground substance. Distributed throughout the disc are chondrocytes, which produce the extracellular matrix. The extracellular substance is a combination of collagen, proteoglycans, and water. The collagen is responsible for the strength of the disc. The proteoglycan molecule is hydrophilic and functions by keeping the fibrous network pressurized with water, so allowing the disc to resist and distribute compressive forces.



Fig. 1. Spine: relation of disc to neurostructures.

The disc is avascular and its nutrition is supplied by diffusion from the ligaments that surround it and the vascular subchondral bone. With aging, there is a decrease in the rate of nutritional diffusion, which interferes with the SYnthesis of collagen and proteoglycans, and results in desiccation of the disc. As the disc dries out, its efficiency in resisting mechanical loads begins to fail. The degenerative process within the disc is thus the result of a combination of mechanical forces superimposed upon the biochemical changes. As the disc fails to function as a shock absorber, increased forces are placed across the posterior facet joint, resulting in a degenerative process. Tears in the anulus may allow nuclear material to herniate through it, causing nerve compression. The degenerative process eventually results in structural changes within the spine, a combination of deterioration and remodeling.

Not all individuals with structural changes appear to be symptomatic, and there are many with considerable pain and disability who have little structural change on radiographs. For these individuals, the cause of their neck and back pain remains an enigma. In recent years, researchers have found numerous pain receptors throughout the structures of the spine. In addition a wide variety of chemical pain mediators have been isolated, including glutamate, bradykinin, histamine, nitric oxide, cytokinin, and substance P. These mediators are found in increased concentration throughout the disc and neural elements in symptomatic patients. They might be the missing piece in the puzzle to explain neck and back pain.

Degenerative changes of the lumbar spine

Degenerative changes within the lumbar spine produce a wide variety of clinical symptoms, but the clinical presentation of patients with back pain can be placed into one of four general categories, based upon clinical history, physical findings, and diagnostic studies. Classification is useful in determining the diagnosis, natural history, and treatment. As the degenerative process is continuous the categories reflect this and there is much overlap between them. In attempting to categorize patients with back pain, it is imperative to obtain a good history and physical examination; in spite of the many diagnostic tests available, these two still the most important in establishing a diagnosis. While conducting the systematic history and physical examination, it is important to be aware of key points or findings. Certain ‘red flags' alert the clinician to the possibility of back pain caused other than by the degenerative process (Table 1). Pain at rest and at night is generally not mechanical pain: one must consider either an infectious process, such as an epidural abscess or osteomyelitis, or a metastatic disease. If a patient presents with pain at rest, it would suggest a visceral cause such as a renal calculus or an intra-abdominal process, possibly an abdominal aneurysm. If the patient has a frank neurologic deficit, urinary retention or incontinence, one must consider the possibility of a space-occupying lesion within the spinal canal, either a very large herniated disc or an epidural abscess or neoplasm.

History of cancer
Trauma
Fever
Weight loss (unexplained)
Intravenous drug use
Immunosuppression
Pain at rest

Table 1 ‘Red flags' to watch for in the history

The clinical categories

Category I: acute low back pain

Clinical picture

The patient with acute low back pain rarely presents before the age of 20 and generally not after the age of 60 years. They give a history of twisting or lifting but generally the amount of trauma is minor. They localize their pain to the lower back with occasional radiation into the buttock but not below the knee. Any attempt at motion increases discomfort. There may be spasm within the paravertebral muscles. Neurological examination rarely shows any deficit in either sensory or motor function.

Diagnostic tests

Diagnostic tests are not indicated in patients with acute low back pain in the absence of ‘red flags' (see Table 1). Radiographs should be obtained in all patients who are over the age of 50 years or who have a history of significant trauma.

Pathophysiology

It is estimated that only 20 per cent of patients who present with acute low back pain can be given a precise pathoanatomic diagnosis. However, because some of those patients go on to develop symptoms of acute disc herniation, then a central herniation or tear in the outer anulus may certainly be the cause in many of them.

Treatment

The patient with acute onset of low back pain often has severe pain. Treatment most often involves reassurance. The syndrome is generally self-limiting. Over a 2-week period, about 70 per cent of patients show clinical improvement; 90 per cent will have improved within 2 months. Treatment should be directed with this natural history in mind. The majority of individuals with acute back pain respond to a short-term treatment program divided into three phases: a period of rest, progressive mobilization, and then a preventative exercise program and education. Non-steroidal anti-inflammatory drugs or occasionally opioids may be necessary initially, as well as a muscle relaxant. There is a suggestion that early treatment with spinal manipulation may help to speed the recovery of some individuals.

Category II: chronic or recurrent low back pain

About 5 per cent of patients who develop low back pain will have symptoms lasting more than 3 months. Some will have recurrent bouts of back pain. There are several syndromes with a specific clinical picture and radiographic findings that can produce low back pain. These include:

1. posterior facet SYndrome;
2. degenerative spinal arthritis;
3. segmental instability;
4. internal derangement of the disc;
5. idiopathic vertebral sclerosis;
6. diffuse idiopathic skeletal hyperostosis;
7. ankylosing spondylitis;
8. fibromyalgia.

With the exception of the last three, all appear to be different manifestations of the same pathophysiology, that is, a degenerative process in the motion segment of the spine. The pain is produced by either mechanical or chemical irritation to the adjacent nociceptive nerve endings. Many of these individuals experience depression, most often as a result of the chronicity of the SYmptoms.

Clinical picture

A patient with chronic pain is typically middle-aged (30- to 60-year age group), and presents with chronic low back pain that may be cyclic or associated only with activity. The pain is typically across the back and may radiate to the upper portion of the lower extremities; it rarely radiates below the knee. Symptoms are typically aggravated by sitting and may gradually resolve with rest. Physical examination will show some restriction of spinal mobility and often an increase or aggravation of pain with extension. Neurological examination normally does not show any deficit in the lower extremities. In the facet syndrome, the patient will experience back and leg symptoms, and will often reproduce their symptoms on extension. Injection of isotonic saline into the facet joints may also reproduce their SYmptoms, which may be relieved by injection of a local anesthetic. The patient with internal derangement of a disc will often have mechanical pain; computed tomography (**CT**) or magnetic resonance imaging (**MRI**) may show a bulging of the anulus, with a dark disc demonstrated on the T_2-weighted MRI (Fig. 2). Provocative discography or injection of fluid into the disc under fluoroscopy may reproduce the symptoms (Fig. 3). Diagnostic tests for segmental instability include flexion–extension of the spine, which is positive when at least 3 mm of displacement is noted on moving from a flexed to extended position. The diagnosis may be confirmed by facet injection if this relieves the pain.



Fig. 2. MRI of dark disc, consistent with degenerative disease.

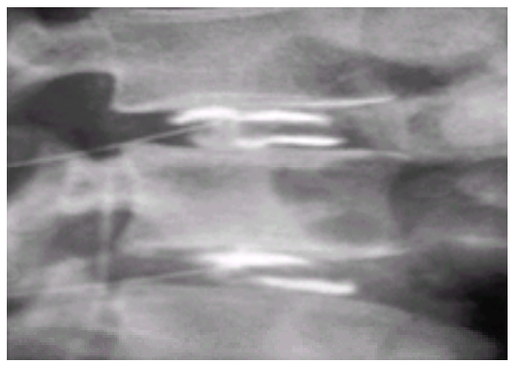


Fig. 3. Radiograph of a discogram.

Diffuse idiopathic spinal hyperostosis is a SYndrome that occurs predominantly in men. The patients typically complain of stiffness. Radiographs show considerable osteophyte formation anteriorly on the spine, with bridging of the vertebrae. Ankylosing spondylitis typically occurs in the male and in up to 2 per cent of the population. These patients often have considerable morning stiffness that tends to clear a little with activity and then become more symptomatic at the end of the day. Radiographs show diffuse ankylosis and the spine will have the typical appearance of ankylosing spondylitis. Idiopathic vertebral sclerosis occurs mostly in women: it produces considerable narrowing of the disc space and rather diffuse sclerosis at the adjacent vertebral end-plates, giving the appearance of a chronic infection of the disc space

without erosion of the end-plates.

Treatment

Patients with chronic recurrent back pain are often not treated surgically. Anti-inflammatory medication is helpful, as there is almost always some inflammatory reaction present. Exercise is important in improving mobility and in attempting to increase the dynamic stability of the spine, with attention directed to both the abdominal and paraspinal muscle groups. Aerobic exercise, such as walking and bicycle riding, and aquatic therapy, are to be encouraged. Aerobic exercise has been found to increase the amount of endorphins and thereby decrease pain. Since many of these patients will experience depression due to the longevity of their symptoms and disability, consultation with a psychiatrist or a psychologist and the use of antidepressant medication may be of great benefit.

The role of surgery Traditionally, surgical attempts to relieve chronic back pain have not been very satisfying. Surgery may be indicated if a painful segment can be clearly identified and the patient has failed in a comprehensive program of exercise and pain management. Localization of the pain generator is difficult. If a definite area of instability is identified and/or provocative discography or facet injection all point to one or two segments of the spine, surgery may be considered. In addition, one might consider surgical treatment if immobilization with a cast or brace gives temporary relief; this would involve a spinal fusion, which can be carried out either from an anterior or a posterior approach (Fig. 4), traditional transverse process fusion or, more recently, interbody fusion with or without the use of instrumentation for stabilization (Fig. 5). There have been few prospective studies of the clinical effectiveness of spinal fusion in patients with low back pain from either lumbar instability or internal derangement of the disc. Therefore, clear indications for surgical treatment of low back pain are still lacking.

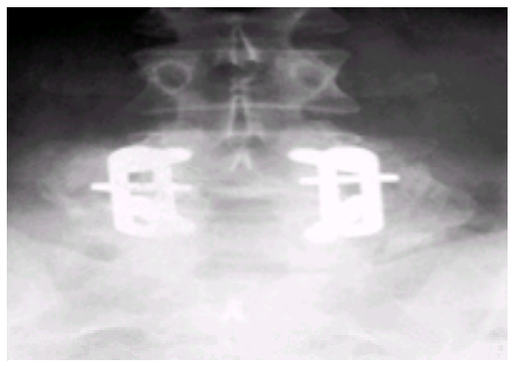


Fig. 4. Radiograph of posterior lateral fusion; posterior plate fixation at L 4–5.

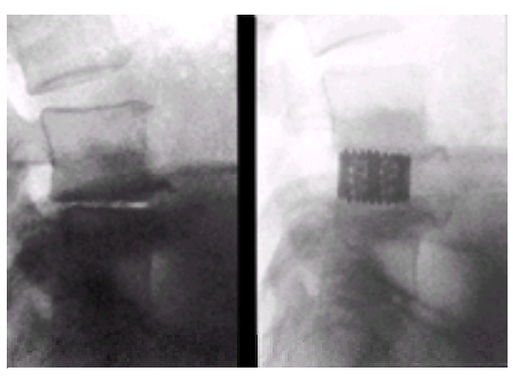


Fig. 5. Postoperative radiograph of interspinal-body fusion using titanium implant.

Category III: sciatica

Sciatica is a term used to describe pain of neurologic origin that radiates down the posterior aspect of the leg in the distribution of sciatic nerve. Besides bringing symptoms of paresthesia, weakness may also be present. Anterior thigh pain may also be of neurologic origin, with lesions affecting the upper lumbar nerve roots. Differential diagnoses of sciatica or lumbar radiculopathy include intraspinal mass proximal to the disc, with lesions affecting the conus or the cauda equina. Lesions at the disc level may include herniated disc (Fig. 6), spinal stenosis, discitis, or tumor. Radiculopathy may also be brought about by extraspinal masses such as a lesion within the pelvis, vascular disease, gynecologic conditions, inflammation, and arthritis about the hip joint. Systemic abnormalities producing a neuritis may also produce a sciatic picture. Lumbar radiculopathy of a sciatic or femoral nerve must be differentiated from referred pain. This pain is usually due to instability or inflammation of a facet joint. Referred pain does not radiate below the knee, may be bilateral, and is not associated with a neurologic deficit.



Fig. 6. Herniated intervertebral disc

Clinical picture of sciatica from a herniated disc

Patients with sciatica due to a herniated disc typically are in the 20- to 50-year age group and about half of them will attribute their symptoms to some incident of trauma. Generally, however, the trauma is of minor significance and the herniation is due to a progressive degenerative change within the disc and eventual rupture. Prior to developing leg symptoms, the majority of patients will develop some back pain, which is typically aggravated by forward bending, lifting, coughing, or sneezing. The pain will often seem to resolve somewhat in a supine position. Physical examination shows that the patient will often stand with flattening of the lumbar spine and will often list to one side to relieve nerve irritation. When standing, the patient will often not put the affected foot permanently on the ground. On examination of the spine, the major limitation is that of flexion; the patient may actually experience some relief of sciatic pain with extension. Examination of the lower extremity may show muscle atrophy if SYmptoms have lasted for some time. In addition, there is often a decreased sensation in the distribution corresponding to the nerve-root dermatome involved. Weakness of specific muscle groups may also be present, along with a decrease or absence of reflex. The most sensitive test in determining nerve-root irritation is that of the straight-leg raise test or the Lasegue sign. The bowstring test is often pathognomonic for nerve compression. When performing these tests, the patient is supine with the knee and hip flexed, and then gradually the knee is extended; this will often produce pain radiating down the leg but it may also radiate up towards the back. If symptoms are in the anterior thigh, you may suspect an upper lumbar disc: a femoral nerve stretch test should be performed, with the patient lying on the side or prone and gently extending the hip and flexing the knee, which will often reproduce the pain in the leg nerve as tension is brought about. If a leg raise is performed on the contralateral side and produces pain on the affected side, the probability that disc herniation is causing the problem rises to approx. 90 per cent and is referred to as a positive cross-leg raise. MRI has become the diagnostic study of choice for herniated disc (Fig. 7). Myelography had long been the ‘gold standard'

and can still be used, but MRI is non-invasive and clear reproduction of the disc in relation to the neural elements is possible. CT can also be employed but it not as sensitive as MRI. Electromyography can be used the localize the level of the nerve-root involvement, but this test gives no information on the exact site or nature of the lesion.



Fig. 7. MRI illustrating herniated intervertebral disc.

Treatment

Sciatica rarely requires immediate surgical treatment unless the symptoms are the result of a massive disc herniation, epidural abscess, or tumor. The other indications for immediate surgical treatment would be a marked neurologic deficit or pain that is uncontrollable with narcotics. Cauda equina syndrome, with sensory and motor deficit and loss of bladder function, occurs acutely and is also a surgical emergency. Half of the patients who present with acute sciatica will have complete recovery within 6 weeks. Initially, unless the patient has marked neurologic deficit, diagnostic studies are not indicated. About 10 per cent of patients with persistent underlying sciatica will require some form of invasive treatment. When one compares patients who have undergone surgery with those who have been treated conservatively, after 2 years, those who have undergone surgery for a herniated disc seem to have better function and fewer symptoms than the nonsurgically treated group. At 4 years, however, the two groups appear about the same with regard to symptoms. Unless the patient has significant, progressive neurologic deficit or severe intractable pain, it is reasonable to wait 3 to 4 weeks before performing diagnostic studies. Treatment initially should consist of limited activity and physical therapy. Epidural steroid injections decrease pain in some patients. Regardless of the treatment employed, the basic principle of nonoperative treatment is to allow the natural course to occur. In about 90 per cent of patients, symptoms will resolve without invasive treatment. If, in 4 to 6 weeks, symptoms have not resolved and diagnostic studies correlate with the clinical picture, invasive procedures are recommended. For treatment of disc herniation these include injection of chymopapain, percutaneous discectomy, microdiscectomy, and traditional discectomy. Which of these procedures is employed appears to be less important than a careful preoperative evaluation of the patient. In the VENN diagram (Fig. 8), the shaded area represents a patient who has a positive straight-leg raise together with objective neurologic and positive diagnostic studies; 90 to 95 per cent of these patients treated surgically will have an excellent result.

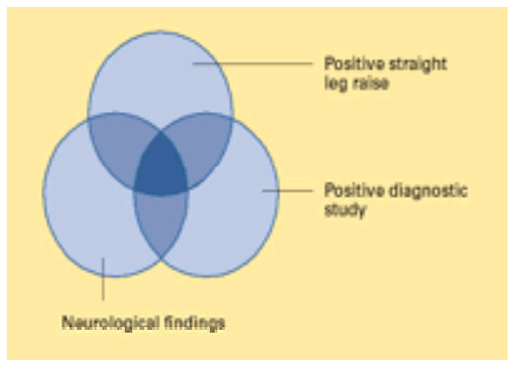


Fig. 8. VENN diagram representing evaluation of the patient with persistent, unrelenting sciatica.

Chemonucleolysis The procedure, which involves the injection of chymopapain enzyme into the disc, was first used in the 1970s and had a brief period of popularity in the early 1980s. Because of some early reports of apparent allergic reactions, it is now rarely used in the United States. It has continued to be used in other parts of the world and, recently, renewed interest has emerged. Studies have suggested that 75 to 80 per cent of patients improve with chymopapain injections. With identification of those who might be allergic to the enzyme, and better diagnostic techniques, it is possible that this procedure may become popular once again.

Percutaneous discectomy This method involves mechanical removal of the disc by inserting a probe laterally into its center, decompressing the nucleus, which allows the herniated portion to retract (Fig. 9). New techniques using arthroscopy and laser have improved the success rate, but it is still only in the 50 to 60 per cent rage, considerably lower than for discectomy. Sophisticated arthroscopic instrumentation and approaches are being developed. With these technological advances, it is very likely that the success rate will improve in the future.



Fig. 9. Percutaneous discectomy.

Microdiscectomy Microdiscectomy involves operative removal of the herniated portion of the disc, with the surgeon wearing magnification loops or using a surgical microscope (Fig. 10). It can be done through a short midline incision, which allows clear visualization of the nerve and minimal surgical trauma; this has decreased the postoperative back pain. The procedure was commonly done under general anesthesia but may be carried out under local anesthesia in a prone, kneeling position.



Fig. 10. Microdiscectomy.

Conventional discectomy This procedure is performed in a manner similar to microdiscectomy but with a slightly larger incision. A small amount of bone may be removed from the lamina and facet joint in order to identify the nerve root. After removal of the disc, the nerve root is examined and followed out into the foramen to be sure there are no further fragments or compression of the nerve and that a lateral recess exists. There appears to be little place for a formal laminectomy or complete removal of the lamina in the treatment of herniated disc. That procedure is still used in the treatment of nerve-root compression in the condition of spinal stenosis, which will be discussed later. The outcome of invasive treatment of disc herniation for sciatica has to do with the experience of the surgeon, accurate knowledge of the anatomy, and the use of proper instrumentation. Poor results may be due to the presence of a retained fragment or failure to decompress the nerve root if it compressed in the foramen. Overall, the best results can be expected using the limited or microdiscectomy. Fusion is rarely indicated in the treatment of sciatica secondary to disc disease. With the best of clinical results and the complete resolution of symptoms, the incidence of recurrence is still in the 5 to 10 per cent range for herniated disc. The reason for this is that the disc is essentially abnormal: the herniated disc has thus degenerated through a combination of biochemical changes and biomechanical stress and such discs may undergo recurrent herniation.

Category IV: spinal stenosis

Spinal stenosis refers to a narrowing of the spinal canal, the nerve-root canal, or the intervertebral foramen within the spine (Fig. 11). It generally affects people over 60 years of age. Narrowing of these structures is due to bone overgrowth, normally within the facet joint. In addition, hypertrophy of the lamina and soft tissues encroaches upon the canal; soft tissues include the ligamentum flavum and the capsule of the facet joint. Stenosis is commonly between the 4th and 5th lumbar vertebrae or may be segmental and diffuse along the entire lumbar spine. It may also have a congenital origin in certain individuals, mostly in those with achondroplasia. Most commonly, stenosis is the end result of degenerative changes within the spine; it may also be the result of scarring after previous surgery such as laminectomy and fusion. A burst-type fracture may also produce stenosis of the lumbar spine.



Fig. 11. CT scan showing marked spinal stenosis.

Clinical picture

Diagnosis of spinal stenosis can easily be made from the history. Probably a more accurate descriptive term would be ‘neurogenic claudication', as there are many people who have relative narrowing of their spinal canal or foramen who are not symptomatic. In its true sense, spinal stenosis is a radiographic finding. Neurogenic claudication refers to patients who present with complaints of pain or numbness, or a feeling of weakness, in the lower extremities. These symptoms are brought on by standing and walking and are relieved by sitting or adopting a flexed position. The differential diagnosis includes vascular claudication, from which the patient finds relief merely by stopping walking and generally will not be symptomatic when standing still. SYmptoms can often be recreated on physical examination by having the patient stand to walk or to extend their spine. Patients with stenosis will often stand in a somewhat forward-flexed position to obtain relief. They may present with one or both legs symptomatic. Examination generally does not reveal any signs of sensory or motor deficit. Reflexes are generally intact. They may have associated knee or hip arthritis. It is important in the examination to check the pulses; some patients may have a combination of neurogenic and vascular claudication.

Diagnostic studies

Plain films will often show a degenerative disc and arthritis of facet joints. It is not uncommon to see a mild subluxation, most often between the 4th and 5th vertebrae, adding to the problem; this is more common in females. MRI is helpful, but if surgical treatment is anticipated, myelography by CT is still the most definitive test for localizing areas of stenosis (Fig. 12). Symptoms of stenosis generally occur in a cross-sectional area of less than 100 mm^2 at the level of L3. Electrodiagnostic studies may be helpful, especially if carried out when a patient has been standing or walking for some time. These studies may also be helpful in attempting to rule out the underlying causes of neuropathy that may be superimposed upon stenosis.

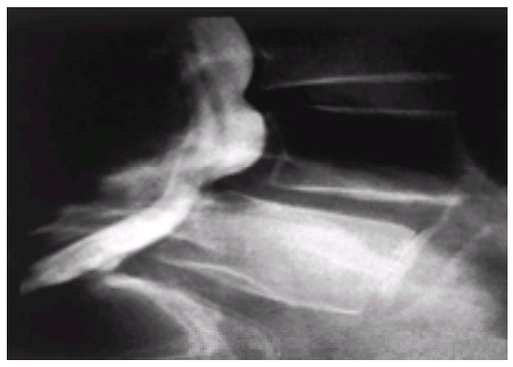


Fig. 12. Myelogram showing narrow canal posteriorly secondary to degenerative spondylolisthesis.

Pathophysiology

The most common form of spinal stenosis is acquired secondarily to degenerative changes. The spinal canal and foramen become narrowed as a result of thickening of the ligamentum flavum, bulging of the intervertebral disc, and degenerative hypertrophy of the facet joint. Once the canal narrows, its cross-section decreases and nerve-root compression can develop. Nerves within the cauda equina do not have myelin sheathes and are therefore more susceptible to compression that other

peripheral nerves. Compression of the nerve root may occur within the central canal under the inferior portion of the superior facet or at the foraminal level.

Treatment

In general, bed rest is of little value unless patients are having acute sciatic irritation. Patients generally find that they are more comfortable carrying out activities in the flexed or sitting position. They will find walking uphill easier than downhill because of the position of the spine, and walking holding on to a shopping cart allows them to walk for a greater distance before becoming symptomatic. Generally, analgesics and muscle relaxants are not needed. Epidural steroid injections bring about some relief of SYmptoms, unfortunately not long term. Braces and corsets do not seem to be helpful. Physical therapy helps patients engage in aerobic activities that do not extend the spine. Swimming or bicycle riding may be of benefit. Symptoms progress and increase in the long term but generally not rapidly. It is very unusual for patients to develop significant neurologic deficit.

Surgical treatment Patients in generally good health who have persistent or increasing functional limitations may be candidates for surgery. Decompression is the mainstay of surgical treatment for stenosis (Fig. 13). Many patients with spinal stenosis also have spinal deformity in the form of a spondylolisthesis and/or degenerative scoliosis (Fig. 14). Surgical decompression by laminectomy and foraminotomy is the treatment of choice if deformity is not present and instability is not detected on flexion–extension films. Discectomy is rarely indicated and may in fact lead to postoperative morbidity by increasing instability. It is important to identify clearly the neural elements and to follow the nerve roots out through the foramen to be sure that adequate decompression has been accomplished. Surgery will be necessary if instability is noted on preoperative flexion–extension films, or if spondylolisthesis or scoliosis are noted. While patients may be treated without fusion if subluxation is present, generally the results are not as good. Spinal instrumentation may be employed and the results suggest an increased rate of fusion in patients with degenerative scoliosis (Fig. 15). In addition to fusion, an attempt must be made to bring about better coronal and sagittal balance.

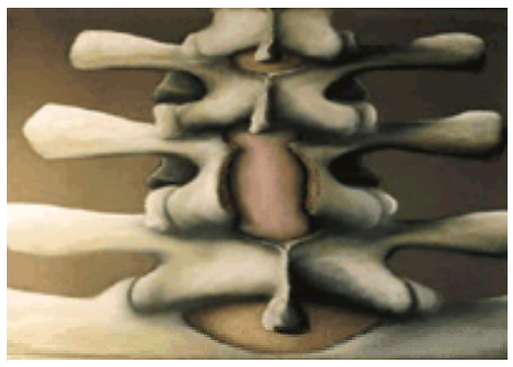


Fig. 13. Lumbar decompression.



Fig. 14. Myelogram showing degenerative scoliosis with canal compromise.



Fig. 15. Postoperative radiographs showing instrumentation and fusion in the treatment of degenerative scoliosis.

Disorders of the cervical spine

The degenerative processes that occur in the cervical spine parallel those in the lumbar area. Terms such as degenerative arthritis, soft disc disease, and spondylosis all really describe manifestations of a degenerative process. Degeneration is an continuous process that may become clinically symptomatic when biomechanical stresses are in excess of normal. Pain will develop as a result of mechanical or chemical irritation to the nociceptive nerve endings within the facet joints, the intraspinous ligaments, and the outer layers of the intervertebral disc. Neurologic deficits, and loss of sensation or strength in addition to pain, may occur if a nerve or nerve roots are compressed. Myelopathy may develop if the spinal cord is compressed either by disc herniation or by degenerative changes within the disc and the surrounding facet joints.

Mechanical neck pain

Clinical manifestations

Patients who develop neck pain may often relate most of their symptoms to an episode of trauma. Even more often, symptoms will develop gradually. Pain may be limited to the neck area but will occasionally be referred to the shoulders, the suboccipital area, and the interscapular area. In addition, a patient may present with vague symptoms such as blurred vision, tinnitus, or dysphasia. Often the discomfort is worse in the morning. Pain occurs during a period of driving or reading. There may be an area of point tenderness in the surrounding musculature, which is referred to as a trigger point. Except for limited range of motion and tenderness, these patients do not present with any sign of neurologic deficit.

Pathophysiology

Cervical spondylosis or mechanical neck pain begins with degeneration of the cervical spine. Mechanical breakdown of the integrity of the disc makes it more susceptible to the effects of minor trauma. Symptoms develop from episodes of trivial injury to the ligaments and joints of the spine. Mild instability may ensue, causing the formation of osteophytes; these can compress not only adjacent nerves but also adjacent vascular or visceral structures, and may produce an inflammatory

response. The pain is often reproduced by injection of hypertonic saline solution into a facet joint or the interspinous ligaments.

Diagnosis

Patients must be questioned for symptoms that may be ‘red flags': these include the presence of night pain, weight loss or fever; a history of significant trauma or of cancer may also be important. Diagnostic studies may include a cervical spine radiograph, which more often than not will show some narrowing of the disc space and possibly some hypertrophic changes at the adjacent facets. Other diagnostic studies such as CT, MRI, or myelography are generally not indicated unless there is no response to treatment in a reasonable period of time, or a change in symptoms.

Treatment of neck pain without radiculopathy

The majority of patients with neck pain due to degenerative process within the spine can be treated with a conscientious program of nonoperative management. The natural history of degenerative disease of the cervical spine is that of gradual resolution of symptoms in the majority. Programs will employ some controlled methods that can be helpful in reducing pain and discomfort, and limit disability.

Immobilization

Reducing the mobility of the spine is useful in both the acute and chronic state. It allows for healing of soft tissues and a decrease in the inflammatory reaction. Immobilization can be done in a number of ways, including the use of a soft collar, bed rest, or traction. The most effective method is a soft cervical collar to maintain the neck in a neutral or somewhat flexed position. Many collars will actually place the neck in an extended position, which will often aggravate inflamed facet joints, and the collar should not be worn for an extended period of time as this may lead to atrophy of the paraspinal muscles. As soon as pain begins to subside, a series of neck exercises should be initiated.

Physical therapy

Once pain resolves, exercise is important. There are two types of recommended exercise, isometric and range of motion. Isometric exercise strengthens the paraspinal musculature without aggravating the underlying disc and joints. The range-of-motion program, which eliminates soft tissue contracture and loss of mobility, is started after the isometric program.

Surgical treatment of non-neurogenic neck pain

In the majority of patients with neck pain, symptoms will resolve over a period of weeks to months. Occasionally, a patient will fail to improve on a conservative program. Diagnostic studies should then include plain radiographs in addition to MRI. MRI may show more than one disc undergoing degenerative change and herniation, and localization of the level that the pain is generated may be difficult, if not impossible. In certain situations, provocative discography may be carried out with an injection of a small amount of saline into the disc; if one or two levels can be identified, then surgical discectomy and fusion may be highly effective in reducing chronic neck pain.

Cervical radiculopathy

When the degenerative process within the cervical spine causes compression or irritation to the nerve root, the patient may or may not experience neck pain but will often experience pain that radiates to the chest, shoulder, arm, or hand. This pain may be associated with weakness and loss of sensation. Nerve-root compression may be brought about by herniation of disc material, known as soft disc, but more often is due to narrowing of the neural foramen as a result of degenerative changes at the disc level at the joint of Luschka in addition to the posterior facet.

Clinical picture

The patient with cervical radiculopathy typically is in the fourth decade of life. Men seem to be affected more frequently than women in a ratio of approximately 4.1:1. Symptoms may begin acutely or with gradual development of pain. Often, there is no precipitating event.

Diagnosis

Many patients present with arm or chest symptoms rather than neck pain. It is important to consider and rule out a cardiogenic background. Other causes may include bursitis or tendinitis about the shoulder or elbow. On physical examination, patients will often show signs of sensory deficit in a specific dermatomal distribution. Examination of the muscle groups may show deficit even in the absence of pain or numbness. Reflex changes may be present, correlating with the specific nerve involved. Differential diagnosis must include spinal cord tumor, tumors of the brachial plexus, neuropathy, Pancoast tumor, thoracic outlet tumor, and peripheral neuropathy. A careful history and physical examination is the most helpful in ruling out these lesions, but diagnostic studies are often necessary. Plain radiographs of the cervical spine include an oblique view to show areas of impingement on the neural foramen. The most definitive study is MRI, which has taken the place of myelography and CT in evaluating the cervical spine. Myelography may still be necessary, however, if the patient is claustrophobic or should not undergo MRI.

Treatment

As with patients with acute or chronic mechanical pain, the patient with cervical radiculopathy has a good overall prognosis for resolution of symptoms. Nonoperative treatment makes use of immobilization, medication, and isometric exercises. During the acute phase, nonsteroidal anti-inflammatory medication or oral steroids and short-term analgesics are recommended. Limitation of activity and use of a soft collar may also be helpful. As symptoms tend to resolve, an aggressive program of stretching and isometric exercises is continued, allowing the patient progressive activity to increase their cardiovascular status. Some 70 to 80 per cent of patients with acute cervical radiculopathy should expect good resolution of symptoms in 2 to 3 months.

Operative treatment

Indications for surgical intervention include progressive neurologic changes and failure of nonoperative treatment over a reasonable period of 4 to 6 weeks. There appears to be little agreement among surgeons about the best approach to the cervical spine, either anterior or posterior, in the treatment of cervical radiculopathy. The anterior approach is best for reaching the midline disc or for patients who experience not only arm but also neck pain, and also provides a method of fusion to stabilize the involved segment. If the lesion is peripheral and involves the foraminal level alone, with no neck pain, then a posterior approach or foraminotomy may be selected. This approach also allows the surgeon access to multiple roots if necessary.

Cervical spondylitic myelopathy

As the degenerative process proceeds within the cervical spine, and especially if the patient has a relative narrow canal, compression of the spinal cord may occur. If this occurs, the patient may be found to have involvement of not only the upper but also the lower extremities. The neurologic pattern may show signs of a lower motor-neuron lesion in the upper extremities; that is, sensory loss or weakness. In the lower extremities, one may detect upper motor-neuron lesions including gait disturbance, bowel or bladder disturbance, or spasticity.

Pathophysiology

Two anatomic factors play a part in the development of cervical myelopathy. The first involves the arterial blood supply to the spinal cord. The cervical cord is supplied from a large anterior spinal artery and two smaller dorsal lateral arteries. These arteries are fed by medullary vessels, which enter the spinal canal through the foramen at each level. Compression of the spinal artery by an osteophyte or disc impairs this blood supply. In addition, the medullary feeders may be compromised by narrowing of the foraminal opening. Another factor that has a role in cervical spinal myelopathy is narrowing of the spinal canal, which results from either protrusion of disc material or osteophyte formation and buckling of the ligamentum flavum posteriorly. Myelopathy in the cervical cord appears to be related to both the mechanical compression and vascular compromise. Patients typically seem to preserve the function of their posterior column, i.e., that of position sense, but there is compromise of the anterior portion of the spinal cord. Demyelinization of the white matter develops. MRI will not only reveal narrowing of the spinal canal, but may also detect changes within the spinal cord consistent with degeneration.

Diagnostic studies

Patients with cervical spondylitic myelopathy may often describe difficulties in walking in addition to a feeling of weakness or numbness in the upper extremities. Symptoms of bladder dysfunction may also be present. The examination reveals sensory and motor deficits in addition to a decreased reflex pattern in the upper extremities and hyper-reflexia, pathological reflexes in the area below the lesion, and may include the Hoffman sign in the upper extremity and the Babinski in the lower. Diagnostic studies should usually include plain films and will show the presence of degenerative change and osteophyte formation (Fig. 16). MRI offers the most definitive examination, giving the clearest picture of the spinal canal and the dent of cord compression. MRI may also help prognostically, with a rather poor prognosis if changes within the cord are present.



Fig. 16. CT scan of cervical spine showing calcified disc with marked narrowing of spinal canal.

Treatment

There is little place for nonoperative treatment in patients with SYmptomatic cervical spondylitic myelopathy. The natural history is that of progression and the patient must be informed of this. As in cervical disc disease, there are two surgical approaches, anterior or posterior. If the patient appears to have most compromise anteriorly on diagnostic studies, then the anterior approach is recommended. It is especially indicated if there is an tendency for straightening or kyphosis to be present. However, if the spine remains in a lordotic position, and especially if the compression appears to be more posterior, than the posterior approach would be indicated. New types of posterior approach are laminectomy, complete removal of lamina, or laminaplasty. The posterior elements are hinged in laminaplasty, allowing for increased canal space. Laminaplasty may also be indicated in a young patient with several levels of stenosis. It does not involve a fusion. Results are in the 90° range for patients operated on from the anterior with corpectomy, interbody fusion, or plating (Fig. 17). Both the anterior and posterior approaches have a reasonably good success rate of 66 to 92 per cent in preventing progression of the neurologic deficit. Unfortunately, in the incidence of reversal of the deficit, improvement of symptoms is less than 50 per cent. Given these results, the earlier patients with symptomatic cervical spondylitic myelopathy are treated with decompression, the better the prognosis.



Fig. 17. Postoperative radiograph: corpectomy, insertion of fibular graft, and anterior spinal plate.

Further reading

Acute low back pain, recurrent low back pain, sciatica

Beaman DN, Graziano GP, Glover RA, Wojtys EM, Chang V. Substance P innervation of lumbar spine facet joints. *Spine* 1993; **18**: 1044–9. [Substance P was found in surgically removed facet joints in patients presenting with degenerative abnormalities.]

Bell GR, Rothman RH. The conservative treatment of sciatica. *Spine* 1984; **9**: 109–54. [Excellent review of nonsurgical treatment of sciatica.]

Benoist M *et al.* Random double blind study to compare low dose and standard dose chymopapain: the treatment of herniated lumbar intravertebral disc. *Spine* 1993; **18**: 28–34. [Study found low-dose chymopapain to be as effective as standard dose in the treatment of herniated disc.]

Buirski G, Silberstein M. Symptomatic lumbar disc in patients with low back pain: magnetic resonance imaging appearances in both symptomatic and controlled population. *Spine* 1993; **18**: 1808–11. [The study found no significant difference in the incident of degenerative disc in symptomatic and asymptomatic patients.]

Bush K, Cowan N, Kadz DZ, Gishen P. The natural history of sciatica associated with disc pathology: a prospective study with clinical and independent radiographic follow up. *Spine* 1992; **17**(10): 1205–12. [Excellent resolution of sciatica without regard to radiograph findings.]

Cuckler JM, Bernini PA, Weisel SW, Booth R, Rothman R, Pickens G. The use of epidural steroids in the treatment of lumbar radicular pain: a prospective randomized double blind study. *Journal of Bone and Joint Surgery* 1985; **67**A: 63–6. [Respective study showed no significant difference between epidural steroids and placebo.]

Deyo RA. Conservative therapy for low back pain: distinguishing usefulness from useless therapy. *Journal of the American Medical Association* 1983; **250**: 1057–62. [Excellent review of various methods of treatment of lower back.]

Deyo RA, Diehl, AK, Rosenthall M. How many days of bed rest for acute low back pain. *New England Journal of Medicine* 1986; **315**: 1064–70. [Best results with only a few days of bed rest.]

Elnaggar IM, Nordin M, Sheikhzadeh A, Parnianpour M, Kahanowitz N. Effects of spinal flexion extension exercises on low back pain and spinal mobility in chronic mechanical low back pain patients. *Spine* 1991; **16**: 967–72. [Both flexion and extension exercises reduced incidents of low back pain.]

Frymoyer JW. Back pain and sciatica. *New England Journal of Medicine* 1986; **315**: 1090–2. [Excellent review of treatment of back pain and sciatica.]

Frymoyer JW. Quality: an international challenge to the diagnosis and treatment of disorders of the lumbar spine. *Spine* 1993; **18**: 2147–52. [Discussion of the quality and cost of treatment of lumbar spine disorders.]

Grub SA, Lipski HJ. Results of lumbosacral fusion for degenerative disc with and without instrumentation: 2–5 year follow up. *Spine* 1992; **17**: 349–55. [Use of instrumentation reduced the incidents of pseudarthrosis.]

Kanhanobitz N. Percutaneous discectomy. *Clinical Orthopedics* 1992; **284**: 75–9. [Review of percutaneous treatment for herniated disc.]

Kirkaldy-Willis WH, Hill RJ. A more precise diagnosis of low back pain. *Spine* 1979; **4**: 102–9. [Outline discussion of the pathophysiology of lower back pain.]

Knox BD, Chatman TM. Anterior lumbar interbody fusion for discogram concordant pain. *Journal of Spinal Disorders* 1993; **6**: 242–4. [Study showed good correlation between discography and

interbody fusion results.]

Schaffer JL, Cambin T. Cutaneous posterolateral lumbar discectomy and decompression with 6.9 mm. cannula. *Journal of Bone and Joint Surgery* 1991; **73**A: 822–31. [New procedure, percutaneous discectomy.]

Spengler DM. Lumbar discectomy: results with limited disc excision and selective foraminotomy. *Spine* 1982; **7**: 604–7. [Excellent results using a traditional discectomy for the treatment of herniated disc.]

Webber H. Lumbar disc herniation: a controlled prospectus study with ten years of observation. *Spine* 1983; **8**: 131–40. [Ten-year follow up of patient with sciatica treated with and without surgery.]

Spinal stenosis

An HS, Gover JM. Spinal stenosis. Historical prospectus, classifications, pathoanatomy. *Seminars in Spine Surgery* 1994; **6**: 69–77. [Excellent discussion and review of spinal stenosis.]

Bell GR, Rothman RH. Conservative treatment of sciatica. *Spine* 1984; **9**: 54. [Discussion of the nonoperative treatment of sciatica.]

Herkowitz HN, Kurz LT. Degenerative lumbar spondylolisthesis with spinal stenosis. *Journal of Bone and Joint Surgery* 1991; **73**A: 802–8. [Patients treated with fusion had a better outcome.]

Katz JN, Lipson SJ, Larson MG, McInnes J, Fossel A, Liang M. The outcome of decompressive laminectomy for degenerative lumbar stenosis. *Journal of Bone and Joint Surgery* 1991; **73**A: 809–16. [Many patients developed recurrent symptoms after decompressive laminectomy for stenosis.]

Degenerative disease of the neck

An HS, Simpson, JM. *Surgery of the cervical spine*. Martin Dunitz, London, 1994. [Clear and concise for the surgical treatment of cervical spine.]

Bohlman HH, Emery SE, Goodfellow DB, Jones P. Robinson anterior cervical discectomy and arthrodesis of cervical radiculopathy. *Journal of Bone and Joint Surgery* 1993; **75**A: 1298–307. [Six-year follow-up of 122 patients with good results using Robinson anterior cervical discectomy and fusion.]

Clark CR. Cervical spondylitic myelopathy: history and physical findings. *Spine* 1988; **13**: 847–9. [Clear description of the history and physical findings of cervical spondylitic myelopathy.]

Dillon W, Booth R, Cuckler J, Balderstone MD, Simione F, Rothman R. Cervical radiculopathy: a review. *Spine* 1986; **11**: 988–91. [A discussion in review of the treatment of cervical radiculopathy.]

Gore DR, Sepic SB, Gardner GM, Murray MP. Neck pain: a long term follow up of 205 patients. *Spine* 1987; **12**: 1. [Ten-year follow up of patients with neck pain. Seventy-nine per cent improved without surgical treatment.]

Herkowitz HN. A comparison of anterior cervical fusions, cervical laminectomy and cervical laminaplasty for the surgical management of multiple level spondylitic radiculopathy. *Spine* 1988; **13**: 774–80. [Anterior cervical decompression and fusion had better results than laminoplasty. Laminectomy had the poorest overall results.]

LaRocca H. Cervical spondylitic myelopathy: natural history. *Spine* 1988; **13**: 854. [A discussion of the natural history of cervical spondylitic myelopathy.]

Watzel FT. Chronic benign cervical pain syndromes: surgical considerations. *Spine* 1992; **17**(10S): 367–74. [Poor results following surgical treatments. Suggestion that dorsal column stimulators may have some place in treatment of chronic neck pain.]

Whitecloud TS, III, Scago MC. Cervical discogenic syndrome: results of operative intervention of patients with positive discography. *Spine* 1987; **12**: 313–16. [Seventy per cent excellent result with cervical discectomy in patients with positive discography.]

46.2.7 Spinal trauma

Philip R. Lucas

Introduction
Principles of treatment of spinal injury
Diagnostic studies
Initial management
Trauma to the cervical spine
Upper cervical spine and atlanto-occipital injury
Fractures of the lower cervical spine, C3–C7
Soft-tissue injuries of the neck
Thoracolumbar spinal injuries
Introduction
General principles of stability
Thoracolumbar trauma
Treatment of specific fractures
Principles of spinal surgery in treatment of thoracolumbar trauma
Further reading

Introduction

Injuries to the neck and back are common, and range in severity from minor soft-tissue injuries involving muscles or ligaments to severe, catastrophic injuries that render the spine unstable and the patient paralyzed. Most injuries fall somewhere in between. The most common cause of spinal injury is a motor vehicle accident, followed by falls and diving accidents. In the Western world, especially the United States of America, gunshot wounds have unfortunately become a common cause of spinal trauma. In the United States alone there are 50 000 spinal fractures every year; about 20 per cent of these involve some type of neurologic deficit. A typical patient in 80 per cent of cases is male and generally under 40 years of age. Most commonly, injuries occur in the area between the 4th and 6th cervical vertebra and the thoracolumbar junction of the spine (Fig. 1).

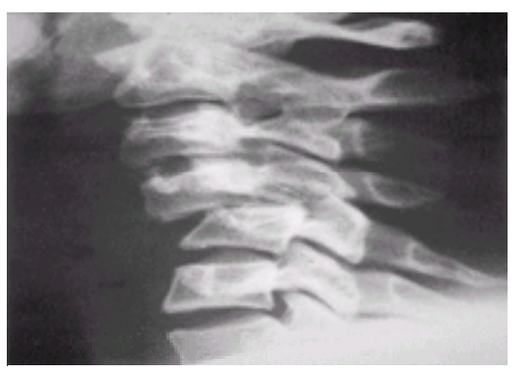


Fig. 1. Lateral radiograph showing 50 per cent subluxation of C4 on C5, consistent with bilateral facet dislocation.

Much progress has been made over the last 10 to 15 years in the treatment of patients with spinal injury. Long-term disability, secondary to neurologic deficit, is still devastating. These are several areas where large strides have been taken but much work still has to be done. With regard to prevention, the advent and requirement of seat-belts and airbags in motor vehicles has resulted in a substantial decline in severe spinal injuries. In addition, the development of emergency medical technicians attending patients at the scene of injury with early stabilization has been invaluable in early treatment and the prevention of severe deficits. New diagnostic studies have given us a better understanding and a clearer picture of the extent of injury aiding a more accurate diagnosis and definitive treatment. Research, especially in the area of pharmacologic treatment of spinal-cord injury, has just scratched the surface. Methylprednisolone given acutely appears to stabilize cell membranes and reduce edema. There are many factors presently being studied for their effect on regrowth of injured nerve tissue and the results are promising.

Principles of treatment of spinal injury

As in all trauma, it is imperative to follow a definite protocol. These steps include initial evaluation, which begins at the scene of the accident and involves following the ABC's of any trauma. The patient must be immobilized and assumed to have had a spine injury (Fig. 2). A history obtained from the patient or witnesses can often be invaluable in attempting to determine the mechanism of injury. Examination of the patient involves a search for abrasions and contusions that may also suggest the mechanism of injury. The neurologic assessment must include a complete sensory and motor examination including reflexes, specifically the bulbocavernosus reflex. It is important that the findings be documented; it is the only way to follow any neurologic deterioration.



Fig. 2. Illustrating the importance of immobilization of the spine until complete evaluation of the trauma patient.

Diagnostic studies

Once the patient has been stabilized, it is important to obtain an initial radiograph of the lateral cervical spine (to include C7) and an open-mouth anteroposterior view. Approximately 85 per cent of significant injuries of the cervical spine will be detected in a lateral radiograph. Anteroposterior and lateral radiographs of the thoracic and lumbar spine are then taken. If an area such as the cervicothoracic junction cannot be well visualized, a computed tomographic (**CT**) scan of this or any other suspect area will be of great help in assessing injury. Magnetic resonance imaging (**MRI**) or myelography are generally not required unless there is a neurologic deficit.

Initial management

Once a diagnosis of spinal injury has been made, it is important to determine if a neurologic deficit exists and whether the spine is stable. These are important factors to note before one sets about treating the patient. Neurologic evaluation must be complete, looking at any sensory deficit, and the motor examination should include grading of muscle strength (Fig. 3) Reflexes should be recorded, including the presence or absence of the bulbocavernosus reflex. During the first 48 h, this reflex may be absent in a patient with profound neurologic deficit who may be in spinal shock. It is important to note the return of the reflex. There can be incomplete spinal injury

with partial preservation of motor and/or sensory function below the zone of injury. Incomplete injuries may include (1) a central cord syndrome, (2) an anterior cord syndrome, (3) Brown–Sequard syndrome, and (4) a post-cord syndrome. Functional return varies with each of these, being excellent in Brown–Sequard and generally poor in the central cord. Neurologic status may also be evaluated according to the Frankel scale (Table 1).

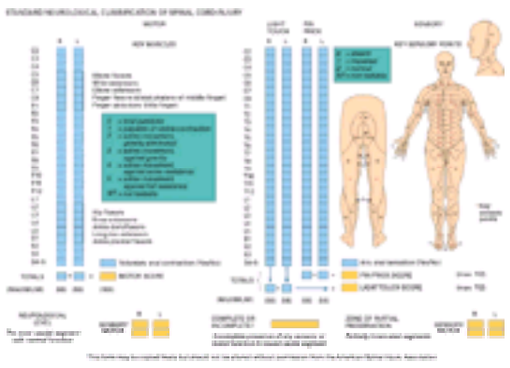


Fig. 3. Neurologic checklist prepared by the American Spinal Injury Association.

Syndrome	Sensory sparing	Prognosis
Central cord	All modalities (distal, proximal)	Fair to good
Anterior cord	Touch, vibration	Poor
Brown–Sequard	Touch bilateral, pain ipsilateral	Good
Posterior	Pain, slight touch	Fair to good

Table 1 Prognosis and neurologic features of the spinal cord syndromes following incomplete spinal-cord injury

Spinal stability

In any spinal injury, it is essential to determine the stability of the spine as best one can. Stability refers to the ability of the spine to function under normal physiologic loads without increased pain, neurologic deficit, or progression of deformity. Determination of stability can be difficult. Evaluation of the patient is based upon the examination and diagnostic studies. Given this information, one should be able to determine the relative stability of the spine. This is best outlined in White and Punjabi's diagnostic checklist for spinal stability (Table 2).

1	Anterior elements = 2
2	Posterior elements = 2
3	Sagittal translation > 3.5 mm or 20 per cent vertebra = 2
4	Sagittal plane rotation > 11° = 2
5	Positive stretch test = 2
6	Cord damage = 2
7	Root damage = 1
8	Abnormal disc narrowing = 1
9	Dangerous loading anticipated = 1

Table 2 White and Punjabi spine stability checklist (score > 5 is unstable)

Trauma to the cervical spine

The neck area of the spine is involved in about half of all spinal injuries. Twenty per cent of patients with cervical spinal injury are hypotensive on their arrival to the emergency room and up to 30 per cent have had trauma involving intraperitoneal structures or the extremities. As previously mentioned, the spine must be stabilized until a complete anteroposterior and lateral radiograph are obtained. Radiographs should include the C7–T1 junction. Reviewing the lateral film, one should look at soft tissues for prevertebral swelling, which may alert the clinician to a possible spinal injury in an otherwise normal-appearing radiograph. The overall alignment of vertebral bodies and spinous processes should be evaluated, including abnormal angulation or displacement. Axial CT with or without reconstruction is invaluable in attempting to delineate better a suspected area of injury, especially one that cannot not be visualized on plain radiographs, such as the cervicothoracic junction. MRI and myelography can be helpful, but need only be employed in patients that have neurologic deficit. Displaced or angulated fractures should be treated with the application of traction; skeletal tongs or a halo are preferred. The rule of thumb is to start with 2.25 kg (5 lb of weight per level of injury), allowing approximately 5 kg for the weight of the head. Using this guideline, a C5–6 injury would initially be treated with 13.5 to 16 kg of traction (30–35 lb). Generally, it is best to start with a small amount of weight and then gradually increase it. During this time, repeat radiographs and clinical evaluation are of the utmost importance.

Upper cervical spine and atlanto-occipital injury

Injury to this region of the cervical spine is apparently rare, which does not necessarily mean that such injuries are truly infrequent but rather that they are not always recognized. They are frequently accompanied by severe head injury and the displacement of fractures at this junction most often is fatal (Fig. 4). An axial load to the top of the head will result in a unilateral/bilateral condylar fracture, whereas hyperextension may lead to atlanto-occipital dislocation. Depending on the injury, radiographs may be difficult to interpret. Occipital–atlantal dislocation injuries are evaluated on the basis of criteria described by Powers. If the injury is a stable fracture, treatment usually consists of immobilization with orthotic support for a period of 6 to 8 weeks. Following this, the spine should be evaluated for signs of instability with flexion–extension films. If one has determined, based upon the Powers' criteria, that ligamentous injury has occurred with subluxation of C1–C2, such injuries are best treated with occiput to C1 or C2 fusion after reduction and traction.

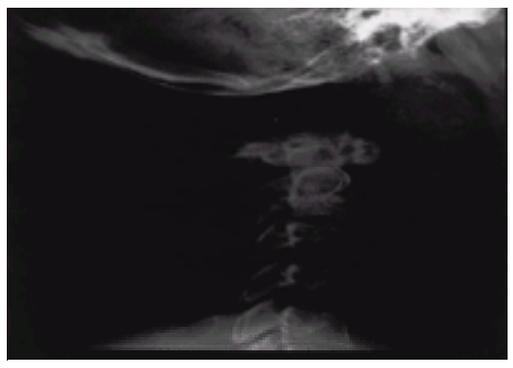


Fig. 4. Radiograph of occipital–atlas displacement, almost always fatal.

Fractures of the atlas

Fractures of C1 account for a quarter of all injuries to the atlantoaxial junction and 10 per cent of all cervical spinal fractures. Injury generally comes about through axial load and results in a break in the ring of C1 with two sites of injury. Fractures are most often noted on open-mouth odontoid views and these may reveal lateral displacement of one or both articular masses. A displacement of the lateral mass greater than 7 mm suggests a rupture of the transverse ligament. Another type of C1 fracture is a Jefferson fracture, in which the ring is broken at four points. In the treatment of nondisplaced or minimally displaced C1 fractures, a cervical orthosis is worn for about 3 months. If 7 mm or more of displacement exists between the lateral mass, treatment includes cervical traction for 4 to 6 weeks and then either further halo immobilization for 2 to 3 months or C1–C2 fusion.

Atlantoaxial rotational injury

This is a purely ligamentous injury to atlantoaxial joints, most commonly seen in victims of motor vehicle accidents. Patients typically present with neck pain and spasm, and generally have an associated torticollis deformity. Diagnosis is made with the help of an open-mouth odontoid view or with a transverse CT scan. Typically, these injuries go untreated for weeks or months, and are generally associated with pain, stiffness, and spasm. Treatment involves axial traction in an attempt to reduce subluxation. Once this is achieved, then maintenance of reduction in a halo vest for 3 months is necessary; if instability is still present, then fusion of C1–C2 would be recommended (Fig. 5).



Fig. 5. Radiograph (postoperative), posterior of C1, C2 wiring.

Fracture of the dens

Up to 14 per cent of cervical spinal fractures involve fractures of the odontoid. Fracture patterns are divided into three types, which are important to know in attempting to develop a prognosis for healing and to provide treatment plans. Type I is an oblique fracture of the tip of the odontoid that probably represents an avulsion injury of the alar apical ligaments. It is a stable injury and patients are treated SYmptomatically with a collar. Type II is the most common and involves a fracture through the base or waist of the odontoid (Fig. 6). This injury has a fairly high rate of nonunion; the factors implicated in this include the patient's age, and displacement of greater than 5 mm or angulation of greater than 10°. Fracture healing also appears related to method of treatment. Treatment with a halo generally yields union of anywhere from 20 to 80 per cent. Patients who are more than 40 years of age and have significantly displaced or angulated fractures should be treated with reduction and posterior spinal fusion. Surgical treatment can include inserting a screw base of C2 into the odontoid, or lateral–posterolateral transarticular screws (Fig. 7), or wiring of C1 to C2 (Fig. 8). In type III fractures, the fracture line extends into the body of C2. These injuries generally have a higher rate of union, given the greater amount of cancellous surface at the fracture site. However, the nonunion rate is still 10 to 15 per cent when treatment is with a halo vest.

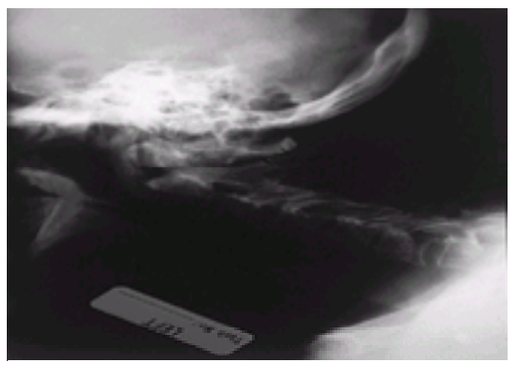


Fig. 6. Lateral radiographs showing posterior displacement of type II odontoid fracture.

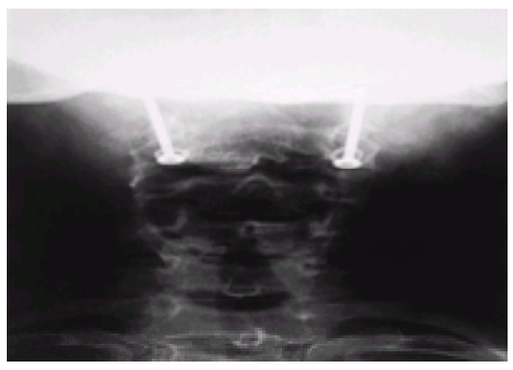


Fig. 7. Radiography showing transarticular screw fixation—C1, C2.

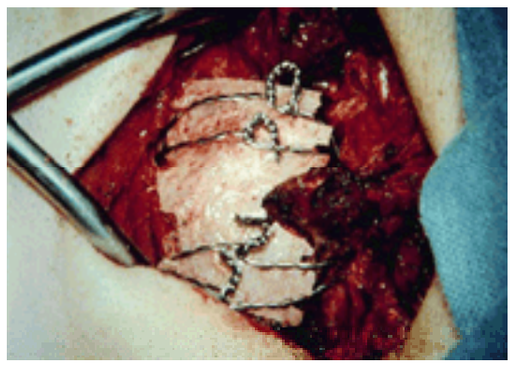


Fig. 8. Interoperative photograph showing bone graft, sublaminar wire—C1, C2.

Traumatic spondylolisthesis of the axis (hangman's fracture)

These fractures occur as a result of extension and axial compression. The fracture line passes through the pedicle of C2 (Fig. 9(a)). Four types have been described, based upon radiographic evaluation: type I with minimal displacement, less than 3 mm; type II, significant displacement, greater than 3 mm and angulation greater than 11° (Fig. 9(b)); type IIA, minimal displacement with angulation greater than 11°; and type III, associated with a facet dislocation. Treatment is dependent upon type. Type I is treated with a halo jacket for 12 weeks. Type II is treated with cervical traction for 2 to 3 weeks to reduce displacement, followed by a halo jacket for 10 to 12 weeks. Type IIA is treat with a halo and extension; type III is treated surgically to reduce the facet and then an anterior C2–C3 fusion or posterior fixation of C2–C3 with internal fixation.

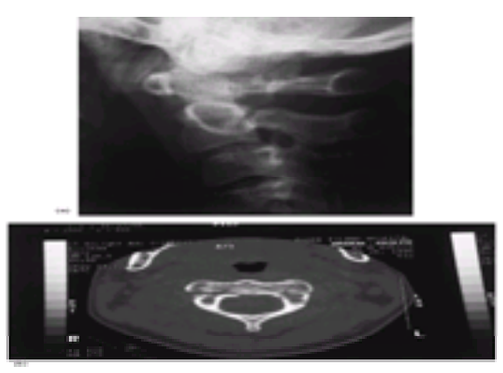


Fig. 9. (a) Lateral radiograph showing fractured C2; (b) CT scan more clearly demonstrating fracture pattern of the axis, type II. Fracture subluxation C4, C5, the result of flexion–distraction mechanism of injury.

Fractures of the lower cervical spine, C3–C7

Several factors are responsible for the different constellations of injuries that occur in this region of the spine than those of the occipital–atlantoaxial complex. First, the anatomy and the articulation of the vertebrae are relatively uniform. Each level or functioning unit is held together by an intervertebral disc anteriorly, and posteriorly by longitudinal ligaments, facet joints, and capsules. Furthermore, the diameter of the neural canal decreases to 70 mm, thereby decreasing the canal:spinal cord ratio and increasing the propensity for neurologic damage. In attempting to determine stability one must define a treatment plan. In the defining of stability, the lower cervical spine is divided into an anterior column, which involves the vertebral body and the disc, and the posterior column, which involves the spinous process, lamina, and facet joint. Once again, White and Punjabi's diagnostic checklist is valued in determining the degree of stability (Table 2).

Several classifications can be used to define fracture and dislocation of the lower cervical spine. The classification system of Allen–Ferguson is noteworthy: it includes distraction–flexion injuries, vertical compression injuries, pressure–flexion injuries, compressive extension, lateral flexion, and distraction–extension (Fig. 10).

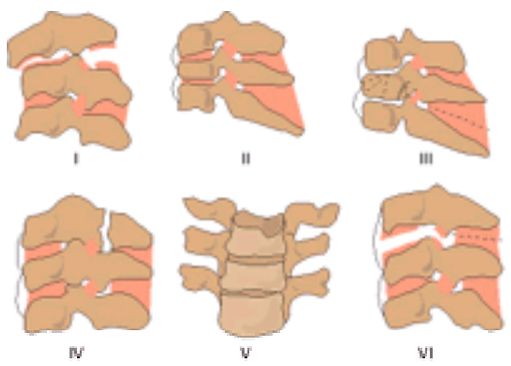


Fig. 10. Classification of fractures and dislocations of the lower cervical spine (Allen-Ferguson). I, distractive flexion (stages 1–4): perched facets, unilateral or bilateral facet dislocation. II, vertical compression (stages 1–3): burst fractures. III, compressive flexion (stages 1–5): compression fractures and tear-drop fractures. IV, compressive extension (stages 1–5): fractures of the posterior elements and associated vertebral bodies. V, lateral flexion (stages 1–2): uncommon. VI, distractive extension (stages 1–2): widening of the disc or retrolisthesis.

Fracture dislocations

These injuries may be unilateral or bilateral, and may be associated with disc herniations with anterior cord compression. Anterior displacement of approximately 25 per cent suggests a unilateral dislocation; 50 per cent dislocation suggests a bilateral facet subluxation. Neurologic involvement occurs in 70 per cent of patients with facet subluxation; this may be single, normal-root syndrome, as seen in the unilateral, or a severe neurologic deficit as a result of bilateral facet dislocation. These injuries are the result of distraction–flexion (Fig. 11). Treatment traditionally has involved the use of cervical tongs or halo traction to bring about reduction. The reduction of bilateral facet dislocations is controversial because of their high incidence of disc herniation. It is suggested by some that if a disc herniation is noted on an MRI obtained before reduction of these injuries, anterior discectomy should be carried out together with a reduction and stabilization. This view is not held by all. It would appear that reduction with skeletal traction is safe as long as it is done with increments of weight and continued clinical observation. If reduction is not possible and open reduction is contemplated, then an MRI should be performed before the reduction.

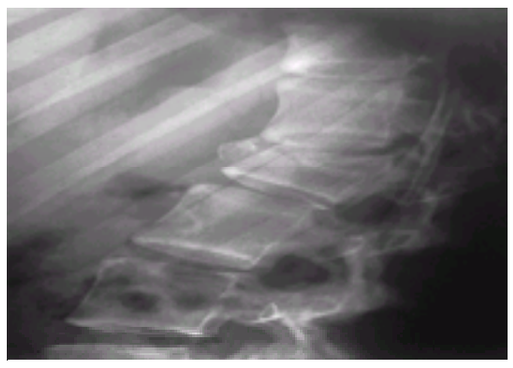


Fig. 11. Lateral view of thoracolumbar fracture as a result of flexion axial load.

Compression fractures of C3–C7 vertebral bodies

These fractures consist of varying degrees of compression of the vertebral body and compression– flexion injury. Treatment is dependent upon the degree of stability, based on White and Punjabi's checklist, and also whether or not neurologic deficit is present. Simple compression fracture, involving just the anterior column, may require just a cervical collar for 6 weeks. If the posterior elements are disrupted, then posterior fusion may be necessary. Vertical compression injuries, such as burst fractures, may be stable if there is minimal displacement, or could be unstable if there is marked spreading of the vertebral body, especially if the spinal canal is compromised, with compression of the cord. In this case, an anterior decompression graft and plating would be the most appropriate treatment.

Teardrop fractures

Teardrop fractures are produced by axial load with compression and rotation, and are always unstable due to significant comminution associated with disruption of the ligamentous complex.

Other cervical spine injuries

Fractures of the spinous process (clay shoveler's fractures) are an avulsion of that process; they are stable and treated symptomatically with a cervical collar. Fractures of the posterior elements include injuries that involve fractures of the lamina, pedicle, and articular processes. The mechanism of injury is generally compressive or distracted extension. This normally can be treated with 8 to 12 weeks of rigid cervical orthosis or halo. Lateral mass fractures that involve a pedicle and ipsilateral lamina are usually the results of lateral flexion. These fractures can be treated with rigid cervical orthosis for 8 to 12 weeks.

Gunshot wounds

There has been a significant increase in injuries to the spine and spinal cord from gunshot wounds over the last 10 to 5 years. In certain large cities of the United States, 25 per cent of spinal cord injuries are related to gunshot wounds. These injuries rarely cause sufficient bone or ligament injury to warrant surgical stabilization. They are often associated with marked neurologic deficit. Whether the bullet should be removed from the spinal canal is controversial. Certainly, in cases of an incomplete lesion with compression of the cord secondary to the retained bullet fragment, removal appears to be indicated. There is some evidence that lead toxicity may develop, together with leaks of cerebrospinal fluid and possible meningitis. A surgical approach is generally not recommended because these complications are rare. Each case has to be treated on its individual features.

Soft-tissue injuries of the neck

These injuries, often referred to as whiplash, may be due to acceleration, extension, or flexion–deceleration (Fig. 12). They routinely involve injuries to the muscles, and to the longitudinal ligaments and the intervertebral disc. Patients often present with neck pain, with radiation to their head, and/or shoulder or arm pain. The other symptoms are dysphasia. Radiographs, including flexion–extension films, should be obtained to rule out occult instability. Treatment involves an initial period of rest and then a progressive program of mobilization. If symptoms persist for over 3 months, an MRI would be appropriate. If internal derangement of a disc is noted, anterior cervical discectomy and fusion may be indicated.



Fig. 12. Illustration of soft-tissue (whiplash) injury of neck.

Thoracolumbar spinal injuries

Introduction

The thoracic and lumbar spines may be divided anatomically into three regions: (1) high thoracic, (2) thoracolumbar, and (3) lumbar. The upper thoracic spine, T2–T10, is a relatively stable area. Stability is provided not only by the disc and structures of the spine, but also by the thoracic rib cage. A high degree of energy is required for injury to occur. A relative high rate of spinal cord injury is associated with thoracic fractures if displacement occurs. The thoracolumbar area, T11–L2, is the most common site of spinal injury. There is a stress concentration at this junction due to the transition from the rigid thoracic region to the more mobile lumbar area, and neurologic injuries may affect the conus at the lower end of the spinal cord or the nerve roots. In the lower lumbar area, L3–L5, the vertebral bodies are rather large and are much more resistant to compressive loading. Due to the lordosis, there appears to be less tendency for kyphotic angulation as a result of compression. Neurologic injury may occur but is less common, due to the space available for the nerve roots above the cauda equina.

General principles of stability

Unlike the cervical area, where the notion of a two-column system is popular, the thoracolumbar region provides a three-column concept that has been popularized by Denis; this system is especially helpful in determining the stability of burst fractures. The anterior column involves the anterior longitudinal ligament, the anterior half of the annulus, and the vertebral body; the middle column involves the posterior half of the vertebral body and the posterior longitudinal ligament; the posterior column involves the pedicles, facets and lamina, spinous process and intraspinous ligament. If the middle column is disrupted, the spine is generally considered unstable (Fig. 13). However, in the thoracic region the same injury to the upper thoracic spine may be stable because of the stabilizing effect of the ribs, while in the thoracolumbar junction the injury may be deemed unstable. White and Punjabi's checklist (Table 2) is once again useful in determining the stability of the spine and treatment. As in all spinal injuries, complete evaluation based upon history, examination, neurologic assessment, and diagnostic studies is of the utmost importance. Documentation of neurologic function initially and in follow-up is important to detect progressive deficit.

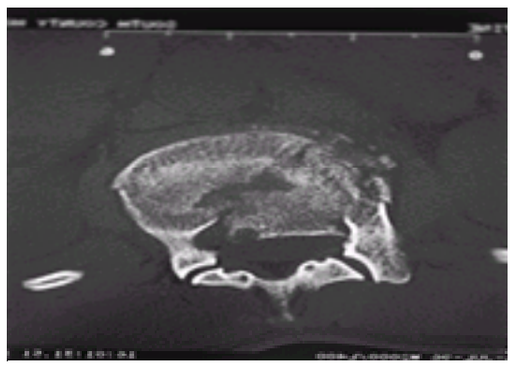


Fig. 13. CT scan illustrating marked comminution/burst-fracture retropulsion of middle column.

Radiographic evaluation

Plain radiographs, anteroposterior, and lateral views initially obtained should be interpreted with regard to any observed collapse of a vertebral body, kyphotic angulation, or translation. In addition, widening of the pedicles may be seen on an anteroposterior film and indicate a burst fracture. The CT scan is invaluable in assessing the degree of canal compromise in burst fractures. More recently, MRI has been used to evaluate the spinal cord in neurologically impaired individuals; in addition, it has value in determining disc herniation and ligamentous injury.

Thoracolumbar trauma

Normally, complex forces across the spine produce the fracture pattern. The forces most commonly associated with this type of injury, especially of the thoracolumbar spine, include axial compression, flexion, lateral compression, flexion–rotation, shear, flexion–distraction, and extension. Depending on location, some forces may produce different fracture patterns. When axial compression occurs upon the thoracic spine, it will generally produce a compression of the anterior portion of the vertebral body. At the thoracolumbar junction and the lower lumbar spine, axial compression will often produce a burst-type fracture.

Flexion

Flexion force causes compression anteriorly. Significant enough distraction may occur posteriorly. Generally, this will produce a wedge fracture in the thoracic or thoracolumbar spine. If more than 50 per cent compression is present in a young adult, this may indicate ligamentous injury posteriorly.

Lateral compression

Lateral compressive forces create a wedging vertebral body and are generally stable injuries.

Flexion–rotation

When exposed to a combination of flexion and rotation, the spine often suffers a severe injury and marked instability, resulting in a combination of bone and ligamentous disruption (Fig. 14). These are highly unstable injuries and generally require spinal stabilization.

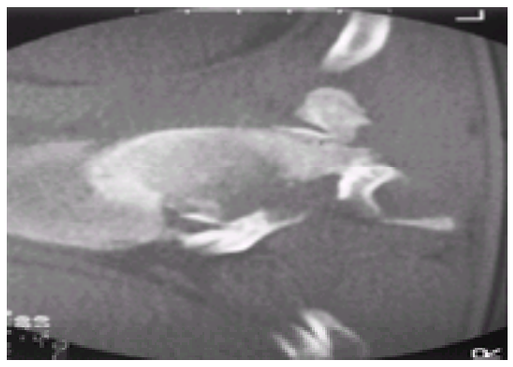


Fig. 14. CT scan showing marked dislocation of T7 on T8.

Flexion–distraction

Injuries due to flexion–distraction forces across the thoracolumbar spine are often referred to as Chance fractures, and are most often seen with seat-belt injuries when a lap belt alone is used. The axis of rotation is moved forward over the spine. In severe cases, it will produce injuries to all three columns. A purely bony injury can generally be treated with bracing in extension. If ligamentous injury is found, then spinal stabilization through surgery is necessary.

Extension injuries

The reverse of the pure flexion injury, extension is applied along the anterior longitudinal ligament and may, in fact, produce disruption of the disc or avulsion fracture and compression forces across the posterior elements. The atural tendency is to assume a kyphotic position. These generally can be treated with bracing, even if two-column involvement exists.

Treatment of specific fractures

Compression fracture

The definition of compression fracture includes compression of the anterior column, preservation of the middle column, and, only in severe cases, with 50 per cent or more vertebral compression is there possible disruption of the posterior ligamentous structures.

Treatment recommended

For patients with less than 40 per cent compression and less than 30° of kyphosis, a hyperextension brace (thoracolumbar sacral orthosis (TLSO)) is generally worn for 3 months. Normally, surgical treatment is recommended only in severe compression of the vertebral body, 50 per cent or more, and kyphotic deformity, over 30°.

Burst fractures

Treatment of burst fractures is dependent upon accurate assessment. The extent of injury determines stability and also whether or not neurologic deficit is present. By definition, every burst fracture must have disruption of the anterior middle column (Fig. 15). If there is less than 40 per cent canal compromise and normal neurologic examination, treatment generally involves a brief period of bed rest and then a well-fitting TLSO brace or cast. Surgery would generally be recommended if the patient has greater than 50 per cent canal compromise or neurologic deficit. However, a fracture that is regarded as unstable does not necessarily require surgical treatment,

as many burst fractures will heal and require several months of bed rest. Whether decompression should be carried out depends upon the neurologic picture. A review of several retrospective and prospective studies suggests that the chance of neurologic improvement is somewhat better with early decompression of a burst fracture, but by no means is this absolute. In summary, burst fractures, even if diagnostic criteria determine an unstable fracture, may be treated nonoperatively. If surgery is recommended, either an anterior or posterior stabilization and fusion is performed. If an incomplete neurologic picture is present, spinal surgeons favor decompression either from an anterior or a posterior approach. Patients with burst fracture and a complete neurologic deficit may be treated with bed rest and bracing, but generally the early immobilization provided by stabilization and fusion would be recommended.



Fig. 15. MRI illustrating retropulsion of middle column impinging on the spinal cord.

Flexion–distraction injuries—Chance fractures

These fractures may be missed on initial evaluation. Care must be taken to check for the tell-tale sign: increased interspinous space. Fractures that purely involve structure may be treated with a hyperextension brace (Fig. 16). However, posterior spinal compression construct is recommended if ligamentous injury occurs posteriorly.

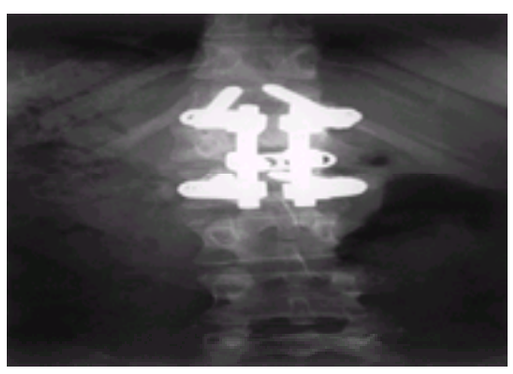


Fig. 16. Radiograph of short-segment stabilization posteriorly for T12, L1 Chance fracture.

Fracture dislocations

These are the most unstable of all injuries and involve flexion–rotation, shear injuries. When evaluating these patients, it is important to remember that a maximum displacement which occurred at the time of injury may be markedly reduced at the time the initial radiographs are taken. One should be particularly suspicious of such injuries, especially if multiple rib fractures or multiple fractures of transverse processes are noted. These injuries are associated with a fairly high incidence of neurologic injury, and early surgery is recommended to reduce the fracture and stabilize for early mobilization.

Minor fractures

This group includes fractures involving the transverse process, and fractures of the pars interarticularis. Fractures of the transverse process generally recur as the result of direct trauma or violent muscle contracture. These are stable injuries. Traumatic fractures of the pars interarticularis normally can be treated with TLSO brace, unless there is marked displacement and instability is visible on flexion– extension films.

Gunshot wounds to the thoracolumbar spine

Generally, gunshot wounds do not produce instability and often may be treated without surgery. Surgery is indicated, however, in the presence of progressive neurologic deterioration or proven neurologic compression as a result of bone, disc, or bullet fragments. Surgery may also be recommended if the gunshot wound perforates the colon and traverses the spine. These instances are best treated by surgical debridement and antibiotics.

Principles of spinal surgery in treatment of thoracolumbar trauma

Over the last 20 years, surgical treatment for thoracolumbar spinal fractures has become popular. Earlier treatment with laminectomy often led to catastrophic results. Before the advent of instrumentation, spinal fusions still required long periods of rest and did not offer advantages over nonoperative treatment, especially in fractures that would eventually stabilize. Spinal instrumentation beginning with Harrington rods are the spinal surgeons' method of internal stabilization, allowing for decompression and fusion, and, most importantly, early mobilization (Fig. 17). There has been an evolution in spinal instrumentation over the years. Most such instrumentation was initially developed for the treatment of deformity, though its use in the treatment of trauma has become increasingly important. The type of instrumentation does not matter greatly as differences between SYstems are minor. What is most important is that they be used as designed, keeping in mind what the spinal surgeon hopes to achieve with regard to stability. Presently, the most popular systems for use posteriorly involve Luque rods and sublaminar wires. Other spinal SYstems include the use of multiple hooks and screws such as the Cotrel–Dubousette®, Texas Scottish Rite®, or Isola®. The use of pedicle screws has allowed instrumentation over relatively short segments. Rod and plate systems such as the Kaneda can be very beneficial, allowing the surgeon to decompress the spine and bring about stability through a single anterior approach (Fig. 18).



Fig. 17. Radiograph illustrating anterior and posterior approach to severe burst fracture for decompression reconstruction of all three columns and stabilization.

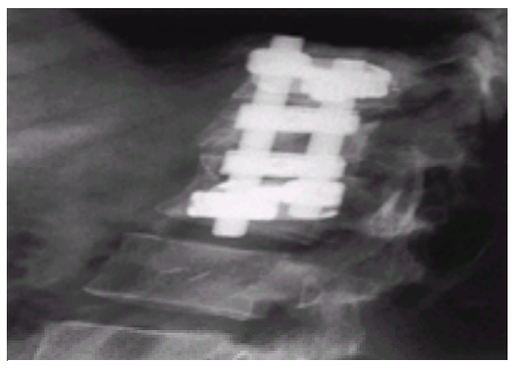


Fig. 18. Kaneda device allows the surgeon to provide excellent stabilization following decompression from an anterior approach.

Further reading

Spine injuries

Anderson LD, D'Alanzo RT. Fracture of the odontoid process of the axis. *Journal of Bone and Joint Surgery* 1974; **56A**: 1663–74. [This is a classic description of odontoid fractures.]

Bohlman HH. Acute fractures and dislocations of the cervical spine: an analysis of 300 hospitalized patients and review of the literature. *Journal of Bone and Joint Surgery* 1979; **61A**: 1119–42. [Excellent review article for fracture/dislocation of the cervical spine.]

Bozboga M, Unalf F, Hepgulk K, Izei N, Turantan I, Turker K. Fractures of the occipital condyle. *Spine* 1992; **17**: 1119–21. [Case report of previously unrecognized fractures.]

Clark CR, White AA, III. Fractures of the dens: a multicenter study. *Journal of Bone and Joint Surgery* 1985; **67A**: 1340–8. [Excellent multicenter study evaluating Dens fractures.]

Cotler JM, Herbison GJ, Nasuti JF, Ditunno J, An H, Wolff B. Closed reduction of traumatic cervical spine dislocation using traction weight up to 140 pounds. *Spine* 1993; **18**: 386–90. [Points pitfalls using closed reduction.]

Fielding WJ, Hawkins RJ. Atlantoaxial rotatory fixation. *Journal of Bone and Joint Surgery* 1977; **59A**: 37–44. [Description of rotatory subluxation of CI-C2.]

Kupchap, An HS, Cotler JM. Gunshot injuries of the spine. *Spine* 1990; **15**: 1058–63. [Large experience with gunshot injuries to the spine.]

Levine AM, Edwards CC. The management of traumatic spondylolisthesis of the axis. *Journal of Bone and Joint Surgery* 1985; **67**: 217–26. [Describes new subclass of hangman's fracture and treatment recommendations.]

Meyer PR. *Surgery of the spine in trauma*. Churchill Livingstone, New York, 1989. [Good reference for treatment of spinal trauma.]

Stauffer ES, Kelley EG. Fracture dislocation of the cervical spine: instability and recurrent deformity following treatment by anterior interbody fusion. *Journal of Bone and Joint Surgery* 1997; **59A**: 45–8. [Problems with anterior fusion for trauma.]

White AA, Johnson RM, Punjabi MM, Southwick W. Biomechanical analysis of clinical stability in cervical spine. *Clinical Orthopedics* 1975; **85**: 85–96. [Classic article in determining spinal instability.]

Thoracolumbar spinal injury

Benson ER. Unstable thoracolumbar fractures with emphasis on burst fractures. *Clinical Orthopedics* 1988; **230**: 14–29. [Clear and concise review of treatment of thoracolumbar fracture.]

Bohlman HH. Current concepts and review of treatment of fractures and dislocations of the thoracic and lumbar spine. *Journal of Bone and Joint Surgery* 1985; **67A**: 165–9. [Review article of thoracolumbar fractures.]

Bracken MB *et al.* A randomized controlled trial of methylprednisolone and naloxone in treatment of acute spinal cord injury: results of second national acute spinal cord injury study. *New England Journal of Medicine* 1990; **322**: 1405–11. [Study suggesting use of steroids for spinal cord injury.]

Carl AL, Tromanhauser SG, Roger DJ. Pedicle screw instrumentation for thoracolumbar burst fractures and fracture dislocations. *Spine* 1992; **17** (Suppl.). [Experience using pedicle screws in thoracolumbar fracture.]

Dennis F. Spinal instability as defined by three column spine concept and acute spinal trauma. *Clinical Orthopedics* 1984; **189**: 65–76.

Dewald RL. Burst fractures of the thoracic and lumbar spine. *Clinical Orthopedics* 1984; **189**: 158. [Overview of treatment of thoracolumbar fractures.]

Edwards C, Levine A. Early rods sleeve stabilization of injured thoracic and lumbar spine. *Orthopedic Clinics of North America* 1988; **17**: 121. [Review including rationale for a new spinal SYstem.]

Ferguson RL, Allen BL. Mechanistic classification of thoracolumbar spine fractures. *Clinical Orthopedics* 1984; **189**: 77. [More complicated classification of thoracolumbar fractures.]

Gertzbien SD. Scoliosis research society: a multicenter spine fracture study. *Spine* 1992; **17**: 528–40. [Large multicenter study comparing surgical and non-surgical treatment of thoracolumbar injuries.]

Green DA *et al.* Flexion distraction injuries of the lumbar spine associated with abdominal injuries. *Journal of Spinal Disorders* 1991; **4**: 312–18. [Noteworthy review of flexion distraction injuries.]

Holdsworth F. Fractures and dislocations and fracture dislocation of the spine. *Journal of Bone and Joint Surgery* 1970; **52A**: 1524. [Classic article and first description of treatment recommendations for fractures of the lumbar spine.]

Kane T, Capenda, Walters R, Zigler J, Adkins R. Spinal cord injury from civilian gunshot wounds: the Rancho experience, 1980–88. *Journal of Spinal Disorders* 1991; **4**: 306–11. [Showed no significant improvement in surgery of gunshot wounds to the spine.]

Kanada K, Abumi K, Fugiya M. Burst fractures with neurologic deficit of the thoracolumbar lumbar spine: results of anterior decompression and stabilization with anterior instrumentation. *Spine* 1984; **9**: 788–95. [Article describes excellent results using the Kanada anterior spinal system.]

Keene JS, Wackwitz DL, Drummond DS, Breed AL. Compression distraction instrumentation of unstable thoracolumbar fractures: anatomic results obtained with each type of injury and method of instrumentation. *Spine* 1986; **11**(9): 895–902.

Krompinger WJ, Frederikson BE, Mino DE, Yuan HA. Conservative treatment of fractures of the thoracic and lumbar spine. *Orthopedic Clinics of North America* 1986; **17**:161. [Study showed many burst fractures can be treated in a non-operative fashion.]

McAfee PC, Yuan HA, Lasda NA. Stable burst fractures. *Spine* 1982; **7**: 365.

McNamara MJ, Stevens GC, Spengler DM. Transpedicular short segment fusions for treatment of lumbar burst fractures. *Journal of Spinal Disorders* 1992; 5: 183–7.

Stevens TC, Devito DP, McNamara MJ. Segmental fixation lumbar burst fractures with Cotrel Dubosset instrumentation. *Journal of Spinal disorders* 1992; **5**: 344–8.

Walters RL, Adkins RH. Removal of bullet fragments retained in the spinal canal. *Spine* 1991; **16**: 934–9.

46.2.8 Inflammatory disorders of the spine

Thomas E. Whitesides

[Rheumatoid arthritis of the cervical spine](#)
[Anesthetic precautions](#)
[Ankylosing spondylitis](#)
[Treatment](#)
[Caution](#)
[Further reading](#)

While many systemic inflammatory diseases affect the axial skeleton with its varied complex of synovial and nonsynovial joints, cancellous and cortical structure, and complex coupled movements, rheumatoid arthritis and ankylosing spondylitis are by far the most common clinically. These two arthritides affect the spine in very different ways, both histologically, biomechanically, and clinically. Rheumatoid arthritis affects primarily the cervical spine and ankylosing spondylitis the entire axial skeleton.

Rheumatoid arthritis of the cervical spine

As in other areas of the body rheumatoid arthritis insidiously (by synovial destruction of joints, ligament, and bone), and generally early in the disease, begins in the O-C2 area and later more distally in the cervical spine.In 80 per cent of rheumatoid arthritics there is some involvement of C1–C2 by progressive destruction of the transverse ligament, allowing anterior subluxation (Fig. 1). This change may compromise the cervical cord by narrowing the space available between the arch of C1 and the posterior aspect of the odontoid or the body of C2. The lateral articulations of O–C1 and C1–C2 may also be destroyed, allowing axial settling between the occiput and C2 with cephalad protrusion of the odontoid process. This may even go through the foramen magnum causing midbrain dysfunction. Pannus may form anterior to the spinal cord causing compression of the cord itself. Later in the disease process the subaxial (C2–T1) area may be involved, causing the synovial joints to be posterior instead of anterior to the spinal cord. Thus subluxation may occur secondary to facet destruction posteriorly which may, in addition to other degenerative changes, cause cord or root compression.



Fig. 1. A 51-year-old woman with severe rheumatoid arthritis has C1–C2 pain and on flexion (a) – extension (b) demonstrates subluxation of 7 mm. While at this time she could be intubated by fiberoptic techniques and have surgery supine, care should be taken.

Early symptoms of upper cervical involvement are local pain; there may be progress later to quadraparesis due to central nervous system dysfunction. This may include involvement of lower cranial nerves or pain and weakness in the cervical roots. However, the subtle findings of clumsiness and spasticity of cord or midbrain compression are often misconstrued as being caused by the rheumatoid involvement of the limbs. Patients with significant upper cervical destruction and/or progressive subluxation at any level should be observed clinically and radiographically every 6–12 months.

Instability that is painful or causes neurologic compression may be an indication for surgery. The most frequent surgical procedure is a posterior fusion of C1–C1 with internal fixation and autogenous bone grafting. Other forms of bone graft or methylmethacrylate used instead of a bone graft have much less favorable results and are contraindicated. Immobilization is often by halo vest. Fusion may be required from the cranium down into the thoracic spine for more extensive disease. In the past, internal fixation was mainly with wire but is now moving toward lateral plate-and-screw and flexible cables. Anterior transoral or retropharyngeal approaches may be useful for anterior decompression and fusion from the occiput down, depending on the problem at hand, and are often concomitant with a posterior fusion.

These patients are all in poor general health involving many systems and are usually immunocompromised. They are at risk for many complications and their care must be gentle, thorough, and often unique and inventive.

Anesthetic precautions

In planning an anesthetic and/or surgery on a patient with rheumatoid arthritis, the status of the cervical spine must be assessed preoperatively and appropriate precautions taken, as inappropriate intubation and handling of the head and cervical spine may be treacherous. Fiberoptic intubation of the awake patient and then positioning for surgery with a check of the neural status before the induction of anesthesia are frequently used, as well as monitoring of spinal cord function (Fig. 2). Intubation is often hampered by rheumatoid destruction of the temporomandibular joints and posterior displacement of the mandible.



Fig. 2. This 40-year-old woman with juvenile rheumatoid arthritis since infancy has required fusion from occiput to T3 for progressive destructive changes. She has micrognathia due to destructive involvement of the temporomandibular joints. Unrecognized, this almost caused a fatality on induction of anesthesia for elective cholecystectomy. This was later accomplished by awake fiberoptic technique.

Ankylosing spondylitis

Ankylosing spondylitis is an inexorably progressive inflammatory disease of the spine and pelvis (axial skeleton) predominantly affecting white (caucasoid) males. It begins typically in the 20s as low back pain. Early on, the sacroiliac joints fuse and the disease progressively moves cephalad. The vertebrae take on a ‘squared off' anterior configuration and osseous bridging through ligaments and eventually disc spaces occurs. The costovertebral joints are involved early, limiting chest expansion significantly. The process may progress up into the cervical spine and even to the skull; however, the occipitocervical and C1–C2 joints are variably spared and

generally are the last to fuse. Kyphotic deformity is frequent. Fifty per cent of the patients eventually also get either shoulder or hip involvement (the hip is more frequently affected). Total fusion of the spine with kyphosis gives them a characteristic slow gait with excessive compensatory arm swinging. The addition of a kyphotic spine to flexion contracture of the hip often leaves them with grotesque posture.

Treatment

Medical treatment is SYmptomatic and, since the disease is chronic and lifelong, any but short-term use of narcotics is inappropriate. The deformity may become gross, with a ‘chin on chest' deformity that can limit swallowing. The rigid thorax and kyphotic lumbar spine make all breathing abdominal. Obesity, tight pants (trousers), body casts or orthoses thus limit respiration. Abdominal surgery is thus a problem as there is limited intra-abdominal space.

Surgical treatment

In general, surgery is of three types. As flexion deformity of the hips associated with hip degeneration is an added problem, total hip replacement is common. Osteotomies of the cervical thoracic area at C7–T1 from the posterior with an anterior controlled osteoclasis are done for cervicothoracic deformity and can correct up to 60°. In the lumbar area, such osteotomies are generally done at L2 or L3. The goal is to allow the patient to stand erect, be able to eat, and to increase intra-abdominal space. Modern internal fixation and halo devices have generally done away with body casts and allow early postoperative mobilization instead of prolonged recumbency until fusion occurs.

Caution

The spine in patients with ankylosing spondylitis is osteopenic and rather brittle, and thus is subject to fracture from falls or other relatively minimal traumas. These breaks are often subtle but all are dangerous if not appropriately treated; progressive deformity and other complications are commonly missed. If a patient with ankylosing spondylitis falls or is in an accident and has spinal pain, it is a fracture until proved otherwise by a negative bone scan and/or computed tomography. The late complications of untreated fracture may be severe.

Anesthesia in ankylosing spondylitis may be difficult and postoperative problems involving an abdominal incision (since the patients are abdominal breathers only) require special care (Fig. 3).

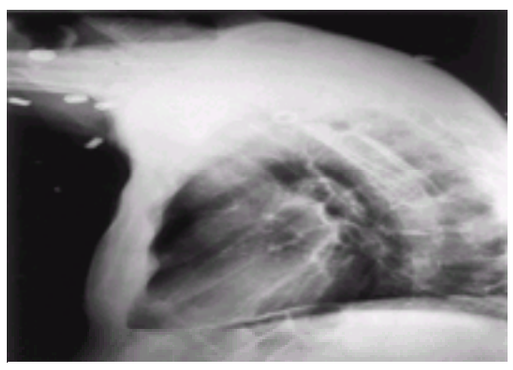


Fig. 3. This frail, 78-year-old man with a 50-year history of ankylosis spondylitis has a totally rigid spine, pelvic, and rib cage. All respiration requires abdominal excursions. Any abdominal surgical procedure would place him at respiratory risk postoperatively and probably could not be done laparoscopically.

Further reading

Boden SP, Dodge LD, Bohlman HH, Rechtine GR. Rheumatoid arthritis of the cervical spine: a long term analysis with predictors of paralysis and recovery. *Journal of Bone and Joint Surgery (A)* 1993; **75**: 1283–97.

Clark CR, Goetz DD, Menezes AH. Arthrodesis of the cervical spine in rheumatoid arthritis. *Journal of Bone and Joint Surgery (A)* 1988; **70**: 382–92.

Lipson SJ. Rheumatoid arthritis of the cervical spine. *Clinical Orthopedics* 1982; 143–9.

Santavirta S *et al.* Treatment of the cervical spine in rheumatoid arthritis. *Journal of Bone and Joint Surgery (A)* 1988; **71**: 568–667.

Simmons EH. Surgery of the spine in rheumatoid arthritis and ankylosing spondylitis. In: Evarts CM, ed. *Surgery of the musculoskeletal system*, Vol. II, p. 85. Churchill Livingstone, New York, 1983.

Zoma A *et al.* Surgical stabilization of the rheumatoid cervical spine: a review of indications and results. *Journal of Bone and Joint Surgery (B)* 1987; **69**: 8–12.

46.2.9 Spinal infections

Mitchell F. Reiter and John G. Heller

Introduction
Pyogenic vertebral osteomyelitis
Clinical presentation
Diagnostic studies
Treatment
Epidural abscess
Clinical presentation and diagnostic studies
Treatment and prognosis
Intervertebral discitis and osteomyelitis of childhood
Clinical presentation and diagnostic studies
Treatment and prognosis
Granulomatous diseases of the spine
Clinical presentation and diagnostic studies
Treatment and prognosis
Postoperative spinal infections
Postoperative discitis
Postoperative wound infection
Conclusion
Further reading

Introduction

With their high potential for morbidity and mortality, infections of the spine demand prompt diagnosis and treatment. Variable clinical presentations and natural histories merit familiarity with the different types of spinal infections. This includes an appreciation for the different ways of classifying such processes. The most basic distinction is based on the host response to the infecting organism. Bacterial agents will typically lead to a pyogenic host response while mycobacteria and fungi produce a granulomatous reaction. Infection can also be classified by anatomic location with possible involvement of the vertebral bodies, intervertebral disc, epidural space, or adjacent soft tissues. The route of infection such as direct extension, postoperative, or hematogenous spread can also be used to classify spinal infections. Finally, age can be a definitive factor with distinctly different natural histories noted for pediatric and adult spinal infections.

Pyogenic vertebral osteomyelitis

Pyogenic vertebral osteomyelitis represents up to 8 per cent of all cases of osteomyelitis. The most common etiology of these infections appears to be from the entrapment of septic emboli in the rich end-arteriole networks found in the metaphyseal region of the vertebral bodies. Extension of the infection into the disc spaces and paravertebral tissues can then occur secondarily. In the past, an overemphasis was placed on the possibility of infection via seeding of the spine with septic emboli from the pelvis carried through Batson's valveless paraspinal venous plexus. Dye injection studies, however, have shown that extremely high pressures are required to cause retrograde flow through this system and that this mode of infection is probably much less common than previously believed.

Any process that can give rise to a bacteremia may predispose a patient to vertebral osteomyelitis. Urinary tract manipulation or infection, cutaneous ulcers or abscesses, and respiratory infection have all been cited as significant risk factors. Contiguous spread of visceral infections can lead to infection of adjacent spinal segments. Penetrating injuries such as gun shot wounds can produce spinal infections with concomitant colonic injury significantly increasing the risk of this possibility. Immunocompromised patients, intravenous drug abusers, and those with diabetes are particularly prone to developing spinal infection. The lumbar spine is the region of the spine involved approximately one-half of the time, followed in frequency by the thoracic spine. The cervical spine is involved in less than 10 per cent of cases.

Most vertebral column infections are due to Gram-positive cocci with *Staphylococcus aureus* accounting for more than 50 per cent of reported cases. Infections arising from the urinary tract are more commonly associated with Gram-negative organisms such as *Escherichia coli, Pseudomonas*, and *Proteus* species. Intravenous drug abusers are particularly prone to infection with *Pseudomonas* species. Superinfection of underlying granulomatous infections by *Staphylococcus epidermidus* following biopsy can be particularly troublesome. One must always keep in mind that local factors and altered host defenses will affect the susceptibility to and the bacteriology of vertebral osteomyelitis.

Clinical presentation

The symptoms and signs of vertebral osteomyelitis are highly variable. Making a timely diagnosis depends on maintaining an appropriate level of suspicion. The presentation may be acute, subacute, or chronic depending on the infecting organism and the host resistance. Back pain is the most common complaint and is present in more than 90 per cent of patients. The pain is typically intense, focal, and unrelated to activity. It may awaken the patient from sleep. It is usually relentless in nature. Associated constitutional symptoms can vary greatly, but may include chills, fever, diaphoresis, anorexia, and malaise. In patients with other illnesses, it may be hard to distinguish the etiology of these SYmptoms from other causes. Less than 50 per cent of patients will present with an elevated body temperature. Neurologic deficits occur in less than 10 per cent of patients with vertebral osteomyelitis. Factors that have been cited as increasing the risk of neurologic deficit include advanced age, a more cephalad level of infection, diabetes, rheumatoid arthritis, and infection with *Staph. aureus*.

The physical examination is frequently non-specific. Palpable tenderness and warmth may be present. Muscle spasm can cause decreased range of motion and loss of lumbar or cervical lordosis. A psoas sign may be present with pain provoked by hip extension. Long-standing or fulminant cases may present with visible structural deformities. A careful neurologic examination should be performed with emphasis on tension signs such as the straight leg raise test. Meningeal signs should also be sought since a subset of patients may have an associated meningitis.

Diagnostic studies

Laboratory studies may be helpful but are non-specific. The erythrocyte sedimentation rate and C-reactive protein are elevated in over 90 per cent of patients. They are useful in supporting diagnosis and monitoring the course of treatment. The leukocyte count is elevated less than half of the time. Blood cultures should be obtained, but these are positive in less than 25 per cent of cases. Diagnostic imaging techniques are essential in the diagnosis and treatment of pyogenic vertebral osteomyelitis. Plain radiographs will typically not be helpful for the first 2 to 3 weeks following the onset of infection. Later changes include osteolysis in the anterior vertebral metaphyseal region, loss of disc space height, end plate erosion, and subchondral reactive bone formation. The radiographs may also reveal soft tissue abnormalities such as a paraspinal mass or psoas shadow. Eventually, extensive bony destruction with pathologic fracture and resulting kyphotic deformity or instability may occur. It may be difficult to distinguish these radiographic findings of infection from those of a neoplastic process such as a tumor or metastasis, but relative sparing of the disc space is more consistent with tumor.

Radionucleotide studies can detect an infectious process before plain radiographic changes are present. Technetium bone scanning is highly sensitive (90 per cent) for pyogenic vertebral osteomyelitis and when combined with gallium scanning has a reported accuracy of 94 per cent. Indium-labeled white blood cell scans are notoriously inaccurate when used to detect spinal infection, having an accuracy of only 31 per cent. It is important to note that radionucleotide scans typically take from several hours to days to obtain and that the results can be significantly altered by previous treatment with antibiotics. Computed tomography (**CT**) scans are most helpful in delineating bone involvement, which may be overestimated by magnetic resonance imaging (**MRI**). With the addition of intrathecal contrast, CT myelography can provide useful information pertaining to the neural anatomy. At the time of lumbar or cervical puncture for myelographic purposes, cerebrospinal fluid should be sent for Gram stain, acid-fast bacilli stain, cultures, cell count, glucose, and protein to rule out meningitis.

MRI is the technology of choice for diagnosing spinal infections. MRI allows non-invasive, rapid, early detection of infection with a reported accuracy of 95 per cent. Additionally, useful information is provided about the anatomic extent of involvement and any soft tissue abscess formation. MRI signal changes characteristic of infection include decreased signal intensity on T_1-weighted sequences of both the vertebral body and disc with an indistinct margin between the two. Conversely, T_2-weighted images have an increased signal intensity in both the vertebral body and disc (Fig. 1). Addition of gadolinium contrast is very helpful in the evaluation of postoperative infections and epidural abscesses. If an infection is suspected, the physician should specify that the MR be done with and without gadolinium to avoid false negatives.

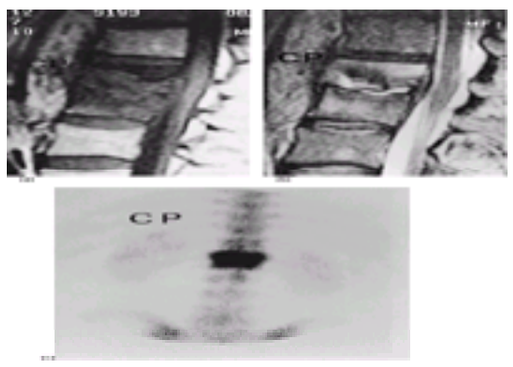


Fig. 1. (a) T_1 - and (b) T_2-weighted sagittal MR sequences demonstrating discitis with surrounding osteomyelitis. On the T_1-weighted image note the decreased signal intensity of both the vertebral body and disc with an indistinct margin between the two. The T_2-weighted images have an increased signal intensity in both the vertebral body and disc. (c) Positive technetium bone scan image of discitis in the same patient.

Definitive diagnosis of vertebral osteomyelitis usually requires a biopsy to obtain a tissue sample from the lesion for bacteriologic and microbiologic evaluation. A situation in which a biopsy can be avoided is when blood cultures are positive in conjunction with appropriate clinical and radiographic findings. Unfortunately, since blood cultures are positive in less than 25 per cent of cases, a biopsy is usually necessary. Needle biopsy under CT or fluoroscopic guidance has been shown to be a safe technique with an accuracy reported between 68 and 86 per cent. When an adequate diagnosis cannot be made with closed biopsy techniques an open biopsy becomes necessary. Samples from the biopsy should be sent for all routine cultures and stains, and must be sent to pathology to rule out malignancy. Rarely, one might encounter a tumor that has become secondarily infected.

Treatment

There are several goals of treatment which must be considered when managing patients with vertebral osteomyelitis. These goals include: (i) establishing a definitive bacteriologic diagnosis; (ii) eradicating infection and preventing recurrence; (iii) preventing or reversing neurologic deficit; (iv) maintaining or re-establishing spinal stability; and (v) providing pain relief. Under a variety of circumstances all of these goals can be achieved non-operatively. Once an organism has been identified by either blood culture or biopsy, treatment begins with intravenous antibiotics. If at all possible, antibiotics should be withheld until an organism is identified. Parenteral antibiotics should usually be continued for 6 weeks and certainly for not less than 4 weeks. During treatment, periodic erythrocyte sedimentation rates and C-reactive protein levels are useful for monitoring the response to therapy. In addition to antibiotics, treatment should include immobilization of the spine. This can be accomplished with molded total-contact orthoses in the lumbar and thoracic spine and a cervicothoracic orthosis or halo vest for cervical or upper thoracic osteomyelitis. Additionally, treatment should include attention to the patient's nutritional status as well as aggressive management of any other comorbid conditions such as diabetes or rheumatoid arthritis.

When planning or carrying out treatment there are several indications for surgical intervention. These include: (i) inability to obtain a bacteriologic diagnosis by closed means; (ii) persistent sepsis or abscess formation; (iii) significant deformity, instability, or advanced bone destruction, including pathologic fracture, caused by the infection; (iv) neurologic deficit caused by compression at spinal cord level or a progressive deficit caused by compression at nerve root level; and (v) failure of medical management. The surgical principles governing the treatment of vertebral osteomyelitis are similar to those for benign, locally aggressive neoplasms. All infected tissue should be thoroughly debrided to render a surgically clean wound. This should include removal of all necrotic bone, disc, and soft tissue as well as decompression of any abscesses. Decompression of the spinal canal is another essential component of treatment. Following debridement and decompression the vertebral column defect must be reconstructed to provide lasting spinal stability.

As the infectious process most often affects the vertebral body, the goals of surgical treatment are readily accomplished by approaching the spine anteriorly. In the lumbar spine, retroperitoneal exposure provides excellent access to the spine, especially in the presence of a psoas abscess. Transperitoneal exposure is most often useful for lesions involving the lumbosacral junction. In the thoracic spine the transthoracic approach is most applicable, although under certain circumstances a costotransverectomy or lateral extracavitary approach may be useful. Cervical lesions are almost always best addressed via a standard anterior cervical approach.

Reconstruction of the spinal column following debridement can be a complex task. Ideally, autogenous iliac crest, rib, or fibula is used in the reconstruction. These may be combined with supplemental fresh frozen allograft as necessary. Anterior instrumentation with or without supplemental posterior segmental fixation is usually necessary. Additionally, all of the standard principles of spinal reconstruction must be meticulously observed when addressing these challenging problems. With timely diagnosis and proper operative or non-operative treatment, the mortality rate of pyogenic vertebral osteomyelitis should be less than 5 per cent. Spontaneous fusion of affected levels also occurs frequently following successful medical treatment and is painless in the majority of cases.

Epidural abscess

Spinal epidural abscess is a rare, but potentially devastating, type of spinal infection. The epidural space can become infected via hematogenous spread from distant sites, contiguous spread from adjacent structures such as vertebral osteomyelitis, or by direct inoculation of the region from spinal surgery or invasive procedures such as epidural anesthetics. The peak incidence of this condition occurs during the fifth or sixth decades of life and is associated with the same known risk factors for other types of spinal infection. The thoracic spine is affected half of the time, followed in frequency by the lumbar and cervical spines, respectively. Abscess formation has been noted to occur posteriorly in 79 per cent of cases and anteriorly in 21 per cent. However, when associated with discitis or osteomyelitis, it is almost always anterior. Concurrent meningitis is present in up to 13 per cent of patients. *Staph. aureus* is the most common organism identified.

Clinical presentation and diagnostic studies

Failure to consider the diagnosis and the high prevalence of axial spine pain in the general population make delay in diagnosis the rule rather than the exception. Severe, localized back pain is the usual initial complaint. Fever and constitutional SYmptoms are present in 50 per cent of patients. The progression of the disease is rather predictable. Over a variable time course, the axial pain will intensify, later progressing to radicular pain and, ultimately, rapid and frequently irreversible spinal cord deficits.

Laboratory findings of sepsis are found in most patients. An elevated leukocyte count, erythrocyte sedimentation rate, and C-reactive protein are all usually present. MRI with gadolinium is the imaging modality of choice to look for an epidural abscess (Fig. 2). MRI is highly sensitive and provides useful information in planning surgical treatment of the infection. Plain radiographs are usually negative unless concomitant vertebral osteomyelitis is present. Radionucleotide studies are not relevant to making this diagnosis. Prior to MRI, myelography with postmyelography CT was the best means of diagnosis and planning treatment. Cerebrospinal fluid should be sent for analysis at the time of myelography to evaluate for meninigitis.

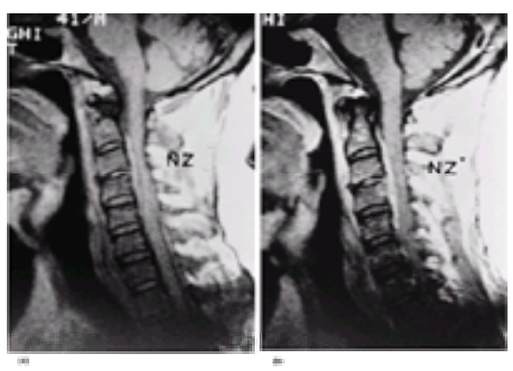


Fig. 2. Ventral cervical epidural abscess associated with hematogenous intervertebral discitis. Sagittal MR images (a) before and (b) after the administration of gadolinium. Note the significant enhancement of the abscess with use of paramagnetic contrast.

Treatment and prognosis

Treatment of an epidural abscess is a surgical emergency. Because the prognosis is directly related to the speed of diagnosis and intervention, delay in treatment should be avoided at all costs. Once diagnosed and localized with some type of neurodiagnostic imaging modality, the abscess should be surgically decompressed. Posterior epidural abscesses can usually be effectively treated with a laminectomy of all affected levels with debridement of all involved tissues. Care should be taken during laminectomy to preserve the facet joints and capsules so as not to destabilize the spine. Anterior epidural abscesses should be approached anteriorly via corpectomies and discectomies of the affected levels. The defect created should then be reconstructed with autogenous bone graft if possible and anterior instrumentation if deemed necessary. Broad-spectrum parenteral antibiotics that cross the blood–brain barrier should be started once a culture has been obtained and then tailored based on the final bacteriologic diagnosis and sensitivity profile. Antibiotics should be continued for 2 to 4 weeks depending on the patient's response.

With prompt diagnosis and appropriate treatment, the mortality rate for epidural abscesses has been reduced to less than 20 per cent. The prognosis for neurologic recovery from paralysis is most dependent on the duration and severity of the paralysis. One review of the literature reported complete recovery in 38 per cent of cases, residual weakness in 29 per cent, permanent paralysis in 21 per cent, and death in 12 per cent. Poor prognostic factors include rapidly progressing paralysis, complete paralysis, and paralysis lasting more than 36 h.

Intervertebral discitis and osteomyelitis of childhood

Historically, there has been some controversy as to the exact etiology of pediatric discitis; however, it is now generally accepted that there is a bacterial component to the process and that in children discitis and vertebral osteomyelitis represent points on a continuum of disease. This continuum may be more accurately called pyogenic infectious spondylitis. Transient bacteremias lead to seeding of either the disc space or vertebral endplate, both of which remain vascularized in early childhood. There is then rapid progression to the classic concomitant vertebra–disc–vertebra unit involvement seen on imaging studies. Pediatric pyogenic infectious spondylitis can be classified into three groups: (i) severely destructive vertebral osteomyelitis in septicemic neonates; (ii) more benign infections of infants and children commonly referred to as discitis; and (iii) a more morbid type of vertebral osteomyelitis found in older children and adolescents similar to adult vertebral osteomyelitis.

Clinical presentation and diagnostic studies

The clinical presentation varies with the patient's age. Children younger than 3 years are unable to articulate their symptoms and generally present with abnormalities of gait or refusal to walk. Children aged 3 to 8 years may complain of abdominal pain which may be suggestive of an acute abdominal process such as appendicitis, mesenteric adenitis, or inflammatory bowel disease. Teenagers frequently will have the more classic complaints of back pain, radicular pain, or spinal deformity. Half of children are febrile and even fewer are SYstemically ill. Physical examination will usually demonstrate pain with local palpation or percussion of the spine. Deep palpation of the abdomen may reveal tenderness in the midline along the vertebral column which may be hard to distinguish from pain associated with other intra-abdominal processes. Neurologic deficits are rare, but tension signs such as a positive straight leg raise or tight hamstrings may be noted.

Laboratory studies are of limited help. The erythrocyte sedimentation rate and C-reactive protein are usually elevated and the white blood cell count is highly variable. Blood cultures are positive in approximately half of patients. Plain radiographs frequently are not helpful unless the infection has been present for 2 to 3 weeks, after which time characteristic destruction of the disc space will be noted. Occasionally, an early chest radiograph may reveal paraspinal soft tissue swelling or a psoas shadow may be seen on a KUB. Radionucleotide imaging is quite helpful with technetium bone scans providing rapid localization of the infectious process. As with the other forms of spinal infection, MRI is the imaging modality of choice. The MRI findings of pediatric infectious spondylitis are virtually pathognomonic and MRI helps exclude the possibility of other diagnoses such as spinal neoplasm.

Treatment and prognosis

Treatment is medical in the majority of pediatric patients. Since *Staph. aureus* is the responsible organism in 80 to 90 per cent of culture-positive cases, empiric intravenous antibiotics to treat a presumptive staphylococcal infection should be initiated. Direct biopsy should be reserved for those cases not responding to treatment or when some other process such as a malignancy is suspected. Oral antibiotics may be substituted after clear clinical improvement occurs, which is usually within 4 to 7 days. A 4-week course of antibiotics is generally accepted. A short period of bed rest may be used for symptomatic relief and bracing is usually not necessary. Patients should be followed on a long-term basis to evaluate for growth arrest and progressive deformity. Surgery is reserved for those patients who fail medical treatment or those who present with advanced disease such as vertebral collapse, neurologic deficit, or abscess formation.

Granulomatous diseases of the spine

The most common pathogen which causes a granulomatous reaction in the spine is *Mycobacterium tuberculosis*, which is responsible for spinal tuberculosis or Pott's disease. Unfortunately, spinal tuberculosis is undergoing a resurgence in developed countries due to an increasing population of immunocompromised patients, including those with AIDS. Approximately 10 per cent of all patients with tuberculosis have skeletal involvement and, of these, 50 per cent have lesions involving the spine. Neurologic deficits are found more commonly in spinal tuberculosis infections than in other types of spinal infection and are present in between 10 and 45 per cent of patients. Pathologically, the typical caseating granulomas with Langerhans giant cells are seen. Acid-fast stains may identify these organisms although polymerase chain reaction techniques are more sensitive. The thoracic spine is most frequently involved, followed by the lumbar spine and least frequently, the cervical spine.

Spinal tuberculosis has several major patterns of involvement. The paradiscal type is the most common and accounts for over one-half of cases. A second pattern includes primarily anterior involvement with an abscess tracking under the anterior longitudinal ligament. A third, central pattern of infection begins in the middle of a vertebral body and may be mistaken for a tumor. Tuberculosis infection proceeds at a slower pace than pyogenic processes and, consequently, radiographic disc space involvement is not seen until later in the course of infection. Extensive bone destruction with significant deformity as well as large paraspinal abscesses are also more commonly seen with tuberculosis than with pyogenic spinal infections.

Clinical presentation and diagnostic studies

Spinal tuberculosis varies widely in its presentation. The most common complaint is that of slowly progressive back pain. Late in the disease process deformity and neurologic deficits may be obvious. Classic constitutional symptoms such as fever, chills, and weight loss may or may not be present. Physical examination reveals spinal tenderness similar to that seen with other types of spinal infections. Laboratory testing including a white blood cell count, erythrocyte sedimentation rate, and C-reactive protein are non-specific in distinguishing tuberculosis from pyogenic infection. Skin testing for purified protein derivative is usually positive, indicating active or previous disease, but may be falsely negative in immunocompromised patients who have become anergic. Chest radiographs should be taken and evaluated for active lung disease.

Plain radiographs may reveal similar changes to those seen with pyogenic vertebral osteomyelitis. Extensive destruction, which appears out of proportion to pain, as well as involvement of many levels of the spine are findings suggestive of tuberculosis infection. As mentioned above, relative preservation of the disc spaces and large abscess formation are characteristic of tuberculosis and other granulomatous infections. Radionucleotide imaging is not usually helpful and one should proceed straight to more direct imaging modalities. CT or MRI scanning are complimentary studies which provide excellent anatomic detail and help plan the next step in the work-up, which is the aspiration biopsy (Fig. 3).



Fig. 3. Axial CT scan showing large paraspinal abscesses (arrowheads) with heterotopic bone characteristic of tuberculous osteomyelitis.

Treatment and prognosis

Following biopsy, one must keep in mind that the differential iagnosi of graulomatos infection is quite extensive. Chronic infection with granulomatous reaction can be caused by atypical bacteria such as *Actinomyces, Nocardia*, and *Brucella*. Atypical mycobacteria must also be considered, especially in immunocompromised patients. Fungal infections such as coccidioidomycosis, blastomycosis, cryptococcus, aspergillosis, and candidiasis must be ruled out in the appropriate clinical and geographic setting. Once a definitive diagnosis of tuberculosis is made, medical treatment is usually begun. In most areas the antimicrobial regimen consists of four drugs including isoniazid, rifampin, ethambutol, and pyrazinamide. Therapy should continue for 6 to 12 months. Consultation with an infectious disease specialist is recommended to help consider regional susceptibility patterns and the most effective treatment regimens. Chemotherapy should be supplemented with external immobilization if deemed necessary.

Indications for surgical intervention include those patients with a significant or worsening neurologic deficit, especially in the cervical spine. Other indications for surgery include extensive destruction with deformity or instability as well as failure of medical treatment. Unlike pyogenic vertebral osteomyelitis, abscess formation is not necessarily an indication for immediate surgical intervention. Decisions for medical rather than surgical treatment are complex and multifactorial. They must be made in the context of the clinical setting as well as the available medical resources in the patient's country. When operative treatment is necessary, the surgical principles are the same as those applied for the management of other types of spinal infection. Anterior debridement with reconstruction using autogenous bone graft when possible, with or without internal fixation or supplemental posterior fusion, is usually the treatment of choice. In the absence of other major comorbidities, the mortality rate for spinal tuberculosis should be less than 5 per cent with a relapse rate of 10 to 20 per cent.

Postoperative spinal infections

Postoperative spinal infections can be a devastating complication that severely compromises the result of any surgical procedure. Careful preoperative assessment, intraoperative technique, and postoperative diligence can help minimize the risk and consequences of this difficult problem. The risk of infection from conventional discectomy is cited at 1 per cent or less. When a fusion is added to the procedure the risk of infection increases to 2 per cent and can go as high as 5 to 6 per cent when instrumentation is used. Cervical procedures tend to have a slightly lower infection rate than lumbar procedures. Use of prophylactic antibiotics has been shown to be the single most important measure for preventing postoperative infections. Proper surgical technique including debridement of necrotic tissues, intermittent release of retractors, and careful hemostasis can also help reduce the risk of infection. Patient-related risk factors for infection include: diabetes, chronic steroid use, rheumatoid arthritis, immunosuppression, and a history of previous infection. Additionally, poor nutritional status has been correlated with an increased risk of postoperative infection. Nutritional status can be determined using parameters such as recent weight loss, total lymphocyte count, serum albumin level, and serum transferrin level. Several types of postoperative infections exist, including postoperative discitis, superficial, and deep wound infections. If treatment of any of these infections is delayed or inadequate, vertebral osteomyelitis with all of its sequelae may result.

Postoperative discitis

After a discectomy, postoperative discitis usually presents as increasing low back pain which begins following an initial period of pain relief. The wound characteristically appears benign except for mild local tenderness. A low-grade fever may be present, but other constitutional symptoms are uncommon. The white blood cell count is usually normal, but the erythrocyte sedimentation rate and C-reactive protein are characteristically elevated. Plain radiographs are of little help in the immediate postoperative period. Several weeks later, radiographs will show the typical changes of a disc space infection. MRI with gadolinium is the imaging modality of choice and allows early accurate diagnosis of this condition.

Treatment begins with percutaneous biopsy to determine the infecting organism. If the organism can be identified, non-surgical treatment may be initiated. Medical management consists of intravenous antibiotics for 2 to 6 weeks, spinal immobilization, and bed rest as needed. Indications for surgical treatment include: failure to determine the infecting organism by percutaneous means, sepsis, neurologic deficit, vertebral collapse, or epidural abscess. Additionally, surgical treatment should be undertaken for patients with a delayed or inadequate response to medical treatment as well as for those who show progressive bone destruction or have intractable pain (Fig. 4). To undertreat this entity is a serious mistake that may result in a significant increase in the ultimate morbidity caused by the infection.

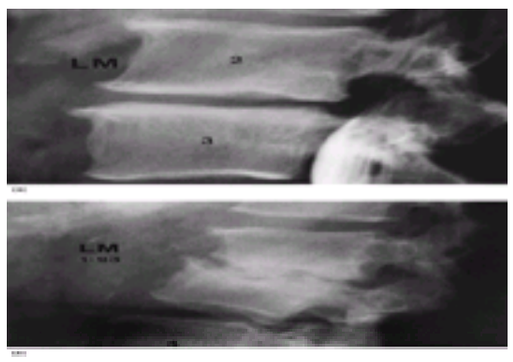


Fig. 4. (a) Preoperative lateral myelogram of L2–3 disc herniation.(b) Lateral radiograph of postoperative disc space infection following treatment with intravenous antibiotics. Note the failure of treatment with extensive bony destruction (osteomyelitis) and kyphotic deformity. Early operative intervention would have prevented this degree of progression.

Postoperative wound infection

Postoperative wound infections can be classified as superficial or deep and typically occur within 2 weeks of surgery, but may present much later. Worsening pain or pain out of proportion to that expected are both warning signs of a wound infection. Constitutional symptoms and elevated white blood cell count, erythrocyte sedimentation rate, and C-reactive protein are all also commonly associated with a postoperative wound infection. Inspection of the wound may reveal a tender, erythematous, or draining incision site, but with deep infections the wound may appear unremarkable. Diagnosis of a postoperative wound infection is usually made clinically. MRI scanning may help localize fluid collections and evaluate for associated osteomyelitis or discitis in certain selected cases. One should not hesitate to perform a bedside or CT-guided sterile wound aspiration when infection is suspected but cannot be confirmed clinically.

Treatment of postoperative wound infections requires an aggressive surgical approach. Medical treatment is unlikely to succeed and is highly discouraged. Patients should be returned to the operating room for irrigation and debridement of the wound and intravenous antibiotics should be started. At the time of debridement, instrumentation and adherent bone graft should be left in place and the wound should usually be closed primarily over drains. Chronically infected wounds may require removal of instrumentation if the fusion is solid. A repeat debridement of the wound should probably be done 48 h later and antibiotics specific for the organism cultured should be continued for 2 weeks in superficial infections and for 6 weeks in deep infections, especially in the presence of bone graft and/or implants.

Conclusion

Infections of the spine have a wide variety of manifestations and continue to complicate the lives of both patients and physicians. Awareness of the various presentations and imaging modalities available can help expedite diagnosis and appropriate management. Treatment in every case must be tailored to the affected structures, the patient, and to the offending organism.

Further reading

Carragee EJ. Pyogenic vertebral osteomyelitis. *Journal of Bone and Joint Surgery* 1997; **79A**: 874–80.

Crawford AH, Kucharzyk DW, Ruda R, Smitherman HC. Discitis in children. *Clinical Orthopedics* 1991; **266**: 70–9.

Currier BL, Eismont FJ. Infections of the spine. In: Rothman RH, Simeone FA, eds, *The spine*, 3rd edn, pp. 1319–80. WB Saunders, Philadelphia, 1992.

Danner RL, Hartman BJ. Update of spinal epidural abscess: 35 cases and review of the literature. *Review of Infectious Disease* 1987; **9**: 265–74.

Eismont FJ, Bohlman HH, Soni PL, Goldberg VM, Freehafer AA. Pyogenic and fungal vertebral osteomyelitis with paralysis. *Journal of Bone and Joint Surgery* 1983; **65A**: 19–29.

Emery SE, Chan DP, Woodward HR. Treatment of hematogenous pyogenic vertebral osteomyelitis with anterior debridement and primary bone grafting. *Spine* 1989; **14**: 284–91.

Heller JG. Infections of the cervical spine. In: An HS, Simpson JM, eds, *Surgery of the cervical spine*, pp. 335–56. Martin Dunitz, London, 1994.

Heller JG. Postoperative infections of the spine. In: Rothman RH, Simeone FA, eds, *The spine*, 3rd edn, pp. 1817–37. WB Saunders, Philadelphia, 1992.

Heller JG, Levine MJ. Spinal infections. In: Garfin SR, Vaccaro AR, eds, *Orthopaedic knowledge update spine*, pp. 257–71. American Acadamy of Orthopedic Surgeons, Rosemont, 1997.

Ho EKW, Leong JCY. Tuberculosis of the spine. In: Weinstein SL, ed., *The pediatric spine: principles and practice*, pp. 837–50. Raven Press, New York, 1994.

Hsu LC, Cheng CL, Leong JC. Pott's paraplegia of late onset: the cause of compression and results after anterior decompression. *Journal of Bone and Joint Surgery* 1988; **70B**: 534–8.

Petty W, Spanier S, Shuster JJ, Silverthorne C. The influence of skeletal implants on incidence of infection. *Journal of Bone and Joint Surgery* 1985; **67A**: 1236–44.

Vincent KA, Benson DR. Differential diagnosis and conservative treatment of infectious diseases. In: Frymoyer JW, ed., *The adult spine: principles and practice*, pp. 763–86. Raven Press, New York, 1991.

Vincent KA, Benson DR. Infectious diseases of the spine: surgical treatment. In: Frymoyer JW, ed., *The adult spine: principles and practice*, pp. 787–810. Raven Press, New York, 1991.

Wenger DR, Davids JR, Ring D. Discitis and osteomyelitis. In: Weinstein SL, ed., *The pediatric spine: principles and practice*, pp. 813–36. Raven Press, New York, 1994.

46.2.10 Spinal cord monitoring

Jerrold Rosenberg

Introduction
Physiology and instrumentation
Methodology
Changes affecting evoked potentials
Clinical settings
Conclusion
Further reading

Introduction

Evoked potentials are the electrical manifestations generated by the nervous system in response to a brief external sensory stimulus. This stimulus may be somatic, dermatomal, or cutaneus.

During orthopedic surgery for spinal deformities the most feared complication is postoperative paraplegia. For adults undergoing unmonitored correction of scoliosis, the loss of motor function in the lower extremities ranges from 1 to 1.7 per cent. A lesser rate of loss occurs in children. In over one-half of these cases the paraparesis is permanent. This impairment may be caused by such correctable factors as circulatory impairment, compression of the cord from bony structures or hematomas, trauma during fitting of orthopedic instrumentation, or mechanical stretching of the spinal cord and peripheral nerve. Evoked potentials provide a way to identify the neurologic impairment early enough to allow prompt correction of the cause.

Until the advent of evoked potential monitoring the best alternative was the wake-up test. The patient is awakened at the end of instrumentation while on the operating room table and with the surgical site exposed. In this semiconscious state, the patient is asked to wiggle their toes as a test of motor function. There are several problems with this technique including the need for an experienced anesthetist: one capable of quick Jaily reversing the anesthesia, awakening the patient, and then reanesthetizing the patient. In addition, there have been false-negative results in which, despite a normal wake-up test, the patient has permanent postoperative neurologic impairment. Nightmares and patient injuries due to agitation have also been reported. The greatest inherent difficulty with the wake-up test, however, is that it is a single moment in time and fails to identify problems that may develop before or after wake-up. In contrast, evoked potential monitoring does not require awakening the patient and can be performed safely throughout the operation as well as in the recovery room. In 1998, state-of-the-art spinal cord surgery should include intraoperative evoked potential monitoring.

Physiology and instrumentation

Dermatomal evoked potentials are initiated by electrical stimulation of dermatomal signature areas and offer the most segmental specificity. While several reports including Coscia (Indiana) report good results, the technique is difficult to perform and recordings acceptable for monitoring were found less than 71 per cent of the time. Dumitru (Texas) and Snowden (Seattle) report a higher degree of sensitivity but not in the operating room setting. In contrast, cutaneous evoked potentials are initiated by electrical stimulation of the cutaneus nerve. They are difficult to obtain and also not yet consistent enough to be utilized for intraoperative procedures. The most common method involves the use of somatosensory evoked potentials (**SEPs**). These potentials are initiated by electrical stimulation of any peripheral sensory nerve. This stimulus excites the largest myelinated afferent fibers and travels to the dorsal root ganglion, then ipsilateral in the posterior column to SYnapse in the dorsal column nuclei of the nucleus cuneatus and nucleus gracilis. The potential crosses in the medial lemniscus to the ventral posterior lateral nucleus of the thalamus. After a second SYnapse in the thalamus the third sensory fiber travels directly to the sensory cortex. The central nervous system has a built-in amplifier effect, increasing the signal as it progresses rostrally. Abnormalities of the evoked potential are most commonly associated with a disorder of touch, vibration, and conscious proprioception.

One immediate question is why we would stimulate a sensory tract to monitor for a potential motor complication? The answer is that we make a great assumption in SEP monitoring, that most insults to the spinal cord significant enough to cause a motor deficit also affect the sensory tract and therefore would be identified by this monitoring technique. There are extensive laboratory data recorded to substantiate this assumption.

Evoked potentials cannot be seen with routine electroencephalography recordings because of their low amplitudes (0.1 to 20 mV) and their admixture with normal background activity of brain a-waves. Separation of these waves can only be accomplished by signal averaging. Stimulation is given repetitively, and measurements are averaged by computer as they arrive at the recording site. In addition, common mode signal rejection, filtering, amplification amplifiers, and multichannel recording systems are required. Electrode placement uses the international electroencephalography 10–20 system.

Equipment standards are detailed by the American Electroencephalographic Society and the American Association of Electromyography and Electro-diagnosis. Voltage is recorded on the vertical axis and is reported as an amplitude or the size of the wave state in microvolts from baseline to peak. Time is recorded on the horizontal axis and is reported as a latency or the time it took from stimulus to recording measured in milliseconds. In the upper extremity this is usually 19 to 20 ms while in the lower it is usually 37 to 40 ms. The morphology of a wave refers to its general appearance.

There are several common areas of confusion regarding these waveforms. The first involves the nomenclature for identifying waveforms. Unfortunately, there is no international standard. The two used most commonly involve numbering the waves sequentially P1, P2, P3, (positive 1, positive 2, etc.) or by the expected normal latency N19, P40 (negative 19 ms, positive 40 ms, etc.). A second confusing aspect involves polarity or whether a wave is positive up or positive down. Again there is no consensus on how to arrange polarity. Generally, electromyographers have adopted a convention of having negativity in the electrode produce an upward deflection. Consistency during surgery, however, is the key factor.

Methodology

There is no universal protocol for intraoperative evoked potential monitoring. We have found stimulation of the posterior tibial nerve behind the medial malleolus to be the most convenient. Preoperatively we place our electrodes bilaterally and then carry out unilateral stimulation periodically (every 15 to 20 min). During the critical stage, as determined by the orthopedic surgeon (usually during wire placement and/or distraction) we monitor continuously. Bilateral stimulation has the advantage of increasing the amplitude of the evoked potential generated, but the disadvantage of missing unilateral changes and/or technical problems. We also stimulate the ulnar nerve t the elbow as a control. A surgical complication would prolong or eliminate the P40 waveform from the leg but leave the N19 waveform from the arm unaffected as this neurophysiologic tract is above the level of the surgery. Many laboratories record the tibial stimulus at the popliteal space peripherally as a control. The short latency period makes this less attractive to our laboratory.

Recording electrode placement upon the cortex is based upon the international electroencephalography 10–20 system. We have found electrode placement over Cz' -Fz' and C4 -C5 provides the best reproducible results. Nevertheless, recording electrodes may be placed anywhere along the neurophysiologic tract. Lueders describes a technique of recording from high cervical intraspinous ligaments to be less affected by anesthesia. Tamaki stimulates in the posterior epidural compartment while recording in the conus medularis. Others utilize a variety of 10 to 15 recording sites simultaneously. We, however, have not found it necessary to use additional recording sites unless we have difficulty with our standard protocol.

More importantly, the surgeon must trust the technology and be prepared to respond to a warning from the monitoring team. One case highlighting this point involved loss of a potential almost immediately after initiating surgery and long before the critical period. When we reported this, the surgeon withdrew all instruments and retractors. He discovered an anomalous blood vessel feeding the spinal cord, inadvertently occluded by a retractor. Had this remained undetected the patient would have suffered a postoperative complication even with a wake-up test.

Changes affecting evoked potentials

Many factors can affect evoked potentials: examples are age, sex, and body size. These, however, can be accounted for by obtaining preoperative studies. Technical problems such as a loose wire or electrical interference from other equipment is usually self-evident. Temperature is another variable, which is significant since it is not unusual for body temperature to be lowered during surgery. Hypothermia causes gradual prolongation of the latency (time) affecting both the lower and upper extremity recordings and should not be a cause for alarm.

Hypotension is frequently used during scoliosis surgery to reduce blood loss; it too may diminish the amplitude. The degree of hypotension must be decided by the anesthetist and the surgeon and is usually safely kept at 55 to 60 mmHg. From an evoked potential standpoint, it is important to maintain hypotension at a steady state during the critical period. In four cases we had a sudden abolition of the evoked potential at a mean blood pressure of 55 mmHg. While we would expect the patient to tolerate this blood pressure, we insisted on raising it to 60 mmHg, where the evoked potential returned unchanged. The surgery proceeded at 60 mmHg and there was no postoperative sequel.

Probably the most significant patient-related factors affecting evoked potentials are drugs and anesthetic agents. Some substantially impair the ability to monitor cortical SEPs. It is therefore important to understand the effects of these agents.

Barbiturates in low doses do not cause significant changes. Moderate to high doses and bolus doses may have a significant effect. In a low dose diazepam is not a problem, but high doses or boluses can cause substantial alterations. Narcotics may cause unpredictable changes.

In general, however, low doses are tolerated while high doses or boluses are variable. Neuromuscular blocking agents do not affect evoked potentials, but, on the contrary, enhance recordings by decreasing movement artifact. Nitrous oxide in concentrations of less than 60 per cent is compatible with recordings of evoked potentials. Even at that concentration, however, we have identified a gradual shift in latency that occurs at 3 to 3.5 h. This is most probably due to tissue saturation. At higher concentrations (> 60 per cent) nitrous oxide may suppress evoked potentials. Halogenated agents in general are incompatible with evoked potential monitoring. This list includes halothane and ethrane. In one of the most comprehensive reviews to date, Bernard (France) details the acceptable concentrations of the various agents used during surgery. Forane, although a halogenated agent, is not absolutely contraindicated if used in low concentrations (0.25 to 0.5 per cent).

These restrictions raise a major difficulty with monitoring in the operating room. Obviously, proper anesthesia is the most important issue. There are, however, many ways to obtain anesthesia. In general we prefer high narcotic, low gas anesthesia (forane < 0.5 per cent, nitrous oxide < 60 per cent).

Clinical settings

There are numerous situations in which monitoring of evoked potentials might be beneficial. These include peripheral nerve releases, brachial plexopathies, thoracic outlet syndrome, radiculopathies, aortic aneurysm repairs, arteriovenous malformations, pedicle screw placement, acetabular surgeries, carotid endarterectomies, cardiac bypass surgery, and spinal cord tumor excision. Evoked potential monitoring, however, is most widely used in scoliosis surgery.

Experimental animal models have been developed to test for evoked potential abnormalities in scoliosis procedures. Machida (Japan) report on SEPs in chickens with experimentally induced scoliosis while Salzman (Maryland) describes the validity of SEPs in rats undergoing Harrington rod distraction.

Nash and Brown (Cleveland) and Gonzales (New York City) have been pioneers in the development of spinal cord monitoring in humans. They report a series of 137 patients who underwent posterior spinal fusion of scoliosis with monitoring. One change in evoked potentials was identified among the 68 patients who received a Harrington rod without segmental spinal surgery. Twelve changes were noted in the 69 patients who received segmental spinal surgery with or without a Harrington rod. Loss of the evoked potential resulted in intervention and a correction of the evoked potential in each case. Spielholz *et al.* reported 55 consecutive cases of Harrington rod surgery and monitoring with no neurologic complications. Mostegl and Bauer report two significant changes out of 61 cases during Luque wiring and loss of function was confirmed by an immediate wake-up test. The surgeons responded, the evoked potential returned, and there were no neurologic sequelae. More recently, Williamson (England) reported 15 significant SEP changes in 60 patients undergoing sublaminar wire procedures. In one of the most comprehensive studies to date, Forbes (England) reports on SEP monitoring of 1168 consecutive cases. They report 119 patients had decreases in SEP amplitude of greater than 50 per cent; 35 of these were restored by repositioning of the recording wires; 52 patients had no postoperative neurologic sequela while 32 had detectable neurologic deficits postoperatively without full resolution of the SEP. They conclude that the use of sublaminal wires, the magnitude of SEP decrement, and limited or absent intraoperative recovery of SEP amplitude were factors which increased the risk of postoperative neurologic deficit. Helmers (Boston) specifically looks at the unique difficulties of pediatric spinal surgeries and the benefits of intraoperative monitoring. Brooke (Connecticut) reports a case where there was abolition of the SEP during Cotrel–Dubousset derotation with confirmation of the neurologic insult with a wake-up test. The derotation was reversed, the SEP returned, and there was no significant postoperative deficit. Bradshaw confirmed that there was complete correlation between the wake-up test and SEP abnormalities in 40 patients.

We have now monitored over 200 cases. False-positive changes due to technical problems, such as loose wires and anesthetic agents, have virtually been eliminated with experience. Nuwer (California) confirms that experienced monitoring teams had fewer false-positive events (0.63 per cent) and 50 per cent fewer neurologic deficits per 100 cases than non-experienced teams. There were 12 cases which we believed to be of neurologic significance. There were four changes due to hypotension, already discussed, and two changes during distraction where the evoked potential returned to baseline after reversal of distraction (Fig. 1). Three times we have had responses decrease by more than 60 per cent in amplitude following clamping of blood vessels during anterior approaches. The evoked response returned to baseline after unclamping and the surgeon chose not to ligate these particular vessels.

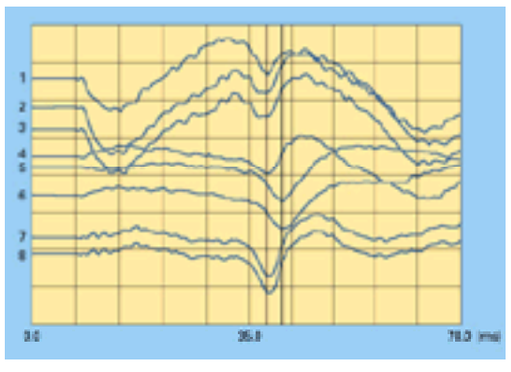


Fig. 1. Waves 1 to 4 predistraction. Waves 5 and 6 shift 2.5 ms during distraction. Waves 7 and 8 return to baseline after release of distraction.

In a similar report from Apel (California) he details the loss of SEP recordings and postoperative paraplegia in three patients following ligation of arterial segmental vessels at the apex of the kyphotic deformity. They then began to monitor for this and identified seven additional SEP changes out of 44 patients reversible by release of vascular clips. Needless to say, they advocate temporary segmental arterial occlusion with SEP recordings to avert ischemic neurologic injury.

Five times we identified amplitude declines and latency prolongation during placement of sublaminal wires. In one of these cases the patient did have a transient paraparesis postoperatively. The only other postoperative complication we incurred was a peroneal nerve injury at the fibular head following a long anterior procedure. This, however, was not a true false negative since we were only monitoring the tibial nerve peripherally. We have modified our protocol and now monitor the peroneal nerve routinely for anterior surgeries. Interestingly, we have had several patients with diminished but present deep sensation where we were unable to obtain preoperative studies. These have included a patient with achondroplastic dwarfism, eight children with significant myelomeningocele and syringomyelia deficits, and three children with muscular dystrophy. One author, Sales de Gauzy (Toulouse), has reported abnormal SEPs in patients with syringomyelia and normal clinical examination. While there is little else in the literature to collaborate these problems preoperatively, several laboratories do confirm periodic difficulties in recording from individuals with neurophysiologic disease.

What constitutes a true change in evoked potential? Certainly a sudden abolition of the evoked potential in the lower extremity with preservation of the upper extremity evoked potential is a significant event. Most neurophysiologists adhere to the 50/10 rule. Simply stated, a rapid 50 per cent loss of amplitude or a 10 per cent prolongation in latency is considered significant. Isolated changes in only one of these three parameters should provoke concern but not panic. We notify the surgeon only when I am convinced that there is a significant waveform change. Our protocol for identifying changes is detailed in Table 1.

A. Check lower extremity every 10–15 min
B. If there is a change:
1. Check impedance of recording electrodes
2. Check integrity of stimulation electrodes
3. Check ground electrode
4. Check with anesthetist for recent alteration of anesthetic agent, blood pressure, or temperature
5. Simultaneously repeat upper extremity recording
6. Repeat other lower extremity recording
7. If change in lower extremity recording persists notify surgeon

Table 1 Protocol

False-positive results or unnecessary alarm is a nuisance, but is substantially more common for inexperienced teams who do not know how to minimize technical problems and normal clinical variability. False-negative results are those in which the evoked potential remains stable, but there is a postoperative neurologic deficit.

While false-negative results are possible, they are extremely rare and usually the result of poor technique. I have reviewed cases where technicians were following background noise or artifacts, (paradoxical 60 Hz waves) not actual, evoked potentials. Nevertheless, monitoring teams have begun to look at motor track stimulation as an even better method to measure for motor tract injury. Magnetic stimulation of the cortex and/or spinal cord as well as direct electrical stimulation of the spinal cord with intraspinus and epidural electrodes have been reported. In one of the largest studies to date, Owen (Missouri) reports on 300 cases (177 children and 123 adults) with administration of SEP and motor evoked monitoring. They concluded that motor evoked monitoring was a more valid indicator of postoperative motor status.

The anesthesia requirements, however, are reversed. With somatosensory evoked potentials you want little gas and significant neuromuscular blockade. With motor evoked monitoring, several authors including Bernard (France) have demonstrated that you can use gases such as isofurane and halothane without problem but cannot use neuromuscular blocking agents. If you want to run both testing protocols concurrently, like Owen, then careful titration and bolus anesthesia techniques are required. Burke (Australia) has recently published a protocol for simultaneous recordings of somatosensory and motor evoked potential monitoring.

Conclusion

Somatosensory evoked potential monitoring during surgical procedures, especially scoliosis, has been demonstrated to be a relatively simple, sensitive, and accurate technique for the prevention of neurologic injury. It allows for continuous monitoring throughout the procedure, unlike the wake-up test and when combined with the new motor evoked potential monitoring techniques is virtually fool proof. While this procedure is expensive, involves highly developed computer equipment, and is costly in terms of physician and technician man hours, compared with the lifelong cost of care for a 13-year-old with paraplegia, it is minimal.

Further reading

Apel DM, Marrero G, King J, Tolo VT. Avoiding paraplegia during anterior spinal surgery. The role of somatosensory evoked potential monitoring. *Spine* 1991: **16**(Suppl. 8): 365–70.

Bradshaw K, Webb JK, Fraser AM. Clinical evaluation of spinal cord monitoring in scoliosis surgery. *Spine* 1984: **9**: 636–43.

Brodkey JS, Richards DE, Blasingame JP, Nulsen FE. Reversible spinal cord trauma in cats: additive effects of direct pressure and ischemia. *Journal of Neurosurgery* 1972; **37**: 591–3.

Brown RH, Nash CL Jr, Berrilla JA, Amaddio MD. Cortical evoked potential monitoring: a system for intraoperative monitoring of spinal cord function. *Spine* 1984; **9**: 256–61.

Burke D, Hicks R, Stephen J, Woodforth I. Assessment of corticalspinal and somatosensory conduction simultaneously during scoliosis surgery. *Electroencephalography and Clinical Neurophysiology* 1992; **85**: 388–96.

Chiappa KH. *Evoked potentials in clinical medicine*. Raven Press, London, 1983.

Ginsburg HH, Shetter AG, Raudzens PA. Postoperative paraplegia with preserved intraoperative somatosensory evoked potentials. *Journal of Neurosurgery* 1985; **63**: 296–300.

Gonzalez EG, Hajdu M, Keim H, Brand L. Quantification of intra-operative somatosensory evoked potentials. *Archives of Physical Medicine and Rehabilitation* 1984; **65**: 721–5.

Newer MR. *Evoked potential monitoring in the operating room*. Raven Press, London, 1986.

Newer MR, Dawson E, Carlson LG, Kanim LE, Sherman JE. Somatosensory evoked potential spinal cord monitoring reduces neurological deficits after scoliosis surgery: results of a large multicenter survey. *Electroencephalography and Clinical Neurophysiology* 1995; **96**: 6–11.

Snowden ML, Haselkorn JK, Kraft GH, Bronstein AD. Dermatomal somatosensory evoked potentials in the diagnosis of lumbosacral spinal stenosis. *Muscle and Nerve* 1992; **15**: 1036–44.

Tamaki T. Current status of spinal cord monitoring. *Spine* 1979; **4**: 467.

Williamson JB, Galasko CS. Spinal cord monitoring during operative correction of neuromuscular scoliosis. *Journal of Bone and Joint Surgery* 1992; **74**: 870–2.

46.3.1 Paediatric fractures

Patricia M. Solga

Introduction
The physis
Child abuse
Pathologic fractures
Fractures of the upper extremities
Clavicle
Scapula
Shoulder dislocations
Proximal humerus
Humeral shaft
Obstetric injuries
Elbow fractures
Supracondylar
Physeal
Dislocation
Fractures of the shaft of the radius and ulna
Hip and pelvic fractures
Hip
Femoral fractures
Midshaft
Distal physis
Pathologic
Fractures of the proximal tibia
Fractures/dislocations of the patella
Fractures of the tibial and fibular shafts
Fractures of the distal tibia–fibula and ankle
Fractures of the foot
Fractures of the toes
Conclusion
Further reading

Introduction

Injury to a child's growing skeleton presents significant challenges to the orthopedic surgeon. The child's skeleton reacts to trauma in seemingly contradictory ways: the bone can fail with trivial force or withstand significant force without fracturing. Bone density, porosity, and morphology change as the child grows. The periosteum is well developed, which endows the bone with stabilizing properties: less comminution occurs when the bone fails, and reduction is easier to obtain and maintain. The periosteum carries with it a rich blood supply, which promotes rapid healing. Healing of fractures is also enhanced by the child's anabolic metabolism. The child's soft tissues are more vulnerable than the adult's but heal more rapidly. However, the most unique characteristic of immature bone is the existence of the physis or ‘growth plate'. This specialized layer of cartilage cells provides a continual supply of scaffolding upon which mature bone forms. The physis also influences fractures patterns as well as bone healing and remodeling. The child's skeleton undergoes significant changes in both size and morphology throughout growth. A thorough understanding of the principles that govern skeletal maturation and response to injury is necessary to render appropriate orthopedic care.

The physis

The physis or growth plate is composed of specialized chondrocytes or cartilage-producing cells that continuously supply a scaffold upon which bone is laid down. Both longitudinal and circumferential bone growth occurs at this site. A physis is located at each end of the long bones, except in the hand and foot, where the physis is present at only one end. In the non-tubular carpal and tarsal bones the physes are circular.

Physeal chondrocytes are arranged in three discrete zones, the reserve, the proliferative, and the hypertrophic. These zones allow the orderly production of bone throughout childhood. The reserve zone provides a renewable source of chondrocytes. The proliferative zone is responsible for cell division. The hypertrophic zone is responsible for cell growth and deterioration, followed by the gradual release and deposition of calcium. Vascular invasion then occurs, bringing with it osteoblasts, the bone-forming cells, which lay down bone upon this newly created scaffolding. These first osteoblasts are subsequently replaced with a new population of cells, osteoclasts, which then resorb the primitive bone matrix. Lastly, a new population of osteoblasts migrates in and produces the finished bone. At skeletal maturity the physis regresses. Without a healthy population of chondrocytes in a healthy milieu, orderly growth cannot occur.

Fractures occur invarious areas of the child's bone: the diaphysis, the physis, the metaphysis, and the epiphysis. Potentially, fractures can occur in any combination of these anatomic areas. Fracture patterns can occur at the interfaces between cartilage, bone, and ligament that are unique to the immature skeleton. It is often the physis that fails first: the resulting fracture patterns have been organized into simple classifications that provide a basis for understanding and treatment. The most commonly used classification is the Salter–Harris (Fig. 1). A Salter–Harris I injury is defined as a fracture along the transverse plane of the physis; it may not be seen on standard radiographs unless stress views are obtained. The Salter–Harris II fracture occurs through the transverse plane of the physis and exits through the metaphysis. The Salter–Harris III fracture is one through the physis and the epiphysis or articular portion of the bone, and the articular surface of the bone is disrupted; this pattern has the potential to produce incongruity and subsequent deterioration of the joint. A Salter–Harris IV fracture occurs through the epiphysis, the physis, and the metaphysis; it therefore also carries the same potential for joint problems. Lastly, a Salter–Harris V fracture is a compression injury to the physis causing destruction of the chondrocytes and disrupting the blood supply to the adjacent bone. The type V fracture pattern is often not recognizable on initial radiographs but is seen after a period of time when the physis closes prematurely and growth ceases. In addition, the healing of fractures near, but not involving, the physis is greatly influenced by the physis itself; these fractures have a great potential to remodel, especially when significant growth still remains.

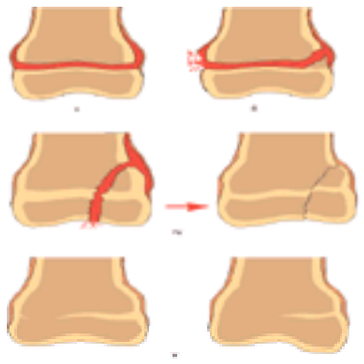


Fig. 1. Salter–Harris classification of growth-plate injuries: (a) type I fracture; (b) type II fracture;(c) type III fracture; (d) type IV fracture; (e) type V fracture.

Child abuse

The orthopedist may be the only physician in a position to identify the signs of child abuse, because some fractures can be subtle and hard to detect. A fracture can be an isolated injury. Fractures may not cause significant pain or distress, or may already be partially healed and therefore susceptible to being overlooked. While certain typical fracture patterns in children are often associated with physical abuse, the possibility that any fracture is the result of abuse or neglect should always be considered, especially in children too young to give a reliable history. Classic teaching holds that fractures in the lower extremities in children who do not yet walk, compression fractures near the metaphysis of the knee and ankle in young children, and spiral fractures of the humerus in small children should be viewed with suspicion. Multiple fractures in various stages of healing or delay in seeking medical attention should similarly be considered suspicious. Communities have established protocols for investigating child abuse. The physician always considers whether the injury is suspicious for abuse or neglect and is required to report these findings to

the designated agency. The potential for lethal injury to an abused child is real, and some studies have reported that up to 10 per cent of abused children eventually suffer a lethal injury.

Pathologic fractures

A pathologic fracture is one that occurs through abnormal bone. It may happen in the generalized or focal skeletal dysplasias, neuromuscular diseases, bone tumors, and osteomyelitis. Some examples are fractures through simple bone cysts, through the osteoporotic bone of nonambulatory children, physeal fractures through the proximal femur in children with rickets, and fractures through the dysplastic bone of children with osteogenesis imperfecta (Fig. 2 and Fig. 3). The treatment goal for pathologic fractures is to restore alignment and length without compromising potential growth. The treatment plan is also guided by the nature of the pathologic entity that caused the fracture.

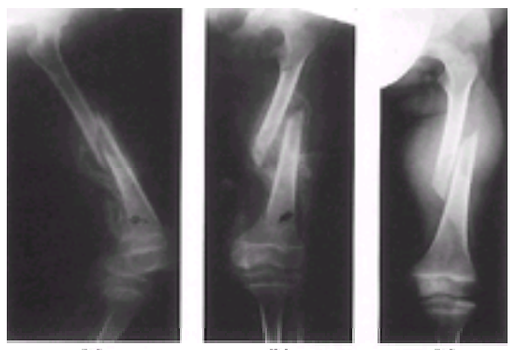


Fig. 2. (a) Pathologic midshaft fracture of the femur in a nonambulatory child with myelomeningocele; the fracture occurred when she crossed her legs while in bed. (b) Supracondylar fracture of the femur following cast removal. (c) Abundant callus formation after cast immobilization, with a new supracondylar fracture.

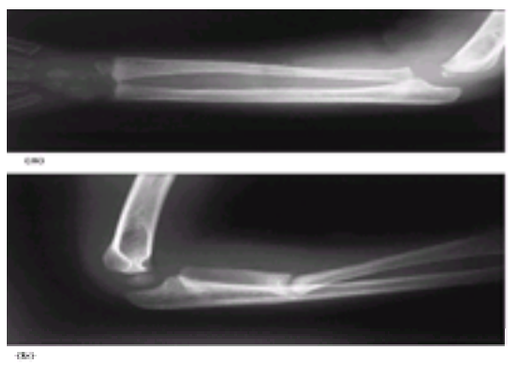


Fig. 3. (a) Neurofibromatosis affecting the radius in a 3-year-old;(b) pathologic fracture producing a pseudarthrosis.

Fractures of the upper extremities

Clavicle

Fracture of the clavicle is common throughout childhood. It is most often caused by a low-energy fall on to an outstretched hand. Depending upon the degree of displacement, pain can be insignificant and the child can present with guarding of the extremity but little pain. The middle third of the clavicle is the most common site. The clavicle's periosteal sleeve, the soft tissue that encases the bone, often prevents significant displacement. Depending upon the alignment seen on radiographs, a figure-of-eight splint will provide adequate reduction and pain control. A simple sling may be all that is necessary to support the weight of the arm in a non-displaced fracture (Fig. 4). The minority of clavicular fractures are widely displaced and sometimes segmental or comminuted (Fig. 5); these occur principally in teenagers engaged in contact sports or involved in motor vehicle accidents. Closed treatment in a sling and swath is usually sufficient to manage those fractures in which the skin is not compromised. Even in these widely displaced fractures, prompt and complete healing is anticipated without any functional impairment. Unlike the healing and gradual remodeling of fractures in other long bones, the clavicle often heals with an abundant callus, which results in a large bump that persists long after healing is completed. When the arm can be actively abducted fully and painlessly, immobilization can be discontinued.



Fig. 4. Minimally displaced clavicle fracture.

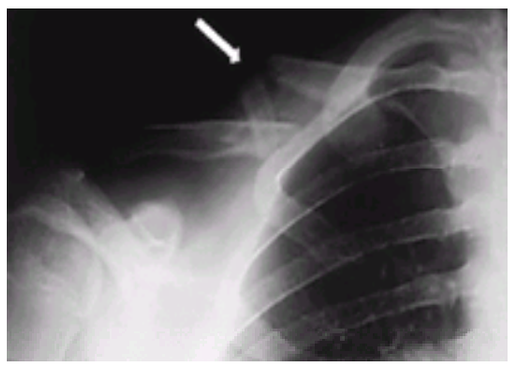


Fig. 5. Segmental midshaft fracture of the clavicle with 90° of displacement managed successfully with closed treatment.

The sternal or acromial ends of the clavicle are less commonly fractured; these fractures can be mistaken for chromioclavicular, acromioclavicular, and sternoclavicular

disruptions. The fracture can be difficult to visualize and is sometimes diagnosed only after periosteal new bone has formed. With involvement of the sternoclavicular joint, it is mandatory to rule out posterior dislocation, best done by computed tomography (**CT**), because this injury can cause respiratory embarrassment.

Scapula

The multiple secondary centers of ossification of the acromion and coracoid process of the scapula are sometimes mistaken for fractures. Radiographic views of the uninvolved side for comparison are often necessary. High-energy trauma to the chest can produce fractures of the body of the scapula, with associated injuries to ribs, the heart and lungs, and the brachial plexus. Glenoid fractures associated with fractures of the proximal humerus and shoulder dislocations are rare but do occur in teenagers. Treatment is dictated by the amount of displacement.

Shoulder dislocations

Anterior dislocation of the shoulder is usually caused by violent trauma such as a direct blow. However, children with generalized ligamentous laxity can have inherent instability of the shoulder and some children are capable of voluntarily subluxing or dislocating their shoulder or experience instability with activities of daily living. In these cases, physical therapy is essential and the provocative maneuvers must be stopped. The traumatic dislocation of the shoulder that usually occurs in teenagers is treated exactly as in an adult: closed reduction, sling-and-swath immobilization for several weeks, and physical therapy. Redislocation is common and surgical stabilization may be necessary.

Proximal humerus

The proximal humerus is vulnerable to Salter–Harris II as well as metaphyseal fractures. Many of these injuries are the result of low-energy trauma and do not require operative intervention. Immobilization with a sling and swath is adequate. Because 80 per cent of the growth of the humerus occurs at the proximal physis, significant remodeling of the fracture is the norm. Residual deformity is rarely a cosmetic problem, owing to the muscular bulk at the proximal humerus. In high-energy trauma, deforming forces at the shoulder girdle render displaced fractures unstable and necessitate closed reduction and temporary fixation with smooth or partially threaded K-wires (Fig. 6). Interposition of the biceps tendon in a widely displaced, irreducible fracture is an indication for open reduction.



Fig. 6. (a) Widely displaced Salter–Harris II fracture of the proximal humerus; (b) closed reduction and provisional fixation with threaded Steinmann pins.

Pathologic fractures occur commonly through previously unrecognized and asymptomatic unicameral bone cysts of the proximal humerus. These fractures, which are usually minimally displaced, lend themselves to simple immobilization for a few weeks; prompt healing is anticipated. Attention is then directed towards treatment of the cyst to prevent re-fracture. Ablation of the cyst can be achieved either with several courses of interosseous injection of methylprednisone, or, if this fails, biopsy, curettage, and bone grafting.

Humeral shaft

Spiral fractures of the humerus can be the result of physical abuse if the arm of a young child is twisted (Fig. 7). High-energy torque can produce both spiral and transverse displaced fractures (Fig. 8). Injury to the radial nerve can accompany this fracture, just as in the adult. As in fractures of the proximal humerus, immobilization is limited by the axilla. A coaptation splint is applied with the axilla well padded. Shortening does not cause the same problem as in fractures in the lower extremity and the need for operative treatment is rare. Older children can be managed in a fracture brace that enhances alignment. Angulation of 10 to 15° in the coronal or sagittal plane is acceptable, but in fractures of the more distal shaft less coronal angulation should be accepted. A fracture of the humeral shaft that heals in varus results in an unacceptable cosmetic result.

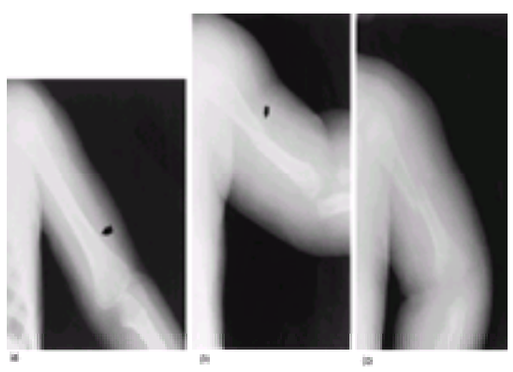


Fig. 7. (a) Non-displaced spiral fracture of the humerus in a 16-month-old who was the victim of child abuse. (b) and (c) Transverse midshaft humoral fracture in a 7-month-old who was the victim of child abuse.

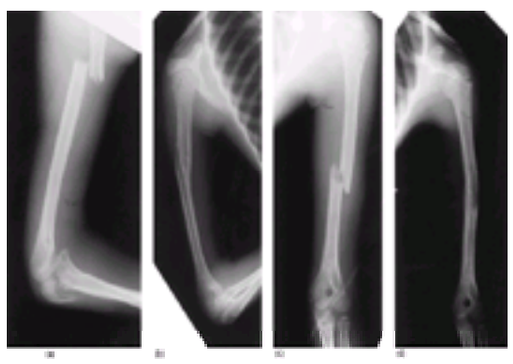


Fig. 8. (a) Ten-year-old boy with displaced olecranon fracture of the right ulna and an ipsilateral proximal transverse fracture of the humerus. The ulnar fracture required open reduction and temporary internal fixation; the humoral fracture required a pin to control the distal fragment. (b) Healed fractures after removal of the hardware. (c) Contralateral midshaft humoral fracture that required closed reduction; (d) left humerus after healing.

Transverse fractures of the shaft from high-energy trauma (Fig. 9) present the same difficulty of immobilization as spiral fractures. They also share a similar tolerance of mild residual deformity associated with proximal metaphyseal fractures.

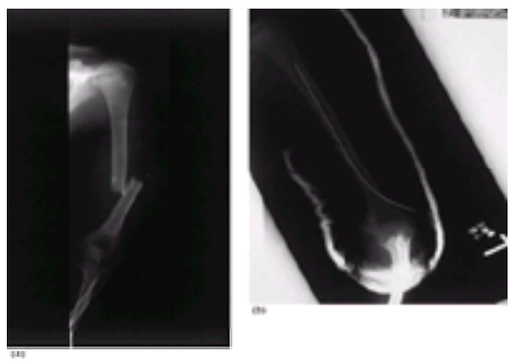


Fig. 9. (a) Grade I open fracture of the humerus; (b) treated with irrigation and debridement, and provisional fixation with a flexible nail.

Obstetric injuries

Birth fractures in the upper extremities, with or without associated injuries to brachial plexus, are associated with difficult vaginal deliveries of large babies and often involve shoulder dystocia. When the newborn infant is not moving an extremity, careful examination and radiographs are indicated. The simplest possibility is a midshaft fracture of the clavicle with a normal neurologic examination. Simple immobilization can be obtained by pinning the sleeve of the undershirt to its bodice. Careful handling of the infant with this type of immobilization provides very adequate pain management. Healing is always extremely rapid and uneventful.

Fracture of the humeral shaft is less common: non- or minimally displaced fractures are treated the same as clavicle fractures, but widely displaced fractures are more difficult to immobilize. The level of the fracture is usually at the axilla and control of the proximal fragment is difficult. A well-padded coaptation splint over the shoulder is used, but extreme caution is needed to prevent skin breakdown. Significant displacement can be tolerated because of the enormous potential for fracture remodeling in infants.

Rarely the entire proximal physis of the humerus can be separated. Because the epiphysis is not ossified at birth the radiographs can be difficult to interpret and an arthrogram may be necessary to make the diagnosis.

Damage to the brachial plexus is the most serious type of obstetric injury to the upper extremity (Fig. 10). The initial examination must differentiate between the pseudoparalysis associated with a fractured clavicle or humerus and an injury to the brachial plexus. It is difficult and unnecessary to localize the exact anatomic lesion. If there is no concurrent osseous injury the parents are taught gentle, passive range-of-motion movements to all joints of the upper extremity. Nerve-conduction studies are condoned in infants. Repeat examinations at 4- to 8-week intervals will adequately define the neurologic lesion and document any recovery. The eponyms of Erb, Klumke, and Erb–Duchenne–Klumke refer to upper, lower, or whole plexus injuries, respectively. Resolution of these injuries is impossible to predict, but some improvement is expected within the first year. After that time there is little potential for spontaneous recovery. Predictable success is not yet obtainable with surgical repair of injured nerve tissue. A program of maintenance physical therapy done by the family is important and will improve the functional outcome if muscle transfers and osteotomies are subsequently needed.

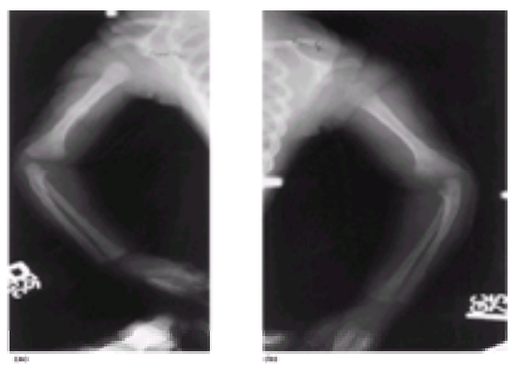


Fig. 10. (a) One-month-old boy with healed, non-displaced fracture of the right humerus following a spontaneous vaginal delivery; he weighted 4.5 kg at birth; (b) his left side has no fracture but he sustained a complete brachial-plexus injury.

Elbow fractures

Supracondylar

Nowhere else in the child's skeleton can a relatively low-energy injury put the limb at such risk and mandate the need for experienced surgical management than in the supracondylar area of the humerus. The cross-section of this metaphyseal portion of the bone is quite small. While this is not a physeal injury, malreduction causes an unacceptable cosmetic result that does not remodel with growth. Supracondylar fractures generally occur between the ages of 2 to 8 years; they are the result of a fall on to an outstretch hand and hyperextended elbow.

The limb is first inspected for signs of an open fracture or tenting of the skin and impending tissue necrosis by the fracture fragments, followed by a meticulous neurovascular examination. Documentation of the examination is imperative. Even minimally displaced fractures can be associated with significant soft-tissue swelling. The grossly deformed elbow should be splinted without any manipulation before radiographs are obtained. Since reduction of the fracture requires restoration of the carrying angle, a comparison anterior–posterior view is mandatory. Displacement is then assessed radiographically in both the coronal and sagittal planes, and Baumann's angle, formed by the physis of the capitellum and its perpendicular to the humeral shaft, is measured. Restoration of this angle prevents a cubitus varus deformity. The lateral view is assessed for flexion–extension deformity. A line drawn along the anterior humeral line should bisect the capitellum. Another line drawn through the midpoint and parallel to the humeral shaft should intersect a line drawn through the midpoint of the capitellum at about 30°.

Treatment is dictated by the neurovascular status of the arm, the integrity of the skin, and the displacement of the fracture as seen radiographically. These fractures have been broadly classified by Gartland based on the amount of displacement. The Gartland type I is nondisplaced. The fracture line may be visible only on an oblique view, or assumed on the basis of a fat-pad sign, or demonstrated by the appearance of periosteal new bone seen on radiographs a week or two later. If careful scrutiny does not demonstrate an impaction injury with some varus displacement, the Gartland I fracture can be safely splinted in 90° of elbow flexion. A posterior splint is used because there is always attendant swelling, even in a nondisplaced fracture. The splint can be changed for a snugger-fitting cast a week later once the initial swelling has resolved. Healing is prompt.

The Gartland type II fracture has a intact posterior cortex, the distal fragment is extended, and some degree of coronal deformity can be present. If 5 to 10° of posterior angulation is not accompanied by a difference in Baumann's angle of 5° or more, and if the neurovascular examination is normal, then the arm can be splinted *in situ.* However, to stabilize the fracture requires immobilization of the elbow in at least 90° of flexion, which can lead to circulatory embarrassment if there is much swelling. In addition, further displacement can occur. Therefore, closed reduction and percutaneous transfixation gives a more predictable result. If a reduction maneuver is needed, it is done under general anesthesia and temporary, smooth K-wire fixation is used to maintain it; two lateral pins or crossed medial and lateral K-wires are used.

The Gartland type III fracture is commonly associated with neurologic injury of almost equal incidence in the ulnar, median, and radial nerves and their branches (Fig. 11). Absence of nerve function should be documented, with spontaneous recovery expected within 3 to 6 months. The absence of the radial pulse by Doppler indicates the need for emergency closed reduction under anesthesia. If no pulse returns but the hand is warm, with good capillary refill, close observation is appropriate. If there are any signs of vascular embarrassment, exploration of the artery is indicated. Any new neurologic deficit found after closed reduction mandates the need for assessment of the position of the transfixion wires and possible exploration of the fracture site.

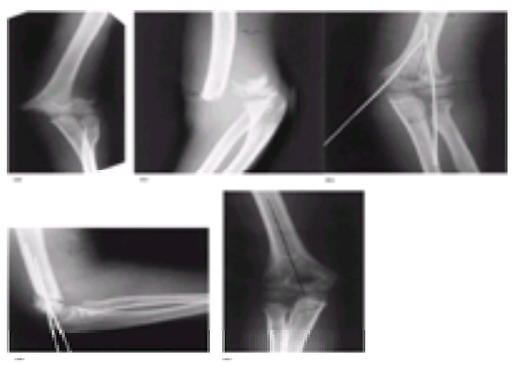


Fig. 11. (a) and (b) Seven-year-old boy with completely displaced supracondylar fracture of the humerus. He presented with no radial or ulnar pulses and complete sensory and motor deficits in the median nerve distribution. The anterior position of the proximal fragment was probably responsible for the neurovascular deficits. (c) and (d) After closed reduction and percutaneous pinning his pulses returned and within 2 days he had regained his median nerve function with only slightly diminished 2-point discrimination in the thumb. Note the restoration of Baumann's angle as compared with(e) the normal side.

Most type III fractures can be treated by closed reduction and percutaneous fixation with smooth K-wires. However, an irreducible fracture will occasionally be encountered, necessitating open reduction by the anteromedial approach of Henry and percutaneous K-wire fixation. All closed and open reductions are maintained in a posterior splint with the elbow at no more than 90°.

Postoperatively, the extremity is then assessed for signs and SYmptoms of compartment syndrome, which, if untreated, leads to the devastating consequence of Volkmann's ischemic contracture. Considerable effort is needed to obtain a reliable examination of a young child's hand.

Physeal

Knowledge of the appearance and location of the secondary centers of ossification and the use of contralateral views are needed to diagnose and treat physeal fractures of the elbow.

Lateral condyle

Fractures of the lateral condyle occur in two ways: a fall on to an outstretched hand with the forearm supinated or a fall on to an outstretched hand with the elbow flexed and a varus force at the elbow. These fractures are intra-articular: they traverse the capitellum and exit the lateral aspect of the distal humeral metaphysis (Fig. 12) The fracture patterns are classified as Milch I, a fracture line lateral to the trochlea, or Milch II, a fracture line into the apex of the trochlea. Treatment depends upon the degree of displacement. Less than 2 mm of displacement can be immobilized in a splint, but frequent visits are necessary during the first 2 weeks to assess further displacement. In order to visualize these small fracture lines, radiographs must be taken out of the splint. Further displacement of the fracture or initial displacement of greater than 2 mm dictate reduction and temporary internal fixation. Joint congruity must be attained. Iatrogenic injury by the stripping of the blood supply of the capitellum leads to avascular necrosis. Unlike other pediatric fractures, displaced lateral condyle fractures that are untreated can fail to unite.



Fig. 12. (a) and (b) Displaced intra-articular fracture of the lateral condyle of the humerus; (c) and (d) treated with open reduction and provisional fixation with smooth K-wires across the physis.

Lateral epicondyle

The secondary center of ossification of the lateral epicondyle appears at about the age of 11 years. Fracture of the lateral epicondyle is a rare, avulsive injury of the extensor muscles. Treatment is simple immobilization.

Medial condyle

Fractures of the medial condyle are rare. The secondary center(s) of ossification of the trochlea appears at about 9 to 10 years of age. Fractures through the unossified medial condyle are, therefore, difficult to appreciate on radiographs and can be missed. An arthrogram or magnetic resonance imaging are performed when medial swelling and tenderness of the elbow is accompanied by negative radiographs.

Medial epicondyle

This fracture is associated with elbow dislocation, fractures of the radial head, ulnar neuropraxia, and displacement of the flexor origin down and into the elbow joint. Displacement of less than 5 mm not involving entrapment of the flexor origin into the joint is treated with a brief period of immobilization followed by active range-of-motion exercises of the elbow. Even in young children there is a tendency for joint stiffness not seen in other elbow fractures. Displaced fractures require open reduction and internal fixation with temporary fixation for about 3 weeks, followed by active range-of-motion exercises. (Fig. 13)

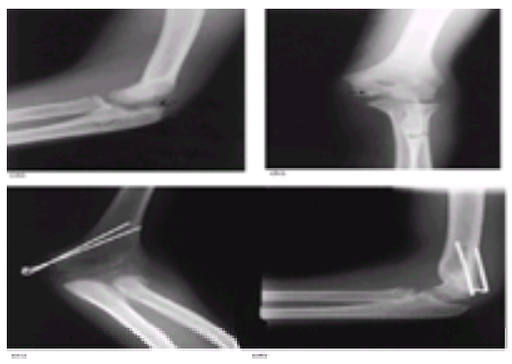


Fig. 13. (a) and (b) Fracture of the medial epicondyle and dislocation of the elbow. Following closed reduction the medial epicondyle is entrapped in the joint, necessitating open reduction and internal fixation. (c) and (d) Radial head fracture and displaced medial epicondylar fracture (in b). Treated with closed reduction of the radial head and open reduction and provisional fixation of the medial epicondylar fracture.

Proximal radius

Fractures of the proximal radius are Salter–Harris types I to IV, but the treatment principles are the same regardless of the type. They are sometimes seen in association with elbow dislocations. Treatment is dictated by the displacement of the fracture. Up to 30° of angulation is acceptable, provided that pronation and supination are not prevented by malalignment of the proximal fragment. Minimally displaced fractures are immobilized for about 5 days and active range-of-motion exercises then begun. For displaced fractures, closed reduction, manipulation with the blunt end of the percutaneous Steinmann pin, or open reduction with or without smooth K-wire fixation are used. Fixation is left for 2 to 3 weeks and active motion then begun (Fig. 14 and Fig. 15). Avascular necrosis and radioulnar SYnostosis are complications that can occur in widely displaced fractures.

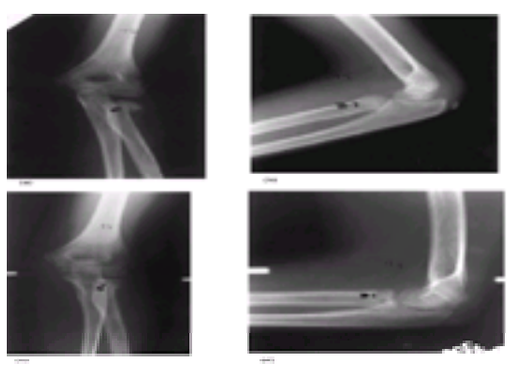


Fig. 14. (a) and (b) Displaced Salter–Harris II fracture of the proximal radial physis in a 9-year-old girl; (c) and (d) following aspiration of the hemarthrosis and closed reduction.

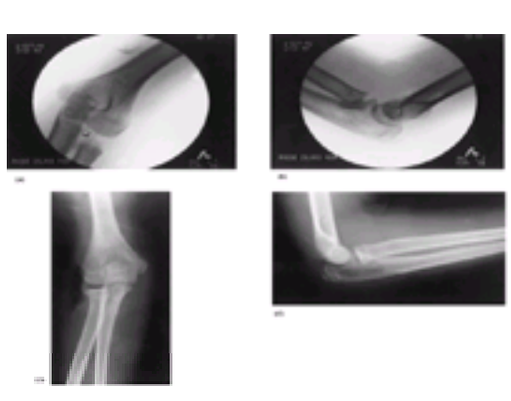


Fig. 15. (a) and (b) Displaced Salter–Harris II fracture of the proximal radial physis in a 10-year-old girl. The elbow lacked pronation and supination. Closed reduction was unsuccessful. (c) and (d) Following open reduction, pronation and supination were normal and the reduction was stable without internal fixation.

Dislocation

Dislocation of the elbow is seen most commonly in adolescents and teenagers. It is usually the result of a fall on to an outstretched hand with the elbow extended, or a fall directly on to a flexed elbow. Associated fractures of the proximal radius and medial epicondyle can occur, and should be looked for before closed reduction is attempted. An uncomplicated dislocation is treated with immobilization for 3 weeks.

Subluxation of the radial head (nursemaid's elbow) occurs when a longitudinal force is applied to the arm of a young child. The radial head can slide away from the annular ligament, owing to the shape of the head and the laxity of the ligament. The diagnosis is made by history and physical examination. Flexing the elbow while supinating can reduce the subluxation. Unfortunately, once the injury has happened it can recur with surprisingly less force and with a less satisfying reduction. The child may require immobilization in a splint for several days, even with successful reduction. Fortunately, subluxation of the radial head ceases with normal growth.

Isolated dislocation of the radial head is rare. It should be differentiated from congenital or old, untreated dislocation of the radial head by noting adaptive changes in the capitellum and radius that are found in long-standing dislocations.

If the diagnosis of an acute traumatic dislocation is made promptly, it is imperative that reduction be achieved by either open or closed means, and repair of the annular ligament may be necessary.

Fractures of the shaft of the radius and ulna

Because of the origins and insertions of arm/forearm muscles, different fracture patterns occur in the proximal, middle, and distal one-third of the forearm. Fractures in the proximal one-third often involve transverse fractures of the radius and it is difficult to obtain anatomic reduction by closed methods. Dislocation of the radial head or fractures of the proximal one-third are often associated with fractures of the ulna, the Monteggia fracture patterns.

To prevent limitation of pronation and supination, midshaft fractures of the radius and ulna require near-anatomic reduction and frequent radiographs to ensure maintenance of reduction. In the chubby, short forearm of a small child, a well-molded cast may not hold the reduction. A second reduction a week or so later may be required to optimize results. Bayonet apposition of the fracture is acceptable but undesirable. An excellent, stable reduction of low-energy greenstick fractures can generally be achieved in a child of 8 to 12 years of age if the shape of the forearm lends itself to cast treatment. Intramedullary fixation with Steinmann pins or flexible intramedullary nails provides excellent provisional fixation for displaced and/or unstable fractures. These nails can be removed without the need for general anesthesia if left outside of the skin. Limited open reduction to remove interposed muscle or periosteum can be necessary when provisional fixation is used (Fig. 16). Teenagers with significantly displaced fractures are treated with open reduction and internal fixation with compression plates.

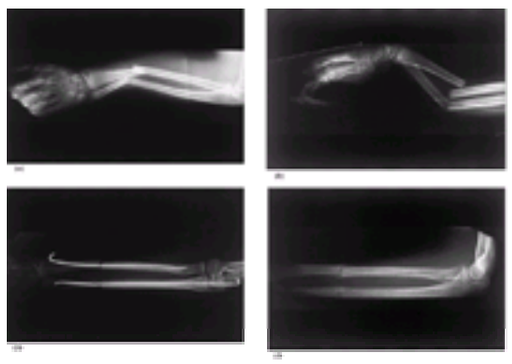


Fig. 16. (a) and (b) Displaced, unstable radius and ulna shaft fractures; (c) treated with limited open reduction and intramedullary flexible nails; (d) healing fractures following removal of nails.

Fractures of the distal one-third occur in the following frequency: minimally displaced compression fractures, transverse fractures of either the radius or both the radius or ulna, and Salter fractures (usually Salter–Harris II) (Fig. 17 and Fig. 18). Nondisplaced compression fractures are managed with a short arm cast for 2 or 3 weeks, except in very young children, in whom a long arm cast is needed to prevent the cast from slipping off. Transverse fractures are often completely displaced and require a vigorous reduction under local anesthetic and intravenous sedation. Occasionally a general anesthetic is required to obtain a satisfactory reduction. Because the proximity of the fracture to the physis allows considerable remodeling, 15° of apex volar or dorsal displacement is tolerated in a child with 4 to 6 years of growth remaining,. Bone apposition of 75 to 80 per cent is sufficient, but radial or ulnar deviation of more than 5° is not acceptable.



Fig. 17. (a) Displace distal radius and ulna fracture; (b) and (c) treated with closed reduction.



Fig. 18. (a) and (b) Completely displaced Salter–Harris II fracture of the distal radial physis in a 10-year-old girl following a horse-riding injury. There were motor and sensory deficits in the median nerve distribution, which improved after closed reduction. Note the arrow showing anterior displacement against the median nerve of the proximal fragment. (c) and (d) Incomplete but acceptable alignment following one attempt at closed reduction. (e) and (f) Complete remodeling after 8 months.

Salter–Harris II fractures of the radius often occur as isolated injuries. The distal radial physis, the ulna, and the entire carpus are often displaced dorsally. Compression of the median nerve is common, but symptoms usually resolve quickly after reduction. An irreducible fracture can occur because of entrapment of periosteum or tendon, in which case an open reduction is required. The formation of a physeal bar can occur in the distal radius, causing angular deformity with further growth. Complete cessation of growth of the distal radius causes a loss of SYmmetry in the relation of the radius to the ulna. Thus, the ulnar variance measured at skeletal maturity will differ from that of the contralateral side, but the significance of this difference has not clearly been established.

Hip and pelvic fractures

Common pelvic fractures in childhood avulsion injuries occur at the origin of tendons: the sartorius from its origin, the anterior superior iliac spine; the rectus femoris from the anterior inferior iliac spine, or the hamstring muscles from the ischial tuberosity. These are usually sports injuries occurring in adolescents or teenagers and are treated with rest, crutches, and nonsteroidal anti-inflammatory drugs. If radiographs are taken several weeks later, abundant callus is seen at the site of avulsion, which can be mistaken for a malignancy.

High-energy trauma to the pelvis is associated with fractures as well as injury to the abdominal and pelvic viscera. The physical examination must include a careful neurologic assessment of the sacral nerve roots and examination of the perineum. Imaging studies must be chosen to visualize the complex shape of the pelvis, such as oblique, inlet and outlet views, and CT scans. Wide pubic diastases require internal or external fixation, and are associated with injuries to the bladder and perineum. Isolated fractures of the pubic rami or minimal pubic diastasis can be managed with bed-to-chair activities and crutch walking, as comfort allows. Significant disruption of the pelvic ring and concomitant soft-tissue injuries can require the same operative management as in the adult patient.

Injury to the triradiate cartilage, the confluence of the three physes of the acetabulum, can cause growth arrest and acetabular dysplasia, which carries significant morbidity for the younger child with considerable growth remaining.

Hip

Fractures of the proximal femur occur in the epiphysis, physis, and metaphysis. Even with prompt treatment, the results can be poor because of resultant avascular necrosis, with varus deformity of the hip leading to poor biomechanics and growth arrest. All displaced fractures of the femoral neck require closed or open reduction and internal fixation, and evacuation of the intracapsular hematoma. Further immobilization in a spica cast to ensure restricted weight-bearing postoperatively is necessary in the young child. Pathologic fractures through the proximal femur may dictate medical management of the underlying condition, such as treatment of rickets or bone grafting for lesions such as bone cysts or fibrous dysplasia, in addition to surgical management of the fracture.

Femoral fractures

Fractures of the femoral shaft can be the result of low- or high-energy trauma. Obstetric fractures can occur and very young children can sustain spiral fractures from a

low-energy torque force. If the child is not yet of walking age the possibility of physical abuse is always raised (Fig. 19). Children also suffer high-energy trauma as the result of motor vehicle accidents while walking or riding a bicycle or while riding in a car. Lastly, typical accidents of childhood, such as sledding, skiing, or playground injuries, commonly cause femoral fractures. The treatment goal is to attain an anatomically aligned bone with no shortening. However, in children, overgrowth of the injured femur of up to 1 cm can occur between the ages of 3 and 10 years. In addition, bayonet apposition in good alignment, without excess shortening, leads to an excellent result in a child.

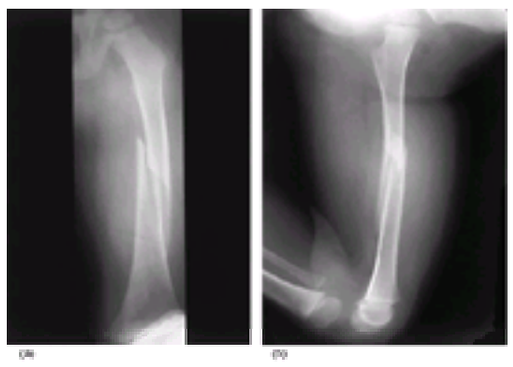


Fig. 19. (a) and (b) Low-energy torque causing a minimally displaced femoral fracture in a 2-year-old, amenable to immediate spica casting.

Fractures of the femoral shaft are conveniently classified by location: proximal, middle, and distal third. This classification, along with the age of the child, the fracture pattern, and the mechanism of injury, dictates treatment. Open fractures require irrigation and debridement, followed by stabilization with plates, external fixators, intramedullary nails, or traction. Children with associated injuries such as to the head, chest, or abdominal viscera cannot tolerate traction and must be managed with internal or external fixation.

Fractures of the proximal third are difficult to manage because the proximal fragment is flexed and rotated by the hip flexors and abductors, making it difficult to control. Skin or skeletal traction for several weeks to allow provisional healing, and spica casting, are needed. If adequate reduction cannot be achieved by traction, then open reduction is necessary.

Midshaft

Minimally displaced and shortened, low-energy spiral fractures are treated with immediate spica casting. Displaced and/or shortened fractures in young children weighing less than 13.5 kg are managed with overhead skin traction of both legs for a week or more until provisional healing occurs. This treatment permits spica casting without loss of reduction. Skin traction requires meticulous attention to the possibility of skin breakdown and circulatory embarrassment. Children older than about 7 years who have non-comminuted fractures can be managed with flexible intramedullary nails as provisional fixation. If necessary this treatment can then be supplemented with a fracture brace incorporating the hip to allow early partial weight-bearing. Treating fractures by this method results in shorter hospital stays and earlier return to school (Fig. 20). More traditionally, fractures can also be treated with skeletal traction through the distal femur, which maintains the hip and the knee, each placed at 90°. Although teenagers can be treated with rigid intramedullary nails, avascular necrosis of the femoral head and growth arrest of the greater trochanter can develop rarely. A more lateral and anterior choice of entry site in the piriformis fossa may decrease the risk of avascular necrosis. Interference with the blood supply to the femoral head by the nail is thought to be the cause of this complication. It is a complication not seen in adults and is an excellent example of the biologic difference between the large, but skeletally immature, femur of a teenager and the femur of an adult. However, there is a substantial benefit to the use of rigid immobilization in teenagers, especially when associated with fracture of the ipsilateral tibia or multiple systemic injuries.

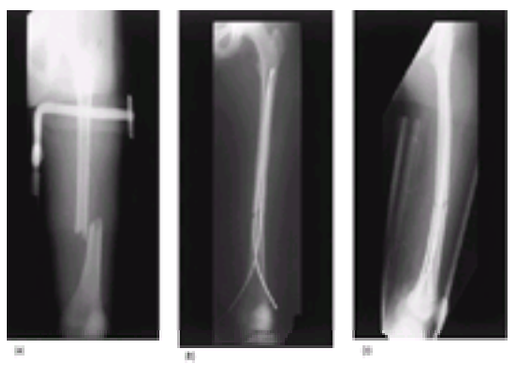


Fig. 20. (a) High-energy midshaft femoral fracture in an 11-year-old; (b) and (c) treated with closed reduction and intramedullary fixation with flexible nails.

Reduction of fractures of the distal third is complicated by the pull of the gastrocnemius on the distal fragment causing flexion at the fracture. Proximal tibial skeletal traction, external fixator, or compression plating are used, depending on the age, size and fracture patterns.

Distal physis

Injury to the distal femoral physis carries a high risk of premature growth arrest. A varus or valgus stress applied to the knee in an adolescent can cause a physeal fracture instead of a soft-tissue injury to the knee joint, such as a torn ligament or meniscus. For example, the energy imparted by a football tackle can potentially cause a fracture of the femoral shaft or a Salter–Harris fracture of the distal femur. Partial or complete growth arrest can occur, leading to shortening or angular deformity of the knee. Stress radiographs may be necessary to diagnose a Salter–Harris I fracture. A Salter–Harris V fracture in one portion of the physis and a Salter–Harris II in another portion lead to an angular deformity secondary to partial growth arrest.

Displaced fractures need reduction and provisional fixation with smooth Steinmann pins. As with other physeal fractures, deformity in the plane of motion of the joint carries a higher remodeling potential.

Pathologic

Children with underlying neuromuscular disorders are prone to fractures of the femur following trivial trauma. Depending on the ambulatory status of the child and the neuromuscular disorder, different, often nonoperative, treatments may be necessary. For example, in a nonambulatory child the bone is gracile with a thin cortex and significant osteoporosis, which are contraindications to skeletal traction or internal fixation. Fixed contractures of the knee and/or the presence of a dislocated hip can both precipitate a fracture and dictate a different treatment. In addition, nonambulatory children with seizure or other neuromuscular disorders do not tolerate well either traction or spica casting. Treatment is directed at obtaining prompt healing in acceptable alignment, without causing further risk or injury to the child while considering their functional needs. In all cases, vigilance in the care of the skin is mandatory, especially in the noncommunicative child or when the limb is insensate. Well-padded splints can often provide adequate immobilization.

Fractures of the proximal tibia

Fractures of the proximal tibia can occur through the joint, such as an avulsion injury of the anterior cruciate ligament, or at the physis, which carries the risk of injury to the neurovascular structures of the popliteal fossa; this can also occur in an adult who sustains a knee dislocation. Unique to the child, a fracture at the bone–ligament–cartilage interfaces of the tibial tubercle can occur, as well as fracture at the metaphysis. Displaced intra-articular fractures are treated with open or limited open reduction and internal fixation. For physeal or metaphyseal fractures, closed reduction and casting generally give good results. Physeal fractures are

monitored for premature growth arrest and metaphyseal fractures are monitored for subsequent valgus deformity.

Fractures/dislocations of the patella

Widely displaced fractures of the patella are rare in children but are treated with tension banding just as in adults. More commonly seen in children is dislocation of the patella following a low-energy force across the patellofemoral joint. Aspiration of the acute hematoma and injection of local anesthetic into the joint are done in conjunction with initial reduction and immobilization. However, after a period of immobilization, assessment of the weight-bearing longitudinal axis, ligamentous stability, and poor secondary medial constraints such as a poor vastus medialis obliquus muscle provide the basis for rehabilitation. In skeletally immature teenagers, chronic subluxations and dislocations require realignment by such as osteotomy of the tibial tubercle.

Fractures of the tibial and fibular shafts

Simple, minimally displaced fractures of the tibial shaft, which are colloquially termed toddler's fractures, occur not only to toddlers but also in children in the first decade of life (Fig. 21). The integrity of the periosteum provides the additional constraint required to obtain and maintain reduction. Provisional healing occurs rapidly, allowing early weight-bearing in the cast. Displaced or unstable shaft fractures that cannot be managed in a cast are treated with external fixators or flexible nails inserted retrogradely (Fig. 22 and Fig. 23). The use of the rigid intramedullary nail is limited by its entry site through the tibial tubercle and therefore this nail can be safely used in a teenager only if there is little remaining growth of that tubercle.

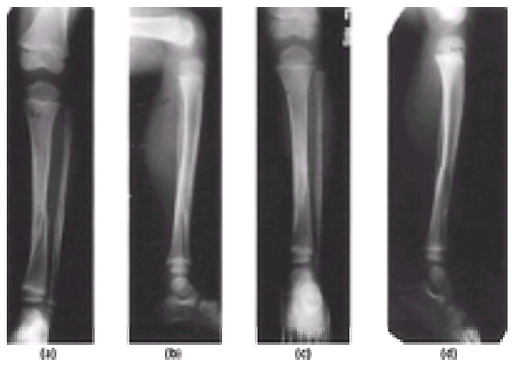


Fig. 21. (a) and (b) Low-velocity spiral fracture in a 3-year-old boy; note the fibula is still intact; (c) and (d) healed fracture.

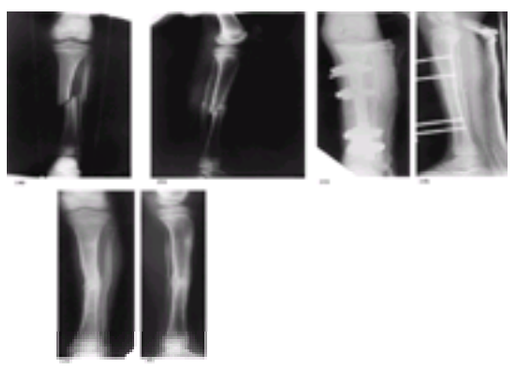


Fig. 22. (a) and (b) Grade II open fractures of the midshaft tibia–fibula in a 10-year-old boy who was struck by a car while riding his bike: (c) and (d) this high-velocity injury required four-compartment fasciotomies after irrigation and debridement; an external fixator was used for 1 month to maintain alignment in this unstable fracture and allow soft-tissue healing; (e) and (f) healed fractures.

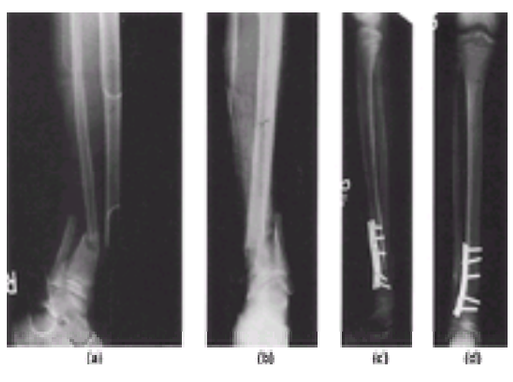


Fig. 23. (a) and (b) Grade III open distal tibia–fibula fractures in an 8-year-old boy who was run over by a car while crossing the street; (c) and (d) he required four-compartment fasciotomies and internal fixation after irrigation and debridement because of severe soft-tissue destruction and instability of the fracture.

Open fractures of the tibia and fibula require the same meticulous care as in the adult: irrigation and debridement of all nonviable tissues, assessment of compartment pressures, and reduction and stabilization. Overgrowth of the tibia and fibula can occur, but usually is clinically insignificant.

Fractures of the distal tibia–fibula and ankle

Fractures of the distal tibial metaphysis are usually amenable to closed reduction and casting (Fig. 24). Propagation of the fracture line into the physis is not always recognized. Physeal injuries, fractures of the malleoli, and syndesmotic injuries to the ankle cause any variety of complex fracture patterns. Treatment must correct displacement, joint incongruity, and stability of the mortise. Open or closed reduction and rigid internal or temporary, smooth K-wire fixation are all treatment strategies (Fig. 25 and Fig. 26). Attention to concomitant tibiotalar dislocation or subluxation and soft-tissue compromise is mandatory.



Fig. 24. Low-energy compression fracture of the distal tibia in a 2-year-old.

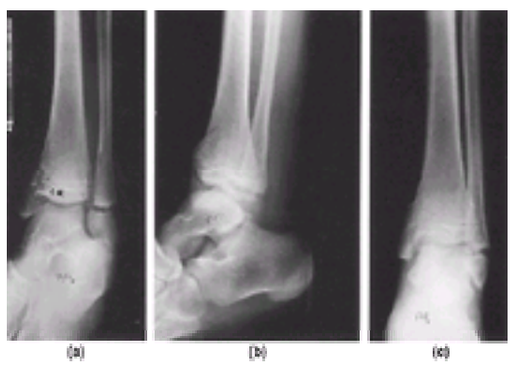


Fig. 25. (a), (b), and (c) Thirteen-year-old boy with a Salter–Harris IV fracture of the distal tibial physis with posteroinferior displacement of the distal fragment.



Fig. 26. (a) and (b) Minimally displaced Salter–Harris IV fracture of the distal tibia in a 9-year-old girl.

Fractures of the foot

Fractures of the hindfoot or dislocation of Lisfranc's joint are rare in children and usually the result of high-energy trauma such as gymnastic injuries (Fig. 27). Both joint stability and congruity must be achieved. Metatarsal fractures are common in children. Unrecognized trauma can produce cause a nondisplaced metatarsal fracture in a child, which presents as a subtle limp. Minimally angulated fractures of a metatarsal shaft are treated with a cast, but multiple fractures with other high-energy injuries and other ipsilateral fractures are managed with internal fixation such as provisional K-wires.

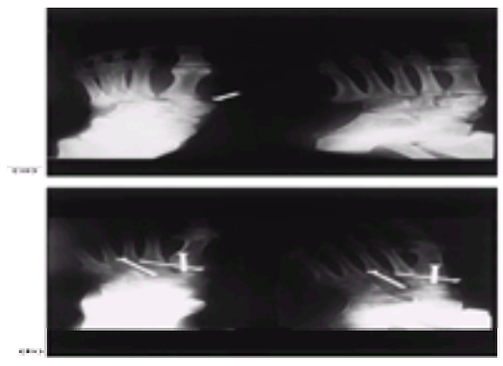


Fig. 27. Lisfranc fracture dislocation of the midfoot in a 15-year-old gymnast treated with open reduction/internal fixation.

Avulsion fractures at the base of the fifth metatarsal must be differentiated from irregular ossification centers or old avulsion injuries; contralateral views are helpful.

Fractures of the toes

While low-energy trauma and minimally displaced phalangeal fractures are common, children also sustain open fractures from water-sport, bicycle, and lawn-mower injuries. Crush injuries to the toes commonly occur in young children when they pull a heavy object down on to the foot. Partial amputation of the distal phalanx and destruction of the nailbed often follow this type of trauma.

Conclusion

Treatment of pediatric fractures requires the expertise of an orthopedic surgeon well versed in the principles governing the growth and development of the immature skeleton. and is both challenging and rewarding.

Further reading

General

Green NE, Swiontkowski MF. *Skeletal trauma in children*, Vol. 3. Saunders, Philadelphia, 1992.

Ogden JA. *Skeletal injury in the child.* Saunders, Philadelphia, 1990.

Rockwood CA, Wilkins KE, King RE. *Fractures in children*, 3rd edn. Lippincott, Philadelphia, 1991.

Tachdjian MO. *Pediatric orthopedics*, Vol. 4. Saunders, Philadelphia, 1990.

Webber BG, Brunner CH, Freuler F. *Treatment of fractures in children and adolescents.* Springer, Berlin, 1980.

Child abuse

Akbarina B, Torg JS, Kirkpatrick J, Sussman S. Manifestations of the battered child syndrome. *Journal of Bone and Joint Surgery* 1974; **56**A: 1159.

Rivara FP, Kamituska MD, Quan L. Injuries to children younger than one year of age. *Pediatrics* 1988; **81**: 93.

Fractures of the elbow

Brown IC, Zinar DM. Traumatic and iatrogenic neurological complications after supracondylar humerus fractures in children. *Journal of Pediatric Orthopedics* 1995; **15**: 440.

Buhl O, Hellberg S. Displaced supracondylar fractures of the humerus in children. *Acta Orthopaedica Scandinavica* 1982; **53**: 67.

Copley LA, Dormans JP, Davidson RS. Vascular injuries and their sequelae in pediatric supracondylar humeral fractures: toward a goal of prevention. *Journal of Pediatric Orthopedics* 1996; **16**(1): 99.

Cummings RJ, Jones ET, Reed FE, Mazur JM. Infantile dislocation of the elbow complicating obstetric palsy. *Journal of Pediatric Orthopedics* 1996; **16**(5): 589.

DeBoeck H, De Smet P, Penders W, De Rydt D. Supracondylar elbow fractures with impaction of the medial condyle in children. *Journal of Pediatric Orthopedics* 1995; **15**(4): 444.

Finnbogason T, Karlsson G, Lindberg L, Mortensson W. Nondisplaced and minimally displaced fractures of the lateral humeral condyle in children: a prospective radiographic investigation of fracture stability. *Journal of Pediatric Orthopedics* 1995; **15**(4): 422.

Foster DE, Sullivan JA, Gross RH. Lateral humeral condylar fractures in children. *Journal of Pediatric Orthopedics* 1985; **5**: 16–22.

Fowles JV, Silman N, Kassab MT. The Monteggia lesion in children: fracture of the ulna and dislocation of the radial head. *Journal of Bone and Joint Surgery* 1983; **65**A: 1276.

Gaddy BC, Strecker WB, Schoenecker PL. Surgical treatment of displaced olecranon fractures in children. *Journal of Pediatric Orthopedics* 1997; **17**(3): 321.

Garbuz DS, Leitch K, Wright JG. The treatment of supracondylar fractures in children with an absent radial pulse. *Journal of Pediatric Orthopedics* 1996; **16**(5): 594.

Gonzalex-Herrans P, Alvarez-Romera A, Burgos J, Rapariz JM, Hevia E. Displaced radial neck fractures in children treated by closed intramedullary pinning (Metaizeau technique). *Journal of Pediatric Orthopedics* 1997; **17**(3): 325.

Hudson DA, DeBeer JD. Isolated traumatic dislocation of the radial head in children. *Journal of Bone and Joint Surgery* 1986; **68**B: 378.

Keenan WNW, Clegg J. Variation of Baumann's angle with age, sex and side: implications for its use in radiological monitoring of supracondylar fracture of the humerus in children. *Journal of Pediatric Orthopedics* 1996; **16**(1): 97.

Letts M, Rumball K, Bauermeister S, McIntyre W, D'Astous J. Fractures of the capitellum in adolescents. *Journal of Pediatric Orthopedics* 1997; **17**(3): 315.

Ortega R, Loder RT, Louis DS. Open reduction and internal fixation of forearm fractures in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 651.

Papavasiliou VA, Nenopoulos S. Ipsilateral injuries of the elbow and forearm in children. *Journal of Pediatric Orthopedics* 1986; **6**: 58–60.

Reigstad A, Hellum C. Volkmann's ischaemic contracture of the forearm. *Injury* 1980; **12**: 148.

Sabharwal S *et al.* Management of pulseless pink hand in pediatric supracondylar fractures of the humerus. *Journal of Pediatric Orthopedics* 1997; **17**(3): 303.

Shaw BA, Murphy KM, Shaw A, Oppenheim WL, Myracle MR. Humerus shaft fractures in young children: accident or abuse? *Journal of Pediatric Orthopedics* 1997; **17**(3): 293.

Vahvanen V, Gripenberg L. Fracture of the radial neck in children: a long-term follow-up study of 43 cases. *Acta Orthopaedica Scandinavica* 1978; **49**: 32–8.

Wiley JJ, Galey JP. Monteggia injuries in children. *Journal of Bone and Joint Surgery* 1985; **67**B: 728.

Wyrsch B, Mencio GA, Green NE. Open reduction and internal fixation of pediatric forearm fracture. *Journal of Pediatric Orthopedics* 1996; **16**(5): 644.

Fractures of the forearm

Kay S, Smith C, Oppenheim WL. Both-bone midshaft forearm fractures in children. *Journal of Pediatric Orthopedics* 1986; **6**: 306.

Fractures of pelvis, hip, and proximal femur

Heiser JM, Oppenheim WL. Fractures of the hip in children: a review of forty cases. *Clinical Orthopedics* 1980; **149**: 177.

Torode I, Zieg D. Pelvic fractures in children. *Journal of Pediatric Orthopedics* 1985; **5**: 76–84.

Fractures of the distal femur and knee

Letts M, Vincent N. Gouw G. The ‘floating knee' in children. *Journal of Bone and Joint Surgery* 1986; **68**B: 442.

Riseborough EJ, Barrett IR, Shapiro F. Growth disturbances following distal femoral physeal fracture-separation. *Journal of Bone and Joint Surgery* 1983; **65**A: 885.

Shapiro, F. Fractures of the femoral shaft in children. *Acta Orthopaedica Scandinavica* 1981; **52**: 649–55.

Fractures of the femur

Aronson DD, Singer RM, Higgins RF. Skeletal traction for fractures of the femoral shaft in children: a long-term study. *Journal of Bone and Joint Surgery* 1987; **69**A: 1435.

Blasier RD, Aronson J, Tursky EA. External fixation of pediatric femur fractures. *Journal of Pediatric Orthopedics* 1997; **17**(3): 342.

Blademore LC, Loder RT, Hensinger RN. Role of international abuse in children 1 to 5 years old with isolated femoral shaft fractures. *Journal of Pediatric Orthopedics* 1996; **16**(5): 585.

DeSanctis N, Gambardella A, Pempinello C, Mallano P, Della Corte S. The use of external fixators in femur fractures in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 613.

Huber RI, Keller HW, Huber PM, Rehm KE. Flexible intramedullary nailing as fracture treatment in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 602.

Thompson JD, Stricker SJ, Williams MM. Fractures of the distal femoral epiphyseal plate. *Journal of Pediatric Orthopedics* 1995; **15**(4): 474.

Fractures of the tibia

Buckley SL, Smith GR, Sponseller PD, Thompson JD, Robertson WW, Griffin PP. Severe (type III) open fractures of the tibia in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 627.

Feldman DS, Otsuka NY, Hedden DM. Extra-articular triplane fracture of the distal tibial epiphysis. *Journal of Pediatric Orthopedics* 1995; **15**(4): 479.

Greiff J, Bergman F. Growth disturbance following fracture of the tibia in children. *Acta Orthopaedica Scandinavica* 1980; **51**: 315.

Kreder HJ, Armstrong P. A review of open tibia fractures in children. *Journal of Pediatric Orthopedics* 1995; **15**(4): 482.

Ogden JA, Ogden DA, Pugh L, Raney EM, Guidera KJ. Tibia valga after proximal metaphyseal fractures in childhood: a normal biologic response. *Journal of Pediatric Orthopedics* 1995; **15**(4): 489.

Robertson P, Karol LA, Rab GT. Open fractures of the tibia and femur in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 621.

Shin AY, Moran ME, Wenger DR. Intramalleolar triplane fractures of the distal tibial epiphysis. *Journal of Pediatric Orthopedics* 1997; **17**(3): 352.

Fractures of the ankle and foot

Cass JR, Peterson HA. Salter–Harris type IV injuries of the distal tibial epiphyseal growth plate, with emphasis on those involving the medial malleolus. *Journal of Bone and Joint Surgery* 1983; **65**A: 1059.

Karrholm J, Hansson LI, Selvik G. Changes in tibiofibular relationships due to growth disturbances after ankle fractures in children. *Journal of Bone and Joint Surgery* 1984; **66**A: 1198.

Kling TF, Jr, Bright RW, Hensinger RN. Distal tibial physeal fractures in children that may require open reduction. *Journal of Bone and Joint Surgery* 1984; **66**A: 647.

Schantz K, Rashmussen F. Calcaneous fractures in the child. *Acta Orthopaedica Scandinavica* 1987; **58**: 507.

Spiegel, PG, Cooperman DR, Laros GS. Epiphyseal fractures of the distal ends of the tibia and fibula: a retrospective study of two hundred and thirty-seven cases in children. *Journal of Bone and Joint Surgery* 1978; **60**: 1046.

Lawn-mower injuries

Farley FA *et al.* Lower extremity lawn-mower injuries in children. *Journal of Pediatric Orthopedics* 1996; **16**(5): 669.

Loder RT, Brown KL, Zaleske DJ, Jones ET. Extremity lawn-mower injuries in children: report by the research committee of the Pediatric Orthopedic Society of North America. *Journal of Pediatric Orthopedics* 1997; **17**(3): 360.

Vosburgh CL, Gruel CR, Heerndon WA, Sullivan JA. Lawn mower injuries of the pediatric foot and ankle: observations on prevention and management. *Journal of Pediatric Orthopedics* 1995; **15**(4): 504

Wrist

Aminian A, Schoenecker PL. Premature closure of the distal radial physis after fracture of the distal radial metaphysis. *Journal of Pediatric Orthopedics* 1995; **15**(4): 495.

46.3.2 Developmental dysplasia of the hip

Michael G. Ehrlich and L. Brett Babat

Treatment
Further reading

Developmental dysplasia of the hip encompasses a spectrum of abnormalities in which any or all of the components, including the capsule, the proximal femur, and the acetabulum, are defective. The condition manifests as subtle clicks to frank dislocation, which usually occurs at delivery or in the perinatal period. The etiology is believed to be related to laxity induced by the maternal hormone relaxin, which enables the cervix to dilate during labor. Though its exact role has yet to be determined, it is thought that relaxin crosses the placenta, presumably affecting the ligaments of the hip capsule. The incidence of developmental dysplasia of the hip in the United States is approximately 10 cases per 1000 live births, with a male to female ratio of 1:6. It has been estimated that developmental dysplasia of the hip is the cause of 20 to 50 per cent of degenerative arthritis of the hip in adults.

Normal hip development results from balanced growth of the acetabular, triradiate, and proximal femoral growth centers in response to a properly located femoral head. The goal of treatment of a dysplastic hip is thus the attainment and maintenance of a concentric reduction. The key to successful management is early treatment. During the first few days of life, and possibly longer, the hip can slip in and out of the acetabulum. When the femoral head is held reduced early, then, as the ligaments tighten, the hip remains concentrically located. If, however, the hip is left dislocated, reduction becomes more difficult as the ligaments tighten.

Additional factors are thought to play a part in the genesis of developmental dysplasia of the hip. Intra-uterine position is important, particularly with oligohydramnios and breech deliveries, when the femoral head may be levered out of the socket. Hyperextension has also been implicated. Our studies have shown that, even when the musculature is removed from the hip of a stillborn fetus, the joint remains stable. This is largely due to a negative intra-articular pressure; if air is introduced into the joint by a needle, the pressure equilibrates, and the hip is easily dislocated. Salter has demonstrated that hyperextension of a joint leads to breaking of the seal, allowing gas to enter. Thus, the old custom of slapping the newborn on the buttocks has been abandoned for fear of dislocating the hip.

At birth, most patients with developmental dysplasia of the hip have ‘dislocatable hips.' If these are allowed to remain out of the socket for several weeks, they become dislocated, and Ortolani's sign becomes negative, as the hip is not easily reduced (Fig. 1). Some hips, however, are dislocated at birth. These are often teratologic in origin, resulting, for example, from arthrogryposis or muscle imbalance. If the hip has been dislocated *in utero*, a false acetabulum may already be present at birth (Fig. 2). Closed treatment usually fails to reduce these hips.



Fig. 1. An infant with dislocation of the left hip and subluxation of the right hip. The left hip has been out for several months. Note the early formation of a false acetabulum on the left.



Fig. 2. Bilateral dislocations in an arthrogrypotic. There is almost no true acetabulum on either side. The false acetabulums were present before birth.

The subluxed hip represents another category of dysplasia. Unfortunately, this term has been used for widely disparate conditions. It most appropriately designates a femoral head that is in the acetabulum, but not perfectly seated (Fig. 3). This condition is most commonly the sequela of an adducted position *in utero*, which persists after birth. The resulting limited abduction prevents the socket from developing well.



Fig. 3. In this radiograph, the right hip is subluxed. The femoral head is displaced laterally, but not dislocated. Note the elevated acetabular index.

Most orthopedic surgeons believe that, at least initially, the bony acetabulum is normal. In hips reduced around the time of birth, radiographs remain normal. If, however, the hip is permitted to remain dislocated, there will be bony changes visible on radiograph. Furthermore, infants with *in utero* dislocations frequently have flattened acetabulums at delivery.

In the newborn, the diagnosis is made with the Ortolani and Barlow maneuvres. The Ortolani maneuvre is a reduction performed by abducting the hip, with a resulting

‘clunk' felt by the examiner. The patient is placed in the supine position. Both hips are flexed and adducted. With the index and middle fingers over the greater trochanter, the hip is slightly abducted while lifting the femoral head forward, into the socket. If the posterior acetabular lip is shallow, the click may be subtle; pressing the hip against the acetabulum during the reduction may accentuate it.

The Barlow maneuvre is used to determine if the hip is dislocatable. Again, the infant is placed supine. The untested hip is flexed to 90 degrees and mid-abduction. The tested hip is in 45 to 60 degrees of flexion and slight adduction. With the examiners hands in the same position as for the Ortolani, the femoral head is gently pressed posteriorly, in an attempt to dislocate the hip. A positive test will be felt as a ‘clunk.' On release of pressure, the hip will re-locate, with a second palpable ‘clunk.' The test is then repeated for the opposite hip.

A ‘soft click' is frequently felt during infancy and early childhood. In the past, this was thought to be due to a tendon, possibly the tensor fascia lata, snapping over the trochanter. Few orthopedists treated these patients, and most patients had no problems. With the increasing use of ultrasound, however, it has been suggested that these clicks actually represent partial dislocation, or subluxation over the cartilaginous limbus, not dislocation *per se*. While the exact significance of this finding is not yet known, some orthopedists treat these hips as if they were truly dislocatable. The approach taken by the senior author is to place these infants in a Pavlik harness (see ‘Treatment', below) for 1 month, in the hope that the hip will tighten. If radiographs are normal at that point, treatment is discontinued and the patient is observed, although the ‘clicks,' may persist for years.

In older children, the Barlow and Ortolani tests are negative. The main diagnostic sign is a lack of abduction. Testing both hips simultaneously can be misleading; the pelvis may rock, resulting in abduction appearing symmetric. The hips should be tested individually, with the opposite hip used to stabilize the pelvis by pushing posteriorly (Fig. 4(a),Fig. 4(b) and Fig. 4(c)).



Fig. 4. Testing abduction in a child with a dislocated right hip. (a) With the right hip stabilized in neutral, the left hip is abducted. (b) The right hip demonstrates marked loss of abduction. (c) When the hips are tested simultaneously, the pelvis rocks, and abduction falsely appears symmetric.

Other physical findings include the Allis or Galleazi sign (Fig. 5) and asymmetry of the gluteal creases (Fig. 6). Galleazi's sign is elicited by examining the supine patient on a firm surface. The hips are flexed, and the buttocks are pressed posteriorly. The dislocated extremity will appear shorter than the other. However, in bilateral dislocations, these tests are unreliable. Abduction is equally limited, and leg lengths and gluteal folds may appear symmetric. A helpful finding in these cases is widening of the perineum. Normally, when the hips are adducted, the perineum is masked by the legs coming together. If the hips are both dislocated, the perineum is more exposed.



Fig. 5. A positive Galleazi sign. The leg with the dislocated left hip appears shorter than the right.



Fig. 6. Asymmetric gluteal folds in a child with a dislocated left hip.

The relationship of the femoral head to the acetabulum is used to diagnose dislocations radiographically. The head is not ossified until 3 months in girls, and until 6 months in boys. Therefore, lines are drawn to determine its position. Hilgenreiner's line is drawn through the tri-radiate cartilage of both acetabulums. Perkins' line is a vertical drawn through the lateral edge of the acetabulum. The femoral metaphysis should be equally inferior to Hilgenreiner's line and medial to Perkins' line (Fig. 7). Shenton's line is a smooth arc drawn along the medial femoral shaft and the superior border of the obterator foramen. Interruption of this line indicates that the femoral head is not concentrically seated in the acetabulum. In addition to these findings on plane radiographs, computed tomography and magnetic resonance imaging studies are occasionally used to establish the exact position of the hip. However, ultrasound is rapidly becoming the standard, as there is no radiation, and dynamic evaluation of the hip is easily performed.



Fig. 7. Radiograph of a normal pelvis with illustration of Hilgenreiner's line (A), Perkins' line (B), Shenton's line (C), and the acetabular index (D).

On the frog-leg lateral view, the hip should point to the tri-radiate cartilage. A line drawn from the tri-radiate cartilage to the lateral edge of the acetabulum defines an angle with Hilgenreiner's line. This is the acetabular index, and is normally about 30 degrees. In the newborn, much of the acetabulum is cartilage, so slight tilting of the pelvis gives widely differing angles. Because the acetabulum is a growth plate, if the head does not make contact with it, there is no suppression of growth. When the hip has been dislocated for some time, the arc of the roof of the acetabulum becomes shallow, and the acetabular index increases above 30 degrees. If the hip remains dislocated, with the head articulating above the acetabulum, an indentation forms in the iliac wall, resulting in a ‘false acetabulum.'

Treatment

The Pavlik harness is the standard treatment for the dislocatable hip discovered at birth (Fig. 8). This device holds the hip reduced by maintaining greater than 90° flexion and moderate abduction. While adduction is prevented, the harness should not be used to hold the hip in maximal abduction, which may cause avascular necrosis. The patient is seen weekly. Radiographs need not be repeated with each exam; progress may be followed with ultrasound.



Fig. 8. An infant in a properly adjusted Pavlik harness.

Abduction improves quickly in subluxed hips; within a month, the harness may be switched to night use only. For dislocated hips, especially those that have been untreated for as long as 6 months, the harness is adjusted to hold the hip in even more flexion. The hip is prevented from adducting, and it is hoped that the child will kick the hip into position. The severe flexion should be reduced as soon as possible to prevent posterior dislocation.

If successful, the harness may be exchanged for an abduction splint after 6 weeks, although some physicians continue to use the harness for the entire 3-month treatment period. At this point, if the acetabulum is developing well, and the hip is no longer dislocatable, treatment may be stopped. When the diagnosis is made in the newborn, treatment in the harness is successful in 85 to 95 per cent of patients. If the hip has not reduced by 3 to 4 weeks, or if the hip still dislocates at 6 weeks, the Pavlik harness is considered to have failed.

The senior author's preference for late dislocations, from 6 months to 1 year, is traction, adductor tenotomy, and spica cast application. This method is tried earlier if reduction is not obtained with the harness. The goal is to obtain gentle closed reduction, as forced reduction frequently leads to avascular necrosis. Various forms of traction may be employed. All are applied with the hip in a flexed position to relax the iliopsoas and prevent it from lying across the acetabulum and blocking reduction. Furthermore, experimental evidence suggests that blood flow to the hip is improved in the flexed position. Though one approach is to use Bryant's traction, with both hips pulled straight up, we prefer Buck's skin traction, with the addition of bolsters under the thighs, to relax the hamstrings at the same time that the hips are flexed. Increasingly, traction has been successfully utilized in a home program.

Two or 3 weeks are usually required for traction to be effective. Radiographs should demonstrate that the hip is being pulled down to the level of the acetabulum. Once the head is opposite the socket, or when it is easily reducible, a subcutaneous adductor tenotomy is performed. While some have abandoned tenotomy, it is useful for relieving pressure on the head, minimizing the risk of avascular necrosis. An arthrogram is performed to ensure the hip is reduced and well seated, without a soft tissue block to reduction (Fig. 9).

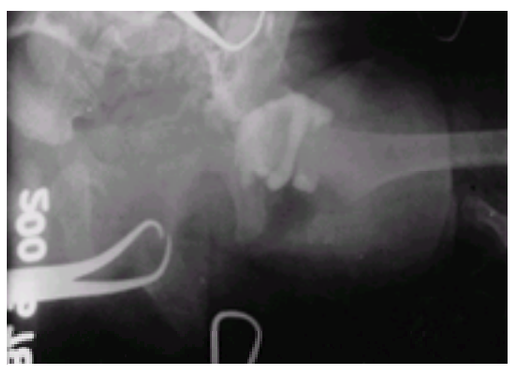


Fig. 9. An arthrogram of a hip with interposed soft tissue blocking full reduction, demonstrated by an increased dye space.

The hip is best seated in 90° or more flexion, and abducted 10° to 20° past the point of reduction. Over-abduction stretches the medial circumflex vessels, increasing the risk of avascular necrosis. As the fat and muscles atrophy during cast immobilization, the hip becomes prone to posterior dislocation. To prevent this, the surgeon must push anteriorly on the trochanter, and the cast should be molded accordingly. Malvitz and Weintstein (1994) found that functional results have been excellent at up to 30 years follow-up when traction and adductor tenotomy are used in concert. The frequency of degenerative changes is markedly decreased compared with hips treated without tenotomy. Stromqvist and Sunden (1989) reported avascular necrosis in fewer than 5 per cent of patients treated between 2 and 12 months of age. They reported good results in 97.5 per cent of patients at 2 to 11 years of follow-up.

When closed reduction cannot be obtained, an open procedure becomes necessary. This is usually the case in teratologic dislocations, those in children over 1 year of age, as well as in early dislocations that prove irreducible, or that have a soft tissue block to concentric seating. The medial approach allows release of the often-obstructing transverse acetabular ligament. While some reported series show a high risk of avascular necrosis with this approach, it is still utilized in many parts of the United States. We prefer an anterior Smith-Peterson approach, between the sartorius and tensor, elevating both heads of the rectus femoris, and cutting the

psoas tendon to permit reduction. This facilitates a capsulorraphy, which holds the head better reduced. The capsule is opened with a ‘T' incision, and the ligamentum teres is used as a guide to the true acetabulum. With the head reduced, the capsule is imbricated to secure the reduction (Fig. 10).

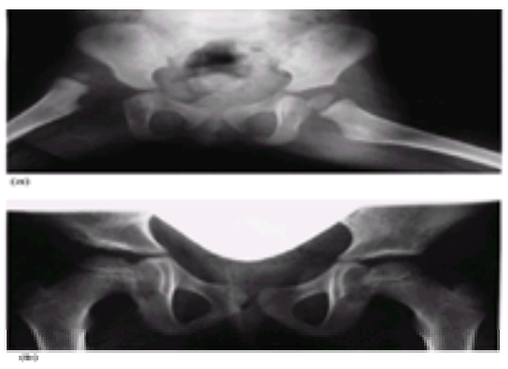


Fig. 10. (a) Unilateral right hip dislocation in an 18-month-old child. The acetabulum is minimally developed, and the femoral head is developmentally delayed. (b) The same child at 5 years of age, after an open reduction and capsulorraphy.

In children over 3 years of age, traction has been forgone in favor of femoral shortening, which is thought to reduce pressure on the head by functionally lengthening the soft tissues. In 1984, Schoenecker and Strecker reported a 54 per cent incidence of avascular necrosis and 31 per cent redislocation rate in children in this age group treated with traction and open reduction, compared with no avascular necrosis and 8 per cent redislocation in those who underwent a femoral osteotomy and open reduction. The cut is made just distal to the greater trochanter, and the hip is then reduced. A small dynamic compression plate is then used to hold the osteotomy. A spica cast is placed for 3 months. Usually, further bracing is not needed, although some prefer to employ it until there is evidence of acetabular deepening. In pathologically lax patients, however, even a well-reduced hip may later have a widened teardrop-head distance. Prolonged bracing is recommended in this group of patients.

For hips that have been dislocated for more than 18 months, there may not be sufficient growth potential for the acetabulum to become normal. The pathology dictates the appropriate procedure. A hip in excessive valgus is treated by a varus osteotomy, ducking the head into the acetabulum. For a shallow acetabulum with a normal femoral head and neck, an acetabuloplasty is performed. When both aspects of the hip are abnormal, both types of procedures are performed.

A common error is made in children with marked anteversion. These hips appear to be well-seated on internal rotation radiographs, although not in neutral or external rotation. Often, the surgeon will perform an external rotation osteotomy, with the intent of forcing the child to hold the hip internally rotated, and, thus, reduced. This does not occur. The correct procedure is to perform a capsulotomy, internally rotate, and seat the hip, then imbricate the capsule, ensuring maintenance of the reduction. A distal femoral osteotomy is then performed, externally rotating the distal segment.

There are multiple osteotomies that may be employed to change the orientation of a dysplastic acetabulum, thereby increasing the coverage of the femoral head. A Salter osteotomy of the innominate bone is generally the senior author's preference. A Gigli saw is placed in the sciatic notch, and a cut is made to a point just superior to the anterior inferior iliac spine. The inferior fragment is then rotated forward, and a wedge of bone from the anterior superior iliac spine is held between the fragments with Kirschner wires or threaded pins (Fig. 11). Salter and Dubos (1974) reported excellent results in 93.6 per cent of their patients over 15 years, and Gulman et al. (1994) had similar (88.4 per cent good to excellent) outcomes.

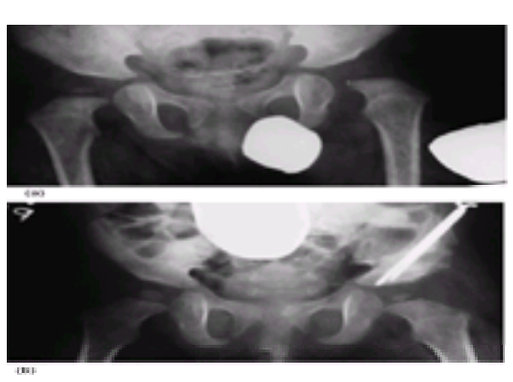


Fig. 11. (a). This child's dislocated left hip was initially treated in excessive abduction at another institution. Her hip remains subluxed, and avascular necrosis has developed. (b) The same child, after open reduction and Salter osteotomy. Note the normalized acetabular index and the healed bone graft, taken from the ipsilateral ilium.

In older children, the pubic symphysis upon which the osteotomized segment hinges is no longer supple. A Sutherland osteotomy, which cuts through the superior and inferior pubic rami just lateral to the symphysis, or a Steele triple osteotomy, with cuts through both the pubis and the ischium (Fig. 12), may then be employed. Osteotomies close to the acetabulum, such as the Dial and Wagner osteotomies, allow the acetabulum to rotate freely, but are technically very difficult, and are rarely necessary.



Fig. 12. Intra-operative radiograph of a Steele triple osteotomy.

Dislocated hips can generally be reduced up to 5 years of age. After this, results are poor. Bilateral dislocations are usually left in place after this point. These hips are usually quite symmetric; treatment frequently results in failure on one side, with resulting asymmetry. In contrast, even in older children, an attempt is usually made to reduce a unilateral dislocation, to provide symmetry.

Further reading

Buchanan JR, Greer RB, Cotler JM. Management strategy for the prevention of avascular necrosis during treatment of congenital dislocation of the hip. *Journal of Bone and Joint Surgery* 1981; **63A**: 140–6. [Effect of treatment options on the occurrence of avascular necrosis.]

Camp J, Herring JA, Dworezynski C. Comparison of inpatient and outpatient traction in developmental dislocation of the hip. *Journal of Pediatric Orthopedics* 1994; **14**: 9–12. [Demonstrated that outpatient traction is as safe and as effective as an inpatient program.]

Gulman B, Tuncay IC, Dabak N, Karaismailoglu N. Salter's innominate osteotomy in the treatment of congenital hip dislocation: a long-term review. *Journal of Pediatric Orthopedics* 1994; **14**: 662–6.

Malvitz TA, Weinstein SL. Closed reduction for congenital dysplasia of the hip: functional and radiographic results after an average of thirty years. *Journal of Bone and Joint Surgery* 1994; **76A**: 1777–92. [Gives long-term follow-up of 152 hips treated by closed reduction.]

Mubarek S, Garfin S, Vance R, McKinnon B, Sutherland D. Pitfalls in the use of the Pavlik harness for treatment of congenital dysplasia, subluxation, and dislocation of the hip. *Journal of Bone and Joint Surgery* 1981; **63A**: 1239–48. [Details common problems with, and the correct usage of, the Pavlik harness.]

Ortolani M. The classic: congenital hip dysplasia in the light of early and very early diagnosis. *Clinical Orthopaedics* 1976; **119**: 6–12. [Forty years of experience with the early treatment of 8000 children with developmental dysplasia of the hip.]

Salter RB, Dubos JP. The first fifteen years' personal experience with innominate osteotomy in the treatment of congenital dislocation and subluxation of the hip. *Clinical Orthopaedics* 1974; **98**; 72–103.

Schoenecker PL, Strecker WB. Congenital dislocation of the hip in children: comparison of the effects of femoral shortening and of skeletal traction in treatment. *Journal of Bone and Joint Surgery* 1984; **66A**: 21–7. [Demonstrated that femoral shortening is preferable to traction in patients more than 3 years of age.]

Stromqvist B, Sunden G. CDH diagnosed at 2 to 12 months of age—treatment and results. *J Pediatric Orthopedics* 1989; **9**: 208–12.

Tachdjian MO. Congenital dysplasia of the hip. In: Tachdlian, MO, ed. *Pediatric orthopedics*, pp. 297–548. WB Saunders, Philadelphia, 1990. [Excellent descriptions and illustrations of techniques and procedures.]

46.3.3 Legg–Calvé–Perthes' disease

Michael G. Ehrlich and Craig P. Eberson

Etiology
Diagnosis
Outcome and treatment
Further reading

Etiology

Legg–Calvé–Perthes' disease represents an idiopathic avascular necrosis of the proximal femur in the growing child. The disease affects boys in a ratio of 4:1 over girls, primarily occurring between the ages of 4 and 8 years. It may, however, be seen in children as young as 2 or as old as 12 years. Twin studies have failed to determine a genetic basis for the disease. The difficulty in arriving at a universally accepted treatment protocol stems from a lack of full understanding of the pathogenesis of the disease; numerous etiologies have been postulated, but to date the underlying cause has yet to be elucidated. The final common pathway in Perthes' disease is avascular necrosis. Investigators have shown that multiple infarctions are necessary to produce the characteristic histopathologic picture of necrotic woven bone overlying necrotic trabeculae. It has not been possible, however, to reproduce the radiographic picture in an animal model.

To understand the pathophysiology of Perthes' disease, it is important to study the evolution of the blood supply of the proximal femur in the developing child. After the first year of life, few children have vessels which cross the physis. Chung, in his necropsy studies of proximal femora from children of various ages, described the presence of two anastomotic rings which supply the head of the femur. The first is extracapsular and is formed by the medial and lateral femoral circumflex arteries. The second is intracapsular, and is more frequently incomplete in boys. This fragile vascular sling, the retinaculum of Weitbrecht, can be easily damaged; while avascular necrosis is the result, the characteristic picture of Perthes' disease is not reproduced. An anterior and medial arcade of vessels is present in the newborn to 2-year-old age group, but often is absent in the 2 to 8-year-old patient. In addition, the distance between the femoral head and neck and the greater trochanter in the child less than 8 years old is narrow, which may be responsible for increased vulnerability to vascular compression. These findings may explain the age and sex distributions seen in Perthes' disease.

It is enticing to subscribe to the notion that Perthes' disease has a traumatic etiology; some authors attribute the increased incidence in boys to their propensity toward traumatic injuries when compared with their female counterparts. The intracapsular hematoma responsible for increased pressure and bone death seen in children with traumatic avascular necrosis, however, is excruciatingly painful; this is not characteristically seen in Perthes' disease. In addition, the hip with Perthes' disease undergoes characteristic radiographic stages not present in simple post-traumatic avascular necrosis. Toxic SYnovitis, a presumably viral affliction, causes pain and restricted motion in the young child's hip. Some authors have postulated an etiologic role of synovitis in Perthes' disease via increased intracapsular pressure. Various long-term studies exist that report an incidence of Perthes' disease developing after toxic SYnovitis which ranges from 1 to greater than 20 per cent. It is likely, however, that the synovitis in these patients is in reality undiagnosed Perthes' disease, rather than a separate, preceding viral illness. Multiple authors have demonstrated abnormalities in the venous drainage system of the proximal femur, and necrosis has been induced in dog models through occluding the venous outflow with silicone injection. Recently, abnormalities of the clotting cascade have been cited. A mutation in the gene for factor V, resulting in resistance to activated protein C and subsequent thrombophilia, has been identified in a group of Perthes' patients and their families. Other studies cite decreased levels of protein C, protein S, and tissue plasminogen activator, as well as increased levels of plasminogen activator inhibitor and lipoprotein A. Multiple endocrine abnormalities, such as delayed bone age, decreased levels of somatomedin C, and thyroid dysfunction, have been identified as well in some cases. The low incidence of bilateral involvement (10 to 12 per cent), however, does not support an endocrine or systemic cause. In summary, the only statement about the etiology of Perthes' disease which can be made with certainty is that the true precipitory factor has not been found; the onset of disease is likely to be a result of multiple factors acting in a predisposed child.

Diagnosis

Most patients with Perthes' disease present between the ages of 4 and 8 years with an asymptomatic limp. Boys are affected four or five times as frequently as girls. The next most common presenting symptom is pain, often involving the hip, groin, or knee. Initially there may be muscle spasm and limited motion, especially abduction and internal rotation. Later findings may include muscle atrophy and fixed contractures, the latter occasionally having to be addressed surgically.

Four characteristic radiographic stages (Fig. 1(a),Fig. 1(b),Fig. 1(c) and Fig. 1(d)) have been described in Perthes' disease, the first of which being the inflammatory stage. The earliest finding may be a widening of the ‘head-teardrop distance', that is, the head is displaced away from the medial wall of the acetabulum. In normal children, this distance is equal within 1 mm of the distance on the uninvolved side. In toxic synovitis, this distance rarely exceeds 2 mm, while in Perthes' disease this distance is increased to a further degree, as a result of synovitis, cartilage overgrowth, change in the shape of the head, and/or hypertrophy of the ligamentum teres. The second finding is usually the ‘crescent sign', best seen on the frog-lateral radiograph. This represents a subchondral fracture as a result of revascularization of the head. While this sign is seen early in the disease presentation, it actually represents femoral head healing and return of blood flow. The second stage, known originally as the avascular stage, is heralded by an increased density in the femoral head; various explanations exist for this radiographic appearance. Calcification of the marrow fat has been suggested, as has collapse and impaction of the necrotic trabeculae. Some feel that the surrounding bone becomes osteopenic, thus the head is only relatively dense. The prevalent opinion today is that this density represents creeping substitution; new bone is laid down on top of necrotic trabeculae. In the third stage, a large portion of the head appears fragmented and is resorbed. The fourth stage involves healing and reconstitution of the femoral head.



Fig. 1. (a) Widening of the ‘head-teardrop distance', seen on the right when compared with the normal left hip. (b) The crescent sign, representing a subchondral fracture, is demonstrated in the right hip. (c) Sclerosis of the right femoral head is evident. (d) The resorption phase, seen ere on the right. (e) The healing phase. Note the residual flattening of the right femoral head.

The differential diagnosis of Legg–Calvé–Perthes' disease is vast and includes blood dyscrasias (such as sickle cell disease), corticosteroid use, and traumatic avascular necrosis. In the 10 to 12 per cent of cases that are bilateral, the presentation is usually sequential; one hip will show involvement after the onset of the disease in the other. Bilateral concomitant presentation suggests an endocrinopathy, such as hypothyroidism, or a familial dysplasia, such as spondyloepiphyseal dysplasia or multiple epiphyseal dysplasia.

In addition to plain radiographs, several other modalities are useful in the diagnosis of Perthes' disease. Magnetic resonance imaging (**MRI**) allows visualization of the involved portion of the head, as well as the shape of the cartilagenous epiphysis (Fig. 2). This imaging technique represents an important tool in allowing the early diagnosis of the disease. Bone scintigraphy has also been described as useful in detecting impaired blood flow to the femoral head. While several authors have cited its prognostic value, we have not found it to be an especially useful tool. At the time of presentation, the patient is usually in the revascularization stage of the disease, and the bone scan does not provide any additional information over the plain films. Its role will most likely be supplanted by MRI. Arthrography (Fig. 3(a),Fig. 3(b) and Fig. 3(c)) enables the surgeon to observe the relationship between the acetabulum and the femoral head during motion when performed under fluoroscopy in the operating room. Using this technique, the cartilagenous portion of the femoral head can be seen, and the amount of true flattening assessed. Computed tomography (**CT**) is helpful in identifying loose bodies or osteochondral fragments within the joint. All of the methods have their supporters, and are often used in conjunction when making the diagnosis of Perthes' disease.



Fig. 2. The characteristic MRI changes of Perthes' disease, seen here in the right hip.

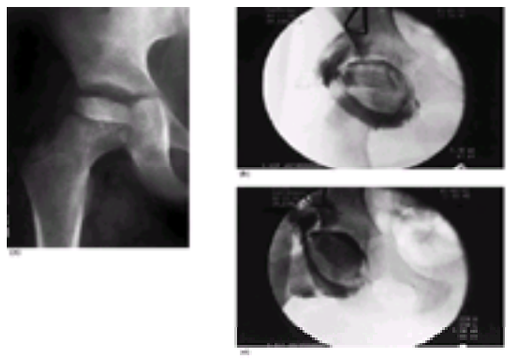


Fig. 3. (a) A 6.5-year-old male presenting with hip pain. A crescent sign can be seen, as well as some flattening of the head. (b) An arthrogram in neutral position reveals severe flattening of the femoral head. In addition, enlargement of the head out of the confines of the acetabulum is seen.(c) An arthrogram in abduction reveals better coverage of the head in this position. Also important is the positioning of the flat part of the femoral head beneath the round part of the acetabulum. This allows remodeling of the head to occur so that its shape matches that of the acetabulum

Outcome and treatment

Since prospective, randomized, double-blinded studies on the outcome of Perthes' disease have not been performed, the natural history of this disorder has to be inferred from a variety of studies using patients who received minimal treatment and from reviews of older patients with osteoarthritis presumed to have had Perthes' disease in the past. The majority of studies show that the most important prognostic factors for developing osteoarthrosis are age of onset and ‘roundness' of the femoral head. In most long-term follow-up series lasting 20 to 40 years, the typical patient has little pain, minimal shortening, and minimal or no impairment in activities of daily living or employment responsibilities. Studies longer than 40 years, however, show a marked reduction in hip function. The vast majority of patients require total hip arthroplasty by the sixth and seventh decades of life. McAndrew and Weinstein found the incidence of disabling osteoarthritis in their group of 35 patients with Perthes' disease to be greater than 50 per cent at 48-year follow-up, with the prevalence of osteoarthritis in this group being 10 times that of the general population.

There is no consensus as to which groups of patients should be treated, or even which of the various modalities is the most likely to be successful. Despite this confusion, there are certain principles of treatment which guide the surgeon. Numerous classification systems exist which grade the amount of involvement of the head; Catterall, Salter, and Herring all have classification systems bearing their names. Caterall grade I represents involvement of only the superolateral head, while grade II is indicative of involvement of up to one-half of the head. In grade III hips nearly all of the head is affected, and grade IV represents whole head involvement. Salter groups Catterall grades I and II as ‘A' in his classification, with III and IV representing Salter ‘B'. Herring felt that the lateral column of the head represents a strut to support weight-bearing. His grade ‘A' hips have no lateral pillar involvement, grade ‘B' represents less than 50 per cent involvement, and grade ‘C' represents greater than 50 per cent involvement. The common theme of all of the classification schemes is that patients with minimal head involvement (i.e. Catterall I, Herring ‘A', Salter ‘A') generally do well without treatment, while the more severely involved heads (i.e. Catterall III and IV, Herring ‘C', Salter ‘B') require intervention. Intermediate patients (Catterall II, Herring ‘B') form a ‘gray zone'. Unfortunately, when the patient is first seen, it is often not possible to determine the amount of eventual head involvement. To assist in decision making, it is helpful to look for the presence of ‘head-at-risk signs', first described by Catterall. Probably the most important of these are calcification lateral to the epiphysis and lateral displacement of the head. Lateral calcification is indicative of a head which will be large when fully ossified. A laterally displaced head suggests either a head which is large and extruded from the acetabulum or abundant synovitis pushing the head out. These factors suggest that the femoral head will not be concentrically seated in the acetabulum at the endstage of the disease, and thus will not be spherical. Stulberg and colleagues described a classification system useful in determining outcomes in patients with Perthes' disease. They found that patients with spherical heads which fit concentrically within the acetabulum (i.e. were congruous), tended to have minimal residual symptoms, as well as the best prognosis. Conversely, patients with aspherical, incongruous heads tended to develop SYmptoms by the end of the fourth decade. Thus, the goal of any treatment regimen should be a spherical, congruous head. Should this not be possible, an aspherical head which is congruent with the acetabulum and allows a full range of motion is optimal (Fig. 4). Finally, if neither of these outcomes is possible, a salvage procedure, such as a Chiari osteotomy, should be performed to allow coverage for the femoral head; that is, the acetabulum should be recontoured to accommodate an enlarged, misshapen femoral head.



Fig. 4. The femoral head, while misshapen and flattened, remains congruent with the acetabulum. Note the large size of the head, which has overgrown the acetabulum laterally.

A consideration which should not be underestimated is that of patient age. The femoral head epiphyseal cartilage is a growth plate responsive to changes in pressure, as is the corresponding acetabular cartilage. As the head subluxes laterally out of the confines of the acetabulum, the uncovered portion no longer is compressed by the acetabulum, and thus overgrows and becomes malformed. The flat portion of the femoral head abuts the flat portion of the acetabulum, and is unable to remodel. If the head is excessively overgrown lateral to the acetabulum, hinged abduction (i.e. the head pivots on the edge of the acetabulum rather than rotating through a smooth arc) and functional retroversion (a large overgrown head anterior to the acetabulum restricts internal rotation, as if the head was excessively retroverted with respect to the femoral shaft) occur. A young child has the ability to remodel the acetabular cartilage in response to a flat head, resulting in a congruous joint. Even in cases of severe involvement of the head, little treatment is required as long as range of motion is maintained; young children have the capacity to remodel the head and improve its sphericity even after the final stage of the disease. It is unusual, therefore, that the young patient presenting with Perthes' disease requires treatment other than maintenance of range of motion. In the child older than 6 years of age, however, much of this ability is lost, and further intervention is considered. Late complications of Perthes' disease include physeal arrest leading to shortened limb length and relative trochanteric overgrowth. Occasionally these conditions require

treatment as well.

Treatment of Perthes' disease has two phases. The first is regaining as much motion as possible. This is achieved through bed rest and traction to relieve associated SYnovitis and muscle spasm, followed by a physical therapy program. The second phase is described as ‘containment'. By keeping the head within the confines of the acetabulum, it is possible to promote a congruent joint. When the flat part of the femoral head lies beneath the round portion of the acetabulum, the pressure is decreased, thus allowing the head to grow until it contacts the acetabular roof. This remodeling promotes congruence, even in the aspherical head.

Abduction bracing has its advocates, and in some centers is widely used. Recent literature, however, suggests that this modality may not be as effective as previously believed. Martinez and colleagues had no good results in any of the 34 hips they treated in an Atlanta Scottish Rite Orthosis, the most commonly used abduction brace. Their group consisted of grade III and IV hips, and thus they felt that bracing was not an acceptable option for patients with severe involvement of the head.

Various surgical methods have been proposed, and the decision lies with the surgeon as to which is the most appropriate for an individual patient. Again, range of motion must be maintained throughout the course of treatment. Femoral osteotomies (Fig. 5(a),Fig. 5(b)) are performed via a lateral approach through the vastus lateralis. The subtrochanteric region of the femur is cut in a ‘dome' fashion, and subsequently fixed in varus with a dynamic compression plate, with the upper screws placed in the neck. Some authors advocate retroverting the femoral head at the same time, based upon MRI or arthrograms demonstrating a more congruent relationship between the head and acetabulum in this position. While femoral osteotomy results in fewer cases of stiff hips than other methods, it is not without complications. If the head is placed in an excessive amount of varus or the greater trochanteric physis overgrows, a valgus osteotomy may be required in the future as a corrective measure. Even with the growth stimulation seen in these patients, as a result of increased blood flow to the area, there may be shortening of the limb. Alternatively, the trochanteric physis may be curetted at the time of the initial surgery or at a later date. Shortening the lever arm of the abductors, as a result of the varus positioning of the femoral head, may result in an abductor lurch, requiring a trochanteric advancement or trochanteric physeal arrest procedure. Barring these complications, a second procedure is still required in order to remove the compression plate.

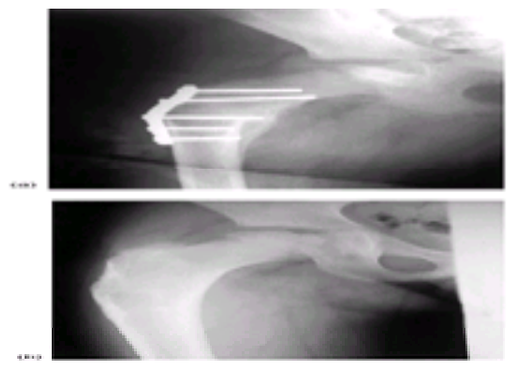


Fig. 5. (a) Radiographs of the same patient in Fig. 3, after varus osteotomy. (b) Following hardware removal the femoral head is seen to have remodeled to a more round shape and is better contained within the acetabulum than was seen on preoperative radiographs.

Pelvic osteotomies are another option, but have been cited to have a higher incidence of recalcitrant postoperative stiffness than other procedures. In addition, most series evaluating the efficacy of pelvic osteotomies include only patients with congruent heads at the initiation of treatment, thus it can be expected that these patients will have good outcomes regardless of the modality chosen. Combined femoral and pelvic osteotomies can be performed when necessary to provide additional coverage for the femoral head (Fig. 6). For patients with disease onset after 10 years of age with little potential to remodel, or patients diagnosed in the reossification or healed stage, a Chiari osteotomy or shelf procedure may be performed for the sole purpose of providing coverage for the head. These patients, as expected, have a poor prognosis. Valgus osteotomy is also done as a salvage procedure in patients presenting with advanced disease. The goal is not to provide coverage, but to place an uninvolved area of the head under the weight-bearing portion of the acetabulum (Fig. 7(a),Fig. 7(b)). If successful, this procedure can relieve symptoms and delay further intervention for several years.



Fig. 6. Combined pelvic and femoral osteotomies provided a congruent joint in this patient.

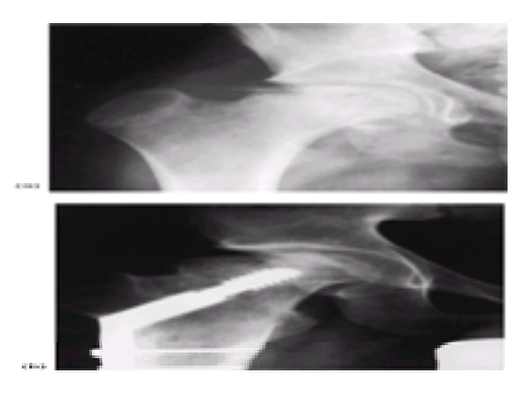


Fig. 7. (a) A 16-year-old patient who presented with advanced disease.(b) Valgus osteotomy performed as a salvage procedure in the same patient. The least involved portion of femoral head articular cartilage, as determined by arthrogram, is placed under the weight-bearing dome of the acetabulum. The patient experienced dramatic relief of pain and increased range of motion. Prognosis is guarded, however, due to the advanced stage of the disease at time of presentation.

Further reading

Catterall A. The natural history of Perthes disease. *Journal of Bone and Joint Surgery* 1971; **53B**: 37. [Describes Cattarall's original classification system and its use as a guide for treatment.]

Chung SMK. The arterial supply of the developing proximal end of the human femur. *Journal of Bone and Joint Surgery* 1976; **58A**: 961. [This landmark paper defined the complex blood supply of the proximal femur at various stages of development.]

Crutcher JP, Staheli LT. Combined osteotomy as a salvage procedure for severe Legg–Calvé–Perthes' disease. *Journal of Pediatric Orthopedics* 1992; **12**: 151. [This papers reviews the role of combined osteotomy as a salvage procedure in patients with severe disease.]

Fulford GE, Lunn PG, Macnicol MF. A prospective study of nonoperative and operative management for Perthes' disease. *Journal of Pediatric Orthopedics* 1993; **13**: 281.

Glueck CJ *et al.* Resistance to activated Protein C and Legg–Perthes' disease. *Clinical Orthopaedics and Related Research* 1997; **338**: 139. [This paper discusses the role of various clotting factor abnormalities in the pathogenesis of Legg–Perthes' disease.]

Henderson RC, Renner JB, Sturdivant MC. Evaluation of magnetic resonance imaging in Legg–Perthes disease: A prospective blinded study. *Journal of Pediatric Orthopedics* 1990; **10**: 289.

Herring JA. *Legg–Calvé–Perthes disease—American Academy of Orthopedic Surgeons' monograph series*. American Academy of Orthopedic Surgeons, Illinois, 1996.

Herring JA, Neustadt JB, Williams JJ, Early JS, Browne R. The lateral pillar classification of Legg–Calvé–Perthes' disease. *Journal of Pediatric Orthopedics* 1992; **12**: 143.

Hoikka V, Lindholm S, Poussa M. Intertrochanteric varus osteotomy in Legg–Calvé–Perthes disease: a report on 112 hips. *Journal of Pediatric Orthopedics* 1986; **6**: 600.

Martinez AG, Weinstein SL, Dietz FR. The weight bearing abduction brace for the treatment of Legg–Perthes disease. *Journal of Bone and Joint Surgery* 1992; **74A**: 12. [A review of 34 hips with severe disease treated with a weight-bearing abduction orthosis. At a mean follow-up of 7 years, no hip had a good result, 12 had a fair result, and 22 had a poor result, according to the criteria of Mose. Therefore, the authors recommend that these devices not be used in the treatment of severely involved hips.]

McAndrew MP, Weinstein SI. A long term follow-up of Legg–Calvé–Perthes disease. *Journal of Bone and Joint Surgery* 1984: **66A**: 860.

Meehan PL, Angel D, Nelson JM. The Scottish Rite abduction orthosis for the treatment of Legg–Perthes disease. *Journal of Bone and Joint Surgery* 1992; **74A**: 2.

Mose K. Methods of measuring in Legg–Calvé–Perthes disease with special regard to prognosis. *Clinical Orthopaedics and Related Research* 1980; **150**: 103. [Mose describes the use of a template of concentric circles to determine sphericity of the femoral head, a predictor of outcome.]

Salter RB. The scientific basis for inominate osteotomy for Legg–Calvé–Perthes disease and the results in children with a bad prognosis. *Orthopaedic Transactions* 1995; **9**: 203.

Salter RB, Thompson GH. Legg–Calvé–Perthes disease. The prognostic significance of the subchondral fracture and a two group classification of the femoral head involvement. *Journal of Bone and Joint Surgery* 1984; **66A**: 479.

Stulberg SD, Cooperman DR, Wallensten R. The natural history of Legg–Calvé–Perthes disease. *Journal of Bone and Joint Surgery* 1981; **63A**: 1095. [Describes a method of predicting outcome based upon sphericity and congruence of the femoral head.]

Wang L, Bowen R, Puniak M, Guille JT, Glutting J. An evaluation of various methods of treatment for Legg–Calvé–Perthes disease. *Clinical Orthopaedics and Related Research* 1995; **314**: 225.

Weinstein SL. Natural history and treatment outcomes of childhood hip disorders. *Clinical Orthopaedics and Related Research* 1997; **344**: 227. [An excellent overview of the childhood hip disorders, including Legg–Calvé–Perthes disease. Reviews pathogenesis, prognostic factors, and long-term follow-up studies.]

46.3.4 Slipped femoral capital epiphysis

Michael G. Ehrlich and Dennis C. Crawford

Aetiology
Clinical characteristics
Treatment
Further reading

Aetiology

Slippage of the capital femoral epiphysis tends to occur between the ages of 10 and 17 years, and is found 2.5 times more frequently in boys compared with girls. Overweight children are particularly at risk, and an association with increased height, and minor trauma is implied by many investigators. Black people are more commonly afflicted than Caucasions and Orientals, and in contrast to early reports appear to be at no increased risk for treatment complications.

Several factors associated with adolescence appear to predispose the growth plate to slippage. Some association with the endocrine alterations of adolescence are suggested but remain unproven. Clear correlation, however, does exists between endocrine disorders and slipped femoral capital epiphysis, particularly hypothyroidism and growth hormone deficiency. Patients with hypothyroidism tend to present prior to the teenage years, and those with growth hormone deficiency considerably later. The majority of patients with slipped capital epiphysis, however, have no evidence of endocrinopathy.

With adolescence, the physis or growth plate, undergoes characteristic changes which predispose it to alterations in anatomic alignment. With accelerated growth comes a reorientation of the growth plate from its horizontal position in young children to a more vertical angle relative to the femoral neck. This proceeds concomitantly with a thinning of the perichondral ring of LaCroix, which is believed to support the growth plate. Animal experimentation suggests slippage after blunt trauma when the perichondral ring is compromised in adolescent puppies. By contrast those with an intact ring suffered femoral neck fractures. The zone of slipping always occurs primarily thorough the zone of hypertrophy in a corrugated undulating fashion. Thus the final common pathway appears to be mechanical failure of the growth plate, due to a weakened state and subsequent inability to resist displacement.

Similar mechanical effects are found in patients with renal failure osteodystrophy. All patients have secondary hyperparathyroidism at the time of diagnosis of the slipped capital femoral epiphysis. Generally, these slips are stable and bilateral (>95 per cent), and occasionally associated with osteonecrosis when steroid immunosuppression is necessary. Under these circumstances, almost all weight-bearing joints can succumb to slippage, including the hip, distal femur, and ankle.

Clinical characteristics

Slips are generally defined as either acute or chronic, and more recently as stable or unstable. Historically, symptoms of pain for less than 3 weeks duration was considered an acute slip, and greater than this chronic. From a practical standpoint many slips will present with severe pain of short duration, which may simply be a manifestion of an acute change from an insidious chronic history. It is essential to elicit subtle signs of low-grade symptoms to differentiate this condition from an acute fracture and surgical emergency. Chronic slips generally present with a limp or a lurch to one side, which may be mild. This may or may not be accompanied by groin, anterior thigh, or knee pain. A stable slip is characterized by the ability to bear weight on the involved limb, regardless of the duration of symptoms. By contrast the unstable slip is too painful to allow weight-bearing with or without crutches. This new classification scheme is probably more prognostic than the traditional acute or chronic system; however, long-term studies are unavailable.

Physical examination reveals a characteristic loss of internal rotation, which manifests as an external rotation of the hip with hip flexion. This is essentially a pathognomonic sign. Typically, this finding will be accompanied by a limited range of hip abduction and flexion when compared with the uninvolved side. Radiographs in anteroposterior orientation are often negative, and a frog lateral projection is mandatory for diagnosis and grading (Fig. 1). Contralateral films are recommended as approximately 30 to 50 per cent of patients have been found to be effected bilaterally. About half of all hips are concurrently involved and the remainder become so within 18 months. This suggests the need for frequent follow up for a minimum of 2 years. Prophylactic pinning is generally not recommended, although some feel it reduces the risk of arthrosis and necessity of repeat radiographs.

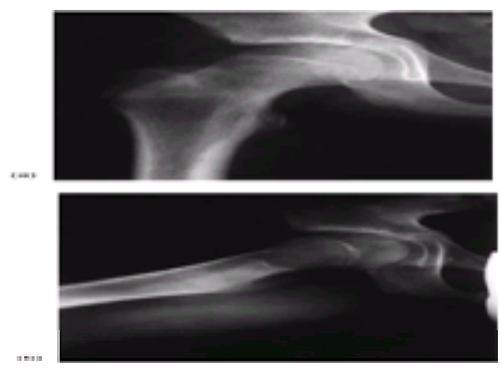


Fig. 1. Radiographs of mild slipped capital femoral epiphysis.(a) Anteroposterior view; (b) frog lateral view.

Generally the plates slip posteriorly and medially, not inferiorly as previously believed. Techniques for grading a slipped plate vary, but essentially revolve around the degree to which the femoral head has disengaged from its original position atop the femoral neck. One commonly used method compares the degree of displacement of the femoral epiphysis compared with the width of the femoral neck (Fig. 2). Another assesses the angle of the epiphysis relative to the femoral neck, and the difference to the opposite side (Fig. 3). The standard for classification of slips can be made from the frog lateral radiograph alone.

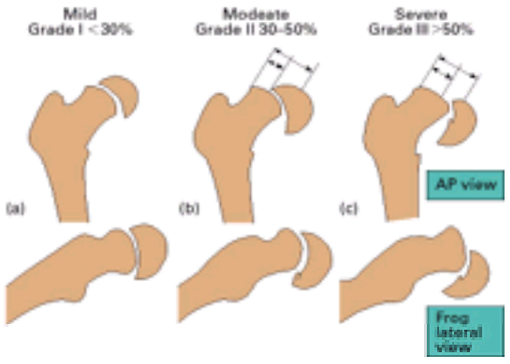


Fig. 2. Degree of slip severity using the method of Jacob. Classification relating the percentage of femoral neck uncovered by slipping physis.(a) *Mild*, less than 30 per cent; (b) *moderate*, between 30 and 50 per cent;(c) *severe*, greater than 50 per cent.

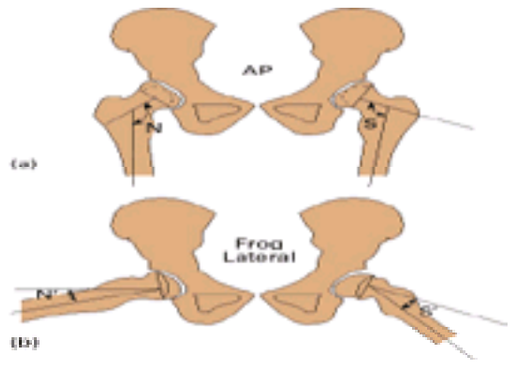


Fig. 3. Degree of slip severity using the method of Southwick. Classification by assessment of the angle of the epiphysis relative to the femoral neck, and the difference to the opposite side. (a) An anteroposterior view of the pelvis; N, normal hip angle at 150°; S, slip physis angle at 120°. (b) A frog lateral view of the pelvis; N1, normal hip angle at 10°; S1, slip physis angle at 50°.

Mild slips are characterized by slippage of less than 30 per cent or a head-shaft angle of less than 30 degrees difference from the uninvolved side; moderate cases from 30 to 50 per cent displacement or a head-shaft angle difference between 30 and 50 degrees; and severe slips are considered those with 50 per cent or more loss of the original overlap between the physis and epiphysis or a head-shaft angle difference greater than 50 degrees. Chronic slips may be seen to have radiographic evidence of callus formation or signs of remodeling such as ‘rounding off' at the superior head–neck junction.

Treatment

Management of the child with this condition is dependent on the extent, stability, and acuity of the slip. Primary goals are to avoid avascular necrosis and chondrolysis, stabilize the slipping process by creating a premature closure of the physis and correct deformity. In general chronic or stable slips are not reduced, whereas acute or unstable slips require reduction prior to fixation.

Mild slips are invariably treated with internal fixation. Some surgeons operate upon presentation, rather than after several hospital days in traction. We prefer to retrieve motion using traction. For chronic slips, this begins with bed rest in balanced suspension to relieve SYnovitis, intra-articular effusion, and improve range of motion. For acute or acute on chronic slips, traction reduction is felt to yield no increased risk of osteonecrosis. In the event motion is severely limited, chondrolysis should be suspected.

Historically, a single, non-threaded central pin was used to fixed most minimally displaced slipped femoral capital epiphysis. Slip recurrence via loss of fixation and continued growth at the epiphysis or displacement of the head altogether, led to abandonment of this method. Attempts at multiple pin fixation, however, were associated with significant iatrogenic chondrolysis and were similarly abandoned. Currently, single instrument central fixation using a cannulated (6 to 7 mm) partially threaded screw, via a percutaneous approach is considered standard of care (Fig. 4). By obtaining a true lateral and anteroposterior radiograph under fluoroscopic guidance, one avoids joint penetration while ensuring fixation across the physis. This technique has reduced inadvertent events and allowed adequate stabilization of most slips. Physeal closure can be expected in 12 to 14 months.



Fig. 4. Frog lateral radiograph showing a partially threaded single cannulated screw within the femoral head.

With acute unstable slips management with single central screw may be inadequate, although single screw fixation is increasingly proving to be safe and effective. We favor two threaded pins (Knowles type) inserted through a percutaneous approach based on the degree of slip. For moderate displacement, a lateral hip starting position is used, an anterior approach is preferred for severe posterior displacement. Subsequently patients are non-weight bearing until the painless range of motion has returned. Careful positioning on a fracture table, without any attempt at manipulative reduction precedes the internal fixation. Gentle reduction can be achieved preoperatively by placing the hip in skin traction. This allows muscle spasm associated with the acute event to subside. For acute slips traction is directed initially into flexion, then abduction is gently increased. Finally, an internal rotation force is applied. Operative repair can then be performed at about 3 weeks post-injury, when partial healing has occurred.

Alternative methods of treatment include spica cast application alone and core decompression with bone graft epiphysiodesis. Treatment in a spica alone does not create early closure of the growth plate and risks recurrence. Some authors recommend it in cases of acute on chronic or chronic slips. Open bone grafting provides for predictable physeal closure (12 to 16 weeks with autograft and 6 months with allograft), but requires a much more significant degree of intervention. Higher overall rates of complications when compared with single-screw fixation, however, have been demonstrated for both techniques. As such, these approaches are generally avoided in the United States.

The two principal complications associated with slipped femoral epiphyseal treatments are avascular necrosis and chondrolysis. Chondrolysis or acute cartilage necrosis, leads to pain, limping, decreased range of motion, and deformity. This may occur in either treated or untreated slipped femoral capital epiphysis. Its exact etiology is unknown, but it is thought to be associated with hardware penetration into the joint or immune reactions within the synovium. Avascular necrosis, similarly, may occur in the presence or absence of treatment. Impeded blood flow to the femoral head is common following aggressive reduction maneuvers in the stable hip, osteotomy, or with improperly placed fixation devices.

After treatment the majority of patients function satisfactorily without serious clinical complaints. Radiographic evidence of osteoarthritis tends to increase with the degree of slip severity, and the length of follow-up. Most patients, however, can expect improved range of hip motion and gait after treatment with satisfactory function into middle age unless chondrolysis or avascular necrosis occur.

Even when pinned *in situ*, severe slips can be expected to improve range of motion secondary to femoral head-neck remodeling. If patients have persistent functional problems, osteotomy should be considered. Re-alignment can be done at the level of the femoral neck and physis (cuneiform), the base of the neck, or the subtrochanteric level (Southwick). Cuneiform osteotomy is anatomically sound, but carries the highest risk of osteonecrosis and is contraindicated after growth plate closure. Basilar osteotomy produce less osteonecrosis, shortens the femoral neck, and generally produces favorable results. The Southwick procedure corrects in both varus and posterior planes with minimal risk of avascular necrosis.

In the event of severe avascular necrosis a fusion (extra- articular and intra-articular) is generally indicated. Fusion at 30 degrees of flexion with the hip in neutral rotation and abduction is adequate for boys, more abduction (5 to 10 degrees) is required for girls. Our approach is via a Smith-Peterson anterior incision, with anterior hip dislocation. Cartilage is removed from both the femoral head and acetabulum prior to rigid fixation of the femora to the pelvis. Initially, a cancelous screw fixes the head to the acetabulum. A large dynamic compression plate is then contoured across the joint from anterior iliac spine to the femoral neck and shaft (Fig. 5). Bone graft is then added about the denuded cortex of the femur to the greater trochanter. A spica cast is applied for about 3 months. Most adolescents can expect a significantly improved activity level if incapacitating pain and or an unacceptable fixed unilateral hip position characterize the preoperative condition.

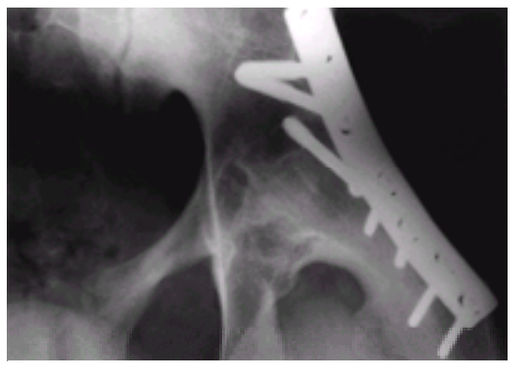


Fig. 5. Radiograph of a hip following fusion with internal fixation following avascular necrosis secondary to severe slipped capital femoral epiphysis.

Treatment of chondrolysis is difficult as the general prognosis and natural history are unclear. As an association with pin penetration of the joint and inflammation is suggested, we recommend removing all hardware and keeping patients non-weight-bearing in traction and on anti-inflammatory agents. Notably, it has been shown that transient violation of the joint during pinning is not associated with increased risk of chondrolysis. The prognosis for hip chondrolysis associated with slipped femoral capital epiphysis is better than for idiopathic chondrolysis, and treatment with the regimen described generally produces improvement.

Further reading

Abraham E *et al.* Treatment of moderate to severe slipped capital femoral epiphysis with extracapsular base of neck osteotomy. *Journal of Pediatric Orthopedics* 1993; **13**: 294–311. [Retrospective review of 36 hips with moderate to severe slipped femoral capital epiphysis, 90 per cent good to excellent results after basilar neck osteotomy at an average of 9 years, no osteonecrosis.]

Aronson D, Loder R. Slipped capital femoral epiphysis: A prospective study of fixation with a single screw. *Journal of Bone and Joint Surgery* 1992*a*; **74A**: 810–19. [Fifty-four of 58 hips treated by a single screw had excellent or good results at 3 years.]

Aronson DD, Loder BT. Slipped capital femoral epiphysis in Black children. *Journal of Pediatric Orthopedics* 1992*b*; **12**: 74–9. [Retrospective review of 74 blacks with 97 slips treated with single or multiple screw fixation found 91 per cent satisfactory results, 3 per cent osteonecrosis, no different than historic controls.]

Aronson DD, Loder RT. Treatment of the unstable (acute) slipped capital femoral epiphysis. *Clinical Orthopaedics and Related Research* 1996; **322**: 99–110. [Excellent review of the new classification and treatment methods for slipped femoral capital epiphysis.]

Betz RR *et al.* Treatment of slipped capital femoral epiphysis: Spica-cast immobilizaton. *Journal of Bone and Joint Surgery* 1990; **72A**: 587–600. [Retrospective review of 37 chronic slipped femoral capital epiphysis treated for 2 to 4 months in spica, with one progression, no osteonecrosis,14 per cent chondrolysis. Spica treatment recommended.]

Blanco JS *et al.* Comparison of single pin vs. multiple pin fixation in treatment of slipped capital epiphysis. *Journal of Pediatric Orthopedics* 1992; **12**: 384–9. [Retrospective review of 43 single pin versus 71 multiple pin fixations for slipped femoral capital epiphysis found no significant difference in time to physeal closure, more complications with multiple.]

Carney BT *et al.* Long-term follow-up of slipped capital femoral epiphysis. *Journal of Bone and Joint Surgery* 1991; **73A**: 667–74. [Retrospective review of 155 hips at 41years stated that pinning *in situ* gave the best long term results regardless of slip severity.]

Crawford AH. Role of osteotomy in the treatment of slipped capital femoral epiphysis. *Journal of Pediatric Orthopedics* 1996; **2**: 102–9. [Current status of re-alignment osteotomies is reviewed.]

Dietz FR. Traction reduction of acute and acute-on-chronic slipped capital femoral epiphysis. *Clinical Orthopaedics and Related Research* 1994; **302**: 101–10. [Retrospective review of 30 acute slips, 13 undergoing traction reduction found no increased rates of osteonecrosis compared with *in-situ* fixation without reduction.]

Fahey JJ, O'Brien ET. Acute slipped capital femoral epiphysis. Review of the literature and report of ten cases. *Journal of Bone and Joint Surgery* 1965; **47A**: 1105–27. [Original description chronic slip as SYmptoms more than 3 weeks.]

Fish JB. Cunieform osteotomy of the femoral neck in the treatment of slipped capital femoral epiphysis: Follow-up note. *Journal of Bone and Joint Surgery* 1994; **76A**: 46–59. [Three of 63 slipped femoral capital epiphysis greater than 30 degrees, treated with cuneiform osteotomy developed osteonecrosis.]

Hagglund G. The contralateral hip in slipped femoral epiphysis. *Journal of Pediatric Orthopedics* 1996; **3**: 158. Risk of contralateral slip is 40 to 80 per cent. [Twenty-five per cent of untreated slips develop arthrosis by age 50. Recommends prophylactic pinning in all cases.]

Irani R *et al.* Epiphyseodesis in slipped capital femoral epiphysis. A comparison of various surgical modalities. *Journal of Pediatric Orthopedics* 1985; **5**: 661. [Review of techniques for open management slipped femoral capital epiphysis.]

Jacobs B. Diagnosis and natural history of slipped femoral capital epiphysis. *Instructional Case Lecture* 1972; **15**: 167–73. [Classification of slip degree based on relation of radiographic displacement.]

Jerre R *et al.* The contralateral hip in patients primarily treated for unilateral slipped upper femoral epiphysis: Long-term follow-up of sixty one hips. *Journal of Bone and Joint Surgery* 1994 **76B**: 563–7. [Fourteen of 61 found with concurrent slip, 11 of 61 with slip at maturity, two of 25 SYmptomatic. Authors do not recommend prophylactic pinning.]

Kennedy JP, Weiner DS: Results of slipped capital femoral epiphysis in the Black population. *Journal of Pediatric Orthopedics* 1990; **10**: 224–7. [Retrospective review of 44 patients treated with pin fixation or bone graft with 93 per cent good or excellent results and no increased risk of complication.]

Loder RT, Hesinger RN. Slipped capital femoral epiphysis associated with renal failure osteodystrophy. *Journal of Pediatric Orthopedics* 1997; **2**: 205. [Retrospective review of 31 children with renal failure associated slipped femoral capital epiphysis.]

Loder RT *et al.* Acute slipped capital femoral epiphysis: The importance of physeal stability. *Journal of Bone and Joint Surgery* 1993; **75A**: 1134. [Introduction of stable versus unstable slip classification.]

Loder RT, Wittenberg B, Desilva G. Slipped capital femoral epiphysis associated with endocrine disorders. *Journal of Pediatric Orthopedics* 1995; **3**: 349–56. [Review of 85 patients with hypothyroidism (40 per cent), growth hormone deficiency (25 per cent). Hypothyroidism age of slipped femoral capital epiphysis less than 10 years, 61 per cent bilateral slipped femoral capital epiphysis with endocrine disease.]

Lubicky JP. Chondrolysis and avascular necrosis: complications of slipped femoral epiphysis. *Journal of Pediatric Orthopedics* 1996; **3**: 162. [Review of these complications of slipped femoral capital epiphysis.]

Meier MC *et al.* Treatment of slipped capital femoral epiphysis with a spica cast. *Journal of Bone and Joint Surgery* 1992; **74A**: 1522–9. [Seventeen hips treated with spica for 3 months, three had further slip, nine osteonecrosis. Authors recommend abandonment of this treatment method.]

Ross PM *et al.* Slipped capital femoral epiphysis long-term results after 10–38 years. *Clinical Orthopedics* 1979; **141**: 176–80. [Correlates the slip severity with degree of osteoarthritis and ultimate hip function.]

Schmidt TL *et al.* Allograft epiphysiodesis for slipped capital femoral epiphysis. *Clinical Orthopaedics and Related Research* 1996; **322**: 61–76. [Review of allograft epiphysiodesis techniques and comparison with autografting for achieving premature physeal closure (6 versus 4 months).]

Schoenecker PL *et al.* Intra-articular hip arthrodesis without subtrochanteric osteotomy in adolescents: technique and short-term follow-up. Am J Orthop 1997; **26**(4); 267–64. [Thirteen of 25 patients with avascular necrosis secondary to slipped femoral capital epiphysis or trauma where treated for incapacitating pain and stiff hips.]

Seigel DB *et al.* Slipped capital femoral epiphysis: A quantitative analysis of motion, gait and femoral remodeling after *in-situ* fixation. *Journal of Bone and Joint Surgery* 1991; **73A**: 659–66. [Retrospective review of 57 hips with *in situ* fixation followed for 2 years found significantly improved gait and hip motion.]

Southwick WO. Osteotomy through the lesser trochanter for slipped capital femoral epiphysis. *Journal of Bone and Joint Surgery* 1967; **49A**: 807–35. [Describes measurements for slip severity, osteotomy procedure.]

Warner WC *et al.* Chondrolysis after slipped femoral epiphysis. *Journal of Pediatric Orthopedics* 1996; **5**: 168–72. [Overview of the latest theories related to the etiology of chondrolysis.]

Weiner D. Pathogenesis of slipped capital femoral epiphysis: current concepts. *Journal of Pediatric Orthopedics* 1996; **5**: 67. [Review and discussion of the potential mechanism for slipped femoral capital epiphysis.]

Wong-Chung J, Strong ML. Physeal remodeling after internal fixation of slipped capital femoral epiphyses. *Journal of Pediatric Orthopedics* 1991; **11**: 2–5. [Change in shaft angle after remodeling in

100 per cent severe slips, and 5 per cent of mild slips pinned *in-situ*. Average remodeling is 12 degrees.]

Zionts LE *et al.* Transient penetration of the hip joint during *in situ* cannulated-screw fixation of slipped capital femoral epiphysis. *Journal of Bone and Joint Surgery* 1991; **73A**: 1054–60. [Thirty slipped femoral capital epiphysis treated by *in situ* fixation with 14 transient joint penetrations and no evidence of chondrolysis at 2 years radiographically.]

46.3.5 Congenital foot disorders

Michael G. Ehrlich and Justin K. Greisberg

Clubfoot (talipes equinovarus)

Definition
Aetiology
Diagnosis
Treatment

Flatfoot

Aetiology
Diagnosis
Treatment

Cavus foot

Definition
Aetiology
Diagnosis
Treatment

Further reading

Clubfoot (talipes equinovarus)

Definition

Clubfoot is a deformity in which the navicular bone and the forefoot are deviated medially relative to the talus (metatarsus adductus), and the calcaneus is locked directly under the talus, so that the heel is stuck in varus and equinus (Fig. 1). As the calcaneus is locked under the talus in equinovarus, it is unable to evert out from underneath the talus. This is the key feature of the deformity. Without unlocking the calcaneus from beneath the talus, the hindfoot cannot be brought into a neutral plantigrade position.



Fig. 1. The bilateral clubfeet of this infant display the characteristic findings. There is equinus, hindfoot varus, and forefoot adduction. It becomes obvious how uncorrected clubfeet can lead to problems with weight bearing.

There are generally considered to be four types of clubfeet. One is a positional deformity, related to the foot being held in an equinovarus position *in utero*. The second type, with which the rest of this section is concerned, is the anatomic clubfoot. Clubfeet may also be associated with muscle absence, such as arthrogryposis, or clear muscle imbalance, such as myelodysplasia.

Aetiology

The incidence in the general population is about 1 in 1000 births, and a first-degree relative of an affected individual has a risk 17 times as high. Some populations, such as Hawaiians and Polynesians, have a much higher incidence, at about 6 in 1000. The inheritance pattern of the anatomic clubfoot is probably multifactorial with a threshold effect, and is subject to environmental influences.

Muscle, bone, and nerve may be abnormal, either individually or collectively, in the anatomic clubfoot. Electromyography has shown peripheral neuropathy in some patients. Direct muscle biopsy, as well as cross-sectional imaging of musculature, has shown abnormality of extrinsic and/or intrinsic muscles, suggesting a primary muscle disturbance. There is a consistent deformity seen in the head of the talus, which is foreshortened and medially deviated. Despite these observations, the underlying etiology still eludes us.

Diagnosis

A clubfoot is obvious clinically, but it is important to determine which type of clubfoot is present. This affects not only the treatment course pursued, but also the expectations for long-term outcome. If the distinction is not made, surgical correction will often fail.

The diagnosis is suggested on manipulation of the foot. A positional clubfoot is supple, and the deformity is easy to correct passively. The arthrogrypotic foot is suggested by the presence of dimpling, indicating loss of subcutaneous fat, and contractures (Fig. 2). Attempts to stimulate muscle activity by tickling the bottom of the foot or stroking over various muscle groups fail. Radiographs of the leg show absent or diminished musculature in the soft tissue shadows. Dislocation of the hip occurs at a higher frequency in the clubfoot patient, but rigid dislocations or abnormalities at other joints may suggest arthrogryposis multiplex congenital, a generalized condition.



Fig. 2. In this infant with arthrogryposis, there is a clubfoot and a marked flexion contracture of the knee.

The deficiency in the anterior tibial muscle so commonly seen in the clubfoot does not by itself indicate a neurologic problem. The anatomic clubfoot cannot be

adequately dorsiflexed (as discussed below), and the anterior tibial muscle functions at a marked mechanical disadvantage. If the toe extensors and the peroneals are functional, the patient probably has no neurologic problem.

Treatment

The initial method of treating all clubfeet is casting. The foot is gently manipulated first, and then plastered. In the early days of clubfoot casting, the foot was manipulated into dorsiflexion before correction of the hindfoot varus component. This caused the forefoot to come up, but not the hindfoot, as the calcaneus is locked in equinus under the talus, causing a rocker-bottom deformity. The key to modern management, whether by casting or by surgery, is to recognize that the key feature of club foot is locking of the calcaneus under the talus. The subtalar joint has an oblique axis, so that the foot goes into dorsiflexion and eversion at the same time, and similarly into varus and inversion. For the heel to be dorsiflexed, the calcaneus cannot be locked under the talus, but must be able to swing out laterally. It follows then, that in casting, one must first evert the heel, and then bring it up and out into dorsiflexion.

The success of casting and manipulation can be assessed on radiographs taken at 3 months. Lateral views are usually done on a dorsiflexion frame, while the anteroposterior view is taken with the patient weight-bearing. If the foot is corrected, the calcaneus and talus cross on the lateral view. Lines drawn through the long axis of the bones form an angle of 30 to 50 degrees (Fig. 3). The anteroposterior view also shows the bones to be disengaged, with the talus pointing to the first metatarsal, forming an angle with the calcaneus of 20 to 40 degrees. The positional clubfoot will be corrected by this time. The anatomic clubfoot may be corrected by prolonged casting. However, ligaments are usually stronger than bone in children, and the prolonged force of casting can cause crushing or deformation of the foot. The bones of the hindfoot have circular growth plates, a fact that is often not appreciated because there is no secondary center of ossification around them. Prolonged casting can cause significant deformities, such as flat-topped tali, and the foot then becomes rigid. Also, clubfeet treated only with prolonged casting are frequently not ompletely corrected. The commonly held impression that clubfeet need bracing for many years is derived from the experience with incompletely corrected feet. Clubfeet treated properly do not need prolonged bracing. Only those with an arthrogrypotic or neurologic cause will need prolonged bracing.

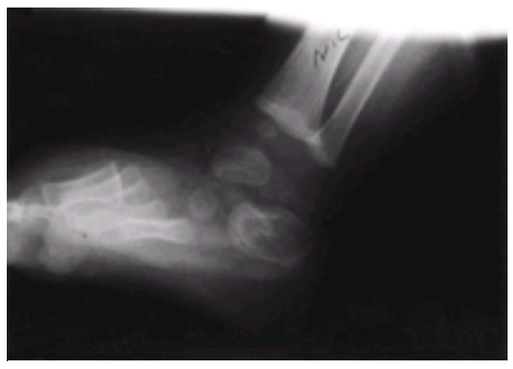


Fig. 3. This lateral radiograph shows a clubfoot treated with Achilles tendon lengthening, which went on to recur. In this postoperative film, the long axis of the talus is still almost parallel to that of the calcaneus (the axes are not drawn in), which indicates incomplete correction. In a corrected foot, and in a normal foot, the lateral talocalcaneal angle should be 30 to 50 degrees.

Casting is usually continued for 3 months, with weekly changes the first month, and then biweekly. Particular care is required in children with insensate feet, such as in myelodysplasia, as they will not offer complaint if the cast is exerting too much local pressure. As mentioned above, the positional clubfoot will be corrected by this time. For those feet which have persistent deformity, surgery is considered at some point within the first year of life. Some surgeons have suggested that operations delayed until 1 year of age have better results. As a temporary measure these surgeons may perform subcutaneous Achilles tenotomies. However, joints will form abnormal contours, and a substantial part of foot growth is finished by that time. We usually operate after 3 months if casting has failed, preferably at 5 or 6 months, when the child is a little bigger.

There are two general groups of surgical procedures. The first includes simple Achilles tendon lengthening, often with posterior ankle and subtalar capsulotomy. This addresses the obvious equinus deformity, but the foot fails to dorsiflex completely because the calcaneus remains locked under the talus. Feet treated this way require prolonged bracing and often recur.

The second group includes a more extensive posteromedial and lateral release. Several surgeons have described variants of this comprehensive release, and ours is described here. An incision is made from the first metatarsal to the dorsum of the talonavicular joint, down around the medial malleolus, dipping over the abductor hallucis, and then up along the interval between the Achilles and the posterior compartment. Some surgeons prefer the Cincinnati incision, which is a horizontal incision that runs medially and laterally around the foot. This incision permits exposure to the lateral ankle structures, but a wide medial exposure also allows access to the lateral ligaments through the subtalar joint.

In an anatomic clubfoot, the Achilles tendon, the posterior tibial tendon, and the tendons of the flexor hallucis longus and digitorum longus are Z-lengthened. (In the arthrogrypotic foot, the tendons are cut but not re-attached, to prevent recurrent deformity.) The posterior tibial neurovascular bundle is identified and followed into the plantar surface of the foot, where the overlying deep muscles and fascial band are incised. The talonavicular joint is freed, and the subtalar joint, including the posterior capsule, is released. One great area of controversy is whether the interosseus ligaments of the subtalar joint should be released. The objections to releasing them are based on the potential of obtaining a severe flatfoot through overcorrection. However, we believe complete release of all of the interosseus ligaments is advisable (Fig. 4). The calcaneus is then positioned under the talus, facing laterally, and with just a trace of valgus. The position is held with a pin through the calcaneus, the talus, and up into the tibia. The talonavicular joint is held reduced by a second pin (Fig. 5). This approach gives a mild flatfoot in our experience, with a trace of medial overhang, but there are no problems with recurrence, and the foot does not require bracing. After 3 months of cast immobilization, the pins are removed, and a program of strengthening is initiated. Some patients may require a University of California at Berkeley type arch support.



Fig. 4. During this comprehensive surgical release, the subtalar joint is entered to release contracted interosseus ligaments which prevent the calcaneus from coming out from underneath the talus.

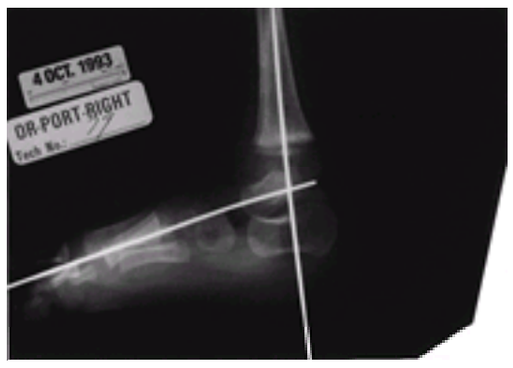


Fig. 5. The corrected clubfoot is temporarily held in position by two wires (and a cast). The long axis of the talus crosses the axis of the calcaneus (not drawn in), suggesting adequate correction.

In recent years, it has been noted that many clubfeet are left with a residual internal rotation deformity (Fig. 6). It had been thought that the rotation was secondary to an associated internal tibial torsion. In actuality, persistent internal rotation is due to the fact that the lateral border of the calcaneus is still bound by the contracted calcaneofibular ligaments posteriorly. This prevents rotation of the calcaneus during surgical correction. This deformity is corrected by releasing this ligament and rotating the calcaneus at the time of pin placement.



Fig. 6. The clubfoot is seen here to be internally rotated relative to the calf. This deformity is due to a contracted calcaneofibular ligament.

Treatment of the anatomic clubfoot never produces a perfectly normal foot. The calf is always smaller than normal and the foot is usually one or two shoe sizes smaller. Most patients do not have muscle deficiencies, and if the condition is well treated, it will not recur, nor does it require long-term bracing. The lack of uniform long-term assessment criteria has made it difficult to provide objective outcome data. Turco (1979) followed his patients with comprehensive posteromedial release, and in a study of 149 feet followed for 2 to 15 years, he found that 84 per cent had good to excellent results using a subjective grading system. Other authors have had similar experiences.

Advocates of limited release, Laaveg and Ponseti (1980) devised their own subjective outcome scale with numerical ratings, and found 74 per cent good to excellent long-term results. However, their patients typically had months of casting followed by years of bracing, with about one-half having relapses requiring further treatment. A more recent outcome analysis by Cooper and Dietz (1995) of patients with limited surgical release revealed 88 per cent good to excellent outcome by their own subjective assessment. One study has directly compared limited with comprehensive release, and (assessing outcome by the Ponseti scale) was unable to find any significant difference in long-term follow-up, although the comprehensive group had fewer operations overall. Of note, all outcome studies report moderate loss of hindfoot and ankle motion.

The Ilizarov technique of distraction histogenesis can be used to achieve a good correction. However, this method accepts the joint deformities and contractures, and is most commonly used for the older patient, often having failed open comprehensive release.

Patients presenting late with uncorrected clubfeet have generally undergone salvage procedures. In children over 8 years old, extensive releases give poor results, because the joints are already formed. In these cases, an attempt is made to accept the basic deformity, but to make the foot more plantigrade. Operations usually consist of metatarsal osteotomies and a calcaneal closing wedge osteotomy. Dwyer originally described the calcaneal osteotomy, and long-term follow-up of his patients show more than 80 per cent having a plantigrade foot with good hindfoot range of motion. (Comprehensive soft tissue releases can be performed on children up to 8 years of age, but often require some bone removal from the lateral side of the foot because of relative overgrowth.)

Uncorrected clubfoot in a mature individual requires a triple arthrodesis.

Flatfoot

The broad category of pediatric flatfoot includes a wide range of disorders with very different clinical presentations. The basic deformity is loss of the longitudinal arch with increased weight bearing on the medial side of the foot, usually with hindfoot eversion as well. The flexible (loose-ligamented) flatfoot is a normal variant, and usually is not symptomatic. In contrast, congenital vertical talus (congenital convex pes valgus) is a severe flatfoot deformity characterized by marked plantarflexion of the talus and dorsal dislocation of the navicular and forefoot relative to the hindfoot. A third disorder, tarsal coalition, may present without a flatfoot at all. Rather, there is a complete absence of hindfoot motion because of the coalition between bones of the hindfoot. A foot with an accessory navicular bone may also present as a painful flatfoot.

Finally, 97 per cent of infants have no longitudinal arch. The bony arch is present, but is filled in with fat.

Aetiology

The flexible flatfoot is probably one end of the spectrum of the normal distribution of arch height, and should not be thought of as a medical disorder unless painful. It is not related to any muscle weakness as once was thought. Electromyography has proven no muscles are active while maintaining the normal arch during stance. Pain occasionally arises from abnormal weight bearing on the medial foot, but studies have shown the vast majority of flexible flatfeet are asymptomatic.

Congenital vertical talus is bilateral in half the cases, and is often associated with central nervous system abnormalities such as myelomeningocele, or muscle imbalance as in arthrogryposis. It may be an isolated deformity or part of a genetic syndrome. The calcaneus and talus are both plantar-flexed, and the navicular rides on the dorsum of the talar head. Plantar ligaments which normally support the longitudinal arch are attenuated. Autosomal dominant inheritance with incomplete penetrance was demonstrated in a family, but this is probably not the rule. The underlying aetiology is not clear.

Tarsal coalitions occur in as much as 1 per cent of the normal population and are often asymptomatic. There is a slight male preponderance. The two most common coalitions occur between talus and calcaneus, and calcaneus and navicular, with the former the most common. About half are bilateral. They arise from failure of segmentation between hindfoot bones *in utero*, but do not become symptomatic until they start to ossify in adolescence. Inheritance may be autosomal dominant with variable penetrance.

An accessory navicular bone is present in 10 per cent of the population, and only occasionally becomes symptomatic. About 90 per cent of these form a stable relationship with the navicular. Only when the accessory bone remains unattached does it become SYmptomatic. Pain arises from a traction injury from the pull of the posterior tibial tendon across the fibrous bridge connecting the accessory bone to the navicular.

Diagnosis

The loose-ligamented (flexible) flatfoot is always bilateral. The feet appear normal when examined while the child is seated on the table. Hindfoot motion is normal, and the foot is supple with no rigid deformity. When the child bears weight, the arch collapses.

Tightness of the heel cord must be assessed when examining a flat foot, as a tight heel cord will pull the hindfoot into eversion (valgus) and cause a flatfoot deformity. With an equinnus contracture, the foot cannot become plantigrade during normal gait. The child can compensate by allowing the subtalar joint to evert. The oblique nature of the joint dorsiflexes the forefoot, so that the foot appears plantigrade. Over time, this stretches out the medial arch. To assess the heel cord, the knee must be extended (to stretch the gastrocnemius) and the hindfoot must be held inverted (varus) while dorsiflexing the ankle. If the hindfoot is allowed to evert, the oblique axis of the subtalar joint will move the forefoot into dorsiflexion, thus concealing tightness of the heel cord.

Several of the flatfoot disorders are characterized by a rigid hindfoot lacking any subtalar motion. The foot with congenital vertical talus will have no hindfoot motion and a prominent talar head medially. The talus and calcaneus are in equinus, while the navicular and forefoot are dorsiflexed and dislocated on to the dorsum of the talus. This combination of hindfoot plantarflexion and forefoot dorsiflexion has been termed congenital rocker-bottom foot and is characteristic (Fig. 7). Radiographs show the deformity, although the dislocated navicular does not ossify until the third year of life (Fig. 8). Magnetic resonance imaging can be used to locate the navicular when the diagnosis is uncertain. The equinus of the calcaneus seen on radiographs distinguishes congenital vertical talus from calcaneovalgus foot, which has a dorsiflexed hindfoot. (The calcaneovalgus foot is a benign condition noted at birth and occasionally confused with congenital vertical talus (Fig. 9).



Fig. 7. The characteristic deformity of congenital vertical talus is seen in this foot, with marked hindfoot plantarflexion and forefoot dorsiflexion. The bottom of the foot takes on a ‘rocker' appearance.



Fig. 8. In this lateral radiograph of a congenital vertical talus, the hindfoot plantarflexion deformity is seen.



Fig. 9. A calcaneovalgus foot has dorsiflexion of both hindfoot and forefoot, and lacks the rocker bottom.

The subtalar joint will become fixed in a valgus position when there is a joint effusion, as seen after trauma or with juvenile rheumatoid arthritis. The valgus position maximizes joint volume. These diagnoses must be kept in mind when evaluating a flatfoot with limited motion.

Lack of hindfoot motion is most characteristic of the foot with tarsal coalition. This foot was once called peroneal spastic flatfoot, although the peroneal muscles are rarely in spasm, and the foot is usually not very flat. Rather, the abnormal coalition between tarsal bones prevents hindfoot motion. Patients develop pain as the coalition begins to ossify in adolesence. The coalition itself may be bony, cartilaginous, or fibrous. Symptoms usually develop insidiously, and on examination is characterized by a rigid hindfoot. Attempts at passive inversion will elicit pain.

Clues to the diagnosis are visible on the lateral radiograph of the foot. The talus may appear foreshortened, with beaking on the dorsum of the talus or navicular. An oblique view may show the abnormal connection between bones in a calcaneocuboid coalition (Fig. 10). Talocalcaneal coalitions are best seen on multiple axial views of the calcaneus shot in angles of 30 to 45 degrees with the floor (Harris–Beath views). These views visualize the subtalar joint. The subtalar coalition is most commonly across the medial facet. Computed tomography scan is the most useful method for identify ing any tarsal coalition and, unlike plain radiographs, can accurately evaluate the size of the coalition (Fig. 11).



Fig. 10. A calcaneocuboid coalition is frequently cartilaginous, so that a direct osseus connection is not visible on an oblique radiograph, but the abnormal interface between the bones is visible.

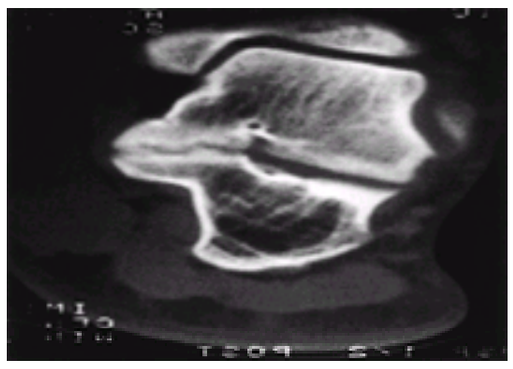


Fig. 11. Computed tomography scan is the best imaging tool to diagnose tarsal coalitions, and here reveals one in the medial portion of the subtalar joint.

The flatfoot caused by traction injury across an accessory navicular bone appears in adolescence. The accesssory bone is visible on plain radiographs (Fig. 12). The patient's pain will be replicated by resisting inversion of the foot. This foot will usually have a supple deformity, with good hindfoot motion.



Fig. 12. These bilateral accessor navicular bones are clearly seen on the anteroposterior radiograph as well-ossified bones distinct from the navicular and medial to it.

Treatment

The flexible flatfoot is probably a variant of normal, and treatment should be directed at education rather than intervention. In 1989, Gould and associates and Wenger and associates both reported results of randomized studies showing that children treated with arch supports had no more improvement in appearance of the arch than untreated flexible flatfeet. Thus, treatment of the painless flexible flatfoot, once other causes of flatfoot have been ruled out, is not warranted.

Very few children with a flexible flatfoot will develop pain. Pain rarely arises from abnormal weight bearing on the talar head. An arch support will allow for more normal weight distribution, although it may have no effect on the ultimate height of the arch. Surgical treatment is rarely, if ever, indicated.

If the Achilles tendon is tight, stretching exercises will be helpful early on. Tight heel cords are common among ‘toe walkers' (children who walk only on their toes). If stretching is not successful, lengthening of the Achilles tendon is done when the flatfoot remains flexible. As the children age, however, the deformity becomes rigid, and correction of deformity can only be achieved by changing the bony anatomy. A lateral opening wedge calcaneal osteotomy will lift the arch and bring the heel out of valgus. Alternatively, the ‘reverse Evans' procedure can be used, wherein the lateral column of the foot is lengthened to restore the medial arch.

We have also used arthoresis, by placing a silicone block in the sinus tarsi, almost into the subtalar joint (but not actually entering the joint). This acts as a ‘door stop', to limit eversion, and has replaced the Grice–Green or subtalar arthrodesis, as it preserves hindfoot motion and blocks the heel valgus that causes the deformity. We have reserved this procedure for pathologic conditions like spasticity and myelodysplasia, where a hindfoot fusion might be detrimental. Our 20-year experience with this has had no problems with silicone SYnovitis, perhaps because the joint space is not entered.

Congenital vertical talus needs surgical correction to regain a plantigrade foot. Early casting can stretch the soft tissues and may improve the appearance of the foot, but a foot with navicular dislocation will never be corrected by casting alone. Surgical correction is usually performed between 6 and 12 months through separate medial and lateral incisions. (A Cincinnatti incision is also acceptable.) Capsulotomy of the talonavicular joint is performed, and the navicular is reduced to the head of the talus. Other medial soft tissues, including interosseus ligaments of the subtalar joint, may need release. On the lateral side, ligaments of the sinus tarsi are released as well. Only when all of these structures are released will the calcaneus sit correctly under the talus. Lengthening of Achilles tendon, peroneals, and toe extensors is often performed as well. Any underlying muscle imbalance needs correction (through tendon transfers) to prevent recurrence. Reduction is maintained with wires across the subtalar and talonavicular joints.

Some surgeons use the posterior tibial tendon as a sling underneath the neck of the talus to support it, while a few actually excise the navicular and wedge it under the talus for support. Although we occasionally reef the posterior tibial tendon, we have not found excision of the navicular to be good policy. A Grice–Green extra-articular subtalar arthrodesis is added for the older patient with severe deformity. The foot is then immobilized in a cast for 3 months. Dodge and associates (1987) reported an average of 14-years of follow-up of their patients treated surgically with comprehensive soft tissue releases, some with subtalar arthrodesis or navicular excision. Eighty-three per cent had no pain at all and only a quarter required bracing or custom shoes. However, they note that many of the patients have marked functional limitations from associated abnormalities, including congenital hip dysplasia and arthrogryposis.

Recently, Seimon (1987) reported results of a less extensive procedure for congenital vertical talus. A limited dorsal and medial capsulotomy of the talonavicular joint is performed, and the navicular is reduced and held with a wire. Dorsal tendons are lengthened if necessary. The Achilles tendon is lengthened through a separate incision, and a second wire holds the subtalar joint reduced. He then immobilizes with a cast for 3 months in total, and no further bracing. He reported a limited patient population all operated on under 2 years of age, and found none needed further surgery and all could wear normal shoes in followup. He did note limited hindfoot range of motion, but no apparent problem clinically. Many surgeons now consider this limited release appropriate for the ‘average' congenital vertical talus. We have found this procedure excellent for the ‘vertically oriented' talus, but in those feet with a true dislocation of the navicular, recurrence is a problem when this limited release is performed.

The foot with a tarsal coalition usually presents with hindfoot pain. Initial treatment includes anti-inflammatory medication and a short leg cast for 6 weeks. Surgery is indicated for those patients whose symptoms persist and interfere with daily activity. For the patient with a calcaneonavicular bar, excision is considered only when there are no degenerative changes on X-ray or computed tomography scan. (Talar beaking alone is not a contraindication to excision.) A dorsolateral incision is made over the sinus tarsi. The origin of the extensor digitorum brevis is elevated, and the coalition is excised with an osteotome (Fig. 13). The origin of the extensor brevis is sutured into the site of the coalition to prevent recurrence. Short-term cast immobilization is used. Several studies have shown good long-term outcome, including that of Swiontkowski *et al.* (1993), which found good pain relief and return of hindfoot motion. Surgery failed in those patients with degenerative changes in the hindfoot.



Fig. 13. After resection of this calcaneocuboid coalition, a clear space can be seen between the bones.

Excision of the talocalcaneal bar can be equally successful. Surgical excision begins with a medial incision, over the sustentaculum tali. Capsulotomy of the subtalar joint is performed, and the coalition (most frequently in the medial facet) is removed with a power burr. Fat is interposed into the gap. Short-term cast immobilization is used. Wilde and associates (1994) evaluated patients' outcome by long-term pain and disability, and found at least half had no disability and minimal to no pain. Those patients with pain and disability had degenerative changes at the subtalar joint at the time of surgery. Scranton (1987) has reported that coalition of greater than 50 per cent of the talocalcaneal joint also leads to a poor outcome after excision. Another reason for poor outcome is failure to recognize a second coalition, as it is not uncommon to have more than one tarsal coalition in the same foot.

Patients who are not candidates for excision of a coalition should receive a triple arthrodesis (Fig. 14).



Fig. 14. A triple arthrodesis was used to treat this child with a subtalar coalition because degenerative changes were present.

Patients with a painful flatfoot from an accessory navicular can be similarly treated with cast immobilization and anti-inflammatory medication. The few who fail non-operative treatment may benefit from the Kidner procedure. The accessory bone is shelled out of the posterior tibial tendon, and the tendon is re-attached to the navicular through a drill hole. Cast immobilization for 6 weeks protects the repair. While this procedure reliably relieves pain, correction of flatfoot deformity is less predictable.

Cavus foot

Definition

Cavus feet, often referred to as high-arched feet, are a complex group of foot deformities. They are all characterized by equinus of the forefoot (plantarflexion of the forefoot relative to the hindfoot). The cavus foot is often associated with clawing of the toes, and is sometimes referred to as a claw foot. As weight bearing is focused on the heel and metatarsal heads, these patients can develop pain (metatarsalgia), callosities, and skin breakdown. The claw toes may rub on the dorsum of the shoe and create similar problems. Many of these patients have sensory neuropathies, which worsens skin breakdown.

Aetiology

The cavus foot is most frequently associated with an underlying neurologic problem. This may be central, secondary to spasticity associated with cerebral palsy or a space-occupying lesion. Patients with Friedrich's ataxia also characteristically develop a cavus foot, usually during the spastic phase. Lesions in the spinal cord are the most common cause of a cavus foot, including poliomyelitis, tethered cord, epidural lipoma, neoplasm, and myelodysplasia. Peripheral neuropathies, most typically Charcot–Marie–Tooth disease (hereditary sensorimotor neuropathy), may present with a cavus foot, as may isolated nerve injuries, perhaps following trauma to the peroneal nerve. Cavus foot may develop after a compartment syndrome of the foot, or may be the result of a poorly corrected club foot. A few patients have no neurologic or myopathic disorder. Regardless of cause, it is an imbalance of muscles in the lower extremity that causes the deformity.

The simplest pattern is isolated weakness of the anterior tibialis. Without this dorsiflexor, the forefoot drops into equinus. If the toe extensors are functional, they may overpull to compensate as the forefoot drops, and the toes are extended. This gives the cavus foot with claw toes. This classic cavus foot is seen in Charcot–Marie–Tooth disease (Fig. 15).



Fig. 15. This high-arched foot is the classic cavus foot. One can see how weight bearing is focused on the metatarsal heads and heel. Mild extension of the metatarsophalangeal joints is seen here, and eventually leads to the claw toe deformity.

Some patients start with an equinus foot from spasticity of the gastrocnemius and soleus. Following an Achilles tendon lengthening, the equinus of the hindfoot is corrected, but the stretched anterior tibialis muscle remains weak, producing a cavus foot. A more extreme variant is seen in the calcaneocavus foot, where the gastrocnemius and soleus are not functional at all. The hindfoot is fixed in dorsiflexion but as the patient walks on the foot, the forefoot may drop into equinus and the toe flexors become tight. This is seen in poliomyelitis and some myelodysplasias (Fig. 16).



Fig. 16. The calcaneocavus foot is characterized by a high arch with dorsiflexion of the hindfoot.

Diagnosis

The cavus foot may often be the first manifestation of an underlying disorder. Although the patient may present to the office to discuss problems with shoe wear, a thorough history and physical examination is essential to uncover the cause of the deformity. There are many good treatment options, but corrective surgery will fail if the deforming forces of persistent muscle imbalance are not addressed.

The history should begin at birth, where low birth weight, prematurity, or delayed milestones suggest cerebral palsy. Family history may reveal Charcot–Marie–Tooth disease or Friedrich's ataxia. Spinal cord diseases, such as tethering or lipoma, are often associated with bedwetting or incontinence. Prior residence in a foreign country suggests poliomyelitis. Determination of previous injuries can assess for nerve damage or compartment syndrome. Finally, inquiries about any previous foot surgery should be made.

A thorough neurologic exam is then performed, being careful to assess accurately sensation down to the lower sacral sensory roots. All lower extremity muscle groups are tested, and the spine is examined for any cutaneous abnormalities. The foot deformity itself can then be examined, first to assess for rigidity of the deformity. The position of the heel is noted, and the plantar fascia is examined for contracture. The block test described by Coleman and Chestnut (1977) is a helpful aid for the cavus foot with a varus heel. The patient is allowed to bear weight by placing the heel and the lateral metatarsal heads on the edge of a block. This allows the plantar-flexed first metatarsal of a cavus foot to drop below the plantar plane of the foot. If the hindfoot remains in a varus position, a fixed deformity is present, and this deformity must be addressed during surgical correction. If the hindfoot does not remain in varus as the first metatarsal drops beside the block, then the hindfoot is relatively flexible, and surgical correction may be limited to the forefoot.

Radiographs are an essential part of the evaluation (Fig. 17). A series of weight-bearing, non-weight-bearing, and active dorsiflexion lateral views will give an indication of the flexibility of the arch. Weight-bearing axial views shows the position of the heel (varus or valgus). Lumbar spine films, electromyography, or magnetic resonance imaging may be useful to evaluate further an underlying condition.



Fig. 17. The lateral radiograph of a cavus foot shows that the long axes of the first metatarsal and talus are divergent. In a normal foot, these axes would be parallel.

Treatment

The treatment strategy for a symptomatic cavus foot varies, depending on the nature of the muscle imbalance, age of the patient, and rigidity of the deformity. Many older children will have problems with shoe wear and foot pain from uneven loading of the foot during weight bearing. The metatarsal heads must be unloaded, and this can often be accomplished with a molded longitudinal arch insert, made of soft and flexible material. Other options include metatarsal pads or rocker bars. All of these orthotics have raised areas just proximal to the metatarsal head for weight bearing, while the area under the metatarsal head is actually unloaded.

Many patients will have a deformity that is too severe to be successfully treated with orthotics, especially in an insensate foot, and surgical correction is then an option. An immature and growing child with no fixed bony deformity needs soft tissue balancing, usually done with tendon transfers. A child with a mature bony deformity needs osteotomy or even fusion, possibly combined with soft tissue balancing. Any bony procedure performed without muscle balancing in a growing child will recur. Similarly, tendon transfers to balance muscle forces will not correct a fixed bony deformity. Furthermore, although fusion is useful for the older patient with a fixed deformity, many of these patients are insensate over the foot, and the skin will break down under repetitive cycles of weight bearing.

Tendon transfers are performed in the immature foot without fixed deformity. The specific transfers depend on the nature of the underlying imbalance. For the patient with anterior tibialis weakness and functional toe extensors, the long toe extensors are transferred to the metatarsal necks through drill holes. These transfers can be performed through two longitudinal dorsal incisions, one between the first and second metatarsal and the other over the fourth. Proper tensioning of the transferred tendons is essential, so that the tendons have adequate tension in the dorsiflexed position (but not very tight).

If the cavus foot has rigid deformity, osteotomies and plantar fascia release are performed 6 weeks before the tendon transfers. The plantar fascia is released through a medial incision, which starts at the medial malleolus and ends at the medial border of the plantar surface. The posterior tibial neurovascular bundle is identified, and all fascia and intrinsic muscles superficial to the bundle on the plantar surface are released. A first metatarsal osteotomy is performed for a fixed equinus of this ray, usually at the same time as the extensor tendon transfer, and through the same dorsal incision. Care must be taken to avoid the proximal metatarsal physis in the growing child. We use a ‘V' or dome osteotomy to correct the equinus of the metatarsal, and fix it with crossed wires. For the cavovarus foot (with the heel fixed in varus), a calcaneal closing wedge osteotomy is performed with the plantar fascia release. This requires a lateral incision, passing from the alcaneal tuberosity to the anterior process. The peroneal tendons and sural nerve are protected, and a closing wedge osteotomy is performed by removing a wedge of bone and bringing the heel to a neutral position.

Patients with more advanced Charcot–Marie–Tooth will have weakness of anterior tibialis, toe extensors, and peroneals. In these patients, the posterior tibialis can be transferred through the interosseus membrane to the dorsum of the foot (Fig. 18). Traditionally, the muscle is transferred to one of the cuneiforms, but we have found that the forefoot continues to drop. By lengthening the posterior tibialis with a graft of one-third of the Achilles tendon, we can attach the transferred muscle to the third metatarsal neck. If there remains a rigid cavus deformity, a midtarsal closing wedge osteotomy can be performed to flatten out the high arch. Although a hindfoot triple arthrodesis can be used to correct the deformity, the combination of tendon transfers and osteotomy provides a plantigrade foot that retains hindfoot motion.



Fig. 18. The posterior tibial tendon has been transferred through the interosseus membrane just above the ankle, and will be passed to the dorsum of the foot to act as a dorsiflexor.

The calcaneocavus foot requires a different strategy, as the anterior tibialis is usually functional. (As mentioned above, the deformity arises from weakness of the gastrocnemius and soleus.) If the posterior tibialis or peroneals are functional, they may be transferred to the calcaneal tuberosity. Unfortunately, these muscles are often weak in the calcaneocavus foot, and transferring them makes them weaker. Another option is to transfer the anterior tibialis through the interosseus membrane to the calcaneal tuberosity to act as a plantarflexor, combined with a transfer of long toe extensors to the metatarsal necks to act as foot dorsiflexors.

Clearly, there is no single formula for correcting a cavus foot. Treatment always requires recognition and correction of the underlying deformity(ies). As a final example, the foot injured by ischemia, such as after compartment syndrome, will have contracture of intrinsic muscles and plantar fascia. This foot needs simple release of contracted plantar tissues, not complicated tendon transfers, for correction.

There are no long-term outcome studies of the cavus foot. Experienced surgeons have found these techniques to provide a functional plantigrade foot in small published series.

Further reading

Clubfoot (talipes equinovarus)

Carroll NC. Pathoanatomy and surgical treatment of the resistant clubfoot. *American Association of Orthopaedic Surgeons Instructional Course Lectures* 1988; **37**: 93. [A good overall summary of surgical treatment of clubfoot.]

Cooper DM, Dietz FR. Treatment of idiopathic clubfoot. *Journal of Bone and Joint Surgery* 1995; **77-A**; 1477–89. [Long term follow-up of patients having limited posterior releases.]

Grant AD, Atar D, Lehman WB. The Ilizarov technique in correction of complex foot deformity. *Clinical Orthopaedics and Related Research* 1992; **280**: 94–103.

Haasbeek JF, Wright JG. A comparison of the long-term results of posterior and comprehensive release in the treatment of clubfoot. *Journal of Pediatric Orthopaedics* 1997; **17**: 29–35. [Compares the limited and comprehensive surgical release for clubfoot, with fair outcome data.]

Irani RN, Sherman MS. The pathological anatomy of clubfoot. *Journal of Bone and Joint Surgery* 1963; **45A**: 45–52.

Ippolito E, Ponseti IV. Congenital clubfoot of the human fetus: a histologic study. *Journal of Bone and Joint Surgery* 1980; **62A**: 8.

Kumar PN, Laing PW, Kienerman L. Medial calcaneal osteotomy for relapsed equinovarus deformity. Long term study of the results of Frederick Dwyer. *Journal of Bone and Joint Surgery* 1993; **75B**: 967–71. [Outcome data for Dwyer's closing wedge calcaneal osteotomy.]

Laaveg SJ, Ponseti IV. Long-term results of treatment of congenital clubfoot. *Journal of Bone and Joint Surgery* 1980; **62A**: 23–31. [Introduces the Ponseti scale for grading outcome, and applies the scale to feet with limited surgical treatment.]

Turco VJ. Resistant congenital clubfoot—one-stage posteromedial release with internal fixation. *Journal of Bone and Joint Surgery* 1979; **61A**: 805–14. [Outcome data for comprehensive surgical release of clubfoot.]

Vesely DG. A method of application of a clubfoot cast. *Clinics in Orthopaedics* 1972; **84**: 47.

Wynn-Davies R. Genetic and environmental factors in the etiology of talipes equinovarus. *Clinics in Orthopaedics* 1972; **84**: 9.

Flatfoot

Colton CL. The surgical management of congenital vertical talus. *Journal of Bone and Joint Surgery* 1973; **55B**: 566–74.

DeVriese L, Dereymaeker G, Molenaers G *et al.* Surgical treatment of tarsal coalitions. *Journal of Pediatric Orthopaedics* 1994; **3**: 96–101.

Dodge LD, Ashley RK, Gilbert RJ. Treatment of the congenital vertical talus: A retrospective review of 36 feet with long-term follow-up. *Foot and Ankle* 1987; **7**: 326–32. [This paper presents good long-term follow-up of comprehensive releases for congenital vertical talus.]

Gould N, Moreland M, Alvarez R *et al.* Development of the child's arch. *Foot and Ankle* 1989; **9**: 241–5.

McCormack TJ, Olney B, Asher M. Talocalcaneal coalition resection: a 10 year follow-up. *Journal of Pediatric Orthopaedics* 1997; **17**: 13–15. [Long-term follow-up showing success of resection of coalition.]

Scranton PE. Treatment of symptomatic talocalcaneal coalition. *Journal of Bone and Joint Surgery* 1987; **69A**: 533–8. [Identifies some reasons for failure of pain relief after resection.]

Seimon LP. Surgical correction of congenital vertical talus under the age of 2 years. *Journal of Pediatric Orthopaedics* 1987; **7**: 405–11. [Description and outcome data of limited surgical release for congenital vertical talus.]

Swiontkowski MF, Scranton PE, Hansen S. Tarsal coalitions: Long-term results of surgical treatment. *Journal of Pediatric Orthopaedics* 1983; **3**: 287–92. [Good long-term outcome data.]

Wenger Dr, Mauldin D, Speck G *et al.* Corrective shoes and inserts as treatment for flexible flatfoot in infants and children. *Journal of Bone and Joint Surgery* 1989; **71A**: 800–10. [One of the first papers to identify the benign course of flexible flatfoot.]

Wilde PH, Torode IP, Dickens DR *et al.* Resection for symptomatic talocalcaneal coalition. *Journal of Bone and Joint Surgery* 1994; **76B**: 797–801.

Cavus foot

Coleman SS, Chestnut WJ. A simple test for hindfoot flexibility in the cavovarus foot. *Clinical Orthopaedics* 1977; **123**: 60. [Describes Coleman's famous block test.]

Dekel S, Weissman SL. Osteotomy of the calcaneus and concomitant plantar stripping in children with talipes cavo-varus. *Journal of Bone and Joint Surgery* 1973; **55B**: 802.

Dwyer FC. Osteotomy of the calcaneum for pes cavus. *Journal of Bone and Joint Surgery* 1959; **41B**: 80. [Describes Dwyer's procedure for closing wedge osteotomy.]

Turner JW, Cooper RR. Posterior transposition of tibialis anterior through the interosseus membrane. *Clinical Orthopaedics* 1971; **79**: 71

46.3.6 Paediatric and adolescent spinal deformity

John B. Emans

Introduction
Normal spinal growth
Assessing spinal deformity
Principles of non-operative treatment
Principles of operative treatment
Idiopathic spinal deformity
Idiopathic scoliosis
Idiopathic kyphosis (Scheurmann's kyphosis)
Congenital spinal deformity
Congenital scoliosis and kyphosis
Other congenital spine deformities
Neuromuscular spinal deformity
Cerebral palsy
Myelodysplasia
Muscular dystrophy
Spondylolysis and spondylolisthesis
Spondylolysis
Spondylolisthesis
Further reading

Introduction

Spinal deformity can develop during embryogenesis (congenital scoliosis, kyphosis, myelodysplasia) or during growth in association with disordered development, neuromuscular disorders, or bony dysplasia. Table 1 lists the more notable causes of spinal deformity. Many spinal deformities worsen with growth, but some also spontaneously improve or stabilize. Although often asymptomatic during growth, spinal deformities may have profound implications for adulthood. Progressive spinal deformity may lead to severe cosmetic distortion, pain, difficulty with seating balance, respiratory failure from restrictive lung disease, and in some congenital deformities, spinal cord impingement. Understanding the natural history of untreated spinal deformity is crucial to choosing appropriate treatment, as deformities of differing etiologies and differing severity vary widely in their behavior during growth and adulthood as well as their response to treatment.

Congenital	
	Scoliosis
	Kyphosis
	Klippel–Feil deformity
	Sacral agenesis
	Segmental spinal dysgenesis
Idiopathic or developmental	
	Idiopathic scoliosis
	Scheuermann's kyphosis
	Spondylolysis
	Spondylolisthesis
Neuromuscular	
	Cerebral palsy
	Myelodysplasia
	Muscular dystrophy
	Spinal muscular atrophy
	Friedreich's ataxia
	Charcot–Marie–Tooth hereditary sensorimotor neuropathy
Bone dysplasias	
	Spondyloepiphyseal dysplasia
	Diastrophic dysplasia
	Osteogenesis imperfecta
Associated with:	
	Neurofibromatosis
Secondary to:	
	Laminectomy
	Thoracotomy
	Burns

Table 1 Partial list of childhood and adolescent spinal deformities

Spinal deformities are complex three-dimensional disorders (Fig. 1) with deformity in the coronal (scoliosis) and sagittal (kyphosis, lordosis) planes as well as axial rotational and sometimes translational deformity. Multidimensional classifications and measurement systems exist, but in clinical practice deformities are referenced in the coronal and sagittal planes and simple estimates of rotational deformity made. Scoliosis refers to curvature with the apex laterally, kyphosis posteriorly, and lordosis anteriorly. For a given curve the etiology, direction, and apical vertebra define the curve. A typical idiopathic scoliosis might be described as ‘right thoracic idiopathic lordoscoliosis with an apex at T5'.

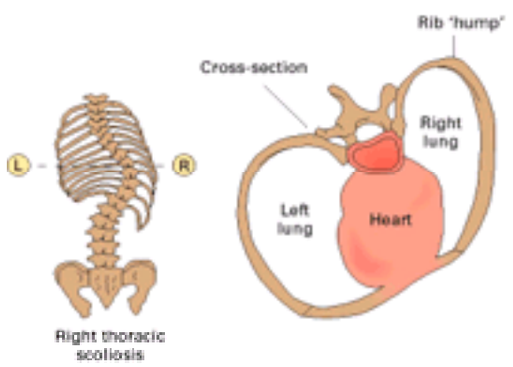


Fig. 1. Cross-section through a right thoracic scoliosis shows distortion of the ribs with resultant rib hump, and rotational deformity of the spine with the spinous processes deviated toward the concavity of the curve. Lung volume on the convex side of the curve diminishes progressively as the deformity worsens.

Normal spinal growth

Formation of the spinal column is an early and fundamental embryologic event. Somite and vertebral development are intimately related, hence visceral or neural malformations are commonly associated with congenital vertebral anomalies. Most vertebrae form from several discrete ossification centers. The cartilaginous neurocentral synchondroses joining early vertebral body ossification centers with those of the lamina are visible in infant spine radiographs and may be confusing. By the age of 5 years neurocentral SYnchondrosis closure is completed and the diameter of the spinal canal has reached nearly adult dimensions.

The normal adult spine is not perfectly straight, but exhibits a normal cervical lordosis, thoracic kyphosis, and lumbar lordosis (Fig. 2). The infant spine is a long, gradual, C-shaped kyphosis. Normal adult cervical lordosis, thoracic kyphosis, and lumbar lordosis develop in sequence as the infant progressively develops head control, sitting, and standing. Thoracic kyphosis increases through childhood to reach a normal adult range of 20 to 50°. Mild scoliosis of less than 10° in the adult is common and considered normal.

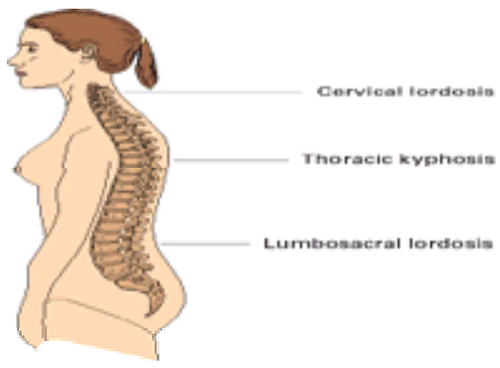


Fig. 2. The normal adult spine is not straight in the sagittal plane, but exhibits a normal cervical lordosis, thoracic kyphosis, and lumbar lordosis.

Assessing spinal deformity

Physical findings

Severe spinal deformity with distortion of the chest and waist is immediately apparent to the observer. Subtle signs of spinal deformity include asymmetry in neck or waist contour, shoulder height, scapula position and prominence, and apparent leg length difference (Fig. 3). Single curves are more visible, as are curves with large rotational deformities and rib ‘humps' (Fig. 4). Imbalance may occur and refers to a shift of the head or trunk laterally as compared with the pelvis.

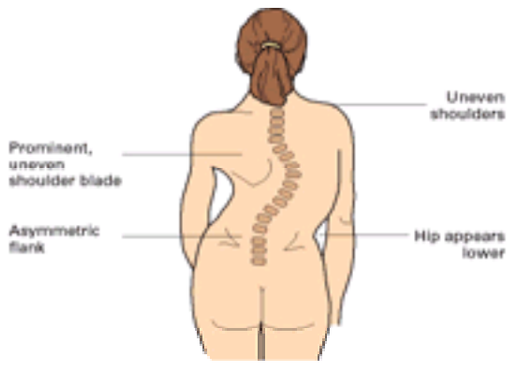


Fig. 3. Typical physical findings in severe idiopathic scoliosis include uneven shoulders, scapular prominence, asymmetric waist creases, and asymmetric hips.

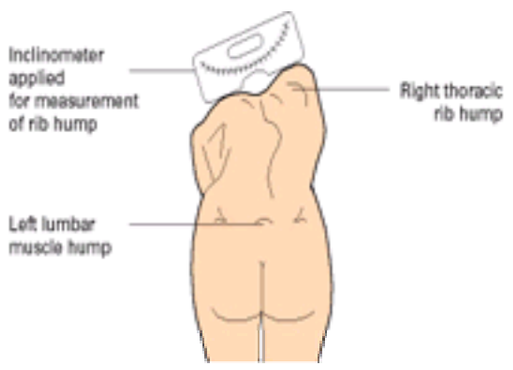


Fig. 4. The forward bend test reveals rib or lumbar muscle humps. An inclinometer can be used to quantify the asymmetry.

Physical examination for spinal deformity includes an assessment of asymmetry and a forward bend test in which the subject bends forward, with hands together, allowing the examiner to see along the back, searching for rib or lumbar humps to measure with an inclinometer (Fig. 4). The magnitude of the rotational asymmetry or hump usually reflects the severity of the curvature, but severe curves may have minimal rotational deformities and vice versa. Deviations from the normal cervical lordosis, thoracic kyphosis, and lumbar lordosis are noted. Leg length is assessed. A screening neurologic examination of reflexes and strength is completed. Signs of root tension such as pain on neck flexion or straight leg raising, tight hamstrings, foot deformity, atrophy of one leg, or associated hairy patches or dimples over the spine may be signs of an underlying spinal cord abnormality. Hip contractures may accompany long-standing severe scoliosis, particularly neuromuscular curves. Flexibility of a scoliosis is assessed by sideways bending, kyphosis and lordosis by flexion and extension.

Radiographic interpretation

Radiographs are an essential part of spinal deformity evaluation. Most simple idiopathic deformities need only standing posteroanterior views, but severe, non-idiopathic, or painful scoliosis and deviations in the sagittal plane (kyphosis, lordosis) also require lateral views. Preoperative planning may dictate bending views. CT scan with three-dimensional reformatting better defines complex deformities. MRI is useful for assessment of neuroanatomy or disc disease. Posteroanterior are generally preferred to anteroposterior views as the absorbed dose of radiation in growing breast and thyroid tissue is diminished with the former exposure. Full-length films, high-speed films, and special grids enhance visualization of the vertebrae.

Curve severity is measured using the Cobb angle in both the coronal and sagittal plane. A perfectly straight spine has a Cobb angle of 0°. The Cobb angle refers to the angle formed between the top of the uppermost tilted vertebra in a curve and the bottom of the lowermost tilted vertebra. Intra- and interobserver error with Cobb angle measurement can be large: 7 and 13° in idiopathic and congenital scoliosis, respectively.

Prediction of progression in scoliosis depends in part on the magnitude of growth remaining. The ‘Risser sign' refers to the extent of ossification of the iliac crest apophysis and is used as a guide to growth remaining when treating idiopathic scoliosis. A hand and wrist radiograph may be obtained for assessment of bone age.

Principles of non-operative treatment

Non-surgical treatment of spinal deformity during growth may have different goals, methods, and efficacy depending on the etiology, severity, or location of the deformity. Active exercises are used in postural disorders and as an adjunct to scoliosis or kyphosis bracing. Exercises may be directed at the spine itself or at extremity contractures such as hip flexion deformities or hamstring contractures. Exercises may consist of stretching or strengthening or postural awareness.

Spine braces or external orthoses such as the Milwaukee or Boston brace are used in progressive idiopathic scoliosis to halt worsening (progression) of the deformity. Except in milder kyphotic deformities and some idiopathic scoliosis it is less common to achieve substantial long-term correction of deformities through orthoses alone. Semi-rigid or total-contact external orthoses are used in neuromuscular scoliosis such as cerebral palsy to facilitate seating and function. Rigid lumbosacral orthoses are used in spondylolysis to relieve pain or encourage healing of a pars interarticularis stress fracture. Orthoses act on the spine via pressure applied to the skin, soft tissues, ribs, and abdomen and the corrective force transmitted to the spine is limited by the tolerance of these tissues. In spinal deformities where patient sensation or neuromuscular control is limited, external orthoses may be poorly tolerated.

Principles of operative treatment

Indications

Pain, continued worsening (progression), neurologic compromise, predicted pulmonary compromise, or severe cosmetic deformity may all be indications for operative treatment of a spinal deformity. Operative treatment may be performed for present problems in the growing spine or in anticipation of problems with the spinal deformity as an adult.

Spinal fusion

Spinal fusions are commonly performed. By achieving a solid fusion or arthrodesis the deformed section of spine is permanently stabilized and will not continue to deform with growth or aging. Choosing the extent of the fusion is critical. The length of spine fused is kept to a minimum to preserve as much normal spinal motion as possible, yet the fusion must include enough of the deformed spine to preclude future worsening of the deformity. The extent of fusion depends upon etiology, curve type, curve severity, and correction gained with instrumentation. Fusion is achieved by exposing and denuding (decorticating) the surfaces of adjacent vertebrae and applying bone graft. Instrumentation to stabilize and correct the deformity is often applied at the same time.

Fusions can be performed posteriorly, anteriorly, or both. Posterior fusions include exposure of the entire lamina and transverse processes, while anterior fusions involve complete intervertebral disc and endplate excision to expose the adjacent raw cancellous surfaces of vertebral bodies. The choice of posterior, anterior, or combined anterior and posterior fusion is determined by curve severity, growth remaining, and the likelihood of fusion. Anterior fusion in addition to posterior fusion may be mandated by a great deal of remaining growth or by severe deformities where excision of discs and associated soft tissue structures such as the costovertebral joints is required to enhance correction of the spine at the time of instrumentation. Absence of posterior bony elements associated with myelodysplasia or after laminectomy mandates anterior fusion and often anterior instrumentation. If the underlying condition makes it less likely that fusion will be achieved (conditions such as neurofibromatosis, myelomeningocele, or post-radiation deformity), then combined anterior and posterior exposure increases the surface area of the fusion mass and likelihood of fusion. The quality of the underlying bone, the closeness and surface area of adjacent vertebrae, rigidity of fixation, and source of bone graft all contribute to the likelihood of fusion. Autogenous iliac crest cancellous or corticocancellous strips are the most osteogenic bone graft available, but freeze-dried allograft bone is utilized when iliac autogenous bone is not practical. Recombinant bone morphogenic protein and other osteo-inductive substances hold promise for the future as bone graft substitutes.

Correction, immobilization, and instrumentation of spinal deformity

Spinal fusion can be performed without deformity correction (*in situ* fusion) but usually correction and immobilization are desired. Pre-or postoperative correction by casting or traction has some special applications, but generally instrumentation is used at the time of spinal fusion for both correction and immobilization. Harrington rod instrumentation, which initiated the modern era of spinal instrumentation, has been supplanted by ‘segmental' instrumentation and multihook contourable rods. The Harrington rod had two points of fixation, at either end of the spinal curve, while segmental instrumentation has multiple points of fixation either through multiple wires, hooks, or screws and often involves more than one rod. Segmental instrumentation produces greater correction and more rigid immobilization and usually eliminates the need for a postoperative brace or cast. Multiple hook types and wires allow fixation posteriorly to the lamina or pedicle or vertebral body via the pedicle, while screws permit fixation to the anterior vertebral body when used with anterior instrumentation or via the pedicle to the vertebral body when used posteriorly.

Sublaminar wires passed at every level are frequently used in paralytic conditions with soft bone. In most paralytic conditions when fusion to the sacrum is desired the instrumentation is placed in the ilium, rather than sacrum. A variety of fixation techniques to the ilium such as screws or rods placed between the inner and outer tables of the ilium (‘Galveston' fixation) allow indirect fixation of the sacrum.

The Harrington rod or other non-segmental instrumentation drew the deformed spine toward a straight rod. Modern segmental fixation is contourable and a combination of distraction, compression, and translatory forces along with rod contouring make it possible to achieve more normal thoracic kyphosis and lumbar lordosis.

Instrumentation without fusion is occasionally performed in the very young. By periodically lengthening the instrumentation, correction of the deformity can be maintained while allowing for some spine growth and awaiting definitive fusion at a later age.

Thoracoplasty with resection of the protuberant posterior ribs is performed for cosmetic improvement of severe deformities. By resecting a section or the proximal end of the rib and allowing the thoracic cage to shift to a more neutral position, an improved cosmetic appearance is achieved.

Spinal release, spinal fusion, and anterior instrumentation can be accomplished via video-assisted thoracoscopic surgery, diminishing the morbidity normally associated with thoracotomy.

Complications of spinal fusion

Spinal fusion eliminates motion and growth over the fused section of the spine. If a long fusion is performed in a very young child the loss of spinal height may be profound. However in those conditions needing early spinal fusion such as severe congenital scoliosis or kyphosis, loss of spine height from fusion is minimal compared with worsening of the deformity through growth. In most circumstances fusing the spine in a shorter, straighter position is preferable to allowing the evolution of a severe, uncorrectable deformity. An approximation of the growth remaining in a normal thoracic or lumbar spine is given by the formula 0.05 cm.(thoracic) or 0.07 cm. (lumbar)/segment per year of growth remaining. Fusing 8 thoracic vertebrae in a 10-year-old girl might result in 0.05 cm × 8 segments × 6 years = 2.4 cm lost, less than one might suspect. Inhibition of growth in diameter of the spinal canal is rarely a problem as the neurocentral synchondroses have finished growth early in childhood.

Fusion concentrates stress over the remaining mobile segments of spine. If the thoracic and most of the lumbar spine are fused leaving few motion segments in the lower lumbar spine, there can be degenerative changes, stress fractures, or spondylolysis below the fusion. Thus unwarranted extent of the fusion is avoided.

If a posterior fusion is performed in a young child with substantial growth remaining, persisting growth in the anterior vertebral bodies may cause the anterior spinal column to lengthen. With the length of the posterior column fixed and tethered by the fusion, the fusion mass may bend into a progressive lordosis or increasing scoliosis and rotation, a phenomenon referred to as the ‘crankshaft phenomenon'. To avoid this side-effect of persisting anterior spinal growth, both anterior and posterior fusions are performed when a great deal of growth remains.

Pseudoarthroses or failures to fuse are a common complication of spinal fusion. Pseudoarthroses occur more commonly in metabolically unfavorable situations, when bone surfaces are inadequate, and when immobilization is incomplete. Repair of the pseudoarthrosis with better bone graft and better fixation often results in a fusion.

In any spinal procedure, a neurologic injury may occur. The origins of neurologic injury associated with spinal instrumentation and fusion are varied. Direct intrusion into the spinal canal by instrumentation or severe translation of one vertebra relative to another can result in compromise of the canal. A change in position of an abnormal or scarred spinal cord such as in myelodysplasia or tethered spinal cord may be associated with neurologic loss. Elongation of the spinal canal may not be tolerated and traction on the vessels supplying the spinal cord or ligation of the segmental vessels associated with anterior exposure may account for some cases of neurologic loss. Neurologic loss can be immediate or delayed, reversible or irreversible, complete or incomplete. Typical precautions taken at the time of spinal instrumentation and fusion include neurophysiologic monitoring of somatosensory evoked potentials or in some cases motor evoked potentials. In co-operative patients the so-called Stagnara ‘wake up' test allows assessment of function under anesthesia, after instrumentation but before the procedure is completed. Postoperative neurologic loss is rare in idiopathic spinal deformities and common in some congenital deformities and bone dysplasias with spinal stenosis.

Idiopathic spinal deformity

Idiopathic scoliosis

Idiopathic scoliosis is the most commonly encountered spinal deformity requiring treatment.

Infantile and juvenile idiopathic scoliosis

Infantile idiopathic scoliosis, occurring under the age of 3 years, is rare and may spontaneously resolve. Juvenile idiopathic scoliosis, occurring in the age range 3 to 10

years, may be associated with an underlying neuropathic abnormality, particularly Chiari I malformation and associated syringomyelia. Juvenile idiopathic scoliosis severe enough to treat warrants a screening MRI of the intraspinal contents to screen for these abnormalities. Juvenile idiopathic scoliosis may respond well to bracing or may relentlessly progress to a more severe deformity needing fusion.

Adolescent idiopathic scoliosis

Etiology, incidence, and natural history

Although ‘idiopathic' in origin there appear to be multiple possible etiologies for adolescent idiopathic scoliosis. The genetics of adolescent idiopathic scoliosis are poorly defined but presumed to be variable in inheritance and expression. A neuropathic etiology is suggested by studies showing differences in vibratory sense, postural equilibrium, and brain stem asymmetries. Adolescent idiopathic scoliosis typically worsens during the rapid early pubertal growth spurt and disordered growth has been cited as an etiology. Most thoracic idiopathic scoliosis is associated with thoracic hypokyphosis or lordosis. The persistence of hypokyphosis late in childhood, particularly in girls, may be associated with the development of progressive adolescent idiopathic scoliosis. An underlying connective tissue or collagen disorder has been suggested by studies demonstrating platelet dysfunction in adolescent scoliosis. It seems unlikely that any one etiology will prevail as the sole cause of ‘idiopathic' scoliosis, rather adolescent idiopathic scoliosis with its typical curve patterns and worsening during the early pubertal growth phase may be a common expression of several subtle disorders.

The incidence of adolescent idiopathic scoliosis depends upon definition. Many normal individuals have a lateral curvature of 10° or less and a high percentage of school children screened for spinal asymmetries may demonstrate a mild scoliosis. Approximately 1 in 500 to 1 in 2000 adolescents may need brace or surgical treatment for idiopathic scoliosis. Adolescent idiopathic scoliosis is more common in girls, appears as a smaller curve in late childhood, and generally worsens with growth. Progression (worsening) of the curvature parallels growth rate but is variable in the individual. Milder curves under 20° often progress no further and even curves as large as 30° with substantial growth remaining may spontaneously stabilize. Most curves over 30° continue to progress with growth. Small curves at maturity usually remain stable through adulthood, while larger curves in excess of approximately 50° frequently continue to worsen slowly during adulthood. Adolescent idiopathic scoliosis is rarely SYmptomatic during adolescence but large progressive curves during adulthood can cause disfigurement and persistent pain, not easily treated by surgical or non-surgical means.

Observation and treatment

The screening, observation, and treatment of adolescent idiopathic scoliosis are directed toward preventing the adult sequelae of idiopathic scoliosis. School screening of 5th to 9th grade children is suggested to search for potentially progressive scoliosis in a phase when curves are still treatable by non-operative means. School screening has become controversial, has been abandoned in some countries, and has been the subject of a United States Public Health Service task force report on preventative services. Unnecessary referral of small curves for radiography can be reduced by use of the ‘inclinometer'. Inclinometer readings of less than 5° angle of trunk rotation on the forward bend test are usually less than 20° Cobb angle of curvature on radiography and do not need radiographic evaluation. If physical findings suggest non-idiopathic scoliosis or there is associated back pain, even smaller curves should be examined radiographically. The natural history of adolescent idiopathic scoliosis during growth is unpredictable in the individual with smaller curves. Follow-up observation of curves in growing children depends upon the magnitude of the curve and growth remaining. Frequent radiographic follow-up is appropriate for larger curves with growth remaining.

Based on natural history studies, brace treatment of idiopathic scoliosis during growth is indicated for juvenile curves of more than 20°, for progressive adolescent curves of more than 25°, and for curves between 30 and 45° with growth remaining. Spinal bracing is of limited effectiveness in curves of greater than 45°, curves high in the thoracic spine, or those associated with hypokyphosis.

The goal of non-operative treatment of adolescent idiopathic scoliosis is cessation of curve progression or improvement such that a fusion is not required. A wide variety of braces are available with different designs and varied degrees of documentation. The Milwaukee brace, which includes a metal superstructure, has the longest track record, followed by the Boston brace. Nighttime only ‘unbending' braces such as the Charleston brace have gained popularity. A recent meta-analysis suggests that the effectiveness of brace treatment of adolescent idiopathic scoliosis is dose related and that part-time braces may not be as effective as full-time braces. A typical ‘full-time' brace program involves wearing a brace for 20 out of 24 h with time out of the brace allowed for activities such as sport, bathing, and exercises. Bracing is continued until skeletal maturity. While wearing the brace there is substantial curve reduction and around 50 per cent reduction of a typical scoliosis curve should be sought in the initial brace application. In-brace curve correction is generally lost over time. After discontinuing the brace at maturity the average result of bracing is cessation of curve progression, not correction. About 30 per cent of patients experience substantial lasting correction, about 50 per cent are the same as at the initiation of bracing, and fully 10 to 20 per cent continue to experience curve progression in spite of bracing and may require spinal fusion. Non-compliance with requested brace wear is a major problem. The efficacy of bracing has been questioned. A prospective study undertaken by the Scoliosis Research Society demonstrates a large, statistically significant difference between brace treatment and observation alone. Electrical stimulation for scoliosis was shown to be statistically ineffective.

Operative treatment for adolescent idiopathic scoliosis is indicated for curves predicted to be troublesome as an adult or severe curves in the growing child uncontrolled by bracing. A curve magnitude of more than 50° is often cited as an indication for fusion, but this indication is balanced with curve appearance, progression after maturity, or pain as additional indications for fusion. Posterior fusion with multihook contourable rod instrumentation is standard (Fig. 5(a) and Fig. 5(b)), but anterior instrumentation and fusion may be preferable for single thoracolumbar or lumbar curves. Usually two-thirds correction of the curve and 50 per cent correction of the associated rib hump or asymmetry can be achieved. Typical risks include a 0.02 rate of pseudoarthrosis and 0.001 risk of paralysis. Autologous predonation of blood and intraoperative spinal cord monitoring are usually employed. The extent of fusion varies according to severity, type, and location of deformity.

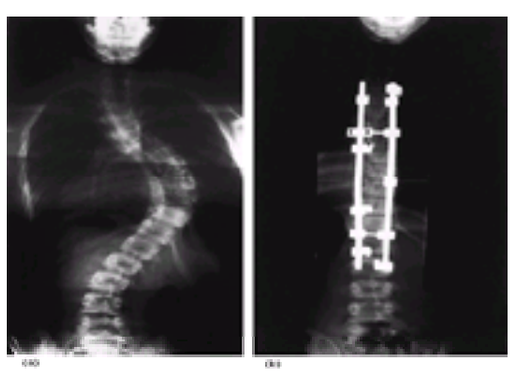


Fig. 5. (a) Shows a moderate, typical right thoracic adolescent idiopathic scoliosis of 60°. (b) Shows the same spine after posterior spine fusion and instrumentation with a segmental multihook contourable rod system.

Idiopathic kyphosis (Scheurmann's kyphosis)

Kyphosis is normal in the thoracic spine, with 20 to 45 or even 60° cited as normal thoracic kyphosis. Normal kyphosis is evenly distributed in the thoracic spine and centered in the mid-thoracic spine. An apex of kyphosis in the low thoracic spine or at the thoracolumbar junction or a magnitude of kyphosis in excess of 45 to 60° is considered abnormal. Idiopathic adolescent thoracic hyperkyphosis (Scheuermann's kyphosis) occurs more in males and shows typical vertebral endplate irregularity and wedging. Scheuermann's kyphosis is less flexible than normal thoracic kyphosis. On forward bending an accentuated hump is noted and on extension of the thoracic spine the kyphosis does not reverse as it would in a normal adolescent. The etiology of Scheuermann's kyphosis is unclear but is typically associated with tight hamstrings and may be associated with back pain.

The natural history of untreated Scheuermann's kyphosis in adulthood is not well documented. Generally, abnormal kyphoses at the thoracolumbar junction are felt to be associated with progressive deformity and pain during adulthood. Evenly distributed mid-thoracic hyperkyphosis may not be a major source of pain during adulthood but may be associated with pain in the lumbosacral area or cervical spine because of compensatory cervical or lumbar hyperlordosis. The incidence of L5 spondylolysis is higher in individuals with Scheuermann's thoracic hyperkyphosis than the general population.

Non-operative treatment for mild Scheuermann's kyphosis includes exercises with variable success. Braces, particularly Milwaukee braces, can be successful for moderate kyphosis between 50 and 70°, but the long-term effectiveness of kyphosis braces has been questioned.

Operative intervention for Scheuermann's kyphosis is generally anterior and posterior fusion although posterior fusion alone may suffice. The indications for prophylactic fusion are unclear but include severe pain, an unacceptable cosmetic deformity, or a severe deformity at the thoracolumbar junction or upper lumbar spine.

Congenital spinal deformity

Congenital spinal deformities originate from an early embryologic malformation and can be subdivided into failures of segmentation, failures of formation, or combinations. Other musculoskeletal, visceral, and neurologic anomalies frequently accompany vertebral anomalies. Patients with congenital spine deformities should have a screening renal ultrasound to rule out an asymptomatic renal collecting anomaly with obstruction. An MRI or other neural imaging study is needed prior to operation on congenital kyphosis or scoliosis or if there are signs of spinal cord tethering or neurologic abnormality. Intraspinal anomalies such as low lying conus, fatty filum terminale, tethered spinal cord, diastematomyelia, diplomyelia, syringomyelia, dermoid, and lipoma are not rare. VACTERL (Vertebral, Anal, Cardiac, Tracheo-Esophageal, Renal, Limb) association includes congenital vertebral anomalies. Sprengel's deformity with a high-riding scapula is commonly accompanied by upper thoracic or cervicothoracic vertebral anomalies.

Congenital scoliosis and kyphosis

Congenital scoliosis and kyphosis can be classified as complete or incomplete failures of segmentation or formation and many combinations thereof. The natural history of congenital scoliosis is variable. If growth potential is unbalanced with more growth potential on one side of the spine than the other, then the curve will progress with growth. If a tethering structure such as a rib fusion or congenital bony intervertebral bar is present, then progression may be rapid. Balanced growth potentials such as found in block vertebrae, balanced hemivertebrae, or fully incarcerated, non-segmented hemivertebrae may show no worsening with growth. Risk of progression is highest during the rapid growth periods of infancy and the first 2 years and again during the preadolescent growth spurt.

Progressive congenital scoliosis should be fused, as brace treatment is generally ineffective. For simple congenital scoliosis, *in situ* fusion without instrumentation is appropriate. For advanced deformity, hemivertebra excision or osteotomy of existing bony bars or areas of congenital fusion coupled with instrumented correction may be more appropriate. Hemivertebra excision is particularly effective at the lumbosacral junction or thoracolumbar junction in young children. Any correction of congenital spinal deformity carries an elevated risk of neurologic injury. Early *in situ* fusion of any progressive curve is generally the preferred treatment.

Congenital kyphosis nearly always worsens with growth. Type I congenital kyphosis with failure of formation may be associated with the early onset of paraplegia when progressive. Type II congenital kyphosis associated with failure of formation is less commonly associated with paraplegia but also worsens with growth. Congenital kyphosis less than 40° in children less than 5 years of age can generally be treated with a low-risk posterior *in situ* fusion and cast immobilization. Continued anterior spine growth after posterior fusion may provide further correction. More severe deformities or those with spinal canal compromise need anterior decompression, fusion, and posterior fusion and instrumentation. Surgery for congenital kyphotic deformities entails a high risk of neurologic injury. Traction for congenital kyphotic deformities is contraindicated, as longitudinal traction draws the already compromised spinal cord more tightly across the apex of the kyphosis. As for congenital scoliosis there is no effective non-operative treatment for congenital kyphosis and early *in situ* fusion is preferable where possible.

Other congenital spine deformities

Klippel–Feil deformity consists of multiple congenital fusions of the cervical and upper thoracic vertebrae. Typically there is an associated short neck and commonly associated Sprengel's deformity of the scapula and webbing of the neck. Because of the restricted cervical motion the remaining mobile cervical segments are subjected to extra stress and may become unstable. Protection from violent activities is often necessary to avoid neurologic injury in children and adults with Klippel–Feil deformity. Follow-up for the development of potentially dangerous cervical instability should continue through adulthood.

Partial or complete sacral or lumbosacral agenesis consists of a complete absence of the bony elements of the sacrum and lumbar vertebrae. Typically the pelvis is diminutive and as a result diverting colostomy or urinary drainage may be necessary. Lower extremity sensation is often intact but motor level paralysis reflects the level of absent vertebrae. Other congenital spinal anomalies include segmental spinal dysgenesis and Jarcho–Levin SYndrome with multiple vertebral anomalies and abnormal ribs. Many other congenital spinal abnormalities may occur singly or in connection with other anomalies

Neuromuscular spinal deformity

Spinal deformity commonly develops in association with neuromuscular conditions occurring during growth. Virtually every neuromuscular condition can be associated with spinal deformity. Abnormal muscle tone, strength, or balance combined with growth can produce profound deformities of the vertebral column. In idiopathic and congenital spinal deformity, treatment is generally directed toward preventing the long-term, adult consequences of deformity. Treatment of neuromuscular deformity emphasizes functional results. Bracing may be directed toward functional gain rather than cessation of curve progression. Spinal fusion will prevent further deformity but also may enhance use of the upper extremities or simplify seating management.

Cerebral palsy

Spinal deformity is common in patients with severe cerebral palsy who are non-ambulatory, but uncommon in the milder, ambulatory patient. Spinal deformity may profoundly affect the non-ambulatory patient, complicating seating and positioning, causing pain or skin break down, or in its most severe forms causing respiratory distress. There is a common and direct relationship between hip deformity and spinal deformity. Asymmetric hip contractures contribute to development of the spinal deformity and vice versa. Commonly a ‘windswept' deformity of the hips with unilateral or bilateral subluxation and an oblique pelvis is associated with the scoliosis. Non-operative treatment of milder deformities consists of positioning and stretching. In moderate deformities an external orthosis, such as a semirigid orthosis, will facilitate seating and function but is probably ineffective in controlling progressive curvature. The natural history of spinal deformity in cerebral palsy is variable. Skeletal maturity is often achieved later and curve progression may continue relentlessly through adulthood. When deformity becomes severe enough to be functionally troublesome in spite of bracing, spinal fusion is indicated. Anterior and posterior fusion with posterior segmental sublaminar wire fixation and dual rods (Fig. 6(a) and Fig. 6(b)), such as a Luque'-Galveston instrumentation, will withstand the abnormal muscle tone and permit early postoperative mobilization. Spinal fusion in cerebral palsy requires close attention to nutrition, as complication rates are directly correlated with preoperative nutritional status. Anterior and posterior surgery can generally be done on the same day, diminishing time in hospital and improving nutritional status.

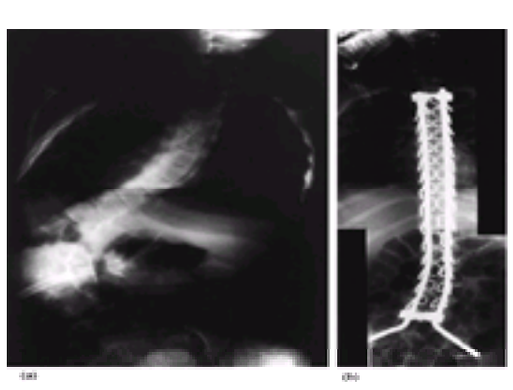


Fig. 6. (a) Shows a severe, collapsing type of curve with resultant seating difficulties and skin break down. (b) Shows the spine after anterior and posterior fusion and posterior instrumentation with Luque'-Galveston segmental sublaminar wire instrumentation. The L-shaped lower portion of the rods provides fixation to the pelvis, and thereby indirectly to the sacrum.

Myelodysplasia

Spinal deformity associated with myelodysplasia and other neural tube defects presents a wide spectrum of deformity. Some patients with asymmetric levels of innervation, a tethered spinal cord, or asymmetric hip contractures develop a long, collapsing deformity while others with associated congenital vertebral anomalies may develop a rigid, localized deformity. Non-surgical treatment of myelodysplasia-related spinal deformity is directed toward facilitating function. Adaptive seating, wheelchair modifications, external spinal orthoses, and exercises are designed to improve independence, mobility, and upper extremity function. When deformity

becomes large and functionally troublesome, fusion is usually indicated. Because the skin overlying most myelodysplasia spines is abnormal, preoperative tissue expanders or planning for local skin flaps is required. Because the accompanying neural elements are abnormal, simultaneous or concomitant spinal cord detethering may be required and because the posterior bony elements of the spine are usually deficient, combined anterior and posterior fusion is usually indicated. Instrumentation is variable depending on what bony elements are present. The functional effect of spinal fusion can be profound, helping the patient who regains the use of an upper extremity previously needed to maintain seating support, or adversely affecting ambulatory patients who lose trunk motion needed for ambulation. Risks of spinal fusion in myelodysplasia include a high incidence of infection, pseudarthrosis, and neurologic injury.

Muscular dystrophy

Nearly all patients with Duchenne's muscular dystrophy experience troubling spinal deformity during their progressive decline in function. As the patient with Duchenne's dystrophy ages and loses ambulatory ability, a spinal deformity usually appears and slowly progresses. Spinal deformity has been particularly troublesome for keeping patients seated comfortably during the non-ambulatory, progressively dependent phase of Duchenne's muscular dystrophy. Attempts to deal with the deformity through adaptive seating or bracing have been less successful than early spine fusion and instrumentation. By the time the spinal deformity is troublesome, respiratory and cardiac function have usually declined too much to tolerate the spinal fusion operation. Current practice in most clinics for Duchenne's muscular dystrophy is to offer spinal fusion at the time the patient either first ceases to walk or when pulmonary function declines to 50 per cent of predicted.

Spondylolysis and spondylolisthesis

Spondylolysis refers to a disruption in the continuity of the vertebral arch at the pars interarticularis (the portion of the vertebral arch joining the superior and inferior articular processes). Spondylolisthesis refers to the forward displacement of one vertebra upon another. Both occur in the lower lumbar spine with the greatest prevalence at L5 and both are developmental in origin.

Spondylolysis

Spondylolysis of L5 (or much less commonly L4) occurs in up to 5 per cent of the normal adult population, and is more common among relatives, certain ethnic groups, and in association with spina bifida occulta of L5 or S1. Spondylolysis is not present at birth, but first develops in the school age child. It is presumed that spondylolysis occurs by a mechanism of pars interarticularis stress, stress fracture, and finally complete fracture with spondylolysis. Fortunately most spondylolysis in adults is asymptomatic, the incidence of low back pain associated with spondylolysis not being different than the general population. Spondylolysis is a common source of pain in the older child and adolescent. Typical physical findings include lumbosacral pain with or without radiation to the buttocks, worse with sitting, back bending, and athletic activities. The incidence of spondylolysis seems higher in athletes involved in activities involving spinal hyperextension such as gymnastics. Plain radiographs may be negative, but spondylolysis may be seen on oblique radiographs and can be definitively diagnosed with a CT scan. If the spondylolysis is still in the ‘stress reaction' or stress fracture stages, before established lysis has occurred or if the lysis has occurred recently, bone scan will reveal increased uptake at the pars interarticularis.

Treatment for symptomatic spondylolysis or stress fracture of the pars interarticularis in the adolescent includes activity modification and temporary brace application. Many stress fractures with impending spondylolysis can be healed in this manner. Spondylolysis in the child and young adolescent may evolve into spondylolisthesis and should be followed periodically with repeat lateral radiographs to rule out progressive slippage. Surgery is rarely indicated in the child or adolescent with spondylolysis, as most can be rendered asymptomatic or minimally SYmptomatic with non-operative treatment. If required, surgery consists of an *in situ* fusion of L5 to S1.

Spondylolisthesis

Spondylolisthesis is less common than spondylolysis, but also is not present at birth and gradually develops during childhood and adolescence. Spondylolisthesis is classified according to etiology (see adult spinal deformity chapter), but in the child and adolescent two types are encountered. The ‘isthmic' type involves elongation or disruption of the pars interarticularis allowing the L5 vertebral body to displace progressively forward while the posterior elements of L5 remain in place. The ‘dysplastic' type involves an insufficiency of the posterior elements and facets between L5 and the sacrum, allowing the L5 vertebral body and its posterior elements to displace forward on the sacrum. Radiographic measurements include slip angle (a measure of kyphosis at the lumbosacral junction) and forward displacement of L5 forward on S1 (grade I to V, or 0 to 100 per cent). Physical findings are highly variable. In mild slips there may be no positive physical findings. More severe symptomatic slips are associated with a palpable step-off between the spinous rocesses of L4 and L5, a shortened waist, visible lumbosacral kyphosis with flattened buttocks, severely tight hamstrings, and a ‘jump' posture consisting of hip and knee flexion. SYmptoms include tight hamstrings and a variable degree of back and leg pain. The natural history of spondylolisthesis in the child and adolescent is variable. Progression of the slippage may be relentless during childhood and adolescence or may stabilize. Treatment for milder degrees of slip (less than 50 per cent displacement) is observation and non-operative treatment including bracing if SYmptomatic. Brace treatment is not felt to prevent further displacement. Treatment of more severe degrees of slippage depends upon growth remaining and symptoms. Asymptomatic severe slips in adults may be observed. Progressive slips in excess of 50 per cent and particularly those that are symptomatic in the adolescent are treated with L4 to the sacrum fusion with or without reduction, nerve root decompression, or instrumentation.

Further reading

Bridwell KH, Lenke LG, Baldus C, Blanke K. Major intraoperative neurologic deficits in pediatric and adult spinal deformity patients. Incidence and etiology at one institution. *Spine* 1998; **23**: 324–31. [Useful, modern data documenting the type and frequency of perioperative neurologic problems expected with a variety of spinal deformities.]

Dimeglio A. Growth of the spine before age 5 years. *Journal of Pediatric Orthopedics*, Part B, 1993; **1**: 102–7. [Important aspects of growth of the spine and the potential height lost by arthrodesis are well documented.]

Dubousset J, Herring JA, Shufflebarger HL. The crankshaft phenomenon. *Journal of Pediatric Orthopedics* 1989; **9**: 541–50. [The first detailed series describing curve progression associated with persistent anterior growth after posterior fusion in younger adolescents and children.]

Emans JB, Kaelin A, Bancel P, Hall JE, Miller ME. The Boston bracing system for idiopathic scoliosis. Follow-up results in 295 patients. *Spine* 1986; **11**: 792–801. [Typical results of Boston brace treatment of adolescent idiopathic scoliosis. Strong correlation noted between in-brace correction and result.]

Fredrickson BE, Baker D, McHolick WJ, Yuan HA, Lubicky JP. The natural history of spondylolysis and spondylolisthesis. *Journal of Bone and Joint Surgery* 1984; **A66**: 699–707. [A classic prevalence and natural history study of spondylolysis and spondylolisthesis.]

Lonstein JE, Carlson JM. The prediction of curve progression in untreated idiopathic scoliosis during growth. *Journal of Bone and Joint Surgery* 1984; **A66**: 1061–71. [Data describing the relationship between growth remaining, curve severity and curve progression. A useful nomogram for predicting curve progression is given.]

Lonstein JE, Winter RB. The Milwaukee brace for the treatment of adolescent idiopathic scoliosis. A review of one thousand and twenty patients. *Journal of Bone and Joint Surgery* 1994; **A76**: 1207–21. [Good documentation of the results of a large series of patients treated with the Milwaukee brace.]

McMaster MJ, David CV. Hemivertebra as a cause of scoliosis. A study of 104 patients. *Journal of Bone and Joint Surgery* 1986; **B68**: 588–95. [Documentation and risk factors for curve progression in congenital scoliosis in association with hemivertebrae.]

Nachemson AL, Peterson LE. Effectiveness of treatment with a brace in girls who have adolescent idiopathic scoliosis. A prospective, controlled study based on data from the Brace Study of the Scoliosis Research Society. *Journal of Bone and Joint Surgery* 1995; **A77**: 815–22. [This prospective study documents the statistical efficacy of brace treatment of adolescent idiopathic scoliosis.]

Rodgers WB, Frim DM, Emans JB. Surgery of the spine in myelodysplasia. An overview. *Clinical Orthopaedics and Related Research* 1997; **338**: 19–35. [A review of indications and surgical techniques in myelodysplasia-related spinal deformity.]

Rowe DE, Bernstein SM, Riddick MF, Adler F, Emans JB, Gardner-Bonneau D. A meta-analysis of the efficacy of non-operative treatments for idiopathic scoliosis. *Journal of Bone and Joint Surgery* 1997; **A79**: 664–74. [This meta-analysis suggests bracing for adolescent idiopathic scoliosis is dose-related in its effectiveness.]

Seitsalo S, Osterman K, Hyvarinen H, Schlenzka D, Poussa-M. Severe spondylolisthesis in children and adolescents. A long-term review of fusion *in situ*. *Journal of Bone and Joint Surgery* 1990; **B72**: 259–65. [*In situ* fusion, without reduction, is shown to give excellent long-term results even in severe grades of slip.]

United States Preventive Services Task Force. Screening for adolescent idiopathic scoliosis. Review article. *Journal of the American Medical Association* 1993; **269**: 2667–72. [A dispassionate review of the available data on the efficacy of school screening questions and the value of scoliosis screening.]

Weinstein SL, Zavala DC, Ponseti IV. Idiopathic scoliosis: long-term follow-up and prognosis in untreated patients. *Journal of Bone and Joint Surgery* 1981; **A63**: 702–12. [Excellent documentation of 194 of 219 untreated patients with idiopathic scoliosis followed through adulthood.]

Winter RB, Moe JH, Lonstein JE. Posterior spinal arthrodesis for congenital scoliosis. An analysis of the cases of two hundred and ninety patients, five to nineteen years old. *Journal of Bone and*

Joint Surgery 1984; **A66**: 1188–97. [The largest published series of congenital scoliosis emphasizes the utility of early fusion.]

Winter RB, Moe JH, Lonstein JE. The surgical treatment of congenital kyphosis. A review of 94 patients age 5 years or older, with 2 years or more follow-up in 77 patients. *Spine* 1985; **10**: 224–31. [A large series of this relatively rare condition stresses the efficacy of early fusion.]

46.4 Management of bone tumours

Dempsey S. Springfield and Henry J. Mankin

Introduction
Classification
Principles of diagnosis and evaluation
Principles and technique of biopsy
Surgical staging of bone tumors
Principles of surgical procedures
Primary bone tumors

Benign
Malignant

Carcinoma metastatic to bone
Myeloma
Further reading

Introduction

Bone tumors are relatively uncommon: only about 250 are diagnosed each year per 100 000 population and the majority of these are benign. The frequency of primary bone malignancies is very low, with only about 2000 new cases per year in the United States of America. Most tumors of bone, whether benign or malignant, can be easily recognized by their appearance on a plain radiograph, corroborated with advanced imaging studies if necessary. Many of the benign ones need no treatment or only observation.

The age at presentation varies widely but, by and large, bone tumors occur in younger patients than do carcinomas. Since the age of the patient may to some degree be a characteristic of the type of tumor, it becomes an important variable when determining a differential diagnosis. The majority of benign bone lesions and two of the major malignancies, osteosarcoma and Ewing's sarcoma, occur predominantly in individuals between 5 and 20 years of age. Giant-cell tumors almost always occur in individuals between the ages of 15 and 40 years. Chondrosarcomas, fibrosarcomas, myelomas, lymphomas, and metastatic disease all have a predilection for those older than 45 years. The patient's sex generally does not help in predictiing the diagnosis, but the giant-cell tumor of bone is more frequent in females while most of the other tumors, and especially osteosarcoma, are far more common in males.

There are some well-established genetic alterations associated with bone tumors. Ewing's sarcoma has a diagnostic stemline chromosomal translocation (between chromosomes 11 and 22); the *RB* and p53 genes are often missing or malfunctioning in patients with osteosarcoma. Hereditary multiple osteocartilaginous exostosis is an autosomal-dominant trait with a known genetic locus but the exact metabolic error that causes the lesion is unknown.

Of greater significance diagnostically is the increased incidence of bone sarcomas in those with hereditary multiple osteocartilaginous exostoses, Ollier's disease (enchondromatosis), Maffucci SYndrome (enchondromatosis with venous malformations), and in survivors of retinoblastoma. Osteosarcomas have also been reported in their siblings and with a greater frequency in their first cousins. Patients with bone infarcts, radiation osteitis, chronic osteomyelitis, and Paget's disease of bone have a higher incidence of bone sarcoma at the site of their local disease than that in the general population. Trauma has been implicated in the etiologic background of bony neoplasms but there is no real evidence that a single, uncomplicated injury can cause them. However, cytokines released at the time of fracture may be implicated in the increased progression of a pre-existing bone tumor.

Classification

The types of tumors occurring within bone reflect its components: osseous tissue, bone marrow, supporting connective tissues, nerves, blood vessels, and fat. The classification of bone tumors is based upon the predominant matrix component and type of cell differentiation within the lesion. Each type of tissue contained within bone may give rise to one or more clinically, radiographically, and histologically distinct, benign or malignant neoplastic lesions, each with its own pattern of biologic behavior (Table 1). It is of some importance, and indeed confusion, to those unfamiliar with these rare lesions that benign and malignant types of tumors occur for each component and may vary in their presentation and biologic behavior. In addition to the bone tumors that have a clearly associated normal tissue there are a few of unknown cell type. For example, the aneurysmal bone cyst is probably a neoplasm with an unknown tissue of origin, but some think it is a reactive process; giant-cell tumor may be of osteoclastic origin but this is really not fully established; and Ewing's sarcoma is probably of neuroectodermal origin but this is also uncertain.



Table 1 Classification of primary bone tumors

Most bone tumors are either locally slow growing and benign or cortically destructive and malignant, but some giant-cell tumors, osteoblastomas, and chondroblastomas, although benign, are locally aggressive and resemble more malignant tumors in that respect.

Principles of diagnosis and evaluation

Most bone tumors can be diagnosed with confidence from the clinical presentation (patient's age, complaints, and physical findings) and a plain radiograph, and the diagnosis then corroborated by special imaging studies such as bone scan, computed tomography (**CT**) or magnetic resonance imaging (**MRI**). Only rarely do blood tests help in the diagnosis and for the most part there are no laboratory markers for tumors of bone. Therefore, all patients suspected of having a bone tumor must have a careful history, a complete physical examination, and plain radiographs in their initial evaluation. After this information is available, further evaluation, if necessary, can be made. In the very young child, osteomyelitis or Langerhans cell histiocytosis are the lesions most likely to be mistaken for a bone tumor; in the adult, metastatic cancer is more likely. If the patient presents with a pathologic fracture it is important to decide if the underlying lesion is a primary bone tumor and if so if it is benign or malignant. When the prebiopsy evaluation cannot determine the nature of the lesion, the bone tumor should be assumed to be a primary malignancy until proved otherwise.

Most patients present either with no SYmptoms or with mild, dull pain. The lesion often turns up unexpectedly on a routine imaging study or bone scan. Those with severe pain, worse at night, should be suspected of having a malignant tumor, while those without pain are more likely to have a benign lesion, but otherwise the presenting symptoms are rarely of diagnostic value. An exception to this rule are the symptoms of an osteoid osteoma, a benign bone tumor, predominantly of childhood, which causes dull and aching pain that is more noticeable at night but almost always quickly and completely relieved by aspirin or a nonsteroidal anti-inflammatory medication.

On physical examination it is rare for a benign tumor, with the exception of osteocartilaginous exostosis, to have a palpable extraosseous mass. While giant-cell tumor of bone commonly has an extraosseous component seen with CT or MRI, it is usually not palpable. Osteosarcoma, Ewing's sarcoma, and chondrosarcoma, on the other hand, usually have a palpable mass, which is often tender to palpation. The malignant tumors may be associated with slight increased temperature of the extremity, swelling, and tenderness. Patients with bone tumors are rarely ill or complain of malaise, and it is uncommon for them to report a weight loss. These findings

suggest osteomyelitis or, occasionally, Ewing's sarcoma in the child, or metastatic disease in an adult.

Even though there are no serum or urine measures that are diagnostic of a specific bone tumor, it is important to obtain a complete blood count and erythrocyte sedimentation rate. These are useful in excluding myeloma, leukemia, and infection, diagnoses commonly in the differential. Determination of calcium and phosphorus may screen for metabolic bone disease. Alkaline phosphatase and lactate dehydrogenase are useful in the assessment of metabolic disease, and may be of prognostic significance in patients with osteosarcoma, lymphoma, or Ewing's sarcoma. Serum and urine immunoelectrophoresis should be obtained if multiple myeloma is suspected.

As mentioned earlier, a plain radiograph (in at least two planes) is essential. Often a specific diagnosis can be made from this study, and even if it cannot, the size and location of the tumor, its relation to the bone, whether there is bone formation or calcification, and the integrity of the cortex can all be seen. These features will help decide how much additional evaluation is needed and which additional tests are likely to be the most useful. A radionuclide bone scan is important, not only in assessing the activity of the lesion relative to its bone production and blood flow, but also in determining the presence or absence of bony lesions at other sites. A chest radiograph should be obtained in any patient with a suspected malignant bone tumor to search for metastatic disease or a primary focus from which a metastasis may have arisen.

If further evaluation is needed, MRI is the single, most common examination to be done. CT is superior to MRI only when looking for small lesions within dense bone (e.g. osteoid osteoma) or for subtle patterns of calcification. In almost every case an MRI or CT study should be obtained before a biopsy is done. Angiograms are rarely useful in the evaluation of bone tumors and are not done except under unusual circumstances.

If a lesion is thought to be metastatic carcinoma, a preliminary attempt to identify the site of the primary tumor is warranted. The most common primary carcinomas to metastasize to bone are breast, prostate, lung, kidney, and thyroid. Thus it is sensible, if metastatic malignancy is thought to be the cause of the bone lesion, to include mammography, chest CT, renal ultrasonography or abdominal CT, prostate-specific antigen, and a thyroid scan in addition to the history and physical examination. If a primary tumor is not found with these, further testing is unlikely to be fruitful.

When the tumor involves the spine, MRI should be used to examine the relation of the tumor to the spinal cord. When the canal is compromised, especially with compression of the cord, the patient may be in need of emergency care.

Principles and technique of biopsy

For most patients with tumors of bone, a biopsy specimen is necessary to establish the diagnosis. The biopsy should be deferred until many, if not all, of the staging studies are complete (see above). This practice avoids the possible interference of local tissue trauma with the imaging studies, and provides information that is useful in planning the location of the biopsy site. A closed (needle) biopsy offers the advantage of a small puncture wound that can be excised at the time of definitive resection. Needle biopsy is most appropriate in lesions of the spine or pelvis, and when the suspected diagnosis is metastasis, infection, a round-cell tumor, or a local recurrence. Accurate placement of the needle is important and should be determined by the surgeon, often in consultation with the radiologist at the time of a CT-directed biopsy.

Open biopsy is considered more reliable and is often preferred in the investigation of musculoskeletal tumors to ensure that sufficient tissue is obtained for accurate diagnosis and to avoid sampling error. Even when properly performed, an incisional biopsy has the disadvantages of creating a larger hematoma, the potential for soft-tissue contamination with tumor cells, and an increased chance of pathologic fracture if the cortex is violated. A good biopsy technique is vital; the procedure must be undertaken by an experienced surgeon, preferably the one who will ultimately carry out the definitive operation. The anatomic choice of the incision site for biopsy of a malignant tumor must take into consideration the extent of the lesion as determined by the staging studies and the possible plans for subsequent surgery. It must be possible to excise fully the tract of the biopsy and adjacent contaminated soft tissues at the time of a local resection or amputation. In almost all instances, transverse incisions in the extremities should be avoided, as should dissections that expose major neurovascular structures or enter an uninvolved anatomic compartment. Hemostasis is important following biopsy of malignant bony lesions: it may be wise to plug the bone ‘window' with polymethylmethacrylate (bone cement) to control hemorrhage and to contain the hematoma. Drains should be avoided or placed close to and in line with the biopsy incision, since drainage tracts become sites of secondary seeding. The use of a tourniquet is optional, but these should not replace the attainment of adequate hemostasis at the time of operation.

Handling of the biopsy specimen is also important. It is wise to obtain a frozen section to ensure that enough tissue has been obtained for diagnosis, and that some tissue for immunohistochemical stains and other special tests (such as cytogenetics) is available. Aerobic and anaerobic bacteriologic cultures should always be obtained, and a small portion of the tissue should be placed in glutaraldehyde for electron microscopy.

With the recent advent of limb-sparing procedures, the complexity of resectional surgery for tumors has increased. A recent study showed that major errors in diagnosis occurred in about 20 per cent of over 300 biopsies. In 5 per cent of patients, problems with the biopsy had led to unnecessary amputation; in 8.5 per cent the prognosis and outcome were thought to have been adversely affected.

Surgical staging of bone tumors

Imaging studies and biopsy provide information that enables a surgical stage for the lesion to be defined. The main variables (and, therefore, prognostic factors) in determining the outcome for bone tumors are the histologic grade (**G**), the anatomic location (**T**), and the presence or absence of distant metastases (**M**). The grade of the lesion is largely a histologic determination based on Broder's classification, but clinical factors such as symptoms, rate of growth, and radiologic characteristics are also taken into account. Certain lesions, such as juxtacortical osteosarcomas and adamantinomas, are almost always low grade (G1). Grading for the most part is subjective and is based on the pathologist's ability to assess cellularity, pleomorphism, and mitotic activity. It is also dependent on the established classification of tumors based on their light-microscopic appearances, which may not predict fully their biologic behavior.

The surgical site (T) refers to the location of the tumor in relation to anatomic compartments that serve as natural barriers to tumor extension. A lesion is deemed intracompartmental (T1) if it lies within an anatomic compartment such as the femur or tibia and is bounded on all sides by barriers to its extension, such as the bony cortex and the articular cartilage. Extracompartmental lesions (T2) are defined as tumors in which the natural planes are transgressed: the lesional tissue involves more than one anatomic compartment (e.g., an osteosarcoma of the femur that has broken through the cortex and presents with a large soft-tissue mass in the anterior compartment of the thigh). Some sites, such as the popliteal space, are by definition extracompartmental (T2) because the major neurovascular structures lie in interfascial tissues. A list of the various surgical compartments is shown in Table 2.

Intracompartmental (T1)	*Extracompartmental (T2)*
Intraosseous	Soft tissue extension
Intra-articular	Soft tissue extension
Superficial to deep fascia	Deep fascial extension
Paraosseous	Intraosseous or extrafascial
Intrafascial compartments:	Extrafascial planes or spaces:
Ray of hand or foot	Mid- and hindfoot
Posterior or anterolateral leg	Popliteal fossa
Anterior, medial, or posterior thigh	Groin—femoral triangle
Buttocks	Intrapelvic
Volar or dorsal forearm	Mid-hand
Anterior or posterior arm	Antecubital fossa
Pelvic spalsr	Axilla
	Periclavicular
	Paraspinal
	Head and neck

Adapted from: Enneking WF, Spanier SS, Goodman MA. A system for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopaedics* 1980; 153: 110.

Table 2 Surgical sites (T)

The final, major factor in determining the surgical stage of any lesion is the presence or absence of metastases (M). The lungs are the most common site of spread for sarcomas, but metastases to lymph nodes, other bones, or viscera carry a similar prognosis. The absence of metastatic disease, as determined by chest CT and bone scan, results in a designation of M0, while the presence of a focus of distant spread leads to a designation of M1.

Based on the assessment of histologic grade (G), site (T), and metastases (M), bone tumors are staged by the following system: all benign primary bone tumors are G0 and, therefore, stage 0 (regardless of T and, of course, M does not apply). Low-grade malignancies (G1) are stage I, which is further divided into intracompartmental (IA) and extracompartmental (IB) lesions. High-grade malignancies (G2) are classified as stage IIA and IIB, depending on the intra- or

extracompartmental location of the lesion, respectively. A lesion of any G or T with distant metastasis (M) is considered stage III (Table 3). This staging system was tested retrospectively in a group of 397 cases of bone and soft-tissue tumors; it was found that prognosis varied, with high statistical significance, between stage I and II and between IIA and IIB; no difference was noted between stage IA and IB lesions. This system appears to be an effective one that not only aids in assessing prognosis but also provides a basis upon which to plan surgical and other treatments.

Stage	*Grade*	*Site*	*Metastasis*
IA	Low (G1)	Intracompartmental (T1)	None (M0)
IB	Low (G1)	Extracompartmental (T2)	None (M0)
IIA	High (G2)	Intracompartmental (T1)	None (M0)
IIB	High (G2)	Extracompartmental (T2)	None (M0)
III	Low (G1)	Intracompartmental (T1)	Yes (M1)
		or	
	High (G2)	Extracompartmental (T2)	

Adapted from: Enneking WF, Spanier SS, Goodman MA. A system for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopaedics* 1980; 153: 111.

Table 3 Surgical stages

Principles of surgical procedures

Surgical resection or amputation is the primary method of obtaining local control of most bone tumors. The oncologic result of the surgical procedure is described in terms of the margins achieved; standard definitions and terminology are available for use in prospective analysis and for comparing results amongst different centers. The margin, defined as the volume of normal tissue surrounding the tumor, can be classified as intralesional (the procedure transects or enters the lesion), marginal (a plane through the reactive zone or ‘pseudocapsule' surrounding the tumor), wide (leaving a significant cuff of normal tissue around the tumor but not including the entire compartment in the resection), and radical (resection of the entire anatomic compartment containing the tumor). These definitions apply both to local resections and amputations (Table 4). Examples of intralesional procedures include incisional biopsy, curettage, debulking of an unresectable lesion, or an amputation through the tumor, all of which leave macroscopic tumor behind. Marginal procedures remove the bulk of tumor tissue, but are likely to leave behind satellite or daughter nodules and microscopic foci or residual tumor. Procedures that gain a wide margin carry the risk of leaving skip lesions in the surrounding normal tissue, as is occasionally seen with osteosarcoma, while with the radical margin, the entire compartment involved is removed and all local disease is presumably eradicated (Table 5).

Margin	*Surgical procedures*	
	Local	*Amputation*
Intralesional	Curettage or debulking	Debulking amputation
Marginal	Marginal excision	Marginal amputation
Wide	Wide local	Wide (through bone) amputation
Radical	Radical local resection	Radical disarticulation

Adapted from: Enneking WF, Spanier SS, Goodman MA. A system for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopaedics* 1980; 153: 111.

Table 4 Surgical procedures classified relative to the margins they attain

Type	*Plane of dissection*	*Result*
Intralesional	Piecemeal debulking or curettage	Leaves macroscopic disease satellites and ‘skip’ lesions
Marginal	Shell out *en bloc* through pseudocapsule or reactive zone	May leave either satellite or ‘skip’ lesions
Wide	Intracompartmental—*en bloc* with cuff of normal tissue	May leave ‘skip’ lesions
Radical	Intracompartmental—*en bloc* entire compartment	No local residual

Adapted from: Enneking WF, Spanier SS, Goodman MA. A system for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopaedics* 1980; 153: 111.

Table 5 Types of surgical margins and the implications relative to residual local disease

Primary bone tumors

Benign

Osteoid osteoma is a benign bony neoplasm of unknown cause that presents as a small, painful lesion with a characteristic clinical, radiographic, and pathologic pattern. Its incidence is highest in the second decade of life, but it is rare in those over the age of 30 years. It is most commonly located in the tibia and femur; in the spine it has a predilection for the posterior elements. The clinical presentation is with a sometimes sharp, boring pain, which is usually worse at night and is very frequently relieved by salicylates. Osteoid osteoma can cause a wide range of physical findings, including local tenderness, scoliosis, overgrowth of a bone, or secondary synovitis if located near a joint. Radiographically the lesion appears as a radiolucent, round or oval focus (the nidus), which may or may not be mineralized (Fig. 1). The nidus is the true lesional tissue and is usually 1 cm or less in diameter, but it is surrounded by a variable amount of reactive bone formation, which may be so extensive as to obscure the lesion. The diagnosis is apparent on radionuclide bone scans, CT, or plain tomograms. Although spontaneous regression may occur, this often takes several years and sclerosis may persist even after symptoms subside. Symptoms may be treated with salicylates or nonsteroidal anti-inflammatory agents, but most lesions are cured by complete excision of the nidus. A unique technique using a percutaneous radiofrequency needle that provides heat ablation has been developed and is under investigation.

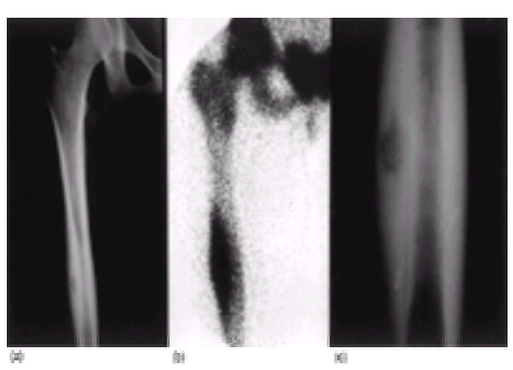


Fig. 1. Osteoid osteoma of the femoral diaphysis. (a) Anteroposterior radiograph showing the diaphyseal lucent nidus surrounded by reactive bone; (b) the lesion and new bone show increased uptake on [^{99}Tc]-MDP bone scan; (c) tomograms show the nidus, which is centrally mineralized, more clearly.

Osteoblastoma is histologically identical to osteoid osteoma. It has a predilection for the posterior elements of the spine (Fig. 2). It is usually not as painful as osteoid osteoma but does not regress spontaneously and requires surgical excision. A minority will behave very aggressively and may be difficult to control locally; the recurrence rate is relatively high with incomplete removal.

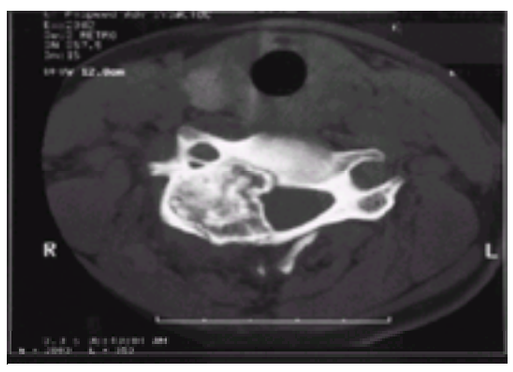


Fig. 2. Osteoblastoma of the cervical spine: on this axial CT scan the lesion is seen to arise from the right pedicle and compromise the canal; it should be removed surgically.

Benign cartilage tumors often present difficulties in differential diagnosis and management. It is often difficult to distinguish benign enchondromas from chondrosarcomas radiographically and even histologically The less common lesions, such as chondroblastoma and chondromyxoid fibroma, have high recurrence rates and are often difficult to approach surgically. Enchondroma, chondroblastoma, and chondromyxoid fibroma, Ollier's disease (multiple enchondro-matosis), and primary chondrosarcoma arise from within the medullary canal. Solitary osteochondroma, juxtacortical chondroma, hereditary multiple osteocartilaginous exostosis, and most secondary chondrosarcomas arise from the surface. Malignant cartilage lesions are unusual in patients below 25 years of age. Malignant tumors are generally larger than benign, are located more proximally (rare below the knees and elbows), and are often painful and tender. The likelihood of malignancy is higher in patients with multiple cartilage tumors.

Enchondroma is a benign cartilage lesion located centrally within the medullary cavity of short, long, or flat bones. It is the most common bone tumor of the hand. The lesions may occur at any age and are often asymptomatic unless accompanied by pathologic fracture. Radiographically, they are lucent and placed centrally within the metaphysis or diaphysis (Fig. 3(a)). Although thinning or scalloping of the cortex may occur, there is usually a moderate amount of sclerotic margination. Rounded or stippled calcification is a frequent finding in adults (Fig. 3(b)). Management by observation is sufficient if the lesions cause no symptoms, deformity, or risk of fracture. Otherwise, treatment is principally by curettage and packing the defect with bone autografts, allografts, or occasionally bone cement (polymethylmethacrylate). Multiple enchondromatosis (Ollier and Maffucci syndromes) is less common than its monostotic counterpart, and although not clearly genetic in origin, occurs with increased frequency in families. The severity ranges from minimal alterations to grotesque deformities of the skeleton. Treatment is aimed at correcting or preventing deformities. These patients (especially those with Maffucci syndrome) have an increased chance of malignancy, and they must be observed carefully throughout their lifetime.



Fig. 3. Enchondroma: (a) a centrally placed lytic lesion of the proximal phalanx of the thumb of this 29-year-old man shows thinning and internal scalloping of the cortex; (b) flocculent calcifications that are ring-like or amorphous stipples are also commonly noted, as in this 46-year-old patient with an enchondroma of the distal femur.

Chondroblastoma is most common in adolescent males, and almost always arises from the epiphysis of a long bone when the physes are open. The proximal humerus is the most common site, with the proximal and distal femur, and proximal tibia, as additional sites. Radiographically, the appearance is that of a round to oval, lucent, epiphyseal lesion, characteristically stippled with punctate calcifications (Fig. 4). Treatment is by curettage, avoiding the adjacent growth plate and articular cartilage. The recurrence rate following intralesional treatment is about 25 per cent. In extremely rare circumstances, benign chondroblastomas may spread to the lungs.

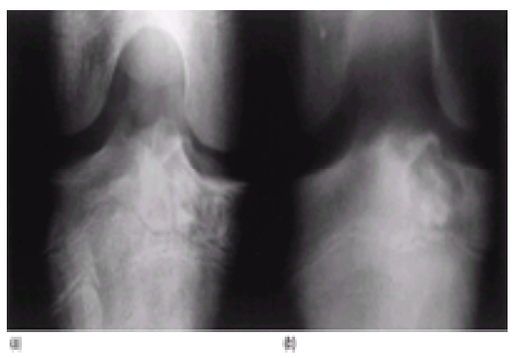


Fig. 4. (a) A 14-year-old boy with a painful chondroblastoma of the medial, proximal tibial epiphysis; (b) a tomogram shows the rounded lesion more clearly, demonstrating the stippled calcification and proximity of the lesion to the growth plate and articular cartilage.

Solitary osteocartilaginous exostosis is a frequent tumor of bone and is probably due to a genetic defect in the embryonic cartilage anlage, or a defect in the restraining periosteum and the ring of Ranvier. The condition occurs at the metaphyseal ends of bones at any age. Growth of these lesions occurs by enchondral ossification and, just as with the normal growth plate, ceases at maturity. Radiographically, exostoses appear as sessile or pedunculated outgrowths of the cortex of the metaphysis, usually projecting away from the joint (Fig. 5). The cortices of the bone of origin and the lesion are continuous and their marrow cavities communicate. The pathologic tissue of the osteocartilaginous exostosis is the cartilage cap, which is responsible for growth by an enchondral sequence not dissimilar from that of a normal growth plate. These lesions usually are not troublesome to the patient unless they cause a painful bursa, undergo a fracture, or interfere with adjacent neurovascular

structures. Malignant degeneration can occur, but does so only after the age of 30 years and is distinctly uncommon, with an incidence of less than 0.1 per cent. Removal is usually for relief of symptoms. Recurrence is unlikely if the cartilage cap is completely excised. Hereditary multiple osteocartilaginous exostosis is inherited as an autosomal-dominant trait; it is characterized by multiple exostotic lesions (Fig. 6) that cause distortion of the skeleton and may lead to severe functional impairment. The main concerns are the management of these deformities in childhood and their monitoring for the possible development of malignancy in adulthood. Malignant degeneration is more common in patients with multiple than solitary lesions.



Fig. 5. (a) A 14-year-old boy with a typical pedunculated osteochondroma of the distal femur first noted after a fall. (b) In another patient a sessile osteochondroma is seen in the distal femoral metaphysis; (c) On MRI the continuity of the cortex of the femur and the osteochondroma is seen.



Fig. 6. Radiographs of the arm (a) and forearm (b) in a 20-year-old man with hereditary osteocartilaginous exostosis, demonstrating the distortion of the bones, abnormal remodeling of the metaphysis, shortening of the ulna, and multiple exostotic lesions.

Fibrous cortical defect and non-ossifying fibroma are the most prevalent of the benign lesions within the bone in childhood, occurring in up to one-third of the population. These lesions, which may not be true neoplasms, occur in the metaphyses of the long bones, especially the distal femur and proximal and distal tibia, but may also develop in the upper extremities (proximal humerus, distal radius, and ulna). Although occasionally persisting into adulthood, most lesions disappear soon after late adolescence and generally produce no symptoms unless a fracture occurs. Their radiographic appearance is usually sufficiently characteristic to make biopsy unnecessary (Fig. 7). No treatment other than observation is required unless the lesion is in danger of causing pathologic fracture or the diagnosis is in doubt, when biopsy, curettage, and bone grafting are indicated.



Fig. 7. Fibrous cortical defect (non-ossifying fibroma) in a skeletally mature patient. The lesion is lytic, trabeculated, and eccentrically placed in the metaphysis of the tibia; its inferior aspect is beginning to fill in with osseous tissue.

Fibrous dysplasia is a disorder of fibro-osseous tissue that probably represents a developmental abnormality rather than a true neoplasm. Three major clinical syndromes exist: monostotic fibrous dysplasia, polyostotic fibrous dysplasia, and McCune–Albright syndrome. In McCune–Albright syndrome, polyostotic fibrous dysplasia is associated with areas of cutaneous pigmentation and an endocrine dysfunction. Precocious puberty is the most common endocrinopathy, but acromegaly, vitamin D-resistant rickets, hyperthyroidism, and Cushing syndrome may be present. A far less common entity is Mazebraud's disease, in which the fibrous dysplasia is associated with soft-tissue myxomas. Radiographically, the bony lesions of fibrous dysplasia have a lucent or ‘ground glass' appearance, and cause thinning of the cortex and endosteal scalloping (Fig. 8). The bone may be enlarged or deformed and its whole length may be involved. Treatment is mainly directed toward prevention or correction of bony deformities, which constitutes one of the most difficult technical challenges in orthopedics, particularly for lesions about the hip (the ‘shepherd crook deformity'). Biopsy of a lesion of fibrous dysplasia is occasionally necessary to distinguish it from more aggressive lesions. Some lesions cease to grow at maturity, but others continue to grow even into adult life. Malignant degeneration is extremely rare (an incidence of much less than 0.01 per cent).

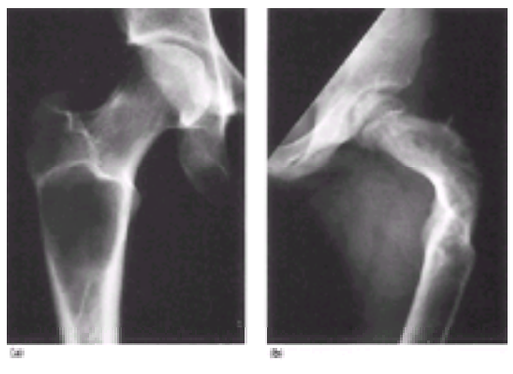


Fig. 8. (a) Fibrous dysplasia of the proximal femur of a 24-year-old woman; the lesion has a ‘ground glass' appearance with cortical thinning and good margination. (b) Typical ‘shepherd's crook' deformity in an 8-year-old girl with polyostotic fibrous dysplasia and precocious puberty (Albright syndrome).

Simple or unicameral bone cysts are lesions of unknown cause, common in the first two decades of life. They are assumed to arise from either a disorder of the growth plate or from a transient circulatory compromise due to a developmental anomaly of the veins of the affected bone. These cysts most commonly manifest themselves in the proximal humerus (55 per cent) or proximal femur (26 per cent) of a growing child, but other bones may be affected. The cysts are thought to resolve spontaneously during the second decade of life. In adults, cysts occur most frequently in the calcaneus or flat bones and are usually discovered incidentally or after pathologic fracture. The typical cyst in childhood appears as a central, radiolucent lesion on the metaphyseal side of the growth plate of a long bone (Fig. 9). The cortex is thinned but intact, and the lesion is usually well demarcated. The cavity is filled with a fluid similar to serum and extracellular fluid; this may be bloody if fracture has occurred. Treatment is by curettage and grafting, or subtotal resection with or without grafting, or, as has been recently advocated, by needle aspiration and instillation of methylprednisolone. This last procedure is associated with the least morbidity and is frequently successful: operative treatment is probably best reserved for patients who fail to respond to steroid injection.



Fig. 9. Unicameral bone cyst of the proximal humeral metaphysis in a 9-year-old boy: the cortex is thinned in places and the lesion is juxtaposed to the growth plate; the cyst is trabeculated and the inferior aspect appears to be filling in, probably due to an earlier fracture.

Aneurysmal bone cyst is a lesion of unknown cause that may occur occasionally as a primary tumor or, more often, is found in conjunction with pre-existing lesions such as giant-cell tumor, chondroblastoma, chondromyxoid fibroma, or fibrous dysplasia. Most aneurysmal bone cysts occur in patients less than 20 years of age, and affect the long bones and spine. Radiographically, these are eccentrically placed, expansile, purely lytic, and pseudotrabeculated tumors that are rather well demarcated from the bone and appear as a ‘blow out' in the cortical surface (Fig. 10) surrounded by a periosteal shell of new bone. Their differentiation from telangiectatic osteosarcoma and giant-cell tumor is important and often difficult. Treatment is either by curettage and grafting or marginal resection, depending upon the location. The recurrence rate is substantial.

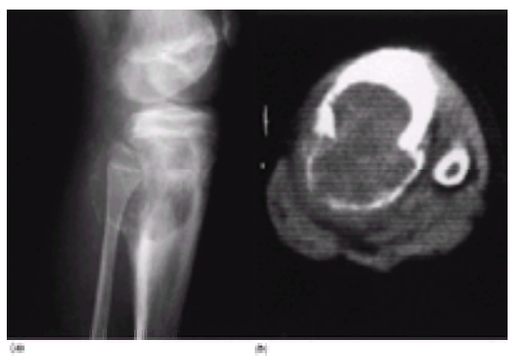


Fig. 10. (a) Aneurysmal bone cyst of the proximal tibial metaphysis in a 13-year-old girl: it is a large, lytic lesion that has destroyed the posterior tibial cortex, but is contained by a thin shell of periosteal bone. (b) The CT scan demonstrated destruction of the posterior tibial cortex, but the soft-tissue extent of the lesion is contained by a rim of periosteal new bone.

Giant-cell tumor of bone is an uncommon, aggressive, locally destructive lesion of the metaphyseoepiphyseal region of the long bones of adults that is benign but can, in unusual cases, metastasize to the lung. These tumors occur most frequently between the ages of 18 and 45 years, and are more common in women. The principal sites of predilection are the distal femur, proximal tibia, distal radius, proximal humerus, proximal femur, and proximal fibula, and occasionally the hand. When giant-cell tumor of bone involves the axial skeleton it arises in the vertebral body. When it arises in a long bone it is eccentrically placed and located in the metaphysis and epiphysis, extending to and sometimes through the subchondral cortex of the adjacent joint (Fig. 11). The cortex is almost always thinned, giving it an expanded appearance, and the lesion is entirely radiolucent. An extraosseous soft-tissue mass may be discernible by CT or MRI.



Fig. 11. (a, b) Anteroposterior and lateral radiographs of a giant-cell tumor of the right femur: a purely lytic lesion located in the metaphyseoepiphyseal aspect of the lateral femoral condyle, which extends down to the subchondral bone plate and is contained by a thin shell of periosteal membrane; note that the lesion is poorly marginated but well delineated on its medullary border. (c) CT scan of the lesion more fully outlines the extent of the lesion; the reactive periosteal shell is not complete but there is no frank soft-tissue mass.

Giant-cell tumor of bone is best treated by a thorough curettage and burring of the normal bone surface. Local recurrence is less with the most aggressive curettage but even with the best curettage the rate is approximately 25 per cent. Adjuvants such as phenol and freezing seem to reduce the incidence of local recurrence. If fracture and joint destruction occur, resection and allograft replacement has been advocated.

Langerhans cell histiocytosis (eosinophilic granuloma) is not truly a bone tumor but is believed to be a form of lipoid granulomatosis, a localized form of the much more aggressive and multiorgan Hand–Schueller–Christian disease (bone lesions, exophthalmos, diabetes insipidus) and Letterer–Siwe SYndrome (Hand–Schueller–Christian disease plus visceral abnormalities and central nervous changes with early death). Histiocytosis confined to one or occasionally several bones is a benign process that appears in the skeleton of a child as a painless, punched-out lesion. It is important to differentiate these lesions from more aggressive or malignant tumors and even from the more aggressive forms of histiocytosis.

Malignant

Malignant bone tumors vary in biologic behavior from locally aggressive lesions that seldom metastasize to highly anaplastic sarcomas with a dismal prognosis. They affect all age groups and all bony sites. Diagnosis is often difficult and treatment is complicated by their sometimes extremely aggressive nature, by their location in areas of functional importance, and by their rarity, which can result in delayed diagnoses.

Osteosarcoma (osteogenic sarcoma), the most frequently encountered malignant lesion of bone, is characterized histologically and often radiographically by the formation of bone or osteoid tissue by a sarcomatous stroma. Osteosarcoma is not a homogeneous disease and compromises several distinct clinicopathologic entities; these must be distinguished, since the biologic behavior of each may be quite different (Table 6). Central osteosarcoma, the most common form, is a high-grade tumor arising in the medullary cavity of the metaphysis or diaphysis of a teenage child. Classic osteosarcoma is the second most common primary malignant tumor of bone (first is myeloma), accounting for about 38 per cent of malignant bone neoplasms. There are two peaks of incidence, one between 10 and 25 years of age: this is a leading cause of cancer morbidity in adolescence, with only leukemia and lymphoma exceeding it in incidence in this age group. The second peak occurs in the fifth and sixth decades: this is probably related to the increased incidence of osteosarcoma in patients with Paget's disease of bone and those who have received radiation therapy. Osteosarcoma is more common in males, with a sex ratio (M:F) of 3:2. Fifty per cent of lesions occur about the knee, the distal femoral metaphysis being the single most common site; the next most frequent sites are the proximal tibia, proximal humerus, and proximal femur. The tumors are usually metaphyseal in location, although diaphyseal osteosarcomas occur; the spine and pelvis are less frequent sites. The flat bones and pelvis are more commonly affected in patients with Paget's disease and postirradiation sarcoma.

Primary osteosarcoma
Classical central osteosarcoma
Low-grade medullary osteosarcoma
Telangiectatic osteosarcoma
Multicentric osteosarcoma
Osteosarcoma in jaws
Secondary osteosarcoma
Paget's disease
Postradiation
In other benign lesions
Juxtacortical osteosarcoma
Parosteal osteosarcoma
Periosteal osteosarcoma

Table 6 Types of osteosarcoma

The characteristic symptoms of pain, local tenderness, a soft-tissue mass, and decreased function may be present for a variable period of time before diagnosis. Physical examination usually discloses a firm, tender mass fixed to the subjacent bone. There are no specific laboratory findings, although the serum alkaline phosphatase and lactate dehydrogenase are often increased.

On radiographs, osteosarcoma varies from a purely lytic to a predominantly blastic lesion, although a combination of both productive and destructive changes is usually noted (Fig. 12). The cortex is nearly always transgressed and a soft-tissue mass is common. The periosteum is elevated and may respond by forming reactive woven bone (Codman's triangle) at the periphery of the lesion. Perpendicular striations (classical ‘sun burst' appearance) and ‘onion skinning' of the periosteum are frequent findings. The lesion usually occupies the metaphysis of the bone but can spread by direct extension into the epiphysis and joint. There may be other foci of tumor within the same bone, separated by areas of normal bone. These ‘skip' lesions can usually be detected by plain radiographs and especially by bone scans, and are associated with a worse prognosis.

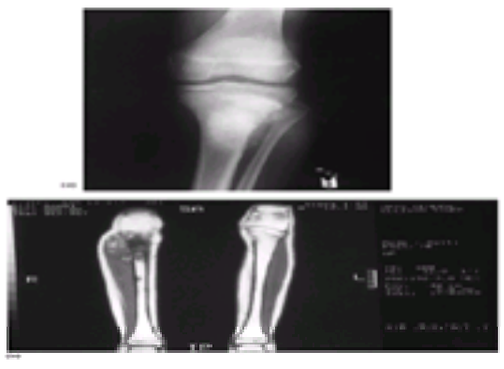


Fig. 12. Osteosarcoma: (a) a typical plain radiograph of an osteosarcoma (image is reversed); there is new bone formation with extension into the soft tissue. (b) On MRI the tumor extension is seen and a ‘skip' lesion in the tibial diaphysis is found.

The prognosis and natural history of osteosarcoma is variable: even tumors histologically classified as high-grade, central osteosarcomas are heterogeneous in their biologic behavior and response to treatment. The overall 5-year survival rate for patients with osteosarcoma treated by amputation alone was historically 15 to 30 per cent. Since the 1970s, however, the picture has changed radically. Current treatment of high-grade osteosarcoma consists of complete surgical eradication and aggressive adjuvant chemotherapy. Most patients receive preoperative or ‘neoadjuvant' chemotherapy for 2 to 3 months, followed by surgical resection (85 per cent) or amputation (15 per cent) of the primary tumor, and approximately 1 year of chemotherapy. Drug regimens include high-dose methotrexate, Adriamycin, cytoxan, ifosfamide, bleomycin, and actinomycin D. If, at the time of surgery after neoadjuvant therapy has been instituted, the ‘kill' is over 95 per cent, the disease-free survival rate is over 80 per cent. Further, the management of patients who develop pulmonary metastases has improved dramatically with the advent of aggressive resection of these lesions by thoracotomy. CT scanning of the lungs allows accurate definition of the extent of disease and resectability of pulmonary lesions. Most lesions are situated in the periphery of the lungs, making wedge resection feasible. Pulmonary resection combined with chemotherapy produces projected 5-year survival rates of greater than 30 per cent. Careful monitoring of patients following treatment of the primary disease is mandatory to allow early detection of relapses.

The various forms of chondrosarcoma of bone perhaps offer the most difficulty in diagnosis and staging, even in centers experienced in dealing with these rare lesions. Management problems are not only attributable to the rarity of the neoplasms, but also to the wide spectrum of histologic and radiographic presentation. Chondrosarcoma has been divided into primary and secondary lesions. Primary chondrosarcomas arise without a pre-existing cartilage tumor; secondary chondrosarcomas arise in a prior benign cartilage tumor. Most commonly, primary chondrosarcoma arises in the medullary canal and secondary chondrosarcoma arises from an exostosis. Subtypes of chondrosarcoma that are usually primary are recognized by their histologic characteristics; they include clear cell, mesenchymal, round cell, and dedifferentiated.

Chondrosarcomas occur principally in adults, with a peak incidence in the third to the sixth decade of life. The sites of predilection are the long bones, especially the femur, pelvis, and humerus, and the lesions are rare below the knees and elbows. The tumors are centrally placed, usually quite large at the time of presentation, and located in the metaphysis or diaphysis (or both). The patient usually complains of dull, aching pain of several months' duration, and often describes a tender mass that has slowly enlarged and distorted the contour of the bone and the extremity. Physical examination often discloses a large soft-tissue mass fixed to the underlying bone. These tumors occur with increased frequency in patients with enchondromatosis (Ollier or Maffucci syndromes).

The typical central chondrosarcoma has a characteristic radiologic appearance (Fig. 13). As defined above, the lesions are large, centrally placed within the bone, metaphyseodiaphyseal in location, and almost invariably produce significant expansion of the cortex. They may produce scalloping of the internal cortex, thinning, and, if high grade, cortical transgression and a soft-tissue mass, which often has a lobular appearance in outline on imaging studies such as CT (Fig. 14) Lower-grade lesions may cause the cortex to appear thicker than normal in some areas, and the central portion of both high- and low-grade tumors may show rounded or ring-shaped calcifications, not dissimilar to those seen in benign cartilage lesions. MRI reveals accurately the extent of the tumor, and is useful in assessing the degree of cortical transgression and the size of the soft-tissue mass, both of which are useful in distinguishing between low- and high-grade lesions.

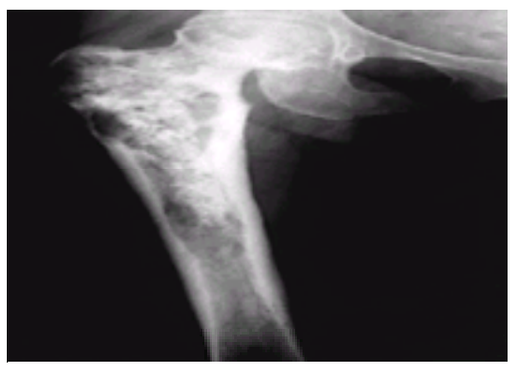


Fig. 13. Chondrosarcoma of the proximal femur in a 44-year-old woman with hip pain of several months' duration. The lesion is lytic with endosteal erosions. Biopsy showed this to be a low-grade chondrosarcoma, hyaline type.

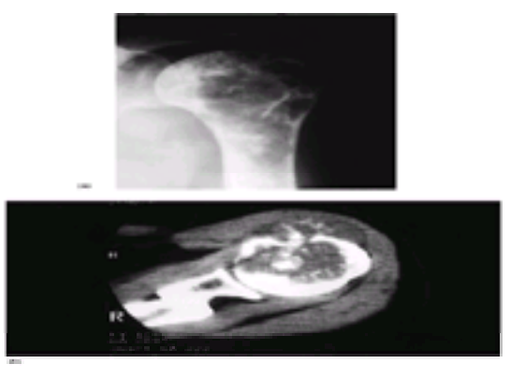


Fig. 14. Chondrosarcoma. (a) On this plain radiograph of the proximal humerus there is a destructive lesion within the humeral head and metaphysis; there are small, punctate calcifications present suggestive of calcified cartilage. (b) On the CT scan of this lesion the calcifications are clearly seen and an extraosseous component can be appreciated.

Secondary or exostotic chondrosarcomas most commonly arise from the cartilage cap of an osteocartilaginous exostosis. The diagnosis usually is made when the cartilage cap is thicker than 2 cm, and especially if it has grown after the patient is 30 years of age. Malignant exostotic tumors are large, proximally placed, and usually present with pain and a tender mass; this may reach an enormous size (particularly in the pelvis) by the time the patient seeks care.

The treatment of chondrosarcoma of bone is surgical resection. Adjuvant chemotherapy may be used in an attempt to control metastatic spread of high-grade lesions, but cannot be substituted for competent wide or radical surgery in the treatment of the local lesion. A wide surgical margin is required to make local recurrence unlikely.

Malignant fibrous tumors of bone are rarer than those forming bone and cartilage, and accounted for less than 4 per cent of malignant bone tumors in the Mayo Clinic series. The two major types are fibrosarcoma and malignant fibrous histiocytoma. Fibrosarcoma of bone is less frequent than is its soft-tissue counterpart. It most commonly presents as a central lesion, but may be periosteal in location; the latter is difficult to distinguish from soft-tissue fibrosarcoma located adjacent to a bone. Many such lesions occur secondarily to other processes such as Paget's disease of bone, giant-cell tumor, fibrous dysplasia, osteomyelitis, previously irradiated bone, and bone infarcts. Fibrosarcomas occur in both sexes and in all age groups, but most are found later in life (fourth to seventh decades). Long bones of the lower extremities contain most of the lesions: approximately half occur in the distal femoral and proximal tibial metaphyses, but any bone may be involved. Because they are purely destructive, pathologic fracture is common. High-grade fibrosarcomas show as lytic, poorly marginated, metaphyseal lesions, and usually demonstrate cortical breakthrough with an associated soft-tissue mass on plain radiographs. Malignant fibrous histiocytoma is similar to fibrosarcoma in presentation and is distinguished by it histologic appearance only. Both fibrosarcoma and malignant fibrous histiocytoma are treated by surgical resection with a wide margin. Fibrosarcomas tend to be of lower grade and most patients with this tumor are not treated with adjuvant chemotherapy, but malignant fibrous histiocytoma is a high-grade tumor with frequent development of pulmonary metastases and adjuvant chemotherapy is recommended. Survival without adjuvant chemotherapy is less than 40 per cent and can be increased to at least 60 per cent with adjuvant chemotherapy.

Treatment is largely surgical and follows the principles discussed under osteosarcoma. Low-grade lesions (stage I) and selected high-grade lesions (stage IIA and IIB) can be resected with wide margins, but stage II lesions may require a radical resection or amputation. Radiotherapy and chemotherapy are of little benefit, although radiotherapy may suppress the progression of disease or deter recurrences in elderly patients who refuse radical surgery. The 5-year survival rates vary from 28 to 34 per cent.

Chordoma is a rare neoplasm believed to arise from retained notochordal remnants; hence it is found predominately in the sacrum or the base of the skull in the region of the spheno-occipital SYnchondrosis and less commonly in the remaining lumbar, thoracic or cervical vertebral bodies. Patients are usually in the fourth to the seventh decade; the condition is slightly more frequent in men. Sacral chordomas are slowly growing and patients may give histories of long-standing, low-grade, mild discomfort in the lower spine and of tenderness on prolonged sitting. The patient usually complains of back pain. As the lesion grows it pushes forward on the rectum, causing constipation; and if untreated this is followed by sensory loss in sacral nerves, weakness of the musculature, impotence, saddle anesthesia, and finally, loss of bladder and bowel control. Lesions located in the clivus grow and invade the adjacent sella, regional cranial nerves, and brainstem; the neurologic findings associated with damage to these structures dominate the picture.

The radiographic hallmark of the sacrococcygeal chordoma is its location, which, by definition, must include the midline of the anterior body of these segments. Chordomas are always radiolucent, and show cortical destruction and poor margination. MRI is the most useful means of defining the nature and extent of these lesions in the sacrum, spine, and spheno-occipital region. Complete resection is the treatment of choice. Intralesional excision is attractive but invariably followed by local recurrence. Irradiation prolongs the disease-free interval and may reduce the incidence of local recurrence, but cannot manage bulk disease.

Round-cell lesions

Bone lesions composed of small, round cells are often difficult to distinguish from one another. Specialized histologic, immunohistologic, electron-microscopic, and even karyotypic studies are necessary to make the diagnosis. The pathologist must therefore be aware of the clinical situation at the time of biopsy so that the tissue can be properly prepared and studied. A frozen-section examination is necessary to establish that one is dealing with a round-cell lesion, but the exact diagnosis usually demands specialized studies. Such lesions include Ewing's sarcoma, malignant lymphoma of bone, myeloma, neuroblastoma, rhabdomyosarcoma, small-cell osteosarcoma, and metastatic small-cell carcinoma of the lung. Osteomyelitis and histiocytosis may also need to be considered in the differential diagnosis. It is important to note that because these lesions present differently, all are not in each differential diagnosis: Ewing's sarcoma, metastatic neuroblastoma, lymphoma, small-cell osteosarcoma, and Langerhans histiocytosis make up the differential diagnosis in the child patient; lymphoma, myeloma, and metastatic carcinoma are in the differential diagnosis for the adult. Rhabdomyosarcoma is a soft-tissue sarcoma of early childhood that rarely is included in the differential but is easily distinguished by special studies.

Ewing's sarcoma is a rare, round-cell lesion accounting for about 6 per cent of all malignant bone tumors. It is lethal, with fewer than 10 per cent of patients surviving without systemic chemotherapy. The tumor principally occurs in the second decade of life, and is rare below 5 years and over 30 years of age. Males are affected more commonly than females. Any bone may be involved, but the pelvis and lower extremities account for 60 per cent of lesions. In the long bones, the lesion shows a predilection for the metaphysis but is often diaphyseal, and, for reasons not clear, the fibula is often involved. Pain, swelling, and the presence of a mass are the usual presenting features, but a history of intermittent fever may precede the diagnosis by several months. An elevated sedimentation rate, leucocytosis, and anemia may be present but are not common. The occurrence of these systemic findings is felt by some to indicate a worse prognosis. Pathologic fractures are present in approximately 2 to 5 per cent of patients.

Analysis of Ewing's sarcoma cells has shown a consistent translocation between chromosomes 11 and 22 t(11;22) (q24;q12), suggesting that many of the lesions previously classified as Ewing's sarcoma are primitive neuroectodermal tumors. The latter presents in bone in an identical fashion to Ewing's sarcoma and is treated

similarly. Many pathologists now consider primitive neuroectodermal tumors and Ewing's sarcoma as one and the same.

The radiographic appearance of Ewing's sarcoma is that of a permeating, lytic, destructive lesion of bone, but there may be areas of reactive bone formation within the lesion and in the adjacent periosteum (Fig. 15). The findings are nonspecific, and often difficult to differentiate from those of osteosarcoma and infection. There is almost always an associated soft-tissue mass and the cortical integrity is compromised. The periosteal reaction often gives an ‘onion skin' appearance, or may produce a perpendicular striations surrounding the tumor. Occasionally, Ewing's sarcoma may occur as a periosteal lesion without medullary involvement, causing a disk-shaped impression on the outer cortex. On MRI a soft-tissue component is usually present at the time of presentation of the tumor.

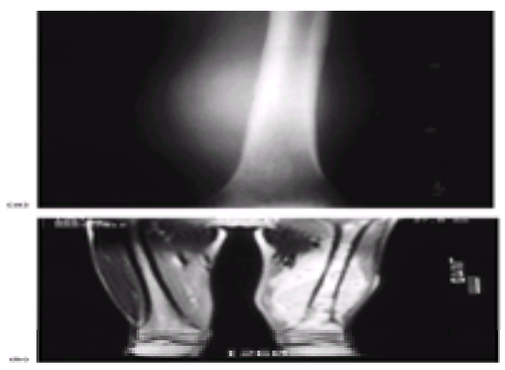


Fig. 15. Ewing's sarcoma. (a) On this plain radiograph of a young man's thigh a soft-tissue density can be seen; close inspection reveals a thickened medial cortex and periosteal reaction. (b) On MRI the extent of the disease is more easily seen; the tumor extends almost to the distal femoral epiphysis and there are two ‘skip' lesions in the proximal metaphysis.

Historically, treatment was with irradiation alone but the prognosis for patients with Ewing's sarcoma not given chemotherapy is poor. In a large series from the Mayo Clinic, the 5- and 10-year survival rates were 16.2 and 13.9 per cent, respectively, for those who did not present with metastatic disease. Patients with axial lesions fared worse than those with lesions of the extremities. Although the lungs are the most common site of metastasis, spread of disease to other bony sites and to lymph nodes is not uncommon. Multiagent chemotherapy, including Adriamcyin, vincristine, actinomycin and Cytoxan, has produced survival rates of about 50 to 65 per cent. Surgical resection is increasingly becoming the treatment of choice for the local lesion, although irradiation is still advocated by some, especially in anatomic locations not easily resected. The proper treatment of the local lesion remains controversial. The incidence of secondary sarcomas in the irradiated bone, an increased risk of local recurrence, and the local morbidity of irradiation are the principal reasons for favoring surgical resection.

Carcinoma metastatic to bone

Carcinoma metastatic to bone is more than twice as frequent as all primary bone tumors, both benign and malignant, in all age groups; in elderly individuals, primary bone tumors are rare and metastatic carcinoma is the most frequent cause of a bone lesion. Thus, a patient in the sixth decade of life who has a destructive lesion of bone should be considered to have a metastasis until proved otherwise, and appropriate studies should be performed as part of the staging.

Although almost any primary carcinoma may metastasize to bone, some are distinctly rare, while others do so with such frequency as to make them highly suspect in any patient thought to have secondary osseous deposits. In men, the most frequent primary site from which a metastatic bony lesion may arise is the prostate (Fig. 16). In women, the most frequent primary site is carcinoma of the breast. For both sexes, metastatic carcinoma from the lung ranks second to the breast and prostatic sites: these tumors produce destructive radiographic changes and are often distributed widely throughout the skeleton. Other primary carcinomas likely to spread to bone include renal-cell carcinoma, thyroid carcinoma, and less commonly, primary carcinoma from the gastrointestinal tract. Patients with renal-cell carcinoma often present with a bone lesion, and the site of origin may be suspected only when a biopsy shows the characteristic histology, or when the urinalysis shows occult hematuria or an imaging study shows a lesion of the kidney.

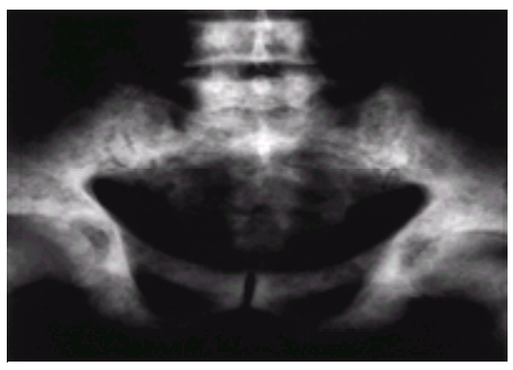


Fig. 16. Typical advanced case of metastatic prostate carcinoma to the spine and pelvis; note the blastic nature of the lesions, which involve every bone pictured.

The treatment of skeletal metastases from primary carcinomas depends on a variety of factors. A prime goal is to alleviate pain and to help the patient maintain optimal quality of life. Pain from metastatic deposits to bone is one, if not the most frequent, cause of pain in a cancer patient. A pathologic fracture is one of the more common causes of diminished mobility and can lead to marked restriction of a patient's activities. In addition, hypercalcemia is made worse by inactivity and can be an adverse consequence of a pathologic fracture. The site and size of the metastatic deposit, and the threat posed to skeletal integrity or adjacent vital structures such as the spinal cord, may dictate emergency surgical treatment and/or radiotherapy. In the long bones, prophylactic nailing, with or without methylmethacrylate cement, is employed to prevent pathologic fracture (Fig. 17). In the spine, an appropriate surgical decompression from posterior or anterior with internal fixation can produce normal neurologic function and a pain-free spine. If the patient has suffered a pathologic fracture of a long bone, open reduction and internal fixation is usually required to stabilize the skeleton prior to radiation therapy, hormonal therapy, or chemotherapy directed at eradicating or controling the lesion. Some tumors regress considerably with appropriate systemic management, and the use of antiestrogens in hormonally responsive breast cancer and estrogens in patients with prostatic carcinoma may produce long-lasting radiographic remission of symptoms and regression of lesions. Radiation therapy is usually very effective in controling metastatic deposits and is indicated for all lesions in the skeleton that are associated with pain or that require surgical treatment.



Fig. 17. A plain radiograph of a patient with a history of breast cancer who has pain in her thigh. The lesion is in the medial aspect of her proximal femur, has weakened the bone, and should be stabilized internally before the bone breaks.

Myeloma

Multiple myeloma (diffuse involvement with greater than 20 per cent plasma cells in a random bone-marrow aspirate) or plasmacytoma (solitary bone lesion with less than 10 per cent plasma cells in a random bone-marrow aspirate) is the single most common primary malignant tumor of bone. The cell of origin is the plasma cell and the neoplastic process is one of the family of disorders called monoclonal gammopathies. This group includes benign as well as malignant conditions, all of which are characterized by proliferation of a single clone of plasma cells that produce increased concentrations of gammaglobulin. Plasma-cell myeloma is for the most part a diffuse disorder affecting the entire bone marrow, but it occasionally presents as a single lesion with no apparent involvement of the rest of the skeleton; it is then referred to as a plasmacytoma. Most patients with ‘solitary' plasmacytoma develop evidence of the disseminated disorder within a few years.

Any bone can be affected, but the spine, ribs, skull, pelvis, and proximal long bones are the most common sites of disease. There tend to be two possible presentations. One is the patient with a solitary or a predominating lesion who complains of localized pain or presents with a pathologic fracture of the bone involved. These patients are usually not systemically ill and appear to be in good health. On the radiograph their bones may appear normal except for a well-defined radiolucent defect (so-called punched-out lesion). The other group have a more diffuse presentation and usually have more widespread disease, often with minimal skeletal lesions. They complain of diffuse pain in the bones. Not infrequently they will have a fractured vertebra or rib after minimal or no trauma. There is usually a history of anorexia, weight loss, malaise, and easy fatiguability of some months' duration, and on physical examination they will often appear chronically ill. The physical findings may suggest profound anemia and hepatosplenomegaly is sometimes present. Radiographs may show evidence of an acute fracture, but of greater importance is the finding of diffuse osteopenia and small, punctate or rounded radiolucencies in the skeleton, sometimes becoming confluent to produce large, lytic areas with severe thinning of the cortices and distortion of the normal contours of the bone. Technetium whole-skeletal bone scans are suggested as a means of screening the skeleton, but approximately 25 per cent of plasmacytoma lesions will not be associated with increased activity on the bone scan. A skeletal survey is recommended in addition to the scan.

Myeloma often involves the spine and when it does it is almost always the vertebral body that is damaged. If there is a soft-tissue component compromising the spinal cord, emergency irradiation is indicated; this should produce a rapid resolution of the extraosseous component and allow recovery of cord function. If there is vertebral collapse, with bone compressing the canal associated with loss of cord function, a surgical decompression is indicated (Fig. 18).

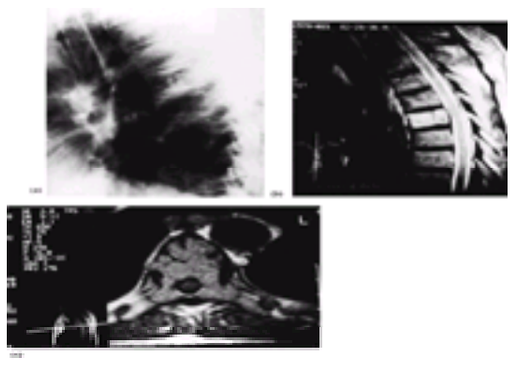


Fig. 18. Myeloma. (a) The spine is osteopenic and there is a subtle compression fracture of two vertebral bodies; the patient complained of back pain. (b) On MRI the signal from a single vertebral body suggests an infiltrative process. (c) On axial MRI extensive replacement of the bone marrow is seen; the canal is compromised. A biopsy was positive for malignant plasma cells.

Laboratory studies are helpful in establishing the diagnosis. Almost 90 per cent of patients with diffuse disease will have a normocytic normochromic anemia and a rapid erythrocyte sedimentation rate (usually above 50). The serum calcium may be elevated in patients with extensive osseous disease (partly on the basis of binding to globulin) and the serum urate is commonly elevated. The serum protein immunoelectrophoretic pattern may be virtually diagnostic: 90 per cent of patients demonstrate the presence of an abnormal, homogeneous ‘M component', usually migrating with the IgG or IgA fractions. The diagnostic test is the bone-marrow biopsy, which is likely to demonstrate a pathognomonic increase in the percentage of plasma cells, even in patients with few bone lesions.

The prognosis for patients with myeloma is poor. Most die within a few years, despite medical management. The use of bisphosphonates has greatly diminished both the symptoms of bone pain and the frequency of bone fractures. Radiation is effective in controling solitary lesions and almost always results in the relief of the pain. Fractures or weakened areas of the bone should be treated with internal fixation. The current chemotherapeutic agents of choice include melphalan and cyclophosphamide with corticosteroids or vincristine, Adriamycin, and Decadron, but remissions obtained by their administration are often short. The suggestion that intensive alkylating therapy with stem-cell transplantation is able to induce prolonged remission in some younger patients is leading to a re-evaluation of alkylating agents as initial therapy in this age group. Antiangiogenic agents are showing promise in clinical trials.

Further reading

Albright F, Butler AM, Hampton AO, Smith P. SYndrome characterized by osteitis fibrosa dissemmata, areas of pigmentation and endocrine dysfunction, with precocious puberty in females. Report of five cases. *New England Journal of Medicine* 1937; **216**: 727–46.[The original article describing the Albright syndrome in fibrous dysplasia.]

Alho A, Connor JF, Mankin HJ, Schiller AL, Campbell CJ. Assessment of malignancy of cartilage tumours using flow cytometry. *Journal of Bone and Joint Surgery* 1983; **65A**: 779–85. [DNA flow cytometry as a method of assessing malignancy.]

Ayala AG, Ro JY, Fanning CV, Flores JP, Yasko AW. Core needle biopsy and fine-needle aspiration in the diagnosis of bone and soft tissue lesions. *Hematologic-Oncologic Clinics of North America* 1995; **9**: 633–51. [Use of needle biopsy in bone and soft tissue tumors.]

Bacci G *et al.* Non-metastatic Ewing's sarcoma: results in 98 patients treated with neoadjuvant chemotherapy. *Italian Journal of Orthopaedic Traumatology* 1991; **17**: 449–65. [Initial excellent results in Ewing's sarcoma with chemotherapy.]

Barry CH. Orthopaedic aspects of Paget's disease of bone. *Arthritis and Rheumatism* 1980; **23**: 1128–30. [Frequency and baleful outcome from Paget's sarcoma.]

Bhardwaj S, Halland JF. Chemotherapy of metastatic cancer in bone. *Clinical Orthopedics and Related Research* 1982; **169**: 28–37. [Some chemotherapy protocols for the treatment of bone metastasis.]

Boriani S, Weinstein JN, Biagini R. Primary bone tumors of the spine. Terminology and surgical staging. *Spine* 1997; **22**: 1036–44. [Staging and characteristics of primary tumors of the vertebra.]

Brady TJ *et al.* NMR imaging of forearms in healthy volunteers and patients with giant-cell tumour of bone. *Radiology* 1982; **144**: 549–52. [The first MRI of a distal radial giant cell tumor.]

Campanacci M. Chondrosarcoma. In: *Bone and soft tissue tumors*, pp. 265–338. Springer, New York, 1990. [An excellent well illustrated description of this difficult tumor.]

Capanna R, Bertoni F, Bettell G, Picci P, Bacchini P, Present D, Campanacci M. Dedifferentiated chondrosarcoma. *Journal of Bone and Joint Surgery* 1988; **70A**: 60–9. [Clinical, histologic, and outcome studies for dedifferentiated chondrosarcoma.]

Cohen J. Unicameral bone cysts. A current SYnthesis of reported cases. *Orthopedic Clinics of North America* 1972; **8**: 715–36. [A review of the characteristics of unicameral bone cyst.]

Dahlin DC, Coventry MB. Osteogenic sarcoma. A study of six hundred cases. *Journal of Bone and Joint Surgery* 1967; **49A**: 101–10. [A very large series of Mayo Clinic-treated osteosarcomas.]

Dahlin DC, Henderson ED. Mesenchymal chondrosarcoma. Further observations in a new entity. *Cancer* 1962; **15**: 410–17. [Description and histology for a quite different tumor from chondrosarcoma.]

Dahlin DC, Capps RE, Johnson EW, Jr. Giant cell tumour: a study of 195 cases. *Cancer* 1970; **25**: 1061–70. [An early description of a large series of giant cell tumors.]

Dahlin DC, Unni KK, Matsuno T. Malignant (fibrous) histiocytoma of bone—fact or fancy? *Cancer* 1977; **39**: 1508–16. [An early review of this tumor, questioning its origin and other possible causes.]

Damron TA, Pritchard DJ. Current combined treatment of high grade osteosarcomas. *Oncology* 1995; **9**: 327–43. [Current practice of Mayo Clinic approach to osteosarcoma.]

Devlin JA, Bowman HE, Mitchell CL. Non-osteogenic fibroma of bone. A review of the literature with the addition of six cases. *Journal of Bone and Joint Surgery* 1955; **37A**: 472–86. [Early report of

non-ossifying fibroma of bone.]

Diller L *et al.* p53 functions as a cell cycle protein in osteosarcomas. *Molecular and Cellular Biology* 1990; **10**: 5772–81. [P-53 aberration as a stimulus to tumor uncontrolled growth.]

Dome JS, Schwartz CL. Osteosarcoma. *Cancer Treatment and Research* 1997; **92**: 215–51. [Current management protocols and projected outcome for osteosarcoma.]

Enneking WF, Spanier SS, Goodman MA. A SYSTEM for the surgical staging of musculoskeletal sarcoma. *Clinical Orthopedics and Related Research* 1980; **153**: 106–20. [Original article describing a very useful method of staging of bone tumors.]

Eyer-Brook AL, Price CHG. Fibrosarcoma of bone. Review of fifty consecutive cases from the Bristol bone tumour registry. *Journal of Bone and Joint Surgery* 1969; **51B**: 20–37. [One of the early reports describing the characteristics of fibrosarcoma of bone.]

Fletcher BD. Imaging pediatric bone sarcomas. Diagnosis and treatment-related issues. *Radiologic Clinics of North America* 1997; **35**: 1477–94. [Imaging characteristics for bone tumors in children.]

Gebhardt MC. Oncogenes and bone tumours. *Current Opinion in Orthopedics* 1990; **1**: 394–9. [The nature and role of oncogenes in a variety of bone tumors.]

Gebhardt MC. Molecular biology of sarcomas. *Orthopedic Clinics of North America* 1996; **27**: 421–9. [Oncogenes, P-53, RB, and other markers for bone tumors.]

Gebhardt MC, Flugstad DI, Springfield DS, Mankin HJ. The use of bone allografts for limb salvage in high-grade extremity osteosarcoma. *Clinical Orthopedics and Related Research* 1991; **270**: 181–96. [Use and outcomes of allograft parts in the treatment of bone sarcomas.]

Gebhardt MC, Lew RA, Bell RS, Baldini N, Litwak G, Mankin HJ. DNA ploidy as a prognostic indicator in human osteosarcoma. *Chirurgia degli Organi d'Movementi* 1990; **75**: 18–21. [DNA flow cytometric determination as a marker for malignancy in osteosarcoma.]

Glaubiger DL, Makuch RW, Schwartz J. Influence of prognostic factors on survival in Ewing's sarcoma. *National Cancer Institute Monographs* 1981; **56**: L285–8. [An original article describing the prognostic features for Ewing's sarcoma.]

Goldenberg RR, Campbell CJ, Bonfiglio M. Giant cell tumour of bone. An analysis of two hundred and eighteen cases. *Journal of Bone and Joint Surgery* 1970; **52A**: 619–64. [A multicenter study of outcome of giant cell tumor – high recurrence rate.]

Goorin AM, Abelson HT, Frei E, III. Osteosarcoma: fifteen years later. *New England Journal of Medicine* 1985; **313**: 1637–43. [The early report of very promising results of chemotherapy for osteosarcoma.]

Goorin AM, Andersen JW. Experience with multiagent chemotherapy for osteosarcoma. Improved outcome. *Clinical Orthopedics and Related Research* 1991; **270**: 22–8. The follow-up on even better results with multiple agents for osteosarcoma.]

Goorin AM *et al.* Weekly high-dose methotrexate and doxorubicin for osteosarcoma: the Dana–Farber Cancer Institute/The Children's Hospital—study III. *Journal of Clinical Oncology* 1987; **5**: 1178–84. [Report on the use of methotrexate and Adriamycin for osteosarcoma.]

Granowetter L. Ewing's sarcoma and extracranial peripheral neuroectodermal tumors. *Current Opinions in Oncology* 1995; **7**: 333–9. [Ewing's tumor as a possible lesion of neural axis origin.]

Han MT *et al.* Aggressive thoracotomy for pulmonary metastatic osteogenic sarcoma in children and young adolescents. *Journal of Pediatric Surgery* 1981; **16**: 928–33. [Use of thoracotomy and removal of lung metastases in osteosarcoma patients.]

Harmon TP, Morton KS. Osteogenic sarcoma in four siblings. *Journal of Bone and Joint Surgery* 1966; **48B**: 493–8. [A family history of four siblings with osteosarcoma lends credence to a genetic basis.]

Higginbotham NL, Phillips RF, Farr HW, Hustu HO. Chordoma, thirty-five year study at Memorial Hospital. *Cancer* 1967; **20**: 1841–50. [History of the chordomas treated at the Memorial Hospital in New York.]

Huvos AG. *Bone tumours. Diagnosis, treatment and prognosis*. Saunders, Philadelphia, 1991. [A classic text defining the extraordinary experience of a New York Memorial Hospital pathologist.]

Inwards CY, Unni KK. Classification and grading of bone sarcomas. *Hematologic-Oncologic Clinics of North America* 1995; **9**: 545–69. [A definition of grading systems for bone tumors.]

Kyle RA. Multiple myeloma. Review of 869 cases. *Mayo Clinic Proceedings* 1975; **50**: 29–40. [A review of multiple myeloma based on the Mayo series of almost 900 cases.]

Lagenstein R, Philip T, Gardner H. Autologous stem cell transplantation for solid tumors in children. *Current Opinions in Pediatrics* 1997; **9**: 55–69. [Use of stem cell transplant in dealing with children's sarcomas.]

Levine AS. *Cancer in the young*, pp. 575–602. Masson, New York, 1982. [A review of the bone tumors in young children in an excellent text.]

Lewis RJ, Ketcham AS. Maffucci's syndrome: functional and neoplastic significance. Case report and review of the literature. *Journal of Bone and Joint Surgery* 1973; **55A**: 1465–79. [Some experience with Maffucci's syndrome, a potential deadly disease.]

Lewis RJ, Marcove RC, Rosen G. Ewing's sarcoma—functional effects of radiation therapy. *Journal of Bone and Joint Surgery* 1977; **59A**: 325–31. [Ewing's tumor treated with X-ray therapy, prior to the wide use of drugs.]

Link MP *et al.* The effects of adjuvant chemotherapy on relapse-free survival in patients with osteosarcoma of the extremity. *New England Journal of Medicine* 1986; **314**: 1600–6. [Another view of the successful use of drugs in patients with osteosarcoma.]

Link MP *et al.* Adjuvant chemotherapy of high-grade osteosarcoma of the extremity. Updated results of the multi-institutional osteosarcoma study. *Clinical Orthopedics and Related Research* 1991; **270**: 8–14. [An update on the previous study by Link, this time with multiple centers.]

Malkin D *et al.* Germline mutations of the p53 tumour-suppressor gene in children and young adults with second malignant neoplasms. *New England Journal of Medicine* 1992; **326**: 1309–15. [The incidence of p53 errors in patients with osteosarcoma and a second tumor.]

Mankin HJ, Connor JF, Schiller AL, Perlmutter N, Alho A, McGuire M. Grading of bone tumours by analysis of nuclear DNA content using flow cytometry. *Journal of Bone and Joint Surgery* 1985; **67A**: 404–13. [DNA flow cytometry as a method of grading bone tumors.]

Mankin HJ, Doppelt S, Tomford W. Clinical experience with allograft implantation. The first ten years. *Clinical Orthopedics and Related Research* 1983; **174**: 69–86. [The first 10 years of allograft transplant and the problems encountered.]

Mankin HJ, Gebhardt MC, Jennings LC, Springfield DS, Tomford WW. Long term results of allograft replacement in the management of bone tumors. *Clinical Orthopaedics* 1996; **324**: 86–97. [The long-term results of allograft transplant; 2- to 3-year follow-up.]

Mankin HJ, Gebhardt MC, Springfield DS, Litwak GJ, Kusazaki K, Rosenberg AE. Flow cytometric studies of human osteosarcoma. *Clinical Orthopedics and Related Research* 1991; **270**: 169–80. [Flow studies for DNA in osteosarcoma as a way of predicting outcome.]

Mankin HJ, Lange TA, Spanier SS. The hazards of biopsy in patients with malignant primary bone tumours and soft tissue tumours. *Journal of Bone and Joint Surgery* 1982; **64A**: 1121–7. [A group study of problems that occurred with biopsy.]

Mankin HJ, Mankin CJ, Simon MA: The hazards of biopsy revisited. *Journal of Bone and Joint Surgery* 1996; **78A**: 656–63. [A 10-year re-look at the biopsy to see if we learned anything – no, we didn't!]

Marcove RC, Mike V, Hajek JV, Levin AG, Hutter RVP. Osteogenic sarcoma under the age of twenty-one. A review of one hundred and forty-five operative cases. *Journal of Bone and Joint Surgery* 1970; **52A:** 411–23. [A study from the Memorial Hospital showing the results of operation for osteosarcoma.]

Mertens WC, Branwell V. Osteosarcoma and other tumors of bone. *Current Opinions in Oncology* 1995; **7**: 349–54. [A review of malignant bone tumors.]

Mindell ER. Chordoma (current concepts review). *Journal of Bone and Joint Surgery* 1981; **63A**: 501–5. [A review article defining the nature, management, and results for chordoma.]

Morris HG, Capanna R, Delben M, Campanacci D. Prosthetic reconstruction of the proximal femur after resection for bone tumors. *Journal of Arthroplasty* 1995; **10**: 293–9. [Use of metallic implants for proximal femoral tumor resection.]

O'Donnell RJ, Springfield DS, Motwani HK, Ready JE, Gebhardt MC, Mankin HJ. Recurrence of giant cell tumors of the long bones after curettage and packing with cement. *Journal of Bone and Joint Surgery* 1994; **76**: 1827–33. [Giant cell tumors treated with curettage and cement – better results than with graft.]

Pitcher JD, Gebhardt MC, Springfield DS, Rosenberg AE, Mankin HJ. Osteosarcoma: the Massachusetts General Hospital–Children's Hospital 20 years experience. In: Novak JF, McMaster JH, ed. *Frontiers of osteosarcoma research*, pp. 17–24. Hogref and Huber, Seattle, 1993. [Osteosarcoma results followed over 20 years including allograft SYstems.]

Pompetti F *et al.* Oncogenic alteration in primary recurrent and metastatic bone tumors. *Journal of Cell Biochemistry* 1996; **63**: 37–50. [Oncogenes in the various malignant tumors are helpful in

predicting behavior.]

Pritchard DJ, Dahlin DC, Dauphine RT, Taylor WF, Beabout JW. Ewing's sarcoma. A clinicopathological and statistical analysis of patients surviving five year or longer. *Journal of Bone and Joint Surgery* 1975; **57A**: 10–16.

Pritchard DJ, Lunke RJ, Taylor WF, Dahlin DC, Medley BE. Chondrosarcoma: a clinicopathologic and statistical analysis. *Cancer* 1980; **45**: 149–57. [A study of a large number of chondrosarcomas from the Mayo Clinic.]

Rosenthal DI, Springfield DS, Gebhardt MC, Rosenberg AE, Mankin HJ. Osteoid osteoma: percutaneous radiofrequency ablation. *Radiology* 1995; **197**: 451–4. [Use of heat ablation for osteoid osteoma – early success.]

Saeter G, Alvegard TA, Elomaa I, Wiebe T, Bjork O, Strander H, Solheim OP. Chemotherapy for osteosarcoma and Ewing's sarcoma. *Acta Orthopaedica Scandinavica* 1997; **273**: 120–5. [Chemotherapeutic successes and failures for high-grade primary bone tumors.]

Samson IR, Springfield DS, Suit HD, Mankin HJ. Operative treatment of sacrococcygeal chordoma. *Journal of Bone and Joint Surgery* 1993; **75A**: 1476–84. [A close look at 21 sacral chordomas treated by posterior sacretomy.]

Scaglietti O, Marchetti PG, Bartolozzi P. Final results obtained in the treatment of bone cysts with methylprednisolone acetate (Depo-Medral) and a discussion of results achieved in other bone lesions. *Clinical Orthopedics and Related Research* 1982; **165**: 33–42. [Use of corticosteroid injection in the treatment of bone cysts.]

Schajowicz F, Derdui JC. Puncture biopsies in lesions of the locomotor system. Review of results in 4,050 cases, including 941 vertebral punctures. *Cancer* 1968; **21**: 531–48. [Use of punctures to obtain material for biopsy – a large series with good results.]

Schajowicz F, Gallardo H. Epiphyseal chondrosarcoma of bone. A clinico-pathological study of sixty-nine cases. *Journal of Bone and Joint Surgery* 1970; **52B**: 205–26. [Chondrosarcoma arising from epiphyseal sites on bone surfaces.]

Scully SP, Teple HT, O'Keefe RJ, Scarborough MT, Mankin HJ, Gebhardt MC. Role of surgical resection in pelvic Ewing's sarcoma. *Journal of Clinical Oncology* 1995; **13**: 2336–41. [Bad result of pelvic Ewing's tumor, independent of treatment.]

Sherry HS, Levy RN, Siffert RS. Metastatic disease of bone in orthopaedic surgery. *Clinical Orthopedics and Related Research* 1982; **169**: 44–52. [Metastatic disease to bone -diagnosis and treatment.]

Shocken JD, Luther WB. Radiation therapy for bone metastasis. *Clinical Orthopedics and Related Research* 1982; **69:** 38–43. [Use of radiation in the treatment of bone metastases – a good technique.]

Silverberg E. Cancer statistics, 1984. *CA: a Cancer Journal for Clinicians*, 1984; **34**: 7–23. [Cancer statistics – how to do it and the meaning of the results.]

Simon MA. Biopsy of musculoskeletal tumours. *Journal of Bone and Joint Surgery* 1982; **64A**: 1253–7. [The biopsy – problems, best techniques, and outcomes.]

Simon MA, Karluk MB. Skeletal metastases of unknown origin. Diagnostic strategy for orthopaedic surgeons. *Clinical Orthopedics and Related Research* 1982; **166**: 96–103. [Bone metastases – how to work them up and results of such studies.]

Simon MA, Aschliman MA, Thomas N, Mankin HJ. Limb-salvage treatment versus amputation for osteosarcoma of the distal end of the femur. *Journal of Bone and Joint Surgery* 1986; **68**: 1331–7. [Review of long-term results comparing limb salvage and amputation.]

Slowick FA, Campbell CJ, Kettelkamp DB. Aneurysmal bone cyst. An analysis of thirteen cases. *Journal of Bone and Joint Surgery* 1968; **50A**: 1142–51. [Aneurysmal bone cyst – a review of cases and the literature.]

Solomon L. Hereditary multiple exostosis. *Journal of Bone and Joint Surgery* 1963; **45B**: 292–304. [Hereditary multiple exostoses – a broad overlook of the problem.]

Spanier SS, Enneking WF, Enriquez P. Primary malignant fibrous histiocytosis of bone. *Cancer* 1975; **36**: 2084–98. [A rare tumor with bad results unless treated aggressively.]

Springfield DS, Gebhardt MC, McGuire MH. Chondrosarcoma: a review. *Instructional Course Lectures* 1996; **45**: 417–24. [A review of the literature and the findings and outcome for chondrosarcoma.]

Springfield DS, Schmidt R, Graham-Pole J, Marcus RB, Jr, Spanier SS, Enneking WF. Surgical treatment for osteosarcoma. *Journal of Bone and Joint Surgery* 1988; **70**: 1124–30. [Surgery for osteosarcoma – planning, technique, and outcome.]

Sweet DL, Mass DP, Simon MA, Shapiro CM. Histiocytosis lymphoma (reticulum-cell sarcoma) of bone. Current strategy for orthopaedic surgeons. *Journal of Bone and Joint Surgery* 1981; **63A**: 79–84. [Non-Hodgkin's lymphoma of bone – diagnostic and teatment SYstem.]

Toguchida J *et al.* Prevalence and spectrum of germline mutations of the p53 gene among patients with sarcoma. *New England Journal of Medicine* 1992; **326**: 1301–8. [P53 mutations in patients with osteosarcoma.]

Unni KK, Dahlin DC, Beabout JW, Sim FH. Chondrosarcoma: clear cell variant. *Journal of Bone and Joint Surgery* 1976; **58A**: 676–83. [Clear cell chondrosarcoma – location, findings, and results of treatment.]

Van der Griend RA. Osteosarcoma and its variants. *Orthopedic Clinics of North America* 1996; **27**: 575–81. [A review of osteosarcoma and especially the variations seen.]

Van der Woude HJ, Bloem JL, Hogendoorn PC. Preoperative evaluation and monitoring chemotherapy in patients with high grade osteogenic and Ewing's sarcoma: review of current imaging modalities. *Skeletal Radiology* 1998; **27**: 57–71. [Review of imaging technology for osteosarcoma treatment.]

Velez-Yanguas MC, Warrrier RP. The evolution of chemotherapeutic agents for the treatment of pediatric musculoskeletal malignancies. *Orthopedic Clinics of North America* 1996; **27**: 545–9. [The history of chemotherapy in malignant bone tumors in children.]

Vlasik R, Sim FH. Ewing's sarcoma. *Orthopedic Clinics of North America* 1996; **27**: 591–603. [Mayo Clinic experience with Ewing's sarcoma.]

Wallace S *et al.* Arterial occlusion of pelvic bone tumour. *Cancer* 1979; **43**: 2–8. [Techniques for vascular interuption in the management of pelvic bone tumors.]

Wick MR, Siegel GP, Unni KK, McLeod RA, Geditzer HG. Sarcomas of bone complicating osteitis deformans (Paget's disease). Fifty years experience. *American Journal of Surgical Pathology* 1981; **5**: 47–59. [Paget's sarcoma with special emphasis on pathology from the Mayo Clinic.]

Wuisman P, Enneking WF. Prognosis for patients who have osteosarcoma with skip metastasis. *Journal of Bone and Joint Surgery* 1990; **72**: 60–8. [Skip metastasis is not uncommon and represents a bad thing in osteosarcoma.]

47.1.1 Abscess, cellulitis, and necrotizing bacterial infections

Ian Oapos;Rourke, Phillip Carson and Gary Lum

Introduction
The inflammatory response
Surgical abscesses
Pathogenesis of abscess formation
Clinical features
Management
Cellulitis
Pathogenesis
Clinical features
Management
Necrotizing bacterial infections
Introduction
Classification and nomenclature
Necrotizing infections primarily affecting skin
Necrotizing infections primarily affecting subcutaneous tissues
Necrotizing infection of muscle
Further reading

Introduction

Abscesses, cellulitis, and necrotizing bacterial infections are seen more commonly in developing countries because of the higher incidence of malnutrition with its resultant immunodepression. Nevertheless, the principles of diagnosis and management are no different to those followed in the developed world.

The inflammatory response

To help understand the pathogenesis, clinical features, and management of abscess and cellulitis, it is important to review briefly the key features of the inflammatory response. This topic is described in detail in the chapter ‘Inflammation and repair' in Robin's *Pathologic basis of disease* (see Further Reading).

Acute inflammation is a protective response by the body to neutralize the initial cause of cellular injury and remove necrotic and damaged tissue. After injury or infection the following responses occur (Box 1).

Surgical abscesses

Pathogenesis of abscess formation

Abscesses are caused by deposits of pyogenic (pus-producing) bacteria in tissues, organs, or confined spaces of the body. This can occur when there is a breach in the continuity of the skin as happens after minor trauma or other conditions (Table 1). Abscesses associated with the alimentary tract are seen when there is a major breach in its integrity such as a ruptured appendix or, in the case of perianal sepsis, an infected anal gland. Abscesses developing in body organs can arise by direct infection (lung) or by bacteraemic spread (liver).

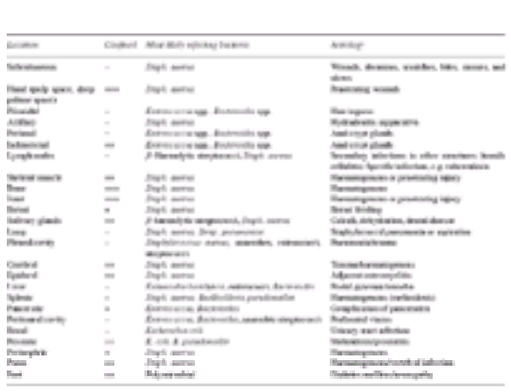


Table 1 Types of abscesses depending on location, compartment, and aetiology

Having penetrated the host's external barrier, the bacteria elicit an intense inflammatory response with an outpouring of plasma and leucocytes. Phagocytosis and killing of the bacteria occur as outlined above. However, when there is infection with micro-organisms that are resistant to phagocytosis and killing, or if there is an area of highly localized necrosis, a quantity of fluid is produced by the inflammatory process consisting of degraded protein, plasma, and dead and living neutrophils, and this is defined as pus. An abscess is formed when a collection of pus is walled off by an inflammatory reaction. This appears histologically as a central region consisting of necrotic leucocytes and tissue cells; outside this is a zone of living neutrophils and outside this there is vascular dilation and fibroblastic proliferation. The main pyogenic organisms from the surgical viewpoint are *Staphylococcus aureus*, b-haemolytic streptococci, *Escherichia coli*, and *Bacteroides* spp.

Clinical features

The clinical features of abscess are systemic and local. The systemic effects depend on the size of the abscess and the virulence of the organism(s) but are similar regardless of the location. The local effects depend on the site.

The systemic effects of inflammation, also known as ‘acute-phase reactions', include fever, increased sleep, increased catabolism, hypotension, and synthesis of C-reactive protein by the liver. Most of these reactions are mediated by the cytokines interleukin (**IL**)1, IL-6, and tumour necrosis factor, which are released from leucocytes during the inflammatory process. In view of the importance of leucocytes in the inflammatory process it is not surprising that leucocytosis is a significant systemic feature.

The location of the abscess will clearly influence the clinical presentation but some or all of the classical local features of inflammation—redness, swelling, heat, pain, and loss of function—will be present. Where the abscess is surrounded by loose, compliant tissue (for example the subcutaneous tissues, lung or peritoneal cavity), localization of pus is brought about by the induced tissue reaction (pyogenic membrane). These abscesses will demonstrate fluctuance and may drain spontaneously to the skin or to a hollow viscus. Abscesses confined by strong fascial planes (for example the finger pulp space or a deep abscess of the hand), however, create high pressure and can lead to ischaemic necrosis of the tissues in the space concerned.

Table 1 outlines the types and locations of various abscesses as well as indicating their causes and whether they are confined by fascial planes.

Box 1

RESPONSE	CHEMICAL MEDIATOR
Increased vascular perfusion of the affected area brought about by arteriolar vasodilatation and opening of new capillary beds	*Vasodilatation is probably brought about by prostaglandins and nitric oxide (NO)*
Increased capillary permeability, which results in escape of protein-rich fluid (exudate) and leucocytes from the circulation into the tissues at the site of the injury	*Histamine, the anaphylactins (C3a, C5a, bradykinin) and the leukotrienes*
Transmigration of leucocytes by a process of chemotaxis	*Brought about chemoattractants derived from rapidly dividing pyogenic bacteria* *The chemoattractants include peptides that are recognized by receptors on the neutrophils* *Endotoxins released from the bacteria stimulate the release of cytokines from macrophages* *Components of the complement system are also important in this process*
Phagocytosis of the offending agent, which includes recognition and attachment, engulfment, and then killing of the organism This process can induce activated leucocytes to release toxic metabolites and proteases	
Tissue damage from the inflammatory process	*Probably caused by the neutrophil and macrophage lysosomal enzymes, oxygen-derived free radicals, and NO*
Fever	*Cytokines, either directly or through the local production of prostaglandins, stimulate the hypothalamus, which acts on the vasomotor system causing sympathetic stimulation and cutaneous vasoconstriction that limits heat dissipation*
Leucocytosis	*Caused by increased release of leucocytes from the bone marrow stimulated by IL-1 and tumour necrosis factor* *With prolonged sepsis there is proliferation of precursors in the bone marrow stimulated by colony-stimulating factors*

Pyomyositis is an abscess developing in skeletal muscle. It is seen mainly in tropical areas of the world and is probably caused by bacteria spreading to a haematoma in the muscle via the circulatory SYstem. The most common organism is *Staph. aureus*.

Abscesses deep within the body (liver, lung, and body cavities) often present with the systemic features of inflammation without localizing signs.

Management

The type and location of the abscess (Table 1) will determine the specific management. The management will therefore differ considerably, but there are overriding principles of management for all abscesses and these include:

1. drainage of pus;
2. removal of necrotic tissue and foreign material;
3. correction of the predisposing cause;
4. antimicrobial cover.

The time-honoured approach to dealing with all abscesses has been a large adequate incision, debridement of the cavity, and continuous free drainage either by serial packing if superficial or by large-calibre drains for deeper cavities. This remains a safe and effective approach (Fig. 1 and Fig. 2).



Fig. 1. Abscess arising in the iliac lymph nodes.

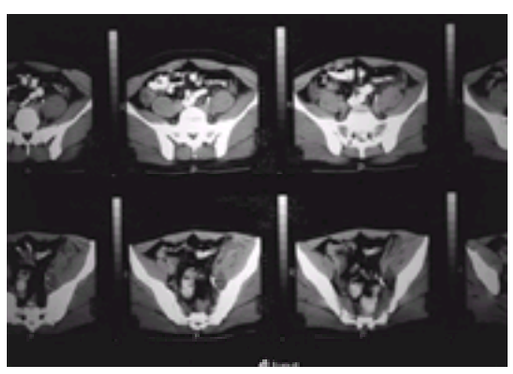


Fig. 2. Computed tomogram of abscess of iliac region that on exploration was the abscess illustrated in Fig. 1.

Over the last few decades, antimicrobials and interventional radiological techniques have allowed considerable evolution of this traditional approach with the aim of reducing treatment-related morbidity. Provided the contents are sufficiently liquid, many deep and visceral abscesses are successfully drained by fine-bore, image-guided, percutaneously placed catheters. Aspiration without drainage under antimicrobial cover is occasionally effective (e.g. pyogenic liver abscess).

Superficial abscesses can also be treated by the closed method as advocated by Ellis. This involves incision and drainage of pus, debridement of overlying non-vital skin, and curettage of the pyogenic membrane. Under appropriate antimicrobial cover the cavity is obliterated by either percutaneous encircling sutures or a closed suction drain (in place for 24 h only) and the skin primarily closed. Controlled trials have demonstrated primary healing in over 80 per cent of cases, with demonstrable advantages of less pain, faster healing, and lower costs when compared to the traditional open packing. Continuing or recurrent suppuration indicates the need for early suture removal and the reinstitution of open packing.

Timing of surgery

Most abscesses should be drained as soon as practicable. However, in the absence of systemic features of toxaemia, it is appropriate to allow superficial abscesses in unconfined spaces to point until fluctuant. Abscesses in confined spaces (Table 1) should be drained as a matter of urgency to avoid extensive soft-tissue destruction.

Antimicrobials and abscess

Localized, superficial abscesses treated by open drainage can be safely treated without antimicrobials. In these cases it is prudent to culture the purulent exudate.

The closed technique of treating abscesses requires high perioperative concentrations of antimicrobials. Percutaneous drainage of deep abscesses will require a 5-day course of antimicrobial cover. The choice of antimicrobial will initially be empirical, but in view of the known pyogenic bacteria for most types of abscesses (Table 1) the initial choice should be reasonably clear. The results of culture and sensitivity will indicate the exact type required. The vast majority of subcutaneous abscesses will be

caused by *Staph. aureus*. If acquired in the community they will most likely be methicillin susceptible. In hospitals they will possibly be methicillin resistant (**MRSA**) depending on the hospital's antibiogram. However, there is an increasing incidence of MRSA acquired in the community. This is being seen increasingly throughout Australia and the rest of the world.

Specific infections such as tuberculosis and melioidosis require prolonged courses of antimicrobials to prevent relapse. Small, multiple, visceral abscesses may respond to antimicrobials alone.

Treatment of underlying cause

All abscesses should be approached with careful consideration of the underlying cause, which must, if possible, be corrected. Unusual locations may suggest a specific aetiological factor, for example melioidosis abscess of the prostate. Visceral perforation such as diverticulitis or perforated appendicitis will require surgical correction. Osteomyelitis may present as an apparent superficial abscess. Foreign bodies and occult malignancy should always be considered.

The poor wound healing and decreased response to infection in patients with diabetes mellitus are additional difficulties. Immunosuppressed patients receiving chemotherapy or who have acquired immune deficiency SYndrome will have a very poor inflammatory response and resistance to infection.

Cellulitis

Pathogenesis

Cellulitis is a diffuse, spreading infection of the deeper layers of the skin and the subcutaneous tissues. By far the most common infective organisms are the b-haemolytic streptococci, e.g. *Strep. pyogenes*(group A streptococcus). *Staphylococcus aureus* is also a major cause of cellulitis after the b-haemolytic streptococci. These bacteria produce enzymes that break down intercellular barriers and promote spread. Despite the fact that cellulitis is thought to be an infective process, bacteria can only be isolated by needle aspiration or biopsy in 30 per cent of cases (technique may cause variations in results). This is probably due to the fact that the bacteria are being rapidly cleared and diluted by the inflammatory process. The release of the cytokines IL-1 and tumour necrosis factor from the Langerhans' cells in the skin increase the phagocytosis and clearance of the bacteria. The heat, redness, and swelling seen in cellulitis are probably due to the small number of residual bacteria and bacterial remnants triggering the inflammatory response and the release of the cytokines and other inflammatory mediators rather than the infection itself.

Clinical features

A patient with cellulitis usually presents with an area of swelling (oedema), erythema, heat, and tenderness (Fig. 3). Most commonly this occurs after minor trauma, such as a laceration or deep abrasion. It is also seen after any break in the skin, such as an area of dermatophyte infection. In many cases there is no obvious break. The red, thickened, oedematous area spreads and is clearly demarcated from the surrounding skin. The patient will have systemic features of inflammation, including malaise, fever, and leucocytosis. If severe, the patient may develop bacteraemia and hypotension.

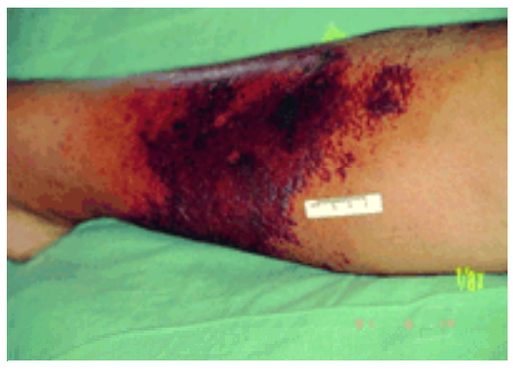


Fig. 3. Advanced cellulitis.

If untreated, lymphangitis develops, with red streaking over the lymphatics draining the area and passing to the local lymph nodes, which also become inflamed (lymphadenitis). Blisters, local abscesses, and areas of local necrosis of the skin may develop in advanced, untreated cases. Cellulitis can predispose to the development of thrombophlebitis, especially in older people.

Predisposition to cellulitis occurs in a number of conditions, including the following.

- In patients with lymphoedema of either the upper or lower limb. The common causes of this would be lymphadenectomy, radiotherapy, or neoplastic involvement of the nodes concerned. This can also involve the vulva in the case of inguinal nodes or the breast and chest wall in axillary nodes.
- Chronic oedema due to venous stasis predisposes to cellulitis, but less frequently.
- Diabetic patients with peripheral neuropathy and ulcer formation in lower limbs are particularly prone to superficial cellulitis around an ulcer (as well as developing deep necrotizing infection).
- Postoperatively, where there may be cellulitis surrounding a surgical wound.
- Orbital cellulitis.

Occasionally, in severe cases of cellulitis, there may be concern that the patient could be suffering from necrotizing fasciitis or gas gangrene (see below).

Erysipelas is uncommon, but when it does occur is more often seen in children and older adults. It is painful and has distinctive, elevated margins that are clearly demarcated from the surrounding skin. It sometimes has a *peau d'orange* appearance (Fig. 4).

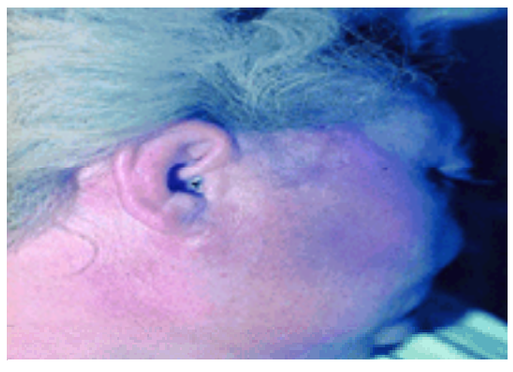


Fig. 4. Erysipelas of the face.

Management

Cellulitis, if recognized and treated early, will often respond to oral antimicrobials and rest in the community setting. Failure of response to these measures or in compromised patients will require intensive, intravenous antimicrobial therapy together with bed rest, elevation of the affected limb, and careful observation in the hospital setting. Any predisposing cause will require prompt and thorough treatment. In particular, fungal infection between the toes will require antifungal therapy.

Wound infections may require drainage and patients with diabetic feet will require full and thorough investigation and management. Patients with leg oedema will require compression stockings.

Most initial therapy is empirical. In the absence of distinctive features outlined in Table 2, the most likely causative agent is either *Strep. pyogenes* or, to lesser extent, *Staph. aureus*, and an antimicrobial regimen should cover both of these bacteria: di(flu)cloxacillin 2 g intravenously (children 25 to 50 mg/kg up to 2 g) every 6 h. In mild cases, oral di(flu)cloxacillin can be used. Dicloxacillin has taken on popularity because flucloxacillin can cause severe hepatitis and cholestatic jaundice, which may be protracted. This reaction is more frequent in older patients and those who take the drug for prolonged periods. However, the intravenous administration of dicloxacillin has been associated with moderate to severe phlebitis, which may require discontinuation of treatment.

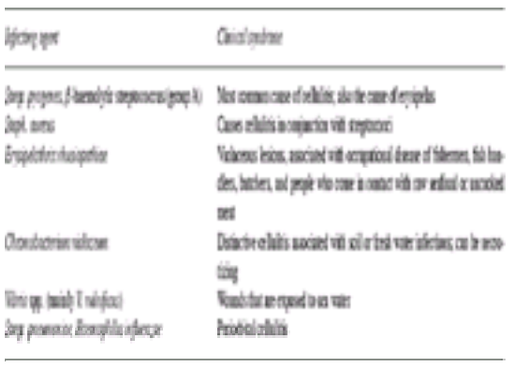


Table 2 Common causes of cellulitis

If the patient is hypersensitive to penicillin, a first-generation cep-halosporin [e.g. cephazolin (cefazolin), cephalexin, or cephalothin] can be used. However, if severe hypersensitivity to penicillin is reported, cross-reaction between b-lactams is possible and then an agent like clindamycin should be considered. Vancomycin can be used if the cellulitis is caused by, or suspected of being caused by, MRSA.

Patients with impaired immunity are susceptible to developing cellulitis from unusual bacteria including *Serratia, Proteus, Strep. pneumoniae* and *Haemophilus influenzae* type b, together with the yeast *Cryptococcus neoformans*.

Failure to respond to intensive antimicrobial therapy, the presence of immunosuppression or of areas of fluctuance indicate the need for needle aspiration to obtain cultures. If the infection and systemic effects are severe or there is any doubt, a full-thickness biopsy to exclude necrotizing fasciitis is indicated

Diabetic pedal wound infections and orbital (pre- and postseptal) cellulitis require urgent surgical intervention together with potent, rapidly acting, broad-spectrum antimicrobial regimens.

Patients with recurrent episodes of cellulitis, especially if associated with lymphoedema, may require prophylactic antimicrobials.

If predisposing conditions such as oedema and fungal foot infections exist, they should be identified and treated concurrently or after the infection has settled (varicose veins/neuropathic ulcers).

Necrotizing bacterial infections

Introduction

Necrotizing bacterial infections are infections of the soft tissues by virulent bacteria that have the ability, usually by the production of toxins, to cause widespread necrosis. The soft tissues can be subcutaneous (e.g. necrotizing fasciitis), muscle (e.g. gas gangrene) and less frequently, skin.

It is apparent that a form of necrotizing infection has been recognized since at least the late eighteenth century, where it was described by a naval surgeon Leonard Gillespie as the ‘malignant ulcer' and was considered to be highly contagious. In the early nineteenth century it was recognized as one of the most dreaded diseases to affect those serving in the navy or the army. It was then known as phagedena (eating away) or hospital gangrene.

Following an outbreak of streptococcal necrotizing fasciitis in the United Kingdom, the disease has been publicized by such sensational names as ‘flesh eating virus, killer bug, or galloping gangrene'. A review of the history is outlined by Loudon.

Classification and nomenclature

Classifying the necrotizing infections of soft tissue is complicated by the fact that the same condition is known by several different names and that some conditions, while usually affecting one layer of soft tissue can, when severe enough, affect all layers. To overcome this confusion an attempt has been made to simplify the classification of necrotizing conditions by defining the layer of primary importance, the clinical syndrome, and the dominant organisms thought responsible. Where two or more names are used for the same condition, they are included in the classification, which is outlined in Table 3.



Table 3 Classification of necrotizing infections according to layer, name organism and clinical syndrome (more common conditions are in bold type)

Necrotizing infections primarily affecting skin

Necrotizing lesions affecting the skin are far less common than those affecting the subcutaneous or muscle layers. There is, in most cases, concurrent necrosis of the subcutaneous layer as well.

Meleney's synergistic gangrene (ulcer)

This condition is also known as progressive bacterial synergistic gangrene.

Predisposing causes, pathogenesis, and aetiological agent

This condition is usually seen following an abdominal operation at the opening of a fistulous tract or around an ileostomy or colostomy. It is occasionally seen adjacent to chronic ulceration on a limb. It is caused by anaerobic or microaerophilic streptococci and *Staph. aureus*(or occasionally by Gram-negative bacilli).

Amoeba can cause a rapid, gangrenous process of the skin of the abdominal wall following surgery. This occurs after operations for conditions involving *Entamoeba histolytica* such as amoebic abscesses or amoebic colitis.

Clinical features

The process commences with an area of redness and swelling around an abdominal wound. It spreads and then slowly ulcerates. A painful ulcer with gangrenous edges develops, and this can grow slowly and relentlessly to enormous size. Occasionally, there is burrowing of infection away from the wound. It differs from necrotizing fasciitis in its slow progress and lack of severe toxicity.

Management

Intravenous antimicrobials in large doses are required, in most cases intravenous benzylpenicillin. Other antimicrobials will depend on the results of culture and susceptibility tests. Most cases do not respond to antimicrobials alone and will require surgical debridement of all necrotic tissue.

Cutaneous involvement in pseudomonal septicaemia

Metastatic skin lesions are seen in the course of severe pseudomonal septicaemia. Necrotic areas appear as indurated, necrotic ulcers. Vesicles and maculopapular lesions also are seen.

Management will consist of antimicrobials in the form of an antipseudomonal b-lactam in combination with an aminoglycoside. The most commonly used b-lactam in febrile neutropenic patients is ceftazidime. The hospital's antibiogram should be consulted if response is slow. Where resistance is a problem a clinical microbiologist and infectious diseases opinion may be required.

Calciphylaxis in chronic renal failure

Patients with endstage renal failure on dialysis not uncommonly develop extensive calcification of subcutaneous small arteries; this causes ischaemic ulcers and eschars of the skin and subcutaneous tissue that become secondarily infected with mixed organisms.

Management of these areas of necrosis is extremely difficult. It involves debridement of the ulcers and antimicrobial therapy. Angiography to exclude a surgically correctable vascular lesion is probably advisable but rarely helpful.

Necrotizing infections primarily affecting subcutaneous tissues

As indicated previously, subcutaneous necrotizing infection can, in severe cases, also affect the cutaneous and muscular layers.

Necrotizing fasciitis

The term necrotizing fasciitis was first used by Wilson in 1952 and it is now a well-recognized condition. It is defined by Bisno thus: ‘Necrotising fasciitis is a deep seated infection of the subcutaneous tissue that results in the progressive destruction fascia and fat.' More recently two distinct types of necrotizing fasciitis have been recognized, based on their different aetiological agents and slightly different clinical presentations.

Necrotising fasciitis type I

Predisposing causes Necrotizing fasciitis type I tends to occur in diabetics, alcoholics, intravenous drug users, immunosuppressed patients, and as a postoperative complication. The portal of entry and aetiology are outlined in Table 4.

Site	Aetiological cause
Limbs	Bite; laceration; intravenous drug use
Abdomen	Abdominal surgery for contamination of peritoneal cavity; incarcerated hernia
Perianal	Ischiorectal abscess; pilonidal abscess
Gynaecological	Bartholin's abscess; vulval abscess in patients with diabetes mellitus
Male genitalia Fournier's gangrene	Urethral trauma; urinary infection
Cervical region	Peritonsillar abscess; dental abscess
Orbit	Periorbital cellulitis; periorbital trauma
Percutaneous instrumentation	Chest tubes; percutaneous endoscopic gastrostomy tubes

Table 4 Portal of entry/aetiology of necrotising fasciitis type I

Pathogenesis and aetiological agent Necrotizing fasciitis type I is a polymicrobial infection caused by a mixture of anaerobes (e.g. *Bacteroides, Peptostreptococcus*), facultative anaerobic cocci (strepto-cocci), and the *Enterobacteriaceae*(e.g. *Escherichia coli, Proteus*, and *Klebsiella*). It can also be caused by marine vibrios (*Vibrio vulnificus*) following a puncture wound by a fish or a laceration in sea water.

Non-clostridial anaerobic cellulitis is a necrotizing soft-tissue infection by anaerobic organisms including *Bacteroides, Peptostreptococcus*, and *Peptococcus*. Although classified as a separate condition from necrotizing fasciitis type I by Bisno and Swartz, its pathogenesis, clinical features, and management make it similar enough to be included in the type I category.

The tissue damage and systemic toxicity are thought to be due to the release of exotoxins from the bacteria and cytokines from leucocytes.

The pathological changes of necrotizing fasciitis include thrombosis of blood vessels, suppuration, and necrosis of the superficial fascia with subcutaneous fat necrosis. There is severe inflammation of the dermis with, initially, no major changes to the epidermis. In most cases, micro-organisms are seen in the tissues. In severe cases there is necrosis of the epidermis and occasionally myonecrosis of underlying skeletal muscle.

Clinical features A patient with necrotizing fasciitis usually presents with the clinical features of cellulitis, including erythema, swelling, and local heat in the affected area. Often the patient is immunosuppressed, an alcoholic, a diabetic, or an intravenous drug user. There may be an association with a contaminated wound or a laparotomy where there was extensive peritoneal soiling. There is pain in the area concerned that is out of proportion to the severity of the cellulitis. Systemic features of toxicity, including fever, tachycardia, and leucocytosis, are also out of proportion to the apparent severity.

If left untreated the skin becomes shiny, hot, and exquisitely tender but discrete margins as occur in Meleney's ulceration do not develop. The skin forms blisters and bullae that contain clear then haemorrhagic fluid. Soft-tissue gas is an uncommon feature of necrotizing fasciitis but is seen when there is anaerobic infection. Finally, as the subcutaneous necrosis continues, with thrombosis of subcutaneous vessels, the overlying skin develops areas of patchy gangrene that can resemble thermal

burns.

During this progression of local changes the patient will become septicaemic with all the features of septic shock including tachycardia, hypotension, confusion, and marked leucocytosis. There could be hypocalcaemia due to fat necrosis and multiorgan failure. If untreated the condition is fatal. It is fatal in 30 per cent of all cases.

The diagnosis is difficult and rests on a high index of suspicion in the clinical settings outlined above. Radiological examination in the form of a plain radiograph or computed tomographic scan may show gas in the tissues. Magnetic resonance imaging can demonstrate soft-tissue fluid and differentiate necrotizing from non-necrotizing cellulitis but its place in the overall management of necrotizing fasciitis has not been confirmed. The diagnostic test is a full-thickness biopsy of the affected area or surgical exploration. If necrotizing fasciitis is present, one would see watery pus (dishwater liquid) coming out of the tissues and there would be easy separation of the skin from the fascia due to extensive necrosis of the subcutaneous tissues. Anatomical pathology shows the features outlined above.

Management Successful management of necrotizing fasciitis includes the following:

- early diagnosis
- surgical debridement
- antimicrobial therapy
- intensive supportive care
- hyperbaric oxygen (possibly).

Early diagnosis with frozen-section tissue biopsy at the bedside and immediate surgical debridement is critical.

Adequate surgical debridement is essential to the successful management of necrotizing fasciitis (Fig. 5 and Fig. 6). This will require radical excision of all necrotic tissue, drainage of involved fascial planes, and extensive fasciotomy. Careful re-evaluation of the wound and formal re-exploration in the operating theatre under general anaesthetic is also required, often on two to three further occasions.

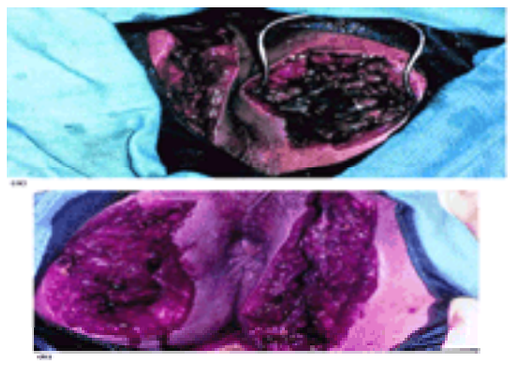


Fig. 5. Necrotizing fasciitis of the buttocks (a) before and (b) after debridement. (By courtesy of Mr John Tracey.)



Fig. 6. Fournier's gangrene of the scrotum after partial debridement. (By courtesy of Mr John Tracey.)

When an extremity is involved, especially in the presence of peripheral vascular disease, amputation may be necessary. When there is extensive perineal necrosis a diverting colostomy may be required.

Broad-spectrum antimicrobial cover will be required in all cases to cover Gram-positives, Gram-negatives, and anaerobes. Current therapeutic guidelines recommend benzylpenicillin, gentamicin, and metronidazole. If *Staph. aureus* is isolated a penicillinase-resistant penicillin (e.g. flucloxacillin or dicloxacillin) or first-generation cephalosporin [e.g. cephalothin, cephalexin, cephazolin (cefazolin)] will be required.

Supportive intensive care will be necessary in all cases. This will entail nutritional feeding (preferably via the enteric route), aggressive fluid resuscitation, and, if required, blood transfusion, possibly endotracheal intubation and ventilation. Some patients will need inotropic support.

Hyperbaric oxygen therapy has many theoretical benefits. However, despite its fairly widespread use for this condition, no randomized, controlled trial has shown its benefit. Certainly if it is to used it must never supplant or take precedence over adequate surgical debridement, antimicrobials, and good supportive care.

Necrotizing fasciitis type II

Necrotizing fasciitis type II is also known as streptococcal gangrene and differs from type I in that it is caused by *Strep. pyogenes*. Although it is seen in settings similar to those of the type I there have been outbreaks of type II in fit, healthy young people with no predisposing illness There is, however, usually a predisposing local lesion such as a varicella infection, an abrasion, laceration, or a surgical procedure. It is also seen after streptococcal throat infections. There are subtypes of *Strep. pyogenes*(M protein types 1, 3, 12 and 28) that are responsible for the highly virulent necrotizing fasciitis type II and streptococcal toxic shock SYndrome.

The infection is often associated with (as in type I) severe systemic symptoms. The early onset of shock and multiorgan failure associated with *Strep. pyogenes* is known as the streptococcal toxic shock syndrome. Exotoxins acting as superantigens that trigger non-specifically the massive release of cytokines are probably responsible for the toxic systemic symptoms (the toxic shock can also occur of without the presence of necrotizing fasciitis).

Clinical features are almost identical to those seen in necrotizing fasciitis type I except that cutaneous involvement occurs earlier in the course of the disease. A full-thickness burn-like eschar rapidly develops over the necrotic fascia.

Management Management principles are very similar to those for necrotizing fasciitis type I and include early diagnosis, surgical debridement, antimicrobial therapy, intensive supportive care, and possibly hyperbaric oxygen. Once it is confirmed that that the infection is caused by streptococci, high-dose penicillin together with clindamycin are the antimicrobials of choice. Before the bacteria and susceptibilities have been identified, broad-spectrum antimicrobial therapy as in type I would be indicated.

Penicillin failures have occurred; these may well be due the size of the bacterial load. In large numbers bacteria reach stationary phase more rapidly and attain a physiological state with a reduced ability to bind penicillin.

Clindamycin is effective because it:

1. is not affected by inoculum size;
2. is a potent suppressor of bacterial toxin synthesis;
3. enhances phagocytosis by inhibiting M-protein synthesis;
4. suppresses synthesis of penicillin-binding proteins that assist in cell-wall SYnthesis;
5. has a longer postantibiotic effect than penicillin;
6. suppresses lipopolysaccharide-induced SYnthesis of tumour necrosis factor by monocytes.

Pedal necrosis in patients with diabetes mellitus

Necrosis and deep infection in the feet of patients with diabetes mellitus is well described as a separate entity Essentially the underlying pathology is a combination of factors seen in diabetes including a peripheral neuropathy, ischaemia due to micro- and macrovascular disease, decreased resistance to infection, and poor wound healing. When limb threatening the infection is usually polymicrobial including streptococci, staphylococci, enterococci, and *Bacteroides*.

Clinical features usually include an infected ulcer followed by deep infection, which causes a hot, swollen foot. Surgical exploration reveals extensive infection, pus formation, and necrosis of the soft tissues. Severe systemic toxicity is not a usual feature.

Management principles are similar to those outlined for necrotizing fasciitis type I:

- early diagnosis
- surgical debridement
- antimicrobial therapy
- hyperbaric oxygen (possibly).

Clostridial anaerobic cellulitis

When dirty, contaminated, devitalized subcutaneous wounds are infected with *Clostridium perfringens* or *Cl. septicum*, a form of necrotizing infection known as clostridial anaerobic cellulitis develops. There is usually gas in the tissues. *Clostridium* spp., instead of infecting the muscle as in classical gas gangrene, affects devitalized subcutaneous tissues. It is a common condition in battlefield injuries. The rapid development of toxic features and pain, and the synergistic gangrene of skin previously described for necrotizing fasciitis type I, are not as prominent in this condition. Because of these differences it is described as a different condition to necrotizing fasciitis type I by Bisno and Swartz.

Clostridial cellulitis has the same management principles as necrotizing fasciitis type I, with particular emphasis on the antimicrobial therapy for *Clostridium* spp., which includes large doses of intravenous penicillin plus clindamycin.

Necrotizing infection of muscle

Infections primarily affecting skeletal muscle include the following.

1. Clostridial gas gangrene
2. Non-clostridial myonecrosis:
   i. anaerobic streptococcal myonecrosis,
   ii. *Strep. pyogenes* necrotizing myositis,
   iii. *Aeromonas hydrophila* myonecrosis,
   iv. infected vascular gangrene.

Clostridial gas gangrene

Gas gangrene is a rapidly progressive, life-threatening infection of skeletal muscle by *Clostridium* spp. (usually *Cl. perfringens*).

Predisposing causes, pathogenesis, and aetiological agent

Gas gangrene most commonly occurs in battlefield injuries and in contaminated civilian wounds such as compound fractures. The key feature in its pathogenesis is the presence of devitalized tissue and foreign material in traumatized wounds with soil or faecal contamination. It occasionally occurs after bowel or biliary-tract surgery and arterial insufficiency in the limbs.

The most common infecting organism is *Cl. perfringens. Clostridium septicum* is also associated with gas gangrene (gastrointestinal neoplasm-associated). The organisms secrete strong toxins that cause rapid and extensive tissue necrosis with few inflammatory cells being present. They release collagenase and hyaluronidase, which break down tissue planes. *Clostridium perfringens* secretes a at least 12 exotoxins: the most important is a-toxin, which degrades lecithin, a major component of cell membranes. Thus there is destruction of muscle cells, platelets, and red cells. a-Toxin can also damage nerve sheathes. q-Toxin is also produced by the bacteria and causes lysis of leucocytes, explaining their absence in the affected tissue. Because of the rapid multiplication of bacteria and the production of such toxins there is rapid spread, necrosis of the muscle, and severe toxaemia.

There is a rare form of gas gangrene spread by the bacteraemic route. It is often associated with pathology of the intestinal tract such as colon cancer and is designated spontaneous, non-traumatic gas gangrene. It is often caused by *Cl. septicum*.

The histopathological features of gas gangrene include coagulative necrosis, cavities in the tissues due to gas, haemolysis, extensive vascular injury, and thrombosis. Numerous Gram-positive bacilli are seen in the tissues with a paucity of neutrophils.

Clinical features

The incubation period is usually 2 to 3 days but can be even more rapid than this. Excruciating pain, out of proportion to the original injury, is one of the earliest symptoms. Local features include the following.

- Tense oedema, local tenderness, a change in the colour of the skin, which is initially pale but as the infection progresses becomes a yellowish bronze colour and develops large bullous vesicles and eventually areas of green to black cutaneous gangrene.
- If there is an open wound or if exploration is undertaken, a dirty serosanguinous discharge is seen issuing from the wound that has a peculiar foul smell. The muscle appears pale and swollen and bulges through the wound. It fails to contract on stimulation or bleed when cut. Later it becomes a greenish-purple colour and friable.
- Gas bubbles may be seen in the discharge and the wound is crepitus to palpation.

These changes can occur very rapidly over a period of 4 to 12 h.

Systemic features include the following:

- The patient appears unwell, pale and sweaty
- Tachycardia
- Shock with hypotension
- Renal failure
- Delirium, drowsiness, and later, coma
- Fever

- Jaundice.

Investigations

- Leucocytosis in a full blood examination
- Gram's stain of the wound or the exudate reveals a large quantity of Gram-positive bacilli, with few leucocytes
- Growth of *Cl. perfringens* on culture
- Gas formation in liquid anaerobic culture
- Radiographs of affected areas show extensive gas in the tissues.

Management

The management principles of gas gangrene are very similar to those for necrotizing fasciitis. Gas gangrene is a rapidly progressive fatal disease and requires immediate diagnosis based on the clinical features outlined above. Urgent surgical exploration of all suspicious wounds is the most important aspect of intervention. Once gas gangrene has been confirmed, radical surgical excision of all necrotic tissue with concurrent fasciotomies is essential. Occasionally, amputation may be necessary.

Antimicrobial therapy in the form of intravenous benzylpenicillin, one to two million units every 2 h, in combination with clindamycin is recommended. If the patient is hypersensitive to penicillin, metronidazole can be used.

Patients who have gas gangrene usually have severe systemic toxaemia and require intensive-care support. This will entail close monitoring of all body functions including pulse, blood pressure, temperature, urine output, and possibly central venous pressure. Aggressive fluid and electrolyte replacement will be required in all patients and blood transfusion in some. When more toxic and shocked, some patients will require endotracheal intubation and assisted ventilation, and in some cases inotropic support. All will require nutritional support, if possible by the enteric route.

Hyperbaric oxygen therapy is used to help control the anaerobic infection and may reduce the extent of gangrene. However, as with necrotizing fasciitis, its use is supplementary to surgical excision of dead muscle and antimicrobial therapy and should not take precedence over these.

Prognosis

The prognosis will depend on the speed and effectiveness of the surgical treatment. If untreated it is uniformly fatal. The mortality in treated cases ranges from 25 to 40 per cent.

Non-clostridial myonecrosis

Anaerobic streptococcal myonecrosis resembles clostridial gas gangrene but is not as rapid in progression. It is caused by an infection of muscle by a mixed group of bacteria, the most predominant of which are the anaerobic streptococci. There usually is a growth of *Staph. aureus* and *Strep. pyogenes* as well. If not treated it progresses to gangrene, toxaemia and shock. Treatment involves surgical debridement, antimicrobials and supportive care.

Strep. pyogenes necrotizing myositis

Streptococcus pyogenes necrotizing myositis is a rare cause of myonecrosis. It usually occurs as a spontaneous streptococcal infection associated with the streptococcal toxic shock syndrome. The clinical features are those of intense pain, swelling, tenderness over the affected muscle, and signs of general toxaemia, in many ways similar to necrotizing fasciitis. The management entails expeditious surgical exploration and debridement, together with antimicrobial therapy.

Aeromonas hydrophila myonecrosis

Aeromonas hydrophila myonecrosis is caused by a Gram-negative bacillus found in a fresh-water environment and is often associated with fish. The condition is usually caused by infection with the organism *A. hydrophila* via a penetrating injury. There is a rapid myonecrosis with many of the clinical features of gas gangrene. Treatment consists of early and extensive surgical debridement along with antimicrobials. *Aeromonas* is susceptible to the aminoglycosides, quaternary fluoroquinolones, and third-generation cephalosporins (although powerful b-lactamases should be sought to confirm sus-ceptibility).

Infected vascular gangrene

Infected vascular gangrene occurs in a limb where there is gangrene caused by arterial insufficiency. There is usually a mixed infection with a *Bacteroides, Proteus*, and anaerobic streptococci. Gas formation and suppuration occur. The infection does not usually spread beyond the extent of the ischaemic muscle to involve the vascularized tissue. Treatment involves amputation of the gangrenous limb under broad-spectrum antimicrobial cover.

Further reading

Abraham N, Doudle M, Carson P. Open versus closed treatment of abscesses: a controlled clinical trial. *Australia and New Zealand Journal of Surgery* 1997; **67**: 173–6. [Primary closure of acute superficial abscesses was associated with an improved outcome.]

Berger L, Osborne D. Treatment pyogenic liver abscesses by percutaneous needle aspiration. *Lancet* 1982; **i**: 132–4. [A report on the effectiveness of aspiration of liver abscess with antibiotic cover.]

Bisno AL, Stevens DL. Streptococcal infections of skin and soft tissues. *New England Journal of Medicine* 1996; **334**: 240–5. [A detailed review of streptococcal infection of skin and soft tissue from cellulitis to necrotizing fasciitis (NF).]

Burge TS, Watson JD. Necrotising fasciitis: be bloody bold and resolute. *British Medical Journal* 1994; **308**: 1453–4. [Definitive, radical excision at the initial operation is essential in the treatment of NF.]

Caputo GM *et al.* Assessment and management of foot disease in patients with diabetes. *New England Journal of Medicine* 1994; **331**: 854–60. [A comprehensive review of diabetic foot disease.]

Chelsom J, Halstensen A, Haga T, Hoiby E. Necrotising fasciitis due to Group A streptococci in western Norway: incidence and clinical features. *Lancet* 1994; **344**: 1111–15. [A detailed report on the case histories of 13 patients with necrotizing fasciitis from Norway.]

Gemmell CG *et al.* Potentiation of opsonization and phagocytosis of *Streptococcus pyogenes* following growth in the presence of clindamycin. *Journal of Clinical Investigation* 1981; **67**: 1249–56. [Clindamycin is effective in the treatment of Type II necrotizing fasciitis because it enhances phagocytosis by inhibiting M protein SYnthesis.]

Green RJ, Dafoe DC, Raffin TA. Necrotising fasciitis. *Chest* 1996; **110**: 219–29. [A review of the history, aetiology, pathophysiology, clinical features and management of necrotizing fasciitis.]

Lewis RT. Soft tissue infections. *World Journal of Surgery* 1998; **22**: 146–51. [An overview of soft tissue infections with a classification recommending: focal v. diffuse and necrotizing v. non-necrotizing.]

Louden I. Necrotising fasciitis, hospital gangrene, and phagedena. *Lancet* 1994; **344**: 1416–18. [An historical review of necrotizing fasciitis dating back to the 18th and early 19th centuries.]

McHenry CR *et al.* Determinants of mortality for necrotising soft tissue infections. *Annals of Surgery* 1995; **221**: 558–65. [Early diagnosis and radical surgical debridement are critical in the management of necrotizing fasciitis.]

Maguire GP, Arthur AD, Boustead PJ, Dwyer B, Currie B. Emerging epidemic of community acquired methicillin-resistant *Staphylococcus aureus* infection in the Northern Territory. *Medical Journal of Australia* 1996; **164**: 721–3. [A report of community acquired methicillin resistant *Staphylococcus aureus* seen in the necrotizing fasciitis of Australia.]

Rahmouni A *et al.* MR imaging in acute infectious cellulitis. *Radiology* 1994; **192**: 493–6. [A study of the role of MRI in the diagnosis of necrotizing fasciitis.]

Robbins SL, Kumar V, Cotran R. *Pathological basis of disease*, 5th edn. Saunders, Philadelphia, 1994. [A major pathology textbook with excellent chapters on the inflammatory response, the immune response, and infection.]

Stamenkovic I, Lew PD. Early recognition of potentially fatal necrotising fasciitis: the use of frozen section biopsy. *New England Journal of Medicine* 1984; **310**: 1689–93. [The benefit of early diagnosis of necrotizing fasciitis by frozen section biopsy.]

Stevens DL, Bryant AE, Hackett SP. Antibiotic effects on bacterial viability, toxin production and host response. *Clinics in Infectious Diseases* 1995; **20**(Suppl. 2): S154–7. [Clindamycin is effective in the treatment of necrotizing fasciitis because of suppression of lipopolysaccharide induced monocyte synthesis of TNF.]

Stevens DL, Maier KA, Mitten JE. Effect of antibiotics on toxin production and viability of *Clostridium perfringens. Antimicrobial Agents and Chemotherapy* 1987; **31**: 213–18. [Toxin suppression and rapid bacterial killing may explain the superiority of clindamycin, rifampicin, and metronidazole over penicillin in experimental gas gangrene.]

Stevens DL, Yan S, Bryant AE. Penicillin-binding protein expression at different growth stages determines penicillin efficacy *in vitro* and *in vivo*: an explanation for the inoculum effect. *Journal of Infectious Diseases* 1993; **167**: 1401–5. [Clindamycin may be more effective in treating Type II necrotizing fasciitis than penicillin because it is not effected by the inoculum effect.]

Swartz MN. Skin and soft tissue infections. In: Mandell GL, Bennett JE, Dolin R, eds. *Mandell, Douglas and Bennett's principles and practice of infectious diseases*, pp. 909–44. Churchill Livingstone, New York, 1995. [An excellent review of skin soft tissue infections in a major reference book on infectious disease.]

Wilson B. Necrotising fasciitis. *American Surgeon* 1952; **18**: 416–31. [The first use of the term ‘necrotizing fasciitis' to describe the condition.]

Yan S, Bohach GA, Stevens DL. Persistent acylation of high-molecular weight penicillin-binding proteins by penicillin induces the post-antibiotic effect in *Streptococcus pyogenes. Journal of Infectious Diseases* 1994; **170**: 609–14. [Clindamycin is effective in treating type II necrotizing fasciitis because it suppresses SYnthesis of penicillin bind proteins which assist in cell wall SYnthesis.]

47.1.2 Typhoid fever and other salmonella infections

Christopher S. Grant, William J. K. Huizinga and A. S. Daar

Introduction
Typhoid fever
Pathogenesis
Immunity
Clinical features
Complications
Diagnosis
Medical treatment
Management of intestinal haemorrhage
Management of typhoid perforation
Mortality
Convalescent and chronic typhoid carriers
Prevention
Paratyphoid fever
Non-typhoidal salmonella infections
Pathogenesis
Clinical features
Further reading

Introduction

Salmonella infections occur worldwide and the practising surgeon should be familiar with their protean manifestations. The surgical complications are infrequent but require prompt diagnosis and treatment. Human infection usually presents in two distinctive clinical forms: typhoid and paratyphoid fevers (sometimes collectively termed enteric fever), and non-typhoidal salmonella infection (enterocolitis) (Table 1).

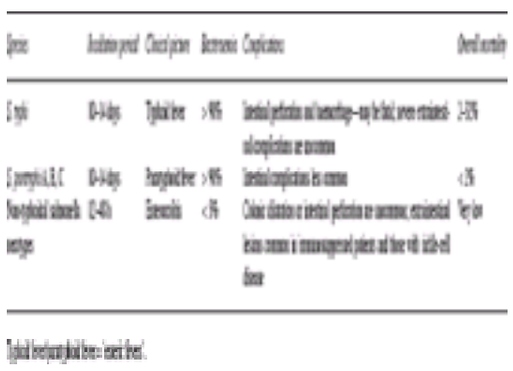


Table 1 Clinicopathological features of salmonella infections

Typhoid fever

Typhoid fever is an important health problem in many parts of the world, particularly in developing countries. An estimated 33 million cases occur each year. The infection is caused by *Salmonella typhi*, a strict human pathogen that is invariably acquired by ingestion of food or water contaminated with excreta from a patient with typhoid or from a convalescing or chronic carrier. *Salmonella typhi*, unlike other salmonella species, possesses a polysaccharide **Vi**(virulence) surface antigen, which enhances its invasiveness. The disease remains endemic because of inadequate sewage disposal and lack of safe water supplies.

Pathogenesis

Some features of the pathogenesis have not been completely elucidated. Essentially, after ingestion and passage through the stomach, the bacteria invade and replicate in the epithelial cells of the terminal ileum. Direct invasion through the M (microfold) cells of Peyer's patches is thought to be more important. The bacteria pass through the epithelial cells and are engulfed by macrophages. Many of them are able to survive and multiply within the macrophages. It is probable that they are transported within the macrophages to the mesenteric lymph nodes, and spread via the lymphatics (through the thoracic duct) and bloodstream to the reticuloendothelial system in the liver, spleen and bone marrow, where continued multiplication takes place during an incubation period of about 10 to 14 days (Fig. 1).

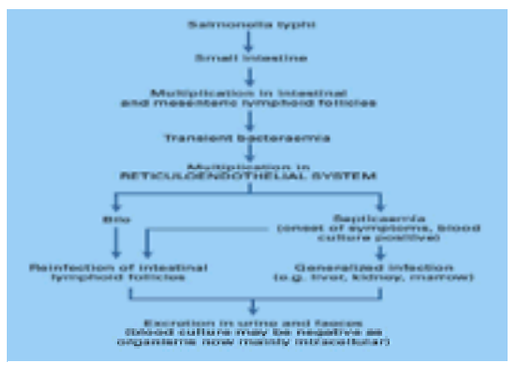


Fig. 1. Pathogenesis of typhoid fever.

Re-entry of bacteria into the bloodstream marks the onset of clinical typhoid fever. During the second or third week of the illness, heavy reinfection of the gut occurs through bile or bacteraemic spread. The bacteria localize in the Peyer's patches in the lower ileum, although the lymphoid tissue in the jejunum, caecum, appendix, and ascending colon may sometimes be affected.

Peyer's patches and solitary lymphoid follicles become oedematous and hyperplastic, and mesenteric lymphadenitis also occurs. The main inflammatory cells are macrophages and lymphocytes; polymorphonuclear leucocytes are virtually absent. Necrosis of Peyer's patches with ulceration of the intestinal mucosa occurs in the long axis of the bowel. This usually heals with minimal fibrosis; otherwise, intestinal haemorrhage from the necrotic patch or perforation of the bowel with resultant peritonitis may occur. These intestinal complications generally arise in the second week of the illness.

Inflammatory lesions are widespread: the liver, spleen, and bone marrow are affected. Other extraintestinal organs, including the kidneys, lungs, skin, brain, bone, joints, myocardium, and heart valves, are often involved. Osteomyelitis may occur in immunosuppressed patients and in those with sickle-cell disease.

Although *S. typhi* is excreted in the bile, acute cholecystitis is rare. However, chronic typhoid cholecystitis often develops after *S. typhi* infection; while this is often SYmptomless, it may be associated with the formation of gallstones. Chronic typhoid carriers excrete large numbers of bacilli in the faeces and are the major reservoir

for spreading the infection.

The pathogenesis of the prolonged fever and toxic symptoms seen in typhoid fever is only partially understood: it is probably related to circulating endotoxins and cytokines such as interleukin-1 and tumour necrosis factor released as a result of the prolonged and diffuse inflammation in the reticuloendothelial system and elsewhere.

Immunity

The nature of immune response to salmonella infections appears complex. Experimental work has shown that both the humoral and cell-mediated immunity are important in resistance. The role of the two forms of immunity varies with the stage of invasion and infection. Cell-mediated immune responses are thought to be more important, especially in the recovery process. Patients immunocompromised by chronic diseases or malnutrition are liable to suffer more complications of typhoid.

Clinical features

Typhoid fever is generally regarded as a disease of older children and young adults. However, recent studies suggest that infection in infants is also common and, for reasons that are obscure, may not present clinically as typhoid fever. The incubation period is about 10 to 14 days after ingestion of the organism. The onset is insidious, with fever, headache, abdominal discomfort or pain, cough, malaise, and anorexia. The fever characteristically increases in a stepwise fashion over many days, but this pattern is not always present. Constipation is common in the initial stages, although pea-soup diarrhoea may occur after the first week. In severe cases, mental changes are often a prominent feature and include clouding of consciousness (*typhos* in Greek means smoke or stupor), coma vigil (in which the patient, muttering, is semiconscious with eyes open), and other psychotic features that could lead to admission to a psychiatric unit. Meningism may be present but meningitis is rare.

Rose spots (a maculopapular rash) may appear, typically on the upper abdomen and lower chest, in the second or third week and fade away after 2 or 3 days. These spots occur in only one-third of patients and are rarely seen in dark-skinned people. Abdominal tenderness and distension are common as the disease progresses. The spleen and the liver are palpable in about 25 per cent of adults and rather more frequently in children. Relative bradycardia occurs in less than 50 per cent of patients.

The likelihood of leucopenia has been exaggerated. More often there is an absence of leucocytosis, with the mean white blood-cell count in the region of 6000 to 7000. Anaemia is due to haemorrhage, marrow suppression (in long-standing infection), and rarely to haemolysis. Thrombocytopenia may occur.

If untreated, the illness lasts for 3 to 4 weeks, and the fever remits in a stepwise fashion. Major sequelae arise through infection of other organs, or more seriously, as a result of intestinal complications, which may be lethal. Differential diagnosis in the early stages includes malaria, brucellosis, amoebic liver abscess, and typhus. When bloody diarrhoea occurs, the clinical picture may resemble bacillary or amoebic dysentery.

Complications

Although typhoid fever affects many organ systems, the most serious complications (Table 2), which are haemorrhage and perforation, occur in the gastrointestinal tract and are responsible for most of the fatalities.

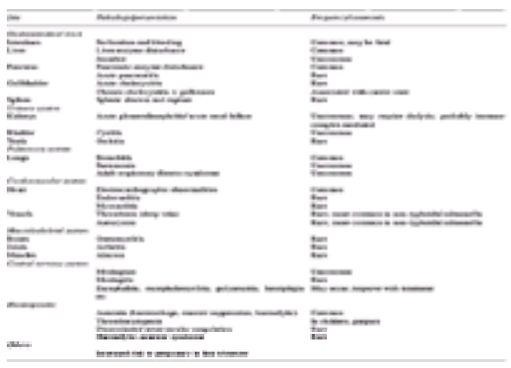


Table 2 Complications associated with typhoid fever

Intestinal haemorrhage

In one study, gross haemorrhage occurred in 0.8 per cent of hospital inpatients, although others have reported an incidence of up to 5 per cent. Haemorrhage usually follows sloughing of typhoid ulcers in the ileum, or less frequently in the caecum and ascending colon, during the second or third week of the disease. It is often insidious and presents as increasing anaemia and tachycardia. Rarely, erosion of large vessels may produce acute, massive, lower gastrointestinal bleeding. The onset of bleeding should alert the physician to the possibility of an imminent intestinal perforation as the two sometimes occur together.

Intestinal perforation

Perforation (Fig. 2) occurs in approx. 5 per cent. However, there are geographical variations in the reported incidence of typhoid perforations (range 0–39 per cent). Perforation typically occurs during the second or third week of the disease and is more common in males than females (4:1 in one large Bangladeshi series and 2.3:1 in a Nigerian series), and in adults than children. There is no clear explanation for these observations. Acute perforation is signalled by sudden or increasing abdominal pain, often beginning in the lower abdomen and spreading. The pain is associated with tenderness, rigidity, guarding and loss of liver dullness, and there is tachycardia and hypotension. Abdominal distension is frequently present, but about one-third of patients may have a scaphoid abdomen in spite of gross peritoneal contamination. These signs, however, may be masked in severely toxic patients, in whom radiographs and repeated abdominal examinations are needed to avoid missing the diagnosis.

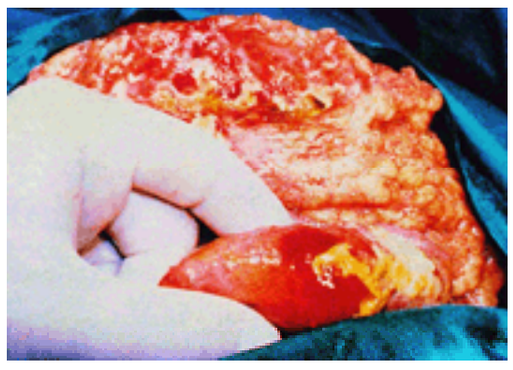


Fig. 2. A solitary perforation on the antimesenteric border of the ileum in an Omani patient with typhoid fever.

Other complications

Many of the other complications are also of surgical interest (Table 2). About 50 per cent of patients show impairment of liver function: this reverts to normal with recovery. Clinical jaundice due to intrahepatic cholestasis is more frequently observed in children. Pancreatic enzyme disturbances are reportedly common, but acute

pancreatitis is rare.

Pneumonia is a serious complication occurring frequently in children. Altered mental status, in the form of confusion, delirium or stupor, is often seen. Seizures are frequent in young children. Many other abnormalities affecting the central nervous SYstem, all relatively rare, have been reported. Acute nephritis and cystitis may occur and acute renal failure, which may require dialysis, is a rare complication. Electrocardiographic abnormalities are sometimes observed; toxic myocarditis and rarely endocarditis are seen more often in children. Deep venous thrombosis occurs in a minority of patients.

Other complications that have become rare since the introduction of antibiotics include acute acalculous cholecystitis, osteomyelitis, arthritis, splenic abscess, and purpura. Most of these extraintestinal complications of typhoid fever are reportedly more frequent in patients who are immunocompromised or suffer from chronic disease or malnutrition. Prolonged (up to 2 years) salmonella septicaemia may occur in patients who have schistosomiasis, as the bacteria grow in the gut and on the skin of the schistosome worm. Adequate management should include specific treatment against both the salmonella and the parasite.

Diagnosis

The diagnosis is often strongly suggested by the clinical presentation and a history of residence or recent visit to an endemic area. In endemic areas an experienced clinician can make the correct diagnosis solely on clinical grounds in up to 80 per cent of patients; a definitive diagnosis of typhoid fever is made by isolation of *S. typhi*, usually from the blood. A recent research report shows that a polymerase chain reaction–microtitre-plate hybridization technique is a useful and rapid diagnostic method where standard culture assays are negative.

Blood cultures are positive in about 80 per cent of patients during the first 2 weeks of the illness; stool, and sometimes urine, cultures become positive, usually from the second week in patients who have not been treated. By the third week, blood cultures often become negative, as the organisms are now mainly intracellular.
Bone-marrow culture is the single, most effective method for the isolation of *S. typhi*. It is positive in 90 per cent of patients, is not affected by recent partial antibiotic therapy, and often remains positive even when the blood cultures have become negative due to antibiotic therapy.

A rising titre of agglutinin [the Widal test for antibodies against the flagella (H) and somatic (O) antigen] over a week helps in making the diagnosis, especially when cultures are negative. However, the test is not completely specific. Newer serodiagnostic tests include an indirect immunofluorescent antibody test for the Vi antigen and detection of IgM antibody to *S. typhi* lipopolysaccharide antigen by enzyme-linked immunosorbent assay. Results can be obtained more rapidly than from bacterial culture and remain positive after antibiotic treatment has been started.

Leucopenia is not generally useful in diagnosis as it occurs in only 12 per cent of patients; an absence of leucocytosis is more common. The presence of both polychromasia and reticulocytosis should alert the physician to intestinal haemorrhage. The diagnosis of intestinal perforation is straightforward in the presence of typical history of typhoid and classical signs of perforation. An upright chest radiograph, when available, demonstrates free subdiaphragmatic gas in up to 75 per cent of patients. The major diagnostic difficulty is in determining whether patients known or considered to have typhoid who have equivocal abdominal signs and no free subdiaphragmatic gas on radiographic examination have suffered a perforation. An abdominal paracentesis is helpful when positive, but a negative tap does not exclude perforation. Repeated abdominal examination may be helpful but if doubt persists laparotomy should not be delayed.

Medical treatment

Antimicrobial therapy (Table 3) gives excellent results in uncomplicated typhoid. Chloramphenicol is no longer considered the drug of choice for typhoid fever because of the emergence of plasmid-mediated multidrug resistance. Since 1989, strains of *S. typhi* resistant to chloramphenicol, ampicillin, and trimethoprim (i.e., multidrug-resistant strains) have caused several outbreaks of typhoid fever in many developing countries, especially in the Indian subcontinent, South-East Asia, and Africa. In a recent report from India, multidrug resistance reached a prevalence of up to 78 per cent of isolates. Multidrug-resistant strains have also been isolated with increasing frequency in developed countries from returning travellers.

Drug	*Dose*	*Duration*
Drug of choice		
Ciprofloxacin	*Adults:* 500 mg twice daily, orally 200–400 mg twice daily, intravenously *Children:* 7.5 mg/kg twice daily, orally 5 mg/kg twice daily, intravenously	14 days
Alternatives		
Ceftriaxone	*Adults:* 2–4 g daily, intravenously *Children:* 100 mg/kg daily, intravenously	14 days
Chloramphenicol	*Adults:* 500 mg 4-hourly, orally or intravenously till fever subsides, then 500 mg 6 hourly *Children:* 50 mg/kg daily, orally in divided doses	14 days
Amoxycillin (amoxicillin)	*Adults:* 1 g 8-hourly, orally Children: 100 g/kg per day, orally in divided doses	14 days
Co-trimoxazole	*Adults:* 160 mg trimethoprim + 800 mg sulphamethoxazole twice daily, orally *Children:* 6 mg trimethoprim + 30 mg sulphamethoxazole/kg per day in two divided doses, orally	14 days

Table 3 Antibiotic treatment of typhoid fever

As a result of the increasing world-wide spread of multidrug-resistant strains, ciprofloxacin, a quinolone, is now considered the drug of choice for typhoid fever, where it is affordable. The newer quinolones, ofloxacin and norfloxacin, are reported to be equally efficacious. The oral route is satisfactory for most patients; intravenous therapy is reserved for the seriously ill patients. Most isolates remain sensitive to the quinolones but strains with chromosomally encoded resistance are emerging. The alternative to the quinolones is to use one of the third-generation cephalosporins, such as ceftriaxone. Chloramphenicol, with its well-known small risk of marrow toxicity, is still widely used in developing countries for susceptible strains because it is effective and inexpensive. Amoxycillin (amoxicillin) or co-trimoxazole are not recommended as a first-line drugs except for patients with known susceptible strains

Once therapy is begun there is a rapid improvement in the patient's general condition, but complete defervescence may take up to 7 days. A very rapid return to normal temperature may cast doubt on the diagnosis of typhoid fever. The minimum recommended length of treatment should be 14 days; some have tried shorter treatment periods but there is then the danger of higher relapse rates. Patients should not be discharged home until they have had three negative stool and urine cultures after the end of drug treatment.

The new quinolone drugs should be used with caution in children and pregnant women as they can cause irreversible damage to cartilage in the strained joints of young experimental animals. However, serious arthropathy has not been reported in children following the extensive use of ciprofloxacin.. Reports suggest that, in addition to the antibiotics, high doses of dexamethasone are valuable in severely ill patients with an altered state of consciousness and shock. An initial intravenous dose of 3 mg/kg over 30 min is followed by eight doses of 1 mg/kg 6-hourly.

Management of intestinal haemorrhage

The treatment is generally non-surgical or expectant, with blood transfusion to correct anaemia, which develops over a few days. Acute haemorrhage should be treated promptly with blood transfusion. If the bleeding is massive and persistent, laparotomy is indicated as an erosion of a large vessel has probably occurred. The affected bowel is often easy to identify (for example, the bowel wall at the site of the ulcers is thin) and to resect. Where facilities exist, mesenteric angiography and embolization of the bleeder may be a suitable alternative.

Management of typhoid perforation

Most authors currently advocate operative rather than non-operative treatment, which in many past series was associated with mortality rates of over 60 per cent. Furthermore, operative treatment is preferred for two reasons. First, typhoid perforation produces fulminating peritonitis and, unlike other perforations, rarely seals spontaneously as the omentum seldom migrates to the area of perforation. Secondly, the clinical diagnosis can be wrong, and peritonitis may be due to a perforated appendix or peptic ulcer.

Patients are usually critically ill, with septicaemia, generalized peritonitis, dehydration, and electrolyte imbalance, especially potassium deficiency. They are best resuscitated in an intensive care unit if the facilities exist. Dehydration and electrolyte imbalance should be treated aggressively to restore normal haemodynamics and

urine output (at least 30 ml/h) within a few hours. Blood transfusion is often required to correct severe anaemia.

Parenteral antibiotic therapy is aimed at *S. typhi* and the common enteric organisms, including anaerobes. A useful regimen is ciprofloxacin and metronidazole or chloramphenicol and metronidazole, but the local antibiotic-resistance pattern must be taken into account. Although steroids have been shown to be useful in the treatment of typhoid fever complicated by severe toxaemia, altered mental status and shock, their role in typhoid perforation has not been evaluated.

Surgical management

At laparotomy, perforation is found on the antimesenteric border of the ileum and it is single in 80 per cent of patients. Two perforations are found in 15 per cent and more than two in 5 per cent. About 90 per cent of ileal perforations are located within 60 cm of the ileocaecal valve and caecal perforations occur in about 2 per cent of patients. Perforations other than ileal and caecal are rare.

The operative management remains controversial (Table 4). The ideal operation would be simple, effective, and tailored to the findings. The procedure of choice is debridement of the margins of the perforation and meticulous closure in two layers of absorbable suture in a transverse direction to avoid narrowing the bowel. Copious peritoneal lavage is required. This approach is within the competence of a district surgeon operating alone in an endemic area. However, when there are more than three perforations that are close together, it is best to resect the affected bowel and perform a primary end-to-end anastomosis. Any areas of apparent impending perforation, if not included in a resection, must be oversewn. Right hemicolectomy is undertaken only for caecal perforations.

[illegible]	[illegible]
Simple closure	Simple but high leak rate in some series
Debridement/wedge excision + simple closure	Simple and effective operation. Not recommended if more than three ileal perforations close together
Ileal resection + primary anastomosis	For multiple ileal perforations
Right hemicolectomy	Only for caecal perforation
Simple closure or ileal resection + end-to-side ileotransverse colostomy	Extensive operation, has decreased morbidity but not mortality
Inversion of perforated ileum	Is extremely critical in moribund patients
Simple peritoneal drainage	Is extremely critical in moribund patients
Oversewing	For areas of impending perforation

Table 4 Operative procedures in typhoid perforation

Following peritoneal lavage, the abdominal wound is closed, usually without drains. If there is gross faecal contamination, the skin wound may be left open to minimize the risk of wound infection. Antityphoid chemotherapy should be continued for at least 14 days.

Postoperative complications

Wound infection is the most common problem, and is often severe. Other serious common complications include respiratory infection, reperforation, peritonitis, faecal fistula, intraperitoneal abscesses, and wound dehiscence. Long-term complications include incisional hernia and adhesive intestinal obstruction.

Mortality

Typhoid fever is a severe disease whose major complications are potentially lethal. Patients at the greatest risk of death are those with severe toxaemia, an altered consciousness, shock, and intestinal perforation. Although the case fatality rate in countries with well-developed health services is less than 2 per cent, in many developing countries, despite the general availability of drugs effective against *S. typhi*, a mortality rate of 4 to 32 per cent is reported.

Intestinal perforation is the leading cause of death. The mortality rates in reported series of typhoid perforation vary widely. In a collective review of reported series from developing countries, the median of the case fatality rates was 43 per cent. Two factors have an important bearing on the prognosis: duration of illness before perforation, and interval between the time of perforation and surgery. Perforations occurring after the third week of the illness are associated with a higher mortality rate than those occurring in the first week. For patients undergoing surgery, the best results are obtained when operations are performed promptly, that is, within 24 h of perforation. When surgery is delayed for up to 48 h after perforation the case fatality rate increases to above 50 per cent; even worse rates are observed when the delay is more than 72 h. The number of perforations does not appear to have a consistent bearing on the postoperative prognosis. Mortality from perforation could therefore be minimized by early presentation, early diagnosis of perforation, improved critical care, and prompt surgical treatment. Unfortunately, in the endemic countries the medical resources required to meet these clinical challenges are scarce.

Convalescent and chronic typhoid carriers

Typhoid patients who have been inadequately treated may become convalescent carriers. Ciprofloxacin has been shown to reduce the convalescent carrier rate. A chronic carrier has usually had a previous subclinical infection and usually excretes *S. typhi* in the stools, but sometimes in the urine, over a long period without any symptoms. The gallbladder or biliary tract is the primary site of chronic carriage, although the infection may also persist in an intrahepatic nidus or in the urinary tract, especially if there is a urinary-tract abnormality. In countries with endemic urinary schistosomiasis, urinary carriers pose a significant health hazard. Eradication of the focus of infection is usually difficult. Ciprofloxacin (750 mg) or norfloxacin (400 mg) twice daily for 28 days are effective, especially in those with chronic gallbladder disease. In patients with gallstones or chronic cholecystitis, cholecystectomy is the most effective way to eliminate the carrier state, but cholecystectomy alone is inadequate if there is associated infection in the biliary or urinary tract. Carriers should not handle food or water that is to be consumed without further cooking or purification.

Prevention

Typhoid fever remains a global health problem. Its eradication in endemic areas requires major efforts to provide safe water supplies and waste disposal. Health education should include proper ways of cooking and storing food, measures that also minimize other salmonella infections. The development of effective oral typhoid vaccines promises effective prophylaxis in the future, but vaccines are unlikely substitutes for the more important public-health measures. Three vaccines are currently available: a parenteral whole-cell vaccine; Ty21a, an effective oral vaccine that is the attenuated, non-pathogenic strain of *S. typhi*, which lacks the Vi polysaccharide; and the parenteral, Vi capsular polysaccharide vaccine.

A recent meta-analysis of the efficacy and toxicity of typhoid vaccines concluded that whole-cell vaccines are more effective than the Ty21a and Vi vaccines but are more frequently associated with adverse events. Whether the added efficacy of the whole-cell vaccines outweighs their toxicity will depend on the setting in which vaccination is used.

Paratyphoid fever

The infection is caused by *S. paratyphi* A, B, and C, of which type B is the most common. The organisms, like *S. typhi*, are also primarily human pathogens, although *S. paratyphi* B has been known to infect cattle.

Paratyphoid fever is similar in most respects to typhoid fever but runs a milder and shorter course. Complications and mortality are also less frequent. Effective vaccines against paratyphoid fever are not available.

Non-typhoidal salmonella infections

Enterocolitis (food poisoning) is the most common form of salmonella infection and is caused by *S. enteritidis* serotypes, which, unlike *S. typhi* and *S. paratyphi*, are

primarily animal pathogens. There are approx. 2000 named serotypes.

Pathogenesis

Infections are usually acquired through the ingestion of food or, less often, of water contaminated by animals, particularly cattle and poultry, for which salmonella is a primary pathogen. A high infecting dose is normally required for salmonella serotype infections, although a lower infective dose may cause illness in patients with gastric hypochlorhydria, immunosuppression, or debilitating diseases. The organisms multiply in the small intestine and produce an acute enterocolitis after an incubation period of about 12 to 48 h. They rarely cause significant bacteraemia.

Clinical features

This varies widely but usually includes abdominal cramps, pain, and diarrhoea that may be bloody and purulent. The infection is normally brief and self-limiting. It is confined to the intestinal tract and often requires no specific therapy. However, in severe cases the bloody diarrhoea may persist and toxic megacolon may rarely occur; the presentation may be confused with ulcerative colitis. Colonic dilation will usually resolve without surgery. Perforation is extremely rare. In other cases, ileal involvement may be prominent with pain and tenderness in the right lower quadrant, mimicking acute appendicitis.

The differential diagnosis includes campylobacter and shigella infections, which have similar incubation periods. An incubation period of less than 6 h, absence of fever, a shorter period of diarrhoea, and absence of blood and pus in the faeces tilts the diagnosis towards the toxin-type food poisoning caused by *Staphylococcus aureus, Clostridium perfringens*, or *Bacillus cereus*.

In contrast to *S. typhi*, bacteraemia occurs in only about 5 per cent of all patients with non-typhoidal salmonella infection (Table 1) but may lead to metastatic infections. Chronic extraintestinal focal diseases may manifest insidiously in any anatomical site long after a bout of enterocolitis and are often associated with other chronic diseases and immunosuppressed status. Osteomyelitis occurs not infrequently in immunosuppressed patients and those with sickle-cell disease. Suppurative complications needing surgery are increasingly being encountered in patients with human immunodeficiency virus infection. The *S. enteritidis* serotypes also seem to have a predilection for atherosclerotic lesions, particularly aneurysms; the infection causes acute softening and rapid enlargement of the aneurysm with catastrophic consequences if not treated promptly. In addition, they may cause septic thrombophlebitis and deep venous thrombosis. In developing countries where malaria and schistosomiasis are endemic, salmonella bacteraemia is an important complication.

Further reading

Archampong EQ. Operative treatment of typhoid perforation of the bowel. *British Medical Journal* 1969; **26**: 207–30.

Azad AK *et al.* Comparison of clinical features and pathologic findings in fatal cases of typhoid fever during the initial and later stages of the disease. *American Journal of Tropical Hygiene* 1997; **56**: 490–3. [This study suggests further studies are needed to understand the mediators of shock and tissue injury during the early period of the illness.]

Bitar R, Tarpley J. Intestinal perforation in typhoid fever: a historical and state-of-the-art review. *Reviews of Infectious Diseases* 1985; **7**: 257–71.

Butler T, Islam A, Kabir I, Jones PK. Patterns of morbidity and mortality in typhoid fever dependent on age and gender: review of 552 hospitalised patients with diarrhoea. *Reviews of Infectious Diseases* 1991; **13**: 85–90.

Butler T, Knight J, Nath SK, Speelman P, Roy SK, Azad MAK. Typhoid fever complicated by intestinal perforation: a persisting fatal disease requiring surgical management. *Reviews of Infectious Diseases* 1985; **7**: 244–56.

Cocolin L, Manzano M, Astori G, Botta GA, Cantoni C, Comi G. A highly sensitive and fast non-radioactive method for the detection of polymerase chain reaction products from Salmonella serovars, such as *Salmonella typhi*, in blood specimens. *FEMS Immunology and Medical Microbiology* 1998; **22**(3): 233–9.

Cohen JL, Bartlett JA, Corey GR. Extra-intestinal manifestations of salmonella infections. *Medicine* 1987; **66**: 349–88.

Edelman R, Levine MM. Summary of international workshop on typhoid fever. *Reviews of Infectious Diseases* 1986; **8**: 329–49.

Eggleston FC, Santoshi B. Treatment of bleeding typhoid ulcer of the terminal ileum and caecum. *World Journal of Surgery* 1981; **5**: 295–7.

Eggleston FC, Santoshi B, Singh CM. Typhoid perforation of the bowel: experiences in 78 cases. *Annals of Surgery* 1979; **190**: 31–5.

Engels EA, Falagas ME, Lau J, Bennish ML. Typhoid fever vaccines: a meta-analysis of studies on efficacy and toxicity. *British Medical Journal* 1998; **316**: 110–16.

Forsyth JRL. Typhoid and paratyphoid. In: Collier L, Balows A, Sussman M, ed. *Topley and Wilson's microbiology and microbial infections*, 9th edn, Vol. 3, pp. 459–78. Arnold, London, 1998.

Gibney EJ. Typhoid perforation. *British Journal of Surgery* 1989; **76**: 887–9.

Hermans P, Gerard M, Laethem YV, Wit SD, Chumeck M. Pancreatic disturbances and typhoid fever. *Scandinavian Journal of Infectious Diseases* 1991; **23**: 201–5.

Hoffman SL *et al.* Reduction of mortality in chloramphenicol treated severe typhoid fever by high dose dexamethasone. *New England Journal of Medicine* 1984; **310**: 82–8. [An important study from Jakarta that shows the value of high doses of dexamethasone in severe typhoid fever.]

Jones BD, Ghori M, Falkow S. *Salmonella typhimurium* initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. *Journal of Experimental Medicine* 1994; **180**: 15–23.

Loefler IJ. The prognostic factors in typhoid ileal perforation: a prospective study of 50 cases. *Journal of the Royal College of Surgeons of Edinburgh* 1998; **43**(5): 359–60.

Mandal BK. Salmonella infections. *Medicine International* 1992; **105**: 4393–7.

Meir DE, Imediegwu OO, Tarpley JL. Perforated typhoid enteritis: operative experience with 108 cases. *American Journal of Surgery* 1989; **157**: 423–7.

Murdoch DA, Banatvala NA, Bone A, Shoismatulloev BI, Ward LR, Threlfall EJ. Epidemic ciprofloxacin-resistant *Salmonella typhi* in Tajikistan. *Lancet* 1998; **351**: 339. [This is the first report of epidemic ciprofloxacin-resistant *S. typhi*.]

Noorani MA, Sial I, Mal V. Typhoid perforation of small bowel: a study of 72 cases. *Journal of the Royal College of Surgeons of Edinburgh* 1997; **42**: 274–6.

Omer AHS *et al.* Unusual complications of salmonella enteriditis group D infection. *Journal of Tropical Medicine and Hygiene* 1988; **91**: 181–2.

Richens J. Typhoid and paratyphoid fevers. In: Weatherall DJ, Ledingham JGG, Warrell DA, ed. *Oxford Textbook of Medicine*, 3rd edn, Vol. 1, pp. 560–8. Oxford University Press, 1996.

Rowe B, Ward LR, Threlfall EJ. Multidrug-resistant *Salmonella typhi*: a world-wide epidemic. *Clinics in Infectious Disease* 1997; **24**(Suppl. 1): 5106–9.

Santillana M. Surgical complications of typhoid fever: enteric perforation. *World Journal of Surgery* 1991; **15**: 170–5.

Sitaram V, Moses BV, Fenn AS, Khanduri P. Typhoid ileal perforations: a retrospective study. *Annals of the Royal College of Surgeons of England* 1990; **72**: 347–9.

Sperber SJ, Schleupner CJ. Salmonellosis during infection with human immunodeficiency virus. *Reviews of Infectious Diseases* 1987; **9**: 925–34.

Stormon MO, McIntyre PB, Morris J, Fasher B. Typhoid fever in children: diagnostic and therapeutic difficulties. *Pediatric Infectious Disease Journal* 1997;**16**(7): 713–14.

Tovey FI. Typhoid enteric perforation. *Tropical Doctor* 1998; **28**(1): 1–3. [A concise review of the operative management of typhoid enteric perforations.]

47.1.3 Abdominal tuberculosis

N. Rangabashyam, B. S. Anand and R. OmPrakash

Definition
History
Epidemiology
Microbiology
Routes of infection and pathogenesis
Sites of intestinal involvement
Pathology
- Intestinal tuberculosis
- Peritoneal tuberculosis

Histopathology
Clinical features
- Peritoneal tuberculosis
- Gastrointestinal tuberculosis

Diagnosis
- Laboratory tests
- Laparoscopy and peritoneal biopsy
- Imaging studies
- Barium studies
- Endoscopy with biopsy
- Serological studies
- Polymerase chain reaction

Treatment
- Medical treatment
- Role of surgery

Specific forms of tuberculosis
- Appendiceal
- Anal canal
- Gastric
- Liver and spleen
- 'Silent' tuberculosis
- Abdominal tuberculosis in childhood
- HIV and tuberculosis

Prognosis
General recommendations
Further reading

Definition

The term abdominal tuberculosis refers to tuberculous infection of the gastrointestinal tract, mesenteric lymph nodes, peritoneum and omentum, and of solid organs related to the gastrointestinal tract such as the liver and spleen (Table 1). Other forms of intra-abdominal tuberculosis, such as of the kidney, urinary tract and genital organs, constitute a different entity and will not be discussed in this section.

I. Peritoneal tuberculosis—acute or chronic
A. Tuberculosis of the peritoneum—chronic form
(i) Wet or ascitic type:
Generalized
Localized (loculated)
(ii) Dry or fibrous type:
Adhesive type
Plastic type
Miliary nodule type
B. Tuberculosis of peritoneal folds and contents
Mesenteric adenitis
Mesenteric cysts
Mesenteric abscesses
Bowel adhesions
Rolled-up omentum
II. Gastrointestinal tuberculosis
Ulcerative
Hypertrophic or hyperplastic
Sclerotic or fibrotic
III. Tuberculosis of solid viscera (e.g. liver and spleen)

Table 1 Various pathological forms of abdominal tuberculosis

History

Tuberculosis was first recognized in the fourth century BC. Hippocrates described a condition resembling tuberculosis in a patient with pulmonary lesions and intestinal disease, and stated that 'phthisical persons die if diarrhoea sets in and it is a mortal symptom'. In the nineteenth century and the early part of twentieth century, tuberculosis was widely prevalent in most parts of the world, and was the major cause of intestinal strictures and bowel obstruction. It is believed that the French emperor Louis XIII succumbed to severe intestinal tuberculosis with ulcerations, perforation, and peritonitis. The modern era in tuberculosis began in 1882 with the identification of the causative organism, *Mycobacterium tuberculosis*, by Robert Koch; this facilitated not only a definitive diagnosis of the illness but also the development of effective antimycobacterial agents. In 1998 the complete genetic sequence of *M. tuberculosis* was identified. The bacterium contains 4.4 million base pairs and 4000 genes, making it the biggest bacterial genome to be fully sequenced thus far, except for *Escherichia coli*. This information may not only help identify virulence factors that could serve as targets for new antibacterial drugs but also provides a variety of new approaches to vaccine development.

Epidemiology

As a result of effective treatment measures and general improvements in sanitation and hygiene, a dramatic decline in the incidence of tuberculosis was noted in the middle of the twentieth century. By the 1970s, it was considered a rare disease in the West. However, starting from the mid-1980s a resurgence of tuberculosis occurred, due to a large extent to the epidemic of acquired immune deficiency disease (**AIDS**). Worldwide the number of new patients with tuberculosis is expected to increase from 8 million annually to over 10 million by year 2000. Intestinal tuberculosis is seen more frequently in people of poor socioeconomic circumstances. Its incidence is higher in patients with caseopneumonic and advanced lung disease than in those with fibrotic lesions and early disease.

In addition to an increase in the incidence of tuberculosis, several other clinical and epidemiological changes have occurred in recent years. The frequency of extrapulmonary tuberculosis has increased, largely an influence of human immunodeficiency virus (**HIV**) infection; over 50 per cent of individuals with HIV and tuberculosis develop extrapulmonary disease compared to 10 to 15 per cent in those without HIV. Studies using DNA fingerprinting suggest that as many as 40 per cent of new patients have recently transmitted infection, in contrast to previous estimates that 90 per cent of active disease is a reactivation of infection acquired years earlier. Finally, there is a marked increase in the emergence of multidrug-resistant *M. tuberculosis* organisms, a cause of much concern for the global control of tuberculosis.

Microbiology

Mycobacterium tuberculosis is responsible for nearly all cases of abdominal tuberculosis. Other pathogenic organisms such as *M. bovis* have been largely eliminated by public-health measures and are rarely encountered today. Several other atypical or anonymous mycobacteria whose pathogenicity is not yet established have been identified. The tubercle bacillus is a Gram-positive, aerobic, non-motile, non-spore-bearing organism that is identified by the Ziehl–Neelson acid-fast differential staining method. The classic method of culturing the organism is in the solid Löwenstein–Jensen medium, which requires an incubation period of 4 to 6 weeks. Liquid culture medium, containing broth base, casein and bovine serum, may provide faster results and is more sensitive than the solid medium. The virulence of *M. tuberculosis* is

established by guinea-pig inoculation.

Routes of infection and pathogenesis

Mycobacterium tuberculosis spreads to the abdomen by several routes. Ingestion of contaminated food may cause primary intestinal tuberculosis; this route of infection has decreased in recent years. Secondary intestinal disease arises from swallowed sputum containing tuberculous bacilli; it is influenced by the virulence and quantity of the bacilli, and by host resistance to the infection. The peritoneum, mesenteric nodes, and the intestine may become infected during the bacteraemic phase that may follow primary pulmonary tuberculosis. Mycobacteria may also spread from diseased adjacent organs, such as the fallopian tubes. When the intestines become infected by lymphatic spread from the mesenteric lymph nodes, the nodal disease is considered as the primary site and intestinal involvement is secondary. This conclusion is supported by the observation that the earliest intestinal lesions are found in the submucosal layer, while the overlying mucosa is normal. In addition, more advanced abnormalities such as caseation necrosis are found in the mesenteric nodes rather than in the intestine. The bacteria may also be disseminated in the bile, since they are sequestrated and excreted from granulomas in the liver.

Sites of intestinal involvement

The terminal ileum and ileocaecal junction are involved most frequently. The other regions affected in order of decreasing frequency are: colon, jejunum, rectum and anal canal, duodenum, stomach, and oesophagus (Table 2). The site of predilection is dictated by factors such as the abundance of lymphoid tissue, the rate of absorption of the intestinal contents, prolonged stasis, which provides longer time of contact with the mucosa, and the digestive activity of the intestinal contents.

Peritoneal	37.6
Gastrointestinal:	56.9
Oesophagus	0.2
Stomach	1.0
Duodenum	1.0
Small bowel	27.0
Ileocaecal area	22.9
Appendix	0.4
Colon and rectum	9.2
Mesenteric lymph nodes	6.2

Results based on collective data from nine studies involving 824 cases.

Table 2 Percentage frequency of involvement of different areas in abdominal tuberculosis

Pathology

Intestinal tuberculosis

Ulcerative lesions

Tuberculous intestinal ulcers are usually deep and are transversely placed in the direction of the lymphatics. Multiple ulcers may be seen, most often in the terminal ileum. Disease progression is associated with the appearance of an inflammatory mass around the bowel. The diseased part of the gut becomes thickened and the serosal surface is studded with tubercles (Fig. 1). There is often a marked increase in mesenteric fat, with fat wrapping around the bowel loops. The regional nodes become enlarged and may caseate, leading to mesenteric abscess formation. Bowel perforation is rare, and is usually confined by the perilesional inflammatory mass.

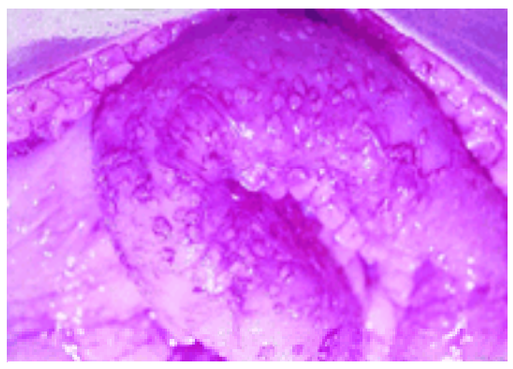


Fig. 1. Grossly thickened small intestine with multiple tubercles over the serosal surface; mesenteric thickening is also apparent.

Hyperplastic lesions

In the hyperplastic form of intestinal tuberculosis, a fibroblastic reaction occurs in the submucosa and subserosa, resulting in marked thickening of the bowel wall; this, together with involvement of the adjacent mesentery, lymph nodes, and the omentum, results in the formation of a mass lesion. Hyperplastic lesions are believed to be the result of reduced bacterial virulence and increased host resistance.

Sclerotic lesions

The sclerotic variety is associated with strictures of the intestine, typically described as ‘napkin-ring strictures' (Fig. 2, Fig. 3), which may be single or multiple (Fig. 4). When multiple, the strictures may occur in a short segment of the bowel or over the entire length of the intestine. Enteroliths can form proximal to the stricture (Fig. 5). In some patients a combination of the different pathological forms may be seen.



Fig. 2. Napkin-ring stricture of ileum: the mesenteric fat encroaching on the intestine is easily seen.

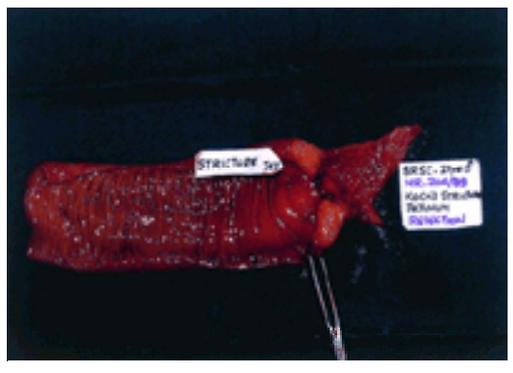


Fig. 3. Cut section of resected ileum showing a dense, ring-like stricture; the mucosa proximal to it is congested and ulcerated.

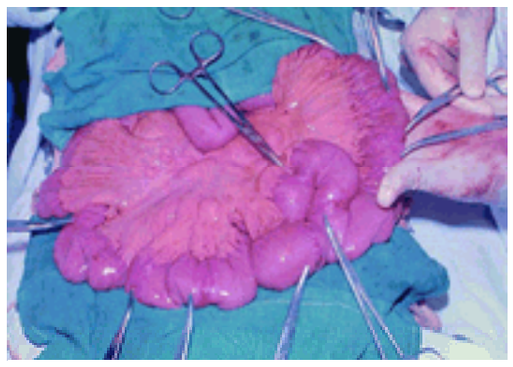


Fig. 4. Multiple strictures involving a long segment of small intestine; the enlarged mesenteric nodes are clearly seen.

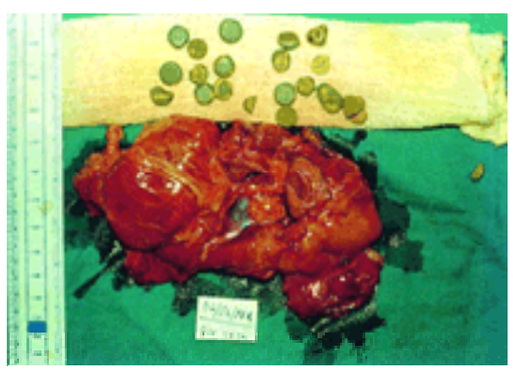


Fig. 5. Enteroliths and a few unabsorbed tablets in the dilated bowel, proximal to a stricture.

Peritoneal tuberculosis

Acute peritonitis

Acute tuberculous peritonitis is extremely rare and is encountered under the following circumstances: in the miliary phase of the disease, on perforation of intestinal disease, and with local dissemination from a ruptured, caseating mesenteric lymph node.

Chronic peritonitis

The chronic form of tuberculous peritonitis is much more common and typically presents as ascites. The fluid is usually clear and straw coloured, but may be sanguinous. Peritoneal adhesions that range from thin and flimsy to dense and thick may occur. In the presence of adhesions the ascitic fluid may become loculated, presenting as a localized cyst. The characteristic lesions are miliary nodules (Fig. 6). When the nodules increase in size and coalesce, plastic adhesions develop (Fig. 7), which may completely obliterate the peritoneal cavity, forming an abdominal cocoon that may encase the intestines (Fig. 8). The omentum thickens to form a transversely placed mass, the so-called rolled-up omentum.

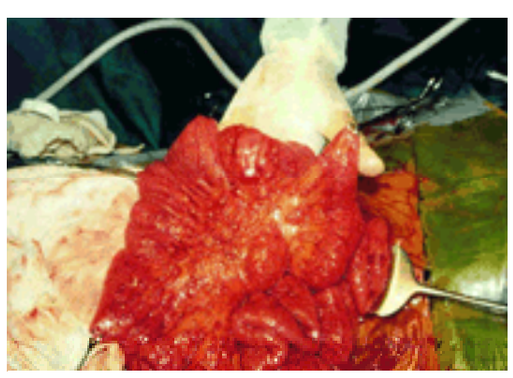


Fig. 6. Extensive miliary nodules spread over the entire mesentery and intestinal serosal surface.

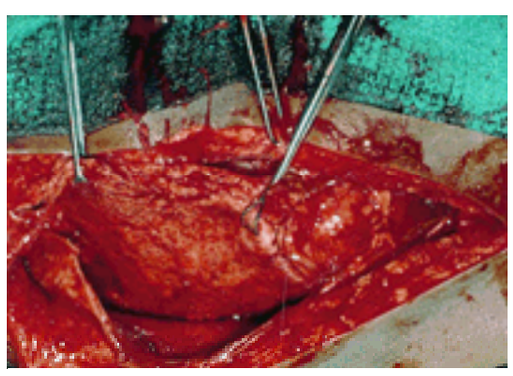


Fig. 7. Dense, plastic adhesions of tuberculous abdomen.



Fig. 8. Cocoon formation (entrapment of variable lengths of small intestine in a bag of membrane).

Histopathology

The typical histological feature of tuberculosis is the caseating granuloma, which may become large and confluent; these are seen in all diseased areas including the bowel, mesenteric nodes, and peritoneum (Fig. 9). The granulomas have a peripheral zone of lymphocytes, plasma cells, and Langhans giant cells with a central area of necrosis. Healing of the granulomas is associated with mucosal regeneration.

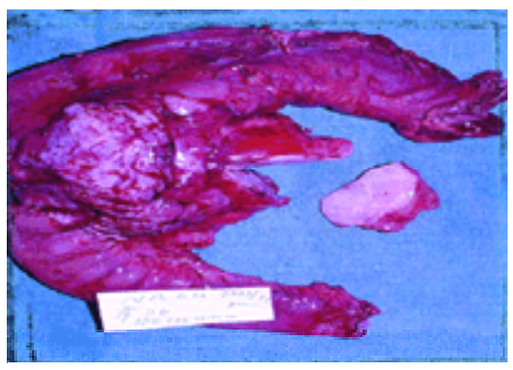


Fig. 9. Caseating granulomas in both the intestine and lymph node.

Clinical features

Abdominal tuberculosis is seen most frequently in patients between the ages of 30 and 50 years. Females outnumber males by 2:1. The onset of illness is usually insidious, and the initial symptoms are often vague and non-specific; this is particularly true of peritoneal tuberculosis. As the disease progresses the individual may develop fever, which is present in two-thirds of the patients, night sweats, malaise, weakness, anorexia, and weight loss. The appearance of specific symptoms depends upon the predominant site of involvement. Analysis of data collected from nine studies involving 824 patients with abdominal tuberculosis showed that the peritoneum was involved in 38 per cent, the gastrointestinal tract in 57 per cent, and the mesenteric lymph nodes in 6 per cent of patients (Table 2).

Peritoneal tuberculosis

Tuberculous peritonitis usually presents with ascites. Less frequently, fluid collection is mild but the fibrotic component is more prominent, resulting in thickening of the peritoneum with adhesions, fluid loculation, and the classic ‘doughy feel' of the abdomen. Patients with peritoneal tuberculosis, especially women, often have coexisting tuberculosis of the pelvic organs, since the genital tract is frequently the portal of entry of the tubercle bacillus. This point is worth remembering, since a good pelvic examination and endometrial biopsy may provide the simplest method of confirming the diagnosis.

Gastrointestinal tuberculosis

The most common presentation in patients with disease of the gastro-intestinal tract is abdominal pain (Table 3). The pain may be dull and vague, but when colicky it suggests intestinal obstruction. In such patients the pain is often exacerbated by eating and relieved by vomiting, but only transiently as it soon recurs. Diarrhoea is another common SYmptom in patients with intestinal involvement. The stools may be watery, small in amount, and mixed with blood when the disease affects predominantly the colon. In primary small-bowel disease the stools are large in amount, foul smelling, and resemble those seen in patients with malabsorption. Other symptoms include flatulence, nausea, altered bowel habit, and borborygmi. Abdominal distension suggests the presence of ascites or persistent subacute intestinal obstruction. Typical duodenal ulcer-like pain may occur when the duodenum is involved.

Abdominal		*Others*	
Abdominal pain	86 (77–94)	Weight loss	63 (35–87)
Vomiting	46 (33–74)	Fever	61 (29–100)
Abdominal distension	37 (28–45)	Anaemia	48 (10–100)
Ascites	37 (19–60)	Amenorrhoea	23 (12–36)
Borborygmi	35 (26–50)	Pulmonary	19 (4–51)
Abdominal mass	33 (17–45)		
Diarrhoea	22 (11–48)		
Constipation	24 (12–46)		
Haematochezia	4 (2–13)		

Average values (per cent) and range (in parentheses).

Table 3 Presenting features in abdominal tuberculosis

Physical examination

Patients with chronic abdominal tuberculosis are often malnourished and anaemic. In some the abdomen may be completely normal on examination, but most demonstrate some abnormal findings. There may be visible peristalsis and the distended bowel loops can be palpated. The abdomen may show diffuse distension and tenderness, or the signs may be more localized, usually in the right lower quadrant. An ileocaecal mass may be felt in the right iliac fossa or higher up in the right lumbar region. A ‘doughy' abdomen suggesting peritoneal disease has become less common in recent years. A rolled-up omentum, when present, is felt as a

transversely place mass in the epigastric region. Loculated ascites, mesenteric cysts, and mesenteric abscesses present as cystic masses.

Patients with ascites may have shifting dullness and a fluid thrill. In patients with large-bowel disease a diffusely thickened and tender colon may be felt. Other findings include hepatosplenomegaly, pelvic abnormalities mimicking gynaecological tumours, and features of gastric-outlet obstruction due to direct involvement of the stomach or extrinsic compression of the duodenum by enlarged mesenteric lymph nodes. Rectal examination may reveal anal fistulas, fissures, or stricture.

Complications

Intestinal obstruction and malabsorption are the most frequent complications (Table 4). Much less common are bowel perforation and massive gastrointestinal haemorrhage. Perforation is usually confined, owing to the presence of surrounding adhesions; as a result, signs of free perforation of the bowel are rarely seen. Acute bleeding from the rectum or haematemesis are rare, but can occasionally be severe and life threatening. Fistulas, both internal between adjacent bowel loops or with other hollow organs (uterus, vagina), and external to the skin, are described but are rare.

Ascites	37 (19–60)
Abdominal mass(s)	33 (17–45)
Non-obstructive and vague	29 (25–38)
Bowel obstruction	21 (15–35) [3–20% of all obstructions]
Malabsorption	14 (5–20)
Bleeding per rectum	4 (2–13)
Bowel perforation	3 (1–10) [5–9% of all perforations]

Average values (per cent) and range (in parentheses).

Table 4 Clinical signs and complications in abdominal tuberculosis

Diagnosis

Since the discovery of the tubercle bacillus it has been possible to make a precise diagnosis of tuberculosis. However, in patients with abdominal tuberculosis the causative organism is often difficult to identify and the diagnosis is generally made by indirect methods.

Laboratory tests

The most common abnormality on routine blood tests is an elevated erythrocyte sedimentation rate, found in over 90 per cent of patients. Other abnormalities include varying degrees of anaemia and leucopenia with relative lymphocytosis. A positive tuberculin skin test (Mantoux test) is an excellent screening test in non-endemic countries, but is of little use in endemic countries because of high rates of positivity in healthy individuals and in those who have received the bacillus Calmette–Guérin (**BCG**) inoculation.

Analysis of ascitic fluid

Routine tests

The ascitic fluid has a high white blood-cell count, with a predominantly lymphocytic response. A total white-cell count of 500/mm^3 or more has a sensitivity of 81 per cent, a specificity of 48 per cent, and a diagnostic accuracy of 46 per cent for tuberculosis. If a high count is associated with a predominantly non-polymorphonuclear response, the specificity and accuracy improve to 82 per cent and 78 per cent, respectively. The diagnostic usefulness of the white-cell count is reduced in the presence of coexisting conditions such as cirrhosis and HIV infection, both of which are associated with an increased incidence of abdominal tuberculosis. In patients with cirrhosis and AIDS the response of white cells is poor and the total white blood-cell count is often within the normal range.

Another characteristic of tuberculous ascites is high total protein, usually 2.5 g/dl or more, and the serum/ascitic fluid albumin gradient (**SAAG**) is less than 1.1. However, high total protein is seen in several conditions, and the sensitivity, specificity, and diagnostic accuracy of this test for tuberculosis are only 65, 78, and 74 per cent, respectively. As with the white blood-cell count, the protein content of ascitic fluid in cirrhotic patients with peritoneal tuberculosis is significantly lower; the concentration is less than 2.5 g/dl in 30 to 50 per cent of patients, and the SAAG is highly variable. Other biochemical tests, such as lactate dehydrogenase (elevated to over 90 units/l), low pH, and an ascitic fluid:blood glucose ratio of less than 0.96, are usually positive in tuberculous ascites, but these tests are non-specific and of little diagnostic value.

Adenosine deaminase

Lately, adenosine deaminase (**ADA**) in ascitic fluid has received much attention. ADA is an enzyme present in several cell types including macrophages, lymphocytes, and erythrocytes. Its concentration in body fluids correlates with the number and degree of stimulation of lymphocytes, and is a marker of host immune response. Several studies have shown that ADA activity is an extremely useful diagnostic test for tuberculous ascites, with specificity and sensitivity of over 95 per cent. However, in the presence of cirrhosis, seen in more than 50 per cent of patients with peritoneal tuberculosis in the West, the sensitivity of ADA drops to 30 per cent; this is related to the abnormal T-lymphocyte activation and proliferation in those with cirrhosis. For similar reasons, ADA activity is likely to be less useful in HIV-positive patients, limiting the effectiveness of this test in subgroups of patients who are highly susceptible to tuberculosis. However, in those without these predisposing conditions, ADA is extremely useful and should be a first-line investigation, obviating the need for more invasive and expensive diagnostic tests. Another test that may prove useful is assay of interferon g (produced by T-lymphocytes) in the ascitic fluid.

Bacterial isolation and culture

Positive identification of *M. tuberculosis* provides the definitive diagnosis of tuberculosis. Acid-fast smears prepared from ascitic fluid have an extremely low yield, with positive results in fewer than 5 per cent of patients. Culture of ascitic fluid for *M. tuberculosis* is more useful and is positive in 20 to 45 per cent of patients. An important disadvantage of culture is that the results take 4 to 6 weeks, which severely limits the clinical usefulness of the test. However, despite this limitation, the current recommendation from the United States Centers for Disease Control is to perform culture and *in vitro* drug-susceptibility testing in all patients in an effort to combat the emergence of multidrug-resistant *M. tuberculosis*. While awaiting the results of culture, antituberculosis treatment should be started, which can be modified later based on drug-sensitivity findings.

Laparoscopy and peritoneal biopsy

Blind needle biopsy of the peritoneum can be performed in the presence of ascites using special needles (Abrams' or Cope's). Most workers note a low complication rate, but fatality has been reported. The main risk of blind biopsy is bowel perforation, particularly in patients with adhesions between bowel loops and the anterior abdominal wall. Open biopsy of parietal peritoneum under local anaesthesia is much safer. A less traumatic and more useful approach is to obtain targeted biopsy specimens from diseased areas such as enlarged lymph nodes under ultrasonographic or computed tomographic (**CT**) guidance. The single most sensitive diagnostic test is laparoscopic examination of the peritoneum. The peritoneal lining loses its smooth, glistening appearance and becomes rough, irregular and dull. The characteristic finding, as described above, is the presence of miliary tubercles. In addition, fibrous adhesions ranging from tiny filaments to thick bands may be seen. Ascitic fluid, if present, should be sent for biochemical analysis and culture. Targeted biopsy specimens should be obtained, preferably from the miliary tubercles. Peritoneal biopsy should be performed even if peritoneal nodules are not seen; histological examination shows caseating granulomas in nearly 90 per cent of patients. The combination of the distinctive visual and histological appearances accurately identifies nearly all patients. However, positive identification of *M. tuberculosis* histologically or by culture is made in only one-half of the patients. The complication rate of laparoscopy is low (less than 5 per cent) but care should be taken in the presence of adhesions. This procedure is especially rewarding in patients with AIDS and cirrhosis, because of a much broader differential diagnosis of ascites and the poor sensitivity of the other tests.

Imaging studies

Plain radiographs

Plain radiographs of the abdomen may show calcification of the mesenteric lymph nodes and calcified granulomas in the spleen, liver or pancreas, but these findings do not imply active disease. Patients in intestinal obstruction may show dilated bowel loops and air–fluid levels. An abnormal chest radiograph indicating pulmonary tuberculosis helps in the differential diagnosis but is seen in fewer than 50 per cent of patients with gastrointestinal disease.

Ultrasonography

Several ultrasonographic findings have been described. The peritoneum assumes a thickened, irregular, echo-poor, sheet-like or nodular appearance. Ultrasound is very sensitive in detecting small quantities of fluid. Because of their small size, peritoneal nodules are rarely detected, but are better visualized in the presence of ascites and appear as tiny, echo-poor deposits. A characteristic finding is the presence of alternating echogenic and echo-free layers produced by the bowel wall, the serosa, and the adjacent bowel loop with interloop fluid collections; this is termed the ‘club sandwich' appearance. In addition, isolated or matted enlarged lymph nodes, with hypoechoic or anechoic centres caused by caseation necrosis, may be seen.

Computed tomography

Similar findings as noted on ultrasound, but with better definition and resolution, are seen with the CT scan. The ascitic fluid has high-density appearances because of its elevated protein content. The thickening and nodularity of the peritoneum and mesentery can be more easily identified with CT than ultrasonography. In patients with intestinal disease there is thickening of the bowel wall and of the ileocaecal valve. The ultrasound and CT findings, although fairly characteristic, are not pathognomonic of abdominal tuberculosis. Similar appearances can be seen in other diseases such as lymphoma, metastatic carcinoma, peritoneal mesothelioma, and pseudomyxoma peritonei.

Barium studies

Barium studies provide useful information on the extent and severity of the intestinal disease. The earliest abnormalities include altered bowel motility, and irregularity and thickening of the mucosal folds. Barium swallow may show compression of the oesophagus by a mediastinal node, with or without mucosal ulceration (Fig. 10). In patients with more advanced disease, mucosal ulcers, deformity of the bowel lumen, and stricture formation are noted (Fig. 11). Rarely, sinus tracts and fistulas may be seen. Thickening of the ileocaecal valve, a wide-open valve accompanied by narrowing of the terminal ileum (Fleischner sign), and a fibrotic terminal ileum opening into a contracted caecum (Sterlin sign) are characteristic of intestinal tuberculosis.

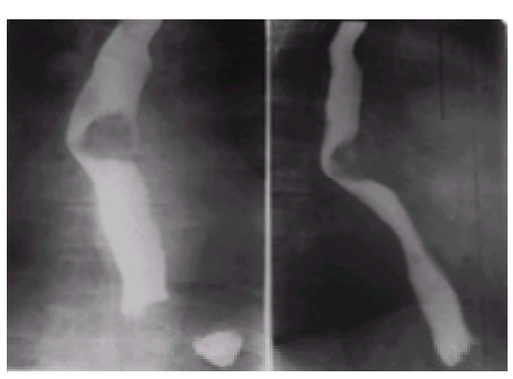


Fig. 10. Barium swallow showing extraneous compression of oesophagus by a mediastinal node.



Fig. 11. Barium meal study of small bowel shows an ileal stricture; the ileum proximal to it is grossly dilated.

Enteroclysis (small-bowel enema) is generally considered the ideal method of assessing the small bowel. This technique involves passing a tube into the distal duodenum, which greatly reduces enthusiasm for the test, both by the radiologist and the patient. Most radiologists therefore prefer the use of small-bowel series. The regular 20 per cent wt/vol barium mixture has a high density, which often prevents proper examination of overlapping bowel loops. Low-density barium preparations (13 per cent wt/vol) containing methylcellulose have been introduced in the belief that they permit better dispersion of a barium that is less prone to flocculation and has greater transradiancy, allowing a ‘see-through' effect with good visualization of overlying loops. Whether these preparations are indeed better remains to be confirmed. For the large bowel the study of choice is air–contrast barium enema.

Overall, barium studies have a low sensitivity and are abnormal in only about 60 per cent of patients with endoscopically proven disease. A rapid transit and lack of barium retention may occur in inflamed segments of the small bowel, resulting in false-negative results. Moreover, the barium abnormalities are non-specific and can be mimicked by other conditions such as Crohn's disease and lymphoma.

Endoscopy with biopsy

Fibreoptic endoscopes have made it possible to visualize directly the gastrointestinal tract. The ileocaecal region can be accessed easily by colonoscopy, and the use of enteroscopes allows examination of the proximal small bowel. The usual endoscopic findings are mucosal ulcerations, nodularity, deformity, narrowing, and stricture of the bowel (Table 5). Rarely, there is diffuse disease of the colon with hyperaemia and friability of the mucosa, mimicking ulcerative colitis. The endoscopic abnormalities in tuberculosis can be mistaken for Crohn's disease. Ulcers seen in tuberculosis are usually transversely located and have sharply defined margins with the surrounding mucosa showing erythema, while in Crohn's disease the ulcers are serpiginous, often longitudinal, with a relatively normal surrounding mucosa (Table 6).

Ulcers	77 (72–85)
Nodules	70 (51–100)
IC valve deformity	48 (27–55)
Multiple sites	29 (10–58)
Strictures	25 (20–27)
Segmental disease	20 (15–26)
Polypoid lesions	14 (13–15)
Diffuse	4 (0–8)

IC, ileocaecal.
Average values (per cent) and range (in parentheses); findings are based on 275 patients collected from seven studies.

Table 5 Endoscopic features of colonic tuberculosis

Features	*Tuberculosis*	*Crohn's disease*
Intestinal ulcers	Transverse; any location	Longitudinal; mesenteric border
Serosal miliary nodules	Numerous	Rare
Perforation	Confined and rare	Rare
Fibrosis	Common	Uncommon
Stricture	Short (< 3cm)	Usually long
Abscess and fistulas	Uncommon	Common
Granulomas	Many, large, well defined	Few
Caseation	Usually present	Absent
Submucosa	Oedematous and widened	Reduced or obliterated
Hyalinization	Common	Rare

Table 6 Features differentiating intestinal tuberculosis from Crohn's disease

Biopsy specimens obtained at endoscopy show granulomas and epithelioid cells in 40 to 74 per cent of patients with intestinal tuberculosis. However, confirmatory findings of caseation necrosis are present in only 8 to 21 per cent, acid-fast bacilli are observed infrequently, and positive cultures are obtained in only 6 to 40 per cent of patients. Although endoscopy provides a definitive diagnosis in only about one-third of the patients, it is very useful in excluding other conditions such as lymphoma, carcinoma, and caecal amoeboma.

Serological studies

Because of the difficulty in identifying *M. tuberculosis* in tissue material, a sensitive and specific serological test should be very helpful. Infection with *M. tuberculosis* stimulates a cell-mediated as well as a humoral immune response. The tuberculin skin test is a manifestation of the cell-mediated immune response. Serological tests based on the detection of specific antibodies to *M. tuberculosis* have been under investigation for several years. *Mycobacterium tuberculosis* has a complex structure and contains numerous antigens that cross-react with each other as well as with antigens of other mycobacterial species. Enzyme-linked immunosorbent assays based on the use of one of several bacterial antigenic preparations are commercially available, but they provide sensitivity, specificity, and diagnostic accuracy of only 80 per cent, and moreover are unable to identify clearly (a) active disease from dormant infection, (b) *M. tuberculosis* infection from other mycobacterial infections, and (c) individuals with previous BCG inoculation.

Polymerase chain reaction

The low positive identification rate for *M. tuberculosis* in intestinal tuberculosis is related to the presence of small numbers of organisms in the tissues. By amplifying tiny quantities of DNA and RNA, the polymerase chain reaction is thus ideally suited for this condition. The technique has been used in a variety of clinical specimens including sputum, cerebrospinal fluid, pleural and peritoneal fluids, and biopsy tissues, with sensitivity, specificity and positive predictivity of 85, 99, and 95 per cent, respectively. Based on studies with pulmonary secretions, the polymerase chain reaction can detect as few as 50 organisms per reaction, which is at least fivefold below the lower limit of detection by culture. The technique has also been successfully used on endoscopically obtained biopsy specimens but much more work is required before it can be recommended for routine use.

Treatment

Medical treatment

The treatment of abdominal tuberculosis is primarily medical. Since acid-fast bacilli and caseation necrosis are seen in a minority of patients, and culture results take several weeks, empirical antituberculous chemotherapy should be initiated in every patient with suspected tuberculosis (Table 7).

Regimens	*Drugs*	*Dosage form*	*Side-effects*
I. Standard therapy	Streptomycin	15–20mg/kg per day IM	VIIth cranial nerve damage, nephrotoxicity
	Ethambutol	25mg/kg per day × 2 weeks 15mg/kg/day × 12–18 months	Optic neuritis, hyperuricaemia, rash
	Isoniazid	7–10mg/kg per day × 12–18 months	Hepatotoxicity; rash; peripheral neuritis; fever
	Rifampicin	10mg/kg per day × 6 months	Hypersensitivity, fever, haemolysis, jaundice ↓, ↑ hepatitis; orange urine
II. Short-term therapy*	Pyrazinamide	1.5g/day × 2 months	
	Isoniazid	300mg/day × 6 months	
	Rifampicin	450mg/day × 6 months	
III. Directly observed therapy†	Pyrazinamide	1.5g/day × 2 months	
	Isoniazid	300mg/day × 2 months 600mg/BIW‡ × 4 months	
	Rifampicin	450mg/day × 2 months 600mg/BIW‡ × 4 months	

IM, intramuscular; BIW, twice weekly.
* In high-resistance areas a fourth drug, ethambutol 15mg/kg per day × 2 months or streptomycin 0.75g/day intramuscular × 2 months, is added.

Table 7 Different regimens of antituberculosis chemotherapy

Standard therapy

Until recently, the usual antituberculous treatment consisted of streptomycin (15–20 mg/kg body wt) intramuscularly daily for 2 months, isoniazid (7–10 mg/kg) orally daily for 12 to 18 months, and rifampicin (10 mg/kg) orally daily or ethambutol (25 mg/kg for 2 weeks, followed by 15 mg/kg) orally daily for 12 to 18 months.

Short-term therapy

The current guidelines put forward by the World Health Organization and the United States Centers for Disease Control recommend a much shorter course of therapy. The initial treatment consists of a three-drug regimen of isoniazid (300 mg), rifampicin (450 mg), and pyrazinamide (1.5 g) in countries where the resistance rate to isoniazid is below 4 per cent. In countries with higher drug-resistance rates, a fourth drug is added, ethambutol (25 mg/kg) or streptomycin (1 g). All drugs are administered daily for 2 months, after which the patient is switched to two drugs: isoniazid 300 mg/day and rifampicin 450 mg/day for 4 months.

Directly observed therapy

If facilities for directly observed therapy are available, initial treatment is the same as in short-term therapy, but after 2 months patients are started on a twice-a-week regimen consisting of isoniazid (600 mg/day) and rifampicin (600 mg/day) for 4 months. Thus, the total length of treatment for both short-term and directly observed therapies is 6 months.

Comment

Although these guidelines are based on studies in pulmonary tuberculosis, there is no evidence that patients with abdominal tuberculosis require longer therapy. Patients should be monitored every 4 to 6 weeks for drug-related side-effects (Table 7). Isoniazid should be discontinued if the liver enzymes (alanine and aspartate amino-transferases) increase threefold over baseline. It is advisable to administer pyridoxine hydrochloride (5–10 mg/day) to all patients to prevent isoniazid-induced peripheral neuropathy. If a coexisting open pulmonary lesion is present, the patient should be isolated for 2 weeks. If not, the patient may be treated at home.

Response to treatment

The clinical response to treatment is excellent. Systemic symptoms such as fever, malaise, and weight loss subside in a few weeks. Mucosal abnormalities take longer, but follow-up barium studies and endoscopy demonstrate regression of the lesions in most individuals. The majority of patients (approx. 70 per cent) with symptoms of subacute bowel obstruction and evidence of intestinal stricture show complete resolution of the radiological abnormality.

Antituberculosis-drug resistance

Multidrug-resistant *M tuberculosis* received little attention until the early 1990s when several outbreaks were reported in patients with HIV. In 1994, the World Health Organization initiated global surveillance for antituberculosis-drug resistance in 35 countries and results of the first 4 years (1994–1997) of this project were recently published. Among patients with no prior treatment, a median of 9.9 per cent (range 2–41 per cent) were resistant to at least one drug and multidrug resistance was seen in 1.4 per cent (0–14.4 per cent). The resistance rates to individual drugs were: isoniazid 7.3 per cent, streptomycin 6.5 per cent, rifampicin 1.8 per cent, and ethambutol 1 per cent. Much higher rates of drug resistance were observed in patients with history of previous antituberculous treatment.

Drug resistance is a major threat to tuberculosis control programmes worldwide. Patients with resistant strains are less likely to be cured, and the treatment is more expensive and more toxic than in those with susceptible organisms. Multidrug resistance occurs more frequently in non-compliant patients, in the presence of HIV infection, and in malnourished individuals. The current emphasis is to create more facilities for directly observed therapy. Studies show that unsupervised treatment has a drug-completion rate of only 25 to 50 per cent, while directly observed therapy results in 85 to 90 per cent cure, with relapse rates of only 5 per cent.

Use of corticosteroids

Studies in the 1980s suggested that concomitant use of corticosteroids might decrease the incidence of adhesions and intestinal obstruction by reducing the deposition of fibrous tissue. However, the role of corticosteroids in abdominal tuberculosis is uncertain, since no controlled, randomized trials have been performed. The routine use of steroids is therefore not recommended.

Role of surgery

Indications

Surgery is reserved for mechanical complications of tuberculosis or when medical therapy fails. Emergency surgery is indicated in the presence of acute complications such as free perforation of the bowel and severe intestinal haemorrhage. The most common indication for surgery is intestinal obstruction secondary to stricture formation. In these patients, unless there is complete obstruction, conservative management is advocated. If symptoms persist, elective surgery is performed at a later stage, with significantly less morbidity. Predictors for surgical intervention are long strictures (12 cm or more in length) and multiple areas of involvement. Other indications for surgery are bowel adhesions, intra-abdominal abscess due to a confined perforation, mesenteric abscess, and internal or external fistulas. Surgery is also appropriate if the diagnosis is in doubt and when malignancy cannot be ruled out with reasonable accuracy.

Surgical procedures

Patients presenting with acute bowel perforation should be treated by resection of the involved segment and primary anastomosis. Simple closure of perforation is followed by a high incidence of reperforation and fistula formation. Drainage tubes are not recommended and, if used, should be removed early because of increased risk of abdominocutaneous fistulas if they are left in place for more than 7 days.

Surgical techniques for patients with bowel obstruction have evolved over time. At one time, bypass procedures such as ileotransverse colostomy and enteroenterostomy were practised commonly, but have now been abandoned because of complications such as blind-loop syndrome, malabsorption, and perforation. Similarly, radical bowel resection and extensive dissection of mesenteric lymph nodes have fallen out of favour because of complications, including the development of short-bowel syndrome. The order of the day is minimal surgery. The procedure of choice for an ileocaecal massis limited resection with 5-cm margins from visibly abnormal tissue in place of the standard right hemicolectomy (Fig. 12). Segmental bowel resection has become rare and is only performed if multiple strictures are located in close proximity. Most surgeons prefer stricturoplasty for lesions involving the small bowel. Other procedures advocated are ileocaecoplasty and coloplasty for ileocaecal and colonic strictures, respectively (Fig. 13).

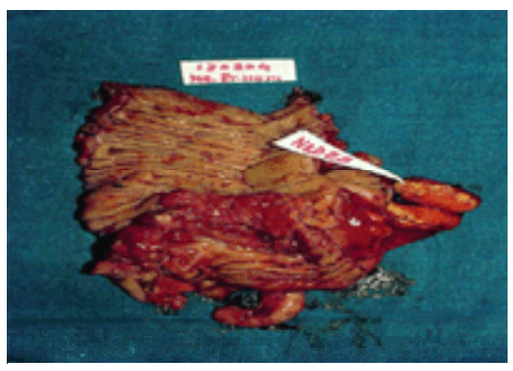


Fig. 12. Limited right hemicolectomy specimen of ileocaecal tuberculosis; the enlarged and turgid appendix is a common finding.

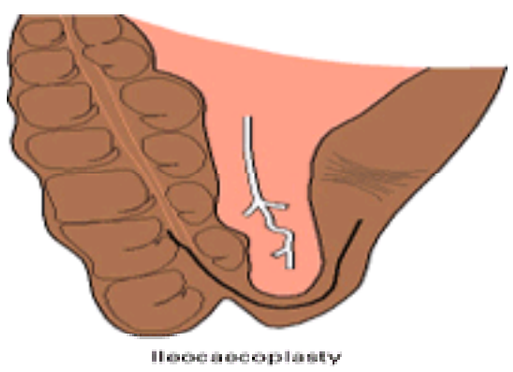


Fig. 13. Like Finney's pyloroplasty, ileocaecoplasty is performed between the distal ileum and caecum following appendectomy.

Approach to intestinal adhesions

Intestinal adhesions require a meticulous surgical approach. In addition to releasing the adhesions, enteric stenting (Fig. 14, Fig. 15) has been found useful in preventing adhesive bowel obstruction. Patients with florid sepsis may be treated by laparostomy, followed by secondary abdominal closure once the sepsis is controlled; this prevents recurrent adhesions and reduces postoperative morbidity and mortality. Patients with enterocutaneous fistulas are treated by resection of the fistulous tract, drainage of any intra-abdominal abscess, and intestinal anastomosis. A multicentre, prospective, randomized trial has shown that the interposition of a barrier film for a sufficient length of time between diseased areas of intestine contributes to adhesion-free healing. Seprafilm and Interceed are a few of the commercially available barrier films proved to be effective. Some of these patients may need stenting in addition to barrier-film placement (Fig. 16, Fig. 17).

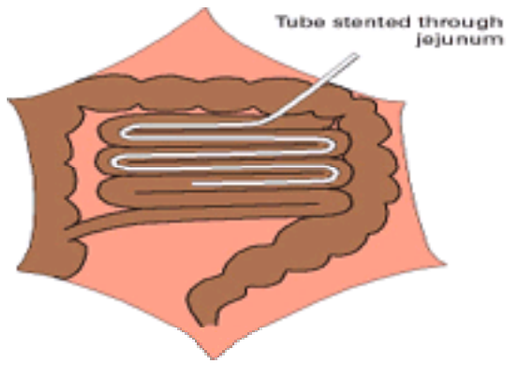


Fig. 14. Small-bowel stenting: here a long tube is threaded through the entire length of small bowel starting 15 to 23 cm distal to the duodenojejunal junction; the bowel is then arranged in the manner shown.

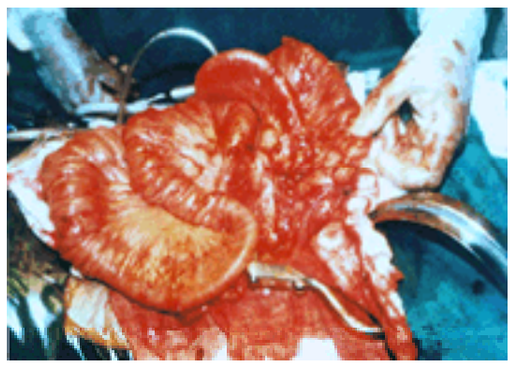


Fig. 15. Enteropexy with Baker's tube (stitchless plication tube) to form non-obstructive adhesion of small intestine.

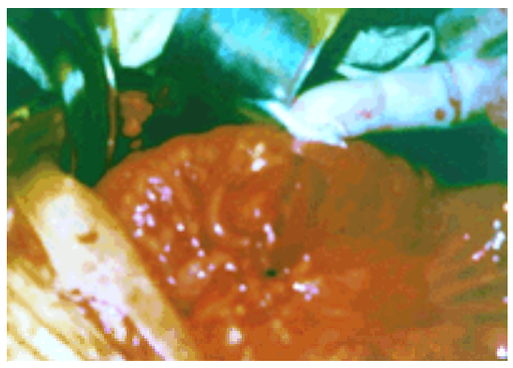


Fig. 16. After the adhesiolysis procedure, bioresorbable membrane (Seprafilm), made of sodium hyaluronate and carboxymethyl cellulose, is laid over the small intestine or peritoneal surface to prevent postoperative intra-abdominal adhesions.

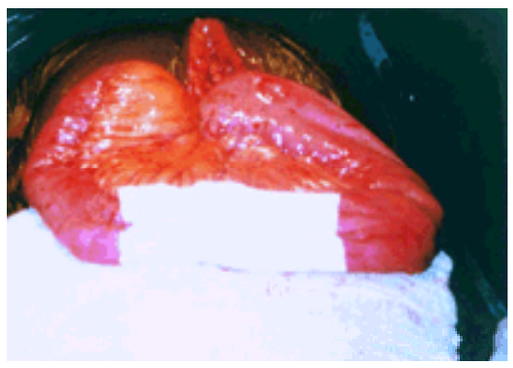


Fig. 17. Placement of absorbable adhesion barrier (Interceed – oxidized regenerated cellulose) on adhesion-prone surface to prevent adhesion.

Specific forms of tuberculosis

Appendiceal

Tuberculosis of the appendix is reported in 0.1 to 3 per cent of patients with tuberculosis. Isolated tuberculosis of the appendix is rare. Appendectomy followed by antituberculosis chemotherapy is the treatment of choice.

Anal canal

Patients with tuberculosis of the anal canal may present with multiple fistulas, fissures, or perianal abscesses. Treatment consists of drainage of the abscess combined

with antituberculous chemotherapy.

Gastric

Involvement of the stomach is rare, occurring in 0.3 to 2.3 per cent of patients with pulmonary tuberculosis. The rarity of this location is related to several factors including the resistance of the gastric mucosa to infection, the lack of lymphoid tissue in the stomach, and the constant flow of contents through the stomach. Tuberculosis of the stomach presents as an ulcerative, granulomatous, or fibrosing lesion (Fig. 18); the last two types may result in gastric-outlet obstruction. At endoscopy, the lesion may be mistaken for peptic ulcer disease, gastric carcinoma, or gastric sarcoidosis. Similar appearances are also seen in syphilis of the stomach. The diagnosis may be confirmed on endoscopic biopsy specimens.

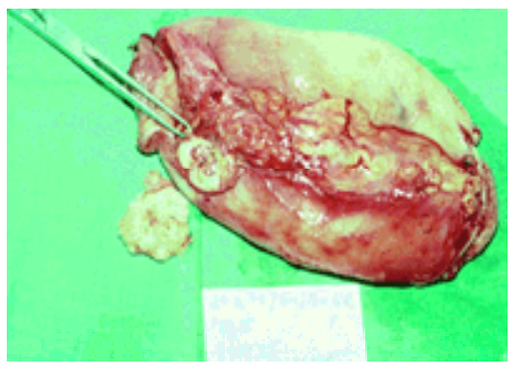


Fig. 18. Partial gastrectomy specimen showing an ulcerative nodular tuberculous lesion.

Antituberculous chemotherapy should be initiated in all patients; this is curative in most, especially those with ulcerative lesions, which can be confirmed on repeat endoscopy and biopsy. Surgical intervention is required if gastric-outlet obstruction persists despite treatment. The usual surgical approach is a partial gastric resection such as a Billroth gastrectomy or a sleeve resection. Gastric cancer presents in a similar fashion, and both tuberculosis and cancer may coexist in the same patient. Therefore, if there is any doubt about the diagnosis, it is best to proceed with frozen-section biopsy followed by surgical resection if indicated.

Liver and spleen

Hepatic tuberculosis has become exceedingly rare these days. The diagnosis is usually made accidentally during exploratory laparotomy or at autopsy in immunocompromised patients. The lesions typically are granulomas, with or without central caseating necrosis (Fig. 19), calcified masses, and biliary strictures. Tuberculous periportal lymph nodes may cause obstructive jaundice by compressing the bile duct. Patients with hepatic tuberculosis usually have hepatomegaly, with or without jaundice. SYmptoms related to abdominal tuberculosis often overshadow those due to liver disease. Liver enzymes, in particular serum alkaline phosphatase, are usually elevated. Tuberculosis should be differentiated from other conditions associated with hepatic granulomas such as leprosy, sarcoidosis, Hodgkin disease, brucellosis, infectious mononucleosis, inflammatory bowel disease, drug-induced liver damage, and syphilis. Chronic active hepatitis may also mimic tuberculosis of the liver. The treatment of hepatic tuberculosis is chemotherapy. It should be remembered that most antituberculous drugs (except ethambutol) are hepatotoxic, and may aggravate the liver damage and worsen the jaundice. These patients therefore should be kept under close observation during antituberculous chemotherapy.



Fig. 19. Multiple tuberculous miliary nodules on the surface of the liver.

Splenic tuberculosis is also rare and may present as a splenic abscess (Fig. 20) or with hypersplenism. The presence of multiple hypoechoic lesions on ultrasonography of the spleen in a HIV-positive patient is highly suggestive of disseminated tuberculosis. The diagnosis is usually made following surgical resection of the diseased spleen.



Fig. 20. Cut section of tuberculosis of spleen with caseation.

‘Silent' tuberculosis

Subclinical or ‘silent' abdominal tuberculosis is seen mainly in patients belonging to upper socioeconomic groupings in the Third World countries. The onset of illness is often precipitated by stressful conditions. The patients present with atypical symptoms, resulting in frequent misdiagnosis. The response to short-term antituberculous drug therapy is excellent.

Abdominal tuberculosis in childhood

Abdominal tuberculosis in children is seen most frequently in immunosuppressed individuals and in those who have not been vaccinated with BCG. Clinically, weight loss and malaise occur in 95 per cent of these children, followed by abdominal distension (83 per cent), abdominal pain (79 per cent), and anaemia (76 per cent). Laboratory tests show leucocytosis and an altered albumin:globulin ratio; pulmonary tuberculosis is see on the chest radiograph. Imaging studies, including plain

radiographs, ultrasonography, CT, and gastrointestinal contrast studies, are helpful in localizing the defect. The diagnosis is confirmed on tissue obtained by endoscopic biopsy or with echo- or CT-guided fine-needle aspiration; samples may be stained by the Ziehl–Nielsen method for acid-fast bacilli, subjected to microbiological culture and drug-sensitivity testing, and examined histopathologically. A short course of chemotherapy is an effective and economical treatment, and is associated with minimal side-effects in children.

HIV and tuberculosis

Abdominal tuberculosis in HIV-infected patients is invariably a manifestation of disseminated disease and results in significant mortality. Extrapulmonary tuberculosis is seen in over 50 per cent of patients. Fever, weight loss, lymphadenopathy, and splenic abscesses are seen more commonly in HIV-infected patients than in those without HIV infection.

The diagnostic techniques are the same as in uninfected individuals and include bacteriological testing of body fluids, abdominal ultrasound and CT scanning combined with guided-needle aspiration biopsies, barium examination, fibreoptic endoscopy, and laparoscopy. Serological tests such as enzyme-linked immunosorbent assay lack sensitivity due to a poor humoral immune response. Most patients respond well to conventional antituberculous drugs used in standard doses. Some experts recommend a longer course of therapy (9 months). Despite treatment, a few patients experience a downhill course, which may be due to drug resistance or the presence of overwhelming infection.

Prognosis

The prognosis of uncomplicated abdominal tuberculosis is good. Most patients respond well to medical therapy and the outcome of surgical management of complications is generally good. Bowel obstruction or perforation associated with plastic adhesions and the development of enterocutaneous fistulas with intra-abdominal abscesses are associated with increased morbidity and poor prognosis.

General recommendations

The main advance in the diagnostic approach to patients with suspected abdominal tuberculosis is the use of endoscopy. Most patients with peritoneal tuberculosis can be diagnosed by laparoscopy, since the peritoneal appearances are fairly distinctive and the histological yield is high. If this facility is not available, adenosine deaminase in ascitic fluid should be assessed because of its high positive predictive value for tuberculosis, especially in patients without coexisting cirrhosis or HIV infection. In contrast to peritoneal tuberculosis, the diagnosis of gastrointestinal tuberculosis is more difficult. All endoscopically accessible lesions should be biopsied and the specimens used for histology, staining for acid-fast bacilli, and culture. Even if a positive diagnosis of tuberculosis is not made, diseases such as carcinoma and lymphoma can be excluded. The next step in endemic countries is a therapeutic trial with antituberculous drugs. The majority of patients, even those with strictures, show a good response. In the West, where tuberculosis has made a major comeback, the main differential diagnosis is with Crohn's disease, and the clinical, radiological, and histological features of the two are often indistinguishable. Here, a specific diagnostic test for *M. tuberculosis* such as polymerase chain reaction, although still a research tool, may prove extremely useful.

Further reading

Alland D *et al.* Transmission of tuberculosis in New York City—an analysis by DNA fingerprinting and conventional epidemiological methods. *New England Journal of Medicine* 1994; **330:** 1710–16. [Study of an inner-city New York community showing that recently transmitted tuberculosis accounts for 40 per cent of the incident cases and two-thirds of drug-resistant cases.]

Anand SS. Hypertrophic ileocaecal tuberculosis in India with a record of fifty hemicolectomies. *Annals of the Royal College of Surgeons of England* 1956; **19:** 205–9. [One of the largest surgical studies on ileocaecal tuberculosis.]

Anand BS, Nanda R, Sachdev GK. Response of tuberculous stricture to antituberculous treatment. *Gut* 1988; **29:** 62–9. [Long-term follow-up study showing that tuberculous strictures respond well to the use of antituberculous therapy.]

Anand BS, Schneider FE, El-Zaatari FAK, Shawar RM, Clarridge JE, Graham DY. Diagnosis of intestinal tuberculosis by polymerase chain reaction on endoscopic biopsy specimens. *American Journal of Gastroenterology* 1994; **89:** 2248–9. [Successful use of the polymerase chain reaction technique on endoscopic mucosal biopsy specimens to diagnose intestinal tuberculosis.]

Becker JM *et al.* Prevention of postoperative abdominal adhesions by a sodium hyaluronate-based bioresorbable membrane; a prospective, randomised, double-blind multicenter study. *Journal of the American College of Surgeons* 1996; **183**; 297–306. [A study on the use of bioreabsorbable materials in the prevention of postoperative abdominal adhesions.]

Bhargava DK, Shriniwas, Chopra P, Nijhawan S, Dasarathy S, Kushwaha AKS. Peritoneal tuberculosis: laparoscopic patterns and its diagnostic accuracy. *American Journal of Gastroenterology* 1992; **87:** 109–12. [Laparoscopic examination in 38 patients with proven peritoneal tuberculosis.]

Clarridge JE, Shawar RM, Shinnick TM, Plikaytis BB. Large-scale use of polymerase chain reaction for detection of *Mycobacterium tuberculosis* in a routine mycobacteriology laboratory. *Journal of Clinical Microbiology* 1993; **31:** 2049–56. [The use of DNA amplification by PCR for the detection of *Mycobacterium tuberculosis* in a variety of clinical specimens.]

Cole ST *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 1998; **393:** 537–44. [Landmark report on the complete genetic sequencing of *Mycobacterium tuberculosis*.]

Daniel TM, Debanne SM. The serodiagnosis of tuberculosis and other mycobacterial diseases by enzyme-linked immunosorbent assay. *American Review of Respiratory Disease* 1987; **135:** 1137–51. [A comprehensive review article on the serodiagnosis of tuberculosis.]

Initial therapy for tuberculosis in the era of multidrug resistance: recommendations of the Advisory Council for the Elimination of Tuberculosis. *Journal of the American Medical Association* 1993; **270:** 694–8. [An advisory council (comprising representatives of several organizations) recommendation on the initial therapy for tuberculosis.]

Moretti A, Sallustic G, Attimonelli R, Arezzo C, Neri D, Lossito A. Gastric localization of sarcoidosis. *Recent Progress in Medicine* 1993; **84**(11): 750–5. [A study of gastric sarcoidosis.]

Morse DI. Directly observed therapy for tuberculosis. *British Medical Journal* 1996; **312:** 719. [A report on different aspects of directly observed therapy for tuberculosis.]

Pablos-Mendez A *et al.* Global surveillance for antituberculosis-drug resistance, 1994–1997. *New England Journal of Medicine* 1998; **338:** 1641–9. [Results of a WHO sponsored study carried out in several countries on the prevalence of drug-resistant *Mycobacterium tuberculosis*.]

Rangabashyam N, Durairaj KV. Ileocaecal tuberculosis. *Antiseptic* 1975; **5:** 1–4. [A report on the clinical aspects of ileocaecal tuberculosis.]

Rangabashyam N, Prakash A. Intestinal tuberculosis In: Rangabashyam N, ed. *Current topics in surgery*, Vol. 1, pp. 1–10. Macmillan Indian, Madras, 1986. [Study of intestinal tuberculosis with special emphasis on surgical aspects of the disease.]

Raviglione MC, Snider DE, Jr, Kochi A. Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic. *Journal of the American Medical Association* 1995; **273:** 220–6. [Extensive analysis of the morbidity and mortality associated with tuberculosis.]

Schanaider A, Madi K. Intra-abdominal tuberculosis in acquired immunodeficiency syndrome—diagnosis and management. *International Surgery* 1995; **80:** 147–51.[A review of the diagnosis and treatment of abdominal tuberculosis in AIDS.]

Singh V, Kumar P, Kamal J, Prakash V, Vaiphei K, Singh K. Clinicocolonoscopic profile of colonic tuberculosis. *American Journal of Gastroenterology* 1996; **91:** 565–8. [Clinical and colonoscopic features of 62 patients with colonic tuberculosis.]

47.1.4 Tropical ulcer

H. Foster

[Introduction](#)
[Epidemiology](#)
[Acute tropical ulcer](#)
- [Clinical features](#)
- [Aetiology and pathogenesis](#)
- [Diagnosis](#)
- [Treatment](#)

[Chronic tropical ulcer](#)
- [Clinical features](#)
- [Pathogenesis](#)
- [Diagnosis](#)
- [Treatment of chronic tropical ulcer](#)

[Management of malignant change in tropical ulcers](#)
- [Diagnosis](#)
- [Prevention](#)
- [Treatment](#)
- [Lymph node metastases](#)
- [Distant metastases](#)

[Further reading](#)

Introduction

Tropical ulcer is a common disease throughout the wet tropics and subtropics. It is responsible for significant acute and chronic morbidity, particularly in rural communities, and its complications are sometimes lethal.

Acute tropical ulcer is a specific disease entity, whereas chronic tropical ulcer is the end result of a number of disease processes, including acute tropical ulceration, trauma, and infection. Tropical ulcer is known by a variety of local names such as naga sore, Aden ulcer, rice pickers' ulcer, Rhodesia sore, or veldt ulcer. Phagedenic ulcer (from the Greek *phagein*, to eat) is a synonym used particularly by French workers to emphasize the progressive nature of the skin necrosis seen in acute tropical ulcer.

Epidemiology

Tropical ulcer is endemic in village communities throughout the tropics and in some subtropical regions. It is most common in the wet tropics and its incidence rises sharply in the wet season. Wherever nutrition, living conditions, and hygiene are poor, a high incidence of acute tropical ulcer will be found, especially amongst children and young adults, who suffer repeated trauma to their legs during their daily work in the field or rain forest. Medical services are often deficient in these areas and patients are rarely able to leave their work for periods of rest or treatment. Repeated trauma and infection compound the problem and lead to a high incidence of chronic ulcer.

Reliable statistics are generally not available, but a high incidence of tropical ulcer has been reported from areas of South America, West Indies, Africa, India, Vietnam, the Philippines, Malaysia, Indonesia, Melanesia, and the West Pacific. Studies from different parts of Africa indicate that up to 30 per cent of presentations to hospital are for tropical ulcer. Presentations to local aid posts and traditional healers are probably even more common. Most studies show the disease is more common in males. Epidemics of acute tropical ulcer have occurred in Africa, Melanesia, and the Pacific Islands. Tropical ulcer was common amongst jungle fighters, and especially prisoners, in the tropics during the Second World War. Squamous-cell carcinoma arising in chronic tropical ulcer is the most common skin cancer amongst dark-skinned inhabitants of the tropics.

Acute tropical ulcer

Clinical features

Acute tropical ulcer results from a full-thickness, necrotizing bacterial infection of the skin. In most cases, a minor puncture wound or dirty laceration provide an entry portal and initiate a spreading bacterial cellulitis that rapidly produces dermal gangrene, probably by a direct cytotoxic effect (Fig. 1).



Fig. 1. Acute necrotizing infection of the skin of the lower limb a few days after a minor penetrating injury.

Initially, a trivial injury to the leg may be disregarded and the patient remain ambulant. Saliva may be used to dress the wound. The necrotizing infection begins in a few days with the development of a painful papule and foul-smelling, watery discharge or local blistering. A central area of gangrene rapidly becomes evident and the process spreads at a variable rate until it is brought to a halt, presumably by the host's defence mechanisms, and demarcation occurs. The patient may be febrile at this stage and there is an acute inflammatory response in surrounding skin. A hard slough has usually formed after about a week and separation begins (Fig. 2). Lymphadenitis is not prominent at any stage. After about 2 weeks, slough separation reveals a layer of foul, yellow or grey–green pus adherent to friable, infected granulation tissue (Fig. 3). In most cases the infection slowly resolves, leaving a bed of healthy granulation tissue surrounded by normal skin, which may show signs of early epidermal growth and migration from the ulcer margin; this constitutes a typical acute tropical ulcer (Fig. 4).



Fig. 2. An acute tropical ulcer following demarcation of necrotic skin with slough separation well advanced.

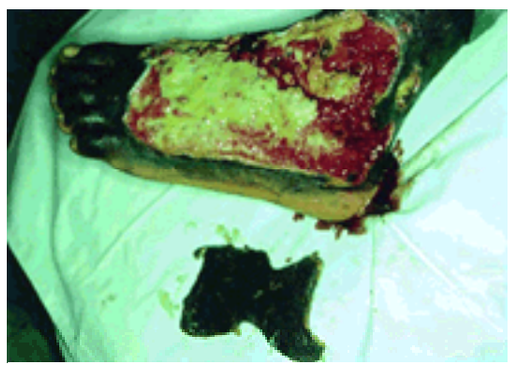


Fig. 3. Slough removal reveals an infected, granulating ulcer floor with foul exudate and pus.

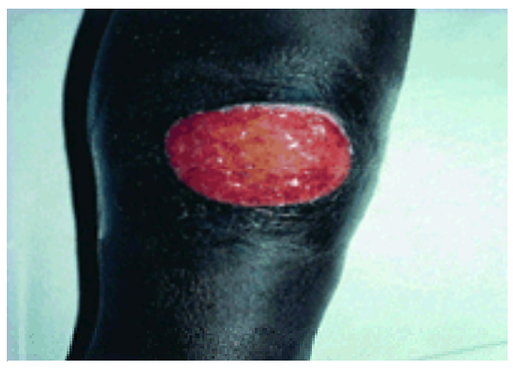


Fig. 4. A typical acute tropical ulcer with a clean, granulating floor and normal surrounding skin.

The ulcer is virtually always on the leg below the knee, common sites being the lower one-third just above the malleoli, on the shin or dorsum of the foot. Tropical ulcers rarely occur above the knee and necrotizing skin infections of the trunk or upper limbs are best regarded as separate clinical conditions. If the ulcer remains unhealed at 3 months, it is regarded as chronic.

Aetiology and pathogenesis

Full-thickness skin necrosis is produced by a synergistic bacterial infection in a host who may be predisposed by poor nutrition or intercurrent illness.

Bacteriology

Since 1899, researchers have observed large numbers of fusiform bacilli and spirochaetes in the early stages of the disease. These were thought to be the same organisms as those found in patients with Vincent's stomatitis, gingivitis and cancrum oris, and known as *Fusobacterium* spp. and *Treponema vincenti*. Similar organisms have been isolated from stagnant water and mud. Acute ulcers have been produced experimentally in man and animals using material from acute tropical ulcers or patients with acute ulcerative gingivitis. It is thought the pathogens may gain access to a skin breach by direct inoculation at the time of injury or with the help of flies feeding on minor leg wounds. Some cases may be infected from the patient's own mouth if saliva is used as a dressing for a leg wound. Coliforms and Gram-positive cocci have also been isolated frequently from early ulcers.

In 1987, workers using anaerobic culture techniques identified and characterized a new species from acute ulcers known as *Fusobacterium ulcerans*. This bacterium is markedly cytotoxic to animal and human cell lines in tissue culture. A combination of aerobic and anaerobic culture techniques with light and electron microscopy has demonstrated anaerobes in combination with aerobes or facultative anaerobes, especially in early ulcers. The anaerobes are usually a combinations of fusobacteria, *Bacteroides* spp., or anaerobic cocci; they are always found in the presence of aerobes. Spirochaetes are also identified by microscopy in early ulcers. It is thought that aerobes and anaerobes act synergistically to produce the dermal gangrene characteristic of acute tropical ulcer. The role of the spirochaetes in the pathogenesis of acute tropical ulcer is not known.

Streptococci, diphtheroids, and other bacteria are isolated with variable frequency from established ulcers; their significance is uncertain.

Predisposing factors

Repeated trauma to the legs provides numerous opportunities for organisms to penetrate the epidermis and perhaps reduces local defence mechanisms. Dirt, damp, and a high ambient temperature may provide an environment rich in the pathogens specific for acute tropical ulcer and suitable conditions for their rapid proliferation.

Population studies support an association with poor nutrition (especially protein and vitamin deficiency), and intercurrent illnesses such as malaria, respiratory infection, and anaemia; however, evidence of a causative effect is generally lacking, and certainly at a clinical level many patients with tropical ulcer are not overtly malnourished or debilitated.

Pathogenesis

The similarity between acute tropical ulcer and other acute necrotizing skin conditions such as cancrum oris, necrotizing fasciitis, Fournier's gangrene, and Meleney's synergistic gangrene is striking. All these conditions are probably produced by a mixed infection with aerobic and anaerobic bacteria. It has been postulated that tissue oxygen tension is reduced by proliferating aerobic bacteria, thereby promoting rapid growth of the anaerobic species. In the case of acute tropical ulcer, toxins produced by the anaerobe *F. ulcerans* probably cause rapid tissue necrosis by a direct cytotoxic effect.

Histologically the acute lesion shows the features of gangrene, with evidence of skin necrosis and infection. An acute inflammatory response is most evident at the periphery of the lesion and is presumably important in halting the spreading infection. After the slough has separated, the ulcer is covered with a layer of fibrin, necrotic debris, and organisms. As the infection resolves, granulation tissue rapidly fills the base of the ulcer and epithelial proliferation and migration begins at the periphery (Fig. 5). Untreated, the acute ulcer will evolve along one of the following three paths. Many ulcers will heal completely by epithelial ingrowth from the periphery (Fig. 5). The new epithelium is thin, lacks adnexal structures, and is often completely devoid of pigment. It has little resistance to further trauma, infection, and ultraviolet radiation but may remain intact over the years. In many cases, initial healing is followed by repeated episodes of trauma, infection, and solar damage, which produce repeated ulceration. In a significant minority, healing is never achieved and chronic ulceration is established (Fig. 6).



Fig. 5. An untreated tropical ulcer healing by granulation and epithelial growth and migration from the skin edge.



Fig. 6. A typical chronic tropical ulcer with an indolent, fibrotic base and evidence of repair in the presence of continuing tissue damage. There are adjacent areas of scarring with depigmented atrophic epithelium.

Diagnosis

Most patients present soon after slough separation with an infected granulating ulcer of the lower leg. The history, site, and appearance of the ulcer are usually typical, and allow an accurate diagnosis on clinical grounds (Table 1). Early presentation with an infected papule or minimal skin necrosis is not so common. A full history and clinical examination of the patient will reveal any predisposing factors such as malnutrition and intercurrent illness.

Condition	Differentiating features
Primary yaws	No skin necrosis Ulcers have a typical yellow crust Rapid response to depot penicillin
Mycobacterial ulcers	Extensive undermining of adjacent skin No skin necrosis
Anaerobic skin infection	Associated anaerobic infection of gastrointestinal tract Microscopy
Non-specific infective and traumatic ulcers	History of trauma, No skin necrosis, Variety of "non-specific" pathogens
Venous	Associated varicose veins/incompetent perforating veins, History of deep venous thrombosis
Neoplastic	Raised, rolled, or everted edge Indurated Histological diagnosis

Table 1 Differential diagnosis of acute tropical ulcer

Special tests

Swabs should be taken from the edge and floor of the ulcer for microscopy, culture (aerobic and anaerobic), and sensitivity. Special stains and dark-ground microscopy may be helpful to demonstrate spirochaetes and fusiform bacilli. Biopsy may be indicated if the clinical diagnosis is uncertain or neoplasm is a possibility. Haematological and other tests may be indicated to elucidate associated conditions.

Treatment

Prevention

Improved nutrition, hygiene, and socioeconomic conditions are suggested as factors required to reduce the incidence of acute tropical ulcer. Use of suitable footwear and long trousers, with improved working conditions, should reduce the incidence of trauma to the legs and subsequent wound contamination.

Prompt, appropriate treatment of minor leg wounds should reduces the incidence of exogenous and endogenous infection progressing to acute tropical ulcer.

Treatment of the acute lesion

Patients who present early with an infected papule should be treated promptly with systemic antibiotics. Crystalline penicillin, often with metronidazole, will abort the infection and prevent skin necrosis and ulceration in many cases.

Most patients present later; once skin necrosis is established, surgical debridement with removal of slough and non-viable tissue is required to control infection. Patients who are septicaemic or have large areas of skin necrosis require urgent admission for bed rest, elevation of the limb, and systemic administration of antibiotics before surgery.

Once any spreading infection has been controlled and necrotic skin removed, local measures are used to eradicate the infection. Debridement is followed by frequent dressings, with rest and elevation of the limb. Antibiotics should be stopped. Many substances have been used to dress tropical ulcers, including paw-paw, pineapple pulp, honey, various enzymes, and antibiotics. Dilute eusol (Edinburgh University solution of lime) is an effective and cheap application for removing residual slough. Isotonic, stabilized sodium hyopochlorite solution is also effective in controlling local infection and removing slough. The dressing of choice should be effective, sterile, cheap, non-toxic, non-allergenic, painless, and easily applied. Frequent changes of moist dressings are the key to rapid control of local infection. Isotonic saline should be used once the ulcer is clean and granulating. Topical antibiotics are not effective and lead to colonization of the ulcer by resistant bacteria; their use is contraindicated.

Healing is completed by providing epithelial cover to the clean, granulating ulcer. Spontaneous epithelialization will occur at about 1 mm/day from the ulcer edge (Fig. 5). Small ulcers may be allowed to heal in this way, but the result is a thin, unstable, depigmented scar that is susceptible to repeated ulceration caused by mechanical trauma, infection, and solar radiation. Autografts of split-thickness skin allow rapid closure of the defect and produce a more stable, pigmented scar. It would be ideal to close these ulcers with normal, full-thickness skin, but this is not currently feasible for most patients with tropical ulcer. Split-thickness skin grafting is relatively cheap, widely available, and has a high rate of successful take. The healed graft lacks hair follicles, sweat and sebaceous glands, but is pigmented and so has at least some

protection from solar damage and carcinogenesis. This is the most widely used technique for achieving cover of acute ulcers (Fig. 11).



Fig. 11. A chronic tropical ulcer treated by split-thickness skin grafting. The central skin graft is pigmented while the surrounding regenerated epithelium is devoid of melanin and highly susceptible to malignant change.

As a general rule, ulcers of 1 to 2 cm or more in diameter should be closed promptly by a split skin graft once infection has been controlled. Patients must be advised to protect the healed graft permanently from trauma, infection, and solar exposure.

Chronic tropical ulcer

Many acute tropical ulcers fail to heal and others breakdown after repeated episodes of trauma and infection. After an arbitrary period of 3 months they may be defined as chronic. Trauma, non-specific infection, and various other disease processes also give rise to chronic leg ulcers clinically indistinguishable from chronic tropical ulcer, except perhaps on the basis of history. For practical reasons, all such non-specific ulcers are usually included in the definition of chronic tropical ulcer.

Clinical features

Chronic ulcers show macroscopic evidence of chronic inflammation, with repair proceeding in the presence of active infection and tissue damage. The granulating ulcer floor is often pale, indolent, and firm. There may be old scars from healed ulcers that have broken down on several occasions. Regenerated epithelium is often thin, atrophic, and devoid of all pigment (Fig. 5). The ulcer base is usually attached to the tibial periosteum, and bone changes due to periosteal reaction and osteitis may be marked. Dense fibrosis may produce crippling contractures or lead to lymphatic obstruction with lymphoedema and secondary skin changes.

After several years, regenerating epithelium at the ulcer edge may show evidence of poorly controlled growth with heaping up of the ulcer margin or migration of abnormal epithelium across the ulcer floor (Fig. 7). Histological studies at this stage will usually reveal pseudoepitheliomatous hyperplasia, which seems to grade imperceptibly into very well-differentiated squamous-cell carcinoma as time passes (Fig. 8).

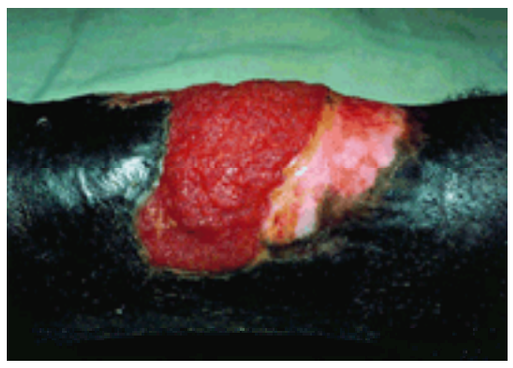


Fig. 7. A chronic tropical ulcer showing depigmented epithelium heaped up at the ulcer edge. Histology showed pseudoepitheliomatous hyperplasia.

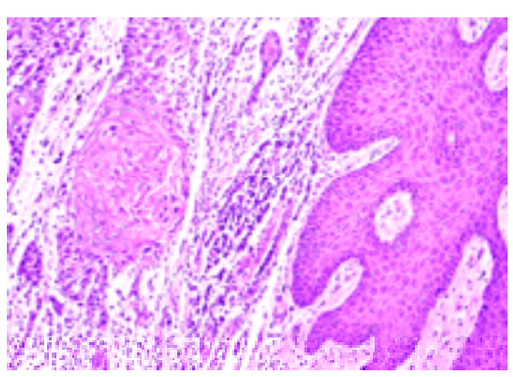


Fig. 8. Histological section of a malignant tropical ulcer showing pseudoepitheliomatous hyperplasia adjacent to squamous-cell carcinoma. Original magnification ×100.

Pathogenesis

Many acute ulcers never heal but become chronic, indolent lesions. Healing is thwarted by repeated injury due to infection, mechanical trauma, and solar radiation. Chronic inflammation, dense fibrosis, bone changes, dystrophic calcification, lymphatic obstruction, and epidermal dysplasia all become evident and render healing unlikely.

Alternatively, repeated episodes of ulceration are followed by more prolonged periods of healing and eventually chronic ulceration. The fragile, depigmented epithelium of a healed ulcer on the lower leg is easily breached by the combined effects of mechanical trauma, infection, and solar radiation. Each episode of ulceration leads to more marked fibrosis and epithelial dysplasia, with less orderly healing. Most ulcers overlie the tibia and cause ‘reactive' bone changes with osteitis and new bone formation.

The mechanism of malignant transformation in tropical ulcer is not known; however, the common antecedent to the development of squamous-cell carcinoma seems to be the depigmented, atrophic epithelium associated with both healed and chronic ulcers (Fig. 7 and Fig. 8). Almost all chronic tropical ulcers show marked pseudoepitheliomatous hyperplasia, which may evolve gradually into frank malignancy. An analogy with malignant change in burn scars, chronic osteomyelitic sinuses, fistulas, and other chronic ulcers is clear; malignant change in chronic tropical ulcers is a form of Marjolin ulcer. Most malignant tropical ulcers are well differentiated histologically. Progression is often slow but relentless, with invasion of underlying bone and eventual pathological fracture. Metastasis to the inguinal lymph nodes is usually manifest only when the primary is advanced, and distant spread is rarely a clinical problem.

Diagnosis

Some patients will give a clear history of preceding acute tropical ulcer. However, many will not be able to give a diagnostic history of the initial ulceration, which may have occurred as much as 50 years previously. Specific causes should be considered in all chronic ulcers (Table 2). Clinical evidence of long-term complications, particularly malignant change, must be sought.

Condition	*Differentiating features*
Chronic yaws	Positive serology
Gummatous (tertiary) syphilis	Positive serology
Mycobacterial ulceration	Skin edge markedly undermined. Positive bacteriology (usually *Mycobacterium ulcerans*, but occasionally *M. tuberculosis*)
Underlying bony disease	Radiographs may reveal osteomyelitis, malunited fracture, or chronic osteomyelitis
Venous ulcers	Varicose veins and/or incompetent perforating veins
Haematological and immune disorders	Features of sickle-cell anaemia, tropical splenomegaly syndrome, Felty syndrome, or other underlying disease
Occlusive arterial disease	Reduced peripheral pulses, arterial Doppler studies, angiography. Associated hypertension and smoking
Lymphatic obstruction	Chronic lymphoedema (congenital or filarial) with secondary infection and ulceration
Leprosy	Ulcers are usually on the sole of the foot and associated with peripheral neuropathy and other stigmata of the disease
Foreign body	History of trauma, radiological or ultrasonographic evidence, surgical exploration on suspicion
Self-mutilation (rare)	Bizarre history, atypical lesion, recurrence despite adequate treatment, observation of self-injury

Table 2 Differential diagnosis of chronic tropical ulcer

Special tests

Swabs from the edge and centre of the ulcer should be taken for bacteriological studies including microscopy, cultures and sensitivity, as well as studies for acid-fast bacilli. Biopsy is indicated where areas of depigmentation and tissue proliferation raise the suspicion of malignant change (Fig. 8, Fig. 9), or if diagnosis is uncertain. Radiography may be indicated to show bone changes, especially malignant erosion, pathological fracture, inflammatory changes, or dystrophic calcification. (Fig. 10).

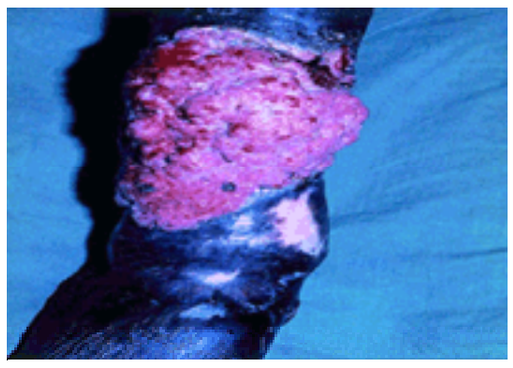


Fig. 9. A large, exophytic carcinoma of the shin arising in a tropical ulcer present for more than 40 years, with adjacent depigmented epithelium. Treated by amputation.

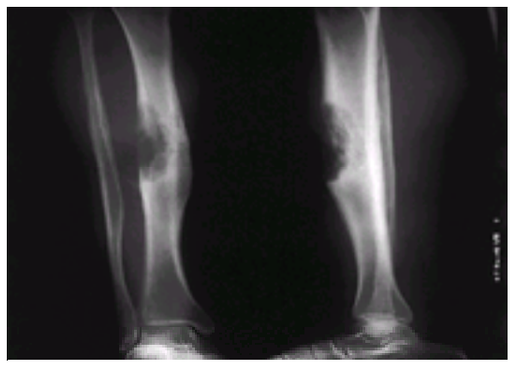


Fig. 10. Radiograph showing erosion of the anterior tibial cortex by a malignant tropical ulcer.

Haematological and other studies may be indicated to elucidate intercurrent illness and nutritional status.

Treatment of chronic tropical ulcer

Prevention

Treatment of acute tropical ulcer that achieves good epithelial cover and the avoidance of subsequent trauma and infection to the leg will prevent chronic or recurrent ulcer.

Treatment of the ulcer

Treatment of uncomplicated chronic tropical ulcer comprises three main aspects: control of infection, control of any factors hindering healing, and provision of a robust epithelial cover that will protect against further ulceration and malignant change.

Control of infection requires admission to hospital for rest in bed and elevation of the limb, with frequent dressing of the ulcer. Some patients require systemic treatment with antibiotics, chosen on the basis of the response of the cultured organisms to antibiotics in the laboratory.

Excision of indolent granulation tissue, areas of dense fibrosis, and severely scarred, depigmented skin may be essential to achieve ultimate healing. Contractures and bone deformity will usually need to be corrected before the ulcer is covered. Bone infection must be controlled.

Epithelial cover is usually achieved by applying a split-thickness skin graft to the granulating bed of the ulcer or after excision of the ulcer. There is a higher rate of graft loss than with acute ulcers, and the healed graft has all the disadvantages mentioned under the treatment of acute ulcers. Delayed graft breakdown and further ulceration are quite common, as split skin grafting does not provide adequate cover in many situations.

Free, full-thickness skin grafts generally will not take on the ulcer bed, and local pedicle flaps in the lower leg are rarely adequate for the large defects present. Cross-leg flaps have been used in many tropical countries, but have several major disadvantages. Composite, free tissue autograft with microvascular anastomosis (e.g., free rectus abdominis graft) is probably the most effective method of providing a robust repair for a large defect of the leg, but is technically demanding and has not been widely used in this context; it is perhaps more appropriate in the treatment of malignant tropical ulcers.

Treatment of chronic tropical ulcer is demanding, particularly where surgical resources are limited. Permanent, stable repair of the defect is essential to prevent long-term morbidity and return the patient to normal activities. Recurrent and chronic ulceration are crippling and associated with a high incidence of malignant change.

There is potential for major improvement in the results currently achieved in treating most patients with chronic tropical ulcer.

Management of malignant change in tropical ulcers

Diagnosis

All chronic tropical ulcers should be regarded as premalignant. The latent period from the first ulcer is rarely less than 5 years and is often about 30 or 40 years. Patients with white patches of atrophic epithelium seem particularly prone to malignant change (Fig. 6, Fig. 7, Fig. 9).

Clinical diagnosis of the typical raised, everted edge or invasive ulceration of squamous-cell carcinoma is not difficult, but the index of suspicion must be high and all suspect areas biopsied. Complete excision of the ulcer for histological examination may be necessary. Occasional confusion with chronic yaws, SYphilitic gumma, or chronic osteomyelitis should be clarified by histology.

The clinical and histological differentiation between squamous-cell carcinoma and pseudoepitheliomatous hyperplasia may be difficult at times. Sampling errors and difficulties with histological interpretation may lead to underdiagnosis of squamous-cell carcinoma; the clinical behaviour of the lesion is the final indicator of its true nature. Eight cases of sarcoma apparently arising in chronic tropical ulcers have been reported, but virtually all malignant tropical ulcers are squamous-cell carcinoma.

Prevention

It is assumed that adequate treatment of acute and chronic ulcers prevents neoplastic change in the long term, but there are no reliable studies to prove this, and squamous-cell carcinoma is occasionally observed in patients who have undergone previous split skin grafting. Protection of chronic and healed ulcers (particularly white patches of skin) from the intense solar radiation of the tropics is thought to be very important in preventing malignant change.

Treatment

The primary carcinoma

On presentation, most patients have large squamous-cell carcinomas with bone attachment or invasion (Fig. 9, Fig. 10). Despite advanced local disease, metastatic disease is not common and potentially curable primary lesions should be treated by surgery in otherwise well patients.

Small, localized lesions may be treated by wide excision and split-thickness skin grafting, but most lesions require more extensive surgery. Amputation has been the mainstay of treatment for large lesions with extensive bone invasion, but the associated morbidity is poorly tolerated in rural communities and alternatives have been sought for suitable patients. Wide excision of the carcinoma with surrounding skin, muscle, and tibial cortex leaves a large defect that is difficult to repair (Fig. 12). One approach is to allow the defect to granulate and eventually achieve epithelial cover with a split skin graft. Although the limb is salvaged, it may be fragile and prone to further ulceration and pathological fracture. In recent years, large defects of the leg have been successfully repaired using free composite autografts with microvascular anastomosis. Techniques such as the free rectus abdominis flap have proven to be versatile, reliable, and technically feasible in a provincial hospital setting; a number of such repairs have been performed in suitable patients with good results (Fig. 13). Widespread use of such limb-saving techniques has great potential for reducing the disability caused by chronic and malignant tropical ulcer.



Fig. 12. A large surgical defect of the lower leg and anterior tibia following wide excision of a malignant tropical ulcer.

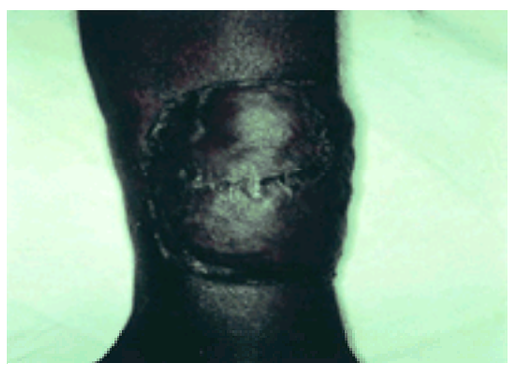


Fig. 13. Satisfactory healing 3 months after wide excision and repair of a malignant tropical ulcer. A free rectus abdominis muscle graft with microvascular anastomosis was used with split skin graft for epithelial cover.

Wide excision and free-flap repair allows limb salvage in patients with moderately advanced carcinomas, but amputation remains indicated for many extensive lesions, pathological fractures, and most undifferentiated carcinomas.

Lymph node metastases

Patients with malignant change in a tropical ulcer may develop enlarged inguinal lymph nodes due to infection or metastasis. Clinical observation and response to treatment with antibiotics may suggest the diagnosis of metastasis to the nodes. Cytology of a fine-needle aspirate from the nodes may be helpful in confirming the diagnosis.

Evidence of disseminated malignancy is usually a contraindication to radical local treatment of the primary lesion or of any inguinal lymph-node metastases. Otherwise, patients with involved but operable nodes should be subjected to radical groin dissection with the aim of controlling the disease locally and obtaining prognostic information by histological examination of the operative specimen. A high risk of local recurrence in the groin is indicated by massive or multiple nodal metastases or invasion through the lymph node capsule; postoperative radiotherapy can reduce the incidence of recurrence in similar patients with high-risk squamous-cell carcinoma, but has not been used widely in patients with malignant tropical ulcer.

Patients whose inguinal nodes are not clinically involved should be followed at regular intervals and advised to return promptly if they become aware of a groin mass. Prophylactic groin dissection is generally not indicated as the incidence of occult nodal metastasesis low; it may be considered for patients with advanced or undifferentiated primary lesions or those who are not likely to return for review. However, the value of this procedure has not been proved in malignant tropical ulcer.

Distant metastases

Treatment of patients with disseminated disease by palliative immuno-, radio-, or chemotherapy is rarely indicated or appropriate in most developing countries; these techniques have little effect on survival, are expensive, often toxic, and may keep the dying patient away from their family. A generous supply of an effective oral analgesic is probably the most valuable treatment that can be offered.

Further reading

Adeloye A, ed. *Davey's companion to surgery in Africa*, 2nd edn, pp. 39–45. Churchill Livingstone, London, 1987. [Clinical aspects of tropical ulcer in Africa.]

Adriaans B. Tropical ulcer—a reappraisal based on recent work. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1988; **82:** 185–9. [Overview of more recent work on pathogenesis of tropical ulcer.]

Adriaans B, Garelick H. Cytotoxicity of *Fusobacterium ulcerans. Journal of Medical Microbiology* 1989; **29:** 177–80. [How the Fusobacterium causes skin necrosis.]

Basset A. Tropical phadenic ulcer. In: Simons RDGP, Marshall J, ed. *Essays on tropical dermatology*, pp. 25–33. Excerpta Medica, Amsterdam, 1969. [A good overview.]

Bulto T, Maskel FH, Fisseha G. Skin lesions in Gambela, with special emphasis on the epidemiology of tropical ulcer. *Ethiopian Medical Journal* 1993; **31**(2): 75–82. [Gives some perspective on the incidence of tropical ulcer in Africa.]

Clements FW. Tropical ulcer with special reference to its aetiology. *Medical Journal of Australia* 1936; **2**(19): 615–43. [An early account of tropical ulcer.]

Connor DH, Neafie RC. Tropical phagedenic ulcer. In: Binford CH, Connor DH, ed. *Pathology of tropical and extraordinary diseases*, Vol. 1, pp. 199–201. Armed Forces Institute of Pathology, Washington DC, 1976. [A good account of histopathology of tropical ulcer.]

Foster HM, Webb SJ. Malignant change in chronic tropical ulcers. *Asian Journal of Surgery* 1984; **7**(2): 232–6. [A clinico-pathological study from Papua New Guinea.]

Loewenthal LJ. Tropical phagedenic ulcer—a review. *International Reviews of Tropical Medicine* 1963; **2:** 267–91. [A good but early overview.]

Morris GE, Hay RJ, Srinivasa A, Bunat A. The diagnosis and management of tropical ulcer in East Sepik Province of Papua New Guinea. *Journal of Tropical Medicine and Hygiene* 1989; **92:** 215–20. [How tropical ulcer is managed in west Papua New Guinea.]

O'Brien HD. Treatment of tropical ulcers. *British Medical Journal* 1951; **2:** 1544–51. [An early account of treatment.]

Robinson DC, Adriaans B, Hay RJ, Yesudian P. The clinical and epidemiological features of tropical ulcer (tropical phagedenic ulcer). *International Journal of Dermatology* 1988; **27**(1): 49–53. [A useful account of acute tropical ulcer.]

Taylor GI, Corlett RJ, Boyd JB. The versatile deep inferior epigastric (inferior rectus abdominis) flap. *British Journal of Plastic and Reconstructive Surgery* 1984; **37**(3): 330–50. [A reliable free vascular flap suitable for repairing surgical defects in the lower leg.]

47.1.5 Otitis media in high-risk populations

Peter S. Morris

Introduction
Natural history
Diagnosis
Treatment
Further reading

Introduction

The World Health Organization has advised that populations where chronic perforation of the tympanic membrane affects more than 4 per cent of children have a massive public-health problem requiring urgent attention. While dysfunction of the eustachian tube is an important aetiological factor, high rates of severe disease are primarily a consequence of the increased colonization of the upper airway and spread of respiratory bacteria (*Streptococcus pneumoniae, Haemophilus influenzae*, and *Moraxella catarrhalis*), and the lack of access to appropriate antibiotic therapy. Middle-ear infections are closely related to the other common suppurative (‘pus-producing') infections of the respiratory tract: rhinosinusitis (nasal discharge) and bronchitis (cough). High-risk populations also experience high rates of pneumonia, bacteraemia, and meningitis. Since bacterial acute respiratory infection is believed to be the most frequent cause of early childhood death worldwide, more effective prevention and treatment programmes are an international priority.

High rates of severe otitis media are best addressed through better primary health care. The following evidence-based management guidelines have been developed for Australian Aboriginal Health Workers. Like all practice guidelines, they should be adapted to suit local needs and resources. Additional details relevant to the surgical aspects of otitis media are provided within the sections on paediatric surgery and head and neck surgery.

Natural history

Early, heavy bacterial colonization of the nasopharynx is associated with the development of otitis media with effusion. This may persist for many years, resulting in a dull, featureless, thickened eardrum. Bulging of the eardrum and perforation (acute otitis media) in the first year of life are also common. These events are relatively asymptomatic, and carers will often regard the health of their infants as satisfactory. The initial perforation usually occurs in the anterior part of the drum and is small. The drum may heal and reperforate several times before chronic discharge is established. If a discharging perforation persists beyond 2 weeks, secondary infection with a range of pathogens including *Pseudomonas, Proteus*, and *Staphylococcus* spp. is likely. These organisms significantly alter the antibiotic-sensitivity profiles, with *Pseudomonas aeruginosa* in particular being susceptible to a limited range of antibiotics. Chronic suppurative otitis media is associated with continuing middle-ear damage and an increase in size of the perforation. In the most severe cases, the whole drum and adjacent ossicles are completely eroded.

Diagnosis (Fig. 1)

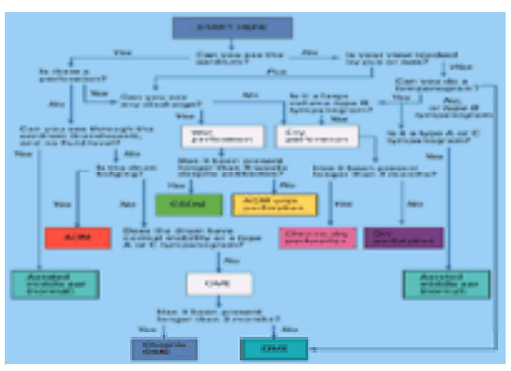


Fig. 1. Flow diagram for the diagnosis of middle-ear disease in children of less than 8 years of age living in a high-risk population. Definitions and management strategies for final ear diagnoses (shown in coloured boxes) are described in Table 1.

The assessment of middle-ear function is the most problematic aspect of effective primary care. Examination occurs with the child sitting comfortably in the carer's (or assistant's) lap with the child's head against the carer's chest. The head and arms should be held firmly enough to limit any unexpected movements. The use of pneumatic otoscopy or tympanometry is necessary to detect accurately the presence of otitis media with effusion, and the ability to clean ears of wax or purulent discharge is an important skill. The presence and extent of suppurative disease, and the level of hearing loss, are the two most important factors determining management.



Table 1 Recommended initial management strategies for different ear states in high-risk populations

AOM = acute otitis media; CSOM = chronic suppurative otitis media; OME = otitis media with effusion

A good understanding of how otitis media can affect hearing is essential. Middle-ear disease is associated with a hearing loss of anywhere between 0 and 60 dB. A hearing loss of 30 dB will mean that whispers will not be heard, and a conversational voice will sound like a whisper. At 60 dB, conversational voice is not heard and shouts are heard at the level of a whisper. Because hearing loss can result in language delay and learning difficulties, experts currently recommend that a hearing loss of greater than 20 dB in the ear with the best hearing that lasts longer than 4 to 6 months should be treated. The younger the child and the greater the degree of hearing loss, the more important it is to provide effective interventions. The first three years of life are critical for language development and all children should achieve the following milestones: 7 to 10 months, babbling sounds (*ba-ba, ma-ma, da-da*); 18 months, understands up to 50 words; 2 years, uses two-word combinations (‘go bye-bye' for example); 3 years, knows several colours, uses short three- to four-word sentences; 4 years, tells stories, sings songs.

Children with normal, aerated middle ears should have hearing thresholds within normal limits (0–15 dB). For children with bilateral otitis media with effusion, the mean hearing loss will be around 25 dB. However, around a third of these children will have hearing within normal limits. For children with bilateral chronic suppurative otitis media, the mean hearing loss is around 35 dB. Very few of these children will have normal hearing and some will have a hearing loss greater than 50 dB. Although the level of hearing loss can be measured accurately in all children, the impact of ear disease on individual children and families is less predictable. Therefore, a child with a hearing loss of 20 dB who also has speech delay or any other form of developmental delay should be managed more aggressively. By the same token, if a child has

chronic suppurative otitis media and a hearing loss of 50 dB but is regarded as perfectly normal by their family, adherence to recommended antibiotic therapy and audiological interventions is likely to be poor. The challenge for primary health-care workers is to provide information in such a way that families have a clear understanding of the potential benefits of treatment.

Treatment

An approach to the initial management of each of the common ear states is described in Table 1. For all conditions there are three areas that can be addressed: (1) ear health and hearing education; (2) medical interventions; (3) audiological interventions. From the clinical perspective, most cases of non-suppurative disease should be managed according to the degree of hearing loss, while suppurative diseases will also require antibiotic therapy.

Ear health and hearing education should be provided to all families even if the child has normal ears (Box 1). Concentrate on explaining how hearing loss can affect learning and language development in young children, and the need to go to the clinic and for antibiotic treatment if the child has an ear infection. Encourage strategies that keep the face and hands of young children free of nasal secretions, thus limiting the spread of infection. Families should be informed that a mild hearing loss (20–40 dB) often cannot be detected by families or teachers, and that behavioural strategies can reduce any adverse developmental effects; for example, talk loudly and clearly, engage eye contact, limit background noise, repeat instructions as necessary, seat at front of class.

Box 1 The principles of managing otitis in a high-risk population.

1. Include the prevention and early treatment of ear disease in maternal and child health programs.
2. For an individual child:
   - Identify suppurative and non-suppurative complications of infection.
   - Consider the presence of other diseases, the duration of ear disease, the level of hearing loss, and the impact of the disease on the child and family.
   - Use the clinical assessment to estimate the likelihood of benefit from the different interventions available.
3. Be aware that this is a chronic disease in most children. Effective communication with the family is essential to ongoing surveillance and adherence with recommended therapy.

Medical interventions include treatment with antibiotics and surgery. Children with suppurative infections should receive appropriate antibiotics until they are completely better. While ear, nose, and throat specialists can advise on diagnosis and the state of the middle ear in young children who are difficult to examine, routinely recommended surgical interventions are limited to the insertion of grommets. Adenoidectomy is a treatment option in children older than 3 years of age who have chronic otitis media with effusion persisting after grommet surgery; and surgical repair of a perforated eardrum (tympanoplasty) is usually delayed until at least 10 years of age.

Audiological interventions should be recommended for all children with a hearing loss of greater than 20 dB that lasts longer than 3 months, or earlier if the hearing loss is affecting learning or language development. The behavioural interventions described above are extremely important. Audiologists are also able to offer a range of assistive hearing devices suitable for group settings or individual children. Soundfield amplification systems can be used in school classrooms to increase the volume of the teacher's voice by 10 to 15 dB. While this level of amplification is not enough for many children, these systems are highly recommended in schools where more than half the children have a hearing loss greater than 15 dB. Children with a hearing loss greater than 30 to 35 dB will also require a personal amplification system. If the child is at school, adapted FM radios allow the microphone worn by the teacher to transmit sounds directly to the child's headset. Children with a persistent hearing loss of greater than 50 dB, or any child with a hearing loss adversely affecting language development, should be fitted with a hearing aid. Since many of these children will have chronic suppurative otitis media, a bone-conduction hearing aid is frequently more appropriate than body-worn, behind-the-ear, or in-the-ear hearing aids. Unfortunately, many countries have insufficient resources to supply even hearing-aid batteries.

More effective treatment of suppurative otitis media is essential to prevent the development of chronic suppurative otitis media. All children with acute otitis media should be reviewed 4 to 7 days after starting treatment, by which time clinical improvement should be obvious. The failure of antibiotic therapy is explained by one or more of the following:

- inadequate adherence to therapy
- inadequate dose
- inadequate spectrum of antibiotic activity
- inadequate length of treatment
- the mucosa is too damaged to allow a complete cure.

At each review, the suppurative infection should be classified as resolved, still present (persistent disease), or getting worse (progressive disease), according to the degree of discharge, bulging or inflammation of the tympanic membrane. Children with persistent disease (excluding perforation) should be provided with a longer course of antibiotics at a higher dose along with strategies to improve adherence. Children with progressive disease or persistent perforation should also have the spectrum of antibiotic activity broadened to cover b-lactamase-producing organisms. Because the extent of mucosal damage will also influence the effectiveness of treatment, antibiotics should be continued for at least 6 weeks before a diagnosis of chronic suppurative otitis media is made. In communities with high rates of suppurative disease and low rates of adherence to therapy, the introduction of supervised dosing through community- or school-based programmes may be required.

Further reading

Acuin J, Smith A, McKenzie I. *Treatment of chronic suppurative otitis media [a Cochrane Review]*, Issue 1, *The Cochrane Library*. Update Software, Oxford, 1999. [Systematic review of randomized, controlled trials.]

Bluestone CD. Epidemiology and pathogenesis of chronic suppurative otitis media: implications for prevention and treatment. *International Journal of Pediatric Otorhinolaryngology* 1998; **42**(3): 207–23. [Narrative review.]

Bluestone CD, Klein JO. *Otitis media in infants and children*, 2nd edn. Saunders, Philadelphia, 1995. [Comprehensive textbook on medical aspects of otitis media.]

Freemantle N *et al.* The treatment of persistent glue ear in children. *Effective Health Care Bulletin No. 4*. NHS Centre for Reviews and Dissemination, University of York, 1992. [Systematic review of randomized, controlled trials of surgery for otitis media with effusion.]

Glaziou PP, Hayem M, Del Mar CB. *Treatments for acute otitis media in children: antibiotic versus placebo [a Cochrane Review]*, Issue 1, *The Cochrane Library*. Update Software, Oxford, 1999. [Systematic review of randomized, placebo-controlled trials.]

Leach AJ, Boswell JB, Asche V, Nienhuys TG, Mathews JD. Bacterial colonization of the nasopharynx predicts very early onset and persistence of otitis media in Australian aboriginal infants. *Pediatric Infectious Diseases Journal* 1994; **13**(11): 983–9. [Longitudinal study describing the onset of severe otitis media in a very high-risk population.]

Morris PS. A systematic review of clinical research addressing the prevalence, aetiology, diagnosis, prognosis and therapy of otitis media in Australian Aboriginal children. *Journal of Paediatrics and Child Health* 1998; **34**: 487–97. [A summary of research findings from a very high-risk population.]

Rosenfeld RM, Bluestone CD. *Evidence-based otitis media*. BC Decker Inc., Hamilton, 1999. [An excellent series of systematic reviews and meta-analyses on various management options for otitis media.]

Stool SE *et al. Otitis media with effusion in young children*, Clinical Practice Guideline, No. 12 (AHCPR Publication No. 94–0622). Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, Rockville, MD, 1994. [State-of-the-art guidelines on the management of otitis media with effusion in children aged 1–3 years.]

Williams RL, Chalmers TC, Stange KC, Chalmers FT, Bowlin SJ. Use of antibiotics in preventing recurrent acute otitis media and in treating otitis media with effusion. A meta-analytic attempt to resolve the brouhaha. *Journal of the American Medical Association* 1993; **270**(11): 1344–51. [Systematic reviews of randomized, controlled trials.]

World Health Organization. Report of a WHO/CIBA Foundation Workshop *Prevention of hearing impairment from chronic otitis media*. WHO, Geneva, 1996. [WHO document addressing the most important preventable cause of hearing impairment in disadvantaged children.]

47.1.6 Leprosy

John Hargrave

[Peripheral nerve involvement](#)
[Specific nerves](#)
[Surgical intervention](#)
[Introduction](#)
[The anaesthetic hand](#)
[Claw hand](#)
[Opponens paralysis](#)
[Footdrop](#)
[Foot ulceration](#)
[Deformity of the face](#)
[Further reading](#)

Leprosy is a disease of skin, mucous membranes of the upper respiratory tract, peripheral nerves, and certain other organs. Previously common throughout the world, it is now confined mainly to the tropics and subtropics. It is caused by the bacillus *Mycobacterium leprae*, which has never been cultured *in vitro.* It has, however, been successfully grown in the footpad of a mouse and in the nine-banded armadillo. Naturally acquired leprosy has been seen in two species of monkeys from West Africa and more recently in a cynomolgus macaque from the Philippines.

The signs and SYmptoms depend on the degree of cell-mediated immunity that patients express towards the bacillus. If there is a well-expressed cell-mediated immunity the disease manifests itself as paucibacillary or tuberculoid leprosy. In this case there are few skin lesions and only one or two peripheral nerves involved. At the other end of the spectrum, where cell-mediated immunity is virtually absent, there is widespread infiltration by the bacillus of skin, mucous membranes of the upper respiratory tract, and peripheral nerves. The patient is then said to have *multibacillary* or lepromatous leprosy. Here the humoral immunity towards the bacillus is high. Between the two poles—pauci- and multibacillary leprosy—there is a whole spectrum of signs and symptoms and the disease is then classified as borderline. Borderline leprosy is immunologically unstable and patients are especially susceptible to reaction. In leprosy reactions there is a rapid increase in the cell-mediated immunity, which is often followed by widespread and sometimes devastating peripheral nerve lesions. Reactions usually occur within the first 6 months of treatment and prompt administration of corticosteroids is essential to prevent further nerve damage or to reverse existing lesions.

For discussion on differential diagnosis, medical complications, and treatment, refer to Further Reading.

Peripheral nerve involvement

Leprosy owes its importance to the surgeon because of its involvement of peripheral nerves. The main nerves affected are the ulnar, median, radial, common peroneal, tibial, and Vth and VIIth cranial nerves. As well, cutaneous nerves are frequently involved, notably those of the upper and lower limbs and also the great auricular nerve in the neck (Fig. 1).



Fig. 1. A grossly enlarged great auricular nerve.

The major peripheral nerves are generally involved at characteristic sites, which are superficial, are sites of nerve entrapment, are generally at or near moving joints, and are sites where the intraneural patterns are more complex.

At the paucibacillary or tuberculoid end of the spectrum, bacilli enter Schwann cells and multiply within them. The nerve is then invaded by lymphocytes and histiocytes, and the whole process goes on to become a chronic granuloma with epithelioid cells and giant cells. This destroys the nerve. It can also progress to form an abscess, or even to calcification of the nerve (Fig. 2).



Fig. 2. Calcification of an ulnar nerve.

At the multibacillary or lepromatous end of the spectrum, bacilli multiply unchecked within the Schwann and perineural cells. When the cells are destroyed by the bacilli, the bacilli are released and engulfed by histiocytes, which become wandering macrophages. These then pass to other nerves and other parts of the body. The sheer mass of invading bacilli and macrophages within the nerve sheath in lepromatous leprosy results in destruction of the nerve.

Destruction of peripheral nerves results in loss of sensation, in paralysis of various muscles, and in anhidrosis of the limbs. Loss of sensation is just as important as paralysis of muscles. It leads to repetitive damage and ulceration, particularly to the cornea, the fingers and the soles of the feet, and can result in neuropathic joints.

Common paralyses are claw hand, loss of opposition of the thumb, wristdrop, claw toes, and footdrop.

Enlarged peripheral nerves are the hallmark of leprosy. Sometimes they are so grossly enlarged that they are confused with other structures. At other times there is only minor enlargement and experience is needed to make a diagnosis. To merely palpate a nerve is not to say that it is enlarged. Where nerves are involved, however, there are generally other clinical signs. But, as the disease progresses towards cure, the nerves often become smaller than normal and they feel much harder

on palpation. Nerve conduction studies can aid in diagnosis.

Specific nerves

Ulnar nerve

P>The ulnar nerve is involved just proximal to the elbow (Fig. 2). Patients lose sensation in the little finger and in the ulnar aspect of the ring finger, and various intrinsics are paralysed. These include the interossei, the two medial lumbrical muscles, part of the flexor brevis of the thumb, and the adductor pollicis. Paralysis of the intrinsics results in clawing of the ring and little fingers, and also instability of the metacarpophalangeal joints of the index and middle fingers. Paralysis of part of the flexor brevis of the thumb and of the adductor pollicis leads to instability of the metacarpophalangeal joint. The thumb attempts to compensate by hyperflexion of the tip and this is manifested in chronic tip flexion. It is seen in Froment's sign, which is diagnostic of such a lesion.

Median nerve

The median nerve is involved in the distal third of the forearm just proximal to the wrist. Its destruction results in paralysis of the two lateral lumbricals, the opponens pollicis, the abductor pollicis brevis, and the remainder of the flexor brevis. If coupled with an ulnar nerve lesion it completes the clawing of the hand and leads to loss of opposition and abduction of the thumb (Fig. 3). Loss of opposition of the thumb is a crippling lesion. Patients also lose sensation in the remainder of the ring, index, and middle fingers, and in the thumb. Rarely, the median nerve is involved proximal to the elbow, where it can be palpated if enlarged. Destruction of the median nerve above the elbow is a serious lesion leading to paralysis of all of the remaining long flexors on the front of the forearm. Coupled with ulnar or radial nerve involvement it severely impairs the function of the upper limb.



Fig. 3. Ulnar/median nerve lesion: clawing of the fingers and loss of opposition of the thumb.

Radial nerve

The main trunk of the radial nerve is involved as it emerges from the spiral groove. It is less frequently involved than the ulnar and median nerves. Radial nerve damage leads to paralysis of the extensors of the forearm and to wristdrop. By way of contrast, the terminal branch is frequently involved and is often palpably enlarged just above the wrist. Loss of sensation from radial nerve damage is of little significance and the terminal branch is often taken as a biopsy to aid in diagnosis.

Common peroneal, tibial, and sural nerves

In the leg, destruction of the common peroneal nerve occurs just above the neck of the fibula. It gives rise to footdrop and loss of sensation on the dorsum of the foot.

The tibial nerve is affected in the region of the ankle and in the lower third of the leg. Its destruction results in paralysis of the intrinsics of the foot and clawing of the toes as well as loss of sensation on the sole (Fig. 4). Loss of sensation on the sole of the foot is frequently accompanied by chronic ulceration. This can and does progress to osteomyelitis of the tarsal and metatarsal bones of the foot, which may eventually lead to loss of the foot or the leg.



Fig. 4. Tibial nerve lesion—loss of sensation on the sole of the foot. After the ulcers have been healed it is essential that the patient wears appropriate footwear.

The sural nerve is also frequently involved and results in loss of sensation over the lateral border of the foot and ankle. This is not significant except in people who sit cross-legged on the ground; they are liable to chronic ulceration over the lateral malleolus.

Impaired sensation within the foot and ankle can result in collapse of tarsal bones, to a Charcot joint, and then frequently to amputation.

Vth and VIIth cranial nerves

The most common branches of the facial nerve to be involved are the temporal and zygomatic, leading to lagophthalmos (an inability to close the eye properly). Such a lesion can be confused with Bell's palsy, but its onset is more likely to be insidious. Coupled with loss of sensation from damage to the ciliary nerves (Vth cranial nerve), lagophthalmos can result in severe corneal ulceration and even blindness.

Aside from the above there are four other sequelae of leprosy that are of surgical significance. They are loss of the eyebrows (madarosis), collapse of the nasal septum, gross enlargement of the earlobes, and gynaecomastia, all of which are seen only in multibacillary or lepromatous leprosy.

Surgical intervention

Introduction

A word of caution is essential. Leprosy surgery should not be undertaken by the novice. Without experience and without adequate physiotherapy, mediocre results or failures are the norm. Failures militate against a good leprosy control programme and jeopardize further surgery.

The anaesthetic hand

Patents lose sensation in the hand in ulnar/median nerve lesions. As a result they unwittingly injure themselves. Repetitive injury can occur in the home and in the workplace. It causes ulceration of the fingers. This leads to chronic osteomyelitis of the phalanges, resorption of bone, and loss of digits. Situations such as that seen in Fig. 5, Fig. 6, Fig. 7, Fig. 8 and Fig. 9 should never occur and the importance of education to prevent such trauma cannot be overemphasized.



Fig. 5.



Fig. 6.



Fig. 7.



Fig. 8.



Fig. 5–9. The use of rough-handled tools by a person with impaired sensation in the fingers can lead to chronic ulceration, osteomyelitis of the phalanges, resorption of digits, and ultimately to loss of fingers. Education to prevent this sort of injury is critically important. See Fig. 5, Fig. 6, Fig. 7, Fig. 8, and Fig. 9.

Claw hand

The fingers become clawed in ulnar/median nerve lesions. In the normal hand the intrinsics act in coordination with the long flexors in making a fist. The intrinsics come into play before the long flexors and stabilize the metacarpophalangeal joints, thus allowing the long flexors to flex the interphalangeal joints and make a proper fist. In the hand in which the intrinsics are paralysed, the long flexors act without their help. They flex the interphalangeal joints but allow hyper-extension of the unstable

metacarpophalangeal joints.

Claw hand is a disabling condition because it seriously weakens grasp, particularly with the ring and little fingers. If they are badly clawed it is impossible, for instance, to get one's hand around a cup. The burden of grasp then falls on the tips of the fingers alone. This can put an intolerable pressure on them and since they are anaesthetic they are highly prone to ulceration (Fig. 10). Ulceration leads to osteomyelitis, resorption of bone, and loss of digits. The repair of claw hand is therefore important both functionally and aesthetically.



Fig. 10. Clawing of the fingers results abnormal pressure on the tips of fingers that are already anaesthetic.

Repair

Claw hand repair is essentially achieved by stabilization of the metacarpophalangeal joints. There are a number of operations, the simplest of which is a superficialis transfer. It can be done through a series of keyholes. In this operation a superficialis tendon is transferred from the palm to the dorsum of the fingers. The two slips of the middle-finger superficialis are divided at the base of the finger, withdrawn into the palm, and divided longitudinally into four, one for each finger. They are then rerouted along the intrinsic pathway, superficial to the deep transverse metacarpal ligament and on to the dorsum of each finger, where they are fixed to the extensor hood over the middle of the proximal phalanx. The superficialis slips are generally routed along the ulnar side of the index and along the radial side of the remaining three fingers. Postoperatively the hand is splinted in a plaster cast with the fingers in the lumbrical position, that is with the metacarpophalangeal joints flexed to 90° and with the interphalageal joints straight.

After 3 weeks the cast is removed and physiotherapy is started. Adequate physiotherapy at this stage is essential. It is important that the patient does not try to flex the fingers at the interphalangeal joints immediately. Wrist extension with concurrent metacarpophalangeal joint flexion exercises, with the fingers in individual splints to keep them straight, are all that is required. When it is clear that the tendon transfer is adducting the fingers towards each other while in individual splints, the splints can be removed by day to allow some early interphalangeal joint flexion. The fingers are then returned to splints at night and after a couple of weeks they can be left off permanently if the clawing of the fingers does not recur.

Claw hand repair should not be attempted if there are fixed contractures at the interphalangeal joints. These contractures must be reduced preoperatively by repeated splinting and physiotherapy. Unfortunately, long-standing fixed contractures of the fingers often result in permanent damage to the dorsal expansion of the extensor apparatus over the proximal interphalangeal joints. The central slip may become detached and the two lateral bands slip down the sides of the joint beyond the fulcrum of the joint, thus perpetuating the deformity. This is difficult to correct. Theoretically the lateral bands can be restored to their original positions and the central slip repaired but the results are frequently disappointing. It underlines the importance of prevention of contractures preoperatively. In other words, physiotherapy to prevent fixed contractures in a claw hand should begin the moment paralysis has been detected. Removal of the superficialis where the profundus is paralysed will leave a patient unable to flex the finger from which the superficialis was taken. A tenodesis to a functional profundus belly can overcome the problem, but it is better to use a different motor tendon as described below.

There are other operations used to correct claw hand. They all work on the same principal as the superficialis transfer, in other words, stabilization of the metacarpophalangeal joints. One elegant transfer is the extensor–flexor many-tailed graft of Brand. In this operation, the extensor carpi radialis longus is used as the motor tendon instead of the superficialis (Fig. 11 and Fig. 12). The tendon is detached at its insertion and rerouted around the radial side to the front of the forearm. A suitable tendon is then fixed to its free end (either a plantaris from the leg or some fascia lata), the tendon transplant is divided into four slips, and the whole routed through the carpal tunnel to the centre of the palm and thence to the fingers as is done in the superficialis transfer. This is technically a little more complex, but has the advantage of leaving the middle-finger superficialis intact.



Fig. 11. Claw hand repair: the motor tendon (extensor carpi radialis longus or palmaris longus) and a plantaris tendon has been attached to it.



Fig. 12. Claw hand repair: the plantaris slips have been routed through the carpal tunnel.

Likewise, the palmaris longus can be used as a motor tendon and this has the advantage of using a relatively unimportant muscle. It should only be used to correct a very mobile claw hand since it is not very strong.

Opponens paralysis

Loss of opposition of the thumb is a serious disability. It can be corrected by transfer of a superficialis tendon (usually from the ring finger) (Fig. 13). As with the superficialis transfer described above, the tendon is divided at the base of the ring finger. It is then withdrawn about 4 cm above the wrist. The division between the two slips is extended, and both are rerouted around the flexor carpi ulnaris and medial to the pisiform. One slip is then routed across the palm to the extensor apparatus over the base of the proximal phalanx of the thumb and is inserted and fixed into it. The other is also passed across the palm (more proximal than the other slip), and is passed around the back of the thumb and inserted into the adductor insertion on the medial side of the metacarpophalangeal joint. The hand is splinted for 3 weeks postoperatively and physiotherapy is then started, initially with the metacarpophalangeal and interphalangeal joints of the thumb splinted straight. When the tendon transfer has been mobilized (which takes a few weeks), the splint can be re-moved and physiotherapy continue without it.



Fig. 13. The same hand as in Fig. 3 following claw hand repair and opponens replacement.

Footdrop

Footdrop can be corrected by transfer of the tibialis posterior. In this operation, the tendon is divided at its insertion, withdrawn on the medial side of the middle of the leg, rerouted between the tibia and fibula (or occasionally around the medial border of the tibia), and inserted into the ligaments over the cuneiforms on the dorsum of the foot. Although the operation is technically simple, it should not be attempted in the elderly patient since retraining is very difficult. It succeeds best in young people. A simple alternative to surgery is to wear a plastic ankle–foot orthosis or toe-raising splint. However, this assumes that the patient wears a shoe, and many women do not like the splint since it can be seen.

Foot ulceration

Ulceration of the sole of the foot is common in tibial nerve lesions. Chronic ulceration leads to osteomyelitis and eventually loss of bones of the foot or even to loss of the whole leg. As well, chronic, un-treated ulceration is a common precursor of secondary amyloidosis and can also lead to secondary malignant changes within the ulcer.

Established ulceration of the sole of the foot is best treated in a plaster cast fitted with a walking rocker. It allows patients to work while the ulcer heals. Sequestra should be removed before application of the cast. Since the limb is anaesthetic the cast must be applied with care to prevent further pressure sores. A well-fitting below-knee cast is better than a heavily padded one since the padding tends to compress, allowing movement within the cast. Plaster casts are usually left on for about 6 weeks. If the ulcer is healed a good shoe or sandal is essential to prevent re-ulceration. Microcellular rubber-soled sandals of appropriate consistency and thickness are ideal but there are also ‘off the shelf' shoes that suffice. Education of patients in the prevention of ulceration is as important as the provision of suitable footwear and can not be overemphasized.

Deformity of the face

Correction of facial deformity is important in some countries where leprosy is a stigma. Nasal septal defects can be treated by cartilage or bone graft. Lagophthalmos can be treated with a tarsorrhaphy (see Fig. 29, Chapter 47.5) or a temporalis fascia transfer, while madarosis or loss of eyebrows can be treated by scalp grafts. Excessive infiltration of the earlobes lends itself readily to removal of excess lobe.

Further reading

Antia NH, Divekar SC, Dastur DK. The facial nerve in leprosy. Clinical and operative aspects. *International Journal of Leprosy* 1966; **34**(2): 103–17.

Fritschi EP. *Reconstructive surgery in leprosy*. John Wright, Bristol, 1971. [A detailed account of surgical techniques.]

Jopling WH, McDougall AC. *Handbook of leprosy*, 4th edn. Heinemann Professional, Oxford, 1988. [A comprehensive handbook including treatment and differential diagnosis.]

McDowell F, Enna CD, comps. *Surgical rehabilitation in leprosy and in other peripheral nerve disorders*. Williams and Wilkins, Baltimore, 1974.

McLeod JG, Hargrave JC, Walsh JC, Booth GC, Gye RS, Barron A. Nerve conduction studies in leprosy. *International Journal or Leprosy* 1975; **43**(1): 21–31.

Mendis M. *Physiotherapy in leprosy*. John Wright, Bristol, 1965. [One of the most comprehensive guides to physiotherapy in leprosy.]

Sunderland S. Ulnar nerve lesions. In: *Nerves and nerve injuries*, 2nd edn, pp. 753–4. Churchill Livingstone: Edinburgh, 1978.

47.1.7 Pyogenic liver abscess

Pai-Ching Sheen and King-Teh Lee

Introduction
Incidence
Clinical features
Diagnosis
Laboratory data
Aetiology
Bacteriology
Complications
Treatment
Mortality and prognosis
Further reading

Introduction

Pyogenic liver abscess is an uncommon but important disease, which is usually fatal unless prompt diagnosis allows early treatment. Despite advances in diagnostic techniques and in antibiotic therapy, the mortality rate associated with pyogenic liver abscess is still high. Diagnosis may not be made until after death, because of the non-specific nature of the disease. The physician therefore needs to be alert to the possibility of pyogenic liver abscess when evaluating patients with fever of unknown origin. In countries where *Entamoeba histolytica* is endemic, a liver abscess is much more likely to be amoebic than pyogenic (see Chapter 30.3).

Incidence

It is difficult to determine the true incidence of pyogenic liver abscess. The frequency determined from autopsy reports varies between 0.29 and 1.47 per cent, while the incidence in the hospital population ranges from 0.008 to 0.016 per cent. In 1938 the average patient age was in the fourth decade; today pyogenic liver abscess predominantly affects those in the fifth and sixth decades.

Clinical features

Fever and abdominal pain are the most common symptoms of pyogenic liver abscess. About 90 per cent of patients experience a spiking fever, which may be followed by chills. Abdominal pain is usually located in the right upper quadrant, and may radiate to the right shoulder or right flank. Epigastralgia or upper abdominal pain may also occur. A ruptured abscess will cause acute diffuse abdominal pain. Non-specific symptoms such as nausea, vomiting, anorexia, and malaise occur with varying frequency.

Hepatomegaly with tenderness in the right subcostal region is the most common physical sign, although its incidence has decreased from 80 per cent to about 50 per cent. Clinically obvious jaundice occurs in one-third of patients and usually indicates biliary tract disease with cholangitis and possible multiple liver abscess; it is considered a grave sign. The symptoms and signs are similar to those of biliary tract diseases such as cholangitis, choledocholithiasis, or even acute cholecystitis. Meticulous differential diagnosis is necessary to facilitate early treatment.

Diagnosis

Ultrasonography has a reported sensitivity for detecting liver abscess in the range of 85 to 95 per cent. It may be the initial test undertaken to exclude a suspected liver abscess (Fig. 1). This technique is non-invasive, avoids radiation exposure, differentiates solid from fluid-filled masses, and is relatively inexpensive. CT also has a 95 to 100 per cent sensitivity for liver abscesses: solitary or multiple abscesses as small as 0.5 cm may be detected by this technique. CT scanning (Fig. 2) will define a liver abscess when ultrasound findings are equivocal. Liver scans have been replaced by ultrasonography: although ultrasonography cannot differentiate solid from cystic masses, it has an accuracy of 80 to 95 per cent. Angiography is recommended only when a febrile type of hepatoma needs to be distinguished from a liver abscess. Most abscesses (60 per cent) are located in the right hepatic lobe. Multiple abscesses and solitary abscesses are equally common.

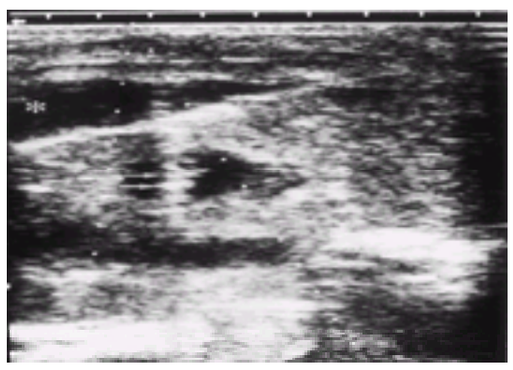


Fig. 1. Abdominal ultrasonography reveals a liver abscess with concomitant subphrenic abscess (*). A drainage catheter (arrows) was inserted for abscess drainage.

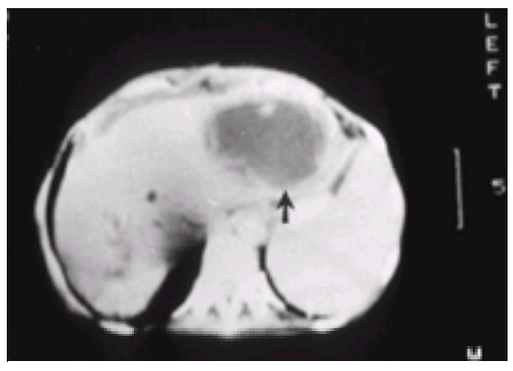


Fig. 2. CT scanning indicates a liver abscess (arrow) at left hepatic lobe.

Pyogenic liver abscess is easily diagnosed by echographic imaging, which reveals a cystic lesion of irregular contour. Confirmation is established when purulent fluid is obtained on ultrasound-guided percutaneous aspiration. The diagnosis can also be made directly at surgery. Pus and the wall of the abscess must be negative for *Entamoeba histolytica*, and levels of antibody to this organism, measured by the indirect haemagglutination test, should be less than 128.

Microabscess, a special entity of pyogenic liver abscess, is secondary to suppurative cholangitis, and its clinical features are similar to those of pyogenic liver abscess. Cholangiography shows an abscess with fig-like club-leaves communicating with the biliary tree (Fig. 3). It is usually cured by percutaneous transhepatic biliary drainage and antibiotics.



Fig. 3. Cholangiography reveals a microabscess (>) communicating with the biliary tree.

Laboratory data

Laboratory investigation does not contribute to the diagnosis. Leucocytosis is present in over 80 per cent of patients, and anaemia is usually present. Varying degrees of abnormality are seen in liver function tests, but these are non-specific. The serum level of alkaline phosphatase is elevated in about 90 per cent of patients. Hyperbilirubinaemia is found in 30 to 50 per cent of patients, and hypoalbuminaemia is present in one-third. This may be an indication of a serious underlying intrahepatic lesion or chronic infection.

Chest radiographs are abnormal in 28 to 70 per cent of patients, showing an elevation of the diaphragm, effusion, infiltration, or atelectasis. Pleural effusion, occurring in 20 to 40 per cent of patients, may indicate a subphrenic inflammatory process. When fever and pleural effusion are present, meticulous abdominal sonography or a CT scan is required to rule out the progression of liver abscess.

Aetiology

Pyogenic liver abscess may originate from biliary tract disease, portal vein, liver trauma, malignancy, direct invasion, hepatic artery, or cryptogenic origins (Table 1). Biliary tract disease is the most common cause, accounting for 33 to 55 per cent of patients in one series. Abscesses that develop from the portal vein (16 to 30 per cent) are generally related to diverticulitis, appendicitis, or pancreatic abscess. Malignancy, perhaps hepatoma itself, causes 8 to 22 per cent of pyogenic liver abscesses; these have a higher mortality rate. Sepsis may cause liver abscess by haematogenous spread through the hepatic artery: four cases of liver abscess secondary to dental abscess, transmitted by this means, have been reported. Pyogenic liver abscess follows liver trauma in 1 to 12 per cent of patients. It is usually caused by superimposed bacterial infection of traumatically devitalized liver tissue in a poorly drained lesion. The incidence of cryptogenic liver abscess has fallen because of advances in diagnostic techniques.

Aetiological factors	%
Biliary	37.5
Portal vein	20.7
Cryptogenic	17.8
Malignancy	12.3
Direct involvement	9.1
Hepatic artery	5.7
Post-trauma	3.0

Table 1 Causes of pyogenic liver abscess in recent decades in a collected series

Bacteriology

The most commonly isolated organism is *Escherichia coli*. Recently however, some authors have reported that *Klebsiella* is becoming more common.

Gram-positive aerobes such as staphylococcus and streptococcus are usually isolated from pus. Polymicrobial infection is present in about 22 per cent of patients, and such patients usually have a higher mortality rate.

The number of anaerobic organisms cultured has steadily increased due to improved microbiological techniques: 19 to 45 per cent of patients are infected by anaerobic bacteria. Successful treatment requires appropriate drug therapy. Mortality rates as low as 5 per cent can be obtained after successful isolation of anaerobic organisms.

Complications

Complications of pyogenic liver abscess are not unusual, and increase the mortality rate. Rubin *et al.* reported fatal complications in 10 of 14 patients with pyogenic liver abscess. Patients with respiratory complications have statistically significant higher mortality rates.

Complications include spontaneous rupture of the abscess. Rupture of the abscess into the abdominal cavity may cause generalized peritonitis, necessitating prompt surgical intervention; this occurs in 6 to 10 per cent of patients. An abscess rupturing into the lungs may cause pyothorax or hepatobronchial fistula: Pitt and Zuidema reported pleuropulmonary changes in 15 per cent of their patients. Subphrenic abscess or subhepatic abscess has been reported in 8 to 10 per cent of patients. Septicaemia occurs in 45 per cent of patients, and is usually fatal. Rare complications include renal failure, meningitis, and rupture into the pericardium, retroperitoneum, or small intestine.

Treatment

Although antibiotics alone have been reported to cure pyogenic liver abscess, drainage either by surgery or by the percutaneous route is usually considered necessary.

Percutaneous catheter drainage has been used for about 15 years, and is now the first line of treatment. Percutaneous drainage performed under ultrasonographic guidance also allows a definitive diagnosis to be made by the demonstration of purulent fluid in the aspirate. The abscess cavity should then be drained by an indwelling catheter. Clinical features such as fever, abdominal pain, and leucocytosis are promptly relieved as the pus is drained and pressure in the abscess cavity is reduced. Percutaneous drainage has several advantages: it is simple and safe, and can usually be performed at the bedside and completed in a few minutes; the procedure is performed under local anaesthesia, avoiding surgical or anaesthetic risks in elderly or high-risk patients; multiple catheters can be installed for multiple abscesses; and reduction of pressure within the abscess cavity by continuous drainage improves blood flow to the adjacent compressed hepatic parenchyma, and enhances the efficacy of antibiotic therapy. The possibility of echinococcal cysts should be excluded before percutaneous drainage is planned, especially when patients are in or from an endemic area.

Surgical drainage should be performed only when percutaneous abscess drainage fails. Such failure may be due to the small calibre of the drainage tube or the high viscosity of the pus. In order to ensure a smooth drainage, a tube of larger bore must replace the previously inserted tube. Surgical intervention is also required if a pyogenic liver abscess ruptures. This usually causes acute abdominal pain. Emergency laparotomy to clean the abdominal cavity and drain the abscess is mandatory. Laparotomy also allows perforation of a hollow organ to be excluded. Surgery is also necessary when the abscess has spread from an abdominal organ. Biliary surgery

should be performed in the presence of biliary stones or stricture. Diseases of the gastrointestinal tract must be treated in addition to the management of the abscess.

Transperitoneal drainage is now used frequently, since it permits assessment and correction of aetiological factors. The abscess is completely evacuated and the roof is usually partially excised. Drains are left in the cavity and others are left in the subhepatic space or adjacent to the abscess cavity. Specimens from the abscess cavity or abscess wall must be examined by a pathologist to exclude the possibility of hepatoma or amoebiasis. A febrile type of hepatoma mimics pyogenic liver abscess, and usually causes a failed percutaneous drainage.

A posterior, superiorly located, solitary abscess can be drained by a posterior extraperitoneal approach, which has the advantage of being simple and safe, especially in critically ill patients. It also prevents purulent fluid from draining into the abdominal cavity.

Either surgical or percutaneous drainage should be complemented by antibiotic therapy. Broad-spectrum antibiotics combined with aminoglycoside are given initially, and changed according to the results of bacterial cultures and sensitivity tests.

Mortality and prognosis

An untreated liver abscess is associated with a mortality rate of up to 79 per cent. However, there has been a significant decrease in mortality in recent years as a result of advances in diagnostic technique and improvements in antibiotic treatment and early drainage. In the past 15 years, the reported mortality rate has ranged from 11.5 to 40 per cent, with a mean of about 20 per cent (Table 2). Factors that have been recognized as causing a higher mortality rate include the presence of multiple abscesses, hypoalbuminaemia (less than 2.5 g/dl), impaired liver function, leucocytosis greater than 20 000 mm^{-3}, and pleural effusion (Table 3).

Study	*No. of cases*	*Mortality rate (%)*
Barbour and Juniper (1972)	33	79
Lazarchick *et al.* (1973)	75	40
Rubin *et al.* (1974)	50	63
Pitt Zuidema (1975)	80	65
Young (1976)	28	30
Sarmaid Davidson (1978)	38	29
Heymann (1979)	27	48.1
Northover (1982)	30	43
Greenstein *et al.* (1984)	38	23.7
McDonald *et al.* (1985)	55	25
Conter *et al.* (1986)	42	40
Gyorffy *et al.* (1987)	26	11.5
Barnes *et al.* (1987)	48	14.6
Farges *et al.* (1988)	46	24
Bansal and Prabhakar (1988)	24	25
Klatchko Schwartz (1989)	33	21.2
Authors (1990)	73	19.2

Table 2 Mortality rate of patients with pyogenic liver abscess in the last two decades

Clinical features
Multiple abscess
Age
Jaundice
Pleural effusion*
Positive blood culture
Laboratory data
Hypoalbuminaemia* <2.5 g/dl
Leucocytosis* >20 000/mm^3
Hyperbilirubinaemia >2.0 mg/dl
Serum aspartate aminotransferase >100 μg/l
Alkaline phosphatase >150 μg/l

* Leucocytosis, hypoalbuminaemia, and pleural effusion are independent significant risk factors for pyogenic liver abscess.

Table 3 Risk factors for pyogenic liver abscess

Further reading

Alfidi RJ, Haaga JR, Havrilla TR, Pepe RG, Cooke SA. Computed tomography of the liver. *American Journal of Surgery* 1980; **4**: 369–80.

Altemeier WA, Schowengerdt CG, Whiteley DH. Abscesses of the liver: surgical consideration. *Archives of Surgery* 1970; **101**: 258–66.

Balasegaram M. Management of hepatic abscess. In: Ravitch MM, ed., *Current problems in surgery*, Vol. 18, pp. 285–340. Year Book Medical Publishers, Chicago, 1981.

Bansal AS, Prabhakar P. Clinical aspects of pyogenic liver abscess: the University Hospital of the West Indies experience. *Journal of Tropical Medicine and Hygiene* 1989; **91**: 87–93.

Barbour GL, Juniper K Jr. A clinical comparision of amebic and pyogenic abscess of the liver in sixty-six patients. *American Journal of Medicine* 1972; **53**: 323–34.

Barnes PF, Decock KM, Reynolds TN, Ralls PW. A comparision of amebic and pyogenic abscess of the liver. *Medicine* 1987; **66**: 472–83.

Bertel CK, Von Heerden JA, Sheedy PG II. Treatment of pyogenic hepatic abscesses. *Archives of Surgery* 1986; **121**: 554–8.

Butter TJ, McCarthy CF. Pyogenic liver abscess. *Gut* 1969; **10**: 389–99.

Cohen JL, Martin FM, Rossi RL, Schoetz DJ Jr. Liver abscess: the need for complete gastrointestinal evaluation. *Archives of Surgery* 1989; **124**: 61–4.

Conter RL, Pitt HA, Tompkins RK. Differentiation of pyogenic from amebic hepatic abscesses. *Surgery, Gynecology, and Obstetrics* 1986; **162**: 114–20.

DeBakey ME, Creech O Jr. Hepatic abscess. In: Schiff L, ed., *Disease of the liver*, 2nd edn, pp. 728–64. JB Lippincott Co., Philadelphia, 1963.

Farges O, Lesse T, Bismuth H. Pyogenic liver abscess. an improvement in prognosis. *British Journal of Surgery* 1988; **75**: 862–5.

Ferrucci JT, Von Sonnenberg E. Intraabdominal abscess: radiological diagnosis and treatment. *Journal of the American Medical Association* 1980; **246**: 2728–33.

Greenstein AJ, Lowenthal D, Hammer GS, Schaffner F. Continuing changing patterns of disease in pyogenic liver abscess. *American Journal of Gastroenterology* 1984; **79**: 217–26.

Gyorffy EJ, Frey CF, Silva J Jr, McGahan J. Pyogenic liver abscess: diagnostic and treatment strategies. *Annals of Surgery* 1987; **206**: 699–705.

Herbert DA *et al.* Pyogenic liver abscess: successful non-surgical treatment. *Lancet* 1982; **ii**: 134–6.

Heymann AD. Clinical aspects of grave pyogenic abscesses of the liver. *Surgery, Gynecology, and Obstetrics* 1979; **149**: 209–13.

Keefer CS. Liver abscess: a review of eighty-five cases. *New England Journal of Medicine* 1934; **24**: 21–4.

Klatchko BA, Schwartz SI. Diagnostic and therapeutic approaches to pyogenic abscess of the liver. *Surgery, Gynecology, and Obstetrics* 1989; **169**: 332–6.

Lazarchick J, Desouza E, Silva NA, Nidrols DR, Washington TA II. Pyogenic liver abscess. *Mayo Clinic Proceedings* 1973; **48**: 349–55.

Lee KT, Kwan AL, Ker CG, Sheen PC. Echo-guided percutaneous transhepatic drainage of liver abscess. *Journal of the Surgical Association of the Republic of China* 1987; **20**: 291–6.

Maher JA Jr, Reynolds TB, Yellin AE. Successful medical treatment of pyogenic liver abscess. *Gastroenterology* 1979; **77**: 618–22.

McDonald AP, Howard RJ. Pyogenic liver abscess. *World Journal of Surgery* 1980; **4**: 369–80.

McDonald MI, Corey GR, Gallis HA, Durack DT. Single and multiple pyogenic liver abscesses. *Medicine* 1984; **63**: 291–302.

Northover JMA, Jones BJM, Dawson JL, Williams R. Difficulties in the diagnosis and management of pyogenic liver abscess. *British Journal of Surgery* 1982; **69**: 48–51.

Ochsner A, DeBakey M, Murray S. Pyogenic abscess of the liver II: an analysis of forty-seven cases with review of the literature. *American Journal of Surgery* 1938; **40**: 292–329.

Pitt HA, Zuidema GD. Factors influencing mortality in the treatment of pyogenic hepatic abscess. *Surgery, Gynecology, and Obstetrics* 1975; **140**: 228–34.

Rubin RH, Swartz MN, Malt R. Hepatic abscess: changes in clinical, bacteriologic and therapeutic aspects. *American Journal of Medicine* 1974; **57**: 601–10.

Rubinson HA, Isikoff MB, Hill MC. Diagnostic imaging of hepatic abscesses: A retrospective analysis. *American Journal of Roentgenology* 1980; **135**: 135–40.

Sabbaj J, Sutter VL, Finegold SM. Anaerobic pyogenic liver abscess. *Annals of Internal Medicine* 1972; **77**: 629–38.

Satiani B, Daridson ED. Hepatic abscesses: improvement in mortality with early diagnosis and treatment. *American Journal of Surgery* 1978; **135**: 647–50.

Sherman JO, Robbins SL. Changing trends in the casuistics of hepatic abscess. *American Journal of Medicine* 1960; **28**: 943–80.

Young AE. The clinical presentation of pyogenic liver abscess. *British Journal of Surgery* 1970; **63**: 216–19.

47.2.1 Amoebiasis

Prasit Watanapa, Arun Pausawasdi and Robin C. N. Williamson

Prevalence
Intestinal amoebiasis
Pathology
Presentation
Diagnosis
Treatment
Surgical complications of amoebic colitis
Liver abscess
Incidence
Clinical manifestations
Complications
Diagnosis
Treatment
Prognosis
Prevention and vaccine development
Further reading

Prevalence

Amoebiasis is generally defined as infestation with *Entamoeba histolytica*, with or without overt clinical symptoms. Dysentery has been recognized since the time of Moses and was discussed by Hippocrates, but the parasite was first identified in faecal specimens in 1869 by Lewis, and the clinical presentation of amoebic dysentery was first described in 1875 by Losch in St Petersburg, Russia. Although the disease is curable, it remains the third leading cause of death from parasitic infestation in the world, surpassed only by malaria and schistosomiasis. Several factors have led to a failure to eradicate the disease.

First, the vast majority of human infestations with *E. histolytica* (at least 80 per cent) are completely asymptomatic. Colonization by the parasite in the large intestine can be detected in 10 to 20 per cent of the world's population, with an average of 12 per cent as reported by the World Health Organization. There are innumerable carriers of the disease in endemic countries including Mexico, India, East and South Africa, and portions of Central and South America.

Second, human beings are not the only species affected by *E. histolytica*, though they are the principal host and reservoir. Certain other animals (especially macaque monkeys) are naturally infested with the parasite. Identical strains of amoeba affect macaques and humans, and cross-infection has been reported.

Third, the diagnosis of amoebiasis by microscopic demonstration of the organism in stools is insensitive and unable to distinguish the invasive parasite *E. histolytica* from the commensal parasite *E. dispar*. Diagnostic accuracy increases when microscopic examination of fresh stool is combined with (1) assays to identify *E. histolytica*-specific antigen or (2) an enzyme-linked immunosorbent assay based on monoclonal antibodies to the galactose-specific adhesin in the stool or (3) (more recently) detection of *E. histolytica*-specific DNA by polymerase chain-reaction amplification. All these tests have several inherent problems and are expensive. Epidemiological studies in an endemic area are therefore difficult, if not possible, to perform.

Many epidemiological studies of *E. histolytica* infestation were performed before the distinction was made between *E. histolytica* and *E. dispar*. Because the two species are morphologically indistinguishable, studies based on stool examinations should reflect infestation with both types. However, the prevalence of *E. histolytica* infestation can be estimated by evaluating the seropositivity in a population, since seroconversion does not occur in people carrying *E. dispar*. Applying these facts to the reported figure of 500 million people worldwide who excrete amoeba in the stool, *E. dispar* infestation is probably 10-fold more common than *E. histolytica* infestation. Nevertheless, only about 10 per cent of the population carrying *E. histolytica* experience symptomatic disease. Among these people, nearly 40 million develop disabling colitis and/or extraintestinal abscesses, resulting in 40 000 deaths annually. Children, particularly neonates, pregnant women, and women in the postpartum period have an increased risk of severe disease and death. In Thailand, at least, amoebic colitis has become less important as a national health problem over the past 10 years. Morbidity and mortality rates have declined overall, and especially in patients with surgical complications of amoebiasis.

Amoebiasis presents in two distinct forms, intestinal and extraintestinal. The intestinal form affects the caecum, ascending colon, sigmoid colon, transverse colon, and rectum; the terminal ileum is generally spared. Extraintestinal amoebiasis is found in the liver, lung, skin, and brain. The most common complication is liver abscess.

Intestinal amoebiasis is almost universal in the tropics, particularly between parallels 40°N and 30°S. The prevalence of infestation and the incidence of disease show geographical variations, probably because transmission and invasiveness vary with ecological and social conditions. The lack of published reports of amoebiasis from some areas of the world probably reflects a lack of medical enthusiasm rather than absence of the parasite. In Thailand, which has a population of 60 million, epidemiological data from the Ministry of Health stated that there were 3000 to 5000 new cases of intestinal amoebiasis annually between 1985 and 1989, and 100 to 120 deaths each year. More information on stool and serological surveys is available from Mexico than from other developing countries in which this parasite is known to endemic. In 1981 between 3 and 5 million of the 69 million Mexicans developed the antibody each year. There were 5 to 6 million clinical cases of amoebiasis and 10 000 to 30 000 deaths. More recent serological studies showed that 8.4 per cent of the population in Mexico City were infested with *E. histolytica* during the last 5 to 10 years.

Intestinal amoebiasis

Pathology

Human beings contract amoebiasis by consuming food or drink contaminated with the cyst of *E. histolytica*. After ingestion the quadrinucleate cysts reach the intestinal tract, where they develop into a metacystic stage and undergo an additional nuclear division; thus eight new uninucleate trophozoites emerge to complete the lifecycle. The trophozoites produce many proteolytic enzymes and cause tissue destruction by forming minute ulcers in the mucosa of the bowel in the ileocaecal and colorectal areas. Subsequent lateral extension of these ulcers through the submucosa with relative sparing of the overlying epithelium gives rise to the classic, flask-shaped ulcer of amoebiasis (Fig. 1). The parasite damages target cells by a contact-dependent mechanism. One of the important molecules involved in mediating amoebic adhesion to intestinal epithelium is a surface galactose-inhibitable adhesin containing 170-kDa and 35-kDa subunits. Genes encoding both subunits have been cloned and sequenced; the 170-kDa subunit has the galactose-binding activity and is encoded by a multigene family. Monoclonal antibodies that bind to epitopes present on the 170-kDa subunit from *E. histolytica* strains but not from *E. dispar* form the basis of faecal antigen-detection assays that distinguish infestation by these two types of amoeba. The latent period before the development of invasive amoebiasis varies: in an outbreak of the disease the usual incubation period is 1 to 4 weeks, but it can be as long as 1 year.

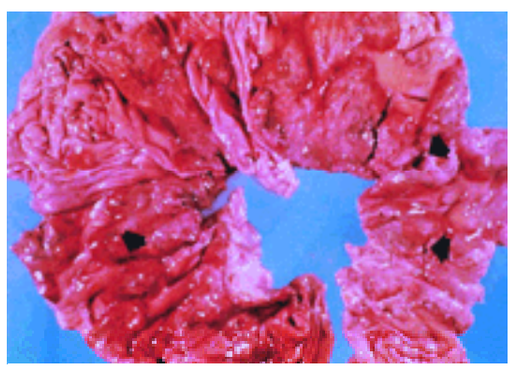


Fig. 1. Resected specimen of perforated amoebic colon shows multiple, ill-defined ulcers (arrows) with extensive oedematous and necrotic mucosa in the area of perforation; normal-looking mucosa is also seen between ulcers (by courtesy of P. Suwanagool, MD).

Presentation

The onset of symptoms may be insidious, with lower abdominal pain. It can also be sudden, with high fever, bloody diarrhoea, foul-smelling stools, dysentery (most patients have tenemus), and the pain of typhloappendicitis. The fulminant form is most common in youths or those who are malnourished, pregnant, with underlying systemic disease, or receiving corticosteroid therapy. The dysentery associated with amoebiasis is frequently confused with that of shigella infection. Amoebiasis may also be confused with infection by *Campylobacter jejuni, Yersinia enterocolitica, Escherichia coli*, and *Vibrio parahaemolyticus*. The mild diarrhoeal syndrome of amoebiasis may be indistinguishable from salmonellosis, giardiasis, toxigenic *E. coli* diarrhoea, other infective diarrhoeas, and irritable bowel syndrome.

Diagnosis

The diagnosis of intestinal amoebiasis still depends primarily on microscopic identification of cysts or trophozoites in the stool or in scrapings or biopsy specimens obtained by sigmoidoscopy or colonoscopy. It is important to remember that all methods of stool examination are tedious, time-consuming, expensive, and require a high degree of skill of the examiner. Since *E. dispar* may be misinterpreted as *E. histolytica* on fresh stool examination, amoebic serology is a useful adjunct in evaluating the patient, particularly in areas where *E. histolytica* is not endemic. However, one has to be cautious about the interpretation of serological tests since antiamoebic antibodies may persist for months or even years after the eradication of the parasite, and the antibodies may be negative early in the course of an acute infestation. In general, the test is positive in 75 to 90 per cent of individuals with symptomatic intestinal infestation. Faecal antigen-detection assays have been developed for commercial use but are not yet widely available. Although DNA-based diagnostic methods for detection of *E. histolytica* can now be performed (see above), they remain research tools at the present time.

Clinical manifestations

Patients with fulminating amoebic colitis have all the features of severe ulcerative proctocolitis. They pass several blood-stained, mucous stools per day. They have severe tenemus, high fever, abdominal pain, a distended tender abdomen, toxaemia, rapid pulse, low blood pressure, and even shock. Colonic perforation is accompanied by signs of peritonitis.

Mucus and blood are found on digital rectal examination. Palpation reveals oedema and ulceration of the rectal mucosa; this is confirmed at proctoscopy. Sigmoidoscopy should be performed with extreme care since insufflation of the colon (or administration of an enema) increases the incidence of perforation. If the disease is confined to the right colon, the rectosigmoid area may appear normal; a limited examination will therefore not exclude the diagnosis of amoebic colitis. Endoscopy shows ulcers all over the rectal mucosa, varying from pinhead size up to 2 cm in diameter. Most ulcers are covered with a yellow exudate.

Stool specimens

Cotton buds should not be used to collect ulcer exudate or mucous stool because amoebic trophozoites may adhere to the cotton and produce a false-negative result. A small spoon or even biopsy forceps yields better results. The specimen should be examined immediately. Because amoebae may be excreted in an acyclic manner or may be unevenly distributed in the faecal specimen, at least three stool specimens should be examined so as to detect 85 to 90 per cent of the infestations. Punch biopsy of the ulcer edge for histological examination can also be performed (Fig. 2).

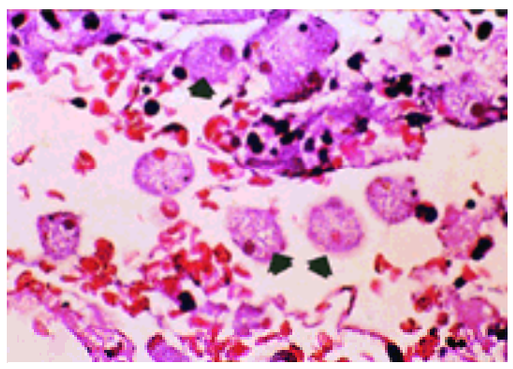


Fig. 2. Multiple, round to oval, amoebic trophozoites containing a small, single nucleus and phagocytized red blood cell in the cytoplasm are demonstrated (arrows). Haematoxylin and eosin-stained section, ×400 (by courtesy of P. Suwanagool, MD).

Radiology

Plain radiographs of the abdomen are important when acute fulminating amoebic colitis is suspected. At an early stage there is generalized gaseous distension of the bowel. Later, when colonic necrosis is imminent, there is marked dilatation of the large bowel from caecum to sigmoid colon (toxic megacolon).

Single- or double-contrast barium enema is a valuable tool for diagnosing invasive amoebiasis. At an initial stage, many superficial small ulcers are usually found, resulting in a fine, granular appearance of the mucosa and discrete spiculation of its margin. When the disease progresses, deep ulcers may penetrate the submucosa giving origin to ‘collar-button' ulcers.

Treatment

The goals of treatment are to treat the invasive disease and to eradicate intestinal carriage of the parasite. Metronidazole is the most popular drug for the treatment of invasive amoebiasis. It is effective, simple to use, and rarely causes side-effects. Despite widespread use of this drug, *E. histolytica* resistance to metronidazole has not been a problem to date. Emetine or dehydroemetine offers a rapid amoebicidal effect and remains an effective drug for patients with severe infestations or fulminant amoebic colitis. After treatment of invasive amoebiasis is complete, a luminal agent is generally begun to eradicate intestinal carriage of the organism. There are three drugs with proven efficacy as luminal amoebicides, iodoquinol, paramomycin, and diloxanide furoate.

Standard regimens for amoebicidal drugs

Metronidazole (400 mg orally) is administered three times a day after meals for 5 days for acute amoebic colitis and for 10 to 12 days for chronic cases. More than 90 per cent of patients respond to this regimen, and less than 10 per cent require a second course of treatment 2 weeks after the first. Emetine or dehydroemetine is not usually required for the treatment of uncomplicated intestinal amoebiasis.

Patients with hepatic amoebiasis (amoebic liver abscess) require 600 mg metronidazole orally four times a day for 1 to 2 days. Ninety-seven per cent of patients respond to this treatment. Parenteral administration of the drug should be reserved for those patients unable to tolerate anything by mouth; the recommended daily dose is 500 mg intravenously every 6 h. The 3 per cent of patients who do not respond to metronidazole, or some patients with severe infections or complicated disease, require emetine or dehydroemetine (1.0–1.5 mg/kg a day for 3 to 5 days in the acute form or for 9 days in the chronic form). Dehydroemetine is now preferred to emetine because of its lesser toxicity. After treatment of invasive disease is complete, an oral luminal amoebicide is given, either iodoquinol (650 mg three times a day after meals for 3 weeks) or paromomycin (30 mg/kg a day for 7 days in three divided doses). An excellent response is obtained in almost all patients. Though diloxanide furoate is not as popular as the two other luminal amoebicides, a recent report on 34 patients with symptomatic intestinal amoebiasis receiving a combined formulation of diloxanide furoate and metronidazole showed SYmptomatic relief in 91 per cent and parasitic clearance in 100 per cent. The combination is given in the form of a tablet containing 500 mg diloxanide furoate and 400 mg metronidazole, which is taken orally three times daily for 5 days.

Operative treatment

Operation is indicated for toxic megacolon or severe bleeding. Before 1970, procedures included total colectomy, segmental colectomy, and colostomy with simple

closure of the perforation. Overall mortality rates reached 70 to 100 per cent, the same as for non-operative treatment.

The optimal management for acute, fulminating amoebic colitis remains controversial. Most authorities agree that operation is indicated only for patients who fail to respond after 48 h of intensive management. Even some patients with peritonitis complicating amoebic colitis but without evidence of obvious colonic perforation can be treated conservatively with antiamoebic and antibacterial therapy, without surgical intervention, because peritoneal irritation is usually due to slow leakage of the colon.

It is best to avoid aggressive surgery and use minimal procedures to deal with the problem. If there is no obvious necrosis of the bowel, simple suture, faecal diversion, and drainage should suffice in severely ill patients. If necrosis is present, segmental resection of the affected segment of bowel is essential, followed by a double-barrelled colostomy.

Surgical complications of amoebic colitis

Incidence

Complications requiring surgical treatment arise in 1 to 4 per cent of patients with intestinal amoebiasis. The most common complication is colonic perforation and peritonitis. A few individuals develop an annular inflammatory mass or amoeboma in the colon, which is sometimes tender and may be indistinguishable radiographically from colonic carcinoma. Another rare but fatal complication is fulminating amoebic colitis with toxic megacolon. A few patients develop a chronic irritable bowel syndrome. Very rarely, severe proctocolitis extends to the perineal skin.

Among 1477 patients undergoing operation for peritonitis at Siriraj Hospital, Bangkok, between 1981 and 1988, there were seven cases (0.47 per cent) of colonic perforation with generalized peritonitis due to amoebiasis and five cases (0.34 per cent) of fulminating amoebic colitis and toxic megacolon. The incidence of these complications has been markedly reduced by the use of metronidazole.

Liver abscess

The liver is the organ most commonly affected by the spread of *E. histolytica* from the primary site of infestation (the large bowel) into the portal vein. The mean incidence of hepatic involvement associated with intestinal amoebiasis is approximately 10 per cent. Liver abscess can develop within days of an attack of amoebic dysentery or may follow after months or even years. Up to 50 per cent of patients have no previous SYmptoms suggestive of intestinal amoebiasis. However, among 45 patients with liver abscess undergoing complete colonoscopy, 10 had a previous history of diarrhoea or had diarrhoea at presentation. Colonic ulcerations were found in 26 of these 45 patients; 23 patients had small (less than 2.5 cm), single or multiple (up to three) ulcers with well-defined margins and minimal or no inflammation of the mucosa, while three had larger and more numerous ulcers with inflammation of the surrounding mucosa (Fig. 3).

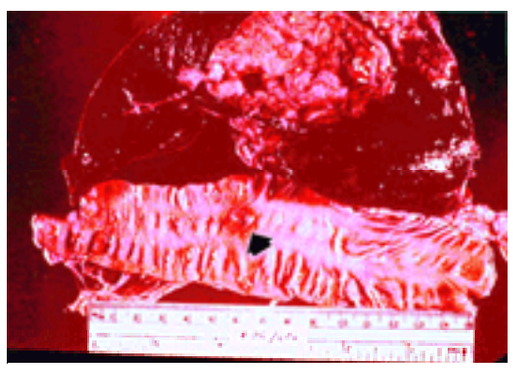


Fig. 3. A large, well-circumscribed amoebic abscess with greyish-brown content of coagulative necrosis is shown in the central portion of the liver; a small, flask-shaped ulcer (arrow) is also found in the ascending colon in the same patient (by courtesy of P. Suwanagool, MD).

Incidence

Among 138 patients with amoebic liver abscess, the age distribution was 8 to 42 years (mean 45 years). Men were affected more often than women, with a ratio of 4.5:1. Amoebic liver abscess is rare in children.

Clinical manifestations

The onset is usually gradual, with abdominal pain (79 per cent), fever (53 per cent), malaise, weight loss, and occasionally jaundice. The pain is often a vague discomfort in the right upper quadrant, and there is point tenderness between the ribs on palpation. Occasionally the symptoms are acute, with fever, pain, and chills. Some 37 per cent of patients complain of a painful and tender abdominal mass. Pain may be referred to the right shoulder and aggravated by deep inspiration. Hiccup reflects diaphragmatic irritation by a so-called dome abscess. The presentation may be with symptoms and signs of a complication: cardiac tamponade, shock, dyspnoea, or cough result from a complicated liver abscess rupturing into the pericardial sac, pleural cavity, lung, or bronchus. Ascites is another potential feature in chronic amoebic peritonitis following rupture of an abscess into the peritoneal cavity. Patients may be toxic, malnourished, and anaemic.

Complications

Rupture of an amoebic liver abscess into adjacent viscera is uncommon, but subdiaphragmatic abscess, empyema thoracis with or without bronchopleural fistula, lung abscess, amoebic pericardium, and hepatogastric, hepatoduodenal, or hepatocolic fistulas can develop. Occasionally the abscess ruptures into the subcutaneous tissues or the rectus sheath, and it can even burst through the skin. Other rare complications include haemobilia, amoebic brain abscess, and rupture into the portal vein or inferior vena cava. Common laboratory findings are anaemia, leucocytosis, and elevation of the erythrocyte sedimentation rate and serum alkaline phosphatase (Table 1).

Complication	*No.*	*%*
Reactionary		
Pleural effusion (right)	12	6.7
Pleural effusion (left)	1	0.6
Rupture into:		
Peritoneal cavity	10	5.5
Pleural cavity (right)	8	4.4
Right pleural and abdominal cavities	1	0.6
Pericardial cavity	1	0.6
Fistula		
Bronchopleural (right)	1	0.6
TOTAL	34	19.0

Table 1 Complications of amoebic liver abscess (*n* = 180)

Diagnosis

The clinical manifestations of liver abscess are seldom pathognomonic. Serological tests are positive in 90 to 95 per cent of cases.

Over the past 10 years the second author (A.P.) has undertaken ultrasonographic examination of the abdomen in over 10 000 patients. In about 70 per cent of cases the hepatobiliary system was examined. Liver abscess ranked as the third most common disease of the liver, exceeded only by diffuse parenchymal diseases and carcinoma. There were 364 cases of amoebic liver abscess diagnosed by ultrasonography: 53 of these patients had a firm clinical diagnosis made before ultrasound examination, 258 cases were suspected, and in 53 cases the clinician did not suspect liver abscess.

Ninety-eight per cent of amoebic liver abscesses are hypoechoic on the ultrasonogram, in contrast to the hyperechoic appearance of hepatocellular carcinoma. In the other 2 per cent of cases the appearances are mixed, and clinical evaluation, serological tests, and ultrasound-guided tissue biopsy for histology are helpful. The ultrasonographic pattern is not particularly helpful in differentiating amoebic from pyogenic liver abscess. In approximately 90 per cent of cases of amoebic liver abscess, the abscess cavity is solitary, contains homogeneous contents, and is situated in the posterosuperior part of the right lobe of the liver.

The usual appearance of amoebic liver abscess on abdominal computed tomographic scanning is a hypodense lesion with attenuation values of 15 to 30 HU, well-defined margins, and alternating hypervascular and hypovascular halos following contrast enhancement. Though magnetic resonance imaging of the liver is now considered the most sensitive method for the detection and characterization of focal lesions, it has only a minor role in amoebic liver abscess. Not only is it expensive, but the findings are not pathognomonic and can be seen in bacterial liver abscess, haematoma, and necrotic neoplasms.

In the authors' experience, ultrasonography can diagnose liver abscess with a sensitivity of 98 per cent, a specificity of 99 per cent, and an accuracy of 99 per cent: for both diagnosing and locating the lesion it is as good as a computed tomogram. The important principle is that general surgeons should learn to use ultrasound just as effectively as physicians can use their stethoscope.

Treatment

Operation is now seldom required for treatment of an amoebic liver abscess: even some ruptured abscesses can be treated conservatively. Patients with ruptured abscess with generalized peritonitis, or with liver abscess associated with a serious intestinal problem such as toxic megacolon, colonic perforation or fulminating colitis, still require operation. Although percutaneous drainage of liver abscess under ultrasonographic guidance is relatively safe, it is usually not necessary since most patients settle with appropriate amoebicidal therapy. However, possible indications for percutaneous aspiration of an amoebic liver abscess as an adjunct to medical therapy include (1) large abscesses (larger than 5 cm in diameter) in which rupture is believed to be imminent, (2) abscesses in the left lobe of the liver (because of the high risk of rupture into the pericardial cavity), and (3) abscesses in patients who do not respond to conservative treatment (i.e., fever and pain persist after 3 to 5 days of amoebicidal therapy).

Prognosis

Most of the deaths in amoebic liver abscess occur when the diagnosis is missed or overlooked. There was no death among the 364 cases of amoebic liver abscess treated at our hospital after ultrasonographic examination of the abdomen.

Prevention and vaccine development

Amoebic infestation can best be prevented by avoiding faecally contaminated food and water. Rapid progress in genetic engineering means that production of a vaccine for amoebiasis is now feasible. Recombinant versions of three amoebic antigens, the 170-kDa galactose-inhibitable adhesin, a serine-rich *E. histolytica* protein (**SREHP**)(a 29-kDa antigen), and a combination of the SREHP and 170-kDa molecules have shown protective efficacy in experimental models of amoebic liver abscess. These developments are still in their infancy and their role in man requires further elucidation.

Further reading

Adams EB, MacLeod IN. Invasive amebiasis: amebic liver abscess and its complications. *Medicine* 1977; **56**: 325–34. [A review of more than 2000 patients with hepatic amoebiasis from a single medical unit in South Africa.]

Balasegaram M. Management of hepatic abscess. *Current Problems in Surgery* 1981; **18**: 281–340. [A classic paper showing the experience of the author in the management of liver abscess.]

Caballero-Salcedo A *et al.* Seroepidemiology of amebiasis in Mexico. *American Journal of Tropical Medicine and Hygiene* 1994; **4**: 412–19. [Study of epidemiology of amoebiasis in Mexico by looking at serum antibodies to *Entamoeba histolytica*.]

Dobell C. Researches on the intestinal protozoa of monkeys and man. IV. An experimental study of the *histolytica* like species of *Entamoeba* living naturally in Macaques. *Parasitology* 1931; **23**: 1–72. [A study showing that infestation by *Entamoeba* occurs not only in man but also in animals such as macaque monkeys.]

Haque R *et al.* Diagnosis of pathogenic *Entamoeba histolytica* infection using a stool ELISA based on monoclonal antibodies to the galactose specific adhesin. *Journal of Infectious Diseases* 1993; **167**: 247–9. [A preliminary study on a promising method for detecting *Entamoeba histolytica* in the stool.]

Haque R, Ali IKM, Akther S, Petri WA Jr. Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of *Entamoeba histolytica* infection. *Journal of Clinical Microbiology* 1998; **36**: 449–52. [All three methods showed excellent correlation, but only the *Entamoeba histolytica* antigen detection test was both rapid and technically simple.]

Healy GR. Immunologic tools in the diagnosis of amoebiasis: epidemiology in the United States. *Reviews of Infectious Diseases* 1986; **8**: 261–72. [Prevalence of amoebiasis was studied using several immunodiagnostic tests in selected groups including children, institutionalized persons, members of extended families, immigrants, and male homosexuals.]

Kanai SR, Knight R. Relapsing amoebic colitis of 12 years standing exacerbated by corticosteroids. *British Medical Journal*, 1969; **2**: 613–14. [A case report of corticosteroid-induced exacerbation of occult amoebic colitis.]

Kimura K, Stoopen M, Reeder MM, Moncada R. Amebiasis: modern diagnosis imaging with pathological and clinical correlation. *Seminars in Roentgenology* 1997; **32**: 250–75. [An extensive review of imaging studies on intestinal and extraintestinal amoebiasis.]

Lee KC *et al.* Analysis of 69 patients with amebic liver abscess. *Journal of Gastroenterology* 1996; **31**: 40–5. [Epidemiology, clinical presentation, pathology, and outcome of treatment were reviewed in 69 patients with amoebic liver abscess at one infectious centre in Japan.]

Lewis EA, Antia AU. Amoebic colitis: review of 295 cases. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1969; **63**: 633–8. [Amoebic colitis is categorized into mild and severe forms. The severe form was more common in children, in pregnancy and puerperium, and in malnourished states.]

Li E, Stanley SL. Protozoa: amebiasis. *Gastrointestinal Clinics of North America* 1996; **25**: 471–92. [A review of amoebiasis, including the lifecycle of the parasite, pathogenesis of colitis, and clinical points of view.]

Mann BJ *et al.* Sequence of a cysteine-rich galactose-specific lectin of *Entamoeba histolytica. Proceedings of the National Academy of Sciences (USA)* 1991; **88**: 3248–52. [A study to identify the sequence of the two subunits of cell surface galactose-specific lectin of the parasite.]

Meerovitch E, Healy GR, Ambroise-Thomas P. Amoebiasis survey in Calcutta (India), Bangkok (Thailand), Medellin (Colombia), and San Jose (Costa Rica). *Canadian Journal of Public Health* 1978; **69**: 286–8. [An epidemiological study of amoebiasis in different parts of the world.]

Petri WA Jr *et al.* Pathogenic and nonpathogenic strains of *Entamoeba histolytica* can be differentiated by monoclonal antibodies to the galactose-specific adherence lectin. *Infection and Immunity* 1990; **58**: 1802–6. [A radioimmunoassay based on monoclonal antibodies to epitopes 1 and 3 of the 170-kDa heavy subunit of lectin is a simple and rapid method for distinguishing pathogenic from non-pathogenic amoeba in culture.]

Powell SJ, Wilmot AJ, Elsdon-Dew R. Further trials of metronidazole in amoebic dysentery and amoebic liver abscess. *Annals of Tropical Medicine and Parasitology* 1967; **61**: 511–20. [A study of optimum dosage and duration of metronidazole treatment in amoebic dysentery and liver abscess.]

Qureshi H, Ali A, Baqai R, Ahmed W. Efficacy of a combined diloxanide furoate-metronidazole preparation in the treatment of amoebiasis and giardiasis. *Journal of International Medical Research* 1997; **25**: 167–70. [A 5-day course of treatment showed promising results.]

Rails PW *et al.* Sonographic findings in hepatic amoebic abscess. *Radiology* 1982; **145**: 123–6. [An analysis of sonographic findings in 143 amoebic liver abscesses in 106 patients.]

Ravdin JI. Amebiasis. *Clinical Infectious Diseases* 1995; **20**: 1453–66. [A review of amebiasis.]

Sachdev GK, Dhol P. Colonic involvement in patients with amebic liver abscess: endoscopic findings. *Gastrointestinal Endoscopy* 1997; **46**: 37–9. [Patterns of colonic involvement were studied in 50

patients and compared to 15 patients with acute amoebic colitis without liver abscess.]

Soong CG *et al.* A recombinant-cysteine rich section of the *Entamoeba histolytica* galactose-inhibitable lectin is efficacious as a subunit vaccine in the gerbil model of amebic liver abscess. *Journal of Infectious Diseases* 1995; **171**: 645–51. [An experimental study on the possibility for developing a vaccine against the parasite.]

Soong CG *et al.* Protection of gerbils from amebic liver abscess by immunization with recombinant *Entamoeba histolytica* 29-kilodalton antigen. *Infection and Immunity* 1995; **63**: 472–7. [The recombinant form of the 29-kDa antigen from the parasite was effective as a subunit vaccine in this animal model of amoebic liver abscess.]

Tannich E, Eert F, Horstmann RD. Primary structure of the 170-kDa surface lectin of pathogenic *Entamoeba histolytica. Proceedings of the National Academy of Sciences (USA)* 1991; **88**: 1849–53. [A study designed to identify the amino-acid sequence of the 170-kDa surface lectin.]

Viranuvatti V, Harinasuta T, Plengvanit U, Choungchareon P, Viranuvatti V. Liver function tests in hepatic amoebiasis based on 274 clinical cases. *American Journal of Gastroenterology* 1963; **39**: 345–61. [Elevation of serum alkaline phosphatase, slightly elevated SGOT and SGPT, a positive iodine test, and BSP retention are good laboratory criteria for the diagnosis of amoebic liver abscess.]

Walsh JA. Problems in recognition and diagnosis of amebiasis: estimation of the global magnitude of morbidity and mortality. *Reviews of Infectious Diseases* 1986; **8**: 228–36. [An article emphasizing the problems of imperfect tests for the diagnosis of amoebiasis.]

WHO Scientific Working Group. Parasitic-related diarrhoeas. *Bulletin of the World Health Organization* 1980; **58**: 819–30. [A review of epidemiology, pathogenesis, clinical manifestations, immunology, diagnosis, and treatment of parasite-related diarrhoeas.]

Zhang T, Stanley SL Jr. Protection of gerbils from amebic liver abscess by immunization with a recombinant protein derived from the 170 kDa surface adhesin of *Entamoeba histolytica. Infection and Immunity* 1994; **62**: 2605–8. [Confirmed the effectiveness of immunization with an *Entamoeba histolytica* specific antigen in the prevention of amoebic liver abscess.]

47.2.2 Schistosomiasis (bilharziasis)

A. S. Daar and Euan M. Scrimgeour

Introduction
Schistosomiasis and the surgeon
The parasites and their life-cycles
Epidemiology
Geographical distribution
Prevalence and intensity of infection
Pathogenesis and pathology
Immune response to infection
Response to egg deposition
Morbidity in schistosomiasis
Genetic aspects
Clinical aspects
History of exposure
Symptoms, signs, and management
Diagnosis of schistosomiasis
Haematological and biochemical tests
Identification of ova in biopsies and stool
Immunodiagnostic tests
Imaging techniques
Chemotherapy
Vaccines
Further reading

Introduction

Schistosomiasis (also called bilharziasis, after Theodor Bilharz, who discovered the parasite in Cairo in 1852) is one of the most widespread of the parasitic diseases that afflict man: between 200 and 300 million people are infected and a further 600 million are at risk. The number of countries endemic for schistosomiasis is increasing, and it continues to be a major public-health problem in most of Africa, in parts of Asia, the Caribbean, and South America. The prevalence rates range from 10 to 80 per cent in endemic areas. Worldwide, up to 500 000 deaths annually are attributable to schistosomiasis. The majority of infections are caused by *Schistosoma haematobium* (estimated 78 million people affected), *S. mansoni* (57 million), and *S. japonicum* (70 million). *Schistosoma intercalatum, S. mekongi* (thousands affected), and *S. malayensis* (several hundred may be affected) have a more limited geographical distribution (Fig. 1).

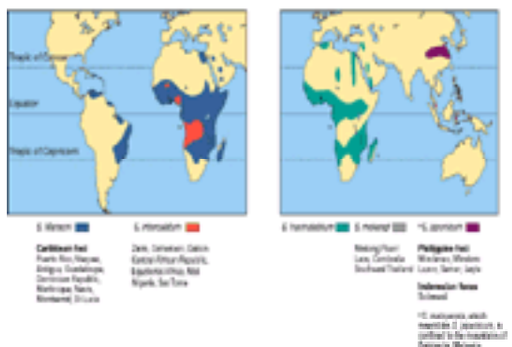


Fig. 1. Geographical distribution of principal pathogenic schistosomes.

Worms of the genus *Schistosoma* comprise several blood parasites of man and other animals; they belong to the family Schistosomatidae, the class Trematoda, and phylum Platyhelminthes (flatworms). The intermediate hosts are various species of freshwater snails, which shed the infective larval form (cercariae) into the surrounding water. Man is exposed on coming into contact with infected water: cercariae penetrate the skin and those that survive ultimately mature into adult worms, which live in pairs in intestinal or vesical veins. There the female worms shed a large number of ova (eggs), some of which are excreted in the host's stool or urine, and when these enter fresh water, the life-cycle can be completed (Fig. 2). The remaining ova are lodged in the host's tissues, where they stimulate a granulomatous response. This response and the resultant fibrosis cause the chronic manifestations of schistosomiasis. *Schistosoma haematobium* affects mainly the urinary tract and is dealt with in detail in Chapter 47.2.3.

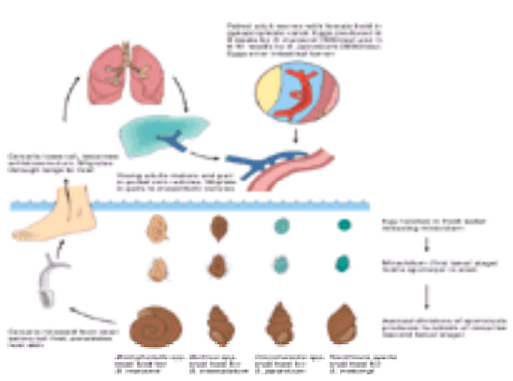


Fig. 2. Lifecycle of schistosomes.

Schistosomiasis and the surgeon

As international travel becomes more widespread, and transmigration of populations from endemic to non-endemic areas increases steadily, the surgeon may encounter schistosomiasis anywhere in the world. Its manifestations are protean, and in its acute and chronic stages it may mimic a number of other diseases (Table 1). It is probably the most common cause of portal hypertension in the world and may present for the first time with massive variceal haemorrhage. The urologist deals with its fibrotic and malignant sequelae, but schistosomiasis also leads to endstage renal failure, to be investigated and treated by the transplant surgeon. The neurosurgeon may see a patient with schistosomiasis of the spinal cord who improves with chemotherapy alone. The general surgeon will be consulted for an acute abdomen, an abdominal mass, or because of bloody diarrhoea, while the cardiothoracic surgeon may be confronted by a massive pulmonary artery in a patient with cor pulmonale. A gynaecologist may not realize the cause of a ruptured ectopic pregnancy until seeing a report from the pathologist. A patient with schistosomiasis may present with a prolonged fever due to chronic salmonellosis. When schistosomiasis is suspected, the surgeon may be requested to perform a biopsy of the rectum or of another affected organ, to help with splenoportography, or to undertake a laparoscopy. When sclerotherapy or variceal ligation do not control variceal haemorrhage, the surgeon may need to perform a shunt or another surgical procedure, and may even contemplate a liver transplant when there is severe functional decompensation. Patients with hepatosplenic schistosomiasis are often chronic carriers of viral hepatitides, perhaps due to a state of relative immunosuppression, and thus present a hazard to the surgeon. Schistosomiasis has also been linked to a number of malignancies, the most well documented of which is squamous-cell carcinoma of the urinary bladder.

Table 1 The course of infection in the different stages of schistosomiasis

The parasites and their life-cycles

Unlike other trematodes that are hermaphroditic, schistosomes are digenetic with an asexual reproductive phase in fresh-water snails, and a sexual reproductive phase in the definitive host (Fig. 2). The elongated, delicate adult worms, which are 0.6 to 2.5 cm in length (depending on the species and sex), reside in pairs in terminal venules of the intestinal veins (or vesical veins in *S. haematobium* infection), attached to the vein wall by the male's ventral sucker. The female lies in the gynaecophoric canal of the male, held permanently *in copulo*. The worms usually live from 3 to 10 years, but exceptionally, as has been documented for *S. mansoni*, they may survive for up to 37 years. This longevity at a time when ova cannot be recovered from stool (or urine) is of particular importance to surgeons, who may encounter puzzling late sequelae.

The female *S. japonicum* lays up to 3000 eggs per day, whereas *S. mansoni* produces 300 per day. *Schistosoma mansoni* has lateral-spined eggs, *S. intercalatum* terminal-spined eggs, but *S. japonicum* and *S. mekongi* eggs only have small lateral knobs. Most of the eggs penetrate the vessel wall and pass into the lumen of the bowel, but others are swept back into the portal vein to the liver. When stools enter fresh water, each egg hatches and releases a free-swimming miracidium that penetrates certain snails, to form the sporocyst, with subsequent complex asexual multiplication. After some weeks, minute, fork-tailed cercariae are released from the snail during daylight, and using enzymes, they are able to penetrate intact human skin. Shedding their tails, they become schistosomules, which travel via the lung, the heart, and the systemic circulation to reach the portal venules in the liver, where they mature. Following copulation in the portal venules, the male and female schistosomes migrate in pairs against the portal flow to their final habitat, the mesenteric veins, or, in the case of *S. haematobium*, the vesical veins (via pelvic venous plexi collaterals). The prepatent period (between exposure to infection and the appearance of ova in urine or stool) is about 4 weeks for *S. mansoni*, from 6 to 10 weeks for *S. japonicum*, and about 12 weeks for *S. haematobium*.

Epidemiology

Schistosomiasis has afflicted man since ancient times, and ova have been found in 3000-year-old Egyptian mummies and also in ancient Chinese mummies. It is a major risk for all who come in contact with fresh surface water containing infected vector snails, particularly peasants (and their families), and agricultural and irrigation workers in rural areas of the tropics and subtropics. Its socioeconomic and public-health importance is second only to malaria among parasitoses. Despite warnings from the World Health Organization, large dam and irrigation schemes continue to be built in many developing countries, sometimes with inadequate measures to prevent proliferation of snails, encouraging the remorseless spread of both snails and epidemic schistosomiasis.

Geographical distribution

Figure 1 illustrates the principal geographical distribution of schistosomiasis. *Schistosoma japonicum* is endemic in the Far East, South-East Asia, and the Philippines. Although zoonotic foci exist in Japan and in Taiwan, human transmission no longer occurs in these countries. A similar species, *S. mekongi*, is present in the Mekong River valley, especially in Laos and Cambodia (Kampuchea), and *S. malayensis* is found in the mountains of peninsular Malaysia. The distribution of *S. mansoni* and *S. haematobium* overlaps in Africa, although *S. mansoni* is less widespread, and dual infection is commonplace. The Nile valley of Egypt is an area of high endemicity for both *S. mansoni* and *S. haematobium*. It has a focal distribution in parts of the Middle East, including Yemen and Saudi Arabia, but has been eradicated from Oman. *Schistosoma mansoni* is thought to have been carried by the slave trade from Africa to the West Indies, Brazil, Venezuela, and Surinam. *Schistosoma intercalatum*, which appears to cause minimal morbidity, is present in parts of central and western Africa.

Schistosoma japonicum is a zoonosis affecting 40 or more definitive hosts, including domestic animals and livestock such as the water buffalo, and it requires snails of the genus *Oncomelania* as the intermediate host. Since it is a zoonosis, and since the snail is amphibious, standard control measures (which include use of molluscicides) are largely ineffective. *Schistosoma mekongi* is also a zoonosis, infecting dogs, and the vector snail, which thrives in rivers rather than lakes, is *Neotricula aperta* (formerly named *Tricula aperta*).

Schistosoma malayensis, which affects an unknown number (probably several hundred) of aboriginal people (the *orang asli*) in the mountains of peninsular Malaysia, appears also to be a zoonosis. *Schistosoma mansoni* is principally confined to humans, with occasional infections in baboons (e.g. in Senegal, Kenya, and Saudi Arabia), in dogs in Kenya, and in rodents in Brazil. *Biomphalaria* spp. snails are the principal intermediate host for *S. mansoni. Schistosoma intercalatum* appears to be restricted to humans. *Schistosoma mattheei* and *S. bovis*, which are parasites of sheep and cattle in Africa, occasionally affect humans, but with minor morbidity.

Prevalence and intensity of infection

In all types of schistosomiasis, infection is contracted when individuals are exposed to the free-swimming cercariae liberated from infected fresh-water snails. In any endemic area, various social, religious, and cultural activities may affect the prevalence and intensity of infection. As a rule, in endemic areas where children play (and defaecate and urinate) in irrigation canals, lakes, and ponds from an early age, the prevalence of infection increases from the age of 3 or 4 years, and with continuing exposure the intensity increases to reach peaks in the 15- to 29-year age range, and then gradually subsides. [The intensity of infection is expressed in terms of eggs per gram (epg) of faeces or per 10 ml urine.] The higher the peak of intensity, the earlier it is reached; peak prevalence is usually a year or two later. This pattern applies generally to *S. mansoni* (and *S. haematobium*) but not to *S. japonicum*, in which there is no sustained decline in intensity of infection with age. In the floating villages in Kampuchea where children are continuously exposed to river water, the peak prevalence of *S. mekongi* infection occurs below the age of 10 years. Community-based surveys on *S. intercalatum* indicate a marked decline in prevalence and intensity in middle-aged individuals, with virtually no infection in elderly people.

When there is a large influx of people into an endemic area, as exemplified by refugee populations during war or civil unrest, epidemic infection may ensue. Military personnel are particularly at risk when they are deployed in endemic areas: in the Middle East; schistosomiasis has been one of the most common infections reported among foreign troops The modern phenomenon of the package tour for holiday-makers proceeding to exotic destinations where they may engage in water sports (e.g. swimming, scuba diving, canoeing, fishing) has produced numerous reports of imported infection in Europe, North America, and Australia.

Pathogenesis and pathology

Immune response to infection

Despite the overall success of schistosomes as parasites, the host does produce an effective immune response and, for example, only a small proportion of schistosomules survive to become adult worms. Both cellular and humoral mechanisms are involved in their elimination from the skin. Infections in the first 12 months of life are rare despite exposure, possibly a result of temporary protection by maternal antibodies transferred through the placenta. In some rodent models, the lungs appear to be an important site for schistosomule elimination, whereas in the guinea-pig, immunity to schistosomules appears to develop in the liver. Immune mechanisms include the ability of IgE antibodies and various cytokines to modulate macrophages, eosinophils, and platelets for effective antibody-dependent, cell-mediated cytotoxic reactions against the schistosomules. Eosinophils are more important in the skin, while macrophages are more important in the lung and liver phases of resistance. Inhibition of parasite migration may be a crucial feature of both skin- and lung-phase immunity in different rodent species.

A variety of antibodies is produced at various stages: these are directed against the structural components of schistosomes, and also against excreted and secreted products of maturing and adult worms. The evasion mechanisms that allow some of the schistosomes to achieve a successful state of parasitism include rapid turnover of membrane components, enzymatic cleavage of attached antibodies and complement (e.g. by serine proteases), tegumental structural development, the coating (or

masquerade) of the parasite surface with host antigenic components (e.g. ABO blood-group antigens, antigens of the major histocompatibility complex, low-density lipoproteins), antigenic mimicry, direct immunosuppressive effects, and the induction of humoral and cellular immunoregulatory mechanisms, including anti-idiotypic networks. A fascinating mechanism is the insertion of the host's decay acceleration factor on to the parasite's surface. Decay acceleration factor normally controls the activation of the host's alternative complement pathway; its presence on the worm's surface ensures that host complement will not be activated there. Furthermore, as the worm matures, its defences against oxidant damage appear to increase progressively.

There is evidence that concomitant immunity occurs in schistosomiasis, although it may be slow to develop. (Concomitant immunity describes the condition seen when an actively infected host partially or completely resists challenge infection by the same type of organism.) A significant degree of protective immunity may develop in children infected with *S. haematobium.*. The age-related decline in intensity and prevalence of infection is most developed after 15 years of age. The increase in IgE antibodies during reinfection (after praziquantel treatment) is likely to augment killing of schistosomules by IgE-dependent macrophages.

Response to egg deposition

Female worms lay a large number of ova every day. When ova are deposited, the body responds with a delayed-type hypersensitivity reaction (modulated by other factors) to form granulomas; these may fail to develop around ectopic eggs in the brain. Most of the morbidity in schistosomiasis is related to this granulomatous response and the subsequent fibrotic reaction initiated through the secretion by the granuloma cells of fibroblast-stimulating cytokines. Collagen synthesis is accompanied by collagen degradation, and the total amount of fibrosis is related to the imbalance between these two processes. The granulomatous response gives rise to early acute features such as intestinal SYmptoms (and haematuria in vesical schistosomiasis), while the resultant fibrosis contributes to late, chronic pathology. After about 4 weeks, the eggs die and granulomatous responses gradually subside, usually with healing and scar formation. In addition, as the disease becomes established, granuloma formation is spontaneously suppressed (modulated), with marked reduction in both size and amount of inflammation around new eggs entering the tissues.

Complications of ectopic egg deposition

The granulomatous response to ectopic egg deposition leads to many of the characteristic complications of schistosomiasis, including periportal fibrosis with portal hypertension in *S. mansoni, S. japonicum*, and *S. haematobium* infection, pulmonary hypertension and cor pulmonale, especially in *S. mansoni* and *S. japonicum* infection, and neuroschistosomiasis, notably in *S. japonicum* and *S. mansoni* infection (see below). Figure 3 indicates the route followed by ectopic ova (and occasionally adult worms). Involvement of the following structures by ova (and in some cases by adult worms) has been reported: peritoneum (retroperitoneal fibrosis may cause ureteric obstruction), gallbladder (wall thickening, calcification), bile ducts (fibrous strict-ures), kidneys, epididymis, testes (mass; giant fibrous pseudo-tumour), seminal vesicles (mass, haematospermia), prostate, salpinges (may cause ruptured ectopic pregnancy), ovaries, uterus, placenta, vagina, cervix (ulcerated masses), breast, adrenals, pancreas, thyroid, lymph nodes, heart, eyes, joints, skin, and parotids. Additional areas involved in schistosomiasis japonica have included pericardium, myocardium, breast, and muscle. A syndrome of massive egg deposition throughout the body, with associated dwarfism in some instances, has been described.

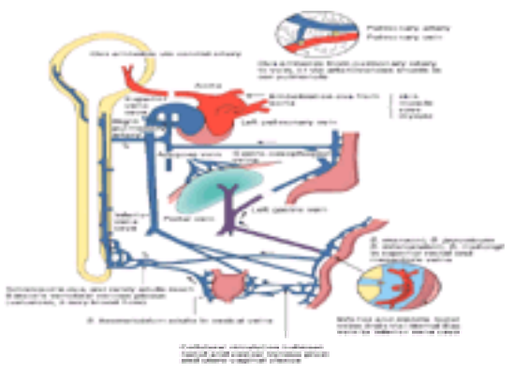


Fig. 3. Diagrammatic representation of routes followed by ectopic ova in schistosomiasis.

Interruption of these complex processes by chemotherapy at an early stage may reverse them and may also augment modulation, with the result that reinfection may not lead to further pathology. Even advanced disease (with persisting viable schistosomes) has been shown to improve to a greater or lesser extent after chemotherapy.

Morbidity in schistosomiasis

The insidious nature of schistosomiasis makes it difficult to measure population morbidity. During the first two decades of life, morbidity correlates with intensity of infection. Studies in the Sudan have shown a 20 to 30 per cent reduction in oxygen intake in *S. mansoni*-infected farmers; in Burundi there was a significant association between intensity of infection and diarrhoea, fatigue, hepatomegaly, and splenomegaly, all of which improved after treatment. The heights of Egyptian children with schistosomal hepatic fibrosis can be more than two standard deviations below the mean for their age and sex, and circulating levels of insulin-like growth factor, free l-thyroxine (T4), and cortisol may be reduced significantly. Delayed puberty, hypogonadism, and amenorrhoea have been reported from several different endemic regions. A gene encoding for pro-opiomelano-cortin, the precursor for both b-endorphin and adrenocorticotrophic hormone, has recently been identified in *S. mansoni.* Both b-endorphin and adrenocorticotrophic hormone are produced by the schistosomes, and since such peptides are implicated in gonad and endocrine regulation or gonadotrophin release, this factor might be related to the hypogonadism reported in humans and animals with chronic schistosomiasis.

Although the majority of infected populations have relatively minor (or even no) evidence of schistosomiasis, a small proportion develops serious morbidity, which poses a grave burden in many endemic countries. *Schistosoma mansoni* and *S. japonicum* are the leading cause of portal hypertension world-wide, and resulting haemorrhage from oesophageal varices may be one of the leading causes of death in endemic areas. Coinfection with viral hepatitis B or C, with the development of chronic infection and cirrhosis, significantly worsens the prognosis in hepatosplenic schistosomiasis. Cor pulmonale and neuroschistosomiasis, caused by all three major species, result in many deaths. *Schistosoma haematobium* causes severe urological disease, which contributes to the high incidence of endstage renal failure, and bladder cancer, in countries such as Egypt, where many uraemic patients die because facilities for dialysis and renal transplantation are relatively unavailable. Acute toxaemic schistosomiasis (most common in *S. japonicum* infection), although rare, may be lethal.

Genetic aspects

Community studies have revealed a high degree of aggregation or ‘overdispersion': i.e., a small proportion of individuals harbour most of the worms, and they are predisposed to severe disease and heavy reinfection following chemotherapy. Severity of disease is affected by coexistent infections, nutritional status, parasitic strain differences, and genetic factors. In Egyptians, hepatosplenic disease is associated with certain HLA haplotypes. The *in vitro* human T-cell responsiveness to *S. japonicum* soluble egg antigen may be determined by a gene in the *HLA-DQ* locus; two alleles, *HLA-DRB1* 1202* and *HLA-DQA*0601*, have recently been shown to be strongly associated with clinical resistance to advanced schistosomiasis japonica in China. Genetic studies in Brazil suggest the presence of a codominant major gene controlling human susceptibility to infection: there is a 0.2 to 0.25 frequency of the deleterious allele. A ‘racial' predisposition may exist in the degree of response to schistosomiasis: for example, in Brazil, whites are particularly prone to develop hepatosplenomegaly, while blacks with an equal infection load are less susceptible; and hepatosplenic and hepatointestinal disease are usually more severe in Egyptian than in Caribbean or South American populations. In populations south of the Sahara, *S. mansoni* is generally considered to result in fewer complications.

Clinical aspects

Table 1 summarizes the course of infection in the different stages of schistosomiasis.

History of exposure

Patients give a history of residence in, or a visit to, an endemic area, where swimming or other skin exposure to fresh water may have occurred. Often, adults from endemic countries cannot be specific about childhood exposure to fresh water, but a returning visitor or holiday-maker may recall a specific incident. Exposure of any part of the skin, however briefly, may result in infection: this includes wading, swimming, scuba diving, fishing, sailing, canoeing, white-water rafting, and water-skiing;

also such mundane pursuits as washing cooking utensils or clothes in lakes or rivers. Drinking infected water is not regarded as a significant risk: cercariae do not appear to penetrate the buccal mucosa or to survive digestive processes. Bathing in fast-running streams, especially at high altitude (e.g. over 2000 m), is unlikely to result in infection.

Symptoms, signs, and management

1. The stage of invasion: cercarial dermatitis (also called swimmer's itch, fisherman's itch, rice-paddy itch, clam-digger's itch)

Initial exposure may produce a mild, papular, pruritic dermatitis, usually affecting the legs, 12 to 24 h after exposure, and persisting for up to 2 weeks. This reaction is more severe in those who have been previously sensitized by exposure to non-pathogenic avian or mammal schistosomal cercariae (which are present worldwide in fresh-water lakes and streams), and in whom erythema, vesicles, and later pustules may form. Residual pigmentation at the site of these lesions may persist for weeks or months. This stage also includes the migration of schistosomules through the lungs, and to the liver, and there may be fever, cough, transient pulmonary infiltrates, myalgia, abdominal pain, and splenomegaly.

2. Acute schistosomiasis (Katayama fever)

Synchronous with the onset of egg laying, an antigen–antibody reaction develops 2 to 9 weeks after infection, especially with *S. japonicum* and *S. mansoni*, and which lasts 1 to 2 months. This is particularly prominent in non-immune individuals; previously exposed persons in endemic areas may have no visible reaction. Fever, chills, sweating, headache, dry cough, nausea, vomiting, abdominal pain, hepatosplenomegaly, lymphadenopathy, and urticaria may occur. Various neurological signs, including fits and transient myelopathy, have been reported. The erythrocyte sedimentation rate is raised and there is a leucocytosis, with eosinophilia ranging from 10 to 70 per cent. Immunoglobulins, especially IgE, IgG, and IgM, are raised. Faecal egg counts may be high or low (or even absent). The differential diagnosis at this stage includes typhoid, brucellosis, infectious mononucleosis, the Churg–Strauss syndrome, and other acute parasitoses including hookworm infection, strongyloidiasis, trichinosis, and amoebiasis.

A more severe, toxaemic form with disseminated granulomas in many organs may develop: laparoscopy frequently discloses granulomas containing eggs on the surface of the liver, intestines, and visceral peritoneum. An ulcerative enterocolitis may result, with perforation and purulent peritonitis. This toxaemic form may be confused with typhoid, malaria, tuberculosis, pneumonia, acute kala-azar, prolonged salmonella septicaemia, disseminated strongyloidiasis (e.g. in the acquired immune deficiency syndrome), trichinosis, amoebic or bacillary dysentery, and other surgical infections of the abdomen.

3. Chronic forms of schistosomiasis

(a) Chronic intestinal schistosomiasis (hepatointestinal schistosomiasis)

The resultant granulomas, fibrosis, and vascular changes, which can involve all areas of the small and large intestine in *S. mansoni, S. japonicum, S. mekongi*, and *S. intercalatum* infection, may eventually become symptomatic. *Schistosoma mansoni* eggs and granulomas have been found throughout the gastrointestinal tract, including the stomach, small and large intestines, pancreas, gallbladder, the peritoneum, and the liver. In Egypt, gross intestinal lesions, especially in the colon, are present at autopsy in 16 per cent of patients with *S. mansoni* infection. *Schistosoma haematobium* eggs are also commonly found in rectal biopsies, in the appendix, and in polyps.

Bowel SYmptoms, including nausea, abdominal pain, anorexia and weight loss, and lassitude and myalgia, occur in about 40 per cent of those with *S. mansoni* infection in Egypt, especially in the 10- to 14-year age group. Bloody diarrhoea with mucus is much more common in Egypt than in Brazil because of the higher incidence of colonic polyposis in Egypt. Blood and protein loss is worse in patients with malnutrition. Pellagra has been found in 24 per cent and finger clubbing in 35 per cent of patients with polyposis.

On examination, the colon may be tender and thickened, and a pericolic mass is present in 16 per cent of patients with polyposis. The polyps may be sessile or pedunculated and may slough, ulcerate, and become secondarily infected. They are most common in the sigmoid colon but may be present anywhere in the large bowel and may vary from a few to many. The mucosa is granular and hyperaemic, and bleeds easily. Barium enema, especially with double contrast, demonstrates the polyps well. Colonic polyps are significantly reduced in size after parasitological cure, and this is accompanied by increases in haemoglobin, serum albumin, and serum iron.

An extensive, pseudoneoplastic form involving the colon, ileum, mesentery, or retroperitoneum may present as an abdominal mass (bilharzioma) with abdominal pain or obstruction. The bowel may be thickened, giving rise to a rigid tube known as ‘schistosomal colonic cord'; this can later become stenotic. Deposition of eggs in the small intestine is less common, but it may lead to partial villous atrophy, heavy round-cell infiltration of the lamina propria, and thickening of the muscularis mucosa. Ballooned, empty villi resembling the pathognomonic lesion of intestinal lymphangiectasia may be seen.

Schistosomiasis and appendicitis Schistosomiasis *per se* rarely causes appendicitis, except when fibrosis occludes the lumen, or if there is heavy infection of the appendix. In endemic areas, schistosome ova are often found in appendices removed for clinical appendicitis, or at autopsy. Appendicitis is an important complication of hepatointestinal *S. japonicum* infection. In *S. haematobium* infection the density of eggs is often higher in the appendix than in adjacent regions, and appendicitis may become SYmptomatic during heavy infections. In Nigeria, 1 per cent of surgically removed appendices examined in 1984 contained schistosome ova. Half of these appendices were removed for clinical appendicitis, and of these 70 per cent showed suppurative changes.

(b) Periportal fibrosis and portal hypertension (hepatosplenic schistosomiasis)

Periportal fibrosis (Symmer's or claypipe-stem fibrosis) is the most common, serious complication of schistosomiasis: in Egypt, autopsy and hospital-record studies indicate that it develops in about 50 per cent of patients with heavy *S. mansoni* infection. It develops when ova, especially of *S. mansoni* and *S. japonicum* (but also *S. haematobium*), embolize the portal veins. Although hepatic fibrosis occurs in *S. mekongi*-infected patients, its contribution is hard to define, since patients in endemic areas are usually infected with *Opisthorcis viverrini* flukes, which live in the biliary tree. *Schistosoma intercalatum* does not appear to produce periportal fibrosis.

The pathogenesis of hepatic disease involves persisting T-cell stimulation by anti-idiotypic antibodies to egg antigens, and recent evidence suggests a possible role of autoreactivity to liver antigens (the hepatic asialoglycoprotein receptor) in hepatosplenic schistosomiasis. After a period of 5 to 10 years the liver becomes hard, with protuberances on its surface, usually with prominence of the left lobe. Section shows the pathognomonic lesions of coarse, periportal fibrosis. Hepatic blood flow is partly maintained by compensatory hypertrophy of the hepatic artery (which may explain the raised wedge hepatic venous pressure in some patients with severe disease). Immunoglobulin may be deposited in the spaces of Disse, and collagen in the subendothelial basement membrane, with capillarization of hepatic sinusoids, and the formation of a barrier between the hepatocytes and the circulating blood.

The haemodynamic disturbances give rise to a presinusoidal type of portal hypertension resulting in the appearance of varices, which may bleed, collateral circulation, which may carry eggs to the lungs (Fig. 3), and splenomegaly, giving rise to hypersplenism. Splenic enlargement is due to hyperplasia of the cellular components in the initial phase and as a result of raised portal pressure in later stages. Associated immunological disturbances lead to other complications, including (in *S. mansoni* infection) immune-complex glomerulonephritis. Hepatocyte function is preserved, unless coexistent hepatitis B or C infection is present, when the features of hepatosplenic schistosomiasis are significantly augmented. It is possible that these patients are more likely to have persistent positivity for hepatitis B surface antigen

The physical signs in the abdomen include an enlarged, firm liver, which may be tender, with left-lobe prominence. The spleen may be grossly enlarged and superficial collaterals may develop. In the late stages the liver may shrink, and ascites and oedema may develop. The main complication is oesophageal variceal haemorrhage, with fulminant or repetitive haematemesis and melaena: rupture often occurs at the gastro-oesophageal junction. Up to about one-third of patients with hepatosplenic disease will suffer a haematemesis.

Differential diagnosis Hepatosplenic schistosomiasis has to be differentiated from malaria, which is common in many of the endemic areas, and which may give rise to hyper-reactive malarial splenomegaly (formerly called tropical splenomegaly syndrome); and from portal venous thrombosis, kala-azar, lymphoma, and Laennec's and postnecrotic cirrhosis.

The management of variceal haemorrhage

Lessons learnt from non-schistosomal patients Recent studies in patients with non-schistosomal portal hypertension have shown that:

- Sclerotherapy or variceal band ligation effectively controls variceal haemorrhage and prevents early rebleeding

- Somatostatin, or its longer-acting synthetic analogue, octreotide (or the prodrug terlipressin) infusion is as effective as sclerotherapy in controlling variceal haemorrhage
- A 5-day continuous infusion of somatostatin was as effective as sclerotherapy in both preventing early variceal rebleeding and maintaining low mortality following acute variceal haemorrhage
- Somatostatin is associated with a lower rate of complications than sclerotherapy
- Early administration of natural somatostatin continued for 120 h, combined with additional bolus injections, is more effective than placebo in the overall control of acute variceal haemorrhage in patients with cirrhosis undergoing sclerotherapy.

The general principles of management in schistosomal variceal haemorrhage are similar to those in portal hypertension from other causes. However, although the onset of haemorrhage is unpredictable, the outcome of uncomplicated haemorrhagic episodes is much better than it is in alcoholic cirrhosis, largely because of the well-preserved hepatocyte function. The general management of acute haemorrhage includes blood transfusions and administration of vasopressin (or another of the newer pharmacological agents). Emergency sclerotherapy may be needed, although large bleeding gastric varices may be technically very difficult to inject. Endoscopic variceal band ligation is also now proving of value, both in the acute situation and as a safer alternative to sclerotherapyas prophylaxis against subsequent bleeds. Balloon tamponade, used when other emergency procedures fail, should be limited to use by experienced personnel, and should preferably be accompanied by endotracheal intubation to avoid aspiration of gastric contents.

Surgical management Repeated major bleeds are an indication for a more invasive surgical procedure (Table 2). Future research may take the direction of ‘pre-primary prophylaxis', i.e., prevention of the development of the varices themselves.

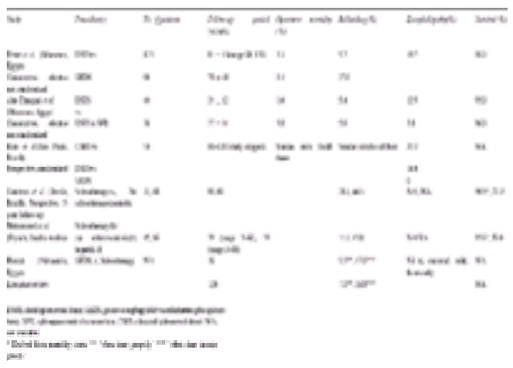


Table 2 Surgical management of schistosomal variceal haemorrhage

Gastro-oesophageal decongestion plus splenectomy

The Hassab procedure of gastro-oesophageal decongestion plus splenectomy is associated with a low incidence of encephalopathy, but the recurrent bleeding rate on its own has been high. However, versions of it are used in Egypt, Sudan, China, Brazil, and elsewhere. Reviews have suggested that the procedure is well suited in schistosomiasis when combined with sclerotherapy. Hassab's own recent review concludes that when done properly (perhiatal devascularization of the lower 7.5 to 10 cm of the oesophagus, complete separation of the stomach from its bed, ligation of the left gastric artery at the lesser curvature, peritonization of the greater curvature, nursing the patient on the right side, suction drainage of the splenic bed, and early respiratory exercises are essential technical points), gastro-oesophageal decongestion plus splenectomy was as effective in controlling bleeding as the distal splenorenal shunt; it produced better hepatic and cardiac function, better quality of life, and better survival with nil or minimal encephalopathy, and combined with sclerotherapy it formed the ideal therapy for bleeding varices in all types of pathology. Combination with sclerotherapy was useful but oesophageal transection was harmful. Variceal rebleeding varied from 5.5 per cent in 3 years to 7 per cent in 10 years when done properly, to 17 per cent and 18.8 per cent when done incompletely; higher rates included minor bleeding of gastritis. Encephalopathy varied from nil to a minimal incidence of mild forms.

Portosystemic shunts and other procedures

The classical portocaval shunt is unacceptable in this condition, since there is a high incidence of portosystemic encephalopathy. The distal splenorenal shunt (Table 2) is currently the best shunt procedure for schistosomal variceal haemorrhage and it is now widely used. Shunt patency rates are high and adequate portal blood flow into the liver is maintained in the majority of patients. Splenopancreatic disconnection as an adjunct to the distal splenorenal shunt appears to be well suited to this condition: although it prolongs the operation and increases the incidence of pancreatitis, it further reduces the incidence of encephalopathy and chronic hyperbilirubinaemia.

Transjugular intrahepatic portosystemic shunting

TIPSS, now in use for (non-schistosomal) portal hypertension and other indications for over a decade, is technically difficult, requires intensive follow-up, has a number of complications (including frequent portosystemic encephalopathy and frequent spontaneous stenoses), and the stents that are used in TIPSS are themselves expensive. Furthermore, the majority of the randomized studies that are available to date show that survival in portal hypertension as a whole is similar between patients receiving TIPSS or endoscopic treatment. For all these reasons it seems unlikely that TIPSS will become important in the overall management of schistosomiasis. Other than in schistosomiasis, TIPSS has been useful for refractory ascites and related complications, such as hepatorenal SYndrome and hepatic hydrothorax, but randomized studies are still lacking. In addition, TIPSS has been applied successfully to patients with Budd–Chiari syndrome, portal venous thrombosis, and before liver transplantation.

Splenectomy

Splenectomy, with or without gastro-oesophageal devascularization, eliminates abdominal discomfort, corrects hypersplenism, improves the general state of the patient, and may reduce portal pressure by about 40 per cent. It has also been claimed to improve or cure infantilism. In Recife, Brazil, children with hepatosplenic schistosomiasis who undergo splenectomy are autoimplanted with 100 g of spleen into the greater omentum. One group in Brazil has performed about 100 segmental splenectomies (with gastro-oesophageal disconnection) for schistosomal and other causes of portal hypertension, with claimed good results, but this has not yet been adequately assessed by other groups. Splenectomized patients living in a malarial area should receive lifelong antimalarial prophylaxis.

Liver transplantation

Orthotopic liver transplantation should be considered in patients who are badly decompensated, or who have very poor functional hepatic reserve, rather than for the portal hypertension *per se*.

Pharmacological management The use of somatostatin (octreotide) and terlipressin for acute variceal haemorrhage has been mentioned above.

Propranolol may be particularly suitable in patients with schistosomiasis awaiting surgery as well as in the longer term. Given over 3 months, doses sufficient to lower the resting pulse by 25 per cent have been reported to lead to a significant and sustained reduction in portal pressure, with a reduction in spleen size, in about 50 per cent of patients. A recent study from the Sudan concluded that propranolol reduces mortality in patients with portal hypertension secondary to schistosomiasis.

Initial randomized trials show that nitrates, such as isosorbide-5-mononitrate, potentiate the effect of b-blockers in reducing portal pressure in non-schistosomal cirrhotics. Nitrovasodialators are also recommended for use with vasopressin to counteract its systemic vasoconstrictor effects, especially coronary vasoconstriction.

(c) Pulmonary schistosomiasis and cor pulmonale

Migration of schistosomules through the lung may be associated with bronchospasm, fever, and cough. During treatment, dying or dead adult worms may sometimes reach the lung, where they can cause a ‘verminous' pneumonia. Spontaneous pneumothorax has also been reported, but is rare. The major pulmonary complication of chronic schistosomiasis is pulmonary hypertension, leading to cor pulmonale. This develops after the onset of significant portal hypertension, when the collateral circulation allows shunting of *S. mansoni* and *S. japonicum* eggs to the pulmonary circulation. *Schistosoma haematobium* eggs, which enter the systemic venous

system directly from the vesical veins, are often found in high concentration in the lungs, but cor pulmonale is rare in *S. haematobium* infection.

Egg deposition results in pulmonary necrotizing arteriolitis, which destroys the intima and occludes blood vessels, while newly formed vessels may progress into an angiomatoid lesion. There is also medial hypertrophy of neighbouring vessels. Significant venoarterial shunting may develop. The increased arteriolar resistance, loss of adaptability, and presence of fibrosis give rise to pulmonary hypertension, which, when severe, may itself lead to further vascular damage. The pulmonary arteries become extremely dilated or even aneurysmal. The end result is cor pulmonale. While a large proportion of patients with *S. mansoni* infection may have some lung disease, only about 15 per cent of those with Symmer's fibrosis develop clinically important chronic lung changes. The peak incidence is between 12 and 35 years, and males are more frequently affected than females. *Schistosoma japonicum* infection may also present with chronic bronchitis, bronchiectasis, and emphysema. With the exception of evidence of schistosomal portal hypertension, the clinical features are similar to other cases of pulmonary hypertension and cor pulmonale. The differential diagnosis includes primary pulmonary hypertension, parenchymal diseases, and multiple pulmonary emboli. A pulmonary biopsy is helpful if the diagnosis is in doubt. Fibreoptic bronchoscopy can be used to diagnose pulmonary schistosomiasis: the findings included redness, swelling, ulceration of the mucosa, and the presence of miliary nodules.

Some patients with hepatosplenomegaly do not develop cor pulmonale, but instead develop cyanosis and clubbing with no or minimal pulmonary hypertension. The cyanosis often develops after splenectomy. Possible explanations for this syndrome include pulmonary arteriovenous shunts, portopulmonary anastomoses, or a reduction in the affinity of haemoglobin for oxygen.

Antischistosomal chemotherapy and supportive care can produce symptomatic improvement in patients with low-grade pulmonary hypertension, but not in patients with cyanosis. Chemotherapy should be given to all actively infected patients, however, if only to reduce further egg deposition.

(d) Neuroschistosomiasis

Brain All three major schistosomes can cause disease in the central nervous system. As noted, Katayama fever may produce immunologically mediated, transient encephalopathy or myelopathy. In chronic, uncomplicated *S. haematobium*, and to a lesser extent, *S. mansoni* infections, deposition of eggs in the brain has been found in 3 to 56 per cent of individuals at autopsy (depending on the population examined); the eggs may be carried to the brain through the vertebral venous plexus or arteriovenous shunts in the lung. The ova of *S. japonicum* (which are smaller than those of the other two) are particularly likely to reach the brain. Rarely, adult worms (but not *S. japonicum*) have been recovered from cerebral vessels. In hepatosplenic schistosomiasis, deposition of *S. mansoni* ova in the brain is more frequent, especially if there is coexistent cor pulmonale, and the route to the brain appears to be via pulmonary arteriovenous shunts (Fig. 3). Cerebral ova deposition may be asymptomatic, but acute encephalitis and subarachnoid haemorrhage can occur.

Clinical evidence of cerebral neuroschistosomiasis is most common in *S. japonicum* infection, affecting between 1.7 to 6.5 per cent of patients admitted to hospital. About 50 per cent of individuals have abnormal electroencephalograms. It is regarded as an important cause of focal epilepsy in the Far East: in the Philippines the most frequent manifestations are jacksonian and psychomotor seizures commencing after 21 years of age. About 20 per cent subsequently develop hemi- or monoparesis, and blindness or hypopituitarism have been reported. Computed tomographic (**CT**) scanning or magnetic resonance imaging (**MRI**) reveal hypodense, multinodular lesions with pronounced contrast enhancement and surrounding oedema. Chemotherapy results in improvement and the disappearance of seizures in most patients. Acute neuroschistosomiasis must be differentiated from cerebral malaria, other cerebral parasitoses including African and American trypanosomiasis, toxoplasmosis, cysticercosis, trichinosis, tuberculosis, and cerebral abscesses. If the diagnosis is confirmed or suspected, chemotherapy should be instituted immediately without embarking on neurosurgery.

Spinal cord Deposition of ova in the spinal cord is less frequent than in the brain, and has been reported in 0.3 to 13.3 per cent of patients with schistosomiasis in Africa and Brazil at autopsy. Egg deposition is more common in *S. haematobium* than in *S. mansoni* or *S. japonicum* infection. In hepatosplenic schistosomiasis caused by *S. mansoni*, myelopathy is rarely encountered. Myelopathy is usually due to an intramedullary granuloma of the conus medullaris, presenting acutely as an areflexic, flaccid paralysis. Laboratory investigations are of limited value except in travellers to endemic areas. Serological tests, including enzyme-linked immunosorbent assay (**ELISA**) and antigen detection, may indicate exposure to schistosomiasis. The cerebrospinal fluid usually reveals a mild pleocytosis with lymphocytosis rather than eosinophilia, and its protein and glucose concentrations are usually normal. Immunological tests of the cerebrospinal fluid, including the circumoval precipitin test, and the ELISA using soluble egg antigen, may support the diagnosis of schistosomal myelopathy. A higher antibody titre in cerebrospinal fluid than in serum may also support neuroschistosomiasis. Depending on the pathology, myelograms may reveal an irregular, partial, or complete block, extending the length of a vertebral body, with irregular ‘trifid' edges. Both CT and MRI may provide supportive evidence. In endemic areas for schistosomiasis, a high index of suspicion is necessary, and even if there are no clinical features to support schistosomiasis, antischistosomal chemotherapy should be started pending the results of investigations. The differential diagnosis includes tuberculosis, tropical spastic paraplegia (caused by the human T-cell lymphotropic virus type 1 (HTLV-1)), acquired immune deficiency SYndrome-associated myelopathy, and neoplasms. Even after many weeks of paraplegia, patients with myelopathy may make a substantial recovery over succeeding months. Explorative surgery is generally reserved for those who fail to respond to chemotherapy, who deteriorate, or in whom the diagnosis remains uncertain. If a granuloma is encountered unexpectedly at surgery, it should not be resected, but a biopsy should be obtained.

(e) Renal involvement in schistosomiasis

Ureteric strictures and urolithiasis contribute to renal damage through hydronephrosis and pyelonephritis. In one study, 14 per cent of 150 Egyptian children with *S. haematobium* infection had hydronephrosis. The kidney, however, may be affected in all three main schistosomal infections.

Immune complexes are deposited consistently in the renal glomeruli in *S. haematobium* infections. The extent to which they contribute to clinical disease is unknown, but in Egypt the nephrotic syndrome is prevalent in areas with high endemicity of *S. haematobium* infection. In an Egyptian investigation, 48 (20 per cent) of 240 patients with active *S. mansoni* infection but no symptoms suggestive of glomerular disease had proteinuria; of these, 15 agreed to renal biopsy and eight were positive: the most common lesion was focal mesangial proliferation, with IgM and C3 deposits. It seems, therefore, that both *S. haematobium* and *S. mansoni* infections contribute to the high prevalence of renal disease in Egypt, where the incidence of endstage renal failure may be as high as 200 new patients per million population per year. In Brazil, where only *S. mansoni* infection occurs, practically all types of glomerulopathy have been seen in hepatosplenic schistosomiasis, but the most frequent is a generalized membranoproliferative glomerulonephritis, presenting with the nephrotic syndrome. Deposits of IgG, IgM, IgE, and IgA have been found. Focal glomerular sclerosis is also common. Schistosomal antigens have on occasion been found in both basement membranes and in the mesangium.

Endstage renal disease due to schistosomiasis can be managed by renal transplantation.The need for transplanation is increasing in many countries where schistosomiasis is endemic, and in this circumstance the donor should be carefully assessed to exclude renal anatomical changes caused by (previous) schistosomiaisis. Schistosomiasis has been passed from a donor to a recipient in a transplanted kidney when the adult worms survived the procedure.

(f) Prolonged salmonella septicaemia

This may occur in patients with schistosomiasis, especially the hepatosplenic form. The salmonellae adhere to the surface of the worm, especially the male, and can also be found in their gut. More than 20 species of *Salmonella* of both human and animal origin can cause this syndrome. Fever develops insidiously, is high, irregular, continuous or intermittent, and is often accompanied by headache, sweating, chills, weight loss, diarrhoea, abdominal pain, and hepatosplenomegaly with mild lymphadenopathy. Epistaxis, petechiae, purpuric lesions, haematuria, and proteinuria may develop. The diagnosis is based on the history of schistosomiasis and prolonged fever (which may extend to 2 years), and is confirmed by blood culture. Salmonella may also be found in stool, urine, bile, or bone marrow. The Widal reaction may be positive for *Salmonella typhi* or *S. paratyphi*. The differential diagnosis from kala-azar can be very difficult. Both the salmonellosis and schistosomiasis must be treated with appropriate chemotherapy, otherwise there is risk of relapse.

(g) Schistosomiasis and the risk of malignancy

Trematodes have frequently been associated with an increased risk of developing malignancies in man. While the role of schistosomiasis in the aetiology of some malignancies is still controversial, there is strong evidence linking *S. haematobium* with squamous-cell and less commonly, transitional cell carcinoma of the bladder. Advanced *S. mansoni* infection has been reported to be associated with follicular lymphoma of the spleen.

Schistosoma japonicum is not regarded as a risk factor for neoplasia; however in one study the onset of colonic cancer was 10 years earlier in infected than in uninfected individuals, and in an autopsy study, the rate of colonic cancer was 25 times greater in infected than in uninfected persons. The carcinoma was reported to be typically a well-differentiated adenocarcinoma with pseudopolyps, and calcified eggs were present in the tissues. In China, patients with the epithelial hyperplastic type of colonic polyps are considered to be at risk for the development of colon carcinoma. Although the incidence of hepatocellular carcinoma appears to be higher in patients infected with *S. japonicum*, evidence of a significant association is lacking.

Diagnosis of schistosomiasis

Haematological and biochemical tests

Eosinophilia with a direct eosinophil count exceeding 600 cells/µml is typically present in Katayama fever and early disease, but may disappear in later stages. Other causes of eosinophilia should be excluded.

Granulocytopenia, thrombocytopenia, and anaemia result from hypersplenism, but the marrow may show failure of granulocyte maturation due possibly to an inhibitory serum factor. Since hepatocyte function is generally well preserved, liver function tests are usually normal. Some patients ultimately develop hepatic decompensation, and liver function tests then become grossly deranged. A form of disseminated intravascular coagulopathy has been described in patients with advanced disease.

Identification of ova in biopsies and stool

Finding viable ova in urine, stool, or biopsy material is the basis of diagnosis. When the ova are live, active movements of the contained ciliated miracidium can be seen under high-power microscopy. When only calcified eggs are present in biopsies, infection is no longer active in that site, but living worms may persist elsewhere. *Schistosoma mansoni* and *S. intercalatum* eggs may also be observed occasionally in urine. Fresh or occult blood in the stool indicates a relatively high intensity of infection, and in this case a single, thick smear should detect ova; however, examination of up to fivethick smears, each 40 mg, is recommended if infection is suspected. In the Kato–Katz modification, the stool is concentrated by passing it through a mesh with the addition of a stain (e.g. malachite green) to facilitate egg identification. When infection is light (which is usually the case in returning holiday-makers), a more certain means of diagnosis is by concentrating up to 2 g of stool using the formol–ether or merthiolate–iodine–formaldehyde techniques, and the preserved specimen can be shipped by post to experienced laboratories if necessary.

All types of schistosomiasis, including urinary, can be diagnosed by examining six rectal mucosa snips obtained with biopsy forceps through a proctoscope, pressing them between two slides, and examining them unstained, under low-power (×10) microscopy. In the numerous examples of ectopic egg deposition in various tissues, diagnosis usually depends on biopsy of the structure involved, and identification of egg material in the centre of granulomas. In intestinal disease, liver biopsy (to detect periovular granulomas) is unnecessary and should generally be avoided.

Immunodiagnostic tests

In prepatent infection, and at times in late established disease, eggs cannot be demonstrated in stool, ova, rectal, or other biopsy. In some complications, for example myelopathy and intracranial disease, it is preferable to make the diagnosis by non-invasive means, including immunodiagnosis. Serodiagnosis for antibodies is most applicable to the prepatent period, or in late disease, when identification of eggs in urine or stool (for various reasons) has not been successful, but false-negative results may occur, and in an endemic area tests may not distinguish between active infection, previous infection, or reinfection. Positive tests may be obtained in individuals who have been exposed to non-pathogenic schistosomes of various avian and mammalian hosts, which occur world-wide. At present, no immunoassay test is superior to another, and the diagnostic value of a single antigen–antibody detection system may vary greatly in different epidemiological settings. Using antigens derived from worms, cercariae, or eggs, tests include ELISA IgG, IgM, and IgE (which indicate past or current infection), indirect haemagglutination assay, and IgM immunofluorescent antibody test to detect recent infection. Following successful treatment, the antibody titre should decline over a period of several months.

Antigens secreted by cercariae, worms, or ova may appear in serum, urine, breast milk, or cerebrospinal fluid. Most useful for detecting the presence of viable worms are the excretory–secretory antigens, also known as gut-associated antigens. When worms regurgitate digested blood, these polypeptides, glycoproteins, proteoglycans, and polysaccharides are released into the circulation. Two proteoglycan gut-associated antigens, the circulating anodic antigen and the circulating cathodic antigen, appear in serum and also urine, and breast milk. Unfortunately, antigen detection tests are of limited value in clinical practice, and those using monoclonal antibodies are only available in reference laboratories. In one study of travellers returning from endemic areas, when infection was light and worms were still maturing, the ELISA for circulating anodic antigen was only 20 per cent sensitive. They are more useful in established infection, when quantitatively, their concentration correlates well with the worm load, and a successful response to therapy can be confirmed by their fall.

Imaging techniques

Ultrasonography

Ultrasonography is a valuable, non-invasive tool in the epidemiological assessment and individual diagnosis of both urinary and hepatosplenic schistosomiasis. Ultrasound of the abdomen demonstrates periportal fibrosis of the liver; the appearances (multiple rounded, sheet, or band-like echogenic areas with a characteristic and fairly central sonolucent area) are pathognomonic (Fig. 4). In schistosomiasis japonica, the fibrosis is described as ‘network,' ‘cobweb,' or ‘fish scale' in appearance. Ultrasonography can determine the diameter of the portal vein, while Doppler studies can determine the direction (usually hepatopetal) and velocity of flow. A portal vein of more than 1.7 cm in diameter and a splenic vein of more than 1.2 cm in diameter indicate that oesophageal varices are likely to be present. Varices are present in more than 80 per cent of patients with hepatosplenic schistosomiasis. Splenoportography (which is hardly necessary if ultrasonographic examination is expertly interpreted) confirms the raised portal pressure, usually to more than 200mm H_2O, but the wedge hepatic venous pressure is usually normal. Ultrasonography will also delineate the size of the spleen, and major collateral vessels.

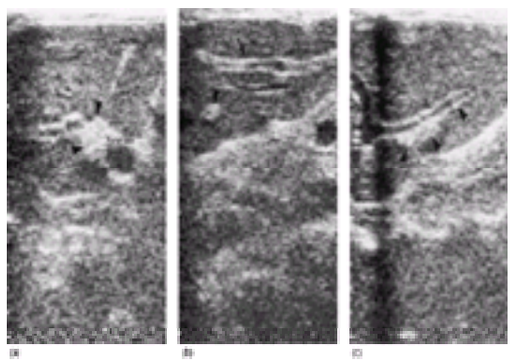


Fig. 4. Ultrasonography of periportal fibrosis.

Other methods

Plain radiology is useful in identifying schistosomal calcification in tissues, but this is better seen on CT. ‘Turtleback' calcification of the liver is pathognomonic for *S. japonicum* infection. Both CT and MRI are helpful in the diagnosis of neuroschistosomiasis.

Barium and endoscopic studies confirm the presence of varices and define their extent. Laparoscopy may be of value in assessing intraperitoneal pathology and obtaining biopsies of granulomas.

Chemotherapy

Chemotherapy is one of the most effective methods of controlling human schistosomiasis, both in individuals and in populations. ‘Swimmer's itch' is generally treated with antihistamines alone, since none of the antischistosomal drugs is effective against either cercariae or schistosomules. Treatment of Katayama fever requires the use of corticosteroids; the place of antischistosomal drugs in this context is controversial. Praziquantel is the most versatile antischistosomiasis drug: it is rapidly lethal to the adults of all species, is easily administered, is generally well tolerated, and is relatively inexpensive. It causes instantaneous muscular contraction in trematodes followed by spastic paralysis. The schistosomes become detached from the endothelium of veins, and ultimately disintegrate. There is also a rapid resolution of cellular reaction and fibrosis in tissues containing eggs. For all types of schistosomiasis except *S. mekongi*, the dose is 40 mg/kg given as a single dose or in two divided doses. For *S. mekongi*, the recommended dosage is 60 mg/kg and repeated doses may be required. Complete cure is achievable in non-endemic areas, but is unnecessary in endemic areas, where a reduction of egg output by 90 per cent is enough to eliminate serious morbidity. Side-effects of praziquantel are usually minor, comprising nausea, abdominal pain, dizziness, and rarely a transient rash. It is usually withheld during the first trimester of pregnancy.

Oxamniquine is an alternative drug with activity against *S. mansoni* and it has been used extensively in South America. It is well tolerated, but occasionally may cause dizziness, drowsiness, and headache. It is not recommended in pregnancy. The dosage in South America is 15 mg/kg given as a single dose. Some oxamniquine resistance has been encountered in Egypt and in East and southern Africa, where a higher dose (30–60 mg/kg) is recommended.

Many apparently irreversible lesions of chronic schistosomiasis, especially in the young, may resolve partly or completely with chemotherapy. Hepatosplenic disease responds well, with a marked reduction in the size of the liver and spleen in about half of the patients treated. Even in hyperendemic regions, where reinfection is inevitable, praziquantel is effective in preventing progressive hepatic fibrosis. The administration of a single dose of praziquantel, 40 mg/kg, once a year for 3 years to individuals with early *S. mansoni* hepatosplenic schistosomiasis living in a hyperendemic area in Madagascar reduced the degree of hepatic fibrosis from 28 per cent to 10 per cent. Adjuvant treatment with colchicine, b-aminopropionitrile, and other agents to attempt reduction of fibrosis remains experimental.

Decompensated liver disease has a poor prognosis but may be safely treated with praziquantel, although improvement is probably due to better care and diet than to therapy alone. Colonic polyps shrink, with improvement in symptoms and in the anaemia and hypoproteinaemia. Neuroschistosomiasis is frequently cured or markedly improved unless treatment is unduly delayed. Drug treatment should also be given before surgery, particularly before shunting procedures for portal hypertension, to avoid prolonged embolization of the lungs by eggs passing through the surgically created shunt.

Vaccines

Although the development of a vaccine for human schistosomiasis has been targeted as a priority by the World Health Organization, an effective vaccine has not been developed as yet. However, the fact that humans have the ability to acquire natural immunity to schistosome infection, together with the successful use of attenuated vaccines in animals both under laboratory and field conditions indicates that a vaccine may be feasible. Attenuated vaccines for schistosomiasis are considered neither safe nor practicable for human use, however, and therefore other approaches are being considered.

Potential vaccines that have been evaluated experimentally include those using antigens from schistosomules, adults, and ova. Immunization with crude or purified antigenic extracts of various life-cycle stages, while producing promising results, has been hampered by the inability to produce enough antigen from laboratory-maintained life-cycle stages for large-scale commercial developments. The recombinant DNA approach to produce large amounts of the relevant antigen is an attractive alternative. A genetically engineered schistosome glutathione-*S*-transferase molecule is being tested in animals as a vaccine. It appears that such a vaccine can provide significant protection in cattle against *S. mattheei* under conditions of low to moderate natural infection.

Clinical studies using two recombinant *S. mansoni* molecules, paramyosin and MAP-4/TPI, and laboratory-based studies on two other molecules, IrV-5 and MAP-3 (Sm23), using the DNA approach for vaccination, have recently been approved for implementation in Egypt as part of the 5-year Schistosomiasis Vaccine Development Programme (which is supported by USAID). Further experimental work with animal models is still needed for peptide/protein vaccine candidates and for DNA vaccines. A major concern is the possibility of exacerbation of the sequelae of chronic infection by the use of vaccines.

Further reading

Abu-Elmaged KM *et al.* Should both schistosomal and non-schistosomal variceal bleeders be disconnected? *World Journal of Surgery* 1991; **15**: 389–98.

Atkin SL, Kamel H, El Hady AMA, El Badawy SA, El-Ghobary A, Dick WC. Schistosomiasis and inflammatory polyarthritis: a clinical, radiological, and laboratory study of 91 patients infected by *Schistosoma mansoni* with particular reference to the diarthrodial joint. *Quarterly Journal of Medicine* 1986; **229**: 479–87.

Avgerinos A, Nevens F, Raptis S, Fevery J. Early administration of somatostatin and efficacy of sclerotherapy in acute oesophageal variceal bleeds: the European Acute Bleeding Oesophageal Variceal Episodes (ABOVE) randomised trial. *Lancet* 1997; **350**:1495–9.

Boisier P, Ramarokoto C-E, Ravaoalimalala VE, Rabarijaona L, Serieye J, Roux J, Esterre P. Reversibility of *Schistosoma mansoni*-associated morbidity after yearly mass praziquantel therapy: ultrasonographic assessment. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1998; **92**: 451–3. [Reviews beneficial effects of annual praziquantel therapy on hepatomegaly and fibrosis.]

Bosch J. Medical treatment of portal hypertension. *Digestion* 1998; 59(5): 547–55.

Brandt CT, Caneca OA, Tavares DJ, Avila L. Surgical hepatosplenic schistosomiasis mansoni in children: a Doppler duplex study of the portal vein and the hepatic artery. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1995; **89**(1): 70–1.

Butterworth AE. Immunity and morbidity in schistosomiasis. In: McAdam KPWJ, ed. *New strategies in parasitology*, pp. 193–210. Churchill Livingstone, Edinburgh, 1989.

Chen MG. *Schistosoma japonicum* and *S. japonicum*-like infections: epidemiology, clinical and pathological aspects. In: Jordan P, Webbe G, Sturrock RF, ed. *Human schistosomiasis*, pp. 237–70. CAB International, Oxford, 1993.

Conn HO. A randomized comparison of three types of surgery in schistosomal portal hypertension: many fewer answers than questions. *Hepatology* 1994; **20**(2): 526–8.

Conn HO, Lebrec D, Terblanche J. The treatment of oesophageal varices: a debate and a discussion. *Journal of Internal Medicine* 1997; **241**(2): 103–8.

Contractor QQ, Contractor TQ, Kher YR. Sclerotherapy in bleeding gastric varices of hepatic schistosomiasis. *Journal of Clinical Gastroenterology* 1996; **23**(2): 97–100.

Cordeiro F. Variceal sclerosis in schistosomotic patients: a 5 year follow-up study. *Gastrointestinal Endoscopy*, 1990; **36**: 475–8.

Daar AS. Renal transplantation in developing countries. In: Morris PJ, ed. *Kidney transplantation: principles and practice*, 4th edn, pp. 478–503. Saunders, Philadelphia, 1992.

Daar AS. Schistosomiasis (bilharziasis). In: Morris PJ, Malt RA, ed. *Oxford textbook of surgery*, 1st edn, pp. 2511–223. Oxford University Press, 1994.

De Bont J, Vercruysse J, Grzych JM, Meeus PF, Capron A. Potential of a recombinant *Schistosoma bovis*-derived glutathione *S*-transferase to protect cattle against experimental and natural *S. mattheei* infection. *Parasitology* 1997; **115** (Pt 3): 249–55.

el Tourabi H, el Amin AA, Shaheen M, Woda SA, Homeida M, Harron DW. Propranolol reduces mortality in patients with portal hypertension secondary to schistosomiasis. *Annals of Tropical Medicine and Parasitology* 1994; **88**(5): 493–500.

Escorsell A *et al.* for the Variceal Bleeding Study Group. Randomized controlled trial of sclerotherapy versus somatostatin infusion in the prevention of early rebleeding following acute variceal hemorrhage in patients with cirrhosis. *Journal of Hepatology* 1998; **29**(5): 779–88.

Evans AC, Martin DJ, Ginsburg BD. Katayama fever in scuba divers. *South Africa Medical Journal* 1991; **79**: 271–4.

Ezzat FA, *et al.* Selective shunt versus non-selective shunt surgery for management of both schistosomal and non-schistosomal variceal bleeders. *Annals of Surgery* 1990; **212**: 97–108.

Feu F, Ruiz del Arbol L, Banares R, Planas R, Bosch J, for the Variceal Bleeding Study Group. Double-blind randomized controlled trial comparing terlipressin and somatostatin for acute variceal hemorrhage.. *Gastroenterology* 1996; **111**(5): 1291–9.

Garcia-Villarreal L *et al.* Transjugular intrahepatic portosystemic shunt versus endoscopic sclerotherapy for the prevention of variceal rebleeding after recent variceal hemorrhage. *Hepatology* 1999; **29**(1): 27–32.

Giboda M, Smith JM. Impact of antifibrotic treatment of the course of *Schistosoma mansoni* infection in murine model. *Memorias do Instituto Oswaldo Cruz* 1997; **92**(5): 693–7.

Hassab MA. Gastroesophageal decongestion and splenectomy in the treatment of esophageal varices in bilharzial cirrhosis: further studies with a report on 355 operations. *Surgery* 1967; **61**(2): 169–76.

Hassab MA. Gastro-esophageal decongestion and splenectomy GEDS (Hassab), in the management of bleeding varices. Review of literature. *International Surgery* 1998; **83**(1): 38–41. [An authoritative review of the technique.]

Helal TE, Danial MF, Ahmed HF. The relationship between hepatitis C virus and schistosomiasis: histopathologic evaluation of liver biopsy specimens. *Human Pathology* 1998; **29**(7): 743–9.

Henderson JM. Surgical methods for the prevention of first and recurrent variceal bleeding. *Zeitschrift für Gastroenterologie Verhandlungsband* 1988; **26** (Suppl. 2): 49–53.

Hipgrave DB, Leyden JA, Walker J, Biggs B-A. Schistosomiasis in Australian travellers to Africa. *Medical Journal of Australia* 1997; **166**: 294–7.

Ishii A, Matsuoka H, Aji T, Ohta N, Arimoto S, Wataya Y, Hayatsu H. Parasite infection and cancer: with special emphasis on *Schistosoma japonicum* infections (Trematoda). A review. *Mutation*

Research 1994; **305**(2): 273–81.

Jordan P, Webbe G, Sturrock RF. *Human schistosomiasis*. CAB International, Oxford, 1993. [Multiauthor review of all major aspects of schistosomiasis; remains a useful reference text of current knowledge up to 1993.]

Kardorff R *et al. Schistosoma mansoni*-related morbidity on Ukerewe Island, Tanzania: clinical, ultrasonographical and biochemical parameters. *Tropical Medicine and International Health* 1997; **2**(3): 230–9.

Leptak CL *et al.* Schistosome egg granulomas and hepatic expression of TNF-a are dependent on immune priming during parasite maturation. *Journal of Immunology* 1997; **158**: 301–7.

Mergo PJ, Ros PR. Imaging of diffuse liver disease. *Radiologic Clinics of North America* 1998; **36**(2): 365–75.

Mohamed A, Elsheikh A, Ghandour Z, Al Karawi M. Impact of hepatitis C virus infection on schistosomal liver disease. *Hepatogastroenterology* 1998; **45**(23): 1492–6.

Mohammed AE, Al Karawi MA, Al Otaibi A, Hamid MA. Results of sclerotherapy in 100 patients. Comparison of the outcome between schistosomiasis and hepatitis B. *Hepatogastroenterology* 1989; **36**: 333–6.

Morris W, Knauer CM. Cardiopulmonary manifestations of schistosomiasis. *Seminars in Respiratory Infections* 1997; **12**(2): 159–70.

Mwatha JK *et al.* High levels of TNF, soluble TNF receptors, soluble ICAM-1, and IFN-gamma, but low levels of IL-5, are associated with hepatosplenic disease in human schistosomiasis mansoni. *Journal of Immunology* 1998; **160**(4): 1992–9.

Ouma JH, Mbugua GG, Butterworth AE. Morbidity in schistosomiasis. *Parasitology Today* 1992; **8**: 55.

Pereira LM *et al.* Specific liver autoreactivity in schistosomiasis mansoni. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1997; **91**(3): 310–14.

Petroianu A, da Silva RG, Simal CJ, de Carvalho DG, da Silva RA. Late postoperative follow-up of patients submitted to subtotal splenectomy. *American Surgeon* 1997; **63**(8): 735–40.

Raia S, da Silva LC, Gayotto LC, Forster SC, Fukushima J, Strauss E. Portal hypertension in schistosomiasis: a long-term follow-up of a randomized trial comparing three types of surgery. *Hepatology* 1994; **20**(2): 398–403.

Raia S, Mies S, Alfieri F. Portal hypertension in Mansonic schistosomiasis. *World Journal of Surgery* 1991; **15**: 176–87.

Richter J *et al.* Hepatosplenic schistosomiasis: comparison of sonographic findings in Brazilian and Sudanese patients—correlation of sonographic findings with clinical symptoms. *Radiology* 1992; **184**(3): 711–16. [Delineation of the sonographic parameters, based on extensive field work in Brazil and Sudan.]

Rossle M, Siegerstetter V, Huber M, Ochs. A The first decade of the transjugular intrahepatic portosystemic shunt (TIPS): state of the art. *Liver* 1998; **18**(2): 73–89. [Critical assessment of the evolving indication for the technique.]

Scrimgeour EM, Daar AS. Schistosomiasis—diagnostic and clinical update. *Australia and New Zealand Journal of Surgery* (in press).

Scrimgeour EM, Gajdusek DC. Involvement of the central nervous system in *Schistosoma mansoni*, and *S. haematobium* infection. A review. *Brain* 1985; **108**: 1023–38. [Review of all published reports (in English, French, German, Spanish, Italian, and Portuguese) of schistosomiasis mansonica and haematobia affecting the central nervous system. Meta-analysis of series of case reports.]

P>**Shekhar KC.** *Schistosoma malayensis*: the biologic, clinical, and pathologic features in man and experimental animals. In: Sun T, ed. *Progress in clinical parasitology*, Vol. 2, pp. 145–78. Field and Wood Medical, New York, 1991.

Shields R *et al.* A prospective randomised controlled trial comparing the efficacy of somatostatin with injection sclerotherapy in the control of bleeding oesophageal varices. *Journal of Hepatology* 1992; **16**(1–2): 128–37.

Stelms FF, van der Werf M, Niang M, Gryseels B. Four years' follow up of hepatosplenic morbidity in a recently emerged focus of *Schistosoma mansoni* in northern Senegal. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1997; **91**: 29–30.

Van Lieshout L *et al.* Analysis of worm burden variation in human *Schistosoma mansoni* infections by determination of serum levels of circulating anodic antigen and circulating cathodic antigen. *Journal of Infectious Diseases* 1995; **172**: 1336–42. [Important contribution to diagnosis of acute and chronic schistosomiasis.]

Vargas HE, Gerber D and Abu-Elmagd K. Management of portal-hypertension related bleeding. *Surgical Clinics of North America 1999*; **79**: 1–22. [A comprehensive review of recent developments and techniques.]

Waine GJ, McManus DP. Schistosomiasis vaccine development—the current picture. *Bioessays* 1997; **19**(5): 435–43

Waine GJ, Ross AG, Williams GM, Sleigh AC, McManus DP. HLA class II antigens are associated with resistance or susceptibility to hepatosplenic disease in a Chinese population infected with *Schistosoma japonicum. International Journal of Parasitology* 1998; **28**(4): 537–42.

Waine GJ, Yang W, Scott JC, McManus DP, Kalinna BH. DNA-based vaccination using *Schistosoma japonicum* (Asian blood-fluke) genes. *Vaccine* 1997; **15**(8): 846–8.

Warren WD, Millikan WJ. The relative role of sclerotherapy vs. surgical procedures in portal hypertension. *Advances in Surgery* 1990; **23**: 1–19.

Wynn TA. Development of an antipathology vaccine for schistosomiasis. *Annals of the New York Academy of Sciences* 1996; **797**: 191–5.

47.2.3 Schistosomiasis (bilharziasis) of the genitourinary tract

Imtiaz Husain

[Pathology and clinical features](#)
[Bladder effects](#)
[The upper tracts](#)
[Other organ involvement](#)
[Management](#)
[Acute urinary bilharziasis](#)
[Chronic bilharzial uropathy](#)
[Carcinoma of the bilharzial bladder](#)
[Further reading](#)

Urinary bilharziasis affects some 200 million people in endemic regions of Africa and south-west Asia (see Chapter 47.2.2). In Europe, a small focus of infection existed in Algarve, southern Portugal, but this is now either dead or dying out. The causative parasite *Schistosoma haematobium* (Fig. 1) is one of three main schistosome species whose primary host is man. The other varieties, *S. mansoni*, which coexists with *S. haematobium* and is also prevalent in South America and some Caribbean islands, and *S. japonicum*, found in the Far East, are responsible for the intestinal forms of the disease. Any area of fresh water that can support the secondary host, an amphibian snail of the *Bulinus* species, is a potential source of infection once the schistosome ova have been deposited in the water by contamination with urine or faeces from an infected individual.



Fig. 1. Scanning electron micrograph of adult male and female *Schistosoma haematobium* worms in copula (by courtesy of Professor Emile A.. Malek and Drs Kuntz, Tulloh, Davidson, and Huang).

The morbidity of bilharziasis relates little to the acute infection, which is now easily and safely treated with effective oral antischistosomal agents, but more to the lingering sequelae of host reaction to the eggs of the parasite, manifested by progressive dysfunction of the urinary tract and the potential development of bladder carcinoma. The impact of the disease on mortality in an area of high endemicity may be judged from an autopsy study performed in Cairo over a 4-month period in 1974 (Table 1). Urinary bilharziasis was recorded in 61 per cent of individuals examined and was the direct or contributory cause of death in almost 10 per cent. There have been several approaches to controlling infestation, but these have failed either because of difficulties with public education or due to the overriding need to preserve the local ecological milieu. Bilharziasis is therefore likely to remain a major health problem for the foreseeable future.

No bilharziasis found:	74 cases
Bilharziasis detected:	116 cases
10.4 %	Direct or underlying cause of death
6.8 %	Contributed to death
23.3 %	Concomitant severe bilharziasis
59.5 %	Incidental finding of bilharziasis

From Smith JH *et al.*, see Further Reading.

Table 1 Frequency and mortality of urinary bilharziasis in 190 consecutive autopsies at Cairo University Hospitals

Pathology and clinical features

The pathological effects of *S. haematobium* infection in man are chiefly the result of an intense and evolving tissue reaction to the eggs. Unlike intestinal bilharziasis, the liver is not significantly affected and the main target organs are the bladder, distal ureters, the prostate, seminal vesicles, and spermatic cord. For reasons that are not entirely clear, bilharziasis rarely affects females, but when it does lesions may also develop in the uterus, vagina, and vulva.

The adult worms reside mainly in the retrovesical plexus of veins, where the eggs are deposited, deep in the wall of the bladder. Each egg develops a miracidium within a week of being shed and excites an intense, local, histiolytic tissue reaction that carries it through the lamina propria and urothelium to be shed into the bladder lumen, together with extravasated blood. The characteristically fusiform and terminally spined eggs are readily identified on urine microscopy (Fig. 2).

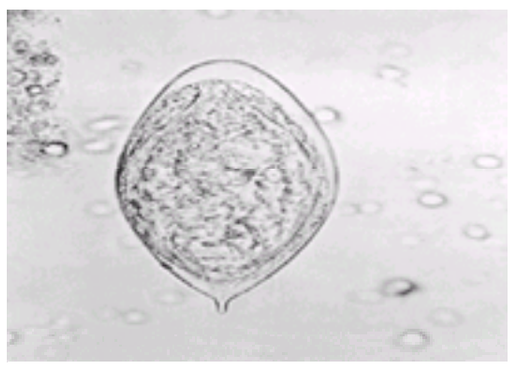


Fig. 2. *Schistosoma haematobium* egg in urine (by courtesy of Professor Emile A. Malek).

When infected urine reaches a body of fresh water, the eggs hatch and the released miracidia swim actively around in search of a suitable snail host. The larval form of

S. haematobium develops exclusively within a specific *Bulinus* species, where it matures in about 30 days. The released cercaria, which are actively mobile in water, are able to penetrate intact human skin to continue the sexual phase of the lifecycle. The larvae, or schistosomulae, gain access to the venous or lymphatic systems, and are thence carried through the heart and lungs into the portal vein. Here, the adult worms mature within a few weeks, the female being carried in the gynaecophoric canal of the male (Fig. 1) during the final stage of their odyssey down to the pelvic veins.

Clinical manifestations are rare before release of the parasite eggs into the urine, usually some 3 months after exposure. Initially, there is painless passage of blood at the end of micturition; later SYmptoms are those of an intense haemorrhagic cystitis with dysuria, pain, and frequency. In some highly endemic areas, blood in the urine is not considered abnormal and is even regarded as a sign of puberty in boys. Patients may occasionally complain of lassitude, body aches, and late afternoon fever. When bladder ulceration occurs, pain is usually intense, being referred to the perineum and the tip of the penis. In this active phase of infestation, there is heavy excretion of eggs in the urine, accompanied by pyuria, and sometimes bacteriuria. Anaemia and eosinophilia may also be evident.

Bladder effects

Ninety per cent of the deposited eggs are distributed in the bladder wall where, initially, a granulomatous lesion develops in the lamina propria or, rarely, deep in the interstitial tissue between muscle bundles. The ova secrete a histiolytic antigen that is tissue-fixed and evokes a cell-mediated immune response in surrounding host tissue. The granulomatous reaction is characterized by an eosinophilic, cellular infiltrate, and the overlying urothelium grows inflamed and proliferative. The extent of the primary lesion and its subsequent evolution through the healing and chronic phases depend on several factors, including the intensity of the tissue egg load, frequency of reinfestation, the nutritional state of the host, the timing of antischistosomal treatment, and any superimposed bacterial infection. The pattern of the developing urothelial lesions is extraordinarily varied and may take the form of atrophic, proliferative, or metaplastic change (Table 2).

Acute
Haemorrhagic cystitis
Acute erosion or ulcer
Pseudotubercle
Granulomatous polyp or papilloma
Chronic
'Sandy patch'
Chronic ulcer
Fibrocalcific polyp
Bladder wall calcification
Bladder contracture
Bladder neck fibrosis
Cystitis cystica or glandularis
Squamous metaplasia or leukoplakia
Carcinoma

Table 2 Bladder lesions in urinary bilharziasis

Atrophic changes, manifested by erosions of the surface epithelium, may progress to chronic, indolent ulceration if the underlying mass of ova has compromised the blood supply. Papillomas are usually evident during the active phases of bilharzial infection and may persist as benign fibrous or fibrocalcific polyps. The subepithelial collections of ova may present ‘pseudotubercles' or ‘sulphur granules' (raised, seed-like, golden-yellow specks) at cystoscopy (Fig. 3). These tubercles are initially surrounded by hyperaemia, but in the chronic stages of the disease the urothelium becomes atrophic and thin, so that the masses of underlying ova show through as a characteristic ‘sandy patch'. These, together with the generalized waxy-yellow appearance of the bladder epithelium, are unmistakable cystoscopic evidence of past bilharzial infection.

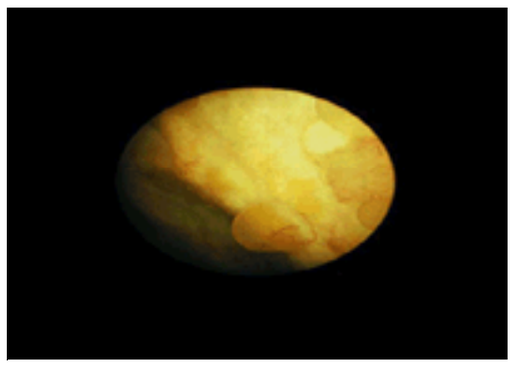


Fig. 3. Bladder endoscopic features of chronic, burnt-out bilharziasis. The urothelium is typically waxy-yellow in appearance and there are cystitis cystica changes overlying the ureteric orifice and trigone. There is also a patch of ‘sulphur granules' at 3 o'clock.

Metaplastic change is most common in areas of high endemicity, and is probably related to the severity and frequency of infestation and to superadded bacterial infection. Squamous metaplastic lesions develop surface keratinization in up to 30 per cent of cases, and may then present as leukoplakia. Columnar metaplasia is typified by predominant glandularis or cystica change, and occasionally there are epithelial enclosures. It is not unusual to find all of these metaplastic changes coexisting in the same bilharzial bladder lesion (Fig. 4). Although cystitis cystica is undoubtedly a benign lesion, both glandular metaplasia and leukoplakia are premalignant in nature. *Schistosoma haematobium* infection is closely associated with bladder cancer, and recent evidence suggests that this relation may be indirect: the irritant stimulus of the erupting schistosome ova may potentiate the effect of a pre-existing latent carcinogenic agent, possibly an environmental or bacterially produced nitrosamine. This sequence produces an irreversible change that is later accelerated to tumour growth by a further factor or by a threshold population of tumour-initiated urothelial cells. Bacterial infection thus plays an important part in the biogenesis of potentially malignant change in the bladder associated with bilharziasis. This inference is supported by reports that the rate of urinary-tract infection is at least 10 times higher in regions of endemic bilharziasis than elsewhere. Two-thirds of bilharzial bladder cancers show squamous-cell differentiation, and squamous metaplasia occurs in around 65 per cent of cases overall. The remaining are transitional-cell tumours, although adenocarcinomas are not uncommon.

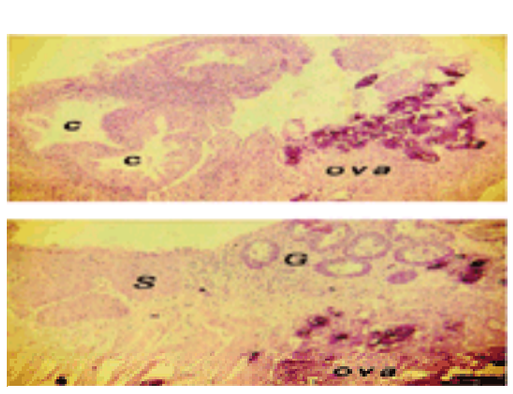


Fig. 4. Patterns of bladder urothelial metaplasia in chronic bilharziasis:(a) cystitis cystica; (b) squamous (S) and glandular (G) metaplasia in the same bladder biopsy specimen. Note the darkly stained dead ova in the deeper layers of the bladder-wall sections.

Severe and prolonged bilharzial cystitis may be complicated by bladder contracture or by bladder outlet obstruction due to involvement of the posterior urethra. The characteristic dystrophic calcification that is often a sign of past bilharzial infection (Fig. 5) causes surprisingly little bladder dysfunction, and these patients usually have

normal bladder capacity and compliance.



Fig. 5. Plain radiograph showing characteristic bilharzial calcification of the bladder and distal ureters.

The upper tracts

The incidence of change in the upper tract appears to vary widely from one endemic region to another, being chiefly influenced by the intensity of infestation and tissue egg burdens. In comparable autopsy studies of schistosomal uropathy, hydronephrosis was present in one-sixth of individuals studied in Nigeria but in more than one-third of those studied in Egypt.

The early exudative and polypoid lesions present as obstruction and ureteral filling defects on urography (Fig. 6). These appear to be more common in children, and usually resolve with adequate antischistosomal treatment. At a later stage, the mucosa becomes roughened and may contain tiny tubercles, like sago grains, or cystica change. In the vicinity of egg-deposition sites, the muscle of the ureteral wall becomes damaged by the toxic effect of the surrounding granulomatous reaction, resulting in segmental dilatation and peristaltic dysfunction (Fig. 7). These focal changes are not usually obstructive and the related kidney is little changed.



Fig. 6. Acute bilharzial ureteropathy, manifesting as ureteritis cystica and polypoid changes in the upper ureters; these resolved with antibilharzial chemotherapy.



Fig. 7. Segmental dilatation of the distal ureters consequent upon past bilharziasis; these changes are non-obstructive and require no surgical consideration.

Repeated or intense infestation is associated with necrosis of the ureteral wall, with disappearance of muscle fibres around sites of ova deposition and consequent dilatation, toxic atonicity, and tortuosity of the affected segment. The loss of muscle contractability, both circular and longitudinal, leads to flaccid elongation with folding and kinking that becomes fixed by periureteric adhesions (Fig. 8). The extending retroperitoneal fibrosis results in the whole ureter becoming encased, thickened, and pulled medially. Until it was realized that dilatation occurred at the site of, rather than above, the diseased ureteric segment, it was widely believed that the principal lesion in the bilharzial ureter was that of stricture, a stenosis with fibrosis. Although up to two-thirds of ureters affected in this way show reflux, the true incidence of stricture is probably not more than 5 per cent. The reflux is usually low pressure and, in the absence of infection, renal damage has not been demonstrated to occur over observation periods of up to 20 years.

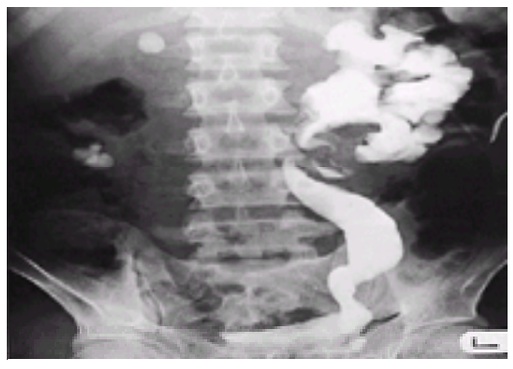


Fig. 8. Descending left nephrostogram demonstrating a tortuous and thickened bilharzial megaureter. Both kidneys were severely damaged and contained stones. This patient required bilateral total ureteral replacement with an interposed loop of ileum.

Ureteric calcification is seen in at least one-third of ureters affected by bilharziasis, and is most common in the distal segment. A characteristic punctate type of calcification may follow cystica change (ureteritis calcinosa). Although stone frequently coexists with bilharzial uropathy, there is no evidence that the incidence is significantly greater in endemic areas. The widely dilated and obstructed ureter is predisposed to the development of stasis stones and often these are of giant size.

In view of the protean nature of these changes, it is surprising that long-term effects on renal function are rare. In regions of high endemicity, however, up to 13.5 per cent of bilharzial patients over the age of 20 years will present with one non-functioning kidney and a few will progress to endstage renal failure.

Other organ involvement

The seminal vesicles are affected in around 80 per cent of patients with bilharzial uropathy: infection probably occurs by reflux from the posterior urethra via the ejaculatory ducts. Haematospermia is a characteristic clinical feature, and rectal examination discloses bilaterally enlarged, firm, and nodular or cystic seminal vesicles. In about one-third of cases, seminal ejaculate contains bilharzial ova. Plain radiographs will sometimes show ‘honeycomb' calcification of the vesicles. Even in severe cases, however, the ampullary canals and ejaculatory ducts remain patent, and bilharziasis is unlikely to play a direct part in the development of obstructive infertility in men. Bilharzial prostatitis is relatively uncommon, and the epididymis and testis are rarely involved.

Involvement of the female genital tract is generally believed to be uncommon. However, a gynaecological survey of women of child-bearing age in an endemic area of Madagascar showed that 15 per cent had schistosome ova in lesions in the vagina or cervix, detected either by biopsy or Papanicoulou smears. Papillomatous lesions may present in the vulva and vagina, usually before the age of puberty. In adults, cervical ulceration is more common, and lesions may rarely occur in the uterus, ovaries, or fallopian tubes. There is no evidence of an association of schistosomiasis with the development of cervical cancer.

Management

Acute urinary bilharziasis

The symptom complex of painful micturition, haematuria, and frequency in a man from an endemic area is virtually diagnostic; diagnosis is quickly confirmed when viable ova are found on urine microscopy. Endoscopy is not useful in the acute disease and the risk of bacterial infection may increase the morbidity.

The introduction of the antischistosomal drug praziquantel has revolutionized the medical treatment of bilharziasis. An isoquinolone–pyrazine derivative, praziquantel has few side-effects and is almost 100 per cent effective against all three principal schistosomes. A single oral dose of 40 mg/kg is recommended for *S. haematobium* infection and this can be safely repeated for the treatment of subsequent reinfestation.

In addition to its schistosomicidal action, praziquantel may also reduce the intensity of the established pathological lesion. Ultrasonographic studies have demonstrated that fibrosis accompanying the hepatomegaly associated with intestinal bilharziasis is markedly diminished by this treatment, and there is also abundant urographic and ultrasonographic evidence of resolution of granulomatous polyps and upper tract obstruction in *S. haematobium* infection, especially in children. There is, therefore, some basis for repeating medical treatment when patients present with florid, indolent lesions and seemingly burnt-out disease, with no evidence of fresh infestation.

Chronic bilharzial uropathy

In endemic regions, patients exposed to repeated infestations usually experience persistent and varying degrees of micturition dysfunction, occasionally with dysuria and perineal discomfort, for the rest of their lives. When symptoms are severe, urodynamic investigation can differentiate among patients with significant bladder-outlet obstruction, high voiding pressures, and poor flow rates, and those showing detrusor dysfunction with low voiding pressures and instability. In the former group, endoscopic incision or surgical Y–V revision of the bladder neck may be indicated. Surgery is not generally helpful in the latter.

Chronic disease of the upper tract usually manifests as ‘tiredness', with vague loin pains or aching in the iliac fossae. Occasionally, patients present with silent renal failure and imaging investigations will reveal grossly dilated ureters and irreversibly damaged, hydronephrotic kidneys. It is not uncommon for patients to be unaware of, or to deny, previous bilharzial infection: the first sign of past disease is often the chance finding of the characteristic ‘onion-leaf' bladder calcification on plain radiographs (Fig. 5).

Cystoscopy is the mainstay in diagnosis of chronic or past disease, the characteristic ‘sandy patches' and thickened, waxy-yellow change of the urothelium being tell-tale. Biopsy of chronic proliferative lesions will usually reveal dead or calcified ova. Although a high proportion of patients will demonstrate raised bilharzial antibody titres, immunological investigations have not proved reliable in diagnosis. Although the sensitivity of IgG enzyme-linked immunosorbent assay using keyhole limpet haemocyanin and soluble egg antigen has been shown to be 91 to 95 per cent in parasitologically negative individuals in endemic areas, specificity was only 40 per cent.

As noted earlier, changes in the upper tract, especially in children, may resolve appreciably with medical treatment. In the majority of the remainder, the infection has a minimal effect on renal function, and is non-obstructive, so that treatment is not required. Stenotic lesions are uncommon and usually affect the distal, juxtavesical ureter. These require resection and reimplantation with a cuffed or ‘nipple' type of neoureterocystostomy. It is unwise to attempt a tunnel reimplantation in the thickened bladder walls of patients with chronic bilharzial lesions, reflux being more acceptable than the risk of restenosis. Surgical tailoring of the wide and thickened ureters seen in these patients is also complicated and carries the risk of restenosis or fistula. Lengthy strictures may require treatment with a Boari flap reconstruction and, despite the thickened bladder wall, these are generally successful. Shorter strictures, less than 2 cm in length, may respond favourably to endourological treatment, either by ureterotomy and stenting or with balloon dilatation.

Extensive surgical lysis and straightening of thickened and tortuous megaureters should only be undertaken if X-ray fluoroscopy demonstrates the presence of reasonably effective ureteric peristalsis. Replacement of a wide and atonic megaureter by a pedicled ileal loop should be considered if a trial period of percutaneous nephrostomy drainage shows that kidney function is salvageable. Total ileal replacement of the ureters in bilharziasis (Fig. 9) has proved functionally rewarding over the long term.



Fig. 9. Ileal replacement of bilharzial megaureters. A single, isoperistaltic loop of terminal ileum is interposed between the dilated renal pelvis or upper ureters and the bladder.

Carcinoma of the bilharzial bladder

Bladder cancer in schistosomiasis commonly occurs between 35 and 45 years of age, approximately a decade earlier than in non-infected individuals. Patients frequently present with the same symptoms of a haemorrhagic cystitis experienced during the course of acute bilharzial infections, and this may explain why the disease is usually at an advanced stage when first diagnosed. About one-quarter of patients have inoperable disease at primary evaluation.

The tumours are massive and of the fungating type, occupying the vault, posterior walls, or lateral walls of the bladder (Fig. 10). Despite their size, however, the frequency of spread to the regional lymph nodes is relatively low, ranging from 15 to 17 per cent in radical cystectomy specimens. The majority of these tumours are of low-grade malignancy and the overall 5-year survival following radical cystectomy approaches 33 per cent. Radiotherapy is of little benefit, because of the associated postbilharzial fibrosis and the fact that most of these tumours are of the squamous-cell variety. Phase 2 trials of intravesical, keyhole limpet haemocyanin immunotherapy for superficial transitional-cell tumours associated with urinary schistosomiasis have shown successful reduction of their recurrence rates from 76.9 to

15.4 per cent.

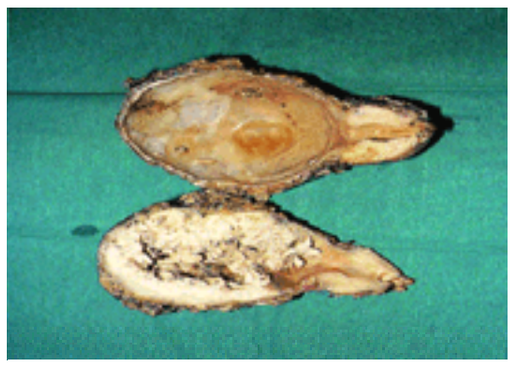


Fig. 10. Radical cystectomy specimen showing a massive verrucous, squamous-cell carcinoma in a bilharzial bladder; there was no involvement of regional lymph nodes (by courtesy of Professor Salah El-Faqih, Riyadh).

Orthotopic bladder substitution is being increasingly used in urinary diversion following radical cystoprostatectomy for bilharzial bladder cancers. However, recent work has demonstrated involvement of the prostatic urethra in 12.5 to 13.5 per cent of transitional- or squamous-cell tumours in such cases. Bladder substitution should not be considered in patients with multifocal or bladder-base tumours.

Further reading

Ghoneim MA *et al.* Cystectomy for carcinoma of the bilharzial bladder: 138 cases 5 years on. *British Journal of Urology* 1979; **51**: 541–4. [Analysis of survival data relative to stage and grade of bilharzial bladder cancers following radical cystectomy.]

Hicks RM. The canopic worm: role of bilharziasis in the aetiology of bladder cancer. *Journal of the Royal Society of Medicine* 1983; **76**: 16–22. [Review of environmental and cellular carcinogenetic factors in aetiology of bilharzial bladder cancer.]

Husain I, ed. *Tropical urology and renal disease*, pp. 211–80. Churchill Livingstone, Edinburgh, 1984. [Epidemiology, pathology and clinical management of genitourinary schistosomiasis.]

Leutscher P *et al. Schistosoma haematobium* induced lesions in the female genital tract in a village in Madagascar. *Acta Tropica* 1997; **66**: 27–33. [Community based survey of female genital schistosomiasis, using parasitological and ultrasound studies.]

Nabeeh A *et al.* Orthotopic bladder substitutes: histopathological risk factors. *Scandinavian Journal of Urology and Nephrology* 1995; **29**: 463–7. [Prostate involvement in 130 radical cystectomies for bilharzial bladder cancer.]

Smith JH, Kamal IA, Elwi A, von Lichtenberg F. A quantitative postmortem analysis of urinary schistosomiasis in Egypt. I. Pathology and pathogenesis. *American Journal of Tropical Medicine and Hygiene* 1974; **23**: 1053–8. [Pathological findings correlated to tissue egg burdens in 190 consecutive autopsies at Cairo.]

Wishahi MM, Ismail IM, Ruebben H, Otto T. Keyhole-limpet hemocyanin immunotherapy in the bilharzial bladder: a new treatment modality? Phase II trial: superficial bladder cancer. *Journal of Urology* 1995; **153**: 926–8. [Efficacy of intravesical keyhole-limpet haemocyanin in 13 patients with superficial transitional cell tumours associated with bilharziasis.]

47.2.4 Hydatid disease

William K. J. Huizinga, Christopher S. Grant and A. S. Daar

Introduction
Types of parasites
Lifecycle of *E. granulosus*
Pathology in the intermediate host
The hydatid cyst
Localization of hydatid cysts
Clinical presentation
Diagnosis
Laboratory tests
Radiology and ultrasound
Surgical treatment
Hepatic cysts
Extrahepatic cysts
Drug treatment
Further reading

Introduction

Hydatid disease is a parasitic infection caused by the larval stage of the cestode *Echinococcus granulosus*. It is a major health problem in endemic areas where cattle and sheep are raised and dogs have access to animal offal. Such regions include Mediterranean countries, Spain, Greece, Turkey, North Africa, the Middle East, but also Australia, New Zealand, South America, Baltic areas, the Philippines, and northern China.

Hydatid disease, however, is not confined only to endemic areas and physicians and surgeons worldwide may encounter the disease sporadically because of increased travel and migration.

Types of parasites

There are several species of the cestode *Echinococcus. E. granulosus* is the commonest form of the disease and has the widest geographical distribution. It exists in two varieties: the pastoral and sylvatic.

P>In the pastoral form, the dog is the definitive host, and sheep, cattle, pigs, and humans are intermediate hosts. It produces the typical features of hydatid disease, mainly with liver involvement.

The sylvatic form is less common and more benign and occurs in North America, particularly in Alaska and Canada. The dog is the definitive host, but intermediate hosts are moose, caribou, and reindeer rather than cattle and sheep. The cystic lesions are found more in the lungs than in the liver.

E. multilocularis (*E. alveolaris, E. sibiricensis*) is found primarily in Germany, Austria, Russia, and in the Arctic regions. The definitive host is the dog or fox and intermediate hosts are various rodents. Infection in humans is more serious than *E. granulosus*, as the parasite produces a multicystic tumour-like infiltrative growth in the liver and, if untreated, carries a mortality approaching 100 per cent.

E. oligarthrus and *E. vogeli* rarely affect humans. Their distribution is limited to countries in Central and South America, including Brazil, Argentina, Colombia, Panama, and Costa Rica.

Lifecycle of *E. granulosus*

The adult parasite is a 5-mm tapeworm, *E. granulosus*, which lives in the small intestine of dogs and other canines. The worm has a pyriform head (scolex) with four suckers and numerous hooklets, a short neck, and only a few segments (proglottids), of which the terminal one releases the ova, 300 to 500 in number. These ova are passed in the faeces and remain viable on the ground for many months. Contaminated grass is ingested by sheep and cattle, and humans may become infected by consuming low growing fruits and vegetables or by handling dogs carrying eggs in hair, saliva, and the anal region. In the jejunum of these intermediate hosts, the larvae (oncospheres) hatch from the eggs, facilitated by digestive enzymes in the duodenum, and migrate through the bowel wall. They end up in mesenteric lymphatics and in venules of the portal circulation, and are then carried to the liver where they are trapped in the sinusoids and develop into hydatid cysts. Some larvae are not filtered in the liver, but remain in the blood to reach the next station, the lungs. In addition, some may pass through the pulmonary circulation and travel to other organs. Larvae transported in the mesenteric lymphatics are carried to the cisterna chyli, the thoracic duct, and into the general circulation, ending up in a variety of distant sites.

When affected sheep and cattle are slaughtered, the viscera and meat may be thrown away and eaten by dogs. The cycle is then closed with the ingested scolices being transformed into adult tapeworms to live again in the canine's small intestine and produce ova (Fig. 1).



Fig. 1. Lifecycle and structure of *E. granulosus*.

Man is an abortive intermediate host since dogs normally do not have access to human viscera.

Methods to interrupt the parasite's lifecycle and prevent the spread of disease include appropriate disposal of infected animal viscera. Humans should be cautious with consumption of potentially contaminated fruits and vegetables in endemic areas. Dogs in affected areas must be regularly dewormed, and humans should wash their hands after contact with these dogs.

Pathology in the intermediate host

In infection by *E. granulosus*, 70 per cent of all hydatid cysts are found in the liver. The right lobe is most frequently involved, probably because most of the portal blood from the upper gastrointestinal tract reaches this area.

The hydatid cyst

This is a spherical-shaped collection of fluid housing innumerable scolices. They are usually single, but multiple cysts may occur. Slowly, over many years, the cysts increase in size. The surrounding parenchyma is displaced, but not invaded, by this expanding colony of scolices which may grow up to a diameter of 30 cm. Surrounding the cyst is an outer layer of adventitia (pericystic layer), which does not belong to the parasite but represents host parenchyma compressed into a fibrovascular layer. Apart from giving some mechanical support to the cyst, the pericystic layer facilitates exchange of nutrients and metabolic products between host and parasite

The true parasitic contribution to the hydatid cyst wall is the laminated membrane. This acelluar and chitinous membrane is glistening white, has the look and feel of a hard boiled egg, is not more than 2 mm thick, very fragile, and ruptures easily. It allows for osmotic transfer of nutrients, as it is in close contact with the inner surface of the pericyst. Any separation between pericyst and laminated membrane results in rupture and fragmentation of the latter.

On the inside of the cyst, the laminated membrane is lined by a very thin, almost transparent layer, the germinal membrane. This cellular layer is the breeding ground for the multiplying scolices. Pedunculated nodes of multiplying cells project into the lumen of the cyst as brood capsules. Scolices develop from these brood capsules and are liberated into the fluid once the connection with the germinal epithelium disrupts. The free floating and gravitating scolices within the cyst fluid are described as hydatid sand. Continuous restructuring of the germinal layer by means of protrusion, invagination, and fragmentation leads to endogenous daughter cysts, and the original univesicular cyst is eventually filled by hundreds of daughter cysts of various sizes.

The germinal membrane lays down on its outer surface the laminated membrane, while on the inside it secretes the hydatid fluid, which is a crystal clear transudate of serum that acts as the amniotic fluid for the countless scolices floating within it. It is isotonic, contains protein, and is antigenic. If released into the circulation, eosinophilia, urticaria, bronchospasm, or anaphylaxis may occur. A hydrostatic pressure of 300 to 900 mmH_2O can build up within the interior of the cysts, enough to cause a near explosive discharge of the hydatid fluid at operation if the cyst suddenly ruptures. Countless viable scolices are then set free over the surrounding field. Some of these may implant on nearby serosal surfaces and lead to recurrence of hydatid cysts.

Sometimes endogenous daughter cysts are formed which lack scolices and are therefore sterile. Separation of germinal from laminated membrane would also cause sterility due to nutritional impediment. After several years of slowly expanding growth, the hydatid cyst does not enlarge any further and the pericystic adventitial wall may calcify. This would also lead to nutritional deprivation and therefore intracystic death of the larvae.

As hydatid cysts grow by expansion, they tend to make their way towards regions of less resistance. In the liver they protrude from the depth of the organ towards the peritoneal surface, particularly beneath the diaphragm. Pressure necrosis of parenchyma and the wall of tubular structures may eventually result in free rupture into a coelomic cavity (peritoneal or pleural) or into bile ducts, bronchial tree, urogenital system, or vessels.

The growth rate of cysts depends on their location. A cyst the size of a tennis ball in the liver has been present for 5 to 10 years, whereas a similar size in the lung is reached in about a year.

Localization of hydatid cysts

Although the liver and lungs are most commonly affected, hydatid cysts may occur virtualy anywhere in the body, and may be found in the peritoneal cavity, heart, brain, skeleton, retroperitoneum, pelvic organs, muscles, and surgical incisions (Table 1). Many cysts are considered primary, like those in liver, lung, kidney, brain, and bone, as the larvae are directly derived from the bloodstream or lymphatics. Occurrence in spleen and pancreas may be the result of retrograde migration in the portal vein during bouts of increased abdominal pressure.

Location	*Frequency (%)*
Liver	55–75
Lung	18–35
Peritoneal cavity	10–16
Kidney	1–4
Spleen	2–3
Skeleton	1–3
Brain	1–2
Heart	0.02–2
Uterus, adnexa	0.5–1.5
Retroperitoneum	0.5–1.5
Pancreas	0.3–0.8
Other (muscle, joints, breast, eye, thyroid, bladder)	0.1–1

Table 1 Common and uncommon sites of infection with *E. granulosus*

The vast majority of abdominal and pelvic cysts are secondary to prior hepatic localization and follow spontaneous rupture or surgical inoculation. Cysts of the myocardium are considered primary, but those of the pericardium secondary. A complication of cardiac cysts is peripheral embolization, with resulting ischaemia and hydatid inoculation.

Clinical presentation

Many carriers of hydatid cysts are entirely asymptomatic and the cysts are incidentally detected during radiological procedures for other reasons. They may be seen as space-occupying lesions in the liver or lung, sometimes with a calcified rim.

Depending on the location of the hydatid cysts (Table 2), patients with symptomatic disease would present with vague abdominal pain, hepatomegaly or an abdominal mass, jaundice, fever, urticaria, anorexia and malaise, coughing, hydatidemesis, headache, neurological, urological, or musculoskeletal symptoms.

Rupture:	Intraperitoneal Into biliary Into pleural space Into bronchus Into other adjacent organ
Pressure effects:	Vertebrae: spinal cord Brain: intracranial pressure Liver: obstructive jaundice, Budd–Chiari syndrome Lung: atelectasis Kidney: obstructive uropathy
Tissue destruction with loss of function:	Bone: fracture Liver: cholangitis, biliary cirrhosis Other: lung, kidney, heart
Secondary infection, spread, recurrence:	Heart: embolization Spontaneous rupture of cyst Iatrogenic: puncture, operation
Allergic manifestations:	Urticaria, bronchospasm, anaphylaxis Eosinophilia

Table 2 Complications of hydatid cysts.

Diagnosis

A history of residence or visits to endemic areas is an important part in the diagnostic work-up, and should raise the suspicion of hydatid disease.

Physical findings are often unimpressive. Hydatid cysts in the liver usually grow in the dome of the right lobe in segments VII and VIII and tend to push the liver downwards. The most common physical finding therefore is diffuse hepatomegaly rather than a mass in the liver. Pulmonary signs depend on pressure exerted by the

cyst on the bronchi, rupture of the cyst into the bronchial tree, and presence of consolidation in surrounding pulmonary tissue. Organ-specific signs arise from a less common involvement of other structures such as brain, kidneys, internal genitalia, skeleton, spleen, heart, thyroid gland, retroperitoneum, and parietes. In fact, hydatid disease may mimic many neoplastic and inflammatory conditions throughout the body and should therefore be included in the differential diagnosis.

Laboratory tests

Liver function tests may be abnormal and biliary obstruction demonstrated by increased levels of alkaline phosphatase and direct bilirubin. Eosinophilia is present in only 25 to 35 per cent of patients and is probably the result of laminated membrane disruption and entrance of antigenic material into the circulation. It is infrequent in patients who have a single, intact, univesicular hydatid cyst. Leucocytosis and microcytic anaemia indicate infection and blood loss.

Skin test

The Casoni skin test was used extensively some years age. For this test, 0.2 ml of sterile hydatid fluid is injected intradermally and the resulting weal compared with a control saline solution after 15 min. The test remains positive for many years after treatment, carries the risk of anaphylaxis, and has a large number of false-positive and false-negative results. It also causes serological tests to become falsely positive. For all these reasons, the Casoni skin test is now better abandoned.

Serological tests

Serological tests are useful not only for diagnosis and follow-up of hydatid disease but may even be used for screening in high risk populations in endemic areas. There are now a large number of these tests available. The complement fixation test is useful, because it becomes negative after cure of a hydatid cyst and can therefore be used to assess the efficacy of treatment. It is not very sensitive, however, and there are many false-negative results. Enzyme-linked immunosorbent assay (**ELISA**), indirect haemagglutination, the radioallergosorbent test, and immunoelectrophoresis are also available. Indirect haemagglutination and ELISA are the most sensitive immunological methods to diagnose human hydatidosis, but a false-positive reaction secondary to cross-reactivity with other parasitic infections (cysticercosis, schistosomiasis) is possible. Immunoelectrophoresis to detect specific antibodies against antigens isolated from hydatid cyst fluid, double diffusion arc-5, is specific. A recent advance has been the introduction of enzyme-linked immunoelectrotransfer blot assays which are more sensitive and specific than other serological tests. The combination of specific ELISA and western blot serology is now recommended.

Radiology and ultrasound

Plain abdominal and chest radiographs may show features of hydatid disease (Fig. 2): a calcified cyst with or without gas–fluid levels; consolidation of lung tissue; a raised, poorly moving right diaphragm; an obscured costophrenic angle in the case of transdiaphragmatic migration of the expanding cyst; hepatomegaly; displacement of adjacent viscera, such as stomach or colon; and bone abnormalities.



Fig. 2. A large multivesicular hydatid cyst in the left lung.

An intravenous urogram may demonstrate lesions in the urinary tract, and selective coeliac angiography may demonstrate stretching and elongation of hepatic vessels as well as avascular areas in the hepatogram.

Ultrasonography is a non-invasive test which can be easily repeated. Pressure effects on nearby ducts and vessels may be readily demonstrated and cyst morphology clearly displayed. It can define a cyst as univesicular or multivesicular and show fragmentation of laminar membrane, cyst wall thickness, and calcification (Fig. 3).

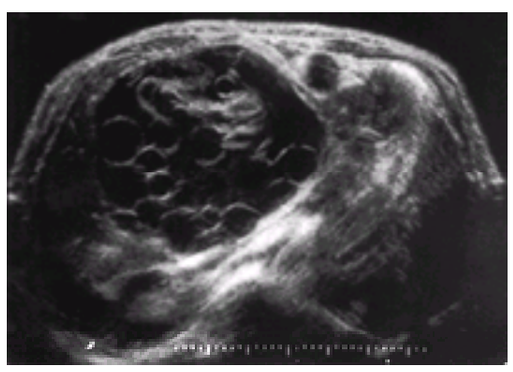


Fig. 3. Ultrasonic image of *E. granulosus* in the liver, showing many daughter cysts within the main cyst.

CT scanning is also a very effective imaging technique. It demonstrates the characteristics of the hydatid cyst and reveals all intrahepatic and extrahepatic lesions (Fig. 4).

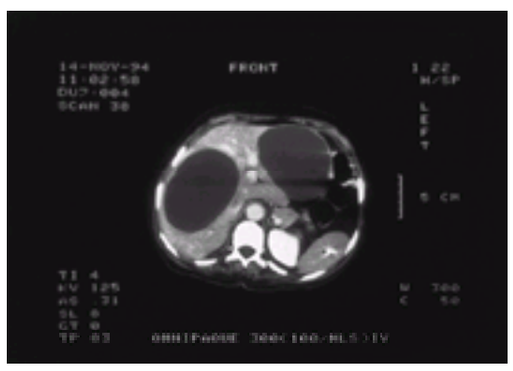


Fig. 4. CT scan showing two hydatid cysts in the liver.

Magnetic resonance imaging provides detailed pictures of the hydatid process, especially daughter cysts, detachment of the membranes, and intrahepatic and extrahepatic rupture. It gives high quality information on intracranial, spinal, and cardiac hydatid processes.

Additional investigations include endoscopic retrograde cholangiopancreatography which delineates communicatons between the cyst and bile ducts and presence of hydatid elements in the biliary tree. A sphincterotomy may be added in case of obstruction to bile flow in the common bile duct.

Retrograde urinary cystography, bronchography, or gastrointestinal contrast studies may be indicated in certain circumstances.

Surgical treatment

Hepatic cysts

Surgical treatment remains the mainstay in the management of hydatid disease and should be individualized according to the number and location of cysts, the presence of complications such as cyst infection and rupture (e.g. into biliary or bronchial tree), and the patient's overall physiological condition.

Spillage of the cyst contents into the peritoneal or pleural cavity should be avoided because of the almost inevitable implantation of viable scolices and development of new cysts on nearby structures. Attempts should be made to aspirate the fluid first and neutralize the cyst contents with scolicidal agents. The laminated and germinal membranes are removed, together with all the brood capsules, daughter cysts, and hydatid scoliceal sand.

As most liver cysts occur in the right lobe, a long curved transverse abdominal incision is made in the right upper quadrant. If needed, this incision may be extended into the chest if the hydatid occupies the dome or posterior aspect of the liver or has transgressed the diaphragm into the pleural space. A search for previously undetected cysts should be undertaken and the retroperitoneum examined. Cysts may be found in the omentum, around the pancreas, in the spleen, kidneys, perivesical, or periuterine areas. Intraoperative ultrasonography may be helpful in determining the extent of disease and in detecting exogenous daughter cysts that might otherwise be overlooked.

Good exposure is important and the lesion should be approached through the most direct route in order to reach its most superficial part first. This area is often visible on the surface of the involved organ as a circular whitish area. The surgical field should be well protected and walled off by appropriately placed gauze pads, soaked in a 3 per cent saline solution to avoid contamination of neighbouring areas in case of accidental spillage.

Trying to enucleate a univesicular hydatid cysts intact without opening it usually fails, as the fragile laminated membrane easily breaks up on separation from the pericystic adventitia. It is therefore generally better first to evacuate the cyst contents by a trocar or large-bore needle.

Special cone-shaped devices are available to prevent spillage. Saidi has devised a funnel to be placed on the dome of the cyst and fused to it cryogenically with carbon dioxide to produce a non-leaking connection. After evacuation, the funnel is removed by thawing and excision of the rim of necrotic tissue around the edge. Another suction cone, designed by Aarons, is placed on top of the cyst and adheres to its surface when suction is applied to a grooved inferior rim via a side arm (Fig. 5). Through the centre of the cone the cyst can now be opened and the contents evacuated.



Fig. 5. The Aarons cone.

Once the cyst has been emptied, scolicidal agents are introduced. Formalin was widely used in the past, but severe complications have occurred such as toxaemia, acute pancreatitis, and sclerosing cholangitis in cysts with biliary communications. Many other products may cause side-effects as well, and few are therefore really harmless. The list includes hypertonic saline (3, 5, 10, 20, or 30 per cent), silver nitrate (0.5 per cent), hydrogen peroxide (10 per cent), povidone iodine (10 per cent), diluted Savlon (0.5 per cent cetrimide with 0.05 per cent chlorhexidine), praziquantel, absolute alcohol (96 per cent), and others. Whether any of these is really 100 per cent scolicidal remains unclear. Our personal preference is a 10 per cent hypertonic saline solution. After some 5 min contact time, the scolicidal fluid is aspirated and the cyst cavity meticulously emptied by suction and instrumentation, taking care of all hydatid material including the laminated membrane (Fig. 6).

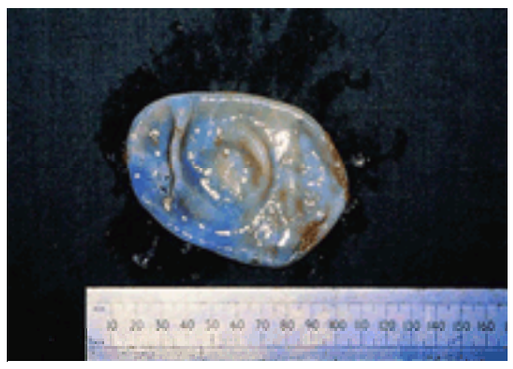


Fig. 6. Appearance of hydatid cyst after removal. The glistening laminated membrane is covered on the inside by germinal epithelium from where the brood capsules and scolices derive.

The pericystic adventitia is a host-derived fibrovascular structure. This cannot be separated from the surrounding parenchyma and should therefore be left *in situ.* Its dissection causes excessive haemorrhage.

The majority of hydatid cysts are multivesicular, rather than univesicular, and the cyst cavity is filled with debris, laminar fragments, and thousands of small and medium-sized daughter cysts. Upon insertion of a needle or trocar very little fluid will come out. The contents of such a cyst are difficult to remove, and very large-bore suction devices are required. Upon evacuation, bile ducts may be seen opening into the cavity. One of these biliary communications may well have been responsible for rupture of the laminated membrane in the first place, months or years previously. The cyst cavity should be cleaned, swabbed dry, and inspected for bile openings. If found, these openings are then closed by transfixion with fine absorbable sutures.

If bile leaks are present, it is advisable to perform cholangiography and explore the common bile duct when there is evidence of hydatid material within. Patients with obstructive jaundice should also undergo this procedure and the bile ducts meticulously cleansed. In addition, the gallbladder should be removed as it may harbour scolices as well.

The remaining cyst cavity can be managed in a number of ways. Small to medium-sized, sterile, univesicular cysts may be emptied and closed primarily. Alternatively, the space can be reduced by infolding (capitonnage) and left open to the peritoneal cavity or closed around a drainage tube which is brought to the outside.

Another good method to obliterate the cavity is to mobilize and introduce a well vascularized pedicle of omentum (omentoplasty). This provides excellent absorptive drainage of any secreted fluid.

External drainage is advisable for complicated cysts with infected contents or biliary communications. The drain may have to remain in place for several weeks, until bile drainage stops and sinography confirms shrinkage of the cavity. Marsupialization to the abdominal wall is not recommended because of poor results. Internal drainage with a Roux-en-Y jejunal loop may be used for deep-seated hepatic cysts with biliary connections.

Hydatid disease is expansive but not invasive and preservation of the surrounding normal parenchyma should be attempted. This can be achieved by removal of the parasite's contributions to the cyst. Incomplete removal, however, will result in recurrence and to prevent this, more radical procedures should be considered. These include total pericystectomy; and segmental or lobar liver resection (Fig. 7), which can be performed safely and expeditiously by experienced surgeons, with optimal vascular and biliary control. In such resections the hydatid cysts are not opened and contamination is therefore less of a problem. However, the feasibility of more radical procedures depends to some extent on the anatomical location of the disease in relation to the hepatic hilum, hepatic veins, and inferior vena cava.

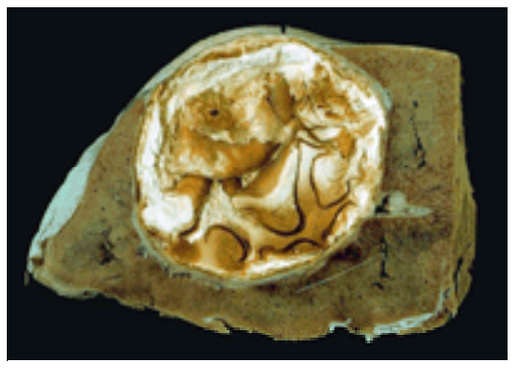


Fig. 7. Pathology specimen of a large multivesicular cyst involving the right lobe of the liver.

Some patients have extensive liver involvement and require multiple operations. Cholangitis, secondary biliary cirrhosis, Budd–Chiari syndrome, and liver failure may ensue and they become candidates for liver transplantation. The transplant operation is more difficult because of the previous surgical interventions.

Resection or liver transplantation is also the best choice for the treatment of *E. multilocularis* because of the lack of demarcation in the liver and its aggressive infiltrative growth.

With the advent of minimal access surgery, less formidable surgical procedures have gained popularity recently. Laparoscopic evacuation of the hydatid cysts can be achieved by a special aspiration–grinder apparatus. This device has an external diameter of 10 mm and is introduced laparoscopically into the lesion. It aspirates the cyst elements and grinds them by means of a propeller driven with a spiral motor. Scolicidal lavage is then performed, followed by wide cystotomy, exploration of the cavity, and extraction of additional parasitic material. Cystography may be added and biliary communications closed with sutures. At the end of the procedure a drain is placed and brought out to the exterior through the nearest port hole. Significant postoperative biliary drainage can be successfully reduced by endoscopic papillotomy. Patients with early-stage cysts are the best candidates for laparoscopic treatment because the cystic cavity is able to collapse totally after laparoscopic evacuation and the possibility of biliary communications is low.

Percutaneous drainage of hepatic hydatid cysts has been proposed as an alternative to surgery. However, this method has several drawbacks. Most cysts are multivesicular and scolicidal agents cannot enter the daughter vesicles. Mechanical removal of all viable or dead cyst elements is mandatory for cure and this is difficult to achieve with a percutaneous catheter. Consequently, foreign material is left in the body and may lead to secondary infection and recurrence. Despite this, some authors have reported favourable results when percutaneous drainage was combined with albendazole therapy. The best results will be obtained in univesicular cysts.

Asymptomatic calcified cysts in liver or lungs may be best left alone when they are discovered incidentally. Radiographic follow-up may be all that is required, especially in patients who are less fit for surgery.

Extrahepatic cysts

Treatment of lung hydatids follows the same principles as for the liver. Adequate exposure is obtained via an anterolateral or posterolateral thoracotomy, and the surgical field is protected against spillage. The pericyst cavity is opened, the contents extracted, and small bronchial openings closed with sutures. The edges of the final opening in the pericyst cavity are overrun for haemostasis and to control air leaks. Some cysts may already have ruptured into a bronchus and the patient presents with haemoptysis and coughs up a salty-tasting fluid with membranous detritus in it. Most of these ruptured cysts are infected and subsequently may turn into a lung abscess. Surgical evacuation and debridement of this infected cavity may still control the situation, but if suture closure of the disrupted bronchus is not satisfactory and much lung tissue is already destroyed, a pulmonary lobectomy becomes unavoidable.

Renal hydatid cysts may present as a palpable mass, with flank pain, haematuria, hydatiduria, or albuminuria. The lesions are solitary in 85 per cent of cases, but may be multifocal (15 per cent) or occur bilaterally (6 per cent). Surgical options are removal of the parasitic components of the cyst, pericystectomy, partial nephrectomy, or total nephrectomy (Fig. 8).

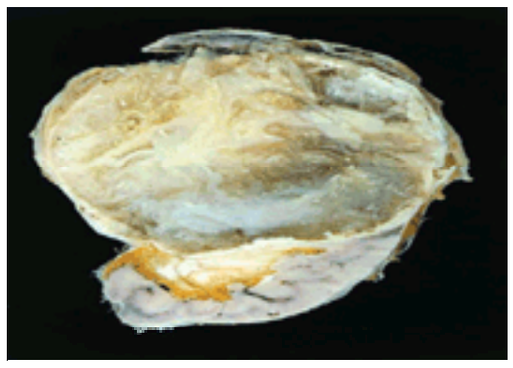


Fig. 8. Pathology specimen of a renal hydatid cyst occupying most of the left kidney.

Splenic cysts are treated by splenectomy and pancreatic cysts in the body or tail by distal pancreatectomy.

Cysts in the brain and heart are difficult to access. They may be managed by enucleation or decompression, in addition to anthelmintic chemotherapy.

Hydatid infection of the skeleton does not produce the typical cystic lesions, as bone tissue is not easily yielding. There is instead a slow erosion of cancellous and cortical bone. The architecture of the bone is destroyed, weakened, and pathological fractures occur. Vertebral involvement may lead to spinal cord compression. Secondary bacterial infection often complicates matters. The best approach is to curette the entire length of the involved bone and irrigate the area with an effective

and non-toxic scolicidal agent. In addition, prolonged anthelmintic treatment is required. For the worst cases of osseous hydatidosis, amputation may be inevitable.

Overall, surgical treament of hydatid disease carries a mortality of 0.3 to 3.8 per cent. The recurrence rate is 5 to 20 per cent.

Drug treatment

Anthelmintic antibiotic therapy is considered adjunctive to surgical intervention in most instances. Administration should begin before operation and continued afterwards, usually for some weeks. It is recommended for complicated hydatid conditions, especially disease involving brain, heart, and bone and in patients unfit for operation. It should also be used in case of recurrent disease.

Three agents are currently available: the benzimidazoles, mebendazole and albendazole, and the isoquinoline derivative, praziquantel.

Mebendazole diffuses through the cyst membrane and interferes with cellular microtubular formation, thus disturbing glucose uptake and energy production in the parasite. Clinical improvement and reduction in cyst size have followed administration of 400 to 600 mg daily for 3 to 4 weeks. However, viable cysts can persist even after 12 months of therapy. Recurrence is not unusual and the concentration of drug achieved in large cysts is uncertain. Side-effects include nausea and vomiting, diarrhoea, urticaria, liver function disturbance, hepatitis, neutropenia, and central nervous system manifestations (headache, vertigo, convulsions). These side-effects, however, are usually mild and transient.

Albendazole is better absorbed from the gut and is probably more effective than mebendazole, but the toxicity pattern is comparable. Various dosage schedules are used. It may be given in 4-week cycles, separated by 2-week drug-free intervals, in a dose of 10 to 15 mg/kg per day. While liver functions and blood counts are monitored, the drug is continued for several cycles until sterilization or regression of the hydatid process is documented by ultrasonography or radiology.

Praziquantel alters the permeability of the plasma membrane to calcium ions and causes derangement of glucose metabolism in the parasite. It kills the protoscolices in the cyst fluid, but does not affect the germinal epithelium from which they arise. Side-effects are few and mild. It may be used intraoperatively to sterilize the cyst cavity. Praziquantel is normally given as an oral dose of 60 mg/kg for 1 day in the treatment of other parasites (schistosomiasis). For *E. granulosus* it may be taken in a lower dose for up to 2 weeks in combination with albendazole.

Further reading

Aarons BG, Kune GA. A suction cone to prevent spillage during hydatid surgery. *Australian and New Zealand Medical Journal* 1983; **53**: 471–2. [Description of a novel useful device.]

Abu-Eshy SA. Some rare presentations of hydatid cyst (*Echinococcus granulosus*). *Journal of the Royal College of Surgeons of Edinburgh* 1998; **43**: 347–52. [A report on some unusual sites of hydatidosis in Saudi Arabian patients.]

Alfieri S *et al.* Radical surgery for liver hydatid disease: a study of 89 consecutive patients. *Hepatogastroenterology* 1997; **44**: 496–500. [A more aggressive approach: total transparenchymal excision. Good results in experienced hands.]

Alper A *et al.* Laparoscopic surgery of hepatic hydatid disease: initial results and early follow up of 16 patients. *World Journal of Surgery* 1995; **19**: 725–8. [Efficacy of minimal access surgery in a limited group of patients.]

Angulo JC, Sanchez-Chapado M, Diego A, Escribano J, Tamayo JC, Martin L. Renal echinococcosis: clinical study of 34 cases. *Journal of Urology* 1997; **157**: 787–94.

Angulo JC, Escribano J, Diego A, Sanchez-Chapado M. Isolated retrovesical and extrarenal retroperitoneal hydatidosis: clinical study of 10 cases and literature review. *Journal of Urology* 1998; **159**: 76–82. [Myriad signs and symptoms in hydatid disease involving the urinary tract.]

Khuroo MS *et al.* Percutaneous drainage compared with surgery for hepatic hydatid cysts. *New England Journal of Medicine* 1997; **337**: 881–7. [A conservative approach to hydatid cysts: aspiration plus chemotherapy.]

Little JM. Hydatid disease. In: Morris PJ, Malt RA, eds, *Oxford textbook of surgery*, 1st edn, pp. 2507–11. Oxford University Press, New York, 1994. [An excellent overview for the surgical community.]

Little JM, Deane SA. Hydatid disease. In: Blumgart LH, ed., *Surgery of the liver and biliary tract*, pp. 955–66. Churchill Livingstone, Edinburgh, 1988. [Surgical aspects of hepatobiliary hydatidosis.]

Moreno-Gonzalez E *et al.* Liver transplantation for *Echinococcus granulosus* hydatid disease. *Transplantation* 1994; **58**: 797–800. [Endstage liver disease resulting from echinococcal infection should be considered for hepatic transplantation.]

Prousalidis J, Tzardinoglou K, Sgouradis L, Katsohis C, Aletras H. Uncommon sites of hydatid disease. *World Journal of Surgery* 1998; **22**: 17–22. [A review of 49 patients with extra-hepatic extrapulmonary hydatidosis.]

Safioleas M, Misiakos E, Manti C, Katsikas D, Skalkeas G. Diagnostic evaluation and surgical management of hydatid disease of the liver. *World Journal of Surgery* 1994; **18**: 859–65. [A classical, innovative technique in prevention of recurrence after surgery.[Reporting a large experience with 132 patients in a Greek surgical centre.]

Saidi F, Nazariah I. Surgical treatment of hydatid cysts by freezing of cyst wall and instillation of 0.5 per cent silver nitrate solution. *New England Journal of Medicine* 1971; **284**: 1346–50. [A classical, innovative technique in prevention of recurrence after surgery.]

Saidi F. Hydatid disease. In: Ahmed M *et al.*, eds, *Surgery for all: a view from the developing world*, pp. 121–32. Ferozsons, Lahore, 1992. [A personal and sensible view from an experienced surgeon.]

Turkcapar AG, Ersoz S, Gungor C, Aydinuraz K, Yerdel MA, Aras N. Surgical treatment of hepatic hydatidosis combined with perioperative treatment with albendazole. *European Journal of Surgery* 1997; **163**: 923–8. [Adjuvant medication improves the outcome of surgery.]

Ulualp KM *et al.* Management of intrabiliary rupture of hydatid cyst of the liver. *World Journal of Surgery* 1995; **19**: 720–4. [Exploration and T-tube drainage of the biliary tree in addition to cystectomy: experience with 36 patients.]

47.2.5 Ascariasis and other intestinal nematode infections

Oluwatope A. Mabogunje and A. S. Daar

Ascaris lumbricoides (roundworm)
Lifecycle and morphology
Host–parasite interaction
Specific immune response
Abdominal complications of ascariasis
Other intestinal nematodes
Ancylostoma duodenale and *Necator americanus* (hookworms)
Strongyloides stercoralis (threadworm)
Enterobius vermicularis (pinworm)
Trichuris trichuria (whipworm)
Anisakis simplex
Capillaria philippinensis
Capillaria hepatica
Laboratory diagnosis of infection by intestinal nematodes
Antihelminthic drugs
Further reading

Intestinal nematodes have a worldwide distribution but are endemic in tropical and subtropical regions (Table 1, Table 2). Of the six nematodes commonly infecting humans, *Ascaris lumbricoides* is a cause of significant surgical morbidity and mortality. In some endemic regions, ascariasis is the most common cause of intestinal obstruction in children and of biliary disease in both children and adults.

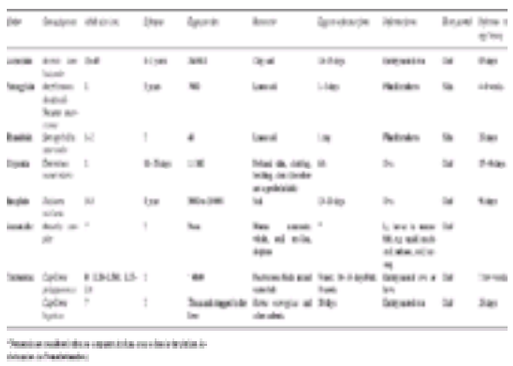


Table 1 Parasitological data*

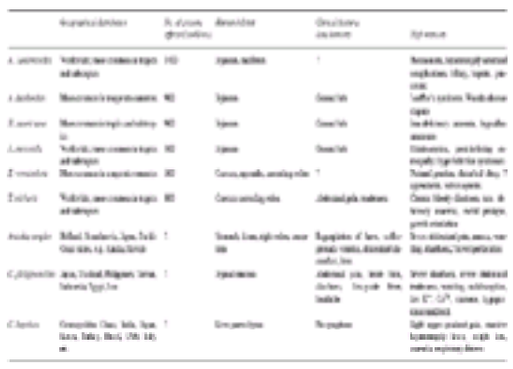


Table 2 Epidemiology and clinical features of intestinal nematodes

Immigration of persons from southern Asia, the Caribbean islands, and Central America contributes to the prevalence of nematode infections in Europe and North America. At the same time case reports of incidental anisakiasis are accumulating in Japan as well as the United States of America and Europe, due to increasing consumption of raw marine fish such as mackerel, herring, and rockfish in the form of sushi, sashimi, and ceviche. Incidental infections by *Capillaria philippinensis* have been reported from Asia and the Middle East in persons who eat uncooked fish from inland waters.

Ascaris lumbricoides (roundworm)

It is estimated that 1.4 billion persons, or a quarter of the human race, are infected by this nematode helminth. Every year up to 2 million cases of clinical disease are recorded, and about 20 000 infected children die from intestinal obstruction and other abdominal complications of ascariasis.

Lifecycle and morphology

Human beings become infected by ingesting food or drink contaminated by soil containing the embryonated ova of *ascarides*. The egg membranes are digested in the stomach and the invasive larvae are released in the duodenum and pass down to the caecum. They penetrate the mucosa and enter the mesenteric venous radicles and the lymphatics. Most larvae pass through the liver sinusoids to the hepatic veins into the inferior vena cava and the right heart. Others pass through the thoracic duct into the subclavian vein and the superior vena cava into the right heart and the pulmonary circulation. There they penetrate the walls of the terminal pulmonary arterioles and alveoli into the bronchoalveolar tree and upwards along the trachea and the pharynx. They are swallowed again down the oesophagus.

During the migration, which lasts 14 days, the larva grows to about 2 mm in length and moults twice. Further growth and maturation in the jejunum results in the production of adult male and female worms. Some 60 to 75 days after ingestion of infective eggs, the new gravid female worm begins to lay its own eggs. The mean lifespan of an adult worm is 1 to 2 years, after which it is spontaneously expelled. In endemic areas, however, there is repeated reinfection, resulting in heavy worm loads, especially in children.

The adult worms live and mate in the human jejunum and mid-ileum. The female is 20 to 49 cm long and 3 to 6 mm in diameter; the male is 15 to 30 cm long and 2 to 4 mm in diameter. Each worm is pinkish-white, rounded and tense, with a striated cuticular surface and two longitudinal streaks. The blunt anterior end has three lips. The vulva opens ventrally in the female and there are two copulatory spicules at the ventrally curved posterior end of the male.

Ascaris feeds on the host's intestinal contents and is equipped with its own battery of digestive enzymes, including amylase, protease, and lipase. The adult worms are facultative anaerobes with a glycogen consumption of 1.3 g/100 g body wt per day. The larval stages have anaerobic metabolism, while the infective eggs in the soil are dormant. Amino acids required for egg production are partly obtained from the gut and partly SYnthesized by the ovaries. Each female lays up to 240 000 eggs per day. These are passed with stool into the soil, where embryonation occurs. The eggs may lie dormant in moist soil for up to 10 years. There are no intermediate hosts.

Host–parasite interaction

The adult *ascaris* has to survive in the host's intestine against many odds. Peristalsis maintains aboral flow of luminal contents, which would expel the worm. Gastric acid, bile salts, and pancreatic enzymes create a hydrolytic milieu, which would digest the worm. Full activation of the mucosal immune system would destroy it. The worm's survival mechanisms have largely been deduced from experimental work with *A. suum* infection of pigs.

The adult worm is not attached to the intestinal mucosa but is braced most of the time by its own muscle power against the intestinal wall and peristaltic activity. Its spindle shape and spiral movements facilitate forward motion and entry into small orifices such as the papilla of Vater. The cuticle, which covers the entire body and its invaginations, is a secreted, multilayered structure of collagens, hyaluronic acid, chondroitin sulphate, and mucopolysaccharides. It is both chemically and physically resistant to digestion. The worm elaborates families of isoinhibitor proteins, which inactivate the host's pancreatic proteases, chymotrypsin, elastase, trypsin, and carboxypeptidases A and B. Many of these host enzymes are simply ingested and complexed with the inhibitors within the worm's enteric tract. Inhibitors of pepsin have also been reported, one of which is an aspartyl protease inhibitor that is now being explored as a vaccine target and a possible avenue for parasite control.

The collagens in the cuticle have many of the characteristics of vertebrate type IV collagens and may be involved in the parasite's molecular mimicry of the host as a mechanism of immune evasion. It is also possible that the mucopolysaccharides render the mucosal immune system unresponsive to *Ascaris* protein antigens in a non-specific fashion, thus engendering oral tolerance for the worm.

Specific immune response

Although the adult *ascaris* worm residing in the intestinal lumen is sheltered from cell-mediated immune effector mechanisms, the migrating larval stages are subject to attack by components of the mucosal immune system; for example, while penetrating the intestinal mucosa some of the infective (L_3) larvae are bound by mucosal IgA and entrapped in the mucous layer, where they are subsequently digested by intestinal proteinases. However, the predominant immunological response to *ascarides* involves eosinophils and IgE antibodies; the serum IgG and IgD are increased to a lesser extent; IgA and IgM are not increased. Ascaris preferentially stimulates the T_H2 subset of CD4+ helper T cells, which secrete interleukin 4 and 5 among other cytokines. Interleukin 4 induces high concentrations of IgE in the blood. Interleukin 5 acts as an eosinophil-activating factor, which augments the production and potentiation of eosinophils for cytotoxicity. The IgE antibodies opsonize the larva. Eosinophils attach to the opsonized larva by specific Fc receptors and degranulate; this releases major basic protein, which lyses the larva by IgE antibody-dependent eosinophil cell-mediated cytotoxicity. Similarly, in the liver sinusoids, other larvae are killed by activated eosinophils, infiltrating neutrophils, and histiocytes, which form granulomatous pseudotubercles (the Splendore–Hoeppli phenomenon).

In the lungs the specific immune response contributes to tissue injury. Accumulation of T_H2 cells in the bronchial mucosa causes an immediate-type hypersensitivity reaction, which results in a peribronchial eosinophilic infiltrate, bronchospasm with increased secretion of mucus, and a serous alveolar exudate. This manifests clinically as Loeffler's syndrome, during which a chest radiograph will show diffuse mottling. The degree of this reaction is related to the number of larvae migrating simultaneously. The frequency of this hypersensitivity reaction is much less in regions where there is continuous infection than where the infection is seasonal.

Although IgE antibodies to *Ascaris* antigens are produced by infected persons, there are relatively few reports of severe anaphylactic reactions in such individuals. In contrast, laboratory personnel become extremely sensitive to volatile allergens of *ascarides* and may respond with bronchial asthma, angioneurotic oedema, urticaria, and sometimes gastrointestinal symptoms.

There is some evidence in man for partial acquired immunity against helminths. In endemic areas, the prevalence of infection declines with age while the presence of antibodies in the serum increases. In a recent study from a village in Papua New Guinea, 95 per cent of children were infected compared with only 21 per cent of adults. Antibody concentrations in the adults were much higher than in the adults from another village with a lower intensity of childhood infection. This suggests that in adults with low egg counts and high antibody titres, although continuous reinfection may occur, immunity prevents the worms from becoming established. Heavy and continuous exposure engenders resistance against superinfection rather than rejection of all worms.

Recent biochemical studies have shown that the parasite can also protect itself from attack by toxic molecules such as oxygen free radicals, nitric oxide, and carbonyls, released by activated inflammatory cells, and by producing a battery of its own defensive enzymes (catalase, superoxide dismutase, glutathione system) that neutralize the toxic molecules. Current research is aimed at developing agents that would inactivate the parasite's defensive enzymes.

Abdominal complications of ascariasis

Intestinal ascariasis

Most infections with *ascarides* are unnoticed by the host until an adult worm is passed in the faeces or through the mouth. However, a patient infected with many worms may complain of vague central abdominal pains, anorexia, intermittent loose stools, and occasional vomiting. The pain in adults may be epigastric, resembling chronic peptic ulcer.

Intestinal obstruction

Partial obstruction of the intestines by intertwining of the worms may cause intestinal colic, nausea and vomiting, and fever; a palpable, mobile abdominal mass may be present. Complete intestinal obstruction usually occurs when a tangled mass of worms is impacted in the distal ileum. This is on occasion precipitated by the use of antihelminthics in a patient with partial obstruction. The patient develops severe, colicky abdominal pain with vomiting and obstipation. The abdomen is distended, with hyperactive bowel sounds, often with visible peristalsis and tenderness. The affected segment may act as a fixed point on which the rest of the bowel twists, causing a volvulus; the closed loop loaded with the worms is thus prone to strangulation and rupture unless prompt laparotomy is performed. Alternatively, intense spasm of the bowel around an obstructing ball of worms may be advanced as the leading point of an ileocaecal intussusception, which may also progress to bowel ischaemia if not relieved. Sometimes the worm obstructs the appendix causing acute appendicitis; the worm may also perforate through the appendix into the peritoneal cavity.

Diagnosis

In an endemic area, ascariasis is always considered in the differential diagnosis of intestinal obstruction. The suspicion is strong in patients with a recent history of passing a worm or of using a purgative or vermifuge before developing the features of intestinal obstruction. Such patients are admitted to hospital for evaluation.

A plain abdominal radiograph may show the worms along with air–fluid levels and dilated bowel loops. Abdominal ultrasonography may also show a mixed echogenic mass, provided the distended loops are not in the way. In doubtful cases a barium meal with follow through will outline the worms, which may also ingest some of the barium.

Management

Oral feeds are stopped, and intravenous fluids and electrolytes begun. A nasogastric tube is inserted and an antispasmodic such as meperidine or hyoscine butylbromide (Buscopan) is given intravenously in order to relax the bowel and allow the worms to disentangle and move into the colon, from where they are likely to be expelled. An antihelminthic such as piperazine or pyrantel is given (via the tube or by mouth with the tube clamped) only after the attack has subsided. This is done so as not to paralyse the worms in the ileum, where the paralysed tangled mass could cause complete obstruction. A large majority of affected patients respond promptly to this regimen by passing numerous worms rectally. Digital disimpaction may facilitate worm evacuation. In some stable and well-hydrated children, diatrizoate meglumine (Gastrografin), which causes osmotic diarrhoea, has been given through the nasogastric tube and has resulted in worm expulsion. Patients are discharged when normal bowel function has returned.

This regimen is, however, likely to fail when bowel obstruction is complicated by volvulus or intussusception. Prompt laparotomy is indicated after appropriate resuscitation in such patients and in those with advanced disease suggestive of strangulated or gangrenous intestinal obstruction. At laparotomy, if the bowel is not gangrenous, and whenever feasible, the worms are manually disentangled and massaged into the colon from where expulsion follows the postoperative administration of an antihelminthic. Otherwise, a judicious enterotomy is made and the worms extracted with minimal spillage of gut contents (Fig. 1). We do not recommend the practice of resecting viable bowel with the contained worms. However, when the bowel segment is not viable, resection is performed, with end-to-end anastomosis. Whenever the bowel is opened or resected, it is important to remove all the worms: a remaining worm may exit through the suture line postoperatively.



Fig. 1. Ascaris worms removed by enterotomy in a 5-year-old Nigerian girl who presented with acute intestinal obstruction unresponsive to conservative measures.

In endemic regions, ascariasis is the cause of up to 25 per cent of intestinal obstructions. It is estimated that in the tropics, 11 per cent of paediatric laparotomies are indicated for complicated intestinal ascariasis. In a children's hospital in South Africa, 12.8 per cent of all acute abdominal emergencies were caused by ascariasis; most of these were managed non-operatively. In another South African series, 22 of 29 children with volvulus required bowel resection. Five (17 per cent) of the 29 died from the complications of gangrenous bowel, endotoxaemia, or anastomotic failure.

In a report from Colombia, ascariasis accounted for 14.4 per cent of 500 children admitted with an acute abdomen to a paediatric surgical unit. Among 107 children with intestinal obstruction *from ascarides*, simple obstruction was present in 34, volvulus with or without gangrene in 63, and intussusception in 7. Because of late presentation, bowel resection was required in 68 children, of whom eight (12 per cent) died. Of the 51 without bowel resection only two (4 per cent) died. Ten other children presented with appendiceal perforation; of these, one died following appendectomy.

Biliary and hepatic ascariasis

Migration of an adult *ascarides* through the sphincter of Oddi into the common bile duct leads to biliary ascariasis. Some drugs, anaesthetic agents, antihelminthics, fever, or spicy foods may trigger migration. Entry into the bile duct is facilitated in patients with a previous sphincterotomy. The worm may spontaneously return to the duodenum, causing no symptoms at all. However, a female worm may be stranded in the bile duct or gallbladder, where it dies after laying some eggs. Such eggs and degenerating cuticular fragments may cause cholangitis, with or without secondary bacterial infection, or become the nidus for gallstones. Several worms invading the bile duct are more likely to block the cystic duct and cause biliary colic or acalculous cholecystitis. Further ascent into the hepatic ducts is associated with acute cholangitis, and an abscess may form within the intrahepatic ducts and liver parenchyma.

In endemic regions, biliary ascariasis may be the most common cause of biliary disease in both children and adults. The frequency and clinical pattern varies in different parts of the tropics; child cases are more frequently described in Latin America but more adult cases have been reported from India. The symptoms of hepatobiliary ascariasis are variable. Recurrent right hypochondrial pain associated with nausea and vomiting is typical for biliary colic. The patient with a sudden, intense pain in the right upper quadrant associated with a low-grade fever (37.5°C) and bilious vomiting (sometimes containing a worm) is more likely to have acute cholecystitis. The invading worms carry some enteric flora into the bile duct and cholangitis is manifested by a high-grade fever (38–40°C), jaundice, and a palpable gallbladder. The liver is enlarged and tender, serum bilirubin and alkaline phosphatase are elevated, and there is a marked leucocytosis. If untreated, septicaemia and endotoxic shock may supervene. A liver abscess is associated with high fever, hepatomegaly, and exquisite intercostal tenderness. Recurrent invasion of the common bile duct by *ascarides* is associated with papillitis and motor dysfunction of the sphincter of Oddi. This leads to impaired biliary drainage, biliary sepsis, and stone formation. The syndrome of recurrent pyogenic cholangitis due to *ascarides* has recently been reported in a few patients in endemic areas of ascariasis.

Diagnosis

Biliary ascariasis can be diagnosed with abdominal ultrasonography. Dilatation of the biliary tree due to obstructing worms can be shown. The cross-section of the bile duct containing a worm shows a bull's-eye configuration; the longitudinal plane shows an echogenic strip with a central, longitudinal, anechogenic tube. The spaghetti sign is caused by several worms being imaged at a time.

A computed tomogram will also show the worm in the dilated bile duct or gallbladder and is useful in locating a liver abscess. The most helpful investigation, however, is duodenoscopy with endoscopic retrograde cholangiopancreatography, which provides for both diagnosis and therapy. The worms can be seen in the duodenum or entering or leaving the ampulla of Vater and can be removed. Bile samples can be obtained and examined for ova or cultured for bacteria. The contrast study will show the position and number of worms. Especially when cholangitis is present, worms can be removed using forceps, Dormia basket, or balloon-tipped catheter. A nasobiliary catheter may be left in place for decompression.

Management

As for intestinal obstruction, fluids are given and nasogastric suction performed while antispasmodics are given intravenously to relax the sphincter of Oddi. This facilitates the return of the worms into the small intestine. At the same time, intravenous antibiotics are given. After the acute attack has subsided, an antihelminthic drug is given in order to expel the worms from the intestines. Most patients improve on this regimen. If not, duodenoscopy is performed and the worms extracted from the ampulla or bile ducts, as described above. This provides rapid relief of SYmptoms. Recurrence of symptoms usually signifies reinfection by worms and the treatment regimen may be repeated as necessary.

Very rarely, non-operative and endoscopic management may fail. Exploration of the common bile duct is required in such instances, especially if cholangitis has resulted in bile-duct stricture. It may be prudent to perform a wide bilioenteric anastomosis with the expectation that any subsequently invading ascarides will be free to move in and out of the biliary tree.

In the Colombian series, biliary ascariasis presented in 19 children under the age of 12 years. The majority responded to non-operative management. In a recent cohort of 500 Indian patients from Kashmir, the mean age was 35 years and the mean duration of illness was 6 years. Biliary colic was present in 280; 64 presented with cholecystitis, 121 with cholangitis, and four with liver abscess. Diagnosis was confirmed by ultrasonography and, where necessary, endoscopic retrograde cholangiopancreatography. All patients were initially treated palliatively and antihelminthic drug treatment was given after the acute symptoms had subsided. Extraction of worms from the duodenum, ampulla, and bile ducts led to rapid relief of biliary colic. In 12 patients, dead worms were removed endoscopically or surgically from the bile ducts. During the follow-up period, seven patients required surgery for the removal of intrahepatic and bile duct stones, which were found to be soft, pigment stones containing a nidus of ascaris segments.

Ascaris pancreatitis

One of the least commonly diagnosed, but often fatal, complications of ascariasis is acute pancreatitis caused by worm impaction in the ampulla and pancreatic duct. The pain is usually epigastric with radiation to the back, and is associated with vomiting and fever. Serum amylase and alkaline phosphatase are usually elevated. Ultrasonography may reveal an enlarged, echo-poor pancreas but the computed tomogram is more likely to show the worm in the duct. Emergency endoscopic retrograde cholangiopancreatography is indicated in all such patients because it shows the correct cause of the pancreatitis and allows prompt extraction of the obstructing worm.

In the series from Kashmir, 31 of 500 patients presented with acute pancreatitis, which was severe in three cases. Many of the patients improved on non-operative management, but 16 required worm extraction. One of these died from haemorrhagic pancreatitis despite laparotomy and extraction of four worms from the pancreatic duct and seven from the bile duct.

In the series from Damascus, ultrasonography and endoscopic retrograde cholangiopancreatography were used to confirm the diagnosis of pancreatobiliary ascariasis in 300 patients. Most patients had previously undergone endoscopic sphincterotomy for the removal of common bile-duct stones or worms. Using a Dormia basket, ascarides were readily cleared from the bile ducts but not from the pancreatic ducts in seven patients. Small volumes of dilute urograffin were used to flush out the worms towards the ampulla and the basket.

In an instructive case report from Europe, a 66-year-old Swiss patient with no history of alcohol abuse or gallstones died from fulminant pancreatitis of undetermined

aetiology. At autopsy, an ascaris worm was found impacted in the ampulla of Vater and pancreatic duct. In view of modern world travel and population migration, clinicians even in non-endemic regions of the world need to be aware of the lethal consequence of undiagnosed ascaris pancreatitis and to consider it in their differential diagnosis.

Ascaris in other locations

The larval form in the pulmonary circulation may fail to penetrate into an alveolus but may enter a pulmonary venous tributary and thence be carried into the left side of the heart. From there it can reach any part of the arterial circulation, including the brain and splanchnic bed. Immature worms have exited from the nasolachrimal duct, for instance.

The adult worm in the jejunum may migrate proximally into the duodenum (Fig. 2) and stomach and cause peptic ulcer-like symptoms and gastric-outlet obstruction. It may ascend the oesophagus into the pharynx and be vomited or be aspirated into the trachea to cause asphyxia, which may be fatal (Fig. 3). It may descend further down into the bronchus to cause atelectasis or lung abscess. It may pass into the nasopharynx and exit through the nostrils, or from the oropharynx and through the eustachian tube into the middle ear and through a perforated tympanum to the auditory canal.

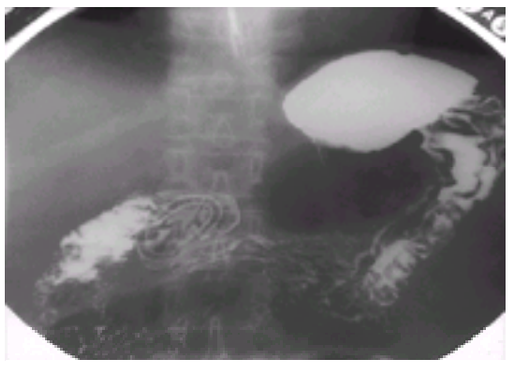


Fig. 2. Barium meal showing two adult ascaris worms in the duodenal bulb of a 37-year-old Filipino woman recently arrived in Oman; she presented with acute and severe epigastric pain and vomiting.

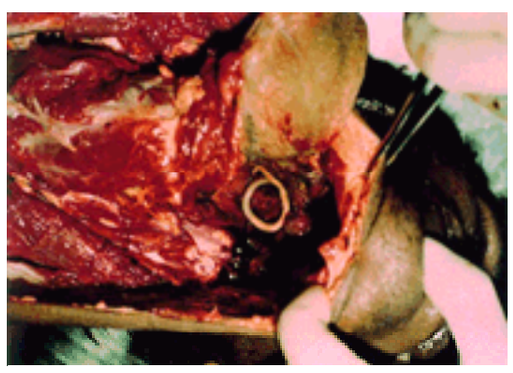


Fig. 3. Autopsy specimen from a 22-year-old Bangladeshi labourer working in Oman. He died of asphyxia following a bout of vomiting. Laryngeal obstruction was caused by aspiration of the ascaris worm shown in the glottis. (By courtesy of Dr C.K. Venkitasubramanian, Forces Pathologist, Royal Oman Police.)

Worms that have caused liver abscess may penetrate the diaphragm into the pleura, causing an empyema. From the liver the worm may also break into the hepatic veins or vena cava, be carried to the right heart and be embolized to the pulmonary artery. The worm may penetrate a Meckel's diverticulum or exit via a patent omphalomesenteric duct at the umbilicus. An occasional worm may perforate a bowel wall that has perhaps been weakened by typhoid or tuberculosis to reach the peritoneal cavity, where it dies, leaving remnants and eggs around which large granulomatous masses develop: these clinically resemble abdominal tuberculosis, schistosomiasis, or lymphoma. Other locations reached by worms include the renal pelvis and urinary bladder, the fallopian tubes, and the uterus via abnormal fistulas.

Other intestinal nematodes

The environmental conditions and cultural practices that favour infection by the other intestinal nematodes are similar to those required by ascarides. The parasitological data are summarized in Table 1; epidemiology and clinical features in Table 2; diagnosis and management in Table 3; and drug therapy in Table 4.

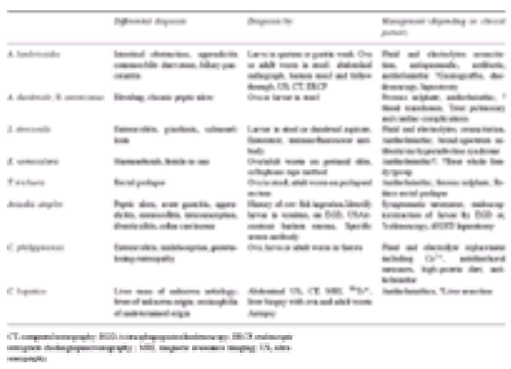


Table 3 Diagnosis and management of intestinal nematode infection

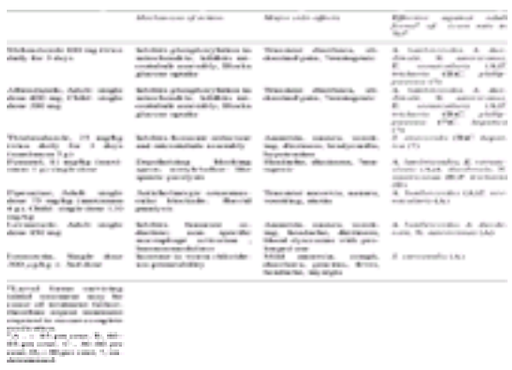


Table 4 Common broad-spectrum antihelminthic drugs

In endemic regions, the most intense infections occur in children, many of whom live on a borderline or inadequate diet. Polyparasitism and other infections add up to a major disease burden whose sequelae include anaemia, malabsorption, protein energy malnutrition, growth retardation, delayed puberty, and impairment of cognition and of academic performance. Although many safe and effective drugs are now available for treatment, continued exposure and reinfection require repeat drug therapies. These may be relatively expensive for the local economies and logistical problems often arise. Public-health control measures for prevention of infections are very important In the long term, improved economic development has the greatest promise of parasite control.

Ancylostoma duodenale and *Necator americanus* (hookworms)

Hookworms are usually considered together. *Ancylostoma* is more prevalent in temperate lands, while *Necator* is perennial in the tropics and subtropics. In communities associated with large dog populations as in Queensland, Australia, and elsewhere, human disease is also caused by *A. caninum*. Of potential surgical interest are the antihaemostatic molecules secreted by *N. americanus* and *A. caninum*, which inhibit human coagulation factors VII and Xa as well as platelet activation and aggregation, as part of the worm's survival strategy. The major clinical disease is blood loss and iron-deficiency anaemia due to the ingestion of blood and the complications of mucosal damage. In patients already on a marginal diet, a heavy worm load leads to chronic anaemia, which may be associated with impaired learning in children, placental ischaemia in pregnant women, and heart failure.

Strongyloides stercoralis (threadworm)

Strongyloidiasis may persist for decades because of its capacity for autoinfection. Recurrent disease has been seen in European veterans of the Second World War who served in Asia 50 years previously and in American veterans of the Vietnam war. The clinical triad of urticaria or larva currens, epigastric pain, and watery diarrhoea may be quite disabling. In a patient with an occult parasite, opportunistic strongyloidiasis may develop due to serious illness such as major burns, lymphoma, leukaemia, and other malignancies; or immunosuppressive therapy for ulcerative colitis, idiopathic thrombocytopenia, systemic lupus erythomatosus or for organ transplantation. If the correct diagnosis is not made, this may progress to the hyperinfection syndrome of disseminated strongyloidiasis. In these circumstances, the host–parasite balance is upset, the rhabditidiform larvae convert to the invasive filariform larvae within the intestines, and massive autoinfection, either internal (intestinal) or external (through perianal skin), results in a build up of large numbers of larvae. These infective filariform larvae then penetrate the submucosa of the bowel wall and migrate throughout the body into organs such as the liver, pancreas, kidney, brain, and endocardium. Secondary Gram-negative polybacterial septicaemia, fluid depletion, massive alveolar haemorrhage, and disseminated intravascular coagulation may develop: a mortality rate of 86 per cent has been recorded. Therefore, all patients who are candidates for immunosuppressive drugs should be screened for strongyloides and treated appropriately. Intriguing, there is an intrinsic, specific antistrongyloides activity by cyclosporin A, the potent immunosuppressive agent used in clinical organ transplantation. The mechanism of this effect is as yet undetermined but probably involves different cellular and molecular pathways independent of the immune system.

Enterobius vermicularis (pinworm)

This parasite is more common in temperate than in tropical and subtropical countries. The rapid expansion of the industry providing out-of-home child care has contributed to the prevalence of pinworms in the Western world. Since it occurs in family or group members, group treatment may be indicated. The ova may be disseminated in dust. Autoinfection commonly occurs by direct transfer of ova from the perianal region to the mouth via the fingers. Retroinfection may also occur: ova that hatch on the perianal skin release larvae that crawl back into the colon, and occasionally into the vulva and genital tract. *Enterobius vermicularis* worms are found sometimes in appendectomy specimens but it is uncertain whether the worms by themselves alone may cause acute appendicitis.

Trichuris trichuria (whipworm)

Heavy infection causes haemorrhagic colitis. Tenesmus, proctitis, and recurrent straining lead to rectal prolapse, which is usually reversible following drug therapy. *Trichuris* dysentery SYndrome has been associated with anaemia and growth retardation, which is reversed after effective drug treatment.

Anisakis simplex

Anisakis is a gastroinstestinal parasite of marine mammals such as whales, seals, and dolphins. Embryonated ova passed into faeces in the ocean hatch into larvae that develop along the food chain through crustaceans, squid and fish, and back to the mammals, where they mature into adult worms. The infective (L_3) larvae reside in the muscles of squid and commercial fish such as herring, mackerel, salmon, and cod. Human infection occurs accidentally when infected fish is eaten raw or undercooked. The larvae may be vomited promptly or may be killed and trapped in an eosinophilic granuloma. Surviving larvae that burrow into the gastric mucosa cause abdominal pain and gastric bleeding. Larvae that reach the ileum and colon may cause an acute abdominal syndrome. These larvae never develop into adult worms in humans (i.e., paratenic hosts), so no eggs are shed. It is important to make an accurate diagnosis because most cases resolve spontaneously and an exploratory laparotomy should be avoided. Anisakiasis is prevalent in Japan, Holland, and Scandinavia but is now spreading to the United States and other European countries owing to the great increase in fish consumption and the popularity of Japanese-style seafood and ‘sushi bars'. Anisakid larvae are killed by heating fish to 65°C for 10 min of freezing to –20°C for 5 days.

Capillaria philippinensis

Capillaria philippinensis is a gastroinstestinal parasite of migrating, fish-eating birds, which discharge embryonated ova into inland waters. These are ingested by crustaceans and small fish within which they hatch into the infective larva (L_3) in 3 weeks. Human capillariasis occurs when infected fish is eaten uncooked. The infective larvae are released in the stomach and moult into adult worms in the jejunal mucosa, where mating occurs. Both oviparous and larviparous females result. The larviparous account for the autoinfection, which causes a large accumulation of worms within a few weeks. Ova and larvae are passed in the faeces. The overwhelming numbers of these worms cause villous atrophy, crypt hyperplasia, and eosinophilic granulomatous inflammation of the jejunum.

Capillaria hepatica

Capillaria hepatica is an intrahepatic parasite of the cosmopolitan rat (*Rattus norvegicus*) and other rodents. The gravid female disintegrates within the liver parenchyma releasing eggs, which are normally retained within the living host. When a predator eats the infected rat or a scavenger (such as a domestic cat or dog) eats the infected dead rat, the eggs pass out unchanged in the faeces of this animal, which acts as a dispersal host. Embryonation of the ova occurs in the soil. Human infection occurs when food contaminated by embryonated ova is ingested. The ova hatch in the intestine and the larvae migrate through the portal venous SYstem into the liver, where they mature into adult male and female. They mate within 28 days, and both die and disintegrate, thus releasing thousands of ova from the gravid female. These ova do not develop further until passed into the soil by the death of the host.

Laboratory diagnosis of infection by intestinal nematodes (see Table 3)

A direct diagnosis of the intestinal nematode infection is usually made by finding the ova or larvae in stools using standard laboratory methods of examination. The Kato and Muira thick-smear stool technique is recommended for the detection of light infections with *Ascaris* or *Trichuris*; the Baermann funnel-filter technique is used to detect larvae in the stool of patients with strongyloidiasis. The adult worm may also be directly identified.

Nematode infections can also be diagnosed indirectly using immunological and biochemical tests such as haemagglutination, indirect immunofluorescence, radioimmunoassay, enzyme-linked immunosorbent assay, or immunoblot techniques. Most of these are not practical in the current clinical situation because of cross-reactivity and the problem of mixed infections. However, immunofluorescent and enzyme-linked immunosorbent assays may be useful in patients with chronic strongyloidiasis, in whom demonstration of the larvae may be difficult.

Antihelminthic drugs (see Table 4)

Many safe and effective drugs are now available; some are effective as single-dose therapy and are therefore useful for mass treatment; some have a broad spectrum of action and are particularly useful in patients infected with multiple parasites.

Further reading

Abbas AK, Lichtman AH, Pober JS, ed. *Cellular and molecular immunology*, 2nd edn, pp. 283–331. Saunders, Philadelphia, 1994.

Beckingham IJ, Cullis SNR, Krige JEJ Bornman PC, Terblanche J. Management of hepatobiliary and pancreatic ascaris infestation in adults after failed medical treatment. *British Journal of Surgery*

1998; **85**: 907–10.

Chappell LH, Wastling JM. Cyclosporin A: anti-parasite drug, modulator of the host–parasite relationship and immunosuppressant. *Parasitology* 1992; **105**: S25–40.

Garcia LS, Bruckner DA, ed. *Diagnostic medical parasitology*, 3rd edn. American Society of Microbiology Press, Washington DC, 1997.

Grencis RK, Cooper ES, *Enterobius, Trichuris, Capillaria* and hookworm including *Ancylostoma caninum.* Gastroenterology Clinics of North America 1996; **25**: 579–97.

Grove DI. Human strongyloidiasis. *Advances in Parasitology* 1996; **38**: 251–309.

Hotez PJ, Pritchard DI. Hookworm infection. *Scientific American* 1995; **272**: 68–74.

Khuroo MS. Ascariasis. *Gastroenterology Clinics of North America* 1996; **25**: 553–77.

Khuroo MS, Zargar SA, Mahajan R. Hepatobiliary and pancreatic ascariasis in India. *Lancet* 1990; **335**: 1503–6.

Kingston IB. Nematode collagen genes. *Parasitology Today* 1991; **7**: 11–15.

Kohatsu H, Zaha O, Shimada K, Chibana T, Yara I, Shimada A, Hasegawa H, Sato Y. A space-occupying lesion in the liver due to capillaria infection. *American Journal of Tropical Medicine and Hygiene* 1995; **52**: 414–18.

Liu LX, Weller PF. Strongyloidiasis and other intestinal nematode infection. *Infectious Diseases Clinics of North America* 1993; **7**: 655–82.

Maddern GJ, Denison AR, Blumgart LH. Fatal ascaris pancreatitis: an uncommon problem in the West. *Gut* 1992; **33**: 402–3.

Ochoa B. Surgical complications of ascariasis. *World Journal of Surgery* 1991; **15**: 222–7.

Salman AB. Management of intestinal obstruction caused by Ascariasis. *Journal of Pediatric Surgery* 1997; **32**: 585–7.

Sandouk F, Haffar S, Zada MM, Graham DY, Anand BS. Pancreatic–biliary ascariasis: experience of 300 cases. *American Journal of Gastroenterology* 1997; **92**: 2264–7.

Schantz PM. The dangers of eating raw fish. *New England Journal of Medicine* 1989: **320**: 1143–5.

Sher A, Coffman RL. Regulation of immunity to parasites by T cells and T cell-derived cytokines. *Annual Review of Immunology* 1992; **10**: 385–409.

Shirahama M, Koga T, Ishibashi H, Uchida S, Ohta Y, Shimoda Y. Intestinal anisakiasis: US in diagnosis. *Radiology* 1992; **185**: 789–93.

Strober W, Fuss IJ. The mucosal immune system. In: Stites DP, Terr AI, Parslow TG, ed. *Medical immunology*, 9th edn, pp. 196–209. Appleton and Lange, Stanford, CT, 1997.

Wasadikar PP, Kulkarni AB. Intestinal obstruction due to ascariasis. *British Journal of Surgery* 1997; **84**: 410–12.

47.2.6 Biliary and intestinal trematodes

Oluwatope A. Mabogunje, Christopher S. Grant and A. S. Daar

[Biliary and intestinal trematodes](#)
[Morphology and lifecycle](#)
[Host–parasite interaction](#)
[Specific immune response](#)
[Opisthorchiasis](#)
[Complications of opisthorchiasis](#)
[Clinical presentation](#)
[Diagnosis](#)
[Management](#)
[Control](#)
[Fascioliasis](#)
[Complications of fascioliasis](#)
[Diagnosis](#)
[Differential diagnosis](#)
[Management](#)
[Prognosis](#)
[Fasciolopsiasis](#)
[Complications of fasciolopsiasis](#)
[Diagnosis and management](#)
[Control of *Fasciola hepatica* and *Fasciolopsis buski*](#)
[Further reading](#)

Biliary and intestinal trematodes

The class Trematoda (flukes) contains several species that are parasites of humans. The schistosomes (blood flukes) are of special biological and clinical importance and are dealt with elsewhere (see Chapter 47.2.2 and Chapter 47.2.3). More than 50 million people are infected by one or more species of the liver, lung, or intestinal fluke. Only the major liver and intestinal flukes are discussed in this chapter. These are *Opisthorchis sinensis* (the Chinese liver fluke, previously known as *Clonorchis sinensis), Opisthorchis viverrini, Opisthorchis felineus* (cat liver fluke), *Fasciola hepatica* (sheep liver fluke), and *Fasciolopsis buski* (the giant intestinal fluke) (Table 1).

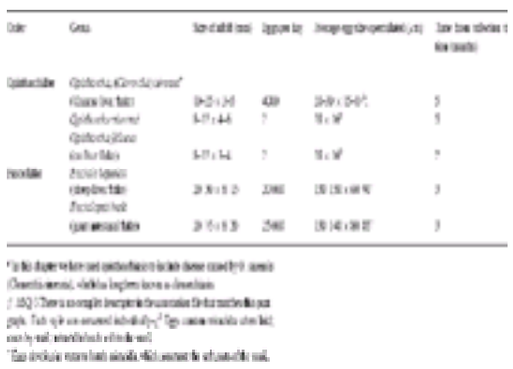


Table 1 Phylum Platyhelminthes, class Trematoda: species of surgical interest (excluding Schistosoma)

Infection by these liver and intestinal flukes may produce hepatic, biliary, and intestinal complications that require surgical intervention.

Morphology and lifecycle

Liver and intestinal flukes have common biological and morphological characteristics. They are digenetic hermaphrodites capable of infecting other mammalian species, which thus enlarges the reservoir of adult flukes from which human infection can occur. Human infection also depends on the epidemiology and abundance of the first intermediate hosts, which are usually freshwater or amphibious snails and of the second intermediate hosts, which are freshwater fish or vegetation (Table 2).

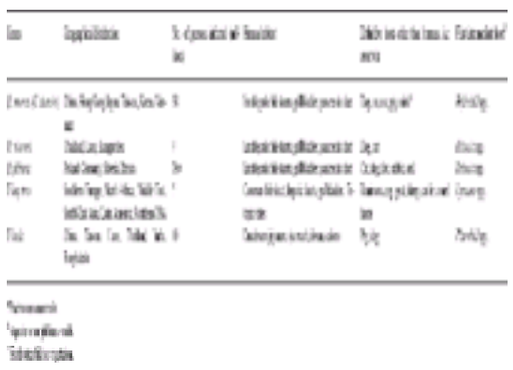


Table 2 Epidemiology of the major liver and intestinal flukes.

The adult worm has a thick oval leaf shape with a glycocalyx tegument which may be covered with spines. Underlying the tegument is a layer of muscle fibres, the contraction of which can alter the shape of the fluke or move it along. An anterior or oral sucker leads to the digestive tract. There is also a ventral sucker; both are muscular organs of locomotion and attachment to host tissues. Each worm has male and female genital organs and produces operculated ova that are either self- or cross-fertilized.

The ova are passed in the host's faeces into water. The miracidia hatch either in the water or after being ingested by the snail intermediate host. Within the snail, each miracidium develops through the stages of sporocyst, redia, and daughter redia; the last multiplies asexually into thousands of free swimming cercariae which leave the snail. They reach the second intermediate host, either a freshwater fish or plant, and encyst as the metacercariae, the infective form.

A definitive host becomes infected by ingesting the metacercariae encysted on raw or undercooked fish or vegetables. The metacercariae excyst in the duodenum and advance into the biliary tracts or descend into the jejunum where the juvenile flukes mature within 3 to 5 months and begin oviposition. The parasitological and epidemiological data are summarized in Table 1 and Table 2.

Through their oral suckers, liver and intestinal flukes feed on blood, mucus, tissues, and intestinal contents of the host. It is estimated that each *Fasciola* ingests 0.2 ml of blood daily. The gastrodermal cells of the fluke secrete proteases that degrade globin into peptides and amino acids. Glucose, fructose, amino acids, and other nutrient molecules are directly absorbed by diffusion and active transport through the tegument.

F. hepatica converts succinate to propionate as an energy-yielding process by recycling coenzyme A. The enzymes involved are vitamin B_{12}-dependent. This process may explain in part the tendency of many intestinal helminths to concentrate vitamin B_{12}, to the detriment of their hosts.

Host–parasite interaction

The miracidia and cercariae are free-swimming larvae with limited lifespans; they must locate the intermediate hosts or perish. In order to ensure transmission, their sensory motor systems are genetically programmed to respond to a complex series of trigger cues specific to each stage of the lifecycle. These include tactic and kinetic responses to light, gravity, vibration, and chemoattractants. Together with high fecundity and chance, such behaviour ensures a high probability of contact between parasites and the intermediate hosts in endemic regions. A better understanding of this phenomenon would provide a more effective avenue for parasite control than would the indiscriminate use of molluscicides.

For the ingested metacercaria, the presence of bile salts and proteolytic enzymes indicates that the duodenum has been reached. Releaser responses vary among the flukes: *Opisthorchis* spp. find and traverse the sphincter of Oddi into the common bile duct; *Fasciola* perforates through the duodenum into the peritoneum and then locates and penetrates the liver to reach the bile ducts, while *Fasciolopsis* moves downstream to the jejunum.

The adult fluke in the bile duct must avoid being flushed out by the tides of bile encountered during the digestive cycles of the host. The ventral suckers of *F. hepatica* respond to the host's gastrointestinal hormones: neuromuscular activities of the suckers increase to strengthen attachment in response to cholecystokinin (which contracts the gallbladder, relaxes the sphincter of Oddi, and increases bile flow) but relax in response to motilin, which is antagonistic to cholecystokinin. *F. hepatica* also produces large quantities of proline by using an arginase enzyme system whose activity is four times that of the host and is not subject to end-product inhibition. Intraperitoneal infusion of proline in rats causes bile duct hyperplasia similar to that observed in the early phase of biliary fascioliasis. Thus, if proline has the same effect on the bile ducts of ruminants and humans, its synthesis would be a factor in pathogenesis.

Specific immune response

The presence of the fluke causes an immune response, which is multifactorial and differs according to the maturity of the fluke, the phase and site of infection, and host genetic factors. The outcome depends in part on the fluke's capacity for evasive action.

The metacercariae of *Opisthorchis* excysting in the duodenum home towards the papilla of Vater. The lumenal content of gastric, biliary, and pancreatic secretions probably limit interaction between the parasites and epithelial cells. Nevertheless, soluble fluke antigens probably trigger a T-cell derived signal to the mucosal immune SYstem, which results predominantly in the production of IgA antibodies by the mucosal follicles of both the intestines and the bile duct. While the serum and bile levels of both total and parasite-specific IgA, IgG, and IgE are elevated, analysis suggests that the IgG and IgE in bile are derived entirely from the serum whereas some of the IgA in the bile is locally synthesized in the hepatobiliary system.

Experiments with the metacercariae of *F. hepatica* suggest that the dog and cat may have a natural resistance, whereas the sheep and goat have low or no resistance. Clinically, humans seem to occupy an intermediate position, showing some resistance to *Fasciola*. During a primary infection, the newly excysted juveniles penetrate the duodenal wall with little response. While migrating through the peritoneum and liver, however, these juveniles trigger communication between the peritoneal, the mucosal, and the systemic immune systems. As a result, the intestinal mucosa is primed against subsequent excysting juveniles. These are more vigorously attacked by parasite-specific IgA and IgG antibodies which reach the lamina propria. Many of them become coated with IgG1 and IgG2a antibodies and are trapped in the liver parenchyma without reaching the bile duct. In a few patients, this incompletely understood response is of sufficient magnitude to eliminate all invading flukes, resulting in spontaneous cure.

The tegument of the juvenile fluke is enveloped in a secreted glycocalyx that bears unique antigens and is continually replaced during the migration through the intestinal wall, peritoneum, and liver parenchyma. It is finally discarded only after the fluke enters the bile duct, where the overall concentration of immunoglobulins is much lower than in serum. As a result of the antigenic changes of the glycocalyx, neither specific antibody plus complement, nor antibody-mediated eosinophil adherence is apparently able to damage the fluke's tegument. Thus, glycocalyx replacement appears to be a mechanism of evading antibody-mediated attack. By the time the immune system responds to the infective stages, the mature parasite expresses new antigens and is no longer a target for immune elimination. This in part accounts for the longevity of these flukes within the host.

In addition, peroxidoxins, a family of distinct antioxidant enzymes newly discovered, have biological activities which include protection against reactive oxygen species and repair of oxidatively damaged molecules. The fact that the peroxidoxin of *F. hepatica* is actively secreted suggests that it has a possible role in the defence of the fluke against reactive oxygen species such as the superoxide radical(O_2^-) generated by the host's cytotoxic immune effectors. Furthermore, excretory and secretory products of the adult fluke are toxic to lymphoid cells *in vitro* and prevent antibody-mediated adherence of peritoneal cells to juvenile flukes. A similar phenomenon *in vivo* would protect against any cell-based immune attack on the fluke.

Infection by *F. hepatica* causes increased serum levels of IgG, IgM, and IgE in humans. Total and specific IgE levels show a positive correlation with egg burden, age, clinical features, and the degree of eosinophilia. IgA levels are usually normal. In both opisthorchiasis and fascioliasis, the role of the increased levels of parasite-specific antibodies in protective immunity is not clear. In endemic regions, multiple infections occur in adults during several years of eating infected raw fish or water vegetables. Egg output therefore increases over time, but reaches a plateau in patients older than 20 years and actually declines beyond the age of 50. The decline may be the consequence of parasite-associated mortality, but is more likely to reflect the slow development of concomitant immunity, whereby after several years of repeated infections, the host harbours a population of adult flukes and becomes relatively resistant to fresh infections.

In animal experiments, calves infected with *Schistosoma bovis* and sheep infected with *Schistosoma mansoni* developed significant resistance to *F. hepatica*. Common antigens have now been reported between *S. mansoni* and *F. hepatica*. Therefore, where both trematodes are endemic, some form of heterologous immunity might develop. It is not definitely known whether such a phenomenon occurs in humans.

Opisthorchiasis (Table 2 and Table 3)

Liver	Gallbladder
Hepatolithiasis	Cholecystitis
Biliary lithiasis	Cholelithiasis
Hepatic abscess	Gallbladder empyema
Lobar atrophy	Gallbladder hydrops
Retention cysts	Salmonella carrier state
Recurrent pyogenic cholangitis[a]	Pancreas
? Biliary portal cirrhosis	Pancreatitis
Mucin-secreting cholangiocarcinoma	Pancreatolithiasis
Fistulas	
Biliogastric	
Biliocutaneous	
Biliobronchial	
Biliopulmonary	

[a]Recurrent pyogenic cholangitis and *O. sinensis* occur in Japan, but some patients with recurrent pyogenic cholangitis have cleared have *O. sinensis* infection.

Table 3 Pathological complications of opisthorchiasis

The term opisthorchiasis encompasses infection by all three *Opisthorchis* species, including infection by *O. sinensis* (also known as clonorchiasis).

The lifecycle of the three *Opisthorchis* species and the various disease syndromes they cause in humans are very similar. In many endemic regions, fish are raised in ponds where human and animal faeces are used as fertilizers. The prevalence of human infection with *O. sinensis* may be as high as 70 per cent in parts of southern China and Hong Kong. Neither the snail nor fish intermediate hosts are indigenous to Hong Kong but infected fish are imported from China daily for local consumption. In Korea, men eat raw fish at all-male parties and so are more often infected than women. In north-east Thailand, the prevalence of *O. viverrini* among snails is 1 per cent, but may reach 100 per cent in certain cyprinoid fish. In some locations 90 per cent of human adults are infected. In all endemic regions, infections begin in childhood but clinical disease usually becomes apparent during the third decade or later.

The metacercariae are ingested in a variety of raw fish preparations such as yue-shan and yue-shan chuk in China and Hong Kong; sashimi, sushi, and sunomono in Japan; and koi pla in Thailand. After excysting in the duodenum, the metacercariae travel through the ampulla of Vater into the common bile duct. There is no tissue-invasive phase. The juvenile flukes have narrow bodies and are thus able to swim far proximally within the intrahepatic ducts, where sexual maturation and oviposition occurs. Grossly dilated ducts may thus become visible just beneath Glisson's capsule. In massive infections, even the common bile duct and gallbladder and pancreatic ducts are filled with flukes.

Complications of opisthorchiasis (Table 3)

Acute syndrome

Visitors to endemic regions who become infected with several metacercariae may develop acute symptoms similar to the Katayama fever of schistosomiasis, which may last 2 to 4 weeks. The syndrome comprises irregular high fever, painful hepatomegaly, anorexia, flatulence, and diarrhoea. There may be lymphadenopathy, leucocytosis, and eosinophilia. This syndrome is not common in residents of the region, who become symptomatic only after several years of accumulated infections.

Chronic biliary phase

The flukes may live up to 30 years in the bile ducts. They cause chronic irritation and injury to ductal epithelium by the mechanical activities of the suckers and by body movements, by the production of toxic metabolic products, by the immune reaction of the host, and by secondary bacterial infection. The extent of ductal injury is related to the number of flukes present and the duration of infection. Up to 20 000 flukes have been recovered from one patient.

Early changes include excessive mucin production, desquamation, inflammatory granulation, and repair with fibrosis and epithelial hyperplasia. Later, localized strictures and dilations of intrahepatic bile ducts lead to clubbing or cysts containing white bile or mucus. In late cases these sacs may coalesce into large multilocular or unilocular cysts of the liver. The wall is formed by the distended bile duct, with marked epithelial adenomatous hyperplasia.

Secondary infection with bacilli such as *Escherichia coli*, which possess b-glucuronidase, causes deconjugation of bilirubin diglucuronide and the formation of calcium bilirubinate stones, with the ova and fragments of dead flukes as the nidus. The presence of bile duct strictures, bacterial infection, and gallstones is conducive to the development of recurrent pyogenic cholangitis. Such attacks may occur in infected persons even years after emigration to non-endemic regions of the world. Recurrent pyogenic cholangitis may lead to biliary cirrhosis, liver impairment, portal hypertension, or to fatal septicaemia. Other pathological changes of the liver reported in complicated opisthorchiasis, especially *O. sinensis* infection, are listed in Table 3. Chronic recurrent infections of the pancreatic duct are associated with periductal fibrosis, ductal dilatation, and foci of mucinous and squamous metaplasia.

Typhoid infection in a patient with chronic opisthorchiasis often leads to the establishment of the salmonella carrier state. The bacilli adhere to the surface of the fluke and replicate, and salmonella is shed in the patient's bile and faeces, producing a source of virulent infection in the community.

Chronic opisthorchiasis and cholangiocarcinoma

The gravest complication of opisthorchiasis, well documented in infection with *O. sinensis* and *O. viverrini*, is mucin-secreting cholangiocarcinoma, which may be a terminal event. Areas endemic for opisthorchiasis consistently report a higher prevalence of cholangiocarcinoma and in younger patients than non-endemic areas. *Opisthorchis* infections are more prevalent in patients with cholangiocarcinoma than in those without. The exact mechanism of carcinogenesis is unknown, but in hamsters, experimental infection with *O. viverrini* acts as a promoter of bile duct carcinoma initiated by *N*-nitrosodimethylamine. Adenomatous hyperplasia and goblet cell metaplasia that occur in human infection may make the bile ducts more susceptible to otherwise subcarcinogenic doses of *N*-nitroso compounds and their precursors, derived either endogenously from macrophages activated by chronic fluke infections or exogenously from traditionally preserved foodstuffs such as salted fish, dried shrimp, or cured pork. In north-east Thailand, where *O. viverrini* infection is endemic, a combination of the tumour markers CA125 and CA19–9 was shown to be potentially useful in the early detection of cholangiocarcinoma in a population at very high risk for the tumour.

A recent study of a similar Thai population demonstrated a striking causative association between acalculous hepatobiliary abnormalities and cancer development among adults, particularly men, with moderate and heavy chronic infection with *O. viverrini*. It is postulated that sludge in a damaged gallbladder increases accumulation of the carcinogenic components of bile and their exposure to the biliary epithelium. One-third of the carcinomas occurred in the extrahepatic bile ducts downstream from the location of the flukes.

Clinical presentation

The majority of infected persons living in endemic areas have no symptoms. Moderate infection is associated with non-specific symptoms such as malaise, anorexia, flatulence, diarrhoea, and occasional biliary colic. Heavy infections due to several thousands of flukes in the bile ducts cause right upper quadrant pain, fever, tender hepatomegaly, and leucocytosis. Complications of the infection produce clinical syndromes resembling those that occur in similar non-parasitic conditions. Thus, in endemic areas, acute cholecystitis, pancreatitis, liver abscess, and pyogenic cholangitis must be suspected to be due to flukes. The differential diagnosis of obstructive jaundice and of hepatomegaly must include, for instance, recurrent pyogenic cholangitis in a febrile patient or cholangiocarcinoma in a non-febrile patient. Tender hepatomegaly without jaundice may result from large liver cysts or cholangiocarcinoma. Other possibilities should be given due consideration in compatible clinical settings (Table 3).

Diagnosis

In the Orient, the diagnosis of infection by liver flukes is usually suspected in all young adults presenting with a compatible clinical syndrome. The definitive diagnosis is made by the identification of ova or adult flukes. Ova may be found in the faeces, as may dead flukes after anthelmintic treatment. Both ova and adult flukes may be recovered in bile obtained by duodenal aspirate, endoscopic retrograde cholangiopancreatography (**ERCP**), percutaneous transphepatic cholangiography, during common bile duct exploration, or at autopsy. Differentiation between the various *Opisthorchis* species is based on geographical location, characteristics of ova, and fluke morphology (see Table 1).

Immunological techniques for diagnosis based on complement fixation and enzyme-linked immunosorbent assays have been developed but are not generally available. New methods for diagnosis based on the detection of parasite DNA and antigens in faeces by dot-blot hybridization techniques are now being developed.

Abdominal ultrasonography and CT scan usually show dilated intrahepatic bile ducts without dilatation of the extrahepatic bile ducts. Sonograms may identify adult flukes as fusiform, weakly echogenic, non-shadowing casts which may be mobile, especially in the gallbladder. Both studies show dilated bile ducts filled with stones in patients with recurrent pyogenic cholangitis. Both identify cystic or solid masses in the liver such as cysts, abscesses, and tumours, and also show lobar atrophy. ERCP shows dilated intrahepatic bile ducts and wavy filamentous filling defects, representing the flukes. Ductal irregularities due to epithelial proliferation are also shown, as are strictures and stones in those with recurrent pyogenic cholangitis. When ERCP is not possible, transhepatic cholangiography may provide similar information.

Hepatic arteriography is useful when a liver resection is planned. Cholangiocarcinoma is relatively avascular with encasement or displacement of the arteries, whereas hepatocellular carcinoma, which is also common in the Orient, shows florid vascularization. Where MRI is available, it is useful in delineating the extent of the carcinoma and its spatial relationship to adjacent structures.

Management

All persons diagnosed as having opisthorchiasis should be treated since chronic infection, even of low intensity and asymptomatic, has the potential for serious complications such as recurrent pyogenic cholangitis and cholangiocarcinoma. The current drug of choice is praziquantel (Table 4).



Table 4 Chemotherapy for liver and intestinal flukes.

Patients with surgical and other complications, such as cholecystitis, cholangitis, and pancreatitis should be managed by standard surgical and endoscopic techniques. ERCP and sphincterotomy may be adequate to extract the flukes, stones, and ova. Balloon dilatation of strictures with or without metal stenting may be attempted after sphincterotomy. Other patients may require cholecystectomy, common bile duct exploration, and evacuation of flukes and stones, followed by T-tube drainage. These procedures may also be performed laparoscopically where such facilities are available.

Liver abscesses and liver cysts may be adequately treated by ultrasound- or CT-guided tube drainage. However, in a report of a Russian series, 17 of 34 patients had cysts of 5 cm diameter or greater and were operated upon either by excision of cyst wall and omentoplasty or by liver resection. The results were satisfactory over a period of 2 to 11 years. Liver resections and bilioenteric reconstructions may be appropriate in patients with complicated recurrent pyogenic cholangitis, and in well-selected candidates, liver resections may also be feasible for cholangiocarcinoma. In a recent series of 30 patients with *O. viverrini*-associated cholangiocarcinoma, the tumour occupied the confluence of the hepatic ducts in all patients, metastatic lymph nodes were present in 21, and carcinomatosis in 4. Tumour resection was performed in 7 patients with a 2-year survival of 43 per cent. Segment III cholangiojejunostomy was performed in 18 and an alternative bypass in 4 for malignant hilar obstruction; all died within 2 years. One patient had exploratory laparotomy only. Thus surgical procedures provided effective palliation with acceptable survival and without operative mortality. This outcome is quite similar to that for patients with cholangiocarcinoma unrelated to liver flukes.

Control

Attempts to control trematode infections by the control of the intermediate snail hosts have been generally unsuccessful. The new approach is to regard the control of food-borne parasitic zoonoses as a food-safety issue involving both a change in lifestyle and improved food technology. Thus, an innovative programme for the control of *O. viverrini* in Thailand is based on the combination of educating the community at large about the dangers of eating raw fish and treating all infected persons with praziquantel. Praziquantel is effective in killing the flukes, and probably decreases the susceptibility to cancer formation, but is unlikely to affect the course of already initiated carcinogenesis.

Fascioliasis (Table 1 and Table 2)

Human infection by the sheep liver fluke occurs mainly in temperate and subtropical countries with large sheep and cattle populations, among which infections are more common than in people. Clinical cases have been reported from more than 40 countries in the past two decades. In a collective review of 2594 persons with fascioliasis, more than half were reported from France, Portugal, Spain, Great Britain, and Tadjikistan. Other reports came from Cuba, Peru, Algeria, Egypt, and Australia. However, the total number of infected persons is unknown because light infections cause no SYmptoms.

Human infection is usually acquired by ingesting the metacercariae encysted on watercress prepared as a green salad. After excysting in the duodenum, the metacercariae burrow through the wall into the peritoneal cavity and travel along the parietes to the liver. They penetrate Glisson's capsule and eat through the parenchyma to enter the bile ducts where they attach with the suckers. This invasive phase takes 5 to 6 weeks. Maturation to adulthood occurs in the hepatic ducts, common bile ducts, and the gallbladder, where the adult fluke is confined by its relatively large body size (Table 1). Oviposition begins 3 months after infection.

Some metacercariae develop in the peritoneal cavity, abdominal wall muscles, or subcutaneous fat, while others enter the blood circulation and are transported to widespread ectopic foci. In sheep, about 600 metacercariae invading the liver cause massive destruction known as ‘liver rot' which is sufficient to kill the animal.

Complications of fascioliasis

The development of symptoms depends on the number of metacercariae that have perforated the intestine to reach the liver, the number that eventually mature in the bile ducts and elsewhere, and the duration of infection.

Acute invasive phase

Two weeks after infection, during the attack on Glisson's capsule by several metacercariae, subcapsular haematoma and haemoperitoneum may occur; in severe infestation this results in haemorrhagic shock requiring emergency laparotomy. There are patchy haemorrhages with hepatocyte necrosis, abscesses, and debris along the tracts, followed by reparative processes such as granulation tissue, regeneration, and fibrosis.

The patient complains of headache, anorexia, nausea and vomiting, pruritus, right hypochondrial pain, and high fever. There is tender hepatomegaly, urticaria, and marked eosinophilia. This may persist for 3 or 4 months.

Early and chronic biliary phase

Because of the size of the adult fluke in relation to the bile ducts, even light infections may sometimes cause transient biliary colic. Adult flukes attached to the ducts may eat through the wall into the parenchyma and vasculature, causing haemobilia. This manifests as haematemesis and melaena associated with epigastric pain, obstructive jaundice, and anaemia. Otherwise, a normocytic, hypochromic anaemia with eosinophilic bone marrow hyperplasia is more typical.

In heavy infections, mechanical irritation by the suckers and injury by the pines on the fluke's body, combined with fluke metabolites such as proline, cause inflammation, hyperplasia, and desquamation of ductal epithelium. Together with fibrosis, these changes lead to thickening and dilatation mainly of the extrahepatic bile ducts. The gallbladder is enlarged and oedematous with thickened walls and contains multiple small stones, with or without blood clots and fibrin plugs. Secondary bacterial infections result in acute cholecystitis or empyema, choledocholithiasis, and cholangitis. At this stage the patient presents with right upper quadrant pain and obstructive jaundice, fever, and chills. There is tender hepatomegaly, eosinophilia, and liver function tests are abnormal. On account of the longevity of the fluke, episodic attacks of cholangitis may occur.

Occasionally, fascioliasis causes acute pancreatitis, which may also be recurrent. This has been ascribed to the blockage of the pancreatic duct at the ampulla of Vater by either the ova or dead adult flukes from the common bile duct. The patient complains of recurrent epigastric pain radiating to the back. Serum amylase levels are usually elevated.

Ectopic fascioliasis

During the acute invasive phase, some metacercariae fail to reach the liver. In an outbreak of acute invasive fascioliasis in the United Kingdom in 1970, only 20 (71 per cent) of the 28 patients subsequently passed eggs in their faeces, indicating that many flukes perished in other tissues without reaching the bile ducts.

Ectopic fascioliasis frequently manifests as painful migratory subcutaneous nodules, which may be mistaken for infected sebaceous cysts. Excision and microscopic examination shows a fluke trapped in a zone of intense eosinophilic reaction surrounded by haemorrhage, fat necrosis, and some of its own ova. Such lesions have been found in the stomach, caecum, appendix, pancreas, spleen, orbit, brain, heart, lungs, pleura, and the epididymis.

Laryngopharyngitis

A SYndrome comprising oedema and bleeding of the pharynx with acute odynophagia and laryngeal obstruction known as ‘halzoun' in Lebanon and ‘marrara' in the Sudan was formerly thought to be caused by the ingestion of juvenile fasciola flukes in the raw liver of sheep or goat. The syndrome is now considered probably to be due to the nymphs of the pentostomid *Linguatula serrata* (tongue worm) or to leeches such as *Limnatis nilotica*.

Diagnosis

The definitive diagnosis of fascioliasis depends on the identification of the eggs in faeces or duodenal aspirates. Dead flukes expelled after anthelmintic therapy or live flukes recovered during ERCP and sphincterotomy or during common bile duct exploration can also be identified.

During the invasive phase before oviposition and in ectopic fascioliasis, serological tests such as indirect haemagglutination, complement fixation, and countercurrent immunoelectrophoresis tests are quite useful, but they show cross-reactivity with other trematodes. The enzyme-linked immunosorbent assay may be more specific. Abdominal ultrasound and CT scans can detect subcapsular haematoma and often show hypodense lesions during liver invasion. CT may also show tortous linear peripheral lesions. During the chronic biliary phase, ultrasound shows extrahepatic bile duct dilatation with irregular wall thickening. Live flukes appear as oval mobile bodies without acoustic shadowing, especially in the gallbladder. ERCP performed to relieve cholangitis usually shows bile duct dilatation with linear crescent-shaped defects in the common bile duct. Liver biopsy sometimes demonstrates the fluke within the parenchyma.

Differential diagnosis

During the acute invasive phase, differential diagnosis includes other causes of fever and tender hepatomegaly such as viral hepatitis and amoebic or pyogenic liver abscess. The combination of hepatomegaly, eosinophilia, and elevated serum alkaline phosphatase, and a positive serological test should suggest the correct diagnosis in a patient who gives the history of recent consumption of watercress salad.

The SYndrome of acute abdominal pain with haemoperitoneum and shock is a surgical emergency requiring laparotomy. At laparotomy, subcapsular haematoma with multiple haemorrhagic points in the liver are clues to the diagnosis. Metacercariae may be identified in the haemoperitoneum. During the chronic biliary phase, cholecystitis and cholangitis may require laparoscopic or open cholecystectomy. The diagnosis of fascioliasis may first be made during the operation. Ectopic fascioliasis is often diagnosed only after excision and histological examination.

Management

The definitive treatment of uncomplicated acute fascioliasis is by means of specific anthelmintics. Triclabendazole is currently the drug of choice but may not be available in all countries, in which case bithionol is quite effective (Table 4).

In complicated fascioliasis, abdominal emergencies such as haemoperitoneum and hypovolaemic shock, haemobilia, and cholangitis require a standard surgical approach, including resuscitation with intravenous fluids and blood transfusion, antibiotics, nasogastric tube suction and irrigation, followed by a laparotomy. The procedures indicated may include cholecystectomy with common bile duct exploration and removal of obstructing flukes. When the diagnosis of haemobilia has been confirmed by hepatic arteriography, selective catheter embolization may suffice, failing which a liver resection may be required.

Prognosis

Infection with *Fasciola* is of low overall prevalence. Although cholangitis may result from obstruction of the common bile duct by the large flukes, recurrent pyogenic cholangitis is not associated with infection by this fluke; chronic fascioliasis, unlike chronic opisthorchiasis, is not associated with cholangiocarcinoma. Death due to fascioliasis is uncommon. However, fatalities do occur among patients who sustain massive haemobilia and in those with severe liver damage.

Fasciolopsiasis (Table 1 and Table 2)

Human infection by the giant intestinal fluke, fasciolopsiasis, is most common in regions where both the domestic pig and water plants are cultivated, as in eastern China and Thailand. The dog serves as a reservoir in Bangladesh. Ova in the faeces of infected dogs, pigs, and humans wash into the ponds. The miracidia hatch and attack planorbid snails. The cercariae emerging from the snails encyst on the stems, leaves, and pods of water plants such as water-caltrop, -chestnut, -bamboo, -cress, and -morning glory. When pigs or people, especially children, ingest the raw vegetables, the metacercariae excyst and attach themselves to the duodenal or jejunal mucosa and develop into adult worms in about 90 days, when oviposition begins.

Complications of fasciolopsiasis

The suckers cause local inflammation, ulceration, bleeding, and abscess formation. A few worms cause no SYmptoms. When hundreds or thousands are present, some spread up into the stomach and others down into the ileum and colon. They cause epigastric pain similar to that of duodenal ulcer. There is also profuse mucous diarrhoea containing undigested food.

Children with large worm burdens may become toxic from ab-sorption of fluke metabolites. Hypoalbuminaemia results from malabsorption and possible protein-losing enteropathy resulting in oedema and ascites. Vitamin B_{12} deficiency has been noted occasionally.

On account of its relatively large size, attachment of several worms within a short segment of bowel may cause partial or complete intestinal obstruction, which may be fatal if unrelieved. In endemic regions the differential diagnosis of intestinal obstruction always includes fasciolopsiasis.

Diagnosis and management

The large eggs are easy to identify in the stool. An adult worm may occasionally be expelled in the stool. Duodenal aspirate may recover both ova and fluke. Uncomplicated fasciolopsiasis is treated with praziquantel 25 mg/kg body weight orally, three doses in 1 day (Table 4).

When intestinal obstruction is present, a nasogastric tube is inserted and intravenous fluid replacement is begun. Praziquantel is given through the nasogastric tube. The drug paralyses the flukes and loosens their attachments to the bowel wall. They are carried into the colon and the obstruction is relieved.

Rarely, surgical relief may be required. Resuscitation is begun and praziquantel is given via the nasogastric tube. During laparotomy, the drug is milked down to the location of the flukes. Subsequently the worms can be massaged into the colon, failing which enterotomy and extraction of the worms is required.

Control of *Fasciola hepatica* and *Fasciolopsis buski*

The use of lavatories and of chemical fertilizers, together with modern techniques of livestock management, is likely to break the transmission of the parasites. Human infection can also be avoided by proper cleansing and cooking of water plants and their products before eating them.

Further reading

Al'Perovich BI, Rodicheva NS, Mitasov VI. Opisthorchiasis cysts of the liver. *Khirugiia-Mosk* 1991; **10**: 96–9.

Apt W *et al.* Treatment of human chronic fascioliasis with triclabendazole: drug efficacy and serologic response. *American Journal of Tropical Medicine and Hygiene* 1995; **52**: 532–5.

Bassiouny HK, Soliman NK, El-Daly SM, Badr NM. Human fascioliasis in Egypt: effect of infection and efficacy of bithionol treatment. *Journal of Tropical Medicine and Hygiene* 1991; **94**: 333–7.

Carpenter HA. Bacterial and parasitic cholangitis. *Mayo Clinic Proceedings* 1998; **73**: 473–8. [A good recent review of the subject.]

Chen MG, Mott KE. Progress in assessment of morbidity due to *Fasciola hepatica* infection. *Tropical Diseases Bulletin* 1990; **87**: R1–R38.

Elkins DB *et al.* Cross-sectional patterns of hepatobiliary abnormalities and possible precursor conditions of cholangiocarcinoma associated with *Opisthorchis viverrini* infection in humans. *American Journal of Tropical Medicine and Hygiene* 1996; **55**: 295–301.

Harinasuta T, Pungpak S, Keystone JS. Trematode infections. *Infectious Disease Clinics of North America* 1993; **7**: 699–716. [An authoritative review of trematode infection.]

Heel KA, Hall JC. Peritoneal defences and peritoneum-associated lymphoid tissue. *British Journal of Surgery* 1996; **83**: 1031–6.

Leung JWC, Sung JY, Chung SCS, Metreweli C. Hepatic clonorchiasis. A study by endoscopic retrograde cholangiopancreatography. *Gastrointestinal Endoscopy* 1989; **35**: 226–31.

Lim JH. Radiologic findings of clonorchiasis. *American Journal of Roentgenology* 1990; **155**: 1001–8. [Excellent and authoritative presentation of radiological features.]

Liu LX, Harinasuta KT. Liver and intestinal flukes. *Gastroenterology Clinics of North America* 1996; **25**: 627–36. [Authoritative and comprehensive coverage of the subject.]

Lloyd DA, Angorn IB. Biliary infestation. In: Blumgart LH, ed., *Surgery of the liver and biliary tract*, pp. 943–53. Churchill Livingstone, Edinburgh, 1988.

Markell EK, Voge M, John DT. *Medical parasitology*, 7th edn, pp. 183–99. WB Saunders, Philadelphia, 1992.

McGonigle S, Dalton JP, James ER. Peroxidoxins: a new antioxidant family. *Parasitology Today* 1998; **14**: 139–45.

Ona FV, Dytoc JNT. Clonorchis-associated cholangiocarcinoma: a report of two cases with unusual manifestations. *Gastroenterology* 1991; **101**: 831–9.

Price TA, Tuazon CU, Simon GL. Fascioliasis: case reports and review. *Clinical Infectious Diseases* 1993; **17**: 426–30.

Pulpeiro JR, Armesto V, Varela J, Corredoira J. Fascioliasis: findings in 15 patients. *British Journal of Radiology* 1991; **64**: 798–801.

Pungpak S, Akai PS, Longnecker BM, Ho M, Befus AD, Bunnag D. Tumour markers in the detection of *Opisthorchis*-associated cholangiocarcinoma. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1991; **85**: 277–9.

Rahman KM, Idris M, Azad Khan AK. A study of fasciolopsiasis in Bangladesh. *Journal of Tropical Medicine and Hygiene* 1981; **84**: 81–6.

Srivatanakul P *et al. Opisthorchis viverrini* infestation and endogenous nitrosamines as risk factors for cholangiocarcinoma in Thailand. *International Journal of Cancer* 1991; **48**: 821–5.

Sukhdeo MVK. Habitat selection by helminths: a hypothesis. *Parasitology Today* 1990; **6**: 234–7.

VanMilligen FJ, Cornelissen JB, Hendricks IM, Gaasenbeck CP, Bokhout BA. Protection of *Fasciola hepatica* with infiltrates of eosinophils, IgG1 and IgG2a antibiotics around the parasites. *Parasite Immunology* 1998; **20**: 285–92.

Veerappan A, Siegel JH, Podany J, Prudente R, Gelb A. *Fasciola hepatica* pancreatitis: endoscopic extraction of liver parasites. *Gastrointestinal Endoscopy* 1991; **37**: 473–5.

Watanapa P. Cholangiocarcinoma in patients with opisthorchiasis. *British Journal of Surgery* 1996; **83**: 1062–4

Wong J, Choi TK. Recurrent pyogenic cholangitis. In: Blumgart LH, ed., *Surgery of the liver and biliary tract*, pp. 977–92. Churchill Livingstone, Edinburgh, 1988. [An impressive chapter by an experienced team.]

Wongratanacheewin S, Bunnag D, Vacusorn N, Sirishinha S. Characterization of humoral immune response in the serum and bile of patients with opisthorchiasis and its application in immunodiagnosis. *American Journal of Tropical Medicine and Hygiene* 1988; **38**: 356–62.

47.2.7 Chyluria

A. P. Pandey

Presentation
Diagnosis
Treatment
Further reading

Chyluria is ‘milky white' urine due to the presence of chyle that enters the urinary drainage system as a result of fistulous communications with the renal lymphatics (Fig. 1). The incidence of chyluria in patients who have filariasis is about 2 per cent. Chyluria is a symptom and not a disease.



Fig. 1. Chylous urine.

Filariasis is the most common cause of chyluria (Table 1). *Wuchereria bancrofti*, a viviparous nematode, accounts for 90 per cent of human filariasis. Man is the only host. The disease is spread by bites from mosquitoes (*Anopheles* or *Culex*); adult worms migrate to the lymphatics, where they cause obstruction. It is rare in Europe but not uncommon in the tropics, particularly in areas where filariasis is common. Parasitic chyluria is geographically distributed from 30°N to 30°S. In 1991, the World Health Organization report committee on filariasis estimated that 90 million persons are infected by filariasis throughout the world and that 750 million persons are at risk. Numerically, filariasis is greatest in China, India, and Indonesia. These three countries account for about two-thirds of the estimated world total of persons infected.

Parasitic
Filariasis
Ascariasis
Malaria
Ancylostoma
Non-parasitic
Tumour
Trauma
Tuberculosis
Pregnancy
Congenital?

Table 1 Aetiology of chyluria

Presentation

Chyluria affects young males and females equally. It has periods of spontaneous remissions and exacerbation, which is often worse after a ‘fatty meal'. Chyluria, haematochyluria (the passage of chylous clots in the urine), and dysuria are common modes of presentation. The loss of protein in urine causes hypoproteinaemia. Chylous clots may produce renal colic and urinary retention.

Diagnosis

A clot is formed in the chyluric specimen of urine on standing, which, on shaking with an equal amount of ether, does not dissolve. The presence of chyle in urine is confirmed by estimation of the urinary colloidal suspension of fat in the form of chylomicrons. Oral ingestion of fat leavened with Sudan Red Three turns the urine pink in those with chyluria but not in normal individuals, a test useful in differentiating chyluria from artefact (Table 2).

Gross pyuria
Phosphaturia
Amorphous urates
Caseous material—?tuberculosis
Pseudochyluria

Table 2 Differential diagnosis of chyluria

The intravenous urogram is usually normal. Lymphangiography shows small nodes with dilated, plexus-shaped lymphatics in para-aortic regions; retrograde filling of lymphatics running towards the hilum of kidneys appears to be the site of fistulization. Intravenous urography combined with lymphangiography usually demonstrates the site of lymphaticourinary fistulas (Fig. 2).

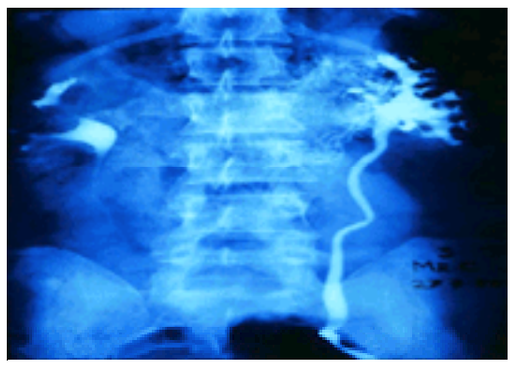


Fig. 2. Lymphangiography combined with intravenous urography in filarial chyluria.

Lymphoscintigraphy often demonstrates the lymphaticorenal communication. Technetium-99-labelled sulphur colloids are the nuclide compound used for this purpose. The isotope appears in the urine promptly and concentrates in the renal area (Fig. 3).

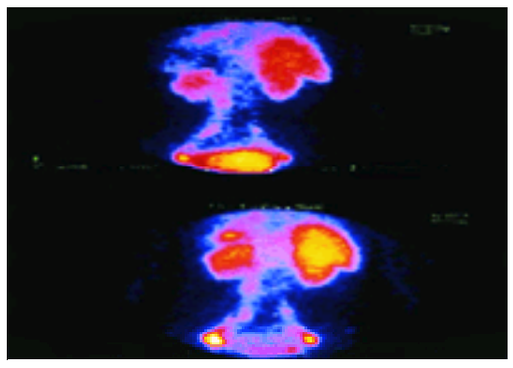


Fig. 3. Isotope studies to demonstrate lymphaticourinary fistulas, before (top) and after (bottom) treatment.

Cystoscopy during the episode of chyluria or 2 h after a ‘fatty meal' often confirms efflux of milky-white urine from one or both the ureteric orifices.

Treatment

Spontaneous remission of chyluria has been reported in about 50 per cent of patients. Conservative measures in its management include diet of low-fat and high-protein content, and large amounts of fluid intake to reduce the incidence of clot colic or clot retention. Medium-chain fatty acids may be substituted for the usual dietary fat as they are directly absorbed into the portal venous system, bypassing the intestinal lymphatics.

Passage of sclerosing agents such as 1 to 2 per cent silver nitrate solution, betadine, 15 to 25 per cent sodium iodide or potassium bromide, or hypertonic glucose or saline into the renal lymphatics through the pyelolymphatic channels, a potential communication that exists between the renal lymphatics and the pelvicalyceal system, produces lymphangitis and finally fibrosis, resulting in obliteration of lymphaticourinary fistulas. Instillation of freshly prepared 1 to 2 per cent silver nitrate solution into the renal pelvis helped to relieve symptoms in 90 per cent of our 200 patients with recurrent chyluria seen during the last 20 years.

Cystoscopy is performed 2 h after a fatty meal consisting of 10 g of butter and a cup of creamy milk. The procedure is performed with local instillation of 2 per cent lignocaine (lidocaine), except in very apprehensive individuals, who receive regional anaesthesia. A balloon ureteric catheter is passed inside the ureteric orifice, effluxing chylous urine. The position of the ureteric catheter inside the renal pelvis is indicated by the sudden gush of urine and confirmed by fluoroscopic imaging. If chyluria is bilateral the more severely affected side is irrigated first, the contralateral side being treated a few weeks later.

The renal pelvis is filled with a measured quantity of distilled water until the patient experiences flank pain. This provides a rough estimate of the renal pelvic capacity. Equal amounts of freshly prepared 1 to 2 per cent silver nitrate solution are gently pushed through the ureteric catheter into the renal pelvis. As soon as the patient feels lumbar pain, the silver nitrate is aspirated out and the renal pelvis is thoroughly lavaged with distilled water. Frusemide (furosemide) is administered to induce diuresis. The ureteric catheter is removed after clear urine starts gushing out. Before removing the cystoscope, the silver nitrate is thoroughly washed out of the bladder. Those with suspected obstruction of the bladder outflow are kept on continuous bladder drainage. Most patients feel sick and experience mild abdominal discomfort immediately or a few hours after the procedure. Those who develop transient haematuria and flank pain respond well to this treatment. Clear urine suggests disappearance of chyle, and absence of urinary chylomicrons signifies obliteration of lymphaticopelvicaliceal fistulas.

Silver nitrate solution occasionally enters the bloodstream through the pyelovenous flow and produces an anaphylactic reaction. High doses of hydrocortisone and fluids help to bring the situation under control. If silver nitrate enters the parenchymal tissue due to improper placement of the ureteric catheter or forceful introduction, cortical and perinephric abscesses develop, necessitating drainage. If there is spillage of silver nitrate through the pelviureteric junction, double-J stenting prevents ureteric stricture. Chemical cystitis, which fails to respond to anticholinergics, requires corticosteroids and repeated bladder distension to prevent bladder contracture. This procedure, besides being easy to perform, seems to be a cost-effective treatment for chyluria in endemic areas. This treatment can be repeated safely without any distortion of renal anatomy or risk to the patient.

Surgical treatment should be reserved for patients with significant problems such as persistent clot colic, clot retention, and with significant hypoproteinaemia and weight loss.

Operative procedures to disconnect the lymphatics from the kidneys involve mobilization of the kidneys, laying bare the renal vessels and upper ureter by dividing and individually ligating all the dilated lymphatics. This can be performed either by open surgical methods or laparoscopically. The patient should be advised to take a high-fat meal (50–100 g of butter) 24 h before the operation for better visualization of lymphatics. Preoperative injection of ‘patent blue' in the pedal web space is to be avoided as a tear in one lymphatic vessel would produce discoloration of tissues in the operative field, making subsequent dissection difficult. Transient renal ischaemia during dissection around the renal artery can be avoided by meticulous dissection and by administering 100 ml of 12.5 per cent mannitol intravenously at the time of hilar dissection. In the renal hilum, the dilated lymphatic vessels, which course along the blood vessels, are carefully dissected and ligated with fine sutures to bare at least 1 cm of the artery and vein.

Another operative procedure performed is the lymphovenous anastomosis, either the transinguinal spermatic or the inguinal lymph node–saphenous vein type. The principle behind this microsurgical lymphovenous anastomosis is that it may increase the drainage of lymph fluid into the venous system, which allows for a rapid decrease in intralymphatic pressure, decreased lymphangiectasis, and facilitation of sealing of the lymphatic fistula in the renal papillae.

Surgical capsulotomy and renal decapsulation have always met with poor results. Renal autotransplantation is indicated for intractable chyluria.

Chyluria commonly recurs in spite of open surgical procedures. Use of higher concentrations of silver nitrate (3–5 per cent) solution is recommended for recurrent cases.

Further reading

Diamond DE, Schapira HE. Chyluria—a review of the literature. *Urology* 1985; **25**: 427–31.

Gomella LG *et al.* Extraperitoneal laparoscopic nephrolysis for treatment of chyluria. *British Journal of Urology* 1998; **81**: 320–1.

Ji Y-Z, Zheng J-H, Chen J-N, Wu Z-D. Microsurgery in the treatment of chyluria and scrotal lymphangial fistula. *British Journal of Urology* 1993; **72**: 952–4.

Karanjaval DK. Lymphangiography in the management of filarial chyluria. *Annals of the Royal College of Surgeons of England* 1970; **46**: 267–76.

47.3 Splenic surgery in the tropics

H. Foster

Introduction
Splenomegaly
Gross splenomegaly (including tropical splenomegaly syndrome, big spleen disease or hyper-reactive malarial splenomegaly)
Splenic rupture
Traumatic rupture
Non-traumatic rupture
Splenic abscess
Splenic torsion (ectopic spleen)
Complications of splenectomy in the tropics
Further reading

Introduction

In the tropics and subtropics, splenic surgery is undertaken not only for conditions found almost exclusively in those regions but also for most splenic conditions that are also found in temperate regions. Splenomegaly, parasitic disease, haematological processes, and trauma are common and occur in various combinations. This discussion is confined to those splenic conditions likely to present to the surgeon and that occur predominantly in the tropics (Table 1).

Splenomegaly:
Moderate
Gross
Rupture:
Traumatic
Non-traumatic
Splenic abscess
Splenic torsion (hypermobile spleen)

Table 1 Surgical presentations of splenic diseases in the tropics

Splenomegaly

A moderate degree of chronic splenomegaly is common in inhabitants of rural and less developed tropical countries, especially in malarious areas. Splenic enlargement is clinically detectable in 10 to more than 50 per cent of the population in parts of Africa, South-East Asia, Melanesia, and the Pacific. Although there are regional variations in aetiology, malaria is the most common cause. Schistosomiasis and visceral leishmaniasis are common in some regions. Evaluation of the individual patient requires consideration of numerous possible diagnoses (Table 2). Moderate splenomegaly *per se* is not a common reason for presentation to the surgeon in the tropics; however, trauma to the abnormal spleen and non-traumatic rupture are important surgical problems.

Moderate splenomegaly
Parasitic, bacterial and viral infection
(e.g. malaria, schistosomiasis, leishmaniasis, trypanosomiasis, typhoid fever, brucellosis, lassa fever)
Portal hypertension
(e.g. cirrhosis, hepatic schistosomiasis, constrictive pericarditis)
Haemolytic diseases
(e.g. sickle-cell disease, thalassaemia)
Splenic abscess
Neoplasms, cysts and hamartomas (e.g. lymphoma, leukaemia, lymphangioma)
Miscellaneous (e.g. lipid-storage diseases, amyloidosis)
Gross splenomegaly (tropical splenomegaly)
Hyper-reactive malarial splenomegaly
Visceral leishmaniasis (kala-azar)
Schistosomiasis
Lymphoma/chronic leukaemia (usually Burkitt lymphoma)

Table 2 Aetiology of moderate and gross splenomegaly in the tropics

Gross splenomegaly (including tropical splenomegaly syndrome, big spleen disease or hyper-reactive malarial splenomegaly)

Splenic enlargement below the umbilicus and across the midline is certainly ‘gross'; some clinicians regard any enlargement more than 10 cm below the costal margin as gross splenomegaly. Such spleens often weigh more than 4 kg, 30 times the mass of a normal spleen (see Fig. 1, Fig. 2 and Fig. 3). A weight of 1500 g is often used as the lower limit of gross splenomegaly.



Fig. 1. A man from the North Solomon Islands with massive splenomegaly.

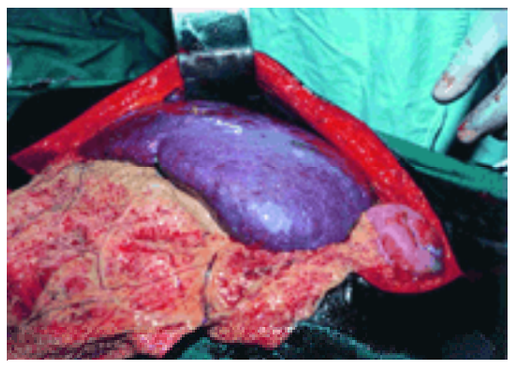


Fig. 2. A massive spleen due to hyper-reactive malarial splenomegaly at laparotomy for bleeding oesophageal varices. The midline incision extends from xiphisternum to pubis.



Fig. 3. A massive spleen weighing over 5 kg, with capsular fibrosis, omental adhesions, and depressed scars due to infarction. Splenectomy was performed for complicated hyper-reactive malarial splenomegaly.

Patients with gross splenomegaly present to surgical clinics most often complaining of a mass, abdominal discomfort, or a complication of their splenic disease. In a minority of these patients a specific cause will be found for their splenomegaly (Table 2). In areas endemic for malaria, the majority will have the features of hyper-reactive malarial splenomegaly (Table 3).

Gross chronic splenomegaly from childhood*
Found only in malarious areas*
Very high IgM and antimalarial antibody*
Response to long-term antimalarial medication*
No detectable, or low-level, parasitaemia
Higher incidence in women
Familial and racial predisposition
Mild hepatomegaly
Abdominal discomfort
General debility and loss of weight
Hypersplenism
Hypervolaemia (dilutional anaemia)
Lymphocyte proliferation (liver, spleen, bone marrow, circulation)
Scanty malarial pigment in phagocytes
Splenic capsular fibrosis and perisplenitis

*Major diagnostic criteria

Table 3 Features of hyper-reactive malarial splenomegaly

The aetiology, pathogenesis, and treatment of tropical splenomegaly SYndrome are now being clarified, principally by studies from Africa and Papua New Guinea in malarious areas. The syndrome is usually associated with an abnormal and perhaps inadequate immune response to chronic malaria in a genetically predisposed host. ‘Hyper-reactive malarial splenomegaly' is a more accurate name for this type of tropical splenomegaly. There is evidence to suggest that these patients have a genetic defect in their cell-mediated immune response to malarial infestation and consequently suffer persistent, low-level parasitaemia, which they are unable to eradicate entirely. This parasitaemia may be a constant stimulus to the immune system, causing an abnormal humoral response with very high concentrations of polyclonal IgM and malaria-specific antibody. The number of T-suppressor cells is reduced. Circulating immune complexes are sequestered by splenic and hepatic phagocytes. Individuals and populations newly exposed to endemic malaria are especially at risk.

Management of hyper-reactive malarial splenomegaly

Diagnosis

In a region where malaria is endemic, hyper-reactive malarial splenomegaly is the most likely cause of gross splenomegaly in children or young adults. Definite diagnosis depends on the exclusion of the other causes of gross splenic enlargement and demonstration of the features of the condition, especially high concentrations of IgM and specific antimalarial antibody, in the absence of acute parasitaemia.

Infectious and neoplastic causes of gross splenomegaly must be actively excluded. Routine blood count and thick films for malaria parasites, needle biopsy of the liver, and endoscopy for oesophageal varices may be helpful.

Complications

Complications of hyper-reactive malarial splenomegaly are listed in Table 4. Sudden death, presumably due to overwhelming malarial or bacterial infection, is the most common complication of the untreated condition.

Sudden death, presumably due to malaria or sepsis
Hypersplenism (especially anaemia)
Portal hypertension (mainly a high-flow effect)
Oesophageal varices (haematemesis and malaena)
Splenic infarct
Splenic 'abscess'
Severe abdominal pain
Leg ulcers
Possible malignant transformation to lymphoma
Splenic rupture (uncommon)

Table 4 Complications of hyper-reactive malarial splenomegaly

Non-surgical treatment

Untreated hyper-reactive malarial splenomegaly has a high mortality, and, although controlled trials of therapy have not been performed, long-term antimalarial medication (and reduced exposure to malaria infestation, where possible) is the current basis of treatment. Splenic size has been shown to diminish in about 70 per cent of patients after 1 year of treatment, with 25 per cent of spleens no longer palpable after this time. Very large, scarred spleens may remain palpable after several years of antimalarial therapy. The adverse features of hyper-reactive malarial splenomegaly diminish as splenomegaly decreases, but relapse if medication is ceased and exposure to malaria persists. Lifelong antimalarial prophylaxis is indicated.

Immunosuppression has been used to treat patients in Africa and Vietnam, but the treatment is hazardous and the results uncertain.

Surgery in hyper-reactive malarial splenomegaly

In patients with uncomplicated hyper-reactive malarial splenomegaly, splenectomy seems to produce initial symptomatic improvement, with correction of anaemia, IgM, and hypersplenism, and with an acceptable operative mortality. However, subsequent mortality rates, apparently due to overwhelming sepsis or malaria, are very high. Although only small numbers of patients have been monitored over a long period, there appears to be a mortality rate of around 30 per cent in the first 12 months after splenectomy. In addition, splenectomy does not seem to alter the underlying immunological defect and lifelong antimalarial medication is needed to prevent relapse

Splenectomy seems indicated in the minority of patients who do not respond to prolonged antimalarial medication and have persistent gross splenomegaly, hypersplenism, abdominal pain, or general debility. Specific complications such as splenic abscess, or suspected malignant change may also necessitate splenectomy.

Splenectomy may also have a place in the management of portal hypertension and bleeding oesophageal varices due to hyper-reactive malarial splenomegaly. The pathogenesis of the raised portal pressure in these patients is not entirely clear, but greatly increased splenic blood flow is certainly a major factor. Patients with hyper-reactive malarial splenomegaly usually do not have cirrhosis or portal fibrosis. The small number of dynamic studies of portal flow performed in these patients have produced varying results, but no convincing evidence of significantly increased hepatic vascular resistance. This correlates with the usual histological finding of sinusoidal lymphocytosis, which would seem unlikely to produce increased hepatic resistance to portal blood flow. Ligation of the splenic artery or splenectomy to reduce excessive splenic venous flow, in conjunction with local obliteration of oesophageal varices, has proved effective in treating patients with massive haematemesis associated with hyper-reactive malarial splenomegaly (Fig. 2 and Fig. 3).

Perioperative antibiotics and antimalarials should be given, and, following splenectomy, patients require lifelong supervision of antimalarial therapy and precautions against overwhelming sepsis. Their long-term management is fraught with major logistic and compliance problems in most developing countries.

Some technical aspects of splenectomy for gross splenomegaly

Blood transfusion should be available and good anaesthesia is essential. Autotransfusion may be appropriate.

Although various incisions are used for splenectomy, a long upper midline incision generally provides the best access, as the mobilized spleen swings into the midline after division of the lienorenal peritoneal reflection, allowing good access to the splenic hilum and short gastric vessels. This incision also allows the splenic artery to be approached through the lesser sac and ligated at the upper border of the pancreas, if this seems advantageous. Most very large spleens can be mobilized by freeing the splenic flexure of the colon and wide division of the peritoneum along the lienorenal reflection. Vascular adhesions between the diaphragm and the spleen are fortunately not common: when present they may constitute a major technical problem that requires good exposure of the left hemidiaphragm; this may be achieved by use of a substernal or costal margin retractor. In exceptional cases, a lateral extension of the midline incision or even left thoracotomy may be helpful. Some surgeons have found that a bilateral subcostal incision provides good access in difficult cases.

Once the spleen has been mobilized into the wound, the hilum can be controlled by digital pressure, and individual vessels isolated and ligated. Closed suction drainage of the splenic bed should be used where technical or haematological factors indicate a risk of postoperative haemorrhage or collection.

Splenic rupture

In the tropics, most spleens ruptured are abnormal; the majority are moderately enlarged by malaria, but other causes of moderate splenomegaly can be associated with rupture (Table 2).

Traumatic rupture

Reliable studies are few and clinical impressions variable, but most clinicians agree that traumatic rupture of the spleen is common in the tropics, particularly in children and young adults whose spleen is moderately enlarged and friable due to malaria. Isolated rupture of the spleen may be more frequent in rural tropical communities where low-energy injuries are more common and the enlarged spleen more vulnerable. This contrasts with the high-energy injuries seen in developed countries, where rupture of the normal, well-protected spleen is more frequently associated with multiple visceral and musculoskeletal injuries.

The importance of preserving splenic function for both children and adults living in the tropics is now well recognized, just as it is in more developed countries. Asplenic patients have an increased risk of overwhelming bacterial sepsis, especially due to pneumococcal or Gram-negative infection, as well as undue susceptibility to severe or fatal malaria. They may also be at increases risk of other protozoal infections and schistosomiasis. Sudden, unexplained death months or years after splenectomy for trauma is not uncommon in the tropics.

The key to management of patients with traumatic rupture of the spleen in the tropics is preservation of the organ if at all possible. This is often possible without surgery, as the incidence of associated injury, persistent haemorrhage, or delayed complications requiring laparotomy is low. Adequate facilities for resuscitation, diagnosis, prolonged accurate observation, and timely surgical intervention are essential for the proper management of patients with a ruptured spleen. These facilities are not always available in developing countries.

Diagnosis is based on clinical findings and confirmation of haemoperitoneun by lavage or ultrasonography if necessary, performed whilst vigorous resuscitation with intravenous fluids and antimalarials is under way. Computed tomography may provide a more detailed image of the injury once the patient is stabilized. Laparoscopy may have a role to confirm haemoperitoneum and detect persistent haemorrhage or injury of other abdominal organs; intraperitoneal blood may be aspirated for autotransfusion.

Patients with continuing haemorrhage or injury of other abdominal organs may require laparotomy in a minority of cases; splenic preservation by repair or partial splenectomy should be successful in about 50 per cent of these cases. The spleen of a patient suffering acute malaria may be enlarged, vascular, and friable, making suture repair technically difficult. The removal of devitalized spleen and the application of haemostatic agents or omental flaps are nevertheless effective in many cases. Major injuries involving hilar vessels usually require splenectomy, as may the need for surgical expediency.

Complications

The early complications of the non-operative management of a ruptured spleen include failure to detect continuing splenic haemorrhage and associated injury requiring laparotomy. In the longer term, delayed rupture of a subcapsular haematoma or post-traumatic cysts may occasionally follow splenic preservation and should also be managed by conservative surgery, such as capsular repair, partial splenectomy, or unroofing and drainage of cysts, whenever possible.

Intraperitoneal implantation of splenic fragments does not seem to provide asplenic patients with protection from severe infection or malaria, and the procedure is no substitute for appropriate lifetime prophylaxis.

Non-traumatic rupture

Non-traumatic rupture of an entirely normal spleen arguably does not occur in the absence of a coagulopathy, and even the diseased spleen rarely ruptures spontaneously. Subcapsular haemorrhage in an acutely inflamed, enlarged, and congested spleen is probably the most common mechanism of pathological rupture.

Trivial trauma may have been forgotten or disregarded.

Acute malaria is the most common cause of non-traumatic splenic rupture in the tropics. The incidence is highest during initial attacks of malaria, but even then it is rare. The chronically enlarged spleen may undergo a degree of capsular fibrosis, which protects it somewhat from rupture.

Infectious causes of spontaneous rupture, such as salmonella, Epstein–Barr virus and hepatitis virus, probably occur more commonly in the tropics, while haematological and neoplastic causes have a similar incidence in tropical and temperate areas.

As with traumatic rupture, the spleen that undergoes ‘spontaneous' rupture in the tropics should be preserved if at all possible.

Splenic abscess

In certain parts of Asia, Africa, and Papua New Guinea there is a condition of uncertain aetiology known as ‘ primary or idiopathic splenic abscess'. This condition is certainly not a true abscess, as the fluid obtained from these lesions is always sterile and may contain cholesterol crystals similar to those seen in the fluid from branchial cysts. The disease is uncommon and occurs mainly where malaria and sickle-cell anaemia coexist. It is less frequent in areas where malaria occurs without sickle-cell disease, such as Papua New Guinea. The condition is distinct from true pyogenic or fungal splenic abscess, from which organisms are cultured in almost all cases. Most pyogenic abscesses are associated with a distant site of infection and presumed haematogenous spread, while a minority follow splenic trauma or direct spread from an adjacent site. True splenic abscesses occur with a low incidence throughout the world, whereas sterile ‘abscesses' closely linked to malaria and sickle-cell anaemia are more common in the tropics.

Sterile or tropical splenic abscess usually has the features of a liquefied splenic infarct and presumably follows occlusion of a segmental artery by aggregations of sickled erythrocytes, or thrombosis in acute malaria. The condition is rare and diagnosis requires an alert clinician. Protracted fever, left upper-quadrant pain, and a mass should suggest the diagnosis. Erect chest and abdominal radiographs may show a raised left hemidiaphragm with a soft tissue mass and an air–fluid level due to gas in the abscess. Contrast radiography of the stomach or left kidney may show displacement of these organs, but ultrasound scan or computed tomography, if available, are the most useful means of diagnosis. Radioactive isotope scanning and selective angiography have also been used.

Treatment requires resuscitation, with antimalarial and usually antibiotic cover. Timely splenectomy to avoid rupture of the abscess has been advocated, but drainage of the abscess with splenic preservation is now preferred in most cases. Drainage at laparotomy is usually the safest and most reliable technique, but percutaneous and laparoscopic drainage should be considered for selected cases. A team approach by surgical and imaging personnel may be appropriate.

Splenic torsion (ectopic spleen)

Elongation of the splenic pedicle and increase in its mass predispose the organ to torsion and subsequent infarction. This rare condition occurs in multiparous women and in children. An elongated pedicle without torsion results in an ectopic or ‘wandering' spleen. This is thought to be congenital and due to a persistent dorsal mesogastrium in children, but may be acquired in multiparous women.

Diagnosis of acute splenic torsion requires some clinical awareness, and is often made only at laparotomy. Splenectomy is indicated if the spleen remains non-viable after untwisting the pedicle, and may be necessary if the mobile organ cannot be safely fixed in the left upper quadrant. Splenic preservation has been achieved by plication of the elongated pedicle with fixation of the spleen to the abdominal wall or left hemidiaphragm.

Complications of splenectomy in the tropics

All the well-known effects and risks of splenectomy, such as haematological changes, injury to adjacent organs, haemorrhage, left lower-lobe collapse, and subphrenic collection, are encountered in the tropics. In addition, acute malaria is common immediately after splenectomy, especially when blood transfusion is used. Infestation of donated blood is common in endemic areas and the infective dose of parasites is small. The risk of acute and chronic malaria after splenectomy in patients with hyper-reactive malarial splenomegaly is probably increased by the genetic defect these patients have in their immune response to the parasite. A therapeutic course of an antimalarial such as chloroquine is indicated perioperatively in all patients undergoing splenectomy; prophylaxis is then necessary for life in malarious regions to reduce the risk of fatal malaria. Asplenic patients from malaria-free countries should be warned against travel to malarious areas, even with appropriate chemoprophylaxis.

Overwhelming babesiosis has been reported as a rare long-term complication of splenectomy.

The incidence of overwhelming bacterial sepsis after splenectomy may be greater in the tropics than temperate zones, and the mortality rate also seems higher, perhaps due to delays and deficiencies in treatment. The relative incidence of the different causative organisms in the tropics is not known. Pneumococcal and other antibacterial vaccines may offer partial protection, but are no substitute for preservation of at least a portion of normally perfused spleen whenever technically possible.

Further reading

Adoloye A, ed. *Davey's companion to surgery in Africa*, pp. 299–308. Churchill Livingstone, London, 1987. [An excellent overview of splenic surgery in Africa.]

Bates I, Bedu-Addo G, Bevan DH, Rutherford TR. Use of immunoglobulin gene rearrangements to show clonal lymphoproliferation in hyper-reactive malarial splenomegaly. *Lancet* 1991; **337**: 505–7. [A genetic study of the pathogenesis of gross malarial splenomegaly.]

Carswell JW. Wandering spleen: 11 cases from Uganda. *British Journal of Surgery* 1974; **61** : 495–7. [A large series of this rare condition.]

Cockshott WP, Weaver EJM. Primary tropical splenic abscess: a misnomer. *British Journal of Surgery* 1962; **49** : 665–9.[The probable pathogenesis of primary splenic abscess.]

Crane G. Recent studies of hyperreactive malarious splenomegaly (tropical splenomegaly SYndrome) in Papua New Guinea. *Papua New Guinea Medical Journal* 1986; **29** :35–40. [An important study on the pathogenesis of malarial splenomegaly.]

Fakunle YM. Tropical splenomegaly syndrome. 1. Africa. *Clinical Haematology* 1981; **10** : 963–75. [A very good overview of the condition.]

Fakunle YM, Greenwood BM. Mortality in tropical splenomegaly syndrome. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 1980; **74** : 419. [One of the few attempts to assess mortality in tropical splenomegaly.]

Foster HM. The management of ruptured spleen in a malarious area—does laparoscopy have a role? *Asian Journal of Surgery* 1988; **11** : 35–8. [Laparoscopy may be useful in assessing the patient with ruptured spleen where less invasive techniques are not available.]

Gibney EJ. Surgical aspects of malaria. *British Journal of Surgery* 1990; **77** : 964–7. [A good overview of malaria for surgeons.]

Papua New Guinea Splenic Injury Study Group. Ruptured spleen in the adult: an account of 205 cases with particular reference to non-operative management. *Australia and New Zealand Journal of Surgery* 1987;**57** : 549–53. [A key study on the management of ruptured spleen in the tropics.]

Schaefer KU, Khan B, Gachihi GS, Kager PA, Muller AS, Verhave JP, McNeill KM. Splenomegaly in Baringo District, Kenya an area endemic for visceral leishmaniasis and malaria. *Tropical and Geographical Medicine* 1995; **47:** 111–14. [The epidemiology of splenomegaly in an area endemic for at least two causative parasites.]

47.4 Obstetric vesicovaginal fistulas

Gordon Williams, Ambaye Michael Wolde

Introduction
Aetiology
Prevalence
Presentation
Investigation
General
Specific investigation
Management
Surgical technique
Abdominal repair
Postoperative management
Results
Long-term management
Further reading

Introduction

Vesicovaginal fistulas are rare in developed countries. When they do occur over 90 per cent result from pelvic surgery and the remainder from cancer or the treatment of it by radiotherapy. In developing countries, over 95 per cent of fistulas are of obstetric aetiology. They are often associated with other vaginal pathology such as urethrovaginal fistulas and rectovaginal fistulas. Not only are they a major surgical challenge, but also an often forgotten social problem.

Aetiology

In developing countries the majority of vesicovaginal fistulas result from complications of a neglected, prolonged, obstructed labour. The mother, often in her early teens (Table 1), delivers without medical assistance attended only by a traditional birth attendant. Villages are isolated and often many days walk from the nearest road and even further away from a hospital.

Age (years)	*No. of patients*	*%*
< 15	18	3
16–20	203	3.3
20–30	106	17.4
> 30	146	24

Table 1 Age of mothers

During normal labour the bladder is displaced upwards, the anterior vaginal wall, bladder base, and urethra are compressed between the fetal head and the posterior surface of the pubis. In a prolonged obstructed labour pressure necrosis of the anterior vaginal wall and underlying bladder neck occurs, though sometimes the area of necrosis is more extensive and the urethra, trigone, and anterior lip of the cervix are involved. Compression of the soft tissues posteriorly against the sacral promontory results in necrosis of the posterior vaginal wall and the underlying rectum. If the mother survives, a macerated fetus is usually expelled 3 to 4 days later and the slough from the devitilized bladder and rectum around 10 days. It is at this time that she becomes incontinent. In a personal series of 609 patients with vesicovaginal fistulas, operated on by Dr Ambaye at the Fistula Hospital in Addis Ababa, 55 per cent of her patients were primiparous, 98 per cent had delivered a dead baby, and 48 per cent had been divorced by their husband. Though the majority of patients are primiparous, vesicovaginal fistulas can occur in patients having had more than 10 previous deliveries.

In some areas of Africa where female mutilation (circumcision) is still common, the deliberately narrowed vaginal introitus results in obstructed labour. Traditional practices in other areas of Africa include cutting of young women with a knife, piece of glass, or a razor blade in an anterior direction to enlarge the vagina. This invariably results in a urethrovaginal fistula and division of the bladder continence mechanism. In Ethiopia the commonest cause of a rectovaginal fistula in a young girl is a result of vaginal intercourse at a very young age.

Prevalence

In the developing world many fistula sufferers are unknown to the medical services. The Addis Ababa Fistula Hospital in Ethiopia, opened in 1975 as a charity hospital to deal specifically with women with vesicovaginal and rectovaginal fistulas, treated 603 women in 1987 and over 1200 in 1997. It is estimated that there are 500 000 untreated cases worldwide.

Presentation

In women who have been cut or develop a fistula as a result of surgery, incontinence may develop from day 1. However, in women who develop a fistula as a result of prolonged obstructed labour (Table 2), the incontinence develops after separation of the necrotic tissues around 10 days. The time to presentation at hospital, however, varies enormously and ranges from less than 3 months to 45 years. On average it is 4 to 9 months.

No. of days	*No. of patients*	*%*
1	133	1.4
2	211	22.8
3	104	36
4	59	17.8
5	32	10.1
6	31	5.5
7	31	5.3
8	5	1.0

Table 2 Duration of labour

Investigation

General

Women who present early have often undergone prolonged periods of pelvic sepsis, they are malnourished, anaemic, and often infested with intestinal parasites. Others will have untreated malaria or tuberculosis and HIV infection is becoming more prevalent.

In women with long-standing fistulas, local infections have usually resolved but the other medical conditions are still present and require treatment before embarking on surgical closure of the fistula.

Specific investigation

The injuries are usually obvious on examination of the vagina using a speculum. When the diagnosis is in doubt the identification of the site of the fistula is best carried out by the instillation of a coloured dye (methylene blue) into the bladder via a urethral catheter with the patient in the lithotomy position and the operating table tilted head down. It is important to ensure that the bladder is fully distended, but also important to ensure that dye does not leak back around the catheter giving the impression of a fistula. The traditional three-swab test is unnecessary and may mask the presence of multiple fistulas. Direct examination of the vagina using a Simm's speculum is usually all that is required. If leakage of clear fluid continues after the bladder has been filled with blue dye, a ureterovaginal fistula is likely. These are uncommon in obstetric fistulas and usually result from previous gynaecological or pelvic surgery. Blue dye presenting from the uterine os is indicative of a uterovaginal fistula, usually following an emergency subtotal hysterectomy for a ruptured uterus. Fortunately, additional radiological investigations, which are usually not available, are unhelpful. Examination under anaesthesia may be required to determine the extent of injury and the available access for a vaginal repair, particularly in those with excessive scarring.

Management

Hospitals which deal with large numbers of women with vesicovaginal fistulas have set up programmes not only to look after the physical needs of their patients but also their emotional, social, and educational requirements.

Patients presenting early (within 3 months) should be treated with appropriate antibiotics to prevent prolonged sepsis and promote tissue healing. Spontaneous healing after a period of bladder drainage, though rare, is well recognized and a trial of catheterization for 7 to 10 days in patients presenting early with small fistulas is justified. Operative repair should not be attempted until at least 3 months has passed. The patient's general condition can be improved and the infection treated, which allows the fistula to become smaller, revascularization of the ischaemic area occurs, and tissue planes are more easy to identify than at subsequent surgery. Repair should not be undertaken until the vaginal mucosa has fully healed and re-epithelialized. Patients who develop bladder stones should have these removed and at least 4 to 6 weeks be allowed to pass to allow the inflamed bladder muscosa to heal.

Traditional therapies to treat vesicovaginal fistulas include strapping the legs together for prolonged periods. This results in contractures of the hips or knees. In addition many women rejected by their families curl up and remain in a fetal position for long periods. Treatment of these contractures is required before surgical exposure to the vagina can take place.

Surgical technique

Most urologists prefer the abdominal approach. The gynaecologists at the Fistula Hospital in Addis Ababa have used the vaginal approach in over 18 000 repairs. Though the fistula may often appear small and easy to close, this is due to tethering of the tissues by scar tissue. There is no role for simple suture. The vagina and bladder must be widely separated. This dissection is helped by the infiltration of a 1:200 000 solution of adrenaline or 10 units of oxytocin diluted in 10 ml saline into the tissue surrounding the fistula. The repair should be tension free and closed in layers using absorbable sutures such as 2.0 chromic catgut or 3.0 vicryl for the bladder and nylon for the vagina. Prior to closure of the fistula, the ureteric orifices should be identified and ureters close to the site of the fistula (65 per cent) protected by the passage of ureteric catheters (Fig. 1), which can be brought out through the urethra. The repair should be started at either end working towards the midline, using interrupted sutures inverting the bladder mucosa. The second layer strengthens the repair by suturing the bladder wall. At this stage the repair should be tested by filling the bladder with a methylene blue solution. A third layer of interrupted mattress sutures is used to evert and close the vaginal wall. If, prior to closure of the vagina, the bladder wall is oozing and there is a large dead space, the vagina should be closed with interrupted catgut, picking up the posterior bladder wall.

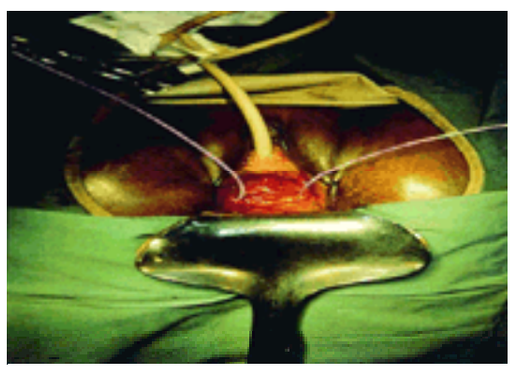


Fig. 1. Vesicovaginal fistulas with ureteric catheters in place.

In patients with large fistulas, or where the urethra and bladder neck has been involved, a Martius graft should be used (Fig. 2). A vertical incision is made for approximately 8 cm over one labia majorum and the labial fat pad mobilized, passed subcutaneously, and anchored with sutures to cover the second layer of the bladder repair. This not only helps vascularize the repair but also maintains competence of the closure mechanism by reducing scarring.



Fig. 2. A Martius graft being drawn into the vagina. Note the ureteric catheters exiting through the urethral meatus.

In patients with multiple fistulas, unless they are widely separated, the fistulas should be converted into one and repaired. If the fistulas are wide apart, two separate repairs should be carried out on the same occasion.

Urethral fistulas are closed by dissecting the paraurethral tissues and suturing them over a urethral catheter in two layers. A Martius graft is essential in this situation.

Abdominal repair

A combined transperitoneal and transvesical procedure is useful for high fistulas. The bladder is opened in the midline and extended downwards to circumcise the fistula. The fistulous track is excised, the vagina is closed in one layer and the bladder in two, again using absorbable interrupted sutures to evert the vaginal mucosa and invert the bladder mucosa. The ureters should be protected prior to closure by the insertion of ureteric catheters. The bladder and vagina should be separated by the interposition of either an omental pedicle graft rotated down into the pelvis on the right gastroepiploic artery or, alternatively, a flap of peritoneum from any available surface can be used and interposed between the bladder and vaginal repair.

Postoperative management

Continuous bladder drainage via a urethral Foley catheter is essential. In patients with a fistula of the bladder neck the balloon should not be inflated, the catheter sutured in place. Open drainage into a kidney dish is successful and associated with very low infection rates and is used in many developing countries due to the non-availability and cost of closed drainage systems. The catheters should be checked hourly during the day and night to confirm free drainage. If urine bypasses the catheter, the catheter should be gently flushed to relieve the blockage. The period of catheter drainage varies according to the type of fistula. With most obstetric fistulas a period of 14 days is adequate. If they leak during these 14 days, the catheter should be left in for 3 weeks.

Though in theory these patients are at significant risk from the development of deep vein thromboses and pulmonary emboli, such an occurrence in developing countries is exceedingly rare. Compression stockings or subcutaneous heparin are usually not available.

Results

At the Fistula Hospital in Addis Ababa, 91.4 per cent of patients with vesicovaginal fistulas were cured of their fistula at their first operation. However, 25 per cent have some degree of incontinence due to persisting stress or urge incontinence. Some patients present with their bladder almost totally destroyed and require treatment by either an enterocystoplasty or urinary diversion. Repair of the fistula is impossible. In other patients it is impossible to identify or reimplant a ureter embedded in scar tissue in the side wall of the pelvis. This results in persistent leakage from the ureterovaginal fistula and not from the treated vesicovaginal fistula.

Because of vaginal stenosis or scarring at the site of the repair, conventional techniques to treat stress incontinence with physiotherapy or sling procedures are unsatisfactory. The injection of urethral bulking agents, which in theory would be effective, are too expensive as would be the use of an artificial urethral sphincter.

Long-term management

Patients are advised to refrain from penetrative sexual intercourse for 3 months. When they become pregnant they are advised to go to the nearest hospital when they feel the baby walking around inside them, that is, around 5 months. (Most women in developing countries have no idea of the date of their last menstrual period.) Though some women who have been cured of a vesicovaginal fistula have a subsequent normal vaginal delivery, they will all invariably develop further vaginal fistulas following a second, third, or fourth pregnancy. They are all advised to have a caesarian section.

Further reading

Hamlin RHJ, Nicholson EC. Reconstruction of urethra totally destroyed in labour. *British Medical Journal* 1969; **ii**: 147–50. [First description of the utilization of the labia minora to reconstruct the urethra.]

Mundy AR. Fistulae. In: Whitfield HN, ed.,*Robin Smith's operative surgery. Genito-urinary surgery*, Vol. 1, 5th edn, pp. 385–91. [Excellent description of transabdominal repair of vesicovaginal fistulas.]

Waaldijk K. *Step by step surgery of vesicovaginal fistulas*. Campion, Edinburgh,1994. [Clearly written, easy to follow guide to transvaginal repair of vesicovaginal fistulas.]

Zacharin RF. *Obstetric fistula*. Springer-Verlag, Vienna, 1988. [Text book covering all aspects of obstetric fistulas.]

47.5 Common surgical eye conditions in the developing world

Gordon J. Johnson and Allen Foster

Introduction
The common blinding conditions
Age-related cataract
Glaucoma
Trachoma/entropion
Ocular leprosy/exposure keratitis
Corneal opacity
Optical iridectomy
Keratoplasty
Pterygium excision
Childhood blindness
Vitamin A deficiency/xerophthalmia
Congenital cataract
Congenital glaucoma
Retinoblastoma
Removal of the eye
Enucleation
Evisceration
Further reading

Introduction

Estimates based on data available in the World Health Organization projected for 1998 suggest that approximately 48 million people in the world are now blind, with a visual acuity in the better eye of less than 3/60. The prevalence of blindness in industrialized countries is in the range 0.05 to 0.3 per cent, and in developing countries 0.5 to 2 per cent, while age/sex standardized rates show blindness to be at least 10 times more common in developing countries. This visual loss represents a massive public health and economic problem in many developing countries. The number of blind people is increasing because of population growth and longevity. It is estimated, however, that 70 to 80 per cent of all blindness is ‘avoidable', that is, can be prevented or cured with resources that could reasonably be made available.

Estimates of prevalence of blindness and numbers of blind for different continents are shown in Table 1, and causes of worldwide blindness are shown in Table 2.

Region	Population (millions) (1998)	Prevalence of blindness (per cent)	Numbers of blind (millions)
Asia	3638	0.85	31
Africa	680	1.3	8
Latin America	520	0.5	3
Europe/Russia/Oceania and North America	1360	0.3	4
Total	5982	0.8	48

Table 1 Estimates of prevalence of blindness and numbers of blind for different continents (1998)

Cause	*Millions*
Cataract	24.0
Glaucoma	6.7
Trachoma/corneal scar	5.5
Diabetic retinopathy	3.0
Childhood blindness	1.5
Trauma	1.5
Onchocerciasis	0.5
Others	5.3
Total	48.0

Table 2 Causes of worldwide blindness estimated in general terms (1998)

In this section, the discussion will emphasize the role of eye surgery in preventing blindness, particularly the situation as it applies to Asia and Africa. Reference will not be made to retinal detachment, vitreous surgery, and orbital exploration, which usually affect only one eye, or to squint, lacrimal, and plastic surgery, which treat conditions that usually do not result in blindness.

The common blinding conditions

Age-related cataract

Cataract (Fig. 1) is far and away the most common cause of blindness in the world, and cataract extraction is the most common operation on any eye service. Worldwide at least 24 million people are blind from cataract and approximately 10 million cataract extractions were performed in 1997. It is only recently that the very large number of patients requiring cataract extraction has been appreciated.

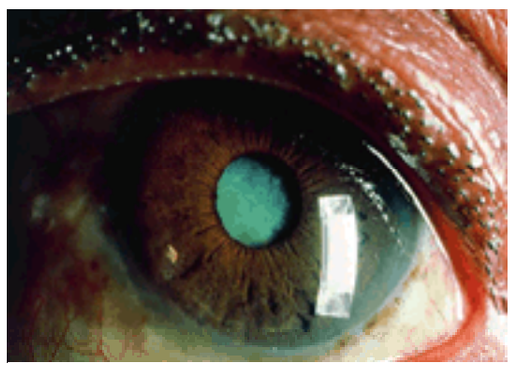


Fig. 1. Dense, age-related cataract (by courtesy of J.D.C. Anderson).

The indications for cataract extraction have changed rapidly in the industrialized world, with the widespread use of intraocular lenses resulting in surgery at levels of only moderately reduced vision (6/12–6/36), but in the developing countries the usual indication for surgery is a vision of less than 6/60 in the better eye. Examination includes a measurement of visual acuity, and if this is less than hand movements, the ability of the eye to perceive the direction from which a light is projected indicates that the peripheral retina is intact. A normal pupil light reflex is evidence that the optic nerve is functional. A second indication for cataract extraction is a swollen, cataractous lens (Fig. 2), which may cause secondary glaucoma if not removed.

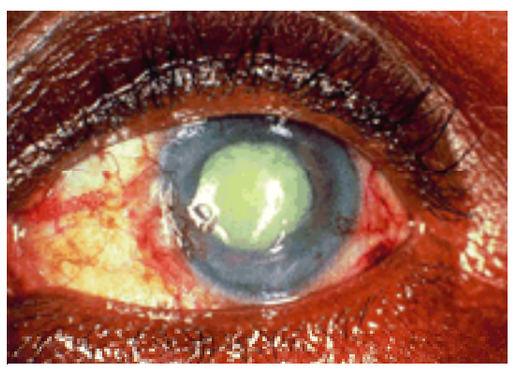


Fig. 2. Swollen (intumescent) cataract giving rise to secondary glaucoma, an indication for cataract extraction (by courtesy of S. Franken).

There are two main methods of cataract extraction: intracapsular extraction, when the lens is removed intact within its capsule, and extracapsular extraction, when the central part of the anterior capsule is removed, the nucleus expressed, and the cortex of the lens aspirated. In extracapsular extraction, the posterior and peripheral parts of the capsule bag remain attached to the ciliary body by the zonules. Phakoemulsification is a particular type of cataract extraction where the lens material is fragmented with an ultrasound probe and then removed through a small incision.

Anaesthesia

Cataract extraction is most commonly done under local anaesthesia. Akinesia of the orbicularis oculi muscle is produced by injection of 4 to 5 ml of 2 per cent lignocaine (lidocaine) or 0.5 per cent bupivicaine near the main trunk of the facial nerve as it runs over the neck of the mandible, in front of the upper part of the ear lobe (O'Brien block) (Fig. 3(a)). Alternatively, 5 ml may be injected from the lateral angle of the eye into the orbicularis muscle above and below the orbit (Van Lint block) (Fig. 3(b)). The nerves running from the apex of the orbit to the globe (sensory) and the extraocular muscles (motor) are paralysed with a retrobulbar injection of 2 ml of the local anaesthetic (Fig. 4). Many surgeons prefer a needle of 3 cm in length, so that the point rests in the muscle cone not too close to the apex of the orbit. If the tip is slightly rounded the chances of penetrating an orbital vein are reduced. After the injection, pressure is applied to the eye for 10 min in order to decompress the vitreous and reduce the intraocular pressure before operation.

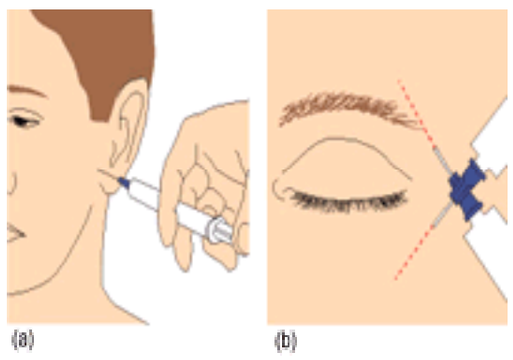


Fig. 3. (a) Injection of facial nerve in front of ear lobe (O'Brien block); (b) infiltration of terminal branches of facial nerve in lids (Van Lint block).

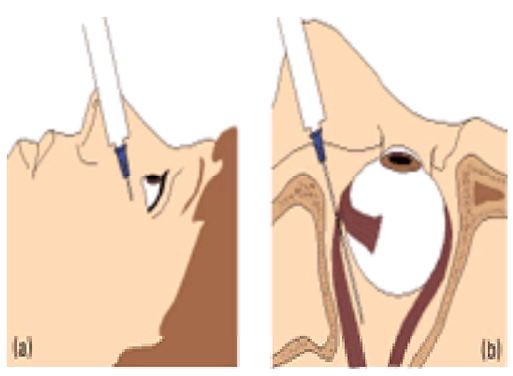


Fig. 4. Retrobulbar injection.

Peribulbar injection of anaesthetic has recently become popular because it reduces the risk of potential problems from retrobulbar injection such as retrobulbar haemorrhage, damage to the optic nerve, or injection into the dural sheath. In the peribulbar method the needle is advanced only as far as the equator of the globe, where approximately 2 ml is injected outside the muscle cone (Fig. 5). A further 1 ml is injected behind the orbital septum as the needle is withdrawn, and another 1 ml in the tissues of the lid in front of the septum. Intermittent pressure is applied to the orbit. In the original description, two injections were made, one superonasally and the other inferolaterally. Some surgeons now find that a single injection is adequate, either adjacent to the medical canthus or inferiolaterally. Mixing 10 000 to 20 000 units of hyaluronidase with the anaesthetic aids its dispersion through the orbit.

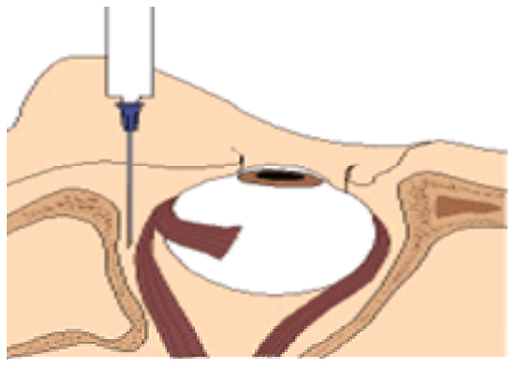


Fig. 5. Peribulbar injection.

Intracapsular cataract extraction

Conventional intracapsular cataract extraction is still the most widely practised technique in many parts of Asia and Africa. Some form of magnification (operating microscope or loupe) and good, focused lighting is recommended. Many different methods are possible at each stage; the following outline is one reliable approach (Fig. 6). The illustrations of cataract operations are as viewed by the surgeon from the head of the bed.

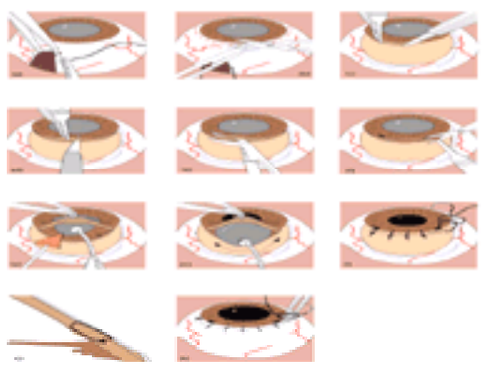


Fig. 6. Intracapsular cataract extraction.

1. *Lids and fixation*. The eyelids are retracted by a speculum or lid sutures. A retraction suture of 4/0 silk is usually inserted behind the insertion of the superior rectus muscle.
2. *Corneoscleral section*. The original von Graefe *ab interno* incision with a cataract knife is quick and reliable in the hands of those trained to use it. A small, limbus-based, conjunctival flap is usually fashioned by the same incision. Alternatively an *ab externo* incision with a razorblade fragment and/or corneal scissors may be used. The corneoscleral junction is first exposed by making a conjunctival flap based at the fornix or at the limbus.
3. *Sutures*. One to three sutures are usually placed before the corneoscleral incision is completed or are inserted immediately after the section is complete. Suture material is commonly 8/0 silk, but may be 9/0 or 10/0 monofilament nylon.
4. *Peripheral iridectomy*. One or two peripheral iridectomies are performed to prevent pupil-block glaucoma due to the vitreous face coming forward after the removal of the lens. If there is any corneal opacity or adhesions of the iris to the lens, a complete sector iridectomy may be performed instead.
5. *Removal of lens*. This is most commonly done with a cryoprobe (Fig. 7). Reliable results can be obtained with a disposable cryoprobe using Freon from a small cylinder. The probe is applied to the upper pole of the lens while the iris is retracted with a cellulose sponge or iris retractor. Alternative methods utilize lens capsule forceps or an erisophake to extract the lens.

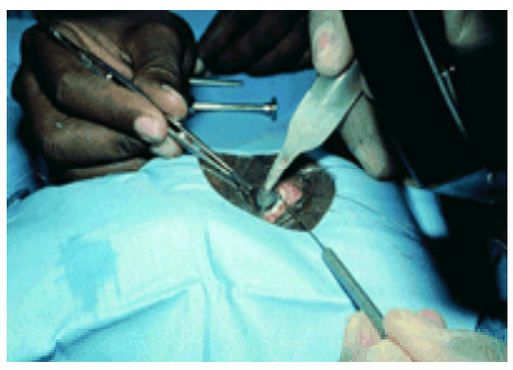


Fig. 7. Intracapsular cataract extraction using a disposable cryoprobe.

6. *Closure*. Any iris tissue in the wound is reposited, the sutures already placed are tied, and additional sutures added to ensure that the wound is watertight.
7. *Anterior chamber*. The anterior chamber may be reformed with fluid if necessary. Air has been used, but has the disadvantage that it may go behind the iris and cause a shallow anterior chamber with possible glaucoma.
8. *Complications*. The most common intraoperative complication is loss of the vitreous as the lens is removed. Using a vitrectomy machine if available, or a series of cellulose sponges and scissors, the anterior vitreous must be removed from the anterior chamber so that the iris falls back; any vitreous in the wound must be removed.

If the capsule breaks (accidental extracapsular extraction) during surgery, then the nucleus should be expressed and as much cortical lens material aspirated from the eye as is safely possible.

Early postoperative complications include rupture of a corneoscleral suture, with prolapse of iris (Fig. 8). The repair consists of excision of the prolapsed iris and resuture of the wound. Other complications are endophthalmitis (infection), iritis (inflammation), hyphaema (bleeding), glaucoma (raised intraocular pressure), and striate keratitis (damage to corneal endothelium).

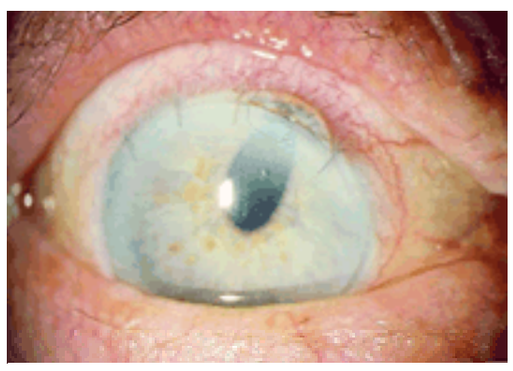


Fig. 8. Rupture of a corneal–scleral suture, with prolapse of part of iris, after cataract extraction.

Late postoperative complications include cystoid macula oedema, retinal detachment, glaucoma, and corneal oedema.

Extracapsular cataract extraction (Fig. 9)

Before operation the pupil must be well dilated with 1 per cent cyclopentolate and 5 per cent phenylephrine drops used with sufficient frequency to produce good mydriasis. In all the subsequent stages in the operation it is essential to protect the endothelium of the cornea from mechanical damage.

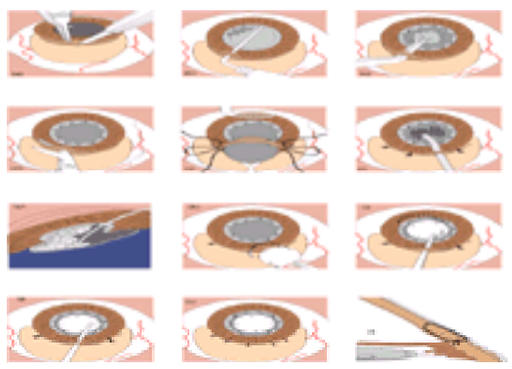


Fig. 9. Extracapsular cataract extraction.

This operation ideally requires an operating microscope with coaxial illumination, but can be done with a ×4.5 magnification loupe with a light mounted on it.

1. *Lids and fixation*. If there is no movement in the extraocular muscles after local anaesthesia, many surgeons dispense with the superior rectus suture.
2. *Section*. A shelved incision with the first part of the section being angled posteriorly is preferred by many surgeons. This first part of the incision is made through 150°, a smaller circumference than the 170 to 180° in intracapsular extraction, but the eye is not yet opened.
3. *Capsulotomy*. The tip and shaft of an insulin-syringe needle are bent to make a cystotome, which is inserted through the section at one point. The anterior chamber is maintained, if necessary, by fluid in the insulin syringe. The centre of the anterior capsule is perforated in a circle in the ‘can opener' technique, or the tip of the needle may be used to tear the capsule under direct vision to make a smooth edge (‘capsulorrhexis').
4. *The corneoscleral section is now completed*. The incision, bevelled downwards, is completed with the blade or corneal scissors through 150°.
5. A viscoelastic substance (hydroxypropylmethylcellulose 2 per cent or Healonid) is injected into the anterior chamber. The anterior capsule is removed with fine, non-toothed forceps.
6. *Expression of the nucleus*. The nucleus may be separated by hydrodissection, balanced salt solution injected from a fine cannula to separate the layers of the cortex. The lens is expressed by pressure on the lower limbus and counter pressure on the upper edge of the corneoscleral incision.
7. *Irrigation–aspiration*. The lens fibres of the cortex are carefully removed under direct vision by irrigation through a two-way cannula with a suitable solution (such as Hartmann's or Ringer's). The opening of the cannula must remain facing the operator and care must be taken not to catch the posterior capsule.
8. *Suturing*. 10/0 monofilament nylon can be used, and the knots buried by turning them into the wound. The sutures must be tied evenly and not tightly if astigmatism is to be avoided.
9. *Complications*. The early complications are similar to those of intracapsular cataract extraction. The most common, long-term complication is thickening of the posterior capsule, which is more likely if lens fibres have been left on the posterior capsule. This can be anticipated with a capsulotomy at the time of surgery or treated with a neodymium–yttrium aluminium garnet (**Nd–YAG**) laser capsulotomy later. Cystoid macula oedema and retinal detachment are reported to occur less frequently with extracapsular than with intracapsular cataract extraction.

Phakoemulsification

This technique was introduced in the late 1960s to break down the nucleus of a cataractous lens by means of ultrasonic vibration applied to a needle at 40 kHz. It is popular in industrialized countries because it enables a smaller incision (3 mm) to be used and therefore reduces the chance of postoperative astigmatism and reduces the need for hospital admission. The cost of the phakoemulsifier, the difficulty of the procedure, and maintenance of the machine at present restrict its suitability in developing countries. Techniques similar to phakoemulsification having the advantages of a small incision but without the use of an expensive phakoemulsifier are now being developed for use in Asia and Africa.

Optical correction of aphakia after cataract extraction

When the lens has been removed, the eye is referred to as being aphakic. In order to obtain good vision the aphakia must be corrected optically. There are three principal methods. The advantages and disadvantages are outlined in Table 3. In Western countries, many ophthalmologists perform a phakoemulsification with implantation of an intraocular lens into the posterior chamber. In many developing countries there is a trend from intracapsular to extracapsular cataract extraction, and, in some situations, to small-incision cataract extraction with or without a phakoemulsifer. Several large-scale clinical trials have been undertaken in Asia to compare intracapsular cataract extraction and spectacles against intracapsular cataract extraction with an anterior-chamber intraocular lens and intracapsular extraction with spectacles against extracapsular extraction with a posterior-chamber lens. The results in Asia have shown a low incidence of blinding complications at 1-year follow-up with posterior-chamber and anterior-chamber intraocular lenses. There is more blindness at 1 year in the spectacle groups due to lost and broken spectacles that have not been replaced. The most important change taking place, therefore, is from cataract surgery with spectacles (Fig. 10) to cataract surgery with an intraocular lens.



Fig. 10. Aphakic spectacle correction gives a magnified image.

	Advantages	Disadvantages
Spectacles	Inexpensive, suitable for unimpacted or extracapsular cataract extraction	Need employing visual rehabilitation poor due to magnification and distortion
Contact lenses	Suitable for unimpacted or extracapsular cataract extraction, good visual rehabilitation	Expensive, need employing, patient intolerance
Intraocular lenses		
Posterior chamber	Excellent visual rehabilitation, good safety record	Only suitable for extracapsular cataract extraction, posterior capsule opacification may eventually cause difficult surgery
Anterior chamber	Excellent visual rehabilitation, can be used with extracapsular or extracapsular extraction	Glaucoma and corneal damage, cataract and pupil distortion

Table 3 Correction of aphakia

Glaucoma

Worldwide, glaucoma is the second cause of blindness, being responsible for approximately 15 per cent of all blindness (i.e., 6.7 million people), although in some parts of West Africa and South-East Asia this proportion is higher.

Definition and classification

Glaucoma is a broad term used to describe a group of different pathophysiological processes that have in common optic-nerve damage as demonstrated by pathological cupping of the optic disc and characteristic visual-field loss. The emphasis in case definition is now on ‘end-organ damage' rather than on risk factors such as intraocular pressure, which is often high but may be within normal limits.

Glaucoma is classified as:

1. primary angle-closure glaucoma;
2. primary open-angle glaucoma (chronic simple glaucoma);
3. secondary glaucomas:
   a. closed angle,
   b. open angle;
4. congenital glaucoma (buphthalmos).

Primary angle-closure glaucoma

This is a very common form of glaucoma worldwide. It is especially prevalent in peoples of East Asian origin. The proportion of primary angle-closure to open-angle glaucoma is 1:10 in the United Kingdom, 1:1 in India, and 4:1 in Myanmar.

The predisposing factors are a shallow anterior chamber (Fig. 11), a smaller corneal diameter, and, to a lesser extent, a shorter axial length. These variables are in part determined genetically. The precipitating factors include a gradual increase in the size of the lens with age, causing the anterior chamber to become shallow and producing partial pupillary block. The acute attack may also be precipitated in a dark room, where the pupil is partially dilated. The iris root is bowed forward to touch the back of the cornea, thus preventing the aqueous from having access to drainage through the trabecular meshwork.

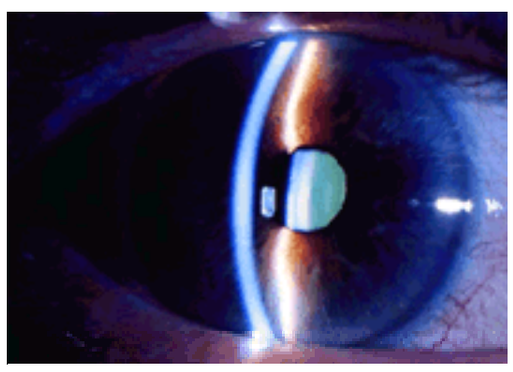


Fig. 11. Shallow anterior chamber shown in the beam of a slit-lamp.

Clinical presentation

The patient may present either with an acute attack with pain (Fig. 12) or with a chronic form with gradual loss of vision. In an acute congestive attack there is typically sudden reduction of vision and severe pain in one eye. This may be accompanied by nausea and vomiting. The eye is red, with dilated ciliary vessels around the limbus. The cornea is hazy, the anterior chamber shallow, and the pupil semidilated and non-reactive. The intraocular pressure is very high, the eye feeling stony hard to the fingers. There may be a history of transient prodromal attacks of blurred vision and discomfort.



Fig. 12. Acute angle-closure glaucoma, with injection of the ciliary vessels around the limbus; the oedema of the cornea is shown by the irregularity of the light reflex.

Immediate management

The traditional treatment has been frequent pilocarpine 2 per cent drops every 5 min for the first hour, then hourly. However, this is now questioned, as it is argued that the pilocarpine increases both accommodation and lens thickness, allowing the lens to move forwards, thus increasing the relative pupillary block.

First, intraocular pressure must be reduced quickly to limit damage to the optic nerve. Second, it is necessary to separate the iris from the corneal periphery by miosis so that permanent adhesions (synechiae) are not formed. These two aims are achieved by the following actions.

1. Laying the patient on their back to allow the lens to fall back;
2. giving acetazolamide 500 mg by intramuscular or very slow intravenous injection and then
3. 250 mg four times per day orally;
4. giving topical b-blocker drops, such as timolol 0.5 per cent twice daily;
5. if such treatment does not lower the intraocular pressure adequately, giving oral glycerol
6. or intravenous mannitol in severe cases, to withdraw fluid from the eye by osmosis.
7. Then giving pilocarpine 0.5 to 2 per cent, one or two drops, four times per day to induce miosis.

Peripheral iridectomy

Definitive treatment is by a peripheral iridectomy (Fig. 13) or laser iridotomy.

1. *Incision*. In the past, a corneoscleral incision was usually made under a conjunctival flap, which could be either limbus- or fornix-based. Nowadays, a corneal section is more common.
2. *The incision is made in two stages*. The first incision is angled backwards, its width is 3 to 4 mm, and it is situated on the cornea 1 mm from the limbus. The second part of the incision is bevelled forwards. The posterior lip is then depressed so that the iris prolapses. If the iris does not present in the wound, it may be picked up close to its root with fine-toothed forceps, taking care not to touch the anterior lens capsule.
3. *Iridectomy*. The iris is drawn slightly out of the wound and a full-thickness piece excised with de Wecker's scissors.
4. *Repositioning of the iris*. The iris can be replaced by stroking the lips of the wound with an iris repositor. A full-thickness iridectomy should now be visible.
5. *Wound closure*. The two-stage corneal wound is usually self-sealing. If on touching with a cellulose swab there is a visible leak, one or two corneal sutures of fine (10/0) suture material can be inserted, and the knots buried.

In units with appropriate facilities it is common to make an iridotomy with either an argon or Nd–YAG laser. A portable Nd–YAG laser that can be used in rural settings is now available.

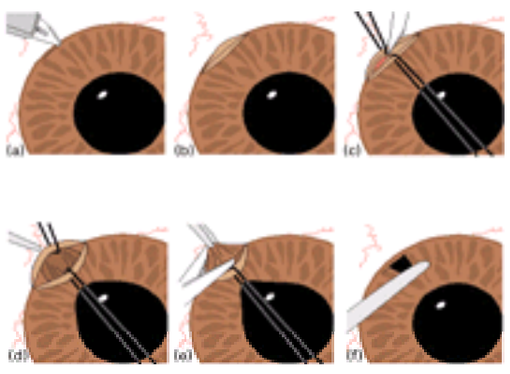


Fig. 13. Peripheral iridectomy in angle-closure glaucoma.

Prophylaxis

The second eye of a patient treated for an acute angle-closure attack should be given either a prophylactic peripheral iridectomy or an iridotomy before the patient is discharged. It is often difficult to persuade a patient to have an operation on an eye that has shown no symptoms, or to a person with incipient angle-closure who has had no symptoms in either eye to undergo such a procedure. However, with the laser iridotomy it is more likely to be accepted than with a formal operation.

Chronic angle-closure glaucoma

This is a form of glaucoma in which the angle gradually becomes closed off by creeping adhesions of the iris root to the peripheral cornea. The slow onset means that this form does not have acute symptoms but presents late, similar to a chronic open-angle glaucoma. Because aqueous drainage is occluded by permanent iris synechiae, trabeculectomy is indicated as for chronic open-angle glaucoma.

Primary open-angle glaucoma

Chronic simple glaucoma is the most common form of glaucoma in European and African populations. It starts at an earlier age in black people and appears to be more aggressive. High prevalences have been reported in some Caribbean islands.

Clinical presentation

The aqueous circulates through the pupil and has access to the drainage mechanism of the trabecular meshwork of the angle as normal, but the rate of drainage is reduced. There is a gradual rise in intraocular pressure, without acute symptoms. The raised pressure predisposes to loss of visual field, with characteristic ‘cupping' of the optic-nerve head (Fig. 14). Central visual loss is only noticed in the advanced stages when the optic nerve has already been severely damaged. Emphasis is therefore placed on early detection of raised intraocular pressure by tonometry and of abnormal or asymmetrical cupping of the optic disc by ophthalmoscopy.

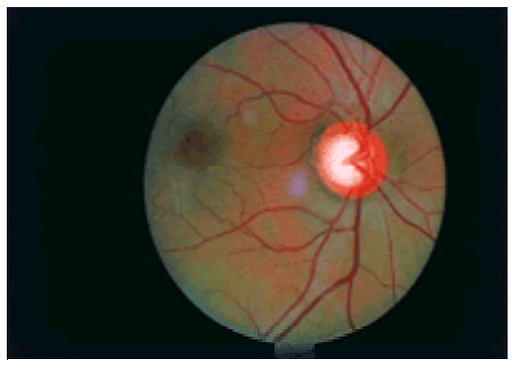


Fig. 14. Fundus photograph showing cupping of the optic disc in chronic open-angle glaucoma.

Diagnosis is confirmed by finding the characteristic defects (‘scotomas') on testing the visual fields.

Trabeculectomy

The trend is to operate rather than to start life-long medical treatment. Certainly, in the African setting, most patients do not continue medical treatment long term, even

if the drugs are available and can be afforded. Surgery is therefore the first line of treatment.

Trabeculectomy is a drainage operation in which a rectangular segment of the corneoscleral tissue is excised (Fig. 15); this is done under a partial-thickness scleral flap. The stages of trabeculectomy are as follows.

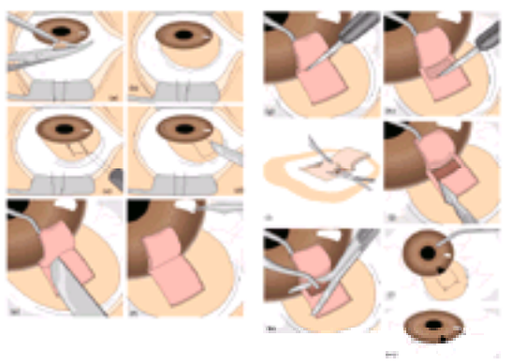


Fig. 15. Trabeculectomy technique.

1. *Lids and fixation*. As for cataract extraction.
2. *Conjunctival flap*. To form a fornix-based flap the conjunctiva is incised at the limbus. A limbus-based conjunctival flap may also be used. The Tenon's capsule is cleared away from over the sclera in the area in which the scleral flap will be raised.
3. *Superficial scleral flap*. A limbus-based flap of superficial sclera is now fashioned. It is usually half-thickness and approximately 5 mm wide at the limbus. The position is marked out on the sclera with a knife or razorblade fragment. Main blood vessels running through the sclera should be avoided or coagulated. The margins of the flap are incised to half the thickness of the sclera. One corner is picked up and the section carried forward with the blade or a Tooke's knife until it is past the limbus into 1 mm of clear cornea. At this point the tissue beneath the scleral flap changes from white to blue–grey opalescent in appearance.
4. *Excision of trabecular tissue*. While the superficial scleral flap is held back by an assistant, a small rectangular block of the deep corneal–scleral tissue is now excised. This should be approximately 3 mm wide in the line of the limbus and 2 mm in radial depth. The area is marked out and incised with a sharp knife and the removal of the block completed with small scissors.
5. *Iridectomy*. The iris either prolapses spontaneously through the trabeculectomy opening, or can be gently grasped with fine forceps so that a small piece can be excised with de Wecker's scissors as in a routine peripheral iridectomy. On stroking of the cornea with an iris repositor, the sphincter should constrict and the iris be repositioned within the anterior chamber.
6. *Closure*. For closure, the scleral flap is sutured at each posterior corner with fine nylon or virgin silk. The conjunctival flap is then sutured firmly in position at the limbus with two sutures of similar material, or, if a limbus-based conjunctival flap was used, a running suture is used to close the conjunctiva.
7. *Postoperative treatment*. Steroid drops and antibiotic drops are used three times a day for 5 to 7 days. Cyclopentolate 1 per cent or tropicamide 1 per cent twice daily for 1 week are recommended to prevent adherence of the iris to the lens.
8. *Complications*. The most common early complication encountered is a delay in reforming of the anterior chamber. This is usually because of excessive drainage, either due to a leak beneath the conjunctiva, demonstrated by the flow of aqueous after fluorescein is placed into the conjunctival sac (the Seidel test), or to the scleral flap being too thin or loose. Occasionally a shallow anterior chamber is due to a choroidal detachment. The anterior chamber will usually reform over several days with conservative management; if not, revision of the flap with additional conjunctival or scleral sutures usually solves the problem. Iritis is common in pigmented eyes, but can usually be controlled by topical steroids and mydriatics.

Trachoma/entropion

The causative organism of trachoma is *Chlamydia trachomatis*, serotypes A, B, and C. The organism infects the conjunctiva, particularly of young children, and produces a mucopurulent discharge that is spread within the family by children's fingers, by contaminated bed clothes and rags, and by flies (Fig. 16). The inflammation is made more severe by repeated infection with chlamydia and superinfection by bacteria. Trachoma becomes a disease of blinding severity under conditions of poverty where there is lack of water for keeping the face clean and lack of sanitation so that the breeding of flies is encouraged.



Fig. 16. Flies of the species *Musca sorbens* feeding on the discharge from the eye of a 4-year-old child infected with trachoma.

Simplified grading scheme

A simplified grading scheme for trachoma that gives good reproducibility between observers in field studies, and can be taught to and used by community health workers, was introduced by the World Health Organization in 1987 (Table 4).

Grade	Definition
TF	Five or more follicles seen on the tarsal conjunctiva of the everted upper lid
TI	Intense inflammation of the upper tarsal conjunctiva, obscuring the view of more than half of the deep tarsal blood vessels on the everted upper eyelid
TS	White lines of subconjunctival scarring visible on the surface of the everted upper eyelid
TT	Trichiasis, at least one eyelash rubbing on the globe, or evidence of recent removal
CO	Corneal opacity, obscuring at least part of the pupil margin

Table 4 Simplified grading scheme for trachoma (1987)

Prevention of blindness

Active infection (TF or TI) is treated with 1 per cent tetracycline ointment instilled twice daily for 6 weeks or with one dose of azithromycin. Repeated reinfection can be

prevented by regular face washing and by community action to improve sanitation and reduce fly density. The actual blinding process is due to vascularization and scarring of the cornea as a result of trichiasis, secondary infection, and corneal ulceration (Fig. 17, Fig. 18, Fig. 19, Fig. 20 and Fig. 21).

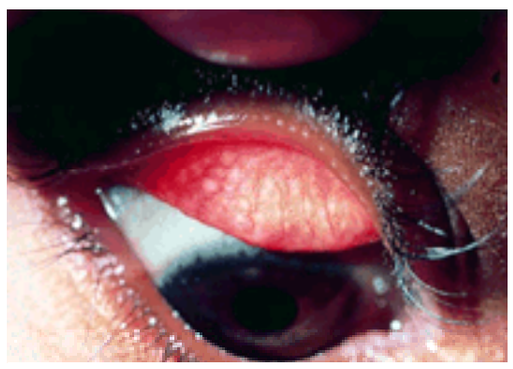


Fig. 17. The follicular stage of trachoma (grade TF) in the tarsal conjunctiva of the upper lid (by courtesy of P. Courtright).



Fig. 18. Intense inflammation of the upper tarsal conjunctiva in trachoma infection, combined with follicle formation (grade TF/TI).



Fig. 19. Scarring of the conjunctiva from repeated trachoma infections and formation of concretions (grade TS).

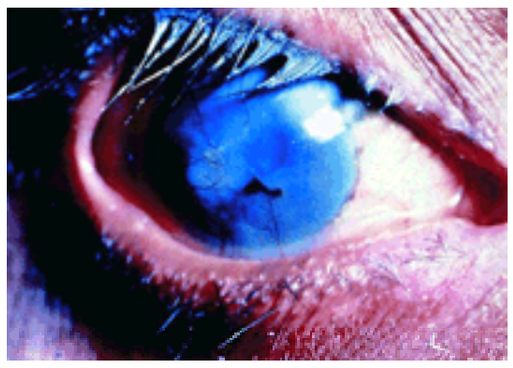


Fig. 20. Trichiasis due to trachoma (grade TT).

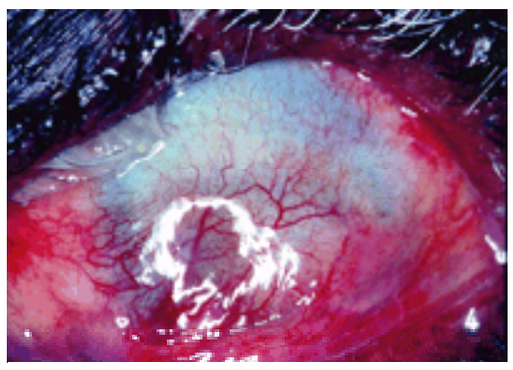


Fig. 21. Corneal opacity and inflammation, associated with trichiasis (grade TT/CO).

Surgery for trichiasis/entropion

A large number of operations have been devised for the correction of trichiasis and entropion as a result of trachoma. There are two main principles: (1) to evert the lid margin so that the eye lashes are directed outwards, and (2) to correct any shortening of the upper lid. Recently, some of these operations have been the subject of a randomized, controlled trial. Bilamellar tarsal rotation was the most successful operation regardless of previous surgery (Fig. 22). It was devised by Ballen in 1964, and is itself a modification of the earlier procedure described by Weis.

1. Apply topical anaesthetic drops to the conjunctiva. Inject 3 ml of local anaesthetic into the upper eyelid between skin and tarsal plate above the eyelashes of the upper eyelid and massage it into the tissues for 1 min.

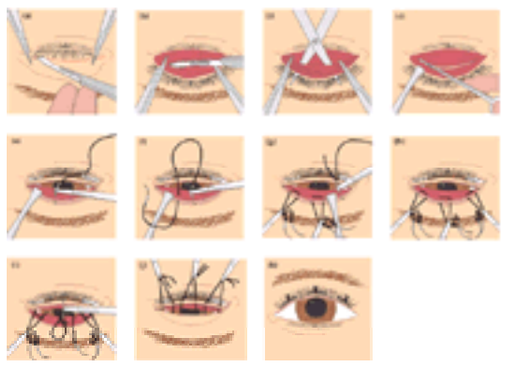


Fig. 22. Bilamellar tarsal rotation.

2. Place haemostat forceps vertically at the lateral end of the upper eyelid (5 mm on to the eyelid) and at the medial end, just lateral to the lacrimal punctum.
3. Incise through the skin and muscle of the upper eyelid between the two haemostats 3 mm above the eyelash margin.
4. Evert the eyelid and, using a scalpel, make an incision 2 mm from the lid margin, cutting through the tarsal plate to join up with the skin incision. Complete the division of the tarsal plate medially and laterally with scissors.
5. Place three mattress sutures with double-armed needles in the eyelid to rotate the distal lid margin outwardly. The suture starts from the inner conjunctival surface of the margin of the proximal tarsal plate. The needle on each arm of the suture is passed through the proximal tarsal plate and is then rotated and passed through the muscle and skin of the distal eyelid margin anterior to the tarsal plate. The three mattress sutures are tied above the eyelash line.
6. The skin is closed with interrupted silk sutures.
7. Topical antibiotic is applied three times daily for 7 days.
8. All sutures are removed after 7 days.

The advantage of bilamellar tarsal rotation is its relative simplicity. It takes approximately 10 min to perform on one eye; it can be taught to ophthalmic assistants and nurses, and performed under field conditions in communities where trichiasis is prevalent. Its disadvantage is the potential for shortening of the upper eyelid. For this reason, in young patients (under 25 years), or in those with severe shortening of the upper eyelid, the more difficult buccal mucous membrane graft is considered an option.

Ocular leprosy/exposure keratitis

Leprosy is a chronic bacterial disease of low-grade infectivity, due to *Mycobacterium leprae*. The disease occurs mainly in developing countries. It is thought to be spread by droplets from the upper respiratory tract, a mechanism encouraged by poverty and crowding. In most people, infection heals spontaneously without producing clinically recognizable disease. Established infections may produce a spectrum of disease from multibacillary leprosy, in which the cellular immunity is low and there are many leprosy bacilli in the tissues, through borderline to paucibacillary leprosy with high cellular immunity and correspondingly low bacterial counts.

Most complications are caused either by a massive bacterial load and secondary atrophy of tissues or by sudden changes in the patient's immune response to the organism. Type I reaction (reversal) occurs in all types of borderline leprosy and is associated with an increase in cell-mediated activity. Type II reaction (erythema nodosum leprosum) occurs in multibacillary leprosy. Type I reaction can lead to paralysis of the ophthalmic branch of the trigeminal nerve and the facial nerve. Type II reaction can lead to acute iritis.

Control of blinding complications

Lagophthalmos (literally ‘hare eye', a failure to close the eyelids for any reason; Fig. 23, Fig. 24) may occur as a result of a type I reaction. If of recent onset (6 months or less), treat vigorously with oral prednisone 30 mg per day, decreasing over 6 months, together with blinking exercises and protective spectacles. If a long-standing lagophthalmos causes exposure keratitis, then surgical treatment is indicated. This is usually by lateral tarsorrhaphy. Temporalis muscle-transfer surgery has given disappointing recovery of spontaneous regular blinking.



Fig. 23. Lagophthalmos (cranial nerve VII) associated with corneal anaesthesia (cranial nerve V) and dacryocystitis on the right side: arrested borderline multibacillary leprosy (by courtesy of M. Hogeweg).



Fig. 24. Forced voluntary lid closure in VII nerve paralysis in leprosy; note ectropion of lower eyelids (same patient as in Fig. 23) (by courtesy of M. Hogeweg).

Corneal anaesthesia, due to involvement of the Vth nerve in type I reactions, may occur in combination with lagophthalmos (Fig. 25). Partial loss of corneal sensation, resulting in an itch, can be especially dangerous if the eye is rubbed with insensitive, deformed hands. In severe cases, tarsorrhaphy is indicated if corneal clarity is threatened.



Fig. 25. Opacity of left cornea due to exposure from lagophthalmos in multibacillary leprosy.

Acute iritis (Fig. 26) is treated with atropine 1 per cent and steroid eye drops. If it occurs in combination with scleritis, topical treatment has to be combined with systemic treatment for erythema nodosum leprosum (type II reaction).

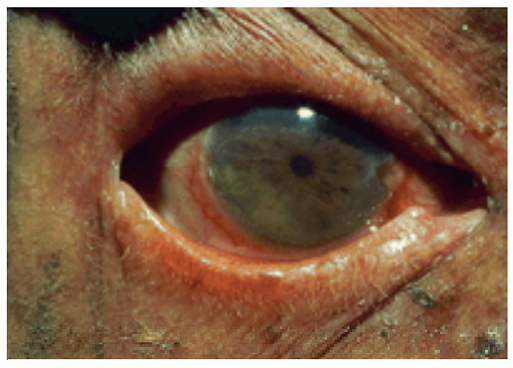


Fig. 26. Iritis in leprosy: note small pupil, slightly hazy view of iris, and injection of ciliary vessels around limbus (by courtesy of H. Limburg).

Cataract may occur secondarily to iritis, but is usually incidental age-related cataract. Routine cataract extraction can be performed without special precautions if the patient is under treatment or has been released from treatment and has not recently experienced any reaction.

Prevention of blinding complications

Prevention includes early systemic multidrug treatment, prompt treatment of type I reactions with systemic steroids, early treatment of any eye complications, and provision of cataract surgical services for leprosy patients with cataracts.

A temporary tarsorrhaphy is indicated for corneal protection when a VIIth nerve palsy is expected to recover, or to assist healing of an indolent corneal ulcer. A permanent tarsorrhaphy is required for established VIIth nerve palsy and when there is severe loss of corneal sensation.

Temporary tarsorrhaphy (Fig. 27)

This can be carried out medially, centrally, or laterally along the eyelids. The length is marked out on the margins of the upper and lower lids and an incision 2 mm deep is made in the ‘grey line' in the middle of the eyelid margin, between the orifices of the meibomian glands behind and the eyelash follicles in front. The marginal epithelium of the lid margins is then excised. The upper and lower eyelids are apposed and sutured together in two layers. The raw surfaces of the tarsal plates of the upper and lower lids are approximated with 6/0 absorbable sutures, making sure that these do not go through on the conjunctival surface where they might abrade the cornea. Double-armed 6/0 or 4/0 non-absorbable mattress sutures are inserted through ‘buttons' (cut from a rubber or plastic cannula), through the skin well back from the lid margin of the upper lid, out through the grey line in the raw area, into the corresponding position of the lower lid, and through the skin of the lower lid to be tied over a similar ‘button' or ‘bolster'. The mattress sutures are removed after 2 to 3 weeks.

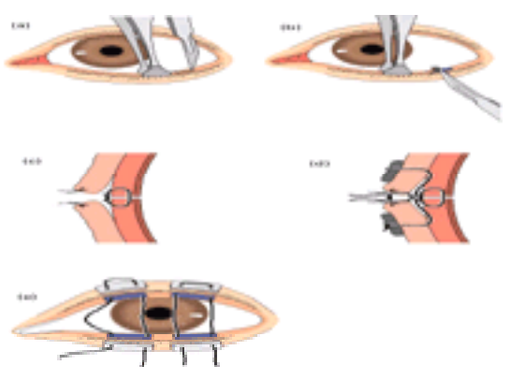


Fig. 27. Temporary tarsorrhaphy.

Permanent tarsorrhaphy (Fig. 28, Fig. 29)

The eyelids are overlapped, with excision of a rim of soft tissue from their margins, to produce a more permanent adhesion that will not undergo retraction. This is most frequently done at the lateral canthus. After the lateral ends of the lids have been split into their two lamellae, a triangle of skin and orbicularis muscle (the anterior lamella) is excised from the lower lid, and a corresponding triangle of tarsal plate and conjunctiva (the posterior lamella) from the upper lid. The upper edge of the tarsoconjunctival lamella of the lower lid may be sutured with fine catgut to the cut edge of the conjunctiva of the upper lid to strengthen the future union. A mattress suture is then brought forward from the conjunctival surface of the lower lid and through the prepared triangle of the upper lid, to be tied finally over a bolster on the skin surface. If a vertical incision is now made at the medial end of the upper lid triangle, a flap is created that can be pulled down to cover the raw tarsal triangle of the lower lid.

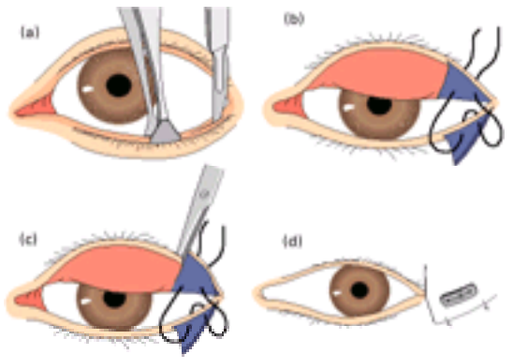


Fig. 28. Permanent tarsorrhaphy.



Fig. 29. Leprosy: permanent lateral tarsorrhaphies to protect corneas from further exposure damage due to lagophthalmos.

Corneal opacity

The major cause of blindness from bilateral corneal opacity is trachoma. Other important causes are suppurative keratitis (although usually unilateral), vitamin A deficiency, climatic keratopathy (Fig. 30, Fig. 31), and exposure due to the lagophthalmos in leprosy. Excimer laser may prove to be the treatment of choice for climatic keratopathy. Pterygium only rarely crosses the pupil to impair vision.

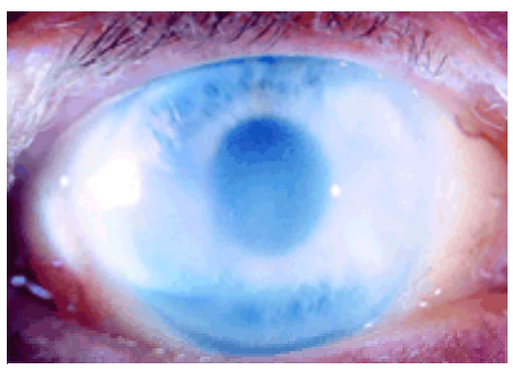


Fig. 30. Climatic keratopathy in an 83-year-old Labrador man, showing the fine, granular ‘droplets' deposited in the superficial corneal stroma.

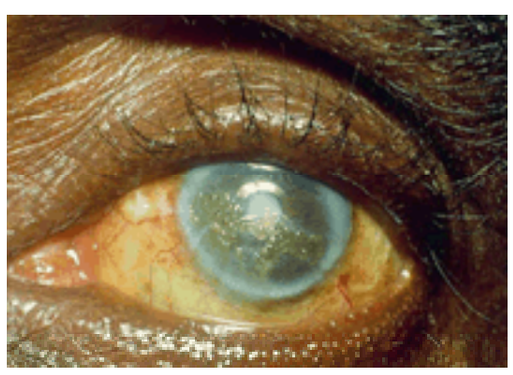


Fig. 31. Advanced climatic keratopathy in a 70-year-old man from Somalia (by courtesy of S. Franken).

Optical iridectomy

Occasionally a patient blind from corneal scarring may obtain useful travel vision by the creation of an artificial pupil (optical iridectomy; see Fig. 40). This is only indicated in a patient who is bilaterally blind and who has a central scar (leucoma) with clear peripheral cornea. Ideally, the optical iridectomy is made inferonasally to be most useful for vision. The incision is made under a conjunctival flap at the limbus. The iris is prolapsed and gently eased out through the wound. An iridectomy from the iris root to the pupil margin is performed with scissors. The limbal incision is closed with sutures.

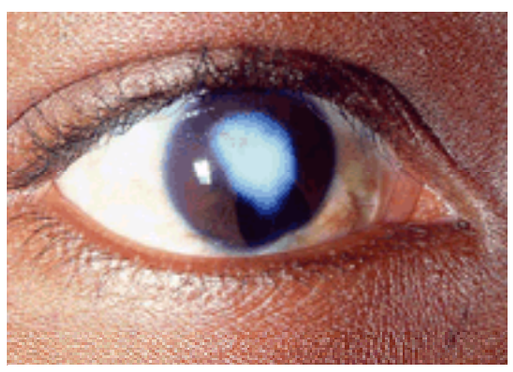


Fig. 40. Optical iridectomy in a child with corneal opacity (by courtesy of C. Gilbert).

Keratoplasty

Corneal grafting (Fig. 32 and see Fig. 41), either full thickness or lamellar, is the definitive treatment for opacities due to corneal dystrophies, corneal scar of various causes, and distortion due to keratoconus. With gradual evolution of technical improvements over the last 30 years (operating microscope, fine sutures, microsurgical instruments, and topical local steroids), the prognosis for survival of corneal grafts is now good for many conditions. Heavily vascularized corneas, absence of lacrimal secretion, or distortion of the eyelid margins all, however, decrease the chance for successful grafting.

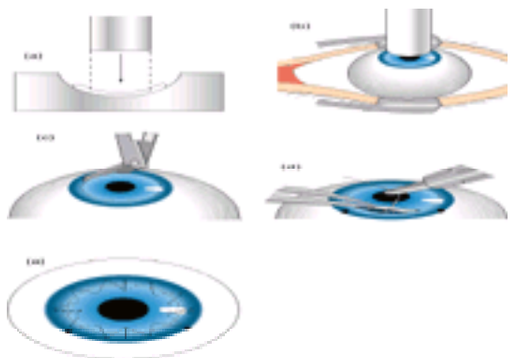


Fig. 32. Keratoplasty.

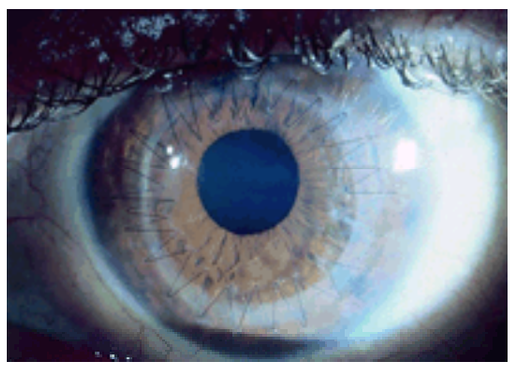


Fig. 41. Clear, full-thickness corneal graft for corneal opacity; continuous 10/0 monofilament nylon suture had been used.

Corneal storage has also improved. The method that has been used for many years is to store the whole eye in a sterile container in a refrigerator at 4°C. Under these conditions the cornea should be used for transplantation as soon as possible to avoid extensive endothelial cell death; the limit is usually regarded as 48 h after death. Short-term storage in tissue-culture media, such as the McCarey–Kaufman solution, now allows storage for up to 4 days. Corneal organ culture allows the donor tissue to be used for up to 30 days after death.

Allograft rejection is substantially reduced by the use of fine monofilament sutures, which reduce the inflammatory reaction; 10/0 nylon sutures are most often used, either interrupted or continuous. An operating microscope and high-quality microsurgical instruments are required to be able to suture the donated cornea into the recipient satisfactorily. The suture knots are buried in the corneal stroma.

In certain circumstances, a localized central opacity can be treated with a rotation autograft, when an eccentric corneal graft of the patient's own cornea rotates the opacity out of the optic axis towards the limbus. In rare circumstances, a corneal exchange can be undertaken between an eye with a clear cornea and total blindness due to posterior-segment disease and the other eye of the same individual that has corneal opacity but is otherwise healthy. There is no possibility of graft rejection with such an autograft. Both these techniques are readily applicable in the developing world where the supply of donor allograft material may be very limited. Recent reports from India and Kenya have demonstrated that, in the developing world, keratoplasty can also be an effective intervention in selected cases.

Pterygium excision

This is a very common condition (Fig. 33). The interaction of different causes is not fully understood. Pterygium very seldom causes blindness.

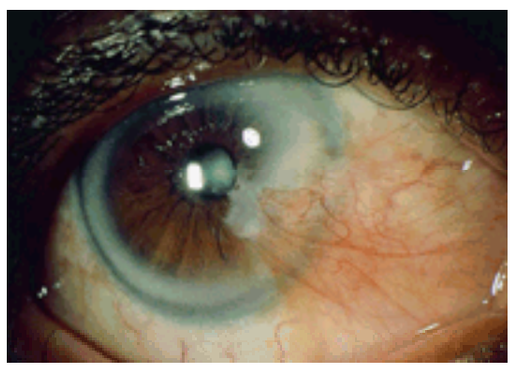


Fig. 33. Treatment of pterygium.

Many different operations have been devised to remove the pterygium and try to prevent a recurrence. A simple technique is to excise the conjunctiva at the base of the pterygium after the head is removed from the cornea, leaving the sclera bare (Fig. 34). If the pterygium recurs, it can be treated with a free conjunctival graft taken from the superior bulbar conjunctiva, after a second excision of the pterygium. The donor area is left to regenerate. This conjunctival autograft technique is now also becoming popular as the primary treatment for pterygium, and in clinical trials has been shown to give rise to significantly fewer recurrences than the bare sclera method.

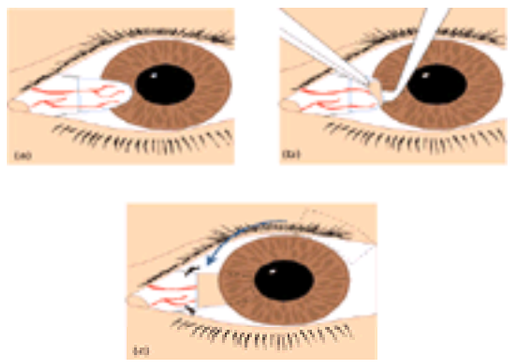


Fig. 34. Pterygium.

Childhood blindness

An estimated 1.5 million children in the world are blind., and approximately 50 per cent of this blindness is thought to be due to vitamin A deficiency.

Vitamin A deficiency/xerophthalmia

Each year some 250 000 children are needlessly blinded because of the effects of deficiency of vitamin A in their diet. Another 250 000 children are left with lesser degrees of permanent visual impairment due to corneal damage. The conditions under which vitamin A deficiency and xerophthalmia occur are poverty, high incidence of precipitating infections (measles and gastroenteritis) in preschool-age children, and lack of maternal education (Fig. 35). Illness and death from measles, lower respiratory-tract infections, and diarrhoea have all been reported as being more common in young children with a low vitamin A status. Conversely, these infections themselves worsen the vitamin A status, thus precipitating blindness.

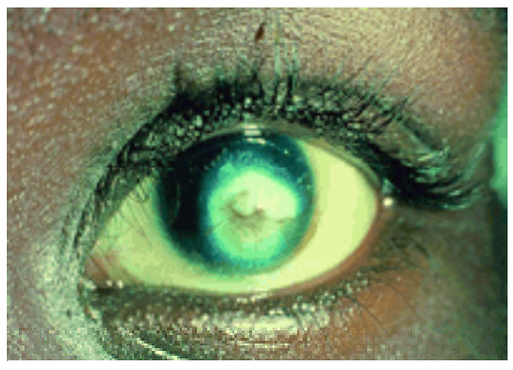


Fig. 35. Corneal scarring in a child following xerophthalmia precipitated by severe diarrhoea (by courtesy of S. Franken).

Eye changes

The older child may complain of night blindness. Signs include dryness of the conjunctiva (xerosis) due to a decrease in the goblet cells and keratinization of the epithelium. Bitot spots are white, foamy, raised areas, usually on the temporal conjunctiva (Fig. 36). The most dangerous signs are dryness of the cornea, ulceration, and then a sudden melting of the cornea, keratomalacia, leading to perforation and often loss of the eye (Fig. 37, Fig. 38). The exact reason for the sudden corneal melting under these conditions is not fully understood. The endstage of blinding malnutrition is corneal scarring, which may be a leucoma impairing vision but allowing the structure of the anterior chamber to remain, or there may be anterior staphyloma (bulging forward of the cornea) or phthisis bulbi (a small, shrunken globe) (Fig. 39).

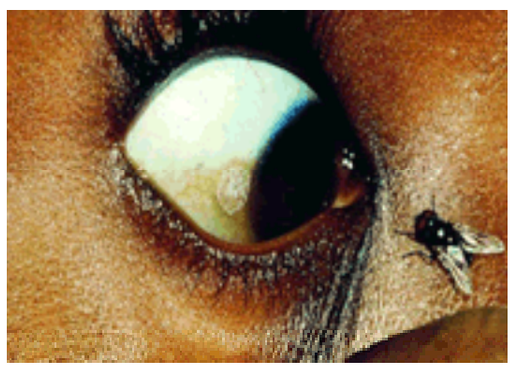


Fig. 36. Xerophthalmia: Bitot spot on the temporal conjunctiva of a 32-month-old child.

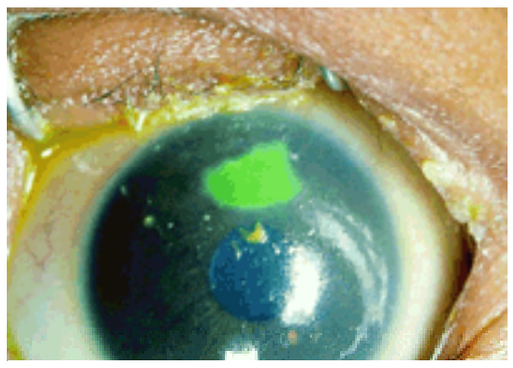


Fig. 37. Xerophthalmia: dryness of the cornea, with early ulceration stained with fluorescein.

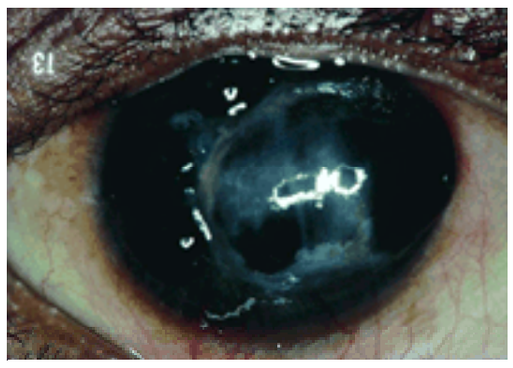


Fig. 38. Xerophthalmia: end-stage, perforation of cornea with protrusion of iris.



Fig. 39. Xerophthalmia: end-stage, bilateral anterior staphylomas. (by courtesy of J.D.C. Anderson).

Management

Corneal involvement is a medical emergency. Three doses of vitamin A in the form of high-dose capsules are given on day 1, day 2, and 1 to 4 weeks later. The dose is 200 000 iu for children aged 1 to 6 years, and 100 000 iu under the age of 1 year. Topical antibiotic ointment is also used. A joint WHO/UNICEF statement has now recommended that this high dosage of vitamin A should be given to all children with measles in communities where measles has a mortality of 1 per cent or more.

Occasionally a blind child or adult with central leucomas may be helped by an optical iridectomy (Fig. 40, Fig. 41). Because of iris adhesions to the cornea, keratoplasty is usually disappointing. Evisceration of the globe may be indicated for a painful or unsightly staphyloma.

Prevention

In the long run, this must consist of improved dietary habits and improved availability of vitamin A-containing foods. Short-term measures include vitamin A supplementation through capsule distribution or food fortification.

Congenital cataract (Fig. 42)



Fig. 42. Bilateral congenital cataracts.

This is the most common cause of treatable childhood blindness after vitamin A deficiency. The major known underlying causes are heredity and congenital acquired rubella syndrome.

Long-term visual loss can occur because of amblyopia, and early surgery is essential, with immediate correction of the refractive error by spectacles, contact lens, or intraocular lens if reasonable results are going to be obtained. In general, the results are much better in bilateral than unilateral congenital cataract. The most widely practised surgical technique is aspiration and irrigation, as for extracapsular cataract extraction in an adult. Since the nucleus is less hard than in an adult and more easily aspirated, a small incision can be made. A posterior capsulotomy may be performed at the end of the procedure. Some surgeons now advocate lensectomy and anterior vitrectomy through the anterior chamber in order to avoid the problem of secondary capsular opacification, which impedes vision and makes accurate refraction and optical correction of the aphakia difficult. The use of intraocular lenses in children over the age of 2 years is now accepted, but for children under this age there is still debate over the best method of visual rehabilitation.

Congenital glaucoma

Congenital glaucoma or bulphthalmos usually presents in early childhood with an enlarged, cloudy cornea (Fig. 43) and photophobia. Assessment, including ophthalmoscopy, tonometry, and gonioscopy (if possible), is made under general anaesthesia. If a membrane of tissue is present (as described by Barkan), then goniotomy is the treatment of choice (Fig. 44). This requires a surgical gonioscopic lens and an irrigating goniotomy knife so that the membrane can be incised under direct vision.



Fig. 43. Enlarged cornea of congenital glaucoma in right eye (by courtesy of A. Mushin).

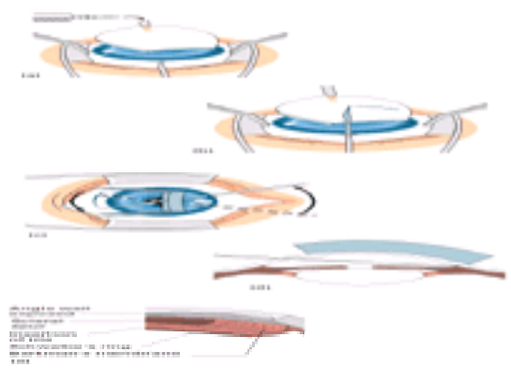


Fig. 44. Barkan goniotomy for congenital glaucoma.

In cases due to other structural abnormalities and where the cloudiness of the cornea prevents a clear view of the angle, a filtration procedure such as trabeculectomy is the management of choice.

Retinoblastoma

Retinoblastoma classically presents as a white pupil, or occasionally as a squint (Fig. 45). There may be a family history, when the condition may be bilateral. In developing countries, patients are often brought in late in the course of disease with secondary glaucoma or severe proptosis. The prognosis for the child's life under these circumstances is poor. Differential diagnosis for proptosis in this age group includes rhabdomyosarcoma, neuroblastoma, orbital abcess, Burkitt lymphoma, and hydatid disease. The second eye must be carefully examined. Should the disease prove to be bilateral, conservative treatment may allow the second eye to be preserved.



Fig. 45. Retinoblastoma shown by pupil reflex (leucocoria) of right eye.

Usually, enucleation is the only treatment possible in developing countries because of the advanced state of the tumour and the limited resources available. The emphasis when training physicians and nurses must therefore be on early detection and immediate referral and treatment.

The enucleated eye should be examined histologically for extension through the cut end of the optic nerve, or extension outside the sclera. In these circumstances, and in the event of orbital recurrence, radiotherapy and adjuvant chemotherapy are indicated if available. Screening of the second eye may allow conservative treatment to be undertaken, with preservation of the eye should the disease prove to be bilateral. Depending on the size of the tumour when discovered, conservative treatment may be external radiotherapy, or radioactive scleral plaques for retinoblastomas up to 10 mm in diameter. Cryotherapy with a retinal cryoprobe, with a triple cycle of freezing and thawing, may be used for tumours up to 7 mm in diameter. Photocoagulation with a xenon-light coagulator for tumours up to 5 mm is also possible; the light is not applied directly to the tumour but to the surrounding retina to cut off the blood supply.

Removal of the eye

Surgical removal of the entire globe is called enucleation. It is indicated when there is a painful blind eye (especially from severe trauma), and for malignant tumours if other treatment is not possible. Removal of the anterior part of the globe and the contents, leaving only the scleral coat and optic nerve, is referred to as evisceration, which usually gives better cosmetic results and ocular mobility. It is indicated when there is suppurative endophthalmitis or painful or unsightly staphyloma. Evisceration is an easier procedure than enucleation and may be performed by a non-specialist for a blind, painful, non-malignant eye.

Enucleation

The procedure is as follows (Fig. 46).

1. *Anaesthesia*. A general anaesthetic is preferred, but the procedure can be done with lid akinesia and retrobulbar local anaesthetic.

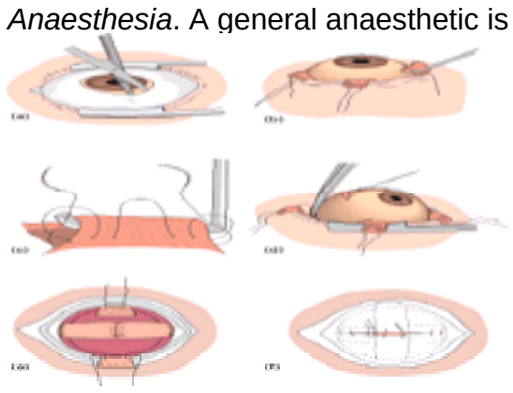


Fig. 46. Enucleation: (a) peritomy of conjunctiva; (b) isolating each rectus muscle; (c) inserting retention sutures in muscle; (d) cutting the optic nerve with scissors and/or removing eye; (e) ball implant inserted, and rectus muscles brought together in front of it; (f) Tenon's capsule and conjunctiva closed in two layers.

2. *Conjunctiva*. The conjunctiva is incised 360° around the limbus, and Tenon's capsule is incised and undermined well back into each quadrant.
3. *Extraocular muscles*. Each rectus muscle is identified and isolated with a muscle hook. 5/0 chromic catgut sutures are passed through each muscle, near its insertion, at least twice, and are locked in place. The rectus muscle is then divided on the globe side of the suture. The oblique muscles are usually divided near their insertions to the globe without inserting sutures. The remaining fascial sheath is dissected from the globe.
4. *Optic nerve*. Curved enucleation scissors are inserted along the lateral orbital wall and opened to feel the tough, rubbery optic nerve between the blades. The optic nerve is cut, in the case of a malignant tumour as close to the apex of the orbit as possible. The globe is removed and the orbit immediately packed with gauze, under pressure from the surgeon's hand, until bleeding stops.
5. *Implant*. Many different types of implant are now available. Most commonly, a silicone or plastic sphere of appropriate size is inserted into the muscle cone. The opposing rectus muscles are tied to each other in front of the implant by means of the absorbable retention sutures already attached. An implant allows some degree of movement of the future artificial eye. Some implants have grooves or slots through which the retention sutures on the rectus muscles can be passed, to improve movement.
6. *Closure*. Tenon's capsule and conjunctiva are each closed separately with a 5/0 continuous absorbable suture. Antibiotic ointment is applied to the conjunctiva, a plastic conformer inserted into the socket, and a pressure bandage applied for 2 days.

When the conjunctiva is healed, the conformer can be replaced with an artificial eye that matches the patient's remaining eye as closely as possible.

Evisceration

The procedure is as follows (Fig. 47).

1. *Anaesthesia*. General anaesthesia is preferable, but if not available the procedure can be performed under local anaesthetic.

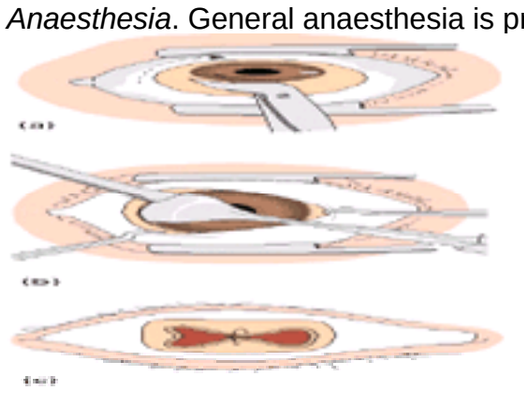


Fig. 47. Evisceration.

2. *Excision of cornea*. The conjunctiva and Tenon's capsule are divided all round the limbus and dissected back anterior to the rectus muscle insertions. A knife is passed through the cornea at the limbus. The corneal circumference is cut with corneal scissors and the cornea removed.
3. *Intraocular contents*. A curved scoop or ‘evisceration spoon' is introduced between the ciliary body and the sclera, and all the contents are scraped from the scleral coat. A dry swab held in forceps removes any remaining choroidal tissue, particularly at the exit of the vortex veins and the optic nerve.
4. *Dressing*. The inside of the sclera is filled with topical antibiotic. A *tulle gras* dressing may be inserted for 48 h. The sclera and conjunctiva may be loosely sutured in two layers with absorbable catgut, but in the case of endophthalmitis it may be better to leave the wound open. A pressure bandage is applied for 48 h. Systemic antibiotics are advisable for 5 days if endophthalmitis was present.
5. *Implant*. If the evisceration is for a cause other than infection, a ball implant that is small enough for the scleral edges to be brought loosely together can be inserted. The conjunctiva is closed with interrupted sutures. As the sclera shrinks down, the resulting fibrous scar may allow good movement of a later artificial eye.

Further reading

Bailey R, Arullendran P, Whittle HC, Mabey DCW. Randomised controlled trial of single dose azithromycin in treatment of trachoma. *Lancet* 1993, **342**: 453–6. [A single dose of oral azithromycin was as effective as 6 weeks' treatment with tetracycline ointment in controlling inflammatory trachoma.]

Ballen PH. A simple procedure for the relief of trichiasis and entropion of the upper lid. *Archives of Ophthalmology* 1964; **72**: 239–40. [the first description of the application of the bilamellar tarsal rotation procedure to entropion of the upper eyelid.]

Cairns JE. Trabeculectomy: preliminary report of a new method. *American Journal of Ophthalmology* 1968; **66**: 673–9. [Classic first description of the operation of trabeculectomy.]

Casey TA, Mayer DJ. *Corneal grafting. Principles and practice*. Saunders, Philadelphia, 1984. [Useful text on surgical methods for corneal grafting.]

Courtright P, Johnson GJ, ed. *Prevention of blindness in leprosy*, revised edn. International Centre for Eye Health, London, 1991. [Practical account of main mechanisms of blindness in patients with leprosy, and how it can be prevented.]

Dandona L, Thomas JN, Janarthanan M, Ragu K, Rao GN. Survival analysis and visual outcome in a large series of corneal transplants in India. *British Journal of Ophthalmology* 1997, **81**: 726–31. [Demonstration of good results for corneal grafting for a range of corneal conditions in Asia.]

Eckstein M, Vijaylakshmi P, Gilbert CE, Foster A. A randomised clinical trial of lensectomy versus lens aspiration and primary capsulotomy for children with bilateral cataract in south India. *British Journal of Ophthalmology* 1999; **83**: 524–9. [There was no difference in visual acuity at 3 years follow-up, but lensectomy avoided the possible need for secondary procedures.]

Gray RH, Johnson J, Freedman A. Climatic droplet keratopathy. *Surveys of Ophthalmology* 1992; **36**: 241–53. [Comprehensive survey of present knowledge of this entity.]

Hennig A, Evans JR, Pradhan D *et al.* Randomised controlled trial of anterior chamber intraocular lenses. *Lancet* 1997; **349**: 1129–33. [This large controlled trial showed that modern anterior chamber intraocular lenses are safe after ICCE in a situation of high volume cataract surgery in Asia.]

Jay JL, Allan D. The benefit of early trabeculectomy versus conventional management in primary open angle glaucoma relative to severity of disease. *Eye* 1989; **3**: 528–35. [One of the early demonstrations of the efficacy of primary filtering surgery for open angle glaucoma, compared with long-term medical therapy.]

Johnson GJ, Forfar A. Prevalence, incidence and distribution of visual impairment. In: Johnson GJ, Miuarrian DC, Weale R, eds. *The epidemiology of eye disease*. Chapman and Hall, London, 1998.

Lewallen S. A randomized trial of conjunctival autografting for pterygium in the tropics. *Ophthalmology* 1989; **96**: 1612–14. [This procedure is effective in the setting of a developing country, as well as in industrialized societies.]

Natchiar GN, Thulasiraj RD, Negrel AD, *et al.* The Madurai Intraocular lens study. 1: A randomised clinical trial comparing complications and vitian outcomes of intracapsular cataract extraction and extracapsular cataract extraction with posterior chamber intraocular lens. *American Journal of Ophthalmology* 1998; **125**: 1–35. [There was no statistically significant difference in safety or visual acuity at 1 year follow-up in a very large well-conducted trial.]

Reacher MH, Huber MJE, Canagaratnam R, Alghassany A. A trial of surgery for trichiasis of the upper lid from trachoma. *British Journal of Ophthalmology* 1990; **74**: 109–13. [Bilamellar tarsal rotation was the most efficacious of five procedures in a randomized trial for trachomatous entropion and trichiasis in Oman.]

Roper-Hall MJ. *Stallard's eye surgery*, 7th edn. Wright, London, 1989. [The most recent edition of the widely-used and practical textbook of ophthalmic surgery.]

Sommer A. *Field guide to the detection and control of xerophthalmia*, 2nd edn. World Health Organization, Geneva, 1982. [The key WHO guide to the detection of xerophthalmia in the community and how to manage it.]

Spaeth GL. *Ophthalmic surgery: principles and practice*, 2nd edn. Saunders, Philadelphia, 1990. [A useful, well-illustrated text on ophthalmic surgery.]

Tan DT, Chee SP, Dear KB, Lim AS. Effect of pterygium morphology on pterygium recurrence in a controlled trial comparing conjunctival autografting with bare sclera excision. *Archives of*

Ophthalmology 1997; **115**: 1235–40. [A randomized, controlled trial of the two treatments in a tropical setting in Singapore.]

Thylefors B, Dawson CR, Jones BR, West SK, Taylor HR. A simple system for the assessment of trachoma and its complications. *Bulletin WHO* 1987; **65**: 477–83. [Description of the simplified WHO grading system for trachoma which is now used throughout the world.]

WHO/Prevention of Blindness Programme. *Prevention of childhood blindness*. World Health Organization, Geneva, 1992. [Outline of the main causes of childhood blindness and how it can be prevented.]

Yorston D, Wood M, Foster A. Penetrating keratoplasty in Africa: graft survival and visual outcome. *Brititish Journal of Ophthalmology* 1996; **80**: 890–4. [Keratoconus was the commonest indication for grafting in a mission hospital in Kenya, and good outcomes were obtained.]

47.6 Intrahepatic stones

C. M. Lo, S. T. Fan and J. Wong

Introduction
Aetiology and surgical pathology
Diagnosis and imaging techniques
Treatment
Emergency
Definitive
Conclusion
Further reading

Introduction

Intrahepatic stones are calculi that are located proximal to the confluence of the right and left hepatic ducts. Stones may arise secondarily to an iatrogenic biliary stricture, primary sclerosing cholangitis, or congenital cystic dilatation of intrahepatic ducts (Caroli's disease) or they may originate from the gallbladder. Primary intrahepatic stones or hepatolithiasis are formed *de novo* in the intrahepatic ducts and are a distinct disease entity characterized by repeated primary bacterial infection of the biliary system with subsequent formation of multiple stones and strictures in any part of the biliary tract. The disease was first described in Hong Kong in 1930 and it came to be known as recurrent pyogenic cholangitis. Other SYnonyms include Asian cholangiohepatitis, Hong Kong disease, primary cholangitis, and Asian infestational cholangitis.

Primary intrahepatic stones are prevalent in East Asia. Within this geographic area, in Hong Kong, Taiwan, the southern part of China, Korea, and Japan, intrahepatic stones account for 5 to 52 per cent of all cases of biliary calculi but the incidence has been in decline. As a result of increasing ease of travel, the disease is now encountered more frequently in the West, particularly in cities with large Asian populations. Unlike gallstone disease in Western countries, there is no sex predominance and the peak age incidence is in the third and fourth decades of life. The disease is more common in rural areas and in the lower socioeconomic classes.

Aetiology and surgical pathology

The exact aetiology of primary intrahepatic stones remains unknown, but the initiating process may be the establishment of infection by enteric bacteria, via the portal venous system, in the biliary radicals. Portal bacteraemia secondary to recurrent enteric infection and a compromised defence mechanism due to malnutrition may account for the higher incidence in rural areas and lower socioeconomic classes. Congenital anomalies such as unusual branching of the right posterior or right anterior duct from the left hepatic duct or excessive mucus from an increased number of mucous glands may promote stasis of bile and stone formation. The role of parasitic infestation, such as *Ascaris lumbricoides* or *Clonorchis sinensis*, is not significant.

Repeated attacks of cholangitis are the primary cause for stone formation and other structural changes in the bile ducts. The organisms involved are mainly enteric bacteria, including Gram-negative rods such as *Escherichia coli*,*Klebsiella, Pseudomonas*, and *Proteus*, and Gram-positive cocci such as *Streptococcus faecalis*. More than 50 per cent of positive bile cultures yield a mixed growth, and anaerobic organisms such as *Clostridium and Bacteroides* are present in about 30 per cent of cases, usually mixed with aerobic organisms. Bacterial b-glucuronidase splits bilirubin diglucuronide into the unconjugated bilirubin, which then combines with calcium ion to form the insoluble calcium bilirubinate precipitate. With time, the precipitate consolidates into stones. Over 90 per cent of intrahepatic stones are calcium bilirubinate stones that are composed largely of calcium bilirubinate, and to a lesser extent cholesterol. They are soft, pigmented, and irregular, and are found mainly in intrahepatic ducts, although 60 per cent of patients have stones in the gallbladder or the common bile duct as well. The left ductal SYstem is involved more often than the right, but in 40 per cent of the patients the disease is bilateral.

Transmural ductal injury from infection results in strictures that are often narrow or web-like near the origin of the hepatic ducts and longer in the segmental ducts. Stone impaction and stricture formation in turn aggravate the infection and these changes perpetuate each other. Occasionally, cholangitic liver abscesses develop, and in advanced cases, severe destruction of the liver parenchyma may occur, leading to atrophy of the involved segment and compensatory hypertrophy of the uninvolved segments.

Histologically, there is extensive neutrophilic infiltration of the portal area and hepatic sinusoids during the acute stage. With repeated attacks of infection, the mucosal lining of the bile duct disappears and there is extensive periductal fibrosis and glandular proliferation. The liver plates become shrunken, and biliary cirrhosis is the final picture in long-standing, severe disease. The bile-duct epithelium may also develop dysplasia and a cholangiocarcinoma is found in 5 to 16 per cent of patients.

Diagnosis and imaging techniques

Primary intrahepatic stones are characterized by an intractable course and frequent recurrence. About 60 per cent of patients present with symptoms of acute cholangitis (fever, abdominal pain, and jaundice). The severity of the acute attack varies from a mild, self-limiting incident with minimal symptoms to a rapidly fatal illness with shock and multiorgan failure. Other patients present with upper abdominal pain, acute pancreatitis, or jaundice alone. Physical signs include fever, abdominal tenderness, and jaundice that may be absent when the disease affects mainly the lobar or segmental ducts. Surgical scars from multiple previous biliary operations are common and an enlarged liver may be palpable. During an acute attack, the results of blood tests essentially reflect acute cholangitis and are not in themselves pathognomonic. There is leucocytosis with a left shift, and an elevated serum bilirubin, alkaline phosphatase, and g-glutamyl transpeptidase. Blood cultures yield a positive growth in less than 50 per cent of the cases. Thorough imaging studies are required to delineate the biliary pathology.

Ultrasonography is non-invasive and the accuracy rate is high in expert hands. Dilatation of the biliary tract, ductal stones, pneumobilia, and liver abscess can be detected. Its application in patients with primary intrahepatic stones, however, is frequently hampered by the presence of multiple abdominal scars from previous operations, and stones may be missed if they form casts within the bile duct or when there is no acoustic shadowing. Interpretation may also be difficult because the pneumobilia from a previous bilioenteric anastomosis causes echoes with acoustic shadows mimicking stones.

Computed tomography is largely free from observer bias, and is essential for complicated cases, particularly those with multiple previous operations. It can demonstrate stones in segregated segmental ducts that are not opacified on the cholangiogram, differentiate stones from pneumobilia, raise the suspicion of a concomitant cholangiocarcinoma, and determine the relative atrophy and hypertrophy of liver segments in planning liver resection. As the density of stones on computed tomography varies, some stones may be obscured against the contrast-enhanced liver parenchyma, and a non-contrast scan should always be examined first.

Direct cholangiography is the standard investigation. The extrahepatic ducts are better visualized on endoscopic retrograde cholangiopancreatography and common-duct stones are effectively managed by endoscopic papillotomy. Percutaneous transhepatic cholangiography is required when an end hepaticojejunostomy has been performed or when precise intrahepatic visualization is difficult because of obstruction of segmental ducts. The cholangiographic features of primary intrahepatic stones include ductal dilatation, loss of parallelism, arrowhead formation (abrupt termination of small ducts), excessive branching (more than four branches from each duct), stones, and strictures. When stones or a tight stricture segregate a segmental duct and result in total non-opacification, cholangiography can be deceptive and failure of treatment or ‘recurrence' of disease will result. Thus, the findings of cholangiography should be correlated with those of ultrasonography or computed tomography to delineate fully the extent of disease and to select the best form of treatment.

Treatment

The basic principle of management is to control infection by energetic conservative measures during an acute attack of cholangitis, and to eradicate all stones, strictures, and destroyed liver segments by elective definitive surgery when the disease is quiescent.

Emergency

The initial conservative measures for acute cholangitis consist of broad-spectrum antibiotics to cover both Gram-positive and Gram-negative bacteria, intravenous fluids, potent analgesics, and careful monitoring of abdominal and vital signs. Emergency therapeutic intervention is necessary in 15 to 30 per cent of cases when the

acute attack fails to resolve, as evidenced by septicaemic shock, persistent fever, and spreading peritonitis.

The primary goal of emergency therapeutic intervention is to decompress the biliary tract and to establish free drainage. Emergency endoscopic papillotomy and nasobiliary drainage or stenting may relieve severe acute cholangitis secondary to common-duct stones that are frequently found in patients with intrahepatic stones. Percutaneous transhepatic biliary drainage provides an alternative option for an obstructed segmental duct. Common-duct exploration with T-tube drainage is the most common operation performed. Stones are removed with forceps, a Fogarty catheter, and gentle saline irrigation, and strictures causing ductal obstruction are dilated. The choledochotomy is closed around a large-bore T-tube, and residual stones are dealt with by postoperative choledochoscopic lithotripsy via the T-tube tract. When the patient is haemodynamically unstable and when the complete pathology of the liver and biliary tract have not been delineated by imaging studies, definitive surgical procedures are often not carried out, and reoperations are frequently required. However, in selected patients with persistent fever but stable vital signs, complete radiological investigations are feasible before emergency surgery, and definitive treatment, including hepatic resection, can be performed in one operation.

Definitive

The definitive operative procedure is planned according to the results of all the imaging studies. The principle is to eradicate all stones and strictures, to establish adequate biliary drainage, and to provide an atraumatic, non-operative access to the biliary tract.

Eradication of stones and strictures

The biliary tract is commonly explored through a choledochotomy or hepaticodochotomy extending to the right or left duct when necessary. When the approach to the common bile duct is difficult because of dense adhesions, peridochal varices, strictures, or left-lobe hypertrophy, the biliary tract can be accessed by the extraperitoneal transduodenal approach for sphincteroplasty or by dissection of the left duct or segment III duct. Sometimes, stones palpable over the surface of the liver or located by intraoperative ultrasound can be removed by direct hepatotomy. Stones are removed with forceps, the Fogarty catheter, saline flushing, and baskets guided by choledochoscopy. Flexible choledochoscopy is indispensable and the whole biliary tract should be inspected. Large, impacted stones situated behind relative strictures or within acutely angulated ducts are fragmented with electrohydraulic or laser lithotripsy under choledochoscopic guidance.

The treatment of a stricture depends on its location, severity, and length. Strictures involving the common bile duct, the hepatic ducts or their branches are relieved by a bilioenteric bypass or reconstruction of the strictured orifice. For more proximal strictures, dilatation using graduated sounds or balloon catheters can be performed, but the recurrence rate is high. Segmental liver resection is the treatment of choice for localized disease when the parenchyma of the liver segment involved by intrahepatic stones and strictures has been destroyed. Other indications include cholangitic abscesses, irreversible biliary strictures, and a suspected or confirmed concomitant cholangiocarcinoma.

Non-operative treatments offer attractive alternatives, particularly for elderly patients with prohibitive operative risk and for those with multiple previous operations. Endoscopic papillotomy may be sufficient for mild disease when stones are predominantly located in the extrahepatic biliary tree, but is largely ineffective for intrahepatic stones because of strictures, peripheral stone impaction, or ductal angulation. Percutaneous transhepatic choledochoscopic lithotripsy allows direct access to a stone-bearing duct. By using a choledochoscope inserted through a dilated percutaneous tract, strictures are dilated, and with the aid of electrohydraulic or laser lithotripsy, stones are fragmented and removed.

Establishment of biliary drainage

Despite various advances in therapeutic techniques and innovative approaches to the biliary tract, recurrence is common in patients with primary intrahepatic stones. Biliary drainage procedures may prevent recurrence by reducing biliary stasis and promoting spontaneous passage of residual or newly formed stones. Sphincteroplasty, as a surgical drainage procedure for patients with predominantly extrahepatic disease, has largely been replaced by endoscopic papillotomy. Supraduodenal choledochoduodenostomy does not provide effective biliary drainage and is often complicated by the development of reflux cholangitis and sump syndrome. End-to-side hepatico- or choledochojejunostomy using a Roux-en-Y jejunal loop is the biliary drainage procedure of choice. When a stricture is present in the more proximal part of the biliary tract, drainage should be achieved by dilatation of the stricture or a bilioenteric anastomosis at a more proximal level.

Provision of atraumatic biliary access

Surgical treatment of residual or recurrent intrahepatic stones imposes a major problem because reoperation becomes increasingly difficult and hazardous. A matured, dilated tract that is established by a percutaneous transhepatic biliary drainage catheter or a T-tube placed operatively provides a temporary access to the biliary tract. Although flexible choledochoscopy through the tract can clear residual or recurrent stones with a success rate over 90 per cent, the procedure can only be undertaken after maturation and dilatation of the tract, and disruption of the fibrous tract is not uncommon. A permanent access to the biliary tract can be achieved with a bilioenteric anastomosis using a Roux-en-Y jejunal loop, the terminal end being brought out as a cutaneous stoma (hepaticocutaneous jejunostomy). Repeated sessions of choledochoscopy can be performed via the stoma until all stones are removed, and strictures dilated before the stoma is closed. When recurrent disease is suspected, the stoma can be reconstituted under local anaesthesia without the need for a laparotomy. The operative procedure is simple and the long-term result excellent.

Conclusion

In summary, primary intrahepatic stones represent a distinct disease entity of obscure aetiology that should not be considered one of interest for Asia alone. The ductal and parenchymal pathology needs to be delineated by modern imaging techniques including ultrasonography, computed tomography, and direct cholangiography. The management of this disease has evolved into a multidisciplinary team approach involving surgery, therapeutic endoscopy, and interventional radiology.

Further reading

Fan ST, Choi TK, Lo CM, Mok FPT, Lai ECS, Wong J. Treatment of hepatolithiasis: improvement of result by a systemic approach. *Surgery* 1991; **109**: 474–80. [A large case series of 137 patients with hepatolithiasis treated with a systematic approach.]

Fan ST, Lai ECS, Wong J. Hepatic resection for hepatolithiasis. *Archives of Surgery* 1993; **128**: 1070–4. [A detailed description of the indications, technique and results of hepatic resection for intrahepatic stones.]

Fan ST, Mok F, Zheng SS, Lai ECS, Lo CM, Wong J. Appraisal of hepaticocutaneous jejunostomy in the management of hepatolithiasis. *American Journal of Surgery* 1993; **165**: 332–5. [Evaluation of the role of hepaticocutaneous jejunostomy in the treatment of residual or recurrent intrahepatic stones.]

Kim MH, Sekijima J, Lee SP. Primary intrahepatic stones. *American Journal of Gastroenterology* 1995; **90**: 540–8. [Review article with special emphasis on the epidemiology and pathogenesis of intrahepatic stones.]

Liu CL, Fan ST, Wong J. Primary biliary stones: diagnosis and management. *World Journal of Surgery* 1998; **22**: 1162–6. [An updated series of 96 patients with intrahepatic stones treated with a systematic approach.]

Lo CM, Lai ECS. Intrahepatic lithiasis and Asian cholangitis. In: Philips EH, ed. *Problems in general surgery*, Vol. 12 *Common bile duct stones in the era of laparoscopy*, pp. 115–23. Lippincott-Raven, Philadelphia, 1996. [A comprehensive overview of intrahepatic stones.]

Pitt HA *et al.* Intrahepatic stones. The transhepatic team approach. *Annals of Surgery* 1994; **219**: 527–37. [A case series of 54 patients with intrahepatic stones treated with a combined radiological and surgical approach with transhepatic stents.]

47.7 Some special aspects of plastic surgery in developing countries

T. E. E. Goodacre

Aspects of plastic surgery
Requirements for plastic surgery
Application of plastic surgery to specific conditions in developing countries
Wound management
Cancer management
Congenital deformity
Hand surgery
Specific infections
Further reading

Aspects of plastic surgery

Reconstructive plastic surgery has a more significant role in the surgical management of trauma, sepsis, neoplasia, and congenital deformity in the developing world than in more highly resourced settings. Hospitals with basic facilities may lack access to many antibiotics, regular opiate analgesics, chemotherapeutic agents, sophisticated dressings, and nurses and therapists skilled in rehabilitation or palliative care. Radiotherapy and other specialized facilities may be only available at a distant capital city and limited to a fraction of those who might benefit from them. In these circumstances, surgical treatment and reconstruction may be the only available option for many conditions, especially in the rural areas of poorer countries. This logistical problem is exacerbated by the late presentation of many patients in such areas: advanced sepsis, malignancy, and contractures are very destructive of normal function and considerable tissue deficiency is commonly encountered. The associated pain, disability, and loss of achievement of potential within society may be ameliorated in many cases if simple reconstructive surgical techniques are used

Requirements for plastic surgery

Regrettably, reconstructive surgical methods are often thought to be complex and plastic surgery is taught as a postgraduate discipline only. This need not be so, and the ability to use skin grafts and simple flaps should be within the remit of any medical-school graduate who might have to practise in a rural hospital setting. With basic anaesthesia, good light, and a sharp knife, most plastic surgery can be undertaken with excellent results when good basic surgical principles are followed. The most important requirement is an understanding of how tissues behave when moved, together with the surgical confidence to solve problems in diverse conditions with well-established plastic methods (Fig. 1).

TABLE cellSpacing=0 cellPadding=0 align=left border=0 hspace="10" vspace="5">



Fig. 1. Sophisticated equipment is not essential for the practice of much good reconstructive plastic surgery. This scene from a poorly resourced hospital in Africa with a draw-over anaesthetic vaporizer and foot suction pump is a setting for satisfactory surgery.

Surgical emphasis: flap rather than graft if possible—single rather than multiple stages preferred.

Although many reconstructive options involve multiple stages, or grafting techniques that are vulnerable in the early postoperative period, single-stage flap reconstructions can avoid many pitfalls and greatly improve success rates. Use of the so-called reconstructive ladder of increasing complexity of procedure (teaching the use of simple grafts before random flaps and so on to more complex flaps) may need to be balanced against the benefits of single-stage operations that facilitate quick recovery, do not need specialized nursing, and make rehabilitation easy. A balance will need to be struck between ‘keeping it simple' and using better methods to reap much greater success.

Application of plastic surgery to specific conditions in developing countries

Wound management

Burn injury

Burn injury is extremely common in the developing world, particularly where cooking is commonly on open fires and within communal living spaces, where inflammable fuels are used in minimally regulated ways, and where public health measures are not yet able to reduce the dangerous use of certain electrical appliances. First aid, particularly the rapid cooling of burns with copious clean cold water, is not widely practised, and the depth of burn is usually greater than seen in the developed world. Without doubt the most pressing need for burn care in the developing world is the widespread adoption of public health measures, including education in burn prevention and first aid.

In rural areas a very simple assessment of burn area and depth may be all that is possible, but if done carefully, a simple formula burns' fluid can be used to give crystalloid fluid intravenously for resuscitation. Many small children die each year from the over-administration of fluid too rapidly in the resuscitation phase, which produces cerebral oedema. Poor populations may also not have tetanus cover, which should be given on admission if the patient's immunization status is unknown. The dense, constricting, burnt tissue around limbs (the eschar) must be divided within 2 h of the burn, is a painless procedure if the skin is full-thickness burnt, and can save a hand or foot distally. Such escharotomies should be taught as first-aid procedures to rural medical assistants (Fig. 2).



Fig. 2. Patient with a full-thickness burn of left forearm with a constricting eschar that is restricting blood flow to distal parts.

What is the best dressing for burns is controversial. Ultimately, intact skin cover should be the goal as soon as possible; for full-thickness burns, this may mean early excision and skin grafting if it can be safely done with a good likelihood of successful take. Superficial burns will heal well if kept clean with undisturbed, clean dressings or frequent washing with clean water. Exposure of burns dries the surface and prevents healing over, but has the advantage of preventing the multiplication of bacteria. Commercial dressings are usually expensive and often unavailable. Many traditional remedies are very acceptable and may even have properties that promote wound healing (such as honey and many leaf preparations); others (such as dried dung used in some societies) contain many harmful organisms and can be catastrophic.

The goal of full skin cover (whether by normal healing or grafting) should be accompanied by the maintenance of a full range of movement by careful splintage across joints and regular exercises. Severe burns, or those of critical body areas (face, hands, genitalia, across all joints) should be grafted, or covered with flap tissue if nerves, tendons, or bone are exposed (Fig. 3, Fig. 4 and Fig. 5).

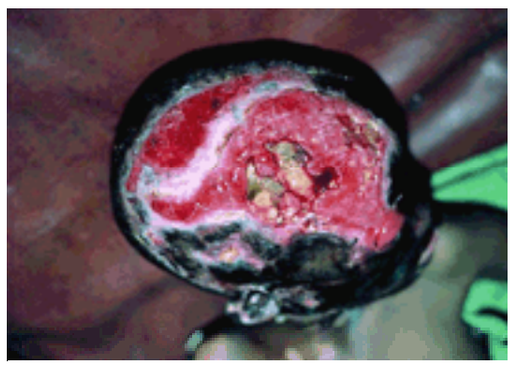


Fig. 3. Exposed cranial bone following severe burn to scalp after falling in a domestic fire.



Fig. 4. Patient illustrated in Fig. 3 having had residual scalp advanced to cover exposed cranium and the secondary defect split-skin grafted.

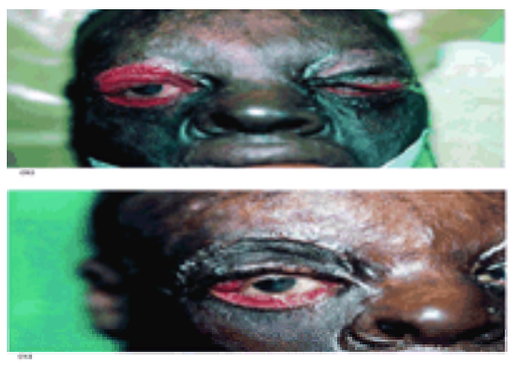


Fig. 5. Severe facial burns causing ectropion of the eyelids (a) before and (b) after split-skin graft to release right upper eyelid (this graft, albeit freshly taken and somewhat bulky, has enabled protective cover of the globe of the eye to be obtained).

Principles of skin grafting in poorly resourced settings

Split-skin grafts can be harvested without complex instruments and under local anaesthetic if no more than about 600 cm^2 is needed. Thin grafts can be taken without guarded blades if the skin is kept taut by an assistant and the knife is sharp and well lubricated. Grafts should be well perforated throughout to allow exudate to escape and expansion if meshed. Graft will take best on a well-prepared, non-oozing surface; freshly excised burns should be left for 15 min for the bleeding to stop before applying graft, which should stick rapidly. In hot climates, many grafts are lost because they are wrapped up for days, during which any contaminating organisms multiply. They would be more successful if left exposed but protected from disturbance by the patient or contamination by flies. Grafts on small children usually need dressing, but changes of dressing and washing should be frequent.

Management of contractures

Burn contracture is a major cause of disability, disfigurement, and failure to achieve potential for large numbers of people throughout the developing world. Contractures that come from inadequately splinted, immobilized parts (such as foot drop after leg injury) can be greatly improved by persistent and careful physiotherapy. However, established contractures following major skin loss should be released surgically in most instances. They recur rapidly if the exposed tissues (which may include nerves, tendons, and joints) are not covered with durable, non-contractile tissue. Flexor surfaces across joints should be lined with thicker split-skin graft, or preferably flap tissue. Linear contractures are best released with multiple sickle flaps or multiple Y–V plasties. These simple local flaps bring in soft, elastic, and pliable tissue from outside the burn scar and make the best use of the tissue store available near to the injury.

Severe contractures, which may even leave a structure such as the wrist at 180° from its normal position, may have to be released in stages, with good splints made after each operation to maintain the benefit obtained, before the new flap or graft skin heals fully and further release can be attempted. This staged method also allows contracted nerves and other blood vessels to stretch out again and avoids the need for complicated nerve and vessel grafts (Fig. 6, Fig. 7).

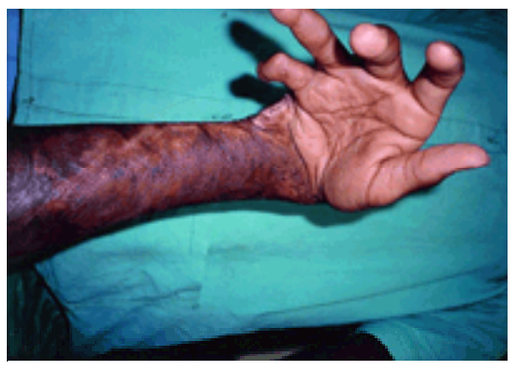


Fig. 6. Severe post-burn contracture of the left hand with the little finger buried in severe ulnar flexion in the wrist.



Fig. 7. Release of linear burn contracture on the lateral aspect of the left lower leg with multiple Y–V plasty.

Limb injury

Severe, compound limb injuries occur more frequently with the increase in car transport and particularly when traffic regulations are poor. Firearm and blast injuries are also common in areas where war is endemic or in which weapons have proliferated and are used in crime. The most important guidance for managing such injuries in developing countries is to adhere strictly to the basic principles of wound excision and washing, and to avoid all temptation to adopt expensive operative techniques of bone fixation and surgical cover, which are often promoted heavily by commercial medical interests. Such methods still produce disastrous consequences when poorly used in developed settings, and require huge investment in training and infrastructure if they are to be adopted safely. The closed treatment of fractures is dealt with elsewhere, but optimal management of soft tissues frequently requires plastic techniques.

Badly contaminated or mangled soft tissues should be excised surgically with a clear margin to undamaged tissue (the so-called pseudotumour approach to debridement), with the exception of nerves and the specialized areas of skin that are irreplaceable: these include the lip and eyelid skin, ear and nose margins, genitalia, palm and sole skin, nail beds, and hair-bearing scalp and eyebrows. After repeated excisions until the wound is clean, the area can be split-skin grafted if granulating or covered with flaps if healthy, clean, deeper structures are involved. During the process of preparing a wound for delayed secondary cover, it is often best washed three to four times a day with sterile saline and kept moist beneath the dressing. Alternatively, a variety of dressings may be used to aid wound cleaning, including larvae (maggots), which eat only dead tissue, simple non-adherent dressings, or honey and plant agents (see burn dressings above).

Principles of wound cover for limbs

Skin can be replaced with split graft so long as the underlying bed has a good blood supply, is not infected, and is not mobile. Exposed bone may granulate if kept moist, but dry bone should be excised and covered with skin with a blood supply (a flap). Areas of skin loss that will result in contracture (such as the tendo Achillis and joint flexor surfaces) also need flap cover. It is important to remember that the tissue next to an area of wound debridement may itself have been severely crushed or involved in the so-called zone of trauma and may respond badly to being moved as a flap. It may also heal with dense, fibrous scar tissue and not bring the benefits of a healthy blood supply to the wounded area that come from cover with a good flap from an uninvolved site.

Local flaps on limbs will usually be axially based, transposition flaps. The long, posterior calf, fasciocutaneous flap (the so-called super flap described by Ponten in 1981) is useful for many upper- and middle-third fractures of the lower leg. Upper-third fractures can also be easily covered with transfer of half the gastrocnemius muscle, which is in turn covered with split-skin graft.

Tissue loss in the lower third of the leg causes the most problems in obtaining cover. The ideal tissue is usually distant muscle, which requires free microvascular transfer that is rarely available in developing countries. Alternatives are distally based fasciocutaneous flaps (only suitable if the distal blood supply is unaffected by the injury), distally based soleus muscle flaps (which are very damaging to the remaining tissues of the lower leg and should be avoided if at all possible), and cross-leg flaps (which can be successful, but require meticulous technique and leg immobilization). These difficult injuries, which are all too common, are probably best prepared for evacuation by careful wound toilet and transferred to a unit with reconstructive facilities if at all possible.

Facial injury

Facial injuries in the developing world are often seen late, when much facial oedema has developed and accurate diagnosis of fractures can be difficult. Careful assessment of penetrating lacerations for underlying fractures, and of disordered bite from fractures of the jaws, is essential. For more severe injuries, assessment of the cranial nerves and eye function can give vital information on maxillary fractures. If the airway has been secured and there is no leakage of cerebrospinal fluid, it is quite acceptable in many rural situations to stabilize the patient for a few days whilst the oedema settles, with subsequent definitive correction of the fractures before associated fibrosis of soft tissues makes bony reduction impossible and the resultant visual and cosmetic disorders resistant to improvement (Fig. 8).

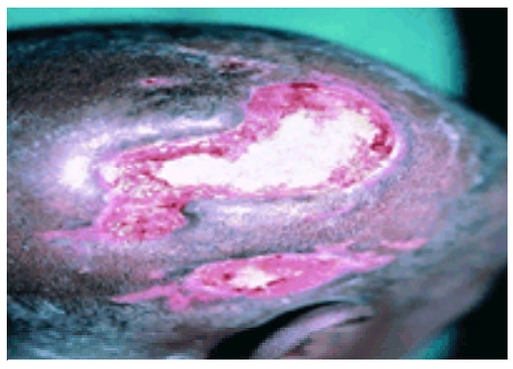


Fig. 8. Scalp loss with exposed pericranium following a bite from a hyena. This defect will require a simple scalp transposition flap to cover the exposed bone, with split-skin grafting of the secondary defect (see also Fig. 5).

Severe jaw loss (usually after gunshot injury) presents a formidable reconstructive problem after the initial goals of wound excision and maintenance of a good airway (usually with a tracheostomy) have been obtained. In the absence of the means to rebuild the mandible and its skin cover immediately (with a microsurgically

transferred osseocutaneous flap), the best strategy is usually to salvage as much skin and tongue as possible, closing in layers to form a ‘sling' to the mouth. Oral continence is difficult to obtain primarily, but can be achieved with a sling of fascia from the tensor fascia lata if sepsis is not a significant problem. Bony reconstruction involves complex surgery whatever method is used, and is dependent upon good cover and lining with soft tissue.

Major loss of soft tissue is seen often after bites from animals or humans. After initial wound excision, reconstruction for animal bites is best undertaken primarily; human bites carry more problems from infection, and repair should be undertaken when this has been controlled at about 2 weeks. Although the best option for wound reconstruction is early surgery (preventing the development of dense contracture), if a rural hospital has no access to surgical expertise at all, then it is better to avoid creating further bad scars on the face in incorrect places that may rule out the ideal procedure for a surgeon tackling the problem later. Ears are best not reconstructed primarily in rural settings. Lips, nose, and eyelids do, however, benefit from early surgery.

War injury

The incidence of landmine and high-velocity gunshot injuries is increasing across the conflict areas of the developing world. Excision of the soft-tissue wound leaves large cavities, often with no bony support. The whole repertoire of plastic procedures may be required to manage the secondary defect after wound healing. The reader is referred to the guidance from the International Committee of the Red Cross produced by Coupland for the best management of landmine injuries, which comprises the characteristic pattern of lower and upper limb loss, genital injury, and blindness. Delayed primary reconstruction of large skin defects of the limbs in war zones is best done with pedicled muscle and fasciocutaneous flaps (Fig. 9).



Fig. 9. Rural farm worker with a bilateral amputation and loss of eyesight after landmine injury.

Cancer management

Skin and soft-tissue cancer

Skin cancer is most frequently seen in areas of chronically scarred or irritated skin, the so-called Marjolin ulcer (Fig. 10). Chronic tropical ulcers (see later) often produce this condition after 10 to 15 years, and it can be prevented by adequate excision and skin grafting of the ulcer prior to cancer development. Once the squamous carcinoma has developed, amputation need not be (as once taught) the inevitable sequel. If there are no signs of spread to regional nodes, it is acceptable to excise the lesion widely and cover the defect with graft or flap.

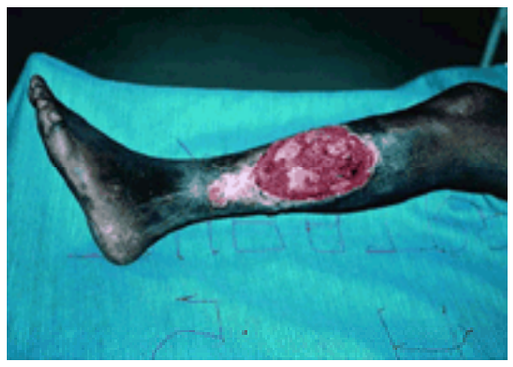


Fig. 10. Large tropical ulcer of lateral lower leg with squamous-cell carcinomatous transformation (Marjolin ulcer).

Squamous carcinoma in children is not infrequently seen in families with xeroderma pigmentosum (dominantly inherited). Such tumours are usually highly aggressive and require urgent wide excision and reconstruction (Fig. 11).



Fig. 11. (a and b) Two siblings with fungating squamous-cell carcinoma of the scalp in a background of underlying xeroderma pigmentosum. Surgical excision and split-skin grafting of the defect is the only available treatment in many rural areas away from access to radiotherapy and may offer good palliation if not cure.

Malignant melanoma is rarely seen in populations with heavily pigmented skin; the usual site in rare cases is the non-pigmented skin of the soles and palms, and the most common type of melanoma is the acral lentiginous lesion. Management is no different from that in any other setting.

Facial cancer

Many of the most obviously neglected cancers seen in the developing world are of the face, because of the difficult problems raised in reconstruction and the absence of radiotherapy, chemotherapy, and sufficient quantities of powerful analgesics. In the young, it is vital to recognize which lesions should not be treated with surgery, particularly the Burkitt lymphoma. Precious supplies of chemotherapeutic agents should be conserved for these cases, in which spectacularly successful treatment can be obtained by medical means alone.

For many other facial tumours, however, the judicious use of radical surgery (maxillectomy, mandibulectomy, glossectomy, or other major resection) can be made

entirely acceptable and possible in small hospitals with limited resources if the means to reconstruct the defect safely are available. Usually, it is preferable in such settings to avoid complex bony reconstruction and it is valuable to remember that a very acceptable quality of life can be achieved without a reconstructed mandible (particularly if the alternative was a painful, ulcerative wound with no other means of palliation) (Fig. 12, Fig. 13).



Fig. 12. Carcinoma of the maxillary antrum: in this isolated rural setting with relatively limited surgical facilities, surgical resection with maxillectomy and skin grafting of the defect is the only feasible palliative/curative option.



Fig. 13. (a) Toung child with massive ulcerating jaw tumour (non-Burkitt) unresponsive to the simple chemotherapeutic agents; (b) palliative resection and deltopectoral flap cover undertaken in a small, isolated bush hospital.

Advanced conditions—palliative surgery

Massively advanced tumours are commonly seen throughout the developing world. Surgical excision with reconstruction is often the sole means of treatment open to the patient and the value of palliation of odious tumour signs cannot be overemphasized. Access to a wide range of flap reconstructions can be liberating for the surgeon who wishes to excise a large tumour. Whilst not advocating an excessively optimistic attitude to surgery, there is frequently much to be offered in rural areas from a bold attempt to excise and reconstruct a tumour that has lead to social ostracism and despair for the sufferer (Fig. 14).



Fig. 14. Massive fungating thyroid cancer. Tumours such as this are unlikely to receive palliative radiotherapy other than in privileged settings. Palliative surgical resection and cover should always be considered even in remote settings, particularly if some blood for transfusing can be obtained.

Congenital deformity

Facial clefting

Facial clefts result in cosmetic and functional deformities that far exceed the physical anomaly, because of the prominent position of the defect. Clefts of the palate are associated with difficulties in sucking at birth and may be responsible for some of the lower incidences in this aspect of clefting in poorer parts of the world without access to feeding support. Lip clefts, however, produce far fewer functional problems and therefore do not usually affect survival. Children and adults with unrepaired clefts are still seen frequently in certain parts of the world. Efforts should be directed towards providing suitably trained personnel in a few centres in each country to give a sustainable service to repair such defects and the multidisciplinary team care required throughout the patients' childhoods if they are to have an acceptable appearance and speak well enough to avoid the stigmatization that so often accompanies this condition.

It should be stressed that repair of a cleft lip with poor technique may often be irredeemable surgically and the inexperienced operator should not be encouraged to ‘dabble'. Hodges has shown that a good service can be provided throughout a country, including teaching for the service to be sustainable, on a tiny budget, because cleft repairs, as with so many other plastic techniques, require little in the way of material resources but are very dependent upon the understanding of an appropriate surgical method. In the meanwhile, some outreach teams do provide good surgical repairs on a sporadic basis in many countries of the world (Interplast, Operation Smile, etc.) (Fig. 15).

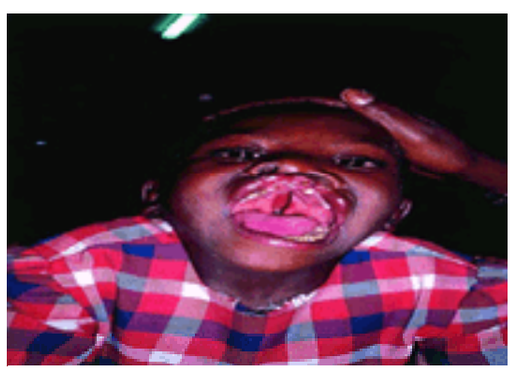


Fig. 15. Unrepaired cleft lip and palate in child considered mentally handicapped because of speech disorder secondary to cleft palate—intellectually entirely normal.

Limb anomalies

Congenital abnormalities of the hands are common and should be corrected, if possible, if the hand can be ‘upgraded' by such an operation. Excision of simple extra digits on the ulnar side of the hand, and separation of simple SYndactyly, can both be undertaken easily in rural hospitals; these account for the majority of hand anomalies. More complex disorders, such as duplex thumbs, missing rays, club hands, and disorders of the brachial plexus, should be referred for specialist treatment only. There is a clear role, however, for tendon transfers for isolated nerve palsies, which may present to hospitals in rural areas. Such procedures can be done with regional anaesthetic blocks, but require careful preoperative physiotherapy and preparation, together with good balancing at surgery followed by meticulous splintage and physiotherapy.

Anomalies of the lower limbs are usually more complex to manage. Correction of club foot is an orthopaedic procedure that is highly beneficial in correctly selected cases and can be done in a well-controlled rural hospital. Minor anomalies in digits or skin of the foot are usually best left intact unless they produce functional problems (which is rare).

Urogenital

The most common urogenital anomaly, hypospadias, can be corrected routinely in rural hospitals without resort to complex prepucial flap techniques. If necessary for social reasons, distal hypospadias can be corrected in a single stage with a meatal advancement [the meatal advancement and glanuloplasty (MAGPI) operation]. More proximal hypospadias should be released and chordee corrected to ensure straight erection. The defect can be very satisfactorily restored with inlay of a full-thickness skin graft, which is tubed and skin-covered at a second operation after 6 months.

More severe, proximal bladder extrophies are very rare and are best referred for specialist care.

Hand surgery

Infections and bites

The hand assumes a level of importance in developing countries over and above that in more affluent areas, because of the high proportion of those whose sheer survival rests upon manual labour. The hand also suffers from more injury and infection when cooking is on open fires, machines are often unprotected, penetrating foreign bodies abound in the workplace, and bites and stings are much more frequent. Delayed presentation of infected hands leads to devastating consequences for tendons in sheaths and palmar spaces, and the lack of splintage for the injured and infected hand rapidly results in lost function. Basic teaching on the value of high elevation of the affected hand and splintage in the ‘neutral' position should be widely spread amongst primary health-care workers as well as hospital providers and the general public. Much more public education of such simple principles is required before the incidence of tragic loss of hand function is reduced throughout the developing world (Fig. 16).



Fig. 16. Neglected infection of the palm of the right hand with ulceration of skin and ‘frozen palm'.

Bites to the hand by venomous snakes are common and, if the patient survives the initial toxic effects, may lead to considerable swelling of the soft tissues; this can be alleviated by elevation, but may require fasciotomy to reduce pressure within the deeper compartments of the hand. Fasciotomy is rarely required within the first 12 to 18 h, but the hand should be inspected regularly to detect early signs of the tense, oedematous swelling that is characteristic of injury with viper bites.

Hand injuries

The management of specific hand injuries is a vast subject beyond the scope of this account. However, the great incidence of hand injuries in the developing world merits a brief comment. The vital importance of the hand to human social function is often not fully appreciated, and the absence of many health-and-safety-at-work precautions and regulations leads to a massive amount of morbidity from this cause in newly industrialized regions. First aid for the hand, including adequate early washing of the wound and splintage in the functional position, should be widely promoted in developing countries. Subsequent wound management should take account of the value of early mobilization with fully covered skin, even if the early management of a contaminated wound has correctly been to leave it open for some days and use delayed primary closure. A repertoire of simple, local and pedicled flaps makes such a treatment plan easy to adopt in most cases and the neglected, open, slowly granulating wound in a painful ‘frozen' hand should be avoidable, even in very basic health centres.

Specific infections

Cancrum oris

This highly destructive condition of the face and jaws is thought to be a form of synergistic gangrene, which can rapidly destroy massive areas of soft tissue and bone. It is most commonly seen in young, malnourished children; early treatment is with antibiotics and nutritional support. Presentation is usually late, however, and survivors of the acute stage should be managed conservatively whilst the necrotic tissues become demarcated and all available vital tissue is preserved. After obvious separation has occurred, surgical toilet and sequestrectomy of the dead bone can be undertaken to reveal the full extent of the destruction that has been inflicted (Fig. 17).



Fig. 17. Child with cancrum oris of chin and mandible.

Reconstruction is then a matter of planning the staged release of ankylosis of the mandible, replacement of the buccal lining, and cover of the cheek defect with a full-thickness flap. In most circumstances where this condition is seen, flaps must be relatively straightforward and follow the principles outlined above. Adams-Ray and James describe the use of platysma neck ‘turn up' flaps for lining and extended deltopectoral flaps for cover in the majority of patients that they saw. After restoration of adequate jaw opening and continent feeding without salivary leakage, dental care is vital since many of these children have severe damage to the teeth.

Tropical ulcers

Ulceration of the lower leg from non-vascular causes is still relatively common in rural areas of poorer countries. It can be seen in children as young as 10 years old and the pathology is still uncertain. Some ulcers are undoubtedly colonized with Vincent's and fusiform bacilli, but once chronically established, they resemble any other form of chronic leg ulcer. Infection with *Mycobacterium ulcerans* produces a different lesion, the Burulli ulcer, which is only seen in well-defined areas. However, these lesions benefit from surgical intervention as well as chemotherapy and are therefore dealt with under this section (Fig. 18).



Fig. 18. Necrotizing fasciitis of lower leg—synergistic gangrene.

Tropical ulcers may exist for many years without healing, and development of squamous-cell carcinoma in the margins (Marjolin ulcer) is common, as described above. Diagnosis without histopathology is difficult and this differential diagnosis should be kept clearly in mind, even in young people.

Surgery can play a significant part in managing these ulcers, since access to dressings with compression, meticulously applied, is a rarity in poorer countries. The established ulcer will usually not accept a skin graft without excision of the fibrous base and this is best undertaken with a tourniquet. Tangential shaving with a skin-graft knife set widely open can preserve as much vital tissue as possible. Following this shave excision, the bed is left for several minutes with compresses applied to reduce bleeding after the larger vessels have been tied off. Perforated or meshed split-skin graft is then applied, which is also haemostatic. Additional graft should be harvested and stored in a conventional refrigerator to replace any early graft failure. Once epithelialized, grafts on the lower leg require some external compressive support, which is often given by a simple elastic tubular bandage.

Mycetoma

Fungal infection of the extremity may lead to the development ofa large destructive mycetoma, one variety of which is common-ly known as the ‘Madura foot'. The chronic granulomatous swellings are relatively painless and so presentation is usually late, especially amongst the barefoot and manual workers who are most prone to the condition. Treatment of the condition includes chemotherapy, but the fungal involvement respects no planes, and may have damaged the bones, tendon sheaths, and regional nodes (Fig. 19, Fig. 20).

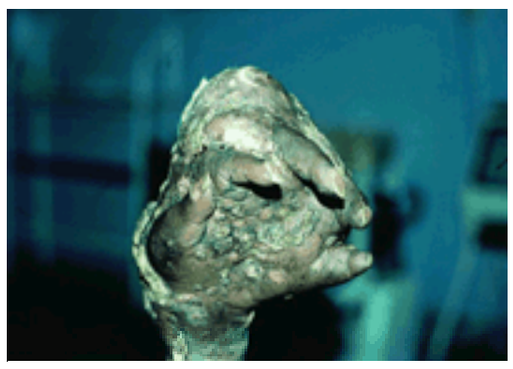


Fig. 19. Mycetoma of left hand: not surgically salvageable prior to amputation.



Fig. 20. Mycetoma of left foot: a less severely involved limb than in Fig. 19—the classical appearance of the so-called Madura foot. This may be relatively asymptomatic but surgical salvage can be attempted with local excision and flap cover.

Surgery is indicated for symptomatic control, and also to prevent distant spread of the disease. Localized lesions can be opened and curetted. Traditional teaching is that major infections which have left the distal limb severely damaged should be amputated. However, some authorities report successful reconstruction of widely excised areas with large, pedicled, lower-limb flaps.

Further reading

Adams-Ray WE, James JH. Cancrum oris: functional and cosmetic reconstruction in patients with ankylosis of the jaws. *British Journal of Plastic Surgery* 1992; **45**: 193–8.

Bodeker G, Parry E. New approaches in traditional medicine. *Tropical Doctor* 1997; **27** (Suppl.1). [Recent advice on traditional materials for dressings.]

Coupland RM. *War wounds of limbs; surgical management.* Butterworth-Heinemann, Oxford, 1993. [Authorized Red Cross advice on managing land mine wounds and other injuries.]

Goodacre TEE. Plastic surgery in a rural African hospital; spectrum and implications. *Annals of the Royal College of Surgeons of England* 1986; **68**: 42–4. [A review of the spectrum of conditions to be found in typical rural practice.]

King M, Bewes P, ed. *Primary surgery*, Vol. 1, *Non-trauma*; Vol. 2, *Trauma.* Oxford Medical Publications, 1990. [The definitive works on ‘low tech' primary care surgery within the developing world.]

Loeffler I. Surgery in Africa in the year 2000. *Proceedings of the Association of Surgeons of East Africa* 1986; **9** (Suppl). [A very pertinent review of the requirements for good surgery in situations with limited resources]

Ponten B. The fasciocutaneous flap: its use in soft tissue defects of the lower leg. *British Journal of Plastic Surgery* 1981; **34**: 215.

47.8 Surgery of advanced disease and late presentation

A. S. Daar and Arjuna Aluwihare

Introduction
Understanding late presentation
Before the surgical consultation
Late presentation: carcinoma of breast as an illustration
Sociology of late presentation
Preventing late presentation and avoiding advanced disease
Avoidance of known aetiological factors: cardiovascular disease as an illustration
Screening: colorectal and breast carcinoma as illustrations
Management of patients presenting with advanced disease
General principles
Palliation versus cure: colorectal malignancies as an illustration
Where do patients die? The hospice movement
Communication and cultural sensitivity
Follow-up (clinical review) for malignant disease: colorectal carcinoma as an example
Specific disease conditions—specific technical aspects
Neonates, infants, and children
Trauma and burns
Acute conditions
Neoplasms
Non-acute surgical conditions
Further reading

Introduction

Surgical disease may present late or in an advanced pathological state because of the intrinsic nature of the condition, because of personal, social, economic and cultural factors, or because of a combination of these factors. At presentation, some conditions are by definition advanced, for example vesicovaginal fistulae and end-stage organ failure. Others only become symptomatic when already fairly advanced, for example colorectal and bronchial carcinoma. All disease may present late because of defects in the health services, for example a lack of enough specialists, lack of health education/awareness, clerical errors, long waiting lists, etc. Because late presentation correlates with poorer outcomes in most situations, it is best prevented, regardless of how good the surgeon, the team, or the facilities may be.

We therefore need to consider (a) how to prevent late presentation where this makes a difference and (b) elicit the principles of management once a condition presents in an advanced state. To do this, we will first look at factors, both personal and sociocultural (Table 1), that contribute to late presentation, that is before the definitive diagnosis and management are established. The second part of the chapter will deal with the management of patients presenting with advanced disease, eliciting the major principles (Table 2), and then examining some specific diseases and categories.

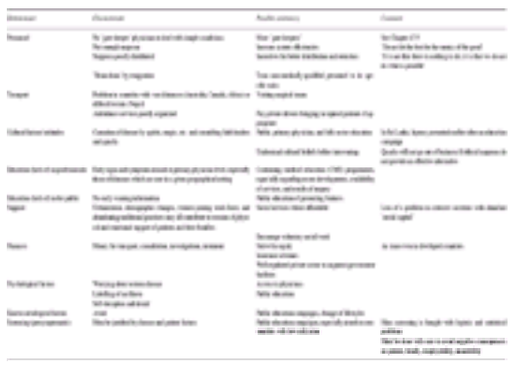


Table 1 Determinants of late presentation

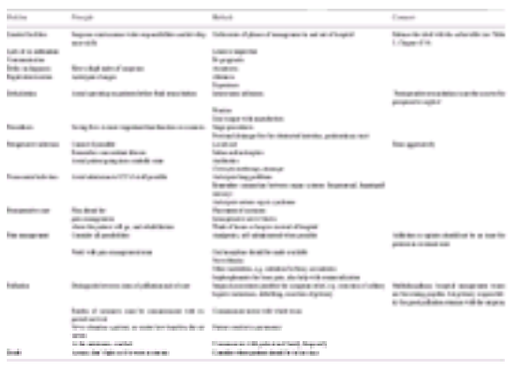


Table 2 Some management principles in patients presenting with advanced surgical disease

Late presentation and advanced disease are problems everywhere in the world; however, the problems are more numerous and of a greater magnitude in countries with limited resources, and the chapter is therefore slanted somewhat towards the realities of developing countries. The surgical principles, however, are applicable wherever in the world a surgeon has to deal with such patients. We have emphasized mainly those points which tend to be ignored or overlooked by very busy surgeons and have tried to use common surgical conditions as illustrations. Because many developing countries have a preponderance of the young (up to 55 per cent under the age of 15 years in some countries) and a major shortage of trained paediatric surgeons, we have deliberately tried to incorporate some more basic paediatric surgical principles in the appropriate sections.

Understanding late presentation

Before the surgical consultation

Personnel

Nordberg *et al.*, working in Africa, have documented what many surgeons in developing countries in Africa, Asia, and South America know only too well: there are just not enough surgeons to deal with the work load. The end result, unfortunately, is non-treatment, late presentation, and advanced disease. There is no easy answer to this complex resource problem, for it is not only a matter of numbers but of distribution within and across countries. Surgeons tend to migrate to other countries, to where the facilities exist to operate well, educate children well, and live comfortably; they move to where they will receive reasonable remuneration and to where they will be treated with dignity and courtesy.

Besides improving efficiencies, one of the solutions that has been recommended is to have several tiers of personnel (not necessarily ‘medically' qualified), each with a

strictly limited role in diagnosis and treatment. Such solutions have been used most successfully in China and Tanzania; in Bangladesh they have been used especially for contraception by sterilization. Surgeons often divide their work into modules, for example one may open an abdomen, another do the difficult part of an operation, and a third close the wound. The different parts of the work are generally performed by persons with different levels of skill. Having several tiers of personnel is only a variation of the module principle. Such personnel can deal not only with simple problems such as abscesses, wounds, and ulcers but, when trained well, with even major conditions. Even if not the ideal solution, the real world dictates that we must not allow the best to be the enemy of the good.

In our experience, inappropriately trained indigenous practitioners and quacks contribute to the late presentation of neoplasms, snake bites, osteomyelitis, elbow dislocation, and many other common conditions. One quack method of managing oral and other surface cancers, for example, is to apply proteolytic enzymes to reduce the size of a tumour. The surgeon is then faced not only with delay in diagnosis, but with the management of iatrogenic disease resulting from incorrect diagnosis and wrong management. Working with such practitioners so as to improve their skills, when possible, is one way to harness their potential for good and minimize their potential for harm. Those who persistently do harm and cannot be educated should be put out of business, but this will only work in the long run if there is an effective alternative provided by ethical surgeons.

The resource problem also applies, at a different level of magnitude, to the developed world, where a state service may not be able to deal with its load of work, leading to long waiting lists. Successive governments in Britain, for example, have addressed this problem with ‘waiting list initiatives' which attempt, at great cost, to compensate for the lack of planning and of basic staffing.

Many countries also lack an effective primary health-care system that serves the ‘gatekeeper' function, dealing with minor problems, and only referring more serious cases to specialist surgeons. The result is that hospital-based surgeons have to deal with a large load of patients who could otherwise have been seen by less qualified doctors.

Education

Surgical disease needs to be represented more in health-care education programmes, for public education can help to dispel fears of surgery and harmful superstitions. Even doctors in primary health-care settings may not see surgical conditions frequently and they often fail to act promptly when faced with unfamiliar conditions. They, therefore, need continual reminding of the importance of early referral of patients with malignancy; they also need to be made aware, most effectively through continuing medical education programmes, of conditions such as urine infection in children with vesicoureteric reflux, which might have innocuous symptoms and signs, but which can have serious long-term consequences. They need to know of advances in diagnosis and management so they can, in turn, pass on the message to their patients, and be more effective at referral.

Late presentation: carcinoma of breast as an illustration

It would be difficult to find a common cause for late presentation of all conditions under all circumstances. We will use carcinoma of the breast as an illustrative example. It is either the commonest, or the second (after carcinoma of cervix), commonest malignancy of adult women in the world. Up to a third of patients in the developed world, and up to two-thirds in the developing world, present with an advanced stage of the disease. It is also a good example because it is visible, it has a relatively predictable prognosis, some surgical principles are well established, and there are variations which have lent themselves to research in terms of how and why patients present late.

USA: ethnic differences in presentation

In the United States, in comparison to white women, Afro-American women with carcinoma of the breast:

- have a 5-year mortality of up to 2.2 times that of white women;

(The poorer prognosis seems to persist, albeit at a less significant level, even when other factors are controlled for, suggesting a more aggressive tumour biology in Afro-Americans. From 1983 to 1987, the 5-year relative survival rate for localized disease was 93 per cent for whites and 83 per cent for Afro-Americans; for regional disease it was 73 per cent and 56 per cent; and for distant disease it was 19 per cent and 12 per cent, respectively.)

- present with advanced disease more frequently;
- present with early disease less frequently;

(In one recent study, at diagnosis, Afro-American women with breast cancer were two times more likely to have stage IV breast cancer and one and one-half times more likely to have stage III breast cancer than white women, and were only one-half as likely to have stage I breast cancer. Similarly, at presentation, they were almost twice as likely as white women to have tumours that were larger than 5 cm or tumours that had extensions to the chest wall or skin.)

- tend to use breast screening facilities less often.

Amongst the factors other than ethnicity that have been documented as contributing to late presentation by all women are: low socio-economic status, social deprivation, social marginalization, poor education levels, poorer awareness of carcinoma of the breast, and, perhaps most importantly, denial.

Cultural factors which have been documented include beliefs that ‘air' causes breast lumps, that the devil causes breast lumps, that unconventional interventions (e.g. chiropractic) would help cure breast lumps, and that women are less attractive to men after a mastectomy.

International experience: determinants of late presentation

The ethnic and cultural differences contributing to late presentation are not restricted to the United States. For example in the United Kingdom, Afro-Caribbean and African women have a significantly lower uptake of screening services than their compatriots: in Camberwell, London, it is 50 per cent or less, compared to a national average of about 70 per cent. At King's College School of Medicine and Dentistry in London a Multidisciplinary Breast Team research project has been set up to study the determinants of delay in seeking treatment for breast SYmptoms. The study is looking at demographic variables, ethnicity, and personality characteristics.

In Pakistan, late presentation has been linked to social taboos, ignorance, and fear of hospitalization. In India, advanced disease at presentation to specialist centres has been linked to lack of awareness, but also to errors of management by referring physicians.

Sociology of late presentation

It is both important and interesting to understand why patients do or do not consult doctors when ill, and what triggers the consultation. The on-physiological factors, which Zola called ‘pathways to the doctor', include availability of medical care, affordability, effectiveness of folk healers, and perception of the problem by the patient and those around him/her.

Common symptoms (e.g. tiredness, headaches) are often considered normal; what is ‘abnormal' is socially constructed and the definition of ill health depends on the underlying concept of health, including social, behavioural, and emotional elements. Zola, in a study in Boston, Massachusetts, found that bodily complaints may be communicated in a ‘restricting,' stoic way, for example by Irish-Americans, or in a ‘generalizing,' more dramatic way by Italian-Americans. Triggers to seek medical help were identified by Zola as an interpersonal crisis; perceived interference with work, physical functioning, or personal relationships; letting another person decide (‘sanctioning'); or after setting a time period for duration of symptoms before seeking medical help. Interestingly, the type of trigger was correlated more with sociocultural factors rather than with severity of illness, leading Zola to conclude that such sociocultural factors actually influence epidemiological statistics, by influencing who seeks medical care for what symptoms.

Even when available and affordable, utilization of medical care also depends on the perceived aetiology, that is whether a condition is believed to originate in the individual, or in the natural, social, or supernatural worlds. We probably underestimate these factors in the West, but surgeons in developing countries quickly learn to account for them in their consultations. Patients, depending on their cultural perceptions, may differentiate between the various roles expected of different healers. This obviously happens even in Western countries: for example Scott found that amongst five ethnic groups in Miami, patients sought relief for their symptoms from conventional doctors, but relied upon a folk-healer to deal with the root causes and to explain these in culturally appropriate terms.

Culture and the causation of illness

The perceived ‘causes' of illness are culturally constructed in the Middle East and Iran, for example the ‘evil eye' (known as ‘*mal ocho*' in Italy and ‘*mal de ojo*' in Spain) as a cause of illness is common amongst all communities, whether Islamic, Christian, Jewish, or Zoroastrian. Ill health can also be ascribed, especially in parts of Africa and the Caribbean, to magic, witchcraft, or sorcery, and in Western society to ‘stress', which is often ascribed to other people. When illness (e.g. a lump in the breast) is ascribed to God (or spirits or ancestral ‘shades') as punishment for some sin, then ‘prayer and repentance, not penicillin, ‘cure sin'. Some spirits, for example in Africa, or in the form of a ‘jinn' in the Middle East, strike much as viruses do, without sin on the patient's part, to cause an illness. Such spirits need to be driven out, or exorcised, from the body. In some societies, illness is perceived to be caused by ‘disequilibria', for example the humoral or hot–cold systems in Latin America, and the *yin–yang* system of China.

While many of the lay models are ‘externalizing', in the sense of causation by external agency, modern biomedical explanations are ‘internalizing', in the sense of finding internal causes within the patient, for example by an oncogene mutation or a genetic predisposition to breast, ovarian, or colorectal neoplasms. However, the scientific, biomedical explanations are achieved mostly at the expense of the important social and psychological components of suffering.

Delay in consultation for cancer and other serious conditions

Hackett *et al.* examined the delay between first sign or symptom for cancer and seeking medical help in 563 patients at the Massachusetts General Hospital, Boston. Only 33.7 per cent were early (less than 4 weeks) ‘responders'; the rest were all late in presenting with their symptoms; 8 per cent waited until they could no longer function independently before giving in to family or community pressure. Those who were more worried of cancer, surprisingly, sought help later, probably to avoid bad news. Candid labelling of the condition as ‘cancer' led to earlier responses. Those from higher socio-economic levels tended to seek treatment earlier, but there was little evidence at that time that cancer education programmes *per se* could be credited for this difference.

In another study, of patients who were well-informed generally, and who then developed acute myocardial infarction, Olin and Hackett found a significant amount of self-deception, whereby the patients tended to explain away their severe chest pain as resulting from less serious conditions such as ‘indigestion.' This denial of serious pathology is a common cause of late presentation of surgical conditions, especially those conditions with high ‘dread' factors, such as carcinoma of the breast.

The patient–doctor consultation: what could go wrong?

The patient–doctor consultation is characterized by ritual and SYmbols, and has been seen by deconstructionists as a power play. Its functions include the presentation (verbally and otherwise) of the illness by the patient; the conversion of illness (vague symptoms and signs) into disease (a pathological entity with a named diagnosis); and the prescription of a management regime understandable by the patient (not always) and by the doctor. The consultation process requires ‘negotiating' a consensus, with patients often minimizing the seriousness of the illness in this social process of converting an ill individual to the status of a ‘patient', with all its attendant rights and obligations. This complex process is beset by a number of problems that may result in an unsatisfactory outcome, and therefore a delayed presentation to another surgeon, later on, when the patient has far-advanced disease. These problems include misunderstandings about the patient's perception of the origin and causation of the pathology, misinterpretation of the ‘language of distress' (the presenting symptom may not be the patient's real problem); and incompatibility of explanatory models for the nature of the illness itself. Another important transactional problem is whether the patient can afford the treatment, assuming facilities for investigation and treatment are available.

Fortunately, medical curricula are beginning to include such sociological and ‘holistic' considerations in training, but unfortunately the way we currently train surgeons diminishes these aspects in favour of a more technological approach.

Preventing late presentation and avoiding advanced disease

Avoidance of known aetiological factors: cardiovascular disease as an illustration

Physical exercise, avoidance of smoking and of bad nutrition, and effective management of diabetes, hypertension, and dyslipidaemia would all contribute to a reduction of coronary and peripheral vascular disease. However, in those who go on to develop cardiovascular disease, late presentation can only be avoided by greater awareness, through public health education of early signs and symptoms, and by provision of primary health care, an effective referral system, adequate numbers of specialists, functioning investigative tools, and levels of outcomes of treatment good enough to make patients with early symptoms feel it worth-while to seek help. These are not easy to provide in countries with limited resources and it is to be expected that in such countries a very substantial proportion of surgeons will be dealing with advanced disease.

Screening: colorectal and breast carcinoma as illustrations

Screening for breast and cervical cancer is widely performed world-wide, while screening for lung carcinoma by chest radiographs appears to have gone out of fashion, although there is renewed recent interest. Screening programmes may be justified in those conditions where the disease is already advanced by the time symptoms develop: colorectal carcinoma is such an example. The other justification is where a common condition with devastating consequences can, by using safe and inexpensive techniques, be picked up at an early stage, and where intervention at an early stage is expected to make a difference in outcome. Carcinoma of the breast partly fits this category, although there is still debate as to which age groups would benefit, and, even now, surprisingly, there is controversy regarding the evidence, which came mainly from early Swedish studies, upon which mammography was adopted as a screening tool.

The main principles of a ‘cost effective' screening programme are that the condition must be common, easily detectable, picked up by tests that are inexpensive and not harmful, that the condition must be treatable, and early treatment should significantly improve the prognosis. Following mass screening there should be fairly simple confirmatory tests. For carcinoma of the colon in the general population without increased risk, the detection of blood in the stool by the haemoccult test (one of the variants of the guaiac test, which detects haematin) has been established because the test is neither too sensitive (therefore it avoids too many false positives) nor too insensitive (avoids missing true positives); it is particularly useful for left-sided colonic lesions, because blood shed from the caecum, for example, would be degraded beyond haematin, and that from the rectum may not have had time to be changed to haematin. Sigmoidoscopy and colonoscopy are available for those found to have a positive result; and for the primary screening of those at high risk.

Management of patients presenting with advanced disease

General principles

Wider responsibilities and surgical skills

The surgeon may be the first and only ‘qualified' doctor to whom the patient comes for help. He or she needs to understand the difficulties the family faces and to take responsibility to orchestrate all the phases of management, whether in the hospital or outside, especially where social services are poor or non-existent.

Under emergency conditions or for advanced acute conditions, diagnosis needs to be based on clinical findings, especially where there is pressure on diagnostic facilities. In rural hospitals, laboratory investigations may be confined to the haemoglobin level and urine sugar only. Fine needle or imprint cytology may have to suffice instead of conventional histology, and radiology may not be available. Telephone or Internet contact for advice is, in practice, rarely possible. Thus, the luxuries of waiting for sophisticated investigations, or even, because of advanced disease, of referring the patient to a more appropriate or experienced colleague, may not exist. Here the surgeon needs to maintain a balance between the ideal and the possible, because waiting for the ideal may lead to enormous community morbidity. Luckily, surgeons in rural hospitals in developing countries are on the whole very capable of ‘making do' and would be very adept at, for example, repairing a hernia with nylon rather than waiting for mesh, or using local anaesthesia (or intravenous ketamine) where appropriate to operate on patients with congenital hypertrophic pyloric stenosis, hernia, anal fistulas, etc.

Staging of procedures is often important and can be life saving. In children, for example those with Hirschsprung's disease, a colostomy provides the time to rebuild the patient physically, mentally, and educationally before major surgery. Urinary obstructions may require proximal drainage to allow for control of infection and restoration of renal function. The principle to remember here is that ‘quick fixes' are the enemy of the good.

Avoidance of known precipitating factors: chronic liver disease as an illustration

In patients with advanced liver disease a number of known factors can precipitate a sudden deterioration of liver function. These include electrolyte imbalance, diuretics, tapping of ascites, intercurrent infections, and major physical trauma, including surgery. However, if major corrective surgery is going to be needed or is unavoidable (e.g. liver transplantation), the surgery should not be delayed until the patient is moribund. Under these circumstances the results of even moderate surgery are significantly poorer, for example the mortality rate for cholecystectomy in patients with advanced cirrhosis can be as high as 80 per cent. Since infection makes both liver function and surgical outcome worse, it is important to undertake precautionary measures for surgery, for example antibiotic prophylaxis. Major sepsis sets in faster in patients with advances liver disease and can lead to multiple organ failure. It is important, therefore, to remember the interconnection between organ systems, for example the hepatorenal and hepatopulmonary syndromes, and to take precautionary measures to protect other organs. Nosocomial infections (defined as those developing after 48 h of admission) affect up to 40 per cent of critical care unit patients, and contribute to about 75 per cent of deaths there. One way to avoid nosocomial infections, which are major problems in most intensive care units, is to avoid admission to intensive care units in the first place by taking precautionary measures to protect vital systems, especially the respiratory system, through pre- and postoperative physiotherapy, appropriate use of antibiotics, placing incisions in less painful positions, using local nerve blocks, and providing adequate analgesia and deep vein thrombosis prophylaxis.

Nutrition

Many patients presenting late with infections or with advanced malignancy are severely anaemic and undernourished, with low serum albumin levels. Restoration by enteral nutrition is best, liquidizing meals if necessary and using tube gastric or jejunostomy feeding. Intravenous alimentation, besides being expensive, is quite impractical in most situations, especially in countries with limited resources. Up to 4000 calories and 125 g of protein per day with a mix of vitamins, iron, and other trace elements are needed—all this is best provided by natural foods with iron supplements for most patients. Thin nasogastric tubes are well tolerated. Relatives can help feed patients. Considering the delays that have often occurred before patients present, time spent restoring weight is very worthwhile if the condition is not life-threatening. Although there is controversy regarding the impact of poor nutrition and low albumin on healing, common sense would dictate that the preoperative patient be fed well, especially when major surgery would lead to further starvation in the immediate postoperative period. In cachectic patients it is important not to miss coexistent fluid and electrolyte shortage.

Palliation versus cure: colorectal malignancies as an illustration

The aims of palliation are very different from those of attempting cure. It is important that all members of the surgical team are aware of what is to be achieved when the patient is to be palliated, and so communication within the team is important. Since every therapeutic manoeuvre has a cost and places a burden on the patient, every investigation or treatment not aimed at the patient's own comfort must be carefully thought out and justified. Depending upon the expected survival of the patient, certain interventions may be inappropriate. In all considerations, the principle should be that the burden of palliative measures must be commensurate with the expected survival. This is not a recipe for inactivity, for surgery does have a major role in palliation.

Resecting a tumour is often better than leaving the tumour *in situ* and this should be done whenever it is safe to do so, for example a right hemicolectomy is indicated for an obstructing carcinoma of the caecum even if there are secondaries in the liver. For symptomatic relief, for example for rectal carcinoma, it is justified to debulk or even to resect (by, say, endoscopic transanal resection) as this may lead to more effective relief for distressing symptoms such as diarrhoea, tenesmus incontinence, and haemorrhage. Local control can also be achieved by laser or cryotherapy.

If the patient is moribund and has not had adequate resuscitation, the first priority is to save his/her life—so for example, a Hartmann's procedure is justified to relieve obstruction, and a more definitive procedure can be performed, if necessary, at a later date. Single hepatic metastases can be resected with relief of pain, possible prolongation of life and the rare cure when the metastasis is solitary, but not when multiple, even if confined to one lobe. Methods of dealing with liver metastases in the future may include ‘electrolytic' therapy using direct electrical current applied by electrodes into the tumour. Primary hepatic tumours have been treated this way by Chinese surgeons for some time now.

Perforation through a tumour almost always means wide peritoneal dissemination of malignant cells. This means that cure is very unlikely and the management should be aimed at effective palliation.

When using chemotherapy, it is important to remember that, generally, the more powerful the cytotoxic agent used, the more serious are the unwanted side effects. Thus, outside of controlled trials with informed and comprehended consent and the ability to withdraw from treatment, these agents should not be given for palliation unless there is very convincing value for a particular patient. 5-Fluorouracil is still the most commonly used agent, giving a response rate of about 15 per cent for colorectal secondaries. Radiation for palliation should be used cautiously, so as not to substitute cystitis for temporary relief of pelvic pain.

Multidisciplinary hospital palliative care teams are increasingly being used. Studies have shown that in patients with malignancies, they are effective in dealing with symptoms such as psychological distress, anorexia, pain, mouth discomfort, constipation, breathlessness, nausea and vomiting. They do not seem to be as effective in managing depression in these dying patients.

Where do patients die? The hospice movement

Although 50 to 70 per cent of cancer patients would prefer to die at home, there has been a trend towards the hospitalization of dying patients in many countries. The hospital is still the most common place of death from cancer, the percentage of cancer patients who die in hospital is diminishing, mainly because of increasing use of hospices and communal establishments, including residential and nursing homes. Given the ageing population, this trend is likely to continue.

The modern hospice movement has been championed by Dr Cicely Saunders more than by anyone else. It had its formal beginnings in the United Kingdom in 1967, with the opening of St Christopher's Hospice in London. Since then, the idea of the hospice, and consequently of good palliative care, has spread widely. The movement has shifted from charismatic, charity-based, and independent hospices to routine and increasingly bureaucratic palliative care. Global differences exist around the degree to which hospice and palliative care are established components of the health-care system. Access appears to increase with the integration of services into mainstream health funding, but at the cost of increased regulation, competition, and a potential loss of the personal attention so necessary for those who are dying. In most developing countries, dying patients are cared for by relatives; the hospice movement has not yet made an impact.

Communication and cultural sensitivity

The surgeon looking after a patient with advanced disease must also remember the family's anxiety and suffering and try to reduce these by arranging for, or providing, comfort and psychological support. As events unfold, repeated communication with the family and patient is required; effective communication includes avoidance of sending mixed messages and the use of familiar language and idiom. Care should be taken to avoid the impact of differences in social status between surgeon and patient (and their family) from interfering with communication. If relatives are to be involved in payment of the bill, or caring for the patient, they would need to be involved in decision-making (with the patient's consent).

It is also important that the surgeon should work with cultural sensitivity, for example when advocating a colostomy, which is almost always refused by practising Muslims, the surgeon needs to know the relationship of faecal soiling to the ritual cleanliness required when praying, and to give appropriate guidance. The surgeon should be truthful without being hurtful, remembering that the truth can be delivered in small doses. There is no need to lie. If the situation reaches a point of hopelessness, the surgeon should know when to stop, for death is part of life's experience and it should not be combated as if it were an enemy.

Follow-up (clinical review) for malignant disease: colorectal carcinoma as an example

This should be performed with clear objectives in mind. Prolonged routine review after surgery for malignant disease is justified if it will lead to early detection of recurrence or metastases, and if early detection will obviate discomfort and improve prognosis, or lead to better palliation. It would also be particularly justified in patients with a high risk of developing polyps and other malignancies, say in patients genetically predisposed to adenomatous polyposis coli or hereditary non-poplyposis colorectal cancer; in these circumstances even uncomfortable and expensive procedures may become worth-while. Serum carcinoembryonic antigen is a good indicator of recurrence in some, but not all, patients; when the level rises, second-look laparotomy may be indicated. Hepatic ultrasound and CT scans will detect small liver secondaries.

Follow-up is, of course, also justified for the psychological needs of the patient, but in that case the same objectives may be achieved by arranging another type of follow-up or psychological assistance.

Specific disease conditions—specific technical aspects

Neonates, infants, and children

Early diagnosis of many congenital conditions can be made by the use of a careful history and physical examination in the evaluation of well babies and children. Failure to thrive is a common feature of delayed and late presentation of most paediatric surgical conditions.

Gastrointestinal conditions

Intestinal atresia and other intestinal anomalies, congenital hypertrophic pyloric stenosis, and megacolon all present late, with vomiting, in cachectic babies. Prolonged delay of therapy in a starving neonate is to be avoided whenever possible. While fluid resuscitation starts, the use of a staged or definitive operation under local anaesthesia must be considered, as babies tolerate this well. In megacolon, a temporary colostomy allows the child to catch up with its developmental milestones before definitive major surgery and also allows for more accurate diagnosis of the extent of the disease. Anorectal anomalies often present late and are almost always correctable; where there is severe overdistention of the bowel, it is recommended that a staged procedure be carried out, with resection of the persistently dilated bowel segment about a year later.

Parents should be reassured that girls with rectovestibular and anovestibular anomalies, and boys with hypospadias, can enjoy sex and have children if these conditions are expertly managed.

Genitourinary conditions

Genitourinary problems such as urethral valves, reflux, and obstruction in the ureter or pelvi–ureteric junction, often present late because the first signs and symptoms of infection have been overlooked, or vomiting has been thought to be of gastrointestinal origin. To prevent this, urine microscopy and culture should be done whenever a baby is ill and the diagnosis has not been established. Vigorous antibacterial therapy and temporary drainage are necessary to improve general health and renal function before definitive operations; they also allow for the child to be referred to an appropriate hospital.

Orthopaedic conditions

The major spinal orthopaedic anomalies are difficult to treat and are best managed by an experienced team. Talipes and dislocated hips are best managed by early, repeated manipulation; however, they often present late, becoming major problems requiring specialized surgery.

Acute osteomyelitis is chronic after 48 h, and while the latter responds better to drainage and long-term antibacterial drugs in children than in adults, compliance can be a major problem. Education of the parents under these circumstances is likely to increase compliance.

Neglected wounds and sepsis are common in children in developing countries, leading to nutritional deprivation and other problems. Malunited fractures mould better than in adults and are usually best left alone. Ankylosis is more difficult and is usually untreatable until joint replacement is performed.

The late sequelae of delayed treatment, especially where there are poor rehabilitation facilities, include scarring and other deformities, and present a vast social and surgical problem. Where facilities exist, bone and joint deformities can be dealt with by the use of osteotomy, bone grafting, and joint replacement with good soft tissue correction. The use of teams of surgeons in organized outreach programmes, visiting places where there are collections of patients, is one solution available to help treat these patients.

Trauma and burns

Although epidemiological data is difficult to come by, common experience shows that neglected trauma and burns are a huge problem in developing countries, where large numbers of patients die in the very acute phase due to lack of adequate treatment. The number who die in the subacute phase after delay in treatment is probably larger. Most victims are either children or young adults; available statistics show that trauma is the leading cause of death between the ages of 10 and 40 years in most countries. The solution to this must be a combination of primary and secondary prevention.

Acute conditions

Late presentation of even minor surgical conditions could endanger the lives of patients.

The range of such conditions is vast and includes acute haemorrhage, abscesses (cutaneous and visceral), wounds, osteomyelitis and septic arthritis, pneumothorax and haemothorax, intracranial haemorrhage, appendicitis, torsion, strangulating and non-strangulating intestinal obstructions, vascular injuries, and arterial emboli. Estimates comparing the incidence of uncomplicated hernias with numbers of patients presenting with its acute complications indicate that there must be large numbers of untreated patients who die of strangulation and obstruction in the community, before reaching hospital. The same is probably true of many other conditions.

The general principles already enunciated also apply here: prevent; arrest bleeding; resuscitate; relieve pain; drain; stage procedures; use local/ regional anaesthesia instead of waiting for general anaesthesia. Adequate, urgent transfusion and replacement of fluid and electrolytes can make the difference between life and death. Remember that saving life takes priority over function and cosmesis. It is essential to try to do something rather than abandon a patient, even if all you can do is comfort.

Neoplasms

Benign neoplasms and other lumps presenting late are problems in rural hospitals only if they are very large, because general, rather than local, anaesthesia may be needed. Drainage of large, infected cysts may also need general anaesthesia; under these circumstances, ketamine is a useful option.

Malignant neoplasms of the surface often present late, even in the developed world. After assessment of spread, alternatives to surgery, such as local treatment by radiotherapy or chemotherapy, should be considered before embarking on heroic and disfiguring operations. This is particularly true for carcinoma of the breast and skin carcinomas. Where facilities for accurate assessment of spread and other modalities are not available, toilet surgery with early grafting can provide symptomatic and cosmetic relief.

Neoplasms obstructing hollow tubes (intestines, ureters) are best treated by staged procedures when the nutritional status of the patient is compromised; in the acute presentation adequate fluid resuscitation, even before fashioning a colostomy, can be life-saving. Nutrition is best administered enterally, if at all possible. One-stage treatment, while often successfully performed in centres with better facilities, may not be at all appropriate in settings with limited resources.

Non-acute surgical conditions

In developing countries, the list of such conditions presenting very late and in an advanced state constitutes a much larger proportion of the work of surgeons than in developed countries. A partial list of commonly encountered conditions would include genitourinary problems, including stones and obstructive uropathy of various aetiologies; inguinal hernia and hydrocoele; perianal conditions (haemorrhoids, fistulas, fissure); gastrointestinal obstruction; ulcers; and goitres, which may be huge and/ or causing complications by obstruction, malignancy, or hyperfunction. Arterial disease and aneurysms often present too late to save limb or life. Cleft lip and palate, burn scars, and similar problems amenable to plastic surgery are often neglected until they become impossible to correct: the social problems the patients face may then persist even after attempted corrective surgery.

Generally, the same principles apply as for acute problems, particularly the principles of always trying to help rather than abandoning. Outreach programmes may be very useful under these circumstances. It is important, always, to strike a balance between expensive technology and availability of basic therapies to large numbers of patients.

Further reading

Aluwihare APR. Adaptation for developing countries in the surgery of congenital anorectal anomalies and Hirschsprung's disease. *Sri Lanka Journal of Surgery* 1975; **2**: 55–60.

Aluwihare APR. Preventive aspects of paediatric surgery. *Ceylon Medical Journal* 1976; **21**:183–7.

Anim JT. Breast cancer in sub-Saharan African women. *African Journal of Medical Science* 1993; **22**: 5–10.

Aranda S. Global perspectives on palliative care. *Cancer Nursing* 1999; **22**: 33–9.

Barnett JR, Coyle P. Social inequality and general practitioner utilisation: assessing the effects of financial barriers on the use of care by low income groups. *New Zealand Medical Journal* 1998; **111**: 66–70.

Caplan LS, Helzlsouer KJ. Delay in breast cancer: a review of the literature. *Public Health Review* 1992–1993; **20**: 187–214. [A very valuable and extensive review of this important subject.]

Carmalt HL, Mann LJ, Kennedy CW, Fletcher JM, Gillett DJ. Carcinoma of the male breast: a review and recommendations for management. *Australia and New Zealand Journal of Surgery* 1998; **68**: 712–15.

Coates RJ, Bransfield DD, Wesley M, Hankey B, Eley JW, Greenberg RS, Flanders D, Hunter CP, Edwards BK, Forman M, *et al.* Differences between black and white women with breast cancer in time from symptom recognition to medical consultation. Black/White Cancer Survival Study Group. *Journal of the National Cancer Institute* 1992; **84**: 938–50. [An important publication from this study group.]

Desai PB. Facilities and challenges for cancer control in India. *Gan-To-Kagaku-Yoho* 1992; **19**(8 Suppl.): 1146–52.

Doglietto GB, Pacelli F, Caprino P, Alfieri S, Carriero C, Malerba M, Crucitti F. Palliative surgery for far-advanced gastric cancer: a retrospective study on 305 consecutive patients. *American Surgeon* 1999; **65**: 352–5.

Edmonds PM, Stuttaford JM, Penny J, Lynch AM, Chamberlain J. Do hospital palliative care teams improve symptom control? Use of a modified STAS as an evaluation tool. *Palliative Medicine* 1998; **12**: 345–51.

Emmink B, Hadley GP, Wiersma R. Infantile hypertrophic pyloric stenosis in a Third World environment. *South African Medical Journal* 1992; **82**: 168–70.

Finch DP. Diagnostic delay in symptomatic colorectal cancer. *Cancer* 1985; **56**: 2120–4.

Gnanaraj J, Gnanaraj L, Shah VK. Bring surgery to remote tribal areas. *Tropical Doctor* 1997; **27**: 163–5.

Goel AK, Seenu V, Shukla NK, Raina V. Breast cancer presentation at a regional cancer centre. *National Medical Journal of India* 1995; **8**: 6–9. [Documents causes of delay and instances of mismanagement in India..]

Guidry JJ, Greisinger A, Aday LA, Winn RJ, Vernon S, Throckmorton TA. Barriers to cancer treatment: a review of published research. *Oncology Nursing Forum* 1996; **9**: 1393–8.

Hackett TP, Gassem NH, Raker JW. Patient delay in cancer. *New England Journal of Medicine* 1973; **289**: 14–20.

Helman CG. *Culture, health and illness*, 2nd edn. Butterworth Heinemann, Oxford, 1990. [An excellent introduction to the subject of medical anthropology.]

Higginson IJ, Astin P, Dolan S. Where do cancer patients die? Ten-year trends in the place of death of cancer patients in England. *Palliative Medicine* 1998; **12**: 353–63. [A good study and review of this subject. Puts hospices is perspective.]

Kleinman A, Eisenberg L, Good B. Culture, illness and care: clinical lessons from anthropologic and cross-cultural research. *Annals of Internal Medicine* 1978; **88**: 251–8. [An important, accessible source. Kleinman, both a psychiatrist and anthropologist, has made a major contribution to explaining the roles of different cultures in understanding disease, and explaining the relevance of social science to biomedicine.]

Koong HN, Chan HS, Nambiar R, Soo KC, Ho J, Ng HS, Ng EH. Gastric cancers in Singapore: poor prognosis arising from late presentation. *Australia and New Zealand Journal of Surgery* 1996; **66**: 813–15.

Lannin DR, Mathews HF, Mitchell J, Swanson MS, Swanson FH, Edwards MS. Influence of socioeconomic and cultural factors on racial differences in late-stage presentation of breast cancer. *Journal of the American Medical Association* 1998; **279**: 1801–7. [An important study.]

Loefler IJP. Surgery in the Third World. In: Watters DAK (ed.) *Surgery in the tropics. Bailliere's Clinical and. Tropical Medicine Communicable Diseases*, Vol 3. Bailliere Tindall, London, 1988, pp. 173–89. [Authoritative, from a Western surgeon who has worked in Africa for a long time.]

Martin WMC, Sengupta SK, Murthy DP, Barua DL. The spectrum of cancer in Papua New Guinea. *Cancer* 1992; **70**: 2942–50.

Matthew PK, Kaptia F, Soaki PJ, Watters DAK. Trauma admissions in the Southern Highlands of Papua New Guinea. *Australia and New Zealand Journal of Surgery* 1996; **66**: 659–3.

Mbonde MP, Amir H, Mbembati NA, Holland R, Schwartz-Albiez R, Kitinya JN, Muguti GI. Experience with breast cancer in Zimbabwe. *Journal of the Royal College of Surgeons Edinburgh* 1993; **38**: 75–8.

Mbonde MP, Amir H, Mbembati NA, Holland R, Schwartz-Albiez R, Kitinya JN. Characterisation of benign lesions and carcinomas of the female breast in a sub-Saharan African population. *Pathological Research Practice* 1998; **194**: 623–9.

Nordberg EM. Incidence and estimated need of caesarean section, inguinal hernia repair and operation for strangulated inguinal hernia in rural Africa. *British Medical Journal* 1984; **259**: 92–3. [An excellent study documenting the enormous surgical load and shortage of surgeons.]

Nundy S. How might we improve surgical services for rural populations in developing countries? *British Medical Journal* 1984; **259**: 71–2.

Olin HS, Hackett TP. The denial of chest pain in 32 patients with acute myocardial infarction. *Journal of the American Medical Association* 1964; **190**: 977–981.

Parkin DM. *Cancer occurrence in developing countries*. Oxford University Press, New York, 1986.

Pergoraro RI, Nirmul D, Bryer JV, Jordaan JP, Joubert SM. Clinical patterns of presentation of breast cancer in women of different race groups in South Africa. *South African Medical Journal* 1985; **8**: 808–10.

Peter-Golden H. Breast cancer: varied perceptions of social support in the illness experience. *Social Science and Medicine* 1982; **16**: 483–91.

Phelan M, Dobbs J, David AS. 'I thought it would go away': patient denial in breast cancer. *Journal of the Royal Society of Medicine* 1992; **85**: 206–7.

Rennie JA, Janka A. Emergency surgery in Ethiopia. *Annals of the Royal College of Surgeons* 1997; **79**(6 Suppl.): 254–6.

Scott CS. Health and healing practices among five ethnic groups in Miami, Florida. *Public Health Report* 1964; **89**: 524–32.

Spooner B. The evil eye in the Middle East. In: Douglas M (ed.) *Witchcraft confessions and accusations*. Tavistock, London, 1970: pp. 311–19.

Underwood P, Underwood Z. New spells for old: expectations and realities of Western Medicine in a remote tribal society in Yemen. In: Stanley NF and Joshe RA (eds). *Changing disease patterns and human behavior*. Academic Press, London, 1981: pp. 271–97.

Usmani K, Moran EM, Haider W, Afzal H, Ahmad N. Breast cancer in pregnant and lactating women. *Journal of Environmental Pathology, Toxicology, and Oncology* 1995; **14**: 227–34.

Velanovich V, Yood MU, Bawle U, Nathanson SD, Strand VF, Talpos GB, Szymanski W, Lewis FR Jr. Racial differences in the presentation and surgical management of breast cancer. *Surgery* 1999; **125**: 375–9.

Watters DAK, Bayley AC. Training doctors and surgeons to meet the surgical needs of Africa. *British Medical Journal* 1987; **295**: 761–3.

Zola IK. Culture and symptoms: an analysis of patients' presenting complaints. *American Sociology Review* 1966; **31**: 615–30.

Zola IK. Pathways to the doctor from person to patient. *Social Science and Medicine* 1973; **7**: 677–89.

47.9 Surgery with limited resources

A. S. Daar, S. Adib Rizvi and S. Anwar Naqvi

Introduction
Poverty and health
A theoretical country with limited resources
Surgery in countries with limited resources
The local level
Surgical spectrum
Malignant diseases in developing countries
The four scourges
The social level
Women's health and surgery
Female genital mutilation
Education and training
Preparing for the various roles of the surgeon
Prevention
Organizational issues
Outreach programmes
Research and publication
Technology issues
Demographics, westernization and transitions: the future
Further reading

Introduction

The world has come to be defined by the binary categories of the developed and the developing, First World and Third World, industrialized and non-industrialized, North and South: on the whole this means a division between rich and poor; between those with a wealth of and those with limited resources.

The reality of this division is that approximately 80 to 90 per cent of the world's population lives in the developing world yet they consume only 10 to 20 per cent of the world's health-care resources; the rest is consumed by the 10 to 20 per cent of the population that lives in the rich, developed world. Only 10 per cent of the 60 billion US dollars spent on health research every year is dedicated to the problems of 90 per cent of the world's population.This is the so-called 10/90 disequilibrium.

In the rich, developed countries surgery takes place underefficient, organized, predictable, clean, uncrowded circumstances; the reality in the poor countries is quite different (see Table 1). However, it is in these poor countries, with their diminished resources, that the simple forms of surgery are a crucial, life- and limb-saving part of the health-care SYstem. But being part of a system means that surgery cannot be divorced from the social, political, and economic realities; to understand the issues facing surgeons in these countries it is necessary to look at both the local scene (summarized here mainly in the tables) and the wider picture.

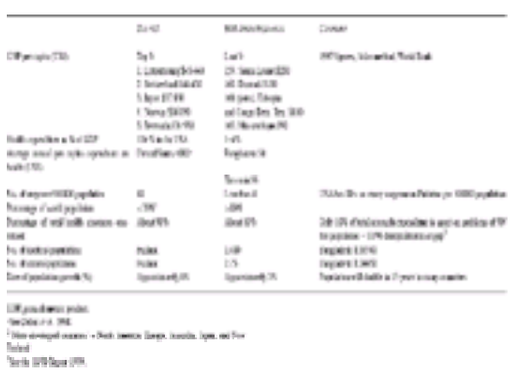


Table 1 Contrasting two worlds

Poverty and health

In the past five decades, on a global basis, life expectancy has risen dramatically to 68 years, gross domestic product in real terms has risen by 2.5 times, food production has more than doubled, and adult literacy has increase substantially. But these blessings have not been distributed equally. A quarter of mankind survives on less than a dollar a day.

Today, with the explosion of costs, particularly of high-technology medicine, per capita health-care expenditure in Europe and North America is about US$2000, while in Tanzania or Bangladesh it is approximately US$5. What this translates to is that of the 10 million children under the age of 5 years who died in 1997, 97 per cent lived in developing countries, where they died mainly from infectious diseases (lower respiratory-tract infections, diarrhoea, malaria etc.) made worse by malnutrition. Most of those premature deaths could be prevented if the resources were available—vaccination alone would have prevented about 2 million deaths. But the resources are not available, particularly in sub-Saharan Africa and parts of Asia and Latin America. In Latin America and the Caribbean, 37 per cent of the population live below the poverty line and 16 per cent in extreme poverty. In Bolivia, Guatemala, and Haiti more than 70 per cent of households live in poverty, the majority of these being in rural areas, particularly among indigenous populations.

But it is in Africa that poverty and health indicators are at their worst. Of the 10 countries with the lowest per capita incomes in the world, nine are in Africa (Nepal being the 10th); of the 10 countries with the highest death rates in children under the age of 5 years, again nine are in Africa (the other being Afghanistan); some of these very countries have the highest fertility rates, the lowest life expectancies, and the highest rates of infant and maternal mortality (see Table 2 and Table 3). Looked at in a different way, the global disease burden is very badly distributed: 93 per cent of the burden from 100 of the most common diseases and injuries is concentrated in low- and middle-income countries, with China, India, and sub-Saharan Africa accounting for fully 60 per cent.

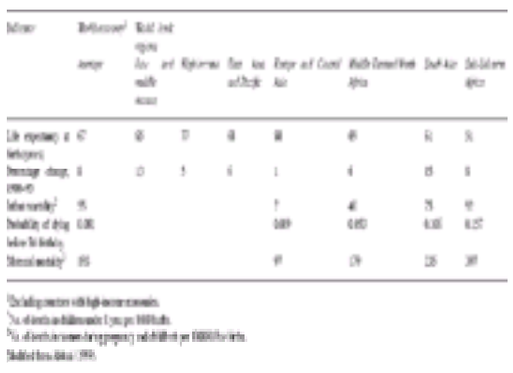


Table 2 World indicators of health (1995)

Ranking	Low GNP per capita	High infant rate in children under 5 years	High malnutrition in children under 5 years	High fertility rate
1 (worst)	Mozambique	Liberia	Bangladesh	Niger
2	Ethiopia	Afghanistan	India	Yemen
3	Tanzania	Sierra Leone	Nepal	Oman
4	Zaire	Guinea-Bissau	Ethiopia	Ethiopia
5	Burundi	Malawi	Mauritania	Somalia
6	Malawi	Guinea	Mozambique	Angola
7	Chad	Somalia	Vietnam	Afghanistan
8	Rwanda	The Gambia	Niger	Mali
9	Sierra Leone	Angola	Eritrea	Burkina Faso
10	Nepal	Rwanda	Afghanistan	Uganda

GNP, gross national product.
Adapted from 'Service strategy' (health, nutrition, and population) (Alleyne 1999).

Table 3 Countries with poorest health, according to World Bank indicators (1996)

The rich countries of the world have made a moral decision to contribute a certain proportion of their annual wealth to aid, but unfortunately only a few countries have consistently lived up to their commitment and some now contribute a far smaller proportion. Nevertheless, they all want to help the poorer countries, yet whatever aid they give is often wasted by corruption, mismanagement, lack of organization, and misguided attraction to inappropriate technology. Globalization of economies, arguably of benefit to all in the long run, has so far meant significant dislocations and increased inequity; and the absolute number of the poor has actually increased over the past two decades.

A theoretical country with limited resources

It would probably be in sub-Saharan Africa. It would spend less than 10 United States dollars per capita per year on health; its per capita income would be less than $300. Population growth would be about 3 per cent or more, and about 50 per cent of the burgeoning population would be under the age of 15 years. The doctor:population ratio would be 1:20 000 or worse (compared to 1:400 in many rich countries); infrastructure would be poor (and deteriorating); the teaching hospital in the capital city would consume about 40 per cent of the health budget. In this composite country a referral system does not exist, transportation is poor, and the Ministry of Health wants to control everything but is unresponsive to the needs of all but the influential few. Per capita food production is falling; 25 per cent or more of sexually active adults are human immunodeficiency virus (**HIV**)-positive and 50 per cent of the population test positive for tuberculosis. All these factors have an impact directly or indirectly on the everyday practice of surgery.

In this poor country there is a transition taking place: more and more people are moving away from the countryside to urban slums, where they now exist without their traditional support structures and where they contribute to increasing crime while taxing already overstretched utilities and health resources. It is expected that nearly half of the population of developing countries will be living in towns by the time this book is published.

Surgery in countries with limited resources

Two levels of discourse are necessary: the local level, with which surgeons are more familiar; and the social level, where we flounder because of the apparent intractability of the political, economic, and structural dimensions that influence our work. Surgeons in these circumstances are usually a dynamic, problem-solving community who could be at the vanguard of change. However, their ability to effect change is smaller than that of others because they often lack interest and involvement in the really important public-health aspects of their work, or they are often too overworked.

The local level

This is the district hospital or the village centre. Table 4 illustrates some of the kinds of problems encountered in an African setting. Here, at best, the difficulties are seen as challenges rather than obstacles, and local surgeons adapt and respond every day in remarkable ways (Table 5) to treat wounds, suture cuts, set fractures, deliver babies, restore sight, prevent disability, and save lives.



Table 4 Trials and tribulations of surgery in countries with limited resources



Table 5 (Table 5) Very practical coping strategies for the surgeon with very limited resources

The surgeon here is also the paediatrician and obstetrician; perhaps a quarter of the work is orthopaedics and obstetrics/gynaecology. Here surgery is not an expensive, high-technology intervention with little relevance to the real needs of the population; it is a crucial component of primary health care, where it has proved to be vital in saving life and limb, every day providing relief from disability or potentially lethal conditions. The World Health Organization (**WHO**) has recognized that the delivery of basic essential surgical services to rural communities is a legitimate primary-health need as well as an important approach to preventative health care. In their cost-effectiveness in preventing premature death and disability, the simple forms of surgery are second only to the most useful vaccines and antimicrobials. Data from Bangladesh, India, and South America indicate that about 10 per cent of all deaths and almost 20 per cent of deaths in young adults can be avoided by simple surgical interventions. However, the shortage of doctors, nurses, facilities, and resources means that there is a huge unmet need for surgical care.

While many of the features that define surgery in these circumstances are recognizable in the richer world, their very magnitude, frequency, and apparent intractability are overwhelming in developing countries. The long distances, poor communications, non-existent referral systems, overcrowded clinics and wards, and severe shortages often discourage patients from seeking early intervention for their maladies. As a result, surgery is often not sought, and when it is, it is often very late. In

many poor, rural villages the burden of chronic disadvantage, poverty, ignorance, and poor sanitation are the *leitmotiv* of everyday life; a surgical disease on top of all this may be the last straw, and as Maurice King has noted ‘when even the struggle to survive may be a losing battle, the fact that a surgical disease is treatable is mostly irrelevant.' Patients come to surgery already infected, malnourished, and beset by concomitant parasitic diseases—all factors that contribute to poorer outcomes, which then discourage future patients from seeking early treatment.

Surgical spectrum

Most of the conditions seen will be similar to those in the developed world: the difference is in their incidence and prevalence, where they are treated, by whom, and with what resources (Table 6). For example, there are 750 million people in the developing world with goitres, and the remarkable thing is that this is still mainly due to iodine deficiency.

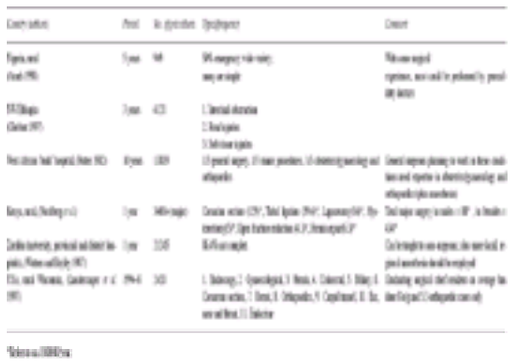


Table 6 Spectrum of surgery in selected developing countries; compared to rural USA

There are, of course, some conditions that are not often seen in the developed world, for example schistosomiasis, Chagas' disease, liver flukes, polio. There is less appendicitis and diverticulitis; fewer carcinomas of the colon, haemorrhoids, and varicose veins; and more urethral strictures, salpingitis, osteomyelitis, amoebiasis, various manifestations of tuberculosis (respiratory, lymph node, abdominal, and in bones), and polio contractures. Major parasitic infections of surgical interest include schistosomiasis (the most common cause of portal hypertension in the world), Chagas' disease, hydatid disease, various intestinal, biliary and hepatic flukes, and worm infestations. Roundworms, for example, are one of the most common causes of intestinal obstruction in parts of Africa.

Avoidable loss of vision is a major feature of surgery in countries with limited resources: children with corneal injuries are not treated; the huge backlog of cataracts cannot be operated on; and onchocerciasis still affects large numbers despite ivermectin being distributed (18 million infected, perhaps 750 000 blinded; the numbers will probably decrease dramatically as ivermectin distribution is extended to more countries in Africa); and while trachoma can be avoided or treated, those with trichiasis are often not operated upon until their corneas are irreversibly scarred.

In many parts of the developing world the surgical burden is increased by civil, political, and military violence. Amputees from landmine injuries are a common site in Angola, Afghanistan, and other areas of recent conflict. In places such as Papua New Guinea the most common cause of surgical death is trauma. Facilities for the early (or late) care of major trauma are related to per capita income (see Table 7). Poorer countries can hardly support ambulance services, let alone sophisticated trauma care, with the result that mortality rates from trauma are very high. Childhood injuries are common in developing countries (Table 8), as they are throughout the world, but their effect is devastating in countries without rehabilitation services. There is a high prevalence of disability from trauma in those who survive: a study from Punjab, India, has shown that for every person who died of an accident, there were eight who were permanently disabled. Apart from affecting individual health directly, disability of such a magnitude affects health indirectly through reduced income where there are no insurance schemes, and it further taxes the meagre resources of family and state.

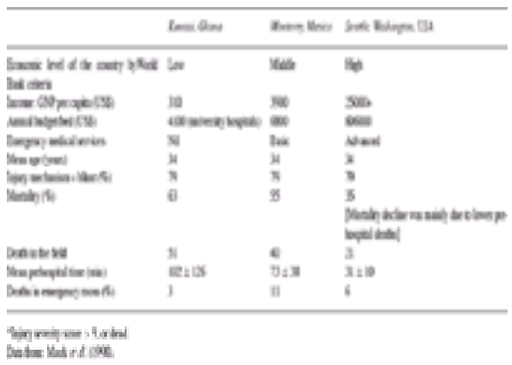


Table 7 Impact of economic resources/development on trauma mortality in severely injured patients*

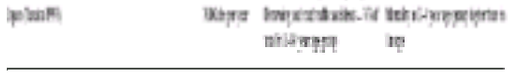


Table 8 Childhood accidental injuries in selected developing countries; compared to Japan

Malignant diseases in developing countries

Reliable figures for incidence and prevalence are very difficult to come by because mortality statistics are unreliable, cancer registries may not exist, and hospital records do not reflect disease burden in many developing countries. The WHO, through its collaborating centre, the International Agency for Research on Cancer (**IARC**) in Lyons, France, does try to gather epidemiological data, but even the databank maintained there is out of date and patchy for many developing countries.

According to IARC 1990 figures (published in 1998/9) the most common cancer in the world is lung cancer, accounting for 18 per cent of cancers of men worldwide, and 21 per cent of cancers in men in the developed countries. Stomach cancer is second in frequency (almost 10 per cent of all new cancers) and breast cancer, by far the most common cancer among women (21 per cent of the total), is third. There are large differences in the relative frequency of different cancers by world area. Other major cancers of developed countries (other than the three already named) are those of the colon/rectum and prostate, and in developing countries, of the uterine cervix and oesophagus. Tobacco smoking and chewing are almost certainly the major preventable causes of cancer worldwide today.

Table 9 is a compilation of a few series reported in separate publications from developing countries.

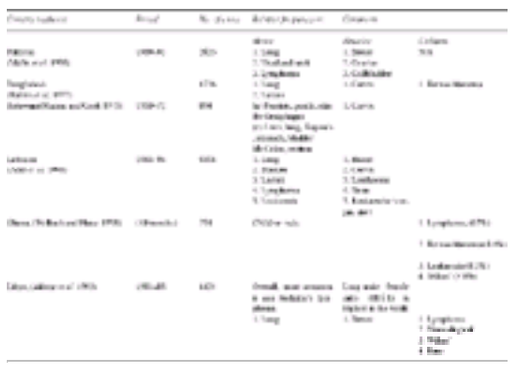


Table 9 Patterns of malignancy in selected developing countries

People of black African origin in the Caribbean and the United States have amongst the highest incidence of prostatic carcinoma, while nasopharyngeal carcinoma is common in South-East Asia. Poor countries may not have all the facilities to manage all malignancies, but some are preventable, for example lung (smoking), head and neck (alcohol, tobacco and betel-nut chewing, smoking), and hepatocellular carcinoma (aflatoxin in mould contaminating unrefrigerated stored grains; hepatitis B vaccination); others are often amenable to definitive treatment if detected early by practical, low-tech, and inexpensive methods, for example oropharynx (leukoplakia), uterine cervix (Pap smear), and breast (self-examination). Suffering can be reduced by having a plan for pain management, and by making oral opiates available.

The four scourges

Surgeons in countries with limited resources need to take greater account of concomitant diseases, dangers to themselves, and diseases contributing to their workload that would otherwise be looked after by different specialists in richer countries.

HIV

Surgeons are increasingly looking after patients infected with HIV. In countries with up to a quarter of the adult population infected, and where screening is unavailable or unaffordable, surgeons must assume that every patient is potentially infected—and yet under conditions of diminished resources, protective measures may also not be available or affordable. Ethical issues arise frequently under these circumstances, but surgeons cannot avoid caring for patients who develop complications of HIV, or for patients with ordinary surgical conditions who happen to be HIV-positive. The magnitude of the problem is increasing, especially in sub-Saharan Africa.

Worldwide, 16 000 new patients are infected with HIV daily; every minute five youngsters between the ages of 15 and 24 years are infected. It has already killed 14 million people. Sub-Saharan Africa may be the worst affected part of the world: about two-thirds of the 35 million people worldwide who were infected by March 1999 lived in sub-Saharan Africa, where more than 2.5 million people died from acquired immune deficiency disease (**AIDS**) in 1998; this is four times as many deaths as for the rest of the world put together. A quarter to a third of the adult populations of Botswana, Namibia, Swaziland, and Zimbabwe are already infected. Armies in many African countries have infection rates of 50 to 70 per cent. In South Africa itself, where the rates approach 15 per cent, more than 1 million will die of AIDS by 2001, bringing life expectancy down to 48 years within the first decade of the new millennium.

In Asia a number of countries already have serious HIV problems. The Asian country that currently seems to have the highest infection rate is Cambodia, where it has been estimated that up to 100 new patients are infected every day out of a population of about 10 million. India, with perhaps the largest road network in the world, has over 5 million long-distance lorry drivers, who, as in Thailand, are easy prey for commercial sex workers along their routes. A recent survey found that 87 per cent of the drivers were sexually promiscuous, having had up to 60 different partners during the previous year; and nearly 90 per cent routinely refuse to use condoms. The drivers pick up the women, and after having sex with them, leave them at another wayside inn of ill repute, where they provide their services to other drivers and local youths—the drivers are thus crucial in spreading sexually transmitted disease and HIV.

Poverty helps to spread HIV, and it, in turn, spreads poverty by decimating economically active adults. When patients die, they leave behind unwanted orphans—8 million in Africa alone, of whom more than 1 million are HIV-positive; such orphans are often left untended, malnourished, and uneducated, thus contributing to poorer health and more poverty later on.

While a vaccine is unlikely in the next 5 to 10 years, some African countries such as Uganda and Senegal have succeeded in sharply reducing the numbers of new infections through public education and other measures to reduce transmission. Success also depends upon changing attitudes amongst the general population, as is happening in Thailand and Brazil

Most of the countries with such heavy infection rates cannot even contemplate diverting very scarce resources to antiretroviral treatment: for Tanzania to fund effective triple-combination therapy and include all indirect costs would amount to $80 per patient or 73 per cent of the entire gross domestic product of the country.

Tuberculosis

The WHO declared tuberculosis a global emergency in 1993. Nearly one-third of the world's population, mainly in developing countries, is infected, and about 10 per cent of those infected develop full-blown tuberculosis. Globally, tuberculosis is the biggest killer of young people and adults. In India and Bangladesh, half the population is infected; in India alone half a million people die every year, and this rate of 10 deaths per million patients is an improvement on the 106 per million 15 years ago when India adopted DOTS (direct observation treatment strategy), which now covers about 14 million people.

India, Bangladesh, Indonesia, Myanmar, and Thailand account for 90 per cent of tuberculosis cases reported in South-East Asia, where 40 per cent of the world's cases and deaths occur.

A frightening development is taking place in Russia, where 10 per cent of the 1 million prisoners have active tuberculosis, half with multidrug-resistant strains. With 25 to 35 per cent of all prisoners being released each year due to the deteriorating economy, 1 million cases of latent tuberculosis are sent into the community every 3 years; each of these, without treatment, will infect 10 to 15 people each year, setting the scene for a major public-health problem that could threaten other countries too. If the current rates continue, the death toll for tuberculosis in Russia will reach about 1.5 million by the year 2000.

Malaria

Surgeons' interaction with malaria is through blood transfusion, the common use of syringes, enlarged spleens, postoperative pyrexia, and as a concomitant condition in surgical patients. It is perhaps the most dangerous disease infecting large numbers of *Homo sapiens*, with up to 500 million infected annually and up to 3 million deaths, the majority being in sub-Saharan African children. The recent development of vaccines, particularly a naked DNA vaccine, raises some hope of controlling malaria, which has continued to take its toll because of parasite resistance to drugs and mosquito resistance to insecticides. Now that chromosome 2 of *Plasmodium falciparum* has been sequenced completely there is also hope of novel, perhaps genetically based, therapeutic approaches. Furthermore, with the discovery and characterization of gametocyte inducers (one of which in the mosquito is xanthurenic acid), it is possible to contemplate rational drug designs either to interrupt the transmission of malaria or to select new mosquito genotypes resistant to infection.

Tobacco

Surgeons throughout the world are called upon to deal with the end-result of tobacco consumption and abuse by having to operate on patients with concomitant respiratory and cardiovascular diseases, and by having to deal directly with sequelae such as bronchial carcinoma, cardiovascular and peripheral vascular diseases. As the developed world is suing tobacco companies, the companies are focusing their marketing strategies on the developing world, where increasing numbers are getting addicted and where it is more difficult to stop smoking, especially by women. WHO, recognizing the huge impact that smoking has on public health, has

launched a major ‘tobacco-free initiative'.

Linkage between AIDS, malaria, and tuberculosis

These three scourges appear to be linked. *In vitro* studies have shown that malaria, by stimulating leucocytes that are the targets for HIV, can increase the replication rate of the virus by 30- to 100-fold. Furthermore, epidemiological studies in Africa suggest that malaria may be a cofactor for HIV, hastening the death of those who carry both infections. Tuberculosis, in turn, is resurgent, and HIV-positive patients are highly exposed and are infected in large numbers, often with multidrug-resistant tuberculosis.

The social level

Surgery does not exist in a vacuum. Foreign aid directed at specific surgical conditions will be wasted if the country tolerates corruption and has a poorly organized and managed health-care system. The main problems with many health-care systems in the developing world have been identified as:

- misallocation of funds to less cost-effective interventions
- inefficient use of funds, as in buying brand-name drugs and expensive technologies
- inappropriate deployment of staff and suboptimal or inappropriate use of beds
- explosion of costs outpacing incomesand perhaps most important of all
- inequity.

The World Bank, and both aid donor and recipient countries, have become increasingly aware of the role of corruption and poor governance in undermining economic stability and contributing to underdevelopment. A number of poor countries in Africa and elsewhere have asked the World Bank for help to establish anticorruption programmes, and donor institutions and countries now insist on professional management and a reduction of political influence in staff appointments. Part of the aid is channelled to improve governance and reduce corruption through deregulation, tax simplification, civil service improvements, public procurement, auditing, legal and judicial reform, and strengthening of mechanisms for public oversight.

The private sector need not be the enemy of health care; in fact, as in many countries, it could play an important part in reducing the burden on state institutions. However, in order for the private sector to flourish there is a need for sound legal and regulatory frameworks: weakness in dealing effectively with regulatory problems in the private sector can cause governments to become excessively involved in providing health services that might otherwise be provided by the private sector for those who can afford it. It is being realized now that governments in poor countries should target their limited resources at sustainable preventative health services, providing basic services to the poor, ensuring against catastrophic illness, overseeing medical education, research and development, and on quality control.

It is indeed an unjust world, but developing countries could make things better by reducing military spending and paying greater attention to the real linkage between elementary education, economic performance, and health. Developed, rich nations can help by honouring their commitments to aid and by recognizing that much of the aid does come back to the donor country indirectly: CIDA, the Canadian International Development Agency, has actually posted on its Internet website the calculation that for every official development assistance dollar, 70 cents returns to Canada through jobs and the purchase of Canadian goods and services, in addition to other tangible benefits to the donor countries. Developing countries, in this context, should be seen as essential partners.

Equity: a central value in public health-care policy

Countries of the world can be grouped according to their national income. However, in every group there are significant variations in the quality of health care. The differences are due to social factors: political unaccountability, corruption, poor organization and management, diversion of resources to military use; and inequity.

Equity, by which is meant, amongst others, equal access to health care; equal priority of care according to medical need rather than social status; political influence or ability to pay; equal quality of care in terms both of interventional effectiveness and patient satisfaction, and equal distribution of the financial burden is emerging as one of the cardinal values in public-health planning. It is difficult to achieve: in almost every country in the world where statistics have been analysed there are inequalities showing a gradient related to personal income, social status, professional status, educational attainment, and even sex.

We know little of how the social and personal factors actually translate into inequalities and different health outcomes: for example, a recent study in the United States of America found that under identical conditions of presentation, women and African Americans are significantly less likely to be referred for coronary angiography (and subsequent interventions) than are white males; and the reasons for this discrepancy appear to be more complex than mere overt sexual or racial bias. Answers to these complex questions will take a long time to be found.

When resources are very limited, inequalities become greatly magnified. Examples would include sending abroad for expensive treatment powerful politicians from countries where the per capita expenditure on health is of the order of $10 per annum. Therefore, it is even more important for poor, developing countries to put equity, and health-care ethics generally, at the top of their agendas and to try to reduce as much as possible the social inequalities in health, to increase the resourcing of primary and district levels of health care, and to reduce the political leverage of rich urban dwellers.

Individual surgeons may feel that they cannot make much of a difference, but we could continually ask ourselves the question ‘am I being as fair as possible in caring for those entrusted to me?'

Women's health and surgery

Sex and procreation seem to bedevil women's health in developing countries and the surgeon has to deal directly with the consequences of the lack of equity, the disregard for human rights, the absence of protective laws, and the inferior roles decreed for women in many traditional societies.

An interesting perspective on what some of these factors can do to women is to count what Amartya Sen, the Indian economics Nobel prize-winner, has identified as ‘100 million missing women': there are at least 100 million women ‘missing' in South Asia, West Asia, and China as calculated from the observed ratio of women to men in much of the rest of the world, typically 105:100. In the worst affected countries, such as India, Pakistan and China, the ratio can be as low as 94:100—the rest are ‘missing.' (Elsewhere in Africa, Asia, and Latin America the ratio varies widely.)

Women are not treated equally or fairly in many parts of the world; this inequality may take the form of poorer access to education, nutrition, and medical care. Where nutrition is poor in girls, one of the consequences later may be obstetric cephalopelvic disproportion, and where access to good care is denied by poverty or culture, the attendant obstructed labour can mean fetal and/or maternal morbidity or death; or it can lead to the complications of vesicovaginal fistulas. One of the most common causes of avoidable death in women of child-bearing age, particularly in Africa, is related to complications of child birth, or often, attempted abortion performed illegally by inexperienced backstreet operators under unhygienic circumstances.

Throughout the world now, carcinoma of the breast seems to be the most common malignancy in adult women (Table 9). The role of the surgeon assumes even greater importance in poorer countries because neither chemotherapy nor radiotherapy are available to the majority of women, particularly in rural areas. Cultural fear of surgery and devaluation of postmastectomy womanhood by the husband or the social group makes the situation worse. Pain management is notoriously poor and oral opiates are rarely available or prescribed, leaving women to suffer unnecessarily. It is not unusual to see women in developing countries present late with fungating lesions.

When even daily survival is a major struggle, where economic circumstances dictate having many children so some can survive as bread winners to help parents in old age, where the husband is dependent on the woman to grow food, do household chores, and look after him and the children, where poor nutrition and overwork sap her body, leaving her vulnerable to infections, and when the husband is perfectly happy and willing to ‘take' another wife when the first is infirm, a lump in the breast or a little soiling assumes less importance. It may well be the last straw, and the fact it may be surgically treatable may be of no real consequence.

Female genital mutilation

A particularly odious condition afflicting women in developing countries, particularly in Africa, is the tradition of female circumcision, now technically called female genital mutilation. It is divided into the following.

- ‘Sunna' circumcision, consisting of the removal of the prepuce and/or the tip of the clitoris (*sunna* means tradition in Arabic)
- Clitoridectomy (also referred to as excision), consists of the removal of the entire clitoris (both prepuce and glans), andthe removal of the adjacent labia

- Infibulation(pharaonic circumcision) where the clitoris, mons veneris, labia majora and minora are extensively excised and the raw areas allowed to heal over the lower vagina. Immediate complications include haemorrhage, infection and urinary retention; later there may be giant implantation dermoid cysts, haematocolpos, difficulty of vaginal examination, obstructed labour, and dysuria. When performed by an inexperienced operator struggling with a pained and terrified young girl, it is even possible to resect the urethra and to have enough tissue damage to get vesicovaginal fistulas. When these complications do occur, there is little expertise to cope with them. An infibulated woman must be cut open to allow intercourse on the wedding night.

At least 2 million girls are at risk of female genital mutilation every year. It is common in both Muslim and Christian societies; it seems related more to custom than canonical law. It persists because of psychosexual and sociological assumptions, through myths, to beliefs concerning hygiene, aesthetics, and religion. In many societies the roots of the tradition are far too remote for it to be amenable to simple exhortations to stop the practice. Female genital mutilation has now been identified as a major women's health problem and strong recommendations have been made to abolish it. A joint statement by the WHO, UNICEF, and UNFPA in 1997 confirmed the universally unacceptable harm caused by female genital mutilation and issued an unqualified call for its elimination in all its forms, which are all irreversible and last a lifetime.

Islamic experts, while confirming male circumcision as a legitimate practice, have proved with sufficient documented evidence that sayings or actions concerning female circumcision ascribed to the Prophet Muhammad have no authenticity. Noting the many risks involved in female circumcision, it was concluded that the practice ‘cannot be legitimate under Islamic law', and further that ‘female circumcision is neither required nor is it an obligation nor a sunna'. Infibulation was considered ‘an odious crime'.

Egypt has already declared the practice illegal, and now Senegal and several other African countries have also banned it, but it would probably require anthropologists, social scientists, educators, and activists working within a culture to make a real difference.

Education and training

Preparing for the various roles of the surgeon

A surgeon working in countries with very limited resources also needs to be a fairly competent obstetricians/gynaecologist, orthopaedist, paediatrician, internist, teacher, and manager. In other words, too much specialization may be a disadvantage. In addition, the surgeon will need to have working knowledge of basic radiology, ultrasonography, equipment maintenance and instrument sterilization, and of anaesthesia, for there are occasions when the surgeon will need to administer the anaesthetic and operate on the patient at the same time.

The surgeon serves his or her patients as well outside the operating rooms as inside it. The long hours, hard work, and enormous responsibilities may take a toll on time, health, and family relationships but the rewards of practising surgery in such circumstances are enormous.

To achieve these aims and yet produce a surgeon capable of further development, it is necessary to apply the important lessons we have learnt about medical education in the past two decades. There needs to be reform so that medical education and the curriculum are based on the community's needs and realities; the student needs to be encouraged to be a life-long learner who is capable of solving problems as they arise, and who is sensitive to the culture and its ethical values.

Because of the great shortage of medically qualified personnel there may be a need to create paramedical personnel to do specific tasks and the medical profession will need to recognize their worth and reward it. Nurses have made major contributions, performing caesarian sections and laparotomies in places such as Zaire (now the Democratic Republic of Congo) when there were no doctors. And today, in over 100 countries, nurse anaesthetists capable of performing all the critical tasks of anaesthesia are making a very important contribution.

Medical training should include rotation through the district hospitals; this is particularly important during residency training. Doctors posted to peripheral hospitals are often forgotten by those in the capital city where the Ministry of Health and the medical school(s) are situated, with the result that they move back to big cities as soon as they can, or leave the country. This is one of the saddest aspects of surgery in most developing countries, which often have an excess of doctors in cities but a big shortage in rural areas. It is likely that if their needs are taken care of, they would stay in rural areas: the ministries and medical schools need to make sure they are visited often, rotate to the capital, participate in Continuing Medical Education programmes, the schooling needs of their children are taken care of, the spouse is helped to find employment, and there are enough material incentives and allowances for working in rural areas.

Going abroad should be restricted to advanced fellowships for subspecialty training rather than for basic residency training—and it should be possible to provide reasons other than contractual obligations (often ignored) for those graduates who do go abroad to come back.

In many of these countries there will exist ‘traditional healers' of one sort or another. Some of them do indeed cause harm, but their very existence through the centuries can be seen as both an indicator of the inadequacies of ‘modern' health care and a testament to their having some value to the community. To get rid of them by legal or administrative fiat without comprehension of their roles and contributions, and without providing an acceptable, affordable alternative, is to court failure. They have survived because of their adaptability, flexibility, and understanding of the dynamics of illness within the community. Perhaps we ought to learn from them before getting rid of them, or we ought to learn to work with them.

Once a district hospital or a village health centre is set up, there continues to be a need for quality assurance: maternal mortality reviews, morbidity and mortality conferences, and reviews of medical records and prescribing habits should be encouraged. National essential drug and equipment lists are very useful; and some have found procedures lists to be useful too.

There is today a well-substantiated call for evidence in making health policy decisions. Even in the developed world this is as yet rather difficult to implement. For those developing countries trying to fit into a global economy and to establish modern, effective health-delivery structures, which increasingly means a central ® district ® village organization, with emphasis on decentralization and community involvement in decision-making, this could be a rather chaotic and confusing enterprise. If possible, what is needed is to establish a well thought out training programme in which health-care managers explore realistically the challenges of translating data from scientific studies into policy and then are given responsibility to implement those policies.

The training of medical students and residents is often beset by suspicions, misunderstandings, and personal rivalries between, on the one hand, the ministry of health responsible for providing health care and which is the ‘owner' of teaching hospitals, and, on the other, the university-based medical schools, which see their role as being mainly educational. But of course the two roles, particularly in surgery, are complementary and cannot be divorced. It is a rare country, such as Oman, that has good enough relationships and structural foundations to ensure effective training without friction. One rather interesting and radical solution is to have, as in Iran, both medical training and health-care delivery under one superministry.

The developed world can make a big difference not by training individual trainees, but by training the trainers.

Outreach programmes can also play an important part in training (see below).

Prevention

Public-health measures (sanitation, waste disposal, health education, good housing, good food handling and nutrition, vaccination programmes, etc.) are more important and cost-efficient than providing health care when people are ill. This is true even of surgical conditions, examples of which include schistosomiasis (the most common cause of portal hypertension in the world), tuberculosis, typhoid, amoebiasis, etc. Wearing crash helmets and seat belts, etter roads, respect for reasonable regulations, road and vehicle maintenance and testing, and other accident-prevention measures obviously make much more sense than treating major trauma and its economic and health consequences. Few developing countries can afford a fully fledged emergency medical service; public education of drivers to care for roadside patients and to reward them for transporting patients to hospital where appropriate can ameliorate some of the consequences of not having an ambulance service at all; advanced trauma life-support (ATLS) training programmes can make a big difference in the emergency room during the critical early period after major trauma.

An interesting paradox is that for the developed world it may be a better investment to fund free vaccine programmes in specific developing countries than at home. Mathematical modelling for hepatitis B has shown that the resources needed to prevent one carrier through universal vaccination in the United Kingdom could prevent 4000 carriers in Bangladesh, of whom statistically four might be expected to emigrate to the United Kingdom. Thus it would be four times more cost-effective for the United Kingdom to sponsor a vaccination programme against hepatitis B in Bangladesh than to introduce its own universal programme.

Perhaps the most important, cost-effective, and efficient way to allocate resources is to put more emphasis on providing elementary education. In a recent editorial in the *British Medical Journal*, Bansal noted that investment in elementary education would not only advance human rights in India but would have a large impact on

health as well. The world map of illiteracy coincides with maps of malnutrition, ill health, and high infant and child mortality. To some extent, education compensates for the effects of poverty on health, irrespective of the availability of health facilities, particularly for women, in whom education is strongly correlated with age at marriage, fertility behaviour, use and demand for family planning, number of children desired, use of antenatal care, delivery in a health facility, vaccination and nutritional status of children, use of oral rehydration solutions, and infant and child mortality. The more years spent in schooling the better the outcome of these variables. There is a robust link between the economic status of women, the degree to which they are literate, in good health, and in jobs that pay a living wage, and the fertility rate. And yet even in India, where the adult literacy rate in 1993 was only 53 per cent and where 73 per cent of rural women of child-bearing age were illiterate, there are pockets of high literacy rates in places such as Goa, Tamil Nadu, and the exemplary state of Kerala, which has nearly 100 per cent literacy.

Another area of prevention is to improve sexual and reproductive health and education, making information and services in family planning more easily available.

Organizational issues

The organizational models that make sense are variations of the primary-health centre ® district hospital ® central hospital models. Some countries have gone a long way in this direction, coupled with decentralization. However, to make this model effective it is important to develop really efficient referral SYstems; to decentralize effectively powers for decision-making and implementation; to increase community involvement in decision-making; and to improve health education to optimize resource utilization. In order for decentralization not to become a meaningless exercise there is a need to study (a) how much choice actually is devolved to the periphery, (b) whether those choices are actually made, and (c) what the effects of those choices are. It is also essential to ensure that services are responsive to users, and that their implementation avoids polarization of services between rich and poor. Finally, there is a need to improve systems of regulation, supervision and monitoring.

Outreach programmes

Outreach programmes for chronic, non-life-threatening disabilities and deformities (cataracts, polio contractures, cardiovascular disease, etc.) are more cost-effective than duplicating major, expensive services in district hospitals. These are of benefit because they utilize resources more efficiently if they are from within the country; when they are from outside they are usually a welcome form of aid.

It is remarkable how much major surgery and anaesthesia can be undertaken in a rural setting. The cardiothoracic outreach programme from the Medical University of South Africa demonstrated, in a pilot project, the safety and cost-effectiveness of cardiothoracic surgery under primitive conditions: the standard of care improved, the local staff acquired good basic skills in patient care, the referral pattern improved, and community confidence in the services was increased. The risk:benefit ratio of the whole exercise was commendable.

Established in 1995, the Pacific Island Project, which is managed by the Australasian College of Surgeons and funded by AusAid, helps 10 widely scattered Pacific island countries (later expanded to include Papua New Guinea) with tertiary medical assistance. Members of several specialist medical colleges and societies volunteered their time to short-term visits of multidisciplinary teams; this was combined with transfer of skills and other educational activities. One hundred and thirty-one visits in 10 disciplines were conducted in the 11 countries by 255 voluntary participants between March 1995 and March 1998; 15 784 patients were seen and 3424 operations performed. Objective, independent assessment of the programme considered it very successful on all counts and it has now been extended for a further 3 years. It will be conducted in parallel with postgraduate educational programmes in the Pacific region and Papua New Guinea.

Other long-standing programmes, some made by flying-doctor services or by ship, have served many thousands.

Research and publication

Planning needs good data. Perhaps more than any research in the developing world what is needed are good, reliable, basic epidemiological studies. Hospital-based data can be misleading. Official statistics for trauma and accident mortality and morbidity often underestimate the true picture.

There are very few institutions in the developed world concentrating on training researchers in the developed world. The one outstanding institution is the International Clinical Epidemiological Network; it has been praised for its training programmes for clinical researchers, social scientists, statisticians, and health economists in developing countries; it has increased capacity for health researchin these countries through multidisciplinary and international collaboration.

Indexed journals and publications in them from developing countries are grossly under- represented; for example the number of journals indexed in Medline from Latin America has progressively declined during the past decade. No journal from Columbia has been indexed since 1991. The total output published in foreign journals from Columbia in the past 10 years has consisted of 521 articles, 219 of which came from four leading universities. Venezuela and Chile have fared better.

Improvement can be expected if there is an increased appreciation of the value of research and education on health care; identification of native and local strengths; collaborations with overseas institutions; and representation of experts (and there are many) from developing countries on editorial boards of international journals.

Technology issues

Even more than in economically privileged countries there is a need in developing countries to emphasize the role of good history-taking and clinical examination. Technology needs to be appropriate and introduced with care. The lack of equipment maintenance is wasteful. There is a need to study and to maximize equipment utilization patterns and to reduce ‘down-time'. Maintenance technician teams need to be trained, and to ensure local supply of parts. Initial purchase costs are less important than long-term averaged costs taking into account all factors. Tender boards need to be honest; and it is important to avoid buying equipment that would be rarely used or add little value. Technological assessment and equipment should be standardized nationally. Donor aid linked to equipment purchase that is likely to lead to technological mismatches should be avoided. Technology needs to be effective, culturally appropriate, affordable (not necessarily cheap), sustainable, politically tenable, and measurable.

Telemedicine has so far not made an impact in developing countries. More information on its utility is needed, and the cost is still too high. Its potential to improve access to health care and contain costs has remained theoretical so far.

Demographics, westernization and transitions: the future

Huge transitions are taking place in epidemiology, economics, demography, life-styles, expectations, cultures, and information relationships.

In many developing countries more than 50 per cent of the population is under the age of 15 years. In developed countries, one of the biggest public-health issues of the future is the demographic change leading to a much larger proportion of elderly people. This is already happening also in many developing countries. While infections predominate as causes of morbidity and mortality, more people are now dying of cardiovascular diseases: this is now the ‘silent epidemic', in which the second cause of death in developing countries after respiratory-tract infections is now ischaemic heart disease. Of the 50 million people who died in 1990, 6.3 million died from coronary arterial disease, and of these 57 per cent lived in developing countries. Smoking rates are increasing while stopping smoking is less successful in these countries. Diabetes is becoming a huge problem.

All these changes will affect surgical care by increasing ‘western' or ‘life-style' types of diseases and by diverting the already meagre resources away from primary surgery. In places such as Nepal, the incidence of Buerger's disease is 50 times higher than it is in North America—it is frightening to imagine the consequences of more smoking and diabetes in the future. The WHO estimates that the number of adults with diabetes in the world will rise from 135 million in 1995 to 300 million in 2025; the majority of this numerical increase will occur in developing countries. In developed countries there will be a 42 per cent increase (from 51 to 72 million) while in developing countries there will be a 170 per cent increase (from 84 to 228 million). By 2025 more than 75 per cent of people with diabetes will live in developing countries compared to 62 per cent in 1995. Countries with the largest numbers of diabetics in 2025 will be India, China, and the United States of America. There will be more diabetics among younger people and more diabetics will live in urban areas.

Hypertension and its sequelae are also becoming epidemic in the Caribbean and in the black population of the United States and United Kingdom.

Further reading

Abantanga FA, Mock CN. Childhood injuries in an urban area of Ghana: a hospital-based study of 677 cases. *Pediatric Surgery International* 1998; **13**(7): 515–18.

Abbasi K. The World Bank and world health: changing sides. *British Medical Journal* 1999; **318**: 865–9. [Important for understanding of World Bank role in health care in developing countries.]

Adib SM, Mufarrij AA, Shamseddine AI, Kahwaji SG, Issa P, el-Saghir NS. Cancer in Lebanon: an epidemiological review of the American University of Beirut Medical Center Tumor Registry (1983–1994). *Annals of Epidemiology* 1998; **8**(1): 46–51.

Akhtar SS, Abu Bakr MA, Dawi SA, Huq IU. Cancer in Libya—a retrospective study (1981–1985). *African Journal of Medicine and Medical Science* 1993; **22**(1): 17–24.

Ameh EA. Surgery in a rural hospital: experience from northern Nigeria. *East African Medical Journal* 1998; **75**(3): 180–3.

Ameh EA, Mbibu HN, Adams LM, Nmadu PT. Role of a general surgeon in obstetrics and gynaecology in a rural setting. *East African Medical Journal* 1998; **75**(1): 27–9.

Attah CA. Surgery in rural Nigeria—problems, challenges and recommendations. *South African Journal of Surgery* 1993; **31**(3): 106–8.

Balnchard RJW, Balnchard MEE, Toussignant P, Ahmed M, Smythe CM. The epidemiology and spectrum of surgical care in district hospitals of Pakistan. *American Journal of Public Health* 1987; **77**: 1439–45.

Bansal RK. Elementary education and its impact on health. [Editorial.] *British Medical Journal* 1999; **318**: 141. [Demonstrates clearly the impact of education, especially for women's health.]

Bener A, el-Rufaie OE, al-Suweidi NE. Pediatric injuries in an Arabian Gulf country. *Injury Prevention* 1997; **3**(3): 224–6.

Bewes PC. Priorities for surgically directed care in Africa. In: Ahmed M *et al.*, ed. *Surgery for all: a view from the developing world*, pp. 37–47. Ferozsons, Lahore, 1992. [An important contribution by experts from several countries.]

Deber R. Swan B. Canadian health expenditures: where do we really stand internationally? *Canadian Medical Association Journal* 1999; **160**: 1730–4.

Editorial. The hidden epidemic of cardiovascular disease. *Lancet* 1998; **352**: 1795.

Ghebrat K. Pattern of surgical admissions in Gondar Teaching Hospital, Ethiopia. *East African Medical Journal* 1997; **74**(12): 812–15.

Global Forum for Health Research. (1999). *The 10/90 Report on Health Research*. http.//www.globalforumhealth.org.report.htm

Gunning-Schepers LJ. Equity on both the scientific and the policy agendas. [Editorial.] *British Medical Journal* 1998; **316**: 1035–6. [Demonstrates need for equity.]

King H, Aubert RE, Herman WH. Global burden of diabetes, 1995–2025: prevalence, numerical estimates, and projections. *Diabetes Care* 1998; **21**(9): 1414–31.

King M, Bewes PC, Cairns J, Thornton J, ed. *Primary surgery*, Vol. 1, *Non-trauma*. Oxford University Press, 1990. [[A classic. Very valuable and inspiring.]

Klein M, Ramoroko SP, Jacobs AG, Bomela MD, Mokhobo KP, Mohlala ML. The Cardiothoracic Outreach Programme—a pilot project. *South African Medical Journal* 1996;**86**(12): 1533–5. [Demonstrates value of outreach programmes in developing countries.]

Landercasper J *et al.* Spectrum of general surgery in rural America. *Archives of Surgery* 1997; **132**(5): 494–6; discussion 496–8.

Logie DE, Benatar SR. Africa in the 21st century: can despair be turned to hope? *British Medical Journal* 1997; **315**: 1444–6.

McAuliffe MS, Henry B. Survey of nurse anesthesia practice, education, and regulation in 96 countries. *American Association of Nurse Anesthetists Journal* 1998; **66**(3): 273–86. [Doctors are not the only providers of health care.]

Macrae SM, Cook BV. A retrospective study of the cancer patterns among hospital in-patients in Botswana 1960–72. *British Journal of Cancer* 1975;**32**(1): 121–33.

Malik IA, Khan WA, Khan ZK. Pattern of malignant tumors observed in a university hospital: a retrospective analysis. *Journal of the Pakistan Medical Association* 1998; **48**(5): 120–2.

Mock CN, Jurkovich GJ, nii-Amon-Kotei D, Arreola-Risa C, Maier RV. Trauma mortality patterns in three nations at different economic levels: implications for global trauma SYSTEM development. *Journal of Trauma* 1998; **44**(5): 804–12; discussion 812–14. [Puts trauma care in perspective, although Ghana is better off than many others.]

Nordberg E, Holmberg S, Kiugu S. Output of major surgery in developing countries: towards a quantitative evaluation and planning tool. *Tropical and Geographic Medicine* 1995; **47**: 206–11. [Excellent demonstration of surgical workload in developing countries.]

Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. *International Journal of Cancer* 1999; **80**(6): 827–41. [Authoritative, but a little outdated, and patchy.]

Potter AR. Surgery in an African bush hospital. *Medical Journal of Australia* 1982; **2**(10): 469–71.

Rahim MA *et al.* Cancer morbidity. *Bangladesh Medical Research Council Bulletin* 1977; **3**(1): 17–21.

Rao KS, Pilli RD, Rao AS, Chalam PS. Sexual lifestyle of long distance lorry drivers in India: Questionnaire Survey. *British Medical Journal* 1999; **318**: 162–3. [Link between social behaviour and health.]

Razzak JA, Luby SP. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan, through the capture–recapture method. *International Journal of Epidemiology* 1998; **27**(5): 866–70.

Reacher MH, Munoz B, Alghassany A, Daar AS, Elbualy M, Taylor HR. A controlled trial of surgery for trachomatous trichiasis of the upper lid. *Archives of Ophthalmology* 1992; **110**(5): 667–74. [Carried out in Oman, but relevant to other countries; one of very few such trials.]

Rosselli D. Latin American biomedical publications: the case of Colombia in Medline. *Medical Education* 1998; **32**(3): 274–7.

Sen A. More than 100 million women are missing. *New York Review of Books*, 20 December, 1990. [Highlights the many deprivations of women in developing countries, a complex subject.]

Speth JM. Consumption in a global village: unequal and unbalanced. In: Speth JG, ed. *Human development report 1998*. Oxford University Press, 1998.

Steiner AK. Surgery and training in surgery in remote rural hospitals. *East African Medical Journal* 1996; **73**(12): 830–1.

Tanaka T. Childhood injuries in Japan. *Acta Paediatrica Japonica* 1993; **35**(3):179–85.

Theile DE, Bennett RC. The Pacific Islands Project: the first 3 years. *Australia and New Zealand Journal of Surgery* 1998; **68**(11): 792–8. [A very successful programme, demonstrates how one country can help many neighbours.]

Voak D, Caffrey EA, Barbara JA, Pollock A, Scott M, Contreras MC. Affordable safety for the blood supply in developed and developing countries. *Transfusion Medicine* 1998; **8**(2): 73–6.

Wasunna AE, Wyper DY. Technology for health in the future. *World Health Statistical Quarterly* 1998; **51**(1): 33–43.

Watters DA, Bayley AC. Training doctors and surgeons to meet the surgical needs of Africa. *British Medical Journal*(Clinical Research Edn) 1987; **295**: 761–3.

Watters DA, Dyke T. Trauma in Papua New Guinea: what do we know and where do we go?*Papua New Guinea Medical Journal* 1996; **39**(2): 121–5.

Weatherall DJ. The future role of molecular and cell biology in medical practice in the tropical countries. *British Medical Bulletin* 1998; **54**(2): 489–501.

Welbeck JE, Hesse AA. Pattern of childhood malignancy in Korle Bu Teaching Hospital, Ghana. *West African Journal of Medicine* 1998; **17**(2): 81–4.

White SM, Thorpe RG, Maine D. Emergency obstetric surgery performed by nurses in Zaire. *Lancet* 1987; **ii**: 612–13.

World Health Organization. *Anaesthesia at the district hospital*. WHO, Geneva, 1988.

World Health Organization. *General surgery at the district hospital*. WHO, Geneva, 1988.

World Health Organization. *Surgery at the district hospital: obstetrics, gynaecology, orthopaedics and traumatology*. WHO, Geneva, 1991. [Important contribution from an often misunderstood organization.]

World Health Organization. *Health education through religion*, No. 8, *Islamic ruling on male and female circumcision. The right path to health*. Non-serial publication of the WHO Regional Office for the Eastern Mediterranean, 1996.

World Health Organization/UNICEF/UNFPA. *Female genital mutilation: a joint statement*. WHO, Geneva, 1997. [Should be required reading from primary school upwards.]

Wright D. Telemedicine and developing countries. A report of study group 2 of the ITU Development Sector. *Journal of Telemedicine and Telecare* 1998; **4** (Suppl. 2): 1–85.

Yach D, Skov Jensen M, Norris A, Evans T. Promoting equity in health. *Promotion of Education* 1998; **5**(2): 7–13; 43. [A good place to start in understanding the need for equity.]

Zwi KJ, Zwi AB, Smettanikov E, Soderlund N, Logan S. Patterns of injury in children and adolescents presenting to a South African township health centre. *Injury Prevention* 1995; **1**(1): 26–30.

47.10 Special paediatric problems in developing countries

Hugh Grant

[Cultural factors](#)
[Coexistent diseases and risk factors](#)
- [Malnutrition](#)
- [Anaemia](#)
- [Human immunodeficiency virus](#)
- [Malaria](#)
- [Parasitic infections](#)

[Acute abdomen](#)
- [Diarrhoea and dysentery](#)
- [Amoebiasis](#)
- [Amoebic liver abscess](#)
- [Typhoid](#)
- [Abdominal tuberculosis](#)
- [Intussusception](#)
- [Ascariasis](#)
- [Schistosomiasis (bilharziasis)](#)
- [Hydatid disease](#)

[Snake bites](#)
- [Clinical features](#)
- [Conclusion](#)

[Further reading](#)

Forty per cent of the world's population are under 20 years of age. The proportion in developing countries is higher and accounts for a major part of the surgical workload. In addition to the common surgical problems seen in the developed world (congenital, traumatic, and neoplastic), many other problems affect children in the poor, developing countries. These problems may be either the consequences of intercurrent diseases, for example malnutrition, anaemia, infestations, or of tropical diseases, for example dysentery, worms, tuberculosis. Frequently, many factors conspire to confuse the primary signs and symptoms and lead to a higher morbidity and mortality than for equivalent disease in developed nations.

Cultural factors

Often surgeons working in remote areas of the developing world are not indigenous to the people among whom they work. Our background, beliefs, and culture are important in the way we feel, interpret, and discuss our symptoms. Failure to understand the local language and practices can prove a barrier to effective treatment. Children often present late: this may be due to poor transport, inadequate finance, delay in getting permission from elders, or because of treatment by traditional healers. This delay often means that surgery has a high rate of complications, reinforcing local beliefs that ‘Western' medicine is dangerous and that hospitals are places where one goes to die.

Treatments and medications by traditional healers vary enormously throughout the world. Many practices are merely misguided, but some are dangerous. Treatments are commonly administered orally, topically, or as an enema; their source is usually a carefully guarded secret but they are usually derived from leaves, roots, barks of trees, or seeds (Fig. 1).



Fig. 1. Traditional healer from South Africa with a selection of her remedies.

- In Southern Africa an enema is the most common ‘traditional' remedy. The indications for an enema vary from headaches, tiredness, and lassitude to abdominal pain, constipation, and diarrhoea. The common constituents are herbs, soaps, and leaves. Occasionally, potentially toxic substances are given, for example battery acid and drain cleaners.
- Cauterization is widely practised in North Africa: hot pokers are used to cauterize the painful organ, potentially confusing important abdominal signs.
- Venesection is practised in East Africa, particularly when a ‘poison' is thought to be the cause of the problem. Shock and haemodilution in the presence of pre-existing anaemia adversely affect outcome.
- Scarification, the making of superficial cuts over the painful organ, commonly produces, in addition to pain, secondary wound infection, depending on which herbs and dressings are applied.
- Steam and suction may be used for constipation or for worm infestations: children are placed over a bucket of boiling water. The negative pressure is intended to ‘suck' the stool or the worms out of the abdomen. It seldom achieves the intended result but invariably results in a major burn to the buttocks and perineum (Fig. 2).



Fig. 2. Burns sustained when a child was placed over a bucket containing boiling water in order to ‘suck out' the worms.

Coexistent diseases and risk factors

Children presenting to the paediatric surgeon in developing countries seldom have only one medical problem. In addition to the surgical pathology, they may also have malnutrition, anaemia, parasitic infections, malaria, infective diarrhoea, chest infections, or human immunodeficiency virus. These all adversely affect the outcome of surgery.

Malnutrition

There are estimated to be 800 million children who are malnourished with growth failure in the developing world. Malnutrition is the most significant reason for the increased morbidity and mortality of these children after surgery. Nutritional marasmus is defined by the World Health Organization (**WHO**) as protein energy malnutrition in a child whose weight has fallen to below 60 per cent of that expected. Affected children look emaciated, starved, and often have coexisting diseases (Fig. 3). Kwashiorkor is protein energy malnutrition associated with a high-carbohydrate diet. It is defined by the WHO as protein energy malnutrition in children whose weight has fallen to between 60 and 80 per cent of that expected, and associated with oedema. Affected children look oedematous, listless, and pot-bellied. They frequently have reddish, friable hair, depigmentation, diarrhoea, Gram-negative sepsis, and urinary-tract infections. They have low plasma proteins, with a raised plasma cortisol, hyponatraemia, atrophy of the pancreas, poor absorption of fat-soluble vitamins, villous atrophy of the small intestine, and fatty livers. They are prone to hypoglycaemia and hypothermia.



Fig. 3. Malnourished child receiving nutritional supplementation before surgery.

Anaemia

Anaemia in the tropics is multifactorial.

- Iron-deficiency anaemia is particularly common in premature babies, children who are breast fed for a long time, children with a diet deficient in iron and protein, and in diarrhoeal diseases.
- Folic acid deficiency occurs in kwashiorkor and in children with persistent diarrhoea.
- Chronic blood loss is frequent when children are infected with parasites, for example hookworm.
- Sickle-cell disease is the most common haemoglobinopathy and may either mimic a surgical problem or complicate it. In the homozygous form (sickle-cell disease), haemoglobin S is present in the cells instead of haemoglobin A. In conditions of stress, sepsis, hypoxia or acidosis, or after anaesthesia, haemoglobin S is precipitated in the red blood cells, resulting in the characteristic shape. These sickle cells are abnormally rigid and aggregate in the capillary beds causing intravascular thrombosis; theymay cause aplastic or painful crises. Pain can be felt in the long bones and joints, or the individual may present with an acute abdomen.

Human immunodeficiency virus

More than a third of the pregnant women in sub-Saharan Africa are infected with human immunodeficiency virus (**HIV**) type 1. Mother–child transmission can be as high as 30 per cent in developing countries. According to the WHO, 590 000 children became infected with HIV-1 in 1997. Children with HIV-1 infection may present incidentally or as a consequence of their HIV illness, which may take the form of infections, for example abscesses, tuberculosis, or fasciitis (Fig. 4), or neoplasia, for example Kaposi sarcoma and certain lymphomas. New surgical presentations of HIV-1 infection have been described, for example acquired rectovaginal fistulas in young girls. Whereas HIV-1 infection itself is not associated with an increase in surgical morbidity, there is an increase in morbidity and mortality in children with the acquired immune-deficiency SYndrome who develop surgical complications.



Fig. 4. Necrotizing fasciitis in a child infected with HIV-1.

Malaria

Malaria is endemic throughout the tropics and kills at least 2 million people per year. The parasites responsible are *Plasmodium malariae, P. vivax, P. ovale*, and *P. falciparum*. Newborn babies have passive immunity from their mothers, which lasts a few months: older children become infected and in severe attacks develop fever, anorexia, and frequently diarrhoea and vomiting. Repeated infections result in hepatosplenomegaly, anaemia, and failure to thrive.

Parasitic infections

All forms of intestinal parasites are endemic in the tropics and the subtropics, where heavy infestation can lead to chronic ill heath, anaemia, failure to thrive, and surgical complications. Surgical practice should be modified to take into account all of these potential problems (Table 1).

Anaemia:	*Immune suppression:*
Poor tissue oxygenation	Postoperative infections
Impaired wound healing	Superinfection
Hyperdynamic circulation	Impaired metabolic response
Malnutrition:	*Diarrhoeal diseases:*
Immune suppression	Malabsorption
Postoperative infections	Fluid losses
Impaired wound healing	Delay in enteral feeding
Hypothermia	*Malaria:*
Hypoglycaemia	Additional septic stress

Table 1 Risk factors for surgery

Acute abdomen

Diarrhoea and dysentery

Diarrhoea and dysentery are endemic in developing countries. They are a cause of chronic ill health but often present with an acute abdomen. *Campylobacter* and *Salmonella* spp. have similar symptoms and signs, and may present with acute abdominal pain, diarrhoea, or blood and mucus in the stool. Pain may be severe enough to mimic an acute abdomen.

Shigella spp. cause bacillary dysentery. After an incubation of 2 to 4 days, illness starts with fever and watery diarrhoea. This is followed by bloody diarrhoea, mucus, tenesmus, and acute abdominal pain, often with signs of peritonitis. Complications are frequent and include toxic megacolon, perforation, ascites, overwhelming septicaemia, haemolytic–uraemic syndrome, and purpura.

Clinical features

Diagnosis is obvious during an outbreak. In sporadic cases it may be difficult to distinguish shigellosis from an acute surgical abdomen: the children are toxic and pyrexial with a distended, tender abdomen. The history is important; urgent microscopy of the stool may be helpful; subsequent culture will detect enteropathogenic organisms. A very raised white blood count (greater than 35×10^9/l), and hyponatraemia (less than 130 mmol/l) are suggestive of shigellosis. A high temperature and gas extending as far as the rectum on a plain abdominal radiograph are suggestive of a medical rather than a surgical cause. The main differential diagnoses are intussusception (which can frequently be felt rectally) and amoebiasis (in which ulcers can be seen on sigmoidoscopy).

Treatment

The treatment of these conditions is vigorous fluid resuscitation, decompression by nasogastric intubation, and appropriate antibiotics. Surgery should be reserved only for complications, for example perforation or severe haemorrhage. Clinical signs of apparent peritonitis in shigellosis may indicate transmural inflammation and are not an indication for laparotomy unless there is a perforation (as demonstrated by free air in the peritoneal cavity).

Amoebiasis

Ten per cent of the world's population are infected with *Entamoeba histolytica*. Children become infected when they ingest cysts from infected food or water. The cysts release trophozoites in the small intestine that infect the colon. Disease only occurs from pathogenic strains that invade the bowel wall and blood vessels, causing vasculitis, thrombosis, and local infarction. If necrosis is transmural, perforation occurs. Affected areas of the colon are plugged by ‘omental wraps' that limit perforation and faecal soiling. Subsequent healing may occur by revascularization and may present later with obstruction due to fibrosis.

Clinical features

Children typically present with bloody diarrhoea, mucus, the passage of slough per rectum, abdominal pain, and tenesmus. Toxic megacolon, haemorrhage, peritonitis, and shock may follow. Often the peritoneal signs are surprisingly unimpressive because of the omental wraps. Diagnosis is made by sigmoidoscopy (ulcers can be seen in the rectum), or by examination of freshly passed stool or slough by microscopy (haematophagous trophozoites will be seen). Abdominal radiographs may be non-specific, but occasionally a fixed loop of colon may be seen. Serological confirmation is by the amoebic gel-diffusion test.

Treatment

The majority of patients settle with medical management. If no complications have occurred, supportive treatment is required: nasogastric suction, intravenous antibiotics (metronidazole 7.5 mg/kg 8-hourly for the amoebiasis, and ampicillin 25 mg/kg 6-hourly, gentamicin 2.5 mg/kg 8-hourly to cover Gram-negative organisms). Diloxanide furoate (20 mg/kg per day in three divided doses for 10 days) eradicates cysts from the lumen of the bowel. If peritonitis has occurred, laparotomy is essential. A urethral catheter should be inserted, and paediatric anaesthetic tubing placed in the rectum for on-table lavage. An upper transverse incision gives good access: frequently there is gross soiling, with patches of colonic infarction surrounded by omental wraps. Total or partial colectomy is accompanied by high mortality and should be avoided if possible. A loop ileostomy placed in the terminal ileum should be raised, and a Foley catheter can be placed in the distal loop for on-table irrigation. Small holes in the colon (demonstrated by leaking irrigation fluid) should be sutured. The peritoneal cavity should be carefully lavaged with saline. Perforated amoebiasis has a high postoperative mortality.

Amoebic liver abscess

Abscesses in the liver develop when amoebae travel up the portal vein from the intestine. Tissue necrosis occurs and the small septic foci may aggregate into large abscesses.

Clinical features

Symptoms are often less dramatic than in amoebic dysentery (anorexia, weight loss, fever, cough, or pleuritic chest pain) and only 10 per cent have a history of dysentery. On examination the patients are often chronically unwell, malnourished and febrile, with tender hepatomegaly, and may have pleural effusions. The differential diagnosis includes pyogenic liver abscess, subphrenic abscess, hepatobiliary ascariasis, and hydatid disease. Diagnosis has been simplified by ultrasonography. Complications include pneumonia, pleural effusions, portal venous thrombosis, pericardial effusions, and rupture into the peritoneal cavity.

Treatment

Treatment is by intravenous metronidazole (7.5 mg/kg 8-hourly) as long as the abscess is not about to rupture. Aspiration is seldom required and may introduce secondary infection. Since the cavity is due to necrosis, the ultrasonographic appearance may persist although the child has clinically recovered.

Typhoid

Asymptomatic carriage of *Salmonella typhi* is common. Infection occurs from contaminated food or water: ingested bacteria attach to the intestinal mucosa and are transported in the lymphatics to the mesenteric nodes. There they multiply and enter the bloodstream via the thoracic duct and are disseminated to the liver, gallbladder, bone marrow, and spleen. Reinfection of the intestine occurs from the bile. Organisms multiply in the macrophages located in the intestinal lymph nodes. If the inflammation persists, necrosis occurs and the lymph nodes ulcerate, leading to bleeding and perforation.

Clinical features

The clinical course is variable: persisting and rising fever with headaches lasts about 3 weeks and is associated with malaise, aches and pains, and cough. Abdominal symptoms include constipation and diarrhoea. Physical signs depend on the severity of the disease. Rose spots that blanch on pressure can be difficult to see in dark skins. The white count is usually normal, with leucopenia and a relative lymphocytosis. The Widal test is diagnostic but interpretation can be difficult because of vaccination and seroprevalence. Cultures should be taken from the blood, marrow, stool, and urine if the diagnosis is suspected. Other serious medical complications include cholecystitis, osteomyelitis, pneumonia, and meningitis.

Treatment

Treatment is primarily medical: intravenous chloramphenicol (12.5–25 mg/kg 6-hourly) or amoxycillin (25 mg/kg 8-hourly) were the most commonly used antibiotics but there is increasing resistance. Third-generation cephalosporins and ciprofloxacin are now frequently used.

Surgery is indicated for complications—peritonitis, perforation, and haemorrhage. Using an upper transverse incision, the small holes in the intestine should be excised locally and closed primarily, and peritoneal lavage of the soiled abdominal cavity performed. Massive resections are seldom required.

Abdominal tuberculosis

Abdominal tuberculosis results from the ingestion of organisms, from infected bile, by haematogenous spread, or directly from adjacent organs. When the organism reaches the intestine it is engulfed by macrophages, which release cytokines causing local tissue damage—endarteritis, microthrombosis, and necrosis. Subsequent healing involves fibrosis and cicatrization.

Abdominal tuberculosis may predominantly affect the intestine, nodes, or peritoneum. Intestinal disease tends to be ulcerative, leading to a perforation; variations include ulceroconstrictive disease which leads to obstruction, or ulcerohypertrophic disease which forms a mass. Peritoneal tuberculosis may appear as thickened peritoneum and omentum with visible nodules and adhesions, or as ascites. Nodal tuberculosis affects the mesenteric and retroperitoneal lymph nodes, which exhibit caseation necrosis.

Clinical features

Abdominal tuberculosis may present in many ways and the surgeon must always consider it in the differential diagnosis. Frequently the signs and symptoms are non-specific. The most common SYmptoms are anorexia, weight loss, distension, abdominal pain, vomiting, and pyrexia. The disease may be associated with obvious pulmonary tuberculosis, in which case the diagnosis is more evident. Examination may reveal a palpable mass or ascites. A positive Mantoux test is diagnostic in developing countries (as few children receive the bacillus Calmette–Guérin vaccine). A negative Mantoux test does not exclude the diagnosis, as malnourished children may be so anergic they are unable to mount a host response. A chest radiograph may demonstrate pulmonary disease. Abdominal radiographs may demonstrate calcified lymph nodes but are usually non-specific. The erythrocyte sedimentation rate is usually elevated.

Treatment

Treatment is primarily medical: chemotherapy depends on local drug resistance, but a widely used regimen is rifampicin 10–20 mg/kg per day, isoniazid 10–20 mg/kg per day, pyrazinamide 15–30 mg/kg per day, and ethambutol 15 mg/kg per day.

Surgery should be reserved for tissue diagnosis or for complications: for example, acute or persistent obstruction, perforation, uncontrolled haemorrhage, or fistulas. Tissue diagnosis is best obtained by either a minilaparotomy or by judicious laparoscopy. There may be adhesions, which make laparoscopy hazardous.

Surgery for acute complications in the untreated patient is difficult. There are frequently dense adhesions, sealed leaks, and fistulas. Minimal dissection should be performed. Local excision with primary closure and exteriorization have both been advocated. In the context of a sick, malnourished child, exteriorization with a controlled fistula is a wiser course. A distal loopogram is necessary before closing the stoma to exclude strictures.

Intussusception

There are differences in the clinical features of intussusception between tropical and temperate countries. In addition to the common, 3- to 9-month-old age group there is also a group of older children who present with colocolic lesions.

Clinical features

Associated diarrhoea is common, there is a high association with parasitic infestations, and a low incidence of primary bowel lesions. It is easily confused with dysentery, amoebiasis, or worm bolus obstruction, and children often present late. Abdominal examination may reveal a mass, but frequently the child is so distended by the long-standing obstruction that a mass is not palpable. Rectal examination reveals a mass in 30 per cent of cases. Sigmoidoscopy will exclude amoebiasis as the cause of the blood and diarrhoea, and a plain radiograph of the abdomen may demonstrate a soft-tissue mass.

Treatment

Initial management includes nasogastric intubation, vigorous resuscitation with intravenous fluids, analgesia and antibiotics (tocover Gram-negative and anaerobic organisms: ampicillin 25mg/kg 6-hourly, gentamicin 1.5 mg/kg 8-hourly, metronidazole 7.5 mg/kg 8-hourly). Reduction by air enema may be appropriate in children who present early and in those who have no evidence of bowel ischaemia. This approach is less commonly feasible in developing countries and laparotomy is frequently required because of complications. An upper transverse incision should be used for good exposure; the intussusception should be squeezed in a retrograde manner. It should never be pulled on, as this leads to serosal tears and damage. Time and patience are crucial. If the bowel looks suspect or if resection appears necessary, bulldog clamps should be applied to the feeding mesenteric vessels before reduction or resection are attempted. This prevents the release of toxins and free radicals that have been localized within the intussusception and may be released on reduction. If the bowel is necrotic, local resection is required and a primary anastomosis or defunctioning ileostomy/colostomy performed if indicated.

Ascariasis

Ascaris lumbricoides inhabits the small intestine from the duodenum to the ileocaecal valve. Infestation occurs where unsanitary conditions prevail: ingested ova hatch in the small intestine; the larvae are carried via the portal circulation to the liver and lungs; they are then coughed up the respiratory tract and swallowed; mature worms subsequently thrive in the small intestine.

Intestinal ascariasis

Clinical features

The common presentations of ascariasis to the surgeon are worm bolus obstruction in the small intestine or in the common bile duct. Since ascariasis is endemic in the tropics, the presence of these worms does not necessarily mean they are the cause of the child's symptoms. The diagnosis of worm bolus obstruction should be considered in a child with obstructive symptoms (bile-stained vomiting, colicky pain, abdominal distension, constipation) with an abdominal mass (which may change in size and location). The differential diagnosis includes intussusception, tuberculosis, and amoebiasis. The diagnosis of worm bolus obstruction should only be accepted if the collection of worms visible on plain abdominal radiograph corresponds with the palpable mass on clinical examination (Fig. 5).



Fig. 5. Plain abdominal radiograph demonstrating a worm bolus obstruction.

Treatment

In the absence of complications, (peritonitis, volvulus, infarction), treatment should initially be conservative. Nasogastric suction, intravenous fluid resuscitation, analgesia, antispasmodics, and antibiotics (e.g. ampicillin, gentamicin, and metronidazole). Round worms move up and down the intestine diurnally and the majority untangle themselves spontaneously. Antihelminthic treatment should not be given until after the obstructive features have settled, otherwise the worms will remain entangled and pressure necrosis may occur in the intestine. Mebendazole (100 mg twice daily for 3 days) and albendazole (200 mg, single dose) kill the worms whereas piperazine (120 mg/kg, single dose) paralyses them.

Exploratory laparotomy is indicated in children whose symptoms do not settle, or in whom there are signs of peritonitis, volvulus or a suspicion of bowel ischaemia. If the bowel is viable an enterotomy should be performed a short distance away from the bolus and the worms removed with sponge-holding forceps. Piperazine should be infused up and down the lumen of the small bowel and the anaesthetist should also give piperazine down the nasogastric tube. Simple closure of the enterotomy should then be performed. If there is a volvulus or infarction the affected bowel should be resected and a primary anastomosis performed.

Hepatobiliary ascariasis

Clinical features

Hepatobiliary ascariasis also presents with colicky abdominal pain and vomiting. It was under-reported before the advent of ultrasonography, but is now a commonly diagnosed condition. The children are seldom jaundiced (because the bile trickles past the worms) but their state may be complicated by liver abscesses and pancreatitis.

Treatment

Treatment should initially be conservative: nasogastric suction, intravenous fluids, antibiotics, antispasmodics, and pethidine. Some suggest the use of oxygen, as the worms are anaerobic. Initially conservative treatment should allow the worms to disentangle and migrate down the common bile duct into the duodenum. Once the symptoms have settled, albendazole, mebendazole, or piperazine can be given orally.

If the symptoms do not settle, endoscopic retrograde cholangiopancreatography (ERCP) is only feasible if there are a few worms in the common bile duct. If the ultrasonogram suggests a huge number of worms, surgical exploration is indicated. Via an upper transverse incision, the common bile duct should be explored and all worms in the duct and liver removed. Surface abscesses of the liver should be explored and any worms removed from within the liver substance. Piperazine should be instilled into the intrahepatic bile ducts and down the common bile duct into the duodenum. In practice it is seldom possible to remove every worm and frequently one or two are left within the liver. If these cause symptoms they can be removed via the T-tube, which should be left in for 10 days and a cholangiogram performed before its removal. Adequate treatment must be given to prevent recolonization of the intestines and common bile duct.

Schistosomiasis (bilharziasis)

Intestinal schistosomiasis is usually caused by *Schistosoma mansoni* or *S. japonicum*. Adult worms reside in the mesenteric vessels and lay eggs that are passed in the stool. In water the eggs hatch into larvae and pass into a water snail, where they mature, and finally penetrate the skin of humans when they bathe in contaminated water. They make their way to the liver and finally settle in the tributaries of the inferior or superior mesenteric veins. Embolization of ova from the mesenteric vessels to the liver results in a fibrotic response.

Clinical features

Acute intestinal symptoms occur 2 to 3 weeks after infection: they include dysentery associated with fever and abdominal pain. Exacerbations occur every few months if not treated. Chronic infection affects the colon mainly. The ova deposited by the adult worms stimulate an inflammatory response that results in granuloma formation, ulceration, and polyps. Chronic liver involvement results in periportal fibrosis and subsequent portal hypertension. Liver function is maintained until the disease becomes advanced.

The diagnosis is made by finding ova in the stools, or by examination of rectal snippings. Periportal fibrosis gives a characteristic appearance on ultrasound.

Treatment

Treatment is medical: praziquantel (40 mg/kg, single dose) kills the adult worms and treats acute disease; it also halts progression of the chronic form of the disease.

Hydatid disease

The adult worm lives in the dog's intestines. Ova are excreted and ingested by people (who are an intermediate host). The ova enter the portal circulation and are most commonly spread to the liver, lungs, and brain. These develop into cysts that are composed of an inner layer (germinal epithelium containing brood capsules) and an outer layer (fibrous response from the host).

Clinical features

Seventy per cent of cysts are found in the liver. They present to the surgeon because of painful hepatomegaly. Diagnosis is by ultrasonography, with serological confirmation by enzyme-linked immunosorbent assay.

Treatment

Treatment is primarily medical: albendazole (5 mg/kg twice daily for 4 weeks) can cause the cysts to regress. Surgery may be required for complications or if the cyst is about to rupture. An upper transverse or subcostal incision gives good exposure. Towels soaked in hypertonic saline should be placed around the operative field to prevent seeding of any spilled brood capsules. The cyst is then aspirated and hypertonic saline injected into the cyst and left for 10 min. The inner (germinal) capsule is then removed (Fig. 6). Any bleeding vessels or large bile ductules are tied off.

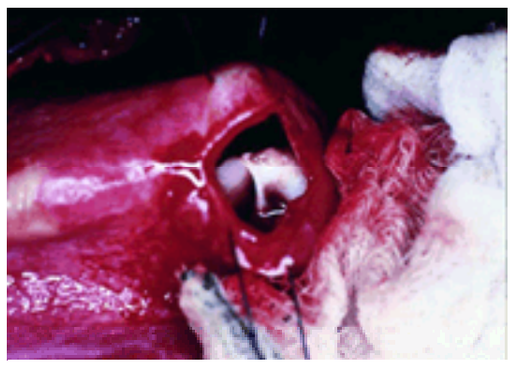


Fig. 6. Hydatid cyst in the liver, demonstrating the inner germinal layer and the saline soaks around the operative field.

Snake bites

Children are frequently bitten by snakes and it is rare to get a clear history or accurate description of the snake. The presence of one or two fang marks in the wound suggests that the bite was by a venomous snake but does not necessarily mean that venom was injected.

Clinical features

There are three types of envenomation:

1. Local cytotoxic venom causes pain, bruising, swelling, and blistering (Fig. 7). Compartment syndrome may occur and is suggested by pain, paraesthesia, pallor, and pain on passive stretching. Absent pulsation is a very late sign.



Fig. 7. Cytotoxic effect of a snake bite.

2. Neurological effects result in vomiting, blurred vision, paraesthesia, salivation, ptosis, loss of upward gaze, and loss of bulbar and respiratory function.
3. The haematological effects of envenomation are characterized by spontaneous haemorrhage and incoagulable blood.

Snake venom usually contains varying concentrations of all three types of venom but they seldom all cause problems in the one patient.

Treatment

Simple cleaning of the wound is essential. Cauterization, suction, or deliberate bleeding should be discouraged. The affected limb should be rested, if necessary with a splint. If there is swelling in the limb, the affected limb should be kept at the level of the heart and not elevated. Tourniquets should never be applied (Fig. 8). Antibiotics and antitetanus prophylaxis should be given as indicated. Careful examination for signs envenomation should be made. If the effects are mild, antivenom is not necessary. If there are signs of severe envenomation, antivenom should be considered. It is vital to determine first whether the serum is appropriate for the species of snake that might be involved. When giving antivenom the clinician must have adrenaline, antihistamines, and steroids ready for possible anaphylactic side-effects.



Fig. 8. This tourniquet did more damage than the snake bite.

Local cytotoxic effects may need to be treated by surgical decompression. Appropriate antivenom should be given in cases of marked swelling. Compartment SYndrome is a potentially serious complication of cytotoxic envenomation. The intracompartmental pressures can be measured with a large bore (19G) needle attached to a transducer. If this is not available it can be determined clinically by pain, paraesthesia, pallor, and pain on passive stretching. Absent pulsation is a late sign and indicates that the limb is beyond salvage. If signs of increased compartmental pressure exist, prompt surgical decompression is necessary (Fig. 9).



Fig. 9. Surgical decompression of compartment syndrome in the forearm following a snake bite.

Conclusion

Surgery in the developing countries is totally different from that practised in Europe and the United States of America. Patients present late, are often in septic shock, and have intercurrent diseases that may be life-threatening in their own right. There may be no experienced anaesthetist to help with the peri- and postoperative care, nor a unit in which to support and monitor critically ill patients. In these circumstances it is imperative to apply the following principles:

- full resuscitation before surgery
- keep surgery simple—only do what is necessary and avoid risky anastomoses
- consider defunctioning stomas rather than massive resections where appropriate
- early enteral feeding
- aggressive physiotherapy and postoperative mobilization

Further reading

Axton J. Paediatrics in developing countries. In: Levene MI, ed. *Jolly's diseases of children*, 6th edn, pp. 529–75. Blackwell, Oxford, 1991. [A broad picture of the medical problems affecting children in the developing word.]

Grant HW, Buccimazza I, Hadley GP. A comparison of colo-colic and ileo-colic intussusception. *Journal of Pediatric Surgery* 1996; **31**(12): 1607–10. [Highlights the different nature of intussusception in the tropics.]

Grant HW, Hadley GP, Wiersma R, Rollins N. Surgical lessons learned from the *Shigella dysenteriae* type 1 epidemic. *Journal of the Royal College of Surgeons of Edinburgh* 1998; **43**: 160–2. [Clinical experience of the surgical problems of shigellosis.]

Jeena PM, Coovadia HM, Bhagwanjee S. Prospective, controlled study of the outcome of HIV-1 antibody positive children admitted to an intensive care unit. *Critical Care* 1996; **24**: 963–7. [Important prospective study of outcome in HIV-1 positive children.]

Mars M, Hadley GP, Aitchison JM. Direct intracompartmental pressure measurement in the management of snakebites in children. *South African Medical Journal* 1991; **80**(5): 227–8. [Practical approach to a clinical problem.]

Millar AJW, Rode H, Cywes S. Abdominal tuberculosis in children: surgical management: a ten year review of 95 cases. *Pediatric Surgery International* 1990; **5**(6): 392–6. [A summary of tuberculosis in children.]

***Report of the global HIV/AIDS epidemic*.** Geneva: WHO, 1998. Report ISBN 92 4 156189 0. [Epidemiology of the HIV/AIDS pandemic.]

Warrell DA. Snake bite and other envenoming. In: Ellis BW, ed. *Hamilton Bailey's emergency surgery*, 12th edn, pp. 850–9. Butterworth Heinemann, Oxford, 1995. [A synopsis of snakebite management.]

Watters D, Kiire C. *Gastroenterology in the tropics and subtropics, a practical approach*, 1st edn. Macmillan, London, 1995. [A useful concise look at surgical problems in the tropics.]

Weaver LT. Nutrition. In: Campbell AGM, Macintosh N, ed. *Forfar and Arneil's textbook of pediatrics*, 5th edn, pp. 1179–230. Churchill Livingstone, Edinburgh, 1998. [A background on nutrition.]

Wiersma R, Hadley GP, Luvuno FM. Amoebiasis in children: pathology, diagnosis and treatment. *Pediatric Surgery International* 1990; **5**(6): 406–10. [Pragmatic approach to surgical management of amoebiasis.]

48.1 Cancer genetics

K. L. Nathanson, T. C. Smyrk, M. S. Brose, H. T. Lynch and B. L. Weber

Introduction
Autosomal dominant cancer susceptibilty syndromes
Breast and ovarian cancer
Colorectal carcinoma
Neurofibromatosis
Retinoblastoma
von Hippel-Lindau (VHL) disease
Cowden and Bannayan-Riley-Ruvalcaba syndromes
Genetics of other common cancers
Malignant melanoma
Pancreatic cancer
Gastric carcinoma
Prostate cancer
Lung cancer
Renal cell carcinoma (RCC)
Conclusion
Overview of patient assessment for hereditary cancer
Further reading

Introduction

The field of cancer genetics has changed dramatically over the past decade. When the first edition of this textbook was written, only one gene, *APC*, in which germline mutations confer a greatly increased susceptibility to cancer had been isolated. Now at least 40 such genes have been identified, with more characterized each year. The purpose of this chapter is to review some of the more common autosomal dominant cancer susceptibility syndromes, and what is known about the genetics of some common cancers (Table 1). Each section reviews the features of the cancers, gives a brief synopsis of the genetics and gives some screening guidelines.



Table 1 A partial list of known cancer susceptibility syndromes

Inherited predisposition to cancer is important for three reasons. First, individuals who are predisposed to cancer are a high-risk population who may benefit from efforts at cancer prevention, or from early diagnosis and treatment of cancer. Secondly, genetic predisposition may tell us about the development of cancers in general. The genes that play a role in the development of hereditary cancers have, in many instances, been demonstrated to be important in the development of sporadic or non-hereditary cancers. Third, many of the genes which, when mutated are ‘cancer genes', may have important normal functions in growth and development. Through research in families with an inherited predisposition to cancer those genes can be identified and characterized.

The most important person in the management of patients and families with hereditary cancer syndromes is the clinician who initially suspects such a diagnosis. Therefore, physicians who are aware of hereditary cancer SYndromes and ask about a family history of cancer are invaluable. Patients and their family members can be screened for mutations in cancer susceptibility genes and those with mutations can be enrolled in surveillance and management programs tailored to the natural history of a particular hereditary disorder, while those lacking the mutation can revert to general population cancer screening guidelines. However, it is important that testing for mutations in cancer susceptibility genes be undertaken by those physicians experienced in, and knowledgeable about, the ramifications of testing and the implications of the mutation test result, whether positive or negative.

Autosomal dominant cancer susceptibilty syndromes

Breast and ovarian cancer

Overview

Breast cancer is the most common form of non-cutaneous cancer in American women. In the United States alone, there are over 186 000 new cases and 46 000 deaths each year. It is estimated that 12.5 per cent of all American women (1 in 8) will develop breast cancer and 3.4 per cent will die from it. Current estimates suggest that family history accounts for approximately 10 to 20 per cent of breast cancer in the United States.

Multiple studies have shown an increased breast cancer risk among first-degree relatives of women with breast cancer. However, breast cancer is a complex disease and can aggregate in families for multiple reasons. Mutations in high penetrance cancer susceptibility genes, such as *BRCA1* and *BRCA2*, are inherited in an autosomal dominant fashion and confer a greatly elevated risk of breast cancer compared to the general population. BRCA1 and BRCA2 both function in cellular response to DNA damage. The DNA damage response pathway may be crucial in cancer development, as allowing mutations to persist in cells could allow them to acquire the cellular phenotype necessary for uncontrolled growth. There are other rarer autosomal dominant cancer susceptibility syndromes, such as Li-Fraumeni, Cowden, and Muir-Torre syndromes (reviewed later in the chapter), which are associated with a greatly elevated risk of breast cancer as well as with other findings in an individual or family. There are presumably other more common breast cancer susceptibility genes, as yet unidentified, in which mutations confer a lower risk than *BRCA1* and *BRCA2*. In addition, variants in other genes, such as the P450 enzymes, may confer small increases in susceptibility to breast cancer. Together, variants in these genes are likely to play a greater role in predisposition to breast cancer at the population level than the more striking susceptibility due to *BRCA1* and *BRCA2* mutations. In addition to the genetic factors that confer increased susceptibility, there may be other non-inherited reasons, which produce clustering of breast cancer in certain families. These could include: (1) geographically limited environmental exposure to carcinogens which might affect an extended family living in close proximity; (2) culturally motivated behavior that alters the risk factor profile, such as age at first live birth or contraceptive choice, and (3) socioeconomic influences that, for example, might result in differing dietary exposure.

BRCA1 and BRCA2

Mutations in both *BRCA1* and *BRCA2* lead to an autosomal dominant cancer susceptibility syndrome conferring an increased risk of breast and ovarian cancer to female mutation carriers. Current estimates of the breast cancer risk in *BRCA1* and *BRCA2* mutation carriers range from 50 to 70 per cent respectively, based on estimates from linkage families and studies in the Ashkenazi Jewish population. The estimated risk of ovarian cancer is lower in *BRCA2* mutation carriers (10–20 per cent) than for *BRCA1* mutation carriers (20–40 per cent). In families segregating *BRCA1* and *BRCA2* mutations there is an increased risk for cancers other than breast and ovarian. In *BRCA1* mutation carriers, there is an increased risk of prostate cancer (4-fold), which has held up across studies. In 173 *BRCA2* families, increased relative risks for prostate (4.65-fold), pancreatic (3.5-fold), stomach (2.6-fold), gallbladder (5-fold), melanoma (2.6-fold), and buccal cavity/pharynx cancer (2.4-fold) cancers were noted. The overall cumulative risk of cancer by age 70 was estimated to be 32 per cent in male carriers and 70 per cent in female carriers of a *BRCA2*

mutation. Taken together these data suggest that while *BRCA2* mutations may confer a lower risk of ovarian cancer than *BRCA1* mutations, they confer a higher risk of 'other' cancers.

The majority of the mutations reported in *BRCA1* are truncating mutations scattered evenly through out the gene. A tabulated list of identified mutations in both genes and details of the detection methodology are available on the Breast Cancer Information Core (BIC) website: http://www.nhgri.nih.gov/Intramural-research/Lab-transfer/Bic/). Truncating mutations presumably render *BRCA1* non-functional, and are therefore thought to be disease associated. However, a number of the mutations reported in *BRCA1* are mis-sense mutations. As these can include polymorphisms and unclassified variants as well as true disease-associated mutations, care must be taken in interpreting the clinical significance of mis-sense mutations. Partial genomic deletions of *BRCA1* also have been found, which are presumably due to the high number of Alu repeats in *BRCA1* introns. In the Dutch population, three large genomic deletions are estimated to account for approximately 30 per cent of all Dutch breast cancer families. In other populations, the proportion of genomic is estimated at 10 to 15 per cent. While the spectrum of *BRCA2* mutations has not been as well defined as that of *BRCA1*, it is clear that there are several similarities to *BRCA1*. Mutations are spread over the entire gene; again frameshift and nonsense mutations are predominant and more than one-half of the mutations appear only once in the BIC. Thirty per cent of the mutations reported are mis-sense mutations, most of which are unclassified variants. While a genomic deletion has been described which encompasses exon 3 of *BRCA2*, presumably this mutation source in *BRCA2* will be less common than in *BRCA1*, as the region surrounding *BRCA2* is not as dense in Alu repeats as *BRCA1*.

BRCA1 and BRCA2 associated tumors

Somatic (non-heritable) mutations in *BRCA1* have not been found in sporadic breast tumors and are rarely found in sporadic ovarian tumors. *BRCA2* is similar, in that mutations are rarely found in sporadic breast and ovarian cancers. The reasons for this surprising finding are unclear, but it is speculated that that other mechanisms may down regulate or inactivate *BRCA1* and *BRCA2* proteins in sporadic tumors.

In order to evaluate the pathology of breast cancers in *BRCA1* and *BRCA2* mutation carriers, a large study by the Breast Cancer Linkage Consortium has been done examining breast cancers from known mutation carriers. Breast cancers from both *BRCA1* and *BRCA2* mutation carriers have a higher overall grade than sporadic cases. The breast cancers of *BRCA1* mutation carriers showed more pleomorphism, continuous pushing margins and lymphocytic infiltration, and a higher mitotic count and less tubule formation than sporadic tumors. A lower mitotic count, less tubule formation, and continuous pushing margins were features of the breast cancers in *BRCA2* mutation carriers. Medullary or atypical medullary carcinoma also was more common in *BRCA1* mutation carriers than in *BRCA2* mutation carriers. *BRCA1* mutation carriers were less likely to have ductal carcinoma *in situ* (DCIS) (45 versus 55 per cent) but the rate in *BRCA2* mutation carriers did not differ from controls. Breast cancers from *BRCA1* mutation carriers also have been found to be less likely to be estrogen and progestrone receptor positive than tumors from *BRCA2* mutation carriers and familial non-*BRCA1,2* cases in two studies. Taken together these data suggest that breast cancers in *BRCA1* mutation carriers are more likely to be higher grade than in sporadic or *BRCA2* mutation associated tumors. Survival studies will need to be done prospectively to avoid survivor bias and have just begun.

Risk assessment and clinical application

Two types of risk assessment models can be used to assist patients with a family history of breast cancer. The first, exemplified by the 'Claus model', predicts lifetime risk of developing breast cancer based solely on family history. The Claus model is based on 4730 patients with histogically proven breast cancer ages 20 to 54, and 4688 controls, which were age and frequency matched to cases from several SEER (Survey, Epidemiology and End Results Program) sites. In this model, breast cancer risk is based on the number and relatedness of affected relatives and their ages of breast cancer diagnosis. Another such model is the 'Gail model' in which risk evaluation is based on number of first degree relatives with breast cancer, number of previous breast biopsies, age at menarche, and first childbirth. The Gail model is best used in Caucasian women undergoing yearly mammography, and may be inaccurate in estimating risk in moderate and high risk breast cancer families, as it does not take into account the age of breast cancer diagnosis and paternal family history of cancer. A drawback to both the Claus and Gail models is that they are entirely based on Caucasian populations, making risk assessment in non-Caucasian populations potentially inaccurate.

The second type of model predicts the probability of finding a *BRCA1* or *BRCA2* mutation in the family, based on family history and ethnic background. With these models, it is important to remember that the probabilities estimated are for affected individuals and must be adjusted accordingly for other family members (e.g. half the likelihood for the unaffected daughter of an affected mother). The 'Couch model' uses ethnicity, average age of onset, family history of breast and/or ovarian cancer, and the presence of breast and ovarian cancer in a single individual to estimate the probability of finding a mutation in *BRCA1*. This model was developed using a high risk clinic population, physician, or self-referred due to the perception of an increased risk of breast cancer. A similar model using ethnicity, age at first diagnosis, family history of breast and/or ovarian cancer, and patient diagnosis has also been published based on data from a commercial testing laboratory. In addition, a third model uses Bayesian theory in conjunction with family history of breast and ovarian cancer in first and second degree relatives and mutation prevalences to calculate the probability of carrying a *BRCA1* or *BRCA2* mutation. In general, a 5 to 10 per cent chance of finding a mutation is considered justification for testing; however, each decision must be made by the individual considering testing.

Surveillance and prophylactic surgery

The efficacy of cancer surveillance or other measures to reduce cancer related mortality in individuals who carry cancer-predisposing mutations is just beginning to be elucidated. Until prospective trials can be completed, recommendations for surveillance in *BRCA1* and *BRCA2* mutation carriers based on expert opinion have been developed by a Cancer Genetics Studies Consortium task force. This group recommends that monthly self-breast examination begins early in adult life, and clinician examination and mammography should be started between ages 25 and 35 every 12 months. In studies of tumors from *BRCA1* mutation carriers there is less DCIS than in sporadic cancers suggesting that perhaps the tumors grow faster. Therefore, every 6-month mammography may be an option in women with *BRCA1* or *BRCA2* mutations. Some women may be concerned about the risk of radiation exposure from frequent mammograms; however, in women at high risk of developing breast cancer, the benefits will greatly outweigh the risks. Women from high-risk breast cancer families are well aware that mammography and clinical breast examination may not detect early lesions. In the face of a striking family history and close personal losses, these women may be unconvinced that mammography and clinical breast examination offers the protection they seek. Such women often inquire about prophylactic mastectomy in the absence of other preventive options. Recent data from the Mayo Clinic suggest that prophylactic mastectomy is effective in reducing breast cancer risk in women from both high- and moderate-risk families. In 425 women from moderate risk families, 39 breast cancer cases were predicted using the Gail model; only 4 cases were seen ($p<0.001$). In 214 women from high risk families, using sisters as the ontrols, a risk reduction of over 90 per cent was seen. However, there are theoretical considerations about prophylactic mastectomy. Current surgical technique does not allow the complete removal of all breast tissue in a prophylactic total mastectomy, and since a germline mutation will be present in all residual breast tissue, individuals may remain at increased risk of breast cancer after surgery. Nonetheless, the amount of residual tissue is small, and the risk of breast cancer is likely to be markedly reduced. Based on currently available evidence, we present women with the available data and support their individual decision about prophylactic mastectomy. In the future, chemopreventive treatment may be able to reduce cancer risk as much as mastectomy.

Unfortunately, screening for ovarian cancer is less effective than screening for breast cancer. While annual or semi-annual ovarian cancer screening is an option using transvaginal ultrasound and serum CA-125 levels, it is clear that most women are still diagnosed with late stage disease. As ovarian screening has not been clearly shown to be efficacious and given the high mortality of ovarian cancer, we currently recommend prophylactic oophorectomy, as does the American College of Obstetrics and Gynecology, when childbearing is completed or at the latest at the time of menopause, as in *BRCA1* mutation carriers the mean familial age of diagnosis of ovarian cancer (49 ± 7 years) appears to be younger than in sporadic cases (63). After oophorectomy, however, there is a low but measurable incidence of peritoneal malignancies about which women should be counseled. Recent studies have shown that another reason to recommend prophylactic oophorectomy is that it appears to reduce the risk of breast cancer as well as ovarian cancer in *BRCA1* mutation carriers.

Other issues

In the first hint that chemoprevention is becoming a reality, the NSABP (National Surgical Adjuvant Breast and Bowel Project) has conducted a clinical trial of tamoxifen versus placebo in women at increased risk of developing breast cancer. Encouragingly, results of the study showed a 49 per cent reduction in breast cancer incidence among the group treated with tamoxifen with an average follow-up time of 47.7 months. Patients were enrolled at or above age 60 or, using the Gail model, as described above, if they were at least 35 years old and had the breast cancer risk of a 60-year-old woman. Lobular carcinoma *in situ* also was an entry criteria for women aged 35 or older. Tamoxifen did increase the risk of three other medical conditions – endometrial cancer, deep vein thrombosis, and pulmonary embolism, all in postmenopausal women only. Interestingly, two European studies also looking at tamoxifen for chemoprevention of breast cancer found no benefit. The studies had different entry criteria (using the Claus model rather than the Gail model), lower numbers of participants, and allowed the use of concomitant hormone replacement therapy and the Italian study had high levels of non-compliance. Further evaluation needs to be done, but tamoxifen should be considered for prevention of breast cancer in high risk women on an individual basis, taking into account the risks and benefits.

In *BRCA1* and *BRCA2* mutation carriers, the use of hormone replacement therapy (HRT) is controversial. When considering use in women with mutations in *BRCA1* and *BRCA2*, it should be noted that HRT provides approximately one-fourth the amount of estrogen as the premenopausal ovary and therefore can presumably be safely used in oophorectomized women at least until age 50 or 55. In addition, a recent study suggests that estrogen use does not incrementally increase breast cancer risk associated with a family history of breast cancer and decreases overall mortality, suggesting that in aggregate it may be beneficial. Oral contraceptive use in *BRCA1* mutation carriers appears to be associated with a significantly decreased risk of ovarian cancer and a small increased risk of breast cancer, consistent with studies in the general population. As with chemoprevention, each patient should be assessed individually, especially with respect to estrogen replacement therapy, as it

provides benefits in terms of cardiovascular disease and osteoporosis.

Colorectal carcinoma

Overview

More than 160 000 new cases of colorectal cancer (CRC) will be diagnosed in 1999. Mutations in genes accounting for known autosomal dominant cancer susceptibility syndromes will account for at least 5 per cent of those, and as in breast cancer, it is likely that variants in lower penetrance gene will contribute to the development of many more. A large prospective study recently confirmed what more than a dozen retrospective studies had asserted: that a family history of CRC confers increased risk of the disease. This section will review four autosomal dominant cancer susceptibility syndromes whose most predominant feature is an increased risk for colon cancer: familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (HNPCC), familial juvenile polyposis (FJP), and Peutz-Jeghers syndrome (PJS).

Familial adenomatous polyposis (FAP)

The incidence of FAP is estimated to be 1 in 6000 to 1 in 13 000; Powell *et al.* estimated that more than 50 000 families in the United States could benefit from genetic counseling for this disorder. Affected individuals carry germline mutations of the *APC* gene on chromosome 5q21-q22. Approximately one-third of FAP patients have no family history and presumably represent new mutations. Many mutations of APC have been described; 80 per cent of them are truncating. There is some correlation between the position of the truncating mutation and phenotype (discussed below). A variant in *APC* (I1307K) has been described in Ashkenazi Jews which is associated with a two-fold increased relative risk for colorectal cancer. The function of APC is complex and it has been postulated to act in multiple pathways including growth factor regulation, cell migration, and adhesion as well as cell cycle regulation.

Individuals with mutations in *APC* have multiple colonic adenomas and, if untreated, will inevitably develop CRC. Adenomas arise in the mid to late teens; 95 per cent of mutation carriers have adenomas by age 35. More than 90 per cent of FAP patients develop duodenal adenomas, but carcinoma supervenes in only about 5 per cent. Gastric polyposis is seen in at least 50 per cent of affected patients; most of the polyps are fundic gland polyps, but gastric adenomas do occur. While fundic gland polyps are considered to be benign, there are rare reports of dysplasia in fundic gland polyps, and reports of carcinoma arising adjacent to, or in continuity with, fundic gland polyps. Gastric carcinoma risk, though, is not appreciably elevated in Western FAP patients. However, Japanese and Korean families with FAP have a three- to four-fold excess risk for gastric carcinoma.

Extraintestinal manifestations of FAP include desmoid tumor, hepatoblastoma, thyroid carcinoma, medulloblastoma, and various benign lesions: sebaceous or epidermoid cysts, lipomas, osteomas, supernumerary teeth, congenital hypertrophy of retinal pigment epithelium (CHRPE), and juvenile nasopharyngeal angiofibromas.

The attenuated variant of FAP (AFAP) presents a particularly difficult diagnostic challenge. Adenomas can be sparse (one or two in some patients, dozens in others) and are often right-sided. The adenomas appear at a later age (35–40 years) than in classic FAP, as do colon carcinomas (55 years). Upper gastrointestinal manifestations (fundic gland polyps and duodenal adenomas) are seen with the same frequency as in classic FAP. The first AFAP families to be described had mutations near the proximal end of the *APC* gene; the phenotype has since been seen in families with mutations at the extreme distal end as well.

Testing and management

Genetic testing is available for FAP. If a mutation is found in an affected family member, other at-risk family members can easily be tested for the mutations. (Genetic testing is not recommended for children less than 10 years old.) If no mutation is found in an affected family member, the test is considered uninformative; a negative result does not rule out FAP, and all at-risk family members must continue endoscopic screening.

Randomized, controlled trials demonstrating the efficacy of screening and management regimens have not been performed, and are not likely to be. Nevertheless, the following recommendations are generally agreed upon: Affected or at-risk individuals should have annual flexible sigmoidoscopy beginning by age 12. If the family history suggests AFAP, then colonoscopy is required, but can begin later (age 20). Prophylactic colectomy should be considered once adenomas have appeared. Endoscopic surveillance of the rectum and anus should be continued after colectomy. A baseline upper endoscopy is advisable by age 20, with follow-up examinations every 2 to 3 years unless symptoms occur. Annual thyroid palpation is suggested, and children at risk for FAP should have serum a-fetoprotein testing every 6 months until 6 years of age in the interest of early detection of hepatoblastoma.

When duodenal adenomas are discovered, they should be removed endoscopically if feasible. Often, though, the adenomas are too numerous to remove; in that case, annual surveillance with biopsy of grossly suspicious lesions is prudent. It is hoped that chemoprevention will help control duodenal adenomas, but results thus far have not been encouraging.

Desmoid tumors can be a difficult management problem. These locally aggressive soft tissue tumors typically arise in the abdominal wall or bowel mesentery. Relentless recurrences are the rule. The difficulty of surgical cure has led to the recommendation that only symptomatic desmoid tumors should be surgically resected.

Hereditary nonpolyposis colorectal cancer (HNPCC)

In the majority of families (60–70 per cent) meeting clinical criteria for HNPCC (Amsterdam criteria, delineated below), a germline mutation in one of the genes responsible for DNA mismatch repair (MMR) can be demonstrated. The incidence has not been definitely established, but a large-scale molecular screening study from Finland indicates that HNPCC accounts for about 2 per cent of the CRC burden in that country.

Among the several genes participating in DNA mismatch repair, germline mutations in *hMLH1* (on chromosome 3p21.3) and *hMSH2* (on 2p22-p21) account for more than 90 per cent of mutations detected in HNPCC families to date. Mutations in *hMSH6* (on 2p16) have been found in a few families, and rare mutations in *hPMS1* (2q31–33) and *hPMS2* (7p22) have been described. *hMLH3* has recently been described but HNPCC families have not yet been screened for germline mutations. All cells of affected individuals carry a nonfunctioning allele of a DNA MMR gene; if the wild-type allele is lost or inactivated, the cell can no longer repair DNA mismatches that inevitably arise during DNA replication. Cells with defective DNA MMR accumulate mutations at a very high rate (as much as 1000 times that of normal cells). Because DNA mismatches are more likely to occur in DNA microsatellites (areas with multiple repeats of one nucleotide or one pair of nucleotides), defective DNA MMR leads to the phenomenon of microsatellite instability, in which the progeny of the defective cells have varying lengths of a given microsatellite.

The predilection of DNA mismatches for mono- and dinucleotide repeats (such as TTTTT … or CACA …) plays a deciding role in the genetic mutations contributing to carcinogenesis. Nearly all colon cancers with microsatellite instability have mutations in the transforming growth factor (type II receptor (*TGF(IIR)*) and *BAX* genes, and those mutations are located in repeat sequences. Mutation of *TGF(IIR)* leads to escape from the growth inhibitory effects of TGF (while mutation of *BAX* interferes with its proapoptotic effect. Thus, carcinogenesis in colon cancers with microsatellite instability may involve mutations in critical genes different from those involved in colon cancers with other genetic alterations (*APC, K-ras, p53*, etc.)

Individuals carrying germline mutations of *hMLH1* or *hMSH2* have a lifetime risk for CRC of the order of 80 per cent. The cancers tend to arise proximal to the splenic flexure (70 per cent), with an average age of diagnosis at 45 years. Multiple synchronous and metachronous colon cancers are a feature of the syndrome, with 30 per cent of patients developing a second colon cancer within 10 years if a limited operation (right hemicolectomy or segmental resection) is done for the initial cancer. (Even when total abdominal colectomy is performed, the rectum is still at risk; Rodriguez-Bigas *et al.* reported that 12 per cent of HNPCC patients had rectal cancer within 12 years after colectomy.) Colon cancer in HNPCC tends to be poorly differentiated (e.g., non-gland forming) and mucinous. There is frequently an exaggerated host lymphoid response to the tumor, seen as lymphoid aggregates at the tumor edge (Crohn's-like reaction) or increased numbers of tumor-infiltrating lymphocytes. HNPCC patients have improved survival relative to patients with sporadic colon cancer. It is not clear whether the relatively improved survival reflects an inherent property of cancers with microsatellite instability or a salutary effect of host immune response.

Other cancers associated with HNPCC

Extracolonic cancers are common in HNPCC. The most frequent is endometrial carcinoma; women with HNPCC have a 20 to 60 per cent lifetime risk. An increased risk for carcinoma of the stomach, ovary, renal pelvis and ureter, small bowel, hepatobiliary tract, and pancreas also exists. Glioblastoma multiforme is seen in some HNPCC families. (Families with colon cancer and brain tumors described under the eponym ‘Turcot's syndrome' have included FAP families with medulloblastoma and HNPCC families with glioblas-toma.) Muir-Torre SYndrome (benign and malignant sebaceous skin tumors in combination with colon cancer and other internal malignancies) has been shown by mutational analysis to be a variant of HNPCC. While the relative risk for breast cancer is not increased in HNPCC, breast cancers

arising in HNPCC demonstrate microsatellite instability, suggesting that breast cancer is a tumor integral to HNPCC, albeit one with lower penetrance.

Diagnosis and testing

The diagnosis of HNPCC is based on pedigree assessment. Still widely used are the Amsterdam criteria, which are: (1) at least three relatives with CRC, one of whom must be a first-degree relative of the other two; (2) involvement of at least two generations and (3) at least one colon cancer diagnosed before age 50. FAP should be excluded and tumors should be verified by pathologic examination. These criteria helped standardize reporting of HNPCC, but had limitations as a case-finding tool, notably the failure to acknowledge any contribution from extracolonic cancer. The importance of extracolonic cancer was demonstrated by Wijnen *et al.*, who found three independent variables predictive of germline mutations in *hMSH2* or *hMLH1*: fulfilment of the Amsterdam criteria; the presence of endometrial cancer in the kindred; and early age at diagnosis of colon cancer.

Gene testing is available on a clinical basis, but is costly and less sensitive than testing for FAP. Several genes need to be considered (clinical testing is currently limited to *hMLH1* and *hMSH2*); hundreds of different mutations have been described and, while many of them are predicted to be truncating mutations (the easiest to detect), the prevalence of truncating mutations is less than in FAP. Also, it appears that one-third of all pathogenic *hMSH2* mutations, at least in the Dutch population, show deletion of almost all of one *hMSH2* allele. PCR-based techniques do not detect this large genomic deletion.

Management

Based on the early age of CRC onset in HNPCC, and its penchant for the proximal colon, full colonoscopy should be initiated by age 20 to 25 in germline mutation carriers and those at risk based on pedigree analysis. Colonoscopy should be performed at least every 2 years. We prefer every other year in high-risk patients who have not had DNA testing and annually in patients with HNPCC germline mutations. This recommendation is based on the findings of Järvinen *et al.*, who identified six CRCs among HNPCC germline mutation carriers who were undergoing colonoscopies every 3 years, and Vasen *et al.*, who discovered five interval cancers in HNPCC patients within 3½ years following a normal colonoscopy. In reviewing this subject, Church suggests that these interval CRCs develop from normal epithelium within 3 years or from adenomas that were missed. It is important to realize that colonoscopy ‘miss' rates are as high as 29 per cent for polyps less than 5 mm in diameter. Patients should be advised that colonoscopy is not a perfect screening procedure and the option of prophylactic colectomy should be discussed.

Subtotal colectomy as a prophylactic measure among HNPCC patients remains controversial, but patients who carry germline mutations should be offered this option as an alternative to lifetime colonoscopic surveillance. Genetic counseling must be provided so that patients can be in a better position to evaluate the various management strategies. Church suggests that prophylactic surgery should be advised for patients likely to show reduced compliance for colonoscopy.

Syngal *et al.* examined the life expectancy and quality-adjusted life expectancy benefits resulting from endoscopic surveillance and prophylactic colectomy among carriers of germline mutations for HNPCC. Both risk-reduction programs showed large gains in life expectancy for mutation carriers, with benefits of 13.5 years for surveillance and 15.6 years for prophylactic proctocolectomy at 25 years of age, compared with no intervention. The benefits of prophylactic colectomy decreased with increasing age.

Women at risk for HNPCC should have annual screening for endometrial cancer beginning at age 25 to 35 years. Endometrial aspiration or transvaginal ultrasound are advised for screening. Prophylactic hysterectomy can be considered, particularly if childbearing is completed.

Other recommended screening measures include annual urinalysis with cytologic examination, beginning at age 25, annual skin surveillance, and periodic upper endoscopy in families with gastric cancer.

Familial juvenile polyposis coli (FJP)

Sporadic juvenile polyps are relatively common (one autopsy study of patients under 21 years reported a prevalence of 1 per cent). Usually solitary, the polyps are innocuous and do not confer increased risk for cancer of the colon. However, the discovery of multiple juvenile polyps in a patient without a family history of polyposis can create a diagnostic dilemma; various numbers in the range 5–10 have been proposed as an upper limit for sporadic juvenile polyps. The clinical diagnosis of FJP requires histologic confirmation of a juvenile polyp, plus the presence of polyposis (i.e., more than 5 or 10 polyps) or a family history of polyposis. Even when there are multiple juvenile polyps and a family history of polyposis, the diagnosis of juvenile polyposis must be made with care, because many other genetic syndromes can include juvenile polyps. Other genetic SYndromes whose manifestations can include juvenile polyps are Cowden SYndrome, Bannayan-Riley-Rulvacaba, Peutz-Jeghers (all reviewed below), nevoid basal cell carcinoma (Gorlin's) syndrome, and hereditary hemorrhagic telangectesia (HHT). As the differential diagnosis contains syndromes with differing implications, it is important that an experienced practitioner examine patients with juvenile polyps.

Two variants of FJP have been described: a rare, usually fatal juvenile polyposis of infancy with diarrhea, protein-losing enteropathy, alopecia, and the more common form, in which juvenile polyps appear during childhood. The inheritance pattern is autosomal dominant. Extraintestinal anomalies have been reported in association with juvenile polyps; these include hydrocephalus, thyroglossal duct cyst, tetralogy of Fallot, coarctation of the aorta, idiopathic hypertrophic subaortic stenosis, and malrotation of the gut.

The etiology of the common form of FJP is genetically heterogeneous. One of the genes for FJP encodes a component of the TGF (signaling pathway, *SMAD4* on 18q21. As much of FJP still remains unexplained, the other members of the *SMAD* gene family have been screened for mutations in families with juvenile polyposis. However, germline mutations in *SMAD2, SMAD3, SMAD5*, and *SMAD1* were not found, presumably because of their overlapping functions. Therefore, the genes underlying much of FJP remain to be identified.

Juvenile polyps may be found in the colorectum (98 per cent), stomach (14 per cent), duodenum (2 per cent), and jejunum and ileum (7 per cent). While the polyps are considered benign, FJP patients are at increased risk for colorectal cancer. The cumulative lifetime risk was estimated by Järvinen to be 50 per cent, an estimate that compares remarkably well to a report from the University of Iowa describing a large FJP kindred in which 16 of 29 (55 per cent) affected individuals developed gastrointestinal cancer. Eleven members of the Iowa kindred had colon cancer and six had gastric cancer.

Affected individuals need regular endoscopic surveillance. Scott-Conner and colleagues recommend upper and lower gastrointestinal endoscopy every 3 years as long as no lesions are detected. Small numbers of polyps can be managed by polypectomy; in these cases, endoscopy should be repeated in 1 year. If there are multiple polyps in the colon, subtotal colectomy is recommended. Multiple gastric polyps, particularly if dysplasia is present, should prompt consideration of gastrectomy.

Children with large numbers of polyps present a difficult management problem. Colectomy is the prudent choice, but we believe regular colonoscopy with removal of the largest polyps until the patient attains puberty is also an option.

Peutz-Jeghers syndrome (PJS)

The gene responsible for PJS is on chromosome 19p13.3 and encodes a nuclear serine threonine kinase, *STK11* (*LKB1*). Abnormal (i.e. inactive) forms of this kinase may lead to defective control of cellular growth and differentiation. However, not all PJS families can be linked to 19p13.3, leading to speculation that there is a second locus for the syndrome. PJS is rare and thought to be about one-tenth as common as FAP.

The diagnosis of PJS requires histologic confirmation of a hamartomatous, Peutz-Jeghers-type polyp. The polyps in PJS can in some cases be differentiated from those of the other hamartomatous syndromes as they contain smooth muscle bundles. As similar polyps can be seen in individuals who do not have PJS, clinical diagnosis also requires at least two of the following: small bowel polyposis; family history of PJS; or pigmented macules of buccal mucosa, lips, fingers, and toes. Genetic testing for the SYndrome is being developed.

Peutz-Jeghers polyps have been found throughout the gastrointestinal tract. The small bowel is the site affected most often, but stomach and colon can be involved; esophageal polyps are rare. The polyps are almost always multiple, but tend to number in dozens rather than hundreds. Peutz-Jeghers polyps have also been described in respiratory mucosa and the urinary tract. One member of the Dutch family originally studied by Peutz suffered from severe nasal polyposis and eventually developed a nasal carcinoma.

PJS carries increased risk for malignancy including colon, stomach, small intestine, ovary, fallopian tube, thyroid, and lung cancers. There appears to be a particularly high incidence of breast and gynecological tumors. Nearly every female PJS patient will have ovarian involvement by sex cord tumor with annular tubules (SCTAT). The tumors are bilateral in at least two-thirds of cases (in contrast to SCTAT in the sporadic setting, which is almost always unilateral). An unusual form of cervical cancer, minimal deviation adenocarcinoma (adenoma malignum), is also characteristic of PJS. This rare tumor accounts for 1 to 3 per cent of all cervical

adenocarcinomas but, in one series, affected 4 of 27 women with PJS.

The surveillance protocol advocated by the St Mark's Polyposis Registry includes yearly hemoglobin and yearly ultrasound of pelvis in females and pancreas in all patients. Testicular ultrasound should also be done in males with feminizing features. Biannual upper and lower endoscopy with small bowel radiograph are recommended. In addition, regular mammography and cervical smear are critical surveillance measures. Tomlinson and Houlston suggest that upper endoscopy, colonoscopy, and small-bowel radiograph begin in the second decade, and mammography at age 25. The gastrointestinal polyps may be associated with bleeding, obstruction, or intussusception. Conservative removal (snare polypectomy) is favored over segmental resection of bowel, to avoid development of a short bowel SYndrome.

Li-Fraumeni syndrome (LFS)

Li-Fraumeni syndrome (LFS), now known to be associated with germline mutations in *p53* and *hCHK2*, was first identified as a syndrome in 1969 with a description of four kindreds in which cousins or siblings had childhood soft-tissue sarcomas and other relatives had excessive cancer occurrence. The SYndrome is rare, accounting for less than 1 per cent of childhood cancers. Mutations clustered in the conserved sequences of the *p53* gene, exons 5 through 9, are found in about 70 per cent of classical LFS families, but *p53* has been ruled out in some classic families, suggesting that this syndrome is genetically heterogeneous. Recently, mutations in *hCHK2* were identified in Li-Fraumini and Li-Fraumini-like families, providing evidence of a second gene responsible for this SYndrome. Genetic testing for *p53* mutations is available.

Li and Fraumeni characterized the tumor spectrum in LFS after noting an increased incidence of relatives with cancer in childhood sarcoma patients. The acronym SBLA (sarcoma, breast, brain tumors, leukemia, lymphoma, laryngeal carcinoma, lung cancer, adrenal cortical carcinoma) also has been used to describe the syndrome. In addition, melanoma, germ cell tumor, pancreatic, gastric, and prostatic carcinomas have also been described in LFS. The clinical diagnosis in a family requires one member with sarcoma under age 45, a first-degree relative with any type of cancer under age 45, and a third affected family member with sarcoma (any age) or other cancer (less than 45 years old). and 90 per cent by age 70. Hisada *et al.* studied the incidence of second and third primary cancers in members of 24 LFS kindreds. The cumulative probability of a second primary was 57 per cent at 30 years after the first cancer diagnosis. Sarcomas and carcinoma of the breast accounted for 46 of the 72 cancers identified. Unfortunately, almost 30 per cent of the tumors in reported families occur before age 15 years.

Annual physical examinations with blood cell counts are advised for LFS patients. Careful attention should be given to sites known to be at risk in LFS. Annual mammography and frequent breast self-examination are particularly important as among 24 LFS families studied in one series, 44 women were diagnosed with breast cancer, of whom 77 per cent were between age 22 and 45 years.

Multiple endocrine neoplasia

Multiple endocrine neoplasia type 1

Multiple endocrine neoplasia type 1 (MEN1) is a rare autosomal dominant disorder associated with several endocrine tumors that cause as much harm by the hormones they oversecrete as by their malignant potential. The tumors most commonly seen include parathyroid tumors, gastrointestinal endocrine tumors, and tumors in the anterior pituitary, with two out of the three in first-degree relatives sufficient to make the diagnosis. Also seen in this syndrome are carcinoids, lipomas, and angiofibromas. Werner first recognized MEN1 as a familial disorder. The gene responsible for this syndrome, *MEN1*, has recently been identified and isolated. The protein product, menin, has been identified, although its function is still unknown.

MEN1 tends to present as primary hyperparathyroidism, and less commonly as Zollinger-Ellison syndrome (ZES), insulinoma, or a pituitary tumor. The majority of those affected will develop hyperparathyroidism; 50 per cent are symptomatic by the age of 25. Patients have hyperplasia of the parathyroid glands and can present with multiple parathyroid adenomas. The majority of patients with ZES or insulinomas will have experienced symptoms of hyperparathyroidism by the time they are diagnosed. Management of these tumors can be difficult due to the nature of the hormones they secrete. In addition, these tumors are often small, multiple, and difficult to remove surgically.

The *MEN1* gene was discovered by positional cloning following its localization on chromosome 11q13. The majority of mutations documented at this time lead to early termination of the protein product suggesting it is a classic tumor suppressor gene. It is now expected that most if not all families with MEN1 will carry mutations in this gene.

Clinical DNA testing exists, and can save an individual from a MEN1 family from periodic biochemical testing of serum calcium, parathyroid hormone, and prolactin, which otherwise is begun as an adolescent. However, genetic testing of individuals at this young age is controversial as it removes their freedom to decide if they desire testing at a later date. In addition, biochemical testing is readily available and often at a lower cost, making the role of clinical DNA testing for *MEN1* uncertain at this time. As we learn more about the *MEN1* gene, and can develop effective treatment strategies based on the knowledge of carrier status, DNA testing may take on a larger role.

Multiple endocrine neoplasia type 2

In the most common type of multiple endocrine neoplasia type 2 (MEN2) families (MEN2A) are predisposed to develop the triad of medullary thyroid carcinoma (MTC), pheochromocytoma (PTC), and parathyroid hyperplasia with hyperparathyroidism (HPT). MTC is the hallmark of all three types of MEN syndromes – MEN2A, MEN2B, and familial medullary thyroid carcinoma (FMTC), all of which are due to mutations in the *RET* proto-oncogene. MEN2A, which encompasses the majority of the cases, usually presents with MTC, which has been documented as early as age 6. Individuals with MEN2A have a 100 per cent lifetime risk of MTC, 50 per cent risk of pheochromocytoma, and a 10 to 15 per cent risk of hyperparathyroidism. MEN2B presents even earlier with MTC and HPT (by approximately 10 years) than MEN2A. MEN2B patients have characteristic developmental abnormalities involving hyperplasia of the autonomic nerves of the intestines leading to megacolon, and characteristic dysmorphic facies due to mucosal neuromas, ganglioneuromatosis, and a marfanoid body habitus. Approximately one-half of patients seen with MEN2B are due to new mutations, presumably due to the earlier onset of MTC and developmental abnormalities in this group. FMTC, the third manifestation of mutations in *RET*, consists of families in which only MTC is present. The MTC in FMTC is often later in onset and has a better prognosis than in other MEN syndromes.

RET is located on chromosome 10q11.2 and has 21 exons. *RET* encodes a receptor tyrosine kinase. Unlike all other known familial cancer syndrome genes, *RET* is a proto-oncogene. Mutations that increase RET activity lead to hyperplasia and the subsequent development of cancer. As elucidated above, different *RET* mutations can generate distinct phenotypic presentations, making it one of the few cancer predisposing genes in which a phenotype-genotype correlation has been documented. Families which exhibit both MEN2A and Hirshprung's disease (HD, associated with congenital absence of sympathetic neurons, distal colon, and rectum) have been shown to be associated with a single mutation in *RET*. Interestingly, families with HD in the absence of the MEN2 syndrome carry different mutations in the *RET* gene that cause loss, rather than gain, of RET function.

Prior to clinical DNA testing, biochemical screening and imaging of individuals at risk for this syndrome began at the age of 4 or 5 and continued into early adulthood. Routine screening consisted of plasma calcitonin (following pentagastrin stimulation), plasma calcium, urine or plasma catecholamines, regular blood pressure measurements, and imaging of the adrenals. Prophylactic thyroidectomy for elevated markers such as calcitonin, was performed prior to the age of 6, because the morbidity was low and the results were curative in those individuals that showed the earliest signs of the syndrome. Screening for the presence of PTC is essential prior to prophylactic surgery or surgery for all MTC. When data from families from several countries were pooled, 97 to 100 per cent of MEN2A, 95 per cent of MEN2B, and 88 per cent of FMTC families had documented mutations in *RET*. In addition, DNA screening for the few common mutations known to be associated with the above phenotypes can exclude germline *RET* mutations in more than 99 per cent of those at risk and patients with sporadic MTC. In a family with a known mutation, DNA testing is 100 per cent accurate, and can save a non-affected individual from unnecessary screening. In MEN2A, prophylactic thyroidectomy for asymptomatic patients on the basis of positive DNA testing alone is suggested between the ages of 5 and 10 years. For FMTC families prophylactic thyroidectomy may be delayed slightly. Less consensus exists for the screening and treatment of MEN2B; however, given the early onset of MTC, thyroidectomy in infancy is generally recommended. The adjacent parathyroid glands may be saved and autotransplanted into the forearm at the time of thyroidectomy. Postoperatively biochemical screening should be continued for residual MTC and PCC. Other preventive options are not agreed upon at this time. It is hoped that with increased information and refinement of our knowledge of RET genotype-phenotype correlations, additional therapies will be possible.

Neurofibromatosis

Neurofibromatosis type 1 (NF1)

Neurofibromatosis type 1 (NF1) is inherited in an autosomal dominant fashion. The disease incidence is one in 3000, with one-third of cases representing new mutations. The implicated gene (*NF1*) is a large gene (59 exons) located on chromosome 17q11.2. Many different mutations of *NF1* have been reported; most (70–80

per cent) are truncating mutations. A protein truncation test is available for clinical genetic testing; however this only picks up approximately 60 per cent of mutations.

Clinical diagnosis of NF1 requires two or more of the following criteria: (1) more than six café-au-lait spots; (2) two neurofibromas or one plexiform neurofibroma; (3) multiple axillary or inguinal freckles; (4) sphenoid wing dysplasia or congenital bowing or thinning of long bone cortex; (5) an optic pathway tumor; (6) more than two Lisch nodules; (7) a first-degree relative with NF1 by these criteria. In addition to the lesions listed above, mutation carriers are at risk for neurofibrosarcoma (3 to 15 per cent of affected individuals), pheochromocytoma, duodenal carcinoid, neuroblastoma, ependymoma, rhabdomyosarcoma, and Wilms' tumor. Children with NF1 are at increased risk for juvenile myelomonocytic leukemia (juvenile chronic myelogenous leukemia), and the monosomy 7 syndrome, a childhood myelodysplasia.

The *NF1* gene encodes a GTPase-activating protein known as neurofibromin. The gene product negatively regulates signals transduced by Ras proteins by catalyzing the conversion of the active Ras.GTP to inactive Ras.GDP. Tumors from NF patients show biochemical evidence of hyperactive Ras. Ras sits in the plasma membrane and acts as a growth signal transmitter, providing an obvious advantage to tumors when it is consitutively activated.

There are few screening recommendations for patients with NF1 and it has been suggested that symptomatic clinical follow-up is the most reasonable course. Blood pressure should be monitored twice a year (pheochromocytoma). Neurologic symptoms (headaches, hearing loss, visual change) should be sought and investigated.

Neurofibromatosis type 2 (NF2)

Neurofibromatosis type 2 (NF2) is also inherited in an autosomal dominant fashion. Disease incidence is one in 35 000, with about one-half of cases representing new mutations. The *NF2* gene is on chromosome 22q12.2. Genetic testing is available clinically.

The consensus criteria for NF2 require either (1) bilateral 8th nerve masses or (2) a first-degree relative with NF2 in a patient with one of the following: unilateral vestibular schwannoma under age 30 or any two of the following: neurofibromas, gliomas, meningiomas, or juvenile posterior subcapsular cataracts. Individuals with unilateral vestibular schwannoma and one of the component tumors or cataracts and those with multiple meningiomas and one of the other component tumors also should be evaluated for NF2.

Patients are at risk for central nervous system (CNS) tumors, the most common of which are vestibular schwannomas, schwannomas at other sites, meningiomas, and ependymomas. While these tumors are often benign and slow growing, their location within the CNS may lead to intracranial and intraspinal involvement with a high rate of morbidity and mortality. Affected individuals also typically develop hearing loss (often bilateral), imbalance, tinnitus, facial weakness and headache, and posterior capsular lens opacities. The average age at onset is in the mid-20s.

NF2 patients may be clinically subdivided into a severe type and a mild subtype, a classification which is based on the age at onset of symptoms, the number and type of tumors, and the duration of disease. The severe type is termed the Wishart form, and SYmptoms are usually manifest prior to 25 years of age. These patients rarely survive past 50 years of age. The Gardner (or mild) subtype, usually presents with symptoms later in life (after 25 years of age), develops a lesser number of more slow-growing tumors, has only bilateral vestibular schwannomas, and generally survives beyond the fifth decade. In general, the severe form of NF2 is due to truncating mutations in the *NF2* gene, while milder disease is due to mis-sense mutations; however, caution must be used when interpreting results as interfamilial variability has been observed.

Annual neurologic examination, to include audiologic and ophthalmologic tests, is advised. All patients with a new diagnosis of NF2 should undergo full spinal MR imaging with and without gadolinium so that maximum prognostic information can be gained. If tumors are found, follow-up imaging is warranted every 6 to 12 months.

Retinoblastoma

Retinoblastoma is the most common intraocular cancer in children, with an incidence of 1 in 20 000 live births, affecting males and females equally. The tumors may be unilateral (of which 15–20 per cent are hereditary) or bilateral (nearly all of which are hereditary). Bilateral tumors show an average age at diagnosis of 12 months, while the average onset of unilateral tumors is 18 months. Most (90 per cent) retinoblastomas are diagnosed before 3 years of age.

Hereditary retinoblastoma is transmitted in an autosomal dominant fashion. The gene (*RB1*) is on chromosome 13q14. Most mutations are highly penetrant (90 per cent); however, some lower penetrant mutations have been reported. In particular some splice-site and mis-sense mutations may be associated with a lower penetrance. The rate of new mutations may be as high as 80 per cent. The gene product functions in cell cycle regulation (G1 to S phase). Retinoblastoma served as the model for Knudsen's two-hit hypothesis. Knudsen reasoned that sporadic retinoblastoma required two somatic mutations, while hereditary retinoblastoma patients had mutation of one allele at the germline level, and needed only one somatic mutation. This theory accounted for the earlier onset and excess bilaterality in the hereditary form of retinoblastoma and forms the basis of our current understanding of hereditary susceptibility to cancer.

Second malignant tumors (non-ocular) are seen in hereditary retinoblastoma. The incidence has been reported as 4.4 per cent at 10 years, 18.3 per cent at 20 years, and 26.1 per cent at 30 years. Radiation therapy for retinoblastoma contributes to this risk, but does not account for all of it. Osteosarcoma and melanoma are the most common second tumors, but brain tumors, soft tissue sarcomas, leukemias, lymphomas, and pinealoblastoma have been reported. Patients with retinoblastoma need lifelong follow-up.

Genetic testing is available and is cost effective. Noorani *et al.* compared molecular and conventional screening of retinoblastoma relatives. The authors demonstrated, '… a significant saving of health care dollars by the molecular route, indicating the benefit of redirecting economic resources to molecular diagnosis in retinoblastoma'. The advantage of molecular screening includes early identification of mutation carriers, which reduces morbidity and mortality in this high-risk population.

von Hippel-Lindau (VHL) disease

VHL disease is transmitted in an autosomal dominant fashion. The disease incidence is approximately one in 36 000. The *VHL* gene is on chromosome 3p25-p26. The gene encodes a protein that binds to elongins (transcription elongation factors) which may regulate hypoxia inducible mRNAs. Hundreds of different mutations have been described, including mis-sense, nonsense, deletion mutations, and genomic rearrangements. Genetic testing (mutation analysis and linkage analysis) is clinically available.

VHL disease is predisposes to many types of tumors, including bilateral multifocal clear-cell renal cell carcinoma (lifetime risk > 70 per cent), retinal and cerebellar hemangioblastoma, endolymphatic sac tumors, and islet-cell tumors. In addition patients get multiple cysts in the kidney, liver, and pancreas. Angiomas and cysts of the spleen are occasionally described. Retinal angiomas may also occur, usually in young adulthood, and may produce visual loss. There are two general types of VHL – type I in which both renal cell carcinoma and pheochromocytoma occur or type II in which renal cell carcinoma can occur alone. Type I patients usually have mis-sense mutations in *VHL* while type II patients have deletions and protein truncation mutations. As pheochromocytoma is an integral lesion in other hereditary syndromes (MEN 2 and NF-1) besides VHL, genetic testing for both *VHL* and *RET* mutations has been advocated for patients with familial, multiple, or early-onset pheochromocytoma. Renal cell carcinoma occurs in up to 75 per cent of affected individuals, with an average age at diagnosis of 40 to 45 years. Pancreatic neoplasms (cystadenocarcinoma, islet cell tumor) cluster in certain families.

Affected individuals are advised to have annual neurologic and ophthalmologic examination, annual red blood cell count (renal cysts and cerebellar hemangioblastoma can produce erythropoietin), regular imaging of central nervous system, kidneys, and pancreas, and annual chemical screening for pheochromocytoma.

Cowden and Bannayan-Riley-Ruvalcaba syndromes

Cowden syndrome (CS) is a rare autosomal dominant syndrome in which patients have a predisposition to both benign and malignant neoplasms. Twenty to thirty per cent of affected women develop breast cancer. Other tumors seen in patients with CS include adenomas and follicular cell carcinomas of the thyroid gland, polyps and adenocarcinomas of the gastrointestinal tract, and ovarian cysts and carcinoma. The characteristic dermatologic features include smooth facial papules (trichilemmomas and verrucae), acral and palmarplantar keratoses, and oral papillomas. The mucocutaneous lesions are pathognomic for CS and are seen in over 90 per cent of patients. Some families also have Lhermitte-Duclos syndrome, a global hypertrophy, or dysplastic ganglioma of the cerebellum.

CS is caused by germline mutations in *PTEN* (*MMAC1/TEP1*). PTEN is a tumor suppressor gene on 10q23.3, which acts as a dual specificity phosphatase and *in vivo* as a phosphoinositide 3-phosphatase, which phosphorylates the lipid second messenger phosphatidylinositol 3,4,5-triphosphate (PIP_3). It is localized to the cytoplasm and appears to be down regulated by transforming growth factor b. The N-terminus of PTEN has homology to chicken tensin and bovine auxillin (cytoskeletal proteins); however, the function of this domain in PTEN is unknown. Germline mutations in *PTEN* have been found in 30 of 37 families with CS in one series, sparing exons 1, 4, and 9. However, some families with CS do not appear to be linked to chromosome 10 suggesting that CS may be genetically heterogeneous. This apparent

heterogeneity may be due to differing criteria delineating the syndrome.

Germline mutations in *PTEN* have also been found in the Bannayan-Riley-Ruvalcaba (BRR) syndrome, which has some features that overlap with CS. Features of BRR not commonly found in CS kindreds include macrocephaly, lipomas, mental retardation, vascular malformations, and pigmented lesions of the glans penis. Recent work by Eng and colleagues has shown a similar spectrum of mutations in CS, BRR, and CS/BRR overlap kindreds, leading to speculation that factors, in addition to mutation status, affect the phenotypic presentation of disease. Mutations found in CS and BBR patients and in tumor specimens have been shown to ablate phosphatase activity *in vitro*, demonstrating their functional significance.

Patients and families with a history suggestive of CS or BRR should be screened for mutations in *PTEN*. Screening recommendations for this group of patients are still being defined but currently include the following suggestions. A meticulous annual physical examination (starting from birth in obvious BRR and CS-BRR overlap families, in the late teens in pure CS families) paying particular attention to the thyroid, breasts, and skin, including urine for microscopic hematuria. Annual mammograpy should start at age 30 or 5 years younger than youngest age of breast cancer diagnosis in the family. A baseline thyroid ultrasound in the late teenage years and annual endometrial examinations using repel biopsies premenopausally and ultrasound postmenopausally beginning at age 35 to 40 or 5 years younger than youngest age of endometrial cancer diagnosis in the family are suggested. In addition, some families also appear to have family-specific cancers such as colorectal cancer or renal cell carcinoma and appropriate screening measures should be instituted in those cases.

Genetics of other common cancers

Malignant melanoma

Inherited risk factors are thought to account for 5 to 7 per cent of melanoma, but gene and mutation frequencies are unknown. Three syndromes whose primary cancer is melanoma have been described—dysplastic nevus syndrome, familial atypical multiple mole-malignant melanoma syndrome (FAMMM), and melanoma-astrocytoma syndrome. Hereditary predisposition to melanoma should be suspected in a patient with invasive melanoma in at least two first-degree relatives. Early age at diagnosis and multiple primary melanomas are typical of melanoma-prone families. While not all melanoma kindreds have dysplastic nevi, 10 to 100 nevi on the upper trunk and limbs, with variability in mole size, shape, and color, are characteristic. Histologically, such nevi show the architectural and cytologic atypia of a dysplastic nevus.

Three putative melanoma-susceptibility loci have been described. The first was CMM1, which was mapped to chromosome 1p36. No candidate gene at this locus has yet been identified. A second susceptibility locus was mapped to 9p21, the location of *CDKN2* (*p16*). Various mutations of *CDKN2* have been described. In Dutch melanoma kindreds, a 19-bp deletion has been implicated. Whelan *et al.* linked melanoma/pancreatic cancer predisposition to a mis-sense mutation (Gly93Trp). Finally, germline mutations in the *CDK4* gene on chromosome 12q14 have been documented in rare melanoma-prone families. CDKs or cyclin-dependant kinases function as regulators of cell cycle progression.

The FAMMM syndrome is associated with pancreatic cancer. Bergman *et al.* found that the risk for pancreatic cancer in these families is elevated by a factor of 13.4. Astrocytomas occur also in excess in some melanoma families.

Genetic testing for mutations in *CDKN2* is available clinically, but its use is limited by debate over how to use the results. Individuals in melanoma-prone kindreds should perform monthly skin self-examination. Comprehensive dermatologic evaluation should be done semi-annually. Any suspicious lesions should be excised. Sunburn should be avoided, and use of ultraviolet A/B-blocking sunscreens is encouraged.

Pancreatic cancer

Until recently the existence of familial pancreatic cancer was confined to anecdotal reports, but large systematic studies have now confirmed these initial reports. Ghadirian *et al.* performed a population-based case-control study in Quebec, Canada, and found that 7.8 per cent of pancreatic cancer patients reported a positive family history of pancreatic cancer, compared to 0.6 per cent of controls. Falk *et al.* studied pancreatic cancer in Louisiana, and found increased risk for pancreatic cancer among persons reporting any cancer in a close relative (odds ratio [OR] = 1.86; 95 per cent CI = 1.42–2.44). The highest risk was seen in those with a history of pancreatic cancer in a close relative (OR = 5.25; 95 per cent CI = 2.08–13.21). The National Familial Pancreas Tumor Registry at Johns Hopkins reported 151 families in which at least two first-degree relatives have pancreatic cancer. In these families even second-degree relatives of affected patients are at increased risk for pancreatic cancer.

Familial aggregation of pancreatic cancer can be due to a variety of known autosomal dominant cancer susceptility syndromes, including HNPCC, Peutz-Jeghers syndrome, FAMMM, and familial breast-ovarian cancer due to mutations in *BRCA2*. Families with hereditary pancreatitis are also at risk for pancreatic cancer. Briefly, hereditary pancreatitis is characterized by recurrent episodes of severe chronic pancreatitis. Affected individuals are at risk for pancreatic pseudocysts, pancreatic exocrine failure, diabetes mellitus, and pancreatic cancer. A gene for hereditary pancreatitis has been mapped to chromosome 7q35. The gene encodes a cationic trypsinogen which, when mutated, fails to inactivate trypsin, resulting in autodigestion of the pancreas. Presumably, chronic epithelial injury and regeneration increase the risk for pancreatic cancer.

Even when known hereditary syndromes are excluded, familial aggregations of pancreatic cancer remain. The inheritance pattern in these families appears to be autosomal dominant. No gene(s) accounting for these familial aggregations have been mapped or identified.

Surveillance of high-risk patients is evolving. Brentnall *et al.* used endoscopic ultrasound, endoscopic retrograde cholangiopancreatography (ERCP), spiral computed tomography, and serum carcinoembryonic antigen and CA19–9 analysis to follow 14 patients from three pancreatic cancer kindreds. Seven patients underwent pancreatectomy on the basis of abnormalities on ultrasonography and ERCP. All seven had histologic dysplasia in the resected specimen; there were no occult carcinomas.

Gastric carcinoma

Gastric cancer clusters have been described anecdotally in many instances, including in Napoleon Bonaparte's family; he, his father and grandfather, four sisters, and a brother all died of gastric cancer at a young age. Case-control studies have shown that first-degree relatives of gastric cancer patients are about three times more likely than the general population to develop gastric cancer. An over-representation of blood type A in gastric cancer patients, and positive concordance in twin studies add to the impression that heredity contributes to gastric cancer risk. As is so often the case, young age at diagnosis is hallmark of hereditary susceptibility to gastric cancer: Mecklin *et al.* found a highly significant ($p < 0.001$) excess of gastric cancer among the parents of patients who developed gastric cancer at less than 40 years of age.

Gastric carcinoma can be divided into two broad categories based on pathologic features. Intestinal-type carcinoma typically forms a well-demarcated mass of gland-forming cells. Chronic atrophic gastritis and intestinal metaplasia are risk factors for intestinal-type gastric cancer. Diffuse-type gastric cancer often produces an ill-defined thickening of the gastric wall, with signet cells and rudimentary gland forms dominating the histologic picture.

One gastric cancer specific susceptibility gene has been described. Guilford *et al.* found germline mutations of *CDH1* on chromosome 16q22.1 in three Maori families with diffuse-type gastric cancer, showing an autosomal dominant pattern with incomplete penetrance. *CDH1* encodes E-cadherin, a cell adhesion protein that may be important in regulating cellular invasion. Downregulation of E-cadherin in sporadic tumors is associated with poor prognosis. Germline mutations of the same gene were later described in eight additional gastric cancer families, including those of European origin. At least one of the families also had early-onset breast cancer. However, *CDH1* was excluded in a large kindred with autosomal dominant familial gastric polyposis/carcinoma, suggesting that gastric cancer may be genetically heterogenous.

This subject has been recently updated by Bevan *et al.* In families with mainly diffuse gastric cancer, given the submucosal involvement, endoscopic ultrasound of the stomach is the screening procedure of choice and should be initiated 5 years earlier than the youngest case in the family.

Gastric cancer also occurs in excess in HNPCC, FAP, PJ, and CS. There is geographic variation in HNPCC and FAP (frequency of gastric cancer being higher in Korean kindreds), and gastric cancer has been declining in Western HNPCC kindreds in parallel with the declining incidence of gastric cancer in Western populations. Interestingly, while the hereditary gastric cancers associated with *E-cadherin* mutations are usually of the diffuse type, gastric cancer in HNPCC is intestinal type.

Patients from families with intestinal-type gastric cancer clustering are advised to have annual upper endoscopy, beginning at an age 5 years younger than the youngest affected patient in the family. Prophylactic gastrectomy in mutation carriers may eventually be an option, once the estimates of gene penetrance and lifetime cancer risk are more thoroughly elucidated.

Prostate cancer

Prostate cancer will affect an estimated 1 in 5 American men over their lifetime. During the period 1973 to 1995, age-adjusted incidence rates were 158/100 000 for Caucasian men and 241/100 000 for African–American men. Family history has been established as a clear risk factor for prostate cancer. Multiple studies have reported that first-degree relatives of men affected with prostate cancer are two to three times more likely to develop prostate cancer compared with men in the general population. The relative risk of prostate cancer is highest among families who develop prostate cancer at an early age, consistent with other cancer susceptibility syndromes. It has been estimated that male relatives of men diagnosed with prostate cancer under age 53 have a lifetime cumulative risk of 40 per cent for developing prostate cancer. A population based case-control study by Whittemore and colleagues of over 1500 cases and 1600 controls in which Caucasians, African–Americans, and Asians were studied reported an odds ratio of 2.5 for men with an affected first-degree relative after adjusting for age and ethnicity. For men with a brother and father or son affected with prostate cancer, the relative risk was estimated to be 6.4. Interestingly, the risk did not significantly change between ethnic groups despite the large differences in incidence suggesting that genetic factors are important in determining risk.

It has been estimated that 5 to 10 per cent of all prostate cancer is due to a rare autosomal dominant prostate cancer susceptibility gene. Several segregation studies have provided supporting evidence that such a gene or genes exist, although there has been some variation in the estimations of prevalence and penetrance for the suggested prostate cancer susceptibility gene(s) in the population, which may depend on the population of ascertainment. For example, estimates from the Mayo Clinic suggest that the frequency of the gene is 0.006 and the penetrance is 89 per cent at age 85, while estimates in the Swedish population suggest that the frequency is 0.0167 and lifetime penetrance 63 per cent.

Linkage studies in high-risk prostate cancer families have provided suggestive evidence of prostate cancer susceptibility genes on chromosomes 1 and X. For HPC1, there has been conflicting evidence of linkage with some groups confirming linkage to HPC1 and others finding no evidence of linkage. Even in the initial study by Smith *et al.*, it was clear that there is locus heterogeneity with linkage seen in only one-third of the families studied. Another study has found linkage of early onset prostate cancer to chromosome 1q42.2 by Berthon and colleagues. A third locus on chromosome 1 has been postulated for brain–prostate families by Jarvik and colleagues. A locus on the X chromosome, HPCX, has been recently identified with a maximum two-point lod score of 4.6, which by heterogeneity estimates accounts for approximately 16 per cent of cases. As yet, no candidate gene has been isolated from any of the hereditary prostate cancer loci.

Studies from *BRCA1* or *BRCA2* mutation carrier families have shown an increased risk for prostate cancer of three- to fourfold in carriers. However, in series of Ashkenazi Jewish prostate cancer patients, the Ashkenazi Jewish founder mutations in *BRCA1* and *BRCA2* were not found in excess over the population. This apparent paradox has yet to be resolved.

Risks and benefits of surveillance are unproven, but regular screening by evaluating serum prostate specific antigen and regular digital rectal examination appear reasonable. In high-risk families, screening should begin before age 50. The utility of transrectal ultrasound and random prostate biopsies has not been evaluated.

Lung cancer

Lung cancer, more than any other solid tumor, is readily identified with an environmental cause, namely tobacco exposure. While no familial lung cancer syndrome has been identified, lung cancer presents a model for how the genetics of toxin metabolism may lead to cancer predisposition. Some non-tobacco related familial clusters have also been associated with an earlier age of diagnosis, but very little has been discovered about the genetic loci that might explain these cases.

Lynch *et al.* evaluated cancer risk in the relatives of 254 consecutively ascertained probands with histologically verified lung cancer, and in relatives of 231 probands with other smoking-associated cancers. There was no strong evidence for increased risk of lung cancer in relatives, but there was a significant increased risk for cancers of all anatomic sites among the relatives of lung cancer probands ($p<0.001$). Conversely, there were no significant excesses of cancer at all anatomic sites in relatives of probands with other smoking-associated carcinomas. The observed increased risk for cancer at all anatomic sites in the relatives of lung cancer probands could result from an underlying hereditary susceptibility to cancer in these families. Studies looking at relatives of non-smoking lung cancer probands by Swanson and colleagues suggest a polygenic model to account for a moderate familial aggregation of cancers in the families of probands with non-smoking lung cancers.

While millions of people smoke each year, not all will develop lung cancer. Tobacco-associated lung cancer has been observed to cluster in some families. This has lead to a great deal of research, both in humans and animals, on genetic determinants of toxin metabolism. The development of lung cancer appears to involve multiple molecular changes. Many lung tumors are found to have somatic mutations in *p53* and other susceptibility genes involved in DNA damage response pathways, although they have not been shown to play a role in hereditary lung cancer. Mouse models of toxin related lung cancers have lead to the identification of a susceptibility gene in that system. This gene, known as *Pas-1*, has been mapped to the mouse chromosome 6. Identification of homologues to *Pas-1* in humans may eventually lead to breakthroughs in our understanding of the pathogenesis of lung cancer and its genetic determinants.

Renal cell carcinoma (RCC)

Familial renal cell carcinoma occurs as two distinct histologic entities: familial papillary renal carcinoma and familial nonpapillary (clear cell) carcinoma. Both have an autosomal dominant pattern of inheritance. Germline mutations in *c-Met*, a proto-oncogene, have been documented in papillary RCC families. The gene for familial clear cell RCC has been mapped to 3p14.2 (sporadic clear cell carcinoma frequently has somatic abnormalities of *3p*), but this gene has not been isolated. Incidences for the two conditions are not known.

As in many hereditary cancer syndromes, early disease onset, multifocal disease, and bilateral disease are clues to the diagnosis. Family history is the mainstay of diagnosis; genetic testing is not available. In familial clusters of clear cell carcinoma, von Hippel-Lindau disease and tuberous sclerosis should be ruled out. Extrarenal neoplasms have been described in families with papillary renal cell carcinoma, including cancers of the stomach, rectum, breast, lung, pancreas, and bile duct. It is not clear whether these associations are significant. Individuals from families with RCC should have regular renal imaging beginning at an age 10 years younger than the age at which the youngest member of the family was diagnosed with renal cancer.

Conclusion

Overview of patient assessment for hereditary cancer

The initial evaluation and assessment of a patient is crucial. Family history and physical examination can be used tentatively to classify an index patient as sporadic, familial, or hereditary (Table 2). Family histories should be updated regularly, sometimes resulting in reclassification of risk status. Screening and genetic testing options depend on the specific syndrome under consideration. When genetic testing is clinically indicated, individuals who are positive for the mutation can enter surveillance and management programs tailored to the natural history of the particular syndrome involved. In some circumstances, prophylactic surgery may be an option. Individuals lacking the mutation can follow general population screening guidelines. Whether positive or negative, patients in this situation require compassionate genetic counseling.

Early age of onset	Late age of onset
Cancer in two or more close relatives	Single individual affected with cancer
Multiple primary tumors (same or different types)	Solitary cancer
Familial constellation of cancers consistent with a particular syndrome (e.g. breast and ovarian cancer)	
Evidence of autosomal dominant inheritance of cancer susceptibility	

Table 2 Hallmarks of hereditary cancer

It is important to be aware of factors that might obfuscate pedigree interpretation. These include incomplete gene penetrance in key relatives, early death prior to

cancer development, incomplete family history, false paternity, cancer occurrences that may be sporadic, limited patient and/or physician cooperation,, lost or destroyed records, and small family size.

We recommend comprehensive education of all at-risk patients as part of genetic counseling. We subscribe to the guidelines proposed by the American Society of Clinical Oncology, which state that predisposition testing should be offered only when (1) there is a strong family history of cancer; (2) the test can be adequately interpreted; (3) the results will influence medical management of the patient or family members; and, (4) laboratories are committed to the validation of testing methodologies.

For those patients who decide to undergo genetic testing, informed consent is a critical part of the process. Despite the potential for learning more about cancer risk, there are risks involved. Our clinic employs three sessions for women undergoing testing for mutations in *BRCA1* and *BRCA2*. The first session is a review of the basics of genetics and inheritance and the risks, benefits, and limitations of testing with a genetics nurse or counselor. After the first session a clinical conference is held, during which the patient's major concerns and predicted probabilities from the various models are reviewed. During the second visit with the patient, the clinical conference findings are reviewed by a physician and the patient makes a decision about testing. That decision may be that testing is not desired for a variety of reasons, to defer testing to a later date, or to proceed. The third visit is for results disclosure and future planning. This process, while sometimes time-consuming for both the patient and health care provider, means that the patient has time to process the complex issues and many opportunities to ask questions.

While educating patients, the health care provider must be aware of the limitations of the testing process. There are several drawbacks to current mutation detection systems, using *BRCA1* and *BRCA2* as examples. First, it is not possible to detect all disease causing mutations in *BRCA1* and *BRCA2*. Current estimates based on families with linkage to *BRCA1* and *BRCA2* suggest that the sensitivity of PCR based testing is 60 per cent in terms of detecting mutations actually present. It is also clear that variable proportions of *BRCA1* mutation carrying families are attributable to non-coding region mutations in different populations. For example, in the Dutch population it is estimated that 30 per cent of *BRCA1* families are due to genomic deletions, whereas in other populations, the numbers may be much lower. In the Ashkenazi Jewish population it is estimated that 90 per cent of mutations are due to the three founder mutations, so the proportion with genomic deletions will be much lower. Genomic deletions are not detectable with the current method of commercial testing (sequencing). Secondly, some of the mutations in *BRCA1* and *BRCA2* detected will be missense mutations. In the absence of a functional test for these mutations, one cannot always predict whether sequence variants are disease-associated. Thirdly, up to 60 per cent of high-risk families are not accounted for by mutations in *BRCA1* and *BRCA2*, resulting in uncertainty when transmitting anything other than a clearly positive result.

In addition to choosing a management plan to prevent cancer or detect cancer in at an early stage, there are other more intangible issues to consider when offering DNA-based cancer susceptibility testing. Insurance discrimination based on genetic testing has been a concern for many people. While a grave concern initially, there is no evidence that this has occurred as a result of breast cancer susceptibility testing. However, those who are self-insured or have coverage as a part of a small group may be at risk. There has been much research into the psychologic ramifications of genetic testing for *BRCA1* and *BRCA2* mutations in particular. Many studies reinforce the importance of a genetic counseling and education program, such as the one outlined above, not only to allow patients to make decisions but also to help address cancer related stress. Interestingly, a recent study has shown that patients that decline testing who scored high on measures of cancer related stress had significantly increased levels of depression. Most of these patients chose not to participate in the educational sessions offered. Patients who tested positive for a mutation did not have any change in their level of depression. These data suggests that the process of testing in the context of adequate counseling helps women with significant family histories deal with their cancer related fears or at the very least does not adversely affect them.

Further reading

Aaltonen LA, Salovaara R, Kristo P, *et al.* Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *New England Journal of Medicine* 1998; **338**: 1481–7. [A population study of HNPCC with screening recommendations.]

Ali IU, Schriml LM, Dean M. Mutational spectra of *PTEN*/*MMAC1* gene: a tumor suppressor with lipid phosphatase activity. Journal of the National Cancer Institute 1999; **91**: 1922–32. [This review looks at the syndromes and cancers where *PTEN* mutations have been found.]

Bevan S, Houlston RS. Genetic predisposition to gastric cancer. *Quarterly Journal of Medicine* 1999; **92**: 5–10. [Review of all aspects of the genetics of gastric cancer.]

Boardman LA, Thibodeau SN, Schaid DJ, *et al.* Increased risk for cancer in patients with the Peutz-Jeghers SYndrome. *Annals of Internal Medicine* 1998; **128**: 896–9. [A study of the frequency of cancers found in patients with PJS.]

Caldas C, Carneiro F, Lynch HT, *et al.* Familial gastric cancer: overview and guidelines for management. *Journal of Medical Genetics* 1999; **36**: 873–80. [Review of genetics and management of gastric cancer.]

Cohen HT. Advances in the molecular basis of renal neoplasia. *Current Opinion in Nephrology and Hypertension* 1999; **8**: 325–31. [A review of the genetics of all types of renal cancer.]

Coppes MJ, Egeler RM. Genetics of Wilms' tumor. *Seminars in Urology and Oncology* 1999; **17**: 2–10. [Review of all genetic aspects of Wilms' tumor,]

Couch FJ, Weber BL. Breast cancer. In: Vogelstein B, Kinzler KW, eds. *The genetic basis of human cancer*. New York: McGraw-Hill, 1998. [Review of breast cancer genetics.]

Eng C. *RET* proto-oncogene in the development of human cancer. *Journal of Clinical Oncology* 1999; **17**: 380–93. [Review of the syndromes caused by mutations in *RET*.]

Gayther SA, Russell P, Harrington P, *et al.* The contribution of germline *BRCA1* and *BRCA2* mutations to familial ovarian cancer: no evidence for other ovarian cancer susceptibility genes. *American Journal of Human Genetics* 1999; **65**: 1021–9. [Recent article reviewing the genetics of hereditary ovarian cancer.]

Gutmann DH, Aylsworth A, Carey JC, *et al.* The diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2. *Journal of the American Medical Association* 1997; **278**: 51–7. [Useful guidelines for the management of NF1 and NF2 patients.]

Lindor NM, Greene MH. Mayo Familial Cancer Program. The concise handbook of family cancer syndromes. *Journal of the National Cancer Institute* 1998; **90**: 1039–71. [An overview of the clinical manifestations of hereditary cancer syndromes.]

Maher ER, Kaelin WG Jr. von Hippel-Lindau disease. *Medicine* (Baltimore) 1997; **76**: 381–91. [A thorough review of VHL.]

Marx S. Multiple endocrine neoplasia type 1. In: Vogelstein B, Kinzler KW, eds. *The genetic basis of human cancer*. New York: McGraw-Hill, 1998: 489–506. [Review of MEN1.]

Morrison PJ, Nevin NC. Multiple endocrine neoplasia type 2B (mucosal neuroma syndrome, Wagenmann-Froboese syndrome). *Journal of Medical Genetics* 1996; **3**: 779–82. [Illustrated review which covers all aspects of MEN.]

Tomlinson IP, Houlston RS. Peutz-Jeghers SYndrome. *Journal of Medical Genetics* 1997; **34**:1007–11. [A nice review with illustrative pictures, although published prior to the identification of LKB1.]

Varley JM, Evans DG, Birch JM. Li-Fraumeni SYndrome–a molecular and clinical review. *British Journal of Cancer* 1997; **76**: 1–14. [A clinically oriented review of LFS.]

Vogelstein B, Kinzler KW, eds. The genetic basis of human cancer. New York: McGraw-Hill, 1998. [An excellent text which covers many aspects of hereditary cancers in depth.]

48.2 Avoidable cancer

Richard Doll

[Luck](#)
[Nature](#)
[Environmental and behavioural causes of cancer](#)
- [Synergism](#)
- [Radiation](#)
- [Smoking](#)
- [Nutrition](#)

[Cancer of specific sites](#)
- [Aerodigestive tract](#)
- [Digestive tract](#)
- [Respiratory tract](#)
- [Urinary tract](#)
- [Reproductive organs](#)
- [Lymphatic and haematopoietic tissues](#)
- [Other cancers](#)

[Further reading](#)

All types of cancer vary in incidence, not only with sex and age but also at each age in different communities. The variation between communities may be 100-fold or more, as with Burkitt's lymphoma, hepatocellular carcinoma, Kaposi's sarcoma, carcinoma of the renal pelvis, and basal cell carcinomas of the skin, or it may be less than twofold as with some childhood cancers and myeloid leukaemia in young adults; but all cancers that are at all common anywhere vary (or have varied) at least fivefold and most 10- or 20-fold. This variation in risk is due to three factors: nature or hereditary constitution, nurture or the influence of the external world and of individuals' behaviour, and luck or the play of chance.

Luck

The last is often overlooked, but it is an important determinant of why one individual develops cancer at (say) 40 years of age while another lives twice as long in apparently similar circumstances, without developing the disease. That this regularly happens is partly due to differences in genetic susceptibility, but it is also due to luck, as is demonstrated in the laboratory when genetically identical animals are kept under conditions as uniform as possible and yet develop cancer at different ages. The reason is easy enough to understand, as the development of cancer is a process that requires several changes to take place in a cell before it is capable of giving rise to a malignant clone, and each change requires the occurrence of a rare event and the failure of the cell's mechanism for repair of the damage. If luck played no part, there would be no possibility of treating a patient with breast cancer by simple excision of the lump, for all the other stem cells in the breast have the same genetic constitution and have been exposed to the same general environment since the individual's birth. If the development of cancer had a simple one to one relationship between genetic susceptibility and exposure, each stem cell would give rise to a malignant clone at about the same time, which fortunately they never do. The fewer the events that are required to happen in a given cell before a malignant clone can appear, the less opportunity there is for luck to play a part and the more likely it is that cancers will occur at about the same age, as is seen in families with familial adenomatous polyposis and *par excellence* in the two eyes of children genetically susceptible to retinoblastoma.

Nature and nurture determine the probability that each individual will develop cancer and luck then determines exactly who will do so and when. But although the role of luck in any individual case is large, it has little effect in determining the number of cases that occur each year in a large town and a negligible effect on the number that occur annually in a whole country; for the larger the population the more certain it is that good and bad luck will average out. In the rest of this chapter luck is, therefore, ignored, as it plays no part in determining the differences in the incidence of cancer in different communities that have led to the conclusion that most cancers are largely avoidable, in the sense that practicable modifications of behaviour and the environment could reduce age-specific incidence rates nearly everywhere by some 80 to 90 per cent.

Nature

The role of nature in determining the probability that an individual develops cancer may be large and is examined in Chapter 48.1. Its role in determining the variation in the incidence of cancer between different communities is, however, small. Exceptions occur, such as the role of skin colour in determining the risk of some types of skin cancer and other hereditary factors may be responsible for a few other community differences such as that between the incidence of chronic lymphatic leukaemia and (to a less extent) multiple myelomatosis in Europeans and in Chinese and Japanese peoples. Nature, however, cannot play any part in determining the changes in the incidence of cancer that are occurring nearly everywhere with time, most markedly in cancers of the lung and stomach and in melanoma, nor can it account for the large changes that occur when migrants move from one country to another.

Environmental and behavioural causes of cancer

These changes are attributable to nurture and study of its effects has led to knowledge of the way at least half the cancers that occur annually throughout the world could be prevented. In examining them, I consider first, the way different causes interact to increase each other's effects; second, three factors that influence the incidence of many different types of cancer; and third, the principal types of cancer and their individual avoidable causes.

Synergism

Some causes act synergistically with others, so that the risks they produce together are greater than the sum of the risks they cause individually when acting alone. This has been shown for combinations of smoking, alcohol, ionizing radiation, asbestos, and some components of diet, but it probably also holds for many other factors. That factors should interact in this way is understandable, when it is borne in mind that the production of cancer is a multistage process and that several different changes have to take place in a cell before a malignant clone is produced. In practice this means that the risk of a particular type of cancer may be greatly reduced in several ways. For example, much the same effect on the risk of developing cancer of the oesophagus can be achieved either by avoiding smoking or by avoiding alcohol, and the avoidance of both has very little more effect than the avoidance of one. The result of this synergism is that when the proportions of cancers that may be avoided by different means are added up the sum may well exceed 100 per cent.

Radiation

Ionizing radiation

One agent that causes practically every type of cancer, is ionizing radiation, which is energetic enough to penetrate to the nucleus of a cell and mutate the DNA. It acts synergistically with other agents so that the effect of a unit dose on a given tissue is approximately proportional to the normal incidence of cancer in that tissue. It follows that a given dose must be anticipated to cause more cancers of the stomach in (say) Japan, where the disease is normally common, and fewer cases in the United States, where the disease is uncommon, and that the reverse will be true for cancer of the lung. The precise quantitative relationship in each tissue has yet to be worked out. For the present, it has been assumed that all but a few tissues are equally susceptible and that the effect of a given dose is to multiply the normal incidence of cancer in that tissue by the same amount. As with other carcinogens, there is a latent period of about 10 years before any effect is observed. The effect then increases for a few years and subsequently persists in a constant proportional relationship with the normal incidence of the disease. The actual numbers of cancers produced will therefore increase progressively with the age of the irradiated subjects. On the basis of this simple relationship it has been estimated from the experience of the survivors of the Hiroshima and Nagasaki atomic bombs that the effect of exposing the whole body to 1 Gy of low linear energy transfer (**LET**) radiation, such as is produced by X-rays, g-rays, and cosmic rays, is to cause a lifetime risk of death from cancer of approximately 12 per cent in an irradiated population with the age distribution of the population of the United Kingdom and larger and smaller risks in direct linear proportion with the size of the dose. With low rates of delivery or with low doses, such as those from normal medical radiography, some allowance can be made for the efficacy of cellular repair and the risk is estimated to be halved to 6×10^{-4} per cGy.

With increasing knowledge, this simple rule will certainly be modified. It is already known that two tissues react differently. The cells that give rise to acute leukaemia and chronic myeloid leukaemia are more susceptible to the induction of cancer than average and the risk produced rises to a maximum earlier (in 5 to 10 years) and

then falls to practically zero after 30 years, while the cells that give rise to chronic lymphatic leukaemia do not appear to be susceptible at all. The thyroid also appears to be more susceptible than average, while tumours induced in the bone increase in incidence with time and then decrease like those induced in the marrow.

The a- and b-radiation emitted by radionuclides deposited on the skin or absorbed by ingestion or inhalation penetrate only short distances into tissues and exert their effects depending on the physiological distribution of the nuclides throughout the body. Dose for dose the effect of b-radiation is much the same as that for X-rays. a-Radiation and neutrons, however, which give up their energy over short distances (high linear energy transfer or high LET radiation) cause relatively much greater effects for a given dose—up to 10 or 20 times that of low LET radiation—and this is allowed for when the radiation dose is expressed in sieverts rather than in grays.

When all types of natural radiation are taken into account it is estimated that the total dose (in conjunction with other agents) accounts for about 5 per cent of all cancer deaths in the United States and somewhat less in the United Kingdom, where radon levels are, on average, lower.

Smoking

Another agent which affects many tissues is tobacco smoke. It contains some 50 chemicals that are known to cause cancer in animals and some that have been shown to cause cancer in humans when they have been exposed to large amounts in the course of their work, such as 2-naphthylamine and benzene. Inhalation is a highly effective way of administering a drug and, consequently, it is hardly surprising to find that cigarette smoking causes cancer in many different organs. Eight types of cancer were recognized to be caused in part by smoking, when the International Agency for Research on Cancer reviewed the subject in 1986 (cancers of the mouth, pharynx (other than nasopharynx), oesophagus, pancreas, larynx, lung, renal pelvis, and bladder) and several other types are also now known to be caused similarly (cancers of the stomach, liver, nose, and body of the kidney and myeloid leukaemia). For some of these cancers, smoking 20 cigarettes a day increases the incidence 15 to 20 times; for others it increases it only 50 per cent. For all of them, however, the consistency of the findings in different communities, the progressive increase in risk with the amount smoked, the reduction in risk in ex-smokers with the time since smoking was stopped, and the lack of any adequate alternative explanation leaves no doubt that cigarette smoking contributes to the production of the disease.

Cigarettes have nearly always contributed more to the risk of cancer than pipes or cigars because their smoke, being acid, is less irritating and more easily inhaled. Nicotine consequently reaches the alveoli where it is quickly absorbed and gives rise to a rapid and (to the addict) gratifying increase in the level in the blood. The smoke from pipes and cigars is generally alkaline and more irritating; but, being alkaline, the nicotine can be absorbed from it in the mouth. Low tar cigarettes are always also low in nicotine and are consequently inhaled more deeply to ensure that the same amount of nicotine is absorbed. Paradoxically, this can spare the bronchi from the deposition of smoke droplets and reduce the risk of bronchial carcinoma, while delivering the same amount of absorbable carcinogens to the pulmonary blood and so causing the same risk to the distant organs such as the pancreas and bladder.

Pipe and cigar smoking, although less hazardous in general, do, however, cause an equal risk of cancers of the mouth, pharynx, oesophagus, and extrinsic larynx and a relatively small risk of cancers of the intrinsic larynx and lung. Pipes also cause an additional risk of cancer of the lip.

The importance of smoking as a cause of these diseases is illustrated by the findings of the American Cancer Society's second cohort study in which information was obtained about the smoking habits of a million American men and women who were then followed for 12 years and the causes of the deaths among them determined. Mortality rates, standardized for age, are shown by smoking habits for 10 types of cancer in Table 1.

Site of primary cancer	*Ratio of mortality rates in cigarette smokers and lifelong non-smokers*	
	Men	Women
Mouth	13.6	10.5
Pharynx	(74.0)[a]	19.0
Larynx	10.6	12.0
Oesophagus	5.6	9.8
Pancreas	2.0	2.3
Lung	23.9	14.0
Bladder	3.9	1.8
Kidney	2.5	1.6
Stomach	2.1	1.4
Liver	2.5	1.6

[a] Atypically high because only one death occurred in a non-smoker, a more typical fig- ure would be about 10.0.

Table 1 Mortality from various types of cancer in cigarette smokers compared with non-smokers

In consequence of its widespread effects, cigarette smoking is the single most important cause of fatal cancer in all economically developed countries. In the United Kingdom it was estimated to have caused 52 per cent of all fatal cancers in men in 1975, a proportion that was reduced to 40 per cent 20 years later; in women the proportion increased over the same period from 12 to 20 per cent. In the United States, cigarette smoking, as opposed to the consumption of tobacco in other ways, became prevalent in men somewhat later and a maximum of 45 per cent was reached only in 1985; in women, however, cigarette smoking became prevalent somewhat earlier and the proportion attributable to smoking by 1995 had risen to 28 per cent.

Nutrition

Nutrition, also, is an important factor in determining the incidence of cancer in many, if not in all, organs. Carcinogens that cause only a small number of cancers in animals may cause several times more in the animals that are fed *ad libitum* rather than kept half starved. How far variations in the total intake of calories, within the normal human limits, determines the incidence of cancer in humans is uncertain. Levels of nutrition that lead to obesity certainly increase the risk of some cancers, most notably cancers of the endometrium and gallbladder and, after the menopause, cancer of the breast. This has a straightforward explanation in the first and third of these cancers, as oestrogens are formed from androstenedione in adipose tissue; as adipose tissue constitutes the main source of oestrogens after the menopause, the greater the amount the greater the oestrogen level in the blood. A high intake of calories in childhood also increases the risks of breast and testis cancer by bringing forward the age of sexual maturity.

Fat

Whether fat, which in Western diets has accounted for some 40 per cent of calories, has a specific effect on the incidence of some cancers, apart from its contribution as a major source of calories, is controversial. Ecological correlations suggested that it might have a specific role in increasing the risk of cancers of the breast, endometrium, ovary, colon, rectum, and prostate and support for this idea was obtained in many case–control studies of patients with and without the various diseases. Follow-up studies have now shown that women who eat above average amounts of fat do not have an increased risk of breast cancer, at least not in developed countries, and that the consumption of olive oil (a mono-unsaturated fat) may even reduce the risk. There is, however, consistent evidence of a positive, but weak, association between the consumption of fat and the risk of cancer of the large bowel, possibly because it increases the secretion of bile acids and consequently the promoting effect of bile acid metabolites in the stools. For the other types of cancer the evidence is inconclusive.

Meat

Several studies have suggested a positive association between the consumption of meat and the risk of cancer of the large bowel. Again, however, the evidence is conflicting and no evident reduction in risk was observed in a combined analysis of five studies of vegetarians.

Starch and fibre

The possible roles of starch and fibre are discussed under cancer of the large bowel.

Vegetables and fruit

The evidence that green and yellow vegetables and fruit help to reduce the risk of a wide variety of cancers is strong. The effect of fruit is most marked on the stomach and brassica vegetables may have some specific effects on the large bowel; but some component of vegetables appears to protect against the development of most of the principal cancers, most consistently cancers of the oesophagus, stomach, and lung, less consistently cancers of the mouth, pharynx, colon, pancreas, and breast,

and not at all cancer of the prostate.

Micronutrients

Vitamin A (retinol), which has been shown to be protective in many animal experiments, is unrelated to the incidence of cancer under the conditions of life of developed countries, but many studies have shown that the risk of cancer increases inversely in proportion to the serum levels of b-carotene. Controlled trials have, however, shown that b-carotene is ineffective in reducing the risk of cancer and it may be that some other carotenoid with which b-carotene is associated is the active agent. At present the best prophylactic advice is to recommend the regular daily consumption of green and yellow vegetables rather than any specific ingredient.

Vitamin C may be the component of fruit and vegetables that reduces the risk of gastric cancer and there is some evidence to suggest that vitamin E, with or without the addition of selenium, may also act as a prophylactic.

Other natural components

Other components of food that might exert a protective effect are the oestrogenic isoflavones in soya food, which have been thought to be protective because of the low incidence of breast and prostate cancer in China and Japan, where large amounts of soya food are consumed. Mechanisms are easy to postulate, since soya has many potentially relevant physiological effects, but the evidence that it exerts a protective effect is weak.

Other components

The grilling and roasting of meat produces a variety of mutagenic compounds, both polycyclic aromatic hydrocarbons and three-ring N-heterocyclic aromatic compounds; but there is no direct evidence that they are the cause of any detectable amount of cancer in humans. Nor is there anything to suggest that the trace pesticides that are found in food cause any material risk. Natural foodstuffs contain many chemicals that are found to be carcinogenic in animals when tested under standard conditions. These chemicals have, in many cases, been developed as natural pesticides and Ames has calculated that an average Western diet contains more than 1000 times as much of these ‘natural pesticides' as of man-made pesticides that may contaminate it. Two specific components that are carcinogenic are aflatoxin, which contributes to the production of liver cancer, and some component of the salted fish that is consumed in China and contributes to the production of nasopharyngeal cancer.

Cancer of specific sites

In the rest of this chapter the environmental and behavioural factors that are known to cause or prevent cancer, or are suspected of causing or preventing it, are considered separately in relation to each of the principal organs in which cancer occurs. The description of each group of cancers is preceded by a table giving the incidence of the main types in England and Wales and the white population of the United States and corresponding figures for some of the high and low incidence areas throughout the world.

The rates given under these latter headings have been recorded by Cancer Registries and are extracted from the publications of the International Agency for Research on Cancer (1992). They are not necessarily the most extreme to have been recorded, but have been selected as typical. For some cancers much more extreme rates have been recorded in special surveys of populations where no regular cancer registry is maintained: for example, in Iran, and most of Africa. Rates have, moreover, been cited only when the registry has recorded more than 1000 cases of all cancers in each sex so as to ensure that the extreme rates are not greatly influenced by random variation. Despite the rigorous standards of registration that the International Agency requires before it accepts data for publication, some of the very low rates are likely to be as low as they are because of incomplete registration. The data given for the United States are derived from the National Cancer Institute's Surveillance, Epidemiology and End Results programme, covering 10 registration areas. Those given for Japan, when they are cited as exceptionally high or low rates, are the arithmetic means of the rates for four prefectures and two cities. More detailed information about causal factors can be found in publications by the International Agency for Research on Cancer (1990) and Schottenfeld and Fraumeni (1996).

Aerodigestive tract (Table 2)

Type of cancer	England and Wales		USA white population		High incidence population	Low incidence population
	M	F	M	F		
Lip	0.5	0.1	2.3+	0.2	Australia, South M 11.0, F 1.5	Japan M 0.1, F 0.1
Salivary glands	0.5	0.4	1.0	0.7		
Nasopharynx	0.4	0.2	0.5	0.2	Hong Kong M 20.5, F 11.2	England and Wales
Tongue and oral cavity	2.4	1.1	5.2	2.0	France, Bas Rhin M 15.6, F 1.9 India, Bombay M 11.3, F 5.0	England and Wales
Other pharynx	1.5	0.5	3.6	1.1	France, Bas Rhin M 19.0, F 0.8 India, Bombay M 13.0, F 3.0	Japan M 1.7, F 0.5

Table 2 Cancers of the aerodigestive tract (age standardized incidence rates per 100 000 per year)

Lip

Carcinoma of the lip was the first type of cancer to be related to an avoidable cause when, more than 200 years ago, it was attributed to smoking pipes. Subsequently it has been found to be produced, though less easily, by smoking in other ways and by exposure to ultraviolet light. In combination, tobacco and ultraviolet light account for the great majority of all cases and the reductions in pipe smoking and in outdoor work account for the current rarity of the disease in the United Kingdom compared with the past.

Tongue, mouth, and pharynx

Carcinomas of the tongue, mouth, and pharynx (other than the nasopharynx) are all produced by the consumption of alcohol and almost equally by all methods of smoking. Alcohol and tobacco act SYnergistically, so that the avoidance of either has almost as much effect as the avoidance of both. The risks are greatly increased by chewing mixtures of tobacco, betel (nut or leaf), and lime, and are so high in some parts of India that they account for 20 per cent of cancers of all types. Cases occur particularly in parts of the mouth in which the quid is held for prolonged periods and this varies between individuals and between communities. Though commonly known as ‘betel chewers' cancers', the risks are not eliminated by the omission of betel. Tobacco in the form of oral snuff is also responsible for some cases in the United States.

Other causes include occupational exposure in the manufacture of mustard gas and (for cancer of the tongue) syphilitic leucoplakia. The reason for the recent increase in the mortality from oral cancer is unknown.

Nasopharynx

Cancer in the nasopharynx is common in south China and in Chinese people who have emigrated from the area. It is moderately common in Malaysians, in Kenya, in parts of North Africa, and in Alaskan Eskimos and American Indians. In all other populations, it is rare. In high incidence populations, the Epstein–Barr virus is integrated into the DNA of the tumours and high levels of antibodies against the virus are found in the patients' serum. Infection with the Epstein–Barr virus is not alone sufficient to cause the disease and its development in the high risk Chinese populations is associated with the consumption of a special kind of salted fish early in life. Nitrosamines extracted from the fish have caused nasopharyngeal cancers in rats.

Salivary glands

Cancers of the salivary glands are uncommon in all populations except Eskimos. No special causes are known apart from a genetic factor that increases the risk of

both cancer of the salivary glands and breast.

Digestive tract (Table 3)

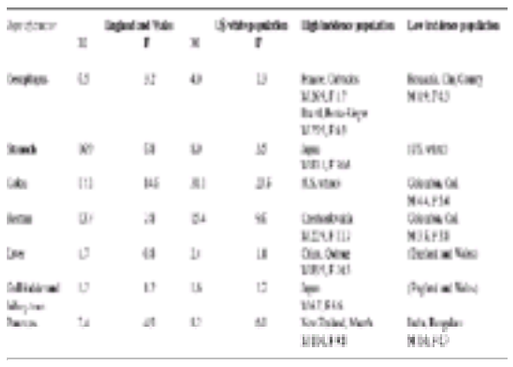


Table 3 Cancers of the digestive tract (age standardized incidence rates per 100 000 per year)

Oesophagus

Carcinoma of the oesophagus, like cancers of the mouth and pharynx, is closely related to both smoking (of cigarettes, cigars, and pipes) and the consumption of alcohol. In the absence of either the risk in the United Kingdom is reduced by about three-quarters and only very little more by the absence of both, as the two factors interact synergistically. Nutritional deficiencies have probably also played a part in the past, particularly in women. A few cases arise in the scars caused by accidental poisoning with corrosive chemicals and a very few cases as a regular complication of a rare hereditary type of tylosis with palmar keratoses.

In Africa and Asia, extremely high rates occur in parts of Shansi and Henan provinces of China, on the east coast of the Caspian Sea, on the east coast of Lake Victoria, and in the Transkei region of South Africa, where the incidence in men and occasionally also in women, may equal the highest rates of cancer of the lung recorded in European and North American cities. In these areas, tobacco and alcohol are of minor importance. When used they increase the risk, but the principal causes are unknown. In China, there is evidence to suggest that mycotoxins produced by *Fusaria* fungi may play a part, but the only established factor, common to all the extremely high incidence areas, is a restricted diet, particularly poor in animal protein and green vegetables. Despite intensive research it has not been possible to implicate any of the nitrosamines that are such a prolific cause of the disease in animal experiments.

In Japan, a few cases have been related to consumption of bracken fern.

Stomach

Cancer of the stomach was for a long time the most common type of cancer in the world, apart from the non-melanomatous skin cancers that are poorly enumerated, but was displaced by cancer of the lung in the mid-1970s. High rates persist in Central and South America, the former USSR, China, and Japan; elsewhere it has become much less common, most notably in the United States. Only in tropical Africa has the incidence been generally low and in some of these areas it is now beginning to become more common.

In the developed world, the disease has been associated with poverty, the incidence being five times as great in unskilled labourers as in members of the professions. No specific cause has been identified, but the disease is closely associated with consumption of food preserved in salt and, to a much smaller extent, with smoked food. Nitrites and nitrates, which can be converted to nitrates *in vivo*, have been thought to cause it because they can interact with secondary amines in the stomach and produce locally acting nitrosamines. The risk of developing the disease is reduced when the diet contains substantial amounts of fruit and fresh vegetables, which are the principal sources of vitamin C, and vitamin C will inhibit the formation of nitrosamines in the stomach. Attempts to demonstrate a direct relationship with nitrites and nitrates have, however, failed.

Pathologically, carcinoma of the stomach is preceded by intestinal metaplasia of the mucosa or by atrophic gastritis, both of which may result from early infection with *Helicobacter pylori*. Infection is associated with a doubled incidence of the disease, but it has yet to be shown that treatment of the infection will reduce the risk.

Chronic gastric ulcers do not become malignant, but an early carcinoma may cause an ulcer. The two diseases are otherwise associated only to the extent that would be expected from their common association with low socio-economic status.

Large bowel

Classed together as cancers of the large bowel, cancers of the colon and rectum constitute the second most common type of cancer in most developed countries. Both are less common in Eastern than in Western Europe and much less common throughout Africa and Asia. The incidence in Japan has, however, increased rapidly in the last 25 years and in Japanese people who migrated to Hawaii it soon came to equal that in the Caucasian population.

Burkitt's idea that cancers of the large bowel occurred because of a deficiency of dietary fibre, that is of vegetable fibre that is not digested by the alimentary enzymes, has proved too simple; but the idea that the risk is increased if very little foodstuff reaches the large bowel is probably correct. Fibre, however, is not the only type of polysaccharide to do so. A variable proportion of starch is also not digested, the proportion varying with its origin and the method of serving (being, for example, relatively high if it comes from bananas and potatoes that are allowed to cool after cooking). Starch that is resistant to digestion in the small bowel and much of the dietary fibre serve as pabulum for bacteria in the large bowel and consequently increase faecal bulk and, in sufficient quantities, modify the pH. Brassica vegetables (cabbage and brussels sprouts) also help to reduce risk. For the effect of other aspects of diet, see nutrition.

Two chronic diseases of the large bowel increase the risk of cancer: namely, ulcerative colitis and rectal infection with *Schistosoma japonicum*. The latter is very common in the Yangtze valley of central China and is the major cause of what is a rare disease elsewhere in the country.

The risk is increased to a small extent by a sedentary lifestyle and reduced similarly by long term use of two drugs: aspirin and oral contraceptives.

Despite all the similarities listed above, cancers of the colon and rectum are in some ways aetiologically distinct. Colonic cancer is more common in women than in men at young ages, particularly when it occurs on the right side, whereas rectal cancer is nearly twice as common in men as in women. The greater incidence of colonic cancer in women is reflected in a relationship with pregnancy which suggests the possibility of a hormonal contribution, the disease being slightly more common in those who have not borne children than in those who have.

Pathologically neither type of cancer arises often *de novo* but both arise predominantly in adenomatous polyps whether present singly, in small numbers, or in the large numbers characteristic of familial adenomatous polyposis and Gardner's SYndrome.

Anus

Anal carcinoma is caused, in part, by infection with the same types of the human papilloma virus that cause carcinoma of the cervix and is, in consequence, commonly associated with anal intercourse.

Liver

Liver cancers are mostly hepatocellular carcinomas and these vary enormously in incidence from one country to another, being rare in most developed countries and the most common type of cancer in men in parts of tropical Africa. In the high incidence areas they are attributable to a combination of infection with the hepatitis B virus and the consumption of aflatoxin, a powerful hepatic carcinogen in animal experiments. Infection with the virus commonly occurs in infancy or childhood and persists in adult life, while aflatoxin is produced in peanuts, maize, and other oil-bearing foods stored in hot and humid conditions and contaminated with *Aspergillus*

flavus. Immunization against the virus is now carried out soon after birth in many high incidence areas and should lead to a sharp reduction in the risk of the disease. Some cases are also caused by the hepatitis C virus, an RNA virus that is spread by blood and is not amenable to control by immunization.

In countries where food contains little or no aflatoxin, most cases are due to cirrhosis, irrespective of whether it is caused by hepatitis, alcohol, or haemochromatosis. Rarely, hepatocellular carcinoma may be caused without cirrhosis by the use of steroid contraceptives or anabolic steroids taken to increase muscular strength.

Cholangiosarcomas are less common and tend to occur at older ages than hepatocellular carcinomas. In China, Korea, Japan, and Thailand they may occur as a complication of infection with the liver flukes, *Clonorchis sinensis* and *Opisthorchis viverrini*.

A third histological variety, variously described as angiosarcoma or reticuloendothelioma, is normally so rare that no more than two or three cases occur annually in the United Kingdom. When, therefore, it was produced by chemicals, the causes were easy to recognize. Two were used in medicine, and their use has ceased: namely, inorganic arsenic and thorium given as a radiological contrast medium in the form of Thorotrast. A third, vinyl chloride, is used extensively in the manufacture of plastics, but exposure to it is now effectively controlled and few if any cases continue to be produced.

Gallbladder and extrahepatic bile ducts

Cancers of the gallbladder and extrahepatic ducts are commonly classed together, which is unfortunate as they differ aetiologically in many ways. The former is more than twice as common in women as in men, is strongly associated with obesity, and is nearly always preceded by cholelithiasis. The second is more common in men than women and is increased in incidence by clonorchiasis and possibly also by chronic ulcerative colitis. Both types are uncommon. The highest incidence of gallbladder cancer in women, about four times the British rate, is recorded in Jewesses in Israel. In the United States the incidence has fallen sharply due principally, if not entirely, to the high rate of cholecystectomy for gallstones.

Pancreas

The disease is generally regarded as a disease of the developed world, but the diagnosis is difficult in the absence of a well-developed medical service and much of the recorded variation between communities and over time is probably due to differences in the availability of medical care. The incidence of the disease is doubled in cigarette smokers (see smoking p.000) and in diabetics and the highest rate (about double that in Britain) is recorded in New Zealand Maoris, who smoke heavily and are prone to obesity, diabetes, hypertension, and myocardial infarction.

Peritoneum

See pleura.

Respiratory tract (Table 4)

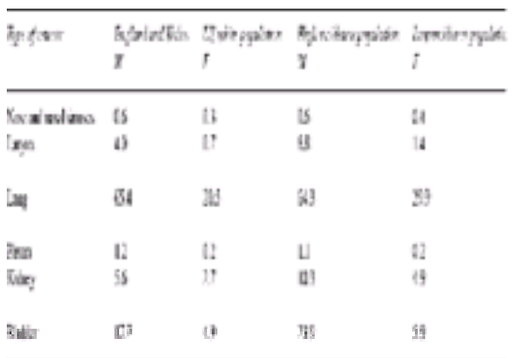


Table 4 Cancers of the respiratory and urinary tracts (age standardized incidence rates per 100 000 per year)

Nose and nasal sinuses

Cancers of the nasal cavity and nasal sinuses are rare in all countries and it has consequently been easy to recognize local occupational hazards. These have caused increased risks in nickel refineries, due probably to several compounds of nickel (including nickel subsulphide, nickel oxide, and nickel sulphate) and also in factories making hardwood furniture, leather goods, and mustard gas or causing exposure to sulphuric acid mists. In some instances the risks have been increased several hundred times, and as many as 5 per cent of heavily exposed workers have contracted the disease. Most cases, whether occupationally induced or not, are squamous carcinomas; but those attributable to hardwood dust have been characteristically adenocarcinomas.

Some few cases are caused by tobacco smoke. The increased risk in cigarette smokers is, however, small, being of the order of 50 per cent.

Larynx

Study of the causes of laryngeal cancer has been complicated by the fact that the causes in different parts of the larynx (the glottis and the supraglottal parts) have differed. Glottal tumours are principally due to smoking, while supraglottal tumours, though related to smoking, are equally if not more closely related to the consumption of alcohol and in India to betel and tobacco chewing. High risks in northern Italy, France, and the United States black population can probably be attributed to the combination of tobacco and alcohol, but it seems likely that some other factor, possibly nutritional, also contributes to the relatively high rates in Brazil (São Paulo), the Middle East, and North Africa.

A few cases have been caused by occupational exposure to mustard gas and sulphuric acid mists.

Lung

Bronchial carcinoma has been the most common type of cancer in the world since the early 1980s (excluding only non-melanomatous skin cancers that are not adequately recorded). It was a rare disease in all countries until shortly after the First World War, when the death rate attributed to lung cancer began to increase rapidly in many countries. Some of the increase was an artefact due to improved methods of diagnosis, but much was real. In England and Wales a peak was reached in the male rate in 1973, when the disease accounted for 36 per cent of all cancer deaths in men and 9 per cent of deaths from all causes. By 1995 the male rate was falling at all ages and had fallen to less than a quarter of the maximum rate under 40 years, to less than a third at ages 40 to 49 years, and to less than a half at ages 50 to 59 years. In women, the rate had fallen to a small extent at young ages, but was continuing to rise at ages 70 years and over and lung cancer was challenging breast cancer as the leading cause of cancer death. Similar changes have taken place in all developed countries, although, in most, the onset of the epidemic has been later and the peak has yet to be reached. Now the increase is spreading to developing countries as well. The rapid increase in incidence, and the commencing decrease in some countries, are attributable almost entirely to changes in the consumption of tobacco, the switch to cigarettes, and, recently to a small extent, to the reduction in the tar content of the smoke. At its peak, smoking has been responsible for nearly 95 per cent of all cases of bronchial carcinoma in men in the United Kingdom and 90 per cent of cases in women. The variation in rates throughout the world can be attributed, again almost entirely, to differences in the evolution of smoking, the exceptionally high incidence in Maori women in New Zealand being due to the adoption of smoking before the end of the nineteenth century. In the absence of smoking, rates are uniformly low in all countries, apart from some populations in China (and possibly in some other developing countries) where much higher rates are associated with indoor pollution in unventilated houses and, in women, also to cooking with rape-seed oil in a wok. On stopping smoking, an increased risk persists, due to the damage done by initiating agents; but the risk stops increasing with age, showing that the smoke must also act as a promoting agent.

Many other causes of lung cancer have been discovered as a result of localized exposure in industry. High concentrations of radon in uranium and some other mines have caused high risks and exceptionally high risks in the past in some mines at Schneeburg and Jachymov in Central Europe. High risks have also been produced by heavy exposure to asbestos and the conditions that used to be found in some nickel refineries. Other bronchial carcinogens to which people have been exposed by their occupation are the polycyclic aromatic hydrocarbons produced by the combustion of fossil fuels, most markedly in the manufacture of coal gas and coke, but also

in iron foundries and aluminium refineries; arsenic trioxide encountered in the manufacture of arsenical pesticides and the refining of copper; hexavalent chromium compounds encountered in the manufacture of chromates from chrome ore; and bischloromethyl ether, sulphuric acid mists, and mustard gas. Some small risk has been produced from exposure to silica and many other causes have been suspected and may have caused proportionally small increases in risk which may, however, have been substantial in absolute terms as the disease is normally so common. These suspected causes include man-made mineral fibres, beryllium and beryllium compounds, diethyl sulphate, and formaldehyde.

Several of the known causative agents are found in the general environment, either naturally or as a result of human activities. The most important is radon gas, the amount of which in the atmosphere varies from one part of the world to another, depending on the nature of the subsoil; in some houses it may build up to levels comparable with those in mines in which an occupational hazard of lung cancer has been observed. Radon and smoking act synergistically so that the effect in smokers is greater than that in non-smokers. Current estimates of its effect suggest that radon may be responsible for 6 per cent of all lung cancers in the United Kingdom and some 10 to 15 per cent in the United States. Atmospheric pollution with coal smoke has been important in many countries in the past and still is in some. At its height, atmospheric pollution with coal smoke was estimated to account for about 10 per cent of all bronchial carcinomas in cities. Coal smoke has now been displaced in the United Kingdom by the combustion products of oil and gas; despite the enormous increase in traffic, the amount of polycyclic aromatic hydrocarbons released is much less and is unlikely to account for more than a small fraction of 1 per cent of cases. Tobacco smoke in the environment is now an equal or greater hazard and is estimated to be responsible for about a quarter of all cases in non-smokers. The hazard from asbestos in the environment from its use as a building and insulating material is trivially small, unless it is disturbed or allowed to become fragmented.

Not all the histological types of bronchial carcinoma are produced equally by all the known causes. Smoking used principally to increase the risk of squamous carcinomas and to have very little effect on the incidence of adenocarcinomas, but with the introduction of low-nicotine cigarettes this has changed. The low nicotine levels are compensated for by deeper inhalation, with the consequence that adenocarcinomas produced by smoking in the periphery of the lung are becoming relatively more common. Occupational hazards have generally caused all the main types of the disease, but radon and bischloromethyl ether have tended characteristically to produce small cell carcinomas, while cooking with a wok has principally caused adenocarcinomas.

Pleura

Cancers of the pleura are nearly all mesotheliomas. About 80 per cent are caused by occupational exposure to asbestos or to similar exposure in unusual domestic or environmental circumstances. Some have been produced by heavy exposure to ionizing radiation and the remainder may be attributed to the small amount of asbestos that occurs ubiquitously in the general environment and to natural radioactivity. In some Turkish villages they have been produced by a similar fibre, erionite, that occurs in the local rock. Mesotheliomas are produced much more readily by the straight fibres of amosite and crocidolite asbestos, which persist in the body for life, than by the curly and less persistent chrysotile fibres. Similar tumours occur less commonly in the peritoneum, due to the same causes; except that they may not be produced by chrysotile at all.

Some mesotheliomas show evidence of infection with a simian-40-like virus, but its aetiological role is uncertain.

Urinary tract

Kidney

The incidence of cancer of the kidney has been increasing slowly in the United Kingdom, now predominantly in the elderly. About a quarter of the common adenocarcinomas of the body of the kidney are due to smoking, as are two-thirds of the relatively rare squamous carcinomas of the renal pelvis. The increase in women may be attributable mainly, if not entirely, to smoking, but not the increase in men.

Renal pelvis cancers are very common in villages along the banks of rivers in Serbia, Bulgaria, and Romania where Balkan nephropathy is rife; in these villages they are about 100 times as common as in the United Kingdom. The cause is unknown. Some cases used to occur as a result of the renal nephropathies induced by analgesic mixtures containing phenacetin.

Renal tumours in childhood are nephroblastomas. The peak incidence is at the age of 1 to 2 years, but some slow growing tumours continue to appear clinically up to 10 years of age.

Bladder

Cigarette smoking is the most important cause of bladder cancer, accounting for about half all cases in the United Kingdom and the United States.

Occupational hazards in the chemical industry were due to at least three chemicals, the use of which has been abandoned (2-naphthyl-amine and 4-aminobiphenyl, both of which are found in tobacco smoke, and benzidine). The hazards produced were in some cases extreme; in one plant all 15 men who distilled 2-naphthylamine developed the disease as did the 5 men in another plant who were employed for 15 years or more in the manufacture of benzidine. The chemicals were used mainly in the manufacture of dyes and people who subsequently used the dyes were also liable to be exposed to them. Occupational hazards of bladder cancer also occurred in the rubber industry with the use of an antioxidant containing 2-naphthylamine and in the retort houses of gasworks, where the same chemical was present in the fumes from coal tar.

Bladder cancer has been caused by two drugs (chlornaphazine, which was used briefly in the treatment of myelomatosis, and cyclophosphamide) and by infestation with *Schistosoma haematobium* in Africa. In the latter case, the disease probably results from the associated persistent bacterial infection in the bladder with the formation of nitrosamines by bacteria *in vivo.*

Several dietary causes have been postulated, partly on the basis of animal experiments (cyclamates and saccharin) and partly on the basis of epidemiological observations (coffee). Intensive investigation has, however, failed to confirm any harmful effect from artificial sweeteners and the weak association recorded with the consumption of coffee may have been due to confounding with smoking.

All these agents cause transitional cell carcinomas, except for schistosomiasis which causes squamous cell carcinomas.

Reproductive organs (Table 5)

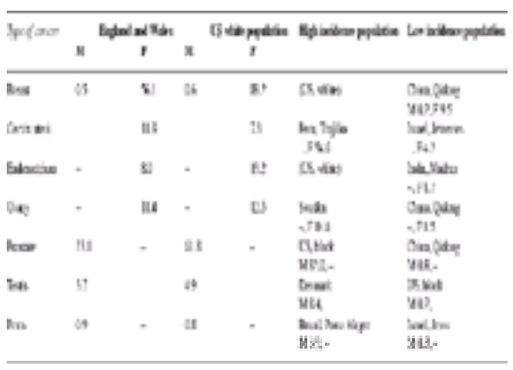


Table 5 Cancers of the reproductive organs (age standardized incidence rates per 100 000 per year)

Breast

Cancer of the breast is the most common cancer in women worldwide (excluding the poorly recorded non-melanomatous cancers of the skin) and accounts for nearly 20 per cent of all cancers in women. It is more common in the developed countries than in the less developed countries of Africa and Asia and is increasing in incidence nearly everywhere, slowly in the former and rapidly in some of the latter. The occurrence of the disease is determined primarily by hormonal factors and is

predisposed to by an early menarche, a late menopause, and nulliparity. Among parous women, the risk is decreased the earlier the age at first full-term pregnancy and independently, but less markedly, by increased parity and prolonged lactation. High levels of oestradiol in the blood predict an increased risk and steroid contraceptives and hormone replacement therapy both increase the risk, but only whilst they are being used and for 5 to 10 years afterwards.

Dietary factors modify the incidence of the disease partly by determining age at menarche and the rate of growth in childhood and partly by determining the degree of obesity. A large body size at all ages and obesity after the menopause increase risk, but before the menopause obesity is associated with late menarche, irregular periods, and a decreased risk. Dietary fat *per se* probably does not increase the risk (see nutrition p. 3423), but the consumption of alcohol does, the risk being increased by about 10 per cent for each 8 g of ethanol consumed per day. Physical inactivity is a further minor factor increasing the risk.

Much, if not all, of the changes in incidence can be explained by changes in the prevalence of these factors.

Cervix uteri

Cancer of the cervix is the second most common cancer in women worldwide. It was much more common in developed countries 100 years ago than it is now, and is still very common in South and Central America, tropical Africa, and India. In the United Kingdom the risk has been increasing again in women who reached sexual maturity after about 1960.

The disease was long thought to be related to childbearing, because the incidence was much greater in multiparous than in nulliparous women. Childbearing, however, plays little or no part and the risk is determined primarily by the risk of infection with certain types of the human papilloma virus, something that is related to the number of sexual partners that a women and, independently, her partner, may have. The viral types responsible are not those that cause vulval warts (types 6 and 11) but types 16, 18, 31, 45, and several others that are less common. Infection with these viruses is frequent and is not enough, in itself, to cause the disease. Two factors that may interact with infection are cigarette smoking (by causing the excretion of mutagens in the cervical mucus) and the use of oral contraceptives. These cannot, however, be the most important cofactors as the disease was prevalent before they were introduced. Diet could be one factor, as the disease is associated with relatively low levels of serum carotene, and some aspect of sexual hygiene could be another as the disease is associated with poverty and, in some communities, with the absence of circumcision.

Most cervical cancers are squamous carcinomas, but both squamous and adenocarcinomas are associated with human papilloma virus infection. The relatively rare adenocarcinomas are somewhat more closely associated with the use of oral contraceptives than the squamous carcinomas.

Endometrium

Endometrial carcinoma is aetiologically almost the reverse of cervical cancer. It is rare in poor countries and common in rich countries; common in nulliparous women and progressively less common as parity increases; unrelated to sexual intercourse and, like cancer of the breast, positively associated with early menarche, late menopause, and postmenopausal obesity. Its incidence is reduced by the use of combined steroid contraceptives and cigarette smoking (which modifies the metabolism of oestrogens) and increased by the use of sequential contraceptives, postmenopausal oestrogens, the presence of oestrogen-secreting ovarian tumours, and the use of tamoxifen for the treatment of cancer of the breast. All these relationships can be simply explained by the extent to which the endometrium is exposed to oestrogens unopposed by progestogens.

Choriocarcinoma

Choriocarcinoma occurs in about 1 in 30 000 pregnancies. Nothing is known about its aetiology except that it is somewhat more common in Chinese, Malaysian, and Philippino populations than in others.

Ovary

Adenocarcinomas of the ovary share many of the aetiological characteristics of endometrial cancer, in that the risk of developing the disease is greater in rich countries than in poor, increases with the length of time between menarche and the menopause, and decreases with increase in the number of pregnancies and the use of combined steroid contraceptives. It is not, however, affected by medication with oestrogens, but is related to the total number of ovulations.

The many other histological types of the disease are rare and have not been clearly related to any of the above factors.

Prostate

Prostate cancer increases in incidence with age more rapidly than any other cancer. It has, in consequence, become more prominent with the increase in the elderly population and now accounts for 10 per cent of all cancers in men. Age-specific incidence rates are, moreover, increasing, partly as a result of screening, as foci of cells indistinguishable from malignant cells are found in a high proportion of apparently normal prostates in elderly men, and partly as a result of a genuine increase in risk, as mortality rates have also increased.

Hormonal factors presumably play an important part in the production of the disease, but none has been identified. Clues to aetiology may be found in the high rates recorded in the United States black population and the low rates in Chinese and Japanese people. These differences may be partly genetic in origin, but they are not wholly genetic as the rates in Japanese men are increasing and are higher in Japanese migrants to the United States than in Japan.

Testis

Testicular cancers may be teratomas or seminomas. Both begin to appear in late adolescence, occur most frequently between 20 and 44 years of age (with teratomas peaking before seminomas), and then become progressively less common with increasing age. Both types have become more common over the last 60 years in Europe and North America, but have continued to be rare in black, Chinese, and Japanese men, wherever they live. The reason for the increase is unknown. The only known aetiological factors are incomplete descent of the testis, which increases the risk 10-fold; maternal obesity during pregnancy, which increases it by about 50 per cent; and early sexual maturity. Incomplete descent of the testis has also become more common in the last two or three decades, but still accounts for only a small proportion of cases.

Penis

Carcinoma of the penis, like carcinoma of the cervix, is closely related to infection with certain types of the human papilloma virus; but infection alone is not enough to produce the disease. The disease can be prevented almost completely by circumcision within the first week or two of life and largely by circumcision later in childhood. Within populations that do not practice circumcision, the variation in incidence may be related to the degree of personal cleanliness.

Lymphatic and haematopoietic tissues (Table 6)

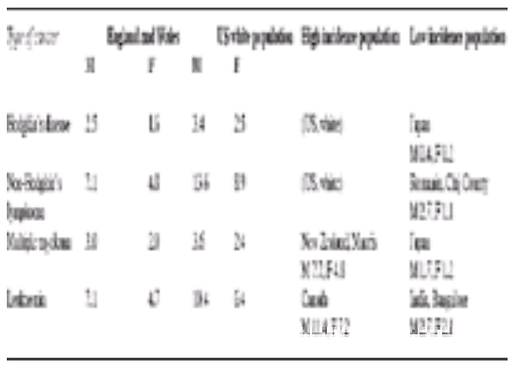


Table 6 Cancers of the lymphatic and haematopoietic tissues (age standardized incidence rates per 100 000 per year)

Hodgkin's disease

Hodgkin's disease consists of a variety of histological and clinical types that should properly be divided into at least two groups, one with a peak incidence in young adult life and one which becomes slowly more common with increasing age. In undeveloped countries the peak in early adult life is absent and there is instead a small peak in childhood. The change in age distribution with a rising standard of living is reminiscent of the change that occurred in the incidence of poliomyelitis and suggests that the young type may be due to an infectious agent that becomes less prevalent as social hygiene improves. Some adult cases are due to infection with the Epstein–Barr virus (human herpes virus type 4).

Non-Hodgkin's lymphoma

Non-Hodgkin's lymphoma is a constellation of different diseases with different aetiologies.

One that can be distinguished histologically, clinically, and epidemiologically is the B-lymphocyte lymphoma named after Burkitt. It affects children throughout the world, but is rare everywhere except in a few places where malarial infestation is heavy. In these areas, the incidence is 100 times that in Europe and North America. In the high incidence areas the disease is due to a combination of infection with the Epstein–Barr virus which becomes integrated into the DNA of the prelymphocytes, and heavy malarial infestation which stimulates their multiplication. Lymphomas due to Epstein–Barr virus infection also constitute the majority of the lymphomas that arise following intensive immunosuppression, whether due to drugs or to the development of AIDS.

Another type of Non-Hodgkin's lymphoma is the lymphoma that constitutes part of the adult T-cell leukaemia/lymphoma syndrome that follows infection with the human T-cell leukaemia/lymphoma virus. The disease is common in the South of Japan and in the Caribbean, but may occur occasionally anywhere.

A third type is the primary upper small intestinal lymphoma that affects young people in many populations with a low standard of living, not only in North Africa and the Middle East (which gave it its earlier name of ‘Mediterranean lymphoma'), but also in Africa south of the Sahara and in Central and South America. Malnutrition, however, is not enough to cause the disease as it does not occur in many countries where malnutrition is common, such as Bangladesh.

The remaining lymphomas, which constitute the majority of those observed in Europe and North America, may need to be further divided. Considered as a group they are increasing in incidence. There is weak evidence to suggest that they may be related to the widespread use of modern pesticides, but no occupational hazard has been established.

Myelomatosis

The recorded incidence has increased greatly in all developed countries over the past 50 years, but it is uncertain how far this is an artefact due to improvements in diagnosis and how far it reflects a real increase in the incidence of the disease. No occupational or other cause for the disease is known.

Leukaemia

Leukaemia, like non-Hodgkin's lymphoma, is a constellation of diseases with different clinical and epidemiological characteristics.

One type, chronic lymphatic leukaemia, increases in incidence with age in the same way as most of the common epithelial cancers. It is common in Europe and North America and rare in India, China, and Japan, due presumably to a difference in genetic susceptibility as the difference persists in migrants from low to high incidence areas. The differences are found to be marked even when the lymphocytes are immunotyped, as it is principally a disease of B lymphocytes in the high incidence areas and of T lymphocytes in the low. No environmental or behavioural cause is known.

Acute lymphoblastic leukaemia occurs at all ages, but half of all cases occur in childhood with a peak incidence at 2 to 3 years of age. The cases causing the childhood peak derive from B lymphocyte precursors and most possess a distinctive antigen and are described as ‘common acute lymphoblastic leukaemias'. The disease is seldom diagnosed in undeveloped countries, but this may be an artefact due to the rapidity with which affected children succumb to infection. There is suggestive, but inconclusive, evidence that a few cases are caused by exposure to magnetic fields of the order of 0.2 μT and above produced by the passage of electricity along high-power cables. Acute lymphoblastic leukaemia in adults is a different disease. One type, which is almost if not entirely limited to migrants in the United Kingdom, is derived from T-lymphocyte precursors and has been described under non-Hodgkin's lymphoma.

Acute myeloid leukaemia increases slowly with age from early childhood onwards and is the principal type of leukaemia in early adult life. In this age group the incidence of the disease is less variable throughout the world than almost all other types of cancer, other than the very rarest. Chronic myeloid leukaemia, in contrast, is very rare in youth, increases in incidence with age more rapidly, and is more common than acute myeloid leukaemia after about 50 years of age. Both types are more readily induced by ionizing radiation than other cancers, as is acute lymphoblastic leukaemia, and a substantial proportion of all three types can be attributed to natural background radiation. Myeloid leukaemia, and particularly its rare variant, erythroleukaemia, is an occupational hazard of workers exposed to benzene. It is slightly more common in cigarette smokers than in non-smokers and is produced by several drugs: chlorambucil, cyclophosphamide, melphalan, methyl-CCNU, myleran, treosulphan, therapy with a combination of nitrogen mustard, vincristine, procarbazine, and prednisone (known as MOPP) and probably due to the nitrogen mustard component and, if given in doses large enough to cause aplastic anaemia, busulphan. The risks are all small in comparison with the therapeutic benefits.

Other cancers (Table 7)

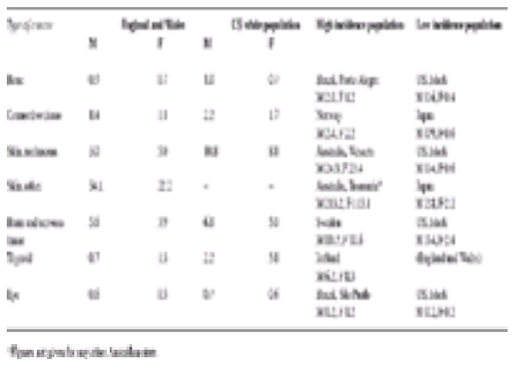


Table 7 Other cancers (age standardized incidence rates per 100 000 per year)

Bone

Most bone cancers are osteosarcomas and these are responsible for the peak incidence that occurs in adolescence and for the subsequent increase in middle and old age. The only known causes are the genetic abnormality that leads to retinoblastoma, which sometimes also leads to osteosarcoma; occupational or medicinal exposure to radium; and whatever it is that causes Paget's osteitis deformans. The last accounts for all the increase in later life and predisposes to cancer to such an extent that osteosarcomas develop in 1 per cent of all affected patients.

Connective tissue

Connective tissue sarcomas are rare throughout the world. A relationship with exposure to chlorophenols and phenoxyacetic acid herbicides has been postulated, but the evidence is inconclusive.

Skin

Basal cell and squamous carcinomas of the skin are common in white populations throughout the world. Nearly all the former and most of the latter occur on the face, head, and neck and are due to exposure to ultraviolet light. The risk decreases with increasing pigmentation and similar cases are almost unknown in black populations, unless they suffer from albinism. As the result of a defect in DNA repair, both types of cancer occur in large numbers on the skin of people with xeroderma pigmentosum.

Squamous carcinomas of the skin also have many other causes. Occupational exposure to polycyclic aromatic hydrocarbons in coal tars and lubricating oils have, in the past, caused many cases on the forearms and scrotum and local customs have caused the ‘kangri cancers' of the abdomen in Kashmir (due to carrying a charcoal stove inside the clothes for warmth) and the ‘dhoti cancers' of the groin in India (due to continued friction from the dhoti cloth). In tropical Africa, squamous carcinomas are a common complication of chronic ulceration of the legs. Squamous carcinomas may sometimes be due to infection with the specific type of the human papilloma virus (type 5) that causes the flat warts of epidermodysplasia verruciformis in association with an inherited defect of cell-mediated immunity and with other types of the virus in patients treated with immunosuppressive drugs.

Melanomas are less common than basal and squamous cell carcinomas but more important because of their higher fatality. In white-skinned populations, their incidence is closely correlated with latitude, being higher the nearer the Equator, and all are related, in one way or another, to exposure to ultraviolet light. Those on the head and neck, like the basal and squamous cell cancers in the same site, are due to chronic exposure. Those elsewhere in the body are more common in office than in outdoor workers, are associated with the frequency of sunburn, and are caused principally by excessive exposure of the untanned skin, particularly in childhood. Changes in customary clothing and the frequency of sunbathing have, in consequence, led to a steady increase in the incidence of the disease in many countries over the last 70 years. In black African populations who go barefoot, melanomas occur at the junction of the pigmented and unpigmented areas.

Kaposi's sarcoma is now classed as a skin cancer rather than a cancer of the connective tissues. Until the 1980s it occurred in two forms. One was exceptionally rare, affected elderly men, and occurred everywhere, but particularly in Jews in, or from, Eastern Europe. The other, which affected principally young men in Central Africa, was so common in some parts that it constituted 8 per cent of all patients with cancer admitted to hospital. Since the early 1980s a third type has appeared in association with AIDS, mostly in men who contracted the disease from homosexual intercourse, less commonly in drug addicts, and very seldom in haemophiliacs. The frequency of the association with AIDS has been much greater in the United States than elsewhere and has diminished with the passage of time. All three types depend primarily on infection by the human herpes virus type 8.

Brain

Brain tumours like so many others are a mixture of different histological types with different epidemiological features. Medulloblastomas occur only in childhood, glioblastomas only in adult life, while astrocytomas occur at all ages. Most types are slightly more common in males, but meningiomas are more common in females. Several occupational hazards have been suspected, particularly in the chemical industry, but none has been confirmed. The minimal variation in incidence between different communities suggests that environmental factors are unlikely to be of much importance. Whether the increase in incidence which has been reported in many countries is real or an artefact of improved diagnosis with the aid of brain scans is uncertain. Some excess has been reported in association with occupational exposure to magnetic fields.

Thyroid

The incidence of thyroid cancer is related to the intake of iodine in a complex way. Relatively high rates occur in association with endemic goitre in iodine deficient areas, such as Switzerland and Colombia, but the highest rates are recorded in areas where the intake is above average, such as Hawaii and Iceland. The histological types, however, differ; papillary carcinomas predominate in the deficient areas and follicular carcinomas predominate in the areas with excess. Medullary and anaplastic carcinomas are rare everywhere.

The only other known causes are ionizing radiation, which causes both papillary and follicular carcinomas, and the hereditary factor which causes medullary carcinomas of the thyroid as part of the multiple endocrine SYndrome.

Eye

Eye cancers in adults are nearly all melanomas. They occur slightly more often than would be expected in rural areas and in people with blue eyes and fair complexions, which suggests an aetiological role for ultraviolet light, despite the filtering effect of the cornea and lens. The retinoblastomas that occur in childhood are commonly hereditary and are discussed in Chapter 48.1.

Further reading

Parkin DM, Muir CS, Whelan SL, Gao YT, Ferlay J, and Powell, J, eds. *Cancer incidence in five continents, vol. VI. IARC Scientific Publication No. 120*. International Agency for Research on Cancer, Lyon, 1992. [For detailed figures for the incidence of cancer of 50 anatomical sites by sex in 50 countries in the late 1980s.]

Schottenfeld D, Fraumeni JF. *Cancer epidemiology and prevention*. Saunders, Philadelphia, 1996. [For a detailed account of the human evidence for known and suspected causes of all the principal types of cancer. For reference.]

Thun MJ, Day-Lalley CA, Calle EE, Flanders WD, Heath CA. Excess mortality among cigarette smokers: changes in a 20-year interval. *American Journal of Public Health* 1995; **85**: 1223–30. [For an account of the risks of the important smoking related diseases by sex and amount smoked and the increased risks with longer duration of past regular cigarette smoking.]

Tomatis L, ed. *Cancer: causes, occurrence and control. IARC Scientific Publication No. 100*. International Agency for Research on Cancer, Lyon, 1987. [For an expanded and readable version of the evidence summarized in this chapter, as available 10 years previously.]

48.3 Chemotherapy of cancer

I. Craig Henderson

Introduction
Uses of systemic therapy
Potential benefits from systemic treatments
Adjuvant therapy
Neoadjuvant therapy: organ preservation
Palliation of symptoms
Important principles in the use of chemotherapy
Patient factors likely to affect response to therapy
Systemic therapy: uses and side-effects
Cytotoxics
Major drug groups
Endocrine therapy
Treatment of specific solid tumors
Lung cancer
Breast cancer
Gastrointestinal tumors
Prostate cancer
Testicular cancer
Ovarian cancer
Evolving systemic therapies
Resistance to cytotoxic therapies
High-dose chemotherapy
Immunotherapy
New therapeutic targets
Further reading

Introduction

When cancers metastasize they usually go to multiple sites. Optimal palliation can often be achieved by treating a few of these sites with surgery or radiation, but a systemic therapy that can reach many sites will usually be more effective. Some sites, such as bone marrow, are not easily accessible to local modalities because of the diffuse nature of these organs and their malignancies. Thus it is not surprising that leukemias were the first malignancies successfully treated with chemotherapy, one of the first malignancies successfully cured with systemic therapy, and the tumor type in which most of the principles for using chemotherapy were established. Other sites of metastases, such as liver and lung, may not be amenable to local treatment with radiotherapy because the optimal dose of radiation to reduce the tumor burden substantially is much greater than the maximally tolerated dose of radiation by normal tissue in that organ.

In general, local treatment of metastases has little or no impact on patient survival. Whether metastases are macro- or microscopic, systemic therapy is more likely than local treatment to prolong survival. Relatively few tumor types can be cured with systemic therapy; it has been estimated that 90 per cent of drug cures occur in 10 per cent of tumor types. However, substantial prolongation of survival is possible, especially when SYstemic therapy is administered immediately following definitive therapy of the tumor primary with surgery and/or radiotherapy.

The effectiveness of systemic therapy varies from one cancer site to another, and the optimal systemic treatments that should be used in each site are also highly variable.

Uses of systemic therapy

In 1896 Beatson demonstrated that removing the ovaries of premenopausal women could result in shrinkage of inoperable breast cancers. This was probably the first successful use of systemic therapy to treat cancer. In the 1940s estrogens were effectively used to treat breast cancer in postmenopausal women. The first cytotoxic therapies resulted from observations on the effects of nerve gases on bone marrow cells during the Second World War, and this was one of the reasons that leukemias were the first target for cytotoxic chemotherapy. Soon after, Sidney Farber successfully induced a complete remission in a child with acute lymphoblastic leukemia using aminopterin, an antifolate and precursor of methotrexate, which he used because he noted these patients had high circulating folate levels. This was the wrong reason but right result. These two forms of systemic therapy, endocrine and chemotherapy, have very different origins and were developed within the context of different paradigms. Endocrine therapy has been considered a removal of or interference with critical tumor-cell growth factors, and it has been widely assumed that this would have primarily a cytostatic rather than cytocidal effect on the tumor cell. Endocrine therapy is effective primarily or exclusively in cells that have a hormone receptor. Since it is rare that all or even most cancer cells in a given tumor are receptor positive, it has been assumed that endocrine therapy would not reduce the number of viable tumor cells to a level sufficient to prolong patient survival, let alone cure disease. We now know from large clinical trials that this is not true. Both ovarian ablation and the antiestrogen, tamoxifen, will substantially prolong the survival of patients with early breast cancer. In contrast, cytotoxic agents were thought to poison the cell and to be primarily cytocidal. By the 1970s there was sufficient evidence to conclude that some cancers could be cured with cytotoxics, most notably acute lymphoblastic leukemia of childhood. However, similar dramatic benefits from chemotherapy in the treatment of the common solid tumors, such as breast, lung, and colon cancer, have not been forthcoming. In fact, the survival benefits from the use of tamoxifen are greater than from cytotoxics when they are used to treat postmenopausal women with early breast cancer. Our current understanding of the various effects of endocrine and chemotherapy suggest that the assumptions and paradigms of the past were oversimplified.

Most tumors consist of multiple clones of cells that are morphologically similar but physiologically different in their sensitivity to various drugs. In the case of prostate and breast cancer, some cells are sensitive to endocrine treatments, others to cytotoxics, and some to both. In almost all tumors there are clones sensitive to one cytotoxic and resistant to some others. Most systemic therapies affect DNA replication or transcription, but expression of the DNA damage may involve disruption of cell cycle checkpoints, growth factors, growth factor receptors, and/or signal transduction pathways. Cell death from most systemic therapies involves a common pathway, apoptosis or programmed cell death. Abnormalities in this pathway may account for resistance to either endocrine or cytotoxic therapy. Better understanding of the molecular events associated with therapy has contributed to the development of fundamental principles for treatment, such as using combinations of systemic therapies, as well as the discovery of new therapeutic targets, such as the HER2/*neu* receptor.

Potential benefits from systemic treatments

When given to patients with metastatic or locally advanced cancer, systemic therapy may palliate SYmptoms, improve survival, or both. The effectiveness of various therapies is usually expressed as a response rate. A partial response is assigned when tumor volume (or more often the area of bidimensional lesions that can be measured on physical examination or radiographs) decreases by 50 per cent or more. Generally, all measurable lesions must decrease for at least 1 month during which no new lesions appear. A complete response is complete disappearance of all measurable and evaluable (i.e. detectable but unmeasurable) lesions. An objective response does not guarantee either palliation of symptoms or prolongation of survival. Conversely, patients without an objective response may experience a decrease in pain, an improvement in their ability to perform work or recreational activities, or an improved sense of well being. Response rates are useful ways of comparing systemic therapies, and there is evidence that treatments with higher response rates result in better outcomes, including improved survival. For example, a meta-analysis of 50 randomized trials conducted in patients with metastatic breast cancer showed that an increase in response rate from about 20 to 79 per cent was associated with a 6-month increase in median survival. It has also been shown that patients randomized to regimens with higher response rates experience a better overall quality of life, even when the more effective treatment is more toxic. Practically, however, there are many instances where the physician must choose between a treatment with high response rate and substantial toxicity and one with a slightly lower response rate and much less toxicity. It is not clear that the decision should be the same for all patients.

Patients with cancer who are facing death from cancer are willing to suffer much greater toxicity from treatment for a modest benefit than either patients without cancer or medical professionals. This is illustrated by a formal study of English patients about to begin chemotherapy as treatment for one of a variety of cancers (Table 1). These patients were willing to undergo considerable toxicity for any one of three possible outcomes: a 1 per cent chance of cure, a 12-month increase in survival, or a 10 per cent chanceof palliation. Other patients, cancer specialists, and general practitioners all felt that greater benefits were needed to justify this much toxicity. This study also demonstrates the patient's vulnerability to any overestimation of benefits from therapy, especially a remote chance of cure, whether this hope is offered by a

medical professional as standard or experimental treatment or by a purveyor of alternative treatments.

	Chance	Prolongation	Chance
	of cure (%)	of life (months)	of relief (%)
Patients with cancer	1	12	10
Patients without cancer	50	36-60	75
Cancer doctors (medical oncologists, radiotherapists)	10	12	50
Oncology nurses	50	24	50
General practitioners	75	24	1

Adapted from Slevin ML, *et al.* Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. *Journal: Medical Journal* [PMC, 2]. [4]:[8]-[8].

Table 1 The minimum benefit felt to be sufficient to justify treatment by 100 patients with cancer about to begin therapy with cytotoxics, 100 patients without cancer, 60 medical oncologists, 88 radiotherapists, 790 general practitioners, and 303 oncology nurses. Before answering, each was told the treatment would probably cause severe nausea, vomiting, hair loss, tiredness, weakness, decreased sexual interest, and would require frequent injections and hospital admission for 3 to 4 days each month. The median values among the responses given by each group are shown in the table.

The most consistent benefit from treatment of metastatic cancer is palliation of symptoms. There are some diseases, such as melanoma and renal cancer, where the net clinical benefit from toxic chemotherapies is questionable. Fortunately, most symptomatic patients can usually decide for themselves whether the decrease in SYmptoms justifies the increase in side-effects from therapy by a therapeutic trial lasting 2 to 3 months. Shorter durations are frequently inadequate. Systemic treatment may cure some non-hematologic tumors even when widely metastatic (Table 2). However, systemic therapy will not cure the vast majority of patients with metastases from common tumors, such as breast, lung, and colon cancer. When life is prolonged, it is usually for periods of less than a year. Very modest improvements in survival may even be seen in patients with endstage disease who have become refractory to most therapies. For example, patients with breast cancer whose tumors had progressed while receiving or shortly after completing several different types of chemotherapy were randomized to receive a more or a less effective form of chemotherapy. The median survival of the group as a whole was about 6 months, but it was a month longer for those randomized to receive the more effective of the two therapies.



Table 2 Benefits from systemic therapy in various tumor types.

Adjuvant therapy

The greatest survival benefits from the use of systemic therapy are achieved when it is administered soon after diagnosis of the primary and before the appearance of distant metastases as an adjuvant to optimal local treatment with surgery and radiotherapy. Patients die of cancer primarily because of metastases that occur months or years prior to diagnosis and that are out of reach of local treatment. The patients most likely to have occult or micrometastases at distant sites are those with regional lymph node involvement, but other factors, such as tumor grade, S-phase fraction, and various genetic mutations, are also useful in identifying patients with a high risk of recurrence. It intuitively makes sense to attempt eradication of these micrometastases in high-risk patients whenever local treatments are used with curative intent.

Micrometastases are also more sensitive than established metastases to the effects of SYstemic therapy. It is generally accepted that the growth of tumors is characterized by a Gompertzian model which predicts logarithmic growth of a tumor or metastatic focus until it reaches about 1 cm in diameter, which is 37 per cent of its maximum size and the size when most tumors can be readily detected on physical or radiologic examination. Thereafter the growth fraction (i.e. the percentage of all tumor cells in a mass undergoing mitosis) and the slope of the tumor growth curve decrease dramatically. This plateau in the growth curve probably occurs because tumors of this size must develop new blood vessels to maintain growth on the outer surface of the tumor and become necrotic because of inadequate oxygenation at the inner core. Since most forms of cytotoxic therapy are more lethal to rapidly dividing cells, micrometastases are more sensitive than clinically apparent metastases to the effects of chemotherapy. In addition, it is difficult to deliver any antineoplastic agent to the relatively avascular inner cores of large tumors. As tumors divide, the chance of developing mutations that impart resistance to either endocrine or cytotoxic therapy also increases.

As with the development of many other applications of chemotherapy, clinical proof of the principle for adjuvant therapy was first demonstrated in pediatric tumors, especially Wilms' tumors. More recently, large clinical trials and meta-analyses or overviews have provided convincing evidence that the survival of adult patients with solid tumors can be significantly prolonged by the use of adjuvant treatment.

The nature of the benefit from adjuvant therapy is often thought to be cure, but there is no incontrovertible evidence of this. If adjuvant therapy imparted a cure to a substantial number of patients, then deaths from cancer as a percentage of all deaths should be fewer in the treatment than the control arms of randomized trials comparing adjuvant chemotherapy with none. In a 20-year analysis of one of the earliest trials comparing adjuvant cyclophosphamide, methotrexate, and 5-fluorouracil with no adjuvant therapy for patients with early breast cancer, there was a survival benefit from adjuvant therapy since only 66 per cent of the treated but 77 per cent of the control patients had died. However, among the patients who had died, all but 10 per cent of the treated and 7 per cent of the control group had died of breast cancer. This suggests that the time of death but not the cause of death had been changed as a result of adjuvant treatment.

The treatment of the primary tumor by surgery has an all or nothing effect. Either the patient is cured because all tumor cells have been removed or the patient dies at about the same time s/he would have died without local therapy because of metastases established at distant sites long before diagnosis that are unaffected by the local treatment. In contrast, treatment with systemic therapy is often partial, even when given as an adjuvant; metastatic or micrometastatic tumor deposits may be substantially reduced without eradication and eventually regrow, causing the patient's death some months, years, or decades later than if systemic therapy had not been given. A 10 per cent survival difference between a treated and untreated population could occur because 10 per cent of the patients were cured, all of the patients had some prolongation of survival and none were cured, or a very few were cured and the majority has some prolongation of life. Based on our current knowledge, the last seems to be the most likely outcome when adjuvant systemic therapy is used to treat common solid tumors. A patient's decision whether to undergo a very toxic treatment will depend on which of these outcomes he or she assumes is more likely and the importance of that particular outcome to the patient. Surveys, such as the one summarized in Table 1, indicate that patients are willing to withstand enormous toxicity for a relatively small chance of cure but require a larger benefit to justify treatment if the only benefit is prolongation of survival without cure. On the other hand, some patients will settle for very modest benefits if they feel quite certain they will be among the winners.

Neoadjuvant therapy: organ preservation

The use of SYstemic therapy before local treatment (neoadjuvant therapy) will reduce the size of (i.e. downstage) responsive tumor types, and this may allow a less radical form of local treatment with preservation of the organ. Theoretically this earlier treatment might also improve survival, as well, but this has generally not been demonstrated in randomized trials. Tumors where neoadjuvant therapy may increase organ preservation are shown in category 8 of Table 2. After neoadjuvant

treatment it is more difficult to define appropriate surgical margins, and there is an increased risk of leaving tumor foci that are no longer grossly apparent but are still microscopically present and capable of regrowth.

Palliation of symptoms

In an attempt to be objective in evaluating the effects of systemic therapies, medical oncologists have placed greater emphasis on measuring changes in the size of tumor lesions than on the effects of these therapies on tumor-related symptoms. Net clinical benefit determinations in which both the relief of symptoms and the toxicity of the therapy are together considered in calculating the improvement or detriment in quality of life resulting from therapy are difficult to find and even more difficult to interpret. However, randomized trials comparing chemotherapy with best supportive care suggest that when a treatment regimen induces a response in 20 per cent or more of the treated patients, palliation of symptoms is likely to occur in a sufficient number to make a therapeutic trial worthwhile. If the therapy is relatively non-toxic, an even lower response rate may be sufficient to justify an attempt at palliation.

Important principles in the use of chemotherapy

The choice of drugs, the doses selected, and the duration of therapy are contingent on the goals to be achieved from treatment. In general, the highest response rates and longest survival will be obtained if the following guidelines are used.

1. Combinations of drugs should be used rather than single agents, particularly if the tumor has not been previously exposed to systemic therapy. (Choriocarcinoma and Burkitt's lymphoma are rare exceptions to this rule, and the possible advantages from two or three drugs given in rotation or sequentially rather than concomitantly is currently under evaluation in some solid tumor types.)
2. All of the drugs in the combination should have some effect on the tumor when used alone. However, there are few instances where more than three drugs have been shown to be superior to a two- or three- drug combination in the treatment of solid tumors.
3. The maximum tolerated dose of each drug should be given since most cytotoxics have a steep dose–response curve under ideal circumstances. However, the threshold dose for seeing a biologic effect, the slope of the dose–response curve, and the point where the dose–response curve plateaus such that further dose escalation provides no or very little additional benefit varies by drug class, the schedule of drug administration, tumor kinetics, and intrinsic tumor sensitivity to the drug.
4. Drugs that are synergistic in their effects on the tumor but not on normal tissues are ideal for a combination. Each drug can then be administered at or near its optimal dose. This will maximize antitumor effect while minimizing the intensity of drug side-effects.
5. The interval between cycles of therapy should be as short as possible to minimize tumor regrowth in this interval but long enough to provide recovery of normal tissues so that the dose of the next cycle of drug does not have to be decreased. In other words, it is best to maximize dose intensity as well as dose.
6. The greatest impact of cytotoxic therapy is seen in the first 4 to 12 cycles, depending on the tumor type. Once the optimal duration of treatment has been reached, additional courses of therapy will increase the cumulative response rate in a cohort of patients but are not likely to improve survival substantially. For example, the earliest curative regimens for testicular cancer used 6 cycles of therapy; subsequent randomized trials have shown there is no improvement in disease-free or overall survival when more than 3 to 4 cycles are used. The earliest adjuvant chemotherapy regimens for breast cancer were administered for 12 to 24 months. Randomized trials have failed to demonstrate an advantage for more than 4 to 6 months of treatment.
7. Treatment should be initiated as early as possible, either in the adjuvant setting or immediately after the onset of symptoms from metastases.

These principles were originally derived from studies in animal models, but most have been confirmed in the clinic, as well. However, they are most applicable to tumors in the first four categories of Table 2. When palliation is the primary or only goal, the guiding principles are somewhat different.

1. For symptomatic patients with tumors that are relatively resistant to cytotoxics, the benefits of combination chemotherapy or higher drug doses are likely to be outweighed by side-effects from this treatment. This group includes those whose tumors have grown while receiving chemotherapy or within 6 months after completion of a course of treatment.
2. When there are small differences in response rates between two regimens, the least toxic will probably provide the greatest palliation.
3. Although combinations of endocrine and chemotherapy induce higher response rates in patients with metastatic prostate and breast cancer, they have not been shown to increase patient survival. Maximal use of endocrine therapy alone followed by cytotoxic treatments when resistance to all forms of endocrine therapy is established provides the longest time without symptoms of disease and toxicity of treatment (**TWiST**).
4. Treat until maximum response has been maintained for several months. At that point, discontinuation of treatment will not compromise survival and may provide the patient with increased TWiST.
5. Prophylactic palliation is not possible. Unless the tumor falls into one of the first four categories of Table 2 it is usually inappropriate to treat an asymptomatic patient with cytotoxics because of a rising tumor marker, such as a carcinoembryonic antigen (CEA).

Patient factors likely to affect response to therapy

In general, the patients most likely to respond to chemotherapy are those who have the most to gain and who, consequently, are most likely to tolerate the side-effects. Naturally, younger patients head this list, but older patients with good renal, pulmonary, and hepatic function will tolerate chemotherapy as well as younger patients if sufficiently motivated. Performance status is a much better indicator of whether a patient will tolerate therapy well. In general, benefits of any type are infrequently seen among patients with solid tumors and an International Union against Cancer (UICC) performance status of 3 and 4 (Table 3). Response rates are lower among patients with UICC status of 1 and 2 compared with those with a status of 0.

0	Able to carry on normal activity
1	Able to live at home with tolerable symptoms
2	Disabling symptoms but less than 50 per cent of time in bed
3	Severely disabled, greater than 50 per cent of time in bed
4	Very ill, confined to bed
5	Dead

Table 3 UICC categories of patient performance status.

Bone marrow reserve

Bone marrow, which has a high growth fraction at all times, is more frequently affected adversely by cytotoxics than any other organ. Patients whose bone marrow is compromised prior to receipt of chemotherapy are often more susceptible to the effects of chemotherapy and require a reduced dose from the first course onwards. Patients with genetic abnormalities that affect leukocyte function, infections including AIDS, or prior exposure to bone marrow toxins including radiotherapy, are in this category.

Hepatic and renal function

Many cytotoxic drugs are metabolized via cytochrome P-450-dependent reactions in the liver, and disorders of hepatic function can result in either loss of efficacy or enhanced toxicity. Cyclophosphamide is inert in its parental form and is metabolized by the liver to the active metabolite, 4-hydroxycyclophosphamide. The anthracyclines and their active metabolites are excreted by both the kidney and the liver. Severe liver dysfunction manifest by hyperbilirubinemia leads to a prolongation of the half-life of the anthracyclines and increased toxicity if the dose or schedule is not adjusted. Abnormalities in renal function may cause delayed excretion of drugs such as methotrexate and carboplatin and serve as another source of increased toxicity.

Cardiovascular history

Patients with decreased cardiac reserve from a prior infarction, hypertension, or radiation to the left chest wall are more susceptible to the cardiotoxic effects of

anthracyclines and anthracenediones.

Patients with a prior history of thromboembolic events have a higher risk of having additional episodes while receiving estrogens, antiestrogens such as tamoxifen, or chemotherapy of any kind.

Systemic therapy: uses and side-effects

Cytotoxics (Table 4(a))

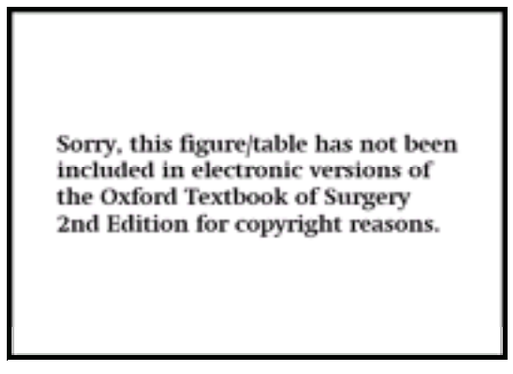


Table 4 (a) Cytotoxic agents, tumors in which they are active, and major side-effects. Drugs included in the list are those used most often to treat solid tumors.

Management of common side-effects

There would be little controversy surrounding the use of cytotoxics to treat cancer were it not for their side-effects, many of which are similar to those caused by the disease itself. Possibly the most ubiquitous and least spoken about is fatigue. While receiving chemotherapy and for several months after completion of a course of treatment, most patients find they end their day early and limit both their physical activities and social routines. To some extent the severity of fatigue is correlated with myelosuppression and anemia. Other common side-effects, such as alopecia, nausea, vomiting, stomatitis, diarrhea, and rash vary with the specific chemotherapy used and the dose and schedule of drug administration.

Vomiting

Some drugs, such as platinum and doxorubicin, induce vomiting severe enough to have required admission to hospital, intravenous hydration, and sedation, prior to the discovery of the serotonin (5-HT_3) antagonists, odansetron, granisetron, and tropisetron. The acute phase is defined as the first 24 h after administration of the cytotoxic and the 5-HT_3 antagonists are very effective during this interval. There are two separate phases of delayed nausea and vomiting, one lasting from the 2nd to the 4th and the second from the 5th to 7th post-therapy days. The 5-HT_3 antagonists are not as effective during the intermediate period. For drugs that commonly induce emesis, cocktails of antiemetics are usually employed. These frequently include corticosteroids. Twenty to 40 per cent of patients with cancer will experience anticipatory vomiting, and this number is much higher among those who have previously had uncontrolled vomiting on one or more cycles of therapy.

Cardiotoxicity

A number of anticancer drugs have been reported to increase the frequency of arrhythmias and non-specific ECG (EKG) abnormalities, including the anthracyclines, paclitaxel, and 5-fluorouracil. Anthracyclines and anthracenediones also cause a cumulative dose-related cardiomyopathy that eventually leads to congestive heart failure. The lesion is diffuse and the histology is unique to this condition, characterized primarily by vacuolization and myofibrillar dropout without a mononuclear cell infiltrate or fibrosis. It can be easily diagnosed with cardiac biopsy, and this is the most accurate means of estimating the amount of additional anthracycline an individual patient can tolerate without congestive heart failure. Overt congestive heart failure is uncommon at cumulative doxorubicin doses of less than 500 mg/m^2, occurs in about 7 per cent of patients given 500 to 600 mg/m^2, and 20 per cent of patients given over 600 mg/m^2. However, the incidence of clinically occult cardiotoxicity in the first few years following treatment with 400 to 500 mg/m^2 cumulative dose may actually be as large as 40 to 50 per cent. It is possible that patients who are cured or live for decades after treatment, such as patients with pediatric and breast cancers, will have a higher incidence of heart failure as they approach the age when other forms of myocardial disease or physiologic stresses are added. Anthracenediones, such as mitoxantrone, and liposomal formulations of doxorubicin have been shown in randomized trials to cause significantly less cardiotoxicity than doxorubicin when equally myelotoxic doses are used. Epirubicin has less cardiotoxicity when identical doses are used, but it has never been shown that epirubicin is less cardiotoxic when equally effective or equally myelotoxic doses of the two drugs are compared. Dexrazoxane administered with doxorubicin reduces cardiotoxicity even when the dexrazoxane is begun after 300 mg/m^2 of doxorubicin has been administered without dexrazoxane.

Neurotoxicity

Neurotoxicity is most often seen with the vinca alkaloids and the taxanes and less frequently with cisplatin. The neuropathies are usually distal and symmetrical. The earliest SYmptoms are numbness and paraesthesias, but with continued drug administration these may progress to weakness. Vincristine may even induce an ileus and, rarely, quadriparesis. The symptoms may persist for months or years after discontinuation of the drug. Central nervous toxicity is rare, presumably because most cytotoxics do not cross the blood–brain barrier. However, short- and long-term toxicities are associated with intrathecal administration of drugs such as methotrexate and cytarabine for the treatment of menigeal carcinomatosis.

Thromboembolic

Estrogen, antiestrogen, progestin, and cytotoxic therapies are all associated with an increased incidence of deep vein thrombosis and pulmonary emboli. This factor should be included in the decision to use these therapies in a situation where the benefits may be marginal. Anticoagulants are occasionally indicated to continue treatment in patients who experience thromboembolic events while receiving anticancer drugs.

Palmar plantar erythema (hand/foot syndrome)

Redness of the palms and soles has been reported as a side-effect of a number drugs including 5-fluorouracil, cytarabine, doxorubicin, etoposide, paclitaxel, docetaxel, hydroxyurea, capecitabine, etretinate, methotrexate, bleomycin, and mercaptopurine.

Other areas of skin, especially pressure points such as the belt line, may also be involved. The first symptom is tingling which usually precedes the rash, but if the dose or schedule of drug is not altered, edema of the hands and feet and, as a result, severe pain and exfoliation may supervene. This SYndrome is most often seen when a drug is continuously administered either intravenously (e.g. continuous infusion of 5-fluorouracil and doxorubicin or pegalated liposomal doxorubicin) or orally (e.g. capecitabine). It is also seen with very high doses of doxorubicin and weekly schedules of paclitaxel.

Second neoplasms

Treatment with some cytotoxics, notably alkylating agents and anthracyclines, is associated with an increased incidence of second malignancies. The absolute risk is small and is related to cumulative dose and, to a less extent, the size of the individual dose. The interval from exposure to the appearance of leukemia may be only a few years; solid tumors are more likely to appear a decade or several decades after exposure. The genotype of the second malignancy is often helpful in establishing that the tumor is iatrogenic. Tumors induced by alkylating agents often have loss of chromosomes 5 and 7; those induced by the epipodophyllotoxins and anthracyclines have a translocation of 11q23 and 21q22. Ironically, some of the drugs with the least acute toxicity, such as chlorambucil and melphalan, are the most likely to induce second malignancies, especially when they are administered continuously for several years. Cyclophosphamide has been linked to bladder cancers.

Infertility

Azoospermia and ovarian failure are related more to the cumulative than to the size of the individual dose of drug. The alkylating agents, including chlorambucil and cyclophosphamide, and cisplatin are the most likely to lead to infertility. Doxorubicin, methotrexate, etoposide, vincristine, and 5-fluorouracil have not been a problem in this regard.

Major drug groups

Alkylating agents

One or another alkylating agent is a core component of most regimens used to treat most human cancers. The majority of these drugs are bifunctional, having two distinct alkyl groups that can link covalently to DNA or protein. They have a predilection for alkylation on the N7 position of guanine in DNA: this can result in abnormal base pairing of the guanine with thymidine or the formation of inter- or intrastrand cross-links between two guanine bases, inhibiting DNA replication. Common toxic effects of alkylating agents include myelosuppression, which is usually dose limiting, alopecia, and gastrointestinal toxicity. Alkylating agents can also cause infertility and have been implicated in the development of secondary leukemia. There is also an increased risk of solid tumors developing, with a peak incidence 6 to 8 years following exposure, although the cumulative incidence continues to increase throughout the patient's lifespan.

Cyclophosphamide and ifosfamide

Cyclophosphamide may be administered intravenously or orally. Intravenous administration leads to higher peak blood levels, and the physician can be certain the patient has received the drug. Toxicities may be slightly more severe but are of shorter duration. The oral route provides more continuous exposure to the drug over a longer period of time and usually causes a longer duration of nausea with very little vomiting. Cyclophosphamide is frequently combinedwith methotrexate and 5-fluorouracil in the popular **CMF** regimen used to treat breast cancer. A randomized trial has demonstrated that oral cyclophosphamide imparts a significant survival advantage over intravenous cyclophosphamide in the CMF regimen. Hemorrhagic cystitis occurs in up to 10 per cent of patients treated with cyclophosphamide and a much higher percentage of those treated with ifosfamide, a congener of cyclophosphamide. For this reason, mesna and a high volume of fluids are routinely given with ifosfamide to prevent cystitis.

Chlorambucil and melphalan

Among the alkylating agents, melphalan and chlorambucil are among the least toxic. Both cause delayed myelosuppression, minimal gastrointestinal toxicity, and almost no alopecia. Neither is now considered the drug of choice for any tumor because other more active agents are available. However, each may be useful in selected situations where a patient is unlikely to tolerate more toxic therapies. Chlorambucil is sometimes substituted for other alkylating agents in lymphoma regimens. Melphalan is frequently used for multiple myeloma and on rare occasions might be used for an ovarian cancer in an elderly patient.

Nitrosoureas

The nitrosoureas are among the very few drugs that cross the blood–brain barrier, and for this reason they are used to treat brain tumors. Their use in the treatment of other tumors is limited in part by severe gastrointestinal toxicity and prolonged myelosuppression.

Antimetabolites

Although most cells are more sensitive to cytotoxics during mitosis, the antimetabolites are especially dependent on the operation of the cell cycle to be effective and are often referred to as ‘cell cycle specific'. In contrast to the alkylating agents, the anthracyclines, and the platinum compounds, which are active in a broad spectrum of tumors, each of the antimetabolites is active in a limited number of tumors. Many of them, including 6-mercaptopurine, 6-thioguanine, cytarabine, 2-chlorodeoxyadenosine, and 2-deoxycoformycin, are used primarily in the treatment of hematologic malignancies.

Methotrexate

Methotrexate is a dihydrofolate reductase inhibitor and also partially inhibits thymidylate synthase. The toxicities of methotrexate are related to both dose and schedule; relatively small doses administered frequently or high doses administered infrequently can lead to severe stomatitis, bone marrow suppression, and renal toxicity. However, the most commonly used doses cause few side-effects. The maximum tolerated dose can be increased by more than 50-fold with almost no toxicity by the administration of reduced folates (leucovorin) beginning 24 h after administration of methotrexate. Increased methotrexate toxicity occurs in patients with hypoalbuminemia, renal failure, or with large third spaces such as ascites or pleural effusions.

5-Fluorouracil

5-Fluorouracil is activated *in vivo* to 5 FdUMP (5 -fluoro-2 -deoxyuridine-5 -monophosphate), which binds to and inhibits thymidylate synthase. The administration of folinic acid (leucovorin) prior to treatment with 5-fluorouracil alters the drug's targets, increasing both its efficacy and toxicity in some tumor types, especially colorectal cancer. Although 5-fluorouracil is usually administered as a bolus, it can also be given by continuous infusion through a portable pump. This substantially alters the toxicity profile and may increase efficacy in some tumors. Patients no loner experience nausea, vomiting, and myelosuppression. Stomatitis and palmar plantar erythema become dose limiting.

Capecitabine

Capecitabine is a fluoropyrimidine carbamate that is converted to 5-fluororucil *in vivo.* It is active against breast and colon cancer, including breast cancer that is refractory to anthracyclines and taxanes. It is administered orally, twice daily for periods of 2 to 6 weeks continuously. The acute symptoms are mild and similar to those seen with continuous infusion of 5-fluororucil. It was first registered in 1998.

Gemcitabine

Gemcitabine (2,2-difluorodeoxyctidine) is a nucleoside analog and a prodrug that is activated in the cell to cytotoxic phosphorylated metabolites. It is similar in structure to and metabolized by the same pathways as cytosine arabinoside but is considerably more effective in solid tumors. The side-effects from this drug are generally quite mild and include myelosuppression, flu-like symptoms, fatigue, and elevations of liver enzymes. Its approved and established uses are for the treatment of pancreatic cancer and non-small cell cancer of the lung, but there is preliminary evidence of activity alone or in combination in a number of other tumor types.

Antitumor antibiotics

This class of drugs, dominated by the anthracyclines, is active against a broad spectrum of different tumor types.

Anthracyclines

The anthracyclines are active in all but four tumor types. They are also among the most toxic causing severe retching which often lasts for 2 to 3 days followed by nausea for another week or two, profound but very transient myelosuppression, baldness 2 weeks after a single injection, and ulceration sufficiently severe to require skin grafting at sites of extravasation. The antitumor activity of the anthracyclines may result from intercalation of DNA, inhibition of topoisomerase II activity, formation of free radicals, or a combination of these and other cellular effects. Topoisomerases (I and II) are important for repair of DNA damage, and for this reason anthracyclines have synergy with other cytotoxics as well as being very active alone. Anthracycline cardiomyopathies may be related to the formation of free radicals. Doxorubicin and its stereoisomer, epirubicin, are the most widely used anthracyclines. Although milligram for milligram epirubicin is less toxic than doxorubicin, it is not established that epirubicin is less toxic at equally effective doses. Daunorubicin is used almost entirely for the treatment of leukemias.

Mitoxantrone

The anthracyclines are tetracyclines with an attached aminosugar. The aminosugar was removed to create a new and less toxic class of drugs, the anthracenediones.

Mitoxantrone is the only drug in this class commonly used to treat cancer. It is very well tolerated and has much less cardiotoxicity, but the activity of mitoxantrone is less than that of doxorubicin. In combination with prednisone, it is the most active cytotoxic for the treatment of prostate cancer. Alone or in combination with 5-fluorouracil and leucovorin it is a reasonable option for patients with breast cancer who are unwilling or unable to tolerate an anthracycline.

Mitomycin C

Patients experience relatively few acute side-effects following treatment with mitomycin C. However, myelosuppression is more prolonged than with almost any other cytotoxic except the nitrosoureas, and this makes it more difficult to use in combination. In addition, it uncommonly causes either pulmonary fibrosis or hemolytic uremic syndrome. These may be irreversible and fatal.

Epipodophyllotoxins

This class of drugs has been in the pharmacopeia for more than 150 years, but it was first used for the treatment of cancer only 25 years ago. Like the anthracyclines, the epipodophyllotoxins inhibit topoisomerase II and produce free radicals. However, they do not usually cause cardiotoxicity. The dose-limiting side-effect is myelosuppression, but 30 to 40 per cent of patients have mild to moderate nausea and vomiting and most have alopecia. These drugs are active against a wide variety of tumors but are used primarily for the treatment of small cell and non-small cell lung cancer, germ cell tumors, and lymphomas.

Platinum compounds

The platinum compounds were first discovered serendipitously in a study of bacteria in an electric field created with platinum electrodes. The parent compound, cisplatin, is cis-diamminedichoroplatinum (II). This substance is relatively inactive, but in environments where chlorine concentrations are low, such as cellular cytoplasm, the two chlorine atoms leave the molecule and it becomes very reactive, binding to DNA in the cell nucleus. The mechanism of action is similar to that of the alkylating agents. The effects of the platinum compounds in cancer cells can be readily identified by the presence of platinum–DNA adducts. Numerous analogs have been generated in which other groups are substituted for the chlorines and/or added to the amino group in an effort to decrease the toxicity or increase the potency of the parent compound. Only carboplatin has been registered so far. The platinum compounds have a spectrum of activity that is as broad as the anthracyclines and the alkylating agents. The tumors that are most responsive to platinum are often relatively less responsive to anthracyclines.

Cisplatin

The spectrum of cisplatin activity includes most epithelial carcinomas. It is also one of the most toxic forms of chemotherapy. In fact, before the introduction of newer antiemetics, patients often had to be admitted to hospital and sedated. Myelosuppression, nephrotoxicity, and/or neurotoxicity can be dose limiting. Nephrotoxicity can be reduced by adequate hydration. Neurotoxicity can be manifest either as a peripheral neuropathy or hearing loss; these symptoms are related to cumulative dose and are only partially reversible. Patients may also manifest hypomagnesemia. Cisplatin is less likely than other alkylating agents to cause infertility, and it has not been implicated in the development of second malignancies.

Carboplatin

Carboplatin causes less nausea and vomiting, renal toxicity, and ototoxicity than cisplatin. Because the pharmacokinetics vary considerably from one patient to another, dosing is based on calculation of the area under the plasma curve for each patient. Carboplatin is routinely substituted for cisplatin, especially when the primary goal of treatment is palliation. Randomized trials demonstrated equivalence between carboplatin and cisplatin in the treatment of ovarian cancer, but two large independent trials in patients with testicular cancer showed that carboplatin was inferior to cisplatin in that disease.

Spindle poisons

This group of drugs either block the assembly (vinca alkaloids) or disassembly (taxanes) of the mitotic spindle. The vincas are among the oldest drugs used to treat cancer, and the taxanes are among the newest. All of the drugs in this group cause some degree of neurotoxicity. This is dose limiting for vincristine but less problematic for the newer spindle poisons such as vindesine, vinorelbine, and the taxanes.

Vinca alkaloids

Vincristine is the oldest member of this group and is active against many tumor types, but usually only at doses that are unacceptably neurotoxic. Because it is not myelotoxic it can be used in many combinations without compromising the dose of other drugs in the regimen. However, in many regimens there is little evidence that vincristine adds much value to the other drugs. Myelosuppression is the dose-limiting toxicity for vinblastine. Vindesine is also much less neurotoxic than vincristine, but its use is limited in part by the lack of efficacy data in many tumor types and the fact that the Food and Drug Administration (**FDA**) has never approved its use in the United States. Vinorelbine is a semisynthetic vinca that is among the most active and least toxic in this group. It appears to be as active as the taxanes in the treatment of lung and breast cancer.

Taxanes

Paclitaxel and docetaxel were originally derived from the bark or needles, respectively, of the yew tree. Now both drugs are semisynthetic. Both are effective in a wide variety of tumors, most notably lung, breast, and ovarian cancers. Myelosuppression is does limiting, but neurotoxicity and stomatitis are also seen with both drugs. Docetaxel causes fluid retention in up to 50 per cent of patients by the time a cumulative dose of 400 mg/m^2 has been reached. This can be ameliorated to some extent by comedication with corticosteroids. The toxicity, and possibly the efficacy, of paclitaxel is related to the schedule of administration. Less myelosuppression and non-hematologic toxicity are seen with shorter infusions. These drugs are used most often in the treatment of ovarian, lung, and breast cancer.

Camptothecins

This new family of drugs, originally derived from a Chinese plant, inhibit topoisomerase I, another DNA repair enzyme closely related to topoisomerase II. Two camptothecins have been registered, but at least another half dozen are in clinical trials. These drugs appear to be most effective when given repeatedly or by continuous infusion during each course of treatment. Since most of these drugs are given intravenously, repeated visits to the clinic may be required. This is unacceptable to many patients. There is surprisingly little overlap in the spectrum of activity of the two drugs now in use.

Topotecan

Topotecan is used primarily for the treatment of ovarian and lung cancer, but some activity has been observed in other tumors as well. Neutropenia is dose limiting, but thrombocytopenia occurs more frequently than with most other myelosuppressive cytotoxics. The drug is active when administered orally, but there is considerable diarrhea when currently available formulations are administered by this route.

Irinotecan

Irinotecan was first developed in Japan. It is of particular interest because it is active against colon cancer, a tumor that responds to relatively few cytotoxics. When administered daily, myelosuppression is the dose-limiting toxicity. When given intermittently at higher doses, diarrhea is dose limiting. Randomized trials have demonstrated that this drug will prolong the lives of patients with colon cancer, including those with tumors resistant to 5-fluorouracil.

Others

Interferons

Interferons are normally released from cells in response to viral infection and provide protection for non-infected cells. Interferons may affect cancer growth by inhibiting cell proliferation, preventing angiogenesis, increasing cellular differentiation, and/or altering immunomodulation. Interferons are important for the management of hairy cell leukemia, Kaposi's sarcoma, chronic myelogenous leukemia, and cutaneous T-cell lymphoma. They are used in the treatment of superficial bladder cancers by transurethral instillation. In recent years there is increasing evidence that they may prolong the time to progression or even survival of some solid tumors including melanoma, renal cell cancer, and lymphoma. The principal toxicities are fever, fatigue, and a flu-like syndrome. The severity of these symptoms is dose dependent, but

they can be very debilitating when high doses of interferon are used.

Interleukin 2

Interleukin 2 (IL-2) is a lymphokine produced by T cells, and its effects on cancer appear to be mediated by immune mechanisms. It has no direct anticancer effect. It is used alone or with adoptive transfer of immune cells for the treatment of melanoma and renal cell cancer. Response rates in both of these tumors are low, but complete remissions do occur. Side-effects can be severe, especially with higher doses, and fatalities were reported in the early studies. These consist of chills, malaise, fluid retention, hypotension, respiratory distress, nausea and vomiting, hepatic dysfunction, and disorientation.

Endocrine therapy (Table 4(b))

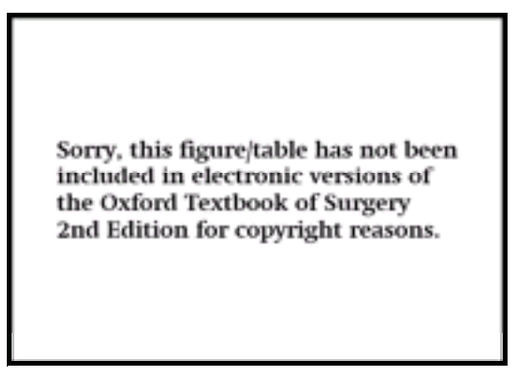


Table 4 (b) Endocrine therapies for breast and prostate cancer with their major side-effects.

Estrogens or androgens function as growth factors for a substantial percentage of all breast and prostate cancers. Removing these growth factors results in cessation of cell growth and, in many cases, cell death by mechanisms still not completely understood. For example, it is puzzling that both the addition and removal of hormones may cause tumor regression. Metastatic breast cancer lesions in a postmenopausal woman will often regress as a result of treatment with estrogens. Usually high doses of estrogen are administered, but it is not established that this is merely a dose effect. When the tumor begins to grow again, simply discontinuing the estrogen therapy may cause another regression. Sometimes additional regressions can be obtained by removing the adrenal or pituitary gland, adding progestins, or administering androgens. Endocrine treatments can palliate symptoms and prolong survival. As a general rule, endocrine treatments differ little in their efficacy and are distinguished from one another primarily by their toxicity profiles. Most forms of endocrine treatment have some effects on both breast and prostate cancer even though the endocrine treatments of choice differ.

Antihormones

Antiestrogens

The antiestrogens are now the most widely used endocrine treatment for breast cancer because they are the least toxic. Tamoxifen and toremifene have antagonist effects on the breast and agonist effects on bone, endometrium, and cholesterol levels. They cause reduction of low-density lipoproteins. Long-term tamoxifen use has been shown to slow osteoporosis and may reduce the risk of death from myocardial infarction. These two drugs decrease antithrombin III levels and increase thromboembolic episodes. More than 90 per cent of treated patients have no side-effects except hot flushes. Menses usually become irregular and sometimes cease in premenopausal women. A small percentage of patients will have mild nausea during the first few weeks of treatment, and up to 10 per cent may experience depression. There is an increased risk of endometrial cancer, most of which will be non-fatal. The risk of liver cancer is increased with tamoxifen, but reported cases are very rare. There may be less risk of liver cancer from toremifene. Tamoxifen has limited activity in a few tumors other than breast cancer, notably ovarian cancer.

New antiestrogens in clinical trials or recently approved include pure estrogen antagonists that have none of the estrogen agonist effects of tamoxifen or toremifene (ICI 182780 or Faslodex); non-steroidal antagonists with a higher antagonist:agonist ratio than tamoxifen or toremifene (droloxifene, TAT-59, and idoxifene); selective estrogen receptor modifiers (SERMS) with agonist effects on bone, cholesterol, and liver but not breast or endometrium (raloxifene).

Antiandrogens

Flutamide and its metabolites, nilutamide and bicalutamide, are non-steroidal antiandrogens that block the binding of dihydrotestosterone to its cytoplasmic receptor. Immediately after initiation of treatment flutamide induces a rise in plasma luteinizing hormone (**LH**), follicle-stimulating hormone (**FSH**), testosterone, and estradiol levels. When used alone for the treatment of advanced prostate cancer, these drugs are associated with a shorter time to progression and worse survival than obtained with orchidectomy. However, potency is preserved to a greater degree with the antiandrogens than with most other forms of endocrine treatment. Flutamide is administered three times daily, while both nilutamide and bicalutamide can be given once daily. Ab-normal light/dark adaptation, alcohol intolerance, and interstitial pneumonitis occur with nilutamide. Randomized comparisons of flutamide and bicalutamide have demonstrated equivalent efficacy and a better safety profile (primarily less diarrhea) for bicalutamide.

Antiprogesterones

The antiprogesterone, mifepistone, was developed initially as an abortifacient, but it has been shown to induce remission in some patients with breast cancer.

Ablation of endocrine function

The first endocrine treatments were surgical ablation (oophorectomy for breast cancer in 1896, orchidectomy for prostate cancer in 1954), and none of the endocrine therapies subsequently developed have been shown to have a more favorable impact on response rates or the survival of patients with these diseases. Both must be considered as primary treatment for metastatic disease, and ovarian ablation may be as effective as adjuvant chemotherapy or tamoxifen for some patients with early breast cancer. The main reasons these therapies are not used more often are the desire of many patients and physicians to avoid surgery, the permanency of the effects, and competition between specialists. Ablation of ovarian function with radiation is as effective as oophorectomy but occurs more slowly. Adrenalectomy and hypophysectomy are rarely performed today because of the high surgical morbidity and mortality of these procedures, especially when performed by those with limited experience, and the long-term complications associated with chronic dependence on hormone replacement.

Medical adrenalectomy and aromatase inhibitors

The adrenal gland is the source of 30 per cent of male androgens and 100 per cent of postmenopausal female androgens. Outside the adrenal gland, the androgen androstenedione may be converted to estrogen by aromatases found throughout the body, especially in fatty tissues and the breast. Aminoglutethimide was originally developed to treat seizure disorders but was removed from the market when it was found to inhibit adrenal function. Subsequently this observation led to its evaluation in patients with metastatic breast and prostate cancer where it was found to be as effective as adrenalectomy. Similarly, ketoconozole's primary use is to treat fungal infections, but when doses six times that used to treat fungi are used it, too, suppresses adrenal function and has been shown to be effective in the management of advanced prostate cancer. Aminoglutethimide is quite toxic causing rash and somnolence; hydrocortisone is usually administered with aminoglutethimide and this contributes additional toxicities. Ketoconozole at high doses is associated with serious hepatotoxicity. Because of these toxicities, drugs that specifically inhibit aromatase, such as letrozole, exemestane, and anastrozole, have been developed. These drugs are associated with minimal toxicity and have largely replaced aminoglutethimide.

Pituitary hormone agonists

Gonadotropin-releasing hormone (**GnRH**) analogs, such as leuprolide, goserelin, and buserelin, have greater affinity for the GnRH receptors in the testis and ovary than natural GnRH, and this may result in an exacerbation of symptoms, or tumor ‘flare', immediately following initiation of this treatment for prostate cancer. However, chronic administration results in blockade of both LH and FSH release from the pituitary. Men have castrate levels of testosterone within a month. Estradiol and

progesterone levels in premenopausal women will fall to those of postmenopausal women in 2 to 4 weeks. Although the GnRH agonists are used primarily for breast cancer in premenopausal women, they will induce a response in about 10 per cent of postmenopausal women. This may be due to some direct effect on breast cancer cells since luteinizing hormone-releasing hormone (**LHRH**) binding sites have been found on 52 to 67 per cent of primary breast cancers. Goserelin has demonstrated antitumor activity against an estrogen-dependent human breast cancer and a prostate cancer cell line. These drugs are usually administered intramuscularly or subcutaneously, and depot preparations that suppress the pituitary for periods up to 3 months are now available. GnRH antagonists that suppress pituitary function more rapidly with a longer duration of action and that do not cause an initial flare are under evaluation in clinical trials.

Adding hormones

At one time diethylstilbestrol was first-line therapy for metastatic prostate cancer and postmenopausal women with metastatic breast cancer. This has changed primarily because of the development of much less toxic alternatives. Today, only the synthetic progesterones or progestins are commonly used to treat cancer, but estrogens and androgens still remain as an important alternative in some patients.

Progestins

Progestins are used to treat breast, endometrial, and prostate cancers as well as the anorexia and cachexia associated with advanced cancer. The precise mechanism of action is not well established but probably involves suppression of GnRH from the pituitary. Frequently the progestins are classified as antiestrogens or antiandrogens. These drugs are well tolerated, but they stimulate appetite, cause weight gain, and, less commonly, cause fluid retention. As a result the progestins are considered more toxic than tamoxifen and the new aromatase inhibitors in the treatment of breast cancer and more toxic than the GnRH agonists and antiandrogens for prostate cancer. Although the drugs may occasionally cause hot flushes, they are probably the most effective and safest method for suppressing these symptoms in postmenopausal women with breast cancer. Both oral and intramuscular formulations are available.

Estrogens

When used to treat cancer, estrogens are given at doses 4 to 30 times those needed for hormonal replacement therapy. The mechanisms through which estrogens induce cancer cell death are not well understood, and it is difficult to mimic this effect *in vitro.* It is assumed that the beneficial anticancer effects are obtainable only with high doses. However, a randomized trial comparing diethylstilbestrol doses ranging from 1.5 to 1500 mg/day failed to show a significant difference in benefit. At higher doses, estrogens are quite toxic, and for this reason they are infrequently used any longer as first-line treatment. However, they should still be considered for patients who have previously responded and then progressed on other endocrine treatments. They are also considerably less expensive than any of the newer treatments.

Androgens

Like estrogens, androgens are infrequently used to treat breast cancer because of unacceptable side-effects. In addition, randomized comparisons with estrogens suggested they were less effective. However, they probably give the patient a greater sense of well being than any other therapy available for patients with the most advanced stages of breast cancer. When used with chemotherapy they counteract some of the bone marrow toxicity. They should be considered for patients with breast cancer who have previously responded and then progressed on other endocrine treatments.

Corticosteroids

Corticosteroids are lympholytic and play an important role in the treatment of leukemia and lymphomas. The combination of vincristine and prednisone will induce an initial complete remission of the most common forms of childhood acute lymphoblastic leukemia, and prednisone is an important element of the curative regimen (MOPP) used to treat Hodgkin's disease. Corticosteroids will also induce remission in some patients with breast cancer. More often, corticosteroids are used in combination with cytotoxics to reduce myelosuppression. These drugs are also an important element in many antiemetic programs.

Somatostatin analogs

Octreotide, a somatostatin analog, is effective in the management of tumors that secrete hormones, such as carcinoid and islet tumors of the pancreas. Forty to 75 per cent of primary breast cancers have somatostatin receptors, and octreotide has been shown to induce tumor regression in a small percentage of these patients, too. It is extremely well tolerated.

Flare

Endocrine therapy, and, less commonly, chemotherapy may cause an exacerbation of tumor signs and symptoms prior to a response. Classically this was described as bone pain and hypercalcemia that increased or appeared for the first time after treatment with estrogens or androgens. Skin lesions may show peritumoral erythema and swelling. Tumor markers, such as the CEA, may rise sharply. New lesions may appear on bone scans and the intensity of older lesions may increase prior to the appearance of recalcification of established lytic lesions in the bones of patients with breast cancer. Only rarely, however, has an unequivocal increase in the rate of measurable tumor growth been documented. Flare occurs commonly when GnRH agonists are used to treat prostate cancer, but this can be obviated by the concomitant administration of antiandrogens. Flare has also been described with tamoxifen, oophorectomy, and adrenalectomy. Flare is distinguished from tumor progression by its transient nature. Unless the symptoms of flare are life threatening, treatment should be continued and the symptoms of flare should be managed by other means until they subside or it is clear there is tumor progression.

Combinations of endocrine agents

In general, there is very little evidence that combining two or more forms of endocrine therapy provides greater benefit than using one. However, since the underlying paradigm for endocrine treatment is the elimination of one or more growth factors, intuitively it makes sense that a regimen that more completely ablates this endocrine function will be proportionately more effective. As a result, the development of each new endocrine treatment leads to new studies to determine if combinations that include the new therapy will be more effective than the old combinations. Recently, promising results have been reported combining goserelin or buserelin with tamoxifen for breast cancer and leuprolide or buserelin with flutamide (‘total androgen blockade') for prostate cancer. Although results from the many randomized trials of total androgen blockade are contradictory, total androgen blockade is widely used to treat prostate cancer.

Treatment of specific solid tumors

Lung cancer

It has long been recognized that local treatment alone has little impact on the survival of patients with small cell cancer of the lung, and for this reason chemotherapy has been used routinely to treat this form of lung cancer for more than three decades. It is much more responsive to chemotherapy than non-small cell lung cancer. While chemotherapy clearly prolongs the survival of this group of patients, recent advances have not been as rapid as once hoped. In non-small cell cancer, randomized trials comparing chemotherapy combinations with best supportive care have now demonstrated that appropriate use of chemotherapy will increase survival and palliate symptoms more effectively and less expensively than the best supportive care without chemotherapy.

Small cell lung cancer (Table 5)

Four of five randomized trials comparing chemotherapy with no chemotherapy demonstrated a favorable survival trend, and two demonstrated a statistically significant improvement in survival for patients treated with chemotherapy. It is assumed that similar studies using current regimens would produce much better results. The median survival of patients in the groups not given chemotherapy was about 5 to 11 months.

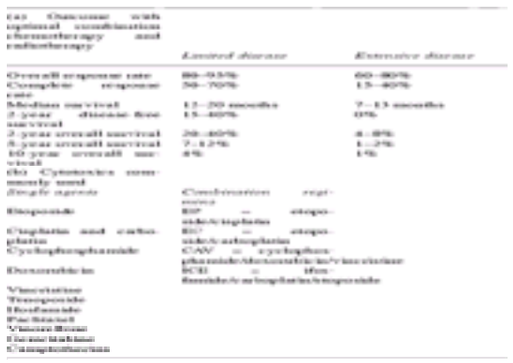


Table 5 Treatment of small cell lung cancer.

More than 15 different cytotoxic agents have been shown to induce regression of small cell lung cancer lesions. The most commonly used are etoposide, cisplatin or carboplatin, cyclophosphamide, doxorubicin, and vincristine. The taxanes, vinorelbine, gemcitabine, and camptothecins, such as topotecan and irinotecan, have only recently been tested in small cell lung cancer, and their impact on treatment outcomes for these patients remains to be seen.

Combinations of these drugs will induce a higher response rate and longer survival than single agents. The most commonly used regimens along with expected outcomes following treatment with combination chemotherapy are shown in Table 5. These survival times are substantially longer than those obtained before the chemotherapy era, but prolonged survival is still very uncommon for patients with small cell lung cancer. The outcome for patients with limited disease is much better than that for those with extensive disease. In elderly patients and those with a poor performance status, a single agent may be employed instead of one of the more toxic combination regimens. For example, in a recent Danish study the response rate to single-agent etoposide or tenoposide was 65 and 71 per cent, respectively, with complete response rates of 24 and 23 per cent.

So far it has not been possible to demonstrate an advantage for combinations that include more than four drugs. Alternating combination regimens, such as EP and CAV (Table 5), may increase the response rate and time to disease progression, but the impact of this strategy on survival has been minimal. The maximum survival benefit from these regimens is achieved after 6 to 9 months of therapy. Uncontrolled trials have suggested there may be an advantage to using higher doses of drug, but this principle has not been established in randomized clinical trials.

Non-small cell lung cancer (Table 6)

Although the median survival of patients with metastatic lung cancer remains dismal in spite of therapy, the consensus interpretation of eight randomized trials is that median survival is significantly increased by at least several weeks; in some trials the increase in median survival reached 5 to 6 months. Symptomatic relief from chemotherapy probably has more substantial clinical value, however. In a study from the United Kingdom, 69 per cent of patients had relief of tumor-related symptoms including reduction in malaise (53 per cent), pain (60 per cent), cough (66 per cent), and dyspnea (59 per cent). Formal assessment of overall quality of life has demonstrated net improvement in spite of the toxicities of treatment.

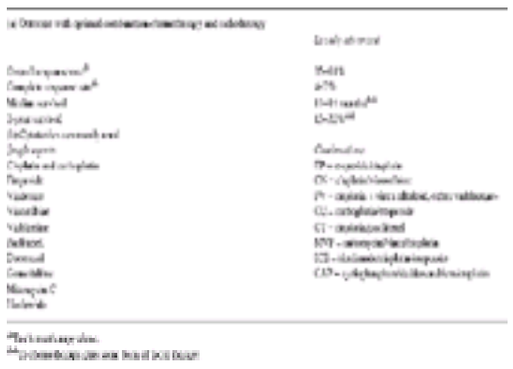


Table 6 Treatment of non-small cell lung cancer.

The introduction of cisplatin (and later, its less toxic congener, carboplatin) substantially increased the benefits from chemotherapy for lung cancer. All of the most popular regimens used today include cisplatin or carboplatin (Table 6). Randomized trials have demonstrated that drug combinations result in a 20 to 25 per cent improvement in 2-year survival compared with single agents. Most of the commonly used combinations are quite toxic. Most of the toxicity is contributed by cisplatin. Because of the cumulative neuro-, oto-, and renal toxicity from cisplatin, treatment is usually limited to six cycles. Younger patients with a low tumor burden and a good performance status are most likely to benefit from the more intensive regimens. For those who do not fall into this category it is reasonable to consider some of the most effective single agents with less toxicity, such as vinorelbine.

A combination of chemotherapy and radiotherapy should now be considered routine practice for locally advanced, unresectable lung cancers. A meta-analysis of 14 randomized trials demonstrated a 12, 13, and 17 per cent reduction in risk of death after 1, 2, and 3 years of follow-up. Among patients alive at the end of 30 months, the likelihood of death between the 31st and the 36th months was decreased by 46 per cent. The difference in median survival of those treated with chemotherapy and radiotherapy compared with radiotherapy alone was 2 months. (Practically, this means some patients may have a shorter survival as a result of chemotherapy while others have their lives prolonged by many months.) The greatest impact was seen among patients with the best prognosis.

It is likely that the survival advantages from the use of preoperative or postoperative adjuvant chemotherapy in patients with operable lung cancer are of a magnitude similar to that seen in patients with locally advanced unresectable disease, but the weight of evidence is still insufficient to consider this a standard therapy for these earlier stages of non-small cell lung cancer.

Breast cancer

Among the common solid tumors, breast cancer is the most responsive to systemic therapy. The tumor's estrogen receptor status is the single most important factor affecting treatment decisions in both the metastatic and adjuvant settings. Systemic therapy will substantially prolong survival when used as an adjuvant immediately after local treatment, will modestly increase survival when used immediately following the diagnosis of metastases, and may even impart a very small survival benefit in patients with extensive prior therapy and a limited life expectancy.

Metastatic breast cancer

Patients with very localized disease, such as an isolated local recurrence or single bone metastasis, and a long interval from diagnosis might reasonably be treated with surgery and/or radiotherapy alone. Although these patients almost always have other sites of metastases, breast cancer often runs an indolent course, and it may be many months or years before additional treatment for palliation of metastases is required.

Treatment with systemic therapy will increase the median survival by 6 to 12 months, and some patients derive years of additional life. However, there is as yet no convincing evidence that the impact of systemic treatment on survival will be greater if it is administered before the patient is SYmptomatic. For this reason, treatment is not recommended for asymptomatic patients with a rising serum marker, such as a CEA.

Patients who are positive for estrogen receptor, especially those with a high level or who are positive for progesterone receptor as well, should first be treated with endocrine therapy. Since a small percentage of patients will have a falsely negative receptor determination, it is also reasonable to try endocrine therapy in patients with the clinical characteristics of those with an endocrine-sensitive tumor. These include premenopausal patients in their late 40s or early 50s, patients more than 5

years after their last menses, and those with a long disease-free interval following diagnosis. There is little evidence that combinations of two or three endocrine therapies provide greater benefit than one endocrine therapy. Combinations of endocrine and chemotherapy induce higher response rates, but randomized trials have failed to demonstrate a survival advantage from this strategy and toxicity is substantially greater. For most receptor-positive patients, both survival and quality of life are maximized by the use of endocrine therapy as long as the patient continues to respond. When patients initially respond and subsequently progress while receiving an endocrine treatment, another should be employed. In general, there is little difference in the effectiveness of different endocrine treatments. Thus, the least toxic is used preferentially. A sequence of treatments based on these principles is outlined in Fig. 1, and expected outcomes from the use of these treatments as first-line treatment are shown in Table 7. Only 20 to 30 per cent of patients who respond to a first endocrine treatment will respond to a second form of hormonal therapy, but a rare patientwill go on to respond five times in sequence and tumor control may occasionally be maintained for periods of 5 to 10 years. Chemotherapy can be started when a patient fails to respond at all to an endocrine therapy.

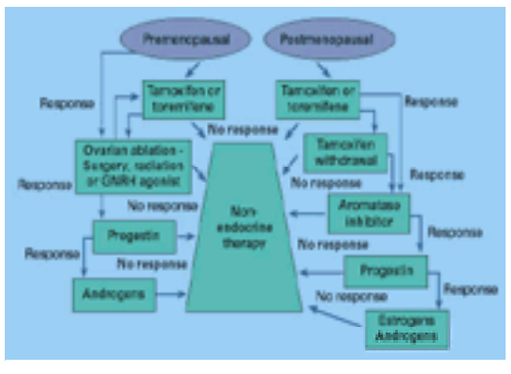


Fig. 1. Treatment sequences for the management of metastatic breast cancer. A patient who responds to one form of endocrine therapy and then progresses should be given a trial of another endocrine treatment until she no longer responds at all. Non-responding patients should be treated with non-endocrine therapy, usually chemotherapy. Among premenopausal women it is reasonable to use either tamoxifen or ovarian ablation with surgery, radiotherapy, or GNRH agonists as the first therapy with cross-over to the other on disease progression after a response. It may be necessary to wait for several months to see a response to tamoxifen withdrawal because of its long half-life. Many physicians and patients are unwilling to wait this long with no active treatment and go on to the next form of endocrine treatment upon progression after an initial response to tamoxifen.

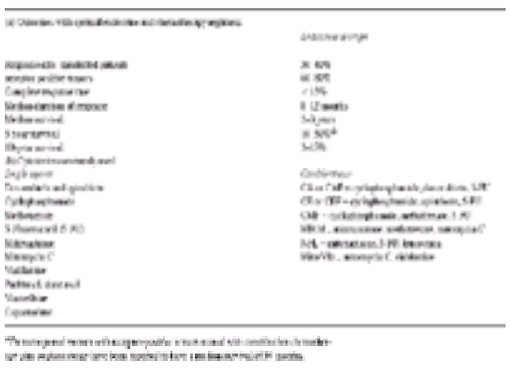


Table 7 Treatment of metastatic breast cancer.

Almost all of the cytotoxic drugs known to be effective in any solid tumor will induce regression of breast cancers. However, the drugs most widely used alone and in combinations are shown in Table 7. Doxorubicin, paclitaxel, and docetaxel are the most active agents. Multiple randomized trials have demonstrated that regimens that include doxorubicin induce a higher response rate and result in a small but statistically significant increase in median survival in the range of 3 to 6 months. Combinations that include both doxorubicin and paclitaxel have been developed, but none have yet been shown to impart an advantage over regimens without a taxane. The median time to response is between 2 and 3 months, but objective responses will occur as late as a year after treatment is initiated among patients who have stable disease at the end of 3 months. There is no convincing evidence that one regimen induces responses faster than any other (or faster than endocrine therapy). When choosing among regimens, the one with the highest response rate will usually be associated with the best quality of life even if it is more toxic. However, regimens such as NFL and MMM (Table 7) induce high response rates and are relatively non-toxic. Response rates and time to disease progression will increase with more prolonged treatment, but no survival advantages are obtained by treating for more than 6 to 9 months.

The response rate to second-, third-, or fourth-line regimens will be lower and the duration of response shorter than that to the first combination chemotherapy regimen used. However, it has been shown that a single drug, such as vinorelbine, administered to patients resistant to other drugs, including doxorubicin, may prolong life by a few weeks in the final stages of the disease. The taxanes are not cross-resistant to doxorubicin, and these drugs are now frequently used, even as single agents, in patients who have failed a doxorubicin regimen. A particularly effective and mild treatment for patients who have failed almost all other drugs, including bolus 5-fluorouracil, is continuous infusion 5-fluorouracil administered through a portable pump. Recently, capecitabine, a 5-fluorouracil congener that can be administered orally daily has been shown to be active in this setting.

Adjuvant systemic therapy for early breast cancer

The greatest survival benefits from systemic therapy for breast cancer are obtained when it is administered soon after diagnosis; this can be before, with, or immediately after local therapy. The first randomized trials evaluating adjuvant ovarian ablation were performed almost 50 years ago, prior to the discovery of the estrogen receptor. They showed that the odds of a woman having a recurrence of disease or dying were reduced about 25 per cent each year for as long as follow-up is available (Table 8). Presumably the benefits are greater in those patients who are estrogen receptor positive.

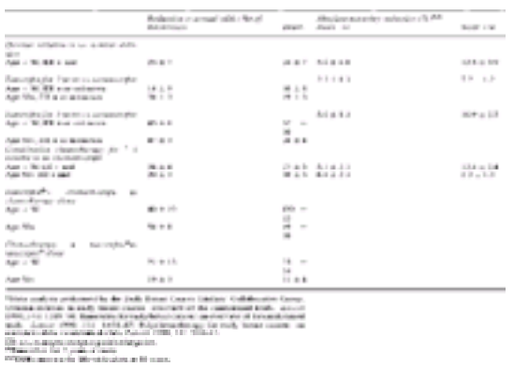


Table 8 Benefits from adjuvant therapy for patients with early breast cancer based on overviews or meta-analyses of all randomized trials performed worldwide*.

Five years of tamoxifen treatment will reduce the annual odds of recurrence by 45 per cent and the annual odds of death by 25 to 30 per cent in patients who are estrogen receptor positive. The most recent evidence indicates the effects of tamoxifen in younger and older women are about the same (Table 8).

Adjuvant chemotherapy is most beneficial for premenopausal women and for those whose tumors are estrogen receptor negative. Among women aged 50 and over, 90 per cent of whom are postmenopausal, the annual reduction in the odds of death while taking adjuvant chemotherapy is only 10 per cent plus or minus 3 per cent, a

value that is only marginally significant.

Adding tamoxifen to chemotherapy provides additional benefit over chemotherapy alone, especially in older women where the benefits of chemotherapy alone are very small. Among younger women the reduction in odds from adding tamoxifen to chemotherapy is not statistically significant, but this may be related to the small number of women studied in these trials and the duration of follow-up. Similarly, chemotherapy appears to provide benefits beyond those obtained with tamoxifen alone, but among younger women in these studies, most of whom were estrogen receptor positive, the added benefit from chemotherapy is not significant.

The reduction in annual odds of recurrence and death is the same among patients with node-negative and -positive disease. It follows, then, that patients with a higher risk of recurrence will have a higher absolute reduction in recurrence and mortality. For example, at 10 years the difference in the survival curves of those treated with 5 years of tamoxifen compared with those given no tamoxifen is 5.6 per cent among those with node-negative and 10.9 per cent among those with node-positive tumors. (Table 8)

No advantage has been demonstrated for continuing adjuvant tamoxifen for more than 5 years or adjuvant chemotherapy for more than 4 months. Most of the data shown in Table 8 are based on older chemotherapy regimens, such as CMF. It has been shown that doxorubicin regimens further reduce the annual odds of recurrence and death by about 12 per cent compared with non-doxorubicin combinations. The addition of paclitaxel may further reduce the odds of recurrence and death by 25 per cent or more compared with doxorubicin regimens.

Adjuvant tamoxifen reduces or delays the onset of a new breast cancer in the contralateral breast, slows or prevents the onset of osteoporosis, reduces serum cholesterol, increases the incidence of endometrial cancer, and may reduce non-cancer mortality. The increased risk of death from endometrial cancer is far outweighed by the reduced risk of death from other causes following tamoxifen therapy. The use of adjuvant therapy before definitive surgery and/or radiotherapy may enable more patients to have breast-conserving local treatment, but it has not been shown to improve survival.

Recommendations regarding the appropriate use of adjuvant treatment in an individual patient require careful consideration of size of the benefits for that patient relative to the toxicities the patient is likely to experience. The available data are sufficient to justify the following guidelines:

Postmenopausal women	
Receptor positive	Tamoxifen. Adding chemotherapy will provide a small additional benefit, but many patients will not feel this is large enough to justify the toxicity
Receptor negative	Chemotherapy will provide a small benefit, but many will consider this too small to justify the treatment even in node- positive patients
Premenopausal women	
Receptor positive	Either tamoxifen alone, ovarian ablation, or chemotherapy alone might be considered in these patients, especially when newer chemotherapy regimens that include doxorubicin and paclitaxel are used
Receptor negative	Chemotherapy, especially in node-positive patients with tumors greater than 1 cm in size

Gastrointestinal tumors

The tumors of the gastrointestinal tract are among the least responsive to systemic therapy. However, in the past few years new drugs have been identified that have at least as much activity as the long established standard for all gastrointestinal tumors, 5-fluorouracil, and it has been demonstrated that early treatment of colon cancer will prolong survival of patients with this disease.

Metastatic colorectal cancer

Since its introduction to the clinic in 1957, 5-fluorouracil has been the mainstay in the treatment of metastatic colorectal cancer. Reported response rates to single-agent therapy have ranged from 8 to 85 per cent. As a result it is difficult to determine the relative value of a new drug for this disease without a randomized trial. In general, higher response rates have been seen with the more dose- and schedule- intense regimens. Continuous intravenous infusions and hepatic artery infusions have each been shown in randomized clinical trials to induce a significantly higher response rate than bolus 5-fluorouracil alone. Neither have consistently been shown to improve survival in individual studies, but recent meta-analyses of all the randomized trials have demonstrated a survival advantage for each of these approaches compared with bolus 5-fluorouracil. The median survival of patients in these studies ranges from 11 to 17 months. Combinations of cytotoxic drugs have also not been repeatedly shown to have a significant advantage.

Leucovorin, a reduced folate, enhances the cytotoxicity of 5-fluorouracil, and randomized trials comparing 5-fluorouracil plus leucovorin with 5-fluorouracil alone have consistently demonstrated higher response rates when leucovorin is added. In several trials the combination resulted in a significant improvement in survival. Toxicity, which includes stomatitis and diarrhea, is also increased substantially.

Irinotecan has recently been registered for the treatment of colon cancer. The response rates are no greater than those seen with 5-fluorouracil, but randomized trials have demonstrated that irinotecan treatment will significantly improve both quality of life and survival of patients whose tumors have become refractory to 5-fluorouracil. Capecitabine, which is converted to 5-fluorouracil *in vivo*, has recently been reported to induce responses in 20 to 25 per cent of patients with metastatic colon cancer. It is registered for the treatment of breast cancer but will almost certainly find a role in the treatment of colon cancer, as well, since the pharmacokinetics of capecitabine mimic continuous infusion therapy.

Adjuvant chemotherapy for colorectal cancer

Two different approaches have been taken in administering adjuvant chemotherapy to patients with colorectal cancer, and both have been successful. The earliest adjuvant chemotherapy trials used single agents, including 5-fluorouracil and thiotepa. No single study demonstrated a survival advantage until a combination of 5-fluorouracil and levamisole was shown to impart a survival advantage over levamisole alone or observation in two separate studies. Results from the largest of these are summarized in Table 9. The contribution of levamisole, an antihelminthic with immunomodulatory properties, has always been a subject of controversy. The mechanism through which levamisole would augment the effects of 5-fluorouracil is not clear. The combination has no greater activity than 5-fluorouracil alone in patients with metastatic colon cancer, and levamisole itself has no activity in the adjuvant setting.

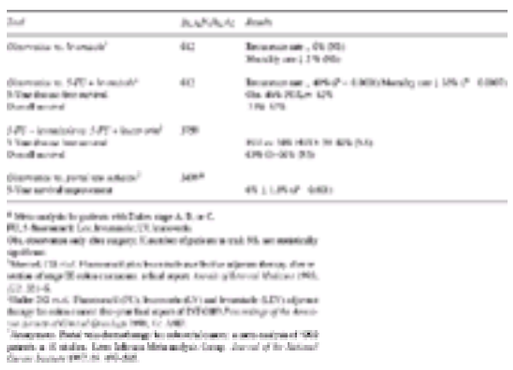


Table 9 Results of randomized trials evaluating adjuvant chemotherapy in patients with stages II and III colon cancer.

Mature results from a multicenter, randomized trial of 3759 have recently been reported (Table 9). Patients received one of four regimens: 12 months of 5-fluorouracil plus levamisole, or 7 to 8 months of either 5-fluorouracil plus low-dose leucovorin, 5-fluorouracil plus high-dose leucovorin, or 5-fluorouracil plus both low-dose leucovorin and levamisole. With the exception of a patient subset with stage III disease in one comparison, 5-fluorouracil plus leucovorin was slightly but not significantly better than 5-fluorouracil plus levamisole. There was no difference in outcome related to the dose of leucovorin used. 5-Fluorouracil plus leucovorin is considerably more toxic, but the leucovorin regimens were used for a shorter period of time.

Yet a third approach to adjuvant chemotherapy for colon cancer is the administration of continuous 5-fluorouracil through the portal vein for a period of 5 to 7 days

immediately following surgical resection of the primary tumor. A meta-analysis of all of the randomized trials designed to evaluate this approach demonstrated a statistically significant survival advantage of 3.6 to 6 per cent, depending on the patient group included in the analysis. This benefit is of a similar magnitude to that achieved with more prolonged courses of 5-fluoro-uracil with either levamisole or leucovorin.

Multiple randomized trials have now demonstrated that a combination of radiotherapy and chemotherapy will reduce local recurrences and improve overall survival of patients with Dukes B and C rectal cancers. The toxicities of these regimens do not outweigh the benefits, since the net time without symptoms of disease or toxicity from therapy (TWiST) was greater for the patients randomized to receive the combined treatment than for those given radiation only. The first results of a trial in which 1696 patients were randomized to receive bolus 5-fluorouracil, 5-fluorouracil with levamisole, 5-fluorouracil with leucovorin, or 5-fluorouracil with both levamisole and leucovorin have recently been reported. There were no differences between the arms, and the authors concluded that even with longer follow-up, no advantage was likely to develop for those randomized to receive levamisole.

Adjuvant chemotherapy of resectable colon cancer and combined radiotherapy plus chemotherapy for resectable rectal cancer prolong patient survival. Different regimens appear to give similar results.

Gastric cancer

The utility of chemotherapy for any group of patients with gastric cancer is frequently questioned. However, four randomized trials have compared a chemotherapy combination with best supportive care, and all have found that median survival is improved from about 3 to 4 months to 8 to 10 months. One of these formally evaluated quality of life and concluded that it was significantly improved by a combination of etoposide and 5-fluorouracil (with or without leucovorin). The median quality-adjusted survival was increased from 2 to 6 months (*P* = 0.03).

As with other gastrointestinal tumors, 5-fluorouracil, used both as a single agent and, more frequently, in a combination regimen, is the mainstay of therapy. Other single agents with nearly as much activity include mitomycin C, cisplatin, and the anthracyclines. None of these will reproducibly induce response in more than 20 per cent of patients. Response rates to combinations of chemotherapy range from 25 to 50 per cent. The most popular of these are 5-fluorouracil, doxorubicin, and mitomycin C (**FAM**); etoposide, doxorubicin, and cisplatin (**EAP**); 5-fluorouracil, doxorubicin, and methotrexate (**FAMTX**); etoposide, leucovorin, and 5-fluorouracil (**ELF**); and epirubicin, cisplatin, and continuous infusion 5-fluorouracil for up to 6 months (**ECF**). Each of these combinations has been shown to have high response rates, often in excess of 50 per cent, in phase II studies. Randomized comparisons have rarely shown a clear advantage. FAM was among the earliest of the combination regimens, and randomized comparison showed it produced no better result than single-agent 5-fluorouracil. Subsequent combinations appear to be superior to FAM, but the differences in response rate and survival are small. Most recently, ECF was compared with FAMTX in a large, multicenter trial in the United Kingdom. Response rates of 45 and 21 per cent (*P* = 0.0002) and median survivals of 8.9 and 5.7 months (*P* = 0.0009), respectively, were seen with ECF and FAMTX. Both regimens were quite toxic with mucositis, nausea/vomiting, alopecia, and three toxic deaths.

The results of uncontrolled trials suggest that preoperative or neoadjuvant chemotherapy will downstage tumors and lead to increased resectability in a substantial percentage of patients with locally advanced disease. Long-term survival has been reported for 20 to 25 per cent of these patients. The early results from recent randomized studies are mixed. The first European study compared neoadjuvant FAMTX with no preoperative chemotherapy and found no significant difference in resectability rates. Adjuvant chemotherapy for early gastric cancer should still be considered experimental.

Pancreatic cancer

In general, pancreatic cancer is among the most resistant tumors. Many drugs have some activity, but usually any gains are at substantial cost in unpleasant side-effects that compromise net quality of life. At least one study suggests that treatment of metastatic pancreatic cancer with combination chemotherapy results in a better survival (40 compared with 9 weeks, *P* = 0.0001) than best supportive care. A combination of postoperative radiotherapy plus 5-fluorouracil has been shown to improve median survival over postoperative radiotherapy alone by a few months in several trials, but this is still a question under study in randomized trials.

Gemcitabine has recently been shown to be superior to 5-fluorouracil in patients with unresectable or metastatic pancreatic cancer. Response rates were 24 and 5 per cent, respectively, (*P* = 0.0022), and patients on gemcitabine had a ‘net clinical benefit' from reduction in symptoms such as weight loss, pain, and overall performance. Survival was also modestly improved from 4.4 months with 5-fluorouracil to 5.7 months (*P* = 0.0025). Gemcitabine will induce a response in about 10 per cent of patients whose tumors are refractory to 5-fluorouracil. Because of these high response rates and its relatively favorable toxicity profile, gemcitabine has rapidly been accepted as the treatment of choice for pancreatic cancer in spite of the modest overall benefits from treatment.

Novel therapies currently under evaluation include LY231514 which inhibits thymidylate synthase, dihydrofolate reductase, and glycinamide ribonucleotide formyltransferase; marimastat, an oral metalloproteinase inhibitor; octreotide; and farnesyl transferase inhibitors which are particularly promising for pancreatic cancers because most of them have K-ras mutations.

Prostate cancer

Prostate cancers are a mixture of androgen-dependent cells that require androgen to survive, androgen- sensitive cells that are not dependent on androgen but whose growth rates may be modified by endocrine treatments, and androgen-independent cells. Some form of endocrine manipulation is the first treatment of choice for the vast majority of patients with metastatic prostate cancer, and 60 to 80 per cent of these patients will undergo a clinical remission. The median duration of response is about 18 months. However, all patients will eventually relapse. Although orchidectomy remains the ‘gold standard', patients increasingly choose medical treatments, such as GnRH agonists and antiandrogens, because of their favorable toxicity profiles. A combination of a GnRH agonist and an antiandrogen is often used to achieve ‘complete androgen blockade', but it is not certain that this improves overall survival compared with sequential use of these therapies. A meta-analysis of 22 randomized trials demonstrated only a 6.4 per cent (± 3.9 per cent) reduction in the annual odds of death. The 5-year survival of those randomized to castration (either orchidectomy or GnRH agonist) plus an antiandrogen was 26.2 per cent compared with 22.8 per cent for castration alone; the difference of 3.5 per cent (± 1.9 per cent) was not statistically significant. Possibly these differences will become significant with longer follow-up. A European Organization for Research and Treatment of Cancer (EORTC) trial was amongst the largest of those included in the meta-analysis. The most recently reported results from that study demonstrated a significant improvement in duration of survival (*P* = 0.04), time to death from prostate cancer (*P* = 0.008), time to first progression (*P* = 0.009), and progression-free survival (*P* = 0.02) for those randomized to receive a combination of goserelin acetate and flutamide compared with orchidectomy alone.

About 20 to 25 per cent of patients who relapse from first-line therapy will respond by merely withdrawal of the hormone therapy whether complete androgen blockage, non-steroidal or steroidal antiandrogens, or estrogens. The reason why this occurs is not well understood. Patients with castrate levels of testosterone may also respond to second-line endocrine treatments. After GnRH agonists and antiestrogens, these may include drugs that suppress adrenal function, such as aminoglutethimide or ketoconazole, progestins, or estrogens.

Elderly patients with prostate cancer do not tolerate chemotherapy well, and most cytotoxics are not very active in this disease. However, comparisons with other tumors are confounded by the difficulty in measuring response in osteoblastic bone lesions, the predominant site of prostatic metastases. Based on objective tumor response, doxorubicin, mitoxantrone, and estramustine (a combination of estrogen and nitrogen mustard) are the most active agents. Paclitaxel and vinorelbine have also been shown to be active using prostate specific antigen (PSA) values or palliation as an end point. Combinations of these drugs with ketoconazole, cyclophosphamide, or a corticosteroid will produce responses in 35 to 60 per cent of patients. In a randomized comparison of mitoxantrone plus prednisone with prednisone alone, there was no differences in survival, but the palliative responses were 29 and 12 per cent, respectively (*P* = 0.01). The durations of benefit were 43 and 18 weeks (*P*< 0.0001). Bone pain may also be relieved in up to 75 per cent of patients with the administration of strontium-89.

Neoadjuvant and adjuvant endocrine treatments are under evaluation and are often used for patients with earlier stages of prostate cancer, especially those with stage C disease. Several uncontrolled studies have demonstrated a lower incidence of positive margins at surgery in patients who have been given preoperative endocrine treatments. However, neither of these approaches has been shown in properly controlled trials to improve patient survival.

Testicular cancer

Testicular cancers, especially germ cell malignancies, are more sensitive to chemotherapy than any other solid tumors. The cure rate for metastatic germ cell cancers is 80 per cent. The first successful combination of drugs employed cisplatin, vinblastine, and bleomycin for six or more cycles. In subsequent studies etoposide was substituted for bleomycin, and it has been shown in randomized trials that shorter courses consisting of three to four cycles will achieve optimal results. It has also been demonstrated in randomized studies that carboplatin is inferior to cisplatin in the management of testicular cancer.

There is controversy regarding the treatment of stage I testicular cancer. With such a high cure rate after metastases are established, it is difficult to justify the use of adjuvant chemotherapy. About 70 per cent of patients with seminoma and 50 per cent of non-seminomatous germ cell tumors will be in this category at diagnosis. Both

careful surveillance and adjuvant therapy are recommended by various experts.

Ovarian cancer

It has also been demonstrated that treatment of both early and advanced stages of ovarian cancer will improve overall survival. The treatment of choice is a combination of cisplatin and paclitaxel. This combination has been shown in large phase III studies to impart better survival than older combinations employing cisplatin and cyclophosphamide. However, the number of drugs that induce responses in patients with ovarian cancer is large (Table 4a, Table 4b), and substantial palliation can be achieved from many of these drugs even when patients have relapsed from their primary drug treatment.

Evolving systemic therapies

Resistance to cytotoxic therapies

In most cases, resistance is multifactorial and multidimensional. This makes it difficult either to design methods for circumventing resistance or even to demonstrate that a proposed mechanism is one important link in a chain.

Drug dose and duration of exposure are important factors for almost all drugs, and inadequate dosage heads the list of potential reasons for the failure of antineoplastic agents (Table 10). Normal tissue toxicity is usually dose limiting for cytotoxic agents, but substantial progress has been made in circumventing this problem with the use of various cytokines (see below). Growing tumor deposits develop necrotic centers. It is difficult to deliver an adequate drug dose to these relatively avascular areas, and many drugs are less effective in this acidic milieu. Tumor cells in these areas proliferate slowly, and for this reason, too, they will be less sensitive to the effects of chemotherapy. Increased oxygenation of these areas could reverse this form of resistance, and this is an ongoing area of research and technologic innovation.

Pharmacologic factors
Delivery of inadequate drug dose for insufficient duration
Tumor sanctuaries: necrotic core, brain, testes
Altered drug transport
Multidrug resistance—MDR1/Pgp
Multidrug resistance-associated protein (MRP)
Lung-resistance protein (LRP)
Intracellular detoxification or deactivation
Glutathione
Reduced intracellular activation
Altered drug target (reduced affinity for target enzymes)
Increase in amount of target
Reduced affinity for target
Increased repair capacity
Alteration of apoptosis pathways

Table 10 Possible mechanisms of resistance to cytotoxic therapies.

Decreased influx of drug across the cell membrane or increased efflux out of the cell represent another source of resistance. Analogs of drugs that more readily enter the cell and that are retained by the cell have been shown, at least in preclinical studies, to be more effective. 10-EDAM, an analog of methotrexate, is one example. Not only does 10-EDAM enter the cell more readily than methotrexate, but it undergoes greater polyglutamation after entry. As a result, there is less efflux. Classic multidrug resistance is another example of altered drug efflux, and this has probably been studied more intensely than other forms of cytotoxic drug resistance.

Multidrug resistance (MDR)

Cells made resistant to single cytotoxic drugs such as doxorubicin or vinblastine *in vitro* frequently exhibit cross-resistance to a range of structurally unrelated cytotoxic drugs derived from natural products. This form of resistance is associated with the presence of a transmembrane protein, p-glycoprotein (**Pgp**), which functions as a drug efflux pump. The gene encoding p-glycoprotein, *mdr1*, has been cloned. Transfection of the *mdr1* gene into recipient cells leads to expression of the resistance phenotype, indicating that this resistance is associated with overexpression of a single gene. Drugs that serve as Pgp substrates are generally lipophilic and enter the cell by passive diffusion. In addition to the anthracyclines and the vinca alkaloids, other natural products in this category include the taxanes, anthraquinones, epipodophyllotoxins, dactinomycin, and to a lesser extent topotecan and mitomycin C.

p-Glycoprotein is found on the luminal surface of the bile canaliculi in the liver, the proximal renal tubule, the apical surface of the colon and small intestine, and capillaries of the adrenal gland and brain. High levels are found in CD34+ hematopoietic stem cells and normal peripheral blood mononuclear cells. In these tissues it is thought to be involved in the excretion of xenobiotics.

High levels of mdr1 mRNA have been found in tumors derived from tissues that normally express high levels of mdr1 mRNA. These tumors are highly resistant to cytotoxic chemotherapy at presentation suggesting a potential role for p-glycoprotein in intrinsic drug resistance. Cancers such as acute lymphoblastic leukemia and myeloma, in contrast, have low levels of mdr1 RNA at presentation and are relatively sensitive to cytotoxic drugs. On relapse, mdr1 levels are increased, and drug resistance is frequently observed, indicating a possible role for this gene in acquired drug resistance.

Solid tumors that express MDR/Pgp have a worse prognosis than those that do not, but the evidence that this is related to drug resistance is much less certain. A recent meta-analysis demonstrated that over 40 per cent of all breast cancers expressed MDR1/Pgp and 57 per cent of those treated with one or more of the MDR1 drugs. Patients with MDR1-positive tumors were four times more likely to fail to respond to treatment with one of the MDR1 drugs.

A number of agents have been found to inhibit Pgp *in vitro* and in preclinical tumor models. These include calcium channel blockers, such as verapamil and nifedipine, drugs that inhibit organ transplant rejection such as cyclosporine, the cyclosporine analog PSC-833, quinidine, quinine, tamoxifen, and progesterone. Several of these drugs have also been evaluated in patients with tumors that express MDR1/Pgp. While a number of uncontrolled trials have provided promising results, the randomized trials conducted to confirm these initial observations have generally been negative. The first agents tested in the clinic substantially increased the toxicity of the chemotherapy, as well. Less toxic and more potent inhibitors of Pgp are currently being evaluated.

Intracellular detoxification mechanisms

Cytotoxic drugs and their metabolic products are frequently highly reactive, and there are many natural mechanisms within the cell designed to detoxify reactive molecules. Glutathione, a tripeptide present in millimolar concentrations in all cells of the body, protects the cell against reactive molecules with a free sulfhydryl group on the cysteine and various free radicals. Glutathione forms conjugates with many cytotoxic drugs, especially alkylating agents and cisplatin. This process is catalyzed by glutathione transferases. Elevated glutathione and glutathione transferase (**GST**) levels have been associated with resistance to cytotoxic drugs both *in vitro* and *in vivo.* Depleting intracellular glutathione levels has been shown to enhance sensitivity to a range of cytotoxic drugs.

Buthionine sulfoximine is a relatively non-toxic inhibitor of one of the key enzymes involved in the formation of glutathione. Phase I clinical trials have demonstrated that it will reduce intracellular glutathione levels in peripheral blood mononuclear cells, and anecdotal clinical evidence suggests that a combination of buthionine sulfoximine and an alkylating agent may be effective in patients who have become resistant to other chemotherapy. Whey protein has also been shown to inhibit glutathione synthesis, and this, too, is under evaluation in the clinic. GST is also a therapeutic target. Ethacrynic acid inhibits GST, and more selective inhibitors are under development.

The affinity of glutathione for heavy metals has been applied in a quite different way to reduce the toxicity, especially the neurotoxicity and nephrotoxicity, of cisplatin. Concomitant administration of these two agents has been demonstrated in randomized trials to be as effective and much less toxic than cisplatin alone.

Some drugs, such as neocarcinostatin, are inactive in their parental form and are activated by glutathione reduction. Since tumor cells generally have higher glutathione levels than normal tissue, this difference could lead to the development of agents that would be activated only in tumor cells or would lead to much higher levels of the active metabolite in tumor cells. Of course, one would anticipate that tumors with low levels of glutathione would be relatively more resistant to these

drugs.

Changes in drug target

Many cytotoxic drugs have more than one target. Resistance can occur at more than one step during drug uptake, binding to target(s), and along the chain of events leading from drug binding to cellular growth arrest. An excess number of binding sites, changes in the structure of the target that result in less drug binding, or alternative pathways that circumvent the need for the inhibited target have all been shown to be important in resistance to one or another cytotoxic agent.

The mechanisms of resistance to the antimetabolites, especially methotrexate, are among the best understood. Methotrexate blocks cellular division by inhibiting dihydrofolate reductase (**DHFR**), an important enzyme for purine synthesis. Methotrexate resistance may result from a defective folate transport SYstem; reduced polyglutamation immediately after transport into the cell which, in turn, allows the drug to leave the cell easily before binding with DHFR; and reduced affinity of methotrexate for its target, DHFR. In the clinic, however, resistance has been shown to result primarily from an excess amount of DHFR because of amplification of the DHFR gene. Increasing the dose of methotrexate can circumvent this resistance by saturating the excess DHFR. Normal tissues can be rescued from the toxic effects of high-dose methotrexate by the administration of leucovorin 24 h after methotrexate. Early studies demonstrated that this was a safe therapy and suggested that high-dose methotrexate with leucovorin rescue was important in the treatment of osteogenic sarcoma, but no controlled trial has yet demonstrated an unequivocal clinical benefit from the use of this strategy for any tumor type.

Another antimetabolite, 5-fluorouracil, targets the enzyme thymidylate SYnthase through an intermediate, 5'FdUMP. The tightness of the binding between 5'FdUMP and the enzyme is affected by the intracellular concentrations of reduced folates. In this case concomitant administration of leucovorin, a reduced folate, with 5-fluorouracil has been shown in well controlled trials to increase substantially the value of 5-fluorouracil in the treatment of colon cancer.

Resistance to topoisomerase II inhibitors, such as doxorubicin and etoposide, develops rapidly. The majority of topoisomerase inhibitors used in clinical practice are transported by p-glycoprotein. However, other mechanisms of resistance to topoisomerase II inhibitors have been demonstrated *in vitro.* These include decreased enzyme activity due to modification of gene expression through hypomethylation of DNA, gene mutations that result in decreased topoisomerase II mRNA expression, and reduced specific activity of the enzyme itself. So far these forms of resistance to topoisomerase II inhibitors have not been shown to be clinically important, but it is plausible that they will become important in the future if and when other forms of resistance to these agents, such as multidrug resistance, are circumvented.

DNA repair

DNA damage that prohibits mitosis and leads to cell death is the goal of many cytotoxic agents, especially the alkylating agents and cisplatin. There is considerable preclinical evidence that increased repair of this damage may be a source of resistance to these agents (as well as to radiation). In the clinic, overexpression of O^6-alkylguanine-DNA alkyltransferase (**AT**), a DNA repair enzyme, has been associated with resistance to the nitrosoureas such as BCNU. Theoretically depletion of AT should lead to increased sensitivity to BCNU. One such agent is streptozocin, and it has been shown in preclinical systems that use of streptozocin before BCNU will enhance tumor response. In the clinic, streptozocin will decrease AT levels in peripheral blood mononuclear cells. More potent and less toxic inhibitors of AT have been developed, and one of these, O-g-benzylguanine, is now being evaluated in the clinic.

Resistance to cisplatin is multifactorial: reduced cellular accumulation of the drug is important in some models, and enhanced cellular detoxification by glutathione or metallothioneins is important in others. However, in some models resistance to cisplatin is associated with enhanced intrastrand cross-link repair. In these model systems it can be shown that agents which inhibit DNA repair are synergistic with cisplatin. These include metoclopramide, nicotinamide, aphidicholin, cytarabine, caffeine, and hydroxyurea. Clinical trials combining one or more of these agents with cisplatin suggest synergy, but no controlled trials have yet proven that this is a viable strategy for increasing cisplatin efficacy.

Resistance to endocrine treatments

The mechanisms of resistance to endocrine therapies are puzzling in part because tumors will respond sequentially to both the addition and withdrawal of the same hormone. The presence of an estrogen receptor is a primary requirement for all forms of endocrine treatment, but it is likely that after that entry into the cell different forms of endocrine treatment affect cell growth through a variety of different mechanisms. Although a small percentage of tumors that are estrogen receptor positive when hormone therapy is initiated become estrogen receptor negative at relapse, the vast majority retain their estrogen receptor positivity. Tamoxifen may even begin to stimulate tumor growth after an initial regression, and presumably this effect is mediated through an intact receptor. In some cases, resistance may result from a variant or mutant estrogen receptor. This is a plausible explanation for tumors that are estrogen receptor positive but fail to respond initially or after many different types of endocrine treatment. Altered expression of receptor-interacting proteins and cross-talk among growth factor signaling pathways are other mechanisms that have been proposed to explain resistance to tamoxifen. Antiestrogens without any agonist activity have been developed, and trials are underway to determine if these agents will be effective against tumors that have become resistant to tamoxifen.

The many mechanisms by which tumors may become resistant to systemic chemotherapy at the cellular level are poorly understood at present. In many, if not most, instances resistance is related to gene mutations. These occur spontaneously as cells divide, and it is hypothesized that they occur at a relatively constant rate in tumors that are not perturbed. Early treatment of any tumor before mutations associated with drug resistance have occurred is likely to yield the best results. However, treatment itself is often mutagenic, and in tumor models it can be shown that a low dose of chemotherapy insufficient to eradicate the tumor quickly will induce resistance, when a higher initial dose of therapy is curative. When survival is an important goal of treatment, optimal doses should be given immediately. Animal models suggest also that when palliation is the goal of treatment, the duration of therapy may be more important than dose of a single course.

High-dose chemotherapy

Until recently, myelosuppression was dose limiting for most cytotoxic drugs. That changed with the development of bone marrow transplantation, and the introduction of granulocyte colony-stimulating factor (**G-CSF** or filgrastim) and granulocyte–macrophage colony-stimulating factor (**GM-CSF** or sargramostim). In animal tumor models, cure can often be achieved by increasing the dose of some anticancer drugs, especially the alkylating agents. In the clinic, a reduction of dose will compromise survival benefits that might accrue from a treatment. It makes sense that survival of patients with cancer might be increased with high-dose chemotherapy supported by either a growth factor or bone marrow transplantation.

Proof of principle for the high-dose concept was obtained when it was demonstrated that allogeneic bone marrow transplantation after marrow ablation with high-dose chemotherapy and total body irradiation resulted in long-term survival or cure in more than 50 per cent of patients with non-lymphoblastic leukemia. Previously this was an incurable cancer with only a small percentage of patients living beyond 2 years.

Autologous bone marrow transplantation (**ABMT**) or infusion of peripheral blood stem cells (**PBSC**, also termed peripheral blood progenitor cells) has been used to test the high-dose concept in other hematologic and non-hematologic malignancies. When ABMT is used, the patient's own marrow is removed before and reinfused after a marrow ablative dose of chemotherapy. Generally this requires general anesthesia. PBSCs are harvested after mobilization of leukocytes with a course of chemotherapy, the administration of G- or GM-CSF, or both. PBSC support is almost always used with ABMT, but ABMT and PBSC infusion have been shown to be equally effective in reconstituting the marrow. Since PBSC support does not require general anesthesia, it has gradually replaced ABMT in many centers.

ABMT and/or PBSC infusion following very high-dose chemotherapy are accepted treatments for lymphomas refractory to conventional-dose chemotherapy and multiple myeloma. Although most of the evidence supporting the use of very high-dose chemotherapy in these diseases is uncontrolled, there is at least one credible randomized trial in each of these diseases demonstrating a survival benefit from high-dose compared with conventional-dose chemotherapy.

Among the solid tumors, the best evidence of benefit from high-dose chemotherapy is derived from studies in patients with testicular cancers. Several randomized trials suggest there may be an advantage for higher doses of cisplatin short of those requiring bone marrow support, but the results from various studies are not consistent. More compelling are studies in patients with testicular cancer who have become refractory to all conventional chemotherapy. Following very high-dose chemotherapy and ABMT, complete response rates of 30 to 40 per cent and long-term survival rates of 15 to 20 per cent have been reported.

The solid tumor most frequently studied and treated with high-dose chemotherapy outside the context of a study is breast cancer. Most of the data supporting this practice come from studies that demonstrate increased complete response rates and longer survival compared with patients treated with conventional chemotherapy in historical controls. However, there is probably a selection bias in these studies. Retrospective analyses of patient cohorts treated with conventional-dose chemotherapy have shown that those who would have been eligible for the very high-dose regimens had higher response rates and longer survival than those who would have been excluded from ABMT trials. Very large, randomized trials comparing high doses of either cyclophosphamide or doxorubicin with more conventional doses of these drugs have failed to demonstrate a significant prolongation of either the time to recurrence or survival. In these studies drugs were escalated only to the extent possible with G-CSF alone. Four randomized trials comparing very high-dose chemotherapy with ABMT and/or PBSC support with conventional-dose chemotherapy in patients with early breast cancer and 10 or more positive lymph nodes failed to demonstrate an advantage for dose escalation. One trial in patients with metastatic disease

showed a substantial and significant improvement in survival from very high-dose chemotherapy and ABMT, but the median survival of the high-dose group was not much longer than most other centers obtain with conventional-dose chemotherapy. Three other studies, including a multicenter trial with nearly 200 patients, failed to demonstrate a survival benefit or showed only a transient benefit for high dose chemotherapy and ABMT/PBSC in patients with metastatic breast cancer. The mature results of these studies, now fully accrued, should determine future use of this approach in the management of solid tumors.

High-dose chemotherapy and ABMT have also been used to treat ovarian and small cell lung cancer. Although promising, the data are not sufficiently compelling to justify this as a routine practice.

From the evidence generated to date, there is a general consensus that tumors primarily resistant to conventional-dose chemotherapy are unlikely to benefit from dose escalation. With the exception of leukemias, lymphomas, and testicular cancer, there is no evidence of benefit from very high-dose chemotherapy and ABMT in patients with end-stage disease who have failed multiple prior chemotherapy regimens.

Immunotherapy

Patients with a variety of immunodeficiency diseases have an increased risk of developing cancer. AIDS is just one example. Virally induced and transplanted cancers in animals can be eradicated by immunotherapy. These observations have invariably led to a search for natural immune responses that might explain why so few patients exposed to known carcinogens actually develop cancer and might, in turn, be used as a therapeutic strategy. At each stage in the development of immunotherapy there have been modest successes that have propelled the field forward. Progress has also been characterized by increasing specificity of the immunotherapy for unique antigens on the tumor cell (Table 11). None the less, the applicability of this approach is still relatively limited, especially for the management of solid tumors.

Non-specific immunotherapy—BCG/MER, levamisole
Cytokines: interleukins, interferons, tumor necrosis factor, and colony-stimulating factors
Passive immunization
Antibodies as therapy
Antibodies as vehicles
Active immunotherapy
Adoptive immunotherapy
Vaccines

Table 11 Approaches to immunotherapy.

Bacille Calmette–Guérin (**BCG**) is commonly used as a vaccine for tuberculosis. BCG or modifications of BCG, such as MER (methanol extracted residue of BCG), were once widely used alone or with chemotherapy to treat melanoma, lung, and breast cancer as well as other solid tumors. Randomized trials have demonstrated the futility of this approach. Direct injection of BCG into tumors, such as melanoma, will cause tumor regression. Instillation of BCG into the bladder stimulates a local inflammatory response, activates immune cells, and causes regression of visible tumors. BCG is commonly used as an alternative to instillation of cytotoxics, such as thiotepa, after resection of superficial bladder cancers to delay or prevent the appearance of new cancers. In addition to local symptoms from the inflammation, patients will infrequently develop tuberculosis from this treatment.

Levamisole, another form of non-specific immunotherapy, has been shown to improve survival when used in conjunction with 5-fluorouracil as adjuvant therapy for patients with Dukes B colon cancer (see above).

Cytokines are soluble proteins produced by immune cells. They facilitate intercellular communication. When cytokines bind to an appropriate receptor, usually on the cell membrane, they activate signal transduction pathways that eventually cause cell division, the production of antibodies, and/or the modulation of cellular behavior. Cytokine production commonly occurs after viral and bacterial infections or after a stimulus from other foreign substances. Interleukins, interferons, tumor necrosis factor, and colony-stimulating factors such as G- and GM-CSF are all cytokines. The cells that produce and respond to a particular type of cytokine are well defined. For example, the primary targets for filgrastim (G-CSF) and sargramostim (GM-CSF) are cells in the granulocytic line. Lymphocytes, especially T cells, most often have receptors for interleukins. Cytokines can function as autocrine or paracrine growth factors.

The mechanisms through which cytokines effect an antitumor response is not always clear. For example, interferons are secreted by virally infected cells and protect non-infected cells from viral infection. They inhibit angiogenesis (see below). They modulate the behavior of immune cells, but they also have a direct antiproliferative effect on some tumor cells. Interferons are employed as first-line treatment for hairy cell leukemia, cutaneous T-cell lymphoma, and Kaposi's sarcoma. They are very active against chronic myelogenous leukemia. a-Interferon has been shown to increase time to recurrence and overall survival when used as an adjuvant to surgery in patients with melanoma with regional lymph node involvement or primary lesions with a depth greater than 4 mm. Responses have been reported from the use of interferon to treat a variety of other tumors, notably renal cell and head and neck cancers, but this has not prolonged survival of these patients. Instillation of interferon is effective treatment for superficial bladder cancers. The primary toxicity from the systemic administration of interferon is flu-like symptoms with aching, fever, and chills.

Renal cell cancers and melanomas are notoriously resistant to the effects of cytotoxic treatment, and in this context the response rates seen with interleukin-2 have been impressive. Fifteen to 25 per cent of patients with renal cell cancer will have an objective response to bolus or continuous infusion therapy. Subcutaneous treatment may be equally effective. The pooled response rate to interleukin-2 in patients with metastatic melanoma is 15 per cent (12 to 27 per cent). Combinations of interleukin and other cytokines, such as tumor necrosis factor, may be more active than one cytokine alone. In the initial trials very high doses with considerable toxicity were used. Lower, less toxic doses may be equally effective, but there are no direct comparisons. The main toxicity is hypotension and vascular leak syndrome.

Immunologists have long searched for antigens that are unique to tumor cells and not found at all on normal cells. So far they have not succeeded. None the less, normal antigens that are overexpressed on the tumor cell have been successfully used for immunotherapy of hematologic malignancies. In some cases the tumor cell appears more responsive to the effects of antibody binding to antigen on its surface than the normal cell does. Passive immunization is the administration of antibodies to the patient. Active immunotherapy is stimulation of the patient's own immune response to the tumor (Table 12). Both have recently been used with some success.

Target	Example
Tumor suppressor genes	Gene therapy to replace p53, retinoblastoma (Rb), and inactivation of other mutated p53
Tumor suppressor gene functions	Antisense therapy of BCL-2, inhibitors of ICE
Oncogenes	
Activated growth factors	Inhibitors of membrane bound ras (e.g., FTI) and intracellular kinases (e.g., Src, Abl, and cyclin dependent kinases)
Activated growth factor receptors on cell surface	Inhibitors of HER2/neu or of EGFR
Signal transduction pathways	Ras antagonists, farnesyl transferase inhibitors
Nuclear transcription factors	Down-regulate overexpression of myc, jun, and/or fos with steroids, tamoxifen
DNA repair genes	Gene therapy to restore ATM, BRCA, PMS2, PMS, functions
Nuclear receptors	Retinoids
Proteases involved in tumor cell invasion and metastases	Matrix metalloproteinase inhibitors, marimastat, ATA744
Angiogenesis	Endostatin, angiostatin, TNP-470
Telomeres	Telomerase inhibitors

Table 12 New anticancer targets and possible therapeutic approaches.

This field has profited by several key technologic advances. First, monoclonal antibodies that bind to only one epitope of a particular human antigen can be produced in large quantities in mouse cells *in vitro.* However, since these are murine proteins, their injection into humans would stimulate a human antibody to the murine antigen (HAMA). To circumvent this problem, monoclonal antibodies are humanized. Almost all of the murine protein is replaced with human immunoglobulins. Only the

minimal amount of murine antibody needed to recognize the antigen is retrained. Antibodies produced in this way can then be used as therapy or further modified by conjugation of the antibody to a toxin. These toxins include radioactive isotopes, potent poisons such as ricin, enzymes, and cytotoxic drugs.

One antigen, CD20, has been successfully targeted using antibody as therapy and as a vehicle for iodine-131. CD20 is on normal B cells and on almost all B-cell non-Hodgkin's lymphomas. It is very specific for this cellular lineage, is expressed in high copy number on lymphoma cells, does not internalize, and has not been known to mutate. Rituximab (Rituxan®) was the first therapeutic monoclonal antibody approved by the FDA for any disease. After rituximab binds to the CD20 on its surface, the lymphoma cell is more easily destroyed by complement in conjunction with antibody-dependent, cell-mediated cytotoxicity effector cells. Up to 50 per cent of patients have been reported to respond to rituximab in studies of patients with relapsed or refractory low-grade or follicular non-Hodgkin's lymphoma. Combinations of rituximab and conventional chemotherapy regimens, such as cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) are being evaluated in ongoing studies. Anti-CD20 bound to iodine-131 (Bexxar®) has also been shown to be very active—possibly more active than rituximab, and it is anticipated that it will also have regulatory approval very soon. Response rates of 80 to 100 per cent with complete response rates in excess of 50 per cent have been observed in small, phase II studies. The median duration of response after a single injection may be as long as 1 year. The side-effects associated with both of these forms of anti-CD20 treatment are mild and consist primarily of flu-like SYmptoms.

These demonstrations of the clinical value of antibodies to CD20 provide proof of principle that might eventually be applied to the treatment of solid tumors, but that achievement has so far been very elusive.

Active immunotherapy has been more effective against solid tumors. The target is the T cell. Tumor antigens will activate T-cell subsets that selectively bind to and lyse tumor cells. In early studies in the development of adoptive immunotherapy, patients were first treated with interleukin-2 to generate ‘lymphokine activated killer' or LAK cells *in vivo.* These cell populations were then expanded *ex vivo* and given back to the patient with more IL-2. Randomized trials failed to demonstrate an advantage to this approach. More recently this concept has been expanded by the development of ‘tumor infiltrating lymphocytes'. The patient's lymphocytes are exposed to both interleukin and tumor antigens before reinfusion. This approach is still undergoing clinical evaluation in small pilot studies.

Cancer vaccines differ from traditional vaccines in several important ways. Traditional vaccines stimulate the expansion of clones of antibody producing lymphocytes to prevent infections. Cancer vaccines sensitize and stimulate the expansion of lymphocyte killer cells that will cause tumor regression. It is thought that lymphocytes are tolerant of tumor antigens because the antigen-presenting cells (**APCs**) lack important costimulatory signals such as B7 and IL-12. Infection with a virus, bacterium, or other foreign substance causes tissue damage and the release of cytokines. These cytokines may be responsible, at least in part, for the appearance of costimulatory signals on APCs. Current research on vaccines is focused on the development of APCs with these costimulatory signals. Dendritic cells are hematopoietic APCs that activate virgin T cells, and they are being used to present tumor antigens to the T cells of the patient with cancer. The major problem with the overall approach to cancer vaccines is, once again, the limited number of antigens that are unique to cancer. For this reason, much of the emphasis continues to be on a laborious process of exposing tumor cells to genetically engineered cell-based vaccines that produce the necessary cytokines or costimulatory molecules. However, antigen-specific cancer vaccines are also being developed using tumor peptides, gangliosides, and nucleic acids. Dendritic cells from the patient have been used to deliver antigens which in turn have resulted in T-cell responses and clinical regression of a few B-cell lymphomas and melanomas. The most promising tumor-specific peptides that have been tested in the clinic use MAGE-3 derived from melanoma cells. In an uncontrolled clinical trial utilizing a combination of high-dose IL-2 and the MAGE-3 vaccine, 41 per cent of patients with metastatic melanoma responded. It remains to be demonstrated in phase III studies whether a similar benefit could be obtained with IL-2 alone.

New therapeutic targets

A new field of science, molecular biology, has emerged during the past few decades. This has contributed substantially to our understanding of the factors involved in the initiation and progression of cancer and has led to the identification of many new therapeutic targets. In general, cancer growth is both the result and the cause of multiple genetic mutations that affect cell division and death. These new therapeutic targets include either the abnormal gene or, more often, gene products (Table 12 and Fig. 2).

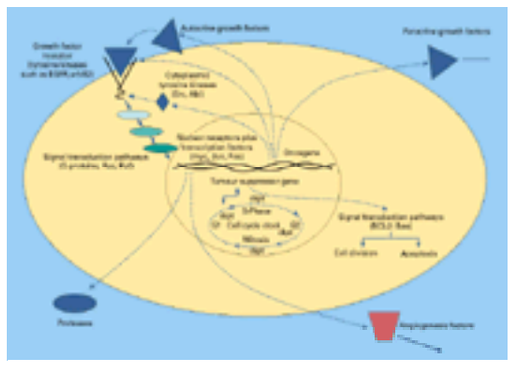


Fig. 2. Signal transduction pathways resulting from oncogene and tumor suppressor gene activation. The initial event in carcinogenesis is an inherited or acquired gene mutation, amplification, or deletion. In a normal cell these DNA abnormalities are detected at one of several checkpoints in the cell cycle clock, and cell death (apoptosis) occurs as the result of a complicated sequence of events (indicated by dotted lines) that involves one or more tumor suppressor genes such as *p53* or *Rb* operating through pathways that often involve the BCL-2:Bax system. An oncogene may cause an increase in membrane-bound growth factor receptors (EGF or the HER2/*neu* receptors), secreted autocrine or paracrine growth factors (HER2/*neu* in this example), intracellular growth factors (Src, Abl), and signal transduction factors (Ras), or it may result in abnormal transcription factors. These changes in either the amount or the type of protein affect the rate of cell growth, adhesion of the cell to its matrix or nearby cells, the ability of the cell to digest basement membrane and invade normal tissues, or the cell's ability to accomplish vital survival functions such as angiogenesis.

Genes that suppress tumor growth do so by detecting gene mutations and, in response to this information, preventing cell division and/or actively promoting apoptosis. The process of detecting DNA abnormalities occurs at a number of checkpoints in the cell cycle clock (Fig. 2). *Rb* (retinoblastoma), *p53*, and *ATM* (ataxia–telangiec-tasia) are among the genes that produce proteins responsible for identifying these abnormalities. For example, activation of *p53* may lead to arrest of cells in G1 and/or cause cell death by signaling through pathways that involve the BCL-2/Bax system. Cells with a mutated, deleted, or inactivated tumor suppressor gene continue to divide in spite of DNA abnormalities and collect additional mutations in other genes with each division. *p53* is the most commonly mutated gene in a variety of tumor types. In most cases this occurs as a late event. Cells with *p53* mutations are more resistant to the effects of radiation and chemotherapy since producing lethal abnormalities in DNA is a critical step in their mechanism of action. Abnormalities in *BCL2* and *Bax* also occur in some tumor types and cause many of the same effects as mutations in p53. Gene therapy aims at restoring the normal function of these genes to suppress tumor growth or, at least, to make the cells more sensitive to the effects of other treatments (see Chapter 48.4). Clinical trials of *p53* gene therapy and *BCL-2* antisense therapy have been started. Other treatments have been designed to replace one or more critical signaling proteins normally produced by tumor suppressor genes.

Oncogene activity may promote continual cell division and uncontrolled growth (Fig. 2). Sometimes mutations in the control region of the gene will lead to the overproduction of a protein, such as epidermal growth factor (**EGF**), or a growth factor receptor that is normally involved in stimulating cell division. In other cases the mutated gene produces an abnormal protein, such as Abl or SRC, that will inappropriately stimulate cell division without a specific external stimulus.

HER2/*neu* (or erbB2) is a member of the EGF/EGFR family. Monoclonal antibodies targeted to this receptor have been shown to be effective in the treatment of metastatic breast cancer. The HER2/*neu* ligand and its receptor are produced in large amounts in some tissues during embryogenesis and in very small amounts in the adult. Tumors in which the *HER2/neu* is frequently amplified and/or overexpressed include glioblastomas and breast, ovarian, lung, bladder, head and neck, cervical, and gastric cancer. Patients whose tumors overexpress *HER2/neu* have a worse prognosis than those with tumors that do not. This tyrosine kinase receptor has both an internal and external domain. The latter can be identified by immunohistochemical assays using either monoclonal or polyclonal antibodies. Between 15 and 20 different antibodies, including one that is commercially available, have been used in published reports evaluating this receptor as a prognostic factor and in determining which anti-HER2/*neu* treatment should be undertaken. The sensitivity amongst these assays varies by more than twofold. Amplification of *HER2/neu* can be determined by fluorescence *in situ* hybridization (**FISH**). A FISH assay is also available commercially, but it has not been determined whether the immunohistochemical or FISH assays are the most useful in the clinic.

In phase II studies, 15 to 20 per cent of patients with metastatic breast cancer had a response to trastuzumab (Herceptin®), a humanized monoclonal antibody that binds to the HER2/*neu* receptor. Almost all of these patients had had extensive prior treatment with cytotoxics. In randomized trials, the response rate among patientswith newly diagnosed metastatic breast cancer was 42 per cent to a combination of doxorubicin and cyclophosphamide and 65 per cent to the same drug

combination along with trastuzumab. Among patients with breast cancer and metastases who had been previously treated with doxorubicin, the response rate to paclitaxel alone was 25 per cent and to a combination of paclitaxel and trastuzumab 57 per cent. Trastuzumab is administered intravenously weekly and induces remarkably few acute side-effects. However, delayed cardiotoxicity identical to that associated with doxorubicin has been observed in 7 per cent of patients treated with trastuzumab alone, 28 per cent of patients treated with trastuzumab plus concomitant doxorubicin, and 11 per cent of patients treated with trastuzumab plus concomitant paclitaxel. This is significantly greater than the incidence of cardiotoxicity observed with either doxorubicin alone (7 per cent) or paclitaxel alone (1 per cent) in randomized comparisons. Since almost all patients in both the phase II and phase III studies of trastuzumab received doxorubicin concomitantly or had been previously treated with doxorubicin, it has been impossible to determine whether trastuzumab itself is cardiotoxic or whether it augments the cardiotoxicity of doxorubicin. Clinical benefits were seen only in patients whose tumors were scored as 3+ using a relatively less sensitive immunohistochemical assay with a range of 0 to 3+.

Biological therapies directed at most of the other targets listed in Table 12 are at much earlier stages of development. Trastuzumab and monoclonal antibodies to EGFR are currently being evaluated in other types of tumors. A number of molecules that inhibit farnesyl transferase, a critical enzyme in the expression of the *Ras* oncogene, are in clinical trials. Drugs that inhibit intracellular tyrosine and cyclin-dependent kinases are in various stages of development.

Nuclear receptors link small molecules, including hormones such as estrogen, progesterone, androgens, and glucosteroids, with specific DNA sequences. While the ligands for hormone receptors were identified long before the receptors, nuclear receptors for which there is no known ligand (‘orphan receptors') have been found. Another class of nuclear receptors binds retinoids, thyroid hormone, and vitamin D. Nuclear receptors are an important part of the signal transduction pathway and modulate transcriptional activity, including transcriptional activity under the direction of oncogenes. Because these changes in transcriptional activity cause cells to take on a more mature phenotype, these ligands have been referred to as ‘differentiation' agents. The value of nuclear receptors as a target for anticancer therapy has been demonstrated with all-*trans* retinoic acid, which is now routinely used as a single agent to induce complete remissions of acute promyelocytic leukemia. The retinoid 13-*cis*-retinoic acid will reduce the incidence of additional primaries but will not suppress the growth of metastases in patients with head and neck cancer. Two of three small randomized trials demonstrated a reduction in the incidence of bladder carcinoma *in situ* with etretinate. Fenretinide (N-[4-hydroxyphenyl] retinamide or **4-HPR**) has been used alone or in combination with tamoxifen to delay the onset of breast cancer in preclinical models. It is currently being evaluated for this purpose in patients with early breast cancer, but the first of the studies evaluating this approach failed to demonstrate any decrease in contralateral breast cancer incidence from the use of 4-HPR.

Cancers are lethal because they are able to invade and displace adjacent non-cancerous tissue, obtain access to the bloodstream or lymphatics, and invade distant organs. All of these processes involve the breakdown or digestion of basement membrane and other tissues that normally separate different tissue types. A number of proteases and collagenases produced by the cancer cell appear to be important in providing cancer cells with these growth advantages. Inhibitors of these enzymes have been developed and are in clinical trials. These new drugs may not be able to induce regression of large tumors but may still be valuable because they are able to prohibit tumor progression. For this reason, large phase III trials in multiple tumor types have been undertaken to evaluable the matrix metalloproteinase inhibitor marimastat since conventional phase II studies are unlikely to demonstrate this benefit.

Once metastases have been established, growth beyond a limited number of cells is dependent on the tumor's ability to ensure its nutritional needs by establishing new blood vessels. The preclinical evidence that this is a pivotal process in cancer development and that it can be inhibited by relatively specific inhibitors is compelling. A number of angiogenesis inhibitors are in phase I and II clinical trials including angiostatin, which is thought to block new blood vessel formation, and endostatin, which blocks new vessel formation and is selectively toxic to tumor blood vessels. A large number of other drugs have also been shown to have antiangiogenetic properties including marimastat, interferon-a, pentosan polysulfate, and low-dose paclitaxel. Proof of principle must still be demonstrated in the clinic, but few biological approaches appear more promising and none are being as intensely studied at the present time.

In the past the paradigm for carcinogenesis has often been an infectious process. It has been assumed that the cancer cells could be totally eradicated by an ‘antibiotic' that took advantage of the unique characteristics of the cancer cell. So far, very few distinctions between normal and cancerous cells are absolute, especially the cancer cells most common in solid tumors. This failure and our new understanding of cancer biology suggest that aging may be a more appropriate paradigm for carcinogenesis than infection. In preclinical studies it can be shown that some cell types have a biological clock based on tandem, repeat sequences of TTAGGG (telomeres) at the ends of chromosomes. As the cell divides, these telomeres shorten, and with this shortening comes cellular senescence. New telomeres require the enzyme telomerase, which is not normally present in adult somatic cells. However, it is present in many tumors. Telomerase inhibitors are being developed in anticipation that cancer cells may not be able to continue unlimited division if this telomerase activity is blocked.

Further reading

A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. *British Journal of Cancer* 1988; **57**: 615–18.

Anonymous. Portal vein chemotherapy for colorectal cancer: a meta-analysis of 4000 patients in 10 studies. Liver Infusion Meta-analysis Group. *Journal of the National Cancer Institute* 1997; **89**: 497–505.

Beatson GW. On the treatment of inoperable cases of carcinoma of the mamma: suggestions for a new method of treatment with illustrative cases. *Lancet* 1896; **2**: 104–7, 162–5.

Bonadonna G, Valagussa P, Moliterni A, Zambetti M, Brambilla C. Adjuvant cyclophosphamide, methotrexate, and fluorouracil in node-positive breast cancer: the results of 20 years of follow-up. *New England Journal of Medicine* 1995; **332**: 901–6.

Carmichael J. Clinical response benefit in patients with advanced pancreatic cancer. Role of gemcitabine. *Digestion* 1997; **58**: 503–7.

Dalton WS. Mechanisms of drug resistance in hematologic malignancies. *Seminars in Hematology* 1997; **34**: 3–8.

Denis LJ *et al.* Maximal androgen blockade: final analysis of EORTC phase III trial 30853. EORTC Genito-Urinary Tract Cancer Cooperative Group and the EORTC Data Center. *European Urology* 1998; **33**: 144–51.

DeVita Jr VT. The relationship between tumor mass and resistance to chemotherapy. *Cancer* 1983; **51**: 1209–20.

Dube S, Heyen F, Jenicek M. Adjuvant chemotherapy in colorectal carcinoma: results of a meta-analysis. *Diseases of the Colon and Rectum* 1997; **40**: 35–41.

Early Breast Cancer Trialists' Collaborative Group. Ovarian ablation in early breast cancer: overview of the randomised trials. *Lancet* 1996; **348**: 1189–96.

Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998; **352**: 930–42.

Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; **351**: 1451–68.

Eisenhauer EA, Vermorken JB. The taxoids. Comparative clinical pharmacology and therapeutic potential. *Drugs* 1998; **55**: 5–30.

Elias AD. Small cell lung cancer: state-of-the-art therapy in 1996. *Chest* 1997; **112**: 251S–258S.

Folkman J. New perspectives in clinical oncology from angiogenesis research. *European Journal of Cancer* 1996; **14**: 2534–9.

Fournier G. Treatment of hormone-refractory prostate carcinoma. *European Urology* 1996; **30**: 32–7. [Review.]

Frei III EF. Selected considerations regarding chemotherapy as adjuvant in cancer treatment. *Cancer Chemotherapy Reports* 1966; **50**: 1–8.

Fuchs CS. Chemotherapy for advanced gastric cancer: where do we stand? *Journal of Clinical Oncology* 1997; **15**: 3299–300.

Garnick MB. Hormonal therapy in the management of prostate cancer: from Huggins to the present. *Urology* 1997; **49**: 5–15.

Gelber RD, Goldhirsch A, Cole BF, Wieand HS, Schroeder G, Krook JE. A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. *Journal of the National Cancer Institute* 1996; **88**: 1039–45.

Glimelius B *et al.* Randomized comparison between chemotherapy plus best supportive care with best supportive care in advanced gastric cancer. *Annals of Oncology* 1997; **8**: 163–8.

Gurney H, Dodwell D, Thatcher N, Tattersall MH. Escalating drug delivery in cancer chemotherapy: a review of concepts and practice—Part 2. *Annals of Oncology* 1993; **4**: 103–15.

Hanahan D. A flanking attack on cancer. *Nature Medicine* 1998; **4**: 13–14.

Haybittle JL. The evidence for cure in female breast cancer. In: Bulbrook RD, ed., *Commentaries on research in breast disease*, Vol. 3, pp. 181–94. Alan R. Liss, New York, 1983.

Henderson I. Adjuvant systemic therapy for early breast cancer. *Cancer* 1994; **74**: 401–9.

Henderson IC. Paradigmatic shifts in the management of breast cancer. *New England Journal of Medicine* 1995; **332**: 951–3.

Henderson IC, Hayes DF, Gelman R. Dose response in the treatment of breast cancer: a critical review. *Journal of Clinical Oncology* 1988; **6**: 1501–15.

Howell A, Dowsett M. Recent advances in endocrine therapy of breast cancer. *British Medical Journal* 1997; **315**: 863–6.

Jones S *et al.* Randomized comparison of vinorelbine and melphalan in anthracycline-refractory advanced breast cancer. *Journal of Clinical Oncology* 1995; **13**: 2567–74.

Kaye SB. Clinical drug resistance: the role of factors other than P-glycoprotein. *American Journal of Medicine* 1995; **99**: 40S–44S.

Klijn JG, Setyono-Han B, Bontenbal M, Seynaeve C, Foekens J. Novel endocrine therapies in breast cancer. *Acta Oncologica* 1996; **35**: 30–7.

Lilenbaum RC, Langenberg P, Dickersin K. Single agent versus combination chemotherapy in patients with advanced nonsmall cell lung carcinoma: a meta-analysis of response, toxicity, and survival. *Cancer* 1998; **82**: 116–26.

Norton L. Implications of kinetic heterogeneity in clinical oncology. *Seminars in Oncology* 1985; **12**: 231–49.

Norton L. A Gompertzian model of human breast cancer growth. *Cancer Research* 1988; **48**: 7067–71.

Norton L, Simon R. The Norton–Simon hypothesis revisited. *Cancer Treatment Reports* 1986; **70**: 163–9.

Osborne CK. Drug therapy: tamoxifen in the treatment of breast cancer. *New England Journal of Medicine* 1998; **339**: 1609–18.

Pardoll DM. Cancer vaccines. *Nature Medicine* 1998; **4**: 525–31.

Pritchard RS, Anthony SP. Chemotherapy plus radiotherapy compared with radiotherapy alone in the treatment of locally advanced, unresectable, non-small-cell lung cancer. A meta-analysis. *Annals of Internal Medicine* 1996; **125**: 723–9.

Prostate Cancer Trialists' Collaborative Group. Maximum androgen blockade in advanced prostate cancer: an overview of 22 randomised trials with 3283 deaths in 5710 patients. *Lancet* 1995; **346**: 265–9.

Schabel FM, Griswold DP, Corbett TH, Laster Jr WR. Increasing the therapeutic response rates to anticancer drugs by applying the basic principles of pharmacology. *Cancer* 1984; **54**: 1160–7.

Shan K, Lincoff AM, Young JB. Anthracycline-induced cardiotoxicity. *Annals of Internal Medicine* 1996; **125**: 47–58.

Skipper HE. Kinetics of mammary tumor cell growth and implications for therapy. *Cancer* 1971; **28**: 1479–99.

Smith IE. Medical management of non-small cell lung cancer. *European Journal of Cancer* 1995; **5**: 818–19.

Tannock IF *et al.* Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized trial with palliative end points. *Journal of Clinical Oncology* 1996; **14**: 1756–64.

Tepper JE *et al.* Adjuvant postoperative fluorouracil-modulated chemotherapy combined with pelvic radiation therapy for rectal cancer: initial results of intergroup 0114. *Journal of Clinical Oncology* 1997; **15**: 2030–9.

Trock BJ, Leonessa F, Clarke R. Multidrug resistance in breast cancer: a meta-analysis of MDR1/gp170 expression and its possible functional significance. *Journal of the National Cancer Institute* 1997; **89**: 917–31.

Wils J. The treatment of advanced gastric cancer. *Seminars in Oncology* 1996; **23**: 397–406.

48.4 Cancer gene therapy

James S. Economou

Introduction
Methods of DNA transfer into cells
Non-viral vectors
Viral vectors
Strategies for cancer gene therapy
Prodrug-converting enzymes
Drug resistance gene therapy
Tumor suppressor gene and oncogene therapy
Cytokine-based cancer gene therapy
Genetic immunotherapy
Concluding remarks
Further reading

Introduction

The concept of gene therapy for inherited diseases is fairly straightforward. The defective gene is replaced with a wild-type gene in appropriate somatic tissues under proper regulatory control. However, while conceptually simple, many practical problems continue to plague the genetic treatment of diseases with single, much less multiple, gene defects. The ability to introduce functional genes with the proper transcriptional regulatory sequences that allows long-term regulated expression still awaits second- and third-generation gene transfer vectors.

On the other hand, the gene therapy of cancer is conceptually quite a bit more complex. The etiology of human cancers at a genetic level may be the result of multiple gene mutations or deletions. For only a minority of cancer is the underlying genetic defect very well understood. An effort to correct all of the genetic alterations in all transformed cells within a tumor is quite impractical. Nevertheless, despite the complexity of the genetic events governing the neoplastic transformation, gene therapy efforts for cancers have flourished and represent the considerable majority of basic, translational, and clinical gene therapy efforts in the last decade. Because cancer gene therapy strategies are directed towards introducing wholly foreign genes into tumors, for which only short-term expression is generally needed, current first-generation vectors are quite suitable. This has allowed the rapid translation of preclinical models into clinical gene therapy trials.

Methods of DNA transfer into cells

There are two general categories of gene vectors—non-viral vectors and recombinant viral vectors.

Non-viral vectors

The use of naked DNA in the form of a eucaryotic expression vector has considerable promise as a method for genetic immunization. Naked plasmid DNA expressing a foreign protein under control of a suitable promoter may be injected into a skeletal muscle or skin (Fig. 1). It can also be biolistically propelled into the skin using a gene gun. DNA is taken up by cells, but remains as an episome (not integrated into the genome), and is transcribed and translated with expression of the vector-encoded protein. This gene product is taken up by host antigen-presenting cells where it is processed and antigenic fragments are presented on the surface of the antigen-presenting cells in the context of major histocompatibility class (MHC) antigens. This causes activation of immune cells which results in generation of antigen-reactive T cells and antibody-producing B cells. The remarkable feature of immune responses to plasmid-encoded vector immunization, as has been shown in experimental animals, is the provision of exceptionally long-lasting immunity. This is a very attractive feature for DNA-based vaccines, particularly with respect to the generation of T-cell responses to define tumor antigens.

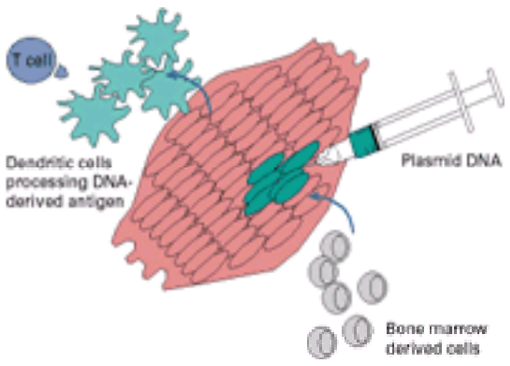


Fig. 1. DNA-based genetic immunization. Plasmid DNA is injected into skeletal muscle where myocytes express the gene product. Dendritic cells derived from bone marrow migrate to the DNA vaccination site, take up the gene product, process the polypeptide, and present antigenic peptides to the immune SYstem.

Liposomes are positively charged lipid membrane that can complex with DNA. These liposome– DNA complexes can fuse with cell membranes permitting the transfer of DNA into cells. The efficiency of liposome-mediated gene transfer is considerably higher than with naked DNA, particularly for single cell transfection *in vitro.* In addition, intratumoral gene transfer of plasmid DNA constructs is significantly improved with the use of cationic liposomes.

Viral vectors

Since the principal job of viruses in nature is the introduction of their nucleic acid into target cells, it is not surprising that recombinant viral vectors have proved to be highly efficient gene transfer vectors. Modifications of retrovirus, adenovirus, adeno-associated virus, lentivirus, poxvirus, and herpesvirus (Table 1 and Fig. 2) are in various stages of development and clinical trial testing.

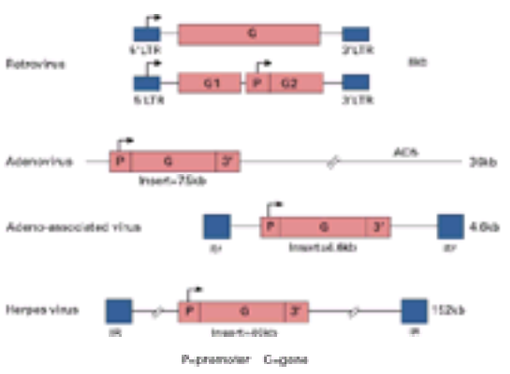


Fig. 2. Viral gene therapy vector backbones. G, gene; p, promoter; 3^{1}, untranslated region; LTR, long terminal repeat; itr, inverted terminal repeat.

Vector	Advantages	Disadvantages
Plasmid DNA	No limitation in gene size Generates good immune response	Limited to muscle and skin Low efficiency
Liposome–DNA	Synthetic carrier No limitation in gene size	Better efficiency
Retrovirus	Integrates into genome Non-immunogenic	Requires replicating target cell Low in vivo efficiency Insert size 8 kb
Adenovirus	High transduction efficiencies Broad range of susceptible target cells Target cell division not required	No integration Immunogenic Insert size 4–5 kb Short duration of expression
Adeno-associated virus	Integrates into genome Broad range of susceptible target cells Target cell division not required	Cumbersome to package Insert size 5 kb
Lentivirus	Integration into genome Insert size 8 kb Long duration of expression	
Herpes simplex virus	Large insert size, 40–50 kb	Toxicity

Table 1 Properties of gene transfer vectors

Retroviruses have an RNA genome which is reversed transcribed into DNA and then integrated into the genome of the infected cell. The retroviral genome is comprised of three genes—*gag, pol*, and *env*. These structural genes are flanked by elements named long terminal repeats (**LTRs**). The LTRs are required for integration into the host genome and also serve as enhancer–promoter sequences that control expression of viral genome. It is possible to replace all three structural genes with one or two so-called transgenes, which may be under control of the viral LTRs or internal promoter elements. The recombinant retroviral backbone must be introduced into a packaging cell line that has been engineered to produce all of the viral structural proteins. The transfected packaging cell line then produces replication-incompetent retroviral particles at very high titer. The released viral particles from these packaging cells are capable of transducing replicating target cells and integrating reversed transcribed (RNA to DNA) transgenes into the cellular genome. Retroviral vectors, because of their ability to stably integrate into target cells, have the potential to allow for long-term expression in transduced cells. However, retroviruses have certain limitations. They can only accommodate approximately 8 kilobases (kb) of DNA. These viruses also cannot transduce non-replicating cells and are relatively inefficient in their ability to transduce cells *in vivo.* For these reasons, recombinant retroviral vectors have found their most useful applications to date in the *ex vivo* transduction of somatic cells which are then returned to the patient.

Adenoviruses are a family of DNA viruses that can infect both dividing and non-dividing cells and are responsible for causing common respiratory infections in humans. Adenoviruses are capable of transducing many cell types with very high efficiency regardless of their cell cycle status. Although these viruses contain over a dozen genes, it is possible to generate replication-defective adenoviral vectors by replacing the E1 gene which is essential for viral replication. First-generation recombinant adenoviruses can be produced by homologous recombination of foreign genes (insert capacity 4 to 5 kb) into the E1 or E3 region; the virus can then be propagated in 293 cells that are engineered to express products of the E1 gene. Very high titers of recombinant adenovirus can be generated, in the order of 10^{11} to 10^{12} viral particles/ml. High transduction efficiencies with robust gene expression can be achieved in most cell tissues except for those of hematopoietic lineage. However, adenoviral vectors exist as episomal elements and are capable only of transient *in vivo* expression. Although virtually any somatic cell can be efficiently transduced *in vivo* and express the transgene, the presence of intact viral genes in the recombinant virus results in low levels of viral gene expression in transduced target cells. These cells all process and express these viral peptides on their cell surface which then act as antigenic targets for immune destruction. Thus, while adenoviral vectors would appear to be entirely unsuitable for stable long-term expression of genes, they work quite well for high-level transient expression in tumors or tissues. For this reason, recombinant adenoviral vectors have gained considerable favor for cytokine-based, tumor suppressor and suicide gene therapy approaches for cancer. Newer adenoviral vectors, crippled by the deletion of additional structural genes, are likely to provide for longer term *in vivo* stability.

Adeno-associated viruses (**AAV**) are small single-stranded DNA viruses which contain two structural genes (*cap* and *rep*), flanked by two inverted terminal repeats. Recombinant AAV vectors can accommodate only 3.5 to 4.0 kb of foreign DNA and their preparation using current technology is labor intensive. Nevertheless, AAV vectors are capable of reasonably high-efficiency transduction and integration both *in vitro* and *in vivo* and have considerable promise for therapies requiring long-term expression.

Lentiviruses, of which the human immunodeficiency virus is a member, are members of the retrovirus family. Lentiviruses contain the three retroviral genes as well as genes for six accessory proteins. An important feature of lentiviruses is that they can infect both dividing and non-dividing cells. These viruses may be disabled by deleting the structural and some of the accessory genes without affecting vector production or transduction efficiency. When lentivirus vectors are injected into somatic cells *in vivo*, they are capable of long-term expression and do not appear to generate an immune response towards transduced cells. These vectors have considerable potential in gene medicine.

The herpes simplex virus is a large double-stranded DNA virus that is capable of establishing a latent infection in the brain. However, the biology of its large genome (150 kb), containing more than 80 genes, has not been fully characterized and its potential for development into useful gene therapy vectors awaits further study. Other viral vectors include vaccinia, pox and vaccinia virus; some are already currently in clinical trial. Vaccinia virus can accommodate large inserts of foreign genes by homologous recombination and while they transduce target cells with high efficiency, this usually results in target cell lysis within several days. Thus, this represents another efficient, but transient, transduction vector.

Strategies for cancer gene therapy

Various strategies have been developed for gene therapy of cancer. These include those using prodrug-converting enzymes, drug resistance genes, tumor suppressor or oncogene therapy, cytokine-based therapies, and genetic immunotherapy.

Prodrug-converting enzymes

Prodrug-converting enzymes convert a non-toxic precursor drug into a toxic metabolite. The prototype enzyme in this category is the herpes simplex virus thymidine kinase (**HSVtk**) gene. HSVtk converts gancyclovir into a monophosphate form that is then converted into its triphosphate form by cellular enzymes and incorporated into the DNA of replicating mammalian cells. This leads to the inhibition of DNA replication and cell death. Since only viral thymidine kinase can efficiently use gancyclovir as a substrate, this drug is used to selectively inhibit herpes virus replication. This enzyme has been used in a number of vector backbones (retroviral, adenoviral, adeno-associated viral) to direct conversion of gancyclovir into toxic metabolites within solid tumors *in vivo. In vivo* transduction of tumors by direct intratumoral vector injection with subsequent systemic administration of gancyclovir can result in pronounced tumor inhibition in a variety of experimental tumors. An interesting feature of HSVtk-based gene therapy is that not all cells within the tumor need to be transduced to achieve tumor regression. There is now good evidence that bystander, untransduced tumor cells may also be inhibited by the uptake of phosphorylated gancyclovir from dying transduced cells and/or destroyed through the generation of antitumor immune responses stimulated by this suicide gene therapy. Thus, this approach may have clinical potential, particularly for tumors of the central nervous system.

Another suicide gene approach has employed the cytosine deaminase (**CD**) gene from *Escherichia coli*. Cytosine deaminase catalyzes the conversion of the prodrug 5-fluorocytosine to the toxic 5-fluorouracil. Expression of the CD gene allows the conversion of 5-fluorocytosine to 5-fluorouracil within transduced tumor cells with high local concentrations of the drug at the tumor cell level. The use of this prodrug-converting enzyme has been used in experimental models of gastrointestinal malignancies.

Although earlier approaches with prodrug-converting enzyme employed vectors in which the suicide gene was expressed constitutively by a strong viral promoter, second-generation vectors have been tested using tumor-specific regulatory sequences. As an example, the regulatory regions of the carcinoembryonic antigen gene have been used in such second-generation vectors to drive the CD gene. This results in expression of the CD gene only in cells producing carcinoembryonic antigen. This represents a potentially important safety measure to prevent expression of this gene in normal replicating tissues.

Drug resistance gene therapy

Another gene therapy strategy is to render normal hematopoietic stem cells resistant to cytotoxic drugs as a means to prevent myelosuppression and to perhaps permit dose intensification systemic therapy. Towards this goal, the **MDR1** gene (multiple drug resistance gene) has been introduced into bone marrow or peripheral blood-derived stem cells. The MDR1 gene produces a P-glycoprotein which functions as a cellular efflux pump that may be responsible for the resistance of some tumor cells to various cytotoxic drugs. MDR-transduced normal bone marrow stem cells are resistant to systemically administered chemotherapeutic agents. Other potential drug resistance genes include dihydrofloate reductase, topoisomerase II, and methylguanine methyltransferases. While this approach is conceptually attractive, it has important caveats. Drug toxicity to non-hematopoietic organs such as the gastrointestinal tract or heart may prove to be dose-limiting. Higher doses of chemotherapy may not result in higher response rates and there is the concern that transduction of a small population of contaminating tumor cells may produce drug

resistant tumors.

Tumor suppressor gene and oncogene therapy

A very promising avenue of cancer gene therapy involves exploiting the p53 tumor suppressor gene status of human cancers. Mutations in p53 are found in a large percentage of human cancers. p53 is a nuclear phosphoprotein transcription factor that controls the expression of proteins involved in the cell cycle. When DNA is damaged, p53 will accumulate in the nucleus and inhibit the cell cycle through a cyclin-dependent kinase inhibitor. p53 can also induce apoptosis or programmed cell death through a variety of pathways. The loss of p53 has been implicated in tumor progression and this can occur by direct genetic mutation/deletion or binding by viral oncoproteins. The restoration of wild-type p53 expression in cells with deleted or defective p53 is sufficient to cause apoptosis or growth arrest. There is considerable evidence that expression of wild-type p53 in established tumors can induce apoptosis *in vivo* as well. This has been accomplished using retroviral, adenoviral, and non-viral gene delivery vectors expressing wild-type p53.

The induction of apoptosis by certain chemotherapeutic agents may require the presence of wild- type p53. The sequential administration of an adenovirus expressing p53 and cisplatin induces SYnergistic tumor regression *in vivo.* p53-based clinical trials are being conducted for a number of malignancies including lung, head and neck, and hepatocellular cancer. There is clear evidence from these phase I trials that adenoviral p53 vectors are capable of efficiently transducing tumor cells *in vivo*, expressing the wild-type p53 product, and mediating clinically significant tumor regressions.

Another novel strategy to exploit the p53 status of tumors is the use of adenoviruses deleted in the E1B region. E1B is essential for viral replication and codes for a protein that binds the p53 protein. Adenoviruses must inhibit p53 in order to replicate in an infected cell. An E1B defective adenovirus, then, could not replicate in a p53 wild-type tumor cell. However, E1B-deleted adenoviruses appear to be capable of replicating and killing tumor cells which have absent p53 function. These viral vectors are currently in clinical trial for head and neck, pancreatic, and hepatocellular cancers.

Oncogenes act in a genetically dominant manner and confer the transforming phenotype in the host cell. Oncogenes are the mutant homologs of proto-oncogenes, which are genes that play critical roles in cell regulation including signal transduction and transcription. Only a single mutant allele is require for the malignant phenotype. One strategy that has been employed to inhibit the function of oncogenes is to inhibit translation of the oncogene messenger RNA using antisense constructs. Antisense constructs have the mirror image of base pair sequences to those in the oncogene messenger RNA. They will bind to the message and inhibit the translation of the gene product. This strategy has been applied with some success to the *ras* family of oncogenes using *ras* antisense vectors.

Cytokine-based cancer gene therapy

There is a compelling evidence that tumor cells present tumor antigens very poorly to the immune system. Induction of tumor antigen-specific T-cell immunity has several requirements. The tumor antigen must be correctly processed and presented in the context of MHC antigen. In addition, the antigen-presenting cell must also express costimulatory molecules such as B.7 to deliver additional signals to the T lymphocytes (Fig. 3). Finally, proinflammatory cytokines provide a favorable environment for the clonal activation and proliferation of antigen-reactive T lymphocytes. Tumor cells, however, frequently downregulate MHC expression, do not express costimulatory ligands, and produce immunosuppressive substances such as transforming growth factor-b, prostaglandins, and interleukin 10 (**IL-10**). It has been recognized for almost half a century that some experimental animal tumors can be rendered immunogenic by mixing tumor cell vaccines with certain bacterial adjuvants. In the last decade, there has been considerable interest in examining the tumor responses generated by tumor cell vaccines genetically engineered to secrete proinflammatory cytokines. This strategy exploits the paracrine nature of the cytokine in which their immunostimulatory effects are most pronounced by regional production at the site of tumor cell antigens. Thus, many recombinant cytokines, which have marginal antitumor activity when delivered systemically, have been found to have considerably enhanced adjuvant effects when produced at the tumor vaccine site. Virtually all the cytokines that have been tested reduce the tumorigenicity of genetically engineered tumor cells. Tumor cells stably or transiently transfected with these cytokines, when produced at sufficient levels, have reduced or absent *in vivo* growth. The cytokines tested include IL-1, IL-2, IL-3, IL-4, IL-6, IL-7, IL-12, interferons-a and -g, tumor necrosis factor-a and -b, and macrophage colony-stimulating factor, granulocyte colony-stimul-ating factor, and granulocyte/macrophage colony-stimulatory factor (**GM-CSF**). This reduced tumorigenicity *in vivo* is caused by the influx of host effector cells such as natural killer cells, macrophages, granulocytes, host antigen-presenting cells, and T lymphocytes. In some experimental models, genuine MHC-restricted T-cell immunity is generated, rendering these animals immune to a rechallenge of the parental untransduced tumor. Among the most effective cytokines in generating immunity is GM-CSF, a cytokine with no immunostimulatory effects on T lymphocytes. When the mechanism of GM-CSF-based tumor vaccine immunization was studied, it was clear that this cytokine attracted and activated host antigen-presenting cells, such as dendritic cells, which migrate to the injection site, take up antigen, process it, and present it the immune system. These experimental findings have spawned clinical trials using autologous tumor cell vaccines transduced cytokine vectors that include IL-12, IL-7, GM-CSF, interferon-g, and IL-12.

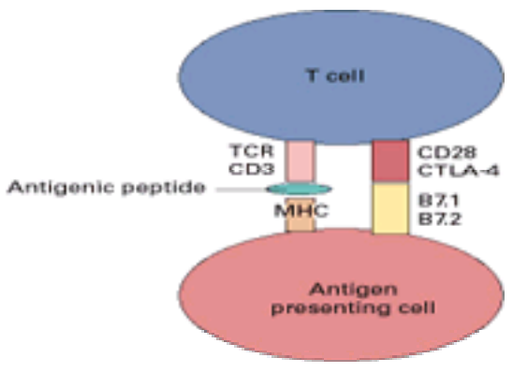


Fig. 3. Activation of T cells by antigen-presenting cells. The T-cell receptor (TCR) is engaged by the antigenic peptide bound to the MHC molecule. Costimulation is provided through ligand interactions.

In addition to transducing tumor cells with cytokines, parallel efforts have involved engineering tumor cells to express important costimulatory molecules such as B7.1 and B7.2. Since these costimulatory molecules provide an important second signal for T-cell activation, these strategies are designed to convert tumor cells into better antigen-presenting cells. These efforts have met with success in experimental models and are in clinical trial.

The *ex vivo* manipulation and transduction of autologous tumor cells, however, is labor intensive and recent efforts have been directed towards *in vivo* transduction of tumor with either cytokine genes and/or costimulatory molecules. Such an approach requires vectors that can generate high transduction efficiencies and levels of cytokine expression. Recombinant adenovirus, adeno-associated virus, and poxvirus vectors are suitable for these strategies. In addition, non-viral gene delivery methods including the gene gun, in which DNA-coded particles are propelled into cells, have been shown to be effective in preclinical animal studies. *In vivo* gene delivery using recombinant viral vectors carries with it certain risks, chief among these is that some vector will be distributed systemically. Even if carefully injected into the tumor, this phenomenon has been amply documented in animal models and accounts for much of the toxicity of these tumor-directed transduction strategies. It is for this reason that transcriptionally specific, second-generation vectors have been developed. As described earlier, these involve replacing the constitutive viral promoters used to drive these genes with the regulatory sequences that are specific for certain tumor histologies. For example, the promoters for the carcinoembryonic antigen, a-fetoprotein, and tyrosinase genes will direct synthesis of their gene products only in tumor cells producing these tumor markers or enzymes. This genetic manipulation should add an important degree of biological safety for these *in vivo* approaches.

Genetic immunotherapy

The immune system recognizes antigens presented as short peptides bound to MHC class I and class II molecules (Fig. 4). The antigenic epitopes of 8 to 11 amino acid in length presented by MHC class I molecules, derived from intracellular proteins digested by the proteasome complex, are transported to the cell surface and bound by the MHC molecule. MHC class I molecules are displayed on the surface of most cells. However, class II molecules have a more restrictive distribution and are mainly expressed on the surface of antigen-presenting cells such as dendritic cells, macrophages, and B cells. CD8 T cells recognize peptides presented by MHC class I molecules through a T-cell receptor complex. CD8 cells are critically involved in mediating antitumor responses. Antigen-specific CD8 cells become activated by the T-cell receptor/MHC peptide interaction, together with help from activated CD4 T cells. This leads to clonal expansion of CD8 cytotoxic T lymphocytes that will specifically lyse target cells expressing this same peptide class I complex on their surface. CD8-mediated cytotoxicity induces programmed cell death primarily through perforin-granzyme B mechanisms. CD4 T cells are activated by specific peptide sequences presented by MHC class II. Two distinct populations of CD4 T cells have been described. T helper 1 (**TH1**) cells produce IL-2, interferon-g, and tumor necrosis factor-a and **TH2** cells produce IL-4, IL-5, IL-6, and IL-10. **TH1** cells mediate delayed-type hypersensitivity responses and **TH2** cells are more potent helpers for antibody production. The importance of each T- helper subtype in antitumor responses can be inferred from the involvement of **TH1** cells in cytolytic responses, while the **TH2** pattern is believed to induce anergy and to downregulate **TH1**

responses. Optimal T-cell activation requires the engagement of costimulatory molecules such as B7.1, B7.2, and CD40 in addition to T-cell receptor triggering.

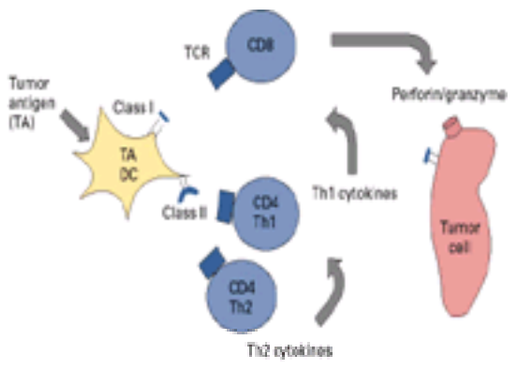


Fig. 4. Activation of T cells by dendritic cells (DC). Dendritic cells take up tumor antigen (TA) and present antigenic peptides on their surface in the context of MHC class I and II molecules. CD8 and CD4 T cells, respectively, are clonally activated and mediate killing of tumor cells expressing the same TA.

Tumor antigens must be presented by host antigen-presenting cells. Among the hematopoietic cells that can present antigen, dendritic cells are the most potent described. Dendritic cells are derived from bone marrow and characterized by dendritic morphology, high mobility, and the ability to present antigen to naive T cells in an MHC-restricted fashion. Dendritic cells express high levels of MHC class I and II molecules, costimulatory molecules such as B7.1 and B7.2, and adhesion molecules. Dendritic cells can take up tumor antigen, process the polypeptide into short antigenic fragments, and present these to the immune system in the context of MHC class I and II (Fig. 4). It is now possible to generate large numbers of dendritic cells from hematopoietic precursors. Both bone marrow and peripheral blood progenitors may be cultured in cytokines such as stem cell factor, GM-CSF, and IL-4 to yield highly enriched, potent, differentiated dendritic cells.

What remains, then, in this equation of genetic immunotherapy is the identification and cloning of defined tumor rejection antigens. Tumor antigens were first convincingly described in malignant melanomas through the use of tumor infiltrating lymphocyte-based expression cloning. Tumor-infiltrating lymphocytes are the T cells infiltrating solid tumors or metastatic deposits. In melanoma, these T cells are enriched with cells having cytotoxic specificity for the tumor they have infiltrated. When these tumor-infiltrating lymphocytes are isolated and expanded in IL-2, they are highly cytotoxic in an MHC-restricted manner for that melanoma. These tumor-infiltrating lymphocytes were then used to identify their respective melanoma-specific antigens and a number of these were identified. These included MART1, gp100, and tyrosinase. Surprisingly, the majority of these tumor rejection antigens were found to be normal, non-mutated differentiation antigens present in normal melanocytes and melanomas. Despite the exposure of the immune system to these melanocyte lineage ‘self' antigens, immunologic tolerance was not generated. This represented an important discovery in tumor immunology, namely, that human tumor rejection antigens could be represented by lineage-specific ‘self' antigens. These findings in melanoma have generated considerable interest in the use of other ‘self' antigens expressed by other human cancers as suitable targets for immunotherapy; for example carcinoembryonic antigen, prostate-specific antigen, Her2-neu, mutated *ras*, wild-type and mutant p53, a-fetoprotein, and others. There is good evidence that T-cell responses can be generated to MHC-restricted epitopes from all of these proteins.

With the identification and cloning of potential tumor rejection antigens, various strategies towards genetically immunizing hosts to these defined antigens are now being tested. The use of GM-CSF-transduced tumor cell vaccines, in which host antigen-presenting cells are attracted to the vaccine site to stimulate immunity, is one form of genetic immunotherapy to undefined tumor antigens. Another genetic immunization strategy involves the intramuscular injection or the biolistic delivery of naked DNA encoded with tumor antigen. This has been shown to be an efficient means of generating both cellular and humoral immunity. The recent acquisition of methods to generate large numbers of potent antigen-presenting dendritic cells *in vitro* has allowed a direct testing of the immunostimulatory properties of these cells. In animal models and now in clinical trials, synthetic tumor antigen peptides, known to bind specific human HLA class I alleles, are loaded onto dendritic cells which are then used to immunize the patient. This peptide-based dendritic cell strategy is capable of inducing T_r cell responses. Even in settings in which the tumor antigens have not been defined, putative tumor antigen can be eluted from the dissected tumor and loaded onto dendritic cells. Finally, an exciting strategy is the possibility of genetically engineered dendritic cells to express defined tumor antigen (Fig. 5). A tumor antigen gene therapy vector, such as an adenoviral vector, may be used to transduce dendritic cells *in vitro.* The expressed tumor antigen will be properly processed by the dendritic cell and the antigenic peptides displayed on the cell surface. These dendritic cells are then administered to the patient to generate tumor antigen peptide-specific responses. This approach has several advantages including the continuous expression of antigenic peptides, the potential of both MHC class I and class II presentation, and the simultaneous presentation of multiple antigenic peptide determinants derived from the tumor antigen polypeptide. These strategies are currently in clinical trial and should provide important insight into autoimmune antitumor responses.

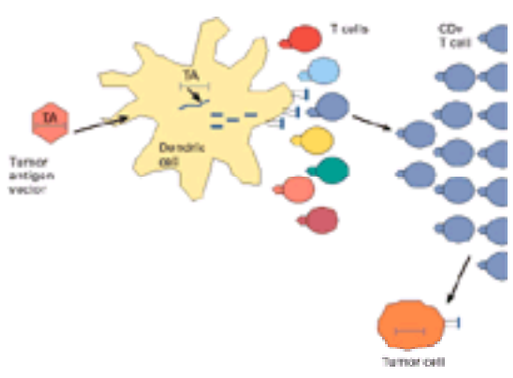


Fig. 5. Dendritic cell-based genetic immunotherapy. Dendritic cells (DC) are engineered to express and process a defined tumor antigen (TA). MHC-restricted antigenic peptides are presented to antigen-specific T cells which are activated and clonally expanded.

Concluding remarks

The application of gene therapy strategies to cancer has considerable promise for several reasons. Current efforts require only short-term transgene expression for which first-generation vectors are suitable. Also, while the underlying biology of most human cancers is only incompletely understood, molecular events required for their destruction either by immune mechanisms or through apoptotic programmed cell death are currently on sounder footing. The number of clinical cancer gene therapy trials being initiated each year increases exponentially and early results have promise.

Further reading

Brenner MK. Applications of gene transfer in hematologic malignancy. *Recent Results in Cancer Research* 1998; **144**: 61–9.

Colombo MP, Forni G. Immunotherapy I: Cytokine gene transfer strategies. *Cancer and Metastasis Reviews* 1997; **16**: 421–32.

Ertl HC, Xiang Z. Novel vaccine approaches. *Journal of Immunology* 1996; **156**: 3579–82.

Jaffe EM, Pardoll DM. Considerations for the clinical development of cytokine gene-transduced tumor cell vaccines. *Methods in Enzymology* 1997; **12**: 143–53.

Lashford LS. Possibilities of gene therapies for cancer. *Annals of Medicine* 1997; **29**: 1–4.

Nielsen LL, Maneval DC. p53 tumor suppressor gene therapy for cancer. *Cancer Gene Therapy* 1998; **5**: 52–63.

Roth JA, Cristiano RJ. Gene therapy for cancer: what have we done and where are we going. *Journal of the National Cancer Institute* 1997; **89**: 21–39.

Tuting T, Storkus WJ, Lotze MT. Gene-based strategies for the immunotherapy of cancer. *Journal of Molecular Medicine* 1997; **75**: 478–91.

Weichselbaum RR, Kufe D. Gene therapy of cancer. *Lancet* 1997; **349**: 10–12.

48.5 Emergencies in cancer

James Carmichael

General
Vascular emergencies
Superior vena caval compression
Cardiac tamponade
Pulmonary emergencies
Acute breathlessness
Pulmonary haemorrhage
Gastrointestinal emergencies
Gastrointestinal haemorrhage
Gastrointestinal obstruction
Gastrointestinal perforation
Neurological emergencies
Spinal cord compression
Raised intracranial pressure
Seizures
Metabolic emergencies
Calcium homeostasis
Disturbance of sodium balance
Renal failure
Tumour lysis
Hypoglycaemia
Haematological and haemostatic emergencies
Neutropenia and infection
Thrombosis
Thrombocytopenia
Conclusion
Further reading

General

Emergencies occur frequently in cancer patients, either as a complication of cancer or its management. The use of more aggressive surgery and more intensive systemic therapy may serve to increase the frequency and severity of these events. The appropriate management of these emergencies is greatly influenced by the age and condition of the patient, the type, extent, and stage of the tumour, its timing relative to specific cancer management, and the wishes of the patient and his or her family. These emergencies can be grouped under six major subheadings.

Vascular emergencies

Superior vena caval compression

Superior vena caval obstruction occurs in between 3 and 8 per cent of patients with lymphoma or lung cancer (particularly the small cell type). Conversely, superior vena caval obstruction has a malignant aetiology in over 90 per cent of cases. Non-malignant causes are rare, although it may be precipitated by central venous catheterization. Common clinical features include shortness of breath and swelling of the face, upper limbs, or trunk. Tachypnoea, venous distension, or oedema of the face and arms are the commonest clinical signs. Untreated superior vena caval compression may lead to thrombosis, neurological toxicity, or respiratory impairment. However, if the obstruction develops slowly, collateral venous channels may develop. These patients may be relatively asymptomatic, with a low risk of serious sequelae.

The diagnosis of superior vena caval obstruction is predominantly clinical, although chest radiology or computed tomographic (**CT**) scanning frequently reveals a mediastinal mass. Venography is rarely indicated and is relatively contraindicated because the raised venous pressure increases the risk of haemorrhage. When superior vena caval obstruction is the primary presenting sign, a tissue diagnosis should be sought prior to starting treatment, to ensure patients receive the most appropriate anticancer therapy. Ideally, tissue should be obtained using the least invasive procedure as superior vena caval obstruction increases the risk of haemorrhage. Rarely, in the presence of rapidly progressing disease, treatment may be justified in the absence of a tissue diagnosis. Therapeutic options include radiation therapy, systemic chemotherapy, and with improvements in invasive radiological techniques, stenting is increasingly considered. Radiation is the most common treatment for superior vena caval obstruction, particularly for drug-resistant tumours. Where the tumour is chemosensitive, and where disease is also present outside the chest, chemotherapy should be considered for primary treatment. Resolution of symptoms are faster and more complete the more sensitive the tumour. Therapeutic response normally begins within 72 h with 90 per cent of patients achieving symptomatic improvement within 7 days, the rate of improvement similar for both radiation and chemotherapy modalities. Failure of response may indicate that venous thrombosis has occurred.

Surgical bypass of superior vena caval obstruction has been successful, but can be hazardous, although sophisticated interventional radiology reduces this risk. Unfortunately there are no large randomized trials comparing these modalities in superior vena caval obstruction.

Cardiac tamponade

Advanced cancer involving the heart and pericardium is unusual, although is reported in approximately 10 per cent of autopsies, and pericardial effusions are a not uncommon autopsy finding. Symptoms of cardiac tamponade can be caused by a malignant pericardial effusion or by constriction due to tumour infiltration or radiation-induced fibrosis. Cardiac tamponade due to a malignant pericardial effusion is relatively rare, but a life-threatening and treatable condition. The symptoms of tamponade are non-specific with breathlessness the most common and frequently only complaint and occasionally chest pain. The characteristic clinical signs comprise a raised jugular venous pressure, tachycardia, hypotension, and marked pulsus paradoxus. Patients with tamponade are at risk of sudden death from a catastrophic drop in cardiac output, cardiac arrhythmia, or cardiac ischaemia. Tamponade may be produced by a rapidly accumulating effusion as small as 200 ml, but over 1 litre may accumulate slowly before symptoms develop. A chest radiograph may show the typical large, globular heart, but the radiograph may be normal with a small effusion. Echocardiography will quickly and accurately diagnose pericardial effusion and should be performed when tamponade is suspected and considered in the presence of unexplained breathlessness in the patient with cancer.

When a significant effusion is detected, immediate pericardiocentesis should be performed, preferably under echocardiographic screening. Placement of a pericardial catheter at the same time will allow complete drainage of the fluid and the instillation of a sclerosant such as tetracycline or thiotepa will prevent reaccumulation should the effusion be malignant. Systemic chemotherapy may prevent recurrence in chemosensitive tumours, while radiotherapy may also be beneficial, particularly if the tumour can be treated without irradiating the whole heart. Surgical intervention, usually to form a pleuropericardial window, is indicated when other methods of control fail and when the patient has a reasonable life expectancy. The recently reported technique of percutaneous balloon pericardiotomy appears as effective but safer than an open surgical procedure and may well become the technique of choice in recurrent malignant effusion. The prognosis of these patients is poor overall, although is dependent on the sensitivity of the underlying cancer. It has been shown that survival in patients with breast cancer who develop tamponade, for instance, is almost double that of patients with lung cancer.

The symptoms and signs of cardiac tamponade caused by tumour infiltration without significant amounts of fluid are similar, with echocardiography revealing a thickened pericardium and little fluid. Effective antitumour therapy with radiation or chemotherapy may relieve the constriction, with surgical options limited to pericardiectomy, although the outcome is generally poor.

Pulmonary emergencies

Acute breathlessness

Respiratory impairment is common in patients with cancer. Metastatic tumour, lymphangitis carcinomatosa, malignant pleural effusion, infection, haemorrhage, emboli, and the effects of drugs and radiation can all cause life-threatening respiratory failure. Accurate diagnosis is most critical where the underlying cause is both treatable

and rapidly progressive. Extensive, SYmptomatic lung metastases or pulmonary infiltrations are generally associated with a poor prognosis. When the underlying tumour can be cured with chemotherapy (for example, lymphomas or germ cell tumours) vigorous resuscitation is warranted, including ventilatory support if necessary. This may be needed particularly following the start of chemotherapy, when pulmonary function may deteriorate as a result of tumour lysis. In responsive but non-curable tumours (for example, breast cancer or prostate cancer) dramatic responses to hormone or chemotherapy may also occur and pulmonary support may be indicated.

A tumour arising in, or metastatic to, the main bronchi may cause life-threatening airways obstruction even when the tumour is small. Patients usually present with stridor or wheeze, which may be mistaken for bronchoconstriction. Respiratory function may deteriorate quickly. Radiotherapy produces rapid relief of SYmptoms in most patients, as will chemotherapy in patients with small cell carcinoma. Where more rapid relief is required, or when the tumour recurs following external beam radiotherapy, endobronchial laser resection, brachytherapy, or the insertion of a tracheobronchial stent can be effective.

Alternatively, many cytotoxic drugs can cause pulmonary reactions. Pulmonary fibrosis, for instance, has been described, particularly after bleomycin and the alkylating agents. Bleomycin can also produce acute breathlessness, for which high-dose prednisolone is indicated to prevent progression to chronic fibrosis. Great care should also be taken in anaesthetizing patients who have been treated with bleomycin, as exposure to high concentrations of oxygen may precipitate an acute, life-threatening pulmonary reaction.

Breathlessness is a very frightening symptom for patients and their relatives and in addition to specific measures listed above detailed attention should be paid to symptomatic control with oxygen and opiates.

Pulmonary haemorrhage

Massive pulmonary haemorrhage can occur in patients with lung cancer due to vascular infiltration commonly seen in centrally sited tumours. Dependent on the prognosis of the patient, surgical resection can be considered; other palliative options include laser coagulation or embolization.

Intrapulmonary and intrabronchial haemorrhage is also seen in patients with thrombocytopenia, either directly tumour-related or due to disseminated intravascular coagulation. Infection, particularly fungal infection, can also result in significant blood loss. Treatment of these complications requires full ventilatory support, control of haemostasis, and therapy of the underlying cause.

Gastrointestinal emergencies

Gastrointestinal haemorrhage

Gastrointestinal haemorrhage is common in patients with cancer, but in only 12 to 17 per cent is it due to the malignancy, with gastritis and peptic ulceration common in these patients. Other contributory factors include treatment- or disease-related thrombocytopenia, chemotherapy-induced mucositis, or tumour infiltration. The management of such cases should be based on standard protocols.

Gastrointestinal obstruction

Obstruction is the commonest gastrointestinal emergency which can be caused by either intrinsic or extrinsic compression by tumours, although non-malignant causes should not be forgotten. Blockage by oesophageal, gastric, and colorectal tumours is frequently intrinsic. However, small bowel obstruction is commonly caused by metastatic peritoneal spread from ovarian or pancreatic cancers.

Treatment of bowel obstruction is dependent on the site of the lesion. Oesphageal obstructions can be relieved by stenting or by the use of external beam radiotherapy. Obstructions at other sites are commonly treated surgically by resection, decompression, or bypass should the patient's prognosis support this. In very sensitive tumours, chemotherapy can be used whilst giving the patient full medical support. In patients not suitable for surgical intervention, full supportive measures should be offered including fluid replacement, antispasmodics, antiemetics, and stool softeners. In refractory patients the use of octreotide can be considered.

Gastrointestinal perforation

Lymphomas are most commonly associated with bleeding and perforation. As many as 33 per cent of gastrointestinal, high-grade, non-Hodgkin's lymphomas perforate either at presentation or during treatment. The greatest risk of perforation is during the early stages of treatment when highly chemosensitive lymphomas show the maximum rate of tumour regression. The risk of perforation and/or bleeding can be greatly reduced by surgical resection of the lymphoma before starting chemotherapy. This should be undertaken, where possible. However, two-thirds of perforations are not due directly to tumour. Other chemosensitive tumours, such as ovarian cancer, may also perforate due to dramatic tumour lysis following the first cycle of chemotherapy. It should therefore not be assumed that perforation is a sign of treatment failure. It may, conversely, herald rapid tumour regression. It should also be remembered that many of these patients are receiving high-dose corticosteroid therapy either as part of their treatment or antiemetic regimens.

Neurological emergencies

There are a variety of neurological complications that occur in patients with cancer, some are caused by the cancer itself and others are complications of therapy and can be caused by either radiation or systemic chemotherapy.

Spinal cord compression

Spinal cord compression occurs most commonly in tumours (such as those of breast, lung, prostate, myeloma) that metastasize to bone and also non-Hodgkin's lymphoma. Early recognition and prompt treatment are essential to prevent permanent neurological damage, disability, and shortened survival.

Pain is the initial symptom in almost all patients, and any back pain in a patient with cancer should be assessed urgently. The pain may be localized to the spine or it may have nerve-root distribution. When thoracic in origin, nerve-root pain may radiate into the chest or abdomen and cause diagnostic confusion. Weakness is also common at presentation. As the compression progresses, paraesthesias, sensory loss, and bladder and bowel dysfunction may develop. The prognosis is critically dependent on the neurological condition of the patient at diagnosis. Sixty per cent of patients who are ambulatory at diagnosis remain so after treatment, while only 20 per cent of those paraplegic at diagnosis improve significantly. Additional adverse prognostic features include loss of bladder function, a rapid-onset lesion, or a high-thoracic lesion.

Magnetic resonance imaging is the investigation of choice in suspected spinal cord compression and gives excellent images of the spinal canal. Where this is not available, myelography is indicated to show the extent and number of lesions.

In all cases, high-dose steroids (usually dexamethasone) should be given while the diagnosis is confirmed and definitive treatment planned. It remains debatable whether very high doses of dexamethasone are superior to standard doses (16 mg/day), with one randomized trial suggesting this. The best treatment for spinal cord compression is unknown as no prospective studies have compared the possible therapies—surgery, radiotherapy, and chemotherapy—either alone or in combination. The choice is influenced by the nature of the tumour (whether radiosensitive or chemosensitive), the level of block, the rapidity of onset, and the clinical skills available locally. Surgery gives the quickest relief of spinal compression and may be the best approach when the compression is progressing rapidly, if there is vertebral collapse, when the histology is unknown, and when the tumour is resistant to or has recurred following radiation therapy. Complete clearance is unlikely and surgery should be followed by radiotherapy to prevent local recurrence. However, there is no evidence that the more rapid relief of compression achieved by surgery results in better outcomes compared with radiotherapy, which is considered by many to be the standard therapy. In highly chemosensitive tumours, particularly when there is systemic disease and the compression is progressing slowly, chemotherapy should be considered, perhaps in conjunction with radiotherapy.

Spinal infection is another, although rarer, cause of spinal cord compression in the patient with cancer. Infection in the epidural space may occur in isolation or may be secondary to vertebral osteomyelitis. In either case it is usually due to haematogenous spread from infection elsewhere. The neurological signs are similar to those described above, but they progress rapidly and occur in association with marked pain, local tenderness, and fever. Urgent surgical decompression is indicated in conjunction with appropriate antibiotic treatment.

Raised intracranial pressure

Increased intracranial pressure is a common problem with both primary and secondary brain tumours. Non-specific symptoms include headache, nausea, and altered mental state. In addition, patients may have localized symptoms dependent on the site of the tumour. Occasionally, patients may present with clinical features of coning, which can also be precipitated by lumbar puncture that is inappropriate in this clinical setting. The use of mannitol and surgical decompression by shunt may be of value. In some tumour types surgical resection of solitary brain metastases should be considered. However, in the majority of patients palliation is the goal, with the use of high-dose steroids and radiation a more appropriate approach.

Seizures

Up to one-third of patients with brain metastases will develop seizures. Fits may be induced by infections or haemorrhage in some cases. In other cases seizures are exacerbated by metabolic disturbances. Some patients develop status epilepticus requiring intravenous anticonvulsant therapy although oral therapy can control fits in the majority of patients. All patients should subsequently receive prophylactic anticonvulsant therapy.

Metabolic emergencies

Calcium homeostasis

The most common metabolic complication of malignancy is hypercalcaemia, which occurs in about 10 to 20 per cent of all cases. The majority of patients with malignant hypercalcaemia have either lung cancer (notably squamous carcinoma) or breast cancer, but myeloma, lymphoma, and renal cell carcinoma also contribute significant numbers. Extensive bone metastases can induce hypercalcaemia by producing locally acting cytokines that induce bone resorption. Other tumours induce hypercalcaemia, even in the absence of bone metastases, by hormonal mechanisms, such as release of parathyroid hormone-related peptide. This peptide has similar functions to parathyroid hormone, probably via direct action on the parathyroid hormone receptor. However, there is limited structural homology between parathyroid hormone and this related peptide. It promotes renal resorption, increases bone resorption of calcium, and increases phosphate excretion.

Patients with hypercalcaemia have thirst, polyuria, nausea, and vomiting, leading to dehydration, hypovolaemia, and diminishing renal function. In severe hypercalcaemia, electrocardiographic abnormalities and a diminished level of consciousness may develop, leading to coma, ventricular arrhythmias, and asystole.

The initial treatment of malignant hypercalcaemia comprises intravenous saline to correct hypovolaemia followed by specific calcium-lowering therapy. The treatment of choice is intravenous bisphosphonates (such as pamidronate or clodronate), which are potent inhibitors of osteoclastic bone resorption and produce a fall in serum calcium over 3 to 6 days. Patients who do not respond to bisphosphonates can be treated with calcitonin or the cytotoxic drug plicamycin (mithramycin). Traditionally, glucocorticosteroids have been used, but several reports show them to be ineffective in hypercalcaemia resulting from solid tumours. They may be helpful in myeloma and lymphoma, but by a direct antitumour effect. Where the underlying tumour is responsive to chemotherapy or hormone therapy, this should be used to prevent recurrence of hypercalcaemia.

Hypocalcaemia also occurs in patients with cancer, but is rarely due to hypoparathyroidism. It is most commonly found in patients with cisplatin renal toxicity resulting in hypomagnesaemia and hypocalcaemia.

Disturbance of sodium balance

Mild hyponatraemia is common in patients with cancer, particularly those with advanced disease. In a few cases hyponatraemia is more marked and is due to the tumour secreting antidiuretic hormone, being most frequent in small cell lung cancer, but having been reported with a wide range of tumours. This syndrome is characterized by a reduced serum osmolarity, suggested by a low serum sodium, in the presence of an inappropriately concentrated urine. In mild cases, patients may be asymptomatic or have only malaise, anorexia, or headaches. When the serum sodium falls below 115 mmol/1, drowsiness, confusion, coma, and occasionally death result.

Restriction of fluid intake to between 500 and 1000 ml/day will result in a slow but steady increase in serum sodium. For patients who are unresponsive to fluid restriction or unable to comply, an antidiuretic hormone antagonist such as demeclocycline should be used. This causes a reversible, dose-dependent, diabetes insipidus and reliably raises serum sodium. It can, however, induce uraemia and should not be used in conjunction with fluid restriction. Secretion of antidiuretic hormone due to small cell lung cancer should also be treated with cytotoxic chemotherapy. This will generally produce prompt resolution, even before tumour shrinkage is seen. Where hyponatraemia is causing coma or convulsions, more rapid correction of serum sodium may be needed. Intravenous saline should be given to correct the serum sodium to 115 mmol/1 at a rate not exceeding 0.5 to 1.0 mmol/1.h. More rapid correction may induce central pontine yelinosis.

Hypernatraemia is much less common in patients with cancer and is most commonly associated with a water losing state. A number of tumours are associated with a diabetes insipidus syndrome, although this represents a low proportion of patients presenting with diabetes insipidus. Some small cell lung cancers are associated with secretion of an ACTH-like peptide that can result in hypernatraemia. The hypernatraemia should be corrected over 2 to 3 days and attempts made to treat the underlying cause.

Renal failure

Renal failure due to tumour progression generally occurs as a terminal event for which no specific therapy is indicated. Where the underlying tumour is chemosensitive, more aggressive intervention is indicated. Bilateral ureteric obstruction may be bypassed by retrograde ureteral stents or by percutaneous nephrostomies, allowing time for cytotoxic chemotherapy to produce tumour shrinkage. Where irreversible renal damage has occurred, but where the tumour is treatable, haemo- or peritoneal dialysis may be indicated, and may be given at the same time as chemotherapy and radiotherapy.

Renal failure may be induced by several commonly used cytotoxic drugs, for example, cisplatin, ifosfamide, or high-dose methotrexate. Appropriate schedules for administering these drugs reduce the risk of significant renal damage as can the use of protective agents such as Amifostine. Close monitoring of renal function should be performed in patients on these drugs, but when renal impairment does occur, alternative, non-nephrotoxic drugs, such as carboplatin should be considered.

Tumour lysis

This syndrome presents as a complex metabolic derangement induced by the rapid lysis of large numbers of tumour cells. This occurs following the treatment of bulky, highly sensitive cancers such as testicular teratoma and lymphomas. Biochemical abnormalities include hyperuricaemia, hyperkalaemia, hypocalcaemia, metabolic acidosis, and renal failure. Treatment of this syndrome when established is frequently unsuccessful, with prevention the best approach. ‘At risk' patients should be adequately hydrated prior to chemotherapy, any acidosis treated with bicarbonate, and the patients should be pretreated and maintained on allopurinol. Incorporation of these measures should prevent this SYndrome in the majority of patients.

Hypoglycaemia

This is an unusual occurrence in malignancy. It can be a feature of insulinomas and occasionally in non-islet cell tumours of the pancreas. In addition a number of sarcomas have been associated with hypoglycaemia. The precise aetiology of this remains unclear, but the production of insulin-like growth factors and decreased production of glucagon may play a role. These patients have fatigue, dizziness, and are often confused or can present in coma. Intravenous glucose is frequently required in the acute phase and these patients may benefit from glucagon therapy if tumour control is not achievable.

Haematological and haemostatic emergencies

Neutropenia and infection

The patient with cancer commonly has reduced resistance to infection for multiple reasons related to their disease and its treatment. In such patients infection is probably the major cause of morbidity and mortality. Although many infections occur in the terminal stages of disease, when vigorous treatment may not be indicated, they may also threaten the survival of patients whose tumours are potentially curable.

The most common infectious emergency is the onset of fever in a patient who is neutropenic as a result of chemotherapy or radiotherapy. Many such patients have a septicaemia which, if not treated with appropriate antibiotics, may progress rapidly with fatal consequences. Febrile, neutropenic patients should be started on broad-spectrum, intravenous antibiotics at the earliest opportunity and before the results of bacteriological investigations are known. Most septicaemias are due to the

patient's endogenous flora, usually mouth or enteric Gram-negative pathogens. The antibiotics chosen should cover such organisms. The synergistic combination of an aminoglycoside and a broad-spectrum penicillin is widely used (for example, piperacillin/gentamicin).

Patients with indwelling, intravenous catheters, particularly central venous catheters, are also at risk of Gram-positive septicaemia. In such patients, or where fever does not respond to first-line antibiotics, vancomycin should be added to cover possible Gram-positive infection.

The management of patients who remain febrile despite standard antibiotics is complex, and firm guidelines are difficult to formulate. Such patients should be carefully re-examined and reassessed. Samples should be sent for repeat bacteriological examination. Fungal infection is a particular concern as it is difficult to diagnose and is a major cause of mortality. Fungal infections are increasingly common the longer the neutropenia persists. For these reasons empirical antifungal therapy should be considered if fever and neutropenia persist. If abnormalities develop on chest radiography, this may suggest infections by unusual organisms such as *Aspergillus, Pneumocystis*, or *Legionella*. The use of bone marrow growth factors such as granulocyte colony-stimulating factor remains controversial in prolonged neutropenic states. Its use can be considered in an acutely infected patient with prolonged neutropenia, but its routine use in this setting has not been shown to be beneficial in randomized studies. The presence of central nervous system signs should also prompt careful assessment of the patient for indications of meningitis or brain abscess.

Thrombosis

It is well recognized that many malignancies are associated with an increased risk of venous thrombosis, although the precise mechanism for this is unclear. As described by Trousseau, thrombosis is often a presenting feature of malignancy. Some thrombotic episodes are due to venous compression, but many are due to increased coagulability. The treatment of such thromboses does not differ from that in non-malignant cases, although they may be more refractory to treatment, or may have certain clinical features making anticoagulation inappropriate, such as cerebral metastases. Recurrent pulmonary embolism refractory to anticoagulation may require the placement of an inferior vena caval filter.

Disseminated intravascular coagulation is common in malignancy. It is frequently mild and chronic requiring no specific therapy, but on occasions presents acutely when rapid intervention is required. It is clinically manifest by impaired coagulation, thrombocytopenia, and haemorrhage, but small-vessel thrombosis may also occur, resulting in ischaemia. The best approach to disseminated intravascular coagulation is to treat the underlying malignancy. Direct treatment of the haemostatic disturbances is complex and unsatisfactory. Bleeding episodes can be treated by intravenous infusion of concentrated clotting factors and platelet transfusion. However, this does not reverse the underlying pathological processes, and the transfused factors may simply be consumed. Heparin can interrupt the processes of disseminated intravascular coagulation and thus restore normal haemostasis, but it is not universally effective and may be dangerous in the presence of severe hypofibrinogenaemia or thrombocytopenia.

Thrombocytopenia

Thrombocytopenia is seen in many patients with cancer either as part of the disease process as in haematological malignancies, due to bone marrow infiltration by solid tumours, as a presenting feature of disseminated intravascular coagulation, and may also follow myelosuppressive, cytotoxic chemotherapy. However, it does not always require active intervention. When the platelet count falls below $10 \times 10^6/1$, platelet transfusion is recommended, although a lower limit has been recommended by some. However, the presence of bleeding, infection, or other haemostatic deficits increases the risk of bleeding and would encourage a higher threshold. Thrombocytopenia following cytotoxic chemotherapy is normally seen as part of a pancytopenia, although certain drugs appear to be more toxic to the platelet lineage, such as carboplatin. In contrast, other drugs, particularly tubulin-binding drugs, appear to be relatively platelet sparing, with combinations of paclitaxel and carboplatin causing less thrombocytopenia than carboplatin itself. Thrombocytopenia is also seen as part of the haemolytic–uraemic syndrome. Haemolytic–uraemic syndrome is associated with certain malignancies untreated, but in addition, some cases are thought to be drug related, this syndrome having been reported following treatment with cytotoxic drugs including mitomycin-C and gemcitabine.

Conclusion

Emergencies arising in the patient with cancer may be complex and, like many oncological problems, are best managed by a multidisciplinary approach. With the increasing effectiveness and intensity of anticancer therapies these events are likely to increase in frequency and it is more important that we control them effectively and in a timely fashion. In order to prevent undue morbidity or mortality, all those managing patients with cancer should be aware of the correct treatment for the common emergencies that arise in these patients.

Further reading

Vascular toxicity

Keane D, Jackson G. Managing recurrent malignant pericardial effusions. *British Medical Journal* 1992; **305**: 729–30.

Lokich JJ, Goodman R. Superior vena cava SYndrome: clinical management. *Journal of the American Medical Association* 1975; **231**: 58–61.

Roswit B, Kaplan G, Jacobson HG. The superior vena cava obstruction in bronchogenic carcinoma. *Radiology* 1953; **61**: 722–37.

Vaitkus PT, Herrmann HC, LeWinter MM. Treatment of malignant pericardial effusion. *Journal of the American Medical Association* 1994; **272**: 59–64.

Weinberg BA, Conces DJ jr, Waller BF. Cardiac manifestations of non cardiac tumors. Part 1: Direct effects. *Clinical Cardiology* 1989; **12**: 289–96.

Pulmonary toxicity

Dumon J-F. A dedicated tracheobronchial stent. *Chest* 1990; **97**: 328–32.

Miller RR, McGregor DH. Hemorrhage from carcinoma of the lung. *Cancer* 1980; **46**: 200–5.

Stover DE, Kaner RJ. Pulmonary complications in cancer patients. *CA: A Cancer Journal for Clinicians* 1996; **46**: 303–20.

Gastrointestinal toxicity

Klein MS, Ennis F, Sherlock P, Winawer SJ. Stress erosions: a major cause of gastrointestinal hemorrhage in patients with malignant disease. *American Journal of Digesive Diseases* 1973; **18**: 167–73.

Lightdale DJ, Kurtz RC, Boyle CC, Sherlock P, Winawer SJ. Cancer and upper gastrointestinal tract hemorrhage: benign causes of bleeding demonstrated by endoscopy. *Journal of the American Medical Association* 1973; **226**: 139–41.

Neurological toxicity

Boogerd W, van der Sande. Diagnosis and treatment of spinal cord compression in malignant disease. *Cancer Treatment Reviews* 1993; **19**: 129–50.

Burch PA, Grossman SA. Treatment of epidural cord compressions from Hodgkin's disease with chemotherapy. A report of two cases and a review of the literature. *American Journal of Medicine* 1988; **84**: 555–8.

Schiff D, Batchelor T, Wen PY. Neurologic emergencies in cancer patients. *Neurologic Clinics of North America* 1998; **16**: 449–83.

Metabolic toxicity

Glover DJ, Glick JH. Metabolic oncologic emergencies. *CA: A Cancer Journal for Clinicians* 1987; **37**: 302–20.

Mosekilde L, Eriksen EF, Charles P. Hypercalcemia of malignancy: pathophysiology, diagnosis and treatment. *Critical Reviews in Oncology/Hematology* 1991; **11**: 1–27.

Slevin ML, Bell R, Cotterill AM, Lister TA. The tumour overkill syndrome. A potentially lethal complication of cancer chemotherapy. *Postgraduate Medical Journal* 1981;**57**: 727–9.

Zapf J *et al.* Regulation of binding proteins for insulin like growth factors in humans. Increased expression of $IGFb_2$ during IGF 1 treatment of healthy adults and in patients with extra-pancreatic tumour

hypoglycemia. *Journal of Clinical Investigation* 1990; **86**: 952–61.

Haematological and haemostasis toxicity

Aderka D, Brown A, Zelikovski A, Pinkhas J. Idiopathic deep vein thrombosis in an apparently healthy patient as a premonitory sign of occult cancer. *Cancer* 1986; **57**: 1846–9.

American Society of Clinical Oncology recommendations for the use of hematopoietic colony stimulating factors: evidence based clinical practice guidelines. *Journal of Clinical Oncology* 1994; **12**: 2471–508.

Consensus Conference: platelet transfusion therapy. *Journal of the American Medical Association* 1987; **257**: 1777–80.

Kies MS, Posch JJ Jr, Giolma JP, Rubin RN. Hemostatic function in cancer patients. *Cancer* 1980; **46**: 831–7.

Klastersky J. Empiric therapy for febrile granulocytopenic patients: no longer a challenge? *Annals of Oncology* 1991; **7**: 454–5.

48.6 Access for chemotherapy

Charles A. Staley

Preoperative evaluation
Operative techniques
Short-term complications
Long-term complications
Maintenance of the devices
Further reading

Patients with cancer commonly require the infusion of chemotherapy, saline, parenteral nutrition, and multiple venipunctures. During the last few decades, safe central venous access devices have been developed to facilitate optimal medical care and improve the quality of life for these patients. The adverse effects of chemotherapeutic agents including phlebitis, venous thrombosis, and fibrosis drastically limit the use of peripheral veins for intermediate- or long-term therapy. Since the introduction of long-term venous access by Broviac and the subsequent modification of this device for use in patients undergoing bone marrow transplant by Hickman in 1979, central venous access devices have become an important factor in the comprehensive care of patients withcancer.

Currently, there are four types of central venous access device to deliver chemotherapy. For short- to intermediate-term therapy, an external, peripherally inserted central catheter (**PICC**) can be inserted at the bedside or procedure room by a qualified nurse. The catheter is placed percutaneously through an anticubital vein and threaded into the superior vena cava. PICC lines are inexpensive and well tolerated by patients. Complications such as venous thrombosis and infection can occur.

Hohn catheters are non-tunneled, Silastic, external central venous catheters which can be used for intermediate-term chemotherapy. The catheter is inserted percutaneously into the central venous system at the bedside or in a procedure room. Similar to PICC lines, Hohn catheters are inexpensive and durable if patients are well instructed on daily maintenance.

The major indication for tunneled central venous access catheters is for the administration of total parenteral nutrition and/or chemotherapy associated with bone marrow transplantation. The catheters are durable, are available in single, double, and triple lumens, and can be left in place for years. A modification of this catheter is the Groshong catheter which has a three-way valve at the tip to prevent reflux of blood back into the catheter.

The most common access device for long-term therapy is an implanted subcutaneous port. These ports are available in single or double lumen, a titanium or plastic reservior, and an attached silicone or polyurathane catheter (Fig. 1). The reservior is completely under the skin and is accessed through the skin with a non-coring needle (Huber needle). The implanted ports and tunneled central venous access catheters are usually inserted in the operating room with fluoroscopic guidance.

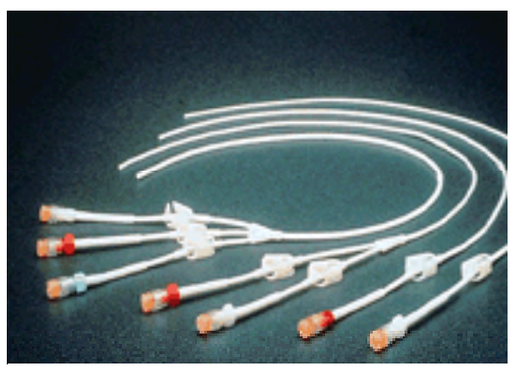


Fig. 1. Single-, double-, and triple-lumen tunneled central venous access catheters. (Manufactured by Sims Deltec Inc.)

Preoperative evaluation

In assessing a patient's need for vascular access for chemotherapy, the physician must know the duration (long term or adjuvant) and mode of administration (bolus or constant infusion) for the planned treatment. Patients undergoing a short course of bolus adjuvant chemotherapy with good veins may be treated through a periperal intravenous line each time at considerably less cost than placing a vascular access device. However, patients with poor veins and/or needing continuous infusion or prolonged treatment should be considered for a vascular access device. Preoperative evaluation should include a history and a physical examination focusing on a history of previous central venous access devices, or anatomic considerations that would change the position or method of line insertion. Patients with previous catheters should be questioned concerning past line or reservior pocket infections and arm swelling indicating possible central vein thrombosis. A recent study demonstrated that in patients with previous thrombosis or infection, 46 per cent had thrombosis of one or more central veins. Preoperative duplex ultrasonography of the central veins should be done in these patients to help direct catheter insertion. Those with previous infections should be examined to rule out any current acute infection and the catheter should be inserted on the opposite side. Insertion of central vein access devices in patients with a previous claviclar fracture, chest wall tumor, or previous breast surgery should be done on the contralateral side.

Preoperative blood work should include a complete blood count with platelet count and differential. Patients with platelet counts below 50 000 should be transfused with platelets just before the catheter insertion. Patients with absolute neutropenia (less than 1000) should be postponed until counts have improved. Routine coagulation studies are not necessary except for patients who have been on coumadin or heparin.

Operative techniques

External PICC catheters are inserted through periperal veins in the arms and the length measured so that the tip ends in the superior vena cava. These catheters can be placed by specially trained nurses.

The Hohn catheter (percutaneous), the tunneled external catheter, and the implanted reservior catheter are all inserted with a similar technique (Fig. 2). After preoperative evaluation, the appropriate side of insertion, left or right, is determined. The patient is placed supine on the operating table with both arms tucked at the side. Sometimes a roll is placed down the middle of the back to lower the shoulders posteriorly. The subclavian vein or internal jugular veins can be accessed percutaneously while the patient is in the Trendelenberg position. Pulsatile flow from the needle can indicate arterial puncture. If any doubt exists, a sterile intravenous extension tubing can be connected to the needle and central pressure measured. Once in the vein, the needle is used to insert a guidewire into the superior vena cava under fluoroscopic guidance. A subcutaneous pocket is made for the port and the external catheter is tunneled up to the insertion site. An introducer with the peel-away catheter is placed over the wire—pushing forward while pulling back on the wire. The wire and introducer should not be advanced as one unit. Pulling back intermittently on the wire while advancing the introducer can prevent perforation of the vein wall. Position of the catheter tip is confirmed by fluoroscopy. Blood return is confirmed and the catheter is flushed with heparinized saline. A postoperative chest radiograph is obtained for positioning and to rule out a pneumothorax.

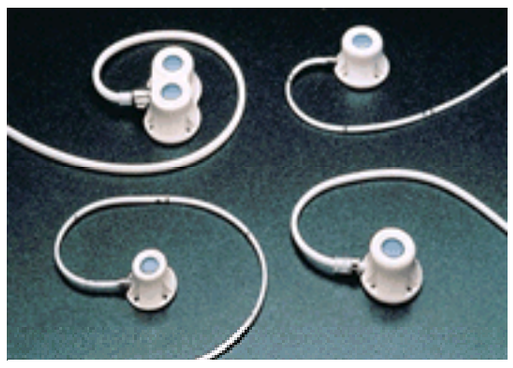


Fig. 2. Single- and double-lumen implantable subcutaneous ports. (Manufactured by Sims Deltec Inc.)

A cutdown approach can be used to gain access to the cephalic, external jugular, or internal jugular veins. The cephalic vein is dissected in the deltopectoral groove. This vein is absent in 10 to 20 per cent of patients and is a small vein in the majority of patients. A cutdown of the external jugular vein should only be done over a visible vessel while the patient is in the Trendelenberg position. After correct positioning of the catheter has been confirmed by fluoroscopy, a suture is placed around the catheter and proximal vein to anchor the catheter. The internal jugular vein may be accessed percutaneously or by a cutdown through an oblique incision anterior to the sternocleidomastoid muscle.

Some surgeons elect to use a fluoroscopy-free insertion technique utilizing an electromagnetic detection system. Avoiding fluoroscopy may potentially save on radiation exposure and cost. Frank *et al.* demonstrated that accurate placement of central access devices could be done without fluoroscopy in 85 per cent of patients.

In patients with superior vena cava occlusion, access through the saphenous vein may be necessary. Optional saphenous vein mapping can be done preoperatively. A small transverse incision is made over the vein and a 2- to 3-cm portion of vein is dissected. The proximal vein is tied off with a suture ligature. A transverse venotomy is made and the catheter inserted into the inferior vena cava under fluoroscopy. The subcutaneous port or external catheter is tunneled to the anterior thigh or lower abdominal wall. In a study by Treiman and Silberman, 86 per cent of patients had successful long-term infusional therapy through a saphenous vein catheter. Infection rates of femoral catheters are similar to subclavian inserted ports. Lower extremity edema due to iliac or caval thrombosis can occur in 10 to 20 per cent of patients and may be prevented by routine treatment with heparin or coumadin in these hypercoagulable patients.

Short-term complications

The complication rates associated with the insertion of central venous access devices are clearly related to the experience of the clinician inserting the catheter.

Potential short-term complications that can occur at the time of catheter insertion include pneumothorax and bleeding. A pneumothorax is caused by a lung parenchymal injury at the time of the venipuncture due to the close proximity of the vessels to the apex of the lung. Varying degrees of pneumothorax with or without clinical symptoms can occur and in the most serious condition can result in a tension pneumothorax. The incidence of a pneumothorax for catheters placed percutaneously is 1 to 5 per cent and can be avoided by using a cutdown insertion technique. Because patients with a pneumothorax can be asymptomatic, a routine chest radiograph is obtained after insertion. A small pneumothorax (less than 30 per cent) in an aymptomatic patient may be observed and checked for stability in 4 h with a repeat chest radiograph. For a larger or symptomatic pneumothorax, tube thoracostomy should be carried out to expand the lung fully.

Bleeding from the insertion of a central line can be local, mediastinal, intrathoracic, or pericardial. Overall the incidence is rare (less than 1 per cent), but can be life threatening. Local bleeding can occur at a venous cutdown site or in the reservior pocket. Most mediastinal or intrathoracic bleeding occurs while inserting the percutaneous introducer sheath over the guidewire. Curving the sheath, intermittently pulling back on the wire, and/or using fluoroscopy during the insertion of the introducer can minimize this complication. Treatment may require a tube thorocostomy or for more serious injuries a thoracotomy. Pericardial tamponade is the most potentially lethal complication related to venous access device insertion. Hypotension, distended neck veins, muffled heart sounds, and cardiopulmonary arrest may occur. An immediate pericardiocentesis, creation of a pericardial window, and repair of the laceration may be necessary. The use of softer catheter materials such as Silastic or polyurethane has decreased the incidence of bleeding complications.

Long-term complications

Potential long-term complications include infection, thrombosis, malfunction, and fracture. The most common complication related to venous access devices is infection (10 to 41 per cent), which can occur locally, regionally (in the tunnel or pocket), or systemically. Localized erythema or induration in most cases can be treated with local wound care and antibiotics. Infections in the subcutaneous tunnel or port pocket are almost never successfully treated without removal of the device. Systemic line sepsis should always be a concern for patients with these vascular access devices who present with signs of an acute or chronic infection. Patients should be treated with broad-spectrum antibiotics and other causes for sepsis ruled out. If there is no improvement in 24 to 48 h, the device should be removed. The ability of antibiotics to eradicate line sepsis is controversial, but most clinicians would remove the device for persistent or recurring sepsis. The most common pathogens include *Staphylococcus epidermidis, Staphylococcus aureus, Escherichia coli*, pseudomonas, and candida. In multiple studies, fungal infections could not be treated without removing the access device. In general, multilumen catheters (compared with single lumen), those inserted using a cutdown technique (compared with percutaneous), and externally placed catheters (compared with subcutaneous ports) have higher incidences of infection. Studies examining the role of prophylactic preoperative antibiotics have not been consistent, but one dose of cephazolin is inexpensive and potentially beneficial. Silver-impregnated cuffs on tunneled devices have not been beneficial in decreasing line sepsis and add to the cost of the catheter.

Catheter-related venous thrombosis can be a devastating complication for patients with cancer who are dependent on venous access devices for therapy. The incidence of thrombosis varies from 0 to 50 per cent. Development of ipsilateral arm swelling, pain, bluish discoloration of the skin, numbness, and/or paresthesias should be investigated with duplex ultrasonography. Of interest is that asymptomatic venous thrombosis is equal to that of symptomatic thrombosis. Many factors contribute to the development of venous thrombosis including catheter size, catheter tip position, biomaterial, venous entry site, an underlying hypercoagulable state, and associated infection. Eastridge and Lefor demonstrated that catheter tips above the T3 level had a significantly increased rate of venous thrombosis. Fluoroscopy at the time of catheter insertion is recommended to assure adequate catheter length and tip position. Treatment of most venous thrombosis requires removal of the catheter and intravenous continuous infusion of heparin. Long term, these patients are usually treated with 3 months of coumadin. In separate studies, low-dose coumadin (1 mg.) and low-molecular-weight heparin (Fragmin, 2500 IU) given for 90 days after catheter insertion significantly decreased the incidence of venous thrombosis. There has been no direct comparison of coumadin with low-molecular-weight heparin. There were no increased rates of bleeding in either study.

Malfunction of the catheter over time usually indicates fibrin at the tip of the catheter. A chest radiograph should be obtained to rule out catheter migration. If the tip is in good position, the catheter should be instilled with 5000 IU of urokinase. A rare cause of catheter malfunction is a ‘pinch-off' that occurs when the catheter enters the subclavian vein too medially thus compressing the catheter between the first rib and clavicle. If this compression is seen on a chest radiograph, the catheter should be removed. If not removed, the catheter can fracture and embolize into the cardiopulmonary circulation.

Maintenance of the devices

Tunneled external catheters require daily flushing with heparinized saline whereas subcutaneous ports only need flushing every 4 to 6 weeks. The most important factor to a functioning venous access device is care of the catheter. Multiple studies have shown that excellent nursing and patient education concerning the care of the devices will significantly decrease complications. With appropriate catheter care, venous access devices can last many month to years without complications.

Further reading

Ahmed Z, Mohyuddin Z. Complications associated with different insertion techniques for Hickman catheters. *Postgraduate Medical Journal* 1998; **74**: 104–7. [A review of complications associated with different insertion techniques.]

Broviac JW, Cole JJ, Schribner BH. A silicone rubber atrial catheter for prolonged parenteral alimentation. *Surgery, Gynecology, Obstetrics* 1973; **136**: 602–6. [Description of using a central venous line for prolonged alimentation.]

Eastridge BJ, Lefor AT. Complications of indwelling venous access devices in cancer patients. *Journal of Clinical Oncology* 1995; **13**: 233–8. [Review of factors associated with central line

complications.]

Frank JL, Halla B, Garb JL, Reed WP. Fluoroscopy-free placement of standard chest wall subcutaneous chronic venous access devices. *Annals of Surgical Oncology* 1997; **4**: 597–602. [A study to evaluate potential benefits of using a fluoroscopy free system for line insertion.]

Groeger JS *et al.* A prospective, randomized evaluation of the effect of silver impregnated subcutaneous cuffs for preventing tunneled chronic venous access catheter infections in cancer patients. *Annal of Surgery* 1993; **218**: 206–10. [Randomized trial evaluating silver impregnated cuffs on central line catheters.]

Henriques HF, Karmy-Jones R, Knoll SM, Copes WS, Giordano JM. Avoiding complications of long-term venous access. *The American Surgeon* 1993; **59**: 555–8. [Complications of long-term venous access devices.]

Hickman RO, Buckner CD, Clift RA, Sanders JE, Stewart P, Thomas ED. A modified right atrial catheter for access to the venous system in marrow transplant recipients. *Surgery, Gynecology, Obstetrics* 1979; **148**: 871–5. [Description of a right atrial catheter for bone marrow transplant patients.]

Kock HJ, Pietsch M, Krause U, Wilke H, Eigler FW. Implantable vascular access systems: experience in 1500 patients with totally implanted central venous port systems. *World Journal of Surgery* 1998; **22**: 12–16. [Experience with 1500 implanted central venous port systems.]

Kraybill WG, Allen BT. Preoperative duplex venous imaging in the assessment of patients with venous access. *Journal of Surgical Oncology* 1993; **52**: 244–8. [A study to determine the role of duplex imaging before central line insertion.]

Mansfield PF, Hohn DC, Fornage BD, Gregurich MA, Ota DM. Complications and failures of subclavian-vein catheterization. *New England Journal of Medicine* 1994; **331**: 1735–8. [Evaluation of complications and failures of subclavian vein catheterization.]

Montreal M *et al.* Upper extremity deep venous thrombosis in cancer patients with venous access devices—prophylaxis with a low molecular weight heparin (Fragmin). *Thrombosis and Haemostasis* 1996; **75**: 251–3. [Low molecular weight heparin prophylaxis in patients with venous access devices.]

Ramirez JM, Miguelena JM, Guemes A, Moncada E, Cabezali R, Sousa R. Fully implanted venous access systems. *British Journal of Surgery* 1993; **80**: 347–8. [A study demonstrating the safety and reliability of implantable access devices.]

Tennenberg S *et al.* A prospective randomized trial of an antibiotic- and antiseptic-coated central venous catheter in the prevention of catheter-related infections. *Archives of Surgery* 1997; **132**: 1348–51. [A randomized trial comparing standard central venous catheters with antibiotic coated catheters.]

Torramade JR *et al.* The complications of central venous access systems: a study of 218 patients. *European Journal of Surgery* 1993; **159**: 323–7. [Complications of central venous devices.]

Treiman GS, Silberman H. Chronic venous access in patients with cancer. *Cancer* 1993; **72**: 760–5. [The use of the saphenous vein for central venous device insertion.]

Whitman ED. Complications associated with the use of central venous access devices. *Current Problems in Surgery* 1996; **33**: 311–78. [Complications of central venous access devices.]

48.7.1 Principles of radiation oncology

David Morris and Joel Tepper

Introduction
Physical factors of ionizing radiation
Units of measurement
Types of radiation used in clinical practice
Clinically related radiation biology
Interaction with biologic material
Cell-survival curves and descriptors of radiation sensitivity
Repair of radiation damage
Repopulation
Oxygen effect
Effects of cell cycle and redistribution
Clinical issues with respect to the physical and biological factors
Normal tissue tolerances
The therapeutic ratio
Relation to other oncologic modalities
Primary radiation therapy
Combined radiation and surgery
Combined radiation and chemotherapy
Summary
Further reading

Introduction

The goal in oncologic management is to eradicate all of the tumor, cause no morbidity, and perform the therapy in a timely fashion. At present, the treatment of cancer patients involves primarily surgery, radiation, chemotherapy, and/or supportive medical care. Radiation therapy has been used clinically for nearly 100 years and remains a major component in such management.

Modern principles of radiation therapy depend on an understanding of the physical factors of ionizing radiation, its biologic effect on tumors and normal tissues, and the patterns of spread of different types of malignancies. Patterns of failure after treatment as well as the interactions of radiation with other types of treatment are of great importance in meeting the goals of oncologic management. In this chapter, we will address the physical and biologic factors important in radiation oncology, the clinical importance of those factors, and the relation of radiation therapy to other oncologic modalities.

Physical factors of ionizing radiation

Radiation is defined as the propagation of energy through space and matter. If the radiation has sufficient energy to eject an orbital electron from an atom during its interaction in matter, it is known as ionizing radiation. With the ionization event there is localized release of a large amount of energy sufficient to break chemical bonds and initiate events that are expressed ultimately as biological damage.

Ionizing radiation is typically categorized as either electromagnetic (X- and g-rays) or particulate (electrons, protons, neutrons, a-particles, and heavy charged particles). Electromagnetic radiation can be considered as both a wave and a particle. X-rays are produced within the nucleus and g-rays are produced outside it, but both are short-wavelength (< 10^{-8} cm) radiations that have no mass or electrical charge. Because of their short wavelength, they are of high frequency and thus have sufficient energy to create ionization. In this sense, they act like particulate radiation. Particulate radiation has a definite mass, a defined position, and a defined momentum at a given point in time. It can also have charge. When it interacts with matter, ionization can occur, resulting in the transfer of energy that can initiate a cascade of biological events.

When radiation interacts with biologic matter, ionization may lead to a direct or an indirect effect on the biological system. Direct effects result from direct interaction with the biologic molecule (most commonly DNA for radiation cell killing); the indirect effect arises when X-rays interact with water molecules and create free radicals, which subsequently interact with DNA molecules and cause damage. If the DNA damage cannot be repaired appropriately, cell death or mutations can result.

Units of measurement

Many attempts have been made to quantify the chemical and biological effects of radiation. Initially, a unit called skin erythema dose (SED), the amount of radiation that would cause reddening of the skin, was used to quantitate radiation dose. However, such a unit is very variable and was therefore abandoned for a more precise unit known as the roentgen (or röntgen; R), a unit of exposure that measures ionizations produced in air by relatively low-energy X- and g-rays [energy less than 3 × 10^6 electronvolts (MeV)]. Because exposure is a measure of ionization in air and our interest is in biologic material, another type of measure is needed.

The concept of absorbed dose was developed to describe radiation uniformly for all types of ionizing radiation, at all energies, and in all materials. Absorbed dose has been defined as the amount of energy absorbed per unit of mass. Currently, the most common unit of absorbed dose is the gray (Gy); 1 Gy is equal to the absorption of 1 J/kg. The Gray replaced the rad, which is equal to the absorption of 100 erg/g (10 μJ/g); since 1 J is equal to 10^7 erg, 1 Gy is equal to 100 rads. The centigray (cGy) is a commonly used unit and is one-hundredth of the value of 1 Gy (i.e., 1 cGy = 1 rad).

Types of radiation used in clinical practice

In clinical practice, radiation can be administered in two different ways, (i) external beam and (ii) brachytherapy.

External beams

External beam therapy can be categorized as photons (e.g. X-rays and g-rays), electrons, or special beams (e.g. neutrons, protons). Most external beam machines deliver treatment from a distance of 80 to 100 cm from the patient. A linear accelerator (the most common external beam treatment machine) is a good example for describing radiation-dose delivery.

High energy electrons generated in a linear accelerator through a wave guide are accelerated into a metal target and are abruptly brought to a stop, resulting in a conversion of the energy of the electron to an electromagnetic waveform known as an X-ray. Linear accelerators in clinical practice produce X-rays with energies between 4 and 30 MeV. These beams have a very uneven distribution so they are passed through a flattening filter that produces a uniform dose intensity across the beam. The X-rays then enter the body and interact with the orbital electrons of atoms in the tissue. These electrons, known as secondary electrons, carry away a portion of the X-ray's energy and are also propelled through the tissue, interacting with other atoms to cause multiple ionizations. With energies greater than 1 MeV, the secondary electrons have sufficient momentum to travel from a few millimeters to a few centimeters (for higher energies) through the tissue. At a certain depth in the tissue, an equilibrium exists between the number of electrons being produced from more superficial depths and the number whose energy has been fully expended. This is the depth in tissue where the maximum dose is delivered and is called d_{max} (the depth of maximum dose). Because the intensity of the X-ray beam is attenuated when it interacts with the tissue, the number of interactions will also decrease at greater depths. As a result, a fall off in dose occurs beyond the d_{max} depth. In addition, at greater distances from the X-ray target in the machine, there is also a decreased intensity of the beam and thus a decreased dose.

In order to demonstrate these concepts, beam characteristics can be plotted on graphs showing the distribution of dose over depth in tissue (per cent depth dose curves) and space (isodose curves). For a per cent depth dose curve, the dose is measured at various depths along the central axis of the X-ray beam in the tissue and recorded as a percentage of a specific reference value, the dose at d_{max}. A percentage depth–dose curve for a 6-MeV linear accelerator is illustrated in Fig. 1. For a 6-MeV beam, the d_{max} is located at a depth of approximately 1.5 cm into the tissue. The portion of the curve from the skin surface to d_{max}, represents the area where the number of X-ray interactions with orbital electrons is increasing. As a result, dose is ‘built up' in this area. This build-up of dose is responsible for the skin sparing that is seen with linear accelerators, since the maximum dose is not observed at the skin surface but rather deeper in the tissue. With an increase in X-ray beam energy, d_{max} is even deeper in the tissue and the higher-energy beams also have greater depth of penetration. For example, for the same size of treatment field, a

6-MeV X-ray beam would deliver 67 per cent of the d_{max} dose at a depth of 10 cm compared to 77 per cent delivered by a 15-MeV X-ray beam. The differences in dose build-up and depth of penetration have implications for treatment, depending upon the location of the tumor. For example, a 15-MeV beam may better treat a deep-seated tumor such as a rectal cancer, while a 6-MeV beam may better treat a more superficial tumor such as a breast cancer.

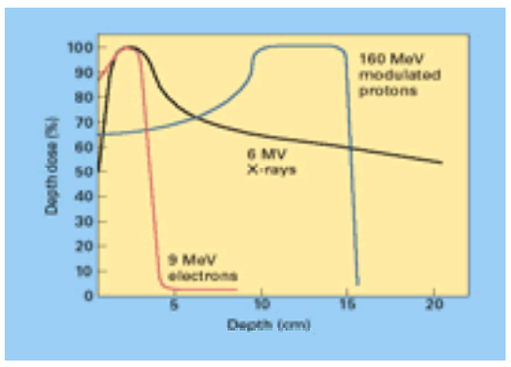


Fig. 1. Depth dose curves along the central axis for 9-MeV electrons, 6-MeV X-rays, and a 160-MeV modulated-energy proton beam.

Linear accelerators can also produce beams of electrons by removing the metal target used to make X-ray beams. Electrons are negatively charged particles with a small mass (approximately 1/2000th the mass of a proton); they therefore have a finite range in tissue, unlike X-ray beams, which lose energy gradually as they traverse tissue. For 9-MeV electrons (also shown in Fig. 1), there is a rapid dose build-up at the skin surface and a steep fall-off of dose with depth. Beyond a depth of 5 cm, virtually no dose is delivered. Depth of penetration is dependent on the energy of the electron beam, with higher-energy beams having greater penetration. Clinically, electrons are used to treat tumors near the skin (typically less than 5 cm deep), and where one does not wish to affect deeper normal tissue. Due to their lack of skin-sparing effects, electrons are rarely used as the sole means of treatment except for cutaneous malignancies.

Special beams such as protons and neutrons are also utilized clinically at a few institutions worldwide. Protons have physical advantages because of their dose distribution. They are charged particles that decelerate in tissue and then have a greater likelihood of interacting with nearby atoms. As a result, protons transfer most of their energy late in their path, resulting in an abrupt peak of ionizations near the end of their path. When modulated for clinical use, they can have a flat dose distribution across the range of interest and a steeper fall-off at the end of the range than electrons. A graph of percentage depth dose for a modulated, 160-MeV proton beam is shown in Fig. 1; this has been used clinically to treat lesions extremely close to important structures, such as uveal (ocular) melanoma and tumors of the skull base.

Neutrons have a dose distribution similar to that of X-rays. They have a role as a clinical modality primarily because of a radiobiological advantage that they have killing cells that are less dependent on oxygen (see below). Clinically, they have been used to treat a wide variety of tumors but perhaps have the greatest advantage in salivary gland tumors and certain sarcomas.

Brachytherapy

Brachytherapy is a method of treatment in which radioactive sources incorporated into intracavitary or interstitial devices are used to deliver radiation over short distances. This method takes advantage of the inverse-square law, which states that, for radiation originating from a point source, the intensity of the radiation to a given position varies inversely with the square of the distance between the source and that position. Therefore, the dose delivered at 2 cm from a point source is one-quarter of that delivered at 1 cm. A high dose of radiation is thereby delivered near the source but a low dose just a few centimeters away, allowing for excellent protection of normal tissues not at risk for tumor involvement.

Radioactive sources for brachytherapy are characterized by their activity, half-life, and type of radiation emitted. The activity describes the rate of decay of a radioisotope. Previously, activity was measured in Curies (Ci), with 1 Ci equal to the number of disintegrations (**dis**) in 1 s of 1 g of radium-226 (3.7×10^{10} dis/s). This measurement has been replaced by the becquerel (Bq), which is defined as 1 dis/s. The half-life of a radioactive substance is defined as the amount of time required for the activity to decay to half its initial value.

Brachytherapy implants can be either permanent or temporary. Permanent implants can be placed at the time of surgery and secured in place in or near the tumor. Commonly used, permanent radioactive sources are palladium-103 and iodine-125 because of their low energy and relatively short half-lives (17 days and 60.2 days, respectively); these are short in comparison to the original brachytherapy source, radium-226, whose half-life is 1600 years. The very low average energies of palladium-103 and iodine-125 (0.021 and 0.028 MeV, respectively) allow permanent placement with minimal concern about radiation shielding, since almost all of the radiation is absorbed in the patient. Currently, permanent implants are used at sites such as the brain, head and neck, and prostate.

In temporary implants, a hollow container or catheters are placed near or in the tumor at the time of surgery and are later loaded with radioactive sources for 2 to 5 days. After the treatment has been delivered, the whole apparatus is removed. The radioactive isotopes used for temporary implants tend to have longer half-lives and higher energies than those used for permanent implants. Cesium-137 is the most common isotope used for temporary intracavitary implants in gynecologic malignancy: it has a half-life of 30 years and an energy of 0.662 MeV. Iridium-192 has a half-life of 74.2 days and an average energy of 0.38 MeV; it is used frequently for interstitial implants. These implants can deliver effective doses of radiation (1500–4000 cGy) at dose rates of 40 to 60 cGy/h over 1½ to 5 days.

Clinically related radiation biology

Interaction with biologic material

When radiation interacts with biologic matter, the physical events of ionization described previously are completed in less than a second. The cascade of biological events triggered by ionization will occur over hours, days, weeks, months, and even years. The free radicals interact with molecules and induce alterations in nucleic acid structure, such as changing or losing a base: this can cause single- or double-strand breaks in DNA; it can also lead to changes in protein and/or lipid structures. This radiation-induced damage can affect the reproductive integrity of the cell such that, at the time of an attempted division, it will lose viability and die. At times, the cell death will be delayed for a number of cycles, but the ability of the cell to undergo prolonged division will still be prevented by the initial ionization and DNA-damaging events. The response of a tissue or organ to this damage leads to the acute and late effects seen clinically in both tumor and normal structures.

Apoptosis can also occur, but not in all cell systems: this is a physiological process in which a cell pursues a genetically determined sequence of events leading to death after being exposed to certain stimuli. The precise mechanism of apoptosis remains unclear, but it can be associated with DNA damage, deprivation of cytokines such as interleukin 3, or cell injury mediated by tumor necrosis factor or Fas receptors. The tumor suppressor gene *p53* and the bcl-2 proteins have a role in the regulation of apoptosis in both tumor and normal tissues.

Radiation-induced damage to a cell can be lethal or non-lethal. A lethal response is the loss of the ability to produce a continuously expanding progeny. Before their eventual lysis, non-viable cells may retain their metabolic integrity and may even divide several times. Although such tumor cells may persist in the tissues for prolonged periods, they are of little consequence for the patient. Non-lethal damage can be repaired, with the cell returning to its preradiation normal or near-normal status.

The factors determining the clinical response of a tissue or organ to radiation include the number and proportion of cells that lose their viability, the tissue's cellularity and its stroma, the time course for pyknosis and lysis of the killed cells, the time course for the removal of cell debris, and the proliferative activity and functional integrity of surviving cells. The time of appearance of gross or SYmptomatic change is principally a function of the cellularity and proliferation kinetics of the specific tissue constituents. For example, changes in bone marrow stem cells are prompt because the marrow comprises large numbers of actively dividing cells that respond rapidly to the radiation insult. In contrast, changes in mature muscle may take many years to develop, since muscle cells do not divide and the damage may be more related to long-term vascular effects than to a direct effect on muscle cells. Radiation effects may thus become apparent in different tissues at different times, even though identical proportions of cells were killed.

Cell-survival curves and descriptors of radiation sensitivity

An understanding of the relation between radiation dose and the probability of cell kill is essential. Cell survival and dose can be plotted on a log-linear scale, with the surviving fraction of cells plotted on the log (*y*-) axis and the dose plotted on the linear (*x*-) axis (Fig. 2). The general shape of the curve has two parts, (i) a shoulder and (ii) an exponential killing region. The shoulder is the curved part in the low-dose area; there is less efficiency of cell killing in this region than in the remainder of the curve, meaning that each increment of radiation dose kills a smaller percentage of cells than at the higher doses. The straight portion of the survival curve (on the log-linear plot) represents the region where cell killing occurs at an exponential rate. The linear nature of the plot means that the same fraction of cells is killed (but not the same number) for each given dose increment.

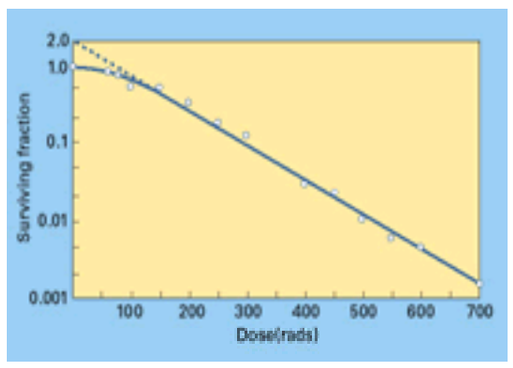


Fig. 2. The radiation cell survival curve, demonstrating an initial shoulder and then a straight line for data plotted on a log-linear graph.

Based on the characteristics of the survival curves, mathematical models and theories have been developed to compare and describe cell killing. One model, the linear-quadratic, was based on the observation that the characteristic survival curve in the region of 1 to 3 Gy per fraction appeared to fit a model where the surviving fraction (SF) was directly related to both the dose and the square of the dose:

SF = *e* to the power of [-(œD + bD^2)], where *D* equals dose.

In this model, ÏD represents the linear component and bD^2 the quadratic (exponential) component of cell killing. There have been attempts to explain these observations from a radiobiological perspective. It has been argued that the ÏD component corresponds to a single lethal event, while the bD^2 component represents cell killing from two separate, interacting events. Because the bD^2 component requires interacting events, it may be repairable under favorable conditions.

When the linear component (ÏD) of cell killing is equal to the quadratic component (bD^2), the ratio Ï/b can be described in terms of *D*. The Ï/b ratio has been widely accepted for describing isoeffects of cellular responses to different dose fractionation schemes. In the clinical setting, it has been observed that the ratio has a value in the range of 1 to 5 for late responding tissues (such as connective tissue and vasculature), and 6 to 20 for early responding tissues (such as mucosa and epithelium) and most tumors. Because of their smaller Ï/b ratios, late responding normal tissues appear to be more sensitive to high doses per fraction. Thus, there may be a clinical benefit of damage to tumor rather than normal tissue from a lower dose per fraction.

Repair of radiation damage

It has been known clinically for many years that when treatment is divided into multiple, small fractions, more total dose can be tolerated by the normal tissues. This is because, as shown in Fig. 3, with multiple fractions the shoulder of the survival curve is reconstituted between fractions of radiation. As a result, if a total dose is delivered in divided doses, it achieves a similar cell kill to a single fraction of a much lower total dose. This occurs because part of the radiation damage produced in a single fraction can be repaired between fractions and thus has no permanent biologic effect. If, however, additional radiation dose is delivered at the same time as the first dose, the extra damage produced can convert the previous sublethal injury into a lethal event.

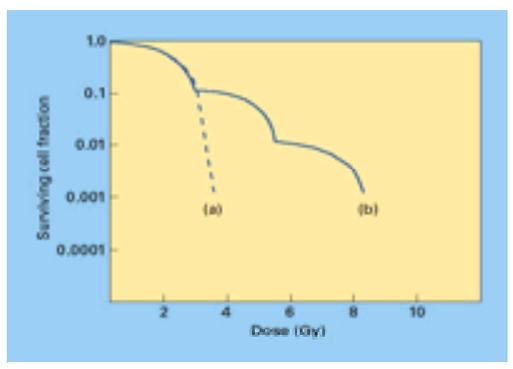


Fig. 3. Comparison of a single-fraction cell-survival curve (a) with a multiple, small fraction survival curve (b); note that while more total dose is delivered with the multiple fractions, the surviving fraction is similar for both curves.

With conventional fractionation schemes (approximately 180–200 cGy/fraction), late responding tissues tend to show greater capacity for repair than acutely responding tissues. While the precise mechanism is unclear, the clinical finding is that late responding tissues are spared more by dose fractionation than are acutely responding tissues. Most tumor cells, as well as mucosa and skin, are considered acutely responding, while most other normal tissues act in a similar way to late responding tissues. As a consequence, with the use of small doses per fraction (approximately 100–200 cGy), a clinical advantage could be achieved by killing tumor cells while sparing long-term morbidity.

Repopulation

If a fractionated course of radiation treatment is given over a prolonged period of time, the remaining cells can divide in between fractions, making it difficult to control the tumor. For acutely responding normal tissues such as mucosa and bone marrow, this repopulation or increased proliferation allows for tissue maintenance. Therefore, protraction of treatment times results in a reduction in acute toxicity. However, repopulation also occurs in tumor cells, and if treatment is too protracted the likelihood of tumor control will decrease because of the need to kill more cells in total.

The repopulation of tumor cells can be of great clinical importance. Withers reviewed the literature on head-and-neck tumours and concluded that after 4 weeks of daily fractionated radiation, significant repopulation occurs and a prolongation of treatment time requires additional radiation dose to overcome the tumor proliferation. Other studies on head-and-neck, lung, and cervical cancer suggest that prolonged treatment times have an effect on local tumor control when total doses are not increased. Concern about accelerated repopulation has been an impetus to consider many altered fractionation schemes.

Oxygen effect

The effect of oxygen on radiosensitivity has been recognized for almost as long as radiation has been used clinically. In 1909, Swartz observed diminished skin reactions to X- and g-rays when the skin was either subject to a strong vacuum or compressed by heavy weights; this effect was attributed initially to decreased blood flow.

In 1955, Thomlinson and Gray performed a histologic study of bronchial carcinomas, which produced indirect evidence that human tumors may contain hypoxic cells. These tumors usually consisted of long ‘cords' with the vascular supply on their periphery. After observing a large number of specimens, Thomlinson and Gray concluded: (1) no tumor cord with a radius of less than 160 µm showed evidence of necrosis; (2) no tumor cord with a radius of more than 200 µm was without a

necrotic center; and (3) regardless of the size of the necrotic center of the cord, the thickness of the rim of seemingly viable cells was never greater than 180 µm, all of which was attributed to the lack of diffusion of oxygen to the center of the cord producing hypoxia and necrosis. In later years, the estimated distance of oxygen diffusion in respiring tissue was decreased to approximately 70 µm. It was also hypothesized that between the viable, fully oxygenated cells and the necrotic cells, there was a region low in oxygen tension that was radioresistant because of the hypoxia. Hypoxia may also exist because of intermittent blood flow in the tumor vasculature.

It is also well established by *in vitro* studies that cells irradiated under hypoxic conditions are less radiosensitive than cells irradiated under fully oxic conditions. A qualitative measure of the different sensitivity is the oxygen enhancement ratio, defined as the ratio of radiation doses under hypoxic conditions to those under aerobic conditions required to achieve the same biological effect. The oxygen enhancement ratio is also affected by variables such as the type of radiation. For example, neutrons have an oxygen enhancement ratio of approximately 1.6 while X-rays have one of approximately 2.5.

Oxygen enhancement ratios of 2.5 are unlikely to confer an ability to control a tumor containing a small number of hypoxic cells (approximately 1 per cent of the tumor volume), since after a few radiation fractions nearly all the oxic cells would be killed and only the radioresistant hypoxic cells would remain. However, it is known that a process called reoxygenation occurs: this is defined as the re-establishment, between radiation dose fractions, of approximately the same hypoxic fraction in a tumor as was present before the first dose. The proposed mechanisms are either that shrinkage of the tumor decreases its overall oxygen utilization, which then allows for diffusion of oxygen to the surviving, previously hypoxic cells, or that the intercapillary distance decreases, allowing for increased diffusion to the hypoxic cells.

Values for pO_2 measured in head-and-neck and uterine cervical cancers have been correlated with the extent of local control achieved by radiation therapy. Tumors with substantial hypoxia had poorer prognoses than those with small amounts of hypoxia. A difference in survival based on the presence/extent of hypoxia has also been demonstrated for patients with carcinoma of the uterine cervix treated with surgery alone, so the oxygen effect may not be specific for radiation therapy.

Effects of cell cycle and redistribution

Radiation sensitivity is a function of the position of the cell in the replication cycle (Fig. 4). Although there is variation between cell lines, cells in the G_2- and M-phase are most sensitive, while those in the late S-phase are the least sensitive. Cells in G_1- and early S-phase show intermediate sensitivity. Proliferating cells also are more sensitive than G_0 cells of normal tissue and quiescent cells of tumors. The effectiveness of a radiation dose therefore depends on the distribution of cells through the cell cycle and the variation of sensitivity with age.

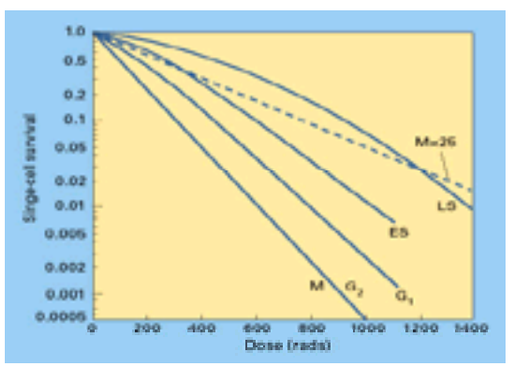


Fig. 4. Cell survival curves determined on V79 Chinese hamster cells as the indicated stages of the replication cycle.

After a single dose of radiation, the cells in the sensitive phases of the cell cycle are killed preferentially. Before a second dose of radiation, the cells in the more resistant phases of the cycle will proceed to more sensitive phases, resulting in additional cell killing. This redistribution of cells to the more sensitive phases may contribute to a similar proportion of cell killing with each dose of radiation rather than an increasing radioresistance.

Clinical issues with respect to the physical and biological factors

Normal tissue tolerances

Any attempt to minimize injury to normal tissues while maximizing tumor control requires knowledge of the radiosensitivity of tumor cells and the doses tolerated by normal tissues. In the majority of clinical settings, the radiation oncologist treats to the tolerance dose of the normal tissue. The reactions of normal tissues have been arbitrarily divided into acute responses, which usually occur during treatment or within 3 months of its completion, and late responses, which occur after that period.

The tissues that respond by producing acute side-effects are often different from those that show late side-effects. Early- and late-responding tissues differ in their cell proliferation kinetics. Stem cells of early-responding tissues are capable of rapid division and often require continued proliferation for their function. Examples are mucosa, intestinal epithelium, skin, germ cells, and bone marrow. For example, the initial radiation doses (approximately 200-cGy fractions for five daily fractions) will damage the mucosa, causing partial denudation and early symptoms of acute mucositis. Re-epithelialization occurs as the remaining mucosal stem cells proliferate rapidly and reconstitute the denuded surface, with resolution of the clinical symptoms. Stem cells in some tissues, such as periosteum and liver, only proliferate in response to damage.

Late-responding tissues are either slowly proliferative in their entirety or reflect damage in a subpopulation of slowly proliferating cells. The damage incurred in these tissues is often the result of damage to the vasculature of the organ. An example of a late-responding tissue is the spinal cord. After a large dose of radiation, clinical signs of myelitis develop approximately 4 to 18 months later. Histological changes correlating with the myelitis include demyelination and arteriolar vascular damage, suggesting that the cause of the damage is associated with oligodendrocytes, vascular endothelium, and Schwann cells. Late effects can also be seen in tissues that are typically rapidly proliferating, such as the oral mucosa. Years after completion of treatment, mechanical injury to the mucosa can result in ulceration and necrosis as a result of ischemia.

Late effects can also be caused by depletion of the stem cell renewal system. Bone-marrow hypoplasia or aplasia can be seen in areas that have received radiation just after completion of treatment and years later. In this case, the hemopoietic cells are either absent or reduced in number, and the change does not appear to be related to a vascular effect, given the presence of sinusoids in the aplastic marrow. Typically, late effects are the dose-limiting factors in radiation therapy.

The volume of an organ treated also is of clinical importance. For example, a patient will tolerate radiation much better if only a small part of the mouth rather than the entire oral cavity and pharynx is treated, because a smaller area of mucositis develops. An organ's functional reserve also needs to be considered. In a patient with lung cancer and marginal pulmonary function, treatment can result in permanent functional damage if the volume of normal lung within the radiation field is too large.

The tolerance doses (**TD**) of normal tissues were systematically analyzed by Rubin and Cassarett in 1972. They described radiation injuries in terms of TD5/5 (the probability of a 5 per cent complication rate within 5 years of treatment) and TD50/5 (the probability of 50 per cent complication rate within 5 years). Listed in Table 1 are estimates of the tolerance doses using standard fractionation (200 cGy per fraction, 5 days per week) for a variety of organs and volumes. However, clinical judgment remains a large part of determining treatment variables.

Organ	TD5/5, values		TD50/5, values	Endpoint			
	1/3	2/3	1/3	2/3	1/3	2/3	
Brain	60 Gy	50 Gy	65 Gy	75 Gy	60 Gy	60 Gy	Necrosis
Spinal cord	50 Gy	47 Gy	60 Gy	70 Gy	55 Gy	55 Gy	Myelitis
Optic structures	As partial volume		55 Gy	As partial volume		65 Gy	Blindness
Eye							
Lens			10 Gy			18 Gy	Cataract
Retina			45 Gy			65 Gy	Blindness
Brachial plexus	60 Gy	55 Gy	60 Gy	70 Gy	55 Gy	55 Gy	Neuropathy
Parotid		50 Gy	50 Gy		45 Gy	50 Gy	Xerostomia
Mandible	65 Gy	60 Gy	60 Gy	70 Gy	60 Gy	70 Gy	Necrosis
Larynx	70 Gy	60 Gy	70 Gy	80 Gy	60 Gy	80 Gy	Necrosis
Lung	30 Gy		17 Gy			24 Gy	Pneumonitis
Whole organ	30 Gy			45 Gy			Fibrosis
Heart	60 Gy	45 Gy	40 Gy	70 Gy	55 Gy	50 Gy	Pericarditis
Esophagus	60 Gy	55 Gy	55 Gy	75 Gy	55 Gy	68 Gy	Stricture
Stomach	60 Gy	55 Gy	50 Gy	70 Gy	65 Gy	65 Gy	Ulceration
Small intestine	50 Gy		45 Gy	55 Gy		50 Gy	Ulceration
Liver	40 Gy	35 Gy	30 Gy	50 Gy		40 Gy	Liver failure
Kidney			18 Gy			28 Gy	Nephritis
Prostate/cervix			120 Gy			65 Gy	Fistula
Kidney			23 Gy			28 Gy	Nephritis
Bladder		60 Gy	65 Gy	Not known			Contracture

Table 1 Normal tissue tolerances to irradiation.

The therapeutic ratio

The therapeutic ratio is the relation between the radiation dose effect on normal tissue compared to the effect on the tumor. As shown in Fig. 5, the probability of injury to normal tissue or tumor control can be plotted on the *y*-axis against dose on the *x*-axis. In most circumstances, the probability curves for injury to normal tissue and tumor control are sigmoidal, with very small changes in probability at both very low and very high doses. When the curve for tumor control is far to the left of that for injury to normal tissue, a tumor can be cured with minimal or no damage to normal tissue. When the two curves are close together or when the curve for tumor control is to the right of the curve for injury to normal tissue, it is more likely that complications will occur in normal tissues without obtaining good tumor control.

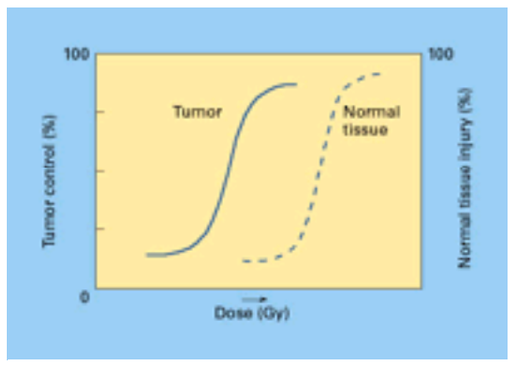


Fig. 5. The therapeutic ratio: in this diagram there is a favorable therapeutic ratio with a minimal risk for injury to normal tissue with a high probability of tumor control.

In clinical practice, a variety of factors affect the therapeutic ratio, including the radiosensitivity of the tumor, its volume and location, the tolerances of normal tissues, the total dose, the fractionation scheme, the technique for irradiation, and the use of combined modalities. Attempts to improve the therapeutic ratio can be based on either moving the curve for injury to normal tissue to the right, often by implementing technical changes in the radiation therapy that allow better targeted beams, or on moving the curve for tumor control to the left, usually by biological manipulations. The major improvements in the therapeutic ratio, and thus the major advances in the use of radiation therapy over the past 50 years, have been based on improvements in the equipment and technology, thereby moving the curve for injury to normal tissues to the right so that higher doses are tolerated.

Physical means to improve the therapeutic ratio

The basic strategy in radiation oncology is to define the target volume in three dimensions; to design an array of radiation beams that will encompass the target volume; to position and immobilize the patient in order to target the beam with a high degree of accuracy; and to employ a dose and fractionation protocol that has the greatest probability of killing the tumor while causing tolerable changes in normal tissues. Such a strategy is the basis for radiation treatment planning.

In order to improve the therapeutic ratio, a clear definition of the location of the tumor and the volume thought to be at risk for tumor involvement needs to be determined. This volume, defined by the clinician, is known as a clinical target volume (**CTV**). Failure to define properly the treatment volume will result in either missing the tumor or treating normal tissue unnecessarily. In practice, a planning target volume (**PTV**) is used, which encompasses the clinical target volume with a margin for set-up and dosimetric variation to ensure adequate dose delivery to the clinical target volume. Diagnostic imaging, placement of surgical clips in the treatment field, physical examination, and an understanding of the patterns of spread of tumors help define the clinical target volume. Once the target has been defined, attempts are made to spare normal tissue while conforming the radiation dose to the clinical target volume.

The delivery of treatment fields for radiation therapy is usually based on marks made on the skin. However, such alignment can produce substantial error since the skin and the deep structures are not directly related. Furthermore, internal structures are often mobile. Failure to consider set-up error and the mobility of internal organs could result in missing the target and unnecessarily irradiating normal tissues.

Immobilization

Immobilization techniques have been developed to limit external mobility during treatment and provide accurate repositioning. Through the use of individually designed molds, the target area can be secured in the same position daily. Tissue localizers (e.g. gold seeds), tattoos, and treatment-room lasers are also used to ensure reproducible set-ups. In head-and-neck cancers and primary brain tumors, head holders and bite blocks are often used to secure the head to provide reproducible positioning.

Simulation

Simulation is the part of treatment planning in which a patient is placed in the proposed treatment position on a diagnostic X-ray machine that has the exact same geometry as the treatment machine, so that treatment beams can be planned precisely. It is used to try to define the direction and shape of the radiation beam in relation to the patient and the tumor volume. Simulation confirms that the treatment field encompasses the tumor volume while sparing critical normal structures. Fluoroscopy is also used in simulation, and assists in determining the treatment set-up and evaluating internal motion (e.g. respiratory movement).

Three-dimensional treatment planning

In the era of computed tomography and magnetic resonance imaging, and of computer-based treatment systems, treatment planning in three dimensions is possible. Not only can structures seen radiographically in two dimensions be visualized in three dimensions but also the soft tissues usually invisible on radiographs can be seen and incorporated directly into the treatment planning process. Three-dimensional planning also allows for evaluation of dose distribution throughout the volume of tumor rather than in a single plane as is conventionally done. Radiation beams entering from arbitrary angles can also be used more easily and with greater accuracy to optimize the treatment plan.

In an attempt to deliver an optimal dose to the tumor and minimize the dose to the surrounding normal tissue, increasing the number of radiation beams, adjusting the dose contributed from each individual beam, and shaping the fields to conform to the target volume are routinely used. For example, two parallel opposed fields treat a central pelvic lesion adequately, with uniform dose to the tumor, but the dose to the peripheral tissues is higher than that to the tumor volume. When multiple beams enter the patient from a variety of directions, it is possible to reduce the dose to the surrounding normal tissues while maintaining an homogeneous dose distribution to the tumor.

Wedge filters and compensators

Where there is significant body curvature or the target volume is irregularly shaped, it is difficult to design radiation beams that provide an homogeneous dose over the target volume because of the varied attenuation of the beams. When critical normal structures are intermingled with the clinical target volume, this poses a problem because high doses to the normal structures can be above tolerance.

Tissue compensators or wedge filters are pieces of metal placed in front of the radiation beam that absorb the beam differentially based upon their thickness and thus are able to distort the isodose curves to a desired shape. They can compensate for body curvature or irregularly shaped targets and so provide improved dose homogeneity while limiting ‘hot spots' in normal structures. This is exemplified in the treatment of an early cancer of the retromolar trigone, where two beams at 90° are used (Fig. 6).



Fig. 6. A comparison of isodose curves with and without wedge filters of a cancer of the right retromolar trigone. Without wedge filters (a) there is significant dose inhomogeneity that could result in either exceeding the tolerance of the normal tissues of mucosa and bone that are within the treatment field or underdosing the tumor. With wedge filters (b), a more uniform dose is delivered that stays within the limits of normal tissue tolerances.

Brachytherapy

Brachytherapy delivers radiation to the target volume at close range. It delivers a high radiation dose near the radioactive source with a rapid dose fall-off. It has a therapeutic advantage where critical structures cannot be avoided effectively from an external-beam field or where larger doses are required for tumor control than can be administered safely with external beam alone. As described previously, a permanent or temporary afterloading device can be placed near or in the tumor. In the case of a bulky, stage IB uterine cervical cancer, the use of brachytherapy in addition to external beam can provide doses of radiation that can be effective in controlling the tumor. The external beam delivers a moderately high dose to the pelvic lymph nodes and the primary tumor, while the implant delivers a high dose ‘boost' to the primary tumor while delivering only low doses to the surrounding, sensitive, normal tissues. Because of easy access to the tumor, the rapid fall-off in dose, and the ability to displace critical structures away from the brachytherapy sources, this treatment can be very effective in delivering minimal tumor doses as high as 85 to 90 Gy with acceptable toxicity in normal tissues.

Biologic means to improve the therapeutic ratio

In the clinical setting, there are essentially four biologic ways in which the therapeutic ratio can be improved: (1) exploiting the radiobiologic differences between tumor and normal cells by altering the fractionation scheme; (2) introducing a modifier that selectively protects normal tissue from radiation damage; (3) introducing a modifier that selectively sensitizes tumor cells to radiation; and (4) combining radiation therapy with chemotherapy and/or surgery to reduce the volume of tumor.

Fractionation

Conventional fractionation of radiation in the curative setting in the United States is generally with 1.8 to 2.5 Gy per fraction, 5 days per week. This fractionation scheme has been utilized because of the tolerable acute side-effects, acceptable late effects, and documented local tumor control in many sites. However, because of biological differences between tumors, it is not always optimal. As a consequence, altered fractionation schemes have been utilized to improve the therapeutic ratio.

Hyperfractionation involves the use of smaller doses per fraction (approximately 110–120 cGy per fraction), delivered two or three times per day with the overall treatment time remaining about the same as in conventional fractionation. Higher total doses are delivered, but, because of the lower effectiveness per cGy at lower doses per fraction, this may not mean a greater biologic effect overall. A 6-h minimum interval is set between fractions to allow for repair of radiation damage. The assumption is that most tumors behave similarly to acutely responding tissues and therefore are more susceptible to total dose and treatment time rather than fraction size. At the same time, late reactions in normal tissues are more susceptible to fraction size and this is kept small. By reducing the fraction size, increasing the total dose, and keeping the treatment time the same, tumor control would be increased with no significant increase in late effects on normal tissues. Prospective and retrospective studies suggest a benefit to such a fractionation scheme in certain disease sites.

Accelerated fractionation utilizes conventional doses per fraction with a shortened overall treatment time (i.e., 4 to 5 weeks rather than 7 to 8 weeks for a head-and-neck cancer). With this regimen, acutely responding normal tissues experience more severe toxicity. The intent of this scheme is to reduce tumor cell repopulation. A variation on this fractionation scheme is accelerated hyperfractionation where there is a slightly smaller dose per fraction (1.5 or 1.6 Gy) delivered two or three times daily, a decrease in the overall treatment time, and generally a higher total dose. Randomized studies in limited-stage small-cell lung cancer, locally advanced non-small-cell lung cancer, and locally advanced head-and-neck cancers have demonstrated advantages in local tumor control.

Split-course radiation involves the use of a mid-course rest period; it is most commonly delivered with large doses per fraction (daily fractions of 250–300 cGy). The overall treatment time is approximately the same as with a conventional fractionation scheme. In general, the split course has the disadvantage of increased late effects secondary to the large fractionation size. Its primary advantage is convenience for the patient; and, if there is systemic disease progression, the second course of radiation treatment can be withheld. As a consequence, such regimens are rarely used for a curative approach in the United States and predominantly has been limited to palliation.

Hypofractionation uses large doses per fraction scheduled on fewer than 5 days per week. However, increased late effects of normal tissue are often seen with this regimen with no clear benefit in tumor control. Its use is reasonable when a patient has difficulty in receiving treatment daily or their life expectancy is short.

Radioprotectors

Radioprotectors are modifiers that attempt to provide normal tissues with a selective advantage over tumor cells by making them more radioresistant. Radioprotection has been attempted by chemical as well as physiological methods. The physiological attempts have tried to take advantage of the oxygen effect and the loss of radiation sensitivity by making normal tissues temporarily hypoxic by reducing the blood flow with tourniquets or vasoactive drugs. Chemical methods of providing radioprotection have focused on increasing the number of free-radical scavengers and increasing repair of free-radical damage in normal tissue. Given the free-radical scavenger effect of glutathione, it was reasoned that the addition of other thiol-containing compounds could have a radioprotective effect. Amifostine is an aminothiol that appears to be excluded from tumor cells and selectively protects certain normal cells. Other free-radical scavengers have been used, such as vitamin C, but have not demonstrated a benefit in randomized trials.

Radiation sensitizers and bioreductive agents

Radiosensitizers are agents that differentially increase the lethal effects of radiation in tumors compared to normal tissues. Clinically relevant sensitizers must demonstrate an increase in tumor control without an excess of complications in normal tissues. Two types of sensitizers have been utilized clinically, halogenated pyrimidines and hypoxic cell sensitizers. The halogenated pyrimidines (e.g. bromo- and iododeoxyuridine) are similar to the DNA precursor thymidines, and thus can be incorporated into DNA as a substitute for thymidine. This substitution either weakens the DNA structure, making it more susceptible to damage, or causes interference with the normal DNA repair process. As the percentage of thymidine replaced increases, the radiation sensitization increases. The halogenated pyrimidines are most likely to be of benefit when tumor cells are cycling more rapidly than normal surrounding tissues. They are considered ‘true sensitizers' because they have little or no

activity by themselves but produce an effect when combined with radiation. These are not being used in standard clinical practice.

Hypoxic cell sensitizers (e.g. misonidazole and nimorazole) exhibit their differential effect because most normal tissues are not hypoxic. By sensitizing the hypoxic component in tumors, one should be able to obtain improved tumor control with no significant increase in injury to normal tissue. In order for these sensitizers to function, they need to be present at the time of radiation. Most clinical trials with these agents have failed to demonstrate an improvement in local control. Possible reasons for these negative results include the failure to define tumors in which hypoxia is a clinical problem and the potential lack of a sensitizing effect of these drugs with conventional fractionation schemes. In one randomized clinical trial, nimorazole improved locoregional tumor control for patients with squamous cell cancer of the supraglottic larynx and pharynx. Other attempts at hypoxic cell sensitization currently under study involve altering hemoglobin affinity for oxygen or modifying blood flow to tumors.

Bioreductive agents kill hypoxic cells rather than sensitize them to radiation. It is believed that hypoxic cells metabolize these drugs to toxic intermediates that cause the lethal effects. In the presence of oxygen, these toxic intermediates are either not formed or unstable. By eliminating a resistant subpopulation of cells, there is a net sensitization of the tumor to radiation. Mitomycin C is a bioreductive agent used in the treatment of anal cancer, although it is unclear if its benefit is related to its hypoxic cell cytotoxicity.

Relation to other oncologic modalities

Primary radiation therapy

The selection of treatments for an individual patient is based on tumor biology, patterns of spread, and the patient's general health. Where known cure rates are similar for either surgery or radiation, the selection should be based on the functional and cosmetic results, the ability to salvage a local failure, and patient preferences. Comorbidity may place a patient at significant risk of operative death, in which case primary radiation may be a safer option.

Primary radiation therapy has a significant role when a tumor cannot be removed surgically because of its anatomic location or the involvement of critical normal structures. It is used for small tumors near or involving critical structures, and also for large, unresectable tumors.

Treatment of epithelial and mesenchymal tumors by radiation alone requires doses of 65 to 75 Gy given in fractions of 1.8 to 2 Gy five times a week. Lymphomas are usually treated with doses of 30 to 50 Gy; many germ-cell tumors require doses in the range of 20 to 40 Gy. Radiation treatment of early disease often produces local control and disease-free survival rates of over 90 per cent. Results are less satisfactory in patients with advanced disease and in those with tumors near or in sensitive structures.

When palliation of SYmptoms is the major goal of treatment, lower total doses are given over a shorter period of time. For example, bone metastases are often treated with fractionation schemes of 20 Gy in five fractions, 30 Gy in ten fractions, or 8 Gy in one fraction.

Combined radiation and surgery

The radiation dose required to inactivate a cell population increases with cell number. There is a low likelihood of radiation therapy alone being successful when epithelial and mesenchymal tumors are large, unless dose is increased to a level associated with significant complications in normal tissues. Often, the radiation dose required can be significantly reduced by resecting the primary tumor mass and using radiation to control the microscopic residual. The sequelae of either aggressive radiation and/or surgery can be avoided, with improved functional and cosmetic results from this combined approach. Clinically, such an approach is used for a variety of malignancies (Table 2).

For organ preservation
Breast
Rectal
Soft-tissue sarcoma
Bladder
For improved locoregional control
Head-and-neck cancer
Breast
Lung
Soft-tissue sarcoma
Rectal
Pancreas
Esophagus
Germ-cell tumors
Vulvar
Endometrial
Medulloblastoma
Positive margins at time of resection

Table 2 Sites where radiation and surgery are used in combination.

Radiation can be delivered pre-, post-, and intraoperatively. The rationale for preoperative treatment includes the ability to eradicate disease beyond or close to the surgical margins as well as the possibility of downstaging a patient so that the surgery can be changed to an organ-preserving procedure. The main disadvantages to preoperative therapy are the risk of prolonged wound healing and of potentially over- or undertreating a cancer because the clinical impression of disease extent differs from the pathological findings. The rationale for postoperative treatment is that if there is gross residual disease, its location will have been established and higher doses can be administered to that area. In addition, the surgery is not delayed and the tumour can be staged pathologically without the confounding effect of radiation. The disadvantages to postoperative treatment are that prolonged recovery from surgery could delay the start of radiation therapy, critical organs might be displaced in the radiation treatment area, and the vascular disruption from surgery might produce hypoxia and impair the effectiveness of the treatment. Intraoperative treatment has the potential to improve the therapeutic ratio in circumstances where external-beam therapy would be difficult to administer to high doses because of the presence of critical normal structures. The size of the target volume could potentially be reduced under direct visualization and sensitive structures could be shielded at the time of surgery. The disadvantages are that there are limitations to field size, access to the tumor bed can be difficult, and the high-risk areas may not be able to tolerate a large single dose of radiation.

Combined radiation and chemotherapy

Combinations of chemotherapy and radiation are used routinely to treat a variety of cancers. The rationale for combining these modalities is based primarily on a single modality's poor ability to control the tumor as well as the type of tumor and its potential sites of failure. The most common rationale is that chemotherapy may be able to obliterate occult metastatic disease outside of the treatment area, thereby improving the cure rate. However, combining chemotherapy with radiation has other advantages. A number of chemotherapeutic agents can sensitize tumor cells to radiation as well as treating systemic disease. Two common chemotherapeutic agents that act as radiation sensitizers are cisplatin and 5-fluorouracil. Cisplatin administered concurrently with radiation has produced improved survival in non-small-cell lung cancer, small-cell lung cancer, locally advanced head-and-neck malignancies, uterine cervical cancer, and esophageal cancer. Concurrent 5-fluorouracil has produced improved results in many gastrointestinal malignancies and head-and-neck cancers.

In addition, certain cancers are likely to contain chemotherapy-resistant cells. These resistant clones could proliferate and metastasize if not treated. For example, in early-stage, intermediate-grade lymphomas, some patients fail to respond completely to chemotherapy but can be cured with the addition of radiation treatment to the initial gross disease.

Summary

Radiation oncology is a specialty that has evolved over the last century and maintains a significant role in the management of cancer patients. It is based on scientific physical and biologic principles. Attempts to improve the therapeutic ratio have been made in the past 40 years with advances in the therapy equipment, radiosensitizers, radioprotectors, alterations in the fractionation schemes, and the use of combined approaches. As our basic scientific knowledge of cancer advances, the role of radiation to cure and palliate patients with cancer may improve.

Further reading

Al-Sarraf M *et al.* Progress report of combined chemoradiotherapy versus radiotherapy alone in patients with esophageal cancer: an intergroup study. *Journal of Clinical Oncology* 1997; **15**(1): 277–84. [A randomized trial demonstrating a benefit to a combined chemotherapy/radiation approach.]

Bloomer WD, Hellman S. Normal tissue responses to radiation. *New England Journal of Medicine* 1975; **293**: 80. [A good review of normal tissue responses to radiation.]

Brahme A. Dosimetric precision requirements in radiation therapy. *Acta Radiologica et Oncologica* 1984; **23**: 379–91.

Brizel DM *et al.* Patterns and variability of tumor oxygenation in human soft tissue sarcomas, cervical carcinomas, and lymph node metastases. *International Journal of Radiation, Oncology, Biology, Physics* 1995; **32**: 1121–5. [A clinical study addressing the effect of tumor hypoxia.]

Brock WA, Baker FL, Geara F, Goepfert H, Peters LJ. *In vitro* radiosensitivity of tumor cells and local tumor control by radiotherapy. [Abstract]. In: Chapman JD, Dewey GF, Whitmore GF, ed. *Congress abstracts*, Vol. 1, *Radiation research. a twentieth century perspective*, pp. 379–91. Academic Press, New York, 1991. [Provides information with respect to radiation sensitivity and tumor control.]

Brown JM. Evidence for acutely hypoxic cells in mouse tumours and a possible mechanism of reoxygenation. *British Journal of Radiology* 1979; **52**: 650–6. [Provides information regarding tumor hypoxia and the concept of reoxygenation.]

Dewey WC, Ling CC, Meyn, RE. Radiation-induced apoptosis: relevance to radiotherapy. *International Journal of Radiation, Oncology, Biology, Physics*, 1995; **33**: 781. [A good review of apoptosis and its relevance to radiation therapy.]

Elkind MM, Sutton H. X-ray damage and recovery in mammalian cells in culture. *Nature* 1959; **184**: 1293–5. [Provides information on radiation damage repair.]

Emami B *et al.* Tolerance of normal tissue to therapeutic irradiation. *International Journal of Radiation, Oncology, Biology, Physics* 1991; **21**: 109–22. [Comprehensive initial attempt at providing dose and volume data.]

Fisher B *et al.* Eight year results of a randomized clinical trial comparing total mastectomy and lumpectomy with or without irradiation in the treatment of breast cancer. *New England Journal of Medicine* 1989; **320**: 822–8. [Seminal article on role of radiation therapy in breast conservation.]

Gatenby RA *et al.* Oxygen distribution in squamous cell carcinoma metastases and its relationship to outcome of radiation therapy. *International Journal of Radiation, Oncology, Biology, Physics* 1988; **14**: 831–8. [Addresses the prognostic implications of tumor hypoxia.]

Goffinet DR *et al.* Bladder cancer: Results of radiation therapy in 384 patients. *Radiology* 1975; **117**: 149–53. [Provides bladder tolerance data.]

Goffinet DR, Choi KY, Brown JM. Combined effects of hyperthermia and ionizing radiation on the adult mouse spinal cord. *Radiation Research* 1977; **72**: 238–45. [Provides spinal cord tolerance data.]

Gray LH. The concentration of oxygen dissolved in tissue at the time of radiation as a factor in radiotherapy. *British Journal of Radiology* 1953; **26**: 638–48. [First to show histologic evidence of tumor hypoxia.]

Hall EJ. *Radiobiology for the radiologist*, 3rd edn. Lippincott, Philadelphia, 1988. [An excellent review of clinically relevant radiobiology.]

Halperin EC *et al.* A double-blind, randomized, prospective trial to evaluate topical vitamin C solution for the prevention of radiation dermatitis. CNS Cancer Consortium. *International Journal of Radiation, Oncology, Biology, Physics* 1993; **26**: 413–16 [A well designed negative trial attempting to use free radical scavengers as radioprotectors.]

Hellman S, Botnick LE. Stem cell depletion: an explanation of the late effects of cytotoxins. *International Journal of Radiation, Oncology, Biology, Physics* 1977; **2**: 181. [An attempt to explain late effects of radiation therapy.]

Hendry JH, Thames HD, ed. *Fractionation in radiotherapy*. Taylor and Francis, London, 1987. [an excellent review of fractionation.]

Hewitt HB, Wilson CW. A survival curve for mammalian leukemia cells irradiated *in vivo. British Journal of Cancer* 1959; **13**: 69–75. [Provides *in vivo* data on survival of irradiated cells.]

Hockel M *et al.* Intratumoral pO_2 predicts survival in advanced cancer of the uterine cervix. *Radiotherapy and Oncology* 1993, **26**: 45–50. [Addresses the prognostic implications of tumor hypoxia.]

Horiot JC *et al.* Hyperfractionation versus conventional fractionation in oropharyngeal carcinoma: final analysis of a randomized trial of the EORTC cooperative group of radiotherapy. *Radiotherapy and Oncology* 1992; **25**: 231–41. [A randomized trial showing a benefit to hyperfraction.]

Horwich A. *Combined radiotherapy and chemotherapy in clinical oncology*. Arnold, London, 1992. [A good review for combined chemoradiotherapy.]

Khan, F. *The physics of radiation therapy*, 2nd edn. Williams and Wilkins, Baltimore, 1994. [An excellent text for clinically relevant radiation physics.]

Kolstad P. Intercapillary distance, oxygen tension and local recurrence in cervix cancer. *Scandinavian Journal of Clinical and Laboratory Investigation* 1968; **106**: 145–57. [Addresses the prognostic implications of tumor hypoxia.]

Lawrence TS *et al.* An application of dose volume histograms to the treatment of intrahepatic malignancies with radiation therapy. *International Journal of Radiation, Oncology, Biology, Physics* 1990; **19**: 1041–7. [A trial that helped define liver tolerance to radiation.]

Lawrence TS *et al.* The use of 3-D dose volume analysis to predict radiation hepatitis. *International Journal of Radiation, Oncology, Biology, Physic*, 1992; **23**: 781–8. [A clinical trial using three-dimensional treatment planning.]

Little JB. Repair of sub-lethal and potentially lethal radiation damage in plateau phase cultures of human cells. *Nature* 1969; **225**: 804–6. [Provides convincing data on sublethal and potentially lethal radiation damage repair.]

McKenna WG *et al.* The role of the H-*ras* oncogene in radiation resistance and metastasis. *International Journal of Radiation, Oncology, Biology, Physics* 1990; **18**: 849–59. [Addresses the role of oncogenes in radiation resistance.]

Miller TM *et al.* Chemotherapy alone compared with chemotherapy and plus radiotherapy for localized intermediate- and high-grade non-Hodgkins lymphoma. *New England Journal of Medicine* 1998; **339**: 21–6. [A randomized trial demonstrating a benefit to combined modality therapy in lymphomas.]

Overgaard J *et al.* A randomized double blind phase III study of nimorazole as a hypoxic radiosensitizer of primary radiation therapy in supraglottic larynx and pharynx carcinoma. Results of the Danish Head and Neck Cancer Study (DAHANCA) protocol 5–85. *Radiotherapy and Oncology* 1998; **46**: 135–46. [A positive randomized trial using a radiation sensitizer.]

Powell S, McMillan TJ. DNA damage and repair following treatment with ionizing radiation. *Radiotherapy and Oncology* 1990; **19**: 95–108. [A good review of radiation effects on DNA and the repair process.]

Puck TT, Marcus PI. A rapid method for viable cell titration and clone production with HeLa cells in tissue culture: the use of X-irradiated cells to supply conditioning factors. *Proceedings of the National Academy of Sciences (USA)* 1955; **41**: 432–7. [The development of a human tumor cell line.]

Puck TT, Marcus PI. Action of X-rays on mammalian cells. *Journal of Experimental Medicine* 1956: **105**: 653–66. [An important article on the effect of radiation on cell survival.]

Puck TT *et al.* Action of X-rays on mammalian cells. II. Survival curves of cells from normal human tissues. *Journal of Experimental Medicine* 1957; **106**: 485–500. [Demonstrates cell survival curves after radiation.]

Sause WT *et al.* Radiation Therapy Oncology Group (RTOG) 88–08 and Eastern Cooperative Oncology Group (ECOG) 4588: preliminary results of a phase III trial in regionally advanced, unresectable non-small-cell lung cancer. *Journal of the National Cancer Institute* 1995; **87**: 198–205.[An excellent study evaluating the effects of combined chemoradiotherapy and different radiation fractionation schedules.]

Sinclair WK. X-ray induced heritable damage (small-colony formation) in cultured mammalian cells. *Radiation Research* 1964; **21**: 584–611. [Provides information on X-ray induced heritable damage.]

Sinclair WK. Cyclic X-ray responses in mammalian cells *in vitro. Radiation Research* 1968; **33**: 620–43. [Provides information on the effect that the cell cyle has on radiation response.]

Suit HD, Shalek RJ, Wette R. Radiation response of C3H mouse mammary carcinoma evaluated in terms of cellular radiation sensitivity. In: Shalek RJ *et al.*, ed. *Cellular radiation biology*, pp. 514–30. Williams and Wilkins, Baltimore, 1965. [An important article evaluating radiation sensitivity.]

Thames HD *et al.* Accelerated fractionation versus hyperfractionation: rationales for several treatment per day. *International Journal of Radiation, Oncology, Biology, Physics* 1983; **9**: 127. [A good review of the rationale for certain fractionation schemes.]

Thomlinson RH, Gray LH. Histological structure of some human lung cancers and possible implications for radiotherapy. *British Journal of Cancer* 1955; **9**: 539–49. [Addresses the possible histologic reason for tumor hypoxia.]

Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation Research* 1961; **14**: 213–22. [Addresses the radiosensitivity of bone marrow

cells.]

Todoroki T, Suit HD. Therapeutic advantage in pre-operative single dose radiation combined with conservative and radical surgery in different size murine fibrosarcomas. *Journal of Surgical Oncology* 1985; **29**: 207–15. [Raises questions with respect to appropriate dose and fractionation.]

Turrisi A *et al.* Observations after 5 year follow up of intergroup trial 0096: 4 cycles of cisplatin (P) etoposide (E) and concurrent 45 Gy thoracic radiotherapy (TRT) given in daily (QD) or twice-daily fractions followed by 25 Gy PCI. Survival differences and patterns of failure. [Abstract]. *Proceedings of American Society of Clinical Oncology* 1998; **17**: 457a. [A randomized trial demonstrating a survival benefit to an altered fractionation scheme.]

Veronesi U *et al.* Comparing radical mastectomy with quadrantectomy, axillary dissection, and radiotherapy in patients with small cancers of the breast. *New England Journal of Medicine* 1981; **305**: 6–11. [A randomized trial demonstrating the feasibility of organ preservation treatment.]

Walsh TN *et al.* A comparison of multimodal therapy and surgery for esophageal adenocarcinoma. *New England Journal of Medicine* 1996; **335**: 462–7. [A randomized trial demonstrating a benefit to combined modality therapy.]

West CM, Davidson SE, Hendry HJ, Hunter RD. Prediction of cervical carcinoma response to radiation therapy. *Lancet* 1991; **338**: 818.

Withers HR. Biologic basis for altered fractionation schemes. *Cancer* 1985; **55**: 2086–95. [An excellent review of the biologic basis for altered fractionation schemes.]

Withers HR, Taylor JMG, Maciejewski B. Treatment volume and tissue tolerance. *International Journal of Radiation, Oncology, Biology, Physics* 1988; **14**: 751–9. [Provides information on treatment volume and tissue tolerance.]

Yang JC *et al.* Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity. *Journal of Clinical Oncology* 1998; **16**: 197–203. [A randomized trial suggesting an improvement in local control with adjuvant radiation.]

48.7.2 Principles of radiation protection

Edward W. Webster

[Nature of ionizing radiation](#)
[Radiation dose and units](#)
[Effects and risks of radiation exposure](#)
[Effects and risks](#)
[Maximum permissible doses](#)
[Sources of X-ray exposure](#)
[Exposure from radioactive sources](#)
[Radiation safety practices](#)
[Radiography](#)
[Fluoroscopy](#)
[Radioactive materials](#)
[Further reading](#)

Nature of ionizing radiation

The radiation emissions produced by X-ray tubes and radioactive substances are known as ionizing radiation because of their ability to disrupt atoms and molecules. The ionizing radiations emitted by radioactive materials include g-rays (which are similar to X-rays), b-rays (which are high-speed electrons), and occasionally a-rays (which are heavy, positively charged particles not employed in medical procedures). Several other types of radiation in medical use such as microwaves, visible light (e.g. from most lasers), and ultrasound do not have the ability to ionize. Ionizing radiation can disrupt DNA in the cell nucleus, causing damage to genes and chromosomes. The ability to ionize is the basis of the unique damage to cells and tissues that X-rays and radioactive substances can create. The hazards of ionizing radiations rapidly became evident after their discovery a century ago, initially in the form of damage to the skin (erythema) and blood fractions, and later as cancer induction and genetic mutation.

Radiation dose and units

Radiation emitted from X-ray tubes and radioactive materials is a form of energy transmitted through space, as are heat and light. When it encounters an absorbing medium, such as the human body, the energy is gradually absorbed. A certain fraction of the energy is absorbed in each layer of tissue. The energy absorbed by a unit mass of absorber is referred to as the absorbed dose.

The passage of X-rays into the body is illustrated in Fig. 1. In this example, it is assumed that 20 per cent of the energy in an X-ray beam is absorbed in the first 1 cm of tissue, and 20 per cent of the remaining energy is absorbed in each successive 1-cm layer. The dose therefore decreases as the depth in tissue increases. When more energy enters the body (Fig. 1b) more energy is absorbed in each tissue layer. The dose received by portions of the body depends on the intensity of the energy flow (energy entering per second), the time for which it flows, and the depth of the tissue below the surface.

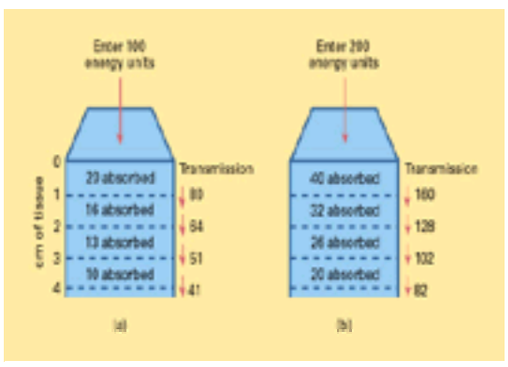


Fig. 1. Ionizing radiation energy flow and absorption in tissue block. A diagnostic X-ray beam with 20 per cent absorption/cm is assumed. The dose in the fourth centimeter is one-half of the dose in the surface layer. In (b) when the energy inflow is doubled, the dose in each layer is doubled.

The unit of absorbed dose universally used today is the gray (Gy), defined as energy deposition of 1 J/kg. This unit has replaced the earlier and smaller ‘rad' unit (= 0.01 Gy). One gray is considered to be a large dose since it can produce acute effects such as a reduction in blood count if received by a large fraction of the body. A typical dose delivered to a target organ in radiation oncology practice is 60 Gy over about 30 treatment days. For small doses such as those delivered in radiographic procedures, the subunit milligray (mGy = 0.001 Gy) is employed. The range of dose received by the skin of a patient from radiography is 1 to 100 mGy. Typical doses received by attending medical personnel are much smaller than those received by patients (e.g. 0.1 to 0.5 mGy/month).

For radiation protection purposes, a quantity known as the ‘equivalent dose' is in general use; this has a different unit, the sievert (Sv) and its subunit the millisievert (mSv). The term ‘equivalent dose' was introduced by the International Commission on Radiological Protection (**ICRP**) because some ionizing radiations, notably neutrons and a-rays, are biologically more effective per gray than X-rays, g-rays, and b-rays. The ‘equivalent dose' is the absorbed dose (Gy) multiplied by a radiation weighting factor w_R that expresses the relative biologic effectiveness. For example, for a-rays $w_R = 20$ and therefore an absorbed dose of 1 Gy gives an equivalent dose of 20 Sv. However, $w_R = 1$ for X-rays and radiations from radioactive substances used in medicine and therefore in this special case an absorbed dose of 1 Gy gives an equivalent dose of 1 Sv; that is the absorbed dose and the equivalent dose are identical. Because of their broad scope that can include neutron and a-ray exposures, maximum equivalent doses for occupational and public exposure to ionizing radiations that are specified in government regulations are denoted in ‘equivalent dose' units, namely the sievert and the millisievert (see Table 1).

Occupational exposures	*Period*	*ICRP* [a]	*NCRP* [b]
Whole body[d]	Annual	–	50
	Cumulative	100 in 5 years	10 × Age in years
Lens of eye	Annual	150	150
Skin, hands, feet	Annual	500	500
Fetus after diagnosis	Monthly		0.5
	Cumulative	1**	
Public exposures			
Whole body[d]	Annual	1	1 Frequent
	* 1/year		5 Infrequent
Lens of eye	Annual	15	15
Skin	Annual	50	50

[a] International Commission on Radiological Protection (1991).
[b] National Council on Radiation Protection, USA (1993).
[d] Limits for effective dose using tabulated weighting factors for organs in Table 1.
** At menstrual/abdominal skin its worth.

Table 1 Maximum permissible dose equivalents in millisieverts (mSv).

Effects and risks of radiation exposure

Radiation effects fall into two main classes: (i) deterministic, which occur only after exposure to large doses exceeding a threshold value and in essentially all such exposed persons after a sufficiently large dose; and (ii) stochastic, which may be caused by any dose (i.e. with no threshold), although the probability that they will

occur is strongly dose dependent and is extremely small at the lowest doses.

Examples of the first class are skin erythema, lens opacities, radiation sickness (nausea, vomiting), sterility, and even death. Effects in this class require doses greatly exceeding those associated with occupational exposure and most diagnostic radiation use. Acute radiation exposures of at least 3 Sv, 2.5 Sv, and 1 Sv are needed to produce a skin erythema, a radiation cataract, and radiation sickness, respectively. Much higher total doses are needed for these effects if the dose is protracted over long periods of time as occurs in occupational exposure. Examples of stochastic effects are the many types of malignant disease that become manifest after a latent period of several years or more after exposure and genetic anomalies that are expressed in future generations. These effects may be produced by the doses received from occupational exposure and diagnostic radiation use, and possibly but not detectably by the ambient background radiation of the order 1 mSv/year which everybody experiences. Fetal development defects may be included in this class of effects: their incidence increases with radiation dose to the fetus, but there are probably dose thresholds of about 0.1 Sv.

The most recent risk estimates for occupational whole-body exposure in a working population are those of the ICRP (1991) and the United States National Council on Radiation Protection (**NCRP**; 1993). The lifetime risk of fatal cancer following exposure to radiation at low dose rates or to intermittent small doses is estimated at 4 per cent/Sv. This translates into four extra deaths due to cancer in 10 000 persons exposed to a 10 mSv (0.01 Sv) dose over the whole body. The risk of serious genetic anomalies appearing in offspring of persons suffering gonadal exposure is estimated at 0.6 per cent/Sv; that is one child in 16 000 will be affected following a gonadal dose of 10 mSv to the parents. The risk of cancer developing after *in utero* exposure to radiation was most recently estimated by Doll and Wakeford to be similar to that of an exposed adult, at about 6 per cent/Sv or 1 in 1600 following a 10-mSv dose. Radiation exposure of a pregnant woman between weeks 8 and 15 of gestation appears to cause a loss of IQ in the child, estimated at 30 IQ points/Sv, but as noted above, a threshold dose of about 0.1 Sv probably pertains.

Effects and risks

The above risk estimates are based largely on the long-term study of the Japanese A-bomb survivors that started in 1950 under the leadership of the United States Atomic Bomb Commission, now the Radiation Effects Research Foundation (**RERF**). This study is limited to the effects including cancer induction of a single radiation exposure to about 100 000 persons with a wide range of radiation doses. Until recently, the many investigations of the health effects in radiation worker groups had inadequate statistical power to confirm or refute the RERF conclusions. However, within the last 5 years, the results of several meta-analyses of combined national groups have been completed. The largest of these studies was conducted with the support of the International Agency for Research on Cancer (**IARC**). It included combined nuclear industry workers numbering over 95 000 persons with an average work history of 22 years in the United Kingdom, Canada, and the United States. The results show no evidence of an association between radiation dose and mortality from all cancers, but mortality from leukemia was significantly associated with external radiation dose. The excess relative risk of 2.18/Sv for leukemia in this major study is compatible with that in the A-bomb study. However, this leukemia result is mainly due to the contribution of one large British facility (Sellafield) involved in chemical separations of radionuclides. Moreover, when the risk estimate was confined to cumulative doses less than 400 mSv (0.4 Sv) there was no significant elevation of leukemia.

Maximum permissible doses

The risk estimates for radiation-induced cancer in Table 1 are several times larger than those published by international review bodies in the early 1980s, partly due to the availability of long-term data from exposed populations, particularly the Japanese A-bomb survivors. There has also been downward revision of the A-bomb doses received, and a change in the projection model for future cancer mortality in study populations. The maximum permissible doses, particularly for lifetime exposure, have also been reduced recently. The most recent dose limits propounded by the ICRP, which guide most countries of the world, and by the NCRP in the United States, are shown in Table 1. The significant new limits from both organizations concern the lifetime limits, which are reduced by factors of 2.5 and about 3.5, respectively, and the limits for exposure of the general public, which are reduced by a factor of 5.

The method by which whole body dose is estimated has also been revised, primarily to take account of the different susceptibilities of internal organs to fatal malignant disease induced by radiation. The new ‘effective dose' method applies dose weighting factors to each exposed organ (Table 2); this is particularly relevant for partial body doses and for internal doses from radionuclides, which are usually markedly non-uniform. Leukemia and cancers of the colon, lung, and stomach are estimated to account for one-half (48 per cent) of the total detrimental effects of radiation.

Tissue/organ	Weighting factor
Bone marrow (red) (10 per cent head and neck)	0.12
Bladder	0.05
Bone surface	0.01
Breast	0.05
Colon	0.12
Esophagus	0.05
Gonads	0.20
Liver	0.05
Lung	0.12
Skin	0.01
Stomach	0.12
Thyroid	0.05
Remainder (e.g. brain, salivary gland)	0.05
Total	1.00

Table 2 Tissue weighting factors for estimating effective dose equivalents.

For occupational X-ray exposures when a lead apron is worn, the effective dose will be due primarily to head and neck dose, which is relatively lightly weighted compared with the dose to the shielded organs under the apron. In this situation the effective dose will be much smaller than the dose recorded by a dosimeter worn at neck or head level. If two radiation dose monitors (such as film badges) are worn at the neck (n) above the apron and the waist (w) under the apron, the effective dose is estimated to be ($0.5\ D_n + 0.025\ D_w$).

Sources of X-ray exposure

As a measure of the significance of a radiation exposure, the normal background radiation dose received by the average person is about 0.08 mSv/month, and the minimum dose above background detectable by a personnel film badge is 0.1 mSv.

In the operating room, X-ray exposure accrues from radiography and C-arm fluoroscopy. In both procedures exposure is almost entirely due to radiation scattered by the part of the patient's body that is in the primary X-ray beam. Leakage radiation that penetrates the housing of the X-ray tube has an intensity less than 1 per cent of that due to scattering.

The dose received from scattered radiation depends on several operational factors: (i) the dose received by the patient—the level of scatter at a given distance from the patient is proportional to the patient dose; (ii) the size of the X-ray field incident upon the patient—the scatter is proportional to the field area; and (iii) the distance of personnel from the patient during the use of the X-ray beam. This last factor is of major significance and depends on the ‘inverse square law' doubling the distance reduces the scatter dose by a factor of 4; a 10-fold increase in distance reduces the scatter by a factor of 100.

Because radiography often does not require personnel to be close to the patient, personnel exposures from this source are normally very small. Fluoroscopy, however, is typically conducted in close proximity to the patient and therefore the major exposures to operating personnel are received by the fluoroscopist. Doses received by operating room personnel from radiography are so low that personnel shielding (lead aprons) is usually not needed, except by persons such as anesthetists, who remain close to the patient in procedures that require multiple abdominal or skull radiographs. The scatter dose to a person 1 m away from a patient is typically about 0.0005 mSv (0.5 µSv) for a single chest film, but 0.005 mSv (5 µSv) for a single abdominal film. At a distance of 3 m the doses are nine times smaller.

Scattered radiation doses at a distance of 30 cm from the side of a patient during fluoroscopy can exceed 12 mSv/h with some orientations of a C-arm fluoroscope (Fig. 2(b)). Protective clothing should always be worn by personnel when a fluoroscope is in use.

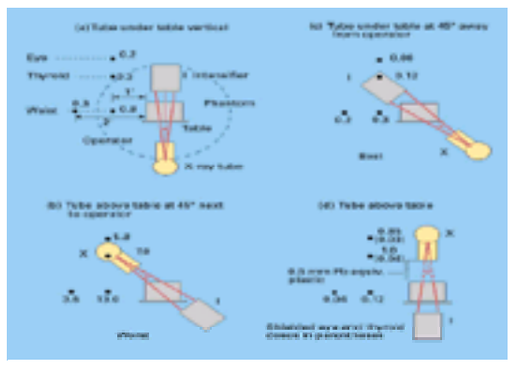


Fig. 2. Dose rates (mSv/h) to the operator of a C-arm fluoroscope for four angulations of the X-ray beam. In (a), (b), and (c) a phantom 30 cm × 30 cm × 20 cm high was employed with X-ray operating factors of 100 kV and 2 mA, with a focus to intensifier distance of 75 cm, and a field size of 15-cm diameter at the intensifier I. In (d) a phantom of similar dimensions was used with a larger field size of approximately 19 cm × 19 cm at the intensifier, but with a tube voltage of 80 kV, a tube current of 1 mA, and a focus to intensifier distance of approximately 120 cm. With these factors head and neck dose rates were about six times smaller. In this study dose rates were also measured with a 0.5-mm lead shield in place between the X-ray tube and operator. Eye dose was reduced by a factor of 20.

Digital image storage with immediate readout offers a convenient low-dose alternative to continuous fluoroscopy in a number of surgical procedures such as hip nailing in orthopedic surgery and electrode or needle positioning in neurosurgery, where a large number of images are required intermittently. This system, usually supplied as an optional accessory to a C-arm fluoroscope, provides a stationary image on a single television frame on demand. The patient dose per image is about 15 times smaller than would be required fora single conventional radiographic film and 30 times less than a fluoroscopic exposure of 1 s. The scattered radiation doses to surrounding personnel are likewise reduced. Typical scatter doses at 1 m from the X-ray axis for 100 images obtained with a 15-cm × 15-cm field at the image intensifier during a cerebral procedure are 0.3 µSv compared with 6 µSv during 1 min of fluoroscopy, a 20-fold reduction.

During the last decade there has been an increase in the use of fluoroscopes with above-table X-ray tubes and below-table image intensifiers for guiding the surgeon in some operative procedures. The versatile C-arm fluoroscope readily lends itself to any angular orientation, including those in which the tube is overhead. The physician's head and neck are necessarily in close proximity to the region of interest in these instances, and the scatter from the patient and the X-ray dose to the operator's head are both maximized. The eye lens, the red bone marrow in the skull and cervical spine, and the thyroid gland are particularly at risk under these circumstances.

To quote a 1985 report from the ICRP: ‘With the operator wearing a protective apron and standing beside the patient, the dose from an over-couch screening set, compared with that from an under-couch set, can be 250 times higher to the hands, 100 times higher to the eyes, and 35 times higher to the whole body. For an operator with a heavy workload the dose to the lens of the eye can greatly exceed the Commission's recommended limit of 150 mSv in a year.'

Figure 2 shows three orientations of a C-arm fluoroscope operating with the tube at 110 kVp and 2 mA, with a 15-cm diameter field size at the face of the intensifier. The dose rates at the eye, thyroid, and waist of the operator are shown at a distance of 30 cm from the side of a wax phantom 30 cm × 30 cm × 20 cm thick, and vary strongly with C-arm orientation. The eye doses vary by a factor of almost 100, the thyroid doses by a factor of 60, and the waist level doses by a factor of about 40. Figure 2(d) shows dose rates to the eye for a fixed fluoroscope with an over-table X-ray tube operating with smaller X-ray factors than the C-arm device and includes the effect of a transparent lead plastic shield hanging from the X-ray tube.

Computed tomography (**CT**) has recently been used intraoperatively, particularly during neurosurgery. The introduction of a mobile CT scanner (Philips Tomoscan M) has made possible the use of postoperative CT in intensive care units in addition to use in operating rooms. The scattered X-ray dose from a series of thirty 5-mm thick CT slices through the brain is of the order of 0.1 mSv at 1 m from the isocenter of the scan plane and 0.01 mSv at 3 m. With conventional CT use in a shielded room, the operator is located at a control station outside the room. In the operating room the attending staff could step behind a mobile lead shield during CT operation or retire to a distance of 3 m or more. The anesthesiologist, who may be closer to the patient, should wear a lead apron if the frequency of CT scans exceeds two/week.

When patients in an intensive care unit receive CT scans using a mobile CT unit, a patient in the adjacent room (who is probably 3 m distant and behind a gypsum board wall transmitting 30 per cent of the incident radiation) will receive a dose of 0.003 mSv (3 µSv). If there are three such scans in 1 week, with the same adjacent patient in the intensive care unit, the dose received would be about 0.01 mSv, which is trivial compared with the general public limit of 1 mSv/year, and therefore no shielding is necessary.

Intraoperative radiation therapy has increased in frequency during the last decade largely due to an expanding spectrum of qualifying malignant disease including cancers of the pancreas, brain, stomach, rectum, and vagina. This practice involves treatment of a tumor by external beam radiotherapy after it has been surgically exposed and, in some cases, after resection. Many major medical centers now have combined surgical/radiotherapy facilities employing very high energy X-rays or electrons. As in all radiotherapy procedures, medical staff do not remain in the room during the radiation treatment phase. Surgeons usually remain in the shielded area outside the therapy room, where they receive a very low and permissible level of occupational exposure.

Exposure from radioactive sources

Each year in the United States, radionuclides are administered about 10 million times for diagnostic purposes: about one patient in eight receives such an examination. A small fraction of patients admitted for surgery will therefore emit g-rays or characteristic X-rays, residual activity from a recent examination. These individuals are a source of radiation exposure to those persons near them for a period that depends on the effective half-life of the radiopharmaceutical in the body. By far the most commonly used diagnostic radionuclide is technetium-99m, which has a physical half-life of only 6 h: in about 2 days (8 half-lives) the radioactivity, even if not excreted, has essentially disappeared. Some radiopharmaceuticals, such as those used to image bone or kidney, are almost entirely excreted within 6 h of administration. Immediately after administration, however, the dose rate at the surface of the patient may range from 0.01 to 0.2 mSv/h; higher dose rates occur when relatively large amounts of radioactivity have been administered, such as 740 MBq (20 mCi) of technetium-99m for a bone scan or a radionuclide angiogram. The dose rate falls rapidly with distance from the patient: by factors of 6 and 30 at distances of 30 cm and 1 m, respectively, from the patient. At 1 m from a patient who has recently received the highest dose, the radiation level will not exceed 0.01 mSv/h.

If surgery is undertaken soon after administration of a radiopharmaceutical, manual contact is made with body fluids and organs that may contain appreciable radioactivity. For example, following the administration of 100 MBq (3 mCi) of technetium-99m sulfur colloid for a liver scan, the dose rate at the surface of the liver is about 0.6 mSv/h. Similarly, following the oral administration of 7.4 MBq (200 µCi) of iodine-123 for a thyroid scan, the dose rate at the surface of the thyroid gland approaches 0.6 mSv/h. Since surgery is rarely performed soon after a diagnostic test, the dose even within an hour of contact must be considered negligible in comparison with the maximum permissible hand dose of 500 mSv/year.

Although exposures to surgeons and nurses from diagnostic radionuclide use may be considered negligible, this may not be the case after administration of radioactive materials for treatment of cancer, or for thyroid ablation. Large doses of iodine-131 with a half-life of 8 days are conventionally used for the treatment of thyroid disorders. Quantities up to 370 MBq (10 mCi) are used to treat hyperthyroidism and up to 5500 MBq (150 mCi) to treat thyroid cancer. A hyperactive thyroid gland with a 50 per cent uptake could contain 185 MBq (5 mCi) during the first day after administration. Neck surgery on or close to the gland would involve g-ray dose rates to the hands at 5 cm from the gland of 4.3 mSv/h. Hand contact with the surface of the exposed gland would produce a finger dose of about 500 mSv/h, most due to the b-rays of low penetration. Surgical gloves reduce this dose by a factor of about 2. The doses received from a patient soon after the administration of a tumoricidal dose are much greater. Within the first few hours after receiving 5500 MBq (150 mCi), the dose/hour at the surface of the patient's abdomen is about 4.5 mSv; immediately over the thyroid it could be 250 mSv. At 30 cm from the body surface it is about 0.75 mSv, and at 1 m it is about 0.15 mSv. As much as 20 per cent of the administered activity may be taken up by metastatic disease in the lungs. The dose to the hands of a surgeon performing chest surgery in the first few days would be about 3 mSv/h if the hands are 10 cm distant from the metastases. Surgery would also involve massive radioactive contamination of the surgical field, instruments, and drapes from spilled blood; this is particularly serious on the first day, with markedly lower contamination levels over the next week. Fortunately about 70 per cent of the radioactivity dose used for cancer treatment is excreted in the first 24 h. It is clear that surgery, if indicated, should be postponed as long as possible.

Radiation exposure of the surgeon and to a smaller degree the operating room staff accompanies the dissection of tissue containing iodine-125 or palladium-103 seeds previously implanted during brachytherapy. The half-life of iodine-125 is 60 days and of palladium-103 it is 17 days. A good example would be prostatectomy within a few months after seed implantation. Surface doses at the prostate would be about 30 mSv/h during this period, with much lower doses at a distance of 30 cm. If, however, the contact time was only 1 min, the hand dose would be less than 0.5 mSv, again far below the hand dose limit of 500 mSv/year.

Radiation safety practices

The cardinal factors that enhance radiation safety are distance from the radiation sources, minimizing the exposure time of the operator or other personnel, and interposing shields between the source and the operator.

Radiography

Although doses are small, personnel who are not required to be close to the patient should retire to a distance of at least 2 m before each image is made. Protective clothing is advised only for persons who cannot remove themselves, and then only when a large series of images is to be made.

Fluoroscopy

Protective garments (aprons 0.3 to 0.5 mm of lead equivalent thickness, including thyroid shields) are mandatory. The operator should use equipment with an under-table tube, if clinically feasible, use an X-ray field as small as possible for visualization of objects of interest, minimize the beam-on time, set the control panel for automatic brightness control of the image (i.e. avoid the manual setting of voltage and current), and keep the image intensifier as close as possible to the patient.

If an over-table X-ray tube is employed (for example, with a C-arm fluoroscope), additional measures should be taken to reduce the head and neck dose of the operator. Ideally, a transparent shield of lead plastic with lead equivalent in the 0.3 to 0.5 mm range should be suspended from the X-ray tube housing at least on the operator's side of the tube; it should be long enough to intercept the scattered radiation arising from the surface of the patient at the X-ray beam entrance portal (Fig. 2d). This will reduce the eye dose rate by a factor of about 30 for 80 kVp X-rays. An alternative is to install a ceiling-mounted lead glass shield for shielding the head and neck. Shielding may be positioned in a flexible fashion between the surgeon and the over-table X-ray tube to intercept the scatter. If such a shield is not provided, an operator with a heavy fluoroscopic workload (more than 0.5 h/week of beam-on time) is advised to wear glasses (goggles) with lead glass 1 to 2 mm thick (0.25 to 0.5 mm of lead equivalent), and not ordinary glass or plastic.

Radioactive materials

Exposure from the intact body of a patient following a diagnostic administration of a radiopharmaceutical requires no precautions. Unless there is a medical emergency, surgery should be delayed for at least 24 h after such an administration to minimize contamination of the surgical facility with radioactive blood or urine and to avoid the exposures of the hand that would attend surgery on some organs with long retention times (technetium-99m sulfur colloid in the liver; iodine-131 in the thyroid).

Following radionuclide therapy, the greatest radiation exposure of a surgeon and his assistants occurs in the rare instance of surgery performed within 24 h of administration of iodine-131 for thyroid cancer treatment. The safest policy is to delay surgery for at least a few days, when more than 80 per cent of the activity has been excreted. Hand exposure over the thyroid gland should be avoided since the maximum permissible annual dose to the hands could be accumulated in a few hours. Excised thyroid tissue from the gland or metastatic disease should be handled with forceps and should be placed in a shielded container with lead walls at least 1 cm thick. Shields against body exposure are impractical, since lead of 3 mm thickness reduces the dose by only a factor of 2. The essence of such a procedure is speed. Following the operation a detailed room survey of radioactive contamination levels should be made by radiation safety personnel and decontamination procedures initiated.

Surgeons operating fluoroscopes frequently or those performing surgery on patients following recent therapeutic radionuclide administrations should wear film badges, preferably two, one on the collar for monitoring head and neck exposure, and one at waist level, which should be under the lead apron in the case of fluoroscopy. A finger ring monitor (if acceptable) under the surgical glove is recommended for the hand primarily used for manipulation in fluoroscopic procedures or in procedures close to radioactive organs following radionuclide therapy.

Further reading

Butler WE *et al.* A mobile computed tomographic scanner with intraoperative and intensive care unit applications. *Neurosurgery* 1998; **42**: 1304–11. [The paper discusses a unique mobile CT scanner (Philips Tomoscan M) for head scanning in the operating room or at the bedside.]

Cardis E *et al.* Effects of low doses and low dose rates of external ionizing radiation: cancer mortality among nuclear industry workers in three countries. *Radiation Research* 1995; **142**: 117–32. [In this largest-ever epidemiological study of 95000 radiation workers in Canada, the UK and the USA with mean dose 40 mSv, there was a small excess of leukemia and o excess of other cancers.]

Committee on the Biological Effects of Ionizing Radiations (BEIR V). *Health effects of exposure to low levels of ionizing radiation*. National Academy Press, Washington, DC, 1990. [The lifetime excess cancer mortality risk (including leukemia) from an acute whole body dose of 100 mSv is estimated as .8 per cent.]

Doll R, Wakeford R. Risk of childhood cancer from fetal irradiation. *British Journal of Radiology* 1997; **70**: 130–9. [A review of epidemiologic studies of childhood cancer in children who were irradiated *in utero*: risk estimate 6 per cent per Sv.]

International Commission on Radiological Protection. Statement from the 1985 Paris Meeting of the Commission. *Radiology* 1985; **156**: 816. [Radiation doses to the operator from fluoroscopy are 35 to 250 times higher from an overcouch X-ray tube compared to an undercouch X-ray tube.]

International Commission on Radiological Protection.*Recommendations of the Commission, 1990*. ICRP Publication 60, Pergamon Press, New York, 1991. [A review of radiation risks by organ and of maximum permissible doses for occupational exposure.]

Liu J, Edwards FM. Radiation exposure to medical personnel during iodine-125 seed implantation of the prostate. *Radiology* 1979; **132**: 748–9. [Measurements of personnel exposure during prostate implantation with 1-125 seeds.]

National Council on Radiation Protection and Measurements.*Precautions in the management of patients who have received therapeutic amounts of radionuclides*. Report No. 37, Bethesda, MD, 1970. [Precautions to minimize radiation exposure to persons inside and outside the hospital close to a patient containing therapeutic amounts of radioactivity.]

National Council on Radiation Protection and Measurements. *Limitation of exposure to ionizing radiation*. Report No. 116, Bethesda, MD, 1993. [Current radiation exposure limits for occupationally exposed persons and the general public.]

Palmer KE, Wright IH. Eye dose limits and the use of over-couch and under-couch X-ray image intensifier systems. *British Journal of Radiology* 1985; **58**: 1221–3. [Eye exposure of physicians from fluoroscopy with X-ray tubes over or under the couch.]

Rosenstein M, Webster EW. Effective dose to personnel wearing protective aprons during fluoroscopy and interventional radiology. *Health Physics* 1994; **67**: 88–9. [This paper derives the formula for calculating effective dose to personnel wearing two radiation monitors (under the apron and above the apron) during X-ray procedures.]

United Nations Scientific Committee on the Effects of Atomic Radiation. *Sources, effects and risks of ionizing radiation. 1988 Report to the General Assembly. Annex E, Genetic effects*. UN Publication E.88.IX.7, United Nations, New York, 1988. [Annex E of this report reviews the risks of genetic damage in the first generation after parental exposure to 10 mGy at low dose rates, concluding that the risk of an abnormal child is of the order 10 per million.]

United Nations Scientific Committee on the Effects of Atomic Radiation. *Sources and effects of ionizing radiation. 1994 Report to the General Assembly. Annex A: Epidemiological studies of radiation carcinogenesis*. UN Publication E.94.IX.11, United Nations, New York, 1994. [This contains an annex updating epidemiological studies of radiation carcinogenesis reviewed in the 1988 Report.[

Webster EW. Radiation protection in nuclear medicine. In: Taveras JM, Ferrucci JT, eds, *Radiology/diagnosis, imaging, intervention*, Vol. 1, Chapter 25, pp. 1–13. Lippincott, Philadelphia, 1987. [This chapter discusses protection of patients and personnel from internal and external exposure to radioactive sources, including exposures from specific procedures.]

Webster EW. Basic principles of X-ray imaging. In: Pohost GM, O'Rourke RA, eds, *Principles and practice of cardiovascular imaging*, pp. 331–65. Little, Brown and Co., Boston, 1991. [This chapter discusses the physical factors affecting resolution, contrast and personnel exposure in radiographic and fluoroscopic procedures.]

49.1 Care of the dying patient

Robert G. Twycross

Appropriate treatment
Symptom relief
Breathlessness
Nausea and vomiting
Bowel obstruction
Dehydration
Confusion
Hope
Spiritual care
As death approaches
Care of the family
Further reading

The majority of dying patients in surgical wards have cancer. The text is written with such patients specifically in mind, although the underlying principles have more general application. When caring for a dying patient, the main goals are relief of pain and other distressing symptoms, and psychological and spiritual support for patients so that they may come to terms with their own death as fully and constructively as possible. Parallel support for the patient's family is also important. For patients who are terminally ill rather than imminently dying, it is necessary to encourage them to continue to live as actively and creatively as possible, thereby facilitating autonomy and enhancing self-esteem. ‘Cocooning' a patient into premature dependency is in nobody's best interests. Care of the dying patient therefore calls for a comprehensive approach, encompassing physical, psychological, social, and spiritual aspects (Box 1). This requires imaginative multiprofessional teamwork.

Box 1. No caption available.

Physical	Symptom relief
Psychological	
Safety	Feeling of security
Understanding	Explanation about symptoms and the disease; opportunity to discuss the process of dying
Self-esteem	Involvement in decision making, particularly as physical dependency on others increases; the opportunity to give as well as to receive
Social	
Love	Expressions of affection; human contact, e.g. touch
Belonging	The need to feel needed and connected, and not a burden
Disengagement	Taking leave of those people or things to whom one is attached; the opportunity to tie up loose ends in business and family matters; to hand on responsibility to others
Spiritual	
Acceptance by others	Regardless of mood, sociability, and appearance
Reconciliation	The opportunity to repair damaged relationships; to seek forgiveness
Self-worth	The need to feel that one's life has been worthwhile
Purpose	The need to feel one's life has had direction and purpose; that life is not futile

Appropriate treatment

Medical care is a continuum, ranging from complete cure at one end to symptom relief and support at the other. Many types of treatment span the entire spectrum of care, notably radiotherapy and, to a lesser extent, chemotherapy and surgery. It is important to keep the therapeutic aim clearly in mind when employing any form of treatment. The key points to bear in mind are the patient's biological prospects, the therapeutic aim of each treatment, the adverse effects of treatment, the benefits of treatment for the patient, and the need not to prescribe a lingering death.

When a person is close to death the primary aim of treatment is comfort. What is appropriate for an acutely ill patient may well be inappropriate in the dying. Nasogastric tubes, intravenous infusions, antibiotics, cardiac resuscitation, and artificial respiration are primarily supportive measures to assist a patient with acute or acute-on-chronic illnesses through the initial critical period towards recovery of health. Such measures are inappropriate in patients who are close to death and for whom there is no expectation of a return to health if they simply prolong the process and distress of dying.

Although unexpected improvement is sometimes seen, there are many occasions when it is appropriate to ‘give death a chance'. All patients must die eventually; ultimately, nature will take its course. The art of medicine is to decide when life sustenance is essentially futile and, therefore, when to allow death to occur without further medical impediment. Although antibiotics are appropriate for the patient with advanced cancer who develops a chest infection while still relatively active and independent, pneumonia should still be ‘the old man's friend'. In other words, if the burden of living has become considerable, it may be more appropriate not to use antibiotics. When it is difficult to make a decision, the ‘2-day rule' should be invoked: if after 2 or 3 days of straightforward symptom relief the patient is clearly holding his own, prescribe an antibiotic but do not if the patient is clearly much worse.

As patients become bedbound, it is sensible to phase out the less important medications. Multivitamins, iron, and potassium supplements are all medium or long-term supportive treatments; they are irrelevant when the patient is close to death. Antihypertensives can often be phased out as weight loss and/or other medication alter the patient's need; for example, morphine reduces the cardiac preload. Continuation of pre-existing diuretic therapy may cause postural hypotension unless the dose is reduced or stopped in the face of a poor fluid intake or repeated vomiting. When death seems close, giving the patient ‘permission' not to eat and discouraging the family from forced-feeding is enerally appropriate.

Symptom relief

A wide range of symptoms may be experienced by patients with advanced cancer, but no symptom occurs invariably. Pain is experienced by only three-quarters of patients (Chapter 49.2). A scientific and SYstematic approach is necessary to achieve the best outcome. The four essential steps are:

- evaluation (‘diagnosis before treatment')
- explanation
- management
- monitoring.

As in other areas of clinical practice, attention to detail is always necessary. Not all symptoms are caused by the cancer. They may also be caused by treatment (past or present) or a concurrent second disorder. Further, some symptoms are caused by multiple factors. It is necessary to identify what these are, correct those that are readily reversible, and initiate appropriate symptom relief measures (analgesics, antiemetics, etc.). In this way, although the underlying disease process remains unchanged, it is generally possible to obtain significant, if not complete, relief. In relation to appetite, weakness and fatigue, however, often the main part of management is gently helping the patient (and family) accept the physical limitations of a terminal illness.

If a drug is indicated for a persistent SYmptom, it should be prescribed regularly on a prophylactic basis. Although this concept is widely applied to analgesics, its relevance to antiemetics and laxatives is not always appreciated. The use of drugs as required instead of regularly by the clockis still the cause of much unrelieved distress. As the needs of patients vary, it is often impossible to predict the optimum dose of many drugs, notably opioids, laxatives, and psychotropic drugs. Adverse effects also jeopardize patient compliance. Adjustments will therefore be necessary, particularly during the first few weeks. This should be anticipated and arrangements made for continuing monitoring.

Although many symptoms respond to a combination of non-drug and drug measures, it is sometimes necessary to compromise in order to avoid unacceptable adverse effects. For example, antimuscarinic effects such as dry mouth or visual disturbance may be limiting factors. In patients with inoperable intestinal obstruction, it is often

better to aim to reduce the incidence of vomiting to once or twice a day rather than to seek absolute control.

Breathlessness

Breathlessness becomes more common as death approaches. The key to successful management at the end of life, however, is at an earlier stage when breathlessness on exertion first becomes a symptom. It is often then that the ‘battle for the mind' is won or lost, with the patient either becoming more and more fearful of a terrifying death or learning what can be done to ease the panic that is a concomitant of acute breathlessness. In most patients breathlessness is caused by several factors, e.g. progressive intrathoracic cancer, chronic obstructive pulmonary disease, anaemia of chronic disease and weakness. Specific treatment should be given for specific causes, e.g. bronchodilators for reversible airways obstruction, diuretics and an angiotensin-converting enzyme inhibitor for cardiac failure, aspiration and pleuradesis for pleural effusion, and radiation for mediastinal obstruction. Acknowledgement of and discussion about the terror associated with acute episodes of breathlessness, e.g. when climbing stairs, is also important. Referral to a physiotherapist for breathing advice and relaxation techniques should be routine. However, when the patient is close to death, breathlessness may be present at rest as well as on exertion, or become apparent on minimal activity. Often the patient is bedbound more because of breathlessness than weakness. In this situation, an opioid (to slow respiratory drive), an anxiolytic and oxygen therapy may all be necessary (Box 2).

Box 2. No caption available.

Bronchodilators

The use of bronchodilators should be reviewed and tested clinically for efficacy; it may be necessary to use a spacer to ensure adequate inhalation from an inhaler, or to convert to a nebulizer.

Opioids

Morphine reduces respiratory drive and can be used to ease the sensation of dyspnoea:

if on morphine for pain, increase the dose by 50 per cent.

if not on oral morphine, 5–6 mg q4h is a good starting dose.

Nebulized morphine is not recommended; it is no better than saline.

Anxiolytics

Diazepam by mouth is a good choice:

5–10 mg stat and bedtime

2–5 mg in the very elderly

reduce dose after several days if drowsy.

Use midazolam subcutaneously for patients who find tablets difficult:

2.5–5 mg stat and p.r.n.

10–20 mg/24 h by subcutaneous infusion.

Oxygen

Oxygen 4 l/min, preferably via nasal prongs, should be tried to see if there is benefit either continuously or before and during activity (e.g. moving from bed to commode or chair). The benefit of oxygen is not dependent on correction of hypoxaemia; a trial of therapy is the only way to determine benefit, not improvement in blood gases.

Nausea and vomiting

Nausea and vomiting are common symptoms in terminal cancer, affecting up to 70 per cent of patients. It is important to choose the most appropriate antiemetic (Table 1). Most patients will obtain relief with one of four first-line antiemetics: metoclopramide, haloperidol, hyoscine butylbromide, and cyclizine (Box 3). In a minority, a second-line antiemetic may need to be substituted for or added to the first-line drug (Box 4).

Position of action	Class	Example
Central nervous system		
Vomiting centre	Antimuscarinic	Hyoscine hydrobromide
	Antihistamine, antimuscarinic*	Cyclizine, dimenhydrinate, promethazine
	$5HT_3$ receptor antagonist	Levomepromazine
Area postrema	D_2 receptor antagonist	Haloperidol, prochlorperazine, levomepromazine, metoclopramide, domperidone
	$5HT_3$ receptor antagonist	Granisetron, ondansetron, tropisetron
Cerebral cortex	Benzodiazepine	Lorazepam
	Cannabinoid	Nabilone
	Corticosteroid	Dexamethasone
Gastrointestinal tract		
Peristalsis	$5HT_4$ receptor antagonist	Metoclopramide, cisapride
	D_2 receptor antagonist	Metoclopramide, domperidone
Antisecretory	Antimuscarinic	Hyoscine butylbromide, glycopyrrolate
	Somatostatin analogue	Octreotide, vapreotide
Vagal $5HT_3$ receptors/afferents	$5HT_3$ receptor antagonist	Granisetron, ondansetron, tropisetron
Anti-inflammatory	Corticosteroid	Dexamethasone

*The antihistaminic, antimuscarinic and phenothiazine/butyrophenone D_2 receptor antagonist actions can be grouped together for convenience.

Table 1 Drugs used to control nausea and vomiting

Box 3. No caption available.

Prokinetic antiemetic

For gastritis, gastric stasis, functional bowel obstruction:

metoclopramide 10 mg q.d.s. orally/subcutaneously or 40–100 mg/24 h by continous subcutaneous infusion.

Antiemetic acting principally in area postrema (chemoreceptor trigger zone)

For most chemical causes of vomiting, e.g. morphine, hypercalcaemia, renal failure:

haloperidol 1.5–5 mg orally at bedtime or b.d., or 5 mg/24 h subcutaneously at bedtime or by continuous subcutaneous infusion. Metoclopramide also has a central antidopaminergic action.

Antipasmodic and antisecretory antiemetic

If bowel colic and/or need to reduce gut secretions:

hyoscine butylbromide 20 mg subcutaneous stat, 80–160 mg/24 h subcutaneous infusion, and 20 mg subcutaneous hourly p.r.n.

Antiemetic acting principally in the vomiting centre

For mechanical bowel obstruction, raised intracranial pressure, motion sickness:

cyclizine 50 mg b.d.–t.d.s. orally or 150 mg/24 h by continuous subcutaneous infusion. (In countries where cyclizine is not available, alternatives include dimenhydrinate and diphenhydramine.)

Do not prescribe a prokinetic drug and an antimuscarinic (anticholinergic) drug concurrently; the final common pathway for prokinetic drugs is cholinergic; antimuscarinic drugs (including cyclizine) block their prokinetic action.

Box 4. No caption available.

Antispasmodic and antisecretory antiemetic

If bowel colic and/or need to reduce gut secretions:

hyoscine butylbromide 20 mg subcutaneous stat, 80–160 mg/24 h subcutaneous infusion, and 20 mg subcutaneous hourly p.r.n.

Levomepromazine

Levomepromazine is a phenothiazine with potent antagonist properties at $5HT_2$ receptors as well as (like other phenothiazines) at α_1-adrenergic, D_2, H_1 and muscarinic receptors. It is the most broad-spectrum antiemetic available, though because of its receptor site affinities and potency it can cause unacceptable drowsiness and severe postural hypotension particularly when used in doses of ≥25mg /24 h subcutaneously. It can be substituted for haloperidol and cyclizine when these narrow-spectrum drugs are inadequate alone or in combination. Typical doses are levomepromazine 6.25–12.5 mg subcutaneously or 12.5–25 mg orally stat and at night, and p.r.n. Doses are adjusted upwards as necessary.

$5HT_3$ receptor antagonists

Introduced initially to control chemotherapeutic vomiting, $5HT_3$-receptor antagonists are of value in other situations where excessive 5HT (serotonin) is released from enterochromaffin cells (abdominal radiation, bowel distension), or from platelets (renal failure), from the raphe nucleus in the brainstem (brainstem radiation, head injury) and postoperatively. Doses parallel those recommended for chemotherapeutic vomiting. If not effective within 3 days, the $5HT_3$-receptor antagonist should be stopped; continued use without the pay-off of therapeutic success is prohibitively expensive.

Corticosteroids

Dexamethasone is an important antiemetic but, apart from prophylaxis against chemotherapeutic vomiting, it is probably underused. Generally, in palliative care, it is added to an existing antiemetic regimen rather than substituted. If successful, its continued use may allow other antiemetics to be discontinued or dose reduced. The dose is somewhat arbitrary, e.g. dexamethasone 4–20 mg once daily or subcutaneously, reducing for maintenance to 2–6 mg once daily when control has been achieved.

Bowel obstruction

Malignant bowel obstruction in terminal cancer generally differs from acute obstruction with its classical quartet of symptoms and signs (see p.000). Because of widespread multiple malignant adhesions, distension is often only mild–moderate and the patient may complain of diarrhoea rather than constipation. Partial (subacute) obstruction is the norm and the patient continues to pass flatus. Bowel sounds vary from hyperactive (with borborygmi) to none; high-pitched tinkling bowel sounds are rare. In many patients (including most of those with cancer of the head of pancreas) the obstruction is more functional than mechanical, and is associated with a disruption of peristalsis. In some cases this is caused by local denervation as a result of neural compression/infiltration by tumour and/or linitis plastica of the intestines. From a management point of view the defining symptom is colic. If this is present, it suggests a significant mechanical component for which hyoscine butylbromide is indicated (e.g. 20 mg subcutaneously at once and 60 to 160 mg/ 24 h by subcutaneous infusion). Glycopyrrolate 600–1200 mcg/24 h can be used instead in countries where hyoscine butylbromide is not available. Some centres prefer the more expensive somatostatin analogue, octreotide 250–1000 ccg/24 h; this is a potent antisecretory drug and reduces colic by decreasing the volume of the intestinal luminal contents.

If there is no colic, a trial of the prokinetic antiemetic metoclopramide is indicated (e.g. 10 mg subcutaneously at once and 40 to 100 mg/24 h by subcutaneous infusion) and is often surprisingly successful. If the antiemetic is appropriately chosen and the dose optimized, it may well be possible to reduce vomiting to once or twice a day (or less if functional) and to relieve debilitating nausea. In dying patients, the automatic use of intravenous fluids and nasogastric tube drainage (‘drip and suck') is inappropriate and almost always unnecessary. Each patient, of course, must be considered individually and it may be necessary to work through a series of therapeutic trials.

With mechanical obstruction of the pylorus, duodenum, or jejunum, a venting gastrostomy may be necessary but the use of an antisecretory drug should be tried first in a patient who, judging by her general condition, seems to be within days or just a few weeks of death (e.g. hyoscine butylbromide as above or octreotide 250 to 500 µg/24 h by subcutaneous infusion). In such a patient, however, a nasogastric tube may also be necessary to relieve intractable nausea and vomiting. The stomach can be emptied before the patient drinks or eats, and possibly left on free drainage overnight. The patient will generally be greatly relieved and often remains remarkably well for several days before deteriorating rapidly as a result of the accumulated loss of fluid and electrolytes.

Dehydration

As the dying patient becomes weaker, intake becomes less. This may relate to dysphagia, nausea, anorexia, lack of energy and interest, and a reduced level of consciousness. As a general rule intravenous fluids are not administered to patients dying in palliative care units. On the other hand, patients dying in a general hospital normally continue to receive intravenous fluids. This reflects a belief that dehydration is distressing, and therefore unacceptable. Although this is true of rapid-onset (acute) dehydration, it appears not to be true of slow-onset (chronic) dehydration at the end of life. Indeed, in patients close to death, a certain amount of dehydration is beneficial because urine output decreases, which results in fewer bed-wetting episodes and less need for urinal, commode, or catheterization. In contrast, parenteral hydration with ³2 litres/day counteracts much of the beneficial antisecretory effect of hyoscine butylbromide and leads to more copious vomiting. This then necessitates the use of a nasogastric tube as seemingly the only means of obviating otherwise intractable nausea and vomiting. Slow-onset dehydration should therefore be regarded as a natural process in the dying that helps to relieve various symptoms in the last days of life.

Dry mouth in dying patients, often considered the most discomforting consequence of dehydration, is not closely correlated with the degree of dehydration; mouth breathing and antimuscarinic drugs are more important factors. Parenteral rehydration does not therefore correct it. Dry mouth can be eased, however, by conscientious mouth care. It is often possible to enlist the family's help in this; a bottle with a spray top allows small volumes (1 to 2 ml) to be introduced into the mouth every 20 to 30 min.

Neuromuscular irritability is a relatively common concomitant of dehydration and electrolyte disturbances in the dying. This leads to multifocal myoclonus (muscle twitching) and general restlessness. Midazolam should be given parenterally to control these, preferably by continuous subcutaneous infusion using a portable syringe driver. Alternatively, diazepam orally or rectally can be used.

Confusion

An acute confusional state (delirium) is associated with mental clouding. This leads to a disturbance of comprehension and bewilderment, and is often associated with drowsiness. Manifestations include poor concentration, impairment of short-term memory, disorientation, misinterpretations, paranoid ideas, hallucinations, rambling incoherent speech, restlessness, and noisy aggressive behaviour. Acute confusion varies in degree and intensity. In the dying, it is generally caused by multiple factors (Box 5). In some patients, symptoms are never more than mild and remain intermittent.

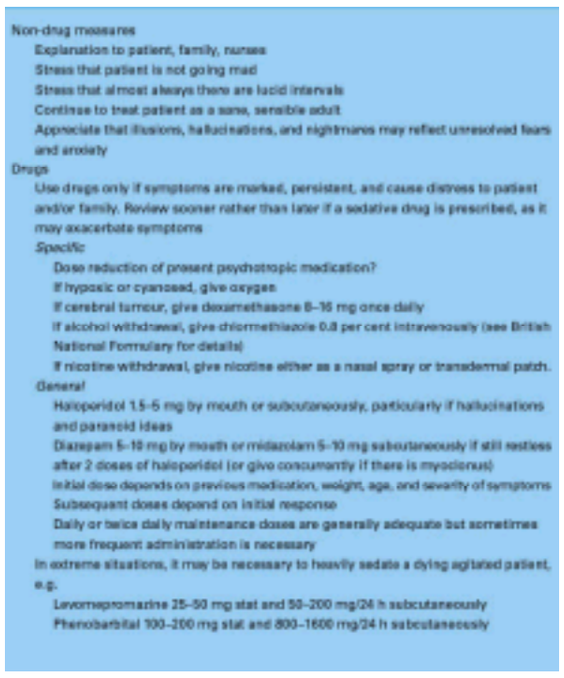
Box 5. No caption available.

Cancer	Drugs
Systemic effect	Sedative
Cerebral involvement	Stimulant
Infection	Antiparkinsonian
Dehydration	Drug withdrawal
Change of environment	Psychotropics
Unfamiliary excessive stimuli	Alcohol
Too hot	Nicotine
Too cold	Biochemical derangement
Wet bed	Hypercalcaemia
Crumbs in bed	Uraemia
Creases in sheets	Hyponatraemia
Pain	Hypoglycaemia
Constipation	Hyperglycaemia
Retention of urine	Cerebral anoxia
Pruritus	Anaemia
Anxiety	Cardiac failure
Depression	Hypoxia
Fatigue	Organ failure
	Hepatic
	Renal

Unfortunately, once a patient is deemed confused, doctors, nurses, friends, and family tend to distance themselves because of embarrassment or even fear.

Explanation and reassurance are therefore vital, and often enable the family and carers to continue to respond supportively to someone who happens to be very ill and who has what may be only a relatively minor disturbance of thought, e.g. disorientation for time and place. Hallucinations, illusions, nightmares, and vivid daydreams often reflect anxiety about death. Anxiolytic medication, together with continuing psychological support, is helpful when the content suggests this. Blunderbuss therapy with chlorpromazine is still a common medical response to this situation. Although this may help to suppress hallucinations, it may increase the tendency to misperceive as a result of further clouding of the sensorium. Haloperidol, which is much less sedative, is the antipsychotic of choice in this situation (Box 6).

Box 6. No caption available.


Non-drug measures
- Explanation to patient, family, nurses
  - Stress that patient is not going mad
  - Stress that almost always there are lucid intervals
  - Continue to treat patient as a sane, sensible adult
  - Appreciate that illusions, hallucinations, and nightmares may reflect unresolved fears and anxiety

Drugs
- Use drugs only if symptoms are marked, persistent, and cause distress to patient and/or family. Review sooner rather than later if a sedative drug is prescribed, as it may exacerbate symptoms
- *Specific*
  - Dose reduction of present psychotropic medication?
  - If hypoxic or cyanosed, give oxygen
  - If cerebral tumour, give dexamethasone 8–16 mg once daily
  - If alcohol withdrawal, give chlormethiazole 0.8 per cent intravenously (see British National Formulary for details)
  - If nicotine withdrawal, give nicotine either as a nasal spray or transdermal patch.
- *General*
  - Haloperidol 1.5–5 mg by mouth or subcutaneously, particularly if hallucinations and paranoid ideas
  - Diazepam 5–10 mg by mouth or midazolam 5–10 mg subcutaneously if still restless after 2 doses of haloperidol (or give concurrently if there is myoclonus)
  - Initial dose depends on previous medication, weight, age, and severity of symptoms
  - Subsequent doses depend on initial response
  - Daily or twice daily maintenance doses are generally adequate but sometimes more frequent administration is necessary

In extreme situations, it may be necessary to heavily sedate a dying agitated patient, e.g.
- Levomepromazine 25–50 mg stat and 50–200 mg/24 h subcutaneously
- Phenobarbital 100–200 mg stat and 800–1600 mg/24 h subcutaneously

Hope

Hope is an expectation greater than zero of achieving a goal. Hope tends to diminish when:

- the patient is mentally isolated by a ‘conspiracy of silence';
- it is implied ‘there is nothing more that can be done';
- pain and other symptoms remain unrelieved or ignored;
- the patient feels alone or unsupported.

Not destroying hope is frequently used as a reason for not informing a patient of the seriousness of his situation. False optimism, however, is a potent destroyer of hope. On the other hand, an unwise catharsis by the doctor of all that is negative, either to the patient or to the family, may make the doctor feel easier but it can irreversibly destroy hope and result in intractable anxiety and despair. It is necessary to apply two parallel principles: ‘never lie to a patient' and ‘avoid total candour.' Most people are best served when realism is tinged with optimism. Gentle, sympathetic, and gradual communication of the truth within the context of continued support and encouragement almost never destroys hope. In fact, good terminal care restores hope by encouraging realistic goals. Telling the patient what you plan to do about his pain and other symptoms helps to counter the living nightmare of never-ending unendurable pain or other physical distress.

Spiritual care

When dying many people take stock of their lives for the first time:

- ‘I have lived a good life';
- ‘I never did anyone any harm';
- ‘Why has it happened to me?'

Although only a minority of patients discuss these issues with their doctor, the majority do so with a nurse or social worker, or with their relatives and close friends. Patients are very perceptive, and they are unlikely to embarrass a doctor if they sense that communication at this level will cause him discomfort. When possible and appropriate, it is right for the doctor to alert the chaplain, priest, or rabbi to the fact that the patient is seriously ill and may appreciate a visit. Although regard for the patient as an individual does not allow the imposition of one's own beliefs, many patients are comforted by the discovery that their doctor has a religious faith.

As death approaches

As patients becomes increasingly weak they are faced with the fact that death is inevitable and imminent. This is a time when support and companionship are of paramount importance. Although the doctor may feel powerless in the face of rapidly approaching death, the patient is usually more realistic. Patients know a miracle cannot be performed, they realize that their time has come, and, provided they are relatively comfortable, continued medical attentiveness is all they expect. The situation need never become negative and hopeless provided the doctor:

- continues to visit;
- quietly indicates that ‘at this stage the important thing is to keep you as comfortable as possible';
- simplifies medication as much as possible;
- arranges for medication to be given by injection (or suppository) when the patient is too weak to swallow;
- continues to inform the family of the changing situation;
- takes steps to control an agitated confusional state even if it results in greater somnolence;
- listens to the nurses.

Care of the family

The care of the family is an integral part of the care of the dying. Family–doctor communication generally needs to be initiated and maintained by the doctor. It is easy to neglect the relatives, who are often reluctant to bother a doctor who they see as busy. There is much to be said, at the time of diagnosis and later, for joint interviews with the patient and spouse. The doctor should also make an opportunity to see both the patient and the close relatives on their own. Further separate or joint interviews can be arranged as necessary.

It is not generally necessary or wise to tell the family the whole truth (as you see it) at one time. If the family and patient are too far ‘out of step' in relation to knowledge about the diagnosis and prognosis, it can create a barrier between the two. They commonly ask for the patient not to be told. This should be seen as the initial shock reaction and not as an excuse for saying nothing to the patient. If the family and patient are to be mutually supportive, the relatives must be helped to move forward from this initial reaction to a position of greater openness and trust. It is important to remember that the family cannot forbid the doctor to discuss diagnosis and prognosis with a patient.

Admission to hospital is often seen as a defeat by the family. It may help to say that you are impressed that they have managed to cope for so long and that now, with the need for day and night care, it is impossible for one (or two) people to continue to cope alone without a break. Frequent visits should be encouraged. Separation anxiety may be reduced by encouraging close relatives to help in the care of the patient, for example by adjusting the pillows, refilling a water jug, helping with a blanket bath, and assisting at mealtimes. Some relatives need to be taught how to visit, to behave as they would at home. For example, to sit and read a book or newspaper, to knit, or to watch television. It should be emphasized that they do not have to keep up a tiring patter of conversation about trivia. If necessary, overnight accommodation should be arranged to enable a close relative to sleep nearby or with the patient in a single room.

Further reading

Ahmedzai S, Davis C. Nebulised drugs in palliative care. *Thorax* 1997; **52** (Suppl. 2): S75–7. [Describes the use of nebulizers in palliative care with cautionary advice about nebulized morphine – by this route it is no better than saline.]

Booth S *et al.* Does oxygen help dyspnea in patients with cancer? *American Journal of Respiratory Critical Care Medicine* 1996; **153**: 1515–18. [Demonstrates that piped air is as good as piped oxygen when administered by nasal prongs to cancer patients breathless at rest.]

Corner J, Plant H, A'Hern R, Bailey C. Non-pharmacological interventions for breathlessness in lung cancer. *Palliative Medicine* 1996; **10**: 299–305. [An important account of non-drug measures to ease breathlessnes in patients with advanced cancer.]

Davis CL. Breathlessness, cough and other respiratory symptoms. *British Medical Journal* 1997; **315**: 931–4. [A short account of the management of respiratory symptoms in advanced cancer.]

Doyle D, Hanks GWC, MacDonald N, eds. *Oxford textbook of palliative medicine*, 2nd edn. Oxford University Press, Oxford, 1997. [A multi-author book covering all aspects of palliative care with individual chapters on most SYmptoms.]

Dunphy K, Finlay I, Rathbone G, Gilbert J, Hicks F. Rehydration in palliative and terminal care: if not—why not? *Palliative Medicine* 1995; **9**: 221–8. [A useful discussion of the ethical and practical aspects of artificial rehydration in dying patients.]

Fainsinger RL, Spachynski K, Hanson J, Bruera E. SYmptom management in terminally ill patients with malignant bowel obstruction. *Journal of Pain and Symptom Management* 1994; **9**: 12–18. [An account of the medical management of bowel obstruction in endstage cancer.]

de Hennezel M. *Intimate death*. Little Brown and Co, Boston, 1997. [A moving account of the insights gained by a clinical psychologist involved in palliative care.]

Kearney M. *Mortally wounded*. Marino, Dublin, 1996. [In this book subtitled ‘Stories of soul pain, death and healing', a physician examines the psychosociospiritual aspects of dying and, with examples, discusses the therapeutic value of image-work.]

Stedeford A. *Facing death: patients, families and professionals*, 2nd edn. Sobell Publications, Oxford, 1994. [An excellent general account of the psychological aspects of dying by a psychiatrist who specialized in palliative care.]

Twycross RG. *Introducing palliative care*, 3rd edn. Radcliffe Medical Press, Oxford, 1999. [An excellent primer and course manual on the essentials of palliative care.]

Twycross RG. *Symptom management in advanced cancer*, 2nd edn. Radcliffe Medical Press, Oxford, 1997. [A wide-ranging account of the management of pain and other symptoms in advanced cancer.]

Twycross R, Back I. Nausea and vomiting in advanced cancer. *European Journal of Palliative Care* 1998; **5**: 39–45. [A consensual paper arising from the 1997 Congress of the European Association for Palliative Care.]

Twycross RG, Barkby GD, Hallwood PM. The use of low dose levomepromazine (methotrimeprazine) in the management of nausea and vomiting. *Progress in Palliative Care* 1997; **5**: 49–53. [An important synopsis of what is known about levomepromazine (methotrimeprazine) and its utility in low doses as a broad-spectrum antiemetic.]

Twycross RG, Wilcock A, Thorp S. *Palliative care formulary*. Radcliffe Medical Press, Oxford 1998. [Brings together pharmacological and pharmaceutical information about all the common symptom relief drugs used in palliative care in the United Kingdom.]

49.2 Cancer pain relief

Robert G. Twycross

Introduction
Case history
Evaluation
Response to analgesics
Pain management
Non-opioids
Weak opioids
Strong opioids
Adjuvant analgesics
Conclusions
Further reading

Introduction

Pain is a major SYmptom in 75 per cent of patients with advanced cancer. A sense of hopelessness and the fear of impending death often adds to the total suffering of the patient. It is necessary, therefore, to think in terms of ‘total pain'. This includes psychological, social, and spiritual factors as well as the underlying physical component.

Case history

A 55-year-old man with recently diagnosed cancer of the oesophagus was still in pain despite receiving modified-release morphine tablets, 6000 mg (100 mg × 60) twice a day. Following admission to a hospice, he became pain-free on morphine 30 mg twice a day and diazepam 10 mg at night. He was able to return home, and converted a spare bedroom into a workshop where he spent many hours creatively and productively. The key to success was listening, explaining, and setting positive rehabilitation goals.

Evaluation

Pain in cancer may be caused directly by the cancer, related to the associated debility (e.g. constipation, bedsores), related to cancer treatment (e.g. chronic postoperative scar pain, chemotherapy mucositis), or caused by a concurrent disorder (e.g. spondylosis, arthritis). Multiple pains are common. About one-third of patients with pain in advanced cancer have two pains and another third have three or more pains. In 10 to 15 per cent of such patients, none of their pains is caused directly by the tumour.

Doctors caring for cancer patients need to acquire the skill of determining the cause of pain on the basis of diagnostic probabilities and pattern recognition. Invasive investigations are increasingly contraindicated as patients move closer to death. A comprehensive list of possible causes of pain is unnecessary because it would be too long to be helpful. What is required is an informed imagination. For example, awareness of functional muscle pain syndromes and of typical patterns of metastatic spread prevent many erroneous conclusions (Fig. 1 and Fig. 2). Pain that is wrongly assumed to be cancerous often becomes invested with all the negative implications of cancer pain, and makes the pain much worse. When there is doubt, investigations should be undertaken urgently. These range from plain radiographs, through ultrasound and contrast radiography, to computed tomography and magnetic resonance imaging. Initial treatment should not be withheld, however, until these have been completed. Analgesics should be started.



Fig. 1. Pain self-portrait by 65-year-old male patient with pancreatic cancer. Sites of pain indicate that it is muscular/myofascial in origin. Explanation, massage, and diazepam at night prescribed.

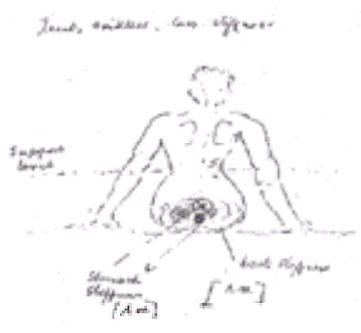


Fig. 2. Second pain-portrait by patient with pancreatic cancer. Myofascial pains have now resolved. New pain probably also non-malignant, related to lying in bed for long periods. Again analgesics were not prescribed. Explanation, sheepskin, and progressive mobilization resulted in relief.

Response to analgesics

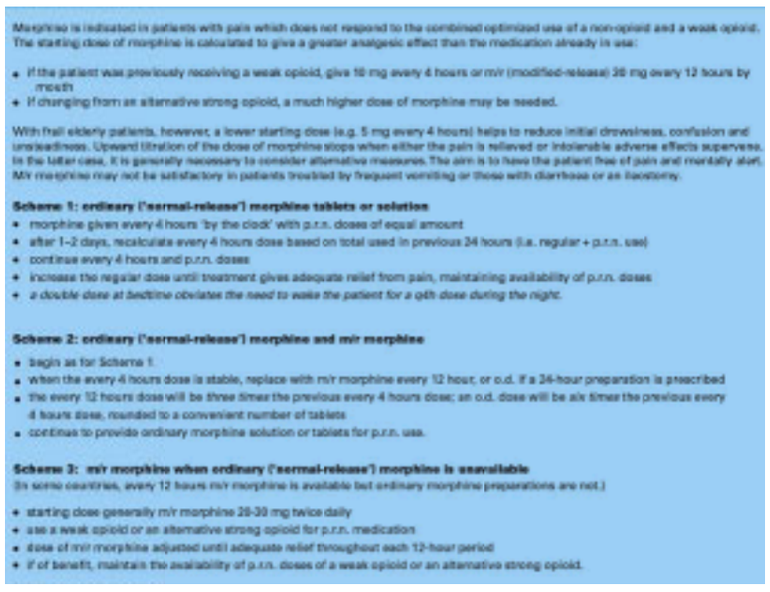
A detailed history of pain medication is part of the evaluation. It includes not only a list of drugs but also doses used, frequency and routes of administration, whether taken regularly or as required, the patient's view on efficacy, adverse effects, duration of use, and reasons for discontinuation. Not all pains respond equally to analgesics, and information about the benefits of specific medication (or lack of it) may help to pin-point the underlying mechanism.

From a therapeutic point of view, pain in cancer patients may be divided into four categories: opioid-responsive, opioid poorly-responsive, opioid non-responsive, and opioid irrelevant. In other words, opioids are of limited value with certain pains, and occasionally are contraindicated (Table 1). Many opioid poorly-responsive pains, however, are more responsive to opioid analgesics given by a spinal route. Opioid-irrelevant pain is pain which reflects psychospiritual distress more than physical injury. Opioid non-responsiveness may be more apparent than real (Box 1). Painful muscle spasm (cramp) is relatively common in cancer and is the prime example of a true opioid non-responsive pain. It is often truncal rather than in the lower limbs, and is commonly associated with underlying bone pain. Anxiety is a potent exacerbating factor.



Table 1 Types of pain and implications for treatment

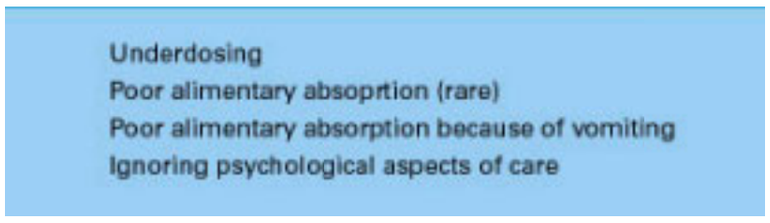
Box 1. No caption available.



Pain management

Pain relief generally requires a multimodal approach (Box 2). The perception of pain requires both consciousness and attention. Pain is worse when it occupies a person's whole attention. Diversional activity (distraction) does more than ‘pass the time'; it diminishes pain. Equally, professional time spent exploring a patient's worries and fears is time well spent, and relates directly to pain management. Indeed, if the patient is very anxious and/or depressed, it may take 2 to 4 weeks to achieve satisfactory relief. Distraction, however, is a concomitant of drug therapy; it is not a substitute.

Box 2. No caption available.



If anticancer treatment is recommended, analgesics should also be given until the treatment ameliorates the pain—which may take several weeks. Radiation should be considered when pain is caused by a bone metastasis or nerve compression. Single dose treatment with 6 to 10 Gy is often possible for bone pain, particularly in the limbs. Radiation gives partial or complete relief in more than 80 per cent of patients with bone pain. Recalcification occurs in most of these. Radiation is generally inappropriate in patients with a very short life expectancy, e.g. less than 2 weeks. Internal fixation or the insertion of a prosthesis should be considered if a pathological fracture of a long bone occurs or threatens. Fixation obviates the need for prolonged bed rest, and pain is much reduced. The decision to treat surgically depends on the patient's general condition.

Since the introduction of spinal analgesia, neurolytic and neurosurgical procedures have become almost obsolete. At many centres, a coeliac axis plexus block with alcohol for epigastric visceral pain is the only neurolytic block still used. Some patients continue to experience pain on movement despite analgesics, other drugs, radiotherapy, and other non-drug measures. The situation, however, is generally improved by modifications to the patient's way of life and environment. This is where a physiotherapist and an occupational therapist are invaluable. Hence, in practice, it is best to aim at progressive pain relief:

- relief at night
- relief at rest during the day
- relief on movement.

Relief should be evaluated in relation to each pain. Re-evaluation is a continuing necessity; old pains may get worse and new ones develop.

For pain caused by the cancer itself, drugs generally give satisfactory relief provided the right drugs are administered in the right dose at the right time intervals. Analgesics comprise three classes, namely: non-opioid (paracetamol and non-steroidal anti-inflammatory drugs); opioid (weak and strong); and adjuvant (e.g. corticosteroids, antidepressants, anticonvulsants, muscle relaxants). The principles governing their use can be summarized as *‘by the mouth'*, *‘by the clock', ‘by the ladder'*(Fig. 3), *‘individualized treatment', ‘use adjuvant drugs'*, and *‘attention to detail'*. Persistent pain requires preventive therapy. This means that analgesics should be given regularly and prophylactically; medication as required is irrational and inhumane (Fig. 4). The right dose of an analgesic is the dose which relieves the pain. Ceiling doses exist for all analgesics, however, either in absolute terms (additional dose increases yield no more pain relief) or relative terms (an increased dose yields more adverse effects than pain relief). Even so, it must be emphasized that, in practice, doses of oral morphine typically range from as little as 5 mg every 4 h up to 200 mg every 4 h, and occasionally much more.

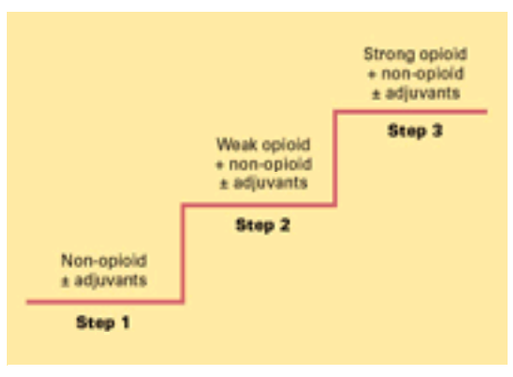


Fig. 3. The World Health Organization analgesic ladder for cancer pain management.

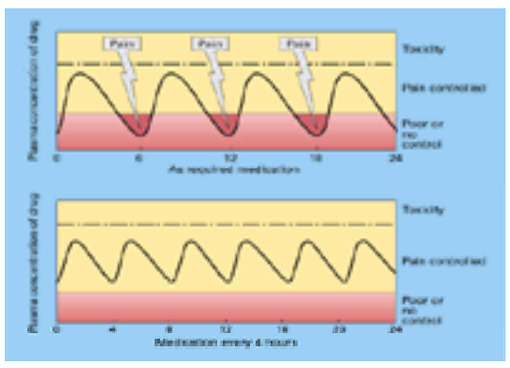


Fig. 4. A comparison of dosing ‘as required and regular morphine every 4 h.

The concept behind the analgesic ladder is ‘broad-spectrum analgesia', i.e. drugs from each of the three classes of analgesic are used appropriately either singly or in combination so as to maximize their analgesic impact (Fig. 5). Generally, however, the benefits of the analgesic ladder should be exploited before introducing or substituting an adjuvant analgesic. Adjuvant analgesics are necessary for more than 50 per cent of patients with neuropathic pain. Psychotropic medication is necessary for highly anxious and clinically depressed patients. A laxative is almost always necessary with an opioid and about 50 per cent of patients need an antiemetic.

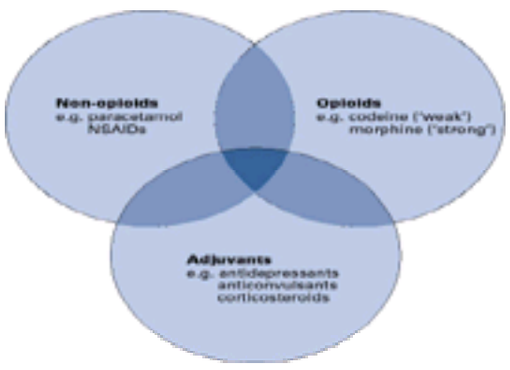


Fig. 5. Broad-spectrum analgesia: analgesics are used appropriately either singly or in combination.

All cancer patients receiving analgesics need close supervision to achieve maximum comfort with minimal adverse effects. Treatment review is sometimes necessary within hours; it is certainly required after 1 to 2 days and at the end of the first week. Subsequently, new pains may develop and old pains re-emerge. A fresh complaint of pain demands re-evaluation, not just a message to increase pain medication, although this may be an important short-term measure.

Non-opioids

Non-opioids comprise the non-steroidal anti-inflammatory drugs (NSAIDs) and paracetamol. NSAIDs are of particular benefit for pains associated with inflammation, e.g. soft tissue infiltration and bone metastases. Because inflammation leads to central sensitization and increased pain, NSAIDs are sometimes important in relieving cancer-related neuropathic pain. Ibuprofen, diclofenac, and naproxen are the most widely used NSAIDs in palliative care. Flurbiprofen is the NSAID of choice at some centres. Availability, fashion, relative cost, patient convenience, and adverse effects profile dictate choice as much as efficacy.

NSAIDs act at several sites, both peripheral and central. The central analgesic effects of NSAIDs have not been fully elucidated and it is possible that clinically important differences may emerge when more data are available. NSAIDs inhibit cyclo-oxygenase (COX), an important enzyme in the arachidonic acid cascade which results in the production of tissue and inflammatory prostaglandins. Inhibition of prostaglandin synthesis does not account for the total analgesic effect of NSAIDs, however, although it appears to explain most of the adverse effects.

COX exists in two forms; COX-1 is constitutive (i.e. present in all normal tissues), whereas COX-2 is normally undetectable in most tissues but massively induced by inflammation (Fig. 6). By producing selective COX-2 inhibitors, it is hoped to reduce the gastric toxicity of NSAIDs. COX-1 inhibition alone, however, does not explain the differential impact of NSAIDs on the gastrointestinal tract; uncoupling of oxidative phosphorylation is possibly as important. COX-2 is normally present in the kidney, although its function there has not yet been elucidated. The possibility that selective COX-2 inhibitors may cause renal dysfunction must not be ignored.

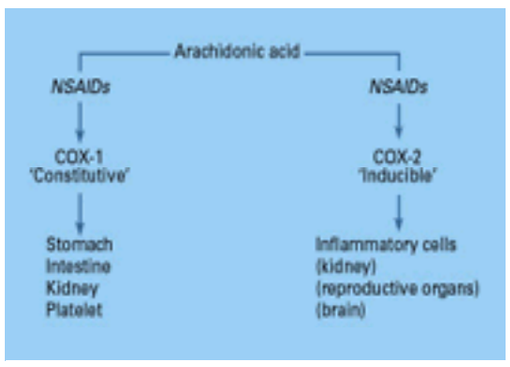


Fig. 6. Important sites of COX-1 and COX-2 production. – indicates inhibition; organs in parenthesis indicate relatively minor constitutive sources of COX-2.

In hot countries, NSAIDs are sometimes discouraged because of the risk of acute renal failure in hypovolaemic patients. Paracetamol is used instead. The main drawback with paracetamol is the frequency of administration (every 4–6 h). However, because NSAIDs and paracetamol differ in their mode of action, the two can be used together with an additive effect. NSAIDs differ in their effect on platelet function. In patients undergoing chemotherapy or with thrombocytopenia from other cause, it is best to use an NSAID which has no effect on platelet function (Table 2).

Drug	Affects platelets	Comment
Aspirin	+	Irreversible platelet dysfunction as a result of acetylation of platelet COX-1
Non-acetylated salicylates	-	No effect at recommended dose (choline magnesium trisalicylate, diflunisal, salsalate)
Non-selective COX inhibitors	+	Reversible platelet dysfunction (includes all other long-established NSAIDs)
Preferential COX-2 inhibitors	-	Not clear yet, unavailable
Selective COX-2 inhibitors	-	Celecoxib, rofecoxib

Table 2 NSAIDs and platelet function

Weak opioids

Weak opioids in combination with non-opioids are step 2 analgesics (Table 3). A large dose of a weak opioid is as effective as a small dose of morphine. This means that step 2 is pharmacologically unnecessary. In many countries, however, it is much easier to obtain and supply a weak opioid preparation. Codeine is about 1/10 as potent as morphine. The typical dose range for codeine is 30 to 60 mg every 4 h. Much of the analgesic effect of codeine depends on biotransformation to morphine. Dihydrocodeine is often used instead in the United Kingdom.

Generic name	Codeine content	Non-opioid content	Net price
Co-codaprin 8/400 tablets	8 mg	aspirin 400 mg	30 = 36p
Co-codaprin 8/400 dispersible tablets	8 mg	aspirin 400 mg	20 = 54p
Co-codamol 8/500 tablets	8 mg	paracetamol 500 mg	20 = 23p
Co-codamol 8/500 dispersible tablets	8 mg	paracetamol 500 mg	20 = 65p
Co-codamol 30/500 tablets	30 mg	paracetamol 500 mg	20 = £1.50
Co-codamol 30/500 effervescent tablets*	30 mg	paracetamol 500 mg	20 = £1.89

*Contains 13.6 mmol Na^+ /tablet; avoid in renal impairment or cardiac failure.

Table 3 Commonly used weak opioid compound preparations (UK)

Strong opioids

Globally, morphine remains the strong opioid of choice; other strong opioids are used mainly when morphine is not readily available or when the patient has intolerable adverse effects with morphine. Morphine is administered as tablets (‘normal-release' 10 mg, 20 mg, 50 mg) or aqueous solutions (10 mg/5 ml, 100 mg/5 ml). An increasing range of modified-release preparations is available—tablets, capsules, and suspensions. There are no generic modified-release morphine tablets and because of differing pharmacokinetic profiles, it may be best to keep individual patients on the same brand. Most are administered twice daily, some once daily. Patients can be started on either an ordinary (‘normal release') or a modified-release preparation (see Box 3). The use of morphine is determined by analgesic need and not by the doctor's estimate of life expectancy—which is often wrong. The correct dose of morphine is what gives adequate relief for 4 h without unacceptable adverse effects. Maximum recommended doses, derived mainly from postoperative parenteral single dose studies, are not applicable for patients with cancer pain.

Box 3. No caption available.

Examination
- To establish trust
- To confirm site
- To identify cause

Explanation to reduce psychological impact of pain

Modification of pathological process
- Radiation therapy
- Hormone therapy
- Chemotherapy
- Surgery

Analgesics
- Non-opioids
- Opioids
- Adjuvant
  - corticosteroids
  - antidepressants
  - anticonvulsants
  - muscle relaxants

Non-drug treatment (physical)
- heat pads
- TENS

Non-drug treatment (psychological)
- relaxation
- cognitive-behavioural therapy
- psychodynamic therapy

Interruption of pain pathways
- Local anaesthesia
  - lidocaine (lignocaine), bupivacaine
- Neurolysis
  - chemical (alcohol, phenol, chlorocresol)
  - cold (cryotherapy)
  - heat (thermocoagulation)
- Neurosurgery
  - cervical spinothalamic
  - tractotomy (cordotomy)

Modification of way of life and environment
- Avoid pain-precipitating activities
- Immobilization of the painful part
  - cervical collar
  - surgical corset
  - slings
  - orthopaedic surgery
- Walking aid
- Wheelchair
- Hoist

When adjusting the dose of morphine, generally increase by 33 to 50 per cent. Two-thirds of patients never need more than 30 mg every 4 h (modified-release morphine 100 mg every 12 h); the rest need up to 200 mg every 4 h (modified-release morphine 600 mg every 12 h) and occasionally more. Extra doses for ‘breakthrough'pain should be the same as the 4-hourly dose of morphine and *are increased when the regular dose is increased*. Instructions must be clear: an extra (as required) dose of morphine does not mean that the next dose of regular morphine is omitted.

An antiemetic should be supplied (e.g. haloperidol 1.5 mg at once and at bedtime) and a laxative (e.g. co-danthrusate or senna and docusate). *Constipation may be more difficult to manage than the pain*. Suppositories and enemas continue to be necessary in about one-third of patients. Warn patients about the possibility of initial drowsiness. If swallowing is difficult or vomiting persists, change from the oral to the subcutaneous route, preferably by infusion using a portable syringe driver. In the United Kingdom one-third of the oral dose of morphine is given as diamorphine; elsewhere half of the oral dose of morphine should be given. Alternatively, morphine may be given rectally (same dose as orally).

It is pharmacological nonsense to prescribe two weak opioids or two strong opioids concurrently. It is sometimes justifiable for a patient receiving morphine to be given another opioid (weak or strong) for breakthrough pain. On the other hand, patients could perhaps more easily be advised to take an extra dose of morphine. Differences between opioids relate in part to differences in receptor affinity. Strong opioids tend to cause the same range of adverse effects (Box 4), although to a varying extent. *Because pain is a physiological antagonist to the central depressant effects of morphine*, strong opioids used correctly do not cause clinically important respiratory depression in cancer patients in pain. In contrast to postoperative patients, cancer patients with pain:

Box 4. No caption available.

Common initial	**Occasional**
Nausea and vomiting	Dry mouth
Drowsiness	Sweating
Unsteadiness	Pruritus
Delirium(confusion)	Hallucinations
	Myoclonus
Common ongoing	**Rare**
Constipation	Respiratory depression
Nausea and vomiting	Psychological dependence

- generally have already been receiving a weak opioid (i.e. are not opioid naïve)
- take medication by mouth (slower absorption, lower peak concentration)
- titrate the dose upwards step by step (less likelihood of an excessive dose being given).

It is therefore extremely rare to need to use naloxone (a specific opioid antagonist) in palliative care. Because of the possibility of an additive sedative effect, however, care needs to be taken when strong opioids and psychotropic drugs are used concurrently.

Tolerance to strong opioids is not a practical problem. The fear of patients developing psychological dependence (addiction) is unfounded and should not limit the use of strong opioids for cancer pain. Caution in this respect should be reserved for patients with a present or past history of substance abuse; even then strong opioids should be used when there is clinical need. Physical dependence does not prevent a reduction in the dose of a strong opioid if the patient's pain ameliorates, e.g. as a result of radiotherapy or a nerve block.

In patients who have intolerable adverse effects with morphine, it may be necessary to substitute an alternative strong opioid (Table 4). The first step, however, is to reduce the dose of morphine (and possibly other drugs) and review. For adverse effects such as cognitive failure, hallucinations or myoclonus, hydromorphone and oxycodone can be used. However, for patients experiencing opioid-induced hyperexcitability, methadone is preferable. Methadone is a mu-receptor agonist, an *N*-methyl-d-aspartate-receptor channel blocker, and a presynaptic blocker of serotonin uptake. Because it is intrinsically more broad-spectrum than morphine, methadone may be more effective for morphine poorly-responsive neuropathic pain.

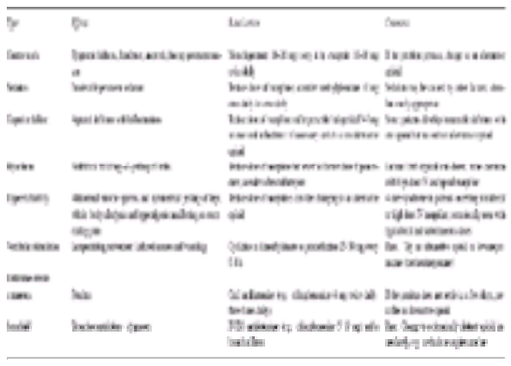


Table 4 Potential intolerable effects of morphine

Transdermal fentanyl is increasingly used because of its convenience for patients. Patches are applied every 3 days. Fentanyl is significantly less constipating than morphine. Pethidine and dextromoramide, however, have short durations of action and are not recommended for regular prophylactic analgesia. Because of its rapid onset of action, some centres use dextromoramide for breakthrough pain in patients taking regular morphine, or as prophylactic additional analgesia before a painful dressing or other procedure. In contrast, at other centres, such procedures are timed to coincide with the peak plasma concentration after a regular dose of morphine as required. Pentazocine should not be used; it is a weak opioid by mouth and often causes psychotomimetic effects (dysphoria, depersonalization, frightening dreams, hallucinations). When converting from an alternative strong opioid to oral morphine, the initial dose depends on the relative potency of the two drugs (Table 5).

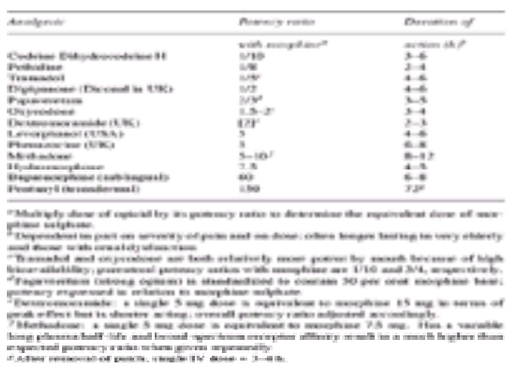


Table 5 Approximate oral analgesic equivalence to morphine

Adjuvant analgesics

These include corticosteroids, tricyclic antidepressants, anticonvulsants, and muscle relaxants. Corticosteroids are important for pain associated with nerve root or nerve trunk compression, and for pain associated with spinal cord compression. Dexamethasone 4 to 8 mg once daily. is prescribed for the former and 12 to 20 mg (or more) for cord compression, followed by radiotherapy.

Nerve injury pain generally results from infiltration of a nerve.The pain is dermatomal in distribution and is typically superficial and burning in character with associated allodynia (i.e. light touch and other non-noxious stimuli cause pain). There may also be spontaneous stabbing pain. Nerve injury pains which do not respond well to morphine may respond to methadone. Methadone is a more difficult drug to use because its plasma half-life ranges from about 8 to 80 h, and specialist advice should be sought. Nerve injury pains are often eased, however, by an antidepressant (e.g. amitriptyline 25 to 75 mg every night) or an anticonvulsant (e.g. sodium valproate 400 to 1000 mg every night) or both, generally used in combination with morphine and a non-opioid. Further options include class 1 anti-arrhythmic drugs (e.g. mexiletine), *N*-methyl-d-aspartate-receptor channel blockers (e.g. ketamine), and spinal analgesia (e.g. morphine + bupivacaine ± clonidine). Some 2 to 3 per cent of patients referred to a specialist palliative care unit need spinal analgesia sooner or later.

Very occasionally a cancer patient presents with a pain which is dependent on the integrity of the regional sympathetic nerves and which responds very poorly to ‘broad-spectrum' analgesic therapy (opioids ± non-opioids ± adjuvant analgesics). The essential features are pain (often burning), a sensory disorder, and relief following a regional sympathetic nerve block. Differentiation between nerve injury burning pain and SYmpathetically maintained burning pain may be difficult because clinical features are not constant and some are common to both conditions. Because sympathetic nerves travel with the arterial SYstem, the distribution of sympathetically maintained pain is vasotopographic rather than neurodermatomal (but this may not be clinically obvious). There may also be radiographic evidence of osteoporosis and of hot spots on isotope bone scans which may be mistaken for osteolytic metastases. If sympathetically maintained pain is suspected in a cancer

patient, a diagnostic sympathetic block should be performed with local anaesthetic. This not only serves to confirm the diagnosis but may also give relief that outlasts the duration of the local anaesthetic effect. A neurolytic block should be considered for lower limb pain that returns.

Muscle spasm pain (cramp) secondary to underlying bone pain and/or skeletal deformity is common in cancer patients. Myofascial trigger point pains also occur. The correct approach is explanation, physical therapies (local heat and massage), diazepam and relaxation therapy, and injection of the trigger points with local anaesthetic and a corticosteroid (e.g. bupivacaine 0.5 per cent and depot methylprednisolone 80 mg). However severe, morphine is not indicated for the relief of cramp and trigger point pains.

Conclusions

Satisfactory pain relief can be achieved in the majority of cancer patients. The patient's pains, however, must be evaluated rapidly and accurately. The doctor must appreciate that pain is a somatopsychic experience, and understand the pathological processes which give rise to pain. He must be aware that some pains experienced by patients with advanced cancer are not caused directly by the cancer, and be able to recognize pain caused by muscle spasm. Familiarity with the use of antidepressants and anticonvulsants in neuropathic pain and knowledge of the potential benefits of corticosteroids are necessary for optimum treatment. Finally, patients prescribed analgesics or other drugs must be closely supervised to assess benefit, monitor adverse effects, and to modify treatment when necessary.

Further reading

Ahmedzai S, Brooks D. Transdermal fentanyl versus sustained-release oral morphine in cancer pain: preference, efficacy and quality of life. *Journal of Pain and Symptom Management* 1997; **13**: 254–61. [The report of a random controlled trial of transdermal fentanyl and modified release morphone; although of equal efficacy, more patients preferred the convenience of a skin patch every 3 days.]

Borgbjerg FM *et al.* Experimental pain stimulates respiration and attenuates morphine-induced respiratory depression: a controlled study in human volunteers. *Pain* 1996; **64**: 123–8. [Confirmed the clinical dictum that pain is the physiological antagonist to the central depressant of morphine (and other opioids.]

Doyle D, Hanks GWC, MacDonald N, eds. *Oxford textbook of palliative medicine*, 2nd edn. Oxford University Press, Oxford, 1997. [Contains several chapters on cancer pain management and provides doctors with a wealth of additional information by authors from both sides of the Atlantic.]

Grond S *et al.* Assessment of cancer pain: a prospective evaluation in 2266 cancer patients referred to a pain service. *Pain* 1996; **64**: 107–14. [Summarizes the results obtained with the WHO three-step analgesic ladder in over 2000 patients with advanced cancer and pain.]

Grond S *et al.* Transdermal fentanyl in the long-term treatment of cancer pain: a prospective study of 50 patients with advanced cancer of the gastrointestinal tract or the head and neck region. *Pain* 1997; **69**: 191–8. [Confirms that fentanyl is less constipating than morphine.]

Grond S *et al.* Assessment and treatment of neuropathic cancer pain following WHO guidelines. *Pain* 1999; **79**: 15–20. [Demonstrates that the results obtained with the WHO three-step analgesic ladder are as good with neuropathic pain as with nociceptive pain.]

Hanks G *et al.* Morphine in cancer pain: modes of administration. *British Medical Journal* 1996; **312**; 823–6. [The report of an expert committee of the European Association for Palliative Care; provides consensual guidelines for the use of morphine in cancer pain.]

McQuay H *et al.* Anticonvulsant drugs for management of pain: a systematic review. *British Medical Journal* 1995; **311**: 1047–52. [The evidence-base for the use of anti-epileptics as analgesics in neuropathic pain.]

McQuay H *et al.* A systematic review of antidepressants in neuropathic pain. *Pain* 1996; **68**: 217–27. [The evidence-base for the use of antidepressants as analgesics in neuropathic pain.]

Morley JS, Makin MK. The use of methadone in cancer pain poorly responsive to other opioids. *Pain Reviews* 1998; **5**: 51–8. [A description of the place and safe use of methadone in cancer pain management.]

Sjogren P *et al.* Disappearance of morphine-induced hyperalgesia after discontinuing or substituting morphine with other opioid antagonists. *Pain* 1994; **59**: 313–16. [Morphine hyperexcitability and its management by switching to an alternative strong opioid.]

Twycross RG. *Pain relief in advanced cancer*. Churchill Livingstone, Edinburgh, 1994. [An extensive account of the author's understanding of pain in cancer and its management.]

Twycross R. *Oral morphine in advanced cancer*, 3rd edn. Beaconsfield Publishers, Beaconsfield, 1997. [60+ common questions and answers about the use of morphine in advanced cancer; a valuable teaching aid.]

Twycross R, Wilcock A, Thorp S. *Palliative care formulary*. Radcliffe Medical Press, Oxford, 1998. [Brings together pharmacological and pharmaceutical information about all the common symptom relief drugs used in palliative care in the United Kingdom.]

World Health Organization. *Cancer pain relief*. World Health Organization, Geneva, 1986. [Contains the original description of the WHO Method for Control of Cancer Pain (three-step analgesic ladder etc.), together with much background material.]

World Health Organization. *Cancer pain relief*: *with a guide to opioid availability*. World Health Organization, Geneva, 1996. [The Revised WHO Method for Relief of Cancer Pain with additional information about obtaining medicinal supplies of morphine and other strong opioids.]

49.3 Relief of chronic non-malignant pain

Henry McQuay

[Pain sensation and transmission](#)
[Psychological factors](#)
[Patients, pain histories, and recording pain](#)
[Pain histories](#)
[Recording pain](#)
[Chronic pain treatment](#)
[Non-opioid analgesics](#)
[Opioids](#)
[Unconventional analgesics](#)
[Nerve transmission block](#)
[Alternatives](#)
[Further reading](#)

This chapter should act as a signpost, directing the interested reader to the best evidence, and to SYstematic reviews and randomized controlled trials where available. The chapter takes the reader through the various interventions used in pain management, be they drugs, injections, operations, psychological, or physical.

Pain sensation and transmission

The easiest way to think of pain in the nervous SYstem is the idea of pain receptors and nerve cables dedicated to the transmission of pain signals, a hard-wired, line-labelled system. This view has always had obvious flaws. The return of pain after an initially successful cordotomy, and the phenomenon of phantom limb pain are two examples. In a hard-wired line-labelled system the pain should not recur after the cordotomy and patients should not feel pain in a limb as they did before the accident or amputation. Such flaws mean that the simple view of a ‘passive' nervous SYstem does not explain all that we see.

For most acute pain the idea of specific cables whose transmission can be blocked is reinforced by the fact that you can perform herniorrhaphy (painlessly) by using local anaesthetic block or indeed a regional block. The receptors and cables of that region are temporarily disabled by the local anaesthetic, and no pain message gets through to the brain. In some chronic pain syndromes the inadequacy of this simple explanation is exposed. An example is phantom pain, because the painful foot or hand, the receptors and the cables, are no longer there. The concept of ‘pain memory' in the spinal cord and brain has to be invoked.

The idea that the nervous SYSTEM can change, or be ‘plastic', has led to the use of the word plasticity, and this concept has had a major impact in both acute and chronic pain. The simplest idea is that a memory of a pain is made and stored in the nervous system. Preventing such a memory being laid down led to the idea of pre-empting postoperative pain; any pain after the operation would be easier to treat if the memory had been minimized. Long-term sequelae, such as the phantom pain which can occur after amputation, might also be prevented. One issue of importance for treatment is at what level of the nervous SYSTEM such memories might be stored. If the memory is stored at a ‘central' level, for instance in the brain, then attacking the pain in the leg or arm where it originated might not do any good. The memory might be held centrally but require continued input from the periphery to sustain it. Attacking the pain in the leg or arm would then have some logic.

Many treatments, or interventions, are used to treat both acute and chronic pain. Chronic pain, not surprisingly, has more twists and turns, because its origins may be more complicated and because the nervous system if damaged or bombarded continuously by pain messages can behave strangely. The concept of plasticity has led to some interventions coming into fashion, and to some going out of fashion. Long-term measures, such as cutting the nerves thought to carry a particular pain message, are going out of fashion. The reasoning is that the nervous system will ‘rewire', the pain will re-emerge, and may well be more difficult to manage than it was in the first place. Better drug control of difficult pains has also reduced the necessity for destructive procedures.

Psychological factors

Pain signals can be amplified or damped by endogenous influences such as mood or endorphins, or exogenous factors, drugs given or the circumstances of the injury. The classic damping-down scenario is the injured soldier who continues, despite a shattered leg, to get himself out of battle. Conversely a stubbed toe when you are tired and miserable is immeasurably more painful than the same injury on a cheerful morning. In chronic pain it is sometimes very hard to disentangle depression from pain. Pain makes depression worse and depression makes pain worse. This pattern is all too familiar to those who manage back pain. The thinking clinician needs to deal with both the pain and the depression.

Patients, pain histories, and recording pain

Surgical referrals to pain centres are usually requests to see a patient after surgery because of pain which has resisted conventional treatment and because further surgery is inappropriate. There are few objective signs which the doctor can use to judge the severity of reported pain. Pain is necessarily subjective. Many patients have no obvious (visible) handicap. The most important principle is that the patient and the doctor are best served if the doctor believes the patient's report. Their problems may be ill-understood, even disbelieved, at work and at home. Chronic pain changes people, affecting their personal and working lives, and ultimately their personalities. Often such changes are reversible with successful treatment. Much time and energy is wasted on procedures designed to ‘catch the patient out'. Labelling patients as malingerers or the pain as psychogenic may be easier than admitting that there is no successful treatment. The differential diagnosis for many pain conditions includes both cancer and non-cancer causes.

Pain histories

The extra emphasis of a pain history compared with a normal medical history is summarized in Table 1.

Site of pain	Where do you feel this pain? Does it go anywhere else? Is it numb where you feel the pain?
Character of pain	What sort of a pain is it? Burning/shooting/stabbing/dull etc.
History of pain	How long have you had this pain? How did it start? Did it come on out of the blue or was it triggered by something?
Relieving factors	Does anything make it better? (position, drugs, distraction, alcohol etc.)
Aggravating factors	Does anything make it worse? (position/exercise/weather etc.)
Pattern	Is there any pattern to the pain? (frequency/severity) Is it worse at any particular time of day?
Sleep disturbance	Do you get off to sleep with no trouble? Does the pain wake you up?
Activities	What does the pain stop you doing which you would otherwise do?
Previous treatments	What methods have been tried already? Did they help the pain?

Table 1 Special features of pain history

Asking whether sensation is normal in the painful area and whether the pain is shooting or stabbing in character may help to identify the pains variously known as dysaesthetic, deafferentation, or neuropathic. It is important to distinguish these because they are unlikely to respond to conventional analgesics. These pains often have little pattern, but are less troublesome when the patient is distracted.

It is important to enquire specifically about the efficacy of each particular drug class. This information prevents the inept prescription of drugs which have failed previously, and may give important clues as to the kind of pain and its sensitivity to different classes of drug.

It is also important to know the dose size, frequency, duration of prescription, and the side-effect problems for each drug which has been prescribed. Dose–response relationships apply with analgesics, and therapeutic failure should not be presumed if the dosage was inadequate; a good example is the use of carbamazepine in trigeminal neuralgia. Is the patient taking drugs other than analgesics? Anticoagulation, for instance, is not only an (almost) absolute contraindication to pain management by injection procedures, but also interacts with some anti-inflammatory drugs.

It is important to be sure whether nerve blocks used previously were technically effective (e.g. did the patient have any numbness after an epidural which included local anaesthetic?), before dismissing them as of no help to this patient. Equally other measures, such as transcutaneous nerve stimulation, may not have been used correctly, and may succeed if the patient receives proper instruction in their use.

Recording pain

For chronic pain, when patients are recording over long periods, we tend to use diaries with word categories, and ask about both pain intensity (none, mild, moderate, or severe) and pain relief (none, slight, moderate, good, or complete). Another useful scale is the patient's ‘global report'. Using the same categories as the pain relief scale, this is a composite view encompassing both pain itself and any adverse effects of treatment. The clinician's global view is a notorious overestimate and should not be used.

Most important and simplest is the idea of including a binary scale because of its clinical relevance. The question is phrased in various forms, around the idea of ‘Is your pain at least 50 per cent relieved?', to which the patient answers yes or no.

The argument for using pain charts as part of normal practice is to improve the quality of care. It is the fact of a chart rather than the form of the chart which is most important. These should be done together with vital signs, and can then be used for both clinical care and for audit.

Some index of function is necessary as a base line so that improvement or deterioration may be monitored. Useful clinical outcome measures are notoriously difficult; using simple pain charts and indices of activity can work well.

Chronic pain treatment

This chapter uses systematic reviews when possible to provide the best available evidence for the various interventions used to relieve pain. The number needed to treat (**NNT**) is used as the measure of clinical significance from quantitative SYstematic reviews. It shows the effort required to achieve a particular therapeutic target. NNT is treatment specific. It describes the difference between active treatment and control. The NNT is given by the equation

$$\mathrm{NNT} = \frac{1}{(\mathrm{IMP}_{act} / \mathrm{TOT}_{act}) - (\mathrm{IMP}_{con} / \mathrm{TOT}_{con})}$$

where:

IMP_{act} = number of patients given active treatment achieving the target

TOT_{act} = total number of patients given the active treatment

IMP_{con} = number of patients given a control treatment achieving the target

TOT_{con} = total number of patients given the control treatment

An NNT of 1 describes an event which occurs in every patient given the treatment but in no patient in a comparator group. This could be described as the ‘perfect' result in, say, a therapeutic trial of an antibiotic compared with placebo. For therapeutic benefit the NNT should be as close as possible to 1; there are few circumstances in which a treatment is close to 100 per cent effective and the control or placebo completely ineffective, so NNTs of 2 or 3 often indicate an effective intervention. For unwanted effects, NNT becomes the NNH (number-needed-to-harm), which should be as large as possible.

Most chronic pain is managed initially with analgesics, but may also involve nerve transmission block and alternative methods.Figure 1 and Figure 2 show the simple comparison with acute pain and the alternatives. As acute pain wanes weaker analgesics are used. If chronic pain increases stronger ones are necessary.

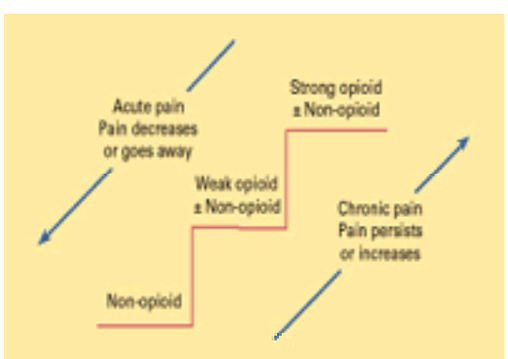


Fig. 1. WHO analgesic ladder in acute and chronic pain.

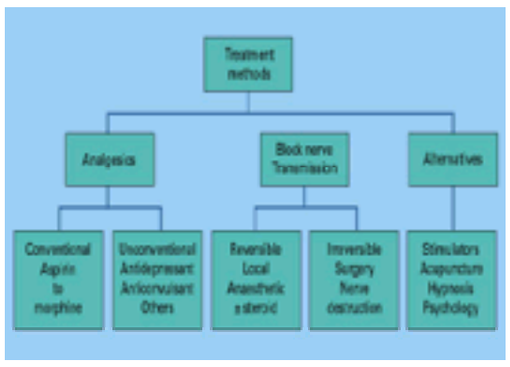


Fig. 2. Treatment methods.

The same analgesics, from non-steroidal anti-inflammatory drugs (**NSAIDs**) through to opioid, are used in chronic as in acute pain. If analgesics relieve the pain to an adequate extent, and with tolerable or controllable adverse effects, then there is little reason to use other interventions. If analgesics are ineffective, other methods have to be considered. If analgesics are effective but cause intolerable or uncontrollable adverse effects, then again other methods should be considered. The effectiveness and the adverse effects of the analgesics are critical.

We know from work with cancer pain that using analgesics according to the WHO ladder (Fig. 1) can relieve pain for 80 per cent of patients. For most of the 80 per cent the relief will be good, for a minority it will only be moderate. This presumes that the pain is managed optimally, and we know from audit that this is often not the case. Optimal management requires that the correct drugs are available, and that they are given in the correct dose by the correct route and at the correct time. This needs staff who are well versed in the problems, and who are available to care for the patient. The second problem is the 20 per cent of patients whose pain is not well managed by intelligent use of analgesic guidelines. The other treatment methods on Fig. 2 are necessary to manage those for whom analgesics fail.

Non-opioid analgesics

Oral NSAIDs, combinations, and others

Choosing the best analgesic for long-term use involves the same decisions as in acute pain. Most comparisons are done using single doses, whereas patients with chronic pain take multiple doses. Figure 3 shows an evolving league table for analgesic performance after surgery.

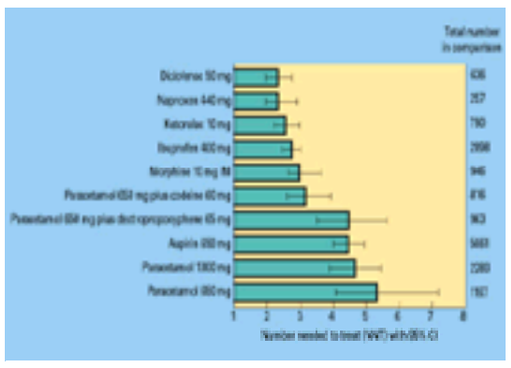


Fig. 3. League table of number needed to treat (NNT) for postoperative pain. NNT for at least 50 per cent pain relief over 4 to 6 h in patients with moderate to severe pain, all oral analgesics except intramuscular morphine, all single-dose studies.

Ibuprofen 400 mg (and other NSAIDs analysed) will produce at least 50 per cent relief of pain for one out of two postoperative patients, paracetamol 1 g for one out of four patients.

Although this efficacy league table (Fig. 3) was derived from acute pain trials we believe it is also valid in chronic pain. The clear guideline is that the first choice is an oral NSAID if there is no contraindication. No single-dose trial has shown any efficacy advantage of one NSAID over another, although this does not fit well with patients' reports on multiple dosing of increased efficacy from NSAIDs with greater anti-inflammatory action. Gastric bleeding is a problem of chronic use. The risk of NSAID-induced gastric bleeding is lowest with ibuprofen, and increases with increasing age. Prophylactic misoprostol or omeprazole should be considered for preventing NSAID-associated gastrointestinal complications when the patient's age is greater than 75 years, there is cardiovascular disease, or a history of peptic ulcer or of gastrointestinal bleeding (NNTs to prevent one serious gastrointestinal complication in 1 year were 105, 58, 11, and 7, respectively). Acute renal failure can be precipitated in those with pre-existing heart or kidney disease, those on loop diuretics, or those who have lost more than 10 per cent of blood volume.

The efficacy dose–response curve for NSAIDs is flatter than the dose–response for adverse effects such as gastrointestinal symptoms, dizziness, and drowsiness. Increasing the dose to improve analgesia is therefore more likely to increase adverse effects than to improve analgesia.

There is an old adage that if patients can swallow it is best to take drugs by mouth. This is very pertinent for pain relief after surgery or trauma. Effective relief can be achieved with oral NSAIDs, and there is no evidence that NSAIDs given by injection perform better than the same drug at the same dose given by mouth.

Centrally acting non-opioids include paracetamol, dipyrone, and nefopam. Dipyrone is widely used in certain countries, and is an effective analgesic which can produce blood dyscrasias. The lack of comparative evidence makes it hard to rank its risk:benefit profile against the other analgesics.

Comparisons

NSAIDs alone produced as good analgesia as single or multiple doses of weak opioids alone or in combination with non-opioid analgesics. Adverse effect incidence and patient dropout rates were the same for multiple doses of NSAIDs or weak opioids in combination with non-opioid analgesics. Conversely, two studies suggest that thereis little advantage in osteoarthritis of either NSAIDs over paracetamol or weak opioids in combination with paracetamol over paracetamol alone.

Topical NSAIDs

Many doctors remain sceptical about the efficacy of topical NSAIDs. This may not be correct. Published randomized controlled trials on chronic pain conditions (mainly knee osteoarthritis) studied over 800 subjects treated with topical NSAIDs and 322 subjects who received placebo. The analgesic response for combined placebo treatment was 30 per cent, and for combined topical non-steroidal anti-inflammatory preparations it was 63 per cent. The number needed to treat was 3.2 (2.6 to 4.1). Rubbing on a topical NSAID is much more effective than rubbing on a placebo.

Opioids

In chronic pain there are two particular problems with opioids. The first is that adequate doses are often not available or are not given, primarily because of fears of addiction. The second is that some (rarer) chronic pain states, particularly when the nervous SYstem is damaged, may not respond fully to opioids.

The main reason that adequate doses are withheld by doctors or nurses is the fear of respiratory depression. Opioids used for people who are not in pain, or in doses larger than necessary to control the pain, can slow and indeed stop breathing. The principle is that the dose has to be titrated to the effect (Fig. 4). The effect is pain relief. If the dose given has not produced pain relief (the patient is still complaining of pain), and it has all been delivered and absorbed, then it is safe to give another dose. This subsequent dose may be smaller than the first. If it too doesn't succeed, the process can be repeated.

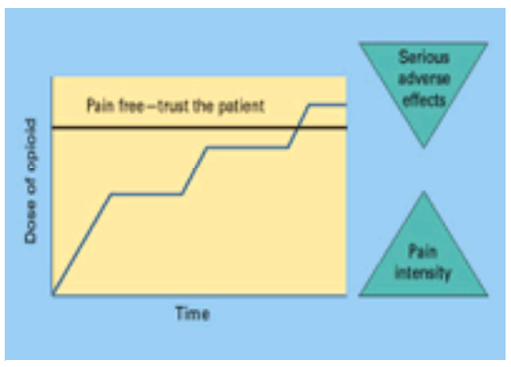


Fig. 4. Opioid titration to effect. The principle is that the dose has to be titrated to the effect. The effect is pain relief. Doses larger than necessary to control the pain can cause serious adverse effects. If the dose given has not produced pain relief because the patient is still complaining of pain, and it has all been delivered and absorbed, then it is safe to give another dose, which may be smaller than the first.

There is no compelling evidence that one opioid is better than another, but there is good evidence that pethidine has a specific disadvantage. Given in multiple doses its metabolite norpethidine can accumulate and act as a central nervous SYstem irritant, ultimately causing convulsions. This is more likely when there is renal dysfunction. This toxic metabolite should preclude using pethidine when multiple doses are needed. The old idea that pethidine is better than other opioids at dealing with colicky pain is no longer tenable.

Morphine (and its family including heroin and codeine) has an active rather than a toxic metabolite, morphine-6-glucuronide. In renal dysfunction this metabolite can accumulate and result, because it is more active than morphine, in greater effect from a given dose. If you are, as you should be, titrating dose against effect, this will not matter. Less morphine will be needed. It can be a problem with unconscious patients being dosed by the clock, and whose renal function is compromised.

Opioids used for people who are not in pain can induce physical and psychological dependence. This does not happen to patients who receive them for pain relief, for instance after an operation or for severe pain from osteoporotic vertebral collapse. Some governments restrict medical availability on the grounds that if the drugs are available medically this will worsen the street addiction problem. There is no evidence for this. The casualties are patients who are deprived of adequate pain relief.

In chronic pain opioids are usually given by mouth. The dose is worked out by titration over a period of days, and then the drug is given regularly, not waiting for the pain to come back. Initial problems with nausea or dizziness commonly settle. If constipation is likely, laxatives are given.

Patients who cannot swallow can try sublingual, transdermal, or suppository dosing. Subcutaneous infusion, usually from a small (external) pump is used for terminal patients who cannot manage these other routes. Rarely the epidural route is used for combination infusion of opioid and local anaesthetic.

If patients' pain starts to increase, the dose is increased. If sensible dose increases do not produce pain relief, or if increasing the opioid dose provokes intolerable or unmanageable adverse effects, then other methods have to be considered, either as well as the opioid or instead of it. A working rule is that if the pain is in a numb area, which is a marker for a damaged nervous system, we would be less confident that opioids would necessarily produce pain relief, and our threshold for using other strategies would be lower.

Unconventional analgesics

Unconventional analgesics are drugs which have other indications in other medical settings and are not normally thought of as analgesics. Treating chronic pain in a tertiary hospital setting we use these drugs for about one-third of our patients. The hallmark is pain in a numb area, neuropathic pain.

When the patient has symptoms and signs of nervous system damage in the area of their pain we expect the response to conventional analgesics to be reduced. Conventional analgesics have often failed already, which is why the patient has been referred. If not, we try the conventional drugs, before we embark on empirical testing to see if any of the unconventional analgesics can provide relief.

Antidepressants

Antidepressants work on the nervous system to relieve depression. We use them in much lower dosage (about half), and we use them to relieve pain. Classically they were used to relieve pain that was burning rather than shooting in character, and anticonvulsants were used for shooting pains. Now we tend to use antidepressants as first line for both types of pain, because we have greater success and because we believe the antidepressants cause fewer adverse effects (but see below).

We use low doses (median 75 mg amitriptyline nocte, maximum 150 mg) compared with those used to control depression. The pain relieving effect happens, if it is going to happen, well within a week, whereas 10 days is the minimum often quoted for an antidepressant effect. The older (tricyclic) antidepressants seem to be better than the selective serotonin reuptake inhibitors as analgesics. The simplest analogy is that these older drugs are like shotguns, acting on multiple transmitter pathways, whereas the newer ones are more like rifles, designed as they are to be more selective and affect only one pathway.

Anticonvulsants

Anticonvulsants have been used for many years to treat the shooting pains of trigeminal neuralgia and of diabetic neuropathy. How they work has never been clear. The catch-all explanation was that they stabilized nerve membranes, preventing them carrying spurious messages. The current fashionable explanation is that these drugs act as antagonists on the *N*-methyl d-aspartate (**NMDA**) mechanism.

Anticonvulsants can provide good relief in neuropathic pain (NNT of 2 to 3, Table 2). Doses required for analgesic effect are close to the anticonvulsant dosing range, and carry a perceived burden of adverse effects. The SYstematic reviews suggest that there is little difference between the antidepressant and anticonvulsant adverse effects. Our main use is the traditional role in trigeminal neuralgia and diabetic neuropathy. We also use anticonvulsants in shooting pain which does not respond to antidepressants. Two examples are phantom limb pain and pain in the head and neck due to tumour.

Condition	*Intervention*	*Outcome (% pain relief)*	*NNT*
Postoperative pain	(good) Ibuprofen 400mg	>50	2
	Paracetamol 1 g	>50	4
	(poor) Codeine 60 mg oral	>50	>10
Back pain	Epidural steroid	>75 at 60 days	>6
Acute sprains etc.	Topical NSAID (good)	>50	2•
Trigeminal neuralgia	Anticonvulsants	>50	2.5
Diabetic neuropathy	Anticonvulsants	>50	2.5
Diabetic neuropathy	Topical capsaicin	>50	4.2
Neuropathic pain	Antidepressants	>50	2.5

Table 2 Number needed to treat (NNT) for some analgesic interventions in chronic pain

Others

Clonidine and other a_2-adrenergic agonists have analgesic effects, both in conventional pain and in neuropathic pain. They extend the duration of local anaesthetic effect and have a synergistic effect with opioids. Their clinical utility is limited by the adverse effects of sedation and hypotension. In neuropathic cancer pain epidural clonidine was effective. Baclofen is used by intrathecal pump to treat the painful spasms of cerebral palsy. Ketamine and dextromethorphan, both drugs with NMDA antagonist action, are being used in severe neuropathic pain.

Nerve transmission block

Reversible

Local anaesthetics

Local anaesthetics block nerve conduction reversibly. When the local anaesthetic wears off the pain returns. That is the pharmacologically correct statement, but another old saying, that a series of local anaesthetic blocks can be used to ‘break the cycle' of pain and effect a cure, now has some empirical support, even if we do not understand the mechanism. Arner and colleagues showed that the duration of pain relief could far outlast the duration of local anaesthetic action, and that prolonged relief could result from a series of blocks. Local anaesthetic blocks can thus be diagnostic and therapeutic. Diagnosis of pain for instance from a ‘trapped' lateral cutaneous nerve of the thigh can be confirmed by local anaesthetic block, and a series of blocks may prevent pain recurring.

Pain clinics use such blocks commonly, for shoulder pain, intercostal neuralgia, rectus sheath nerve entrapment, postoperative scar pains, and other peripheral

neuralgias (Table 3). What is not clear is the extent to which adding steroid to the local anaesthetic makes a difference, either prolonging the duration of effect of a particular procedure or increasing the chance of success of a series of blocks.

Block	Common indications
Trigger point	Focal pain (e.g. in muscle)
Peripheral	
Intercostal	Pain in dermatomal distribution
Sacral nerves	
Rectus sheath	
Extradural	Uni- or bilateral pain (lumbosacral, cervical, thoracic etc.)
Midline perineal pain	
Intrathecal	Unilateral pain (neurolytic injection for pain due to malignancy, limbs, chest etc.)
(Midline perineal pain)	
Autonomic	
Intravenous regional sympathectomy	Reflex sympathetic dystrophy
Stellate ganglion	Reflex sympathetic dystrophy Arm pain Brachial plexus nerve compression
Lumbar sympathetic	Reflex sympathetic dystrophy Lumbosacral plexus nerve compression Vascular insufficiency lower limb Perineal pain
Coeliac plexus	Abdominal pain

Table 3 Common nerve blocks

Fibromyalgia

Similar injections are done for the trigger points of fibromyalgia, but there do not appear to be any controlled comparisons of injections with other treatments.

Intravenous regional sympathectomy

Intravenous regional sympathetic blocks are used widely in patients with reflex sympathetic dystrophy. A systematic review of seven randomized controlled trials of these blocks found that none of the four guanethidine trials showed significant analgesic effect. Two reports, one using ketanserin and one bretylium with 17 patients in total, showed some advantage of intravenous regional sympathetic blocks over control. Adding guanethidine in these blocks does not appear to be more effective than local anaesthetic alone.

Epidural steroids and facet joint blocks

Two other common (for back pain) pain clinic procedures are epidural steroid injection and facet nerve blocks. Two systematic reviews have addressed the effectiveness of epidural steroid injections for sciatica and back pain. Both examined the randomized controlled trials published up to the end of 1994.

The analysis by Koes and colleagues from Amsterdam goes into great depth examining methodological quality of the trials and how this has been scored by the reviewers. This type of review is, frankly, disappointing and does not produce any meta-analytic judgements on which to work, and few enlightening ideas about the future research agenda other than some anodyne comments about possible trial design.

The best that this review can do is to tell us that, of the four studies with the highest methodological quality assessed by their particular scoring system, two had positive outcomes for epidural steroids (judged by their authors) and two had negative outcomes.

In contrast, a meta-analysis of the same trials by Watts and Silagy is an important step forward in showing that epidural corticosteroids have an analgesic effect on sciatica compared with control. Their analysis, using odds ratios, answered the question ‘Do epidural steroids work?'.

To address the question ‘How well do they work?', to try to assess the extent of the benefit given by the steroids, we reanalysed their data, and, adding a new trial (Carette *et al.*), calculated the NNT (Table 4), for at least 75 per cent pain relief for short-term outcomes (1 to 60 days) and at least 50 per cent pain relief for long-term outcomes (12 weeks to 1 year).

Trial	Improved on epidural/steroid	Improved on control	Relative benefit (95% CI)	NNT (95% CI)
Short term relief (1–60 days)				
Bellissa (97)	18/26	10/24	1.5 (0.8–2.8)	12 (3–∞)
Berrizbeitia et al. (1994)	9/41	5/19	2.1 (0.9–5.1)	3.3 (1.5–∞)
Buirski et al. (1991)	8/87	2/31	1.76 (0.98–13.7)	7.1 (1.3–7.5)
Carette et al. (1997)	23/76	23/78	1.1 (0.3–2.6)	29 (3.5–∞)
Cuellar et al. (1985)	13/19	8/16	1.1 (0.5–2.6)	10 (4–∞)
Dilke et al. (1973)	21/35	11/36	2.0 (1.1–3.6)	3.3 (1.9–14)
Klenerman et al. (1994)	15/19	12/44	1.1 (0.8–1.5)	17 (3–∞)
Mathews et al. (1987)	14/21	10/32	1.2 (0.8–1.8)	10 (3–∞)
Papadaki et al. (1992)	18/28	8/30	2.4 (1.3–4.6)	2.4 (1.6–7.2)
Ridley et al. (1988)	17/19	3/16	4.8 (1.7–13)	1.4 (1.1–2.1)
Snaith et al. (1977)	8/27	5/28	1.4 (0.5–3.8)	11 (3–∞)
Combined short term (CI 95%)	169/349	110/345	1.5 (1.1–1.9)	7.3 (4.7–16)
Long term relief (12–52 weeks)				
Buirski et al. (1991)	10/41	7/31	1.1 (0.5–2.2)	5.0 (1.8–∞)
Carette et al. (1997)	41/76	43/77	1.0 (0.7–1.3)	∞
Cuellar et al. (1985)	11/46	4/27	1.6 (0.5–4.6)	10 (9.1–∞)
Dilke et al. (1973)	16/45	8/38	1.8 (0.9–3.7)	6.3 (2.8–∞)
Mathews et al. (1987)	9/23	3/24	0.95 (0.5–1.8)	∞
Snaith et al. (1988)	76/117	98/108	1.4 (1.1–1.7)	5.6 (3.5–14)
Combined long term	163/348	174/395	1.3 (1.1–1.5)	13 (6.6–314)

Relative benefit and NNT were calculated for this table.

Table 4 Epidural corticosteroids for sciatica

Short-term relief There were 11 trials that gave short-term relief data (more than 75 per cent pain relief), with 319 patients given epidural steroids, and 345 given placebo. Only three of these studies were themselves statistically significant, but overall there was a statistically significant benefit (1.5, with 95 per cent confidence intervals 1.2 to 1.9). The NNT for short-term (1 to 60 days) greater than 75 per cent pain relief from the 10 trials with short-term outcomes combined was just under 7.3, with 95 per cent confidence intervals from 4.7 to 16. This means that for seven patients treated with epidural steroid, one will obtain more than 75 per cent pain relief short-term who would not have done had they received the control treatment (placebo or local anaesthetic).

Long-term relief There were six trials that gave long-term relief data with 315 patients given epidural steroids, and 395 given placebo. Only one of these studies was itself statistically significant, but overall there was a statistically significant benefit (1.3, with 95 per cent confidence intervals 1.1 to 1.5).

The NNT for long-term (12 weeks up to 1 year) improvement from the six trials combined was about 13 for 50 per cent pain relief, with 95 per cent confidence intervals from 6.6 to 314. This means that for 13 patients treated with epidural steroid, one will obtain more pain relief over this longer term period who would not have done had they received the control treatment (placebo or local anaesthetic).

Conclusion These NNT values at first sight appear disappointing. Here is an intervention that shows statistically significant improvement compared with control, and yet the clinical benefit, the number needed to treat for one patient to reach the chosen end-point, is 7 for short-term benefit and 13 for long-term. The short-term end-point, however, is quite a high hurdle. Using an easier hurdle of 50 per cent relief rather than 75 per cent, the ‘best' NNT achieved by drug treatment of neuropathic pain was just under 3. Patients may choose the epidural if it means they do not have to take medication, particularly if it gives a higher level of relief, even though there is a 1 in 7 chance of this level of response.

The long-term NNT of 13 is perhaps not surprising. Occasional patients in most clinics report a ‘cure' as a result of a steroid epidural, but the majority of epidural steroid successes return for repeat epidurals. That one patient has relief lasting between 12 weeks and a year for 12 treated with epidural steroid fits with experience.

For patients with chronic disease, and in this case chronic painful disease, interventions may be attractive even if their success rate is far lower than would be acceptable in, say, the management of postoperative pain. This means that the interpretation of measures of clinical benefit, such as NNTs, has to be context dependent.

There is considerable current controversy about the potential for epidural steroid to produce long-term neurological sequelae. Intrathecal injection of steroid can produce neurological sequelae. It is therefore important that intrathecal injection is avoided.

Classically, facet joint injection with local anaesthetic and steroid is indicated when pain is worse when sitting, and pain is provoked by lateral rotation and spine extension. Recent studies suggest that whether or not the injection is actually in the facet joint makes little difference, and indeed cast some doubt on long-term utility. Short-lived success (less than 6 weeks) with local anaesthetic and steroid is said to be improved by use of cryoanalgesia or radiofrequency blocks to the nerves to the joints.

Irreversible

The destructive procedures are aimed at cutting, burning, or damaging the nerve fibres carrying the pain signals. The flaw in the logic is that the nervous system can all too often rewire, finding a way around the lesion. If that happens, and the pain returns, then it may be even more difficult to manage—severe neuropathic pain can result. In general, neurolytic blocks in non-malignant pain are not recommended, because they do not last forever and recurrent pain may be more difficult to manage, and because of the morbidity. In cancer pain these neurolytic block procedures do have a place, when there is a prognosis of less than 3 months, or where alternatives such as painstaking drug control or long-term epidural infusion are not possible. Similar distinction between cancer and non-cancer pain holds for coeliac plexus block in pancreatic pain. Pain associated with pancreatic cancer responds well to coeliac plexus block, and it may also help those with abdominal or perineal pain from tumour in the pelvis. In chronic pancreatitis results are much less convincing.

The limitation is the potential for motor and sphincter damage. This risk is greater with bilateral and repeat procedures, and greater the lower the cord level of the block. Extradural neurolytics have limited efficacy. While claims have been made that the paravertebral approach is preferable, patchy results may be attributed to unpredictable injectate spread. Results with epidural infusion of a combination of local anaesthetic and opioid are superior to neurolytic blocks, providing good analgesia with minimal irreversible morbidity.

Surgery

The relevant neurosurgical interventions for chronic pain include dorsal column stimulation, rhizotomy, cordotomy, and dorsal root entry zone lesions. The indications are usually non-malignant neuropathic pain which has failed to respond to pharmacological measures. The difficulties of trials of uncommon surgical procedures are well known. These procedures are usually documented by glowing case series. Longer term outcomes may not be so good.

Alternatives

TENS, acupuncture

The rationale for transcutaneous nerve stimulation (**TENS**) is the gate theory. If the spinal cord is bombarded with impulses from the TENS machine then it is distracted from transmitting the pathological pain signal. We know from systematic reviews that TENS has limited efficacy in acute (and postoperative) pain, and also in labour pain. Most trials in chronic pain are of very short duration (low ‘dose') TENS and the pooled results are unconvincing.

What we do know is that attention to detail makes considerable difference to TENS efficacy in chronic pain, and that the dose of TENS may need to be much greater than tested in the trials to date. Patients need to be told that it is useless expecting success unless the machine is connected for at least an hour at a time. They need to be told where to put the electrodes, how to put them on, how to manipulate the stimulus to best advantage, and indeed to turn the machine on.

At least three systematic reviews discuss acupuncture in chronic non-malignant pain. These show an effect in poor quality trials, but that effect in clinical practice is often short-lived (3 days), and is therefore expensive in time. It is difficult to know what is the real place of acupuncture, like other complementary interventions, because of the lack of trials comparing complementary with mainstream procedures.

Physiotherapy and variants

Pain clinics keep a very open mind about other interventions. If patients benefit from alternatives we are only too pleased. The evidence from back pain, however, suggests that on rigorous outcome measures physiotherapy and other forms of manipulation have but limited success. Such analyses often did not include any measure of quality of life. If they make the patient feel better and they are cheap, then it is a decision for the third party payer whether or not these physiotherapy manoeuvres should be offered.

Behavioural management

Back schools through to behavioural management programmes offer a range of help for patients, help to cope with their (usually back) pain problems. Making decisions about the benefits of psychologically based treatments of medical problems is not easy, and especially difficult to compare with other treatments and to measure relative benefit and cost. Patients whose pain has proved intractable to all reasonable medical and other interventions are chronic consumers of health care—general practitioner or hospital clinic time, analgesic and psychotropic drugs, repeated admissions, and sometimes surgery. If rehabilitation treatment enables these patients to carry on more satisfying lives with minimum medical help, how can it be most effectively and economically offered?

Further reading

Source for systematic reviews of pain treatments

McQuay HJ, Moore RA. *An evidence-based resource for pain relief*. Oxford: Oxford University Press, 1998.

NSAID adverse effects and efficacy

Eisenberg E, Berkey CS, Carr DB, Mosteller F, Chalmers TC. Efficacy and safety of nonsteroidal antiinflammatory drugs for cancer pain: a meta-analysis. *Journal of Clinical Oncology* 1994; **12**: 2756–65.

Hawkey CJ, Karrasch JA, Szczepanski L. Omeprazole compared with misoprostol for ulcers associated with nonsteroidal antiinflammatory drugs. *New England Journal of Medicine*, 1998; **338**:727–34.

Henry D, *et al.* Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis. *British Medical Journal* 1996; **312**:1563–6.

Yeomans ND, Tulassay Z, Juhasz Lea. A comparison of omeprazole with ranitidine for ulcers associated with nonsteroidal antiinflammatory drugs. *New England Journal of Medicine* 1998; **338**:719–26.

Shield MJ, Morant SV. Misoprostol in patients taking non-steroidal anti-inflammatory drugs. *British Medical Journal* 1996; **312**:846.

Silverstein FE, *et al.* Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. *Annals of Internal Medicine* 1995; **123**:241–9.

Clonidine

Eisenach JC, DuPen S, Dubois M, Miguel R, Allin D. Epidural clonidine analgesia for intractable cancer pain. The Epidural Clonidine Study Group. *Pain* 1995; **61**:391–9.

Nerve blocks

Arner S, Lindblom U, Meyerson BA, Molander C. Prolonged relief of neuralgia after regional anesthetic blocks. A call for further experimental and systematic clinical studies. *Pain* 1990; **43**:287–97.

Epidurals for sciatica

Carette S, Leclaire R, Marcoux S *et al.* Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. *New England Journal of Medicine* 1997; **336**:1634–40.

Koes BW, Scholten RPM, Mens JMA, Bouter LM. Efficacy of epidural steroid injections for low-back pain and sciatica: a systematic review of randomized clinical trials. *Pain* 1995; **63**:279–88.

Watts RW, Silagy CA. A meta-analysis on the efficacy of epidural corticosteroids in the treatment of sciatica. *Anaesthesia and Intensive Care* 1995; **23**:564–9.

Facet joint injections

Carette S, *et al.* A controlled trial of corticosteroid injections into facet joints for chronic low back pain. *New England Journal of Medicine* 1991; **325**:1002–7.

Lilius G, Laasonen EM, Myllynen P, Harilainen A, Grönlund G. Lumbar facet joint syndrome. *Journal of Bone and Joint Surgery* 1989; **71**:681–4.

Coeliac plexus

Eisenberg E, Carr DB, Chalmers TC. Neurolytic celiac plexus block for treatment of cancer pain: a meta-analysis. *Anesthesia and Analgesia* 1995; **80**: 290–5.

TENS

Reeve J, Menon D, Corabian P. Transcutaneous electrical nerve stimulation (TENS): a technology assessment. *International Journal of Technology Assessment in Health Care* 1996; **12**: 299–324.

50.1 Principles and practice of plastic surgery

Michael J. Yaremchuk, G. Gregory Gallico III

General principles
Wound healing
Facial injuries—soft tissue
Diagnosis
Wound preparation
Suture techniques
Scar revision
Windshield (windscreen) injuries
Traumatic tattooing
Bites
Wounds with tissue loss
Skin lesions
Benign skin lesions
Basal-cell carcinoma
Squamous-cell carcinoma
Melanoma
Soft-tissue coverage
Skin grafts
Composite grafts
Skin-graft substitutes
Flaps
Facial reconstruction by anatomic site
Eyelid
Nose
Ear
Breast surgery
Augmentation
Reduction
Reconstruction
Subcutaneous mastectomy
Gynecomastia
Abdominal wall reconstruction
Rectus abdominis muscle flap
Tensor fascia lata flap
Gracilis muscle flap
Microsurgery
History
Microvascular technique
Replantation
Indications
Preservation of the ischemic part
Sequence of surgical repair
Replantation by anatomic site
Viability and functional results
Free-tissue transfer
History
Flaps
Technique
Lower extremity reconstruction
Acute coverage of leg wounds
Osteomyelitis
Head-and-neck reconstruction
Scalp and forehead
Floor of mouth
Mandible
Maxilla
Hemifacial soft tissue
Cervical esophagus
Facial nerve grafting
Fractures of the facial skeleton
Facial trauma
Treatment by anatomic area
Cosmetic surgery
Relation of reconstructive to cosmetic surgery
Rhinoplasty
Face lift
Blepharoplasty and brow lift
Minimally invasive plastic surgery
Alloplastic facial implants
Chemical peel and dermabrasion
Laser
Body contouring and suction lipectomy
Further reading

Plastic surgery is a discipline that mostly concerns surgical problems of the integument, particularly that of the face. The plastic surgical correction of deformities of congenital, post-traumatic, or postablative origin is usually intended not only to optimize function but also to restore form. Most procedures involve the rearrangement of local tissues or the transfer of more distant ones, such as grafts and flaps. The plastic surgeon seeks to optimize conditions for the wound-healing process by attention to technique and detail.

General principles

Wound healing (see also Section 6)

Wound healing is a non-specific response to injury. It involves the biologic processes of inflammation, collagen metabolism, and contraction in an overlapping, integrated continuum. Wound healing is divided temporally into three phases, an inflammatory, a fibroblastic, and a remodelling phase. The nature and condition of the tissues as well as the mechanism of wound closure determine the relative duration of these phases and the end-result of the healing process. Against this background, wound healing is classically divided into three types: primary intention, delayed primary intention, and secondary intention (Fig. 1).

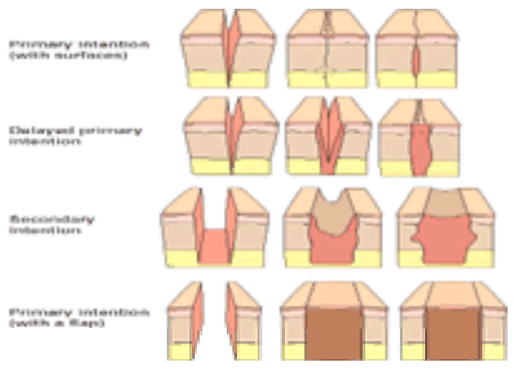


Fig. 1. Diagrammatic representation of wound closure by primary, delayed primary, and secondary intention.

Primary intention

The coaptation of sharply incised wound edges within hours of injury is referred to as primary repair and usually results in healing by primary intention. Because little or no tissue is lost and the injury is confined to the wound edges, the healing phases are relatively well defined. The inflammatory phase, which begins at the time of injury, largely subsides by 5 days. The fibroblastic phase, in which collagen synthesis by fibroblasts is the main process, overlaps the initial phase by 2 to 3 days and dominates from days 5 to 14. Wound strength increases rapidly with collagen content. The final phase of maturation is a protracted one; during this phase, which begins at about day 14 and lasts for up to 6 months, collagen formation is balanced by collagen degradation. The end-result of the healing process is the bridging of the soft-tissue gap with collagen and its resurfacing with epithelium.

Delayed primary intention

In delayed primary intention, the healing edges of the wound are brought together several days after wounding. This delay allows inflammatory and immune processes to control wound contamination. Healing by delayed primary intention is appropriate when initial wound cleansing and debridement is unlikely to be sufficient to avoid infection if the wound were closed primarily. Delayed primary closure is effected by placing sutures between the wound edges at the initial debridement, but leaving them untied until day 5. It does not delay the development of wound strength, but more scar tissue develops between the native wound edges, and the appearance and function of soft tissues may be compromised.

Secondary intention

Wounds heavily contaminated by bacteria or with significant necrotic tissue may be left unsutured and allowed to heal by secondary intention. They gradually become filled with capillaries and fibrous tissue (‘granulation tissue'), and the inflammatory and fibroblastic periods of wound healing are markedly prolonged. These wounds eventually are closed, to a varying extent, by the processes of contraction and epithelialization. Wound contraction is believed to be controlled by the myofibroblast, a fibroblast with smooth-muscle components. Although healing by secondary intention may be the optimal means to achieve wound closure, in certain instances the end-result is an increased amount of scar and a thin epithelial covering.

When tissue loss prevents direct closure and healing by secondary intention is impossible or would result in significant compromise in function or appearance, skin grafts or flaps are required to close the wound.

Facial injuries—soft tissue

Diagnosis

Injuries of the facial soft tissues are usually self-evident. Those resulting from blunt trauma should always first be considered a sign of underlying significant injury to the facial skeleton or brain. Appropriate physical examination and diagnostic imaging should be undertaken.

The repair of soft-tissue injuries to the face should restore both function and appearance. To this end, it is especially important to minimize the amount of scar tissue, to restore anatomic landmarks, and to avoid contour irregularities. The wound is ideally closed as soon as possible after injury, thereby minimizing distortion secondary to edema, and decreasing the effects of initial contamination and the opportunity for secondary contamination. Lacerations accompanying facial fractures should be closed after fracture fixation, if performed acutely. However, if fracture therapy is to be delayed, they should be closed acutely to avoid contamination of the wound and possibly the fracture site. Clean soft-tissue injuries that have been cleansed well initially can be closed up to 24 h after injury. This interval can be extended if the wound margins are excised prior to closure.

Wound preparation

Wound cleansing and irrigation is intended to remove foreign bodies and decrease the numbers of bacteria. Normal saline solution, which is readily available and causes no tissue damage, is best used for irrigation. Because small particulate matter can only be removed by mechanical disruption and not by volume dilution, irrigation is best done with a fluid jet delivering 7 lb of pressure per square inch. This goal can be accomplished by injecting sterile saline from a handheld 3- to 5-ml SYringe through an 18-G needle and for larger wounds by using a jet lavage system. A bulb syringe does not generate sufficient pressure to debride the wound.

Debridement is more conservative in the face than in other areas of the body: severely damaged skin often survives because of the rich facial blood supply, and anatomic landmarks are virtually impossible to reconstruct secondarily. When tissue is in relative excess, such as in the cheek or forehead, ragged lacerations may be excised and a ‘new surgical wound' created.

Antibiotics are appropriate when a surgically clean wound cannot be obtained, when lacerations communicate with the mouth or paranasal sinuses, and when systemic or local factors known to diminish host resistance are present. The need for tetanus immunization depends on the nature of the wound and the patient's previous immunization history.

Suture techniques

For orientation, ‘cardinal' sutures are placed to realign anatomic landmarks, such as lid margins, eyebrows, vermilion borders, or significant wrinkles (Fig. 2). To avoid contour deformities, wounds are closed in anatomic layers and dead spaces are obliterated. To this end, muscle, subcutaneous tissue, and the deep dermis are approximated with 4-0 or 5-0 polyglycolic acid sutures. Skin edges are then precisely aligned with slight eversion, using interrupted sutures of 6-0 nylon or prolene (Fig. 3). Loupe magnification is often helpful.



Fig. 2. Facial features must be carefully realigned prior to suturing lacerations. ‘Cardinal' sutures are placed at definable anatomic landmarks for orientation. For this

reason, the eyebrow (which will grow back) should never be shaved as part of the wound preparation.

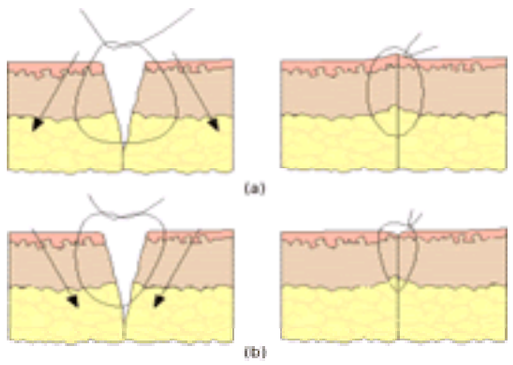


Fig. 3. Suture technique showing inverted dermal suture and everting skin suture: (a) eversion; (b) inversion.

Facial sutures are removed between 3 and 5 days after surgery to avoid leaving suture marks. The wound may be reinforced at this time with skin tapes. The creation of suture marks is largely determined by the tightness with which sutures are tied and by the length of time they are left in.

Scar revision

Because the wound-healing process is long, scar revision is seldom performed sooner than 6 months to 1 year after injury. Hypertrophic scars and keloids result from imbalances during the proliferative and the maturation phases of wound healing. Hypertrophic scars are raised and thickened, but remain within the borders of the original scar; they undergo some maturation and improvement with time. Hypertrophic scars often respond to pressure, which causes realignment and remodeling of the collagen bundles within the scar. If they are the result of a wound that healed by secondary intention or of a laceration across a flexion crease, they may improve by excision and appropriate reclosure.

Keloids grow beyond the limits of the original scar. Direct intralesional injection of triamcinolone acetonide (10 mg/ml) may decrease their bulk. Injections can be given every 3 to 4 weeks for up to 6 months.

Lacerations in the cheek should always alert the surgeon to the possibility of injury to the parotid gland, Stensen's duct, or branches of the facial nerve. Prior to administering an anesthetic, the function of the facial nerve should always be tested. Lacerations of the parotid gland need not be repaired, but are managed by layered wound closure and drainage; persistent parotid fistula is exceedingly uncommon. Management of duct lacerations is determined by their location. Lacerations within the substance of the gland are best managed by ligation, allowing reflex atrophy. Lacerations to the portion overlying the masseter muscle may be repaired over a polyethylene splint. Those distal to the masseter are managed by ligating the distal cut and then directing the proximal end into the oral cavity (Fig. 4).

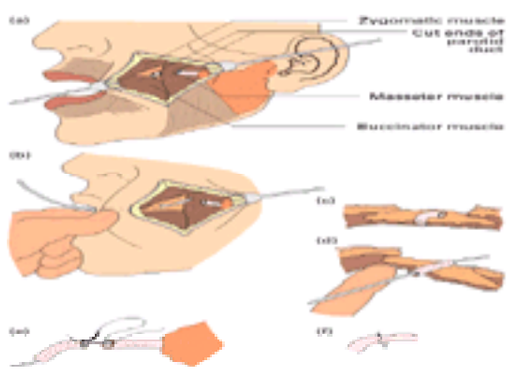


Fig. 4. Lacerations of the cheek may require exploration and anastomosis of the parotid duct.

Buccal branches of the facial nerve run parallel to Stensen's duct and are usually involved when the duct has been lacerated. As with injuries to other branches of the facial nerve, the nerve should be repaired at the time of presentation using microsurgical techniques. If emergency repair cannot be performed, the nerve ends should be tagged at presentation, since the distal divided ends lose their ability to respond after 48 h, making later localization difficult.

Windshield (windscreen) injuries

The plastic used to laminate windshields distributes impact forces and retains the resulting broken pieces. Windshield impact causes a mosaic of tiny lacerations and avulsion flaps that mimics the fracture pattern of the windshield itself. Effective management requires removal of all pieces and meticulous approximation of all lacerations. If allowed to heal without repair, a cobblestone appearance that is impossible to correct later will result.

Traumatic tattooing

Foreign bodies embedded in soft-tissue wounds must be removed at the time of presentation. In most cases this can be accomplished by vigorous scrubbing with a surgical scrub brush, but dermabrasion or excision may be required. Delayed removal of traumatic tattooing is difficult, if not impossible.

Bites

Bite injuries to the face can be disfiguring because of the crush avulsion involved. They are treated as dirty facial lacerations, with vigorous irrigation and debridement. If possible, the wound is recreated surgically by removing 1 to 2 mm of its margins both at the skin surface and in its depths. If this would result in significant disfigurement, only clearly devitalized tissues are removed. These wounds are then closed in a standard fashion. Amputated parts are cleansed, debrided, and replaced as composite grafts. If the amputated part has vessels of a sufficient caliber, microvascular replantation should be considered.

Prophylactic antibiotics are administered. Because *Pasteurella* is the most frequent pathogen in dog and cat bites, penicillin or its equivalent should be given to those who are not allergic. Both streptococcal and staphylococcal organisms are prevalent in human bites: cephalosporin, erythromycin, or penicillin and a penicillinase-resistant penicillin may be administered. Guidelines for tetanus and rabies prophylaxis should be consulted.

Wounds with tissue loss

Facial lacerations seldom cause tissue loss. The placement of ‘cardinal' sutures to restore landmarks will usually show that little, if any, tissue is missing. If tissue cannot be reapproximated at the time of immediate repair, flap reconstruction is rarely indicated. If the defect is significant, an emergency skin graft should be considered to allow definitive reconstruction later under ideal circumstances.

Skin lesions

Benign skin lesions

Patients frequently request the excision of benign lesions; occasionally, these masquerade as malignancies and excision is required to clarify the diagnosis. If there is any question about the diagnosis, biopsy examination is necessary. These lesions are usually excised as an elliptical or lenticular shape, which allows a straight-line closure. The patient must realize that removal of the benign lesion will cause a scar. The orientation of the ellipse is dictated by wrinkle lines, lines of least tension on the face, contour lines, and the outlines of facial features (Fig. 5). If a more elaborate reconstructive technique than primary closure is required, it may be best to leave the benign lesion alone.



Fig. 5. The extant lines on the face formed by the wrinkles, tension lines, facial contours, and facial features direct the orientation of lenticular excisions.

Basal-cell carcinoma

Basal-cell carcinoma is a skin malignancy derived from the basal epithelial cells. These slow-growing lesions rarely metastasize. They grow as flat, diffuse plaques, morpheaform cancers, or raised, well-circumscribed nodules. Morpheaform lesions are best excised with intraoperative control of margins. The nodular lesions have a clear border that allows them to be excised with a reasonable expectation of removal by gross inspection. Subsequent pathological inspection of the margins of all lesions is necessary to be sure that removal is complete. Tumor-positive margins after excision do not invariably mean the lesion will recur, but repeat excision is usually recommended.

Squamous-cell carcinoma

Squamous-cell skin malignancies, originating from the differentiated epithelial cells, are more serious than basal-cell carcinoma since metastasis is possible. Like basal-cell carcinomas, squamous-cell cancers are caused in part by sun exposure and, therefore, will be found on chronically exposed skin surfaces. The face, especially the lower lip, and the back of the neck and hands, are often involved. Because the histologic margins may be difficult to determine by gross inspection and because invasion into the subcutaneous tissue is possible, the resection should have a wider margin than is required for basal-cell cancers. Intraoperative examination of the resection margins by frozen section is advised.

Because of the increased risk of metastasis, the draining lymph nodes should be palpated for enlargement. Clinically enlarged nodes suggest the need for a standard lymph-node dissection in addition to the primary resection.

Melanoma (see also Section 38.5)

The most serious of the skin cancers, melanoma, requires rapid resection and lymph-node dissection depending on the depth of invasion of the primary lesion. Melanoma is derived from the melanocytes interspersed among the basal-cell layer of the epithelium. The lesions occur in a superficially spreading or a nodular invasive pattern; the superficial lesion may be the predecessor of the nodular.

Melanomas, like other skin cancers, occur on sun-exposed surfaces. They may also occur in mucosa, the retina, the palmar and plantar skin, and subungually. The prognosis for metastasis is dependent on the depth of the primary invasion. There have been two classifications proposed, Clark's and Breslow's. Clark's (Fig. 6) divides the primary lesion into five levels: level I is confined to the epithelium; level II passes through the basement membrane; level III passes to the level of the junction of papillary and reticular dermis; level IV passes into the reticular dermis; and level V invades the subcutaneous tissue. Breslow's system measures the depth of penetration of the melanoma from the epithelial surface. Metastasis is rare for lesions less than 1 mm in depth; from 1.5 or 1.67 to 3.6 mm, there is a significant and increasing risk of local nodal metastasis, and therapeutic or prophylactic lymph-node dissection is recommended; deeper lesions have a high incidence of metastasis, and local nodal dissections are often futile.

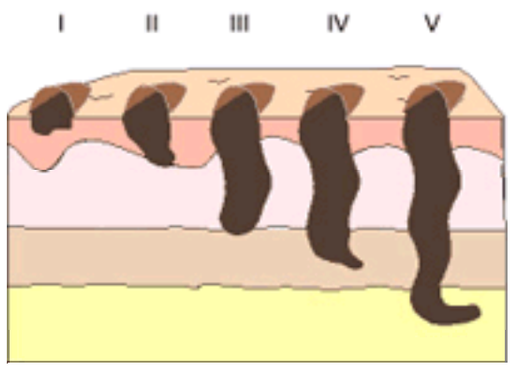


Fig. 6. The depth of melanoma within the skin as defined by the systems of Clark or Breslow is statistically prognostic for cure.

Melanoma is excised with a wider margin than squamous-cell carcinoma. A margin of 1.5 cm is recommended on the face near vital features; a margin of 3 cm is recommended elsewhere. Although many adjunctive treatments have been tried, including irradiation, chemotherapy, and immunotherapy, none has been successful. Early detection and complete initial surgical resection are the keys to caring for patients with melanoma.

Soft-tissue coverage (see Fig. 7)

Skin grafts

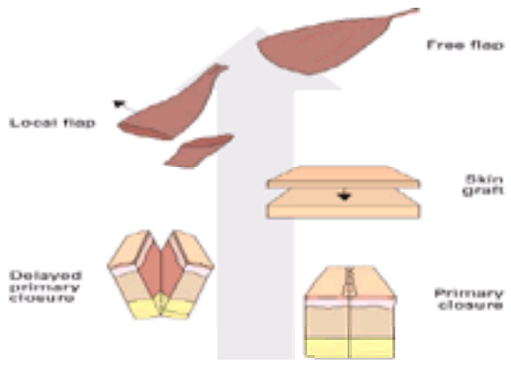


Fig. 7. Reconstructive alternatives for wound closure; the method will be determined by complexity of wound.

Split-thickness skin grafts are thin sections of skin, consisting of varying amounts of dermis and its overlying epidermis, which are totally detached from the body and transplanted to a recipient site (Fig. 8). The skin graft is, therefore, totally dependent on the vascularity of the recipient site for its survival. Any tissue bed with an exposed capillary circulation will potentially support a skin graft. The graft must be in close contact with the recipient bed so that a plasmatic circulation (serum inbibition) can support it until there is revascularization (inosculation). These processes are called the ‘taking' of the graft. The immobilization and intimate contact of the graft with the bed are accomplished by the use of dressings. In the extremities, a circumferential dressing is used; in the face and other areas, tie-over stent dressings are usually employed (Fig. 9). The ‘take' of the graft is also determined by the type and amount of bacterial contamination. Pseudomonads and streptococci are particularly detrimental to ‘take'. Tissues containing fewer than 10^5 organisms per gram by quantitative culture usually allow a graft to ‘take'.

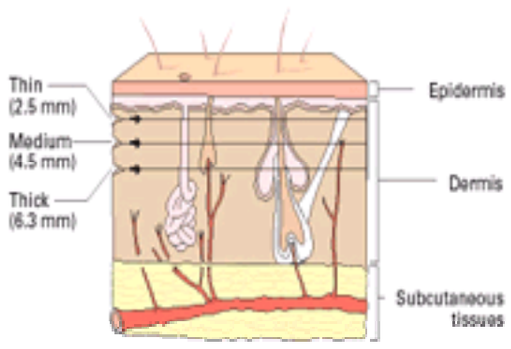


Fig. 8. Cross-section of skin showing various thicknesses of grafts.

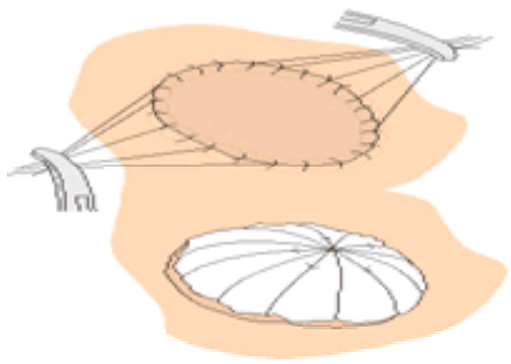


Fig. 9. Stent used to immobilize a full-thickness graft.

Thin grafts are more likely to take than thick grafts because there is less tissue to be supported by imbibition and a larger number of blood-vessel openings on their undersurface, a reflection of the increasingly branching vascular pattern from the deep dermis to the surface. Split-thickness grafts can be harvested with a drum dermatome, a power dermatome, or a freehand knife. Split-thickness grafts are often cut to form a mesh to allow better conformation to the wound, to cover a larger surface area, and to allow drainage of any exudate. The open areas of meshed grafts are allowed to close by re-epithelialization and contraction (Fig. 10).



Fig. 10. Mesh graft applied to a clean, granulating wound.

Full-thickness skin grafts are harvested to contain the entire thickness of the papillary and reticular dermis. Full-thickness grafts, with their greater metabolic demands, require optimal conditions for ‘take'. They are usually used to resurface surgically created wounds, and they provide reconstruction of the best quality, contour, color match, durability, and with lack of wound contraction. Unlike the donor sites for split-thickness grafts, which heal by re-epithelializ-ation, those for full-thickness grafts require surgical closure. No special instruments are required to harvest full-thickness grafts. Useful donor sites include the postauricular area, the upper eyelid, the groin, and the flank.

Composite grafts

In addition to skin and subcutaneous tissue, composite grafts include another tissue element, most commonly cartilage or hair follicles. Because of the volume of tissue transferred, composite grafts need a well-vascularized bed with optimal conditions for ‘take'. A common composite graft is the free edge of the ear used to reconstruct a full-thickness defect of the alar border. Since the graft must survive to its core by imbibition and diffusion until inosculation is completed, its dimensions are limited. The graft should be designed so that no part is greater than 5 mm from any part of the nutrient bed (Fig. 11).

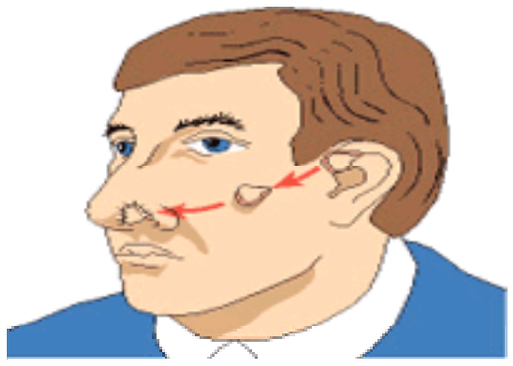


Fig. 11. Composite graft harvested from an auricle and transplanted to a defect of the nasal border.

Composite grafts with hair follicles are frequently used to replace the hair lost with male-pattern baldness. ‘Micrografts' consisting of one or two hair follicles may be used to reconstruct the eyebrow (Fig. 12).

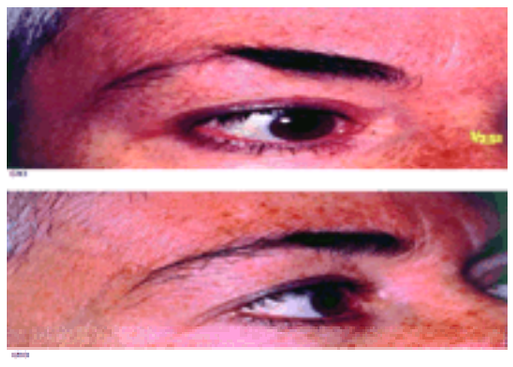


Fig. 12. Eyebrow reconstructed with ‘micrografts' consisting of one or two hair follicles: (a) preoperative; (b) postoperative.

Skin-graft substitutes

Occasionally, there is an insufficient supply of skin-graft donor sites, such as in a patient with a large burn, and a substitute is required for temporary or permanent wound coverage. An open wound is a portal for bacterial invasion and for loss of water vapor and protein: any substitute must fulfil the role of a barrier. Some substitutes have been designed to attempt not only to create a barrier but also to produce a ‘skin' with the supple biomechanic properties of the normal integument.

Cadaver skin, which can be harvested as a split-thickness graft, is the best temporary substitute. The graft will adhere to the wound and become vascularized, but the barrier provided is only temporary: it will be rejected in the first 2 to 3 weeks after placement, depending on the immunologic capability of the host. Cadaver allograft skin can be stored frozen to be available for wound coverage when needed. Frozen allografts are particularly useful for the immediate closure of large, excised burn wounds, but have also been used as biologic dressings on venous stasis ulcers.

The barrier function of skin can also be temporarily simulated by an alloplastic sheet. Silicone sheets with or without small pores may be placed on the wound and encouraged to adhere to it by materials attached to their undersurfaces. A model for this type of synthetic skin is Biobrane, a porous Silastic sheet with an under layer of fine nylon coils. Silastic sheets, being foreign bodies, are prone to infection; they can only be left in place for 3 to 6 weeks, after which they must be replaced with a permanent skin.

Various types of collagen sponge have also been used as temporary skin substitutes. Skin is a bilaminar structure of epidermis and dermis. Its barrier function resides primarily in the epidermis; the dermis contributes to elasticity and durability. In an effort to produce a dermal substitute that might approximate the biomechanical behavior of a full-thickness skin graft, collagen sponges covalently linked to glycosaminoglycans have been developed. The outer surface of these dermal substitutes is covered temporarily with Silastic. After the collagen sponge has been incorporated into the wound (3-6 weeks), the outer Silastic layer is removed and is replaced with a thin, split-thickness skin graft. Whether these materials will lead to development of a permanent dermal substitute is unknown; their use is still in an experimental phase.

Human epithelial cells can be grown in tissue culture to produce coherent sheets that can be transferred as a pure epithelial autograft or allograft (Fig. 13). From a small biopsy specimen of 4 cm^2, up to 2 m^2 can be grown in 3 weeks. The rapid growth of epithelial cells in tissue culture allows such grafts to be used for the closure of large burns. The grafts can be placed on the excised burn wound and will adhere and produce a permanent skin substitute. Long-term histologic examination of cultured epithelial grafts as a permanent skin shows the differentiation of a multilaminar epithelium with the development of a basement membrane, anchoring fibrils, and rete ridges and pegs. The differentiation of subjacent collagen in response to the epithelial grafts may indicate that a dermis does not need to be provided by a collagen sponge.

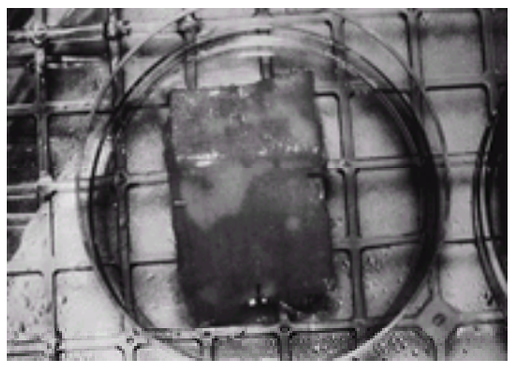


Fig. 13. Human epithelial cells grown in tissue culture can be transferred as a skin-graft substitute.

Flaps

Flaps are transferable blocks of tissue that maintain their viability from their own circulation. They are, therefore, used when the recipient bed does not have sufficient capillary circulation or when the bulk, color, or contour of the specific flap tissue are desired. Flap closure of a surgically created wound allows healing by primary intention, with all of its advantages.

Flap design is based on a knowledge of the vascular anatomy of the skin and on geometric principles. The skin receives its blood supply from the dermal-subdermal plexus. In most areas of the body, these vessels are supplied by vertically oriented, perforating branches from large vessels running in the underlying fascia or muscle. The muscle or fascia, in turn, are supplied by one or more vascular pedicles (Fig. 14). In a few areas of the body, the dermal-subdermal plexus is supplied by vessels

that course directly in the subcutaneous tissues.

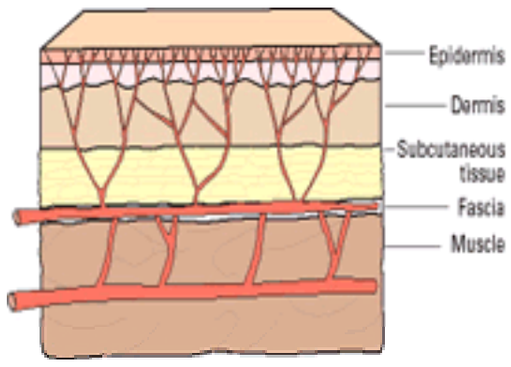


Fig. 14. Cross-section of the soft tissues to demonstrate the blood supply to the skin. The skin is supplied by a dermal–subdermal plexus fed by vertically oriented perforators arising from axially oriented vessels running in the fascia, muscle, and in some areas (not shown) the subcutaneous tissues.

Flaps can be classified on the basis of their blood supply, the type of tissue transferred, their proximity to the wound, and their geometric design.

Patterns of flap blood supply

Random flaps

Flaps designed without regard to the underlying vascular anatomy are termed random. They have a blood supply based on the subdermal plexus. Except for the richly vascularized areas of the face and scalp, their length:width ratio should not exceed 1:1. Flaps of greater length:width ratio are elevated in stages to alter their pattern of nutritional blood flow, a process termed ‘delaying the flap'.

Axial flaps

Axial flaps are supplied by large vascular pedicles that run in the subcutaneous tissue. As long as the pedicle is included in the flap, they can be designed without regard to the length:width ratio.

The groin flap is an axial flap that is particularly useful for hand reconstruction. It is supplied by the superficial circumflex iliac artery, a branch of the common femoral or profunda femoris arteries (Fig. 15). The superficial circumflex iliac runs laterally, parallel to the inguinal ligament, and will support the skin overlying this area to the midlateral line. The deltopectoral flap, based on branches of the internal mammary artery, and the forehead flap, based on the supratrochlear vessels, are other useful axial cutaneous flaps.

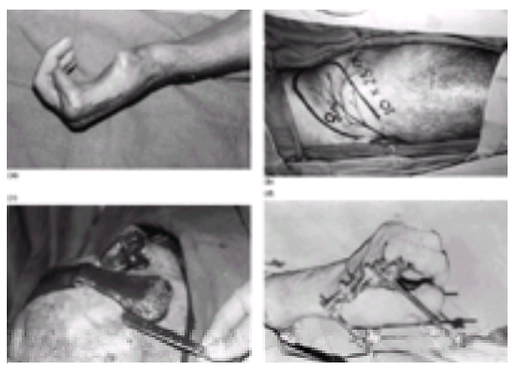


Fig. 15. Groin flap: (a) traumatic loss of the thumb; (b) groin flap based on an axial artery; (c) flap on the hand covering a bone graft; (d) reconstructed thumb.

Muscle and myocutaneous flaps

Most cutaneous areas receive a blood supply from the underlying muscle. Knowledge of the vascular anatomy of muscles allows one to transfer flaps of the muscles alone covered by a skin graft or of the overlying skin and subcutaneous tissue as well. Because the pattern of blood supply to muscles determines their potential usefulness as a flap, the vascular anatomy of muscles has been classified in terms of the number of pedicles, their size, and their point of entry (Fig. 16). This classification allows the surgeon to predict the muscle's reliability, arc of rotation, or potential for free-tissue transfer. Their reliable blood supply and their bulk, which allows them to obliterate dead space, account for the prominent role of muscle and myocutaneous flaps in reconstruction.

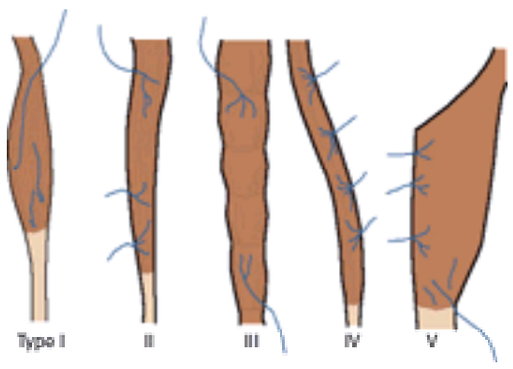


Fig. 16. Mathes and Nahai classification of muscle flaps: type I, single dominant vascular pedicle; type II, dominant proximal and smaller segmental; type III, two large dominant pedicles; type IV, multiple vascular pedicles; type V, dominant pedicle at insertion and multiple segmental pedicles at origin.

Fasciocutaneous flaps

An intercommunicating system of vessels derived from perforators from the underlying muscle runs in the muscular fascia, and in certain areas of the body these fascial vessels will provide a sufficient blood supply to the overlying skin. Either the fascia alone or the fascia with its overlying skin may be transferred.

Flap proximity to wound

Local flaps

Flaps derived from tissues immediately next to the wound are termed local flaps. Obviously, a local flap is an easier tool to choose for a particular reconstruction because no new incision is made in another part of the body. However, reconstructive choices with a local flap may be limited because sufficient skin with the appropriate blood supply may not be available. Small facial defects are commonly reconstructed with local flaps; their design is discussed below.

Distant flaps

When local tissue is insufficient to cover an adjacent defect, a flap must be transferred from a remote location. Traditionally, this was done in one of three ways: the flap donor site and defect recipient site were brought near to one another, as in the case of the cross-leg flap; the recipient site was brought near to the donor site, as when a hand defect is covered by a groin flap; or the flap was brought to the recipient site attached temporarily to the blood supply provided by an intermediate carrier site, such as the wrist. All of these techniques require prolonged, sometimes awkward, positioning of the patient and several operations to be successful.

Free flaps

Any flap whose blood supply is derived from an artery at least 1 mm in diameter may be completely detached from the donor site and transferred to a distant recipient site, with microvascular anastomosis to restore the blood supply. This process of flap transfer is called free-tissue transfer; the flap is called a free flap (Fig. 17). The flap may consist of any type of tissue reliably supplied by an artery and vein. Free flaps allow the reconstruction of defects where no local tissue is available for transfer and where use of a distant flap might require more than one operation. The variety of types of tissue that can be transferred, the obscurity of the donor site, and the ingenuity of the reconstruction frequently provide a result whose excellence justifies the length and risk of the operation.

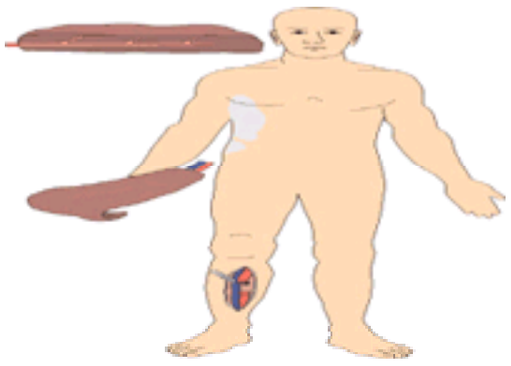


Fig. 17. Diagrammatic representation of a free-tissue transfer whereby expendable autogenous tissue reliably supplied by an artery and a vein is totally detached from the donor site and transplanted to a recipient site, with revascularization accomplished by microsurgical anastomosis of donor and recipient artery and vein.

Flap geometric design

In addition to categorization by blood supply, local flaps can be categorized by their geometric design, that is, by the manner through which areas of relative skin excess are transferred to areas of dearth. The design and movement of these flaps from the donor to the recipient site is dependent on a balance of blood supply and tissue tension. Excess tension will decrease blood supply and lead to flap necrosis.

Advancement flaps

Advancement flaps depend on the laxity of their constituent skin to provide excess tissue when separated from its underlying structures. They are most useful for small defects when skin is in excess and possesses an inherently good blood supply, such as in the aging face (Fig. 18a).

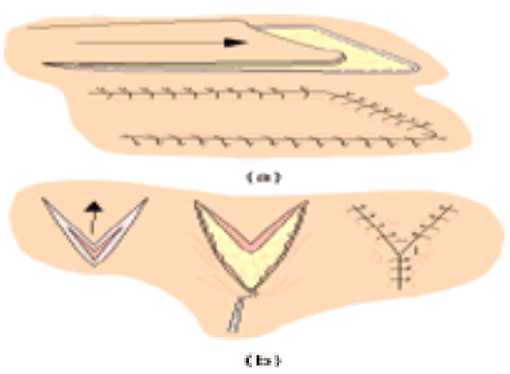


Fig. 18. (a) Advancement flap; (b) V-Y advancement flap.

V-Y flap V-Y advancement flap is a modification of the advancement. A small amount of length along the vertical axis is gained from movement of tissue from the horizontal axis. The V-shaped flap is advanced and its donor site is closed to make the vertical limb of the Y (Fig. 18b).

Rotation and transposition flaps

When tissues are rotated into defects they are called rotation flaps; when moved laterally into a defect, they are called transposition flaps. Many flaps combine, in varying degree, the principles of rotation and transposition. One of the two motions may predominate, depending on the size of the primary defect and the laxity of the adjacent tissues. The secondary defect may be closed by primary suture or with a skin graft.

In principle, rotation flaps are designed so that the defect and flap lie in a semicircle. Pure rotation flaps depend on tissue laxity and a redistribution of tension to close the secondary defect. If the primary defect is sufficiently large, a back-cut is made to allow the flap also to be transposed. The secondary defect is usually best managed with a skin graft to avoid excess tension (Fig. 19).

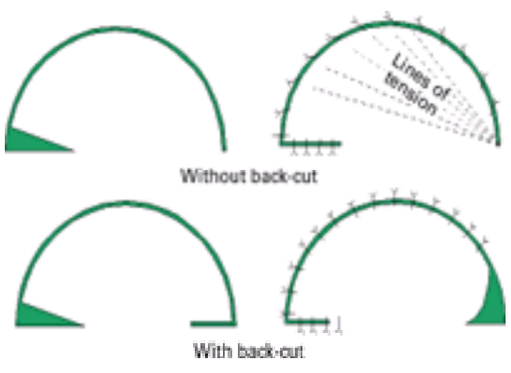


Fig. 19. Rotation flap.

Transposition flaps have predominantly lateral motion, even though they pivot on an axis, a rotational movement. They are designed as rectangles adjacent to the defect and are moved laterally to fill it. They are usually designed to allow primary closure of the secondary defect, but may require skin grafts (Fig. 20). A back-cut may be necessary to reduce tension at the base of the flap and allow its rotation into the defect., but the back-cut reduces the circulation to the flap.

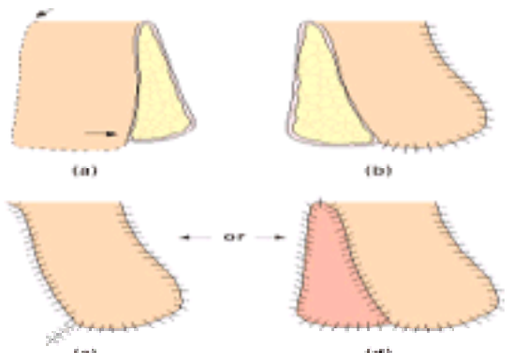


Fig. 20. Transposition flap: (a) flap outlined to close defect; (b) flap transposed into defect resulting in ‘secondary' defect; (c) secondary defect closed primarily; (d) secondary defect closed with skin graft.

Z-plasty A Z-plasty is two opposing transposition flaps that exchange places and thereby alter the direction and length of scars. The name is derived from the Z-shape of the outline of the two flaps. Z-plasties are useful because after transposition there is a gain in length along the central axis of the original Z and a change in its direction by 90°. The increase in length is useful in the treatment of contractures and the change in direction is often useful in the revision of hypertrophic scars (Fig. 21).

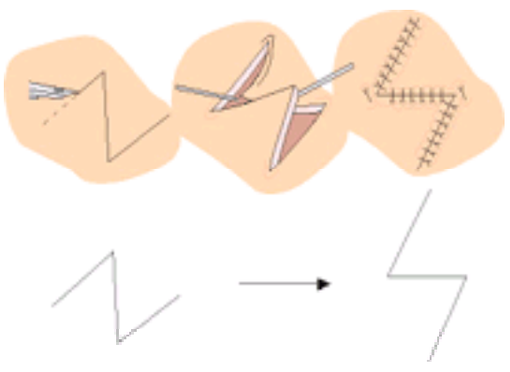


Fig. 21. Z-plasty.

The variables involved in constructing the Z are orientation, limb length, and angle size. When lengthening contractures, the vertical limb is placed in the axis of the contracture. To lengthen a specific scar, the central limb is drawn over the scar and the two limbs are drawn of the same length and at an angle of 60° to the central member. To change the direction of a scar, the proposed resultant central member is drawn across the scar and of the same length. The two limbs are then drawn from the ends of the new central member to the ends of the old central member, with some variation of angle allowed. For a given limb length, increasing the angle increases the amount of length gained. Clinical experience shows that angles of 60° are most effective. Greater angles produce flaps that are difficult to transpose, while smaller angles are less effective in gaining length and are progressively less well vascularized. For a given angle, the length of the limb determines the amount of lengthening. This length is determined by the amount of tissue available on either side.

Composite flap tissues

The requirements of the reconstruction dictate the types of tissue that will be included in the flap. Most often it contains skin and the underlying subcutaneous tissue. This tissue may be supported by its random or axial blood supply, or may be carried by the subjacent fascia or muscle. These last two structures may contribute to the reconstruction with their bulk or function. Occasionally, muscle will be transferred alone to provide bulk to fill a large cavity or to be a functional muscle unit. In special circumstances, the flap may be designed to contain bone, nerves, intestine, or omentum. The number of different tissues within a flap is limited only by the requirements of the reconstruction and the ingenuity of the surgeon.

Facial reconstruction by anatomic site

Eyelid

Defects in the eyelids resulting either from trauma or tumor resection should be reconstructed immediately to provide protective cover for the globe.

Loss of up to 25 per cent of the lid may be reconstructed by primary closure, which must be carefully layered to maintain lid function. The conjunctiva is closed with fine, buried, absorbable sutures. The tarsal plate is approximated separately with more slowly absorbed sutures. The orbicularis oculi is closed with fine, absorbable sutures. The gray line must be precisely lined up to avoid notching. Three 6-0 silk sutures at the inner and outer edges of the gray line and through the meibomian glands in the center anatomically align the edges. The skin is closed with 6-0 nylon sutures (Fig. 22). When the defect is just greater than 25 per cent of the lower lid, a lateral canthotomy may aid the primary closure. When up to half of the lower lid is lost, the tissue lateral to the eye must be elevated as a rotation flap. A curved incision rising above the lateral canthus allows the temple skin to be rotated medially. The excess conjunctiva in the lateral fornix will provide a lining for the flap after the septum orbitale and inferior lateral canthus have been divided.

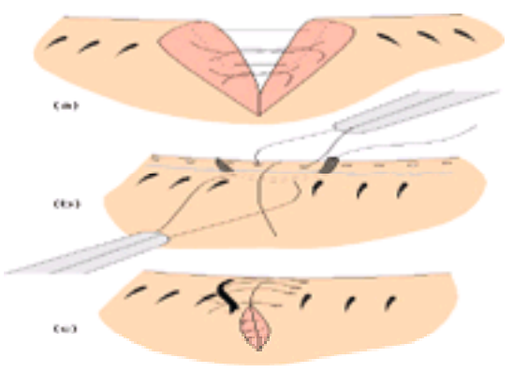


Fig. 22. Closure of lacerations or small resections of eyelids requires precise approximation of anatomic structures.

For defects affecting more than 50 per cent of the lower lid, the lateral incision may be continued down to the preauricular area to rotate a large cheek flap. In these cases, there will be insufficient conjunctiva for lining, and a free graft of septal cartilage with mucoperichondrium on one side provides lid support and a conjunctival replacement. Alternatively, an advancement flap from the upper lid, including part of the tarsus and the conjunctiva, can be brought to the lower lid, particularly for narrow horizontal defects. The outer skin surface is replaced with a full-thickness skin graft or with a pennant flap from the upper lid.

Upper-lid defects of up to 25 per cent may be closed similarly to lower-lid defects. For larger defects, tissue is borrowed from the lower lid in the form of a cross-lid flap. The lacrimal punctum of the lower lid should be preserved, but the borrowed portion of lower lid can be replaced by the techniques mentioned above, depending on the size of the cross-lid flap.

Nose

Reconstruction of the nose is one of the oldest plastic surgical procedures. Nasal amputations, a not uncommon form of punishment, were reconstructed in India in approximately 500 BC by Susruta. Nasal reconstruction was adopted by Western European medicine in sixteenth-century Italy, where a flap of skin from the upper arm as well as the ‘Indian' forehead flap were used.

Restoration of a functional and esthetic nose needs a careful analysis of the missing parts. The cantilevered part of the nose is a sandwich of lining, support, and cover, all of which must be considered in the reconstruction. The contours, features, and margins of the face separate it into distinct areas that are generally best reconstructed as a whole (see Fig. 5).

Partial

Full-thickness loss of the dorsal skin of the nose due to tumor resection or trauma is best restored by a local flap first and a full-thickness or composite skin graft later. The nasal skin may be thick and flaps may not transpose easily. A defect of the tip of the nose may be restored by elevating a flap of the entire nasal dorsum including a V cut into the glabellar skin. The flap is transposed to fill the defect and the glabellar V is closed in the shape of a Y. A graft from the concha of the ear can be used to replace any missing cartilage.

A side-wall defect is best replaced by a nasolabial flap. The flap is based on the angular arterial blood supply and may be designed with a superior or inferior base. Nasal reconstructions will usually be based on the superior supply and can be cut with an island paddle of skin to tailor the flap inset and the donor-site closure.

Total

Subtotal or total reconstruction of the nose is best effected by transposition of a forehead flap. This procedure is one of the oldest in plastic surgery. A flap based on the supratrochlear vessels may be turned 180° to cover the nasal structures (Fig. 23).

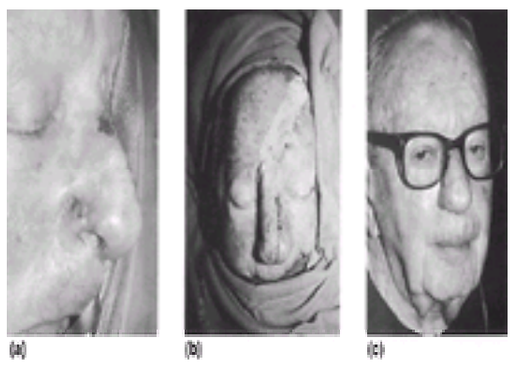


Fig. 23. Nasal reconstruction with a forehead flap: (a) nasal tumour; (b) forehead flap rotated into place; (c) result.

Ear

Reconstruction of all or part of the ear for either a congenital or a post-traumatic defect requires a cartilage framework and a thin-flap skin cover. The best source of cartilage graft is the confluence of the lower costal cartilages. Depending on the size of cartilage required, the cartilage may be harvested from the same or opposite side. A slightly exaggerated framework is carved from the rib to simulate the helical and antihelical folds. The key to the operation is the provision of a thin skin flap with enough redundancy to reveal the architecture of the underlying cartilage graft.

Breast surgery

Augmentation

Small breasts may be enlarged surgically by the placement of an implant behind the mammary tissue or behind the subjacent pectoral muscle. The indications for the procedure are the cosmetic interests of the patient or marked asymmetry due to dystrophy of one breast (Fig. 24).

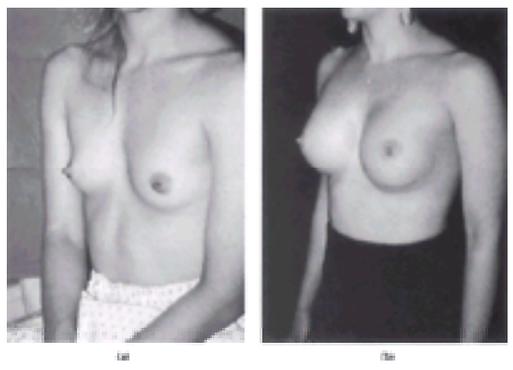


Fig. 24. Breast augmentation by subpectoral implant placement: (a) preoperative; (b) postoperative.

The operation may be performed either under general anesthesia, or with local anesthesia and intravenous sedation. The route of access is through an incision in the areola, just above the inframammary crease, or within the axilla. The dissection is carried down to a plane under the pectoralis major and above the pectoralis minor muscles, where a pocket substantially larger than the proposed implant is made. The pocket used to be made above the pectoralis major, but the incidence of scar capsule contraction seems to be higher in that position. Implants for breast augmentation or reconstruction are usually Silastic outer envelopes filled with either a gel of

the same material or with saline.

The most frequent complication is scar capsule contraction around the implant, causing the breast to be rounder and firmer than desired. There may also be decreased nipple sensation, hematoma formation within the pocket, a hypertrophied scar at the incision, or asymmetry of implant position.

Reduction

Overly large breasts may cause neck and back pain, grooving of the shoulders from the brassiere straps, and intertrigo. A breast reduction is performed either to alleviate these symptoms or to improve the appearance. There are no anthropomorphic criteria for breast hypertrophy, but it is generally accepted that a reduction will remove in excess of 300 g per breast. There is a limit to how large a reduction may be effected without producing a flattened or square appearance of the breast. General anesthesia is required.

The pattern for breast reduction is best drawn on the patient in a standing position before surgery. Many patterns and operative designs have been devised. The design described here will be based on the Wise template for breast shape and on a combination of a bipedicle–augmented inferior pedicle for blood supply to the transposed nipple.

The central pedicle is de-epithelialized so that it can be buried subcutaneously. The nipple, with a reduced diameter, is left as an island on the pedicle. Medial and lateral segments of breast tissue and fat are removed. A suction lipectomy may be performed to reduce axillary fat. The flaps of breast so created are tailored to leave symmetric amounts of tissue on both sides. The incisions are closed in several layers over suction drains, with the nipple folded on to the center of the ‘keyhole' (Fig. 25).

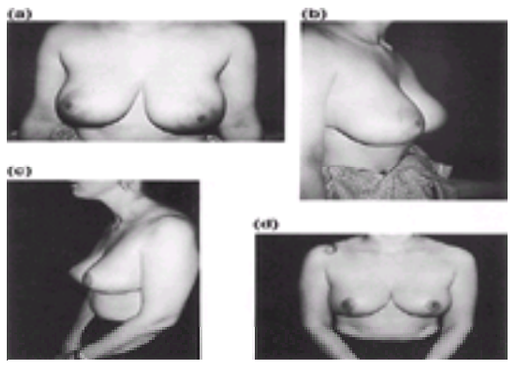


Fig. 25. Breast reduction is performed by advancing the nipple superiorly on a vascularized pedicle with resection of the inferior tissues: (a) and (b) preoperative; (c) and (d) postoperative.

Possible complications include hematoma, decreased nipple sensation, hypertrophied scars, asymmetry, and nipple loss due to ischemia. Patient satisfaction with the change in breast shape and weight is usually high.

Reconstruction

A breast reconstruction after mastectomy is not a necessity since adequate external prostheses are available. However, the convenience and appearance of the reconstructed breast are superior. Recent advances in implant technology and types of operation promise reconstructions that more accurately simulate the opposite breast.

The operation may be performed immediately after the mastectomy under the same anesthetic, or at any time in the future. The immediate procedure has the advantage of saving a later operation and perhaps in shoring up the patient's psyche. It also preserves more of the native breast skin that would otherwise be discarded during the mastectomy closure. Preservation of this conically shaped skin may allow a superior cosmetic result. The delayed operation may give the woman more time to decide which type of reconstruction she desires, if any, as well as preventing complications of the reconstruction interfering with required postoperative radiation or chemotherapy.

The simplest breast reconstruction is an implant placed in a submuscular pocket, similar to augmentation. In contrast to augmentation, the pocket is created completely beneath muscle (pectoralis major, serratus anterior, and rectus abdominis) to protect the implant from overlying flap necrosis and to minimize scar capsule contraction. This type of reconstruction suffers from its inability to simulate anything but the smallest and least ptotic breast.

The ellipse of skin resected in the mastectomy may be replaced by a transposed island myocutaneous flap from latissimus dorsi with an implant placed under the latissimus dorsi and pectoralis major. A larger and more ptotic breast may be created, but at the expense of a back scar. There is little functional deficit from the transposition of the muscle. The latissimus dorsi flap is particularly useful when the breast has been irradiated preoperatively.

The resected skin can also be ‘replaced' by skin expansion before implant placement. This procedure has the advantage of providing additional skin without requiring a new donor site. At the initial operation a Silastic envelope with a subcutaneous filling portal is inserted. After surgery, saline is injected through this portal at weekly or biweekly intervals. Much like a pregnant abdomen, the overlying skin is stretched to a larger surface area. The breast is often expanded to a volume greater than that ultimately desired, to allow for skin hysteresis. Some 1 to 3 months after the placing of the expander, a permanent implant is sited in the expanded pocket, with some attempt to create an inframammary crease. The breast so reconstructed is larger than that made by an implant alone, but, despite some hint of ptosis, will not match an opposite pendulous breast.

If the opposite breast is large or ptotic, final symmetry is enhanced by performing a mastopexy or breast reduction on that side. This operation is done at the time of breast reconstruction, or at a later date to maximize symmetry.

The breast is most naturally reconstructed with a reasonable match to even a large or ptotic opposite breast by a transverse island myocutaneous flap from the rectus abdominis. The operation is suited to the patient who does not want a permanent implant or who would not want any surgery on the opposite, large or ptotic breast. It has the further advantage of providing a ‘tummy tuck' at the time of breast reconstruction. However, the scale, duration, and extent of the procedure should not be underestimated.

A transverse island of skin from the lower abdomen is elevated, based on one or two of the rectus abdominis muscles. The upper abdominal skin is elevated, and the flap is passed through this tunnel to the reconstruction site. Only slightly more than one-half of the lower abdominal skin receives sufficient blood supply from the ipsilateral rectus abdominis. The flap is turned into the site of the new breast to recreate an axillary fold as well as the pendulous form of the breast (Fig. 26).

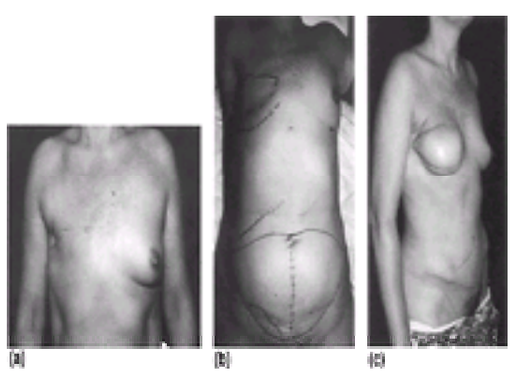


Fig. 26. A flap of lower abdominal skin and fat may be carried on the rectus abdominis muscle to reconstruct the breast: (a) mastectomy defect; (b) transverse rectus

abdominis myocutaneous flap design; (c) postoperative appearance.

If the patient is obese, even the ipsilateral skin may not entirely survive on one rectus abdominis muscle. In that case, two muscles may be transposed together to support the flap. However, the double transfer weakens the abdominal wall. Alternatively, the flap can be transferred as a free flap based on the inferior epigastric vessels anastomosed to branches of the axillary vessels. The flap may also be transferred on one muscle with a microvascular anastomosis of the inferior epigastric vessels.

Breast reconstruction by transverse island myocutaneous flap from the rectus abdominis is an extensive operation. The blood loss from the abdominal dissection is moderate, and an autologous unit of blood is often required. The anterior fascia of the rectus abdominis superior to the arcuate line is partially or completely resected, which results in some laxity of the abdominal wall; this defect can be supported with a patch of synthetic mesh. Although it was initially thought that ventral hernias would not be seen, observation over time is revealing an incidence of 5 to 10 per cent.

Subcutaneous mastectomy

The breast tissue may be removed with preservation of the entire breast skin and nipple in the treatment of benign breast disease. Incapacitating mastodynia is the primary indication for this procedure; the indications have been extended to removal of the breasts with premalignant disease. Although one would expect a decrease in the incidence of breast cancer in the breasts so operated, that decrease has not yet been documented without question. In addition, the presence of an implant may hamper the detection of breast cancer in the small vestige of breast tissue left after surgery.

The operation is performed in a similar manner to breast reduction. The nipple is maintained primarily by a superior pedicle and the breast is degloved through an incision in the inframammary crease. A total submuscular pocket is made to accept an implant.

Gynecomastia

Hypertrophy of the male breast during adolescence usually resolves spontaneously in several years. Endocrinologic evaluation is unwarranted. Persistent breast enlargement in later years may be reduced by a combination of resection and suction lipectomy. Unilateral breast enlargement in older men should be biopsied because of the rare occurrence of male breast cancer.

The operation can be performed under general anesthesia, or local anesthesia and sedation. If there is true glandular hypertrophy, an incision is made around the inferior portion of the areola. A portion of the glandular (firm and white) tissue is resected. Care should be taken not to resect the entire mass since the nipple will indent without some support. For the fatty breast and for the perimeter of the breast with true hypertrophy, suction lipectomy is performed to reduce the surrounding fat and to ‘feather' the edge of the excision into the surrounding area. Some surgeons believe that the entire procedure can be done with suction alone. However, since true glandular tissue resists even the scalpel, suction alone is not likely to be able to remove it. All tissue removed should be sent for pathologic examination.

Abdominal wall reconstruction

Large defects of the abdominal wall resulting from trauma, infection, extirpative surgery for cancer, resection of fistulas, and late sequelae of radiation therapy can be difficult to close. Fistulas may be to bowel or bladder, and may need concomitant repair in combined procedures. The local tissues are often infected and friable, requiring extensive preoperative debridement. They will be inadequate for primary repair, leaving exposed viscera or a skin closure without fascial support that may result in herniation.

The ideal reconstruction of the abdominal wall should restore its normal layers and contain vascularized fascia, subcutaneous tissue, and skin; these are preferably supplied by a single reconstructive unit in a one-stage procedure. Knowledge of the vascular anatomy of the abdominal wall allows the design of safe elective incisions and of appropriate flaps.

Many methods are available. Zone I comprises the upper abdomen and epigastrium: it is defined as the area from xiphisternal junction superiorly to the umbilicus inferiorly and to the midaxillary line laterally. Although this area is central on the body surface, it is one of the more difficult areas for which to provide flap coverage, particularly when the superior epigastric arteries have been sacrificed in the defect. Myocutaneous flaps are most easily transposed from the center outward or from the extremities toward the center. There are few good flaps in the center of the body that can be transposed locally. Flaps useful in repairing zone I defects include the rectus abdominis muscle flap, the extended deep inferior epigastric flap, the external oblique muscle flap, the lateral intercostal flap, the latissimus dorsi muscle flap, and the greater omentum.

Zone II comprises the mid-abdomen, which includes the periumbilical area from 2 cm above the umbilicus to 2 cm above the pubis and laterally to include both flanks. One of the most reliable and versatile flaps for the reconstruction of zone II defects is the tensor fascia lata flap, which can supply fascial support as well as skin coverage. Other possible flaps and donor sites include the rectus abdominis, the rectus femoris, the extended deep inferior epigastric flap, the groin flap, the posterior thigh flap, and the external oblique flap.

Zone III consists of the suprapubic and groin areas. Defects in this zone present the combined reconstructive problems of potential herniation, contaminated pubic bone, and exposed femoral vessels. The flaps available for defects in zone III include the tensor fascia lata flap, the groin flap, the gracilis muscle flap, the sartorius muscle flap, the rectus abdominis flap, and the vastus lateralis muscle flap.

Rectus abdominis muscle flap

The rectus abdominis muscle extends from the lower costal margin and the cartilages of the sixth, seventh, and eighth ribs to the pubic tubercle on the pubic crest. The two rectus muscles are bilaterally SYmmetric around the midline. The dominant blood supply is from the deep inferior and superior epigastric systems, and each can usually support the entire muscle. The muscle can therefore be reflected superiorly on to the chest wall for reconstruction as well as inferiorly into the perineum and groin. Its motor and sensory nerves derive from multiple intercostal nerves. If the superior epigastric artery is intact, the rectus abdominis flap is the best choice for repair of epigastric defects. Based on the superior pedicle, with a point of rotation just below the xiphisternum, the flap will cover the lateral abdomen, upper flank, lower chest, and sternum. It may be especially helpful in filling defects about the xiphisternum, where pectoralis major flaps will not easily reach. The rectus abdominis flap can be used with its overlying skin for reconstruction of the chest wall, taking skin from the lower abdomen and transferring it to the chest. The rectus abdominis primarily provides soft-tissue coverage and is a weak support for the integrity of the abdominal wall, even when its anterior sheath is included. It can be used in conjunction with a supporting Prolene or Marlex mesh if necessary.

The posterior rectus sheath above the arcuate line (semilunar line of Douglas) should be preserved to facilitate closure of the donor site and avoid production of a hernia. Below the arcuate line, if the anterior sheath is taken with the muscle flap, the defect should be supported by synthetic mesh to avoid hernia formation. Donor sites can usually be closed primarily.

Tensor fascia lata flap

The tensor fascia lata muscle originates from the anterior superior iliac spine and from the greater trochanter of the femur. The fascia lata is an extension of this small muscle and inserts into the lateral aspect of the knee, where it functions as a stabilizer. If one draws a line from the greater trochanter to the midlateral aspect of the knee, it should bisect the center of the tensor fascia lata muscle as well as the flap. The anterior and posterior limits of the tensor fascia lata flap are the rectus femoris and the biceps femoris muscle, respectively. The dominant blood supply is from the lateral circumflex vessels, which enter the muscle approx. 10 cm below the anterior superior iliac spine and the inguinal ligament. The motor nerve of the tensor fascia lata muscle branches from the gluteal nerve, and the sensory nerve to the overlying skin derives from the lateral cutaneous nerve of the thigh and sensory branches of T12 (Fig. 27).

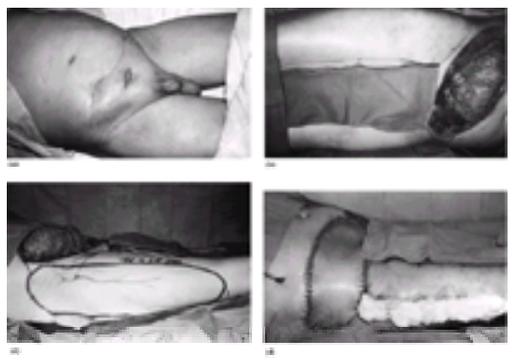


Fig. 27. The tensor fascia lata (TFL) myocutaneous flap is useful for lower abdominal reconstruction: (a) abdominal wall metastasis; (b) abdominal wall resection; (c) TFL flap design; (d) abdominal wall reconstruction.

Wangensteen described the use of this flap as a pedicled island for replacement of the abdominal wall. The muscle and its fascial extension can be elevated on the basis of the vascular pedicle, with a variably sized skin paddle either with a proximal subcutaneous base or as a true skin island. The tensor fascia lata flap will reach to just above the umbilicus and is the flap of choice for most defects below this level. The cephalad reach of the flap can be extended to zone I by dissecting the lateral circumflex vascular pedicle back to the femoral vessels, ligating the side branches to the rectus femoris and vastus lateralis muscles, and tunneling the entire island flap under the rectus femoris muscle to exit over the femoral vessels. There is little or no functional loss associated with harvest of the tensor fascia lata muscle. The donor site can be closed directly when the skin paddle is less than 8 cm in width.

In addition to its primary role in the closure of zone II defects, the tensor fascia lata flap is also a good flap for the reconstruction of pubic defects and for covering ischial pressure sores.

Gracilis muscle flap

The gracilis muscle originates from the pubic tubercle on the inferior ramus of the pubis and inserts on to the medial knee. Its main function is a flexor of the knee as well as an accessory adductor of the thigh. It has a proximal blood supply from the profundus femoris artery, which enters the muscle approx. 10 cm below the inguinal ligament. Occasionally, the dominant vessels will enter as low as 16 to 17 cm below the inguinal ligament, which limits the arc of rotation of this flap, making it useless for reconstruction of the abdominal wall. Doppler evaluation of the level of vascular ingress into the muscle is advised before operation. There is a distal minor vascular pedicle, which can be divided without vascular compromise to the distal muscle. The motor nerve to the gracilis muscle comes from the femoral nerve, and the sensory nerve to the skin paddle arises from the medial cutaneous nerve of the thigh.

The blood supply to the subcutaneous tissue and skin overlying this muscle comes from vessels wrapping around this muscle in the intermuscular septa and not from direct muscular perforators. Because of this septal blood supply, the skin paddle of this flap, particularly in the obese patient, is not as reliable as that of other myocutaneous flaps.

The gracilis muscle is generally used for primary defects of the perineum, vagina, scrotum, and penis. After pelvic exenteration, paired gracilis myocutaneous flaps can be joined in the midline to close the pelvic floor and to recreate a vaginal sheath. This flap may also be used to cover irradiated tissue and exposed vessels in the ipsilateral groin.

Microsurgery

History

Microvascular surgery and microneurosurgery developed as a combination of the microsurgical techniques of ophthalmology and the macrovascular techniques derived from the work of Alexis Carrel. In plastic surgical reconstruction the microscope is used primarily for limb replantation, nerve surgery, and free-tissue transfer.

Microvascular surgery, the anastomosis of 1-mm vessels, began in the laboratory in 1960. Early demonstrations that small vessels could be successfully anastomosed led very quickly to clinical revascularization and replantation. The first arm replantation was performed by Malt; the first clinically successful revascularization of a thumb was performed by Kleinert in 1963; and the first true replantation of a thumb was performed by Komatsu and Tamai in 1965. In the next decade, large series of successful limb replantations were reported from around the world.

Microvascular technique

The instruments of microvascular surgery include the double-head binocular operating microscope, the microsurgical instruments of ophthalmology, and very fine nylon sutures with swaged-on needles. It was the development of these sutures that enabled microvascular surgery to begin.

Vascular anastomosis is begun by placing the vessel ends approximate to each other in a nontraumatic clamp. Initial interrupted nylon sutures are placed 120° on the circumference of the vessels. After an additional one or two more sutures have been placed in the ‘front wall', the clamp and vessels are turned over. Since the ‘front wall' sutures were placed relatively close together, the flaccid ‘back wall' can be sutured without danger of suturing the lumen. Veins, because of the lower intraluminal pressure, require fewer sutures than do arteries (Fig. 28).

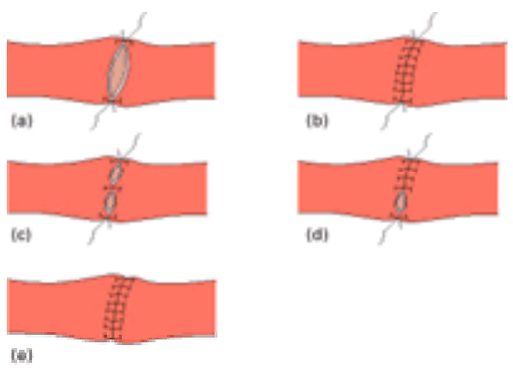


Fig. 28. The technique of microvascular anastomosis.

Replantation

Replantation of amputations has grown in the past 30 years from a laboratory experimental model to a common technique in trauma surgery. Microvascular surgery and replantation have developed together, the one stimulated by the other.

Indications

All patients who have an amputated part should be considered candidates for an attempt at replantation. Replantation is contraindicated only if the patient's poor condition would discourage a prolonged operation or if it anticipated that the replanted part would function less well than a prosthesis.

In general, young, healthy patients with a clean (guillotine) amputation have the best result. The older the patient, the less good function is likely to be. Major amputations at the upper arm, forearm, or hand should be replanted if possible because of the devastating functional loss. Smaller amputated parts frequently provide the best reconstruction for the injury when replanted. The thumb and multiple fingers should be replaced if possible. Even smaller parts of single fingers may be replanted in some patients.

Preservation of the ischemic part

The amputated part should be cooled from the time it is retrieved by the medical team. The cooling is usually effected by wrapping the part in moist gauze and placing it in a plastic bag immersed in iced saline. Cold ischemia preserves a distal part for up to 24 h or more; proximal amputations that contain a large bulk of muscle can only be preserved for 6 to 8 h.

Sequence of surgical repair

The bone is usually shortened proximally and distally to allow easy approximation of all structures. The sequence of repair is carried from deeper to more superficial and from larger to more microscopic structures. The bones are fixed with interosseous wires, Kirschner wires, or fixation plates. The extensor and flexor tendons are repaired. The vascular repairs may begin with the artery or vein first, although initial venous repair will minimize blood loss when the arterial repair is unclamped. The nerve repairs are performed next, followed by the soft-tissue closure.

Replantation by anatomic site

Finger and thumb

The ideal reconstruction for an amputated thumb is replantation. Even if the replanted thumb does not move well, a sensate thumb post provides good function. Very distal amputations of the thumb are replanted, as thumb length is important. As the amputation approaches the eponychium, it may be difficult to find suitable veins for anastomosis. In this case the nailbed or the soft tissue can be allowed to bleed to decompress the part (Fig. 29).

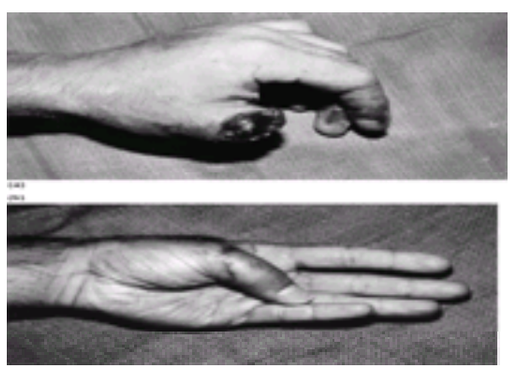


Fig. 29. (a) An amputated thumb is one of the prime indications for replantation. (b) Result after replantation.

Amputation of many fingers at all levels except the most distal deserves an attempt at replantation. From inspection of the parts it may be difficult to determine which will survive or which will be the most functional. Single fingers amputated distal to the proximal interphalangeal joint provide excellent function after replantation. Fingers amputated through ‘no man's land' because of the devastating combined injury to all structures including the flexor tendons may be expected to require a prolonged course of physical therapy to regain motion. Single fingers replanted in ‘no man's land' may be bypassed in future hand activity, but multiple replanted fingers in the well-motivated patient usually function better than the alternatives.

Hand

Amputations through the wrist or distal forearm produce some of the best functional restorations of all replantations. The forearm bones are fixed with plates. All tendons and nerves are repaired. The structures are large, and two arteries and several veins should be repaired if possible. Amputations within the carpal bones may require proximal row carpectomy for fixation (Fig. 30).

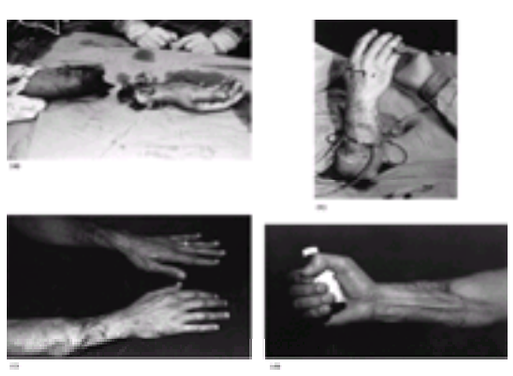


Fig. 30. Hand replantations through the wrist or forearm provide some of the best functional results.

Viability and functional results

Success in revascularizing the amputated part depends to a large degree on the mechanism of injury. Clean, guillotine amputations should have a rate of anastomotic patency in excess of 90 per cent. A crushing component to the mechanism widens the zone of injury, increasing the risk of vascular thrombosis. Avulsing injuries will produce a zone of injury, particularly within the vessels, that may make revascularization impossible, even with long vein grafts. The overall rate of viability should be around 80 per cent if the replantation team is attempting difficult as well as easy replantations.

Functional results are dependent on the mechanism of injury and the age and motivation of the patient. Young patients will regain more motion than old patients.

Free-tissue transfer

History

Following the success of replantation, several investigators determined that microsurgery could be used to transfer flaps from one area of the body to another. Such flaps were initially shown to be possible by Goldwyn and Krizek in the laboratory. In 1973, Daniel performed the first successful free flap in a patient by transferring the groin flap based on the superficial circumflex iliac vessels.

Flaps

Axial arterial flaps and myocutaneous flaps are usually chosen for free-tissue transfer because their arteries and veins are of sufficient caliber and provide a blood supply to a sufficiently large area to function as a reconstructive transfer.

The groin flap was the most commonly chosen flap in the early era of free-tissue transfer. Maxwell reported on the use of the latissimus dorsi muscle, either alone or with an island skin paddle, and emphasized its large size, malleability, and long vascular leash as desirable characteristics. The latissimus dorsi flap remains as one of the most reliable of free-tissue transfers (see Fig. 17).

Over the ensuing years, multiple flaps for microsurgical transfer have been devised by ingenious surgeons confronted by difficult reconstructive problems. The flaps whose vascular supplies had been researched during the time when myocutaneous flaps were being developed were sequentially employed as free flaps. In addition, several new flaps, such as the great toe for thumb reconstruction, were devised in direct response to the microsurgical potential.

Technique

The operation of free-tissue transfer requires the preparation of the flap in one area of the body and the preparation of the recipient site in another. The procedure is frequently performed by two surgical teams working simultaneously.

The donor flap is dissected out fully, with the vessels skeletonized but not ligated, so that it remains perfused until the recipient vessels are ready for anastomosis. The surgical crew working on the recipient site prepares the wound to be reconstructed; this may involve the excision of an ulcer, a tumor, or osteomyelitic bone. Suitably nearby vessels of the appropriate caliber and blood flow for the vessel size and perfusion requirements of the flap are then dissected out. When both operative sites are ready, the microscope is positioned, the flap divided, and the anastomosis performed.

The microsurgical anastomosis of free flaps is more often done end-to-side rather than end-to-end as commonly employed in replantation surgery. Since the recipient, or inflow, vessels may be important or essential to maintaining the distal portion of the limb being reconstructed, the end-to-side anastomosis will allow flow to the flap and the limb. In some cases, it may be possible that the flap will ‘steal' arterial flow from the distal limb through its end-to-side anastomosis. On the other hand, in operations on limbs that are chronically ischemic from arterial insufficiency, this flow may, in fact, revascularize the distal limb through neovascular connections to the bed in which the flap is inset.

Lower extremity reconstruction

Acute coverage of leg wounds

High-velocity injuries to the lower extremity frequently result in open fractures that are further complicated by soft-tissue loss and perhaps by distal devascularization. Aggressive early debridement and internal fracture fixation necessary for limb salvage can only be performed when soft-tissue coverage can be provided in the acute or subacute phase. The microsurgical techniques initiated by replantation and free-tissue transfer have revolutionized the salvage of lower limbs.

The central tenets of this approach include the fixation of bone fractures with preservation of limb length, maintenance or restoration of normal distal blood flow, and closure of the wound by ample, well-vascularized soft tissue. This surgery should only be performed if sensibility can be expected in the foot and if the anticipated limb function will be better than would be expected for an amputation and prosthesis. The function of the lower limb with a prosthesis is quite good, particularly for a below-knee amputation. Thus the anticipated function of the reconstructed part must be very good to be worth the multiple surgical procedures and prolonged rehabilitation usually required.

Osteomyelitis

Chronic osteomyelitis of the lower limb is best treated by complete excision of the affected bone followed by coverage with a well-vascularized flap. For osteomyelitis in sites as distal as the midtibia there are adequate local muscle flaps that can be transposed into position. However, for distal osteomyelitis close to the malleolus, flaps for soft-tissue coverage are not available and free-tissue transfer is required.

The essential part of the surgical therapy is the debridement of infected and necrotic bone. Initial surgical debridement is performed before antibiotic administration to be sure that the offending organisms are identified by culture. The complete excision of all involved bone is assured by repeated surgical procedures (sometimes three are required), with examination of the bone surfaces in the wound between procedures for the development of granulation tissue. Occasionally, the extent of the infection necessitates the removal of the complete circumference or a segment of the bone. In this case, an external fixator must be applied to hold the bones in alignment and the limb at proper length. At the least the limb is maintained in a protective cast or splint and in a non-weight-bearing state to avoid pathologic fracture. Once the bone is covered by granulation tissue, the wound may undergo final debridement and closure with a free-muscle flap. Usually this operation is preceded by an arteriogram to check that the recipient vessels are sufficient to support the flap (Fig. 31).



Fig. 31. Free flap for osteomyelitis: (a) osteomyelitic tibial defect; (b) gracilis flap design; (c) flap ready for microvascular anastomosis; (d) healed wound.

Postoperative care is similar to that for replantation. Anticoagulants may be given intravenously. The patient is maintained at bedrest with the affected limb elevated. Suction drains are kept in the donor and recipient wounds until drainage is minimal or until the flap is deemed to be well adherent to the bone in the wound bed. The myocutaneous or muscle flap can be examined directly to be sure that anastomotic thrombosis has not occurred. In addition, the subcutaneous or vessel surface temperature can be followed for quantitative evidence of thrombosis. If thrombosis is suspected, the patient should be returned as soon as possible to the operating room for direct examination of the anastomosis and any repairs that are necessary. Free flaps have an average initial viability rate of about 90 per cent; of the 10 per cent that undergo early thrombosis perhaps half can be salvaged.

Head-and-neck reconstruction

Scalp and forehead

Trauma or resection for cancer may result in large defects of skull and scalp. Dural resection, if required, is usually performed in concert with a neurosurgeon. In a noninfected surgical field, the bone defect may be restored at the same time as wound closure. If the wound is contaminated or if there is concern about potential flap necrosis, the bone defect may be reconstructed at a later date.

Local transposition flaps of the entire thickness of the scalp are based on the rich vascular supply from the anastomotic network of the superficial temporal and occipital vessels. Large flaps may be transposed safely to cover the defect, with the donor site closed with a split-thickness skin graft. Because the galea is the least elastic layer of the scalp, the flap can be stretched to a larger surface or greater reach by scoring the underlying galea. Care must be taken in the course of the scoring to avoid injury to the vessels that run immediately above the galea.

Skin expansion is particularly useful for scalp closure. The skull provides a firm base on which to stretch the overlying skin, and of course, the flap transfers

hair-bearing tissue without leaving a glabrous donor site (Fig. 32).

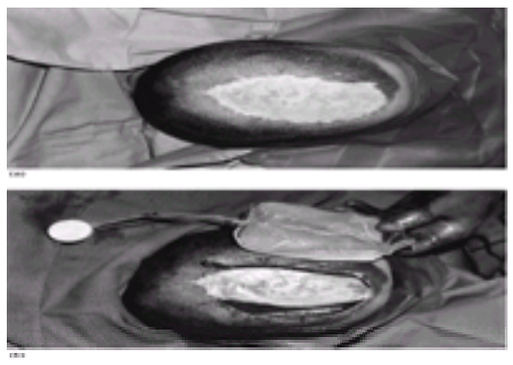


Fig. 32. Scalp expansion: (a) scalp defect; (b) removing expanders leaves redundant skin for advancement.

An adjacent, but not local, flap that may be used for scalp reconstruction is the trapezius myocutaneous flap. Based on the circumflex cervical vessels and the trapezius muscle, a paddle of skin can be brought from the midback to the cheek, orbit, or scalp. This flap is particularly suitable for occipital reconstruction. The donor site can be closed primarily or skin grafted.

As with other sites at the body's periphery where local flaps may not be sufficient for large defects, the scalp is frequently the site of reconstruction by free-tissue transfer. Free flaps that provide the broad, flat characteristics required for the scalp include the latissimus dorsi muscle, the scapular flap, and the omentum covered by a split-thickness skin graft. The superficial temporal vessels are frequently small and spastic, and vein grafts may be necessary to derive an inflow from the external carotid (Fig. 33).

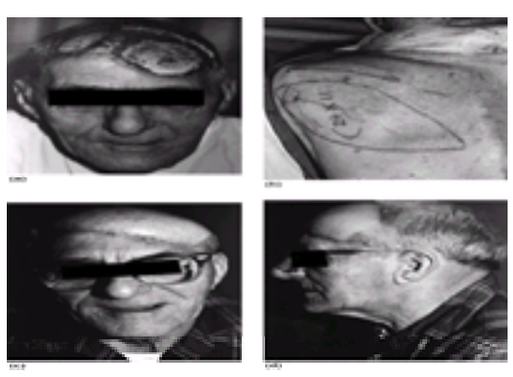


Fig. 33. Free flap for head-and-neck reconstruction: (a) electric injury to scalp; (b) scapular flsp design; (c) and (d) postoperative.

Floor of mouth

Resection of cancer of the floor of the mouth may leave a defect affecting portions of the tongue, the floor of the mouth, the mandible, and the cervical lymph nodes. An ideal flap for reconstructing this wound is the pectoralis major myocutaneous flap. An island of skin can be carried on the pectoralis major muscle with its blood supply from the thoracoacromial vessels. The flap and muscle are turned back on themselves through the dissected neck to provide coverage of the exposed carotid artery. The skin paddle is inset into the mucosal and tongue defect (Fig. 34).

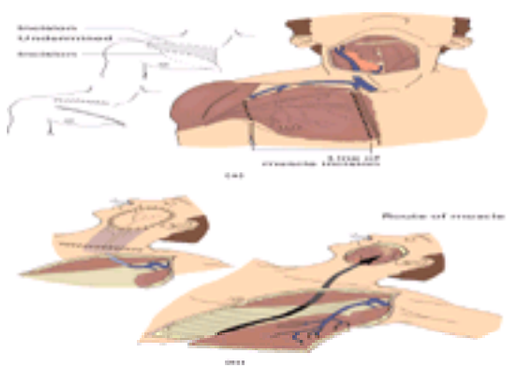


Fig. 34. (a) Pectoralis major flap; (b) flap tunneled under chest and neck to the mouth.

Mandible

A mandibular defect may be reconstructed at the time of resection and primary flap closure, or at a later time. If not reconstructed, a mandibular defect will allow deviation of the mentum to the defect side and will prevent the fitting of functional dentures. However, it is important to emphasize that not every defect requires reconstruction. The bulk of the pectoralis flap will minimize deviation, and a bone graft will not necessarily allow mastication.

Mandibular defects can be simply reconstructed with a plate and screws. For lateral, short-segment losses, the plate functions well. Unfortunately, plates in the chin or in an irradiated area may become exposed, and their removal is then required.

For short segments of bone loss (less than 6 cm), a nonvascularized bone graft from the rib or iliac crest will heal. Defects greater than 6 cm long require a vascularized bone graft. A section of vascularized rib may be carried with the pectoralis major myocutaneous flap. Large defects are, however, best reconstructed by microvascular transfer of free-bone flaps. The best free-bone grafts are from the iliac crest and the fibula. The iliac crest will replace up to 15 cm of absent mandible; the fibula is needed for longer reconstructions. The iliac crest has a configuration suitable to the particular shape of the mandibular ramus and angle. The fibula provides a long piece of bone with a thick cross-section suitable for dental implants (Fig. 35).



Fig. 35. Free fibula and radial forearm flap to total maxillectomy defect: (a) preoperative; (b) osteocutaneous fibular flap for maxilla; (c) additional radial forearm flap for nasal floor and roof of mouth; (d) postoperative.

Maxilla

While not all maxillectomy defects require reconstruction, the hollowness of the cheek, the absence of the malar prominence, and the incontinent palate can be corrected with an osteocutaneous free flap from the scapula. The circumflex scapular vessel penetrates the triangular space to supply the lateral border of the scapula and the overlying skin. After osteotomy, the bone spans the infraorbital defect, the de-epithelialized flap fills the cheek, and the skin on the tail of the flap closes the palate (Fig. 36). If there is also a defect of the external cheek skin, there is an effective need for three skin flaps to cover the external cheek, the soft and hard palate, and the lateral nasal wall. These types of defects can be replaced by designing a large free flap with multiple separate petals of skin separated by de-epithelialized bands. In this fashion, one flap is divided into several flaps to be inset into each separate area. The latissimus dorsi myocutaneous flap or the scapular free flap is particularly suitable for this type of reconstruction. The scapular flap has the added advantage that a section of bone from the lateral border of the scapula can be carried with the vascular bundle and used to reconstruct the zygoma or maxilla. The maxillary defect can also be reconstructed using a radial forearm flap and a separate graft of vascularized fibular bone anastomosed to the radial forearm flap. This technique is particularly useful for bilateral maxillary defects.



Fig. 36. Free scapular osteocutaneous flap to cheek defect: (a) gunshot wound of skin, palate, and mucosa; (b) osteocutaneous scapular flap; (c) flap inset; (d) postoperative with ocular prosthesis.

Hemifacial soft tissue

Soft-tissue deformities of the face, particularly of a hemifacial nature, may be congenital, acquired, or postsurgical. These can be corrected by the microvascular free-tissue transfer of soft tissues from other parts of the body. The first free-tissue transfers attempted for treatment of hemifacial microstomia were free omental grafts. Free-tissue transfer flaps of omentum based on a single vascular pedicle were transferred to the face in a subcutaneous position, with multiple tongues of omental fat carefully placed in the abnormal contours, and revascularized by microvascular and venous anastomosis to the superficial temporal system. Although the initial restoration of facial contour was excellent, the long-term result was one of gravitational drop of the omental fat to the lower face.

A better flap for subcutaneous augmentation is the de-epithelialized scapular free flap. This flap is based on the circumflex scapular branch of the subscapular artery. A large flap of upper back skin and fat can be harvested to restore contour to the forehead, cheek, and jawline. This flap is de-epithelialized so that it can be completely buried under the normal facial skin. The flap can be cut to fit the necessary contours. After suturing in place, the flap is less susceptible to the gravitational change seen with free omental transfers. The flap is frequently so full that a subsequent procedure is needed to decrease the prominence of the augmented soft tissues.

Cervical esophagus

Cervical esophageal reconstruction is ideally effected with a free transfer of a segment of jejunum. Although the colon or the stomach can be advanced superiorly to the cervical esophagus, the vascularity of these tissues at the level of the neck is decreased. A loop of jejunum up to 16 or 18 cm in length can be harvested, based on a single proximal artery and vein supplying the arcade to the entire segment. The segment is transferred to the cervical esophagus and placed in an antegrade direction for normal peristalsis. The artery and vein are repaired using microvascular techniques to restore blood supply to the intestine. The proximal and distal anastomoses are performed in a standard fashion. Intestinal continuity is returned by a standard anastomosis. Because the cervical anastomoses are performed to well-vascularized intestine, the risk of fistula or stricture is low. Oral alimentation can be resumed in 7 to 10 days.

Facial nerve grafting

Total parotidectomy generally entails truncal resection of the facial nerve. The ideal reconstruction is immediate sural nerve grafting. Multiple cables of sural nerve are grafted in an interfascicular fashion from the trunk to all branches.

Fractures of the facial skeleton

Facial trauma

The management of major fractures of the facial skeleton has undergone a dramatic revolution in the last 15 years. Several new diagnostic and therapeutic modalities have evolved, including computed tomographic (**CT**) scans, which make possible the precise definition of fracture patterns, extended open reduction techniques, rigid fixation systems, and acute bone grafting. These techniques make the restoration of anatomy and function realistic goals, even after the most severe injuries.

History

The mechanism of injury is important in estimating not only the extent of injury to the facial skeleton, but also the likelihood of injury to the central nervous system and other major organ systems.

Physical examination

Shortly after injury and before the onset of soft-tissue swelling, architectural deformity with facial asymmetry may be obvious. This abnormality is often best noted by looking from above. Numbness in the distribution of a branch of the trigeminal nerve suggests fracture of the bone through which the branch passes. Hence, numbness in the distribution of the infraorbital nerve suggests a fracture in the zygomatic complex, in that of the supraorbital nerve a frontal bone fracture, and in that of the

mental nerve a mandibular fracture.

Malocclusion

Fractures of the maxilla or the mandible distort the normal relation between the dental arches and therefore the presence of unexpected malocclusion should initiate a search for fractures in one or both jaws (Fig. 37).

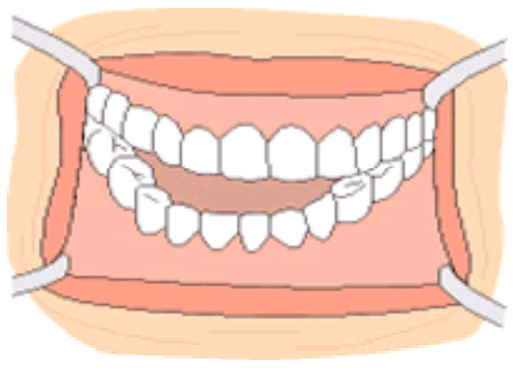


Fig. 37. Malocclusion seen in either maxillary or mandibular fractures.

Mobility

Palpable crepitation and motion should be searched for by gently palpating all contours of the facial skeleton. Midface mobility is elicited by stabilizing the frontal area with one hand and gently grasping and rocking the maxillary alveolus with the other. Motion indicates fracture at one of the LeFort maxillary lines (Fig. 38).

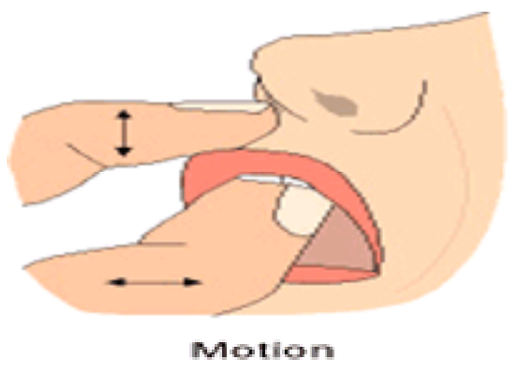


Fig. 38. Midface mobility seen with transverse maxillary perspective.

Radiologic diagnosis

Radiologic examination defines the location and extent of fracture, and aids in determining the need for and method of therapy. In the acute setting, plain films are limited by the need for precise patient positioning and by overlapping bone outlines, making interpretation difficult. The Water's mento-occipital position, the submental vertex, and oblique views of the mandible are all helpful. The single, most revealing plain film is the reversed Water's view, which requires little patient positioning and can be obtained with portable equipment in almost all patients (Fig. 39). Thin-section CT scanning has become the ‘gold standard' for facial skeletal imaging (Fig. 40); it provides precise information about fracture patterns and displacement. In addition, the status of the orbital and intracranial soft tissues is also made clear.
Three-dimensional CT scans generated from summation of averages of two-dimensional slices lose detail but provide good perspective (Fig. 41). Patients suffering high-energy trauma should have the entire cervical spine, including C1–C2 and C6–C7, visualized. If not visualized, a cervical injury should be assumed and the head immobilized with a cervical collar.

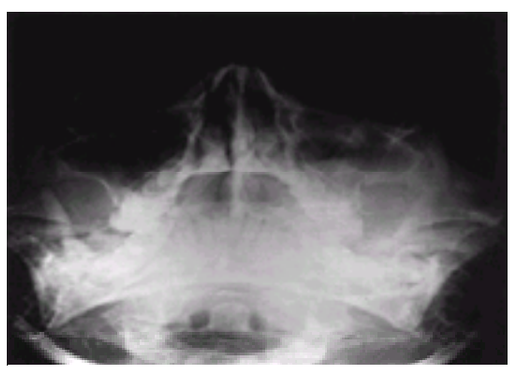


Fig. 39. Reversed Water's view (mento-occipital position) provides the most information on the facial skeleton from single-view portable radiography; there is increased magnification of the facial bones due to the increased distance between the face and the film.

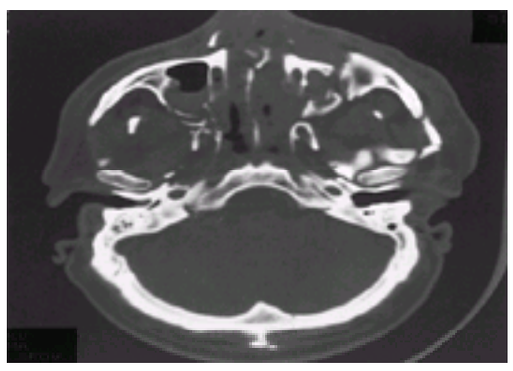


Fig. 40. Axial CT scan provides the most precise information about the facial skeleton.



Fig. 41. Three-dimensional CT scan generated from summation of two-dimensional images results in loss of detail but provides excellent perspective.

Treatment timing

Clinical experience has shown that superior results are achieved when fracture reduction and fixation is performed in the acute phase. Treatment delayed after 10 days becomes limited by contraction of the contused soft-tissue envelope over the malaligned skeletal infrastructure, bone resorption, and bone healing. Emergency therapy is ideal when experienced teams and safe monitoring are available. Hemorrhage, significant risk to vision, ventilatory insufficiency, and elevated intracranial pressure are contraindications to immediate therapy.

Reduction and fixation

Closed reduction techniques may be suitable for uncomplicated nasal, zygomatic, and mandibular fractures. Fractures with displacement and comminution are usually best managed with extended, open reduction and fixation techniques.

Surgical incision

Surgical incisions used to approach the facial skeleton are borrowed from esthetic surgery; they include the bicoronal, lower blepharoplasty, and intraoral gingivobuccal sulcular incisions. Mandibular fractures anterior to the mandibular angle are usually explored intraorally. The angle, ramus, and complicated fractures usually require extraoral incisions (Fig. 42). Through these incisions, subperiosteal dissection is performed to allow exposure of all fracture segments and to attain an anatomic reduction.



Fig. 42. Cutaneous (solid line) incisions available for open reduction and internal fixation of facial skeleton; intraoral incisions are also used (dotted lines).

Whether the scars from elective incisions remain visible is much influenced by their placement. Incisions should be placed within, or parallel to, skin lines with minimal tension that lie perpendicular to the overlying musculature. In the face these are obvious at ‘wrinkle lines' of the muscles of facial expression. Placement of incisions at junctions of unlike tissues such as the brow and forehead skin, or hairline and forehead skin, makes the resultant scar less conspicuous.

Wire osteosynthesis is being gradually replaced by the use of miniaturized plates and screws borrowed from orthopedic surgery. Screws that can gain a purchase on both cortices and are of sufficient scale to counteract the forces of the muscles of mastication are used in the mandible. Stability increases when compression osteosynthesis by the special design of the plate face is employed. In the upper face, small-scale ‘miniplates' and even smaller ‘microplates', which gain purchase on a single cortex and provide three-dimensional stability, are employed.

Bone grafts

When bone is lost or extensively comminuted, the contour is restored by replacing the severely damaged or missing bone with bone grafts at the time of acute reconstruction. The outer table of the skull, the ribs, and the iliac crest are preferred donor sites. Acute bone grafting is never performed in the mandible, where it is accompanied by a high incidence of infection.

In addition to providing three-dimensional stability of the reconstructed skeleton, plate-and-screw fixation allows the release of maxillomandibular fixation immediately after operation, in most cases decreasing the need for tracheostomy, improving oral hygiene, and allowing better postoperative nutrition as well as decreasing hospital stay. Disadvantages of rigid fixation include the expense of the equipment, a steep learning curve for the operator, and the unforgiving nature of the fixation technique.

Treatment by anatomic area

Nose

The nasal bones are the most commonly fractured bones of the facial skeleton. When these fractures are caused by high-energy forces, they are frequently associated with fractures of contiguous areas including the maxilla, orbital, frontal, and cranial base. Plain radiographs are of little help in diagnosis. CT scans aid in diagnosis of injury to adjacent structures in high-energy trauma.

Nasal fractures are grouped into two general types, depending on the direction of the impact forces. Lateral impact forces may fracture both nasal bones and the cartilaginous septum while deviating them to the opposite side. Frontal impact forces cause telescoping of the nasal septum and splaying of the nasal bones; this results in nasal shortening, dorsal saddling, and upper nasal widening. If impact forces are severe enough, fractures extend beyond the nasal bones into the medial orbital and even into the cranial base (Fig. 43). Ideally, nasal fractures are treated immediately after injury, before swelling makes adequacy of reduction difficult to determine; treatment is usually done under local anesthesia, using local infiltration with xylocaine with epinephrine (adrenaline) and cocaine packs to anesthetize the nasal mucosa topically. Septal hematoma should always be drained acutely to avoid problems from pressure necrosis and/or secondary infection.

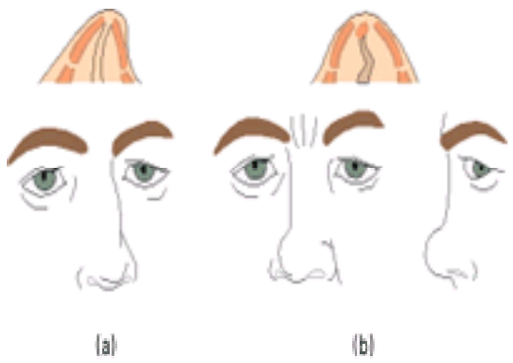


Fig. 43. Patterns of nasal fracture: (a) lateral impact; (b) frontal impact.

In laterally deviated fractures, a Kelly clamp or a knife handle are used to centralize the nasal bones. A greenstick fracture on the opposite side must be completed to allow its repositioning. Frontal impact injuries that result in nasal collapse often require Asch forceps placed on either side of the septum to allow elevation of the nasal bones into their proper position. Fractures that extend into the nasoethmoid region may need bone-graft support.

After reduction, a protective dorsonasal splint maintains bony position. Nasal packing with nonadherent antibiotic-coated gauze is used to maintain the reduction of the septum and to limit hematoma formation; in comminuted injuries, it prevents collapse of the nasal bones by providing counterpressure against the external splint.

Nasoethmoid

High-energy blunt trauma to the central face may result in fractures beyond the limits of the nasal skeleton into the medial orbits, frontal, and ethmoid sinuses. When that portion of the medial orbital rim which bears the medial canthal tendon is fractured with the potential for displacement, acute treatment is required to avoid telecanthus. The characteristic deformities of the fractures when not adequately treated are telecanthus, nasal shortening, and loss of nasal support (Fig. 44). Fractures of this region are often accompanied by injuries to adjacent structures.



Fig. 44. Clinical appearance of untreated naso-orbital-ethmoidal fracture; note increased intercanthal distance, increased nasal width, and decreased nasal height: (a) frontal, (b) lateral.

Proper management requires open reduction and internal fixation, taking care to reposition the medial canthal ligament. Reduction is usually accomplished with transnasal canthopexy. Acute bone graft of the nasal dorsum is usually required to restore nasal height. No attempts are made to assure the patency of the lacrimal system. Proper reduction and internal fixation reduce the need for lacrimal surgery to less than 20 per cent of patients with this injury.

Zygoma

Mistakenly labeled a ‘tripod', the zygoma has articulations with five adjacent structures: the frontal bone, the arch, the greater wing of the sphenoid (lateral orbit), the medial maxilla (orbital rim), and the maxillary alveolus (lateral buttress). Characteristically, frontal impact causes malar depression and arch bowing; lateral impacts cause arch depression. Fractures usually extend into the internal orbit, and if severe, disrupt the floor and medial wall enough to increase the orbital volume, resulting in enophthalmos and dystopia of the ocular globe. Entrapment of the extraocular muscles in the fracture site may cause diplopia. Anesthesia in the distribution of the infraorbital nerve results from impingement on the nerve. Several radiographic views are necessary to evaluate the zygoma. The Water's view shows displacement at the infraorbital rim and fluid in the maxillary sinus. The Caldwell view is necessary to show displacement at the zygomaticofrontal suture, while only the submental vertex view will show posterior displacement of the malar complex. CT scans define the entire area, including the status of the internal orbit and its structures.

Isolated fractures of the zygomatic arch usually have a classic W-shape; they can be treated by closed reduction using the Gillies' approach: a small scalp incision is made within the temporal hair-bearing skin and carried to the deep temporal fascia, then a long instrument is passed beneath this fascia and beneath the zygomatic arch to allow its elevation. Closed reduction is usually stable.

Closed reduction of fractures of the zygomatic body can be performed in selected patients whose fractures are not comminuted or unstable. Closed reduction can be successful in patients who have only infraorbital rim and zygomatic maxillary displacement. Closed reduction should be performed within 72 h after injury and patients should be monitored closely for subsequent displacement.

Open reduction and internal fixation are performed through a transconjunctival incision (sometimes with lateral canthotomy), lateral extension of an upper-lid blepharoplasty incision, and an intraoral gingivobuccal incision, as required. Alignment is confirmed at the zygomaticofrontal suture, the infraorbital rim, and the lateral buttress arch prior to intraosseous wiring or, preferably, plate-and-screw fixation at one or two of these points. Three-point fixation, and preferably plate-and-screw fixation, resist the downward pull of the masseter muscle. In the past, antral packing or K-wire fixation were used to prevent displacement.

Orbital floor

Fractures of the orbital floor inevitably accompany fractures of the zygomatic complex; less frequently they occur as isolated ‘blow-out' fractures. Surgery is indicated for persistent entrapment as determined by a forced duction test, enophthalmos, or vertical dystopia. CT scanning is important in assessment. Defects greater than 1.5 cm result in enophthalmos if not treated. Treatment is intended to restore the anatomy of the internal orbit. After any fat that may have prolapsed into the maxillary antrum has been retrieved, the defect is defined and reconstructed with an alloplastic implant or autogenous bone or cartilage. Defects involving more than the floor are best managed with autogenous bone. Ophthalmic consultation should be obtained before and after surgery (Fig. 45).

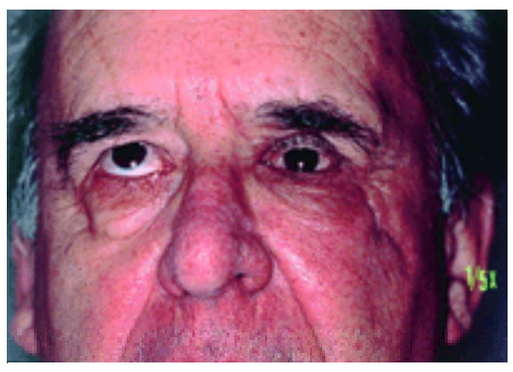


Fig. 45. Patient with enophthalmos and diplopia after post-traumatic disruption of the internal orbit.

Maxilla

Little can be added to the classic work of LeFort, who described the areas of weakness through which midface fractures tended to pass (Fig. 46). The use of CT scans preoperatively, and of extended open reduction techniques in treatment, have shown that fractures rarely occur in ‘pure' Lefort patterns. Rather they tend to occur in combinations. The more severely damaged side tends to have more comminution and a higher level of fracture.

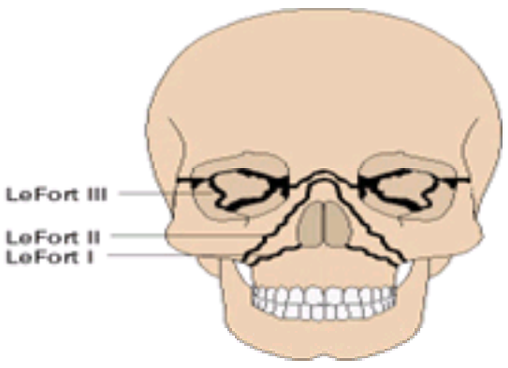


Fig. 46. Classification of midface fractures as proposed by LeFort.

The classic deformity of untreated Lefort fractures is elongation and retrusion of the midface—the ‘dish face deformity'. Malocclusion and mobility of the midface are the clinical findings. Treatment consists of first restoring the occlusion by placing the jaws in maxillomandibular fixation; this restores the proper facial projection. Height is restored by realigning the four maxillary buttresses, using the least injured one as a guide. Osteosynthesis is accomplished with intraosseous wires or, better, with miniplates, which are placed at each buttress (Fig. 47). Bone defects greater than 0.5 cm are filled with bone grafts. All devitalized sinus mucosa is debrided. The maxillary sinuses are not drained.

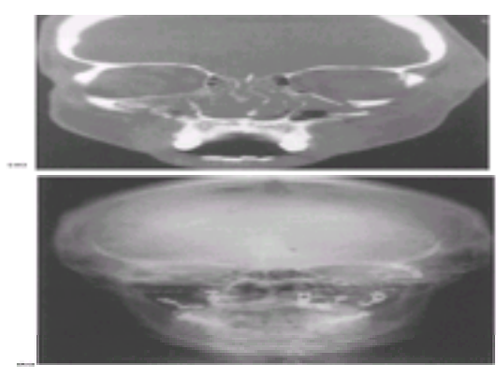


Fig. 47. (a) Coronal CT scan showing a complex LeFort fracture of the facial skeleton before reduction; (b) plain radiograph showing miniplates at each of the four maxillary buttresses in a complex midface fracture after reduction and fixation.

When plate-and-screw fixation is used, maxillomandibular fixation is removed and the occlusion examined. If malocclusion is found, the fixation has distracted the mandibular condyles; because plate-and-screw fixation is non-yielding, the plates must be removed and the fixation process repeated. When rigid fixation is used, maxillomandibular fixation can usually be removed after surgery. Use of extended open reduction and intraosseous wiring alone, or maxillomandibular fixation and suspension wires, requires jaw immobilization for 4 to 6 weeks. The use of extended, open reduction and rigid fixation techniques has replaced suspension wires and external fixation in many centers.

Mandible

Mandibular fractures tend to occur in structurally weak areas of the bone, including the subcondylar area, the angle, and the cuspid (canine) region. In the dentate mandible the subcondylar area is most frequently fractured; in the edentulous mandible, fractures occur most commonly in the body and angle. Fracture displacement is determined by the direction of the blow, the pattern of the fracture, and the direction of pull of the strong muscles of mastication. The unique shape of the mandible results in a high incidence of bilateral fractures; these often occur in relatively standard combinations: symphysis and bilateral subcondylar, parasymphysis and contralateral subcondylar, angle and contralateral body.

Signs of mandibular fracture include malocclusion, local pain, swelling, and crepitation on palpation. Standard radiographs confirm and delineate the fractures. Few fractures escape detection by Panorex (panoral tomographic) or CT examination.

Restoration, or a functional reapproximation, of the occlusion is the goal of therapy. Maxillomandibular fixation is the standard treatment for the majority of mandibular fractures. Arch bars are ligated to the teeth. Elastic traction will overcome the muscle pull to correct small degrees of displacement. Closed reduction using maxillomandibular fixation alone is most likely to be successful with fractures whose muscle force is opposed by the direction of the fracture line, thereby preventing displacement. In unfavorable fractures, the fracture lines do not oppose the muscle pull and displacement tends to occur when maxillomandibular fixation is used alone. Open reduction and internal fixation are, therefore, necessary. In the edentulous mandible, dentures or dental splints may be used to restore the maxillary and mandibular occlusion and provide a means for fixation. These appliances are joined to the mandible by circum-mandibular wiring and to the maxilla by fixing to the alveolar process or with suspension wiring. Maxillomandibular fixation is maintained for 4 to 6 weeks, with the exception of subcondylar fractures, in which jaw excursion must be initiated at 3 to 4 weeks to prevent fibrous ankylosis of the temporomandibular joint, which is usually damaged in subcondylar injuries. Subcondylar fractures are rarely opened for treatment in adults and virtually never so in children.

Open reduction is indicated when fracture displacement cannot be corrected and stabilized by maxillomandibular fixation alone. Intra-osseous wires should be positioned at right angles to the fracture line for best stability. Stabilization is provided postoperatively by splinting with fixation for 4 to 6 weeks. Plate-and-screw stabilization has increased the indications for open reduction with internal fixation since it obviates the need for maxillomandibular fixation in the postoperative period for many patients (Fig. 48). It also has the potential for better results with complex injuries. Despite the increased exposure required, there is no increase of infection when used correctly. Plate-and-screw fixation should only be used when the surgeon is well versed in the principles and techniques of its application.



Fig. 48. Symphyseal fracture treated with a bicortical mandibular compression plate.

When teeth are involved in the fracture line, they are removed if they pose a risk for infection or necrosis. All injuries with teeth in the fracture line and all open fractures should be treated with prophylactic antibiotics.

Cosmetic surgery

Relation of reconstructive to cosmetic surgery

Cosmetic surgery is frequently considered as a series of surgical procedures somewhat apart from reconstructive plastic surgery. Indeed, there are surgical procedures, such as will be discussed below, that are performed only for the improvement in appearance from an acceptable ‘normal' to a more ‘coveted normal'. Reconstructive procedures are likely to have a dominant functional goal. However, all reconstructive procedures have in addition a strong ‘cosmetic' motivation and strive to produce a pleasing appearance. By the same token, cosmetic procedures require the same expertise in plastic surgical techniques as reconstructive procedures, and it may be quite difficult to say whether any one procedure is cosmetic or reconstructive. Is a breast reduction or the restoration of facial motion after paralysis of the facial nerve to be considered cosmetic or reconstructive? For purposes of explanation only, the following surgical procedures are categorized as cosmetic. However, all plastic surgical procedures have both cosmetic and reconstructive aspects and the best plastic surgeons will be accomplished in both.

Rhinoplasty

The external appearance of the nose may be altered in patients with congenital, developmental, or traumatically acquired abnormalities. Rhinoplasty is usually performed after the nasal skeleton has matured in adolescence. It is indicated when the patient wishes to change a perceived irregularity of nasal contour and the surgeon regards this as worth the risks involved. The surgeon must be ready to probe the patient's inner feelings about the appearance of their nose, which may have strong psychosocial aspects.

The procedure is usually an outpatient one, using local anesthesia with intravenous sedation. The goal is to alter the underlying bone and cartilage skeleton, and to allow the skin to contract around this new form (Fig. 49). Surgery is performed as much as possible through the nostrils with intranasal incisions.

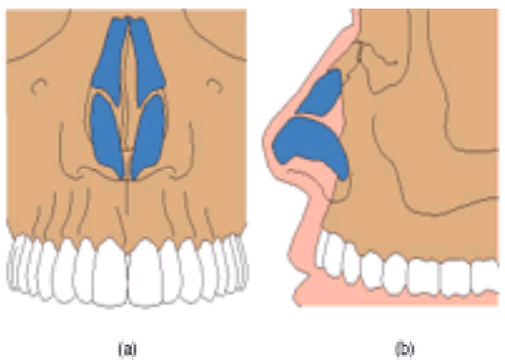


Fig. 49. The surgical anatomy of the nasal skeleton, including bones and cartilage, is reduced in rhinoplasty to alter the external appearance of the nose.

The midline dorsal prominence is reduced using a rasp or osteotome for the bone and a scalpel for the cartilage. The medial borders of the upper lateral cartilages may require resection as they also produce some dorsal prominence. The upper border of the lower lateral cartilages is resected either with the initial intranasal incision or under direct vision with the cartilages everted through another incision just under the nostril rim. The nasal bones, which have been separated from each other by the initial bone rasping, are now cut at their junction with the maxilla so that they may be inclined toward each other to narrow the nose and close the dorsal ‘roof'. The esthetic goal of rhinoplasty is the production of a balanced nose that does not appear to have had surgery (Fig. 50).



Fig. 50. Rhinoplasty is a surgical technique to change a prominent nose (a, b) to one which is in better balance with the facial features (c, d).

Post-traumatic nasal deformity is often associated with septal deviation, which may interfere with nasal breathing and alter the timbre of the voice. This septal deformity can be corrected at the same time as a rhinoplasty. Prior to the osteotomy of the nasal bones, the mucoperichondrium is elevated from both sides of the septum through an incision in the membranous septum. The deviated portion of the quadrilateral cartilage is resected or sculpted to correct its curvature. In addition, the surgeon must correct the usual concomitant deformity in the perpendicular plate of the ethmoid and the vomer to assure that this posterior portion of the septum is not left obstructing the airway. Occasionally, the inferior turbinates will need to be removed from a prominent position in the airway by outfracture or resection. Although rhinitis sicca can result from an airway made too patent, the most common complication of septal surgery is failure to clear the airway of all obstructions.

Face lift

Excess facial skin and fat, particularly in the lower face, can be removed or reduced by a face-lift procedure. Two symmetric incisions are made from the scalp to the preauricular skin to the postauricular sulcus and into the posterior scalp to allow the elevation of bilateral cheek flaps, which are then advanced posteriorly to resect an ellipse of facial and scalp skin and tighten the facial tissues along a vector parallel to the jawline. The cheek flaps are elevated in a subcutaneous plane across the

cheeks, jawline, and down the neck to the level of the hyoid. Frequently, the platysma muscle and the submuscular aponeurotic system are elevated as a separate layer in the face lift, and transposed superiorly and posteriorly to tighten the deeper tissues. A submental incision may be made specifically to excise submental fat and to excise or suture the medial borders of the platysma muscle. The supraplatysmal fat within the neck may be removed by direct excision or by suction lipectomy. The platysma flap and the skin flaps are then pulled posteriorly and superiorly under tension and sutured. The major tightening of the face lift occurs from the level of the oral commissure to the hyoid; there is little change to the nasolabial folds. The major complications of face lifting include hematoma with loss of pre- or postauricular skin and injury to branches of the facial nerve.

Blepharoplasty and brow lift

Excess skin in the upper and lower eyelids and periorbital fat can be corrected by a blepharoplasty or a blepharoplasty combined with a brow lift. The lower lid may have excess or herniated periorbital fat, even in a young person; this can be a familial characteristic that worsens with age. Young people will occasionally have a lower-lid blepharoplasty to remove this excess tissue, which will not reaccumulate until quite late in life. It is performed through a subciliary incision carried out laterally into one of the laterally radiating wrinkles. The incision can also be created through the conjunctival fornix. Through the subciliary incision the skin and orbicularis oculi muscle are elevated as a single flap. The dissection is carried down to the orbital septum, which is entered in the medial, middle, and lateral fat pockets. These fat pockets are removed back to a level of the infraorbital rim. The skin and muscle are then tailored to remove excess and the wound is closed.

The upper-lid blepharoplasty usually removes a larger ellipse of skin from the supratarsal fold superiorly. A section of orbicularis oculi muscle is also removed to enhance fixation of the supratarsal fold to the levator aponeurosis. The orbital septum is then entered and the medial and lateral pockets of periorbital fat are excised back to the level of the supraorbital rim (Fig. 51).



Fig. 51. A blepharoplasty will remove excess upper and lower eyelid skin and fat: (a) preoperative; (b) postoperative.

The major complications of blepharoplasty relate to exposure of the sclera through excessive tissue release or scar contraction. Rarely, a hematoma in the deep periorbital fat may lead to temporary or permanent blindness.

The effects of upper-lid blepharoplasty can be significantly improved by a brow lift, which also elevates the forehead tissues. Muscle excision is often performed to minimize furrows and wrinkles on the brow. The brow lift is performed through a coronal incision, a prehairline incision, or through two separate elliptical incisions just above the brow; the coronal incision is the most common. The subcutaneous tissue of the forehead is elevated superficial to the pericranium, with careful preservation of the supraorbital nerves. The corrugator muscles are resected medially just under the supratrochlear vessels and nerves. The brow is then elevated to improve the contours of the lateral upper lid and the wound closed. The major potential complication to this procedure comes from injury to the supraorbital or supratrochlear nerves and decreased forehead sensation, or from injury to the branches of the facial nerve to the frontalis muscle and an inability to elevate the forehead and eyebrow on that side.

Minimally invasive plastic surgery

Technologic advances have made possible minimally invasive procedures throughout the specialties of surgery. Small access incisions ideally minimize morbidity and cost. Endoscopically assisted procedures have revolutionized intra-abdominal, intrathoracic, and sinus surgery. The integration of endoscopic techniques into the existing armamentarium of plastic surgery began with a great wave of enthusiasm in the early 1990s. By the end of the decade, its role has evolved into an adjunctive one in certain areas. This lesser impact is due to three fundamental reasons. (1) Whereas the abdomen, thorax, and sinuses are cavities ideal for endoscopic visualization, the integument, which is closely applied, requires the creation of a cavity for endoscopy, which can be cumbersome. (2) Smaller access incisions do not have the same impact in plastic procedures as in general thoracic procedures. A small incision in the abdomen or chest avoids the otherwise necessary division of large muscle groups, while a small incision in the skin only allows a smaller visible scar and less damage to cutaneous sensation. (3) ‘Rejuvenating' cosmetic procedures almost invariably require removal of redundant skin, for example rhytidectomy and standard abdominoplasty; therefore, lengthy incisions and the resultant lengthy scars are intrinsic to these procedures.

Experience has shown that access through small incisions and the resultant smaller scars can have a positive impact. For example, muscles such as the rectus abdominis and latissimus dorsi can be harvested for free transfer to small access incisions. Similarly, cutaneous nerves and tendons can be harvested through a series of small access incisions, avoiding the donor-site scars. The most frequently performed cosmetic procedures assisted by the endoscope are the brow lift and breast augmentation accessed through the axilla.

Alloplastic facial implants

Abnormalities of facial contour can be improved by various types of alloplastic facial implant. The most common implant used is placed over the chin to correct microgenia. The implants are usually made of Silastic or other plastic materials. The alloplastic implant is placed through an incision either in the buccal sulcus or the submental area. Care must be taken to preserve sensation via the mental nerve to the lower lip. The incision is closed in multiple layers to insure healing without infection. Although the implant is placed through the mouth, infection is rare. The main complication is decreased lip sensation and long-term bone erosion. The cosmetic result is generally excellent. When combined with rhinoplasty, a major improvement in the balance of the profile can be achieved (Fig. 52).



Fig. 52. Rhinoplasty and chin augmentation together can improve the profile: (a) before; (b) after.

Prominent malar eminences can be created by the placement of malar implants. These implants can be placed in a lateral, middle, or medial position to change the contours of the cheek bones. The implants may be of a teardrop shape or specifically contoured to mimic the infraorbital rim and zygomatic arch. They are placed through an intraoral incision passing in a supraperiosteal position lateral to the infraorbital nerve and superior to the zygomaticus major muscle on to the malar prominence of the zygomatic bone. They can also be placed through a standard subciliary blepharoplasty incision. The potential complications include damage to the

infraorbital nerve and decreased sensation in the upper lip or asymmetry of the implants.

Chemical peel and dermabrasion

Indented scars on the skin, wrinkles, and other surface irregularities can be somewhat improved by dermabrasion or chemical peel. Dermabrasion is a method of abrading the outer layer of the skin, including the epidermis into the mid-dermis. After the dermabrasion the open wound is allowed to re-epithelialize spontaneously. In the course of this re-epithelialization there is an increase in the thickness of the collagen bundles within the deep dermal tissue. Dermabrasion is particularly suited to acne scars. The effect is to bevel the sharp edges of the acne scar so that a less prominent shadow is cast. The scar is not truly removed, but is altered in a way that it may be less visible. The actual abrasion is usually done with a mechanical wheel with a wire brush or a carborundum drum. The procedure may be performed under local or general anesthesia. For maximal effect it is important to carry the abrasion to the deep dermis. Of course, epithelial remnants must be left within the hair follicles or sweat glands to allow for epithelialization. After surgery the dermabraded area is bandaged and allowed to epithelialize under this bandage. The epithelium will regrow in 4 to 7 days. The area of abrasion remains quite erythematous for several months. Eventually, however, the abraded area is usually paler than the surrounding skin. The potential complications of dermabrasion include hypertrophic scars, milia, and hyperpigmentation. The dermabrasion can be repeated at a later time with some additional slight improvement in the scars.

Chemical peel is more suited to the correction of fine facial wrinkles or hyperpigmented areas. In chemical peel a 50 per cent solution of phenol or a weaker solution of trichloroacetic acid is applied to the area and then covered with an occlusive dressing. The caustic material penetrates the skin to a variable depth, causing epithelial necrosis and reorientation of the collagen structure of the dermis on healing. The occlusive dressing is left for 48 h and then removed with the necrotic outer layer of epidermis. The underlying wound is then dressed and allowed to epithelialize spontaneously. The epithelium usually regenerates in 4 to 10 days. The wound remains erythematous for 3 to 6 months. After healing, the depth and prominence of normal facial wrinkles is significantly decreased.

A by-product of the chemical peel is hypopigmentation of the skin. This effect can be used therapeutically to remove dark facial pigmentation such as cloasma. The chemical peeling solution is applied to the pigmented area and dressed in the same way. The adjacent areas of normal facial skin are not necessarily treated. The pigmented area will frequently be bleached significantly.

The complications of chemical peel include delayed healing, prolonged erythema, hypertrophic scarring, and distortion of facial features.

Laser

Laser is an acronym for *l*ight *a*mplification by the *s*timulated *e*mission of *r*adiation. Lasers are devices that produce intense beams of pure colors of light and have their surgical effect by heating tissue. The laser medium is the material used to emit the laser light and for which the laser is named. Different mediums emit characteristic colors of light, which are used for various medical applications. These mediums are chosen for their ability to react selectively with colored materials in the tissue that act as the target (chromophores) for the laser light. Anderson coined the term ‘selective photothermolysis' to describe site-specific, thermally mediated injury of microscopic, pigmented tissue targets by selectively absorbed pulses of radiation. By matching the laser medium with a specific target and manipulating its delivery, lasers have the potential to deliver heat energy with great selectivity (Fig. 53). In addition to being of one wavelength, laser light is highly collimated. A collimated beam is parallel and, therefore, does not diverge like light from an ordinary source. The parallel nature of laser light allows it to be focused into tiny spots, thereby making possible highly precise surgical manipulations. This precision has proved invaluable and revolutionary for certain ophthalmologic applications and use in other areas of surgery. Lasers have become part of the plastic surgeon's and dermatologist's clinical armamentarium. They have been shown to be the best available treatment for certain birthmarks, pigmented skin lesions, and tattoos (Table 1). Their role as a cutting tool and as a skin resurfacer is being defined (Fig. 54).

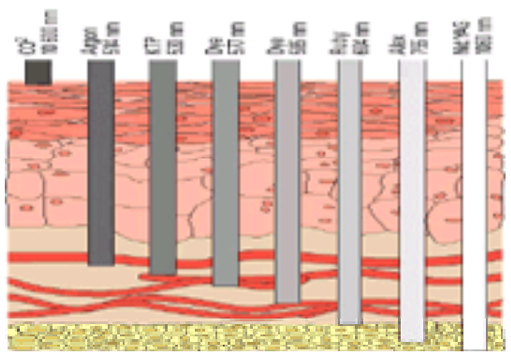


Fig. 53. Schematic diagram representing approximate level of penetration for various lasers (Nd:YAG, neodynium: yttrium-aluminum-garnet; CO_2, carbon dioxide).



Fig. 54. Professional tattoo removed with ruby (Q-switched) laser: (a) pretreatment; (b) post-treatment, five sessions (by courtesy of Joop M. Grevelink, MD, Department of Dermatology, Massachusetts General Hospital.)

Condition	*Laser*	*Wavelength (nm)*
Vascular lesions		
Superficial	Flashlamp	585
	Pumped pulsed dye	
	Copper vapor	
Deep	Argon	488, 514
	KTP	532
	Nd:YAG	1064
Tattoos		
Decorative	Q-switched ruby	694
Traumatic	Q-switched YAG	1064/532
Pigmented lesions	Q-switched Alexandrite	755
Vaporization/ surgical excision	CO_2	10 600
	Nd:YAG	1064
	KTP	532
Rhytids	CO_2	10 600
	Erbium	2940

KTP, potassium titanyl phosphate; Nd:YAG, neodymium: yttrium-aluminum-garnet.

Table 1 Laser type and wavelengths for various clinical problems

Body contouring and suction lipectomy

Excess subcutaneous tissue can be removed from several areas of the body to improve esthetic contours. The subcutaneous fat may be removed by suctioning under

the skin or, if there is also excess skin, by direct excision of fat and skin.

The most common procedure involving direct excision of fat and skin is an abdominoplasty. Excess lower abdominal fat and skin, particularly common in postpartum women, can be removed from the level of the umbilicus down to the pubis and closed primarily. The upper abdominal skin is undermined just superficial to the anterior abdominal fascia so that a large upper abdominal advancement flap is created, which is sutured at the level of the inguinal creases and the upper pubis. The umbilicus is left in its original position, but is brought out through a new opening in the center of the flap. When the abdominal flap is elevated and after the lower abdominal ellipse of skin and fat has been resected, any diastasis recti can be plicated in the midline to further tighten the waistline. This is a major surgical procedure with major blood loss. There is a risk of skin loss in the lower abdomen and appreciable scarring.

Other areas of the body that can be contoured by excision of skin and fat include the medial upper arms, the lower midback around the circumference of the hips, and the posterior lower buttocks and thighs. These procedures are less common and do produce significant scarring.

When there is excess subcutaneous tissue without lax skin, or in a patient whose skin may be expected to contract after fat removal, a solid-tube cannula can be introduced into the subcutaneous tissue and, when placed on suction close to a vacuum, the subcutaneous fat can be removed specifically without depriving the overlying skin of a blood supply. Generally this suctioning technique is carried out by passing the cannula through the subcutaneous tissue several times in slightly different directions so that the fat is removed evenly. The suctioning is at the deep premuscular fascial level so that surface irregularities are minimized. The total amount of fat removed must be monitored carefully as the blood loss can be impressive within the material removed by suction. There is much edema within the tissues after surgery. The areas most frequently treated by suction lipectomy include the lateral and posterior thighs and buttocks, the medial knee, the lower abdomen, and the submental tissue. The potential complications of this procedure include surface irregularities, decreased skin sensation, and major fascial infection.

Further reading

Anderson RR, Parrish JA. Selective photothermolysis: precise microsurgery by selective absorption of pulsed irradiation. *Science* 1983; **220**: 524–7. [Classic paper defining concepts of clinical laser utility.]

Breslow A. Thickness, cross-sectional areas, and depth of invasion in the prognosis of cutaneous melanoma. *Annals of Surgery* 1970; **172**: 902–8. [Classic article correlating tumor thickness with prognosis.]

Burget GC, Menick FJ. Nasal support and lining: the marriage of beauty and blood supply. *Plastic and Reconstructive Surgery* 1989; **84**: 189–203. [Techniques for obtaining elegant nasal reconstruction.]

Byrd HS, Spicer TE, Gerny G, III. Management of open tibial fractures. *Plastic and Reconstructive Surgery* 1985; **76**: 719–28. [One of the first to emphasize the importance of adequate debridement, fracture stabilization, and early muscle flap coverage.]

Cormack GC, Lambert BGH. A classification of fascio-cutaneous flaps according to their patterns of vascularization. *British Journal of Plastic Surgery* 1984; **37**: 80–7. [Original description of fasciocutaneous flaps and their circulation.]

Daniel RK, Taylor GI. Distant transfer of an island flap by microvascular anastomoses. A clinical technique. *Plastic and Reconstructive Surgery* 1973; **52**: 111. [Describes first successful clinical free-tissue transfer.]

Francel TJ, Van der Kolk CA, Hoopes JE, Manson PN, Yaremchuk MJ. Microvascular soft-tissue transplantation for reconstruction of acute open tibial fractures: timing of coverage and long-term functional results. *Plastic and Reconstructive Surgery*, 1992; **89**: 478–89. [Objective assessment of results.]

Gruss JS, Van Wyck L, Phillips JH, Antonyshyn O. The importance of the zygomatic arch in complex midfacial fracture repair and correction of post-traumatic orbitozygomatic deformities. *Plastic and Reconstructive Surgery* 1990; **85**: 878–90. [Describes importance of restoring a crucial three-dimensional landmark in complex facial fractures.]

Hartrampf CR, Bennett GK. Autogenous tissue reconstruction in the mastectomy patient: a critical review of 300 patients. *Annals of Surgery* 1987; **205**: 508. [Critical analysis of most effective techniques using autogenous tissue for breast reconstruction.]

Malt RA, McKhann CF. Replantation of severed arms. *Journal of the American Medical Association* 1964; **189**: 716. [Reports first successful extremity replantation.]

Manson PN *et al.* Midface fractures: advantages of immediate extended open reduction and bone grafting. *Plastic and Reconstructive Surgery* 1985; **76**: 1–10. [Revolutionized concepts for acute facial skeletal fracture repair.]

Mathes SJ, Nahai F. *Clinical applications for muscle and musculocutaneous flaps*. Mosby, St. Louis, 1982. [Revolutionized concepts of flap reconstruction.]

Yaremchuk MJ, Gruss JS, Manson PN, ed. *Rigid fixation of the craniomaxillofacial skeleton*, 1st edn. Butterworth-Heinemann, Boston, 1992. [Defines the use of rigid fixation in facial skeleton surgery.]

50.2 Craniofacial reconstruction for congenital, traumatic, and neoplastic conditions

Michael D. Poole

Introduction
History and background of craniofacial surgery
Scope of the subspecialty
Basic concepts
Preoperative preparation and investigation
Anaesthesia
Surgical exposures
Osteotomies
Bone grafts
Reconstruction of the skull base
Postoperative management
Congenital conditions
Craniosynostoses
Simple craniosynostoses
Complex syndromal SYnostoses
Clefting conditions
Encephaloceles
Other congenital conditions
Traumatic conditions
Acute craniofacial trauma
Post-traumatic deformity
Tumours
Tumour-like conditions
Benign tumours
Malignant tumours
Future of craniofacial surgery
Further reading

Introduction

History and background of craniofacial surgery

Some of the operations that now form the armamentarium of craniofacial surgery have been performed by different specialist groups for many years. Craniectomy and, to a degree, cranial remodelling in craniosynostosis, has been carried out by neurosurgeons for a long time. Similarly, major osteotomies on the facial skeleton have been undertaken by plastic and maxillofacial surgeons in the past and are nothing new: Gillies carried out a Le Fort III facial osteotomy in 1942. It is the integration of techniques in a team approach which has made craniofacial surgery a subspecialty. A craniofacial team should comprise a nucleus of craniofacial surgeon, neurosurgeon, anaesthetist, and ophthalmologist. Help should also be readily available from other specialties, such as maxillofacial surgery and otolaryngology.

This joint approach allows safer and comprehensive reconstruction of difficult deformities and defects which would previously only have been treated in a superficial way to camouflage the problem.

The development of craniofacial surgery has depended not only on teamwork, but also on a number of other advances in surgical technology. Computed tomographic imaging has made it possible for surgeons to visualize and understand better the deformities on which they can work. New knowledge of the biology of bone grafts, the introduction of miniplating systems for rigid bone fixation, and advances in anaesthesia have all added momentum to the growth of craniofacial surgery.

Scope of the subspecialty

The common thread which runs through craniofacial surgery is that it is surgery of, and adjacent to, the skull base. This interface between neurosurgery, plastic and maxillofacial surgery, otolaryngology, and ophthalmology was a relative surgical ‘no man's land' until joint efforts were made to develop techniques to tackle some of the problems involving this anatomical area. As methods have evolved many conditions in and around the orbits, the base and vault of the skull, and the facial skeleton and soft tissues have also been addressed by craniofacial surgery, and the scope of the specialty has widened. Techniques and knowledge have been applied to the management of deformities such as Treacher Collins syndrome, hemifacial microsomia, cleft lip and palate, and trauma of the orbit and face, where the skull base itself is not involved in the surgical procedures. The broad scope of modern craniofacial surgery is shown in Table 1. Three main groups of conditions are treated: congenital deformities, trauma, and tumours involving the craniofacial region.

Congenital deformities	
	Craniosynostoses
	Craniofacial clefts
	Craniofacial encephaloceles
	Other craniofacial deformities
Traumatic conditions	
	Acute craniofacial trauma
	Late secondary deformities
Neoplastic conditions	
	Tumour-like conditions
	Benign tumours
	Malignant tumours
Other problems in the craniofacial region	

Table 1 Scope of craniofacial surgery

Many conditions dealt with by craniofacial surgeons are rare. This is particularly true of the more difficult congenital abnormalities. It makes sense, therefore, for a few centres to undertake this work, so that a team's expertise and safety can be maintained; one centre for 10 to 20 million population is advisable. Many of the techniques of craniofacial surgery can also be useful to surgeons working outside craniofacial units, particularly in the areas of trauma and tumours. These will be discussed further in this chapter.

Basic concepts

Preoperative preparation and investigation

Radiographic imaging should be undertaken to define the deformity or defect and to ascertain whether intracranial abnormalities coexist. A thorough assessment of fitness for major surgery is always required, and preparation of blood for transfusion, preoperative baseline visual testing, dental assessment, and psychological preparation may also be appropriate. A team of medical and paramedical personnel who may or may not be involved in the actual surgery is needed for adequate preoperative preparation. Decisions may need to be made about the best timing for surgery. In the growing child, for example, some procedures may have a limiting effect on further growth, whereas others may promote more normal growth. The whole area of surgery on the craniofacial soft tissues and skeleton during growth is a very complex one. The psychological benefits of early surgical treatment of deformity in children, if any, are not known as yet.

Anaesthesia

This has all the hazards associated with anaesthesia for neurosurgical procedures, but other factors are often present, such as difficult airways problems. In small children blood loss can be relatively massive, and unless precautions are taken, hypothermia may occur. Craniofacial surgery is no more for the ‘occasional' anaesthetist than it is for the ‘occasional' surgeon—a high level of experience is needed to provide good operating conditions for the surgical team and to maintain the safety of the patient.

Surgical exposures

These include the bicoronal incision across the top of the scalp, other scalp incisions, and incisions in the face, for example in the lower eyelid to give exposure to the orbital floor. All of the cranio-orbital skeleton can be exposed through the bicoronal approach by turning down the forehead scalp flap, and using subperiosteal dissection as the supraorbital margin is approached (Fig. 1). By dissecting in this plane into the orbits the eyes and other orbital contents are protected within a sleeve of periosteum attached anteriorly only at the lacrimal sac at its junction with the nasolacrimal duct, but the orbital skeleton is widely exposed. The nasal and zygomatic areas can be similarly exposed by subperiosteal dissection. Laterally, care needs to be taken to avoid facial nerve branches, particularly the frontal branch. The supraorbital nerves are preserved intact by their careful removal from any bony tunnel or notch. Exposure of the skull base within the cranium is achieved by an appropriately placed craniotomy followed by extradural dissection to delineate the area of bone required.

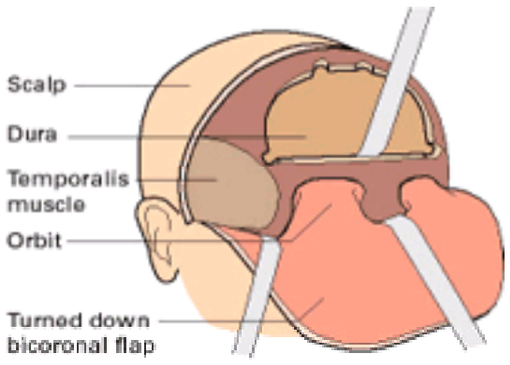


Fig. 1. Diagram of the exposure used in craniofacial surgery. The scalp is turned down forwards after making a bicoronal incision across the top of the head.

Osteotomies

These are performed by making cuts in the bones of the craniofacial skeleton and moving the separated segment to the desired position where it is fixed. In order to carry out osteotomies on the skull base and adjacent areas safely, good exposure of the bone is needed, with gentle retraction of orbital contents and brain and dura away from the area in question.

Bone grafts

These are frequently used in craniofacial procedures to repair defects left by the movement of skeletal segments or after excision of bone. The most readily available donor site for bone grafts is the cranium itself. Following a craniotomy, the bone flap is split through the diploe and bone grafts used from the inner table. If no craniotomy is needed cranial bone grafts can be taken from the outer table. This donor site is painless by comparison with the iliac crest or rib grafts, and there is some evidence that grafts taken from bone which develops in membrane, such as cranium, will undergo less postoperative resorption than grafts taken from bone of endochondral origin. If insufficient bone is available from the cranium because it is too thin, or if a large amount is required, bone grafts from other sites may also be needed. Bone grafts and mobilized skeletal segments are fixed into position by wires or miniplates and screws. When metallic miniplates and screws are used on the outer surface of the infant skull for fixation, with growth of the skull these tend to migrate through deeper into the bone and the screws can protrude on the inner surface of the skull through the dura. This has resulted in metal plates being largely abandoned in infant craniofacial surgery. Bioabsorbable plates and screws are currently under investigation.

In infants under 1 year of age large craniectomies usually reossify quite readily from the osteogenic dura, and grafts are not needed to fill defects.

In the future, harvesting of bone grafts from the cranium or other parts of the body may not be needed if current work on tissue engineering to produce bone bears fruit. The use of a matrix for bone deposition and growth factors to promote osteogenesis appears to have considerable potential.

Reconstruction of the skull base

Bone and soft tissue are used to form a barrier if the nasal cavity, nasopharynx, or paranasal sinuses are opened into communication with the cranium. Absence of such a barrier is associated with a risk of ascending infection into the intracranial cavity, early or late after surgery. Bone grafts are obtained as above. Soft tissues can take the form of free pericranial grafts, vascularized flaps of local tissues such as galea or temporalis muscle and fascia, or tissues brought to the area from elsewhere in the body and vascularized by microvascular anastomoses.

Postoperative management

Careful monitoring of blood and fluid losses is required: replacement is particularly important in small children. Other management takes the form of general supportive measures and airways management, antibiotic cover, and possibly anticonvulsant prophylaxis if the cranium has been opened, and after considerable brain retraction, steroids to minimize cerebral oedema. Appropriate intensive care facilities, ready access to computed tomography scanning in the event of neurosurgical complications, and the easy availability of team members should they be needed, are all essential for safe postoperative care of these patients.

The potential problems which can develop during and after surgery are listed in Table 2. Many of these can be fatal unless they are attended to promptly. Even with constant vigilance by experienced teams there is occasional mortality of about 1 per cent and an incidence of serious complications of about 10 per cent if all operations in the scope of craniofacial surgery are included. Some groups of patients have a higher risk of complications: this is particularly true of those undergoing radical resections of large tumours around the skull base, and some children with complex SYndromal craniofacial deformities such as Apert's syndrome.

Airways obstruction
Cerebral oedema
Intracranial haemorrhage
Hydrocephalus
Brain damage
Epilepsy
Meningitis
Brain abscess
Bone infection
Cerebrospinal leak
Diabetes insipidus
Visual loss
Diplopia
Unresolved or worse deformity

Table 2 Some potential complications of craniofacial operations

Congenital conditions

Craniosynostoses

These conditions are the result of premature closure (synostosis) of the sutures between bones in the cranial vault or base by ossification. Depending on the site of SYnostosis, a deformity of the skull often occurs as the growing brain pushes on the bones, which can still grow at the sutures. So-called ‘simple' craniosynostoses involve only one or two sutures in the vault; complex cases involve multiple sutures in the vault and skull base. There is a genetic background to many of these conditions, in particular the syndromal complex craniofacial SYnostoses (sometimes called craniofacial dysostoses) and the fibroblast growth factor receptor group of genes are often involved. These syndromes often include deformities of the limbs, such as SYndactyly in Apert's syndrome. The exact biochemical mechanisms underlying expression of the phenotype in these syndromes is, to date, unknown. There appears to be a relationship between the behaviour of cranial sutures and the underlying dura which is not understood as yet. (Table 3)

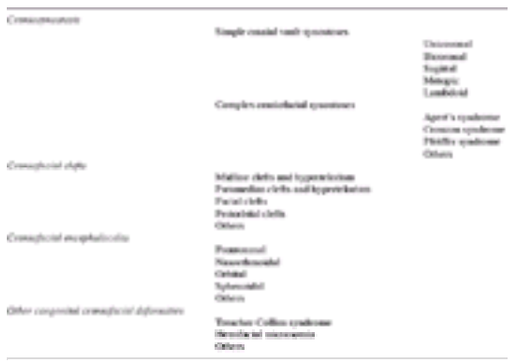


Table 3 Congenital craniofacial malformations

The chief clinical features of craniosynostoses are deformity, and raised intracranial pressure, which affects a significant number of infants with these problems. Mental development in some conditions such as Apert's syndrome may be delayed for reasons as yet not fully understood, as well as due to any effects of raised intracranial pressure. Using sophisticated scanning techniques alterations in cerebral perfusion underlying craniosynostotic areas of the skull have been demonstrated. The significance of these is not yet delineated.

Treatment is initially directed at early release of the brain from restriction of growth by craniectomy or cranial remodelling. This lowers raised intracranial pressure. Such procedures should be performed during the first year of life for maximum benefit. Simple craniectomies are not usually undertaken; these have been replaced by more complex operations to remodel the cranium and attempt to increase the intracranial volume. Secondary procedures may be required for the correction of residual or recurrent deformities.

Simple craniosynostoses

These conditions affect only one or two of the cranial sutures. Unimpaired growth at normal unfused sutures causes a cranial deformity to develop with growth. The deformity is usually characteristic, depending on which suture is fused (Fig. 2). In only a small proportion of children is intracranial pressure increased. There may be imbalance of the extraocular muscles in association with an orbital deformity in some children. The growth of the mid-face and lower face is usually normal. Untreated, the deformity will persist into adult life.

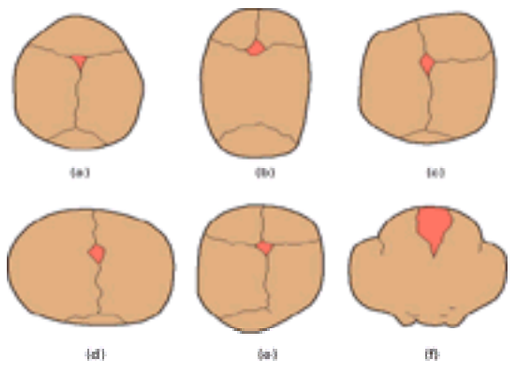


Fig. 2. Different cranial deformities, viewed from above with the forehead at the top, which occur according to which suture(s) is synostosed. (a) Metopic synostosis; (b) sagittal synostosis; (c) left coronal SYnostosis; (d) bilateral coronal synostosis; (e) right lambdoid SYnostosis; (f) multiple suture involvement as in Apert's SYndrome.

Treatment is best carried out during the first year of life if possible, and is aimed at excising the fused suture and correcting any cranial deformity which may be present. If the sagittal suture is affected operation within the first 2 months of life is worthwhile: simple craniectomy is usually sufficient treatment at this stage. If, however, the coronal sutures are affected on one or both sides surgery needs to be delayed. Not only is the suture excised but the forehead is remodelled and advanced by removal of the frontal bone and supraorbital bony skeleton and refixing these skeletal segments in a symmetrical, advanced position. Such surgery is carried out through a bicoronal incision. In small infants this can result in considerable operative blood loss and loss of heat.

Correction of simple craniosynostoses during the first year of life usually allows satisfactory subsequent growth of the cranium, and no further treatment is needed in the majority of cases (Fig. 3(a),Fig. 3(b)). There has been a considerable amount of confusion over the last few years about posterior plagiocephaly, that is, unilateral flattening of one side of the occiput, and the differential diagnosis of this and the place of lambdoid suture synostosis in this deformity. Occipital plagiocephaly has been far more common in recent years since recommendations that infants sleep on their back to lessen incidence of sudden infant death. This is very rarely associated with lambdoid SYnostosis and if there is doubt, radiological imaging of the sutures will usually give the answer. Even if lambdoid synostosis is present, surgical treatment is not always required.

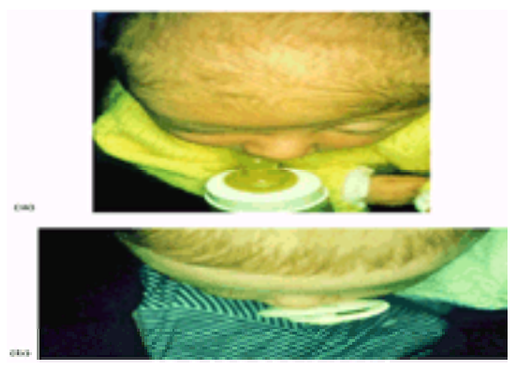


Fig. 3. (a) A child's head with unilateral coronal synostosis prior to correction, viewed from above. The forehead deformity is seen. (b) The same child postoperatively.

Complex syndromal synostoses

These conditions are usually caused by genetic abnormalities and many of these involve mutations in the fibroblast growth factor receptor family of genes. The craniosynostosis can involve several cranial sutures. As well as the cranial vault sutures being involved, those in the skull base and the growth centres between the cranial base and the facial skeleton are often affected. This results in deformity of the mid-face as well as the cranium. Shallow orbital cavities result in proptosis of the eyes, underdeveloped nasal passages and nasopharynx cause airways problems and a dish face deformity is associated with malocclusion of the teeth. Typical examples are Apert's and Crouzon's syndromes which are autosomal dominant conditions (Fig. 4). The incidence of raised intracranial pressure is higher in these syndromes than in simple craniosynostoses. A proportion of patients with some conditions, such as Apert's syndrome, have mental retardation.



Fig. 4. A child with the typical facial deformity of Apert's syndrome.

In most of these conditions the cranial part of the synostosis affects both coronal sutures. Surgical correction in infancy aims to advance the frontal bone and supraorbital area as far as possible, (usually about 2 cm) through a bicoronal incision. This has the effect of releasing the fused coronal sutures, correcting the recessed frontal bone and supraorbital region, and increasing the anteroposterior dimensions of the anterior cranial fossa. It may also increase the intracranial volume. Whether early surgery improves the potential for mental development is not known. It does, however, lower raised intracranial pressure to normal levels. As a high intracranial pressure during development and growth of the brain carries a poor outlook for mental development, early surgery would appear to be worthwhile. It has added benefits of improving the appearance of the anterior cranium and supraorbital region.

The recessed mid-facial skeleton often needs to be advanced at a later date, although this is sometimes undertaken in young children if there are pressing indications, such as excessive proptosis such that the eyelids do not close over the eyes causing exposure, inadequate nasal airways for nasal breathing, or cosmetic deformity caused by the dish face. If the surgery is being performed simply for cosmetic reasons it is usually left until late childhood.

Facial advancement is usually undertaken with osteotomy at the Le Fort III level, which is a separation of the facial skeleton from the cranial base (Fig. 5). It can be combined with frontal advancement as a frontofacial monoblock advancement procedure, but this has a high incidence of complications such as infection. The procedure is, therefore, often split into two separate operations: one to advance the forehead and a second to advance the face. In this way communication between the intracranial cavity and the nasal cavity and nasopharynx can be avoided and the complication rate lowered. Facial advancement in children needs to be repeated, or some further orthognathic procedure needs to be carried out at the end of growth, to obtain optimal dental occlusion. This further surgery, however, may not be at the Le Fort III level and a lower level advancement, such as a Le Fort I procedure, may be sufficient. Repeating facial advancement at the same level is made difficult by the scarring of the previous operation.

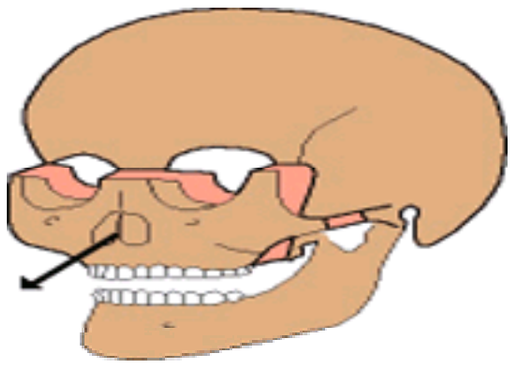


Fig. 5. A diagrammatic representation of a mid-facial advancement at the le Fort III level. All the osteotomies can be performed through the bicoronal exposure.

Facial advancement in children is a major operation, with considerable blood loss; postoperative airways complications are always to be feared. These procedures are usually carried out in specialized centres (Fig. 6(a),Fig. 6(b)). The main indications to undertake facial advancement in young children is airways obstruction or severe exophthalmos such that the eyelids cannot close over the eyeballs. The airways problems that these children may have can cause obstructive sleep apnoea and be difficult to manage. Tracheostomy or facial advancement, or both, may be required. In the future, it is likely that the magnitude of such procedures can be lessened by the use of distraction osteogenesis such as has been used in long bone lengthening in orthopaedics. These techniques can be applied to the craniofacial skeleton and have been used for the purposes of advancing the mid-face (or mandible) with suitably designed appliances which are implanted on the skull and osteotomies carried out, and the advancement subsequent to that is done gradually over a period of time. Much of the blood loss and the need for harvesting bone grafts can thus be avoided.



Fig. 6. (a) A patient with Crouzon syndrome before mid-facial advancement. (b) The same patient after le Fort III advancement.

Many children with craniofacial synostosis syndromes have abnormalities of the hands and feet. Complex syndactyly present in Apert's syndrome for example frequently needs to be corrected in a number of staged operations. Abnormal nails on the feet may also need surgical attention. These children also need attention from orthoptists and ophthalmic surgeons for squints caused by extraocular muscle imbalance, from orthodontists to correct malocclusion and other dental problems, and management by otolaryngologists for airways and ear abnormalities. Children, most of all, need the expertise of a whole treatment team.

Clefting conditions

A craniofacial cleft can be defined as a zone of dysplasia or an abnormality of development producing a zone of hypoplasia or absence of the skeleton or the soft tissues, or both. Clefts are usually surrounded by hypoplastic tissues. The most typical cleft is that of the lip and palate. Craniofacial clefts, however, tend to occur at a higher level in the craniofacial skeleton. Tessier (1976) has classified these by relating them to the orbit anatomically, and numbering them as they progress periorbitally. This classification SYstem, while not related to the developmental fault behind the deformity, is useful because it describes the anatomy of the cleft quite accurately.

With growth, the deformity of patients with a craniofacial cleft tends to remain the same, although in some there is the potential for the condition to worsen.

The most typical craniofacial cleft is the group of mid-line or paramedian clefting SYndromes which result in hypertelorism, or widening of the distance between the orbits with a wide, abnormal nose in the middle of the face. This condition was one of those to receive early attention in the development of craniofacial surgery because of the serious deformity. It is possible to reduce the wide distance between the orbits because, although the orbits are angulated away from one another in a lateral direction, their apices usually show a fairly normal separation. The anterior parts of the orbits containing the globes can therefore be rotated closer together by appropriate osteotomies and resection of the central bone and tissues. The apices of the orbits containing the optic nerves, the third, fourth, and sixth cranial nerves, and the origins of the extraocular rectus muscles can be left undisturbed. This surgery is undertaken through a bicoronal incision: the orbits can be moved either by a periorbital osteotomy and movement of the orbit as a ‘box', or by facial bipartition, in which the orbits and hemifacial skeletons are moved as single units closer to each other with resection of a wedge-shaped piece of the central facial skeleton to narrow the nose. Nasal reconstruction is frequently necessary in these patients (Fig. 7 and Fig. 8).

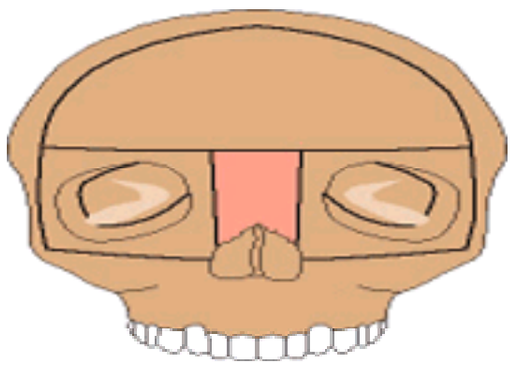


Fig. 7. Diagram of the osteotomies performed to move the functional orbits closer together for hypertelorism. In this procedure the orbits are moved as two ‘boxes' after the pink area is resected.

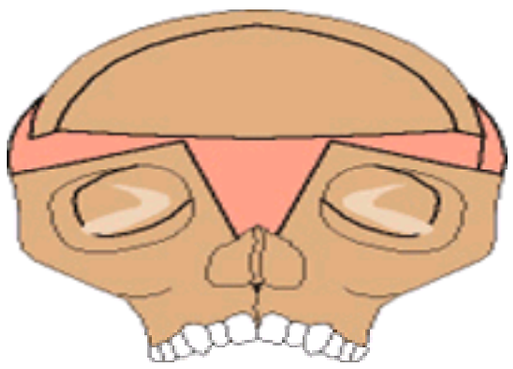


Fig. 8. Diagram of the facial bipartition operation. The two hemifacial skeletal segments are moved together, so narrowing the interorbital distance.

Many clefting syndromes, in particular hypertelorism, are corrected for aesthetic rather than functional reasons (Fig. 9(a),Fig. 9(b)). Visual disturbances such as the lack of stereoscopic vision are not improved by such surgery. If clefts involve the eyelids, surgery can help with cover of the eye and does, therefore, have functional implications.



Fig. 9. (a) A child with a mid-line ‘cleft' of the face causing hypertelorism and a deformed nose. (b) The child after surgery using facial bipartition.

Operations to correct hypertelorism or orbital dystopia are major operations in children and, because of the rarity of these conditions, should be carried out in specialized units. In such units the incidence of serious complications is about 10 to 12 per cent, and there is a mortality rate of around 1 per cent. Major surgery around the orbital skeleton has potential complications of visual loss, postoperative extraocular muscle disturbance, infection of bone which may lead to loss of skeletal segments, and major intracranial problems such as infection or haemorrhage.

Assessments of the aesthetic outcome of such operations have rarely been undertaken. The few studies that have been published indicate that such surgery is worthwhile, but that these patients still look deformed after their operation and cannot usually be made normal. The psychological benefits of undertaking surgery early in childhood have not been investigated adequately.

Encephaloceles

Encephaloceles are hernias of cerebral tissue and the covering meninges through a bony defect in the cranium. Around the craniofacial region this defect can occur between the frontal and nasal bones, between the nasal bones and ethmoids, or through the skull base further posteriorly. Encephaloceles can be corrected by craniofacial exposure and reconstruction with excision of the usually abnormal, herniated brain, and repair of the dura and cranial base or cranium and of the overlying soft tissues. If carried out within the first year of life, subsequent growth of many of these children is satisfactory and nothing further needs to be done (Fig. 10(a),Fig. 10(b)). Some patients present later and with extensive deformity due to distorted growth of the affected area of the craniofacial region. These need more major repositioning of the orbital or facial skeleton in addition to treatment of the encephalocele itself.



Fig. 10. (a) A baby with a large frontonasal encephalocele. (b) The same child years after early removal of the encephalocele and repair.

Other congenital conditions

There are other congenital abnormalities of the craniofacial region which may not need joint neurosurgical/transcranial procedures for their correction, but which benefit considerably from the techniques of craniofacial surgery. These include Treacher Collins syndrome and hemifacial microsomia, both of which affect the mid and lower facial skeleton and soft tissues.

The facial skeleton in Treacher Collins syndrome is hypoplastic or cleft in the area of the malar bones (Fig. 11). This can be repaired during childhood using vascularized or free bone grafts to augment the skeleton in those areas. Other soft tissue components of the deformity such as the colobomas of the lower eyelid may need correction simultaneously. The complex upper and lower jaw deformity which exists in these patients can also be corrected, but best results are obtained if this is left until adolescence.

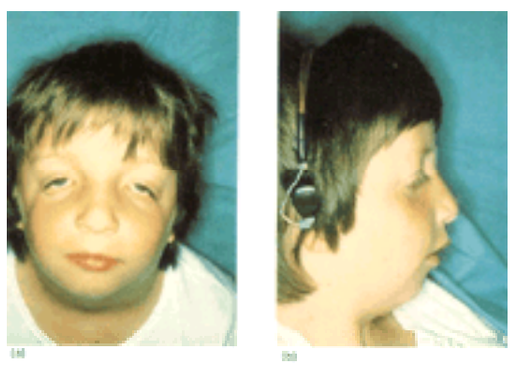


Fig. 11. (a) An unoperated patient with Treacher Collins SYndrome—front view showing the malar hypoplasia. (b) Profile showing the underdeveloped mandible.

Hemifacial microsomia is an asymmetrical hypoplasia of the face centred round the structures derived from the first and second branchial arches. It usually consists of hypoplasia of the external ear, of the vertical ramus of the mandible, and of the temporomandibular joint on the affected side; the overlying soft tissues may also be hypoplastic. Surgery can be carried out during childhood to improve the appearance but will need to be repeated in adolescence (Fig. 12(a),Fig. 12(b)). Skeletal surgery on the mandible and maxilla is usually necessary, many patients requiring soft tissue augmentation of the cheek tissues, as well as some kind of reconstruction of the abnormal pinna. This can be done using autogenous cartilage graft from the costal margin; alternatively, a prosthetic ear can be fitted to osteo-integrated implants in the bone.



Fig. 12. (a) Hemifacial microsomia affecting the left side. (b) The same child after surgery to the mandible and cheek soft tissues.

As already mentioned, distraction osteogenesis is useful in the craniofacial skeleton for the lengthening of bony structures, in particular the mandible, and as has already been mentioned can be applied to the mid-face as well. Mandibular distraction osteogenesis is now well established and in the future will play a major part in the management of hemifacial microsomia and conditions such as Treacher Collins syndrome and other conditions involving retrognathia. Relatively minimal surgery followed by distraction will produce lengthening of the bone to the required extent.

Traumatic conditions

Acute craniofacial trauma

Patients with acute craniofacial trauma benefit considerably from the application of craniofacial surgical principles in their management. Procedures carried out jointly by a neurosurgeon and a reconstructive surgeon will improve the postoperative appearance, lessen the need for secondary surgery, result in a lower incidence of dangerous late intracranial infections, and shorten the patient's stay in hospital.

Patients with craniofacial trauma need the usual early attention to airways and respiratory problems. Any acute intracranial complications such as haematomas need to be dealt with urgently and the soft tissues closed. The craniofacial fractures can be treated several days later, when the acute injury to the brain has had some time to settle down, and when there has been time to assess the injury properly with radiographs and computed tomography scans, and full ophthalmic and dental assessment. When all of this information is available a joint procedure can be planned to repair all the fractures at a single procedure and also repair all the basal dural tears and any associated soft tissue problems, such as telecanthus caused by detachment of the medial canthal ligaments from their bony anchorages.

First, the fractures all need to be exposed. Much of this work can be done through a bicoronal incision, avoiding the creation of surgical scars on the face. It may, however, be necessary to use subciliary incisions in the lower eyelids to gain access to the orbital floor and infraorbital margins. This is advisable in trauma patients to avoid excessive retraction of the injured orbital contents. Craniotomy is usually required for repair of the dural tears in the anterior cranial fossa; the craniotomy flap can be split and the inner layer used as bone graft material for the reconstruction. Dural tears are delineated. If the dura cannot be directly sutured, it is patched with pericranium: this can be done extradurally without having to make wide intradural exposures. Reconstruction of bone may be needed in the anterior cranial fossa, the

supraorbital region, the nasal or nasoethmoidal region, the malar areas, and orbital floors. When the posterior wall of the frontal sinus is extensively fractured the whole frontal sinus is cranialized by removal of the posterior walls and all the mucosa, with plugging of the frontonasal ducts using cancellous bone from the diploe of the craniotomy bone flap. Other facial injuries such as those of soft tissues, and maxillary and mandibular fractures, if present, are then attended to in the same procedure (Fig. 13(a),Fig. 13(b)).



Fig. 13. (a) Traumatic right orbital dystopia as a result of periorbital fractures. (b) After repair in the semiacute phase.

Post-traumatic deformity

The late management of untreated deformities following craniofacial trauma is difficult. The displaced bones and soft tissue tend to become fixed by scar tissue in the deformed position, and correction at a late stage is much more difficult than it is acutely when there isno scar tissue. While considerable improvements can be made in some patients with late corrections the results are never satisfactory as when early reconstructive surgery is practised. The surgery is dif-ficult and has a higher complication rate and is therefore best carried out in specialized centres.

Tumours

A variety of tumours and tumour-like conditions affecting the craniofacial region are amenable to surgical resection and reconstruction of the resulting defects using the principles of craniofacial exposure and reconstruction (Table 4).

Tumour-like conditions		
	Bone dysplasias	
	Neurofibromatosis	
	Vascular anomalies	
Benign tumours		
	Bone	
	Soft tissues	
		Epithelial
		Connective tissues
Malignant tumours		
	Bone	
	Soft tissues	
		Epithelial
		Connective tissues

Table 4 Craniofacial tumours

Tumour-like conditions

Bone dysplasias, in particular fibrous dysplasia, are the most common of these. Fibrous dysplasia results in an enlargement of areas of the skeleton, sometimes confined to the craniofacial region or sometimes polyostotic in nature. The enlargement of bone can be progressive or can cease and begin again at a later date. In the craniofacial region fibrous dysplasia usually affects either the maxilla or the supraorbital region. Ideally, surgical correction is aimed at removal of all the affected bone but sometimes this is so extensive that total removal is impossible. Areas of bone which cannot be removed and replaced by reconstruction can then only be managed by shaving and sculpturing procedures. Sometimes, however, the entire affected area can be removed and replaced by bone graft.

Indications for surgery in fibrous dysplasia are deformity, pain, or encroachment of bony enlargement on foramens such as the optic foramen. If the latter occurs bone needs to be removed adjacent to the optic canal so that the optic nerve can be decompressed. The results in terms of improvement in appearance can be very well worthwhile in these patients (Fig. 14(a),Fig. 14(b)), and visual deterioration caused by optic nerve decompression can often be arrested.



Fig. 14. (a) Cranio-orbital involvement by fibrous dysplasia of bone. (b) The patient after resection and reconstruction of the affected areas of bone.

The other major tumour-like condition is neurofibromatosis affecting the cranio-orbital region. This is a difficult condition to treat, and surgery carries a high complication rate. Although the results are not particularly satisfactory in aesthetic terms (Fig. 15(a),Fig. 15(b)), some improvement can be made. Typically, cranio-orbital neurofibromatosis involves a partial absence of the wings of the sphenoid bone, which allows the orbital contents to pulsate because of the unrestrained pressure of the intracranial contents behind through the defect in the posterior orbit. This results in an increasing proptosis, and eventually in loss of vision. There may also be infiltration of the soft tissues of the lateral part of the upper lid and other periorbital areas by neurofibromatous tissue, causing grotesque deformity. As well as repair of the missing bone at the back of the orbit, these soft tissue enlargements may need attention.



Fig. 15. (a) A patient with left cranio-orbital involvement by neurofibromatosis. (b) The patient after two staged repair procedures showing the limited improvement possible.

Vascular anomalies are tumour-like conditions which can also be difficult to treat surgically, but in some cases surgery may be well worthwhile. Operative procedures can be made easier and safer in some vascular lesion patients by preliminary embolization via radiologically placed intravascular catheters to reduce blood supply. Total removal of the lesion is the idea, but in many cases this is not realistic.

Benign tumours

One of the most common benign tumours in the craniofacial region is intraosseous meningioma of the sphenoid wing. This tumour usually presents with proptosis. The involved bone can be removed using craniofacial exposures and the resulting defect replaced by a bone graft (Fig. 16(a),Fig. 16(b)). A wide variety of other benign tumours occur around the anterior cranial base and orbit: these can be resected and reconstructed, usually with quite acceptable results in terms of aesthetic outcome (Fig. 17(a),Fig. 17(b)), recurrence rate, and complication rates.

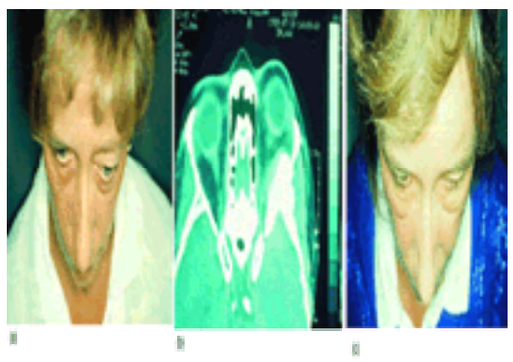


Fig. 16. (a) A patient presenting with proptosis as a result of an intraosseous meningioma in the greater wing of the sphenoid. (b) An axial computed tomography scan through the orbits showing the tumour in the bone and the adjacent soft tissues of the orbit and meninges. (c) The patient's proptosis corrected by removal of the tumour.

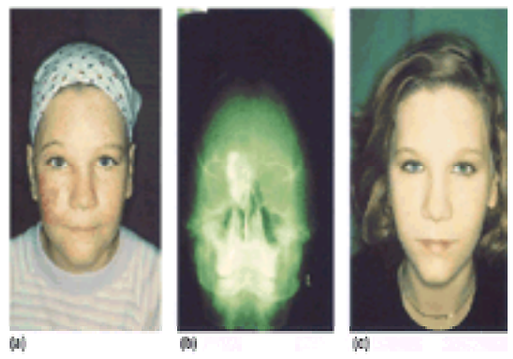


Fig. 17. (a) A young patient with a tumour in the right medial brow area. The port wine stain is coincidental. (b) Plain radiograph shows an osteoma in the frontal sinus. (c) Years after removal of the tumour and reconstruction.

Malignant tumours

Both epithelial and connective tissue malignancies can affect the craniofacial area. Some of these are amenable to resection using craniofacial exposures and techniques. The large defects created the need to be repaired by bony and soft tissue reconstruction of the skull base, cranium, facial skeleton, and associated soft tissues.

Because of the size of the resulting defects and the importance of separating the intracranial contents from the nasopharynx, nasal cavity, and air sinuses, flap transfers are frequently necessary, and free flap surgery using microvascular anastomosis aids considerably in the reconstruction these patients. The nature of these tumours and their significant recurrence rate means that such surgery may only be palliative, and its effects on the patient's quality of life need to be borne in mind when planning treatment. However, if surgery can be carried out without major postoperative disability it is often worthwhile, even if the patient is not cured of the tumour. Fungating masses can be removed, odour and pain alleviated, and quality of life improved in some incurable patients.

Full preoperative assessment of the extent of these lesions using modern radiological techniques is necessary to predict the required extent of the resection and the likely outcome. The histological type of the tumour has a lot of influence on the outcome. For example, aesthesioneuroblastomas have a relatively good prognosis by comparison with squamous cell carcinomas, adenocarcinomas, or connective tissue malignancies.

Attempted excision of craniofacial tumours is contraindicated when the major cerebral vessels would need to be sacrificed, where the only remaining seeing eye would need to be sacrificed, where the tumour is extensively involving the cavernous sinus, or where it would appear for other reasons that attempted extirpation of the tumour is going to disable the patient to the extent that postoperative quality of life would be severely impaired. Management of these tumours requires extensive preoperative discussion with the patients and their families.

In general, about 50 per cent of patients with advanced malignant tumours of the anterior skull base of sufficient extent to require a craniofacial procedure will survive and remain tumour-free postoperatively. The complication rate of these operations is quite high, however, at about 20 per cent.

Future of craniofacial surgery

It is likely that with more long-term studies, and more information on the psychological effects of surgery for craniofacial deformities, surgeons will be able to predict better which patients are most likely to benefit. Refinements in techniques, improvements in ease and safety, and a better understanding of the nature of the deformities will improve outcome. Concentration of craniofacial surgery into centralized units gives the opportunity for basic research into the cause of some of these

problems, and will hopefully provide some answers.

Some of the advances in knowledge and techniques that have been mentioned in this chapter will influence the future development of craniofacial surgery a lot. For example, rapidly increasing information about the genetic basis of many of the craniosynostosis syndromes and continuing study of the pathogenesis of these deformities, together with further advances in imaging will provide those caring for patients with these problems with a lot of much needed knowledge. Further development of techniques such as distraction osteogenesis, minimally invasive surgery, and tissue engineering, which are developing areas in reconstructive surgery will also play a part.

Further reading

Cohen MM, ed. *Craniosynostosis. Diagnosis, evaluation, and management*. Raven Press, New York, 1986. [A comprehensive textbook devoted to the subject of caniosynostosis.]

Gorlin R. Fibroblast growth factors, their receptors and receptor disorders. *Journal of Cranio-maxillo-facial Surgery 19*97, **25**, 69–79. [A review of the status of fibroblast growth factors and their receptors as related to craniosynostoses.]

Marchac D, Renier D. *Craniofacial surgery for craniosynostosis*. Little, Brown and Company, Boston, 1982. [A textbook on surgical treatment of craniosynostosis.]

Marchac D, Renier D, Broumand S. Timing of treatment for craniosynostosis and faciocraniosynostosis: a 20 year experience. *British Journal of Plastic Surgery 19*94, **47**, 211–22. [A paper looking at the long-term follow-up of children treated surgically for craniosynostoses.]

Mulliken J, Warman M. Molecular genetics and craniofacial surgery. *Plastic and Reconstructive Surgery 19*96, **97**, 666–75. [A review of what is known of molecular genetics and craniofacial surgery in 1996.]

Poole MD. Complications in craniofacial surgery. *British Journal of Plastic Surgery 19*88**; 41**:608–13. [A review of complications in craniofacial surgery.]

Poole MD, Briggs M. Cranio-orbital trauma: a team approach to management. *Annals of the Royal College of Surgeons England 19*89; **71**: 187–94. [An outline of the team approach to trauma management in the craniofacial region.]

Salyer KE. *Techniques in aesthetic craniofacial surgery*. JB Lippincott, Philadelphia, 1989. [A small textbook on surgical techniques in craniofacial surgery.]

Shah J, Kraus D, Bilsky M, Gutin P, Harrison L, Strong E. Craniofacial resection for malignant tumours involving the anterior skull base. *Archives of Otolaryngology, Head and Neck Surgery 19*97; **123**: 1312–17. [A review of craniofacial resection for malignant tumours in the anterior skull base and results.]

Stricker M, Van der Meulen J, Raphael B, Mazzola R, eds. *Craniofacial malformations*. Churchill Livingstone, Edinburgh, 1990. [A large and comprehensive textbook on craniofacial malformations.]

Tavakoli K, Stewart K, Poole MD. Distraction osteogenesis in craniofacial surgery: A review. *Annals of Plastic Surgery 19*98; **40**: 88–90. [A review of the status of distraction osteogenesis in craniofacial surgery.]

Tessier P. Anatomical classification of facial, cranio-facial and later-facial clefts. *Journal of Maxillo-facial Surgery 19*76; **4**: 69–92. [The original description of the Tessier Classification of craniofacial clefts.]

50.3 Acute burns

Robert L. Sheridan and Ronald G. Tompkins

Overview
Classification of burns
Burn depth
Burn extent
Special injuries and illnesses
Initial assessment and treatment
Assessment
Initial therapy
Escharotomies and fasciotomies
Transfer considerations
Further treatment
Electrolyte and other abnormalities
Pigment nephropathy
Inhalation injuries
Definitive surgical strategy
Metabolic issues and nutrition
Changes in the overall metabolic rate
Nutritional requirements
Route of nutritional support
Wound care and infections
Topical agents
Prophylactic antibiotics
Treatment of invasive infection
Complications
Pulmonary
Gastrointestinal and biliary
Renal
Cardiovascular
Neurologic
Burn aftercare
Further reading

Overview

Despite growing emphasis on prevention, 2 million people are burned, 80 000 are admitted to hospital, and 6500 lose their lives each year in the United States of America. Of those who survive serious injuries, many suffer long-term disability. Survival of the seriously burned has improved over the past few decades. In the 1950s and early 1960s, survival from a burn injury involving 30 per cent of the body surface area at any age was almost unprecedented. Today, children survive burns affecting more than 90 per cent of the body surface area, and young and middle-aged adults with more than 70 per cent burns survive routinely. These dramatic improvements have come from the recognition that destroyed and damaged tissues must be removed promptly after the injury and that the wound must be physiologically closed immediately thereafter. As the size of the burn injury increases, surgical strategies must be altered to accomplish the goal of tissue excision and immediate wound closure.

Classification of burns

The direct injury caused by a burn produces an initial volume of irreversibly damaged tissue, surrounding which is a region of tissue that has suffered reversible injury. The final depth and extent of the injury depends upon the degree to which these adjacent, sublethally injured tissues recover: their survival is influenced by the host's physiologic response, which includes tissue edema and ischemia, the host's inflammatory response, and by environmental factors, including infections and the efficacy of topical treatment. Modification of this process is an exciting area of investigation. Determining the depth and extent of the wound requires frequent observation by a knowledgeable clinician over the first 3 to 5 days following injury so that the damage can be assessed accurately; accurate estimation of depth is important in treatment and prognosis.

When a burn wound heals spontaneously the dead tissue separates as new epithelium covers the injured area. When burns are superficial, regeneration occurs rapidly from uninjured epidermal elements, hair follicles, and sweat glands within the injured tissue; since dermis remains intact, little scarring results unless infection occurs. When burns are deep, however, the epidermis and much of the dermis are destroyed, and epithelialization starts from the edges of the wound or from the scattered integumental remains. Because this process is exceptionally slow, excessive granulation tissue forms before the wound can be covered by epidermis. Wounds such as these eventually contract dramatically; unless skin grafting is prompt, disfiguring or disabling scars result.

Burn depth

Grading of a burn as first, second, third, or fourth degree refers to the depth of the irreversible tissue damage (Fig. 1). The depth of injury can be determined from the appearance of the wound (Table 1). Distinguishing first-degree and superficial second-degree injuries is usually simple; operative intervention is rarely required. A distinction between deep second-degree and third-degree burns can often not be made until after 3 to 5 days of observation, but in either case surgery is the optimal treatment. Fourth-degree burns are readily apparent upon observation; they are always treated surgically.

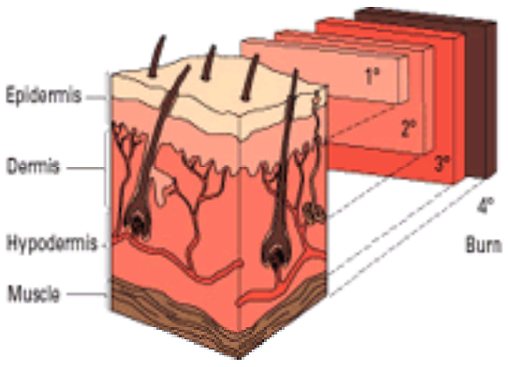


Fig. 1. Schematic demonstrating the correlation of depth-of-injury nomenclature with anatomic structures of the skin.

First degree
Red, erythematous
Very sensitive to touch
Very painful
Usually moist
No blisters
Surface markedly and widely blanches to light pressure
Second degree
Erythematous or whitish with a fibrinous exudate
Wound base is sensitive to touch
Painful
Commonly have blisters
Surface may blanch to pressure
Third degree
Surface may be:
White and pliable
Black, charred, and leathery
Pale and mistaken for normal skin
Bright red from hemoglobin fixed in the subdermis
Generally anesthetic or hypoesthetic
Subdermal vessels do not blanch
No blisters
Hair easily pulled from its follicle
Fourth degree
Involves deep tissues including fascia, muscle, bone, and tendons

Table 1 Selected characteristics suggesting depth of injury

Elderly individuals and young children require additional consideration since they have thin skin and are therefore more likely to suffer full-thickness injuries when exposed to burning agents that would create only partial-thickness injuries in others. This point is particularly important when assessing scald burns, which occur relatively frequently in toddlers. These indeterminate injuries are usually deeper than is first apparent, and this becomes obvious after 3 to 5 days of observation.

Burn extent

The overall extent of the burn injury is best estimated by recording the affected areas on a burn diagram and estimating the percentage of the body surface area using the ‘rules of nines' in the adult (Table 2) or using the Lund–Browder chart in children (Fig. 2). Special consideration is necessary in children, in whom the head forms a much greater percentage and the lower extremities a lower percentage of the total body surface area than in adults.

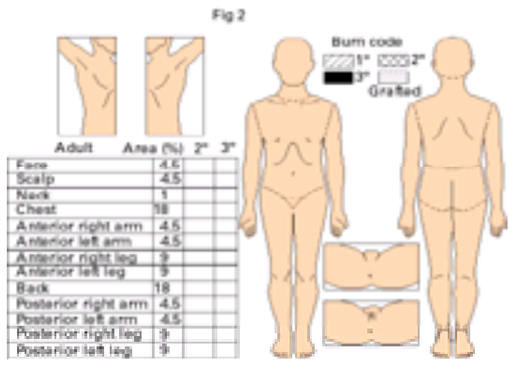


Fig. 2. Schematic that allows documentation of the areas of the body surface involved in the cutaneous injury and a convenient table to determine the total extent of the body surface involved.

	Body surface area	*(%)*
Head and neck		9
Face and neck	4.5	
Scalp	4.5	
Arm, forearm, and hand[a]		18
Anterior arm, forearm, hand	4.5	
Posterior arm, forearm, hand	4.5	
Volar hand	1	
Leg and foot[b]		18
Anterior	9	
Posterior	9	
Anterior trunk		18
Chest	9	
Abdomen	9	
Posterior trunk		18
Upper back	9	
Lower back[c]	9	
Perineum		1

[a] Hand in adults.
[b] Includes skin fold. [c] Includes inferior gluteal fold. [d] Includes buttocks.

Table 2 Rule of nines*

The physiologic impact of the injury or the severity of the burn can be judged from the quantity of tissue involved in the injury (Table 3). This assessment is represented by the percentage of the body surface area burned and by the depth of the burn. A small injury will generally require little treatment and often does not greatly alter the cardiovascular and metabolic state. Moderate and large injuries require resuscitation of the patient and have potentially dire cardiovascular and metabolic consequences.

Body surface area* (%)	Classification	Metabolic and cardiovascular response
< 15	Small	+
15–49	Moderate	++
50–89	Large	+++
≥ 70	Massive	+++

*Refers to cutaneous involvement by second- and third-degree injury.

Table 3 Classification of burns

Special injuries and illnesses

Electrical burns occur when an electric current passes through the body; there may be severe internal damage even when the most obvious cutaneous manifestations are only small wounds of entrance and exit. The two major additional effects of electrical injury are the effect of the current upon normal cardiac electrical conductivity and the fact that electric current tends to cause more damage to deep than to superficial tissues. Continuous cardiac monitoring is mandatory for at least the first 24 h after an electrical burn. Damage to deep tissues can be assessed by frequent observation by knowledgeable observers and ultimately by sequential operative debridement of devitalized tissues in the wound and in adjacent subfascial compartments.

Chemicals continue to destroy tissue while they remain in contact with it; the best treatment for chemical injuries is to flush the surface immediately with copious amounts of water and/or saline to remove or dilute the agent. Powdered chemicals may be brushed from the skin and then flushed from the body surface. Chemicals in contact with the eyes should be thoroughly irrigated with clean water or saline if available. Following such exposure, an evaluation by an ophthalmologist is required.

Chemical agents designed for warfare that require further consideration include nerve agents (Tabun, Sarin, Soman, and VX), blistering agents (mustard and nitrogen mustard), choking agents (phosgene), blood poison or anoxia agents (hydrocyanic acid and cyanogen chloride), and phosphorus. Nerve agents are volatile liquid organophosphates that may be inhaled or absorbed through the skin at ambient temperatures. They act primarily by inhibiting acetylcholinesterase at cholinergic receptor sites and result in cholinergic (muscarinic and nicotinic) hyperactivity without serious skin injury. Their actions can be blocked by atropine, pyridostigmine bromide (Mestinon), and anticonvulsants. Choking agents are intended to be inhaled and produce irritation of the upper airway, bronchospasm, and bronchorrhea without skin injury. The anoxia agents are also inhaled and interfere with oxygen utilization by the tissues; they do not cause cutaneous injury, but lead to deep coma and death within minutes. Other warfare agents, such as mustard, nitrogen mustard, and phosphorus, burn skin: exposure should be treated by copious water flushes to dilute and remove the agent.

A number of diseases result in difficult wounds and organ failures. These conditions include toxic epidermal necrolysis, or Stevens-Johnson syndrome, and purpura fulminans; such patients are optimally managed in burn units.

Initial assessment and treatment

Assessment

The initial approach to management of the patient with a major burn is similar to that for any other patient experiencing major trauma, as suggested by the American College of Surgeons in the Advanced Trauma Life-Support Course. Assessment begins with a primary survey including evaluation of the A, B, C of resuscitation (*A*irway with immobilization of the cervical spine, *B*reathing, and *C*ircula-tion). Later, a complete secondary survey should be conducted from head to toe to identify associated injuries: although the burn may often be the most obvious injury, other serious and life-threatening injuries can also be present. The mechanism of inury and past medical history should be elicited.

Primary survey

As mucosal edema evolves in those with extensive body burns or inhalation injury, obstruction of the upper airway can develop rapidly, particularly in small children (Table 4). If the airway is adequate and the patient is breathing, administration of oxygen by facemask or by nasal cannulas may be sufficient. However, if the patient is not breathing well because of either an obstructed upper airway or a depressed level of consciousness, endotracheal intubation is necessary (Fig. 3). Emergency tracheostomy should be avoided unless it is absolutely necessary. In those with potential injury to the cervical spine, all manipulations of the airway must be performed while maintaining immobilization of the head and neck in the axis of the body until the cervical spine has been evaluated fully.

A B C
Provide an adequate Airway with C-spine immobilization
Provide sufficient Breathing including ventilation if necessary
Evaluate Circulation considering clinical signs and symptoms of shock and temporizable obvious bleeding
Determine the presence of associated life-threatening conditions (e.g. obstructed airway, cardiac tamponade, tension pneumothorax, etc.)
Initiate fluid resuscitation by large-bore venous cannulas
Remove jewellery and clothing
Place urinary and nasogastric catheters
Order appropriate radiographic and laboratory evaluations

*Activities necessary as the initial approach to the seriously injured patient.

Table 4 Primary survey*



Fig. 3. The soft-tissue edema associated with a large burn can compromise the airway. It is essential that the airway be guarded by early endotracheal intubation, as it may become progressively more difficult to control it as edema progresses.

As part of the primary survey, adequacy of the peripheral circulation should be determined early after the burn injury by palpation of peripheral pulses. All the patient's clothing should promptly be removed, allowing rapid determination of the extent of injury and adequacy of peripheral circulation while stopping the burning process by removing smouldering or chemical-laden clothes. Rings, watches, and other jewellery should be removed from any injured limbs: constriction of an oedematous extremity by jewellery can result in distal ischemia.

Fluid is transferred to the area of injury from the vascular space within minutes of thermal injury. This fluid loss may result in hypovolemic shock; the greater the surface area involved, the more extensive the hypovolemia. Initial volume resuscitation should begin immediately, following the introduction of at least two large-bore intravenous catheters—ideally into unburned upper or lower extremities, but through burned tissues if necessary. In hospitals equipped to manage complications of central venous cannulation, placement of a central venous catheter is preferred when extremities are circumferentially burned; ‘cut-downs' should be avoided unless absolutely necessary. Since patients with moderate burns often have an ileus that persists beyond the early days of treatment, a nasogastric tube should be used to decompress the bowel. A bladder catheter is necessary to allow measurement of urine output accurately.

A reasonable estimate of the rate of fluid administration may be obtained from the formula:

2–4 ml of lactated Ringer's solution × body wt in kg × the per cent body surface burn.

Of this volume, 50 per cent must be given in the first 8 h following the injury (*note*: this first 8-h period begins at the time of injury, not with the initiation of resuscitation). The remaining fluid is given over the next 16 h. In this calculation, the percentage body surface area burned includes the sum of partial- and full-thickness injury: the fluid requirement in the second 24 h is usually approximately 50 per cent of the calculated resuscitation volume. Monitoring of vital signs is required to maintain an adequate urine output of at least 0.5 ml/kg per hour (1.0 ml/kg per hour in young children). This guideline provides an initial estimate of fluid replacement, but only close monitoring during initial treatment can direct the treatment of an individual patient.

Secondary survey

The secondary survey begins with a detailed physical examination from head to toe, which aims to detect any associated injuries (Table 5). A history is taken during the examination, and details of the accident should be obtained, including information about the burning agent, where the injury happened (closed versus open space), the possibility of smoke inhalation, exposure to noxious chemicals, and the possibility of any related trauma. Any history of pre-existing illnesses such as diabetes, hypertension, and cardiac or renal disease should be elicited. Use of regular medications, alcohol, or drugs should be identified. Allergies or other sensitivities should be identified, and a history of tetanus immunizations should be obtained.

Obtain a medical history
Perform a head-to-toe physical examination
Record a detailed burn diagram
Administer analgesics intravenously
Administer tetanus prophylaxis
Cover the wounds
Evaluate and perform indicated escharotomies and fasciotomies
Consider transfer to other facility as necessary

*Activities that follow the primary survey in order to evaluate the burned patient completely and to develop a treatment plan.

Table 5 Secondary survey*

Initial therapy

Following both the primary and secondary surveys, definitive resuscitation therapy should be instituted. Morphine may be used for pain management, but full-thickness burns are painless because nerve fibers have been destroyed. Narcotics and other parenteral medications that are administered should be given intravenously, not intramuscularly, since drugs are poorly absorbed from muscular tissues during hypovolemic shock, but they may be absorbed rapidly once circulation is restored.

Patients who have undergone previous active immunization for tetanus and have had a toxoid booster injection within 1 year of the injury require no further prophylaxis unless they have an associated tetanus-prone wound, when they should receive a toxoid booster injection. Patients who received their most recent booster injection more than 1 year before injury should be given a booster dose of toxoid; those who have not received prior immunization should be given human antitetanus globulin and an immunizing dose of toxoid, followed by the usual immunizing dosage schedule of toxoid established for nonimmunized people.

Initially the burn wound should simply be covered with a clean, dry sheet with no creams or other medications. Covering the wound excludes air from partial-thickness injuries and may diminish pain. Application of ice is neither useful nor recommended. Initial laboratory studies that should be undertaken include the hematocrit, glucose, electrolytes, blood urea nitrogen, urinalysis, and chest radiograph. In patients with large burns or possible inhalation injuries, arterial blood gases, an electrocardiogram, and a measurement of blood carboxyhemoglobin should also be obtained.

Escharotomies and fasciotomies

Edema may develop underneath circumferential burns of extremities and compromise the arterial circulation to the more distal aspects. Early after the injury, the adequacy of the peripheral circulation can usually be assessed by palpation of the peripheral pulses, but these pulses frequently become impossible to identify as edema develops under a circumferential eschar. Increased compartment pressures can completely obstruct arterial inflow, leading to distal ischemia, necrosis, and gangrene. Signs and symptoms of peripheral ischemia can be difficult to identify in patients with large burns, who are often intubated, receiving narcotics, and have peripheral edema due to administration of resuscitation fluid. The classic signs and symptoms of peripheral ischemia (pain, paraesthesias, pallor, pulselessness, and paralysis) may therefore be masked, and Doppler ultrasonography is the only reliable method for its early detection. When vascular compromise occurs, escharotomies (incisions made through burned epidermis and dermis) are necessary to restore both arterial and venous circulation (Fig. 4).



Fig. 4. Escharotomies must be performed promptly to avoid ischemic injury; this requires continuous monitoring of blood flow to burned extremities.

Circumferential chest burns may similarly restrict excursion of the chest wall and minute alveolar ventilation; escharotomies of the chest wall may be required when excessive inspiratory pressure is required to ventilate the lungs. Escharotomies should be carried out in these patients after consultation with regional burn care centers.

Fasciotomy is usually not necessary, except in patients with very deep burns in the extremities and burns from electrical injuries. The volar compartments of the forearm and anterior, lateral, and posterior compartments of the leg are the important regions to consider for decompression by fasciotomy. The clinical indication for fasciotomy is similar to that for escharotomy, but the confining tissue is the deep fascia and not the eschar. Escharotomy will often decompress the compartmental pressures sufficiently to obviate the need for fasciotomy. The use of devices to determine the need for fasciotomy based upon compartmental pressures is investigational and the decision to perform a fasciotomy should be made on clinical grounds.

Transfer considerations

About 85 per cent of burns are small and can be treated in outpatient facilities. Patients with more severe burns require admission and consideration for treatment in a specialized burn facility after initial assessment and treatment has been provided. The types of injuries that should be referred to a specialized treatment center are shown in Table 6.

1. Second and third-degree burns greater than 10% body surface area (BSA) in patients under 10 or over 50 years of age
2. Second and third-degree burns greater than 20% BSA in patients between 10 and 50 years of age
3. Second and third-degree burns with serious threat to functional and cosmetic impairment that involve the face, hands, feet, genitalia, perineum, and other major joints
4. Third-degree burns greater than 5% BSA
5. Specialized injuries such as electrical burns, including lightning and chemical burns, with serious threat of functional or cosmetic impairment
6. Significant inhalation injuries
7. Circumferential burns of the extremities or the chest
8. Pre-existing medical disorders that complicate management, prolong recovery, or affect mortality
9. Child patients in hospitals without qualified personnel or equipment for the care of children
10. Concomitant trauma in which the burn injury poses the greatest risk of mortality; however, if the trauma poses the greater immediate risk, the patient should receive treatment in the initial trauma center

*Criteria established by the American Burn Association that identify those requiring care in a facility frequently visited for such patients.

Table 6 Transfer considerations

Further treatment

Electrolyte and other abnormalities

Since most cardiac arrhythmias are caused by hypovolemia, hypoxia, acidosis, or hyperkalemia, these metabolic disorders should be sought and corrected in addition to the use of drugs to rectify cardiac problems as necessary. Patients with pre-existing cardiovascular and renal disease present special problems. Fluid, electrolyte, and colloid administration should be limited to amounts sufficient to produce a urinary output of 0.5 ml/kg per hour in adults and 1 ml/kg per hour in children. Pulse, blood pressure, temperature, the electrocardiogram, and arterial blood gases should be monitored, particularly in elderly patients and in those with moderate or larger injuries.

Hypokalemia is common in the early resuscitation period: many patients present with depleted K^+ stores secondary to prior diuretic therapy and K^+ is not generally included in the early vigorous fluid replacement. Serum potassium should be maintained above 4 mEq/l. If 0.5 per cent silver nitrate topical solution is used as an antibacterial agent, one should remember that this hypotonic solution leeches electrolytes from the tissues into the wet dressings; additional supplementation of these electrolytes may be necessary to avoid severe hyponatremia, hypochloremia, and hypochloremic alkalosis.

Hypoalbuminemia results from a combination of the dilutional effects of crystalloid replacement therapy and enhanced loss of protein into the subeschar edema fluid. Parenteral administration of colloid solution should be continued throughout the resuscitation period to maintain albumin at or above 1.5 g/dl. Since most calcium in the serum is reversibly bound to albumin, hypocalcemia may be a result of hypoalbuminemia. Although the ionized fraction of serum calcium is usually normal, it should be measured. Replacement quantities of calcium, phosphate, and magnesium may be given daily.

Generalized poor tissue perfusion resulting from hypovolemia or cardiac failure causes metabolic acidosis. A blood pH of less than 7.2 should prompt consideration of treatment with intravenous sodium bicarbonate; the underlying cause of the metabolic acidosis must be identified and treated vigorously. Focal poor tissue perfusion can result from constricting eschar or fascia; it should be treated surgically. Metabolic acidosis can also result from the inhibition of carbonic anhydrase activity by topical sulfamylon; the contribution of this to the overall metabolic acidosis can be dramatic.

Because thermally injured skin loses core body heat more rapidly than normal skin, maintenance of body temperature is important. Hypothermia below 97°F (36°C) can be associated with poor peripheral perfusion and cardiac arrhythmias. Rewarming patients from temperatures below 91°F (33°C) can be associated with fatal arrhythmias; these patients should be warmed slowly and monitored continuously.

Pigment nephropathy

Myoglobinuria can result from ischemic constriction of muscle, deep thermal injury, or electrical burns of muscle. It should be treated by alkalinization of the urine with sodium bicarbonate, with frequent monitoring of both serum and urinary pH to allow adjustments to the treatment; the goal is a urinary pH above 8. Intravenous mannitol may be needed to promote osmotic diuresis if myoglobinuria is severe. Hemoglobinuria may result from erythrocyte destruction following burns; its treatment is identical to that for myoglobinuria. Both abnormalities may result in renal tubular necrosis if not promptly and accurately managed (Fig. 5).

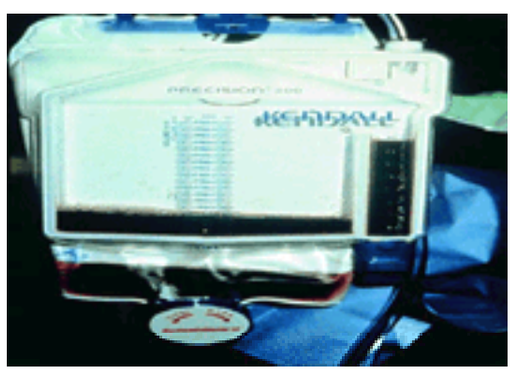


Fig. 5. Pigment nephropathy will occur if pigmenturia is not promptly recognized and treated.

Inhalation injuries

Injuries of the respiratory tract form a heterogeneous group of pulmonary complications of burns and they increase the risk of mortality by 20 per cent for a given cutaneous injury. Respiratory injuries are clinically divided into asphyxia, pharyngeal–glottic, and upper airway injuries (upper airway obstruction), and respiratory bronchiolar–alveolar injury (gas-exchange difficulties).

Asphyxia seriously contributes to mortality from burn injuries: there are 3000 to 4000 deaths from carbon monoxide poisoning or asphyxia in the United States each year, and the majority of these patients die at the scene of the accident. Death occurs from the decreased availability of hemoglobin for oxygen binding and, therefore, a decreased oxygen-carrying capacity and oxygen delivery. The target organs are the brain and myocardium; victims present with myocardial ischemia and neurologic depression. Cyanide may also contribute to asphyxia and should be considered in patients with a history of exposure to burning plastics. Selected, hemodynamically stable patients with serious exposure to carbon monoxide may benefit from hyperbaric oxygen therapy if it can be safely administered.

Glottic and airway injuries may be present in up to one-third of burn patients and may result from a combination of thermal and chemical damage. These injuries have a variable presentation: onset of airway edema and resultant obstruction is occasionally sudden, but an insidious and delayed onset of respiratory complications, including tracheobronchial infection, is more common. Few gases have sufficient heat capacity to produce thermal damage below the level of the trachea and direct thermal injury usually only affects the uppermost airway. The exceptions to this are superheated steam and explosions occurring within the airway itself.

The diagnosis of inhalation injury is usually made by clinical suspicion and bronchoscopy (Fig. 6); chest radiographs and arterial blood gases may initially be normal. Inhalation injury can be confirmed by ^{133}Xe lung scans and by pulmonary function testing. The treatment of acute pulmonary injuries and their respiratory complications is entirely symptomatic: little is known about the mechanism of injury or repair. The consensus is against the continuous use of corticosteroids or prophylactic antibiotics for inhalation injuries in the early postburn period. Lung-protective mechanical ventilation optimizes outcome, supported by chest physio-therapy, toilet bronchoscopy, and vigilant surveillance for evolving pulmonary infection.

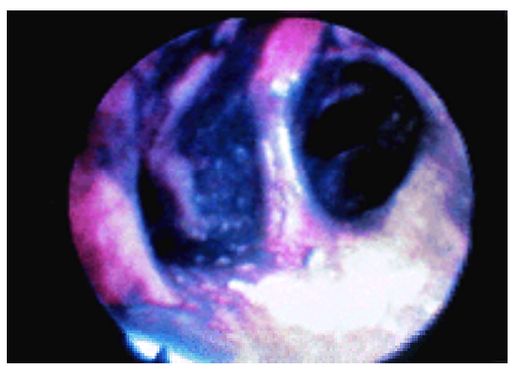


Fig. 6. Bronchoscopic findings in those with inhalation injury include carbonaceous debris, mucosal pallor, and ulceration.

Definitive surgical strategy

Prompt excision and immediate wound closure removes devitalized tissue and replaces it with either autografts or acceptable skin substitutes such as allografts or artificial skin. Prompt excision and immediate closure consists of the surgical removal of all areas of the burn (both deep dermal and full-thickness injury) that will require more than 3 weeks to heal spontaneously. Staged surgical procedures begin as soon as possible after injury and are continued until all devitalized tissue (eschar) is excised. In massive burn injuries the first surgical procedure includes harvesting of all available donor sites, excising the largest burned areas that have the highest likelihood of successful graft take, and immediately closing the excised area with autografts. In patients with very large injuries, autograft is quickly exhausted and wounds are then closed with human allograft or Integra© (Fig. 7). Excision and autografting of the back is often chosen as the first procedure because this graft shows reliable take and greatly simplifies the patient's clinical care once a graft has taken. The physical state of the resuscitated patient does not recover after the injury but steadily deteriorates until wound closure is complete. Therefore the operative procedures should begin as soon after injury as possible, when the patient is physiologically closer to normal: extensive operative procedures performed in the first week after injury are frequently safer than those undertaken in the third week. It is equally important to harvest the autograft donor sites soon after injury. Since donor sites heal and may be reharvested in 2 weeks, the sites harvested on the first day after injury can usually be reharvested on about day 14. This plan minimizes both the time required for autograft closure and length of hospital stay.



Fig. 7. In those patients with very large injuries, autograft is quickly exhausted and wounds may then be closed with human allograft or Integra®. Integra®, shown here, consists of an inner layer of bovine collagen and an outer layer of silicone. The silicone is later replaced with a thin autograft.

Metabolic issues and nutrition

There are two critical problems for the nutritional management of an injured patient. First, what level of caloric intake will be required to allow them to meet the caloric needs in response to their injury? Second, what mixture of substrates (proteins, carbohydrates, and fats) will meet their caloric requirements and optimally satisfy metabolic needs without incurring a negative nitrogen balance?

Changes in the overall metabolic rate

Oxygen-consumption studies have shown that the metabolic rate of burn patients does not exceed twice the patient's normal basal metabolic rate as predicted by the Harris–Benedict table of correlations for adults and children above 2 years old. Usually, the metabolic rate does not exceed a 50 per cent increase above the normal basal rate. When wounds are promptly excised and closed, the contribution of burn size to hypermetabolism is small and for clinical purposes can be neglected in determining the daily caloric requirements of burn patients. However, nutritional support of the seriously burned should be closely monitored, ideally with regular indirect calorimetry and nitrogen balance studies, to ensure proper provision of substrate to support the expected hypermetabolic state.

Nutritional requirements

The total caloric requirement is met by the provision of carbohydrate, protein, and fat. A carbohydrate delivery rate of 5 mg/kg per minute provides enough calories to minimize the utilization of amino acids as an energy source and approximates the maximum rate of glucose oxidation for an injured patient remaining in bed. Additional exogenous glucose delivered to a patient at rest is not used for ATP production but is simply converted to fat. At the infusion rate of 5 mg/kg per minute, the respiratory quotient is just below unity, indicating that glucose is being oxidized to CO_2, H_2O, and energy and is not being stored as fat.

Protein should be infused at a rate of 1.5 to 2.5 g/kg per day, depending on the size of injury and the presence of sepsis. This rate will maintain a positive nitrogen balance in adults and in children, but neither the exact protein requirements nor the optimal mixture of amino acids required by seriously injured patients are known. Unfortunately, studies of nitrogen balance do not produce the exact information necessary to determine the quantity and composition of the proteins required. Until rates of synthesis of muscle protein, collagen components of host defense, and other proteins can be accurately measured *in vivo*, this protein-replacement rate remains only an estimate.

Calculating the caloric equivalent received by a seriously burned patient given glucose at 5 mg/kg per minute and protein at 2.5 g/kg per day shows that their caloric requirement (calculated as basal metabolic rate × 2) is not achieved. Fats are therefore given to meet the remaining caloric requirement, via either the enteral or parenteral routes. These fats supply calories and minimize the need for mobilization of endogenous proteins for energy and gluconeogenesis; they also provide essential fatty acids. Parenteral lipids in the form of intravenous fat emulsions should be administered with caution in burn patients: this has been associated with lung dysfunction and thrombocytopenia, particularly in infants and children.

Route of nutritional support

Early, aggressive metabolic support with amino acids begins within hours after the injury in patients with burns affecting greater than 20 per cent of the body surface area, pre-existing malnutrition, patients with complications such as sepsis or associated injuries, and those admitted late after a burn and who have already lost more than 10 per cent of their premorbid weight. Nutritional support is begun as soon as possible after the injury, but not later than 24 h after completion of the immediate postburn resuscitation phase or within 12 h after injury. This support is accomplished by oral feeding, enteral feeding, or total parenteral nutrition, depending on the patient's condition. The enteral route is preferred since it is associated with fewer complications and is cheaper than parenteral feeding. However, anorexia, facial burns, or dysphagia may make oral feeding difficult or impossible and enteral feedings may need to be given via tubes. If gastric emptying is delayed but intestinal absorption is normal, continuous enteral feeding can be given through postpyloric feeding tubes. Total parenteral nutrition is used if intestinal motility or absorption are abnormal or if multiple operative procedures are anticipated over the first weeks of the admission, inhibiting normal gastrointestinal absorptive function.

Wound care and infections

Infection of the burn wound is a major cause of complications and death in burn patients: the best approach to the problem is the prevention of wound infection. Infection is most likely to affect a large, open wound containing necrotic tissue; susceptibility is increased by the lowered host resistance that results from serious trauma, and this is more important than the virulence of most infecting bacteria in determining the seriousness of the infection. Decreased host resistance must be corrected or prevented. Necrotic tissue must be removed and wounds properly closed. Secondary derangements in physiology and metabolism leading to caloric and protein starvation must be corrected. Cross-contamination of wounds must be prevented, and antibiotic treatment to prevent invasive infections should be administered only at times of increased risk.

Prompt excision of necrotic tissue and immediate closure of the wound with skin graft is essential in preventing infection. Protection of the wound from cross-contamination is promoted by maintaining the patient in a isolated environment and by following universal precautions. An additional and important adjunct is the use of topical antibacterial agents such as 0.5 per cent silver nitrate, 10 per cent mafenide acetate (sulfamylon), or silver sulfadiazine to reduce bacterial colonization of the burn wound.

Topical agents

Before application of a topical agent, all grease, oil, loose skin, burned clothing, and other contaminants should be removed. The advantages and disadvantages of

commonly used topical agents are shown in Table 7. If 0.5 per cent silver nitrate is to be used, a thick layer of wide-mesh gauze dressing is placed on the wound and saturated with the agent, which not only markedly decreases bacterial growth on the burn wound, but also minimizes evaporative water loss, thereby conserving energy. Silver nitrate is not used on the perineum or face because of the potential for ingestion and likelihood of staining of mucous membranes. For these areas, silver sulfadiazine has been used. Dressings should be soaked with the solution every 2 h to maintain a 0.5 per cent concentration at the wound surface. A lower concentration is not bacteriostatic; a higher concentration can damage viable skin. Rarely, methemoglobinemia can be produced when nitrates are reduced by certain unusual strains of *Escherichia coli* and *Klebsiella* in wounds.

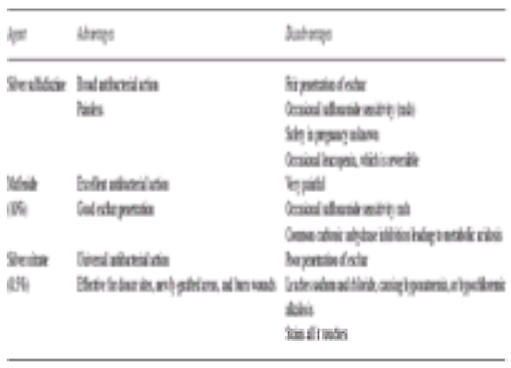


Table 7 Commonly used topical wound agents

The two other topical agents frequently used are creams of sulfamylon and silver sulfadiazine. They are applied twice daily directly to the burn wound after the initial removal of wound contaminants. All previously applied sulfamylon or silver sulfadiazine must be removed before a fresh layer is applied. As previously mentioned, sulfamylon is a carbonic anhydrase inhibitor and can cause a severe metabolic acidosis. Allergic reactions and hemolysis may develop in patients deficient in glucose 6-phosphate treated with either sulfamylon or silver sulfadiazine.

Prophylactic antibiotics

Routine administration of systemic antibiotics throughout the burn illness does not prevent burn wound infection and should not be done. The use of antibiotics for indiscriminate periods in an effort to avoid infection is an ineffective and dangerous approach that alters the normal flora and causes allergic reactions, renal, otic, or other organ injury, and wound colonization by resistant bacterial strains. Nevertheless, antibiotic administration can, in specific instances and for short intervals, supplement the patient's natural ability to fight invasive infection. Antibiotics are of considerable importance for two general indications: in prevention of infection during specific periods of reduced host resistance and in established infections.

Prophylactic antibiotics may be useful against the high incidence of bacteremia that occurs during and after excision of colonized burn eschar. Treatment should begin immediately before the operation and last through the immediate postoperative period, until normal cardiovascular hemodynamics are restored (usually within 24 h) and other normal physiologic signs return. The perioperative antibiotic given should be chosen on the results of previous cultures from the burn wound and the sensitivities of the organisms. If these are unavailable, general antimicrobial coverage for both Gram-positive cocci and Gram-negative rods is recommended. Intravenous antibiotics should be directed toward the commonly encountered *Staphylococcus aureus, Pseudomonas aeruginosa, E. coli, Enterobacter, Klebsiella, Acinetobacter*, and *Proteus* spp. Serum concentrations of both vancomycin and aminoglycoside should be measured when these antibiotics are used perioperatively but continued for more than 48 to 72 h.

Treatment of invasive infection

Identification of invasive infections requiring antibiotic treatment is often difficult because burn wounds are not normally sterile and, therefore, bacteriologic testing of the wound will not alone distinguish between colonization and invasive infection requiring treatment. Once a diagnosis of invasive infection has been established, however, prompt SYstemic antibiotic therapy is required. Indications for antibiotic treatment in burn patients are outlined in Table 8.

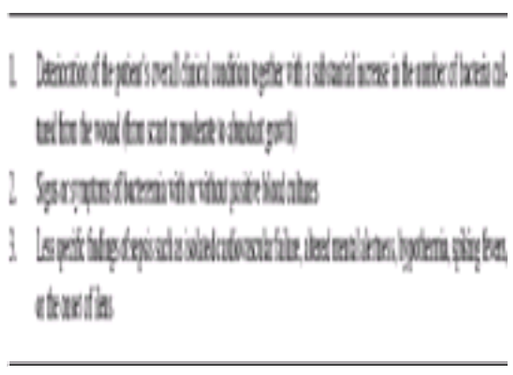


Table 8 Indications for antibiotics in burn patients

The best indicator of invasive infection of a burn wound is histologic evidence of bacterial invasion of viable tissue at the base of the wound. Biopsy should include the viable wound base, and the specimen should be subjected to grinding and bacterial quantitation, as well as to routine histomorphologic examination, including the proper stains for bacteria and yeast. Quantitative cultures from burn wounds and histologic evaluation are difficult to perform in the course of routine burn care and are not usually useful clinically.

The infecting organism can usually be predicted by referring to the most recent surveillance cultures of the burn wound, urine, and respiratory-tract secretions. In large burn injuries, surveillance cultures of body fluids should be taken regularly (at least weekly): the last strain cultured from the burn wound is often responsible forthe invasive infection. More definitive information can be obtained from blood cultures obtained at the first clinical indication of invasive infection.

Pathogenic Gram-negative bacteria, especially *E. coli* and *Klebsiella* and *Enterobacter* spp., are usually responsible for invasive infections. These organisms are best managed by individual sensitivity testing of the particular isolate. If the pathogen cannot be identified, both an aminoglycoside and penicillinase-resistant penicillin should be given to allow wide-spectrum coverage until more specific antibiotic therapy can be designed. In patients with large, open wounds, and especially in children, drug leakage from the wound greatly complicates accurate therapy based on body size, and serum should be monitored to ensure that therapeutic drug concentrations are maintained.

Complications

Pulmonary

Bronchogenic infections or pneumonia occur frequently, usually accompany inhalation injury, and are commonly due to the organism colonizing the burn wound. Prophylactic corticosteroid therapy is detrimental, and preventive antibiotics are probably ineffective after inhalation injury. Daily sputum cultures are appropriate in the susceptible patient and dictate the choice of antibiotic if pneumonia does occur. Attention to pulmonary therapy and toilet is also indicated.

The adult respiratory distress SYndrome occurs frequently in thermally injured patients, but is particularly difficult to distinguish from inhalation injury. In addition, cardiogenic pulmonary edema, bronchopneumonia, and severe tracheobronchial infection need to be excluded. The typical chest radiographic findings and pulmonary gas-exchange abnormalities usually confirm the diagnosis in the absence of significant inhalation injury and infectious processes. Treatment is supportive, as in other

critically ill patients with associated organ failures. Pulmonary toilet is particularly important in these patients.

Pulmonary edema may occur following inhalation injury or fluid overload, particularly in patients lacking cardiac reserve. Close attention should be given to arterial O_2 tension, respiratory mechanics, and the administered fluid load. Intubation, positive end-expiratory pressure, and pharmacologic intervention may become necessary. When pulmonary edema occurs immediately after the burn, there is probably extensive parenchymal chemical damage and the prognosis is guarded. Atelectatic changes and pneumonitis may also occur. Atelectasis and aspiration can be prevented and treated as in other patients admitted to hospital.

Gastrointestinal and biliary

Curling first noted the association between bleeding duodenal ulcers and burn injury in 1842. The incidence of diagnosed gastric or duodenal ulceration in burn patients was about 10 per cent in 1970; however, ulcer-related complications have markedly decreased in the last decade, probably due to the advent of continuous tube feedings and exacting control of gastric pH. The pathophysiology of the initial mucosal injury appears to be related to mucosal hypoxia, which increases susceptibility to damage by normal concentrations of gastric acid. This hypoxia may be due to diminished organ blood flow or submucosal arteriovenous shunting.

The other notable gastrointestinal complication is impaired motility involving the gastrointestinal tract and the biliary system. Acute gastric dilation and intestinal paralytic ileus are commonly seen; they are probably the result of frequent anesthesia, sepsis, fluid overload, and electrolyte imbalances. Delayed gastric emptying and ileus frequently limit the success of enteral alimentation. Acute acalculous cholecystitis is common in these critically ill patients. It usually manifests as sepsis, pain and mass in the right upper quadrant, and abnormalities of liver function. An ultrasonographic examination or radionuclide scan usually supports the diagnosis. Cholecystitis can often be treated by antibiotics plus percutaneous cholecystostomy, or laparoscopic or open cholecystectomy.

Renal

Acute renal failure may be secondary to hypoperfusion and hypoxia occurring before plasma volume was replaced in resuscitation. Failure may also be exacerbated by precipitation of free hemoglobin from damaged red blood cells or muscular myoglobin from crush or electrical injuries; or it may be a result of nephrotoxic drugs, particularly antimicrobial agents, that are administered to these patients. These insults may also be superimposed on pre-existing renal compromise. Oliguric or nonoliguric acute tubular necrosis can result, with the additive attendant clinical problems of acute renal failure in combination with management of the burn injury. Careful attention to intravascular volume will minimize renal dysfunction.

Cardiovascular

Congestive heart failure occurs either in the acute phase of the burn injury or during the mobilization of the peripheral edema. Endocarditis may also complicate burn sepsis and should be kept in mind as an infrequent cause of infection. The use of digitalis and antiarrhythmics may become necessary in specific patients. Rapid atrial fibrillation is a common arrythmia in elderly patients during burn resuscitation.

Neurologic

Burn encephalopathy encompasses a wide range of syndromes of cerebral compromise whose causes include water intoxication, acute hypertension, drug narcosis, septicemia, hyperpyrexia, electrolyte shifts, and dehydration. Autopsy at the endstage of this encephalopathic picture reveals cerebral edema and uncal or cerebellar herniation.

Burn aftercare

The outlook for seriously burned individuals continues to improve. To achieve optimal functional and cosmetic outcomes, and to maximize return to working life and general reintegration requires more than a successful acute stay in hospital. It is ideal if patients who suffer serious burns are enrolled in a coordinated program of burn-specific aftercare; this includes psychologic support, rehabilitation, scar management, and reconstructive surgery. With such participation, the majority of seriously burned patient can enjoy a quality of survival heretofore unprecedented.

Further reading

Ryan CM, Schoenfeld DA, Thorpe WP, Sheridan RL, Cassem EH, Tompkins RG. Objective estimates of the probability of death from burn injuries. *New England Journal of Medicine* 1998; **338**: 362–6. [A predictive formula for burn mortality is described.]

Sheridan RL *et al.* Permissive hypercapnia as a ventilatory strategy in burned children: effect on barotrauma, pneumonia, and mortality. *Journal of Trauma* 1995; **39**: 854–9. [Lung protective ventilation strategy is consistent with high survival rates.]

Sheridan RL, Tompkins RG. Skin substitutes in burns. *Burns* 1999; **25**: 97–103. [A review of the recent trends in skin substitutes.]

Sheridan RL, Baryza MJ, Ryan CM, Schulz JT, Tompkins RG. Acute hand burns in children: management and long term outcome based on a ten year experience with 698 injured hands. *Annals of Surgery* 1999; **229**: 558–64. [Describes an organized approach to serious hand burns and associated long-term results.]

Sheridan RL, Hinson MI, Liang MM, Ryan CM, Mulligan J, Tompkins RG. Long-term outcome of children surviving massive burns. *Journal of the American Medical Association* 2000; **283**: 69–73. [Study documenting that most survivors of massive burns have satisfying long-term outcomes.]

Sheridan RL, Remensnyder JP, Schnitzer JJ, Ryan CM, Schulz JT, Tompkins RG. Current expectations for survival in pediatric burns. *Archives of Pediatrics* 2000 (in press). [Documents improvement in survival among burned children.]

Tompkins RG, Burke JF. Burn wound closure using permanent skin replacement materials. *World Journal of Surgery* 1992; **16**: 47–52. [Skin substitutes are reviewed.]

Tompkins RG *et al.* Prompt eschar excision: a treatment system contributing to reduced burn mortality. A statistical evaluation of burn care at the Massachusetts General Hospital (1974–1984). *Annals of Surgery* 1986; **204**: 272–81. [Early eschar excision reduces mortality in adults.]

Tompkins RG *et al.* Significant reductions in mortality for children with burn injuries through the use of prompt eschar excision. *Annals of Surgery* 1988; **208**: 577–85. [Early eschar excision reduces mortality in children.]

Tompkins RG *et al.* Increased survival after massive thermal injuries in adults: preliminary report using artificial skin. *Critical Care Medicine* 1989; **17**: 734–40. [Artifical skin is useful in management of patients with massive burns.]

51 How should surgical trainees be selected?

Thomas F. Dodson

Conclusions
Further reading

> When a new flight (i.e., a class) of trainees arrived at Pensacola, they were brought into an auditorium for a little lecture. An officer would tell them: ‘Take a look at the man on either side of you.' Quite a few actually swivelled their heads this way and that, in the interest of appearing diligent. Then the officer would say: ‘one of the three of you is not going to make it!'—meaning, not get his wings. That was the opening theme, the *motif* of primary training. We already know that one-third of you do not have the right stuff—it only remains to find out who.
>
> Tom Wolfe, *The Right Stuff*, 1979

All surgical educators, regardless of their country of origin, experience some anxiety about the selection of surgical candidates. It is a labor-intensive process: receiving and organizing applications, letters of recommendation, grades, dean's letters, chairman's letters, and others; filing this material in individual folders; and then the time-consuming process of pouring over each application trying to discern by some word or phrase that candidate who might be ‘right' for your own program. Like so many issues that we face today as we enter the third millennium, it is perhaps instructive to turn back the pages a bit and reflect on the times and the men that brought forth the modern surgical residency.

The surgical resident of today owes his/her position to his European forefathers, specifically those from France and Germany. In writing on the need to retain the ‘surgical generalist' in the early 1970s, Wagensteen (1972) stated that the surgical residency was French in origin, and noted that as early as 1613, five ‘compagnons chirurgiens' (surgical house officers) were housed at the Hotel Dieu, along with a ‘Master Surgeon'. He further noted that by 1625, two of the five surgical house officers were required to sleep in the hospital, closely mirroring the stresses and sleep deprivation of the modern surgical resident! Over time, the number of house officers gradually increased to 12, and after 6 years of service in the hospital, the resident would become a master surgeon and begin his own practice of surgery. Parenthetically, it should be noted that the famous French surgeon, Ambrose Pare, spent 3 years at the Hotel Dieu, a time which preceded his efforts at the siege of Turin when he utilized ‘a digestive' to treat the wounds of soldiers.

It also seems pertinent to remark at this point on the contributions of William Cheselden, a surgeon at St. Thomas' Hospital in London, who was instrumental in improving the educational standards of English surgeons in the late 1700s. This resulted, in 1745, in the ‘separation' of the Surgeons from the Barbers, and later; in 1800, the Surgeons were granted a Royal Charter by George III which created the Royal College of Surgeons. Perhaps of more importance to the English people, Cheselden was a mentor to John Hunter, the famous British surgeon who introduced scientific methods to the study of surgery (Fig. 1).



Fig. 1. John Hunter, student of William Cheseldon. Hunter introduced scientific methods to the study of surgery.

Skip forward several hundred miles and about one hundred years from the formation of a company of surgeons in England, and the surgical residency's next evolutionary step came from Germany. Bernard von Langenbeck was born in 1810, received his medical degree from Gottingen in 1835, and was appointed professor of Surgery at Kiel in 1842. In 1847, he succeeded Dieffenbach in the chair of surgery at Berlin, and has been recognized as ‘the father of modern surgical training programs'. Although he wrote a classical textbook, performed feats of surgical legerdemain such as complete excision of the scapula and repair of cleft palate, and discovered *Candida albicans*, he is remembered primarily as a master teacher and educator. Befitting his role as teacher, his pupils were sought after and called to chairs of surgery throughout Europe. One of his pupils, Theodor Billroth, eventually eclipsed his old master, both in surgical skills and in refinement of the surgical residency program. He, too, attracted men of great promise, and allowed students to begin surgical training only after completion of medical studies and preliminary experience working in a hospital. He felt that operations on cadavers and experimental animals were a necessary prelude for the young surgical trainee, and that 2 to 3 years of assistantship in a surgical department were necessary before independent practice. Billroth's pupils eventually occupied chairs of surgery in Heidelberg, Vienna, Prague, and other centers.

The next individual of importance to the surgical residency program came from an unlikely country. William Stewart Halsted was born in New York City in 1852, the son of a wealthy dry-goods wholesaler (Fig. 2). He graduated from Yale University in 1874 and in 1877 received his medical degree from the College of Physicians and Surgeons in New York. He then did an 18-month internship at Bellevue Hospital where he became interested in surgery, and in 1878, at the urging of his friend, William H. Welch, he left for Europe to study at centers of excellence there.



Fig. 2. William Stewart Halsted. He introduced the German SYstem of surgical training to the United States.

As noted by Sabiston (1987), ‘the development of the surgical residency training program in the United States is clearly related to the Langenbeck-Billroth school…'. While visiting the great centers of Europe, Halsted was particularly impressed with the German system of surgical training. They emphasized integration of the basic sciences with clinical teaching; competition among the surgical trainees with selection for advanced positions going only to ‘the best and the brightest;' and finally, selection of the House Surgeon (or Chief Resident) from the group of ‘assistants'. This ‘first assistant', as the position was then known, would keep his position until a chair of surgery became available or, if tired of waiting, he could then go into practice.

Halsted returned to New York in 1880 and opened a private practice limited to surgery. His old friend, William Welch, was instrumental in bringing Halsted to Hopkins, and when the Johns Hopkins Hospital opened in 1889 and William Macewen of Glasgow refused the position as Chief of Surgery at the new hospital, Halsted was

appointed as ‘acting surgeon' for 1 year. He became Professor of Surgery at Hopkins and surgeon-in-chief of the hospital 3 years later. Under his direction, residents at Hopkins were selected only after careful scrutiny, and were expected to compete for senior slots in the manner of a ‘pyramid'. During the 33 years in which he served as Chief of Surgery at Hopkins, Halsted appointed 17 resident surgeons and 55 men served as assistant resident surgeons during this same period. Many of these individuals went on to become chairmen of departments or were otherwise influential in the early development of the surgical residency in the United States.

Speaking before his friends and colleagues at Yale in 1904, Halsted had this to say on the training of the surgeon:

> It was our intention originally to adopt as closely as feasible the German plan, which, in the main, is the same for all the principal clinics of the German universities. The house surgeon, or first assistant, as he is called in Germany, is selected, after several years of service, from a number of well-tried assistants.
>
> There is no regular advancement from the bottom to the top of the staff of resident assistants. Only a small proportion of these venture to entertain the hope of becoming first assistant…
>
> We need a system, and we shall surely have it, which will produce not only surgeons but surgeons of the highest type, men who will stimulate the first youths of our country to study surgery and to devote their energies and their lives to raising the standard of surgical science.

Given that this text has an international audience, it might be worthwhile to compare and contrast systems of surgical training in Great Britain, Germany, and the United States. To the reader of the foregoing history, it will be no surprise that these countries all share ‘significant similarities.' It is apparent, as noted in Table 1 and Table 2, that there are more similarities than differences between the three countries. Perhaps the most remarkable differences are: (a) fewer years of training in Germany and the United States as compared with Great Britain; (b) less supervision in the British system as compared with the other two countries; and (c) less night call, better pay, and longer vacations in the German SYstem. As the authors note, ‘the question of which country offers the best training in surgery is difficult to answer and depends to some extent on the point of view of the observer.' The following discussion will, by necessity, have an American flavor to it, but the issues are common to all surgery residency programs. How do we pick the best people? How do we pick those people who will endure the stresses and rigors of a surgical residency, perform at a high level with relatively low pay, stay the course, and join their community or academic program with the skills necessary to provide state of the art surgical care?

	Britain	Germany	United States
General surgery residents	1511[a]	1743[c]	7799
Female residents (per cent)	9.2[a]	14.3[c]	13.6
Duration of training (years)	8–20	6–10	5–9
Number of operations	<3000[d]	4500[e]	8995
Ratio, supervised versus unsupervised operations[f]	50:50	90:10	90:10
Night calls per month	10–15	6–8	10–16
Vacation (weeks)	5 plus study leave	6 plus overtime	3–4
Monthly net income (dollars)	2300	3000	3500
Training environment	Cordial	Hierarchy	Powerful work ethic

[a]Figures based on estimates, [b]figures refer to England and Wales, [c]figures refer to West Germany only, [d]individual data, [e]reporting period 1990–91.

Table 1 General surgical training in Britain, Germany, and the United States 1989 to 1990

Item	Surgery	Assessment
Time	Long supervised training, practical experience, theoretical training, specialist training	Continuous evaluation, examinations
Content	Practical and theoretical training, experience	Training goals and their control
Objective	Specialist surgeon	Specialist surgeon

Table 2 Comparative evaluation of the surgical training in Britain, Germany, and the United States

We will examine five areas of interest relative to resident selection: (a) medical school performance; (b) the role of the interviews; (c) letters of recommendation; (d) grades and United States Medical Licensing Examination (USMLE) scores; and (e) the role of psychomotor testing.

There are five studies that look carefully at the relationship of performance in medical school compared with residency performance. Two of these studies were done over a decade ago, and the remaining three were completed in the present decade. Erlandson and his colleagues at the University of Michigan undertook a retrospective study of surgical residents from 1975 to 1981. They sought to identify house officer applicant data that would correlate with performance in the first 2 years of training as measured by faculty evaluation and the American Board of Surgery In-Training Examination (ABSITE) scores. Interestingly, but perhaps not surprisingly, those residents who had been elected to AOA (medical school honorary society) had higher ABSITE scores and clinical evaluation knowledge scores for the first 2 years of their residencies. When interpersonal outcome measures were evaluated, however, *none* of the applicant data had any significance. They also found statistical significance for medical school honors (but not surgery honors) and for house officers who graduated from medical schools with letter or numerical scores (as opposed to a pass/fail grading system). Three years later, Kron *et al.* from Virginia reported on a study of 62 residents from that institution who entered the general surgery residency over a 10-year period. They found that four factors predicted successful completion of their program with an 89 per cent accuracy: AOA, high class rank, clinical honors, and publications. However, *no* subjective or objective data correlated with later chief resident performance.

Taking a somewhat different tack, Brown *et al.*, from the University of California, San Francisco, looked at medical school graduates who performed poorly compared with graduates who performed well. They asked a critical question: ‘Could these poorly received graduates have been recognized before graduation, and perhaps better groomed for their careers?' They concluded that the two groups were actually quite similar in most measurements, and that most of the poorly received graduates had experienced problems which were ‘personal and motivational' rather than skill or knowledge based. They concluded that the medical school records contained ‘little evidence' that might serve to discriminate between residents who performed well versus those who performed poorly. They also expressed the following thoughtful remark:

> It is perhaps more surprising that most residents can function with as much competence and grace as they do than that a few should fail to meet the expectations of their preceptors for one reason or another.

The final two studies were completed in 1997, but provide little clarity to this issue. Papp and colleagues from the University of Louisville asked 23 colleagues to evaluate the general surgery residents who completed their surgical training from 1990 to 1994. As they noted, the analysis of potential residents ‘is based on the assumption that human behavior is somewhat predictable and that the best predictor of future performance is past performance.' Unfortunately, ‘no single predictor of success' during surgical residency could be found in the applications to the program. As they pointed out, most applicants in this country have an abundance of information about their cognitive skills—they have markedly less so about other, perhaps more important, aspects of their qualifications.

In another careful analysis of residents who have finished a general surgery training program, McGreevy and Kollmorgen reviewed the files of 63 University of Utah residents who completed their residency program and of 22 who did not finish. The only success that could be predicted from an analysis of their medical school performance was the ability to take standardized tests. As they pointed out, although there is a general agreement about the need to measure non-cognitive factors, it is still unclear exactly what should be measured and how that should be done.

If the medical school record offers little in the way of solace to the surgical educator, what of the interview? Surely that time-tested method can be counted upon to

provide hints or clues to the good surgical resident. Not so, said Komives *et al.*, from the Beth Israel Hospital in Boston, and they even concluded that the interview may serve more to aid the applicant than to provide helpful information to the interviewer. Two studies, however, suggest that the interview continues to be a major influence on the selection and ranking of prospective applicants. While acknowledging that objectivity, reliability, and validity of the interview are ‘debatable,' the authors state that it remains an important aspect of the selection process. Relative to the interview itself, there continues to be controversy about the manner in which it is conducted. Some authorities have suggested that an interviewer should prepare by reading the applicant's file prior to the interview, while others have suggested that the interview must be conducted ‘blindly' in order to reduce bias.

A careful analysis of letters of recommendation was presented at the Annual Meeting of the Association of Program Directors in Surgery in 1990. At that meeting, Greenburg and colleagues pointed out two features of letters of recommendation that seemed to have ‘positive' characteristics: (a) a desire on the part of the letter writer to want the applicant to stay in his own program, and (b) a sense that the letter writer truly ‘knew' the applicant with a resultant careful and detailed appraisal of his/her skills. Conversely, ‘brevity and generality' come across as negative features, and they create a sense of unease about the writer's interest in or actual knowledge of the applicant. The authors acknowledge that current letters of recommendation have ‘little specificity and therefore little predictive power' with respect to performance in a surgery residency program. In a delightful and somewhat whimsical paper on this topic, Friedman disparaged letters of recommendation as being a ‘fantasy land', and he recommended that we ‘commit truth' in order to improve the process.

In the United States, the information in the typical application has actually been declining in the past two decades owing to the increased emphasis on decreasing competition among medical students and the increased acceptance of the ‘pass/fail' method of grading. This change was noted by Moss and colleagues, and they concluded that surgery residency training programs would be better off if they selected applicants from schools that continue to provide class standings. In a survey of general surgery program directors, Dietrick *et al.* noted that nearly 90 per cent of respondents preferred to evaluate transcripts that use grades rather than pass/fail analyses. Because of this lack of objective evidence of academic performance, some program directors are turning to the USMLE in an attempt to circumvent the lack of grades. In one study that looks directly at this issue, the authors) made the following comments:

> … the use of the USMLE results as the sole criterion for final selection is not justified. Because clinical performance is multidimensional and because cognitive knowledge, as measured by the USMLE, tends to represent only one facet, satisfactory performance on the USMLE is a necessary but not sufficient condition for adequate clinical performance.

Psychomotor testing has received some interest, and was explored in a symposium held at the Royal College of Surgeons in 1987. The opening remarks were made by Mr Malcolm Gough, who noted the difficulties inherent in using previous academic achievement, interviews, and letters of recommendation as a basis from which to select surgical residents. He noted that ‘more formal personnel assessment and aptitude tests' were available, and that they had been adopted by both industry and some branches of the military. Candidate selection for surgical training in the Netherlands was also reviewed, and the process was noted to involve three separate procedures: (a) completion of a 12-page registration form; (b) psychotechnical testing; and (c) a panel interview with five surgeons interviewing each applicant for 40 min. The initial results were stated to be ‘effective' at reducing the loss of surgeons from the training programs with quality and aptitude remaining ‘very satisfactory'. However, in a recent presidential address to members of the Society of Clinical Vascular Surgery, John D. Corson, MBChB, FRCS (Erg), FACS, stated that, despite initial enthusiasm by the Dutch surgeons, the use of aptitude tests had been ‘abandoned'.

An excellent overview of this entire process is provided to us in the textbook, *Medical education: a surgical perspective*. In this book, which is a compilation of talks given during a conference on surgical education at the University of Michigan in 1986, leading educators from the Department of Surgery at Southern Illinois University discoursed on the topic of ‘the resident selection process'. Their conclusions are provocative and consonant with the findings of this chapter:

> medical school record—‘membership in AOA (medical school honorary society) did predict excellence during the residency years'
>
> medical school grades—grades are only indicative of ‘abilities to recall information and do not reflect attitudes, motivation, and sincerity'
>
> NBME/USMLE scores—‘NBME scores have no consistent correlation with residency performance'
>
> letters of recommendation—‘ineffective in providing a true appraisal of the student's strengths and weaknesses'
>
> interviews—‘faculty ratings of applicants based on an interview had no significant predictive value'
>
> psychomotor testing—‘it would seem at present unnecessary to include tests of personality or psychomotor function as a way to discriminate among prospective surgical residents'

Given the paucity of truly predictive information, what are programs in our country doing at the present time? In a recent look at two university programs, the University of Mississippi and the University of Iowa, Scott-Conner and colleagues looked at the actual factors that played a part in the selection of surgical residents. Both of these programs placed a significant emphasis on the importance of the interview, although they acknowledged the subjective nature of the process. The authors also noted that AOA membership and USMLE scores seemed to have added weight in the present climate with its absence of ‘consistent grading systems'. And they concluded with a lament about the ‘lack of objective information that correlates with performance', but with no suggestions for improvement.

Conclusions

Probably no issue takes up more of the surgical educator's time than the issue of residency selection. It is somewhat disheartening to note that there are so few ‘hard data' on which to base such important decisions, but as noted by Dean, ‘few would argue with the results'. Good men and women continue to apply to programs in surgery in numbers greater than positions available, and the results continue to generate pride from their respective surgical educators. Further studies are obviously needed to make the entire process of residency selection a more objective undertaking.

Further reading

Bartlett RH, Zelenock GB, Strodel WE, Harper NL, Turcotte JG, eds. *Medical education: a surgical perspective*. Lewis Publishers, Chelsea, Michigan, 1986: 313–31. [One of the few texts available about surgical education based on a conference held at the University of Michigan in 1986.]

Berner ES, Brooks CM, Erdmann JB. Use of the USMLE to select residents. *Academic Medicine* 1993; **68**: 753–79. [This article suggests that the USMLE shows a ‘generally positive relationship' between this examination and subsequent clinical performance.]

Brown E, Rosinski EF, Altman DF. Comparing medical school graduates who perform poorly in residency with graduates who perform well. *Academic Medicine* 1993; **68**: 806–8. [The medical school records contained little evidence that might have predicted the graduates' poor reception as residents.]

Buckman RF. The surgical principles of John Hunter. *Surgery, Gynecology and Obstetrics* 1987; **164**: 479–84. [John Hunter, the ‘'founder of scientific surgery', trained under William Cheselden, an important figure in improving the educational standards of English surgeons in the late 1700s.]

Carter BN. The fruition of Halsted's concept of surgical training. *Surgery* 1952; **32**: 518–26. [William Halsted brought an ‘approximation' of the German approach to Johns Hopkins Hospital, and developed the foremost surgical residency of his day. His influence lives on in the lives of the third generation of Halstedians.]

Corson JD. Passing the torch. *American Joumal of Surgery* 1997; **174**: 96–101. [This article comments on computers, telemedicine, the internet, virtual reality and other ‘new technologies,' but concludes with the statement that ‘attention to detail' and ‘the ability to go the extra mile' will still be important attributes of the excellent surgeon.]

Dean RE. Selection of candidates for surgical residency. *Problems in General Surgery* 1986; **3**: 375–82. [Dean and his colleagues from Michigan State University present their ‘logical and pragmatic method' for selecting residents.]

Dean RE, Dean KB, Nicholas WR, Scholten DJ. The interviewing process as it relates to the selection of candidates for general surgical residency programs. *Current Surgery* 1987; **44**: 1–6. [The results of a survey of general surgery residency program directors in 1985–1986. It concluded that the applicant interview was the most important factor in the ranking of candidates for surgical residencies.]

Dietrick JA, Weaver MT, Merrick HW. Pass/fail grading: A disadvantage for students applying for residency. *American Journal of Surgery* 1991; **162**: 63–66. [The article suggests that the current use of pass/fail grading puts the medical student at a disadvantage in competing for general surgery residency positions.]

Erlandson EE *et al.* Resident selection: Applicant selection criteria compared with performance. *Surgery* 1982; **92**: 270–5. [A retrospective study of applicant data from 1975 to 1981. Certain data did correlate with performance in terms of cognitive knowledge, but interpersonal skills were impossible to predict.]

Friedman RB. Fantasy land. *New England Journal of Medicine* 1983; **308**: 651–3. [A wry and somewhat disgruntled plea to ‘commit truth' when writing letters of recommendation.]

Gong, Jr. H, Parker NH, Apgar FA, Shank C. Influence of the interview on ranking in the residency selection process. *Medical Education* 1984; **18**: 366–9. [This study shows that the interview can have a ‘major influence' on the interviewer and his or her ranking of student applicants.]

Gough M. Personality assessment techniques and aptitude testing as aids to the selection of surgical trainees. *Symposium of the Royal College of Surgeons* 1988; **70**: 265–6. [The opening remarks of the chairman of a symposium on the possible role of formal personnel selection techniques and aptitude testing.]

Greenburg AG, Doyle J, McClure DK. Letters of recommendation for surgical residencies: What they say and what they mean. *Journal of Surgical Research* 1994; **56**: 192–8. [This paper suggests that letters of recommendation have little specificity and therefore little predictive power with respect to screening and selection of candidates for a surgical residency.]

Halsted WS. The training of the surgeon. *Bulletin of the Johns Hopkins Hospital* 1904; **15**: 267–75. [This paper summarizes the thoughts of the leading American surgeon of his day on how residents should be trained. As he noted, it was his intention to adopt ‘as closely as feasible' the German plan as his model.]

John K-D, Modlin IM. A brief historical perspective and a comparison of the current systems of surgical training in Great Britain, Germany and the United States of America. *Surgery, Gynecology and Obstetrics* 1993; **177**: 622–32. [A fascinating look at the similarities and differences between Britain, Germany, and the United States. It suggests that we are in a ‘new period' of transition, and that education will necessarily be modified in the future as surgeons adjust to these changes.]

Keeman JN, Lagaay MB. Candidate selection for surgical training in The Netherlands. *Symposium of the Royal College of Surgeons* 1988; **88**: 275–7. [The results of personnel selection consultants on behalf of the Association of Surgeons of the Netherlands. They concluded that the new procedure was ‘an effective means' of reducing the number of new surgeons leaving the training program.]

Komives E, Weiss ST, Rosa RM. The applicant interview as a predictor of resident performance. *Journal of Medical Education* 1984; **59**: 425–6. [The data suggested that the residency interview was of greater value to applicants than to house staff selection committees.]

Kron IL, Kaiser DL, Nolan SP, Rudolf LE, Muller, Jr. WH, Jones RS. Can success in the surgical residency be predicted from preresidency evaluation? *Annals of Surgery* 1985; **202**: 694–5. [They were able to predict the likelihood of successful completion of the residency program with 89 per cent accuracy based on membership in ADA, achievement of honors, high class rank, and previous publications.]

McGreevy JM, Kollmorgen DR. Medical school performance does not predict success in a surgery residency. *Current Surgery* 1997; **54**: 438–41. [Performance in medical school as measured by traditional indicators does not correlate with performance in a surgical residency.]

Moss TJ, Deland EC, Maloney JV, Jr. Selection of medical students for graduate training: Pass/fail versus grades. *New England Journal of Medicine* 1978; **299**: 25–7. [This article suggests that a residency program that ‘seeks excellence' would be advised to select applicants from schools that provide a specific class standing as part of the total evaluation of the student.]

Papp KK, Polk, HC, Jr, Richardson JD. The relationship between criteria used to select residents and performance during residency. *American Journal of Surgery* 1997; **173**: 326–9. [The authors were unable to find any single predictor of of success during surgical residency in applicants' admission folders.]

Robin AP, Bombeck CT, Pollack R, Nyhus LM. Introduction of bias in residencycandidate interviews. *Surgery* 1991; **110**: 253–8. [This paper concludes that interviews should be conducted in a ‘blinded' fashion to reduce bias by candidate information.]

Robinson JO. The barber-surgeons of London. *Archives of Surgery* 1984; **119**: 1171–5. [One of the great milestones in English surgery was the union of the Company of Barbers and the Fellowship of Surgeons in 1540.Interestingly, the Royal Barbers were held in greater standing than the Surgeons at this point in time.]

Roses DF. Profile: William Stewart Halsted, MD. *Traditions in Surgery* 1994; **1**: 2–4. [A brief history of the life and times of the important American surgeon and educator.]

Rutkow IM. William Stewart Halsted and the Germanic influence on education and training programs in surgery. *Surgery, Gynecology and Obstetrics* 1978; **147**: 602–6. [The story of how the German system became the model for residency training in the United States.]

Rutkow IM. *Surgery: an illustrated history*, pp. 263, 393. Mosby, St. Louis, 1993. [Five thousand years of surgical history presented in pictures and words.]

Rutkow IM, Imbembo AL, Suidema GD. The application and interviewing process for surgical house officership. *Surgery* 1979; **85**: 184–90. [The authors make several suggestions about making the interview process more effective.]

Sabiston DC, Jr. Milestones in surgery. In: Sabiston DC, Jr, ed. *Essentials of surgery*, pp. 1–9. WB Saunders, Philadelphia, 1987. [In this text, Bernhard von Langenbeck 1s credited with being ‘the father of modern surgical training programs'.]

Scott-Conner CER Rhodes RS, Aggarwal AK, Bass JD. Selecting residents for a surgical residency: Are we doing what we say we are doing? *Current Surgery* 1997: **54**: 42–6. [This article notes that decreasing availability of objective information has been counterbalanced in the United States by increasing reliance on USMLE scores. She calls the present SYSTEM ‘chaotic'.]

Wangensteen OH. Historical aspects of some surgical training programs: A plea to retain the surgical generalist in the academic arena. *Surgery* 1972; **72**: 692–8. [This paper suggests that the surgical residency is French in origin, and that it dates back to 1613 with five surgical house officers being lodged and fed at the Hôtel Dieu.]